University of Glasgow

1st January 2016
12.00 noon – 13.30 p.m.
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.

END OF QUESTION PAPER.

1. Consider the following hierarchical model:

$$\begin{aligned} y_{ij}|\boldsymbol{\beta},\phi &\sim \text{Ga}(2,\beta_i), \qquad i=1,\ldots,I;\, j=1,\ldots,n_i, \text{ independently}, \\ \beta_i|\phi &\sim \text{Ga}(3,\phi), \qquad i=1,\ldots,I, \text{ independently}, \\ \phi &\sim \text{Expo}(1), \end{aligned}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_I)$.

(i) Show that the joint p.d.f of all the random quantities in the model is

$$p(\boldsymbol{y},\boldsymbol{\beta},\phi) = 2^{-I}\left(\prod_{i=1}^{I}\prod_{j=1}^{n_i} y_{ij}\right)\left(\prod_{i=1}^{I}\beta_i^{2n_i+2}\right)\exp\left(-\left[\phi + \sum_{i=1}^{I}\beta_i(n_i\bar{y}_i+\phi)\right]\right)\phi^{3I},$$

where $\boldsymbol{y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{I1}, \ldots, y_{In_I})$ and $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. **[5 MARKS]**

(ii) Hence, or otherwise, find the full conditional distributions of the β_is and of ϕ, identifying what probability distributions they are. **[4,2 MARKS]**

(iii) Describe an algorithm to perform Gibbs sampling of the joint posterior distribution of $\boldsymbol{\beta}$ amd ϕ, using the results derived in part (b). **[4 MARKS]**

(iv) Explain how you would estimate a central 95% posterior interval for β_1 from the output of the Gibbs sampler? When might it be more sensible to try instead to produce a 95% highest posterior density region (HPDR)? **[2,2 MARKS]**

(v) If two parameters are very strongly negatively correlated in the posterior distribution, does the Gibbs sampler

A. work efficiently,

B. work inefficiently,

C. work just as efficiently as if they weren't correlated,

D. any of the above depending on context?

[1 MARK]

2. 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:

END OF QUESTION PAPER.

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)
  }
  inv.alpha ~ dgamma(0.5, 0.5)
  inv.beta ~ dgamma(0.5, 0.5)
  alpha <- 1/inv.alpha
  beta <- 1/inv.beta
}
```

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

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

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

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

```
node     mean    sd      MC error 2.5%   median  97.5%  start sample
alpha    4.116   4.1     0.09016  0.7035 2.983   14.34  1000  99001
beta     16.32   17.12   0.3767   2.179  11.64   58.0   1000  99001
theta[1] 0.2538  0.0427  2.034E-4 0.1762 0.2521  0.3423 1000  99001
theta[2] 0.2799  0.04556 2.616E-4 0.1962 0.2781  0.3741 1000  99001
theta[3] 0.08879 0.02903 2.652E-4 0.0404 0.08599 0.153  1000  99001
theta[4] 0.2516  0.04171 1.846E-4 0.1753 0.2497  0.3387 1000  99001
theta[5] 0.1691  0.03338 1.217E-4 0.1085 0.1676  0.2386 1000  99001
theta[6] 0.1654  0.03296 1.304E-4 0.1056 0.1638  0.2343 1000  99001
```

(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 listed below, state whether it 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 also how you would go about computing it in `WinBUGS`.

END OF QUESTION PAPER.

(a) The posterior mean of $\theta_6 - \theta_5$. **[2 MARKS]**

(b) The posterior standard deviation of $\theta_6 - \theta_5$. **[2 MARKS]**

(c) The posterior median of the prior probability of success in one experiment $\alpha/(\alpha+\beta)$. **[2 MARKS]**

(d) The central 95% posterior interval for the log odds, $\log(\theta/[1-\theta])$, in experiment 3. **[2 MARKS]**

(iv) Clinicians judge that, to be worth pursuing this therapy, it needs to have a probability of success (in the sense of improving liver function) at least as big as 10%. Explain how you would use `WinBUGS` to estimate the posterior probability that the success probability in any of the experiments is at least 10%. (You may assume there is a function `min()` in `WinBUGS` which, given a vector `vec[]` as an argument, returns the smallest entry in `vec`.) **[4 MARKS]**

3. (i) Consider a random variable $X \sim \mathrm{Ga}(\alpha, \beta)$. Suppose $Z = X^{-1}$. Z is said to have an Inv-gamma(α, β) distribution. Show that the p.d.f. of Z is

$$P_Z(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-(\alpha+1)} \exp(-\beta/z).$$

[3 MARKS]

(ii) Suppose n pieces of i.i.d. data y_i are normally distributed with a *known* mean μ but an *unknown* variance σ^2, i.e.

$$y_i|\sigma^2 \sim \mathrm{N}(\mu, \sigma^2) \qquad i = 1, \ldots, n \text{ (independently and } \mu \text{ known).}$$

(a) Show that the likelihood of σ^2, $L(\sigma^2)$, satisfies

$$L(\sigma^2) = (2\pi)^{-n/2} \sigma^{-n} \exp\left(-\frac{n\nu}{2\sigma^2}\right),$$

where $\nu = \frac{1}{n}\sum_{i=1}^n (y_i - \mu)^2$. **[2 MARKS]**

(b) Suppose that the prior distribution on σ^2 is taken an Inv-gamma(α, β) for some choice of α and β. Show that

$$\sigma^2|\boldsymbol{y} \sim \text{Inv-gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{n\nu}{2}\right),$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$. **[3 MARKS]**

END OF QUESTION PAPER.

(c) Use the above result to explain what is meant by a *conjugate prior*. **[2 MARKS]**

(d) Derive the Jeffreys' prior for σ^2. **[6 MARKS]**

(e) Prove whether the Jeffreys' prior above is proper or improper. **[2 MARKS]**

(f) Show that the Jeffreys' prior is a limiting case of the conjugate Inv-gamma prior. **[2 MARKS]**

4. (i) Consider a loss function for estimating a parameter θ with the value ('action') a with the following form:

$$L(\theta, a) = \begin{cases} \alpha(\theta - a), & \theta - a \geq 0; \\ \beta(a - \theta), & \theta - a < 0, \end{cases}$$

where α and β are two positive constants.

(a) Describe what this specific loss function is capturing about the decision process if $\alpha > \beta$. **[2 MARKS]**

(b) Suppose that the posterior distribution of θ is $\text{Un}[0, 0.5]$. Show that the Bayes expected loss is

$$\rho(\pi, a) = \begin{cases} \alpha(0.25 - a) & a \leq 0; \\ (\alpha + \beta)a^2 - \alpha a + 0.25\alpha & 0 < a \leq 0.5; \\ \beta(a - 0.25) & a > 0.5. \end{cases}$$

[6 MARKS]

(c) Show that the Bayes action $a^\pi = \frac{1}{2}\alpha/(\alpha + \beta)$ and sketch the Bayes expected loss as a function of a. **[3,3 MARKS]**

(ii) Suppose the loss function for estimating a parameter θ with the value a is the weighted squared-error loss:

$$L(\theta, a) = \omega(\theta)(\theta - a)^2 \qquad \text{for some } \omega(\theta) > 0.$$

Show that the Bayes action is:

$$a^\pi = \frac{E[\omega(\theta)\theta|y]}{E[\omega(\theta)|y]}.$$

[5 MARKS]

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

END OF QUESTION PAPER.