XXXX 2017
xx.xx – x.xx

EXAMINATION FOR THE DEGREES OF M.Sci., M.Sc. and M.Res.

Bayesian Statistics (Level M)

"Hand calculators with simple basic functions (log, exp, square root, etc.) may be used in examinations. No calculator which can store or display text or graphics may be used, and any student found using such will be reported to the Clerk of Senate".

NOTE: Candidates should attempt ALL questions.

CONTINUED OVERLEAF/

1. Consider the Poisson model with parameter θ:

$$p(y|\theta) = \frac{\theta^y}{y!}\exp(-\theta), \qquad \theta > 0, y = 0, 1, \ldots.$$

   (a) Write down the likelihood function appropriate to n i.i.d. observations $y_1, \ldots, y_n$. **[2 MARKS]**

   (b) Explain the Bayesian concept of a conjugate prior. Show that the $\text{Ga}(\alpha, \beta)$ distribution is conjugate for i.i.d. Poisson data $y_1, \ldots, y_n$ using the likelihood derived in the previous question. **[2,3 MARKS]**

   (c) Derive the Jeffreys' prior $p(\theta)$ for the parameter θ in this Poisson model. **[7 MARKS]**

   (d) Is the Jeffreys' prior for θ a proper prior distribution? Explain. **[2 MARKS]**

   (e) Is the posterior distribution that results from using the Jeffreys' prior in this problem a proper distribution? Explain. **[4 MARKS]**

2. Consider the hierarchical model

$$\begin{aligned} y_{ij}|\theta,\beta &\sim \text{Exp}(\theta_j), \qquad i = 1, \ldots, n_j;\ j = 1, \ldots, J, \text{ independently;} \\ \theta_j|\beta &\sim \text{Ga}(\alpha, \beta), \qquad j = 1, \ldots, J, \text{ independently;} \\ \beta &\sim \text{Exp}(\psi), \end{aligned}$$

   where $\theta = (\theta_1, \ldots, \theta_J)$ and α and ψ are two fixed positive real numbers.

   (a) Show that the joint probability density function of all the random quantities in the model is

$$p(\beta, \theta, y) = \psi \frac{\beta^{\alpha J}}{[\Gamma(\alpha)]^J}\left[\prod_{j=1}^{J}\theta_j^{\alpha+n_j-1}\right]\exp\left\{-\left[\psi\beta + \beta\sum_{j=1}^{J}\theta_j + \sum_{j=1}^{J}n_j\bar{y}_j\theta_j\right]\right\},$$

   where $y = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{J1}, \ldots, y_{Jn_J})$ and $\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$. **[4 MARKS]**

   (b) Find the full conditional distributions of β and θ. **[5 MARKS]**

   (c) Explain how the full conditional distributions could be used to implement a Gibbs sampler to draw from $p(\beta, \theta|y)$. **[3 MARKS]**

   (d) In an Empirical Bayes approach, one would drop from the model the higher-level prior on β (third line) and instead estimate β from the data. Explain how you would do that. [Hint: recall that the expected value of a $\text{Ga}(\alpha, \beta)$ random variable is α/β.] **[4 MARKS]**

CONTINUED OVERLEAF/

(e) Suppose that a sample $(\theta^{(t)}, \beta^{(t)})$, $t = 1, \ldots, T$, from the joint posterior distribution of θ and β is available. Explain how you can use it to compute an estimate of the posterior predictive distribution of $\tilde{y}_j$, a future observation from the jth group of observations. **[4 MARKS]**

3. The effectiveness of a proposed gene therapy for a genetic condition that affects the liver was explored in mice (prior to potential application in humans). In the ith of six replicate experiments, n_i $(i = 1, \ldots, 6)$ mice with the liver condition were administered with the gene therapy and after a certain period of time the number y_i of mice with liver function improved by a certain amount was determined, with the following results:

Experiment, i	1	2	3	4	5	6
Sample size, n_i	91	88	102	96	110	113
Number improved, y_i	24	26	7	25	18	18

To explore the effect of the treatment, a Bayesian hierarchical model was fitted in `WinBUGS` with the following model code:

```
model {
  for (i in 1:6) {
    y[i] ~ dbin(theta[i], n[i])
    theta[i] ~ dbeta(alpha, beta)
  }
  alpha ~ dexp(1)
  beta ~ dexp(1)
}
```

where `y[i]` corresponds to y_i and `n[i]` corresponds to n_i.

(a) Convert the `WinBUGS` model specification into standard statistical notation, making it clear in your answer which random quantities are independent. **[4 MARKS]**

(b) Describe the model in words and state one advantage of explaining the data with a model which is hierarchical. **[3,1 MARKS]**

(c) Some of the `WinBUGS` output is shown below:

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	1.086	0.4834	0.002433	0.3689	1.012	2.227	20001	100000
beta	3.084	1.518	0.007458	0.8676	2.839	6.698	20001	100000
theta[1]	0.2633	0.04501	1.532E-4	0.1802	0.2616	0.3562	20001	100000
theta[2]	0.294	0.04726	1.508E-4	0.206	0.2926	0.3908	20001	100000
theta[3]	0.07596	0.02578	8.242E-5	0.03364	0.07325	0.1336	20001	100000

CONTINUED OVERLEAF/

theta[4]	0.2603	0.04365	1.525E-4	0.1792	0.2587	0.3499	20001	100000
theta[5]	0.1672	0.03489	1.093E-4	0.1049	0.1652	0.2409	20001	100000
theta[6]	0.163	0.034	1.103E-4	0.1021	0.1611	0.2346	20001	100000

(Below, the symbols θ_i, α and β will be used to represent the variables called `theta[i]`, `alpha` and `beta`, respectively, in `WinBUGS`.)

For each of the posterior summaries, (i. to iv.), listed below, state whether each can be determined using the `WinBUGS` output above. If your answer is "yes", explain why and compute the estimate. If your answer is "no", explain why not and explain how it could be computed in `WinBUGS`.

i. The posterior mean of $\theta_6 - \theta_5$. **[2 MARKS]**

ii. The posterior variance of $\theta_6 - \theta_5$. **[2 MARKS]**

iii. The posterior mean of α/β. **[2 MARKS]**

iv. The central 95% posterior interval for the log odds, $\log(\theta/[1-\theta])$, in experiment 2. **[2 MARKS]**

(d) Explain how you would modify the model to use independent inverse gamma priors on α and β. Inverse gamma distribution is not supported by `WinBUGS` directly, but it supports the Gamma distribution using function `dgamma(a,b)`. **[4 MARKS]**

4. (a) Imagine you are running a radio manufacturing company. You receive a shipment of transistors from a supplier to be used in the assembly of the final product. Checking the performance of each transistor in the shipment is too expensive, so you establish a sampling plan to either accept or reject the shipment. A random sample of 100 transistors from the whole shipment will be tested. Based upon the number y of defective transistors in that random sample, you will take one of the two decisions: a_1 = *accept the shipment*, or a_2 = *reject the shipment*.

As the test sample is relatively small in comparison to the shipment size, we are safe to assume:

$$y \sim Bi(100, \theta)$$

where θ is the proportion of the defective transistors in the shipment.

From your previous experience with this supplier, you know that θ is relatively small, and assume *a priori*:

$$\theta \sim Be(0.05, 1)$$

You select the following loss function:

$$L(\theta, a) = \begin{cases} 10\ \theta, & a = a_1 \\ 1, & a = a_2 \end{cases}$$

CONTINUED OVERLEAF/

as if the shipment is rejected (a_2 decided) the loss is a constant cost due to inconvenience, delay, and testing of a replacement shipment; while if the shipment is accepted (a_1 decided) the loss is proportional to θ, since θ will also reflect the proportion of defective radios produced. The factor 10 indicates the relative costs involved in two types of errors.

You found on Wikipedia that

$$\int_0^1 x^{a-1}(1-x)^{b-1}dx = B(a,b), \quad \text{for } a \in \mathbb{R}, b \in \mathbb{R}.$$

and you found an excerpt of the statistical table for B(a,b):

$B(a,b)$	b				
a	1	98	99	100	101
0.05	20.0000	15.4850	15.4771	15.4693	15.4616
1.05	0.9524	0.0079	0.0078	0.0077	0.0077
2.05	0.4878	8.3708×10^{-5}	8.1993×10^{-5}	8.0330×10^{-5}	7.8716×10^{-5}
3.05	0.3279	1.7152×10^{-6}	1.6634×10^{-6}	1.6136×10^{-6}	1.5659×10^{-6}
4.05	0.2469136	5.1769×10^{-8}	4.9714×10^{-8}	4.7761×10^{-8}	4.5902×10^{-8}

You receive a shipment that has $y = 2$ defective transistors in a random sample of 100.

i. What is your posterior for θ given this latest test result with $y = 2$?

[2 MARKS]

ii. Find the Bayes expected loss $p(\pi, a)$ for the decision on accepting this shipment.

[5 MARKS]

iii. Should you accept or reject this shipment? **[1 MARK]**

(b) What is the Bayes action for the absolute error loss $L(\theta, a) = |\theta - a|$? **[1 MARK]**

(c) What is the Bayes action for the squared error loss $L(\theta, a) = (\theta - a)^2$? **[1 MARK]**

(d) It is common to report results of an experiment rounded to certain precision. Consider studying average waiting time for a Route 6 bus at a bus stop next to the Botanic Gardens in Glasgow during commute hours. It is natural to consider an exponential model for this problem:

$$y_i \overset{i.i.d}{\sim} exp(\theta)$$

where y_i are the observed waiting times, and θ is the rate of the exponential distribution, $\mathbb{E}[y] = 1/\theta$.

Assume non-informative Gamma prior on θ:

$$\theta \sim Ga(1, 0)$$

CONTINUED OVERLEAF/

I recorded three waiting times between buses as 5, 6, 10 minutes, rounded to a minute.

i. Infer the posterior distribution of θ considering observed waiting times as exact (not rounded). What is the posterior mean waiting time?

[2 MARKS]

ii. Infer the posterior distribution of θ taking rounding into account. What is the posterior mean waiting time in this case?

[5,3 MARKS]

Total: 80

END OF QUESTION PAPER.