UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Monday 5 June 2006   1.30 to 4.30

STATISTICAL THEORY

*Attempt **FOUR** questions, not more than **TWO** of which should be from Section B.*
*There are **TEN** questions in total.*
Marks for each question are indicated on the paper in square brackets.
Each question is worth a total of 20 marks.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script paper

SPECIAL REQUIREMENTS
None

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

Section A

1 Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with distribution function F. Define the empirical distribution function $\hat{F}_n$. State and prove the Glivenko–Cantelli theorem. [10]

Define the pth sample quantile $\hat{F}_n^{-1}(p)$. Subject to a smoothness condition which you should specify, write down the asymptotic distribution of the sample median, $\hat{F}_n^{-1}(1/2)$. [3]

In each of the two cases below, compare the asymptotic variance of $n^{1/2}\ \hat{F}_n^{-1}(1/2)$ with that of $n^{1/2}\bar{X}_n$, where $\bar{X}_n = n^{-1}(X_1 + \ldots + X_n)$:

(i) $F = \Phi$, the standard normal distribution function [3]

(ii) F has density $f(x) = 6x(1-x)$ for $x \in (0,1)$. [4]

2 Let $Y_1, \ldots, Y_n$ be independent and identically distributed with model function $f(y;\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^d$, and let θ_0 denote the true parameter value. Derive the asymptotic distribution of the maximum likelihood estimator $\hat{\theta}_n$. [8]

[*You may assume that the usual regularity conditions hold. In particular, you may assume a Taylor expansion for the score function $U(\theta)$, of the form*

$$0 = U(\hat{\theta}_n) = U(\theta_0) - j(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(n^{1/2}),$$

as $n \to \infty$, where $j(\theta)$ is the observed information matrix at θ.]

Describe how this asymptotic result is related to the Wald test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. Now suppose that $\theta = (\psi, \lambda)$, where only ψ is of interest. Describe the Wald test of $H_0 : \psi = \psi_0$ against $H_1 : \psi \neq \psi_0$. [4]

Let $Y_1, \ldots, Y_n$ be independent and identically distributed with the inverse Gaussian density

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

Show that the maximum likelihood estimator of ψ is

$$\hat{\psi} = \left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{Y_i} - \frac{1}{\bar{Y}}\right)\right\}^{-1},$$

where $\bar{Y} = n^{-1}(Y_1 + \ldots + Y_n)$. [4]

Using the fact that $\mathbb{E}_{\psi,\lambda}(Y_1) = \lambda$, show further that the Wald statistics for testing $H_0 : \psi = \psi_0$ against $H_1 : \psi \neq \psi_0$ coincide in the two cases where λ is known and where λ is unknown. [4]

3 Let $X_1, \ldots, X_n$ be independent and identically distributed with distribution function F, and let $X_{(n)} = \max_i X_i$. If G is a non-degenerate distribution function, what does it mean for F to belong to the domain of attraction $D(G)$ of G? What does it mean for G to be max-stable? Prove that $D(G)$ is non-empty if and only if G is max-stable. [7]

[*You may assume that if (F_n) is a sequence of distribution functions satisfying $F_n(a_n x + b_n) \xrightarrow{d} G_1(x)$ as $n \to \infty$ and $F_n(\alpha_n x + \beta_n) \xrightarrow{d} G_2(x)$, for non-degenerate G_1, G_2, then $G_1(x) = G_2(ax + b)$, for some $a \in (0, \infty), b \in \mathbb{R}$.*]

Let $F(x) = 1 - 1/(x \log x)$ for $x > x_0$, where $x_0 \log x_0 = 1$. By quoting a result about regular variation, or otherwise, find a non-degenerate distribution function G such that $F \in D(G)$. Give expressions for constants $a_n > 0$ and b_n such that, for all $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\frac{X_{(n)} - b_n}{a_n} \leqslant x\Big) \to G(x),$$

as $n \to \infty$. [4]

By writing down an equation satisfied by $F(a_n)$, show first that there exists $n_0 \in \mathbb{N}$ such that $a_n < n$ for $n \geqslant n_0$. Show further that $a_n > n/\log n$ for $n \geqslant n_0$, and finally that

$$a_n < \frac{n}{\log n - \log\log n}$$

for $n \geqslant n_0$. Deduce that, for all $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\frac{X_{(n)} \log n}{n} \leqslant x\Big) \to G(x) \qquad [9]$$

as $n \to \infty$.

4 Write an essay on exponential families, which should include the following:

(i) The definition of a full natural exponential family of order p

(ii) A calculation of the moment generating function of a random variable Y with density in full natural exponential family form, and of expressions for the mean vector and covariance matrix of Y

(iii) The general definition of an exponential family of order p, and of a (p, q) curved exponential family, together with an example of the latter

(iv) An explanation of the existence and uniqueness of maximum likelihood estimators in regular natural exponential families. [20]

5 Let f be a bounded density with a bounded, continuous second derivative f'' satisfying $\int_{-\infty}^{\infty} f''(x)^2 \, dx < \infty$, and let $X_1, \ldots, X_n$ be independent and identically distributed with density f. Define the kernel density estimator $\hat{f}_h(x)$ with kernel K and bandwidth h. Under conditions on h and K which you should specify, derive the leading term of an asymptotic expansion for the bias of $\hat{f}_h(x)$ as a point estimator of $f(x)$. [10]

Observing that $\mathrm{Var}\{\hat{f}_h(x)\} = (nh)^{-1} R(K) f(x) + o\{1/(nh)\}$, where $R(K) = \int_{-\infty}^{\infty} K(z)^2 \, dz$, and provided that $f''(x) \neq 0$, find the bandwidth $h_{AMSE}(x)$ which minimises the asymptotic mean squared error of $\hat{f}_h(x)$ at the point x. Write down (or compute) the asymptotically optimal mean integrated squared error bandwidth, h_{AMISE}. [3]

For $f(x) = \phi(x)$, the standard normal density, show that

$$\inf_{x \in \mathbb{R} \setminus \{-1,1\}} \frac{h_{AMSE}(x)}{h_{AMISE}} = \left(\frac{9e^5}{8192} \right)^{1/10}. \quad [7]$$

[*You may find it helpful to note that* $R(\phi'') = \frac{3}{8\sqrt{\pi}}$.]

6 Let $g : (a,b) \to \mathbb{R}$ be a smooth function with a unique minimum at $\tilde{y} \in (a,b)$ satisfying $g''(\tilde{y}) > 0$. Sketch a derivation of Laplace's method for approximating

$$g_n = \int_a^b e^{-ng(y)} \, dy. \quad [7]$$

[*You may treat error terms informally. An explicit expression for the* $O(n^{-1})$ *term is not required.*]

By making an appropriate substitution, use Laplace's method to approximate

$$\Gamma(n+1) = \int_0^\infty y^n e^{-y} \, dy. \quad [7]$$

Let $p(\theta)$ denote a prior for a parameter $\theta \in \Theta \subseteq \mathbb{R}$, and let $Y_1, \ldots, Y_n$ be independent and identically distributed with conditional density $f(y|\theta)$. Explain how Laplace's method may be used to approximate the posterior expectation of a function $g(\theta)$ of interest. [6]

Section B

7 Consider the linear regression

$$Y = X\beta + \epsilon$$

where Y is an n-dimensional observation vector, X is an $n \times p$ matrix of rank p, and ϵ is an n-dimensional vector with components $\epsilon_1, \ldots, \epsilon_n$. Here, $\epsilon_1, \ldots, \epsilon_n$ are normally and independently distributed, each with mean zero and variance σ^2; we write this as $\epsilon \sim N_n(0, \sigma^2 I_n)$.

(a) Define $Q(\beta) = (Y - X\beta)^T(Y - X\beta)$. Find an expression for $\hat{\beta}$, the least squares estimator for β and state without proof the joint distribution of $\hat{\beta}$ and $Q(\hat{\beta})$. [10]

(b) Define $\hat{\epsilon} = Y - X\hat{\beta}$. Find the distribution of $\hat{\epsilon}$. [6]

(c) Suppose $\beta^T = (\beta_1^T : \beta_2^T)$. How would you test $H_0 : \beta_2 = 0$? [4]

8 Suppose that we have independent observations $Y_1, \ldots, Y_n$ and that we assume the model

$$\omega : Y_i \text{ is Poisson, parameter } \mu_i \text{ and } \log(\mu_i) = \beta_0 + \beta_1 x_i,$$

where $x_1, \ldots, x_n$ are given scalar covariates.

(a) Find the equations for the maximum likelihood estimators $\hat{\beta}_0, \hat{\beta}_1$ and state without proof the approximate distribution of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$. [9]

(b) How is the deviance for ω calculated? If you found that this deviance took the value 32.4, where $n = 35$, what would you conclude? [5]

(c) Discuss briefly how your answers to the above would change if the model ω is replaced by the model [6]

$$\omega' : \mu_i = \beta_0 + \beta_1 x_i \; .$$

9 Write an account, with appropriate examples of the decision theory approach to inference. Your account should include discussion of *all* of the following:

(i) the main elements of a decision theory problem;

(ii) the Bayes and minimax principles;

(iii) admissibility;

(iv) finite decision problems;

(v) decision theory approaches to point estimation and hypothesis testing.

(Proofs of results are not expected.) [20]

10 Write an account of the main results in the frequentist (Neyman-Pearson) theory of optimal hypothesis testing. Your account should include discussion of all of the following:

(i) size of a test;

(ii) Neyman-Pearson lemma;

(iii) Uniformly most powerful tests; and

(iv) unbiased tests.

(Proofs of results are not expected.) [20]

END OF PAPER