Master Thesis Project

Comparison of modes of convergence in a particle system related to the Boltzmann equation

Mikael Petersson

LiTH - MAT - EX - - 2010 / 27 - - SE
Comparison of modes of convergence in a particle system related to the Boltzmann equation

Matematiska institutionen, Linköpings Universitet

Mikael Petersson

Lith - MAT - EX - - 2010 / 27 - - SE
Comparison of modes of convergence in a particle system related to the Boltzmann equation

Mikael Petersson

The distribution of particles in a rarefied gas in a vessel can be described by the Boltzmann equation. As an approximation of the solution to this equation, Caprino, Pulvirenti and Wagner [3] constructed a random $N$-particle system.

In the equilibrium case, they prove in [3] that the $L^1$-distance between the density function of $k$ particles in the $N$-particle process and the $k$-fold product of the solution to the stationary Boltzmann equation is of order $1/N$. They do this in order to show that the $N$-particle system converges to the system described by the stationary Boltzmann equation as the number of particles tends to infinity.

This is different from the standard approach of describing convergence of an $N$-particle system. Usually, convergence in distribution of random measures or weak convergence of measures over the space of probability measures is used. The purpose of the present thesis is to compare different modes of convergence of the $N$-particle system as $N$ tends to infinity assuming stationarity.

Random measures, Stochastic particle systems, The Boltzmann equation.
Abstract

The distribution of particles in a rarefied gas in a vessel can be described by the Boltzmann equation. As an approximation of the solution to this equation, Caprino, Pulvirenti and Wagner [3] constructed a random $N$-particle system.

In the equilibrium case, they prove in [3] that the $L^1$-distance between the density function of $k$ particles in the $N$-particle process and the $k$-fold product of the solution to the stationary Boltzmann equation is of order $1/N$. They do this in order to show that the $N$-particle system converges to the system described by the stationary Boltzmann equation as the number of particles tends to infinity.

This is different from the standard approach of describing convergence of an $N$-particle system. Usually, convergence in distribution of random measures or weak convergence of measures over the space of probability measures is used. The purpose of the present thesis is to compare different modes of convergence of the $N$-particle system as $N$ tends to infinity assuming stationarity.

**Keywords:** Random measures, Stochastic particle systems, The Boltzmann equation.
Acknowledgements

I would like to thank my supervisor and examiner Jörg-Uwe Löbus for a very good cooperation during this project. I also want to thank my opponent David Lam for his valuable opinions of the report. Moreover, I want to express my gratitude to Prashant Kara who recommended me for this project. I am very grateful for the fantastic years with my friends at the mathematics program at Linköpings Universitet. My thoughts also go out to my family and other friends.

Mikael Petersson
### Notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B(S)$</td>
<td>Bounded real valued functions on $S$</td>
</tr>
<tr>
<td>$\mathcal{B}(S)$</td>
<td>The Borel $\sigma$-algebra over $S$</td>
</tr>
<tr>
<td>$\mathcal{B}^c(S)$</td>
<td>Relatively compact Borel measurable subsets of $S$</td>
</tr>
<tr>
<td>$C(S)$</td>
<td>Continuous real valued functions on $S$</td>
</tr>
<tr>
<td>$C_b(S)$</td>
<td>Bounded continuous real valued functions on $S$</td>
</tr>
<tr>
<td>$C_b^+(S)$</td>
<td>Continuous non-negative real valued functions on $S$ with compact support</td>
</tr>
<tr>
<td>$\mathbb{C}$</td>
<td>The complex numbers</td>
</tr>
<tr>
<td>$\partial E$</td>
<td>The boundary of $E$</td>
</tr>
<tr>
<td>$E(X)$</td>
<td>The expected value of $X$</td>
</tr>
<tr>
<td>$L^1(S)$</td>
<td>Integrable functions on $S$</td>
</tr>
<tr>
<td>$L^+ (S)$</td>
<td>Measurable functions from $S$ to $[0, \infty]$</td>
</tr>
<tr>
<td>$\mathcal{M}(S)$</td>
<td>Locally finite measures on $S$</td>
</tr>
<tr>
<td>$\mathbb{N}$</td>
<td>The natural numbers, {1, 2, \ldots}</td>
</tr>
<tr>
<td>$P{\xi \in E}$</td>
<td>The probability of the event {$\xi \in E$}</td>
</tr>
<tr>
<td>$\mathcal{P}(S)$</td>
<td>Borel probability measures on $S$</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>The real numbers</td>
</tr>
<tr>
<td>$\mathbb{R}_+$</td>
<td>The non-negative real numbers, $[0, \infty)$</td>
</tr>
<tr>
<td>$\sigma(E)$</td>
<td>The $\sigma$-algebra generated by $E$</td>
</tr>
<tr>
<td>$2^S$</td>
<td>All subsets of $S$</td>
</tr>
<tr>
<td>$\emptyset$</td>
<td>The empty set</td>
</tr>
<tr>
<td>$A \subset S$</td>
<td>$A$ is a subset of, or equal to $S$</td>
</tr>
<tr>
<td>$\xi \overset{d}{=} \eta$</td>
<td>$\xi$ and $\eta$ have the same distribution</td>
</tr>
<tr>
<td>$\xi_n \overset{d}{\rightarrow} \xi$</td>
<td>$\xi_n$ converges in distribution to $\xi$</td>
</tr>
<tr>
<td>$\xi_n \overset{p}{\rightarrow} \xi$</td>
<td>$\xi_n$ converges in probability to $\xi$</td>
</tr>
<tr>
<td>$\mu_n \overset{v}{\rightarrow} \mu$</td>
<td>$\mu_n$ converges vaguely to $\mu$</td>
</tr>
<tr>
<td>$P_n \overset{w}{\rightarrow} P$</td>
<td>$P_n$ converges weakly to $P$</td>
</tr>
<tr>
<td>$|f|_{L^1}$</td>
<td>The $L^1$-norm of $f$</td>
</tr>
<tr>
<td>$|f|_\infty$</td>
<td>The uniform norm of $f$</td>
</tr>
</tbody>
</table>
## Contents

1 Introduction 1

2 Preliminaries 3
   2.1 Metric Spaces and Normed Vector Spaces 3
   2.2 $\sigma$-algebras and Measures 6
   2.3 Integration of Real Valued Functions 9
   2.4 Random Elements 12

3 Convergence of Probability Measures 15
   3.1 Prohorov's Metric 15
   3.2 Prohorov's Theorem 17
   3.3 Weak Convergence of Probability Measures 20
   3.4 Convergence Determining Sets 24

4 Convergence of Random Measures 27
   4.1 Convergence in Distribution of Random Elements 27
   4.2 Random Measures 30
   4.3 Convergence in Distribution of Random Measures 33

5 Convergence of Particle Systems 37
   5.1 The Boltzmann Equation 37
   5.2 The $N$-particle Process 39

6 Comparison of Modes of Convergence 43
   6.1 Comparison of Convergence I 43
   6.2 Comparison of Convergence II 46
   6.3 Speed of Convergence 50
Chapter 1

Introduction

The distribution of particles in a rarefied gas in a vessel can be described by the Boltzmann equation. The solution to this equation is a function \( f(x, t) \) that for each \( t \in [0, T] \) is the probability density function for a random vector \( X \) which describes the state of one particle at time \( t \). The state of one particle is given by its position and its velocity.

To approximate the solution to the Boltzmann equation, Caprino, Pulvirenti and Wagner [3] constructed a system containing \( N \) particles. This system is described by an initial boundary value problem. Its solution is the joint density function for the state of the \( N \) particles. Usually for an \( N \)-particle system at a fixed time \( t \geq 0 \), its limit as the number of particles tends to infinity is expressed

(a) in terms of a weak limit of random measures or

(b) in terms of a weak limit of a sequence of measures over the space of probability measures.

The literature [3] is different. It uses an

(c) "\( L^1 \)-distance between the \( k \)-particle density and the \( k \)-fold product of the solution to the stationary Boltzmann equation".

The objective of this thesis is to clarify the relationship among the modes (a)-(c). In fact, the convergences (a) and (b) are very well established. In this text, a presentation of the convergences (a) and (b) and their relations based on Ethier, Kurtz [4] and Kallenberg [7] is given. We describe the solutions to the Boltzmann equation as the limit of an \( N \)-particle system in the form of [3]. Furthermore, we compare the convergence (c) with (a) and (b).

The outline of the report is as follows. Chapter 2 collects definitions and theorems which will be needed later. Some basic theory about measures and how these are used to define integration is presented. At Linköpings universitet, these topics are not subject to undergraduate education. In order to make the present thesis paper accessible to the master students at the university of Linköping, we include this part. All results stated in this chapter are used in the following chapters. With a few exceptions, the proofs of the results are given by references.

Chapter 3 deals with the theory of weak convergence of measures on arbitrary spaces. As a starting point, we look at the Prohorov metric on the space of
Chapter 1. Introduction

probability measures over some metric space $S$. A characterization of the compact subsets of this space of measures is given by Prohorov’s theorem through the notion on tightness. Another important result is that when $S$ is separable, then weak convergence is equivalent to convergence in the Prohorov metric. A number of ways to verify this convergence is presented. The chapter is based on Ethier, Kurtz [4].

Then in chapter 4 this theory is specialized to measures on spaces of measures. First, the theory of weak convergence is paraphrased as the theory of convergence in distribution. Then random measures are defined and we go over some important results. The final theorem of this chapter gives two ways of verifying convergence in distribution of random measures. The theory of the chapter is built up with the aim of presenting a rigorous proof of this result. Here we use Kallenberg [7].

In chapter 5, the mathematical formulation of the system of particles is given and how this can be approximated by an $N$-particle system. We describe how [3] defines convergence of these systems. This mode of convergence does not involve the theory of chapter 3 and 4.

Finally, chapter 6 compares different modes of convergence of the $N$-particle systems. Two other possible convergence types that involve the theory presented in chapter 3 and 4 are described. We prove two results that give relations between the different modes of convergence. The proofs of these results use ideas from articles of Sznitman [8] and Grigorescu, Kang [6]. By using the calculations in these proofs, we establish two results regarding the speed of the convergence.
Chapter 2

Preliminaries

This chapter collects some background material which will be used later in this thesis. The main purpose of the chapter is to establish the notation and to give an introduction to the reader not familiar with measure and integration theory. Section 1 introduces metric spaces which are used to define distances between elements in abstract spaces. Section 2 contains some basic definitions from measure theory and describes how the important Lebesgue measure is constructed. Section 3 defines integration of real valued functions with respect to a measure. Finally, section 4 gives an introduction to how measure theory are used in probability theory. The material in this chapter is mainly based on Folland [5].

2.1 Metric Spaces and Normed Vector Spaces

This section collects definitions and theorems about metric spaces and normed vector spaces that will be used later.

Metric Spaces

Denote by $\mathbb{R}_+$ the set of all non-negative real numbers, that is $\mathbb{R}_+ = [0, \infty)$. For an arbitrary set $S$, a function $d : S \times S \to \mathbb{R}_+$ is said to be a metric on $S$ if the following holds.

- $d(x, y) = 0$ if and only if $x = y$.
- $d(x, y) = d(y, x)$ for all $x, y \in S$.
- $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in S$.

If $d$ is a metric on $S$, then $(S, d)$, or simply $S$ if the metric is understood, is called a metric space. In a metric space $(S, d)$ a sequence of elements $x_1, x_2, \ldots \in S$ is said to converge to $x \in S$ if for every $\varepsilon > 0$ there exists a positive integer $N$ such that $d(x_n, x) < \varepsilon$ for all $n \geq N$. A sequence $x_1, x_2, \ldots \in S$ is called a Cauchy sequence if for every $\varepsilon > 0$ there exist a positive integer $N$ such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geq N$. A set $E \subset S$ is called complete if every
Cauchy sequence in $E$ converges to some element that belongs to $E$. An open ball centered at $x$ with radius $r > 0$ is defined as the set

$$B(x, r) = \{y \in S : d(x, y) < r\}.$$ 

A subset $E$ of $S$ is said to be open if for every $x \in E$ there exists $r > 0$ such that $B(x, r) \subset E$. For example, every open ball is open so the name is justified. The set $E$ is closed if its complement $E^c$ is open. In this text, the inclusion symbol includes the possibility that the sets are equal. The intersection of all closed sets that contains $E$ is the smallest closed set that contains $E$. It is called the closure of $E$ and is denoted by $\overline{E}$. The boundary of $E$ is defined as

$$\partial E = E \cap E^c$$

and the interior of $E$ is defined as $E^\circ = \overline{E} \setminus E$. A neighbourhood of $x \in S$ is a set $E \subset S$ such that $x \in E^\circ$. If $E$ is a subset of $S$ such that $\overline{E} = S$, it is said to be dense in $S$. A metric space $(S, d)$ is separable if it contains a dense countable subset. The following theorem gives a condition for a subset of a metric space to be complete.

**Theorem 2.1.** If $(S, d)$ is a complete metric space and $E \subset S$ is closed, then the metric space $(E, d)$ is complete.

**Proof.** See section 0.6 in [5].

Let $E$ be a subset of $S$. A family of sets $\{V_\alpha\}_{\alpha \in A}$, where $A$ is some arbitrary index set, is said to be a covering of $E$ if $E \subset \bigcup_{\alpha \in A} V_\alpha$. If for every $\varepsilon > 0$ there exists a finite family of open balls with radius $\varepsilon$ that covers $E$, then $E$ is called totally bounded. A set $K$ is said to be compact if for every family of open sets $\{G_\alpha\}_{\alpha \in A}$ that covers $K$, there exists a finite set $F \subset A$ such that $\{G_\alpha\}_{\alpha \in F}$ covers $K$. If $E$ is a set such that $\overline{E}$ is compact, then it is said to be relatively compact. The following theorem gives two other characterizations of a compact set in a metric space.

**Theorem 2.2.** If $(S, d)$ is a metric space and $K \subset S$, then the following statements are equivalent.

(a) $K$ is compact.

(b) $K$ is complete and totally bounded.

(c) For every sequence of elements in $K$, there exists a subsequence that converges to some element that belongs to $K$.

**Proof.** See section 0.6 in [5].

The next theorem gives a condition for a set to be compact in the special case when $S = \mathbb{R}^d$.

**Theorem 2.3.** If $K$ is a closed and bounded subset of $\mathbb{R}^d$, then $K$ is compact.

**Proof.** See section 0.6 in [5].
Normed Vector Spaces

A norm on a space \( X \) is a function \( \| \cdot \| : X \to \mathbb{R}_+ \) that satisfies the following.

- \( \| x + y \| \leq \| x \| + \| y \| \) for all \( x, y \in X \).
- \( \| \lambda x \| = |\lambda| \| x \| \) for all \( x \in X \) and \( \lambda \in \mathbb{C} \).
- If \( \| x \| = 0 \), then \( x = 0 \).

A space \( X \) such that if \( x, y \in X \) and \( \alpha, \beta \in \mathbb{R} \), then \( \alpha x + \beta y \in X \), is called a vector space over \( \mathbb{R} \). A vector space together with a norm is called a normed vector space. Let \( C(S) \) denote the space of all continuous real valued functions on \( S \) and define

\[
\| f \|_\infty = \sup_{x \in S} |f(x)|, \quad \text{for all } f \in C(S).
\]

This is called the uniform norm and it can be shown that \((C(S), \| \cdot \|_\infty)\) is a normed vector space. Every subspace of \( C(S) \) together with the uniform norm is also a normed vector space.

The Stone-Weierstrass Theorem

To state the next theorem, we will need the following definitions. If \( f \) and \( g \) are two functions on \( S \), their product \( fg \) is defined as

\[
fg(x) = f(x)g(x) \quad \text{for all } x \in S.
\]

A subspace \( B \) of a normed vector space such that if \( f, g \in B \), then \( fg \in B \), is called an algebra. The function \( 1 \) on a space \( S \) is defined by

\[
1(x) = 1, \quad \text{for all } x \in S,
\]

and a constant function on \( S \) is a function of the form \( c1 \), where \( c \) is a constant. If \( B \subset C(S) \) is a set such that for all \( x, y \in S \), where \( x \neq y \), there exists a function \( h \in B \) such that \( h(x) \neq h(y) \), then it is said to separate points.

The following result is known as the Stone-Weierstrass theorem and the version presented here is the one given in Yosida [9].

**Theorem 2.4** (The Stone-Weierstrass Theorem). Let \( K \) be a compact set and \( B \) a subset of \( C(K) \) that satisfies the following conditions.

- \( B \) is an algebra.
- \( B \) contains the constant functions.
- If \( f_1, f_2, \ldots \in B \) and \( f \in C(K) \) are functions such that \( f_n \to f \) uniformly, then \( f \in B \).

Then \( B \) is dense in \( C(K) \) if and only if \( B \) separates the points of \( K \).

**Proof.** See section 0.2 in [9].

In the proof of one of the results of this thesis, the following corollary of the Stone-Weierstrass theorem will be used.
Corollary 2.5. Let $K$ be a compact set and $B$ a subset of $C(K)$ that satisfies the following conditions.

- $B$ is an algebra.
- $B$ contains the constant functions.
- $B$ separates the points of $K$.

Then $B$ is dense in $C(K)$.

Proof. Let $B$ be an algebra and suppose that $\alpha, \beta \in \mathbb{R}$ and $f, g \in \overline{B}$. Then there exists sequences of functions $\{f_n\}$ and $\{g_n\}$ in $B$ such that $f_n \to f$ and $g_n \to g$ uniformly. Since addition and multiplication are continuous operations it follows that $\alpha f_n + \beta g_n \to \alpha f + \beta g$, and $f_n g_n \to fg$ uniformly, so that $\alpha f + \beta g \in \overline{B}$ and $fg \in \overline{B}$ and hence $\overline{B}$ is an algebra. If $B$ contains the constant functions, then so does $\overline{B}$, and if $B$ separates the points of $K$, then so does $\overline{B}$. By applying $\overline{B}$ to the Stone-Weierstrass theorem it follows that $B$ is dense in $C(K)$.

2.2 $\sigma$-algebras and Measures

This section presents some basic definitions and theorems about $\sigma$-algebras and measures.

$\sigma$-algebras

First, some basic definitions. An algebra over a set $S$ is a non-empty collection $\mathcal{A}$ of subsets of $S$ that satisfies the following.

- If $E \in \mathcal{A}$, then $E^c \in \mathcal{A}$.
- If $E_1, \ldots, E_n \in \mathcal{A}$, then $\bigcup_{i=1}^n E_i \in \mathcal{A}$.

A $\sigma$-algebra over $S$ is an algebra $\mathcal{F}$ such that if $E_1, E_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^\infty E_i \in \mathcal{F}$. If $\mathcal{F}$ is a $\sigma$-algebra over $S$, then $(S, \mathcal{F})$ is called a measurable space and the sets in $\mathcal{F}$ are called measurable sets. Let $(S, \mathcal{F})$ be a measurable space. It follows from the definition that if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcap_{i=1}^\infty A_i \in \mathcal{F}$ since

$$\bigcap_{i=1}^\infty A_i = \left( \bigcup_{i=1}^\infty A_i^c \right)^c.$$ 

Moreover, the empty set $\emptyset$ and $S$ always belong to the $\sigma$-algebra since $\emptyset = A \cap A^c$ and $S = A \cup A^c$ for any set $A \subset S$. It can also be shown that the intersection of any family of $\sigma$-algebras is again a $\sigma$-algebra. If $\mathcal{E}$ is a family of subsets of $S$, there exists a smallest $\sigma$-algebra over $S$ that contains $\mathcal{E}$, namely the intersection of all $\sigma$-algebras containing $\mathcal{E}$. It is called the $\sigma$-algebra generated by $\mathcal{E}$ and is denoted by $\sigma(\mathcal{E})$. The Borel $\sigma$-algebra over $S$ is the $\sigma$-algebra generated by the family of all open subsets of $S$ and is denoted by $\mathcal{B}(S)$. 
The Monotone Class Theorem

For the next result, we need two more definitions. A family of subsets $C$ of some space $S$ is called a $\pi$-system if $A, B \in C$ implies that $A \cap B \in C$. A family of subsets $D$ of $S$ is called a $\lambda$-system if it satisfies the following conditions.

- $S \in D$.
- If $A, B \in D$ and $B \subset A$, then $A \setminus B \in D$.
- If $A_1, A_2, \ldots \in D$ and $A_1 \subset A_2 \subset \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in D$.

There exists different formulations of the following theorem. The one stated here is taken from Kallenberg [7].

**Theorem 2.6** (Monotone Class Theorem). If $C$ is a $\pi$-system and $D$ is a $\lambda$-system in some space $S$ such that $C \subset D$, then $\sigma(C) \subset D$.

**Proof.** See chapter 1 in [7].

Measures

A measure on $(S, \mathcal{F})$, or simply on $S$ if the $\sigma$-algebra is understood, is a function $\mu : \mathcal{F} \rightarrow [0, \infty]$ which satisfies the following.

- $\mu(\emptyset) = 0$.
- If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

If $\mu$ is a measure on $(S, \mathcal{F})$, then $(S, \mathcal{F}, \mu)$ is called a measure space. A measure $\mu$ is said to be $\sigma$-finite if there exists $A_1, A_2, \ldots \in \mathcal{F}$ such that $S = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$, for all $i$. A measure defined on the Borel $\sigma$-algebra is called a Borel measure. If a statement about points $x \in S$, for instance convergence, holds for all $x$ except for those in a set $N \subset S$ such that $\mu(N) = 0$, the statement is said to hold almost everywhere (a.e.) or for almost every $x$. A simple but important example of a measure is the Dirac measure. If $(S, \mathcal{F})$ is a measurable space and $x \in S$, the Dirac measure at $x$ is for any $A \in \mathcal{F}$ defined as

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Some basic properties of measures are summarized in the following theorem.

**Theorem 2.7.** If $(S, \mathcal{F}, \mu)$ is a measure space, then the following statements hold.

(a) If $A, B \in \mathcal{F}$ and $A \subset B$, then $\mu(A) \leq \mu(B)$.

(b) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

(c) If $A_1, A_2, \ldots \in \mathcal{F}$ and $A_1 \subset A_2 \subset \ldots$, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mu(A_i)$.

(d) If $A_1, A_2, \ldots \in \mathcal{F}$, $A_1 \supset A_2 \supset \ldots$ and $\mu(A_1) < \infty$, then $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mu(A_i)$.

**Proof.** See section 1.3 in [5].
From the first statement of theorem 2.7 it follows that if \( \mu(E) = 0 \) and \( F \subset E \), then \( \mu(F) = 0 \) if \( F \in \mathcal{F} \). However, it need not be the case that \( F \in \mathcal{F} \).

A measure such that the corresponding \( \sigma \)-algebra contains all subsets of sets of measure zero is called a complete measure and it follows from the next theorem that a measure always can be extended to a complete measure.

**Theorem 2.8.** Let \((S, \mathcal{F}, \mu)\) be a measure space and define \( \mathcal{N} = \{ N \in \mathcal{F} : \mu(N) = 0 \} \). Let

\[
\mathcal{F} = \{ E \cup F : E \in \mathcal{F}, \quad F \subset N \text{ for some } N \in \mathcal{N} \},
\]

and define \( \overline{\mu} : \mathcal{F} \to [0, \infty] \) by

\[
\overline{\mu}(E \cup F) = \mu(E).
\]

Then \( \mathcal{F} \) is a \( \sigma \)-algebra and \( \overline{\mu} \) is the unique extension of \( \mu \) to a complete measure on \( \mathcal{F} \).

**Proof.** See section 1.3 in [5]. \( \square \)

In theorem 2.8, \( \overline{\mu} \) is called the completion of \( \mu \) and \( \mathcal{F} \) is called the completion of \( \mathcal{F} \) with respect to \( \mu \). The next theorem shows how measures can be constructed. For this, two more definitions are needed. If \( \mathcal{A} \) is an algebra, then a function \( \mu_0 : \mathcal{A} \to [0, \infty] \) is called a premeasure if the following holds.

- \( \mu_0(\emptyset) = 0 \).
- If \( A_1, A_2, \ldots \in \mathcal{A} \) are disjoint and \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \), then \( \mu_0(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu_0(A_i) \).

In analogy to measures, a premeasure \( \mu_0 \) is said to be \( \sigma \)-finite if there exists \( A_1, A_2, \ldots \in \mathcal{A} \) such that \( S = \bigcup_{i=1}^{\infty} A_i \) and \( \mu(A_i) < \infty \), for all \( i \).

**Theorem 2.9.** Let \( \mathcal{A} \) be an algebra over \( S \) and \( \mu_0 \) a premeasure on \( \mathcal{A} \). For each \( E \in S \) let

\[
\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu_0(A_i) : A_1, A_2, \ldots \in \mathcal{A}, \quad E \subset \bigcup_{i=1}^{\infty} A_i \right\},
\]

and define

\[
\mu(E) = \mu^*(E), \quad \text{for all } E \in \sigma(\mathcal{A}).
\]

Then \( \mu \) is a measure on \( \sigma(\mathcal{A}) \) and its restriction to \( \mathcal{A} \) is \( \mu_0 \). Moreover, if \( \mu_0 \) is \( \sigma \)-finite, then \( \mu \) is the unique extension of \( \mu_0 \) to a measure on \( \sigma(\mathcal{A}) \).

**Proof.** See section 1.4 in [5]. \( \square \)

**The Lebesgue Measure on \( \mathbb{R}^d \)**

One of the most important measures in applications is the Lebesgue measure on the euclidean space. Since it is used in this paper, the construction is given, but without any justification of the results. The complete theory can be found in for example Folland [5]. First, the construction of the Lebesgue measure on \( \mathbb{R} \) is given and then this is extended to the \( d \)-dimensional case. To begin,
2.3. Integration of Real Valued Functions

define an h-interval as a subset of \( \mathbb{R} \) of the form \((a, b), (a, \infty)\) or \(\emptyset\), where \(-\infty \leq a < b < \infty\). Let \( \mathcal{A} \) denote the set of all finite disjoint unions of h-intervals. If \((a_1, b_1), \ldots, (a_n, b_n)\) are disjoint h-intervals, define

\[
\mu_0 \left( \bigcup_{i=1}^n (a_i, b_i) \right) = \sum_{i=1}^n (b_i - a_i),
\]

and let \( \mu_0(\emptyset) = 0 \). Then it can be shown that \( \mu_0 \) defines a \( \sigma \)-finite premeasure on \( \mathcal{A} \). By theorem 2.9 there exists a unique extension to a measure \( \mu \) on \( \sigma(\mathcal{A}) \). Moreover, it can be shown that \( \sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}) \), so \( \mu \) is the unique Borel measure on \( \mathbb{R} \) such that \( \mu((a, b)) = b - a \) for all h-intervals \((a, b)\). The Lebesgue measure \( \lambda \) is the completion of \( \mu \) and is defined on \( \mathcal{L} \), the completion of \( \mathcal{B}(\mathbb{R}) \) with respect to \( \mu \). The sets in \( \mathcal{L} \) are called Lebesgue measurable sets. Moving on to the \( d \)-dimensional case, we define a rectangle in \( \mathbb{R}^d \) as a set of the form

\[
A_1 \times \ldots \times A_d, \quad \text{where } A_1, \ldots, A_d \in \mathcal{B}(\mathbb{R}).
\]

Let \( \mathcal{A}^d \) denote the collection of finite disjoint unions of \( d \)-dimensional rectangles. If \( E \in \mathcal{A}^d \) is the finite disjoint union of the rectangles \( A_1^1 \times \ldots \times A_d^1, \ldots, A_1^n \times \ldots \times A_d^n \), let

\[
\pi_0(E) = \sum_{i=1}^n \lambda(A_1^i) \cdots \lambda(A_d^i),
\]

where \( \lambda \) is the Lebesgue measure on \( \mathbb{R} \). Then it can be shown that \( \pi_0 \) defines a \( \sigma \)-finite premeasure on \( \mathcal{A}^d \). As in the construction of the one-dimensional Lebesgue measure, there exists a unique extension of \( \pi_0 \) to a measure \( \mu^d \) on \( \sigma(\mathcal{A}^d) \) and it can be shown that \( \sigma(\mathcal{A}^d) = \mathcal{B}(\mathbb{R}^d) \). This measure, constructed as in theorem 2.9, is the unique measure on \( \mathcal{B}(\mathbb{R}^d) \) such that

\[
\mu^d(A_1 \times \ldots \times A_d) = \lambda(A_1) \cdots \lambda(A_d) \quad \text{for all } A_1, \ldots, A_d \in \mathcal{B}(\mathbb{R}).
\]

The \( d \)-dimensional Lebesgue measure \( \lambda^d \) is the completion of \( \mu^d \) and is defined on \( \mathcal{L}^d \), the completion of \( \mathcal{B}(\mathbb{R}^d) \) with respect to \( \mu^d \).

2.3 Integration of Real Valued Functions

This section defines integration of a real valued function with respect to a measure. For any set \( S \), denote by \( 2^S \) the family of all subsets of \( S \). Let \( S \) and \( X \) be two arbitrary sets. For any mapping \( f : S \to X \), the mapping \( f^{-1} : 2^X \to 2^S \) is defined by

\[
f^{-1}(E) = \{ s \in S : f(s) \in E \}.
\]

If \( (S, \mathcal{F}) \) and \( (X, \mathcal{X}) \) are measurable spaces, a function \( f : S \to X \) is said to be \( (\mathcal{F}, \mathcal{X}) \)-measurable if

\[
f^{-1}(E) \in \mathcal{F} \quad \text{for all } E \in \mathcal{X}.
\]

If \( \mathcal{F} \) and \( \mathcal{X} \) are understood, then \( f \) is called just measurable and if nothing else is said, it is assumed that the \( \sigma \)-algebras are the Borel \( \sigma \)-algebras. If \( f \) and \( g \) are measurable functions, then \( f + g \) and \( fg \) are measurable as well. It can also be shown that if \( f_1, f_2, \ldots \) is a sequence of measurable functions such that
10 Chapter 2. Preliminaries

If \( f_n \to f \) pointwise, then \( f \) is measurable. For a measurable space \((S, \mathcal{F})\), the indicator function of a set \( E \in \mathcal{F} \) is defined as

\[
1_E(x) = \begin{cases} 
1 & \text{if } x \in E \\
0 & \text{otherwise}
\end{cases}
\]

A simple function on \( S \) is a function \( \phi \) that can be written in the form

\[
\phi = \sum_{i=1}^{n} a_i 1_{E_i}, \quad a_1, \ldots, a_n \in \mathbb{C}, \quad E_1, \ldots, E_n \in \mathcal{F}.
\]

It can be shown that all simple functions are measurable. Let \((S, \mathcal{F}, \mu)\) be an arbitrary measure space and define

\[
L^+(S) = \{ f : S \to [0, \infty] : f \text{ is measurable} \}.
\]

If \( \phi = \sum_{i=1}^{n} a_i 1_{E_i} \) is a simple function in \( L^+ \), the integral of \( \phi \) with respect to the measure \( \mu \) is defined as

\[
\int \phi \, d\mu = \sum_{i=1}^{n} a_i \mu(E_i).
\]

This definition is extended to all functions in \( L^+(S) \) as follows.

\[
\int f \, d\mu = \sup \left\{ \int \phi \, d\mu : 0 \leq \phi \leq f, \quad \phi \text{ simple} \right\}.
\]

**Theorem 2.10** (Monotone Convergence Theorem). Let \( f, f_1, f_2, \ldots \) be functions in \( L^+ \) such that \( f_1 \leq f_2 \leq \ldots \) and \( f_n \to f \) pointwise. Then

\[
\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu.
\]

**Proof.** See section 2.2 in [5]. \( \square \)

The functions \( f^+(x) = \max\{f(x), 0\} \) and \( f^-(x) = \min\{-f(x), 0\} \) are called the positive and negative parts of \( f \), respectively. It holds that \( f = f^+ - f^- \) and it can be shown that if \( f \) is measurable, so is \( f^+ \) and \( f^- \). The integral of a real valued measurable function \( f \) is defined as

\[
\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.
\]

If both terms on the right hand side are finite, then \( f \) is said to be integrable. Integration over a subset \( E \) of \( S \) is defined as

\[
\int_E f \, d\mu = \int f 1_E \, d\mu.
\]

We now introduce a class of functions that will play an important role in this thesis. Let \((S, \mathcal{F}, \mu)\) be a measure space and for any real valued function \( f \) on \( S \) let \( \|f\|_{L^1} = \int |f| \, d\mu \). Define

\[
L^1(S, \mathcal{F}, \mu) = \{ f : S \to \mathbb{R} : f \text{ is measurable and } \|f\|_{L^1} < \infty \}.
\]
This space is often abbreviated $L^1(S)$ when the $\sigma$-algebra and the measure are understood. It can be shown that $(L^1(S), \| \cdot \|_{L^1})$ is a normed vector space. Two functions $f$ and $g$ in $L^1(S)$ are defined to be equal if $f(x) = g(x)$ for almost every $x \in S$. The following theorem is another important result about limits of integrals.

**Theorem 2.11** (Dominated Convergence Theorem). Let $f_1, f_2, \ldots$ be a sequence in $L^1(S)$. Suppose that $f_n \to f$ a.e. and that there exists $g \in L^1(S)$ such that $|f_n| \leq g$ a.e. for all $n$. Then $f \in L^1(S)$ and

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu.$$  

**Proof.** See section 2.3 in [5].

The next theorem is a special case of the Fubini theorem. The reference in the proof is to the more general case outlined in [5].

**Theorem 2.12** (Fubini). If $f \in L^1(\mathbb{R}^{d+e})$, then

$$\int_{\mathbb{R}^{d+e}} f(x, y) d\lambda^{d+e}(x, y) = \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^e} f(x, y) d\lambda^e(y) \right) d\lambda^d(x)$$

$$= \int_{\mathbb{R}^e} \left( \int_{\mathbb{R}^d} f(x, y) d\lambda^d(x) \right) d\lambda^e(y).$$

**Proof.** See section 2.5 in [5].

For the remainder of this text we assume that if the measure in the integral is not written out explicitly, then the integration is carried out with respect to the Lebesgue measure, that is,

$$\int f(x) \, dx = \int f(x) \, d\lambda(x).$$

**Theorem 2.13** (Hölder’s Inequality). Let $f$ and $g$ be measurable functions on $(S, \mathcal{F}, \mu)$. If $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, then

$$\int |fg| \, d\mu \leq \left( \int |f|^p \, d\mu \right)^{1/p} \left( \int |g|^q \, d\mu \right)^{1/q}.$$  

**Proof.** See section 6.1 in [5].

Using the following result, it is sometimes possible to write an integral with respect to an arbitrary measure as an integral with respect to the Lebesgue measure.

**Theorem 2.14.** If $f$ is a measurable function on $(S, \mathcal{F}, \mu)$ and $0 < p < \infty$, then

$$\int |f|^p \, d\mu = p \int_0^\infty t^{p-1} \mu\{x \in S : |f(x)| > t\} \, dt.$$  

**Proof.** See section 6.4 in [5].
2.4 Random Elements

This section gives a basic introduction to how measure theory is used in probability theory, based on Billingsley [1]. First, a number of definitions is given. A probability measure on a measurable space $S$, is a measure $\mu$ such that $\mu(S) = 1$. A probability space is a measure space $(S, \mathcal{F}, \mu)$ where $\mu$ is a probability measure. An event space is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where the elements of $\Omega$ and $\mathcal{F}$ are called outcomes and events, respectively. The probability measure $\mathbb{P}$ assigns probabilities to events. A random element is a measurable mapping $\xi$ from the event space $\Omega$ to a metric space $S$. The distribution of $\xi$ is a Borel probability measure on $S$ denoted by $\mathbb{P} \circ \xi^{-1}$ such that for all $E \in \mathcal{B}(S)$,

$$\mathbb{P} \circ \xi^{-1}(E) = \mathbb{P}(\xi^{-1}(E)) = \mathbb{P}\{\omega \in \Omega : \xi(\omega) \in E\} = \mathbb{P}\{\xi \in E\}.$$  

For any set $E \in \mathcal{B}(S)$ it is said that $\xi \in E$ almost surely (a.s.) if $\mathbb{P}\{\xi \in E\} = 1$ and a random element is called non-random if there exists $\xi_0 \in S$ such that $\xi = \xi_0$ a.s. The term random element is used when $S$ is an arbitrary metric space. In some special cases we will use different terms depending on the nature of $S$.

- If $S = \mathbb{R}$, then $\xi$ will be called a random variable.
- If $S = \mathbb{R}^d$, then $\xi$ will be called a random vector.
- If $S$ is a space of measures, $\xi$ will be called a random measure.

A function $f \in L^+(\mathbb{R}^d)$ such that $\int f(x) \, dx = 1$ is called a probability density function, or simply density function. The next result, which is taken from Billingsley [2], shows how the distribution of a random vector can be defined by a density function.

**Theorem 2.15.** Let $f$ be a probability density function and $\nu$ an arbitrary measure. If the mapping $\mu$ is defined by $\mu(E) = \int_E f \, d\nu$, for all $E \in \mathcal{F}$, then $\mu$ is a probability measure. Moreover, if $g$ is a measurable function which is integrable with respect to $\mu$, then

$$\int g \, d\mu = \int fg \, d\nu.$$

**Proof.** See section 16 in [2].
Lemma 2.16. Using the notation above, if \( f : S \to \mathbb{R} \) is a measurable function, then
\[
\int_{\Omega} f(X(\omega)) \, dP(\omega) = \int_{S} f(x) \, dP \circ X^{-1}(x),
\]
whenever either side is well defined.

Proof. See section 10.1 in [5].

Using this result, the expected value of \( X \) can be written as
\[
\mathbb{E}(X) = \int_{S} x \, dP \circ X^{-1}(x).
\]
Moreover, if \( X \) is defined in terms of a density function \( f \), with respect to a measure \( \nu \), and \( \mathbb{E}(X) < \infty \), then it follows from theorem 2.15 that
\[
\mathbb{E}(X) = \int_{S} xf(x) \, d\nu(x).
\]

An expression for the expected value in the case when the random variable is non-negative is given in the following theorem.

Theorem 2.17. If \( X \) is a non-negative random variable, then
\[
\mathbb{E}(X) = \int_{0}^{\infty} \mathbb{P}\{X > t\} \, dt.
\]

Proof. This is an immediate consequence of theorem 2.14.

Now we introduce a basic type of convergence in probability theory. Let \( \xi, \xi_1, \xi_2, \ldots \) be random elements defined on the probability space \((\Omega, \mathcal{F}, P)\) and taking values in the metric space \((S, d)\). The sequence \(\xi_1, \xi_2, \ldots\) is said to converge in probability to \( \xi \) if for every \( \varepsilon > 0 \),
\[
\lim_{n \to \infty} \mathbb{P}\{d(\xi_n, \xi) > \varepsilon\} = 0.
\]

This is denoted by \( \xi_n \overset{P}{\to} \xi \).

Theorem 2.18 (Chebyshev’s Inequality). Let \( X \) be a random variable defined on \((\Omega, \mathcal{F}, P)\) and let \( p \in (0, \infty) \). If \( \mathbb{E}(|X|^p) < \infty \), then for any \( \varepsilon > 0 \),
\[
\mathbb{P}\{|X| > \varepsilon\} \leq \frac{\mathbb{E}(|X|^p)}{\varepsilon^p}.
\]

Proof. If \( A(\varepsilon) = \{\omega \in \Omega : |X(\omega)| > \varepsilon\} \), then
\[
\mathbb{E}(|X|^p) = \int X(\omega)^p \, dP(\omega) \geq \int_{A(\varepsilon)} X(\omega)^p \, dP(\omega) \geq \varepsilon^p \mathbb{P}(A(\varepsilon)).
\]

Theorem 2.19. If \( X \) is a random variable on \((\Omega, \mathcal{F}, P)\) and \( 1 \leq p < \infty \), then
\[
(\mathbb{E}(|X|))^p \leq \mathbb{E}(|X|^p).
\]
Proof. If $p = 1$ we have equality and for $1 < p < \infty$ it follows from Hölder’s inequality that

$$
\mathbb{E}(X) = \int |X \cdot 1| \, d\mathbb{P} \leq \left( \int |X|^p \, d\mathbb{P} \right)^{1/p} \left( \int |1|^q \, d\mathbb{P} \right)^{1/q} = (\mathbb{E}(|X|^p))^{1/p},
$$

where $q = p/(p-1)$.

\[
\square
\]
Chapter 3

Convergence of Probability Measures

This chapter presents a theory for convergence of sequences of Borel probability measures defined on a metric space \((S,d)\). Section 1 introduces the Prohorov metric \(\rho\) defined on \(\mathcal{P}(S)\), the space of all Borel probability measures on \(S\). Section 2 gives a characterization of the compact subsets of \(\mathcal{P}(S)\). Section 3 shows that convergence in the Prohorov metric is equivalent to weak convergence of probability measures if \(S\) is separable and gives useful tools for verifying this convergence. Finally, section 4 introduces the concept of separating and convergence determining sets which can be used as another tool for verifying weak convergence. The chapter is based on section 3.1-3.4 in Ethier, Kurtz [4].

3.1 Prohorov’s Metric

This section shows how to construct a metric space where the elements are Borel probability measures. For any \(A \in \mathcal{B}(S)\) define

\[
A^\varepsilon = \left\{ x \in S : \inf_{y \in A} d(x,y) < \varepsilon \right\}. \tag{3.1}
\]

Let \(\mathcal{C}\) denote the collection of all closed subsets of \(\mathcal{B}(S)\) and define the function \(\rho : \mathcal{P}(S) \times \mathcal{P}(S) \to [0,1]\) by

\[
\rho(P, Q) = \inf\{ \varepsilon > 0 : P(F) \leq Q(F^\varepsilon) + \varepsilon \text{ for all } F \in \mathcal{C} \}. \tag{3.2}
\]

Then \((\mathcal{P}(S), \rho)\) is a metric space. To show this the following two lemmas will be used.

Lemma 3.1. Let \(P,Q \in \mathcal{P}(S)\) and \(\alpha, \beta > 0\). If

\[
P(F) \leq Q(F^\alpha) + \beta \quad \text{for all } F \in \mathcal{C}, \tag{3.3}
\]

then

\[
Q(F) \leq P(F^\alpha) + \beta \quad \text{for all } F \in \mathcal{C}.
\]
Proof. Let \( F_1 \in \mathcal{C} \) be arbitrary and set \( F_2 = (F_1^c)^c \), where \( c \) denotes the complement. Then \( F_2 \in \mathcal{C} \) since \( F_1^c \) is an open set and it also holds that \( F_1 \subset (F_2^c)^c \), so
\[
P(F_1^c) = 1 - P(F_2) \geq 1 - Q(F_2^c) - \beta = Q((F_2^c)^c) - \beta \geq Q(F_1) - \beta.
\]
The first inequality follows from (3.3) and the second is true because \( F_1 \subset (F_2^c)^c \).

Lemma 3.2. If \( P, Q \in \mathcal{P}(S) \) and \( P(F) = Q(F) \) for all \( F \in \mathcal{C} \), then \( P = Q \).

Proof. See section 1.1 in [1].

Theorem 3.3. \((\mathcal{P}(S), \rho)\) is a metric space.

Proof. It follows from lemma 3.1 that \( \rho(P, Q) = \rho(Q, P) \) for all \( P, Q \in \mathcal{P}(S) \). If \( P = Q \), equation (3.2) is true for all \( \varepsilon > 0 \), which implies that \( \rho(P, Q) = 0 \). Conversely, if \( \rho(P, Q) = 0 \), then
\[
P(F) \leq Q(F^c) + \varepsilon, \quad \text{for all } \varepsilon > 0 \text{ and } F \in \mathcal{C}.
\]
From this it follows by lemma 3.1 that
\[
Q(F) \leq P(F^c) + \varepsilon, \quad \text{for all } \varepsilon > 0 \text{ and } F \in \mathcal{C}.
\]
These two equations together imply that \( P(F) = Q(F) \) for all \( F \in \mathcal{C} \). By lemma 3.2 it follows that \( P = Q \). To prove the triangle inequality, let \( P, Q, R \in \mathcal{P}(S) \) and suppose that \( \delta > \rho(P, Q) \) and \( \varepsilon > \rho(Q, R) \). Then,
\[
P(F) \leq Q(F^{\delta}) + \delta \leq Q(F^{\delta}) + \delta \leq R((F^{\delta})^c) + \delta + \varepsilon \leq R(F^{\delta + \varepsilon}) + \delta + \varepsilon,
\]
for all \( F \in \mathcal{C} \). From this it follows that \( \rho(P, R) \leq \delta + \varepsilon \). Since this is true for all \( \delta > \rho(P, Q) \) and \( \varepsilon > \rho(Q, R) \),
\[
\rho(P, R) \leq \rho(P, Q) + \rho(Q, R), \quad \text{for all } P, Q, R \in \mathcal{P}(S),
\]
so \( \rho \) satisfies the properties of a metric and hence \((\mathcal{P}(S), \rho)\) is a metric space.

The following theorem gives another formula for this metric.

Theorem 3.4. Let \( P, Q \in \mathcal{P}(S) \) and define
\[
\mathcal{M}(P, Q) = \{ \mu \in \mathcal{P}(S \times S) : \mu(A \times S) = P(A), \mu(S \times A) = Q(A), A \in \mathcal{B}(S) \}.
\]
Then
\[
\rho(P, Q) = \inf_{\mu \in \mathcal{M}(P, Q)} \inf \{ \varepsilon > 0 : \mu\{(x, y) : d(x, y) \geq \varepsilon \} \leq \varepsilon \}.
\]

Proof. Suppose that \( \varepsilon_0 \in \{ \varepsilon > 0 : \mu\{(x, y) : d(x, y) \geq \varepsilon \} \leq \varepsilon \} \) and \( \mu \in \mathcal{M}(P, Q) \). Then for any \( F \in \mathcal{C} \),
\[
P(F) = \mu(F \times S) = \mu((F \times S) \cap \{(x, y) : d(x, y) < \varepsilon_0 \}) + \mu((F \times S) \cap \{(x, y) : d(x, y) \geq \varepsilon_0 \}) \leq \mu(F \times F^{\varepsilon_0}) + \mu\{(x, y) : d(x, y) \geq \varepsilon_0 \} \leq \mu(S) \times F^{\varepsilon_0} + \varepsilon_0 = Q(F^{\varepsilon_0}) + \varepsilon_0.
\]
so \( \varepsilon_0 \in \{ \varepsilon > 0 : P(F) \leq Q(F^c) + \varepsilon, F \in \mathcal{C} \} \) and hence \( \rho(P, Q) \leq \varepsilon_0 \). From this it follows that

\[
\rho(P, Q) \leq \inf_{\mu \in \mathcal{M}(P, Q)} \inf \{ \varepsilon > 0 : \mu \{ (x, y) : d(x, y) \geq \varepsilon \} \leq \varepsilon \}.
\]

For the proof of the opposite inequality, see section 3.1 in [4].

As an application of this theorem, a relation between convergence in the Prohorov metric and convergence in probability of random elements can be proved.

**Corollary 3.5.** Let \((S, d)\) be separable and suppose that \(\xi, \xi_1, \xi_2, \ldots\) are \(S\)-valued random elements defined on the same probability space \((\Omega, \mathcal{F}, P)\), with distributions \(P, P_1, P_2, \ldots\), respectively. If \(\xi_n \xrightarrow{P} \xi\), then \(\lim_{n \to \infty} \rho(P_n, P) = 0\).

**Proof.** If \(\xi_n \xrightarrow{P} \xi\), then for every \(\varepsilon > 0\),

\[
\lim_{n \to \infty} P \{ d(\xi_n, \xi) > \varepsilon \} = 0.
\]

If \(\mu_n \in \mathcal{P}(S \times S)\) denotes the joint distribution of \((\xi_n, \xi)\) for all \(n \in \mathbb{N}\), then

\[
\lim_{n \to \infty} \mu_n \{ (x, y) : d(x, y) > \varepsilon \} = 0.
\]

It follows from theorem 3.4 that

\[
\rho(P_n, P) \leq \inf \{ \varepsilon > 0 : \mu_n \{ (x, y) : d(x, y) \geq \varepsilon \} \leq \varepsilon \}.
\]

Since the right hand side tends to zero as \(n \to \infty\), the result follows.

The last theorem of this section will be needed for later purposes.

**Theorem 3.6.** If \((S, d)\) is separable, then \((\mathcal{P}(S), \rho)\) is separable. If \((S, d)\) is complete and separable, then \((\mathcal{P}(S), \rho)\) is complete and separable.

**Proof.** See section 3.1 in [4].

### 3.2 Prohorov’s Theorem

In this section, a characterization of the compact subsets of the metric space \((\mathcal{P}(S), \rho)\) is given through the notion of tightness. The key result is Prohorov’s theorem. Tightness is defined as follows. A probability measure \(P \in \mathcal{P}(S)\) is said to be **tight** if for every \(\varepsilon > 0\) there exists a compact set \(K\) such that \(P(K) \geq 1 - \varepsilon\). A family of probability measures \(\mathcal{M} \subset \mathcal{P}(S)\) is said to be **tight** if for every \(\varepsilon > 0\) there exists a compact set \(K\) such that

\[
\inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon.
\]

To prove Prohorov’s theorem we will use the following lemma.

**Lemma 3.7.** If \((S, d)\) is complete and separable, then each \(P \in \mathcal{P}(S)\) is tight.
Proof. Let \( \varepsilon > 0 \) be arbitrary. Since \( S \) is separable, it contains a dense countable subset \( \{x_1, x_2, \ldots \} \). For each \( n = 1, 2, \ldots \), it is possible to choose a positive integer \( N_n \) such that

\[
P \left( \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \right) \geq 1 - \frac{\varepsilon}{2^n}.
\]

Let \( K \) be the closure of

\[
\bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}).
\]

It follows from theorem 2.1 that \( K \) is complete since \( (S,d) \) is complete and \( K \subset S \) is closed. Let \( \delta > 0 \) be arbitrary and choose a positive integer \( \alpha \) such that \( \frac{1}{\alpha} < \delta \). Then

\[
\bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \subset \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{\alpha}) \subset \bigcup_{i=1}^{N_n} B(x_i, \delta),
\]

so it follows that \( K \) is totally bounded. Since \( K \) is complete and totally bounded, it follows from theorem 2.2 that it is compact and

\[
P(K) \geq P \left( \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \right)
\]

\[
= 1 - P \left( \bigcup_{n=1}^{\infty} \left( \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \right) \right)
\]

\[
\geq 1 - \sum_{n=1}^{\infty} P \left( \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \right)
\]

\[
= 1 - \sum_{n=1}^{\infty} \left( 1 - P \left( \bigcup_{i=1}^{N_n} B(x_i, \frac{1}{n}) \right) \right)
\]

\[
\geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon.
\]

\( \square \)

Theorem 3.8 (Prohorov’s Theorem). If \( (S,d) \) is complete and separable, then the following statements are equivalent.

(a) \( \mathcal{M} \) is tight.

(b) For every \( \varepsilon > 0 \) there exists a compact set \( K \) such that

\[
\inf_{P \in \mathcal{M}} P(K^\varepsilon) \geq 1 - \varepsilon,
\]

where \( K^\varepsilon \) is defined as in (3.1).

(c) \( \mathcal{M} \) is relatively compact.
3.2. Prohorov’s Theorem

Proof. (a) implies (b): Since $K \subset K^\varepsilon$ and $\mathcal{M}$ is tight, it follows that for every $\varepsilon > 0$ there exists a compact set $K$ such that

$$\inf_{P \in \mathcal{M}} P(K^\varepsilon) \geq \inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon.$$ 

(b) implies (c): We will first show that $\mathcal{M}$ is totally bounded, that is, for every $\delta > 0$, there exists a finite set $\mathcal{N}$ such that

$$\mathcal{M} \subset \bigcup_{P \in \mathcal{N}} B(P, \delta) = \bigcup_{P \in \mathcal{N}} \{Q \in \mathcal{M} : \rho(P, Q) < \delta\}.$$

Let $\delta > 0$ be arbitrary and for $\varepsilon \in (0, \delta/2)$ choose a compact set $K$ such that (3.4) holds. Since $K$ is compact, it is totally bounded so there exists a finite set $\{x_1, \ldots, x_n\} \subset K$ such that $K^\varepsilon \subset \bigcup_{i=1}^n B(x_i, \varepsilon)$ and from this it follows that $K^\varepsilon \subset \bigcup_{i=1}^n B(x_i, 2\varepsilon)$. Let $x_0 \in S$ be arbitrary and $m \geq n/\varepsilon$. Define

$$\mathcal{N} = \left\{ \frac{1}{m} \sum_{i=0}^n k_i \delta_{x_i} : k_0, \ldots, k_n \in \{0, \ldots, m\}, \sum_{i=0}^n k_i = m \right\},$$

where $\delta_{x_i}$ denotes the Dirac measure. Let $E_1 = B(x_1, 2\varepsilon)$ and

$$E_i = B(x_i, 2\varepsilon) \setminus \bigcup_{k=1}^{i-1} B(x_k, 2\varepsilon), \text{ for } i = 2, \ldots, n.$$

For arbitrary $Q \in \mathcal{M}$ define $k_i = \lfloor mQ(E_i) \rfloor$ for $i = 1, \ldots, m$, where $\lfloor \cdot \rfloor$ denotes the integer part, and let $k_0 = 1 - \sum_{i=1}^m k_i$. If we set $P = \frac{1}{m} \sum_{i=0}^n k_i \delta_{x_i}$, then $P \in \mathcal{N}$ and for any $F \in \mathcal{G}$,

$$Q(F) = Q(F \cap K^\varepsilon) + Q(F \cap (K^\varepsilon)^c)$$

$$\leq Q \left( \bigcup_{i=1}^n (F \cap E_i) \right) + Q((K^\varepsilon)^c)$$

$$\leq Q \left( \bigcup_{F \cap E_i \neq \emptyset} E_i \right) + \varepsilon$$

$$= \sum_{F \cap E_i \neq \emptyset} Q(E_i) + \varepsilon$$

$$\leq \sum_{F \cap E_i \neq \emptyset} \left( \lfloor mQ(E_i) \rfloor + 1 \right) + \varepsilon$$

$$\leq \sum_{F \cap E_i \neq \emptyset} \frac{k_i}{m} + 2\varepsilon.$$

It follows from the construction of the sets $E_1, \ldots, E_n$ that

$$\{x_i : i \in \{1, \ldots, n\}, \ F \cap E_i \neq \emptyset\} \subset F^{2\varepsilon}.$$

Using this,

$$\sum_{F \cap E_i \neq \emptyset} \frac{k_i}{m} = \frac{1}{m} \sum_{F \cap E_i \neq \emptyset} k_i \delta_{x_i}(F^{2\varepsilon}) \leq \frac{1}{m} \sum_{i=0}^n k_i \delta_{x_i}(F^{2\varepsilon}) = P(F^{2\varepsilon}),$$

$$\sum_{F \cap E_i \neq \emptyset} \frac{k_i}{m} \leq \frac{1}{m} \sum_{F \cap E_i \neq \emptyset} k_i \delta_{x_i}(F^{2\varepsilon}) = P(F^{2\varepsilon}).$$
so $Q(F) \leq P(F^{2\varepsilon}) + 2\varepsilon$ for all $F \in \mathcal{C}$ and this proves that $\rho(P, Q) \leq 2\varepsilon < \delta$.

It can be concluded that $\mathcal{M}$ is totally bounded which implies that the closure of $\mathcal{M}$ is totally bounded. By theorem 3.6, $(\mathcal{P}(S), \rho)$ is complete since $(S, d)$ is separable and complete, so it follows that the closure of $\mathcal{M}$ is complete. Putting all this together, we have that the closure of $\mathcal{M}$ is complete and totally bounded, which implies that $\mathcal{M}$ is relatively compact.

(c) implies (a): Let $\varepsilon > 0$ and $Q \in \mathcal{M}$ be arbitrary. Since $\mathcal{M}$ is relatively compact, it is totally bounded, so for any $n = 1, 2, \ldots$, we can choose a finite set $\mathcal{K}_n$ such that there exists $P_n \in \mathcal{K}_n$ satisfying $\rho(P_n, Q) < \varepsilon/2^{n+1}$. Moreover, it follows from lemma 3.7 that there exists compact sets $K_1, K_2, \ldots$ such that

$$P_n(K_n) \geq 1 - \frac{\varepsilon}{2^{n+1}}, \quad \text{for all } n = 1, 2, \ldots$$

This equation and that $\rho(P_n, Q) < \varepsilon/2^{n+1}$ gives

$$Q\left(\bigcap_{n=1}^{\infty} K_n^{\varepsilon/2^{n+1}}\right) \geq P_n(K_n) - \frac{\varepsilon}{2^{n+1}} \geq 1 - \frac{\varepsilon}{2^n}, \quad \text{for all } n = 1, 2, \ldots$$

Let $K$ be the closure of $\bigcap_{n=1}^{\infty} K_n^{\varepsilon/2^{n+1}}$. Arguing as in the proof of lemma 3.7, it follows that $K$ is compact and

$$Q(K) \geq Q\left(\bigcap_{n=1}^{\infty} K_n^{\varepsilon/2^{n+1}}\right) = 1 - Q\left(\bigcup_{n=1}^{\infty} (K_n^{\varepsilon/2^{n+1}})^c\right) \geq 1 - \sum_{n=1}^{\infty} \left(1 - Q\left(K_n^{\varepsilon/2^{n+1}}\right)\right) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon.$$ 

Since $Q \in \mathcal{M}$ was arbitrary, it holds for this $K$ that

$$\inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon,$$

so $\mathcal{M}$ is tight.

The following result states that compactness can be verified through tightness even without the assumption that $(S, d)$ is complete and separable.

**Corollary 3.9.** Let $(S, d)$ be an arbitrary metric space. If $\mathcal{M}$ is tight, then $\mathcal{M}$ is relatively compact.

**Proof.** See section 3.2 in [4].

3.3 Weak Convergence of Probability Measures

This section introduces the concept of weak convergence of probability measures and shows that this is equivalent to convergence in the Prohorov metric if the
metric space \((S, d)\) is separable. To define weak convergence, let \(C_b(S)\) denote the space of all real valued bounded continuous functions defined on \(S\), endowed with the uniform norm \(\|f\|_\infty = \sup_{x \in S} f(x)\). A sequence of probability measures \(P_1, P_2, \ldots \in \mathcal{P}(S)\) is said to converge weakly to \(P \in \mathcal{P}(S)\) if

\[
\lim_{n \to \infty} \int f \, dP_n = \int f \, dP, \quad \text{for all } f \in C_b(S). \tag{3.5}
\]

This is denoted by \(P_n \xrightarrow{w} P\). In the next theorem, several ways of verifying weak convergence is given. The following definition will be used in this result.

A set \(A \subset S\) is said to be a \(P\)-continuity set if \(A \in \mathcal{B}(S)\) and \(P(\partial A) = 0\).

**Theorem 3.10.** Let \((S, d)\) be arbitrary and let \(P, P_1, P_2, \ldots \in \mathcal{P}(S)\). Then, of the following statements, (b) through (f) are all equivalent and all of them are implied by (a). If \((S, d)\) is separable, then all 6 statements are equivalent.

(a) \(\lim_{n \to \infty} \rho(P_n, P) = 0\).

(b) \(P_n \xrightarrow{w} P\).

(c) \(\lim_{n \to \infty} \int f \, dP_n = \int f \, dP\) for all uniformly continuous \(f \in C_b(S)\).

(d) \(\lim \sup_{n \to \infty} P_n(F) \leq P(F)\) for all closed sets \(F \subset S\).

(e) \(\lim \inf_{n \to \infty} P_n(G) \geq P(G)\) for all open sets \(G \subset S\).

(f) \(\lim_{n \to \infty} P_n(A) = P(A)\) for all \(P\)-continuity sets \(A \subset S\).

**Proof.** (a) implies (b): Throughout this proof, we let \(\|f\| = \|f\|_\infty\). Define \(\varepsilon_n = \rho(P_n, P) + 1/n\) and let \(f \geq 0\) be a function in \(C_b(S)\). Using theorem 2.14 and that \(f\) is bounded gives

\[
\int f \, dP_n = \int_0^{\|f\|} P_n\{x : f(x) > t\} \, dt \\
\leq \int_0^{\|f\|} \left( P\{x : f(x) \geq t\} + \varepsilon_n \right) \, dt \\
= \int_0^{\|f\|} P\{x : f(x) \geq t\} \, dt + \varepsilon_n \|f\|, \quad \text{for all } n = 1, 2, \ldots
\]

Since the left hand side is a bounded sequence of points and the right hand side is a decreasing sequence of points it holds that

\[
\lim_{n \to \infty} \sup \int f \, dP_n \leq \lim_{n \to \infty} \int_0^{\|f\|} P\{x : f(x) \geq t\} \, dt \\
= \int_0^{\|f\|} P\{x : f(x) \geq t\} \, dt \\
= \int f \, dP, \tag{3.6}
\]

where the first equality follows from the dominated convergence theorem. Now, let \(f \in C_b(S)\) be arbitrary. Using (3.6) and that \(\|f\| + f\) and \(\|f\| - f\) are
non-negative functions it follows that
\[
\limsup_{n \to \infty} \int (\|f\| + f) \, dP_n \leq \int (\|f\| + f) \, dP
\]
\[
\limsup_{n \to \infty} \int (\|f\| - f) \, dP_n \leq \int (\|f\| - f) \, dP
\]
Note that \( \int \|f\| \, dP_n = \|f\| P_n(S) = \|f\| \cdot 1 \), for all \( n \), so these two inequalities can be rewritten as
\[
\|f\| + \limsup_{n \to \infty} \int f \, dP_n \leq \|f\| + \int f \, dP
\]
\[
\|f\| - \liminf_{n \to \infty} \int f \, dP_n \leq \|f\| - \int f \, dP
\]
Since \( \|f\| \) is finite, it can be subtracted from both sides of the inequalities, which implies that
\[
\limsup_{n \to \infty} \int f \, dP_n \leq \int f \, dP \leq \liminf_{n \to \infty} \int f \, dP_n.
\]
From this it follows that \( \lim_{n \to \infty} \int f \, dP_n = \int f \, dP \).

(b) implies (c): If \( P_n \xrightarrow{w} P \), then by definition, \( \lim_{n \to \infty} \int f \, dP_n = \int f \, dP \) for all \( f \in C_b(S) \). In particular, this holds for all uniformly continuous \( f \in C_b(S) \).

(c) implies (d): Let \( F \subset S \) be closed and for every \( \varepsilon > 0 \) define
\[
f_\varepsilon(x) = \max \left\{ 1 - \frac{d(x,F)}{\varepsilon}, 0 \right\}, \quad \text{for all } x \in S,
\]
where \( d(x,F) = \inf_{y \in F} d(x,y) \). Then \( f_\varepsilon \) is uniformly continuous and belongs to \( C_b(S) \). Since \( f_\varepsilon(x) = 1 \) for all \( x \in F \), it holds that \( 1_F(x) \leq f_\varepsilon(x) \) for all \( x \in S \) and hence
\[
P_n(F) = \int 1_F \, dP_n \leq \int f_\varepsilon \, dP_n.
\]
The sequence \( \{P_n(F)\} \) is bounded, so we have that
\[
\limsup_{n \to \infty} P_n(F) \leq \limsup_{n \to \infty} \int f_\varepsilon \, dP_n = \int f_\varepsilon \, dP,
\]
where the equality follows from (c). Since this holds for all \( \varepsilon > 0 \),
\[
\limsup_{n \to \infty} P_n(F) \leq \lim_{\varepsilon \to 0} \int f_\varepsilon \, dP = \int 1_F \, dP = P(F),
\]
because \( f_\varepsilon \) converges uniformly to \( 1_F \) as \( \varepsilon \to 0 \).

(d) implies (e): If \( G \subset S \) is open, then
\[
\liminf_{n \to \infty} P_n(G) = \liminf_{n \to \infty} (1 - P_n(G^c)) = 1 - \limsup_{n \to \infty} P_n(G^c) \geq 1 - P(G^c) = P(G),
\]
where the inequality follows from (d).

(e) implies (f): If \( A \subset S \) is a \( P \)-continuity set, then

\[
\limsup_{n \to \infty} P_n(A) \leq \limsup_{n \to \infty} P_n(\overline{A})
= \limsup_{n \to \infty} (1 - P_n((\overline{A})^c))
\leq 1 - P((\overline{A})^c)
= P(A)
\]

and

\[
\liminf_{n \to \infty} P_n(A) \geq \liminf_{n \to \infty} P_n(A^c) \geq P(A^c) = P(A).
\]

From this it can be concluded that \( \lim_{n \to \infty} P_n(A) = P(A) \).

(f) implies (b): Let \( f \geq 0 \) be a function in \( C_b(S) \). Then \( P\{x : f(x) = t\} = 0 \) for all but at most countably many \( t \in (0,\|f\|) \). Since

\[
\partial\{x : f(x) \geq t\} \subset \{x : f(x) = t\},
\]

it follows that \( \{x : f(x) \geq t\} \) is a \( P \)-continuity set for almost every \( t \in (0,\|f\|) \).

Using the dominated convergence theorem and (f) gives

\[
\lim_{n \to \infty} \int f \, dP_n = \lim_{n \to \infty} \int_0^{\|f\|} P_n\{x : f(x) \geq t\} \, dt
= \int_0^{\|f\|} P\{x : f(x) \geq t\} \, dt
= \int f \, dP,
\]

for all non-negative functions in \( C_b(S) \). By the linearity of the integral, this holds for all \( f \in C_b(S) \), and hence \( P_n \xrightarrow{w} P \).

(e) implies (a): This part of the proof is under the assumption that \((S,d)\) is separable. Let \( \varepsilon > 0 \) be arbitrary. Since \( S \) is separable there exists \( E_1, E_2, \ldots \in \mathcal{B}(S) \) such that \( S = \bigcup_{i=1}^{\infty} E_i \), \( E_i \cap E_j = \emptyset \) for all \( i \neq j \) and

\[
\sup\{d(x,y) : x, y \in E_i\} < \frac{\varepsilon}{2}, \quad \text{for all } i \in \mathbb{N}.
\] (3.7)

Let \( N \) be the smallest positive integer such that

\[
P\left(\bigcup_{i=1}^{N} E_i\right) > 1 - \frac{\varepsilon}{2}.
\]

Define \( \mathcal{G} \) as the collection of open sets of the form \( \bigcup_{i \in I} E_i^{\varepsilon/2} \) where \( I \subset \{1,\ldots,N\} \) and let \( G_1, \ldots, G_k \) be an enumeration of the sets in \( \mathcal{G} \). It follows from
condition (e) that for each $i = 1, ..., k$ there exists a positive integer $n_i$ such that $P(G_i) \leq P_n(G_i) + \varepsilon/2$, for all $n \geq n_i$. If $n_0 = \max\{n_1, ..., n_k\}$, then
\[
P(G) \leq P_n(G) + \frac{\varepsilon}{2}, \quad \text{for all } G \in \mathcal{G} \text{ and } n \geq n_0.
\]
Let $F \in \mathcal{G}$ be arbitrary and define
\[
F_0 = \bigcup\{E_i : 1 \leq i \leq N, E_i \cap F \neq \emptyset\}.
\]
Since $F \in \mathcal{G}$, it holds that
\[
P(F) = P\left(F \cap \left(\bigcup_{i=1}^{N} E_i\right)\right) + P\left(F \cap \left(\bigcup_{i=1}^{N} E_i^c\right)\right)
\]
\[
\leq P(F_0) + P\left(\left(\bigcup_{i=1}^{N} E_i\right)^c\right)
\]
\[
\leq P_n\left(F_0^{\varepsilon/2}\right) + \frac{\varepsilon}{2}
\]
\[
\leq P_n\left(F_0^{\varepsilon/2}\right) + \varepsilon
\]
\[
\leq P_n(F^{\varepsilon/2}) + \varepsilon, \quad \text{for all } n \geq n_0,
\]
where the last inequality is true because equation (3.7) implies that $F_0^{\varepsilon/2} \subset F^\varepsilon$.

It follows that $\rho(P_n, P) \leq \varepsilon$, so for every $\varepsilon > 0$, there exists a positive integer $n_0$ such that $\rho(P_n, P) \leq \varepsilon$ for all $n \geq n_0$, which means that
\[
\lim_{n \to \infty} \rho(P_n, P) = 0.
\]

\[\square\]

3.4 Convergence Determining Sets

To conclude weak convergence of a sequence of Borel probability measures it is not necessary to verify the convergence in (3.5) for all $f \in C_b(S)$. This section gives a characterization of sets $M \subset C_b(S)$ such that if the convergence in (3.5) holds for all $f \in M$, then it also holds for all $f \in C_b(S)$. Let $B(S)$ denote the space of all bounded real valued functions defined on $S$. A sequence of functions $f_1, f_2, ... \in B(S)$ is said to converge **boundedly and pointwise** to $f \in B(S)$ if $f_n$ converges pointwise to $f$ and $\sup_n \|f_n\| < \infty$. This is denoted by
\[
\text{bp-lim}_{n \to \infty} f_n = f. \quad (3.8)
\]
A set $M \subset B(S)$ is called **bp-closed** if whenever $f_1, f_2, ... \in M$, $f \in B(S)$ and (3.8) holds, then $f \in M$. The **bp-closure** of $M$ is the smallest bp-closed set containing $M$. A set $M \in B(S)$ is said to be **bp-dense** in $B(S)$ if the bp-closure of $M$ is equal to $B(S)$. If $M$ is bp-dense in $B(S)$ and $f \in B(S)$ there need not exist a sequence $f_1, f_2, ... \in M$ such that (3.8) holds. Separating and convergence determining sets are defined as follows. A set $M \subset C_b(S)$ is called **separating** if whenever $P, Q \in \mathcal{P}(S)$ and
\[
\int f \, dP = \int f \, dQ, \quad \text{for all } f \in M,
\]
we have \( P = Q \). A set \( M \subset C_b(S) \) is called *convergence determining* if whenever \( P, P_1, P_2, \ldots \in \mathcal{P}(S) \) and
\[
\lim_{n \to \infty} \int f \, dP_n = \int f \, dP, \quad \text{for all } f \in M, \tag{3.9}
\]
we have \( P_n \overset{w}{\to} P \). If a set is convergence determining, then it is also separating. The converse is false in general but the following theorem gives sufficient conditions for when it does hold.

**Lemma 3.11.** Let \( \{P_1, P_2, \ldots\} \subset \mathcal{P}(S) \) be relatively compact, let \( P \in \mathcal{P}(S) \) and suppose that \( M \subset C_b(S) \) is separating. If (3.9) holds, then \( P_n \overset{w}{\to} P \).

**Proof.** See section 3.4 in [4].

Recall from section 2.1 that a set \( M \subset C_b(S) \) is said to *separate points* if for all \( x, y \in S \), where \( x \neq y \), there exists a function \( h \in M \) such that \( h(x) \neq h(y) \). Furthermore, a set \( M \subset C_b(S) \) is said to *strongly separate points* if for all \( x \in S \) and \( \delta > 0 \), there exists a finite set \( \{h_1, \ldots, h_k\} \subset M \) such that
\[
\inf_{y : d(y, x) \geq \delta} \max_{i \in \{1, \ldots, k\}} |h_i(y) - h_i(x)| > 0.
\]

The concluding theorem of this chapter relates these definitions to separating and convergence determining sets. For this theorem, recall the definition of an algebra of functions defined in section 2.1.

**Theorem 3.12.** Let \((S, d)\) be complete and separable. If \( M \subset C_b(S) \) is an algebra, then the following statements hold.

(a) If \( M \) separates points, then \( M \) is separating.

(b) If \( M \) strongly separates points, then \( M \) is convergence determining.

**Proof.** See section 3.4 in [4].
Chapter 4

Convergence of Random Measures

In this chapter, a theory for convergence of sequences of random measures is presented. The results in this section will in particular hold for random probability measures, but the theory is given for more general random measures. Section 1 is devoted to the study of convergence of random elements in a metric space $S$. Then in section 2, this theory is specialized to the case when $S$ is a space of measures and some important uniqueness results are given. Finally, section 3 gives the key theorem for verifying convergence of a sequence of random measures. The chapter is based on Kallenberg [7].

4.1 Convergence in Distribution of Random Elements

In the previous chapter, convergence of probability measures were studied. Sometimes it is more convenient to consider convergence of random elements and in this section the corresponding theory for this is given. A sequence of random elements $\xi_1, \xi_2, \ldots$ is said to converge in distribution to the random element $\xi$ if $P \circ \xi^{-1}_n \xrightarrow{w} P \circ \xi^{-1}$, or equivalently

$$\lim_{n \to \infty} E(f(\xi_n)) = E(f(\xi)),$$

for all $f \in C_b(S)$. This is denoted by $\xi_n \xrightarrow{d} \xi$. The following theorem, which is important for later purposes, gives a relation between convergence in distribution and convergence in probability.

**Theorem 4.1.** If $\xi_1, \xi_2, \ldots$ are random elements in a metric space $(S, d)$ and $\xi$ is a non-random element in $(S, d)$, then $\xi_n \xrightarrow{d} \xi$ if and only if $\xi_n \xrightarrow{p} \xi$.

**Proof.** Let $\mu_n = P \circ \xi_n^{-1}$ for $n \in \mathbb{N}$ and let $\mu = P \circ \xi^{-1}$. By definition, $\xi_n \xrightarrow{d} \xi$ if and only if $\mu_n \xrightarrow{w} \mu$, which by theorem 3.10 is equivalent to that $\mu_n(A) \to \mu(A)$ for all $\mu$-continuity sets $A \subset S$. Since $\xi$ is non-random, $\mu = \delta_{x_0}$ for some $x_0 \in S$, so it is enough to show that $\xi_n \xrightarrow{w} \xi$ if and only if

$$\mu_n(A) \to \delta_{x_0}(A), \quad \text{for all } \delta_{x_0} \text{-continuity sets } A \subset S. \quad (4.1)$$

First assume that $\xi_n \overset{p}{\to} \xi$ and let $A \subset S$ be a $\delta_{x_0}$-continuity set. Then $x_0 \notin \partial A$, so if $x_0 \in A$, there exists $\varepsilon_0 > 0$ such that $B(x_0, \varepsilon_0) \subset A$ and

$$|\mu_n(A) - \delta_{x_0}(A)| = |\mu_n(A) - 1|$$

$$= \mu_n(A^c)$$

$$\leq \mu_n(B(x_0, \varepsilon_0)^c)$$

$$= P\{\xi_n \in B(x_0, \varepsilon_0)^c\}$$

$$= P\{d(\xi_n, x_0) \geq \varepsilon_0\} \to 0 \text{ as } n \to \infty.$$

If $x_0 \notin A$, then since $x_0 \notin \partial A$, there exists $\varepsilon_0 > 0$ such that $B(x_0, \varepsilon_0)^c$ and

$$|\mu_n(A) - \delta_{x_0}(A)| = |\mu_n(A) - 0|$$

$$\leq \mu_n(\overline{A})$$

$$= 1 - \mu_n((\overline{A})^c)$$

$$\leq 1 - \mu_n(B(x_0, \varepsilon_0))$$

$$= \mu_n(B(x_0, \varepsilon_0)^c)$$

$$= P\{d(\xi_n, x_0) \geq \varepsilon_0\} \to 0 \text{ as } n \to \infty.$$

Conversely, assume that (4.1) holds. Then for any $\varepsilon > 0$, the limit

$$\lim_{n \to \infty} P\{d(\xi_n, x_0) \geq \varepsilon\} = \lim_{n \to \infty} \mu_n(B(x_0, \varepsilon)^c) = \delta_{x_0}(B(x_0, \varepsilon)^c) = 0,$$

since $B(x_0, \varepsilon)^c$ is a $\delta_{x_0}$-continuity set for every $\varepsilon > 0$.

The concept of tightness of random elements is defined as follows. A sequence of random elements in a metric space $(S, d)$ is said to be tight if for every $\varepsilon > 0$, there exists a compact set $K \subset S$ such that

$$\liminf_{n \to \infty} P\{\xi_n \in K\} \geq 1 - \varepsilon.$$

In the case when $S$ is separable and complete, we can replace the ‘liminf’ by ‘inf’. The next result is Prohorov’s theorem formulated for sequences of random elements. For this, the following definition is needed. A sequence of random elements $\xi_1, \xi_2, ...$ is said to be relatively compact in distribution if every subsequence has a further subsequence that converges in distribution.

**Theorem 4.2.** If $\xi_1, \xi_2, ...$ are random elements in a complete and separable metric space $(S, d)$, then $\{\xi_n\}$ is tight if and only if $\{\xi_n\}$ is relatively compact in distribution.

**Proof.** By definition, $\{\xi_n\}$ is tight if and only if $\mathcal{M} = \{ P \circ \xi_1, P \circ \xi_2, ... \}$ is tight. According to theorem 3.8, this is equivalent to relative compactness of $\mathcal{M}$ with respect to the Prohorov metric. Since $(S, d)$ is separable, it follows from theorem 3.10 that convergence in the Prohorov metric is equivalent to weak convergence. This means that every sequence in $\mathcal{M}$ has a subsequence that converges weakly and this is equivalent to that $\{\xi_n\}$ is relatively compact in distribution.
Lemma 4.3. Let $S$ and $T$ be complete and separable metric spaces and let $f : S \to T$ be a continuous mapping. If $\{\xi_n\}$ is a sequence of random elements which is tight in $S$, then $\{f(\xi_n)\}$ is tight in $T$.

Proof. For any function $g \in C_b(T)$ it holds that $g \circ f \in C_b(S)$. So if $\xi_n \xrightarrow{d} \xi$, then for any $g \in C_b(T)$,

$$
\lim_{n \to \infty} E(g \circ f(\xi_n)) = E(g \circ f(\xi)),
$$

which is the same as $f(\xi_n) \xrightarrow{d} f(\xi)$. This implies that $\{f(\xi_n)\}$ is relatively compact in distribution if $\{\xi_n\}$ is. By Prohorov’s theorem, relative compactness is equivalent to tightness for complete separable metric spaces, so the result follows from this.

The following theorem is used in the proof of the main theorem for convergence of random measures.

Theorem 4.4. Let $S$ and $T$ be metric spaces and let $\xi, \xi_1, \xi_2, \ldots$ be random elements in $S$ such that $\xi_n \xrightarrow{d} \xi$. Suppose that $f, f_1, f_2, \ldots : S \to T$ are measurable functions and that $C \subset S$ is a measurable set such that $\xi \in C$ a.s. Then $f_n(\xi_n) \xrightarrow{d} f(\xi)$ if $f_n(s_n) \to f(s)$ as $s_n \to s \in C$.

Proof. Let $G \subset T$ be an open set and suppose that $s \in f^{-1}(G) \cap C$. Since $G$ is open it follows from the assumptions in the theorem that there exists a neighbourhood $N$ of $s$ and a number $m \in \mathbb{N}$ such that $f_k(s') \in G$ for all $k \geq m$ and $s' \in N$. This implies that $N \subset \bigcap_{k=m}^{\infty} f_k^{-1}(G)$ and since $N$ is a neighbourhood of $s$ it follows that $s \in T_m^\infty$, where

$$
T_m^\infty = \bigcap_{k=m}^{\infty} f_k^{-1}(G).
$$

Such $m$ exist for arbitrary $s \in (f^{-1}(G) \cap C)$, so we have that

$$
f^{-1}(G) \cap C \subset \bigcup_{m=1}^{\infty} T_m^\infty.
$$

Let $\mu, \mu_1, \mu_2, \ldots$ be the distributions of $\xi, \xi_1, \xi_2, \ldots$, respectively. Theorem 2.7 (c) and the assumption that $\xi \in C$ almost surely gives

$$
\mu(f^{-1}(G)) = \mu((f^{-1}(G)) \cap C) \leq \mu \left( \bigcup_{m=1}^{\infty} T_m^\infty \right) = \sup_m \mu(T_m^\infty).
$$

(4.2)

By theorem 3.10 (e) and that $T_m^\infty \subset f_k^{-1}(G)$ for sufficiently large $n$, it follows that

$$
\mu(T_m^\infty) \leq \liminf_{n \to \infty} \mu_n(T_m^\infty) \leq \liminf_{n \to \infty} \mu_n(f_k^{-1}(G)).
$$

(4.3)

Equations (4.2) and (4.3) together yields

$$
\mu \circ f^{-1}(G) \leq \liminf_{n \to \infty} \mu_n \circ f^{-1}(G).
$$

Using theorem 3.10 (e) again it follows that $\mu_n \circ f^{-1} \xrightarrow{w} \mu \circ f^{-1}$, which is equivalent to that $f_n(\xi_n) \xrightarrow{d} f(\xi)$ since $\mu \circ f^{-1}, \mu_1 \circ f_1^{-1}, \mu_2 \circ f_2^{-1}, \ldots$ are the distributions of $f(\xi), f_1(\xi_1), f_2(\xi_2), \ldots$, respectively. □
4.2 Random Measures

Now the focus is turned to the case when the random elements are measure valued. In Kallenberg [7], the random measures considered are assumed to take values on a topological space $S$, which is locally compact, second countable and Hausdorff. To avoid any discussion about topology, this section will only consider random measures on a metric space $(S,d)$, where $S$ is an open subset of the euclidean space $\mathbb{R}^d$. This class of spaces form a special case of the ones satisfying the conditions mentioned above. Let $S \subset \mathbb{R}^d$ be open and denote by $\mathcal{B}(S)$ the class of all relatively compact subsets of $S$. A measure $\mu$ on $S$ is said to be locally finite if

$$\mu(B) < \infty, \quad \text{for all } B \in \mathcal{B}(S).$$

Denote by $\mathcal{M}(S)$ be the space of all locally finite measures defined on $S$. If $f$ is a continuous function, then the support of $f$ is defined as

$$\text{supp } f = \{x : f(x) \neq 0\}.$$

Let $C^+_K(S)$ denote the family of continuous functions $f : S \rightarrow \mathbb{R}_+$ with compact support. Before stating the next theorem, the concept of vague convergence is needed. A sequence of measures $\mu_1, \mu_2, ... \in \mathcal{M}(S)$ is said to converge vaguely to $\mu \in \mathcal{M}(S)$ if

$$\int f \, d\mu_n \rightarrow \int f \, d\mu, \quad \text{for all } f \in C^+_K(S).$$

Note that $C^+_K(S) \subset C_b(S)$, so if $\mu_n \xrightarrow{w} \mu$, then $\mu_n \xrightarrow{v} \mu$.

**Theorem 4.5.** Let $f_1, f_2, ...$ be dense in $C^+_K(S)$, where $S \subset \mathbb{R}^d$ is open, and define

$$\rho(\mu, \nu) = \sum_{k=1}^{\infty} \frac{1}{2^k} \min \left\{ \left\| \int f_k \, d\mu - \int f_k \, d\nu \right\|, 1 \right\}, \quad \text{for all } \mu, \nu \in \mathcal{M}(S).$$

Then $(\mathcal{M}(S), \rho)$ is a complete and separable metric space and convergence in this metric is equivalent to vague convergence.

**Proof.** See section A2 in [7].

Now everything is set up for the following definition. A random measure is a random element in $(\mathcal{M}(S), \rho)$, that is, a measurable mapping from the event space $(\Omega, \mathcal{F}, P)$ to $(\mathcal{M}(S), \rho)$. If $\mu \in \mathcal{M}(S)$, define

$$\mathcal{B}(\mu)(S) = \left\{ B \in \mathcal{B}(S) : \mu(\partial B) = 0 \right\}$$

and if $\xi$ is a random measure, define

$$\mathcal{B}(\xi)(S) = \left\{ B \in \mathcal{B}(S) : \xi(\partial B) = 0 \quad \text{a.s.} \right\}.$$

A set $M \subset \mathcal{M}(S)$ is called vaguely relatively compact if every sequence $\mu_1, \mu_2, ... \in M$ has a subsequence that converges vaguely to some element $\mu \in \mathcal{M}(S)$. From theorem 4.5 it follows that a set $M$ is vaguely relatively compact if and only if $\overline{M}$ is compact with respect to the metric $\rho$. Two properties of vague convergence are stated in following theorem.
Theorem 4.6. If $S$ is an open subset of $\mathbb{R}^d$, then the following holds.

(a) A set $M \subset \mathcal{M}(S)$ is vaguely relative compact if and only if

$$\sup_{\mu \in M} \int f \, d\mu < \infty \quad \text{for all } f \in C_K^+(S).$$

(b) If $\mu_n \xrightarrow{v} \mu$, then $\mu_n(B) \to \mu(B)$ for all $B \in \hat{\mathcal{B}}_\mu(S)$.

Proof. See section A2 in [7].

To prove an important uniqueness result for random measures, we will use the following result.

Theorem 4.7. If $m \in \mathcal{M}(S)$, the Borel $\sigma$-algebra over $\mathcal{M}(S)$ is generated by the sets

$$\left\{ \{\mu : \mu(B) \in A\} : A \in \mathcal{B}(\mathbb{R}_+), \quad B \in \hat{\mathcal{B}}_m(S) \right\}.$$

Proof. See section A2 in [7].

The following two uniqueness results will be used in the proof of the main convergence result of this chapter.

Lemma 4.8. If $\xi$ and $\eta$ are two random measures on some open set $S \subset \mathbb{R}^d$ such that

$$(\xi(B_1), \ldots, \xi(B_k)) \overset{d}{=} (\eta(B_1), \ldots, \eta(B_k)), \quad \text{for all } B_1, \ldots, B_k \in \hat{\mathcal{B}}_{\xi+\eta}(S), \quad k \in \mathbb{N},$$

then $\xi \overset{d}{=} \eta$.

Proof. The monotone class theorem stated in section 2.2 will be used to prove this result. Define

$$\mathcal{D} = \left\{ M \in \mathcal{B}(\mathcal{M}(S)) : \mathbb{P} \circ \xi^{-1}(M) = \mathbb{P} \circ \eta^{-1}(M) \right\}.$$

To show that this class of subsets is a $\lambda$-system, first note that $\mathcal{M}(S) \in \mathcal{D}$ since

$$\mathbb{P} \circ \xi^{-1}(\mathcal{M}(S)) = \mathbb{P} \circ \eta^{-1}(\mathcal{M}(S)) = 1.$$

Then let $M_1, M_2 \in \mathcal{D}$ with $M_2 \subset M_1$. By basic properties of measures,

$$\mathbb{P} \circ \xi^{-1}(M_1 \setminus M_2) = \mathbb{P} \circ \xi^{-1}(M_1) - \mathbb{P} \circ \xi^{-1}(M_2) = \mathbb{P} \circ \eta^{-1}(M_1) - \mathbb{P} \circ \eta^{-1}(M_2) = \mathbb{P} \circ \eta^{-1}(M_1 \setminus M_2),$$

so $M_1 \setminus M_2 \in \mathcal{D}$. To prove the last property of a $\lambda$-system, let $M_1, M_2, \ldots$ be an increasing sequence of sets from $\mathcal{D}$. By theorem 2.7 (c),

$$\mathbb{P} \circ \xi^{-1} \left( \bigcup_{i=1}^{\infty} M_i \right) = \lim_{i \to \infty} \mathbb{P} \circ \xi^{-1}(M_i)
= \lim_{i \to \infty} \mathbb{P} \circ \eta^{-1}(M_i)
= \mathbb{P} \circ \eta^{-1} \left( \bigcup_{i=1}^{\infty} M_i \right),$$
so \( \bigcup_{i=1}^{\infty} M_i \in \mathcal{D} \). Then define
\[
\mathcal{C} = \{ \{ \mu : \mu(B_1) \in A_1, \ldots, \mu(B_k) \in A_k \} : A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R}_+) \} \\
B_1, \ldots, B_k \in \bigotimes_{\pi+\eta}(S), \quad k \in \mathbb{N} \}.
\]

To show that \( \mathcal{C} \) is a \( \pi \)-system, let \( M_1 \) and \( M_2 \) be two arbitrary sets in \( \mathcal{C} \), that is
\[
M_i = \{ \mu : \mu(B_1) \in A_1^i, \ldots, \mu(B_k) \in A_k^i \}, \quad i = 1, 2.
\]
Then the intersection of \( M_1 \) and \( M_2 \) is
\[
\{ \mu : \mu(B_1^1) \in A_1^1, \ldots, \mu(B_k^1) \in A_k^1, \mu(B_1^2) \in A_1^2, \ldots, \mu(B_k^2) \in A_k^2 \}.
\]
This set belongs to \( \mathcal{C} \), so it follows that \( \mathcal{C} \) is a \( \pi \)-system. If \( M \in \mathcal{C} \), then by the assumption in the lemma,
\[
\mathbb{P} \circ \xi^{-1}(M) = \mathbb{P}\{ \xi \in \{ \mu : \mu(B_1) \in A_1, \ldots, \mu(B_k) \in A_k \}\}
= \mathbb{P}\{ \xi(B_1) \in A_1, \ldots, \xi(B_k) \in A_k \}
= \mathbb{P}\{ \eta(B_1) \in A_1, \ldots, \eta(B_k) \in A_k \}
= \mathbb{P}\{ \eta \in \{ \mu : \mu(B_1) \in A_1, \ldots, \mu(B_k) \in A_k \}\}
= \mathbb{P} \circ \eta^{-1}(M),
\]
so \( M \in \mathcal{D} \) and hence \( \mathcal{C} \subset \mathcal{D} \). It follows from the monotone class theorem that \( \sigma(\mathcal{C}) \subset \mathcal{D} \) and from theorem 4.7 that \( \sigma(\mathcal{C}) = \mathcal{B}(\mathcal{M}(S)) \), so it can be concluded that \( \xi \overset{d}{=} \eta \).

\[\Box\]

**Lemma 4.9.** If \( \xi \) and \( \eta \) are two random measures on some open set \( S \subset \mathbb{R}^d \) such that
\[
\int f \, d\xi \overset{d}{=} \int f \, d\eta \quad \text{for all } f \in C_K^+(S),
\]
then \( \xi \overset{d}{=} \eta \).

**Proof.** First define the \( \sigma \)-algebras \( \mathcal{F} \) and \( \mathcal{G} \) as follows.
\[
\mathcal{F} = \sigma \left\{ \left\{ \mu : \int f \, d\mu \in A \right\} : f \in C_K^+(S), \quad A \in \mathcal{B}(\mathbb{R}_+) \right\},
\]
\[
\mathcal{G} = \sigma \left\{ \{ \mu : \mu(B) \in A \} : B \in \mathcal{B}(S), \quad A \in \mathcal{B}(\mathbb{R}_+) \right\}.
\]
Almost by following the proof of lemma 4.8 line by line, it can be shown that
\[
\mathbb{P} \circ \xi^{-1}(M) = \mathbb{P} \circ \eta^{-1}(M), \quad \text{for all } M \in \mathcal{F}.
\]
By theorem 4.7, the Borel \( \sigma \)-algebra over \( \mathcal{M}(S) \) is generated by a subset of \( \mathcal{G} \), so it is enough to show that \( \mathcal{G} \subset \mathcal{F} \). For any \( f \in C_K^+(S) \), define the mapping \( \pi_f \) by
\[
\pi_f(\mu) = \int f \, d\mu, \quad \text{for all } \mu \in \mathcal{M}(S)
\]
and for any \( B \in \mathcal{B}(S) \), define the mapping \( \pi_B \) by
\[
\pi_B(\mu) = \mu(B), \quad \text{for all } \mu \in \mathcal{M}(S).
\]
To prove that $G \subset F$, it is enough to show that $\pi_B$ is $F$-measurable for any $B \in \mathcal{B}(S)$. First let $K \subset S$ be an arbitrary compact set and choose a sequence of functions $f_1, f_2, \ldots \in C_K(S)$ such that $f_n \downarrow 1_K$. Then $\pi_{f_1}, \pi_{f_2}, \ldots$ are $F$-measurable functions that converge pointwise to $\pi_K$ on $\mathcal{M}(S)$, so it follows that

$$\pi_K \text{ is } F\text{-measurable for any compact } K \subset S. \quad (4.4)$$

Now let $K \subset S$ be a fixed compact set. The monotone class theorem will be used to prove that $\pi_B$ is $F$-measurable for any Borel set $B \subset K$. Define

$$D = \{ B \in \mathcal{B}(K) : \pi_B \text{ is } F\text{-measurable.} \}.$$

To prove that this is a $\lambda$-system, first note that it follows from (4.4) that $K \in D$. Then let $B_1, B_2, \ldots$ be a sequence of increasing sets in $D$ and let $B' = \bigcup_{i=1}^{\infty} B_i$. By theorem 2.7 (c),

$$\lim_{i \to \infty} \pi_{B_i}(\mu) = \lim_{i \to \infty} \mu(B_i) = \mu(B') = \pi_{B'}(\mu),$$

for any $\mu \in \mathcal{M}(S)$, so $\pi_{B'}$ is $F$-measurable and hence $B' \in D$. To prove the third property of a $\lambda$-system, let $A$ and $B$ be Borel subsets of $K$ such that $B \subset A$ and note that

$$\pi_{A \setminus B}(\mu) = \mu(A \setminus B) = \mu(A) - \mu(B) = \pi_A(\mu) - \pi_B(\mu),$$

for any $\mu \in \mathcal{M}(S)$. Since the difference of two measurable functions is again measurable it follows that $A \setminus B \in D$. Now define

$$C = \{ B \in \mathcal{B}(K) : B \text{ is compact} \}.$$

This family of subsets is a $\pi$-system since the intersection of any two compact sets is again a compact set and (4.4) implies that $C \subset D$. By the monotone class theorem, $\sigma(C) \subset D$ and since $\sigma(C) = \mathcal{B}(K)$, it follows that $\pi_B$ is $F$-measurable for any Borel subset $B$ of $K$. In particular, this holds for any $B \in \mathcal{B}(K)$. Since $K$ was arbitrary, this proves the lemma. \qed

### 4.3 Convergence in Distribution of Random Measures

The proof of the key theorem for convergence of random measures uses the following tightness criterion.

**Lemma 4.10.** If $\xi_1, \xi_2, \ldots$ are random measures on some open set $S \subset \mathbb{R}^d$, then the following three statements are equivalent.

(a) $\{\xi_n\}$ is relatively compact in distribution.

(b) $\{ \int f \, d\xi_n \}$ is tight in $\mathbb{R}_+$ for every $f \in C_K^+(S)$.

(c) $\{\xi_n(B)\}$ is tight in $\mathbb{R}_+$ for every $B \in \mathcal{B}(S)$. 

Proof. (a) implies (b): Assume that \( \{\xi_n\} \) is relatively compact in distribution. Since \( (\mathcal{M}(S), \rho) \) is complete and separable, it follows from theorem 4.2 that relative compactness in distribution is equivalent to tightness for \( \{\xi_n\} \). Let \( f \in C^+_K(S) \) be arbitrary and define the mapping \( \pi_f : \mathcal{M}(S) \to \mathbb{R}_+ \) by \( \pi_f(\mu) = \int f \, d\mu \). Since this mapping is continuous, it follows from lemma 4.3 that \( \{\pi_f(\xi_n)\} \) is tight in \( \mathbb{R}_+ \).

(b) implies (c): Let \( B \in \tilde{\mathcal{B}}(S) \) be arbitrary. Choose open relatively compact sets \( G_1, G_2, \ldots \in \tilde{\mathcal{B}}(S) \) such that \( B \subset \bigcup_{i=1}^{\infty} G_i \). Since \( B \) is compact, there exists \( i_1, \ldots, i_m \) such that \( C = G_{i_1} \cup \ldots \cup G_{i_m} \) covers \( B \). Here \( C \) is relatively compact because a finite union of compact sets is again compact. Then we can choose \( f \in C^+_K(S) \) such that \( \text{supp} \, f = C \) and \( f = 1 \) on \( B \). Note that \( \xi_n(B) \leq \int f \, d\xi_n \) for any \( n \). From this and the assumption, it follows that for every \( \varepsilon > 0 \), there exists a positive real number \( r \) satisfying

\[
\inf_n \mathbb{P}\{\xi_n(B) \leq r\} \geq \inf_n \mathbb{P}\left\{ \int f \, d\xi_n \leq r \right\} \geq 1 - \varepsilon,
\]

so \( \{\xi_n(B)\} \) is tight.

(c) implies (a): Let \( G_1, G_2, \ldots \in \tilde{\mathcal{B}}(S) \) be open sets such that \( S \subset \bigcup_{k=1}^{\infty} G_k \). Then for each \( \varepsilon > 0 \) and \( k \in \mathbb{N} \), there exists a positive real number \( r_k \) such that

\[
\inf_n \mathbb{P}\{\xi_n(G_k) \leq r_k\} \geq 1 - \varepsilon 2^{-k}.
\]

Define \( \bar{A} = \bigcap_{k=1}^{\infty} \{\mu : \mu(G_k) \leq r_k\} \) and let \( f \in C^+_K(S) \) be arbitrary. Since \( f \) has compact support there exists only a finite collection of sets \( G_{i_1}, \ldots, G_{i_m} \) where \( f \neq 0 \). From this it follows that

\[
\sup_{\mu \in \bar{A}} \int f \, d\mu \leq \sup_{\mu \in \bar{A}} \sum_{k=1}^{m} \int_{G_{i_k}} f \, d\mu \leq \sup_{\mu \in \bar{A}} \sum_{k=1}^{m} \|f\| \mu(G_{i_k}) \leq \|f\| \sum_{k=1}^{m} r_k < \infty,
\]

so the set \( \bar{A} = \bigcap_{k=1}^{\infty} \{\mu : \mu(G_k) \leq r_k\} \) is relatively compact by the first statement of theorem 4.6. This implies that \( \bar{A} \), the closure of \( A \), is compact and

\[
\inf_n \mathbb{P}\{\xi_n \in \bar{A}\} \geq \inf_n \mathbb{P}\{\xi_n \in A\}
\]

\[
= 1 - \sup_n \mathbb{P}\{\xi_n \in A^c\}
\]

\[
= 1 - \sup_n \mathbb{P}\left\{ \xi_n \in \bigcup_{k=1}^{\infty} \{\mu : \mu(G_k) > r_k\} \right\}
\]

\[
\geq 1 - \sup_n \sum_{k=1}^{\infty} \mathbb{P}\{\xi_n \in \mu(G_k) > r_k\}
\]

\[
\geq 1 - \sum_{k=1}^{\infty} \sup_n \mathbb{P}\{\xi_n(G_k) > r_k\}
\]

\[
= 1 - \sum_{k=1}^{\infty} \left( 1 - \inf_n \mathbb{P}\{\xi_n(G_k) \leq r_k\} \right)
\]

\[
\geq 1 - \sum_{k=1}^{\infty} \varepsilon 2^{-k} = 1 - \varepsilon.
\]
so \( \{\xi_n\} \) is tight and hence relatively compact in distribution. \( \square \)

**Corollary 4.11.** If \( \xi_1, \xi_2, ... \) are random probability measures on some open set \( S \subset \mathbb{R}^d \), then \( \{\xi_n\} \) is relatively compact in distribution.

**Proof.** For any \( B \in \mathcal{B}(S) \), \( \mathbb{P}\{\xi_n(B) \in [0, 1]\} = 1 \) for all \( n \in \mathbb{N} \), so \( \{\xi_n(B)\} \) is tight in \( \mathbb{R}_+ \). It follows from lemma 4.10 that \( \{\xi_n\} \) is relatively compact in distribution. \( \square \)

The next theorem is the main result for convergence of random measures.

**Theorem 4.12.** If \( \xi, \xi_1, \xi_2, ... \) are random measures on some open set \( S \subset \mathbb{R}^d \), then the following statements are equivalent.

(a) \( \xi_n \overset{d}{\to} \xi \).

(b) \( \int f \, d\xi_n \overset{d}{\to} \int f \, d\xi \) for all \( f \in C^+_K(S) \).

(c) \( (\xi_n(B_1), ..., \xi_n(B_k)) \overset{d}{\to} (\xi(B_1), ..., \xi(B_k)) \), \( B_1, ..., B_k \in \mathcal{B}(S) \), \( k \in \mathbb{N} \).

**Proof.** (a) implies (b): Let \( f \in C^+_K(S) \) be arbitrary and define the function \( g \) by

\[
g(\mu) = \int f \, d\mu \quad \text{for all} \quad \mu \in \mathcal{M}(S).
\]

Let \( C = \mathcal{M}(S) \) and suppose that \( \mu_1, \mu_2, ... \in \mathcal{M}(S) \) and \( \mu \in C \) are arbitrary measures such that \( \hat{\mu}(\mu_1, \mu) \rightarrow 0 \). By theorem 3.10, \( g(\mu_n) \rightarrow g(\mu) \) and since \( \xi \in C \) a.s. it follows from theorem 4.4 that \( g(\xi_n) \overset{d}{\to} g(\xi) \), that is,

\[
\int f \, d\xi_n \overset{d}{\to} \int f \, d\xi.
\]

(a) implies (c): Let \( B_1, ..., B_k \in \mathcal{B}(S) \) be arbitrary. Define the function \( g : \mathcal{M}(S) \to \mathbb{R}_+^k \) by \( g(\mu) = (\mu(B_1), ..., \mu(B_k)) \) and let

\[
C = \{\mu \in \mathcal{M}(S) : \mu(\partial B_i) = 0 \quad \text{for all} \quad i = 1, ..., k\}.
\]

Suppose that \( \mu_1, \mu_2, ... \in \mathcal{M}(S) \) and \( \mu \in C \) are arbitrary measures such that \( \hat{\mu}(\mu_1, \mu) \rightarrow 0 \). By theorem 3.10, \( \mu_n(B_i) \rightarrow \mu(B_i) \) for all \( i = 1, ..., k \), so \( g(\mu_n) \rightarrow g(\mu) \). Since \( \xi \in C \) a.s. it follows from theorem 4.4 that \( g(\xi_n) \overset{d}{\to} g(\xi) \), that is,

\[
(\xi_n(B_1), ..., \xi_n(B_k)) \overset{d}{\to} (\xi(B_1), ..., \xi(B_k)).
\]

(b) implies (a): By assumption, \( \{\int f \, d\xi_n\} \) is relatively compact in distribution and hence tight for all \( f \in C^+_K(S) \). From this and lemma 4.10 it follows that \( \{\xi_n\} \) is relatively compact in distribution. Now, assume that \( \xi_n \nRightarrow \xi \). Then for every \( \varepsilon > 0 \), there exists a function \( f \in C_b(\mathcal{M}(S)) \) and a subsequence \( N' \subset \mathbb{N} \) such that

\[
|E(f(\xi_n)) - E(f(\xi))| > \varepsilon \quad \text{for all} \quad n \in N'.
\]

Since \( \{\xi_n\} \) is relatively compact in distribution, there exists a further subsequence \( N'' \subset N' \) and a random measure \( \eta \) such that

\[
\xi_n \overset{d}{\to} \eta \quad \text{along} \quad N''.
\]
Let \( f \in C^+_K(S) \) be arbitrary. The assumption implies that \( \int f \, d\xi_n \xrightarrow{d} \int f \, d\xi \) along \( N'' \). Since \( \xi_n \xrightarrow{d} \eta \) along \( N'' \), it follows from the first part of the proof that \( \int f \, d\xi_n \xrightarrow{d} \int f \, d\eta \) along \( N'' \) and hence
\[
\int f \, d\xi \xrightarrow{d} \int f \, d\eta \quad \text{for all } f \in C^+_K(S).
\]
By lemma 4.9, this implies that \( \xi \xrightarrow{d} \eta \). From this and (4.6), it follows that \( \xi_n \xrightarrow{d} \xi \) along \( N'' \), but this is a contradiction because of (4.5), so it can be concluded that \( \xi_n \xrightarrow{d} \xi \).

(c) implies (a): It follows from the assumption that \( \{\xi_n(B)\} \) is relatively compact in distribution and hence tight for all \( B \in \hat{\mathcal{B}}(S) \). Then lemma 4.10 implies that \( \{\xi_n\} \) is relatively compact in distribution. Arguing as in the proof that (b) implies (a), it remains to show that if
\[
(\xi(B_1), \ldots, \xi(B_k)) \overset{d}{=} (\eta(B_1), \ldots, \eta(B_k)), \quad \text{for all } B_1, \ldots, B_k \in \hat{\mathcal{B}}_{\xi+\eta}(S), \quad k \in \mathbb{N},
\]
then \( \xi \overset{d}{=} \eta \), but this follows from lemma 4.8. \( \square \)
Chapter 5

Convergence of Particle Systems

In this chapter, a system of particles in motion in a bounded region is considered. The state of a particle at time $t$ is given by its location and its velocity. This is a random process and the statistical distribution of the particles is given by a density function which is the solution to the Boltzmann equation described in this chapter. To approximate this solution, the paper [3] considers systems of $N$ particles and shows how these systems in some sense converge to the particle system described by the Boltzmann equation. In section 1, the Boltzmann equation considered in [3] is outlined. Section 2 presents the construction of the $N$-particles systems and how convergence of these systems are defined.

5.1 The Boltzmann Equation

Let $D \subset \mathbb{R}^d$ be a bounded open set with smooth boundary, where $d = 2$ or $d = 3$, and let

$$V = \{ v \in \mathbb{R}^d : 0 < v_{\text{min}} < \|v\| < v_{\text{max}} \}.$$  

Here $D$ is called the physical space and $V$ is called the velocity space. The probability density function for the distribution of one particle at time $t$ is for each $t$ given by $p(y, v, t)$, where $(y, v, t) \in D \times V \times [0, T]$. For any $v, v_1 \in V$ and unit vector $e \in \mathbb{R}^d$ define

$$v^* = v + \langle e, (v_1 - v) \rangle e,$$

$$v^*_1 = v_1 - \langle e, (v_1 - v) \rangle e,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{R}^d$ defined by $\langle a, b \rangle = \sum_{i=1}^{d} a_i b_i$ for all $a, b \in \mathbb{R}^d$. Let $S^{d-1}$ be the unit sphere and let

$$S^{d-1}_+ = \{ e \in S^{d-1} : \langle e, (v - v_1) \rangle \geq 0 \}.$$

The collision kernel is a function $B : \mathbb{R}^d \times \mathbb{R}^d \times S^{d-1} \rightarrow \mathbb{R}_+$ such that

$$B(v, v_1, e) \leq c_1 < \infty,$$
for some constant $c_1$. Furthermore, $h_\beta \in L^\infty$ is a symmetric function vanishing for $|y - z| \geq \beta > 0$ which satisfies $\int h_\beta(y, z) \, dz = 1$. Using all this, the collision operator $Q$ is defined as

$$Q(p, p)(y, v, t) = \int_D \int_V \int_{S^{d-1}_+} B(v, v_1, e) h_\beta(y, z) \mathbf{1}_{\{(v^*, v_1^*) \in V \times V\}}$$

$$\times \left( p(y, v^*, t)p(z, v_1^*, t) - p(y, v, t)p(z, v_1, t) \right) \, dz \, dv_1 \, de.$$

Let $n(y)$ denote the outward normal at point $y \in \partial D$. The incoming flux at point $y$ and time $t$ is given by

$$J(y, t) = \int_{\{v, n(y)\} \geq 0} (v, n(y)) p(y, v, t) \, dv.$$

The function $M$ is a bounded positive function defined on the set

$$\{(y, v) \in D \times V : y \in \partial D, \quad \langle v, n(y) \rangle \leq 0\},$$

such that

$$\int_{\{v, n(y)\} \leq 0} |\langle v, n(y) \rangle| M(y, v) \, dv = 1.$$

Using all this, the Boltzmann equation can be presented.

**Differential Version**

Let $\lambda$ be a real parameter. The following is the differential version of a Boltzmann equation of cutoff type with initial and boundary conditions.

$$\partial_t p(y, v, t) + \langle v, \nabla_y \rangle p(y, v, t) = \lambda Q(p, p)(y, v, t),$$

$$p(y, v, 0) = p_0(y, v) \geq 0,$$

$$p(y, v, t) = J(y, t) M(y, v), \quad y \in \partial D, \quad \langle v, n(y) \rangle \leq 0,$$

where $\nabla_y = (\partial_{y_1}, ..., \partial_{y_d})$. Physically, these equations describe a rarefied gas in a vessel with diffusive boundary conditions.

**Integrated Version**

The Knudsen semigroup $S(t)$ is the solution the following initial boundary value problem.

$$\left( \partial_t + \langle v, \nabla_y \rangle \right) S(t)p_0(y, v) = 0,$$

$$S(t)p_0(y, v) = J(y, t) M(y, v), \quad y \in \partial D, \quad \langle v, n(y) \rangle \leq 0.$$

The integrated version of the Boltzmann equation above is

$$p(y, v, t) = S(t)p_0(y, v) + \lambda \int_0^t S(t - s) Q(p, p)(y, v, s) \, ds.$$
The Stationary Boltzmann Equation

If the particle system is in equilibrium, the distribution of the particles will not depend on the time $t$ and then $\partial_t p(y,v,t) = 0$. In this case, let $p(y,v,t) = g(y,v)$. The stationary version of the Boltzmann equation above is then
\[
\langle v, \nabla_y \rangle g(y,v) = \lambda Q(g,g)(y,v),
\]
together with boundary conditions corresponding to the time dependent case, see equation (5.1), and the normalization property
\[
\int \int g(y,v) dy dv = 1.
\]
The following theorem establishes uniqueness for the stationary Boltzmann equation.

**Theorem 5.1.** There exists $\lambda_0 > 0$ such that for any $\lambda \leq \lambda_0$ there exists a unique probability density function $g$ such that
\[
g = S(t)g + \lambda \int_0^t S(t-s)Q(g,g) ds, \quad t \in \mathbb{R}_+.
\]
Moreover, for this density $g$, there exists a constant $c$ such that
\[
\|p(t) - g\|_{L^1} \leq e^{ct},
\]
where $p(t)$ is any solution to (5.2).

**Proof.** See appendix in [3].

5.2 The $N$-particle Process

This section presents how the particle system described by the stationary Boltzmann equation can be approximated by a system containing $N$ particles and how the paper [3] defines convergence of these $N$-particle systems as $N \to \infty$.

Define
\[
D = \{ x \in \mathbb{R}^{2d} : x_1,\ldots,x_d \in D, \quad x_{d+1},\ldots,x_{2d} \in V \}
\]
and note that $D$ is bounded since both $D$ and $V$ is. Let $x_N \in D^N$ be defined by
\[
x_N = (x_1,\ldots,x_N), \quad x_i = (y_i,v_i), \quad i = 1,\ldots,N.
\]
For any function $f$ and $x_N \in D^N$ define
\[
G_N^{free}(f)(x_N) = \sum_{i=1}^N \langle v_i, \nabla_y \rangle f(x_N).
\]
For $1 \leq i < j \leq N$ let
\[
q(x_i,x_j,c) = h_\beta(y_i,y_j)B(v_i,v_j,c)1_{\{ (v_i^*,v_j^*) \in V \times V \}}
\]
and
\[
x_N^{(i,j)} = (x_1,\ldots,x_{i-1},y_i,v_i^*,\ldots,x_{j-1},y_j,v_j^*,\ldots,x_N).
\]
Using this, for any function \( f \) and \( x_N \in D^N \) define
\[
G_N^{\text{jump}}(f)(x_N) = \sum_{1 \leq i < j \leq N} \int_{S^{d-1}} \left(f(x_N^{(i,j)}) - f(x_N)\right) q(x_i, x_j, e) \, de
\]
and
\[
G_N(f)(x_N) = G_N^{\text{free}}(f)(x_N) + \frac{\lambda}{N} G_N^{\text{jump}}(f)(x_N).
\]
For all \( i = 1, \ldots, N \) let
\[
x_N(i) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)
\]
and
\[
J^N_i(y_i, t, x_N(i)) = \int_{\{v, n(y_i)\geq 0\}} \langle v_i, n(y_i) \rangle f^N(x_N, t).
\]
The \( N \)-particle process is defined as the solution the the following initial boundary value problem.
\[
\partial_t f^N(x_N, t) + \sum_{i=1}^N \langle v_i, \nabla y_i \rangle f^N(x_N, t) = \frac{\lambda}{N} G_N^{\text{jump}}(f^N)(x_N),
\]
\[
f^N(x_N, 0) = f^N(x_N),
\]
\[
f^N(x_1, \ldots, x_i, \ldots, x_N, t) = J^N_i(y_i, t, x_N(i)) M(y_i, v_i), \quad y_i \in \partial D,
\]
\[
\langle v_i, n(y_i) \rangle \leq 0, \quad i = 1, \ldots, N.
\]
As in the case of the Boltzmann equation, the stationary version of the above problem is obtained by putting \( \partial_t f^N(x_N, t) = 0 \). The stationary version is
\[
\sum_{i=1}^N \langle v_i, \nabla y_i \rangle f^N(x_N) = \frac{\lambda}{N} G_N^{\text{jump}}(f^N)(x_N),
\]
with boundary conditions corresponding to the time dependent case, see equation (5.3). The following theorem establishes existence and uniqueness to this problem.

**Theorem 5.2.** For any \( N \in \mathbb{N} \), there exists a unique probability density function \( f^N \) which is invariant under the \( N \)-particle process.

**Proof.** See appendix in [3]. \( \square \)

This means that the states of the particles in an \( N \)-particle system in equilibrium can be described by a random vector \( X^N = (X^N_1, \ldots, X^N_N) \), where the distribution is given by the density function \( f^N \). The random vectors \( X^N_1, \ldots, X^N_N \) take values in \( D \) and are identically distributed, but not necessarily independent. The question is if the sequence of particle systems
\[
X^1 = (X^1_1)
\]
\[
X^2 = (X^2_1, X^2_2)
\]
\[
X^3 = (X^3_1, X^3_2, X^3_3)
\]
\[
\vdots
\]
in some sense converge to the particle system described by the stationary Boltzmann equation as $N \to \infty$. To answer this question, the paper [3] compares $\tilde{f}^N$ with the solution to the stationary Boltzmann equation. For this purpose, the $k$-particle density and $k$-fold product is introduced. The $k$-particle density is defined by

$$\tilde{f}^N_k(x_k) = \int \cdots \int \tilde{f}^N(x_N) \, dx_{k+1} \cdots dx_N, \quad \text{if } k = 1, \ldots, N-1,$$

and $\tilde{f}^N_k(x_k) = \tilde{f}^N(x_N)$ if $k = N$. Let $g$ be the solution to the stationary Boltzmann equation. The $k$-fold product of $g$ is defined as

$$g_k(x_k) = \prod_{i=1}^{k} g(x_i), \quad x_k = (x_1, \ldots, x_k) \in D^k.$$

Using this, it is said in [3] that the particle systems $X^N$ converge to the particle system described by the stationary Boltzmann equation if

$$\lim_{N \to \infty} \| \tilde{f}^N_k - g_k \|_{L^1} = 0, \quad \text{for all } k \in \mathbb{N}.$$

The following theorem is the main result of the paper [3] for the stationary case.

**Theorem 5.3.** There exists $\lambda_0 > 0$ such that for any $\lambda \leq \lambda_0$ and $k \in \mathbb{N}$, there exists a constant $c$ satisfying

$$\| \tilde{f}^N_k - g_k \|_{L^1} \leq \frac{c^k}{N}, \quad \text{for all } N > k,$$

where $c$ does not depend on $\lambda, k, N$.

**Proof.** See section 3 in [3].
Chapter 6

Comparison of Modes of Convergence

In previous chapter, convergence of the particle systems $X_1, X_2, ...$ were defined as in [3]. In this chapter this convergence is compared with two other convergence types for particle systems. In section 1 and 2, two results that give relations between the different types of convergences is proved. Section 3 discusses the speed of the convergence. To avoid too many repetitions, in the remainder of this chapter we will use the following notation.

- If $k, N \in \mathbb{N}$, where $k \leq N$, then $\tilde{f}_N^k$ is the $k$-particle density in the system containing $N$ particles as described in section 5.2.
- The density function $g$ is the solution to the stationary Boltzmann equation.
- For any $k \in \mathbb{N}$, the $k$-fold product of $g$ is denoted by $g_k$. Recall that it is defined by
  \[ g_k(x_1, ..., x_k) = \prod_{i=1}^{k} g(x_i), \quad x_1, ..., x_k \in D. \]

6.1 Comparison of Convergence I

For each particle system $X^N = (X^N_1, ..., X^N_N)$, the random empirical measure $\xi_N$ is defined by

\[ \xi_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^N_i}, \quad N = 1, 2, ... \]

So $\xi_1, \xi_2, ...$ is a sequence of random measures such that for any $A \in \mathcal{B}(D)$,

\[ \xi_N(A) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^N_i}(A) \]

is a random variable describing the proportion of particles in $A$ for the $N$-particle process. This motivates that we can say that the particle systems $X^1, X^2, ...$
converge if $\xi_N$ converges in distribution to some probability measure. The next result is a relation between this mode of convergence and the one described in previous chapter. To prove this we follow and modify an idea outlined in Grigorescu, Kang [6], section 5.2, corollary 3, and Sznitman [8], section 1.2, proposition 2.2.

**Theorem 6.1.** Let $\mu$ be the measure generated by $g$, the solution to the stationary Boltzmann equation, with respect to the Lebesgue measure. If

$$
\lim_{N \to \infty} \| f_N^k - g_k \|_{L^1} = 0, \quad \text{for } k \in \{1, 2\}, \quad (6.1)
$$

then

$$
\frac{1}{N} \sum_{i=1}^{N} \delta_{X_N^i} \overset{d}{\rightarrow} \mu.
$$

**Proof.** By theorem 4.12, $\xi_N \overset{d}{\rightarrow} \mu$ is equivalent to that

$$
\int \phi \xi_N \overset{d}{\rightarrow} \int \phi d\mu
$$

for all $\phi \in C^+_K(D)$. Since $\mu$ is a non-random measure it follows from theorem 4.1 that it is equivalent to show this convergence in probability, that is,

$$
\lim_{N \to \infty} P \left\{ \left| \int \phi d\xi_N - \int \phi d\mu \right| > \varepsilon \right\} = 0, \quad \text{for all } \phi \in C^+_K(D) \text{ and } \varepsilon > 0.
$$

Note that if $\phi \in C^+_K(D)$, it always holds that

$$
\int \phi d\mu = \int \phi(x)g(x) dx \leq \|\phi\|_{\infty} \int g(x) dx = \|\phi\|_{\infty} \cdot 1 < \infty,
$$

since all functions in $C^+_K(D)$ are bounded. By Chebychev’s inequality,

$$
P \left\{ \left| \int \phi d\xi_N - \int \phi d\mu \right| > \varepsilon \right\} \leq \frac{\mathbb{E} \left( \left| \int \phi d\xi_N - \int \phi d\mu \right|^2 \right)}{\varepsilon^2}, \quad (6.2)
$$

for every $\varepsilon > 0$, so it is enough to prove that

$$
\lim_{N \to \infty} \mathbb{E} \left( \left| \int \phi d\xi_N - \int \phi d\mu \right|^2 \right) = 0 \quad \text{for all } \phi \in C^+_K(D).
$$

To show this, first note that

$$
\int \phi d\xi_N = \frac{1}{N} \sum_{i=1}^{N} \phi(X_i^N).
$$

From this it follows that

$$
\mathbb{E} \left( \left| \int \phi d\xi_N - \int \phi d\mu \right|^2 \right) = \mathbb{E} \left( \left| \frac{1}{N} \sum_{i=1}^{N} \left( \phi(X_i^N) - \int \phi d\mu \right) \right|^2 \right)
$$

$$
= \mathbb{E}(S_1^N + S_2^N),
$$

where

$$
S_1^N = \frac{1}{N^2} \sum_{i=1}^{N} \left( \phi(X_i^N) - \int \phi d\mu \right)^2
$$

and

$$
S_2^N = \frac{1}{N^2} \sum_{i=1}^{N} \left( \phi(X_i^N) - \int \phi d\mu \right)^2
$$

are the terms contributing to the convergence.
and \[ S_2^N = \frac{1}{N^2} \sum_{1 \leq i \neq j \leq N} \left( \phi(X_i^N) - \int \phi \, d\mu \right) \left( \phi(X_j^N) - \int \phi \, d\mu \right). \]

Now it remains to show that \( E(S_1^N) \) and \( E(S_2^N) \) tend to zero as \( N \to \infty \). Since \( X_1^N, \ldots, X_N^N \) are identically distributed, it holds that

\[
E(S_1^N) = \frac{1}{N} \left( \phi(X_1^N) - \int \phi \, d\mu \right)^2.
\]

This tends to zero as \( N \to \infty \), since

\[
E(S_1^N) \leq \frac{1}{N} \left( \mathbb{E}(\phi(X_1^N)^2) + \left( \int \phi \, d\mu \right)^2 \right) = \frac{1}{N} \left( \int \phi(x)^2 \tilde{f}_1^N(x) \, dx + \left( \int \phi(x) g(x) \, dx \right)^2 \right) \leq \frac{2\|\phi\|_\infty^2}{N}.
\]

The random vectors \( \{(X_i^N, X_j^N) : 1 \leq i \neq j \leq N\} \) are identically distributed so it follows that

\[
E(S_2^N) = \frac{N(N-1)}{N^2} \mathbb{E} \left( \left( \phi(X_1^N) - \int \phi \, d\mu \right) \left( \phi(X_2^N) - \int \phi \, d\mu \right) \right) = \frac{N(N-1)}{N^2} \mathbb{E}(h(X_1^N, X_2^N)).
\]

Here, \( \lim_{N \to \infty} (N(N-1))/N^2 = 1 \), so if the expected value on the right hand side tends to zero as \( N \to \infty \), then \( \lim_{N \to \infty} \mathbb{E}(S_2^N) = 0 \). To show this, note that the expectation can be expanded as

\[
\mathbb{E}(h(X_1^N, X_2^N)) = \mathbb{E}(\phi(X_1^N)\phi(X_2^N)) - 2 \left( \int \phi \, d\mu \right) \mathbb{E}(\phi(X_1^N)) + \left( \int \phi \, d\mu \right)^2.
\]

By Fubini’s theorem,

\[
\left( \int \phi \, d\mu \right)^2 = \left( \int \phi(x_1) g(x_1) \, dx_1 \right) \left( \int \phi(x_2) g(x_2) \, dx_2 \right) = \int \int \phi(x_1) \phi(x_2) g_2(x) \, dx.
\]

From this it follows that

\[
\left| \mathbb{E}(\phi(X_1^N)\phi(X_2^N)) - \left( \int \phi \, d\mu \right)^2 \right| = \left| \int \int \phi(x_1) \phi(x_2) \left( \tilde{f}_2^N(x) - g_2(x) \right) \, dx \right| \leq \|\phi\|_\infty^2 \|\tilde{f}_2^N - g_2\|_{L^1},
\]

and

\[
\left| \mathbb{E}(\phi(X_1^N)) - \int \phi \, d\mu \right| = \left| \int \phi(x_1) \left( \tilde{f}_1^N(x) - g_1(x) \right) \, dx \right| \leq \|\phi\|_\infty \|\tilde{f}_1^N - g_1\|_{L^1}.
\]
Using this,

\[ |\mathbb{E}(h(X_1^N, X_2^N))| \leq \left| \mathbb{E}(\phi(X_1^N)\phi(X_2^N)) - \left( \int \phi \text{d}\mu \right)^2 \right| + 2 \left| \int \phi \text{d}\mu \right| \left| \mathbb{E}(\phi(X_1^N)) - \int \phi \text{d}\mu \right| \]

\[ \leq \|\phi\|_\infty^2 \left( \|\hat{f}_2 - g_2\|_{L^1} + 2\|\hat{f}_1 - g_1\|_{L^1} \right), \]

so it follows from the assumption in the theorem that \( \mathbb{E}(h(X_1^N, X_2^N)) \to 0 \) as \( N \to \infty \).

\[ \square \]

### 6.2 Comparison of Convergence II

As described in section 2.4, the density functions \( \hat{f}_k \) and \( g_k \) generate measures on \( D^k \). For fixed \( k \in \mathbb{N} \), let \( \mu_k, \mu_1^k, \mu_2^k, \ldots \) be the measures generated by \( g_k, \hat{f}_k^{k+1}, \hat{f}_k^{k+2}, \ldots \) respectively, with respect to the Lebesgue measure. We can say that the particle systems \( X^1, X^2, \ldots \) converge if \( \mu_k^N \xrightarrow{w} \mu_k \) for all \( k \in \mathbb{N} \), or equivalently

\[ \lim_{N \to \infty} \int_{D^k} \phi(x) \hat{f}_k^N(x) \text{d}x = \int_{D^k} \phi(x) g_k(x) \text{d}x, \quad \phi \in C_b(D^k), \quad k \in \mathbb{N}. \]

The following theorem gives a relation between this mode of convergence and the one described in chapter 5.

**Theorem 6.2.** Let \( \mu_k, \mu_1^k, \mu_2^k, \ldots \) be the measures generated by \( g_k, \hat{f}_k^{k+1}, \hat{f}_k^{k+2}, \ldots \) respectively, and let \( \mu \) be the measure generated by \( g \), the solution to the stationary Boltzmann equation. If

\[ \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i^N} \overset{d}{\to} \mu, \]

then

\[ \mu_k^N \overset{w}{\to} \mu_k, \quad \text{for all} \quad k \in \mathbb{N}. \]

To prove this, we adopt an idea of Sznitman [8], section 1.2, proposition 2.2. The following two lemmas will be used in the proof.

**Lemma 6.3.** For any \( k \in \mathbb{N} \), let

\[ A_k = \left\{ \varphi_1(x_1) \cdots \varphi_k(x_k) \in C(\overline{D^k}) : \varphi_1, \ldots, \varphi_k \in C(D), \quad x_1, \ldots, x_k \in D \right\}. \]

If \( \xi_N \overset{d}{\to} \mu \), then for any fixed \( k \in \mathbb{N} \),

\[ \lim_{N \to \infty} \int_{D^k} \varphi \text{d}\mu_k^N = \int_{D^k} \varphi \text{d}\mu_k, \quad \text{for all} \quad \varphi \in A_k. \]

(6.4)
6.2. Comparison of Convergence II

Proof. Let \( k \in \mathbb{N} \) and \( \varphi(x) = \varphi_1(x_1) \cdots \varphi_k(x_k) \in A_k \) be arbitrary. It holds that

\[
\left| \int \varphi \, d\mu_k^N - \int \varphi \, d\mu_k \right| \leq \left| \int \varphi_1(x_1) \cdots \varphi_k(x_k) f_k^N(x) \, dx - \mathbb{E} \left[ \prod_{i=1}^{k} \varphi_i \, d\xi_N \right] \right| + \left| \mathbb{E} \left[ \prod_{i=1}^{k} \varphi_i \, d\xi_N \right] - \int \varphi_1(x_1) \cdots \varphi_k(x_k) g_k(x) \, dx \right| = T_1^N + T_2^N.
\]

For any choice of \( k \in \mathbb{N} \) and \( N \in \mathbb{N} \), where \( k \leq N \), let \( T_{k,N} \) denote the class of all functions from \( \{1, \ldots, k\} \) to \( \{1, \ldots, N\} \). Define a subclass of \( T_{k,N} \) by

\[
\hat{T}_{k,N} = \{ \sigma \in T_{k,N} : \sigma(i) \neq \sigma(j), \text{ for all } i \neq j \},
\]

that is, all injective functions from \( \{1, \ldots, k\} \) to \( \{1, \ldots, N\} \). For any choice of \( \sigma \in \hat{T}_{k,N} \), we have that \( (X^N_{\sigma(1)}, \ldots, X^N_{\sigma(k)}) \) is a random vector in \( D^k \). Since the random vectors \( X^N_1, \ldots, X^N_N \) are identically distributed for any fixed \( N \), the random vectors in

\[
\left\{ (X^N_{\sigma(1)}, \ldots, X^N_{\sigma(k)}) : \sigma \in \hat{T}_{k,N} \right\},
\]

are identically distributed for any fixed \( k \) and \( N \). Using this and that \( \hat{T}_{k,N} \) contains \( \frac{N!}{(N-k)!} \) elements, it follows that

\[
\mathbb{E} \left( \varphi_1(X^N_1) \cdots \varphi_k(X^N_k) \right) = \frac{(N-k)!}{N!} \sum_{\sigma \in \hat{T}_{k,N}} \mathbb{E} \left( \varphi_1(X^N_{\sigma(1)}) \cdots \varphi_k(X^N_{\sigma(k)}) \right).
\]

Also note that

\[
\mathbb{E} \left( \prod_{i=1}^{k} \varphi_i \, d\xi_N \right) = \mathbb{E} \left( \prod_{i=1}^{k} \left( \frac{1}{N} \sum_{j=1}^{N} \varphi_i(X^N_j) \right) \right) = \frac{1}{N^k} \sum_{\sigma \in T_{k,N}} \mathbb{E} \left( \varphi_1(X^N_{\sigma(1)}) \cdots \varphi_k(X^N_{\sigma(k)}) \right).
\]

If \( M = \max\{\|\varphi_1\|_{\infty}, \ldots, \|\varphi_k\|_{\infty}\} \), then for any \( \sigma \in T_{k,N} \),

\[
\left| \mathbb{E} \left( \varphi_1(X^N_{\sigma(1)}), \ldots, \varphi_k(X^N_{\sigma(k)}) \right) \right| \leq \sup_{x \in D^m} |\varphi_1(x_{\sigma(1)}), \ldots, \varphi_k(x_{\sigma(k)})| \leq \|\varphi_1\|_{\infty} \cdots \|\varphi_k\|_{\infty} \leq M^k.
\]

Using this, that \( T_{k,N} \) contains \( N^k \) elements and that \( T_{k,N} \setminus \hat{T}_{k,N} \) contains
$N^k - N!/(N-k)!$ elements, the following is obtained.

$$T_1^N = \left| \mathbb{E} \left( \varphi_1(X_1^N) \cdots \varphi_k(X_k^N) \right) - \mathbb{E} \left( \prod_{i=1}^{k} \left( \frac{1}{N} \sum_{j=1}^{N} \varphi_i(X_i^N) \right) \right) \right|$$

$$= \left| \left( \frac{(N-k)!}{N!} - \frac{1}{N^k} \right) \sum_{\sigma \in \hat{T}_{k,N}} \mathbb{E} \left( \varphi_1(X_{\sigma(1)}^N) \cdots \varphi_k(X_{\sigma(k)}^N) \right) \right|$$

$$\leq \left( \frac{(N-k)!}{N!} - \frac{1}{N^k} \right) \sum_{\sigma \in \hat{T}_{k,N}} M^k + \frac{1}{N^k} \sum_{\sigma \in T_{k,N} \setminus \hat{T}_{k,N}} M^k$$

$$= M^k \left( \frac{(N-k)!}{N!} - \frac{1}{N^k} \right) \frac{N!}{(N-k)!} + M^k \frac{1}{(N-k)!}$$

$$= 2M^k \left( 1 - \frac{N!}{N^k(N-k)!} \right) \to 0, \text{ as } N \to \infty.$$  

To prove that $T_2^N \to 0$ as $N \to \infty$, define the mapping $f : \mathcal{P}(S) \to \mathbb{R}$ by

$$f(P) = \prod_{i=1}^{k} \int_{D} \varphi_i \, dP.$$  

Note that it follows from Fubini’s theorem that

$$\int_{D^k} \varphi_1(x_1) \cdots \varphi_k(x_k) g(x) \, dx = \prod_{i=1}^{k} \int_{D} \varphi_i(x_i) g(x_i) \, dx_i = \prod_{i=1}^{k} \int_{D} \varphi_i \, d\mu.$$  

Using this,

$$T_2^N = \left| \mathbb{E} \left( \prod_{i=1}^{k} \int_{D} \varphi_i \, d\xi_N \right) - \prod_{i=1}^{k} \int_{D} \varphi_i \, d\mu \right| = \left| \mathbb{E}(f(\xi_N)) - \mathbb{E}(f(\mu)) \right| \to 0,$$

as $N \to \infty$, since $\xi_N \overset{d}{\to} \mu$ and $f \in C_{b}(\mathbb{D}^k)$.  

In the next lemma, the sets $A_k$ are defined as in lemma 6.3.

**Lemma 6.4.** For any $k \in \mathbb{N}$, the set

$$B_k = \left\{ \sum_{i=1}^{m} \alpha_i \varphi_i : m \in \mathbb{N}, \quad \alpha_1, \ldots, \alpha_m \in \mathbb{R}, \quad \varphi_1, \ldots, \varphi_m \in A_k \right\}$$

is dense in $C(\mathbb{D}^k)$.

**Proof.** Note that $\mathbb{D}^k$ is a closed and bounded subset of $\mathbb{R}^k$, so by theorem 2.3 it is compact. First we show that $B_k$, which is a subset of $C(\mathbb{D}^k)$, satisfies the three conditions stated in the corollary of the Stone-Weierstrass theorem.
(corollary 2.5). First assume that \( f \) and \( g \) are arbitrary functions in \( B_k \). Then \( f + g \in B_k \) and \( fg \) will be a finite linear combination with terms of the form
\[
c \cdot \varphi \varphi', \quad \text{where } c \in \mathbb{R} \text{ and } \varphi, \varphi' \in A_k.
\]
It follows that \( f g \in B_k \) since \( \varphi \varphi' \in A_k \). The function in \( B \) obtained by taking \( m = 1, \alpha_1 = 1 \) and \( \varphi^1 = 1 \) is the constant function on \( \overline{D} \), so also the second condition is satisfied. To prove that \( B_k \) is separating, let \( x = (x_1, \ldots, x_k) \) and \( y = (y_1, \ldots, y_k) \) be two arbitrary points in \( \overline{D} \) such that \( x \neq y \). Choose \( i \in \{1, \ldots, k\} \) such that \( x_i \neq y_i \). Then there exists \( \psi \in C_b(\overline{D}) \) such that \( \psi(x_i) = 1 \) and \( \psi(y_i) = 0 \). Define
\[
\psi'(x) = \psi(x_i), \quad \text{for all } x \in \overline{D}.
\]
Then \( \psi' \in B_k \) and \( \psi'(x) \neq \psi'(y) \), so it follows that \( B_k \) separates the points of \( \overline{D} \). Now it follows from corollary 2.5 that \( B_k \) is dense in \( C_b(\overline{D}) \)

**Proof of theorem 6.2.** Since \( D^k \) has no Lebesgue measure on the boundary, it is sufficient to prove that \( \mu^w_k \to \mu_k \) on \( \overline{D} \) for all \( k \in \mathbb{N} \), that is,
\[
D^N_k(\phi) = \left| \int_{\overline{D}^c} \phi(x) \tilde{f}^N_k(x) \, dx - \int_{\overline{D}^c} \phi(x) g_k(x) \, dx \right|
\]
tends to zero as \( N \to \infty \) for all \( k \in \mathbb{N} \) and \( \phi \in C_b(\overline{D}) \). To do this, let \( \phi \in C_b(\overline{D}) \) and \( \varepsilon > 0 \) be arbitrary. It follows from lemma 6.4 that we can choose \( \varphi = \sum_{i=1}^m \alpha_i \varphi_i \in B_k \) such that
\[
||\phi - \varphi||_\infty < \frac{\varepsilon}{3}.
\]
Let \( s = \sum_{i=1}^m |\alpha_i| \). Since (6.4) holds, there exists \( n_1, \ldots, n_m \in \mathbb{N} \) such that for \( i = 1, \ldots, m \) it holds that
\[
\left| \int_{\overline{D}^c} \varphi_i(x) \tilde{f}^N_k(x) \, dx - \int_{\overline{D}^c} \varphi_i(x) g_k(x) \, dx \right| < \frac{\varepsilon}{3s}, \quad \text{for all } N \geq n_i.
\]
Using this,
\[
D^N_k(\phi) = \left| \int_{D^c} (\phi(x) - \varphi(x)) \tilde{f}^N_k(x) \, dx \right|
\]
\[
+ \int_{\overline{D}^c} (\phi(x) - \varphi(x)) \, g_k(x) \, dx
\]
\[
- \left( \int_{\overline{D}^c} \varphi(x) \tilde{f}^N_k(x) \, dx - \int_{\overline{D}^c} \varphi(x) g_k(x) \, dx \right)
\]
\[
\leq 2 ||\phi - \varphi||_\infty + \sum_{i=1}^m |\alpha_i| \left( \int_{\overline{D}^c} \varphi_i(x) \tilde{f}^N_k(x) \, dx - \int_{\overline{D}^c} \varphi_i(x) g_k(x) \, dx \right)
\]
\[
< \frac{2\varepsilon}{3} + \sum_{i=1}^m |\alpha_i| \left| \int_{\overline{D}^c} \varphi_i(x) \tilde{f}^N_k(x) \, dx - \int_{\overline{D}^c} \varphi_i(x) g_k(x) \, dx \right|
\]
\[
= \frac{2\varepsilon}{3} + \frac{\varepsilon}{3s} \sum_{i=1}^m |\alpha_i| = \varepsilon, \quad \text{for all } N \geq n_0,
\]
where \( n_0 = \max\{n_1, \ldots, n_m\} \). □
6.3 Speed of Convergence

Using the result stated in theorem 5.3, bounds of the limits in section 6.2 can be obtained. To simplify the notation, first set

\[ Y_N = \int \phi \, d\xi_N \quad \text{and} \quad y = \int \phi \, d\mu. \]

Using the notation and the calculations from the proof of theorem 6.1,

\[ \mathbb{E}(|Y_N - y|^2) = \mathbb{E}(S_N^1) + \mathbb{E}(S_N^2), \]

where

\[ \mathbb{E}(S_N^1) \leq \frac{2\|\phi\|_\infty^2}{N}, \quad (6.6) \]

and

\[ \mathbb{E}(S_N^2) \leq \frac{N(N-1)\|\phi\|_\infty^2}{N^2} \left( \|\tilde{f}_N^2 - g_2\|_{L^1} + 2\|\tilde{f}_N^1 - g_1\|_{L^1} \right). \]

Applying theorem 5.3 to the last equation yields

\[ \mathbb{E}(S_N^2) \leq \frac{N(N-1)\|\phi\|_\infty^2}{N^2} \left( \frac{c_2^2}{N} + \frac{2c_2}{N} \right) \]

\[ = \left( 1 - \frac{1}{N} \right) \left( \frac{c_2^2 + 2c_2}{N} \right) \]

\[ \leq \frac{(c_2^2 + 2c_2)\|\phi\|_\infty^2}{N}, \quad \text{for all } N > 2. \quad (6.7) \]

Putting (6.6) and (6.7) together gives

\[ \mathbb{E}(|Y_N - y|^2) \leq \frac{(c_2^2 + 2c_2 + 2)\|\phi\|_\infty^2}{N} = \frac{C(\phi)}{N}. \quad (6.8) \]

Using all this and equation (6.2) from the proof of theorem 6.1, the following bound is obtained.

\[ \mathbb{P} \left\{ \left| \int \phi \, d\xi_N - \int \phi \, d\mu \right| > \varepsilon \right\} \leq \frac{C(\phi)}{\varepsilon^2 N}, \quad \text{for every } \varepsilon > 0. \]

The problem here is that \( \varepsilon \) can be arbitrarily small. This motivates to instead study the difference between \( \mathbb{E}(Y_N) \) and \( y \), for which the following result holds.

**Theorem 6.5.** Suppose that for \( k \in \{1, 2\} \), it holds that

\[ \|\tilde{f}_k^N - g_k\|_{L^1} \leq \frac{c_k}{N}, \quad \text{for all } N > k, \]

where \( c_1 \) and \( c_2 \) are constants. Then for any \( \phi \in C_0(D) \) and \( \delta \in (0, \frac{1}{2}) \), there exists \( C(\phi, \delta) > 0 \) such that

\[ \left| \mathbb{E} \left( \int \phi \, d\xi_N \right) - \int \phi \, d\mu \right| \leq \frac{C(\phi, \delta)}{N^{1/2 - \delta}}, \quad \text{for all } N > 2. \]
6.3. Speed of Convergence

Proof. Let $Y_N$ and $y$ be defined as above. By theorem 2.17,

$$|E(Y_N) - y| \leq E(|Y_N - y|) = \int_0^\infty \mathbb{P}(|Y_N - y| > t) \, dt.$$  

Let $\delta \in (0, \frac{1}{2})$ be arbitrary. Using Chebychev’s inequality with $p = 1 - 2\delta$, theorem 2.19 with $p = 2/(1 - 2\delta)$ and equation (6.8), the following is obtained.

$$\int_0^1 \mathbb{P}(|Y_N - y| > t) \, dt \leq \int_0^1 \mathbb{E}(|Y_N - y|^{1-2\delta})^{(1-2\delta)/2} \, dt$$

$$\leq \frac{1}{2\delta} \left( \frac{m\|\phi\|_\infty^2}{N} \right)^{1/2 - \delta},$$

where $m$ is a constant that depends on $c_1$ and $c_2$. Applying Chebychev’s inequality with $p = 2$ and again using equation (6.8) gives

$$\int_1^\infty \mathbb{P}(|Y_N - y| > t) \, dt \leq \frac{m\|\phi\|_\infty^2}{N}.$$  

Putting this together yields

$$|E(Y_N) - y| \leq \frac{1}{2\delta} \left( \frac{m\|\phi\|_\infty^2}{N} \right)^{1/2 - \delta} + \frac{m\|\phi\|_\infty^2}{N}$$

$$\leq \frac{1}{2\delta} \frac{m\|\phi\|_\infty^2}{N^{1/2 - \delta}} + \frac{m\|\phi\|_\infty^2}{N^{1/2 - \beta}}$$

$$= C(\phi, \delta) N^{1/2 - \beta},$$

for all $N > 2$.

Now the focus is turned to the case when the order of the convergence presented in section 6.1 is known. In this case, the following result holds.

**Theorem 6.6.** Let $k \in \mathbb{N}$ and $\phi \in C_b(D^k)$. Suppose that

$$|E(\phi(\xi^N)) - E(\phi(\mu))| \leq \frac{C_1(\phi)^k}{N^\alpha},$$

for all $N \geq k$, where $\alpha > 0$ is a constant. Then

$$\left| \int_{D^k} \phi \, d\mu^N - \int_{D^k} \phi \, d\mu \right| \leq \frac{C_2(\phi)^k}{N^\beta},$$

for all $N \geq k$, where $\beta = \min\{\alpha, 1\}$.

To prove this, the following lemma will be used.

**Lemma 6.7.** If $k \in \mathbb{N}$ is fixed, then

$$1 - \frac{N!}{N^k(N-k)!} \leq \frac{k(k-1)}{2N},$$

for all $N \geq k$.  

Proof. The statement is true for \( k = 1 \), because then both sides are equal to zero. Suppose that the relation is true for \( k = m \), where \( m \in \mathbb{N} \). Then, for any \( N \geq m + 1 \),

\[
1 - \frac{N!}{N^{m+1}(N-m-1)!} = 1 - \frac{N-m}{N} \cdot \frac{N!}{N^m(N-m)!} \\
\leq 1 - \left(1 - \frac{m}{N}\right) \left(1 - \frac{m(m-1)}{2N}\right) \\
= \frac{m(m-1)}{2N} + \frac{m}{N} - \frac{m^2(m-1)}{2N^2} \\
\leq \frac{m(m-1) + 2m}{2N} = \frac{(m+1)m}{2N}.
\]

The result now follows by induction. \( \square \)

**Proof of theorem 6.6.** Let \( \varphi \in A_k \), where \( A_k \) is defined as in equation (6.3). Using the calculations from the proof of lemma 6.3, it follows that

\[
\left| \int_{D_k^N} \varphi d\mu_k^N - \int_{D_k} \varphi d\mu_k \right| \leq T_1^N + T_2^N,
\]

where

\[
T_1^N \leq 2M^k \left(1 - \frac{N!}{N^k(N-k)!}\right)
\]

and

\[
T_2^N \leq |E(f(\xi_N)) - E(f(\mu))|.
\]

The mapping \( f \) was defined in equation (6.5). By lemma 6.7 and the assumption in the theorem,

\[
\left| \int_{D_k^N} \varphi d\mu_k^N - \int_{D_k} \varphi d\mu_k \right| \leq \frac{M^k k(k-1)}{N^\alpha} + \frac{C_1(f)^k}{N^\beta} \\
\leq \frac{(2M)^k + C_1(f)^k}{N^\beta} \\
\leq \frac{(2M + C_1(f))^k}{N^\beta}.
\]

Now let \( \phi \in C_b(D_k^N) \). By the same arguments as in the proof of theorem 6.2, it follows that there exists functions \( \varphi^1, \ldots, \varphi^m \in A_k \) such that for every \( \varepsilon > 0 \),

\[
\left| \int_{D_k} \phi d\mu_k^N - \int_{D_k^N} \phi d\mu_k \right| \leq \varepsilon + \sum_{i=1}^m |\alpha_i| \left| \int_{D_k} \varphi^i d\mu_k^N - \int_{D_k} \varphi^i d\mu_k \right|.
\]

Then using that \( D^k \) has no measure on the boundary, equation (6.9) and that \( \varepsilon > 0 \) was arbitrary, the following is obtained.

\[
\left| \int_{D_k} \phi d\mu_k^N - \int_{D_k} \phi d\mu_k \right| \leq \frac{(2M + C_1(f))^k}{N^\beta} \sum_{i=1}^m |\alpha_i| \leq \frac{C_2(\phi)^k}{N^\beta}.
\]

The constant \( C_2 \) depends on \( \phi \) since the choices of \( \alpha_i, \varphi^i, f \) and \( M \) only depend on \( \phi \). \( \square \)
Bibliography


Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years from the date of publication barring exceptional circumstances. The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtidiga ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår. Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art. Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skyld mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är krävande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

© 2010, Mikael Petersson