H.1 Exercises on Chapter 1

1. A card came is played with 52 cards divided equally between four players, North, South, East and West, all arrangements being equally likely. Thirteen of the cards are referred to as trumps. If you know that North and South have ten trumps between them, what is the probability that all three remaining trumps are in the same hand? If it is known that the king of trumps is included among the other three, what is the probability that one player has the king and the other the remaining two trumps?

2. (a) Under what circumstances is an event A independent of itself?

(b) By considering events concerned with independent tosses of a red die and a blue die, or otherwise. give examples of events A, B and C which are not independent, but nevertheless are such that every pair of them is independent.

(c) By considering events concerned with three independent tosses of a coin and supposing that A and B both represent tossing a head on the first trial, give examples of events A, B and C which are such that $\mathsf{P}(ABC) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C)$ although no pair of them is independent.

3. Whether certain mice are black or brown depends on a pair of genes, each of which is either B or b. If both members of the pair are alike, the mouse is said to be homozygous, and if they are different it is said to be heterozygous. The mouse is brown only if it is homozygous bb. The offspring of a pair of mice have two such genes, one from each parent, and if the parent is heterozygous, the inherited gene is equally likely to be B or b. Suppose that a black mouse results from a mating between two heterozygotes.

(a) What are the probabilities that this mouse is homozygous and that it is heterozygous?

Now suppose that this mouse is mated with a brown mouse, resulting in seven offspring, all of which turn out to be black.

(b) Use Bayes' Theorem to find the probability that the black mouse was homozygous BB.

(c) Recalculate the same probability by regarding the seven offspring as seven observations made sequentially, treating the posterior after each observation as the prior for the next (cf. Fisher, 1959, Section II.2).

4. The example on Bayes' Theorem in Section 1.2 concerning the biology of twins was based on the assumption that births of boys and girls occur equally frequently, and yet it has been known for a very long time that fewer girls are born than boys (cf. Arbuthnot, 1710). Suppose that the probability of a girl is p, so that

$$\begin{array}{lll} \mathsf{P}(GG|M) = p, & \mathsf{P}(BB|M) = 1-p, & \mathsf{P}(GB|M) = 0, \\ \mathsf{P}(GG|D) = p^2, & \mathsf{P}(BB|D) = (1-p)^2, & \mathsf{P}(GB|D) = 2p(1-p). \end{array}$$

Find the proportion of monozygotic twins in the whole population of twins in terms of p and the sex distribution among all twins.

5. Suppose a red and a blue die are tossed. Let x be the sum of the number showing on the red die and twice the number showing on the blue die. Find the density function and the distribution function of x.

6. Suppose that $k \sim \mathrm{B}(n, \pi)$ where n is large and π is small but $n\pi = \lambda$ has an intermediate value. Use the exponential limit $(1 + x/n)^n \to \mathrm{e}^x$ to show that $\mathsf{P}(k = 0) \cong \mathrm{e}^{-\lambda}$ and $\mathsf{P}(k = 1) \cong \lambda\mathrm{e}^{-\lambda}$. Extend this result to show that k is such that

$$p(k) \cong \frac{\lambda^k}{k!} \exp(-\lambda)$$

that is, k is approximately distributed as a Poisson variable of mean λ (cf. Appendix A).

7. Suppose that m and n have independent Poisson distributions of means λ and μ respectively (see question 6 and that $k = m+n$. has a chi-squared density on one degree of freedom as defined in Appendix A.

8. Modify the formula for the density of a one-to-one funtion $g(x)$ of a random variable x to find an expression for the density of x^2 in terms of that of x, in both the continuous and discrete case. Hence show that the square of a standard normal distribution has a chi-squared distribution on one degree of freedom as defined in Appendix A.

9. Suppose that $x_1, x_2, \dots, x_n$ are independently and all have the same continuous distribution, with density $f(x)$ and distribution function $F(x)$. Find the distribution functions of

$$M = \max\{x_1, x_2, \dots, x_n\} \quad \text{and} \quad m = \min\{x_1, x_2, \dots, x_n\}$$

in terms of $F(x)$, and so find expressions for the density functions of M and m.

10. Suppose that u and v are independently uniformly distributed on the interval [0, 1], so that the divide the interval into three sub-intervals. Find the

joint density function of the lengths of the first two sub-intervals.

11. Show that two continuous random variables x and y are independent (that is, $p(x,y) = p(x)p(y)$ for all x and y) if and only if their joint distribution function $F(x,y)$ satisfies $F(x,y) = F(x)F(y)$ for all x and y. Prove that the same thing is true for discrete random variables. [This is an example of a result which is easier to prove in the continuous case.]

12. Suppose that the random variable x has a negative binomial distribution $\mathrm{NB}(n,\pi)$ of index n and parameter π, so that

$$p(x) = \binom{n+x-1}{x}\pi^n(1-\pi)^x$$

Find the mean and variance of x and check that your answer agrees with that given in Appendix A.

13. A random variable X is said to have a chi-squared distribution on ν degrees of freedom if it has the same distribution as

$$Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$$

where Z_1, Z_2, ..., Z_ν are independent standard normal variates. Use the facts that $\mathsf{E}Z_i = 0$, $\mathsf{E}Z_i^2 = 1$ and $\mathsf{E}Z_i^4 = 3$ to find the mean and variance of X. Confirm these values using the probability density of X, which is

$$p(X) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} X^{\nu/2-1}\exp(-\tfrac{1}{2}X) \qquad (0 < X < \infty)$$

(see Appendix A).

14. The *skewness* of a random variable x is defined as $\gamma_1 = \mu_3/(\mu_2)^{\frac{3}{2}}$ where

$$\mu_n = \mathsf{E}(x - \mathsf{E}x)^n$$

(but note that some authors work in terms of $\beta_1 = \gamma_1^2$). Find the skewness of a random variable X with a binomial distribution $B(n,\pi)$ of index n and parameter π.

15. Suppose that a continuous random variable X has mean μ and variance ϕ. By writing

$$\phi = \int (x-\mu)^2 p(x)\,\mathrm{d}x \geqslant \int_{\{x;\,|x-\mu|\geqslant c\}} (x-\mu)^2 p(x)\,\mathrm{d}x$$

and using a lower bound for the integrand in the latter integral, prove that

$$\mathsf{P}(|x-\mu| \geqslant c) \leqslant \frac{\phi}{c^2}.$$

Show that the result also holds for discrete random variables. [This result is known as Čebyšev's Inequality (the name is spelt in many other ways, including Chebyshev and Tchebycheff).]

16. Suppose that x and y are such that

$$\mathsf{P}(x=0, y=1) = \mathsf{P}(x=0, y=-1) = \mathsf{P}(x=1, y=0) = \mathsf{P}(x=-1, y=0) = \tfrac{1}{4}.$$

Show that x and y are uncorrelated but that they are *not* independent.

17. Let x and y have a bivariate normal distribution and suppose that x and y both have mean 0 and variance 1, so that their marginal distributions are standard normal and their joint density is

$$p(x,y) = \left\{2\pi\sqrt{(1-\rho^2)}\right\}^{-1} \exp\left\{-\tfrac{1}{2}(x^2 - 2\rho xy + y^2)/(1-\rho^2)\right\}.$$

Show that if the correlation coefficient between x and y is ρ, then that between x^2 and y^2 is ρ^2.

18. Suppose that x has a Poisson distribution (see question 6) $\mathrm{P}(\lambda)$ of mean λ and that, for given x, y has a binomial distribution $\mathrm{B}(x,\pi)$ of index x and parameter π.

(a) Show that the unconditional distribution of y is Poisson of mean

$$\lambda\pi = \mathsf{E}_{\widetilde{x}}\mathsf{E}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x}).$$

(b) Verify that the formula

$$\mathcal{V}\,\widetilde{y} = \mathsf{E}_{\widetilde{x}}\mathcal{V}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x}) + \mathcal{V}_{\widetilde{x}}\mathsf{E}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x})$$

derived in Section 1.5 holds in this case.

19. Define

$$I = \int_0^\infty \exp(-\tfrac{1}{2}z^2)\,dz$$

and show (by setting $z = xy$ and then substituting z for y) that

$$I = \int_0^\infty \exp(-\tfrac{1}{2}(xy)^2)\,y\,dx = \int_0^\infty \exp(-\tfrac{1}{2}(zx)^2)\,z\,dx.$$

Deduce that

$$I^2 = \int_0^\infty \int_0^\infty \exp\{-\tfrac{1}{2}(x^2+1)z^2\}\, z\, dz\, dx.$$

By substituting $(1+x^2)z^2 = 2t$ so that $z\,dz = dt/(1+x^2)$ show that $I = \sqrt{\pi/2}$ so that the density of the standard normal distribution as defined in Section 1.3 does integrate to unity and so is indeed a density. (This method is due to Laplace, 1812, Section 24.)

H.2 Exercises on Chapter 2

1. Suppose that $k \sim \mathrm{B}(n, \pi)$. Find the standardized likelihood as a function of π for given k. Which of the distributions listed in Appendix A does this represent?

2. Suppose we are given the twelve observations from a normal distribution:

15.644,	16.437,	17.287,	14.448,	15.308,	15.169,
18.123,	17.635,	17.259,	16.311,	15.390,	17.252.

and we are told that the variance $\phi = 1$. Find a 90% HDR for the posterior distribution of the mean assuming the usual reference prior.

3. With the same data as in the previous question, what is the predictive distribution for a possible future observation x?

4. A random sample of size n is to be taken from an $\mathrm{N}(\theta, \phi)$ distribution where ϕ is known. How large must n be to reduce the posterior variance of ϕ to the fraction ϕ/k of its original value (where $k > 1$)?

5. Your prior beliefs about a quantity θ are such that

$$p(\theta) = \begin{cases} 1 & (\theta \geqslant 0) \\ 0 & (\theta < 0). \end{cases}$$

A random sample of size 25 is taken from an $\mathrm{N}(\theta, 1)$ distribution and the mean of the observations is observed to be 0.33. Find a 95% HDR for θ.

6. Suppose that you have prior beliefs about an unknown quantity θ which can be approximated by an $\mathrm{N}(\lambda, \phi)$ distribution, while my beliefs can be approximated by an $\mathrm{N}(\mu, \psi)$ distribution. Suppose further that the reasons that have led us to these conclusions do not overlap with one another. What distribution should represent our beliefs about θ when we take into account all the information available to both of us?

7. Prove the theorem quoted without proof in Section 2.4.

8. Under what circumstances can a likelihood arising from a distribution in the exponential family be expressed in data translated form?

9. Suppose that you are interested in investigating how variable the performance of schoolchildren on a new mathematics test, and that you begin by trying this test out on children in twelve similar schools. It turns out that the average standard deviation is about 10 marks. You then want to try the test on a thirteenth school, which is fairly similar to those you have already investigated, and you reckon that the data on the other schools gives you a prior for the variance in this new school which has a mean of 100 and is worth 8 direct observations on the school. What is the posterior distribution for the variance if you then observe a sample of size 30 from the school of which the standard deviation is 13.2? Give an interval in which the variance lies with 90% posterior probability.

10. The following are the dried weights of a number of plants (in grammes) from a batch of seeds:

$$4.17,\ 5.58,\ 5.18,\ 6.11,\ 4.50,\ 4.61,\ 5.17,\ 4.53,\ 5.33,\ 5.14.$$

Give 90% HDRs for the mean and variance of the population from which they come.

11. Find a sufficient statistic for μ given an n-sample $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ from the exponential distribution

$$p(x|\mu) = \mu^{-1}\exp(-x/\mu) \qquad (0 < x < \infty)$$

where the parameter μ can take any value in $0 < \mu < \infty$.

12. Find a (two-dimensional) sufficient statistic for (α, β) given an n-sample $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ from the two-parameter gamma distribution

$$p(x|\alpha,\beta) = \{\beta^\alpha\Gamma(\alpha)\}^{-1}x^{\alpha-1}\exp(-x/\beta) \qquad (0 < x < \infty)$$

where the parameters α and β can take any values in $0 < \alpha < \infty$, $0 < \beta < \infty$.

13. Find a family of conjugate priors for the likelihood $l(\beta|x) = p(x\,|\,\alpha,\beta)$ where $p(x\,|\,\alpha,\beta)$ is as in the previous question, but α is known.

14. Show that the tangent of a random angle (that is, one which is uniformly distributed on $[0, 2\pi)$) has a Cauchy distribution $\mathrm{C}(0, 1)$.

15. Suppose that the vector $\boldsymbol{x} = (x, y, z)$ has a trinomial distribution depending on the index n and the parameter $\boldsymbol{\pi} = (\pi, \rho, \sigma)$ where $\pi + \rho + \sigma = 1$, that is

$$p(x|\boldsymbol{\pi}) = \frac{n!}{x!\,y!\,z!}\pi^x \rho^y \sigma^z \qquad (x + y + z = n).$$

Show that this distribution is in the two-parameter exponential family.

16. Suppose that the results of a certain test are known, on the basis of general theory, to be normally distributed about the same mean μ with the same variance ϕ, neither of which is known. Suppose further that your prior beliefs about (μ, ϕ) can be represented by a normal/chi-squared distribution with

$$\nu_0 = 4, \qquad S_0 = 350, \qquad n_0 = 1, \qquad \theta_0 = 85.$$

Now suppose that 100 observations are obtained from the population with mean 89 and sample variance $s^2 = 30$. Find the posterior distribution of (μ, ϕ). Compare 50% prior and posterior HDRs for μ.

17. Suppose that your prior for θ is a $\frac{2}{3} : \frac{1}{3}$ mixture of $\mathrm{N}(0, 1)$ and $\mathrm{N}(1, 1)$ and that a single observation $x \sim \mathrm{N}(\theta, 1)$ turns out to equal 2. What is your posterior probability that $\theta > 1$?

18. Establish the formula

$$(n_0^{-1} + n^{-1})^{-1}(\overline{x} - \theta_0)^2 = n\overline{x}^2 + n_0\theta_0^2 - n_1\theta_1^2$$

where $n_1 = n_0 + n$ and $\theta_1 = (n_0\theta_0 + n\overline{x})/n_1$, which was quoted in Section 2.13 as providing a formula for the parameter S_1 of the posterior distribution in the case where both mean and variance are unknown which is less susceptible to rounding errors.

H.3 Exercises on Chapter 3

1. Laplace claimed that the probability that an event which has occurred n times, and has not hitherto failed, will occur again is $(n+1)/(n+2)$ [see Laplace (1774)], which is sometimes known as *Laplace's rule of succession*. Suggest grounds for this assertion.

2. Find a suitable interval of 90% posterior probability to quote in a case when your posterior distribution for an unknown parameter π is $\mathrm{Be}(20, 12)$, and compare this interval with similar intervals for the cases of $\mathrm{Be}(20.5, 12.5)$ and $\mathrm{Be}(21, 13)$ posteriors. Comment on the relevance of the results to the choice of a reference prior for the binomial distribution.

3. Suppose that your prior beliefs about the probability π of success in Bernoulli trials have mean $1/3$ and variance $1/32$. Give a 95% posterior HDR for π given that you have observed 8 successes in 20 trials.

4. Suppose that you have a prior distribution for the probability π of success in a certain kind of gambling game which has mean 0.4, and that you regard your prior information as equivalent to 12 trials. You then play the game 25 times and win 12 times. What is your posterior distribution for π?

5. Suppose that you are interested in the proportion of females in a certain organisation and that as a first step in your investigation you intend to find out the sex of the first 11 members on the membership list. Before doing so, you have prior beliefs which you regard as equivalent to 25% of this data, and your prior beliefs suggest that a third of the membership is female.

Suggest a suitable prior distribution and find its standard deviation.

Suppose that 3 of the first 11 members turn out to be female; find your posterior distribution and give a 50% posterior HDR for this distribution.

Find the mean, median and mode of the posterior distribution.

Would it surprise you to learn that in fact 86 of the total number of 433 members are female?

6. Show that if $g(x) = \sinh^{-1}\sqrt{(x/n)}$ then

$$g'(x) = \tfrac{1}{2}n^{-1}[(x/n)\{1 + (x/n)\}]^{-\frac{1}{2}}.$$

Deduce that if $x \sim \mathrm{NB}(n, \pi)$ has a negative binomial distribution of index n and parameter π and $z = g(x)$ then $\mathsf{E}z \cong \sinh^{-1}\sqrt{(x/n)}$ and $\mathcal{V}z \cong 1/4n$. What does this suggest as a reference prior for π?

7. The following data were collected by von Bortkiewicz (1898) on the number of men killed by a horse in certain Prussian army corps in twenty years, the unit being one army corps for one year:

Number of deaths:	0	1	2	3	4	5 and more
Number of units:	144	91	32	11	2	0.

Give an interval in which the mean number λ of such deaths in a particular army corps in a particular year lies with 95% probability.

8. Recalculate the answer to the previous question assuming that you had a prior distribution for λ of mean 0.66 and standard deviation 0.115.

9. Find the Jeffreys prior for the parameter α of the Maxwell distribution

$$p(x|\alpha) = \sqrt{\frac{2}{\pi}}\alpha^{3/2}x^2\exp(-\tfrac{1}{2}\alpha x^2)$$

and find a transformation of this parameter in which the corresponding prior is uniform.

10. Use the two-dimensional version of Jeffreys' rule to determine a prior for the trinomial distribution

$$p(x, y, z|\pi, \rho) \propto \pi^x \rho^y (1 - \pi - \rho)^z.$$

(cf. question 15 on Chapter 2).

11. Suppose that x has a Pareto distribution $\text{Pa}(\xi, \gamma)$ where ξ is known but γ is unknown, that is,

$$p(x|\gamma) = \gamma\xi^\gamma x^{-\gamma-1} I_{(\xi,\infty)}(x).$$

Use Jeffreys' rule to find a suitable reference prior for γ.

12. Consider a uniform distribution with $\gamma = 2$. How large a random sample must be taken from the uniform distribution in order that the coefficient of variation (that is, the standard deviation divided by the mean) of the length $\beta - \alpha$ of the interval should be reduced to 0.01 or less?

13. Suppose that observations $x_1, x_2, \ldots, x_n$ are available from a density

$$p(x|\theta) = (c + 1)\theta^{-(c+1)} x^c \qquad (0 < x < \theta).$$

Explain how you would make inferences about the parameter θ using a conjugate prior.

14. What could you conclude if you observed *two* tramcars numbered, say, 71 and 100?

15. In Section 3.8 we discussed Newcomb's observation that the front pages of a well-used table of logarithms tend to get

dirtier than the back pages do. What if we had an *antilogarithm* table, that is, a table giving the value of x when $\log_{10} x$ is given? Which pages of such a table would be the dirtiest?

16. We sometimes investigate distributions on a circle (for example, von Mises' distribution which is discussed in Section 3.9 on "The circular normal distribution"). Find a Haar prior for a location parameter on the circle (such as μ in the case of von Mises' distribution).

17. Suppose that the prior distribution $p(\mu, \sigma)$ for the parameters μ and σ of

a Cauchy distribution

$$p(x|\mu, \sigma) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x - \mu)^2}$$

is uniform in μ and σ, and that two observations $x_1 = 2$ and $x_2 = 6$ are available from this distribution. Calculate the value of the posterior density $p(\mu, \sigma|\boldsymbol{x})$ (ignoring the factor $1/\pi^2$) to two decimal places for $\mu = 0, 2, 4, 6, 8$ and $\sigma = 1, 2, 3, 4, 5$. Use Simpson's rule to approximate the posterior marginal density of μ, and hence go on to find an approximation to the posterior probability that $3 < \mu < 5$.

18. Show that if the log-likelihood $L(\theta|x)$ is a concave function of θ for each scalar x (that is, $L''(\theta|x) \leqslant 0$ for all θ), then the likelihood function $L(\theta|\boldsymbol{x})$ for θ given an n-sample $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ has a unique maximum. Prove that this is the case if the observations x_i come from a logistic density

$$p(\boldsymbol{x}|\theta) = \exp(\theta - x)/\{1 + \exp(\theta - x)\}^2 \qquad (-\infty < x < \infty)$$

where θ is an unknown real parameter. Fill in the details of the Newton-Raphson method and the method of scoring for finding the position of the maximum, and suggest a suitable starting point for the algorithms.

[In many applications of Gibbs sampling, which we consider later in Section 9.4, all full conditional densities are log-concave (see Gilks *et al.*, 1996, Section 5.3.3), so the study of such densities is of real interest.]

19. Show that if an experiment consists of two observations, then the total information it provides is the information provided by one observation plus the mean amount provided by the second given the first.

20. Find the entropy $H\{p(\theta)\}$ of a (negative) exponential distribution with density $p(\theta) = \beta^{-1} \exp(-\theta/\beta)$.

H.4 Exercises on Chapter 4

1. Show that if the prior probability π_0 of a hypothesis is close to unity, then the posterior probability p_0 satisfies $1 - p_0 \cong (1 - \pi_0)B^{-1}$ and more exactly $1 - p_0 \cong (1 - \pi_0)B^{-1} + (1 - \pi_0)^2(B^{-1} - B^{-2})$.

2. Watkins (1986, Section 13.3) reports that theory predicted the existence of a Z particle of mass 93.3 ± 0.9 GeV, while first experimental results showed its

mass to be 93.0 ± 1.8 GeV. Find the prior and posterior odds and the Bayes ratio for the hypothesis that its mass is less than 93.0 GeV.

3. An experimental station wishes to test whether a growth hormone will increase the yield of wheat above the average value of 100 units per plot produced under currently standard conditions. Twelve plots treated with the hormone give the yields:

140, 103, 73, 171, 137, 91, 81, 157, 146, 69, 121, 134.

Find the P-value for the hypothesis under consideration.

4. In a genetic experiment, theory predicts that if two genes are on different chromosomes, then the probability of a certain event will be 3/16. In an actual trial, the event occurs 56 times in 300. Use Lindley's method to decide whether there is enough evidence to reject the hypothesis that the genes are on the same chromosome.

5. With the data in the example in Section 3.4 on "The Poisson distribution", would it be appropriate to reject the hypothesis that the true mean equalled the prior mean (that is, that $\lambda = 3$). [Use Lindley's method.]

6. Suppose that the standard test statistic $z = (\overline{x} - \theta_0)/\sqrt{(\phi/n)}$ takes the value $z = 2.5$ and that the sample size is $n = 100$. How close to θ_0 does a value of θ have to be for the value of the normal likelihood function at $\overline{x}$ to be within 10% of its value at $\theta = \theta_0$?

7. Show that the Bayes factor for a test of a point null hypothesis for the normal distribution (where the prior under the alternative hypothesis is also normal) can be expanded in a power series in $\lambda = \phi/n\psi$ as

$$B = \lambda^{-\frac{1}{2}} \exp(-\tfrac{1}{2}z^2)\{1 + \tfrac{1}{2}\lambda(z^2+1) + \dots\}.$$

8. Suppose that $x_1, x_2, \dots, x_n \sim \mathrm{N}(0, \phi)$. Show over the interval $(\phi - \varepsilon, \phi + \varepsilon)$ the likelihood varies by a factor of approximately

$$\exp\left\{\frac{\varepsilon}{\phi}\left(\frac{\sum x_i^2/n}{\phi} - 1\right)\right\}.$$

9. At the beginning of Section 4.5, we saw that under the alternative hypothesis that $\theta \sim \mathrm{N}(\theta_0, \psi)$ the predictive density for $\overline{x}$ was $\mathrm{N}(\theta_0, \psi + \phi/n)$, so that

$$p_1(\overline{x}) = \{2\pi(\psi + \phi/n)\}^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\overline{x} - \theta_0)^2/(\psi + \phi/n)]$$

Show that a maximum of this density considered as a function of ψ occurs when $\psi = (z^2 - 1)\phi/n$, which gives a possible value for ψ if $z \geqslant 1$. Hence show that if $z \geqslant 1$ then for any such alternative hypothesis the Bayes factor satisfies

$$B \geqslant \sqrt{\mathrm{e}}\, z \exp(-\tfrac{1}{2}z^2)$$

and deduce a bound for p_0 (depending on the value of π_0).

10. In the situation discussed in Section 4.5, for a given P-value (so equivalently for a given z) and assuming that $\phi{=}\psi$, at what value of n is the posterior probability of the null hypothesis a minimum.

11. Mendel (1865) reported finding 1850 angular wrinkled seeds to 5474 round or roundish in an experiment in which his theory predicted a ratio of 1 : 3. Use the method employed for Weldon's dice data in Section 4.5 to test whether his theory is confirmed by the data. [However, Fisher (1936) cast some doubt on the genuineness of the data.]

12. A window is broken in forcing entry to a house. The refractive index of a piece of glass found at the scene of the crime is x, which is supposed $\mathrm{N}(\theta_1, \phi)$. The refractive index of a piece of glass found on a suspect is y, which is supposed $N(\theta_2, \phi)$. In the process of establishing the guilt or innocence of the suspect, we are interested in investigating whether $\mathrm{H}_0 : \theta_1 = \theta_2$ is true or not. The prior distributions of θ_1 and θ_2 are both $N(\mu, \psi)$ where $\psi \gg \phi$. Write

$$u = x - y, \qquad z = \tfrac{1}{2}(x + y).$$

Show that, if H_0 is true and $\theta_1 = \theta_2 = \theta$, then θ, $x - \theta$ and $y - \theta$ are independent and

$$\theta \sim \mathrm{N}(\mu, \psi), \qquad x - \theta \sim \mathrm{N}(0, \phi), \qquad y - \theta \sim \mathrm{N}(0, \phi).$$

By writing $u = (x-\theta)-(y-\theta)$ and $z = \theta+\frac{1}{2}(x-\theta)+\frac{1}{2}(y-\theta)$, go on to show that u has an $\mathrm{N}(0, 2\phi)$ distribution and that z has an $\mathrm{N}(\mu, \frac{1}{2}\phi+\psi)$, so approximately an $\mathrm{N}(\mu, \psi)$, distribution. Conversely, show that if H_0 is false and θ_1 and θ_2 are assumed independent, then θ_1, θ_2, $x - \theta_1$ and $y - \theta_2$ are all independent and

$$\theta_1 \sim \mathrm{N}(\mu, \psi), \quad \theta_2 \sim \mathrm{N}(\mu, \psi), \quad x - \theta_1 \sim \mathrm{N}(0, \phi), \quad y - \theta_2 \sim \mathrm{N}(0, \phi).$$

By writing

$$\begin{aligned} u &= \theta_1 - \theta_2 + (x - \theta_1) - (y - \theta_2), \\ z &= \tfrac{1}{2}\{\theta_1 + \theta_2 + (x - \theta_1) + (y - \theta_2)\} \end{aligned}$$

show that in this case u has an $\mathrm{N}(0, 2(\phi + \psi))$, so approximately an $\mathrm{N}(0, 2\psi)$, distribution, while z has an $\mathrm{N}(\mu, \frac{1}{2}(\phi + \psi))$, so approximately an $\mathrm{N}(\mu, \frac{1}{2}\psi)$, distribution. Conclude that the Bayes factor is approximately

$$B = \sqrt{(\psi/2\phi)}\, \exp[-\tfrac{1}{2}u^2/2\phi + \tfrac{1}{2}(z - \mu)^2/\psi].$$

Suppose that the ratio $\sqrt{(\psi/\phi)}$ of the standard deviations is 100 and that $u = 2 \times \sqrt{(2\phi)}$, so that the difference between x and y represents two standard deviations, and that $z = \mu$, so that both specimens are of commonly occurring glass.

Show that a classical test would reject H_0 at the 5% level, but that $B = 9.57$, so that the odds in favour of H_0 are multiplied by a factor just below 10.

[This problem is due to Lindley (1977); see also Shafer (1982). Lindley comments that, "What the [classical] test fails to take into account is the extraordinary coincidence of x and y being so close together were the two pieces of glass truly different".]

13. Lindley (1957) originally discussed his paradox under slightly different assumptions from those made in this book. Follow through the reasoning used in Section 4.5 with $\rho_1(\theta)$ representing a uniform distribution on the interval $(\theta_0 - \frac{1}{2}\tau, \theta_0 + \frac{1}{2}\tau)$ to find the corresponding Bayes factor assuming that $\tau^2 \gg \phi/n$, so that an $\mathrm{N}(\mu, \phi/n)$ variable lies in this interval with very high probability. Check that your answers are unlikely to disagree with those found in Section 4.5 under the assumption that $\rho_1(\theta)$ represents a normal density.

14. Express in your own words the arguments given by Jeffreys (1961, Section 5.2) in favour of a Cauchy distribution

$$\rho_1(\theta) = \frac{1}{\pi} \frac{\sqrt{\psi}}{\psi + (\theta - \theta_0)^2}$$

in the problem discussed in the previous question.

15. Suppose that x has a binomial distribution $B(n, \theta)$ of index n and parameter θ, and that it is desired to test $\mathrm{H}_0 : \theta = \theta_0$ against the alternative hypothesis $\mathrm{H}_1 : \theta \neq \theta_0$.

(a) Find lower bounds on the posterior probability of H_0 and on the Bayes factor for H_0 versus H_1, bounds which are valid for any $\rho_1(\theta)$.

(b) If $n = 20$, $\theta_0 = \frac{1}{2}$ and $x = 15$ is observed, calculate the (two-tailed) P-value and the lower bound on the posterior probability when the prior probability π_0 of the null hypothesis is $\frac{1}{2}$.

16. Twelve observations from a normal distribution of mean θ and variance ϕ are available, of which the sample mean is 1.2 and the sample variance is 1.1. Compare the Bayes factors in favour of the null hypothesis that $\theta = \theta_0$ assuming (a) that ϕ is unknown and (b) that it is known that $\phi = 1$.

17. Suppose that in testing a point null hypothesis you find a value of the usual Student's t statistic of 2.4 on 8 degrees of freedom. Would the methodology of Section 4.6 require you to "think again"?

18. Which entries in the table in Section 4.5 on "Point null hypotheses for the normal distribution would, according to the methodology of Section 4.6, cause you to "think again"?

H.5 Exercises on Chapter 5

1. Two analysts measure the percentage of ammonia in a chemical process over 9 days and find the following discrepancies between their results:

Day	1	2	3	4	5	6	7	8	9
Analyst A	12.04	12.37	12.35	12.43	12.34	12.36	12.48	12.33	12.33
Analyst B	12.18	12.37	12.38	12.36	12.47	12.48	12.57	12.28	12.42

Investigate the mean discrepancy θ between their results and in particular give an interval in which you are 90% sure that it lies.

2. With the same data as in the previous question, test the hypothesis that there is no discrepancy between the two analysts.

3. Suppose that you have grounds for believing that observations x_i, y_i for $i = 1, 2, \ldots, n$ are such that $x_i \sim \mathrm{N}(\theta, \phi_i)$ and also $y_i \sim \mathrm{N}(\theta, \phi_i)$, but that you are not prepared to assume that the ϕ_i are equal. What statistic would you expect to base inferences about θ on?

4. How much difference would it make to the analysis of the data in Section 5.1 on rat diet if we took $\omega = \frac{1}{2}(\phi + \psi)$ instead of $\omega = \phi + \psi$.

5. Two analysts in the same laboratory made repeated determinations of the percentage of fibre in soya cotton cake, the results being as shown below:

Analyst A	12.38	12.53	12.25	12.37	12.48	12.58	12.43	12.43	12.30
Analyst B	12.25	12.45	12.31	12.31	12.30	12.20	12.25	12.25	12.26
	12.42	12.17	12.09						

Investigate the mean discrepancy θ between their mean determinations and in particular give an interval in which you are 90% sure that it lies

(a) assuming that it is known from past experience that the standard deviation of both sets of observations is 0.1, and

(b) assuming simply that it is known that the standard deviations of the two sets of observations are equal.

6. A random sample $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ is available from an $\mathrm{N}(\lambda, \phi)$ distribution and a second independent random sample $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ is available from an $\mathrm{N}(\mu, 2\phi)$ distribution. Obtain, under the usual assumptions, the posterior distributions of $\lambda - \mu$ and of ϕ.

7. Verify the formula for S_1 given towards the end of Section 5.2.

8. The following data consists of the lengths in mm of cuckoo's eggs found in nests belonging to the dunnock and to the reed warbler:

Dunnock	22.0	23.9	20.9	23.8	25.0	24.0	21.7	23.8	22.8	23.1
Reed warbler	23.2	22.0	22.2	21.2	21.6	21.9	22.0	22.9	22.8	

Investigate the difference θ between these lengths without making any particular assumptions about the variances of the two populations, and in particular give an interval in which you are 90% sure that it lies.

9. Show that if $m = n$ then the expression f_1^2/f_2 in Patil's approximation reduces to
$$\frac{4(m-5)}{3+\cos 4\theta}.$$

10. Suppose that T_x, T_y and θ are defined as in Section 5.3 and that
$$T = T_x \sin\theta - T_y \cos\theta, \qquad U = T_x \cos\theta + T_y \sin\theta$$
Show that the transformation from (T_x, T_y) to (T, U) has unit Jacobian and hence show that the density of T satisfies
$$p(T|\boldsymbol{x}, \boldsymbol{y}) \propto \int_0^\infty [1 + (T\sin\theta + U\cos\theta)^2/\nu_x]^{-(\nu(x)+1)/2} \times [1 + (-T\cos\theta + U\sin\theta)^2/\nu_y]^{-(\nu(y)+1)/2}\,\mathrm{d}U.$$

11. Show that if $x \sim \mathrm{F}_{\nu_1,\nu_2}$ then
$$\frac{\nu_1 x}{\nu_2 + \nu_1 x} \sim \mathrm{Be}(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2).$$

12. Two different microscopic methods, A and B, are available for the measurement of very small dimensions in microns. As a result of several such measurements on the same object, estimates of variance are available as follows:

Method	A	B
No. of observations	$m = 15$	$n = 25$
Estimated variance	$s_1^2 = 7.533$	$s_2^2 = 1.112$

Give an interval in which you are 95% sure that the ratio of the variances lies.

13. Measurement errors when using two different instruments are more or less symmetrically distributed and are believed to be reasonably well approximated by a normal distribution. Ten measurements with each show a sample standard deviation three times as large with one instrument as with the other. Give an interval in which you are 99% sure that the ratio of the true standard deviations lies.

14. Repeat the analysis of Di Raimondo's data in Section 5.6 on the effects of penicillin of mice, this time assuming that you have prior knowledge worth about six observations in each case suggesting that the mean chance of survival is about a half with the standard injection but about two-thirds with the penicillin injection.

15. The table below [quoted from Jeffreys (1961, Section 5.1)] gives the relationship between grammatical gender in Welsh and psychoanalytical symbolism according to Freud:

Psycho. \ Gram.	M	F
M	45	30
F	28	29
Total	73	59

Find the posterior probability that the log odds-ratio is positive and compare it with the comparable probability found by using the inverse root-sine transformation.

16. Show that if $\pi \cong \rho$ then the log odds-ratio is such that

$$\Lambda - \Lambda' \cong (\pi - \rho)/\{\pi(1 - \pi)\}.$$

17. A report issued in 1966 about the effect of radiation on patients with inoperable lung cancer compared the effect of radiation treatment with placebos.

The numbers surviving after a year were:

	Radiation	Placebos
No. of cases	308	246
No. surviving	56	34

What are the approximate posterior odds that the one-year survival rate of irradiated patients is at least 0.01 greater than that of those who were not irradiated?

18. Suppose that $x \sim \mathrm{P}(8.5)$, i.e. x is Poisson of mean 8.5, and $y \sim \mathrm{P}(11.0)$. What is the approximate distribution of $x - y$?

H.6 Exercises on Chapter 6

1. The sample correlation coefficient between length and weight of a species of frog was determined at each of a number of sites. The results were as follows:

Site	1	2	3	4	5
Number of frogs	12	45	23	19	30
Correlation	0.631	0.712	0.445	0.696	0.535

Find an interval in which you are 95% sure that the correlation coefficient lies.

2. Three groups of children were given two tests. The numbers of children and the sample correlation coefficients between the two test scores in each group were as follows:

Number of children	45	34	49
Correlation	0.489	0.545	0.601

Is there any evidence that the association between the two tests differs in the three groups?

3. Suppose you have sample correlation coefficients $r_1, r_2, \dots, r_k$ on the basis of sample sizes $n_1, n_2, \dots, n_k$. Give a 95% posterior confidence interval for $\zeta = \tanh^{-1}\rho$.

4. From the approximation

$$p(\rho|\boldsymbol{x}, \boldsymbol{y}) \propto (1-\rho^2)^{n/2}(1-\rho r)^{-n}$$

which holds for large n, deduce an expression for the log-likelihood $L(\rho|\boldsymbol{x}, \boldsymbol{y})$ and hence show that the maximum likelihood occurs when $\rho = r$. An approximation to the information can now be made by replacing r by ρ in the second derivative of the likelihood, since ρ is near r with high probability. Show that this approximation suggests a prior density of the form

$$p(\rho) \propto (1-\rho^2)^{-1}.$$

5. Use the fact that

$$\int_0^\infty (\cosh t + \cos\theta)^{-1}\,\mathrm{d}t = \theta/\sin\theta$$

(cf. Edwards, 1921, art. 180) to show that

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}\frac{d^{n-2}}{\mathrm{d}(\rho r)^{n-2}}\left(\frac{\operatorname{arc}\cos(-\rho r)}{\sqrt{(1-\rho^2 r^2)}}\right).$$

6. Show that in the special case where the sample correlation coefficient $r = 0$ and the prior takes the special form $p(\rho) \propto (1-\rho^2)^k$ the variable

$$\sqrt{(k+n+1)}\rho/(1-\rho^2)$$

has a Student's t distribution on $k+n+1$ degrees of freedom.

7. By writing

$$\begin{aligned}\omega^{-1}(\omega+\omega^{-1}-2\rho r)^{-(n-1)} = \omega^{n-2}&(1-\rho^2)^{-(n-1)}\\ &\times[1+(\omega-\rho r)^2(1-\rho^2 r^2)^{-1}]^{-(n-1)}\end{aligned}$$

and using repeated integration by parts, show that the posterior distribution of ρ can be expressed as a finite series involving powers of

$$\sqrt{(1-\rho r)/(1+\rho r)}$$

and Student's t integrals.

8. By substituting

$$\cosh t - \rho r = \frac{1-\rho r}{1-u}$$

in the form

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}\int_0^\infty (\cosh t - \rho r)^{-(n-1)}\,\mathrm{d}t$$

for the posterior density of the correlation coefficient and then expanding

$$[1-\tfrac{1}{2}(1+\rho r)u]^{-\frac{1}{2}}$$

as a power series in u, show that the integral can be expressed as a series of beta functions. Hence deduce that

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}(1-\rho r)^{-n+(3/2)}S_n(\rho r)$$

where

$$S_n(\rho r) = 1 + \sum_{l=1}^{\infty} \frac{1}{l!}\left(\frac{1+\rho r}{8}\right)^l \prod_{s=1}^{l} \frac{(2s-1)^2}{(n-\frac{3}{2}+s)}.$$

9. Fill in the details of the derivation of the prior

$$p(\phi,\psi,\rho) \propto (\phi\psi)^{-1}(1-\rho^2)^{-3/2}$$

from Jeffreys' rule as outlined at the end of Section 6.1.

10. The data below consist of the estimated gestational ages (in weeks) and weights (in grammes) of twelve female babies:

Age	40	36	40	38	42	39	40	37	36	38	39	40
Weight	3317	2729	2935	2754	3210	2817	3126	2539	2412	2991	2875	3231

Give an interval in which you are 90% sure that the gestational age of a particular such baby will lie if its weight is 3000 grammes, and give a similar interval in which the mean weight of all such babies lies.

11. Show directly from the definitions that, in the notation of Section 6.3,

$$S_{ee} = \sum\{y_i - a - b(x_i - \overline{x})\}^2.$$

12. Observations y_i for $i = -m, -m+1, \dots, m$ are available which satisfy the regression model

$$y_i \sim \mathrm{N}(\alpha + \beta u_i + \gamma v_i,\, \phi)$$

where $u_i = i$ and $v_i = i^2 - \frac{1}{2}m(m+1)$. Adopting the standard reference prior $p(\alpha,\beta,\gamma,\phi) \propto 1/\phi$, show that the posterior distribution of α is such that

$$\frac{\alpha - \overline{y}}{s/\sqrt{n}} \sim \mathrm{t}_{n-3}$$

where $n = 2m+1$, $s^2 = S_{ee}/(n-3)$ and

$$S_{ee} = S_{yy} - S_{uy}^2/S_{uu} - S_{vy}^2/S_{vv}$$

in which S_{yy}, S_{uy}, etc., are defined by

$$S_{yy} = \sum(y_i - \overline{y})^2, \qquad S_{uy} = \sum(u_i - \overline{u}))(y_i - \overline{y}).$$

[*Hint:* Note that $\sum u_i = \sum v_i = \sum u_i v_i = 0$, and hence $\overline{u} = \overline{v} = 0$ and $S_{uy} = 0$.]

13. Fisher (1925b, Section 41) quotes an experiment on the accuracy of counting soil bacteria. In it, a soil sample was divided into four parallel samples, and from each of theses after dilution seven plates were inoculated. The number of colonies on each plate is shown below. Do the results from the four samples agree within the limits of random sampling?

Plate \ Sample	A	B	C	D
1	72	74	78	69
2	69	72	74	67
3	63	70	70	66
4	59	69	58	64
5	59	66	58	64
6	53	58	56	58
7	51	52	56	54

14. In the case of the data on scab disease quoted in Section 6.5, find a contrast measuring the effect of the season in which sulphur is applied and give an appropriate HDR for this contrast.

15. The data below [from Wishart and Sanders (1955, Table 5.6)] represent the weight of green produce in pounds made on an old pasture. There were three main treatments, including a control (O) consisting of the untreated land. In the other cases the effect of a grass-land rejuvenator (R) was compared with the use of the harrow (H). The blocks were therefore composed of three plots each, and the experiment consisted of six randomized blocks placed side by side. The plan and yields were as follows:

O. H. R.	R. H. O.	O. R. H.	O. R. H.	H. O. R.	O. H. R.
813 647 713	814 759 795	705 652 598	774 617 559	580 687 539	581 480 437

Derive an appropriate two-way analysis of variance.

16. Express the two-way layout as a particular case of the general linear model.

17. Show that the matrix $\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ which arises in the theory of the general linear model is a *generalized inverse* of the (usually non-square) matrix $\boldsymbol{A}$ in that

(a) $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{A}$

(b) $\boldsymbol{A}^{+}\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{A}^{+}$

(c) $(\boldsymbol{A}\boldsymbol{A}^{+})^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{+}$

(d) $(\boldsymbol{A}^+\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}^+\boldsymbol{A}$

18. Express the bivariate linear regression model in terms of the original parameters $\boldsymbol{\eta} = (\eta_0, \eta_1)^{\mathrm{T}}$ and the matrix $\boldsymbol{A}_0$ and use the general linear model to find the posterior distribution of $\boldsymbol{\eta}$.

H.7 Exercises on Chapter 7

1. Show that in any experiment E in which there is a possible value y for the random variable $\widetilde{x}$ such that $p_{\widetilde{x}}(y|\theta) = 0$, then if z is any other possible value of $\widetilde{x}$, the statistic $t = t(x)$ defined by

$$t(x) = \begin{cases} z & \text{if } x = y \\ x & \text{if } x \neq y \end{cases}$$

is sufficient for θ given x. Hence show that if $\widetilde{x}$ is a continuous random variable, then a naïve application of the weak sufficiency principle as defined in Section 7.1 would result in $\mathrm{Ev}\{E, y, \theta\} = \mathrm{Ev}\{E, z, \theta\}$ for any two possible values y and z of $\widetilde{x}$.

2. Consider an experiment $E = \{\widetilde{x}, \theta, p(x|\theta)\}$. We say that *censoring* (strictly speaking, fixed censoring) occurs with censoring mechanism g (a known function of x) when, instead of $\widetilde{x}$, one observes $y = g(x)$. A typical example occurs when we report x if $x < k$ for some fixed k, but otherwise simply report that $x \geqslant k$. As a result, the experiment really performed is $E^g = \{\widetilde{y}, \theta, p(y|\theta)\}$. A second method with censoring mechanism h is said to be *equivalent* to the first when

$$g(x) = g(x') \quad \text{if and only if} \quad h(x) = h(x').$$

As a special case, if g is one-to-one then the mechanism is said to be equivalent to no censoring. Show that if two censoring mechanisms are equivalent, then the likelihood principle implies that

$$\mathrm{Ev}\{E^g, x, \theta\} = \mathrm{Ev}\{E^h, x, \theta\}.$$

3. Suppose that the density function $p(x|\theta)$ is defined as follows for $x = 1, 2, 3, \ldots$ and $\theta = 1, 2, 3, \ldots$. If θ is even, then

$$p(x|\theta) = \begin{cases} \frac{1}{3} & \text{if } x = \theta/2,\ 2\theta \text{ or } 2\theta + 1 \\ 0 & \text{otherwise} \end{cases}$$

if θ is odd but $\theta \neq 1$, then

$$p(x|\theta) = \begin{cases} \frac{1}{3} & \text{if } x = (\theta-1)/2,\ 2\theta \text{ or } 2\theta+1 \\ 0 & \text{otherwise} \end{cases}$$

while if $\theta = 1$ then

$$p(x|\theta) = \begin{cases} \frac{1}{3} & \text{if } x = \theta,\ 2\theta \text{ or } 2\theta+1 \\ 0 & \text{otherwise} \end{cases}$$

Show that, for any x the data intuitively give equal support to the three possible values of θ compatible with that observation, and hence that on likelihood grounds any of the three would be a suitable estimate. Consider, therefore, the three possible estimators d_1, d_2 and d_3 corresponding to the smallest, middle and largest possible θ. Show that

$$p(d_2 = 1) = \begin{cases} \frac{1}{3} & \text{when } \theta \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

$$p(d_3 = 1) = \begin{cases} \frac{1}{3} & \text{when } \theta \text{ is odd but } \theta \neq 1 \\ 0 & \text{otherwise} \end{cases}$$

but that

$$p(d_1 = 1) = \begin{cases} 1 & \text{when } \theta = 1 \\ \frac{2}{3} & \text{otherwise} \end{cases}$$

Does this apparent discrepancy cause any problems for a Bayesian analysis? (due to G. Monette and D. A. S. Fraser).

4. A drunken soldier, starting at an intersection O in a city which has square blocks, staggers around a random path trailing a taut string. Eventually he stops at an intersection (after walking at least one block) and buries a treasure. Let θ denote the path of the string from O to the treasure. Letting N, S, E and W stand for a path segment one block long in the indicated direction, so that θ can be expressed as a sequence of such letters, say $\theta = NNESWSWW$. (Note that NS, SN, EW and WE cannot appear as the taut string would be rewound). After burying the treasure, the soldier walks one block further in a random direction (still keeping the string taut). Let X denote this augmented path, so that X is one of θN, θS, θE and θW, each with probability $\frac{1}{4}$. You observe X and are then to find the treasure. Show that if you use a reference prior $p(\theta) \propto 1$ for all possible paths θ, then all four possible values of θ given X are equally likely. Note, however, that intuition would suggest that θ is three times as likely to extend the path as to backtrack, suggesting that one particular value of θ is more likely than the others after X is observed. (Due to M. Stone).

5. Suppose that, starting with a fortune of f_0 units, you bet a units each time on evens at roulette (so that you have a probability of 18/37 of winning at Monte Carlo or 18/38 at Las Vegas) and keep a record of your fortune f_n and the difference d_n between the number of times you win and the number of times you lose in n games. Which of the following are stopping times?

(a) The last time n at which $f_n \geqslant f_0$?

(b) The first time that you win in three successive games?

(c) The value of n for which $f_n = \max_{\{0 \leqslant k < \infty\}} f_k$?

6. Suppose that $x_1, x_2, \ldots$ is a sequential sample from an $\mathrm{N}(\theta, 1)$ distribution and it is desired to test $\mathrm{H}_0 : \theta = \theta_0$ versus $\mathrm{H}_1 : \theta \neq \theta_0$. The experimenter reports that he used a proper stopping rule and obtained the data 3, −1, 2, 1.

(a) What could a frequentist conclude?

(b) What could a Bayesian conclude?

7. Let $x_1, x_2, \ldots$ be a sequential sample from a Poisson distribution $\mathrm{P}(\lambda)$. Suppose that the stopping rule is to stop sampling at time $n \geqslant 2$ with probability

$$\sum_{i=1}^{n-1} x_i \Big/ \sum_{i=1}^{n} x_i$$

for $n = 2, 3, \ldots$ (define $0/0 = 1$). Suppose that the first five observations are 3, 1, 2, 5, 7 and that sampling then stops. Find the likelihood function for λ. (Berger, 1985).

8. Show that the mean of the beta-Pascal distribution

$$p(S|R, r, s) = \binom{S}{s} \frac{\mathrm{B}(r'' + s, R'' - r'' + S - s)}{\mathrm{B}(r'' - 1, R'' - r'')}$$

is given by the formula in Section 7.3, namely,

$$\mathsf{E}S = (s + 1)\left(\frac{R'' - 2}{r'' - 2}\right) - 1$$

9. Suppose that you intend to observe the number x of successes in n Bernoulli trials and the number y of failures before the nth success after the first n trials, so that $x \sim \mathrm{B}(n, \pi)$ and $y \sim \mathrm{NB}(n, \pi)$. Find the likelihood function $L(\pi|x, y)$ and deduce the reference prior that Jeffreys' rule would suggest for this case.

10. The negative of loss is sometimes referred to as *utility*. Consider a gambling game very unlike most in that you are bound to win at least £2, and accordingly in order to be allowed to play, you must pay an entry fee of £e. A coin is tossed until it comes up heads, and if this occurs for the first time on the nth toss, you receive £2^n. Assuming that the utility to you of making a gain of £x is $u(x)$, find the expected utility of this game, and then discuss whether it is plausible that $u(x)$ is directly proportional to x. [The gamble discussed here is known as the *St Petersburg Paradox*. A fuller discussion of it can be found in Leonard and Hsu (1999, Chapter 4).]

11. Suppose that you want to estimate the parameter π of a binomial distribution $\mathrm{B}(n, \pi)$. Show that if the loss function is

$$L(\theta, a) = (\theta - a)^2/\{\theta(1-\theta)\}$$

then the Bayes rule corresponding to a uniform (that is, $\mathrm{Be}(1,1)$) prior for π is given by $d(x) = x/n$ for any x such that $0 < x < n$, that is, the maximum likelihood estimator. Is $d(x) = x/n$ a Bayes rule if $x = 0$ or $x = n$?

12. Let $x \sim \mathrm{B}(n, \pi)$ and $y \sim \mathrm{B}(n, \rho)$ have independent binomial distributions of the same index but possibly different parameters. Find the Bayes rule corresponding to the loss

$$L((\pi, \rho), a) = (\pi - \rho - a)^2$$

when the priors for π and ρ are independent uniform distributions.

13. Investigate possible point estimators for π on the basis of the posterior distribution in the example in the subsection of Section 2.10 headed "Mixtures of conjugate densities".

14. Find the Bayes rule corresponding to the loss function

$$L(\theta, a) = \begin{cases} u(a - \theta) & \text{if } a \leqslant \theta \\ v(\theta - a) & \text{if } a \geqslant \theta. \end{cases}$$

15. Suppose that your prior for the proportion π of defective items supplied by a manufacturer is given by the beta distribution $\mathrm{Be}(2, 12)$, and that you then observe that none of a random sample of size 6 is defective. Find the posterior distribution and use it to carry out a test of the hypothesis $\mathrm{H}_0 : \pi < 0.1$ using

(a) a "$0 - 1$" loss function, and

(b) the loss function

$a\backslash\theta$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	1
a_1	2	0

16. Suppose there is a loss function $L(\theta, a)$ defined by

$a\backslash\theta$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	10
a_1	10	0
a_2	3	3

On the basis of an observation x you have to take action a_0, a_1 or a_2. For what values of the posterior probabilities p_0 and p_1 of the hypotheses $\mathrm{H}_0 : \theta \in \Theta_0$ and $\mathrm{H}_1 : \theta \in \Theta_1$ would you take each of the possible actions?

17. A child is given an intelligence test. We assume that the test result x is $\mathrm{N}(\theta, 100)$ where θ is the true intelligence quotient of the child, as measured by the test (in other words, if the child took a large number of similar tests, the average score would be θ). Assume also that, in the population as a whole, θ is distributed according to an $\mathrm{N}(100, 225)$ distribution. If it is desired, on the basis of the intelligence quotient, to decide whether to put the child into a slow, average or fast group for reading, the actions available are:

a_1 : Put in slow group, that is, decide $\theta \in \Theta_1 = (0, 90)$
a_1 : Put in average group, that is, decide $\theta \in \Theta_2 = [90, 100]$
a_1 : Put in fast group, that is, decide $\theta \in \Theta_3 = (100, \infty)$.

A loss function $L(\theta, a)$ of the following form might be deemed appropriate:

$a\backslash\theta$	$\theta \in \Theta_1$	$\theta \in \Theta_2$	$\theta \in \Theta_3$
a_1	0	$\theta - 90$	$2(\theta - 90)$
a_2	$90 - \theta$	0	$\theta - 110$
a_3	$2(110 - \theta)$	$110 - \theta$	0

Assume that you observe that the test result $x = 115$. By using tables of the normal distribution and the fact that if $\phi(t)$ is the density function of the standard normal distribution, then $\int t\phi(t)\,\mathrm{d}t = -\phi(t)$, find is the appropriate action to take on the basis of this observation. [See Berger (1985, Sections 4.2, 4.3, 4.4)].

18. In Section 7.8, a point estimator δ_n for the current value λ of the parameter of a Poisson distribution was found. Adapt the argument to deal with the case where the underlying distribution is geometric, that is

$$p(x|\pi) = \pi(1-\pi)^x.$$

Generalize to the case of a negative binomial distribution, that is,

$$p(x|\pi) = \binom{n+x-1}{x}\pi^n(1-\pi)^x.$$

H.8 Exercises on Chapter 8

1. Show that the prior

$$p(\alpha,\beta) \propto (\alpha+\beta)^{-5/2}$$

suggested in connection with the example on risk of tumour in a group of rats is equivalent to a density uniform in $\bigl(\alpha/(\alpha+\beta),\,(\alpha+\beta)^{-1/2}\bigr)$.

2. Observations $x_1, x_2, \dots, x_n$ are independently distributed given parameters $\theta_1, \theta_2, \dots, \theta_n$ according to the Poisson distribution $p(x_i|\boldsymbol{\theta}) = \theta_i^{x_i}\exp(-\theta_i)/x_i!$. The prior distribution for $\boldsymbol{\theta}$ is constructed hierarchically. First the θ_is are assumed to be independently identically distributed given a hyperparameter ϕ according to the exponential distribution $p(\theta_i|\phi) = \phi\exp(-\phi\theta_i)$ for $\theta_i \geqslant 0$ and then ϕ is given the improper uniform prior $p(\phi) \propto 1$ for $\phi \geqslant 0$. Provided that $n\overline{x} > 1$, prove that the posterior distribution of $z = 1/(1+\phi)$ has the beta form

$$p(z|\boldsymbol{x}) \propto z^{n\overline{x}-2}(1-z)^n.$$

Thereby show that the posterior means of the θ_i are shrunk by a factor $(n\overline{x}-1)/(n\overline{x}+n)$ relative to the usual classical procedure which estimates each of the θ_i by x_i.

What happens if $n\overline{x} \leqslant 1$?

3. Carry out the Bayesian analysis for known overall mean developed in Section 8.2 above (a) with the loss function replaced by a weighted mean

$$L(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{r} w_i(\theta_i - \widehat{\theta}_i)^2,$$

and (b) with it replaced by

$$L(\boldsymbol{\theta},\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{r} |\theta_i - \widehat{\theta}_i|.$$

4. Compare the effect of the Efron-Morris estimator on the baseball data in Section 8.3 with the effect of a James-Stein estimator which shrinks the values of π_i towards $\pi_0 = 0.25$ or equivalently shrinks the values of X_i towards $\mu = 2\sqrt{n}\sin^{-1}\sqrt{\pi_0}$.

5. The *Helmert transformation* is defined by the matrix

$$\boldsymbol{A} = \begin{pmatrix} r^{-1/2} & 2^{-1/2} & 6^{-1/2} & 12^{-1/2} & \dots & \{r(r-1)\}^{-1/2} \\ r^{-1/2} & -2^{-1/2} & 6^{-1/2} & 12^{-1/2} & \dots & \{r(r-1)\}^{-1/2} \\ r^{-1/2} & 0 & -2\times 6^{-1/2} & 12^{-1/2} & \dots & \{r(r-1)\}^{-1/2} \\ r^{-1/2} & 0 & 0 & -3\times 12^{-1/2} & \dots & \{r(r-1)\}^{-1/2} \\ r^{-1/2} & 0 & 0 & 0 & \dots & \{r(r-1)\}^{-1/2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{-1/2} & 0 & 0 & 0 & \dots & -(r-1)^{1/2}r^{-1/2} \end{pmatrix}$$

so that the element a_{ij} in row i, column j is

$$a_{ij} = \begin{cases} r^{-1/2} & (j=1) \\ \{j(j-1)\}^{-1/2} & (i<j) \\ 0 & (i>j>1) \\ -(j-1)^{1/2}j^{-1/2} & (i=j>1). \end{cases}$$

It is also useful to write $\boldsymbol{\alpha}_j$ for the (column) vector which consists of the jth column of the matrix $\boldsymbol{A}$. Show that if the variates X_i are independently $\mathrm{N}(\theta_i, 1)$, then the variates $W_j = \boldsymbol{\alpha}_j{}^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu}) = \sum_i a_{ij}(X_i-\mu_j)$ are independently normally distributed with unit variance and such that $\mathsf{E}W_j = 0$ for $j>1$ and

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = \sum_j W_j^2 = \sum_i (X_i-\mu_i)^2 = (\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu}).$$

By taking $a_{ij} \propto \theta_j - \mu_j$ for $i>j$, $a_{ij}=0$ for $i<j$ and a_{jj} such that $\sum_j a_{ij} = 0$, extend this result to the general case and show that $\mathsf{E}\,W_1 \propto \gamma = \sum_i(\theta_i-\mu_i)^2$. Deduce that the distribution of a non-central chi-squared variate depends only of r and γ.

6. Show that $R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\boldsymbol{JS+}}) < R(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}^{\boldsymbol{JS}})$ where

$$\widehat{\boldsymbol{\theta}}^{\boldsymbol{JS+}} = \boldsymbol{\mu} + \max\left[\left(1-\frac{r-2}{S_1}\right), 0\right](\boldsymbol{X}-\boldsymbol{\mu})$$

(Lehmann 1983, Section 4.6, Theorem 6.2).

7. Writing

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}, \qquad \widehat{\boldsymbol{\theta}}_k = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$$

for the least-squares and ridge regression estimators for regression coefficients $\boldsymbol{\theta}$, show that

$$\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k = k(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\widehat{\boldsymbol{\theta}}_k$$

and that the bias of $\widehat{\boldsymbol{\theta}}_k$ is

$$\boldsymbol{b}(k) = \{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}-\boldsymbol{I}\}\boldsymbol{\theta}$$

while its variance-covariance matrix is

$$\mathcal{V}\widehat{\boldsymbol{\theta}}_k = \phi(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}+k\boldsymbol{I})^{-1}.$$

Deduce expressions for the sum $\mathcal{G}(k)$ of the squares of the biases and for the sum $\mathcal{F}(k)$ of the variances of the regression coefficients, and hence show that the mean square error is

$$MSE_k = \mathsf{E}(\widehat{\theta}_b - \theta)^{\mathrm{T}}(\widehat{\theta}_k - \theta) = \mathcal{F}(k) + \mathcal{G}(k).$$

Assuming that $\mathcal{F}(k)$ is continuous and monotonic decreasing with $\mathcal{F}'(0) = 0$ and that $\mathcal{G}(k)$ is continuous and monotonic increasing with $\mathcal{G}(k) = \mathcal{G}'(k) = 0$, deduce that there always exists a k such that $MSE_k < MSE_0$ (Theobald, 1974).

8. Show that the matrix $\boldsymbol{H}$ in Section 8.6 satisfies $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{B} = \boldsymbol{0}$ and that if $\boldsymbol{B}$ is square and non-singular then $\boldsymbol{H}^{-1}$ vanishes.

9. Consider the following particular case of the two way layout. Suppose that eight plots are harvested on four of which one variety has been sown, while a different variety has been sown on the other four. Of the four plots with each variety, two different fertilizers have been used on two each. The yield will be normally distributed with a mean θ dependent on the fertiliser and the variety and with variance ϕ. It is supposed *a priori* that the mean for plots yields sown with the two different varieties are independently normally distributed with mean α and variance ψ_α, while the effect of the two different fertilizers will add an amount which is independently normally distributed with mean β and variance ψ_β. This fits into the situation described in Section 8.6 with Φ being ϕ times an 8×8 identity matrix and

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}; \qquad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}; \qquad \boldsymbol{\Psi} = \begin{pmatrix} \psi_\alpha & 0 & 0 & 0 \\ 0 & \psi_\alpha & 0 & 0 \\ 0 & 0 & \psi_\beta & 0 \\ 0 & 0 & 0 & \psi_\beta \end{pmatrix}.$$

Find the matrix $\boldsymbol{K}^{-1}$ needed to find the posterior of θ.

10. Generalize the theory developed in Section 8.6 to deal with the case where $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{A\theta}, \boldsymbol{\Phi})$ and $\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{B\mu}, \boldsymbol{\Psi})$ and knowledge of $\boldsymbol{\mu}$ is vague to deal with the case where $\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{C\nu}, \boldsymbol{K})$ (Lindley and Smith, 1972).

11. Find the elements of the variance-covariance matrix $\boldsymbol{\Sigma}$ for the one way model in the case where $n_i = n$ for all i.

H.9 Exercises on Chapter 9

1. Find the value of $\int_0^1 \mathrm{e}^x\,\mathrm{d}x$ by crude Monte Carlo integration using a sample size of $n = 10$ values from a uniform distribution $\mathrm{U}(0,1)$ taken from tables of random numbers [use, for example, groups of random digits from Lindley and Scott (1995, Table 27) or Neave (1978, Table 8.1)]. Repeat the experiment 10 times and compute the overall mean and the sample standard deviation of the values you obtain. What is the theoretical value of the population standard deviation and how does the value you obtained compare with it?

2. Suppose that, in a Markov chain with just two states, the probabilities of going from state i to state j in one time unit are given by the entries of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{pmatrix}$$

in which i represents the row and j the column. Show that the probability of getting from state i to state j in t time units is given by the tth power of the matrix $\boldsymbol{A}$ and that

$$\boldsymbol{A}^t = \begin{pmatrix} 2/5 & 3/5 \\ 2/5 & 3/5 \end{pmatrix} + \left(-\tfrac{1}{4}\right)^t \begin{pmatrix} 3/5 & -3/5 \\ -2/5 & 2/5 \end{pmatrix}.$$

Deduce that, irrespective of the state the chain started in, after a long time it will be in the first state with probability $2/5$ and in the second state with probability $3/5$.

3. Smith (1969, Section 21.10) quotes an example on genetic linkage in which we have observations $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ with cell probabilities

$$\left(\tfrac{1}{4} + \tfrac{1}{4}\eta,\ \tfrac{1}{4}\eta,\ \tfrac{1}{4}(1-\eta),\ \tfrac{1}{4}(1-\eta) + \tfrac{1}{4}\right).$$

The values quoted are $x_1 = 461$, $x_2 = 130$, $x_3 = 161$ and $x_4 = 515$. Divide x_1 into y_0 and y_1 and x_4 into y_4 and y_5 to produce augmented data $\boldsymbol{y} = (y_0, y_1, y_2, y_3, y_4, y_5)$ and use the *EM* algorithm to estimate η.

4. Dempster *et al.* (1977) define a generalized *EM* algorithm (abbreviated as a *GEM* algorithm) as one in which $Q(\theta^{(t+1)}, \theta^{(t)}) \geqslant Q(\theta^{(t)}, \theta^{(t)})$. Give reasons for believing that *GEM* algorithms converge to the posterior mode.

5. In question 16 in Chapter 2 we supposed that the results of a certain test were known, on the basis of general theory, to be normally distributed about the same mean μ with the same variance ϕ, neither of which is known. In that

question we went on to suppose that your prior beliefs about (μ, ϕ) could be represented by a normal/chi-squared distribution with

$$\nu_0 = 4, \qquad S_0 = 350, \qquad n_0 = 1, \qquad \theta_0 = 85.$$

Find a semi-conjugate prior which has marginal distributions that are close to the marginal distributions of the normal/chi-squared prior but is such that the mean and variance are independent *a priori*. Now suppose as previously that 100 observations are obtained from the population with mean 89 and sample variance $s^2 = 30$. Find the posterior distribution of (μ, ϕ). Compare the posterior mean obtained by the *EM* algorithm with that obtained from the fully conjugate prior.

6. A textile company weaves a fabric on a large number of looms. Four looms selected at random from those available, and four observations of the tensile strength of fabric woven on each of these looms are available (there is no significance to the order of the observations from each of the looms), and the resulting data are given below:

Loom	Observations			
1	98	97	99	96
2	91	90	93	92
3	96	95	97	95
4	95	96	99	98

Estimate the means for each of the looms, the overall mean, the variance of observations from the same loom, and the variance of means from different looms in the population.

7. Write computer programs in C++ equivalent to the programs in R in this chapter.

8. Use the data augmentation algorithm to estimate the posterior density of the parameter η in the linkage model in question 3 above.

9. Suppose that $y \,|\, \pi \sim \mathrm{B}(n, \pi)$ and $\pi \,|\, y \sim \mathrm{Be}(y + \alpha, n - y + \beta)$ where n is a Poisson variable of mean λ as opposed to being fixed as in Section 9.4. Use the Gibbs sampler (chained data augmentation) to find the unconditional distribution of n in the case where $\lambda = 16$. $\alpha = 2$ and $\beta = 4$ (cf. Casella and George, 1992).

10. Find the mean and variance of the posterior distribution of θ for the data in question 5 above using the prior you derived in answer to that question by means of the Gibbs sampler (chained data augmentation).

11. The data below represent the weights of $r = 30$ young rats measured weekly for $n = 5$ weeks as quoted by Gelfand *et al.* (1990), Tanner (1996, Table 1.3 and Section 6.2.1), Carlin and Louis (2000, Example 5.6):

Rat\Week	1	2	3	4	5	Rat\Week	1	2	3	4	5
1	151	199	246	283	320	16	160	207	248	288	324
2	145	199	249	293	354	17	142	187	234	280	316
3	147	214	263	312	328	18	156	203	243	283	317
4	155	200	237	272	297	19	157	212	259	307	336
5	135	188	230	280	323	20	152	203	246	286	321
6	159	210	252	298	331	21	154	205	253	298	334
7	141	189	231	275	305	22	139	190	225	267	302
8	159	201	248	297	338	23	146	191	229	272	302
9	177	236	285	340	376	24	157	211	250	285	323
10	134	182	220	260	296	25	132	185	237	286	331
11	160	208	261	313	352	26	160	207	257	303	345
12	143	188	220	273	314	27	169	216	261	295	333
13	154	200	244	289	325	28	157	205	248	289	316
14	171	221	270	326	358	29	137	180	219	258	291
15	163	216	242	281	312	30	153	200	244	286	324

The weight of the ith rat in week j is denoted x_{ij} and we suppose that weight growth is linear, that is,

$$x_{ij} \sim \mathrm{N}(\alpha_i + \beta_i j,\, \phi),$$

but that the slope and intercept vary from rat to rat. We further suppose that α_i and β_i have a bivariate normal distribution, so that

$$\boldsymbol{\theta}_i = \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \sim \mathrm{N}(\boldsymbol{\theta}_0,\, \boldsymbol{\Sigma}) \quad \text{where} \quad \boldsymbol{\theta}_0 = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix},$$

and thus we have a random effects model. At the third stage, we suppose that

$$p(\boldsymbol{V}\,|\,\nu, \boldsymbol{\Omega}) \propto \frac{|\boldsymbol{V}|^{(\nu-k-1)/2}}{|\boldsymbol{\Omega}|^{\nu/2}} \exp\left[-\tfrac{1}{2}\mathrm{Trace}(\boldsymbol{\Omega}^{-1}\boldsymbol{V})\right].$$

Methods of sampling from this distribution are described in Odell and Feiveson (1966), Kennedy and Gentle (1990, Section 6.5.10) and Gelfand *et al.* (1990). [This example was omitted from the main text because we have avoided use of the Wishart distribution elsewhere in the book. A slightly simpler model in which $\boldsymbol{\Sigma}$ is assumed to be diagonal is to be found as the example 'Rats' distributed with WinBUGS.]

Explain in detail how you would use the Gibbs sampler to estimate the posterior distributions of α_0 and β_0, and if possible carry out this procedure.

12. Use the Metropolis-Hastings algorithm to estimate the posterior density of the parameter η in the linkage model in Sections 9.2 and 9.3 using candidate values generated from a uniform distribution on $(0, 1)$ [cf. Tanner (1996, Section 6.5.2)].

13. Write a WinBUGS program to analyze the data on wheat yield considered towards the end of Section 2.13 and in Section 9.3.

14. In bioassays the response may vary with a covariate termed the *dose*. A typical example involving a binary response is given in the table below, where R is the number of beetles killed after five hours exposure to gaseous carbon disulphide at various concentrations (data from Bliss, 1935, quoted by Dobson, 2002, Example 7.3.1).

Dose x_i ($\log_{10}\mathrm{CS}_2\mathrm{mg\ l}^{-2}$)	Number of insects, n_i	Number killed, r_i
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Fit a logistic regression model and plot the proportion killed against dose and the fitted line.

H.10 Exercises on Chapter 10

1. Show that in importance sampling the choice

$$p(x) = \frac{|f(x)|q(x)}{\int |f(\xi)|q(\xi)\,\mathrm{d}\xi}.$$

minimizes $\mathcal{V}w(x)$ even in cases where $f(x)$ is not of constant sign.

2. Suppose that $x \sim \mathrm{C}(0,1)$ has a Cauchy distribution. It is easily shown that $\eta = \mathsf{P}(x > 2) = \tan^{-1}(\frac{1}{2})/\pi = 0.147\,583\,6$, but we will consider Monte Carlo methods of evaluating this probability.

(a) Show that if k is the number of values taken from a random sample of size n with a Cauchy distribution, then k/n is an estimate with variance $0.125\,802\,7/n$.

(b) Let $p(x) = 2/x^2$ so that $\int_x^\infty p(\xi)\,\mathrm{d}\xi = 2/x$. Show that if $x \sim \mathrm{U}(0,1)$ is uniformly distributed over the unit interval then $y = 2/x$ has the density $p(x)$ and that all values of y satisfy $y \geqslant 2$ and hence that

$$\sum_{i=1}^{n} \frac{1}{2\pi}\frac{y_i^2}{1+y_i^2}$$

gives an estimate of η by importance sampling.

(c) Deduce that if x_1, x_2, …, x_n are independent U(0, 1) variates then

$$\widehat{\eta} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2\pi}\frac{4}{4+x_i^2}$$

gives an estimate of η.

(d) Check that $\widehat{\eta}$ is an unbiased estimate of η and show that

$$\mathsf{E}\widehat{\eta}^2 = \frac{\tan^{-1}(\frac{1}{2})+\frac{2}{5}}{4\pi^2}$$

and deduce that

$$\mathcal{V}\widehat{\eta} = 0.000\,095\,5$$

so that this estimator has a notably smaller variance than the estimate considered in (a).

3. Apply sampling importance resampling starting from random variables uniformly distributed over $(0, 1)$ to estimate the mean and variance of a beta distribution Be(2, 3).

4. Use the sample found in the previous section to find a 90% HDR for Be(2, 3) and compare the resultant limits with the values found using the methodology of Section 3.1. Why do the values differ?

5. Apply the methodology used in the numerical example in Section **??** to the dataset used in both Exercise 16 on Chapter 2 and Exercise 5 on Chapter 9.

6. Find the Kullback-Leibler divergence $\mathcal{I}(q : p)$ when p is a binomial distribution B(n, π) and q is a binomial distribution B(n, ρ). When does $\mathcal{I}(q : p) = \mathcal{I}(p : q)$?

7. Find the Kullback-Leibler divergence $\mathcal{I}(q : p)$ when p is a normal distribution N(μ, ϕ) and q is a normal distribution N(ν, ψ).

8. Let p be the density $2(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$ $(x > 0)$ of the modulus $|z|$ of a standard normal distribution and let q be the density $\beta^{-1}\exp(-x/\beta)$ $(x > 0)$ of an E(β) distribution. Find the value of β such that q is as close an approximation to p as possible in the sense that $\mathcal{I}(q : p)$ is a minimum.

9. The paper by Corduneaunu and Bishop (2001) referred to in Section **??** can be found on the web at

`http://research.microsoft.com/pubs/67239/bishop-aistats01.pdf.`

Härdle's data set is available in R by going `data(faithful)`. Fill in the details of the analysis of a mixture of multivariate normals given in that section.

10. Carry out the calculations in Section 10.4 for the genetic linkage data quoted by Smith which was given in Exercise 3 on Chapter 9.

11. A group of n students sit two exams. Exam one is on history and exam two is on chemistry. Let x_i and y_i denote the ith student's score in the history and chemistry exams, respectively. The following linear regression model is proposed for the relationship between the two exam scores:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, 2, \dots, n)$$

where $\varepsilon_i \sim \mathrm{N}(0, 1/\tau)$.

Assume that α, β and τ are unknown parameters to be estimated and $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$.

Describe a reversible jump MCMC algorithm including discussion of the acceptance probability, to move between the four competing models:

1. $y_i = \alpha + \varepsilon_i$;
2. $y_i = \alpha + \beta x_i + \varepsilon_i$;
3. $y_i = \alpha + \lambda t_i + \varepsilon_i$;
4. $y_i = \alpha + \beta x_i + \lambda t_i + \varepsilon_i$.

Note that if z is a random variable with probability density function f given by

$$f(z) \propto \exp\left(-\tfrac{1}{2}A\left(z^2 - 2Bz\right)\right),$$

then $z \sim N(B, 1/A)$ [due to P. Neal].