M. Phil. in STATISTICAL SCIENCE

Monday, 7 June, 2010   9:00 am to 11:00 am

NONPARAMETRIC STATISTICAL THEORY

*Attempt no more than **THREE** questions.*

*There are **FOUR** 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

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with distribution function $F : \mathbb{R} \to [0, 1]$. Define the empirical distribution function $F_n : \mathbb{R} \to [0, 1]$. Define the (standard) Brownian bridge $\mathbb{G}$ and the F-Brownian bridge process $\mathbb{G}_F$. Carefully state Donsker's central limit theorem for $F_n - F$.

Assuming that F is continuous, use Donsker's central limit theorem to prove that

$$\sqrt{n} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow{d} \max_{t \in [0,1]} |\mathbb{G}(t)|$$

$n \to \infty$. [You may use the continuous mapping theorem for convergence in distribution in metric spaces.]

Explain how this result can be used to construct an asymptotic confidence band for F centered at F_n.

2

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with probability density function $f : \mathbb{R} \to [0, \infty)$. Define the kernel density estimator $f_n^K(x, h)$ of f with bandwidth h.

What is a kernel of order l? Let $\{\phi_m\}_{m \in \mathbb{N}}$ be the orthonormal basis of Legendre polynomials defined by

$$\phi_0(x) := 2^{-1/2}, \qquad \phi_m(x) = \sqrt{\frac{2m + 1}{2}} \frac{1}{2^m \, m!} \frac{d^m}{dx^m} [(x^2 - 1)^m]$$

for $x \in [-1, 1]$ and $m \in \mathbb{N}$. Using these polynomials (or otherwise), construct a kernel of order l. [You may use standard properties of spaces of polynomials in your answer.]

Suppose f is three times differentiable and that f and $D^3 f$ are bounded functions. Devise a kernel K and a bandwidth h_n depending on n such that for every $x \in \mathbb{R}$

$$E \left| f_n^K(x, h) - f(x) \right| \leqslant C n^{-\frac{3}{7}}$$

for some constant C independent of n. Give an example of a probability density function $f : \mathbb{R} \to [0, \infty)$ for which

$$E \left| f_n^K(x, h) - f(x) \right| \leqslant C n^{-\frac{1}{2}}$$

for every $x \in \mathbb{R}$.

Nonparametric Statistical Theory

3

If $m : \mathbb{R} \to \mathbb{R}$ is a function for which $\int_{\mathbb{R}} (m(x))^2 \, dx < \infty$, and if ϕ and ψ are the generating functions of a wavelet basis, define carefully the wavelet series and the wavelet projection of m.

Considering random variables Y_i, design points x_i and a regression function m, define the *fixed design regression model*. For design points $x_i = i/n$, define the wavelet regression estimator of m. By analogy to the wavelet thresholding density estimator, argue how one could implement a *thresholded* wavelet regression estimator, and motivate a reasonable choice of thresholds.

4

Let S be a normed space with norm $\| \cdot \|_S$, and let Φ be a real-valued mapping defined on S. Define the notion of Hadamard-differentiability of Φ at a point $s_0 \in S$. Let r_n be real numbers diverging to infinity 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 taking values 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 Skorohod's almost sure representation theorem in the proof, provided it is carefully stated.]

Suppose now $\Phi : L^\infty \to \mathbb{R}$ is Hadamard differentiable on the space L^∞ of bounded real-valued functions defined on $\mathbb{R}$ equipped with the supremum norm $\|f\|_\infty = \sup_{t \in \mathbb{R}} |f(t)|$. Suppose you are given a random sample from some unknown distribution function $F : \mathbb{R} \to [0, 1]$. Construct an estimator T_n for $\Phi(F)$ such that $\sqrt{n}(T_n - \Phi(F))$ is stochastically bounded, that is, for every $\varepsilon > 0$ there exists $n_0 := n_0(\varepsilon) \in \mathbb{N}$ and $M := M(\varepsilon)$ finite such that $\Pr(\sqrt{n}(T_n - \Phi(F)) \leqslant M) > 1 - \varepsilon$ for every $n \geqslant n_0$. [You may use Donsker's central limit theorem in the answer.]

END OF PAPER

Nonparametric Statistical Theory