UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Wednesday, 2 June, 2021 12:00 pm to 2:00 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 Define what is meant by saying that a mean-zero random variable X is *sub-Gaussian* with parameter $\sigma^2 > 0$. Prove that such a random variable X satisfies

$$\max\big\{\mathbb{P}(X \geqslant x), \mathbb{P}(X \leqslant -x)\big\} \leqslant e^{-x^2/(2\sigma^2)}$$

for all $x \geqslant 0$. Prove further that it also satisfies

$$\mathrm{Var}(X) \leqslant \sigma^2.$$

What is meant by saying that a mean-zero random variable X is *sub-Gamma* on the right tail with variance factor σ and scale factor c? State Bernstein's inequality.

Let $X_1, \ldots, X_n$ be independent, mean-zero random variables that are sub-Gaussian with parameter σ^2, and let $a_1, \ldots, a_n \in \mathbb{R}$. Let $a_+ := \big(\max(a_i, 0)\big)_{i=1}^n$ and $a_- := \big(\max(-a_i, 0)\big)_{i=1}^n$, and for $u = (u_1, \ldots, u_n)^\top \in \mathbb{R}^n$, let $\|u\|_2 := \big(\sum_{i=1}^n u_i^2\big)^{1/2}$ and $\|u\|_\infty := \max_{i=1,\ldots,n} |u_i|$. Prove that

$$\mathbb{P}\biggl(\frac{1}{n}\sum_{i=1}^n \big\{a_i X_i^2 - \mathbb{E}(a_i X_i^2)\big\} \geqslant x\biggr) \leqslant \exp\biggl(\frac{-n^2 x^2}{2(\sigma^4 v + cn\sigma^2 x)}\biggr)$$

for all $x \geqslant 0$, where $v := 16(\|a_+\|_2^2 + \|a_-\|_2^2)$ and $c := \max(2\|a_+\|_\infty, \|a_-\|_\infty)$.

2 Let $\mathcal{F}$ denote the set of all differentiable densities on $\mathbb{R}$, and let $X_1, \ldots, X_n$ be independent, real-valued random variables with density $f \in \mathcal{F}$. In the context of kernel density estimation, define what is meant by a *kernel* and a *kernel density estimator* $\hat{f}_n \equiv \hat{f}_{n,h,K}$ of f with bandwidth h and kernel K. Define what it means for a kernel to be of *order* $\ell \in \mathbb{N}$.

Let K be a continuously differentiable kernel that vanishes outside $[-1, 1]$, and consider $\hat{f}_n'$ as an estimator of f'. Derive a bound of the form

$$\int_{-\infty}^{\infty} \operatorname{Var} \hat{f}_n'(x)\, dx \leqslant \frac{1}{nh^{\alpha}} C_1(K),$$

where the universal constant $\alpha > 0$ and the function $C_1(K)$ of the kernel should be specified.

For $\beta, L > 0$, define the Nikol'ski class of functions $\mathcal{N}(\beta, L)$. Prove that if $f \in \mathcal{F}$ is such that $f' \in \mathcal{N}(\beta, L)$, and if K is of order $\ell := \lceil \beta \rceil$, then

$$\int_{-\infty}^{\infty} \operatorname{Bias}^2 \hat{f}_n'(x)\, dx \leqslant C_2(\beta, L, K) h^{\gamma},$$

where the universal constant $\gamma > 0$ and function $C_2(\beta, L, K)$ should be specified. [You may assume Taylor's theorem with an appropriate remainder term, as well as the generalised Minkowski inequality.]

Define the *Mean Integrated Squared Error* (MISE) of $\hat{f}_{n,h,K}'$ as an estimator of f', and determine $\delta > 0$, depending only on β, such that

$$\inf_{h>0} \sup_{f \in \mathcal{F}: f' \in \mathcal{N}(\beta, L)} \operatorname{MISE}(\hat{f}_{n,h,K}') \leqslant \frac{C_3(\beta, L, K)}{n^{\delta}}$$

for all $n \in \mathbb{N}$. [You need not specify $C_3(\beta, L, K)$.]

Comment briefly on the relative difficulty of estimating f and f' with respect to the Mean Integrated Squared Error criterion when they each belong to $\mathcal{N}(\beta, L)$.

3 Consider the fixed design nonparametric regression model

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

where $\epsilon_1, \ldots, \epsilon_n$ are independent, with $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = 1$ for $i = 1, \ldots, n$. Define the *local polynomial estimator* $\hat{m}_n(\cdot; p, h, K)$ of degree p with bandwidth h and kernel K. Under a positive definiteness condition that you should specify and then assume throughout, show that it is a linear function of $Y = (Y_1, \ldots, Y_n)^\top$.

Show that the local linear estimator $\hat{m}_n(\cdot; 1, h, K)$ can be expressed in the form

$$\hat{m}_n(x; 1, h, K) = \frac{1}{nh} \sum_{i=1}^{n} \frac{s_2(x) - s_1(x)(x_i - x)}{s_2(x)s_0(x) - s_1^2(x)} K\left(\frac{x_i - x}{h}\right) Y_i,$$

for functions $s_r(x) \equiv s_{r,h,K}(x; x_1, \ldots, x_n)$ with $r \in \{0, 1, \ldots\}$ that you should specify.

Now suppose that $m(x) = e^x$, that $x_i = i/n$ for $i = 1, \ldots, n$, and that $K(u) = \mathbb{1}_{\{|u| \leqslant 1\}}/2$. Writing $s_r := s_r(0)$, for $r \in \{0, 1, 2, 3\}$ derive bounds on $\left|s_r - \frac{h^r}{2(r+1)}\right|$, and hence show that there exists a universal constant $c > 0$ such that $s_2 s_0 - s_1^2 \geqslant ch^2$ for all $h > 0$ and $n \in \mathbb{N}$ with $nh \geqslant 32$. [You may use the facts that $\sum_{i=1}^{m} i^2 = m(m+1)(2m+1)/6$ and $\sum_{i=1}^{m} i^3 = m^2(m+1)^2/4$ for $m \in \mathbb{N}$.]

Hence or otherwise prove that there exists a universal constant $C > 0$ such that, provided $nh \geqslant 32$, we have

$$\left|\mathrm{Bias}\, \hat{m}_n(0; 1, h, K)\right| \leqslant Ch^2.$$

What is the corresponding upper bound for $\left|\mathrm{Bias}\, \hat{m}_n(0; 0, h, K)\right|$? [You may assume the form of the Nadaraya–Watson estimator.]

4 Define the *total variation distance* between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$. When P is absolutely continuous with respect to Q, define the Kullback–Leibler divergence from Q to P. State Pinsker's inequality. State and prove Le Cam's two point lemma.

For $\beta, L > 0$, define the Hölder class of functions $\mathcal{H}(\beta, L)$ on $[0,1]$. For $i = 1, \ldots, n$, let

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

where $x_i = i/n$ and where $\epsilon_1, \ldots, \epsilon_n$ are independent $N(0,1)$ random variables. Prove that there exists $c > 0$, depending only on β, such that

$$\inf_{x_0 \in [0,1]} \inf_{\hat{\theta}_n \in \hat{\Theta}} \sup_{m \in \mathcal{H}(\beta, L)} \mathbb{E}\big[\big\{\hat{\theta}_n(Y_1, \ldots, Y_n) - m(x_0)\big\}^2\big] \geqslant c \max\left\{\frac{1}{n}, \min\left(\frac{L^{2/(2\beta+1)}}{n^{2\beta/(2\beta+1)}}, 1\right)\right\},$$

where $\hat{\Theta}$ denotes the set of Borel measurable functions from $\mathbb{R}^n$ to $\mathbb{R}$. [You may assume the form of the Kullback–Leibler divergence between two normal distributions, and the fact that the function $K(u) := e^{-1/(1-u^2)} \mathbb{1}_{\{|u| \leqslant 1\}}$ is infinitely differentiable on $\mathbb{R}$ with $\|K^{(r)}\|_\infty < \infty$ for every non-negative integer r.]

END OF PAPER