Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

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

M2S2

Statistical Modelling I

Date: Monday, 21st May 2007 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.

Statistical tables will not be available.

A Formula Sheet is provided on pages 6-7.

1. Let $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$ denote a random sample from the exponential distribution that has the form for parameter $\alpha > -1$:

$$f_Y(y) = \frac{1}{1+\alpha} \exp\left(\frac{-y}{1+\alpha}\right), \quad y > 0.$$

(i) (a) Write down the likelihood of $\mathbf{Y}$.

(b) Determine the maximum likelihood estimator of α. Denote this estimator $\widehat{\alpha}$.

(c) Find the expectation of $\widehat{\alpha}$.

(ii) (a) Define the bias of a generic estimator $\widehat{\theta}$. Is $\widehat{\alpha}$ unbiased for α?

(b) Denote the minimum of the random sample by $V = \min_{1 \leq i \leq n} \{Y_i\}$. Find the probability density function, and the expectation of V.

(c) Construct an unbiased estimator $\widetilde{\alpha}$ for α of the form $\widetilde{\alpha} = a_n V + b_n$, by identifying appropriate values for a_n and b_n.

2. A random variable K follows the negative binomial distribution, defined for a fixed $r \in \mathbb{N}$ and $0 < p < 1$, by:

$$P(K = k) = \binom{r+k-1}{k} p^r (1-p)^k, \; k = 0, \; 1, \ldots. \tag{1}$$

(i) Describe the method of moments procedure, for any random sample $K_1, \ldots, K_n$, for estimating the mean, denoted μ_K, of the distribution $\{K_i\}$ were drawn from.

(ii) Calculate the expectation of K, whose distribution is specified by equation (1).

(iii) A random sample of size n is drawn from the distribution in equation (1). Derive the method of moments estimator of $\theta = (1-p)/p$, denoting this estimator $\widehat{\theta}$.

(iv) Determine the mean and variance of $\widehat{\theta}$. You may use the expression for the variance of a negative binomial given in the formula sheet.

(v) If the pmf (probability mass function) of equation (1) is rewritten as a function of θ it becomes:

$$P(K = k) = \binom{r+k-1}{k} \left(\frac{1}{1+\theta}\right)^r \left(\frac{\theta}{1+\theta}\right)^k, \; k = 0, \ldots, n.$$

Determine the MLE of θ and compare this with the method of moments estimator. You need not check that the stationary point corresponds to a maximum.

3. Let $\mathbf{U} = [U_1, \ldots, U_n]^T$ denote a random sample from a Gaussian distribution with (unknown) mean μ and (known) variance σ_o^2. Assume you wish to pursue a Bayesian analysis of the data. A prior with hyperparameters (μ_0, τ) is suggested:

$$p_\mu(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\mu-\mu_0)^2}$$

(i) Calculate the posterior distribution for μ based on the random sample $\mathbf{U}$, and the prior $p_\mu(\mu)$.

(ii) Under square error loss, give the optimal point estimate of μ.

(iii) Interpret the point summary in terms of the Maximum Likelihood estimator of the mean $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n U_i$ and the prior mean μ_0.

(iv) (a) Interpret the role of μ_0 and τ in the prior specification, and indicate any values of μ_0 that are particularly appropriate to chose, and why these are appropriate.

(b) What degree of precision are we expressing for the prior as $\tau \to 0$ and $\tau \to \infty$?

(c) If we could use two different priors $p_\mu^{(A)}(\mu)$ and $p_\mu^{(B)}(\mu)$, depending on if we consulted expert A or B, with the some set of hyperparameters $(\mu_o^{(A)}, \tau^{(A)})$ and $(\mu_o^{(B)}, \tau^{(B)})$, where $\tau^{(B)} >> \tau^{(A)}$, which should we prefer? Discuss the preference in terms of the posterior variance and give your reasoning.

4. The heights of inhabitants of city A are modelled as coming from the distribution:

$$f_A(a) = \frac{1}{\Gamma(\alpha_0)\theta_A}\left(\frac{a}{\theta_A}\right)^{\alpha_0-1} e^{-\frac{a}{\theta_A}},\ a>0,\ \mathrm{E}\left(A\right) = \alpha_0\theta_A,$$

where α_0 (known) and θ_A (unknown) are positive parameters. A random sample of size n heights, denoted $\mathbf{A} = [A_1,\ A_2,\ \ldots,\ A_n]^T$, are collected from the population at large to test the hypothesis that the population mean is (at least) 6 feet tall.

(i) (a) Using Moment Generating Functions (MGFs) (or otherwise) determine the distribution of the sum of all observed heights $S_A = \sum_{i=1}^n A_i$. You may use the fact that the MGF of any A_i is given by $M_{A_i}(t) = 1/(1-\theta_A t)^{\alpha_0}$. Deduce the distribution of the sample mean $\overline{A} = \frac{1}{n}S_A$.

(b) Determine the distribution of $X = 2n\overline{A}/\theta_A$.

(c) Using X as a pivotal quantity, test the hypothesis that:

$$H_0: \quad E\left\{A\right\} = \alpha_0\theta_A = 6$$
$$H_1: \quad E\left\{A\right\} = \alpha_0\theta_A < 6.$$

(ii) The Mayor of city B makes the claim that the inhabitant of city B are equally tall as those of city A. A random sample size m of heights is therefore collected from the city, collected in vector $\mathbf{B} = [B_1,\ B_2,\ \ldots,\ B_m]^T$. The distribution of B_i is given by:

$$f_B(b) = \frac{1}{\Gamma(\alpha_0)\theta_B}\left(\frac{b}{\theta_B}\right)^{\alpha_0-1} e^{-\frac{b}{\theta_B}},\ b>0,$$

where we assume that $\alpha_0 = 2k$, takes the same value for both distributions, and that this value is known. Using $T = \frac{\overline{A}\theta_B}{\overline{B}\theta_A}$ as a pivotal quantity test the hypothesis that

$$E\left\{A\right\}/E\left\{B\right\} = \frac{\theta_A}{\theta_B} = 1, \quad \text{vs} \quad E\left\{A\right\}/E\left\{B\right\} \neq 1.$$

State any assumptions that you use carefully.

5. The marks in a statistics exam are recorded for a sample of n mathematics students, and are denoted, $Z_1, \ldots, Z_n$, respectively. It is anticipated that the mark of student i, can be explained in terms of the previous year's average mark of student i, x_i, and the number of lecture hours student i attended, namely h_i. To this purpose a model of the form:

$$Z_i = \beta_1 + \beta_2 x_i + \beta_3 h_i + \epsilon_i, \; i = 1, \ldots, n$$

is proposed.

(i) State the Second Order Assumptions.

(ii) State and prove the Gauss-Markov Theorem.

(iii) Assume you are given a set of observations $\{Z_i, \; x_i, \; h_i\}$. Specify the matrices needed to calculate $\widehat{\boldsymbol{\beta}}$, the least squares estimator of $\boldsymbol{\beta}$, and provide the form of $\widehat{\boldsymbol{\beta}}$, in terms of these matrices. Do NOT calculate or simplify the exact algebraic form of the estimators, but leave the estimators in terms of the specified matrices.

(iv) It is taken as given that the previous year's averages are required in the model. State the hypothesis 'Attending lectures has no effect on the exam mark', in terms of the parameters of the model. With Normal Theory Assumptions, outline how to test this assumption from the data, vs an alternative that lecture attendance *improves* the mark. State any results you use carefully, and motive your derivation.

DISCRETE DISTRIBUTIONS

	RANGE $\mathbb{X}$	PARAMETERS	MASS FUNCTION f_X	CDF F_X	$\mathsf{E}_{f_X}[X]$	$\mathsf{Var}_{f_X}[X]$	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	$\exp\left\{\lambda\left(e^t-1\right)\right\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\dfrac{1}{\theta}$	$\dfrac{(1-\theta)}{\theta^2}$	$\dfrac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\dfrac{n}{\theta}$	$\dfrac{n(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
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

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) \qquad M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathsf{E}_{f_Y}[Y] = \mu + \sigma\mathsf{E}_{f_X}[X] \qquad \mathsf{Var}_{f_Y}[Y] = \sigma^2\mathsf{Var}_{f_X}[X]$$

CONTINUOUS DISTRIBUTIONS

	$\mathbb{X}$	PARAMS.	PDF f_X	CDF F_X	$\mathsf{E}_{f_X}[X]$	$\mathsf{Var}_{f_X}[X]$	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	$e^{\{\mu t+\sigma^2 t^2/2\}}$
$Student(\nu)$	$\mathbb{R}$	$\nu\in\mathbb{R}^+$	$\dfrac{(\pi\nu)^{-\frac{1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\left\{1+\frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		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 $\alpha>2$)	
$Beta(\alpha,\beta)$	$(0,1)$	$\alpha,\beta\in\mathbb{R}^+$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$		$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	