UNIVERSITY OF CAMBRIDGE

MAMA/210, NST3AS/210, MAAS/210

MAT3

MATHEMATICAL TRIPOS Part III

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 Let X be a random variable with $\mathbb{E}(X) = 0$.

(a) Define what it is means for X to be *sub-Poisson in the right tail with variance parameter* σ^2. Define what it is means for X to be *sub-Gamma in the right tail with variance parameter* σ^2 *and scale parameter* c.

(b) Prove that if X is sub-Poisson in the right tail with variance parameter σ^2, then it is sub-Gamma in the right tail for appropriate variance and scale parameters that you should specify.

(c) Prove that if $\mathbb{E}(X^2) \leqslant \sigma^2$ and $\mathbb{E}(X_+^q) \leqslant q!\sigma^2 c^{q-2}/2$ for all integers $q \geqslant 3$, then X is sub-Gamma in the right tail with variance parameter σ^2 and scale parameter c.

[You may use the fact that $e^u - 1 - u \leqslant u^2/2$ for $u \leqslant 0$.]

(d) Define what it is means for Y to be *sub-Gaussian with variance parameter* σ^2. Prove that for every $q > 0$,

$$\mathbb{E}(|Y|^q) \leqslant 2 \cdot \Gamma\Big(\frac{q}{2} + 1\Big)(2\sigma^2)^{q/2}.$$

(e) Now let Y and Z be sub-Gaussian with variance parameters σ_Y^2 and σ_Z^2 respectively. Prove that $YZ - \mathbb{E}(YZ)$ is sub-Gamma in the right tail with variance parameter $64\sigma_Y^2\sigma_Z^2$ and scale parameter $4\sigma_Y\sigma_Z$. Prove further that if $\mathbb{E}(YZ) \geqslant 0$, then the variance parameter and scale parameter can be improved to $16\sigma_Y^2\sigma_Z^2$ and $2\sigma_Y\sigma_Z$ respectively.

[You may use the inequality $(a+b)^r \leqslant 2^{r-1}(a^r + b^r)$ for $a, b \geqslant 0$ and $r \geqslant 1$.]

2 Let f be a density on $\mathbb{R}$, and let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f$. Define what is meant by a *kernel* K, and given a bandwidth $h > 0$, define the *scaled kernel* K_h. Define a *kernel density estimator* $\hat{f}_n \equiv \hat{f}_{n,h,K}$ of f with bandwidth h and kernel K.

Let $h > 0$, define $f_h : \mathbb{R} \to [0, \infty]$ by $f_h(x) := (2h)^{-1} \int_{x-h}^{x+h} f$ and suppose that K vanishes outside $[-1, 1]$. Prove that

$$\mathbb{E} \int_{-\infty}^{\infty} \bigl|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\bigr|\, dx \leqslant \|K\|_\infty \int_{-\infty}^{\infty} \min\biggl\{\frac{2^{1/2} f_h^{1/2}}{(nh)^{1/2}}, 4f_h\biggr\}.$$

Prove further that for $\rho \in [0, 1/2]$ and $\delta > \rho/(1-\rho)$,

$$\begin{aligned}\mathbb{E} \int_{-\infty}^{\infty} \bigl|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)\bigr|\, dx &\leqslant \frac{2^{2-3\rho}\|K\|_\infty}{(nh)^\rho} \int_{-\infty}^{\infty} f_h^{1-\rho} \\ &\leqslant \frac{2^{2-3\rho}\|K\|_\infty C_{\rho,\delta}^\rho (1+h)^{\delta(1-\rho)}}{(nh)^\rho} \biggl(\int_{-\infty}^{\infty} (1+|x|)^\delta f(x)\, dx\biggr)^{1-\rho},\end{aligned}$$

where $C_{\rho,\delta} := \int_{-\infty}^{\infty} (1+|x|)^{-\delta(1-\rho)/\rho}\, dx = 2\rho/\bigl((1-\rho)\delta - \rho\bigr) \in (0, \infty)$.

[*Hint: You may use the fact that if $U \sim U[-1,1]$ and X_1 are independent, then $hU + X_1$ has density f_h.*]

3 Suppose that $n = n_1^d$ for some $n_1 \in \mathbb{N}$, and for $i = (i_1, \ldots, i_d)^\top \in [n_1]^d$, define $x_i = (x_{i1}, \ldots, x_{id})^\top \in \{1/n_1, 2/n_1, \ldots, 1\}^d$ by $x_{ij} := i_j/n_1$ for $j \in [d]$. Consider the nonparametric regression model

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

for $i \in [n_1]^d$, where $(\epsilon_i)_{i\in[n_1]^d}$ are independent with $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) \leqslant \sigma^2$ for all i.

For $\alpha = (\alpha_1, \ldots, \alpha_d)^\top \in \mathbb{N}_0^d$, we write $\alpha! := \prod_{j=1}^d \alpha_j!$ and $|\alpha| := \sum_{j=1}^d \alpha_j$, and if $\beta = (\beta_1, \ldots, \beta_d)^\top \in \mathbb{N}_0^d$, we say α *is smaller than* β *in the lexicographic order* if there exists $j^* \in [d]$ such that $\alpha_j = \beta_j$ for $j < j^*$ and $\alpha_{j^*} < \beta_{j^*}$. For $p \in \mathbb{N}_0$ and $p^* := \binom{d+p}{p}$, define $Q : \mathbb{R}^d \to \mathbb{R}^{p^*}$ by $Q(u) := \bigl(\frac{1}{\alpha!}\prod_{j=1}^d u_j^{\alpha_j}\bigr)_{|\alpha|\leqslant p}$, with components ordered in increasing lexicographic order in α, and for $h > 0$, let $Q_h(\cdot) := Q(\cdot/h)$.

Given a kernel K on $\mathbb{R}$, the *local polynomial estimator* of $m(\cdot)$, denoted $\hat{m}_n(\cdot) \equiv \hat{m}_n(\cdot; p, h, K)$, is defined at $x_0 = (x_{01}, \ldots, x_{0d})^\top \in [0,1]^d$ by $\hat{m}_n(x_0) := e_1^\top \hat{\beta}$, where

$$\hat{\beta} \in \operatorname*{argmin}_{\beta\in\mathbb{R}^{p^*}} \sum_{i\in[n_1]^d} \bigl\{Y_i - \beta^\top Q_h(x_i - x_0)\bigr\}^2 \prod_{j=1}^d K\Bigl(\frac{x_{ij} - x_{0j}}{h}\Bigr),$$

and where $e_1 \in \mathbb{R}^{p^*}$ denotes the first standard basis vector.

Define what is meant by a *linear estimator* of m, and, assuming the positiveness of the minimum eigenvalue λ_0 of a matrix that you should specify, show that $\hat{m}_n$ is a linear estimator of m.

For $\beta, L > 0$, and with $\beta_0 := \lceil \beta \rceil - 1$, suppose that m is β_0-times differentiable on $[0,1]^d$, and that (in multi-index notation for partial derivatives),

$$|m^{(\alpha)}(x) - m^{(\alpha)}(x')| \leqslant L\|x - x'\|_\infty^{\beta-\beta_0}$$

for all $x, x' \in [0,1]^d$ and $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = \beta_0$. Assume further that K vanishes outside $[-1,1]$ and that $p \geqslant \beta_0$. By first establishing a reproducing property and a bound on appropriately-defined effective kernel weights, prove that for $h \geqslant 1/(2n_1)$,

$$\mathrm{Var}\,\hat{m}_n(x_0; p, h, K) \leqslant \frac{(4e)^d\|K\|_\infty^{2d}\sigma^2}{\lambda_0^2 n h^{\gamma_1}} \quad \text{and} \quad \bigl|\mathrm{Bias}\,\hat{m}_n(x_0; p, h, K)\bigr| \leqslant \frac{(4e^{1/2})^d d^{\beta_0} L\|K\|_\infty^d}{\lambda_0 \beta_0!} h^{\gamma_2}$$

for some $\gamma_1 > 0$, depending only on d, and $\gamma_2 > 0$, depending only on β, that should be specified.

[*Hint: You may use the fact that*

$$\sum_{\alpha\in\mathbb{N}_0^d:|\alpha|=\ell} \frac{1}{\alpha!} = \frac{d^\ell}{\ell!}$$

for $\ell \in \mathbb{N}_0$.]

4 Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, let $(\mathcal{Y}, \mathcal{B})$ be a measurable space and let $g : \mathcal{X} \to \mathcal{Y}$ be measurable. Define the *pushforward* measure of μ under g.

State the Lebesgue decomposition theorem. What is meant by an f*-divergence* between two probability measures P and Q on $(\mathcal{X}, \mathcal{A})$?

State and prove the data processing inequality. [You may assume the Radon–Nikodym theorem.]

Define the *squared Hellinger distance* $\mathrm{H}^2(P, Q)$ between P and Q, and state an alternative expression for it when P and Q have densities f and g respectively with respect to a σ-finite measure μ. Prove that if $P_1, \ldots, P_M$ are probability measures on $\mathcal{X}$ and $A_1, \ldots, A_M \in \mathcal{A}$ form a partition of $\mathcal{X}$, then

$$\frac{1}{M} \sum_{j=1}^{M} P_j(A_j) \leqslant \frac{1}{M} + \inf_{Q \in \mathcal{Q}} \sqrt{\frac{1}{M} \sum_{j=1}^{M} \mathrm{H}^2(P_j, Q)} \times \sqrt{1 - \frac{1}{4M} \sum_{j=1}^{M} \mathrm{H}^2(P_j, Q)},$$

where $\mathcal{Q}$ denotes the set of all probability distributions on $\mathcal{X}$. [You may use the fact that f-divergences are jointly convex.]

[*Hint: You may wish to use the fact that* $\sqrt{pq} + \sqrt{(1-p)(1-q)} \leqslant \sqrt{1-(p-q)^2}$ *for* $p, q \in [0, 1]$.]

END OF PAPER