ANALYSIS OF VARIANCE FOR FUNCTIONAL DATA

Abdelhak Zoglat

March, 1994

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch

Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-612-00512-7

Canadä

ANALYSIS OF VARIANCE FOR FUNCTIONAL DATA

by

Abdelhak Zoglat

A Ph.D. Thesis

Submitted to School of Graduate Studies and Research

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy in Mathematics*

University of Ottawa

Ottawa, Ontario

Canada

March, 1994

* The Ph.D Program is a joint program with

Carleton University, administered by the Ottawa-Carleton

Institute of Mathematics and Statistics

Acknowledgement

I am grateful to Professor André Robert Dabrowski, my supervisor, for his valuable guidance and advice. During the years I have been working under his supervision, I have highly appreciated his infinite patience and constant interest for my work.

I wish to thank the University of Ottawa, the Canadian International Development Agency, and the Green Plan for providing me with financial support.

Abstract

In this dissertation we present an extension to the well known theory of multivariate analysis of variance. In various situations data are continuous stochastic functions of time or space. The speed of pollutants diffusing through a river, the real amplitude of a signal received from a broadcasting satellite, or the hydraulic conductivity rates at a given region are examples of such processes. After the mathematical background we develop tools for analyzing such data. Namely, we develop estimators, tests, and confidence sets for the parameters of interest. We extend these results, obtained under the normality assumption, and show that they are still valid if this assumption is relaxed. Some examples of applications of our techniques are given. We also outline how the latter can apply to random and mixed models for continuous data. In the appendix, we give some programs which we use to compute the distributions of some of our tests statistics.

Contents

Chapter 0

INTRODUCTION

Statistics have become an extremely powerful research tool for scientists, engineers, economists and every researcher whose work is based on observations. In numerous life experiences, the "disorder" may have a perfectly rational explanation at one scale. Yet, there may be a smaller scale where the data do not fit the theory. This suggests the investigation of this new residual uncertainty. Here arises the need for Statistics, the science which attemps to model order in disorder.

Disciplines like geology, image processing, ecology and many others are concerned with data that are continuous over time, space or simultaneously over both time and space. For such data, models that involve statistical dependence are more realistic (data that are close together, in time or space are more likely to be correlated). Moreover, the observation for each subject is best represented as a function. Our purpose in this work is to develop some statistical tools for analyzing such data. It is useful to consider an example supporting our approach.

Suppose that measurements of the levels of pollutants are taken continuously over a month for a large number of streams. Each stream gives rise to a function describing the level of pollutants over time. The analysis of that function can be done through time series. If however, we wish to compare these functions across 30 streams, northern versus southern streams say, a standard approach would be to compute a few summary values for each stream, and to use multivariate analysis of variance to decide whether or not the groups are similar. The fact that we have only 30 streams restricts severely the number of summary values we may use in the characterization of each stream. This condensation of the data into only a few measures may lose potentially useful information.

In this work we have developed procedures which permit the user to use all the data from each stream in his analysis. Each data point is now a function in time rather than a real value. The theoretic computations are done in an abstract space. The practical computations, however, only require numerical integrations. This new approach requires that the user have some idea about the error structure of his data. In many situations, this is accessible from time series methods. In other cases one may only be able to state a form for the error structure, without knowing the exact scale. This presents no problem to our techniques.

This work has two major parts. The first one contains our main theoretical contributions. We first choose a framework and establish tools for our analysis. Then we use them to extend results on estimation and testing from the analysis of variance theory. The applications of our new approach constitute the second part of the thesis. The main application is a geological example where we show that our techniques apply to a wide class of random processes. We consider some simulated data and obtain tables of critical values for certain cases.

Chapter one is devoted to the mathematical background needed for our work. The abstract space mentioned in the previous paragraph is $\mathbf{L}_2$, the space of all square integrable functions. It is a Hilbert space, and thus some useful results on stochastic processes like the Karhünen-Loève expansion are available in a very handy form. It is not an artificial framework; almost all the "natural" random phenomena are described by stochastic elements of $\mathbf{L}_2$.

In chapter two we present two methods of estimation, namely the least squares and maximum likelihood methods. We show that the estimators we obtain are "optimal" in a certain sense. Based on these estimators we develop, in chapter three, tests of linear hypotheses and confidence intervals. These results extend the classical ones from linear models theory.

The results of chapter three require normality. In chapter four we show that the normality asumption can be relaxed. In other words, we show that the results of chapter three are valid asymptotically.

In chapter five we give some examples of applications of our techniques. We use simulated data to test their strength. The results and programs of simulations are given in the appendices A and B.

The previous chapters deal with "fixed effects" models, i.e., models where each observation is a sum of a non-random mean function and a random error. In chapter six we give an overview of how our techniques apply to "random" and "mixed effects" models.

Chapter 1

GOALS AND SETTING

In this chapter we present some results and notions from functional analysis and the general theory of random processes. We first recall some results and properties of *covariance operators*, and illustrate these notions with some examples. A family of semi-norms related to covariance operators is introduced in the second section. The last section presents the model we adopt in this work.

1.1 Preliminaries

Let $(S, \mathcal{S}, \mu)$ be a measure space, with μ a σ-finite measure. Let $\mathbf{L}_2(S,\mu)$ denote the separable Hilbert space:

$$\mathbf{L}_2(S,\mu) = \left\{f\colon S \to \mathbf{R}; \int_S f^2(s)d\mu(s) < \infty\right\},$$

endowed with its usual inner product:

$$\forall f,g \in \mathbf{L}_2(S,\mu) \quad < f,g >= \int_S f(s)g(s)d\mu(s),$$

and its usual norm

$$\forall f \in \mathbf{L}_2(S,\mu) \quad \|f\|_2^2 =< f,f > .$$

In this work we are using some tools from functional analysis, namely " Integral Operators". For the reader's convenience, in this section we will recall some classical results on linear operators. Most of the results, although valid in a more general setting, are stated only in the form we need them.

Let $\kappa : S \times S \to \mathbf{R}$ be a function that is symmetric , i.e., for every $(s,t) \in S \times S$, $\kappa(s,t) = \kappa(t,s)$, and $\mu \times \mu$-square integrable:

$$\int_{S\times S} \kappa^2(s,t)d\mu(s)d\mu(t) < \infty. \tag{1.1}$$

Define the operator

$$T_\kappa \ : \ \mathbf{L}_2(S,\mu) \to \mathbf{L}_2(S,\mu)$$
$$f \longmapsto T_\kappa f,$$

where $T_\kappa f$ is a the real-valued function defined on S by

$$\forall t \in S \quad T_\kappa f(t) = \int_S \kappa(s,t) f(s) d\mu(s).$$

The operator T_κ is called the integral operator with kernel κ. Before stating some important properties of T_κ, recall some definitions and results that can be found in the literature on "Operators in Hilbert Spaces" theory (see for example: [1]). The following result states that (1.1) can be relaxed.

Theorem 1.1.1 [1, Theorem 1.6, p. 28] *If there exists a constant c such that*

$$\int_S |\kappa(s,t)| d\mu(t) \leq c \quad \mu - almost\, everywhere,$$

then T_κ is a continuous linear operator.

It can easily be seen that T_κ is *self-adjoint*, i.e., $< T_\kappa f, g > = < f, T_\kappa g >$ for all f, $g \in \mathbf{L}_2(S,\mu)$. The following results state some useful properties of T_κ.

Definition 1.1.1 [1, Definition 4.1, p. 41] *A linear operator T is compact if the image of the unit ball by T has a compact closure.*

Proposition 1.1.1 [1, Proposition 4.7, p. 43] *The operator T_κ is compact, continuous and self-adjoint.*

The key result is the following proposition.

Proposition 1.1.2 [1, Theorem 5.1 and Corollary 5.3, pp. 46-47] *If T is a compact self-adjoint operator on a Hilbert space $\mathcal{H}$, then T has only a countable number of distinct eigenvalues. Moreover there exist a sequence (λ_k) of real numbers and an orthonormal basis (φ_k) for $(\ker T)^\perp$ such that for all $h \in \mathcal{H}$*

$$T h = \sum_{k \geq 1} \lambda_k < h, \varphi_k > \varphi_k. \tag{1.2}$$

Remark 1.1.1 Proposition 1.1.2 *implies in particular that*

1. *For each $k \geq 1, \varphi_k$ is an eigenfunction of T_κ associated with the eigenvalue λ_k.*

2. For each $f \in \mathbf{L}_2(S,\mu)$ *there exists* $h \in \ker T_\kappa$ *and* $\varphi \in (\ker T_\kappa)^\perp$ *such that:*

$$\varphi = \sum_{k\geq 1} < f, \varphi_k > \varphi_k, \textit{and} \quad f = h + \varphi,$$

and thus, by (1.2),

$$T_\kappa f = T_\kappa \varphi = \sum_{k\geq 1} \lambda_k < f, \varphi_k > \varphi_k.$$

The operators we consider in our work appear in a natural way. Most of the processes of interest have a covariance function that satisfies (1.1), and thus induce integral operators called "Covariance Operators". In particular we are interested in covariance operators associated with Gaussian processes, defined in the next section.

1.1.1 Covariance Operators

Let $\mathcal{B}_{\mathbf{L}_2(S,\mu)}$ be the Borel σ-algebra of $\mathbf{L}_2(S,\mu)$. Recall the following definitions.

Definition 1.1.2 *A measurable function* $X : \big(\Omega, \mathcal{A}, \mathbf{P}\big) \to \big(\mathbf{L}_2(S,\mu), \mathcal{B}_{\mathbf{L}_2(S,\mu)}\big)$ *is called an* $\mathbf{L}_2(S,\mu)$*-valued random variable.*

i) It is called a mean zero random variable if for every $f \in \mathbf{L}_2(S,\mu)$*, the expectation of the real random variable* $< X, f >$ *is null, i.e.,* $\mathbf{E} < X, f >= 0$*,*

ii) and is said to be Gaussian if for every $f \in \mathbf{L}_2(S,\mu)$*, the real random variable* $< X, f >$ *is Gaussian.*

Remark 1.1.2 *By Fubini's theorem*

$$\forall f \in \mathbf{L}_2(S,\mu) \quad \mathbf{E} < f, X >= \int_S f(s)\mathbf{E}\left[X(s)\right] d\mu(s).$$

Therefore an $\mathbf{L}_2(S,\mu)$*-valued random variable is mean zero if and only if for* μ*-almost every* $s \in S$*,* $\mathbf{E}\left[X(s)\right] = 0$*. We will use this latter characterization in this work.*

Let X be a mean zero $\mathbf{L}_2(S,\mu)$-valued random variable, and let κ denote the *covariance* function of X, i.e., $\kappa : S \times S \to \mathbf{R}$ defined by:

$$\kappa(s,t) = \mathbf{E}\big(X(s)X(t)\big).$$

In this setting we will always consider processes for which κ is $\mu \times \mu$-square integrable. Thus T_κ, the *covariance operator* of X is linear, self-adjoint and by Proposition 1.1.1 compact. The following result will be used later. See [14] for example.

Proposition 1.1.3 *Suppose that* κ *is* $\mu \times \mu$*-square integrable. For every* $f, g \in \mathbf{L}_2(S,\mu)$*,*

$$< T_\kappa f, g >= \mathbf{E}[< X, f >< X, g >]. \tag{1.3}$$

Proof

Since κ is $\mu \times \mu$-square integrable, we have by Fubini's theorem

$$\int_S \int_S \kappa^2(s,s)d\mu(s) < \infty.$$

Let A^c denote the complement of $A = \{s \in S; \kappa(s,s) \leq 1\}$. For every $f \in \mathbf{L}_2(S,\mu)$, we have from the definition of A and by the Cauchy-Schwarz's inequality,

$$\begin{aligned}\int_S |\kappa(s,s)|^{1/2}|f(s)|d\mu(s) &\leq \int_A |f(s)|d\mu(s) + \int_{A^c} |\kappa(s,s)||f(s)|d\mu(s) \\ &\leq \|f\|_2 + \left[\int_S |\kappa(s,s)|^2 d\mu(s)\right]^{1/2} \|f\|_2 \\ &< \infty.\end{aligned}$$

Using the Cauchy-Schwarz inequality and Fubini-Tonelli's theorem, we have

$$\begin{aligned}&\int_S \int_S \mathbf{E}|X(s)X(t)f(s)g(t))|d\mu(s)d\mu(t) \\ &\leq \int_S \int_S |\kappa(s,s)|^{1/2}|\kappa(t,t)|^{1/2}|f(s)g(t))|d\mu(s)d\mu(t) \\ &= \int_S |\kappa(s,s)|^{1/2}|f(s)|d\mu(s) \int_S |\kappa(t,t)|^{1/2}|g(t)|d\mu(t) \\ &< \infty.\end{aligned}$$

Therefore we can apply Fubini's theorem and get

$$\begin{aligned}\mathbf{E} < X,f >< X,g > &= \int_S \int_S \kappa(s,t)f(s)g(t)d\mu(s)d\mu(t) \\ &= < T_\kappa f, g >\end{aligned}$$

as claimed. □

In the sequel $(\varphi_k)_{k\geq 1}$ will denote an orthogonal basis of a Hilbert subspace, H_κ, of $(\ker T_\kappa)^\perp = \{f \in \mathbf{L}_2(S,\mu), < f,h >= 0 \;\; \forall h \in \ker T_\kappa\}$, and $(\lambda_k)_{k\geq 1}$ the sequence of eigenvalues of T_κ. Their existence is guaranteed by Proposition 1.1.2. On the other hand note that, by (1.3), the operator T_κ is positive semi-definite, so that its eigenvalues are non-negative.

The Hilbert space H_κ, is called the reproducing kernel Hilbert space associated with the kernel κ. More precisely we have the following definition.

Definition 1.1.3 [14, , Definition 4E, p. 965] *A Hilbert space H_κ is said to be a reproducing kernel Hilbert space, with kernel κ, if its elements are functions on some set S, and if there is a kernel κ on $S \times S$ having the following two properties:*

1. *For every $s \in S$, $\kappa(.,s) \in H_\kappa$, and*

2. for every $h \in H_\kappa$, $< h, \kappa(.,s) >_{H_\kappa} = h(s)$.

Let H_κ denote the closure in $\mathbf{L}_2(S,\mu)$ of the subspace spanned by (φ_k), i.e.,

$$H_\kappa = \{f \in \mathbf{L}_2(S,\mu); f = \sum_{k\geq 1} < f, \varphi_k > \varphi_k\}.$$

Define on H_κ the inner product $< .,. >_{H_\kappa}$ given by

$$\forall f, g \in H_\kappa \quad < f, g >_{H_\kappa} = < T_\kappa f, g > .$$

The first property in the definition holds because of Mercer's theorem, (see for example [7] Theorem 1, p. 62);

$$\forall s, t \in S \quad \kappa(s,t) = \sum_{k\geq 1} \lambda_k \varphi_k(s)\varphi_k(t).$$

For the second one, note that for every $s \in S$

$$\begin{aligned} < T_\kappa g, \kappa(.,s) > &= \sum_{k\geq 1} \lambda_k \varphi_k(s) < T_\kappa g, \varphi_k > \\ &= \sum_{k\geq 1} T_\kappa \varphi_k(s) < g, T_\kappa \varphi_k > \\ &= g(s). \end{aligned}$$

The last equality holds because $(T_\kappa \varphi_k)$ is a basis of H_κ.

The covariance operators associated with Gaussian processes are of particular interest for us. We present here some of their properties.

The Case of Gaussian Processes

Let X be a Gaussian mean zero random variable with values in $\mathbf{L}_2(S,\mu)$.

Lemma 1.1.1 *The function κ^2 is $\mu \times \mu$-integrable.*

Proof

Using the definition of κ, Cauchy-Schwarz inequality and Fubini-Tonelli's theorem, we have that

$$\begin{aligned} \int_{S\times S} \kappa^2(s,t)d\mu(s)d\mu(t) &= \int_{S\times S} \mathbf{E}\Big[X(s)X(t)\Big]^2 d\mu(s)d\mu(t) \\ &\leq \int_{S\times S} \mathbf{E}\Big[X^2(s)X^2(t)\Big]d\mu(s)d\mu(t) \\ &= \mathbf{E}\Big[< |X|, |X| >\Big]^2 \\ &= \mathbf{E}\|X\|^4. \end{aligned}$$

Since X is an $\mathbf{L}_2(S,\mu)$-valued Gaussian r.v, $\mathbf{E}\|X\|^4$ is finite. □

The following expansion, known as the Karhünen-Loève expansion, plays an important role in our work.

Proposition 1.1.4 [7, Theorem 2, p. 64] *There exists a sequence* (X_k) *of independent and identically distributed normal random variables such that:*

$$\begin{aligned} X_k &= \frac{1}{\sqrt{\lambda_k}} < X, \varphi_k >, \quad \text{and} \\ X &= \sum_{k\geq 1} \sqrt{\lambda_k} X_k \varphi_k, \quad a.s. \end{aligned}$$

In ([7], Theorem 2, p. 64), this result is stated for general second-order processes $(X(t), t \in [a,b])$, where $[a,b]$ is an interval. But the result remains true in the present setting.

Proof

The elements of the sequence $\left(< X, \varphi_k >\right)_{k\geq 1}$ are independent Gaussian real random variables such that for each $k \geq 1$,

$$\begin{aligned} \mathbf{E}[< X, \varphi_k >] &= 0, \quad \textit{and} \\ \mathbf{E}[< X, \varphi_k >^2] &= < T_\kappa \varphi_k, \varphi_k > \\ &= \lambda_k. \end{aligned}$$

On the other hand we know, see Remark 1.1.1, that there exists $Z \in \ker T_\kappa$ such that:

$$X = Z + \sum_{k\geq 1} < X, \varphi_k > \varphi_k.$$

But $Z \in \ker T_\kappa$ implies that $< T_\kappa Z, Z >= 0$, and thus

$$\begin{aligned} \mathbf{E}[< Z, Z >^2] &= \mathbf{E}[< X, Z >^2] \\ &= < T_\kappa Z, Z > \\ &= 0. \end{aligned}$$

Therefore $\|Z\|^2 = 0$ almost surely and thus

$$X = \sum_{k\geq 1} < X, \varphi_k > \varphi_k \quad \text{a.s.}$$

□

We end this section by some remarks that will be used later.

Remark 1.1.3 *1. For each* $(s,t) \in S \times S$ *we have that:*

$$\kappa(s,t) = \sum_{k\geq 1} \lambda_k \varphi_k(s) \varphi_k(t).$$

2. The series $\sum_{k\geq 1} \lambda_k^2$ *is convergent.*

The first assertion is Mercer's theorem (see for example [7] Theorem 1 p. 62). The second one follows from

$$\begin{aligned} \sum_{k\geq 1} \lambda_k^2 &\leq \Big(\sum_{k\geq 1} \lambda_k\Big)^2 \\ &= [\mathbf{E} < X, X >]^2 \\ &< \infty. \end{aligned}$$

1.1.2 Examples

To illustrate the Karhünen-Loève expansion, here are some examples.

One parameter time

Suppose that $(S, \mathcal{S}, \mu) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$, the unit interval endowed with its Borel σ-algebra and Lebesgue measure.

The Brownian motion process

Let $(W(s), s \in [0,1])$ denote the Brownian motion process. The covariance function of W is given by

$$\kappa(s,t) = s \wedge t,$$

where $a \wedge b$ is the minimum of a and b. The eigenvalues and eigenfunctions of the covariance operator associated with W are given by

$$\lambda_k = ((k+1/2)\pi)^{-2}, \quad \text{and} \quad \varphi_k(s) = \sqrt{2}\sin\left((k+\frac{1}{2})\pi s\right).$$

There exists a sequence (X_k) of independent and identically distributed $\mathcal{N}(0,1)$ random variables such that

$$\forall s \in [0,1] \quad W(s) = \sum_{k\geq 1} \lambda_k X_k \varphi_k(s) \quad a.s.$$

The reproducing kernel Hilbert space associated with W is given by

$$H_W = \left\{ \sum_{k\geq 1} a_k \varphi_k; \quad (a_k) \in \ell_2 \right\},$$

where $\ell_2 = \{(a_k) \subset \mathbf{R}; \sum_{k\geq 1} a_k^2 < \infty\}$.

The Brownian bridge

Let $(U(s), s \in [0,1])$ denote the Brownian bridge process. The covariance function of U is given by

$$\kappa(s,t) = s \wedge t - st.$$

The eigenvalues and eigenfunctions for the covariance operator associated with U are given by

$$\lambda_k^* = (k\pi)^{-2}, \quad \text{and} \quad \varphi_k^*(s) = \sqrt{2}\sin(k\pi s).$$

Furthermore, there exists a sequence (X_k) of independent and identically distributed $\mathcal{N}(0,1)$ random variables such that

$$\forall s \in [0,1] \quad U(s) = \sum_{k\geq 1} \lambda_k^* X_k \varphi_k^*(s) \quad a.s.$$

The reproducing kernel Hilbert space associated with U is given by

$$H_U = \left\{ \sum_{k\geq 1} a_k \varphi_k^*; \quad (a_k) \in \ell_2 \right\}.$$

Two parameter time

Suppose that $(S, \mathcal{S}, \mu) = ([0,1] \times [0,1], \mathcal{B}_{[0,1]\times[0,1]}, \lambda \times \lambda)$, the unit interval endowed with its Borel σ-algebra and Lebesgue's product measure. As examples of Gaussian processes on $[0,1] \times [0,1]$, we present the two-parameter Wiener process, and the Kiefer process. For more details on the properties of these processes, we refer the reader to [3] and the references therein.

Two parameter Wiener process

Let $(\mathbf{W}(s,t), (s,t) \in [0,1] \times [0,1])$ denote the two-parameter Wiener process. Its covariance function is given by

$$\kappa(\mathrm{s}_1, \mathrm{s}_2) = (s_1 \wedge s_2)(t_1 \wedge t_2),$$

where $\mathrm{s}_i = (s_i, t_i)$, for $i = 1, 2$.

Let (λ_k) and (φ_k) denote the eigenvalues and eigenfunctions of the covariance operator associated with the Brownian motion process on $[0,1]$. The eigenvalues and eigenfunctions of the covariance operator associated with $\mathbf{W}$ are given by

$$\lambda_{k,k'} = \lambda_k \lambda_{k'}, \quad \text{and} \quad \varphi_{k,k'}(s,t) = \varphi_k(s)\varphi_{k'}(t).$$

In particular there exists a sequence $(X_{k,k'})$ of independent and identically distributed $\mathcal{N}(0,1)$ random variables such that

$$\forall \mathbf{s}, \mathbf{t} \in [0,1] \times [0,1] \quad \mathbf{W}(\mathbf{s}, \mathbf{t}) = \sum_{k,k'\geq 1} \lambda_{k,k'} X_{k,k'} \varphi_{k,k'}(\mathbf{s}, \mathbf{t}) \quad a.s.$$

To see this, since Gaussian processes are characterized by their covariance functions, it suffices to use the definition of the $\varphi_{k,k'}$ and the fact that $\sum_{k\geq 1} \lambda_k \varphi_k(s_1)\varphi_k(s_2) = s_1 \wedge s_2$.

The reproducing kernel Hilbert space associated with $\mathbf{W}$ is given by

$$H_{\mathbf{W}} = \left\{ \sum_{k,k'\geq 1} a_{k,k'} \varphi_{k,k'}; \quad (a_{k,k'}) \in \ell_2 \right\}.$$

The Kiefer process

Let $(\mathbf{K}(s,t),(s,t)\in[0,1]\times[0,1])$ denote the Kiefer process. Its covariance function is given by

$$\kappa(\mathbf{s}_1,\mathbf{s}_2)=(s_1\wedge s_2-s_1s_2)(t_1\wedge t_2),$$

where $\mathbf{s}_i=(s_i,t_i)$, for $i=1,2$.

Let (λ_k), (φ_k), (λ_k^*) and (φ_k^*) be as above, the eigenvalues and eigenfunctions of the covariance operator associated respectively with the Brownian motion and the Brownian bridge processes on $[0,1]$.

The eigenvalues and eigenfunctions of the covariance operator associated with $\mathbf{K}$ are given by

$$\lambda_{k,k'}=\lambda_k\lambda_{k'}^*,\quad \text{ and } \quad \varphi_{k,k'}^*(s,t)=\varphi_k(s)\varphi_{k'}^*(t).$$

Using the same arguments as above, we can see that the sequence $(\varphi_{k,k'}^*)$ is an orthonormal basis of $\mathbf{L}_2([0,1]\times[0,1])$. Thus there exists a sequence $(X_{k,k'})$ of independent and identically distributed $\mathcal{N}(0,1)$ random variables such that

$$\forall s,t\in[0,1]\times[0,1]\quad \mathbf{K}(s,t)=\sum_{k,k'1}\lambda_{k,k'}^*X_{k,k'}\varphi_{k,k'}^*(s,t)\qquad a.s.$$

The reproducing kernel Hilbert space associated with $\mathbf{K}$ is given by

$$H_{\mathbf{K}}=\left\{\sum_{k,k'\geq 1}a_{k,k'}\varphi_{k,k'}^*;\quad (a_{k,k'})\in\ell_2\right\}.$$

On the separable Hilbert space $\mathbf{L}_2(S,\mu)$, one can define different inner products, and thus different "norms". In the next section we will define norms that are associated with integral operators. Throughout we will use at our convenience either these norms or the usual one $\|.\|_2$.

1.2 The Choice of Norms

The operator T_κ induces on $\mathbf{L}_2(S,\mu)$ an inner product $<.,.>_{T_\kappa}$ defined by

$$\forall f,g\in\mathbf{L}_2(S,\mu),\quad <f,g>_{T_\kappa}=<f,T_\kappa g>.$$

This inner product induces on $\mathbf{L}_2(S,\mu)$ the semi-norm given by

$$\forall f\in\mathbf{L}_2(S,\mu)\quad \|f\|_{T_\kappa}^2=<f,f>_{T_\kappa}.$$

Its restriction to the reproducing kernel Hilbert space H_κ is a norm. Using the definition of T_κ and Schwarz's inequality it can easily be seen that this seminorm is continuous with respect to the usual one,

$$\forall f\in\mathbf{L}^2(S,\mu)\quad \|f\|_{T_\kappa}^2\leq\|f\|_2^2\int_{S\times S}\kappa^2(s,t)d\mu(s)d\mu(t). \tag{1.4}$$

This implies in particular that any sequence that converges in $(\mathbf{L}_2(S,\mu), \|.\|_2)$ converges also in $(\mathbf{L}_2(S,\mu), \|.\|_{T_\kappa})$. The computations with $\|.\|_{T_\kappa}$ do not present any extra difficulties. Furthermore, using this seminorm we will show that the least squares and the maximum likelihood methods lead to the same estimator. This can not be established for the usual norm.

It is often convenient to use vectorial notation. We will consider vectors in the Hilbert space $\mathbf{L}_2^r(S,\mu) = \underbrace{\mathbf{L}_2(S,\mu) \times \ldots \times \mathbf{L}_2(S,\mu)}_{r\ times}$, endowed with its canonical inner product

$$\mathbf{f} = (f_1, \ldots, f_r)', \mathbf{g} = (g_1, \ldots, g_r)' \in \mathbf{L}_2^r(S,\mu); \quad < \mathbf{f}, \mathbf{g} > = \sum_{i=1}^{r} < f_i, g_i >,$$

or with the inner product $< .,. >_{T_\kappa}$ defined by

$$\mathbf{f} = (f_1, \ldots, f_r)', \mathbf{g} = (g_1, \ldots, g_r)' \in \mathbf{L}_2^r(S,\mu); \quad < \mathbf{f}, \mathbf{g} >_{T_\kappa} = \sum_{i=1}^{r} < f_i, g_i >_{T_\kappa} .$$

Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)'$ be an element of $\mathbf{L}_2^r(S,\mu)$. For each $i = 1, \ldots, r$, and $k \geq 1, \xi_{i(k)}$ will denote the component of the projection of ξ_i on φ_k, i.e.,

$$\xi_{i(k)} = < \xi_i, \varphi_k > \varphi_k,$$

where φ_k an eigenvector of T_κ.

The element of $\mathbf{R}^r$, whose components are $\xi_{1(k)}, \ldots, \xi_{r(k)}$ will be denoted by $\xi_{(k)}$, i.e.,

$$\boldsymbol{\xi}_{(k)} = (\xi_{1(k)}, \ldots, \xi_{r(k)})'.$$

Using the linearity of projections it can be easily seen that for any $p \times r$ matrix $\mathbf{A}$ and any $\mathbf{Z} \in \mathbf{L}_2^r(S,\mu)$ we have

$$(\mathbf{AZ})_{(k)} = \mathbf{AZ}_{(k)}.$$

The following result will be used later, its proof is straightforward.

Lemma 1.2.1 *Let* $\mathbf{Z} \in \mathbf{L}_2^r(S,\mu)$, *then* $\|\mathbf{Z}\|_{T_\kappa}^2 = \sum \lambda_k \mathbf{Z}'_{(k)} \mathbf{Z}_{(k)}$, *in particular, if* $\mathbf{A}$ *is a* $p \times q$ *matrix then* $\|\mathbf{AZ}\|_{T_\kappa}^2 = \sum \lambda_k \mathbf{Z}'_{(k)} \mathbf{A}' \mathbf{A} \mathbf{Z}_{(k)}$

1.3 The Model

The model we adopt in this work has its origins in classical linear models theory. The observations are functions of time, and each observation $Y(t)$ is the sum of a deterministic function m, and a scaled random function ϵ:

$$Y(t) = m(t) + \sigma\epsilon(t), \quad t \in [0,1], \tag{1.5}$$

where $m(t)$ is the mean-value function , $\epsilon(t)$ the fluctuation or error, and σ is a positive and unknown scalar. The mean-value function m is a linear combination of some fixed but unknown functions:

$$m(t) = \sum_{j=1}^{p} x_j \beta_j(t), \quad t \in [0,1], \tag{1.6}$$

where the x_j are known real numbers, and the β_j are unknown functions. They are assumed to belong to the Hilbert space $\mathbf{L}_2(S,\mu)$. Put

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

The equation (1.5) can be written as

$$Y(t) = \mathbf{x}'\boldsymbol{\beta}(t) + \sigma\epsilon(t).$$

For $t \in [0,1]$ fixed, it reduces to the classical linear model form. Put

$$y = Y(t), \quad b_j = \beta_j(t), \quad \text{and} \quad e = \epsilon(t),$$

so that

$$y = \sum_{j=1} x_j b_j + \sigma e.$$

The dependent variable y, the independent variables x_j, the parameters b_j and the error e are real numbers.
More generally, let $t_1, t_2, \ldots, t_q \in [0,1]$ be fixed, and put

$$\mathbf{y} = \begin{bmatrix} Y(t_1) \\ \vdots \\ Y(t_q) \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} \beta_1(t_1) & \ldots & \beta_1(t_q) \\ & \vdots & \\ \beta_p(t_1) & \ldots & \beta_p(t_q) \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \text{ and } \mathbf{e} = \begin{bmatrix} \epsilon(t_1) \\ \vdots \\ \epsilon(t_q) \end{bmatrix}.$$

This is summarized by

$$\mathbf{y}' = \mathbf{x}'\mathbf{b} + \sigma\mathbf{e}', \tag{1.7}$$

which is the form of the classical multivariate linear model. The distinction between multivariate models and standard (univariate) linear models is that the former involves more than one dependent variable. Furthermore, while the observations are supposed to be independent, the components of the error within each observation can be correlated. Equation (1.5), is in this sense a generalization of equation (1.7).

For simplicity of notation and without any loss of generality we will suppose that $(S, \mathcal{S}, \mu) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$, the unit interval endowed with its Borel σ-algebra and the Lebesgue measure. In chapter 4, we will see how these results extend to the general case.

From now on $\mathbf{L}_2$ will denote the Hilbert space $\mathbf{L}_2([0,1],\lambda)$.

Chapter 2

ESTIMATION

The most commonly used methods in the theory of estimation are the least squares and the maximum likelihood methods. In the theory of linear models, under the normality assumption, they lead to the same estimators that have some optimality properties. Their justification as satisfactory estimation procedures is given in many standard statistical texts. Roughly speaking, the least squares method is based on minimizing the squared norm of the error. In the Euclidian space this can be achieved by computing derivatives and equating them to zero. In the infinite dimensional case there is no equivalent to this procedure. However, in a Hilbert space, (or more generally in a Banach space having a countable basis), it is possible to apply the least squares procedure. In sections one and two we show how it applies in $\mathbf{L}_2$, and prove that the estimators are optimum in a certain sense. The method of maximum likelihood is more restrictive than the least squares method. More precisely it requires a distribution density with respect to some measure. Lebesgue measure is the most frequently used in classical linear models, and it is efficient under the normality assumption. In the functional case there is no equivalent to Lebesgue's measure, and thus even under the normality assumption a distribution density might not exist. In sections three and four we present some cases where the maximum likelihood method is applicable.

Suppose that we observe n realizations of the random function Y of (1.5) with the mean-value function m of (1.6). The ith observation is then given by

$$\forall t \in [0,1],\ Y_i(t) = m_i(t) + \sigma\epsilon_i(t), \quad \text{with} \quad m_i(t) = \sum_{j=1}^{p} x_{ij}\beta_j(t),$$

where the β_j are unknown functions and the x_{ij} are known real numbers.
Using vectorial notation, the model can be written

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \sigma\boldsymbol{\epsilon}(t); \quad t \in [0,1], \tag{2.1}$$

where σ is an unknown scalar, ϵ is an $\mathbf{L}_2^n$-valued mean zero random variable, $\boldsymbol{\beta}$ is a non-random element of $\mathbf{L}_2^p$ to be estimated, and $\mathbf{X} = (x_{ij})$ is an $n \times p$ matrix whose

entries are known real numbers.
The observations will be assumed independent and identically distributed so that the components of $\boldsymbol{\epsilon}$ are independent copies of an $\mathbf{L}_2$-valued mean zero random variable, ϵ say. Let κ be the covariance function of ϵ:

$$s,t \in [0,1] \quad \kappa(s,t) = \mathbf{E}\Big[\epsilon(s)\epsilon(t)\Big].$$

Throughout we will suppose that κ *is square-integrable* so that it defines a self-adjoint bounded and compact linear operator T_κ, the covariance operator associated with ϵ.

For any $t \in [0,1]$ fixed, the equations (2.1) are those of a classical linear model and so estimators of $\boldsymbol{\beta}(t)$ are available. Our aim is to find an estimator of the vector $\boldsymbol{\beta}$, an element of $\mathbf{L}_2^p$, that coincides for t fixed with the classical ones, and to make inferences using it. To this end we will reduce the equations (2.1) to equations of classical form: this is the goal of the next section.

2.1 Reduction to the Finite Dimensional Case

Let $\{(\lambda_k, \varphi_k), k \geq 1\}$ denote the sequence of pairs of eigenvalues-eigenfunctions of the covariance operator T_κ. The sequence $(\varphi_k)_{k\geq 1}$ enables us to reduce the model (2.1) to a system of classical models. In particular if ϵ is Gaussian, this reduction preserves independence across projections of realizations, and allows us to identify, as will be seen, the law of some estimators. Before stating the reduced model, we need to introduce some more notation.

For notational simplicity, from now on T will denote T_κ and H_T will denote H_κ, the reproducing kernel Hilbert space associated with κ. Furthermore we will suppose in the sequel that the mean functions $m_i \in H_T$, i.e., the observations belong to the reproducing kernel Hilbert space associated with the covariance function of the error.

Remark 2.1.1 *There is no loss of generality in supposing that $m \in H_T$. Indeed, each observation can be written as*

$$Y = Y_{H_T} + Y_{(H_T)^\perp},$$

where Y_{H_T} is the projection of Y on H_T and $Y_{(H_T)^\perp} = Y - Y_{H_T}$. Recall that by the Karhünen-Loève expansion, $\epsilon \in H_T$ almost surely. Therefore

$$m_{(H_T)^\perp} = Y_{(H_T)^\perp} \quad \text{a.s.},$$

and thus we only need to make inferences about m_{H_T}.

Recall that T induces on $\mathbf{L}_2$ a new inner product

$$\forall f,g \in \mathbf{L}_2 \quad < f,g >_T = < Tf, g >,$$

which defines a norm on H_κ denoted by $\|.\|_T$. The extension to $\mathbf{L}_2^r$ is made in the obvious way, i.e.,

$$\forall \mathbf{f}, \mathbf{g} \in \mathbf{L}_2^r \quad < \mathbf{f}, \mathbf{g} >_T = \sum_{i=1}^{r} < Tf_i, g_i > .$$

Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r)'$ be an element of $\mathbf{L}_2^r$. For each $i = 1, \ldots, r$, and $k \geq 1$

$$\xi_{i(k)} = < \xi_i, \varphi_k > \varphi_k.$$

The elements of $\mathbf{R}^r$, whose components are the $\xi_{i(k)}$ will be denoted by $\boldsymbol{\xi}_{(k)}$, i.e.,

$$\boldsymbol{\xi}_{(k)} = (\xi_{1(k)}, \ldots, \xi_{r(k)})'.$$

Using the linearity of projections it can easily be seen that the model (2.1) leads to

$$\mathbf{Y}_{(k)} = \mathbf{X}\boldsymbol{\beta}_{(k)} + \sigma\epsilon_{(k)}, \quad \forall k \geq 1. \tag{2.2}$$

On the other hand, if for each $i = 1, \ldots, n$, β_i belongs to H_T, the models (2.1) and (2.2) are equivalent.

We will always assume that for each i, β_i belongs to H_T. In the case of Gaussian error, it follows that for each i, the probability measures $\mathcal{L}(Y_i)$ and $\mathcal{L}(\sigma\epsilon_i)$ are equivalent, i.e., the Radon-Nikodym derivative of one with respect to the other exists. We will use this property, (see the discussion at the end of this chapter) to derive a maximum likelihood estimator of $\boldsymbol{\beta}$ which, as in the finite dimensional case, coincides with the estimator obtained by the method of least squares.

2.2 The Least Squares Method

The method of least squares consists of finding the vector $\hat{\boldsymbol{\beta}}$ that minimizes the T-norm, associated with $< .,. >_T$, of the error $\mathbf{e}(\boldsymbol{\beta}) = (\mathbf{Y}_1 - (\mathbf{X}\boldsymbol{\beta})_1, \ldots, \mathbf{Y}_n - (\mathbf{X}\boldsymbol{\beta})_n)$ when viewed as a function of $\boldsymbol{\beta}$. Since σ does not depend on β, dividing $\mathbf{e}(\boldsymbol{\beta})$ by σ if necessary, we will suppose that $\sigma = 1$.

We have that

$$\begin{aligned} \|\mathbf{e}(\boldsymbol{\beta})\|_T^2 &= < \mathbf{e}(\boldsymbol{\beta}), \mathbf{e}(\boldsymbol{\beta}) >_T \\ &= \sum_{i=1}^{n} < \mathbf{e}_i(\boldsymbol{\beta}), \mathbf{e}_i(\boldsymbol{\beta}) >_T \\ &= \sum_{i=1}^{n} \|\mathbf{e}_i(\boldsymbol{\beta})\|_T^2. \end{aligned}$$

We know from Proposition1.1.4 that

$$\|\mathbf{e}_i(\boldsymbol{\beta})\|_T^2 = \sum_{k\geq 1} < \mathbf{e}_i(\boldsymbol{\beta}), \varphi_k >_T^2 .$$

On the other hand we have that

$$\mathbf{e}_i(\boldsymbol{\beta}) = \mathbf{Y}_i - (\mathbf{X}\boldsymbol{\beta})_i,$$

therefore

$$\begin{aligned} < \mathbf{e}_i(\boldsymbol{\beta}), \varphi_k >_T &= < \mathbf{Y}_i, \varphi_k >_T - < (\mathbf{X}\boldsymbol{\beta})_i, \varphi_k >_T \\ &= < T\mathbf{Y}_i, \varphi_k > - < T(\mathbf{X}\boldsymbol{\beta})_i, \varphi_k > \\ &= \lambda_k [\mathbf{Y}_{i(k)} - (\mathbf{X}\boldsymbol{\beta})_{i(k)}], \end{aligned}$$

and therefore

$$\begin{aligned} \sum_{i=1}^{n} \|\mathbf{e}_i(\boldsymbol{\beta})\|_T^2 &= \sum_{k\geq 1}\sum_{i=1}^{n} \lambda_k^2 [\mathbf{Y}_{i(k)} - (\mathbf{X}\boldsymbol{\beta})_{i(k)}]^2 \\ &= \sum_{k\geq 1} \lambda_k^2 (\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)})'(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)}). \end{aligned}$$

This shows that the $\mathbf{L}_2^n$-norm of $\mathbf{e}(\boldsymbol{\beta})$ is minimized if, and only if, for each $k \geq 1$, the Euclidian norm of $\mathbf{e}_{(k)}(\boldsymbol{\beta})$ is minimized.
Let $k \geq 1$ be fixed. Viewed as a function of $\boldsymbol{\beta}_{(k)}$, the norm of $\mathbf{e}_{(k)}(\boldsymbol{\beta})$ is minimized at

$$\hat{\boldsymbol{\beta}}_{(k)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_{(k)}.$$

Proposition 2.2.1 *The vector whose components are linear combinations of the observations,*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

is an estimator of $\boldsymbol{\beta}$ that minimizes the T-norm of $\mathbf{Y} - \mathbf{X}\mathbf{b}$ when viewed as a function of $\mathbf{b}$.

The BLUE Property

From classical linear models theory it is known that under the Gauss-Markov conditions, for each $k \geq 1, \hat{\boldsymbol{\beta}}_{(k)}$, the estimator of $\boldsymbol{\beta}_{(k)}$, is the BLUE, i.e.,

$$\forall \mathbf{a} \in \mathbf{R}^p, \forall \mathbf{b} \in \mathbf{R}^n, \mathbf{E}\left[\mathbf{b}'\mathbf{Y}_{(k)}\right] = \mathbf{a}'\boldsymbol{\beta}_{(k)} \Rightarrow var\left[\mathbf{a}'\hat{\boldsymbol{\beta}}_{(k)}\right] \leq var\left[\mathbf{b}'\mathbf{Y}_{(k)}\right]$$

In other words, among unbiased estimators of $\mathbf{a}'\boldsymbol{\beta}_{(k)}$ that are linear combinations of $\mathbf{Y}_{(k)}$, $\mathbf{a}'\hat{\boldsymbol{\beta}}_{(k)}$ is the one that has the minimum variance. This also can be formulated as follows:

If $\mathbf{f} \in \mathbf{R}^n$ and $\mathbf{g} \in \mathbf{R}^p$ are two continuous linear forms, on their respective spaces, such that

$$\mathbf{E}[< \mathbf{f}, \mathbf{Y}_{(k)} >] = < \mathbf{g}, \boldsymbol{\beta}_{(k)} > \text{ then, } var[< \mathbf{g}, \hat{\boldsymbol{\beta}}_{(k)} >] \leq var[< \mathbf{f}, \mathbf{Y}_{(k)} >].$$

But $\mathbf{Y}_{(k)}$ and $\boldsymbol{\beta}_{(k)}$ are also elements of $\mathbf{L}_2^n$ and $\mathbf{L}_2^p$ respectively, and continuous linear forms on $\mathbf{R}^r$, $r = n$ or p, can be viewed as particular elements of $\mathbf{L}_2^r$. A natural generalization of the "BLUE" property could therefore be:

If $\mathbf{f} \in \mathbf{L}_2^n$ and $\mathbf{g} \in \mathbf{L}_2^p$ are two continuous linear forms, on their respective spaces, such that $\mathbf{E}[< \mathbf{f}, \mathbf{Y} >] =< \mathbf{g}, \boldsymbol{\beta} >$, then $var[< \mathbf{g}, \hat{\boldsymbol{\beta}} >] \leq var[< \mathbf{f}, \mathbf{Y} >]$.

Let $\mathbf{g} \in \mathbf{L}_2^p$. The question is:
For which $\mathbf{f}^* \in \mathbf{L}_2^n$ does

$$var[< \mathbf{f}^*, \mathbf{Y} >] = min\{var[< \mathbf{h}, \mathbf{Y} >] \,, \mathbf{E}[< \mathbf{h}, \mathbf{Y} >] =< \mathbf{g}, \boldsymbol{\beta} > \,, \mathbf{h} \in \mathbf{L}_2^n\}?$$

The answer to this is given by the next theorem.

Theorem 2.2.1 *Let $\mathbf{f} \in \mathbf{L}_2^n$ and $\mathbf{g} \in \mathbf{L}_2^p$ be such that:*

$$\mathbf{E}[< \mathbf{f}, \mathbf{Y} >] =< \mathbf{g}, \boldsymbol{\beta} >,$$

then we have

$$var[< \mathbf{g}, \hat{\boldsymbol{\beta}} >] \leq var[< \mathbf{f}, \mathbf{Y} >].$$

Remark 2.2.1 1. *Since $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $< \mathbf{g}, \hat{\boldsymbol{\beta}} >=< \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}, \mathbf{Y} >$. Furthermore by linearity, we have $\mathbf{E} < \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}, \mathbf{Y} >=< \mathbf{g}, \boldsymbol{\beta} >$. Thus the above result implies that $\mathbf{f}^* = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}$.*

2. *The result of the theorem is given in terms of the usual inner product in $\mathbf{L}_2$ to emphasize that it generalizes the known result in the finite dimensional case. The same arguments used for the proof can be used if $< .,. >$ is replaced by $< .,. >_T$.*

To prove this theorem we need the following lemmas.

Lemma 2.2.1 *For all $k \geq 1$, and for $i = 1, \ldots, n$ we have $var(Y_{i(k)}) = \sigma^2 \lambda_k$.*

Proof

It is clear that $var(Y_{i(k)}) = \sigma^2 var(\epsilon_{i(k)})$. On the other hand we have that

$$var(\epsilon_{i(k)}) = \mathbf{E}[< \epsilon_i, \varphi_k >] =< T\varphi_k, \varphi_k >,$$

where the last equality follows from equation (1.3). Since φ_k is the orthonormal eigenfunction of T associated with λ_k, the lemma follows. □

Lemma 2.2.2 *For every $\mathbf{f} \in \mathbf{L}_2^n$, we have*

$$var[< \mathbf{f}, \mathbf{Y} >] = \sum_{k \geq 1} var[< \mathbf{f}_{(k)}, \mathbf{Y}_{(k)} >].$$

Proof

From the definition of the inner product in $\mathbf{L}_2^n$, $< \mathbf{f}, \mathbf{Y} >= \sum_{i=1}^n < f_i, Y_i >$, and by independence of the Y_i we have $var[< \mathbf{f}, \mathbf{Y} >] = \sum_{i=1}^n var[< f_i, Y_i >]$. But

$$\begin{aligned} var[< f_i, Y_i >] &= \sigma^2 \mathbf{E}[< f_i, \epsilon_i >^2] \\ &= \sigma^2 < T f_i, f_i >, \end{aligned}$$

where the last equality follows from equation (1.3). Now using the expansion of f_i in the orthonormal basis (φ_k), and the fact that the φ_k are eigenfunctions of T we have

$$var[< f_i, Y_i >] = \sigma^2 \sum_{k \geq 1} \lambda_k f_{i(k)}^2,$$

where $f_{i(k)} =< f_i, \varphi_k >$. This shows that

$$\begin{aligned} var[< \mathbf{f}, \mathbf{Y} >] &= \sigma^2 \sum_{i=1}^n \sum_{k \geq 1} \lambda_k f_{i(k)}^2 \\ &= \sigma^2 \sum_{k \geq 1} \sum_{i=1}^n \lambda_k f_{i(k)}^2 \\ &= \sigma^2 \sum_{k \geq 1} \lambda_k f'_{(k)} f_{(k)}. \end{aligned}$$

Note that the second equality holds because the terms of the series are non-negative and the third equality from the inner product in $\mathbf{R}^n$. On the other hand, since $Y_{1(k)}, \ldots, Y_{n(k)}$ are independent, $cov[\mathbf{Y}_{(k)}] = \sigma^2 \lambda_k \mathbf{I}$. Thus

$$\begin{aligned} var[< \mathbf{f}, \mathbf{Y} >] &= \sum_{k \geq 1} f'_{(k)} cov[\mathbf{Y}_{(k)}] f_{(k)} \\ &= \sum_{k \geq 1} var[< \mathbf{f}_{(k)}, \mathbf{Y}_{(k)} >]. \end{aligned}$$

□

Proof of Theorem 2.2.1

Let $\mathbf{f}^* \in \mathbf{L}_2^n$ be such that

$$var[< \mathbf{f}^*, \mathbf{Y} >] = min\{var[< \mathbf{h}, \mathbf{Y} >] \,, \mathbf{E}[< \mathbf{h}, \mathbf{Y} >] =< \mathbf{g}, \boldsymbol{\beta} > \,, \mathbf{h} \in \mathbf{L}_2^n\}.$$

Since the equality

$$\mathbf{E}[< \mathbf{f}^*, \mathbf{Y} >] =< \mathbf{f}^*, \mathbf{X}\boldsymbol{\beta} >=< \mathbf{g}, \boldsymbol{\beta} >,$$

should hold for all $\boldsymbol{\beta}$, $\mathbf{f}^*$ has to satisfy $\mathbf{X}'\mathbf{f}^* = \mathbf{g}$. In other words, it should satisfy

$$\forall k \geq 1 \qquad \mathbf{X}'\mathbf{f}^*{}_{(k)} = \mathbf{g}_{(k)}. \tag{2.3}$$

By Lemma 2.2.2, $var[<\mathbf{f},\mathbf{Y}>] = \sum_{k\geq 1} var[<\mathbf{f}_{(k)},\mathbf{Y}_{(k)}>]$. Thus, to minimize this variance, it suffices to minimize each term of the series. Let $k \geq 1$ be fixed. We will use the Lagrange multiplier vector $2\boldsymbol{\theta}_{(k)}$ to minimize the kth term of the series subject to the conditions (2.3). We have that

$$\begin{aligned} var[<\mathbf{f}^*_{(k)},\mathbf{Y}_{(k)}>] &= \mathbf{f}^{*'}_{(k)} var[\mathbf{Y}_{(k)}]\,\mathbf{f}^*_{(k)}] \\ &= \sigma^2 \mathbf{f}^{*'}_{(k)} \lambda_k \mathbf{I}\,\mathbf{f}^*_{(k)}. \end{aligned}$$

Thus, since λ_k and σ^2 are constants, we have to minimize

$$w = \mathbf{f}^{*'}_{(k)}\mathbf{I}\,\mathbf{f}^*_{(k)} - 2\boldsymbol{\theta}'_{(k)}\,(\mathbf{X}'\mathbf{f}^*_{(k)} - \mathbf{g}_{(k)}),$$

with respect to $\mathbf{f}^*$ and $\boldsymbol{\theta}_{(k)}$. Clearly $\partial w/\partial \boldsymbol{\theta} = 0$ gives $\mathbf{X}'\mathbf{f}^*_{(k)} = \mathbf{g}_{(k)}$ and $\partial w/\partial \mathbf{f}^*_{(k)} = 0$ gives $\lambda_k \mathbf{f}^*_{(k)} = \mathbf{X}\boldsymbol{\theta}_{(k)}$. Replacing in (2.3), we get

$$(\mathbf{X}'\mathbf{X})\boldsymbol{\theta}_{(k)} = \mathbf{g}_{(k)},$$

which implies that

$$\boldsymbol{\theta}_{(k)} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}_{(k)}.$$

Therefore, we have that

$$\mathbf{f}^*_{(k)} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}_{(k)}.$$

And since this is true for all $k \geq 1$, we have

$$\mathbf{f}^* = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{g}.$$

The theorem is proved. □

The following section is devoted to another method of estimation. We will see that it leads to the same estimator $\hat{\boldsymbol{\beta}}$.

2.3 The Maximum Likelihood Method

Here we consider the model given by (1.5):

$$Y(t) = m(t) + \sigma\epsilon(t); \quad t \in [0,1],$$

where m is an element of $\mathbf{L}_2$ and ϵ is a mean zero $\mathbf{L}_2$-valued random variable which is Gaussian. Put $\mathcal{L}(Y) = \nu$ and $\mathcal{L}(\sigma\epsilon) = \mu$. These are two probability measures on $(\mathbf{L}_2, \mathcal{B}_{\mathbf{L}_2})$. Let T, as before denote the covariance operator of ϵ, and let $\sqrt{T}$ denote the operator defined by

$$\forall k \geq 1, \quad \sqrt{T}\varphi_k = \sqrt{\lambda_k}\,\varphi_k.$$

The distributions of two Gaussian processes which have the same covariance operator are either equivalent or orthogonal. A necessary and sufficient condition for two Gaussian probabilities to be equivalent is given in the following theorem.

Theorem 2.3.1 [7, Theorem 1, p. 271] *In order that two probability measures* $\mathbf{P}_0$ *and* $\mathbf{P}_1$ *with means* $m^{(0)}$ *and* $m^{(1)}$ *and common covariances be equivalent, it is necessary and sufficient that*

$$\sum_{k\geq 1} \frac{(m_k^{(0)} - m_k^{(1)})^2}{\lambda_k} < \infty. \tag{2.4}$$

The only alternatives are that they are equivalent or orthogonal.

Remark 2.3.1 *1. According to our notation , for* $i = 0$ or 1, $m_k^{(i)} =< m^{(i)}, \varphi_k >$.

2. As mentioned in [7, Remark 2] *, condition 2.4 is equivalent to* $m^{(0)} - m^{(1)} \in$ *range*$(\sqrt{T})$*. A proof of this assertion can be found, for example, in [10, Theorem 3.1, p. 118].*

In both [10] and [7] one can find an expression for the Radon-Nikodym derivative of one probability with respect to the other when they are equivalent. The proof given in [10, Theorem 3.1, p. 118] uses abstract Wiener spaces, while the one given in [7] is based on Kakutani's theorem.

A necessary and sufficient condition for the Radon-Nikodym derivative of ν with respect to μ to exist as we have seen is that m belongs to *range*$(\sqrt{T})$. This condition is actually equivalent to $m \in H_T$, the reproducing kernel Hilbert space associated with the covariance functions of ϵ, (see for example [6]). We have seen in Remark 2.1.1 that for estimating m there is no loss of generality in supposing that $m \in H_T$. Under this condition, in the next proposition, we give the Radon-Nikodym derivative $d\nu/d\mu$. It can be proved using elementary tools following [6]. Such a proof is given below for the reader's convenience.

Proposition 2.3.1 *Assume the probability measure* ν *is absolutely continuous with respect to the probability measure* μ *then we have*

$$\frac{d\nu}{d\mu}(x) = \exp\left[-\frac{1}{2}\sum_{k\geq 1} m_k^{*2} + \sum_{k\geq 1} m_k^* x_k^*\right], \quad \forall x \in H_T,$$

where

$$m_j^* =< m, \varphi_j^* >,\ x_j^* =< x, \varphi_j^* >,\ \text{and}\ \varphi_j^* = \sigma\sqrt{\lambda_j}\,\varphi_j.$$

Note that by definition of the inner product $< .,. >_T$, we have that

$$\begin{aligned} \frac{d\nu}{d\mu}(x) &= \exp\left[-\frac{\sigma^2}{2}\left(\sum_{k\geq 1}\lambda_k m_k^2 + \sum_{k\geq 1}\lambda_k m_k x_k\right)\right], \quad \forall x \in H_T \\ &= \exp\left[-\frac{\sigma^2}{2}(< m, m >_T - 2 < m, x >_T)\right]. \end{aligned}$$

Proof

Let $(\mathbf{X}_j)_{j\geq 1}$ be independent identically distributed normal $\mathcal{N}(0,1)$ random variables such that:

$$\begin{aligned}\sigma\epsilon &= \sigma\sum_{k\geq 1}\sqrt{\lambda_j}\mathbf{X}_j\varphi_j\\ &= \sum_{k\geq 1}\mathbf{X}_j\varphi_j^*,\end{aligned}$$

and put

$$U_k(x)=\sum_{j=1}^{k}[m_j^{*2}-2m_j^*x_j^*].$$

We have that

$$\begin{aligned}\int_{\mathbf{L}_2}\exp\Big[-U_k(x)\Big]\,d\mu(x) &= \int_\Omega \exp\Big[-\frac{1}{2}U(\epsilon(\omega))\Big]\,d\mathbf{P}(\omega)\\ &= \frac{1}{(2\pi)^{k/2}}\int_{\mathbf{R}^k}\exp\Big[-\sum_{j=1}^{k}(m_j^{*2}-2m_j^*x_j+\frac{1}{2}x_j^2)\Big]\,d\mathbf{x}\\ &= \exp\Big[\sum_{j=1}^{k}m_j^{*2}\Big]\int_{\mathbf{R}^k}\exp\Big[-\frac{1}{2}\sum_{j=1}^{k}(2m_j^*-x_j)^2\Big]\frac{d\mathbf{x}}{(\sqrt{2\pi})^k}\\ &= \exp\Big[\sum_{j=1}^{k}m_j^{*2}\Big],\end{aligned}$$

where $d\mathbf{x}=dx_1\dots dx_k$. Since $\sum_{j\geq 1}m_j^{*2}\leq\|m\|<\infty$, the sequence $(\exp\big[-\frac{1}{2}U_k(.)\big])_{k\geq 1}$ is μ-uniformly integrable. Indeed, using the Cauchy-Schwarz and Chebechev inequalities it can easily be seen that

$$\lim_{\alpha\to\infty}\sup_{k\geq 1}\int_{\exp(-\frac{U_k(x)}{2})>\alpha}\exp(-\frac{U_k(x)}{2})\,d\mu(x)=0.$$

Let $\mathcal{B}_k$ denote the σ-algebra generated by $(\varphi_1^*,\dots,\varphi_k^*)$, i.e., $\mathcal{B}_k=\sigma(\varphi_1,\dots,\varphi_k)$. For each $A\in\mathcal{B}_k$ there exists $B\in\mathcal{B}_{\mathbf{R}^n}$ such that:

$$A=\Big\{f\in\mathbf{L}_2;(<f,\varphi_1^*>,\dots,<f,\varphi_k^*>)\in B\Big\},$$

and we have

$$\begin{aligned}\int_A\exp\Big[-\frac{1}{2}U_k(y)\Big]\,d\mu(y) &= \int_\Omega\exp\left[-\frac{1}{2}\sum_{j=1}^{k}(m_j^{*2}-2m_j^*\epsilon_j^*)\right]\mathbf{1}_B(\epsilon_j^*,\dots,\epsilon_j^*)d\mathbf{P}\\ &= \frac{1}{(\sqrt{2\pi})^k}\int_{x\in B}\exp\Big[-\frac{1}{2}\sum_{j=1}^{k}(m_j^{*2}-2m_j^*x_j+x_j^2)\Big]\,d\mathbf{x}\end{aligned}$$

$$= \frac{1}{(\sqrt{2\pi})^k}\int_{x\in B}\exp\Big[-\frac{1}{2}\sum_{j=1}^{k}(m_j^* - x_j)^2\Big]\,d\mathbf{x}$$
$$= \frac{1}{(\sqrt{2\pi})^k}\int_{x+z\in B}\exp\Big[-\frac{1}{2}\sum_{j=1}^{k}x_j^2\Big]\,d\mathbf{x},$$

where $z = (m_1^*, \ldots, m_j^*)'$ and $d\mathbf{x} = dx_1 \ldots dx_k$. Therefore we have that:

$$\int_A \exp\Big[-\frac{1}{2}U_k(y)\Big]\,d\mu(y) = \nu(A).$$

Let $A \in \mathcal{B}_k$, we have that:

$$\begin{aligned}\int_A \exp\Big[-\frac{1}{2}\sum_{j\geq 1} m_j^{*2} + \sum_{j\geq 1} m_j^* x_j\Big]\,d\mu(x) &= \int_A \lim_{N\to\infty}\exp\Big[-\frac{1}{2}U_N(x)\Big]\,d\mu(x)\\ &= \lim_{N\to\infty}\int_A \exp\Big[-\frac{1}{2}U_N(x)\Big]\,d\mu(x).\end{aligned}$$

The last term is equal to

$$\begin{aligned}&\lim_{N\to\infty}\int_{B\times R^{N-k}}\exp\Big[-\frac{1}{2}\sum_{j=1}^{N}(m_j^{*2} - 2m_j^* x_j + x_j^2)\Big]\frac{dx_1\ldots dx_N}{(2\pi)^N}\\ =\ &\int_B \exp\Big[-\frac{1}{2}\sum_{j=1}^{k}(m_j^{*2} - 2m_j^* x_j + x_j^2)\Big]\frac{dx_1\ldots dx_k}{(\sqrt{2\pi})^k}\times\\ &\lim_{N\to\infty}\int_{R^{N-k}}\exp\Big[-\frac{1}{2}\sum_{j=k+1}^{N}(m_j^{*2} - 2m_j^* x_j + x_j^2)\Big]\frac{dx_{k+1}\ldots dx_N}{(\sqrt{2\pi})^{N-k}}\\ =\ &\int_B \exp\Big[-\frac{1}{2}\sum_{j=1}^{k}(m_j^{*2} - 2m_j^* x_j + x_j^2)\Big]\frac{dx_1\ldots dx_k}{(\sqrt{2\pi})^k}\times 1\\ =\ &\int_A \exp[-\frac{1}{2}U_k(x)]d\mu(x)\\ =\ &\nu(A).\end{aligned}$$

The probability measure ν' defined by

$$A \in \mathcal{B}_{\mathbf{L}_2}, \nu'(A) = \int_A \exp\Big[-\frac{1}{2}\sum_{j\geq 1} m_j^{*2} + \sum_{j\geq 1} m_j^* x_j^*\Big]\,d\mu(x),$$

and ν coincide on the algebra $\bigcup_k \mathcal{B}_k$, and thus on the σ-algebra $\mathcal{B}_{\mathbf{L}_2} = \sigma\Big(\bigcup_k \mathcal{B}_k\Big)$. □

Remark 2.3.2 *The reason that we are considering the Radon-Nikodym derivative of $\mathcal{L}(Y)$ with respect to $\mathcal{L}(\sigma\epsilon)$ rather than $\mathcal{L}(\epsilon)$ is that $\mathcal{L}(Y)$ and $\mathcal{L}(\epsilon)$ are orthogonal. We postpone the details of this fact to the end of this section.*

Recall that the model (2.1) can be written

$$\mathbf{Y}(t) = \mathbf{m}(t) + \sigma\boldsymbol{\epsilon}(t); \quad t \in [0,1],$$

where $\mathbf{m} = (m_1, \ldots, m_n)'$ is a non-random element of $\mathbf{L}_2^n$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is an $\mathbf{L}_2^n$-valued Gaussian random variable which components are independent copies of ϵ. The likelihood function ℓ is defined by:

$$\ell(\mathbf{m}, \mathbf{x}) = \exp\Big[-\frac{1}{2}\sum_{j\geq 1}\sum_{i=1}^{n} m_{i(j)}^{*2} + \sum_{j\geq 1}\sum_{i=1}^{n} m_{i(j)}^{*} x_{i(j)}^{*}\Big], \quad \mathbf{x} \in \mathbf{L}_2^n.$$

As usual, put $\boldsymbol{\xi}_{(j)} = (\xi_{1(j)}, \ldots, \xi_{n(j)})'$, then we have

$$\ell(\mathbf{m}, \mathbf{x}) = \prod_{j\geq 1} \exp\Big\{-\frac{1}{2}\big[\mathbf{m}_{(j)}^{*'}\mathbf{m}_{(j)}^{*} - 2\mathbf{m}_{(j)}^{*}\mathbf{x}_{(j)}^{*}\big]\Big\}, \quad \mathbf{x} \in \mathbf{L}_2^n,$$

which also can be written

$$\ell(\mathbf{m}, \mathbf{x}) = \prod_{j\geq 1} \exp\Big\{-\frac{\sigma^2\lambda_j}{2}\big[\mathbf{m}_{(j)}'\mathbf{m}_{(j)} - 2\mathbf{m}_{(j)}\mathbf{x}_{(j)}\big]\Big\}, \quad \mathbf{x} \in \mathbf{L}_2^n.$$

In the case where

$$\mathbf{m}(t) = \mathbf{X}\boldsymbol{\beta}(t), \quad t \in [0,1],$$

we have that

$$\ell(\boldsymbol{\beta}, \mathbf{y}) = \prod_{j\geq 1} \exp\Big\{-\frac{\sigma^2\lambda_j}{2}\big[\boldsymbol{\beta}_{(j)}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_{(j)} - 2\boldsymbol{\beta}_{(j)}'\mathbf{X}'\mathbf{Y}_{(j)}\big]\Big\}, \quad \mathbf{y} \in \mathbf{L}_2^n.$$

Note that, for every $j \geq 1$,

$$\boldsymbol{\beta}_{(j)}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}_{(j)} - 2\boldsymbol{\beta}_{(j)}'\mathbf{X}'\mathbf{Y}_{(j)} = (\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)}) - \mathbf{Y}_{(j)}'\mathbf{Y}_{(j)},$$

so that maximizing $\ell(\boldsymbol{\beta}, \mathbf{y})$ with respect to $\boldsymbol{\beta}$, independently of σ, is equivalent to minimizing

$$(\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)}).$$

Hence we have the following result

Proposition 2.3.2 *Viewed as a function of* $\boldsymbol{\beta}$, $\ell(\boldsymbol{\beta}, \mathbf{y})$ *is maximized for*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The two methods of estimation, LS and ML, as in the finite dimensional case lead to the same estimator of $\boldsymbol{\beta}$. However while the former is always applicable the latter is not. More precisely,

Remark 2.3.3 *1. As in the finite dimensional case, The least squares method does not require normality. Moreover, it does not even require any knowledge about the covariance structure.*

2. Unlike the finite dimensional case, in infinite dimensional spaces there is no direct equivalent to Lebesgue measure in $\mathbf{R}^n$*. However, if the error is Gaussian this difficulty can be circumvented by letting the Radon-Nikodym derivative (of the observation distribution with respect to the error distribution) play the role of the likelihood function. If the error is not Gaussian it is not easy to remove the difficulty.*

3. The inner product $< .,. >_T$ *is not an artificial one. It appears naturally in stochastic processes studies using the reproducing kernel Hilbert space approach. Furthermore, the likelihood function is maximized at*

$$\sum_{j\geq 1} \lambda \mathbf{Y}'_{(j)}\mathbf{Y}_{(j)} - \sum_{j\geq 1} \lambda_j(\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)})'(\mathbf{Y}_{(j)} - \mathbf{X}\boldsymbol{\beta}_{(j)}).$$

In terms of the inner product $< .,. >_T$*, one can easily see that this expression is analogous to the statistic SSR usually found in analysis of variance or regression models.*

Discussion

Now we present another aspect of the difference between the finite and infinite dimensional cases. It is clear that in the finite dimensional case the law of a Gaussian random variable X is equivalent to the law of σX for any positive real number σ, i.e., $d\mathcal{L}(X)/d\mathcal{L}(\sigma X)$ exists. Here we show that if ϵ is an $\mathbf{L}_2$-valued Gaussian random variable, then $\mathcal{L}(\epsilon)$ and $\mathcal{L}(\sigma\epsilon)$ are orthogonal.

Let (Z_k) be a sequence of *i.i.d.* N(0,1) random variables, (X_k) a sequence of *i.i.d.* $N(0, \sigma^2)$ random variables. Let $Z = (Z_k)$ and $X = (X_k)$. They are Gaussian random variables with values in $\mathbf{R}^\mathbf{N}$, the space of all real sequences. Using Kakutani's theorem on product measures [8], it can be shown that the probability measures $\mathbf{P}_Z = \mathcal{L}(Z)$ and $\mathbf{P}_X = \mathcal{L}(X)$ induced by Z and X on $\mathbf{R}^\mathbf{N}$ are orthogonal. A proof can also be found in [7, example 2, p. 115]. This proof is based on the following version of [7, Theorem 1, p. 107]. Before stating the version that we need, recall that by the Radon-Nikodym theorem, there exists $N \in \mathcal{B}_{\mathbf{R}^\mathbf{N}}$ such that $\mathbf{P}_Z\{N\} = 0$, and a non-negative function f integrable with respect to $\mathbf{P}_Z$ such that for any $E \in \mathcal{B}_{\mathbf{R}^\mathbf{N}}$

$$\mathbf{P}_X\{E\} = \int_E f(x)\mathbf{P}_Z(dx) + \mathbf{P}_X(E \cap N).$$

Theorem 2.3.2 (see Grenander, Theorem 1 p 107) *Let g_n^0 denote the distribution density of $\mathcal{L}(Z_1, \ldots, Z_n)$, and g_n^1 denote the distribution density of $\mathcal{L}(X_1, \ldots, X_n)$. Put*

$$f_n(x) = \frac{g_n^1(x^n)}{g_n^0(x^n)},$$

where $x = (x_1, x_2, \ldots) \in \mathbf{R}^{\mathbf{N}}$ and $X^n = (x_1, \ldots, x_n)$. then

$$\lim_{n \to \infty} f_n(x) = \infty \Longrightarrow \mathbf{P}_X\{N\} = 1$$

i.e., $\mathbf{P}_X$ and $\mathbf{P}_Z$ are orthogonal.

By definition of orthogonality of measures, there exist two measurable sets $E, F \in \mathcal{B}_{\mathbf{R}^{\mathbf{N}}}$ such that $E \cap F = \emptyset$, $\mathbf{P}\{Z \in E\} = 1$, and $\mathbf{P}\{X \in E\} = 0$. For $A \in \mathcal{B}_{\mathbf{R}^{\mathbf{N}}}$ define $A' \in \mathcal{B}_{\mathbf{L}_2}$ by

$$A' = \{f \in \mathbf{L}_2, (< f, \varphi_k^* >) \in A\} \quad \text{with} \quad \varphi_k^* = \sqrt{\lambda_k}\varphi_k.$$

Using the Karhünen-Loève expansion, we have $\mathbf{P}\{\epsilon \in E'\} = \mathbf{P}\{Z \in E\} = 1$, and $\mathbf{P}\{\sigma\epsilon \in E'\} = \mathbf{P}\{X \in E\} = 0$. Thus, since $E' \cap F' = \emptyset$, $\mathcal{L}(\epsilon)$ and $\mathcal{L}(\sigma\epsilon)$ are orthogonal.

These remarks end this chapter. In the next chapter we use $\hat{\beta}$ to develop tests for linear hypothesis.

Chapter 3

TESTING LINEAR HYPOTHESES

The mean-value function is assumed to be a linear combination of unknown functions $\beta_1, \ldots, \beta_p$, so that the observations are summarized by the form (2.1):

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \sigma\boldsymbol{\epsilon}(t), \quad t \in [0,1].$$

Our aim, in this section, is to make inferences on $\boldsymbol{\beta}$ by establishing tests for the hypothesis H:$\mathbf{K}\boldsymbol{\beta} = \mathbf{C}$. Throughout this chapter we suppose that $\boldsymbol{\epsilon}$ is Gaussian. As mentioned in Remark 2.1.1, there is no loss of generality in supposing that $\beta_1, \ldots, \beta_p \in H_T$. Thus we will suppose that $\boldsymbol{\beta} \in H_T^p$.

For the sake of clarity, we first treat the case where $\mathbf{K}$ and $\mathbf{C}$ are respectively the identity and the null matrices. The general case will also be treated.

3.1 Testing

We have seen that the method of least squares and the maximum likelihood method lead to the same estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Following the notation from classical linear models theory, we will define the residual error sum of squares SSE, the sum of squares due to regression SSR, and establish some of their properties. Using the definition of $\hat{\boldsymbol{\beta}}$, it is easy to see that

$$\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y},$$

where $\mathbf{I}$ is the identity matrix. Let $\mathbf{M} = \left(m_{ij}\right)_{i,j}$ denote the matrix $(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$. It is idempotent,(i.e., $\mathbf{MM} = \mathbf{M}$), symmetric and such that $\mathbf{MX} = \mathbf{O}$. Therefore, for each $k \geq 1$, $\sigma^{-2}\mathbf{Y}'_{(k)}\mathbf{M}\mathbf{Y}_{(k)}$ is a non-central $\chi^{2\prime}\Big(r(\mathbf{M}), \frac{1}{2\sigma^2}\boldsymbol{\mu}'_{(\mathrm{k})}\mathbf{M}\boldsymbol{\mu}_{(k)}\Big)$ random variable, where $r(\mathbf{M})$ is the rank of $\mathbf{M}$ and

$$\mathbf{Y}_{(k)} = \Big(< Y_1, \varphi_k >, \cdots, < Y_n, \varphi_k >\Big)'$$

is a Gaussian vector $\mathcal{N}\Big(\boldsymbol{\mu}_{(k)}, \sigma^2\mathbf{I}\Big)$ with mean,

$$\begin{aligned} \boldsymbol{\mu}_{(k)} &= \mathbf{E}\mathbf{Y}_{(k)} \\ &= \mathbf{X}\boldsymbol{\beta}_{(k)}. \end{aligned}$$

Therefore, $\mathbf{M}\boldsymbol{\mu}_{(k)} = \mathbf{0}$ and thus $\mathbf{Y}'_{(k)}\mathbf{M}\mathbf{Y}_{(\mathrm{k})}$ is a $\sigma^2\chi^2_{r(\mathrm{M})}$ random variable.

3.1.1 The SSE and SSR

By analogy with the classical linear models theory we define *SSE* by:

$$\begin{aligned} SSE &= \Big\langle \mathbf{Y} - \hat{\mathbf{Y}}, \mathbf{Y} - \hat{\mathbf{Y}} \Big\rangle_T \\ &= \Big\langle \mathbf{M}\mathbf{Y}, \mathbf{M}\mathbf{Y} \Big\rangle_T \\ &= \Big\langle \mathbf{Y}, \mathbf{M}\mathbf{Y} \Big\rangle_T, \end{aligned}$$

where $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.
Recall that, for each $i = 1, \ldots, n$, we have that:

$$Y_i = \sum_{k \geq 1} < Y_i, \varphi_k > \varphi_k,$$

where $(\varphi_k)_{k\geq 1}$ is the orthogonal basis of H_T the reproducing kernel Hilbert space associated with ϵ_1. On the other hand, from the definition of the inner product $< ., . >_T$, we have that:

$$\begin{aligned} \Big\langle \mathbf{Y}, \mathbf{M}\mathbf{Y} \Big\rangle_T &= \sum_{i=1}^{n} \Big\langle Y_i, \big(\mathbf{M}\mathbf{Y}\big)_i \Big\rangle_T \\ &= \sum_{i=1}^{n} \Big\langle Y_i, \sum_{j=1}^{n} m_{ij} Y_j \Big\rangle_T \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \Big\langle Y_i, m_{ij} Y_j \Big\rangle_T. \end{aligned}$$

Using the expansion of the Y_i and the orthonormality of the φ_k, the general term of the double sum becomes

$$\begin{aligned} \Big\langle Y_i, m_{ij} Y_j \Big\rangle_T &= \Big\langle \sum_{k\geq 1} < Y_i, \varphi_k > T\varphi_k, m_{ij} \sum_{l \geq 1} < Y_j, \varphi_l > \varphi_l \Big\rangle \\ &= \sum_{k \geq 1} \lambda_k < Y_i, \varphi_k > m_{ij} < Y_j, \varphi_k > . \end{aligned}$$

Substituting this in the previous formulae, we get

$$\begin{aligned}\left\langle \mathbf{Y}, \mathbf{MY} \right\rangle_T &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\geq 1} \lambda_k < Y_i, \varphi_k > m_{ij} < Y_j, \varphi_k > \\ &= \sum_{k\geq 1}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_k < Y_i, \varphi_k > m_{ij} < Y_j, \varphi_k > \\ &= \sum_{k\geq 1} \lambda_k \mathbf{Y}'_{(k)} \mathbf{M} \mathbf{Y}_{(k)}.\end{aligned}$$

This shows that

$$SSE = \sum_{k\geq 1} \lambda_k \mathbf{Y}'_{(k)} \mathbf{M} \mathbf{Y}_{(k)}.$$

We also define the total sum of squares SST to be

$$SST = \left\langle \mathbf{Y}, \mathbf{Y} \right\rangle_T.$$

Then the SSR is given by

$$\begin{aligned}SSR &= SST - SSE \\ &= \left\langle \mathbf{Y}, \mathbf{X}\hat{\beta} \right\rangle_T \\ &= \left\langle \mathbf{Y}, \mathbf{BY} \right\rangle_T,\end{aligned}$$

where $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is a symmetric and idempotent matrix. Furthermore, it is orthogonal to $\mathbf{M}$. We can now state and prove the following results.

Proposition 3.1.1 *1. There exists a sequence (η_k) of independent random variables with the same distribution $\chi^2_{r(\mathbf{M})}$, such that the series $\sigma^2 \sum_{k\geq 1} \lambda_k^2 \eta_k$ converges to SSE almost surely.*

2. There exists a sequence (ξ_k) of independent real random variables such that each ξ_k has a non-central $\chi^{2\prime}\big(r(\mathbf{B}), (2\sigma^2\lambda_k)^{-1}\boldsymbol{\mu}'_{(k)}\mathbf{B}\boldsymbol{\mu}_{(k)}\big)$ distribution, and the series $\sigma^2 \sum_{k\geq 1} \lambda_k^2 \xi_k$ converges to SSR almost surely. Also,

3. The SSR and SSE are independent.

Proof

1. We have seen that

$$\begin{aligned}SSE &= \sum_{k\geq 1} \lambda_k \mathbf{Y}'_{(k)} \mathbf{M} \mathbf{Y}_{(k)} \\ &= \sigma^2 \sum_{k\geq 1} \lambda_k^2 \Big(\frac{1}{\sigma\sqrt{\lambda_k}} \mathbf{Y}_{(k)}\Big)' \mathbf{M} \Big(\frac{1}{\sigma\sqrt{\lambda_k}} \mathbf{Y}_{(k)}\Big).\end{aligned}$$

Put

$$\eta_k = \Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big)'\mathbf{M}\Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big).$$

It is a $\chi^2\Big(r(\mathbf{M}), \frac{1}{2\sigma^2\lambda_k}\boldsymbol{\mu}'_{(k)}\mathbf{M}\boldsymbol{\mu}_{(k)}\Big)$ distributed random variable. Recall that $\boldsymbol{\mu}'_{(k)}\mathbf{M} = \mathbf{0}$. On the other hand we have, $\mathbf{E}[SSE] = \mathbf{E} < \mathbf{Y}, \mathbf{MY} >_T$ is finite and therefore SSE is almost surely finite.

2. Using the same arguments as for SSE we have that

$$\begin{aligned} SSR &= \sum_{k\geq 1} \lambda_k \mathbf{Y}'_{(k)}\mathbf{B}\mathbf{Y}_{(k)} \\ &= \sigma^2 \sum_{k\geq 1} \lambda_k^2 \Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big)'\mathbf{B}\Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big). \end{aligned}$$

Put

$$\xi_k = \Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big)'\mathbf{B}\Big(\frac{1}{\sigma\sqrt{\lambda_k}}\mathbf{Y}_{(k)}\Big).$$

It is a non-central $\chi^{2\prime}\Big(r(\mathbf{B}), (2\sigma^2\lambda_k)^{-1}\boldsymbol{\mu}'_{(k)}\mathbf{B}\boldsymbol{\mu}_{(k)}\Big)$ random variable. On the other hand we have $\mathbf{E}[SSR] = \mathbf{E} < \mathbf{Y}, \mathbf{BY} >_T$ is finite and therefore SSR is almost surely finite.

3. It suffices to show that the sequences $(\eta_k)_{k\geq 1}$ and $(\xi_k)_{k\geq 1}$ are mutually independent.
Since $\Big(\mathbf{Y}_{(k)}\Big)_{k\geq 1}$ is a sequence of mutually independent random vectors and since η_k and ξ_k are functions of $\mathbf{Y}_{(k)}$ alone, it suffices to show that for each $k \geq 1$, η_k and ξ_k are independent. We have $\eta_k = (\sigma\lambda_k)^{-1}\mathbf{Y}'_{(k)}\mathbf{M}\mathbf{Y}_{(k)}$ and $\xi_k = (\sigma\lambda_k)^{-1}\mathbf{Y}'_{(k)}\mathbf{B}\mathbf{Y}_{(k)}$ where $\mathbf{Y}_{(k)}$ is a Gaussian $\mathcal{N}\Big(\boldsymbol{\mu}_{(k)}, \sigma^2\lambda_k \mathrm{I}\Big)$ random vector. But, $\mathbf{BM} = \mathbf{MB} = \mathbf{0}$, thus η_k and ξ_k are independent. □

Corollary 3.1.1 *The statistic*

$$\frac{SSE}{r(\mathbf{M})\sum_k \lambda_k^2}$$

is an unbiased estimator for σ^2. Moreover we have

$$\lim_{n\to\infty} \frac{SSE}{r(\mathbf{M})\sum_k \lambda_k^2} = \sigma^2 \quad \text{a.s.}$$

Proof

By the first assertion of Proposition 3.2.1, and with the same notation, we have almost surely that $SSE = \sigma^2 \sum_k \lambda_k^2 \eta_k$. Therefore $\mathbf{E}[SSE] = r(\mathbf{M})\sigma^2 \sum_k \lambda_k^2$. Since the series terms are non-negative we have

$$\begin{aligned}
\lim_{n\to\infty} \sum_{k\geq 1} \lambda_k^2 \frac{\eta_k}{r(\mathbf{M})} &= \lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=1}^{N} \lambda_k^2 \frac{\eta_k}{r(\mathbf{M})} \\
&= \lim_{N\to\infty} \lim_{n\to\infty} \sum_{k=1}^{N} \lambda_k^2 \frac{\eta_k}{r(\mathbf{M})} \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} \lambda_k^2 \lim_{n\to\infty} \frac{\eta_k}{r(\mathbf{M})}.
\end{aligned}$$

On the other hand, $r(\mathbf{M}) = n - r(\mathbf{X}) = n - p$. Thus by the strong law of large numbers, for any integer $N \succeq 1$, we have almost surely

$$\sum_{k=1}^{N} \lambda_k^2 \lim_{n\to\infty} \frac{\eta_k}{r(\mathbf{M})} = \sum_{k\geq 1} \lambda_k^2.$$

This ends the proof. □

3.1.2 A particular case

To develop a test for the hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, we will use the maximum likelihood method. The likelihood function, as we have seen in §2.2.3, is given by

$$\ell(\boldsymbol{\beta}, \mathbf{y}) = \prod_{j\geq 1} \exp\Big\{\frac{\sigma^2\lambda_j}{2}\big[\beta'_{(j)}\mathbf{X}'\mathbf{X}\beta_{(j)} - 2\beta'_j\mathbf{X}'\mathbf{y}_{(j)}\big]\Big\}.$$

As a function of β, ℓ is maximized at $\hat{\boldsymbol{\beta}} = \big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}'\mathbf{y}$. We have that

$$\begin{aligned}
\ell(\hat{\boldsymbol{\beta}}, \mathbf{y}) &= \prod_{j\geq 1} \exp\Big\{\frac{\sigma^2\lambda_j}{2}\mathbf{y}'_{(j)}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(j)}\Big\} \\
&= \exp\Big\{\sum_j \frac{\sigma^2\lambda_j}{2}\mathbf{y}'_{(j)}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(j)}\Big\} \\
&= \exp\Big\{\frac{\sigma^2}{2} < \mathbf{y}, \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} >_T\Big\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\sup_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}, \mathbf{y}) &= \exp(\frac{\sigma^2}{2} SSR), \quad \text{and} \\
\sup_{\boldsymbol{\beta}=\mathbf{0}} \ell(\boldsymbol{\beta}, \mathbf{y}) &= 1,
\end{aligned}$$

therefore we have the following *maximum likelihood test.*

Theorem 3.1.1 *For testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, at level α, there exists a test ψ given by:*

$$\psi = \begin{cases} 1 & \text{if } S(\mathbf{Y}) > C(H_0, \alpha), \\ 0 & \text{otherwise} \end{cases}, \tag{3.1}$$

where $S(\mathbf{Y}) = SSR$ if σ is known, and $S(\mathbf{Y}) = SSR/SSE$ otherwise. The constant $C(H_0, \alpha)$ is such that $\mathbf{P}\{S(\mathbf{Y}) > C(H_0, \alpha), \boldsymbol{\beta} = \mathbf{0}\} = \alpha$.

3.1.3 The General Case

Consider the general hypothesis $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C}$, where $\mathbf{K}$ is an $m \times p$ matrix and $\mathbf{C}$ is an $m \times 1$ vector of given functions. The only restriction on $\mathbf{K}$ is that it is of full row rank, i.e., $rank(\mathbf{K}) = p$. This simply means that the linear functions of $\boldsymbol{\beta}$ which form the hypothesis H_0 are linearly independent. For testing this kind of hypothesis we need to compute an estimator of $\boldsymbol{\beta}$ under H_0. We will see that the *constrained* least squares method provides such an estimator. It will be denoted by $\tilde{\boldsymbol{\beta}}$.

Constrained Least Squares

The constrained least squares estimates are obtained by minimizing $\|\mathbf{e}(\boldsymbol{\beta})\|_T^2$ under the condition $\mathbf{K}\boldsymbol{\beta} - \mathbf{C} = \mathbf{0}$, which is equivalent to

$$\mathbf{K}\boldsymbol{\beta}_{(k)} - \mathbf{C}_{(k)} = \mathbf{0} \quad \forall k \geq 1. \tag{3.2}$$

Recall that

$$\|\mathbf{e}(\boldsymbol{\beta})\|_T^2 = \sum_{k \geq 1} \lambda_k^2 (\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)})'(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)}).$$

Therefore, to minimize $\|\mathbf{e}(\boldsymbol{\beta})\|_T^2$ subject to (3.2), it suffices to minimize, for each $k \geq 1$,

$$(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)})'(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)})$$

subject to

$$\mathbf{K}\boldsymbol{\beta}_{(k)} - \mathbf{C}_{(k)} = \mathbf{0}.$$

Using Lagrange multiplier vectors, it can easily be seen that

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}' \left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1} (\mathbf{K}\hat{\boldsymbol{\beta}} - \mathbf{C}). \tag{3.3}$$

Remark 3.1.1 *1. Note that since $rank(\mathbf{K}) = p < m = rank((\mathbf{X}'\mathbf{X})^{-1})$, we have that $rank(\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}') = p$. Thus $\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'$ is invertible since it is of full rank.*

2. *Recall that, in Chapter 2, the maximum likelihood estimator has been obtained by minimizing, for each $k \geq 1$,*

$$(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)})'(\mathbf{Y}_{(k)} - \mathbf{X}\boldsymbol{\beta}_{(k)}).$$

Therefore, the constrained maximum likelihood obtained by maximizing the likelihood function subject to (3.2) *is equal to* $\tilde{\boldsymbol{\beta}}$.

Let SSE_{H_0} denote the residual error sum of squares under the hypothesis $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C}$, i.e.,

$$SSE_{H_0} = \|\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}\|_T^2.$$

Writing

$$\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}),$$

and using the fact that $\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}$ we can see that

$$SSE_{H_0} = \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|_T^2 + \|\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\|_T^2. \tag{3.4}$$

From (3.3), we have that

$$\begin{aligned}
\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}(\mathbf{K}\hat{\boldsymbol{\beta}} - \mathbf{C}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{C}\right] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[\mathbf{Y} - \mathbf{X}\mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{C}\right].
\end{aligned}$$

Thus $\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \mathbf{D}(\mathbf{Y} - \mathbf{C}^*)$, with

$$\begin{aligned}
\mathbf{D} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \quad \text{and} \\
\mathbf{C}^* &= \mathbf{X}\mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}\mathbf{C}.
\end{aligned}$$

Note that $\mathbf{D}$ is idempotent, symmetric and $\mathbf{DM} = \mathbf{MD} = \mathbf{0}$. In this form we can now give the distribution of $\mathbf{Q} = \|\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\|_T^2$ and show that it is stochastically independent of SSE.

Proposition 3.1.2 1. *The real random variables* $\mathbf{Q}$ *and* SSE *are stochastically independent.*

2. *There exists a sequence* (ζ_k) *of independent real random variables such that each* ζ_k *has a non-central* $\chi^{2\prime}(r(\mathbf{D}), \delta_k)$ *distribution, and such that the series* $\sigma^2 \sum_k \lambda_k^2 \zeta_k$ *converges to* $\mathbf{Q}$ *almost surely. Further*

$$\delta_k = (2\sigma^2\lambda_k)^{-1}(\mathbf{E}\mathbf{Y}_{(k)} - \mathbf{C}^*_{(k)})'\mathbf{D}(\mathbf{E}\mathbf{Y}_{(k)} - \mathbf{C}^*_{(k)}),$$

so that under $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C}$, $\delta_k = 0$, *for every* $k \geq 1$.

Proof

The last assertion is straightforward, the remaining assertions can be proved as was Proposition 3.1.1. □

Define the sum of squares due to regression under the hypothesis $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C}$ by $SSR_{H_0} = SST - SSE_{H_0}$. From (3.4), we obtain

$$SSR_{H_0} - SSR = \mathbf{Q},$$

and thus we have the following result.

Theorem 3.1.2 *For testing $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{C}$ versus $H_1 : \mathbf{K}\boldsymbol{\beta} \neq \mathbf{C}$, at level α, there exists a test ψ given by:*

$$\psi = \begin{cases} 1 & \text{if } S_{H_0}(\mathbf{Y}) > C(H_0, \alpha), \\ 0 & \text{otherwise} \end{cases}, \tag{3.5}$$

where $S_{H_0}(\mathbf{Y}) = SSR_{H_o} - SSR$ if σ is known, and $S_{H_0}(\mathbf{Y}) = (SSR_{H_o} - SSR)/SSE$ otherwise. The constant $C(H_0, \alpha)$ is such that $\mathbf{P}\{S_{H_0}(\mathbf{Y}) > C(H_0, \alpha), \mathbf{K}\boldsymbol{\beta} = \mathbf{C}\} = \alpha$.

3.2 Confidence Sets

Recall that the model is given by

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \sigma\epsilon(t), \quad t \in [0, 1],$$

where σ is an unknown scalar.
Following the definitions in [15, §1.4, p. 13], we define estimable functions as follows:

Definition 3.2.1 *A parametric function Φ is defined to be a linear function of the unknown parameters $\{\beta_1, \ldots, \beta_p\}$ with known constant coefficients $\{c_1, \ldots, c_p\}$, i.e.,*

$$\Phi = \sum_{j=1}^{p} c_j \beta_j. \tag{3.6}$$

In vectorial notation, equation (3.6) can be written $\Phi = \mathbf{c}'\boldsymbol{\beta}$, with $\mathbf{c} = (c_1, \ldots, c_p)'$.

Definition 3.2.2 *A parametric function $\Phi = \mathbf{c}'\boldsymbol{\beta}$ is called an estimable function if it has an unbiased linear estimate, in other words, if there exists a vector $\mathbf{a} \in \mathbf{R}^n$ of constants such that*

$$\mathbf{E}[\mathbf{a}'\mathbf{Y}] = \Phi,$$

identically in $\boldsymbol{\beta}$ (i.e. no matter what the true values of $\{\beta_1, \ldots, \beta_p\}$ are).

Remark 3.2.1 Theorem 1 in [15, p. 13] *states that* Φ *is estimable if and only if there exists* $\mathbf{a} \in \mathbf{R}^n$ *such that*

$$\mathbf{c}' = \mathbf{a}'\mathbf{X}$$

Let $\mathbf{\Phi} = (\Phi_1, \ldots, \Phi_q)$ *be an element of* $\mathbf{L}_1^q$ *whose components are linearly independent estimable functions. Thus there exists a* $q \times p$ *matrix* $\mathbf{C}$ *and a* $q \times n$ *matrix* $\mathbf{A}$ *such that* $\mathbf{\Phi}=\mathbf{C}\boldsymbol{\beta}$ *and* $\mathbf{C}=\mathbf{A}\mathbf{X}'$. *Furthermore,* $\mathbf{C}$ *and* $\mathbf{A}'\mathbf{A}$ *are of rank* q.

The following theorem is an analogue to the one stated in [15, p. 26].

Theorem 3.2.1 *Let* $\mathbf{\Phi}$ *be as in the remark. If* ϵ *is Gaussian, then the least squares estimator* $\hat{\mathbf{\Phi}} = \mathbf{C}\hat{\boldsymbol{\beta}}$, *of* Φ *is Gaussian and is stochastically independent of* $SSE = \|\mathbf{Y} - \hat{\mathbf{Y}}\|$.

Proof
Recall that the least squares estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

so that

$$\begin{aligned} \hat{\mathbf{\Phi}} &= \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \end{aligned}$$

and thus $\hat{\mathbf{\Phi}}$ is Gaussian.
To prove the independence of $\boldsymbol{\xi}$ and SSE, recall the notation

$$\boldsymbol{\xi}_{(k)} = (<\xi_1, \varphi_k>, \ldots, <\xi_r, \varphi_k>)', \quad \forall \boldsymbol{\xi} \in \mathbf{L}_2^r.$$

We have (see for example [16], Theorem 3, p. 59) that $\mathbf{Y}_{(k)}$ and $\hat{\mathbf{\Phi}}_{(k)}$ are independent since

$$\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \mathbf{0}.$$

On the other hand, it is easy to see that for any pair (k, k') of integers such that $k \neq k'$, $\mathbf{Y}_{(k)}$ and $\hat{\mathbf{\Phi}}_{(k')}$ are independent. □

We have that

$$\begin{aligned} \hat{\Phi} - \Phi &= \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{A}\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{X}\boldsymbol{\beta} - \sigma\epsilon] - \mathbf{A}\mathbf{X}\boldsymbol{\beta} \\ &= \sigma\mathbf{A}\mathbf{B}\epsilon, \end{aligned}$$

where $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Since $\mathbf{A}\mathbf{B}\mathbf{A}'$ is symmetric and of full rank it is non-singular. Let $(\mathbf{A}\mathbf{B}\mathbf{A}')^{-1/2}$ denote the matrix whose square is $(\mathbf{A}\mathbf{B}\mathbf{A}')^{-1}$, and define

$$\mathbf{D} = \mathbf{B}\mathbf{A}'(\mathbf{A}\mathbf{B}\mathbf{A}')^{-1}\mathbf{A}\mathbf{B}.$$

We have the following result.

Proposition 3.2.1 *There exists a sequence* (ζ_k) *of independent random variables with the same distribution* $\chi^2_{r(\mathbf{D})}$*, such that*

$$\sum_{k\geq 1} \lambda_k \zeta_k = \sigma^{-2} \|(\mathbf{ABA}')^{-1/2}(\hat{\Phi} - \Phi)\|_T^2 \quad \text{a.s..}$$

Proof

By lemma 1.3.1 we have that

$$\|(\mathbf{ABA}')^{-1/2}(\hat{\Phi} - \Phi)\|_T^2 = \sigma^2 \sum_{k\geq 1} \epsilon'_{(k)} \mathbf{D} \epsilon_{(k)}.$$

On the other hand $\mathbf{D}$ is idempotent and the $\epsilon_{(k)}$ are independent and identically distributed $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Therefore the $\epsilon'_{(k)}\mathbf{D}\epsilon_{(k)}$ are independent and identically distributed $\chi^2_{r(\mathbf{D})}$. That the series converges a.s. to $\sigma^{-2}\|(\mathbf{ABA}')^{-1/2}(\hat{\Phi} - \Phi)\|_T^2$ follows from $\hat{\Phi} - \Phi = \sigma\mathbf{AB}\epsilon$. □

The proposition and theorem above together with the fact that $SSE/\lambda^2 r(\mathbf{M})$ (with $\lambda^2 = \sum_k \lambda_k^2$) is an unbiased estimator for σ^2 provide a confidence set for Φ.

Corollary 3.2.1 *With the same notation as before, there exists a* $(100 - \alpha)\%$ *confidence set* C_Φ *for* Φ *given by*

$$C_\Phi = \{\mathbf{f} \in \mathbf{L}_2^q, \|(\mathbf{ABA}')^{-1/2}(\hat{\Phi} - \mathbf{f})\|_T^2 \leq c_\alpha SSE/\lambda^2 r(\mathbf{M})\}.$$

The constant c_α *is given by*

$$\mathbf{P}\left\{\frac{\sum_{k\geq 1}\lambda_k\zeta_k}{\sum_{k\geq 1}\lambda_k\eta_k} \geq \frac{c_\alpha}{\lambda^2 r(\mathbf{M})}\right\} = \alpha,$$

where (ζ_k) *is as above,* (η_k) *is as in Proposition 3.2.1 and such that the two sequences are independent.*

The results we have seen in this chapter have been obtained under the normality assumption. In the next chapter we will see how these results extend to some non-normal cases.

Chapter 4

EXTENSIONS

In this chapter we consider the model given by (2.1)

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \sigma\epsilon(t), \quad t \in [0,1],$$

but we do not suppose that the error is Gaussian. However, its covariance function is assumed to be square-integrable so that it defines a self-adjoint and compact operator T_ϵ. On the other hand, for testing hypothesis, the normality assumption is replaced by an asymptotic condition on the design matrix that ensures the central limit theorem.

4.1 Consistency and Normality of the Estimators

Replacing ϵ by $\sigma\epsilon$, if necessary, we will suppose in this section that $\sigma = 1$. For making inferences about $\boldsymbol{\beta}$, it is the normality of the estimators rather than the observations that is required. Since the estimators are linear combinations of the observations, it is possible to invoke the *central limit theorem* to show that, under some conditions, the estimators are asymptotically Gaussian even if the the observations are not.

Before stating the result recall that the least squares estimator, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, of $\boldsymbol{\beta}$ satisfies

$$\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon.$$

The constrained least squares estimator of $\boldsymbol{\beta}$, under the constraint $\mathbf{K}\boldsymbol{\beta}$=$\mathbf{C}$, is given by

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\left[\mathbf{K}\hat{\boldsymbol{\beta}} - \mathbf{C}\right],$$

where $\mathbf{K}$ is a given matrix of full raw rank, and $\mathbf{C}$ is a given vector. Note that

$$\mathbf{K}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) = \mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon + \mathbf{K}\boldsymbol{\beta} - \mathbf{C}.$$

Also recall that $\mathbf{B}$=(b_{ij}) denotes the $n \times n$ matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

Theorem 4.1.1 *There exists a mean zero, $\mathbf{L}_2^p$-valued Gaussian random variable $\mathbf{G} = (G_1, \ldots, G_p)'$ with stochastically independent components, the same covariance function as ϵ_1, i.e.,*

$$\forall s, t \in [0,1], \quad cov(G_1(s), G_1(t)) = cov(\epsilon_1(s), \epsilon_1(t)),$$

and such that if

$$\lim_{n\to\infty} \max_{i=1}^{n} |b_{jj}| = 0,$$

then

$$[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\left[\mathbf{K}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) - (\mathbf{K}\boldsymbol{\beta} - \mathbf{C})\right] \xrightarrow{\mathcal{L}} \mathbf{G} \quad \text{as} \quad n \to \infty,$$

This theorem is due to Srivastava [19] in the finite dimensional case (see for example [17]). We will give its proof in the Hilbert space case for the reader's convenience. But first note that if we take **K**=**I** and **C**=$\boldsymbol{\beta}$, i.e., without any constraint, the conclusion of the theorem becomes

$$(\mathbf{X}'\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \mathbf{G} \quad as \quad n \to \infty. \tag{4.1}$$

Using this we prove the following consistency result.

Theorem 4.1.2 *With the same notation as (4.1) we have that*

$$\lim_{n\to\infty} \max_{j} |b_{jj}| = 0 \Longrightarrow \sup_{i=1}^{n} \left(\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)_i \to 0 \text{ in probability} \quad n \to \infty.$$

Proof of Theorem 4.1.2
We have that

$$\begin{aligned}
\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}(\mathbf{X}'\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\left[\mathbf{G} + (\mathbf{X}'\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mathbf{G}\right] \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G} + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{r},
\end{aligned}$$

where $\mathbf{r} = (\mathbf{X}'\mathbf{X})^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \mathbf{G}$. By *Skorohod's theorem*, (see [18] or [4]), we can suppose that the convergence in (4.1) occurs almost surely so that **r** converges to zero almost surely. Therefore we have only to prove that

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G} \overset{a.s}{\to} 0 \quad as \quad n \to \infty.$$

Let $t \in [0,1]$ be fixed. Let (a'_{ij}) denote the $n \times p$ matrix $\mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}$. For $i = 1, \ldots, n$, we have,

$$\mathbf{E}\left[\sum_{j=1}^{p} a'_{ij} G_j(t)\right]^2 = \sum_{j=1}^{p} a'^2_{ij}\, \mathbf{E}\left[G_1(t)\right]^2.$$

On the other hand, since $\mathbf{G}(t)$ has independent components,

$$\begin{aligned} cov\left(\mathbf{A}'\mathbf{G}(t)\right) &= \mathbf{A}'\mathbf{A}\mathbf{E}\left[G_1(t)\right]^2 \\ &= \mathbf{E}\left[G_1(t)\right]^2 (b_{ij}) \end{aligned}$$

where (b_{ij}) denotes the $n \times n$ matrix $\mathbf{A}'\mathbf{A}$. Therefore

$$\lim_{n\to\infty} \sum_{j=1}^{p} a_{ij}'^2 = \lim_{n\to\infty} b_{ii} = 0.$$

Let $\delta > 0$, and $N \geq 1$,

$$\begin{aligned} \sup_{i=1}^{n} \left[\sum_{j=1}^{p} a'_{ij} G_j(t) \right]^2 &\leq \sum_{j=1}^{p} G_j^2(t) \sup_{i=1}^{n} \sum_{j=1}^{p} a_{ij}'^2 \\ &= \sup_{i=1}^{n} b_{ii} \sum_{j=1}^{p} G_j^2(t). \end{aligned}$$

Therefore

$$\lim_{n\to\infty} max_{i=1}^{n} b_{jj} = 0 \Longrightarrow \lim_{n\to\infty} \sup_{i=1}^{n} \sum_{j=1}^{p} a'_{ij} G_j(t) = 0 \quad \text{a.s.}$$

This proves that, $\forall t \in [0,1]$,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G}(t) \to 0 \quad as \;\; n \to \infty.$$

We have proved that for each $t \in [0,1]$, there exists a negligible Ω_t such that

$$\forall \omega \in \Omega - \Omega_t \Longrightarrow \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G}(t) \to 0 \quad as \;\; n \to \infty.$$

Let $S = \mathbf{Q} \cap [0,1]$, and put $\Omega_0 = \cup_{t\in S}\Omega_t$, which is a negligible set such that

$$\forall \omega \in \Omega - \Omega_0 \Longrightarrow \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G}(t) \to 0 \quad as \;\; n \to \infty, \quad \forall t \in S.$$

Since S is dense in $[0,1]$, and $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{G}$ is almost surely continuous, we have the result. □

Proof of Theorem 4.1.1
Let s be an integer and $(\mathbf{A}_n)_{n\geq 1}$ a sequence of matrices such that for every $n \geq 1$,

- $\mathbf{A}_n = (a_{i,j})$ is a $s \times n$ matrix where $a_{i,j} = a_{ij}(n)$, and
- $\mathbf{A}_n\mathbf{A}'_n = \mathbf{I}_{s\times s}$.

The second condition can be expressed as

$$\sum_{j=1}^{n} a_{ij}^2 = 1 \quad for\ i = 1, \ldots, s, \quad and \quad \sum_{j=1}^{n} a_{ij}a_{kj} = 0 \quad for\ i \neq k. \tag{4.2}$$

Let $(b_{jj}^*)_{j=1,\cdots,n}$ denote the diagonal elements of $\mathbf{A}_n'\mathbf{A}_n$, and $(\epsilon_n)_{n\geq 1}$ be a sequence of independent and identically distributed $\mathbf{L}_2$-valued random variables such that:

$$\mathbf{E} < f, \epsilon >= 0\ \forall f \in \mathbf{L}_2 \quad and \quad \mathbf{E}\|\epsilon_1\|^2 < \infty.$$

We have the following central limit theorem.

Proposition 4.1.1 *If* $\max_{j=1}^{n} |b_{jj}^*| \to 0$, *as* $n \to \infty$, *then* $(\mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)')_{n\geq 1}$ *converges in distribution to an* $\mathbf{L}_2^s$*-valued mean zero Gaussian random variable,* G *say. Furthermore,* G *has the same covariance structure as* ϵ_1.

Proof
We have that $\mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)' = (\sum_{j=1}^{n} a_{1j}\epsilon_j, \ldots, \sum_{j=1}^{n} a_{pj}\epsilon_j)'$ is an element of $\mathbf{L}_2^s$. We will show that there exists a random variable $\mathbf{G}$ in $\mathbf{L}_2^s$ such that

$$\forall \mathbf{f} \in \mathbf{L}_2^s, \quad < \mathbf{f}, \mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)' > \xrightarrow{\mathcal{L}} < \mathbf{f}, \mathbf{G} > . \tag{4.3}$$

Condition (4.3) is actually equivalent to

$$\mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)' \xrightarrow{\mathcal{L}} \mathbf{G}.$$

Indeed, a necessary and sufficient condition for a sequence $(\xi_k) \subset \mathbf{L}_2$ to converge in distribution to $\xi \in \mathbf{L}_2$ is

$$\forall f \in \mathbf{L}_2, \quad < f, \xi_k > \xrightarrow{\mathcal{L}} < f, \xi > .$$

Let $\mathbf{f} = (f_1, \ldots, f_s)'$ be an arbitrary, but fixed, element of $\mathbf{L}_2^s$, we have

$$\begin{aligned} < \mathbf{f}, \mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)' > &= \sum_{i=1}^{s} < f_i, \sum_{j=1}^{n} a_{ij}\epsilon_j > \\ &= \sum_{j=1}^{n}\sum_{i=1}^{s} a_{ij} < f_i, \epsilon_j > . \end{aligned}$$

Put $Z_j = \sum_{i=1}^{s} a_{ij} < f_i, \epsilon_j >$. The Z_j are mean zero, independent and we have

$$\mathbf{E}[Z_j]^2 = \sum_{i=1}^{s} a_{ij}^2 \mathbf{E}[< f_i, \epsilon_1 >]^2 + \sum_{i\neq k} a_{ij}a_{kj}\mathbf{E}[< f_i, \epsilon_1 >< f_k, \epsilon_1 >].$$

Put $\sigma_j^2 = \mathbf{E}[Z_j]^2$, and $v_n^2 = \sum_{j=1}^n \sigma_j^2$. By (4.2) we have $v_n^2 = \sum_{i=1}^s \mathbf{E}[< f_i, \epsilon_1 >]^2$. Indeed

$$\begin{aligned} v_n^2 &= \sum_{j=1}^n \sum_{i=1}^s a_{ij}^2 \mathbf{E}[< f_i, \epsilon_1 >]^2 \\ &\quad + \sum_{j=1}^n \sum_{i \neq k} a_{ij} a_{kj} \mathbf{E}[< f_i, \epsilon_1 >< f_k, \epsilon_1 >] \\ &= \sum_{i=1}^s \mathbf{E}[< f_i, \epsilon_1 >]^2 \underbrace{\sum_{j=1}^n a_{ij}^2}_{=1} \\ &\quad + \sum_{i \neq k} \mathbf{E}[< f_i, \epsilon_1 >< f_k, \epsilon_1 >] \underbrace{\sum_{j=1}^n a_{ij} a_{kj}}_{=0} \\ &= \sum_{i=1}^s \mathbf{E}[< f_i, \epsilon_1 >]^2. \end{aligned}$$

Since v_n^2 does not depend on n it will be denoted by v^2.
The convergence in (4.3) will hold if the Z_j fulfill Lindeberg 's condition:

$$\forall u > 0 \quad \lim_{n \to \infty} \frac{1}{v^2} \sum_{j=1}^n \int_{|Z_j| > u\,v} Z_j^2 d\mathbf{P} = 0.$$

Note that

$$\begin{aligned} |Z_j| &\leq \sum_{i=1}^s |a_{ij}| \, |< f_i, \epsilon_j >| \\ &\leq \|\epsilon_j\| \, (\sum_{i=1}^s a_{ij}^2 \sum_{i=1}^s \|f_i\|^2)^{1/2}, \end{aligned}$$

and that $\sum_{i=1}^n a_{ij}^2 = b_{jj}^*$. Therefore

$$\{|Z_j| > u\,v\} \subset \{\|\epsilon_j\| > u\,v(\|\mathbf{f}\|^2 \max_j b_{jj}^*)^{-1/2}\}, \quad \text{and}$$

$$\begin{aligned} \int_{|Z_j| > u\,v} Z_j^2 d\mathbf{P} &\leq \int_{\|\epsilon_j\| > \delta_n} Z_j^2 d\mathbf{P} \\ &= \mathbf{E}[Z_j \mathbf{1}_{\|\epsilon_j\| > \delta_n}]^2, \end{aligned}$$

where $\delta_n = u\,v(\|\mathbf{f}\|^2 \max_{j=1}^n b_{jj}^*)^{-1/2}$. On the other hand we have that

$$\mathbf{E}[Z_j \mathbf{1}_{\|\epsilon_j\| > \delta_n}]^2 = \mathbf{E}[(\sum_{i=1}^s a_{ij}^2 < f_i, \epsilon_j >^2 +$$

$$\sum_{i\neq k} a_{ij}a_{kj} < f_i, \epsilon_j >< f_k, \epsilon_j >)\mathbf{1}_{\|\epsilon_j\|>\delta_n}]$$

$$= \sum_{i=1}^{s} a_{ij}^2 \mathbf{E}[< f_i, \epsilon_j >^2 \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}] +$$

$$\sum_{i\neq k} a_{ij}a_{kj}\mathbf{E}[< f_i, \epsilon_1 >< f_k, \epsilon_1 > \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}],$$

and hence by the condition (4.2), we have

$$\sum_{i=1}^{n} \mathbf{E}[Z_j \mathbf{1}_{\|\epsilon_j\|>\delta_n}]^2 = \sum_{i=1}^{s} \mathbf{E}[< f_i, \epsilon_1 >^2 \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}] \underbrace{\sum_{j=1}^{n} a_{ij}^2}_{=1} +$$

$$\sum_{i\neq k} \mathbf{E}[< f_i, \epsilon_1 >< f_k, \epsilon_1 > \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}] \underbrace{\sum_{j=1}^{n} a_{ij}a_{kj}}_{=0}$$

$$= \sum_{i=1}^{s} \mathbf{E}[< f_i, \epsilon_1 >^2 \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}].$$

Recall that

$$\begin{aligned} \lim_{n\to\infty} \delta_n &= \lim_{n\to\infty} u\, v(\|\mathbf{f}\|^2 \max_{j=1}^{2} b_{jj}^*)^{-1/2} \\ &= u\, v \lim_{n\to\infty} (\|\mathbf{f}\|^2 \max_{j=1}^{2} b_{jj}^*)^{-1/2} \\ &= \infty, \end{aligned}$$

and that for every $g \in \mathbf{L}_2$, $\mathbf{E}[< g, \epsilon_1 >]^2 < \infty$. Therefore

$$\lim_{n\to\infty} \sum_{i=1}^{s} \mathbf{E}[< f_i, \epsilon_1 >^2 \ \mathbf{1}_{\|\epsilon_1\|>\delta_n}] = 0.$$

Since Lindeberg's condition is fulfilled we have that

$$\sum_{j=1}^{n} \sum_{i=1}^{s} a_{ij} < f_i, \epsilon_j > \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2) \quad as\ n \to \infty,$$

where $v^2 = \sum_{i=1}^{s} \mathbf{E}[< f_i, \epsilon_1 >]^2$. Let $G_1, \ldots, G_p$ be independent and identically distributed $\mathbf{L}_2$-valued mean zero Gaussian random variables having the same structure of covariance as ϵ_1, and put $\mathbf{G} = (G_1, \ldots, G_s)'$. The real random variable

$$< \mathbf{f}, \mathbf{G} > = \sum_{i=1}^{s} < f_i, G_i >$$

is Gaussian $\mathcal{N}(0, v^2)$. This shows that

$$< \mathbf{f}, \mathbf{A}_n(\epsilon_1, \ldots, \epsilon_n)' > \xrightarrow{\mathcal{L}} < \mathbf{f}, \mathbf{G} > \quad as\ n \to \infty,$$

and since $\mathbf{f}$ is arbitrary, the proposition is proved.

Proof of Theorem 4.1.1 resumed

To see that Theorem 4.1.1 holds, note that

$$[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\left[\mathbf{K}(\tilde{\beta}-\hat{\beta})-(\mathbf{K}\beta-\mathbf{C})\right]$$
$$=[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon,$$

and put

$$\mathbf{A}_n=[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

so that

$$\mathbf{A}_n\mathbf{A}_n'=\mathbf{I}_{\mathrm{s}\times\mathrm{s}}.$$

It only remains to show that

$$\lim_{n\to\infty}\max_{j=1}^{n}|b_{jj}|=0\Longrightarrow\lim_{n\to\infty}\max_{j=1}^{n}|b^*_{jj}|=0,$$

where

$$\begin{aligned}\mathbf{B}&=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'=(b_{jj}),\quad\text{and}\\ \mathbf{A}_n'\mathbf{A}_n&=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'=(b^*_{jj}).\end{aligned}$$

This involves simple linear algebra that we recall for the reader's convenience. Noting that

$$\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'=\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

we can write $\mathbf{A}_n'\mathbf{A}_n=\mathbf{B}_1\mathbf{B}$, with

$$\mathbf{B}_1=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Now since $\mathbf{B}_1=(b_{1,ij})$ is symmetric and idempotent, its eigenvalues are either 0 or 1. Furthermore there exist a matrix $\mathbf{M}_1=(m_{1,ij})$ which is orthonormal ($\mathbf{M}_1\mathbf{M}_1'=\mathbf{M}_1'\mathbf{M}_1=\mathbf{I}$), and a diagonal matrix $\mathbf{D}_1=(d_{1,ij})$ whose diagonal elements are the eigenvalues of $\mathbf{B}_1$, such that $\mathbf{B}_1=\mathbf{M}_1\mathbf{D}_1\mathbf{M}_1'$.

Using the symmetry of $\mathbf{B}$ and the Cauchy-Schwarz inequality we have

$$\begin{aligned}b^*_{ii}&=\sum_{j=1}^{p}b_{1,ij}\,b_{ij}\\&\le[\sum_{j=1}^{p}b^2_{1,ij}]^{1/2}[\sum_{j=1}^{p}b^2_{ij}]^{1/2}\\&=\sqrt{b_{1,ii}}\sqrt{b_{ii}}\\&\le\sqrt{b_{ii}}.\end{aligned}$$

The last inequality holds because

$$b_{1,ii}=\sum_{j=1}^{p}d_{1,jj}\,b^2_{1,ij}\le 1.$$

This ends the proof of Theorem 4.1.1. □

4.2 SSR and SSE for Non-Gaussian Error

As mentioned in the introduction, we consider the model given by:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sigma\epsilon,$$

with an error that is not necessarily Gaussian. However, we suppose that its $\mathbf{L}_2$-norm has a finite second moment, i.e., $\mathbf{E}\|\epsilon\|_2^2 < \infty$. This is a necessary condition for a central limit theorem to hold in $\mathbf{L}_2$, see [20]. This condition, ensures that the covariance operator T associated with ϵ is compact, (see Theorem 1.1.1 and Proposition 1.1.1), and thus has a countable set of eigenvalues. Let (λ_k) and (φ_k) denote respectively, as usual, the eigenvalues and the eigenfunctions of T. For the sake of clarity, we treat first the unconstrained case and then the constrained case.

The Unconstrained Case

Let $\mathbf{A}$ denote the $n \times p$ matrix given by $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1/2}\mathbf{X}'$. We have that

$$\mathbf{A}\mathbf{A}' = \mathbf{I}_{p\times p}, \quad \text{and} \quad \mathbf{A}'\mathbf{A} = \mathbf{B},$$

where $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Put $\mathbf{B} = (b_{ij})_{1\leq i,j\leq n}$.

As mentioned in the previous section, to make inferences about the parameters, it is the normality of the estimators that is needed. Such "normality" has been obtained under the condition

$$\lim_{n\to\infty} \max_{i=1}^{n} b_{ii} = 0. \tag{4.4}$$

Throughout this section we will suppose that (4.4) holds.

The least squares estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and the SSE and SSR are given by

$$SSE = <\mathbf{MY}, \mathbf{Y}>_T, \quad \text{and} \quad SSR = <\mathbf{BY}, \mathbf{Y}>_T,$$

where $\mathbf{M} = \mathbf{I_{nxn}} - \mathbf{B}$.

Asymptotic distribution of SSR

From the definition of SSR, and since $\mathbf{B}$ is symmetric and idempotent, (i.e. $\mathbf{BB}=\mathbf{B}$), we have

$$\begin{aligned} SSR &= \;=<\mathbf{BY}, \mathbf{BY}>_T \\ &= <\mathbf{BX}\boldsymbol{\beta} + \sigma\mathbf{B}\epsilon, \mathbf{BX}\boldsymbol{\beta} + \sigma\mathbf{B}\epsilon>_T \\ &= <\mathbf{BX}\boldsymbol{\beta}, \mathbf{BX}\boldsymbol{\beta}>_T + 2\sigma <\mathbf{BX}\boldsymbol{\beta}, \mathbf{B}\epsilon>_T + \sigma^2 <\mathbf{B}\epsilon, \mathbf{B}\epsilon>_T \end{aligned}$$

Under the null hypothesis $H_0 : \boldsymbol{\beta} = 0$, only the last term is of interest. Note that the last term can be written as

$$\begin{aligned} < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T &= < \mathbf{A}'\mathbf{A}\epsilon, \mathbf{A}'\mathbf{A}\epsilon >_T \\ &= < \mathbf{A}\mathbf{A}'\mathbf{A}\epsilon, \mathbf{A}\epsilon >_T \\ &= < \mathbf{I}_{\mathbf{p}\times\mathbf{p}}\mathbf{A}\epsilon, \mathbf{A}\epsilon >_T . \end{aligned}$$

By Proposition 4.2.1, and since the inner product $< .,. >_T$ is continuous with respect to the usual norm (see (1.4)) $\|.\|_2$, we have that

$$< \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T \xrightarrow{\mathcal{L}} < \mathbf{G}, \mathbf{G} >_T, \qquad \text{as} \ \to \infty,$$

where $\mathbf{G}=(G_1, \ldots, G_p)'$ is a mean zero Gaussian $r.v$ with independent and identically components satisfying

$$\forall s, t \in [0,1], \quad cov(G_1(s), G_1(t)) = cov(\epsilon_1(s), \epsilon_1(t)).$$

Proposition 4.2.1 *There exists a sequence (η_k) of independent random variables with the same distribution χ_p^2 such that, if $\boldsymbol{\beta}=\mathbf{0}$,*

$$SSR \xrightarrow{\mathcal{L}} \sigma^2 \sum_{k \geq 1} \lambda_k^2 \eta_k \qquad \text{as} \ \to \infty.$$

Proof

We know that there exists a sequence (η_k) of independent random variables with the same distribution χ_p^2 such that the series $\sum_{k\geq 1} \lambda_k^2 \eta_k$ converges almost surely to $< \mathbf{G}, \mathbf{G} >_T$. Since, for $\boldsymbol{\beta} = \mathbf{0}$, SSR converges in distribution to $\sigma^2 < \mathbf{G}, \mathbf{G} >_T$, The result follows. □

The asymptotic distribution of SSR would be completely known if we knew σ^2. For large sample sizes we can approximate σ^2 by $SSE/(nc)$, where c is a constant that depends only on the error covariance, i.e., T.

Approximation of σ^2

We have that

$$\begin{aligned} \sigma^{-2} SSE &= < \mathbf{M}\epsilon, \mathbf{M}\epsilon >_T \\ &= < \epsilon, \epsilon >_T -2 < \epsilon, \mathbf{B}\epsilon >_T + < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T \\ &= < \epsilon, \epsilon >_T - < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T . \end{aligned}$$

The last equality holds because $\mathbf{B}$ is symmetric and idempotent. We have the following lemma.

Lemma 4.2.1 *If* (4.4) *holds then*

$$\lim_{n\to\infty} \frac{1}{n} < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T = 0 \quad \text{a.s.}$$

Proof

Recall that $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} = (b_{ij})_{1\leq i,j\leq n}$, and that trace($\mathbf{B}$)$= \sum_{i=1}^{n} b_{ii} = p$. Since $\mathbf{B}$ is symmetric and idempotent, we have that

$$\begin{aligned}\frac{1}{n} < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T &= \frac{1}{n} < \mathbf{B}\epsilon, \epsilon >_T \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} < \epsilon_i, \epsilon_j >_T \\ &= \frac{1}{n}\sum_{i=1}^{n} b_{ii} < \epsilon_i, \epsilon_i >_T + \frac{1}{n}\sum_{1\leq i\neq j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T .\end{aligned}$$

For the first term in the right hand side note that

$$0 \leq \frac{1}{n}\sum_{i=1}^{n} b_{ii} < \epsilon_i, \epsilon_i >_T \leq \max_{i=1}^{n} b_{ii} \; \frac{1}{n}\sum_{i=1}^{n} < \epsilon_i, \epsilon_i >_T .$$

By the strong law of large numbers and (4.4), the last term converges to 0 almost surely. Thus

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} b_{ii} < \epsilon_i, \epsilon_i >_T = 0 \quad \text{a.s.}$$

It remains to show that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{1\leq i\neq j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T = 0 \quad \text{a.s.}$$

Define, for $n \geq 1$, S_n and $\mathcal{F}_n$ by

$$S_n = \sum_{1\leq i\neq j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T, \text{ and } \mathcal{F}_n = \sigma(\epsilon_1, \ldots, \epsilon_n).$$

In Appendix C, we show that $(S_n, \mathcal{F}_n)_{n\geq 1}$ is a martingale, i.e.

$$S_n = \mathbf{E}[S_{n+1}/\mathcal{F}_n] \quad \forall n \geq 1,$$

and that there exists a constant $C > 0$ such that $\mathbf{E}[S_n^2] < C$. This implies in particular, (see for example [5], Theorem 2 p. 242), that the sequence $(S_n)_n$ converges almost surely and thus

$$\lim_{n\to\infty} \frac{1}{n} S_n = \lim_{n\to\infty} \frac{1}{n}\sum_{1\leq i\neq j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T = 0 \quad \text{a.s.}$$

□

We can now state and prove the following result.

Proposition 4.2.2 *With probability one we have that*

$$\lim_{n\to\infty} \frac{SSE}{n} = \sigma^2 \sum_{k\geq 1} \lambda_k^2.$$

Proof

From the definition of the inner product $< .,. >_T$ in $\mathbf{L}_2^n$, we have that

$$\frac{1}{n} < \epsilon, \epsilon >_T = \frac{1}{n}\sum_{i=1}^{n} < \epsilon_i, \epsilon_i >_T .$$

Thus by the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n} < \epsilon, \epsilon >_T = \mathbf{E}[< \epsilon_1, \epsilon_1 >_T].$$

Using the definition of $< .,. >_T$, Fubini's theorem, Mercer's expansion of $\kappa(.,.)$, and the orthonormality of (φ_k) we have that

$$\begin{aligned}
\mathbf{E}[< \epsilon_1, \epsilon_1 >_T] &= \mathbf{E}\left[\int_{[0,1]\times[0,1]} \kappa(s,t)\epsilon_1(s)\epsilon_1(t)ds\, dt\right] \\
&= \int_{[0,1]\times[0,1]} \kappa^2(s,t)ds\, dt \\
&= \sum_{k,l\geq 1} \lambda_k\lambda_l \int_{[0,1]\times[0,1]} \varphi_k(s)\varphi_k(t)\varphi_l(s)\varphi_l(t)ds\, dt \\
&= \sum_{k\geq 1} \lambda_k^2.
\end{aligned}$$

Since $SSE/\sigma^2 = < \epsilon, \epsilon >_T - < \mathbf{B}\epsilon, \mathbf{B}\epsilon >_T$, this and lemma 4.2.1 give the result. □

The following result is a version of Theorem 3.1.1 for the non-Gaussian error case.

Theorem 4.2.1 *For testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, at level α, there exists a test ψ given by:*

$$\psi = \begin{cases} 1 & \text{if } S(\mathbf{Y}) > C(H_0, \alpha), \\ 0 & \text{otherwise} \end{cases}, \tag{4.5}$$

where $S(\mathbf{Y}) = SSR$ if σ is known, and $S(\mathbf{Y}) = n\, SSR/SSE$ otherwise. For n large enough, the constant $C(H_0, \alpha)$ can be chosen (approximately) by

$$\mathbf{P}\left\{\frac{\sum_{k\geq 1}\lambda_k^2\eta_k}{\sum_{k\geq 1}\lambda_k^2} > C(H_0, \alpha)\right\} = \alpha,$$

where the η_k are as in Proposition 4.2.1.

The Constrained Case

In this section we show that we also have a version of Theorem 3.1.2 for the non-Gaussian case.

Theorem 4.2.2 *For testing $H_0 : \mathbf{K}\beta = \mathbf{C}$ versus $H_1 : \mathbf{K}\beta \neq \mathbf{C}$, at level α, there exists a test ψ given by:*

$$\psi = \begin{cases} 1 & \text{if } S(\mathbf{Q}) > C(H_0, \alpha), \\ 0 & \text{otherwise} \end{cases}, \tag{4.6}$$

where $\mathbf{Q} = \|\mathbf{X}(\hat{\beta} - \tilde{\beta})\|_T^2$, and $S(\mathbf{Q}) = \mathbf{Q}$ if σ is known, and $S(\mathbf{Q}) = n\,\mathbf{Q}/SSE$ otherwise. For n large enough, the constant $C(H_0, \alpha)$ can be chosen (approximately) by

$$\mathbf{P}\left\{\frac{\sum_{k\geq 1}\lambda_k^2\eta_k}{\sum_{k\geq 1}\lambda_k^2} > C(H_0,\alpha)\right\} = \alpha,$$

where the η_k are as in Proposition 4.2.1

This theorem can be proved using Proposition 4.2.2 and the following result on the asymptotic distribution of $\mathbf{Q}$.

Proposition 4.2.3 *Let (η_k) be, as in* Proposition 4.2.1, *a sequence of independent random variables with the same distribution χ_p^2. If $\mathbf{K}\beta = \mathbf{C}$ then*

$$\mathbf{Q} \xrightarrow{\mathcal{L}} \sigma^2 \sum_{k\geq 1} \lambda_k^2 \eta_k \qquad \text{as } \to \infty.$$

Proof

Recall that if $\mathbf{K}\beta = \mathbf{C}$ then we have

$$\begin{aligned} \mathbf{X}(\tilde{\beta} - \hat{\beta}) &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\ &= \mathbf{A}_n'\mathbf{A}_n\epsilon, \end{aligned}$$

where $\mathbf{A}_n$, as in the proof of Theorem 4.1.1, denotes $[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}']^{-1/2}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Noting that

$$\mathbf{Q} = < \mathbf{A}_n\epsilon, \mathbf{A}_n\epsilon >_T,$$

the proof follows from Theorem 4.1.1. □

4.3 Generalization to the $\mathbf{L}_2(S, \mathcal{S}, \mu)$ Case

Let S be a compact subset of $\mathbf{R}^p$, and μ be a σ-finite measure on $(S, \mathcal{S})$. Thus $\mathbf{L}_2(S, \mathcal{S}, \mu)$ is a separable Hilbert space. Since all separable infinite dimensional Hilbert spaces are isomorphic, there exists an isomorphism (i.e., a linear surjection) $\mathcal{U} : \mathbf{L}_2(S, \mathcal{S}, \mu) \longrightarrow \mathbf{L}_2$ such that

$$< \mathcal{U}f, \mathcal{U}g >_{\mathbf{L}_2} = < f, g >_{\mathbf{L}_2(S,\mathcal{S},\mu)},$$

where $< ., . >_{\mathcal{H}}$ denotes the inner product in $\mathcal{H}$. This shows that any estimation or testing problem in $\mathbf{L}_2(S, \mathcal{S}, \mu)$ can be reduced to a problem in $\mathbf{L}_2$ where the previous techniques apply. This is an abstract approach to see that there is no loss of generality in considering $\mathbf{L}_2$. A pragmatic approach consists in noting that whenever any property of $\mathbf{L}_2$ has been involved it could have been replaced by its equivalent in $\mathbf{L}_2(S, \mathcal{S}, \mu)$. Results on covariance operators like compacity, or on Gaussian random variables like the Karhünen-Loève expansion, which play an important role in determination of our statistics distributions in $\mathbf{L}_2$, are also valid in $\mathbf{L}_2(S, \mathcal{S}, \mu)$. This more general setting may be important in applications as is suggested by the following example.

Assume that we observe some kind of pollutants which are diffusing through non-homogeneous soil. Let $Y(t)$ denote the observation at location $t \in S$, the amount of pollutants say. The values of Y could increase and decrease in a greater than linear manner (the soil components in some regions are more permeable than others). In permeable regions, changes in value are more extreme than the entire range of values in the unpermeable ones. It is a common practice to transform the data (by taking its logarithms for example) to circumvent such irregularity. Another way to do so could be to transform the parameter space S itself to another parameter space S' so that $(Y(t'), t' \in S')$ is in a certain sense "regular". If S' is $[0,1]$ for example, then one has to replace Lebesgue measure, in the tests we have seen in the previous chapter, by an adequate measure that takes into consideration the initial irregularity of the data. Briefly, although the Y may not have a simple Gaussian structure with respect to Lebesgue's measure on $[0,1]$, it may be quite simple (e.g. Brownian motion) with respect to μ (the "permeability" measure) on $[0,1]$. Thus we also avoid computing eigenvalues for a complex Gaussian process.

These remarks end this chapter. In the next chapter, we will present some examples and some simulation results where we apply the theory presented in the previous chapters.

Chapter 5

APPLICATIONS

In this chapter we present some examples of the application of our techniques. In the first section we discuss how these techniques are a natural generalization of classical multivariate analysis of variance in the case where the error covariance is not completely unknown. In the second section, theoretical and simulated examples show how our new approach provides tests that are more powerful than classical ones. In the third section we apply our techniques to a classical problem from empirical processes theory. The fourth section is devoted to a class of processes that play an important role in the applications of the theory of statistics to geology. The list of examples we present here could naturally not be exhaustive. Yet it would look incomplete if it does not include an example with Brownian motion, the most famous Gaussian process known to all scientists. The last section is devoted to a regression model with a Brownian error.

5.1 Relation to Multivariate Analysis of Variance

5.1.1 The Classical Approach

Recall that our model is given by

$$\mathbf{Y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \sigma\epsilon(t), \quad t \in [0, 1].$$

Suppose that the data is collected only at some points, $t_1, \ldots, t_q$ say. Then it is possible to apply classical multivariate analysis of variance techniques to make inferences about $\boldsymbol{\beta}$. Let first introduce some notation.

For $i = 1, \ldots, n$, and $j = 1, \ldots, q$ put

$$y_{ij} = Y_i(t_j), \quad b_{ij} = \beta_i(t_j), \quad \text{and} \quad e_{ij} = \epsilon_i(t_j).$$

The model is then, in vectorial notation, written as

$$\begin{bmatrix} y_{11}, \dots, y_{1q} \\ \vdots \\ y_{n1}, \dots, y_{nq} \end{bmatrix} = \begin{bmatrix} X_{11}, \dots, X_{1p} \\ \vdots \\ X_{n1}, \dots, X_{np} \end{bmatrix} \begin{bmatrix} b_{11}, \dots, b_{1q} \\ \vdots \\ b_{p1}, \dots, b_{pq} \end{bmatrix} + \sigma \begin{bmatrix} e_{11}, \dots, e_{1q} \\ \vdots \\ e_{n1}, \dots, e_{nq} \end{bmatrix},$$

It is convenient to put for $i = 1, \dots, n$, and $j = 1, \dots, q$

$$\mathbf{y}_i' = [y_{i1}, \dots, y_{iq}],\ \mathbf{X}_i' = [X_{i1}, \dots, X_{ip}],\ \mathbf{b}_j' = [b_{1j}, \dots, b_{pj}],\ \mathbf{e}_i' = [e_{i1}, \dots, e_{iq}],$$

so that

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_n' \end{bmatrix}, \quad \mathbf{b} = [\mathbf{b}_1, \dots, \mathbf{b}_q], \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1' \\ \vdots \\ \mathbf{X}_n' \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1' \\ \vdots \\ \mathbf{e}_n' \end{bmatrix}.$$

In a condensed vectorial notation, the model is written

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sigma\mathbf{e}.$$

The random vectors $\mathbf{e}_1', \dots, \mathbf{e}_n'$ are mean zero independent and identically distributed. Let Σ denote the $q \times q$ covariance matrix of $\mathbf{e}_1$, i.e., $\Sigma = (\sigma_{ij}) = (cov(e_{1i}, e_{1j}))$. The least squares estimator of $\mathbf{b}$ is given by

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Recall that $\hat{\mathbf{b}}$ is BLUE. In the finite dimensional case the least squares estimator has a geometrical interpretation, (see for example [15]). It is the orthogonal projection of the vector of observations on the subspace of $\mathbf{R}^p$ spanned by the columns of the design matrix $\mathbf{X}$.

If the error $\mathbf{e}_1'$ is also Gaussian with a non singular covariance matrix Σ, then the observations are Gaussian and the likelihood function is given by

$$\ell(\mathbf{y}, \mathbf{b}, \Sigma) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-q/2} |\Sigma|^{-1/2} \exp\left[-(\mathbf{y}_i - \mathbf{b}'\mathbf{X}_i)'\Sigma^{-1}(\mathbf{y}_i - \mathbf{b}'\mathbf{X}_i)/2\sigma^2\right].$$

The maximum likelihood method leads to the same estimator $\hat{\mathbf{b}}$. The maximum likelihood estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{nq} \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{b}}'\mathbf{X}_i)'\Sigma^{-1}(\mathbf{y}_i - \hat{\mathbf{b}}'\mathbf{X}_i).$$

Define the $n \times 1$ vector $\rho_i = (0, \dots, 0, 1, 0, \dots, 0)'$ with 1 at the ith place.

$$\hat{\sigma}^2 = \frac{1}{nq} \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i)'\Sigma^{-1}(\mathbf{y}_i - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i)$$

$$\begin{aligned}
&= \frac{1}{nq}\sum_{i=1}^{n}\rho_i'(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y})\Sigma^{-1}(\mathbf{y}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y})'\rho_i \\
&= \frac{1}{nq}\sum_{i=1}^{n}\rho_i'(\mathbf{I}-\mathbf{B})\mathbf{y}\Sigma^{-1}\mathbf{y}'(\mathbf{I}-\mathbf{B})\rho_i \\
&= \text{trace}\left[\mathbf{M}\mathbf{y}\Sigma^{-1}\mathbf{y}'\mathbf{M}\right] \\
&= \text{trace}\left[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}\right],
\end{aligned}$$

As usual, $\mathbf{I}$ is the $n \times n$ identity matrix, $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{M} = \mathbf{I} - \mathbf{B}$.

Various tests of linear hypotheses exist. One of the most commonly used is the *likelihood ratio test.* This is simply the overall maximum of the likelihood divided by the maximum of the likelihood under the null hypothesis. The tests we have developed in chapter 3 are maximum likelihood ratio tests. In this chapter we will throw some light on the relationship of our tests to the maximum likelihood ratio tests in multivariate analysis of variance.

Testing $H_0 : \mathbf{b} = \mathbf{0}$

The likelihood ratio test rejects H_0 for large values of the statistic

$$S(\mathbf{y}) = \frac{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{y}]}{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}]}.$$

Since $\mathbf{I} = \mathbf{M} + \mathbf{B}$, and $\mathbf{M}\ \mathbf{B}{=}\mathbf{B}\ \mathbf{M}{=}0$, we have that $\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{B}\mathbf{y} + \mathbf{y}'\mathbf{M}\mathbf{y}$. Thus the statistic $S(\mathbf{y})$ can be written as

$$S(\mathbf{y}) = 1 + \frac{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{B}\mathbf{y}]}{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}]}.$$

and hence the H_0 is rejected for large values of the ratio

$$\frac{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{B}\mathbf{y}]}{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}]}.$$

By analogy with the classical linear models notation put

$$SSR^* = \text{trace}\left[\Sigma^{-1}\mathbf{y}'\mathbf{B}\mathbf{y}\right], \tag{5.1}$$

$$SSE^* = \text{trace}\left[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}\right]. \tag{5.2}$$

In the next section, we show that a similar test can be obtained using the theory developed in chapters 2 and 3.

5.1.2 A New Approach

Let μ be the atomic measure on $\left([0,1],\mathcal{B}_{[0,1]}\right)$ defined by

$$\forall t \in [0,1] \quad \mu(t) = \begin{cases} 1 & \text{if} \quad t \in \{t_1,\ldots,t_q\} \\ 0 & \text{otherwise} \end{cases}$$

As usual let κ, and T denote the covariance function and the covariance operator of ϵ_1. Note that if two functions f, g are such that

$$\forall i \in \{1,2,\ldots,q\} \quad f(t_i) = g(t_i),$$

they are equal μ-almost everywhere. Thus f and g represent the same element of $\mathbf{L}_2([0,1],\mu)$. In other words $\mathbf{L}_2([0,1],\mu)$ can be identified with $\mathbf{R}^q$, i.e., every function $f \in \mathbf{L}_2([0,1],\mu)$ can can be identified with the column vector $(f(t_1),\ldots,f(t_q))'$. Viewed as an operator from $\mathbf{R}^q$ to $\mathbf{R}^q$, T is the linear transformation associated with the matrix Σ, i.e.,

$$T\,f = \Sigma\,(f(t_1),\ldots,f(t_q))'.$$

The expression of SSE,(see section 3.2.1), is given by

$$\begin{aligned} SSE &= \langle \mathbf{Y}, \mathbf{MY}\rangle_T \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \langle Y_i, m_{ij} Y_j\rangle_T , \end{aligned}$$

where $\mathbf{M} = (m_{ij})$. From the definition of the inner product $< .,. >_T$, and because $\mathbf{M}$ is symmetric we have that

$$\begin{aligned} SSE &= \sum_{i=1}^{n} \left\langle T\,Y_i, \sum_{j=1}^{n} m_{ij} Y_j \right\rangle \\ &= \sum_{i=1}^{n} \langle T\,Y_i, (\mathbf{y}'\mathbf{M})_i\rangle \\ &= \sum_{i=1}^{n} \langle (\Sigma\,\mathbf{y}')_i, (\mathbf{y}'\mathbf{M})_i\rangle , \end{aligned}$$

where $\mathbf{y}'$ is the matrix defined in the previous section, $(\Sigma\,\mathbf{y}')_i$, and $(\mathbf{y}'\mathbf{M})_i$ denote the ith columns of the products. From the definition of the inner product in $\mathbf{R}^n$, we have

$$SSE = \text{trace}\,[\Sigma\,\mathbf{y}'\mathbf{My}]\,.$$

Using the same arguments, we can see that

$$SSR = \text{trace}\,[\Sigma\,\mathbf{y}'\mathbf{By}]\,.$$

We have seen in chapter 3 that the hypothesis $H_0 : \boldsymbol{\beta} = 0$ is rejected for large values of the ratio test SSR/SSE. This ratio is similar to the one of the previous section except that Σ^{-1} is replaced by Σ. The reason for this difference is that the statistic SSR/SSE is related to the Radon-Nikodym derivative of $\mathcal{L}(Y_1)$ with respect to $\mathcal{L}(\sigma\epsilon)$, while SSR^*/SSE^*comes from the density of of $\mathcal{L}(\mathbf{y}_1')$ with respect to Lebesgue measure on $\mathbf{R}^n$.

5.2 Examples on Power

It is intuitively obvious that the more data one has the better are the conclusions. Our techniques, unlike the classical approach, enable us to use all the information we get from our data. The first example is theoretical and shows explicitly, in a particular case, that our tests are asymptotically more powerful that the classical ones. The second example based on simulated data shows that our methods provides tests that are more powerful that the ones from classical multivariate analysis of variance (MANOVA).

5.2.1 A Theoretical Example

Consider the model

$$\mathbf{Y} = \mathbf{1}\beta + \epsilon, \tag{5.3}$$

where $\mathbf{1} = (1, \dots, 1)' \in \mathbf{R}^n$, $\beta = \theta\, a$ where $a \in \mathbf{L}_2$ is a known function, $\theta \in \mathbf{R}$ is the unknown to be estimated, and ϵ is an $n \times 1$ random vector whose components are independent copies of the mean zero, $\mathbf{L}_2$-valued Gaussian random variable ϵ.
Our aim is to test at level α the hypothesis $H_0 : \theta a = 0$, and compare the power of our test to another one which is based on linear projections of the same observations and known to be uniformly most powerful unbiased level α for testing $H_0' : \theta = 0$ versus $H_1' : \theta \neq 0$.

The model (5.3) can be reduced to a real one by linear forms. This corresponds to the practical problem encountered in functional data when one wants to replace the functional data by real valued data.

We consider the following projection on $\mathbf{R}$:

$$\forall i = 1, \dots, n \qquad <\mathbf{Y}_i, b>_T = \theta <a, b>_T + <\epsilon_i, b>_T,$$

where $b \in \mathbf{L}_2$ will be chosen later.

The real random variables $y_i =<\mathbf{Y}_i, b>_T$ are independent identically distributed normal $\mathcal{N}(\mu_b, \sigma_b^2)$, where $\mu_b = \theta <a, b>_T$, and $\sigma_b^2 = \mathbf{E}[<\epsilon_i, b>_T^2]$. It is known, that for testing at level α H_0' versus H_1', there exists, (see, for example [11]), a Uniformly Most Powerful Unbiased level α test, based on $y_1, \dots, y_n$, given by:

$$\psi^* = \begin{cases} 1 & \text{if } \quad \frac{\sqrt{n}|\bar{y}|}{\sigma_b} > z_{\alpha/2} \\ 0 & \quad \text{otherwise} \end{cases},$$

where $\bar{y} =<\bar{\mathbf{Y}}, b>_T$, and $z_{\alpha/2}$ is such that $\mathbf{P}\{\mathcal{N}(0,1) > z_{\alpha/2}\} = \alpha/2$.
Recall that for testing $H_0 : \boldsymbol{\beta} = 0$ against $H_1 : \boldsymbol{\beta} \neq 0$, we have obtained a test ψ

given by

$$\psi = \begin{cases} 1 & \text{if} \quad SSR > c \\ 0 & \quad \text{otherwise} \end{cases},$$

where c is such that $\mathbf{P}\{SSR > c\} = \alpha$.

Comparison of the powers of ψ and ψ^*

We have

$$\begin{aligned} n < \bar{\mathbf{Y}}, b >_T^2 &= n[\theta < a, b >_T + < \bar{\epsilon}, b >_T]^2 \\ &= n[\theta^2 < a, b >_T^2 + 2\theta < a, b >_T < \bar{\epsilon}, b >_T^2], \end{aligned}$$

and

$$\begin{aligned} SRR &= n < \bar{Y}, \bar{Y} >_T \\ &= n[\theta^2 < a, a >_T^2 + 2\theta < \bar{\epsilon}, a >_T + < \bar{\epsilon}, \bar{\epsilon} >_T]^2. \end{aligned}$$

For $\theta \neq 0$, we have

$$\frac{SSR}{n < \bar{\mathbf{Y}}, b >_T^2} = \frac{n[\theta^2 < a, a >_T^2 + 2\theta < \bar{\epsilon}, a >_T + < \bar{\epsilon}, \bar{\epsilon} >_T]^2}{\theta^2 < a, b >_T^2 + 2\theta < a, b >_T < \bar{\epsilon}, b >_T^2},$$

and by the SLLN

$$\frac{SSR}{n < \bar{\mathbf{Y}}, b >_T^2} \longrightarrow \frac{\|a\|_T^2}{< a, b >_T^2} \quad a.s.$$

Therefore for n large and $\theta \neq 0$, we have:

$$\begin{aligned} \mathbf{P}\{SSR > c\} &\simeq \mathbf{P}\left\{ n < \bar{\mathbf{Y}}, b >_T^2 \frac{\|a\|_T^2}{< a, b >_T^2} > c \right\} \\ &= \mathbf{P}\left\{ \frac{n < \bar{\mathbf{Y}}, b >_T^2}{\sigma_b^2} > \frac{< a, b >_T^2}{\|a\|_T^2 \sigma_b^2} c \right\}. \end{aligned}$$

If we can choose $b \in \mathbf{L}_2$ such that

$$c \leq \frac{\|a\|_T^2 \sigma_b^2}{< a, b >_T^2} z_{\alpha/2}, \tag{5.4}$$

then ψ will be at least as powerful as ψ^*.

Choosing $b \in \mathbf{L}_2$

We have that:

$$\sigma_b^2 = \|Tb\|_T^2,$$

thus the condition (5.4) becomes

$$< a^*, b^* >^2 \leq \frac{z_{\alpha/2}^2}{c},$$

where $a^* = a/\|a\|_T^2$, and $b^* = Tb/\|Tb\|_T^2$. It is always possible to choose such a b^*, indeed one can take

$$b^* = c_1\, a^* + c_2\, (a^*)^{\perp},$$

with $c_1^2 \leq z_{\alpha/2}^2\, c^{-1}$, and choose c_2 such that $\|b^*\|_T^2 = 1$.

These comparisons of power functions need to be interpreted carefully. In this example, the optimal test based on projections can be obtained using the exact form of the mean function. In other words, if the function $t \longrightarrow a(t)$ is known, one would project on the function $t \longrightarrow a(t)/\|a\|^2$ rather than the function b^*. But if the statistician chose his projection onto $\mathbf{R}$ poorly so that (5.4) is satisfied, our test will be asymptotically more powerful that the UMPU test based on this projection.

In the following simulated example, the choice of projection points $t_1, \ldots, t_q$ is influenced by our knowledge of the form of the mean function. A full comparison of power should randomize the choice of projection points.

5.2.2 A Simulated Example

As mentioned in the beginning of this chapter, our techniques are a natural generalization of MANOVA. Our aim in this section is to compare the power of our test statistics and the power of the test provided by the MANOVA. To this end we consider the model given by

$$Y(t) = \theta\sqrt{2}\,\sin(\frac{3\pi t}{2}) + \sigma\,\epsilon(t),$$

where ϵ is a Brownian motion, θ and σ are real numbers. Figure 1 shows one realization of this random process.

The mean function

$$\begin{array}{ccc} [0,1] & \longrightarrow & \mathbf{R} \\ t & \longrightarrow & \sqrt{2}\,\sin(\frac{3\pi t}{2}), \end{array}$$

is the first eigenfunction φ_1 of T_ϵ. It has only been chosen for some computational reasons. In vectorial notation the model is written

$$\mathbf{Y} = \mathbf{X}\beta + \sigma\epsilon,$$

where $\mathbf{X} = (1, \ldots, 1)' \in \mathbf{R}^n$, $\beta(t) = \theta\sqrt{2}\sin(\frac{3\pi t}{2})$ and ϵ is an element of $\mathbf{L}_2^n$ whose components are independent and identically distributed like ϵ.

We simulate a random sample to test the null hypothesis $H_0 : \theta = 0$ against the alternatives $H_1 : \theta = i/10$, for $i = 1, \ldots, 5$.

Recall that our test rejects the null hypothesis for large values of the statistic

$$Q(Y) = SSR/SSE.$$

With the notation of § 5.1.1, the MANOVA test rejects the null hypothesis for large values of the statistic

$$S(\mathbf{y}) = \frac{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{y}]}{\text{trace}\,[\Sigma^{-1}\mathbf{y}'\mathbf{M}\mathbf{y}]},$$

where

$$\mathbf{y} = (\mathbf{y_1}', \ldots, \mathbf{y_n}')', \quad \text{and} \quad \mathbf{y_i}' = (Y_i(t_1), \ldots, Y_i(t_q)), \quad i = 1, \ldots, n.$$

The size of our simulated random sample is $n = 30$. For the MANOVA test, we considered the cases $q = 4$ and $q = 10$. Figures 2, 3, and 4 show the power curves of our test and the MANOVA test at different levels of significance.

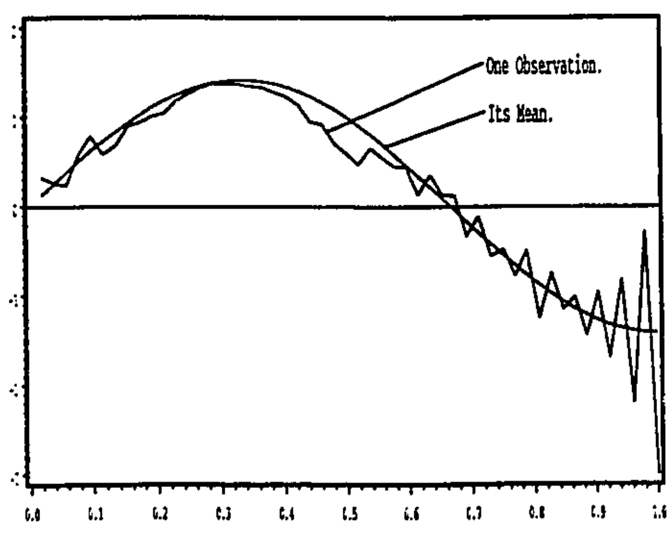


Figure 1.

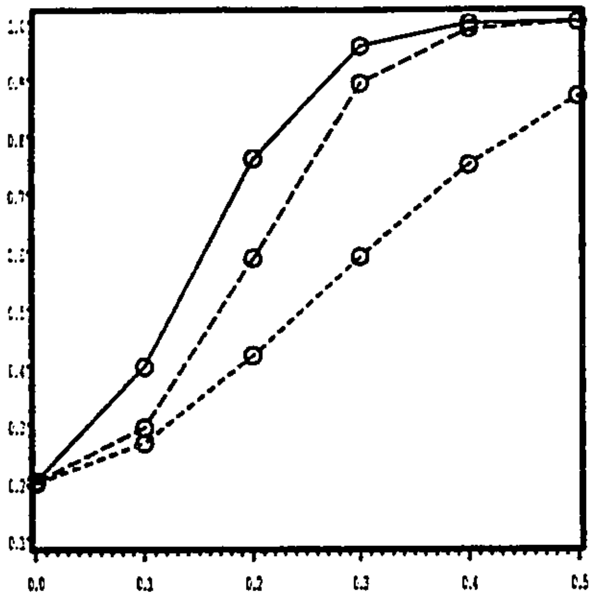


Figure 2: Powers of the 20% tests.

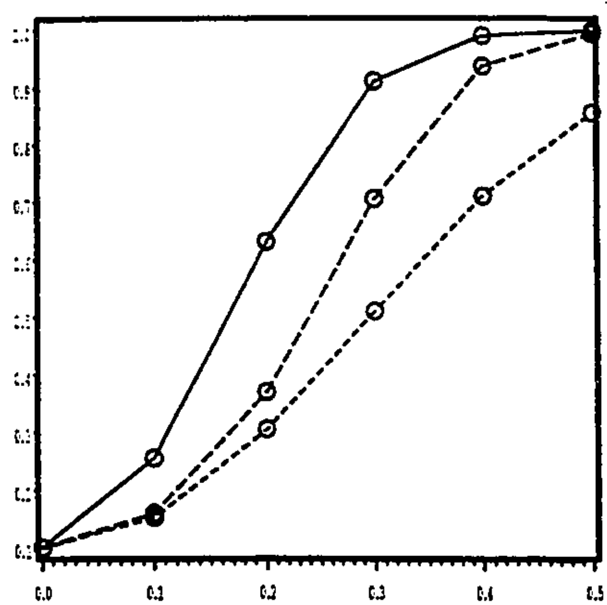


Figure 3: Powers of the 10% tests.

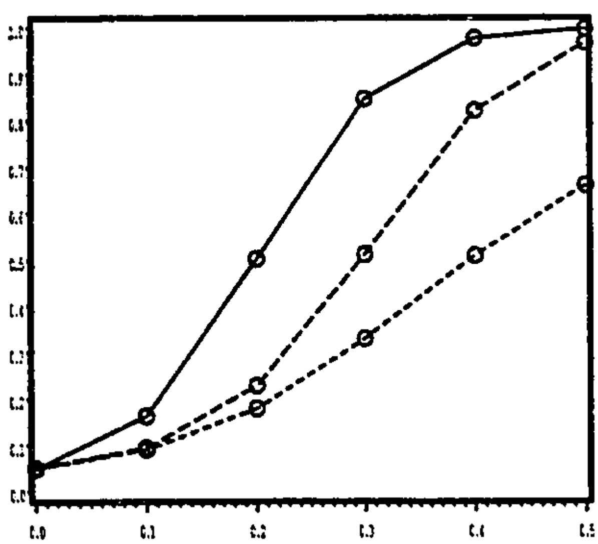


Figure 4: Powers of the 5% tests.

Solid Line: Power of Our test.

Dashed Line: Power of the MANOVA test, (g=10).

Dotted Line: Power of the MANOVA test, (g=4).

5.3 Testing for Equality of Two Distributions

The model is given by

$$Y_{ij}(t) = \beta_i(t) + \epsilon_{ij}(t) \quad j = 1, \ldots, n_i, \quad i = 1, 2, \text{ and } t \in [0,1],$$

where the errors ϵ_{ij} are mean zero and independent. Furthermore we assume that for each $i = 1, 2$, the random variables $\epsilon_{i1}, \ldots, \epsilon_{in_i}$ are identically distributed. Let T denote the covariance operator associated with $\epsilon_{11} - \epsilon_{21}$, and $\|.\|_T$ the semi-norm it induces on $\mathbf{L}_2$.

In vectorial notation, the model is written

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \text{ with } \mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{bmatrix} \text{ and } \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \text{ and } \mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{bmatrix} \text{ with } \mathbf{1}_{n_i} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbf{R}^{n_i}.$$

Our aim is to test if the two groups have the same mean function, i.e., we test the hypothesis $H_0 : \mathbf{K}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{K} = [1 \quad -1]$.

For testing $H_0 : \mathbf{K}\boldsymbol{\beta} = 0$ versus $H_1 : \mathbf{K}\boldsymbol{\beta} \neq 0$, by Theorem 3.1.2, there exists a test ψ given by

$$\psi = \begin{cases} 1 & \text{if } \mathbf{Q} > C_\alpha \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{Q} = \|\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\|_T^2$, and C_α is such that

$$\mathbf{P}\{\mathbf{Q} > C_\alpha, \mathbf{K}\boldsymbol{\beta} = \mathbf{0}\} = \alpha.$$

Recall that, in this case,

$$\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}' \left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1} \mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Note that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix} \text{ and } \mathbf{K}\mathbf{K}' = 2.$$

For simplicity of calculations, assume that $n_1 = n_2 = n$. Thus

$$\begin{aligned}\mathbf{X}(\hat{\beta} - \tilde{\beta}) &= \frac{1}{n}\mathbf{X}\mathbf{K}'[\mathbf{K}\mathbf{K}']^{-1}\mathbf{K}\mathbf{X}'\epsilon \\ &= \frac{1}{2n}\mathbf{X}\mathbf{K}'\mathbf{K}\mathbf{X}'\mathbf{Y} \\ &= \frac{1}{2}\begin{bmatrix} (Y_{1.} - Y_{2.})\mathbf{1}_n \\ (Y_{1.} - Y_{2.})\mathbf{1}_n \end{bmatrix},\end{aligned}$$

where

$$Y_{i.} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}.$$

The statistic $\mathbf{Q}$ is thus given by

$$\mathbf{Q} = \|\sqrt{n}\,(Y_{1.} - Y_{2.})\|_T^2.$$

To determine critical values of the test ψ, we need the probability distribution of the statistic $\mathbf{Q}$ under the null hypothesis. The central limit theorem provides an asymptotic behaviour, in distribution, of $\mathbf{Q}$. More precisely we have the following result, which is a particular case of the general central limit theorem in Hilbert space.

Lemma 5.3.1 *Suppose that the* $\mathbf{L}_2$*-valued random variable* ϵ_{11} *and* ϵ_{21} *are such that*

$$\mathbf{E}\|\epsilon_{11}\|_2^2 < \infty, \quad \text{and} \quad \mathbf{E}\|\epsilon_{21}\|_2^2 < \infty.$$

Then there exists an $\mathbf{L}_2$*- valued Gaussian mean zero random variable* G *with the same covariance structure as* $\epsilon_{11} - \epsilon_{21}$ *and such that*

$$\sqrt{n}\,(\epsilon_{1.} - \epsilon_{2.}) \xrightarrow{\mathcal{L}} G, \quad \text{as } n \to \infty.$$

Remark 5.3.1 *1. Note that under the null hypothesis,* $\mathbf{Q} = \|\sqrt{n}\,(\epsilon_{1.} - \epsilon_{2.})\|_T^2$.

2. Letting $\epsilon_i = \epsilon_{1i} - \epsilon_{2i}$ *for* $i = 1, \ldots, n$, *it can easily be seen that the above lemma is a particular case of* Proposition 4.1.1.

There are numerous results concerning the problem of testing for equality of two distributions. Two test statistics have become famous for their rich history, namely the Kolmogorov-Smirnov statistic K, and the Cramér-von Mises statistic W. They are given by

$$\begin{aligned}K &= \sup_x |\bar{F}_1(x) - \bar{F}_2(x)|, \quad \text{and} \\ W &= n\int_{-\infty}^{\infty} |\bar{F}_1(x) - \bar{F}_2(x)|^2\, dF(x),\end{aligned}$$

where

$$\bar{F}_1(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{(-\infty,x]}(X_{ij}), \quad i = 1, 2.$$

In the remainder of this section we show how our techniques can be applied to testing for the equality of two distributions. We also present histograms of our statistic $\mathbf{Q}$ and those of Cramér-von Mises and Kolmogorov-Smirnov Statistics.

An Application

Let F denote a continuous cumulative distribution function. Let $X_{i1},\dots,X_{in}$, for $i=1,2$, be independent and identically distributed real random variables. Let F_i denote their common cumulative distribution function. Assume that $\{X_{11},\dots,X_{1n}\}$ and $\{X_{21},\dots,X_{2n}\}$ are independent. We are interested in testing if $F_1=F_2=F$.

For $i=1,2$ and $j=1,\dots,n$, define

$$Y_{ij}(t)=\mathbf{1}_{[0,t]}(F(X_{ij})),\quad t\in[0,1].$$

To fit the model above let $\beta_i=\mathbf{E}[Y_{i1}]$ and $\epsilon_{ij}=Y_{ij}-\beta_i$. Therefore $F_1=F_2$ if and only if $\beta_1=\beta_2$. Recall that

$$\mathbf{Q}=\|\sqrt{n}(Y_{1.}-Y_{2.})\|_T^2,$$

where T is the covariance operator associated with $\epsilon_{11}-\epsilon_{21}$.

To identify the limit, in distribution, of $\mathbf{Q}$ we need to know the covariance function of $\epsilon_{11}-\epsilon_{21}$. We have that

$$\begin{aligned}\mathbf{E}\left[(\epsilon_{11}(s)-\epsilon_{21}(s))(\epsilon_{11}(t)-\epsilon_{21}(t))\right] &= \mathbf{E}\left[\epsilon_{11}(s)\epsilon_{11}(t)\right]+\mathbf{E}\left[\epsilon_{21}(s)\epsilon_{21}(t)\right]\\ &= \mathbf{E}\left[\mathbf{1}_{[0,s\wedge t]}(F(X_{11}))\right]-\beta_1(s)\beta_1(t)\\ &\quad+\mathbf{E}\left[\mathbf{1}_{[0,s\wedge t]}(F(X_{21}))\right]-\beta_2(s)\beta_2(t).\end{aligned}$$

Therefore, by Lemma 1.1, we have the following result:

Under the hypothesis $H_0: F_1=F_2=F$, we have

$$\mathbf{Q}\xrightarrow{\mathcal{L}}2\|U\|_T^2,\quad \text{as } n\to\infty, \tag{5.5}$$

where U denotes the Brownian bridge.

Indeed, under H_0 we have

$$\mathbf{E}\left[(\epsilon_{11}(s)-\epsilon_{21}(s))(\epsilon_{11}(t)-\epsilon_{21}(t))\right]=2(s\wedge t-st).$$

Therefore by Lemma 5.3.1,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\epsilon_{1j}-\epsilon_{2j})\xrightarrow{\mathcal{L}}\sqrt{2}\,U,\quad \text{as } n\to\infty.$$

By (1.4) the semi-norm $\|.\|_T$ is continuous with respect to the usual norm $\|.\|_2$. Thus (5.5) holds.

A Simulated Example

As an illustration we apply these techniques to simulated data. Let F be the distribution of a random variable X which is uniformly distributed on $[0,1]$, and $\sqrt{F}$ the distribution of X^2. For different values of n, we simulated two independent sets of data $y_{11},\ldots,y_{1n}$ and $y_{21},\ldots,y_{2n}$, under the null hypothesis $H_0 : F_1 = F_2 = F$ and under the alternative $H_1 : F_1 = F$ and $F_2 = \sqrt{F}$. In both cases we computed the statistic $\mathbf{Q}$. To have an idea about the distribution of $\mathbf{Q}$ under the null hypothesis, and also under the alternative, we repeated this operation 1000 times. The histograms are shown in Appendix A.

From the theoretical point of view our statistic $\mathbf{Q}$ looks like the Cramér-von Mises statistic W: Both are integrals of the observations. It is therefore not surprising to see that in the present simulated case they lead to the same results. They probably do so in general. Comparing $\mathbf{Q}$ to the Kolmogorov-Smirnov test K, the histograms show that $\mathbf{Q}$ is doing slightly better than K for small sample sizes, but this could only be due the distribution we chose for our simulations. As we mentioned before, $\mathbf{Q}$ and W might be comparable. However, it is hard to predict any comparison between $\mathbf{Q}$ and K. Even for K and W, there is no evidence which one is better despite their rich history.

5.4 Geological Example

Continuous models in geology describe properties, like hydraulic conductivity, that vary continuously. In such models one has to specify statistical properties like the mean, variability about the mean and spatial correlation between neighboring values.

Suppose that one is interested in hydraulic conductivity rate in a given region. As a spatial function, the rate is continuous and must be spatially correlated over short distances. Before describing how our techniques can be applied, here is a very brief outline of the method known in geostatistics as the *Kriging* method. It is a " minimum -mean-squared-error" method of spatial prediction. Matheron [12] named this method of optimal spatial linear prediction after D.G. Krige [9] who, in the 1950s, developed empirical methods in mining. However, the formulation of the *Kriging method* did not come from Krige's work. See for example [13], [2] and the references therein for the extent of of the early work of Krige.

Let $Y(t)$ denote the rate at location t. The simplest form of Kriging is punctual Kriging which consists in estimating $Y(t)$ by $\sum_i \alpha_i Y(t_i)$, where the α_i are weights and the $Y(t_i)$ are measurements at locations t_i. There are an infinity of possible combinations of weights that could be chosen. There is, however, only one combination that will give a minimum combination error. It is this combination of weights that Kriging attempts to find. In other words, the problem reduces to minimizing, with respect to the α_i, the error

$$\epsilon(t) = Y(t) - \sum_i \alpha_i Y(t_i).$$

Let $Y_k(t)$ denote the estimate of $Y(t)$ obtained by Kriging, thus we have

$$Y_k(t) = Y(t) - \epsilon_k(t),$$

where ϵ_k is the error due to Kriging. In this form, the problem looks like the one we are considering in this work. Furthermore our estimate for Y, (m in our setting), is also based on minimizing the error. Suppose that estimates of Y, i.e, Y_k, are available from n independent regions presenting some given p characteristics. Assume that we want to know the effects of these characteristics, denoted by $\beta_1, \ldots, \beta_p$, on the value of Y. The β_i are spatial unknown continuous functions. Let $x_{ij} = 1$ or 0, for $i = 1, \ldots, n$ and $j = 1, \ldots, p$, according to the presence or absence of the jth characteristic at the ith region. Now let Y_i and ϵ_i denote respectively Y_k and ϵ_k which are obtained by punctual Kriging from the ith region. Therefore,

$$Y_i(t) = \sum_{j=1}^{p} x_{ij}\beta_j(t) + \epsilon_i(t), \quad i = 1, \ldots, n.$$

The distribution of the error can be estimated using time series analysis techniques, and its covariance function is usually estimated from covariograms and semivariaograms. Assume that we only know, from time series analysis methods, that the error is $\sigma\epsilon$ for some unknown scale parameter σ, but the form of the covariance function of ϵ is known. This is the general case to which our techniques apply.

The error covariance structures most commonly assumed in geostatistics are those of Gaussian and second order stationary processes. In this section we consider models with Gaussian second order stationary error.

We first recall some results from Fourier analysis. We use these results to determine eigenvalues and eigenvectors of an error covariance operator when the error is a Gaussian second order stationary process. In the second subsection we show how our techniques apply in this case and then apply them to a simulated example. Our simulation is based on a program developed by M.J.L. Robin from the Department of Geology, University of Ottawa.

5.4.1 Mathematical Background

It is well known that an even function f on $[-1, 1]$, i. e., $f(x) = f(-x)$ for every $x \in [0, 1]$, has the Fourier expansion

$$f(x) = \frac{a_0}{2} + \sum_{n\geq 1} a_n \cos(n\pi x), \qquad a_n = 2\int_0^1 f(x)\cos(n\pi x)\,dx. \tag{5.6}$$

Suppose that f is square-integrable and define the operator $T_f : \mathbf{L}_2 \longrightarrow \mathbf{L}_2$ by

$$\forall \varphi \in \mathbf{L}_2, \qquad T_f\varphi(x) = \int_0^1 f(x-y)\varphi(y)\,dy. \tag{5.7}$$

Define, for $k \geq 0$,

$$\begin{aligned} \varphi_0 &\equiv 1 & (5.8)\\ \lambda_k &= \frac{a_k}{2} & (5.9)\\ \varphi_k(x) &= \sqrt{2}\cos(k\pi x), \quad \forall x \in [0,1]. & (5.10) \end{aligned}$$

Using the definition of T_f, equation (5.7), and the Fourier expansion of f, equations (5.6), it can easily be seen that we have the following result

Lemma 5.4.1 *For all $k \geq 0$ we have*

$$T_f\varphi_k = \lambda_k\varphi_k.$$

Furthermore, the φ_k are orthonormal.

In the next section we will see that there is a wide class of Gaussian processes whose covariance operators have their eigenvalues and eigenfunctions given by the expressions above.

5.4.2 Second Order Stationary Processes

Definition 5.4.1 *Let $\left(\epsilon(s), s \in S \subset \mathbf{R}\right)$ be a random process. It is said to be a mean zero second order stationary process if*

1. $\mathbf{E}[\epsilon(s)] = 0 \quad \forall s \in S$, *and*

2. *For any $s, t \in S$, $\mathbf{E}[\epsilon(t)]^2$ exists and*

$$cov\left(\epsilon(s), \epsilon(t)\right) = \kappa(t-s),$$

for some covariance function $\kappa(.)$ that depends only on $(t-s)$.

Let $\left(\epsilon(s), s \in [0,1]\right)$ be a mean zero second order stationary process. For every $t \in [0,1]$, we have $cov\left(\epsilon(0), \epsilon(t)\right) = \kappa(t)$, and $cov\left(\epsilon(t), \epsilon(0)\right) = \kappa(-t)$. Therefore, $\kappa(.)$ is an *even* and square-integrable function on $[-1,1]$. Its Fourier expansion is given by

$$\kappa(x) = \frac{a_0}{2} + \sum_{n\geq 1} a_n\cos(n\pi x), \qquad a_n = 2\int_0^1 \kappa(x)\cos(n\pi x)\,dx.$$

Let T_ϵ denote the *covariance* operator associated with ϵ, i.e.,

$$\forall \varphi \in \mathbf{L}_2, \qquad T_\epsilon\varphi(x) = \int_0^1 \kappa(x-y)\varphi(y)\,dy. \tag{5.11}$$

The operator T_ϵ is linear, self adjoint and compact. The pairs (λ_k, φ_k) given by (5.8), (5.9), and (5.10) are eigenvalues and eigenvectors of T_ϵ . Since we have

$$\kappa(x) = \frac{a_0}{2} + \sum_{n\geq 1} a_n \cos(n\pi x), \quad \forall x \in [0,1],$$

there exist a sequence (X_k) of independent and identically distributed normal $\mathcal{N}(0,1)$ random variables such that

$$\epsilon(t) = \sum_{k\geq 0} \sqrt{\lambda_k}\, X_k \cos(k\pi t) \quad \forall t \in [0,1].$$

In other words the φ_k are the only eigenvectors of T_ϵ since they span the reproducing kernel Hilbert space of ϵ (see for example [6, Proposition 1.2.3 p. 22]).

Example

Measurements of hydraulic conductivity of the soil are taken at two locations, location 1 and location 2 say. The model is given by

$$Y_{ij}(t) = \beta_j(t) + \epsilon_{ij}(t) \quad j = 1,\ldots,n, \quad i = 1,2, \text{ and } t \in [0,1].$$

In vectorial notation, it is written

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \sigma \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix},$$

where,

$$\mathbf{1}_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbf{R}^n, \; \mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in} \end{bmatrix}, \text{ and } \epsilon_i = \begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in} \end{bmatrix} \quad \text{for } i = 1,2.$$

Our aim is to test if the means at location 1 and location 2, are the same, i.e., $\beta_1 = \beta_2$. The null hypothesis is $H_0 : \mathbf{K}\boldsymbol{\beta} = 0$, with

$$\mathbf{K} = [1 \quad -1], \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Assume that $\epsilon_{i1}, \ldots, \epsilon_{in}$, $i = 1,2$, are independent and identically distributed. Furthermore we suppose that ϵ_{11} is a second-order stationary mean zero Gaussian process whose covariance function is given by

$$\kappa(s,t) = \exp(-|s-t|) \quad \forall s,t \in [0,1].$$

Recall that the statistic $\mathbf{Q}$ is given by

$$\mathbf{Q} = \|\mathbf{X}(\hat{\beta} - \tilde{\beta}\|_T^2.$$

In the present case it reduces to

$$\begin{aligned}\mathbf{Q} &= \|\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\left[\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}'\right]^{-1}\mathbf{K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\|_T^2 \\ &= n\|\frac{1}{n}\sum_{j=1}^{n}(Y_{1j} - Y_{2j})\|_T^2 \\ &= \|\sqrt{n}(Y_{1.} - Y_{2.})\|_T^2.\end{aligned}$$

Under the null hypothesis, we have that

$$\mathbf{Q} = n\|\epsilon_{1.} - \epsilon_{2.}\|_T^2,$$

thus

$$\mathcal{L}(\mathbf{Q}) = \mathcal{L}(2\sum_{k\geq 0}\lambda_k^2\xi_k),$$

where the ξ_k are independent random variables which are identically distributed as a χ_1^2 with one degree of freedom, the λ_k, the eigenvalues, are given by

$$\lambda_k = \frac{1 - (-1)^k \exp(-1)}{1 + (\pi k)^2}.$$

For testing $H_0 : \mathbf{K}\boldsymbol{\beta} = 0$ versus $H_1 : \mathbf{K}\boldsymbol{\beta} \neq 0$, by Theorem 4.2.2, there exists a test ψ given by

$$\psi = \begin{cases} 1 & \text{if } \frac{\mathbf{Q}}{SSE} > C_\alpha \\ 0 & \text{otherwise.}\end{cases}$$

Recall that by Proposition 3.1.1,

$$SSE = \sum_{k\geq 0}\lambda_k^2\eta_k \quad \text{a.s.},$$

where the η_k are independent random variables which are identically distributed as a $\chi^2_{r(M)}$ with $r(\mathbf{M})$ degree of freedom. In this case, the rank of the matrix $\mathbf{M}$ is $n-2$.

We simulate the cases where $\beta_1 = 0$, and $\beta_2 \equiv \gamma$ for $\gamma = 0, 1/4, 1/2, 3/4$, and 1.

In Appendix A, we give tables of critical values of ψ when $n = 50$. We also draw the power of the test as a function of γ, and the histograms of our statistics values. These values have been obtained by a simulation program given in Appendix B.

5.5 An Example of Regression Models

Other examples of Gaussian processes which are frequently used in other domains like chemistry and physics are Brownian motion and Brownian bridge. In physics, for instance, the movements of microscopic particles suspended in a liquid or gas are well described by a Brownian motion. Consider a particle which is diffusing through a liquid, and suppose that we observe its trajectory. Assume that we are interested in the effect of the particle weight on its trajectory. We suppose that the trajectory is modeled by

$$Y(t) = \beta_0(t) + x\beta_1(t) + \sigma\epsilon(t),$$

where β_0 is the "mean trajectory", x is the weight of the particle, $x\beta_1$ is the effect due to the weight, and $\sigma\epsilon$ represents the fluctuations due to the collisions between the observed particle and other particles in the liquid. We want to test the hypothesis $H_0 : \beta_1 = 0$. For this we observe the trajectories of n particles with different weights. The model is given by

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = [\mathbf{1}_n \;\; \mathbf{x}_n] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \sigma \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix},$$

where $\mathbf{1}_n = (1, \ldots, 1)'$, and $\mathbf{x}_n = (x_1, \ldots, x_n)' \in \mathbf{R}^n$. For simplicity of computations, suppose that $\mathbf{1}_n'\mathbf{x}_n = \mathbf{x}_n'\mathbf{1}_n = 0$, i.e., $\mathbf{1}_n$ and $\mathbf{x}_n$ are orthogonal in $\mathbf{R}^n$. Let

$$\mathbf{X} = [\mathbf{1}_n \;\; \mathbf{x}_n], \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \text{and} \quad \mathbf{K} = [0 \;\; 1].$$

The hypothesis to test becomes $H_0 : \mathbf{K}\boldsymbol{\beta} = 0$. By Theorem 3.1.2, we reject H_0 for large values of $\mathbf{Q}$/SSE. As before, $\mathbf{Q}$ and SSE are given by

$$\begin{aligned} \mathbf{Q} &= \|\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})\|_T^2 \\ SSE &= \sigma^2\|(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\epsilon\|_T^2. \end{aligned}$$

Recall that by Proposition 3.1.2, $\mathbf{Q}$ and SSE are independent. Since, under H_0, the ratio $\mathbf{Q}$/SSE does not depend on σ, we suppose that $\sigma = 1$. After some algebra, one can see that

$$\begin{aligned} \mathbf{Q} &= \sum_{i=1}^n \left\| \sum_{j=1}^n \frac{x_i x_j}{\|\mathbf{x}_n\|^2} Y_j \right\|_T^2 \\ &= \sum_{i=1}^n \frac{|x_i|^2}{\|\mathbf{x}_n\|^2} \left\| \sum_{j=1}^n \frac{x_j}{\|\mathbf{x}_n\|} Y_j \right\|_T^2 \\ &= \left\| \sum_{j=1}^n \frac{x_j}{\|\mathbf{x}_n\|} Y_j \right\|_T^2. \end{aligned}$$

where $\|\mathbf{x}_n\|$ is the Euclidian norm of $\mathbf{x}_n$. Therefore there is no loss of generality in assuming that $\|\mathbf{x}_n\| = 1$. Assume that $\|\mathbf{x}_n\| = 1$, we have that

$$\begin{aligned}
\sum_{j=1}^{n} x_j Y_j &= \mathbf{x}'_n \mathbf{Y} \\
&= \mathbf{x}'_n \mathbf{1}_n \beta_0 + \mathbf{x}'_n \mathbf{x}_n \beta_1 + \mathbf{x}'_n \epsilon \\
&= \beta_1 + \mathbf{x}'_n \epsilon,
\end{aligned}$$

because $\mathbf{x}'_n \mathbf{1}_n = 0$, and $\|\mathbf{x}_n\| = 1$. Noting that $\mathcal{L}(\mathbf{x}'_n \epsilon) = \mathcal{L}(\epsilon_1)$, we have under H_0 that

$$\mathcal{L}(\mathbf{Q}) = \mathcal{L}(\sum_{k\geq 1} \lambda_k^2 X_k^2),$$

where the X_k are independent and identically distributed $\mathcal{N}(0,1)$ random variables, and for every $k \geq 1$

$$\lambda_k = \frac{1}{((k+1/2)\pi)^2}.$$

Under H_1 we have that

$$\mathcal{L}(\mathbf{Q}) = \mathcal{L}(\sum_{k\geq 1} \lambda_k^2 \left(X_k + \beta_{1(k)}\right)^2),$$

where

$$\beta_{1(k)} = \sqrt{2} \int_0^1 \beta_1(s) \sin(\frac{2k+1}{2}\pi s), \quad \forall k \geq 1.$$

Since $\mathcal{L}(\mathbf{Q})$ does not depend on β_0, we take $\beta_0 = 0$ in our simulated example. More precisely, we simulate for $n = 6$

$$\mathbf{Y} = \mathbf{x}_6 \beta_1 + \epsilon,$$

with

$$\mathbf{x}_6 = \frac{1}{\sqrt{28}}(1, 2, 3, -1, -2, -3)', \quad \text{and} \quad \beta_1 \equiv 3.$$

The histograms and a table with critical values are given in Appendix A.

Chapter 6

RANDOM AND MIXED EFFECTS MODEL

The models we considered in the previous chapters are known in classical theory of linear models as *fixed effect models.* The observations in those models are composed of two parts, a non-random component called the mean and a random component which is the error. In various situations, the outcomes depend on several parameters which are themselves random. In such cases it is more appropriate to look at the observations as sums of two or more random variables. In this chapter we consider models dealing with these situations. The theory here is not as general as is the case in chapters two and three, and the results are similar to those known in the real case. The techniques we present here are essentially the same as in the previous chapters. Our purpose is only to show how these techniques extend to the functional random and mixed effects models.

6.1 The One-Way Layout

Suppose that our data are measurements of the propagation speed of a radioactive cloud over different regions. The propagation speed can depend on factors like the geography of the region, the weather and eventually on the nature of radiations. In this section we will consider the case of one factor only. Thus suppose that a single cloud is observed n times over I independent regions. Let Y_{ij} denote the jth observation at the ith region. It can be modeled by

$$Y_{ij} = m_i + e_{ij}.$$

The component e_{ij} can be thought of as the measurement error, while m_i is the effect due to the region. The latter is the sum of the "true" speed and a random term, i.e., $m_i = \mu + a_i$. The random terms a_i and e_{ij} are assumed to be mean zero independent processes.

The model is given by

$$Y_{ij}(t) = \mu(t) + a_i(t) + e_{ij}(t), \quad i = 1, \cdots, I, \; j = 1, \cdots, n, \quad t \in [0,1].$$

Least squares estimators of μ and the a_i can be obtained by minimizing the sum of the squared norms of errors due to measurements. Let ESS denote this sum, i.e.,

$$\text{ESS} = \sum_{ij} \|Y_{ij} - \mu - a_i\|_2^2.$$

Let $(\varphi_k)_{k\geq 1}$ be an orthonormal basis of $\mathbf{L}_2$, we have that

$$Y_{ij} = \sum_{k\geq 1} Y_{ijk}\varphi_k, \quad \mu = \sum_{k\geq 1} \mu_k\varphi_k, \quad \text{and} \quad a_i = \sum_{k\geq 1} Y_{ik}\varphi_k,$$

where

$$Y_{ijk} = \int_0^1 Y_{ij}(s)\varphi_k(s)\,ds, \; \mu_k = \int_0^1 \mu(s)\varphi_k(s)\,ds, \text{ and } a_{ik} = \int_0^1 a_i(s)\varphi_k(s)\,ds.$$

Hence

$$\text{ESS} = \sum_{k\geq 1}\sum_{ij}\left[Y_{ijk} - \mu_k - a_{ik}\right]^2,$$

and therefore minimizing ESS is equivalent to minimizing each term of the series. Before stating the normal equations, recall the very convenient *dot notation* which we will use through this chapter.

$$\xi_{ij.} = \frac{1}{\nu}\sum_{k\in\{k_1,\ldots,k_\nu\}} \xi_{ijk}, \quad \xi_{i..} = \frac{1}{\nu'}\sum_{j\in\{j_1,\ldots,j_{\nu'}\}} \xi_{ij.}, \text{ etc...}$$

The normal equations are

$$\begin{aligned} \hat{\mu}_k &= Y_{..k} - \hat{a}_{.k}, \\ \hat{a}_{ik} &= Y_{i.k} - \hat{\mu}_k. \end{aligned}$$

Since they hold for every $k \geq 1$ they are equivalent to

$$\begin{aligned} \hat{\mu} &= Y_{..} - \hat{a}_., \\ \hat{a}_i &= Y_{i.} - \hat{\mu}. \end{aligned}$$

These are I linearly independent equations with $I+1$ unknowns. A solution is given by

$$\hat{\mu} = Y_{..}, \quad \text{and} \quad \hat{a}_i = Y_{i.} - Y_{..}.$$

For hypotheses testing, we need to make some more assumptions about the distributions of the error and the factor effect. We suppose that

1. $(a_i)_{i\leq I}$ are independent and identically distributed as $\sigma_1\epsilon_1$,

2. $(e_{ij})_{i\leq I, j\leq n}$ are independent and identically distributed as $\sigma_2\epsilon_2$,

3. σ_1, σ_2 are unknown and ϵ_1, ϵ_2 are independent, mean zero Gaussian processes.

With these assumptions, for $i = 1, 2$, let $(\lambda_{ik}, \varphi_{ik})_k$ denote the sequences of pairs of eigenvalue-eigenvector associated with the covariance operator T_{ϵ_i}. Without any loss of generality, for $i = 1, 2$, we assume that $(\varphi_{ik})_k$ is an orthonormal sequence. By the Karhünen-Loève expansion, see Proposition 1.1.4, for all i and j we have that

$$\begin{aligned} a_i &= \sigma_1 \sum_{k\geq 1} \sqrt{\lambda_{1k}}\, X_{ik}\varphi_{1k} \quad \text{a.s. and,} && (6.1) \\ e_{ij} &= \sigma_2 \sum_{k\geq 1} \sqrt{\lambda_{2k}}\, X_{ijk}\varphi_{2k} \quad \text{a.s.} && (6.2) \end{aligned}$$

where $(X_{ik})_{ik}$ and $(X_{ijk})_{ijk}$ are independent and identically distributed standard normal.

By analogy with the real case we define SS_a, the sum of squares due to the random factor, by

$$\mathrm{SS}_a = \sum_{ij} \|\hat{a}_i\|_2^2.$$

Using the expression of $\hat{a}_i$, (6.1) and (6.2), straightforward calculations show that we have, almost surely,

$$\begin{aligned} \mathrm{SS}_a &= n \sum_{i=1}^{I} \|(a_i - a_.) + (e_{i.} - e_{..})\|_2^2 \\ &= n\Big[\sigma_1^2 \sum_{k\geq 1} \lambda_{1k}\eta_{1k} + \sigma_2^2 \sum_{k\geq 1} \lambda_{2k}\eta'_{2k} + 2R(a,e)\Big], \end{aligned}$$

where (η_{1k}) are independent random variables which are identically distributed as χ^2_{I-1} , (η'_{1k}) are independent and identically distributed as χ^2_{I-1}, and

$$R(a,e) = \sum_i < a_i - a_., e_{i.} - e_{..} > .$$

Also let SS_e denote the sum of squares of the errors due to fitting the model, defined by

$$\mathrm{SS}_e = \sum_{ij} \|Y_{ij} - \hat{Y}_{ij}\|_2^2.$$

Similarly, it can easily be seen that we have, almost surely,

$$\begin{aligned} \mathrm{SS}_e &= \sum_{ij} \|e_{ij} - e_{i.}\|_2^2 \\ &= \sigma_2^2 \sum_{k\geq 1} \lambda_{2k} \sum_{ij} \Big(X_{ijk} - X_{i.k} \Big)^2 \\ &= \sigma_2^2 \sum_{k\geq 1} \lambda_{2k} \eta_{2k}, \end{aligned}$$

where (η_{2k}) are independent and identically distributed as $\chi^2_{I(n-1)}$. Furthermore we have the following result

Proposition 6.1.1 *The statistics SS_a and SS_e are stochastically independent.*

Proof

Recall that the a_i and e_{ij} are independent. Therefore the sequences $(\eta_{1k})_k$ and $(\eta_{2k})_k$ are independent, and it remains only to show that $(\eta'_{2k})_k$ and $(\eta_{2k})_k$ are independent. Since the X_{ijk} are independent, $\eta'_{2k'}$ and η_{2k} are independent for $k \neq k'$. It remains only to show that $(\eta'_{2k})_k$ and $(\eta_{2k})_k$ are independent for every k. To prove this we need the following result, see for example [15, Appendix VI, Theorem 2, p. 420],

Theorem 6.1.1 (Cochran's Theorem) *Let $Z_1, \dots, Z_m$ be independently $\mathcal{N}(0,1)$, and let $Q_1, \dots Q_s$ be quadratic forms in the Z_i such that*

$$\sum_i Z_i^2 = Q_1 + \cdots + Q_s.$$

Let $m_j = rank(Q_j)$. Then $Q_1, \dots Q_s$ have independent chi-square distributions with $m_1, \dots, m_s$ degrees of freedom, respectively, if and only if $\sum_j m_j = m$.

For k fixed, note that

$$\sum_{ij} X_{ijk}^2 = IJX_{..k} + J\sum_i (X_{i.k} - X_{..k})^2 + \sum_{ij} (X_{ijk} - X_{i.k})^2.$$

Let Q_1, Q_2 and Q_3 denote respectively the first, second and third term of the right hand expression. Then by Cochran's Theorem we have that $\sum_i (X_{i.k} - X_{..k})^2$ and $\sum_{ij}(X_{ijk} - X_{i.k})^2$ are independent, i.e., $(\eta'_{2k})_k$ and $(\eta_{2k})_k$ are independent. □

The statistics SS_a and SS_e provide estimators for σ_1 and σ_2, this is shown in the next section.

6.1.1 Estimation of σ_1 and σ_2

From the expressions of SS_a and SS_e above, we have that

$$\begin{aligned}
\mathbf{E}[\mathrm{SS}_a] &= n\Big[\sigma_1^2 \sum_{k\geq 1} \lambda_{1k}\mathbf{E}[\eta_{1k}] + \sigma_2^2 \sum_{k\geq 1} \lambda_{2k}\mathbf{E}[\eta'_{2k}]\Big], \\
&= n(I-1)[\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2], \quad \text{and} \\
\mathbf{E}[\mathrm{SS}_e] &= \sigma_2^2 \sum_{k\geq 1} \lambda_{2k}\mathbf{E}[\eta_{1k}]\Big], \\
&= I(n-1)[\lambda_2\sigma_2^2],
\end{aligned}$$

where $\lambda_\nu = \sum_{k\geq 1} \lambda_{\nu k}$, $\nu = 1, 2$. Therefore, unbiased estimators of σ_1 and σ_2 are given by

$$\begin{aligned}
\lambda_2\hat{\sigma}_2^2 &= \frac{\mathrm{SS}_e}{I(n-1)}, \\
\lambda_1\hat{\sigma}_1^2 &= \frac{\mathrm{SS}_a}{I-1} - \frac{\mathrm{SS}_e}{I(n-1)}
\end{aligned}$$

Using classical notation, put

$$\mathrm{MS}_a = \frac{\mathrm{SS}_a}{(I-1)}, \quad \textit{and} \quad \mathrm{MS}_e = \frac{\mathrm{SS}_e}{I(n-1)},$$

so that

$$\hat{\sigma}_2^2 = \frac{\mathrm{MS}_e}{\lambda_2}, \quad \textit{and} \quad \hat{\sigma}_1^2 = \frac{\mathrm{MS}_a - \mathrm{MS}_e}{\lambda_1}.$$

Here again, as in the real case, the estimator could turn out to be negative. Theoretically this problem can be avoided by taking large samples. Indeed note that

$$\frac{\mathrm{MS}_a - \mathrm{MS}_e}{n} = \frac{1}{I-1}\sum_i \|a_i - a_.\|_2^2 + r_{in},$$

with

$$\begin{aligned}
r_{in} &= \frac{1}{I-1}\Big[\sum_i \|e_{i.} - e_{..}\|_2^2 + 2\sum_i < a_i - a_., e_{i.} - e_{..} >\Big] - \\
&\quad \frac{1}{In(n-1)}\sum_{ij} \|e_{ij} - e_{i.}\|_2^2.
\end{aligned}$$

By the strong law of large numbers, for $i = 1, \ldots, I$, the sequence $(r_{in})_n$ converges almost surely to 0. Therefore, for n large enough,

$$\hat{\sigma}_1^2 \simeq \frac{1}{I-1}\sum_i \|a_i - a_.\|_2^2 \geq 0.$$

6.1.2 Testing for the Random Factor Effect

The absence of the random factor effect can be expressed by $\sigma_1 = 0$. Let H_a denote the hypothesis: $H_a : \sigma_1 = 0$. Exactly as in the real case we have the following test:

Proposition 6.1.2 *For testing the hypothesis $H_a : \sigma_1 = 0$ versus the alternative $\sigma_1 \neq 0$, there exists a level-α test $\psi_{H_a,\alpha}$ given by*

$$\psi_{H_a,\alpha} = \begin{cases} 1 & \text{if } \frac{\text{MS}_a}{\text{MS}_e} > C_{H_a,\alpha} \\ 0 & \text{otherwise,} \end{cases}$$

where $C_{H_a,\alpha}$ is a non-negative constant such that

$$\mathbf{P}\Big\{\frac{\text{MS}_a}{\text{MS}_e} > C_{H_a,\alpha}, \sigma_1 = 0\Big\} \leq \alpha.$$

Note that under the null hypothesis H_a we have that

$$\frac{\text{MS}_a}{\text{MS}_e} = n\frac{I(n-1)}{I-1}\frac{\sum_{k\geq 1}\lambda_{2k}\eta'_{2k}}{\sum_{k\geq 1}\lambda_{1k}\eta_{1k}},$$

where the η'_{2k} are independent and identically distributed χ^2_{I-1}, the η_{2k} are independent and identically distributed $\chi^2_{I(I-1)}$, and the two sequences (η'_{2k}) and η_{2k} are independent.

6.2 The Two-Way Layout

Suppose now that n measurements of the same cloud are taken at I independent regions under J independent weather conditions. Each observation can be written as

$$Y = m + e, \tag{6.3}$$

where e is the measurements error, and m is the effect of the regions and weather conditions. As usual the error is assumed to be a mean zero process. We assume that m and e are stochastically independent. Since m depends on two independent factors we can suppose that it is defined on a product probability space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, \mathbf{P}_1 \times \mathbf{P}_2)$. Define the deterministic function μ by

$$\forall t \in [0,1] \quad \mu(t) = \int_{\Omega_1\times\Omega_2} m(t)\, d\mathbf{P}_1 \times \mathbf{P}_2.$$

For $t \in [0,1]$, let

$$\begin{aligned} a(t) &= \int_{\Omega_2} m(t)\, d\mathbf{P}_2 - \mu(t), \\ b(t) &= \int_{\Omega_1} m(t)\, d\mathbf{P}_1 - \mu(t), \text{ and} \\ c(t) &= m - \int_{\Omega_2} m(t)\, d\mathbf{P}_2 - \int_{\Omega_1} m(t) d\mathbf{P}_1 + \mu(t). \end{aligned}$$

Note that

$$\mathbf{E}_{\Omega_1}[c] = \mathbf{E}_{\Omega_2}[c] = \mathbf{E}[c] = 0,$$

where $\mathbf{E}_{\Omega_i}$ denotes the expectation with respect to $\mathbf{P}_i$. The equation (6.3) becomes

$$Y(t) = \mu(t) + a(t) + b(t) + c(t) + e.$$

The function μ is the overall mean, a is the error due to the first random factor, b is the error due to the second random factor, and c is the interaction effect.
The processes a, b and c are all mean zero, and a and b are independent. Moreover a and c are uncorrelated. Indeed,

$$\begin{aligned}
\mathbf{E}[a\,c] &= \mathbf{E}_{\Omega_1}[\mathbf{E}_{\Omega_2}[a\,c]] \\
&= \mathbf{E}_{\Omega_1}[a\,\mathbf{E}_{\Omega_2}[c]] \\
&= 0.
\end{aligned}$$

Similar arguments show that b and c are also uncorrelated.

Let Y_{ijk} denote the kth observation at the ith region with the jth weather condition. The observations can then be written as

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk},$$

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, n$. The component μ is the overall mean, a_i and b_j are the random effects due respectively to the region and the weather condition, and c_{ij} is the interaction random effect .

For hypotheses testing, we have to add some normality assumptions. More precisely we suppose that $(a_i)_i, (b_j)_j, (c_{ij})_{ij}$, and $(e_{ijk})_{ijk}$ are such that:

1. $(a_i)_i$ are independent and identically distributed as $\sigma_1\epsilon_1$,

2. $(b_j)_j$ are independent and identically distributed as $\sigma_2\epsilon_2$,

3. $(c_{ij})_{ij}$ are independent and identically distributed as $\sigma_3\epsilon_3$, and

4. $(e_{ijk})_{ijk}$ are independent and identically distributed as $\sigma_4\epsilon_4$,

where $\sigma_1, \sigma_2, \sigma_3$, and σ_4 are unknown scalars, and $\epsilon_1, \epsilon_2, \epsilon_3$, and ϵ_4 are mean zero Gaussian processes. We have seen that $\epsilon_1, \epsilon_2, \epsilon_3$ are uncorrelated. We assume that they are jointly Gaussian, and thus independent.

The least squares estimators of μ, a_i, b_j, and c_{ij} are obtained by minimizing the sum of squares of the errors given by:

$$\text{ESS} = \sum_{ijk} \|Y_{ijk} - \mu - a_i - b_j - c_{ij}\|_2^2.$$

Let (φ_l) denote an orthonormal basis of $\mathbf{L}_2$. Since

$$\forall f \in \mathbf{L}_2 \quad \|f\|_2^2 = \sum_l f_l^2, \quad \text{with} \quad f_k = \int_0^1 f(s)\varphi_k(s)ds, \tag{6.4}$$

we have that

$$ESS = \sum_l \sum_{ijk} (Y_{ijkl} - \mu_l - a_{il} - b_{jl} - c_{ijl})^2.$$

Computing the derivatives and equating to zero, we see that the least squares estimators are solutions for the normal equations:

$$\begin{aligned}
\hat{\mu} &= Y_{...} - \hat{a}_{.} - \hat{b}_{.} - \hat{c}_{..} \\
\hat{a}_i &= Y_{i..} - \hat{\mu} - \hat{b}_{.} - \hat{c}_{i.} \\
\hat{b}_j &= Y_{.j.} - \hat{\mu} - \hat{a}_{.} - \hat{c}_{.j} \\
\hat{c}_{ij} &= Y_{ij.} - \hat{\mu} - \hat{a}_i - \hat{b}_j.
\end{aligned}$$

One solution to these equations is given by:

$$\begin{aligned}
\hat{\mu} &= Y_{...}, \\
\hat{a}_i &= Y_{i..} - Y_{...}, \\
\hat{b}_j &= Y_{.j.} - Y_{...}, \text{and} \\
\hat{c}_{ij} &= Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...}.
\end{aligned}$$

These solutions lead to estimators for $(\sigma_i)_{i=1,\dots,4}$. Before giving these estimators we introduce some more notation.

For $i = 1, \dots, 4$, let $(\lambda_{ik}, \varphi_{ik})_k$ denote the pairs of eigenvalue-eigenvector associated with the covariance operator T_{ϵ_i}. For every i, j and k, using Karhünen-Loève expansion we have that

$$\begin{aligned}
a_i &= \sigma_1 \sum_{l\geq 1} \sqrt{\lambda_{1l}}\, X_{1il}\varphi_{1l}, \\
b_j &= \sigma_2 \sum_{l\geq 1} \sqrt{\lambda_{2l}}\, X_{2jl}\varphi_{2l}, \\
c_{ij} &= \sigma_3 \sum_{l\geq 1} \sqrt{\lambda_{3l}}\, X_{3ijl}\varphi_{3l}, and \\
e_{ijk} &= \sigma_4 \sum_{k\geq 1} \sqrt{\lambda_{4l}}\, X_{4ijkl}\varphi_{4l},
\end{aligned}$$

where the X are independent and identically distributed standard normal random variables.

Using equation (6.4) and the known results on sums of squares of Gaussian random variables it can easily be seen that we have

$$\begin{aligned}
\sum_i \|a_i - a_.\|_2^2 &= \sigma_1^2 \sum_{l\geq 1} \lambda_{1l}\eta_{1l,[I-1]}, && (6.5)\\
\sum_j \|b_j - b_.\|_2^2 &= \sigma_2^2 \sum_{l\geq 1} \lambda_{2l}\eta_{2l,[J-1]}, && (6.6)\\
J\sum_i \|c_{i.} - c_{..}\|_2^2 &= \sigma_3^2 \sum_{l\geq 1} \lambda_{3l}\eta_{3l,[I-1]}, && (6.7)\\
I\sum_j \|c_{.j} - c_{..}\|_2^2 &= \sigma_3^2 \sum_{l\geq 1} \lambda_{3l}\eta^{(1)}_{3l,[J-1]}, && (6.8)\\
\sum_i \|c_{ij} - c_{i.} - c_{.j} + c_{..}\|_2^2 &= \sigma_3^2 \sum_{l\geq 1} \lambda_{3l}\eta^{(2)}_{3l,[(I-1)(J-1)]}, && (6.9)\\
Jn\sum_i \|e_{i..} - e_{...}\|_2^2 &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l}\eta^{(1)}_{4l,[I-1]}, && (6.10)\\
In\sum_j \|e_{.j.} - e_{...}\|_2^2 &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l}\eta^{(2)}_{4l,[J-1]}, && (6.11)\\
n\sum_{ij} \|e_{ij.} - e_{i..}e_{.j.} + e_{...}\|_2^2 &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l}\eta^{(3)}_{4l,[(I-1)(J-1)]}, && (6.12)\\
\sum_{ijk} \|e_{ijk} - e_{ij.}\|_2^2 &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l}\eta_{4l,[IJ(n-1)]}, && (6.13)
\end{aligned}$$

where the η's are independent real random variables distributed as a χ^2. The subscripts between square brackets are the degrees of freedom.

6.2.1 Estimation of $\sigma_1, \sigma_2, \sigma_3$ and σ_4

Define $\mathrm{SS}_a, \mathrm{SS}_b, \mathrm{SS}_c$, and SS_e by

$$\begin{aligned}
\mathrm{SS}_a &= \sum_{ijk} \|\hat{a}_i\|_2^2,\\
\mathrm{SS}_b &= \sum_{ijk} \|\hat{b}_i\|_2^2,\\
\mathrm{SS}_c &= \sum_{ijk} \|\hat{c}_i\|_2^2, and\\
\mathrm{SS}_e &= \sum_{ijk} \|Y_{ijk} - \hat{Y}_{ijk}\|_2^2.
\end{aligned}$$

Replacing the estimators by their values, these equations become

$$\mathrm{SS}_a = Jn\sum_i \|(a_i - a_.) + (c_{i.} - c_{..}) + (e_{i..} - e_{...})\|_2^2,$$

$$
\begin{aligned}
\mathrm{SS}_b &= In\sum_j \|(b_i - b_.) + (c_{.j} - c_{..}) + (e_{.j.} - e_{...})\|_2^2, \\
\mathrm{SS}_c &= n\sum_{ij} \|(c_{ij} - c_{i.} - c_{.j} + c_{..}) + (e_{ij.} - e_{i..} - e_{.j.} + e_{...})\|_2^2, \\
\mathrm{SS}_e &= \sum_{ijk} \|e_{ijk} - Y_{ij.}\|_2^2.
\end{aligned}
$$

Note that SS_a can also be written as

$$
\begin{aligned}
\mathrm{SS}_a &= Jn\Big[\sum_i \|a_i - a_.\|_2^2 + \sum_i \|c_{i.} - c_{..}\|_2^2 + 2\sum_i < a_i - a_., c_{i.} - c_{..} > \\
&\quad + \sum_i \|e_{i..} - e_{...}\|_2^2 + 2\sum_i < a_i - a_. + c_{i.} - c_{..}, e_{i..} - e_{...} >\Big].
\end{aligned}
$$

Using the expressions (6.5), (6.7), (6.10) and independence we obtain

$$
\mathbf{E}[\mathrm{SS}_a] = (I-1)\Big[Jn\sigma_1^2\lambda_1 + n\sigma_3^2\lambda_3 + \sigma_4^2\lambda_4\Big].
$$

Using the same arguments for SS_b, and SS_c we obtain

$$
\begin{aligned}
\mathbf{E}[\mathrm{SS}_b] &= (J-1)\Big[In\sigma_2^2\lambda_2 + n\sigma_3^2\lambda_3 + \sigma_4^2\lambda_4\Big], \\
\mathbf{E}[\mathrm{SS}_c] &= (I-1)(J-1)\Big[n\sigma_3^2\lambda_3 + \sigma_4^2\lambda_4\Big], \textit{ and} \\
\mathbf{E}[\mathrm{SS}_e] &= IJ(n-1)\sigma_4^2\lambda_4,
\end{aligned}
$$

with $\lambda_\nu = \sum_l \lambda_{\nu l}$, for $\nu = 1, \ldots, 4$.
Therefore we have

$$
\begin{aligned}
\lambda_4\hat{\sigma}_4^2 &= \frac{\mathrm{SS}_e}{IJ(n-1)} \doteq \mathrm{MS}_e, \\
\lambda_3\hat{\sigma}_3^2 &= \frac{1}{n}\Big[\frac{\mathrm{SS}_c}{(I-1)(J-1)} - \mathrm{MS}_e\Big] \doteq \frac{1}{n}[\mathrm{MS}_c - \mathrm{MS}_e], \\
\lambda_2\hat{\sigma}_2^2 &= \frac{1}{In}\Big[\frac{\mathrm{SS}_b}{J-1} - \mathrm{MS}_e\Big] \doteq \frac{1}{In}[\mathrm{MS}_b - \mathrm{MS}_c], \\
\lambda_1\hat{\sigma}_1^2 &= \frac{1}{In}\Big[\frac{\mathrm{SS}_a}{I-1} - \mathrm{MS}_c\Big] \doteq \frac{1}{Jn}[\mathrm{MS}_a - \mathrm{MS}_c].
\end{aligned}
$$

Remark 6.2.1 *1. By the strong law of large numbers, we have that*

$$
\lim_{n\to\infty} \frac{\mathrm{SS}_e}{n(n-1)} = 0 \quad a.s.,
$$

therefore, $\hat{\sigma}_3^2$ is almost surely non-negative as $n \longrightarrow \infty$.

2. By the same arguments we have that, for I large enough, $\hat{\sigma}_2^2$ is almost surely non-negative, and for J large enough $\hat{\sigma}_1^2$ is almost surely non-negative.

6.2.2 Testing for the Factors Effects

The absence of the first factor effect is expressed by $\sigma_1 = 0$, $\sigma_2 = 0$ expresses that there is no second factor effect, and $\sigma_3 = 0$ means that there is no interaction effect. Consider the hypotheses

$$\mathrm{H}_a : \sigma_1 = 0, \quad \mathrm{H}_b : \sigma_2 = 0, \quad \text{and } \mathrm{H}_c : \sigma_3 = 0.$$

The hypothesis H_c

Under H_c, the SS_c becomes

$$\begin{aligned} \mathrm{SS}_{c,H_c} &= n \sum_{ij} \|e_{ij.} - e_{i..} - e_{.j.} + e_{...}\|_2^2 \\ &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l} \eta^{(3)}_{4l,[(I-1)(J-1)]}, \end{aligned}$$

where $(\eta^{(3)}_{4l,[(I-1)(J-1)]})_l$ is a sequence of independent real random variables all distributed as a $\chi^2_{(I-1)(J-1)}$. Furthermore, by Cochran's theorem, SS_c and SS_e are independent.

Proposition 6.2.1 *For testing the hypothesis $H_c : \sigma_1 = 0$ versus the alternative $\sigma_1 \neq 0$, there exists a level-α test $\psi_{H_c,\alpha}$ given by*

$$\psi_{H_c,\alpha} = \begin{cases} 1 & \text{if } \frac{\mathrm{SS}_c}{\mathrm{SS}_e} > C_{H_c,\alpha} \\ 0 & \text{otherwise,} \end{cases}$$

where $C_{H_c,\alpha}$ is a non-negative constant such that

$$\mathbf{P}\Big\{\sum_l \lambda_{4l}\eta^{(3)}_{4l,[I-1)(J-1)]} / \sum_l \lambda_{4l}\eta_{4l,[I-1)(J-1)]} > C_{H_c,\alpha}\Big\} \leq \alpha.$$

The hypothesis H_a

We will develop two tests for H_a; one under H_c and the other under $\bar{H}_c$, the alternative of H_c.

a- Test for H_a under H_c

Under both H_a and H_c we have

$$\begin{aligned} \mathrm{SS}_{a,H_aH_c} &= Jn \sum_i \|e_{i..} - e_{...}\|_2^2 \\ &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l} \eta^{(1)}_{4l,[I-1]}, \quad \textit{and} \\ \mathrm{SS}_{c,H_aH_c} &= n \sum_{ij} \|e_{ij.} - e_{i..} - e_{.j.} + e_{...}\|_2^2 \\ &= \sigma_4^2 \sum_{l\geq 1} \lambda_{4l} \eta^{(3)}_{4l,[(I-1)(J-1)]}, \end{aligned}$$

where the $\eta^{(1)}_{4l,[I-1]}$ are , independent and identically distributed as χ^2_{I-1}, and the $\eta^{(3)}_{4l,[(I-1)(J-1)]}$ are independent and identically distributed as $\chi^2_{(I-1)(J-1)}$. Furthermore, by Cochran's theorem, SS_{a,H_aH_c} and SS_{c,H_aH_c} are independent.

Proposition 6.2.2 *Under $H_c : \sigma_3 = 0$, for testing the hypothesis $H_a : \sigma_1 = 0$ versus the alternative $\sigma_1 \neq 0$, there exists a level-α test $\psi_{H_aH_c,\alpha}$ given by*

$$\psi_{H_aH_c,\alpha} = \begin{cases} 1 & \text{if } \frac{\mathrm{SS}_a}{\mathrm{SS}_c} > C_{H_aH_c,\alpha} \\ 0 & \text{otherwise,} \end{cases}$$

where $C_{H_aH_c,\alpha}$ is a non-negative constant such that

$$\mathbf{P}\Big\{\sum_l \lambda_{4l}\eta^{(1)}_{4l,[(I-1)(J-1)]} / \sum_l \lambda_{4l}\eta^{(3)}_{4l,[I-1)(J-1)]} > C_{H_aH_c,\alpha}\Big\} \leq \alpha.$$

b- Test for H_a under $\bar{H}_c$

By the strong law of large numbers, we have that

$$\lim_{n\to\infty} \|e_{i..} - e_{...}\|_2^2 = 0, \text{ and } \lim_{n\to\infty} \|e_{ij.} - e_{i..} - e_{.j.} + e_{...}\|_2^2 = 0 \quad \text{a.s..}$$

Therefore

$$\lim_{n\to\infty} \frac{\mathrm{SS}_{a,H_a}}{\mathrm{SS}_c} = J\frac{\sum_i \|c_{i.} - c_{..}\|_2^2}{\sum_{ij}\|c_{ij} - c_{i.} - c_{.j} + c_{..}\|_2^2} \quad a.s.$$

The numerator and denominator of the limit ratio are independent and we have

$$J\frac{\sum_i \|c_{i.} - c_{..}\|_2^2}{\sum_{ij}\|c_{ij} - c_{i.} - c_{.j} + c_{..}\|_2^2} \stackrel{\mathcal{L}}{=} J\frac{\sum_l \lambda_{2l}\eta^{(1)}_{2l,[I-1)(J-1)]}}{\sum_l \lambda_{2l}\eta_{2l}}.$$

Thus we have the following "asymptotic" test

Proposition 6.2.3 *Under $\bar{H}_c$, for testing the hypothesis $H_a : \sigma_1 = 0$ versus the alternative $\sigma_1 \neq 0$, there exists a level-α test $\psi_{H_a\bar{H}_c,\alpha}$ given by*

$$\psi_{H_a\bar{H}_c,\alpha} = \begin{cases} 1 & \text{if } \frac{\mathrm{SS}_a}{\mathrm{SS}_c} > C_{H_a\bar{H}_c,\alpha} \\ 0 & \text{otherwise.} \end{cases}$$

For n large $C_{H_a\bar{H}_c,\alpha}$ can be chosen such that

$$\mathbf{P}\Big\{\sum_l \lambda_{2l}\eta^{(1)}_{2l} / \sum_l \lambda_{2l}\eta_{2l} > C_{H_a\bar{H}_c,\alpha}\Big\} \leq \alpha.$$

The hypothesis H_b

The effects a and b play symmetric roles. Thus, tests for H_b are obtained using similar arguments as under H_a.

6.3 Mixed Models

Suppose now that the region effect is not random, i.e., the effect a_i is not random. Let b_j and c_{ij} be as in the previous section. Without loss of generality we can suppose that our model is given by

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}, \quad i = 1, \ldots, I,\ j = 1, \ldots, J, \text{ and } k = 1, \ldots, n,$$

where b_j, c_{ij} and e_{ijk} are as before and a_i is not random with $a_. = 0$. As before, least squares estimators for μ, a_i, b_j and c_{ij} can be obtained by minimizing ESS and solving the normal equations. Let SS_b, SS_c, and SS_e be as in the previous section. Since $a_. = 0$, SS_a becomes

$$\begin{aligned} \mathrm{SS}_a &= \sum_{ijk} \|\hat{a}_i\|_2^2 \\ &= Jn \sum_{ij} \|a_i + (c_{i.} - c_{..}) + (e_{i..} - e_{...})\|_2^2. \end{aligned}$$

Estimation and Testing

The estimators $\hat{\sigma}_2^2$, $\hat{\sigma}_3^2$, and $\hat{\sigma}_4^2$ given in the random-effect models case are still valid in this case. On the other hand , we have that

$$\mathbf{E}[\mathrm{SS}_a] = Jn \sum_i \|a_i\|_2^2 + n(I-1)\sigma_3^2\lambda_3 + (I-1)\sigma_4^2\lambda_4,$$

where $\lambda_\nu = \sum_l \lambda_{\nu l}$ for $\nu = 3, 4$. Thus an estimator of $\sum_i \|a_i\|_2^2$ is given by

$$\begin{aligned} \widehat{\sum_i \|a_i\|_2^2} &= \frac{1}{Jn}\Big(\mathrm{SS}_a + (I-1)(\mathrm{MS}_c - \mathrm{MS}_e) - (I-1)\mathrm{MS}_c\Big) \\ &= \frac{1}{Jn}\Big(\mathrm{SS}_a + (I-1)\mathrm{MS}_c\Big) \end{aligned}$$

Because $a_. = 0$, the hypothesis $H_a : a_1 = \cdots = a_I$ is equivalent to "$a_i = 0$, for all i". Therefore it can be tested using the same arguments as in the previous section, replacing SS_a by SS_μ.

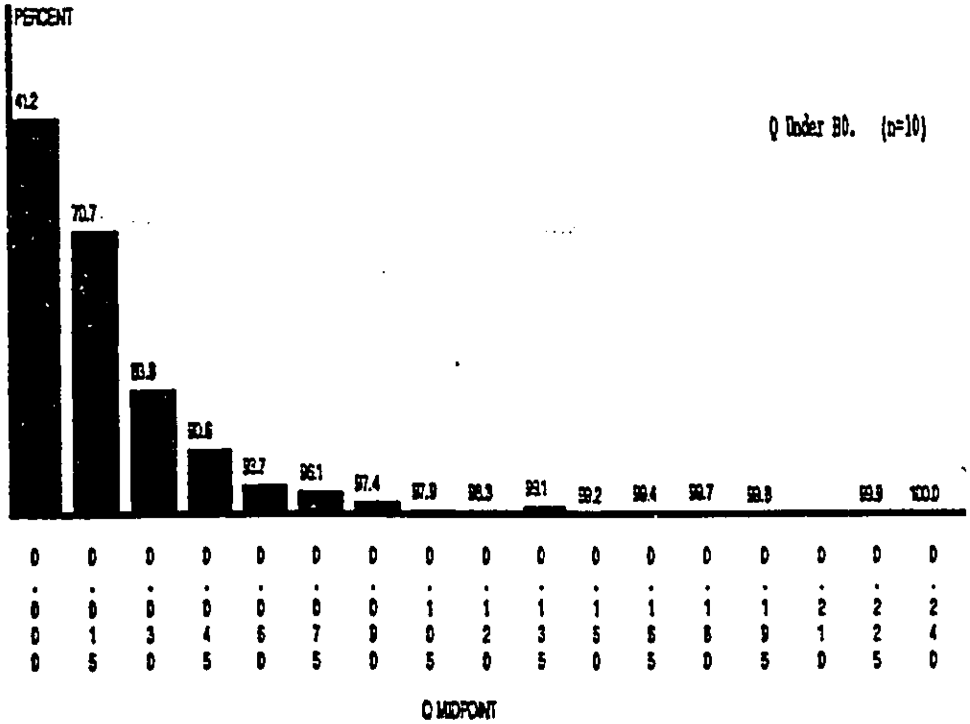
There is a rich literature on random and mixed effects models for the real case. The results we presented here could not be considered as a generalization of the existing ones. As we mentioned in the beginning of this chapter, here we only attempt to outline how the techniques we developed in chapters two and three can be extended to a more general case.

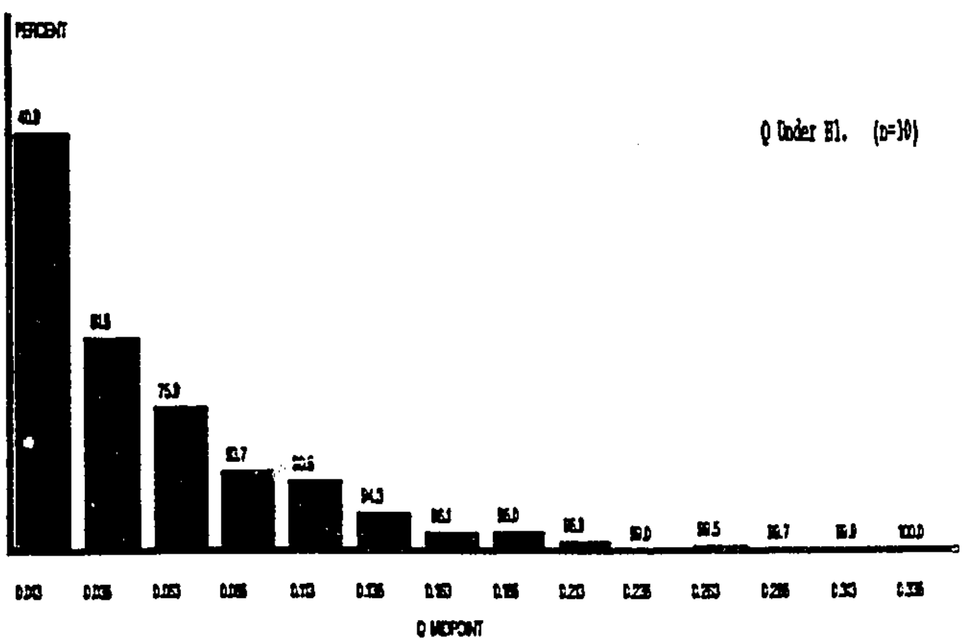
Appendix A

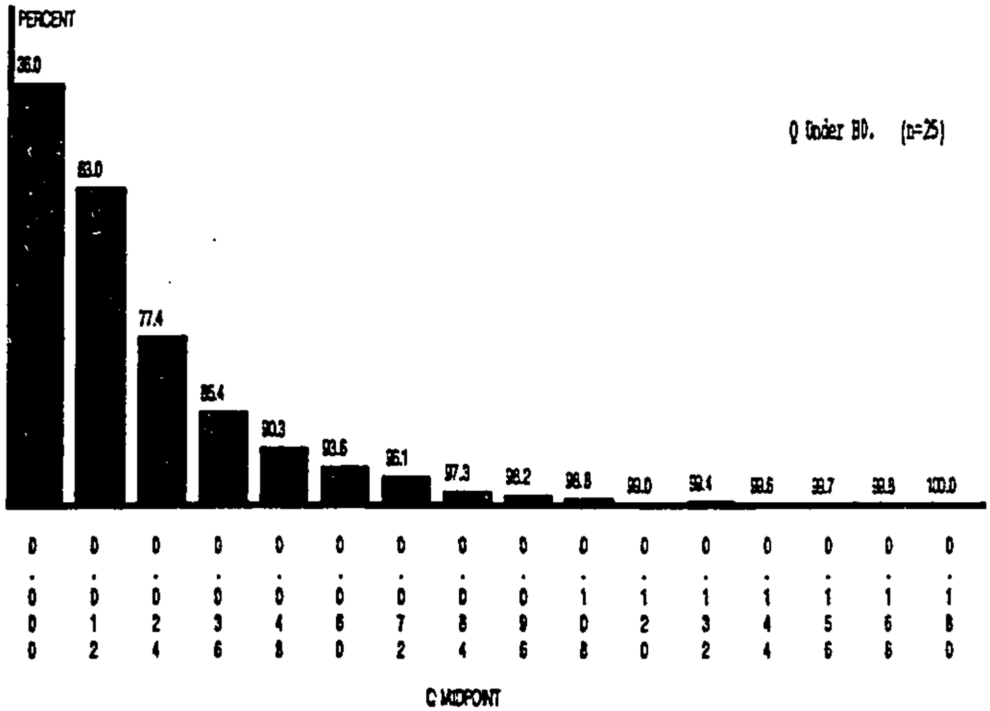
TABLES AND HISTOGRAMS

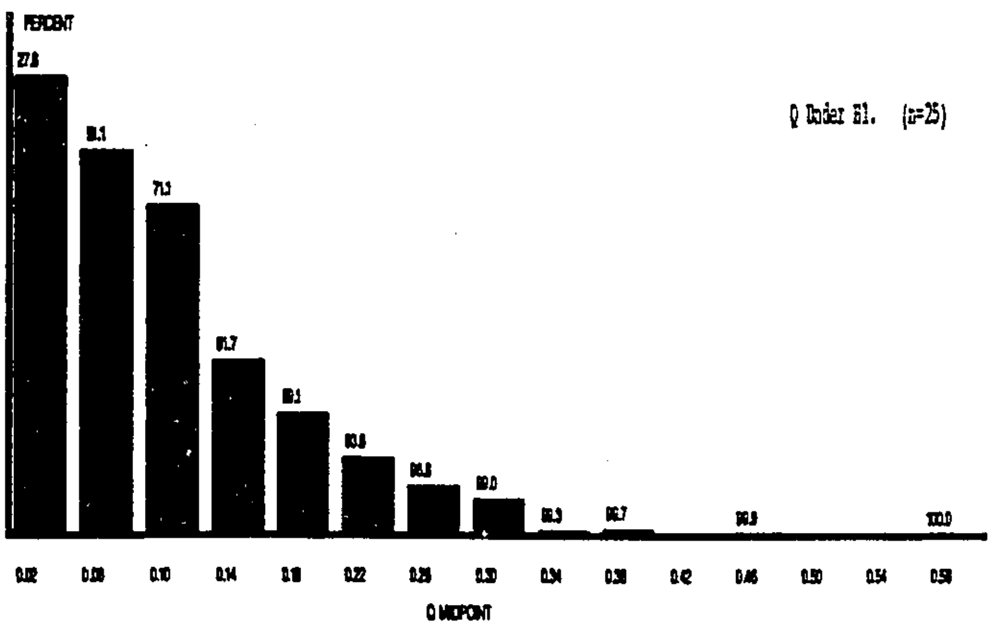
A.1 The Equality of Distributions Example

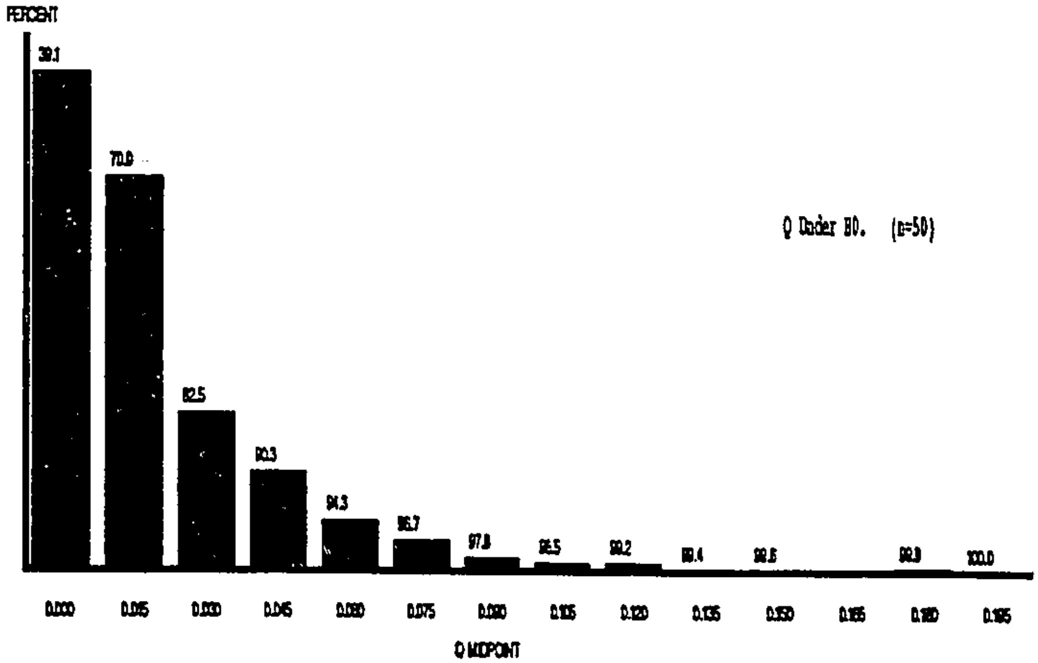
We consider two independent samples each of size n. We use the program given in Appendix B to simulate 1000 realizations under both H_0 :*The two groups are identically distributed* and H_1 :*They are not.* More precisely, we consider for the alternative the case where the first group comes from F, the uniform distribution on [0,1], and the second from $\sqrt{F}$. The following figures are histograms of our statistic **Q**, the Cramer-von Mises statistic W and the Kolmogorov-Smirnov statistic K.

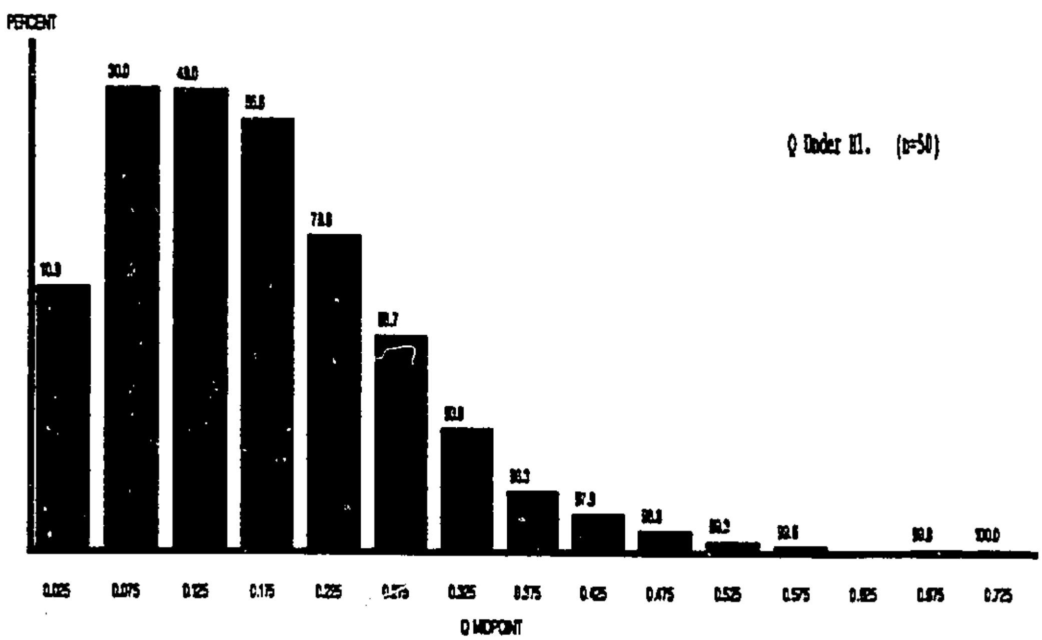

PERCENT
41.2
70.7
97.4
99.2
100.0
Q Under H0. (n=10)
0.000
0.015
0.030
0.045
0.060
0.075
0.090
0.105
0.120
0.135
0.150
0.165
0.180
0.195
0.210
0.225
0.240
Q MIDPOINT

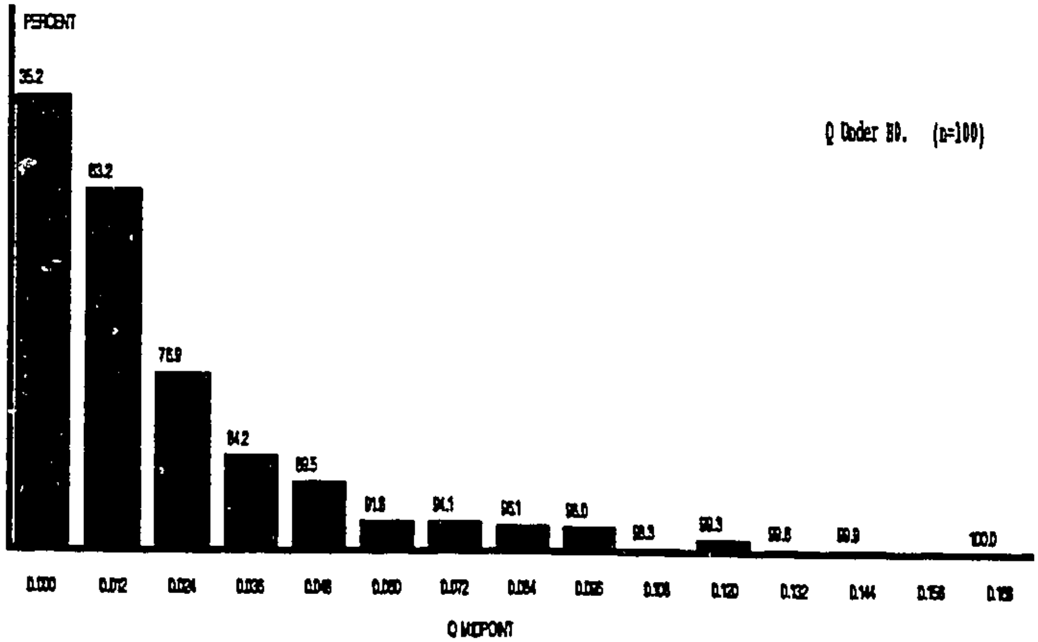

PERCENT
Q Under H1. (n=10)
99.5
100.0

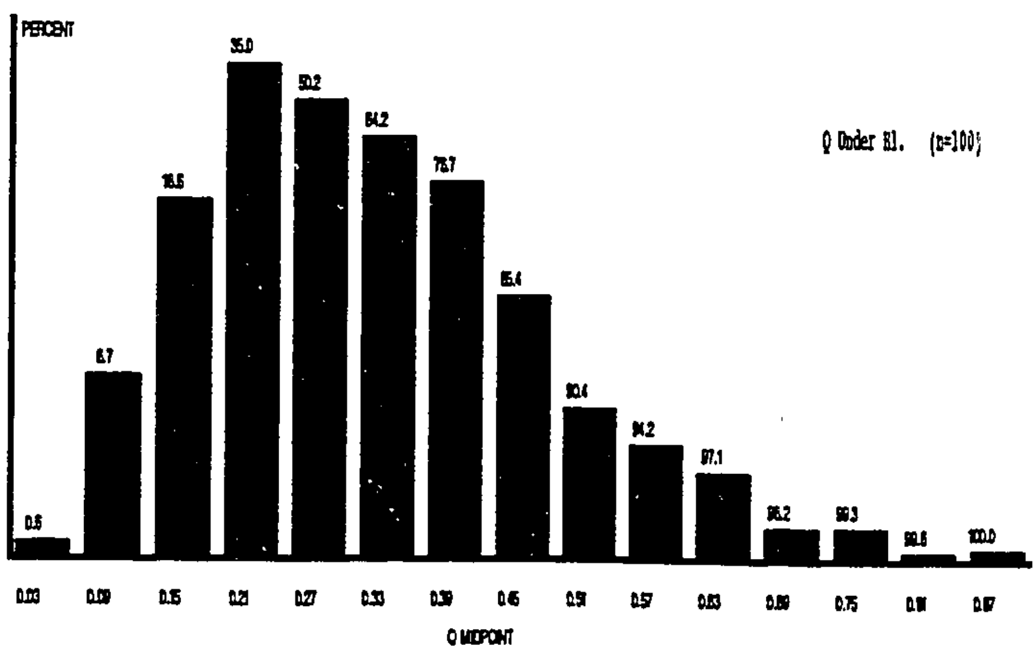

PERCENT
36.0
63.0
77.4
85.4
90.3
93.6
96.1
97.3
98.2
98.8
99.0
99.4
99.6
99.7
99.9
100.0
Q Under H0. (n=25)
Q MIDPOINT

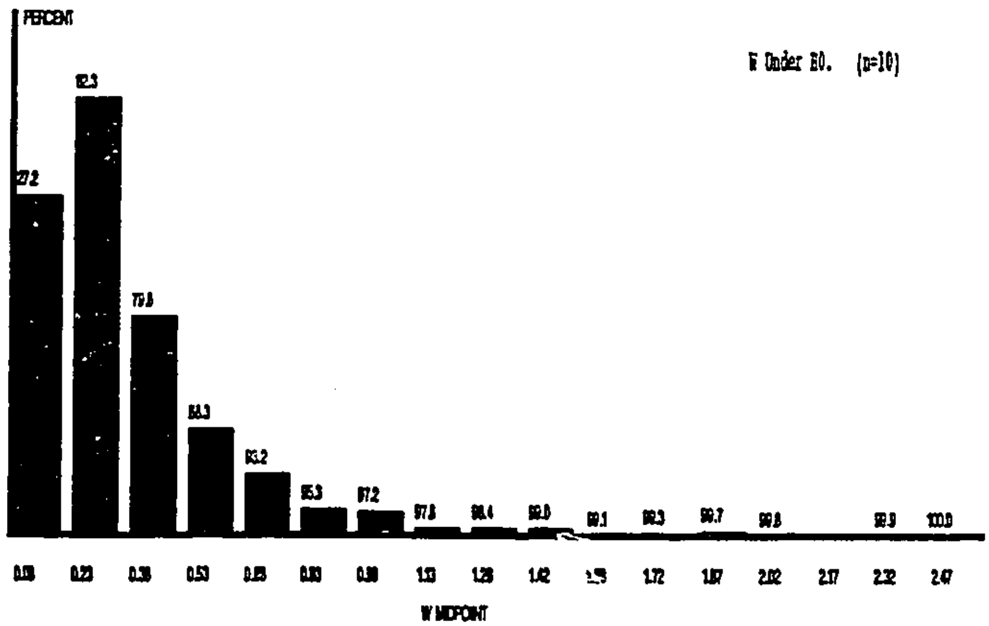

PERCENT
71.1
100.0
Q Under H1. (n=25)
0.02
0.10
0.14
0.22
0.30
0.34
0.42
0.46
0.50
0.54
Q MIDPOINT

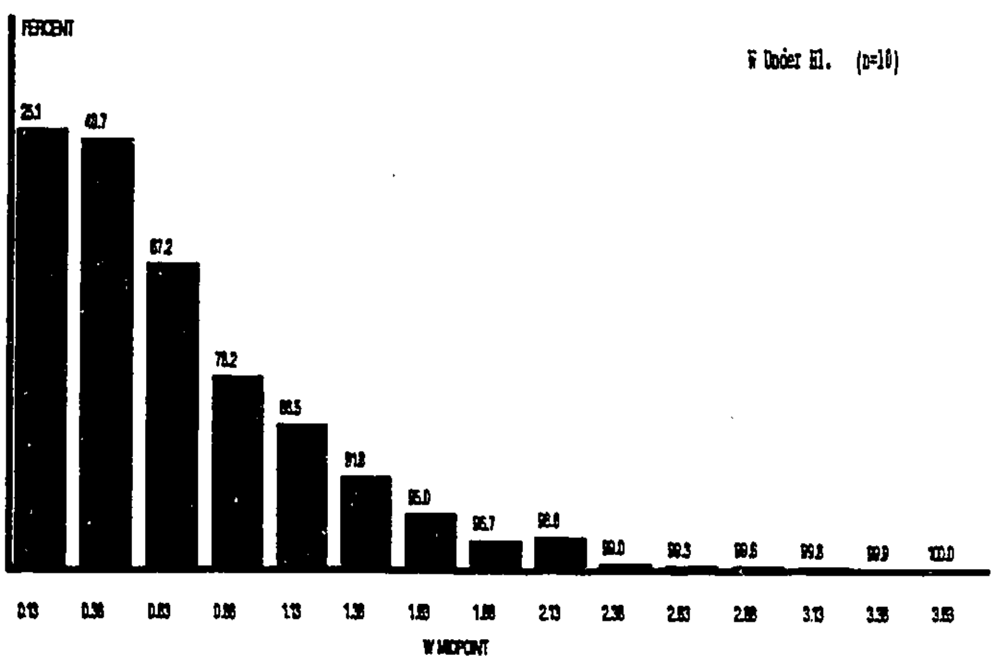

PERCENT
82.5
94.3
96.7
100.0
Q Under H0. (n=50)
0.000
0.015
0.030
0.045
0.060
0.075
0.090
0.105
0.120
0.135
0.150
0.165
0.180
0.195
Q MIDPOINT

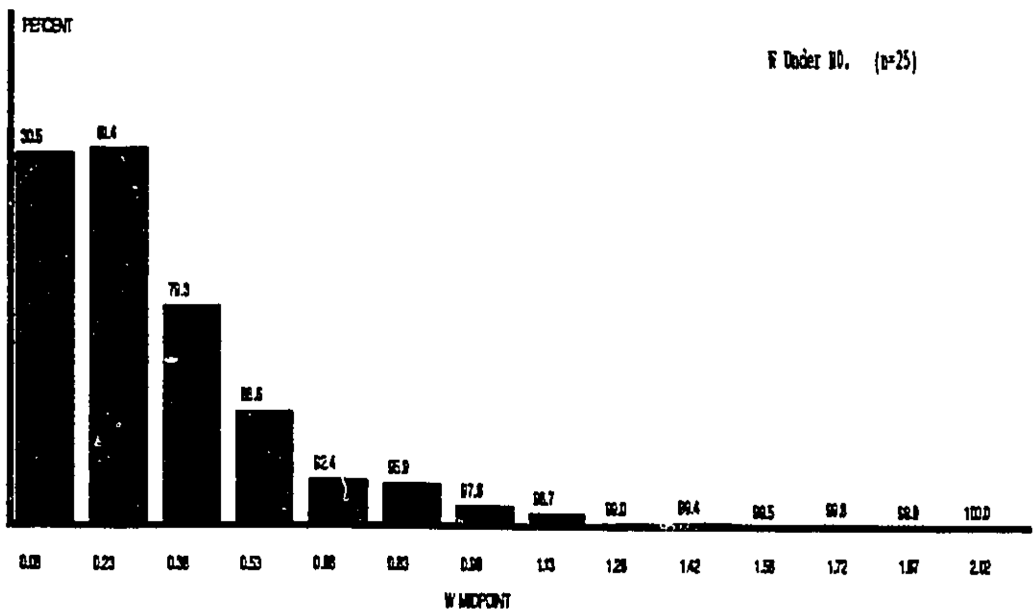

PERCENT
100.0
Q Under H1. (n=50)
0.025
0.075
0.125
0.175
0.225
0.325
0.375
0.425
0.475
0.525
0.575
0.675
0.725
Q MIDPOINT

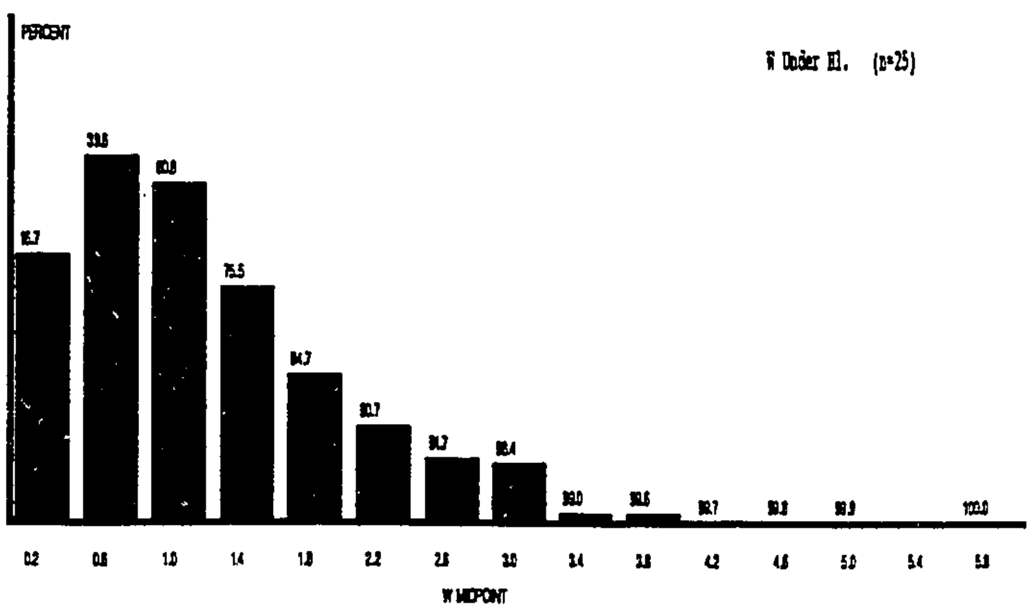

PERCENT
Q Under H0. (n=100)
Q MIDPOINT

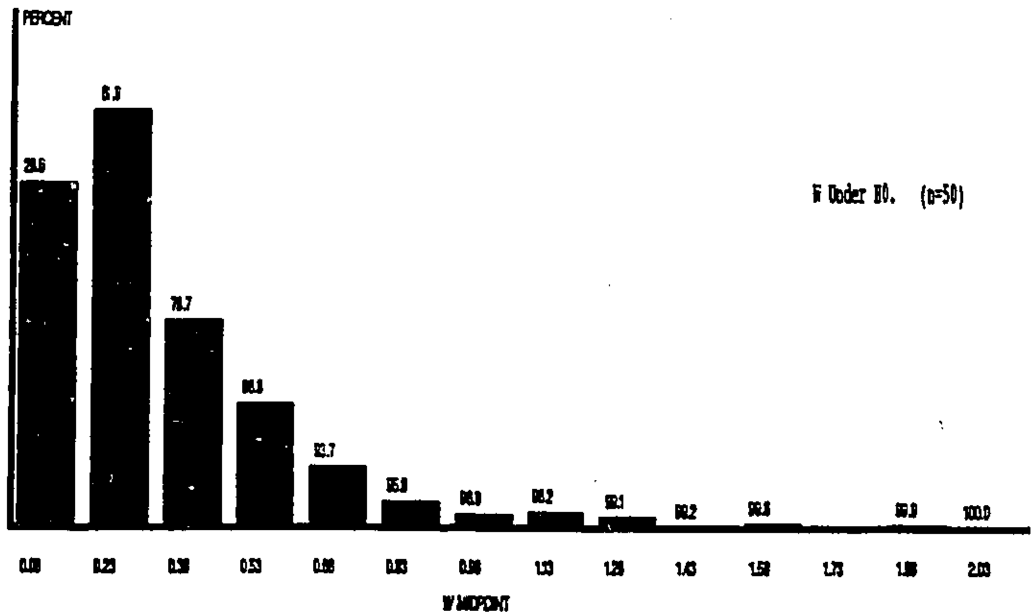

PERCENT
Q Under H1. (n=100)
Q MIDPOINT

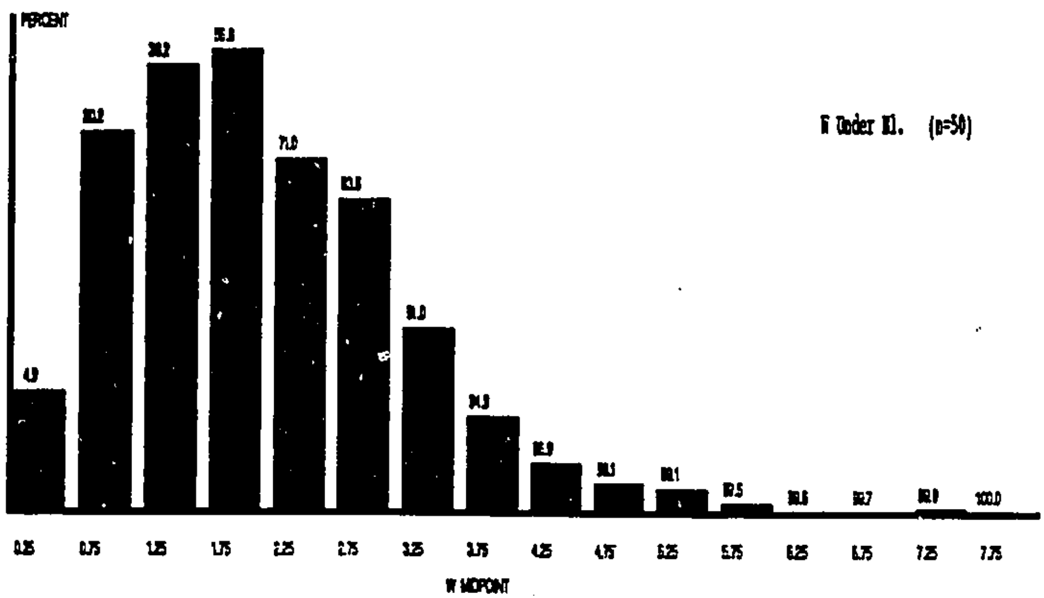

PERCENT
W Under H0. (n=10)
W MIDPOINT

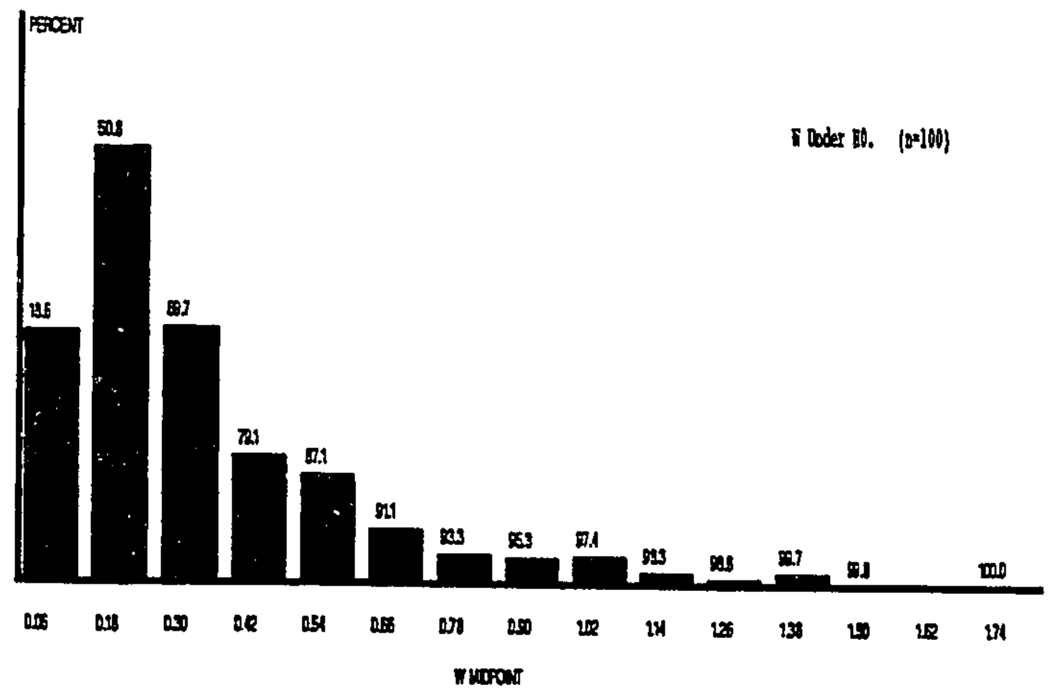

PERCENT
W Under H1. (n=10)
W MIDPOINT

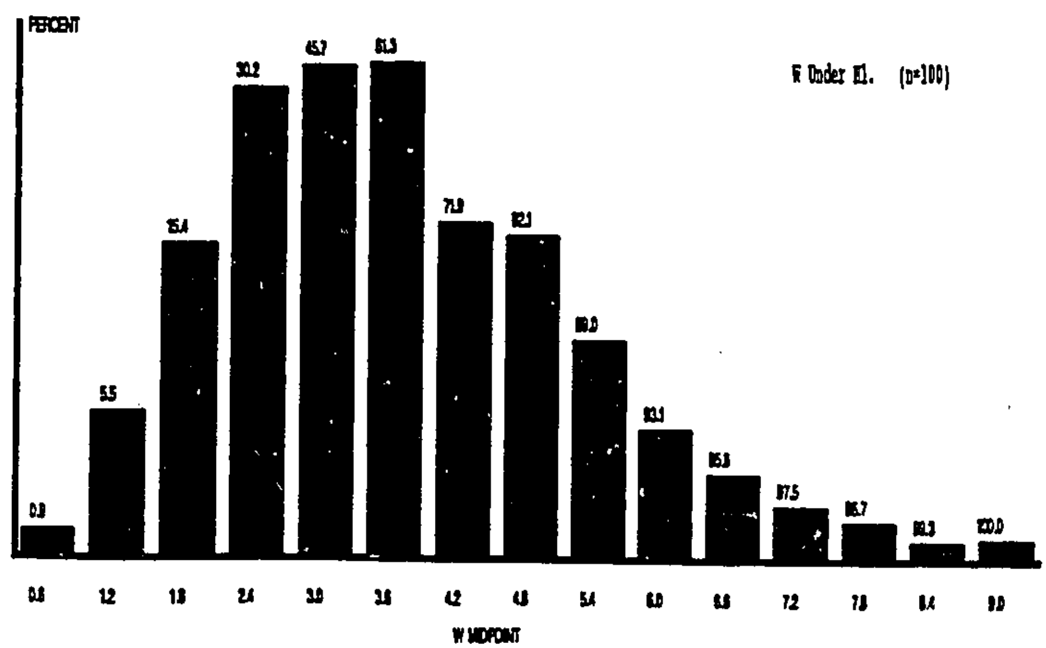

PERCENT
79.3
95.9
100.0
0.23
0.53
1.13
1.28
1.42
1.72
2.02
W MIDPOINT

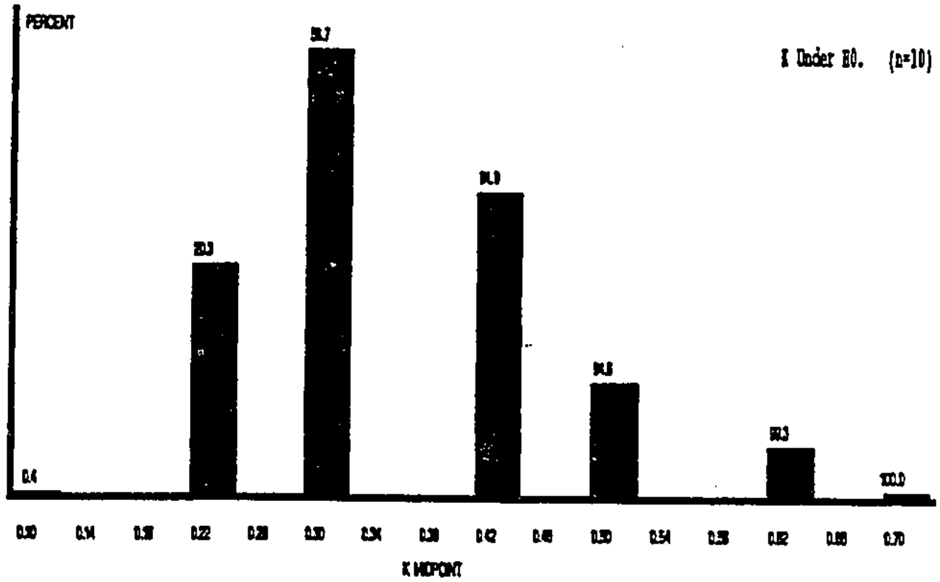

PERCENT
75.5
100.0
0.2
1.0
1.4
2.2
3.0
3.4
4.2
5.0
5.4
W MIDPOINT

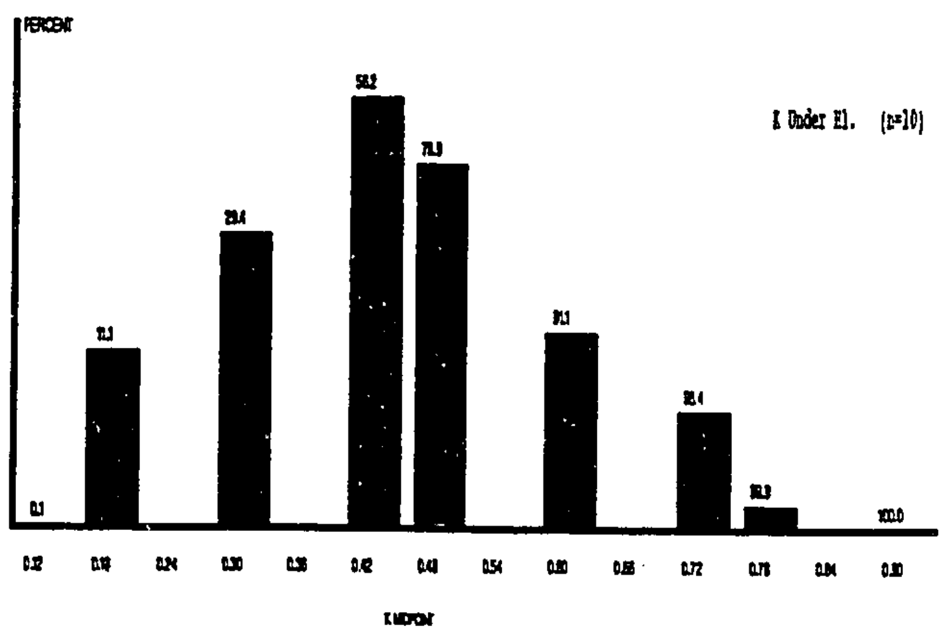

PERCENT
W Under H0. (n=50)
100.0
W MIDPOINT

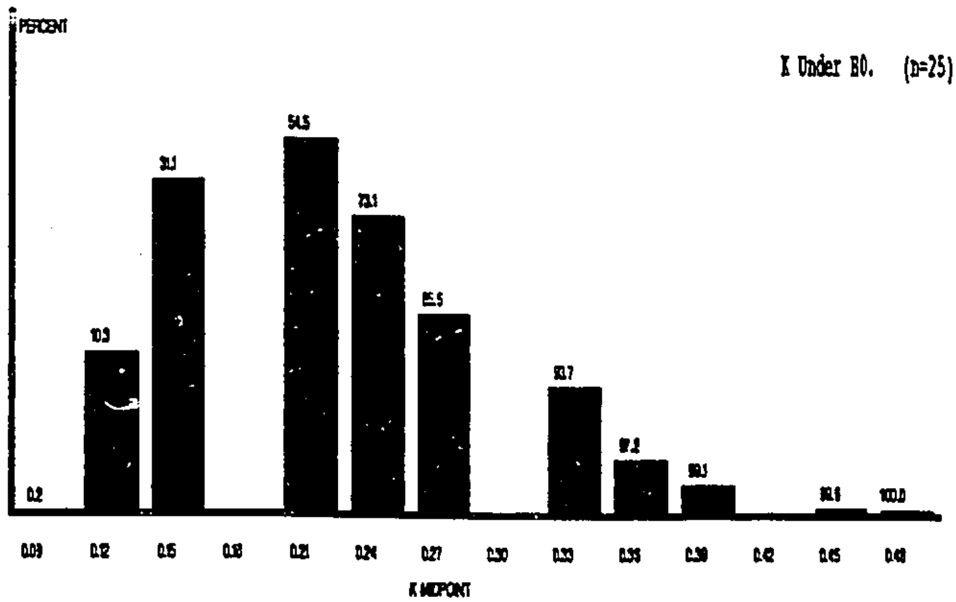

PERCENT
W Under H1. (n=50)
100.0
0.25
0.75
1.25
1.75
2.25
2.75
3.25
3.75
4.25
4.75
5.25
5.75
6.25
6.75
7.25
7.75
W MIDPOINT

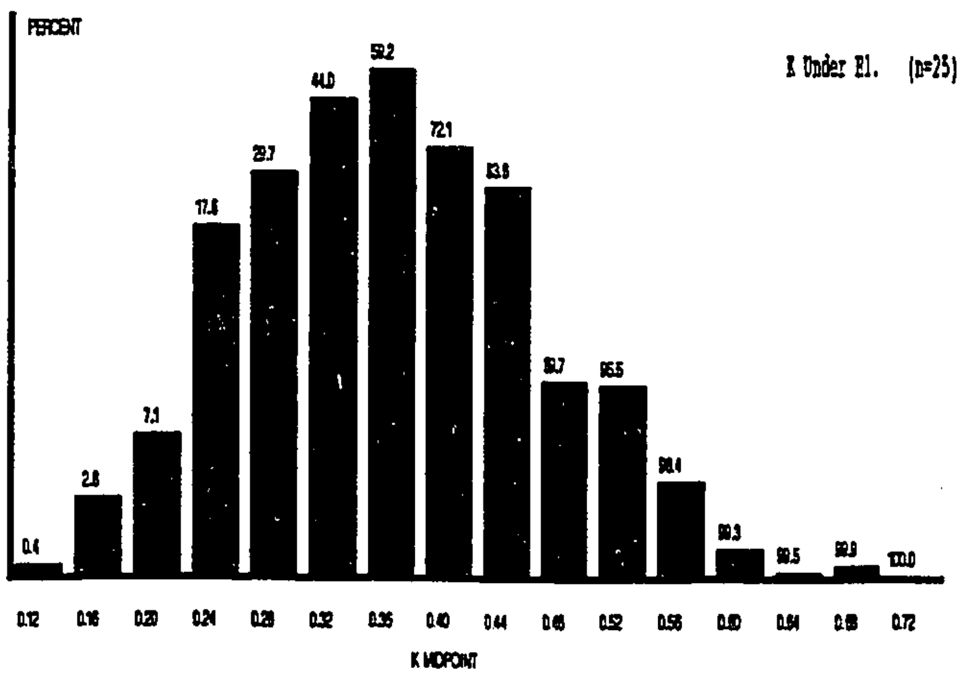

PERCENT
W Under H0. (n=100)
0.06
0.18
0.30
0.42
0.54
0.66
0.78
0.90
1.02
1.14
1.26
1.38
1.50
1.62
1.74
W MIDPOINT

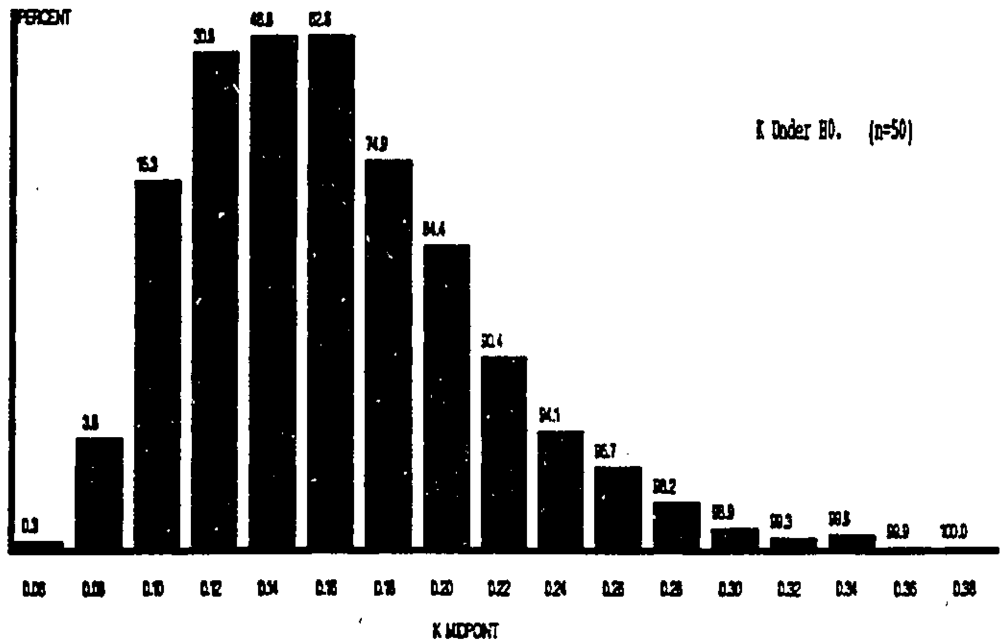

PERCENT
W Under H1. (n=100)
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
6.6
7.2
7.8
8.4
9.0
W MIDPOINT

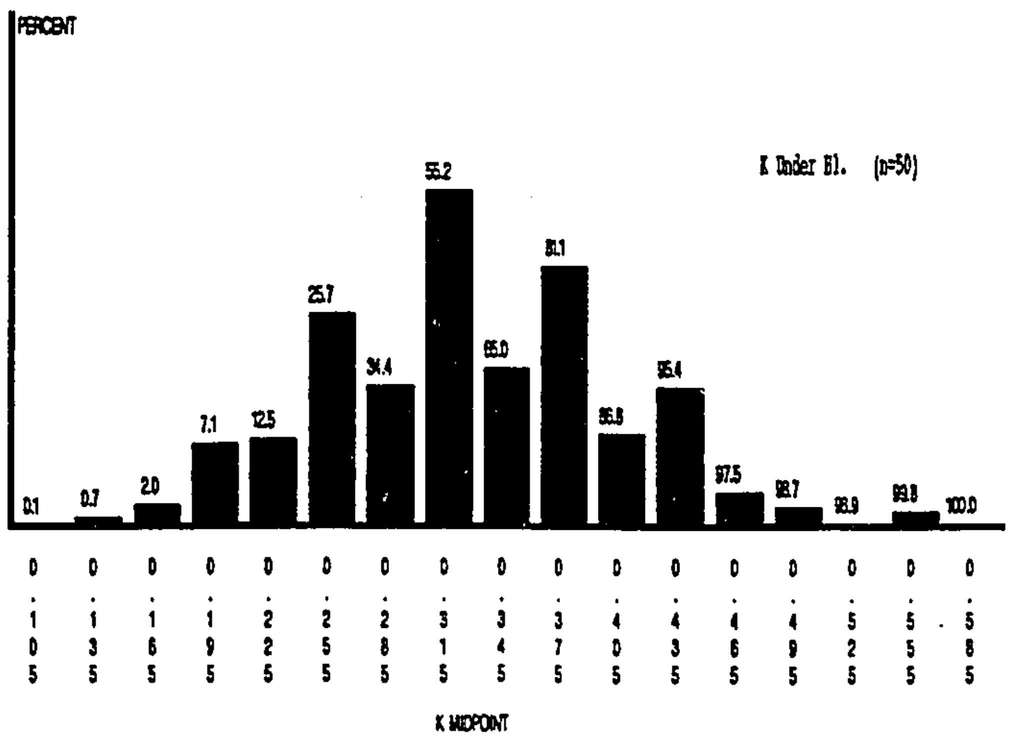

PERCENT
Under H0. (n=10)
K MIDPOINT

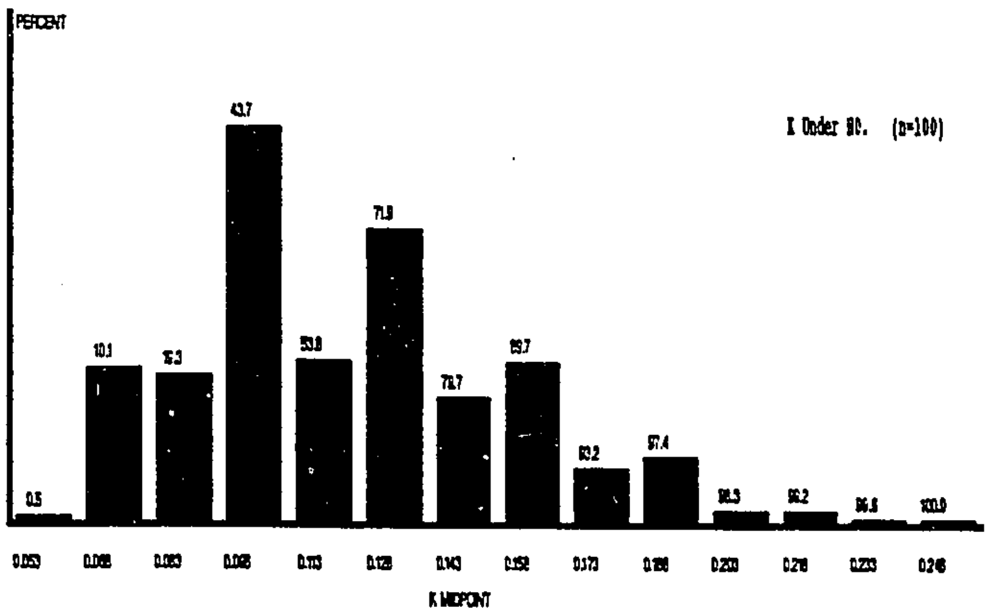

PERCENT
Under H1. (n=10)

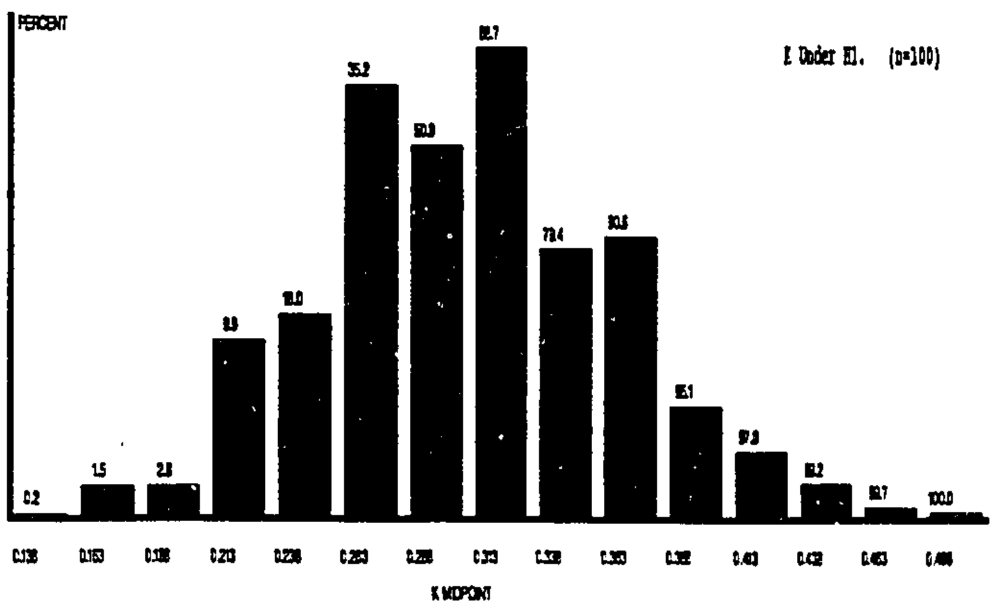

PERCENT
K Under H0. (n=25)
0.2
10.3
54.5
100.0
0.12
0.15
0.21
0.24
0.27
0.30
0.33
0.42
K MIDPOINT


PERCENT
K Under H1. (n=25)
0.4
7.1
29.7
44.0
59.2
72.1
100.0
0.12
0.16
0.20
0.24
0.32
0.36
0.40
0.44
0.52
0.56
0.60
0.72
K MIDPOINT


PERCENT
K Under H0. (n=50)
K MIDPOINT


PERCENT
K Under H1. (n=50)
K MIDPOINT


PERCENT
K Under H0. (n=100)
K MIDPOINT


PERCENT
K Under H1. (n=100)
K MIDPOINT

A.2 The Geological Example

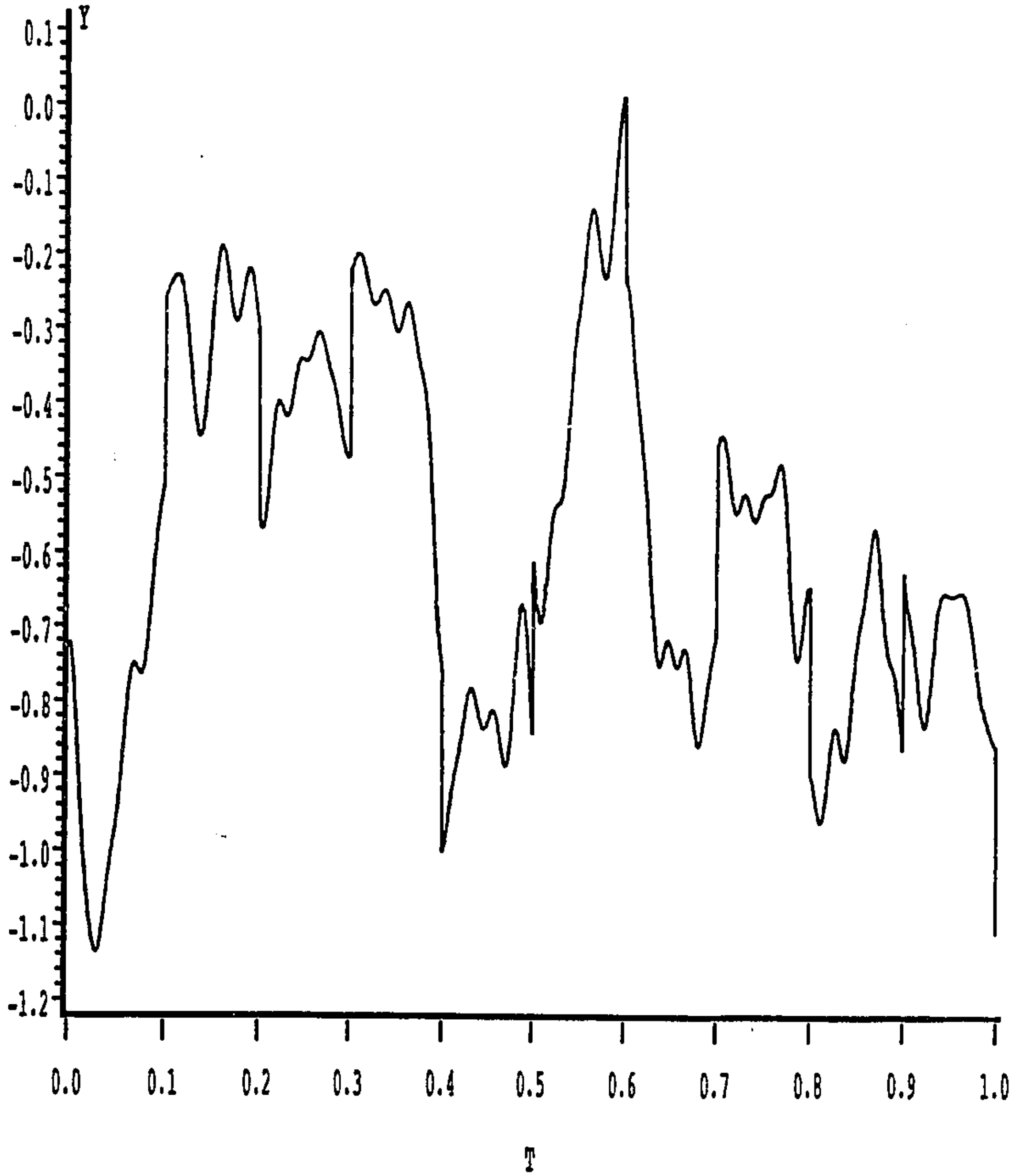
We consider two samples of size 50 each and simulate 60000 realizations and compute **Q**, SSE and their ratio. We do this for different values of γ defined by

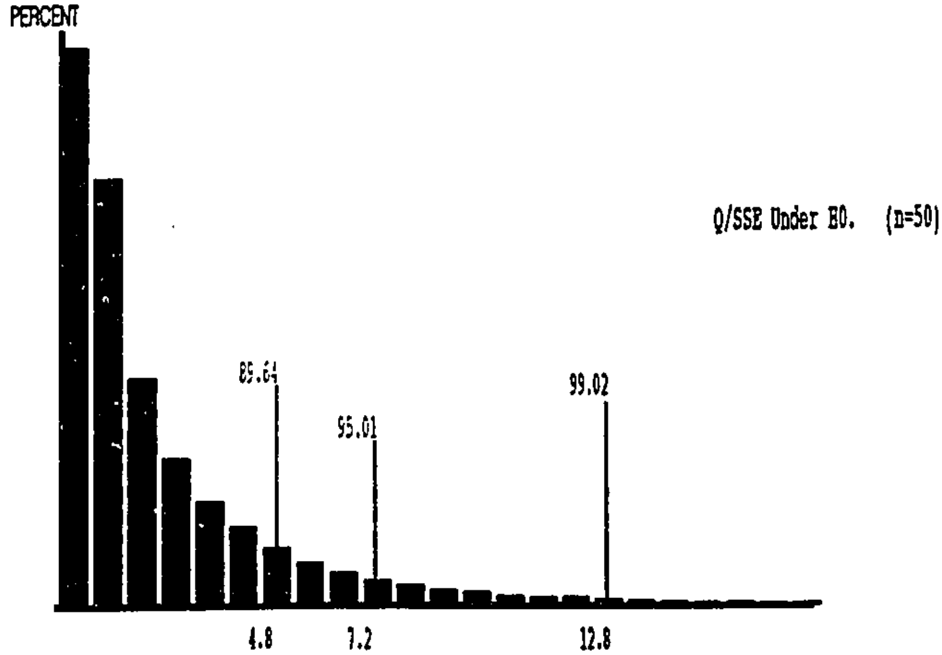
$$\forall t \in [0,1] \quad \beta_2(t) = \begin{cases} \gamma & \text{if } [5t] \text{ is odd} \\ 0 & \text{if } [5t] \text{ is even,} \end{cases}$$

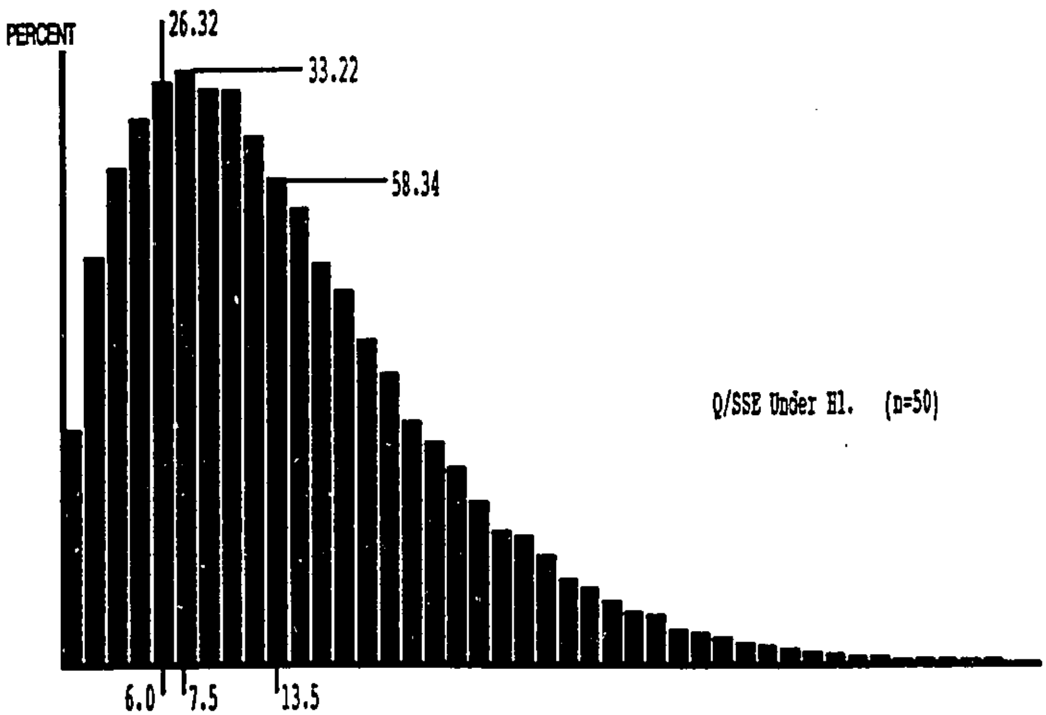
where $[x]$ is the integer part of x. We simulate the cases $\gamma = 0, 1/4, 1/2, 1/3$, and 1. The following table gives approximate critical values of the statistic $\mathbf{Q}/SSE$.

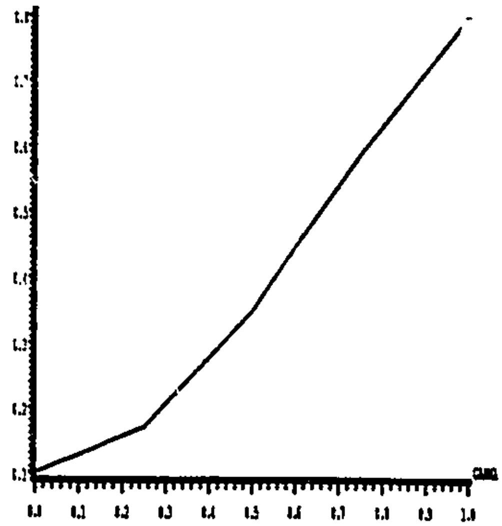
t	$\mathbf{P}\{\mathbf{Q}/\text{SSE} \leq \text{t}, H_0\}$
5.34	0.9004
5.36	0.9010
7.60	0.9499
7.62	0.9503
13.02	0.9899
13.06	0.9900

The following figures show respectively the graph of one observation for $\gamma = 1$, the histograms of our statistic **Q**/SSE under both $H_0 : \beta_1 = 0, \gamma = 0$ and $H_1 : \beta_1 = 0, \gamma = 1$, and the plots of the test power, as a function of γ, for the commonly used levels of significance. The last figure shows the histograms of **Q**/SSE for respectively $\gamma = 1/4, 1/2$ and $3/4$.

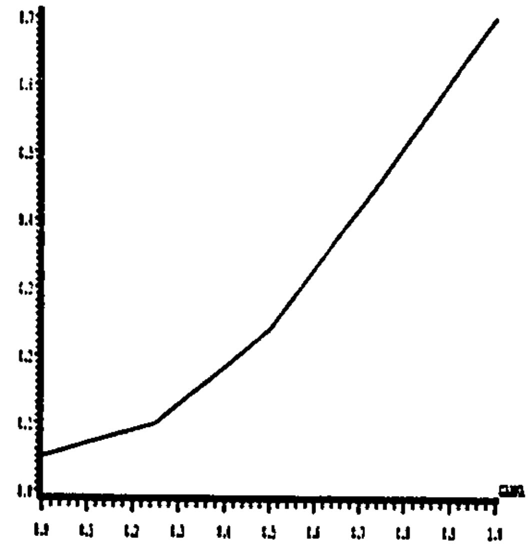

Y
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1.0
-1.1
-1.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
T


PERCENT
Q/SSE Under H0. (n=50)
89.64
95.01
99.02
4.8
7.2
12.8

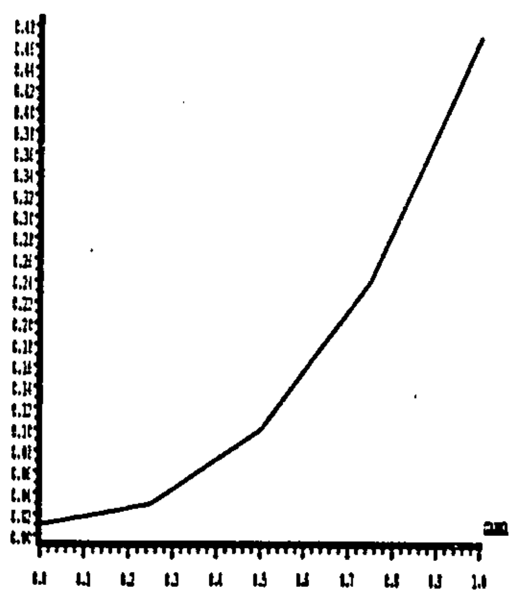

PERCENT
26.32
33.22
58.34
Q/SSE Under H1. (n=50)
6.0
7.5
13.5



(a)

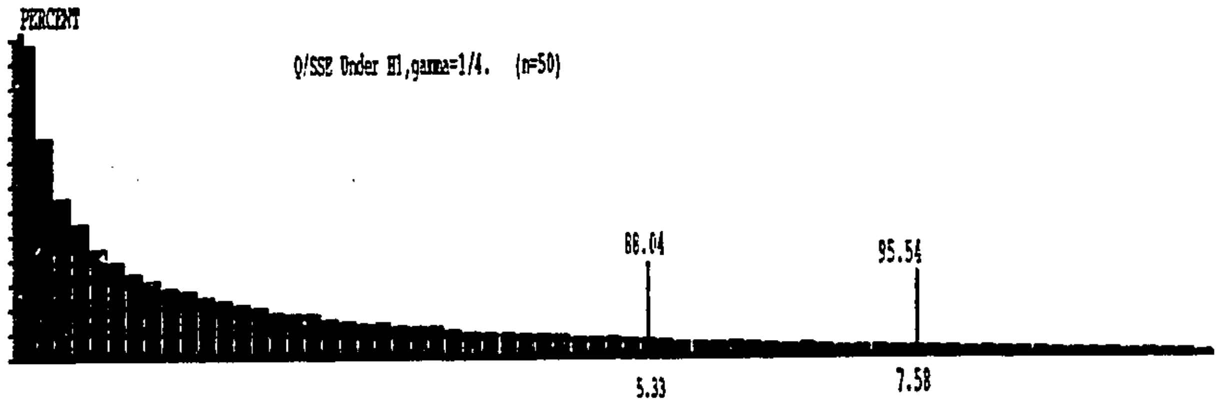


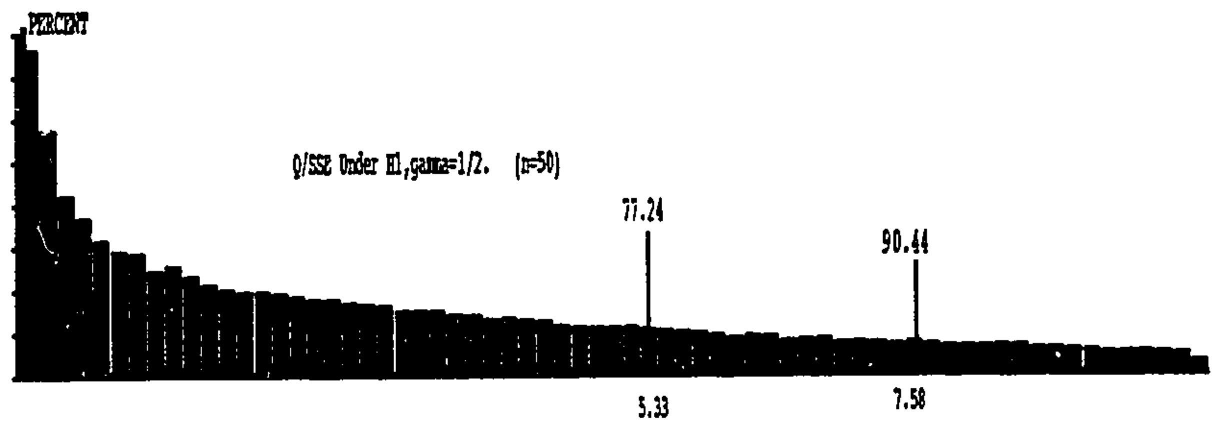
(b)

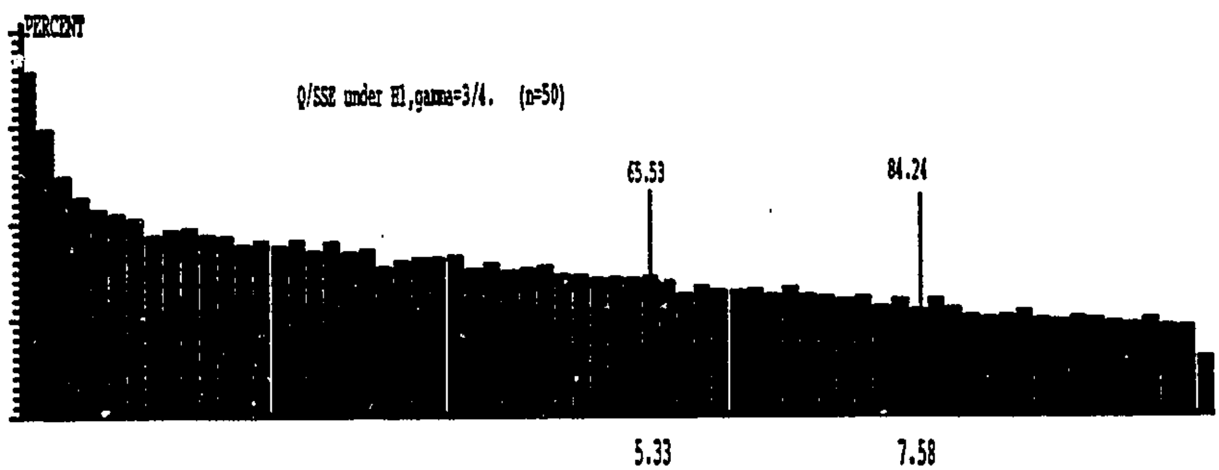


(c)

The graphs (a),(b), and (c) represent respectively the power of tests at levels 0.10, 0.05, and 0.01.


PERCENT
Q/SSE Under H1,gamma=1/4. (n=50)
88.04
95.54
5.33
7.58


PERCENT
Q/SSE Under H1,gamma=1/2. (n=50)
77.24
90.44
5.33
7.58


PERCENT
Q/SSE under H1,gamma=3/4. (n=50)
65.53
84.24
5.33
7.58

A.3 The Regression Example

Here we simulate the model

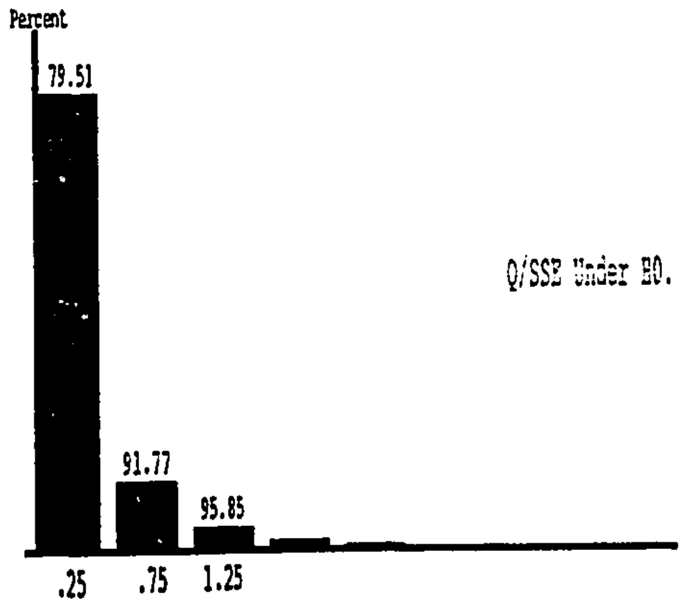
$$\mathbf{Y} = \mathrm{x}_6\beta_1 + \sigma\epsilon, \quad \text{with} \quad \beta_1 \equiv 3, \quad \mathrm{x}_6 = \frac{1}{\sqrt{28}} \begin{bmatrix} 1 \\ 2 \\ 3 \\ -1 \\ -2 \\ -3 \end{bmatrix},$$

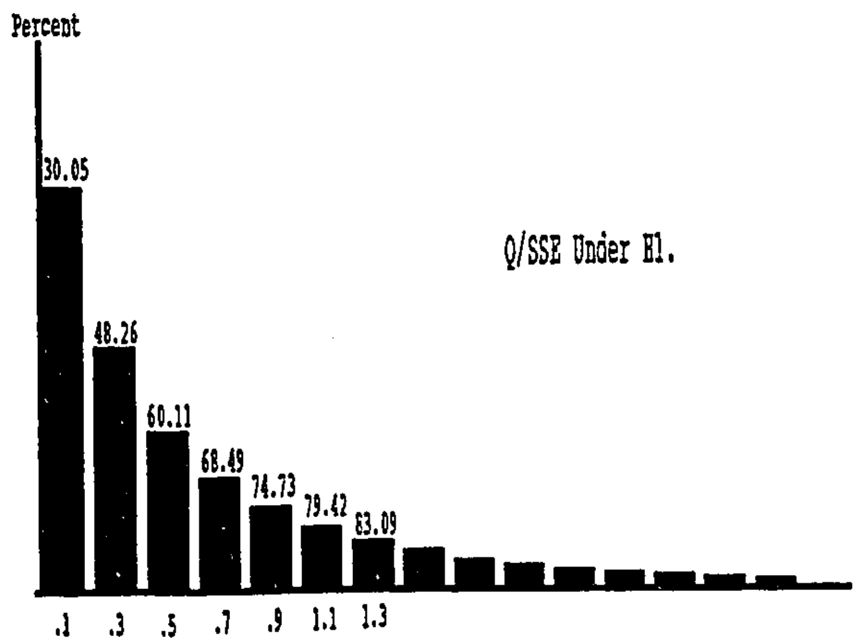
and ϵ is the Brownian motion. Since σ is irrelevant for our test statistic, we take $\sigma = 1$.

The following table gives approximate critical values for this example.

t	$\mathbf{P}\{\mathbf{Q}/\mathrm{SSE} \leq \mathrm{t}, H_0\}$
0.88	0.9003
1.36	0.9500
2.82	0.9899
2.84	0.9901

The following figure shows the histograms of $\mathbf{Q}/\mathrm{SSE}$ under H_0 and H_1.


Percent
79.51
91.77
95.85
Q/SSE Under H0.
.25
.75
1.25


Percent
30.05
48.26
60.11
68.49
74.73
79.42
83.09
Q/SSE Under H1.
.1
.3
.5
.7
.9
1.1
1.3

Appendix B

PROGRAMS

B.1 Main Subroutines

```
{**********************************************************}
var
  gasdevIset, Uniforminext, Uniforminextp: integer;
  idum,gasdevGset: real  ;
  Uniformma: array[1..55] of real ;

{**********************************************************}

function Uniform (var idum: real): real;
```

Returns a uniform random deviate between 0.0 and 1.0. Set idum to any negative value to initialize the sequence.

```
const
    mbig = 4.0e6;
    mseed = 1618033.0;
    mz = 0.0;
    fac = 2.5e-7;
var
   i, ii, k: integer;
   mj, mk: real ;

begin {MAIN}
 if idum < 0 then
 begin
   mj := mseed + idum;
   if mj > 0.0 then
    mj := mj - mbig * trunc(mj / mbig)
```

```
   else
    mj := mbig - abs(mj) + mbig * trunc(abs(mj / mbig));
   Uniformma[55] := mj;
   mk := 1;
   for i := 1 to 54 do
   begin
     ii := 21 * i mod 55;
     Uniformma[ii] := mk;
     mk := mj - mk;
     if mk < mz then
     mk := mk + mbig;
     mj := Uniformma[ii]
   end;
   for k := 1 to 4 do
   begin
     for i := 1 to 55 do
     begin
       Uniformma[i] := Uniformma[i] -
       Uniformma[1 + ((i + 30) mod 55)];
       if Uniformma[i] < mz then
Uniformma[i] := Uniformma[i] + mbig
      end
    end;
    Uniforminext := 0;
    Uniforminextp := 31;
    idum := 1
 end;
 Uniforminext := Uniforminext + 1;
 if Uniforminext = 56 then
 Uniforminext := 1;
 Uniforminextp := Uniforminextp + 1;
 if Uniforminextp = 56 then
 Uniforminextp := 1;
 mj := Uniformma[Uniforminext] - Uniformma[Uniforminextp];
 if mj < mz then
  mj := mj + mbig;
 Uniformma[Uniforminext] := mj;
 Uniform := mj * fac;
{ idum:=idum-0.01;WRITELN('IDUM  ',IDUM:4:4); }
end{MAIN};
{**********************************************************}
var
```

```
     GasdivIset:integer; {It has to be initialized at zero}
     GasdivGset:real;
    **********

function gasdev (var idum: real): real;
```

Returns a normally distributed deviate with mean zero and unit variance using the previous subroutine.

```
var
   fac, r, v1, v2: real;

begin{MAIN}
 if gasdevIset = 0 then
  begin
    repeat
    v1 := 2.0 * ran3(idum) - 1.0;
    v2 := 2.0 * ran3(idum) - 1.0;
    r := sqr(v1) + sqr(v2);
    until (r < 1.0) and (r > 0.0);
    fac := sqrt(-2.0 * ln(r) / r);
    gasdevGset := v1 * fac;
    gasdev := v2 * fac;
    gasdevIset := 1
  end
 else
  begin
    gasdevIset := 0;
    gasdev := gasdevGset;
    idum:=-2*idum;
  end;
end{MAIN};

{*********************************************************}

 PROCEDURE gaussj (var a:ArrayNPbyNP;var b:ArrayNPbyMP;n,m:integer);

          {SOLVES A*X=B, THE OUTPUTS ARE A:=A_INVERSEAND B:=SOLUTION}

  var
   big, dum, pivinv: real;
   i, icol, irow, j, k, l, ll: integer;
   indxc, indxr, ipiv: IntegerArrayNP;
```

```
            BEGIN{GAUSSJ}
for j := 1 to n do
 ipiv[j] := 0;
for i := 1 to n do
 begin
  big := 0.0;
  for j := 1 to n do
   if ipiv[j] <> 1 then
    for k := 1 to n do
     if ipiv[k] = 0 then
     if abs(a[j, k]) >= big then
     begin
     big := abs(a[j, k]);
     irow := j;
     icol := k
     end
     else if ipiv[k] > 1 then
     begin
     writeln('pause 1 in GAUSSJ - singular matrix');
     readln
     end;
  ipiv[icol] := ipiv[icol] + 1;
  if irow <> icol then
   begin
    for l := 1 to n do
     begin
     dum := a[irow, l];
     a[irow, l] := a[icol, l];
     a[icol, l] := dum
     end;
    for l := 1 to m do
     begin
     dum := b[irow, l];
     b[irow, l] := b[icol, l];
     b[icol, l] := dum
     end
   end;
  indxr[i] := irow;
  indxc[i] := icol;
  if a[icol, icol] = 0.0 then
   begin
```

```
      writeln('pause 2 in GAUSSJ - singular matrix');
      readln
     end;
    pivinv := 1.0 / a[icol, icol];
    a[icol, icol] := 1.0;
    for l := 1 to n do
     a[icol, l] := a[icol, l] * pivinv;
    for l := 1 to m do
     b[icol, l] := b[icol, l] * pivinv;
    for ll := 1 to n do
     if ll <> icol then
      begin
       dum := a[ll, icol];
       a[ll, icol] := 0.0;
       for l := 1 to n do
       a[ll, l] := a[ll, l] - a[icol, l] * dum;
       for l := 1 to m do
       b[ll, l] := b[ll, l] - b[icol, l] * dum
      end
   end;
  for l := n downto 1 do
   if indxr[l] <> indxc[l] then
    for k := 1 to n do
     begin
      dum := a[k, indxr[l]];
      a[k, indxr[l]] := a[k, indxc[l]];
      a[k, indxc[l]] := dum;
     end;
                        END{GAUSSJ};
{*********************************************************}
```

The previous subroutines can be found in any standard textbook on computation programming like Numerical Recipes In Pascal by W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling. Cambridge University Press.

```
{*********************************************************}
PROCEDURE GAUSSIAN(var GAUS:real);

BEGIN
     GASDEVISET:=0;
     GAUS:=GASDEV(IDUM);
END;
```

```
{*********************************************************}

FUNCTION CHI(DEG:integer):REAL;
```

Returns central Chi squared deviates with DEG degrees of freedom

```
  var
     l,rn: integer;
     S,Gaus:real;

BEGIN{CHI}
    rn:=DEG;
    S:=0;
    for l:=1 to DEG do
      begin
GAUSSIAN(Gaus);
S:=S+SQR(Gaus);
      end;
     CHI:= S;
END{CHI};
{*********************************************************}
FUNCTION CHI1(DEG:integer; meank:real):REAL;
```

Returns non-central Chi squared deviates with "DEG" degrees of freedom, meank is the non-centrality parameter.

```
var
   l,rn: integer;
   S:real;
   Gaus:real;

BEGIN{CHI}
      rn:=DEG;
      S:=0;
      for l:=1 to DEG do
begin
  GAUSSIAN(Gaus);
  S:=S+SQR(Gaus+meank);
end;
       CHI1:= S;
END{CHI};
{*********************************************************}
function TRAPZD (a,b: real;var vect: Array1PbyMP): real;
```

ESTIMATES THE INTEGRAL BETWEEN a AND b FOR THE GIVEN VALUES:VECT

```
var
  i,subdivsize: integer;
  surf: real;

BEGIN{TRAPZD}

    surf:=(b-a)*vect[1]/(subdivsize);
    for i := 2 to subdivsize do
      surf := surf+(b - a) * (vect[i]+vect[i - 1]) / (2 * subdivsize);
    trapzd :=surf;
END{TRAPZD};
{*********************************************************}
```

B.2 Examples On Power: A Simulated Example

```
program MANOVAstatistic;

const

  simulsize= ;    rand= ;
  np= ;                    {THE MAXIMUM SAMPLE SIZE}
  mp= ;                     {THE MAXIMUM SUBDIVISION SIZE}

type
  Array1PbyMP = array[0..mp] of real;
  ArrayNPbyNP = array[1..np, 1..np] of real;
  ArrayNPbyMP = array[1..np, 1..mp] of real;
  ArrayMPbyMP = array[1..mp, 1..mp] of real;
  ArrayMPbyNP = array[1..mp, 1..np] of real;
  integerArrayNP = array[1..np] of integer;
{USED BY PROCEDURE GAUSSJ}

var
   j,i,k,count              :integer;
   Frat0,Frat1              :text;
   a,b,theta,pi             :real;
   gasdevIset, ran3inext, ran3inextp: integer;
   idum,gasdevGset,GAUS     :real  ;
```

```
    ran3ma                    :array[1..55] of real ;
    CM_inverse                :ArrayNPbyMP;
    CM                        :ArrayNPbyNP;
    Y                         :ArrayNPbyMP;
    t                         :ARRAY1PbyMP;

{*******************************************************}
  PROCEDURE erreur(var err:ARRAY1PbyMP);

       {err=(e(t1),...,e(tm))'}
  var
   X:real;
   j:integer;
  begin
    err[0]:=0;
    for j:=1 to mp do
      begin
GAUSSIAN(X);
err[j]:=t[j-1]*err[j-1]+(t[j]-t[j-1])*X;
      end;
  end;
{*******************************************************}
  PROCEDURE Obs0(var Yvect:ARRAY1PbyMP);

       {Yvect=error, i.e. mean=0}

  var
    j:integer;
    eps:ARRAY1PbyMP;

  begin
  Yvect[0]:=0;
  erreur(eps);
  for j:=1 to mp do
  Yvect[j]:=eps[j];
  end;
{*******************************************************}
  PROCEDURE Obs1(var Yvect:ARRAY1PbyMP);
      {Yvect=mean +error}
  var
    j:integer;
    eps:ARRAY1PbyMP;
```

```
  begin
  Yvect[0]:=0;
  erreur(eps);
  for j:=1 to mp do
  Yvect[j]:=theta*t[j]+eps[j];
  end;
{*****************************************************}
 PROCEDURE Ymat0(var ZY:ARRAYNPbyMP);

   {Ymat=matrix of observations under H0:mean=0}

  var
    i,j:integer;
    Y  :ARRAY1PbyMP;

  begin
    for i:=1 to np do
      begin
Obs0(Y);
for j:=1 to mp do
  ZY[i,j]:=Y[j];
      end;
  end;
{*****************************************************}
 PROCEDURE Ymat1(var ZY:ARRAYNPbyMP);

   {Ymat=matrix of observations under
       H1: mean=sqrt(2)*theta*sin(3pit/2)}
   var
    i,j:integer;
    Y  :ARRAY1PbyMP;

  begin
    for i:=1 to np do
      begin
Obs1(Y);
for j:=1 to mp do
  ZY[i,j]:=Y[j];
      end;
  end;
{*****************************************************}
```

```
PROCEDURE Trace(var Z:arrayNPbyMP;var a:real;var b:real);

 var
   c       :real;
   i,j,k   :integer;
   Zdot    :Array1PbyMP;
   ZtZ,ZtBZ:ArrayMPbyNP;          {t for Transpose}



 begin
   for i:=1 to mp do           {computes Z'Z}
    for j:=1 to mp do
     begin
ZtZ[i,j]:=0;
for k:=1 to np do
 ZtZ[i,j]:=ZtZ[i,j]+Z[k,i]*Z[k,j];
     end;


   Zdot[0]:=0;
   for i:=1 to mp do           {Zdot is a row of BZ since }
    begin                      {  B=1/np*(1',..,1')       }
      Zdot[i]:=0;
      for k:=1 to np do
 Zdot[i]:=Zdot[i]+Z[k,i]/np;
    end;

    for i:=1 to mp do              {Computes the entries of Z'BZ}
     for j:=1 to mp do
  ZtBZ[i,j]:=Zdot[i]*Zdot[j];


    a:=0;                        {Computes trace(CM_inverse Z'Z)}
    for i:=1 to mp do
      for k:=1 to mp do
 a:=a+CM_inverse[i,k]*ZtZ[k,i];


    b:=0;                        {Computes trace(CM_inverse Z'MZ)}
    for i:=1 to mp do            { =trace(CM_inverse[Z'Z-ZBZ])    }
      for k:=1 to mp do
```

```
b:=b+CM_inverse[i,k]*(ZtBZ[k,i]);
    b:=a-b
   end;

{*****************************************************}

                     BEGIN{MAIN PROGRAM}
    pi:=4*arctan(1);
    idum:=-rand;
    t[0]:=0;
    for i:=1 to mp do
      t[i]:=sqrt(2)*sin(3*pi*i/(2*mp));


    for i:=1 to mp do                   {THIS IS THE COV. MATRIX}
      for j:=1 to mp do
begin
  if i=j then
  begin
  CM[i,j]:=1;
  CM_inverse[i,j]:=1;
  end
  else
  begin
  if (i=j+1) or (i=j-1) then CM[i,j]:=-1
  else
  CM[i,j]:=0;
  CM_inverse[i,j]:=0;
  end;
{CM_inverse=Idmp, for computational raisons}
end;

    gaussj(CM,CM_inverse,mp,mp);

{gaussj(CM,CM_inverse,mp,mp) Computes CM_inverse, and calls it CM}

    for i:=1 to mp do          {thus we need this}
      for j:=1 to mp do
begin
CM_inverse[i,j]:=CM[i,j];
end;
```

```
      rewrite(Frat1);
       for i:=1 to simulsize do
begin
  for count:= 0 to 5 do
  begin
   theta:=count/10;
   Ymat1(Y);
   trace(Y,a,b);
   write(Frat1,a/b:10:6);
  end;
 writeln(Frat1);
end;
      close(Frat1);


      END{MAIN PROGRAM}.
{************************************************************}
 PROGRAM Qstatistic;

  const
    simulsize= ;    samsize=  ; rand= ;
    subdivsize= ;sumsize= ;
    mp=subdivsize;DFQ=1;DFSSE=samsize-2;
    {DFQ=rank(D) & DFSEE=rank(M)}
  type
   LAMBVECT=array[1..sumsize] of real;
   OBSVECT =array[1..simulsize] of real;
   GAUSVECT=array[1..dfsse] of shortreal;

   Array1PbyMP = array[0..mp] of real;
  var
    k,i,DQ,DS,COUNT:integer;
    val0,val1: text;
    pi,t,Q0,Q1,theta:real;
    X,BETA:LAMBVECT;
    RATIO:OBSVECT;
    LAMB2:LAMBVECT;
{*********************************************************}

  FUNCTION CHI(DEG:integer):REAL;
```

```
{IT GENERATES A CHI-SQUARE(DEG) DEVIATE }

var
 l,rn: integer;
 S:real;
 Gaus:real;

BEGIN{CHI}
    rn:=DEG;
    S:=0;
    for l:=1 to DEG do
    begin
    GAUSSIAN(Gaus);
    S:=S+SQR(Gaus);
    end;
    CHI:= S;
END{CHI};

{*********************************************************}

FUNCTION CHI1(DEG:integer; meank:real):REAL;

{IT GENERATES A CHI-SQUARE(DEG) DEVIATE }

var
 l,rn: integer;
 S:real;
 Gaus:real;

BEGIN{CHI}
    rn:=DEG;
    S:=0;
    for l:=1 to DEG do
    begin
    GAUSSIAN(Gaus);
    S:=S+SQR(Gaus+meank);
    end;
    CHI1:= S;
END{CHI};

{*****************************************************************}
```

```
  PROCEDURE QSTAT (DEG:integer;VAR Q0:REAL);

   {QSTAT IS AN INFINITE SUM OF WEIGHTED CHI-SQUARES WITH n=DEG}
   {DEGREES OF FREEDOM}
  var
     S0,S1:real;
     k:integer;

  BEGIN{}
     S0:=0;
     for k:=1 to sumsize do
       S0:=S0+LAMB2[k]*CHI(DEG);
     Q0:=S0;
  END{};
{****************************************************************}

  PROCEDURE QSTAT1 (DEG:integer;VAR  Q1:REAL);

   {QSTAT IS AN INFINITE SUM OF WEIGHTED CHI-SQUARES WITH n=DEG}
   {DEGREES OF FREEDOM}
  var
     S1:real;
     k :integer;

  BEGIN{}
     S1:=0;
     for k:=1 to sumsize do
       S1:=S1+LAMB2[k]*CHI1(DEG,theta*beta[k]);
     Q1:=S1;
  END{};

{****************************************************************}
  function TRAPZD (a,b: real;var vect: Array1PbyMP): real;

    {ESTIMATES THE INTEGRAL BETWEEN a AND b FOR THE GIVEN VALUES:VECT}


    var
     i: integer;
      surf: real;

      BEGIN{FUNCTION TRAPZD}
```

```
    surf:=(b-a)*vect[1]/(subdivsize);
    for i := 2 to subdivsize do
      surf := surf+(b - a) * (vect[i]+vect[i - 1]) / (2 * subdivsize);
    trapzd :=surf;
       END{FUNCTION TRAPZD};
{***************************************************************}
   PROCEDURE MEANH1(var beta:LAMBVECT);

   var
     k,l:integer;
     a,b:real;
     mean,vect1:array1PbyMP;

   BEGIN{MEANH1}
     a:=0;b:=1;
     for l:=1 to subdivsize do
       mean[l]:=theta*l/subdivsize;
     for k:=1 to sumsize do
     begin
     for l:=1 to subdivsize do
     vect1[l]:=sin((2*k+1)*pi*l/(2*subdivsize))*mean[l];
     beta[k]:=sqrt(2)*trapzd(a,b,vect1);
     end;
  END{MEANH1};
{************************************************************}

  BEGIN{MAIN PROGRAM}

    rewrite(val0);rewrite(val1);
    pi:=4*arctan(1);
    idum:=-abs(rand);



    FOR k:=1 TO SUMSIZE DO
    begin
      LAMB2[k]:=1/SQR(SQR((2*k+1)*pi/2));
      beta[k] :=sqrt(2)*cos(k*pi)*sqr(2/((2*k+1)*pi));
     { writeln(val1,k,'    ',beta[k]);
      writeln(val0,k,'   ',LAMB2[k]); }
    end;
```

```
    DQ:=DFQ;COUNT:=1;

    for k:=1 to simulsize do
    begin
      QSTAT(DQ,Q0);
      writeln(val0,Q0:10:4);
      for count:=1 to 5 do
begin
  theta:=count/10;
  QSTAT1(DQ,Q1);
  write(val1,Q1:10:4);
end;
writeln(val1);
    end;
  close(val0);close(val1);
  END{MAIN PROGRAM}.
{***********************************************************}
```

B.3 The Equality of Distributions Example

```
program EMPIRICAL;

const
    samsize=10;
    np=samsize;                    {THE MAXIMUM SAMPLE SIZE}
    subdivsize=500;
    mp=SUBDIVSIZE;           {THE MAXIMUM SUBDIVISION SIZE}
type
    ArrayNPbyMP = array[1..np, 1..mp] of real;
    Array1PbyMP = array[1..mp] of real;

var

   'filename1','filename2','filename3':text;
   rn,i,j,k:integer;
   boundinf, boundsup: real;        {INTERVAL BOUNDS}
   rand,a1,a,S,Q: real;
{***********************************************************}
procedure data( var ZS: Array1PbyMP);
```

This procedure generates data and returns a vector ZS of values of $Y_{1.} - Y_{2.}$

```
var
```

```
  i,j,k:integer;
  a:real;
  W1,W2,Y1,Y2,ZS2: Array1PbyMP;

BEGIN{DATA}
  a:=0;
  for i:=1 to samsize do
   begin
       W1[i]:=ran3(idum);
       W2[i]:=ran3(idum);
   end;
  for j:=1 to subdivsize do
   Begin
       Y1[j]:=0;Y2[j]:=0;
       for i:=1 to samsize do
 begin
     if Sqr(W2[i])<=j/subdivsize then
     Y2[j]:=1+Y2[j]
     else
     Y2[j]:=Y2[j];
     if W1[i]<=j/subdivsize then
     Y1[j]:=1+Y1[j]
     else  Y1[j]:=Y1[j];
 end;
 ZS[j]:=(Y1[j]-Y2[j])/samsize;
 if abs(ZS[j])>a then a:=abs(ZS[j]);
 ZS2[J]:=SQR(ZS[J]);
   End;
   WRITELN('filename2',a:6:6);
   WRITELN('filename3',SAMSIZE*TRAPZD(BOUNDINF,BOUNDSUP,ZS2):6:6);

END{DATA};
{**********************************************************}

function KERNEL (s, t: real): real;

begin
    if s<=t then kernel:=s*(1-t) else kernel:=t*(1-s);
       {THIS THE COVARIANCE FUNCTION OF THE ERROR}
end;
{**********************************************************}
procedure norm(var nor:real);
```

This procedure returns the T*-norm of the "vector" ZS, computed before.*

```
var
   k,l: integer;
   ZS,vect1,vect2:Array1PbyMP ;

BEGIN{NORM}
   data(ZS);
   for k:=1 to subdivsize do
   begin
for l:=1 to subdivsize do
vect1[l]:=kernel(k/subdivsize,l/subdivsize)*ZS[l];
vect2[k]:=trapzd(boundinf,boundsup,vect1)*ZS[k];
   end;
   nor:=trapzd(boundinf,boundsup,vect2);
END{NORM};
{**********************************************************}
BEGIN{MAIN PROGRAM}
      IDUM:=-0.1;
      rewrite('filename1');
      rewrite('filename2');
      rewrite('filename3');
      boundinf:=0;
      boundsup:=1;
      rn:=samsize;
      for k:=1 to 1000 do
 BEGIN
 norm(S);
 Q:=S*samsize;
 writeln('filename1',Q:6:6);
 END;
      CLOSE('filename1'); CLOSE('filename2'); CLOSE('filename3');
END{MAIN PROGRAM}.
```

B.4 The Geological Example

```
PROGRAM Geol;

const
    simulsize=20000;    samsize=50 ; rand=0.50;
    subdivsize=1000;sumsize=100;
    mp=subdivsize;DFQ=1;DFSSE=samsize-2;
```

```
   {DFQ=rank(D) & DFSEE=rank(M)}
type
   LAMBVECT=array[0..sumsize] of real;
   OBSVECT =array[1..simulsize] of real;
   GAUSVECT=array[1..dfsse] of shortreal;

   Array1PbyMP = array[0..mp] of real;
var
    k,i,DQ,DS,COUNT,rankM:integer;
    gval50e,file2,fg1,fg2,fg3: text;
    pi,t,Q0,Q1,SSE,scale:real;
    BETA:LAMBVECT;
    RATIO:OBSVECT;
    LAMB2:LAMBVECT;
{**********************************************************}
PROCEDURE SSTAT(DEG:INTEGER;VAR SSE:REAL);

   {SSE IS AN INFINITE SUM OF WEIGHTED CHI-SQUARES WITH n=DEG}
  {DEGREES OF FREEDOM}

var
    S:real;
    k,i:integer;

BEGIN{SSE}
    S:=LAMB2[0]*CHI(DEG);
    for k:=1 to sumsize do
      S:=S+LAMB2[k]*CHI(DEG);
    SSE:=S;
END{SSE};
{**********************************************************}

PROCEDURE QSTAT (DEG:integer;VAR Q0,Q1:REAL);
```

QSTAT IS AN INFINITE SUM OF WEIGHTED CHI-SQUARES WITH n=DEG DEGREES OF FREEDOM

```
var
    S0,S1:real;
    k:integer;

BEGIN{SSE}
```

```
    S0:=LAMB2[0]*CHI(DEG);
    S1:=LAMB2[0]*CHI1(DEG,beta[0]);
    for k:=1 to sumsize do
    begin
S0:=S0+LAMB2[k]*CHI(DEG);
S1:=S1+LAMB2[k]*CHI1(DEG,beta[k]);
    end;
    Q0:=S0;Q1:=S1;
END{SSE};
{**********************************************************}
PROCEDURE MEANH1(var beta:LAMBVECT);

var
    k,l:integer;
    a,b:real;
    mean,vect1:array1PbyMP;

BEGIN{MEANH1}
    a:=0;b:=1;
    for l:=1 to subdivsize do
    begin
    if  odd(l div 100) then
       mean[l]:=sqrt(samsize/2)*scale
    else
       mean[l]:=0;
    end;
    beta[0]:=trapzd(a,b,mean);
    for k:=1 to sumsize do
    begin
    for l:=1 to subdivsize do
    vect1[l]:=cos(k*pi*l/subdivsize)*mean[l];
    vect1[k]:=trapzd(a,b,vect1);
    end;
END{MEANH1};
{**********************************************************}
BEGIN{MAIN PROGRAM}

    rewrite(gval50e);
    pi:=4*arctan(1);rankM:=2*(2*samsize-2);
    LAMB2[0]:=SQR(1-1/EXP(1));
    idum:=-abs(rand);
    MEANH1(BETA);
```

```
    FOR k:=1 TO SUMSIZE DO
    begin
      IF ODD(k) THEN
      LAMB2[k]:=SQR((1+1/EXP(1))/(1+SQR(k*pi)))
      ELSE
      LAMB2[k]:=SQR((1-1/EXP(1))/(1+SQR(k*pi)));
    end;
    DQ:=DFQ;DS:=DFSSE;COUNT:=1;

    for k:=1 to simulsize do
    begin
    scale:=0.25;
    MEANH1(BETA);
    {THE RATIO Q/SSE=(Q/n)/(SSE/n)=2*samsize*QSTAT/SSE}
      QSTAT(DQ,Q0,Q1);SSTAT(DS,SSE);
      if SSE <> 0 THEN
      write(gval50e,rankM*Q0/SSE:10:4, rankM*Q1/SSE:10:4);
    for i:=2 to 4 do
    begin
    scale:=i/4;
    MEANH1(BETA);
    {THE RATIO Q/SSE=(Q/n)/(SSE/n)=2*samsize*QSTAT/SSE}
      QSTAT(DQ,Q0,Q1);SSTAT(DS,SSE);
      if SSE <> 0 THEN
      write(gval50e, rankM*Q1/SSE:10:4);
    end;
    writeln(gval50e);
    end;
  close(gval50e);
END{MAIN PROGRAM}.
```

B.5 The Regression Example

The program used here is essentially the same as the previous one. One has only to choose the samsize (the sample size), and the eigenvalues (squared) LAMB2[k]. We omit the details.

Appendix C

Proof of Lemma 4.2.1

In the following, we show that $(S_n, \mathcal{F}_n)_{n\geq 1}$ is an $\mathbf{L}_2$-bounded martingale, i.e., $S_n = \mathbf{E}[S_{n+1}/\mathcal{F}_n]$, and there exists a constant $C > 0$ such that $\mathbf{E}[S_n^2] < C$, for every $n \geq 1$.

From the definition of S_n, we have that

$$\begin{aligned} S_n &= \sum_{1\leq i\neq j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T \\ &= 2 \sum_{1\leq i<j\leq n} b_{ij} < \epsilon_i, \epsilon_j >_T \\ &= 2 \sum_{1\leq j\leq n} < A_j, \epsilon_j >_T, \end{aligned}$$

where $A_1 = 0$ and

$$A_j = \sum_{i=1}^{j-1} b_{ij}\epsilon_i.$$

Using the independence of the ϵ_i, straightforward calculations show that $(S_n, \mathcal{F}_n)$ is a martingale. To prove that

$$\sup_{n\geq 1} \mathbf{E}[S_n^2] < C = 4\,\mathbf{E}\left[< \epsilon_1, \epsilon_2 >_T^2\right] trace(\mathbf{B}),$$

note that

$$S_n^2 = 4\left[\sum_{j=1}^{n} < A_j, \epsilon_j >_T^2 + \sum_{1\leq j<k\neq n} < A_j, \epsilon_j >_T < A_k, \epsilon_k >_T\right].$$

By independence we have that, for $j \neq k$,

$$\mathbf{E}[< A_j, \epsilon_j >_T < A_k, \epsilon_k >_T] = 0.$$

Indeed, suppose that $j < k$, then

$$\begin{aligned}
\mathbf{E} & \left[< A_j, \epsilon_j >_T < A_k, \epsilon_k >_T\right] \\
&= \mathbf{E}\left[< A_j, \epsilon_j >_T \int_{[0,1]^2} \kappa(s,t)\, A_k(s)\, \epsilon_k(t)\, ds\, dt\right] \\
&= \int_{[0,1]^2} \kappa(s,t)\, \mathbf{E}\left[< A_j, \epsilon_j >_T \ A_k(s)\right] \mathbf{E}\left[\epsilon_k(t)\right] \ ds\, dt \\
&= 0,
\end{aligned}$$

since $< A_j, \epsilon_j >_T \ A_k$ and ϵ_k are independent, and $\mathbf{E}\left[\epsilon_k(t)\right] = 0$, for all $t \in [0,1]$. This shows that

$$\begin{aligned}
\mathbf{E}[S_n^2] &= 4\,\mathbf{E}\left[\sum_{j=1}^{n} < A_j, \epsilon_j >_T^2\right] \\
&= 4\sum_{j=1}^{n}\left\{\mathbf{E}\left[\sum_{i=1}^{j-1} b_{ij} < \epsilon_i, \epsilon_j >_T\right]^2\right\} \\
&= 4\sum_{j=1}^{n}\left\{\mathbf{E}\left[\sum_{i=1}^{j-1} b_{ij}^2 < \epsilon_i, \epsilon_j >_T^2\right]\right\} + \\
&\qquad 4\sum_{j=1}^{n}\left\{\mathbf{E}\left[\sum_{1\le i\neq k\le j-1} b_{ij}b_{kj} < \epsilon_i, \epsilon_j >_T < \epsilon_k, \epsilon_j >_T\right]\right\}.
\end{aligned}$$

By independence again, the last term is null, and thus

$$\begin{aligned}
\mathbf{E}[S_n^2] &= 4\sum_{j=1}^{n}\left\{\mathbf{E}\left[\sum_{i=1}^{j-1} b_{ij}^2 < \epsilon_i, \epsilon_j >_T^2\right]\right\} \\
&= 4\,\mathbf{E}\left[< \epsilon_1, \epsilon_2 >_T^2\right] \sum_{j=1}^{n}\sum_{i=1}^{j-1} b_{ij}^2 \\
&\le 4\,\mathbf{E}\left[< \epsilon_1, \epsilon_2 >_T^2\right] \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ij}^2
\end{aligned}$$

Recall that $\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X} = (b_{ij})_{1\le i,j\le n}$, is symmetric and idempotent, i.e., $b_{ij} = b_{ji}$ and $\mathbf{BB}$=$\mathbf{B}$. Therefore

$$\sum_{i=1}^{j-1} b_{ij}^2 = b_{jj}, \quad \forall j = 1, \ldots, n.$$

Moreover trace($\mathbf{B}$)= $\sum_{i=1}^{n} b_{ii} = p < \infty$, thus

$$\mathbf{E}[S_n^2] < \mathbf{E}\left[< \epsilon_1, \epsilon_2 >_T^2\right] p < \infty.$$

□

This completes the proof.

Bibliography

[1] J.B. CONWAY (1989): A Course in Functional Analysis. Second Edition. Springer-Verlag.

[2] N. CRESSIE (1990): The Origins of Kriging. Mathematical Geology, **22**, pp. 239-252.

[3] M.CSÖRGŐ AND P.RÉVÉSZ (1981): Strong Approximations in Probability and Statistics. Academic Press.

[4] R.M. DUDLEY (1968): Distances of Probability Measures and Random Variables. Ann. Math. Stat. **39**, pp. 1563-1572.

[5] W. FELLER (1971): An Introduction to Probability Theory and Its Applications. Volume II, Second Edition. John Wiley & Sons,Inc.

[6] X. FERNIQUE (1985): Gaussian Random Vectors Their Reproducing Kernel and Hilbert Spaces. Technical Report Series of the Laboratory for Research in Statistics and Probability, No. 34. University of Ottawa and Carleton University.

[7] U. GRENANDER (1981): Abstract Inference. John Wiley & Sons, Inc.

[8] S. KAKUTANI (1948)): On Equivalence of Infinite Product Measures. Ann. Math. **49**, pp. 214-224.

[9] D.G. KRIGE (1951): A Statistical Approach to Some Basic Mine Problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, **52**, pp. 119-139.

[10] H. KUO (1970): Gaussian Measures in Banach Spaces. Lecture Notes in Mathematics **463**. Springer-Verlag.

[11] E.L. LEHMANN (1986): Testing Statistical Hypotheses. Second Edition. John Wiley & Sons, Inc.

[12] G. MATHERON (1963): Principles of Geostatistics. Economic Geology, **58**, pp. 1246-1266.

[13] G. MATHERON (1971): The Theory of Regionalized Variables and Its Applications. Cahiers du centre de Morphologie Mathématique, **No. 5**. Fontainebleau, France.

[14] E. PARZEN (1989): An Approach to Time Series Analysis. Ann. Math. Stat., **32**, pp. 951-989.

[15] H. SCHEFFÉ (1959): The Analysis of Variance. John Wiley & Sons, Inc.

[16] S.R. SEARLE (1971): Linear Models. John Wiley & Sons, Inc.

[17] A. SEN AND M. SRIVASTAVA (1990): Regession Analysis. Theory, Methods, and Applications. Springer-Verlag.

[18] A.V SKOROKHOD (1956): Limit Theorems for Stochastic Processes. Theor. Prob. Appl., **1**, pp. 261-290.

[19] M. SRIVASTAVA (1971): On Fixed Width Confidence bounds for Regression Parameters. Ann. Math. Stat. **42**, pp. 1403-1411.

[20] S.R.S. VARADHAN (1962): Limits Theorems for Sums of Independent Random Variables with Values in a Hilbert Space. Sankhyà A 24, pp. 213-238.