CHL 5223H Applied Bayesian Methods

Homework 2
Monday, February 6, 2017
DUE: MONDAY, March 6, 2017

1. (Data from Kleinbaum, Kupper, Muller, and Nizam, pg108) A biologist wished to explore the relationship of temperature on the growth process of a certain type of tissue. Using the same parent batch, the scientist cultured 5 different cell lines for 4 different temperatures (for a total of 20 different lines). The total number of cells sampled for the different temperatures are given in the following table: (Note: numbers are $\times 10^6$ after 7 days.)

Temp=40	Temp=60	Temp=80	Temp=100
1.13	1.75	2.30	3.18
1.20	1.45	2.15	3.10
1.00	1.55	2.25	3.28
0.91	1.64	2.40	3.35
1.05	1.60	2.49	3.12

For this homework, please assume we have the following model. Let the observation for the j-th sample of the i-th temperture level be Y_{ij} (Label the temperature levels: 1, 2, 3, and 4). Let the these observations be normally distributed about the i-th temperature effect mean μ_i with a common precision τ. Let the temperature effects means be normally distributed about a grand mean μ_0 with precision τ_0. Assume that the precisions τ and τ_0 have gamma priors with parameter values $\alpha = 1$ and $\beta = 1^2$. Also, let μ_0 have a normal prior with mean 0 and precision 1^2. Therefore, we have the following model:

$$
\begin{aligned}
Y_{ij}|\mu_i, \tau &\sim \text{Normal}(\mu_i, \tau) \\
\mu_i|\mu_0, \tau_0 &\sim \text{Normal}(\mu_0, \tau_0) \\
\mu_0 &\sim \text{Normal}(0, 1^2) \\
\tau &\sim \text{Gamma}(1, 1^2) \\
\tau_0 &\sim \text{Gamma}(1, 1^2).
\end{aligned}
$$

Do the following:

(a) Please give the distribution of:

i. $f(\mu_i|\text{DATA}, \mu_0, \tau, \tau_0)$

ii. $f(\mu_0|\text{DATA}, \mu_1, \ldots, \mu_4, \tau, \tau_0)$

iii. $f(\tau|\text{DATA}, \mu_1, \ldots, \mu_4, \mu_0, \tau_0)$

iv. $f(\tau_0|\text{DATA}, \mu_1, \ldots, \mu_4, \mu_0, \tau)$

(b) Use R to run a Markov chain Monte Carlo algorithm on this data. (DO NOT USE WinBugs, Openbugs, Jags or any of their relatives. Th purpose is for you to do some basic programing.) You only need to run the MCMC for about 20,000 iterations for this example. Please supply your computer code.

(c) Please give the posterior distribution (a plot) for the mean of each temperature effect (the μ_i's) as well as some of the summary statistics of these posterior distributions for the drug effect means (mean, standard deviation, and some type of 95% credible region is sufficient).

(d) What is the posterior distribution of the difference in number of cells grown at a temperature of 40 versus 80? What is the posterior probability that there will be more cells grown at a temperature of 40 versus 80?

2. The following table gives the number of down syndrome births and live births and live births for different birth order of the births and for different age categories of the mother. The data was taken from a book by Holford (2002, pg 100) and is originally from Mantel and Stark (1968, Biometrics). The data is for the number of down syndrome births in Michigan from 1950 to 1964. The data and the description of the data is in the data file `DownBirthData.txt`. Do the following with this data.

   (a) Do a Bayesian Poisson regression on this data. Model the logarithm of the rate of down syndrome per live births as a linear function of age and birth order. Model both the age category and the birth order. Please give the Bugs code for this data. Provide the univariate posterior distributions of each level of age category and birth order and also give the posterior distribution of the precisions used in your model. (If you wish, you may report the "standard deviations" instead of the precisions.)

   (b) Provide a brief justification that your model has converged and that you have "burned-in" your simulation enough. (Note: you may give "thinned" plots of your simulations to avoid printing out plots which are suppose to represent several thousands or hundreds of thousand points. For the purpose of this question, you can provide simple diagnostics like trace plots and autocorrelation plots.)

   (c) Provide your belief as to the increase probability of a child being born with Down syndrome to a mother who is over 40 versus a mother who is between 20 and 24 years of age. Also, provide your belief as the the increase probability of a child being born with Down syndrome to the first child in the birth order versus the 5th or later child in the birth order. (That is, you would say that you believe that if the mother is 40+ years of age, then their would have an increase risk of xxxx times the 20-24 year old.) For both of these parameters, please include both a "point" estimate of this increase risk and also state some interval estimate. Please identify the type of point and interval estimate that you are using.

Maternal Age	Birth Order 1	2	3	4	5+
# of Down Syndrome Births					
<20	107	25	3	1	0
20-24	141	150	71	26	8
25-29	60	110	114	64	63
30-34	40	84	103	89	112
35-39	39	82	108	137	262
40+	25	39	75	96	295
# of Live Births					
<20	230,061	72,202	15,050	2,293	327
20-24	329,449	326,701	175,702	68,800	30,666
25-29	114,920	208,667	207,081	132,434	123,419
30-34	39,487	83,228	117,300	98,301	149,919
35-39	14,208	28,466	45,026	46,075	104,088
40+	3,052	5,375	8,660	9,834	34,392

3. In this problem we will analysis an experiment which studies the effect of a dose of a drug on the growth of rats. The data is in the file `BigRatDat.txt` and is described below. This data consists of the growth of 50 rats, where 10 rats were randomly assigned to five different doses of a drug. (The dose levels are 0, .5, 1, 4, and 8 units of the drug.) The weights of the rats were obtained each week for 11 weeks. The data file is sent along with the file for this homework. The data file has the following structure:

```
Column 1: Doses levels
Column 2: Rat number (note: the rats are different for the different dose levels.
Column 3: Week 1 weight
Column 4: Week 2 weight
Column 5: Week 3 weight
Column 6: Week 4 weight
Column 7: Week 5 weight
Column 8: Week 6 weight
Column 9: Week 7 weight
Column 10: Week 8 weight
Column 11: Week 9 weight
Column 12: Week 10 weight
Column 13: Week 11 weight
```

The first several lines of the file `BigRatDat.txt` are below:

```
0    1   54  60  63  74   77   89   93 100  108  114 124
0    2   69  75  81  90   97  120  114 119  126  138 143
0    3   77  81  87  94  101  110  117 124  134  141 151
```

0	4	64	69	77	83	88	96	104	109	120	123	131
0	5	51	58	62	71	74	81	88	93	99	103	113
0	6	64	71	77	89	90	100	106	114	122	134	139
0	7	80	91	97	101	111	119	129	131	137	147	154
0	8	79	85	89	99	104	105	116	121	132	139	147
0	9	77	82	88	92	101	109	119	127	135	144	158
0	10	79	84	91	98	107	114	119	131	137	146	155
.5	1	62	71	75	79	87	91	100	105	111	121	124
.5	2	68	73	81	89	94	101	110	114	123	132	139
.5	3	94	102	109	110	128	133	147	151	153	171	184
.5	4	81	90	95	102	109	120	128	137	141	154	160

Do the following:

(a) Model this data with a growth curve model. So, for the ith rat, you should fit a regression line with the basic form: $Y_{ij} = \beta_{0i} + \beta_{1i} * \text{week}_j + \epsilon_{ij}$. For the β's, you should model them as a linear relationship with the dose level. For example, you should model β_{0i} as a normal distribution with some precision and with $E[\beta_{0i}] = \beta_{00} + \beta_{01}\text{DoseLevel}_i$. The coeffecient β_{1i} is modelled similarly. (That is, the β_{0i} and β_{1i} each follow a normal distribution with some mean and precision.) Pick priors which are not restrictive. Show the WinBugs/OpenBUGS/JAGS code used to model this data. (If you run this through R, then please provide this code also.)

(b) Provide some preliminary evidence that the model looks like it converged. You do not have to do the more advanced statistics (for the purpose of this assignment). It is sufficient to show trace plots and some auto correlation values. (Don't do this for each individual rat's β parameters, it is sufficient to show these values for the parameters of the hierarchical parameters.)

(c) Provide the posterior distribution for the parameters of the distribution of the β's That is, the β_{0i}'s, β_{1i}'s and the β_{00}, β_{01}, β_{10}, and β_{11} parameters. (For this sub-question, you don't need to supply the density estimation. You may just provide the usual summary statistics for the distributions.) (This is for the mean and precision parameters which are discussed in part (a) of this question.) Also, provide the posterior distribution for the precision of ϵ_{ij} (which is defined in (a) above.

(d) Is there any effect due to the drug dose level. Please provide your posterior belief that there is a difference. In justifying your answer, you should include the appropriate posterior distribution.

(e) In a short paragraph, summarize your belief in the effect of the drug dose level. In this summary, you should state what the strength of the evidence from the posterior distributions.