UNIVERSITY OF CAMBRIDGE

MAMA/225, NST3AS/225, MAAS/225

MAT3

MATHEMATICAL TRIPOS Part III

Thursday, 8 June, 2023 1:30pm to 3:30pm

PAPER 225

FUNCTIONAL DATA ANALYSIS

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 element of $L^2[0,1]$ such that $\mathbb{E}\|X\|^2 < \infty$, $\mathbb{E}X = \mu$ and with covariance operator $C_X(\cdot) = \mathbb{E}(\langle (X-\mu), \cdot\rangle (X-\mu))$.

(a) State and Prove the Karhunen-Loève Theorem.

[Mercer's Theorem can be used without proof as long as it is stated and you do not need to prove a continuous covariance function is a Mercer Kernel].

(b) Let

$$c_X(s,t) = \min(s,t) - st \qquad s,t \in [0,1]$$

where c_X is the kernel of C_X. Use the Karhunen-Loève Theorem to find an expansion of X in terms of the eigencomponents of C_X.

2 Let $X, X_1, \ldots, X_n$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|X\|^4 < \infty$, $\mathbb{E}X = \mu$ and with covariance operator C_X. Let $X^*, X_1^*, \ldots, X_n^*$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|X^*\|^4 < \infty$, $\mathbb{E}X^* = \mu^*$ and with covariance operator $C_{X^*} = C_X$. Assume that the two samples are independent of each other.

By considering the eigendecomposition of C_X, for some $K \in \mathbb{N}$, find a K-dimensional test to determine whether $\mu = \mu^*$. Determine its asymptotic properties as $n \to \infty$ under the null hypothesis $H_0 : \mu = \mu^*$ and under the alternative $H_A : \mu \neq \mu^*$.

[You may assume that the first $K+1$ eigenvalues of C_X are all distinct. If you state them, you may use the convergence properties of eigenvalues, eigenfunctions and any version of the central limit theorem without proof].

3

(a) Let X be a random element of $L^2[0,1]$ such that $\mathbb{E}\|X\|^2 < \infty$, $\mathbb{E}X = \nu$. Let h be a random element of $L^2[0,1]$ such that $\mathbb{E}\|h\|^2 < \infty$, $\mathbb{E}h(t) = t, t \in [0,1]$, and where $h(0) = 0$, $h(1) = 1$, and h is strictly monotone. Assume that h and h^{-1} are differentiable. Define $Y = X \circ h^{-1}$, where $(X \circ h)(t) = X(h(t))$. Define $\mathbb{E}Y = \mu$.

(i) Show that $\mathbb{E}\|Y-\mu\|^2$ can be decomposed into terms involving $A\mathbb{E}\|X-\nu\|^2$, $A\|\nu\|^2$, and $\|\mu\|^2$ and find the constant A in terms of the relationship between functions of X and h. When does $A = 1$?

(ii) Let $\gamma : [0,1] \to [0,1]$ be a diffeomorphism, $\gamma(0) = 0, \gamma(1) = 1$. Define $Q(h(t)) = \frac{h'(t)}{\sqrt{|h'(t)|}}$, where $'$ denotes the derivative. Let $h_1, \ldots, h_n$ be realizations of h. Show that

$$\|Q(h_i) - Q(h_j)\| = \|Q(h_i \circ \gamma) - Q(h_j \circ \gamma)\|, \quad 1 \leqslant i, j \leqslant n$$

.

(b) (i) Define the Procrustes Distance

(ii) Let C_1 and C_2 be covariance operators on $L^2[0,1]$. Assume that there exists L_i^p such that $L_i^p(L_i^p)^* = C_i^p$, $i = 1,2$, and where * indicates the adjoint operator. Assume $L_i^p \to L_i$ in the Hilbert-Schmidt norm as $p \to \infty$ where $L_i L_i^* = C_i$. Prove that

$$d_P(C_1^p, C_2^p)^2 \to d_P(C_1, C_2)^2 \quad \text{as } p \to \infty$$

where $d_P(C_1, C_2)^2 = \|L_1\|_{\text{HS}}^2 + \|L_2\|_{\text{HS}}^2 - 2\sup_{R \in O\{L^2[0,1]\}} tr(R^* L_2^* L_1)$ is the space of unitary operators on $L^2[0,1]$ and $\|\cdot\|_{\text{HS}}$ is the Hilbert-Schmidt norm, and with $d_P(C_1^p, C_2^p)^2$ defined analogously.

4 Let $X, X_1, \ldots, X_n$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|X\|^4 < \infty$, $\mathbb{E}X = 0$ and with positive definite covariance operator C_X. Let $\epsilon_i \sim N(0, \sigma^2)$, and assume the sets $\{X_i\}$ and $\{\epsilon_i\}$ are independent. Let

$$Y_i = \int_0^1 X_i(t)\beta(t)dt + \epsilon_i, \quad i = 1, \ldots, n.$$

(a) Assume an orthonormal basis of fixed basis functions $\{B_k\}_{k=1}^{\infty}$ of $L^2[0,1]$. For some $K \in \mathbb{N}$, let $\beta(t)$ be expressed as a truncated sum of the basis functions $\beta(t) = \sum_{k=1}^{K} c_k B_k(t)$. This truncation is an approximation to $\beta(t) = \sum_{k=1}^{\infty} c_k B_k(t)$.

(i) Find the bias from the least squares estimator using the approximation and for K fixed, show that, in general, it is not consistent as $n \to \infty$.
[You may use without proof that for a suitably chosen matrix $\tilde{X}$, $\frac{1}{n}\tilde{X}^T\epsilon \xrightarrow{p} 0$ and $\frac{1}{n}\tilde{X}^T\tilde{X} \xrightarrow{p} \Sigma_{\tilde{X}}$ for some positive definite matrix $\Sigma_{\tilde{X}}$].

(ii) If C_X is known, explain (briefly) the difference if the fixed basis functions $\{B_k(t)\}_{k=1}^{\infty}$ were chosen as the eigenbasis of C_X.

(b) Assume $\{B_k\}_{k=1}^{\infty}$ is an orthonormal basis of $L^2[0,1]$ with second derivatives, and for some fixed $K \in \mathbb{N}$, let $\beta(t) = \sum_{k=1}^{K} c_k B_k(t)$. For some $\rho \in \mathbb{R}$, let the loss function L,

$$L(\beta) = \sum_{i=1}^{n} |Y_i - \langle X_i, \beta\rangle|^2 + \rho\|\beta''\|^2,$$

where $\beta''(t)$ is the second derivative of $\beta(t)$.

Find an estimator for the coefficients c_k, which minimize the loss function L.

[You may assume it is possible to calculate $\langle X_i, B_k\rangle$].

END OF PAPER