UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 6 June, 2016 1:30 pm to 3:30 pm

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

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

For some positive integers n and m, we observe n balls placed in m bins, labelled from 1 to m. Let $S_1, \dots, S_d$ non-empty disjoint subsets of $\{1, \dots, m\}$, each with cardinality $k < m$. Under distribution $\mathbf{P}_0$, the balls are placed independently and uniformly in the m bins. Under distribution $\mathbf{P}_1$, an element j of $\{1, \dots, d\}$ is drawn uniformly at random, and conditionally on j, the balls are placed in the m bins with the following distribution $\mathbf{Q}_j$, independently: We flip a coin with probability of heads $\pi \in (0,1)$. If it falls on heads, the ball is placed uniformly in one of the k bins with a label in S_j, and if it falls on tails, uniformly in the m bins.

We denote by $b_i \in \{1, \dots, m\}$ the label of the bin in which ball i is placed. The number of balls in bins with index in S_j is $c_j = |\{i \,|\, b_i \in S_j\}|$. We write $\varepsilon = k/m$.

(a) For any $j \in \{1, \dots, d\}$, give the distribution of c_j, under $\mathbf{P}_0$ and $\mathbf{Q}_j$ using the parameter ε.

(b) Show without proof that if X has binomial distribution $B(n,p)$, then $X - np$ is sub-Gaussian with parameter $n/4$. Use this to derive without proof a bound on the probability that $X - np > t$, for all $t > 0$. State any theorem that you use.

(c) Fix $\delta \in (0,1)$. Using the statistic $C = \max_{1 \leqslant j \leqslant d} c_j$, give a test ψ such that whenever

$$\pi(1-\varepsilon) > \sqrt{\frac{\log(d)}{2n}} + 2\sqrt{\frac{\log(1/\delta)}{2n}}\,,$$

we have

$$\mathbf{P}_0(\psi = 1) \vee \mathbf{P}_1(\psi = 0) \leqslant \delta\,.$$

(d) Define the chi-square divergence between two distributions. Write $\chi^2(\mathbf{P}_1, \mathbf{P}_0)$ in terms of the likelihood ratios $\mathbf{Q}_j/\mathbf{P}_0$.

(e) Using without proof the fact that

$$\chi^2(\mathbf{P}_1, \mathbf{P}_0) = \Big(1 - \frac{1}{d}\Big)(1-\pi^2)^n + \frac{1}{d}\big(1 + \pi^2(1/\varepsilon - 1)\big)^n - 1\,,$$

show that for $\nu \in (0, 1/2)$, when

$$\pi\big(\varepsilon^{-1} - 1\big)^{1/2} \leqslant \sqrt{\frac{\log(16\nu^2 d)}{n}}\,,$$

we have

$$\inf_{\psi} \mathbf{P}_0(\psi = 1) \vee \mathbf{P}_1(\psi = 0) \geqslant \frac{1}{2} - \nu\,,$$

where the infimum is taken over all tests ψ.

2

Let $d \geqslant 2$ be an integer and S be an unknown subset of $\{1, \ldots, d\}$ of k consecutive integers (modulo d). Let $u(S)$ be the unit vector of $\mathbb{R}^d$ defined by $u(S) = \mathbf{1}_S/\sqrt{k}$ (where $\mathbf{1}_S$ is the vector with ones in S, zeroes elsewhere), and by $\mathcal{S}$ the collection of these d subsets.

For $\theta > 0$, let Y be a real-valued random variable with distribution $\mathcal{N}(0, \theta)$ and Z be a random vector of $\mathbb{R}^d$ with distribution $\mathcal{N}(0, I_d)$. Let $X_1, \ldots, X_n$ be n i.i.d. random vectors of $\mathbb{R}^d$ such that

$$X_i = Y_i\, u(S) + Z_i\,,$$

where $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ are $2n$ independent copies of respectively, Y and Z.

(a) Give the distribution $\mathbf{P}_{\theta,S}$ of X_1. What is the expectation of $\hat{\Sigma} = \sum_{i=1}^n X_i X_i^\top/n$?

(b) Let X be a sub-Gaussian random variable with parameter σ^2. Give, without proof, a bound on the probability that $X^2 - \mathbb{E}[X^2] > t$, for all $t > 0$. State the theorems that you use.

(c) Fix $\delta \in (0, 1)$, let $n \geqslant \log(d/\delta)$. Using the statistic $\Lambda = \max_{S \in \mathcal{S}} u(S)^\top \hat{\Sigma} u(S)$, give a test ψ such that for some constant $C > 0$, whenever

$$\theta > C\sqrt{\frac{\log(d/\delta)}{n}}\,,$$

we have

$$\mathbf{P}_0^{\otimes n}(\psi = 1) \vee \max_{S \in \mathcal{S}} \mathbf{P}_{\theta,S}^{\otimes n}(\psi = 0) \leqslant \delta\,,$$

where we write $\mathbf{P}_0 = \mathbf{P}_{0,S}$ without ambiguity, and $\otimes n$ denotes n independent samples.

(d) Define the chi-square divergence between two distributions. Using without proof the fact that for $\theta < 1/2$,

$$\mathbb{E}_0\left[\frac{\mathrm{d}\mathbf{P}_S}{\mathrm{d}\mathbf{P}_0}\frac{\mathrm{d}\mathbf{P}_T}{\mathrm{d}\mathbf{P}_0}\right] = \Big(1 - \theta^2\big(u(S)^\top u(T)\big)^2\Big)^{-1/2}\,,$$

show that for $\theta < 1/2$

$$\chi^2(\mathbf{P}_{\theta,n}, \mathbf{P}_0^{\otimes n}) = \mathbb{E}_R\Big[\Big(1 - \frac{\theta^2 R^2}{k^2}\Big)^{-n/2}\Big] - 1 \leqslant \mathbb{E}_R\Big[e^{\frac{n\theta^2 R}{k}}\Big] - 1\,,$$

where R is the cardinality of the intersection of two random subsets of $\{1, \ldots, d\}$ of size k drawn independently and uniformly. [Hint: you can use that $1/(1-t) \leqslant e^{\frac{t}{1-t}}$ for $t < 1/2$]

(e) For $\varepsilon \in (0, 1)$, $\theta < 1/2$, and $k \leqslant d^{1-\varepsilon}$, show that whenever

$$\theta \leqslant \sqrt{\frac{\varepsilon \log(\alpha d)}{n}}\,,$$

we have for some constant $\nu_{\varepsilon,\alpha}$ that depends on ε and α, taking the infimum is taken over all tests ψ.

$$\inf_\psi \mathbf{P}_0^{\otimes n}(\psi = 1) \vee \max_{S \in \mathcal{S}} \mathbf{P}_{\theta,S}^{\otimes n}(\psi = 0) \geqslant \frac{1}{2} - \nu_{\varepsilon,\alpha}\,.$$

3

Let G be a given connected graph on n vertices labelled 1 through n, with a set of m edges E. We seek to estimate a signal $\theta^* \in \mathbb{R}^n$, where θ^*_i is associated to vertex i. It is assumed that this signal is "smooth over the graph", i.e. that the coefficients of θ^* are not too different for two vertices that are connected. This is quantified by the function $s : \mathbb{R}^n \to \mathbb{R}$, defined by

$$s(\theta) = \sum_{(i,j)\in E} |\theta_i - \theta_j| \,.$$

We observe y, a noisy version of θ^*. For each $1 \leqslant i \leqslant n$, we have for $z \in \mathsf{sG}_n(1)$

$$y_i = \theta^*_i + z_i \,.$$

A $m \times n$ matrix D is called an *incidence matrix* of G if for each $e = (i, j) \in E$, the row D_e satisfies $D_{e,i} = 1$, $D_{e,j} = -1$, and 0 everywhere else. As an example, for a square graph (cycle graph on four vertices), an example of incidence matrix is

$$\begin{pmatrix} 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \end{pmatrix}$$

For $\lambda > 0$, the estimator $\hat\theta$ is defined by $\hat\theta \in \operatorname{argmin}_{\theta\in\mathbb{R}^n} \frac{1}{n}\|y-\theta\|_2^2 + \lambda\|D\theta\|_1$.

(a) Let D be an incidence matrix of G. Write $s(\theta)$ as a function of $D\theta$. Show that this expression is valid for any incidence matrix. Show that for any $\theta \in \mathbb{R}^n$ such that $D\theta = 0$, we have $\theta = t\mathbf{1}$, for some $t \in \mathbb{R}$.

(b) Show that

$$\frac{1}{n}\|\theta^* - \hat\theta\|_2^2 \leqslant \frac{2}{n}\langle z, \hat\theta - \theta^*\rangle + \lambda\|D\theta^*\|_1 - \lambda\|D\hat\theta\|_1 \,.$$

(c) Let $D^\dagger$ be the $n \times m$ pseudoinverse of D, i.e. such that $D^\dagger D = I_n - \Pi_{\mathbf{1}}$, where $\Pi_{\mathbf{1}} = \mathbf{1}\mathbf{1}^\top/n$ is the orthonormal projector on the span of $\mathbf{1}$. Let λ be such that $\|(D^\dagger)^\top z\|_\infty \leqslant n\lambda/2$ with probability $1-\delta$. Show that it holds with probability $1-\delta$ that

$$\frac{1}{n}\|\theta^* - \hat\theta\|_2^2 \leqslant 2\lambda s(\theta^*) + \frac{2}{n}\|\Pi_{\mathbf{1}} z\|_2\, \|\theta^* - \hat\theta\|_2 \,.$$

(d) Let $A \in \mathbb{R}^{k\times m}$ be a matrix whose columns satisfy $\|A^{(i)}\|_2 \leqslant 1$, and g a random vector of $\mathbb{R}^k$ that is sub-Gaussian with parameter τ^2. Stating the theorems that you use, show that with probability $1-\delta$, we have

$$\|A^\top g\|_\infty \leqslant \tau\sqrt{\log(2m)} + \tau\sqrt{\log(1/\delta)} \,.$$

(e) Let $L = \max_{1\leqslant e\leqslant m} \|D^\dagger_{\cdot,e}\|_2$ be the maximum of the ℓ_2 norm of the columns of $D^\dagger$ (or of the rows of $(D^\dagger)^\top$). Show that for a λ independent of θ^* to be determined, it holds with probability $1-2\delta$ that

$$\frac{1}{n}\|\theta^* - \hat\theta\|_2^2 \leqslant \frac{4}{1-t^2}\frac{s(\theta^*)L}{n}\sqrt{\log(2m)} + \frac{4}{1-t^2}\frac{s(\theta^*)L}{n}\sqrt{\log(1/\delta)} + \frac{1}{t^2(1-t^2)}\frac{c_\delta}{n} \,,$$

for some constant c_δ that only depends on δ, and some real $t \in (0,1)$.

[Hint: for all reals t, a, b, it holds that $2ab \leqslant t^2a^2 + b^2/t^2$.]

4

Let n be an even integer, and $\mathcal{P}$ the collection of subsets of $\{1, \ldots, n\}$ with $n/2$ elements. For an unknown set $S \in \mathcal{P}$, a random graph G on n vertices is drawn by placing edges independently between distinct vertices i, j with probability $1/2$ if $(i,j) \subset S$ or $(i,j) \subset S^c$, and probability $1/2 - t/\sqrt{n}$ otherwise, where t is a fixed positive constant smaller than $\sqrt{n}/6$. We denote by A the symmetric *adjacency matrix* of the graph G, where $A_{ij} = 1$ if there is an edge between i and j, 0 otherwise.

(a) What is the expectation of A, denoted by $A_0 = \mathbb{E}[A]$? Draw a schematic picture when $S = \{1, \ldots, n/2\}$. Show that $A_0 + I_n/2$ has rank 2, and give its eigen-decomposition. The matrix M is defined by

$$M = A + I_n/2 - \Big(\frac{1}{2} - \frac{t}{2\sqrt{n}}\Big)U\,,$$

where $U_{ij} = 1$ for all i, j. What is the expectation of M, denoted by M_0?

(b) State the Davis–Kahan curvature lemma. We define the estimator $\hat{v}$ by $\hat{v} \in \operatorname{argmin}_{\|v\|_2 = 1} v^\top M v$. Show that

$$\mathbb{E}\big[\|\hat{v}\hat{v}^\top - vv^\top\|_F\big] \leqslant \frac{c}{t}$$

for some constant $c > 0$.

(c) Let $\hat{S} = \big\{i \in \{1, \ldots, n\} \,|\, \hat{v}_i > 0\big\}$, and let $\Delta(\hat{S}, S) = \min_{\sigma \in \{-1,1\}} |\mathbf{1}_{\hat{S}} - \sigma \mathbf{1}_S|$. Explain what $\Delta(\hat{S}, S)$ represents, in terms of the partitions (S, S^c) and $(\hat{S}, \hat{S}^c)$. Show that

$$\mathbb{E}[\Delta(\hat{S}, S)] \leqslant n\frac{2c}{t}\,.$$

(d) For any $S \in \mathcal{S}$, we denote by $\mathbf{P}_S$ the distribution of G. For any $S, S' \in \mathcal{P}$, show that

$$\begin{aligned}\mathsf{KL}(\mathbf{P}_S, \mathbf{P}_{S'}) = &\sum_{(i,j) \in \partial S \setminus \partial S'} \mathsf{KL}(\mathsf{Ber}(1/2), \mathsf{Ber}(1/2 - t/\sqrt{n})) \\ &+ \sum_{(i,j) \in \partial S' \setminus \partial S} \mathsf{KL}(\mathsf{Ber}(1/2 - t/\sqrt{n}), \mathsf{Ber}(1/2))\,,\end{aligned}$$

for properly defined sets ∂S and $\partial S'$. Show that for $p, q \in (0,1)$, we have

$$\mathsf{KL}(\mathsf{Ber}(p), \mathsf{Ber}(q)) \leqslant \frac{(p-q)^2}{q(1-q)}\,.$$

(e) State Fano's inequality. You can assume without proof the existence of $\mathcal{M}$, a subset of $\mathcal{P}$ such that for all distinct $S, S' \in \mathcal{M}$, $\Delta(S, S') \geqslant n(1/2 - \varepsilon)$, and $|\mathcal{M}| \geqslant e^{\varepsilon^2 n}$, for some $\varepsilon \in (0, 1/2)$. Using this set, show that for some constant $c' > 0$

$$\inf_{\hat{\Sigma}} \max_{S \in \mathcal{P}} \mathbf{P}_S\big(\Delta(\hat{\Sigma}, S) \geqslant n(1/2 - 2\varepsilon)\big) \geqslant 1 - \frac{c' t^2}{\varepsilon^2}\,,$$

where the infimum is taken over all measurable estimators $\hat{\Sigma}$ of S.

END OF PAPER