Imperial College
London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2006

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

M2S2

Statistical Modelling

Date: Wednesday, 17th May 2006   Time: 10 am − 12 noon

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.

A Formula Sheet is provided on pages 7-8.

1. (i) Write down a description of the method of moments estimation procedure for an arbitrary set of parameters collected in a vector $\boldsymbol{\theta}$ of length k, based on the random sample $Z_1, \ldots, Z_n$. Each Z_i, $i = 1, \ldots, n$ is independently drawn from the distribution $f_Z(z; \boldsymbol{\theta})$.

(ii) The Pearson type III distribution has, for fixed values of $\alpha \in \mathbb{R}$, $\beta > 0$ and $p > 0$, the form

$$f_X(x) = \begin{cases} \frac{1}{\beta\Gamma(p)} \left(\frac{x-\alpha}{\beta}\right)^{p-1} e^{-\left(\frac{x-\alpha}{\beta}\right)} & \text{if} \quad x > \alpha \\ 0 & \text{if} \quad x > \alpha \end{cases}.$$

Assume that a random sample of size n, denoted $X_1, \ldots, X_n$, is collected from this distribution.

(a) Determine the mean of any of the X_i drawn from the Pearson type III distribution.

(b) It can be shown that the variance of a realisation X_i from the Pearson type III is given by

$$\text{Var}\,[X_i] = p\beta^2. \tag{1}$$

Determine $E\,[X_i^2]$ by using part (a) and equation (1).

(c) Determine the method of moments estimators of α and β for a fixed known value of p, denoting them by $\widehat{\alpha}_M$ and $\widehat{\beta}_M$, respectively.

(d) Form an estimator of the mean of X based on your solution to part (a), and your estimates for α and β. Show that this estimator is unbiased for the mean.

2. The half-normal distribution for a random variable Y has probability density function

$$f_Y(y) = \frac{2\sqrt{a}}{\pi} e^{-\frac{y^2 a}{\pi}}, \;\; y > 0, \; a > 0$$

A random sample of size m, $Y_1, \ldots, Y_m$, is collected from the half-normal distribution.

(i) Write down the likelihood of the data.

(ii) Write down the log-likelihood of the data.

(iii) Determine the maximum likelihood estimator (MLE) of a, denoting this estimator $\widehat{a}_{MLE}$.

(iv) It can be shown that $Y_i^2 \sim Gamma(\frac{1}{2}, \frac{a}{\pi})$. Using moment generating functions, or otherwise, determine the distribution of $V = \sum_{i=1}^{m} Y_i^2$. You may use the form of the MGF of a Gamma random variable from the formula sheet, and any known result about the properties of moment generating functions, as long as you state the result used carefully.

(v) For any p it is assumed that $0 < p < m/2$ and $m > 4$, and in this case calculate $\mathrm{E}\left[V^{-p}\right]$. Thus deduce the mean and variance of $\widehat{a}_{MLE}$.

(vi) Comment on the form of the bias and variance of $\widehat{a}_{MLE}$, for moderate and large values of m.

3. The absolute returns on m different assets of approximately equivalent magnitude are modelled as exponentially distributed random variables, from the same distribution with population mean $1/\lambda$. The assets are assumed to be independent, and the ith asset is denoted by A_i, for $i = 1, \ldots, m$.

(i) (a) Without proof, write down the distribution of $C = \sum_{i=1}^{m} A_i$.

(b) Determine the distribution of $D = \gamma C$, where γ is a positive constant.

(c) Write down the definition of a pivotal quantity for parameter θ based on sample $X_1, \ldots X_m$.

(d) By choosing a suitable value of γ (or otherwise) write down a pivotal quantity for λ, and say why this is a pivotal quantity.

(e) A random sample, $B_1, \ldots, B_n$, is collected of returns on a different day, again modelled as exponentially distributed, but with population mean $1/\lambda'$. Write down a pivotal quantity for λ'.

(ii) The two samples, $A_1, \ldots, A_m$ and $B_1, \ldots, B_n$, are assumed to be independent. We wish to decide between two different hypotheses:

$$H_0 \quad : \quad \lambda = \lambda'$$
$$H_1 \quad : \quad \lambda = 2\lambda'.$$

Describe how to test this at level α, starting from a pivotal quantity for λ/λ', based on the ratio of the two previous pivotal quantities. You may assume that standard distributional results are valid, as long as you state them carefully.

4. The survival times of rats in a sample of n rats are measured. The survival time of the ith rat is denoted by T_i. The ith rat has been exposed to a substance that is a potential toxin at dosage x_i, and T_i has an expectation that is modelled by:

$$\mathrm{E}\left(T_i\right) = \beta_0 + \beta_1 x_i^{-1}, \quad \mathrm{var}\left(T_i\right) = \sigma_i^2, \quad i = 1, \ldots, n. \tag{2}$$

(i) Write down the model in vector and matrix form, defining the vector of dependent observations by $\mathbf{T}$, the design matrix by $\mathbf{X}$ and the vector of parameters by $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 \end{bmatrix}^\top$.

(ii) Write down the formal definition of Second Order Assumptions (SOA). You may for this part, and for the rest of the question make SOA on $\{T_i\}_{i=1}^n$.
Form an estimate of $\boldsymbol{\beta}$; give an explicit expression for the estimate in terms of $\overline{x^{-1}}$, $\overline{x^{-2}}$, $S_{X^{-1},X^{-1}}$, defining such terms appropriately. Justify your choice of estimator, and state any theorems you use.

(iii) Calculate the variance matrix of $\widehat{\boldsymbol{\beta}}$, and denote this $\boldsymbol{A}$.

(iv) The hypothesis of 'the substance is not a toxin,' *i.e.* that an increased level of x does not tend to reduce the mean of the survival time, is to be tested based on the random sample of survival times.

  (a) Write down the null and alternative hypothesis for this test.

  (b) Under normal theory assumptions we may note that

$$\widehat{\boldsymbol{\beta}} \sim N_2\left(\boldsymbol{\beta}, \boldsymbol{A}\right).$$

  Assume that σ_i^2 is known from another study of the rats. Describe how to test the hypothesis: 'the substance is not a toxin.'

5. The temperature at a given weather station in Sussex over time is measured. The readings correspond to Y_i, where $i = 1, \ldots, 3n$ is the time point of the measurements, all made between the start of April and early June. The mean temperature is modelled as increasing linearly over time by

$$\mathrm{E}\left(Y_i\right) = \alpha_1 + \alpha_2 i = \mu_i, \ i = 1, \ldots, 3n.$$

Three different weather-station attendants make the measurements, where attendant 1 took the measurements $i = 1, \ldots, n$, attendant 2 took the measurements $i = n + 1, \ldots, 2n$, and attendant 3 took the measurements $i = 2n + 1, \ldots, 3n$. It is known that the variance of attendant 2 is four times that of attendant 1 and 3 as he has a bit of a problem with his alcohol consumption; attendants 1 & 3 do not drink and their variances are known to be equal, given by some fixed unknown constant σ^2. We assume that the Y_i are not correlated.

(i) Write down the model in vector and matrix form, defining the vector of observations by $\mathbf{Y} = [Y_1, \ldots, Y_{3n}]^T$, the design matrix by $\mathbf{X}$ and the vector of parameters by $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2]^T$.

(ii) Write down the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{Y}$.

(iii) Define a suitable matrix

$$\boldsymbol{D} = \mathrm{diag}\left(d_1, d_2, \ldots, d_{3n}\right),$$

with entries d_i to be specified, so that

$$\mathbf{Z} = \boldsymbol{D}\mathbf{Y}$$

and

$$\mathrm{cov}\left(\mathbf{Z}\right) = c\mathbf{I_n}.$$

(iv) Calculate the minimum variance, unbiased linear estimator of $\boldsymbol{\alpha}$, expressing this estimator in terms of $\mathbf{X}$, $\boldsymbol{D}$ and $\mathbf{Y}$.

(v) The mean value of the temperature is to estimated at time point $i = t$. The estimator is formed by

$$\widehat{\mu}_t = \widehat{\alpha}_0 + \widehat{\alpha}_1 t.$$

Calculate the mean and variance of $\widehat{\mu}_t$. You should denote

$$
\begin{aligned}
s_1(n) &= \sum_{t=1}^{n} t + \sum_{i=2n+1}^{3n} t + \frac{1}{4}\sum_{i=n+1}^{2n} t, \\
s_2(n) &= \sum_{t=1}^{n} t^2 + \sum_{i=2n+1}^{3n} t^2 + \frac{1}{4}\sum_{i=n+1}^{2n} t^2, \\
s_3(n) &= \frac{9n}{4}s_2(n) - s_1^2(n).
\end{aligned}
$$

DISCRETE DISTRIBUTIONS

	RANGE $\mathbb{X}$	PARAMETERS	MASS FUNCTION f_X	CDF F_X	$E_{f_X}[X]$	$\text{Var}_{f_X}[X]$	MGF M_X
$Bernoulli(\theta)$	$\{0,1\}$	$\theta \in (0,1)$	$\theta^x(1-\theta)^{1-x}$		θ	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\dfrac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
or	$\{0,1,2,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n+x-1}{x}\theta^n(1-\theta)^x$		$\dfrac{n(1-\theta)}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta}{1-e^t(1-\theta)}\right)^n$

For **CONTINUOUS** distributions (see over), define the **GAMMA FUNCTION**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

and the LOCATION/SCALE transformation $Y = \mu + \sigma X$ gives

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) \qquad M_Y(t) = e^{\mu t}M_X(\sigma t)$$

$$E_{f_Y}[Y] = \mu + \sigma E_{f_X}[X] \qquad \text{Var}_{f_Y}[Y] = \sigma^2 \text{Var}_{f_X}[X]$$

CONTINUOUS DISTRIBUTIONS

	$\mathbb{X}$	PARAMS.	PDF f_X	CDF F_X	$\mathrm{E}_{f_X}[X]$	$\mathrm{Var}_{f_X}[X]$	MGF M_X
$Uniform(\alpha,\beta)$ (standard model $\alpha=0, \beta=1$)	(α,β)	$\alpha < \beta \in \mathbb{R}$	$\dfrac{1}{\beta-\alpha}$	$\dfrac{x-\alpha}{\beta-\alpha}$	$\dfrac{(\alpha+\beta)}{2}$	$\dfrac{(\beta-\alpha)^2}{12}$	$\dfrac{e^{\beta t}-e^{\alpha t}}{t(\beta-\alpha)}$
$Exponential(\lambda)$ (standard model $\lambda=1$)	$\mathbb{R}^+$	$\lambda \in \mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1-e^{-\lambda x}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda-t}\right)$
$Gamma(\alpha,\beta)$ (standard model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta \in \mathbb{R}^+$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$		$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta-t}\right)^\alpha$
$Weibull(\alpha,\beta)$ (standard model $\beta=1$)	$\mathbb{R}^+$	$\alpha,\beta \in \mathbb{R}^+$	$\alpha\beta x^{\alpha-1}e^{-\beta x^\alpha}$	$1-e^{-\beta x^\alpha}$	$\dfrac{\Gamma(1+1/\alpha)}{\beta^{1/\alpha}}$	$\dfrac{\Gamma(1+2/\alpha)-\Gamma(1+1/\alpha)^2}{\beta^{2/\alpha}}$	
$Normal(\mu,\sigma^2)$ (standard model $\mu=0, \sigma=1$)	$\mathbb{R}$	$\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$e^{\{\mu t+\sigma^2 t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu \in \mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\dfrac{\nu}{2}\right)\left\{1+\dfrac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu>1$)	$\dfrac{\nu}{\nu-2}$ (if $\nu>2$)	
$Pareto(\theta,\alpha)$	$\mathbb{R}^+$	$\theta,\alpha \in \mathbb{R}^+$	$\dfrac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$	$1-\left(\dfrac{\theta}{\theta+x}\right)^\alpha$	$\dfrac{\theta}{\alpha-1}$ (if $\alpha>1$)	$\dfrac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}$ (if $\alpha>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta \in \mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	