UNIVERSITY OF LONDON

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2003

This paper is also taken for the relevant examination for the Associateship.

M2S2 Statistical Modelling I

DATE: Thursday, 22nd May 2003 TIME: 2 pm – 4 pm

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used. Statistical tables will not be available.

1. *a)* Define, in general terms, a random sample and a maximum likelihood estimator. Define any terminology used.

b) Consider $\underline{X} = (X_1, \dots, X_n)^\top$, a random sample from the general Laplace distribution with probability density function (pdf)

$$f(x) = \frac{1}{2\phi} \exp\left(-\frac{|x-\theta|}{\phi}\right) \quad \text{for} \quad x \in \mathbb{R},\ \phi > 0,$$

and let $\underline{Y} = (Y_1, \dots, Y_n)^\top$ be a random sample from the Laplace distribution with $\phi = 1$.
Assume that $\underline{X}$ and $\underline{Y}$ are independent.

i) What is the joint likelihood function of the two random samples, $\underline{X}$ and $\underline{Y}$?

ii) Assuming that $\theta = \theta_0$, some known constant, estimate ϕ using the maximum likelihood method, denoting your estimator by $\widehat{\phi}$.
Why would you expect $\widehat{\phi}$ not to depend on $\underline{Y}$?

c) *i)* Find the first and second moments of X_i, $i = 1, \dots, n$.

ii) Given that $\phi = \phi_0$, some known constant, find $\widehat{\theta}_X$, the method of moments estimator of θ using only the random sample $\underline{X}$.
Deduce the method of moments estimator of θ using only the random sample $\underline{Y}$, denoting this by $\widehat{\theta}_Y$.

iii) Let $\widehat{\theta} = 0.5(\widehat{\theta}_X + \widehat{\theta}_Y)$ be an alternative estimator of θ. Calculate the variance of $\widehat{\theta}$.
Determine which of the three estimators should be preferred.

2. *a)* *i)* Assume that $B_1 \sim \chi^2_{m_1}$ and $B_2 \sim \chi^2_{m_2}$ independently.

Using moment generating functions (mgfs), or otherwise, show that $B_3 = B_1 + B_2 \sim \chi^2_{m_1+m_2}$.

By induction, or otherwise, show that if $B_i \sim \chi^2_{m_i}$ for $i = 1, \ldots, n$, independently, then

$$C = \sum_{i=1}^{n} B_i \sim \chi^2_{\sum m_i}.$$

You may assume that the mgf of χ^2_ν has the form $M(t) = (2-t)^{-\nu/2}$.

ii) Using the method used in *a) i)*, or otherwise, show that if $Q_i \sim N(\mu_i, \sigma_i^2)$, for $i = 1, \ldots, n$, independently, then

$$\sum_{i=1}^{n} Q_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

You may use the form of the mgf of $N(\mu, \sigma^2)$ without deriving it.

b) *i)* Consider $X_1, \ldots, X_n$, a random sample of size n, where

$$X_i \sim \frac{a}{\nu}\chi^2_\nu \quad \text{for } i = 1, \ldots, n.$$

When ν is large and known, prove that the approximate distribution of X_i is $N(\mu, \sigma^2)$, stating the values of μ and σ^2 (you may assume that the central limit theorem holds).

ii) Use *a) ii)* to find the approximate distribution of $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

c) Let (independently)

$$X_{ij} \sim \frac{a_j}{\nu}\chi^2_\nu, \; i = 1, \ldots, n; \; j = 1, \ldots, m.$$

We can consider two estimators of a_j. For the first we assume that the a_j are approximately equal and use an estimator based on all the X_{ij}:

$$\widehat{a}_j = \frac{1}{n}\frac{1}{m}\sum_{i,j} X_{ij}.$$

For the second, the estimator is based only on $X_{1j}, \ldots, X_{nj}$:

$$\widetilde{a}_j = \frac{1}{n}\sum_{i} X_{ij}.$$

Find the mean square errors when estimating a_j using the two methods. Discuss the relative merits of the two estimators in terms of the variability among the a_j.

3. *a)* Define, in general terms, the expected posterior loss of any parameter. Assuming square error loss, find the point estimate $\widetilde{\theta}$ that minimises this function.

b) Consider modelling the value of a signal at $t = 1, \ldots, n$ by

$$X_t \sim N\left(\mu_t, \frac{1}{\lambda}\right),$$

where we assume that the X_t are independent.
Let $\underline{\mu} = (\mu_1, \ldots, \mu_n)^\top$, and suppose that we put a prior on the parameters of the form

$$p(\underline{\mu}, \lambda) = \left(\prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\mu_t^2/2\tau^2}\right) \quad \delta(\lambda - \lambda_0).$$

i) Describe in words what the prior information corresponds to.
Find the posterior distribution.

ii) Find the posterior mean of μ_t. How is this affected by changing the value of τ^2?
Why is the posterior mean of λ not interesting to calculate?

iii) Assume now that $\mu_t = \mu$ for $t = 1, \ldots, n$, and suppose that we change the prior to

$$p(\mu, \lambda | \alpha, \beta) \propto \lambda^{\frac{1}{2}} e^{-\lambda(2\alpha - 1)\frac{\mu^2}{2}} \cdot \lambda^{\alpha-1} e^{-\frac{\lambda}{\beta}}.$$

Find the joint posterior distribution of μ and λ (up to constants of proportionality).
Find the joint posterior mode (you do not have to show that the stationary point corresponds to a maximum).

4. Consider the linear model

$$E(Y_i) = \beta_0 \, |x_i|^{1/2} + \beta_1 \, |x_i|^{3/2}, \quad x_i > 0,$$

for $i = 1, \ldots, n$, where $n > 2$ and

$$var(Y_i) = \sigma^2 x_i^3.$$

a) Why is $x_i > 0$ a necessary restriction? Find the MVULEs (Minimum Variance Unbiased Linear Estimators) of β_0 and β_1, transforming the model if necessary. State the theorem (but do not prove it) that ensures that these estimators are MVULEs.

b) Let

$$x_i = i - 0.9 \quad \text{for} \quad i = 1, \ldots, 10.$$

i) Define, in general, the leverage of observation i in a linear model setting with design matrix $\mathbf{U}$.

ii) In *a)* a linear model was defined to find the MVULEs. Find the leverage of the ith observation using the design matrix of this linear model in terms of x_i WITHOUT substituting in the numbers.

iii) Considering the value of x_i, is there any reason to assume that any one observation has very high leverage?
What does $var(Y_i) = \sigma^2 x_i^3$ say about this observation?

c) Find the covariance matrix of the MVULE $\underline{\widehat{\beta}}$. Assume now that, instead of x-values $\underline{v}$, we might observe at $\underline{q}$, where

$$\begin{aligned} \underline{v} &= (0.1, 1.1, 2.1, \ldots, 9.1), \\ \underline{q} &= (1.1, 2.1, 3.1, \ldots, 10.1). \end{aligned}$$

Which of the two sets of observations is preferred and why?
You may assume that

$$\begin{aligned} \textstyle\sum v_i^{-1} &= 12.69, & \textstyle\sum q_i^{-1} &= 1.34, \\ \textstyle\sum v_i^{-2} &= 101.33, & \textstyle\sum q_i^{-2} &= 2.79, \\ 10\textstyle\sum v_i^{-2} - (\sum v_i^{-1})^2 &= 852.37, & 10\textstyle\sum q_i^{-2} - (\sum q_i^{-1})^2 &= 5.63. \end{aligned}$$

5. Consider the linear model

$$E_{\underline{Y}|\underline{\beta}}\left(\underline{Y}|\underline{\beta}\right) = \beta_0 + \beta_1 p_1(x_i) + \beta_2 p_2(x_i), \quad i = 1, \ldots, n,$$

where the usual normal theory assumptions are made for the errors and

$$\sum_{i=1}^{n} p_1(x_i) = \sum_{i=1}^{n} p_2(x_i) = \sum_{i=1}^{n} p_1(x_i) p_2(x_i) = 0.$$

Let $\underline{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ and $\underline{Y} = (Y_1, \ldots, Y_n)^\top$.

a) State the usual normal theory assumptions and the corresponding second order assumptions. In what way do they differ?

b) Find the minimum variance unbiased linear estimate of $\underline{\beta}$.

c) Suppose that we consider the model above in a Bayesian framework, with σ^2 known and $\beta_2 = 0$. Assume that we assign a prior distribution of the form

$$p_{\beta_0,\beta_1}(\beta) = N\left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \tau^2 & \rho \\ \rho & \tau^2 \end{pmatrix}\right).$$

Write down the likelihood function and use Bayes' theorem to find the posterior density of $\beta_0|\beta_1$.

d) Find the posterior mode of $\beta_0|\beta_1$.