University of Alberta

Extensions of Skorohod's almost sure representation theorem

by

Nancy Hernandez Ceron

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Mathematics

Department of Mathematical and Statistical Sciences

Fall 2010
Edmonton, Alberta

Examining Committee

Norman C. Beaulieu, Department of Electrical and Computer Engineering.

Alexander Litvak, Department of Mathematical and Statistical Sciences.

Byron Schmuland, Department of Mathematical and Statistical Sciences.

Abstract

A well known result in probability is that convergence almost surely (a.s.) of a sequence of random elements implies weak convergence of their laws. The Ukrainian mathematician Anatoliy Volodymyrovych Skorohod proved the lemma known as *Skorohod's a.s. representation Theorem,* a partial converse of this result.

In this work we discuss the notion of continuous representations, which allows us to provide generalizations of Skorohod's Theorem. In Chapter 2, we explore Blackwell and Dubins's extension [3] and Fernique's extension [10].

In Chapter 3 we present Cortissoz's result [5], a variant of Skorokhod's Theorem. It is shown that given a continuous path in $\mathcal{M}(S)$ it can be associated a continuous path with fixed endpoints in the space of S-valued random elements on a nonatomic probability space, endowed with the topology of convergence in probability.

In Chapter 4 we modify Blackwell and Dubins representation for particular cases of S, such as certain subsets of $\mathbb{R}$ or $\mathbb{R}^n$.

Acknowledgments

I would like to thank all people who have helped me, this research project would not have been possible without their support.

My deepest gratitude goes to my supervisor Dr. Byron Schmuland, his wide knowledge, invaluable guidance and support enabled me to develop an understanding of the subject. Always willing to explain my questions patiently, discuss and make interesting explorations in my work, he motivated me throughout this work. The energy and enthusiasm for research I learned from him is my lifetime benefit.

Deep gratitude is also due to the members of the supervisory committee, Dr. Norman C. Beaulieu and Dr. Alexander Litvak without whose knowledge and assistance this study would not have been successful.

The financial support provided by the department of Mathematical and Statistical Sciences at the University of Alberta is gratefully acknowledged.

During my time in Edmonton, I was delighted to interact with professors, students, classmates and friends. I wish to thank them for providing a stimulating environment for me to learn and grow, a home away from home. Their emotional support and kind encouragement will be always appreciated. I am especially grateful to Jonathan Chavez for helping me get through the difficult times, his love and the constructive advices in life and mathematics he gave me deserve special mention.

Finally, I wish to thank my parents, Virginia Ceron and Victor Hernandez, for giving me their constant love and support throughout my life. They worked industriously and spare no effort to provide the best for our family. To my brothers, Omar and Ivan, I want to express my love and gratitude, I feel proud of them and I just want them to feel the same way towards me. My apologies to my extended family, aunts, uncles, grandparents and cousins for not mentioning their names. I have no suitable words to describe my everlasting love to them. To them I dedicate this thesis.

Por último, quiero dar las gracias a mis padres, Virginia Cerón y Víctor Hernández, por darme su amor y apoyo constante a lo largo de mi vida. Ellos han trabajado arduamente, y sin escatimar esfuerzos han dado lo mejor a nuestra familia. A mis hermanos, Omar e Iván, les quiero expresar mi amor y gratitud, me siento muy orgullosa de ellos y mi único deseo es que sientan lo mismo hacia mí. Mis disculpas a mi familia, tías, tíos, abuelos y primos por no mencionar sus nombres. Es difícil encontrar las palabras adecuadas para describir mi amor eterno hacia ellos. A ellos les dedico esta tesis.

Nancy Hernandez-Ceron, June 2010.

.

Contents

Chapter 1

Preliminaries

In this chapter we recall some of the basic tools and results needed in probability theory. We recommend [1], [7] and [12] for further details in these topics.

1.1 Topology and Metric Spaces

1.1.1 General Topology

One of the most important concepts in this work is the concept of convergence. Because of that, we include a section to discuss the basic notions of general topology. The easiest and most common example of convergence is the convergence of real numbers. Intuitively, one can imagine a sequence of numbers, getting closer and closer to another number. But, if we are talking about probability measures, how can we define convergence in such spaces? For that reason we need a more general definition of convergence, continuity and open sets.

Let X be a non empty set, a class $\mathscr{T}$ of subsets of X is called a **topology** if

a) $\emptyset, X$ are elements of $\mathscr{T}$,

b) for any $U, V \in \mathscr{T}$ we have $U \cap V \in \mathscr{T}$, and

c) for any $\mathscr{U} \subset \mathscr{T}$, $\bigcup_{U \in \mathscr{U}} U \in \mathscr{T}$.

The elements of $\mathscr{T}$ are called **open** sets and their complements are called **closed** sets. The pair $(X, \mathscr{T})$ is a **topological space.** The **interior** and **closure** of a set A, $A \subset X$, are defined as follows

$$\begin{aligned} \operatorname{int} A &:= \bigcup \{U \in \mathscr{T} : U \subset A\}, \\ \overline{A} &:= \bigcap \{V : V \text{ is closed and } A \subset V\}, \end{aligned}$$

and the **boundary** of A is the set $\partial A = \overline{A} \setminus \operatorname{int} A$.

Let x be an element of X and $V \subset X$, V is a **neighborhood** of x if $x \in \operatorname{int} V$, in other words, if there exists an open set O containing x such that $O \subset V$. A **basis** for a topology $\mathscr{T}$ is any collection $\mathscr{U} \subset \mathscr{T}$ such that, for all $V \in \mathscr{T}$,

$$V = \bigcup \{U \in \mathscr{U} : U \subset V\}.$$

So, any open set can be expressed as the union of open sets in its basis. Indeed, if the collection $\mathscr{U}$ satisfies

1. $\mathscr{U}$ covers X, i.e. $X = \bigcup \{U : U \in \mathscr{U}\}$,
2. If U_1, U_2 are elements of $\mathscr{U}$ with non empty intersection and for each $x \in U_1 \cap U_2$ there exists $U_3 \in \mathscr{U}$ containing x such that $U_3 \subset U_1 \cap U_2$,

then $\mathscr{U}$ is the basis for some topology $\mathscr{T}$.

We say that X has a countable basis at the point x if there is a countable collection $\mathscr{U}_x$ of open sets containing x such that any neighborhood V of x contains at least one element of $\mathscr{U}_x$. A space that has a countable basis at each of its points is said to satisfy the first countability axiom, and it is called a **first countable** topological space. On the other hand, $(X, \mathscr{T})$ satisfies the second countability axiom if it has a countable basis. A space that satisfies this axiom is called a **second countable** space. Clearly, any second countable space is first countable. A subset D of X is **dense** if $\overline{D} = X$. $(X, \mathscr{T})$ is called **separable** if there exists a countable dense subset of X.

A subset V of the topological space X is called **compact** if any open cover (a collection of open sets whose union contains V) has a finite subcover of V. If for every $x \in V$ there exists an open set O containing x such that $O \subset V'$, where $V' \subset V$ is a compact set, then V is called **locally compact**.

Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of elements of X. We say that x_n converges to $x \in X$, in symbols $x_n \to x$, if for every open set U containing x, there exists a $N \in \mathbb{N}$ such that $x_n \in U$, $\forall n \geq N$.

Given two topological spaces $(X, \mathscr{T})$ and $(X', \mathscr{T}')$, a function $f : X \to X'$ is **continuous** if, for every $U' \in \mathscr{T}'$, $f^{-1}(U') \in \mathscr{T}$.

A function $d : X \times X \to [0, \infty)$ is said to be a **metric** if

a) for all $x, y \in X$, $d(x, y) = d(y, x)$,

b) for all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$, and

c) $d(x, y) = 0$ iff $x = y$.

(X, d) is called a **metric space.** Notice that, given a metric space, it is possible to construct a topological space. Denote by $B(x, r)$, $x \in X$, the set

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

and call it the open ball with center at x and radius r. Then, the set

$$\mathscr{U} := \{B(x, r) : x \in X, r > 0\}$$

is a basis of a topology, the topology induced by the metric d. In fact,

$$\mathscr{U}_x := \{B(x, r) : x \in X, r > 0, r \in \mathbb{Q}\}$$

is a countable basis at x. Hence, any metrizable space is first countable. Moreover, if the metric space is separable then it must be second countable. Let $D = \{x_n\}_{n=1}^{\infty}$ be a dense subset of X, then

$$\mathscr{U}^* = \bigcup_{n=1}^{\infty} \mathscr{U}_{x_n}$$

is a countable basis of the topology.

In a metric space (X, d), $x_n \to x$ iff for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that
$$d(x_n, x) < \epsilon, \quad \text{for all } n \geq N.$$
A sequence $\{x_n\}_{n \in \mathbb{N}}$ satisfying
$$\lim_{n \to \infty} \left[\sup_{m \geq n} d(x_n, x_m) \right] = 0$$
is called a **Cauchy sequence**. The metric space (X, d) is **complete** if for every Cauchy sequence $\{x_n\}_{n=1}^{\infty}$ there exists a $x \in X$ such that $x_n \to x$.

In metric spaces some topological concepts seem to be easier to visualize and work with, for example if (X, d) and (X', d') are two metric spaces and $\mathscr{T}$, $\mathscr{T}'$ are the topologies induced by d and d' respectively, a function $f : X \to X'$ is continuous at x if for every $\epsilon > 0$ there exists $\delta > 0$ such that
$$d(x, y) < \delta \;\; \text{implies} \;\; d'(f(x), f(y)) < \epsilon.$$
Then, f is a continuous function if f is continuous at x for all $x \in X$. So, we do not have to worry about the inverse images of all open subsets in X', we just verify this new criteria, which in some cases may be easier to check or simply easier to imagine.

Notice that this definition allows us to have a particular δ for each point x, in other words, δ may depend on x. If there's no such dependence, i.e. if for every ϵ there exists δ such that $d'(f(x), f(y)) < \epsilon$ if $d(x, y) < \delta$ $\forall x, y \in X$, then f is called **uniformly continuous** from (X, d) to (X', d').

Remark: In a metric space the following can be proved

- f is continuous at x iff for every $x_n \to x$ one has $f(x_n) \to f(x)$.
- If $f : X \to X'$ is continuous and V is a compact subset of X, then $f(V)$ is compact in X'.

Compact spaces have desirable properties, that is why there exists some "methods" to enlarge a topological space in such way that the result is a compact space. This process is called compactification. We only study one of this methods, the **Alexandroff one point compactification**. We start with an example: the real line with its usual topology. We know that every closed bounded subset of $\mathbb{R}$ is compact. When we have an unbounded set it is possible to construct a divergent sequence, a sequence that somehow, escapes from the real line. So, naturally one may think about a solution: adding two new points ∞ and $-\infty$. However, it may be enough to add only the point ∞ and make any sequence with unbounded absolute values go to this point.

In general, for a topological space $(X, \mathscr{T})$ the one point compactification X' of X is obtained by adding one extra point called infinity and denoted by ∞. Then, define a new topology $\mathscr{T}'$ as follows

$$\mathscr{T}' := \mathscr{T} \cup \{O \cup \{\infty\} : O \in \mathscr{T},\ X \backslash O \text{ is closed and compact}\}.$$

It can be easily verified that $\mathscr{T}'$ is indeed a topology and that $(X', \mathscr{T}')$ is compact.

1.1.2 Order Topology

To simplify some of the proofs in Section 2.3 we present some basic definitions and facts about a particular topological space that can be defined in a totally ordered space: the order topology. We also discuss ordinal numbers and ordinal spaces but we only work with countable ordinals.

Let A be a non empty set. A relation $C \subset A \times A$ is said to be a **total order relation** on A if

1. (Comparability) For all $x, y \in A$ such that $x \neq y$ either xCy or yCx.
2. (Non reflexibility) For no $x \in A$ does xCx hold.
3. (Transitivity) If xCy and yCz then xCz.

If such relation exists A is called an ordered set with respect to C. Commonly, we denote C by "<" and we say that x is smaller than y if $x < y$. A subset B of A is bounded if there exists $a \in A$ such that $b < a$ holds for all $b \in B$. This element a of A is called an upper bound of B. The set A has the **least upper bound property** if for every bounded nonempty subset of A there exists an upper bound $a \in A$ such that if c is another upper bound of the set, then $a < c$. If every nonempty subset of A has a smallest element, then A is a **well ordered set**.

Some examples of totally ordered sets are $\mathbb{R}$ and $\mathbb{Q}$ with the usual "$<$". Any finite set is well ordered, indeed $\mathbb{N} = \{1, 2, \ldots\}$ and any subset of it is well ordered.

In a totally ordered set $(X, <)$ the following subsets of X are called intervals

$$\begin{aligned}(a, b) &:= \{x \in X : a < x < b\} \\ (a, \infty) &:= \{x \in X : a < x\} \\ (-\infty, a) &:= \{x \in X : x < a\}.\end{aligned}$$

If $(a, b) = \emptyset$, then a is the predecessor of b.

Consider the class $\mathscr{I}$ of all such subsets of X, it can be easily verified that $\mathscr{I}$ satisfies all the properties to be a basis of a topology in X. The topology generated in this way is called the **order topology** of X, with respect to $<$. Standard topologies in $\mathbb{R}, \mathbb{N}, \mathbb{Q}$ and $\mathbb{Z}$ are order topologies.

Lemma 1.1.1 *Let X be a totally ordered set having the least upper bound property. In the order topology, each closed interval in X is compact.*

1.1.3 Additional Topics

In this section we discuss the concept of ordinals number, which in plain words is a generalization of the natural numbers. We also introduce the functions called partition of unity. Both topics are a key factor for Fernique's extension of Skorohod's a.s. representation Theorem (Section 2.3).

Formally an ordinal number is a well ordered set. In fact every well ordered set is uniquely order isomorphic to an ordinal number. A standard definition says that a set S is an ordinal if and only if S is well ordered (with respect to set membership) and every element of S is also a subset of S. The first infinite ordinal is ω, the set of all natural numbers. There are three usual operations defined on ordinals: addition, multiplication, and exponentiation. Let's define the **addition**. The union of two disjoint well ordered sets S and T is also well ordered if we define the following order relation: the well ordered set S is written "to the left" of the well ordered set T and every element of S is smaller than every element of T. The sets S and T themselves keep the ordering they already have. If the ordinal numbers are finite then we get, somehow, the usual addition in $\mathbb{N}$. If we try to visualize the ordinal $\omega + \omega$, two copies of the natural numbers ordered as described before, this is what we get:

$$0 < 1 < 2 < 3 < \cdots < 0' < 1' < 2' < \cdots$$

This is different from ω because in ω only 0 does not have a direct predecessor while in $\omega + \omega$ the two elements 0 and $0'$ do not have direct predecessors. On the other hand $3 + \omega = \omega$:

$$0 < 1 < 2 < 0' < 1' < 2' < \cdots,$$

while $\omega + 3 \neq \omega$:

$$0 < 1 < 2 < \cdots < 0' < 1' < 2'.$$

The **product** of two ordinal numbers is given by its cartesian product. $S \times T$ is a well ordered set with an order that puts the least significant position first. This is $(s_1, t_1) < (s_2, t_2)$ iff $t_1 < t_2$ (in T) OR if $t_1 = t_2$ and $s_1 < s_2$. This operation is associative and generalizes the multiplication of natural numbers. For example $\omega \cdot 2$ would look like

$$(0,0) < (1,0) < (2,0) < \cdots < (0,1) < (1,1) < (2,1) < \cdots$$

Notice that $\omega \cdot 2 = \omega + \omega$. On the other hand $2 \cdot \omega = \omega$:

$$(0,0) < (1,0) < (0,1) < (1,1) < \cdots < (0,k) < (1,k) < \cdots$$

Finally we only define **exponentiation** of well ordered sets when the exponent is a finite set. In this case,

$$\alpha^n = \underbrace{\alpha \cdot \alpha \cdots \alpha}_{n \text{ times}},$$

and the power is the product of iterated multiplication. For instance, $\omega^2 = \omega \cdot \omega$.

Let λ be an ordinal number, one can consider the spaces

$$[0, \lambda) = \{\alpha : \alpha < \lambda\} \text{ and } [0, \lambda] = \{\alpha : \alpha \leq \lambda\}$$

endowed with the order topology. Such spaces are called **ordinal spaces**. If λ is a finite ordinal then the spaces $[0, \lambda)$ and $[0, \lambda]$ are discrete spaces. When λ is a limit ordinal (it has no predecessor) $[0, \lambda]$ is the one point compactification of $[0, \lambda)$. Another way to see that $[0, \lambda]$ is a compact set, is using Lemma 1.1.1.

Let X be an ordinal space, a function $F : X \to Y$ can be seen as an ordinal indexed Y-valued sequence. F is continuous iff for any limit ordinal $\lambda \in X$, $F(\lambda)$ takes the value of the limit of $\{F(\alpha),\ \alpha < \lambda\}$.

Now we introduce the concept of partition of unity. As usual, let $(X, \mathscr{T})$ be a topological space. A collection of functions $\{\varphi_n\}_{n=1}^{\infty}$ is said to be a **partition of unity** if

1. for all $n = 1, 2, \ldots$, the function $\varphi_n : X \to [0, 1]$ is continuous, and
2. for all $x \in X$ one has $\sum\limits_{n=1}^{\infty} \varphi_n(x) = 1$.

The partition is **locally finite** if for every $x \in X$ there exists a neighborhood V_x of x such that

$$\{n \in \mathbb{N} : V_x \cap \operatorname{supp}(\varphi_n) \neq \emptyset\} \text{ is finite,}$$

where $\operatorname{supp}(\varphi)$, the support of φ, is the closed set

$$\overline{\{x \in X : \varphi(x) \neq 0\}}.$$

Let $\{U_n\}_{n=1}^{\infty}$ be a countable collection of open sets covering X, if $\operatorname{supp}(\varphi_n) \subset U_n$ for every n, then the partition is **dominated by** the cover $\{U_n\}_{n=1}^{\infty}$. The existence of such functions is guaranteed for metric spaces (see [9]).

Theorem 1.1.2 *Let $(X, \mathscr{T})$ be a metrizable topological space. Let $\{U_n\}_{n=1}^{\infty}$ be a countable open cover of X, then there exists a locally finite continuous partition of unity $\{\varphi_n\}_{n=1}^{\infty}$ dominated by $\{U_n\}_{n=1}^{\infty}$.*

1.2 Polish Spaces

Polish spaces arose as a topological concept, but they have been used in probability theory as a natural generalization of spaces such as the real line or the interval (0,1). If the topological space $(S, \mathscr{T})$ is metrizable, separable and complete (with respect to this metric), then it is called a **Polish space**. Notice that a Polish space is both first and second countable. The term "Polish space" was first introduced by French mathematicians and illustrates the work of Polish topologists and logicians like Sierpiński, Kuratowski and Tarski, who extensively studied this kind of structures.

Example Any finite or countable set with the discrete topology is a Polish space.

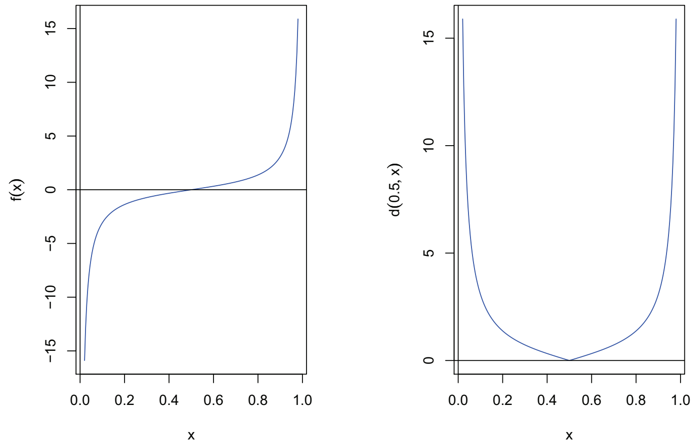
Example The real line with its usual topology $\mathscr{T}$ is a Polish space, since the euclidean metric d generates the topology and $(\mathbb{R}, d)$ is complete. Notice that $\mathbb{Q}$ is a dense subset of $\mathbb{R}$, i.e. $(\mathbb{R}, \mathscr{T})$ is separable.

A topological space $(S, \mathscr{T})$ may be generated by more than one metric. If for one of these metrics $(S, \mathscr{T})$ is not complete, it does not mean that we cannot find another metric for which the condition of completeness is satisfied. For instance, the interval $(0, 1)$ with its usual topology is not complete for the euclidean metric. However, before we disqualify it as a Polish space we provide the following example.

Example The open interval (0,1) with its usual topology is a Polish space. Consider the functions $f : (0, 1) \to \mathbb{R}$ and $d : (0, 1) \times (0, 1) \to [0, \infty)$, given by

$$f(x) = \tan\left(\pi x - \frac{\pi}{2}\right) \ , \quad d(x, y) = |f(x) - f(y)|$$

The first graphic below shows how the function f behaves in $(0, 1)$ and the second is the graphic of $d\left(\frac{1}{2}, x\right)$, $x \in (0, 1)$. Clearly d is a metric on $(0, 1)$.



Moreover, if $r > 0$

$$\{y : d(x, y) < r\} = \{y : |f(x) - f(y)| < r\} = (a, b),$$

where $a = f^{-1}(f(x) - r)$ and $b = f^{-1}(f(x) + r)$. Therefore, d generates the usual topology in the unit interval. In addition, $(0, 1)$ is separable and complete with respect to this metric (there is no Cauchy sequence converging to 0 or 1).

Now, we give some examples of topological spaces which are not Polish spaces. The statement "$(S, \mathscr{T})$ is not a Polish space" is quite strong. To say this, we must show that for any metric d generating $\mathscr{T}$, (S, d) is not complete. On the other hand, separability does not depend on the metric, it is enough to find a countable $D \subset S$ such that $\overline{D} = S$.

Example $\mathbb{Q}$ with its usual topology $\mathscr{T}_{\mathbb{Q}}$ is not a Polish space. The Baire category Theorem states that in a complete metric space a set with non empty interior is of second category (it cannot be written as a countable union of nowhere dense sets). A set E is nowhere dense if $\operatorname{int} \overline{E} = \emptyset$. In this example,

$$\mathbb{Q} = \bigcup_{n=1}^{\infty} \{q_n\},$$

where (in the usual topology) $\operatorname{int}\overline{\{q_n\}} = \operatorname{int}\{q_n\} = \emptyset$, while $\operatorname{int}\mathbb{Q} = \mathbb{Q}$. It follows that $\mathbb{Q}$ is of first category, which contradicts the Baire Category Theorem. Therefore $(\mathbb{Q}, \mathscr{T}_{\mathbb{Q}})$ is not a Polish space.

Example *Upper limit topology.* This is a very interesting topology defined in the set $\mathbb{R}$ of real numbers. This topology is generated by

$$\mathscr{U} = \{(a, b] : a, b \in \mathbb{R}, a \leq b\}.$$

Clearly $\mathscr{U}$ covers $\mathbb{R}$. If U_1, U_2 are non empty elements of $\mathscr{U}$, say $U_1 = (a_1, b_1]$ and $U_2 = (a_2, b_2]$ then $U_1 \cap U_2 = (\max(a_1, a_2), \min(b_1, b_2)] \in \mathscr{U}$. Hence $\mathscr{U}$ is indeed the basis of some topology, which we denote by $\mathscr{T}_u$. Since

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right]$$

the upper topology is finer than the usual topology. $\mathscr{T}_u$ is disconnected, because the intervals of the form $(a, b]$ are both open and closed sets. To see this, notice that

$$(a, b] = ((-\infty, a] \cup (b, \infty))^c, \quad \text{and}$$

$$(-\infty, a] = \bigcup_{n=1}^{\infty}(-n, a], \quad (b, -\infty) = \bigcup_{n=1}^{\infty}(b, n].$$

It can be proven that the upper limit topology is first countable and separable (indeed $\mathbb{Q}$ is dense in this topology) but not second countable. Therefore, $\mathscr{T}_u$ is not metrizable, thus $(\mathbb{R}, \mathscr{T}_u)$ is not a Polish space.

1.3 Probability

In this section we include some results in probability theory that may be used in the next sections.

1.3.1 Measurable Functions

Let $\mathcal{F}$ be a σ-algebra of subsets of Ω, then $(\Omega, \mathcal{F})$ is said to be a **measurable space**. If $(X, \mathscr{T})$ is a topological space, in order to obtain a measurable space, we define $\mathcal{B}(X) = \sigma(\mathscr{T})$, the σ-algebra generated by $\mathscr{T}$. $\mathcal{B}(X)$ is called the **Borel σ-algebra** of X.

Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces, a function $f : \Omega \to \Omega'$ is a measurable function if

$$f^{-1}(\mathcal{F}') = \{f^{-1}(B) : B \in \mathcal{F}'\} \subset \mathcal{F}.$$

Moreover, if the σ-algebra $\mathcal{F}'$ is generated by the class of sets $\mathcal{A}$, i.e. $\mathcal{F}' = \sigma(\mathcal{A})$, then

$$f^{-1}(\mathcal{A}) \subset \mathcal{F}$$

implies the measurability of f. As a consequence, we obtain the next lemma

Lemma 1.3.1 *Let (S, d) be a metric space. The function $f : (\Omega, \mathcal{F}) \to (S, \mathcal{B}(S))$ is measurable iff*

$$f^{-1}(B) \in \mathcal{F}$$

for all open balls B.

Example A widely used probability space is $(U, \mathcal{B}, \lambda)$ where $U = (0, 1)$, $\mathcal{B}$ the Borel sets, and λ is the Lebesgue measure. We will refer to this as the **Lebesgue probability space**.

We state, without proof, a very useful result in probability: the Borel Cantelli Lemma. It will be used in Section 2.1.

Lemma 1.3.2 (Borel Cantelli Lemma) *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of measurable sets and μ a probability measure. If*

$$\sum_{n=1}^{\infty} \mu(A_n) < \infty,$$

then

$$\mu\left(\limsup_{n\to\infty} A_n\right) = 0, \quad \textit{which is the same as} \quad \mu\left(\liminf_{n\to\infty} A_n^c\right) = 1,$$

where

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \quad \textit{and} \quad \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i.$$

Sometimes we initially work with the two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, but become interested in the product space $\Omega \times \Omega'$. A way to define measurable subsets of this product space is defining $\mathcal{F} \otimes \mathcal{F}'$, the σ-algebra generated by the class of sets of the form $A \times B$, $A \in \mathcal{F}$ and $B \in \mathcal{F}'$, which are called rectangles. Then we consider the measurable space $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$. If f is a function from $\Omega \times \Omega'$ to a third measurable space $(\Omega^*, \mathcal{F}^*)$, f is **jointly measurable** if

$$f^{-1}(\mathcal{F}^*) = \{f^{-1}(B) : B \in \mathcal{F}^*\} \subset \mathcal{F} \otimes \mathcal{F}'.$$

When proving joint measurability for a function, things may get complicated. Luckily for us, there are results that can help us to verify this property, like the next theorem

Theorem 1.3.3 *The map $f : \Omega \times [a, b] \to \mathbb{R}$ is jointly measurable if:*

1. *$f_t : \Omega \to \mathbb{R}$ is measurable for all $t \in [a, b]$, where $f_t(\omega) = f(\omega, t)$.*
2. *$f_\omega : [a, b] \to \mathbb{R}$ continuous for all $\omega \in \Omega$, where $f_\omega(t) = f(\omega, t)$.*

Proof. For each $n \in \mathbb{N}$, let

$$a_{i,n} = a + \frac{i(b-a)}{n}, \quad \text{for all } i = 0, 1, \dots, n.$$

Now, for $t \in [a, b]$, let $\tau_n(t) = \min\{a_{i,n} : a_{i,n} \geq t\}$. Then we have

$$\tau_n(t) \to t \quad \text{as} \quad n \to \infty,$$

so that $f(\omega, \tau_n(t)) \to f(\omega, t)$, by hypothesis. Since the limit of measurable functions is measurable, we only need to show that $f(\omega, \tau_n(t))$ is (jointly) measurable for all n.

Take any $B \in \mathcal{B}(\mathbb{R})$, then the set $\{(\omega, t) : f(\omega, \tau_n(t)) \in B\}$ may be written as

$$(\,\{\omega : f(\omega, a) \in B\} \times \{a\}\,) \cup \left[\bigcup_{i=1}^{n} (\,\{\omega : f(\omega, a_{i,n}) \in B\} \times (a_{i-1,n}, a_{i,n}]\,)\right],$$

which is a union of rectangles in $\mathcal{F} \otimes \mathcal{B}[a, b]$. It follows that $(\omega, t) \mapsto f(\omega, t)$ is a measurable function. ∎

Remark: Clearly the proposition is still true even if we only have right or left continuity and the proof is similar.

Let X be a measurable function from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$. Such function is called a **random element** in Ω'. If $\Omega' = \mathbb{R}$, then X is said to be a **random variable.** If the domain of a random element is endowed with a probability measure P, then X induces a probability measure in its range $(\Omega', \mathcal{F}')$. The **distribution** of X, $P_X : \mathcal{F}' \to [0, 1]$ is the probability measure given by

$$P_X(B) := P \circ X^{-1}(B) = P(X^{-1}(B)).$$

Sometimes P_X is also called the **law** of X. We write $X \overset{D}{=} \mu$ to say that the distribution of X is equal to μ, or $X \overset{D}{=} Y$ meaning that the law of X is equal to the law of Y.

When $\Omega' = \mathbb{R}$ we can exploit some of the properties of the real numbers, like its order structure. In the next section we study the cumulative distribution function of a random variable, which is a function that characterizes uniquely its distribution.

As we saw before, we can relate a probability measure to each random variable, of course this relation is not one to one. In fact, there are infinitely many different random variables with the same distribution. So, given a probability measure in S it is natural to ask ourselves if there exists a S-valued random variable X having this distribution. The next theorems, stated without proof, answer this question. (See Section 1.3.4 for a definition of Borel space.)

Theorem 1.3.4 *(Existence, Borel) For any probability measures $\{\mu_n\}_{n=1}^{\infty}$ on some Borel spaces $S_1, S_2, \dots$, there exist some independent random elements $\{X_n\}_{n=1}^{\infty}$ on the Lebesgue probability space with $X_n \overset{D}{=} \mu_n$ for all n.*

Theorem 1.3.5 *(Existence, infinite product measures, Lomnicki and Ulam) Consider a collection of probability spaces $(S_t, \mathcal{F}_t, \mu_t)$ indexed by some nonempty set T. There exist some (independent) random elements X_t in S_t with distributions μ_t for all $t \in T$.*

The first theorem can be found in [11] (Theorem 2.19, p. 33) as well as the second one (Corollary 5.18, p. 93).

To finish the section we define two types of convergence of random elements, which are the central topic of the following chapters. Let X and $\{X_n\}_{n=1}^{\infty}$ be random elements from the probability space $(\Omega, \mathcal{F}, P)$ to the Polish space S. We say that $\{X_n\}$ **converges almost surely** (a.s.) to X if

$$P\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

On the other hand, $\{X_n\}$ **converges in distribution** to X ($X_n \xrightarrow{D} X$) if for every continuous and bounded function $f : S \to \mathbb{R}$

$$\lim_{n\to\infty} \int_{\Omega} f(X_n)\, dP = \int_{\Omega} f(X)\, dP.$$

In Section 1.3.4 we will see that there is a simplified and equivalent definition for convergence in distribution in the particular case $S = \mathbb{R}$. Also, in Theorem 1.3.19 we show that convergence a.s. is stronger than convergence in distribution.

1.3.2 Pseudoinverse of a c.d.f.

Let X be a random variable, its **cumulative distribution function** (c.d.f.) is the function $F : \mathbb{R} \to [0, 1]$ given by

$$F(x) := P(\{\omega : X(\omega) \leq x\}) = P(X \leq x).$$

Some properties of F are:

- F is nondecreasing and right continuous,
- $\lim_{x\to-\infty} F(x) = 0$ and
- $\lim_{x\to\infty} F(x) = 1$.

Since the cumulative distribution function (c.d.f.) may not be an injective function we are not able to define its inverse function, however, we can define its **pseudoinverse**. The pseudoinverse $F^{(-1)}$ of a c.d.f F is a function from $(0, 1)$ to $\mathbb{R}$ given by

$$F^{(-1)}(u) := \inf\{x : F(x) \geq u\}.$$

This definition can be found in [1] (Theorem 7.7.2, p.334).

Remark: If F is strictly increasing and continuous then $F^{(-1)} = F^{-1}$, the ordinary inverse.

Since F is a right continuous function, the set $\{x : F(x) \geq u\}$ is of the form $[x^*, \infty)$ and $F^{(-1)}(u) = \inf[x^*, \infty) = x^*$. Hence $F^{(-1)}(u)$ belongs to the set $\{x : F(x) \geq u\}$, which implies $u \leq F[F^{(-1)}(u)]$.

Now, suppose that $F^{(-1)}(u) \leq a$, then $F[F^{(-1)}(u)] \leq F(a)$ because F is nondecreasing. By the previous line, we can conclude that $u \leq F(a)$. On the other hand, if $u \leq F(a)$ we must have $F^{(-1)}(u) \leq a$, because a is an element of $\{x : F(x) \geq u\}$. We just proved

Lemma 1.3.6 *$F^{(-1)}(u) \leq a$ iff $u \leq F(a)$.*

Using this fact we can prove the following

Theorem 1.3.7 *Let $F^{(-1)}$ be the pseudoinverse of a distribution function F. Then*

1. *$F^{(-1)}$ is an increasing function.*
2. *$F^{(-1)}$ is left continuous.*

Proof. The first statement is just a consequence of the previous lemma. To prove the second statement let $u \in (0, 1)$ and $\{u_n\}_n$ be any sequence such that $\{u_n\} \uparrow u$, then $\{F^{(-1)}(u_n)\}$ is also an increasing sequence and $F^{(-1)}(u_n) \leq F^{(-1)}(u)$ for all n (because $u_n \leq u$). Then,

$$\lim_{n\to\infty} F^{(-1)}(u_n) \leq F^{(-1)}(u).$$

On the other hand,

$$\lim_{n\to\infty} F[F^{(-1)}(u_n)] \leq F\left(\lim_{n\to\infty} F^{(-1)}(u_n)\right),$$

because $F^{(-1)}(u_n)$ is an increasing sequence and F is an increasing function. Since $u_n \leq F[F^{(-1)}(u_n)]$ we have

$$u = \lim_{n\to\infty} u_n \leq \lim_{n\to\infty} F[F^{(-1)}(u_n)] \leq F\left(\lim_{n\to\infty} F^{(-1)}(u_n)\right).$$

Then, since $F^{(-1)}(u) \le a$ iff $u \le F(a)$, we have

$$F^{(-1)}(u) \le \lim_{n\to\infty} F^{(-1)}(u_n).$$

It follows that

$$F^{(-1)}(u) = \lim_{n\to\infty} F^{(-1)}(u_n),$$

which means that F is left continuous. ∎

In the last example of Section 1.2 we talked about the upper limit topology $\mathscr{T}_u$ in $\mathbb{R}$. Before we discuss the relationship between $\mathscr{T}_u$ and the pseudoinverse of a cumulative distribution function, we recall some of its properties:

- $\mathscr{T}_u$ is finer than the usual topology: $(a, b) \in \mathscr{T}_u$.
- $\mathscr{T}_u$ is first countable: the intervals $(p, x]$ (where $p \in \mathbb{Q}$, $p < x$) form a countable basis at x.
- $\mathscr{T}_u$ is separable: $\mathbb{Q}$ is a dense set.
- $\mathscr{T}_u$ is not second countable thus not metrizable.

Lemma 1.3.8 *$\mathscr{T}_u$ is strongly Lindelöf, this is, every open cover contains a countable subcover.*

Proof. Let $\{U_\alpha\}_\alpha$ be a collection of open sets in $\mathscr{T}_u$ and V_α be the interior of U_α in the usual topology. Then, let

$$U = \bigcup_\alpha U_\alpha \qquad \text{and} \qquad V = \bigcup_\alpha V_\alpha.$$

We claim that $A = U \backslash V$ is countable. To see this, take $x \in A$. Find y so close to x that $(y, x] \subset U$. Then $(y, x) \subset V$, which implies

$$(y, x) \cap A = \emptyset.$$

Now we can conclude that A is countable because the elements of A are "separated": for every $x \in A$ there exists y_x such that $(y_x, x) \subset A^c$. Finally, since the usual topology of $\mathbb{R}$ is strongly Lindelöf, we may rewrite

$$V = \bigcup_{i=1}^{\infty} V_{\alpha_i},$$

and notice that $U = A \cup V$. ∎

The range of $F^{(-1)}$, $\operatorname{Ran} F^{(-1)}$, is the set

$$\{F^{(-1)}(u) \in \mathbb{R} : u \in (0,1]\}.$$

Notice that we included $F^{(-1)}(1)$ in its range, provided that the value is a real number. In case 1 is not attained by F in $\mathbb{R}$ we assign $F^{(-1)}(1) = \infty$ but $\infty \notin \operatorname{Ran} F^{(-1)}$. Clearly $F^{(-1)}(u) \uparrow F^{(-1)}(1)$ whenever $u \to 1$. The next theorem gives a relation between $\operatorname{Ran} F^{(-1)}$ and the upper limit topology.

Theorem 1.3.9 *Let F be the cumulative distribution function associated to the probability measure μ, that is, $F(x) = \mu(-\infty, x]$. Then*

$$\mathit{Ran}\, F^{(-1)} = \bigcap \{C \subset \mathbb{R} : \mu(C) = 1,\ C \text{ is closed in } \mathscr{T}_u\},$$

and

$$\mu\left(\mathit{Ran}\, F^{(-1)}\right) = 1.$$

Proof. In this theorem what we see is that the range of $F^{(-1)}$ is the smallest closed set with full probability. First we show $\operatorname{Ran} F^{(-1)} \subset \mathcal{C}$, where

$$\mathcal{C} = \bigcap \{C \subset \mathbb{R} : \mu(C) = 1,\ C \text{ is closed in } \mathscr{T}_u\}.$$

Let $x = F^{(-1)}(u), u \in (0,1], x \neq \infty$, then we have

$$\mu(-\infty, x] \geq u \quad \text{and} \quad \mu(-\infty, x - \epsilon] < u \quad \forall \epsilon > 0,$$

which implies $\mu(x-\epsilon, x] > 0$. Let C be a closed set, if $x \notin C$ then $(x-\epsilon, x] \subset C^c$ for some $\epsilon > 0$ (x belongs to the open set C^c). Then we must have $\mu(C) < 1$. Thus $x \in \mathcal{C}$.

On the other hand, let $x \in \mathcal{C}$. Then $\mu(x-\epsilon, x] > 0$ for all $\epsilon > 0$ (otherwise x would not be in $\mathcal{C}$). Let $u = F(x)$, we claim

$$F^{(-1)}(u) = x.$$

Since $\mu(-\infty, x] = u$ we know for sure that $F^{(-1)}(u) \leq x$. If $F^{(-1)}(u) < x$ then

$$\mu(F^{(-1)}(u), x] = 0,$$

which is a contradiction.

To prove $\mu\left(\operatorname{Ran} F^{(-1)}\right) = 1$ use the fact that $\mathscr{T}_u$ is strongly Lindelöf, then rewrite the intersection as a countable intersection. ■

To finish the section we present a straightforward consequence of the last theorem, but first we define B' to be the closure of $B \subset \mathbb{R}$ in $\mathscr{T}_u$.

Lemma 1.3.10 *If X is a random variable with $P(X \in B) = 1$, $B \in \mathcal{B}(\mathbb{R})$, distribution μ and cumulative distribution function F then*

$$Ran\, F^{(-1)} \subset B'.$$

1.3.3 Functional Monotone Class Theorem

Before we introduce the Functional Monotone Class Theorem, which will be used in the next section, we state without proof **Dynkin's** π, λ**-systems Lemma**. To do so, we need some definitions.

Let Ω be a non empty set. A π**-system** is any class of subsets of Ω that is closed under finite intersections. Similarly, a class $\mathcal{D}$ of subsets of Ω is a λ**-system** if

- $\Omega \in \mathcal{D}$,
- if $A, B \in \mathcal{D}$ and $A \subset B$, then $B \backslash A \in \mathcal{D}$.
- for any increasing sequence $\{A_n\}$ of elements of $\mathcal{D}$, we have $\lim A_n \in \mathcal{D}$.

It is clear that any σ-algebra is both a π and λ-system. Define $\pi(\mathcal{I})$ to be the smallest π-system containing $\mathcal{I}$. Similarly, we define $\lambda(\mathcal{I})$ and $\sigma(\mathcal{I})$.

Lemma 1.3.11 (Dynkin's π, λ-systems Lemma) *If $\mathcal{I}$ is a π-system, then*

$$\lambda(\mathcal{I}) = \sigma(\mathcal{I}).$$

Theorem 1.3.12 (Functional Monotone Class Theorem) *Let $\mathcal{H}$ be a collection of bounded real valued functions from Ω satisfying the following conditions*

a) $\mathcal{H}$ is a vector space over $\mathbb{R}$, *b)* $\mathcal{H}$ *contains the constant function* 1 *(hence every constant function),*

c) if $\{f_n\} \subset \mathcal{H}$ *is an increasing sequence with* $\sup_n \sup_\omega |f_n(\omega)| < \infty$*, then* $\lim_n f_n \in \mathcal{H}$*, and*

d) $\mathcal{H}$ *contains the indicator function of every set in some* π*-system* $\mathcal{I}$*.*

Then $\mathcal{H}$ *contains every bounded* $\sigma(\mathcal{I})$ *measurable real valued function on* Ω*.*

Proof. Let $\mathcal{D}$ be the class of subsets F of Ω such that $\mathbb{I}_F \in \mathcal{H}$. Since $1 \in \mathcal{H}$, $\Omega \in \mathcal{D}$. If $A, B \in \mathcal{D}$ and $A \subset B$, then $\mathbb{I}_{B-A} = \mathbb{I}_B - \mathbb{I}_A \in \mathcal{H}$, by condition a. On the other hand, let $\{A_n\}$ be an increasing sequence of sets. Since

$$\mathbb{I}_{\lim A_n} = \lim \mathbb{I}_{A_n}$$

(and, by condition c) we can conclude that $\mathcal{D}$ is a λ-system. It means, by the previous lemma, that $\sigma(\mathcal{I}) \subset \mathcal{D}$.

Now, let $f \geq 0$ be a $\sigma(\mathcal{I})$-measurable and bounded function. As a matter of fact, f may be approximated as an increasing sequence of simple functions, which are nothing else than linear combinations of indicator functions. Therefore $f \in \mathcal{H}$. Finally, if f is bounded but not necessarily non negative, we can always rewrite f as $f^+ - f^-$, where

$$f^+ = f \cdot \mathbb{I}_{\{f \geq 0\}} \quad \text{and} \quad f^- = -f \cdot \mathbb{I}_{\{f<0\}}.$$

Clearly, f^+ and f^- are non negative and bounded. It follows that f^+, f^- and f are elements of $\mathcal{H}$ and the proof is complete. ∎

1.3.4 Weak Convergence

From now on S and S' are used to denote Polish spaces. As before, $\mathcal{B}(S)$ denotes the Borel σ-algebra of S, and the elements of $\mathcal{B}(S)$ are called Borel sets. A measurable space that is Borel isomorphic to a Borel subset T of $[0, 1]$ (there exists a bijection $f : S \to T$ such that both f and f^{-1} are measurable) is called a **Borel space**. It can be proven that any Polish space endowed with its Borel σ-algebra is a Borel space. A measure on the measurable space $(S, \mathcal{B}(S))$ is called a Borel measure.

Define

$$\mathcal{M}(S) = \{\mu : \mu \text{ is a probability measure on } (S, \mathcal{B}(S))\}.$$

Let Φ be the function from $\mathcal{M}(S) \times B_b(S)$ to $\mathbb{R}$ given by

$$\Phi(\mu, f) = \int_S f \, d\mu, \tag{1.1}$$

where $B_b(S)$ is the set of all bounded measurable functions from S to $\mathbb{R}$. Similarly, $C_b(S) \subset B_b(S)$ denotes the set of all bounded and continuous functions. Clearly each element of $B_b(S)$ is integrable, so that Φ is well defined. As usual, the f-section of Φ is given by

$$\Phi_f : \mathcal{M}(S) \to \mathbb{R}, \quad \Phi_f(\mu) = \Phi(\mu, f).$$

Similarly, we define the μ-section of Φ, Φ_μ. Let $\mathscr{T}_{\mathcal{M}(S)}$ be the smallest topology that makes Φ_f a continuous function for every $f \in C_b(S)$, this is, the smallest topology containing

$$\mathscr{C} := \{\Phi_f^{-1}(O) : f \in C_b(S),\ O \subset \mathbb{R} \text{ is an open set}\}.$$

Indeed, $\mathscr{T}_{\mathcal{M}(S)}$ is generated by the smaller set

$$\mathscr{C} := \{\Phi_f^{-1}(O) : f \in C_b(S),\ O \in \mathscr{U}_{\mathbb{R}}\},$$

where $\mathscr{U}_{\mathbb{R}}$ is a countable basis of the topology in $\mathbb{R}$. As before, we can also define the Borel σ-algebra of $\mathcal{M}(S)$, denoted by $\mathcal{B}(\mathcal{M}(S))$.

Example 1 The map $\phi : \mathcal{M}(S) \to \mathbb{R}$, given by

$$\phi(\mu) = \mu(O), \quad O \text{ is an open set},$$

is measurable since it can be written as the limit of continuous (hence measurable) functions. Set

$$f_n(x) = \min\{1, nd(x, O^c)\}, \quad \text{where } d(x, O^c) = \inf_{y \in O^c} d(x, y).$$

Clearly, $f_n \in C_b(S)$ for all n. Then

$$\phi(\mu) = \int \mathbb{I}_O \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu,$$

because $0 \leq f_n \uparrow \mathbb{I}_O$.

Example 2 The map $\mu \mapsto \mu(A)$, $A \in \mathcal{B}(S)$ is also measurable. It is important to note that we can no longer use the approximation argument. However, we use the Functional Monotone Class Theorem (Theorem 1.3.12) with

$$\mathcal{I} = \{O \subset S : O \text{ is open }\} \text{ and } \mathcal{H} = \{f : \mu \mapsto \Phi_f(\mu) \text{ is measurable }\},$$

since (by previous example) every indicator function of an open set belongs to $\mathcal{H}$. The first and second condition of Theorem 1.3.12 are clearly satisfied, while the Bounded Convergence Theorem proves that $\mathcal{H}$ is closed under increasing sequences of functions, which is the third condition. Therefore $B_b(S)$, the set of all measurable and bounded functions from S to $\mathbb{R}$, is contained in $\mathcal{H}$. In particular $\mu(A) = \int \mathbb{I}_A \, d\mu$ defines a measurable map.

Example 3 In the previous example we explained how to prove that

$$\mu \mapsto \int f \, d\mu$$

is measurable if $f : S \to \mathbb{R}$ is measurable and bounded.

A subbasis $\mathscr{S}$ of a topology is a collection of sets that cover the entire space. From a subbasis we can create a basis for a topology as follows

$$\mathscr{U} = \left\{ \bigcap_{i=1}^{n} V_i : V_i \in \mathscr{S},\ n \in \mathbb{N} \right\}.$$

Consider the following open neighborhoods of $\mu \in \mathcal{M}(S)$

$$\begin{aligned}
\mathscr{S}_{1,\mu} &= \left\{ \nu : \left| \int f \, d\nu - \int f \, d\mu \right| < \epsilon,\ f \in C_b(S) \right\} \\
\mathscr{S}_{2,\mu} &= \{\nu : \nu(F) < \mu(F) + \epsilon,\ F \text{ closed set}\} \\
\mathscr{S}_{3,\mu} &= \{\nu : \nu(G) > \mu(G) - \epsilon,\ G \text{ open set}\} \\
\mathscr{S}_{4,\mu} &= \{\nu : |\nu(A) - \mu(A)| < \epsilon,\ A \ \mu\text{-continuity set}\},
\end{aligned}$$

where a μ-continuity set A is such that $\mu(\partial A) = 0$.

Then, let $\mathscr{S}_i$, for $i \in \{1, 2, 3, 4\}$, be the subbasis obtained after taking the union of the neighborhoods at μ for all $\mu \in \mathcal{M}(S)$. Similarly we define the basis $\mathscr{U}_i$, $i = 1, 2, 3$ and 4, generated by the subbasis $\mathscr{S}_i$. In the following theorem we show that the four bases generate the same topology, $\mathscr{T}_{\mathcal{M}(S)}$.

Theorem 1.3.13 *The topologies generated by each $\mathscr{U}_1, \mathscr{U}_2, \mathscr{U}_3$ and $\mathscr{U}_4$ coincide and are equal to $\mathscr{T}_{\mathcal{M}(S)}$.*

Proof. Let $\mathscr{T}_i$ be the topology generated by $\mathscr{U}_i$. Choose any $f \in C_b(S)$ and $\epsilon > 0$. Let I be the open interval $\left(\int f\, d\mu - \epsilon, \int f\, d\mu + \epsilon\right)$, then

$$\left\{\nu : \left|\int f\, d\nu - \int f\, d\mu\right| < \epsilon\right\} = \Phi_f^{-1}(I).$$

It follows that $\mathscr{T}_1 = \mathscr{T}_{\mathcal{M}(S)}$, ($\Phi$ as in (1.1)) On the other hand, it is easy to see that a set in $\mathscr{S}_2$ coincide with a set in $\mathscr{S}_3$ (just take $F = G^c$). So, we only have to show

- $\mathscr{T}_2 = \mathscr{T}_4$. Let $U \in \mathscr{S}_4$, say $U = \{\nu : |\nu(A) - \mu(A)| < \epsilon,\ A\ \mu\text{-continuity set}\}$ and let

$$V = \{\nu : \nu(\overline{A}) < \mu(\overline{A}) + \epsilon\} \cap \{\nu : \nu(\overline{A^c}) < \mu(\overline{A^c}) + \epsilon\} \in \mathscr{U}_2.$$

  Notice $\mu \in V$. We claim that $V \subset U$, i.e. any set in $\mathscr{S}_{4,\mu}$ contains a set in $\mathscr{U}_2$. Take $\nu \in V$, then

$$\nu(A) \le \nu(\overline{A}) < \mu(\overline{A}) + \epsilon = \mu(A) + \epsilon$$

  and

$$\nu(A^c) \le \nu(\overline{A^c}) < \mu(\overline{A^c}) + \epsilon = \mu(A^c) + \epsilon,$$

  which implies $\nu(A) > \mu(A) - \epsilon$. It follows, as we wanted to show, that

$$|\nu(A) - \mu(A)| < \epsilon.$$

  Now, let $U = \{\nu : \nu(F) < \mu(F) + \epsilon,\ F \text{ closed set}\} \in \mathscr{S}_2$. Find δ such that $F_\delta = \{s \in S : d(s, F) < \delta\}$ is a μ-continuity set and $\mu(F_\delta) < \mu(F) + \frac{\epsilon}{2}$. Let

$$V = \left\{\nu : |\nu(F_\delta) - \mu(F_\delta)| < \frac{\epsilon}{2}\right\},$$

  clearly if $\nu \in V$ then $\nu(F) \le \nu(F_\delta) < \mu(F) + \epsilon$. In other words, $V \subset U$.

- $\mathscr{T}_1 = \mathscr{T}_2$. First we show that every set in $\mathscr{S}_{2,\mu}$ contains a set in $\mathscr{S}_{1,\mu}$ and then we show the other direction. Let $U = \{\nu : \nu(F) < \mu(F) + \epsilon,\ F \text{ closed set}\}$. Take F_δ such that $\mu(F_\delta) < \mu(F) + \frac{\epsilon}{2}$. Choose $f \in C_b(S)$ so that

$$f(s) = \begin{cases} 1 & \text{if} \quad s \in F, \\ 0 & \text{if} \quad s \notin F_\delta, \\ t \in [0,1] & \text{if} \quad s \in F_\delta \backslash F. \end{cases}$$

Let

$$V = \left\{ \nu : \left| \int f \, d\nu - \int f \, d\mu \right| < \frac{\epsilon}{2} \right\}.$$

To show $V \subset U$ take $\nu \in V$, clearly

$$\nu(F) \leq \int f \, d\nu < \mu(F) + \epsilon, \quad \text{i.e. } \nu \in U.$$

Finally, let $U = \left\{ \nu : \left| \int f \, d\nu - \int f \, d\mu \right| < 2\epsilon,\ f \in C_b(S) \right\}$. We may assume $0 \leq f \leq 1$, otherwise transform by adding and multiplying by appropriate constants. Pick $k \in \mathbb{N}$ such that $\frac{1}{k} < \epsilon$. Define $F_i = \left\{ x : \frac{i}{k} \leq f(x) \right\}$. One can show that for $m \in \mathcal{M}(S)$

$$\frac{1}{k} \sum_{i=1}^{k} m(F_i) \leq \int f \ dm < \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} m(F_i).$$

We define

$$V = \bigcap_{i=1}^{k} \left\{ \nu : \nu(F_i) < \mu(F_i) + \frac{1}{k} \right\} \in \mathscr{B}_2.$$

Our claim is that $V \subset U$. Take $\nu \in V$, then

$$\begin{aligned} \int f \, d\nu &< \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \nu(F_i) \leq \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \left(\mu(F_i) + \frac{1}{k} \right) \\ &= \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \mu(F_i) + \frac{1}{k} < \sum_{i=1}^{k} \mu(F_i) + 2\epsilon \\ &\leq \int f \, d\mu + 2\epsilon. \end{aligned}$$

The same argument applied to $1 - f$ will show $\int f \, d\mu - 2\epsilon < \int f \, d\nu$, which completes the proof. ∎

In fact, it is possible to reduce the subbasis $\mathscr{S}_2$ and $\mathscr{S}_3$ and still generate the same topology $\mathscr{T}_{\mathcal{M}(S)}$.

Lemma 1.3.14 *The topology $\mathscr{T}_{\mathcal{M}(S)}$ is generated by the subbasis*

$$\mathscr{S}_3' = \{\{\nu : \nu(G) > \mu(G) - \epsilon\}, \quad \mu \in \mathcal{M}(S),\ G \in \mathscr{U}'\},$$

where $\mathscr{U}'$ is the set of finite unions of elements in the countable basis $\{G_i\}_{i=1}^{\infty}$ of S.

Proof. Clearly $\mathscr{T}_{\mathcal{M}(S)}$ is finer than the topology generated by this subbasis, hence we only have to show that any set of the form

$$U = \{\nu : \nu(G) > \mu(G) - 2\epsilon\}, \ \ G \text{ open set}, \ \ \mu \in \mathcal{M}(S)$$

contains a set in V in $\mathscr{S}_2'$ with $\mu \in V$. Let $G' \subset G$ be a set in $\mathscr{U}'$ satisfying $\mu(G \backslash G') < \epsilon$. Then we show that

$$V = \{\nu : \nu(G') > \mu(G') - \epsilon\} \subset U.$$

To see this, take $\nu \in V$, then

$$\nu(G') > \mu(G') - \epsilon > \mu(G') + \mu(G \backslash G') - 2\epsilon = \mu(G) - 2\epsilon.$$

It implies $\nu(G) > \mu(G) - 2\epsilon$, i.e. $\nu \in U$. ∎

Now, let μ_n and μ be probability measures on $(S, \mathcal{B}(S))$. We say that μ_n **converges weakly** to μ, in symbols $\mu_n \Rightarrow \mu$, if

$$\lim_{n\to\infty} \int_S f \, d\mu_n = \int_S f \, d\mu, \ \text{ for all } f \in C_b(S).$$

Clearly, the previous definition is closely related to the definition of convergence in distribution, in fact one can see, using change of variables, that $X_n \xrightarrow{D} X$ iff $\mu_{X_n} \Rightarrow \mu_X$, where $X_n \stackrel{D}{=} \mu_{X_n}$ and $X \stackrel{D}{=} \mu_X$.

As one can imagine, the definition of convergence we just gave is consistent with the topological space $(\mathcal{M}(S), \mathscr{T}_{\mathcal{M}(S)})$.

Lemma 1.3.15 *Let $\{\mu_n\}_{n=1}^{\infty}$, μ be a sequence and an element of $\mathcal{M}(S)$ respectively. Then $\mu_n \Rightarrow \mu$ iff $\{\mu_n\}$ $\mathscr{T}_{\mathcal{M}(S)}$-converges to μ.*

Proof. Suppose $\{\mu_n\}$ $\mathscr{T}_{\mathcal{M}(S)}$-converges to μ, and let $f \in C_b(S)$. Plugging into the $\mathscr{T}_{\mathcal{M}(S)}$ continuous function Φ_f, we get $\int f\, d\mu_n \to \int f\, d\mu$.

On the other hand, suppose $\int f d\,\mu_i \to \int f\, d\mu$, for all $f \in C_b(S)$. Let V be a neighborhood of μ. Find a set of the form

$$U = \bigcap_{i=1}^{m} \Phi_{f_i}^{-1}(U_i), \quad f_i \in C_b(S),\, U_i \text{ open in } \mathbb{R}$$

satisfying $\mu \in U$ and $U \subset V$. Let k_i, $i \in \{1, \ldots, m\}$, be such that for all $n \geq k_i$ we have $\mu_n \in \Phi_{f_i}^{-1}(U_i)$. Then take the maximum of this numbers, i.e. let $k = \max\{k_1, \ldots, k_m\}$. Clearly, if $n \geq k$ then $\mu_n \in \bigcap_{i=1}^{m} \Phi_{f_i}^{-1}(U_i) \subset V$. We just proved that for every neighborhood V of μ there exists k so that $\{\mu_n\}_{n \geq k} \subset V$. ∎

Example Let $x \in S$. Let $\delta_x : \mathcal{B}(S) \to \{0, 1\}$ be the probability measure given by

$$\delta_x(A) = \begin{cases} 0 & \text{if} \quad x \notin A, \\ 1 & \text{if} \quad x \in A. \end{cases} \tag{1.2}$$

Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in S converging to x, then $\delta_{x_n} \Rightarrow \delta_x$.

The following theorem give us alternative definitions of weak convergence.

Theorem 1.3.16 (Portmanteau Theorem) *For probability measures μ_n and μ on $(S, \mathcal{B}(S))$ the following are equivalent:*

1. $\lim\limits_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$*, for all f in $C_b(S)$*
2. *For any open set U we have $\liminf\limits_{n\to\infty} \mu_n(U) \geq \mu(U)$*
3. *For any closed set F we have $\limsup\limits_{n\to\infty} \mu_n(F) \leq \mu(F)$*

4. For any continuity set A of μ we have $\lim_{n\to\infty} \mu_n(A) = \mu(A)$

Proof. **1. $\Rightarrow$ 2.** Let U be an open set and F its complement. Define $f_k(x) = \min\{1, kd(x,F)\}$, where $d(x,F) = \inf_{y\in F} d(x,y)$, and notice that $0 \leq f_k \uparrow \mathbb{I}_U$. Moreover, every f_k is a bounded and continuous function from S to $\mathbb{R}$. Hence

$$\lim_{n\to\infty} \int f_k \, d\mu_n = \int f_k \, d\mu.$$

On the other hand, since $\mu_n(U) \geq \int f_k \, d\mu_n$, we have

$$\liminf_{n\to\infty} \mu_n(U) \geq \liminf_{n\to\infty} \int f_k \, d\mu_n = \lim_{n\to\infty} \int f_k \, d\mu_n = \int f_k \, d\mu.$$

Finally, letting $k \to \infty$ gives

$$\mu(U) = \int \mathbb{I}_U \, d\mu = \lim_{k\to\infty} \int f_k \, d\mu \leq \liminf_{n\to\infty} \mu_n(U),$$

since $\int f_k \, d\mu \leq \liminf \mu_n(U)$.

2. $\Leftrightarrow$ 3. It is easily verified taking complements.

3. $\Rightarrow$ 4. Let A be a continuity set of μ, this is $\mu(\partial A) = 0$. Recall: $\operatorname{int} A \subset A \subset \overline{A}$ and $\partial A = \overline{A} \setminus \operatorname{int} A$. Then,

$$\begin{aligned}
\mu(\operatorname{int} A) &\leq \liminf_{n\to\infty} \mu_n(\operatorname{int} A) \leq \liminf_{n\to\infty} \mu_n(A) \\
&\leq \limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\overline{A}) \\
&\leq \mu(\overline{A}),
\end{aligned}$$

since $\operatorname{int} A$ is an open set and $\overline{A}$ is closed. Moreover, $\overline{A} = \operatorname{int} A \cup \partial A$ implies $\mu(\overline{A}) = \mu(\operatorname{int} A) + \mu(\partial A)$. Since $\mu(\partial A) = 0$, then

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

4. $\Rightarrow$ 1. Let $f \in C_b(S)$, define $F_y := f^{-1}(\{y\})$. Notice that if $y_1 \neq y_2$ then $F_{y_1} \cap F_{y_2} = \emptyset$. Also, $\mu(F_y) > 0$ for at most countably many values of y. In fact, without loss of generality, we may assume $\mu(F_0) = 0$ (otherwise add an appropriate constant).

On the other hand, for every $\epsilon > 0$ and all $k \in \mathbb{Z}$ define

$$B_{k,\epsilon} := f^{-1}\Big([k\epsilon, (k+1)\epsilon)\Big).$$

It follows that $\partial B_{k,\epsilon} \subset F_{k\epsilon} \cap F_{(k+1)\epsilon}$, which implies

$$\mu(\partial B_{k,\epsilon}) \leq \mu(F_{k\epsilon}) + \mu(F_{(k+1)\epsilon}).$$

If $k = 0$, we have $\mu(\partial B_{0,\epsilon}) \leq \mu(F_\epsilon)$, since $\mu(F_0) = 0$. Now, let $\{y_1, y_2, \ldots\}$ be the set of points such that $\mu(F_{y_i}) > 0$. Define C as the set

$$\left\{\epsilon : \epsilon = \frac{y_i}{k},\ k \in \mathbb{Z} - \{0\},\ i \geq 1\right\}.$$

Notice that $\mu(F_{k\epsilon}) > 0$ for some k, if and only if $\epsilon \in C$. Clearly C is countable and, if $\epsilon' \in C^c, \epsilon' > 0$ then

$$\mu(\partial B_{k,\epsilon'}) = 0,\ \text{for all } k \in \mathbb{Z}.$$

There exists a sequence $\{\epsilon_j\}$ convergent to zero so that $\{\epsilon_j\} \subset C^c, \epsilon_j > 0$ (otherwise C would not be countable). Hence B_{k,ϵ_j} is a continuity set of μ (for all k and all j) and $\mu_n(B_{k,\epsilon_j}) \to \mu(B_{k,\epsilon_j})$ as n goes to infinity.

Since f is bounded, for a given ϵ, $B_{k,\epsilon} = \emptyset$ for all but finitely many $k's$. Now, for a fixed ϵ_j

$$\sum_k k\epsilon_j \mu(B_{k,\epsilon_j}) \geq \int f\, d\mu - \epsilon_j,$$

where the summation is taken over all those (finitely many) $k's$ such that B_{k,ϵ_j} is not an empty set. Clearly,

$$\sum_k k\epsilon_j \mu(B_{k,\epsilon_j}) = \lim_{n\to\infty} \sum_k k\epsilon_j \mu_n(B_{k,\epsilon_j}),$$

and

$$\lim_{n\to\infty} \sum_k k\epsilon_j \mu_n(B_{k,\epsilon_j}) = \liminf_{n\to\infty} \sum_k k\epsilon_j \mu_n(B_{k,\epsilon_j}) \leq \liminf_{n\to\infty} \int f\, d\mu_n.$$

On the other hand

$$\limsup_{n\to\infty} \int f\, d\mu_n \leq \limsup_{n\to\infty} \sum_k (k+1)\epsilon_j \mu_n(B_{k,\epsilon_j}) = \lim_{n\to\infty} \sum_k (k+1)\epsilon_j \mu_n(B_{k,\epsilon_j}).$$

Moreover,

$$\lim_{n\to\infty} \sum_k (k+1)\epsilon_j \mu_n(B_{k,\epsilon_j}) = \sum_k (k+1)\epsilon_j \mu(B_{k,\epsilon_j}) \leq \int f\, d\mu + \epsilon_j.$$

Then, putting everything together,

$$\int f\, d\mu - \epsilon_j \leq \liminf_{n\to\infty} \int f\, d\mu_n \leq \limsup_{n\to\infty} \int f\, d\mu_n \leq \int f\, d\mu + \epsilon_j.$$

Finally, make $\epsilon_j \to 0$, to conclude

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

∎

In fact we can relax the first condition in the previous theorem, it is enough to verify if $\int f\, d\mu_n \to \int f\, d\mu$ for just countably many functions f.

Lemma 1.3.17 *Let $\mathscr{U} = \{U_n : n \in \mathbb{N}\}$ be a basis of the topology in S and*

$$\hat{\mathscr{U}} = \left\{ \bigcup_{i\in I} U_i : I \subset \mathbb{N},\ I \text{ is finite} \right\}.$$

If $\int f\, d\mu_n \to \int f\, d\mu$ for every function $f : S \to \mathbb{R}$ of the form

$$f(x) = \min\{1, kd(x, U^c)\}, \quad k \in \mathbb{N}, \quad U \in \hat{\mathscr{U}}$$

then $\mu_n \Rightarrow \mu$.

Proof. First of all, we want to emphasize that the $\hat{\mathscr{U}}$ is countable. As part of the proof of the previous theorem (**1.** $\Rightarrow$ **2.**) we proved that the hypothesis of the lemma implies $\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$ for all $U \in \hat{\mathscr{U}}$.

Let $V \subset S$ be any open set. Let $\{U_k\}_{k=1}^{\infty} \subset \mathscr{U}$ be such that

$$V = \limsup_{i\to\infty} V_i\,, \qquad \text{where } V_i = \bigcup_{k=1}^{i} U_k.$$

Clearly $\mu_n(V_i) \le \mu_n(V)$, hence

$$\liminf_{n\to\infty} \mu_n(V_i) \le \liminf_{n\to\infty} \mu_n(V).$$

By hypothesis $\mu(V_i) \le \liminf_{n\to\infty} \mu_n(V_i)$, therefore $\mu(V_i) \le \liminf_{n\to\infty} \mu_n(V)$.

It implies

$$\mu(V) = \lim_{i\to\infty} \mu(V_i) \le \liminf_{n\to\infty} \mu_n(V).$$

Therefore, by the Portmanteau theorem, $\mu_n \Rightarrow \mu$ ∎

Let

$$\mathcal{F} = \{f : f(x) = \min\{1, kd(x, U^c)\},\ k \in \mathbb{N},\ U \in \hat{\mathscr{U}}\}, \tag{1.3}$$

where $\hat{\mathscr{U}}$ as in the last lemma. Using the previous result and Lemma 1.3.13 we can obtain a simplified version of one of the subbasis that generates the topology $\mathscr{T}_{\mathcal{M}(S)}$ in $\mathcal{M}(S)$.

Lemma 1.3.18 *$\mathscr{T}_{\mathcal{M}(S)}$ is generated by the countable set*

$$\{\Phi_f^{-1}(O) : f \in \mathcal{F},\ O \in \mathscr{U}_{\mathbb{R}}\},$$

$\mathcal{F}$ as in (1.3), and $\mathscr{U}_{\mathbb{R}}$ denotes a countable basis of $\mathbb{R}$.

The next theorem shows that convergence almost surely is stronger than weak convergence, in the sense that if a sequence of random variables converges a.s., then the sequence of measures induced by them converges weakly.

Theorem 1.3.19 *Let $\{X_n\}$ and X be random elements from $(\Omega, \mathcal{F}, P)$ to $(S, \mathcal{B}(S))$. Suppose that $X_n \to X$ a.s. and let μ_n, μ be the distributions of X_n and X respectively. Then $\mu_n \Rightarrow \mu$.*

Proof. Let $f \in C_b(S)$, we need to prove

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

Since $\mu_n = P \circ X_n^{-1}$, we can write $\int f\, d\mu_n = \int f(X_n)\, dP$. Similarly, $\int f\, d\mu = \int f(X)\, dP$. On the other hand, since f is a continuous function we have that

$f(X_n) \to f(X)$ a.s. and the fact that f is bounded implies that $f(X_n)$ and $f(X)$ are also bounded. Hence, by the Bounded Convergence Theorem we have

$$\lim_{n\to\infty} \int f(X_n)\, dP = \int f(X)\, dP,$$

which means that $\mu_n \Rightarrow \mu$. ∎

In particular, if μ_n and μ are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we have an alternative and easier way to define weak convergence. In the next lemma we only prove that our original definition implies this new one. Later, in Section 2.1, we prove the converse of this result.

Theorem 1.3.20 *Let μ_n and μ be probability measures on $\mathbb{R}$ and F_n, F their respective c.d.f. Then, $\mu_n \Rightarrow \mu$ implies $F_n(x) \to F(x)$ for every continuity point x of F.*

Proof. By the Portmanteau Theorem

$$\limsup_{n\to\infty} F_n(x) \leq F(x) \quad \text{and} \quad \limsup_{n\to\infty} \mu_n([x,\infty)) \leq \mu([x,\infty)).$$

However, if x is a continuity point of F, then we must have

$$\limsup_{n\to\infty}[1 - F_n(x)] \leq \limsup_{n\to\infty} \mu_n([x,\infty)) \leq \mu([x,\infty)) = 1 - F(x),$$

which is the same as

$$\liminf_{n\to\infty} F_n(x) \geq F(x).$$

It follows that

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x).$$

Clearly, $F_n(x) \to F(x)$, whenever x is a continuity point of F. ∎

In the following theorem, taken from [2] (Theorem 5.1, p. 30), S and S' are Polish spaces, but is easy to see that it would be enough to work in metric spaces and the theorem would remain true.

Theorem 1.3.21 *Let $\{\mu_n\}_n$ and μ be probability measures on S, let X be a measurable function from S to S'. If $\mu_n \Rightarrow \mu$ and $\mu(D_X) = 0$, then*

$$\mu_n \circ X^{-1} \Rightarrow \mu \circ X^{-1},$$

where D_X is the discontinuity set of X.

Proof. We shall show that for any closed subset of S', lets say F, we have

$$\limsup_{n\to\infty} \mu_n X^{-1}(F) \leq \mu X^{-1}(F).$$

Since $\mu_n \Rightarrow \mu$ we have

$$\limsup_{n\to\infty} \mu_n(X^{-1}(F)) \leq \limsup_{n\to\infty} \mu_n \left(\overline{X^{-1}(F)}\right) \leq \mu \left(\overline{X^{-1}(F)}\right),$$

because $X^{-1}(F) \subset \overline{X^{-1}(F)}$. Therefore it is enough to prove that $\mu\Big(X^{-1}(F)\Big) = \mu \left(\overline{X^{-1}(F)}\right)$. Notice that

$$\overline{X^{-1}(F)} \subset D_X \cup X^{-1}(F),$$

By hypothesis $\mu(D_X) = 0$, which means that $\mu(X^{-1}(F)) = \mu \left(\overline{X^{-1}(F)}\right)$, as we wanted to prove. ∎

The following theorem will be used in Section 2.2.2. Let $\{f_n\}_{n=1}^{\infty}$ be the countable subset of $C_b(S)$ defined in Lemma 1.3.17. This set contains functions of the form

$$f(x) = \min\{1, kd(x, U^c)\},$$

where U is a certain kind of open set and $k \in \mathbb{N}$. Φ as defined in (1.1).

Theorem 1.3.22 *Let S, S' be two Polish spaces and X a function from $(\mathcal{M}(S'), \mathcal{B}[\mathcal{M}(S')])$ to $(\mathcal{M}(S), \mathcal{B}[\mathcal{M}(S)])$. Then, X is measurable iff $\Phi_{f_n} \circ X : \mathcal{M}(S') \to \mathbb{R}$ is measurable for all $n \in \mathbb{N}$.*

Proof. Because of the fact that $\mathscr{T}_{\mathcal{M}(S)}$ is generated by a countable subbasis, the σ-algebra $\mathcal{B}(\mathcal{M}(S))$ is generated by the same set, say

$$\mathscr{S} = \{\Phi_{f_n}^{-1}(O) : f \in \{f_n\}_{i=1}^{\infty},\ O \in \mathscr{U}_{\mathbb{R}}\}$$

($\mathscr{U}_{\mathbb{R}}$ is a countable basis of $\mathbb{R}$). If X is measurable, clearly $\Phi_{f_n} \circ X$ is measurable, since it is the composition of two measurable functions.

On the other hand, suppose $\Phi_{f_n} \circ X$ is measurable for all $n \in \mathbb{N}$. Then, for any $O \in \mathscr{U}_{\mathbb{R}} \subset \mathcal{B}(\mathbb{R})$, the set

$$(\Phi_{f_n} \circ X)^{-1}(O) = X^{-1}\Big(\Phi_{f_n}^{-1}(O)\Big),$$

is measurable. In other words, if $B \in \mathscr{S}$ then $X^{-1}(B)$ is measurable, i.e.

$$X^{-1}(\mathscr{S}) \subset \mathcal{B}[\mathcal{M}(S')].$$

Since $\mathscr{S}$ generates the σ-algebra $\mathcal{B}(\mathcal{M}(S))$, we can conclude that X is a measurable function. ∎

1.3.5 Prohorov Metric

So far we have defined the topological space $(\mathcal{M}(S), \mathscr{T}_{\mathcal{M}(S)})$. It can be proven that it is metrizable. Let's define

$$F_\epsilon = \{x \in S : d(x, F) < \epsilon\}.$$

If μ, ν are two elements of $\mathcal{M}(S)$, we set

$$d_P(\mu, \nu) = \inf\{\epsilon > 0 : \text{ for all closed sets } F, \mu(F) \leq \nu(F_\epsilon) + \epsilon\}. \tag{1.4}$$

Lemma 1.3.23 $d_P : \mathcal{M}(S) \times \mathcal{M}(S) \to [0, 1]$ *is a metric in* $\mathcal{M}(S)$.

Proof. Let

$$E_1 = \{\epsilon > 0 : \mu(F) \leq \nu(F_\epsilon) + \epsilon \ \text{ for all closed sets } F\}$$

$$E_2 = \{\epsilon > 0 : \nu(F) \leq \mu(F_\epsilon) + \epsilon \ \text{ for all closed sets } F\},$$

Then we show $E_1 = E_2$. Take any $\epsilon \in E_1$ and let F be any closed subset of S, let $G = S\backslash F_\epsilon$. Clearly, G is a closed set and $F \subset S\backslash G_\epsilon$. It follows that

$$\mu(F_\epsilon) = 1 - \mu(G) \geq 1 - \nu(G_\epsilon) - \epsilon \geq \nu(F) - \epsilon,$$

This is,

$$\nu(F) \leq \mu(F_\epsilon) + \epsilon.$$

Hence, $E_1 \subset E_2$. Similarly, we show $E_2 \subset E_1$. It follows immediately that $d_P(\mu, \nu) = d_P(\nu, \mu)$.

To prove the triangle inequality let $\mu, \nu, \lambda \in \mathcal{M}(S)$ and suppose $d_P(\mu, \nu) < \delta$ and $d_P(\nu, \lambda) < \epsilon$. Then, for any closed set F

$$\begin{aligned} \mu(F) &\leq \nu(F_\delta) + \delta \leq \nu\left(\overline{F_\delta}\right) + \delta \\ &\leq \lambda\left((\overline{F_\delta})_\epsilon\right) + \delta + \epsilon \leq \lambda(F_{\delta+\epsilon}) + \delta + \epsilon, \end{aligned}$$

which implies

$$d_P(\mu, \lambda) \leq d_P(\mu, \nu) + d_P(\nu, \lambda).$$

Finally, suppose $d_P(\mu, \nu) = 0$. It means that $\mu(F) = \nu(F)$ for every closed set F. Let $\mathcal{D} = \{B \in \mathcal{B}(S) : \mu(B) = \nu(B)\}$. In order to prove $\mu = \nu$ we must show $\mathcal{B}(S) \subset \mathcal{D}$. Notice that

1. $S \in \mathcal{D}$,
2. if $A, B \in \mathcal{D}$ and $A \subset B$,

   $$\mu(B\backslash A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B\backslash A),$$

   i.e. $B\backslash A \in \mathcal{D}$
3. if $\{B_n\}_n \subset \mathcal{D}$ is an increasing sequence then

   $$\mu\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \nu(B_n) = \nu\left(\lim_{n\to\infty} B_n\right),$$

   i.e. $\lim\limits_{n\to\infty} B_n \in \mathcal{D}$.

In other words, $\mathcal{D}$ is a λ-system. Moreover, $\mathcal{F} \subset \mathcal{D}$, where $\mathcal{F}$ is the set of all closed subsets of S. Since $\mathcal{F}$ is a π-system, by Dynkin's π, λ-systems (Lemma 1.3.11) we have

$$\mathcal{B}(S) = \sigma(\mathcal{F}) = \lambda(\mathcal{F}) \subset \mathcal{D}.$$

Therefore $\mu = \nu$. ∎

The metric d_P is called the **Prohorov metric** and it can be proven that the metric space $(\mathcal{M}(S), d_P)$ is separable and complete. For instance, if $D \subset S$ is a countable dense, then

$$\mathcal{D} = \left\{ \mu \in \mathcal{M}(S) : \mu = \sum_{n=1}^{N} a_n \delta_{x_n},\ \sum_{n=1}^{N} a_n = 1,\ a_n \in \mathbb{Q},\ x_n \in D \right\},$$

is a countable dense subset of $\mathcal{M}(S)$.

We close this chapter with a probabilistic interpretation of the Prohorov metric, provided in [8] (Chapter 3, Theorem 1.2). This interpretation will be used in Chapter 3.

Theorem 1.3.24 *Let $\mu, \nu \in \mathcal{M}(S)$. Define $\mathcal{P}(\mu, \nu)$ to be the set of all probability measures $m \in \mathcal{M}(S \times S)$ with marginals μ and ν. This is, for all $A \in \mathcal{B}(S)$ one has $m(A \times S) = \mu(A)$ and $m(S \times A) = \nu(A)$. Then*

$$d_P(\mu, \nu) = \inf_{m \in \mathcal{P}(\mu,\nu)} \inf\{\epsilon > 0 : m(\{(x, y) : d(x, y) \geq \epsilon\}) \leq \epsilon\}.$$

It is known that the Prohorov metric gives the same topology on $\mathcal{M}(S)$ as the topology of weak convergence.

Chapter 2

Skorohod's Theorem and Extensions

2.1 Skorohod's a.s. Representation Theorem

In Section 1.3.1 and 1.3.4 we introduced the concept of convergence a.s. and convergence in distribution of random elements as well as weak convergence of measures. We also proved, in Theorem 1.3.19, that convergence almost surely implies weak convergence of the distributions of such random elements. Skorohod's almost sure representation theorem establishes that there exist random variables for which weak convergence of their measures and convergence a.s. hold simultaneously.

In 1956 A. V. Skorohod published his paper "Limit Theorems for Stochastic Processes" [13] containing the proof of this converse result for a complete, metric and separable space. Later, in 1968, R. M. Dudley weakened the initial hypothesis and prove this result without completeness [6]. In this section we include both results.

Before exploring a more general space S let's begin with $\mathbb{R}$. Suppose we want to prove the previous statement: "if $\mu_n, \mu \in \mathcal{M}(\mathbb{R})$ and $\mu_n \Rightarrow \mu$, then there are $X_n \to X$ a.s. having distributions μ_n and μ". The first thing we should notice is that, given the structure of $\mathcal{B}(\mathbb{R})$, instead of looking at the measures μ_n and μ we can study their cumulative distributions functions, say F_n and F. Then, we have to construct $X_n, X : (0,1) \to \mathbb{R}$ with the given distributions.

Naturally, the first candidate we have in mind are the pseudoinverses of F_n and F (see Section 1.3.2)

$$\begin{aligned} F_n^{(-1)}(u) &= \inf\{x : F_n(x) \geq u\}\,, \quad u \in (0,1) \\ F^{(-1)}(u) &= \inf\{x : F(x) \geq u\}\,, \quad u \in (0,1). \end{aligned}$$

Since $F_n(x) \to F(x)$ for all but countable many values of x one can imagine that $F_n^{(-1)}(u) \to F^{(-1)}(u)$ hold for almost every $u \in (0,1)$. It would imply that following this method we can find the random variables we are looking for. In the following lemma we prove this statement.

Lemma 2.1.1 *Suppose $\mu_n, \mu \in \mathcal{M}(\mathbb{R})$ and $\mu_n \Rightarrow \mu$. Let $X, X_n : (U, \mathcal{B}, \lambda) \to \mathbb{R}$ be given by $X(u) = F^{(-1)}(u)$ and $X_n(u) = F_n^{(-1)}(u)$. Then X has distribution μ, X_n has distribution μ_n and*

$$X_n \to X \quad a.s.$$

Proof. To show that the law of X_n is μ_n it is enough to show that the cumulative distribution function of X_n is indeed F_n (similarly with X). In Lemma 1.3.6 we proved that $F^{(-1)}(u) \leq a$ iff $u \leq F(a)$. Therefore,

$$\begin{aligned} \lambda(X_n \leq x) &= \lambda(\{u : X_n(u) \leq x\}) = \lambda(\{u : F_n^{(-1)}(u) \leq x\}) \\ &= \lambda(\{u : u \leq F_n(x)\}) = \lambda(0, F_n(x)] \\ &= F_n(x). \end{aligned}$$

It shows that X_n and X have right distributions.

Now, to show $X_n \to X$ a.s. consider the functions Y_n and Y from $(0,1)$ to $\mathbb{R}$ defined as follows

$$Y_n(u) = \inf\{x : F_n(x) > u\} \quad \text{and} \quad Y(u) = \inf\{x : F(x) > u\}.$$

We can see that Y_n is quite similar to X_n, the same happens with Y and X. Indeed, is easy to see that

$$\lambda(X_n = Y_n) = \lambda(X = Y) = 1,$$

since the discontinuity points of X_n and X are countable. Now, fix any u in the set $\{Y(u) = X(u)\}$. Let v be a continuity point of F such that $F(v) > u$. As proved in Theorem 1.3.20, $F_n(v) \to F(v)$. It implies that for a large enough n we must have $Y_n(u) \leq v$. Therefore

$$\limsup_{n\to\infty} Y_n(u) \leq v.$$

Then, choose $v \downarrow Y(u)$ (it is possible because all but countably many v's are continuity points of F). It follows that

$$\limsup_{n\to\infty} Y_n(u) \leq Y(u).$$

By similar arguments

$$X(u) \leq \liminf_{n\to\infty} X_n(u).$$

Since $X_n(u) \leq Y_n(u)$, we have

$$X(u) \leq \liminf_{n\to\infty} X_n(u) \leq \limsup_{n\to\infty} X_n(u) \leq \limsup_{n\to\infty} Y_n(u) \leq Y(u).$$

Since $X(u) = Y(u)$ we have $X_n(u) \to X(u)$ for such u. Again, since $\lambda(X = Y) = 1$, we can conclude

$$\lambda(\{u : X_n(u) \to X(u)\}) = 1,$$

which proves our claim. ∎

Using the properties of $\mathbb{R}$, like the order of its elements, we were able to find X_n and X relatively easily. Using this result we can complete Lemma 1.3.20.

Lemma 2.1.2 *Let $\{\mu_n\}_{n=1}^{\infty}$ and μ be probability measures on $\mathbb{R}$. Let F_n, F be their cumulative distribution functions. Then $\mu_n \Rightarrow \mu$ iff $F_n(x) \to F(x)$ for all continuity points x of F.*

Proof. Half of the lemma has been proved in Lemma 1.3.20. We only need to show that $F_n(x) \to F(x)$ for all continuity points x of F implies $\mu_n \Rightarrow \mu$. By the previous result take $X_n = F_n^{(-1)}$ and $X = F^{(-1)}$, then $X_n \stackrel{D}{=} \mu_n$, $X \stackrel{D}{=} \mu$ and $X_n \to X$ a.s. Since convergence a.s. is stronger than convergence in distribution (Theorem 1.3.19), we must have $\mu_n \Rightarrow \mu$, which completes the proof. ∎

When we change from $\mathbb{R}$ to a more general space S it is not so clear how can we produce random elements with the required properties. For instance, we can no longer rely on cumulative distribution functions or their pseudoinverses. In the following theorem we explore Skorohod's proof for a complete, metric and separable space.

Theorem 2.1.3 ([13], 3.1.1, p. 281) *Let S be complete, metric and separable. Assume that the probability measures in S $\{\mu_n\}_{n=1}^{\infty}$ converge weakly to μ_0. Then there exist S-valued random elements $\{X_n\}_{n=0}^{\infty}$ on the Lebesgue probability space, having distributions μ_n $n = 0, 1, \ldots,$ such that*

$$X_n \to X_0 \quad a.s.$$

Proof. We construct Borel sets $S_{i_1,i_2,\ldots,i_k}$, $i_j, k \in \mathbb{N}$ in the following way. Let $\{x_i^k\}_{i=1}^{\infty}$ be a sequence of points such that every point of S lies at a distance no greater than $2^{-(k+1)}$ from at least one point of the sequence. As before $B(x, r)$ is the open ball with center at x and radius r. We can choose an $r_k \in \left(\frac{1}{2^{k+1}}, \frac{1}{2^k}\right)$ such that

$$\mu\left(\partial B(x_i^k, r_k)\right) = 0 \qquad \text{for all } i, \tag{2.1}$$

which can be done because there is at most a countable set of r where (2.1) is positive for each i. Let

$$D_i^k = B(x_i^k, r_k) \backslash \left(\bigcup_{j=1}^{i-1} B(x_j^k, r_k) \right),$$

$$S_{i_1,i_2,\ldots,i_k} = D_{i_1}^1 \cap D_{i_2}^2 \cap \cdots \cap D_{i_k}^k.$$

It is easy to see that the sets we obtained satisfy the following properties

- $S_{i_1,i_2,\ldots,i_k} \cap S_{i'_1,i'_2,\ldots,i'_k} = \emptyset$ if $i_j \neq i'_j$ for at least one j,
- $\bigcup\limits_{i_k=1}^{\infty} S_{i_1,i_2,\ldots,i_{k-1},i_k} = S_{i_1,i_2,\ldots,i_{k-1}}$, $\bigcup\limits_{i=1}^{\infty} S_i = S$,
- The diameter of each $S_{i_1,i_2,\ldots,i_k}$ is less than 2^{-k},
- $\mu\left(\partial S_{i_1,i_2,\ldots,i_k}\right) = 0$.

We denote by $\Delta^n_{i_1,i_2,\ldots,i_k}$, $(n = 0, 1, \ldots)$ intervals of the segment $[0, 1]$ defined as follows.

- $\Delta^n_{i_1,i_2,\ldots,i_k}$ and $\Delta^n_{i_1,i_2,\ldots,i'_k}$ have no intersection if $i_k \neq i'_k$,
- the union of $\{\Delta^n_{i_1,i_2,\ldots,i_k}\}_{i_k=1}^{\infty}$ is the interval $\Delta^n_{i_1,i_2,\ldots,i_{k-1}}$, which implies

$$\Delta^n_{i_1,i_2,\ldots,i_k} \subset \Delta^n_{i_1,i_2,\ldots,i_{k-1}},$$

- $\Delta^n_{i_1,i_2,\ldots,i_k}$ lies to the left of $\Delta^n_{i_1,i_2,\ldots,i'_k}$ whenever $i_k < i'_k$.
- the length of $\Delta^n_{i_1,i_2,\ldots,i_k}$ is equal to $\mu_n(S_{i_1,i_2,\ldots,i_k})$.

From each set $S_{i_1,i_2,\ldots,i_k}$ we now choose one point $x_{i_1,i_2,\ldots,i_k} \in S$. For $n = 0, 1, \ldots$ we define the functions $X^m_n : [0, 1] \to S$ by

$$X^m_n(t) = x_{i_1,i_2,\ldots,i_m} \quad \text{for } t \in \Delta^n_{i_1,i_2,\ldots,i_m}.$$

It is easy to see that $d(X^m_n(t), X^{m+p}_n(t)) \leq 2^{-m}$. Since S is a complete space, the limit X_n of $\{X^m_n\}_m$ exists. Since the length of $\Delta^n_{i_1,i_2,\ldots,i_k}$ tends to the length of $\Delta^0_{i_1,i_2,\ldots,i_k}$, for all interior points t of the intervals $\Delta^0_{i_1,i_2,\ldots,i_m}$ we have

$$\lim_{n\to\infty} d(X_0(t), X_n(t)) \leq 2^{-m},$$

and therefore for all $t \in [0, 1]$ except possibly a countable set (the boundaries of the intervals), we have

$$\lim_{n\to\infty} X_n(t) = X_0(t).$$

To complete the proof we must only show that the distribution P_{X_n} of each X_n is indeed μ_n. By construction P_{X_n} and μ_n coincide for all sets in

$$\mathcal{C} = \{S_{i_1,i_2,\ldots,i_k}, k \in \mathbb{N}\}.$$

In fact $\mathcal{C}$ is a π-system because the intersection of two sets is either empty or the "smallest" set (in case one is contained in the other). Hence $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$. Also notice that the class of sets where P_{X_n} and μ_n coincide is a λ-system, thus, if we show $\lambda(\mathcal{C}) = \mathcal{B}(S)$ we are done.

To do so, we show that for every open $O \subset S$ there exists a set in the (countable) class $\mathcal{C}$ inside O: take $x \in O$ and find $k \in \mathbb{N}$ so that $B(x, 2^{-k}) \subset O$, then we can certainly find $S_{i_1,i_2,\ldots,i_k}$ that contains x and so is completely contained in O. Then, it is easy to see that

$$O = \bigcup \{A \in \mathcal{C},\ A \subset O\}$$

(prove by contradiction). It follows that $O \in \lambda(\mathcal{C})$. Therefore we can conclude that $\sigma(\mathcal{C}) = \mathcal{B}(S)$, and the proof is complete. ∎

Now, we prove the same result without completeness in the space S.

Theorem 2.1.4 ([6], Theorem 3, p. 1569 and [11], Theorem 3.30, p.56) *Let $\mu_1, \mu_2, \ldots$ and μ be measures in the separable metric space (S, d) such that $\mu_n \Rightarrow \mu$. Then on the Lebesgue probability space there exist some random elements $\{X_n\}_n$ and X, from $(0,1)$ to S, with induced measures μ_n and μ respectively, satisfying $X_n \to X$ a.s.*

Proof. For this proof we start with a random element $X : (U, \lambda) \to S$ with distribution μ. The existence of such random element is given by the existence theorems (see Section 1.3.1, Theorem 1.3.5).

First we prove the theorem when S is finite. Without loss of generality, assume $S = \{1, \ldots, N\}$. Let $p_i = \lambda(X = i) = \mu(\{i\})$ and $p_{i,n} = \mu_n(\{i\})$. Clearly, $\mu_n \Rightarrow \mu$ iff $p_{i,n} \to p_i$ for all $i = 1, \ldots, N$. Let $U : (0,1) \to (0,1)$ be uniform and independent of X. We now construct some random elements $\bar{X}_n$ so that $\bar{X}_n \overset{D}{=} \mu_n$. Notice that

$$\lambda\left(\{X = i\} \cap \left\{U \le \frac{p_{i,n}}{p_i}\right\}\right) = p_i \cdot \min\left\{\frac{p_{i,n}}{p_i}, 1\right\} \le p_{i,n}.$$

Thus we may construct $\bar{X}_n$ satisfying

$$\{X = i\} \cap \left\{U \le \frac{p_{i,n}}{p_i}\right\} \subset \{\bar{X}_n = i\}.$$

Since $p_{i,n} \to p_i$ for each i, we get $\bar{X}_n \to X$ a.s.

For general S, fix any $p \in \mathbb{N}$ and choose a partition of S into μ-continuity sets $B_1, B_2, \ldots \in \mathcal{B}(S)$ of diameter less or equal than 2^{-p}. Next choose k so large that $\mu(B_0) < 2^{-p}$, where

$$B_0 = \left(\bigcup_{i=1}^{k} B_i \right)^c .$$

This is possible since $\bigcup_i B_i \uparrow S$. Let $I = \{0, 1, \ldots, k\}$. Define $h_n : (0,1) \to I$ so that $\lambda(\{h_n = i\}) = \mu_n(B_i)$ and put $h = \sum_{i=0}^{k} i\, \mathbb{I}_{(X \in B_i)}$. It's easy to see that $\lambda \circ h_n \Rightarrow \lambda \circ h$ ($B_0, B_1, \ldots, B_k$ are μ-continuity sets). Moreover, since I is a finite set, we may assume $h_n \to h$ a.s. (by the previous case).

Let us further introduce some independent random elements $Y_n^i : (0,1) \to B_i$, $i = 0, 1, \ldots, k$ with distributions given by

$$\lambda\left(Y_n^i \in B\right) = \frac{\mu_n(B \cap B_i)}{\mu_n(B_i)}.$$

Then define

$$X_n^p(u) = \begin{cases} Y_n^0(u) & \text{if} \quad u \in \{h_n = 0\}, \\ Y_n^1(u) & \text{if} \quad u \in \{h_n = 1\}, \\ & \vdots \\ Y_n^k(u) & \text{if} \quad u \in \{h_n = k\}. \end{cases}$$

To prove that X_n^p has distribution μ_n notice that, for any $B \in \mathcal{B}(S)$,

$$\{X_n^p \in B\} = \bigcup_{i=0}^{k} \{X_n^p \in B\} \cap \{h_n = i\} = \bigcup_{i=0}^{k} \{Y_n^i \in B\} \cap \{h_n = i\},$$

hence

$$\begin{aligned} \lambda\{X_n^p \in B\} &= \sum_{i=0}^{k} \lambda(\{Y_n^i \in B\} \cap \{h_n = i\}) = \sum_{i=0}^{k} \lambda(\{Y_n^i \in B\})\lambda(\{h_n = i\}) \\ &= \sum_{i=0}^{k} \frac{\mu_n(B \cap B_i)}{\mu_n(B_i)} \mu_n(B_i) = \mu_n(B). \end{aligned}$$

From the construction is clear that

$$\{h_n = h\} \cap \{X \notin B_0\} \subset \{d(X_n^p, X) \leq 2^{-p}\}, \quad \text{for all } n, p.$$

Since h_n, h have finite range, $h_n \to h$ a.s. and $P(X \in B_0) < 2^{-p}$, there exists for every p some $n_p \in \mathbb{N}$ with

$$P\left(\bigcup_{n \geq n_p} \{d(X_n^p, X) > 2^{-p}\}\right) < 2^{-p}, \quad \text{for all } p,$$

which is equivalent to $P(d(X_n^p, X) > 2^{-p},\ n \geq n_p) < 2^{-p}$. We may further assume that $n_1 < n_2 < \cdots$. By the Borel Cantelli Lemma we get that

$$P\left(\sup_{n \geq n_p} d(X_n^p, X) \leq 2^{-p} \text{ for all but finitely many } p\right) = 1.$$

Now define $X_n = X_n^p$ for $n_p \leq n < n_{p+1}$, and note that X_n has distribution μ_n. Moreover $X_n \to X$ a.s. ■

2.2 Extensions

In the rest of the chapter S is a Polish space, $\mathcal{M}(S)$ the space of all probability measures on (the Borel sets of) S endowed with the weak convergence topology and $(U, \mathcal{B}, \lambda)$ is the Lebesgue probability space. A function $\rho : \mathcal{M}(S) \times U \to S$ is a **representation** of $\mathcal{M}(S)$ if

1. ρ is Borel measurable, and
2. for every $\mu \in \mathcal{M}(S)$ the function $\rho_\mu : U \to S$, defined by

$$\rho_\mu(u) = \rho(\mu, u),$$

has distribution μ (with respect to λ on $U = (0,1)$).

In 1983, Blackwell and Dubins published a paper claiming the existence of a representation satisfying specific conditions. The main property of such representations is related with a "particular" kind of continuity. Notice that, when we say $\rho(\cdot, u)$ is continuous at μ we mean that $\mu_n \Rightarrow \mu$ implies $\rho(\mu_n, u)$ converges to $\rho(\mu, u)$ in S.

In my thesis we study this and other types of continuity, we name them for simplicity:

- (C1) There is a null set N so that for $u \notin N$, $\mu \mapsto \rho(\mu, u)$ is continuous.
- (C2) For each $\mu \in \mathcal{M}(S)$ there is a null set N_μ so that for $u \notin N_\mu$ the map $m \mapsto \rho(m, u)$ is continuous at μ.
- (C3) For each $\mu \in \mathcal{M}(S)$ and sequence $\mu_n \Rightarrow \mu$ there is a null set $N_{\mu,\{\mu_n\}}$ so that for $u \notin N_{\mu,\{\mu_n\}}$, $\rho(\mu_n, u) \to \rho(\mu, u)$.

Clearly (C1) is the strongest type of continuity and (C3) is the weakest among the three of them. However, a representation with any of this continuity properties is a generalization of Skorohod's result by taking $X_n(u) := \rho(\mu_n, u)$ and $X(u) := \rho(\mu, u)$. Moreover, ρ is defined in the whole space $\mathcal{M}(S)$, so that the limit random element with distribution μ does not change if we change the sequence $\{\mu_n\}$ converging to μ.

For instance, the existence of a (C2)-continuous representation for $S = \mathbb{R}$ can be easily verified, we just let ρ^*_μ be the pseudoinverse function of the cumulative distribution function of μ.

Theorem 2.2.1 *Denote by $\mathcal{M}^* = \mathcal{M}(\mathbb{R})$. Let $\rho^* : \mathcal{M}^* \times U \to \mathbb{R}$, be the function defined by:*

$$\rho^*(\mu, u) = \inf\{x : \mu((-\infty, x]) \geq u\},\ u \in (0, 1).$$

Then, ρ^* is a (C2)-continuous representation of $\mathcal{M}^*$. This is,

1. *ρ^* is a measurable function,*
2. *$\rho^*_\mu = \rho * (\mu, u)$ has distribution μ (with respect to λ on U) and,*
3. *except of a null subset N_μ of U, $\rho^*(\cdot, u)$ is continuous at μ.*

Proof. First of all notice that we can write

$$\rho^*(\mu, u) = F_\mu^{(-1)}(u).$$

To prove measurability we use Theorem 1.3.3. Since $F_\mu^{(-1)}(u)$ is left continuous, we only need to prove that $\rho^*_u(\cdot)$ is measurable for all u. Notice that

$$\begin{aligned} A &= \rho_u^{*-1}((-\infty, a]) = \{\mu \in \mathcal{M}^* : \rho^*_u(\mu) \leq a\} \\ &= \{\mu \in \mathcal{M}^* : \rho^*(\mu, u) \leq a\} = \{\mu \in \mathcal{M}^* : F_\mu^{(-1)}(u) \leq a\}. \end{aligned}$$

In Lemma 1.3.6 we proved $F_\mu^{(-1)}(u) \leq a$ iff $u \leq F(a)$. Therefore,

$$\begin{aligned} A &= \{\mu \in \mathcal{M}^* : F_\mu(a) \geq u\} = \{\mu \in \mathcal{M}^* : \mu((-\infty, a]) \geq u\} \\ &= \left\{\mu \in \mathcal{M}^* : \int f \, d\mu \geq u\right\} = \Phi_f^{-1}([u, \infty)), \end{aligned}$$

where $f = \mathbb{I}_{(-\infty,a]}$ is a bounded function from S to $\mathbb{R}$ and Φ is given by 1.1. By the second example in Section 1.3.4, $\mu \mapsto \Phi_f(\mu)$ is a measurable function, which means that the set A is measurable. Therefore ρ_u^* is measurable and we can conclude that ρ^* is a measurable function.

To complete the proof, we refer to Lemma 2.1.1. ∎

2.3 Blackwell and Dubins's Extension

Blackwell and Dubins sketched a proof for the existence of a (C2)-continuous representation of $\mathcal{M}(S)$, where S is an arbitrary Polish space, moreover, they constructed and gave an explicit formula for such function ρ (see [3]). A question we had in mind when we first studied the paper was a mistake my supervisor found, the function ρ was not well defined: one of the functions that defines ρ was evaluated outside its domain. Our goal was to fix this problem redefining the function. In this section we explain and prove the statements claimed by Blackwell and Dubins, while studying the representation introduced by them.

For the general case, we add a new parameter a so that our probability space is $A \times U$, where $A = [1, 2]$ and $U = (0, 1)$. Let D be a countable dense subset of S and let

$$\{H_n\}_{n=1}^{\infty} \tag{2.2}$$

be an enumeration of all open spheres with centers in D and with rational radius. Fix $a \in A$ and denote by H_n^a the open sphere with center equal to the center of H_n and with radius equal to the product of a with the radius of H_n, so that the sequence $H_1^a, H_2^a, \ldots$ also determines the topology of S. This is because, if $\mathcal{O}$ is any open set, we have $\mathcal{O} = \bigcup_{i \in \mathcal{C}} H_i^a$, where $\mathcal{C} = \{i : H_i^a \subset \mathcal{O}\}$.

For $(a, x) \in A \times S$ define

$$c(a, x) = \sum_{n=1}^{\infty} \frac{\mathbb{I}_{H_n^a}(x)}{3^n}.$$

Lemma 2.3.1 *$c : A \times S \to [0, 1/2]$ is an injective function.*

Proof. Suppose $c(a_1, x_1) = c(a_2, x_2)$, then

$$\sum_{n=1}^{\infty} \frac{\mathbb{I}_{H_n^{a_1}}(x_1) - \mathbb{I}_{H_n^{a_2}}(x_2)}{3^n} = \sum_{n=1}^{\infty} \frac{i_n}{3^n} = 0,$$

where $i_n = \mathbb{I}_{H_n^{a_1}}(x_1) - \mathbb{I}_{H_n^{a_2}}(x_2)$. Suppose that $i_n \neq 0$ for some $n \in \mathbb{N}$. Let n^* be the first number such that $i_{n^*} \neq 0$. Without loss of generality suppose $i_{n^*} = 1$. Then,

$$\sum_{n=1}^{\infty} \frac{i_n}{3^n} = \sum_{n=1}^{n^*-1} \frac{i_n}{3^n} + \frac{1}{3^{n^*}} + \sum_{n=n^*+1}^{\infty} \frac{i_n}{3^n} = 0.$$

If we multiply by 3^{n^*} we get

$$1 + \sum_{n=1}^{\infty} \frac{i_{n+n^*}}{3^n} = 0.$$

But $-\frac{1}{2} \leq \sum \frac{1}{3^n} i_{n+n^*} \leq \frac{1}{2}$. This contradiction shows that there is no such n^*, therefore $i_n = 0$ for all $n \in \mathbb{N}$.

Suppose $x_1 \neq x_2$ and, without loss of generality, $a_1 \leq a_2$. Find $H_m^{a_2}$ in $\{H_n^{a_2}\}_n$ so small that $x_2 \in H_m^{a_2}$ but $x_1 \notin H_m^{a_2}$. Since $H_m^{a_1} \subset H_m^{a_2}$ we get

$$\mathbb{I}_{H_m^{a_1}}(x_1) = 0 \quad \text{while} \quad \mathbb{I}_{H_m^{a_2}}(x_2) = 1,$$

which is a contradiction. Therefore x_1 has to be equal to x_2.

Now, assume $a_1 < a_2$ and let $x = x_1 = x_2$. Take $y \in S$, a center of one of the open balls $\{H_n\}_n$, such that $d(x, y) = \delta \neq 0$. Then, find $r \in \mathbb{Q}$ so that

$$a_1 < \frac{\delta}{r} < a_2.$$

Let H_k be the open ball with center y and radius r. Clearly $x \notin H_k^{a_1}$ but $x \in H_k^{a_2}$. Again this is a contradiction since, by hypothesis, we must have $\mathbb{I}_{H_k^{a_1}}(x) = \mathbb{I}_{H_k^{a_2}}(x)$ for all k. It follows that $a_1 = a_2$ and the proof is complete. ∎

Lemma 2.3.2 *The function c_a, given by $x \mapsto c(a,x)$, is continuous in the complement of $\bigcup_n \partial H_n^a$.*

Proof. Let $\{x_k\}$ be a sequence in S such that $x_k \to x$, with $x \notin \bigcup_n \partial H_n^a$. Our claim is that $c_a(x_k) \to c_a(x)$ as $n \to \infty$. Fix any $\epsilon > 0$. Let $N \in \mathbb{N}$ be such that $\sum_{n=N+1}^{\infty} \frac{1}{3^n} < \epsilon$. Then,

$$
\begin{aligned}
|c_a(x_k) - c_a(x)| &\leq \left| \sum_{n=1}^{N} \frac{\mathbb{I}_{H_n^a}(x_k) - \mathbb{I}_{H_n^a}(x)}{3^n} \right| + \left| \sum_{n=N+1}^{\infty} \frac{1}{3^n} \right| \\
&< \left| \sum_{n=1}^{N} \frac{\mathbb{I}_{H_n^a}(x_k) - \mathbb{I}_{H_n^a}(x)}{3^n} \right| + \epsilon.
\end{aligned}
$$

We claim that for a large enough k, $\mathbb{I}_{H_n^a}(x_k) = \mathbb{I}_{H_n^a}(x)$. Since $x_k \to x$ the equality is true if $x \in H_n^a$. On the other hand, suppose $x \notin H_n^a$. The fact that $x \notin \bigcup \partial H_n^a$ implies $x \notin \partial H_n^a$. Hence $x \in \operatorname{int}[(H_n^a)^c]$. It follows that eventually $x_k \in (H_n^a)^c$ and $\mathbb{I}_{H_n^a}(x_k) = \mathbb{I}_{H_n^a}(x) = 0$. Hence, for all $n \in \mathbb{N}$ exists a k'_n such that $\mathbb{I}_{H_n^a}(x_k) = \mathbb{I}_{H_n^a}(x)$ for all $k \geq k'_n$. Let k^* be the maximum of $\{k'_1, k'_2, \ldots, k'_N\}$. It follows that

$$
\left| \sum_{n=1}^{N} \frac{\mathbb{I}_{H_n^a}(x_k) - \mathbb{I}_{H_n^a}(x)}{3^n} \right| = 0,
$$

for all $k \geq k^*$. As a consequence $|c_a(x_k) - c_a(x)| < \epsilon$. So that $c_a(x_k) \to c_a(x)$, as we wanted to prove. ∎

Moreover, c_a is a measurable function because it may be expressed as the limit of measurable functions, namely

$$
c_a = \lim_{k \to \infty} \sum_{n=1}^{k} \frac{\mathbb{I}_{H_n^a}}{3^n}.
$$

Hence, we may define $J_a(\mu) \in \mathcal{M}(\mathbb{R})$ to be the distribution of c_a on the probability space $(S, \mathcal{B}, \mu)$, that is, $J_a(\mu) = \mu \circ c_a^{-1}$.

Lemma 2.3.3 *For a fixed $x \in S$ the function $a \mapsto c_a(x)$ is left continuous.*

Proof. Fix any $x \in S$ and $a \in (1, 2]$. Let $\{a_n\}_{n=1}^{\infty}$ be an increasing sequence converging to a. We need to prove that, if $\epsilon > 0$, then there exists $N^* \in \mathbb{N}$ such that

$$|c_{a_n}(x) - c_a(x)| < \epsilon \quad \text{for all } n \geq N^*.$$

Let N be an integer such that $\sum_{i=N}^{\infty} \frac{1}{3^i} < \epsilon$. Then,

$$\begin{aligned}
|c_{a_n}(x) - c_a(x)| &= \left| \sum_{i=1}^{N-1} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] + \sum_{i=N}^{\infty} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] \right| \\
&\leq \left| \sum_{i=1}^{N-1} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] \right| + \left| \sum_{i=N}^{\infty} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] \right| \\
&\leq \left| \sum_{i=1}^{N-1} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] \right| + \sum_{i=N}^{\infty} \left| \frac{1}{3^i} \mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right| \\
&< \left| \sum_{i=1}^{N-1} \frac{1}{3^i} \left[\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) \right] \right| + \epsilon.
\end{aligned}$$

Now,

$$\mathbb{I}_{H_i^a}(x) = \begin{cases} 0 & \text{if} \quad d(x, y_i) \geq a r_i, \\ 1 & \text{if} \quad d(x, y_i) < a r_i, \end{cases}$$

where y_i and r_i are the center and radius of H_i.

Suppose $\mathbb{I}_{H_i^a}(x) = 0$, since $a_n \uparrow a$, we have $a_n r_i \leq a r_i$ and

$$\mathbb{I}_{H_i^{a_n}}(x) = \mathbb{I}_{H_i^a}(x) = 0.$$

If $\mathbb{I}_{H_i^a}(x) = 1$, define $\epsilon_i = a r_i - d(x, y_i) > 0$. Since $a_n \uparrow a$, there exists m_i such that $|a_n - a| = a - a_n < \frac{\epsilon_i}{r_i}$ for all $n \geq m_i$. Hence $a_n > a - \frac{\epsilon_i}{r_i}$ and

$$a_n r_i > a r_i - \epsilon_i = a r_i - a r_i + d(x, y_i) = d(x, y_i),$$

it follows that $\mathbb{I}_{H_i^{a_n}}(x) = 1$, for all $n \geq m_i$. Thus, for large n, we have $\mathbb{I}_{H_i^{a_n}}(x) - \mathbb{I}_{H_i^a}(x) = 0$ for all $i < N$, which give us

$$|c_{a_n}(x) - c_a(x)| \leq \epsilon,$$

as we required. ■

Since c is injective, it possesses an inverse $c^{-1} : c(A \times S) \to A \times S$. Define h to be the projection of c^{-1} on its second coordinate. By Kuratowski's Theorem ([8], Theorem 10.5), the image of c is a Borel subset of $[0, 1/2]$ and the inverse map c^{-1} is Borel measurable. It follows that h is a Borel function.

Lemma 2.3.4 *Let $\{y_n\}_{n=1}^{\infty}$ be sequence converging to y, where $y_n, y \in c(A \times S)$. Then $h(y_n) \to h(y)$. In other words, h is continuous in $c(A \times S)$.*

Proof. Let $\{(a_n, x_n)\}_{n=1}^{\infty}$ be the correspondent sequence in $A \times S$, this is $c(a_n, x_n) = y_n$. Let (a, x) be such that $c(a, x) = y$, notice that $h(y_n) = x_n$ and $h(y) = x$. Define $\{i_k\}_{k=1}^{\infty}$ and $\{i_{nk}\}_{k=1}^{\infty}$ to be the ternary representation of y and y_n, so that

$$y = \sum_{k=1}^{\infty} \frac{i_k}{3^k} \quad \text{and} \quad y_n = \sum_{k=1}^{\infty} \frac{i_{nk}}{3^k}, \ \forall n \in \mathbb{N}.$$

We can also see i_{nk} as $\mathbb{I}_{H_k^{a_n}}(x_n)$ and i_k as $\mathbb{I}_{H_k^a}(x)$. The fact that $y_n \to y$ must imply that, eventually these first "ternary digits" of the elements in the sequence are equal to the digits in y. To prove it, let N be the first integer such that $i_{nN} \neq i_N$, then

$$\begin{aligned}
|y_n - y| &= \left| \sum_{k=1}^{\infty} \left(\frac{i_{nk} - i_k}{3^k} \right) \right| = \left| \frac{i_{nN} - i_N}{3^N} - \sum_{k=N+1}^{\infty} \left(\frac{i_k - i_{nk}}{3^k} \right) \right| \\
&\geq \left| \frac{i_{nN} - i_N}{3^N} \right| - \left| \sum_{k=N+1}^{\infty} \left(\frac{i_{nk} - i_k}{3^k} \right) \right| = \frac{1}{3^N} - \left| \sum_{k=N+1}^{\infty} \left(\frac{i_{nk} - i_k}{3^k} \right) \right| \\
&\geq \frac{1}{2 \cdot 3^N},
\end{aligned}$$

because $\sum_{k \geq N+1} \frac{1}{3^k} \leq \frac{1}{2 \cdot 3^N}$. Since $|y_n - y| \to 0$, we have $N \to \infty$, so that each term in the ternary representation of y_n and y will eventually be the same.

Now, we will prove that $x_n \to x$. Fix any $\epsilon > 0$. Take $z \in S$, the center of one of the open balls in $\{H_n\}_{n=1}^{\infty}$, such that $d(x,z) < \frac{\epsilon}{8}$. Let $k \in \mathbb{N}$ be such that $\frac{1}{k} < \frac{\epsilon}{8}$. Let $N \in \mathbb{N}$ be the index of $B\left(z, \frac{1}{k}\right)$ in $\{H_n\}_n$. It follows that

$$H_N = B\left(z, \frac{1}{k}\right) \subset H_N^{a'} \subset B(x, \epsilon), \quad \text{for all } a' \in A.$$

Since $\mathbb{I}_{H_N^a}(x) = 1$, we have that exists a large enough n^* such that $x_n \in H_N^{a_n}$ for all $n \geq n^*$. It follows that $x_n \in B(x, \epsilon)$, which implies

$$d(x_n, x) < \epsilon, \quad \text{for all } n \geq n^*.$$

Thus,

$$h\left(\lim_{n\to\infty} y_n\right) = h(y) = x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} h(x_n),$$

as we claimed before. ■

Finally, we introduce a variation of the function ρ constructed by Blackwell and Dubins. Define ρ, thus:

$$\rho(\mu, a, u) = \bar{h}[\rho^*(J_a(\mu), u)], \tag{2.3}$$

where the subscript a indicates the dependence of J on a, and $\bar{h}$ is given by

$$\bar{h}(y) = \begin{cases} h(y) & \text{if} \quad y \in c(A \times S) \\ x^* & \text{if} \quad y \notin c(A \times S). \end{cases}$$

where x^* is any element of S.

The change made to the original Blackwell and Dubins ρ consists in extending h, using $\bar{h}$ instead. This is due to the fact that the range of ρ^* may be larger than $c(A \times S)$, see Section 1.3.2.

In the following sections we show that the new ρ is a (C3)-continuous representation of $\mathcal{M}(S)$, but may fail to be (C2)-continuous.

2.3.1 Properties of ρ

As before, $\rho : \mathcal{M}(S) \times A \times (0,1) \to S$ is given by

$$\rho(\mu, a, u) = \bar{h}[\rho^*(J_a(\mu), u)].$$

Since ρ is a composition of several functions, it is enough to prove that the functions $\bar{h}$, ρ^*, and J_a are (jointly) measurable to prove that ρ is indeed measurable. Recall that the measurability of ρ^* and $\bar{h}$ has been already proved.

Lemma 2.3.5 $J : \mathcal{M}(S) \times A \to \mathcal{M}(\mathbb{R})$ *is jointly measurable.*

Proof. By Theorem 1.3.3, it is enough to prove that $J_a : \mathcal{M}(S) \to \mathcal{M}(\mathbb{R})$ is measurable for all $a \in A$, and $J_\mu : A \to \mathcal{M}(\mathbb{R})$ is a left continuous function.

- J_a is measurable. By Theorem 1.3.22 we only need to verify the measurability of $\varphi_f : \mathcal{M}(S) \to \mathbb{R}$, where $\varphi_f(\mu) = \int f dJ_a(\mu)$, $f \in C_b(\mathbb{R})$. Recall the definition of $J_a(\mu)$: the distribution of c_a under the measure μ. Then
$$\varphi_f(\mu) = \int_{\mathbb{R}} f dJ_a(\mu) = \int_S f(c_a) d\mu.$$
Since $f \in C_b(\mathbb{R})$, $f \circ c_a$ is a bounded function from S to $\mathbb{R}$. By the second example in Section 1.3.4 it is clear that φ_f is a measurable function.

- J_μ is left continuous. Let $a \in A$, take $\{a_n\}$, any increasing sequence that converges to a. Our claim is that $J_\mu(a_n)$ converges to $J_\mu(a)$, in the topology of $\mathcal{M}(\mathbb{R})$. Recall Theorem 1.3.19, which establishes that if $f_n, f : S \to \mathbb{R}$ are random variables such that $f_n \to f$ a.s. then its distribution functions converge weakly. Lemma 2.3.3 shows that $c_{a_n} \to c_a$ a.s., which implies $J_\mu(a_n) \to J_\mu(a)$.

This completes the proof. ∎

In summary, we have that all components of ρ ($\bar{h}, \rho^*$ and J_a) are measurable functions, which implies the measurability of ρ.

Now, we analyze the distribution of the random element $\rho_\mu : A \times U \to S$. As usual, the a-section of ρ is defined as the function $\rho_a : \mathcal{M}(S) \times (0,1) \to S$ determined by $\rho_a(\mu, u) = \rho(\mu, a, u)$. First, we will prove that for any $a \in A$ $\rho_a(\mu, \cdot)$ has distribution μ.

To analyze the distribution of $\rho_a(\mu, \cdot)$ define $Y_\mu : (0,1) \to S$ as the function given by $Y_\mu(u) = \rho_a(\mu, u)$. First of all, notice that Y_μ is a Borel function since it is a section of a measurable function, so it is a random element from $(0,1)$ to S. Then its distribution λ_{Y_μ} is given by

$$\lambda_{Y_\mu}(B) = \lambda(\{u : Y_\mu(u) \in B\}) = \lambda(Y_\mu^{-1}(B)),\ \ B \in \mathcal{B}(S),$$

where, as usual, λ is the Lebesgue measure in $(0,1)$. The class

$$\mathcal{C} = \{B \in \mathcal{B}(S) : x^* \notin B\}$$

is a π-system that generates $\mathcal{B}(S)$, so if we show that $\lambda(Y_\mu^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{C}$ it would imply that $\lambda \circ Y_\mu^{-1} = \mu$ in $\mathcal{B}(S)$.

Let $B \in \mathcal{C}$, then

$$\begin{aligned} Y_\mu^{-1}(B) &= \{u : Y_\mu(u) \in B\} \\ &= \{u : \bar{h}[\rho^*(J_a(\mu), u)] \in B\} \\ &= \{u : \bar{h}[\rho^*(J_a(\mu), u)] \in B, \rho^*(J_a(\mu), u) \in c(A \times S)\} \\ &\quad \cup \{u : \bar{h}[\rho^*(J_a(\mu), u)] \in B, \rho^*(J_a(\mu), u) \notin c(A \times S)\} \\ &= \{u : h[\rho^*(J_a(\mu), u)] \in B, \rho^*(J_a(\mu), u) \in c(A \times S)\} \\ &= \{u : \rho^*(J_a(\mu), u) \in c(A \times B)\}. \end{aligned}$$

Notice that, by Kuratowski's Theorem ([8], Theorem 10.5), $c(A \times B)$ is a Borel set.

On the other hand, since ρ^* is a representation of $\mathcal{M}(\mathbb{R})$, we have

$$\lambda(Y_\mu^{-1}(B)) = J_a(\mu)[c(A \times B)].$$

Moreover,

$$\begin{aligned} J_a(\mu)[c(A \times B)] &= (\mu \circ c_a^{-1})(c(A \times B)) \\ &= \mu(\{x : c_a(x) \in c(A \times B)\}) \\ &= \mu(B), \end{aligned}$$

where the last equality is justified by c being an injective function. We may now conclude that the distribution of Y_μ is actually μ.

Finally, since the conditional distribution of ρ given a is μ for each a, its unconditional distribution is μ.

2.3.2 (C3)-continuity of ρ

We start with an additional assumption, then we'll show that such assumption holds almost surely.

As before, let a be any number in A. Suppose that $\mu(\partial H_n^a) = 0$ for all $n \in \mathbb{N}$, our claim is that ρ_a is (C3)-continuous at μ, i.e. if $\mu_k \Rightarrow \mu$, then there exists $N^a_{\mu,\{\mu_n\}}$ with $\lambda(N^a_{\mu,\{\mu_n\}}) = 0$ such that $\rho_a(\mu_k, u)$ converges to $\rho_a(\mu, u)$ for every $u \notin N^a_{\mu,\{\mu_n\}}$.

Now, just for a moment, lets suppose that $J_a(\mu_k) \Rightarrow J_a(\mu)$, then it would imply $\rho^*(J_a(\mu_k), u) \to \rho^*(J_a(\mu), u)$ for almost every u since ρ^* is a (C2)-continuous representation of $\mathcal{M}(\mathbb{R})$. Define the null set $N = \{u : \rho^*(J_a(\mu_k), u) \not\to \rho^*(J_a(\mu), u)\}$.

If $u \notin N$ and $\rho^*(J_a(\mu_k), u), \rho^*(J_a(\mu), u) \in c(A \times S)$ then

$$\rho_a(\mu_k, u) = h[\rho^*(J_a(\mu_k), u)] \to h[\rho^*(J_a(\mu), u)] = \rho_a(\mu, u),$$

by Lemma 2.3.4. It is important to see that

$$\begin{aligned}
\lambda(\{u : \rho^*(J_a(\mu_k), u) \in c(A \times S)\}) &= J_a(\mu_k)[c(A \times S)] \\
&= \mu_k(\{x : c_a(x) \in c(A \times S)\}) \\
&= \mu_k(S) = 1,
\end{aligned}$$

for all k. Hence $\lambda(\{u : \rho^*(J_a(\mu_k), u), \rho^*(J_a(\mu), u) \in c(A \times S)\}) = 1$. Now let

$$N^a_{\mu,\{\mu_n\}} = N \cup \{u : \rho^*(J_a(\mu_k), u), \rho^*(J_a(\mu), u) \notin c(A \times S)\}.$$

Then, it is easy to see that $\rho_a(\mu_k, u)$ converges to $\rho_a(\mu, u)$ for every $u \notin N^a_{\mu,\{\mu_n\}}$. In conclusion, if we prove that $\mu_k \Rightarrow \mu$ implies $J_a(\mu_k) \Rightarrow J_a(\mu)$, for Lebesgue almost every $a \in A$, then we are done.

As proved before, in Lemma 2.3.2, c_a is continuous in the complement of $\bigcup_n \partial H_n^a$ and, by hypothesis,

$$\mu\left(\bigcup_n \partial H_n^a\right) \leq \sum_n \mu(\partial H_n^a) = 0.$$

Therefore, we can apply theorem 1.3.21 to get $\mu_k \circ c_a^{-1} \Rightarrow \mu \circ c_a^{-1}$. In other words,

$$J_a(\mu_k) \Rightarrow J_a(\mu).$$

It follows that, under the condition $\mu(\partial H_n^a) = 0 \quad \forall n$, the conditional probability given a of the discontinuity set of ρ_u for μ and $\{\mu_n\}$ is zero.

If we prove that, for any μ, the measure of the boundaries of H_n^a is zero for all but countably many a, we are done, since the unconditional probability of the discontinuity set of the map $\mu \mapsto \rho(\mu, a, u)$ for $\mu, \{\mu_n\}$ will be zero. It implies that the representation is (C3)-continuous at each μ.

To prove this initial hypothesis notice first that $\partial H_n^{a_1} \cap \partial H_n^{a_2} = \emptyset$ for $a_1 \neq a_2$. It implies that for a fixed n, the set of $\{a : \mu(\partial H_n^a) > 0\}$ is indeed countable. Then, the equality

$$\{a : \mu(\partial H_n^a) > 0, n \in \mathbb{N}\} = \bigcup_n \{a : \mu(\partial H_n^a) > 0\}$$

shows that it is possible to rewrite the first set as a countable union of countable sets, which proves our claim.

2.4 Fernique's Extension

In 1988 a paper was published by X. Fernique ([10]) giving an explicit way to construct a (C2)-continuous representation of $\mathcal{M}(S)$, for any Polish space S. We start with some definitions and notation. As before, S is a Polish space with metric d. Recall that for $\epsilon > 0$, $F \subset S$

$$F_\epsilon = \{x \in S : d(x, F) < \epsilon\}.$$

The Prohorov metric on $\mathcal{M}(S)$, denoted by d_P, is given by

$$d_P(\mu, \nu) = \inf \{\epsilon > 0 : \text{ for all closed set } F, \ \mu(F) \leq \nu(F_\epsilon) + \epsilon\},$$

see (1.4), Section 1.3.5. In this section we explain Fernique's (C2)-continuous representation of $\mathcal{M}(S)$. Blackwell and Dubins representation uses a function from Polish to a more particularly studied space, the interval $[0, 1]$. Such construction is too "discontinuous" and fails to capture the structure of S. Fernique regularize it using continuous partitions of unity, (see Section 1.1.3). Actually, this representation does not use a second parameter as the first representation where the domain of ρ is $\mathcal{M}(S) \times ([1, 2] \times [0, 1])$.

In this section $(U, \mathcal{F}, \lambda)$ denotes $[0,1)$ with its usual topology and the Lebesgue measure. For any integer $n \geq 0$, let

$$\mathcal{T}_n = \{O_{n,k}\,,\ k = 0, 1, 2, 3, \ldots\}$$

be a sequence, ordered by k, of open sets of diameter less or equal to 2^{-n} covering S. Let $F_n = \{f_t, t \in \mathcal{T}_n\}$ be a continuous partition of unity dominated by $\mathcal{T}_n$ (existence is given by Theorem 1.1.2), provided with the order induced by its index. The partitions F_n are in general not comparable, in the sense that there is not a clear relationship between the sets in $\mathcal{T}_n$ for different values of n. Let

$$\mathcal{G}^n = F_1 \times \cdots \times F_n.$$

For simplicity, sometimes we write $(a_1, \ldots, a_n) = (f_{a_1}, \ldots, f_{a_n})$, where $f_{a_i} \in F_i$. We equip $\mathcal{G}^n$ with the dictionary (or lexicographic) order associated with the orders of its factors: $(a_1, a_2, \ldots, a_n) < (b_1, b_2, \ldots, b_n)$ iff any of the following is true

- $a_1 < b_1$
- $a_1 = b_1$ and $a_2 < b_2$
- $a_1 = b_1$, $a_2 = b_2$ and $a_3 < b_3$

  $\vdots$
- $a_i = b_i$ for $i = 1, \ldots, n-1$ and $a_n < b_n$.

We may identify $\mathcal{G}^n$ with the ordinal space $[0, \omega^n)$ as follows,

$$\begin{aligned}
(0,0,\ldots,0,0,0) &= 0 \\
(0,0,\ldots,0,0,1) &= 1 \\
(0,0,\ldots,0,0,2) &= 2 \\
&\vdots \\
(0,0,\ldots,0,0,k) &= k \\
&\vdots \\
(0,0,\ldots,0,1,0) &= \omega \\
(0,0,\ldots,0,1,1) &= \omega + 1
\end{aligned}$$

$$
\begin{array}{rcl}
(0,0,\ldots,0,1,2) & = & \omega+2 \\
& \vdots & \\
(0,0,\ldots,0,2,0) & = & \omega+\omega=\omega\cdot 2 \\
& \vdots & \\
(0,0,\ldots,0,k,m) & = & \omega\cdot k+m \\
& \vdots & \\
(0,0,\ldots,1,0,0) & = & \omega\cdot\omega=\omega^2
\end{array}
$$

and, in general

$$(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} \omega^{n-i} \cdot a_i.$$

At any element g of $\mathcal{G}^n$, we associate the product

$$p(g) = f_{a_1} \cdot f_{a_2} \cdots f_{a_n}.$$

As in Section 1.1.3 the support of $p(g)$ is the set $\overline{\{x \in S : p(g)\,(x) \neq 0\}}$. Denote by $t(g)$ an element in the support of $p(g)$. Notice that

$$\operatorname{supp} p(g) \subset B(t(g), 2^{-n}), \tag{2.4}$$

where $B(t(g), 2^{-n})$ represents the open ball with center at $t(g)$ and radius 2^{-n}. For every pair (m, n) of integers larger or equal to 0, let $q^{(m+n,n)}$ be the projection of $\mathcal{G}^{n+m}$ on $\mathcal{G}^n$, i.e.

$$q^{(n+m,n)}(f_{a_1}, f_{a_2}, \ldots, f_{a_{n+m}}) = (f_{a_1}, f_{a_2}, \ldots, f_{a_n}), \qquad f_{a_i} \in F_i.$$

Lemma 2.4.1 *For any $n \in \mathbb{N}$ the set $\{p(g), g \in \mathcal{G}^n\}$ is a continuous partition of unity. Moreover, $\{p(g), g \in \mathcal{G}^{n+1}\}$ is finer than $\{p(g), g \in \mathcal{G}^n\}$, meaning that*

$$p(g) = \sum \{p(h) : h \in \mathcal{G}^{n+1}, q^{(n+1,n)}(h) = g\}, \;\; \forall g \in \mathcal{G}^n. \tag{2.5}$$

Proof. Clearly, $p(g)$ is continuous and has range in $[0, 1]$. Fix any $x \in S$, since F_1 is a partition of unity $f(x) = 0$ for all but finitely many $f \in F_1$.

The same is true for $k = 1, \ldots, n$, therefore $p(g)\,(x) = 0$ for all but finitely many $g \in \mathcal{G}^n$.

To prove (2.5), let $g = (f_{a_1}, \ldots, f_{a_n}) \in \mathcal{G}^n$ where $f_{a_i} \in F_i$, then

$$\left\{h \in \mathcal{G}^{n+1} : q^{(n+1,n)}(h) = g\right\} = \{(f_{a_1}, \ldots, f_{a_n}, h_i) : h_i \in F_{n+1}\}.$$

It follows that

$$\begin{aligned}
\sum \left\{p(h) : h \in \mathcal{G}^{n+1}, q^{(n+1,n)}(h) = g\right\} &= \sum_{i=0}^{\infty} p(f_{a_1}, \ldots, f_{a_n}, h_i)\,(x) \\
&= \sum_{i=0}^{\infty} f_{a_1}(x) \cdots f_{a_n}(x) \cdot h_i(x) \\
&= f_{a_1}(x) \cdots f_{a_n}(x) \sum_{i=0}^{\infty} h_i(x) \\
&= f_{a_1}(x) \cdots f_{a_n}(x) \cdot 1 = p(g)\,(x).
\end{aligned}$$

Finally, to show $\sum\limits_{g \in \mathcal{G}^n} p(g)\,(x) = 1$, notice that

$$\begin{aligned}
\sum_{g \in \mathcal{G}^n} p(g)\,(x) &= \sum_{g \in \mathcal{G}^{n-1}} \sum_{f \in F_n} p(g)\,(x) \cdot f(x) = \sum_{g \in \mathcal{G}^{n-1}} p(g)\,(x), \qquad \text{by (2.5)} \\
&= \sum_{g \in \mathcal{G}^{n-2}} \sum_{f \in F_{n-1}} p(g)\,(x) \cdot f(x) = \sum_{g \in \mathcal{G}^{n-2}} p(g)\,(x) \\
&= \cdots = \sum_{g \in \mathcal{G}^1} p(g)\,(x) = \sum_{g \in F_1} g(x) = 1,
\end{aligned}$$

which completes the proof. ■

For any element g of $\mathcal{G}^n$, set

$$\Sigma_n(g) = \sum_{h \in \mathcal{G}^n,\, h < g} p(h).$$

Clearly, $\Sigma_n : \mathcal{G}^n \to [0, 1)$ is an increasing sequence (recall that we equipped $\mathcal{G}^n$ with the dictionary order).

Lemma 2.4.2 *For any integer $n \geq 1$ and all $g \in \mathcal{G}^n$, it is possible to simplify*

$$\Sigma_n(g) = \Sigma_{n-1}(g') + p(g') \sum_{f \in F_n,\, f < g_n} f, \tag{2.6}$$

where $g' = q^{(n,n-1)}(g)$ and g_n is the n^{th} entry of g.

Proof. Suppose that g has no predecessor in $\mathcal{G}^n$, i.e. g is of the form $\sum_{i=1}^{n-1} \omega^i \cdot a_i$, $a_i \neq 0$ for at least one i. In that case

$$g = (\underbrace{f_{a_1}, \ldots, f_{a_{n-1}}}_{g'}, f_0) = (g', f_0),$$

and the equation (2.6) is trivially satisfied.

Now, suppose that g has predecessor in $\mathcal{G}^n$, i.e. g is of the form (g', k). Notice that, by $g = (g', k)$ we mean that the nth entry in $g \in \mathcal{G}^n = F_1 \times \cdots \times F_n$ is the function $f_k \in F_n = \{f_i\}_{i=0}^{\infty}$. In this case

$$\begin{aligned}
\Sigma_n(g) &= \sum_{h \in \mathcal{G}^n,\, h < g} p(h) = \sum_{h < (g',0)} p(h) + \sum_{i=0}^{k-1} p((g', i)) \\
&= \sum_{h \in \mathcal{G}^{n-1},\, h < g'} p(h) + \sum_{i=0}^{k-1} p(g') \cdot f_i, \qquad f_i \in F_n \\
&= \Sigma_{n-1}(g') + p(g') \cdot \sum_{i=0}^{k-1} f_i,
\end{aligned}$$

as we wanted to show. ∎

By induction, this implies in particular that the functions $\Sigma_n(g)$ are continuous and bounded on S. As we mentioned before, some elements of $\mathcal{G}^n$ have no precedent, but every element g has a successor g^+. Let μ be a probability measure on S. Fix an integer $n \geq 0$ and denote by $\mu_n \in \mathcal{M}(S)$ the probability measures defined by

$$\mu_n = \sum_{g \in \mathcal{G}^n} \left(\int p(g)\, d\mu \right) \delta_{t(g)}.$$

Our claim is that the sequence in $\mathcal{M}(S)$ constructed in this way converges weakly to μ.

Lemma 2.4.3 *For all n, $d_P(\mu_n,\mu) \le 2^{-n}$.*

Proof. Let $F \subset S$ be a closed set, in order to prove the lemma we must show $\mu_n(F) \le \mu(F_{2^{-n}}) + 2^{-n}$. Define the following sets

$$\begin{aligned} A_F &= \{g \in \mathcal{G}^n : t(g) \in F\} \\ B_F &= \{g \in \mathcal{G}^n : \operatorname{supp} p(g) \cap F \neq \emptyset\}, \end{aligned}$$

clearly $A_F \subset B_F$. Then

$$\mu_n(F) = \sum_{g \in A_F} \left(\int p(g) d\mu \right) \le \sum_{g \in B_F} \left(\int p(g) d\mu \right) = \int \sum_{g \in B_F} p(g) d\mu.$$

By (2.4), $g \in B_F$ implies $\operatorname{supp} p(g) \subset F_{2^{-n}}$. Then

$$\int \sum_{g \in B_F} p(g) d\mu = \int_{F_{2^{-n}}} \sum_{g \in B_F} p(g) d\mu \le \int_{F_{2^{-n}}} 1 d\mu = \mu(F_{2^{-n}}).$$

In summary

$$\mu_n(F) \le \mu(F_{2^{-n}}) < \mu(F_{2^{-n}}) + 2^{-n},$$

for all closed set F. ∎

A simple recurrence, based on (2.6) after integration and the fact that, as g increases $\Sigma_n(g) \uparrow 1$ (because $\sum_{g \in \mathcal{G}^n} p(g)$ is the constant 1 in S), show that for every $x \in [0,1)$ there exists a unique element g of $\mathcal{G}^n$, denoted by g_x, such that

$$\int \Sigma_n(g_x)\, d\mu \le x < \int \Sigma_n(g_x^+)\, d\mu.$$

Define the function X_n from $[0,1)$ to S by putting $X_n(x) = t(g_x)$ for all x in $[0,1)$.

Lemma 2.4.4 *X_n is a well defined measurable function. Moreover $X_n \overset{D}{=} \mu_n$.*

Proof. First notice that $\{t(g) : g \in \mathcal{G}^n\}$ is a countable set and

$$X_n^{-1}\{t(g)\} = \{x \in [0,1) : X_n(x) = t(g)\} = \bigcup_{\{h:t(h)=t(g)\}} \left[\int \Sigma_n(h)\, d\mu, \int \Sigma_n(h^+)\, d\mu\right),$$

therefore

$$\begin{aligned} \lambda(X_n = t(g)) &= \sum_{\{h:t(h)=t(g)\}} \left(\int \Sigma_n(h^+)\, d\mu - \int \Sigma_n(h)\, d\mu\right) \\ &= \sum_{\{h:t(h)=t(g)\}} \int [\Sigma_n(h) + p(h) - \Sigma_n(h)]\, d\mu \\ &= \sum_{\{h:t(h)=t(g)\}} \int p(h)\, d\mu \\ &= \mu_n(t(g)). \end{aligned}$$

∎

Lemma 2.4.5 *For every $x \in [0,1)$ the sequence $\{X_n(x)\}_{n=1}^{\infty}$ is a Cauchy sequence in the complete space (S,d). Then it converges to a limit $X(x)$, satisfying*

$$d\Big(X_n(x), X(x)\Big) \le 2^{-n}, \qquad \forall x \in [0,1),\ \forall n \in \mathbb{N}.$$

Proof. Suppose $X_n(x) = t(g)$ and $X_{n+1}(x) = t(h)$, then

$$\int \Sigma_n(g)\, d\mu \le x < \int \Sigma_n(g^+)\, d\mu \quad \text{and} \quad \int \Sigma_{n+1}(h)\, d\mu \le x < \int \Sigma_{n+1}(h^+)\, d\mu.$$

Let

$$h_i = (g, f_i) = (\underbrace{g_{a_1}, \ldots, g_{a_n}}_{g}, f_i), \qquad f_i \in F_{n+1} = \{f_i\}_{i=0}^{\infty}.$$

By (2.6), $\Sigma_{n+1}(h_0) = \Sigma_n(g) \le x$ and $\Sigma_{n+1}(h_i) = \Sigma_n(g) + p(g) \sum_{j<i} f_j$. It follows that

$$\lim_{k\to\infty} \Sigma_{n+1}(h_k) = \Sigma_n(g) + p(g) \sum_j f_j = \Sigma_n(g^+).$$

Since $x < \Sigma_n(g^+)$, there exists k so that

$$\Sigma_{n+1}(h_k) \leq x < \Sigma_{n+1}(h_k^+),$$

which implies $q^{(n+1,n)}(h) = g$. Hence $\operatorname{supp} p(h) \subset \operatorname{supp} p(g) \subset B(t(g), 2^{-n})$. Notice that the last contention is given by (2.4). Therefore

$$d(t(h), t(g)) = d(X_n(x), X_{n+1}(x)) \leq 2^{-n}.$$

Similarly, if $X_{n+2}(x) = t(h')$ then

$$q^{(n+2,n)}(h') = q^{(n+1,n)}[q^{(n+2,n+1)}(h')] = q^{(n+1,n)}(h) = g.$$

In general, $d(X_n(x), X_{n+m}(x)) \leq 2^{-n}$. Finally, let

$$X(x) = \lim_{n\to\infty} X_n(x).$$

It follows that $d(X_n(x), X(x)) \leq 2^{-n}$. ∎

The function $X : [0,1) \to S$ defined above is measurable, and because of the relationship between its law and the sequence $\{\mu_n\}$ (which converges weakly to μ) we know $X \stackrel{D}{=} \mu$. We set $X = X(\mu)$ and $X_n = X(\mu, n)$.

For any integer $n \geq 0$ and any probability measure μ on S, let

$$N(\mu, n) = \left\{ \int \Sigma_n(g)\, d\mu : g \in \mathcal{G}^n \right\},$$

Lemma 2.4.6 *The set $N(\mu, n)$ is a closed subset in $[0,1)$ and*

$$N(\mu) = \bigcup_{n\in\mathbb{N}} N(\mu, n)$$

is a countable null subset of $[0,1)$.

Proof. Let $\Psi : \mathcal{G}^n \cup \{\omega^n\} \to [0,1]$ be the function defined by

$$\begin{aligned} \Psi(g) &= \int \sum_{f<g} p(f)\, d\mu = \int \Sigma_n(g)\, d\mu, \qquad g \in \mathcal{G}^n, \\ \Psi(\omega^n) &= \int \sum_{f\in\mathcal{G}^n} p(f)\, d\mu = 1. \end{aligned}$$

Since $p(g) \geq 0$ in S for all $g \in \mathcal{G}^n$, the function can be interpreted as an increasing ordinal indexed sequence. In order to prove continuity of Φ we only have to show $\Phi(g) \uparrow \Phi(f)$ whenever $g \uparrow f$, where f is a limit ordinal. Notice that f must be either ω^n or an element of $\mathcal{G}^n$ of the form

$$(\underbrace{f_{a_1}}_{\in F_1}, \ldots, \underbrace{f_{a_{n-1}}}_{\in F_{n-1}}, \underbrace{f_0}_{\text{first in } F_n}) = (a_1, \ldots, a_{n-1}, 0) = \sum_{i=1}^{n-1} \omega^{n-i} \cdot a_i$$

when $a_i \neq 0$ for at least one i. Suppose $f \neq \omega^n$. Take any $g < f$, then

$$\Sigma_n(f) - \Sigma_n(g) = \sum_{h \in [g,f)} p(h).$$

Fix $x \in S$, since $\{p(g) : g \in \mathcal{G}^n\}$ is a continuous partition of unity we have that $p(g)\,(x) = 0$ for all but finitely many g. It follows that if g is close enough to f then $[\Sigma_n(f) - \Sigma_n(g)](x) = 0$. In other words, $\Sigma_n(f) - \Sigma_n(g)$ decreases pointwise to zero as $g \uparrow f$. Then

$$\Phi(f) - \Phi(g) = \int [\Sigma_n(f) - \Sigma_n(g)]\, d\mu \to 0, \quad \text{as } g \uparrow f.$$

Similarly, it can be proved that Φ is continuous at $f = \omega^n$. We use the fact that $[0, \omega^n]$ is a compact set to prove that the direct image of Φ is compact, thus closed, in $[0,1]$. That is,

$$\begin{aligned} \Phi(\mathcal{G}^n \cup \{\omega^n\}) &= \left\{ \int \Sigma_n(g)\, d\mu : g \in \mathcal{G}^n \right\} \cup \{1\} \\ &= N(\mu, n) \cup \{1\} \end{aligned}$$

is closed, which implies that $N(\mu, n)$ is closed in $[0, 1)$.

Finally, it is immediate that $N(\mu, n)$ and $N(\mu)$ are countable null sets. ∎

Let μ be an element of $\mathcal{M}(S)$ and let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence converging weakly to μ. Take any $x \notin N(\mu)$ and any $\epsilon > 0$. Then choose an integer n such that $2^{-n+1} < \epsilon$. In particular $x \notin N(\mu, n)$, therefore there exists $\delta > 0$ and $g \in \mathcal{G}^n$ such that

$$\int \Sigma_n(g)\, d\mu < x - \delta < x + \delta < \int \Sigma_n(g^+)\, d\mu.$$

The sums $\Sigma_n(g)$ and $\Sigma_n(g^+)$ are bounded continuous functions, therefore $\mu_i \Rightarrow \mu$ implies $\int \Sigma_n(g)\, d\mu_i \to \int \Sigma_n(g)\, d\mu$ and $\int \Sigma_n(g^+)\, d\mu_i \to \int \Sigma_n(g^+)\, d\mu$. Then, there exists $k \in \mathbb{N}$ so that

$$\int \Sigma_n(g)\, d\mu_i \leq x < \int \Sigma_n(g^+)\, d\mu_i \quad \text{for all } i \geq k,$$

which give us the equality $X(\mu_i, n)(x) = X(\mu, n)(x)$. Moreover,

$$\begin{aligned} d\Big(X(\mu_i)(x), X(\mu)(x)\Big) &\leq d\Big(X(\mu_i)(x), X(\mu_i, n)(x)\Big) + d\Big(X(\mu_i, n)(x), X(\mu)(x)\Big) \\ &= d\Big(X(\mu_i)(x), X(\mu_i, n)(x)\Big) + d\Big(X(\mu, n)(x), X(\mu)(x)\Big) \\ &= 2^{-n} + 2^{-n} = 2^{-n+1} \leq \epsilon. \end{aligned}$$

It shows that the sequence $\{X(\mu_i)(x)\}_{i=1}^{\infty}$ converges to $X(\mu)(x)$.

Finally, Fernique's (C2)-continuous representation $\rho : \mathcal{M}(S) \times [0, 1) \to S$ is given by

$$\rho(\mu, x) = X(\mu, x).$$

The previous lemmas confirm that ρ has the properties required to be (C2)-continuous. The only thing we must show is the measurability of ρ.

Lemma 2.4.7 *The function $\rho : \mathcal{M}(S) \times [0, 1) \to S$ is jointly measurable.*

Proof. Since $\rho(\mu, x) = \lim_{n\to\infty} X_n(\mu, x)$, where $X_n(\mu, x) = X(\mu, n)(x)$, it is enough to prove that $X_n : \mathcal{M}(S) \times [0, 1) \to S$ is jointly measurable for all n. Fix any $n \in \mathbb{N}$ and, for simplicity, let $\hat{\rho}(\mu, x) = X_n(\mu, x)$. Then we show

- $\hat{\rho}_x : \mathcal{M}(S) \to S$ is measurable,
- $\hat{\rho}_\mu : [0, 1) \to S$ is right continuous.

By Theorem 1.3.3 the joint measurability of $\hat{\rho}$ is implied.

To show the first part recall $\hat{\rho}(\mu, x) = t(g_x)$, where g_x is the unique element of $\mathcal{G}^n$ such that

$$\int \Sigma_n(g_x)\, d\mu \leq x < \int \Sigma_n(g_x^+)\, d\mu.$$

Since the range of $\hat{\rho}_x$ is a subset of $\{t(g) : g \in \mathcal{G}^n\} \subset S$, we can conclude that $\hat{\rho}_x$ only takes countable many values in S. Hence, to verify measurability, we only need to check whether or not the set $\hat{\rho}_x^{-1}\{t(g)\}$ is measurable for all $g \in \mathcal{G}^n$. Notice that,

$$\begin{aligned}\hat{\rho}_x^{-1}\{t(g)\}) &= \{\mu \in \mathcal{M}(S) : \hat{\rho}_x(\mu) = t(g)\} \\ &= \left\{\mu \in \mathcal{M}(S) : \int \Sigma_n(f)\, d\mu \leq x < \int \Sigma_n(f^+)\, dm, \quad t(f) = t(g)\right\}.\end{aligned}$$

On the other hand, the set

$$\left\{\mu \in \mathcal{M}(S) : \int \Sigma_n(f)\, d\mu \leq x < \int \Sigma_n(f^+)\, d\mu\right\}$$

can be rewritten as

$$\left\{\mu \in \mathcal{M}(S) : \int \Sigma_n(f)\, d\mu \leq x\right\} \bigcap \left\{\mu \in \mathcal{M}(S) : x < \int \Sigma_n(f^+)\, d\mu\right\},$$

the intersection of two measurable subsets of $\mathcal{M}(S)$ (example 3, Section 1.3.4). This proves that $\hat{\rho}_x^{-1}$ is measurable.

Now, we prove the second part. Take any $\mu \in \mathcal{M}(S)$ and fix $x \in [0, 1)$, we must show that for any decreasing sequence $\{x_k\}_{k=1}^{\infty}$ converging to x we have $\hat{\rho}_\mu(x_k) \to \hat{\rho}_\mu(x)$. Let

$$\epsilon = \int \Sigma_n(g_x^+)\, d\mu - x > 0.$$

Find k^* such that $|x - x_k| = x_k - x < \epsilon$ for all $k \geq k^*$. Since $\int \Sigma_n(g_x)\, d\mu \leq x \leq x_k$, we can conclude

$$\int \Sigma_n(g_x)\, d\mu \leq x_k < \int \Sigma_n(g_x^+)\, d\mu \qquad \text{for all } k \geq k^*.$$

The fact that $x \mapsto g_x$ is uniquely determined implies $g_{x_k} = g_x$ for all $k \geq k^*$. That is, $\hat{\rho}_\mu(x_n) = \hat{\rho}_\mu(x)$ for a large enough k, and the proof is complete. ∎

Chapter 3

Liftings with Fixed Boundary Values

3.1 Construction of Liftings with Fixed Boundary

In this chapter we present another variation of Skorohod's theorem by Jean Cortissoz (see [5]). We begin with some notation: let $L^0(\Omega, S)$ be the space of S-valued random elements from Ω, where S is a Polish space. Suppose α is a continuous path from $[0,1]$ to $\mathcal{M}(S)$, then we ask ourselves: is it possible to find a continuous $\overline{\alpha} : [0,1] \to L^0(\Omega, S)$ so that the distribution of $\overline{\alpha}(t)$ is $\alpha(t)$ for all $t \in [0,1]$? Notice that we ask continuity for the path $\overline{\alpha}$ but we have not yet specified a topology in $L^0(\Omega, S)$. If by "continuous" we mean that whenever $t_n \to t$ we must have $\overline{\alpha}(t_n) \to \overline{\alpha}(t)$ a.s., then knowing the existence of a (C2)-continuous representation ρ and putting $\Omega = [0,1]$, we can easily take

$$\overline{\alpha}(t) = \rho(\alpha(t), \cdot) \tag{3.1}$$

to obtain the desired result. Indeed, there is no need to restrict ourselves to the case $\Omega = [0,1]$. It is possible to generalize equation (3.1) when $(\Omega, \mathcal{F}, P)$ is a **non atomic** probability space. A subset A of Ω is an **atom** if $\forall B \in \mathcal{F}, B \subset A$ we have either $P(B) = P(A)$ or $P(B) = 0$. If there is no atom then the probability space is said to be non atomic.

Lemma 3.1.1 *Let $(\Omega, \mathcal{F}, P)$ be non atomic. For any $t \in [0,1]$ there exists $E_t \in \mathcal{F}$ satisfying $P(E_t) = t$.*

A proof for the lemma above can be found in [4] (Corollary 1.12.10, p.56) or [14].

Lemma 3.1.2 *If Ω is non atomic then there exists a $[0,1]$-uniform random variable X from Ω. That is, the distribution of X, $P \circ X^{-1}$, is equal to the Lebesgue measure λ in $[0,1]$.*

Proof. Let $\{q_n\}_{n=1}^{\infty}$ be an enumeration of $\mathbb{Q} \cap (0,1)$. Inductively we construct $E_{q_1}, E_{q_2}, \ldots$ so that

$$P(E_s) = s \quad \text{and} \quad E_s \subset E_t \text{ iff } s \leq t.$$

To do so, start with E_{q_1} satisfying $P(E_{q_1}) = q_1$. If $q_2 < q_1$ take $E_{q_2} \subset E_{q_1}$ with $P(E_{q_2}) = q_2$. If $q_2 > q_1$ find $A \subset \Omega \backslash E_{q_1}$ with $P(A) = q_2 - q_1$ and take $E_{q_2} = E_{q_1} \cup A$. Following that method we can get sets with the desired properties. In addition we ask $E_0 = \emptyset$ and $E_1 = \Omega$.

Define $X : \Omega \to [0,1]$ as

$$X(\omega) = \inf\{q : \omega \in E_q\}.$$

The set $X^{-1}[0,q]$ is given by $\bigcap\limits_{t>q,\, t\in\mathbb{Q}} E_t$. It follows that X is indeed a measurable function with distribution λ, because for all $s > q$

$$E_q \subset \bigcap_{t>q,\, t\in\mathbb{Q}} E_t \subset E_s,$$

then $P\left(\bigcap\limits_{t>q,\, t\in\mathbb{Q}} E_t\right) = q.$ ∎

As a consequence, (3.1) can be transformed into

$$\overline{\alpha}(t) = \rho(\alpha(t), X(\omega)).$$

Clearly it meets the requirements described in the first paragraph. From now on Ω is non atomic.

Another question that rises when looking for a path $\overline{\alpha}$ is if one can fix the endpoints at $t = 0$ and $t = 1$. The previous path does not satisfy this extra condition, because once we evaluate ρ at $\alpha(t) \in \mathcal{M}(S)$ we get a random element whose law is indeed $\alpha(t)$, but we do not get to choose the random element we wanted from all possible $X \in L^0(\Omega, S)$ having this law. It makes us wonder if convergence a.s. in $L^0(\Omega, S)$ is too "strong" and if we should consider other topologies in this set.

Let $X_n, X : (\Omega, \mathcal{F}, P) \to S$, we say that $\{X_n\}_n$ **converges in probability** to X ($X_n \overset{P}{\to} X$) if for every $\epsilon > 0$

$$\lim_{n\to\infty} P\{\omega : d(X_n(\omega), X(\omega)) \geq \epsilon\} = 0.$$

As before, let $\alpha : [0,1] \to \mathcal{M}(S)$ be a continuous path. A **lifting of** α is a continuous function $\overline{\alpha} : [0,1] \to L^0(\Omega, S)$ ($L^0(\Omega, S)$ endowed with the topology of the convergence in probability) such that, for all $t \in [0,1]$, the law of $\overline{\alpha}(t)$ is $\alpha(t)$.

Cortissoz was able to prove the existence of liftings with fixed endpoints at 0 and 1. That is, if X_0 and X_1 are S-valued random elements defined on $(\Omega, \mathcal{F}, P)$, a non atomic probability space, with distributions $\alpha(0)$ and $\alpha(1)$ respectively, then we can construct a lifting $\overline{\alpha}$ for which $\overline{\alpha}(0) = X_0$ and $\overline{\alpha}(1) = X_1$. This chapter is devoted to proving this statement.

Let $\hat{d} : L^0(\Omega, S) \times L^0(\Omega, S) \to [0,1]$ be given by

$$\hat{d}(X,Y) = \inf\{\, \epsilon > 0 : P\{\omega : d(X(\omega), Y(\omega)) \geq \epsilon\} \leq \epsilon \,\}.$$

The metric $\hat{d}$ is called the Ky Fan metric, [7] (p. 226).

Lemma 3.1.3 *Given a sequence $\{X_n\}_{n=1}^{\infty}$ and X an element of $L^0(\Omega, S)$*

$$X_n \overset{P}{\to} X \quad \textit{iff} \quad \lim_{n\to\infty} \hat{d}(X_n, X) = 0.$$

Proof. Suppose $\hat{d}(X_n, X) \to 0$. We must show $P\{d(X_n, X) \geq \epsilon\} \to 0$ for any $\epsilon > 0$. Take $\epsilon' \in (0, \epsilon)$, since $\lim_{n\to\infty} \hat{d}(X_n, X) = 0$ there exists $N \in \mathbb{N}$ so that $P\{d(X_n, X) \geq \epsilon'\} \leq \epsilon'$ for all $n \geq N$. Then

$$P\{d(X_n, X) \geq \epsilon\} \leq P\{d(X_n, X) \geq \epsilon'\} \leq \epsilon', \quad \forall n \geq N.$$

Therefore,

$$\lim_{n\to\infty} P\{d(X_n, X) \geq \epsilon\} = 0,$$

which implies $X_n \xrightarrow{P} X$.

On the other hand, suppose $X_n \xrightarrow{P} X$ and fix any $\epsilon^* > 0$. Since $P\{d(X_n, X) \geq \epsilon^*\} \to 0$ we can find $N \in \mathbb{N}$ so that $\forall n \geq N$

$$P\{d(X_n, X) \geq \epsilon^*\} \leq \epsilon^*.$$

Hence,

$$\hat{d}(X_n, X) = \inf\{\, \epsilon > 0 : P\{\omega : d(X_n(\omega), X(\omega)) \geq \epsilon\} \leq \epsilon \,\} \leq \epsilon^*.$$

It follows that $\hat{d}(X_n, X) \to 0$. ∎

The Prohorov metric d_P in $\mathcal{M}(S)$ and $\hat{d}$ in $L^0(\Omega, S)$ are related through Theorem 1.3.24, which establishes

$$d_P(\mu, \nu) = \inf_{m \in \mathcal{P}(\mu,\nu)} \inf\{\epsilon > 0 : m(\{(x, y) : d(x, y) \geq \epsilon\}) \leq \epsilon\}.$$

The relation looks more evident in the equality

$$d_P(\mu, \nu) = \inf\{\hat{d}(X, Y) : \text{ law of } X = \mu, \text{ law of } Y = \nu\}. \tag{3.2}$$

As a consequence, convergence in probability of X_n to X implies weak convergence of their distributions. So, if we see "law" as a function from $L^0(\Omega)$ to $\mathcal{M}(S)$ this function is continuous.

We say that $\mu \in \mathcal{M}(S)$ is **finitely supported** if it can be rewritten as

$$\mu = \sum_{i=1}^{n} c_i \delta_{a_i}\,, \quad c_i \geq 0, \;\; \{a_i\}_{i=1}^n \subset S.$$

Let $\{A_i\}_{i=1}^n$ be a measurable partition of Ω. The random element X that assigns a_i for all elements of A_i, written as

$$X = \sum_{i=1}^{n} a_i \mathbb{I}_{A_i},$$

is said to be a **simple** S-valued random element.

Lemma 3.1.4 *Let μ and ν be finitely supported measures. Let*

$$0 \leq d_P(\mu, \nu) < \epsilon.$$

Then, given X a random variable such that law$(X) = \mu$, there exists Y whose law is ν satisfying

$$\hat{d}(X, Y) < \epsilon.$$

Proof. Since μ and ν are finitely supported, by (3.2) we can find

$$X' = \sum_{i=1}^{m} a_i \mathbb{I}_{A_i'} \quad \text{and} \quad Y' = \sum_{i=1}^{m} a_i \mathbb{I}_{B_i'},$$

where $\{a_1, \dots, a_m\}$ is the union of the supports of μ and ν, so that law $(X') = \mu$, law $(Y') = \nu$ and $\hat{d}(X', Y') < \epsilon$. Write $X = \sum_{i=1}^{m} a_i \mathbb{I}_{A_i}$. By Lemma 3.1.1, we can find measurable sets $\{B_{ij}\}_{j=1}^{m}$, $i = 1, \dots, m$, such that

$$B_{ij} \subset A_i, P(A_i \cap B_{ij}) = P(A_i' \cap B_j') \quad \text{and} \quad A_i = \bigcup_{j=1}^{m} B_{ij}.$$

Then, take

$$B_j = \bigcup_{i=1}^{m} B_{ij}\,, \qquad Y = \sum_{j=1}^{m} a_j \mathbb{I}_{B_j}.$$

Since $P(B_j) = P(B_j')$, we must have law $(Y) = \nu$. Moreover $\hat{d}(X, Y) = \hat{d}(X', Y') < \epsilon$. ∎

Lemma 3.1.5 *Assume $d_P(law(X), law(Y)) < \epsilon$, for a given $\epsilon > 0$. Then there is Y' such that $law(Y') = law(Y)$ and $\hat{d}(X, Y') < \epsilon$.*

Proof. The proof will be split in three cases:

- X and Y are simple. Already proved in previous lemma.

- X is arbitrary and Y is simple. Take $\delta > 0$ small enough so that

  $$d_P(\text{law}\,(X), \text{law}\,(Y)) < \epsilon - \delta.$$

  Since the set of simple random variables is dense in $L^0(\Omega, S)$, and the function "law" is continuous, we can choose X' a simple random variable such that

  $$\hat{d}(X, X') < \delta \qquad \text{and} \qquad d_P(\text{law}\,(X'), \text{law}\,(Y)) < \epsilon - \delta.$$

  By the previous case, we can find $Y' \stackrel{D}{=} Y$ so that $\hat{d}(X', Y') < \epsilon - \delta$. Therefore

  $$\hat{d}(X, Y') \leq \hat{d}(X, X') + \hat{d}(X', Y') < \delta + \epsilon - \delta = \epsilon.$$

- X and Y arbitrary. Let $\delta > 0$ be such that

  $$2\delta < \epsilon - d_P(\text{law}\,(X),\ \text{law}\,(Y)).$$

  Find a sequence $\{Y_n\}_n$ of simple random elements converging to Y such that

  $$\hat{d}(Y_n, Y_{n+1}) < \frac{1}{2^{n+1}} \qquad \text{and} \qquad d_P(\text{law}\,(Y_n), \text{law}\,(Y)) < \delta.$$

  This can be done because the continuity of the function law. Let $N \in \mathbb{N}$ so that $2^{-N} < \delta$ and, at the same time, $\hat{d}(Y_N, Y) < \delta$. Construct a new sequence $\{Y_j'\}_{j \geq N}$ as follows:
  when $j = N$, choose Y_N' such that

  $$\text{law}\,(Y_N') = \text{law}\,(Y_N) \qquad \text{and} \qquad \hat{d}(Y_N', X) < d_P(\text{law}\,(X), \text{law}\,(Y)) + \delta.$$

  which can be done because

  $$\begin{aligned} d_P(\text{law}\,(Y_N), \text{law}\,(X)) &\leq d_P(\text{law}\,(Y_N), \text{law}\,(Y)) + d_P(\text{law}\,(Y), \text{law}\,(X)) \\ &< d_P(\text{law}\,(Y), \text{law}\,(X)) + \delta. \end{aligned}$$

  Once $\{Y_j'\}_{j=N}^{N+m}$ have been chosen, pick Y_{N+m+1}' satisfying

  $$\text{law}\,(Y_{N+m+1}') = \text{law}\,(Y_{N+m+1}) \quad \text{and} \quad \hat{d}(Y_{N+m+1}', Y_{N+m}') < \frac{1}{2^{N+m+1}}.$$

The existence of such random elements follows from

$$\begin{aligned} d_P(\text{law}\,(Y_{N+m+1}), \text{law}\,(Y'_{N+m})) &= d_P(\text{law}\,(Y_{N+m+1}), \text{law}\,(Y_{N+m})) \\ &< \frac{1}{2^{N+m+1}}. \end{aligned}$$

By construction, the new sequence $\{Y'_j\}$ is convergent, and for its limit Y' we can verify law $(Y') = $ law (Y). Moreover

$$\begin{aligned} \hat{d}(X, Y') &\leq \hat{d}(X, Y'_N) + \hat{d}(Y'_N, Y) \leq \hat{d}(X, Y'_N) + \sum_j \frac{\delta}{2^j} \\ &< d_P(\text{law}\,(X), \text{law}\,(Y)) + \delta + \delta < \epsilon. \end{aligned}$$

This completes the proof. ■

As part of the proof of Lemma 3.1.2, it has been shown that there exists a family of measurable sets $\{A_t\}_{t\in[0,\delta]}$ so that $A_s \subset A_t$ whenever $s \leq t$ and $P(A_t) = t$. Such family is called a $[0, \delta]$-family and it will be very useful when proving the existence of liftings with fixed endpoints.

Let $X = \sum_{j=1}^{m} a_j \mathbb{I}_{A_j}$ and $Y = \sum_{j=1}^{m} a_j \mathbb{I}_{B_j}$ be two simple random elements. We set

$$E_{ij} = A_i \cap B_j \qquad \text{and} \qquad e_{ij} = P(E_{ij}).$$

Let $\{[E_{ij}]_t\}$ be a $[0, e_{ij}]$-family of E_{ij}. A **segment joining** X to Y is defined as $\alpha_{X,Y} : [a, b] \to L^0(\Omega, S)$ given by

$$\alpha_{X,Y}(t) = \sum_{i=1}^{m} a_i \mathbb{I}_{E_{ii} \cup \left(\bigcup_{k=1,\, k\neq i} [E_{ki}]_{\left(\frac{t-a}{b-a}\right) e_{ki}} \right)} + \sum_{i=1}^{m} \sum_{j=1,\, j\neq i}^{m} a_i \mathbb{I}_{E_{ij} \setminus [E_{ij}]_{\left(\frac{t-a}{b-a}\right) e_{ij}}}.$$

Next we describe some important properties of these segments.

Lemma 3.1.6 *$\overline{\alpha} := \alpha_{X,Y}$ thus defined is a continuous function with $\alpha_{X,Y}(a) = X$ and $\alpha_{X,Y}(b) = Y$. Moreover,*

$$\alpha(t) := \mathit{law}\,(\overline{\alpha}(t)) = \left(\frac{b-t}{b-a}\right) \mathit{law}\,(X) + \left(\frac{t-a}{b-a}\right) \mathit{law}\,(Y).$$

Proof. Let $\epsilon > 0$ and $s < t$. Then we have

$$\begin{aligned} P(\, d(\overline{\alpha}(t), \overline{\alpha}(s)) \geq \epsilon\,) &\leq \sum_{\{(i,k): d(a_i,a_k)\geq\epsilon\}} P\left([E_{ik}]_{(t-s)e_{ik}}\right) \\ &\leq (t-s) \sum_{\{(\dots)\}} e_{ik} \leq t-s. \end{aligned}$$

An immediate consequence of the inequality is the continuity of $\alpha_{X,Y}$.

Now we show that law $(\overline{\alpha}) = \alpha$. Without loss of generality we may assume $a = 0$ and $b = 1$. In order to prove the last statement we only have to show that the coefficient of δ_{a_i} in law $(\overline{\alpha})$ is $(1-t)P(A_i) + tP(B_i)$. Let's fix $i = 1$. Then

$$\begin{aligned} P(\overline{\alpha} = a_1) &= P(E_{11}) + \sum_{j=2}^{m} P\left([E_{j1}]_{t \cdot e_{j1}}\right) + \sum_{j=2}^{m} P\left(E_{1j} \backslash [E_{1j}]_{t \cdot e_{1j}}\right) \\ &= e_{11} + \sum_{j=2}^{m} t e_{j1} + \sum_{j=2}^{m} (1-t) e_{1j} \\ &= t \sum_{j=1}^{m} e_{j1} + (1-t) \sum_{j=1}^{m} e_{1j} \\ &= tP(B_1) + (1-t)P(A_1). \end{aligned}$$

The same can be shown when $i = 2, \dots, m$. ∎

To help us find liftings with fixed endpoints, continuous maps from $[0,1]$ to $\mathcal{M}(S)$ will be approximated using a special family of paths we call **polygonals**. $\beta : [0,1] \to \mathcal{M}(S)$ is called a polygonal with vertices at $\mu_0, \dots, \mu_{n+1}$ if there is a partition $t_0 = 0, t_1, \dots, t_n, t_{n+1} = 1$ of $[0,1]$ such that β restricted to $[t_i, t_{i+1}]$ is given by

$$\beta(t) = \left(\frac{t_{i+1} - t}{t_{i+1} - t_i}\right) \mu_i + \left(\frac{t - t_i}{t_{i+1} - t_i}\right) \mu_{i+1}.$$

The function defined as above is called the segment (with domain $[t_i, t_{i+1}]$) joining μ_i to μ_{i+1}.

An easy consequence of Lemma 3.1.6 is the following fact about polygonals.

Lemma 3.1.7 *Let $\alpha : [0,1] \to \mathcal{M}(S)$ be a polygonal with vertices at measures of finite support. Given X_0 and X_1 such that law$(X_0) = \alpha(0)$ and law$(X_1) = \alpha(1)$, there is a lifting $\overline{\alpha} : [0,1] \to L^0(\Omega, S)$ such that $\overline{\alpha}(0) = X_0$ and $\overline{\alpha}(1) = X_1$.*

To show that polygonals can be used to approximate continuous maps from the unit interval to $\mathcal{M}(S)$ we will show that they form a dense set in such space. To make the proof of this statement easier, we make the following observation

Lemma 3.1.8 *Let $\mu, \nu \in \mathcal{M}(S)$. For $t \in [0,1]$ we have*

$$d_P(\nu, t\mu + (1-t)\nu) \leq d_P(\nu, \mu).$$

Proof. Let $\epsilon > d_P(\nu, \mu)$, then $\mu(F) \leq \nu(F_\epsilon) + \epsilon$ for all $F \subset S$ closed. Then we have

$$\begin{aligned} t\mu(F) + (1-t)\nu(F) &\leq t\nu(F_\epsilon) + t\epsilon + (1-t)\nu(F_\epsilon) + (1-t)\epsilon \\ &= \nu(F_\epsilon) + \epsilon. \end{aligned}$$

It means that $d_P(\nu, t\mu + (1-t)\nu) \leq \epsilon$. Since this is true for any $\epsilon \geq d_P(\nu, \mu)$ the statement of the lemma follows. ∎

Now we prove the density property of polygonals.

Lemma 3.1.9 *Given $\alpha : [0,1] \to \mathcal{M}(S)$ a continuous function and $\epsilon > 0$, there is a polygonal β with vertices at measures of finite support such that*

$$\sup_{t \in [0,1]} d_P(\alpha(t), \beta(t)) \leq \epsilon.$$

Proof. Let $\epsilon > 0$ be given. Since α is continuous and $[0,1]$ is compact, we can find $\delta > 0$ such that whenever $|s-t| < \delta$ we have $d_P(\alpha(t), \alpha(s)) < \frac{\epsilon}{5}$ (by the uniform continuity of α). Let $N \in \mathbb{N}$ be large enough so that $\frac{1}{N} < \delta$, and define a partition of the interval $[0,1]$ by $t_i = \frac{i}{N}$ for $i = 0, 1, \ldots, N$. For each i pick a finitely supported measure μ_i such that

$$d_P(\mu_i, \alpha(t_i)) \leq \frac{\epsilon}{5}.$$

Let β be the polygonal defined by the segments $\beta : [t_i, t_{i+1}] \to \mathcal{M}(S)$ with endpoints μ_i and μ_{i+1}. For each $t \in [t_i, t_{i+1}]$ we have

$$\begin{aligned}
d_P(\alpha(t), \beta(t)) &\leq d_P(\alpha(t), \alpha(t_i)) + d_P(\alpha(t_i), \mu_i) + d_P(\mu_i, \beta(t)) \\
&\leq d_P(\alpha(t), \alpha(t_i)) + d_P(\alpha(t_i), \mu_i) + d_P(\mu_i, \mu_{i+1}) \\
&\leq d_P(\alpha(t), \alpha(t_i)) + d_P(\alpha(t_i), \mu_i) + d_P(\mu_i, \alpha(t_i)) \\
&\quad + d_P(\alpha(t_i), \alpha(t_{i+1})) + d_P(\alpha(t_{i+1}), \mu_{i+1}) \\
&< \frac{\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} = \epsilon.
\end{aligned}$$

Clearly

$$\sup_{t\in[0,1]} d_P(\alpha(t), \beta(t)) \leq \epsilon,$$

as we wanted to show. ∎

Lemma 3.1.10 *Let μ and ν be two finitely supported measures. If*

$$X_\mu = \sum_{j=1}^{m} a_j \mathbb{I}_{A_j} \quad \textit{and} \quad X_\nu = \sum_{j=1}^{m} a_j \mathbb{I}_{B_j}$$

are such that $law(X_\mu) = \mu$ and $law(X_\nu) = \nu$ then

$$\hat{d}(X_\mu, \overline{\alpha}(t)) \leq \hat{d}(X_\mu, X_\nu),$$

where $\overline{\alpha} = \alpha_{X_\mu, X_\nu}$ is the segment joining X_μ to X_ν.

Proof. Given $\epsilon > 0$ we have

$$\begin{aligned}
P(d(X_\mu, X_\nu) \geq \epsilon) &= \sum_{\{(i,j):d(a_i,a_j)\geq\epsilon\}} P(E_{ij}) \\
&\geq \sum_{\{(i,j):d(a_i,a_j)\geq\epsilon\}} P([E_{ij}]_{te_{i,j}}) \\
&= P(d(X_\mu, \overline{\alpha}(t)) \geq \epsilon).
\end{aligned}$$

The conclusion of the lemma follows. ∎

Using the previous lemma we can now show

Lemma 3.1.11 *Let $\alpha : [0,1] \to \mathcal{M}(S)$ be a continuous path. Fix any $\epsilon > 0$ and let β be an arbitrary polygonal with vertices at measures of finite support and such that*

$$\sup_{t\in[0,1]} d_P(\alpha(t), \beta(t)) < \epsilon.$$

Then, given any continuous lifting $\overline{\alpha}$ of α, there is a lifting $\overline{\beta}$ of β such that

$$\sup_{t\in[0,1]} \hat{d}\left(\overline{\alpha}(t), \overline{\beta}(t)\right) < 5\epsilon.$$

Proof. Let $\overline{\alpha}$ be a continuous lifting of α. Take a partition $t_0 = 0 < t_1 < \cdots < t_{n+1} = 1$ of the unit interval, in such a way that

$$\hat{d}(X_i, \overline{\alpha}(t)) < \epsilon \quad \text{ where } \quad X_i = \overline{\alpha}(t_i), \;\text{ for all } t_i \le t \le t_{i+1}.$$

Choose Y_i for $i = 0, 1, \ldots, n+1$ so that $\text{law}\,(Y_i) = \beta(t_i)$ and $\hat{d}(X_i, Y_i) < \epsilon$, which can be done by Lemma 3.1.5. Then we have

$$\hat{d}(Y_i, Y_{i+1}) \le \hat{d}(Y_i, X_i) + \hat{d}(X_i, X_{i+1}) + \hat{d}(X_{i+1}, Y_{i+1}) < 3\epsilon.$$

Construct a lifting $\overline{\beta}$ of β, such that $\overline{\beta}$ restricted to the segment $[t_i, t_{i+1}]$ is a lifting of $\beta : [t_i, t_{i+1}] \to \mathcal{M}(S)$ with $\overline{\beta}(t_i) = Y_i$ (use segments joining Y_i to Y_{i+1}). Then $\overline{\beta}$ is continuous and for $t \in [t_i, t_{i+1})$ we have

$$\begin{aligned}
\hat{d}(\overline{\alpha}(t), \overline{\beta}(t)) &\le \hat{d}(\overline{\alpha}(t), X_i) + \hat{d}(X_i, Y_i) + \hat{d}(Y_i, \overline{\beta}(t)) \\
&\le \hat{d}(\overline{\alpha}(t), X_i) + \hat{d}(X_i, Y_i) + \hat{d}(Y_i, Y_{i+1}) \qquad \text{(by Lemma 3.1.10)} \\
&\le \epsilon + \epsilon + 3\epsilon = 5\epsilon. \qquad \blacksquare
\end{aligned}$$

Finally, we prove the existence of liftings with fixed endpoints.

Theorem 3.1.12 *Let $\alpha : [0,1] \to \mathcal{M}(S)$ be a continuous function. Let X_0 and X_1 be two elements of $L^0(\Omega, S)$ such that law $(X_0) = \alpha(0)$ and law $(X_1) = \alpha(1)$. Then, there exists a continuous path $\overline{\alpha} : [0,1] \to L^0(\Omega, S)$, ($L^0(\Omega, S)$ endowed with the topology of the convergence in probability) such that*

$$\overline{\alpha}(0) = X_0\,, \quad \overline{\alpha}(1) = X_1 \quad \textit{and} \quad \textit{law}\,(\overline{\alpha}) = \alpha.$$

Proof. Let $\{X_n^0\}$ and $\{X_n^1\}$ be two sequences of simple random elements in $L^0(\Omega, S)$ satisfying $X_n^0 \to X_0$, $X_n^1 \to X_1$,

$$\hat{d}(X_n^0, X_{n+1}^0) < \frac{1}{5^{n+1}} \quad \text{and} \quad \hat{d}(X_n^1, X_{n+1}^1) < \frac{1}{5^{n+1}}.$$

Take a sequence of polygonals $\{\alpha_n\}_n$ with vertices at measures of finite support, in particular $\alpha_n(0) = \text{law}\,(X_n^0)$ and $\alpha_n(1) = \text{law}\,(X_n^1)$, such that

$$\alpha_n \to \alpha \quad \text{and} \quad \sup_{t\in[0,1]} d_P(\alpha_n(t), \alpha_{n+1}(t)) < \frac{1}{5^{n+1}}.$$

Now, by induction we construct the liftings $\{\overline{\alpha}_n\}_n$. At the n^{th} step, by Lemma 3.1.7, we can find $\overline{\alpha}_n$ associated to α_n with $\overline{\alpha}_n(0) = X_n^0$ and $\overline{\alpha}_n(1) = X_n^1$. Then, by 3.1.11, we can find $\overline{\alpha}_{n+1}$ associated to α_{n+1} so that

$$\sup_{t\in[0,1]} \hat{d}(\overline{\alpha}_n(t), \overline{\alpha}_{n+1}(t)) < \frac{1}{5^n},$$

which can be done because $\sup_{t\in[0,1]} d_P(\alpha_n(t), \alpha_{n+1}(t)) < \frac{1}{5^{n+1}}$. In fact, we can even take $\overline{\alpha}_{n+1}(0) = X_n^0$ and $\overline{\alpha}_{n+1}(1) = X_n^1$ because $\hat{d}(X_n^0, X_{n+1}^0) < \frac{1}{5^{n+1}}$ and $\hat{d}(X_n^1, X_{n+1}^1) < \frac{1}{5^{n+1SSSSS}}$. Hence

$$\overline{\alpha}_n(0) \to X_0 \quad \text{and} \quad \overline{\alpha}_n(1) \to X_1.$$

It is clear by construction that $\{\overline{\alpha}_n\}$ is a convergent sequence. Let $\overline{\alpha}$ be its limit. Then, since the convergence is uniform, $\overline{\alpha}$ is continuous. By the continuity of the function law we must have

$$\text{law}\,(\overline{\alpha}(t)) = \alpha(t)\ , \qquad \forall t \in [0,1].$$

This finishes the proof. ∎

Chapter 4

Results

4.1 (C1)-continuity

In this section we prove that there is no (C1)-continuous representation unless S contains only one point.

Theorem 4.1.1 *Let S be a Polish space containing more that one element. Let ρ be a representation of $\mathcal{M}(S)$. Then ρ cannot be (C1)-continuous.*

Proof. Let x_1, x_2 be two different elements of S. Define $\mathcal{L} \subset \mathcal{M}(S)$ as

$$\mathcal{L} = \{\mu_\alpha : \mu_\alpha = \alpha\delta_{x_1} + (1-\alpha)\delta_{x_2},\ \alpha \in [0,1]\},$$

where $\delta_{x_1}, \delta_{x_2} \in \mathcal{M}(X)$ are given by (1.2), Section 1.3.4.

Suppose that $\rho : \mathcal{M}(S) \times U \to S$ is a (C1)-continuous representation. Without loss of generality, we assume that $\rho_u : \mathcal{M}(S) \to S$ is continuous for all $u \in U$. It follows that $\rho(\mu_\alpha, \cdot) : U \to S$ has distribution μ_α and the set $\{u : \rho_{\mu_\alpha}(u) \in \{x_1, x_2\}\}$ has λ-measure equal to 1. However, it may be false that

$$\lambda\left(\{u : \rho_{\mu_\alpha}(u) \in \{x_1, x_2\},\ \alpha \in [0,1]\}\right) = 1.$$

In contrast, since $\mathbb{Q} \cap [0,1]$ is countable

$$U' = \{u : \rho_{\mu_\alpha}(u) \in \{x_1, x_2\},\ \alpha \in \mathbb{Q} \cap [0,1]\}$$

satisfies $\lambda(U') = 1$.

We want to analyze how many $u \in U' \subset (0,1)$ are mapped to x_1 and how many to x_2 for each $\mu_\alpha \in \mathcal{L}$, but more importantly, how the transition from one point to the other is going to be. Fix $u \in U'$, then consider the sets

$$A_1 = \{\alpha : \rho_u(\mu_\alpha) = x_1,\ \alpha \in \mathbb{Q} \cap [0,1]\},$$

$$A_2 = \{\alpha : \rho_u(\mu_\alpha) = x_2,\ \alpha \in \mathbb{Q} \cap [0,1]\}.$$

Clearly $A_1 \cup A_2 = \mathbb{Q} \cap [0,1]$ and $A_1 \cap A_2 = \emptyset$. In addition, it is easy to see that $\overline{A_1} \cup \overline{A_2} = [0,1]$.

Now, consider the sets

$$B_1 = \{\alpha : \rho_u(\mu_\alpha) = x_1,\ \alpha \in [0,1]\}$$

$$B_2 = \{\alpha : \rho_u(\mu_\alpha) = x_2,\ \alpha \in [0,1]\}.$$

It's also clear that $A_1 \subset B_1$, $A_2 \subset B_2$, $B_1 \cup B_2 \subset [0,1]$, and $B_1 \cap B_2 = \emptyset$. In order to prove that $B_1 \cup B_2$ is indeed the interval $[0,1]$ we are going to show that $\overline{A_1} \subset B_1$. Take any $\alpha \in \overline{A_1}$, then there exists $\{\alpha_n\} \subset A_1$ such that $\lim_n \alpha_n = \alpha$. As a consequence $\mu_{\alpha_n} \Rightarrow \mu_\alpha$ and, since ρ_u is continuous,

$$\lim_{n\to\infty} \rho_u(\mu_{\alpha_n}) = \rho_u(\mu_\alpha).$$

Notice that the fact that $\rho_u(\mu_{\alpha_n}) = x_1$ for all $n \in \mathbb{N}$ (because $\alpha_n \in A_1$), implies $\rho_u(\mu_\alpha) = x_1$. Hence $\alpha \in B_1$.

Similarly we can prove $\overline{A_2} \subset B_2$, so that

$$\overline{A_1} \cap \overline{A_2} = \emptyset, \quad B_1 = \overline{A_1} \quad \text{and} \quad B_2 = \overline{A_2}.$$

It is important to see that B_1 and B_2 are both closed (and open!) disjoint subsets of $[0,1]$ whose union is $[0,1]$ itself. Since this interval is connected either B_1 or B_2 is an empty set. In other words, $\rho_u : \mathcal{M}(S) \to S$ is a constant function. However, since ρ_{μ_α} must have distribution μ_α

$$\lambda(\{u : \rho_u \text{ is constant }\}) = 0.$$

In conclusion, the hypothesis of ρ_u being (C1)-continuous is false when S has more than one point. ∎

4.2 (C2)-continuity

4.2.1 A Counterexample

As stated before, our version of ρ, given by (2.3), may fail to be (C2)-continuous. Consider the following counterexample: let S be the infinite dimensional Hilbert space

$$S = \left\{ x = (x_1, x_2, \ldots) : x_i \in \mathbb{R}, \sum_{i=1}^{\infty} x_i^2 < \infty \right\}$$

with orthonormal basis $\{e_i\}_{i=1}^{\infty}$, $e_i = (\delta_{i,j})_j$ where $\delta_{i,j} = 1$ iff $i = j$. Let D be a countable dense subset of S given by

$$D = \{y = (y_1, y_2, \ldots) : y_i \in \mathbb{Q},\ y_i = 0 \quad \forall i \geq n,\ \text{for some } n \in \mathbb{N}\}.$$

Let $m : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ give the enumeration of the spheres $G_1, G_2, \ldots$, that is $G_{m(i,j)} = B(y_j, r_i)$, where $y_j \in D$ and $r_i \in \mathbb{Q}$. Then the function c_a can be expressed as

$$c_a(x) = \sum_{i,j} \frac{1}{3^{m(i,j)}} \mathbb{I}_{d(x,y_j)<ar_i}.$$

For any $\alpha \geq 0, x \in S$ we have

$$\|\alpha e_n - x\|^2 = \alpha^2 - 2\alpha x_n + \|x\|^2 \to \alpha^2 + \|x\|^2.$$

Now, define the function f given by

$$f(\alpha) = \sum_{i,j} \frac{1}{3^{m(i,j)}} \mathbb{I}_{\alpha^2 + ||y_j||^2 < a^2 r_i^2},$$

and compare with

$$c_a(\alpha e_n) = \sum_{i,j} \frac{1}{3^{m(i,j)}} \mathbb{I}_{\alpha^2 + ||y_j||^2 - 2\alpha y_{j,n} < a^2 r_i^2},$$

where $y_{j,n}$ is the n^{th} term of y_j. Notice that every y_j contains only a finite amount of non zero elements. Hence for a fixed m we can find a large enough n so that $2\alpha y_{j,n} = 0$ for all $j = 1, 2, \ldots, m$. It follows that

$$\lim_{n\to\infty} c_a(\alpha e_n) = f(\alpha).$$

It's easy to see that, for $\alpha > 0$, $f(\alpha) \notin c_a(S)$. Otherwise, because c_a^{-1} is continuous on $c_a(S)$, we would have that the sequence $\{\alpha e_n\}_n$ is convergent, which is a contradiction. However, $f(0) = c_a(0) \in c_a(S)$.

The function f is right continuous and strictly decreasing. To see this, take any sequence $\alpha_n \downarrow \alpha$, for a large enough n

$$\alpha^2 + ||y_j||^2 < a^2 r_i^2 \quad \text{implies} \quad \alpha_n^2 + ||y_j||^2 < a^2 r_i^2,$$

$$\alpha^2 + ||y_j||^2 \geq a^2 r_i^2 \quad \text{implies} \quad \alpha_n^2 + ||y_j||^2 \geq a^2 r_i^2,$$

in other words $\mathbb{I}_{\alpha^2+||y_j||^2} = \mathbb{I}_{\alpha_n^2+||y_j||^2}$. Then choose an appropriate (finite) amount of j, i in J so that

$$|f(\alpha_n) - f(\alpha)| = \left| \sum_{i,j \notin J} \frac{1}{3^{m(i,j)}} \left(\mathbb{I}_{\alpha^2+||y_j||^2 < a^2 r_i^2} - \mathbb{I}_{\alpha_n^2+||y_j||^2 < a^2 r_i^2} \right) \right| \leq \epsilon,$$

for any $\epsilon > 0$. In other words $f(\alpha_n) \to f(\alpha)$. We have shown that $f(\alpha_n)$ increases strictly to $f(\alpha)$ and $f(\alpha_n) \in \overline{c_a(S)}$ for all n. Thus, for every $\alpha > 0$ there exists an increasing sequence $\{w_n\} \subset c_a(X)$ whose limit is $f(\alpha)$. Let $x_n \in S$ be such that $c_a(x_n) = w_n$, i.e. $x_n = h(w_n)$. Without loss of generality we may assume $||x_n|| \leq 2\alpha$ and $w_n \leq f(2\alpha)$ for all n.

Fix $u \in (0, 1)$ and define the probability measure

$$\mu_\alpha = u \left(\sum_{n=1}^{\infty} \frac{1}{2^n} \delta_{x_n} \right) + (1-u)\delta_0,$$

where δ_x is given by (1.2), Section 1.3.4. Notice that $\mu_\alpha \Rightarrow \delta_0$ as $\alpha \to 0$. Also notice that

$$J_a(\mu_\alpha) = u \left(\sum_{n=1}^{\infty} \frac{1}{2^n} \delta_{w_n} \right) + (1-u)\delta_{c_a(0)}.$$

We arranged things so that $\rho^*[J_a(\mu_\alpha), u] = f(\alpha) \notin c_a(S)$. This means that

$$\rho_a(\mu_\alpha, u) = \bar{h}[\rho^*(J_a(\mu_\alpha), u)] = x^*,$$

which does not converge to $\rho_a(\delta_0, u) = \bar{h}[\rho^*(J_a(\delta_0), u)] = 0$ as $\alpha \to 0$, for this we require $x^* \neq 0$. In other words, every value of u and a is a discontinuity point for some sequence converging to δ_0. This proves that our new version of ρ fails to be (C2)-continuous.

Remark: In a similar way we can take $w = (1, 0, 0, \ldots)$ instead of 0 and get a sequence of measures converging to δ_w for which the representation is not continuous at u for any u.

4.2.2 (C2)-continuity on $\mathbb{R}$

(C2)-continuity seems to be a very strong kind of continuity for a representation of $\mathcal{M}(S)$, however when $S = \mathbb{R}$, ρ^* has this property. If, instead of taking the complete real line we only take a Borel subset B of it, we would like to know whether or not is possible to construct a (C2)-continuous representation of $\mathcal{M}(S)$. Then, our intuition will probably lead us to try with ρ^* restricted to the set B. The first problem we find is that the values taken by $\rho^*(\mu)$, the pseudoinverse of μ, may be outside B, (see Section 1.3.2). So, it will be necessary to modify the function in order to get a well defined representation. Still, the way we redefine ρ^* outside B has to somehow preserve continuity. In the next two lemmas we show the existence of such a representation of $\mathcal{M}(B)$ when B satisfies certain properties. Before that, we recall that B' represents the closure of $B \subset \mathbb{R}$ in the upper limit topology $\mathscr{T}_u$, while $\overline{B}$ represents the closure in the usual topology.

Lemma 4.2.1 *Let B a Borel set such that $\left(\overline{B' \backslash B}\right) \cap B = \emptyset$. Then the function $\overline{\rho^*} : \mathcal{M}(B) \times (0, 1) \to B$ given by*

$$\overline{\rho^*}(\mu, u) = F_\mu^{(-1)}(u)\mathbb{I}_{F_\mu^{(-1)}(u) \in B} + b^*\mathbb{I}_{F_\mu^{(-1)}(u) \in B' \backslash B},$$

where b^* is an arbitrary element of B, is a (C2)-continuous representation of $\mathcal{M}(B)$.

Proof. First of all, notice that joint measurability of $\overline{\rho^*}$ follows from the measurability of its components. Now, take any $\mu \in \mathcal{M}(B)$ and define

$$N_\mu = \{u \in (0, 1) : F_\mu^{(-1)}(u) \text{ is discontinuous at } u\} \cup \{F_\mu^{(-1)}(u) \in B' \backslash B\}.$$

Lemma 1.3.10 guarantees that $\operatorname{Ran} F_\mu^{(-1)} \subset B'$. Clearly $\lambda(N_\mu) = 0$.

We claim that for a fixed $u \notin N_\mu$

$$\overline{\rho^*}(\mu_n, u) \to \overline{\rho^*}(\mu, u) \quad \text{whenever} \quad \mu_n \Rightarrow \mu.$$

Before we prove this, recall that $F_{\mu_n}^{(-1)}(u) \to F_{\mu}^{(-1)}(u)$ for every $u \notin N_\mu$. Let

$$C_1 = \{n \in \mathbb{N} : F_{\mu_n}^{(-1)}(u) \in B\} \quad \text{and} \quad C_2 = \{n \in \mathbb{N} : F_{\mu_n}^{(-1)}(u) \in B' \backslash B\}.$$

Clearly $C_1 \cap C_2 = \emptyset$ and $C_1 \cup C_2 = \mathbb{N}$. Suppose that C_2 is an infinite set. Then, let's rewrite $C_2 = \{n(k)\}_k$ so that $n(k) < n(k+1)$. Then

$$\lim_{k \to \infty} F_{\mu_{n(k)}}^{(-1)}(u) = F_{\mu}^{(-1)}(u),$$

but $F_{\mu_{n(k)}}^{(-1)}(u) \in B' \backslash B$ and $F_{\mu}^{(-1)}(u) \in B$, hence $F_{\mu}^{(-1)}(u) \in \left(\overline{B' \backslash B}\right) \cap B$, which is a contradiction. Therefore C_2 must be finite. Note that no matter what sequence $\{\mu_n\}$ we take, this set is always finite. This means that there exists N such that $\overline{\rho^*}(\mu_n, u) = F_{\mu_n}^{(-1)}(u)$ for all $n \geq N$. Hence

$$\begin{aligned} \lim_{n \to \infty} \overline{\rho^*}(\mu_n, u) &= \lim_{n \geq N} \overline{\rho^*}(\mu_n, u) = \lim_{n \geq N} F_{\mu_n}^{(-1)}(u) \\ &= F_{\mu}^{(-1)}(u) = \overline{\rho^*}(\mu, u), \end{aligned}$$

and the proof is complete. ∎

The previous lemma give us a construction of a (C2)-continuous representation of $\mathcal{M}(B)$, when B satisfies an additional property. For instance, if B is an open set then

$$B' \backslash B \subset \partial B \quad \text{and} \quad \left(\overline{B' \backslash B}\right) \cap B = \emptyset.$$

Similarly, when B is closed then we have $B' = \overline{B} = B$ ($\mathscr{T}_u$ is finer than the usual topology) and $B' \backslash B = \emptyset$, in which case $\overline{\rho_\mu^*}$ is just $F^{(-1)}$ restricted to the set B.

In fact, the last lemma can be generalized to the next lemma

Lemma 4.2.2 *Let B a Borel set such that*

$$\left(\overline{B' \backslash B}\right) \cap B \quad \textit{is a closed set.}$$

Then the function $\overline{\rho^*} : \mathcal{M}(B) \times (0,1) \to B$ *given by*

$$\overline{\rho^*}(\mu, u) = F_\mu^{(-1)}(u)\mathbb{I}_{F_\mu^{(-1)}(u)\in B} + f\left(F_\mu^{(-1)}(u)\right)\mathbb{I}_{F_\mu^{(-1)}(u)\in B'\setminus B} \tag{4.1}$$

(where $f : \mathbb{R} \to B$ *is the projection in* $\left(\overline{B'\setminus B}\right) \cap B$*) is a (C2)-continuous representation of* $\mathcal{M}(B)$*.*

Proof. First of all, let's define the function $f : \mathbb{R} \to B$ so that

$$|x - f(x)| = \inf\left\{ |x - y| : y \in \left(\overline{B'\setminus B}\right) \cap B\right\}.$$

In plain words, $f(x)$ is the closest (or one of the closest) point in $\left(\overline{B'\setminus B}\right)\cap B$ to x. Since $\left(\overline{B'\setminus B}\right) \cap B$ is a closed set, the function is well defined. Clearly, if $\{x_n\}_{n=1}^\infty \subset B'\setminus B$, $x \in B$ and $x_n \to x$ then $x \in \left(\overline{B'\setminus B}\right) \cap B$. Thus

$$|f(x_n) - x| \le |f(x_n) - x_n| + |x - x_n| \le 2|x - x_n| \to 0,$$

implying

$$\lim_{n\to\infty} f(x_n) = x. \tag{4.2}$$

Take any sequence $\{\mu_n\} \subset \mathcal{M}(B)$ converging to $\mu \in \mathcal{M}(B)$ and $\overline{\rho^*}(\mu, u)$ as in (4.1). Let

$$N_\mu = \{u \in (0,1) : F_\mu^{(-1)} \text{ is discontinuous at } u\} \cup \{F_\mu^{(-1)}(u) \in B'\setminus B\}.$$

Again, $\lambda(N_\mu) = 0$. We claim that, for a fixed $u \notin N_\mu$, $\overline{\rho^*}(\mu_n, u) \to \overline{\rho^*}(\mu, u)$. To prove this, define

$$C_1 = \{n \in \mathbb{N} : F_{\mu_n}^{(-1)}(u) \in B\} \quad \text{and} \quad C_2 = \{n \in \mathbb{N} : F_{\mu_n}^{(-1)}(u) \in B'\setminus B\}.$$

Clearly $C_1 \cap C_2 = \emptyset$ and $C_1 \cup C_2 = \mathbb{N}$. If C_2 is a finite set then

$$\begin{aligned}\lim_{n\to\infty} \overline{\rho^*}(\mu_n, u) &= \lim_{n\to\infty, n\in C_1} \overline{\rho^*}(\mu_n, u) = \lim_{n\to\infty} F_{\mu_n}^{(-1)}(u) \\ &= F_\mu^{(-1)}(u) = \overline{\rho^*}(\mu, u).\end{aligned}$$

Let's suppose now that C_2 is an infinite set. Then for $n \in C_2$ we must have $F_{\mu_n}^{(-1)}(u) \in B' \backslash B$ and $F_{\mu}^{(-1)}(u) \in B$. Recall that $F_{\mu_n}^{(-1)}(u) \to F_{\mu}^{(-1)}(u)$. Hence $F_{\mu}^{(-1)}(u) \in \left(\overline{B' \backslash B}\right) \cap B$, and by (4.2)

$$\lim_{n\to\infty, n\in C_2} \overline{\rho^*}(\mu_n, u) = \lim_{n\to\infty} f\left(F_{\mu_n}^{(-1)}(u)\right) = F_{\mu}^{(-1)}(u) = \overline{\rho^*}(\mu, u).$$

Since $F_{\mu_n}^{(-1)}(u)$ and $f\left(F_{\mu_n}^{(-1)}(u)\right)$ both converge to $F_{\mu}^{(-1)}(u) = \overline{\rho^*}(\mu, u)$ the representation is (C2)-continuous.

Finally, joint measurability follows from the measurability of the components of $\overline{\rho^*}$. ∎

4.2.3 (C2)-continuity in a Locally Compact Polish Space

In Section 4.2.1 of this chapter we modified the original Blackwell and Dubins representation but we lost (C2)-continuity. However, in Section 4.2.2 we prove that under some conditions of B a Borel subset of $\mathbb{R}$ we may redefine $\rho^*(\mu, u) = F_{\mu}^{(-1)}(u)$ to get a (C2)-continuous representation of $\mathcal{M}(B)$.

In the rest of the section we prove that if S is a locally compact Polish space, then we can find a (C2)-continuous representation. Before we prove this, we show some other results. In the next lemma $\{H_n\}_n$ is an enumeration of open balls with centers in a countable dense subset of S and rational radius, as in (2.2).

Lemma 4.2.3 *Let S be a locally compact Polish space. The basis $\{H_n\}_n$ of the topology in S may be reduced to a collection $\{H_{n(k)}\}_k$ where $\overline{H^a_{n(k)}}$ is compact for all k and all $a \in [1,2]$.*

Proof. First of all, notice that if $\overline{H^2_n}$ is compact, then it remains true for $\overline{H^a_n}$, when $a \in [1,2]$. As usual, let $D = \{x_n\}_n$ be the (countable dense) set of all centers of $\{H_n\}_n$. Let

$$s(n) = \sup\left\{\frac{1}{k} : \overline{B\left(x_n, \frac{2}{k}\right)} \text{ is compact, } k \in \mathbb{N}\right\}.$$

Clearly, $s(n) > 0$ because S is locally compact. Then define

$$C_n = \{\, B(x_n, r) : r \leq s(n), r \in \mathbb{Q} \,\}.$$

It is very easy to see that

$$\bigcup_n C_n \subset \{H_n\}_n.$$

Then we take $\{H_{n(k)}\}_k = \bigcup_n C_n$. We must show now that this is a basis of the topology in X.

Take any open set U and choose $x \in U$. Let $r \in \mathbb{R}$ be small enough to make the the closure of $B(x, r)$ compact and $B(x, r) \subset U$. Because $D = \{x_n\}$ is dense, we can find $x_{n^*} \in D$ such that $d(x, x_{n^*}) \leq \frac{r}{4}$. Then $x \in B(x_{n^*}, r^*)$ where $r^* \in \left(\frac{r}{4}, \frac{r}{2}\right)$ and $r^* \in \mathbb{Q}$. Clearly

$$\overline{B(x_{n^*}, r^*)} \subset \overline{B(x, r)}$$

as a consequence, $r^* \leq s(n^*)$ and $B(x_{n^*}, r^*) \in \{H_{n(k)}\}_k$. We just proved that for any $x \in U$ there exists an element in $\{H_{n(k)}\}_k$ contained in U and at the same time containing x. ∎

For simplicity, in this section $\{H_{n(k)}\}_k$ will be denoted just by $\{H_n\}_n$ but it is important to keep in mind the the closure of this sets are compact sets. Now, we redefine the function c introduced in Section 2.2, let $\bar{c} : [1, 2] \times S \to \mathbb{R}$ be given by

$$\bar{c}(x) = \sum_{n=1}^{\infty} \frac{1}{3^n} \mathbb{I}_{H_n^a}(x),$$

where, as established before $\{H_n^a\}$ is based on the "reduced" basis. Because $\{H_n\}_n$ is a basis of the topology in S $\bar{c}$ has all the properties that c has. For instance, $\bar{c}_a$ is an injective, continuous function in the complement of $\bigcup_n \partial H_n^a$ and measurable.

This slight change in c will allow us to obtain an continuous extension of the (already continuous in $\bar{c}([1, 2] \times S)$) inverse of $\bar{c}$.

Lemma 4.2.4 *There exists a continuous function $g : \bar{c}(A \times S)' \to S$ such that*

$$g[\bar{c}(a, x)] = x \quad \textit{for all} \quad x \in S,\ a \in [1, 2].$$

Proof. In this proof h represents the projection in S of the inverse of $\bar{c}$. In Lemma 2.3.4 we showed that such function is continuous in $\bar{c}([1,2]\times S)$. So, we only need to find a way to extend h outside $\bar{c}([1,2]\times S)$.

Let $y \in \bar{c}(A\times S)'\backslash\bar{c}(A\times S)$. Then, there exista a sequence $\{y_n\}_n$ in $c(A\times S)$ converging to y. Let (a_n, x_n) be the (unique) pair of numbers that give us $\bar{c}(a_n, x_n) = y_n$. Denote by $\{\alpha_i\}_{i=1}^{\infty}$ and $\{\alpha_{n,i}\}_{i=1}^{\infty}$ the ternary expansion of y and y_n respectively, that is

$$y = \sum_{i=1}^{\infty} \frac{\alpha_i}{3^i}, \quad y_n = \sum_{i=1}^{\infty} \frac{\alpha_{n,i}}{3^i}, \quad \alpha_i, \alpha_{n,i} = 0, 1 \quad \text{for all} \quad i, n.$$

To see that α_i cannot be equal to 2 for any i, notice that $\bar{c}(A\times S)$ is a subset of the Cantor-like set

$$\mathcal{C} = \left\{ y \in \mathbb{R} : y = \sum_n \frac{\beta_n}{3^n},\ \beta_n = 0, 1,\ n \in \mathbb{N} \right\},$$

which is already closed. Hence $\bar{c}(A\times S)' \subset \mathcal{C}$ (the upper limit topology is finer than the usual topology).

Now, since $y_n \to y$, we must have $\alpha_{n,i} \to \alpha_i$, for all n, i, in fact $\alpha_{n,i} = \alpha_i$ for all sufficiently large n. The fact that $y \notin \bar{c}(A\times S)$ implies that, for any $a \in A$

$$\bigcap_{i=1}^{\infty} (H_i^a)^{\alpha_i} = \emptyset,$$

where

$$(H_i^a)^{\alpha_i} = \begin{cases} H_i^a & \text{if} \quad \alpha_i = 1 \\ (H_i^a)^c & \text{if} \quad \alpha_i = 0. \end{cases}$$

However, for any $n \in \mathbb{N}$, there exists a large enough m so that

$$x_k \in \bigcap_{i=1}^{n} (H_i^{a_k})^{\alpha_i} \subset \bigcap_{i=1}^{n} \overline{(H_i^*)^{\alpha_i}}, \quad \text{for all } k \geq m,$$

where $H_i^* = H_i^2$ if $\alpha_i = 1$ and $H_i^* = H_i^1$ if $\alpha_i = 0$. Since $\bar{c}(A\times S)$ is contained in $(0,\infty)$, which is closed in the upper limit topology, we have $\bar{c}(A\times S)' \subset (0,\infty)$ and so $y \neq 0$. The sets $\bigcap_{i=1}^{n} \overline{(H_i^*)^{\alpha_i}}$ are closed and eventually compact. We say eventually compact because $y \neq 0$ implies the existence of

k such that $\alpha_k = 1$, and $\overline{H_k^*}$ is compact. Hence, these compact sets have the finite intersection property. Therefore

$$\lim_{n\to\infty} \bigcap_{i=1}^{n} \overline{(H_i^*)^{\alpha_i}} = \bigcap_{i=1}^{\infty} \overline{(H_i^*)^{\alpha_i}} \neq \emptyset.$$

Suppose that x_1 and x_2 are both elements of this set. If $x_1 \neq x_2$, find H_i so that $x_1 \in H_i^1 \subset H_i^2$ but $x_2 \notin \overline{H_i^2}$. So, if $\alpha_i = 1$ then $x_1 \in (H_i^*)^{\alpha_i} = H_i^2$ while $x_2 \notin (H_i^*)^{\alpha_i} = H_i^2$. On the other hand, if $\alpha_i = 0$ then $x_1 \notin (H_i^*)^{\alpha_i} = (H_i^1)^c$ while $x_2 \in \overline{(H_i^*)^{\alpha_i}} = (H_i^1)^c$. This contradicts the fact that both x_1 and x_2 belong to $\bigcap_i \overline{(H_i^*)^{\alpha_i}}$. Therefore we may take the unique element x of this set and define $g(y) = x$, while $g(y) = h(y)$ if $y \in \bar{c}(A \times S)$.

The function $g : \bar{c}(A\times S)' \to S$ can be seen as an extension we are looking for. Of course $\bar{c}(g(y))$ may not be equal to y, but we claim that if $\{z_n\}_n$ is a sequence in $\bar{c}(A \times S)'$ that converges to $z \in \bar{c}(A \times S)'$ then

$$\lim_{n\to\infty} g(z_n) = g(z).$$

To prove our claim, let $\{\beta_i\}$ and $\{\beta_{n,i}\}$ be the ternary expansion of z and z_n respectively. Then

$$\{g(z)\} \in \bigcap_{i=1}^{\infty} \overline{(H_i^*)^{\beta_i}},$$

similarly,

$$\{g(z_n)\} \in \bigcap_{i=1}^{\infty} \overline{(H_i^*)^{\beta_{n,i}}},$$

and $\beta_{n,i} \to \beta_i$. Fix any $\epsilon > 0$, we must show $d(g(z_n), g(z)) < \epsilon$ for a large enough n. Since $\{H_n^2\}_n$ is a basis of the topology in X, there exists H_k^2 such that $g(z) \in H_k^2$ and the radius of H_k^2 is less than ϵ. Then, find $m \in \mathbb{N}$ large enough so that

$$\{g(z_i)\}_{i=m}^{\infty} \subset \bigcap_{i=1}^{k} \overline{(H_i^*)^{\beta_i}} \subset \overline{H_k^2}.$$

Therefore

$$d(g(z_n), g(z)) \leq \epsilon, \;\; \forall n \geq m,$$

which proves $g(z_n) \to g(z)$. ∎

Lemma 4.2.5 *Let S be a locally compact Polish space, in particular S may be $\mathbb{R}^n$. Then there exists a (C2)-continuous representation of $\mathcal{M}(S)$.*

Proof. Define $\rho : \mathcal{M}(S) \times [1,2] \times (0,1) \to S$ as

$$\rho(m, a, u) = g(\rho^*(\bar{J}_a(m), u)),$$

where $\bar{J}_a(\mu)$ is the distribution of $\bar{c}_a : S \to \mathbb{R}$ in the probability space $(S, \mathcal{B}(S), \mu)$. Notice that $\bar{J}_a$ has the same properties of J_a, as in Section 2.2.

Suppose $\mu_n \Rightarrow \mu$, then if $\mu\left(\bigcup_n \partial H_n^a\right) = 0$ (condition that holds for almost every a), $J_a(\mu_n) \Rightarrow J_a(\mu)$. Since ρ^* is a (C2)-continuous representation of $\mathcal{M}(\mathbb{R})$, for any $u \notin N_\mu$, where

$$N_\mu = \{u : \rho^*(\bar{J}_a(\mu_n), u) \to \rho^*(\bar{J}_a(\mu), u)\}^c,$$

we must have

$$g(\rho^*(\bar{J}_a(\mu_n), u)) \to g(\rho^*(\bar{J}_a(\mu), u))$$

Clearly $\lambda(N_\mu) = 0$, moreover N_μ is a countable set.

On the other hand, notice that all the components of ρ are jointly measurable. Overall, we can conclude that ρ is a continuous representation of $\mathcal{M}(S)$. ∎

Bibliography

[1] Ash, R.B. and Doléans-Dade, C.A. (2000) *Probability & Measure Theory, Second Edition*. Academic Press.

[2] Billingsley, P. (1968) *Convergence of Probability Measures*. John Wiley & Sons.

[3] Blackwell, D., Dubins, L.E. (1983) *An extension of Skorohod's almost sure representation theorem*, Proc. Amer. Math. Soc. , 89, pp. 691-692.

[4] Bogachev, V. I. (2007) *Measure Theory*. Springer.

[5] Cortissoz, J.C. (2007) *On the Skorokhod representation theorem*, Proc. Amer. Math. Soc. 135, pp. 3995-4007.

[6] Dudley, R.M. (1968) *Distances of probability measures and random variables*, Ann. Math. Statist. 39, pp. 1563-1572.

[7] Dudley, R.M. (1989) *Real analysis and probability*. Brooks/Cole Pub. Co.

[8] Ethier, S. and Kurtz, T. (1986) *Markov Processes: Characterization and Convergence*. John Wiley & Sons.

[9] Fathi, A. (1997) *Partitions of unity for countable covers*, Amer. Math. Monthly, Vol. 104, No. 8, pp. 720-723.

[10] Fernique, X. (1988) *Un modèle presque sûr pour la convergence en loi*, C. R. Acad. Sci. Paris Sér. I 306 , pp. 335-338.

[11] Kallenberg, O. (1997) *Foundations of Modern Probability*. Springer.

[12] Munkres, J.R. (1975) *Topology: a first course*. Prentice Hall, Inc.

[13] Skorohod, A.V. (1956) *Limit theorems for stochastic processes*, Theory Probab. Appl. 1, pp. 261-290.

[14] Troitskiĭ, V.G. (1994) *Real partitions of measure spaces*, (Russian) Sibirskiĭ Matematicheskiĭ Zhurnal, Vol. 35, No. 1, pp. 207-209; translation in Siberian Math. J. 35, no. 1, pp. 189-191.