UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

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

M2S3 Statistical Theory I

Date: Tuesday, 22nd May 2007 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.

Formula sheets are given on pages 5 and 6.

1. (i) (a) Write down an expression for the probability mass function (pmf) or probability density function of the one-parameter exponential family of distributions (EF).

(b) Give two reasons for the importance of the EF in parameter estimation.

(ii) The generalised power series distributions (GPS) are discrete distributions having their pmf in the form

$$f_{X|\theta}(x \mid \theta) \;=\; a_x\, \theta^x e^{\beta(\theta)} \qquad (x = 0, 1, \ldots),$$

with $\theta > 0$, $a_x \geq 0$, where for each value of x the a_x are given constants, and where $\beta(\theta)$ is a normalising constant making the total probability 1.

(a) Show that GPS distributions are special cases of the EF.

(b) If $\beta(\theta)$ is twice differentiable, show that we can write the score $U(\theta)$ and Fisher information $I(\theta)$ for a single random variable X from GPS in a regular estimation problem for a single unknown parameter θ as

$$U(\theta) \;=\; \frac{X}{\theta} \;+ \beta'(\theta), \qquad I(\theta) \;=\; \frac{E(X)}{\theta^2} \;- \beta''(\theta)\,.$$

(c) Using the expression for $U(\theta)$ given in (b), find an expression for $E_{X|\theta}(X)$ in terms of θ and $\beta(\theta)$.

(d) Show that $\operatorname{var}_{X|\theta}(X) = -\theta\beta'(\theta) - \theta^2\beta''(\theta)$. *Give your reasoning.*

(e) Show that *Binomial* (n, ψ) is a special case of GPS by identifying a_x, θ and $\beta(\theta)$.

2. We have a random sample $\boldsymbol{x} = \{x_1\,, x_2\,, \ldots\,, x_n\}$ from *Poisson* (μ) $(\mu > 0)$.

(a) Show that the sample mean $\overline{x}$ is a sufficient statistic for μ.
Show that $\overline{x}$ is the maximum likelihood estimate of μ.

(b) For fixed $k \neq 0$, let

$$t(\boldsymbol{x}) \;=\; \begin{cases} 1 & (x_1 = k), \\ 0 & (x_1 \neq k). \end{cases}$$

Show that $t(\boldsymbol{x})$ is unbiased for $\theta \;=\; \dfrac{\mu^k}{k!}\; e^{-\mu}$.

(c) Apply the Rao-Blackwell Theorem to show that $\dbinom{s}{k}\left(\dfrac{1}{n}\right)^k\left(1 - \dfrac{1}{n}\right)^{s-k}$, where $s = n\overline{x}$, is an unbiased estimate of θ.
Explain, without proof, but stating a further property of $\overline{x}$ and theorems used, why we may take this estimate to be a uniformly minimum variance unbiased estimate of θ.

(d) Comment on the relationship of this estimate to the maximum likelihood estimate for θ as n becomes large for fixed k.

3. (i) For an interval estimation problem for a single unknown parameter, explain briefly what is meant by a *pivotal quantity* and a *best confidence set* (*ie the shortest*).

(ii) $X_1, X_2, \ldots, X_n$ are independent random variables, where X_k has the probability density function

$$f_{X_k|\theta}(x_k \mid \theta) = \begin{cases} \exp\{-(x_k - k\theta)\} & (x_k \geq k\theta), \\ 0 & \text{(otherwise)}, \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) From observed values $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ write down the likelihood function $\ell(\theta|\boldsymbol{x})$.
Show that $t = \min\limits_k \left(\dfrac{x_k}{k}\right)$ is sufficient for θ.

(b) Show that $Y_k = \dfrac{X_k}{k} - \theta$ is *Exponential* (k) for each k.
Find the distribution of $Z = \min\limits_k(Y_k)$.
Identify Z as a pivotal quantity, and hence find the sampling distribution of t.

(c) Show that the best $100(1-\alpha)\%$ confidence set for θ is given by $P(0 < Z < \xi)$, where ξ is to be found as a function of α.
Hence find the best confidence set for θ.

4. We have a random sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ from *Pareto* $(\theta, 2)$ which has the probability density function

$$f_{X|\theta}(x \mid \theta) = \begin{cases} \dfrac{\theta 2^\theta}{x^{\theta+1}} & (2 \leq x < \infty), \\ 0 & \text{(otherwise)}, \end{cases}$$

where $\theta > 0$ is an unknown parameter.

(a) Show that the maximum likelihood estimate $\widehat{\theta}$ of θ is $\left\{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \ln\left(\dfrac{1}{2}x_i\right)\right\}^{-1}$.

(b) To test $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$, show that the likelihood ratio test statistic

$$\lambda(\boldsymbol{x}) = \frac{\ell(\widehat{\theta} \mid \boldsymbol{x})}{\ell(1 \mid \boldsymbol{x})},$$

where $\ell(\theta|\boldsymbol{x})$ is the likelihood function, is such that

$$\Lambda(\boldsymbol{x}) = \ln \lambda(\boldsymbol{x}) = n\left(\frac{1}{\widehat{\theta}} + \ln \widehat{\theta} - 1\right).$$

(c) By showing that $\Lambda(\boldsymbol{x})$ has a unique minimum at $\widehat{\theta} = 1$, show that the test rejects H_0 if $\widehat{\theta}$ is too large or too small relative to value 1.
Show that the acceptance set can be written in the form

$$c_1 < \sum_i \ln x_i < c_2 .$$

(d) By showing that $\ln\left(\frac{1}{2}X\right)$ is *Exponential* (θ), describe precisely how you would obtain the values of c_1 and c_2 for a size-α test.

5. We have a random sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$ from $N(\theta, 1)$, where θ is an unknown parameter.

(a) Use the improper prior probability density function $\pi(\theta) \propto 1$ $(-\infty < \theta < \infty)$ to identify the posterior probability density function $\pi(\theta|\boldsymbol{x})$ as the probability density function (pdf) of a normal distribution.

Write down the posterior mean, posterior variance, and a 95% credible region for θ.
[*You may use that* $\Phi(1.96) = 0.975$, *where* Φ *is the cumulative distribution function for* $N(0,1)$.]

(b) Suppose that we have prior knowledge that $\theta > 0$. Using the improper prior pdf

$$\pi(\theta) \propto \begin{cases} 1 & (0 < \theta < \infty), \\ 0 & \text{(otherwise)}, \end{cases}$$

obtain the posterior pdf $\pi(\theta|\boldsymbol{x})$.

Find a 95% credible region for θ of the form $(0, \xi)$, where ξ is to be obtained in terms of Φ and the data.

Distribution	$f(x \mid \theta)$	$x \in \mathbb{X}$	$\theta \in \Theta$
Bernoulli (θ)	$\theta^x(1-\theta)^{1-x}$	$x = 0, 1$	$0 < \theta < 1$
(Discrete) Uniform (k)	$\frac{1}{k}$	$x = 1, 2, \ldots, k$	$k = 1, 2, \ldots$
Binomial(n, θ)	$\binom{n}{x} \theta^x(1-\theta)^{n-x}$	$x = 0, 1, \ldots, n$	$0 < \theta < 1$
Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$	$x = 0, 1, 2, \ldots$	$\lambda > 0$
Geometric(θ)	$(1-\theta)^{x-1}\theta$	$x = 1, 2, \ldots$	$0 < \theta < 1$
Negative Binomial (ν, θ)	$\binom{x+\nu-1}{x}(1-\theta)^x\theta^\nu$	$x = 0, 1, 2, \ldots$	$0 < \theta < 1,\ \nu > 0$
Uniform(α, β)	$\frac{1}{\beta - \alpha}$	$\alpha < x < \beta$	$\alpha < \beta$
Exponential(λ)	$\lambda \exp(-\lambda x)$	$x > 0$	$\lambda > 0$
Gamma(ν, λ)	$\frac{1}{\Gamma(\nu)}\ \lambda(\lambda x)^{\nu-1}\exp(-\lambda x)$	$x > 0$	$\lambda > 0,\ \nu > 0$
Cauchy (α, β)	$\frac{1}{\pi\beta}\left\{1 + \left(\frac{x-\alpha}{\beta}\right)^2\right\}^{-1}$	$-\infty < x < \infty$	$\beta > 0$
$N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	$-\infty < x < \infty$	$\sigma > 0$
Beta (α, β)	$\frac{1}{B(\alpha, \beta)}\ x^{\alpha-1}(1-x)^{\beta-1}$	$0 < x < 1$	$\alpha > 0,\ \beta > 0$
Weibull (α, β)	$\beta\alpha x^{\alpha-1}\exp(-\beta x^\alpha)$	$x > 0$	$\alpha > 0,\ \beta > 0$
Gumbel (α, β)	$\frac{1}{\beta}\ \exp\left(-\frac{x-\alpha}{\beta}\right)\exp\left\{-\exp\left(-\frac{x-\alpha}{\beta}\right)\right\}$	$-\infty < x < \infty$	$\beta > 0$
Pareto (θ, α)	$\frac{\theta\alpha^\theta}{x^{\theta+1}}$	$x > \alpha$	$\theta > 0,\ \alpha > 0$
Chi-square χ^2_k	$\frac{1}{2^{k/2}\ \Gamma(k/2)}\ x^{(k/2)-1}\exp\left(-\frac{1}{2}x\right)$	$x > 0$	$k = 1, 2, \ldots$
Student t_m	$\frac{\Gamma((m+1)/2)}{\Gamma(m/2)\sqrt{\pi m}}\ \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}$	$-\infty < x < \infty$	$m = 1, 2, \ldots$

Distribution	$E(X)$	$\mathrm{var}\,(X)$	$G_X(z)$ or $M_X(t)$
Bernoulli (θ)	θ	$\theta(1-\theta)$	$1-\theta+\theta z$
Uniform(k)	$(k+1)/2$	$(k^2-1)/12$	$z(1-z^k)/\{k(1-z)\}$
Binomial(n,θ)	$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta z)^n$
Poisson(λ)	λ	λ	$\exp\{-\lambda(1-z)\}$
Geometric(θ)	$\dfrac{1}{\theta}$	$\dfrac{1-\theta}{\theta^2}$	$\dfrac{\theta z}{1-z(1-\theta)}$
Negative Binomial (ν,θ)	$\dfrac{\nu(1-\theta)}{\theta}$	$\dfrac{\nu(1-\theta)}{\theta^2}$	$\left(\dfrac{\theta z}{1-z(1-\theta)}\right)^\nu$
Uniform(α,β)	$(\alpha+\beta)/2$	$(\beta-\alpha)^2/12$	$\dfrac{e^{\beta t}-e^{\alpha t}}{(\beta-\alpha)t}$
Exponential(λ)	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda-t)$
Gamma(ν,λ)	ν/λ	ν/λ^2	$\{\lambda/(\lambda-t)\}^\nu$
Cauchy (α,β)	none	none	none
$N(\mu,\sigma^2)$	μ	σ^2	$\exp(\mu t+\frac{1}{2}\sigma^2t^2)$
Beta (α,β)	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	${}_1F_1(\alpha;\beta;t)$
Weibull (α,β)	$\beta^{-1/\alpha}\Gamma\left(1+\dfrac{1}{\alpha}\right)$	$\beta^{-2/\alpha}\left\{\Gamma\left(1+\dfrac{2}{\alpha}\right) - \left[\Gamma\left(1+\dfrac{1}{\alpha}\right)\right]^2\right\}$	none
Gumbel (α,β)	$\alpha+\beta\gamma \quad (\gamma=0.5772\cdots)$	$\pi^2\beta^2/6$	$e^{\alpha t}\,\Gamma(1-\beta t) \quad (t<1/\beta)$
Pareto (θ,α)	$\dfrac{\alpha\theta}{\theta-1} \quad (\theta>1)$	$\dfrac{\alpha^2\,\theta}{(\theta-1)^2(\theta-2)} \quad (\theta>2)$	none
Chi-square χ^2_k	k	$2k$	$(1-2t)^{-k/2}$
Student t_m	$0 \quad (m=2,3,\ldots)$	$\dfrac{m}{m-2} \quad (m=3,4,\ldots)$	none