UNIVERSITY OF
CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Thursday, 2 June, 2011 1:30 pm to 4:30 pm

PAPER 32

STATISTICAL THEORY

*Attempt no more than **FOUR** questions.*

*There are **SIX** questions in total.*

The questions carry equal weight.

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.

1

Suppose $\Theta \subset \mathbb{R}^p$ is compact, and assume $Q : \Theta \to \mathbb{R}$ is a continuous function. Let $(\Omega, \mathcal{A}, P)$ be a probability space and suppose $Q_n(\theta) \equiv Q_n(\theta, \omega), Q_n : \Theta \times \Omega \to \mathbb{R}$ is a random function on Θ, continuous in $\theta \in \Theta$ for every fixed $\omega \in \Omega$. Assume that θ_0 is the unique minimizer of Q over Θ. Prove that if

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \to^P 0 \tag{1}$$

as $n \to \infty$, then any solution $\hat{\theta}_n$ of

$$\min_{\theta \in \Theta} Q_n(\theta)$$

converges to θ_0 in probability as $n \to \infty$. Show by means of a counterexample that the conclusion may not hold if (1) is replaced by the weaker condition $Q_n(\theta) \to^P Q(\theta)$ for every $\theta \in \Theta$ as $n \to \infty$.

Consider next a statistical model consisting of probability densities $\{f(\theta, \cdot) : \theta \in \Theta\}$ with respect to Lebesgue measure, and let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables from some density $f(\theta_0, \cdot)$ on $\mathcal{Y} \subset \mathbb{R}$, where $\theta_0 \in \Theta$. Define the log-likelihood function $l_n(\theta)$ associated with this model, and define the maximum likelihood estimator $\hat{\theta}_n$ of θ. Assuming that $\int_{\mathcal{Y}} |\log f(\theta, y)| f(\theta_0, y) dy < \infty$ for every $\theta \in \Theta$, that $f(\theta, y) > 0$ for every $\theta \in \Theta$ and every $y \in \mathcal{Y}$, and that

$$f(\theta_0, \cdot) = f(\theta_1, \cdot) \text{ Lebesgue almost everywhere} \Leftrightarrow \theta_0 = \theta_1,$$

show that

$$Q(\theta) = -E_{\theta_0} l_n(\theta)$$

has a unique minimiser θ_0 in Θ.

2

Let $X_1, \ldots, X_n$ be i.i.d. with arbitrary probability density function $f : \mathbb{R} \to [0, \infty)$, and denote by E expectation under the joint distribution of $X_1, \ldots, X_n$. Define the kernel density estimator $f_n(h, \cdot)$ of f based on bandwidth $h > 0$ and kernel K. Let the kernel equal $K(x) = 1_{[-1/2,1/2]}(x)$, and assuming $h_n \to 0$ but $nh_n \to \infty$ as $n \to \infty$, show that

$$E \int_{\mathbb{R}} |f_n(h_n, x) - f(x)| \, dx \to 0$$

as $n \to \infty$. Derive the rate of convergence of this quantity to zero in the case where f equals the standard normal density function and when $h_n \sim 1/\log n$.

[You may use results from measure theory, provided they are clearly stated, including that $(2a_n)^{-1} \int_{x-a_n}^{x+a_n} f(y) dy \to f(x)$ for any sequence $a_n \to 0$ and almost every $x \in \mathbb{R}$, the dominated convergence theorem, and that the mapping $a \mapsto \int_{\mathbb{R}} |f(x+a) - f(x)| dx$ is continuous at zero.]

3

Suppose we observe $Y_1, \ldots, Y_n$ independent identically distributed random variables with probability density $f(\theta, \cdot), \theta \in \Theta \subset \mathbb{R}^p$, and consider the testing problem $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta \setminus \{\theta_0\}$, where θ_0 is an interior point of Θ. Define the likelihood ratio test statistic Λ_n for this testing problem, and express it in terms of the maximum likelihood estimator $\hat{\theta}_n$. Give a proof of the asymptotic distribution of Λ_n under the null-hypothesis H_0 (necessary regularity conditions need not be explicitly stated).

Define further the score statistic and the score test. How would you choose critical values for the rejection regions of the score test?

4

Let $X_1, \ldots, X_n$ be independent random variables taking values in $[-1, 1]$ satisfying $EX_i = 0$ for every $i = 1, \ldots, n$. Prove that for every $u > 0$ and $n \in \mathbb{N}$ the inequality

$$\Pr\left\{\sum_{i=1}^{n} X_i > u\right\} \leqslant \exp\left(-\frac{u^2}{2n}\right) \tag{1}$$

holds true.

Without using the central limit theorem, prove that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ is stochastically bounded. [A sequence W_n of random variables is stochastically bounded if for every $\epsilon > 0$ there exists a finite constant $M(\epsilon)$ such that $P(|W_n| > M(\epsilon)) < \epsilon$.]

If $Z_1, \ldots, Z_n$ are independent and identically distributed random variables taking values in $[-\frac{1}{2}, \frac{1}{2}]$, construct a conservative level $1 - \alpha$-confidence interval C_n for the mean EZ such that the diameter $|C_n|$ of C_n has diameter of order $M(\alpha)/\sqrt{n}$, where $M(\alpha)$ grows of the order $\sqrt{-2 \log \alpha}$ as α approaches zero. [A conservative confidence interval is a random interval C_n such that $\Pr(EZ \in C_n) \geqslant 1 - \alpha$ for every $n \in \mathbb{N}$.]

5

Define the concept of a compactly supported scaling function $\phi : \mathbb{R} \to \mathbb{R}$ and the associated compactly supported wavelet $\psi : \mathbb{R} \to \mathbb{R}$. Define carefully the concept of the corresponding wavelet basis of the space $L^2(\mathbb{R})$ of square-integrable real-valued functions defined on $\mathbb{R}$.

Prove that if $f : \mathbb{R} \to \mathbb{R}$ is S times differentiable with bounded S-th derivative on $\mathbb{R}$, if ϕ is S-times differentiable, and if

$$K_j(f) = 2^j \int_{\mathbb{R}} K(2^j x, 2^j y) f(y) dy, \; K(x,y) = \sum_{k \in \mathbb{Z}} \phi(x-k)\phi(y-k),$$

then, for every $x \in \mathbb{R}$,

$$|K_j(f)(x) - f(x)| \leqslant C 2^{-jS}$$

where C is some constant independent of j. [You may use the fact that $\int_{\mathbb{R}} K(x, x-u) u^{\alpha} du$ equals 1 if $\alpha = 0$ and 0 if $0 < \alpha \leqslant S$.]

6

Define the fixed and random design nonparametric regression model. Define the Nadaraya-Watson estimator.

Define the concept of a cubic spline on $[0,1]$ with breakpoints $0 < x_1 < x_2 < \cdots < x_n < 1$. A cubic spline g is called natural if $D^2 g(0) = D^2 g(1) = D^3 g(0) = D^3 g(1) = 0$. Let $Y_1, \ldots, Y_n$ be independent random variables. Let m be any minimizer of $Q(m) = \sum_{i=1}^n (Y_i - m(x_i))^2 + \lambda J(m)$ over the set of twice differentiable functions m defined on $[0,1]$, where $\lambda > 0$ and $J(m) = \int_0^1 (D^2 m(x))^2 dx$. Discuss briefly the heuristic behind this penalized nonparametric regression estimator. Show that m can be taken to be a natural cubic spline . [You may use the fact that for any set of numbers $z_1, ..., z_n$ we can find a unique natural cubic spline g such that $g(x_i) = z_i$.]

END OF PAPER