MATHEMATICAL TRIPOS Part III

Wednesday 4 June 2008 1.30 to 3.30

PAPER 44

APPLIED BAYESIAN STATISTICS

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

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

The questions carry equal weight.

When asked to provide code, you need only give the approximate form for the statements: precise syntax is not expected.

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 Much of biomedical science is concerned with finding true relationships between 'exposures' (say genes) and 'outcomes' (say diseases). Ioannidis (2005) published a much-cited paper *Why most published research findings are false*, in which he supposed that relationships being studied were either 'true' (denoted T) or 'false' (denoted F): a false relationship corresponds to a null hypothesis H_0, while a true relationship corresponds to an alternative hypothesis H_1. He assumed, in any particular field, a ratio R between the proportion of true relationships and the proportion of false relationships. Furthermore, he assumed that each study was reported as being either 'positive' (denoted P), or 'negative' (denoted N), and that all studies were designed with Type 1 error α (the probability that a false relationship is detected in the study - a 'false positive result') and Type II error β (the probability that a relationship that truly does exist is not detected in the study - a 'false negative result').

(a) What is the probability that a relationship is false, given that a study reported it as positive?

(b) Show that R has to be at least $\alpha/(1-\beta)$, in order for a reported positive result to be more likely true than false.

Suppose a study is designed to have a test statistic Y with a $\mathrm{N}(0,1)$ distribution under the null hypothesis that there is no relationship, and the alternative hypothesis is that Y has a $\mathrm{N}(2.49,1)$ [NB $\Phi(1.65) \approx 0.95, \Phi(-0.84) \approx 0.2$, where Φ is the standard normal cumulative distribution function].

(c) If a positive result is declared if $Y > 1.65$, show that $\alpha = 0.05, \beta = 0.20$.

(d) If we observe y, show that the Bayes factor for the alternative hypothesis against the null hypotheses is $e^{2.49(y-1.245)}$.

(e) In a succeeding paper, Goodman and Greenland criticised Ioannidis' analysis, saying that it is not sensible to summarise results by just whether they are 'significant' or not. In the light of the preceding analysis, do you think this is a reasonable criticism?

Suppose that n independent studies, each designed with Type I error α and Type II error β, are performed to test the same relationship.

(f) What is the chance that, if the relationship does not exist, at least one will show a positive result?

(g) If all we know is that there is at least one positive result among the n studies, what are the odds that this is a true finding?

(h) What happens to these odds as n increases? Why does Ioannidis claim that '*The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true*'?

(i) Goodman and Greenland also say it is unrealistic to assume that all we know is that there is at least one positive result among n studies. If we read a published account that claims a positive relationship based on the *full* results of many studies carried out in a field, should that increase or decrease our confidence in the positive finding, relative to what we would feel after reading a single study?

2 Suppose y_i is the numbers of cases of MRSA (a bacterial infection) in a single year, in the ith of n hospitals, each drawn respectively from a Poisson distribution with mean θ_i. The θ_i's are assumed drawn from an exponential distribution $p(\theta_i|\psi) = \psi e^{-\psi\theta_i}$; $\theta_i > 0$.

(a) Draw a directed graph for this model.

(b) When using Gibbs sampling for this model, prove that the full conditional distribution for θ_i depends only on y_i and ψ, and derive its form. [The Gamma distribution has the form $p(x) = b^a x^{a-1} e^{-bx}/\Gamma(a)$.]

(c) What does y_i contribute to the likelihood for ψ? Show that contribution of y_i to the likelihood for $\phi = 1/(1+\psi)$ is $(1-\phi)\phi^{y_i}$.

(d) What is the conjugate prior for ϕ?

Suppose the logarithm of the hyper-parameter ψ is given an improper uniform prior

$$p(\log \psi) \propto 1 .$$

(e) What is the induced prior for ϕ?

(f) Find the posterior distribution for ϕ using this prior.

(g) Show that the posterior expectation of θ_i is a weighted average of y_i and $\overline{y}$.

(h) In real hospitals would you expect the assumptions about the prior distribution for the θ_i's to be correct?

(i) What might be a more reasonable model?

3 The testosterone/epitestosterone (T/E) ratio is used as a biomarker for detecting athletes' abuse of testosterone, and a Bayesian model has been suggested for monitoring repeated measures of the logarithm of an athlete's T/E ratio.

An athlete i is assumed, if he/she is not taking additional testosterone, to have log(T/E) measurements drawn independently from a normal distribution with mean θ_i and standard deviation σ_i. We assumed that the population of θ_i's are independently drawn from a $N(\mu, \tau^2)$ distribution.

Suppose for the moment that we know values for μ and τ. We select an athlete i for whom we have no data, and assume we know σ_i.

(a) What is the predictive distribution for Y_{i1}, the first observation to be taken on athlete i?

(b) We want to set an upper limit u_{i1}, so that the predictive probability that an observation lies above that limit, if the athlete is not abusing testosterone, is 1/1000. Write down an explicit expression for u_{i1} in terms of μ, τ, σ_i and Φ, where Φ is the standard normal distribution function.

Suppose we now take a measurement y_{i1} on athlete i.

(c) What is the posterior distribution for θ_i?

(d) Denoting the posterior mean and variance of θ_i by m_{i1} and v_{i1} respectively, what is the predictive distribution for Y_{i2}, a future second measurement?

(e) What monitoring limit u_{i2} would you set for Y_{i2} , such that there was a 1/1000 predicted probability that an 'innocent' athlete would exceed that limit?

(f) Explain in words how the monitoring limits will change as more data becomes available on athlete i.

In fact the parameters μ, τ and σ's are unknown, but there is a historical database available of repeated measurements on athletes who were not abusing testosterone, denoted $y_{j1}, ..., y_{jn_j}$, for j in 1 to J. The following code has been suggested for analysing the database.

```
for (j in 1:J) {

   for (k in 1:n [j]) {

       Y [j,k] ~ dnorm (theta [j], invsigma2 [j])
   }

   theta [j]           ~ dnorm (mu,  invtau2)
   log (invsigma2 [j]) <-    - log.sigma2 [j]
   log.sigma2 [j]      ~ dnorm (phi, invpsi2)     # (1)
}

mu   ~ dunif (-100, 100)
invtau2 <- 1 / (tau*tau) ; tau ~ dunif (0, 100)   # (2)
phi  ~ dunif (-100, 100)
invpsi2 <- 1 / (psi*psi) ; psi ~ dunif (0, 100)   # (3)
```

(g) Explain why this code may be reasonable, with special reference to the numbered lines of code.

Question continued/...

(h) Suppose you now have three measurements available on a new athlete, and you want to check whether the third observation is 'extreme' relative to the first two (*i.e.* outside a 99.9% prediction interval). How would you adapt the code to carry out this analysis?

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

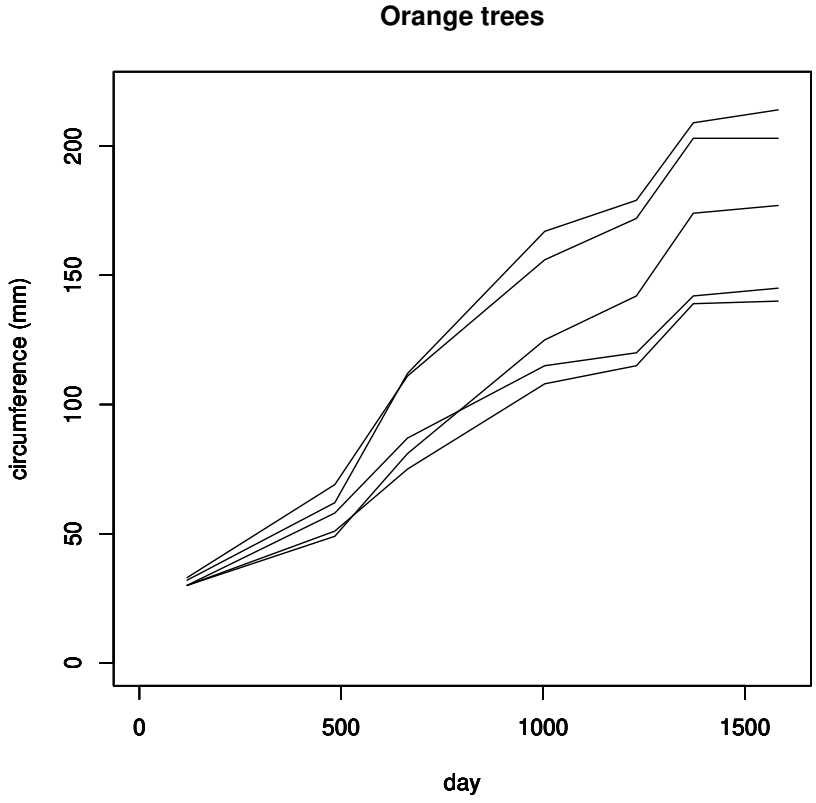
```
theta   [i]  ~ dnorm   (mu , invtau2 [i])
invtau2 [i] <- lambda  [i] / (4*tau*tau)
lambda  [i]  ~ dchisqr (4)
```

(i) What prior distribution will this imply for each $(\theta_i - \mu)/\tau$?

(j) Why might this be an appropriate model?

(k) How might you assess if it was a more appropriate model for the historical data?

4 The graph below shows the circumferences in mm of 5 orange trees, each measured on day x.



An investigator uses the following WinBUGS code to fit an independent 'logistic growth curve' to each tree (Model A).

```
for (i in 1:5) {

   for (j in 1:7) {

      Y [i,j] ~ dnorm (m [i,j], invsigma2)
      m [i,j] <- phi [i,1] / ( 1 + phi [i,2]
                               * exp ( phi [i,3]
                               * ( x [j] - mean (x []) ) / sd (x []) ) )
   }

   log ( phi [i,1] ) <-   theta [i,1]
   log ( phi [i,2] ) <- - theta [i,2]
         phi [i,3]   <- - theta [i,3]

   for (p in 1:3) {                  # *

      theta [i,p] ~ dunif (-100,100)  # *
   }                                  # *
}                                     # *

invsigma2 ~ dgamma (1.0E-3, 1.0E-3)
sigma <- 1 / sqrt (invsigma2)
```

(a) Why is x transformed?

(b) Interpret the parameters `phi[i,1]`, and 1/ (1 + `phi[i,2]`). Why do you think this is called a logistic curve?

(c) Do you think the prior distributions are reasonable?

The investigator decides to replace the starred lines of code in Model A by the following (Model B).

```
.....
theta [i,1:3] ~ dmnorm (mu [], Omega [,])
}
for (p in 1:3) { mu[p] ~ dunif (-100,100) }
Omega [1:3, 1:3] ~ dwish (R [,], 3)
R [1,1] <- 1; R [1,2] <- 0; R [1,3] <- 0;
R [2,1] <- 0; R [2,2] <- 1; R [2,3] <- 0;
R [3,1] <- 0; R [3,2] <- 0; R [3,3] <- 1
```

Some of the output is shown below:

	theta [,1]		theta [,2]		theta [,3]	
	mean	sd	mean	sd	mean	sd
theta [1,]	5.063	0.090	0.753	0.357	1.443	0.281
theta [2,]	5.392	0.059	0.679	0.252	1.601	0.223
theta [3,]	5.057	0.103	0.557	0.354	1.412	0.280
theta [4,]	5.427	0.053	0.673	0.234	1.715	0.227
theta [5,]	5.302	0.098	0.291	0.289	1.498	0.241

(d) Comment on the appropriateness of the new code

(e) Why might it be reasonable to assume that all trees have the same values for `theta[i,2]` and `theta[i,3]`?

(f) How might you change the code to reflect this assumption (keeping `theta[i,1]` as a random effect) ? Call this Model C.

After 50,000 iterations the following output for DICs were obtained for all three fitted models:

```
Dbar = post.mean of -2logL;
          Dbar    pD       DIC
Model A   248.1   12.0     260.1
Model B   244.2   13.9     258.1
Model C   247.3    8.2     255.6
```

(g) Comment on the how the values of `Dbar` (the mean deviance), `pD` (the effective number of parameters) and `DIC` depend on the models.

(h) What might be considered unsatisfactory about `Dbar` and `pD` for Model A? Why might this have happened?

(i) In the final model, say very briefly how you would go about testing the normality assumption for the residual error. Would you also be testing the adequacy of the normal assumption for the random effects?

END OF PAPER