MAS3301 Bayesian Statistics

M. Farrow School of Mathematics and Statistics Newcastle University

Semester 2, 2008-9

13 Sequential updating

13.1 Theory

We have seen how we can change our beliefs about an unknown parameter θ from a prior distribution with density $f^{(0)}(\theta)$ to a posterior distribution with density $f^{(1)}(\theta)$ when we observe data y, using

$$f^{(1)}(\theta) \propto f^{(0)}(\theta) L(\theta; y)$$

where the likelihood $L(\theta; y) = f_Y(y \mid \theta)$, the probability or probability density of y given θ .

Now suppose that we repeat the process. We start with a prior distribution with density $f^{(0)}(\theta)$. Then we observe data $y^{(1)}$ and our beliefs change so that they are represented by $f^{(1)}(\theta)$. Suppose that we now observe further data $y^{(2)}$ so that our density for θ becomes $f^{(2)}(\theta)$.

Before we observe either $y^{(1)}$ or $y^{(2)}$, the joint probability (density) of θ , $y^{(1)}$, $y^{(2)}$ is

$$f^{(0)}(\theta)f_1(y^{(1)} \mid \theta)f_2(y^{(2)} \mid \theta, y^{(1)})$$

where $f_1(y^{(1)} | \theta)$ is the conditional probability (density) of $y^{(1)}$ given θ and $f_2(y^{(2)} | \theta, y^{(1)})$ is the conditional probability (density) of $y^{(2)}$ given θ and $y^{(1)}$.

Clearly

$$f^{(1)}(\theta) \propto f^{(0)}(\theta) f_1(y^{(1)} \mid \theta) = f^{(0)}(\theta) L_1(\theta; y^{(1)})$$

and

$$\begin{aligned} f^{(2)}(\theta) &\propto f^{(0)}(\theta) f_1(y^{(1)} \mid \theta) f_2(y^{(2)} \mid \theta, y^{(1)}) = f^{(0)}(\theta) L_{1,2}(\theta; y^{(1)}, y^{(2)}) \\ &\propto f^{(1)}(\theta) f_2(y^{(2)} \mid \theta, y^{(1)}) = f^{(1)}(\theta) L_2(\theta; y^{(2)} \mid y^{(1)}) \end{aligned}$$

In many cases $y^{(1)}$ and $y^{(2)}$ will be independent given θ . In such a case $f_2(y^{(2)} \mid \theta, y^{(1)}) = f_2(y^{(2)} \mid \theta)$ and

$$L_{1,2}(\theta; \ y^{(1)}, y^{(2)}) = L_1(\theta; \ y^{(1)}) L_2(\theta; \ y^{(2)}).$$
(11)

Our "posterior" distribution after observing $y^{(1)}$ is our "prior" distribution before we observe $y^{(2)}$. The only complication when $y^{(1)}$ and $y^{(2)}$ are not independent given θ is that the second likelihood $L_2(\theta; y^{(2)} | y^{(1)})$ must allow for the dependence of $y^{(2)}$ on $y^{(1)}$.

Note that $f^{(2)}(\theta)$ is the same whether we update our beliefs first by $y^{(1)}$ and then by $y^{(2)}$ or first by $y^{(2)}$ and then by $y^{(1)}$. This is obvious when $y^{(1)}$ and $y^{(2)}$ are independent given θ and (11) applies. In the case where they are not it is simply a matter of whether we factorise the joint probability (density) of $y^{(1)}$ and $y^{(2)}$ given θ into $f_1(y^{(1)} | \theta)f_2(y^{(2)} | \theta, y^{(1)})$, as above, or into $f_2^*(y^{(2)} | \theta)f_1^*(y^{(1)} | \theta, y^{(2)})$.

13.2 Examples

13.2.1 Example 1: Binomial

Consider the example of section 3.4 where each animal may have a particular gene or not. The prior pdf $f^{(0)}(\theta)$ is proportional to

 $\theta^{a-1}(1-\theta)^{b-1}.$

We observe 20 animals, 3 of which have the gene, giving a likelihood

$$L_1(\theta; y^{(1)}) \propto \theta^3 (1-\theta)^{17}.$$

Thus the posterior pdf at this stage, $f^{(1)}(\theta)$, is proportional to

$$\theta^{a+3-1}(1-\theta)^{b+17-1},$$

the pdf of a beta(a + 3, b + 17) distribution.

Now suppose that we examine another 30 animals and 7 of these have the gene. These observations are reasonably supposed to be independent of $y^{(1)}$ given θ . So

$$L_2(\theta; y^{(2)}) \propto \theta^7 (1-\theta)^{23}$$

and our new pdf for θ , $f^{(2)}(\theta)$, is proportional to

$$\theta^{a+10-1}(1-\theta)^{b+40-1}$$
.

the pdf of a beta(a + 10, b + 40) distribution.

13.2.2 Example 2: Sampling inspection

This is a more complicated example, in which $y^{(1)}$ and $y^{(2)}$ are not independent given θ . Consider the acceptance sampling example in section 6.5. A batch of m items is manufactured. It contains an unknown number, d, of defectives. We inspect a sample of $n_1 < m$ of the items and find r_1 defectives among them. Suppose now that we select a further sample of $n_2 < m - n_1$ items from the remaining $m - n_1$ items in the batch and r_2 defectives are found in this new sample. Note that the unknown quantity here is d which therefore plays the role of θ .

The likelihoods are

$$L_1(d; r_1) = \frac{\binom{d}{r_1}\binom{m-d}{n_1-r_1}}{\binom{m}{n_1}}$$

for the first sample and

$$L_2(d; r_2 | r_1) = \frac{\begin{pmatrix} d - r_1 \\ r_2 \end{pmatrix} \begin{pmatrix} (m - d) - (n_1 - r_1) \\ n_2 - r_2 \end{pmatrix}}{\begin{pmatrix} m - n_1 \\ n_2 \end{pmatrix}}$$

for the second sample. Notice that $L_2(d; r_2 | r_1)$ does depend on r_1 in this case. Observe also that the overall likelihood is

$$\begin{split} L_{1,2}(d; r_1, r_2) &= \frac{\binom{d}{r_1}\binom{m-d}{n_1-r_1}}{\binom{m}{n_1}} \frac{\binom{d-r_1}{r_2}\binom{(m-d)-(n_1-r_1)}{n_2-r_2}}{\binom{m-n_1}{n_2}} \\ &= \frac{d!}{r_1!(d-r_1)!} \times \frac{(m-d)!}{(n_1-r_1)![(m-d)-(n_1-r_1)]!} \times \left\{\frac{m!}{n_1!(m-n_1)!}\right\}^{-1} \\ &\times \frac{(d-r_1)!}{r_2!(d-r_1-r_2)!} \times \frac{[(m-d)-(n_1-r_1)]!}{(n_2-r_2)![(m-d)-(n_1+n_2-r_1-r_2)]!} \\ &\times \left\{\frac{(m-n_1)!}{n_2!(m-n_1-n_2)!}\right\}^{-1} \end{split}$$

Deleting factors which do not involve d we see that

$$\begin{split} L_{1,2}(d; \ r_1, r_2) &\propto \frac{d!}{(r_1 + r_2)!(d - r_1 - r_2)!} \\ &\times \frac{(m - d)!}{(n_1 + n_2 - r_1 - r_2)![(m - d) - (n_1 + n_2 - r_1 - r_2)]!} \\ &\times \left\{ \frac{m!}{(n_1 + n_2)!(m - n_1 - n_2)!} \right\}^{-1} \\ &\propto \frac{\left(\begin{array}{c} d \\ r_1 + r_2 \end{array} \right) \left(\begin{array}{c} m - d \\ n_1 + n_2 - r_1 - r_2 \end{array} \right)}{\left(\begin{array}{c} m \\ n_1 + n_2 \end{array} \right)}, \end{split}$$

the likelihood from a single sample of $n_1 + n_2$ containing $r_1 + r_2$ defectives.

14 Inference for Normal Distributions I

14.1 Introduction

The normal distribution is, of course, an important model in statistics. So far we have not considered inference for a normal distribution but have established the general principles of Bayesian inference using examples with other distributions. The normal distribution, of course, has two parameters and therefore we need a bivariate prior distribution.

In this lecture we will consider inference when we have a sample from a single normal population and we know the value of the variance parameter. In the following lectures we will consider more complicated problems.

14.2 Precision

In non-Bayesian statistics we usually think of the two parameters of a normal distribution as the mean (usually μ) and the variance (usually σ^2). In Bayesian statistics it is often more convenient to work in terms of the mean and the *precision*. The precision is just the reciprocal of the variance. Thus, for a normal $N(\mu, \sigma^2)$ distribution, the precision is

$$\tau = \sigma^{-2}$$

and we would often write the distribution as $N(\mu, \tau^{-1})$. Using precision τ the probability density function becomes

$$f(y \mid \mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(y-\mu)^2\right\}.$$

14.3 Likelihood

Suppose we are going to observe data y_1, \ldots, y_n where $Y_i \sim N(\mu, \tau^{-1})$ and, given μ, τ , the observations are independent.

The likelihood is

$$L(\mu, \tau; y) = \prod_{i=1}^{n} \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\tau}{2}(y_{i}-\mu)^{2}\right\}$$
$$= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_{i}-\mu)^{2}\right\}$$
$$= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_{i}-\bar{y}+\bar{y}-\mu)^{2}\right\}$$
$$= \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{n}(y_{i}-\bar{y})^{2}+n(\bar{y}-\mu)^{2}\right]\right\}$$

$$L(\mu, \tau; y) = \exp\left\{-\frac{n\tau}{2}(\bar{y}-\mu)^2\right\} \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}S_d\right\}$$
(12)
where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$ and $S_d = \sum_{i=1}^n (y_i - \bar{y})^2.$

14.4 Inference with known precision, normal prior for mean

In non-Bayesian statistics, when the variance is known tests and confidence intervals are based on the normal distribution. When the variance is not known the tests and confidence intervals have to be based on Student's *t*-distribution. As we will see, a similar situation applies in Bayesian statistics. First we will deal with the case where the variance, or precision, is known.

Suppose that our prior distribution for μ is normal with prior mean M_0 and prior precision P_0 . That is, it is $N(M_0, P_0^{-1})$. This is a conjugate prior and the posterior distribution is normal, as follows.

Posterior : In this case the posterior distribution is normal with mean M_1 and precision P_1 where

$$P_1 = P_0 + P_d,$$

$$M_1 = \frac{P_0 M_0 + P_d \bar{y}}{P_0 + P_d}$$
where $P_d = n\tau = \left(\frac{\sigma^2}{n}\right)^{-1}$

Notice that the posterior precision is the sum of the prior precision and the "data precision" P_d where the data precision is the reciprocal of the sampling variance of \bar{Y} and the posterior mean is a weighted average of the prior mean and the sample mean, with weights given by the prior precision and the data precision. Thus the amounts of weight we give to our prior beliefs about μ and the evidence from the data depend on the relative sizes of our prior precision (measuring our prior certainty) and the data precision (reflecting the sample size and the "error" precision).

Proof : The proof is straightforward.

The prior density is proportional to

$$\exp\left\{-\frac{P_0}{2}(\mu-M_0)^2\right\}$$

The likelihood is proportional to

$$\exp\left\{-\frac{n\tau}{2}(\bar{y}-\mu)^2\right\}.$$

The posterior density is proportional to

$$\exp\left\{-\frac{P_{0}}{2}(\mu-M_{0})^{2}\right\} \times \exp\left\{-\frac{n\tau}{2}(\bar{y}-\mu)^{2}\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[P_{0}(\mu-M_{0})^{2}+n\tau(\bar{y}-\mu)^{2}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[(P_{0}+n\tau)\mu^{2}-2(P_{0}M_{0}+n\tau\bar{y})\mu+P_{0}M_{0}^{2}+n\tau\bar{y}^{2}\right]\right\}$$

$$\propto \exp\left\{-\frac{(P_{0}+n\tau)}{2}\left[\mu^{2}-2\left(\frac{P_{0}M_{0}+n\tau\bar{y}}{P_{0}+n\tau}\right)\mu+\left(\frac{P_{0}M_{0}+n\tau\bar{y}}{P_{0}+n\tau}\right)^{2}\right]\right\}$$

$$\propto \exp\left\{-\frac{(P_{0}+n\tau)}{2}(\mu-M_{1})^{2}\right\}$$

Example : Suppose our prior mean for μ is $M_0 = 10$ and our prior standard deviation for μ is 4.

A prior 95% interval for μ would then be $10\pm1.96\times4.$ That is 2.16 $<\mu<17.84.$ The prior variance is $4^2 = 16$ and the prior precision $P_0 = 1/16 = 0.0625$. Suppose the error standard deviation is $\sigma = 0.8$ so $\sigma^2 = 0.64$ and $\tau = 1/0.64 = 1.5625$. Suppose that n = 15 and $\sum(y) = 94.5$. Then $\bar{y} = 94.5/15 = 6.3$. So the posterior precision is

 $P_1 = 0.0625 + 15 \times 1.5625 = 23.5$

so the posterior variance is 1/23.5 = 0.04255 and the posterior standard deviation is $\sqrt{0.04255} =$ 0.2063.

The posterior mean is

$$M_1 = \frac{0.0625 \times 10 + 15 \times 1.5625 \times 6.3}{23.5} = 6.3098.$$

A 95% posterior interval for μ is $6.3098 \pm 1.96 \times 0.2063$. That is $5.9055 < \mu < 6.7142$.

14.5 Problems 4

1. I recorded the attendance of students at tutorials for a module. Suppose that we can, in some sense, regard the students as a sample from some population of students so that, for example, we can learn about the likely behaviour of next year's students by observing this year's. At the time I recorded the data we had had tutorials in Week 2 and Week 4. Let the probability that a student attends in both weeks be θ_{11} , the probability that a student attends in week 2 but not Week 4 be θ_{10} and so on. The data are as follows.

Attendance	Probability	Observed frequency
Week 2 and Week 4	$ heta_{11}$	$n_{11} = 25$
Week 2 but not Week 4	$ heta_{10}$	$n_{10} = 7$
Week 4 but not Week 2	θ_{01}	$n_{01} = 6$
Neither week	$ heta_{00}$	$n_{00} = 13$

Suppose that the prior distribution for $(\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$ is a Dirichlet distribution with density proportional to

$$\theta_{11}^3 \theta_{10} \theta_{01} \theta_{00}^2.$$

- (a) Find the prior means and prior variances of θ_{11} , θ_{10} , θ_{01} , θ_{00} .
- (b) Find the posterior distribution.
- (c) Find the posterior means and posterior variances of θ_{11} , θ_{10} , θ_{01} , θ_{00} .
- (d) Using the R function hpdbeta which may be obtained from the Web page (or otherwise), find a 95% posterior hpd interval, based on the exact posterior distribution, for θ_{00} .
- 2. Suppose that we have J samples and, given the parameters, observation i in sample j is

$$y_{i,j} \sim N(\mu_j, \tau^{-1})$$

for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$. Let $\underline{\mu} = (\mu_1, \ldots, \mu_J)^T$, let $\underline{\bar{y}} = (\bar{y}_1, \ldots, \bar{y}_J)^T$, and let

$$S = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{i,j} - \bar{y}_j)^2,$$

where

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{i,j}.$$

Show that \bar{y} and S are sufficient for μ and τ .

3. We make *n* observations y_1, \ldots, y_n , which, given that values of parameters α , β , are independent observations from a gamma(α, β) distribution. Show that the statistics T_1 , T_2 are sufficient for α , β where

$$T_1 = \sum_{i=1}^n y_i \qquad \text{and} \qquad T_2 = \prod_{i=1}^n y_i$$

4. Davies and Goldsmith (1972) give the following data on piston ring failures in steam-driven compressors. There were four identical compressors in the same compressor house, each oriented the same way, and each had three legs. The data give the number of failures in each leg of each compressor over a period of some years.

Compressor	North	Centre	South
Number	Leg	Leg	Leg
1	17	17	12
2	11	9	13
3	11	8	19
4	14	7	28

Let the number of failures in leg j (North: j = 1, Centre: j = 2, South: j = 3) of compressor i be $X_{i,j}$. Suppose that we regard the total number of failures, N = 166, as fixed and regard the numbers $X_{1,1}, \ldots, X_{4,3}$ as being an observation from a multinomial $(N, \theta_{1,1}, \ldots, \theta_{4,3})$ distribution. Suppose that our prior distribution for $\theta_{1,1}, \ldots, \theta_{4,3}$ is a Dirichlet $(a_{1,1}, \ldots, a_{4,3})$ distribution with $a_{i,j} = 2.0$ for all i, j.

- (a) Find the posterior distribution for $\theta_{1,1}, \ldots, \theta_{4,3}$.
- (b) Find the posterior mean for each of $\theta_{1,1}, \ldots, \theta_{4,3}$.
- (c) For each of $\theta_{1,1}, \ldots, \theta_{4,3}$, find the symmetric 95% posterior interval, compare these intervals and comment.

<u>Note</u>: A symmetric 95% interval for θ is simply an interval (k_1, k_2) such that $\Pr(\theta < k_1) = \Pr(\theta > k_2) = 0.025$. You will need to use R to evaluate these intervals.

5. Ten measurements are made using a scientific instrument. Given the unknown value of a quantity θ , the natural logarithms of the measurements are independent and normally distributed with mean log θ and known standard deviation 0.05.

Our prior distribution is such that $\log \theta$ has a normal distribution with mean 2.5 and standard deviation 0.5.

The logarithms of the measurements are as follows.

 $2.99 \quad 3.03 \quad 3.04 \quad 3.01 \quad 3.12 \quad 2.98 \quad 3.03 \quad 2.98 \quad 3.07 \quad 3.10$

- (a) Find the posterior distribution of $\log \theta$.
- (b) Find a symmetric 95% posterior interval for $\log \theta$.
- (c) Find a symmetric 95% posterior interval for θ .
- (d) Find the posterior probability that $\theta < 20.0$.
- 6. Walser (1969) gave the following data on the month of giving birth for 700 women giving birth for the first time. The births took place at the University Hospital of Basel, Switzerland.

Month No. of births		Month		No. of births	Month		No. of births	
1	January	66	5	May	64	9	September	54
2	February	63	6	June	74	10	October	51
3	March	64	7	July	70	11	November	45
4	April	48	8	August	59	12	December	42

We have unknown parameters $\theta_1, \ldots, \theta_{12}$ where, given the values of these parameters, the probability that one of these births takes place in month j is θ_j and January is month 1, February is month 2 and so on through to December which is month 12. Given the parameters, the birth dates are assumed to be independent.

Our prior distribution for $\theta_1, \ldots, \theta_{12}$ is a Dirichlet distribution with parameters $a_1 = a_2 = \cdots = a_{12} = 2$.

- (a) Find the posterior distribution of $\theta_1, \ldots, \theta_{12}$.
- (b) For each of j = 1, ..., 12, find the posterior mean of θ_j .
- (c) For each of j = 1, ..., 12, find the posterior probability that $\theta_j > 1/12$ and comment on the results.
- (d) Find the joint posterior distribution of $\theta_1, \theta_2, \tilde{\theta}_2$, where $\tilde{\theta}_2 = 1 \theta_1 \theta_2$.

Note: You may use R for the calculations but give the commands which you use with your solution.

- 7. Potatoes arrive at a crisp factory in large batches. Samples are taken from each batch for quality checking. Assume that each potato can be classified as "good" or "bad" and that, given the value of a parameter θ , potatoes are independent and each has probability θ of being "bad."
 - (a) Suppose that m samples, each of fixed size n, are chosen and that the numbers of bad potatoes found are x_1, \ldots, x_m . Show that

$$s = \sum_{i=1}^{m} x_i$$

is sufficient for θ .

(b) Suppose that potatoes are examined one at a time until a fixed number r of bad potatoes is found. Let the number of potatoes examined when the r^{th} bad potato is found be y. This process is repeated m times and the values of y are y_1, \ldots, y_m . Show that

$$t = \sum_{i=1}^{m} y_i$$

is sufficient for θ .

- (c) Suppose that we have a prior distribution for θ which is a beta(a, b) distribution. A two-stage inspection procedure is adopted. In Stage 1 potatoes are examined one at a time until a fixed number r of bad potatoes is found. The rth bad potato found is the yth to be examined. In Stage 2 a further n potatoes are examined and x of these are found to be bad.
 - i. Find the posterior distribution of θ after Stage 1.
 - ii. Find the posterior distribution of θ after Stage 1 and Stage 2.

Homework 4

Solutions to Questions 5, 6, 7 of Problems 4 are to be submitted in the Homework Letterbox no later than 4.00pm on Monday April 20th.

Reference

- Davies, O.L. and Goldsmith, P.L. (eds.), 1972. Statistical Methods in Research and Production, 4th edition, Oliver and Boyd Edinburgh.
- Walser, P., 1969. Untersuchung über die Verteilung der Geburtstermine bei die mehrgebärenden Frau. Helvetica Paediatrica Acta. Suppl. XX ad vol. 24, fasc. 3, 1-30.