RESTRICTED OPEN BOOK EXAMINATION
Data provided: Neave's Statistical Tables

MAS6062

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2012–2013**

MAS6062 Bayesian Statistics and Clinical Trials **3 hours**

Restricted Open Book Examination.
Candidates may bring to the examination lecture notes and associated lecture material (but no textbooks) plus a calculator which conforms to University regulations.
Marks will be awarded for your best ***five*** *answers. Total marks 100.*

Please leave this exam paper on your desk
Do not remove it from the hall

Registration number from U-Card (9 digits)
to be completed by student

Blank

1 The true length (in millimetres, mm) of an important engineering component is denoted by θ. It is possible to take measurements x_j which are conditionally independent, given θ, such that $x_j \sim N(\theta, \sigma^2)$ with known standard deviation (in mm) $\sigma = 0.02$.

(i) Given a prior distribution $N(m, v)$ for θ, state carefully the posterior distribution for θ after n measurements $x_1, \ldots x_n$ as above. (You do not need to derive the result). ***(2 marks)***

(ii) If measurements $x_1, \ldots x_5$ are taken, giving values

$$62.607, 62.594, 62.582, 62.610, 62.602,$$

calculate the posterior distribution for θ:

(a) if the prior distribution for θ is $N(62.8, 0.04^2)$, based on past information for similar components, and

(b) if a limiting 'flat' prior is used for θ. ***(4 marks)***

(iii) Calculate the posterior probability that $\theta < 62.6$ based on the posterior distribution from (ii)(a). ***(2 marks)***

(iv) An engineer who believes that the information in (ii)(a) is not directly relevant wants to formulate a prior distribution based on her own experience. Before seeing the data, she believes that θ is likely to be close to 62.5, and has probability 0.9 of being between 62.3 and 62.7. Formulate a suitable prior to represent her views, and give one example of a probability that could be 'fed back' to her to check the appropriateness of the prior.
(6 marks)

(v) The following Winbugs model represents a generalisation of the situation described above. (Variables not defined in the model can be assumed to be given fixed numerical values.) Explain in what way it generalises the situation, and draw a Directed Acyclic Graph to represent the new model.

```
model
{
p <- 1/v
theta ~ dnorm(m,p)
sigma2 <- 1/tau
tau ~ dgamma(a,b)
for (j in 1:n)
{
  x[j] ~ dnorm(theta,tau)
}
}
```

(6 marks)

2 A scientist is interested in estimating the decay rate θ of a rare isotope, relative to a known rate; she designs a series of experiments giving observations $X_1, X_2, \ldots$ with $X_i \sim Poisson(\theta)$, so that

$$P(X_i = x|\theta) = \frac{\theta^x \exp(-\theta)}{x!}$$

with X_i and X_j conditionally independent given θ, for $i \neq j$. Before carrying out the experiments, she considers her prior beliefs about θ and decides that they can represented by an $Exponential(0.4)$ distribution (i.e. an exponential distribution with rate 0.4). She wants to update her beliefs using data x and give a point estimate $\tilde{\theta}$ for θ, but is unsure of the appropriate form for her loss function.

Recall that if θ has the $Gamma(a, b)$ distribution then its probability density function is

$$f(\theta) = \frac{b^a \theta^{a-1} \exp(-b\theta)}{\Gamma(a)}$$

for $\theta > 0$, and it has mean a/b and variance a/b^2; and that the $Exponential(b)$ distribution is the special case with a=1, and has cumulative distribution function

$$F(\theta) = 1 - \exp(-b\theta)$$

for $\theta > 0$.

(i) If she observes $X_1, \ldots, X_{24}$ with $\sum_{i=1}^{24} x_i = 44$, show that her posterior distribution for θ is $Gamma(a^\star, b^\star)$ where $a^\star = 45, b^\star = 24.4$. ***(4 marks)***

(ii) Under the assumption of a quadratic loss function

$$L_Q(\theta, \tilde{\theta}) = (\theta - \tilde{\theta})^2$$

what would be the scientist's point estimate, and associated expected loss,

(a) given her prior distribution and

(b) given her posterior distribution? ***(4 marks)***

(iii) Under the assumption of an absolute loss function

$$L_A(\theta, \tilde{\theta}) = |\theta - \tilde{\theta}|$$

what would be the appropriate point estimate given the scientist's prior distribution? ***(3 marks)***

(iv) Under the assumption of a zero-one loss function

$$L_Z(\theta, \tilde{\theta}) = \begin{cases} 0 & |\theta - \tilde{\theta}| \leq c \\ 1 & |\theta - \tilde{\theta}| > c, \end{cases}$$

what would be the scientist's prior point estimate (as a function of the constant c) and the associated expected loss? If c is small, derive her posterior point estimate and, by taking the posterior density to be approximately constant close to the estimate, obtain an approximate expression for the expected loss. ***(9 marks)***

3 (i) (a) Define $X \sim Binomial(n, \theta)$ and $Y \sim Binomial(m, \theta)$ to be conditionally independent, conditional on the value of θ, and let θ have a $Beta(a, b)$ prior distribution. Write down the posterior distribution for θ given X, and the predictive distribution for Y given X. (You do not need to derive these results.) ***(2 marks)***

(b) A geneticist is unsure about the proportion θ of individuals in a (large) population who carry a particular gene. His prior distribution for θ, based on experience of other similar genes, is $Beta(1/2, 1/2)$. He tests three randomly sampled individuals and finds that none of them carry the gene. Calculate his posterior distribution for θ given X, his predictive probability that the next individual sampled carries the gene and his predictive probability that the next two individuals sampled both carry the gene. ***(5 marks)***

(ii) Observations X and Y each have the distribution $N(0, \theta)$ and are conditionally independent given θ. The parameter θ has prior distribution given by the inverse gamma distribution with parameters d and a, written $IG(d, a)$, and so has density

$$f(\theta) = \frac{a^d \theta^{-(d+1)}}{\Gamma(d)} \exp\left(-\frac{a}{\theta}\right),$$

for $\theta > 0$.

(a) Show that the posterior distribution for θ given X is also of the form $IG(D, A)$ and give expressions for D and A. ***(5 marks)***

(b) Derive the predictive distribution for Y given X and comment briefly on its shape, compared with the distribution of $Y|\theta$. ***(8 marks)***

4 A trial of a new drug for the acute treatment of stroke against placebo is being planned in an Emergency Department setting. Many of the patients arriving are quite ill (some unconscious). The primary efficacy outcome is improvement in patients' quality of life at 1 month.

(i) What are the ethical considerations in recruiting patients into the trial? **(2 marks)**

(ii) If a mean improvement of 5 units in quality of life is required for the trial with an anticipated population standard deviation of 20 units, what is the required evaluable sample size per arm (for 90% power and two sided significance level of 5%)? If 15% of patients are expected to drop out at by 1 month what sample size would be required to ensure sufficient numbers of patients? **(4 marks)**

(iii) The investigator reports that due to budget constraints the maximum sample size can only be 350 patients recruited per arm to ensure a sufficient number evaluable. What impact does this new sample size have on the power assuming again 15% of patients are not in the trial at 1 month? **(3 marks)**

Note for the remaining questions take the sample size to be 350 patients recruited per arm to ensure a sufficient number evaluable.

(iv) Suppose efficacy will be also assessed with O'Brien-Fleming rules for the efficacy assessments at the interim analyses. There will be 3 interim analyses. Comment on the effect these efficacy analyses will have on the evaluable sample size and why there is an effect. **(3 marks)**

(v) It has been suggested that the trial could be undertaken as a cluster randomised trial. If the intra-class correlation is 0.02 and the cluster size was expected to be about 30 patients, what would the evaluable sample size be? Comment on your results. **(2 marks)**

(vi) The randomisation is 1-to-1 and a sample size re-estimation is undertaken half-way through the trial. The total variance observed in the trial is 395 (units) estimated from the two groups combined. What is the estimate of the evaluable sample size now? **(3 marks)**

(vii) A key secondary outcome is the proportion of patients who have responded well to treatment as assessed by a physician. It is anticipated that 70% of patients on the placebo arm will respond while 80% on the test drug are anticipated to respond. For a two sided level of statistical significance at 5% is the current trial big enough? The planned analysis is a chi-squared test. **(3 marks)**

5 The following is some output from R from a subset of a cluster randomised trial with 2 treatments (psychological therapy, (g=1) and control (g=0)) and 4 clusters per arm. The outcome is a Health Related Quality of Life (HRQoL) score (range 0-10) at 3 months.

```
Formula: y ~ 1 + g + (1 | cl)
   AIC   BIC logLik deviance REMLdev
 83.73 87.71 -37.86    75.73   72.89

Random effects:
 Groups   Name        Variance Std.Dev.
 cl       (Intercept) 1.8120   1.3461
 Residual             1.5096   1.2287

Number of obs: 20, groups: cl, 8

Fixed effects:
            Estimate Std. Error t value
(Intercept)   5.4590     0.7913   6.898
g            -1.9918     1.1191  -1.780
```

(i) What is the mean HRQoL score for the treatment group and control group at three months? **(2 marks)**

(ii) What is the type of statistical model fitted here? (**1 mark)**

(iii) Write down two assumptions about the random effect. **(2 marks)**

(iv) Is the difference between treatment groups statistically significant? What is the range of plausible values for difference between treatments? **(2 marks)**

(v) Define an intra-class correlation (ICC) in words and explain how it affects the design and analysis of a cluster randomised trial. **(2 marks)**

(vi) Explain how the ICC is derived from the random effects and estimate it for this study. **(3 marks)**

(vii) Give the main reason using a cluster trial compared with an individually randomised trial and discuss possible sources of bias. **(4 marks)**

(viii) Discuss the merits of matched versus unmatched cluster trials. (**4 marks)**

6 (i) In Good Clinical Practice which fundamental document governs the ethics of clinical trials? **(1 mark)**

(ii) Discuss the different responsibilities of the researcher, the sponsor and the care organisation in running a clinical trial. **(4 marks)**

(iii) Who can consent a patient? Comment on any particular difficulties for consenting a patient. **(4 marks)**

(iv) What is the difference between an adverse event, a serious adverse event, a suspected adverse reaction and a suspected serious adverse reaction (SUSAR)? Discuss problems of detecting a serious adverse event in a clinical trial. (**4 marks)**

(v) Define and discuss the problems of using: a composite variable, a surrogate variable and multiple primary variables. **(4 marks)**

(vi) For safety monitoring in the trial, what consideration would there need to be with respect to blinding and independence? Who is responsible for establishing relevant monitoring? (**3 marks**)

End of Question Paper