MAS3301 Bayesian Statistics: Project

Semester 2

2008-9

1 Background

When analysing proportions (eg. with a binomial model) we can often use a beta prior distribution. However sometimes, particularly in more complicated cases, it is convenient to use a different form of prior distribution. Typically we transform the proportion to put it onto a $(-\infty,\ \infty)$ scale, rather than $(0,\ 1)$, and then give the transformed proportion a normal distribution. In particular this makes it easy to give a relationship to two or more proportions in our prior distribution.

One form of transformation which is often used is known as "logits." If our proportion is θ then we transform this to

$$\eta = \log\left(\frac{\theta}{1-\theta}\right) \tag{1}$$

and then give η a normal prior distribution.

This project is concerned with analyses of this type.

2 Data

Each student will use a different data set. Each data set is identified by a reference number. A list of students and reference numbers and a separate list giving the data set for each reference number are provided in sections 5 and 6 below.

3 Tasks

1. Find an expression for the pdf of θ when η is given by (1) and $\eta \sim N(m,\ v)$. You may wish to refer to Section 5.4 of the lecture notes.
2. Many people regard a beta$(1,1)$ distribution as a "noninformative" prior since the pdf is a constant. It might be thought that, if we use a logit transformation and then give η a $N(0,\ v)$ prior distribution, then this will be "noninformative" if we make the variance v large. However things are not as simple as this because, for large enough v, the prior distribution for θ becomes bimodal.
   (a) Use R to plot graphs of the pdf of θ when η is given by (1) and $\eta \sim N(0,\ v)$ for a range of values of v in $1 \le v \le 4$ and hence deduce, at least approximately, the value of v at which the distribution of θ becomes bimodal.
   (b) Find analytically the exact value of v at which the distribution of θ becomes bimodal when $\eta \sim N(0,\ v)$. *Hint: The density of θ is symmetric about $\theta = 1/2$ so look at the second derivative of the log density at this point.*
3. Each patient in a sample of n_1 patients with a certain chronic illness is given a treatment. The number x_1 who show a particular response is recorded. Your values of n_1 and x_1 are in your data set. Given the value of a parameter θ_1, we regard x_1 as an observation from the binomial(n_1, θ_1) distribution. The value of θ_1 is, however, unknown so we give it a prior

distribution by first transforming to η_1 using (1) and then giving η_1 a normal $n(m_1,\ v_1)$ prior distribution.

(a) In our prior beliefs, the lower quartile of θ_1 is 0.4 and the upper quartile is 0.8. Use these to find the lower and upper quartiles of η_1 and hence find the values of m_1 and v_1.

(b) Use numerical methods to find the posterior density of θ_1 given your data and plot a graph showing both the prior and posterior densities.

(c) Find the posterior probability that $\theta_1 > 0.75$.

4. In another experiment with the same disease two further groups of patients are used. Group 2 contains n_2 patients and is given treatment T_2. Group 3 contains n_3 patients and is given treatment T_3. The number in Group j who show the response is x_j. Given the values of parameters $\theta_2,\ \theta_3,\ x_j$ is regarded as an observation from the binomial$(n_j,\ \theta_j)$ distribution and x_2, x_3 are independent.

   Our prior distribution for $\theta_2,\ \theta_3$ is as follows. We transform $\theta_2,\ \theta_3$ to $\eta_2,\ \eta_3$ using (1) and then give $\eta_1,\ \eta_2$ a bivariate normal prior distribution with parameters as follows.

$$\begin{aligned} \mathrm{E}(\eta_2) = \mathrm{E}(\eta_3 &= 0.4 \\ \mathrm{var}(\eta_2) = \mathrm{var}(\eta_3) &= 0.6 \\ \mathrm{covar}(\eta_2, \eta_3) &= 0.3 \end{aligned}$$

(a) Use R to find and plot the joint posterior density of θ_2 and θ_3.

(b) Find and plot the prior and posterior probability density functions of the log relative risk γ where

$$\begin{aligned} \gamma &= \log\left(\frac{\theta_2/(1-\theta_2)}{\theta_3/(1-\theta_3)}\right) \\ &= \log\left(\frac{\theta_2}{1-\theta_2}\right) - \log\left(\frac{\theta_3}{1-\theta_3}\right) \\ &= \eta_2 - \eta_3 \end{aligned}$$

(c) Comment on your results.

Hint: Part 4 of Practical 1 should help with Task 4. You will need to modify the R functions appropriately.

4 Submission

Reports are to be submitted to the General Office of the School of Mathematics and Statistics no later than 4.00pm on **Thursday 30th April** (Week 10). As this counts for 10% of the module mark you will need to hand in your report at the reception window.

5 Data Set Reference Numbers

Agnew,	Thomas	1
Armstrong,	Suzanne	2
Askew,	Leanne	3
Bannister,	John Jeffrey	4
Batey,	Aidan Joseph	5
Brooks,	Ciaran Anthony Foley	6
Brown,	Kate Elizabeth	7
Browne,	Colin John	8
Busby,	John	9
Chan,	Athena	10
Cheung,	Ho Yee	11
Cheyne-Vidal,	Nicola	12
Clarkson,	John-Frederic	13
Curwen,	Robert Anthony	14
D'Souza,	Francesca Kate	15
Dew,	Christopher	16
Egan,	Josephine	17
Ferguson,	Katy	18
Garvey,	Emma Jayne	19
Gill,	Martin	20
Goldthorpe,	Rowen	21
Gray,	Laura	22
Hallmark,	Laura Catherine	23
Jamison,	Deborah	24
Jayasuriya,	Gregory Anthony	25
Keene,	Stuart Peter	26
Kieselack,	Simon Nicholas	27
Lawson-Matthew,	Emma Jane	28
Longworth,	Jessica Daisy	29
MacGilchrist,	Graeme Alastair	30
Mann,	Kay Debby	31
McKinnon,	Alison	32
McParland,	Iain James	33
Millman,	Jill Fairless	34
Moncaster,	Sam	35
Munro,	Joseph	36
Nichols,	Ben	37
Payne,	Charlotte Elizabeth	38
Phillips,	Kate	39
Proom,	Rebecca Jane	40
Riley,	Anthony David	41
Roberts,	Catherine	42
Robertson,	Patrick	43
Smith,	James	44
Smith,	Warren Andrew	45
Tee,	Jane Katherine	46
Wang,	Yue	47
Wilkinson,	Nina	48
Wood,	Rachael Louise	49
Woodward,	Joe	50
Wrigley,	Amy	51

6 Data Sets

Ref.no.	n_1	x_1	n_2	x_2	n_3	x_3	Ref.no.	n_1	x_1	n_2	x_2	n_3	x_3
1	25	18	55	32	20	9	34	27	22	53	28	48	41
2	26	19	54	28	21	16	35	28	20	52	32	47	37
3	27	24	53	30	22	17	36	29	21	51	33	46	36
4	28	24	52	28	23	16	37	30	24	50	25	45	36
5	29	25	51	33	24	19	38	31	19	49	33	44	36
6	30	20	50	30	25	21	39	32	24	48	24	43	29
7	31	25	49	36	26	20	40	33	23	47	24	42	28
8	32	23	48	32	27	23	41	34	28	46	19	41	31
9	33	21	47	27	28	23	42	35	28	45	29	40	34
10	34	28	46	19	29	21	43	36	28	44	30	39	29
11	35	22	45	35	30	26	44	37	27	43	23	38	31
12	36	25	44	19	31	24	45	38	30	42	23	37	31
13	37	24	43	24	32	29	46	39	32	41	25	36	27
14	38	30	42	24	33	29	47	40	28	40	24	35	28
15	39	30	41	27	34	27	48	41	31	39	25	34	24
16	40	30	40	22	35	27	49	42	34	38	23	33	27
17	41	28	39	19	36	31	50	43	35	37	23	32	27
18	42	34	38	19	37	28	51	44	30	36	24	31	24
19	43	31	37	23	38	30	52	45	29	35	23	30	28
20	44	31	36	21	39	29	53	46	35	34	19	29	25
21	45	39	35	18	40	28	54	47	34	33	17	28	27
22	46	34	34	19	41	34	55	48	40	32	15	27	22
23	47	33	33	23	42	30	56	49	34	31	23	26	16
24	48	35	32	24	43	31	57	50	40	30	19	25	22
25	49	36	31	15	44	34	58	51	33	29	16	24	18
26	50	44	30	22	45	37	59	52	37	28	16	23	16
27	51	34	29	21	46	37	60	53	38	27	16	22	17
28	52	42	28	18	47	36	61	54	39	26	14	21	20
29	53	44	27	15	48	39	62	55	38	25	19	20	18
30	54	40	26	18	49	34	63	31	27	34	15	33	28
31	55	41	25	14	50	42	64	32	25	33	17	34	27
32	25	21	55	35	50	41	65	33	26	32	19	33	26
33	26	23	54	34	49	42	66	34	26	31	16	34	28