MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Central Limit Theorems and Invariance Principles for Martingale Difference Arrays

verfasst von / submitted by

Mag. Jürgen Steininger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc.)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Assoz. Prof. Dr. Roland Zweimüller

To
Levi & Noah

Contents

Abstract

Following a paper of D.L. McLeish (see [9]) we prove Central Limit Theorems (CLT's) and Functional Central Limit Theorems (FCLT's) - also called invariance principles - for martingale difference arrays. In this context, we give an introduction to the theory of weak convergence in general metric spaces. We prove a general CLT for martingale difference arrays according to Theorem 2.3 in [9]. Before proving the corresponding FCLT we study the space $\mathbb{D}[0,1]$ *of càdlàg-functions on* $[0,1]$ *with Skorohod's* $\mathcal{J}_1$*-topology. The results are used to present a proof of a general FCLT for martingale difference arrays according to Theorem 3.2 in [9].*

1 Introduction

One of the main purposes of this thesis is to fill the gap between introductory literature in probability theory and the mentioned paper [9] which is hard to read if one is not a "specialist" in probability theory. Consequently the work should help to make the topic approachable to a wider public. Anyone with basic knowledge of probability theory (and measure theory as well) should be able to read the text without many problems. Some basic knowledge of topology is also advantageous.

1.1 Notation

For the rest of the text we fix some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is some non-empty set, $\mathcal{F}$ a σ-algebra on Ω and $\mathbb{P}$ a probability measure on $\mathcal{F}$. We call a map X from Ω to another measurable space $(E, \mathcal{E})$ *random element* of E if it is $\mathcal{F}$-$\mathcal{E}$-measurable. If E is a metric space then it is usually equipped with the Borel σ-algebra denoted by $\mathcal{B}_E$. In the special case of $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}_\mathbb{R})$ we call X *random variable.* If we say that the sequence of random elements $(X_n)_{n\geq 1}$ is an *iid-sequence* we mean that the random elements are independent and identically distributed. If X is normally distributed with expected value μ and variance σ^2 we write $X \sim \mathcal{N}(\mu, \sigma^2)$. By $\mathbb{C}[0,1]$ we denote the space of continuous functions $x : [0,1] \longrightarrow \mathbb{R}$ which can be seen as a subspace of $\mathbb{R}^{[0,1]}$, the space of all functions defined on $[0,1]$. Different modes of convergence will appear in the text. As common in probability theory we will denote them by labeling the arrows, with one exception: weak convergence. We will denote it with a double arrow, i.e. writing $X_n \Rightarrow X$.

1.2 Motivation

Although the CLT is nowadays one of the standard results in probability theory and simple versions are even part in the curriculum in secondary school it is still a fascinating result having applications in many fields. A classical version reads as follows:

Theorem (Classical Central Limit Theorem).
Let $(X_n)_{n\geq 1}$ be an iid-sequence of random variables with mean $\mathbb{E}[X_1] = \mu$ and variance $\mathrm{Var}[X_1] = \sigma^2 < \infty$. Then

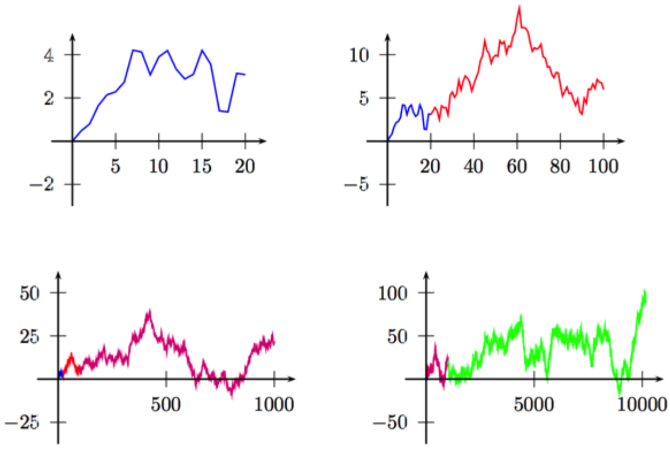
$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu) \Rightarrow Z \sim \mathcal{N}(0, \sigma^2).$$

Basic assumptions are the identical distribution and the independence of the random variables. Using some conditions on the sequence which guarantee that the contribution of any individual random variable is not to big one can get rid of the assumption on identical distribution. Even better, it's also possible to get rid of the independence assumption. Usually, this is done by introducing the concept of (triangular) arrays of random variables. Nevertheless, the independence in each row is still necessary. In this text we want to go a step further and ask whether if it is possible to get completely rid of the independence. For sure the random variables then must have a nice structure of dependence, in this text they will be so-called martingale difference arrays. There are different versions of CLTs for such martingale difference arrays. Our aim is to prove those of [9] but we will also mention versions with other assumptions.

All CLTs have one thing in common: They make a statement about the asymptotic behaviour of the sum $S_n = \sum_{i=1}^{n} X_i$ of random variables. We also want to move to the next level and ask if there is some regularity in the whole "path". For motivation we consider a one-dimensional random walk

$$S_n := \sum_{i=1}^{n} \xi_i,$$

where the ξ_i's are an iid-sequence of random variables normalized such that $\mathbb{E}[\xi_i] = 0$ and $\mathrm{Var}[\xi_i] = 1$ for all i. Let us plot many steps of the random walk, where the points are connected by linear interpolation:



It seems that there is a limit process if we do the right rescaling. Therefore we need to fit those paths first into a common area, like the interval $[0,1]$. To make this argument rigorous one defines stochastic processes $\left(S_t^{(n)}\right)_{t\in[0,1]}$, $n \geq 1$, via

$$S_t^{(n)} := \frac{1}{\sqrt{n}}\left(\xi_1 + \ldots + \xi_{\lfloor nt \rfloor}\right) + \frac{nt - \lfloor nt \rfloor}{\sqrt{n}} \xi_{\lfloor nt \rfloor + 1}. \tag{1.1}$$

The next Theorem is certainly the most famous stochastic process limit theorem.

Theorem (Donsker's Theorem).
Let $S_t^{(n)}$ be defined as in (1.1). Then

$$\left(S_t^{(n)}\right)_{t\in[0,1]} \Rightarrow (W_t)_{t\in[0,1]},$$

where $(W_t)_{t\in[0,1]}$ is a (standard) Brownian motion on $[0,1]$.

We want to generalize this Theorem in two ways: Firstly, corresponding to the CLT we will again only assume that the ξ_i's form a martingale difference array and secondly, we will construct $S_t^{(n)}$ in a more general way such that the paths are not continuous anymore - a generalization which comes with many technical issues but is necessary for many applications.

1.3 Organization

Chapter 2 has an introductory character to weak convergence in general metric spaces. It provides all the necessary results which are needed for the later chapters. In Chapter 3 CLT's for martingale difference arrays are the main object. Before going to the corresponding FCLT's we need some understanding of the space $\mathbb{D}[0,1]$ of right-continuous functions with left-hand limits on the interval $[0,1]$. Therefore we give in Chapter 4 first an introduction to weak convergence on the subspace $\mathbb{C}[0,1] \subset \mathbb{D}[0,1]$ of continuous functions since it is easier to handle. In Chapter 5 we will show how these results can be generalized to $\mathbb{D}[0,1]$. Finally, Chapter 6 is devoted to FCLT's for martingale difference arrays.

2 Weak Convergence

In the whole work we are concerned with weak convergence (or convergence in distribution), a mode of convergence which, in the case of random variables, is already part in every introductory course of probability theory. The reason why we want to study it here again is that we need a more general concept of weak convergence which is well-suited not only for random variables but also for random elements on more general spaces. Most of the results are taken from the sections 1-3 and 5 of the "bible" of weak convergence [1]. Other helpful literature was [4], [8] and the lecture notes [6]. For a better reading we state also some basic definitions and well-known results which are later needed without proofing them.

2.1 Towards a general definition of Weak Convergence

To motivate the general definition we look first at a random variable X and a sequence of random variables $(X_n)_{n\geq 1}$. Recall that the *distribution* of a random variable is defined to be the induced probability measure $\mathbb{P} \circ X^{-1}$, denoted by $\mathbb{P}_X$. Weak convergence is traditionally defined by convergence of their *distribution functions* $F_n(t) := \mathbb{P}[X_n \leq t]$ and $F(t) := \mathbb{P}[X \leq t]$:

Definition 2.1. X_n *converges in distribution* to X if $F_n(t) \longrightarrow F(t)$ for all continuity points t of F.

Even though the assumption about continuity points is not necessary if F is continuous, there are cases where the assumption plays a role.

Example 2.2. Let $(X_n)_{n\geq 1}$ and X be random variables with distribution functions $F_n(t) := \mathbb{1}_{[\frac{1}{n},\infty)}(t)$ and $F(t) = \mathbb{1}_{[0,\infty)}(t)$. Then we have $X_n \Rightarrow X$ even though the convergence fails at the discontinuity point $t = 0$.

Although the concept of distribution functions is well suited for simple arguments it has some disadvantages for more general concepts. It is not well suited to calculations involving sums of independent random variables. Take for example the proof of the classical CLT. Usually it does not check pointwise convergence of distribution functions but uses instead characteristic functions. For us even worse, the concept of distribution functions is tied to the real line (or the Euclidian space) and can not be used in more general spaces of random elements (for example the space $\mathbb{C}[0, 1]$).Therefore we want to extend weak convergence in a way that reduces to the above definition in the case for random variables.

Recall that F is continuous at a point x if and only if the set $\{x\}$ has $\mathbb{P}_X$-measure 0, i.e. $\mathbb{P}_X[\{x\}] = \mathbb{P}[X = x] = 0$. Let $B = (a, b]$ and X_n converging in distribution to X. Then

$$\begin{aligned}\mathbb{P}_{X_n}[B] = \mathbb{P}[a < X_n \leq b] = F_n(b) - F_n(a)\\ \longrightarrow F(b) - F(a) = \mathbb{P}[a < X \leq b] = \mathbb{P}_X[B],\end{aligned}$$

whenever $\mathbb{P}_X[\{a, b\}] = \mathbb{P}[X = a \text{ or } X = b] = 0$. Thus, converging in distribution implies $\mathbb{P}_{X_n}[B] \longrightarrow \mathbb{P}_X[B]$ for all Borel sets $B = (a, b]$ with $\mathbb{P}_X[\{a, b\}] = 0$. Therefore, we have motivated a definition of weak convergence in terms of convergence of probability measures.

Let ∂A denote the boundary of a subset $A \in \mathcal{B}_{\mathbb{R}}$. Then we can say that $\mathbb{P}_{X_n}$ converges weakly to $\mathbb{P}_X$ if and only if

$$\mathbb{P}_{X_n}[A] \longrightarrow \mathbb{P}_X[A] \text{ whenever } \mathbb{P}[\partial A] = 0.$$

It is straightforward to see that this definition is equivalent to the traditional definition with distribution functions. But the concept of weak convergence of probability measures has the advantage that it can be formulated for a general metric space, which is the real reason for preferring the latter concept.

2.2 Weak Convergence in Metric Spaces

Having motivated a general definition we now want to study weak convergence in metric spaces. Therefore, let (S, d) be a metric space. For $x \in S$ we denote by

$$B(x, \varepsilon) = \{y \in S \mid d(x, y) < \varepsilon\}$$

the *open ball* of radius ε. Since we want to study the convergence of probability measures on this space we need to equip S with a σ-algebra. An obvious candidate is the Borel σ-algebra

$$\mathcal{B}_S := \sigma(S, d) = \sigma\left(\{G \subset S : G \text{ open}\}\right).$$

Hence, $\mathcal{B}_S$ is the smallest σ-algebra containing all open sets.

Remark 2.3. For studying weak convergence we actually do not need random elements. We start therefore the chapter by looking only at probability measures at the space $(S, \mathcal{B}_S)$ and we will later identify these measures as the distributions of certain random elements.

We denote the *set of probability measures* on $(S, \mathcal{B}_S)$ by $\mathcal{P}(S)$.

Remark 2.4. To define the convergence of a sequence $(Q_n)_{n\geq 1} \in \mathcal{P}(S)$ to a limit Q a naive first attempt would actually be to require that

$$Q_n[A] = \int_S \mathbb{1}_A dQ_n \to Q[A] = \int_S \mathbb{1}_A dQ$$

for all $A \in \mathcal{B}_S$. But, if we take for example $Q_n := \delta_{\frac{1}{n}}$ where $\delta_{\cdot}$ denotes the Dirac-measure defined by $\delta_x[A] = \mathbb{1}_A(x)$ we want intuitively that Q_n converges to $Q := \delta_0$, however $Q_n[\{0\}] = 0 \nrightarrow Q[\{0\}] = 1$. Therefore we need a weaker notion of convergence as motivated above where we have required the convergence only for sets A with $Q[\partial A] = 0$. Another solution would be to replace the discontinuous functions $\mathbb{1}_A$ by continuous functions.

We call a set $A \in \mathcal{B}_S$ whose boundary ∂A satisfies $Q[\partial A] = 0$ a *Q-continuity set.*

Definition 2.5. Let $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(S)$. We say that Q_n converges *weakly* to Q if

$$Q_n[A] \longrightarrow Q[A] \text{ for all } Q\text{-continuity sets } A \in \mathcal{B}_S,$$

and write $Q_n \Rightarrow Q$.

Most often in the literature another equivalent definition is given. In fact, the following Theorem provides useful conditions equivalent to weak convergence. Any of them could serve as the definition.

Theorem 2.6 (Portmanteau Theorem).
Let $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(S)$. Then the following are equivalent:

i) $Q_n[A] \longrightarrow Q[A]$ *for all Q-continuity sets $A \in \mathcal{B}_S$,*

ii) $\int_S f\, dQ_n \longrightarrow \int_S f\, dQ$ *for all bounded, continuous functions f,*

iii) $\int_S f\, dQ_n \longrightarrow \int_S f\, dQ$ *for all bounded, uniformly continuous functions f,*

iv) $\int_S f\, dQ_n \longrightarrow \int_S f\, dQ$ *for all bounded, Lipschitz-functions f,*

v) $\limsup_{n\to\infty} Q_n[F] \leq Q[F]$ *for all closed F,*

vi) $\liminf_{n\to\infty} Q_n[G] \geq Q[G]$ *for all open G.*

Proof. Cf. [1], Theorem 2.1. □

The definition most used is the condition (ii) of the. Portmanteau's Theorem do not only provide many useful tools for proving weak convergence but also has a very important corollary: Let therefore $(\tilde{S}, \mathcal{B}_{\tilde{S}})$ be another metric space and $\varphi : S \to \tilde{S}$ a $\mathcal{B}_S$-$\mathcal{B}_{\tilde{S}}$-measurable map. If $Q \in \mathcal{P}(S)$ then $Q \circ \varphi^{-1},\ Q \circ \varphi^{-1}[\tilde{A}] := Q[\varphi^{-1}(\tilde{A})]$ with $\tilde{A} \in \mathcal{B}_{\tilde{S}}$ defines a probability measure on $(\tilde{S}, \mathcal{B}_{\tilde{S}})$. An important question is if we can conclude from $Q_n \Rightarrow Q$ that $Q_n \circ \varphi^{-1} \Rightarrow Q \circ \varphi^{-1}$. The answer is yes if φ has some additional property.

Theorem 2.7 (**Continuous Mapping Theorem**).
Let $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(S)$ and $N(\varphi)$ be the set of continuity-points of φ. If $Q_n \Rightarrow Q$ and $Q[N(\varphi)] = 1$, then $Q_n \circ \varphi^{-1} \Rightarrow Q \circ \varphi^{-1}$.

Proof. Cf. [1], Theorem 2.7. □

In particular, the Continuous Mapping Theorem holds for any continuous function. It has extensive consequences in the theory of stochastic processes. If one proves a limit result it is enough to take for example a continuous function φ and get a new limit result.

The next Lemma is often useful for examples (or counter examples):

Lemma 2.8. *For $x \in S$ let δ_x be the Dirac-measure on $(S, \mathcal{B}_S)$. Then we have*

$$\delta_{x_n} \Rightarrow \delta_{x_0} \quad \Leftrightarrow \quad x_n \to x_0$$

Proof. ($\Leftarrow$): If $x_n \to x_0$ and f is a bounded and continuous function, then

$$\int_S f \, d\delta_{x_n} = f(x_n) \to f(x_0) = \int_S f \, d\delta_{x_0}.$$

($\Rightarrow$): Suppose $x_n \nrightarrow x_0$. Then $\exists \varepsilon > 0$, s.t.

$$d(x_n, x_0) > \varepsilon$$

for infinitely many n. Define for $x \in S$ the function

$$f_\varepsilon(x) := \max\left\{0, 1 - \frac{d(x_n, x_0)}{\varepsilon}\right\}$$

The function f_ε is bounded and continuous, $f_\varepsilon(x_0) = 1$ and $f_\varepsilon(x_n) = 0$ for infinitely many n. Therefore

$$\int_S f_\varepsilon \, d\delta_{x_n} = f_\varepsilon(x_n) \nrightarrow f_\varepsilon(x_0) = \int_S f_\varepsilon \, d\delta_{x_0}$$

and the result follows. □

In practise it's often hard to prove the convergence for all Q-continuity sets $A \in \mathcal{B}_S$. Another strategy would be to prove $Q_n[A] \to Q[A]$ for all Q-continuity sets A of a subclass $\mathcal{H} \subset \mathcal{B}_S$ and try to conclude that $Q_n \Rightarrow Q$.

Definition 2.9. We call $\mathcal{H} \subset \mathcal{B}_S$ a *separating class* if any two probability measures Q_1 and Q_2 that agree on $\mathcal{H}$ agree also on the whole of $\mathcal{B}_S$, i.e. if

$$Q_1[A] = Q_2[A] \ \ \forall A \in \mathcal{H} \ \Rightarrow \ Q_1[A] = Q_2[A] \ \ \forall A \in \mathcal{B}_S.$$

We call $\mathcal{H} \subset \mathcal{B}_S$ a *convergence-determining class* if for every $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(S)$, convergence of $Q_n[A] \to Q[A]$ for all Q-continuity sets $A \in \mathcal{H}$ implies $Q_n \Rightarrow Q$.

The next Lemma gives a very useful condition for a subclass $\mathcal{H} \subset \mathcal{B}_S$ to be a separating class.

Lemma 2.10. *$\mathcal{H} \subset \mathcal{B}_S$ is a separating class if it is a π-system and $\sigma(\mathcal{H}) = \mathcal{B}_S$.*

Proof. Let $Q_1, Q_2 \in \mathcal{P}(S)$, s.t. $Q_1[A] = Q_2[A]$ for every $A \in \mathcal{H}$ and let

$$\Gamma \coloneqq \{A \in \mathcal{B}_S : Q_1[A] = Q_2[A]\}$$

be the set of "good sets". We show that Γ is a λ-system:

- Clearly, $S \in \Gamma$ since $Q_1[S] = Q_2[S] = 1$.
- Let $A \in \Gamma$. We have $Q_1[A^c] = 1 - Q_1[A] = 1 - Q_2[A] = Q_2[A^c]$ and therefore $A^c \in \Gamma$.
- If $(A_n)_{n\geq 1}$ are disjoint sets in Γ we have

  $$Q_1\left[\bigcup_{n\geq 1} A_n\right] = \sum_{n\geq 1} Q_1[A_n] = \sum_{n\geq 1} Q_2[A_n] = Q_2\left[\bigcup_{n\geq 1} A_n\right]$$

  and therefore the union is in Γ.

Since $\mathcal{H} \subset \Gamma$ and Γ is a λ-system we get by Dynkin's π-λ-Lemma that $\sigma(\mathcal{H}) \subset \Gamma \subset \mathcal{B}_S$. By assumption we have $\sigma(\mathcal{H}) = \mathcal{B}_S$ and therefore $\Gamma = \mathcal{B}_S$. □

Example 2.11. Let $(S, \mathcal{B}_S) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. The set $\mathcal{H} = \{(-\infty, x] : x \in \mathbb{R}\}$ is a separating class. Indeed, $\mathcal{H}$ is clearly a π-system and $\sigma(\mathcal{H}) = \mathcal{B}_{\mathbb{R}}$. It is also a convergence-determining class since this is exactly the traditional definition.

By the definition it is clear that a convergence-determining class is always a separating class, the converse need not to be true. Unfortunately, in complicate spaces it's not possible to find convergence-determining classes but only separating classes. Our most important separating classes will be collections of preimages of certain functions. We look therefore at a.s. continuous (and therefore measurable) functions

$$f_i : (S, \mathcal{B}_S) \to (S_i, \mathcal{B}_{S_i})$$

for $i \in I$ where I is some index set. We can then look at the collection

$$\mathcal{H} = \left\{f_i^{-1}(B) \mid B \in \mathcal{B}_{S_i}, i \in I\right\} \subset \mathcal{B}_S. \tag{2.2}$$

If $\mathcal{H}$ is a convergence-determining class, to prove $Q_n \Rightarrow Q$ it would be enough to show that $Q_n \circ f_i^{-1} \Rightarrow Q \circ f_i^{-1}$ for all $i \in I$. However, if $\mathcal{H}$ is just a separating class we can not conclude directly that $Q_n \Rightarrow Q$ follows from $Q_n \circ f_i^{-1} \Rightarrow Q \circ f_i^{-1}$. But maybe it's possible if the sequence $(Q_n)_{n\geq 1}$ has additional structure.

Definition 2.12. Let $\Pi \subset \mathcal{P}(S)$ be a family of probability measures. We call Π *relatively sequentially compact* if every sequence of elements of Π contains a weakly convergent subsequence.

Note that the limiting probability measures might be different for different subsequences. The annex "relative" refers to the fact that the limiting measure of such a converging subsequence may lie outside of Π. Most of the time we are concerned with the relative sequential compactness of sequences $(Q_n)_{n\geq 1}$. Following the definition this means that every subsequence $(Q_{n_k})_{k\geq 1}$ contains another subsubsequence $(Q_{n_{k(m)}})_{m\geq 1}$ such that $Q_{n_{k(m)}} \Rightarrow Q$ for $m \to \infty$ and some probability measure Q. Now we can prove a recipe for proving weak convergence:

Theorem 2.13 (Recipe for weak convergence).
Let $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(S)$ and $f_i : (S, \mathcal{B}_S) \to (S_i, \mathcal{B}_{S_i})$ a.s. continuous functions. If we have

1. *$(Q_n)_{n\geq 1}$ is relatively sequentially compact,*
2. *$Q_n \circ f_i^{-1} \Rightarrow Q \circ f_i^{-1}$ for all $i \in I$,*
3. *$\mathcal{H} = \left\{ f_i^{-1}(B) \mid B \in \mathcal{B}_{S_i}, i \in I \right\}$ is a separating class,*

then $Q_n \Rightarrow Q$.

Proof. Suppose $Q_n \not\Rightarrow Q$. Then we have by Portmanteau's Theorem 2.6 that $\int_S f\ dQ_n \not\to \int_S f\ dQ$ for some bounded and continuous f. Hence, $\exists \varepsilon > 0$ and some subsequence with

$$\left| \int_S f\ dQ_{n_k} - \int_S f\ dQ \right| > \varepsilon \tag{2.3}$$

for all k. By (1) there exists a probability measure $\mu \in \mathcal{P}(S)$ and a subsequence of (Q_{n_k}) such that $Q_{n_{k(m)}} \Rightarrow \mu$. From (2.3) it follows

$$\left| \int_S f\ d\mu - \int_S f\ dQ \right| \geq \varepsilon$$

and hence, $\mu \neq Q$.

But, since the functions f_i are almost surely continuous it follows $Q_{n_{k(m)}} \circ f_i^{-1} \Rightarrow \mu \circ f_i^{-1}$ by the Continuous Mapping Theorem 2.7. Hence, by (2) we have

$$Q \circ f_i^{-1} = \mu \circ f_i^{-1}$$

for all $i \in I$. Therefore $Q = \mu$ on $\mathcal{H}$ and by (3) we have $Q = \mu$, a contradiction. $\square$

Although relative sequential compactness is a crucial property it is a little bit nasty to work with. Luckily, there is another condition for relative sequential compactness which is easier to verify in practice.

Definition 2.14. Let $\Pi \subset \mathcal{P}(S)$ be a family of probability measures. We call Π *tight* if for any $\varepsilon > 0$ there exists a compact set K such that

$$Q[K^c] < \varepsilon \ \ \forall Q \in \Pi.$$

Note that some authors are speaking of *uniform tightness* in the above definition since the condition must hold for all $Q \in \Pi$, i.e. $\sup_{Q\in\Pi} Q[K^c] < \varepsilon$.

Remark 2.15. If $(S, \mathcal{B}_S) = (\mathbb{R}, \mathcal{B}_\mathbb{R})$, tightness of Π is obviously equivalent to the fact that for all $\varepsilon > 0$ there must exists $M > 0$ such that

$$Q\left[[-M, M]^c\right] < \varepsilon \ \ \forall Q \in \Pi$$

or equivalently

$$\lim_{M\to\infty} \sup_{Q\in\Pi} Q\left[[-M, M]^c\right] = 0$$

This gives a good intuition about tightness. A given collection of measures is tight if it doesn't "escape to infinity". Note also that in the case of a sequence $(Q_n)_{n\geq 1}$ of probability measures this is the same as

$$\lim_{M\to\infty} \limsup_{n\to\infty} Q_n\left[[-M, M]^c\right] = 0.$$

In general weak convergence can be handled on any metric space but in this work we will be concerned with so-called polish spaces most of the time.

Definition 2.16. A metric space $(S, \mathcal{B}_S)$ is called *polish* if it is separable and complete.

Lemma 2.17. *If $(S, \mathcal{B}_S)$ is polish any $Q \in \mathcal{P}(S)$ is tight.*

Proof. Let $\{a_1, a_2, \ldots\}$ be a countable dense subset of S. Then for each $\delta > 0$ we have $S = \bigcup_{k\geq 1} B(a_k, \delta)$. Hence, $Q[S] = 1 = \lim_{n\to\infty} Q\left[\bigcup_{k=1}^n B(a_k, \delta)\right]$. Let $\varepsilon > 0$. Then there is for each $m \geq 1$ an n_m such that

$$Q\left[\bigcup_{k=1}^{n_m} \underbrace{B\left(a_k, \tfrac{1}{m}\right)}_{=: A_{k,m}}\right] \geq 1 - \frac{\varepsilon}{2^m},$$

otherwise we would have $Q[\bigcup_{k\geq 1} A_{k,m}] = 1 < 1 - \varepsilon \cdot 2^{-m}$, a contradiction. Now we define

$$K \coloneqq \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{n_m} \bar{A}_{k,m}.$$

Since any finite union and any intersection of closed sets are closed, K is closed. Furthermore, we have for each $\delta > 0$ that

$$K \subset \bigcup_{k=1}^{n_m} \bar{A}_{k,m} \subset \bigcup_{k=1}^{n_m} B(a_k, \delta)$$

if we choose $m > \delta^{-1}$. Hence, K is totally bounded. Since S is complete it follows that K is compact. We get

$$\begin{aligned}
Q\left[K^c\right] &= Q\left[\bigcup_{m=1}^{\infty} \left\{\bigcup_{k=1}^{n_m} \bar{A}_{k,m}\right\}^c\right] \\
&\leq \sum_{m=1}^{\infty} Q\left[\left\{\bigcup_{k=1}^{n_m} \bar{A}_{k,m}\right\}^c\right] \\
&= \sum_{m=1}^{\infty} 1 - Q\left[\bigcup_{k=1}^{n_m} \bar{A}_{k,m}\right] \\
&\leq \sum_{m=1}^{\infty} 1 - Q\left[\bigcup_{k=1}^{n_m} A_{k,m}\right] \\
&\leq \sum_{m=1}^{\infty} \frac{\varepsilon}{2^m} \\
&= \varepsilon,
\end{aligned}$$

which concludes the proof. □

The next famous result due to Prohorov gives the connection between the two important concepts of tightness and relative sequential compactness.

Theorem 2.18 (**Prohorov's Theorem**).
Let $\Pi \subset \mathcal{P}(S)$. Then we have

1. Π *tight* $\Rightarrow$ Π *relatively sequentially compact*
2. *If* $(S, \mathcal{B}_S)$ *is polish and* Π *relatively sequentially compact* $\Rightarrow$ Π *tight.*

Proof. Cf. [1], Theorem 5.1 for the first part and Theorem 5.2 for the second part. □

Hence, we can replace the relative sequential compactness assumption in our recipe 2.13 with the tightness condition.

2.3 Random Elements on Metric Spaces

The whole theory of weak convergence can be paraphrased as the theory of convergence in distribution. We look therefore at a $\mathcal{F}$-$\mathcal{B}_S$-measurable map $X : \Omega \to S$. The distribution $\mathbb{P} \circ X^{-1}$ is again denoted by $\mathbb{P}_X$ and contains all the essential information about X. It is therefore clear what we mean by the distribution of a single random variable, but later we will speak about the distributions of stochastic processes. We will therefore change our point of view and will "see" stochastic processes also as a single random object.

2.3.1 The Distribution of a Stochastic Process

We recall first the traditional definition:

Definition 2.19. A stochastic process $\mathbb{X} = \{X_t : t \in T\}$ is a collection of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in a common measurable space $(E, \mathcal{E})$, indexed by a set T.

Note that X_t depends on both t and the point $\omega \in \Omega$ at which it is evaluated, i.e. $X_t = X(\omega, t)$. For fixed value of t, $X_t(\omega)$ is straightforward a random element. For each fixed $\omega \in \Omega$, however, the process $\mathbb{X}$ defines a function $\mathbb{X}_\cdot(\omega) : T \to E, t \mapsto X_t(\omega)$ which is called a sample path of the process. Hence, we can look at the process $\mathbb{X}$ as a functioned valued random variable. To make this precise we need a measure space of functions and a measurable mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to that function space. The natural set to use is the product space

$$(E^T, \mathcal{E}^T) = \left(\prod_{t \in T} E, \bigotimes_{t \in T} \mathcal{E} \right).$$

Recall that $\mathcal{E}^T$ is the smallest σ-algebra that makes the projections

$$\begin{aligned} p_t : & E^T \to E \\ & x \longmapsto p_t(x) = x(t) \end{aligned}$$

measurable. Note also that the σ-algebra $\mathcal{E}^T$ is generated by the set of one-dimensional cylinders $\mathcal{Z}_1 = \left\{ p_t^{-1}(A) \mid A \in \mathcal{E}, t \in T \right\} = \left\{ x \in E^T \mid p_t(x) \in A \text{ with } A \in \mathcal{E}, t \in T \right\}$ or also the set of finite-dimensional cylinders $\mathcal{Z}_{fin} = \left\{ p_t^{-1}(A) \mid A \in \mathcal{E}^T, t \in T \right\}$, i.e. we have $\sigma(\mathcal{Z}_1) = \sigma(\mathcal{Z}_{fin}) = \mathcal{E}^T$.

Proposition 2.20. *$\mathbb{X}$ is $\mathcal{F}$-$\mathcal{E}^T$-measurable if and only if X_t is $\mathcal{F}$-$\mathcal{E}$-measurable for every $t \in T$*

Proof. Since a map $\mathbb{X}$ is $\mathcal{F}$-$\mathcal{E}^T$-measurable if $\mathbb{X}^{-1}(B) \in \mathcal{F}$ for all B in a class that generates $\mathcal{E}^T$, to check measurability it is enough to consider B of the form $p_t^{-1}(A)$ with

$A \in \mathcal{E}$. But

$$
\begin{aligned}
\mathbb{X}^{-1}\left(p_t^{-1}(A)\right) &= \left\{\omega \in \Omega \mid \mathbb{X}(\omega) \in p_t^{-1} \in (A)\right\} \\
&= \{\omega \in \Omega \mid p_t\left(\mathbb{X}(\omega)\right) \in A\} \\
&= \{\omega \in \Omega \mid X_t(\omega) \in A\},
\end{aligned}
$$

which is in $\mathcal{F}$ whenever X_t is $\mathcal{F}$-$\mathcal{E}$-measurable. The converse implication follows in the same way. $\square$

This Proposition is very natural, since if $\mathbb{X}$ is measurable all the marginals X_t should be in order to go along with the primary definition. The projection operator is therefore also a convenient tool for recovering the individual coordinates: For $t \in T$, $X_t = \pi_t \circ \mathbb{X}$ gives us a collection of random variables, i.e. a stochastic process in the sense of our first definition.

$$
\begin{tikzcd}
(\Omega, \mathcal{F}, \mathbb{P}) \arrow[rr, "X_t"] \arrow[dr, "\mathbb{X}"'] & & (E, \mathcal{E}) \arrow[dl, "p_t^{-1}"] \\
& (E^T, \mathcal{E}^T) &
\end{tikzcd}
$$

Remark 2.21. If T is uncountable then the σ-algebra $\mathcal{E}^T$ is rather crude. Subsets of E^T which are not determined by countably many $t \in T$ are never in $\mathcal{E}^T$.

Most of the time we want our path of the stochastic process to have certain properties. In other words, the paths of $\mathbb{X}$ belongs most of the time to a subset of E^T. Write the set of all such functions as $S \subset E^T$. Notice that S has not be an element of $\mathcal{E}^T$.

Example 2.22. Let $T = [0,1]$, $E = \mathbb{R}$, $\mathcal{E} = \mathcal{B}_{\mathbb{R}}$ and look at the subspace $\mathbb{C}[0,1] \subset \mathbb{R}^{[0,1]}$. One can show that $\mathbb{C}[0,1] \notin \mathcal{E}^T$.

Therefore, one can instead look at the trace-σ-algebra $\mathcal{S} = S \cap \mathcal{E}^T$, defined by

$$\mathcal{S} = S \cap \mathcal{E}^T = \{S \cap E : E \in \mathcal{E}^T\},$$

to get the measurable space $(S, \mathcal{S})$. Note that it follows directly from the definition of the trace-σ-algebra that $\mathcal{S}$ coincides with the σ-algebra generated by the projections restricted to S, denoted by $p_t\big|_S$, i.e.

$$\mathcal{S} = S \cap \sigma\left(p_t \mid t \in T\right) = \sigma\left(p_t\big|_S \mid t \in T\right).$$

We can therefore think of $\mathbb{X}$ as a map from $(\Omega, \mathcal{F}, \mathbb{P})$ to a subspace S of all functions from T to E. Proposition 2.20 gives immediately:

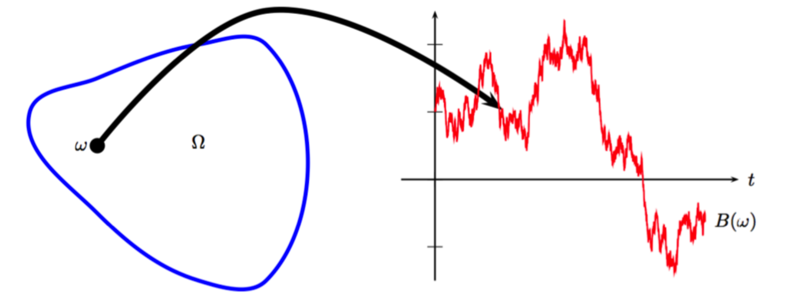
Corollary 2.23. *$\mathbb{X} : \Omega \to S \subset E^T$ is $\mathcal{F}$-$S \cap \mathcal{E}^T$-measurable if and only if $X_t : \Omega \to E$ is $\mathcal{F}$-$\mathcal{E}$-measurable.*

Proof. Since $\mathbb{X}(\omega) \in S$ we have $\mathbb{X}^{-1}(B) = \{\omega \in \Omega \mid \mathbb{X}(\omega) \in B\} \in \mathcal{F}$ if and only if $\mathbb{X}$ is $\mathcal{F}$-$\mathcal{E}^T$-measurable. Now apply Proposition 2.20. $\square$

Our subspace S will be always a metric space. Therefore the space carries an other natural σ-algebra, namely the Borel σ-algebra $\mathcal{B}_S$ generated by the open set's of its metric. We will later put quite some effort into proofing that these Borel-σ-algebras coincide with $\mathcal{S}$ in our spaces of interest.

The following example gives us an intuition how we can think of a process as a random function:

Example 2.24. Look at a Brownian Motion $\mathbb{B} = (B_t)_{t\geq 0}$, a collection of random variables $B_t : \Omega \to \mathbb{R}$. All the sample paths lie in the space $S = \mathbb{C}[0,\infty)$, the space of real continuous functions on $[0,\infty)$. We can therefore think of $\mathbb{B}$ as a random element of $\mathbb{C}[0,\infty)$, i.e. as a map $\mathbb{B} : \Omega \to \mathbb{C}[0,\infty)$. Some random experiment is performed on Ω and the outcome $\mathbb{B}(\omega)$ is a function on $\mathbb{C}[0,\infty)$:



Once we view the stochastic process $\mathbb{X}$ as a single random element we can define the probability distribution as usual.

Definition 2.25. Let $\mathbb{X}$ be an $(E,\mathcal{E})$-valued stochastic process. Then the induced probability measure $\mathbb{P}_{\mathbb{X}} = \mathbb{X} \circ \mathbb{P}^{-1}$ on the path space $\left(E^T, \mathcal{E}^T\right)$ (or on some subspace $(S, S \cap \mathcal{E}^T)$) is called the probability distribution of $\mathbb{X}$.

2.3.2 Convergence of Random Elements

Definition 2.26. A sequence $(X_n)_{n\geq 1}$ of random elements *converges in distribution* to a random element X if $\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$. In this case we write $X_n \Rightarrow X$.

Note that $\mathbb{P}$ is always a probability measure on an arbitrary measurable space $(\Omega, \mathcal{F})$, whereas the distribution of a random element $\mathbb{P}_X$ is a probability measure on the metric space $(S, \mathcal{B}_S)$.

$$\begin{tikzcd}
(\Omega, \mathcal{F}) \arrow[rr, "X"] \arrow[dr, "\mathbb{P}"'] & & (S, \mathcal{B}_S) \arrow[dl, "\mathbb{P}_X = \mathbb{P} \circ X^{-1}"] \\
& {[0,1]} &
\end{tikzcd}$$

All the results from the section before can now be carried over to random elements. No new ideas are needed, it's all just a matter of notation. Just to get used to the notation we give two examples:

Example 2.27 (Notation for random elements).
Condition (ii) of Portmanteaus Theorem: Note that

$$\int_S f \; d\mathbb{P}_X = \int_\Omega f(X) \; d\mathbb{P} = \mathbb{E}[f(X)].$$

Therefore we have $X_n \Rightarrow X$ if $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for all bounded and continuous functions f.

Continuous Mapping Theorem: Note that

$$\mathbb{P}_{\varphi(X)} = \mathbb{P} \circ (\varphi \circ X)^{-1} = \mathbb{P} \circ X^{-1} \circ \varphi^{-1} = \mathbb{P}_X \circ \varphi^{-1}.$$

If therefore $X_n \Rightarrow X$ and $\varphi : S \to \tilde{S}$ measurable for some other metric space $(\tilde{S}, \mathbb{B}_{\tilde{S}})$ with $\mathbb{P}[X \in N(\varphi)] = 1$, then $\varphi(X_n) \Rightarrow \varphi(X)$.

We briefly remind the definition of other modes of converges in the context of metric spaces:

Definition 2.28. A sequence of random elements $(X_n)_{n\geq 1}$ is said to *converge in probability* to a random element X, denoted by $X_n \xrightarrow{\mathbb{P}} X$ if $\forall \varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left[d(X_n, X) \geq \varepsilon\right] = 0$$

A sequence of random elements $(X_n)_{n\geq 1}$ is said to converge *almost surely* to a random element X, denoted by $X_n \xrightarrow{a.s.} X$ if

$$\mathbb{P}\left[\lim_{n\to\infty} d(X_n, X) = 0\right] = 1$$

Note that these two modes of convergence only make sense if all the random elements are defined on the same probability space.

Remark 2.29. The statement of the Continuous Mapping Theorem 2.7 also holds for these convergence-types.

Remember that we have the following relation between the different notions of convergence:

$$\boxed{X_n \xrightarrow{a.s} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \Rightarrow X}$$

If $S = \mathbb{R}$ we have the additional concept of L^p-convergence.

Definition 2.30. A sequence of random variables $(X_n)_{n\geq 1}$ is said to converge in L^p for $p \geq 1$ to a random variable X, denoted by $X_n \xrightarrow{L^p} X$ if $\mathbb{E}[|X_n|^p]$ and $\mathbb{E}[|X|^p]$ exist and

$$\mathbb{E}\left[|X_n - X|^p\right] \longrightarrow 0.$$

For us the most important cases are $p = 1$ or $p = 2$. This notion of convergence is also strong in the sense that it implies convergence in probability and convergence in distribution. For $q > p$ we have:

$$\boxed{X_n \xrightarrow{L^q} X \implies X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \Rightarrow X}$$

A useful result for deducing weak convergence of one sequence of random elements from another sequence is the following Lemma.

Lemma 2.31 (**Converging together Lemma**).
Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be two sequences of random elements with $d(X_n, Y_n) \to 0$ in probability and $X_n \Rightarrow Z$. Then $Y_n \Rightarrow Z$.

Proof. We use part (iii) of Portmanteau's Theorem. Therefore, let f be a bounded function ($|f(x)| \leq M$) which is also Lipschitz with Lipschitz constant K. We have for $\varepsilon > 0$

$$\begin{aligned}
|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(X_n)]| &\leq \mathbb{E}\left[|f(Y_n) - f(X_n)|\right] \\
&= \mathbb{E}\left[|f(Y_n) - f(X_n)|\mathbb{1}_{\{d(X_n,Y_n)<\varepsilon\}}\right] + \mathbb{E}\left[|f(Y_n) - f(X_n)|\mathbb{1}_{\{d(X_n,Y_n)\geq\varepsilon\}}\right] \\
&\leq \mathbb{E}\left[K \cdot d(X_n, Y_n) \cdot \mathbb{1}_{\{d(X_n,Y_n)<\varepsilon\}}\right] + \mathbb{E}\left[2M \cdot \mathbb{1}_{\{d(X_n,Y_n)\geq\varepsilon\}}\right] \\
&\leq K \cdot \varepsilon \cdot \mathbb{P}[d(X_n, Y_n) < \varepsilon] + 2M \cdot \mathbb{P}[d(X_n, Y_n) \geq \varepsilon] \\
&\leq K \cdot \varepsilon + 2M \cdot \mathbb{P}[d(X_n, Y_n) \geq \varepsilon].
\end{aligned}$$

Therefore we get

$$\begin{aligned}
|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Z)]| &\leq |\mathbb{E}[f(Y_n)] - \mathbb{E}[f(X_n)]| + |\mathbb{E}[f(X_n)] - \mathbb{E}[f(Z)]| \\
&\leq K \cdot \varepsilon + 2M \cdot \mathbb{P}[d(X_n, Y_n) \geq \varepsilon] + |\mathbb{E}[f(X_n)] - \mathbb{E}[f(Z)]|\,.
\end{aligned}$$

For $n \to \infty$ the second term goes to 0 since $X_n \xrightarrow{\mathbb{P}} Y_n$ and the third term goes to 0 since $X_n \Rightarrow Z$. Since ε was arbitrary we conclude that $\mathbb{E}[f(Y_n)] \longrightarrow \mathbb{E}[f(Z)]$. □

3 CLT's for Martingale Difference Arrays

In this section we specify our metric space $(S, \mathcal{B}_S)$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ where we always assume that $\mathbb{R}$ is equipped with the natural metric defined by $d(x, y) \coloneqq |x-y|$. Note that $(\mathbb{R}, d)$ is then a polish space. Hence, all the familiar tools can be used, especially the concept of characteristic functions will be of great help. The chapter consists of three sections. In the first section we provide some basic results which can be skipped if the reader is familiar with the concept of characteristic functions, uniform integrability and the notion of martingale difference arrays. The main course is derived from section 11 in [10] and section 5 in [16]. The second section gives a first CLT for martingale difference arrays due to [2] after a short reminder of the Lindeberg-Feller CLT. Finally, the third section gives the proof of a general CLT for martingale difference arrays corresponding to Theorem 2.3 in [9] and an application of the result. To elaborate the proof we used section 5 of [4], section 3 of [5] and the notes [15]. For the connection to the CLT of [2] we used in particular [18].

3.1 Preliminaries

3.1.1 Characteristic Functions

Recall that the characteristic function of a random variable X is a function $\varphi : \mathbb{R} \to \mathbb{C}$ defined by

$$\varphi(t) = \mathbb{E}\left[e^{\mathrm{i}tX}\right] = \int_{\Omega} e^{\mathrm{i}tX}\, d\mathbb{P} = \int_{\mathbb{R}} e^{\mathrm{i}tx}\, d\mathbb{P}_{\mathbb{X}}$$

The most useful result of characteristic functions which we will use quite often is the following famous Theorem:

Theorem 3.1 (Lévy's Continuity Theorem).
Let $(X_n)_{n\geq 1}$ and X be random variables with characteristic functions φ_{X_n} and φ_X. Then the following are equivalent:

- $X_n \Rightarrow X$
- $\varphi_{X_n}(t) \longrightarrow \varphi_X(t) \;\; \forall t \in \mathbb{R}$

Proof. Cf. [10], Theorem 7.12 (in German) or [7], Theorem 5.3. □

Note that this result also holds for random vectors in $\mathbb{R}^d$. Hence, to prove weak convergence in $\mathbb{R}$ or $\mathbb{R}^d$ it is enough to check pointwise convergence of the corresponding characteristic functions.

3.1.2 Uniform Integrability

Assume that we know $X_n \xrightarrow{a.s.} X$. It is often of great importance in many applications to determine conditions ensuring that $\mathbb{E}[X_n] \to \mathbb{E}[X]$. The Monotone and Dominated Convergence Theorems give sufficient conditions ensuring that $\mathbb{E}[|X_n - X|] \to 0$, hence L^1-convergence which is even stronger since we have $|\mathbb{E}[X_n] - \mathbb{E}[X]| \leq \mathbb{E}[|X_n - X|]$. Nevertheless, the assumptions of these two well-known Theorems are not necessary. In general the needed condition is the so-called uniform integrability of the sequence $(X_n)_{n\geq 1}$.

Remark 3.2. As we have seen in the first chapter there is no connection between almost sure and L^1-convergence. In applications it's often hard to show L^1-convergence directly. Hence, our target is to obtain a result which gives us a recipe for proving L^1-convergence.

Definition 3.3. A family $(X_n)_{n\geq 1}$ of integrable random variables is called *uniformly integrable* if

$$\sup_{n\geq 1} \mathbb{E}\left[|X_n| \mathbb{1}_{\{|X_n|\geq c\}}\right] \stackrel{c\to\infty}{\longrightarrow} 0.$$

Remark 3.4. [**Properties of uniform integrability**]

(i) For an integrable random variable X the one-member family $\{X\}$ is uniformly integrable.

(ii) Every finite family of integrable random variables $(X_n)_{n<n_0}$ with $n_0 \in \mathbb{N}$ is uniformly integrable.

(iii) If the family $(X_n)_{n\geq 1}$ of integrable random variables is dominated by an integrable random variable Y, i.e. $|X_n| \leq Y$ for all $n \in \mathbb{N}$ then $(X_n)_{n\geq 1}$ is uniformly integrable. In particular, if $(X_n)_{n\geq 1}$ is bounded by a constant it is uniformly integrable.

(iv) Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be two families of integrable random variables. If $(Y_n)_{n\geq 1}$ is uniformly integrable and $|X_n| \leq |Y_n|$ for all $n \in \mathbb{N}$, then $(X_n)_{n\geq 1}$ is uniformly integrable.

(v) If $\mathcal{G}$ and $\mathcal{H}$ are two uniformly integrable families then the collections $\{X+Y : X \in \mathcal{G}, Y \in \mathcal{H}\}$, $\{X - Y : X \in \mathcal{G}, Y \in \mathcal{H}\}$ and $\{|X| : X \in \mathcal{G}\}$ are also uniformly integrable families.

Definition 3.5. We say that $(X_n)_{n\geq 1}$ is *(uniformly) bounded* in L^p if

$$\sup_{n\geq 1} \mathbb{E}\left[|X_n|^p\right] < \infty.$$

Lemma 3.6. *If $(X_n)_{n\geq 1}$ is uniformly integrable, then it is bounded in L^1.*

Proof. For an uniformly integrable family of random variables $(X_n)_{n\geq 1}$ we can choose $K > 0$ such that $\sup_{n\geq 1} \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>K\}}\right] \leq 1$. Since for all $X_n, n \geq 1$, we have

$$\begin{aligned}\mathbb{E}[|X_n|] &= \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>K\}}\right] + \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|\leq K\}}\right] \\ &\leq \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>K\}}\right] + K\end{aligned}$$

we get $\sup_{n\geq 1} \mathbb{E}[|X_n|] \leq 1 + K$. Therefore $(X_n)_{n\geq 1}$ is bounded in L^1. □

The converse of the statement is not true.

Example 3.7. Let $\Omega := [0,1], \mathcal{F} := \mathcal{B}_{[0,1]}, \mathbb{P} := \lambda|_{[0,1]}$ and $X_n := n \cdot \mathbb{1}_{[0,\frac{1}{n}]}$ for $n \geq 1$. We have $\mathbb{E}[|X_n|] = n \cdot \mathbb{P}\left[[0,\frac{1}{n}]\right] = 1$ for all $n \geq 1$, and therefore $(X_n)_{n\geq 1}$ is bounded in L^1. But for given $c > 0$ we have $\mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|\geq c\}}\right] = \mathbb{E}[|X_n|] = 1$ for $n > c$. Thus $(X_n)_{n\geq 1}$ is not uniformly integrable.

Theorem 3.8. *For a sequence $(X_n)_{n\geq 1}$ of integrable random variables the following are equivalent:*

- *(i) The sequence $(X_n)_{n\geq 1}$ is uniformly integrable.*
- *(ii) $(X_n)_{n\geq 1}$ is bounded in L^1 and $\forall \varepsilon > 0 \; \exists \delta > 0$ such that for all $A \in \mathcal{F}$ we have*

$$\mathbb{P}[A] < \delta \;\Rightarrow\; \sup_{n\geq 1} \mathbb{E}\left[|X_n|\mathbb{1}_A\right] < \varepsilon \tag{3.4}$$

Proof. (i) $\Rightarrow$ (ii): We know from Lemma 3.6 that we get boundedness in L^1 from uniform integrability. To show (3.4) let $\varepsilon > 0$. We choose $c(\varepsilon)$ such that for every $n \geq 1$

$$\mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>c(\varepsilon)\}}\right] < \frac{\varepsilon}{2}.$$

With $\delta := \frac{\varepsilon}{2c(\varepsilon)}$ we get $\forall A \in \mathcal{F}$ with $\mathbb{P}[A] < \delta$:

$$\begin{aligned}\mathbb{E}\left[|X_n|\mathbb{1}_A\right] &= \mathbb{E}\left[|X_n|\mathbb{1}_A\mathbb{1}_{\{|X_n|\leq c(\varepsilon)\}}\right] + \mathbb{E}\left[|X_n|\mathbb{1}_A\mathbb{1}_{\{|X_n|>c(\varepsilon)\}}\right] \\ &\leq c(\varepsilon) \cdot \mathbb{P}[A] + \mathbb{E}\left[|X_n|\mathbb{1}_{\{|X_n|>c(\varepsilon)\}}\right] \\ &< c(\varepsilon) \cdot \frac{\varepsilon}{2c(\varepsilon)} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

for every $n \geq 1$ and therefore (3.4) follows.
(ii) $\Rightarrow$ (i): Let $\varepsilon > 0$. We choose $\delta > 0$ such that (3.4) holds and set

$$c_0 := \frac{1}{\delta} \cdot \sup_{n\geq 1} \mathbb{E}\left[|X_n|\right] < \infty.$$

Using Markov's inequality we get $\forall n \geq 1$ and $c \geq c_0$

$$\mathbb{P}\left[|X_n| \geq c\right] \leq \frac{\mathbb{E}\left[|X_n|\right]}{c} < \delta.$$

Therefore, we can use (3.4) for the event $\{|X_n| \geq c\}$ to get for all $c \geq c_0$

$$\sup_{n\geq 1} \mathbb{E}\left[|X_n| \mathbb{1}_{\{|X_n|\geq c\}}\right] < \varepsilon.$$

Hence, $(X_n)_{n\geq 1}$ is uniformly integrable. $\square$

The example before the Theorem illustrates that L^1-boundedness is not sufficient for uniform integrability. Every slightly stronger growing function than the identity leads to uniform integrability as the following Theorem shows:

Theorem 3.9. *Let $(X_n)_{n\geq 1}$ be a family of integrable random variables and $\phi : \mathbb{R} \to \mathbb{R}^+$ a measurable function with $\lim_{t\to\infty} \frac{\phi(t)}{t} = \infty$. If*

$$M := \sup_{n\geq 1} \mathbb{E}\left[\phi\left(|X_n|\right)\right] < \infty$$

then $(X_n)_{n\geq 1}$ is uniformly integrable.

Proof. Let $\varepsilon > 0$ and set $K := \frac{M}{\varepsilon}$. By assumption there exists c_0 such that $\frac{\phi(t)}{t} \geq K$ for all $t \geq c_0$. Therefore we get for all $c \geq c_0$ and $n \geq 1$

$$\begin{aligned}
\mathbb{E}\left[|X_n| \mathbb{1}_{\{|X_n|\geq c\}}\right] &\leq \mathbb{E}\left[\frac{1}{K}\phi\left(|X_n|\right) \mathbb{1}_{\{|X_n|\geq c\}}\right] \\
&\leq \mathbb{E}\left[\frac{1}{K}\phi\left(|X_n|\right)\right] \\
&\leq \frac{M}{K} = \varepsilon,
\end{aligned}$$

which implies uniform integrability of $(X_n)_{n\geq 1}$. $\square$

Corollary 3.10. *Let $(X_n)_{n\geq 1}$ be bounded in L^p for $p > 1$, then $(X_n)_{n\geq 1}$ is uniformly integrable.*

Proof. Follows from Theorem 3.9 with $\phi(t) := |t|^p$. $\square$

Now we are able to state the main Theorem of this subsection. It can be seen as an improved Dominated Convergence Theorem.

Theorem 3.11 (**Recipe for proving L^1-convergence**).
For a random variable X and a sequence $(X_n)_{n\geq 1}$ of integrable random variables the following are equivalent:

(i) The sequence $(X_n)_{n\geq 1}$ is uniformly integrable and $X_n \xrightarrow{\mathbb{P}} X$.

(ii) $X \in L^1$ and $X_n \xrightarrow{L^1} X$

Proof. (i) $\Rightarrow$ (ii): Since $X_n \xrightarrow{\mathbb{P}} X$ it follows that there is a subsequence $(X_{n_k})_{k\geq 1}$ such that $X_{n_k} \xrightarrow{a.s.} X$. Since $x \mapsto |x|$ is continuous we have $|X_{n_k}| \xrightarrow{a.s.} |X|$ by Remark 2.29. Since $(|X_{n_k}|)_{k\geq 1}$ is a positive valued sequence of random variables we can use Fatou's Lemma to obtain

$$\mathbb{E}\left[|X|\right] \leq \liminf_{k\to\infty} \mathbb{E}\left[|X_{n_k}|\right] \leq \sup_{k\geq 1} \mathbb{E}\left[|X_{n_k}|\right] \leq \sup_{n\geq 1} \mathbb{E}\left[|X_n|\right] < \infty, \tag{3.5}$$

where the last inequality follows from Lemma 3.6. Hence, $X \in L^1$.
Let $\varepsilon > 0$. Since $(X_n)_{n\geq 1}$ and $\{X\}$ are uniform integrable the family $\{|X_n|+|X| : n \geq 1\}$ is uniformly integrable by Remark 3.4(v). By Theorem 3.8 there exists $\delta > 0$ such that

$$\mathbb{E}\left[(|X_n| + |X|)\,\mathbb{1}_{\{|X_n - X| > \frac{\varepsilon}{2}\}}\right] < \frac{\varepsilon}{2}$$

whenever

$$\mathbb{P}\left[|X_n - X| > \frac{\varepsilon}{2}\right] < \delta.$$

Since $X_n \xrightarrow{\mathbb{P}} X$ this holds for $n \geq n_0(\delta)$. Therefore, we get for all $n \geq n_0(\delta)$

$$\begin{aligned}
\mathbb{E}\left[|X_n - X|\right] &= \mathbb{E}\left[|X_n - X|\mathbb{1}_{\{|X_n - X| \leq \frac{\varepsilon}{2}\}}\right] + \mathbb{E}\left[|X_n - X|\mathbb{1}_{\{|X_n - X| > \frac{\varepsilon}{2}\}}\right] \\
&\leq \frac{\varepsilon}{2} + \mathbb{E}\left[(|X_n| + |X|)\,\mathbb{1}_{\{|X_n - X| > \frac{\varepsilon}{2}\}}\right] \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Since ε was arbitrary we get $X_n \xrightarrow{L^1} X$.

(ii) $\Rightarrow$ (i): Since L^1-convergence implies convergence in probability we only need to show uniform integrability of the sequence $(X_n)_{n\geq 1}$. Therefore, we use Theorem 3.8: First note that $(X_n)_{n\geq 1}$ is bounded in L^1 as it converges in L^1. Note also that for $A \in \mathcal{F}$ we have

$$\begin{aligned}
\mathbb{E}\left[|X_n|\mathbb{1}_A\right] = \mathbb{E}\left[|X_n - X + X|\mathbb{1}_A\right] &\leq \mathbb{E}\left[|X_n - X|\mathbb{1}_A\right] + \mathbb{E}\left[|X|\mathbb{1}_A\right] \\
&\leq \mathbb{E}\left[|X_n - X|\right] + \mathbb{E}\left[|X|\mathbb{1}_A\right]
\end{aligned}$$

Now let $\varepsilon > 0$ be given and choose n_0 such that $\mathbb{E}\left[|X_n - X|\right] < \frac{\varepsilon}{2}$ for $n \geq n_0$. Furthermore, since $\{X\}$ is uniformly integrable we can use Theorem 3.8 and choose $\delta_1 > 0$ s.t. $\mathbb{E}\left[|X|\mathbb{1}_A\right] < \frac{\varepsilon}{2}$ if $\mathbb{P}[A] < \delta_1$. Therefore we get for $n \geq n_0$ and $\mathbb{P}[A] < \delta_1$

$$\mathbb{E}\left[|X_n|\mathbb{1}_A\right] < \varepsilon.$$

Hence, $(X_n)_{n\geq n_0}$ is uniformly integrable. Since the finite family $(X_n)_{1\leq n<n_0}$ is uniformly integrable we can choose $\delta_2 > 0$ such that

$$\sup_{n<n_0} \mathbb{E}\left[|X_n|\mathbb{1}_A\right] < \varepsilon$$

if $\mathbb{P}[A] < \delta_2$. With $\delta := \delta_1 \wedge \delta_2$ we have for all $n \in \mathbb{N}$ that $\mathbb{E}\left[|X_n|\mathbb{1}_A\right] < \varepsilon$ if $\mathbb{P}[A] < \delta$ and the result follows from Theorem 3.8. $\square$

3.1.3 Martingale Difference Arrays

In the following we will consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 1}, \mathbb{P})$.

Definition 3.12. A real stochastic process $(X_n)_{n\geq 1}$ is a *martingale difference* (relative to the filtration $(\mathcal{F}_n)_{n\geq 1}$), if

(1) $(X_n)_{n\geq 1}$ is adapted to $(\mathcal{F}_n)_{n\geq 1}$,

(2) $\mathbb{E}[|X_n|] < \infty$ for all $n \geq 1$,

(3) $\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] = 0$ a.s. for all $n \geq 1$.

Note that a martingale difference $(X_n)_{n\geq 1}$ satisfies $\mathbb{E}[X_n] = \mathbb{E}\left[\mathbb{E}[X_n|\mathcal{F}_{n-1}]\right] = \mathbb{E}[0] = 0$ for all $n \geq 1$.

Remark 3.13. A martingale difference represents more or less the same stochastic property as a martingale. Indeed, if $(X_n)_{n\geq 1}$ is a martingale difference, then $M_n := \sum_{k=1}^n X_k$ with $n \geq 1$ is a martingale (relative to the same filtration). Conversely, if $(M_n)_{n\geq 1}$ is a martingale, then $X_n := M_n - M_{n-1}$ for $n \geq 2$ and $X_1 := M_1$ is a martingale difference.

Although martingale differences are in general not independent they are at least uncorrelated. Assuming $\mathbb{E}[X_k^2] < \infty$ for all $k \geq 1$ we get for $i < j$

$$\mathbb{E}[X_i X_j] = \mathbb{E}\left[\mathbb{E}[X_i X_j|\mathcal{F}_i]\right] = \mathbb{E}[X_i \underbrace{\mathbb{E}[X_j|\mathcal{F}_i]}_{=0}] = 0.$$

We shall study (the more general) so-called *martingale difference arrays*. These are arrays of random variables $(X_{ni})_{i=1}^{k_n} = \left((X_{ni})_{i=1}^{k_n}\right)_{n\geq 1}$

$$\begin{array}{ccccc} X_{11} & \dots & X_{1k_1} & & \\ X_{21} & X_{22} & \dots & X_{2k_2} & \\ X_{31} & X_{32} & X_{33} & \dots & X_{3k_3} \\ \vdots & \vdots & & & \end{array}$$

such that each row in the array forms a martingale difference. Therefore, let $(\mathcal{F}_{ni})_{i=1}^{k_n}$ be an array of sub-σ-algebras of $\mathcal{F}$ with $\mathcal{F}_{n,i-1} \subset \mathcal{F}_{ni}$. The assumptions for being a martingale difference array are therefore

(1) $(X_{ni})_{i=1}^{k_n}$ is adapted to $(\mathcal{F}_{ni})_{i=1}^{k_n}$, i.e. X_{ni} is $\mathcal{F}_{ni}$-measurable for all $n \geq 1$ and $1 \leq i \leq k_n$,

(2) $\mathbb{E}[|X_{ni}|] < \infty$ for all $n \geq 1$ and $1 \leq i \leq k_n$,

(3) $\mathbb{E}[X_{ni} \mid \mathcal{F}_{n,i-1}] = 0$ a.s. for all $n \geq 1$ and $1 \leq i \leq k_n$.

3.2 A first CLT for Martingale Difference Arrays

Before proving CLT's for martingale difference arrays we want to recall some well-known results to see that they arise naturally as a generalization of these Theorems. Like we said in the first paragraph it is possible to state a CLT for random variables which are neither necessarily identically distributed nor independent in a certain sense. The suitable framework for the statement of such more general results than the classical CLT is that of (triangular) arrays of random variables.

We start with the stronger assumption that the elements of each row are independent. We may then ask under which conditions the row sums $\sum_{i=1}^{k_n} X_{ni}$ converge in distribution to a normally distributed limit. There is no loss in generality by assuming that all the random variables in the array are centered and the sum of the variances in every row adds up to 1, i.e.

- $\mathbb{E}[X_{ni}] = 0$ for all n and i,
- $\sum_{i=1}^{k_n} \mathbb{E}[X_{ni}^2] = 1$ for all $n \geq 1$.

Since the random variables do not have to be identically distributed we need a condition which ensure roughly speaking that none of these random variables plays a too dominant role. Two conditions, the so-called Lindeberg- and Lyapunov-conditions, are known. We will only elaborate the first one since we use it later for generalizations.

Definition 3.14. An array of random variables $(X_{ni})_{i=1}^{k_n}$ with second moments satisfy the *Lindeberg-condition* if

$$\forall \varepsilon > 0 : \quad \sum_{i=1}^{k_n} \mathbb{E}\left[X_{ni}^2 \mathbb{1}_{\{|X_{ni}| \geq \varepsilon\}}\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{L}$$

The Lindeberg-condition makes precise in which sense the random variables have to be small. The contribution to the row variance from the terms with absolute value greater than ε becomes negligible as you go down the rows. The next Theorem is the most famous result for arrays of random variables:

Theorem 3.15 (Lindeberg-Feller-CLT).
Let $(X_{ni})_{i=1}^{k_n}$ be an array of random variables satisfying

(1) $X_{n1}, \ldots, X_{nk_n}$ are independent for all $n \geq 1$,

(2) (L).

Then $\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z \sim \mathcal{N}(0,1)$.

Proof. Cf. [10], Theorem 7.38 (in German) or [8], Theorem 15.43. □

Note that this Theorem generalizes the classical CLT. Indeed, for an iid-sequence of random variables with $\mathbb{E}[X_1] = 0$ and $\operatorname{Var}[X_1] = 1$ set $k_n := n$ and $X_{ni} := \frac{X_i}{\sqrt{n}}$. We have $\mathbb{E}[X_{ni}] = 0$ and $\sum_{i=1}^{k_n} \mathbb{E}[X_{ni}^2] = 1$. For $\varepsilon > 0$ we get

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}\left[\frac{X_i^2}{n} \mathbb{1}_{\left\{\left|\frac{X_i}{\sqrt{n}}\right| > \varepsilon\right\}}\right] &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i^2 \mathbb{1}_{\{|X_i| > \varepsilon\sqrt{n}\}}\right] \\ &= \mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1| > \varepsilon\sqrt{n}\}}\right].\end{aligned}$$

Since $X_1^2 \mathbb{1}_{\{|X_1| > \varepsilon\sqrt{n}\}} \xrightarrow{\mathbb{P}} 0$, $\left|X_1^2 \mathbb{1}_{\{|X_1| > \varepsilon\sqrt{n}\}}\right| \leq X_1^2$ and $\mathbb{E}[X_1^2] < \infty$ we can apply the Dominated Convergence Theorem to see that (L) is satisfied.

Even though we want to prove later a much more general result we look first at a Theorem due to Brown (see [2]) because this version can be better be understood as a generalization of the Lindeberg-Feller-Theorem 3.15. Instead of the independence in each row we now assume that each row forms a martingale difference sequence. Clearly, by weakening the assumptions we get an extra condition which deals with the conditional variances.

Theorem 3.16 (Brown, 1971).
Let $(X_{ni})_{i=1}^{k_n}$ be a martingale difference array satisfying

(B1) (L),

(B2) $\sum_{i=1}^{k_n} \mathbb{E}[X_{ni}^2 \mid \mathcal{F}_{n,i-1}] \xrightarrow{\mathbb{P}} 1.$

Then $\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z \sim N(0,1)$.

Proof. For a direct proof see [2]. □

We skipped the proof since we will show later that this Theorem follows from the main result of this chapter.

Remark 3.17. If the X_{ni} are independent for all $n \geq 1$, the Theorem becomes the Lindeberg-Feller-CLT.

3.3 General CLT for Martingale Difference Arrays

Before proving the main result of this chapter we need two general Lemmas.

Lemma 3.18. *Let $(U_n)_{n\geq 1}$ and $(T_n)_{n\geq 1}$ be sequences of random variables s.t. for $a \in \mathbb{R}$*

(i) $U_n \xrightarrow{\mathbb{P}} a$,

(ii) $(T_n)_{n\geq 1}$ *is uniformly integrable,*

(iii) $(T_n \cdot U_n)_{n\geq 1}$ *is uniformly integrable,*

(iv) $\mathbb{E}[T_n] \longrightarrow 1$.

Then $\mathbb{E}[T_n \cdot U_n] \longrightarrow a$.

Proof. We write $T_n \cdot U_n = T_n \cdot (U_n - a) + a \cdot T_n$. Since $\mathbb{E}[T_n] \longrightarrow 1$ we only have to show $\mathbb{E}[T_n \cdot (U_n - a)] \longrightarrow 0$. Therefore we show first that $T_n \cdot (U_n - a) \xrightarrow{\mathbb{P}} 0$
Since $(T_n)_{n\geq 1}$ is uniformly integrable we have particularly

$$\sup_{n\geq 1} \mathbb{E}\left[\mathbb{1}_{\{|T_n|\geq c\}}\right] = \sup_{n\geq 1} \mathbb{P}[|T_n| > c] \xrightarrow{c\to\infty} 0. \tag{3.6}$$

Therefore, we get

$$\begin{aligned}
\mathbb{P}\left[|T_n \cdot (U_n - a)| \geq \varepsilon|\right] &= \mathbb{P}\left[|T_n \cdot (U_n - a) \geq \varepsilon|, |T_n| > c\right] + \mathbb{P}\left[|T_n \cdot (U_n - a) \geq \varepsilon|, |T_n| \leq c\right] \\
&\leq \mathbb{P}[|T_n| > c] + \mathbb{P}\left[c \cdot |U_n - a| \geq \varepsilon|\right] \\
&\leq \sup_{n\geq 1} \mathbb{P}[|T_n| > c] + \mathbb{P}\left[|U_n - a| \geq \tilde{\varepsilon}\right]
\end{aligned}$$

with $\tilde{\varepsilon} = \frac{\varepsilon}{c}$. For $\delta > 0$ we can choose $c \in \mathbb{R}$ such that $\sup_{n\geq 1} \mathbb{P}[|T_n| > c] < \frac{\delta}{2}$ because of (3.6) and n big enough such that $\mathbb{P}\left[|U_n - a| \geq \tilde{\varepsilon}\right] < \frac{\delta}{2}$ since $U_n \xrightarrow{\mathbb{P}} a$. Therefore we get $T_n \cdot (U_n - a) \xrightarrow{\mathbb{P}} 0$.
As $(T_n \cdot U_n)_{n\geq 1}$ and $(a \cdot T_n)_{n\geq 1}$ are uniformly integrable $T_n \cdot (U_n - a)$ is uniformly integrable by Remark 3.4(v). From Theorem 3.11 it follows $T_n \cdot (U_n - a) \xrightarrow{L^1} 0$ and therefore $\mathbb{E}[T_n \cdot (U_n - a)] \longrightarrow 0$. □

Lemma 3.19. *Let $(X_{ni})_{i=1}^{k_n}$ be an array of random variables. For $n \geq 1$ and $t \in \mathbb{R}$ define $T_n(t) := \prod_{i=1}^{k_n} (1 + \mathrm{i}tX_{ni})$, where $\mathrm{i}^2 = -1$. Assume that*

(a) $\mathbb{E}[T_n(t)] \longrightarrow 1$ *for all* $t \in \mathbb{R}$,

(b) $(T_n(t))_{n\geq 1}$ *is uniformly integrable for all* $t \in \mathbb{R}$,

(c) $\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{\mathbb{P}} 1$,

(d) $\max_{1\leq i\leq k_n} |X_{ni}| \xrightarrow{\mathbb{P}} 0$.

Then $\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z \sim N(0, 1)$.

Proof. Use Taylor-expansion to get

$$\log(1+\mathrm{i}x) = \mathrm{i}x + \frac{1}{2}x^2 + r(x),$$

and therefore equivalently

$$e^{\mathrm{i}x} = (1+\mathrm{i}x)e^{-\left(\frac{1}{2}x^2 + r(x)\right)}$$

where the error term satisfies $|r(x)| \leq |x|^3$ for $|x| < 1$. Using this approximation for the complex valued random variable $e^{\mathrm{i}tS_n}$ with $S_n := \sum_{i=1}^{k_n} X_{ni}$ we obtain

$$\begin{aligned} e^{\mathrm{i}tS_n} &= \prod_{i=1}^{k_n} e^{\mathrm{i}tX_{ni}} \\ &= \prod_{i=1}^{k_n} (1+\mathrm{i}tX_{ni}) \exp\left\{-\frac{1}{2}t^2X_{ni}^2 - r(tX_{ni})\right\} \\ &= T_n(t) \cdot \underbrace{\exp\left\{-\frac{t^2}{2}\sum_{i=1}^{k_n} X_{ni}^2 - \sum_{i=1}^{k_n} r(tX_{ni})\right\}}_{=:\, U_n(t)} \end{aligned}$$

Hence, if we can show $\mathbb{E}\left[T_n \cdot U_n\right] \to e^{-\frac{t^2}{2}}$ the result will follow from Lévy's Continuity Theorem 3.1. Therefore, we will show that the conditions of Lemma 3.18 with $a = e^{-\frac{t^2}{2}}$ are satisfied.

In view of (d), for every $t \in \mathbb{R}$, $\max_{1\leq i\leq k_n} |tX_{ni}| < 1$ with probability tending to 1 as $n \to \infty$. For the error term in U_n we get therefore for large enough n

$$\begin{aligned} \left|\sum_{i=1}^{k_n} r(tX_{ni})\right| &\leq \sum_{i=1}^{k_n} |r(tX_{ni})| \\ &\leq |t|^3 \sum_{i=1}^{k_n} |X_{ni}|^3 \\ &\leq |t|^3 \left(\max_{1\leq i\leq k_n} |X_{ni}|\right) \sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{\mathbb{P}} 0, \end{aligned}$$

where we used (c) and (d) for the convergence to 0. By (c) we have $U_n \xrightarrow{\mathbb{P}} e^{-\frac{t^2}{2}}$. This shows condition (i). The conditions (ii) and (iv) are our present conditions (b) and (a), respectively. Since

$$|T_n \cdot U_n| = |e^{\mathrm{i}tS_n}| = 1$$

the sequence $(T_n \cdot U_n)_{n\geq 1}$ is uniformly integrable by Remark 3.4(iii) which shows condition (iii). $\square$

Now we apply this general result to a martingale difference array.

Theorem 3.20 (McLeish, 1974).
Let $(X_{ni})_{i=1}^{k_n}$ be a martingale difference array with

(1) $(\max_{1\leq i\leq k_n} |X_{ni}|)_{n\geq 1}$ *is bounded in* L^2*,*

(2) $\max_{1\leq i\leq k_n} |X_{ni}| \xrightarrow{\mathbb{P}} 0$*,*

(3) $\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{\mathbb{P}} 1$*.*

Then $\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z \sim \mathcal{N}(0,1)$*.*

Proof. We start with a cutoff by defining an array $(Z_{ni})_{i=1}^{k_n}$ by $Z_{n1} := X_{n1}$ and

$$Z_{ni} := X_{ni} \cdot \mathbb{1}_{\left\{\sum_{r=1}^{i-1} X_{nr}^2 \leq 2\right\}}$$

for $2 \leq i \leq k_n$. Since

$$\mathbb{E}[Z_{ni} \mid \mathcal{F}_{n,i-1}] = \mathbb{1}_{\{\sum_{r=1}^{i-1} X_{nr}^2 \leq 2\}} \mathbb{E}[X_{ni} \mid \mathcal{F}_{n,i-1}] = 0,$$

$(Z_{ni})_{i=1}^{k_n}$ is also a martingale difference array. We define

$$J_n := \inf\left\{ i : \sum_{r=1}^{i} X_{nr}^2 > 2 \right\} \wedge k_n.$$

Since the sum in the definition of the Z_{ni}'s only goes up to $i-1$, J_n is the last index in the $n-$th row with $X_{nJ_n} = Z_{nJ_n}$, the remains $Z_{n,J_n+1}, Z_{n,J_n+2}, \ldots$ are cut off to 0. Then

$$\begin{aligned}\mathbb{P}\left[X_{nr} \neq Z_{nr} \text{ for some } r \leq k_n\right] &= \mathbb{P}[J_n < k_n] \\ &\leq \mathbb{P}\left[\sum_{r=1}^{k_n} X_{nr}^2 > 2\right] \longrightarrow 0, \end{aligned} \tag{3.7}$$

since by condition (3) we have $\sum_{i=1}^{k_n} X_{ni}^2 \xrightarrow{\mathbb{P}} 1$. Let $\tilde{S}_n := \sum_{r=1}^{k_n} Z_{nr}$. We just proved $S_n - \tilde{S}_n \xrightarrow{\mathbb{P}} 0$. By the Converging Together Lemma 2.31 it is therefore enough to prove the result for $\tilde{S}_n$. Therefore, we show that $(Z_{ni})_{i=1}^{k_n}$ satisfies the condition of Lemma 3.19:

(c) and (d): By 3.7 the martingale difference array $(Z_{ni})_{i=1}^{k_n}$ also satisfy the conditions of the Theorem and therefore (c) and (d) are our present conditions (3) and (2), respectively.

(b): Let $T_n(t) := \prod_{i=1}^{k_n} (1 + \mathrm{i}tZ_{ni})$. Using the definition of J_n and the inequality $|1 + \mathrm{i}tx|^2 = 1 + t^2x^2 \leq e^{t^2x^2}$ for $x \in \mathbb{R}$ we get

$$\begin{aligned}\mathbb{E}[|T_n(t)|^2] &= \mathbb{E}\left[\prod_{i=1}^{k_n} \left(1 + t^2 Z_{ni}^2\right)\right] \\ &= \mathbb{E}\left[\left\{\prod_{i=1}^{J_n-1} \left(1 + t^2 X_{ni}^2\right)\right\} \cdot \left(1 + t^2 X_{nJ_n}^2\right)\right]\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{E}\left[\exp\left\{t^2 \underbrace{\sum_{i=1}^{J_n-1} X_{ni}^2}_{\le 2}\right\} \cdot (1+t^2X_{nJ_n}^2)\right] \\
&\le e^{2t^2}\left(1+t^2\mathbb{E}[X_{nJ_n}^2]\right) \\
&\le e^{2t^2}\left(1+t^2\mathbb{E}\left[\max_{1\le i\le k_n} X_{ni}^2\right]\right) =: M < \infty
\end{aligned}
$$

by (1). Since $(T_n)_{n\ge1}$ is bounded in L^2 it is uniformly integrable by Corollary 3.10.

(a): Since every $T_n(t)$ is in particular integrable we get

$$
\begin{aligned}
\mathbb{E}[T_n(t)] &= \mathbb{E}\left[\mathbb{E}[T_n(t) \mid \mathcal{F}_{n,k_n-1}]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{k_n-1}(1+\mathrm{i}tZ_{ni}) \cdot \mathbb{E}\left[(1+\mathrm{i}tZ_{nk_n} \mid \mathcal{F}_{n,k_n-1}\right]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{k_n-1}(1+\mathrm{i}tZ_{ni})\right] \\
&\ \vdots \\
&= \mathbb{E}[1+\mathrm{i}tZ_{n1}] = 1+\mathrm{i}t\mathbb{E}[Z_{n1}] = 1,
\end{aligned}
$$

where we used that $(Z_{ni})_{i=1}^{k_n}$ is a martingale difference array. □

Remark 3.21. If we take instead of (3) that $\sum_{i=1}^{k_n} X_{ni} \xrightarrow{\mathbb{P}} \sigma^2$ the proof can be done in the exactly same way and we get that $\sum_{i=1}^{k_n} X_{ni} \Rightarrow Z \sim \mathcal{N}(0,\sigma^2)$.

Remark 3.22. Conditions (1) and (2) of the Theorem can be replaced by

(1') $\max_{1\le i\le k_n}|X_{ni}| \xrightarrow{L^1} 0$.

The proofs for (c) and (d) are identical as above (for (d) note that L^1 convergence implies convergence in probability). For (b) we use the same inequality as above to get

$$
\begin{aligned}
|T_n(t)| &= \left\{\prod_{i=1}^{J_n-1}|1+\mathrm{i}tX_{ni}|\right\} \cdot |1+\mathrm{i}tX_{nJ_n}| \\
&= \left\{\prod_{i=1}^{J_n-1}\left(1+t^2X_{ni}^2\right)^{\frac{1}{2}}\right\} \cdot |1+\mathrm{i}tX_{nJ_n}| \\
&\le \exp\left\{\frac{t^2}{2}\sum_{i=1}^{J_n-1}X_{ni}^2\right\} \cdot (1+|t||X_{nJ_n}|) \\
&\le e^{t^2} \cdot \left(1+|t|\max_{1\le i\le k_n}|X_{ni}|\right).
\end{aligned}
$$

Since $\max_{1\le i\le k_n}|X_{ni}|$ converges in L^1 to 0 we can use Theorem 3.11 to see that the family $\{\max_{1\le i\le k_n}|X_{ni}|\}$ is uniformly integrable. Therefore, $(T_n)_{n\ge1}$ is uniformly integrable by Remark 3.4(iii). The proof for (a) is identical as above.

Next, we want to show that the Theorem of Brown 3.16 is really a special case of McLeish's Theorem 3.20. Therefore, we assume that (B1)=(L) and (B2) hold.

(B1) $\Rightarrow$ (1): We have

$$\begin{aligned}
X_{ni}^2 &\leq \varepsilon^2 + X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} \\
\Rightarrow X_{ni}^2 &\leq \varepsilon^2 + \sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} \quad \forall i \in \{1, \dots, k_n\} \\
\Rightarrow \max_{1\leq i\leq k_n} X_{ni}^2 &\leq \varepsilon^2 + \sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} \\
\Rightarrow \mathbb{E}\left[\max_{1\leq i\leq k_n} X_{ni}^2\right] &\leq \varepsilon^2 + \sum_{i=1}^{k_n} \mathbb{E}\left[X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}}\right]
\end{aligned}$$

Let $\xi > 0$. Since (L) is satisfied we can choose $\varepsilon > 0$ s.t. $2\varepsilon^2 < \xi$ and n_0 s.t. for $n \geq n_0$: $\sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} < \varepsilon^2$. Therefore, $\sup_{n\geq 1} \mathbb{E}\left[\max_{1\leq i\leq k_n} X_{ni}^2\right] < \xi < \infty$.

(B1) $\Rightarrow$ (2): Observe that (events are the same!)

$$\mathbb{P}\left[\max_{1\leq i\leq k_n} |X_{ni}| > \varepsilon\right] = \mathbb{P}\left[\sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} > \varepsilon^2\right].$$

The Lindeberg-condition (L) states that $\sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} \xrightarrow{L^1} 0$, which implies in particular that $\sum_{i=1}^{k_n} X_{ni}^2 \mathbb{1}_{\{|X_{ni}|>\varepsilon\}} \xrightarrow{\mathbb{P}} 0$. Hence, $\mathbb{P}\left[\max_{1\leq i\leq k_n} |X_{ni}| > \varepsilon\right] \longrightarrow 0$.

(B1)+(B2) $\Rightarrow$ (3): For arbitrary $\varepsilon > 0$ we put

$$Z_{ni} := X_{ni} \cdot \mathbb{1}_{\left\{|X_{ni}|\leq\varepsilon,\, \sum_{r=1}^{i} \mathbb{E}\left[X_{nr}^2 \mid \mathcal{F}_{n,r-1}\right] \leq 2\right\}}.$$

As in the proof of Theorem 3.20 we observe that

$$\mathbb{P}\left[X_{nr} \neq Z_{nr} \text{ for some } r \leq k_n\right] \longrightarrow 0$$

and

$$\sum_{r=1}^{k_n} \mathbb{E}\left[X_{nr}^2 - Z_{nr}^2 \mid \mathcal{F}_{n,r-1}\right] \xrightarrow{\mathbb{P}} 0$$

for $n \to \infty$. Moreover, we have

$$\begin{aligned}
\mathbb{E}\left[\left\{\sum_{r=1}^{k_n} Z_{nr}^2 - \mathbb{E}\left[Z_{nr}^2 \mid \mathcal{F}_{n,r-1}\right]\right\}^2\right] &= \sum_{r=1}^{k_n} \left\{\mathbb{E}\left[Z_{nr}^4\right] - \mathbb{E}\left[\mathbb{E}\left[Z_{nr}^2 \mid \mathcal{F}_{n,r-1}\right]\right]^2\right\} \\
&\leq \varepsilon^2 \mathbb{E}\left[\sum_{r=1}^{k_n} \mathbb{E}\left[Z_{nr}^2 \mid \mathcal{F}_{n,r-1}\right]\right] \\
&\leq 2\varepsilon^2.
\end{aligned}$$

Therefore we get for $\eta > 0$

$$\limsup_{n\to\infty} \mathbb{P}\left[\sum_{r=1}^{k_n} \left\{X_{nr}^2 - \mathbb{E}\left[X_{nr}^2 \mid \mathcal{F}_{n,r-1}\right]\right\} > \eta\right] \leq \frac{2\varepsilon^2}{\eta}.$$

Since ε was arbitrary this limit must be 0 and condition (3) follows from (B2).

To see that this general CLT has some nice applications we want to prove a CLT where we specify our array to a sequence of ergodic and stationary random variables:

Theorem 3.23. *Let $(Z_n)_{n\geq 1}$ be a stationary and ergodic sequence of random variables with $\mathbb{E}[Z_1^2] = \sigma^2 < \infty$ and $\mathbb{E}[Z_{n+1} \mid \mathcal{F}_n] = 0$ where $(\mathcal{F}_n)_{n\geq 1}$ is a filtration. Then we have*

$$\frac{1}{\sqrt{n}}(Z_1 + \ldots + Z_n) \Rightarrow Z \sim N(0, \sigma^2)$$

Proof. Let $X_{ni} := \frac{Z_i}{\sqrt{n}}$ and $\mathcal{F}_{ni} := \mathcal{F}_i$ for $1 \leq i \leq n := k_n$. Then $(X_{ni})_{i=1}^{k_n}$ is a martingale difference array with respect to the filtration $(\mathcal{F}_{ni})$. Therefore it is enough to show that the conditions (1') (from the Remark) and (3) with σ^2 instead of 1 from McLeish's Theorem 3.20 are satisfied:

(1'): We need to show that $\mathbb{E}[\max_{1\leq i\leq n} |X_{ni}|] \to 0$.

Claim: For a sequence $(T_k)_{k\geq 1}$ of identical distributed (not necessarily independent) random variables with $\mathbb{E}[T_1] < \infty$ we have $\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\max_{1\leq i\leq n} |T_i|] = 0$
Write

$$\begin{aligned} |T_i| &= |T_i|\mathbb{1}_{\{|T_i|\leq M\}} + |T_i|\mathbb{1}_{\{|T_i|>M\}} \\ &=: A_i + B_i \end{aligned}$$

Therefore we have

$$\max_{1\leq i\leq n} |T_i| \leq \max_{1\leq i\leq n} A_i + \max_{1\leq i\leq n} B_i$$

Clearly we have

$$\frac{1}{n}\mathbb{E}\left[\max_{1\leq i\leq n} A_i\right] \to 0.$$

For the second term note that we have for any random variable $H \geq 0$ that $\mathbb{E}[H] = \int_0^\infty \mathbb{P}[H \geq x]\, dx$. Hence, we get

$$\begin{aligned} \mathbb{E}\left[\max_{1\leq i\leq n} B_i\right] &= \int_0^\infty \mathbb{P}\left[\max_{1\leq i\leq n} \geq x\right]\, dx \\ &= \int_0^\infty \mathbb{P}\left[\bigcup_{i=1}^{n}\{B_i \geq x\}\right]\, dx \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{i=1}^{n} \int_0^\infty \mathbb{P}[B_i \geq x]\, dx \\
&= n \cdot \int_0^\infty \mathbb{P}[B_1 \geq x]\, dx \\
&= n \cdot \mathbb{E}\left[|T_1| \mathbb{1}_{\{|T_1|>M\}}\right]
\end{aligned}
$$

Since $\mathbb{E}[|T_1|] < \infty$ and M was arbitrary the claim follows.

Since the sequence (Z_i^2) is identically distributed, $\mathbb{E}[Z_1^2] < \infty$ by assumption we can use the claim with $T_i \coloneqq Z_i^2$ to obtain

$$
\mathbb{E}\left[\max_{1\leq i\leq n} \left|\frac{Z_i^2}{n}\right|\right] \longrightarrow 0
$$

and therefore

$$
\mathbb{E}\left[\max_{1\leq i\leq n} |X_{ni}|\right] = \mathbb{E}\left[\max_{1\leq i\leq n} \left|\frac{Z_i}{\sqrt{n}}\right|\right] \to 0.
$$

(3): We need to show that $\sum_{i=1}^n X_{ni}^2 \xrightarrow{\mathbb{P}} \sigma^2$.

By the Ergodic Theorem we have $\frac{1}{n}\sum_{i=1}^n Z_i \xrightarrow{a.s.} \mathbb{E}[Z_1]$. Note that for a stationary and ergodic sequence $(Y_j)_{j\geq 1}$ and measurable functions $f \geq 0$ the sequence $(f(Y_j))_{j\geq 1}$ is also stationary and ergodic. Using this fact with $f(x) = x^2$ we obtain

$$
\sum_{i=1}^n X_{ni}^2 = \frac{1}{n}\sum_{i=1}^n Z_i^2 \xrightarrow{a.s.} \mathbb{E}[Z_1^2] = \sigma^2
$$

and therefore also convergence in probability. $\square$

4 Interlude: Weak Convergence in $\mathbb{C}[0,1]$

Even though the càdlàg-space $\mathbb{D}[0,1]$ will eventually be our space of interest we will first develop some results for the space $\mathbb{C}[0,1]$ since it is easier to handle. The main reason for doing so is that all concepts can be carried over to the non-continuous case. In fact, in the next chapter we will mainly be concerned to generalize the results from this chapter. Note that we choose without loss of generality the interval $[0,1]$, nothing changes by looking at any other interval $[0,T]$ with $T < \infty$. Having in mind Theorem 2.13 we aim to find a separating class and some results about tightness. Helpful literature was [1], [6] and [14]. For elaborating the tightness conditions [3] was of great help.

4.1 A Separating Class for $\mathcal{B}_{\mathbb{C}}$

When specifying our general metric space $(S, \mathcal{B}_S)$ to the space $\mathbb{C}[0,1]$ we equip $\mathbb{C}[0,1]$ as usual with the *uniform metric* defining the distance between two points x and y (actually two continuous functions on $[0,1]$) as

$$\rho(x,y) = \sup_{0\leq t\leq 1} |x(t) - y(t)| = ||x-y||_\infty.$$

We define $\mathcal{B}_{\mathbb{C}}$ as the Borel σ-algebra on $\mathbb{C}[0,1]$, i.e. to be the smallest σ-algebra containing the open sets generated by ρ.

Remark 4.1. Note that every $x \in \mathbb{C}[0,1]$ is uniformly continuous since it is defined on the compact interval $[0,1]$.

Remark 4.2 (**($\mathbb{C}[0,1], \rho)$ is polish**).
As we will see separability and completeness facilitate derivation of sufficient conditions for weak convergence a lot. In this chapter we do not need to worry about it: $(\mathbb{C}[0,1], \rho)$ is complete because a uniform limit of continuous functions is continuous and it is separable because the set of polynomials on $[0,1]$ is dense in $\mathbb{C}[0,1]$ by the Weierstrass Approximation Theorem and the set of polynomials with rational coefficients is countable and dense in the set of all polynomials, hence also dense in $\mathbb{C}[0,1]$.

When studying infinite dimensional spaces the coordinate projections are often a helpful tool. We define therefore:

Definition 4.3. The mappings

$$\pi_t : (\mathbb{C}[0,1], \mathcal{B}_{\mathbb{C}}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$
$$x \longmapsto \pi_t(x) = x(t)$$

for $t \in [0,1]$ are called *one-dimensional projections.*

A separating class for $\mathcal{B}_{\mathbb{C}}$ is by definition a subset of $\mathcal{B}_{\mathbb{C}}$. Hence, if we want to define a family of sets as in (2.2) the map need to be measurable.

Lemma 4.4. *π_t is continuous and therefore $\mathcal{B}_{\mathbb{C}}$-$\mathcal{B}_{\mathbb{R}}$-measurable.*

Proof. Let $t \in [0,1]$. Since we have

$$|\pi_t(x) - \pi_t(y)| = |x(t) - y(t)| \leq \rho(x,y)$$

for all $x, y \in \mathbb{C}[0,1]$ it follows that π_t is (Lipschitz-)continuous for every t and therefore measurable. $\square$

In other words, the collection of preimages $\pi_t^{-1}(H)$ for $H \in \mathcal{B}_{\mathbb{R}}$ is a subset of $\mathcal{B}_{\mathbb{C}}$. We define therefore:

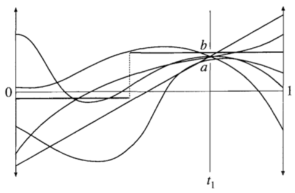
Definition 4.5. The family of sets given by

$$\mathcal{C}_0 := \left\{\pi_t^{-1}(H) : H \in \mathcal{B}_{\mathbb{R}}, t \in [0,1]\right\}$$

are called *one-dimensional* sets of $\mathbb{C}[0,1]$.

To get an idea of the sets in $\mathcal{C}_0$ we give an example:

Example 4.6. Let H be an interval $[a,b] \in \mathcal{B}_{\mathbb{R}}$. The set $A = \pi_{t_1}^{-1}([a,b]) \in \mathcal{C}$ with $t_1 \in [0,1]$ consists then of all functions that passing through a "gate" of width $b - a$ in a barrier erected at the point t_1.



They are a first serious candidate for a separating class. We will see that they generate $\mathcal{B}_{\mathbb{C}}$ but unfortunately they do not build a π-system as the following counterexample shows:

Example 4.7 ($\mathcal{C}_0$ is not a π-system).
Define the sets

$$A = \{x \in \mathbb{C}[0,1] \mid x(0) > 0\} = \pi_0^{-1}((0,\infty)) \text{ and}$$
$$B = \{x \in \mathbb{C}[0,1] \mid x(1) > 0\} = \pi_1^{-1}((0,\infty)).$$

Then $A \cap B$ cannot be written as $\{x \in \mathbb{C}[0,1] \mid x(t) \in H\} = \pi_t^{-1}(H)$ for any $H \in \mathcal{B}_{\mathbb{R}}$ and $t \in [0,1]$.

We go therefore a step further and define the larger collection of all finite-dimensional sets to avoid examples as above.

Definition 4.8. Let $\{t_1, \ldots, t_d\}$ be any finite collection of points in $[0,1]$ and $d \in \mathbb{N}$. The mappings

$$\begin{aligned} \pi_{t_1,\ldots,t_d} : (\mathbb{C}[0,1], \mathcal{B}_{\mathbb{C}}) &\to \left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right) \\ x &\longmapsto (x(t_1), \ldots, x(t_d)) \end{aligned}$$

are called *d-dimensional projections.*

They are again continuous since a mapping into $\mathbb{R}^d$ is continuous if every component mapping is which we know from Lemma 4.4. Therefore the d-dimensional projections are also $\mathcal{B}_{\mathbb{C}}$-$\mathcal{B}_{\mathbb{R}^d}$-measurable.

Definition 4.9. We call

$$\mathcal{C} := \{\pi^{-1}_{t_1,\ldots,t_d}(H) : d \in \mathbb{N}, 0 \le t_1 \le t_2 \le \ldots \le t_d \le 1, H \in \mathcal{B}_{\mathbb{R}^d}\}$$

the *finite-dimensional sets* of $\mathbb{C}[0,1]$.

Example 4.10. If H is a d-dimensional rectangle we can think of a set $A = \pi^{-1}_{t_1,\ldots,t_d}(H)$ similar as in Example 4.6 as the family of functions passing through "gates" erected at the points $t_1, \ldots, t_d$.

Note that we could have written the intersection in Example 4.7 as $\pi^{-1}_{0,1}(H)$ with $H = (0,\infty) \times (0,\infty) \in \mathcal{B}_{\mathbb{R}^2}$. More generally we have:

Proposition 4.11. *$\mathcal{C}$ is a π-system.*

Proof. Note that the index set defining a finite-dimensional set can always be enlarged. Indeed, if we want to enlarge t_1, t_2 to t_1, τ, t_2 with $t_1 < \tau < t_2$ we define the projection $\phi : \mathbb{R}^3 \to \mathbb{R}^2$ by $\phi(x,y,z) = (x,z)$ to obtain $\pi_{t_1,t_2} = \phi \circ \pi_{t_1,\tau,t_2}$. Hence, $\pi^{-1}_{t_1,t_2}(H) = \pi_{t_1,\tau,t_2}\left(\phi^{-1}(H)\right)$ and of course $\phi^{-1}(H) \in \mathcal{B}_{\mathbb{R}^3}$ if $H \in \mathcal{B}_{\mathbb{R}^2}$. The general case follows in the same way, just with more notational issues. It follows that two sets A and $\tilde{A}$ in $\mathcal{C}$ can be represented as $A = \pi_{t_1,\ldots,t_d}(H)$ and $\tilde{A} = \pi_{t_1,\ldots,t_d}(\tilde{H})$ for the same d-tuple $(t_1, \ldots, t_d)$. Therefore, $A \cap \tilde{A} = \pi_{t_1,\ldots,t_d}(H \cap \tilde{H}) \in \mathcal{C}$. □

In order to prove that they are a separating class we need by Lemma 2.10 in addition that they also generate $\mathcal{B}_{\mathbb{C}}$. Note therefore that $\sigma(\mathcal{C}_0) = \sigma(\mathcal{C})$. It is clear that $\sigma(\mathcal{C}_0) \subseteq \sigma(\mathcal{C})$. For the other direction note that the collection of all $B \in \mathcal{B}_{\mathbb{R}^d}$ such that $\pi^{-1}_{t_1,\ldots,t_d}(H)$ belongs to $\sigma(\mathcal{C}_0)$ is a σ-algebra and contains sets of the type $B = B_1 \times \ldots \times B_d$ where each $B_i \in \mathcal{B}_{\mathbb{R}}$. Hence, for notational issues it is enough to prove that $\mathcal{C}_0$ generates $\mathcal{B}_{\mathbb{C}}$.

Theorem 4.12. *We have $\sigma(\mathcal{C}) = \mathcal{B}_{\mathbb{C}}$. In particular, $\mathcal{C}$ is a separating class for $\mathcal{B}_{\mathbb{C}}$.*

Proof. $\sigma(\mathcal{C}) \subseteq \mathcal{B}_{\mathbb{C}}$: Since the projection π_t is measurable we have $\mathcal{C}_0 \subseteq \mathcal{B}_{\mathbb{C}}$ and therefore $\sigma(\mathcal{C}) \subseteq \mathcal{B}_{\mathbb{C}}$.

$\mathcal{B}_{\mathbb{C}} \subseteq \sigma(\mathcal{C})$: Since $\mathbb{C}[0,1]$ is separable any open set can be written as a countable union of open balls. Therefore it suffices to prove that any open ball in $\mathbb{C}[0,1]$ is in $\sigma(\mathcal{C})$. Observe that

$$\begin{aligned}\overline{B(x,\varepsilon)} &= \{y \in \mathbb{C}[0,1] : \rho(x,y) \leq \varepsilon\} \\ &= \bigcap_{r \in \mathbb{Q}\cap[0,1]} \{y \in \mathbb{C}[0,1] : |x(r) - y(r)| \leq \varepsilon\} \\ &= \bigcap_{r \in \mathbb{Q}\cap[0,1]} \pi_r^{-1}\left([x(r)-\varepsilon, x(r)+\varepsilon]\right).\end{aligned}$$

Since $\pi_r^{-1}\left([x(r)-\varepsilon, x(r)+\varepsilon]\right) \in \sigma(\mathcal{C})$ it follows that $\overline{B(x,\varepsilon)} \in \sigma(\mathcal{C})$. But every open ball can be written as

$$B(x,\varepsilon) = \bigcup_{n\geq 1} \overline{B(x, \varepsilon - \tfrac{1}{n})}$$

and therefore $\sigma(\mathcal{C})$ contains all open sets. □

Having found a separating class we look now at sequences of probability measures $(Q_n)_{n\geq 1} \in \mathcal{P}\left(\mathbb{C}[0,1]\right)$.

Definition 4.13. Let $Q \in \mathcal{P}\left(\mathbb{C}[0,1]\right)$. The probability measures $Q \circ \pi_{t_1,\ldots,t_d}^{-1}$ for $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$ and $d \in \mathbb{N}$ are called *finite-dimensional distributions* of Q.

Note that $Q \circ \pi_{t_1,\ldots,t_d}^{-1}$ are probability measures on the space $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$. The fact that $\mathcal{C}$ is a separating class means, in other words, that the finite-dimensional distributions uniquely determine these probability measures. At this point we see again that $\mathcal{C}_0$ cannot be a separating class since this would mean that a probability measure on the space $\mathbb{C}[0,1]$ would completely be determined by its one-dimensional marginal distributions which of course cannot be true.

Now assume that $Q_n \Rightarrow Q$. Since $\pi_{t_1,\ldots,t_d}$ is continuous we can use the Continuous Mapping Theorem 2.7 to get

$$Q_n \circ \pi_{t_1,\ldots,t_d}^{-1} \Rightarrow Q \circ \pi_{t_1,\ldots,t_d}^{-1}.$$

Now we could ask if the converse is also true. In other words, is $\mathcal{C}$ maybe even a convergence-determining class? Unfortunately this is false as the following Lemma shows.

Lemma 4.14. *$\mathcal{C}$ is not a convergence-determining class.*

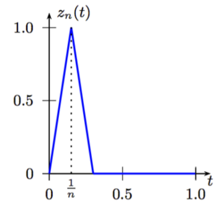
Proof. We give a counterexample. Take $z(t) = 0$ and

$$z_n(t) = nt\mathbb{1}_{[0,\frac{1}{n}]}(t) + (2 - nt)\mathbb{1}_{[\frac{1}{n},\frac{2}{n}]}(t)$$

for all $t \in [0, 1]$. Now define $Q_n := \delta_{z_n}$ and $Q := \delta_z$. Choose d and $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$. Then we have for any n, s.t. $\frac{2}{n}$ is smaller than the least non-zero t_i that $\pi_{t_1,\ldots,t_d}(z_n(t)) = (0, 0, \ldots, 0) = \pi_{t_1,\ldots,t_d}(z(t))$ and therefore

$$Q_n[\pi^{-1}_{t_1,\ldots,t_d}(H)] = Q[\pi^{-1}_{t_1,\ldots,t_d}(H)] \; \forall H \in \mathcal{B}_{\mathbb{R}^d}$$

Hence, $Q_n \circ \pi^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\ldots,t_d}$. But, by Lemma 2.8 we have $\delta_{z_n} = Q_n \not\Rightarrow Q = \delta_z$ since $z_n \not\to z = 0$. □



Nevertheless, since $\mathcal{C}$ is at least a separating class we can give a recipe for weak convergence in $\mathbb{C}[0, 1]$ by adding tightness of the sequence of probability measures. Note that we have equivalence in Prohorov's Theorem 2.18 since $(\mathbb{C}[0, 1], \mathcal{B}_{\mathbb{C}})$ is a polish space.

Theorem 4.15. *If $(Q_n)_{n\geq 1}$ is tight and $Q_n \circ \pi^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\ldots,t_d}$ for all $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$ and $d \in \mathbb{N}$, then we have $Q_n \Rightarrow Q$.*

Proof. Follows directly from Theorem 2.13 and Proposition 4.12. □

4.2 Tightness

As we have seen in the last Theorem we need a good understanding of tightness and therefore of compactness on $(\mathbb{C}[0, 1], \mathcal{B}_{\mathbb{C}})$ if we want to prove weak convergence. We are facing the problem that it is not easy to see if a given subset is compact or not. Recall that the Heine-Borel Theorem states that, in $\mathbb{R}^d$, a set that is closed and bounded is compact. In contrast, in infinite-dimensional normed vector spaces like $\mathbb{C}[0, 1]$ closed and bounded sets need not be compact. But maybe it is possible to get a similar result by adding an extra condition. Our aim is therefore a result which characterizes compactness in $\mathbb{C}[0, 1]$. We also consider sets which are not closed. Recall therefore that a set $\mathcal{A}$ is relatively compact if its closure $\bar{\mathcal{A}}$ is compact. Note also that a subset $\mathcal{A}$ of $\mathbb{C}[0, 1]$ is actually a family of functions.

Definition 4.16. Let $x : [0, 1] \to \mathbb{R}$ be an arbitrary function (not necessarily in $\mathbb{C}[0, 1]$). Given $\delta > 0$ we define the *modulus of continuity* by

$$w(x, \delta) = \sup_{\substack{t,s\in[0,1] \\ |t-s|<\delta}} |x(t) - x(s)|.$$

The modulus of continuity expresses information about the sensitivity of x, i.e. tells us something about the oscillation of x.

Remark 4.17 (Properties of $\boldsymbol{w(x, \delta)}$).

(i) For fixed δ we may think of $w(x, \delta)$ as a function on the domain $\mathbb{C}[0, 1]$. Since

$$|w(x, \delta) - w(y, \delta)| \leq 2\rho(x, y),$$

the modulus of continuity $w(x, \delta)$ is continuous on $\mathbb{C}[0, 1]$ and therefore also measurable.

(ii) By the definition it is clear that $w(x, \delta_1) \leq w(x, \delta_2)$ for any $\delta_1 < \delta_2$, i.e. w is an increasing function.

(iii) We have the following equivalence:

$$x \in \mathbb{C}[0, 1] \Leftrightarrow \lim_{\delta \to 0} w(x, \delta) = 0.$$

Indeed, for ($\Leftarrow$) note that $w(x, \delta) < \varepsilon$ implies in particular $|x(s) - x(t)| < \varepsilon$ for all s, t with $|s - t| < \delta$, hence the definition of uniform continuity.
For ($\Rightarrow$) let $x \in \mathbb{C}[0, 1]$. Suppose that $\lim_{\delta \to 0} w(x, \delta) \neq 0$. Since w is monotonic there is an $\varepsilon > 0$ such that $w(x, \delta) \geq \varepsilon$ for all $\delta > 0$. But by uniform continuity we can take $\frac{\varepsilon}{2} > 0$ and find a $\delta > 0$ such that $|t - s| < \delta$ implies $|x(t) - x(s)| < \frac{\varepsilon}{2}$, a contradiction.

Since we want to deal with families of function we need a stronger type of continuity.

Definition 4.18. A set $\mathcal{A} \subset \mathbb{C}[0, 1]$ is called *uniformly equicontinuous* if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t.} |x(t) - x(s)| < \varepsilon \text{ for all } x \in \mathcal{A} \text{ and } t, s \in [0, 1] \text{with } |t - s| < \delta.$$

We can write the definition in a more compact way as

$$\forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t.} \sup_{x \in \mathcal{A}} w(x, \delta) < \varepsilon$$

or as

$$\lim_{\delta \to 0} \sup_{x \in \mathcal{A}} w(x, \delta) = 0$$

Note that this definition of uniform equicontinuity is more general then just saying that every $x \in \mathcal{A}$ is uniformly continuous. Given an $\varepsilon > 0$ there must be a δ that will work for *every* $x \in \mathcal{A}$.

Definition 4.19. A set $\mathcal{A} \subset \mathbb{C}[0, 1]$ is called *uniformly bounded* (sometimes also called *equibounded*) if there is a constant C such that $|x(t)| \leq C$ for all $x \in \mathcal{A}$ and $t \in [0, 1]$.

Again, in a more compact way we can write

$$\sup_{x\in\mathcal{A}}\left(\sup_{t\in[0,1]}|x(t)|\right)<\infty.$$

Now, the famous Theorem from Arzelà-Ascoli states that a set $\mathcal{A}$ is relatively compact if and only if $\mathcal{A}$ is uniformly equicontinuous and uniformly bounded. Most of the time a weaker version is given which follows directly from the following Lemma.

Lemma 4.20. *Suppose $\mathcal{A}\subset\mathbb{C}[0,1]$ is such that*

(i) $\sup_{x\in\mathcal{A}}|x(0)|=M<\infty$

(ii) $\mathcal{A}$ is uniformly equicontinuous

Then $\mathcal{A}$ is uniformly bounded.

Proof. Given any $x\in\mathcal{A}$, $k\in\mathbb{N}$ and $t\in[0,1]$ we have

$$\begin{aligned}|x(t)|&=\left|x(0)+\sum_{i=1}^{k}\left(x\left(\frac{it}{k}\right)-x\left(\frac{(i-1)t}{k}\right)\right)\right|\\&\leq|x(0)|+\left|\sum_{i=1}^{k}\left(x\left(\frac{it}{k}\right)-x\left(\frac{(i-1)t}{k}\right)\right)\right|\end{aligned}$$

From (ii) we know that for every $\varepsilon>0$ exists $k\in\mathbb{N}$ large enough s.t. if $\delta\leq\frac{1}{k}$, then

$$\sup_{x\in\mathcal{A}}\left(\sup_{\substack{t,s\in[0,1]\\|t-s|<\delta}}|x(t)-x(s)|\right)<\varepsilon$$

Combining this with (i) and taking the supremum with respect to all $x\in\mathcal{A}$ on both sides in the inequality above we obtain

$$\sup_{x\in\mathcal{A}}|x(t)|<M+k\varepsilon<\infty.$$

Since this holds for all $t\in[0,1]$, $\mathcal{A}$ is equibounded. □

Hence, in the presence of uniform equicontinuity we can replace uniform boundedness by the weaker assumption of uniform boundedness at the origin.

Theorem 4.21 (Arzelà-Ascoli).
A set $\mathcal{A}\subset\mathbb{C}[0,1]$ is relatively compact if and only if

(i) $\sup_{x\in\mathcal{A}}|x(0)|<\infty$,

(ii) $\mathcal{A}$ is uniformly equicontinuous.

Proof. Cf. [1], Theorem 7.2 or [6], Theorem 18.3 for a more elaborate proof. □

Example 4.22. Let $\mathcal{A} := \{z_n : n \in \mathbb{N}\}$, where z_n is as in the counterexample of Proposition 4.12. We have $w(z_n, \delta) = 1$ for $\delta \geq \frac{1}{n}$ and therefore $\sup_{x\in\mathcal{A}} w(x, \delta) = 1$ for $\delta > 0$. Hence, $\mathcal{A}$ is not relatively compact since (ii) of the Arzelà-Ascoli Theorem is violated.

Now we want to translate this general result into a stochastic version, talking about probability measures instead of set's. The message of the following Theorem is that tightness of measures on $\mathbb{C}[0, 1]$ is equivalent to boundedness at the origin and continuity arising with sufficiently high probability in the limit.

Theorem 4.23 (Stochastic Arzelà-Ascoli).
Let $(Q_n)_{n\geq1} \in \mathcal{P}(\mathbb{C}[0, 1])$. Then $(Q_n)_{n\geq1}$ is tight if and only if the following two conditions hold:

(a) $\lim_{M\to\infty} \limsup_{n\to\infty} Q_n[\{x : |x(0)| > M\}] = 0$,

(b) $\forall\varepsilon > 0\ \lim_{\delta\to0} \limsup_{n\to\infty} Q_n [\{x : w(x, \delta) \geq \varepsilon\}] = 0$.

Remark 4.24. Condition (a) says that for all $\eta > 0$ there exists (large) $M = M_\eta$ and $N = N_\eta$ such that for all $n \geq N$ we have

$$Q_n [\{x : |x(0)| \geq M\}] \leq \eta.$$

Note also that

$$\begin{aligned} Q_n [\{x : |x(0)| > M\}] &= Q_n [\{x : x(0) \in [-M, M]^c\}] \\ &= Q_n [\{x : \pi_0(x) \in [-M, M]^c\}] \\ &= Q_n \left[\pi_0^{-1} ([-M, M]^c)\right] \\ &= \left(Q_n \circ \pi_0^{-1}\right) ([-M, M]^c) . \end{aligned}$$

Hence, (a) is the same as saying that the family of probability measures $\left(Q_n \circ \pi_0^{-1}\right)_{n\geq1}$, which are measures on the space $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, is tight.

Proof. ($\Rightarrow$) Suppose that $(Q_n)_{n\geq1}$ is tight. Given $\eta > 0$ we choose therefore a compact set K such that $Q_n[K] > 1 - \eta$ for all $n \geq 1$. By Theorem 4.21 there exists $M < \infty$ and $\delta \in (0, 1)$ such that

$$K \subseteq \{x : |x(0)| \leq M\} \cap \{x : w(x, \delta) < \varepsilon\}$$

for any $\varepsilon > 0$. Applying the De Morgan law we get

$$\begin{aligned} \eta \geq Q_n [K^c] &\geq Q_n [\{x : |x(0)| > M\}] \cup \{Q_n [\{x : w(x, \delta) \geq \varepsilon\}] \\ &\geq \max \{Q_n[\{x : |x(0)| > M\}], Q_n[\{x : w(x, \delta) \geq \varepsilon\}]\} . \end{aligned}$$

Hence, conditions (a) and (b) hold for all $n \geq 1$.

($\Leftarrow$) We define for $k \geq 1$ the sets

$$A_k \coloneqq \left\{x : w(x, \delta_k) < \tfrac{1}{k}\right\}$$

where $(\delta_k)_{k\geq 1}$ is a sequence chosen so that for $\tau > 0$ we have

$$\sup_{n\geq N} Q_n[A_k] > 1 - \frac{\tau}{2^{k+1}}.$$

for some finite $N \geq 1$. This is possible by condition (b) of the Theorem. Moreover, using condition (a) we can define the set

$$B \coloneqq \{x : |x(0)| \leq M\}$$

where we choose M such that

$$\sup_{n\geq N} Q_n[B] > 1 - \frac{\tau}{2}.$$

Now we define the set K as the closure of the intersection $\bigcap_{k\geq 1} A_k \cap B$. The conditions of Theorem 4.21 hold therefore for the set $\mathcal{A} = K$. Thus, K is (relatively) compact and we get

$$\begin{aligned}
\sup_{n\geq N} Q_n\left[K^c\right] &\leq \sup_{n\geq N} Q_n\left[\bigcup_{k\geq 1} A_k^c \cup B^c\right] \\
&\leq \sup_{n\geq N}\left\{\sum_{k\geq 1} Q_n\left[A_k^c\right] + Q_n\left[B^c\right]\right\} \\
&\leq \sum_{k\geq 1}\sup_{n\geq N} Q_n\left[A_k^c\right] + \sup_{n\geq N} Q_n\left[B^c\right] \\
&\leq \sum_{k\geq 1}\frac{\tau}{2^{k+1}} + \frac{\tau}{2} \\
&= \tau.
\end{aligned}$$

Since τ was arbitrary it follows that $(Q_n)_{n\geq N}$ is tight. To get the tightness of the whole sequence note that $\mathbb{C}[0,1]$ is polish and therefore every individual probability measure Q_n is tight by Lemma 2.17. This holds in particular for $n \in \{1, \ldots, N-1\}$ and therefore $(Q_n)_{n\geq 1}$ is tight. $\square$

Remark 4.25. Note that in the presence of convergence of the finite dimensional distributions we get condition (a) of Theorem 4.23 for free. Indeed, if $Q_n \circ \pi_{t_1,\ldots,t_d}^{-1} \Rightarrow Q \circ \pi_{t_1,\ldots,t_d}^{-1}$ we have in particular $Q_n \circ \pi_0^{-1} \Rightarrow Q \circ \pi_0^{-1}$. From Prohorov's Theorem 2.18 it follows that the sequence $\left(Q_n \circ \pi_0^{-1}\right)_{n\geq 1}$ is tight which is the same as condition (a) by Remark 4.24.

Combining the results we can give now a Theorem which can be seen as the cornerstone of the theory of weak convergence in $\mathbb{C}[0,1]$.

Theorem 4.26. *Let* $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(\mathbb{C}[0,1])$. *If*

1. $Q_n \circ \pi^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\ldots,t_d}$ *for all* $d \geq 1$ *and* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$ *and*

2. $\forall \varepsilon > 0$ $\lim_{\delta \to 0} \limsup_{n\to\infty} Q_n\left[\{x : w(x,\delta) \geq \varepsilon\}\right] = 0$,

then $Q_n \Rightarrow Q$.

Proof. By Remark 4.25 both assumptions of the Stochastic Arzelà-Ascoli Theorem 4.23 are satisfied. Therefore $(Q_n)_{n\geq 1}$ is tight. Using Theorem 4.15 we get $Q_n \Rightarrow Q$. $\square$

5 The Space $\mathbb{D}[0,1]$

In this chapter we introduce the earlier mentioned càdlàg-space $\mathbb{D}[0,1]$ of real functions on $[0,1]$ that are right-continuous and have left-hand limits. The whole chapter can be seen as a preliminary for the next chapter where we want to study the asymptotic behaviour of such functions. Therefore it is indispensable to have a fine grasp of the space $\mathbb{D}[0,1]$ itself. Note again that we choose the unit interval for notational issues, every result holds for any interval $[0,T]$ with $T<\infty$. Our target is to transform the results from last chapter to the non-continuous case. Before doing so we will first study some basic properties of the space. Since many proofs in the last chapter depended on the separability and completeness we will search in the second section for an appropriate metric for $\mathbb{D}[0,1]$ before studying weak convergence by finding a separating class and examining tightness. The main course follows [1]. For a clear structure [11] was also of great help. Although most of the results are covered by [1] the proofs are most often taken from other literature for a better reading. Therefore we used especially [3] and [14].

Remark 5.1. In the most important Theorem of this thesis we prove that a stochastic process whose paths are lying in the space $\mathbb{D}[0,1]$ is converging to a Brownian Motion, a process with continuous paths. In this chapter we are going sometimes a step beyond and state or prove results which can also be used in the case where the limit process has non-continuous paths.

5.1 Basic Properties of Càdlàg-functions

For a function $x\in\mathbb{R}^{[0,1]}$ we define

$$x(t^+) \coloneqq \lim_{s\to t^+} x(s),\ \ t\in[0,1)\quad \text{(right-hand limit)}$$

$$x(t^-) \coloneqq \lim_{s\to t^-} x(s),\ \ t\in(0,1]\quad \text{(left-hand limit)}$$

Definition 5.2. The space

$$\mathbb{D}[0,1] \coloneqq \left\{x\in\mathbb{R}^{[0,1]} \mid x(t^+)=x(t)\ \forall t\in[0,1) \text{ and } x(t^-) \text{ exists } \forall t\in(0,1]\right\}$$

consisting of all functions that are right-continuous and have left-hand limits is called *càdlàg-space*.

Remark 5.3. The space is also often called *Skorohod-space*, named after the famous Ukrainian mathematician Anatoli Skorohod. We call a function $x\in\mathbb{D}[0,1]$ *càdlàg-function*. The name is a french acronym and stands for **c**ontinue **à** **d**roite, **l**imites **à** **g**auche.

Example 5.4 (**Basic examples of càdlàg-functions**).

(i) Any continuous function is clearly a càdlàg-function and therefore $\mathbb{C}[0,1]$ is a subspace of $\mathbb{D}[0,1]$.

(ii) A prototype of a non-continuous càdlàg-function is a step function having a jump at some point $a \in (0,1)$ whose value at a is defined to be the value *after* the jump, i.e. $x(t) = \mathbb{1}_{[a,1]}(t)$. To check that $x \in \mathbb{D}[0,1]$ note that x is continuous everywhere apart from the point a and we have

$$x(a^+) = \lim_{t \to a^+} x(t) = \lim_{t \to a^+} 1 = 1 = x(a) \quad \text{and} \quad x(a^-) = \lim_{t \to a^-} x(t) = \lim_{t \to a^-} 0 = 0.$$

Hence, $x(a^+) = x(a)$ and $x(a^-)$ exists.

(iii) More generally, the step function

$$x(t) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{[t_{i-1},t_i)}$$

with $\alpha_i \in \mathbb{R}$ and $0 = t_0 < t_1 < \ldots < t_n = 1$ and $x(1)$ arbitrary is a càdlàg-function. They are therefore natural models for the paths of stochastic processes with jumps.

We see that the main difference to the continuous case are the jumps which complicates things a lot. Luckily, càdlàg-functions cannot jump to wildly as we see in the next Lemma and Corollary. Therefore, we define the *jump size* at t for $x \in \mathbb{D}[0,1]$ and $t \in (0,1)$ with

$$\Delta x(t) \coloneqq x(t^+) - x(t^-).$$

Obviously, x has a discontinuity at t if and only if the jump size $\Delta x(t) \neq 0$. Hence, a càdlàg-function can only have jump discontinuities, also called discontinuities of the first kind. There is no possibility of isolated values $x(t)$ at discontinuity points $t \in [0,1)$ because that would contradict the right-continuity. However, there is the possibility that $x(1)$ is isolated.

We define the *set of jumps* of x by

$$K(x) \coloneqq \{t \in [0,1] : \Delta x(t) \neq 0\},$$

and *the set of jumps of x exceeding a given number ε* by

$$K_\varepsilon(x) \coloneqq \{t \in [0,1] : |\Delta x(t)| > \varepsilon\}.$$

The following Lemma is of central importance for the derivation of essential properties of càdlàg-functions and shows how the maximum number of jumps is limited in a certain way.

Lemma 5.5. *For each $x \in \mathbb{D}[0,1]$ and every $\varepsilon > 0$ exists a finite partition $0 = t_0 < t_1 < \ldots < t_k = 1$ such that we have for all $i = 1, \ldots, k$*

$$\sup_{s,t \in [t_{i-1}, t_i)} |x(s) - x(t)| < \varepsilon \tag{5.8}$$

Proof. Let

$$\tau := \sup_{0 \leq t \leq 1} \{[0,t) \text{ can be decomposed into finitely many intervals satisfying 5.8}\}$$

Since $x(0) = x(0^+)$ there is some $\tilde{s}$ such that $|x(s) - x(0)| < \frac{\varepsilon}{2}$ for all $s \leq \tilde{s}$. By the definition of τ we have therefore $\tau \geq \tilde{s} > 0$.
As τ is the supremum we can find a sequence $\tau_n \nearrow \tau$ such that the intervals $[0, \tau_n]$ can be decomposed for every n satisfying (5.8). By positivity of τ it's left-hand limit $x(\tau^-)$ exists and there is some n_0 such that $|x(\tau_n) - x(\tau^-)| < \varepsilon$ for all $n \geq n_0$. Therefore $[0, \tau)$ can be decomposed (take for example $t_0 = 0$, $t_1, \ldots, t_k$ given by the sequence τ_n and $t_{k+1} = \tau$)
To conclude we need to show $\tau = 1$. Assuming $\tau < 1$, right-continuity $x(\tau) = x(\tau^+)$ implies that there exists $\delta > 0$ s.t. for $0 \leq t' - t < \delta$ we have $|x(t') - x(t)| < \varepsilon$. Consequently, the result holds for $[0, t')$ which contradicts the fact that τ is the supremum unless $\tau = 1$. □

From the Lemma we get some important results about càdlàg-functions.

Corollary 5.6. *Let $x \in \mathbb{D}[0,1]$. Then we have*

(a) $K_\varepsilon(x)$ is finite for all $\varepsilon > 0$,

(b) $K(x)$ is at most countable,

(c) $\sup_{0 \leq t \leq 1} |x(t)| < \infty$.

Proof. (a): Assume $|K_\varepsilon(x)| = \infty$. Then we can find a sequence $(s_n)_{n \geq 1}$ in $[0,1]$ such that $s_i \neq s_j$ for $i \neq j$ and $|\Delta x(s_n)| \geq \varepsilon$ for all n. From Lemma 5.5 it follows that there must be a finite decomposition of $[0,1]$ with property (5.8). Hence, there must be a s_m with $s_m \neq t_i$ for all i which lies in one of the intervals $[t_{i-1}, t_i)$. Let $s_m \in [t_{p-1}, t_p)$ for $p \in \{1, \ldots, k\}$. Then we have

$$\sup_{s,t \in [t_{p-1}, t_p)} |x(s) - x(t)| \geq \Delta |x(s_m)| \geq \varepsilon,$$

which is a contradiction to Lemma 5.5.

(b): Since

$$\{t \in [0,1] : \Delta x(t) \neq 0\} = \bigcup_{m \in \mathbb{N}} \{t \in [0,1] : |x(t)| > \tfrac{1}{m}\}$$

it follows from (a) that $K(x)$ is a countable union of finite sets and therefore itself at most countable.

(c): If $\{t_0, \ldots, t_k\}$ is a decomposition of $[0,1]$ satisfying (5.8) we have

$$\sup_{0 \leq t \leq 1} |x(t)| \leq \sum_{j=1}^{k} |\Delta x(t_j)| + k \cdot \varepsilon < \infty,$$

where the first term is the amount of jumps which occur at the k points obtained by the Lemma 5.5, while the second term is the increment of x within these k intervals. $\square$

5.2 Two Metrics on $\mathbb{D}[0,1]$

As we have seen we have $\mathbb{C}[0,1] \subset \mathbb{D}[0,1]$. It was easy to define a topology on $\mathbb{C}[0,1]$ since we could induce from the most natural metric, namely the uniform metric ρ. Therefore, the first attempt would be to take ρ again as our metric. Unfortunately, for càdlàg-functions things are a bit more complicated. Two functions might be very similar, but have a discontinuity of similar (or equal) magnitude at slightly different places. The distance in the uniform metric is therefore very big.

Example 5.7. Let $x(t) = \mathbb{1}_{[a,1]}$ and $y(t) = \mathbb{1}_{[b,1]}$ for $t \in [0,1]$ and $a, b \in (0,1)$. If $a \neq b$, then we have $\rho(x,y) = 1$ even when a is very close to b.

From the example we can also deduce another drawback:

Proposition 5.8. *The space* $(\mathbb{D}[0,1], \rho)$ *is not separable.*

Proof. Define as in the example the functions $x_\alpha = \mathbb{1}_{[\alpha,1]}$ for $\alpha \in [0,1]$. If the space $(\mathbb{D}[0,1], \rho)$ were separable, any subset would be separable, in particular $\{x_\alpha\}_{\alpha \in [0,1]}$. But $\{x_\alpha\}_{\alpha \in [0,1]}$ is discrete and uncountable and therefore not separable. $\square$

This means that most of the convergence theory developed in the last chapter for $\mathbb{C}[0,1]$ will not apply to $(\mathbb{D}[0,1], \rho)$. We will deal with this problem by finding an another metric which turns $\mathbb{D}[0,1]$ into a polish space.

Remark 5.9. Note that this is not the only way to deal with this problem. Another approach is to work on the space $(\mathbb{D}[0,1], \rho)$ and exclude the pathological cases from the σ-field under consideration. See for example [12] for details. In this work we will use the approach pioneered by Skorohod by metrizing $\mathbb{D}[0,1]$ as a separable complete space. Although this is more technical it has the advantage that the concepts of more familiar spaces like $\mathbb{C}[0,1]$ can be modified.

The idea is to define a metric which also allows uniformly small deformations of the time scale. Before giving a general definition we want to look at a similar example as before.

Example 5.10. Let $x(t) = \mathbb{1}_{[\frac{1}{2},1]}(t)$ and $x_n(t) = \left(1 - \frac{1}{n}\right) \mathbb{1}_{[\frac{1}{2}-\frac{1}{n},1]}(t)$. Certainly we want that x_n converge to x. But even though the jump sizes and the jump-times of $x_n(t)$ converge to that of $x(t)$ we have $\rho(x, x_n) = 1 - \frac{1}{n}$ for every $n \geq 1$. We want to avoid that two jumps which are close to each other increase the distance of the functions. Therefore, a possible solution would be to allow a jump of the function x_n to move to a very close but different place, like the jump of x. Therefore, we allow deformations of the time scala to overlay vicinal jumps. Formally, we define functions $\lambda_n : [0,1] \to [0,1]$ given by

$$\lambda_n(t) = \begin{cases} \left(1 - \frac{2}{n}\right) t, \text{ if } t \in [0, \frac{1}{2}) \\ \left(1 + \frac{2}{n}\right) t - \frac{2}{n}, \text{ if } t \in [\frac{1}{2}, 1] \end{cases}$$

These functions are continuous, strictly increasing and we have $\lambda_n(0) = 0, \lambda_n(1) = 1$. Moreover, $\sup_{0\leq t\leq 1} |\lambda_n(t) - t| \to 0$ for $n \to \infty$. Now we get $x_n(\lambda_n(t)) = \left(1 - \frac{1}{n}\right) x(t)$ and therefore

$$\rho(x_n \circ \lambda_n, x) = \frac{1}{n} \overset{n\to\infty}{\longrightarrow} 0.$$

In other words, we rescaled the axis $[0,1]$ by a small amount and made x_n close to x.

This example motivates the following definition:

Definition 5.11. Let

$$\Lambda := \{\lambda : [0,1] \to [0,1] : \lambda \text{ continuous, strictly increasing, bijective}\}$$

be the set of *time-shifts*.

If $\lambda \in \Lambda$, note that $\lambda(0) = 0$ and $\lambda(1) = 1$. Now, we define for $x, y \in \mathbb{D}[0,1]$ the metric $d_S(x,y)$ to be the infimum of all $\varepsilon > 0$ such that there $\exists \lambda \in \Lambda$ with

$$\sup_{0\leq t\leq 1} |\lambda(t) - t| < \varepsilon$$

and

$$\sup_{0\leq t\leq 1} |x(t) - y(\lambda(t))| < \varepsilon.$$

Let id be the identity map, i.e. $id(t) = t$ for all $t \in [0,1]$. Then we can write in a more compact form

$$d_S(x,y) = \inf_{\lambda\in\Lambda} \left\{||\lambda - id||_\infty \vee ||x - y \circ \lambda||_\infty\right\}.$$

We call $d_S(x,y)$ the *Skorohod-metric*.

We can think of Λ as the set of all increasing graphs connecting the opposite corners of the unit square. Skorohod proposed actually four metrics, denoted as $\mathcal{J}_1, \mathcal{J}_2, \mathcal{M}_1$ and $\mathcal{M}_2$. As we shall not be concerned with the others we refer to d_S as "the" Skorohod metric.

Remark 5.12 (Properties of $\boldsymbol{d_S}$).

(i) If we take $\lambda = id$ in our definition of $d_S(x, y)$ we get $d_S(x, y) \leq \rho(x, y)$.

(ii) Since each $\lambda \in \Lambda$ fixes 0 and 1 we have

$$||x - y \circ \lambda||_\infty \geq \max\{|x(0) - y(0)|, |x(1) - y(1)|\}$$

and therefore $\max\{|x(0) - y(0)|, |x(1) - y(1)|\}$ is a lower bound for $d_S(x, y)$.

It is straightforward to check that d_S is indeed a metric. While in the uniform metric ρ two functions are close if their *vertical* abscission is uniformly small, the Skorohod metric d_S also allows that the *horizontal* abscission is small which might conform better to our intuition of vicinity in the case of càdlàg-functions. Not only the intuition is right, the next theorem explains the theoretical interest of d_S.

Proposition 5.13. *The space* $(\mathbb{D}[0, 1], d_S)$ *is separable.*

Proof. For $n \in \mathbb{N}$ let A_n be the collection of piecewise constant functions on the intervals $\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right)$ for $i \in \{1, \dots, 2^n\}$ which are only assume rational values. If we denote by A the union of the sequence $(A_n)_{n\geq 1}$, then A is a set of functions taking rational values at a set of points indexed on the dyadic rationals and hence countable. We want to show that A is dense in $\mathbb{D}[0, 1]$.
For $x \in \mathbb{D}[0, 1]$ and $\varepsilon > 0$ there exists by Lemma 5.5 a finite partition $\{t_0, \dots, t_k\}$ of $[0, 1]$ such that for all $i \in \{1, \dots, k\}$ we have

$$\sup_{s,t\in[t_{i-1},t_i)} |x(s) - x(t)| < \varepsilon.$$

Now let y be a piecewise constant function constructed on the same intervals $[t_{i-1}, t_i)$ taking rational values $y_1, \dots, y_k$ where y_i differs by no more than ε from a value assumed by x on that interval $[t_{i-1}, t_i)$. Then we have $d_S(x, y) < 2\varepsilon$. For $n \geq 1$ we choose now an $z \in A_n$ such that $z_j = y_i$ when $\frac{j}{2^n} \in [t_{i-1}, t_i)$. Since the dyadic rationals are dense in $[0, 1]$ we have $d_S(x, y) \to 0$ as $n \to \infty$. Hence, $d_S(x, z) \leq d_S(x, y) + d_S(y, z)$ is tending to a value not exceeding 2ε. By taking k large enough ε can be arbitrary small. It other words, for every $x \in \mathbb{D}[0, 1]$ and $\delta > 0$ we can find $z \in A$ such that $d_S(x, z) < \delta$. The function x is therefore a closure point of A and the countable set A therefore dense in $\mathbb{D}[0, 1]$. □

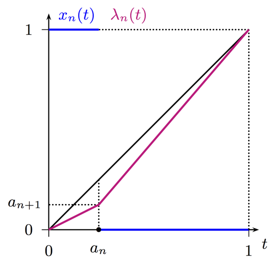
Remark 5.14. Note how this proof would fail under ρ. If x has discontinuities at the partitions points t_i the uniform distance $\rho(x, y)$ would be small only if the two sets of intervals overlap exactly such that $t_i = \frac{j}{2^n}$ for some j. If for example t_i were irrational this could not occur and x would fail to be a closure point of A.

Nevertheless, working with d_S will clearly complicate matters somewhat. The projections will be for example not continuous in general which complicates showing measurability of them. And there is another serious problem:

Proposition 5.15. *The space* $(\mathbb{D}[0,1], d_S)$ *is not complete.*

Proof. We give a counterexample. Let $a_n = \frac{1}{2^n}$ and $x_n(t) = \mathbb{1}_{[0,a_n)}(t)$ for $n \geq 1$. Let $\lambda_n(a_n) = a_{n+1}$ and λ_n affine on $[0, a_n]$ and on $[a_n, 1]$. Then we have $||\lambda_n - id||_\infty = a_{n+1}$. Furthermore, we have for all $n \geq 1$ that $x_{n+1} \circ \lambda_n = \mathbb{1}_{\{[0,a_{n+1})\}} \circ \lambda_n = x_n$ and therefore $||x_{n+1} \circ \lambda_n - x_n||_\infty = 0$. By definition of d_S we have then $d_S(x_n, x_{n+1}) \leq a_{n+1} = 2^{-(n+1)}$ and for $m \geq n$ we get

$$d_S(x_n, x_m) \leq d_S(x_n, x_{n+1}) + \ldots + d_S(x_{m-1}, x_m) = \sum_{k=n+1}^{m} 2^{-k} \leq 2^{-n}.$$



Thus, $(x_n)_{n\geq 1}$ forms a Cauchy sequence. Now, if a limit exists it must coincide with the pointwise limit of x_n, which is $x \equiv 0$ for all $t > 0$. But for all $\lambda \in \Lambda$ we have $||x_n \circ \lambda - x||_\infty = ||x_n \circ \lambda||_\infty = 1$ and therefore $d_S(x_n, x) = 1$. □

Hence, $(\mathbb{D}[0,1], d_S)$ is not polish and therefore the results from last chapter cannot be used. As $\mathbb{D}[0,1]$ is at least separable under d_S we don't want to drop this metric but rather find another metric which generates the same topology as d_S and therefore preserves separability.

Definition 5.16. Two metrics d_1 and d_2 on a set S are *equivalent* if, for each $x \in S$ and $\varepsilon > 0$, there is a $\delta > 0$ such that

$$d_1(x, y) < \delta \;\Rightarrow\; d_2(x, y) < \varepsilon$$
$$d_2(x, y) < \delta \;\Rightarrow\; d_1(x, y) < \varepsilon$$

for each $y \in S$.

Note that if two metrics d_1 and d_2 are equivalent a set is open with respect to d_1 if and only if it is open with respect to d_2. This is exactly what we want because then they generate the same topology. Sequences as in Proposition 5.15 should not be Cauchy sequences in this new metric anymore.

Remark 5.17. In the literature one finds creative alternatives to d_S. We will follow the approach from Billingsley (see [1]) as it is the most used result, denoting the new metric therefore as d_B.

The lack of completeness under d_S is roughly speaking due to the fact that the derivative of λ can be arbitrarily large at some points. From a topological point of view there is no need to use the supremum-norm on $\lambda - id$ in our definition of the metric. Since we want to regulate smoothness of the reparameterisation we could also use the following norm: For $\lambda \in \Lambda$ let

$$||\lambda|| := \sup_{s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \quad (\leq \infty).$$

Note that $||\lambda|| : \Lambda \to \mathbb{R}^+$ is a functional measuring the maximum deviation of the gradient of 1. For $\lambda(t) = t$ we have in particular $||\lambda|| = 0$. Now the definition of d_B requires that the slope of each chord is close to 1 or equivalently that the logarithm is close to 0. The logarithm enters the definition just because it is analytically more convenient. Although $||\lambda||$ can be infinite for $\lambda \in \Lambda$ these elements do not play a role in the following definition of our new metric

$$d_B(x, y) := \inf_{\lambda \in \Lambda} \left\{ ||\lambda||, ||x - \lambda \circ y||_\infty \right\}.$$

Again, it's not hard to prove that d_B is indeed a metric, see for example [3], Theorem 28.6.

Proposition 5.18. *The metrics d_S and d_B are equivalent and the space $(\mathbb{D}[0,1], d_B)$ is polish.*

Proof. Cf. [1], Theorem 12.1 and 12.2 or [3], Theorem 28.7 and 28.8. □

Definition 5.19. The topology induced by d_S or d_B is called *Skorohod-topology* denoted by $\mathcal{J}_1$.

Remark 5.20. Given a sequence $(x_n)_{n\leq 1}$, we have $d_B(x_n, x) \to 0$ if and only if $d_S(x_n, x) \to 0$, whenever $x \in \mathbb{D}[0,1]$. Note that this does not imply that $(x_n)_{n\geq 1}$ is a Cauchy sequence in $(\mathbb{D}[0,1], d_B)$ whenever it is a Cauchy sequence in $(\mathbb{D}[0,1], d_S)$ because the latter space is not complete and a sequence may have its limit outside the space.

Note that the Skorohod topology relativized to $\mathbb{C}[0,1]$ coincides with the uniform topology, i.e. d_S, d_B and ρ are equivalent on $\mathbb{C}[0,1]$.

5.3 Weak Convergence in $\mathbb{D}[0,1]$

Having found a metric under which $\mathbb{D}[0,1]$ is polish we are in a similar position as in $\mathbb{C}[0,1]$. Our target is therefore a version of Theorem 4.26 applying to measures on $\mathcal{P}(\mathbb{D}[0,1])$. The lack of continuity forces us to some new ideas. As before, we will first search for a separating class and will then establish a criterion for tightness.

5.3.1 A Separating Class for $\mathcal{B}_{\mathbb{D}}$

We denote by $\mathcal{B}_{(\mathbb{D}[0,1],d_B)} = \mathcal{B}_{\mathbb{D}}$ the Borel-σ-algebra on $(\mathbb{D}[0,1], d_B)$ which is by Proposition 5.18 the same as $\mathcal{B}_{(\mathbb{D}[0,1],d_S)}$. Recall that a sequence of functions in $(\mathbb{D}[0,1], d_B)$ converges by definition and Remark 5.20 if there $\exists \lambda_n \in \Lambda$ such that

$$\sup_{0\le t\le 1} |\lambda_n(t) - t| \overset{n\to\infty}{\longrightarrow} 0 \;\text{ and }\; \sup_{0\le t\le 1} |x_n(t) - x(\lambda_n(t))| \overset{n\to\infty}{\longrightarrow} 0$$

Taking $\lambda_n = id$ again, we see that

$$\rho(x_n, x) \to 0 \;\Rightarrow\; d_S(x_n, x) \to 0$$

The converse need not to be true. From $d_S(x_n, x) \to 0$ not even pointwise convergence for all $t \in [0,1]$ follows. But we can at least expect pointwise convergence for all continuity points. Uniform convergence follows only with an extra assumption:

Lemma 5.21. *Let $x_n, x \in \mathbb{D}[0,1]$ with $d_S(x_n, x) \to 0$. Then we have*

(i) $x_n(t) \to x(t)$ for all continuity points t of x,

(ii) if x is additionally continuous we have $x_n(t) \to x(t)$ uniformly.

Proof. (i): Let $\lambda_n \in \Lambda$ be such that $||\lambda_n - id||_\infty \to 0$ and $||x_n - x \circ \lambda_n||_\infty \to 0$. Triangle inequality gives

$$\begin{aligned}|x_n(t) - x(t)| &\le |x_n(t) - x(\lambda_n(t))| + |x(\lambda_n(t)) - x(t)| \\ &\le ||x_n - x \circ \lambda_n||_\infty + |x(\lambda_n(t)) - x(t)|.\end{aligned}$$

Now, by the choice of $\lambda_n \in \Lambda$ the first summand on the right-hand-side goes to 0. Since $||\lambda_n - id||_\infty \to 0$ we have $\lambda_n \to id$ and since x is continuous at t the second summand also goes to 0.

(ii): As in (i) we get

$$||x_n - x||_\infty \le ||x_n - x \circ \lambda_n||_\infty + ||x \circ \lambda_n - x||_\infty.$$

Since x is continuous on $[0,1]$ and therefore uniformly continuous the result follows. □

Our target is again to use Corollary 2.13. Having the case of $\mathbb{C}[0,1]$ in mind we hope that the finite-dimensional distributions plays a similar role in $\mathbb{D}[0,1]$. In $\mathbb{C}[0,1]$ it was easy to define the finite-dimensional sets since $\pi_{t_1,\dots,t_d}$ were continuous and therefore measurable. Therefore, if we want to define again finite-dimensional sets in $\mathbb{D}[0,1]$ we first have to check that the projections are measurable. Unfortunately we can not use continuity in this case as the following example shows. In order to avoid confusion with the projections defined on $\mathbb{C}[0,1]$ we denote the projections on $\mathbb{D}[0,1]$ by $\tilde{\pi}_t$ and $\tilde{\pi}_{t_1,\dots,t_d}$.

Example 5.22. Let $t \in (0,1)$. We define $x_n := \mathbb{1}_{[t+\frac{1}{n},1]} \in \mathbb{D}[0,1]$ and $x := \mathbb{1}_{[t,1]} \in \mathbb{D}[0,1]$. Then we have $d_S(x_n, x) \to 0$ and therefore $d_B(x_n, x) \to 0$. But since $\tilde{\pi}_t(x_n) = x_n(t) = 0$ and $\tilde{\pi}_t(x) = x(t) = 1$ it follows that $\tilde{\pi}_t$ is not continuous.

Theorem 5.23. *We have*

(i) $\tilde{\pi}_0$ *and* $\tilde{\pi}_1$ *are continuous,*

(ii) If $0 < t < 1$ *then:* $\tilde{\pi}_t$ *is continuous at* x *iff* x *is continuous at* t*,*

(iii) $\tilde{\pi}_t$ *is* $\mathcal{B}_{\mathbb{D}}/\mathcal{B}_{\mathbb{R}}$*-measurable and* $\tilde{\pi}_{t_1,\dots,t_d}$ *is* $\mathcal{B}_{\mathbb{D}}/\mathcal{B}_{\mathbb{R}^d}$*-measurable for all* $d \geq 1$ *and* $t_1, \dots, t_d \in [0,1]$.

Proof. (i): Since each $\lambda \in \Lambda$ fixes 0 and 1 we have for $i \in \{0,1\}$

$$d_S(x,y) \geq \inf_{\lambda\in\Lambda} ||x - y \circ \lambda||_\infty \geq |x(i) - y(i)| = |\pi_i(x) - \pi_i(y)| \, .$$

Therefore $\tilde{\pi}_0$ and $\tilde{\pi}_1$ are (Lipschitz)-continuous.

(ii): ($\Leftarrow$) If x_n converges to x in the Skorohod metric and x is continuous at t we have by Lemma 5.21 that $x_n(t) = \tilde{\pi}_t(x_n) \to x(t) = \tilde{\pi}_t(x)$ and hence $\tilde{\pi}_t(x)$ is continuous. ($\Rightarrow$): Assume that x is discontinuous at t. Then we have to show that $\tilde{\pi}_t(x)$ is discontinuous at x. Since x is a càdlàg-function it must have a jump discontinuity at t. Hence, $x(t^-) \neq x(t)$. Let $x_\varepsilon(t)$ be a shifted version of x defined by

$$x_\varepsilon(t) = \begin{cases} x(0), \text{ if } t \leq \varepsilon \\ x(t-\varepsilon) \text{ if } t > \varepsilon \end{cases}$$

Using the fact that x is right-continuous at 0 we get $d_S(x, x_\varepsilon) \leq \varepsilon$. Therefore, we have $\lim_{\varepsilon\to 0^+} x_\varepsilon = x$ in the Skorohod metric. But we get

$$\lim_{\varepsilon\to 0^+} \tilde{\pi}_t(x_\varepsilon) = \lim_{\varepsilon 0^+} x_\varepsilon(t) = x(t^-) \neq x(t) = \tilde{\pi}_t(x)$$

Therefore, $\tilde{\pi}_t(x)$ is not continuous at x.

(iii): A mapping into $\mathbb{R}^d$ is measurable if each component mapping is. Therefore, it suffices to show that $\tilde{\pi}_t$ is measurable to obtain both results. Since $\tilde{\pi}_1$ is continuous by (i) we can assume $t < 1$ and define for $x \in \mathbb{D}[0,1]$

$$h_\varepsilon(x) := \frac{1}{\varepsilon} \int_t^{t+\varepsilon} x(s) \; ds.$$

If $x_n \to x$ in the Skorohod topology we have by Lemma 5.21 pointwise convergence $x_n(t) \to x(t)$ for continuity points t of x, hence for all points outside of Lebesgue measure 0. Therefore, $x_n \xrightarrow{a.s.} x$. Since $(x_n)_{n\geq 1}$ is uniformly bounded we can use the Dominated Convergence Theorem to obtain

$$h_\varepsilon(x_n) \to h_\varepsilon(x).$$

Thus, h_ε is continuous and therefore measurable for every ε, implying that it's limit is measurable. By right-continuity of x we have therefore

$$h_\varepsilon(x) \xrightarrow{\varepsilon\to 0} x(t) = \tilde{\pi}_t(x).$$

Hence, $\tilde{\pi}_t$ is measurable. □

Knowing that the projection map is measurable we can now define as in $\mathbb{C}[0,1]$ the class of one- and finite-dimensional sets.

Definition 5.24. We define by

$$\mathcal{D}_0 := \{\tilde{\pi}_t^{-1}(H) : H \in \mathcal{B}_\mathbb{R}, t \in [0,1]\},$$
$$\mathcal{D} := \{\tilde{\pi}_{t_1,\dots,t_d}^{-1}(H) : d \in \mathbb{N}, 0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1, H \in \mathcal{B}_{\mathbb{R}^d}\}.$$

the *one*- and *finite-dimensional sets* in $\mathbb{D}[0,1]$.

We have again $\sigma(\mathcal{D}_0) = \sigma(\mathcal{D})$. By the same proof as in $\mathbb{C}[0,1]$ we get that $\mathcal{D}$ is a π-system. In order to show that $\mathcal{D}$ is again a separating class we need to show that $\mathcal{D}$ generates $\mathcal{B}_\mathbb{D}$.

Theorem 5.25. *We have $\sigma(\mathcal{D}) = \mathcal{B}_\mathbb{D}$. In particular, $\mathcal{D}$ is a separating class for $\mathcal{B}_\mathbb{D}$.*

Proof. For $x \in \mathbb{D}[0,1]$, $\varepsilon > 0$ and any collection $\{t_1,\dots,t_d\} \subseteq [0,1]$ we define

$$H_d(x,\varepsilon) := \left\{ y \in \mathbb{D}[0,1] : \exists \lambda \in \Lambda \text{ s.t. } ||\lambda|| < \varepsilon,\ \max_{1\leq i\leq d} |y(t_i) - x(\lambda(t_i)| < \varepsilon \right\}.$$

It can be shown (cf. [3], Lemma 28.9) that $H_d(x,\varepsilon) \in \mathcal{D}$. Note that an open ball in $(\mathbb{D}[0,1], d_B)$ is a set of the form

$$\begin{aligned} S(x,\varepsilon) &= \{y \in \mathbb{D}[0,1] : d_B(x,y) < \varepsilon\} \\ &= \left\{ y \in \mathbb{D}[0,1] : \exists \lambda \in \Lambda \text{ s.t. } ||\lambda|| < \varepsilon,\ \sup_{t\in[0,1]} |y(t) - x(\lambda(t))| < \varepsilon \right\} \end{aligned}$$

for $x \in \mathbb{D}[0,1]$ and $\varepsilon > 0$. Since $(\mathbb{D}[0,1], d_B)$ is separable these sets generate $\mathcal{B}_\mathbb{D}$. It is therefore enough to show that these sets can be constructed by countable unions and complements (and therefore also countable intersections) of sets in $\mathcal{D}$. Let therefore

$$H(x,\varepsilon) = \bigcap_{k\geq 1} H_k(x,\varepsilon),$$

where the $H_k(x,\varepsilon)$ are sets defined above with the extra condition that $d = 2^k - 1$ and $t_i = \frac{i}{2^k}$, so that the set $\{t_1, \ldots, t_d\}$ converges on the dyadic rationals as $k \to \infty$. Now consider $y \in H(x,\varepsilon)$. Since $y \in H_k(x,\varepsilon)$ for every k, we may choose a sequence $(\lambda_k)_{k\geq 1}$ such that for each $k \geq 1$ we have $||\lambda_k|| < \varepsilon$ and

$$\max_{1\leq i\leq 2^k-1} \left| y\left(\frac{i}{2^k}\right) - x\left(\lambda\left(\frac{i}{2^k}\right)\right)\right| < \varepsilon \tag{5.9}$$

Note that λ_k has (incidentally) all the properties of a cumulative distribution function on $[0,1]$. Therefore we can use Helly's Selection Theorem (cf. [8], Theorem 13.33) to find a subsequence $(\lambda_{k_n})_{n\geq 1}$ and a non-decreasing function λ with $\lambda(0) = 0$ and $\lambda(1) = 1$ such that

$$\lim_{n\to\infty} \lambda_{k_n}(t) = \lambda(t) \tag{5.10}$$

for all continuity points t of λ. To show that $\lambda \in \Lambda$ take two distinct continuity points s and t of λ. We have

$$\left|\log \frac{\lambda(t) - \lambda(s)}{t-s}\right| = \lim_{n\to\infty} \left|\log \frac{\lambda_{k_n}(t) - \lambda_{k_n}(s)}{t-s}\right| \leq \varepsilon$$

since $||\lambda_k|| < \varepsilon$ for all k. This relation makes a jump in λ impossible (implying that the convergence (5.10) holds for all t) and it implies that λ is strictly increasing, thus $\lambda \in \Lambda$. The inequality also implies

$$||\lambda|| \leq \varepsilon \tag{5.11}$$

Moreover, using the right-continuity of x on $[0,1)$ together with (5.9) and (5.10) we have either $|y(t) - x(\lambda(t))| \leq \varepsilon$ or $|y(t) - x(\lambda(t)^-)| \leq \varepsilon$ for every t in the dyadic rationals. But since the dyadic rationals are dense in $[0,1]$ this is equivalent to

$$\sup_{t\in[0,1]} |y(t) - x(\lambda(t))| \leq \varepsilon. \tag{5.12}$$

Note that both inequalities (5.11) and (5.12) are not strict anymore. Nevertheless, by the definition of $S(x,\varepsilon)$ we get at least $y \in \bar{S}(x,\varepsilon)$ and therefore $H(x,\varepsilon) \subseteq \bar{S}(x,\varepsilon)$. Now we put $\varepsilon = r - \frac{1}{n}$ and take the countable union to obtain

$$\bigcup_{n\geq 1} H\left(x, r - \tfrac{1}{n}\right) \subseteq \bigcup_{n\geq 1} \bar{S}\left(x, r - \tfrac{1}{n}\right) = S(x,r).$$

By the definition of $H_k(x,\varepsilon)$ it is also clear that $S(x,\varepsilon) \subseteq H_k(x,\varepsilon)$ for all $\varepsilon > 0$ and $k \geq 1$. Therefore, it follows that we have also

$$S(x,r) = \bigcup_{n\geq 1} S\left(x, r - \tfrac{1}{n}\right) \subseteq \bigcup_{n\geq 1} H\left(x, r - \tfrac{1}{n}\right).$$

Thus, for any $x \in \mathbb{D}[0,1]$ and $r > 0$ we have

$$S(x,r) = \bigcup_{n\geq 1} \bigcap_{k\geq 1} H_k\left(x, r - \tfrac{1}{n}\right),$$

where $H_k\left(x, r - \frac{1}{n}\right) \in \mathcal{D}$. □

Remark 5.26. By Remark 5.12(i) we have $\mathcal{B}_{(\mathbb{D}[0,1],d_B)} \subseteq \mathcal{B}_{(\mathbb{D}[0,1],\rho)}$ and therefore by Theorem 5.25 that $\sigma(\mathcal{D}) \subseteq \mathcal{B}_{(\mathbb{D}[0,1],\rho)}$. It can be shown that this is a strict subset. Hence, $\mathcal{D}$ would not be a separating class for $\mathcal{B}_{(\mathbb{D}[0,1],\rho)}$.

We consider now probability measures on $(\mathbb{D}[0,1], \mathcal{B}_{\mathbb{D}})$. Just as in the case of $\mathbb{C}[0,1]$ the finite dimensional distributions of a probability measure $Q \in \mathcal{P}(\mathbb{D}[0,1])$ are defined as the measures $Q \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d}$. But since these projections $\tilde{\pi}^{-1}_{t_1,\ldots,t_d}$ are not everywhere continuous we cannot use the Continuous Mapping Theorem 2.7 to conclude from $Q_n \Rightarrow Q$ that $Q_n \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d}$.

Example 5.27. Let Q_n be the Dirac measure on $\mathbb{1}_{[0,\frac{1}{2}+\frac{1}{n})}$ and Q the Dirac measure on $\mathbb{1}_{[0,\frac{1}{2})}$. By Lemma 2.8 we have $Q_n \Rightarrow Q$, but $Q_n \circ \tilde{\pi}_{\frac{1}{2}} \not\Rightarrow Q \circ \tilde{\pi}_{\frac{1}{2}}$.

On the other hand, the example given in 4.14 also applies here, i.e. from $Q_n \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d}$ we cannot conclude that $Q_n \Rightarrow Q$ again.

Recall our version of the Continuous Mapping Theorem 2.7. That it holds for continuous functions was just a special case of a more general result. It also applies for a discontinuous function as long as the set of discontinuity points has Q-measure 0.
To generalize the recipe to $\mathbb{D}[0,1]$ we need therefore a little detour. Note that we go a step beyond our needs and will therefore skip some proofs. For $T \subset [0,1]$ we look at the collection

$$\mathcal{D}(T) \coloneqq \{\tilde{\pi}^{-1}_{t_1,\ldots,t_d}(H) : d \in \mathbb{N}, 0 \le t_1 \le t_2 \le \ldots \le t_d \in T, H \in \mathcal{B}_{\mathbb{R}^d}\},$$

which is a subset of $\mathcal{D}$. Furthermore, we define for a probability measure $Q \in \mathcal{P}(\mathbb{D}[0,1])$ the set

$$T_Q \coloneqq \{t \in [0,1] : \pi_t \text{ is } Q\text{-a.s. continuous}\}.$$

By Theorem 5.23 (i) the points 0 and 1 lie in T_Q. For $0 < t < 1$ we have by Theorem 5.23 (ii) that

$$t \in T_Q \;\Leftrightarrow Q[J(t)] = 0,$$

where $J(t) \coloneqq \{x \in \mathbb{D}[0,1] : \Delta x(t) \neq 0\}$.

Lemma 5.28. *We have $Q[J(t)] > 0$ for at most countably many t.*

Proof. By Corollary 5.6 a point $x \in \mathbb{D}[0,1]$ can have at most countably many jumps. Let

$$J_\varepsilon(t) \coloneqq \{x \in \mathbb{D}[0,1] : |\Delta x(t)| > \varepsilon\}.$$

For fixed ε and δ suppose that $Q\left[J_\varepsilon(t_n)\right] \ge \delta$ for infinitely many different t_n. Then we get

$$Q\left[J_\varepsilon(t_n) \text{ i.o.}\right] = Q\left[\limsup_{n\to\infty} J_\varepsilon(t_n)\right] \geq \limsup_{n\to\infty} Q\left[J_\varepsilon(t_n)\right] \geq \delta$$

contradicting the fact that for a single $x \in \mathbb{D}[0,1]$ the jumps can exceed ε at only finitely many points. Thus we have $Q\left[J_\varepsilon(t)\right] > 0$ for at most countably many t. Since for $\varepsilon \to 0^+$ we have

$$Q\left[J_\varepsilon(t)\right] \nearrow Q[J(t)]$$

the claim follows. □

Corollary 5.29. *T_Q contains 0 and 1 and its complement in $[0,1]$ is at most countable.*

Proof. Since $Q\left[J_\varepsilon(t)\right] \nearrow Q[J(t)]$ for $\varepsilon \to 0^+$ we have

$$\{t \in [0,1] : Q[J(t)] > 0\} = \bigcup_{n\geq 1} \left\{t \in [0,1] : Q\left[J_{\frac{1}{n}}(t)\right] > 0\right\}$$

and therefore the complement of T_Q in $[0,1]$ is a countable union of sets which are at most countable. □

Hence, if all $t_1, \ldots, t_d$ lie in T_Q the function $\tilde{\pi}_{t_1,\ldots,t_d}$ is continuous on a set of Q-measure 1 and it follows from $Q_n \Rightarrow Q$ by the Continuous Mapping Theorem 2.7 that $Q_n \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \tilde{\pi}^{-1}_{t_1,\ldots,t_d}$. But now we have another problem. Since the collection $\mathcal{D}(T_Q)$ is a subset of $\mathcal{D}$ we do not know if it is still a separating class for $\mathcal{B}_{\mathbb{D}}$. Luckely, we have the following result:

Proposition 5.30. *If T contains 1 and is dense in $[0,1]$, then $\mathcal{D}(T)$ is a separating class. In particular, $\mathcal{D}(T_Q)$ is a separating class.*

Proof. Cf. [1], Theorem 12.5 or [14], Theorem 7.2. Having proved the first part it follows directly from 5.29 that $\mathcal{D}(T_Q)$ is a separating class. □

As a consequence we get know the recipe for $\mathbb{D}[0,1]$:

Theorem 5.31. *If $(Q_n)_{n\geq 1}$ is tight and $Q_n \circ \pi^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\ldots,t_d}$, whenever all $t_1, \ldots, t_d$ lie in T_Q, then we have $Q_n \Rightarrow Q$.*

Proof. Follows directly from Theorem 2.13 and Proposition 5.30. □

Note that if the limit process has continuous paths we have $T_Q = [0,1]$. In this case we can take any dense $T \subseteq [0,1]$, because automatically $T \subseteq T_Q$.

5.3.2 Tightness

To analyse tightness in $\mathbb{D}[0,1]$ we need first a tool to characterize compact sets. The main tool in the continuous case was the Arzelà-Ascoli Theorem stating that uniform boundedness plus uniform equicontinuity is sufficient (and also necessary) for a set to be relatively compact. We want to generalize this result for sets in the space $\mathbb{D}[0,1]$. Therefore, we will again need uniform boundedness but obviously uniform equicontinuity won't work as a criterion for discontinuous functions. Since $w(x,\delta)$ was a convenient tool for discriminating between functions in $\mathbb{C}[0,1]$ and functions outside the space we seek some analogue to the modulus of continuity that ignore jumps in a certain way. It will be again a helpful tool to discriminate between càdlàg functions and those with arbitrary discontinuities.

Definition 5.32. For $\delta > 0$ let Π_δ denote a partition $\{t_0, \dots, t_r\}$ of $[0,1]$ with the property $\min_{1\le i\le r}(t_i - t_{i-1}) > \delta$. We define the càdlàg-modulus by

$$\xi(x,\delta) = \inf_{\Pi_\delta} \left\{ \max_{1\le i\le r} \left\{ \sup_{s,t\in[t_{i-1},t_i)} |x(t) - x(s)| \right\} \right\}.$$

Remark 5.33 (Properties of $\boldsymbol{\xi(x,\delta)}$).

(i) The càdlàg-modulus weakens the modulus of continuity $w(x,\delta)$ in that $\xi(x,\delta)$ can be small even if the function jumps a lot. Indeed, intuitively this is clear since for small enough δ we can place the points of the partition t_i in such a way that jumps of the function x take place over the boundaries between adjacent parts of the mesh. Note also that the value of $x(1)$ does not affect the càdlàg-modulus $\xi(x,\delta)$.

(ii) For $\delta < \frac{1}{2}$ we have $\xi(x,\delta) \le w(x,\delta)$.
Indeed, for $\delta < \frac{1}{2}$ we can find a partition Π_δ of $[0,1]$ into subintervals $[t_{i-1}, t_i)$ such that $\delta < t_i - t_{i-1} \le 2\delta$ and therefore $\xi(x,\delta) \le w(x,\delta)$.
So obviously we have for any $x \in \mathbb{C}[0,1]$ that $\lim_{\delta\to 0} \xi(x,\delta) = 0$.

(iii) Using Lemma 5.5 one can easily show that

$$x \in \mathbb{D}[0,1] \;\Leftrightarrow\; \lim_{\delta\to 0} \xi(x,\delta) = 0.$$

Hence, the càdlàg-modulus $\xi(x,\delta)$ plays a similar role in $\mathbb{D}[0,1]$ as the modulus of continuity $w(x,\delta)$ in $\mathbb{C}[0,1]$.

We are now in a position to prove a generalized version of the Arzelà-Ascoli-Theorem. Before doing so we recall some basic notions for metric spaces which we will use later in the proof.

Definition 5.34. Let (S,d) be a metric space. A set $E \subseteq A \subseteq S$ is called an *ε-net* for A if

$$A \subseteq \bigcup_{x \in E} B_d(x,\varepsilon).$$

If $\forall \varepsilon > 0$ there is a finite ε-net for A the set A is called *totally bounded.*

If a set is relatively compact it is totally bounded (cf. [3], Theorem 5.5). Moreover, in a complete metric space the concepts of relative compactness and totally boundedness are equivalent, i.e. a set A is relatively compact if and only if it is totally bounded. (cf. [3], Theorem 5.13).

Theorem 5.35 (**Arzelà-Ascoli in $\mathbb{D}[0,1]$**).
A set $\mathcal{A} \subset \mathbb{D}[0,1]$ is relatively compact if and only if

(i) $\mathcal{A}$ is uniformly bounded

(ii) $\lim_{\delta\to 0}\sup_{x\in\mathcal{A}} \xi(x,\delta) = 0.$

Remark 5.36. For us the the only important part is the sufficiency since we will use it to prove tightness. For necessity we refer to [1] or to [3] for a more elaborate proof.

Proof. Since $(\mathbb{D}[0,1], d_B)$ is complete it is enough to show that the two conditions imply that $\mathcal{A}$ is totally bounded (with respect to d_B). Let us first show that $\mathcal{A}$ is totally bounded with respect to the more tractable metric d_S:
Let $\sup_{x\in\mathcal{A}}\sup_{t\in[0,1]}|x(t)| = M$ and fix $\varepsilon > 0$. By (ii) we can choose m as the smallest integer such that

$$\frac{1}{m} < \frac{\varepsilon}{2}$$

and

$$\sup_{x\in\mathcal{A}} \xi\left(x, \tfrac{1}{m}\right) < \frac{\varepsilon}{2}.$$

Now we construct the finite collection E_m of piecewise constant functions in $\mathbb{D}[0,1]$ with discontinuity points at $t = \frac{j}{m}$ for $j \in \{0,\ldots,m-1\}$. The values at the discontinuity points are drawn from the set $\left\{M\left(\frac{2u}{v}-1\right) : u \in \{0,\ldots,v\}\right\}$ where v is an integer exceeding $\frac{2M}{\varepsilon}$. This set contains $v+1$ real values between $-M$ and M which are less than ε apart, so that E_m has $(v+1)^m$ different elements. We show that E_m is an ε-net for $\mathcal{A}$:
By the definition of m we can choose for $x \in \mathcal{A}$ a partition $\Pi_{\frac{1}{m}} = \{t_0,\ldots,t_r\}$ as defined above with the property $\min_{1\le i\le r}(t_i - t_{i-1}) > \frac{1}{m}$ such that

$$\max_{1\le i\le r}\left\{\sup_{s,t\in[t_{i-1},t_i)}|x(t)-x(s)|\right\} < \frac{\varepsilon}{2}. \tag{5.13}$$

For $i \in \{0, \ldots, r-1\}$ let j_i be the integer such that

$$\frac{j_i}{m} \leq t_i < \frac{j_i + 1}{m}.$$

Since the t_i are at a distance more than $\frac{1}{m}$ there is at most one j_i in any of the intervals $\left[\frac{j}{m}, \frac{j+1}{m}\right)$ for $j \in \{0, \ldots, m-1\}$. Now we choose a piecewise linear function $\lambda \in \Lambda$ with vertices $\lambda\left(\frac{j_i}{m}\right) = t_i$ for $i \in \{0, \ldots, r\}$. We have $\left|t_i - \frac{j_i}{m}\right| \leq \frac{1}{m}$ and $\max_{0 \leq i \leq r} \left|\lambda\left(\frac{j_i}{m}\right) - \frac{j_i}{m}\right| \leq \frac{\varepsilon}{2}$. Together with the linearity of λ this implies

$$\sup_{t \in [0,1]} |\lambda(t) - t| \leq \frac{\varepsilon}{2}. \tag{5.14}$$

The function λ maps by construction points in $\left[\frac{j}{m}, \frac{j+1}{m}\right)$ into $[t_i, t_{i+1})$ whenever $j_i \leq j \leq j_{i+1}$. Since x varies by at most $\frac{\varepsilon}{2}$ over the intervals $[t_i, t_{i+1})$ the function $x \circ \lambda$ can at most vary $\frac{\varepsilon}{2}$ over the intervals $\left[\frac{j}{m}, \frac{j+1}{m}\right)$. Therefore, we can choose $y \in E_m$ such that

$$\left|y\left(\tfrac{j}{m}\right) - x\left(\lambda\left(\tfrac{j}{m}\right)\right)\right| < \frac{\varepsilon}{2} \tag{5.15}$$

for $j \in \{0, \ldots, m-1\}$. Since $y(t) = y\left(\frac{j}{m}\right)$ for $t \in \left[\frac{j}{m}, \frac{j+1}{m}\right)$ we get by (5.13) and (5.15)

$$\begin{aligned}\sup_{t \in [0,1]} |y(t) - x(\lambda(t))| &\leq \max_{0 \leq j \leq m-1} \left\{ \left|y\left(\tfrac{j}{m}\right) - x\left(\lambda\left(\tfrac{j}{m}\right)\right)\right| + \sup_{t \in \left[\frac{j}{m}, \frac{j+1}{m}\right)} \left|x\left(\lambda\left(\tfrac{j}{m}\right)\right) - x(\lambda(t)\right| \right\} \\ &< \varepsilon.\end{aligned}$$

Together with (5.14) this implies that $d_S(x, y) \leq \varepsilon$, showing that E_m is an ε-net for A. This proves that A is totally bounded in $(\mathbb{D}[0,1], d_S)$. But since d_S and d_B are equivalent A is also totally bounded in $(\mathbb{D}[0,1], d_B)$. In particular, if E_m is an ε-net for A in $(\mathbb{D}[0,1], d_S)$ then we can find κ such that it is an κ-net for A in $(\mathbb{D}[0,1], d_B)$, where κ can be set arbitrarily small. Since $(\mathbb{D}[0,1], d_B)$ is complete, $\bar{A}$ is compact. □

Again, we can translate this into a stochastic version for probability measures which directly parallels Theorem 4.23. Note that the conditions are exactly the same as in $\mathbb{C}[0,1]$ with $\sup_{t \in [0,1]} |x(t)|$ instead of $|x(0)|$ and ξ in place of w.

Theorem 5.37 (Stochastic Arzelà-Ascoli in $\mathbb{D}[0,1]$).
Let $(Q_n)_{n \geq 1} \in \mathcal{P}(\mathbb{D}[0,1])$. Then $(Q_n)_{n \geq 1}$ is tight if and only if the following two conditions hold:

(a) $\lim_{a \to \infty} \limsup_{n \to \infty} Q_n[\{x : \sup_{t \in [0,1]} |x(t)| > a\}] = 0,$

(b) $\forall \varepsilon > 0 \ \lim_{\delta \to 0} \limsup_{n \to \infty} Q_n\left[\{x : \xi(x, \delta) \geq \varepsilon\}\right] = 0.$

Proof. Change in all the definitions of the proof of Theorem 4.23 $|x(0)|$ by $\sup_{t \in [0,1]} |x(t)|$ and $w(x, \delta)$ by $\xi(x, \delta)$. Since $\mathbb{D}[0,1]$ is polish the proof follows that of 4.23 word for word. □

In the last chapter the first condition of Theorem 4.23 was automatically fulfilled if we assumed convergence of the finite-dimensional distributions. This is not the case in $\mathbb{D}[0,1]$. Hence, if we want to prove weak convergence we need to check $Q_n \circ \pi^{-1}_{t_1,\dots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\dots,t_d}$, whenever all $t_1,\dots,t_d$ lie in T_Q and *both* conditions of Theorem 5.37. In the literature most authors find some other equivalent conditions which are easier to check.

Since we deal in this work exclusively with the situation where the limit process is continuous we omit the details and refer to [1] for the more general case. We look therefore only for sufficient conditions in the special case where the limit is continuous. Note that we have done basically the essential work for the more general result where the limit is not continuous.

Hence, how to characterize a sequence in $\mathbb{D}[0,1]$ which is converging to a limit in $\mathbb{C}[0,1]$? The modulus of continuity w will be the natural medium for answering this question. It is therefore natural to ask what happens if we change the conditions in Theorem 5.37 with them of the continuous analogue Theorem 4.23.

Theorem 5.38. *Let $(Q_n)_{n\geq 1} \in \mathcal{P}(\mathbb{D}[0,1])$. If the two conditions*

(1) $\lim_{a\to\infty} \limsup_{n\to\infty} Q_n[\{x : |x(0)| > a\}] = 0$,

(2) $\forall \varepsilon > 0 \ \lim_{\delta\to 0} \limsup_{n\to\infty} Q_n\left[\{x : w(x,\delta) \geq \varepsilon\}\right] = 0$

hold, then $(Q_n)_{n\geq 1}$ is tight. Furthermore, if Q is the weak limit of a subsequence $(Q_{n_k})_{k\geq 1}$, then $Q\left[\{x : x \in \mathbb{C}[0,1]\}\right] = 1$.

Note that we have no equivalence anymore but just the sufficiency part.

Proof. We want to prove this result by checking the conditions of Theorem 5.37. Since $\xi(x,\delta) \leq w(x,\delta)$ for $\delta < \frac{1}{2}$ (see Remark 5.33(ii)) condition (b) of Theorem 5.37 is satisfied.
To see that (a) of 5.37 is satisfied let $k = \left\lfloor \frac{1}{\delta} \right\rfloor + 1$ where $\delta > 0$ is specified by condition (2). It follows by (2) that

$$Q_n\left[\left\{x : \left|x\left(\frac{ti}{k}\right) - x\left(\frac{t(i-1)}{k}\right)\right| \geq \varepsilon\right\}\right] \leq \eta$$

for $i \in \{1,\dots,k\}$ and $t \in [0,1]$. In Lemma 4.20 we have noted that

$$|x(t)| \leq |x(0)| + \sum_{i=1}^{k} \left|x\left(\frac{ti}{k}\right) - x\left(\frac{t(i-1)}{k}\right)\right|,$$

where each of the k intervals has width less than δ. Using (1) and (2) we get

$$Q_n\left[\left\{x : \sup_{t\in[0,1]} |x(t)| > M + k\varepsilon\right\}\right] \leq Q_n\left[\{x : |x(0)| > M\}\right] \leq \eta,$$

proving (i) of Theorem 5.37. Hence, $(Q_n)_{n\geq 1}$ is tight.

To show the other assertion, let μ be the weak limit along a subsequence, i.e. $Q_{n_k} \Rightarrow \mu$. We define $A = \{x : w(x,\delta) \geq \varepsilon\}$ and consider the open set A°, the interior of A. By Theorem 2.6(vi) we have

$$\mu\left[A^\circ\right] \leq \liminf_{n\to\infty} Q_{n_k}\left[A^\circ\right] \leq \eta.$$

Clearly we also have $\mu[B] \leq \eta$ for any $B \subset A^\circ$. Since ε and η are arbitrary we can choose a decreasing sequence $(\delta_j)_{j\geq 1}$ such that $\mu\left[B_j\right] \geq \frac{1}{j}$, where $B_j = \left\{x : w(x,\delta_j) \geq \frac{1}{j}\right\}$. For each $m \geq 1$ we have therefore $\mu\left[\bigcap_{j=m}^\infty B_j\right] = 0$. Now recall that

$$B \coloneqq \liminf_{j\to\infty} B_j = \bigcup_{m=1}^\infty \bigcap_{j=m}^\infty B_j.$$

Therefore we get by subadditivity that $\mu[B] = 0$. If $x \notin B$ it means that it is element of the set

$$B^c = \bigcap_{m=1}^\infty \bigcup_{j=m}^\infty B_j^c = \left\{x : w(x,\delta_j) < \tfrac{1}{j} \text{ for some } j \geq m \text{ and } m \in \mathbb{N}\right\}.$$

Since $(\delta_j)_{j\geq 1}$ is decreasing and therefore monotonic we have $\lim_{\delta\to 0} w(x,\delta) = 0$ for such x. By Remark 4.17(iii) it follows that $x \in \mathbb{C}[0,1]$. Hence $B^c \subset \mathbb{C}[0,1]$ and since $\mu\left[B^c\right] = 1$ the result follows. $\square$

As in $\mathbb{C}[0,1]$ the first condition follows from the convergence of the finite-dimensional distributions. Combining the results we get now our most important result for proving weak convergence in $\mathbb{D}[0,1]$

Theorem 5.39. *Let* $(Q_n)_{n\geq 1}, Q \in \mathcal{P}(\mathbb{D}[0,1])$. *If*

1. $Q_n \circ \pi^{-1}_{t_1,\ldots,t_d} \Rightarrow Q \circ \pi^{-1}_{t_1,\ldots,t_d}$ *for all* $d \geq 1$ *and* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$ *and*

2. $\forall \varepsilon > 0\ \lim_{\delta\to 0} \limsup_{n\to\infty} Q_n\left[x : w(x,\delta) \geq \varepsilon\right] = 0$,

then $Q_n \Rightarrow Q$ *and* $Q\left[\{x : x \in \mathbb{C}[0,1]\}\right] = 1$.

Proof. Follows directly from Theorem 5.31 and Theorem 5.38. $\square$

6 Invariance Principles

In chapter 3 we were concerned with the asymptotic behaviour of the partial sums S_n. We now go a step further and ask whether there is some regularity in the "whole path". To talk about convergence in a reasonable way we need to fit those paths into a common area, like the interval $[0,1]$. We produce therefore random elements of $\mathbb{C}[0,1]$ or $\mathbb{D}[0,1]$. In the first section we give a short and intuitive sketch of the famous invariance principle of Donsker and the corresponding result in $\mathbb{D}[0,1]$. For this part [1], [11] and [13] was helpful. In the second section we generalize these results to the the case where the random elements are built with martingale difference arrays. It is therefore the corresponding Functional Central Limit Theorem to the CLT 3.20 which is the cornerstone of this thesis and the main result of [9]. For the proof of this invariance principle we used in particular [4].

6.1 Motivation: Donsker's FCLT

All the results in chapter 4 and 5 were given in terms of arbitrary probability measures on the spaces $\mathbb{C}[0,1]$ and $\mathbb{D}[0,1]$. We will now interpret them as the distribution of certain random elements on these spaces. Therefore, recall section 3 of chapter 2 where we defined the distribution of a stochastic process and note that we deal here actually with a sequence of stochastic processes, interpreting every stochastic process as a random element of the space $\mathbb{C}[0,1]$ or $\mathbb{D}[0,1]$. We will denote them therefore by uppercase double-barred letters. To prove convergence of these processes one always uses the main results from the last chapters. Each of these results can be routinely reformulated in terms of random elements.

Note therefore that for a sequence $\left(\mathbb{W}^{[n]}\right)_{n\geq 1}$ of random elements of $\mathbb{C}[0,1]$ or $\mathbb{D}[0,1]$ convergence of finite dimensional distributions means

$$\mathbb{P}_{\mathbb{W}^{[n]}} \circ \pi^{-1}_{t_1,\ldots,t_d} = \mathbb{P} \circ \left(\pi_{t_1,\ldots,t_d} \circ \mathbb{W}^{[n]}\right)^{-1} = \mathbb{P}_{\pi_{t_1,\ldots,t_d} \circ \mathbb{W}^{[n]}} \Rightarrow \mathbb{P}_{\pi_{t_1,\ldots,t_d} \circ \mathbb{W}}.$$

Since $\pi_{t_1,\ldots,t_d}\left(\mathbb{W}^{[n]}\right) = \left(\mathbb{W}^{[n]}(t_1),\ldots,\mathbb{W}^{[n]}(t_d)\right)$ we have that the finite dimensional distributions of $\mathbb{W}^{[n]}$ converge to that of $\mathbb{W}$ if

$$\left(\mathbb{W}^{[n]}(t_1),\ldots,\mathbb{W}^{[n]}(t_d)\right) \Rightarrow \left(\mathbb{W}(t_1),\ldots,\mathbb{W}(t_d)\right).$$

for all $d \geq 1$ and $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$.

Note also that if $A = \{x : w(x,\delta) \geq \varepsilon\}$, then $\mathbb{W}^{-1}(A) = \{\omega \in \Omega : w(\mathbb{W},\delta) \geq \varepsilon\}$.
We get therefore the following Theorem, which we use later to prove the general invariance principle.

Theorem 6.1. *Let* $\mathbb{W}, \left(\mathbb{W}^{[n]}\right)_{n\geq 1}$ *be random elements in* $\mathbb{D}[0,1]$ *(or in* $\mathbb{C}[0,1]$*). If, for all* $d \geq 1$ *and* $0 \leq t_1 \leq t_2 \leq \ldots \leq t_d \leq 1$ *we have*

$$\left(\mathbb{W}^{[n]}(t_1), \ldots, \mathbb{W}^{[n]}(t_d)\right) \Rightarrow (\mathbb{W}(t_1), \ldots, \mathbb{W}(t_d))$$

and for all $\varepsilon > 0$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\left[\sup_{\substack{t,s\in[0,1]\\ |t-s|<\delta}} \left|\mathbb{W}^{[n]}(t) - \mathbb{W}^{[n]}(s)\right| > \varepsilon\right] = 0$$

then $\mathbb{W}^{[n]} \Rightarrow \mathbb{W}$.

Proof. Using the notational remarks above this Theorem is just a paraphrase of Theorem 5.39 (or Theorem 4.26). □

If we want to achieve convergence of stochastic processes on $[0,1]$ we need as in the Central Limit Theorems to scale them in an appropriate way.

To find the right rescaling let $X_1, X_2, \ldots$ be i.i.d. centered random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite variance $\sigma^2 > 0$. We fix a positive integer n and try to define a rescaled random walk $S_t^{(n)}$ with time steps of size $\frac{1}{n}$ by

$$S_{\frac{k}{n}} = c_n \cdot S_k$$

for $k \in \{0, 1, \ldots n\}$ and some constants $c_n > 0$. If t is a multiple of $\frac{1}{n}$, then

$$\mathrm{Var}[S_t^{(n)}] = c_n^2 \,\mathrm{Var}[S_{nt}] = c_n^2 nt\sigma^2.$$

Therefore, if we want to achieve convergence of $S_t^{(n)}$ we need to choose c_n proportional to $n^{-\frac{1}{2}}$.

Hence, if we define stochastic processes $\mathbb{S}^{[n]} = \left(S_t^{(n)}\right)_{t\in[0,1]}, n \geq 1$, this consideration makes clear how to define the values for all multiples of $\frac{1}{n}$. To obtain a continuous-time process we have to make clear what happens on the intervals $\left(\frac{k}{n}, \frac{k+1}{n}\right)$. Two common embeddings for the discrete sequence $\left(\frac{1}{\sqrt{n}} S_k\right)$ are the following:

$$S_t^{(n)} := \frac{1}{\sigma\sqrt{n}} \left(X_1 + \ldots + X_{\lfloor nt\rfloor}\right) + \frac{nt - \lfloor nt\rfloor}{\sqrt{n}} X_{\lfloor nt\rfloor+1} \tag{6.16}$$

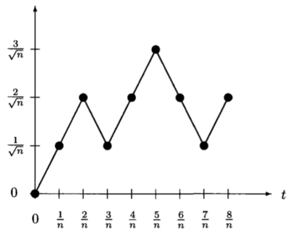
and

$$Y_t^{(n)} := \frac{1}{\sigma\sqrt{n}} \left(X_1 + \ldots + X_{\lfloor nt\rfloor}\right) \tag{6.17}$$

Note that for fixed n the process $\mathbb{S}^{[n]} = \left(S_t^{(n)}\right)_{t\in[0,1]}$ is a random element of $\mathbb{C}[0,1]$ while $\mathbb{Y}^{[n]} = \left(Y_t^{(n)}\right)_{t\in[0,1]}$ is a random element of $\mathbb{D}[0,1]$. Notice that we have

$$S^{(n)}_{\frac{k}{n}} = \frac{1}{\sigma\sqrt{n}} S_k$$

for $k \in \{0, 1, \ldots, n\}$. Hence, for fixed n the paths of the processes $\mathbb{S}^{[n]}$ are continuous functions defined by linear interpolation between its values $\frac{1}{\sigma\sqrt{n}} S_k$ at the points $\frac{k}{n}$. In particular, $\left(\mathbb{S}^{[n]}\right)_{n\geq 1}$ is like a random walk except with time sped up by a factor of n and space shrunk by a factor of $\sigma\sqrt{n}$. The following picture shows the situation for $\sigma^2 = 1$.



The next Theorem is the first and most famous FCLT which was proved 1951 from Donsker.

Theorem 6.2 (Donsker's Theorem).
Let $S^{(n)}_t$ be defined as in (6.16). Then we have

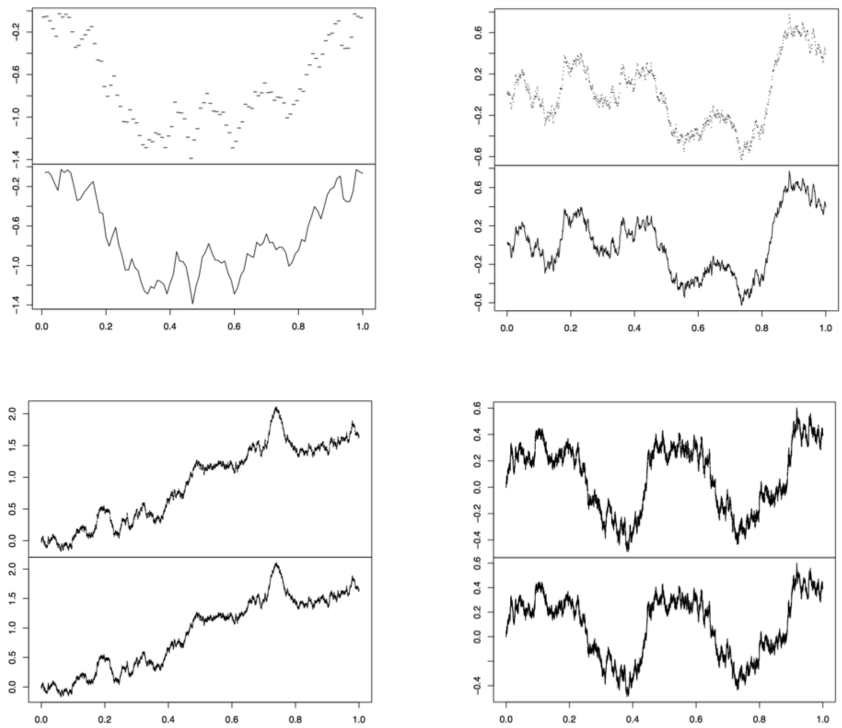
$$\mathbb{S}^{[n]} \Rightarrow \mathbb{W},$$

where $\mathbb{W}$ is a (standard) Brownian motion on $[0, 1]$.

Proof. Cf. [1], Theorem 8.2. □

This result corresponds with our intuition since the paths of a rescaled random walk for large n looks like a paths of a Brownian motion. It is also very often used for proving the existence of Brownian Motion.

However, it is not essential to choose a linear interpolation between the points. Another option is to use constant interpolation by defining stochastic processes $\mathbb{Y}^{[n]}$ via (6.17) making the paths only right-continuous. Although this approach leeds to some additional technical complexities as we have seen in the last chapter it is more useful for certain applications. The following pictures indicates that even though we start with càdlàg-paths we will get a continuous Brownian Motion path as well in the limit. In the pictures we plotted paths of $\mathbb{Y}^{[n]}$ in the upper half and paths of $\mathbb{S}^{[n]}$ in the lower half for $n \in \{100, 1000, 10000, 100000\}$:



We see that for large n they are nearly the same. In fact we did all the work in chapter 5 to extend Donkser's Theorem 6.2 to the case where we start with the càdlàg-function $\left(Y_t^{(n)}\right)_{t\in[0,1]}$. Note therefore that the Wiener measure $\mathbb{W}$ (it is common to denote the measure and the random element by $\mathbb{W}$) defined on $(\mathbb{C}[0,1],\mathcal{B}_{\mathbb{C}})$ can be extended to $(\mathbb{D}[0,1],\mathcal{B}_{\mathbb{D}})$. Since ρ and d_B coincides on $(\mathbb{C}[0,1],\mathcal{B}_{\mathbb{C}})$ we may "embed" $(\mathbb{C}[0,1],\mathcal{B}_{\mathbb{C}})$ in the space $(\mathbb{D}[0,1],\mathcal{B}_{\mathbb{D}})$ and a probability measure on the former space can be extended to the latter with support in $\mathbb{C}[0,1]$. In particular, $\mathbb{W}$ is defined on $(\mathbb{D}[0,1],\mathcal{B}_{\mathbb{D}})$ by simple augmenting the standard conditions on $\mathbb{W}$ by $\mathbb{W}[\mathbb{C}[0,1]] = 1$. For $A \in \mathcal{B}_{\mathbb{D}}$ we have therefore $\mathbb{W}[A] = \mathbb{W}[A \cap \mathbb{C}[0,1]]$.

Theorem 6.3 (Donsker's Theorem in $\mathbb{D}[0,1]$).
Let $Y_t^{(n)}$ be defined as in (6.17). Then we have

$$\mathbb{Y}^{[n]} \Rightarrow \mathbb{W},$$

where $\mathbb{W}$ is a (standard) Brownian motion on $[0,1]$.

Proof. Having done the work in chapter 5 this proof can be in fact done with only minor changes of the proof of Donsker's Theorem. Cf. [1], Theorem 14.1. $\square$

6.2 FCLT for Martingale Difference Arrays

Now we want to go a step further, weakening the assumptions on the random variables X_i's. We need two Lemmas for the proof.

The first Lemma is a very useful tool when we want to establish finite-dimensional convergence. It states that problems which involve random vectors in $\mathbb{R}^d$ can be reduced to problems involving only ordinary random variables in $\mathbb{R}$.

Lemma 6.4 (**Cramér-Wold device**).
Let $(\mathbf{Y_n})_{n\geq 1} = ((Y_{n1}, \dots, Y_{nd}))_{n\geq 1}$ *be a sequence of random vectors in* $\mathbb{R}^d$ *and* $\mathbf{Y} = (Y_1, \dots, Y_d) \in \mathbb{R}^d$*. Then we have*

$$(Y_{n1}, \dots, Y_{nd}) \Rightarrow (Y_1, \dots, Y_d) \quad \Leftrightarrow \quad \sum_{k=1}^{d} a_k Y_{nk} \Rightarrow \sum_{k=1}^{d} a_k Y_k \;\; \forall (a_1, \dots, a_d) \in \mathbb{R}^d.$$

Proof. ($\Rightarrow$) Define $h : \mathbb{R}^d \to \mathbb{R}$ with $h(x_1, \dots, x_d) = a_1 x_1 + \ldots + a_d x_d$. Since h is continuous the result follows from the Continuous Mapping Theorem 2.7.

($\Leftarrow$) By Lévy's Continuity Theorem 3.1 we have for every $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ and $s \in \mathbb{R}$

$$\mathbb{E}\left[\exp\left\{is\sum_{k=1}^{d} a_k Y_{nk}\right\}\right] \to \mathbb{E}\left[\exp\left\{is\sum_{k=1}^{d} a_k Y_k\right\}\right].$$

Taking $s = 1$ we see that

$$\mathbb{E}\left[e^{i\mathbf{aY_n}}\right] \to \mathbb{E}\left[e^{i\mathbf{aY}}\right]$$

Since $\mathbf{a}$ was arbitrary, the result follows from Lévy's Continuity Theorem 3.1. $\square$

The following general Lemma for martingales will be a helpful inequality for proving tightness.

Lemma 6.5. *Let* $(M_k)_{k=0}^N$ *be a martingale with* $M_0 = 0$*. Then we have for* $c > 0$

$$\mathbb{P}\left[\max_{1\leq k\leq N} |M_k| > 2c\right] \leq \mathbb{E}\left[\frac{2}{c}|M_N| \mathbb{1}_{\{|M_N|>c\}}\right]$$

Proof. We denote by $U_{[a,b]}(X)$ the number of times a sequence $(X_i)_{i=1}^N$ crosses from $\leq a$ to $\geq b$ up to a fixed time N. The upcrossing-inequality for (sub-)martingales states that

$$\mathbb{E}\left[U_{[a,b](X)}\right] \leq \frac{\mathbb{E}\left[(X_N - a)^+\right] - \mathbb{E}\left[(X_0 - a)^+\right]}{b - a},$$

where Y^+ denotes $\max\{Y, 0\}$ for a random variable Y. Now we have

$$\begin{aligned}
\mathbb{P}\left[\min_{1\leq k\leq N} M_k < -2c\right] &= \mathbb{P}\left[\min_{1\leq k\leq N} M_k < -2c \text{ and } M_N \geq -c\right] \\
&\quad + \mathbb{P}\left[\min_{1\leq k\leq N} M_k < -2c \text{ and } M_N < -c\right] \\
&\leq \mathbb{P}\left[U_{[-2c,-c]}(M) > 0\right] + \mathbb{P}[M_N < -c] \\
&\leq \mathbb{E}\left[U_{[-2c,-c]}(M)\right] + \mathbb{P}[M_N < -c].
\end{aligned}$$

Therefore it follows that

$$\begin{aligned}
\mathbb{P}\left[\max_{1\leq k\leq N} M_k > 2c\right] &= \mathbb{P}\left[\min_{1\leq k\leq N} -M_k < -2c\right] \\
&\leq \mathbb{E}\left[U_{[-2c,-c]}(-M)\right] + \mathbb{P}[M_N > c]
\end{aligned}$$

and hence

$$\mathbb{P}\left[\max_{1\leq k\leq N} |M_k| > 2c\right] \leq \mathbb{P}[|M_N| > c] + \mathbb{E}\left[U_{[-2c,-c]}(M)\right] + \mathbb{E}\left[U_{[-2c,-c]}(-M)\right] \tag{6.18}$$

Since $(M_i)_{i=1}^N$ and $(-M_i)_{i=1}^N$ are both martingales with first term zero we get by the upcrossing-inequality

$$\begin{aligned}
\mathbb{E}\left[U_{[-2c,-c]}(M)\right] + \mathbb{E}\left[U_{[-2c,-c]}(-M)\right] &\leq \frac{\mathbb{E}\left[(M_N - 2c)^+\right] - 2c}{c} + \frac{\mathbb{E}\left[(-M_N - 2c)^+\right] - 2c}{c} \\
&= \frac{1}{c}\mathbb{E}\left[M_N \mathbb{1}_{\{M_N \geq -2c\}} - M_N \mathbb{1}_{\{M_N \leq 2c\}}\right] \\
&\qquad - 2\left(\mathbb{P}[M_N < -2c] + \mathbb{P}[M_N > 2c]\right) \\
&\leq \frac{1}{c}\mathbb{E}\left[M_N \mathbb{1}_{\{M_N \geq 2c\}} - M_N \mathbb{1}_{\{M_N \leq -2c\}}\right] \\
&= \frac{1}{c}\mathbb{E}\left[|M_N| \mathbb{1}_{\{|M_N| \geq 2c\}}\right] \\
&\leq \frac{1}{c}\mathbb{E}\left[|M_N| \mathbb{1}_{\{|M_N| \geq c\}}\right].
\end{aligned}$$

For the first term in (6.18) we have also

$$\mathbb{P}[|M_N| > c] \leq \frac{1}{c}\mathbb{E}\left[|M_N| \mathbb{1}_{\{|M_N| \geq c\}}\right]$$

and therefore the result follows. □

We now consider an array of random variables $(X_{ni})_{i=1}^{k_n}$ and define

$$\mathbb{W}^{(n)}(\tau) = \sum_{i=1}^{k_n(\tau)} X_{ni},$$

where $k_n(\tau)$ is a sequence of integer-valued, non-decreasing, right-continuous functions defined on $[0, \tau]$, hence just a little generalisation of $\lfloor nt \rfloor$. Theorem 3.20 provides conditions under which

$$\mathbb{W}^{[n]}(\tau) = \sum_{i=1}^{k_n(\tau)} X_{ni} \Rightarrow Z \sim N(0, 1)$$

for any fixed τ. We give now the functional version of this Theorem. Note that we prove the result for the interval $[0,T]$ for arbitrary finite T.

Theorem 6.6 (McLeish, 1974).
Let $(X_{ni})_{i=1}^{k_n}$ be a martingale difference array satisfying for any $t \in [0,T]$

(i) $\max_{1\leq i\leq k_n(t)} |X_{ni}| \xrightarrow{L^2} 0$ *and*

(ii) $\sum_{i=1}^{k_n(t)} X_{ni}^2 \xrightarrow{\mathbb{P}} t$.

Then we have $\mathbb{W}^{[n]} \Rightarrow \mathbb{W}$ on $\mathbb{D}[0,T]$.

Proof. We need to check the two conditions of Theorem 6.1. Therefore, we begin as in the corresponding CLT with a reduction step and establish some properties of the new process:

1. Reduction step
We define by a cutoff a related martingale difference array by

$$Z_{ni} := X_{ni} \cdot \mathbb{1}_{\left\{\sum_{r=1}^{i-1} X_{nr}^2 \leq T+1\right\}}.$$

and set $\widetilde{\mathbb{W}}^{[n]}(t) := \sum_{i=1}^{k_n(t)} Z_{ni}$. As in the proof of Theorem 3.20 we have

$$\mathbb{P}[X_{ni} \neq Z_{ni} \text{ for some } i \leq k_n(t)] \longrightarrow 0 \tag{6.19}$$

and therefore

$$\mathbb{P}[\mathbb{W}^{[n]}(t) \neq \widetilde{\mathbb{W}}^{[n]}(t) \text{ for any } t \in [0,T]] \longrightarrow 0.$$

By Lemma 2.31 it is therefore enough to prove $(\widetilde{\mathbb{W}}^{[n]}(t))_{t\in[0,T]} \Rightarrow (\mathbb{W}_t)_{t\in[0,T]}$.

2. Useful properties of $(\widetilde{\mathbb{W}}^{[n]}(t))_{t\in[0,T]}$:
By condition (i) it follows $\max_{1\leq i\leq k_n(t)} |X_{ni}| \xrightarrow{L^1} 0$ and together with (6.19) we get

$$\max_{1\leq i\leq k_n(t)} |Z_{ni}| \xrightarrow{\mathbb{P}} 0. \tag{6.20}$$

Next, observe that

$$\sum_{i=1}^{k_n(t)} Z_{ni}^2 \leq T + 1 + \max_{1\leq i\leq k_n(T)} X_{ni}^2$$

By condition (1) the right-hand side converges to $T+1$ in L^1, hence is uniformly integrable and therefore $\sum_{i=1}^{k_n(t)} Z_{ni}^2$ is uniformly integrable. By condition (ii) and (6.19) we have $\sum_{i=1}^{k_n(t)} Z_{ni}^2 \xrightarrow{\mathbb{P}} t$. Therefore we can conclude from Theorem 3.11 L^1 convergence of $\sum_{i=1}^{k_n(t)} Z_{ni}^2$, i.e.

$$\mathbb{E}\left[\left|\sum_{i=1}^{k_n(t)} Z_{ni}^2 - t\right|\right] \longrightarrow 0 \tag{6.21}$$

for any $t \in [0,T]$.

3. Convergence of the finite-dimensional distributions of $(\widetilde{\mathbb{W}}^{[n]}(t))_{t\in[0,T]}$:
Let $0 = t_0 < t_1 < \ldots < t_d \leq T$ and $a_1, \ldots, a_d$ be arbitrary constants. For each $i \in (k_n(t_{j-1}), k_n(t_j))$ we set

$$Y_{ni} \coloneqq a_j Z_{ni}$$

with $i = 1, \ldots, d$. Then we get

$$\begin{aligned}\sum_{j=1}^{d} a_j \left(\widetilde{\mathbb{W}}^{[n]}(t_j) - \widetilde{\mathbb{W}}^{[n]}(t_{j-1})\right) &= \sum_{j=1}^{d} a_j \left(\sum_{i=1}^{k_n(t_j)} Z_{ni} - \sum_{i=1}^{k_n(t_{j-1})} Z_{ni}\right) \\ &= \sum_{j=1}^{d} \sum_{i=k_n(t_{j-1})+1}^{k_n(t_j)} a_j Z_{ni} \\ &= \sum_{i=1}^{k_n(t_d)} Y_{ni}.\end{aligned}$$

To get convergence of the finite-dimensional distributions we want to use Theorem 3.20 and Remark 3.21. Therefore we check that $\sum_{i=1}^{k_n(t_d)} Y_{ni}$ satisfy all the assumptions:

(1) and (3): By definition of Y_{ni} and (6.21) we get

$$\sum_{i=1}^{k_n(t_d)} Y_{ni}^2 \xrightarrow{L^1} \sigma^2 = \sum_{j=1}^{d} (a_j)^2 (t_j - t_{j-1})$$

and therefore also convergence in probability. From the L^1-convergence of $\sum_{i=1}^{k_n(t_d)} Y_{ni}^2$ it follows also that $\sup_{n\geq 1} \mathbb{E}\left[\max_{1\leq i\leq k_n(t_d)} Y_{ni}^2\right] < \infty$.

(2): From (6.20) it follows directly that $\max_{1\leq i\leq k_n(t_d)} |Y_{ni}| \xrightarrow{\mathbb{P}} 0$.

Thus, we get

$$\sum_{j=1}^{d} a_j \left(\widetilde{\mathbb{W}}^{[n]}(t_j) - \widetilde{\mathbb{W}}^{[n]}(t_{j-1})\right) \Rightarrow Z \sim N(0, \sigma^2)$$

with $\sigma^2 = \sum_{j=1}^{d} (a_j)^2 (t_j - t_{j-1})$. By the Cramér-Wold-device 6.4 we get the convergence of the finite-dimensional distributions.

4. Tightness of $(\widetilde{\mathbb{W}}^{[n]}(t))_{t\in[0,T]}$:
We want to show the tightness condition of Theorem 6.1. First observe that

$$\mathbb{P}\left[\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq T}} \left|\widetilde{\mathbb{W}}^{[n]}(s) - \widetilde{\mathbb{W}}^{[n]}(t)\right| > \varepsilon\right] = \mathbb{P}\left[\sup_{\substack{|s-t|<\delta \\ 0\leq s,t\leq T}} \left|\sum_{i=k_n(s)+1}^{k_n(t)} Z_{ni}\right| > \varepsilon\right]$$

$$\leq \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \mathbb{P}\left[\sup_{j\delta \leq t \leq (j+1)\delta} \left| \sum_{i=k_n(j\delta)}^{k_n(t)} Z_{ni} \right| > \frac{\varepsilon}{3}\right].$$

Now, since (Z_{ni}) is a martingale difference array the sequence

$$0, Z_{nk_n(j\delta)}, Z_{nk_n(j\delta)} + Z_{nk_n(j\delta)+1}, \ldots, \sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}$$

is a martingale which we identify with the martingale from Lemma 6.5 to obtain

$$\sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \mathbb{P}\left[\sup_{j\delta \leq t \leq (j+1)\delta} \left| \sum_{i=k_n(j\delta)}^{k_n(t)} Z_{ni} \right| > \frac{\varepsilon}{3}\right] \leq \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \mathbb{E}\left[\frac{12}{\varepsilon} \left| \sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni} \right| \mathbb{1}_{\left\{\left|\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}\right| \geq \frac{\varepsilon}{6}\right\}}\right]$$

$$\leq \frac{12}{\varepsilon} \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \left\{ \left\{ \mathbb{E}\left[\left(\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}\right)^2\right]\right\}^{\frac{1}{2}} \left\{ \mathbb{E}\left[\left(\mathbb{1}_{\left\{\left|\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}\right| \geq \frac{\varepsilon}{6}\right\}}\right)^2\right]\right\}^{\frac{1}{2}} \right\}$$

$$= \frac{12}{\varepsilon} \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \left\{ \left\{ \mathbb{E}\left[\left(\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}\right)^2\right]\right\}^{\frac{1}{2}} \left\{ \mathbb{P}\left[\left|\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta} Z_{ni}\right| \geq \frac{\varepsilon}{6}\right]\right\}^{\frac{1}{2}} \right\},$$

where we used the Cauchy-Schwartz inequality. Now we look what happens in the limit with the individual terms in the last expression. Using the fact that martingale difference arrays are uncorrelated we get for the first term:

$$\mathbb{E}\left[\left(\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta)} Z_{ni}\right)^2\right] = \mathbb{E}\left[\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta)} Z_{ni}^2\right] \to (j+1)\delta - j\delta = \delta$$

where we used property (6.21) to obtain the limit. By the arguments of step 3 we have

$$\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta)} Z_{ni} \Rightarrow M \sim N(0, (j+1)\delta - j\delta) = N(0, \delta)$$

and therefore

$$\mathbb{P}\left[\left|\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta)} Z_{ni}\right| \geq \frac{\varepsilon}{6}\right] = 1 - \mathbb{P}\left[\left|\sum_{i=k_n(j\delta)}^{k_n((j+1)\delta)} Z_{ni}\right| < \frac{\varepsilon}{6}\right]$$

$$\longrightarrow 1 - \left[2 \cdot \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}}\right) - 1\right] = 2\left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}}\right)\right],$$

where Φ denotes the cumulative distribution function of the normal distribution with mean 0 and variance δ. It remains to show that

$$\lim_{\delta \to 0} \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \sqrt{\delta} \left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}}\right)\right]^{\frac{1}{2}} = 0.$$

For $\delta \leq 1$ we have

$$\left| \sum_{j=0}^{\lfloor \frac{T}{\delta} \rfloor} \sqrt{\delta} \left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}} \right) \right]^{\frac{1}{2}} \right| \leq (T+1) \frac{1}{\sqrt{\delta}} \left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}} \right) \right]^{\frac{1}{2}}.$$

Therefore it is enough to prove

$$\lim_{\delta \to 0} \frac{1}{\delta} \left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}} \right) \right] = 0.$$

Now, since the denominator δ and the numerator $\left[1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}} \right) \right]$ goes to 0 for $\delta \to 0$ we can use L'Hôpital's rule to obtain

$$\begin{aligned} \lim_{\delta \to 0} \frac{1 - \Phi\left(\frac{\varepsilon}{6\sqrt{\delta}} \right)}{\delta} &= \lim_{\delta \to 0} \Phi'\left(\tfrac{\varepsilon}{6\sqrt{\delta}} \right) \delta^{-\frac{3}{2}} \cdot \frac{\varepsilon}{12} \\ &= \lim_{\delta \to 0} \frac{\varepsilon}{12} \delta^{-\frac{3}{2}} \frac{e^{-\frac{\varepsilon^2}{72\delta}}}{\sqrt{2\pi}} \\ &= \frac{\varepsilon}{12\sqrt{2\pi}} \lim_{x \to \infty} x^{\frac{3}{2}} e^{-\frac{\varepsilon^2 x}{72}} = 0, \end{aligned}$$

which proves the tightness condition of Theorem 6.1. $\square$

Remark 6.7. As we have seen this proof works for every interval $[0, T]$ with $T < \infty$. By [17] we have therefore proved that the convergence also holds on the space $\mathbb{D}[0, \infty)$. To see how the Skorohod-theory form chapter 5 can be extended to the space $\mathbb{D}[0, \infty)$ see [1], chapter 16.

7 Bibliography

[1] P. Billingsley: *Convergence of Probability Measures*, 2nd edition, John Wiley and Sons, Inc., New York, 1999.

[2] B. M. Brown: *Martingale central limit theorems*, The Annals of Math. Statistics, 59-66, 1971.

[3] J. Davidson: *Stochastic Limit Theory. An Introduction for Econometricians*, Oxford University Press, Oxford, 1994.

[4] T. R. Fleming, D. P. Harrington: *Counting Processes and Survival Analysis*, John Wiley and Sons, Inc., New York, 1990.

[5] P. Hall, C. C. Heyde: *Martingale Limit Theory and Its Application*, Academic Press, New York, 1980.

[6] N. Henze: *Lecture Notes: Asymptotic stochastics*, Karlsruher Institut fr Technologie, 2014.

[7] O. Kallenberg: *Foundations of Modern Probability*, 2nd edition, Springer, New York, 2002.

[8] A. Klenke: *Probability Theory. A Comprehensive Course*, 2nd edition, Springer, New York, 2014.

[9] D. L. McLeish: *Dependent central limit theorems and invariance principles*, Annals of Probability 2, 620-628, 1974.

[10] D. Meintrup, S. Schäffler: *Stochastik. Theorie und Anwendungen*, Springer, Berlin/Heidelberg, 2005.

[11] B. U. Park: *Lecture Notes: Theory of Statistics. Weak convergence*, Seoul National University, 2016.

[12] D. Pollard: *Convergence of Stochastic Processes*, Springer, New York, 1984.

[13] J. S. Rosenthal: *A First Look at Rigorous Probability Theory*, 2nd edition, World Scientific Printers, Singapore, 2006.

[14] S. Sagitov: *Weak Convergence of Probability Measures*, Chalmers University of Technology and Gothenburg University, 2015.

[15] S. Sethuraman: *A martingale central limit theorem*, University of Arizona, 2002.

[16] A. Sokol, A. R. Nielsen: *Advanced Probability*, Department of Mathematical Sciences/University of Copenhagen, Copenhagen, 2013.

[17] C. Stone: *Weak convergence of stochastic processes defined on semi-infinite time intervals*, Proc. Math. Soc. 14, 694-696.

[18] M. Taniguchi, Y. Kakizawa: *Asymptotic Theory of Statistical Inference fore Time Series*, Springer, New York, 2000.

Deutsche Zusammenfassung

Zentrale Grenzwertsätze stellen ein wesentliches Fundament der modernen Wahrscheinlichkeitstheorie dar. Einfachste Versionen wurden bereits im 17. Jahrhundert von DeMoivre und Laplace bewiesen. Im Laufe der Zeit bemühte man sich die Vorraussetzungen abzuschwächen und allgemeinere Versionen, die vor allem auch in Anwendungsbereichen benötigt werden, zu beweisen. Im 20. Jahrhundert ging man einen Schritt weiter und fragte nicht nur nach der Verteilung des Endresultats, sondern interessierte sich für den gesamten „Weg“ dorthin. Um derartige Aussagen präzise formulieren zu können, muss man die schwache Konvergenz von Wahrscheinlichkeitsmßen auf allgemeineren Räumen studieren. Der wohl berühmteste Funktionale Grenzwertsatz (auch Invarianzprinzip genannt) stammt von Donsker und wurde 1951 bewiesen. Die vorliegende Arbeit beschäftigt sich mit Verallgemeinerungen von zentralen und funktionalen Grenzwertsätzen auf sogenannte Martingaldifferenzschemata, die auf der Grundlage von [9] basieren.

Das erwähnte Paper von McLeish [9] ist bereits 1974 erschienen, ist jedoch für Personen, die keine „Experten“ im Bereich der modernen Wahrscheinlichkeitstheorie sind, schwer lesbar. Ziel der Arbeit ist es daher, die Thematik einer größeren Öffentlichkeit zugängig zu machen. Vorkenntnisse aus (höherer) Wahrscheinlichkeitstheorie und Maßtheorie sind ausreichend, um diese Arbeit im Detail zu lesen.

Nach einer kurzen Einführung wird im zweiten Kapitel die schwache Konvergenz auf allgemeine metrische Räume übertragen und wesentliche Eigenschaften davon besprochen. Im dritten Kapitel werden die Vorraussetzungen des Zentralen Grenzwertsatzes fortlaufend abgeschwächt und eine sehr allgemeine Version entsprechend Theorem 2.3 in [9] bewiesen. Um entsprechende Funktionale Grenzwertsätze zu behandeln bedarf es an einem soliden Verständnis der zugrundeliegenden Räume. Obwohl das Hauptresultat im Raum $\mathbb{D}[0,1]$ der auf $[0,1]$ definierten rechtsstetigen Funktionen mit existierendem linksseitigen Grenzwert bewiesen wird, besprechen wir im vierten Kapitel zuerst den Raum der stetigen Funktionen $\mathbb{C}[0,1]$. Grund dafür ist, dass alle Konzepte mit Mehraufwand im fünften Kapitel auf den Raum $\mathbb{D}[0,1]$ übertragen werden können. Dabei wird zunächst einiger Aufwand betrieben, um eine passende Metrik auf $\mathbb{D}[0,1]$ zu finden, die den Raum separabel und vollständig macht bevor die stetigen Resultate angewendet werden können. Im sechsten Kapitel wird die Theorie der vorigen Kapitel angewendet. Ausgehend von Donsker's Theorem beweisen wir entsprechend Theorem 3.2. in [9] ein Invarianzprinzip für Martingaldifferenzschemata.