UNIVERSITY OF CAMBRIDGE

MAMA/210, NST3AS/210, MAAS/210

MAT3

MATHEMATICAL TRIPOS Part III

Monday, 12 June, 2023 1:30 pm to 3:30 pm

PAPER 210

TOPICS IN STATISTICAL THEORY

Before you begin please read these instructions carefully

Candidates have TWO HOURS to complete the written examination.

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
Rough 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 Given a distribution function F on $\mathbb{R}$, define the *quantile function* $F^{-1} : (0,1] \to (-\infty, \infty]$.

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$. Define the *empirical distribution function* $\mathbb{F}_n$ and, for $j \in [n]$, define the jth *order statistic* $X_{(j)}$. Express $X_{(j)}$ in terms of $\mathbb{F}_n^{-1}$, evaluated at an appropriate point.

State and prove Bennett's inequality.

Let $U_1, \ldots, U_n \overset{\text{iid}}{\sim} U[0,1]$. For $j \in [n]$, state the distribution of the jth order statistic $U_{(j)}$, as well as $\mathbb{E}(U_{(j)})$. Prove that

$$\mathbb{P}\biggl(U_{(j)} - \frac{j}{n+1} \leqslant -x\biggr) \leqslant \Bigl(\frac{enp}{j}\Bigr)^j$$

for every $x \in \bigl[0, \frac{j}{n+1}\bigr)$, where $p := \frac{j}{n+1} - x \in \bigl(0, \frac{j}{n+1}\bigr]$.

2 For $\beta, L > 0$, define the Hölder class $\mathcal{F}(\beta, L)$ of densities on $\mathbb{R}$. In the context of kernel density estimation, define what is meant by a *kernel*, and define the *order* of a kernel.

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f \in \mathcal{F}(\beta, L)$ and let $\hat{\mathcal{F}}_n$ denote the set of Borel measurable functions from $\mathbb{R} \times \mathbb{R}^n$ to $\mathbb{R}$. Prove that there exists $C > 0$, depending only on β, such that

$$\inf_{\hat{f}_n \in \hat{\mathcal{F}}_n} \sup_{x \in \mathbb{R}} \sup_{f \in \mathcal{F}(\beta,L)} \mathbb{E}_f\bigl[\bigl\{\hat{f}_n(x; X_1, \ldots, X_n) - f(x)\bigr\}^2\bigr] \leqslant C L^{2/(\beta+1)} n^{-2\beta/(2\beta+1)}.$$

[You may assume the existence of a bounded kernel K of arbitrarily large order satisfying $\int_{-\infty}^{\infty} |u|^\beta |K(u)| \, du < \infty$.]

3 Consider a vector $Y = (Y_1, \ldots, Y_n)^\top$ of responses from the nonparametric regression model

$$Y_i = m(x_i) + \epsilon_i,$$

where $x_i = i/n$ for $i \in [n]$, where $m : [0,1] \to \mathbb{R}$ and where $\epsilon_1, \ldots, \epsilon_n$ are independent with $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) \leqslant \sigma^2$ for $i \in [n]$. Fix $x \in (0,1)$, let K denote a bounded kernel that vanishes outside $[-1,1]$, let $p \in \mathbb{N}_0$ and let $h > 0$. Show that, for suitable matrices $X \in \mathbb{R}^{n \times (p+1)}$ and $W \in \mathbb{R}^{n \times n}$, and subject to a positive definiteness condition that you should state and then assume throughout, the local polynomial estimator of $m(x)$ of degree p, bandwidth h and kernel K can be expressed as an appropriate component of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^\top$, where $\hat{\beta} = (X^\top W X)^{-1} X^\top W Y$.

Now suppose that m is differentiable and consider $\hat{m}_n'(x) := \hat{\beta}_1/h$ as an estimator of $m'(x)$. Writing $\hat{m}_n'(x) = n^{-1} \sum_{i=1}^n v_{p,i}(x) Y_i$, prove that whenever R is a polynomial of degree at most p, we have

$$\frac{1}{n} \sum_{i=1}^n v_{p,i}(x) R(x_i) = R'(x).$$

Prove further that, for a suitable $\lambda_0 > 0$,

$$\max_{i \in [n]} \frac{1}{n} |v_{p,i}(x)| \leqslant \frac{2\|K\|_\infty}{\lambda_0 n h^2}$$

and

$$\frac{1}{n} \sum_{i=1}^n |v_{p,i}(x)| \leqslant \frac{2\|K\|_\infty}{\lambda_0 n h^2} \sum_{i=1}^n \mathbb{1}_{\{|x_i - x| \leqslant h\}}.$$

Finally, assume that m belongs to the Hölder class $\mathcal{H}(\beta, L)$ for some $\beta > 1$ and $L > 0$. Prove that when $p \geqslant \lceil \beta \rceil - 1$ and $h \geqslant 1/(2n)$, we have

$$\mathrm{Var}\, \hat{m}_n'(x) \leqslant \frac{16\|K\|_\infty^2 \sigma^2}{\lambda_0^2 n h^3} \quad \text{and} \quad \left|\mathrm{Bias}\, \hat{m}_n'(x)\right| \leqslant \frac{8L\|K\|_\infty}{\lambda_0 \beta_0!} h^{\beta - 1}.$$

4 Let P and Q denote probability measures on a common measurable space. Define the *total variation distance* $\mathrm{TV}(P,Q)$ and the *Kullback–Leibler divergence* $\mathrm{KL}(P,Q)$.

For $\mu \in \mathbb{R}$, write Laplace(μ) for the Laplace distribution with mean μ, having density $x \mapsto e^{-|x-\mu|}/2$ with respect to Lebesgue measure on $\mathbb{R}$. Prove that if $P = \text{Laplace}(0)$ and $Q = \text{Laplace}(\mu)$, then $\mathrm{KL}(P,Q) = e^{-|\mu|} - 1 + |\mu|$.

State and prove Assouad's lemma.

For $n \in \mathbb{N}$, let $\mathcal{M}_n := \{\theta = (\theta_1, \ldots, \theta_n)^\top \in \mathbb{R}^n : \theta_i \leqslant \theta_j \text{ for } i < j\}$. For $\theta = (\theta_1, \ldots, \theta_n)^\top \in \mathcal{M}_n \cap [0,1]^n$, consider the isotonic regression model $Y_i = \theta_i + \epsilon_i$ for $i \in [n]$ and $n \geqslant 2$, where $\epsilon_1, \ldots, \epsilon_n \overset{\text{iid}}{\sim} \text{Laplace}(0)$. Writing $\hat{\Theta}$ for the set of Borel measurable functions from $\mathbb{R}^n$ to $\mathbb{R}^n$, prove that there exists a universal constant $c > 0$ such that

$$\inf_{\hat{\theta} \in \hat{\Theta}} \sup_{\theta \in \mathcal{M}_n \cap [0,1]^n} \frac{1}{n} \mathbb{E}_\theta\bigl(\|\hat{\theta}(Y_1, \ldots, Y_n) - \theta\|^2\bigr) \geqslant c \cdot n^{-2/3}.$$

[Pinsker's inequality may be used without proof.]

END OF PAPER