UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Friday, 10 June, 2022 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 Let X be a random element of $L^2[0,1]$ such that $\mathbb{E}\|X\|^2 \leqslant \infty$, $\mathbb{E}X = \mu$ and with covariance operator $C_X(\cdot) = \mathbb{E}(\langle X, \cdot \rangle X)$.

(a) Prove that C_X is symmetric, non-negative definite and trace class.

(b) Show that the Identity operator (the operator which takes any element of $L^2[0,1]$ to itself) is not a covariance operator

(c) Show that not every bounded, symmetric, Hilbert-Schmidt operator is a covariance operator

(d) Now let $X_1, \ldots, X_n$ be i.i.d. random elements of $L^2[0,1]$ with the same distribution as X above.

Show that $\mathbb{E}(\hat{\mu}) = \mu$ and $\mathbb{E}\|\hat{\mu} - \mu\|^2 = O(\frac{1}{n})$, where $\hat{\mu}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$.

(e) Let (λ_k, ϕ_k) and $(\lambda_{k'}, \phi_{k'})$ be the $k, k' \in \mathbb{N}$ (not necessarily distinct) eigenvalue and associated eigenfunction pairs of C_X.

Find $\mathrm{Cov}(\langle \hat{\mu}, \phi_k \rangle, \langle \hat{\mu}, \phi_{k'} \rangle)$.

2 Let $X, X_1, \ldots, X_n$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|X\|^4 \leqslant \infty$, $\mathbb{E}X = \mu$ and with covariance operator C_X. Let $\{\lambda_k, \phi_k\}_{k=1}^{\infty}$ be the eigenvalues/eigenfunctions of C_X, and assume $\lambda_1 > \lambda_2 > \ldots$.

(a) Let $H_0 : \mu = \mu_0$ and $H_A : \mu \neq \mu_0$

Find the large sample behaviour of the test statistic

$$T_{PC} = n \sum_{k=1}^{K} \frac{\langle \hat{\mu} - \mu_0, \hat{\phi}_k \rangle^2}{\hat{\lambda}_k}$$

under the null and alternative hypotheses. Here $\hat{\mu}$, $\hat{\phi}_k$, $\hat{\lambda}_k$ are the sample versions of μ, ϕ_k, λ_k, respectively, and $K \in \mathbb{N}$.

[You may assume that the first $K + 1$ eigenvalues of covariance operators of X are distinct, and you may also use without proof the convergence properties of eigenvalues, eigenfunctions and any version of the central limit theorem]

(b) Assume $\mu \neq \mu_0$. Define $C^*(\cdot) = C_X(\cdot) + \langle \mu - \mu_0, \cdot \rangle (\mu - \mu_0)$, and let $\{\lambda_k^*, \phi_k^*\}_{k=1}^{\infty}$ be the eigenvalues/eigenfunctions of C^*, and assume $\lambda_1^* > \lambda_2^* > \ldots$.

Show that if, for some k, $1 \leqslant k \leqslant K$, $\langle \mu - \mu_0, \phi_k \rangle \neq 0$, then $\exists l, 1 \leqslant l \leqslant K$, such that $\langle \mu - \mu_0, \phi_l^* \rangle \neq 0$

(c) Again assume $\mu \neq \mu_0$. Let $(\mu - \mu_0) = b\phi_{K+1}$, for some $b \in \mathbb{R}, b \neq 0$. Explain why a test for the hypothesis of part (a) based on the first K eigenfunctions of C^*, as defined in part (b), might be preferable to a test based on the first K eigenfunctions of C_X.

[There is no need to construct the test for part (c)].

3 Let C_1 and C_2 be covariance operators on a separable Hilbert Space

(a) Define the square-root distance $d_R(C_1, C_2)$ and Procrustes distance $d_P(C_1, C_2)$.

(b) Prove the Procrustes distance is given by

$$d_P^2(C_1, C_2) = \|L_1\|_{HS}^2 + \|L_2\|_{HS}^2 - 2\sum_{k=1}^{\infty} \sigma_k$$

where $C_i = L_i L_i^*$, L_i^* is the adjoint of L, σ_k are the singular values of $L_2^* L_1$, and where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm.

(c) Assume that there exists sequences $L_i^{(p)}$ such that $L_i^{(p)}(L_i^{(p)})^* = C_i^{(p)}$ and $L_i^{(p)} \to L_i$ in the Hilbert-Schmidt norm as $p \to \infty$ where $L_i L_i^* = C_i$. Prove that the Procrustes distance

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

(d) Find a pair C_1 and C_2, $C_1 \neq C_2$, where $d_P(C_1, C_2) = d_R(C_1, C_2)$.

4 Let $X, X_1, \ldots, X_n$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|X\|^4 \leqslant \infty$, $\mathbb{E}X = 0$. Let $\epsilon, \epsilon_1, \ldots, \epsilon_n$ be i.i.d. random elements of $L^2[0,1]$ such that $\mathbb{E}\|\epsilon\|^4 \leqslant \infty$, $\mathbb{E}(\epsilon) = 0$, and let $X, X_1, \ldots, X_n$ be mutually independent of $\epsilon, \epsilon_1, \ldots, \epsilon_n$. For each i, let (Y_i, X_i, ϵ_i) follow the same population model as (Y, X, ϵ) namely,

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

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

Let $H_0 : \beta = 0$ and $H_A : \beta \neq 0$.

Define a finite-dimensional statistic, based on the first K eigenfunctions of the covariance operator of X and the first L eigenfunctions of the covariance operator of Y to test the null hypothesis. Determine the asymptotic properties, as $n \to \infty$, of the test under the null hypotheses.

[You may assume that the eigenvalues of covariance operators of X and Y are all distinct, and you may also use without proof the convergence properties of eigenvalues, eigenfunctions and any version of the central limit theorem].

END OF PAPER