NEWCASTLE UNIVERSITY

SCHOOL OF MATHEMATICS & STATISTICS

SEMESTER 2 2007/2008

MAS3301

Bayesian Statistics

Time allowed: 1 hour 30 minutes

Credit will be given for ALL answers to questions in Section A, and for the best TWO answers to questions in Section B. No credit will be given for other answers and students are strongly advised not to spend time producing answers for which they will receive no credit.

Marks for each question are indicated. However you are advised that marks indicate the relative weight of individual questions, they do not correspond directly to marks on the University scale.

There are FOUR questions in Section A and THREE questions in Section B.

Calculators may be used. Statistical tables will be provided.

Beta distribution

If $X \sim \text{beta}(a, b)$ then it has density

$$f(x \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

for $0 < x < 1$, $a > 0$, $b > 0$.

Also $\mathrm{E}(X) = a/(a+b)$ and $\mathrm{var}(X) = ab/\{(a+b)^2(a+b+1)\}$.

Gamma distribution

If $X \sim \text{gamma}(a, b)$ then it has density

$$f(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

for $x > 0$, $a > 0$, $b > 0$.

Also $\mathrm{E}(X) = a/b$ and $\mathrm{var}(X) = a/b^2$.

Binomial distribution

If $X \sim \text{bin}(n, \theta)$ then it has probability function

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

for $x = 0, 1, 2, \ldots, n$ and $0 < \theta < 1$.

Also $\mathrm{E}(X) = n\theta$ and $\mathrm{var}(X) = n\theta(1-\theta)$.

Geometric distribution

If $X \sim \text{geometric}(\theta)$ then it has probability function

$$f(x \mid \theta) = (1-\theta)^{x-1}\theta$$

for $x = 1, 2, \ldots$ and $0 < \theta < 1$.

Also $\mathrm{E}(X) = 1/\theta$ and $\mathrm{var}(X) = (1-\theta)/\theta^2$.

Poisson distribution

If $X \sim \text{Poisson}(\lambda)$ then it has probability function

$$f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

for $x = 0, 1, 2, \ldots,\ \lambda > 0$.

Also $\text{E}(X) = \lambda$ and $\text{var}(X) = \lambda$.

Normal distribution

If $X \sim N(\mu, \sigma^2)$ then it has density

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$.

Also $\text{E}(X) = \mu$ and $\text{var}(X) = \sigma^2$.

Dirichlet distribution

If $(X_1, \ldots, X_m) \sim \text{Dirichlet}(a_1, \ldots, a_m)$ then the joint density of $X_1, \ldots, X_m$ is

$$f(x_1, \ldots, x_m) = \frac{\Gamma(A)}{\prod_{i=1}^m \Gamma(a_i)} \prod_{i=1}^m x_i^{a_i - 1},$$

where $A = \sum_{i=1}^m a_i$, for $a_i > 0$ and $0 < x_i < 1$.

$$\text{Also} \quad \text{E}(X_i) = \frac{a_i}{A}, \qquad \text{var}(X_i) = \frac{a_i}{A(A+1)} - \frac{a_i^2}{A^2(A+1)}$$

$$\text{and} \quad \text{covar}(X_i, X_j) = -\frac{a_i a_j}{A^2(A+1)} \quad \text{where} \quad i \neq j.$$

Student's t-distribution

Let τ and μ have the following joint distribution: the marginal distribution of τ is

$$\tau \sim \text{gamma}(A/2,\ B/2)$$

and the conditional distribution of μ given τ is

$$\mu \mid \tau \sim N(M,\ [C\tau]^{-1}).$$

Note that

$$t = \frac{\mu - M}{\sqrt{s^2/C}}$$

has a Student's t-distribution on A degrees of freedom, that is $t \sim t_A$, where $s^2 = B/A$.

SECTION A

A1. (a) Explain briefly what is meant by saying that, in a particular Bayesian statistics problem, a prior distribution is "conjugate".

(b) In a market research study, a random sample of n members of the public is to be chosen and each person in the sample will be asked whether he or she has used a certain product. We suppose that, given the value of the unknown population proportion θ, each person chosen has probability θ of answering "Yes" and each person's answer is independent of the others. Our prior distribution for θ is a beta(a, b) distribution. Let the number answering "Yes" be x.

(i) Explain briefly why the likelihood function for θ given the data x is

$$L(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}.$$

(ii) Find the posterior distribution of θ.

(iii) Given that $a = 2$, $b = 3$, $n = 50$ and $x = 32$, evaluate the posterior mean and the posterior standard deviation.

[10 marks]

A2. A sample of n portable CD players is tested. Each player is subjected to vibration and played continuously until it fails. The failure times, in hours, are $t_1, t_2, \ldots, t_n$. We suppose that, given the value of a parameter θ, the failure times are independent and each has an exponential distribution with probability density function

$$f_t(t \mid \theta) = \begin{cases} 0 & (t < 0) \\ \theta e^{-\theta t} & (0 \le t < \infty). \end{cases}$$

Our prior distribution for θ is a gamma(a, b) distribution with mean 0.01 and standard deviation 0.005.

(a) Find the values of a and b.

(b) Write down the likelihood function for θ given the data $t_1, t_2, \ldots, t_n$.

(c) The data are such that $n = 10$ and

$$\sum_{i=1}^{10} t_i = 1347.$$

(i) Find the posterior distribution of θ.

(ii) Find values k_1, k_2 such that, in the posterior distribution,

$$\Pr(\theta < k_1) = \Pr(\theta > k_2) = 0.05.$$

You may use the fact that, if $X \sim$ gamma(a, b) and $Y = 2bX$, then $Y \sim \chi^2_{2a}$.

[10 marks]

A3. Apples are inspected to see whether they show a blemish of a particular kind. We suppose that, given the value of the unknown population proportion θ, each apple chosen has probability θ of showing the blemish and each apple is independent of the others. Our prior distribution for θ has the following probability density function.

$$f(\theta) = \begin{cases} k_0(1-\theta^2) & (0 < \theta < 1) \\ 0 & \text{otherwise.} \end{cases}$$

A sample of twelve apples is examined and, of these, five show the blemish.

(a) Find the value of k_0.

(b) Find a function $h(\theta)$ such that the posterior density for θ is $k_1 h(\theta)$ for some constant k_1.

(c) Find the value of k_1.

[10 marks]

A4. (a) Explain the term "Bayes factor" in the context of using data to compare two simple hypotheses.

(b) In a plant breeding experiment, plants are grown from the seeds resulting from a cross-pollination. Each plant will either produce red flowers or purple flowers. Given the value of a parameter θ, let the probability that red flowers are produced be θ in each case and suppose that the plants are independent given θ. Depending on the unknown genetic makeup of the parent plants, the value of θ can be either 0.50 or 0.75. Consider the hypotheses

$$\begin{aligned} H_1 &: \Pr(\text{red}) = 0.50, \\ H_2 &: \Pr(\text{red}) = 0.75. \end{aligned}$$

Ten plants are observed and, of these, four produced red flowers.

(i) Find the Bayes factor in favour of H_1.

(ii) Find the smallest prior probability for H_1 which would give a posterior probability of H_1 greater than the posterior probability of H_2.

[10 marks]

SECTION B

B5. (a) (i) State, but do not prove, the factorisation theorem with reference to sufficiency of a statistic t for the parameter θ of a model, given data y.

(ii) Suppose that we observe data, $y_1, \ldots, y_n$. These are observations on variables which, given the values of μ, τ, are independent and from a normal distribution with mean μ and variance τ^{-1}. Show that $\bar{y}$ and S_d are sufficient statistics for μ and τ, given data $y_1, \ldots, y_n$, where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{and} \qquad S_d = \sum_{i=1}^{n}(y_i - \bar{y})^2.$$

(b) Orange juice is put into cartons by a machine. The nominal amount of juice in each carton is 1 litre (i.e. 1000ml). It is known that the actual quantities vary by small amounts and, in fact, given the value of a parameter μ, the amounts, in ml, put into the cartons are independent and, to a good approximation, normally distributed with mean μ and standard deviation 0.5. It is thought that the machine may need to be adjusted. The value of μ is unknown. The prior distribution for μ is normal with mean 1000 and standard deviation 2.

The amounts of juice y_i in ten cartons are measured and found to be (in ml)

1002.7	1002.7	1002.5	1001.2	1002.3
1002.4	1002.3	1002.3	1001.8	1002.0

These give $\bar{y} = 1002.22$ and $S_d = 1.856$, where $\bar{y}$ and S_d are defined as above.

(i) The posterior distribution of μ is a normal distribution. Find its mean and variance.

(ii) Find the posterior 95% hpd (highest probability density) interval for μ.

(iii) Find the probability, after having seen the data, that a further carton will contain more than 1002 ml of juice.

(c) Consider the same situation as in part (b) except that, this time, the population standard deviation of the amounts is unknown. Given the values of parameters μ, τ, the amounts, in ml, put into cartons are independent and, to a good approximation, normally distributed with mean μ and standard deviation σ, where $\sigma^2 = \tau^{-1}$. The values of μ and τ are unknown. The prior distribution of τ is a gamma(4, 1) distribution and the conditional prior distribution of μ given τ is a normal distribution with mean 1000 and variance $16\tau^{-1}$. The data are as in part (b).

(i) Find the marginal posterior distribution of μ.

(ii) Find the 95% marginal posterior hpd (highest probability density) interval for μ.

(iii) Find the 95% hpd interval, after having seen the data, for the amount of juice in a further carton.

You may use the following formulae, given in the notation used in the course:

$$
\begin{aligned}
A_1 &= A_0 + n, \\
B_1 &= B_0 + B_d, \\
C_1 &= C_0 + n, \\
M_1 &= \frac{C_0 M_0 + n\bar{y}}{C_0 + n}, \\
B_d &= S_d + R, \\
R &= \left(\frac{1}{C_0} + \frac{1}{n}\right)^{-1} (\bar{y} - M_0)^2.
\end{aligned}
$$

[30 marks]

B6. (a) Explain what is meant by saying that a continuous prior distribution for a scalar parameter is "proper".

(b) A sample of n observations, $y_1, \ldots, y_n$, is taken where, given the value of a parameter θ, the observations are independent and from a Poisson distribution with mean θ.

The observed data, with $n = 10$, are

45	52	51	48	62
63	57	66	57	57

Question B6 continued on next page

(i) Find the Jeffreys prior distribution in this case.

(ii) State whether this Jeffreys prior distribution is proper and give the reason for your answer.

(iii) Find the posterior distribution when this Jeffreys prior distribution is used.

(iv) Find the posterior distribution for θ when the prior distribution is gamma$(1, 0.01)$.

(v) Find approximately the posterior distribution when the prior distribution is uniform on the interval $(0, 200)$.

(vi) Briefly compare these three prior distributions and their corresponding posterior distributions. Give a possible advantage and a possible disadvantage for each prior distribution.

(c) A sample of n observations, $y_1, \ldots, y_n$, where n is large, is taken where, given the value of a parameter θ, the observations are independent and from a Poisson distribution with mean $\exp(\theta)$.

The prior distribution for θ is such that the prior density may be considered to be effectively constant over the region of interest and it has negligible effect on the posterior distribution when n is large.

(i) Show that the likelihood for θ given the data $y_1, \ldots, y_n$ is

$$L(\theta) = \frac{\exp\left\{n\left(\theta\bar{y} - e^{\theta}\right)\right\}}{\prod_{i=1}^{n} y_i!}$$

where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

(ii) Find an approximate posterior distribution for θ.

(iii) Evaluate an approximate 95% posterior hpd (highest probability density) interval for θ when $n = 100$ and $\sum_{i=1}^{n} y_i = 9453$.

[30 marks]

B7. (a) Our beliefs about the unknown value of a scalar parameter θ are expressed by the continuous probability distribution with probability density function $f_\theta(\theta)$. The distribution of an observable quantity X depends on θ and the conditional distribution of X given θ has probability density function $f_X(x \mid \theta)$. Write down an expression for the probability density function of the predictive distribution of X.

(b) Our beliefs about the unknown value of a scalar parameter θ are represented by a gamma(a, b) distribution. Given the value of θ, the variables T_1 and T_2 are independent and each has an exponential distribution with probability density function

$$f_T(t \mid \theta) = \theta e^{-\theta t} \qquad\qquad (0 < t < \infty).$$

(i) Show that the predictive probability density function of T_1 is

$$f_{\text{pred}}(t) = \frac{a}{b}\left(\frac{b}{b+t}\right)^{a+1} \qquad\qquad (0 < t < \infty).$$

(ii) Find the joint predictive probability density function of T_1 and T_2.

(c) (i) Show that, when θ has a beta(a, b) distribution,

$$\mathrm{E}(\theta^2) = \frac{\Gamma(a+b)\Gamma(a+2)}{\Gamma(a+b+2)\Gamma(a)} = \frac{(a+1)a}{(a+b+1)(a+b)}.$$

(ii) Our distribution for a scalar parameter θ, where $0 < \theta < 1$, has a probability density function which is proportional to

$$k_1\theta^{a_1-1}(1-\theta)^{b_1-1} + k_2\theta^{a_2-1}(1-\theta)^{b_2-1}$$

for $0 < \theta < 1$ and which is zero otherwise. Show that, for $0 < \theta < 1$, the probability density function has the form

$$f(\theta) = p_1 f_1(\theta) + p_2 f_2(\theta),$$

where $f_1(\theta)$ and $f_2(\theta)$ are the probability density functions of a beta(a_1, b_1) distribution and a beta(a_2, b_2) distribution respectively, and that, for $j = 1, 2$,

$$p_j = \frac{q_j}{q_1 + q_2}$$

Question B7 continued on next page

and
$$q_j = k_j \frac{\Gamma(a_j)\Gamma(b_j)}{\Gamma(a_j + b_j)}.$$

(d) Given the value of a parameter θ, the number of people in a random sample of n from the population who have a particular gene has a binomial(n, θ) distribution. The prior distribution for θ has probability density function

$$f^{(0)}(\theta) = \begin{cases} k_0[1 + (1-\theta)^2] & (0 \le \theta \le 1) \\ 0 & (\text{otherwise}). \end{cases}$$

(i) Find the value of k_0.

(ii) Show that this prior distribution is a mixture of two beta distributions with probability density function

$$f^{(0)}(\theta) = p_1^{(0)} f_1^{(0)}(\theta) + p_2^{(0)} f_2^{(0)}(\theta),$$

where $f_1^{(0)}(\theta)$ and $f_2^{(0)}(\theta)$ are beta probability density functions, and find the two beta distributions and the values of $p_1^{(0)}$ and $p_2^{(0)}$.

(iii) Find the prior mean and prior variance of θ.

(iv) In a small pilot study, a sample with $n = 10$ is taken and the observed number with the gene is $x = 4$.
Find the posterior distribution of θ and express its probability density function in the form

$$f^{(1)}(\theta) = p_1^{(1)} f_1^{(1)}(\theta) + p_2^{(1)} f_2^{(1)}(\theta),$$

where $f_1^{(1)}(\theta)$ and $f_2^{(1)}(\theta)$ are beta probability density functions, and find the two beta distributions and the values of $p_1^{(1)}$ and $p_2^{(1)}$.

[**30 marks**]

THE END