UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Thursday, 7 June, 2012 9:00 am to 12:00 pm

PAPER 39

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

For $\mathcal{H}$ a class of real-valued functions defined on $\mathbb{R}^d$, state a uniform in $\mathcal{H}$ law of large numbers under a bracketing covering assumption on $\mathcal{H}$. [You do not have to prove the result.]

Suppose now Θ is a bounded and closed subset of $\mathbb{R}^p$, and let $q(\theta, x) : \Theta \times \mathbb{R}^d \to \mathbb{R}$ be continuous in θ for each x and measurable in x for each θ. If $X, X_1, ..., X_n$ are i.i.d. random vectors in $\mathbb{R}^d$ and if

$$E \sup_{\theta \in \Theta} |q(\theta, X)| < \infty \tag{1}$$

prove that, as $n \to \infty$,

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} q(\theta, X_i) - E q(\theta, X) \right| \to 0 \quad \text{almost surely}$$

[*Hint: You may use the bracketing uniform law of large numbers without proof. You may also use without proof the Heine-Borel and the dominated convergence theorems in your justification.*]

Considering $q(\theta, x) = \log f(\theta, x)$ and an exponential family of order 1,

$$f(\theta, x) = e^{\theta x - K(\theta)} f_0(x), \theta \in \Theta, \quad x \in \mathbb{R},$$

with f_0 a fixed probability density and K the cumulant generating function, devise weak assumptions on K, f_0 such that the domination condition (1) is satisfied.

2

Let $\mathcal{P}_\Theta = \{f(\theta, \cdot) : \theta \in \Theta\}$ be a parametric statistical model of probability density functions $f(\theta, y) : \Theta \times \mathbb{R} \to [0, \infty)$ indexed by an open subset Θ of $\mathbb{R}^p$. Define what it means for $\mathcal{P}_\Theta$ to be locally asymptotically normal (LAN). Informally discuss conditions on the first and second derivatives of the mapping $\theta \mapsto \log f(\theta, y)$ that ensure that $\mathcal{P}_\Theta$ is LAN, and sketch a proof of this argument.

For two sequences of probability measures P_n, Q_n on a measurable space, define the concept of mutual contiguity of P_n and Q_n. For $h \in \mathbb{R}^p$ let now $P^n_{\theta + h/\sqrt{n}}$ and P^n_θ be the product probability measures describing the joint distribution of i.i.d. samples $Y_1, \ldots, Y_n$ drawn from $P_{\theta + h/\sqrt{n}}$ and P_θ in $\mathcal{P}_\Theta$, respectively. Prove that if $\mathcal{P}_\Theta$ is LAN then $P_{\theta + h/\sqrt{n}}$ and P_θ are mutually contiguous for every $\theta \in \Theta$. Deduce that if an estimator $\hat{\theta}_n = \hat{\theta}(Y_1, \ldots, Y_n)$ is consistent for θ under P_θ then it is also consistent for θ under $P_{\theta + h/\sqrt{n}}$. [You may use without proof Le Cam's first lemma, provided it is clearly stated.]

3

Consider $X_1, \ldots, X_n$ i.i.d. random variables drawn from an unknown probability density f. Define the kernel density estimator $f_n^K(h)$ for f with kernel K and bandwidth h. Assuming that K is compactly supported and symmetric and that f is square integrable and twice differentiable with $\|D^2 f\|_2^2 = \int_{\mathbb{R}} (D^2 f(x))^2 dx < \infty$, carefully prove that the mean-integrated squared error (MISE) $E \int_{\mathbb{R}} (f_n^K(h,x) - f(x))^2 dx$ can be bounded by

$$\frac{1}{nh} \int_{\mathbb{R}} K^2(u) du + (1/3) h^4 \|D^2 f\|_2^2 \left(\int_{\mathbb{R}} u^2 K(u) du \right)^2 .$$

[You may use without proof Fubini's and Taylor's theorem in your justification.]

Give the rate of convergence in MISE obtained from optimising this risk bound. Assuming that f is three times continuously differentiable with bounded derivatives, discuss a heuristic to estimate $\|D^2 f\|_2^2$ from the sample. Can you expect that your estimate attains a rate of estimation for $\|D^2 f\|_2^2$ of order $1/\sqrt{n}$?

4

Suppose we are given a model $\mathcal{P}$ of probability densities, and a random sample $X_1, \ldots, X_n, n > 2$, from $f \in \mathcal{P}$, where $\mathcal{P}$ is equipped with some metric d. Let $P_f \equiv P_f^n$ be the joint distribution of $X_1, \ldots, X_n$, and denote by E_f expectation with respect to P_f. Consider *any* estimator $f_n(x) = f(x; X_1, \ldots, X_n)$ for $f(x)$. For r_n a sequence of positive real numbers and f_0, f_1 two probability densities in $\mathcal{P}$ such that $d(f_0, f_1) \geqslant 2r_n$ for every n, prove that for every $\eta > 0$ and every $n \in \mathbb{N}$

$$\inf_{f_n} \sup_{f \in \mathcal{P}} r_n^{-1} E_f d(f_n, f) \geqslant \frac{1-\eta}{2} \left(1 - \frac{E_{f_0}|Z-1|}{\eta} \right)$$

where Z is the likelihood ratio

$$\frac{dP_{f_1}}{dP_{f_0}} = \prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)}.$$

Now consider

$$\Psi = \{\psi : [0,1] \to \mathbb{R}, \sup_{x \in [0,1]} |\psi(x)| \leqslant 1, \int_0^1 \psi = 0\}$$

and consider the family of probability densities $\mathcal{P}_\Psi = \{f = 1 + \psi : \psi \in \Psi\}$ on $[0,1]$ equipped with the L^2-metric $d^2(f,g) = \int_0^1 (f(x) - g(x))^2 dx$. Prove that for every $n \in \mathbb{N}$

$$\inf_{f_n} \sup_{f \in \mathcal{P}_\Psi} \sqrt{n} E_f d(f_n, f) \geqslant c > 0$$

for some $c > 0$.

5

Define the fixed and random design regression models. Carefully define the Nadaraya-Watson estimator and the local polynomial regression estimator. What is the relationship between these two estimators?

Let now $\hat{m}_n(x)$ be the local polynomial regression estimator of order ℓ of a regression function $m : [0,1] \to \mathbb{R}$, based on a bounded compactly supported kernel K and a bandwidth h. Suppose m has s bounded derivatives and that $nh \geqslant 1$ for all n. Prove that for some constant $c > 0$,

$$E|\hat{m}_n(x) - m(x)| \leqslant c[(nh)^{-1/2} + h^s]$$

if one has a sample $(Y_1, x_1), \ldots, (Y_n, x_n)$ with equally spaced fixed design points $x_i = i/n, i = 1, \ldots, n$.

[*Hint: You may use the fact that* $\hat{m}_n(x) = \sum_{i=1}^n W_{ni}(x) Y_i$ *for*

$$W_{ni}(x) = \frac{1}{nh} U^T(0) B^{-1} U\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right),$$

where $B \equiv B_n$ *is an invertible matrix with eigenvalues bounded away from zero uniformly in* n*, and where* $U^T(t) = (1, t, t^2/2!, \ldots, t^\ell/\ell!)$*. You may further use without proof that*

$$\sum_{i=1}^n Q(x_i) W_{ni}(x) = Q(x)$$

for any polynomial Q *of degree less than or equal to* ℓ *and every* $x \in [0,1]$*.*]

6

Let $S_D \subset S$ where S is a normed space with norm $\|\cdot\|_S$ and let Φ be a real-valued mapping defined on S_D. Define the notion of Hadamard-differentiability of Φ at a point $s_0 \in S_D$. Let $r_n \to \infty$ as $n \to \infty$, and let X_n be random variables taking values in S such that $r_n(X_n - s_0)$ converges in distribution to some random variable X in S as $n \to \infty$. Derive the asymptotic distribution of $r_n(\Phi(X_n) - \Phi(s_0))$ as $n \to \infty$. [You may use without proof Skorohod's almost sure representation theorem in metric spaces in the proof, provided it is carefully stated.]

Let now $S = C(\mathbb{R})$ be the space of bounded continuous functions on $\mathbb{R}$, equipped with the supremum norm. Prove that for every $u > 0$ the mapping

$$(F, G) \mapsto \int_0^u G(x)DF(x)dx$$

is Hadamard-differentiable on $S_D \times S_D \subset C(\mathbb{R}) \times C(\mathbb{R})$ where

$$S_D = \{F \in C^1(\mathbb{R}) : \int |F'(x)|dx \leqslant 1\}$$

and where $C^1(\mathbb{R})$ is the space of bounded continuously differentiable functions on $\mathbb{R}$ with a bounded derivative.

END OF PAPER