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

The University Of Sheffield.

MAS6004

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

MAS6004 Inference **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 $x_1, \dots x_n$ as above. (You do not need to derive the result). ***(2 marks)***

(ii) If measurements $x_1, \dots x_5$ 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 (i) The load on 10 web servers is measured and observed to be $x_1, \ldots, x_{10}$. A *Gamma*(a, b) with density:

$$f(x) = \frac{b^a x^{a-1} \exp(-bx)}{\Gamma(a)} \quad x > 0$$

is believed to fit the data.

(a) Derive the profile log-likelihood for a. ***(7 marks)***

(b) For our observed dataset of the load on 10 web servers, we find that $\sum_{i=1}^{10} x_i = 6.09$ and $\sum_{i=1}^{10} \log x_i = -6.212$. The profile likelihood function for a is maximised at $a = 4.159$. Give the maximum likelihood estimate for b. ***(2 marks)***

(c) Use the profile deviance function to test the null hypothesis that $a = 4$.
Note: $\Gamma(4) = 6$ *and* $\Gamma(4.159) = 7.352$. ***(5 marks)***

(ii) Suppose importance sampling, using a normal approximation as the importance density, is to be used to sample from a *Beta*$(4, 3)$ distribution with density

$$f_\theta(\theta) = \begin{cases} 60\,\theta^3(1-\theta)^2 & \text{for } \theta \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

By considering a Taylor series expansion of $\log f(\theta)$ about the mode of θ, obtain the mean and variance of the importance density. ***(6 marks)***

5 Suppose X_1 and X_2 are independent random variables with a $N(\mu_1, \sigma^2)$ distribution while Y_1 and Y_2 are independent random variables with a $N(\mu_2, \sigma^2)$ distribution. The common variance σ^2 is presumed known. Let us define $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$.

(i) Show that the complete data log-likelihood

$$\log\{f(X, Y|\mu_1, \mu_2)\} = M - \frac{1}{2\sigma^2}\left[x_1^2 + x_2^2 + y_1^2 + y_2^2 - 2\mu_1(x_1 + x_2) - 2\mu_2(y_1 + y_2) + 2\mu_1^2 + 2\mu_2^2\right],$$

where M is some constant, can be written as

$$l(X, Y|\theta) = \sum_{i=1}^{2} A_i(\theta)B_i(X, Y) + C(X, Y) + D(\theta)$$

where $\theta = (\mu_1, \mu_2)$. Give expressions for $A_1(\theta)$, $A_2(\theta)$, $B_1(X, Y)$, $B_2(X, Y)$ and $D(\theta)$. ***(4 marks)***

(ii) Suppose that only three observations are available: X_1, Y_1 and $Z = X_2 + Y_2$.

(a) Show that

$$X_2|Z \sim N\left(\frac{\mu_1 + z - \mu_2}{2}, \frac{\sigma^2}{2}\right).$$

Hint: Use the fact that

$$f(x_2|Z = z) \propto f_{(X_2,Z)}(x_2, z) = f_{(X_2,Y_2)}(x_2, z - x_2)$$

since $Z = z|X_2 = x_2$ *if and only if* $Y_2 = z - x_2$.

(6 marks)

(b) The maximum likelihood estimates of μ_1 and μ_2 given X_1, Y_1 and Z are to be obtained using the EM algorithm. Let μ_1^{old} and μ_2^{old} denote the current estimates of $\hat{\mu}_1$ and $\hat{\mu}_2$. By maximising

$$Q(\mu_1, \mu_2|\mu_1^{old}, \mu_2^{old}) = E\left[\log\{f(X, Y|\mu_1, \mu_2)\}|X_1, Y_1, Z, \mu_1^{old}, \mu_2^{old}\right]$$

with respect to μ_1 and μ_2, derive improved estimates of $\hat{\mu}_1$ and $\hat{\mu}_2$. This can be done as follows:

- Use your result from part (a) to find

$$E\left[\log\{f(X, Y|\mu_1, \mu_2)\}|X_1, Y_1, Z, \mu_1^{old}, \mu_2^{old}\right].$$

- Solve

$$\frac{\partial Q}{\partial \mu_i} = 0$$

for $i = 1, 2$.

(10 marks)

6 (i) It is possible to rescale the standard Cauchy distribution to give it a `location` x_0 and a `scale` γ. The rescaled density is then

$$f(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

In a physics experiment, the energy of an unstable quantum state is recorded on n occasions and stored in a vector `x`. These data are known to have extremely heavy tails. The following R analysis is then applied:

```
> theta.hat <- median(x)
> scale.hat <- IQR(x)/2
>
> n <- length(x)
> M <- 1000
> z.star <- rep(NA, M)
>
> for(i in 1:M) {
+   x.star <- rcauchy(n, location = theta.hat, scale = scale.hat)
+   z.star[i] <- median(x.star)
+ }
>
> quantile(z.star, c(0.005, 0.995))
    0.5%    99.5%
1.105624 3.638157
```

(a) Explain carefully the procedure that has been performed here, and state what the output in the last line represents. ***(5 marks)***

(b) Why do you think we are using `median` and `IQR` as the parameter estimates instead of `mean` and `sd`? ***(1 mark)***

(c) Comment on the accuracy of the procedure in relation to the size of n and M. ***(3 marks)***

(d) How would you find a 90% confidence interval for the location using the non-parametric bootstrap? ***(4 marks)***

 Question 6 continued on next page

6 (continued)

(ii) Suppose we have 30 observations (x_i, y_i) thought to arise from a model

$$y_i = \alpha x_i + \epsilon_i$$

where $\log \epsilon_i \sim N(0, \sigma^2)$. Let $\hat{\alpha}$ denote the least squares estimator of α. We wish to test a null hypothesis $H_0 : \alpha = 0$ against a two-sided alternative.

(a) Explain why a conventional parametric test comparing the test statistic

$$T = \frac{|\hat{\alpha}|}{\sqrt{Var(\hat{\alpha})}}$$

against a t-distribution is not suitable in this instance. ***(2 marks)***

(b) Explain the following alternative analysis including the conclusion of the investigation. Make sure you explain why the test statistic $T^\star = \left|\sum_{i=1}^{30} x_i y_i\right|$ is being used instead of the T statistic in part a).

```
> t.obs <- abs(sum(x*y))
>
> M <- 1000
> t.star <- rep(NA, M)
>
> for(i in 1:M) {
+  y.resamp <- sample(y, replace = FALSE)
+  t.star[i] <- abs(sum(x*y.resamp))
+ }
>
> mean(t.star > t.obs)
[1] 0.031
```

(5 marks)

End of Question Paper