Data provided: Neave's Tables

The University Of Sheffield.

SCHOOL OF MATHEMATICS AND STATISTICS **Spring Semester 2013–2014**

MAS6062 Bayesian Methods 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 (i) A zoologist wants to learn about the true weight in grammes, θ, of an animal of a newly discovered species. Based on a visual assessment and knowledge of related species, his best guess for θ is 1000, and he thinks that it is 95% probable that θ lies between 900 and 1100.

(a) Find a suitable normal distribution to represent this prior distribution for θ. ***(2 marks)***

(b) He then takes a measurement X of the weight of the animal in grammes, using equipment known to have errors with mean zero and standard deviation 40, so that $X \sim N(\theta, 40^2)$. State the zoologist's posterior distribution for θ after observing $X = x$. ***(2 marks)***

(c) If $x = 920$, calculate his posterior mean and variance for θ. Find values z_L and z_U such that his posterior probabilities for $\theta < z_L$ and $\theta > z_U$ are both equal to 0.25. How do these values compare with the corresponding quantiles of his prior distribution? ***(6 marks)***

(ii) A unitless physical constant ψ is well known from a combination of different experimental results; experts are agreed on a prior distribution which is normal with mean $\mu = 0.23120$ and standard deviation $\tau = 1.5 \times 10^{-4}$.

(a) If an estimate $\hat{\psi}$ is to be made from this prior, based on a quadratic loss function, give the optimal value of the estimate and the expected loss incurred. ***(3 marks)***

(b) If observations $X_1, \ldots, X_n$ are to be made, with $X_i \sim N(\psi, \sigma^2)$ where $\sigma = 2 \times 10^{-3}$ (and the observations are conditionally independent given ψ), how large must n be to allow an estimate with *half* the quadratic loss of that based on the prior? ***(4 marks)***

(c) What is the predictive distribution for the first measurement, X_1, based on the prior information only? ***(3 marks)***

2 A horticulturalist is interested in the probability θ that a seed of a particular variety germinates successfully. Her prior distribution for the germination probability can be represented by the Beta(a, b) distribution; in an experiment she then observes n (conditionally independent) seeds, of which x germinate successfully.

(i) Write down her posterior distribution for θ. ***(1 mark)***

(ii) If her prior is determined by $a = 3, b = 1$, and she observes $x = 7$ successes with $n = 10$ seeds, give her posterior distribution and posterior mean and variance for θ. ***(3 marks)***

(iii) What is her predictive probability that the next seed observed, after the experiment above, germinates sucessfully? What is her predictive probability that *all* the seeds in a further batch of 10 would germinate? What would the corresponding probabilities have been based only on her prior beliefs? ***(5 marks)***

(iv) She wishes to set up a further experiment based on m seeds. Show that her probability for one or more seeds germinating is

$$1 - \frac{(m+3)\times(m+2)\times\cdots\times 4}{(m+13)\times(m+12)\times\cdots\times 14}$$

and hence show that she would require $m \geq 5$ to ensure that the probability of one or more seeds germinating is at least 0.99. ***(6 marks)***

(v) A less experienced colleague wants to repeat the above analyses but has little knowledge of germination rates for seeds of this sort. Give two different distributions that would be suitable for representing this prior ignorance, and comment briefly on their advantages and disadvantages. Explain what differences you would expect this change to make to the numerical results in (ii); no further calculation is required. ***(5 marks)***

3 (i) The table below shows data on the numbers of work-related accidents $A_1, \ldots, A_8$ occurring within a sample of similar-sized companies in the same industry, recorded over a year, as part of a study about accident rates in the industry as a whole, which is made up of a much larger number of companies.

Index i	1	2	3	4	5	6	7	8
A_i	54	19	44	60	49	51	20	70

The WinBUGS code below implements a model intended to help with the interpretation of these data.

```
model
{
for (j in 1:N)
{
L[j]~dnorm(M,P)
R[j]<-exp(L[j])
A[j]~dpois(R[j])
}
M~dnorm(3,0.25)
P~dgamma(1,4)
V<-1/P
S<-sqrt(V)
}
```

The model is to be run using the following data:

```
list(N=8,A=c(54,19,44,60,49,51,20,70))
```

Write down the model in mathematical terms, and draw a directed acyclic graph to represent its structure. ***(7 marks)***

(ii) A simpler model could be expressed in WinBUGS as follows.

```
model
{
for (j in 1:N)
{
L[j]~dnorm(0,0.001)
R[j]<-exp(L[j])
A[j]~dpois(R[j])
}
}
```

Explain briefly the key statistical differences between the models and their implications for the analysis of the data in (i). ***(3 marks)***

3 (continued)

(iii) The table below shows statistical summaries (in WinBUGS) of some of the output from running the model in (i).

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
L[1]	3.974	0.1363	0.001321	3.699	3.975	4.229	1001	10000
L[2]	2.96	0.2265	0.002139	2.495	2.967	3.376	1001	10000
L[3]	3.773	0.1504	0.001427	3.465	3.776	4.054	1001	10000
L[4]	4.08	0.1306	0.001223	3.815	4.083	4.327	1001	10000
L[5]	3.878	0.1439	0.001263	3.588	3.882	4.157	1001	10000
L[6]	3.919	0.1392	0.001389	3.637	3.923	4.188	1001	10000
L[7]	3.005	0.2189	0.00203	2.553	3.013	3.414	1001	10000
L[8]	4.232	0.1198	0.001179	3.991	4.233	4.463	1001	10000
M	3.705	0.4147	0.004599	2.831	3.708	4.524	1001	10000
R[1]	53.69	7.289	0.07072	40.42	53.25	68.67	1001	10000
R[2]	19.79	4.42	0.04223	12.13	19.44	29.26	1001	10000
R[3]	43.99	6.578	0.06294	31.97	43.66	57.65	1001	10000
R[4]	59.66	7.763	0.07137	45.39	59.34	75.75	1001	10000
R[5]	48.85	6.995	0.06093	36.16	48.51	63.88	1001	10000
R[6]	50.85	7.048	0.07051	37.96	50.54	65.91	1001	10000
R[7]	20.66	4.465	0.04041	12.85	20.36	30.39	1001	10000
R[8]	69.33	8.275	0.08152	54.12	68.94	86.7	1001	10000
S	1.135	0.3796	0.005883	0.5375	1.086	1.993	1001	10000

(a) Based on the table, give 95% central posterior intervals for the annual accident rate for company 1, and for the underlying average accident rate for the industry. ***(4 marks)***

(b) What can you say about the variability of accident rates across the industry? ***(3 marks)***

(c) How do the prior and posterior distributions for the quantity `M` compare? Explain whether you think the prior distribution for `M` seems reasonable. ***(3 marks)***

4 A recent meta-analysis (BMJ, 2013: 347:f6471) looked at trials on the effect of probiotic supplementation during pregnancy on the prevention of asthma and wheeze in the child subsequently.

One outcome was whether a doctor had diagnosed asthma in the child. The results for four selected trials are given in the following table.

	Probiotic		**Placebo**		OR_i	θ_i	wt_i
Study	Asthma cases	Total	Asthma cases	Total			
1	9	77	3	82	3.49	1.249	2.1195
2	5	78	8	78	0.60	−0.512	2.8330
3	14	95	11	96	1.34	0.289	5.3634
4	8	211	12	204	a	b	c

OR_i = Odds ratio for study i, $\theta_i = \ln(OR_i)$, wt_i = inverse of variance.

(i) Find the values a, b and c in the table. ***(3 marks)***

(ii) Prove that the variance of the fixed effect estimate is the inverse of the sum of the weights. ***(3 marks)***

(iii) Find the fixed effects estimate of the treatment effect and its standard error. ***(2 marks)***

(iv) Test the hypothesis that all the θ_i are zero. ***(2 marks)***

(v) Carry out a test of heterogeneity. ***(3 marks)***

(vi) Find the inconsistency statistic I^2 and comment. ***(2 marks)***

(vii) Write a short report stating what evidence exists from this analysis that probiotics in pregnancy can prevent a diagnosis of asthma in a child. Comment on the size of the effect. ***(5 marks)***

5 A researcher is planning a clinical trial to compare a new treatment to standard treatment for multiple sclerosis. The endpoint is time to relapse or death. They have looked at published trial data on the standard treatment and estimate the median time to relapse or death to be 6 months.

(i) Give 3 factors the researcher should consider when using published data to estimate the median time to relapse on a standard treatment. ***(3 marks)***

The researcher thinks the new treatment will extend the time to relapse or death by 50% and is planning a clinical trial to investigate this.

(ii) Assuming proportional hazards, write down the associated null and alternative hypothesis in terms of the hazard ratio of experimental to standard for time to relapse or death to test the researcher's belief. ***(1 mark)***

(iii) Determine the number of events required to test this hypothesis with 2.5% 1-sided type I error and 90% power. ***(4 marks)***

(iv) Write down the number of events required to test the hypothesis that new treatment will extend the time to relapse or death by 100% with the same type I error and power. ***(1 mark)***

(v) Assuming: a median time to relapse of death on the standard treatment is 6 months, the new treatment will extend the time to relapse or death by 50% , exponential time to relapse or death, uniform accrual of patients over a 6 month period and a minimum 6 month follow-up period, calculate :

(a) The probability of a relapse or death over the trial period (i.e. 6 months accrual plus 6 months minimum follow-up) ***(3 marks)***

(b) The number of patients required to recruit into the trial. ***(2 marks)***

(c) Re-calculate (a) and (b) using the further assumptions of a constant dropout rate of 0.02 patients per month and patient accrual following a truncated exponential with p.d.f $f(a) = \dfrac{\gamma e^{-\gamma a}}{1 - e^{-\gamma A}}$, where A is the time the trial is expected to run and a is the time a patient enters the trial, measured from the start with $0 < a \leq A = 6$ months and $\gamma = -1.5$. ***(6 marks)***

6 An investigator wishes to undertake an assessment of dose response of a drug in asthma, so that three doses of the new drug at 25mg, 50mg and 75mg will be investigated against placebo. They decide they require 32 patients.

(i) Construct a Williams square for this design of sequences for the treatment arms for patients to be randomised to. ***(2 marks)***

(ii) Due to practical constraints there can only be three periods for this design. Construct sequences for the treatment arms for this design for patients to be randomised to. ***(2 marks)***

(iii) For the same practical constraints there can only be three periods for this design, but the investigator wants everyone to receive the placebo. Construct sequences for the treatment arms for this design for patients to be randomised to. ***(3 marks)***

The results for a pharmacokinetic study on the active treatment is given in the following table.

	Intravenous
Cmax (units)	14
T1/2 (h)[1]	13
AUC0-12(units)	75
AUC0-∞ (units)	114

[1]T1/2 is the elimination half-life.

(iv) If this was a single dose study what would be the minimum time between periods based on the pharmacokinetics for the cross-over study? ***(1 mark)***

(v) If the treatment is to be given as a repeat dose with a dosing interval of $\tau = 12$ hours, what would be the predicted values of Cmax and AUC0-τ at steady state? ***(2 marks)***

The investigator then decides to conduct a cluster randomised trial of the 50 mg dose of the drug vs placebo, where general practitioners (GPs) are randomised to give all their patients only one treatment.

(vi) If each GP has on average 10 patients with asthma willing to be randomised, and the intra-cluster correlation is 0.1, what is the design effect? ***(2 marks)***

(vii) If the number of patients required for a non-clustered trial is 64, what is the number of patients and GPs required for a cluster trial? ***(2 marks)***

(viii) If we expect 20% of patients to drop out, what number of patients and GPs should we recruit? ***(2 marks)***

(ix) Discuss briefly the merits of an individually randomised trial vs a cluster trial in this case. ***(4 marks)***

End of Question Paper