UNIVERSITY OF CAMBRIDGE

MAMA/225, NST3AS/225, MAAS/225

MAT3

# MATHEMATICAL TRIPOS Part III

Tuesday 4 June 2024 9:00 am to 11:00 am

## 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**

(a) Let $X$ be a random element of $L^2[0,1]$ such that $\mathbb{E}\|X\|^2 < \infty$, $\mathbb{E}X = 0$ and with covariance operator $C_X(\cdot) = \mathbb{E}(\langle X, \cdot\rangle X)$. Let $c_X$ be the kernel of $C_X$. Determine whether each of the following is a covariance operator or not, and if so, find its eigendecomposition.

(i) $U$ with kernel $u(s,t) = \min(s,t), \quad 0 \leqslant s,t \leqslant 1$

(ii) $U = \mathrm{Id}$, where Id is the identity operator

(iii) $U$ with kernel $u(s,t) = (s+t)^2, \quad 0 \leqslant s,t \leqslant 1$

(b) 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 covariance operator $C_X$.

For a $\mu_0 \in L^2[0,1]$, not necessarily equal to 0, find the asymptotic properties of the estimator

$$\hat{c}_{\mu_0}(s,t) = \frac{1}{n}\sum_{i=1}^{n}(X_i(t) - \mu_0(t))(X_i(s) - \mu_0(s))$$

for the kernel $c_X(s,t)$ 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 $C_X$. Consider the problem of testing the hypotheses

$$H_0 : \mu = 0 \quad \text{vs} \quad H_1 : \mu \neq 0$$

(a) Using a function of $\|\frac{1}{n}\sum_{i=1}^{n} X_i\|$, give a test of the above hypothesis, and find its properties under the null and alternative hypotheses.

(b) Using Functional Principal Component Analysis, give, for some $K \in \mathbb{N}$, a $K$ dimensional test of the above hypothesis, and find its properties under the null and alternative hypotheses.

(c) Give a permutation-type test to test the above hypothesis, and explain how its p-value would be found, giving any assumption you need to make to find this. (Its properties under the null and alternative do not need to be given).

(d) Give a short discussion of the advantages and disadvantages of your three proposed tests in (a)-(c).

[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** Let $C_1, C_2, \ldots, C_n$, $n > 1$ be covariance operators on a separable Hilbert Space

(a) Let $d_L(C_i, C_j) = \|C_i - C_j\|_{HS}$, where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm. Find $C$ which minimizes

$$\sum_{k=1}^{n} d_L(C_k, C)^2$$

(b) Define Procrustes distance $d_P(C_i, C_j)$ and prove the Procrustes distance is given by

$$d_P(C_i, C_j)^2 = \|L_i\|_{HS}^2 + \|L_j\|_{HS}^2 - 2\sum_{k=1}^{\infty} \sigma_k$$

where $C_i = L_i L_i^*$, $L_i^*$ is the adjoint of $L_i$, with analogous definitions for $C_j$, and $\sigma_k$ are the singular values of $L_j^* L_i$.

(c) Let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal basis for $L^2[0,1]$. Let $C_1(\cdot) = \lambda_1 \langle e_1, \cdot \rangle e_1$, and $C_2(\cdot) = \lambda_2 \langle e_2, \cdot \rangle e_2$ for some $\lambda_1, \lambda_2 > 0$. Find $d_L(C_1, C_2)$, $d_R(C_1, C_2)$ and $d_P(C_1, C_2)$, where $d_R(C_1, C_2)$ is the square-root distance.

**4** Let $X$ be a random element of $L^2[0,1]$ such that $\mathbb{E}\|X\|^2 < \infty$, $\mathbb{E}X = 0$. Let $\epsilon$ be a random element of $L^2[0,1]$ such that $\mathbb{E}\|\epsilon\|^2 < \infty$, $\mathbb{E}(\epsilon) = 0$, and let $X, \epsilon$ be mutually independent. Let $(Y, X, \epsilon)$ follow the model,

$$Y(t) = \int_0^1 \beta(t,s) X(s) ds + \epsilon(t), \quad t \in [0,1]$$

where $\beta \in L^2([0,1] \times [0,1])$, and $\int_0^1 \int_0^1 \beta^2(t,s) dt ds < \infty$.

Find an expression for

$$\int_0^1 \mathrm{Var}(\mathbb{E}(Y(t)|X)) dt$$

in terms of the principal component functions and scores of $X$ and $Y$.

## END OF PAPER