Some Statistical Methods for Directly and Indirectly Observed Functional Data

by

Johnny Pang

A Dissertation

in

Mathematics

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in Mathematics
Approved

Dr. Frits Ruymgaart

Dr. Robert Paige

Dr. Alex Wang

Dr. Fred Hartmeister

Dean of the Graduate School

August, 2008

ACKNOWLEDGEMENTS

Getting my doctoral is always a dream in my life. For the past years at Texas Tech, I have been learning so much that I could not imagine. I remember the first semester at Tech, I could only sleep about four hours daily because of all the school work and teaching duty. Now I can still remember the anxiety and excitement on the first time handling the responsibilities of being both student and teacher. I am glad that I have worked and studied hard, and had mastered almost all my classes material. These learning and teaching experiences at tech also prepare me to be a great teacher and leader. I am very happy that I have received much support from professors, friends, and classmates in the TTU mathematics department. Without them, I don't think I could survive until now.

In here, I would like to especially like to thank Dr. Frits Ruymgaart for his tireless support and inspiration. With his loving guidance, I could finish this dissertation. Also, I learned more analysis and theoretical statistics than I thought I could ever understand. His advice and support will always be treasured and never forgotten. He is truly one of the most humble men that I have ever had and he exceeds all expectations of what an advisor should be.

TABLE OF CONTENTS

ABSTRACT

In this dissertation, we will be concerned with the statistical inference regarding linear models with functional data. For the sake of generality these functional data will be considered as sample elements in an abstract infinite dimensional Hilbert space. In the special instance of the one-sample problem, both directly and indirectly observed functions will be included. It should be stressed that the linear model mentioned above each sample elements itself is a function, so that we have more information than in cases where the data consist of a number of sampled function values.

In Chapter 1, we will review some useful properties and formulas of arbitrary random variables and Gaussian random variables in Hilbert spaces. It should be noted that a Gaussian measure will be employed as a dominating measure because there doesn't exist a shift invariant (i.e. Lebesgue) measure on an infinite dimensional Hilbert space.

In Chapter 2, linear model in Hilbert space will be considered. We will borrow the notation from the univariate linear model and use matrices to arrive at a convenient notation for linear models in Hilbert spaces. We will show that our estimator of the function parameter has approximately a Gaussian distribution for large sample size.

In Chapter 3, three special cases of the main model introduced in Chapter 2 will be considered. First, the simplest version of the one-sample problem in Hilbert spaces will be introduced together with an application of neighborhood hypotheses. Second, the indirect-one-sample problem in Hilbert spaces will be considered. We will exploit the spectral-cut-off type regularized inverse and consider the MISE of the estimator as a means to investigate its quality. In fact, we will prove that the estimator is rate-optimal. Finally, multi-sample problem will be briefly considered along the same lines as the direct one-sample problem.

CHAPTER 1

INTRODUCTION OF PROPERTIES IN HILBERT SPACES $\mathbb{H}$

1.1 Random Variables

Throughtout this dissertation, we will focus on functional data in infinite dimensional Hilbert space. Let us introduce some useful properties of Hilbert-space-valued random variables in this section. A Hilbert space is equipped with a scalar product, in terms of which we can express both the length of vectors and angles between vectors. With these rich and useful properties of Hilbert spaces, we want to define mean and covariance operator of random elements in this section. Let $(\Omega, \mathcal{W}, \mathbb{P})$ be a probability space and throughout $\mathbb{H}$ will be a separable Hilbert space over the real numbers equipped with the σ-field $\mathcal{B}$ of Borel sets, and of infinite dimension. A mapping $X : \Omega \to \mathbb{H}$ is called a random element if it is $(\mathcal{W}, \mathcal{B})$-measurable. Let $\mathbb{P}_X = P$ be the induced probability measure on $(\mathbb{H}, \mathcal{B})$. Throughout it will be assumed that $\mathbb{E}\|X\|^2 < \infty$. Under this assumption both the mean and the covariance operator of X can be defined. The mean

$$\mathbb{E}X = \mu \tag{1.1}$$

is the element in $\mathbb{H}$ uniquely determined by

$$\mathbb{E}\langle x, X\rangle = \langle x, \mu\rangle, \quad \forall x \in \mathbb{H} \tag{1.2}$$

and the covariance operator Σ is the operator mapping $\mathbb{H}$ into itself, uniquely determined by

$$\begin{aligned} \mathbb{E}\langle a, X-\mu\rangle\langle X-\mu, b\rangle &= \mathbb{E}\langle a, X\rangle\langle X, b\rangle - \langle a, \mu\rangle\langle \mu, b\rangle \\ &= \langle a, \Sigma b\rangle, \quad \text{for } a, b \in \mathbb{H}. \end{aligned} \tag{1.3}$$

These results follow from the Riesz representation theorem.

Note that it is not hard to see that Σ is bounded if we apply the Schwarz inequality. In fact, we have

$$\begin{aligned} |\langle a, \Sigma b\rangle| &= |\mathbb{E} \langle a, X-\mu\rangle \langle b, X-\mu\rangle | \\ &\leq \|a\|\|b\| \ \mathbb{E}\|X-\mu\|^2, \quad \forall\ a, b, \in \mathbb{H}. \end{aligned} \tag{1.4}$$

If we choose $a = \Sigma b$, the inequality yields

$$\|\Sigma b\|^2 \leq \|b\|\|\Sigma b\| \ \mathbb{E}\|X\|^2, \tag{1.5}$$

and hence

$$\|\Sigma b\| \leq \|b\| \ \mathbb{E}\|X\|^2, \quad \forall b \in \mathbb{H}, \tag{1.6}$$

which proves the boundedness.

Definition 1.1.1 The Banach space of all bounded linear operators $T : \mathbb{H} \to \mathbb{H}$ will be denoted by $\mathcal{L}$, and the norm in that space will be denoted by $\| \cdot \|_{\mathcal{L}}$.

Consider

$$\begin{aligned} \sum_{i=1}^{N} \langle e_i, \Sigma e_i\rangle &= \sum_{i=1}^{N} \mathbb{E} \langle e_i, X-\mu\rangle^2 \\ &= \sum_{i=1}^{N} \mathbb{E}\{\langle e_i, X\rangle - \langle e_i, \mu\rangle\}^2 \\ &= \mathbb{E} \sum_{i=1}^{N} \langle e_i, X\rangle^2 - \sum_{i=1}^{N} \langle e_i, \mu\rangle^2 \end{aligned} \tag{1.7}$$

where $\{e_i\}$ is an orthonormal basis in $\mathbb{H}$ and $N \geq 1$. By the monotone convergence theorem, taking limits for $N \to \infty$, both sides, we obtain

$$\sum_{i=1}^{\infty} \langle e_i, \Sigma e_i\rangle = \mathbb{E} \sum_{i=1}^{\infty} \langle e_i, X\rangle^2 - \sum_{i=1}^{\infty} \langle e_i, \mu\rangle^2 . \tag{1.8}$$

Using the Parseval equality, we have

$$\sum_{i=1}^{\infty} \langle e_i, \Sigma e_i \rangle = \mathbb{E}\|X\|^2 - \|\mu\|^2. \tag{1.9}$$

This sum is called the trace and the above formula shows that the trace of Σ is finite and independent of the choice of orthonormal basis. Since Σ has finite trace, it is also compact which means that Σ maps every bounded set in $\mathbb{H}$ onto a relatively compact set in $\mathbb{H}$. It is also easy to see that the operator is Hermitian. Notice that

$$\begin{aligned} \langle a, \Sigma b \rangle &= \mathbb{E} \langle a, X - \mu \rangle \langle b, X - \mu \rangle \\ &= \mathbb{E} \langle b, X - \mu \rangle \langle a, X - \mu \rangle \\ &= \langle b, \Sigma a \rangle \\ &= \langle \Sigma a, b \rangle . \end{aligned} \tag{1.10}$$

Any such Σ has a sequence of eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \downarrow 0$ with corresponding orthonormal basis of eigenvectors $e_1, e_2, \cdots$ such that Σ can be written as

$$\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 e_k \otimes e_k. \tag{1.11}$$

In particular, the relation

$$\begin{aligned} \langle \Sigma y, y \rangle &= \int_{\mathbb{H}} \langle x, y \rangle^2 dP(x) - \langle \mu, y \rangle \\ &= \int_{\mathbb{H}} \langle x - \mu, y \rangle^2 dP(x) \end{aligned} \tag{1.12}$$

holds for all $y \in \mathbb{H}$.

Proposition 1.1.1 The covariance operator Σ associated with the probability measure P defined by the above has the following properties:
a. Σ is compact, that is, Σ maps every bounded set in $\mathbb{H}$ onto a relatively compact set in $\mathbb{H}$.
b. Σ is Hermitian, that is, $\Sigma = \Sigma^*$, where Σ^* is the adjoint of Σ.
c. Σ is positive semi-definitive, that is, $\langle \Sigma x, x \rangle \geq 0, \forall\ x \in \mathbb{H}$.

d. Σ has finite trace.

Definition 1.1.2 An operator on $\mathbb{H}$ satisfying the above properties $\mathbf{a} - \mathbf{d}$. is known as a covariance operator.

1.2 Gaussian Random Variables

In this section we extend the notion of normal distributions to infinite-dimensional Hilbert spaces.

Definition 1.2.1 Let X be a random element in $\mathbb{H}$ and $\mathbb{P}_X = P$. The characteristic functional or Fourier transform of X or P is defined as

$$\varphi(t) = \mathbb{E}e^{i\langle t,X\rangle} = \int_{x\in\mathbb{H}} e^{i\langle t,x\rangle} dP(x), \quad t \in \mathbb{H}. \tag{1.13}$$

This characteristic functional defined above is continuous, positive definite with $\varphi(0) = 1$. Notice that positive definite means that

$$\sum_{j=1}^{N}\sum_{k=1}^{N} z_j \bar{z}_k \varphi(t_j - t_k) \geq 0, \quad \forall z_1, \cdots, z_N \in \mathbb{C}, \forall t_1, \cdots, t_N \in \mathbb{R}, \text{ and } \forall N \in \mathbb{N}. \tag{1.14}$$

Conversely, we may wonder when a continuous, positive definite functional $\varphi : \mathbb{H} \to \mathbb{C}$ with $\varphi(0) = 1$ defines a probability measure on $\mathbb{H}$. This is true for $\mathbb{R}$, but not in general for infinite-dimensional $\mathbb{H}$ we consider here. Actually, a famous result, viz. the Minlos-Sazonov theorem is needed. This theorem claims that such a functional is indeed the Fourier transform of a probability measure if and only if, for given $\epsilon > 0$, there exists a covariance operator Σ_ϵ on $\mathbb{H}$, such that

$$1 - \text{Re } \varphi(t) < \epsilon, \text{ whenever } \langle \Sigma_\epsilon t, t\rangle < 1. \tag{1.15}$$

As an application of the Minlos-Sazonov theorem we define and derive some basic properties of a Gaussian measures on Hilbert spaces.

Proposition 1.2.1 Let $(\mathbb{H}, \mathcal{B})$ be a real, separable Hilbert space. Also let φ be a complex-valued function defined by the formula

$$\varphi(t) = e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Sigma t\rangle}, t \in \mathbb{H}, \tag{1.16}$$

where $\mu \in \mathbb{H}$, and Σ is a covariance operator on $\mathbb{H}$. Then φ is the Fourier transform of a probability measure P on $\mathcal{B}$ called Gaussian with mean μ and covariance operator Σ (notation: $\mathcal{G}(\mu, \Sigma)$), satisfying the condition

$$\int_{\mathbb{H}} \|x\|^2 dP(x) < \infty \tag{1.17}$$

Moreover, in this case, μ is the mean vector and Σ is the covariance operator associated with the probability measure P.

Proof: In here, we will only show that there exists a probability measure P on $\mathcal{B}$ such that φ is the Fourier transform of P. Since φ is continuous on $\mathbb{H}$ and $\varphi(0) = 1$ and Σ is a bounded, positive, self-adjoint operator on $\mathbb{H}$, φ considered as a function on any finite-dimensional subspace M of $\mathbb{H}$ is the Fourier transform of a finite-dimensional (that is, multivariate normal) Gaussian probability distribution. Hence, in particular, φ is positive definite on M. It follows from the continuity property of φ that φ is positive definite on $\mathbb{H}$. Let us now show that the condition of the Minlos-Sazonov theorem is satisfied. In fact, for $t \in \mathbb{H}$

$$\begin{aligned} 0 \le 1 - \text{Re } \varphi(t) &= 1 - \cos\langle \mu, t\rangle e^{-\frac{1}{2}\langle \Sigma t, t\rangle} \\ &= 1 - e^{-\frac{1}{2}\langle \Sigma t, t\rangle} + (1 - \cos\langle \mu, t\rangle) e^{-\frac{1}{2}\langle \Sigma t, t\rangle} \\ &\le \frac{1}{2}(\langle \Sigma t, t\rangle + \langle \mu, t\rangle^2). \end{aligned} \tag{1.18}$$

If we define the operator T_μ on $\mathbb{H}$ by setting

$$T_\mu t = \langle \mu, t\rangle \mu, \tag{1.19}$$

then we see easily that T_μ is a covariance operator on $\mathbb{H}$ such that $\text{tr}T_\mu = \|\mu\|^2$. Let $\epsilon > 0$. Define the operator Σ_ϵ on $\mathbb{H}$ by the relation $\Sigma_\epsilon = \frac{1}{2\epsilon}(\Sigma + T_\mu)$. Clearly Σ_ϵ is a covariance operator such that $\text{tr}\Sigma_\epsilon = \frac{1}{2\epsilon}(\text{tr}\Sigma + \|\mu\|^2) < \infty$. Let $t \in \mathbb{H}$ be such that $\langle \Sigma_\epsilon t, t\rangle < 1$. Then $\frac{1}{2\epsilon}(\langle \Sigma t, t\rangle + \langle \mu, t\rangle^2) < 1$, which implies that $0 \le 1-\text{Re } \varphi(t) < \epsilon$. It follows from the Minlos-Sazonov theorem that there exists a probability measure P on $\mathcal{B}$ such that $\varphi = \hat{P}$.

Proposition 1.2.2 Let $(\mathbb{H}, \mathcal{B})$ be a real, separable Hilbert space. An $\mathbb{H}$-valued random variable X is $\mathcal{G}(\mu, \Sigma)$ if and only if for every $t \in \mathbb{H}$ the real-valued random variable $\langle X, t\rangle$ is $\mathcal{N}(\langle \mu, t\rangle, \langle t, \Sigma t\rangle)$.

Proof: We will proof the only if part here. Suppose that X is Gaussian. Then its characteristic functional φ is of the form $\varphi(t) = \mathbb{E}e^{i\langle X,t\rangle} = e^{i\langle \mu,t\rangle - \frac{1}{2}\langle \Sigma t,t\rangle}$, for $t \in \mathbb{H}$, where $\mu \in \mathbb{H}$ is the mean vector, and Σ defined on $\mathbb{H}$ is the covariance operator associated with the probability distribution P_X of X. Then for $k \in \mathbb{R}$ we have $\mathbb{E}e^{ik\langle X,t\rangle} = \varphi(kt) = e^{ik\langle \mu,y\rangle - \frac{1}{2}k^2\langle \Sigma t,t\rangle}$ so that $\langle X, t\rangle$ is a normal random variable with mean $\langle \mu, t\rangle$ and variance $\langle \Sigma t, t\rangle$.

In Euclidean spaces the normal distribution can also be defined by means of its density with respect to Lebesgue measure, which is a very natural choice for a dominating measure. It should be noted, however, that on an infinite dimensional Hilbert space Lebesgue measure (invariant with respect to shift) doesn't exist.

The simplest type of measure to employ as a dominating measure is one of the Gaussian measures. Fixing one such Gaussian measure P_0, we may want to find the density of another Gaussian measure P_1 with respect to P_0. However, there another difficulty arises: according to a famous result two Gaussain measures with the same covariance structure are either equivalent or orthogonal to each other. This situation is also very different from the Euclidean case where two such measures are always equivalent. In other words, P_1 does not always have a density with respect to P_0. When densities do exist, however, they enable us to find sufficient statistics in the usual way, as will turn out to be useful in Section 3.2. Let us now formulate a necessary and sufficient condition for equivalence.

Let P_0 be Gaussian with mean $\mu_0 \in \mathbb{H}$ and covariance operator Σ, and let P_1 be Gaussian with mean $\mu_1 \in \mathbb{H}$ and the same covariance operator Σ. As we have seen in (1.11), we may expand Σ in its eigenprojections and write

$$\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 e_k \otimes e_k. \tag{1.20}$$

An important tool in the study of equivalence of measures on infinite dimensional spaces is Kakutani's lemma; see Grenander (1981). In the following form, this lemma is of sufficient generality for our purposes.

Let $P_0 = P_{01} \times P_{02} \times \cdots$ be an infinite product of probability measures P_{0k} on $\mathbb{R}$, and similarly, let $P_1 = P_{11} \times P_{12} \times \cdots$ be an infinite product of probability measures P_{1k} on $\mathbb{R}$. Both P_0 and P_1 are then probability measures on $\mathbb{R}^{\mathbb{N}}$. Suppose that $P_{0k} \sim P_{1k}$ for all k, and let f_{0k} and f_{1k} be the densities of P_{0k} and P_{1k} with respect to some dominating measure, for instance $\mu_k = P_{0k} + P_{1k}$. The Hellinger affinity ρ_k of P_{0k} and P_{1k} is defined as

$$\rho_k = \int_{\mathbb{R}} \sqrt{f_{0k}(x) f_{1k}(x)} d\mu_k(x), \tag{1.21}$$

and doesn't depend on the choice of μ_k. According to Kakutani's lemma, we have

$$P_0 \sim P_1 \Longleftrightarrow \prod_{k=1}^{\infty} \rho_k > 0. \tag{1.22}$$

We now continue with a sketch of the proof of the equivalence-orthogonality statement above. Although (1.22) is for measures on $\mathbb{R}^{\mathbb{N}}$, it is not too hard to see that it also applies to the measures P_0 and P_1 on $\mathbb{H}$. In fact, in this situation the k-th marginals are obtained as the distributions of $\langle X, e_k \rangle$, where the e_k and the elements of the basis in (1.20), and $\mathbb{P}_X$ is either P_0 or P_1. It is immediate that

$$\langle X, e_k \rangle \stackrel{d}{=} \mathcal{N}(\langle \mu_j, e_k \rangle, \sigma_k^2), \quad \text{if } X \stackrel{d}{=} P_j, \tag{1.23}$$

and

$$\mathbb{E}\langle e_k, X - \mu_j \rangle \langle X - \mu_j, e_m \rangle = \sigma_k^2 \delta_{km}. \tag{1.24}$$

Because of the joint normality, the marginals therefore have to be independent. Hence the affinity in (1.21) can be written

$$\begin{aligned} \rho_k &= \frac{1}{\sqrt{2\pi\sigma_k^2}} \int_{\mathbb{R}} e^{\frac{-1}{4}\left(\frac{(x-\mu_{0k})^2}{\sigma_k^2} + \frac{(x-\mu_{1k})^2}{\sigma_k^2}\right)} dx \\ &= e^{-\frac{1}{8\sigma_k^2}(\mu_{0k}-\mu_{1k})^2} \end{aligned} \tag{1.25}$$

That implies,

$$\prod_{k=1}^{\infty} \rho_k = e^{-\frac{1}{8}\sum_{k=1}^{\infty}\frac{(\mu_{0k}-\mu_{1k})^2}{\sigma_k^2}} > 0. \tag{1.26}$$

Hence the product on the left is strictly positive if and only if the sum in the exponent is finite, so that we have the following result that can be found in Feldman (1958) and Hájek (1958).

Theorem 1.2.1 For the two Gaussian measures P_0 and P_1 with means μ_0 and μ_1 and common covariance operator Σ to be equivalent, it is necessary and sufficient that

$$\sum_{k=1}^{\infty} \frac{(\mu_{0k}-\mu_{1k})^2}{\sigma_k^2} < \infty. \tag{1.27}$$

The only alternative is that they are perpendicular.

Theorem 1.2.2 Under the condition (1.27), the Radon-Nikodym derivative for P_1 w.r.t. P_0 equals

$$f(x) = \frac{dP_1}{dP_0}(x) = e^{\sum_{k=1}^{\infty}\frac{\mu_{1k}-\mu_{0k}}{\sigma_k^2}(x_k-\frac{\mu_{1k}+\mu_{0k}}{2})}, \quad x = \sum_{k=1}^{\infty} x_k e_k \in \mathbb{H}. \tag{1.28}$$

Proof: Since (1.27) holds, the question is how the Radon-Nikodym derivative $f = \frac{dP_1}{dP_0}$ can be explicitly computed. It is clear from (1.23) and (1.24) that the density of $P_{11} \times \cdots \times P_{1n}$ with respect to $P_{01} \times \cdots \times P_{0n}$ equals

$$f_n(x) = \frac{\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_k^2}}\ e^{-\frac{1}{2\sigma_k^2}(x_k-\mu_{1k})^2}}{\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_k^2}}\ e^{-\frac{1}{2\sigma_k^2}(x_k-\mu_{0k})^2}} \tag{1.29}$$

Since $f_n \longrightarrow f$ in probability P_0 and P_1, the limit can be expressed as

$$f(x) = e^{\sum_{k=1}^{\infty}\frac{\mu_{1k}-\mu_{0k}}{\sigma_k^2}(x_k-\frac{\mu_{1k}+\mu_{0k}}{2})} \tag{1.30}$$

and this yields the result.

1.3 Sampling from an Arbitrary Distribution

In this section, we will consider random samples from almost arbitrary distributions on Hilbert space, focusing in particular on the sample mean and sample covariance operator. Since exact distributions are virtually impossible to obtain, we will pay particular attention to asymptotics. Some general notions will be introduced, and applied to obtain the asymptotic distributions of the aforementioned sample quantities.

Definition 1.3.1 A family $\mathcal{P}$ of probability measures on $(\mathbb{H}, \mathcal{B})$ is called tight if for each $\epsilon > 0$ there exists a compact set $K_\epsilon \subset \mathbb{H}$ such that

$$\sup_{P \in \mathcal{P}} P(\mathbb{H} \setminus K_\epsilon) < \epsilon \tag{1.31}$$

A sequence $(T_n)_{n \geq 1}$ of $\mathbb{H}$-valued random variables is called tight if the collection $\mathcal{P} = \{\mathbb{P}_{T_n} = P_n,\ n \in \mathbb{N}\}$ of induced probability measures in tight in the sense defined above.

Definition 1.3.2 Let $P,\ P_1,\ P_2, \cdots$ be probability measures on $(\mathbb{H}, \mathcal{B})$. We say that the P_n converge in distribution to P, notation $P_n \xrightarrow{d} P$, as $n \to \infty$, if

$$\int_{\mathbb{H}} f dP_n \longrightarrow \int_{\mathbb{H}} f dP \quad \forall f \in C_0(\mathbb{H}), \tag{1.32}$$

where $C_0(\mathbb{H})$ is the class of all bounded continuous, real valued functions on $\mathbb{H}$. If $T,\ T_1,\ T_2, \cdots$ are $\mathbb{H}$-valued random variables that induce the probability measures $P,\ P_1,\ P_2, \cdots$, we also say that the T_n converge in distribution to T since $T_n \xrightarrow{d} T$, as $n \to \infty$ in $\mathbb{H}$, if $P_n \xrightarrow{d} P$.

Theorem 1.3.1 Continuous Mapping Theorem. Let $\mathbb{H}$ be a separable Hilbert space over the real and $\varphi : \mathbb{H} \longrightarrow \mathbb{R}$ continuous. Then, we have

$$\varphi(T_n) \xrightarrow{d} \varphi(T), \quad \text{as} \quad n \to \infty, \quad \text{in } (\mathbb{R}, \mathcal{B}). \tag{1.33}$$

Now suppose that $\mathcal{P} = \{P_1, P_2, \cdots\}$ is tight and that there exists a function

$\varphi : \mathbb{H} \to \mathbb{C}$ such that ofr each $t \in \mathbb{H}$

$$\varphi_n(t) = \int_{x\in\mathbb{H}} e^{i\langle t,x\rangle} dP_n(x) \longrightarrow \varphi(t), \quad \text{as} \quad n \to \infty, \tag{1.34}$$

then there exists a probability measure P on $(\mathbb{H}, \mathcal{B})$ such that

$$P_n \xrightarrow{d} P, \quad \text{as} \quad n \to \infty, \tag{1.35}$$

and

$$\varphi(t) = \int_{x\in\mathbb{H}} e^{i\langle t,x\rangle} dP(x), \quad t \in \mathbb{H}. \tag{1.36}$$

This follows from Laha & Rohatgi (1979), Proposition 7.4.2.

Quite often probability distributions are induced by random variables. We will now formulate a convenient sufficient condition for convergence in distribution of $\mathbb{H}$-valued random variables. The proof is almost immediate from the preceding remark.

Theorem 1.3.2 Let $(T_n)_{n\geq 1}$ be a sequence of $\mathbb{H}$-valued random variables such that

$$(T_n)_{n\geq 1} \text{ is tight } \& \ \langle T_n, h\rangle \xrightarrow{d} U_h, \text{ as } \ n \to \infty \ \text{ in } \mathbb{R}, \ \forall\ h \in \mathbb{H}, \tag{1.37}$$

where U_h is some real valued random variable. Then there exists an $\mathbb{H}$-valued random variable T such that

$$T_n \xrightarrow{d} T, \text{ as } \ n \to \infty, \text{ in } \mathbb{H}. \tag{1.38}$$

A simple sufficient condition for tightness turns out to be the requirement that

$$\sup_{n\in\mathbb{N}} \sum_{m=M}^{\infty} \mathbb{E}\langle T_n, e_m\rangle^2 \longrightarrow 0, \text{ as } \ M \to \infty, \tag{1.39}$$

where $e_1, e_2, \cdots$ is an orthonormal basis of $\mathbb{H}$. In fact, the property is independent of the choice of basis. In conjunction with Theorem 1.3.2, this yields the following result.

Theorem 1.3.3 Let $(T_n)_{n\geq 1}$ be a sequence of $\mathbb{H}$-valued random variables such that for some orthogonal basis $e_1, e_2, \cdots$ of $\mathbb{H}$ we have

$$\sup_{n\in\mathbb{N}} \sum_{m=M}^{\infty} \mathbb{E}\langle T_n, e_m\rangle^2 \longrightarrow 0, \text{ as } M \to \infty, \tag{1.40}$$

and such that for each $h \in \mathbb{H}$ there exists a real valued random variable U_h such that

$$\langle T_n, h\rangle \xrightarrow{d} U_h, \text{ as } n \to \infty, \text{ in } \mathbb{R}. \tag{1.41}$$

Then there exists an $\mathbb{H}$-valued random variable T such that

$$T_n \xrightarrow{d} T, \text{ as } n \to \infty, \text{ in } \mathbb{H}. \tag{1.42}$$

This last result is particularly useful to establish convergence in distribution in not too complicated cases. We will apply this result later in Chapter 2. Here we will first illustrate its usefulness by establishing a version of the central limit theorem which is at once very simple and of great importance, since it yields the asymptotic distribution of the suitably standardized sample mean.

Let $X_1, \cdots, X_n$ be independent copies of a random element $X : \Omega \longrightarrow \mathbb{H}$ with probability distribution $\mathbb{P}_X = P$. Then, these n random elements are also called a random sample of size n from P. If $\mathbb{E}\|X\|^2 < \infty$, so that the mean μ and covariance operator Σ of X are well defined, we may write

$$X_1, \cdots, X_n \text{ i.i.d. } (\mu, \Sigma). \tag{1.43}$$

The sample mean is, just as in the finite dimensional case, defined as

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.44}$$

Since the expectation is linear, it follows that

$$\mathbb{E}(\overline{X}) = \mu. \tag{1.45}$$

Because, by the independence of the X_i

$$\begin{aligned}\mathbb{E}\left\langle x, \overline{X}-\mu\right\rangle\left\langle y, \overline{X}-\mu\right\rangle &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left\langle x, X_i-\mu\right\rangle\left\langle y, X_j-\mu\right\rangle \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\left\langle x, X_i-\mu\right\rangle\left\langle y, X_i-\mu\right\rangle \\ &= \frac{1}{n}\left\langle x, \Sigma y\right\rangle, \quad \forall x, y \in \mathbb{H},\end{aligned} \tag{1.46}$$

it follows that the covariance operator of $\overline{X}$ equals $\frac{1}{n}\Sigma$. When we define the random element

$$\begin{aligned}G_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \\ &= \sqrt{n}(\bar{X}-\mu),\end{aligned} \tag{1.47}$$

we have the central limit theorem in $\mathbb{H}$.

Theorem 1.3.4 Central Limit Theorem in $\mathbb{H}$. Let $X_1, X_2, \cdots$ be independent copies of a random variable X in $\mathbb{H}$ with $\mathbb{E}\|X\|^2 < \infty$, and let $\mathbb{E}(X) = \mu$ and $\mathbb{E}(X-\mu)\otimes(X-\mu) = \Sigma$. Then we have

$$\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} G \overset{d}{=} \mathcal{G}(0, \Sigma), \quad \text{as } n \to \infty, \text{ in } \mathbb{H}. \tag{1.48}$$

Proof: The proof is based on Theorem 1.3.2. Let $e_1, e_2, \cdots$ be the orthonormal basis of eigenvectors of Σ. Then we have

$$\begin{aligned} \sup_{n\in\mathbb{N}} \sum_{m=M}^{\infty} \mathbb{E}\langle \sqrt{n}(\bar{X}-\mu), e_m\rangle^2 &= \sup_{n\in\mathbb{N}} \sum_{m=M}^{\infty} n\mathbb{E}\langle \bar{X}-\mu, e_m\rangle^2 \\ &= \sup_{n\in\mathbb{N}} \sum_{m=M}^{\infty} n\frac{1}{n}\langle e_m, \Sigma e_m\rangle \\ &= \sum_{m=M}^{\infty} \langle e_m, \Sigma e_m\rangle \\ &= \sum_{m=M}^{\infty} \sigma_m^2 \longrightarrow 0, \text{ as } M\to\infty, \end{aligned} \tag{1.49}$$

because Σ has finite trace. This settles the tightness.

Next, let us choose an arbitrary $h \in \mathbb{H}$, and note that the $\langle X_i - \mu, h\rangle$ are i.i.d. real valued random variables with mean 0 and variance $\langle h, \Sigma h\rangle$. Hence the central limit theorem in one dimension yields the existence of a real valued random variable U_h such that

$$\langle \sqrt{n}(\bar{X}-\mu), h\rangle \xrightarrow{d} U_h \stackrel{d}{=} \mathcal{N}(0, \langle h, \Sigma h\rangle), \quad \text{as } n\to\infty, \text{ in } \mathbb{R}. \tag{1.50}$$

This settles the convergence requirement.

Consequently, there exists a random element G in $\mathbb{H}$, such that

$$\sqrt{n}(\bar{X}-\mu) \xrightarrow{d} G, \quad \text{as } n\to\infty, \text{ in } \mathbb{H}. \tag{1.51}$$

It follows from the continuous mapping theorem (Theorem 1.3.1) that

$$\langle \sqrt{n}(\bar{X}-\mu), h\rangle \xrightarrow{d} \langle G, h\rangle, \quad \text{as } n\to\infty, \text{ in } \mathbb{N}. \tag{1.52}$$

But this means that

$$\langle G, h\rangle \stackrel{d}{=} \mathcal{N}(0, \langle h, \Sigma h\rangle). \tag{1.53}$$

It now follows from Proposition 1.2.2 that

$$G \overset{d}{=} \mathcal{G}(0, \Sigma), \tag{1.54}$$

and the proof is complete.

Corollary 1.3.1 Weak Law of Large Number (WLLN).

$$\mathbb{E}\{\|\overline{X} - \mu\| \geq \epsilon\} \longrightarrow 0, \quad \text{as} \quad n \to \infty, \quad \forall \epsilon > 0; \tag{1.55}$$

in other words, we have

$$\overline{X} \xrightarrow{p} \mu, \quad \text{as} \quad n \to \infty. \tag{1.56}$$

Definition 1.3.2 A mapping $\varphi : \mathbb{H} \longrightarrow \mathbb{R}$ is Fréchet differentiable at $\mu \in \mathbb{H}$ with respect to $\mathbb{H}$ if there exists a bounded, linear functional, represented by φ_μ^*, such that

$$|\varphi(\mu + h) - \varphi(\mu) - \langle h, \varphi_\mu^* \rangle| = o(\|h\|_{\mathbb{H}}), \quad \text{as} \quad h \to 0. \tag{1.57}$$

Corollary 1.3.2 Delta method. Let $\varphi : \mathbb{H} \longrightarrow \mathbb{R}$ be Fréchet differentiable at μ with respect to $\mathbb{H}$. Denote the derivative at μ by φ_μ^*. Then we have

$$\sqrt{n}\{\varphi(\overline{X}) - \varphi(\mu)\} \xrightarrow{d} \langle G, \varphi_\mu^* \rangle, \quad \text{as} \quad n \to \infty, \quad \text{in } (\mathbb{R}, \mathcal{B}). \tag{1.58}$$

Proof: The differentiability entails that

$$\begin{aligned} |\sqrt{n}\{\varphi(\overline{X}) - \varphi(\mu) - \varphi_\mu^*(\overline{X} - \mu)\}| &= |\sqrt{n}\{\varphi(\mu + \overline{X} - \mu) - \varphi(\mu) - \varphi_\mu^*(\overline{X} - \mu)\}| \\ &= \sqrt{n}\ o(\|\overline{X} - \mu\|_{\mathbb{H}}). \end{aligned} \tag{1.59}$$

Because $\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} G$, and $o(x) = \|x\|$, is continuous. It follows from Theorem

1.3.1 that $\|\sqrt{n}(\overline{X}-\mu)\|_{\mathbb{H}} = O_p(1)$. Consequently,

$$\begin{aligned} \sqrt{n}\ o_p(\|\overline{X}-\mu\|_{\mathbb{H}}) &= o_p(\sqrt{n}(\overline{X}-\mu)_{\mathbb{H}}) \\ &= o_p(1), \quad \text{as} \quad n \to \infty. \end{aligned} \tag{1.60}$$

Since φ_μ^* is continuous and linear, it follows from Theorem 1.3.1 again that

$$\varphi_\mu^*(\overline{X}-\mu) = \varphi_\mu^*\sqrt{n}(\overline{X}-\mu) \xrightarrow{d} \left\langle G, \varphi_\mu^* \right\rangle, \quad \text{as} \quad n \to \infty, \quad \text{in } (\mathbb{H}, \mathcal{B}). \tag{1.61}$$

Combination yields the desired result.

CHAPTER 2

GENERAL LINEAR MODEL IN $\mathbb{H}$

Let us introduce a new notation

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix}_{q\times\infty} \in \mathbb{H}^q \text{ and } b_i \in \mathbb{H}. \tag{2.1}$$

This notation has similar meaning and properties of matrices in finite Euclidean spaces but we add ∞ symbol to it to remind us of the infinite-dimensional properties of Hilbert spaces and we also define

$$A_{p\times q}\ \mathbf{b}_{q\times\infty} = \begin{pmatrix} \sum_{j=i}^{q} a_{1j}b_j \\ \vdots \\ \sum_{j=i}^{q} a_{pj}b_j \end{pmatrix}_{p\times\infty} \in \mathbb{H}^q \tag{2.2}$$

in which

$$A_{p\times q} = \begin{pmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pq} \end{pmatrix} \text{ and } a_{ij} \in \mathbb{R}. \tag{2.3}$$

The model that we are considering in this section is

$$\mathbf{X}_{n\times\infty} = D_{n\times p}\ \mathbf{f}_{p\times\infty} + \epsilon_{n\times\infty} \tag{2.4}$$

where $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^*$ has i.i.d. $(0, \Sigma)$ components in $\mathbb{H}$. Similar to the univariate linear model, we will consider the estimator $\hat{\mathbf{f}}$ of $\mathbf{f}$ given by

$$\begin{aligned} \hat{\mathbf{f}}_{p\times\infty} = \begin{pmatrix} \hat{f}_1 \\ \vdots \\ \hat{f}_p \end{pmatrix}_{p\times\infty} &= (D^*D)^{-1}_{p\times p}\ D^*_{p\times n} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}_{n\times\infty} \\ &= (D^*D)^{-1}_{p\times p}\ D^*_{p\times n}\mathbf{X}_{n\times\infty} \in \mathbb{H}^P, \end{aligned} \tag{2.5}$$

exploiting the notational conventions introduced above. The main purpose of this section is to show that this estimator has approximately a Gaussian distribution for large sample size n. In order to prepare for this result, let us recall first prove the Cramér-Wold device and the Hájek-Sidák central limit theorem for $\mathbb{H}$-valued random variables.

Theorem 2.1.1 Cramér-Wold in $\mathbb{H}^p$. Let $\mathbf{H} = (H_1, \cdots, H_p)^*$ be a random variable in $\mathbb{H}^p$, $\mathbf{T_n} = (T_{n1}, \cdots, T_{np})^*$, $n \in \mathbb{N}$, a sequence of random variables in $\mathbb{H}^p$ such that

$$\lambda^*\mathbf{T_n} \xrightarrow{d} \lambda^*\mathbf{H}, \quad \text{as} \quad n \to \infty, \quad \text{in } \mathbb{H} \quad \text{and} \quad \forall\ \lambda \in \mathbb{R}^p. \tag{2.6}$$

Then we have

$$\mathbf{T_n} \xrightarrow{d} \mathbf{H}, \quad \text{as} \quad n \to \infty, \quad \text{in } \mathbb{H}^p. \tag{2.7}$$

Proof: Let $\mathbf{h} = (h_1, \cdots, h_p)^* \in \mathbb{H}^p$. Assumption (2.6) entails in particular that

$$T_{nj} \xrightarrow{d} H_j, \quad \text{as} \quad n \to \infty, \text{in } \mathbb{H} \tag{2.8}$$

Applying the continuous mapping theorem, then yields

$$\langle T_{nj}, h_j\rangle \xrightarrow{d} \langle H_j, h_j\rangle, \quad \text{as} \quad n \to \infty, \quad \text{in } \mathbb{R}, \quad \text{for} \quad j = 1, \cdots, p, \tag{2.9}$$

and

$$\langle \mathbf{T_n}, \mathbf{h} \rangle_p = \sum_{j=1}^{p} \langle T_{nj}, h_j \rangle \xrightarrow{d} \sum_{j=1}^{p} \langle H_j, h_j \rangle = \langle \mathbf{H}, \mathbf{h} \rangle_p, \quad \text{as } n \to \infty, \quad \text{in } \mathbb{R}, \tag{2.10}$$

This settle one of the conditions of (1.37) of Theorem 1.3.2.

To verify the tightness, let us first observe that (2.8) implies that $(T_{nj} \in K_j)_{n\geq 1}$ is tight for each j. Hence, given an arbitrary $\epsilon > 0$, ther exists compact sets $K_j \subset \mathbb{H}$ such that

$$\mathbb{P}\{T_{nj} \in K_j\} \geq 1 - \frac{1}{p}\epsilon, \quad \forall n \in \mathbb{N}. \tag{2.11}$$

But then the set $\mathbb{K} = K_1 \times \cdots \times K_p$ is compact in $\mathbb{H}^p$, and

$$\begin{aligned} \mathbb{P}\{\mathbf{T_n} \notin \mathbb{K}\} &= \mathbb{P}(\bigcup_{j=1}^{p} \{T_{nj} \notin K_j\}) \\ &\leq \sum_{j=1}^{p} \mathbb{P}\{T_{nj} \notin K_j\} \\ &\leq p \cdot \frac{1}{p} \cdot \epsilon \\ &= \epsilon. \end{aligned} \tag{2.12}$$

Tightness of $(\mathbf{T_n})_{n\geq 1}$ follows and the condition (1.37) of Theorem 1.3.2 is apparently fulfilled, and the proof is complete with an application of that theorem.

Theorem 2.1.2 Hájek-Sidák Central Limit Theorem in $\mathbb{H}$. Let $a_{n1}, \cdots, a_{nn}$ be a triangular array of real numbers satisfying the Noether condition, i.e. such that

$$\max_{1\leq i\leq n} a_{ni}^2 / (\sum_{i=1}^{n} a_{ni}^2) \longrightarrow 0, \quad \text{as } n \to \infty. \tag{2.13}$$

Let $\xi_1, \xi_2, \cdots$ be a sequence of i.i.d. random variables in $\mathbb{H}$ with $\mathbb{E}\|\xi_i\|^2 < \infty$, and

$$\mathbb{E}\xi_i = \mu \in \mathbb{H}, \quad \mathbb{E}(\xi_i - \mu) \otimes (\xi_i - \mu) = \Sigma. \tag{2.14}$$

Then there exists a Gaussian random variable $G \in \mathbb{H}$, with $\mathbb{E}G = 0$ and $\mathbb{E}G \otimes G = \Sigma$, such that

$$T_n = \sum_{i=1}^{n} \frac{a_{ni}}{\sqrt{\sum_{i=1}^{n} a_{ni}^2}} (\xi_i - \mu) \xrightarrow{d} G, \quad \text{as } n \to \infty, \quad \text{in } \mathbb{H}. \tag{2.15}$$

Proof: Let us write

$$c_i = \frac{a_{ni}}{\sqrt{\sum_{i=1}^{n} a_{ni}^2}}, \tag{2.16}$$

and note that

$$\sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n} \frac{a_{ni}^2}{\sum_{i=1}^{n} a_{ni}^2} = 1. \tag{2.17}$$

For an arbitrary $h \in \mathbb{H}$, the $\langle \xi_i - \mu, h \rangle$ are i.i.d. real valued random variables with mean 0 and variance $\langle h, \Sigma h \rangle$. The traditional Hájek-Sidák central limit theorem entails at once that

$$\langle T_n, h \rangle = \sum_{i=1}^{n} c_i \langle \xi_i - \mu, h \rangle \xrightarrow{d} \mathcal{N}(0, \langle h, \Sigma h \rangle), \quad \text{as } n \to \infty, \tag{2.18}$$

so that condition (1.37) of Theorem 1.3.2 is satisfied. Rather than verifying condition (1.37) of that theorem, let us consider the sufficient conditions (1.40) and (1.41) for the tightness in Theorem 1.3.3. Given an orthonormal basis $e_1, e_2, \cdots$ of

$\mathbb{H}$, we have

$$\begin{aligned}
\sum_{m=M}^{\infty} \mathbb{E}\langle T_n, e_m\rangle^2 &= \sum_{m=M}^{\infty} \mathbb{E}\langle \sum_{i=1}^{n} c_i(\xi_i - \mu), e_m\rangle^2 \\
&= \sum_{m=M}^{\infty} \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \mathbb{E}\langle e_m, \xi_i - \mu\rangle\langle \xi_j - \mu, e_m\rangle \\
&= (\sum_{i=1}^{n} c_i^2) \sum_{m=M}^{\infty} \langle e_m, \Sigma e_m\rangle \\
&= \sum_{m=M}^{\infty} \langle e_m, \Sigma e_m\rangle \longrightarrow 0, \quad \text{as} \quad M \to \infty,
\end{aligned} \tag{2.19}$$

because Σ has finite trace. Tightness follows because the last expression in (2.19) does not depend on n. We may now conclude from Theorem 2.1.1 that there exists a random variable G in $\mathbb{H}$ such that

$$T_n \xrightarrow{d} G \quad \text{as} \quad n \to \infty, \quad \text{in } \mathbb{H}. \tag{2.20}$$

It remains to show that G is Gaussian $\mathcal{G}(0, \Sigma)$. According to (2.18), G has the property

$$\langle G, h\rangle \stackrel{d}{=} \mathcal{N}(0, \langle h, \Sigma h\rangle), \quad \forall h \in \mathbb{H}. \tag{2.21}$$

But then G must be Gaussian by Theorem 1.3.4. Knowing that G is Gaussian, (2.21) finally entails that mean and covariance operators have to be 0 and Σ respectively.

We are now ready for the main theorem of this section. Recall that d_i^*, the i-th row of the matrix D. As for the asymptotics in the univariate linear model, it will be assumed that

$$\max_{i=1,\cdots,n} d_i^*(D^*D)^{-1}d_i \longrightarrow 0, \quad \text{as} \quad n \to \infty, \tag{2.22}$$

$$n(D^*D)^{-1} \longrightarrow V, \quad \text{as} \quad n \to \infty \tag{2.23}$$

where V is a strictly positive symmetric $p \times p$ matrix. Recall that the random errors $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. with mean 0 and covariance operator Σ. For brevity, let us write

$$V \otimes_p \Sigma = \begin{pmatrix} V_{11}\Sigma & \cdots & V_{1p}\Sigma \\ \vdots & & \vdots \\ V_{p1}\Sigma & \cdots & V_{pp}\Sigma \end{pmatrix} \tag{2.24}$$

The right hand side of (2.24) is a covariance operator on $\mathbb{H}^p$.

Theorem 2.1.3 Under the assumptions (2.22) and (2.23), there exists a Gaussian random variable $\mathbf{G}$ in $\mathbb{H}^p$, such that

$$\sqrt{n}(\hat{\mathbf{f}} - \mathbf{f}) \xrightarrow{d} \mathbf{G}, \quad \text{as} \;\; n \to \infty, \quad \text{in } \mathbb{H}^p. \tag{2.25}$$

Moreover, $\mathbf{G}$ has mean 0 and covariance operator $V \otimes_p \Sigma$.

Proof: we can write $\sqrt{n}(\hat{\mathbf{f}} - \mathbf{f}) = \sqrt{n}(D^*D)^{-1}D^*\epsilon$, so that for an arbitrary $\lambda \in \mathbb{R}^p$ we have

$$\begin{aligned} \sqrt{n}\lambda^*(\hat{\mathbf{f}} - \mathbf{f}) &= \sqrt{n}\sum_{j=1}^{p} \lambda_j(\hat{f}_j - f_j) \\ &= \sqrt{n}\lambda^*(D^*D)^{-1}D^*\epsilon \\ &= a_n^*(\lambda)\epsilon \\ &= \sum_{i=1}^{n} a_{ni}(\lambda)\epsilon_i \end{aligned} \tag{2.26}$$

where $a_n(\lambda)$ is a vector of n real numbers given by

$$a_n(\lambda) = \sqrt{n}D(D^*D)^{-1}\lambda, \tag{2.27}$$

and

$$a_{ni}(\lambda) = \sqrt{n}\; d_i^*(D^*D)^{-1}\lambda. \tag{2.28}$$

Next observe that

$$\begin{aligned}
\max_{\lambda\neq 0}\frac{\max_{i=1,\cdots,n} a_{ni}^2(\lambda)}{\sum_{i=1}^n a_{ni}^2(\lambda)} &= \max_{i=1,\cdots,n}\max_{\lambda\neq 0}\frac{n\lambda^*(D^*D)^{-1}(d_i d_i^*)(D^*D)^{-1}\lambda}{n\lambda^*(D^*D)^{-1}\lambda} \\
&= \max_{i=1,\cdots,n}\{\text{largest eigenvalue of}(D^*D)^{-\frac{1}{2}}(d_i d_i^*)(D^*D)^{-\frac{1}{2}}\} \\
&= \max_{i=1,\cdots,n}\operatorname{tr}((D^*D)^{-\frac{1}{2}}(d_i d_i^*)(D^*D)^{-\frac{1}{2}}) \\
&= \max_{i=1,\cdots,n} d_i^*(D^*D)^{-1}d_i \longrightarrow 0, \quad \text{as} \quad n\to\infty,
\end{aligned} \tag{2.29}$$

by assumption (2.22). Here we have the matrix $d_i d_i^*$ has range of dimension at most 1 (in fact it is equal to 1, since $d_i \neq 0$ because D is of full rank p). This means that for each λ, the numbers $a_{ni}(\lambda)$ satisfy the Noether condition so that, by Theorem 2.1.2, it follows that

$$\sum_{i=1}^n \frac{a_{ni}(\lambda)}{\sqrt{\sum_{i=1}^n a_{ni}^2(\lambda)}}\,\epsilon_i \xrightarrow{d} \mathcal{G}(0,\Sigma), \quad \text{as} \quad n\to\infty, \ \text{in } \mathbb{H}. \tag{2.30}$$

Because $\sum_{i=1}^n a_{ni}^2(\lambda) = n\lambda^*(D^*D)^{-1}\lambda$ and it is immediate from (2.26), (2.30), and assumption (2.23) that

$$\sqrt{n}\sum_{j=1}^p \lambda_j(\hat{f}_j - f_j) \xrightarrow{d} \mathcal{G}(0,(\lambda^*V\lambda)\Sigma). \tag{2.31}$$

Let us now introduce a random variable $\mathbf{G}$ such that

$$\mathbf{G} = (G_1,\cdots,G_p)^* \stackrel{d}{=} \mathcal{G}(\mathbf{0}, V\otimes_p \Sigma); \tag{2.32}$$

see also (2.24). It is clear that

$$\begin{aligned}
\mathbb{E}(\sum_{j=1}^p \lambda_j G_j)\otimes(\sum_{k=1}^p \lambda_k G_k) &= \sum_{j=1}^p\sum_{k=1}^p \lambda_j(\mathbb{E}G_j\otimes G_k)\lambda \\
&= \sum_{j=1}^p\sum_{k=1}^p(\lambda V_{jk}\lambda_k)\Sigma \\
&= (\lambda^*V\lambda)\Sigma.
\end{aligned} \tag{2.33}$$

In other words, we may write

$$\sqrt{n}\sum_{j=1}^{p}\lambda_j(\hat{f}_j - f_j) \xrightarrow{d} \sum_{j=1}^{p}\lambda_j G_j, \quad \text{as} \quad n \to \infty, \quad \text{in } \mathbb{H}. \tag{2.34}$$

An application of Theorem 2.1.1 yields the claims of Theorem 2.1.3.

Let next consider the estimation of the common covariance operator Σ of the errors ϵ_i. Again the construction of this estimator is very similar to that of covariance matrix in multivariate analysis. Let us introduce the matrix

$$H = D(D^*D)^{-1}D^*, \tag{2.35}$$

which is the matrix of the orthogonal projection onto the subspace spanned by the columns of D (the so called hat matrix). Then $I_n - H$ is the projection onto the orthogonal complement of that subspace. This geometric interpretation entails at once that

$$(I_n - H)D = 0. \tag{2.36}$$

Furthermore, let us introduce the

$$\textit{predicted values} \;\; \hat{\mathbf{X}} = D\hat{\mathbf{f}} = H\mathbf{X}, \tag{2.37}$$

$$\textit{residuals} \;\; \hat{\epsilon} = \mathbf{X} - \hat{\mathbf{X}} = (I_n - H)\mathbf{X}. \tag{2.38}$$

Note that

$$\begin{aligned}\hat{\epsilon} &= (I_n - H)(D\mathbf{f} + \epsilon)\\ &= (I_n - H)\epsilon,\end{aligned} \tag{2.39}$$

because of (2.36), and that

$$\hat{\epsilon} = \begin{pmatrix} \hat{\epsilon_1} \\ \vdots \\ \hat{\epsilon_n} \end{pmatrix}_{p\times\infty}, \quad \hat{\epsilon_{\mathbf{i}}} = X_i - d_i^* \hat{\mathbf{f}}, \tag{2.40}$$

where again d_i^* is the i-th row of D. It should not be hard to show that

$$\hat{\Sigma} = \frac{1}{n-p} \sum_{i=1}^{n} \hat{\epsilon_i} \otimes \hat{\epsilon_i} \xrightarrow{p} \Sigma, \quad \text{as} \quad n \to \infty, \quad \text{in} \quad \mathbb{H}^2. \tag{2.41}$$

A canonical way to formulate a hypothesis testing problem is to partition the design matrix in two parts: $D = (D_1|D_2)$, where D_1 is $n \times p(1)$, D is $n \times p(2)$ and both are of full rank $p(1)$ and $p(2)$ respectively. Let us also partition $\mathbf{f}$ in two parts $\mathbf{f}_1$ and $\mathbf{f}_2$, where $\mathbf{f}_1$ is $p(1) \times \infty$ and $p(2) \times \infty$. We now have

$$\mathbf{X} = D_1\mathbf{f_1} + D_2\mathbf{f_2} + \epsilon. \tag{2.42}$$

The usual null hypothesis to be tested can be formulated as

$$H_0 : \mathbf{f_2} = 0. \tag{2.43}$$

For reason already explained above when dealing with special cases of the current model, we may prefer to consider the neighborhood hypothesis

$$H_\delta : \|\mathbf{f_2}\|^2_{p(2)} \leq \delta^2, \quad \text{for some} \quad \delta > 0. \tag{2.44}$$

Assuming that D_2 satisfies conditions similar to those for D, i.e.

$$\max_{i=1,\cdots,n} d_{2i}^*(D_2^*D_2)^{-1}d_{2i} \longrightarrow 0, \quad \text{as} \quad n \to \infty, \tag{2.45}$$

where d_{2i}^* is the i-th row of the matrix D_2, and

$$n(D_2^*D_2)^{-1} \longrightarrow V_2, \quad \text{as} \quad n \to \infty, \tag{2.46}$$

where V_2 is a strictly positive symmetric $p(2) \times p(2)$ matrix, we may infer at once from Theorem 2.1.3 that

$$\sqrt{n}(\hat{\mathbf{f}}_{\mathbf{2}} - \mathbf{f}_{\mathbf{2}}) \xrightarrow{d} \mathbf{G}_{\mathbf{2}}, \quad \text{as} \quad n \to \infty, \quad \text{in} \quad \mathbb{H}^{p(2)}, \tag{2.47}$$

where

$$\hat{\mathbf{f}}_{\mathbf{2}} = (D_2^* D_2)^{-1} D_2^* \mathbf{X}, \quad \text{and} \quad \mathbf{G}_{\mathbf{2}} \stackrel{d}{=} \mathcal{G}(0, V_2 \otimes_{p(2)} \Sigma). \tag{2.48}$$

Let us consider the functional

$$\varphi(\mathbf{x}) = \|\mathbf{x}\|_{p(2)}^2, \quad \mathbf{x} \in \mathbb{H}^{p(2)}. \tag{2.49}$$

Notice that φ has a Fréchet derivative at $\mathbf{0} \in \mathbb{H}^{p(2)}$, given by

$$\varphi_{\mathbf{0}}'(\mathbf{x}) = 2\langle \mathbf{x}, \mathbf{0} \rangle_{p(2)} = 2 \sum_{j=1}^{p(2)} \langle x_j, a_j \rangle. \tag{2.50}$$

It follows from the delta method in Corollary 1.3.3 that

$$\sqrt{n}(\|\hat{\mathbf{f}}_{\mathbf{2}}\|_{p(2)}^2 - \|\mathbf{f}_{\mathbf{2}}\|_{p(2)}^2) \xrightarrow{d} 2\langle \mathbf{G}_{\mathbf{2}}, \mathbf{f}_{\mathbf{2}} \rangle_{p(2)}, \quad \text{as} \quad n \to \infty, \tag{2.51}$$

We see from (2.48) that

$$2\langle \mathbf{G}_{\mathbf{2}}, \mathbf{f}_{\mathbf{2}} \rangle_{p(2)} \stackrel{d}{=} \mathcal{N}(0, v^2), \tag{2.52}$$

where

$$\begin{aligned} v^2 &= 4\mathbb{E}\{2 \sum_{j=1}^{p(2)} \langle G_{2j}, f_{2_j} \rangle \}^2 \\ &= 4 \sum_{j=1}^{p(2)} \sum_{k=1}^{p(2)} \langle f_{2j}, V_{2|jk} \Sigma f_{2k} \rangle, \end{aligned} \tag{2.53}$$

and $V_{2|jk}$ is the jk element of the matrix V_2. It should be possible to consistently estimate this variance by an estimator $\hat{v}^2$ that can be obtained from (2.53) by using the estimators $\hat{\mathbf{f}}_{\mathbf{2}}$ and $\hat{\Sigma}$ in (2.48) and (2.39) respectively.

Assuming $v^2 > 0$, we arrive at

$$\sqrt{n}(\|\hat{\mathbf{f}}_{\mathbf{2}}\|^2_{p(2)} - \|\mathbf{f}_{\mathbf{2}}\|^2_{p(2)})/\hat{v} \xrightarrow{d} \mathcal{N}(0,1), \tag{2.54}$$

With the help of (2.54), we can obtain an asymptotic level α test by rejecting H_δ in (2.44) if and only if

$$\sqrt{n}(\|\hat{\mathbf{f}}_{\mathbf{2}}\|^2_{p(2)} - \delta^2)/\hat{v} > \Phi^{-1}(1-\alpha), \quad 0 < \alpha < 1, \tag{2.55}$$

where Φ is the standard normal c.d.f.. We will not provide the details for this general procedure but rather focus on some important special cases in Section 3.1 and 3.3 below.

CHAPTER 3

SPECIAL CASES IN $\mathbb{H}$

3.1 One-Sample Problem with Neighborhood Hypothesis

In this section, we will define estimators for the mean and covariance operator in $\mathbb{H}$. Also we develop a test for the one-sample problem in $\mathbb{H}$. Let us consider a random sample $X_1, \cdots, X_n$ of independent copies of a Hilbert space valued random variable X with $\mathbb{E}\|X\|^2 < \infty$. Denote $\mu = \mathbb{E}X$ and let Σ be the covariance operator of X. Suppose we want to test the null hypothesis

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0, \tag{3.1}$$

for a given vector $\mu_0 \in \mathbb{H}$. Let us introduce the notation

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.2}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}) \otimes (X_i - \bar{X}) \tag{3.3}$$

for sample mean and sample covariance operator respectively.

An important result in the development for the test procedure below is the well known fact that under the null hypothesis.

$$\sqrt{n}(\bar{X} - \mu_0) \xrightarrow{d} G, \quad \text{as} \quad n \to \infty, \quad \text{in} \quad \mathbb{H}. \tag{3.4}$$

Let us assume that Σ has spectral decomposition

$$\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 e_k \otimes e_k, \tag{3.5}$$

where the eigenvalues $\sigma_1^2 \geq \sigma_2^2 \geq \cdots$ are repeated according to their finite multiplicities, and the $e_1, e_2, \cdots$ form an orthogonal basis of eigenvectors. The mean of G and covariance operator of G are 0 and Σ respectively. In fact, the mean and covariance operator of G are the same as the mean and covariance of the generic term $X - \mu_0$ of the standardized sum on the left in (3.4).

As a test statistic, we propose

$$T_n = n\|(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\{(\bar{X} - \mu_0) \otimes (\bar{X} - \mu_0)\}(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\|_{\mathcal{L}}, \tag{3.6}$$

which is derived from finite dimension situation. It follows from (3.4) that, under H_0,

$$(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0) \xrightarrow{d} (\alpha I + \Sigma)^{-\frac{1}{2}}G = H_\alpha, \quad \text{as } n \to \infty, \ \text{in } \mathbb{H}, \tag{3.7}$$

and, since the mapping $\varphi : \mathbb{H} \to \mathcal{L}$ given by

$$\varphi(x) = x \otimes x, \quad x \in \mathbb{H}, \tag{3.8}$$

is continuous, we also have

$$\begin{aligned}\{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\} \otimes \{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\} \\ \xrightarrow{d} H_\alpha \otimes H_\alpha, \quad \text{as } n \to \infty \text{ in } \mathcal{L}.\end{aligned} \tag{3.9}$$

The continuity of φ in (3.9) follows from

$$\begin{aligned}\|x \otimes x - y \otimes y\|_{\mathcal{L}} &= \|(x - y) \otimes x + y \otimes (x - y)\|_{\mathcal{L}} \\ &\leq \|(x - y) \otimes x\|_{\mathcal{L}} + \|y \otimes (x - y)\|_{\mathcal{L}} \\ &= \|x - y\|\|x\| + \|y\|\|x + y\| \longrightarrow 0, \quad \text{as} \ \ y \to x.\end{aligned} \tag{3.10}$$

Next observe that the mapping $\varphi(T) = \|T\|$, is continuous for $T \in \mathcal{L}$, so that (3.9) entails

$$\begin{aligned}\|\{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\} \otimes \{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\}\|_{\mathcal{L}} \\ \xrightarrow{d} \|H_\alpha \otimes H_\alpha\|_{\mathcal{L}}, \quad \text{as} \ \ n \to \infty.\end{aligned} \tag{3.11}$$

The continuity of this mapping follows from

$$|\varphi(u) - \varphi(v)| = |\|u\|_{\mathcal{L}} - \|v\|_{\mathcal{L}}| \leq \|u - v\|_{\mathcal{L}}, \quad u, v \in \mathcal{L}. \tag{3.12}$$

We can rewrite the left hand side of (3.11) and see that in fact it equals (3.6). Indeed,

$$\begin{aligned} &\|\{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\} \otimes \{(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\sqrt{n}(\bar{X} - \mu_0)\}\|_{\mathcal{L}} \\ &\qquad = n\|(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\{(\bar{X} - \mu_0) \otimes (\bar{X} - \mu_0)\}(\alpha I + \hat{\Sigma})^{-\frac{1}{2}}\|_{\mathcal{L}}, \end{aligned} \tag{3.13}$$

so that we have shown that

$$T_n \xrightarrow{d} \|H_\alpha \otimes H_\alpha\|_{\mathcal{L}} = \|H_\alpha\|^2. \tag{3.14}$$

In order to specify the limiting distribution i.e. the distribution of $\|H_\alpha\|^2$, let us note that G is a Gaussian random element in $\mathbb{H}$ with mean $\mathbb{E}G = 0$ and with the covariance operator Σ, the common covariance operator of the data. Recall (3.5) and the fact that the process G has the Karhunen Loève expansion

$$G = \sum_{k=1}^{\infty} \sqrt{\sigma_k^2} Z_k e_k, \tag{3.15}$$

where the $Z_1, Z_2, \cdots$ are i.i.d. $\mathcal{N}(0,1)$. Indeed, we have $\mathbb{E}G = 0$, and the covariance operator of G equals Σ. To see the latter, note that on the one hand

$$\begin{aligned} \mathbb{E}\langle a, \sum_{k=1}^{\infty} \sigma_k Z_k e_k\rangle\langle\sum_{j=1}^{\infty} \sigma_j Z_j e_j, b\rangle &= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \sigma_k\sigma_j\langle a, e_k\rangle\langle e_j, b\rangle\mathbb{E}Z_k Z_j \\ &= \sum_{k=1}^{\infty} \sigma_k^2\langle a, e_k\rangle\langle e_k, b\rangle, \end{aligned} \tag{3.16}$$

and on the other hand

$$
\begin{aligned}
\langle a, \Sigma\rangle &= \langle a, \sum_{k=1}^{\infty} \sigma_k^2 (e_k \otimes e_k) b\rangle \\
&= \sum_{k=1}^{\infty} \sigma_k^2 \langle a, e_k\rangle \langle e_k, b\rangle,
\end{aligned}
\tag{3.17}
$$

for any $a, b \in \mathbb{H}$. This proves the result according to (1.3). Consequently,

$$
\begin{aligned}
H_\alpha &= (\alpha I + \Sigma)^{-\frac{1}{2}} G \\
&= \sum_{k=1}^{\infty} \sqrt{\frac{\sigma_k^2}{(\alpha + \sigma_k^2)}} Z_k e_k,
\end{aligned}
\tag{3.18}
$$

and, finally,

$$
\|H_\alpha\|^2 = \sum_{k=1}^{\infty} \frac{\sigma_k^2}{(\alpha + \sigma_k^2)} Z_k^2. \tag{3.19}
$$

To work with this limiting distribution in practice, we need to know the σ_k^2, or at least be able to estimate them. There are other ways to attack the problem, However, one of the ways is that we can construct an asymptotic size α test by modifying the usual hypothesis. This modified hypothesis may make more sense from an applied point of view and leads, moreover, to a simpler statistics and simpler asymptotics. Suppose that we want to test the hypothesis that $\mu \in M \subset \mathbb{H}$, is a subspace of finite dimension $m \in \{0, 1, 2, \cdots\}$. Note that $M = \{0\}$ leads to $\mu = 0$. Let us denote the orthogonal projection onto M by Π, and onto $M^\perp$ by $\Pi^\perp$. It is useful to observe that

$$
\begin{aligned}
\left\langle \Pi^\perp x, \Pi^\perp y \right\rangle &= \left\langle x, (\Pi^\perp) \Pi^\perp y \right\rangle \\
&= \left\langle x, \Pi^\perp y \right\rangle, \quad \forall x, y \in \mathbb{H},
\end{aligned}
\tag{3.20}
$$

because projection $\Pi^\perp$ is Hermitian and idempotent. Furthermore, let us introduce the functional

$$\varphi_M(x) = \|x - M\|^2, \quad x \in \mathbb{H}, \tag{3.21}$$

representing the squared distance of a point in $\mathbb{H}$ to M (finite dimensional subspaces are closed). Note that $\varphi_M(x) = \left\|\Pi^{\perp}x\right\|^2$. The Neighborhood Hypothesis to be tested is

$$H_\delta : \mu \in M_\delta \cup B_\delta, \quad \text{for some} \;\; \delta > 0, \tag{3.22}$$

where
$M_\delta = \{x \in \mathbb{H} : \varphi_M(x) < \delta^2\}$ & $B_\delta = \{x \in \mathbb{H} : \varphi_M(x) = \delta^2, \left\langle \Pi^{\perp}x, \Sigma\Pi^{\perp}x \right\rangle > 0\}$.
The alternative to (3.22) is

$$A_\delta : \;\mu \; \in M_\delta^c \cap B_\delta^c. \tag{3.23}$$

The usual hypothesis would have been $\mu \in M$. It should be noted that H_δ contains $\{\varphi_M < \delta^2\}$ and that A_δ contains $\{\varphi_M > \delta^2\}$. These are the important components of the hypothesis; the set B_δ is added to the null hypothesis because the asymptotic power on that set is precisely α, as will be seen below.

For testing hypotheses like (3.22) see Dette & Munk (1998). These authors also observe that testing

$$H'_\delta : \mu \in M_\delta^c \cup B_\delta \;\; \text{versus} \;\; A'_\delta : \mu \in M_\delta \cap B_\delta^c, \tag{3.24}$$

can be done in essentially the same manner. This may be very useful in practice. When, for instance, M is the subspace of all polynomials of degree at most $m - 1$, it is more appropriate to test (3.24) if one wants to establish that the mean value function is close to such a polynomial. In the traditional set-up interchanging null hypothesis and alternative would be virtually impossible due to mathematical difficulties, just as this is the case in the classical goodness-of-fit problems.

The reason that it is mathematically easier to deal with the present hypothesis is that the test statistic, which is based on

$$\varphi_M(\overline{X}) - \delta^2, \tag{3.25}$$

has a simple normal distribution in the limit for large sample sizes.

Lemma 3.1.1 We have

$$\sqrt{n}(\varphi_M(\overline{X}) - \varphi_M(\mu)) \xrightarrow{d} \mathcal{N}(0, v^2), \quad \text{as} \quad n \to \infty, \tag{3.26}$$

and

$$v^2 = 4\left\langle \Pi^{\perp}\mu, \Sigma\Pi^{\perp}\mu \right\rangle. \tag{3.27}$$

If $v^2 = 0$, the limiting distribution $\mathcal{N}(0, v^2)$ is to be interpreted as the distribution which is degenerate at 0.

Proof: The central limit theorem for $\mathbb{H}$-valued random variables yields the existence of a $\mathcal{G}(0, \Sigma)$ random element G, such that

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} G, \quad \text{as} \quad n \to \infty, \quad \text{in} \quad (\mathbb{H}, \mathbb{B}_{\mathbb{H}}). \tag{3.28}$$

It is easy to see that $\varphi_M : \mathbb{H} \to \mathbb{R}$ is Fréchet differentiable at any $\mu \in \mathbb{H}$, tangentially to $\mathbb{H}$, with derivative the linear functional

$$\varphi^*_{M,\mu} h = 2\left\langle \Pi^{\perp}\mu, h \right\rangle, \quad h \in \mathbb{H}. \tag{3.29}$$

It is because

$$\begin{aligned}|\varphi_M(x+h)-\varphi_M(x)-\varphi^*_{M,x}h| &= |\|\Pi^\perp(x+h)\|^2-\|\Pi^\perp x\|^2-2\left\langle\Pi^\perp x,h\right\rangle|\\ &= |\left\langle\Pi^\perp(x+h),x+h\right\rangle-\left\langle\Pi^\perp x,x\right\rangle-2\left\langle\Pi^\perp x,h\right\rangle|\\ &= |\left\langle\Pi^\perp x,x\right\rangle+2\left\langle\Pi^\perp x,h\right\rangle\\ &\quad+\left\langle\Pi^\perp h,h\right\rangle-\left\langle\Pi^\perp x,x\right\rangle-2\left\langle\Pi^\perp x,h\right\rangle|\\ &= o(\|\Pi^\perp h\|),\quad \forall h\in\mathbb{H}.\end{aligned} \tag{3.30}$$

According to the Delta Method, we may conclude

$$\sqrt{n}\{\varphi_M(\overline{X})-\varphi_M(\mu)\}\xrightarrow{d}2\left\langle\Pi^\perp\mu,G\right\rangle. \tag{3.31}$$

The random variable on the right in (3.31) is normal because G is gaussian, and clearly its mean is 0. Therefore, its variance equals

$$\mathbb{E}\;2\left\langle\Pi^\perp\mu,G\right\rangle 2\left\langle\Pi^\perp\mu,G\right\rangle=4\left\langle\Pi^\perp\mu,\Sigma\Pi^\perp\mu\right\rangle, \tag{3.32}$$

according to the definition of Σ.

As an estimator for v^2, we propose

$$\widehat{v}^2=4\langle\Pi^\perp\overline{X},\widehat{\Sigma}\Pi^\perp\overline{X}\rangle. \tag{3.33}$$

Note that this estimator is obtained from the expression on the right of (3.27) and replacing μ with $\overline{X}$, and Σ with $\widehat{\Sigma}$. We will now show that this estimator is consistent.

Lemma 3.1.2. We have

$$\widehat{v}^2\xrightarrow{p}v^2,\quad\text{as}\quad n\to\infty. \tag{3.34}$$

Proof: By simple algebra, we find

$$\begin{aligned}
\langle \Pi^{\perp}\overline{X}, \widehat{\Sigma}\Pi^{\perp}\overline{X}\rangle &= \langle \Pi^{\perp}\overline{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}) \otimes (X_i - \overline{X})\ \Pi^{\perp}\overline{X}\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\langle \Pi^{\perp}\overline{X}, \left\langle \Pi^{\perp}\overline{X}, (X_i - \overline{X})\right\rangle (X_i - \overline{X})\right\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\langle \Pi^{\perp}\overline{X}, X_i - \overline{X}\right\rangle \left\langle \Pi^{\perp}\overline{X}, X_i - \overline{X}\right\rangle \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\langle \Pi^{\perp}\overline{X}, X_i - \overline{X}\right\rangle^2 \qquad (3.35)\\
&= \frac{1}{n}\sum_{i=1}^{n}\{\left\langle X_i, \Pi^{\perp}\overline{X}\right\rangle - \left\langle \Pi^{\perp}\overline{X}, \Pi^{\perp}\overline{X}\right\rangle\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, \Pi^{\perp}\overline{X}\right\rangle^2 - \left\langle \Pi^{\perp}\overline{X}, \Pi^{\perp}\overline{X}\right\rangle^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, \Pi^{\perp}\overline{X}\right\rangle^2 - \|\Pi^{\perp}\overline{X}\|^4.
\end{aligned}$$

According to the weak law of large numbers in $\mathbb{H}$ and the continuity of the projection operator, we have $\|\Pi^{\perp}\overline{X}\|^4 \xrightarrow{p} \|\Pi^{\perp}\mu\|^4$. The weak law of large numbers for real valued random variables yields $\frac{1}{n}\sum_{i=1}^{n}\left\langle \Pi^{\perp}X_i, \Pi^{\perp}\mu\right\rangle^2 \xrightarrow{p} \mathbb{E}\left\langle \Pi^{\perp}X, \Pi^{\perp}\mu\right\rangle^2$. Furthermore, we have

$$\begin{aligned}
\left|\ \frac{1}{n}\sum_{i=1}^{n}\{\left\langle X_i, \Pi^{\perp}\overline{X}\right\rangle^2 - \left\langle X_i, \Pi^{\perp}\mu\right\rangle^2\}\ \right| &= \left|\ \frac{1}{n}\sum_{i=1}^{n}\left\langle X_i, \Pi^{\perp}(\overline{X} - \mu)\right\rangle \left\langle X_i, \Pi^{\perp}(\overline{X} + \mu)\right\rangle\ \right| \\
&\le \{\frac{1}{n}\sum_{i=1}^{n}\|X_i\|^2\}\ \|\Pi^{\perp}(\overline{X} - \mu)\|\|\Pi^{\perp}(\overline{X} + \mu)\| \\
&\xrightarrow{p} 0.
\end{aligned} \qquad (3.36)$$

From assumption, $\frac{1}{n}\sum_{i=1}^{n}\|X_i\|^2$ is bounded. Also, $\Pi^{\perp}\overline{X} \xrightarrow{p} \Pi^{\perp}\mu$ and $\|\Pi^{\perp}\overline{X} + \Pi^{\perp}\mu\| \xrightarrow{p} 2\|\Pi^{\perp}\mu\|$. The lemma follows from straightforward combination of all the above.

Now let us consider the asympototics under the null hypothesis. For $0 < \alpha < 1$, let $z_{1-\alpha}$ denote the quartile of order $1-\alpha$ of the standard normal distribution. Focusing on the testing problem (3.22) and (3.23), let us decide to reject the null hypothesis when $\sqrt{n}(\varphi_M(\overline{X}) - \delta^2)/\widehat{v} \ > \ z_{1-\alpha}$. The corresponding power function is then

$$\beta_n(\mu) = \mathbb{P}\{\sqrt{n}(\varphi_M(\overline{X}) - \delta^2)/\widehat{v} \ > \ z_{1-\alpha}\}, \tag{3.37}$$

when $\mu \in \mathbb{H}$ is the true parameter.

Theorem 3.1.1: Asymptotics under the null hypothesis and fixed alternatives. The power function in (3.37) satisfies

$$\lim_{n\to\infty} \beta_n(\mu) = \begin{cases} 0, & \varphi_M(\mu) < \delta^2, \\ \alpha, & \varphi_M(\mu) = \delta^2, \quad v^2 > 0, \\ 1, & \varphi_M(\mu) > \delta^2. \end{cases} \tag{3.38}$$

Hence the test has asymptotic size α, and is consistent against the alternatives $\mu : \varphi_M(\mu) > \delta^2$.

Proof: If $v^2 > 0$, it is immediate from Lemma's 3.1.1 and 3.1.2 that

$$\sqrt{n}(\varphi_M(\overline{X}) - \varphi_M(\mu))/\widehat{v} \xrightarrow{d} \mathcal{N}(0,1). \tag{3.39}$$

The result now follows in the usual way by observing that as $n \to \infty$ if $\varphi_M(\mu) < \delta^2$, then $\sqrt{n}(\varphi_M(\mu) - \delta^2)$ tends to go to $-\infty$ (asymptotic size $= 0$). If $\varphi_M(\mu) = \delta^2$, then $\sqrt{n}(\varphi_M(\mu) - \delta^2)$ tends to go to 0 (asymptotic size $= \alpha$). Finally, if $\varphi_M(\mu) > \delta^2$, then $\sqrt{n}(\varphi_M(\mu) - \delta^2)$ tends to go to ∞ (asymptotic size $= 1$). Notice that even if $v^2 = 0$, we still have that $\sqrt{n}(\varphi_M(\mu) - \delta^2)/\widehat{v}$ tends in probability to ∞ when $\varphi_M(\mu) > \delta^2$ or to $-\infty$ when $\varphi_M(\mu) < \delta^2$.

To describe the sampling situation under local alternatives (including the null hypothesis), we assume now that

$$X_1, \cdots, X_n \ \text{ are i.i.d. } \ (\mu_{n,t}, \Sigma), \tag{3.40}$$

where Σ is as usual and

$$\mu_{n,t} = \mu + \frac{t}{\sqrt{n}}\gamma, \quad t \geq 0, \mu \in B_\delta, \quad \text{and} \quad \gamma \in \mathbb{H} : \Pi^\perp \gamma \neq 0. \tag{3.41}$$

Under the assumptions, it follows that $\mu_{n,0} = \mu$ satisfies H_δ, and $\mu_{n,t}$ satisfies A_δ for each $t > 0$.

Theorem 3.1.2 Asymptotic power

We have, for fixed $t > 0$

$$\lim_{n\to\infty} \beta_n(\mu_{n,t}) = 1 - \Phi(z_{1-\alpha} - 2t\frac{\left\langle \Pi^\perp \mu, \Pi^\perp \gamma \right\rangle}{v}), \tag{3.42}$$

where Φ denote the standard normal c.d.f..

Proof: We may write $X_i = X_i' + (t/\sqrt{n})\gamma$, where the X_i' are i.i.d.(μ, Σ). It is easy to see from this representation that we still have

$$\widehat{v}_n^2 \xrightarrow{p(t)} v^2 > 0, \quad \text{as} \quad n \to \infty, \quad \forall t > 0. \tag{3.43}$$

Exploiting once more the Fréchet differentiability of φ_M, we obtain

$$\begin{aligned}
\sqrt{n}\frac{\varphi_M(\overline{X}) - \delta^2}{\widehat{v}} &= \sqrt{n}\frac{\varphi_M(\overline{X}' + \frac{t}{\sqrt{n}}\gamma) - \delta^2}{\widehat{v}} \\
&= \sqrt{n}\frac{\varphi_M(\overline{X}') + 2\left\langle \Pi^\perp \overline{X}', \frac{t}{\sqrt{n}}\gamma \right\rangle + o(\|\frac{t}{\sqrt{n}}\gamma\|) - \delta^2}{\widehat{v}} \\
&= \frac{\sqrt{n}(\varphi_M(\overline{X}') - \delta^2)}{\widehat{v}} + \frac{2\sqrt{n}\left\langle \Pi^\perp \overline{X}', \ \frac{t}{\sqrt{n}}\gamma \right\rangle}{\widehat{v}} + \frac{o(1)}{\widehat{v}} \\
&= \frac{\sqrt{n}(\varphi_M(\overline{X}') - \varphi_M(\mu))}{\widehat{v}} + \frac{2\left\langle \Pi^\perp \overline{X}', \ t\gamma \right\rangle}{\widehat{v}} + \frac{o(1)}{\widehat{v}} \\
&\xrightarrow{d} \frac{2\left\langle \Pi^\perp \mu, G \right\rangle}{v} + \frac{2\left\langle \Pi^\perp \mu, \ t\gamma \right\rangle}{v} \\
&\stackrel{d}{=} \mathcal{N}(2t\frac{\left\langle \Pi^\perp \mu, \ \Pi^\perp \gamma \right\rangle}{v}, 1), \quad \text{as} \quad n \to \infty,
\end{aligned} \tag{3.44}$$

and the result follows.

Remark 1. To corroborate the remark about interchanging null hypothesis and alternative made as we mentioned before, just note that an asymptotic size α test for testing H'_δ versus A'_δ in (3.24) is obtained by rejecting H'_δ when

$$\sqrt{n}\frac{\varphi_{\mathcal{M}}(\overline{x}) - \delta^2}{\widehat{v}} < z_\alpha \tag{3.45}$$

Remark 2 The expression in (3.42) remains valid for $t = 0$ or γ with $\Pi^\perp\gamma = 0$. In either case, the corresponding mean satisfies the null hypothesis assumption and the limit in (3.41) equals α.

3.2 Indirect One-Sample Problem

We will consider the situation where $X_1, \cdots, X_n$ are independent copies of a random element

$$X : \Omega \to \mathbb{H}, \tag{3.46}$$

where

$$X = Kf + \epsilon, \tag{3.47}$$

with $K : \mathbb{H} \to \mathbb{H}$, a compact linear operator, which is assumed to be injective and ϵ is a random element in $\mathbb{H}$ such that

$$\mathbb{E}\|\epsilon\|^2 < \infty, \quad \text{and} \quad \mathbb{E}\epsilon = 0. \tag{3.48}$$

Let us precondition the equation (3.47) with K^*, and consider

$$\begin{aligned} Y &= K^*X \\ &= K^*Kf + K^*\epsilon \\ &= Rf + \eta, \end{aligned} \tag{3.49}$$

where now $R : \mathbb{H} \to \mathbb{H}$ is compact, injective and strictly positive Hermitian, and

$$\eta = K^*\epsilon,\ \mathbb{E}\eta = 0 \text{ and } \mathbb{E}\|\eta\|^2 \leq \infty. \tag{3.50}$$

According to the spectral theorem, this operator R has eigenvalues (repeated according to their finite multiplicity)

$$\rho_1 \geq \rho_2 \cdots \downarrow 0, \tag{3.51}$$

with a corresponding orthonormal basis of eigenvectors

$$e_1, e_2, \cdots . \tag{3.52}$$

Moreover, we can write

$$R = \sum_{k=1}^{\infty} \rho_k e_k \otimes e_k \tag{3.53}$$

It is well-known that a function of R can be obtained as follows

$$\psi(R) = \sum_{k=1}^{\infty} \psi(\rho_k) e_k \otimes e_k, \tag{3.54}$$

with in particular

$$R^{-1} = \sum_{k=1}^{\infty} \frac{1}{\rho_k} e_k \otimes e_k. \tag{3.55}$$

This clearly shows that R^{-1} is unbounded because $\frac{1}{\rho_k} \uparrow \infty$, as $k \uparrow \infty$. We will see that therefore in our subsequent estimation procedure, we will have to regularize R^{-1}.

Our data are

$$Y_i = Rf + \eta_i \tag{3.56}$$

where $\eta_1, \cdots, \eta_n$ are iid with the same distribution as $K^*\epsilon = \eta$. Consequently,

$$\bar{Y} = Rf + \bar{\eta}, \tag{3.57}$$

and we intend to find out f by $R^{-1}\bar{Y}$. However, $\bar{Y}$ is not exactly Rf, and R^{-1} is not bounded, we will replace R^{-1} by a regularized version. Although several choices are available, we will restrict ourselves to the spectral-cut-off-type regularized inverse

$$R_m^{-1} = \sum_{k=1}^{m} \frac{1}{\rho_k} e_k \otimes e_k, \tag{3.58}$$

where the selection of m will be discussed later. Thus we arrive at the estimator

$$\begin{aligned} \hat{f}_{[m]} &= R_m^{-1}\bar{Y} \\ &= R_m^{-1}(Rf + \bar{\eta}) \\ &= (R_m^{-1}R)f + R_m^{-1}\bar{\eta}. \end{aligned} \tag{3.59}$$

Let us note that (3.50) entails $\mathbb{E}\bar{\eta} = 0$ and hence

$$\begin{aligned} f_{[m]} &= E\hat{f}_{[m]} \\ &= (R_m^{-1}R)f. \end{aligned} \tag{3.60}$$

To ensure the quality of this estimator, we will look at its MISE, i.e. $\mathbb{E}\|\hat{f}_{[m]} - f\|^2$. Henceforth, we will assume that

$$K = K^* = \sqrt{R}. \tag{3.61}$$

The MISE satisfies the well-known variance-bias decomposition

$$\mathbb{E}\|\hat{f}_{[m]} - f\|^2 = \mathbb{E}\|\hat{f}_{[m]} - f_{[m]}\|^2 + \|f_{[m]} - f\|^2. \tag{3.62}$$

To obtain a more explicit expression, let us employ the expressions (3.53) and (3.58).

The variance part then can be written as

$$\begin{aligned} \mathbb{E}\|\hat{f}_{[m]} - f_{[m]}\|^2 &= \mathbb{E}\|(R_m^{-1}R)f + R_m^{-1}\bar{\eta} - (R_m^{-1}R)f\|^2 = \mathbb{E}\|R_m^{-1}\bar{\eta}\|^2 \\ &= \mathbb{E}\|\sum_{k=1}^{m}\frac{1}{\rho_k}(e_k \otimes e_k)\bar{\eta}\|^2 = \mathbb{E}\|\sum_{k=1}^{m}\frac{1}{\rho_k}\langle\bar{\eta}, e_k\rangle e_k\|^2 \\ &= \mathbb{E}\|\sum_{k=1}^{m}\frac{1}{\rho_k}\langle K^*\bar{\epsilon}, e_k\rangle e_k\|^2 = \mathbb{E}\|\sum_{k=1}^{m}\frac{1}{\rho_k}\langle\bar{\epsilon}, Ke_k\rangle e_k\|^2 \\ &= \mathbb{E}\|\sum_{k=1}^{m}\frac{1}{\sqrt{\rho_k}}\langle\bar{\epsilon}, e_k\rangle e_k\|^2 = \mathbb{E}\sum_{k=1}^{m}\frac{\langle\bar{\epsilon}, e_k\rangle^2}{\rho_k} = \sum_{k=1}^{m}\frac{\mathbb{E}\langle\bar{\epsilon}, e_k\rangle^2}{\rho_k} \\ &= \frac{1}{n}\sum_{k=1}^{m}\frac{Var\langle\epsilon, e_k\rangle}{\rho_k}. \end{aligned} \tag{3.63}$$

The bias part equals,

$$\begin{aligned}
\|f_{[m]} - f\|^2 &= \|R_m^{-1}(Rf) - f\|^2 \\
&= \|\sum_{k=1}^{m} \frac{1}{\rho_k}\langle Rf, e_k\rangle e_k - f\|^2 \\
&= \|\sum_{k=1}^{m} \frac{1}{\rho_k}\langle f, Re_k\rangle e_k - \sum_{k=1}^{\infty}\langle f, e_k\rangle e_k\|^2 \\
&= \|\sum_{k=1}^{m}\langle f, e_k\rangle e_k - \sum_{k=1}^{\infty}\langle f, e_k\rangle e_k\|^2 \\
&= \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2.
\end{aligned} \tag{3.64}$$

Theorem 3.2.1 For the MISE, we have

$$\mathbb{E}\|\hat{f}_{[m]} - f\|^2 \leq \frac{1}{n}\mathbb{E}\|\epsilon\|^2 \sum_{k=1}^{m} \frac{1}{\rho_k} + \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2. \tag{3.65}$$

Proof: We see from (3.62) to (3.64) that

$$\begin{aligned}
\mathbb{E}\|\hat{f}_{[m]} - f\|^2 &= \sum_{k=1}^{m} \frac{\mathbb{E}\langle \bar{\epsilon}, e_k\rangle^2}{\rho_k} + \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2 \\
&= \frac{1}{n}\sum_{k=1}^{m} \frac{Var\langle \epsilon, e_k\rangle}{\rho_k} + \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2 \\
&= \frac{1}{n}\sum_{k=1}^{m} \frac{\mathbb{E}\langle \epsilon, e_k\rangle^2}{\rho_k} + \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2 \\
&\leq \frac{1}{n}\mathbb{E}\|\epsilon\|^2 \sum_{k=1}^{m} \frac{1}{\rho_k} + \sum_{k=m+1}^{\infty} \langle f, e_k\rangle^2,
\end{aligned} \tag{3.66}$$

since $\langle \epsilon, e_k\rangle^2 \leq \|\epsilon\|^2\|e_k\|^2 = \|\epsilon\|^2$.

Let us specify this upper bound under special assumptions on the operator R and the input function f. Let us assume that

$$\rho_k \sim Ck^{-\gamma}, \quad \text{for} \quad \gamma \geq 0, \tag{3.67}$$

and that f is in the smoothness class

$$\Gamma_\nu = \{f \in \mathbb{H} : |\langle f, e_k \rangle| \leq Ck^{-\nu}\}, \quad \nu > \frac{1}{2}. \tag{3.68}$$

Then, we have

$$\sup_{f \in \Gamma_\nu} \mathbb{E}\|\hat{f}_{[m]} - f\|^2 \leq C\{\frac{1}{n}\sum_{k=1}^{m} k^\gamma + \sum_{k=m+1}^{\infty} k^{-2\nu}\}. \tag{3.69}$$

At this point, we may try to find an optimal choice of m (the smoothing parameter) by balancing these two series on the right. For this purpose, we realize that

$$\sum_{k=1}^{m} k^\gamma \approx \int_0^m x^\gamma dx \approx Cm^{\gamma+1}, \tag{3.70}$$

and, similarly, that

$$\begin{aligned} \sum_{k=m+1}^{\infty} k^{-2\nu} &\approx \int_m^\infty x^{-2\nu} dx \\ &\approx Cm^{1-2\nu}. \end{aligned} \tag{3.71}$$

Let us assume that $m = m(n) = n^\delta$, for some $\delta > 0$. In order for $\sum_{k=1}^{m} k^\gamma$ to be approximately equal to $\sum_{k=m+1}^{\infty} k^{-2\nu}$, we need

$$n^{-1}n^{\delta(\gamma+1)} = n^{\delta(1-2\nu)}, \tag{3.72}$$

that is

$$\begin{aligned} \delta(\gamma+1)-1 &= \delta(1-2\nu) \\ \delta(2\nu+\gamma) &= 1 \\ \delta &= \frac{1}{2\nu+\gamma}. \end{aligned} \tag{3.73}$$

The above yields the following.

Theorem 3.2.2 Under (3.67) and (3.68), we have

$$\sup_{f\in\Gamma_\nu} \mathbb{E}\|\hat{f}_{[m(n)]}-f\|^2 \leq Cn^{-\frac{2\nu-1}{2\nu+\gamma}}. \tag{3.74}$$

Now let us consider a lower bound of the estimator. At this point, let us make some simplifying assumptions. First, we still assume as for Theorem 3.2.2 (see (3.61)) that K is positive Hermitian, then

$$\begin{aligned} K &= \sqrt{R} \\ &= \sum_{k=1}^{\infty} \sqrt{\rho_k} e_k \otimes e_k. \end{aligned} \tag{3.75}$$

Secondly, we also assume that ε(i.e. all the ε_i) have a Karhunen-Loève expansion

$$\varepsilon = \sum_{k=1}^{\infty} \sqrt{\sigma_k^2} Z_k e_k, \tag{3.76}$$

where Z_1,Z_2,$\cdots$ are i.i.d. with $\mathcal{N}(0,1)$ and hence

$$\sigma_k Z_k \stackrel{d}{=} \mathcal{N}(0,\sigma_k^2). \tag{3.77}$$

This means that the covariance operator Σ of ϵ has the expansion $\Sigma = \sum_{k=1}^{\infty} \sigma_k^2 e_k \otimes e_k$ as in (3.5). It is important to note that both Σ and R have the same basis of eigenvectors $e_1, e_2, \cdots$.

Under the above assumptions, we have

$$X \stackrel{d}{=} \mathcal{G}(Kf, \Sigma). \tag{3.78}$$

In order for $\mathbb{P}_X = P_f$ to be equivalent to $P_0 \stackrel{d}{=} \mathcal{G}(0, \Sigma)$ according to Theorem 1.2.1 it is sufficient that

$$\sum_{k=1}^{\infty} \frac{(\langle Kf, e_k \rangle - 0)^2}{\sigma_k^2} = \sum_{k=1}^{\infty} \frac{\rho_k f_k^2}{\sigma_k^2} < \infty. \tag{3.79}$$

Under this assumption, we have

$$\frac{dP_f}{dP_0} = e^{\sum_{k=1}^{\infty} \frac{\sqrt{\rho_k} f_k}{\sigma_k^2}(x_k - \frac{1}{2}\frac{\sqrt{\rho_k} f_k}{\sigma_k^2})} \tag{3.80}$$

if we define the notation $\langle f, e_k \rangle = f_k$. Also, if $X_1, \cdots, X_n$ are independent copies of X, the joint density of these random elements w.r.t P_0^n becomes

$$\begin{aligned}
\frac{dP_f^n}{dP_0^n} &= (\frac{dP_f^n}{dP_0^n})(x_{11}, x_{12}, \cdots; x_{21}, x_{22}, \cdots; \cdots; x_{n1}, x_{n2}, \cdots) \\
&= \prod_{i=1}^{n} e^{\sum_{k=1}^{\infty} \frac{\sqrt{\rho_k} f_k}{\sigma_k^2}(x_{ik} - \frac{1}{2}\frac{\sqrt{\rho_k} f_k}{\sigma_k^2})} \\
&= e^{\sum_{k=1}^{\infty} \frac{\sqrt{\rho_k} f_k}{\sigma_k^2}(\sum_{i=1}^{\infty} x_{ik} - \frac{n}{2}\frac{\sqrt{\rho_k} f_k}{\sigma_k^2})} \\
&= e^{\sum_{k=1}^{\infty} \frac{n\sqrt{\rho_k} f_k}{\sigma_k^2}(\bar{x}_k - \frac{1}{2}\frac{\sqrt{\rho_k} f_k}{\sigma_k^2})} \\
&= e^{\frac{n\sqrt{\rho_k} f_k}{\sigma_k^2}(\bar{x}_k - \frac{1}{2}\frac{\sqrt{\rho_k} f_k}{\sigma_k^2})} e^{\sum_{m \neq k} \frac{n\sqrt{\rho_m} f_m}{\sigma_m^2}(\bar{x}_m - \frac{1}{2}\frac{\sqrt{\rho_m} f_m}{\sigma_m^2})} \\
&= \kappa_{f_k}(\bar{x}_k) h_{f_{(k)}}(x)
\end{aligned} \tag{3.81}$$

where $\langle f, e_k \rangle = f_k$ and $f_{(k)} = (f_1, \cdots, f_{k-1}, f_{k+1}, \cdots)$. Hence, we can see that $\bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ik}$ is sufficient for f_k for each $f_{(k)}$.

Let $T = T(X_1, \cdots, X_n)$ be an arbitrary estimator of $f \in \mathbb{H}$ such that $\mathbb{E}\|T\|^2 < \infty$, and define $\langle T, e_k \rangle = T_k$. Then the risk of T is given in the following lemma.

Lemma 3.2.1

$$\begin{aligned}\mathbb{E}\|T-f\|^2 &= \lim_{m\to\infty}\sum_{k=1}^{m}\mathbb{E}\langle T-f,e_k\rangle^2 \\ &= \lim_{m\to\infty}\sum_{k=1}^{m}\mathbb{E}(T_k-f_k)^2.\end{aligned} \tag{3.82}$$

Note that we may consider $T_k = \langle T, e_k\rangle$ to be the estimator of the real parameter f_k. Because of the sufficiency, we may restrict the estimators T_k of f_k to functions $h(\bar{X}_k)$ of

$$\bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ik}, \tag{3.83}$$

where $X_{1k},\cdots,X_{nk}$ are i.i.d. $\mathcal{N}(\sqrt{\rho_k}f_k,\sigma_k^2)$, and hence

$$\bar{X}_k \overset{d}{=} \mathcal{N}(\sqrt{\rho_k}f_k,\frac{1}{n}\sigma_k^2),\quad f_k\in I_k=[-Ck^{-\nu},+Ck^{-\nu}]. \tag{3.84}$$

The lower bound to be used is based on the van Trees inequality (Gill & Levit (1995)) and requires a prior on each of the intervals I_k. Let $\pi:\mathbb{R}\to[0,\infty)$ be continousely diffentiable outside $(-1,1)$, and such that $\int_{-1}^{1}\pi(t)dt=1$. Assume that its Fisher information satisfies

$$0<\mathcal{I}(\pi)=\int_{-1}^{1}\{\frac{\pi'(t)}{\pi(t)}\}^2\pi(t)dt<\infty. \tag{3.85}$$

Then define

$$\pi_k(t)=\frac{1}{Ck^{-\nu}}\pi(\frac{t}{Ck^{-\nu}}),\quad t\in I_k. \tag{3.86}$$

This is a density on I_k with Fisher information

$$\begin{aligned} I(\pi_k) &= \int_{-Ck^{-\nu}}^{+Ck^{-\nu}} \{\frac{\pi_k'(t)}{\pi_k(t)}\}^2 \pi_k(t)dt \\ &= (\frac{k^\nu}{C})^2 \mathcal{I}(\pi). \end{aligned} \tag{3.87}$$

Next let us compute the information in $\bar{X}_k$. Because its density equals

$$p_{f_k}(x) = \frac{1}{\sqrt{2\pi \frac{1}{n}\sigma_k^2}} \exp^{-\frac{1}{2}\frac{(x-\sqrt{\rho_k} f_k)^2}{\frac{1}{n}\sigma_k^2}}, \tag{3.88}$$

and

$$\frac{\partial}{\partial f_k}\{-\frac{1}{2}\frac{(x-\sqrt{\rho_k}f_k)^2}{\frac{1}{n}\sigma_k^2}\} = \frac{x-\sqrt{\rho_k}f_k}{\frac{1}{n}\sigma_k^2}\sqrt{\rho_k}, \tag{3.89}$$

we have

$$\begin{aligned} \mathcal{I}(f_k) &= \mathbb{E}\{\frac{\partial}{\partial f_k}\log p_{f_k}(\bar{X}_k)\}^2 \\ &= \mathbb{E}(\frac{\bar{X}_k - \sqrt{\rho_k}f_k}{\frac{1}{n}\sigma_k^2})^2 \rho_k \\ &= \frac{n^2}{\sigma_k^4}\frac{1}{n}\sigma_k^2 \rho_k \\ &= \frac{n}{\sigma_k^2}\rho_k. \end{aligned} \tag{3.90}$$

An application of the Van Trees inequality now yields

$$\begin{aligned} \int_{I_k} \mathbb{E}\{T_k - f_k\}^2 \pi_k(f_k) df_k &\geq \frac{1}{(\frac{k^\nu}{C})^2\mathcal{I}(\pi) + \frac{n}{\sigma_k^2}\rho_k} \\ &\geq \frac{C}{k^{2\nu} + nk^{\beta-\gamma}} \end{aligned} \tag{3.91}$$

if we assume

$$\sigma_k^2 \sim Ck^{-\beta},\ \beta > 1, \tag{3.92}$$

and in addition to (3.67).

Now an actual lower bound is

$$\begin{aligned}
&\sup_{f\in\Gamma_\nu} \mathbb{E}\|T-f\|^2 \\
&\geq \sup_{f\in\Gamma_\nu} \sum_{k=1}^{m} \mathbb{E}\langle T_k - f_k\rangle^2 \\
&\geq \int_{I_1}\cdots\int_{I_m} \sum_{k=1}^{m} \mathbb{E}\langle T_k - f_k, e_k\rangle^2 \pi_1\cdots\pi_m df_1\cdots df_m \\
&= \sum_{k=1}^{m} \int_{I_1}\cdots\int_{I_m} \mathbb{E}\langle T_k - f_k, e_k\rangle^2 \pi_1(f_1)\cdots\pi_m(f_m) df_1\cdots df_m \\
&\geq C\sum_{k=1}^{m} \frac{1}{k^{2\nu}+nk^{\beta-\gamma}}.
\end{aligned} \tag{3.93}$$

Because this holds true for every $m\in\mathbb{N}$ and T, we even have

$$\begin{aligned}
\inf_T \sup_{f\in\Gamma_\nu} \mathbb{E}\|T-f\|^2 &\geq C\sum_{k=1}^{\infty} \frac{1}{k^{2\nu}+nk^{\beta-\gamma}} \\
&\geq C\int_1^\infty \frac{1}{x^{2\nu}+nx^{\beta-\gamma}}.
\end{aligned} \tag{3.94}$$

Let us now first assume that $\gamma-\beta > -1$. In this case, the last integral can be written

$$\begin{aligned}
\int_1^\infty \frac{x^{\gamma-\beta}}{n+x^{2\nu+\gamma-\beta}}dx &\geq \frac{C}{n}\int_0^\infty \frac{x^{\gamma-\beta}}{1+(xn^{-\frac{1}{2\nu+\gamma-\beta}})^{2\nu+\gamma-\beta}}dx \\
&= Cn^{-\frac{2\nu-1}{2\nu+\gamma-\beta}} \int_0^\infty \frac{y^{\gamma-\beta}}{1+y^{2\nu+\gamma-\beta}}dy \\
&= Cn^{-\frac{2\nu-1}{2\nu+\gamma-\beta}}.
\end{aligned} \tag{3.95}$$

For $\gamma - \beta \leq -1$, (3.94) becomes

$$\begin{aligned}\inf_T \sup_{f\in\Gamma_\nu} \mathbb{E}\|T - f\|^2 &\geq C \int_1^\infty \frac{1}{x^{2\nu} + nx^{\beta-\gamma}} \\ &> C \int_1^\infty \frac{1}{nx^{2\nu} + nx^{\beta-\gamma}} \\ &\geq \frac{C}{n}.\end{aligned} \tag{3.96}$$

Returning to the upper bound in (3.65), it should be noted that no special assumptions on the error distribution were made. If we impose the same conditions as for the lower bound, in particular (3.92), we see that

$$\begin{aligned}\mathbb{E}\langle \epsilon, e_k\rangle^2 &= \langle e_k, \Sigma e_k\rangle \\ &= \sigma_k^2 \\ &\sim Ck^{-\beta}.\end{aligned} \tag{3.97}$$

Substituting this in (3.66) yields the upper bound

$$\begin{aligned}\sup_{f\in\Gamma_\nu} \mathbb{E}\|\hat{f}_{m(n)} - f\|^2 &\leq C\{\frac{1}{n}\sum_{k=1}^{m} k^{\gamma-\beta} + \sum_{k=m+1}^{\infty} k^{-2\nu}\} \\ &\leq C\{\frac{1}{n}\int_1^m x^{\gamma-\beta}dx + \int_m^\infty x^{-2\nu}dx\}.\end{aligned} \tag{3.98}$$

Again, assuming the $\gamma - \beta > -1$ we obtain an upper bound

$$\frac{C}{n}m^{\gamma-\beta+1} + Cm^{-2\nu+1}. \tag{3.99}$$

Setting $m = n^\delta$ and equating the orders of these two terms we need

$$\begin{aligned}-1 + \delta(\gamma - \beta + 1) &= \delta(1 - 2\nu), \\ \delta &= \frac{1}{2\nu + \gamma - \beta}.\end{aligned} \tag{3.100}$$

This yields

$$\sup_{f\in\Gamma_\nu} \mathbb{E}\|\hat{f}_{m(n)} - f\|^2 \le Cn^{-\frac{2\nu-1}{2\nu+\gamma-\beta}}. \tag{3.101}$$

For $\gamma - \beta \le -1$, (3.98) becomes

$$\begin{aligned}\sup_{f\in\Gamma_\nu} \mathbb{E}\|\hat{f}_{m(n)} - f\|^2 &\le C\{\frac{1}{n}\int_1^m x^{\gamma-\beta}dx + \int_m^\infty x^{-2\nu}dx\} \\ &\le \frac{C}{n}.\end{aligned} \tag{3.102}$$

For each case, note that this is the same order of magnitude as for the lower bound. This means that we have the following result.

Theorem 3.2.3 Under assumption (3.67), (3.76), and (3.92), we have that the order of magnitude of the supremum of the MISE of the estimators $\hat{f}_{m(n)}$ equals that of the lower bound. In other words, these estimators are rate-optimal. For $\gamma - \beta > -1$, MISE=$Cn^{-\frac{2\nu-1}{2\nu+\gamma-\beta}}$ for suitable choice of $m(n)$. For $\gamma - \beta \le -1$, MISE=$\frac{C}{n}$ because when β get larger σ_k^2 goes to zero faster and Σ is similar to finite dimension.

3.3 Multi-Sample Problem

Let $X_{j1}, \cdots, X_{jn_j}$ be i.i.d. with mean μ_j and covariance operator Σ_j, where $n_j \in \mathbb{N}$, s.t. $\sum_j n_j = n$, and let these random elements satisfy the moment condition that $\mathbb{E}\|X\|^4 < \infty$: all of this for $j = 1, \cdots, p$. Moreover these p samples are supposed to be mutually independent, and their sample sizes satisfy

$$\frac{n_j}{n} = \lambda_j + o(\frac{1}{\sqrt{n}}) \text{ as } n = n_1 + \cdots + n_p \to \infty, \tag{3.103}$$

and

$$\lambda_j \in (0,1), \quad j = 1, \cdots, p. \tag{3.104}$$

Let us define

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}, \quad \bar{X} = \frac{1}{p} \sum_{j=1}^{p} \frac{n_j}{n} \bar{X}_j, \quad j = 1, \cdots, p. \tag{3.105}$$

Furthermore, let the functionals $\psi_n : \mathbb{H}^p \to \mathbb{R}$ be given by

$$\psi_n(x_1, \cdots, x_p) = \sum_{j=1}^{p} \|\frac{n_j}{n} x_j - \bar{x}_n\|^2, \tag{3.106}$$

where $x_1, \cdots, x_p \in \mathbb{H}$ and $\bar{x}_n = \frac{1}{p} \sum_{j=1}^{p} \frac{n_j}{n} x_j$. Defining $\psi : \mathbb{H}^p \to \mathbb{R}$ by

$$\psi(x_1, \cdots, x_p) = \sum_{j=1}^{p} \|\lambda_j x_j - \bar{x}\|^2, \tag{3.107}$$

where $\bar{x} = \frac{1}{p} \sum_{j=1}^{p} \lambda_j x_j$, it is readily verified that

$$\sqrt{n}\{\psi_n(x_1, \cdots, x_p) - \psi(x_1, \cdots, x_p)\} \longrightarrow 0, \quad \text{as } n \to \infty, \tag{3.108}$$

provided that condition (3.103) and (3.104) are fulfilled.

The neighborhood hypothesis in this model can be loosely formulated as approximate equality of the means. More precisely the null hypothesis

$$H_{p,\delta} : \mu = (\mu_1, \cdots, \mu_p)^* \in M_{p,\delta} \cup B_{p,\delta}, \tag{3.109}$$

where $M_{p,\delta} = \{x \in \mathbb{H}^p : \psi(x) < \delta^2\}$ and
$B_{p,\delta} = \{x \in \mathbb{H}^p : \psi(x) = \delta^2,\ \sum_{j=1}^p \lambda_j \langle \lambda_j x_j - \bar{x},\ \Sigma_j(\lambda_j x_j - \bar{x})\rangle > 0\}$, will be tested against the alternative

$$A_{p,\delta} : \mu = (\mu_1, \cdots, \mu_p)^* \in M^c_{p,\delta} \cap B^c_{p,\delta}, \tag{3.110}$$

Let us introduce some further notation and set

$$\tau_p^2 = 4 \sum_{j=1}^p \lambda_j \langle \lambda_j \mu_j - \bar{\mu},\ \Sigma_j(\lambda_j \mu_j - \bar{\mu})\rangle, \quad \text{where } \bar{\mu} = \frac{1}{p} \sum_{j=1}^p \lambda_j \mu_j. \tag{3.111}$$

Writing $\bar{\Sigma}_j$ for the sample covarinace operatior of the jth sample, the quantity in (3.111) will be estimated by

$$\hat{\tau}_p^2 = 4 \sum_{j=1}^p \lambda_j \langle \lambda_j \bar{X}_j - \bar{X},\ \bar{\Sigma}_j(\lambda_j \bar{X}_j - \bar{X})\rangle. \tag{3.112}$$

Theorem 3.3.1 The test that rejects $H_{p,\delta}$ for

$$\sqrt{n}\{\psi_n(\bar{X}_1, \cdots, \bar{X}_p) - \psi_n(\mu_1, \cdots, \mu_p)\}/\hat{\tau}^2_{p,n} > z_{1-\alpha}, \quad 0 < \alpha < 1 \tag{3.113}$$

has asymptotic size α, and is consistent against fixed alternatives $\mu = (\mu_1, \cdots, \mu_p)^*$ with $\psi(\mu) > \delta^2$.

Proof: Because the p samples are independent the central limit theorem in (1.48) yields

$$\sqrt{n} \begin{pmatrix} \bar{X}_1 - \mu_1 \\ \vdots \\ \bar{X}_p - \mu_p \end{pmatrix}_{p \times \infty} \xrightarrow{d} \begin{pmatrix} G_1 \\ \vdots \\ G_p \end{pmatrix}_{p \times \infty} \in \mathbb{H}^P, \tag{3.114}$$

where $G_1, \cdots, G_p$ are independent Gaussian random elements in $\mathbb{H}$, and

$$G_j \overset{d}{=} \mathcal{G}(0, \ \frac{1}{\lambda_j}\Sigma_j) \tag{3.115}$$

It follows from (3.108) that

$$\sqrt{n}\ [\psi_n(\bar{X}_1, \cdots, \bar{X}_p) - \psi_n(\mu_1, \cdots, \mu_p) - \{\psi(\bar{X}_1, \cdots, \bar{X}_p) - \psi(\mu_1, \cdots, \mu_p)\}] \longrightarrow o(1). \tag{3.116}$$

Moreover, a slight modification of Lemma 3.1.2 yields that

$$\langle \bar{X}_j - \bar{X}, \bar{\Sigma}_j(\bar{X}_j - \bar{X})\rangle \xrightarrow{p} \langle \mu_j - \bar{\mu}, \sum_j (\mu_j - \bar{\mu})\rangle \tag{3.117}$$

and hence

$$\hat{\tau}^2_{n,p} \xrightarrow{p} \tau^2_p. \tag{3.118}$$

This means that the statistic on the left in (3.113) and the one obtained by replacing ψ_n with ψ in that expression will exhibit the same first order asymptotics. The proof with be continued with the latter, simpler version. A simple calculation shows that $\psi : \mathbb{H}^p \longrightarrow \mathbb{R}$ is Fréchet differentiable at any $x \in \mathbb{H}^p$, tangentially to $\mathbb{H}^p$. Writing $\bar{h} = \frac{1}{p}\sum_{j=1}^{p} \lambda_j h_j$, for any $h_1, \cdots, h_p \in \mathbb{H}$, its derivative is equal to

$$2\sum_{j=1}^{p}\langle \lambda_j x_j - \bar{x}, \lambda_j h_j - \bar{h}\rangle = 2\sum_{j=1}^{p}\langle \lambda_j x_j - \bar{x}, \lambda_j h_j\rangle. \tag{3.119}$$

Application of the delta method with the functional ψ in the basic result (3.114) yields

$$\sqrt{n}\{\psi(\bar{X}_1, \cdots, \bar{X}_p) - \psi(\mu_1, \cdots, \mu_p)\} \xrightarrow{d} 2\sum_{j=1}^{p}\langle \lambda_j \mu_j - \bar{\mu}, \lambda_j G_j\rangle. \tag{3.120}$$

According to (3.115) we have

$$\lambda_j G_j \stackrel{d}{=} \mathcal{G}(0, \lambda_j \Sigma_j), \tag{3.121}$$

and because of the independence of the $G - j$ it follows that

$$2 \sum_{j=1}^{p} \langle \lambda_j \mu_j - \bar{\mu}, \lambda_j G_j \rangle \stackrel{d}{=} \mathcal{N}(0, \tau_p^2), \tag{3.122}$$

where τ_p^2 is defined in (3.111). Exploiting the consistency in (3.118) the proof can be concluded in much the same way as that of Theorem 3.1.1. Just as in that theorem we need here that $\tau_p^2 > 0$ at the alternative considered in order to ensure consistency.

CHAPTER 4

CONCLUSION AND FUTURE WORK

In this dissertation, we have introduced and presented a general linear functional model and have shown that the estimator of the functional parameter has a Gaussian distribution in the limit. Special emphasis has been given to the special cases of the one-sample problem. Both the direct version using neighborhood hypotheses and the indirect version are included. In the indirect model, the proposed estimator is shown to have rate-optimal MISE. The multi-sample problem is only briefly included, and we believe a more elaborate study would be desirable.

BIBLIOGRAPHY

[1] Vander Vaart, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.

[2] Grenander, U. (1981). *Abstract Inference*, Wiley, New York.

[3] Debnath, L. & Mikusiński, P. (1999). *Introduction to Hilbert Spaces with Applications*. 2nd Edition. Academic Press.

[4] Kirsch, A. (1996). *An Introduction to the Mathematical Theorey of Inverse Problems*. Springer, New York.

[5] Laha, R. G & Rohatgi, V.K. (1979). *Probability Theorey.* Wiley, New York.

[6] Hájek, J & S̃idak, Z. (1967). *The Theory of Rank Tests*. Academic Press, New York.

[7] Sen, P.K. & Singer, J. M. (1993). *Large Sample Methods in Statistics*. Chapman & Hall, London.

[8] Feldman, J. (1958). Equivalence and perpendicularity of gaussian process. Pac. J Math. 8, 699-708.

[9] Hájek, J. (1958). On a property of normal distribution of any stochastic process. Cz. Math. J. 8, 610-617.

[10] Gill, R.D. & Levit, B.Y. (1995). Applications of the van Trees inquality: a Bayesian Cramér-Rao bound Bernoulli 1, 59-79.

[11] Munk A., Paige R., Pang J., Patrangenaru V., & Ruymgaart F. (2008). The one- and multi-sample problem for functional data with application to projective shape analysis. J of Multivariate Analysis. 99, 815-833.