On von Mises Functionals with Emphasis on Trace Class Kernels

by

José M. González-Barrios


Submitted to the
Department of Mathematics
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May, 1990

© Massachusetts Institute of Technology, 1990
All rights reserved.

Signature of Author

Department of Mathematics
May, 1990

Certified by

Richard M. Dudley
Professor of Mathematics
Thesis Supervisor

Accepted by

Sigurdur Helgason, Chairman
Departmental Graduate Committee
Department of Mathematics

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
AUG 23 1990
LIBRARIES
ARCHIVES
On von Mises Functionals with Emphasis

on Trace Class Kernels.

by

José M. González-Barrios

Submitted to the Department of Mathematics on May, 1990

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Abstract

Consider a measurable space \((X, \mathcal{A})\) and a function \(K : X^2 \to \mathbb{R}\) and \(\mu, \nu\) two finite signed measures on \((X, \mathcal{A})\). Define an integral operator \(T_K\) with kernel \(K\) by:

\[
T_K(\mu, \nu) = \iint K(x, y) d\mu(x) d\nu(y).
\]

We are interested in the particular case in which \(\mu\) and \(\nu\) are differences of probability measures. We obtain results on the behavior of \(T_K\), using metric entropy conditions, in the case in which \(K\) is a trace class kernel. We also introduce some results on trace class operators.

If we let \(T(P) := T(P, P)\), for a probability measure \(P\), defined on \(I = [0, 1]\), and we consider \(P_n\), the empirical measure of a sample of size \(n\), we give asymptotic distributions of \(T(P_n)\) under relatively mild conditions.
Acknowledgements

To my thesis adviser Professor Richard M. Dudley, I express my respect and deep appreciation. Professor Dudley introduced me to many interesting topics in Probability Theory. This work would not have been possible without his useful suggestions and patient guidance. I consider myself very fortunate for having had the opportunity to work with Professor Dudley. Thank you very much.

I would like to thank the Mexican government and specifically its agency CONACYT, for giving me the opportunity of studying at M.I.T.

I want to thank Regina Pereira de Souza for her helpful advices for the printing of this thesis.

Finally, I would like to thank all my friends in Boston for all the support they provided.
“A mis padres con mucho cariño”
## Contents

1 Operators ......................................................... 8  
   1.1 Review ......................................................... 8  
   1.2 Schmidt Class Operators ..................................... 13  
   1.3 Integral Operators ............................................ 16  
   1.4 Trace Class Operators ....................................... 18  
   1.5 Metric Entropy ................................................ 21  

2 Trace Class Operators .......................................... 28  
   2.1 Kernels of Trace Class Operators ............................ 28  
   2.2 Series Representation of Trace Class Kernels ............ 31  
   2.3 Convergence a.s. and in $L^1$ ................................ 38  

3 Central Limit Theorem .......................................... 46  
   3.1 Bases .......................................................... 46  
   3.2 A Central Limit Theorem ..................................... 52  
   3.3 Measurability and Majorizing Measures ..................... 57  

4 von Mises Functionals ........................................... 64  
   4.1 Basic Concepts ............................................... 64  
   4.2 Differentiability ............................................. 69  
   4.3 Asymptotic distribution of $T(P_n)$ ......................... 75


Introduction

Consider \((X, \mathcal{A})\) a measurable space and \(K : X^2 \to \mathbb{R}\) a function, which in principle is not assumed to be symmetric. We are interested in the behavior of the functional \(T_K(P)\) defined by

\[
T_K(P) = \int \int K(x, y)dP(x)dP(y)
\]

where \(P\) is any probability measure on \((X, \mathcal{A})\). Such functionals are called “von Mises functionals” and there are extensive known results about them in statistical literature, since von Mises proposed the study of several statistics and their asymptotic behavior using the concept of Gateaux differentiability in 1947. However, most of the work done on von Mises functionals has been dedicated to the unidimensional case, with the exception of von Mises himself and Filippova in 1961, the case of multidimensional kernels has been somewhat neglected until recent studies, in which the asymptotic behavior of such functionals is of basic interest. In the following research we analyze only the bidimensional case and we mainly find results when \(K(\cdot, \cdot)\) is the kernel of a trace class operator.

The first chapter of this thesis is a compilation of results on operators which includes: basic definitions and theorems about linear operators in the first section. The second section includes the main results on Schmidt class operators \((\sigma c)\) and concludes with the completeness of the space with respect to its metric \(\sigma(A - B)\). In the third section we will talk about integral operators and their relation to Schmidt operators. The fourth section introduces trace class operators and several results related to them. Through this section all the results are taken from Schatten [1960]. The fifth and last section defines the concept of metric entropy and together with the results of Mitjagin [1961], we give a characterization of a trace class operator in terms of an entropy condition; later results involving entropy will be used in the following chapters.

In the first section of the second chapter we give a list of known results about conditions on a kernel \(K(x, y)\), that imply that \(A_K\), the integral operator it generates, is a trace class operator, and also an entropy condition on \(K\), that depends on a Smithies-Stinespring theorem. The second section deals mainly with the representation of a trace class kernel, in terms of a series of products of \(L^2\) functions, such that the series is convergent in \(L^2\); and the section ends with an example which shows that the conditions on the representation Theorem are sharp. The last section of chapter II is about convergence a.s. and in \(L^1\) of the series given by the representation Theorem, not only on \([0, 1]^2\), but also on the diagonal, which in most cases represents a problem, since its measure is generally zero.

The third chapter answers the question: Which conditions on \(K(x, y)\) imply that \(K(x, y) = \sum a_if_i(x)g_i(y)\), where \(\sum |a_i||f_i|\|g_i|\| < \infty\). The purpose of this condition is that it does not depend on any measure, and therefore, we can find “universal”
results. To do so we use another paper of Mitjagin published in 1964. In the first section we make a short analysis of some other possible basic systems, besides the usual Fourier series. In a second section we give a CLT under metric entropy with $L^2$-bracketing, to do this we use a strong result proved by Ossiander in 1987. In the third and last section of this chapter, we use a later paper, Andersen et al. [1988], which proves that the separability condition in Ossiander's Theorem can be removed, and that we can replace $L^2$-brackets by what they call $\Delta_{2,\infty}$-brackets and the entropy condition by the weaker majorizing measure condition. We finish this section and the chapter by relating both papers.

In the fourth chapter we talk about the functional

$$T_K(\mu, \nu) = \iint K(x, y) d\mu(x) d\nu(y)$$

where $\mu, \nu$ are signed measures and more specifically differences of probability measures. The first section is an introduction to von Mises functionals and the basic concepts of differentiability in the sense of Gateaux, compact or Hadamard and Fréchet as in Reeds [1976]. In the second section we use Fréchet differentiability with respect to $\| \cdot \|_F$, that is the norm generated by an specific family of functions $F$, and we find conditions on $F$ in order to have the Fréchet differentiability of $T_K(P)$ and also obtain $\sqrt{n} \| P_n - P \|_F = O_p(1)$. In the last section we find the asymptotic distribution of $T_K(P_n)$ using results of Serfling [1980].
Chapter 1

Operators

The first four sections of this chapter are a review of some known facts about operators on Hilbert spaces. For facts stated but not proved here, refer to Schatten [1960] for proofs.

1.1 Review

Definition 1.1.1 Consider a complex Hilbert space $H$ with inner product $(\cdot, \cdot)$. A family $\{f_j\}_{j \in J} \subset H$ is summable with sum $f$ (denoted by $\sum_{j \in J} f_j = f$) iff for every $\epsilon > 0$, there exists $J_1 \subset J$ a finite subset of indices, such that for every $J_1 \subset J_0$, finite:

$$\| \sum_{j \in J_0} f_j - f \| < \epsilon$$

where $\|h\| = (h, h)^{1/2}$ for $h \in H$.

Definition 1.1.2 $\{\varphi_j\}_{j \in J} \subset H$ is a basis for $H$ iff $\{\varphi_j\}_{j \in J}$ is complete, i.e. $(f, \varphi_j) = 0$ for every $j \in J$ implies $f = 0$, and $\{\varphi_j\}$ is orthonormal, i.e. $(\varphi_j, \varphi_k) = \delta_{j,k}$.

The next theorem about bases of Hilbert spaces is well known:

Theorem 1.1.3 Let $\{\varphi_j\}_{j \in J} \subset H$ be an orthonormal family of vectors in $H$. Then the following are equivalent:

i) $\{\varphi_j\}_{j \in J}$ is a basis

ii) $(f, \varphi_j) = 0$ for all $j$ implies $f = 0$

iii) For every $f \in H$ we have $f = \sum_{j \in J}(f, \varphi_j)\varphi_j$ (Fourier series expansion)

iv) For every $f, g \in J$ $\sum_{j \in J}(f, \varphi_j)(\varphi_j, g) = (f, g)$ (Parseval’s identity)

v) For every $f \in H$ $\|f\|^2 = \sum_{j \in J}|(f, \varphi_j)|^2$.
Definition 1.1.4 A bounded linear transformation $A$ from $H$ into $H$ is called an operator and its norm is denoted by $\|A\|$.

Definition 1.1.5 An operator $A$ is invertible iff there exists an operator $B$ such that $AB=BA=I$, the identity operator, and we write $B = A^{-1}$.

Definition 1.1.6 The set of all complex numbers $\lambda$ such that $A-\lambda I$ is not invertible, is called the spectrum of $A$ and is denoted by $\Lambda(A)$.

In fact $\Lambda(A)$ is a closed subset of the disk $|z| \leq \|A\|$.

Definition 1.1.7 The adjoint of $A$ is denoted by $A^*$, and is defined as the operator such that $(Af, g) = (f, A^*g)$ for every $f, g \in H$. $A$ is called selfadjoint or Hermitian iff $A = A^*$.

If $A$ is selfadjoint then $\Lambda(A) \subset \mathbb{R}$.

Definition 1.1.8 $A$ is called positive iff $(Af, f) \geq 0$ for all $f \in H$.

Every positive operator on a complex Hilbert space is then selfadjoint.

Definition 1.1.9 $A$ is normal iff $AA^* = A^*A$.

Definition 1.1.10 $A$ is unitary iff $AA^* = A^*A = I$.

Definition 1.1.11 Let $\mathcal{M} \subset H$ be a linear subspace of $H$. An operator $W$ is called partially isometric iff it is isometric on $\mathcal{M}$ and equal to 0 on $H \ominus \mathcal{M} = \{y \in H | (x, y) = 0 \ \forall x \in \mathcal{M}\}$, $\mathcal{M}$ is called the initial set of $W$ and the range $\mathcal{R}$ of $W$ is its final set.

By definition of the adjoint operator $A^{**} = A$ and $A^*A$ is positive, since $(A^*Af, f) = (Af, A^{**}f) = (Af, Af) \geq 0$. For every positive operator $A$ there exists a unique positive operator $B$ such that $A = B^2$.

Definition 1.1.12 Since $A^*A$ is positive we can define $[A] = (A^*A)^{1/2}$.

It is easily checked that $[A] = [A^*]$ iff $A$ is normal.

If $A$ is of finite rank, that is, its range is of finite dimension, then $A^*A, AA^*$, $[A]$ and $[A^*]$ are also of finite rank.
Theorem 1.1.13 Let $A$ be an operator. There exists a partially isometric operator $W$ whose initial set is the closure of the range of $[A]$ and the final set is the closure of the range of $A$, satisfying:

i) $A = W[A]$

ii) $[A] = W^*A$

iii) $A^* = W^*[A^*]$


Also these representations are unique in the sense that if $A = W_1B_1$ where $B_1 \geq 0$ and $W_1$ is partially isometric, with initial set the closure of the range of $B_1$, then $B_1 = [A]$ and $W_1 = W$. If $A$ is of finite rank then we can assume $W$ is unitary (not necessarily unique however).

Proof: see Schatten page 4.

Definition 1.1.14 Given $\varphi$ and $\psi$ in $H$ we define the operator $\varphi \otimes \overline{\psi}$ by:

$$(\varphi \otimes \overline{\psi})f = (f, \psi)\varphi.$$ 

With this definition it is clear that $\varphi \otimes \overline{\psi}$ is an operator.

Theorem 1.1.15 Let $\{\varphi_j\}_{j \in J}$ and $\{\psi_j\}_{j \in J}$ be two orthonormal families of vectors and $\{\lambda_j\}_{j \in J}$ be a family of complex numbers, then $\{\lambda_j(f, \psi_j)\varphi_j\}_{j \in J}$ is summable for every $f \in H$ iff $\{\lambda_j\}$ is bounded. If $\{\lambda_j\}$ is bounded the sum $\sum_{j \in J} \lambda_j(f, \psi_j)\varphi_j$ defines an operator denoted by:

$$\sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi_j}$$

whose norm is given by $\sup_{j \in J} |\lambda_j|$. 

Proof: see Schatten page 8.

Corollary 1.1.16 $\sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi_j}$ is zero iff every $\lambda_j = 0$, and so $\sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi_j} = \sum_{j \in J} \mu_j \varphi_j \otimes \overline{\psi_j}$ iff $\mu_j = \lambda_j$ for all $j$.

Theorem 1.1.17 Let $A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi_j}$. Then:

i) $A^* = \sum_{j \in J} \overline{\lambda_j} \psi_j \otimes \overline{\varphi_j}$

ii) $A^*A = \sum_{j \in J} |\lambda_j|^2 \psi_j \otimes \overline{\psi_j}$

iii) $[A] = \sum_{j \in J} \lambda_j |\psi_j \otimes \overline{\psi_j}|$. 

10
The operator $\sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\varphi_j}$ is normal, and is Hermitian iff all $\lambda_j$ are real.

**Theorem 1.1.18** The operator $A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\varphi_j}$ has an inverse iff

i) $\{\varphi_j\}_{j \in J}$ is a basis

ii) $\{\psi_j\}_{j \in J}$ is a basis

iii) $\{\lambda_j\}_{j \in J}$ is not only bounded but bounded away from 0

and in that case $A^{-1} = \sum_{j \in J} (1/\lambda_j) \psi_j \otimes \overline{\varphi_j}$.

**Proof:** See Schatten page 11.

Notice that for a basis $\{\varphi_j\}$ we have:

$$\sum_{j \in J} \varphi_j \otimes \overline{\varphi_j} = I.$$ 

If we have a representation for $A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\varphi_j}$ we can find the spectrum $\Lambda(A) = \{\lambda \in \mathbb{C} | A - \lambda I \text{ is not invertible} \}$:

$$A - \lambda I = \sum_{j \in J} (\lambda_j - \lambda) \varphi_j \otimes \overline{\varphi_j},$$

so $\Lambda(A)$ is the closure of the set of the $\lambda_j$.

If $\lambda \notin \Lambda(A)$ then $(A - \lambda I)^{-1} = \sum_{j \in J} \frac{1}{\lambda_j - \lambda} \varphi_j \otimes \overline{\varphi_j}$.

For the case $\{\varphi_j\}_{j \in J}$ not being complete, the same can be done by adding $\{\omega_i\}_{i \in I}$ such that $\{\varphi_j\}_{j \in J} \cup \{\omega_i\}_{i \in I}$ is a basis, and noticing that the range of $\sum_{j \in J} \varphi_j \otimes \overline{\varphi_j}$ is the subspace spanned by $\{\varphi_j\}_{j \in J}$ and is equal to the range of $A$, so that:

$$A - \lambda I = \sum_{j \in J} (\lambda_j - \lambda) \varphi_j \otimes \overline{\varphi_j} + \sum_{i \in I} (-\lambda) \omega_i \otimes \overline{\omega_i}.$$ 

As a remark, given a non-empty compact set $\Lambda$ of complex numbers, we can construct a normal operator on a given infinite dimensional space, having $\Lambda$ as its spectrum. To do so just take $\lambda_1, \lambda_2, \ldots$ a dense sequence in $\Lambda$, take an orthonormal sequence $\varphi_1, \varphi_2, \ldots$ and if necessary extend it to be a basis by adding $\{\omega_i\}_{i \in I}$, then let $A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\varphi_j} + \sum_{i \in I} (\lambda_1) \omega_i \otimes \overline{\omega_i}$ and if $\Lambda$ is a set of real numbers then $A$ is selfadjoint.

Let $\mathcal{U}$ be the algebra of all bounded operators on $H$.

There are a lot of equivalent definitions for compact (or completely continuous) operators. We adopt the following:
Definition 1.1.19 An operator $A$ is compact if it takes weakly convergent sequences into strongly convergent sequences, in symbols $f_n \rightharpoonup f$ implies $Af_n \to Af$. The class of all compact operators will be denoted by $C$. 

This definition is equivalent to the usual one that states that $A$ of the unit ball is totally bounded. As an immediate consequence of the definition we have the following:

Lemma 1.1.20 $C$ with usual sum, product and scalar multiplication is a two sided ideal in the algebra $U$ i.e.

i) $A \in C \Rightarrow \lambda A \in C$ for any complex number

ii) $A, B \in C \Rightarrow (A + B) \in C$

iii) $A \in C, X \in U \Rightarrow (AX) \in C$ and $(XA) \in C$.

Proof: It follows directly from the definition of compact operators. \hfill \Box

From the polar decomposition of operators, Theorem 1.1.13 and lemma 1.1.20, we have the next:

Lemma 1.1.21 The following are equivalent:

i) $A \in C$

ii) $A^* \in C$

iii) $[A] \in C$

iv) $[A^*] \in C$.

Lemma 1.1.22 Let $A_n \in C$ for $n = 1, 2, \ldots$ and suppose $\|A - A_n\|_{n \to \infty} \to 0$, then $A \in C$.


As an important corollary we have the following:

Corollary 1.1.23 Let $\{\varphi_i\}_{i \in I}$ and $\{\psi_i\}_{i \in I}$ be two orthonormal sequences and $\{\lambda_i\}_{i \in I}$ a sequence convergent to 0 of complex numbers, then:

$$\sum_{i \in I} \lambda_i \varphi_i \otimes \overline{\psi_i} \in C.$$ 

Proof: Consider $A_n = \sum_{i=1}^{n} \lambda_i \varphi_i \otimes \overline{\psi_i}$ and apply Lemma 1.1.22. \hfill \Box
Theorem 1.1.24 The compact operators $A$ are precisely those admitting a polar representation $A = \sum \lambda_i \varphi_i \otimes \psi_i$ where both $\{\varphi_i\}$ and $\{\psi_i\}$ are orthonormal sequences and all the $\lambda_i$'s are positive. The sum has either a finite or a denumerably infinite number of terms; in the last case, we have also $\lambda_i \to 0$. The above representation is unique in the sense that the $\lambda_i$'s are necessarily all the positive eigenvalues (each appearing the number of times that equals its multiplicity) of $[A]$.

Proof: See Schatten page 19.

The next theorem includes several of the results about $C$:

Theorem 1.1.25 Consider the set $C$ of all compact operators. With the usual definition of sum, product and scalar multiple for operators, $C$ is a selfadjoint two-sided ideal in the algebra $\mathcal{U}$. The algebra $C$ is normed. Moreover, the resulting normed algebra is complete, that is, $C$ is a Banach algebra.

1.2 Schmidt Class Operators

Lemma 1.2.1 Let $A$ be an operator on $H$ and $\{\varphi_j\}_{j \in J}$, $\{\psi_j\}_{j \in J}$ two bases of $H$. Then the families:

$$\{\|A\varphi_j\|^2\}, \{\|A^*\psi_j\|^2\}, \{|(A\varphi_j, \psi_i)|^2\}$$

of nonnegative numbers are all summable or not, and the sums are equal and do not depend on the bases if they are summable.

Proof: By Theorem 1.1.3 v) $\|A\varphi_j\|^2 = \sum_i |(A\varphi_j, \psi_i)|^2$. Then

$$\sum_j \|A\varphi_j\|^2 = \sum_{i,j} |(A\varphi_j, \psi_i)|^2 = \sum_{i,j} |(\varphi_j, A^*\psi_i)|^2 = \sum_i |(A^*\psi_i, A^*\varphi_j)|^2 = \sum_i \|A^*\psi_i\|^2.$$

In the same way we have (using $A^{**} = A$):

$$\sum_i \|A^*\psi_i\|^2 = \sum_{i,j} |(A^*\psi_i, \psi_j)|^2 = \sum_{i,j} |(\psi_i, A^{**}\psi_j)|^2 = \sum_{i,j} |(\psi_i, A^*\psi_j)|^2 = \sum_i \|A^*\psi_i\|^2 = \sum_i \|A^*\psi_i\|^2 = \sum_i \|A\varphi_j\|^2.$$

From all these equalities $\sum_j \|A\varphi_j\|^2 = \sum_i \|A\psi_j\|^2$. \qed

Definition 1.2.2 If the families above are summable we denote the common value of the sum by $\sigma(A)^2$, in other cases let $\sigma(A) = \infty$.

The class of operators $A$ such that $\sigma(A) < \infty$ forms the Schmidt class denoted by $(\sigma c)$.

Lemma 1.2.3 For every operator $A$, $\|A\| \leq \sigma(A)$. 

13
Proof: There is nothing to prove if \( \sigma(A) = \infty \), so suppose \( \sigma(A) < \infty \). To show that \( \|A\varphi\| \leq \sigma(A) \) for every \( \varphi \) such that \( \|\varphi\| = 1 \), let \( \{\varphi_j\} \) be a basis for \( H \) with \( \varphi \) one of its elements, then we have: \( \|A\varphi\| \leq (\sum_j \|A\varphi_j\|^2)^{1/2} = \sigma(A) \).

Lemma 1.2.4 For the class \((\sigma c)\) we have:

i) \( A \in (\sigma c) \) iff \( A^* \in (\sigma c) \) and \( \sigma(A) = \sigma(A^*) \).

ii) If \( A \in (\sigma c) \) and \( \lambda \) is a complex number then \( \lambda A \in (\sigma c) \) and \( \sigma(\lambda A) = |\lambda|\sigma(A) \).

iii) \( A, B \in (\sigma c) \) implies \( A + B \in (\sigma c) \) and \( \sigma(A + B) \leq \sigma(A) + \sigma(B) \).

iv) \( A \in (\sigma c) \), \( X \in \mathcal{U} \) implies \( AX, XA \in (\sigma c) \) and \( \sigma(XA), \sigma(AX) \leq \|X\|\sigma(A) \).

v) If \( \varphi, \psi \in H \) then \( (\varphi \otimes \overline{\psi}) \in (\sigma c) \) and \( \sigma(\varphi \otimes \overline{\psi}) = \|\varphi\||\psi\| \).

vi) \( A \in (\sigma c) \) iff \( [A] \in (\sigma c) \) and \( \sigma(A) = \sigma([A]) \).

Proof: i) and ii) follow from the definition of \((\sigma c)\) and \( \sigma(\cdot) \).

iii) Let \( \{\varphi_j\}_{j \in J} \) be a basis of \( H \) and let \( K \subset J \) finite, then

\[
(\sum_{j \in K}(\|A + B\|\varphi_j\|^2)^{1/2} \leq (\sum_{j \in K}(\|A\varphi_j\|^2 + \|B\varphi_j\|^2)^{1/2} \leq (\sum_{j \in K}\|A\varphi_j\|^2)^{1/2} + (\sum_{j \in K}\|B\varphi_j\|^2)^{1/2} \leq \sigma(A) + \sigma(B).
\]

Since this is true for every finite set \( K \subset J \) the result follows.

iv) \( \sigma(AX) = \sigma(AX^*) = \sigma(X^*A^*) \) by i), but

\[
\|X\|A\varphi_j\|^2 \leq \|X\|\|A\varphi_j\|^2
\]

so \( \sigma(XA) \leq \|X\|\sigma(A) \), and then:

\[
\sigma(X^*A^*) \leq \|X^*\|\sigma(A^*) = \|X\|\sigma(A).
\]

v) Let \( \{\varphi_j\}_{j \in J} \) be a basis of \( H \). By Theorem 1.3 v) we have:

\[
\|\varphi\|^2\|\psi\|^2 = \|\varphi\|^2\sum_{j \in J}(\varphi_j, \psi)^2 = \sum_{j \in J}(\varphi_j, \psi)\varphi_j^2 = \sum_{j \in J}(\varphi \otimes \overline{\psi})\varphi_j^2 = (\sigma(\varphi \otimes \overline{\psi}))^2.
\]

vi) By Theorem 1.1.13 we have \( A = W[A] \) and \( [A] = W^*A \) and apply iv). \( \square \)

By v) and iii) all operators of finite rank belong to \((\sigma c)\), in fact by ii) and iii) \((\sigma c)\) is a linear space.
Lemma 1.2.5 Let $A, B \in (\sigma c)$ and let $\{\varphi_j\}_{j \in J}$ be a basis for $H$. Then the family
\[ \{(A\varphi_j, B\varphi_j)\} \]
is summable, and so is $\{(A\varphi_j, B\varphi_j)\}$, and the latter sum is independent of $\{\varphi_j\}_{j \in J}$.

Proof: See Schatten page 30.

Definition 1.2.6 If $A, B \in (\sigma c)$ and $\{\varphi_j\}_{j \in J}$ is a basis for $H$ we define:
\[ (A, B) = \sum_{j \in J} (A\varphi_j, B\varphi_j). \]

Lemma 1.2.7 Let $A, A_1, B, B_1 \in (\sigma c), X \in \mathcal{U}$, and $c$ a complex number. Then:
i) $(B, A) = \overline{(A, B)}$
ii) $(cA, B) = c(A, B)$
ii') $(A, cB) = \overline{c}(A, B)$
iii) $(A_1 + A_2, B) = (A_1, B) + (A_2, B)$
iii') $(A, B_1 + B_2) = (A, B_1) + (A, B_2)$
iv) $(A, A) \geq 0, (A, A) = 0$ iff $A=0$
v) $(A^*, B^*) = (A, B)$
vi) $(X A, B) = (A, X^* B)$
vi') $(X A, B) = (A, BX^*)$
vii) For $\varphi, \psi, f, g \in H$ $(\varphi \otimes \overline{\psi}, f \otimes \overline{g}) = (\varphi, f)(g, \psi)$.


From i) to iv) $(\cdot, \cdot)$ is an inner product in $(\sigma c)$ and $(A, A)^{1/2} = \sigma(A)$ is the norm that it defines.

In particular $|(A, B)| \leq \sigma(A)\sigma(B)$ by Schwarz's inequality.

Lemma 1.2.8 $(\sigma c)$ is a complete space with respect to its metric $\sigma(A - B)$.

Proof: See Schatten page 32.

Theorem 1.2.9 If $A \in (\sigma c)$ then $A \in \mathcal{C}$.

Proof: See Schatten page 32.
Theorem 1.2.10  A compact operator with polar representation \( A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi}_j \in (\sigma c) \) iff
\[
\sum_{j \in J} \lambda_j^2 < \infty
\]
which is the series of nonzero eigenvalues of \( A^* A \), and in this case
\[
\sigma(A) = (\sum_{j \in J} \lambda_j^2)^{1/2}.
\]

Proof: See Schatten page 33.

Theorem 1.2.11  Let \( \{\varphi_j\}_{j \in J} \) be a basis in \( H \). The set \( (\sigma c) \) of all operators \( A \) for which \( \{\|A \varphi_j\|^2\} \) is summable, is a linear space. There
\[
\sigma(A) = (\sum_{j \in J} \|A \varphi_j\|^2)^{1/2}
\]
is a norm. The resulting normed linear space is complete, hence a Banach space; it contains the operators of finite rank as a dense subset. For any pair of operators \( A \) and \( B \) in \( (\sigma c) \), the family \( \{(A \varphi_j, B \varphi_j)\} \) of complex numbers is summable. Its sum \( (A, B) \) defines an inner product in \( (\sigma c) \) and \( (A, A)^{1/2} = \sigma(A) \) is the norm that goes with it. Thus, \( (\sigma c) \) is a Hilbert space (independent of the chosen basis \( \{\varphi_j\}_{j \in J} \)). The operators in \( (\sigma c) \) are necessarily compact. They are precisely those compact operators (in the polar form) \( A = \sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi}_j \) for which \( \sum_{j \in J} \lambda_j^2 < \infty \); we also have \( \sigma(A) = (\sum_{j \in J} \lambda_j^2)^{1/2} \). Moreover, \( (\sigma c) \) is an ideal in the algebra \( \mathcal{U} \). Under its own norm \( (\sigma c) \) is a Banach algebra, in fact, also a norm-ideal.

Proof: This theorem is just a compilation of all the previous results.

\( \square \)

1.3  Integral Operators

Consider \( (\Omega, A, \mu) \) a measure space and consider the space of \( \mu \)-square integrable functions.

Definition 1.3.1
\[
L^2(\mu) = \{ f : \Omega \rightarrow C | \ f \text{ is measurable and } \int |f|^2 d\mu < \infty \},
\]
\[
L^2(\mu \otimes \mu) = \{ f : \Omega \otimes \Omega \rightarrow C | \ f \text{ is measurable and } \int \int |f(x, y)|^2 d\mu(x) d\mu(y) < \infty \}
\]
with usual inner product, i.e.
\[
(f, g) = \int f \overline{g} d\mu \quad \text{for} \quad f, g \in L^2(\mu)
\]
\[
\langle f, g \rangle = \int \int f(x) \overline{g(y)} d\mu(x) d\mu(y) \quad \text{for} \quad f, g \in L^2(\mu \otimes \mu)
\]
with corresponding norms
\[
\|f\| = (f, f)^{1/2} \ \text{in} \ L^2(\mu)
\]
\[
\|f\| = \langle f, f \rangle^{1/2} \ \text{in} \ L^2(\mu \otimes \mu).
\]
It is well known that $L^2(\mu), L^2(\mu \otimes \mu)$ with their inner products become Hilbert spaces.

**Definition 1.3.2** For $K(x,y) \in L^2(\mu \otimes \mu)$ define:

$$A_K(f) = \int K(x,y)f(y)d\mu(y) \quad \text{for} \quad f \in L^2(\mu).$$

Then $A_K$ defines an operator in $L^2(\mu)$ and is called an integral operator with kernel $K(x,y)$. It is direct to check that $(A_K)^* = A_K^*$ if

$$K^*(x,y) = \overline{K(y,x)} \quad \text{a.e.}[\mu].$$

If $A_{K_1}, A_{K_2}$ are integral operators with kernels $K_1, K_2$ then for every complex number $c, cA_{K_1}$ and $A_{K_1} + A_{K_2}$ are integral operators with kernels $cK_1$ and $K_1 + K_2$ respectively, moreover $A_{K_1}A_{K_2}$ is an integral operator with kernel $\int K_1(x,z)K_2(z,y)d\mu(z)$ as can be verified by easy calculations.

By Fubini’s theorem and Schwarz’s inequality

$$\|A_K\| \leq \|K\|$$

and in the case of products

$$\|\int K_1(x,z)K_2(z,y)d\mu(z)\| \leq \|K_1\| \cdot \|K_2\|.$$

**Theorem 1.3.3** An operator $A_K$ on $L^2(\mu)$ is of integral type iff $A_K \in (\sigma c)$, and if $K$ generates $A_K$, then

$$\sigma(A_K) = \|K\|.$$  

The correspondence $A_K \leftrightarrow K$ is a linear isometry between $L^2(\mu \otimes \mu)$ and the Schmidt operators on $L^2(\mu)$.

**Proof:** See Schatten page 35.

**Corollary 1.3.4** If $K(x,y) \in L^2(\mu \otimes \mu)$ is such that $K(x,y) = \overline{K(y,x)}$ then we can write:

$$K(x,y) = \sum_{j \in J} \lambda_j \varphi_j(x)\overline{\varphi_j(y)}$$

where the $\lambda_j$s are the non-zero eigenvalues of the integral operator $A_K$, which are real, and $\{\varphi_j(x)\}$ is a corresponding sequence of eigenvectors. The convergence of the sequence is in the sense of $L^2(\mu \otimes \mu)$.  

17
Proof: See Schatten page 36.

Analogously we have:

Corollary 1.3.5 If \( K(x, y) \in L^2(\mu \otimes \mu) \) we can rewrite

\[
K(x, y) = \sum_{j \in J} \lambda_j \varphi_j(x) \overline{\psi_j(y)}
\]

where all the \( \lambda_j \)'s are greater than 0 and \( \{\varphi_j(x)\} \) and \( \{\psi_j(x)\} \) are orthonormal sequences in \( L^2(\mu) \). The convergence of the series is in the sense of \( L^2(\mu \otimes \mu) \), and the representation is unique in the sense that the \( \lambda_j \)'s are necessarily the positive eigenvalues of \( [A_K] \).

Proof: The result follows from Theorem 1.1.13 and the proof given in Schatten page 36.

\( \square \)

1.4 Trace Class Operators

Lemma 1.4.1 Let \( A = BC \) be the product of two operators in \( \sigma_c \) and \( \{\varphi_j\}_{j \in J} \) a basis in \( H \). Then \( \{(A\varphi_j, \varphi_j)\}_{j \in J} \) is summable and so is \( \{(A\varphi_j, \varphi_j)\} \), and the latter sum is independent of \( \{\varphi_j\}_{j \in J} \).

Proof: See Schatten page 37.

Definition 1.4.2 The products of two operators in \( \sigma_c \) form the trace class \( \tau_c \) and for \( A \in \tau_c \)

\[
t(A) = \sum_{j \in J} (A\varphi_j, \varphi_j)
\]

is called the trace of \( A \).

In fact if \( B, C \in \sigma_c \) and \( A = C^*B \) then \( t(A) = (B, C) \) and if \( B_1 \) and \( C_1 \in \sigma_c \), then \( C^*B = C_1^*B_1 \) implies \( (B, C) = (B_1, C_1) \), so \( t \) does not depend on the representation.

If \( A \) is an operator, we know \( [A] \geq 0 \) and so \( [A]^{1/2} \) is well defined, and we have the next:

Lemma 1.4.3 Consider \( \{\varphi_j\}_{j \in J} \) a fixed basis, then the following are equivalent:

i) \( A \in \tau_c \)

ii) \( [A] \in \tau_c \)

iii) \( [A]^{1/2} \in \sigma_c \)

iv) \( \sum_{j \in J} ([A]\varphi_j, \varphi_j) < \infty. \)
Proof: See Schatten page 37.

Lemma 1.4.4 For the class $(\tau c)$ we have:

i) $A \in (\tau c)$ iff $A^* \in (\tau c)$

ii) $A \in (\tau c)$ implies $(\lambda A) \in (\tau c)$ for every complex number $\lambda$

iii) $A, B \in (\tau c)$ implies $(A + B) \in (\tau c)$

iv) $A \in (\tau c)$ and $X \in \mathcal{U}$ implies $AX$ and $XA \in (\tau c)$.

Proof: See Schatten page 38.

Lemma 1.4.5 The trace $t(\cdot)$ has the following properties:

i) For any $A \in (\tau c)$, $A^* \in (\tau c)$ and $t(A^*) = t(A)$.

ii) $t(cA) = ct(A)$ for any complex number $c$ and $A \in (\tau c)$

iii) $t(A + B) = t(A) + t(B)$ for any $A, B \in (\tau c)$

iv) $t(AB) = t(BA)$ with either $A$ or $B$ in $(\tau c)$.

Proof: See Schatten page 38.

Definition 1.4.6 For $A \in (\tau c)$ let $\tau(A) = t([A])$.

Lemma 1.4.7 $A \in (\tau c)$ iff $[A] \in (\tau c)$ and $\tau(A) = \tau([A])$.

Proof: Refer to Schatten page 39.

Lemma 1.4.8 $A \in (\tau c)$ iff $[A]^{1/2} \in (\sigma c)$ and

$$\tau(A) = (\sigma([A]^{1/2}))^2.$$  


Lemma 1.4.9 If $A \in (\tau c)$ and $X$ is any operator, then $X[A] \in (\tau c)$ and

$$|t(X[A])| \leq \|X\|\tau(A).$$

Lemma 1.4.10 Let $A, B \in (\tau c), X \in U$ and $\lambda$ a complex number. Then

i) $\tau (A^*) = \tau (A)$

ii) $\tau (\lambda A) = |\lambda| \tau (A)$

iii) $\tau (A + B) \leq \tau (A) + \tau (B)$

iv) $\tau (A) \geq 0$, $\tau (A) = 0 \Rightarrow A = 0$

v) $\tau (AX)$ and $\tau (XA)$ are less or equal $\|X\| \tau (A)$

vi) $|t(A)| \leq \tau (A)$.

Proof: See Schatten page 40.

Corollary 1.4.11 If $A \in (\tau c)$ and $\{\varphi_j\}_{j \in J}, \{\psi_j\}_{j \in J}$ are two bases, then $\{|(A\varphi_j, \psi_j)|\}$ is summable. Hence $\{(A\varphi_j, \psi_j)\}$ is also summable, moreover:

$$|\sum_{j \in J} (A\varphi_j, \psi_j)| \leq \tau (A).$$

Proof: See Schatten page 40.

Lemma 1.4.12 If $\varphi, \psi \in H$

i) $\varphi \otimes \overline{\psi} \in (\tau c)$

ii) $t(\varphi \otimes \overline{\psi}) = (\varphi, \psi)$

iii) $\tau (\varphi \otimes \overline{\psi}) = \|\varphi\| \cdot ||\psi||$.

Proof: Refer to Schatten page 41.

Theorem 1.4.13 If $A \in (\tau c)$ then $A \in C$.

Proof: See Schatten page 41.

Theorem 1.4.14 Let $A \in C$ and let $\sum_{j \in J} \lambda_j \varphi_j \otimes \overline{\psi}_j$ be a polar representation of $A$, then: $A \in (\sigma c)$ iff $\sum_{j \in J} \lambda_j^2 < \infty$ and in this case $\sigma (A) = (\sum_{j \in J} (\lambda_j)^2)^{1/2}$. Furthermore $A \in (\tau c)$ iff $\sum_{j \in J} \lambda_j < \infty$, and in this case

$$\tau (A) = \sum_{j \in J} \lambda_j.$$

Proof: See Schatten page 41.

Theorem 1.4.15 If $A \in (\tau c)$ and $\epsilon > 0$ there exists $B$ of finite rank such that $\tau (A - B) < \epsilon$.
Proof: See Schatten page 41.

Theorem 1.4.16 For $A \in (\tau c)$ we have the next inequalities:

$$\|A\| \leq \sigma(A) \leq \tau(A).$$

Proof: See Schatten page 42.

To conclude this section we give a summary of all the previous results on the class of trace class operators:

Theorem 1.4.17 Consider the class $(\tau c)$ of all products of two operators in the class $(\sigma c)$. This class is the same as the class of all operators $A$ such that $\sum_{j \in J}|[A]|_{\varphi_j, \varphi_j} < \infty$ for a given basis $\{\varphi_j\}_{j \in J}$. With usual sum and scalar multiplication $(\tau c)$ is a linear space. And it is also normed, with norm given by the sum above, which is independent of the basis, and we called it $\tau(A)$. The resulting normed linear space is complete; it contains the operators of finite rank as a dense subset. $(\tau c) \subset (\sigma c)$, so all trace class operators are compact; and they are the operators having the polar form $\sum_{j \in J}\lambda_j\varphi_j \otimes \bar{\psi}_j$ with $\sum_{j \in J}\lambda_j < \infty$, and then we have $\tau(A) = \sum_{j \in J}\lambda_j$. Moreover $(\tau c)$ is a two-sided ideal in the algebra $\mathcal{U}$, and a Banach algebra under its own norm. In fact $(\tau c)$ is also a norm ideal.

Proof: See all the results in this section and for the completeness see Schatten page 47.

1.5 Metric Entropy

In this section we introduce the notion of “Metric Entropy” (also called $\varepsilon$—entropy). For simplicity we will use the notation of Mityagin [1961], but we have to keep in mind that metric entropy can be defined in any metric space; for example see Dudley [1984], Chapter 6. The purpose of this section is to find a characterization of trace class operators in terms of metric entropy.

Definition 1.5.1 A set $A$ in a linear space $E$ is called a balanced convex set if for any numbers $\lambda$ and $\mu$ such that $|\lambda| + |\mu| \leq 1$ we have $\lambda A + \mu A \subset A$.

Definition 1.5.2 Let $S$ be a balanced convex set in a linear space $E$ and $K$ any subset of $E$, we write:

$$N(K, S) = \inf\{n : K \subset \bigcup_{i=1}^{n}(x_k + S), \quad x_k \in E \text{ for } k = 1, \ldots, n\}$$
i.e., \( N(K,S) \) is the smallest number of translations of the set \( S \) which cover the set \( K \).

And now let:

\[
M(K, S) = \sup\{n : x_i - x_j \notin S, \ i \neq j, \ x_i \in K; i, j = 1, \ldots, n\}
\]

i.e. \( M(K,S) \) is the largest number of elements of the set \( K \) whose distance apart is greater than 1, with the distance understood to be:

\[
|x - y|_S = \sup\{\lambda : \lambda(x - y) \notin S, \lambda > 0\}.
\]

The \( \epsilon \)-entropy of \( K \) with respect to the set \( S \) is defined by:

\[
H(K, \epsilon S) = \log N(K, \epsilon S)
\]

and the \( \epsilon \)-capacity of \( K \) with respect to the set \( S \) is defined by:

\[
\Sigma(K, \epsilon S) = \log M(K, \epsilon S).
\]

These two concepts are closely related as it is shown in the next:

**Lemma 1.5.3** For every set \( K \subset \mathcal{E} \) and every balanced convex set \( S \) we have the inequality:

\[
M(K, 2S) \leq N(K, S) \leq M(K, S).
\]

**Proof:** For a more general proof we refer to Dudley [1984], Thm. 6.0.1, page 39. \( \Box \)

**Definition 1.5.4** A \( p \)-ellipsoid in the space \( l_p, 1 \leq p < \infty \) is a set of the form:

\[
\Xi = \{ \xi \in l_p : (\sum_{n=1}^{\infty} |\xi_n a_n|^p)^{1/p} \leq 1\}
\]

where \( a_n \) is a monotonically increasing sequence of positive numbers such that \( a_n \geq 1 \) and \( a_n \to \infty \) as \( n \to \infty \). And we also define:

\[
m(t) = \sup\{n : a_n \leq t\}.
\]

**Theorem 1.5.5** The \( \epsilon \)-entropy \( H(\Xi, \epsilon^{1/r}S) \) of a \( p \)-ellipsoid \( \Xi \) with respect to the unit ball \( S = \{ \xi \in l_p : (\sum_{n=1}^{\infty} |\xi_n|^p)^{1/p} \leq 1\} \) satisfies the inequalities:

\[
m(2\epsilon) \log 4 + \int_0^1 \frac{m(t)}{t} dt \geq H(\Xi, \epsilon^{1/r}S) \geq \int_0^{\frac{1}{2\epsilon}} \frac{m(t)}{t} dt.
\]
Proof: See Mitjagin [1961], page 71.

In the case $p = 2$ we have some corollaries, but first let us define:

$$I(s) = \int_0^s \frac{m(t)}{t} dt.$$

**Corollary 1.5.6** The $\epsilon$-entropy of a 2-ellipsoid $\Sigma$ satisfies the inequalities:

$$m\left(\frac{2}{\epsilon}\right) \log(8/\epsilon) \geq H(\Sigma, \epsilon^{1/r} S) \geq m\left(\frac{1}{2\epsilon^2}\right).$$

**Proof:** See Mitjagin [1961], page 73.

**Corollary 1.5.7** The $\epsilon$-entropy of a 2-ellipsoid satisfies the inequalities:

$$I\left(\frac{8}{\epsilon}\right) \geq H(\Sigma, \epsilon^{1/r} S) \geq I\left(\frac{1}{2\epsilon}\right).$$

**Proof:** Refer to Mitjagin [1961], page 74.

We want to find a characterization of trace class operators in terms of entropy, our claim is the following:

**Theorem 1.5.8** An operator $A$ is trace class iff

$$\int_0^1 H(A(S), \epsilon S) d\epsilon < \infty$$

where $S$ is the unit ball.

The Theorem will be proved a little later. First, consider $A$ an operator. By Theorem 1.1.13 $A = W[A]$ where $[A] = (A^*A)^{1/2}$ and $W$ is a partially isometric operator with initial set the closure of the range of $[A]$ and final set the closure of the range of $A$ (see Definition 1.11).

So, if $S$ is the unit ball then $A(S) = W[A](S)$. Since $W$ is partially isometric the $\epsilon$-entropy is not altered if we consider $[A](S)$ instead of $A(S)$, that is, if $A(S) = \Xi$ and $[A](S) = \Xi_1$ then

$$H(\Xi, \epsilon^{1/r} S) = H(\Xi_1, \epsilon S) \quad \forall \epsilon > 0.$$

And in fact $\Xi$ is a $p$-ellipsoid for $p = 2$.

Choose a basis of eigenvectors $\{\varphi_j\}_{j \in J}$ of $[A]$ such that $[A] \varphi_j = \lambda_j \varphi_j$ where $\lambda_j \searrow 0$ as $j \to \infty$ (ordering the eigenvalues), then:

$$[A](S) = \{ x : \sum \frac{x_j^2}{\lambda_j^2} \leq 1 \}$$
if \( x = \sum_{j \in J} x_j \varphi_j \).

The operator \( [A] \) is trace class iff \( \sum_{j \in J} \lambda_j < \infty \). Now \( A \in \{r\} \) iff \([A] \in \{r\}\) by Lemma 1.4.3, so we have \( A \in \{r\} \) iff \( \sum_{j \in J} \lambda_j < \infty \).

From now on \( J \) will be the set of positive integers. And we define:

\[
a_n = \frac{1}{\lambda_n}.
\]

Since \( \sum_n \lambda_n < \infty \), then at most a finite number of \( a_n \leq 1 \), so without loss of generality we will assume that for every integer \( n \), \( a_n > 1 \). (Since \( \lambda_n \) is decreasing to 0, then \( a_n \) is increasing to \( \infty \)).

In order to prove Theorem 1.5.8 we need some lemmas. Let \( \{a_n\}_{n \geq 1} \) be a increasing sequence such that \( a_n > 1, \forall n \), and \( r > 0 \) such that \( \sum_n \frac{1}{a_n} \leq \infty \), then for \( m(t) \) as defined above we have:

**Lemma 1.5.9**

\[
\int_0^1 \int_0^{c/\lambda^r} \frac{m(t)}{t} dt \, d\lambda \, d\epsilon = \begin{cases} \int_1^c \frac{m(t)}{t} dt + c^r \int_0^\infty \frac{m(t)}{t^{r+1}} dt & \text{for } c \geq 1 \\ c^r \int_1^\infty \frac{m(t)}{t^{r+1}} dt & \text{for } c < 1. \end{cases}
\]

**Proof:** Since \( a_n > 1 \) then we can take \( m(t) = 0 \) for \( 0 < t \leq 1 \) from the definition of \( m(t) \) in 1.5.4.

1) If \( c \geq 1 \)

\[
\int_0^1 \int_0^{c/\lambda^r} \frac{m(t)}{t} dt \, d\lambda \, d\epsilon = (\text{by Fubini's theorem})
\]

\[
\int_1^c \frac{m(t)}{t} d\lambda \, d\epsilon + \int_c^\infty \frac{m(t)}{t^{r+1}} dt \, d\lambda \, d\epsilon = \int_1^c \frac{m(t)}{t} dt + c^r \int_0^\infty \frac{m(t)}{t^{r+1}} dt.
\]

2) If \( c < 1 \)

\[
\int_0^1 \int_0^{c/\lambda^r} \frac{m(t)}{t} dt \, d\lambda \, d\epsilon = \int_1^\infty \frac{m(t)}{t^{r+1}} dt \, d\lambda \, d\epsilon = c^r \int_1^\infty \frac{m(t)}{t^{r+1}} dt.
\]

\[\square\]

**Lemma 1.5.10**

\[
\frac{1}{2^r} \int_1^\infty \frac{m(t)}{t^{r+1}} dt \leq \int_0^1 H(\Xi, \epsilon^{1/r} S) \, d\epsilon \leq \int_1^8 \frac{m(t)}{t^{r+1}} dt + 8^r \int_8^\infty \frac{m(t)}{t^{r+1}} dt.
\]

24
Proof: By Corollary 1.5.7

\[ I\left(8/e^{1/r}\right) = \int_0^{8/e^{1/r}} \frac{m(t)}{t^r} dt - H(\Xi, e^{1/r} S) \geq \int_0^{1/2e^{1/r}} \frac{m(t)}{t^r} dt = I\left(1/2e^{1/r}\right) \]

so

\[ \int_0^1 \int_0^{8/e^{1/r}} \frac{m(t)}{t^r} dt \, d\epsilon \geq \int_0^1 H(\Xi, e^{1/r} S) \, d\epsilon \geq \int_0^1 \int_0^{1/2e^{1/r}} \frac{m(t)}{t^r} dt \, d\epsilon. \]

By Lemma 1.5.9 with \( c = 8 > 1 \) and \( c = 1/2 < 1 \) we have:

\[ \frac{1}{2r} \int_1^{\infty} \frac{m(t)}{t^{r+1}} dt \leq \int_0^1 H(\Xi, e^{1/r} S) \, d\epsilon \leq \int_0^8 \frac{m(t)}{t^r} dt + 8^{r} \int_0^\infty \frac{m(t)}{t^{r+1}} dt. \]

\[ \Box \]

Corollary 1.5.11

\[ \int_0^1 H(\Xi, e^{1/r} S) \, d\epsilon < \infty \quad \text{iff} \quad \int_1^\infty \frac{m(t)}{t^{r+1}} dt < \infty. \]

Proof:

\[ \int_1^8 \frac{m(t)}{t^r} dt < \infty \quad \text{since} \quad a_n \not\to \infty \]

and

\[ \int_0^\infty \frac{m(t)}{t^{r+1}} dt \leq \int_1^\infty \frac{m(t)}{t^{r+1}} dt. \]

\[ \Box \]

Lemma 1.5.12 Let \( a_n \) and \( m(t) \) be as defined above. Then:

\[ (1 - \frac{1}{2^r}) \sum_{k=0}^{\infty} \frac{m(2k)}{2^r k} \leq \sum_{k=0}^{\infty} \frac{1}{a_n^k} \leq (2^r - 1) \sum_{k=0}^{\infty} \frac{m(2k)}{2^r k}. \]

Proof: If we define:

\[ A_k = \{ n : 2^k < a_n \leq 2^{k+1} \}, \quad k \text{ an integer} \]

and remembering that every \( a_n \) was assumed to be greater than 1, we have:

\[ \sum_{n=1}^{\infty} \frac{1}{a_n^r} = \sum_{k=0}^{\infty} \sum_{n \in A_k} \frac{1}{a_n^r} \geq \sum_{k=0}^{\infty} \frac{1}{2^r(k+1)^r} \text{card} A_k = \]

\[ \sum_{k=0}^{\infty} \frac{m(2^{k+1}) - m(2^k)}{2^r(k+1)} = \sum_{k=0}^{\infty} \frac{m(2^{k+1})}{2^r(k+1)} - \frac{1}{2^r} \sum_{k=0}^{\infty} \frac{m(2^k)}{2^r k}. \]

25
\[
\sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} - \frac{1}{2^r} \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} = (1 - \frac{1}{2^r}) \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}}
\]

which gives the first inequality, similarly for the other inequality we have:

\[
\sum_{n=1}^{\infty} \frac{1}{a_n^r} = \sum_{n \in A_k} \frac{1}{a_n^r} \leq \sum_{k=0}^{\infty} \frac{1}{2^{rk}} \text{card}A_k =
\]

\[
\sum_{k=0}^{\infty} \frac{m(2^{k+1}) - m(2^k)}{2^{rk}} = 2^r \sum_{k=0}^{\infty} \frac{m(2^{k+1})}{2^{r(k+1)}} - \sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk}} =
\]

\[
2^r \sum_{k=1}^{\infty} \frac{m(2^k)}{2^k} - \sum_{k=0}^{\infty} \frac{m(2^k)}{2^k} = (2^r - 1) \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}}.
\]

\[\square\]

**Lemma 1.5.13** Let \(m(t)\) be as in Definition 1.5.4. Then

\[
\frac{(2^r - 1)}{r2^r} \sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk}} \leq \int_1^{\infty} \frac{m(t)}{t^{r+1}} dt \leq \frac{(2^r - 1)}{r} \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}}
\]

**Proof:**

\[
\int_1^{\infty} \frac{m(t)}{t^{r+1}} dt = \sum_{k=0}^{\infty} \int_{2^k}^{2^{k+1}} \frac{m(t)}{t^{r+1}} dt \quad \text{(since \(m(t)\) is increasing)}
\]

\[
\leq \sum_{k=0}^{\infty} m(2^{k+1}) \int_{2^k}^{2^{k+1}} \frac{dt}{t^{r+1}} = \sum_{k=0}^{\infty} m(2^{k+1}) \left( \frac{1}{r} \left[ \frac{1}{2^{k+1}} \right] \right)
\]

\[
= \frac{1}{r} \sum_{k=0}^{\infty} m(2^{k+1}) \left( \frac{1}{2^{rk}} - \frac{1}{2^{rk+1}} \right) = \frac{(2^r - 1)}{r} \sum_{k=0}^{\infty} \frac{m(2^{k+1})}{2^{r(k+1)}} = \frac{(2^r - 1)}{r} \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}}
\]

giving the right hand inequality. Similarly

\[
\int_1^{\infty} \frac{m(t)}{t^{r+1}} dt = \sum_{k=0}^{\infty} \int_{2^k}^{2^{k+1}} \frac{m(t)}{t^{r+1}} dt \geq \sum_{k=0}^{\infty} m(2^k) \int_{2^k}^{2^{k+1}} \frac{dt}{t^{r+1}}
\]

\[
= \frac{(2^r - 1)}{r} \sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk+1}} = \frac{(2^r - 1)}{r2^r} \sum_{k=0}^{\infty} \frac{m(2^k)}{2^k}.
\]

\[\square\]

**Lemma 1.5.14** \(\sum_{n=1}^{\infty} 1/a_n^n\) is convergent iff \(\int_1^{\infty} \frac{m(t)}{t^{r+1}} dt < \infty\).
Proof: Suppose that $\sum_{n=1}^{\infty} 1/a_n$ is convergent. By Lemma 1.5.12 we know that

$$(1 - \frac{1}{2^r}) \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} \leq \sum_{n=1}^{\infty} \frac{1}{a_n^r} < \infty$$

so $0 \leq \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} < \infty$ since $m(t) \geq 0$ for $t \geq 0$. Now by Lemma 1.5.13

$$\int_{1}^{\infty} \frac{m(t)}{t^{r+1}} dt \leq \frac{(2^r - 1)}{r} \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} < \infty.$$ 

For the other implication suppose $\int_{1}^{\infty} \frac{m(t)}{t^{r+1}} dt < \infty$, then by Lemma 1.5.13

$$\frac{(2^r - 1)}{r2^r} \sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk}} \leq \int_{1}^{\infty} \frac{m(t)}{t^{r+1}} dt < \infty$$

so that $\sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk}} < \infty$ and by Lemma 1.5.12

$$\sum_{n=1}^{\infty} \frac{1}{a_n^r} \leq (2^r - 1) \sum_{k=1}^{\infty} \frac{m(2^k)}{2^{rk}} \leq (2^r - 1) \sum_{k=0}^{\infty} \frac{m(2^k)}{2^{rk}} < \infty$$

which finishes the proof.

Lemma 1.5.15 $\int_{0}^{1} H(\Xi, \epsilon^{1/2} S) d\epsilon < \infty$ iff $\sum_{n} \frac{1}{a_n^r} < \infty$.

Proof:

$$\int_{0}^{1} H(\Xi, \epsilon^{1/2} S) d\epsilon < \infty \iff \int_{1}^{\infty} \frac{m(t)}{t^{r+1}} dt < \infty \iff \sum_{n} \frac{1}{a_n^r} < \infty$$

applying Corollary 1.5.11 and Lemma 1.5.14.

Now we can easily prove Theorem 1.5.8, just take $r = 1$ in the previous Lemma and $A(S)$ as the ellipsoid $\Xi$. And even more we have also the following:

Theorem 1.5.16 $A$ is a Schmidt operator iff $\int_{0}^{1} H(A(S), \epsilon^{1/2} S) d\epsilon < \infty$.

Proof: Apply Lemma 1.5.15 with $A(S) = \Xi$ and $r = 2$. 

27
Chapter 2

Trace Class Operators

2.1 Kernels of Trace Class Operators

In the first chapter we recalled the definition of a trace class operator and we gave the main results about the space of trace class operators, as well as a new characterization of trace class operators in terms of their $\epsilon$-entropy.

In this section we list some of the known conditions on $K(x, y)$ that make the integral operator

$$(TKf)(x) = \int_a^b K(x, y)f(y)dy$$

where $K \in L^2([a, b]^2)$ and $f$ is any function in $L^2([a, b])$ a trace class operator (also known in mathematical literature as "nuclear operator").

We first begin with some basic definitions:

**Definition 2.1.1** Let $K(x, y) \in L^2([a, b]^2)$.

a) $K(x, y)$ is nonnegative definite if for every $f \in L^2([a, b])$

$$\int_a^b \int_a^b K(x, y)f(x)f(y)dxdy \geq 0.$$

b) $K(x, y)$ is symmetric or Hermitian iff $K(x, y) = \overline{K(y, x)}$ for each $x$ and $y$ in $[a, b]$.

c) $K(x, y) \in \text{Lip}_\alpha$ in the variable $x$ iff

$$\Delta K := |K(x + h, y) - K(x, y)| \leq |h|^{\alpha}A(y)$$

where $A(y)$ is nonnegative, $A \in L^2([a, b])$ and $x, y \in [a, b]$. (Likewise define $K(x, y) \in \text{Lip}_\alpha$ in the variable $y$).

d) $K(x, y) \in \text{Lip}_{(\alpha, p)}$ in the variable $x$ if its $L^p$ modulus of continuity $\omega_p(K, \delta)$ satisfies:

$$\omega_p(K, \delta) := \sup_{0 < h \leq \delta} \left[ \int_a^b |(\Delta K)^p|dz \right]^{1/p} < \delta^\alpha A(y)$$

where $\Delta$ and $A$ are as in c). (Likewise define $K(x, y) \in \text{Lip}_{(\alpha, p)}$ in the variable $y$).
The first sufficient condition we mention is due to Mercer [1909], and probably is the best known among the rest.

**Theorem 2.1.2** (Mercer) \( T_K \) is a trace class operator if \( K(x, y) \) is continuous, symmetric and non-negative (positive) definite.

**Proof:** Refer to Courant [1953], page 138.

Other known results are:

**Theorem 2.1.3** (Chang) \( T_K \) is a trace class operator iff

\[
K(x,y) = \int_a^b K_1(x,z)K_2(z,y)dz
\]

where \( K_1 \) and \( K_2 \in L^2([a,b]^2) \).

**Proof:** See Chang [1947], this also follows from the definition 1.4.2 of trace class operators, Theorem 1.3.3 and the remarks before it. \( \square \)

**Theorem 2.1.4** (Smithies) \( T_K \) is a trace class operator if \( K(x, y) \in \text{Lip}(\alpha, p) \) in either of its variables, with \( p \geq 1 \) and \( \alpha \geq 1/\min(2,p) \).

**Proof:** See Smithies [1937].

**Theorem 2.1.5** (Smithies-Stinespring) \( T_K \) is a trace class operator if \( K(x, y) \in \text{Lip}_\alpha \) in either of its variables and \( \alpha > 1/2 \).

**Proof:** Refer to Stinespring [1958].

**Theorem 2.1.6** (Cochran) \( T_K \) is a trace class operator if \( K(x, y) \in \text{Lip}_{(\alpha, p)} \) in \( x \) or \( y \) and \( K(x, y) \in \text{Lip}(\beta, q) \) in \( y \) or \( x \) respectively, for some \( 1 \leq p \leq q \) with:

i) for \( \alpha \leq \beta \), \( \beta > 1/\min(2,q) \);

ii) for \( \beta < \alpha \leq \beta + 1/p - 1/q \),

\[
\begin{cases}
\beta > 1/q & q \leq 2 \\
\alpha p(q-2) + \beta q(2-p) > q - p & p \leq 2 < q \\
\alpha > 2 & p > 2;
\end{cases}
\]

iii) and for \( \beta + 1/p - 1/q \), \( \alpha > 1/\min(2,p) \).
Proof: See Cochran [1977].

Previous results in the direction of the last theorem can be found in Cochran [1976,a], [1976,b], Steel [1977] and Cheng [1942].

Theorem 1.5 proves to be very useful because, as we will see later in this chapter, it will lead us to an important representation of $K(x, y)$ in terms of a convergent series in $L^2(\mu \otimes \mu)$.

Remark 2.1.7 All the previous definitions and theorems remain valid for

$K \in L^2([a, b]^2, \mu \otimes \mu)$

and the operator

$$(Tkf)(x) = \int_a^b K(x, y)f(y)d\mu(x)$$

where $\mu$ is a measure on $[a, b]$.

For more information about this remark, see Stinespring [1958].

Since in further parts of this thesis we will refer to the set of functions which are measurable and square $\mu$-integrable in $[0,1]$ rather than $[a, b]$, we will adopt the following notation:

$L^2(\mu \otimes \mu) = L^2([0,1]^2, \mu \otimes \mu)$ and $L^2(\mu) = L^2([0,1], \mu)$

keeping in mind that this just simplifies notation.

Now if we suppose that $K \in L^2(\mu \otimes \mu)$ and it satisfies the condition of Theorem 2.1.5, we define

$$\|K(x, \cdot) - K(z, \cdot)\|_2 = (\int_0^1 |K(x, y) - K(z, y)|^2d\mu(y))^{1/2}$$

and we let $M^2 = \int_0^1 A^2(y)d\mu(y)$ where $A$ is the non-negative, square-$\mu$ integrable function given in the definition of $\text{Lip}_\alpha$. Then for $h < (\epsilon/M)^{1/\alpha}$ we have: that if $|x - x_0| \leq h$

$$\|K(x, \cdot) - K(x_0, \cdot)\|_2^2 = \int_0^1 |K(x, y) - K(x_0, y)|^2d\mu(y) \leq \int_0^1 h^{2\alpha}A^2(y)d\mu(y) = M^2h^{2\alpha} < \epsilon^2$$

which proves the following:

Lemma 2.1.8 If $K \in L^2(\mu \otimes \mu)$ and $K \in \text{Lip}_\alpha$ for some $\alpha > 1/2$ then:

i) $K$ is the kernel of a trace class operator, and

ii) $\forall \epsilon > 0 \exists N(\epsilon) = N$ and $x_1, \ldots, x_N \in [0,1]$ such that $\forall x \in [0,1]$ $\exists j \in \{1, \ldots, N\}$ that satisfies $\|K(x, \cdot) - K(x_j, \cdot)\|_2 < \epsilon$, and even more $N$ can be taken such that $N \leq C/\epsilon^{1/\alpha}$ where $C$ is a constant.
2.2 Series Representation of Trace Class Kernels

In this section we introduce a sufficient condition for a representation of a kernel $K(z, y)$ in terms of a series of products of functions in $L^2$, that will be shown to be equivalent to the trace class property at the end of this section.

**Theorem 2.2.1** Let $K(z, y) \in L^2(X \otimes X, P \otimes P)$ where $(X, \mathcal{A})$ is a measurable space and $P$ is a probability measure in it. Assume that:

(*) For every $\epsilon > 0$ there exists $N(\epsilon) = N$ and $x_1, \ldots, x_N \in X$ such that for any $x \in X$ there exists $j \in \{1, \ldots, N\}$ such that $\|K(x, \cdot) - K(x_j, \cdot)\|_2 < \epsilon$, where $N \leq C/\epsilon^r$ for some $r < 2$.

Then $K(x,y)$ can be written as:

$$K(x, y) = \sum_j f_j g_j$$

where the series converges in $L^2$, $f_j, g_j \in L^2(P)$ and they satisfy:

$$\sum_j \|f_j\|_2 \|g_j\|_2 < \infty.$$  

**Proof:** Consider $\epsilon_k = 1/2^k$, $k \in \{0,1,\ldots\}$, and let $B_0 = \{x_{1,0},\ldots,x_{N(0),0}\} \subset X$ such that (*) holds for $\epsilon_0 = 1$ and we define $N(0) := N(\epsilon_0)$, so $N(0) \leq C$.

In general for $k \in \mathbb{N}$ let $B(k) = \{x_{1,k},\ldots,x_{N(k),k}\} \subset X$ such that (*) is satisfied for $\epsilon_k = 1/2^k$ and we define $N(k) := N(\epsilon_k)$. Then $N(k) \leq C2^{kr}$.

Now for $k \in \mathbb{N}$ fixed, we define $h_k : X \to \{1,\ldots,N(k)\}$ as follows:

$$h_k(x) = \begin{cases} 
  j & \text{if } x_{j,k} \in B_k \text{ is the closest element of } B_k \text{ to } x \\
  \text{i.e. } d(x, x_{j,k}) = \min_{i \in \{1,\ldots,N(k)\}} d(x, x_{i,k}) \\
  \text{and take the smallest } j \text{ if ties occur.} 
\end{cases}$$

Let $A_{j,k} = \{x \in X|h_k(x) = j\}$ for $j = 1,\ldots,N(k)$. If we ask $N(k)$ to be minimal, then we have that $\{A_{j,k}\}_{j=1,\ldots,N(k)}$ are non-empty disjoint sets whose union is $X$ by (*), therefore a partition of $X$. Now for $k$ fixed let:

$$x_{(k)} := x_{h_k,k}$$

$$x_{(k,\ldots,i)} := (\ldots(x_{(k)})(k-1)\ldots)(i) \text{ for } k > i \geq 0.$$ 

And now consider the following sequence for $x \in X$:

$$X \to B_k \to B_{k-1} \to \cdots \to B_0 \text{ such that}$$

31
and define

\[ f_{k,0} : X \to B_0 \text{ such that } f_{k,0}(x) = x_{(k,...,0)}, \]

then

\[
d(x, f_{k,0}(x)) \leq d(x, x_{(k)}) + d(x_{(k)}, x_{(k,k-1)}) + \cdots + d(x_{(k,...,1)}, x_{(k,...,0)})
\]

\[
\leq 1/2^k + 1/2^{k-1} + \cdots + 1 \leq \sum_{k=0}^{\infty} 1/2^k = 2
\]

for each \( k \geq 0 \).

Now suppose \( k > i \geq 0 \) and take \( x \in X \) and consider the sequence

\[ X \to B_k \to B_{k-1} \to \cdots \to B_i \text{ such that}
\]

\[ x \to x_{(k)} \to x_{(k,k-1)} \to \cdots \to x_{(k,...,i)} \]

and define

\[ f_{k,i} : X \to B_i \text{ such that } f_{k,i}(x) = x_{(k,...,i)}, \]

then

\[
d(x, f_{k,i}(x)) \leq d(x, x_{(k)}) + \cdots + d(x_{(k,...,i+1)}, x_{(k,...,i)})
\]

\[
\leq 1/2^k + 1/2^{k-1} + \cdots + 1/2^i \leq \sum_{i=0}^{\infty} 1/2^i = 2/2^i.
\]

Note: From the definition of \( f_{k,i} \), it is easy to see that:

if \( k > i > l \geq 0 \) then \( f_{i,l}(f_{k,i}(x)) = f_{k,l}(x) \quad \forall x \in X. \)

Now consider a fixed \( x \in X \) and the sequence \( \{f_{k,0}(x)\}_{k=0}^{\infty} \). The result is a sequence that takes values in a finite set \( B_0 \), so there must be a subsequence \( \{f_{k,n,0}(x)\}_{n=1}^{\infty} \) such that for each \( m, n \in \mathbb{N} \)

\[ f_{k(n),0}(x) \equiv f_{k(n),0}(x), \]

an element in \( B_0 \), therefore we can define a function \( f_0 : X \to B_0 \) by:

\[ f_0(x) = \begin{cases} 
\text{the least } j \in \{1, \ldots, N(0)\} \text{ for which } f_{k,0}(x) = x_{j,0} \\
\text{for infinitely many } k's.
\end{cases} \]

This function is well defined by the note above.

Let \( C_{j,0} = \{x \in X | f_0(x) = j\}, \quad j \in R(f_0) = \text{ range of } f_0. \)

Then \( \{C_{j,0}\}_{j \in R(f_0)} \) is a partition of \( X. \)
For fixed \( x \in X \) consider the subsequence \( k(n) \) such that

\[
f_0(x) = f_{k(n),0}(x) = f_{k(m),0}(x) \quad \forall m, n \in \mathbb{N}
\]

and now take the sequence \( \{f_{k(n),1}(x)\}_{n=1}^{\infty} \), which is a sequence that takes values in \( B_1 \), a finite set. Therefore, there must be a subsequence of \( k(n) \), let us say \( k(n, m) \), such that for each \( l, m \in \{1, 2, \ldots\} \)

\[
f_{k(n,m),1}(x) = f_{k(m),1}(x),
\]

an element of \( B_1 \). So we can define a function \( f_1 : X \to B_1 \) by:

\[
f_1 = \left\{ \begin{array}{ll}
\text{the least } j \in \{1, \ldots, N(1)\} & \text{for which } f_{k(n),1}(x) = x_{j,1} \\
\text{for infinitely many } k(n)'s & 
\end{array} \right.
\]

and let \( C_{j,1} = \{x \in X | f_1(x) = j\} \), \( j \in R(f_1) = \text{range of } f_1 \). Then \( \{C_{j,1}\}_{j \in R(f_1)} \) is a partition of \( X \). In the same way we define inductively \( C_{j,2}, C_{j,3}, \ldots \). Now we want to prove that \( \{C_{j,k}\}_{j \in R(f_k)} \) is a subpartition of \( \{C_{j,k-1}\}_{j \in R(f_{k-1})} \); we just check this statement for \( k = 1 \). Consider \( j \in R(f_1) \) and \( x, y \in C_{j,1} \), we want to find \( i \in R(f_0) \) such that \( x, y \in C_{i,0} \). First for \( x \) consider the subsequence \( k(n) \) such that:

\[
\forall m, n \in \{1, 2, \ldots\} \quad f_{k(n),0}(x) = f_{k(m),0}(x) = f_0(x)
\]

and for \( y \) consider the subsequence \( l(n) \) such that:

\[
\forall m, n \in \{1, 2, \ldots\} \quad f_{l(n),0}(x) = f_{l(m),0}(x) = f_0(x).
\]

Since \( x, y \in C_{j,1} \) then \( f_1(x) = f_1(y) = j \in B_1 \). Then take the subsequence of \( k(n), k(n, m) \) such that \( f_{k(n,m),1}(x) = x_{j,1} \quad \forall m \in \{1, 2, \ldots\} \). And the subsequence of \( l(n), l(n, m) \) such that \( f_{l(n,m),1}(y) = x_{j,1} \quad \forall m \in \{1, 2, \ldots\} \).

Then for each \( m \in \{1, 2, \ldots\} \)

\[
f_{l(n,m),1}(y) = f_{k(n,m),1}(x) = x_{j,1},
\]

and by the note above we have:

\[
f_{k(n,m),0}(x) = f_{1,0}(f_{k(n,m),1}(x)) = f_{1,0}(x_{j,1}) = f_{1,0}(f_{l(n,m),1}(y)) = f_{l(n,m),1}(y). \quad (**)
\]

But \( f_{1,0}(x_{j,1}) = x_{(1,0)} \) and for each \( n \in \{1, 2, \ldots\} \)

\[
f_{k(n),0}(x) = f_0(x).
\]

Therefore, \( \forall m \in \{1, 2, \ldots\} \)

\[
f_{k(n,m),0}(x) = f_0(x).
\]

In the same way \( \forall m \in \{1, 2, \ldots\} \)

\[
f_{l(n,m),0}(y) = f_0(y) \quad \text{and by (**)} \text{ we have:}
\]

\[
f_0(x) = f_0(y) = x_{(1,0)}.
\]

And so, \( x \) and \( y \) belong to the same \( C_{i,0} \) for some \( i \in R(f_0) \). Therefore, \( \forall j \in R(f_1) \) \( C_{j,1} \subset C_{i,0} \) for some \( i \in R(f_0) \), then \( \{C_{j,i}\}_{j \in R(f_1)} \) is a refinement of the
partition \(\{C_{j,0}\}_{j \in R(f_0)}\). In a similar way we can check that \(\{C_{j,k}\}_{j \in R(f_k)}\) is a refinement of the partition \(\{C_{j,k-1}\}_{j \in R(f_{k-1})}\). Since \(R(f_k) \subset B_k\) then:

\[
\text{card} R(f_k) \leq \text{card} B_k \leq C 2^k
\]

from the definition of \(B_k\) and (*)

Now for the partition \(\{C_{j,k}\}_{j \in R(f_k)}\) of \(X\), let

\[
K_k(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)K(x, y)
\]

and since \(\{C_{j,k}\}\) is a partition then

\[
K(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)K(x, y)
\]

so,

\[
K(x, y) - K_k(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)(K(x, y) - K(x, k, y))
\]

And if \(x \in C_{j,k}\) then \(f_k(x) = j\), so \(j\) is the least value in the set \(\{1, 2, \ldots, N(k)\}\) for which \(f_k(n_0, n_1, \ldots, n_{k-1}, k)(x) = x_{j,k}\) (where \(f_k(n_0, n_1, \ldots, n_{k-1}, k)(x)\) is the inductive generalization for \(k \geq 2\) mentioned before) for infinitely many values.

So

\[
K_k(x, y) - K_{k-1}(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)K(x, y) - \sum_{j=1}^{N(k-1)} 1_{C_{j,k-1}}(x)K(x, y, k-1, y).
\]

But \(C_{j,k-1} = \bigcup C_{i,k}\) where the union runs over all the \((i, k)\) such that \(C_{i,k} \subset C_{j,k-1}\). Then

\[
K_k(x, y) - K_{k-1}(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)K(x, y) - \sum_{j=1}^{N(k-1)} \sum_{\{(i, k) : C_{i,k} \subset C_{j,k-1}\}} 1_{C_{i,k}}(x)K(x, y, k-1, y).
\]

The term in parentheses is a function depending on \((j, k)\), let us say \(g_{j,k}\).

So \(K_k(x, y) - K_{k-1}(x, y) = \sum_{j=1}^{N(k)} 1_{C_{j,k}}(x)g_{j,k}(y)\), and by construction \(\|g_{j,k}\|_2^2 \leq (2/2^k)^2\).

Then

\[
\sum_{j=1}^{N(k)} \|1_{C_{j,k}}(x)\|_2 \|g_{j,k}\|_2
\]
\[
\leq (\sum_{j=1}^{N(k)} ||1_{C_{j,k}}(x)||_2^2)^{1/2} (\sum_{j=1}^{N(k)} ||g_{j,k}||_2^2)^{1/2}
\leq (\sum_{j=1}^{N(k)} P(C_{j,k}))^{1/2} (\sum_{j=1}^{N(k)} (1/2^{k-1})^2)^{1/2}
\leq 1 \cdot (C2^{hr} \cdot 1/2^{2k-2})^{1/2} \leq M(\frac{1}{2})^{(2-r)k/2}.
\]

Here we use Cauchy's inequality, the fact that \( P \) is a probability measure and the bound for \( N(k) \) and the fact that \( M \) is a positive constant.

So \( \sum_{k=1}^{\infty} \sum_{j=1}^{N(k)} ||1_{C_{j,k}}(x)||_2 ||g_{j,k}||_2 \leq M \sum_{k=1}^{\infty} \rho^k \), where \( \rho = \frac{1}{4}(2-r)/2 \).

And since \( (2-r)/2 > 0 \) we have:

\[
\sum_{k=1}^{\infty} \sum_{j=1}^{N(k)} ||1_{C_{j,k}}(x)||_2 ||g_{j,k}(y)||_2 < \infty.
\]

And finally we let \( k \to \infty \), so that \( K_k \to K \) in \( L^2 \).

As an immediate consequence of this Theorem we have the following:

**Corollary 2.2.2** If \( K(x,y) \in L^2([0,1]^2, P \otimes P) \) where \((X,A)\) is a measurable space and \( P \) is a probability measure in it, and \( K(x,y) \in Lip_{\alpha} \) for some \( \alpha > 1/2 \), then the conclusion of Theorem 2.2.1 follows.

**Proof:** First apply Lemma 2.1.8 and then Theorem 2.2.1.

**Example** Suppose \( X \) is a random variable with an arbitrary, continuous distribution function \( F(x) = P(X \leq x) \) and let:

\[
\Omega_0 = \{ F : F \text{ is absolutely continuous and } F(0) = \frac{1}{2}, \text{ uniquely} \}.
\]

Consider random samples \( X_1, \ldots, X_n \) from \( F(x - \theta_x) \) and \( Y_1, \ldots, Y_n \) from \( F(y - \theta_y) \), \( F \in \Omega_0 \). And let \( \Delta = \theta_y - \theta_x \). If we want to test \( H_0 : \Delta = 0 \) vs \( H_1 : \Delta > 0 \) we can use the Mann-Whitney-Wilcoxon rank statistic \( W \) defined by:

\[
W = \int F_n(y) dG_n(y)
\]

where \( F_n \) and \( G_n \) are the empirical distribution functions of \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \) respectively, but:

\[
\int F_n(y) dG_n(x) = \int \int 1_{x \leq y}(x,y) dF_n(x) dG_n(y)
\]

where \( 1_{x \leq y} \) is the indicator function of the set \( \{(x,y) : x \leq y \} \), so if we define \( K(x,y) = 1_{x \leq y} \) we have:

\[
W = \int \int K(x,y) dF_n(x) dG_n(y).
\]
For more information about this topic see Hettmansperger [1984], page 132. So, we are interested in the behaviour of the kernel \( K(x, y) = 1_{0 \leq x \leq y \leq 1} \), as a particular case. First we notice that we can expand \( K \) in the following form:

\[
K(x, y) = 1_{[0,1/2]}(x)1_{[1/2,1]}(y) + 1_{[0,1/4]}(x)1_{[1/4,1/2]}(y) + 1_{[1/2,3/4]}(x)1_{[3/4,1]}(y) + \cdots
\]

and each of the indicator functions in this expression are clearly elements of \( L^2 \), but

\[
\sum \| \cdot \|_2 = \sqrt{1/2/1/2} + 2(\sqrt{1/4/1/4}) + \cdots = 1/2 + 1/2 + 1/2 + \cdots = \infty.
\]

So, even if \( K(x, y) \) has a representation as a sum of products of \( L^2 \) functions the conclusion of Theorem 2.2.1 does not follow. It is then natural to ask if the integral operator generated by \( K(x, y) \) is a trace class operator. The answer is no, \( A_K \), the integral operator generated by \( K \), is not trace class, as we will observe next:

Let \( \varphi_j(x), \ j \in J \), be a subset of the usual Fourier system, and denote by \( A_K \) the integral operator generated by \( K \), then we have:

\[
(A_K \varphi_j, \varphi_j) = \int \int 1_{0 \leq x \leq y \leq 1}(x, y)\varphi_j(x)\overline{\varphi_j(y)}dx\,dy = \int_0^1 \int_0^y e^{i2\pi(x-y)}dx\,dy
\]

\[
= \int_0^1 \int_0^y \cos(2\pi j(x - y)) + i \sin(2\pi j(x - y))dx\,dy
\]

\[
= \int_0^1 \frac{1}{2\pi j} \sin(2\pi jy) - \frac{i}{2\pi j}(1 - \cos(2\pi jy))dy
\]

\[
= -\frac{i}{2\pi j},
\]

then

\[
\sum_{j=1}^{\infty} (A_K \varphi_j, \varphi_j) = \sum_{j=1}^{\infty} \frac{-i}{2\pi j}
\]

which is not convergent, so \( A_K \) is not a trace class operator.

Now for the same \( K(x, y) \) if we fix two points \( x_j \) and \( x_i \) in \([0,1]\)

\[
K(x_i, y) = \begin{cases} 
1 & \text{for } y \geq x_i, \\
0 & \text{for } y < x_i,
\end{cases}
\]

\[
K(x_j, y) = \begin{cases} 
1 & \text{for } y \geq x_j, \\
0 & \text{for } y < x_j.
\end{cases}
\]

So, if we assume that \( 0 \leq x_i \leq x_j \leq 1 \)

\[
K(x_i, y) - K(x_j, y) = \begin{cases} 
1 & \text{for } x_i \leq y < x_j, \\
0 & \text{otherwise}.
\end{cases}
\]
Therefore,

\[ \|K(x_i, \cdot) - K(x_j, \cdot)\|_2 = \left( \int_0^1 (K(x_i, y) - K(x_j, y))^2 dy \right)^{1/2} = \left( \int_{x_i}^{x_j} 1 dy \right)^{1/2} = \sqrt{x_j - x_i} \]

and then for any \( \epsilon > 0 \)

\[ \|K(x_i, \cdot) - K(x_j, \cdot)\|_2 < \epsilon \iff \sqrt{|x_i - x_j|} < \epsilon \iff |x_i - x_j| < \epsilon^2 \]

which shows that \( r < 2 \) in condition (*) of Theorem 2.2.1 is sharp.

Now we present the following:

**Theorem 2.2.3** If \( K(x, y) \in L^2(\mu \otimes \mu) \) and \( A_K \) is the integral operator with kernel \( K \), then:

\( A_K \) is trace class if \( \exists f_j, g_j \in L^2(\mu) \) such that \( K(x, y) = \sum_j f_j(x)g_j(y) \) and \( \sum_j \|f_j\|_2 \|g_j\|_2 < \infty \).

**Proof:** Suppose \( A_K \) is trace class. By Lemma 1.4.3 \( [A_K] \) is also trace class. So if

\[ A_K = \sum_j \lambda_j \varphi_j \otimes \overline{\psi_j} \]

then by Lemma 1.1.7

\[ [A_K] = \sum_j |\lambda_j| \varphi_j \otimes \overline{\psi_j} \]

and then

\[ K(x, y) = \sum_j |\lambda_j| \varphi_j(x) \overline{\psi_j(y)} \]

and

\[ \sum_j \|g_j\|_2 \|\psi_j\|_2 < \infty. \]

Now assume that \( \exists f_j, g_j \in L^2(\mu) \) such that \( K(x, y) = \sum_j f_j(x)g_j(y) \) where

\[ \sum_j \|f_j\|_2 \|g_j\|_2 < \infty. \]

For each \( j \)

\[ \tau(f_j \otimes g_j) = \|f_j\|_2 \|g_j\|_2. \]

So for each \( N \) fixed positive integer we have:

\[ \sum_{j=1}^N f_j \otimes g_j \]

is a trace class operator and since \( (\tau c) \) is complete with respect to the norm \( \tau(\cdot) \) then \( A_K \) is trace class. \( \Box \)
2.3 Convergence a.s. and in $L^1$

Suppose $K(\cdot, \cdot)$ in $L^2(P \otimes P)$ is such that for every $\epsilon > 0$ there exists $N(\epsilon) = N$ and $x_1, \ldots, x_N \in X$ such that for any $x \in X$ there exists $j \in \{1, \ldots, N\}$ such that $\|K(x, \cdot) - K(x_j, \cdot)\|_2 < \epsilon$ where $N \leq C/\epsilon^r$ for some $r < 2$ and $C$ a positive constant.

We proved before (Theorem 2.2.1) that $K(x, y)$ can be written as:

$$K(x, y) = \sum_{k=1}^{\infty} h_k(x) j_k(y)$$

where $\sum_k \|h_k\|_2 \|j_k\|_2 < \infty$. Let

$$f_k(x) = \frac{h_k(x)}{\|h_k\|_2} \quad \text{and} \quad g_k(y) = \frac{j_k(y)}{\|j_k\|_2}$$

so if $\alpha_k = \|h_k\|_2 \|j_k\|_2$, then

$$K(x, y) = \sum_{k=1}^{\infty} \alpha_k f_k(x) g_k(y) \quad \text{where} \quad \sum_{k} \alpha_k < \infty$$

and for every $k$, $\|f_k\|_2 = \|g_k\|_2 = 1$. Furthermore we have the following:

Lemma 2.3.1 The series $\sum_{k=1}^{\infty} \beta_k f_k(x) g_k(y)$ converges absolutely for all $x, y$ such that $\sum_k \beta_k f_k^2(x) < \infty$ and $\sum_k \beta_k g_k^2(y) < \infty$.

Proof: Let $m \geq n \geq 1$ integers, then

$$|\sum_{k=n}^{m} \beta_k f_k(x) g_k(y)|^2 = |\sum_{k=n}^{m} \sqrt{\beta_k} f_k(x) \sqrt{\beta_k} g_k(y)|^2$$

$$\leq (\sum_{k=n}^{m} \beta_k f_k^2(x))(\sum_{k=n}^{m} \beta_k g_k^2(y))$$

by Cauchy's inequality following the proof of Mercer's Theorem as in Courant[1953], page 138. Therefore, by Cauchy's criteria if $\sum_k \beta_k f_k^2(x) < \infty$ and $\sum_k \beta_k g_k^2(y) < \infty$ the series $\sum_{k=1}^{\infty} \beta_k f_k(x) g_k(y)$ converges absolutely. \qed

Now if we take $\beta_k = \alpha_k$ and $f_k, g_k$ as defined before the last lemma, we have:

The series $\sum_k \alpha_k f_k^2(x)$ converges a.s. by monotone convergence, and in $L^1$, since $\int f_k^2(x) dP(x) = 1$ for all $k$, $\alpha_k > 0$ and $\sum_k \alpha_k < \infty$.

Using the same argument we can prove that $\sum_k \alpha_k g_k^2(y)$ converges a.s. and in $L^1$, so we have proved the next:
Lemma 2.3.2 If $K(\cdot, \cdot)$ is the kernel of a trace class operator and $\alpha_k, f_k, g_k$ are defined as above; then
\[
\sum_k \alpha_k f_k(x)g_k(y) \text{ converges almost surely and in } L^2 \text{ to } K(x,y).
\]
(In fact the series converges absolutely, not necessarily to $K(x,y)$ however).

Since we know already that $\sum_{j=1}^{\infty} \alpha_j f_j(x)g_j(x)$ converges almost surely and in $L^2$ to $K(x,y)$, the next thing we want to know is when we have convergence of the series above in the diagonal $D = \{(x,y) \mid x = y\}$. Since this set may have measure 0, we could have the series not being convergent on $D$. Therefore, we want to find conditions on $K$ that guarantee such convergence. In order to find adequate conditions we first partition the area around $D$ into little squares. (Here we assume again that $K$ is defined in $[0,1]^2$.) Let us define:

\[
I_{m,1} = [0, 1/2^m]
\]
\[
I_{m,k} = [(k-1)/2^m, k/2^m] \quad k = 2, 3, \ldots, 2^m
\]

for every fixed $m$ positive integer. Now consider a fixed $t \in [0,1]$ and fixed $m$. Let $k(t) = k_t$ be the integer in \{1, 2, \ldots, $2^m$\} such that $t \in I_{m,k_t}$. Then clearly

\[
\{t\} = \bigcap_{m=1}^{\infty} I_{m,k_t}.
\]

We define for $m$ fixed and $t \in I_{m,k_t}$

\[
K_m(t) = 4^m \int_{I_{m,k_t}} \int_{I_{m,k_t}} K(x,y) \, dx \, dy,
\]

that is, $K_m$ is the average of $K(x,y)$ on $I_{m,k_t} \times I_{m,k_t}$.

If we assume $K$ is continuous on $[0,1]^2$ then we can assume also that $f_j$ and $g_j$ are continuous, and in this case we have that $K_m(t)$ converges to $K(t,t)$ as $m \to \infty$. Now let us define analogously:

\[
f_{j,m}(t) = 2^m \int_{I_{m,k_t}} f_j(x) \, dx,
\]
\[
g_{j,m}(t) = 2^m \int_{I_{m,k_t}} g_j(x) \, dx.
\]

Then we also have as $m \to \infty$ that $f_{j,m}(t) \to f_j(t)$ and $g_{j,m}(t) \to g_j(t)$.

For each $m$ fixed, $f_{j,m}$ and $g_{j,m}$ are step functions, which are constant on each $I_{m,k_t}$, and by continuity of $f_j$ and $g_j$, \{$f_{j,m}\}_{m=1}^{\infty}$ and \{$g_{j,m}\}_{m=1}^{\infty}$ converge uniformly to $f_j$ and $g_j$ respectively. Therefore the product $\alpha_j f_{j,m}(x)g_{j,m}(y)$ converges uniformly to
\( \alpha_j f_j(x)g_j(y) \), in particular any finite sum of the form \( \sum_{j=1}^{l} \alpha_j f_{j,m}(x)g_{j,m}(y) \) converges uniformly to \( \sum_{j=1}^{l} \alpha_j f_j(x)g_j(y) \), that is:

\[
\sum_{j=1}^{l} \alpha_j \int_{I_{m,k_1}} \int_{I_{m,k_2}} f_j(x)g_j(y) \, dx \, dy
\]

converges uniformly to \( \sum_{j=1}^{l} \alpha_j f_j(x)g_j(y) \) for each \( l \) positive integer. To prove the almost sure convergence we first notice that

\[
\sum_{j} \alpha_j (f_j - f_{j,m})^2(t) \xrightarrow{m \to \infty} 0
\]

in \( L^1 \) since \( \sum_{j} \alpha_j E f_j^2 < \infty \) and \( E(f_j - f_{j,m})^2 \leq E f_j^2 \) since \( f_j - f_{j,m} \) \perp f_{j,m} \) and \( E(f_j - f_{j,m})^2 \xrightarrow{m \to \infty} 0 \) by uniform convergence, for each \( j \). So we have dominated convergence for sums.

Now, we can take a subsequence \( m_k \) which goes to \( \infty \) fast enough so the \( L^1 \) norm convergence implies almost sure convergence. Then we have for almost all \( t \)

\[
\{ \sqrt{\alpha_j f_{j,m}(t)} \}_{j \geq 1} \xrightarrow{m \to \infty} \{ \sqrt{\alpha_j f_j(t)} \}_{j \geq 1} \quad \text{in} \quad L^2
\]

and likewise

\[
\{ \sqrt{\alpha_j g_{j,m}(t)} \}_{j \geq 1} \xrightarrow{m \to \infty} \{ \sqrt{\alpha_j g_j(t)} \}_{j \geq 1} \quad \text{in} \quad L^2.
\]

then since the inner product is jointly continuous, we have

\[
K(t, t) \leftarrow K_m(t, t) = \sum_j \alpha_j f_{j,m}(t)g_{j,m}(t) \rightarrow \sum_j \alpha_j f_j(t)g_j(t)
\]

as \( m \to \infty \). So we have proved the following:

**Theorem 2.3.3** Let \( K(x, y) \) defined on \([0,1]^2\) be as in Lemma 2.3.2 and a continuous function. Then we have

\[
K(x, x) = \sum_j \alpha_j f_j(x)g_j(x)
\]

and the series converges almost surely and in \( L^1 \).

Here we notice that for a given \( P \) if we define

\[
K_m(t) = \frac{1}{(P \times P)(I_{m,k_1} \times I_{m,k_2})} \int_{I_{m,k_1}} \int_{I_{m,k_2}} K(x, y) \, dP(x) \, dP(y)
\]

where \( P \) is such that \( P(I_{m,k_1}) > 0 \) for each \( m \) and \( k \) we also have \( K(x, x) = \sum_j \alpha_j f_j(x)g_j(x) \) a.s.(\( P \)). So far we have seen that if \( K \) is continuous and the kernel of a trace class operator, the problem with the diagonal is avoided. Now we will try to find alternative conditions that will give us the same result.
As above we define a partition of $[0,1]$ into the $I_{m_{k_{t}}}$'s, and let us assume that $K$ satisfies:

$$|K(t + h, y) - K(t, y)| \leq |h|^{\alpha} A(y), \quad 0 < \alpha < 1$$

where $A$ is nonnegative and $A \in L^{2}(P)$, and

$$|K(t, y + \eta) - K(t, y)| \leq |\eta|^\beta B(t), \quad 0 < \beta < 1$$

where $B$ is nonnegative and $B \in L^{2}(P)$ that is, $K \in Lip_{\alpha}$ in $x$ and $K \in Lip_{\beta}$ in $y$.

Let us take a fixed value of $t \in [0,1]$. If $t \in I_{m_{k_{t}}}$ for a fixed value of $m$ as defined above, then letting $h$ vary in $I_{m_{k_{t}}}$ we have

$$K(t, t + \eta) - \frac{1}{2m_{\alpha}} A(t + \eta) \leq K(t, t + \eta) - |h|^{\alpha} A(t + \eta) \leq K(t + h, t + \eta) - \frac{1}{2m_{\alpha}} A(t + \eta)$$

$$K(t + h, t + \eta) \leq K(t, t + \eta) + |h|^{\alpha} A(t + \eta) \leq K(t, t + \eta) + \frac{1}{2m_{\alpha}} A(t + \eta)$$

since $t + h \in I_{m_{k_{t}}}$ implies $|h| \leq 1/2^{m}$. Now integrating $K(t + h, t + \eta)$ over the values of $t + h$ that belong to $I_{m_{k_{t}}}$, and averaging over the measure of the same interval

$$K(t, t + \eta) - \frac{1}{2m_{\alpha}} A(t + \eta) \leq \frac{1}{P(I_{m_{k_{t}}}k_{t})} \int_{I_{m_{k_{t}}}} K(t + h, t + \eta) dP(h)$$

$$\leq K(t, t + \eta) + \frac{1}{2m_{\alpha}} A(t + \eta).$$

Now if we let vary $\eta$ on $I_{m_{k_{t}}}$ we have in a similar way:

$$K(t, t) - \frac{1}{2m_{\beta}} B(t) \leq K(t, t + \eta) \leq K(t, t) + \frac{1}{2m_{\beta}} B(t)$$

and $B(t)$ is finite a.s. ($P$) since $B \in L^{2}(P)$, therefore putting the last two expressions together we get:

$$K(t, t) - \frac{1}{2m_{\beta}} B(t) - \frac{1}{2m_{\alpha}} A(t + \eta) \leq \frac{1}{P(I_{m_{k_{t}}k_{t}})} \int_{I_{m_{k_{t}}k_{t}}} K(t + h, t + \eta) dP(h)$$

$$\leq K(t, t) + \frac{1}{2m_{\beta}} B(t) + \frac{1}{2m_{\alpha}} A(t + \eta).$$

We now integrate with respect to $t$ in such a way that $t + \eta$ remains in $I_{m_{k_{t}}}$, and after averaging we obtain:

$$K(t, t) - \frac{1}{2m_{\beta}} B(t) - \frac{1}{2m_{\alpha}} \frac{1}{P(I_{m_{k_{t}}k_{t}})} \int_{I_{m_{k_{t}}k_{t}}} A(t + \eta) dP(\eta) \leq$$

$$\frac{1}{(P \times P)(I_{m_{k_{t}}k_{t}} \times I_{m_{k_{t}}})} \int_{I_{m_{k_{t}}k_{t}}k_{t}} \int_{I_{m_{k_{t}}k_{t}}k_{t}} K(t + h, t + \eta) dP(\eta) dP(\eta) = K_{m}(t)$$

41
\[ K(t, t) + \frac{1}{2m^2} B(t) + \frac{1}{2ma P(I_{m,k})} \int_{I_{m,k}} A(t + \eta) dP(\eta). \]

Since \( A \in L^2(P) \), \((1/P(I_{m,k})) \int_{I_{m,k}} A(t + \eta) dP(\eta) < \infty\) for each \( m \). Now if we keep \( t \) fixed and we let \( m \to \infty \) we will have that \( K_m(t) \to K(t, t) \) a.s.(\( P \)) so we can state the following:

Lemma 2.3.4 If \( K : [0,1]^2 \to \mathbb{R} \) is such that \( K \in \text{Lip}_\alpha \) in \( x \) for some \( 0 < \alpha < 1 \) and \( K \in \text{Lip}_\beta \) in \( y \) for some \( 0 < \beta < 1 \), then \( K_m(t) \) as defined above converges a.s.(\( P \)) to \( K(t, t) \).

Remark 2.3.5 If we want \( K \) to be the kernel of a trace class operator all we need to ask is \( \alpha > 1/2 \) or \( \beta > 1/2 \) in the last Lemma (here we refer to the Smithies-Stinespring condition Theorem 2.1.5).

Let us define using the same notation of this section:

\[ f_{j,m}(t) = \frac{1}{P(I_{m,k})} \int_{I_{m,k}} f_j(x) dP(x) \]
\[ g_{j,m}(t) = \frac{1}{P(I_{m,k})} \int_{I_{m,k}} g_j(x) dP(x) \]

and let \( A_m \) be the \( \sigma \)-algebra generated by \( I_{m,k}, k = 1, \ldots, 2^m \), then \( A_m \subset A_{m+1} \) for \( m \geq 1 \). We know that \( f_j \in L^1(P) \), and \( E(f_j | A_m) = f_{j,m} \), since \( f_{j,m} \) is clearly \( A_m \) measurable and for any \( I_{m,k}, k = 1, \ldots, 2^m \)

\[ \int_{I_{m,k}} f_{j,m}(t) dP(t) = \int_{I_{m,k}} \frac{1}{P(I_{m,k})} \int_{I_{m,k}} f_j(x) dP(x) dP(t) = \int_{I_{m,k}} f_j(x) dP(x) \]

so \( E(f_j | A_m) = f_{j,m} \). Therefore \( \{f_{j,m}, A_m\} \) is a martingale, for

\[ E(f_{j,m+1} | A_m) = E(E(f_j | A_{m+1}) | A_m) \]
\[ = E(f_j | A_m) = f_{j,m}. \]

since \( A_{m+1} \subset A_m \).

Now applying two well known results in martingales, see for example Ash [1972] pages 298-300, we have the following lemma:

Lemma 2.3.6 Let \( f_{j,m}(t) \) and \( g_{j,m}(t) \) defined as above. Then for each \( j \in \{1,2,\ldots\} \)

\[ f_{j,m} \to E(f_j | A_\infty) = f_j \quad \text{and} \quad g_{j,m} \to E(g_j | A_\infty) = g_j \]

and the convergence is almost surely and in \( L^1 \), and \( A_\infty \) is the usual \( \sigma \)-algebra in \([0,1]\).
I thank Prof. Richard M. Dudley for the next example: Example of $K(x, y)$ continuous symmetric kernel of a trace class operator $K(x, y) = \sum \lambda_n \varphi_n(x) \varphi_n(y)$ where the $\{\varphi_n\}$ are orthonormal and continuous and $\sum_n |\lambda_n| < \infty$ but $\sum_n |\lambda_n| \varphi^2_n$ is unbounded on the diagonal. (This example shows why the hypothesis of Mercer’s Theorem of positive definiteness is needed).

Take $x, y \in [0, 1]$, the Lebesgue measure $\lambda$ and disjoint intervals $I_k$ such that $\lambda(I_k) = 1/2k^2$ for $k \geq 1$. Since $\sum_k 1/2k^2 < 1$ this is always possible. For example define $I_1 = [0, 1/2), I_2 = [1/2, 5/8), \ldots$. Let $f_k$ = triangle function on the left half of $I_k$ such that $\sup_{x \in [0, 1]} f_k(x) = 1$. in our example we have:

$$f_1(x) = \begin{cases} 8x & \text{if } x \in [0, 1/8) \\ 2 - 8x & \text{if } x \in [1/8, 1/4) \\ 0 & \text{elsewhere}; \end{cases}$$

$f_2, \ldots$ are defined in a similar way. Let $g_k$ = triangle function on the right half of $I_k$ such that $\sup_{y \in [0, 1]} g_k(y) = k$ for $k \geq 1$ and such that $\int g_k^2 d\lambda = \int f_k^2 d\lambda$. Then $f_k$ and $g_k$ are continuous $L^2$ functions on $[0, 1]$. Define

$$K(x, y) = \sum k \frac{1}{2k^{3/2}}(f_k(x)g_k(y) + g_k(x)f_k(y)).$$

Then $\sup(f_kg_k) = 1 \cdot k = k$ and they have disjoint supports by construction and also $2k/k^{3/2} \to 0$ as $k \to \infty$. Then the series converges uniformly and absolutely to $K(x, y)$ and by continuity of $f_k, g_k$, $K$ is continuous and symmetric by construction. Besides $(f_k + g_k) \perp (f_k - g_k)$ since $\int f_k^2 d\lambda = \int g_k^2$. If we let

$$(f \otimes g)(x, y) := f(x)g(y)$$

we have

$$(f_k \otimes g_k) + (g_k \otimes f_k) = \frac{1}{2}[(f_k + g_k) \otimes (f_k + g_k) - (f_k - g_k) \otimes (f_k - g_k)].$$

Then we can write

$$K(x, y) = \sum_{k=1}^{\infty} \frac{1}{2k^{3/2}}[((f_k + g_k) \otimes (f_k + g_k) - (f_k - g_k) \otimes (f_k - g_k))](x, y).$$

And for different $k$'s the terms have disjoint supports and

$$\int (f_k + g_k)^2 d\lambda = \int (f_k - g_k)^2 d\lambda = 2 \int f_k^2 d\lambda = 1/(6k^2).$$

Therefore, if we define $h_k^+ = \sqrt{6} k(f_k + g_k)$ and $h_k^- = \sqrt{6} k(f_k - g_k)$, they are orthonormal and

$$K(x, y) = \sum \frac{1}{2k^{3/2}}[\frac{1}{6k^2}(h_k^+ \otimes h_k^+) - \frac{1}{6k^2}(h_k^- \otimes h_k^-)](x, y).$$
\[
= \sum \frac{1}{12k^{7/2}} [(h_k^+ \otimes h_k^+) - (h_k^- \otimes h_k^-)].
\]
If we just relabel and we call \( \lambda_j = \pm \frac{1}{12k^{7/2}} \) we have
\[
K(x, y) = \sum_j \lambda_j (\varphi_j \otimes \varphi_j)(x, y)
\]
where \( \varphi_j = h_k^+ \otimes h_k^+ \) or \( \varphi_j = h_k^- \otimes h_k^- \). So \( K \) is the kernel of a trace class operator.

But on the diagonal
\[
\sum_j |\lambda_j| \varphi_j \otimes \varphi_j \geq \sum_k \frac{1}{k^{3/2}} g_k^2
\]
and \( \sup_x g_k(x) = k \) so, \( \sup_k \sum \frac{1}{(k^{3/2})} g_k^2 \geq k^{1/2} \) \( \forall k \), and so \( \sum_j |\lambda_j| \varphi_j \otimes \varphi_j \) is unbounded on the diagonal.

Let \( \beta(\cdot) \) the brownian bridge and \( \beta(f) \) the process indexed by functions \( f \in L^2 \), as defined in Dudley [1984], then we have:

Lemma 2.3.7 Let \( K(x, y) \in L^2(P \times P) \). If \( K \) can be written as
\[
K(x, y) = \sum_m f_m(x)g_m(y)
\]
where \( \sum_m \|f_m\|_2 \|g_m\|_2 < \infty \), then
\[
\sum_m E \left| \int f_m(x)g_m(y) d\beta(x)d\beta(y) \right| < \infty.
\]

Proof: \( \int f_m d\beta \sim N(0, VarP f_m) \) and \( \int g_m d\beta \sim N(0, VarP g_m) \), where \( VarP f_m = \int f_m^2 dP - (\int f_m dP)^2 \) and \( VarP g_m = \int g_m^2 dP - (\int g_m dP)^2 \). Let \( \text{cov}_P(f, g) = \int fg dP \). Then using projections we can write \( g_m = c_m f_m + h_m \) for \( m \geq 1 \), where \( h_m \perp f_m \) i.e. \( \text{cov}_P(h_m, f_m) = 0 \).

We will use the following known result:

if \( X \sim N(0, \sigma^2) \) then \( E|X| = \sqrt{\frac{2}{\pi}} \sigma \).

If we fix \( m \), we have:
\[
E|\beta(f_m)\beta(g_m)| = E|\beta(f_m)\beta(c_m f_m + h_m)| = E|c_m \beta^2(f_m) + \beta(f_m) \beta(h_m)|
\]
\[
\leq |c_m|E\beta^2(f_m) + E|\beta(f_m)|E|\beta(h_m)| = |c_m|E_P(f_m^2) + \frac{2}{\pi} \sqrt{VarP(f_m) \sqrt{VarP(h_m)}}
\]
for every \( m \). From the projection we have \( \|c_m f_m\|_2 \leq \|g_m\|_2 \) and \( \|h_m\|_2 \leq \|g_m\|_2 \), so \( |c_m||f_m||_2^2 \leq \|f_m\|_2 \|g_m\|_2 \). On the other hand \( VarP(h_m) \leq E h_m^2 \) so \( \sqrt{VarP(h_m)} \leq \|h_m\|_2 \leq \|g_m\|_2 \), then \( \sqrt{VarP(f_m) \sqrt{VarP(h_m)}} \leq \|f_m\|_2 \|g_m\|_2 \). If we sum over all \( m \)'s we get:
\[
\sum_m E|\beta(f_m)\beta(g_m)| \leq \sum_m \{|c_m|E_P(f_m^2) + \frac{2}{\pi} \sqrt{VarP(f_m) \sqrt{VarP(g_m)}}\}
\]

44
\[ 
\leq \sum \left( \|f_m\|_2 \|g_m\|_2 + \frac{2}{\pi} \|f_m\|_2 \|g_m\|_2 \right) \\
= \frac{2 + \pi}{\pi} \sum_m \|f_m\|_2 \|g_m\|_2 < \infty.
\]
Chapter 3

Central Limit Theorem

3.1 Bases

Let us consider now a fixed kernel $K(x,y)$ and let $\mathcal{F}$ be the class of functions $K(\cdot,y)$ with $y$ fixed. Let $\mu, \nu$ be signed measures, and consider another class of functions $\mathcal{G}$ that consists of all functions $y \mapsto \int K(x,y)d\mu(x)$.

Definition 3.1.1

$$\|\mu\|_{\mathcal{F}} := \sup_y \{|\mu(K(\cdot,y))|\} = \sup_y \{\int K(x,y)d\mu(x)!\},$$

analogously

$$\|\nu\|_{\mathcal{G}} = \sup\{|\nu(f)| : f \in \mathcal{G}\}$$

$$= \sup\{|\nu(\int K(x,y)d\mu(x))| : ||\mu||_{\mathcal{F}} < 1\}$$

$$= \sup\{||\int\int K(x,y)d\mu(x))d\nu(y)| : ||\mu||_{\mathcal{F}} < 1\}.$$ 

We want to know when $\mathcal{F}$ is a Donsker class for every $P$, that is a universal Donsker class as defined in Dudley [1987], page 1308. For this purpose we consider Pollard’s entropy condition as defined in Dudley [1987], page 1310. First consider $f, g \in \mathcal{F}$, then there exist $y_1, y_2$ in the domain of $K$ such that $f(\cdot) = K(\cdot, y_1)$ and $g(\cdot) = K(\cdot, y_2)$. For $Q$ a probability measure we have

Definition 3.1.2 For $f, g \in \mathcal{F}$ we define

$$e_{Q,2}(f, g) := e_{Q,2}(y_1, y_2) := (\int |K(x, y_1) - K(x, y_2)|^2 dQ(x))^{1/2}.$$
We already proved that if $K(x, y)$ is the kernel of a trace class operator then we can represent $K(x, y) = \sum_j f_j(x)g_j(x)$ where $\sum_j \|f_j\|_2 \|g_j\|_2 < \infty$ ($L^2$ norms for $Q$). So, if we suppose $K$ has this representation then:

$$e_{Q,2}(f, g) = e_{Q,2}(y_1, y_2) = \|K(\cdot, y_1) - K(\cdot, y_2)\|_2$$

$$= \left(\int |\sum_j f_j(x)g_j(y_1) - \sum_j f_j(x)g_j(y_2)|^2dQ(x)\right)^{1/2}$$

$$= \left(\int |\sum_j f_j(x)(g_j(y_1) - g_j(y_2))|^2dQ(x)\right)^{1/2}.$$

**Definition 3.1.3**

$$D^{(2)}(\epsilon, \mathcal{F}, Q) := D(\epsilon, \mathcal{F}, e_{Q,2})$$

$$= \sup\{m : \text{for some } y_1, \ldots, y_m \in X, e_{Q,2}(y_i, y_j) > \epsilon \text{ for } 1 \leq i < j \leq m\}$$

and

$$D^{(3)}(\epsilon, \mathcal{F}) = \sup D^{(2)}(\epsilon, \mathcal{F}, Q)$$

where the supremum is taken over all probability measures $Q$ concentrated on finite sets.

**Definition 3.1.4** We say that $\mathcal{F}$ satisfies Pollard’s entropy condition iff

$$\int_0^1 (\log D^{(2)}(\epsilon, \mathcal{F}))^{1/2}d\epsilon < \infty.$$

If $\mathcal{F}$ is a uniformly bounded class of functions satisfying some mild measurability conditions and Pollard’s entropy condition, then $\mathcal{F}$ is a universal Donsker class by a Theorem of Pollard [1982], see also Dudley [1984], Theorem 11.3.1. According to the definition of $D^{(2)}(\epsilon, \mathcal{F})$ and if we suppose that $K$ satisfies the condition of Theorem 2.2.1 for every $Q$ probability measure and $N(\epsilon)$ is independent of $Q$, then

$$D^{(2)}(\epsilon, \mathcal{F}) \leq N(\epsilon) \leq C/\epsilon^r \quad (*)$$

so the condition described above is satisfied, and $\mathcal{F}$ is a universal Donsker class. Unfortunately, in general the value of $N(\epsilon)$ depends strongly on $Q$. Therefore, we need a stronger condition on $K$ to have a universal Donsker class.

Note: We observe that $\{y_1, \ldots, y_N\}$ given by the condition of Theorem 2.2.1, is an $\epsilon$-net as defined in Dudley [1984], page 39. So using his notation:

$$N(\epsilon, X, e_{Q,2}) \leq D(\epsilon, \mathcal{F}, e_{Q,2})$$

where $D(\epsilon, X, e_{Q,2}) = \text{largest } n, \text{ such that for some } y_1, \ldots, y_n \in X, e_{Q,2}(y_i, y_j) > \epsilon$ whenever $i \neq j$. And taking $N(\epsilon, X, e_{Q,2}) = \text{smallest } n, \text{ such that } X \subset \bigcup_{i=1}^n B_i$
for some $B_i$'s such that $\text{Diam}B_i \leq 2\varepsilon$, then we have that $N(\varepsilon, X, e_{Q,2}) \leq C/\varepsilon'$ and applying the result in Dudley [1984], Theorem 6.0.1 we have:

$$D(2\varepsilon, X, e_{Q,2}) \leq N(\varepsilon, X, e_{Q,2}) \leq C/\varepsilon'$$

and so (*) follows.

The first problem we want to solve is to find conditions on $K(\cdot, \cdot)$, such that $A_K$, the integral operator it generates, is trace class independently of $P$, the probability measure. In order to do this we propose, that if $K$ has an absolutely convergent development in supremum norm, then $A_K$ is trace class for every $P$.

We start our analysis with the usual trigonometric Fourier series in $\mathbb{R}$:

Let $f$ be a real function and let $S[f]$ be its Fourier series so

$$S[f] = \sum_{-\infty}^{\infty} c_n e^{inx}$$

where $c_n = (1/2\pi) \int_{-\pi}^{\pi} f(t) e^{-int} dt$, so

$$S[f] = (1/2)a_0 + (a_1 \cos(x) + b_1 \sin(x)) + \cdots$$

where $c_n = (1/2)(a_n - ib_n)$ and $c_{-n} = (1/2)(a_n + ib_n)$.

Let

$$b_n(y) = (1/2\pi) \int_{-\pi}^{\pi} K(t,y) e^{-int} dt$$

for $K(x, y) \in L^2([-\pi, \pi] \times [-\pi, \pi])$. Then $b_n(y) = (1/2\pi)(A_K e^{int})(y)$ for the Lebesgue measure.

**Definition 3.1.5** Let $f$ be a function on $\mathbb{R}$ and let $\omega(\delta)$ its modulus of continuity, that is:

$$\omega(\delta) := \omega(\delta, f) := \sup\{|f(x_2) - f(x_1)| : |x_2 - x_1| \leq \delta\}.$$

If for some $\alpha > 0$ $\omega(\delta) \leq C\delta^{\alpha}$ where $C$ is independent of $\delta$, $f$ satisfies a Lipschitz condition of order $\alpha$. In symbols $f \in \Lambda_\alpha$, $0 < \alpha \leq 1$.

By a theorem of Bernstein [1914], as stated in Zygmund [1959], page 240, if $K(\cdot, y) \in \Lambda_\alpha$ for $\alpha > 1/2$, then $S[K(\cdot, y)]$ converges absolutely, and for $\alpha = 1/2$ the result does not follow necessarily, for a counterexample see Zygmund [1959], pages 197 and 243.

So if $K(\cdot, y) \in \Lambda_\alpha$ for some $\alpha > 1/2$ then

$$\sum_{-\infty}^{\infty} (1/2\pi)|(A_K e^{int})(y)| < \infty.$$
Now write \( b_n(y) \) in the form \( a_n g_n(y) \) with \( a_n = 1/2\pi \) and \( g_n(y) = (A_K e^{-int})(y) \). Then
\[
\sum_{n} |a_n| \|g_n\|_{\infty} \|f_n\|_{\infty} = \sum_{n} (1/2\pi) \sup_{y} \left| \int_{-\pi}^{\pi} K(t,y) e^{-int} dt \right|
\]
which, under only the assumptions of Bernstein's Theorem is not necessarily finite. So, we need stronger conditions on \( K \) to have an absolutely convergent series for \( K(x,y) \).

In the rest of this section we give some results about bases and the absolute convergence of series, starting with results in \( \mathbb{R} \) and giving some extensions to the multidimensional case in \( \mathbb{R}^d \). We start with some results on Schauder bases. Most of the results can be found in Semadeni [1982].

**Definition 3.1.6** A Schauder basis in an infinite dimensional Banach space \( (F, \| \cdot \|) \) is a sequence \( (e_n) \) of elements of \( F \), such that for each \( f \in F \) there is a unique sequence of scalars \( (a_n) \) with
\[
f = \sum_{n=1}^{\infty} a_n e_n, \quad i.e. \quad \|f - \sum_{n=1}^{m} a_n e_n\| \to 0 \quad \text{as} \quad m \to \infty.
\]
A basic sequence in \( (F, \| \cdot \|) \) is a sequence that is a Schauder basis in its closed linear span.

From now on, when we refer to a basis we mean a Schauder basis.

It is important to notice that a basis is a sequence \( (e_n)_{n \in \mathbb{N}} \) with a prescribed order of terms. So, if we write \( (e_v)_{v \in V} \) or \( \sum_{v} a_v e_v \) we really mean \( (e_{v_n})_{n \in \mathbb{N}} \) and \( \sum_{n=1}^{\infty} a_{v_n} e_{v_n} \), where the order in \( V \) has been prescribed. The coefficients \( a_j \) depend naturally on \( f \). So, we define \( e^*_{n}(\sum_{j=1}^{\infty} a_{j} e_{j}) = a_{n} \), then we can write \( f = \sum_{j=1}^{\infty} e^*_{j}(f)e_{j} \). From the uniqueness of \( a_{n} \) it follows that \( e^* \) is a linear functional on \( F \), which satisfies:
\[
e^*_{n}(e_{m}) = \delta_{n,m}.
\]
The partial sums of a given \( f \) will be denoted by \( s_{n} = \sum_{j=1}^{n} a_{j} e_{j} \), so \( s_{n} \) is a linear operator \( s_{n} : F \to F \) such that \( s_{n}(f) = \sum_{j=1}^{n} e^*_{j}(f)e_{j} \), that satisfies:

i) The range of \( s_{n} \) is \( n \)-dimensional.

ii) \( s_{n} \circ s_{m} = s_{m} \circ s_{n} = s_{n} \) if \( m \geq n \).

**Definition 3.1.7** A basis \( (e_{n}) \) in \( F \) is normalized iff \( \|e_{n}\| = 1 \) for every \( n \), and seminormalized iff \( 0 < \inf \|e_{n}\| \leq \sup \|e_{n}\| < \infty \).

**Lemma 3.1.8** Let \( (e_{n}) \) be a basis of \( F \). Then for \( f \in F \) of the form \( f = \sum_{n=1}^{\infty} a_{n} e_{n} \), the number
\[
\|f\| = \sup_{n} \| \sum_{j=1}^{n} a_{j} e_{j} \|
\]
is finite, and \( \| \cdot \| \) is a norm in \( F \) equivalent to the initial norm in \( F \). Specifically \( \exists c > 0 \) such that \( \|f\| \leq \|f\| \leq c\|f\| \).
Proof: See Semadeni page 2.

**Theorem 3.1.9** Let \((e_n)\) be a sequence in \(F\). Then the following are equivalent:

i) \((e_n)\) is basic.

ii) \(e_j \neq 0\) for each \(j\) and there exists \(c\) such that for every sequence \((a_j)\) of scalars and \(n < m\) we have

\[
\| \sum_{j=1}^{n} a_j e_j \| \leq c \| \sum_{j=1}^{m} a_j e_j \|.
\]

iii) \(e_j \neq 0\) and there exists \(\gamma > 0\) such that for every \(n, k \in \mathbb{N}\) and every \((a_j)\) we have

\[
\| \sum_{j=1}^{n} a_j e_j \| = 1 \quad \text{implies} \quad \| \sum_{j=1}^{n+k} a_j e_j \| \geq \gamma.
\]

Proof: See Semadeni page 3.

**Proposition 3.1.10** Let \((e_n)\) be a basis of \(F\). Then the linear operator \(s_n\) is a bounded linear projection operator onto \(\text{span} \ (e_1, \ldots, e_n)\) and \(\sup_n \|s_n\| < \infty\). Moreover each \(e_n(\cdot)\) is a bounded linear functional. If \((e_n)\) is seminormalized then \(\sup_n \|e^*_n\| < \infty\).


**Proposition 3.1.11** Let \((e_n)\) be a basis of \(F\). Then \((e^*_n)\) is a basic sequence in the dual space \(F'\). Moreover the basis constants of \((e_n)\) and \((e^*_n)\) are equal, where the basis constant is \(\sup_n \|s_n\|\).


**Definition 3.1.12** A basic sequence \((e_n)\) is monotone iff for any scalars \((a_n)\)

\[
\| \sum_{j=1}^{n} a_j e_j \| \leq \| \sum_{j=1}^{n+1} a_j e_j \| \quad \text{for} \quad n = 1, 2, \ldots,
\]

\((e_n)\) is monotone iff \(\|s_n\| = 1\) for each \(n\), that is \(\|s_n(f)\| \leq \|f\|\) for each \(f \in F\).

**Definition 3.1.13** Let \(X\) be a locally compact space such that \(\gamma X\), its one-point compactification, is metrizable, and let \(t_n \in X, \ n = 1, 2, \ldots\). A basis \((e_n)\) in \(C_0(X)\), the subspace of functions vanishing at infinity, is called interpolating with nodes \((t_n)\) iff for each \(f \in C_0(X)\) and each \(n\) we have \(s_n(t_m) = f(t_m)\) for \(m = 1, 2, \ldots\), where

\[
s_n = \sum_{k=1}^{n} e^*_k(f) e_k = s_n(f) \quad \text{for} \quad m \leq n.
\]
Proposition 3.1.14 Let \((e_n)\) be a sequence in \(C_0(X)\). Suppose that for each \(f\) in \(C_0(X)\) there exists at least one sequence \((a_n)\) such that \(f = \sum_{n=1}^{\infty} a_n e_n\). Let \((t_n)\) be a sequence in \(X\) such that \(t_n \neq t_m\) for \(n \neq m\). Then the following are equivalent:

i) \((e_n)\) is an interpolating basis with nodes \((t_n)\).

ii) \(e_n(t_n) \neq 0\) and \(e_n(t_m) = 0\) for \(n > m\).

Proof: Refer to Semadeni page 9.

Definition 3.1.15 A series \(\sum f_n\) in a Banach space \(F\) is unconditionally convergent if it is convergent for any permutation of terms, i.e. \(\sum f_{\pi(n)}\) is convergent for every bijection \(\pi : \mathbb{N} \to \mathbb{N}\). A basis \((e_n)\) in \(F\) is unconditional iff for any \(f = \sum a_n e_n\) the convergence of the series is unconditional. 

\[ \sum f_n \text{ is absolutely convergent iff } \sum \|f_n\| < \infty. \]

There are many different basic systems, which have been studied carefully; among them we can mention: Haar, the hat functions, Faber-Schauder, Franklin, pyramidal, etc. For details about all these systems refer to Semadeni [1982].

The specific problem of dealing with absolute convergence in any orthonormal system has been treated extensively (see McLaughlin [1973], Bljumin and Kotljar [1970], Cheng [1942] among many). But as noted by Mitjagin [1964], page 1083 Theorem 1, the condition \(\alpha > 1/2\) in Bernstein's Theorem cannot be improved for any orthogonal system:

Theorem 3.1.16 Let \((e_n)\) be any orthogonal system over \([0, 2\pi]\); there exists a function \(\phi : [0, 2\pi] \to \mathbb{R}\) with \(\phi \in \Lambda_{1/2}\) such that the series of its Fourier coefficients in the system \((e_n)\) is not absolutely convergent, i.e.,

\[ \sum_1^\infty \left| \int_0^{2\pi} \phi(t)e_n(t)dt \right| = \infty. \]

Proof: See Mitjagin [1964], page 1083.

To find conditions on a kernel \(K(x, y)\) to have an absolutely convergent series we will use another result of Mitjagin [1964], but first we need one definition.

Definition 3.1.17 Let \(\alpha = (\alpha_1, \alpha_2)\) be a pair of positive numbers such that \(\alpha_j = p_j + \gamma_j\) for \(j = 1, 2\), where \(p_j\) 's are positive integers and \(0 < \gamma_j \leq 1\) for \(j = 1, 2\). Let \(C^\alpha\) the space of functions \(K\) of two variables that are periodic in each variable and they satisfy:

\[ \left| \frac{\partial^n K(t_1 + u_1, t_2)}{\partial t_1^{p_1}} - \frac{\partial^n K(t_1, t_2)}{\partial t_1^{p_1}} \right| \leq M_K |u_1|^n \]
and

\[
\left| \frac{\partial^p K(t_1, t_2 + u_2)}{\partial t_2^p} - \frac{\partial^p K(t_1, t_2)}{\partial t_2^p} \right| \leq M_K |u_2|^\gamma.
\]

And let \(1/\alpha_0 = 1/\alpha_1 + 1/\alpha_2\).

**Theorem 3.1.18** If \(\alpha_0 > 1/2\), then every function \(K \in C^\alpha\) has an absolutely convergent series of Fourier coefficients in the trigonometrical system. But if \(\alpha_0 \leq 1/2\), then for every complete orthonormal system \(\Psi = \{\psi_n(t_1, t_2)\}\), \(0 \leq t_j \leq 2\pi\), there exists a function \(\phi \in C^\alpha\) such that \(|\sum |\int \phi_n dt_1 dt_2| = \infty\).

**Proof:** See Mitjagin [1964], page 1085.

**Remark 3.1.19** Therefore, if we want \(\alpha_0\) to be greater than \(1/2\), and we suppose for simplicity that \(p = p_1 = p_2\) and \(\gamma = \gamma_1 = \gamma_2\), all we need, is to ask \(p + \gamma > 1\), i.e., if \(K\) has first partial derivatives and they satisfy a Lipschitz condition of order \(0 < \gamma \leq 1\). Then \(K\) has an absolutely convergent Fourier series.

### 3.2 A Central Limit Theorem

Let \(K(x, y) \in L^2([0, 1]^2, P \times P)\) be a function which is periodic in both variables and suppose it satisfies:

\[
|\frac{\partial}{\partial x} K(x + u, y) - \frac{\partial}{\partial x} K(x, y)| \leq M|u|^\gamma
\]

and similarly

\[
|\frac{\partial}{\partial y} K(x, y + u) - \frac{\partial}{\partial y} K(x, y)| \leq M|u|^\gamma
\]

where \(0 < \gamma \leq 1\), \(M\) is a positive constant. Then \(K\) has an absolutely convergent series of Fourier coefficients in the trigonometrical system by Theorem 3.1.18.

Let \(\alpha > 0\) be a positive number and let \(\beta\) be the greatest integer less than \(\alpha\), for \(x := (x_1, \ldots, x_d) \in \mathbb{R}^d\) and \(p := (p_1, \ldots, p_d) \in \mathbb{N}^d\) let, \(|p| := p_1 + \cdots + p_d\), \(x^p := x_1^{p_1} x_2^{p_2} \cdots x_d^{p_d}\) and \(D^p := \frac{\partial^{|p|}}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}}\). Following the idea of Kolmogorov and Tikhomirov as in Dudley [1984], page 51, we define the space of differentiable functions having bounded derivatives through order \(\alpha\) defined in an open set \(U \subset \mathbb{R}^d\) as the set:

**Definition 3.2.1** \(C_\alpha = \{f : U \rightarrow \mathbb{R} : f\) has all partial derivatives \(D^p f\) of order \(|p| \leq \beta\) defined on \(U\) and its partial derivatives of order \(\beta\) satisfy a Lipschitz condition of order \(\alpha - \beta\}\).
Now we define a norm in $C_a$.

**Definition 3.2.2** Let $f \in C_a$ define the norm of $f$, $\|f\|_{a,U} := \|f\|_{a}$ by:

$$
\|f\|_{a} := \max_{|p| \leq \beta} \sup_{x \in U} \{|D^p f(x)|\} + \max_{|p| = \alpha} \sup_{x \neq y} \frac{|D^p f(x) - D^p f(y)|}{|x - y|^\alpha - \beta}
$$

where $|u| := (u_1^2 + \cdots + u_n^2)^{1/2}$, $u \in \mathbb{R}^d$ and $x, y \in U$.

Let $I^d$ denote the unit cube $\{x \in \mathbb{R}^d: 0 \leq x_j \leq 1 \text{ for } j = 1, \ldots, d\}$.

Let $G_{a,M,d}$ denote the set of all functions $f : I^d \to \mathbb{R}$ such that $f \in C_a$ and $\|f\|_{a} \leq M$.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $S = [0,1]$. Consider a kernel $K(x,y)$ with $K \in L^2(S \times S, P \times P)$ and the family $\mathcal{F} = \{K(s, \cdot) : s \in S\}$.

In the rest of this section we want to find conditions on $K(x,y)$ so that we can apply a central limit theorem under metric entropy with $L^2$ bracketing for $\mathcal{F}$. To obtain such a result, we will use Ossiander's paper [1987]. In order to apply her results, first we have to find an appropriate metric $\rho$ such that, the $L^2$ metric of the difference of any two elements in $\mathcal{F}$ is bounded by $\rho$, independently of $P$. Secondly we have to find conditions on $K$, such that the metric entropy with bracketing of $K$ with respect to $\rho$ satisfies an integrability condition, and finally apply Theorem 3.1 of Ossiander [1987], page 904, to get a central limit theorem for $S_n(s) = (1/\sqrt{n}) \sum_{i=1}^n K(s, V_i)$, where $\{V_i : i \geq 1\}$ is a sequence of independent copies of $V$, a random variable defined on $(\Omega, \mathcal{A}, P)$.

So, to begin we need a metric $\rho(s,t)$ with $s,t \in S$, such that

$$
\|K(s, \cdot) - K(t, \cdot)\|_{2,P} = \left( \int |K(s,y) - K(t,y)|^2 dP(y) \right)^{1/2} \leq \rho(s,t)
$$

for every probability measure $P$ and $K(s,\cdot), K(t,\cdot) \in \mathcal{F}$. The natural selection here is the sup norm, so, let us define

$$
\rho(s,t) = \|K(s, \cdot) - K(t, \cdot)\|_{\infty} = \sup_{y \in S} |K(s,y) - K(t,y)| = \rho_1(K_s, K_t)
$$

where $K_u = K(u, \cdot) \in \mathcal{F}$. Then $\rho$ is clearly a pseudometric in the space $S$ and certainly satisfies

$$(I) \quad \|K(s, \cdot) - K(t, \cdot)\|_{2,P} \leq \rho(s,t)$$

for every probability $P$ measure on $S$.

(II) Let us assume that for each $s \in S$, $E_P K(s,V) = 0$.

(III) Since $K \in L^2(S \times S, P \times P)$, then we also have that $E_P K^2(s,V) < \infty$, for all $s \in S$. Conditions (I),(II) and (III) correspond to conditions (2.3),(2.1) and (2.2) in Ossiander's paper. Condition (I) gives a Lipschitz condition for each $K_s \in \mathcal{F}$ with respect to the metric $\rho$ in $L^2(P)$, for every probability measure $P$. We will assume
also that the process is separable, that is, there exists a countable dense set $D_S \subset S$ and a measurable set $N \subset \Omega$ with $P(N) = 0$, such that for any open set $G \subset S$ and any closed set $F \subset \mathbb{R}$.

$$\{K(s, V) \in F \text{ for all } s \in G \cap D_S\} \setminus \{K(s, V) \in F \text{ for all } s \in G\} \subset N.$$ 

The sets $N$ and $D_S$ obviously depend on $P$, so that, to apply these results we must have a separable process $\mathcal{F}$ for every $P$.

In the next section we will talk about the separability condition.

Now we recall some definitions on entropy, as in Dudley [1984], page 39. Let $(S, \rho)$ be a metric (pseudometric) space and $B \subset S$.

**Definition 3.2.3** If $\delta > 0$, a $\delta$-net for $B$, is a set $\{x_1, \ldots, x_n\} \subset S$, such that for every $x \in B$ there exists $i \in \{1, \ldots, n\}$ that satisfies $\rho(x, x_i) \leq \delta$.

**Definition 3.2.4**
\[
\text{diam}B := \sup_{x,y \in B} \rho(x, y).
\]

**Definition 3.2.5**
\[
N(\delta, B, \rho) := \left\{ \begin{array}{ll}
\text{smallest } n, \text{ such that } B \subset \bigcup_{i=1}^{n} B_i \\
\text{and for each } i \in \{1, \ldots, n\} \quad \text{diam}B_i \leq 2\delta.
\end{array} \right\}
\]

**Definition 3.2.6**
\[
D(\delta, B, \rho) := \left\{ \begin{array}{ll}
\text{largest } n, \text{ such that for some } x_1, \ldots, x_n \in B \\
\rho(x_i, x_j) > \epsilon \text{ for } i \neq j, i,j \in \{1, \ldots, n\}.
\end{array} \right\}
\]

**Definition 3.2.7**
\[
\mathcal{N}(\delta, B, S, \rho) := \text{minimal number of points in a } \delta\text{-net, for } B \text{ in } S.
\]

Given a probability space $(\Omega, \mathcal{A}, P)$, $1 \leq q \leq \infty$ and a family of functions $\mathcal{F} \subset L^q(\Omega, \mathcal{A}, P)$ with usual norm $\|f\|_q = (\int |f|^q dP)^{1/q}$, we define

**Definition 3.2.8**
\[
N_{(q)}^{(\delta)}(\mathcal{F}, P) := \left\{ \begin{array}{ll}
\text{smallest } m, \text{ such that for some } f_1, \ldots, f_m \in L^q(\Omega, \mathcal{A}, P) \\
\mathcal{F} \subset \bigcup_{i,j} \{f_i, f_j : \|f_i - f_j\|_q \leq \delta\}.
\end{array} \right\}
\]

54
where \([f_i, f_j] = \{ f : f_i \leq f \leq f_j \}\) is called a bracket and \(\log N_{[1]}^{(q)}(\delta, \mathcal{F}, P)\) is called metric entropy with bracketing.

One theorem stated by Dudley [1984], page 39, relates the above definitions, in fact, this is the theorem that we mention in the proof of Lemma 1.5.3.

**Theorem 3.2.9** For any \(\delta > 0\) and \(B \subset S\)

\[
D(2\delta, B, \rho) \leq N(\delta, B, \rho) \leq N(\delta, B, S, \rho) \leq N(\delta, B, B, \rho) \leq D(\delta, B, \rho).
\]

Proceeding with the second step, we need to find conditions on \(K\), in order to obtain an integrability property of the metric entropy with bracketing of \(\mathcal{F}\), as defined above. To do so, we will need some previous results, we refer to Dudley [1984] for proofs.

**Definition 3.2.10** Let \(f, g\) be two real functions defined for small \(y > 0\), then \(g \propto f\) means that

\[
0 < \liminf_{y \downarrow 0} \left( \frac{g}{f} \right)(y) = \limsup_{y \downarrow 0} \left( \frac{g}{f} \right)(y) < \infty.
\]

For the metric entropy in the sup norm \(d_{\sup}(f, g) = \sup_x |f(x) - g(x)|\), we have Kolmogorov’s Theorem:

**Theorem 3.2.11** For \(0 < M < \infty\), as \(\delta \downarrow 0 \log D(\delta, \mathcal{G}_{\alpha, M, 2}, d_{\sup}) \propto \delta^{-2/\alpha}\).

For some \(J := J(\alpha, M)\), any law \(P\) on \(S^2\), \(1 \leq q \leq \infty\) and \(0 < \delta < 1\)

\[
\log N_{[1]}^{(q)}(\delta, \mathcal{G}_{\alpha, M, 2}, P) \leq J\delta^{-2/\alpha}.
\]

**Proof:** See Dudley [1984], page 51.

Thus, Kolmogorov’s Theorem says that the metric entropy of \(\mathcal{G}_{\alpha, M, 2}\) with bracketing is at most of order \(\delta^{-2/\alpha}\), for any \(0 < \delta < 1\).

Now, returning to the conditions on Ossiander’s paper, we need to assume that \(K(s, \cdot)\) can be bounded above and below, by “simple” functions in \(S\), which are themselves close in \(L^2(P)\). For \(\delta > 0\), let \(S(\delta/2)\) be a \(\delta/2\)-net for \(S = [0, 1]\) with the metric \(\rho(s, t)\) as defined in (I), then for \(s \in S\), there exists \(s^\delta \in S(\delta/2)\) such that

\[
(IV) \quad \rho(s, s^\delta) < \delta/2 < \delta,
\]

so

\[
\sup_{y \in [0, 1]} |K(s, y) - K(s^\delta, y)| < \delta/2.
\]

Let

\[
K^\delta(s^\delta, y) = K(s^\delta, y) - \delta/2
\]
and
\[ K^\delta_y(s^\delta, y) = K(s^\delta, y) + \delta/2 \]
for \( y \in S \).

Since \( K(s^\delta, y) \) is a measurable function and \( K(s^\delta, \cdot) \in L^2(P) \), then so do \( K^I_\delta \) and \( K^I_y \).

And by their definition we have

\[(V) \quad K^I_\delta(s^\delta, y) \leq K(s, y) \leq K^I_y(s^\delta, y)\]
a.s., for any \( P \) probability measure.

And we also have

\[(VI) \quad (EP(K^I_\delta(s^\delta, V) - K^I_\delta(s^\delta, V))^2)^{1/2} < \delta \]

independently of \( P \). And we say that \( K^I_\delta \) and \( K^I_y \) are upper and lower \( \delta \)-approximations to \( K(s, \cdot) \) in \( L^2(P) \), for every \( P \), as defined in Ossiander [1987], page 900.

We are now ready to define another kind of metric entropy.

**Definition 3.2.12** Let

\[ \nu^B(\delta, S, \rho) = \min \{ \text{card} \mathcal{S}(\delta): (IV), (V) \text{ and } (VI) \text{ are satisfied} \} \]

and the metric entropy with bracketing of \( f \) with respect to \( \rho \)

\[ H^B(\delta, S, \rho) = \log \nu^B(\delta, S, \rho). \]

The condition that we need now to apply a central limit theorem, is that

\[ \int_0^1 (H^B(\delta, S, \rho))^{1/2} d\delta < \infty, \]

since in that case

\[ S_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n K(s, V_i) \rightarrow_L Z(s) \]

where \( \{Z(s) : s \in S\} \), is the mean 0 gaussian process with covariance

\[ \text{cov}(Z(s), Z(t)) = \text{cov}(K(s, V), K(t, V)) \]

for \( s, t \in S \), by Theorem 3.1 Ossiander [1987], page 904.

Here we notice that we can relate the entropies \( H^B(\delta, S, \rho) \) and \( \log N^{(\infty)}(\delta, \mathcal{G}_{\alpha, M2}, P) \), since

\[ \rho(s_i, s_j) = \|f_{s_i} - f_{s_j}\| = \sup_{v \in [0,1]} |K(s_i, y) - K(s_j, y)| \]

we have that \( N^{(\infty)}(\delta, \mathcal{G}_{\alpha, M2}, P) \geq \nu^B(\delta, S, \rho) \).

So our problem reduces to finding out for which value of \( \alpha \) we have \( \int_0^1 (J\delta^{-2/\alpha})^{1/2} d\delta < \)
$\infty$, because of Theorem 3.2.11.

Recall that $\alpha = n + \gamma$ where $n \in \{0, 1, \ldots\}$ and $0 < \gamma < 1$, so we need $\alpha = n + \gamma$ with minimal $n$, obviously such that $\int_0^1 \delta^{-1/\alpha} d\delta < \infty$. This condition is satisfied for any $\alpha > 1$, so that, all we need is $\alpha = 1 + \gamma$ for $0 < \gamma < 1$.

Summing up, if $K \in C_{1+\gamma}$ for $0 < \gamma < 1$, then $S_n \rightarrow_L Z$, and we have the following

**Theorem 3.2.13** Let $K : [0, 1]^2 \rightarrow \mathbb{R}$ such that $K \in C_{1+\gamma}$ for $0 < \gamma < 1$, then

i) $K$ has an absolutely convergent trigonometric series, and so it is the kernel of a trace class operator for any $P$ probability measure

ii) If $K$ satisfies (II) and the separability condition on $F$ holds, then

$$S_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} K(s, V_i) \rightarrow_L Z(s)$$

where $Z$ is the mean zero gaussian process with covariance

$$\text{cov}(Z(s), Z(t)) = \text{cov}(K(s, V), K(t, V))$$

for $s, t \in S$.

### 3.3 Measurability and Majorizing Measures

As we mentioned in the last section, Ossiander's Theorem has an assumption on measurability of $F$, that is, $F$ has to be separable, but is not clear in the proof of her central limit theorem, where the separability of $F$ is needed. It seemed that this condition could be removed without altering the main result of her Theorem. Fortunately this is so, as it was proved in a later paper of Andersen et al. [1988], which proves that Ossiander's central limit theorem can be improved replacing $L^2$-brackets by $\Delta_{2,\infty}$-brackets, and the entropy condition by the weaker majorizing measure condition. In this section we present the main results of the later paper.

Let $(T, d)$ be a pseudometric space and $\mu$ a Borel probability measure on $(T, d)$.

**Definition 3.3.1** $\mu$ is a majorizing measure for $(T, d)$ iff

$$\sup_{t \in T} \int_0^\infty \left[ \log(\mu\{B_d(t, \epsilon)\}) \right]^{-1/2} d\epsilon < \infty,$$

where $B_d(t, \epsilon) = \{s \in T : d(s, t) \leq \epsilon\}$. 

57
Definition 3.3.2 \( \mu \) is a discrete majorizing measure if there exists a (countable) set \( S \subseteq T, \quad S = \{\pi_q t : t \in T, q \in \mathbb{N}\} \), which supports \( \mu \) and satisfies
\[
d(t, \pi_q t) \leq 1/2^q \quad t \in T, q \in \mathbb{N}
\]
and
\[
\sup_{t \in T} \sum_{q=1}^{\infty} \frac{1}{2^q} \left[ \log(\mu(\pi_q t))^{-1} \right]^{1/2} < \infty.
\]

And we can relate these two definitions by the following:

Theorem 3.3.3 If \((T,d)\) has a majorizing measure, then it also has a discrete majorizing measure. If the majorizing measure \( \nu \) satisfies
\[
\limsup_{t \in T} \int_0^\delta \left[ \log(\nu(B_d(t, \epsilon)))^{-1} \right]^{1/2} d\epsilon = 0,
\]
then the discrete majorizing measure \( \mu \) can be chosen to satisfy
\[
\limsup_{k \to \infty} \sum_{t \in T} \frac{1}{2^q} \left[ \log(\mu(\pi_q t))^{-1} \right]^{1/2} = 0.
\]

Proof: See Andersen et al. [1988], page 275 lemma 2.1.

Let \((S,S)\) be a measure space, let \( \{P_n_j : j = 1, \ldots, n, \quad n \in \mathbb{N}\} \) be probability measures on \((S,S)\) and let \( \mathcal{F} \subset \cap_{n,j} L_1(S,S,P_{n,j}) \) be such that
\[
\sup_{f \in \mathcal{F}} |f(s)| < \infty \quad \forall s \in S.
\]

Let \( (\Omega^n, \Sigma_n, P_{jr}) = (S^n, S^n, P_{n1} \otimes \cdots \otimes P_{nn}) \times ([0,1], B, \lambda) \), where \( \lambda \) is the usual Lebesgue measure on \([0,1]\), let \( X_{nj} : \Omega_n \to S \) the coordinates projections and let \( \{a_n\} \) be a sequence of real positive numbers, then:

Definition 3.3.4 \( \mathcal{F} \) satisfies the central limit theorem with centering at expectations with respect to \( \{P_{nj}\} \) and \( \{a_n\} - \mathcal{F} \in CLT\{P_{nj}; a_n\} \) for short - if there exists a (centered) Radon measure \( \gamma \) on \( \ell^\infty(\mathcal{F}) \) such that for all \( H : \ell^\infty(\mathcal{F}) \to \mathbb{R} \) bounded and continuous
\[
E^s H(1/a_n \sum_{j=1}^n (\delta_{X_{nj}} - P_{nj})) \to_{n \to \infty} \int H d\gamma.
\]

In our case all functions in \( \mathcal{F} \) will be \( S \)-measurable, and we will consider the case \( a_n = \sqrt{n} \).
Theorem 3.3.5 Let $\mathcal{F} \subset \cap_{n,j} L_1(S,S,P_{nj})$ such that $\sup_{f \in \mathcal{F}} |f(s)| < \infty \ \forall s \in S$. Assume that for all $(f_1,\ldots,f_k) \in \mathcal{F}, k \in \mathbb{N}$ the finite dimensional distributions

$$\{L[1/a_n \sum_{j=1}^n (f_j(X_{nj}) - P_{nj} f_j)]\}_{n=1}^\infty$$

converge weakly in $\mathbb{R}^k$. Assume further that there are maps $\pi_q : \mathcal{F} \to \mathcal{F}, q \in \mathbb{N}$ such that

i) $\text{card}\{\pi_q f : f \in \mathcal{F}\} < \infty, \ \forall q \in \mathbb{N}$

and for all $\epsilon > 0$, with $\{\epsilon_i\}$ a Rademacher sequence defined on $([0,1],B,\lambda)$

ii) $\lim_{q \to \infty} \limsup_n P_{n} \{1/a_n \sum_{j=1}^n \epsilon_j (f - \pi_q f)(X_{nj})\} > \epsilon = 0$

and

iii) $\lim \limsup_n \sup_{f \in \mathcal{F}} P_{n} \{1/a_n \sum_{j=1}^n [(f - \pi_q f)(X_{nj}) - P_{nj} (f - \pi_q f)] > \epsilon\} = 0$

then $\mathcal{F} \in CLT\{P_{nj},a_n\}$.

Proof: See Andersen et al. [1988], page 282, Theorem 2.9.

Lemma 3.3.6 Let $\{Z_i\}_{i=1}^k, k \leq \infty$ be a sequence of independent nonnegative real random variables and let

$$\|\{Z_i\}\|_{2,\infty} := (\sup_{a>0} a^2 \sum_{i=1}^k 1_{[Z_i > a]})^{1/2}$$

which is equivalent to the $l_{2,\infty}$ norm of $\{Z_i\}$. The following inequalities hold:

a) $\|\{Z_i\}\|_{2,\infty}^2 \leq \sup_{a>0} a \sum_{i=1}^k Z_i 1_{[Z_i > a]} \leq 2 \|\{Z_i\}\|_{2,\infty}^2$

and

b) if $K := \sup_{t>0} t^2 \sum_{i=1}^k P\{Z_i > t\} < \infty$

and

$z_i(\omega) \leq b < \infty \ \ \omega - a.s., i=1,\ldots,k.$

then for $c > eK$,

$$P\{\|\{Z_i\}\|_{2,\infty}^2 > c\} \leq (1 - eK/c)^{-1} \exp(-cb^2 \log(c/eK)).$$
Proof: See Andersen et al [1988], page 286, lemma 2.16.

In the proof of this last lemma, they use an inequality for binomial probabilities given by (Gine and Zinn [1984]), which states:
If \( PA_j = p_j, j = 1, \ldots, n \) and the \( A_j \)'s are independent then
\[
P(\sum_{j=1}^{n} 1_{A_j} \geq l) \leq \left( e \sum_{j=1}^{n} p_j / l \right)^l.
\]

Using all these results, we can prove the basic inequality needed for the proof of the central limit theorem.

**Theorem 3.3.7** Let \((\Omega, \Sigma, \Pr)\) be a probability space, \((S, S)\) a measurable space and \(\{X_i\}_{i=1}^{k}\) a sequence of independent \(S\)-valued random variables with laws \(\mathcal{L}(X_i) = P_i\).
Let \(\mathcal{F} \subset \cap_{i=1}^{k} \mathcal{L}_0(S, S, P_i)\). For some \(q_0 \in \mathbb{N}\) assume that \(\forall q \geq q_0\) there exists a set \(T_q\), and functions \(t_q : \mathcal{F} \to T_q, \pi_q : T_q \to \cap_{i=1}^{k} \mathcal{L}_0(S, S, P_i)\), \(\gamma_q : T_q \to \mathbb{R}_+ \setminus \{0\}\) and \(\Delta_q \in T_q \to \mathcal{L}_0(S, S, P_i)\) for \(1 \leq i \leq k\), such that, for \(\pi_q = \pi_q \circ t_q\), \(\gamma_q = \gamma_q \circ t_q\) and \(\Delta_q = \Delta_q \circ t_q\), we have
\[
|\langle f - \pi_q f \rangle| \leq \Delta_q^{-i} f \quad \text{for} \quad f \in \mathcal{F}, 1 \leq i \leq k,
\]
\[
\beta_{q_0} := \sup_{f \in \mathcal{F}} \sum_{q \geq q_0} (1/2^q) \gamma_q f < \infty,
\]
\[
\sum_{i=1}^{k} E((\pi_q f - \pi_{q-1} f)(X_i))^2 \leq K/2^{2q}, \quad f \in \mathcal{F},
\]
\[
\sup_{t > 0} t^2 \sum_{i=1}^{k} \Pr\{\Delta_i f(X_i) > t\} \leq K/2^{2q}, \quad f \in \mathcal{F},
\]
\[
\Delta_q^{-i} f \downarrow \text{ as } q \uparrow \quad \text{for } f \in \mathcal{F}, 1 \leq i \leq k, \text{ and}
\]
\[
t_q^{-1} f = t_q^{-1} g \quad \text{for all } f, g \in \mathcal{F} \text{ such that } t_q f = t_q g
\]
where \(K < \infty\) is a constant. Then for all \(\alpha \in \mathbb{R}_+\) and \(q_1 > q_0\) such that
\[
\alpha \geq (K e^2 \beta_{q_0}) \vee ((Kk)^{1/2} / 2^{q_1})
\]
we have
\[
\Pr^*\{\left\| \sum_{i=1}^{k} \varepsilon_i (f - \pi_{q_0} f) 1_{\{\Delta_q \leq 1/2^{q_0+1} \gamma_{q_0+1} f\}}(X_i) \right\|_{\mathcal{F}} > 12\alpha\}
\leq 3 \sum_{q=q_0+1}^{q_1} \sum_{t \in T_q} \exp\left(-\frac{3}{14} \alpha \beta_{q_0}^{-1} \gamma_q(t)^2\right),
\]
where \(\{\varepsilon_i\}\) is a Rademacher sequence independent of \(\{X_i\}\).
Proof: Andersen et al. [1938], page 288, Thm. 3.1.

The proof uses lemma 3.3.7 and Bernstein's inequality. Now we are ready to give a central limit theorem for non-i.i.d. random variables.

Theorem 3.3.8 Let \( F \subset \cap_{n,j} L_1(S,S,P_{nj}) \) satisfying \( \sup_{f \in F} |f(s)| < \infty \) \( \forall s \in S \) and let \( F \) be the envelope function of \( F \). Assume

(i) For every \( k \in \mathbb{N} \), and \( f_1, \ldots, f_k \in F \), the finite dimensional distributions

\[
\{ \mathcal{L}[{(1/a_n)^{1/n}} \sum_{i=1}^{n}(f_i(X_{nj}) - P_{nj}f_i)_{i=1}^{k}] \}_{n=1}^{\infty}
\]

converge weakly;

(ii)

\[
\sum_{j=1}^{n} P_{nj}^{*} \{ F > ta_n \} \rightarrow_{n \to \infty} 0 \quad \forall t > 0
\]

and that there exists a pseudo-distance \( \rho \) on \( F \) dominated by \( d_G \) (where \( d_G(f,g) = \sqrt{E(G(f) - G(g))^2} \), \( G \) a centered gaussian process indexed by \( F \)) with bounded \( d_G \)-uniformly continuous paths, such that (iii)

\[
(1/a_n^2) \sum_{j=1}^{n} P_{nj}^*(f-g)^2 \leq \rho^2(f,g) \quad \forall f, g \in F
\]

and

(iv) for all \( f \in F \) and \( \epsilon > 0 \)

\[
\sup_{t > 0} \epsilon^2 \sum_{j=1}^{n} P_{nj}^* \left\{ \sup_{s \in B_s(f,t)} |f - g| > t \lambda_n \right\} \leq \epsilon^2.
\]

Then \( F \in CLT\{P_{nj}; a_n\} \) and the limiting measure is gaussian.

Proof: The existence of \( \mu \) a discrete majorizing measure for \( d = d_G \) is guaranteed by Talagrand's Theorem 14 in his paper of 1987, so, if for \( T := \{ \pi_q f : f \in F, q \in \mathbb{N} \} \) satisfying the conditions of definition 3.3.2, we have that the \( \pi_q \)'s satisfy conditions i) and iii) of Theorem 3.3.5, so the only condition to be checked is ii), but this follows from Theorem 3.3.7.

We have the following corollary:

Corollary 3.3.9 Let \( F \subset L_2(S,S,P) \) and let \( F \) be its envelope function. Assume \( \|PF\|_F < \infty \) and
i) \[ \lim_{t \to \infty} t^2 P\{F^* > t\} = 0 \]

ii) \[ F \text{ is } P\text{-pregaussian} \]

iii) there exists a bounded and $d_q$-uniformly continuous centered Gaussian process $G$, such that for all $\epsilon > 0$ and all $f \in F$

\[ \Lambda^2_{d,\infty} \left( \sup_{g \in B_d(f, t)} |f - g|^* \right) \leq \epsilon \]

then $F \in \text{CLT}(P)$.

**Proof:** follows directly from Theorem 3.3.8. \[\square\]

We now are ready to give a theorem that relates this paper with Ossiander [1987], that is, we will give a “bracketing” form of Theorem 3.3.8.

**Theorem 3.3.10** Let $F \subset L_1(S, \mathcal{S}, P)$ and let $F$ be its envelope function. Assume

\[ \|Pf\|_\mathcal{F} < \infty \text{ and} \]

i) \[ \lim_{t \to \infty} t^2 P(F^* > t) = 0, \]

ii) \[ F \text{ is } P\text{-pregaussian}, \]

iii) for all $q \in \mathbb{N}$ and $f \in F$ there exist measurable functions $l_q(f)$ and $u_q(f)$ on $(S, \mathcal{S})$ and there exists a finite measure $\mu$ on the pairs $\{(l_q(f), u_q(f)) : q \in \mathbb{N}, f \in F\}$ such that

(a) \[ l_q(f) \leq f \leq u_q(f), \text{ for all } f \in F \text{ and } q \in \mathbb{N}, \]

(b) \[ \sup_{f \in F} \Lambda^2_{d,\infty} (u_q(f) - l_q(f)) \leq 1/2^q, \text{ for all } q \in \mathbb{N}, \]

(c) \[ \lim_{r \to \infty} \sup_{f \in F} \sum_{q=r}^{\infty} 2^{-q} \left[ \log(\mu((l_q(f), u_q(f))^{-1}\right) = 0 \]

and the sup is finite for $r = 1$.

Then $F \in \text{CLT}(P)$.

**Proof:** Assume the hypotheses in Corollary 3.3.9 hold. Let $\{\pi_q f\}$ and $\nu$ satisfy the conditions of Definition 3.3.2 and Theorem 3.3.3. Define

\[ u_q(f) = \pi_q+1 f + \left( \sup_{g \in B_d(\pi_{q+1} f, 2^{-q-1})} |\pi_{q+1} f - g|^* \right), \]
\[ l_q(f) = \pi_{q+1} f - \left( \sup_{g \in B_2(\pi_{q+1} f, 2^{-q-1})} |\pi_{q+1} f - g| \right). \]

and \( \mu(\{ l_q(f), u_q(f) \}) = \nu(\pi_{q+1} f) \), for all \( f \in \mathcal{F} \). Then \( l_q(f), u_q(f) \) and \( \mu \) satisfy condition (iii) in Theorem 3.3.10. In the other direction, assume condition (iii) in Thm. 3.3.10 holds. Define

\[ h_q(f) = \{(l_r(f), u_r(f))\}_{r=1}^{\infty} \quad \forall f \in \mathcal{F}, q \in \mathbb{N} \]

and \( \mu_1(h_q(f)) = 2^{-q} \prod_{r=1}^{\infty} \mu(\{(l_r(f), u_r(f))\}) \). Then \( \mu_1 \) satisfies the conditions in Thm. 3.3.3 with \( \pi_q \) replaced by \( h_q(f) \). Define

\[ \rho_1(f, g) = 2^{-r(f,g)+1} \quad \text{where} \quad r(f, g) = \inf\{ q : h_q(f) \neq h_q(g) \}. \]

Choose a function \( h \) in each class \( \{ g : h_q(f) = h_q(g) \} \), \( q \in \mathbb{N}, f \in \mathcal{F} \), and define \( \pi_{q-1} f = h \). Let \( \nu(\pi_{q-1} f) = \mu_1(h_q f) \), \( q \in \mathbb{N}, f \in \mathcal{F} \). Since \( \rho_1(f, \pi_q f) \leq 2^{-q} \) and \( \nu \) verifies conditions on 3.3.3, it follows that \( \nu \) is a majorizing measure for \( \rho_1 \). So, \( \rho_1 \) is dominated by the pseudo-distance \( d_G \) of a Gaussian process \( G \) with bounded and \( \rho \)-uniformly continuous sample paths. Finally the \( \Lambda_{2,\infty} \)-condition in Theorem 3.3.9 is obvious from the definition of \( \rho(\rho_1) \) and the \( h_q \)'s and from (iii)(c) in Theorem 3.3.10. \( \square \)

Therefore, Theorem 3.3.10 implies the result in Ossiander [1987] just replacing \( L^2 \)-brackets by \( \Lambda_{2,\infty} \)-brackets, and the entropy condition by the majorizing measure condition.
Chapter 4

von Mises Functionals

4.1 Basic Concepts

Consider a measurable space \((X, \mathcal{A})\) and a function \(K : X^2 \to \mathbb{R}\) and two finite signed measures \(\mu, \nu\) on \((X, \mathcal{A})\).

Define a functional \(T\) with kernel \(K\) as follows:

\[
T(\mu, \nu) := T_{K}(\mu, \nu) := \int \int K(x, y) d\mu(x) d\nu(y).
\]

We are interested in the case in which \(\mu\) and \(\nu\) are probability measures, or differences of probability measures. Let us consider first the case in which \(\mu = \nu = P\), a fixed probability measure. In such a case we denote \(T(P) := T(P, P)\), and we are concerned with the properties of this functional.

Functionals of this form were first proposed by von Mises in 1947, and since then, they have been studied by many researchers, among them we can mention: Filippova [1961], Reeds [1976], Fernholz [1983] and many others.

von Mises first noticed that several known statistics can be expressed as statistical functionals with diverse kernels of different dimensions, in fact most of the known statistics can be expressed in this form, for examples we refer to the work of Reeds [1976], Fernholz [1983] and Filippova [1961]. Many advances have been made in Probability Theory since then, which allow us to give a different approach to the problem of finding limiting distributions and establishing "universal" results independently of \(P\), the probability law of the population.

von Mises considered a statistical functional \(T(P)\), as a direct function of the probability measure \(P\), and using the concept of Gateaux differentiability (defined below), he developed \(T(P)\) as a Taylor series, in order to express the difference \(T(Q) - T(P)\) as a first derivative term \(\int f_P d(Q - P)\) plus a remainder \(r(Q, P)\).
For the empirical measure $P_n$ determined by a sample of order $n$, letting $Q = P_n$ we have

$$T(P_n) - T(P) = \int f_P d(P_n - P) + r(P_n, P)$$

where in principle $\int f_P d(P_n - P)$ is of order $1/\sqrt{n}$ and $r(P_n, P) = o_P(1/\sqrt{n})$ as $n \to \infty$. If so,

$$n^{1/2}(T(P_n) - T(P)) = n^{1/2} \int f_P d(P_n - P) + o_P(1)$$

and the asymptotic distribution of $T(P_n)$ is determined by the first derivative term.

In order to start our analysis we first give the definition of three kinds of differentiability following Reeds [1976], page 45, Chapter II.

**Definition 4.1.1** Let $B_1, B_2$ be topological vector spaces and $L(B_1, B_2)$ the set of all continuous linear transformations from $B_1$ to $B_2$. For $\phi : B_1 \to B_2$, $x \in B_1$ and $d\phi(x) \in L(B_1, B_2)$ define $Rem : B_1 \to B_2$ by

$$Rem(x + h) := \phi(x + h) - \phi(x) - d\phi(x)h$$

and $Q : \mathbb{R} \times B_1 \to B_2$ by

$$Q(t, h) = \begin{cases} 
0 & \text{for } t = 0 \\
\frac{Rem(x + th)}{t} & \text{for } t \neq 0.
\end{cases}$$

Then we will say

(1) $\phi$ is Gateaux differentiable, or that $\phi$ has a directional derivative at $x$, with derivative $d\phi(x)$ if

$$\forall h \in B_1 \quad Q(t, h) \to 0 \text{ as } t \to 0.$$ 

(2) $\phi$ is Hadamard or compactly differentiable at $x$ with derivative $d\phi(x)$ if

$$\forall K \subset B_1 \text{ compact, } Q(t, h) \to 0 \text{ uniformly in } h \in K.$$ 

(3) $\phi$ is Fréchet differentiable at $x$ with derivative $d\phi(x)$ if

$$\forall B \subset B_1 \text{ bounded, } Q(t, h) \to 0 \text{ uniformly in } h \in B.$$ 

It is clear from these definitions that

$\phi$ Fréchet differentiable $\Rightarrow$ $\phi$ compactly differentiable $\Rightarrow$ $\phi$ Gateaux differentiable.

As noted by Reeds, Gateaux differentiability is the weakest one and Fréchet differentiability is the strongest. In fact, functions that are only Gateaux differentiable are
not too useful for our purposes. For examples of functions that are Gateaux differentiable, but not Fréchet or compactly differentiable and a function that is compactly differentiable, but not Fréchet differentiable, see Reeds [1976], pages 48-51.

We can also define the concept of differentiability when \( \phi \) is only defined on some convex subset, and this will be the definition that we will adopt.

Consider \( \mathcal{P} \) the space of all probability measures on a measurable space \((X, \mathcal{A})\), let \( P, Q \in \mathcal{P} \), two probability measures, and take the “line segment” joining \( P \) and \( Q \), that is, the set of all probability measures that belong to \( \{(1 - \lambda)P + \lambda Q : 0 \leq \lambda \leq 1\} \).

**Definition 4.1.2** If \( T \) is a functional defined for \( P + \lambda(Q - P) \) for small \( \lambda \) and if
\[
\frac{dT(P; Q - P)}{d\lambda} := \lim_{\lambda \to 0+} \frac{T(P + \lambda(Q - P)) - T(P)}{\lambda}
\]
exists, then we call \( dT(P; Q - P) \) the Gateaux derivative of \( T \) at \( P \) in the direction of \( Q \).

In a similar way we can define higher order derivatives by
\[
d_k T(P; Q - P) := \frac{d^k}{d\lambda^k} T(P + \lambda(Q - P))|_{\lambda=0+}.
\]
In the case of \( K : X^2 \to \mathbb{R} \) and for the statistical functional generated by \( K \), \( T(P) = \iint K(x, y)dP(x)dP(y) \), we have
\[
dT(P; Q - P) = \frac{d}{d\lambda} T(P + \lambda(Q - P))|_{\lambda=0+} =
\]
\[
\frac{d}{d\lambda} \left( \iint K(x, y)d(P + \lambda(Q - P))(x)d(P + \lambda(Q - P))(y) \right)|_{\lambda=0+} =
\]
\[
\frac{d}{d\lambda} \left\{ \iint K(x, y)dP(x)dP(y) + \lambda \iint K(x, y)dP(x)d(Q - P)(y) 
+ \lambda \iint K(x, y)d(Q - P)(x)dP(y) + \lambda^2 \iint K(x, y)d(Q - P)(x)d(Q - P)(y) \right\}|_{\lambda=0+} =
\]
\[
\iint K(x, y)dP(x)d(Q - P)(y) + \iint K(x, y)d(Q - P)(x)dP(y).
\]
In the rest of this chapter, we will use the stronger sense of Fréchet differentiability, and we will need a wider definition, in order to do so we will use Serfling [1980].

Let \( \mathcal{D} \) be the space of all signed measures which can be expressed as a difference of two probability measures times a constant, i.e. the set of all finite signed measures of total mass zero,
\[
\mathcal{D} = \{ c(P - Q) : P, Q \in \mathcal{P}, c \in \mathbb{R} \}.
\]
If we equip \( \mathcal{D} \) with a norm \( \| \cdot \| \), then we can define:
**Definition 4.1.3** For $T$ defined as above, we say that $T$ has a differential at $P \in \mathcal{P}$ with respect to $\| \cdot \|$, if there exists a functional $T(P; c(Q - P))$ defined for $c(Q - P) \in \mathcal{D}$ and linear in $c(Q - P)$, such that

$$T(Q) - T(P) - T(P; Q - P) = o(\|Q - P\|)$$

as $\|Q - P\| \to 0$. Then $T(P; Q - P)$ is called the Fréchet differential.

In order to relate this last definition with the previous one, we have the following:

**Lemma 4.1.4** If $T$ has a differential at $P$ with respect to $\| \cdot \|$, then for any $Q$

$$dT(P, Q - P) = T(P; Q - P)$$

**Proof:** See Serfling [1980], page 218, lemma A.

Therefore, to show that $T$ is Fréchet differentiable with respect to a norm $\| \cdot \|$ on $\mathcal{D}$, we have to show that:

$$T(Q) - T(P) - dT(P; Q - P)$$

$$= \iint K(x, y)dQ(x)dQ(y) - \iint K(x, y)dP(x)dP(y)$$

$$- \iint K(x, y)dP(x)d(Q - P)(y) - \iint K(x, y)d(Q - P)(x)d(Q - P)(y)$$

$$= \iint K(x, y)d(Q - P)(x)d(Q - P)(y)$$

is $o(\|Q - P\|)$.

So, for $T$ to be differentiable, we need to have a norm $\| \cdot \|$ “large” enough to check that:

$$\iint K(x, y)d(Q - P)(x)d(Q - P)(y)$$

is $o(\|Q - P\|)$.

On the other hand, if we take $Q = P_n$ the empirical measure concentrated on $n$ observations with law $P$, we have that if $\sqrt{n}\|P_n - P\| = O_p(1)$ then $\sqrt{n}(T(P_n) - T(P) - dT(P; P_n - P))$ converges to 0 in probability, but $\sqrt{n}\|P_n - P\| = O_p(1)$ is only satisfied if the norm $\| \cdot \|$ is “small”. Then we have a conflicting problem with the size of the norm $\| \cdot \|$. We will try to find a norm $\| \cdot \|$ on $\mathcal{D}$ such that both conditions are satisfied. But first we give an example of a functional $T$ which is not Fréchet differentiable for $\| \cdot \|_{\text{sup}}$.

**Example** Let $K(x, y) = 1_{x \leq y}$, $x, y \in [0, 1]$ and consider the functional

$$T(P) = \iint K(x, y)dP(x)dP(y)$$
\[
= \int_0^1 \int_0^v dP(x) dP(y) = \int_0^1 F_P(y) dP(y)
\]
where \(F_P(\cdot)\) is the distribution function of \(P\).

Let \(P, Q \in \mathcal{P}\) and let \(G = F_Q, F = F_P, H = G - F\), then
\[
T(Q - P, Q - P) = \int_0^1 H(y) dH(y).
\]
We will see that \(T(Q - P, Q - P)\) is not \(o(\|H\|_{\text{sup}})\) as \(\|H\|_{\text{sup}} \to 0\). First we notice that \(H(0^-) = H(1) = 0\), and total variation \((H) \leq 2\). For \(n \geq 1\) let
\[
P_n = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\left(\frac{j}{2n}\right)}
\]
and
\[
Q_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\left(\frac{j}{n}\right)}.
\]
Then for \(n\) fixed we have:
\[
F_{P_n}(x) = \begin{cases} 
0 & \text{for } x < \frac{1}{2n} \\
\frac{i+1}{n} & \text{for } \frac{1}{2n} \leq x < \frac{1}{2n} + \frac{i+1}{n} \\
1 & \text{for } x \geq \frac{1}{2n} + \frac{i+1}{n}
\end{cases}
\]
and
\[
F_{Q_n}(x) = \begin{cases} 
0 & \text{for } x < \frac{1}{n} \\
\frac{i}{n} & \text{for } \frac{i}{n} \leq x < \frac{i+1}{n} \\
1 & \text{for } x \geq 1.
\end{cases}
\]
Therefore
\[
H_n(x) = (F_{Q_n} - F_{P_n})(x) = \begin{cases} 
0 & \text{for } x < \frac{1}{2n} \\
-\frac{1}{n} & \text{for } \frac{1}{2n} \leq x < \frac{2}{2n} \\
0 & \text{for } \frac{2}{2n} \leq x < \frac{3}{2n} \\
-\frac{1}{n} & \text{for } \frac{3}{2n} \leq x < \frac{4}{2n} \\
\vdots & \vdots \\
0 & \text{for } x \geq 1
\end{cases}
\]
From these expressions it is easy to see that:
\[
\int H_n dQ_n = \sum_{j=1}^{n} H_n(j/n) 1/n = 0
\]
and
\[
\int H_n dP_n = \sum_{j=0}^{n-1} H_n(1/2n + j/n) 1/n = (1/n) \sum_{j=0}^{n-1} (-1/n) = -1/n.
\]
So, \(\int H_n d(Q_n - P_n) = 1/n = \|H_n\|_{\text{sup}}\), but
\[
|T(Q_n - P_n, Q_n - P_n)| = |\int H_n(y) dH_n(y)| = 1/n \neq o(\|H_n\|_{\text{sup}}).
\]
Therefore, \(T\) is not differentiable for \(\|\cdot\|_{\text{sup}}\).

68
4.2 Differentiability

We saw in the last section that for $K$ a kernel defined in $X^2$ and the functional $T(P) = \iint K(x, y) dP(x) dP(y)$, we can define a differentiability notion in terms of a norm $\| \cdot \|_V$ in the space $\mathcal{D}$, that is, all finite signed measures of total mass zero, in the measurable space $(X, \mathcal{A})$. We want to find a norm $\| \cdot \|$, such that $T$ is Fréchet differentiable with respect to it, see Definition 4.1.3.

In the last years, many results in Probability Theory have been found using a norm specified by a certain family of measurable functions $\mathcal{F}$ that is uniformly bounded, we can define:

Definition 4.2.1 For $\mathcal{F}$ as above, and $P, Q \in \mathcal{P}$. Let

$$\|P - Q\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\int f(x) d(P - Q)(x)|.$$ 

Depending on the selection of $\mathcal{F}$ we can have important properties such as $\mathcal{F}$ being a "universal Donsker class", that is, for every probability measure $P$, the normalized and centered empirical measure $\sqrt{n}(P_n - P)$ converges in law, with respect to uniform convergence over $\mathcal{F}$, to the limiting "Brownian bridge" process. So, another question is, when is $\mathcal{F}$ a universal Donsker class?

For a given $\mathcal{F} \subset \mathcal{F}_1$ and its corresponding norm $\| \cdot \|_{\mathcal{F}}$, and $T$ the functional induced by $K$, if we want $T$ to be Fréchet differentiable with respect to the norm $\| \cdot \|_{\mathcal{F}}$, we already know that we need

$$\iint K(x, y) d(Q - P)(x) d(Q - P)(y) = o(\|Q - P\|_{\mathcal{F}}).$$

It seems natural to ask that for every $y \in X$, $\int K(x, y) d(Q - P)(x) \in \mathcal{F}$, but if we use $\| \cdot \|_{\mathcal{F}}$ as defined above then

$$\alpha = |\int (\int K(x, y) d(Q - P)(x)) d(Q - P)(y)| \leq \sup_{f \in \mathcal{F}} |\int f(y) d(Q - P)(y)| = \|Q - P\|_{\mathcal{F}},$$

but we can not guarantee that $\alpha = o(\|Q - P\|_{\mathcal{F}})$.

So far we have noticed that considering a trace class operator $T_K$, generated by a kernel $K(x, y)$, gives important results, so we will try to consider a class of functions $\mathcal{F}$ which includes some of those kernels, and a norm $\| \cdot \|$ such that $T$ is Fréchet differentiable with respect to that norm, to do so we recall Definition 3.2.1 and the family defined by

$$\mathcal{G}_{\alpha, M} = \{K : [0,1]^2 \to \mathbb{R} \mid K \in C_\alpha \text{ and } \|K\|_\alpha \leq M \}$$
where $\alpha = 1+\gamma$, with $0 < \gamma < 1$ and $f \in C_{\alpha}$ means that $K$ has first partial derivatives and they satisfy a Lipschitz condition of order $\gamma$, and the norm is the one we defined in $C_{\alpha}$ by

$$
\|K\|_{\alpha} = \max_{p \leq 1} \sup_{x \in \mathbb{R}^d} \{ |D^p K(x)| \} + \max_{p=1} \sup_{x \neq y} \left\{ \frac{|D^p K(x) - D^p K(y)|}{|x - y|^\gamma} \right\}
$$

where $|u| = (u_1^2 + u_2^2)^{1/2}, u \in \mathbb{R}^2$ and $x, y \in \mathbb{R}^2$.

Let $\Gamma$ be the set of all laws on $I^2$ of the form $(1/n) \sum_{j=1}^n \delta_{x(j)}$ for some $x(j) \in I^2, j = 1, \ldots, n$, and let $F(x,y) = M$ on $I^2$. For $\epsilon > 0$, and $Q \in \Gamma$ let

**Definition 4.2.2**

$$
D_F^2(\epsilon, Q, G_{1+\gamma,M,2}) := \sup \left\{ m : \text{for some } K_1, \ldots, K_m \in G_{1+\gamma,M,2}, \text{ and for } i \neq j, \int_0^1 |K_i - K_j|^2 dQ > \epsilon^2 \int_0^1 F^2 dQ = \epsilon^2 M^2 \right\}
$$

By Theorem 3.2.11, for $0 < M < \infty$, as $\epsilon \downarrow 0$

$$
\log D(\epsilon, G_{1+\gamma,M,2}, d_{sup}) \propto \epsilon^{\frac{2}{1+\gamma}}.
$$

Let $D_F^2(\epsilon, G_{1+\gamma,M,2}) = \sup_{Q \in \Gamma} D_F^2(\epsilon, Q, G_{1+\gamma,M,2})$. we will use a proposition proved in Dudley [1984], page 106, Prop. 11.1.4, that states the following inequality:

$$
\log D_F^2(\epsilon, G_{1+\gamma,M,2}) \leq \log D(\epsilon, G_{1+\gamma,M,2}, d_{sup}) \propto \epsilon^{\frac{2}{1+\gamma}}.
$$

Now we notice that $\alpha = 1 + \gamma > 1$, and we apply the next theorem:

**Theorem 4.2.3** If $\alpha > d/2$ then $G_{\alpha,M,d}$ is a functional Donsker class for any $M < \infty$, $d = 1, 2, \ldots$ and any law $P$ on $I^d$.

**Proof:** Use the above inequality taking $d/\alpha$ instead of $2/\alpha$ for the general case, and integrate from zero to one the square root of the log expression on the left. The result follows from Pollard's central limit theorem, see Dudley [1984], page 117, Thm. 11.3.1.

Therefore, $G_{1+\gamma,M,2}$, for $0 < \gamma < 1$ is not only a Donsker class, but a universal Donsker class.

**Remark 4.2.4** The last Theorem is sharp, since for $\alpha = d/2$, $G_{d/2,M,d}$ is not a Donsker class. See Dudley [1984], page 68, Thm. 8.1.1.

In the rest of this section we will define a class of functions $\mathcal{F}$, such that, if $K(x,y) \in \mathcal{F}$, then we will have that $T_K(P)$ is Fréchet differentiable with respect to a certain norm. Let us define
Definition 4.2.5

\[ \mathcal{H}_{\gamma,M,2} = \left\{ K : I^2 \to \mathbb{R} : \right. \]
\[ |K(x,y) - K(u,v)| \leq M(|x-u|^\gamma + |y-v|^\gamma) \]
\[ \text{and } |K(x,y)| \leq M \quad \forall (x,y),(u,v) \in I^2 \}

where \( 0 < \gamma \leq 1, 0 < M < \infty \).

and

Definition 4.2.6

\[ \mathcal{G}_{\alpha,M,1} = \left\{ g : I \to \mathbb{R} : \right. \]
\[ |g(x) - g(y)| \leq M|x-y|^\alpha \]
\[ \text{and } |g(x)| \leq M \quad \forall x,y \in I \}

where \( 0 < \alpha \leq 1, 0 < M < \infty \).

From the definition above it is clear that, if \( \beta \) is such that \( 0 < \alpha < \alpha + \beta \leq 1 \), then

(a) \( \mathcal{G}_{\alpha + \beta,M,1} \subset \mathcal{G}_{\alpha,M,1} \)

and if \( 0 < M \leq N < \infty \), then

(b) \( \mathcal{G}_{\alpha,M,1} \subset \mathcal{G}_{\alpha,N,1} \).

Definition 4.2.7 Let \( \mu \) be a signed measure on \( I \) and define

\[ \|\mu\|_\alpha^* := \sup_{h \in \mathcal{G}_{\alpha,1,1}} \{|\int h d\mu|\}. \]

Remark 4.2.8 \( \mathcal{G}_{\alpha,M,1} \) is the same family given after definition 3.2.2, and by Theorem 4.2.3 if \( \alpha > 1/2 \) then \( \mathcal{G}_{\alpha,M,1} \) is a universal Donsker class for \( 0 < M < \infty \).

Now we will show that if \( \gamma > 1/2 \) then \( T_K \) is Fréchet differentiable for \( \| \cdot \|_{\mathcal{G}_{\alpha,1,1}} \).

Lemma 4.2.9 If \( K \in \mathcal{H}_{\alpha+\beta,M,2} \), where \( \alpha, \beta > 0 \) and \( 0 < M < \infty \), and \( \mu, \nu \) are signed measures on \( I \), then

\[ | \iint K(x,y) d\mu(x)d\nu(y) | \leq 2M\|\mu\|_\alpha^* \|\nu\|_\beta^*. \]

Proof: Let \( \alpha, \beta > 0, M > 0 \) and \( K \in \mathcal{H}_{\alpha+\beta,M,2} \) and consider the difference of the values of the function \( K \) on the vertices of the rectangle generated by \( (x,y) \) and \( (u,v) \in I^2 \).

\[ |K(x,y) - K(x,v) - K(u,y) + K(u,v)| \]
\[ \leq |K(x,y) - K(x,v)| + |K(u,v) - K(u,y)| \]
\[ \leq M|y-v|^\alpha + M|y-v|^\beta = 2M|y-v|^\alpha + \beta. \]
On the other hand
\[ |K(x,y) - K(x,v) - K(u,y) + K(u,v)| \]
\[ \leq |K(x,y) - K(u,y)| + |K(u,v) - K(x,v)| \]
\[ \leq M|x - u|^{\alpha + \beta} + M|x - u|^4 = 2M|x - u|^{\alpha + \beta} \]
so
\[ |K(x,y) - K(x,v) - K(u,y) + K(u,v)| \leq 2M \min\{|x - u|^{\alpha + \beta}, |y - v|^{\alpha + \beta}\}. \]

But for \( w, z \geq 0, \min\{z^{\alpha + \beta}, w^{\alpha + \beta}\} \leq x^{\alpha}w^{\beta}, \) since \( \min\{z^{\alpha + \beta}, w^{\alpha + \beta}\} = (\min\{z, w\})^{\alpha + \beta}, \) and if for example \( z = \min\{z, w\} \) then \( z^{\beta} \leq w^{\beta} \) and so \( \min\{z, w\}^{\alpha + \beta} = z^{\alpha}z^{\beta} \leq z^{\alpha}w^{\beta} \), therefore
\[ |K(x,y) - K(x,v) - K(u,y) + K(u,v)| \leq 2M|x - u|^4|y - v|^\beta. \]

Now let \( f(x) := f_{(y,v)}(x) = K(x,y) - K(x,v) \) then
\[ |f(x) - f(u)| = |K(x,y) - K(x,v) - K(u,y) + K(u,v)| \leq (2M|y - v|^\beta)|x - u|^4, \]
for every \( x, u \in I, \) and
\[ |f(x)| = |K(x,y) - K(x,v)| \leq M|y - v|^4 \leq 2M|y - v|^\beta \]
since \( |y - v| \leq 1, \) for \((y,v) \in I^2.\) So, from the last two expressions we have that
\[ f(x) = f_{(y,v)} \in G_{\alpha, (2M|y - v|^\beta), 1}, \quad \forall (y,v) \in I^2, \]
and in particular \( f_{(y,v)} \in G_{\alpha, 2M, 1} \) for every \((y,v) \in I^2,\) from remark (b). Besides
\[ |\int f_{(y,v)}(x)d\mu(x)| = |\int (K(x,y) - K(x,v))d\mu(x)| \leq 2M|y - v|^\beta \|\mu\|_\alpha, \]
from the definition of \( \| \cdot \|_\alpha, \) for every \((y,v) \in I^2.\) So that, if we let \( \varphi(y) = \int K(x,y)d\mu(x) \) then
\[ |\varphi(y) - \varphi(v)| = |\int (K(x,y) - K(x,v))d\mu(x)| \leq 2M|y - v|^\beta \|\mu\|_\alpha \quad (1) \]
and since \( K(\cdot, y) \in G_{\alpha + \beta, M, 1}, \) for any \( y \in I \) fixed, so, \( |\varphi(y)| = |\int K(\cdot, y)d\mu(\cdot)| \leq M \|\mu\|^{\alpha + \beta}_\alpha, \) but by remark (a) \( G_{\alpha + \beta, M, 1} \subset G_{\alpha, M, 1}, \) so \( |\varphi(y)| \leq M \|\mu\|_\alpha \) and then \( \varphi \in G_{\beta, 2M \|\mu\|_\alpha, 1} \) by (1), and finally
\[ |\int K(x,y)d\mu(x)d\nu(y)| = |\int \varphi(y)d\nu(y)| \leq 2M \|\mu\|^{\alpha}_\alpha \|\nu\|_\beta \]
as we wanted to prove. \( \square \)
Now consider any $0 < \alpha \leq 1$ and define 

$$\rho_\alpha(P, Q) = \|P - Q\|_\alpha^\ast,$$

for $P, Q \in \mathcal{P}$.

We know that $\|\cdot\|_\alpha^\ast$ metrizes convergence of laws, so that if $P_n$ is the empirical measure of a sample of size $n$ and $\alpha, \beta \in (0, 1]$ then

$$\|P_n - P\|_\alpha^\ast \to 0 \quad \text{iff} \quad \|P_n - P\|_\beta^\ast \to 0.$$ 

Notice that for $\alpha = 1$, $\|\cdot\|_1^\ast$ is just the $BL^\ast$ metric.

Suppose that $K \in \mathcal{H}_{\gamma, M, 2}$ for some $\gamma > 1/2$, let $1/2 < \alpha < \gamma$ and $\beta = \gamma - \alpha > 0$, then

$$\left| \int \int K(x, y) d(Q - P)(x) d(Q - P)(y) \right| \leq 2M \|Q - P\|_\alpha^\ast \|Q - P\|_\beta^\ast$$

by the previous Lemma, so

$$\left| \int \int K(x, y) d(Q - P)(x) d(Q - P)(y) \right| = o(\|Q - P\|_\alpha^\ast)$$

as $\|Q - P\|_\alpha^\ast \to 0$, and putting all this together with Definition 4.1.3 we have proved the following:

**Theorem 4.2.10** Let $T_K(P) = \int \int K(x, y) dP(x) dP(y)$ and $K \in \mathcal{H}_{\gamma, M, 2}$ for some $\gamma > 1/2$. Then $T_K(\cdot)$ is Fréchet differentiable at $P$ for $\|\cdot\|_\alpha^\ast$, where $\|\cdot\|_\alpha^\ast = \|\cdot\|_{G_{\alpha, 1, 1}}$, for $\alpha$ as defined above. Even more $G_{\alpha, 1, 1}$ is a universal Donsker class.

**Remark 4.2.11** If we consider any $\gamma \leq 1/2$, then $G_{\alpha, 1, 1}$ for $\alpha \leq 1/2$ is not a Donsker class by remark 4.2.4.

**Remark 4.2.12** If $0 < \gamma \leq 1/2$ we can still have differentiability of $T_K$ at $P$ with respect to $\|\cdot\|_\alpha^\ast$ with any $0 < \alpha < \gamma$.

**Remark 4.2.13** If $K \in \mathcal{H}_{\gamma, M, 2}$ for $\gamma > 1/2, 0 < M < \infty$ then

$$|K(x, y) - K(u, y)| \leq M |x - u|$$

which by Theorem 2.1.5, implies that $T_K$ is a trace class operator. So, $\mathcal{H}_{\gamma, M, 2} \subset$ class of all kernels of trace class operators.

Actually the conditions in Theorem 4.2.10 can be weakened slightly, for this purpose we will need the following:

**Lemma 4.2.14** Let $a, b \in [0, 1]$ and $\varphi, \varepsilon > 0$, such that $\varphi + \varepsilon \leq 1$, then

$$\min(a, b) \leq a^\varphi b^\varepsilon.$$
Proof: Suppose \(0 < a < b < 1\), then
\[
\min(a, b) = a \leq a^{\phi + \epsilon} = a^{\varphi a^\epsilon} < a^{\varphi b^\epsilon}.
\]
Now suppose \(0 \leq b \leq a \leq 1\), then
\[
\min(a, b) = b \leq b^{\phi + \epsilon} = b^{\varphi b^\epsilon} < a^{\varphi b^\epsilon}.
\]
\(\square\)

Now suppose \(\alpha > 1/2\), take \(w\) such that \(\alpha > w > 1/2\) and define
\[
\varphi = w/\alpha,
\]
then \(0 < \varphi < 1\) and take
\[
\epsilon > 0 \quad \text{such that} \quad \varphi + \epsilon \leq 1.
\]
For any \(\alpha > 1/2\) and any small \(\beta > 0\), we define the following family of functions:

**Definition 4.2.15**

\[
\mathcal{G}_{\alpha,N,2,\beta,M} = \left\{ K : I^2 \to \mathbb{R} \mid \begin{array}{l}
|K(x, y) - K(u, y)| \leq N|x - u|^{\alpha}, \\
\text{where } N \text{ is nonnegative,} \\
\text{and } |K(x, y) - K(x, v)| \leq M|y - v|^{\beta} \\
\text{with } x, y, u, v \in I \text{ and } M > 0.
\end{array} \right\}
\]

For \(M > 0\) we define as in 4.2.6 \(\mathcal{G}_{\alpha,M,1}\), which is a universal Donsker class for any \(\alpha > 1/2\) and \(0 < M < \infty\). If we take \(K \in \mathcal{G}_{\alpha,N,2,\beta,M}\), then for \(x, y, u, v \in I\)
\[
|K(x, y) - K(x, v) - K(u, y) - K(u, v)| \\
\leq |K(x, y) - K(u, y)| + |K(x, v) - K(u, v)| \\
\leq |x - u|^{\alpha}A(y) + |x - u|^{\alpha}A(y) \leq 2N|x - u|^{\alpha}.
\]

So, if we let \(f(x) := f_{(x,y)}(x) = K(x, y) - K(x, v)\) for \(y, v \in I \text{ fixed, then}
\[
|f(x) - f(y)| \leq 2N|x - u|^{\alpha},
\]
and we also have
\[
|K(x, y) - K(x, v) - K(u, y) - K(u, v)| \\
\leq |K(x, y) - K(x, v)| + |K(u, y) - K(u, v)| \\
\leq 2M|y - v|^{\beta}.
\]

Let \(\Delta = \max\{M, cN\}\). Then
\[
|K(x, y) - K(x, v) - K(u, y) - K(u, v)|
\]

74
Now we apply Lemma 4.2.14 taking \( a = |x - u|^{\alpha} \), \( b = |y - v|^{\beta} \), \( \varphi \) and \( \epsilon \) as defined above, then we have

\[
2\Delta \min(|x - u|^{\alpha}, |y - v|^{\beta}) \leq 2\Delta|x - u|^\alpha|y - v|^\beta \epsilon
\]

where \( w > 1/2 \) and \( \beta \epsilon > 0 \). Therefore,

\[
|K(x, y) - K(x, v) - K(u, y) - K(u, v)| \leq 2\Delta|x - u|^w|y - v|^\beta \epsilon.
\]

Now we refer to Lemma 3.2.9, and we follow the argument after (***) in the exact same way taking \( \Delta \) instead of \( M \), \( w \) instead of \( a \) and \( \beta \) instead of \( \beta \) to get:

**Theorem 4.2.16** Let \( T_K(P) = \int \int K(x, y)P(x)P(y) \), where \( K \in \mathcal{G}_{\alpha,N,2,\beta,M} \) for some \( \alpha > 1/2 \) and any \( 0 < N < \infty, 0 < M < \infty \) and \( \beta > 0 \). Then for \( \alpha > w > 1/2 \), \( T_K(\cdot) \) is Fréchet differentiable \( \mathcal{L} \)-\( P \) for \( \| \cdot \|_w = \| \cdot \|_{G_{1,1}} \) and even more \( G_{w,1,1} \) is a universal Donsker class.

**Remark 4.2.17** Just as before, any element of \( \mathcal{G}_{\alpha,N,2,\beta,M} \) is the kernel of a trace class operator by Theorem 2.1.5, in fact to get the class \( \mathcal{G}_{\alpha,N,2,\beta,M} \) from the original hypotheses in the last theorem, all we need is a Lipschitz condition on the second variable, of any positive order and the boundedness of \( A(y) \) required for the universality of the class.

### 4.3 Asymptotic distribution of \( T(P_n) \)

Consider a kernel \( K : I^2 \to \mathbb{R} \), and the functional \( T_K \) it defines for \( P \) any probability measure on \( I \).

\[
T_K(P) = \int \int K(x, y)P(x)P(y).
\]

Recall that the differential of \( T_K \) is given by:

\[
dT(P; Q - P) = \frac{d}{d\lambda}T(P + \lambda(Q - P))|_{\lambda=0^+}
\]

\[
= \int \int K(x, y)d(Q - P)(x)dP(y) + \int \int K(x, y)dP(x)d(Q - P)(y),
\]

and

\[
T(Q) - T(P) = dT(P; Q - P) + \text{Rem}_1
\]

where \( \text{Rem}_1 = \int \int K(x, y)d(Q - P)(x)d(P - Q)(y) \).
Lemma 4.3.1 If $T_K$ has a differential at $P$ with respect to $\| \cdot \|$, and we let $\{ X_i \}$ be observations on $P$ (not necessarily independent) such that $\sqrt{n} \| P_n - P \| = O_p(1)$, where $P_n$ is the empirical measure, then

$$\sqrt{n} \text{Rem}_{1,n} \xrightarrow{P} 0.$$  

Proof: See Serfling [1980], page 218, lemma B.

Remark 4.3.2 We already proved that for $\| \cdot \|_{G_{w,1,1}}$ with $w > 1/2$, $T_K$ has a differential at $P$, even more since $w > 1/2$, $G_{w,1,1}$ is a universal Donsker class, therefore,

$$\sqrt{n} \| P_n - P \|_{G_{w,1,1}} = O_p(1).$$

Then applying Lemma 4.3.1 we have that $\sqrt{n} \text{Rem}_{1,n} \xrightarrow{P} 0$.

If we ask $K$ to be symmetric

$$dT(P; Q - P) = 2 \int \int K(x, y)dP(x)d(Q - P)(y)$$

and if we let $t_1[P; y] = 2 \int K(x, y)dP(x)$, then $dT(P; Q - P) = \int t_1[P; y]d(Q - P)(y)$.

It is possible now to represent $dT(P; P_n - P)$ in the form of a $V$ statistic as defined in Serfling [1980], page 174, to do so we use the following representation:

$$\int t_1[P; y]d(Q - P)(y) = \int \tilde{h}(y)dQ(y)$$

where

$$\tilde{h}(y) = t_1[P; y] - \int t_1[P; y]dP(y)$$

$$= 2 \int K(x, y)dP(x) - 2 \int \int K(x, y)dP(x)dP(y),$$

in particular we have

$$\int \tilde{h}(y)dP(y) = 0 = EP(\tilde{h}(y)).$$

Then

$$dT(P; P_n - P) = \int \tilde{h}(y)dP_n(y) = \frac{1}{n} \sum_{j=1}^{n} \tilde{h}(y_j) = V_{1,n},$$

in fact $\tilde{h}(y) = \tilde{h}(P; y)$. Now

$$\text{Var}_P(\tilde{h}(P; y)) = EP(\tilde{h}(P; y))^2$$

$$= \int (2 \int K(x, y)dP(x) - 2 \int \int K(x, y)dP(x)dP(y))^2dP(y)$$

$$= 4[\int (\int K(x, y)dP(x))^2dP(y) - C^2].$$

So, we can state the following:
Theorem 4.3.3 Let \( \{X_i\}_{i=1}^{\infty} \) be independent observations having law \( P \) on \( I \). Let \( T_K \) the functional induced by \( K \in G_{\alpha,N,2,\beta,M}, \alpha > 1/2, \beta, N, M > 0 \) and assume also that \( K \) is symmetric, let \( \sigma^2(T, P) = \text{Var}_P(h(P; y)) \) and assume that \( 0 < \sigma^2(T, P) < \infty \). Then
\[
T(P_n) \text{ is asymptotically } N(T(P), \frac{1}{n}\sigma^2(T, P)).
\]

Proof: By Theorem A, Serfling [1980], page 226, all we need to see is that \( \sqrt{n}\text{Rem}_{1,n} \xrightarrow{P} 0 \), but this is true if \( K \in G_{\alpha,N,2,\beta,M} \), by remark 4.3.2. \( \square \)
Bibliography


