UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 9 June, 2014 9:00 am to 11:00 am

PAPER 34

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 F be a distribution function on $\mathbb{R}$. Define the *quantile function.* Prove that the quantile function is non-decreasing and left-continuous.

Let $U_1, \ldots, U_n$ be independent and identically distributed $U(0,1)$ random variables, and let $U_{(1)} < \ldots < U_{(n)}$ be the corresponding order statistics. For a fixed $j \in \{1, \ldots, n\}$, compute the density of $U_{(j)}$.

Now let $X_1, \ldots, X_n$ be independent and identically distributed with continuous distribution function F, and let $X_{(1)} < \ldots < X_{(n)}$ be the corresponding order statistics. For $j < n/2$, show that the interval $(X_{(j)}, X_{(n-j+1)}]$ is a $(1-\alpha)$-level confidence interval for the median, $F^{-1}(1/2)$, where

$$\alpha = 2n\binom{n-1}{j-1}\int_0^{1/2} x^{n-j}(1-x)^{j-1}\,dx.$$

[*Hint: You may use the fact that if X has continuous distribution function F, then $F(X) \sim U(0,1)$.*]

2

Let f be a density on $\mathbb{R}$. Define what is meant by a *kernel density estimator* $\hat{f}_h$ of f. Define the Mean Integrated Squared Error of $\hat{f}_h$, denoted $MISE(\hat{f}_h)$.

Let $X_1, \ldots, X_n$ be independent Exp(1) random variables, and consider the kernel density estimator $\hat{f}_h$ with kernel $K(x) = \mathbb{1}_{\{|x| \leqslant 1/2\}}$. Show that

$$\mathbb{E}\{\hat{f}_h(x)\} = \frac{1}{h}\big[\exp\{-\max(x - h/2, 0)\} - \exp\{-\max(x + h/2, 0)\}\big].$$

Deduce that there exist $h_0 > 0$ and $c > 0$ such that for all $h \in (0, h_0]$, we have

$$\int_{-\infty}^{\infty} \mathrm{Bias}^2\{\hat{f}_h(x)\}\,dx \geqslant ch.$$

Show further that

$$\int_{-\infty}^{\infty} \mathrm{Var}\{\hat{f}_h(x)\}\,dx = \frac{1}{nh} - \frac{1}{n}\int_{-\infty}^{\infty} \{\mathbb{E}\hat{f}_h(x)\}^2\,dx.$$

You are given that if $h \to 0$ as $n \to \infty$, then

$$\int_{-\infty}^{\infty} \mathrm{Bias}^2\{\hat{f}_h(x)\}\,dx = \frac{h}{12} + o(h) \quad \text{and} \quad \int_{-\infty}^{\infty} \{\mathbb{E}\hat{f}_h(x)\}^2\,dx = O(1).$$

Show that

$$n^{1/2} \inf_{h>0} MISE(\hat{f}_h) \to \frac{1}{\sqrt{3}}$$

as $n \to \infty$.

3

Consider the fixed design nonparametric regression model

$$Y_i = m(i/n) + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ_i are independent and identically distributed with $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = 1$. Define the *local polynomial* estimator $\hat{m}_h(\cdot; p)$ of m of degree p, and with kernel K and bandwidth h. Give an explicit formula for the local constant estimator $\hat{m}_h(\cdot; 0)$.

Fix $x_0 \in (0,1)$. Assume that m is differentiable at x_0, that K is bounded, continuous, symmetric and supported on $[-1,1]$, and that $h \to 0$ but $nh \to \infty$ as $n \to \infty$. You are also given that for $x_0 \in (0,1)$ and $r \geqslant 0$,

$$s_{r,h}(x_0) := \frac{1}{n}\sum_{i=1}^{n}(i/n - x_0)^r K_h(i/n - x_0) = h^r \mu_r(K) + o(h^r)$$

$$t_{r,h}(x_0) := \frac{1}{n}\sum_{i=1}^{n}(i/n - x_0)^r K_h^2(i/n - x_0) = h^{r-1}\mu_r(K^2) + o(h^{r-1}),$$

as $n \to \infty$, where $\mu_r(L) := \int_{-\infty}^{\infty} x^r L(x)\, dx$. Writing MSE for the mean squared error, prove that

$$MSE\,\{\hat{m}_h(x_0; 0)\} = \frac{R(K)}{nh} + o\left(h^2 + \frac{1}{nh}\right),$$

where $R(K) := \int_{-\infty}^{\infty} K^2(x)\, dx$. By considering $h_\epsilon = \epsilon^{-1/3} n^{-1/3}$, where $\epsilon > 0$ is arbitrary, or otherwise, deduce that

$$n^{2/3} \inf_{h>0} MSE\{\hat{m}_h(x_0; 0)\} \to 0.$$

Now let $\mathcal{F}$ denote the class of real-valued functions on $[0,1]$ that are differentiable at x_0. Show that there exists $c > 0$ such that for any estimator $\tilde{m}$ of m and any $x_0 \in [0,1]$, we have

$$\sup_{m \in \mathcal{F}} \mathbb{E}\{\tilde{m}(x_0) - m(x_0)\}^2 \geqslant c n^{-2/3}$$

for sufficiently large $n \in \mathbb{N}$.

[You may use Le Cam's two-point lemma without proof.]

4 Let F be a distribution function on $\mathbb{R}$. What does it mean for F to be *non-degenerate*? In the context of extreme value theory, define what it means for F to be *max-stable*, and for F to belong to the *domain of attraction* of a distribution function G.

State the extremal types (Fisher–Tippet–Gnedenko) theorem. In each of the following three cases, determine the domain of attraction to which the distribution function corresponding to the given density belongs: (i) $f_1(x) := 2(1-x)\mathbb{1}_{\{x\in(0,1)\}}$, (ii) $f_2(x) := \lambda e^{-\lambda x}\mathbb{1}_{\{x>0\}}$, (iii) $f_3(x) := \lambda/(1+x)^{\lambda+1}\mathbb{1}_{\{x>0\}}$, where $\lambda > 0$.

[Sufficient conditions to belong to a domain of attraction should be stated, but need not be proved.]

Let $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ denote the order statistics of a random sample from a distribution function F. Suppose that

$$\lim_{t\to\infty}\int_t^\infty \frac{1-F(x)}{1-F(t)}\frac{dx}{x} = \frac{1}{\alpha}, \tag{1}$$

which implies that F belongs to the domain of attraction of a Fréchet distribution with shape parameter $\alpha > 0$. For some fixed $k \in \mathbb{N}$, Hill's estimator of $1/\alpha$ is defined by

$$\hat{\gamma}_H := \frac{1}{k}\sum_{j=1}^{k}\log\left(\frac{X_{(n-j+1)}}{X_{(n-k)}}\right).$$

Show that Hill's estimator is obtained by substituting the empirical distribution function for F in (1) and replacing the limit in t with the fixed choice $t = X_{(n-k)}$.

END OF PAPER