UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Wednesday, 3 June, 2009 1:30 pm to 3:30 pm

PAPER 33

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 Suppose $X_1, ..., X_n$ are independent and identically distributed random variables with cumulative distribution function $F : \mathbb{R} \to \mathbb{R}$. Define the *empirical distribution function* F_n of the sample.

Given two (measurable) real-valued functions l, u on $\mathbb{R}$, a *bracket* is the set of functions $[l, u] := \{f : \mathbb{R} \to \mathbb{R} : l(x) \leqslant f(x) \leqslant u(x) \text{ for all } x \in \mathbb{R}\}$. Suppose $\mathcal{H}$ is a class of measurable functions from $\mathbb{R}$ to $\mathbb{R}$ such that, for every $\varepsilon > 0$, there exist $N(\varepsilon) < \infty$ brackets $[l_i, u_i]_{i=1}^{N(\varepsilon)}$ that satisfy the following conditions: i) for every i, $E|l_i(X)| < \infty$, $E|u_i(X)| < \infty$, $E|u_i(X) - l_i(X)| < \varepsilon$, and ii) for every $h \in \mathcal{H}$ there exists i with $h \in [l_i, u_i]$. Prove the *uniform law of large numbers*

$$\sup_{h\in\mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} (h(X_i) - Eh(X)) \right| \to 0 \quad \textit{almost surely}$$

as $n \to \infty$.

Deduce from the above result that

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \to 0 \quad \textit{almost surely}$$

as $n \to \infty$.

Furthermore, give an example of a class $\mathcal{H}$ of continuous functions $h : \mathbb{R} \to \mathbb{R}$ with uncountably many elements for which the uniform law of large numbers holds. [You may use results from functional analysis, such as the Ascoli-Arzela theorem, in the justification.]

2 Given an independent and identically distributed sample $X_1, ..., X_n$ from the probability density function $f : \mathbb{R} \to \mathbb{R}$, define, for $x \in \mathbb{R}$, the kernel density estimator $f_n^K(x, h)$ with bandwidth $h > 0$ and kernel K. Discuss briefly a motivation for this estimator.

Suppose that f is differentiable on $\mathbb{R}$ with bounded derivative and that $h = h_n$ satisfies $nh_n^3 \to 0$ as $n \to \infty$. Assume that the kernel $K : \mathbb{R} \to \mathbb{R}$ is a nonnegative, bounded and compactly supported function. Prove that, for every $x \in \mathbb{R}$,

$$\sqrt{nh_n}(f_n^K(x, h_n) - f(x)) \to^d N(0, f(x)\|K\|_2^2)$$

as $n \to \infty$, where $\|K\|_2^2 = \int_{\mathbb{R}} K^2(x)dx$. [You may assume the Lindeberg-Feller central limit theorem, provided it is carefully stated.]

Suppose you are given the quantiles of the $N(0, f(x)\|K\|_2^2)$ distribution. Describe how to construct a confidence interval for $f(x)$ of asymptotic coverage $1 - \alpha$, based on the above limit theorem.

3 What is the *wavelet series* of a square-integrable function $f : \mathbb{R} \to \mathbb{R}$? How can it be used to approximate the function?

Considering the Haar wavelet, denote by $K_j(f)$ the projection (with respect to the inner product $< f, g >= \int_{\mathbb{R}} f(x)g(x)dx$) of a locally integrable function $f : \mathbb{R} \to \mathbb{R}$ onto the space V_j of functions that are piecewise constant on the intervals $(k/2^j, (k+1)/2^j]$, $k \in \mathbb{Z}$. Prove that, if $f : \mathbb{R} \to \mathbb{R}$ is bounded and differentiable with a bounded derivative, then there exists a constant c independent of j and x such that $|K_j(f)(x) - f(x)| \leqslant c2^{-j}$ for every $x \in \mathbb{R}$.

Suppose you are given a sample of independent and identically distributed random variables $X_1, ..., X_n$ with common probability density function $f : \mathbb{R} \to \mathbb{R}$, where f is differentiable with bounded derivative. How can you use wavelets to estimate f? Show that one can construct a density estimator $f_n^W(x)$ based on Haar wavelets such that the pointwise risk satisfies $E|f_n^W(x) - f(x)| = O(n^{-1/3})$.

4 Suppose you are given n independent and identically distributed copies of the random vector (X, Y) with joint probability density function $f(x, y)$, marginal density for X given by f^X and suppose $m(x) = E(Y|X = x)$. Define, for $x \in \mathbb{R}$, the *Nadaraya-Watson estimator* $\hat{m}_n(h, x)$ based on the kernel K and bandwidth h.

Again, let $x \in \mathbb{R}$ and suppose $m(x)$ is bounded and twice continuously differentiable at x, that the conditional variance function $V(x) = \mathrm{Var}(Y|X = x)$ is bounded on $\mathbb{R}$ and continuous at x, and that f^X is bounded, continuous on $\mathbb{R}$, continuously differentiable at x, and satisfies $f^X(x) > 0$. Suppose further that the kernel is $K(x) = 1_{[-1/2,1/2]}(x)$. If $h = h_n \simeq n^{-1/5}$, prove that

$$E|\hat{m}_n(h_n, x) - m(x)| = O(n^{-2/5})$$

as $n \to \infty$. [You may use in the proof the auxiliary result that

$$E(|\hat{m}_n(h_n, x) - m(x)|1\{\hat{f}_n^X(x) \leqslant \delta\}) = o(n^{-2/5})$$

for some $\delta > 0$, where $\hat{f}_n^X(x) = (nh_n)^{-1} \sum_{i=1}^n K((x - X_i)/h_n)$. You may use further that the ordinary kernel density estimator satisfies $E(f^X(x) - \hat{f}_n^X(x))^2 = o(1)$ as $n \to \infty$.]

END OF PAPER