UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 8 June, 2015 1:30 pm to 3:30 pm

PAPER 36

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

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 P, Q be probability measures on a measure space $(\Omega, \mathcal{A})$ with common dominating measure μ, and denote the corresponding probability density functions by $dP/d\mu$ and $dQ/d\mu$, respectively.

Define the Kullback-Leibler divergence $K(P,Q)$ between P and Q. Prove that

$$\int_\Omega \left| \frac{dP}{d\mu} - \frac{dQ}{d\mu} \right| d\mu \leqslant \sqrt{2K(P,Q)}.$$

Now define the *Hellinger distance* as

$$H^2(P,Q) = \int_\Omega \left(\sqrt{\frac{dP}{d\mu}} - \sqrt{\frac{dQ}{d\mu}} \right)^2 d\mu.$$

Prove that

$$H(P,Q) \leqslant \sqrt{K(P,Q)}.$$

[Hint: You may use $-\log(x+1) \geqslant -x$ for all $x > -1$ without proof.]

2

Define the Gaussian white noise model with drift function f and noise level $1/\sqrt{n}, n \in \mathbb{N}$.

Let $g_1, \ldots, g_N$ be random variables each with distribution $N(0,1)$. Show that, for every $N \in \mathbb{N}$,

$$E \max_{k=1,\ldots N} |g_k| \leqslant \sqrt{2\log(2N)}.$$

For a dyadic partition of $[0,1]$ with grid points

$$\frac{k}{2^J}, \quad k = 0, \ldots, 2^J - 1, \; J \in \mathbb{N},$$

define the Haar-wavelet projection $\Pi_{V_J}(f)$ of f. Given observations in the Gaussian white noise model, construct an unbiased estimator $\widehat{\Pi_{V_J}(f)}$ of $\Pi_{V_J}(f)$ and show that for every $n \in \mathbb{N}$ and $J \in \mathbb{N}$,

$$E\|\widehat{\Pi_{V_J}(f)} - \Pi_{V_J}(f)\|_\infty \leqslant \sqrt{\frac{2^J(2J+2)\log 2}{n}}.$$

[Here $\|h\|_\infty$ denotes the supremum norm $\sup_{x\in[0,1]} |h(x)|$ of a function $h : [0,1] \to \mathbb{R}$.]

3

State Hoeffding's inequality.

Denote by $\Omega = \{-1, 1\}^n$ the discrete hypercube, and endow it with the Hamming distance

$$\rho(\omega, \omega') = \sum_{i=1}^{n} 1_{[\omega_i \neq \omega_i']},$$

where $\omega = (\omega_1, \ldots, \omega_n), \omega' = (\omega_1', \ldots, \omega_n')$ are points in Ω. For $n \geqslant 8$, prove that there exists a set $S \subset \Omega$ of cardinality $|S| \geqslant e^{n/4}$ that is $n/8$-separated for the ρ-distance, that is, S satisfies $\min_{s,s' \in S, s \neq s'} \rho(s, s') \geqslant n/8$. [You may use Hoeffding's inequality without proof.]

4

Let $p, n \in \mathbb{N}$ and let the $n \times p$ random matrix $X = (X_{ij})$ consist of i.i.d. $N(0,1)$ entries. Define

$$\hat{\Sigma} = \frac{1}{n} X^T X$$

and

$$\mathbb{R}_k^p = \{\theta \in \mathbb{R}^p : \theta_j = 0 \ \forall \ j > k\}, \ k < p.$$

Assume that $n \geqslant Ck \log p$ for some constant C. Show that for C large enough there exists a constant $C' > 0$ such that

$$P\left(\sup_{\theta \in \mathbb{R}_k^p, \theta \neq 0} \left| \frac{\theta^T \hat{\Sigma} \theta - \theta^T \theta}{\theta^T \theta} \right| > \frac{1}{2} \right) \leqslant 2 \exp\{-C' k \log p\}.$$

[You may use the following inequality without proof in your argument: For $g_1, \ldots, g_n$ i.i.d. $N(0,1)$ random variables and $Z = \sum_{i=1}^n (g_i^2 - 1)$,

$$P(|Z| > 4(\sqrt{nz} + z) \leqslant 2e^{-z}.$$

You may further use the fact that one requires at most $(A/\delta)^k$ balls of radius not exceeding $0 < \delta < A$ to cover the unit ball of a k-dimensional Euclidean space, where A is a numerical constant.]

END OF PAPER