Fréchet-differentiation of functions of operators with application to testing the equality of two covariance operators

Xiaoyi Ji¹, Frits H Ruymgaart²

Department of Mathematics & Statistics, Texas Tech University
E-mail: xiao.ji@ttu.edu, h.ruymgaart@ttu.edu

Abstract. It is well-known that the sample covariance operator converges in distribution in the Hilbert space of Hilbert-Schmidt operators, and that this result entails the asymptotic distribution of simple eigenvalues and corresponding eigenvectors. Several estimators and test statistics for the analysis of functional data require the asymptotic distribution of eigenvalues and eigenvectors of certain functions of sample covariance operators. To obtain such a result, it turns out that the asymptotic distribution of such a function of the sample covariance operator is a prerequisite. In this paper we briefly review the Fréchet-derivative of functions of operators and an ensuing delta-method to solve this problem. The results are applied to obtain the asymptotic distribution of a statistic for testing the equality of two covariance operators.

1. Introduction

If, in a suitable framework, the asymptotic distribution of a random operator is known, from this basic result the asymptotic distribution of its largest eigenvalue and corresponding eigenvector can be derived. See, for instance, Watson (1983) for matrices and Dauxois et al. (1982) for operators on Hilbert spaces. In several practical situations, however, the asymptotic distribution of the largest eigenvalue and corresponding eigenvector is required not of the random operator itself, but of an analytic function of this operator (in the sense of functional calculus; see Lax (2002)). By the same token, the asymptotic distribution of this function of the operator would provide the main tool to solve that problem.

A first example is provided by what is generally considered to be the prototype of an inverse problem, viz, the practical solution of an integral equation of the first kind, where the output is observed with error. Regularization of the inverse of the integral operator is usually required. In recent econometric studies (Hall & Horowitz (2005), Florens (2003)) the situation is considered where the integral operator itself is unknown, but estimable from the data. The actual recovery of the input is then based on a regularized inverse of this random operator, and a regularized inverse of Tikhonov type would, indeed, be an analytic function of the random operator.

A second example is the estimation of functional canonical correlations (Leurgans et al. (1993), He et al. (2002)). If suitable regularization is applied such an estimated canonical correlation is the largest eigenvalue of a product of operators containing Tikhonov inverses of the
square root of sample covariance operators, which is again an analytic function of that operator. A major step towards the asymptotic distribution of the estimated canonical correlation is to find the asymptotic distribution of the Tikhonov inverse of the square root of a sample covariance operator (Cupidon et al. (2006, a, b)).

The third example, to be considered here, is also in the area of functional data analysis and concerns testing the hypothesis of equality of two covariance operators. In Section 3 a suitable test statistic for this problem will be derived. This statistic turns out to be the maximum eigenvalue of a product of certain analytic functions of sample covariance operators, containing once more an analytic function of such an operator.

The asymptotic distribution of the sample covariance operator and its ramifications is well-known and can be found in Dauxois et al. (1982). In order to find the required asymptotic distribution of an analytic function of the sample covariance operator, a kind of delta-method is needed (Cupidon et al. (2006 b)); this delta-method can be obtained via the Fréchet-derivative of analytic functions of bounded operators (Gilliam et al. (2007)). The Fréchet-derivative is taken tangentially to the space of all bounded operators, so that the increment does not have to commute with the operator at which the derivative is evaluated. This freedom is important in statistical applications, and leads to a more complicated expression of the derivative than when the increment and the operator would commute. These results are briefly reviewed in Section 2. The basic asymptotics for the test statistic are derived in Section 4, and an asymptotic size-α test is proposed in Sectoin 5. The paper is concluded with some remarks in Section 6.

2. Review of covariance operators

Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space, \( \mathbb{H} \) an infinite dimensional, separable Hilbert space with inner product \( \langle \cdot, \cdot \rangle \) and \( \sigma \)-field of Borel sets \( \mathcal{B}_\mathbb{H} \), and \( X : \Omega \rightarrow \mathbb{H} \) an \( (\mathcal{F}, \mathcal{B}_\mathbb{H}) \)-measurable random element in \( \mathbb{H} \) with \( \mathbb{E} \|X\|^4 < \infty \). The mean \( \mu \) of \( X \) is uniquely determined by the relation

\[
\mathbb{E} \langle f, X \rangle = \langle f, \mu \rangle \quad \forall f \in \mathbb{H},
\]

(2.1)

Let \( \mathcal{L}(HS) \) denote the Hilbert space of all Hilbert-Schmidt operators mapping \( \mathbb{H} \) into itself, equipped with the inner product

\[
\langle T, U \rangle_{HS} = \sum_{k=1}^{\infty} \langle Te_k, Ue_k \rangle, \quad T, U \in \mathcal{L}(HS),
\]

(2.2)

where \( e_1, e_2, \cdots \) is an orthonormal basis of \( \mathbb{H} \). It is well-known that this inner product does not depend on the choice of basis (Lax (2002)). Under the present conditions \( X - \mu \otimes (X - \mu) \) is a random element of \( \mathcal{L}(HS) \) with mean \( S \), the covariance operator of \( X \), uniquely determined by

\[
\mathbb{E} \langle T, (X - \mu) \otimes (X - \mu) \rangle_{HS} = \langle T, S \rangle_{HS} \quad \forall T \in \mathcal{L}(HS).
\]

(2.3)

Equivalently, \( S \) is uniquely determined by

\[
\mathbb{E} \langle f, X - \mu \rangle \langle X - \mu, g \rangle = \langle f, Sg \rangle \quad \forall f, g \in \mathbb{H};
\]

(2.4)

see, for instance, Laha & Rohatgi (1979).

The operator \( S \) is nonnegative Hermitian and even of trace class (and consequently in \( \mathcal{L}(HS) \)). We will assume that \( S \) is strictly positive. In this case \( S \) can be written as

\[
S = \sum_{k=1}^{\infty} \sigma_k P_k
\]

(2.5)
where the $\sigma_k$ are the distinct eigenvalues satisfying
\[ \sigma_1 > \sigma_2 > \ldots \downarrow 0, \] (2.6)
and where the $P_k$ are the corresponding eigenprojections.

Next, let $X_1, \ldots, X_n$ be independent copies of $X$, and define the sample mean
\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \] (2.7)
and the sample covariance operator
\[ \hat{S} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}) \otimes (X_i - \bar{X}). \] (2.8)

The inner product in $L(HS)$ is introduced in (2.2). Let the tensor product in $L(HS)$ be denoted by $\otimes_{HS}$, and let us introduce
\[ S_{HS} = \mathbb{E}\{(X - \mu) \otimes (X - \mu) - S\} \otimes_{HS} \{(X - \mu) \otimes (X - \mu) - S\}, \] (2.9)
the covariance operator of $(X - \mu) \otimes (X - \mu)$. The central limit theorem in separable Hilbert spaces (see, for instance, Laha & Rohetgi (1979)) entails the following result that can be found in Dauxois et al. (1982).

**Theorem 2.1.** There exists a Gaussian operator $G$ in $L(HS)$ with mean 0 and covariance operator $S_{HS}$ in (2.9), such that
\[ \sqrt{n}(\hat{S} - S) \xrightarrow{d} G, \text{ as } n \to \infty, \text{ in } L(HS). \] (2.10)

In the sequel a “delta-method” for sample covariance operators will be needed. A mathematical tool from which this method can be derived is the Fréchet-derivative of an analytic function of an operator $T \in \mathcal{L}$, tangentially to $\mathcal{L}$, where $\mathcal{L}$ is the Banach space of all bounded operators on $H$. Let $\sigma(T)$ be the closed bounded spectrum of $T$, $\Omega$ an open region in the complex plane $\mathbb{C}$ with smooth boundary $\partial\Omega$, such that
\[ \sigma(T) \subset \Omega, \text{ distance } (\partial\Omega, \sigma(T)) > 0, \] (2.11)
Furthermore, let $D \supset \bar{\Omega}$ be an open neighborhood of $\bar{\Omega}$, and
\[ \varphi : D \to \mathcal{C}, \text{ analytic.} \] (2.12)

The resolvent $R(z) = (zI - T)^{-1}$ is analytic for $z \in \rho(T) = [\sigma(T)]^c$, the open resolvent set of $T$. According to the usual functional calculus we have
\[ \varphi(T) = \frac{1}{2\pi i} \oint_{\partial\Omega} \varphi(z) R(z) dz. \] (2.13)
In the special case where $T = S$ with expansion (2.5), relation (2.13) reduces to
\[ \varphi(S) = \sum_{k=1}^{\infty} \varphi(\sigma_k) P_k. \] (2.14)
For the above see, for instance, Dunford & Schwarz (1957) or Lax (2002).

Next, let $\tilde{T} = T + \Pi$, for $\Pi \in \mathcal{L}$, where we do not assume that $T$ and $\Pi$ commute. Let $\| \cdot \|_{\mathcal{L}}$ denote the ordinary operator norm on $\mathcal{L}$. For $\| \Pi \|_{\mathcal{L}}$ sufficiently small we also have $\sigma(\tilde{T}) \subset \Omega$ and $\varphi(\tilde{T})$ can be defined similar to $\varphi(T)$ in (2.13). The following result can be found in Gilliam et al. (2007).

**Theorem 2.2.** We have

$$\varphi(T + \Pi) = \varphi(T) + \varphi_T' \Pi + O(\| \Pi \|_{\mathcal{L}}^2),$$  

(2.15)

as $\| \Pi \| \to 0$, where

$$\varphi_T' \Pi = \frac{1}{2\pi i} \oint_{\partial \Omega} \varphi(z)R(z)\Pi R(z)dz.$$  

(2.16)

In the special case where $T = S$ with expansion (2.5), relation (2.16) reduces to

$$\varphi_S' \Pi = \sum_j \varphi'(\sigma_j)P_j\Pi P_j + \sum_{j \neq k} \frac{\varphi(\sigma_k) - \varphi(\sigma_j)}{\sigma_k - \sigma_j}P_j\Pi P_k.$$  

(2.17)

The delta-method can now be obtained by exploiting Theorem 2.2 with $T = S$ and $\Pi = \widehat{S} - S$. This yields the following result (see Cupidon et al. (2006, b)).

**Theorem 2.3.** With $\mathcal{G}$ as in (2.10), we have

$$\sqrt{n}\{\varphi(\widehat{S}) - \varphi(S)\} \xrightarrow{d} \varphi_S' \mathcal{G}, \text{ as } n \to \infty, \text{ in } \mathcal{L}_{HS},$$  

(2.18)

where $\varphi_S'$ is given in (2.17).

3. The testing problem and test statistic

Let $X_{j,1}, X_{j,2}, \ldots, X_{j,n(j)}$, be iid random elements in $\mathbb{H}$ such that $\mathbb{E}\|X_{j,i}\|^4 < \infty$, having mean $\mu_j$ and a covariance operator $S_j$, for $j = 1, 2$. We want to test the null hypothesis

$$H_0 : S_1 = S_2 = S > O, \text{ versus } H_1 : S_1 > S_2 > O,$$  

(3.1)

where $O$ is the zero operator and the usual ordering of nonnegative Hermitian operators is used.

Let $\hat{S}_j$ respectively $\hat{S}_j$ be the empirical mean and covariance operator of the $j-th$ sample as defined in general in (2.7) and (2.8) with $n = n_j$. To obtain a test statistic, let us apply Roy’s (1953) union-intersection principle. For this purpose choose $f \in \mathbb{H}$, $f \neq 0$, and consider

$$\mu_j(f) = \langle f, \mu_j \rangle, \quad s_j^2(f) = \langle f, S_j f \rangle.$$  

(3.2)

Note that the real valued random variables

$$X_{j,i}(f) = \langle f, X_{j,i} \rangle, \quad i = 1, 2, \ldots, n(j), \quad j = 1, 2,$$  

(3.3)

are iid with mean $\mu_j(f)$ and variance $s_j^2(f)$. In order to test the hypothesis

$$H_0(f) : s_1^2(f) = s_2^2(f) > 0, \text{ versus } H_1(f) : s_1^2(f) > s_2^2(f) > 0,$$  

(3.4)
we may use the statistic
\[ T(f) = \frac{\hat{s}_1^2(f)}{\hat{s}_2^2(f)}, \quad \hat{s}_j^2(f) = \langle f, \hat{S}_j f \rangle. \] (3.5)

In the case where the samples are from normal distributions this statistic would be the likelihood ratio test statistic.

Because \( H_0 = \bigcap_{f \neq 0} H_0(f) \), a good test statistic for \( H_0 \) seems to be \( \sup_{f \neq 0} T(f) \). This works out all right in the Euclidean case, but in the present situation with data in an infinite dimensional Hilbert space, the sup cannot be used without modification because the operator \( \hat{S}_2 \) has finite dimensional kernel (Riesz & Sz-Nagy(1990)).

This entails a degeneracy that can be remedied by regularization. Let us replace \( \hat{S}_2 \) with the operator \( \epsilon I + \hat{S}_2 \), where \( \epsilon > 0 \) is a fixed but arbitrary regularization parameter, and \( I \) is the identity operator. The modified test statistic that will be employed is
\[ T_\epsilon = \max_{f \neq 0} \frac{\langle f, \hat{S}_1 f \rangle}{\langle f, (\epsilon I + \hat{S}_2) f \rangle} \]
where
\[ \hat{R}_\epsilon = (\epsilon I + \hat{S}_2)^{-1/2} \hat{S}_1 (\epsilon I + \hat{S}_2)^{-1/2}, \] (3.7)

The population version of \( R_\epsilon \) is
\[ R_\epsilon = (\epsilon I + S_2)^{-1/2} S_1 (\epsilon I + S_2)^{-1/2}, \] (3.8)

**Remark 3.1.** Because \( S_1 \) is trace class and the regularized inverse \( (\epsilon I + S_2)^{-1/2} \) is bounded, the operator \( R_\epsilon \) is still trace class and \( R_\epsilon \in \mathcal{L}(HS) \); see, for instance Lax (2002). A similar remark holds true for \( \hat{R}_\epsilon \).

Under \( H_0 \) (3.8) simplifies to
\[ R_\epsilon = (\epsilon I + S)^{-1/2} S (\epsilon I + S)^{-1/2} \]
\[ = \sum_{k=1}^{\infty} \frac{\sigma_k}{\epsilon + \sigma_k} P_k, \] (3.9)
in view of (2.5). It follows that under \( H_0 \) the largest eigenvalue of \( R_\epsilon \) equals
\[ \| R_\epsilon \| = \frac{\sigma_1}{\epsilon + \sigma_1} = \frac{\| S \|}{\epsilon + \| S \|}. \] (3.10)

Within the present asymptotic framework, it seems appropriate to employ a suitably scaled version of the centered variable
\[ \| \hat{R}_\epsilon \| - \| R_\epsilon \|, \] (3.11)
as a test statistic. It should be noted, however, that even under \( H_0 \) the quantity \( \| R_\epsilon \| \) is not known because \( S \) is unknown, as we see from (3.10). At least three modifications could be contemplated.
First of all, in (3.6) one might replace \( \hat{S}_1 \) with \((\epsilon I + \hat{S}_1)\) and deal with \( \hat{R}_e' = (\epsilon I + \hat{S}_2)^{-1/2}(\epsilon I + \hat{S}_1)(\epsilon I + \hat{S}_2)^{-1/2} \) rather than with \( \hat{R}_e \) and \( R_e \). Then under \( H_0 \) we would have \( R_e' = I \) and \( \|R_e'\| = 1 \) known. A significant disadvantage, however, is that \( R_e' \) and \( \hat{R}_e' \) are still bounded but no longer compact, let alone Hilbert-Schmidt or trace class. This means that the convergence in distribution of \( \hat{R}_e' \) cannot be studied in the Hilbert space \( \mathcal{L}(HS) \).

A second modification could be to estimate \( \|R_e\| \) in (3.10) directly by \( \|\hat{S}||/(\epsilon + ||\hat{S}||) \) where

\[
\hat{S} = \frac{n(1)}{n} \hat{S}_1 + \frac{n(2)}{n} \hat{S}_2 = n(1) + n(2).
\]

We might now employ the test statistic

\[
\|\hat{R}_e\| - \frac{||\hat{S}||}{\epsilon + ||\hat{S}||}.
\]

Its seems that its asymptotics are rather complicated.

A third possibility, that will be adopted here, is to modify the null hypothesis somewhat. This modified hypothesis will be described in Section 4, and a test procedure will be given in Section 5.

4. Basic asymptotics under the null hypothesis

It can be seen, for instance from Watson (1983) for matrices, and Dauxois et al. (1982) or Gillian et al. (2007) for operators on Hilbert spaces, that if under suitable regularity conditions a random operator converges in distribution in \( \mathcal{L}(HS) \), then the asymptotic distribution of its simple eigenvalues and corresponding eigenvectors follows as a corollary. Therefore our first concern is to establish the convergence in distribution of \( \sqrt{n}(\hat{R}_e - R_e) \).

Let us write \( n = n(1) + n(2) \), assume that

\[
\lim_{n \to \infty} \frac{n(j)}{n} = \lambda_j \in (0, 1), \quad j = 1, 2,
\]

and introduce the function

\[
\varphi_\epsilon(z) = \frac{1}{\sqrt{\epsilon + z}}, \quad z \in \mathbb{C} \setminus \{ -\epsilon \}.
\]

This function satisfies all the conditions for application of Theorem 2.3. The functional derivative evaluated at \( S \) for this function equals

\[
\varphi'_{\epsilon,S}\Pi = -\frac{1}{2} \sum_j \frac{1}{(\epsilon + \sigma_j)^{3/2}} P_j \Pi P_j
\]

\[
+ \sum_{j \neq k} \sum \frac{(\epsilon + \sigma_j)^{1/2} - (\epsilon + \sigma_k)^{1/2}}{(\sigma_k - \sigma_j)(\epsilon + \sigma_j)^{1/2} (\epsilon + \sigma_k)^{1/2}} P_j \Pi P_k.
\]

The modified null hypothesis to be henceforth considered is

\[
H_{0,C} : S_1 = S_2 = S > 0, \quad \|S\| \leq C \in (0, \infty),
\]

the largest eigenvalue \( \sigma_1 \) of \( S \) in (2.5) is simple,

so that \( P_1 = p_1 \otimes p_1 \) for some \( p_1 \in \mathbb{H} \) with \( \|p_1\| = 1 \).

\[
(4.4)
\]
Let us write $S_{HS,j}$ for the operator in (2.9) with $X$ replaced by $X_{j,1}$ and $\mu$ by $\mu_j$. The following two lemmas are immediate.

**Lemma 4.1.** Under $H_{0,C}$ the largest eigenvalue $\sigma_1/(\epsilon + \sigma_1)$ of $R_\epsilon$ in (3.9) is simple, and has eigenprojection $p_1 \otimes p_1$.

**Lemma 4.2.** Under $H_{0,C}$ we have, for $j = 1, 2$,

$$
\sqrt{n}(\hat{S}_j - S_j) \xrightarrow{d} G_j, \text{ as } n \to \infty, \text{ in } \mathcal{L}(HS),
$$

$$
\sqrt{n}\{\varphi_\epsilon(\hat{S}_j) - \varphi_\epsilon(S_j)\} \xrightarrow{d} \varphi_\epsilon' G_j, \text{ as } n \to \infty, \text{ in } \mathcal{L}(HS),
$$

where the $G_j$ are zero mean Gaussian random elements with $G_1 \perp G_2$, $\mathbb{E}G_j \otimes HS G_j = \frac{1}{\lambda_j} S_{HS,j}$.

To simplify the notation for the expressions that will play a role below, let us introduce the function

$$
\psi_\epsilon(z) = \frac{z}{\sqrt{\epsilon + z}}, \text{ } z \in \mathbb{C} \setminus \{-\epsilon\}.
$$

and note that

$$
\psi_\epsilon(S) = S\varphi_\epsilon(S) = \sum_{k=1}^{\infty} \frac{\sigma_k}{\sqrt{\epsilon + \sigma_k}} P_k.
$$

Furthermore, in the notation (4.6) let us write

$$
\mathcal{R}_1 = \varphi_\epsilon(S) G_1 \varphi_\epsilon(S),
$$

$$
\mathcal{R}_2 = \psi_\epsilon(S) (\varphi_\epsilon' G_2).
$$

Both $\mathcal{R}_1$ and $\mathcal{R}_2$ are zero mean elements of $\mathcal{L}(HS)$, $\mathcal{R}_1$ is Hermitian, each factor of $\mathcal{R}_2$ is Hermitian, and

$$
\mathcal{R}_1 \perp \mathcal{R}_2.
$$

Finally, let us write

$$
\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_2^*,
$$

so that $\mathcal{R}$ is Hermitian.

**Theorem 4.1.** Under $H_{0,C}$ we have

$$
\sqrt{n}(\hat{R}_\epsilon - R_\epsilon) \xrightarrow{d} \mathcal{R}, \text{ as } n \to \infty, \text{ in } \mathcal{L}(HS).
$$

Proof. The left-hand side in (4.14) can be decomposed into (cf. (3.7) and (3.9))

$$
\sqrt{n}(\hat{R}_\epsilon - R_\epsilon) = \varphi_\epsilon(S) \sqrt{n}(\hat{S}_1 - S_1) \varphi_\epsilon(\hat{S}_2) + \varphi_\epsilon(S) S \sqrt{n}\{\varphi_\epsilon(\hat{S}_2) - \varphi_\epsilon(S)\} + \sqrt{n}\{\varphi_\epsilon(\hat{S}_2) - \varphi_\epsilon(S)\} \hat{S}_1 \varphi_\epsilon(\hat{S}_2).
$$

The first term on the right in (4.15) converges in distribution to $\mathcal{R}_1$, the second to $\mathcal{R}_2$, and the third to $\mathcal{R}_2^*$. 

7
Now we are in a position to apply the principle mentioned in the beginning of this section, and obtain the asymptotic distribution of the difference of the largest eigenvalue of \( \hat{R} \) and of \( R \). In particular Lemma 4.1 guarantees that in the present case all conditions are fulfilled to proceed in a way similar to that in Proposition 8 of Dauxois et al. (1982) or Theorem 4.2 of Cupidon et al. (2006, b), or to apply Corollary 3.3 in Gilliam et al. (2007). This yields at once the following result.

**Theorem 4.2.** Under the null hypothesis we have
\[
\sqrt{n}(\|\hat{R}\| - \|R\|) \xrightarrow{d} N(0, \nu^2), \quad \text{as } n \rightarrow \infty, \tag{4.16}
\]
where, with \( p_1 \) as in Assumption 4.1,
\[
\nu^2 = E(\langle R_{p_1}, p_1 \rangle)^2. \tag{4.17}
\]

In practice the variance \( \nu^2 \) is unknown. In order to obtain a useful test statistic rescaling with a consistent estimator of \( \nu = \sqrt{\nu^2} \) will be needed. In the next section the expression for \( \nu^2 \) will be made more explicit by substituting the expression for \( R \), and a suitable estimator will then be given.

5. **An asymptotic size-\( \alpha \) test**

In this section we will establish the main result of this paper, viz. the existence of an asymptotic size-\( \alpha \) test for the null hypothesis \( H_{0,C} \) in (4.4), The only remaining problem is the estimation of \( \nu^2 \) in (4.17). Because \( E\langle R_j \rangle = 0 \) it follows from (4.12) and (4.13) that
\[
\nu^2 = E(\langle R_{p_1}, p_1 \rangle)^2 + 4E(\langle R_{2p_1}, p_1 \rangle)^2. \tag{5.1}
\]

Using (4.10), (4.11) we obtain
\[
\nu^2 = E(\langle \varphi(S)G_1\varphi(S)p_1, p_1 \rangle)^2 + 4E(\langle \psi(S)(\varphi'G_2)p_1, p_1 \rangle)^2. \tag{5.2}
\]

For brevity, Let us write (cf. (4.3))
\[
Q = \sum_{j=2}^{\infty} \frac{(\epsilon + \sigma_j)^{1/2} - (\epsilon + \sigma_1)^{1/2}}{(\sigma_1 - \sigma_j)(\epsilon + \sigma_j)^{1/2}(\epsilon + \sigma_1)^{1/2}} P_j. \tag{5.3}
\]

Exploiting the fact that in each term of (5.2) every factor is Hermitian, that \( P_jp_1 = \delta_{j1}p_1 \), and that \( Qp_1 = 0 \), we arrive at
\[
\nu^2 = \frac{1}{(\epsilon + \sigma_1)^{1/2}} E(\langle G_1p_1, p_1 \rangle)^2 \tag{5.4}
\]
\[
+ \frac{4 - \sigma_1^2}{\epsilon + \sigma_1} E(\frac{1}{\epsilon + \sigma_1)} G_2p_1 + QG_2p_1, p_1 \rangle \]
\[
= \gamma_1^2 \delta_1^2 + \gamma_2^2 \delta_2^2. \tag{5.5}
\]

where
\[
\gamma_1^2 = \frac{1}{(\epsilon + \sigma_1)^{1/2}}, \quad \gamma_2^2 = \frac{4\sigma_1^2}{(\epsilon + \sigma_1)^{1/4}}.
\]
\[ \delta_j^2 = \frac{1}{\lambda_j} \langle p_1 \otimes p_1, S_{HS,j} p_1 \otimes p_1 \rangle_{HS} = \frac{1}{\lambda_j} \langle p_1, (S_{HS,j} p_1 \otimes p_1) p_1 \rangle. \] (5.6)

To estimate \( \upsilon_1^2 \), we can estimate \( \sigma_1 \) by \( |\hat{S}| \), where \( \hat{S} \) is defined in (3.12). Moreover, because the eigenvalue \( \sigma_1 \) of \( S \) is simple (by assumption) with eigenvector \( p_1 \), according to theory, the largest eigenvalue \( |\hat{S}| \) of \( \hat{S} \) is also simple with high probability for large \( n \); the corresponding eigenvector \( \hat{p}_1 \) is an estimator of \( p_1 \). Finally, \( S_{HS,j} \) can be estimated by

\[ \hat{S}_{HS,j} = \frac{1}{n(j)} \sum_{i=1}^{n(j)} \{(X_{j,i} - \bar{X}_j) \otimes (X_{j,i} - \bar{X}_j) - \hat{S}_j\} \otimes H_S \{(X_{j,i} - \bar{X}_j) \otimes (X_{j,i} - \bar{X}_j) - \hat{S}_j\}, \] (5.7)

and \( \lambda_j \) replaced with \( n(j)/n \). Substitution of the above in (5.4) yields a consistent estimator \( \hat{\upsilon}_1^2 \) of \( \upsilon_1^2 \).

**Theorem 5.1.** An asymptotic size-\( \alpha \) test for testing the null hypothesis \( H_{0,C} : S_1 = S_2 = S > 0, \|S\| \leq C \in (0, \infty) \), versus \( H_1 : S_1 > S_2 > 0 \) is obtained by rejecting the null hypothesis when

\[ T_{\epsilon,C} = \sqrt{n} \frac{\|\hat{R}_\epsilon\| - \frac{C}{\epsilon + C}}{\hat{\upsilon}_\epsilon} > \Phi^{-1}(1 - \alpha), \] (5.8)

0 < \( \alpha \) < 1, where \( \Phi \) is the standard normal cdf.

Proof. The proof is entirely routine. For given \( S \in H_{0,C} \) and corresponding \( R_\epsilon \) as in (3.9) we have

\[ \mathbb{P}\{T_{\epsilon,C} > \Phi^{-1}(1 - \alpha)\} = \mathbb{P}\{\sqrt{n} \frac{\|\hat{R}_\epsilon\| - \|R_\epsilon\|}{\hat{\upsilon}_\epsilon} > \Phi^{-1}(1 - \alpha) + \sqrt{n} \frac{\frac{C}{\epsilon + C}}{\hat{\upsilon}_\epsilon} \}. \] (5.9)

Relation (4.14) and the fact that \( \hat{\upsilon}_\epsilon \) is a consistent estimator of \( \upsilon_\epsilon \) entails that \( \sqrt{n} \frac{\|\hat{R}_\epsilon\| - \|R_\epsilon\|}{\hat{\upsilon}_\epsilon} \overset{d}{\rightarrow} N(0,1) \). If \( \|S\| < C \), then \( \|R_\epsilon\| < \frac{C}{\epsilon + C} \), and it follows that

\[ \mathbb{P}\{T_{\epsilon,C} > \Phi^{-1}(1 - \alpha)\} \to 0, \text{ as } n \to \infty. \] (5.10)

If \( \|S\| = C \), then \( \|R_\epsilon\| = \frac{C}{\epsilon + C} \), and we obtain

\[ \mathbb{P}\{T_{\epsilon,C} > \Phi^{-1}(1 - \alpha)\} \to 1 - \Phi(\Phi^{-1}(1 - \alpha)) = \alpha, \text{ as } n \to \infty. \] (5.11)

and the theorem follows.

6. Concluding remarks

Once the test procedure has been established, consistency and asymptotic power are in principle routine extensions. The consistency is based on the fact that in a situation where \( S_1 > S_2 > 0 \) a result like (4.16) still remains true albeit with an asymptotic variance \( \upsilon_\epsilon^2 \), say, different from \( \upsilon_1^2 \). This is because (4.5) and (4.6) remain true with the limiting random elements \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) still independent, but no longer related in distribution to \( \mathcal{G} \) under the null hypothesis. Consistency should then follow for alternatives with

\[ \|R_\epsilon\| > \frac{C}{\epsilon + C}, \] (6.1)
as (5.9) suggests.

The asymptotic power could be computed for local alternatives given by

\[
\begin{align*}
S_1 &= S_{1n} = S + \frac{t}{\sqrt{n}} S_0, \quad t > 0, \\
S_2 &= S,
\end{align*}
\]  

(6.2)

for some covariance operator \( S_0 > 0 \). A slight generalization of (4.5) to triangular arrays of iid \( (\mu_1, S_{1n}) \) sample elements will be required.

An important question is how the regularization parameter \( \epsilon \) should be chosen. A suitable answer to this question may involve both theory and extensive simulations and is beyond the scope of this paper.

Acknowledgments

1 Research supported by NSF grant DMS-0605167
2 Research supported by The Summer Dissertation Award Program, Texas Tech University

References

[1] Cupidon J, Eubank R, Gilliam D and Ruymgaart F 2006 a Some properties of canonical correlations and variates in infinite dimensions Multivariate Analysis to appear
[2] Cupidon J, Eubank R, Gilliam D and Ruymgaart F 2006 b The delta-method for analytic functions of random operators with application to functional data Bernoulli to appear
[3] Dauxois J, Pousse A and Romain Y 1982 Asymptotic theory for the principle component analysis of a vector random function: some applications to statistical inference Multivariate Analysis 12 136-154
[4] Dunford N and Schwartz J T 1957 Linear Operators Part I: General Theory (New York: Interscience)
[5] Florens J P 2003 Inverse problems and structural econometrics: the example of instrumental variables Advances in Economics and Econometrics: Theory and Applications vol 2, ed Dewatripont M and Hanson L-P et al (Cambridge) pp 284-311
[6] Gilliam D, Hohage T, Ji X and Ruymgaart F 2007 The Fréchet-derivative of an analytic function of a bounded operator with some applications Tech. Rept. Dept. Math. and Statist. Texas Tech University
[7] Hall P and Horowitz J L 2005 Nonparametric methods for inference in the presence of instrumental variables Ann. Statist. 33 2904-2929
[8] He G, Mülller H-G and Wang J L 2002 Functional canonical analysis for square integrable stochastic processes Multivariate Analysis 85 54-77
[9] Laha R G and Rohatgi V K 1979 Probability Theory (New York: Wiley)
[10] Lax P D 2002 Functional Analysis (New York: Wiley-Interscience)
[11] Leurgans S E, Moyeed R A and Silverman B W 1993 Canonical correlation analysis when the data are curves Roy. Statist. Soc. B 55 725-740
[12] Riesz and Sz.-Nagy B 1990 Functional Analysis (New York: Dover)
[13] Roy S N 1953 On a heuristic method of test construction and its use in multivariate analysis Ann. Math. Statist. 24 220-238
[14] Watson G S 1983 Statistics on Spheres (New York: Wiley)