

Research Article

Some Asymptotic Theory for Functional Regression and Classification

Frits Ruymgaart, Jing Wang, Shih-Hsuan Wei, and Li Yu

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA

Correspondence should be addressed to Frits Ruymgaart, h.ruymgaart@ttu.edu

Received 10 October 2011; Accepted 2 November 2011

Academic Editor: Wing Keung Wong

Copyright © 2011 Frits Ruymgaart et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Exploiting an expansion for analytic functions of operators, the asymptotic distribution of an estimator of the functional regression parameter is obtained in a rather simple way; the result is applied to testing linear hypotheses. The expansion is also used to obtain a quick proof for the asymptotic optimality of a functional classification rule, given Gaussian populations.

1. Introduction

Certain functions of the covariance operator (such as the square root of a regularized inverse) are important components of many statistics employed for functional data analysis. If Σ is a covariances operator on a Hilbert space, $\hat{\Sigma}$ a sample analogue of this operator, and φ a function on the complex plane, which is analytic on a domain containing a contour around the spectrum of Σ , a tool of generic importance is the comparison of $\varphi(\hat{\Sigma})$ and $\varphi(\Sigma)$ by means of a Taylor expansion:

$$\varphi(\hat{\Sigma}) = \varphi(\Sigma) + \dot{\varphi}_{\Sigma}(\hat{\Sigma} - \Sigma) + \text{remainder}. \quad (1.1)$$

(It should be noted that $\dot{\varphi}_{\Sigma}$ is not in general equal to $\varphi'(\Sigma)$, where φ' is the numerical derivative of φ ; see also Section 3.) In this paper, two further applications of the approximation in (1.1) will be given, both related to functional regression.

The first application (Section 4) concerns the functional regression estimator itself. Hall and Horowitz [1] have shown that the IMSE of their estimator, based on a Tikhonov type regularized inverse, is rate optimal. In this paper, as a complementary result, the general asymptotic distribution is obtained, with potential application to testing linear hypotheses of

arbitrary finite dimension, mentioned in Cardot et al. [2] as an open problem: these authors concentrate on testing a simple null hypotheses. Cardot et al. [3] establish convergence in probability and almost sure convergence of their estimator which is based on spectral cutoff regularization of the inverse of the sample covariance operator. In the present paper, the covariance structure of the Gaussian limit will be completely specified. The proof turns out to be routine thanks to a “delta-method” for $\varphi(\widehat{\Sigma}) - \varphi(\Sigma)$, which is almost immediate from (1.1).

The second application (Section 5) concerns functional classification, according to a slight modification of a method by Hastie et al. [4], exploiting penalized functional regression. It will be shown that this method is asymptotically optimal (Bayes) when the two populations are represented by equivalent Gaussian distributions with the same covariance operator. The simple proof is based on an upper bound for the norm of $\varphi(\widehat{\Sigma}) - \varphi(\Sigma)$, which follows at once from (1.1).

Let us conclude this section with some comments and further references. The expansion in (1.1) can be found in Gilliam et al. [5], and the ensuing delta method is derived and applied to regularized canonical correlation in Cupidon et al. [6]. For functional canonical correlation see also Eubank and Hsing [7], He et al. [8], and Leurgans et al. [9]. When the perturbation ($\widehat{\Sigma} - \Sigma$ in the present case) commutes with Σ the expansion (1.1) can already be found in Dunford & Schwartz [10, Chapter VII], and the derivative does indeed reduce to the numerical derivative. This condition is fulfilled only in very special cases, for instance, when the random function, whose covariance operator is Σ , is a second order stationary process on the unit interval. In this situation, the eigenfunctions are known and only the eigenvalues are to be estimated. This special case, that will not be considered here, is discussed in Johannes [11] who in particular deals with regression function estimators and their IMSE in Sobolev norms, when the regression is such a stationary process. General information about functional data analysis can be found in the monographs by Ramsay and Silverman [12] and Ferraty and Vieu [13]. Functional time series are considered in Bosq [14]; see also Mas [15].

2. Preliminaries and Introduction of the Models

2.1. Preliminaries

As will be seen in the examples below, it is expedient to consider functional data as elements in an abstract Hilbert space \mathbb{H} of infinite dimension, separable, and over the real numbers. Inner product and norm in \mathbb{H} will be denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X : \Omega \rightarrow \mathbb{H}$ a Hilbert space valued random variable (i.e., measurable with respect to the σ -field of Borel sets $\mathcal{B}_{\mathbb{H}}$ in \mathbb{H}), and $\eta : \Omega \rightarrow \mathbb{R}$ a real valued random variable. For all that follows it will be sufficient to assume that

$$\mathbb{E}\|X\|^4 < \infty, \quad \mathbb{E}\eta^2 < \infty. \quad (2.1)$$

The mean and covariance operator of X will be denoted by

$$\mathbb{E}X = \mu_X, \quad \mathbb{E}(X - \mu_X) \otimes (X - \mu_X) = \Sigma_{X,X}, \quad (2.2)$$

respectively, where $a \otimes b$ is the tensor product in \mathbb{H} . The Riesz representation theorem guarantees that these quantities are uniquely determined by the relations

$$\begin{aligned}\mathbb{E}\langle a, X \rangle &= \langle a, \mu_X \rangle, \quad \forall a \in \mathbb{H}, \\ \mathbb{E}\langle a, X - \mu_X \rangle \langle X - \mu_X, b \rangle &= \langle a, \Sigma_{X,X} b \rangle, \quad \forall a \in \mathbb{H} \quad \forall b \in \mathbb{H},\end{aligned}\tag{2.3}$$

see Laha & Rohatgi [16]. Throughout $\Sigma_{X,X}$ is assumed to be one-to-one.

Let \mathcal{L} denote the Banach space of all bounded linear operators $T : \mathbb{H} \rightarrow \mathbb{H}$ equipped with the norm $\|\cdot\|_{\mathcal{L}}$. An operator $U \in \mathcal{L}$ is called Hilbert-Schmidt if

$$\sum_{k=1}^{\infty} \|Ue_k\|^2 < \infty,\tag{2.4}$$

for any orthonormal basis e_1, e_2, \dots of \mathbb{H} . (The number in (2.4) is in fact independent of the choice of basis.) The subspace $\mathcal{L}_{\text{HS}} \subset \mathcal{L}$ of all Hilbert-Schmidt operators is a Hilbert space in its own right with the inner product

$$\langle U, V \rangle_{\text{HS}} = \sum_{k=1}^{\infty} \langle Ue_k, Ve_k \rangle,\tag{2.5}$$

again independent of the choice of basis. This inner product yields the norm

$$\|U\|_{\text{HS}}^2 = \sum_{k=1}^{\infty} \|Ue_k\|^2,\tag{2.6}$$

which is the number in (2.4). The tensor product for elements $a, b \in \mathbb{H}$ will be denoted by $a \otimes b$, and that for elements $U, V \in \mathcal{L}_{\text{HS}}$ by $U \otimes_{\text{HS}} V$.

The two problems to be considered in this paper both deal with cases where the best linear predictor of η in terms of X is linear:

$$\mathbb{E}(\eta | X) = \alpha + \langle X, f \rangle, \quad \alpha \in \mathbb{R}, \quad f \in \mathbb{H}.\tag{2.7}$$

Just as in the univariate case (Rao [17, Section 4g]), we have the relation

$$\Sigma_{X,X} f = \mathbb{E}(\eta - \mu_{\eta})(X - \eta_X) = \Sigma_{X,\eta}.\tag{2.8}$$

It should be noted that if $\Sigma_{X,X}$ is one-to-one and $\Sigma_{X,\eta}$ in its range, we can solve (2.8) and obtain

$$f = \Sigma_{X,X}^{-1} \Sigma_{X,\eta}.\tag{2.9}$$

Since the underlying distribution is arbitrary, the empirical distribution, given a sample $(X_1, \eta_1), \dots, (X_n, \eta_n)$ of independent copies of (X, η) , can be substituted for it. The

minimization property is now the least squares property, the same formulas are obtained with μ_X , $\Sigma_{X,X}$, μ_η , and $\Sigma_{X,\eta}$ replaced with their estimators

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad (2.10)$$

$$\hat{\Sigma}_{X,X} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) \otimes (X_i - \bar{X}), \quad (2.11)$$

$$\hat{\mu}_\eta = \frac{1}{n} \sum_{i=1}^n \eta_i = \bar{\eta}, \quad (2.12)$$

$$\hat{\Sigma}_{X,\eta} = \frac{1}{n} \sum_{i=1}^n (\eta_i - \bar{\eta}) \times (X_i - \bar{X}). \quad (2.13)$$

Let us next specify the two problems.

2.2. Functional Regression Estimation

The model here is

$$\eta = \alpha + \langle X, f \rangle + \varepsilon, \quad (2.14)$$

where ε is a real valued error variable and the following assumption is satisfied.

Assumption 2.1. The error variable has a finite second moment, and

$$\varepsilon \perp\!\!\!\perp X, \quad \mathbb{E}\varepsilon = 0, \quad \text{Var } \varepsilon = v^2 < \infty. \quad (2.15)$$

Example 2.2. The functional regression model in Hall and Horowitz [1] is essentially obtained by choosing $\mathbb{H} = L^2(0, 1)$, so that the generic observation is given by

$$\eta = \alpha + \int_0^1 X(t)f(t) dt + \varepsilon. \quad (2.16)$$

Example 2.3. Mas and Pumo [18] argue that in the situation of Example 2.2, the derivative X' of X may contain important information and should therefore be included. Hence, these authors suggest to choose for \mathbb{H} the Sobolev space $W^{2,1}(0,1)$ in which case the generic observation satisfies

$$\eta = \alpha + \int_0^1 X(t)f(t)dt + \int_0^1 X'(t)f'(t)dt + \varepsilon. \quad (2.17)$$

Example 2.4. Just as in the univariate case, we have that the model

$$\eta = \alpha + \langle X, f \rangle^2 + \varepsilon, \quad (2.18)$$

quadratic in the inner product of \mathbb{H} , is in fact linear in the inner product of \mathcal{L}_{HS} , because

$$\langle X, f \rangle^2 = \langle X \otimes X, f \otimes f \rangle_{\text{HS}}. \quad (2.19)$$

We will not pursue this example here.

In the infinite dimensional case $\widehat{\Sigma}_{X,X}$ cannot be one-to-one, and in order to estimate f from the sample version of (2.9), a regularized inverse of Tikhonov type will be used, as in Hall & Horowitz [1]. Thus, we arrive at the estimator (see also (2.11) and (2.13))

$$\begin{aligned} \widehat{f}_\delta &= (\delta I + \widehat{\Sigma}_{X,X})^{-1} \left(\frac{1}{n} \sum_{i=1}^n \eta_i (X_i - \bar{X}) \right) \\ &= (\delta I + \widehat{\Sigma}_{X,X})^{-1} \widehat{\Sigma}_{X,\eta}, \quad \text{for some } \delta > 0. \end{aligned} \quad (2.20)$$

Let us also introduce

$$f_\delta = (\delta I + \Sigma)^{-1} \Sigma f, \quad f \in \mathbb{H}. \quad (2.21)$$

In Section 4, the asymptotic distribution of this estimator will be obtained, and the result will be applied to testing.

2.3. Functional Classification

The method discussed here is essentially the one in Hastie et al. [4] and Hastie et al. [19, Sections 4.2 and 4.3]. Let P_1 and P_2 be two probability distributions on $(\mathbb{H}, \mathcal{B}_{\mathbb{H}})$ with means μ_1 and μ_2 and common covariance operator Σ . Consider a random element $(I, X) : \Omega \rightarrow \{1, 2\} \times \mathbb{H}$ with distribution determined by

$$\begin{aligned} \mathbb{P}\{X \in B \mid I = j\} &= P_j(B), \quad B \in \mathcal{B}_{\mathbb{H}}, \\ \mathbb{P}\{I = j\} &= \pi_j \geq 0, \quad \pi_1 + \pi_2 = 1. \end{aligned} \quad (2.22)$$

In this case, the distribution of X is $\pi_1 P_1 + \pi_2 P_2$, with mean

$$\mu_X = \pi_1 \mu_1 + \pi_2 \mu_2, \quad (2.23)$$

and covariance operator

$$\Sigma_{X,X} = \Sigma + \pi_1 \pi_2 (\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2). \quad (2.24)$$

Hastie et al. [19] now introduce the indicator response variables $\eta_j = 1_{\{j\}}(I)$, $j = 1, 2$, and assume that the η_j satisfies (2.7) for $\alpha_j \in \mathbb{R}$ and $f_j \in \mathbb{H}$. Note that

$$\mu_{\eta_j} = \mathbb{E}\eta_j = \pi_j, \quad (2.25)$$

$$\Sigma_{X, \eta_j} = \mathbb{E}\eta_j(X - \mu_X) = \begin{cases} \pi_1\pi_2(\mu_1 - \mu_2), & j = 1, \\ \pi_1\pi_2(\mu_2 - \mu_1), & j = 2. \end{cases} \quad (2.26)$$

Since η_j is Bernoulli, we have, of course, $\mathbb{E}(\eta_j | X) = \mathbb{P}\{I = j | X\}$. Precisely as for matrices (Rao & Toutenburg [20, Theorem A.18]), the inverse of the operator in (2.24) equals

$$\Sigma_{X, X}^{-1} = \Sigma^{-1} - \gamma \cdot \Sigma^{-1}((\mu_1 - \mu_2) \otimes (\mu_1 - \mu_2))\Sigma^{-1}, \quad (2.27)$$

where $\gamma = \pi_1\pi_2 / (1 + \pi_1\pi_2\langle \mu_1 - \mu_2, \Sigma^{-1}(\mu_1 - \mu_2) \rangle)$, provided that the following assumption is satisfied.

Assumption 2.5. The vector $\mu_1 - \mu_2$ lies in the range of Σ , that is,

$$\Sigma^{-1}(\mu_1 - \mu_2) \text{ is well defined.} \quad (2.28)$$

It will also be assumed that

$$\pi_1 = \pi_2 = \frac{1}{2}. \quad (2.29)$$

Assuming (2.28), (2.8) can be solved and yields after some algebra

$$f_j = \Sigma_{X, X}^{-1} \Sigma_{X, \eta_j} = \begin{cases} \gamma \Sigma^{-1}(\mu_1 - \mu_2), & j = 1, \\ \gamma \Sigma^{-1}(\mu_2 - \mu_1), & j = 2. \end{cases} \quad (2.30)$$

If only X , and not I , is observed, the rule in Hastie et al. [19] assigns X to P_1 if and only if

$$\mathbb{E}(\eta_1 | X) > \mathbb{E}(\eta_2 | X), \quad (2.31)$$

$$\iff$$

$$\left\langle X - \mu_X, \Sigma^{-1}(\mu_1 - \mu_2) \right\rangle > \frac{(\pi_2 - \pi_1)}{2\gamma}. \quad (2.32)$$

Because of assumption (2.29), the rule reduces to

$$\left\langle X - \frac{1}{2}(\mu_1 + \mu_2), \Sigma^{-1}(\mu_1 - \mu_2) \right\rangle > 0. \quad (2.33)$$

Hastie et al. [19] claim that in the finite dimensional case, their rule reduces to Fisher's linear discriminant rule and to the usual rule when the distributions are normal. This remains in fact true in the present infinite dimensional case. Let us assume that

$$P_j = \mathcal{G}(\mu_j, \Sigma), \quad j = 1, 2, \quad (2.34)$$

where $\mathcal{G}(\mu, \Sigma)$ denotes a Gaussian distribution with mean μ and covariance operator Σ . It is well known [21–23] that under Assumption 2.5 these Gaussian distributions are equivalent. This is important since there is no “Lebesgue measure” on \mathbb{H} [24]. However, now the densities of P_1 and P_2 with respect to P_1 can be considered; it is well known that

$$\frac{dP_1}{dP_2}(x) = e^{-\langle x - (1/2)(\mu_1 + \mu_2), \Sigma^{-1}(\mu_1 - \mu_2) \rangle}, \quad x \in \mathbb{H}. \quad (2.35)$$

This leads at once to (2.33) as an optimal (Bayes) rule, equal in appearance to the one for the finite dimensional case.

In most practical situations, μ_1 , μ_2 , and Σ are not known, but a training sample $(I_1, X_1), \dots, (I_n, X_n)$ of independent copies of (I, X) is given. Let

$$\bar{X}_j = \frac{1}{n_j} \sum_{i \in \mathcal{J}_j} X_i, \quad \mathcal{J}_j = \{i : I_i = j\}, \quad \#\mathcal{J}_j = n_j, \quad (2.36)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^2 \sum_{i \in \mathcal{J}_j} (X_i - \bar{X}_j) \otimes (X_i - \bar{X}_j), \quad (2.37)$$

and we have (cf. (2.24))

$$\hat{\Sigma}_{X,X} = \hat{\Sigma} + \frac{n_1 n_2}{n} (\bar{X}_1 - \bar{X}_2) \otimes (\bar{X}_1 - \bar{X}_2). \quad (2.38)$$

Once again the operator $\hat{\Sigma}$ (and $\hat{\Sigma}_{X,X}$ for that matter) cannot be one-to-one. In order to obtain an empirical analogue of the rule (2.32), Hastie et al. [4] employ penalized regression, and Hastie et al. [19] also suggest to use a regularized inverse. (The methods are related.) Here the latter method will be used and X will be assigned to P_1 if and only if

$$\left\langle X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2), (\delta I + \hat{\Sigma})^{-1}(\bar{X}_1 - \bar{X}_2) \right\rangle > 0. \quad (2.39)$$

Section 5 is devoted to showing that this rule is asymptotically optimal when Assumption (2.5) is fulfilled.

3. A Review of Some Relevant Operator Theory

It is well known [16] that the covariance operator Σ is nonnegative, Hermitian, of finite trace, and hence Hilbert-Schmidt and therefore compact. The assumption that Σ is one-to-one is

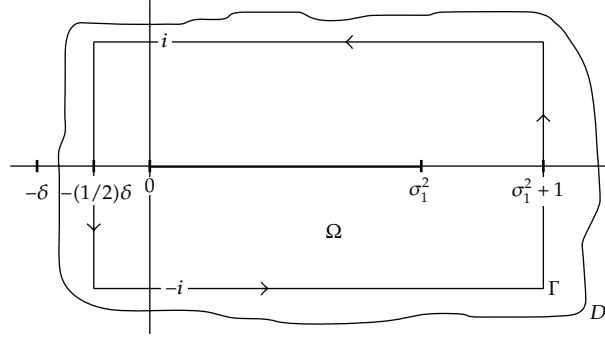


Figure 1

equivalent to assuming that Σ is strictly positive. Consequently, Σ has eigenvalues $\sigma_1^2 > \sigma_2^2 > \dots \downarrow 0$, all of finite multiplicity. If we let P_1, P_2, \dots be the corresponding eigenprojections, so that $\Sigma P_k = \sigma_k^2 P_k$, we have the spectral representation:

$$\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 P_k, \quad \text{with } \sum_{k=1}^{\infty} P_k = I. \quad (3.1)$$

The spectrum of Σ equals $\sigma(\Sigma) = \{0, \sigma_1^2, \sigma_2^2, \dots\} \subset [0, \sigma_1^2]$. Let us introduce a rectangular contour Γ around the spectrum as in Figure 1, where $\delta > 0$ is the regularization parameter in (2.20), and let Ω be the open region enclosed by Γ . Furthermore, let $D \supset (\Omega \cup \Gamma) = \overline{\Omega}$ be an open neighborhood of $\overline{\Omega}$ and suppose that

$$\varphi : D \rightarrow \mathbb{C} \text{ is analytic.} \quad (3.2)$$

We are interested in approximations of $\varphi(\tilde{\Sigma}) = \varphi(\Sigma + \Pi)$, where $\Pi \in \mathcal{L}$ is a perturbation. The application we have in mind arises for $\Pi = \hat{\Pi} = \hat{\Sigma} - \Sigma$ and yields an approximation of $\varphi(\hat{\Sigma})$; see also Watson [25] for the matrix case. Therefore, we will *not* in general assume that Π and Σ will commute. In the special case where X is stationary, as considered in Johannes [11], there exists a simpler estimator Σ of Σ , such that Σ and Π do commute, which results in a simpler theory; see also Remark (4.1).

The resolvent of Σ ,

$$R(z) = (zI - \Sigma)^{-1}, \quad z \in \rho(\Sigma) = [\sigma(\Sigma)]^c, \quad (3.3)$$

is analytic on the resolvent set $\rho(\Sigma)$, and the operator

$$\varphi(\Sigma) = \frac{1}{2\pi i} \oint_{\Gamma} \varphi(z) R(z) dz \quad (3.4)$$

is well defined. For the present operator Σ , as given in (3.1), the resolvent equals more explicitly

$$R(z) = \sum_{k=1}^{\infty} \frac{1}{z - \sigma_k^2} P_k. \quad (3.5)$$

Substitution of (3.5) in (3.4) and application of the Cauchy integral formula yields

$$\varphi(\Sigma) = \sum_{k=1}^{\infty} \varphi(\sigma_k^2) P_k. \quad (3.6)$$

Example 3.1. The two functions

$$\varphi_1(z) = \frac{1}{\delta + z}, \quad \varphi_2(z) = \frac{z}{\delta + z}, \quad z \in \mathbb{C} \setminus \{-\delta\}, \quad (3.7)$$

are analytic on their domain that satisfies the conditions. With the help of these functions we may write (cf. (2.20) and (2.21))

$$\widehat{f}_\delta - f_\delta = (\varphi_1(\Sigma)) \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i (X_i - \bar{X}) \right) + (\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma)) f, \quad f \in \mathbb{H}, \quad (3.8)$$

and (cf. (2.39))

$$(\delta I + \widehat{\Sigma})^{-1} (\bar{X}_1 - \bar{X}_2) = \varphi_1(\widehat{\Sigma}) (\bar{X}_1 - \bar{X}_2). \quad (3.9)$$

Regarding the following brief summary and slight extension of some of the results in [5], we also refer to Dunford & Schwartz [10], Kato [26], and Watson [25]. Henceforth, we will assume that

$$\|\Pi\|_{\mathcal{L}} \leq \frac{\delta}{4}. \quad (3.10)$$

For such perturbations, we have $\sigma(\widetilde{\Sigma}) = \sigma(\Sigma + \Pi) \subset \Omega$, so that the resolvent set of $\widetilde{\Sigma}$ satisfies

$$\rho(\widetilde{\Sigma}) = \rho(\Sigma + \Pi) \supset \Omega^c \supset \Gamma. \quad (3.11)$$

It should also be noted that

$$\|R(z)\|_{\mathcal{L}} = \sup_{k \in \mathbb{N}} \frac{1}{|z - \sigma_k^2|} \leq \frac{2}{\delta} \quad \forall z \in \Omega^c. \quad (3.12)$$

The basic expansion (similar to Watson [25])

$$\tilde{R}(z) = (zI - \tilde{\Sigma})^{-1} = R(z) + \sum_{k=1}^{\infty} R(z)(\Pi R(z))^k, \quad z \in \Omega^c, \quad (3.13)$$

can be written as

$$\tilde{R}(z) = R(z) + R(z)\Pi R(z)(I - \Pi R(z))^{-1}, \quad (3.14)$$

useful for analyzing the error probability for $\delta \rightarrow 0$, as $n \rightarrow \infty$, and also as

$$\tilde{R}(z) = R(z) + R(z)\Pi R(z) + R(z)(\Pi R(z))^2(I - \Pi R(z))^{-1}, \quad (3.15)$$

useful for analyzing the convergence in distribution of the estimators.

Let us decompose the contour Γ into two parts $\Gamma_0 = \{-(1/2)\delta + iy : -1 \leq y \leq 1\}$ and $\Gamma_1 = \Gamma \setminus \Gamma_0$, write $M_\varphi = \max_{z \in \Gamma} |\varphi(z)|$, and observe that (3.10) and (3.12) entail that

$$\left\| (I - \Pi R(z))^{-1} \right\|_{\mathcal{L}} \leq 2, \quad z \in \Omega^c. \quad (3.16)$$

We now have

$$\begin{aligned} & \left\| \frac{1}{2\pi i} \oint_{\Gamma} \varphi(z) R(z) (\Pi R(z))^k (I - \Pi R(z))^{-1} dz \right\|_{\mathcal{L}} \\ & \leq \frac{1}{\pi} M_\varphi \|\Pi\|_{\mathcal{L}}^k \oint_{\Gamma} \|R(z)\|_{\mathcal{L}}^{k+1} dz \\ & \leq \frac{1}{\pi} M_\varphi \|\Pi\|_{\mathcal{L}}^k \left\{ \int_{-1}^{+1} \left(\frac{1}{4} \delta^2 + y^2 \right)^{-(1/2)(k+1)} dy + \left| \int_{\Gamma_1} 1 dz \right| \right\} \\ & \leq \frac{2}{\pi} M_\varphi \|\Pi\|_{\mathcal{L}}^k \left\{ \left(\frac{4}{\delta} \right)^k + 5 + 2\|\Sigma\|_{\mathcal{L}} \right\}, \quad k \in \mathbb{N}. \end{aligned} \quad (3.17)$$

Multiplying both sides by $\varphi(z)$, taking $(1/2\pi i) \oint_{\Gamma}$, and using $0 < C < \infty$ as a generic constant that does not depend on Π or δ , (3.14) and (3.15) yield the following.

Theorem 3.2. *Provided that (3.10) is fulfilled, one has*

$$\|\varphi(\Sigma + \Pi) - \varphi(\Sigma)\|_{\mathcal{L}} \leq CM_\varphi \frac{\|\Pi\|_{\mathcal{L}}}{\delta}, \quad (3.18)$$

$$\|\varphi(\Sigma + \Pi) - \varphi(\Sigma) - \dot{\varphi}_\Sigma \Pi\|_{\mathcal{L}} \leq CM_\varphi \left(\frac{\|\Pi\|_{\mathcal{L}}}{\delta} \right)^2, \quad (3.19)$$

where $\dot{\varphi}_\Sigma : \mathcal{L} \rightarrow \mathcal{L}$ is a bounded operator, given by

$$\dot{\varphi}_\Sigma \Pi = \sum_k \varphi'(\sigma_k^2) P_k \Pi P_k + \sum_{j \neq k} \sum \frac{\varphi(\sigma_k^2) - \varphi(\sigma_j^2)}{\sigma_k^2 - \sigma_j^2} P_j \Pi P_k. \quad (3.20)$$

Remark 3.3. If Σ and Π commute, so will P_k and Π , and $R(z)$ and Π , and the expressions simplify considerably. In particular, (3.20) reduces to

$$\dot{\varphi}_\Sigma \Pi = \left(\sum_k \varphi'(\sigma_k^2) P_k \right) \Pi, \quad (3.21)$$

that is, $\dot{\varphi}_\Sigma = \varphi'(\Sigma)$, where φ' is the numerical derivative of φ , and $\varphi'(\Sigma)$ should be understood in the sense of “functional calculus” as in (3.6). For the commuting case, see Dunford & Schwartz [10].

Let us now present some basic facts that are useful to subsequent statistical applications. Dauxois et al. [27] have shown that there exists a Gaussian random element $\mathcal{G}_\Sigma : \Omega \rightarrow \mathcal{L}_{HS}$, such that $\sqrt{n}(\hat{\Sigma} - \Sigma) \xrightarrow{d} \mathcal{G}_\Sigma$, as $n \rightarrow \infty$, in \mathcal{L}_{HS} . Because the identity map from \mathcal{L}_{HS} in \mathcal{L} is continuous, we may state

$$\sqrt{n}(\hat{\Sigma} - \Sigma) \xrightarrow{d} \mathcal{G}_\Sigma, \quad \text{as } n \rightarrow \infty, \text{ in } \mathcal{L}_{HS} \implies \text{in } \mathcal{L}, \quad (3.22)$$

by the continuous mapping theorem. This entails that

$$\|\hat{\Pi}\|_{\mathcal{L}} = \|\hat{\Sigma} - \Sigma\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{as } n \rightarrow \infty, \quad (3.23)$$

so that condition (3.10) is fulfilled with arbitrary high probability for n sufficiently large. Expansions (3.14) and (3.15) and the resulting inequalities hold true for $\tilde{\Sigma}$ replaced with $\hat{\Sigma}(\omega) = \Sigma + \hat{\Pi}(\omega)$ for $\omega \in \{\|\hat{\Pi}\|_{\mathcal{L}} \leq \delta/4\}$.

Example 3.4. Application to asymptotic distribution theory. In this application $\delta > 0$ will be kept fixed: see also Section 4.1. It is based on the delta method for functions of operators [6] which follows easily from (3.19). In conjunction with (3.22) this yields

$$\sqrt{n}(\varphi_2(\hat{\Sigma}) - \varphi_2(\Sigma)) \xrightarrow{d} \dot{\varphi}_{2,\Sigma} \mathcal{G}_\Sigma, \quad \text{as } n \rightarrow \infty, \text{ in } \mathcal{L}. \quad (3.24)$$

In turn this yields

$$\sqrt{n}(\varphi_2(\hat{\Sigma}) - \varphi_2(\Sigma))f \xrightarrow{d} (\dot{\varphi}_{2,\Sigma} \mathcal{G}_\Sigma)f, \quad \text{as } n \rightarrow \infty, \text{ in } \mathbb{H}, \quad (3.25)$$

for any $f \in \mathbb{H}$, by the continuous mapping theorem. This result will be used in Section 4.

Example 3.5. Application to classification. Here we will let $\delta = \delta(n) \downarrow 0$, as $n \rightarrow \infty$, and write $\varphi_{1,n}(z) = 1/(\delta(n) + z)$ to stress this dependence on sample sizes. Since

$$\max_{z \in \mathbb{I}} |\varphi_{1,n}(z)| \leq \frac{1}{\delta(n)}, \quad (3.26)$$

it is immediate from (3.18) that

$$\left\| \varphi_{1,n}(\widehat{\Sigma}) - \varphi_{1,n}(\Sigma) \right\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\delta^2(n)\sqrt{n}}\right), \quad \text{as } n \rightarrow \infty, \quad (3.27)$$

a result that will be used in Section 5.

4. Asymptotics for the Regression Estimator

4.1. The Asymptotic Distribution

The central limit theorem in Hilbert spaces entails at once

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i (X_i - \mu) \xrightarrow{d} \mathcal{G}_0, \quad \text{as } n \rightarrow \infty, \text{ in } \mathbb{H}, \quad (4.1)$$

where \mathcal{G}_0 is a zero mean Gaussian random variable in \mathbb{H} , and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \{(X_i - \mu) \otimes (X_i - \mu) - \Sigma\} \xrightarrow{d} \mathcal{G}_\Sigma, \quad \text{as } n \rightarrow \infty, \text{ in } \mathcal{L}_{\text{HS}}, \quad (4.2)$$

where \mathcal{G}_Σ is a zero mean random variable in \mathcal{L}_{HS} . These convergence results remain true with μ replaced by \bar{X} and, because $\varepsilon \perp\!\!\!\perp X$ by assumption (2.15), we also have that jointly

$$\begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i (X_i - \bar{X}) \\ \sqrt{n}(\widehat{\Sigma} - \Sigma) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \mathcal{G}_0 \\ \mathcal{G}_\Sigma \end{bmatrix}, \quad \text{as } n \rightarrow \infty, \text{ in } \mathbb{H} \times \mathcal{L}_{\text{HS}}, \quad (4.3)$$

where \mathcal{G}_Σ is the same in (3.22), and

$$\mathcal{G}_0 \perp\!\!\!\perp \mathcal{G}_\Sigma. \quad (4.4)$$

Because the limiting variables are generated by the sums of iid variables on the left in (4.1) and (4.2) we have

$$\begin{aligned}\mathbb{E}\mathcal{G}_0 \otimes \mathcal{G}_0 &= \mathbb{E}\{\varepsilon(X - \mu)\} \otimes \{\varepsilon(X - \mu)\} \\ &= (\mathbb{E}\varepsilon^2)\mathbb{E}(X - \mu) \otimes (X - \mu) \\ &= v^2\Sigma,\end{aligned}\tag{4.5}$$

$$\mathbb{E}\mathcal{G}_{\Sigma} \otimes_{\text{HS}} \mathcal{G}_{\Sigma} = \mathbb{E}\{(X - \mu) \otimes (X - \mu) - \Sigma\} \otimes_{\text{HS}} \{(X - \mu) \otimes (X - \mu) - \Sigma\},\tag{4.6}$$

for the respective covariance structures. These are important to further specify the limiting distribution of the regression estimator as will be seen in Section 4.2.

Let us write, for brevity,

$$\widehat{f}_{\delta} - f_{\delta} = U_n + V_n,\tag{4.7}$$

where, according to (2.20) and (2.21),

$$\begin{aligned}U_n &= (\varphi_1(\widehat{\Sigma})) \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i (X_i - \bar{X}) \right), \\ V_n &= (\varphi_2(\widehat{\Sigma}) - \varphi_2(\Sigma))f.\end{aligned}\tag{4.8}$$

Note that φ_1 and φ_2 depend on δ .

With statistical applications in mind, it would be interesting if there would exist numbers $a_n \uparrow \infty$ and $\delta(n) \downarrow 0$, as $n \rightarrow \infty$, such that

$$a_n(\widehat{f}_{\delta(n)} - f) \xrightarrow{d} \mathcal{L}, \quad \text{as } n \rightarrow \infty, \text{ in } \mathbb{H},\tag{4.9}$$

where \mathcal{L} is a nondegenerate random vector. It has been shown in Cardot et al. [28], however, that such a convergence in distribution when we center at f is not in general possible.

Theorem 4.1. *For fixed $\delta > 0$, one has*

$$\sqrt{n}(\widehat{f}_{\delta} - f_{\delta}) \xrightarrow{d} \mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2, \quad \text{as } n \rightarrow \infty, \text{ in } \mathbb{H},\tag{4.10}$$

where $\mathcal{L}_1 = \varphi_1(\Sigma)\mathcal{G}_0$ and $\mathcal{L}_2 = (\varphi_{2,\Sigma}\mathcal{G}_{\Sigma})f$ are zero mean Gaussian random elements, and $\mathcal{L}_1 \perp\!\!\!\perp \mathcal{L}_2$.

Further information about the structure of the covariance operator of the random vector \mathcal{L} on the right in (4.10) will be needed in order to be able to exploit the theorem for statistical inference. This will be addressed in the next section.

4.2. Further Specification of the Limiting Distribution

It follows from (4.5) that \mathcal{G}_0 has a Karhunen-Loève expansion

$$\mathcal{G}_0 = \sum_{j=1}^{\infty} v \sigma_j Z_j p_j, \quad (4.11)$$

where the real valued random variables

$$Z_j (j \in \mathbb{N}) \text{ are iid } N(0, 1). \quad (4.12)$$

Accordingly \mathcal{H}_1 in (4.10) can be further specified as

$$\mathcal{H}_1 = \varphi_1(\Sigma) \mathcal{G}_0 = v \sum_{j=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} Z_j p_j. \quad (4.13)$$

The Gaussian operator in (4.2) has been investigated in Dauxois et al. [27], and here we will briefly summarize some of their results in our notation. By evaluating the inner product in \mathcal{L}_{HS} in the basis $p_1, p_2 \dots$ it follows from (4.6) that

$$\begin{aligned} & \mathbb{E} \langle p_j \otimes p_k, \mathcal{G}_{\Sigma} \rangle_{\text{HS}} \langle \mathcal{G}_{\Sigma}, p_{\alpha} \otimes p_{\beta} \rangle_{\text{HS}} \\ &= \mathbb{E} \langle p_j, \langle X - \mu, p_k \rangle (X - \mu) - \sigma_k^2 p_k \rangle \times \langle p_{\alpha}, \langle X - \mu, p_{\beta} \rangle (X - \mu) - \sigma_{\beta}^2 p_{\beta} \rangle \\ &= \mathbb{E} \langle X - \mu, p_j \rangle \langle X - \mu, p_k \rangle \langle X - \mu, p_{\alpha} \rangle \langle X - \mu, p_{\beta} \rangle - \delta_{j,k} \delta_{\alpha,\beta} \sigma_k^2 \sigma_{\beta}^2. \end{aligned} \quad (4.14)$$

This last expression does not in general further simplify. However, if we assume that the regressor X satisfies

$$X \stackrel{d}{=} \text{Gaussian}(\mu, \Sigma), \quad (4.15)$$

it can be easily seen that the

$$\langle X - \mu, p_j \rangle \stackrel{d}{=} N(0, \sigma_j^2) \text{ are independent,} \quad (4.16)$$

so that the expression in (4.14) equals zero if $(j, k) \neq (\alpha, \beta)$. As in Dauxois et al. [27], we obtain in this case

$$\mathbb{E} \langle p_j \otimes p_k, \mathcal{G}_{\Sigma} \rangle_{\text{HS}} \langle \mathcal{G}_{\Sigma}, p_{\alpha} \otimes p_{\beta} \rangle_{\text{HS}} = \begin{cases} 0, & (j, k) \neq (\alpha, \beta), \\ v_{j,k}^2, & (j, k) = (\alpha, \beta), \end{cases} \quad (4.17)$$

where

$$v_{j,k}^2 = \begin{cases} 2\sigma_j^4, & j = k, \\ \sigma_j^2 \cdot \sigma_k^2, & j \neq k. \end{cases} \quad (4.18)$$

Consequently the $p_j \otimes p_k$ ($j \in \mathbb{N}, k \in \mathbb{N}$) are an orthonormal basis of eigenvectors of the covariance operator of \mathcal{G}_Σ with eigenvalues $v_{j,k}^2$. Hence \mathcal{G}_Σ has the Karhunen-Loève expansion (in \mathcal{L}_{HS})

$$\mathcal{G}_\Sigma = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{j,k} Z_{j,k} p_j \otimes p_k, \quad (4.19)$$

where the random variables

$$Z_{j,k} (j \in \mathbb{N}, k \in \mathbb{N}) \text{ are iid } N(0, 1). \quad (4.20)$$

Let us, for brevity, write (see (3.7) for φ_2)

$$w_{j,k} = \begin{cases} \frac{\varphi_2(\sigma_k^2) - \varphi_2(\sigma_j^2)}{\sigma_k^2 - \sigma_j^2}, & j \neq k \\ \varphi_2'(\sigma_j^2), & j = k \end{cases} = \frac{\delta}{(\delta + \sigma_j^2)(\delta + \sigma_k^2)}, \quad \forall j, k \in \mathbb{N}, \quad (4.21)$$

$$\begin{aligned} \mathcal{H}_2 &= (\varphi_{2,\Sigma} \mathcal{G})f \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \mathcal{G}_\Sigma P_k f \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \left(\sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} v_{\alpha,\beta} Z_{\alpha,\beta} p_\alpha \otimes p_\beta \right) P_k f \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} v_{j,k} Z_{j,k} \langle f, p_k \rangle p_j. \end{aligned} \quad (4.22)$$

Summarizing, we have the following result.

Theorem 4.2. *The random element \mathcal{H}_1 , on the right in (4.10) can be represented by (4.13). If one assumes that the regressor $X \stackrel{d}{=} \text{Gaussian}(\mu, \Sigma)$, the random element \mathcal{H}_2 on the right in (4.10) can be represented by (4.22), where the $Z_{j,k}$ in (4.20) are stochastically independent of the Z_j in (4.12).*

4.3. Asymptotics under the Null Hypothesis

Let us recall that f_δ is related to f according to (2.21) so that the equivalence

$$H_0 : f_\delta = 0 \iff f = 0, \quad (4.23)$$

where again $\delta > 0$ is fixed, holds true. The following is immediate from Theorem 4.1.

Theorem 4.3. *Under the null hypothesis in (4.23), one has*

$$n \left\| \widehat{f}_\delta \right\|^2 \xrightarrow{d} \|\mathcal{L}\|^2 = \|\mathcal{L}_1\|^2, \quad \text{as } n \rightarrow \infty, \quad (4.24)$$

where

$$\|\mathcal{L}_1\|^2 \stackrel{d}{=} \sigma^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} Z_j^2. \quad (4.25)$$

An immediate generalization of the hypothesis in (4.23) is

$$H_0 : f_\delta = f_{\delta,0} = (\varphi_2(\Sigma)) f_0 \iff f = f_0, \quad (4.26)$$

for some given $f_0 \in \mathbb{H}$. This hypothesis is in principle of interest for confidence sets. Of course, testing (4.26) can be reduced to testing (4.23) by replacing the η_i with

$$\eta'_i = \eta_i - \langle X_i, f_{\delta,0} \rangle = \alpha + \langle X_i, f - f_{\delta,0} \rangle + \varepsilon_i, \quad (4.27)$$

and then using the estimator

$$\widehat{f}'_\delta = (\delta I + \widehat{\Sigma})^{-1} \left(\frac{1}{n} \sum_{i=1}^n \eta'_i (X_i - \bar{X}) \right). \quad (4.28)$$

Since $f - f_{\delta,0} = 0$ the following is immediate.

Theorem 4.4. *Assuming (4.27), one has*

$$n \left\| \widehat{f}'_\delta \right\|^2 \xrightarrow{d} \|\mathcal{L}_1\|^2, \quad \text{as } n \rightarrow \infty, \quad (4.29)$$

where $\|\mathcal{L}_1\|^2$ has the same distribution as in (4.25). Related results can be found in Cardot et al. [2].

Another generalization of the hypothesis in (4.23) is

$$H_0 : f_\delta \in \mathbb{M} = [q_1, \dots, q_M], \quad (4.30)$$

where q_1, \dots, q_M are orthonormal vectors in \mathbb{H} . Let Q and Q^\perp denote the orthogonal projection onto \mathbb{M} and \mathbb{M}^\perp respectively, and note that $f_\delta \in \mathbb{M}$ if and only if $\|Q^\perp f_\delta\| = 0$. A test statistic might be based on $\|Q^\perp \widehat{f}_\delta\|^2$ and we have the following.

Theorem 4.5. *Under H_0 in (4.30), one has*

$$n \left\| Q^\perp \widehat{f}_\delta \right\|^2 \xrightarrow{d} \left\| Q^\perp \mathcal{L} \right\|^2, \quad \text{as } n \rightarrow \infty. \quad (4.31)$$

The distribution on the right in (4.31) is rather complicated if q_1, \dots, q_M remain arbitrary. But if we are willing to assume (4.20), it follows from (4.27) that

$$\begin{aligned}
\|Q^\perp \mathcal{H}\|^2 &= \|Q^\perp \mathcal{H}_1\|^2 + \|Q^\perp \mathcal{H}_2\|^2 + 2\langle Q^\perp \mathcal{H}_1, Q^\perp \mathcal{H}_2 \rangle \\
&= v^2 \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} \frac{\sigma_k}{\delta + \sigma_k^2} Z_j Z_k \langle Q^\perp p_j, Q^\perp p_k \rangle \\
&\quad + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} w_{j,k} w_{\alpha,\beta} v_{j,k} v_{\alpha,\beta} Z_{j,k} Z_{\alpha,\beta} \langle f, p_k \rangle \langle f, p_\beta \rangle \langle Q^\perp p_j, Q^\perp p_\alpha \rangle \\
&\quad + 2v \sum_{j=1}^{\infty} \frac{\sigma_j}{\delta + \sigma_j^2} Z_j \sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} w_{\alpha,\beta} v_{\alpha,\beta} Z_{\alpha,\beta} \langle f, p_\beta \rangle \langle Q^\perp p_j, Q^\perp p_\alpha \rangle.
\end{aligned} \tag{4.32}$$

A simplification is possible if we are willing to modify the hypothesis in (4.30) and use a so-called neighborhood hypothesis. This notion has a rather long history and has been investigated by Hodges & Lehmann [29] for certain parametric models. Dette & Munk [30] have rekindled the interest in it by an application in nonparametric regression. In the present context we might replace (4.30) with the neighborhood hypothesis

$$H_{0,\varepsilon} : \|Q^\perp f_\delta\|^2 \leq \varepsilon^2, \quad \text{for some } \varepsilon > 0. \tag{4.33}$$

It is known from the literature that the advantage of using a neighborhood hypothesis is not only that such a hypothesis might be more realistic and that the asymptotics are much simpler, but also that without extra complication we might interchange null hypothesis and alternative. This means in the current situation that we might as well test the null hypothesis

$$H'_{0,\varepsilon} : \|Q^\perp f_\delta\|^2 \geq \varepsilon^2, \quad \text{for some } \varepsilon > 0, \tag{4.34}$$

which could be more suitable, in particular in goodness-of-fit problems.

The functional $g \mapsto \|Q^\perp g\|^2$, $g \in \mathbb{H}$, has a Fréchet derivative at f_δ given by the functional $g \mapsto 2\langle g, Q^\perp f_\delta \rangle$, $g \in \mathbb{H}$. Therefore, the delta method in conjunction with Theorem 4.1 entails the following result.

Theorem 4.6. *One has*

$$\sqrt{n} \left\{ \|Q^\perp \hat{f}_\delta\|^2 - \|Q^\perp f_\delta\|^2 \right\} \xrightarrow{d} 2\langle \mathcal{H}, Q^\perp f_\delta \rangle, \quad \text{as } n \rightarrow \infty. \tag{4.35}$$

The limiting distribution on the right in (4.35) is normal with mean zero and complicated variance

$$\begin{aligned}
\Delta^2 &= 4\mathbb{E}\langle \mathcal{H}, Q^\perp f_\delta \rangle^2 \\
&= 4\left\{ \mathbb{E}\langle \mathcal{H}_1, Q^\perp f_\delta \rangle^2 + \mathbb{E}\langle \mathcal{H}_2, Q^\perp f_\delta \rangle^2 \right\} \\
&= 4v^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} \langle p_j, Q^\perp f_\delta \rangle^2 \\
&\quad + 4\mathbb{E}\left\langle \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k} P_j \left(\sum_{\alpha=1}^{\infty} \sum_{\beta=1}^{\infty} v_{\alpha,\beta} Z_{\alpha,\beta} p_\alpha \otimes p_\beta \right) P_k f, Q^\perp f_\delta \right\rangle^2 \\
&= 4v^2 \sum_{j=1}^{\infty} \frac{\sigma_j^2}{(\delta + \sigma_j^2)^2} \langle p_j, Q^\perp f_\delta \rangle^2 \\
&\quad + 4 \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} w_{j,k}^2 v_{j,k}^2 \langle f, p_k \rangle^2 \langle p_j, Q^\perp f_\delta \rangle^2.
\end{aligned} \tag{4.36}$$

Remark 4.7. As we see from the expressions in (4.24), (4.32), and (4.36), the limiting distributions depend on infinitely many parameters that must be suitably estimated in order to be in a position to use the statistics for actual testing. Estimators for the individual parameters are not too hard to obtain. The eigenvalues σ_j^2 and eigenvectors p_j of Σ , for instance, can in principle be estimated by the corresponding quantities of $\hat{\Sigma}$. Although in any practical situation only a finite number of these parameters can be estimated, theoretically this number must increase with the sample size and some kind of uniform consistency will be needed for a suitable approximation of the limiting distribution. This interesting question of uniform consistency seems to require quite some technicalities and will not be addressed in this paper.

Remark 4.8. In this paper we have dealt with the situation where Σ is entirely unknown. It has been observed in Johannes [11] that if X is a stationary process on the unit interval, the eigenfunctions p_j of the covariance operator are always the same, known system of trigonometric functions, and only its eigenvalues σ_j^2 are unknown. Knowing the p_j leads to several simplifications. In the first place, Σ can now be estimated by the expression on the right in (3.1) with only the σ_k^2 replaced with estimators. If $\tilde{\Sigma}$ is this estimator, it is clear that Σ and $\tilde{\Pi} = \tilde{\Sigma} - \Sigma$ commute, so that the derivative $\dot{\varphi}_{2,\Sigma}$ now simplifies considerably (see Remark 3.3). Secondly, we might consider the special case of H_0 in (4.30), where $q_j = p_j$, $j = 1, \dots, M$. We now have

$$f_\delta \in \mathbb{M} = [p_1, \dots, p_M] \iff f \in \mathbb{M}, \tag{4.37}$$

so that even for fixed δ we can test the actual regression function. In the third place, under the null hypothesis in (4.37), the number of unknown parameters in (4.32) reduces considerably because now $Q^\perp p_j = 0$ for $j = 1, \dots, M$. When the p_j are known, in addition to all the changes mentioned above, also the limiting distribution of Σ differs from that of $\hat{\Sigma}$. Considering all these modifications that are needed, it seems better not to include this important special case in this paper.

4.4. Asymptotics under Local Alternatives

Again we assume that X is Gaussian. Suppose that

$$f = f_n = \tilde{f} + \frac{1}{\sqrt{n}}g, \quad \text{for } \tilde{f}, g \in \mathbb{H}. \quad (4.38)$$

For such f_n only minor changes in the asymptotics are required, because the conditions on the X_i and ε_i are still the same and do not change with n . Let us write

$$f_\delta = f_{n,\delta} = \tilde{f}_\delta + \frac{1}{\sqrt{n}}g_\delta, \quad (4.39)$$

where $\tilde{f}_\delta = (\delta I + \Sigma)^{-1}\tilde{f}$, $g_\delta = (\delta I + \Sigma)^{-1}\Sigma g$. The following is immediate from a minor modification of Theorem 4.1.

Theorem 4.9. For $f_\delta = f_{n,\delta}$ as in (4.39), one has

$$\sqrt{n}(\hat{f}_\delta - \tilde{f}_\delta) \xrightarrow{d} g_\delta + \mathcal{A}_1 + \tilde{\mathcal{A}}_2, \quad (4.40)$$

where $\mathcal{A}_1 = \varphi_1(\Sigma)\mathcal{G}_0$ is the same as in (4.13), $\tilde{\mathcal{A}}_2$ is obtained from \mathcal{A}_2 in (4.22) by replacing f with \tilde{f} , and $\mathcal{A}_1 \perp\!\!\!\perp \tilde{\mathcal{A}}_2$.

By way of an example, let us apply this result to testing the neighborhood hypothesis and consider the asymptotics of the test statistics in (4.35) under the local alternatives $f_{n,\delta}$ in (4.39) with

$$\|Q^\perp \tilde{f}_\delta\|^2 = \varepsilon^2, \quad g_\delta \perp \mathbb{M}, \quad \langle \tilde{f}_\delta, g_\delta \rangle > 0. \quad (4.41)$$

Note that under such alternatives, $\hat{f}_{\delta(n)}$ is still a consistent estimator of \tilde{f} for any $\delta(n) \downarrow 0$, as $n \rightarrow \infty$, and so is \hat{f}_δ for f_δ . The following is immediate from Theorem 4.9. To conclude this section, let us first assume that parameters can be suitably estimated. We then arrive at the limiting distribution of a test statistics that allows the construction of an asymptotic level- α test whose asymptotic power can be computed.

Theorem 4.10. For $f_\delta = f_{n,\delta}$, as in (4.40) and (4.41), one has

$$T_n = \frac{\sqrt{n}}{\hat{\Delta}} \left(\|Q^\perp \hat{f}_\delta\|^2 - \varepsilon^2 \right) \xrightarrow{d} N \left(\frac{2\langle g_\delta, Q^\perp \tilde{f}_\delta \rangle}{\Delta}, 1 \right), \quad \text{as } n \rightarrow \infty, \quad (4.42)$$

assuming that $\hat{\Delta}$ is a consistent estimator of Δ in (4.36). Note that the limiting distribution is $N(0, 1)$ under \mathcal{A}_0 (i.e., $g_\delta = 0$).

5. Asymptotic Optimality of the Classification Rule

In addition to Assumption 2.5 and (2.34), it will be assumed that the smoothness parameter $\delta = \delta(n)$ in (2.39) satisfies

$$\delta(n) \rightarrow 0, \quad \delta(n) \gg n^{-1/4}, \quad \text{as } n \rightarrow \infty. \quad (5.1)$$

We will also assume that the sizes of the training samples n_1 and n_2 (see (2.36)) are deterministic and satisfy ($n = n_1 + n_2$)

$$0 < \liminf_{n \rightarrow \infty} \frac{n_j}{n} \leq \limsup_{n \rightarrow \infty} \frac{n_j}{n} < 1. \quad (5.2)$$

Let us recall from (3.7) that $\varphi_1(z) = \varphi_{1,n}(z) = 1/\{\delta(n) + z\}$, $z \neq -\delta(n)$.

Under these conditions the probability of misclassification equals $\mathbb{P}\{\langle X - (1/2)(\bar{X}_1 + \bar{X}_2), \varphi_{1,n}(\hat{\Sigma})(\bar{X}_1 - \bar{X}_2) \rangle > 0 \mid X \stackrel{d}{=} \mathcal{G}(\mu_2, \Sigma)\}$. Let us note that

$$\begin{aligned} & \left| \left\langle X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2), \varphi_{1,n}(\hat{\Sigma})(\bar{X}_1 - \bar{X}_2) \right\rangle - \left\langle X - \frac{1}{2}(\mu_1 + \mu_2), \varphi_{1,n}(\Sigma)(\mu_1 - \mu_2) \right\rangle \right| \\ & \leq \frac{1}{2} \left\{ \|\bar{X}_1 - \mu_1\| + \|\bar{X}_2 - \mu_2\| \right\} \|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \|\mu_1 - \mu_2\| \\ & \quad + \left\| X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \right\| \times \left[\|\varphi_{1,n}(\hat{\Sigma}) - \varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \|\bar{X}_1 - \bar{X}_2\| \right. \\ & \quad \left. + \left\{ \|\bar{X}_1 - \mu_1\| + \|\bar{X}_2 - \mu_2\| \right\} \|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} \right]. \end{aligned} \quad (5.3)$$

Since $\|\bar{X}_j - \mu_j\| = \mathcal{O}_p(n^{-1/2})$, $\|\varphi_{1,n}(\Sigma)\|_{\mathcal{L}} = \mathcal{O}(\delta^{-1}(n))$, and, according to (3.21),

$$\|\varphi_{1,n}(\hat{\Sigma}) - \varphi_{1,n}(\Sigma)\|_{\mathcal{L}} = \mathcal{O}_p\left(\frac{1}{\delta^2(n)\sqrt{n}}\right), \quad (5.4)$$

it follows from (5.1) that the limit of the misclassification probability equals

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left\langle X - \mu_2 - \frac{1}{2}(\mu_1 - \mu_2), \varphi_{1,n}(\Sigma)(\mu_1 - \mu_2) \right\rangle > 0 \right\} \\ & = 1 - \Phi \left(\frac{1}{2} \left\langle \mu_1 - \mu_2, \Sigma^{-1}(\mu_1 - \mu_2) \right\rangle \right), \end{aligned} \quad (5.5)$$

where Φ is the standard normal cdf.

For (5.5) we have used the well-known property of regularized inverses that $\|(\delta + \Sigma)^{-1}\Sigma f - f\| \rightarrow 0$, as $\delta \rightarrow 0$, for all $f \in \mathbb{H}$, and the fact that we may choose $f = \Sigma^{-1}(\mu_1 - \mu_2)$ by Assumption 2.5. Since rule (2.33) is optimal when parameters are known, we have obtained the following result.

Theorem 5.1. *Let $P_j \stackrel{d}{=} \mathcal{G}(\mu_j, \Sigma)$ and let Assumption 2.5 and (5.1) be satisfied. Then the classification rule (2.39) is asymptotically optimal.*

References

- [1] P. Hall and J. L. Horowitz, "Methodology and convergence rates for functional linear regression," *The Annals of Statistics*, vol. 35, no. 1, pp. 70–91, 2007.
- [2] H. Cardot, F. Ferraty, A. Mas, and P. Sarda, "Testing hypotheses in the functional linear model," *Scandinavian Journal of Statistics*, vol. 30, no. 1, pp. 241–255, 2003.
- [3] H. Cardot, F. Ferraty, and P. Sarda, "Functional linear model," *Statistics & Probability Letters*, vol. 45, no. 1, pp. 11–22, 1999.
- [4] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *The Annals of Statistics*, vol. 23, no. 1, pp. 73–102, 1995.
- [5] D. S. Gilliam, T. Hohage, X. Ji, and F. Ruymgaart, "The Fréchet derivative of an analytic function of a bounded operator with some applications," *International Journal of Mathematics and Mathematical Sciences*, vol. 2009, Article ID 239025, 17 pages, 2009.
- [6] J. Cupidon, D. S. Gilliam, R. Eubank, and F. Ruymgaart, "The delta method for analytic functions of random operators with application to functional data," *Bernoulli*, vol. 13, no. 4, pp. 1179–1194, 2007.
- [7] R. L. Eubank and T. Hsing, "Canonical correlation for stochastic processes," *Stochastic Processes and their Applications*, vol. 118, no. 9, pp. 1634–1661, 2008.
- [8] G. He, H.-G. Müller, and J.-L. Wang, "Functional canonical analysis for square integrable stochastic processes," *Journal of Multivariate Analysis*, vol. 85, no. 1, pp. 54–77, 2003.
- [9] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman, "Canonical correlation analysis when the data are curves," *Journal of the Royal Statistical Society. Series B*, vol. 55, no. 3, pp. 725–740, 1993.
- [10] N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory*, Interscience, New York, NY, USA, 1957.
- [11] J. Johannes, "Privileged communication," 2008.
- [12] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2005.
- [13] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2006.
- [14] D. Bosq, *Linear Processes in Function Spaces*, vol. 149 of *Lecture Notes in Statistics*, Springer-Verlag, New York, NY, USA, 2000.
- [15] A. Mas, *Estimation d'opérateurs de corrélation de processus fonctionnels: Lois limites, tests, déviations modérées*, Ph.D. thesis, Université Paris VI, 2000.
- [16] R. G. Laha and V. K. Rohatgi, *Probability Theory*, John Wiley & Sons, New York, NY, USA, 1979, Wiley Series in Probability and Mathematical Statistics.
- [17] C. R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York, NY, USA, 1965.
- [18] A. Mas and B. Pumo, "Functional linear regression with derivatives," Tech. Rep., Institut de Modélisation Mathématique de Montpellier, 2006.
- [19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer-Verlag, New York, NY, USA, 2001.
- [20] C. R. Rao and H. Toutenburg, *Linear Models*, Springer Series in Statistics, Springer, New York, NY, USA, 1995.
- [21] J. Feldman, "Equivalence and perpendicularity of Gaussian processes," *Pacific Journal of Mathematics*, vol. 8, pp. 699–708, 1958.
- [22] Y. Hájek, "On a property of normal distribution of any stochastic process," *Czechoslovak Mathematical Journal*, vol. 8, no. 83, pp. 610–617, 1958.
- [23] U. Grenander, *Abstract Inference*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 1981.
- [24] A. V. Skorohod, *Integration in Hilbert Space*, Springer, New York, NY, USA, 1974.
- [25] G. S. Watson, *Statistics on Spheres*, John Wiley & Sons, New York, NY, USA, 1983.
- [26] T. Kato, *Perturbation Theory for Linear Operators*, Die Grundlehren der mathematischen Wissenschaften, Band 132, Springer-Verlag, New York, NY, USA, 1966.
- [27] J. Dauxois, A. Pousse, and Y. Romain, "Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference," *Journal of Multivariate Analysis*, vol. 12, no. 1, pp. 136–154, 1982.
- [28] H. Cardot, A. Mas, and P. Sarda, "CLT in functional linear regression models," *Probability Theory and Related Fields*, vol. 138, no. 3-4, pp. 325–361, 2007.

- [29] J. L. Hodges, Jr. and E. L. Lehmann, "Testing the approximate validity of statistical hypotheses," *Journal of the Royal Statistical Society. Series B*, vol. 16, pp. 261–268, 1954.
- [30] H. Dette and A. Munk, "Validation of linear regression models," *The Annals of Statistics*, vol. 26, no. 2, pp. 778–800, 1998.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

