UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Thursday 7 June 2001 1.30 to 4.30

PAPER 32

STATISTICAL THEORY

Attempt any ***FOUR*** *questions. The questions carry equal weight.*

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1 Let a d-dimensional parameter vector θ be partitioned as $\theta = (\psi, \lambda)$.

Explain what is meant by *orthogonality* of ψ and λ.

Discuss briefly the consequences of parameter orthogonality for maximum likelihood estimation.

Suppose that Y is distributed according to a density of the form

$$P_Y(y;\theta) = a(\lambda, y) exp\{\lambda t(y;\psi)\}.$$

Show that ψ and λ are orthogonal.

2 Write a brief account of the concept and properties of *profile likelihood*.

Define what is meant by *modified profile likelihood*.

Let $Y_1, \ldots, Y_n$ be independent, identically distributed according to an inverse Gaussian distribution with density

$$\{\psi/(2\pi y^3)\}^{1/2} \; exp \; \{-\frac{\psi}{2\lambda^2 y} \; (y-\lambda)^2\}, \quad y > 0$$

where $\psi > 0$ and $\lambda > 0$. The parameter of interest is ψ.

Find the form of the profile log-likelihood function and of the modified profile log-likelihood.

3 (i) Let $Y_1, \ldots, Y_n$ be independent, identically distributed random variables with density $f_Y(y)$ and cumulant generating function $K_Y(t)$.

Describe in detail the *saddlepoint approximation* to the density of

$$\overline{Y} = n^{-1} \sum_{i=1}^{n} Y_i.$$

(ii) Let $Y_1, \ldots, Y_n$ be independent random variables each with a Laplace density

$$f_Y(y) = \exp\{-|y|\}/2, \; -\infty < y < \infty.$$

Show that the cumulant generating function is $K_Y(t) = -\log(1-t^2), \; |t| < 1$, and derive the form of the saddlepoint approximation to the density of $\overline{Y}$.

4 Explain what is meant by an *M-estimator* of a parameter θ, based on a given ψ function. Show that the influence function is proportional to ψ and derive an expression for the asymptotic variance of the M-estimator at a distribution F.

A location model on $\mathbb{R}$, with parameter space $\mathbb{R}$, is given by $F_\theta(x) = F(x-\theta)$, and an M-estimator is to be constructed using a ψ function of the form $\psi(x,\theta) = \psi(x-\theta)$. Let $IF(x;\psi,F)$ and $V(\psi,F)$ denote the influence function and asymptotic variance, respectively, at a distribution F, and let Φ denote the standard normal distribution. Show that the problem of choosing an odd, non-decreasing ψ function which minimises $V(\psi,\Phi)$ among all estimators with

$$|IF(x;\psi,\Phi)| \leqslant C < \infty,$$

for given $C \geqslant \sqrt{\pi/2}$, is solved by

$$\psi(x) = \max\{-K, \min\{x, K\}\},$$

with $C = K/\{2\Phi(K)-1\}$.

5 (i) What is meant by a *maximal invariant statistic* with respect to a group of transformations on a sample space?

State and prove a result which establishes the importance of maximal invariants in the construction of non-parametric tests.

(ii) Let $X_1,\ldots,X_n$ be independent, identically distributed with continuous distribution function F_X, and $Y_1,\ldots,Y_m$ be independent, identically distributed from a continuous distribution function F_Y.

Describe the *Wilcoxon test* of $H_0 : F_X(z) = F_Y(z), \forall z$, against $H_1 : F_X(z) \geqslant F_Y(z), \forall z$, and justify the test in terms of the discussion in (i) above.

State, without proof, the asymptotic null distribution of the Wilcoxon test statistic.

6 Write an account of *one* of the following:

(i) Edgeworth and Laplace approximations;

(ii) The p^*-formula for the density of the maximum likelihood estimator;

(iii) Exponential families and transformation models.