UNIVERSITY OF CAMBRIDGE

MAT3, MAMA

MATHEMATICAL TRIPOS Part III

Tuesday, 11 June, 2019 9:00 am to 11:00 am

PAPER 210

TOPICS IN 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
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 $X = (X_1, \ldots, X_d)$ be a centred Gaussian random vector in $\mathbb{R}^d$ and assume $EX_i^2 \leqslant 1$ for all $i = 1, \ldots, d$. Denote by $\|x\|_\infty = \max_{i \leqslant d} |x_i|$ the maximum norm. Show that for all $u \geqslant 0$ we have

$$\Pr\left(\|X\|_\infty > \sqrt{2\log 2d} + u\right) \leqslant e^{-\frac{2u^2}{\pi^2}}.$$

[Hint: You may use the following inequality from lectures without proof: For any Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$ with gradient vector ∇f, and any $\lambda > 0$, we have

$$E^X\left[e^{\lambda(f(X) - Ef(X))}\right] \leqslant E^{X,Y}\left[e^{\frac{\lambda\pi}{2}\langle \nabla f(X), Y\rangle_{\mathbb{R}^d}}\right]$$

where Y is an independent copy of X, and where $\langle \cdot, \cdot \rangle_{\mathbb{R}^d}$ denotes the standard Euclidean inner product on $\mathbb{R}^d$.]

2 Let Π be a product prior distribution in $\mathbb{R}^K$ whose marginal distributions all have identical probability density function $\pi : \mathbb{R} \to (0, \infty)$, and such that $|\log \pi(x) - \log \pi(y)| \leqslant c|x - y|$ for some constant $c > 0$ and all $x, y \in \mathbb{R}$. Suppose observations

$$Y = \left(Y_k = \theta_k + \frac{1}{\sqrt{n}} g_k : k \in \mathbb{N}\right) \sim P_\theta^Y, \quad n \in \mathbb{N},$$

arise in the Gaussian sequence space model, and denote by $\Pi(\cdot|Y)$ the resulting posterior distribution on $\mathbb{R}^K$. Consider inference on the functional

$$T(\theta) = 2\theta_2 - \theta_1 + 4\theta_4$$

via a one-sided posterior credible interval $C_n = \{t \in \mathbb{R} : t - T(Y) \leqslant R_n\}$ for $T(\theta)$ where R_n is such that $\Pi(C_n|Y) = 1 - \alpha$ for some $\alpha > 0$. Show that for every θ_0 such that $\sum_k \theta_{0,k}^2 < \infty$ we have

$$P_{\theta_0}^Y(T(\theta_0) \in C_n) \to 1 - \alpha \text{ as } n \to \infty$$

and find the limit in probability of $\sqrt{n}R_n$.

[Hint: You may use without the proof the fact that if $E_{P_n} e^{tX} \to E_P e^{tX}$ almost surely for all $t \in \mathbb{R}$, a sequence of possibly random probability measures P_n on $\mathbb{R}$ and P the law of a normal distribution on $\mathbb{R}$, then also $\sup_{s \in \mathbb{R}} |P_n(X \leqslant s) - P(X \leqslant s)| \to 0$ almost surely.]

3 Suppose you are given observations

$$Y = \big(Y_k = \theta_k + \frac{1}{\sqrt{n}} g_k : k \in \mathbb{N}\big) \sim P_\theta^Y, \quad n \in \mathbb{N}, \ \theta \in \ell_2,$$

in the Gaussian sequence space model, where

$$\ell_2 = \{(\theta_k : k \in \mathbb{N}) : \|\theta\|_{\ell_2}^2 \equiv \sum_{k \in \mathbb{N}} \theta_k^2 < \infty\}.$$

For given $\theta_0 \in \ell_2$, consider the testing problem

$$H_0 : \theta = \theta_0 \ \ vs. \ H_1 : \theta \in \Theta_n : \|\theta - \theta_0\|_{\ell_2} \geqslant M\varepsilon_n, \quad M > 0, \varepsilon_n > 0,$$

where $\Theta_n = \{\theta \in \ell_2 : \theta_k = 0 \ \forall \ k > K_n, \|\theta\|_{\ell_2} \leqslant n\}$. Suppose $K_n \to \infty$ in such a way that

$$\varepsilon_n^2 = \frac{K_n}{n} \log n \to 0$$

as $n \to \infty$. Show that there exists a test $\Psi_n(Y)$ (a function of Y taking only values $0, 1$) such that, for every $c_1 > 0$ we can find M large enough such that the type-I and type-II errors of Ψ_n are bounded as

$$\max\big[E_{\theta_0}\Psi_n, \sup_{\theta \in H_1} E_\theta(1 - \Psi_n)\big] \leqslant e^{-c_1 K_n \log K_n}, \quad n \in \mathbb{N}.$$

[Hint: You may use without proof results about covering numbers of balls in Euclidean spaces, as well as basic concentration inequalities for one-dimensional Gaussian random variables, provided they are clearly stated.]

4 Suppose you are given observations

$$Y = \big(Y_k = \theta_k + \frac{1}{\sqrt{n}} g_k : k \in \mathbb{N}\big) \sim P_\theta^Y, \quad n \in \mathbb{N}, \ \theta \in \ell_2,$$

in the Gaussian sequence space model, where

$$\ell_2 = \{(\theta_k : k \in \mathbb{N}) : \|\theta\|_{\ell_2}^2 \equiv \sum_{k \in \mathbb{N}} \theta_k^2 < \infty\}.$$

Suppose Θ is a compact subset of ℓ_2 such that for every $\varepsilon > 0$ it can be covered by $\exp\Big[(1/\varepsilon)^{1/4}\Big]$-many ℓ_2-balls of radius ε. Define the least squares estimator $\hat{\theta} = \hat{\theta}(Y)$ for Θ and prove that for every $\theta_0 \in \Theta$ we have

$$P_{\theta_0}^Y \left(\|\hat{\theta} - \theta_0\|_{\ell_2} \geqslant c n^{-4/9} \right) \to 0$$

as $n \to \infty$ for $c > 0$ a large enough constant independent of n.

[Hint: You may use concentration results for Gaussian processes from lectures without proof, provided they are clearly stated.]

END OF PAPER