UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Friday, 6 June, 2014 1:30 pm to 3:30 pm

PAPER 35

APPLIED BAYESIAN STATISTICS

Attempt no more than ***THREE*** *questions.*

There are ***FOUR*** *questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1

Starting at time $t = 0$, a machine produces widgets as a Poisson process with a rate λ widgets per hour, so that the number of widgets in time $(0, t)$ hours has a Poisson distribution with mean λt. It is observed until T hours have elapsed, and in that time it produces a total of n widgets at $t_1, t_2, ..., t_n$ hours.

(a) What is the likelihood for λ provided by the data?

(b) What does it mean to say that a Gamma (a, b) distribution for λ is conjugate for this likelihood?

(c) If you assume a Gamma(a, b) prior for λ, what is the posterior distribution?

(d) Let M be the number of widgets that will be produced in the next hour. Conditional on λ, what is the probability $p(M = 0|\lambda)$ of no widgets in the next hour?

(e) Show that $p_0 = p(M = 0|a, b, n)$, the current predictive probability of no widgets in the next hour, can be written $\left(1 - \frac{1}{b+T+1}\right)^{n+a}$. Show that, as $T, n \to \infty$ and $n/T \to \hat{\lambda}$, p_0 tends to $e^{-n/T}$. Suggest why this is reasonable.

(f) Suppose you are given values for a, b, n. Describe, using rough BUGS code (exact syntax is not necessary), how you would estimate the posterior probability that $p_0 < 0.5$.

(g) Suppose you set $\lambda = 1$ and set the machine going at time 0. You now find that at time T the machine is in fact operating at a rate of 2 widgets per hour, since at some point θ in the interval $(0, T)$ the hourly rate suddenly changed from 1 to 2. If $j(\theta)$ is such that $t_j < \theta < t_{j+1}$, what is the likelihood for θ?

(h) Describe, using rough BUGS code, how you might make an inference on θ, by using the 'zeros' trick or otherwise. You can assume a uniform prior for θ.

[A Poisson(μ) distribution has density $p(y|\mu) = \frac{\mu^y}{y!}e^{-\mu}; y = 0, 1, ...$ A Gamma(a, b) distribution has density $p(\lambda|a, b) = \frac{b^a}{\Gamma(a)}\ \lambda^{a-1}\ e^{-\lambda b};\ \lambda \in (0, \infty)$, with mean a/b and variance a/b^2.]

2

Assume a lady is given 8 cups of tea, 4 of which have had the tea put in first, and for 4 the milk has been put in first. She does not know there are 4 of each, and is just asked to say whether the tea or milk were first, with the following results.

	Says 'Milk first'	Says 'Tea first'	
Truly Milk first	3	1	4
Truly Tea first	1	3	4

Assume that, when the Milk truly has been put in first, the number of times she says 'Milk first' is Binomial$(4, p_M)$, and if the Tea truly has been put in first, the number of times she says 'Milk first' is Binomial$(4, p_T)$.

(a) Define the Jeffreys prior $p_J(\theta)$ for a general one-parameter sampling distribution $p_Y(y|\theta)$.

(b) Derive the Jeffreys prior for a Binomial(n, θ) distribution.

(c) Assuming independent Jeffreys priors for both p_M and p_T, what are the posterior distributions for p_M and p_T, and their posterior means?

(d) How would you make an inference on $p_M - p_T$ using MCMC methods?

(e) Someone says that p_M and p_T should be correlated, to reflect that an individual is generally more or less likely to say that the milk went in first, whatever the truth. They suggest the following method of producing a correlated distribution for p_M and p_T.

1. Generate p_M from a Beta(α, β) prior distribution
2. Generate a random variable $X = x$ from a Binomial(n, p_M) distribution, where n is fixed by the investigator
3. Compute the posterior distribution Beta$(\alpha + x, \beta + n - x)$
4. Generate p_T from this Beta$(\alpha + x, \beta + n - x)$ prior distribution

Explain briefly why large n will induce a large correlation between p_M and p_T.

(f) Using the law of the iterated expectation (see below), show that $E[X] = \frac{n\alpha}{\alpha+\beta}$.

(g) Using the law of the iterated expectation, show that $E[p_T] = \frac{\alpha}{\alpha+\beta}$.

(h) Show that the joint density $p(p_T, X, p_M) \propto \binom{n}{x} (p_T p_M)^{x+\alpha-1}[(1 - p_T)(1 - p_M)]^{n-x+\beta-1}$

(i) What does it mean to say 2 random variables X and Y are exchangeable? If X and Y are exchangeable, show they have the same marginal distribution.

(j) Explain why p_M and p_T are exchangeable. Hence argue that p_T has a Beta(α, β) prior distribution, the same as p_M.

[NB A Beta(a, b) distribution has density $p(\theta|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ \theta^{a-1}\ (1-\theta)^{b-1};\ \ \theta \in (0,1)$. Its mean is $a/(a+b)$. The law of the iterated expectation states that for random variables (X, Y), $E[E[X|Y]] = E[X]$.

3

Assume $Y_1, .., Y_n$ are observations from a Normal$(\beta, 1)$ distribution so that $\overline{Y} \sim$ Normal$(\beta, \frac{1}{n})$. Two competing hypotheses are being considered: H_0 says that β is negligibly small, and H_1 that β may be any size. These hypotheses are represented by alternative prior distributions $p_0(\beta)$ = Normal$(0, \frac{1}{n_0})$, where n_0 is large, and $p_1(\beta)$ = Normal$(0, \frac{c^2}{n_0})$, where c is large.

(a) Show that c is the ratio of the two prior densities at $\beta = 0$.

(b) For $i = 0, 1$, what are the posterior distributions $p_i(\beta|\overline{y})$? [NB You can just state these].

(c) For $i = 0, 1$, what are the predictive distributions $p_i(\overline{y}) = \int p(\overline{y}|\beta)p_i(\beta)d\beta$? [NB You do not need to actually do the integrals].

(d) For an observed $\overline{y}$, what is the Bayes factor B_{01} between hypotheses H_0 and H_1? If $\overline{y} = 0$, show that $B_{01} = \sqrt{(\frac{n_0}{n} + c^2)/(\frac{n_0}{n} + 1)}$.

(e) Suppose we observe $\overline{y} = 3/\sqrt{n}$. Why would we consider this 'statistically significant' evidence that $\beta \neq 0$? It has been suggested that a reasonable assumption might be $n_0/n = 100$, and $c = 100$. What would be the behaviour of the posterior distributions $p_i(\beta|\overline{y})$ for $\overline{y} = 3/\sqrt{n}$?

(f) If we observe $\overline{y} = 3/\sqrt{n}$, show that the Bayes factor will favour H_0.

(g) Suppose we assign prior probabilities $p(H_0) = p(H_1) = 0.5$. In terms of the posterior distributions $p_i(\beta|\overline{y})$ and B_{01}, what is the overall posterior distribution for β?

(h) Explain why this posterior distribution will be pulled towards 0 for 'small' $\overline{y}$ of $O(\frac{1}{\sqrt{n}})$, but for large $\overline{y}$ will centre on $\overline{y}$.

(i) Explain how you might extend this idea to a multiple regression context in which you wanted to select variables only if they had a substantial effect, and otherwise reduce their coefficients to near 0.

4

A clinical trial randomises n_T patients to a new drug and n_C to a control treatment, and observes r_T successful responses in the treated group and r_C in the control: these are assumed to be observations from Binomial distributions with underlying chance θ_T, θ_C of success in the treated and control groups respectively. Suppose we assume model logit $\theta_C = \log \frac{\theta_C}{1-\theta_C} = \alpha - \beta/2$, logit $\theta_T = \alpha + \beta/2$.

(a) Show that β is the log-odds-ratio associated with the treatment.

(b) How, roughly, might you interpret α?

(c) Suppose we thought an odds-ratio outside the range (1/8,8) was rather implausible. What prior on β might be appropriate? [NB. $e^2 = 7.4$]

(d) Consider the following data from 6 (real) randomised trials on beta-blocker drugs.

Study	**Mortality: deaths/total**	
	Treated	*Control*
1	3/38	3/39
2	7/114	14/116
3	5/69	11/93
4	102/1533	127/1520
5	32/209	40/218
6	22/680	39/674

We now want to put together this series of $J = 6$ studies on the same drug, where the jth study has r_{Tj}/n_{Tj} successful responses in the treated group and r_{Cj}/n_{Cj} in the control group. We assume logit $\theta_{Cj} == \alpha_j - \beta_j/2$, logit $\theta_{Tj} = \alpha_j + \beta_j/2$. What does it mean to say that we assume the β_js are exchangeable, and when might this be a reasonable assumption?

(e) Suppose we assume $\beta_j \sim \text{Normal}(\mu, \tau^2)$. You consider it very unlikely that the true odds-ratios e^{β_j} would vary from trial by more than a factor of 50. What might be reasonable priors for μ and τ, if you have this weak prior information? Why would it be inappropriate to assume a Jeffreys prior $p(\tau) \propto 1/\tau$?

(f) We now assume the α_i's have independent locally uniform priors. Write rough BUGS code for this analysis

(g) We then fit two additional models: a 'Common' model in which it is assumed that all the β_j's are identical, and an 'Independent' model in which each β_j is given an independent locally uniform prior. The following output from WinBUGS is obtained.

```
Dbar = post.mean of -2logL;
Dhat = -2LogL at post.mean of stochastic nodes

Model            Dbar    Dhat      pD     DIC
-------------------------------------------------------------------
```

Common	64.8	57.8	7.0	71.8
Exchangeable	61.9	53.2	8.7	70.5
Independent	64.9	53.1	11.9	76.8

How would you interpret the `Dhat`, `pD` and the DIC values for these three models?

(h) The assumption that the β's are normally distributed is questionable, and the following modification to the prior distribution has been suggested:

```
beta[j]     ~  dnorm(mu , invtau2[j])
invtau2[j] <-  lambda[j]/(4*psi*psi)
lambda[j]   ~  dchisqr(4)
```

By considering the distribution of $\frac{(\beta_j - \mu)}{\psi}$ or otherwise, prove that this specification will induce, conditional on μ and ψ, a t_4 distribution for the β's.

(i) When might this be an appropriate model?

(j) How might you assess if it were a more appropriate model for the data?

END OF PAPER