UNIVERSITAT DE BARCELONA

ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

CONVERGENCE TO THE BROWNIAN MOTION

Author:
Marc Cano i Cànovas

Supervisor:
Dr. Carles Rovira Escofet

Facultat de Matemàtiques i Informàtica

June 28, 2023

Abstract

The Brownian motion is a stochastic process that models the motion of particles suspended in a liquid or a gas. In mathematics, it also plays a vital role in stochastic calculus.

This thesis consists in the proving of three different results of convergence towards the Brownian motion.

The first one is proving the Donsker's theorem, for which different notions of convergence, such as weakly convergence or convergence in distribution, are introduced.

The second result consists in the proving of a certain type of stochastic processes converging in distribution towards the Brownian motion.

For the last result, uniform transport processes are presented and then it is showed that they converge almost surely to the Brownian motion. In addition, a couple of results that extend this almost sure convergence are mentioned.

2020 Mathematics Subject Classification. 60B05, 60B10, 60F05, 60G50, 60G51, 60G55, 60H05, 60J10, 60J65

Acknowledgements

Vull agrair a la Noelia tot el suport incondicional.

També agrair al Carles tota la feina feta durant aquests mesos fent aquesta tesi.

Per últim, agrair a la meva família el recolzament que em brinden sempre.

Contents

1 Introduction

I first got interest in stochastic processes when I was an undergraduate student and I took a subject on, precisely, an introduction to stochastic processes.

After doing this subject, I felt like I needed to do some more work related on this area of mathematics so I did my bachelor's thesis on renewal stochastic processes. While studying this processes, I enrolled to the Advanced Mathematics master in order to be able to extend my knowledge in stochastic analysis as my first priority.

Therefore, it made sense to do this master's thesis about some topic related to stochastic calculus.

Then, I talked with my tutor, Carles Rovira, about what to do and one of the topics that we discussed as a possible master's thesis was the one that we finally commit to: proving different results on the convergence to the Brownian motion.

The project

During the 19th century, scientists started developing the discipline of statistical mechanics. Basically, they started treating physical systems mathematically. For example, they regarded containers filled with gas as collections of many moving particles.

In 1859, James Clerk Maxwell presented a work on the kinetic theory of gases where he assumed that the gas particles move in random directions at random velocities.

This was the starting point for the development of the statistical physics during the second half of the 19th century. During this period of time, Thorvald N. Thiele, in 1880, published a paper where he described the mathematics behind the *Brownian motion*.

The Brownian motion describes the random movement of particles suspended in a liquid or a gas. It was first described by the botanist Robert Brown, in 1827, while he was looking at pollen particles through a microscope.

Despite he described it, as we have said previously, the mathematics behind this motion were not addressed until the end of the 19th century. Furthermore, it was not until 1900 when Louis Bachelier modeled for the first time, and under the supervision of Henri Poincaré, the stochastic process that we now know as the Brownian motion or the Wiener process.

This Brownian motion will be the protagonist of this thesis.

What we will do is prove different results regarding some type of convergence towards this Brownian motion.

One of the results that we are going to see is a classical result which is the Donsker's theorem. For this result we will follow the first chapter, and part of the second, of the book *Convergence of Probability Measures* by Patrick Billingsley [2].

In order to state and prove this result we are going to discuss about different notions of convergence such as the weak convergence, the convergence in distribution or the convergence in probability.

We will also define the concept of tightness. Moreover, we will discuss this concept of tightness, and also the notion of weak convergence, in the set of continuous functions on $[0, 1]$.

Finally, we will see the definitions of the Wiener measure and the Brownian motion.

With all this, we will be able to state the Donsker's theorem and prove that the stochastic processes that it defines converge towards the Brownian motion in distribution.

Notice that, for all the previous concepts, we will use the book *Curs de Probabilitats* by David Nualart and Marta Sanz [10] to reinforce these two chapters of the *Convergence of Probability Measures* book.

Another result that we will prove is the convergence in distribution of a particular type of stochastic processes. This processes that we will define were presented by Mark Kac but was Daniel Stroock who explicitly proved their convergence. This result can be found in the work of Stroock, *Topics in Stochastic Differential Equations* [13]. Even so, we will follow a presentation made by Xavier Bardina on December 18, 2014 at Bucuresti called *On the Kac-Stroock Approximations.*

The last result that we will see is the almost sure convergence of the uniform transport processes towards the Brownian motion.

In order to prove this result we will follow the paper written by Richard J. Griego, David Heath and Alberto Ruiz-Moncayo, *Almost Sure Convergence of Uniform Transport Processes to Brownian Motion* [5].

We will use classical books as *An Introduction to Probability Theory and its Applications* by William Feller [3] and *Studies in the Theory of Random Processes* by Anatoliy V. Skorokhod [12] to complement the proof of the main paper [5].

Furthermore, we will see a couple of results that extend the result of almost sure convergence. To do so we will use some results that we can find in the paper *Rate of Convergence of Uniform Transport Processes to Brownian Motion and Application to Stochastic Integrals* by Luis G. Gorostiza and Richard J. Griego [4] and in the paper *On the Convergence of Ordinary Integrals to Stochastic Integrals* by Eugene Wong and Moshe Zakai [14].

Structure

This thesis consists of three parts without taking into account the Introduction and the Appendices.

In the first part we are going to prove the Donsker's theorem. We will discuss about the different notions of convergence that we have said before and we will also see concepts as tightness, the Wiener measure or the Brownian motion.

The second and the third part are both in the same section, under the name of *Uniform Transport Processes*.

What we are going to do in these sections is to prove two different results of different types of convergence towards the Brownian motion.

In the second part we are proving the convergence in distribution of a certain type of stochastic processes that were discussed by Kac and Stroock. We will basically study the construction of the processes they studied and then we will prove that they converge in distribution to the Brownian motion.

In the third part we are going to prove a stronger result which involves a generalization of the processes studied in the second part. We will prove that they converge almost surely towards the Brownian motion.

To finish, we will see a couple of results that are an extension of this last notion of almost sure convergence to the Brownian motion.

At the end of the thesis we will find a section of Appendices that contain different sections which complement the topics that we study during this work. Their purpose will be to fill the gaps that we may leave during the thesis.

2 The Donsker's Theorem

We start the thesis with this first part where the objective is to state and prove the Donsker's theorem. In order to achieve this goal, we will first work on some preliminaries to be able to understand everything involved in the theorem.

For the whole section we will denote the set of continuous functions on $[0,1]$ by $\mathcal{C} := \mathcal{C}([0,1])$. Also, we will define the following distance:

$$\rho(x,y) := \sup_t |x(t) - y(t)| . \tag{2.1}$$

2.1 Weak Convergence

The first thing that we want to study is the notion of weak convergence in probability. We extract the definition from the book of David Nualart and Marta Sanz, *Curs de Probabilitats* [10].

Definition 2.1. *Let $\{\mathbb{P}_n\}_{n\geq 1}$ be a sequence of probabilities in $\mathbb{R}$. We will say that this sequence* ***converges weakly*** *to a probability $\mathbb{P}$, $\mathbb{P}_n \Rightarrow \mathbb{P}$, if*

$$\lim_n \int_{\mathbb{R}} f \; d\mathbb{P}_n = \int_{\mathbb{R}} f \; d\mathbb{P},$$

for every function $f : \mathbb{R} \to \mathbb{R}$ continuous and bounded.

Instead of studying the convergence with probabilities in $\mathbb{R}$, we want to see how can we translate this definition but considering a metric space and considering the probabilities on the class of Borel sets in S.

Let S be a metric space and $\mathscr{P}$ be the smallest σ-field containing all the open sets. We will say that $\mathbb{P}$ is a probability measure on $\mathscr{P}$ if it is a non-negative, countably additive set function[2] with $\mathbb{P}(S) = 1$.

Now, we can define the notion of weak convergence.

Definition 2.2. *Let $\{\mathbb{P}_n\}_{n\geq 1}$ be a sequence of probabilities in $\mathscr{P}$ and let $\mathbb{P}$ also be a probability measure in $\mathscr{P}$. We will say that this sequence* ***converges weakly*** *to $\mathbb{P}$, i.e. $\mathbb{P}_n \Rightarrow \mathbb{P}$, if*

$$\lim_n \int_S f \; d\mathbb{P}_n = \int_S f \; d\mathbb{P},$$

for f a real, continuous and bounded function on S.

The following theorem provides us equivalent ways to state the notion of weak convergence. To state the theorem we first need to define the notion of $\mathbb{P}$-continuity sets.

[2] A **set function** is a function mapping sets to numbers where the value for the union of two disjoint sets is the sum of its values on these sets, i.e. $f(A \cup B) = f(A) + f(B)$. This set function f is **countably additive** if for any given collection of sets $\{E_k\}_{k=1}^n$ where f is defined, then $f\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} f(E_k)$.

Definition 2.3. *We will say that* A *in* $\mathscr{P}$ *is a* $\mathbb{P}$***-continuity set*** *if its boundary,* ∂A*, satisfies* $\mathbb{P}(\partial A) = 0$.

With that being said, we can state the following theorem.

Theorem 2.4. *Let* $\mathbb{P}_n$*,* $\mathbb{P}$ *be probability measures on* $(S, \mathscr{P})$*. Then, the following conditions are equivalent:*

(i) $\mathbb{P}_n \Rightarrow \mathbb{P}$.

(ii) $\lim_n \int f \, d\mathbb{P}_n = \int d\mathbb{P}$ *for every* f *real, bounded and uniformly continuous.*

(iii) $\limsup_n \mathbb{P}_n(F) \leq \mathbb{P}(F)$ *for all closed* F.

(iv) $\liminf_n \mathbb{P}_n(G) \geq \mathbb{P}(G)$ *for all open* G.

(v) $\lim_n \mathbb{P}_n(A) = \mathbb{P}(A)$ *for all* $\mathbb{P}$*-continuity sets* A.

Proof. Notice first that $(i) \implies (ii)$ is trivial because of the definition that we have given for weakly convergence.

Let us prove now $(ii) \implies (iii)$.

We assume (ii) and suppose that we have F closed.

For every δ we can find an $\varepsilon > 0$ small enough such that $G := \{x \colon \rho(x, F) < \varepsilon\}$, where ρ is the distance we have previously defined, satisfies $\mathbb{P}(G) < \mathbb{P}(F) + \delta$, since the sets G decrease to F as $\varepsilon \downarrow 0$.

Let us define now the function $f(x) = \varphi\left(\frac{1}{\varepsilon}\rho(x, F)\right)$ where $\varphi(t)$ is the continuous real function defined as it follows:

$$\varphi(t) := \begin{cases} 1 & \text{if} \quad t \leq 0, \\ 1 - t & \text{if} \quad 0 \leq t \leq 1, \\ 0 & \text{if} \quad 1 \leq t. \end{cases}$$

Then, f satisfies:

- f is uniformly continuous on S,
- $f(x) = 1$ if $x \in F$,
- $f(x) = 0$ if $x \in G^c$,
- $0 \leq f(x) \leq 1, \quad \forall x$.

Notice that we have the following relations:

$$\mathbb{P}_n(F) = \int_F f \, d\mathbb{P}_n \leq \int f \, d\mathbb{P}_n,$$

and

$$\int f \, d\mathbb{P} = \int_G f \, d\mathbb{P} \leq \mathbb{P}(G) \leq \mathbb{P}(F) + \delta.$$

Then, using these relations and combining them with the condition (ii), which we are assuming, we obtain that

$$\limsup_n \mathbb{P}_n(F) \leq \lim_n \int f \ d\mathbb{P}_n = \int f \ d\mathbb{P} < \mathbb{P}(F) + \delta.$$

Hence, since δ is arbitrary, we have proved (iii).

To continue, we are going to prove $(iii) \implies (i)$.

Let us assume (iii) and that f is continuous and bounded on S.

We want to prove first that

$$\limsup_n \int f \ d\mathbb{P}_n \leq \int f \ d\mathbb{P}. \tag{2.2}$$

If we linearly transform f with a positive coefficient for the first-degree term, i.e. $\bar{f} = af + b$ with $a > 0$, then we can reduce to the case where $0 < \bar{f}(x) < 1$ for all x by choosing proper coefficients. We will abuse notation and use f as $\bar{f}$.

Now, let us fix $k \in \mathbb{N}$ and define $F_i \coloneqq \left\{x \colon \frac{i}{k} \leq f(x)\right\}$ for $i = 0, \dots, k$.

Since $0 < f(x) < 1$,

$$\sum_{i=1}^{k} \frac{i-1}{k} \mathbb{P}\left[\left\{x : \frac{i-1}{k} \leq f(x) \leq \frac{i}{k}\right\}\right] \leq \int f \ d\mathbb{P} < \sum_{i=1}^{k} \frac{i}{k} \mathbb{P}\left[\left\{x : \frac{i-1}{k} \leq f(x) \leq \frac{i}{k}\right\}\right].$$

Notice that those sums, in fact, are

$$\sum_{i=1}^{k} \frac{i-1}{k} \mathbb{P}\left[\left\{x : \frac{i-1}{k} \leq f(x) \leq \frac{i}{k}\right\}\right] = \sum_{i=1}^{k} \frac{i-1}{k} \left[\mathbb{P}(F_{i-1} - F_i)\right] = \frac{1}{k} \sum_{i=1}^{k} P(F_i),$$

and

$$\sum_{i=1}^{k} \frac{i}{k} \mathbb{P}\left[\left\{x : \frac{i-1}{k} \leq f(x) \leq \frac{i}{k}\right\}\right] = \sum_{i=1}^{k} \frac{i}{k} \left[\mathbb{P}(F_{i-1} - F_i)\right] = \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(F_i).$$

Hence, the previous inequalities become

$$\frac{1}{k} \sum_{i=1}^{k} P(F_i) \leq \int f \ d\mathbb{P} < \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k} \mathbb{P}(F_i). \tag{2.3}$$

Also, since (iii) holds, we have that $\limsup_n \mathbb{P}_n(F_i) \leq \mathbb{P}(F_i)$ for every i. Hence, applying (2.3) we obtain that

$$\limsup_n \int f \ d\mathbb{P}_n \leq \frac{1}{k} + \int f \ d\mathbb{P}.$$

Letting $k \to \infty$ we obtain (2.2).

If we apply (2.2) to $-f$ we can get $\liminf_n \int f \, d\mathbb{P} \geq \int f \, d\mathbb{P}$. This, jointly with (2.2), proves the weak convergence.

The prove of $(iii) \iff (iv)$ can be done by complementation.

For example, let us assume (iii). Notice that, for each F closed, there exists a $G = S \backslash F$ open and, therefore, $\mathbb{P}(F) = 1 - \mathbb{P}(G)$ and $\mathbb{P}_n(F) = 1 - \mathbb{P}_n(G)$.

Hence, we have that

$$\limsup_n \mathbb{P}_n(F) \leq 1 - \mathbb{P}(G).$$

Moreover,

$$\limsup_n \mathbb{P}_n(F) = \limsup_n \left(1 - \mathbb{P}_n(G)\right) \geq \liminf_n \left(1 - \mathbb{P}_n(G)\right).$$

Therefore,

$$\begin{aligned} \liminf_n \left(1 - \mathbb{P}_n(G)\right) \leq 1 - \mathbb{P}(G) \quad &\iff \quad 1 - \liminf_n \mathbb{P}_n(G) \leq 1 - \mathbb{P}(G) \\ &\iff \quad \mathbb{P}(G) \leq \liminf_n \mathbb{P}_n(G). \end{aligned}$$

With a similar argument, we can prove the other implication.

Now, it just remains to prove $(iii) \iff (v)$.

We first assume (iii) (and hence, (iv)) . Let $A \in S$ be a $\mathbb{P}$-continuity set and let $\mathring{A}$ be its interior and $\bar{A}$ its closure.

Then, for each A, we have the following inequalities:

$$\mathbb{P}(\bar{A}) \geq \limsup_n \mathbb{P}_n(\bar{A}) \geq \limsup_n \mathbb{P}_n(A) \geq \liminf_n \mathbb{P}_n(A) \geq \liminf_n \mathbb{P}_n(\mathring{A}) \geq \mathbb{P}(\mathring{A}).$$

Since we are assuming that A is a $\mathbb{P}$-continuity set, we have that $\mathbb{P}(\partial A) = 0$, where ∂A is the boundary of A. Then, $\mathbb{P}(\bar{A}) = \mathbb{P}(\mathring{A})$ and, therefore,

$$\lim_n \mathbb{P}_n(A) = \mathbb{P}(A).$$

Finally, let us assume (v).

We have that

$$\partial \left\{x \colon \rho(x, F) \leq \delta\right\} \subseteq \left\{x \colon \rho(x, F) = \delta\right\}.$$

Then, these boundaries are disjoint for distinct δ. This means that at most countably many of them can have positive $\mathbb{P}$-measure.

Hence, for some sequence $\{\delta_k\}_{k \geq 0}$ such that $\delta_k \to 0$, the sets $F_k \coloneqq \{x \colon \rho(x, F) \leq \delta_k\}$ are $\mathbb{P}$-continuity sets.

Since (v) holds by assumption,

$$\limsup_n \mathbb{P}_n(F) \leq \lim_n \mathbb{P}_n(F_k) = \mathbb{P}(F_k) \text{ for each } k.$$

Then, if F is closed, $F_k \downarrow F$ and hence, (iii) holds.

□

With all that being said, we are going to define a notion which is useful when we are dealing with weak convergence. This is the notion of **tightness**.

Definition 2.5. *We will say that a probability measure* $\mathbb{P}$ *in* $(S, \mathscr{P})$ *is* ***tight*** *if for each* $\varepsilon > 0$ *there exists a compact set* K *such that* $\mathbb{P}(K) > 1 - \varepsilon$*.*

Another way to characterize tightness is using the following theorem.

Theorem 2.6. *If* S *is separable*[3] *and complete*[4]*, then each probability measure on* $(S, \mathscr{P})$ *is tight.*

In order to prove this result, we need the following theorem, whose proof is in the **Appendices** (**Appendix B**).

Theorem 2.7. *For an arbitrary set* A *in* S *these three conditions are equivalent:*

(i) $\bar{A}$ *is compact.*

(ii) Each sequence in A *has a convergent subsequence, the limit of which necessarily lies in* $\bar{A}$*.*

(iii) A *is totally bounded*[5] *and* $\bar{A}$ *is complete.*

Proof of Theorem 2.6. Since S is separable, for each n, there exists a sequence of $\frac{1}{n}$-spheres covering S, let us say $A_{n_1}, A_{n_2}, \ldots$.

Let us choose an $i_n = i(n)$ so that

$$\mathbb{P}\left(\bigcup_{i \leq i_n} A_{n_i}\right) > 1 - \frac{\varepsilon}{2^n}.$$

Using Theorem 2.7, since we are assuming that S is complete, by this technical result we have that the totally bounded set $\bigcap_{n \geq 1} \bigcup_{i \leq i_n} A_{n_i}$ has compact closure K. Therefore, $\mathbb{P}(K) > 1 - \varepsilon$, which proves the tightness.

□

2.2 Weak Convergence and Tightness in $\mathcal{C}$

Since we will work with functions in $\mathcal{C}$, we want to see some results related to the notions of weak convergence and tightness on $\mathcal{C}$.

The first thing that we are going to see is a theorem which gives us the notion of weak convergence in $\mathcal{C}$ using the weak convergence that we have already studied.

[3] S is separable if it contains a countable, dense subset. That is that there exists a sequence $\{x_n\}_{n \geq 1}$ of elements of S such that every nonempty open subset contains at least one element of the sequence.

[4] A S set is complete if every Cauchy sequence of points in S has its limit also in S.

[5] A set A in a metric space S is **totally bounded** if there is a finite collection of open spheres with radius ε such that its centers are in A and the union of these balls contains A.

Theorem 2.8. *Let* $\mathbb{P}_n$*,* $\mathbb{P}$ *be probability measures on* $(\mathcal{C}, \mathscr{C})$*. If the finite-dimensional distributions of* $\mathbb{P}_n$ *converge weakly to those of* $\mathbb{P}$*, and if* $\{\mathbb{P}_n\}$ *is tight, then* $\mathbb{P}_n \Rightarrow \mathbb{P}$*.*

Remark 2.9. *Notice that* $\mathscr{C}$ *is the class of Borel sets of* $\mathcal{C}$*.*

The proof of this theorem follows from **Prokhorov's Theorem**, which says that the compactness of a family of probability measures on $(\mathcal{C}, \mathscr{C})$ is equivalent to the tightness of the family (see **Appendix D**).

Now, let us see the notion of tightness in $\mathcal{C}$. First we define the **modulus of continuity** of an element x of $\mathcal{C}$ as

$$w_x(\delta) = w(x, \delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1. \tag{2.4}$$

For the following theorems, let $\{\mathbb{P}_n\}$ be a sequence of probability measures on $(\mathcal{C}, \mathscr{C})$.

Theorem 2.10. *The sequence* $\{\mathbb{P}_n\}$ *is tight if and only if the following two conditions hold:*

(i) For each positive η *there exists an* a *such that*

$$\mathbb{P}_n [\{x \colon |x(0)| > a\}] \leq \eta, \quad \textit{for all } n \geq 1. \tag{2.5}$$

(ii) For each $\varepsilon, \eta > 0$ *there exist* $0 < \delta < 1$ *and* $n_0 \in \mathbb{N}$ *such that*

$$\mathbb{P}_n [\{x \colon w_x(\delta) \geq \varepsilon\}] \leq \eta, \quad \textit{for all } n \geq n_0. \tag{2.6}$$

Remark 2.11. *Notice that, since* $w(\cdot, \delta)$ *is continuous then it is measurable . Hence, we can rewrite (2.6) as*

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}_n [\{x \colon w_x(\delta) \geq \varepsilon\}] = 0$$

Proof. We first suppose that $\{\mathbb{P}_n\}$ is tight. With ε and η fixed we choose K a compact set such that $\mathbb{P}_n(K) > 1 - \eta$ for all n. By the Arzelà-Ascoli theorem[6] we have that $K \subset \{x \colon |x(0)| \leq a\}$ for a large enough and $K \subset \{x \colon w_x(\delta) < \varepsilon\}$ for δ small enough. Therefore, we have (i) and (ii) (setting $n_0 = 1$ for (ii)).

Notice that, since a single probability measure $\mathbb{P}$ in $(\mathcal{C}, \mathscr{C})$ is tight, it follows by the necessity of (ii) that for each ε, η there is a δ such that $\mathbb{P}_n [\{x \colon w_x(\delta) \geq \varepsilon\}] \leq \eta$. Hence, if $\{\mathbb{P}_n\}$ satisfies (ii), we may ensure that (2.6) holds for the finitely many n preceding n_0 by decreasing δ if necessary. Thus, we can assume, for sufficiency, that n_0 is always 1.

For the other implication, let us assume that $\{\mathbb{P}_n\}$ satisfies (i) and (ii), with $n_0 = 1$.

[6] **Arzelà-Ascoli Thm:** A subset $A \subset \mathcal{C}$ has compact closure iff $\sup_{x \in A} |x(0)| < \infty$ and $\lim_{\delta \to 0} \sup_{x \in A} w_x(\delta) = 0$.

Then, given η, let us choose a such that if $A = \{x \colon |x(0)| \leq a\}$ then $\mathbb{P}_n(A) \geq 1 - \frac{1}{2}\eta$ for all n. Also, choose δ_k such that if $A_k = \left\{x \colon w_x(\delta_k) < \frac{1}{k}\right\}$ then $\mathbb{P}_n(A_k) \geq 1 - \frac{\eta}{2^{k+1}}$ for all n.

If K is the closure of $A = \bigcap_{k=1}^{\infty} A_k$, then $\mathbb{P}_n(K) \geq 1 - \eta$ for all n and K is compact (by Arzelà-Ascoli theorem). Therefore, $\{\mathbb{P}_n\}$ is tight.

□

The next theorem goes a step beyond this previous theorem to characterize the tightness in $\mathcal{C}$.

Theorem 2.12. *The sequence $\{\mathbb{P}_n\}$ is tight if these two conditions are satisfied:*

(i) For each $\eta > 0$ there exists an a such that

$$\mathbb{P}_n\left[\{x \colon |x(0)| > a\}\right] \leq \eta, \quad \text{for all } n \geq 1. \tag{2.7}$$

(ii) For each $\varepsilon, \eta > 0$ there exist $0 < \delta < 1$ and $n_0 \in \mathbb{N}$ such that

$$\frac{1}{\delta}\mathbb{P}_n\left[\left\{x \colon \sup_{t \leq s \leq t+\delta} |x(s) - x(t)| \geq \varepsilon\right\}\right] \leq \eta, \quad \text{for all } n \geq n_0, \tag{2.8}$$

for all t.

Remark 2.13. *On (ii) we restrict t to $0 \leq t \leq 1$. Also, if $t > 1 - \delta$, we restrict s to $t \leq s \leq 1$.*

Note also that (2.8) is formally satisfied if $\delta > \frac{1}{\eta}$ but we require $\delta < 1$.

Proof. Let us fix δ and let

$$A_t := \left\{x \colon \sup_{t \leq s \leq t+\delta} |x(s) - x(t)| \geq \varepsilon\right\}.$$

Now, let s and t lie in intervals of the form $[i\delta, (i+1)\delta]$.

If $|s - t| < \delta$, then these intervals either coincide or abut. Hence, it follows that

$$\mathbb{P}_n\left[\{x \colon w_x(\delta) \geq 3\varepsilon\}\right] \leq \mathbb{P}_n\left[\bigcup_{i < \frac{1}{\delta}} A_{i\delta}\right].$$

Now, since

$$\mathbb{P}_n\left[\bigcup_{i < \frac{1}{\delta}} A_{i\delta}\right] \leq \sum_{i < \frac{1}{\delta}} \mathbb{P}_n(A_{i\delta}),$$

we have that (2.8) implies that $\mathbb{P}_n\left[x \colon w_x(\delta) \geq 3\varepsilon\right] \leq (1 + \left[\frac{1}{\delta}\right])\delta\eta < 2\eta$.

Therefore, (ii) implies the condition (ii) on Theorem 2.10. Moreover, since (i) is the same as in Theorem 2.10, we have the double implication proved.

□

2.3 Convergence in Distribution

Now that we have seen the definition of weak convergence, we are able to define the notion of convergence in distribution.

The first thing we are going to do is to describe what are the random elements in a probability space and what are the probability distributions.

From now on, on this section, we will consider $(\Omega, \mathscr{B}, \mathcal{P})$ a probability space and S a metric space.

Definition 2.14. *Let X be a mapping from $(\Omega, \mathscr{B}, \mathcal{P})$ to S. If X is measurable then we say that it is a* ***random element****.*

Remark 2.15. *We say that X is defined on its domain Ω (or the probability space $(\Omega, \mathscr{B}, \mathcal{P})$) and that is defined in its range S. We can also call X a random element of S.*

For example, if $S = \mathbb{R}$ we will say that X is a **random variable**; but if $S = \mathbb{R}^k$, then X is a **random vector**; and lastly, if $S = \mathcal{C}$, then we will say that X is a **random function**. In fact, we see these functions depending on time and we call them **stochastic processes**. Those random functions are the ones that we will work with when proving the Donsker's theorem.

Before defining the notion of convergence in distribution we need to define what a distribution is.

Definition 2.16. *The* ***distribution*** *of X is the probability measure $\mathbb{P} = \mathcal{P}X^{-1}$ on $(S, \mathscr{P})$. That is:*

$$\mathbb{P}(A) = \mathcal{P}\left[X^{-1}A\right] = \mathcal{P}\left[\{\omega\colon X(\omega) \in A\}\right] = \mathcal{P}\left[\{X \in A\}\right], \quad A \in \mathscr{P}.$$

Note that $\mathcal{P}$ is a probability measure on a space of an arbitrary nature and, on the other hand, $\mathbb{P}$ is always defined on a metric space.

The distribution $\mathbb{P}$ contains all relevant information about the random element X. If h is a measurable function on S, then,

$$\int h(X)\ d\mathcal{P} = \int h\ d\mathbb{P},$$

in the sense that both integrals exist, or not, together and have the same value. Then,

$$\mathbb{E}\left(h(X)\right) = \int h\ d\mathbb{P}.$$

Notice also that each probability measure on each metric space is the distribution of some random element on some probability space. Therefore, given $\mathbb{P}$ on $(S, \mathscr{P})$, if we take $(\Omega, \mathscr{B}, \mathcal{P}) = (S, \mathscr{P}, \mathbb{P})$ and X as the identity, i.e. $X(\omega) = \omega$ for $\omega \in \Omega = S$, then X is a random element on Ω with values in S and distribution $\mathbb{P}$. Thus, the class of distributions coincides with the class of probability measures on metric spaces but we will call a measure on a metric space a distribution only when it is the distribution of some random element already under discussion.

Definition 2.17. *Let $\{X_n\}_{n\geq 0}$ be a sequence of random elements. We say that this sequence* ***converges in distribution*** *to the random element X,*

$$X_n \xrightarrow{\mathscr{D}} X,$$

if the distributions of the X_n's, say $\mathbb{P}_n$, converge weakly to the distribution of X, say $\mathbb{P}$, i.e. $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Notice that this definition only makes sense if the image space S and its topology are the same for all the random elements but the underlying probability spaces may be distinct.

We usually do not mention this underlying spaces because their structures enter into the argument only by way of the distributions on S they induce. Therefore, we write $\mathcal{P}[X_n \in A]$ when we should write $\mathcal{P}_n[X_n \in A]$ and we write $\mathbb{E}(f(X_n))$ where we should write $\int f(X_n)d\mathcal{P}_n$ or $\mathbb{E}_n(f(X_n))$.

Hence, since $\int_S f(x)\mathbb{P}(dx) = \int_\Omega f(X)d\mathcal{P}$, and similarly for $\int f(X_n)d\mathcal{P}_n$, we have that:

$$X_n \xrightarrow{\mathscr{D}} X \iff \lim_n \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)),$$

for every f real, continuous and bounded on S.

As we have seen for weak convergence, on Theorem 2.4, we can also find different ways to state the convergence in distribution. To do so, we just have to define the equivalent notion of $\mathbb{P}$-continuity sets but for distributions.

Definition 2.18. *Let $A \in \mathscr{P}$. We will say that A is an* **X*-continuity set*** *if*

$$\mathbb{P}[\{X \in \partial A\}] = 0.$$

Thus, we can state the following theorem.

Theorem 2.19. *Let $\{X_n\}$ be a sequence of random elements and let X be a random element too. Then, the following conditions are equivalent:*

(i) $X_n \xrightarrow{\mathscr{D}} X$.

(ii) $\lim_n \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ *for all f real, bounded and uniformly continuous function.*

(iii) $\limsup_n \mathcal{P}[\{X_n \in F\}] \leq \mathcal{P}[\{X \in F\}]$ *for all F closed.*

(iv) $\liminf_n \mathcal{P}[\{X_n \in G\}] \geq \mathcal{P}[\{X \in G\}]$ *for all G open.*

(v) $\lim_n \mathcal{P}[\{X_n \in A\}] = \mathcal{P}[\{X \in A\}]$ *for all X-continuity sets A.*

All these equivalences follow by Theorem 2.4.

To close this section of Convergence in Distribution, let us see an hybrid way to state this convergence.

If X_n are random elements of S with corresponding distributions $\mathbb{P}_n$ and if $\mathbb{P}$ is a probability measure $(S, \mathscr{P})$, we say that X_n **converges in distribution** to $\mathbb{P}$, i.e.,

$$X_n \xrightarrow{\mathscr{D}} \mathbb{P},$$

in case that $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Notice now that, if $h : S \to S'$ is a measurable mapping between two metric spaces, then each probability measure $\mathbb{P}$ on $(S, \mathscr{P})$ induces a unique probability measure on $(S', \mathscr{P}')$, $\mathbb{P}h^{-1}$, defined by $\mathbb{P}h^{-1}(A) = \mathbb{P}(h^{-1}(A))$ for $A \in \mathscr{P}'$.

If we assume that h is continuous, then we have that $f(h(x))$ is continuous and bounded on S whenever $f(y)$ is continuous and bounded on S'. Therefore, $\mathbb{P}_n \Rightarrow \mathbb{P}$ implies that

$$\int f(h(x))\mathbb{P}_n(dx) \longrightarrow \int f(h(x))\mathbb{P}(dx)$$

which becomes

$$\int f(y)\mathbb{P}_n h^{-1}(dy) \longrightarrow \int f(y)\mathbb{P}h^{-1}(dy),$$

and, therefore, $\mathbb{P}_n h^{-1} \Rightarrow \mathbb{P}h^{-1}$.

We can even weaken the continuity assumption to obtain a similar result.

If we assume only that h is measurable and let D_h be the set of discontinuities of h ($D_h \in \mathscr{P}$) then we have this theorem.

Theorem 2.20. *If $\mathbb{P}_n \Rightarrow \mathbb{P}$ and $\mathbb{P}(D_h) = 0$, then $\mathbb{P}_n h^{-1} \Rightarrow \mathbb{P}h^{-1}$.*

Proof. We will show that if F is a closed subset of S', then

$$\limsup_{n\to\infty} \mathbb{P}_n h^{-1}(F) \leq \mathbb{P}h^{-1}(F).$$

Since $\mathbb{P}_n \Rightarrow \mathbb{P}$, we have that

$$\limsup_n \mathbb{P}_n(h^{-1}F) \leq \limsup_n \mathbb{P}_n \left(\overline{h^{-1}F}\right) \leq \mathbb{P}\left(\overline{h^{-1}F}\right).$$

Then, it suffices to prove $\mathbb{P}\left(\overline{h^{-1}F}\right) = \mathbb{P}(h^{-1}F)$. But this follows from the assumption $\mathbb{P}(D_h) = 0$ and the fact that $\overline{h^{-1}F} \subset D_h \cup h^{-1}F$.

□

An immediate corollary of this theorem is:

Corollary 2.21. *If $X_n \xrightarrow{\mathscr{D}} X$ and $\mathcal{P}\left[\{X \in D_h\}\right] = 0$, then $h(X_n) \xrightarrow{\mathscr{D}} h(X)$.*

2.4 Convergence in Probability

In this section we want to introduce briefly the notion of convergence in probability and see a result that will let us prove the Donsker's Theorem.

First, recall that we define the distance

$$\rho(x, y) = \sup_t |x(t) - y(t)| .$$

We say that the random elements X_n **converge in probability** to $a \in S$ if for each $\varepsilon > 0$,

$$\mathcal{P}[\{\rho(X_n, a) \geq \varepsilon\}] \to 0.$$

We will write $X_n \xrightarrow{\mathcal{P}} a$.

Note that if a is a constant-valued random element, then $X_n \xrightarrow{\mathcal{P}} a$ is equivalent to say that $X_n \xrightarrow{\mathscr{D}} a$.

Also, without assumptions on a, we can say that $X_n \xrightarrow{\mathcal{P}} a$ if and only if X_n converges weakly to the probability measure corresponding to a mass of 1 at the point a.

Now, if X_n and Y_n have a common domain, we can consider the distance between them, $\rho(X_n, Y_n)$, which has value $\rho(X_n(\omega), Y_n(\omega))$ at ω. If S is separable, then $\rho(X_n, Y_n)$ is a random variable.

For the following result we assume that X_n and Y_n have a common domain for every n and that S is separable.

Theorem 2.22. *If $X_n \xrightarrow{\mathscr{D}} X$ and $\rho(X_n, Y_n) \xrightarrow{\mathcal{P}} 0$, then $Y_n \xrightarrow{\mathscr{D}} X$.*

Proof. Let F be a set and $F_\varepsilon := \{x \colon \rho(x, F) \leq \varepsilon\}$.

Then,

$$\mathcal{P}[\{Y_n \in F\}] \leq \mathcal{P}[\{\rho(X_n, Y_n) \geq \varepsilon\}] + \mathcal{P}[\{X_n \in F_\varepsilon\}] .$$

Since F_ε is closed, the hypothesis of the theorem implies that

$$\limsup_n \mathcal{P}[\{Y_n \in F\}] = \limsup_n \mathcal{P}[\{X_n \in F_\varepsilon\}] \leq \mathcal{P}[\{X \in F_\varepsilon\}] .$$

If F is closed, $F \downarrow F_\varepsilon$ as $\varepsilon \downarrow 0$ and therefore, using Theorem 2.4 we obtain the result we wanted.

□

2.5 The Donsker's Theorem

In this section we will, finally, state and prove the Donsker's theorem. Before we do so, we are going to discuss a couple of things that are involved in the statement and proof of the theorem.

The first thing that we want to discuss involves a particular type of stochastic processes. These are random functions defined as it follows.

Let $\xi_1, \xi_2, \ldots$ be random variables on the probability space $(\Omega, \mathscr{B}, \mathcal{P})$. For the moment we do not require them to be independent but later, on the theorem, we will see which characteristics we need them to satisfy.

To continue, we define the sequence of partial sums $S_n = \xi_1 + \cdots + \xi_n$ with $S_0 = 0$. Then, we construct X_n as it follows.

Considering the points $\frac{i}{n} \in [0, 1]$, for $i = 1, \ldots, n$, the values of X_n will be

$$X_n\left(\frac{i}{n}, \omega\right) := \frac{1}{\sigma\sqrt{n}} S_i(\omega). \tag{2.9}$$

On the other hand, for the remaining points $t \in [0, 1]$, we define $X_n(t, \omega)$ by linear interpolation:

If $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$, then

$$\begin{aligned} X_n(t, \omega) &= \frac{\frac{i}{n} - t}{\frac{1}{n}} X_n\left(\frac{i-1}{n}\right) + \frac{t - \frac{i-1}{n}}{\frac{1}{n}} X_n\left(\frac{i}{n}\right) \\ &= \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega) + n\left(t - \frac{i-1}{n}\right) \frac{1}{\sigma\sqrt{n}} \xi_i(\omega). \end{aligned} \tag{2.10}$$

Furthermore, since $i - 1 = [nt]$, if $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$ we can define X_n as

$$X_n(t, \omega) := \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega). \tag{2.11}$$

With this definition we obtain a function defined on $[0, 1]$ which at the points $\frac{1}{n}, \ldots, \frac{n-1}{n}, 1$ has the values defined at (2.9) and that it is linear on each interval $\left[\frac{i-1}{n}, \frac{i}{n}\right]$ for $i = 1, \ldots, n$. An example of such processes is the one we have in the next figure.

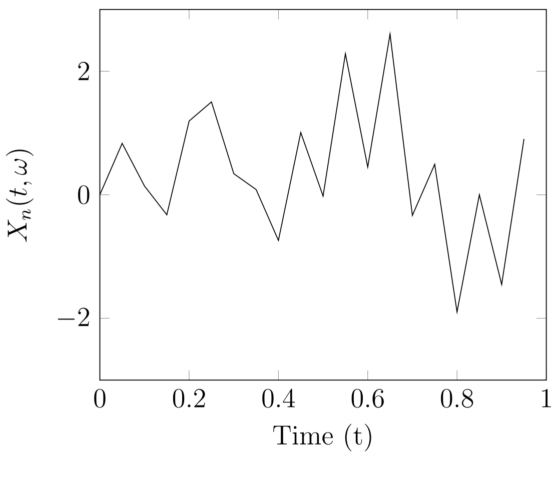

Figure 1: Representation of $X_n(t, \omega)$
with $\xi_i \sim N(0, \sigma^2)$, $\sigma = \sqrt{2}$ and $n = 20$.

Notice that $X_n(t)$ is a random variable for each t since the ξ_i's (and therefore the S_i's) are random variables. Thus, the X_n are stochastic processes.

Moreover, we are interested in studying if these sequences of random functions are tight or not. Notice that, since $X_0 = 0$, the sequence $\{X_n(0)\}$ is tight.

Now, we can recast the Theorem 2.12 in the following way:

$\{X_n\}$ is tight if and only if $\{X_n(0)\}$ is tight and if for each $\varepsilon, \eta > 0$ there exist $0 < \delta < 1$ and $n_0 \in \mathbb{N}$ such that

$$\frac{1}{\delta}\mathcal{P}\left[\sup_{t\leq s\leq t+\delta}|X_n(s) - X_n(t)| \geq \varepsilon\right] \leq \eta, \quad n_0 \geq n, \ 0 \leq t \leq 1. \tag{2.12}$$

Moreover, if $t = \frac{k}{n}$ and $t + \delta = \frac{j}{n}$, $k, j \in \mathbb{Z}$, we obtain

$$\frac{1}{\delta}\mathcal{P}\left[\sup_{t\leq s\leq t+\delta}|X_n(s) - X_n(t)| \geq \varepsilon\right] \leq \eta \Longleftrightarrow$$

$$\Longleftrightarrow \quad \frac{1}{\delta}\mathcal{P}\left[\sup_{t\leq s\leq t+\delta}\left|\frac{1}{\sigma\sqrt{n}}S_{[ns]} + (ns - [ns])\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} - \frac{1}{\sigma\sqrt{n}}S_{[nt]} - \right.\right.$$

$$\left.\left. - \ (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}\right| \geq \varepsilon\right] \leq \eta$$

$$\Longleftrightarrow \quad \frac{1}{\delta}\mathcal{P}\left[\sup_{\frac{k}{n}\leq s\leq \frac{j}{n}}\left|\frac{1}{\sigma\sqrt{n}}S_{[ns]} + (ns - [ns])\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} - \frac{1}{\sigma\sqrt{n}}S_{[k]} - \right.\right.$$

$$\left.\left. - \ (k - [k])\frac{1}{\sigma\sqrt{n}}\xi_{[k]+1}\right| \geq \varepsilon\right] \leq \eta$$

$$\Longleftrightarrow \quad \frac{1}{\delta}\mathcal{P}\left[\sup_{\frac{k}{n}\leq s\leq \frac{j}{n}}\left|\frac{1}{\sigma\sqrt{n}}S_{[ns]} + (ns - [ns])\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} - \frac{1}{\sigma\sqrt{n}}S_{k} - \right.\right.$$

$$\left.\left. - \ (k - k)\frac{1}{\sigma\sqrt{n}}\xi_{k+1}\right| \geq \varepsilon\right] \leq \eta$$

$$\Longleftrightarrow \quad \frac{1}{\delta}\mathcal{P}\left[\sup_{\frac{k}{n}\leq s\leq \frac{j}{n}}\left|\frac{1}{\sigma\sqrt{n}}S_{[ns]} + (ns - [ns])\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} - \frac{1}{\sigma\sqrt{n}}S_{k}\right| \geq \varepsilon\right] \leq \eta.$$

Since we are taking $t = \frac{k}{n}$, we have that the s in the supremum satisfies the condition $\frac{k}{n} \leq s \leq \frac{k+n\delta}{n}$. Therefore, we can take an $i \in \mathbb{Z}$ such that $s = \frac{k+i}{n}$ where i is, at most, $n\delta$. Then, instead of taking the supremum with the condition $t \leq s \leq t + \delta$

we can take the maximum with the condition $i \le n\delta$. Hence,

$$\frac{1}{\delta}\mathcal{P}\left[\sup_{\frac{k}{n}\le s\le \frac{k+n\delta}{n}}\left|\frac{1}{\sigma\sqrt{n}}S_{[ns]}+(ns-[ns])\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1}-\frac{1}{\sigma\sqrt{n}}S_k\right|\ge\varepsilon\right]\le\eta$$

$$\iff \frac{1}{\delta}\mathcal{P}\left[\max_{i\le n\delta}\left|\frac{1}{\sigma\sqrt{n}}S_{[k+i]}+(k+i-[k+i])\frac{1}{\sigma\sqrt{n}}\xi_{[k+i]+1}-\frac{1}{\sigma\sqrt{n}}S_k\right|\ge\varepsilon\right]\le\eta$$

$$\iff \frac{1}{\delta}\mathcal{P}\left[\max_{i\le n\delta}\left|\frac{1}{\sigma\sqrt{n}}S_{k+i}+(k+i-(k+i))\frac{1}{\sigma\sqrt{n}}\xi_{k+i+1}-\frac{1}{\sigma\sqrt{n}}S_k\right|\ge\varepsilon\right]\le\eta$$

$$\iff \frac{1}{\delta}\mathcal{P}\left[\max_{i\le n\delta}\left|\frac{1}{\sigma\sqrt{n}}S_{k+i}-\frac{1}{\sigma\sqrt{n}}S_k\right|\ge\varepsilon\right]\le\eta.$$

Therefore, (2.12) is reduced to

$$\frac{1}{\delta}\mathcal{P}\left[\max_{i\le n\delta}\frac{1}{\sigma\sqrt{n}}\left|S_{k+i}-S_k\right|\ge\varepsilon\right]\le\eta. \tag{2.13}$$

This will help us proving the following result.

Theorem 2.23. *Suppose $\{X_n\}$ is defined by (2.11). Then, the sequence is tight if for each $\varepsilon > 0$ there exist $\lambda > 1$ and $n_0 \in \mathbb{N}$ such that, if $n \ge n_0$, then*

$$\mathcal{P}\left[\max_{i\le n}\left|S_{k+i}-S_k\right|\ge\lambda\sigma\sqrt{n}\right]\le\frac{\varepsilon}{\lambda^2} \tag{2.14}$$

holds for all k.

Remark 2.24. *Requiring $\lambda > 1$ corresponds to the requirement of $\delta < 1$ that we did on Theorem 2.12.*

Proof. Given ε and η we want to produce $0 < \delta < 1$ and n_0 for which (2.13) holds for all k if $n \ge n_0$. Since (2.13) becomes more stringent as ε and η decrease, we may assume $\varepsilon > 0$ and $\eta < 1$.

By hypothesis, with $\eta\varepsilon^2$ in place of ε, there exists $\lambda > 1$ and n_1 such that

$$\mathcal{P}\left[\max_{i\le n}\left|S_{k+i}-S_k\right|\ge\lambda\sigma\sqrt{n}\right]\le\frac{\eta\varepsilon^2}{\lambda^2},\quad \text{for}\quad n\ge n_1,\ k\ge 1. \tag{2.15}$$

Let us take $\delta = \frac{\varepsilon^2}{\lambda^2}$. Since $\lambda > 1 > \varepsilon$, we obtain $0 < \delta < 1$.

Let $n_0 \in \mathbb{N}$ such that $n_0 \ge \frac{n_1}{\delta}$. Notice that, if $n \ge n_0$, then $[n\delta] \ge n_1$. Hence, it follows from (2.15) that

$$\mathcal{P}\left[\max_{i\le[n\delta]}\left|S_{k+i}-S_k\right|\ge\lambda\sigma\sqrt{[n\delta]}\right]\le\frac{\eta\varepsilon^2}{\lambda^2}.$$

Finally, since $\lambda\sqrt{[n\delta]} \le \varepsilon\sqrt{n}$ and $\frac{\eta\varepsilon^2}{\lambda^2} = \eta\delta$, (2.13) holds for all k if $n \ge n_0$.

Therefore,we have proved the theorem.

□

The second thing we need to discuss is to define what the **Wiener Measure** is.

We will denote it with W and basically it is a probability measure on $(\mathcal{C}, \mathscr{C})$ which satisfies the following two properties.

Properties 2.25.

(i) For each t, the random variable x_t is normally distributed under W with zero mean and variance t. That is:

$$W\left[x_t \leq \alpha\right] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{\frac{-u^2}{2t}} du. \tag{2.16}$$

Note that if $t = 0$ we interpret $W\left[x_0 = 0\right] = 1$.

(ii) The stochastic process $\{x_t \colon 0 \leq t \leq 1\}$ has independent increments under W. That is that, if $0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq 1$, then the random variables

$$x_{t_1} - x_{t_0}, x_{t_2} - x_{t_1}, \ldots, x_{t_k} - x_{t_{k-1}}$$

are independent under W.

Notice that, if W has these two properties and we take $s \leq t$, then we have that $x_t \sim N(0, t)$ and $x_s \sim N(0, s)$. Moreover, x_t has to be the sum of the independent random variables x_s and $x_t - x_s$. Therefore, $x_t - x_s$ has to be normally distributed with zero mean and variance $t - s$. Thus, when (2.16) holds, we have

$$W\left[x_{t_i} - x_{t_{i-1}} \leq \alpha_i,\ i = 1, \ldots, k\right] = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \int_{-\infty}^{\alpha_i} e^{\frac{-u^2}{2(t_i - t_{i-1})}} du. \tag{2.17}$$

In particular, the increments are *independent* (i.e. for $t_1 < t_2 < \cdots < t_k$, $k \geq 1$, $x_{t_2} - x_{t_1}, \ldots, x_{t_k} - x_{t_{k-1}}$ are independent) and *stationary* (i.e. the distribution of $x_t - x_s$ under W depends only on the difference $t - s$ and not on the particular values of t or s).

If we interpret $x_t = x(t)$ as the position of a moving particle at time t, then x gives the history of the particle's motion from the time $t = 0$ to the time $t = 1$. What the Wiener measure W does is give to these paths x a distribution which allows us to describe the Brownian motion. We can formalize the definition of the Brownian motion as it follows.

Definition 2.26. *A stochastic process $\{x_t,\ t \geq 0\}$ is a **standard Brownian motion** if the following properties are satisfied:*

(i) $x_0 = 0$, almost surely.

(ii) It has stationary and independent increments.

(iii) For all $0 \leq s < t$, $x_t - x_s \sim N(0, t - s)$.

The appearance of the standard Brownian motion is the one we can find in the following figure.

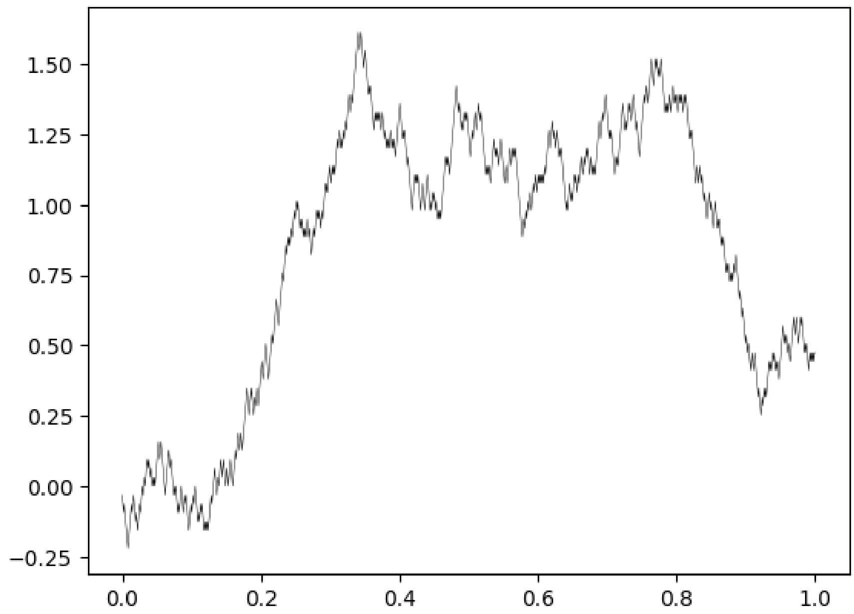

Figure 2: Trajectory of a Standard Brownian Motion.

With all that being said, we finally are able to state the **Donsker's theorem**.

Theorem 2.27 (**Donsker**). *Let $\xi_1, \ldots, \xi_n, \ldots$ be independent identically distributed random variables with zero mean and positive variance σ^2. Define $S_n = \xi_1 + \cdots + \xi_n$ the partial sums of the ξ_i's and define a random element X_n of $\mathcal{C}$ as it follows:*

$$X_n(t, \omega) := \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega). \tag{2.18}$$

Then, these random functions satisfy

$$X_n \xrightarrow{\mathscr{D}} W.$$

Remark 2.28. *Notice that by saying $X_n \xrightarrow{\mathscr{D}} W$ we are just stating that the stochastic process X_n converges in distribution towards the standard Brownian motion.*

Before proving the theorem, we may recall the **Central Limit Theorem of Lévy-Lindeberg** [10]:

Theorem 2.29. *Let $\{\xi_n, n \geq 1\}$ a sequence of independent identically distributed random variables with mean m and variance $\sigma^2 < \infty$. If $S_n = \xi_1 + \cdots + \xi_n$, then*

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{D}} N(0,1).$$

Also, we need two more results that we will use during the prove. These are the following theorem and lemma.

Theorem 2.30. *Let $\mathbb{P}'_n$ and $\mathbb{P}'$ be probability measures on $(S', \mathscr{P}')$ and $\mathbb{P}''_n$ and $\mathbb{P}''$ be probability measures on $(S'', \mathscr{P}'')$. If $S = S' \times S''$ is separable, then $\mathbb{P}'_n \times \mathbb{P}''_n \Rightarrow \mathbb{P}' \times \mathbb{P}''$ if and only if $\mathbb{P}'_n \Rightarrow \mathbb{P}'$ and $\mathbb{P}''_n \Rightarrow \mathbb{P}''$.*

We will prove this result on the **Appendix C**.

Lemma 2.31. *Let $\xi_1, \dots, \xi_m$ be independent random variables with zero mean and finite variances $\sigma_1^2, \dots, \sigma_m^2$ respectively. Set $S_i = \xi_1 + \dots + \xi_i$ and $s_i^2 = \sigma_1^2 + \dots + \sigma_i^2$. Then,*

$$\mathcal{P}\left[\max_{i \le m} |S_i| \ge \lambda s_m\right] \le 2\mathcal{P}\left[|S_m| \ge (\lambda - \sqrt{2}) s_m\right]. \tag{2.19}$$

Notice that for $\lambda \le \sqrt{2}$ the result is trivial.

Proof of the Lemma. Let us consider the following sets,

$$E_i := \left\{\max_{j<i} |S_j| < \lambda s_m \le |S_i|\right\}.$$

Clearly, we have that

$$\mathcal{P}\left[\max_{i \le m} |S_i| \ge \lambda s_m\right] \le \mathcal{P}\left[|S_m| \ge (\lambda - \sqrt{2}) s_m\right] + \sum_{i=1}^{m-1} \mathcal{P}\left[E_i \cap \left\{|S_m| < (\lambda - \sqrt{2}) s_m\right\}\right].$$

Notice that $|S_i| \ge \lambda s_m$ and $|S_m| < (\lambda - \sqrt{2}) s_m$ imply that $|S_m - S_i| \ge \sqrt{2} s_m$. Then, by the independence of the ξ_i and by Chebyshev's inequality[7], we have that the sum in the previous inequality is, at most,

$$\begin{aligned}
\sum_{i=1}^{m-1} \mathcal{P}\left[E_i \cap \left\{|S_m| < (\lambda - \sqrt{2}) s_m\right\}\right] &\le \sum_{i=1}^{m-1} \mathcal{P}(E_i)\, \mathcal{P}\left[|S_m - S_i| \ge \sqrt{2} s_m\right] \\
&\le \sum_{i=1}^{m-1} \mathcal{P}(E_i) \frac{1}{2 s_m^2} \sum_{k=i+1}^{m} \sigma_k^2 \\
&\le \frac{1}{2} \sum_{i=1}^{m-1} \mathcal{P}(E_i) \\
&\le \frac{1}{2} \mathcal{P}\left[\max_{i \le m} |S_i| \ge \lambda s_m\right].
\end{aligned}$$

Hence, combining this inequality and the previous one, we obtain (2.19).

□

We have everything we need to prove the Donsker's Theorem so let us proceed with the proof.

[7] Chebyshev's Inequality: $\mathcal{P}\left[|X_n - X| > \varepsilon\right] \le \frac{1}{\varepsilon^p} \mathbb{E}\left[|X_n - X|^p\right]$.

Proof of the Donsker's Theorem. Our objective is to use Theorem 2.8 to prove the convergence.

The first thing we are going to do is to see that the finite-dimensional distributions of the X_n converge to those of W.

We first consider a single time s. Then, we want to prove

$$X_n(s) \xrightarrow{\mathscr{D}} W_s. \tag{2.20}$$

Notice that we have

$$\begin{aligned} \left|X_n(s) - \tfrac{1}{\sigma\sqrt{n}}S_{[ns]}\right| &= \left|\tfrac{1}{\sigma\sqrt{n}}S_{[ns]} + (ns-[ns])\tfrac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} - \tfrac{1}{\sigma\sqrt{n}}S_{[ns]}\right| \\ &= \left|(ns-[ns])\tfrac{1}{\sigma\sqrt{n}}\xi_{[ns]+1}\right| \\ &\leq \tfrac{1}{\sigma\sqrt{n}}\xi_{[ns]+1}, \end{aligned}$$

since $|ns-[ns]| \leq 1$.

Then, we can apply the Chebyshev's inequality to obtain the following:

$$\begin{aligned} \mathcal{P}\left[\tfrac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} > \varepsilon\right] &\leq \frac{1}{\varepsilon^2}\mathbb{E}\left(\left|\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1}\right|^2\right) \\ &= \frac{1}{\varepsilon^2}\mathbb{E}\left(\frac{1}{\sigma^2 n}\left|\xi_{[ns]+1}\right|^2\right) \xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Hence,

$$\frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} \xrightarrow{\mathcal{P}} 0.$$

Therefore, (2.20) will follow from Theorem 2.22 if we prove that

$$\frac{1}{\sigma\sqrt{n}}S_{[ns]} \xrightarrow[n\to\infty]{\mathscr{D}} W_s.$$

Let us prove it then.

By the Central Limit Theorem of Lévy-Lindeberg (Theorem 2.29) we have that

$$\frac{1}{\sigma\sqrt{[ns]}}S_{[ns]} \xrightarrow[n\to\infty]{\mathscr{D}} N(0,1),$$

where we are using that the ξ's have zero mean. To continue, we can work with this expression and obtain the following:

$$\frac{1}{\sigma\sqrt{[ns]}}S_{[ns]} = \frac{\sqrt{n}S_{[ns]}}{\sigma\sqrt{[ns]}\sqrt{n}} = \frac{S_{[ns]}}{\sigma\sqrt{\frac{[ns]}{n}}\sqrt{n}} \xrightarrow[n\to\infty]{\mathscr{D}} N(0,1).$$

Therefore,

$$\frac{\sqrt{s}}{\sigma\sqrt{\frac{[ns]}{n}}\sqrt{n}}S_{[ns]} \xrightarrow[n\to\infty]{\mathscr{D}} N(0,s) \sim W_s.$$

Now, since $\frac{[ns]}{n} \xrightarrow[n\to\infty]{} s$, we obtain that

$$\frac{\sqrt{s}}{\sigma\sqrt{\frac{[ns]}{n}}\sqrt{n}}S_{[ns]} \xrightarrow[n\to\infty]{} \frac{\sqrt{s}}{\sigma\sqrt{s}\sqrt{n}}S_{[ns]} = \frac{1}{\sigma\sqrt{n}}S_{[ns]}.$$

Hence,

$$\frac{1}{\sigma\sqrt{n}}S_{[ns]} \xrightarrow{\mathscr{D}} W_s,$$

as we wanted to see. Thus, we have proved that

$$X_n \xrightarrow{\mathscr{D}} W_s.$$

What we are going to do, in order to continue with the proof, is to see that, for two times s, t such that $s < t$,

$$(X_n(s), X_n(t)) \xrightarrow{\mathscr{D}} (W_s, W_t).$$

Note that if we prove that

$$(X_n(s), X_n(t) - X_n(s)) \xrightarrow{\mathscr{D}} (W_s, W_t - W_s)$$

then, by Corollary 2.21, we will have the result. Let us prove it then.

Notice that, using the relation

$$\left| X_n(s) - \frac{1}{\sigma\sqrt{n}}S_{[ns]} \right| \leq \frac{1}{\sigma\sqrt{n}}\xi_{[ns]+1} \xrightarrow{\mathcal{P}} 0,$$

as we did before, we can reduce the problem into proving

$$\left(\frac{1}{\sigma\sqrt{n}}S_{[ns]}, \ \frac{1}{\sigma\sqrt{n}}(S_{[nt]} - S_{[ns]}) \right) \xrightarrow{\mathscr{D}} (W_s, W_t - W_s).$$

Again, by Theorem 2.29, we have, separately,

$$\frac{S_{[ns]}}{\sigma\sqrt{n}} \xrightarrow{\mathscr{D}} W_s \quad \text{and} \quad \frac{S_{[nt]} - S_{[ns]}}{\sigma\sqrt{n}} \xrightarrow{\mathscr{D}} W_{t-s}.$$

Notice that, since the ξ_i's are independent, the components on the left are independent. Then, by the Theorem 2.30, we have proved the result for two time points.

We have seen the convergence of the finite-dimensional distributions for one and two time points. For three or more time points we can work as we have done it for two time points. Therefore, we have seen that the finite-dimensional distributions of the X_n's converge weakly to those of W.

Note that, now, it just remains to prove tightness. To do so, we will use the Lemma 2.31. We can apply it to the random variables on Donsker's Theorem. Therefore, by Lemma 2.31, we have that

$$\mathcal{P}\left[\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right] \leq 2\mathcal{P}\left[|S_n| \geq (\lambda-\sqrt{2})\sigma\sqrt{n}\right].$$

Moreover, for $\lambda > 2\sqrt{2}$, we have that $\frac{\lambda}{2} < \lambda - \sqrt{2}$. Hence, to simplify the previous inequality, we can take

$$\mathcal{P}\left[\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right] \leq 2\mathcal{P}\left[|S_n| \geq \frac{1}{2}\lambda\sigma\sqrt{n}\right].$$

By Theorem 2.29 and Chebyshev's inequality (with $p = 3$), we have that

$$\mathcal{P}\left[|S_n| \geq \frac{1}{2}\lambda\sigma\sqrt{n}\right] = \mathcal{P}\left[\frac{|S_n|}{\sigma\sqrt{n}} \geq \frac{1}{2}\lambda\right] \xrightarrow{\mathscr{D}} \mathcal{P}\left[|N| \geq \frac{1}{2}\lambda\right] < \frac{8}{\lambda^3}\mathbb{E}\left(|N|^3\right),$$

where N is a random variable with standard normal distribution, $N \sim N(0,1)$.

Hence, we have that

$$\mathcal{P}\left[\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right] < 2\cdot\frac{8}{\lambda^3}\mathbb{E}\left(|N|^3\right) = \frac{16}{\lambda^3}\mathbb{E}\left(|N|^3\right).$$

Therefore, for a fixed $\varepsilon > 0$ and for a λ big enough, we have that

$$\limsup_{n\to\infty}\mathcal{P}\left[\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right] \leq \limsup_{n\to\infty} 2\mathcal{P}\left[|S_n| \geq \frac{1}{2}\lambda\sigma\sqrt{n}\right] \leq \frac{16}{\lambda^3}\mathbb{E}\left(|N|^3\right) < \frac{\varepsilon}{\lambda^2}.$$

We can use this last bound because, for ε fixed, since $\mathbb{E}\left(|N|^3\right)$ is just a value, we have

$$\frac{16}{\lambda}\mathbb{E}\left(|N|^3\right) < \varepsilon.$$

Then, we just have to take λ such that

$$\lambda > \frac{16\mathbb{E}\left(|N|^3\right)}{\varepsilon}.$$

Therefore, for fixed ε and λ big enough, we have

$$\limsup_{n\to\infty}\mathcal{P}\left[\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right] < \frac{\varepsilon}{\lambda^2}.$$

Thus, by Theorem 2.6 we obtain tightness and hence, by Theorem 2.8, we have proved the Donsker's Theorem.

□

3 Uniform Tansport Processes

We have proved a classical result which shows the convergence to the Brownian motion of a certain type of stochastic processes. What we want to do now is to study another type of stochastic processes which also converge to the Brownian motion. In fact, our aim is to prove a notion of weaker convergence (convergence in distribution) and also a stronger convergence (an almost sure convergence).

But, before talking about the convergences, let us see what are these stochastic processes.

We can find a definition of the uniform transport processes in the introduction of [1] but, as it says in itself, the origin of such processes can be found in [7], where Kac, trying to obtain a solution of the telegrapher's equation

$$\frac{1}{v}\frac{\partial^2 F}{\partial t^2} = v\frac{\partial^2 F}{\partial x^2} - \frac{2a}{v}\frac{\partial F}{\partial t}, \tag{3.1}$$

with $a, v > 0$ and where $F(x, 0) = \varphi(x)$ with $\varphi(x)$ a smooth enough function such that $\left(\frac{\partial F}{\partial t}\right)_{t=0} = 0$, introduced the processes

$$x(t) = v\int_0^t (-1)^{N_a(r)} dr, \tag{3.2}$$

where $N_a = \{N_a(t), t \leq 0\}$ is a Poisson process of intensity a.

Furthermore, he noticed that if, in (3.1), the parameters a and v tend to infinity with $\frac{2a}{v^2} = \frac{1}{D}$ constant, then the solution of the equation converges to the solution of what we know as the heat equation:

$$\frac{1}{D}\frac{\partial F}{\partial t} = \frac{\partial^2 F}{\partial x^2}. \tag{3.3}$$

Stroock, in [13], proved that a modification of the processes in (3.2) converges in law (or in distribution) to a standard Brownian motion. Such modification is of the following form:

$$x_n = \left\{ x_n(t) := \frac{1}{\sqrt{n}}\int_0^{nt} (-1)^{N(u)} du,\ t \in [0, T] \right\},\ \forall n \in \mathbb{N}, \tag{3.4}$$

where $\{N(t),\ t \geq 0\}$ is a standard Poisson process.

If we look at these processes, we can obtain

$$x_n(t) = \frac{1}{\sqrt{n}}\int_0^{nt} (-1)^{N(u)} du = \sqrt{n}\int_0^t (-1)^{N(ns)} ds = \sqrt{n}\int_0^t (-1)^{N_n(s)} ds.$$

Then, these $x_n(t)$ are the processes that Kac considered with $a = n$ and $v = \sqrt{n}$ and, also, the constant satisfyng $\frac{2a}{v^2} = \frac{1}{D}$ is $D = \frac{1}{2}$.

The fact that x_n converges in law to a standard Brownian motion is that if we consider its image law, $\{\mathbb{P}_n\}$, in the Banach space of continuous functions on $[0, T]$, $\mathcal{C}([0, T])$, then $\{\mathbb{P}_n\}$ converges weakly, as $n \uparrow \infty$, towards to the Wiener measure.

This result of convergence in law to the Brownian motion is the first one that we will prove.

But we will not stay here, we will go further. We can find, in mathematical literature, generalizations of this result of convergence of Stroock. A way of generalizing it is to see a stronger way of convergence. We will study a paper written by Griego, Heath and Ruiz-Moncayo which proves that a modification of the processes in (3.4) converges strongly and uniformly on bounded time intervals to a Brownian motion.

These modifications that Griego, Heath and Ruiz-Moncayo consider for this stronger result are what we call **uniform transport processes**. These can be represented as

$$\tilde{x}_n(t) := \frac{1}{\sqrt{n}}(-1)^A \int_0^{nt} (-1)^{N(u)} du, \tag{3.5}$$

where $A \sim Bernoulli(\frac{1}{2})$ is independent of the Poisson process N.

With all that being said, let us start with the result of convergence in distribution proved by Stroock.

3.1 Convergence in Distribution

In this section we will prove the first result that we have commented. To do so, we will first analyze the processes we have defined on (3.4).

Notice that such processes have a Poisson process involved on its definition. Therefore, we are going to do first a brief introduction (based on [11]) on Poisson processes and then we will continue analyzing these processes considered by Stroock.

Let us begin this introduction to the Poisson process by defining what **counting processes** are.

Definition 3.1. *We say that the stochastic process $\{N(t), t \geq 0\}$ is a* ***counting process*** *if it satisfies:*

(i) $N(0) = 0$.

(ii) $N(t)$ *is an integer for every* $t \geq 0$.

(iii) $N(t) \geq 0$ *for every* $t \geq 0$.

(iv) If $s < t$ *then* $N(s) < N(t)$.

Remark 3.2. *In general, if $s < t$, we can interpret the increments $N(t) - N(s)$ as the number of events that have occurred during the time interval $(s, t]$.*

We are interested in these counting processes because the Poisson process is just a counting process that satisfies some particular properties. Then, let us see how do we define the Poisson process.

Definition 3.3. *The counting process* $\{N_\lambda(t),\ t \geq 0\}$ *is a* ***Poisson process parameterized by*** λ, $\lambda > 0$, *if it satisfies:*

(i) $N_\lambda(0) = 0$ *and* $N_\lambda(t) \geq 0$.

(ii) $N_\lambda(t)$ *is an integer for every* $t \geq 0$.

(iii) The process has independent increments, i.e., for all $t_1 < t_2 < \cdots < t_m$, $N_\lambda(t_2) - N_\lambda(t_1), \ldots, N_\lambda(t_m) - N_\lambda(t_{m-1})$ *are independent.*

(iv) For every $t, s \geq 0$, $N_\lambda(t+s) - N_\lambda(t) \sim$ *Poiss*(λs). *That is,*

$$\mathbb{P}\left[N_\lambda(t+s) - N_\lambda(t) = n\right] = e^{-\lambda s}\frac{(\lambda s)^n}{n!}.$$

Remark 3.4. *When we have* λ *being a constant, we say that the Poisson process is homogeneous. On the other hand, if the* λ *is a function depending on time,* $\lambda(t)$, *we say that it is non-homogeneous.*

The following figure shows how the trajectory of a Poisson process looks like.

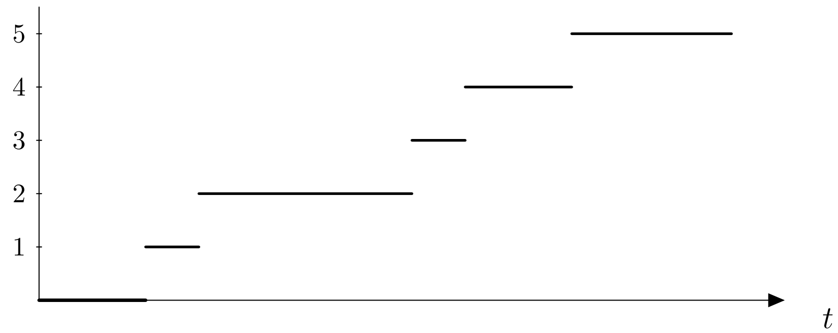

Figure 3: Trajectory of a Poisson process $N(u)$.

Now that we have introduced what a Poisson process is, let us picture what those processes that Stroock considered are.

We start with a Poisson process. We will consider the one that we have defined before in Figure 3. Then, the representation of $(-1)^{N(u)}$ will be a process that is 1 whenever the value of the Poisson process is even and it is -1 whenever the value of the Poisson process is odd. That is what the following figure represents.

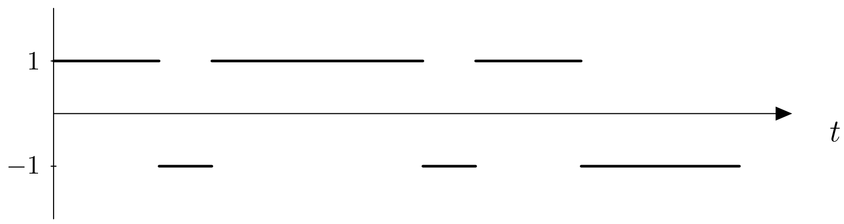

Figure 4: Trajectory of the process $(-1)^{N(u)}$.

Then, if we consider the integral of this process, we obtain a new process which looks like the graphic showed in Figure 5.

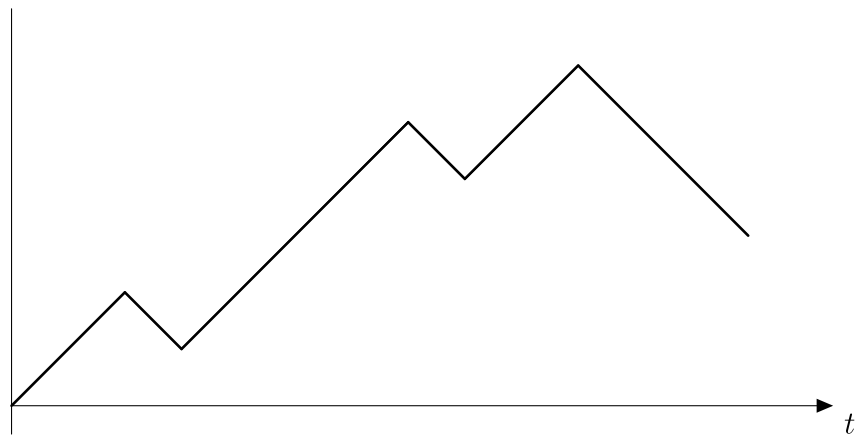

Figure 5: Trajectory of the process $\int_0^t (-1)^{N(u)} du$.

Notice that, so far, we have been picturing how the processes $\int_0^t (-1)^{N(u)} du$ look like. In order to have the processes considered by Stroock it just remains to multiply these processes by $\frac{1}{\sqrt{n}}$. Basically what this multiplication by this term does is to impose that the time passes faster on the process, so we have more ups and downs in smaller time intervals.

With all this reasoning, we have pictured the processes that we need to consider.

The result that we want to prove in this section is the following theorem presented by Stroock [13].

Theorem 3.5. *Consider* $\{N(t),\ t \geq 0\}$ *a Poisson process and define, for all* $n \in \mathbb{N}$, *the continuous processes*

$$x_n = \left\{ x_n(t) \colon \frac{1}{\sqrt{n}} \int_0^{nt} (-1)^{N(u)} du, \quad t \in [0,T] \right\}. \tag{3.6}$$

If $\{\mathbb{P}_n\}$ *are the distributions of the* x_n *in the Banach space* $\mathcal{C}([0,T])$ *of continuous functions on* $[0,T]$, *then the sequence* $\{\mathbb{P}_n\}$ *converges in distribution, as* $n \uparrow \infty$, *towards the Wiener measure.*

Remark 3.6. *To simplify the proof we will assume that the Poisson process has intensity* 1.

To prove the theorem, we will use Theorem 2.8, as we have done in the proof of the Donsker's theorem. Therefore, we have to see that the sequence of distributions of the stochastic processes we are studying, $\{\mathbb{P}_n\}$, is tight and converges weakly to the law of a Brownian motion, let us say $\mathbb{P}$.

Therefore, let us proceed with the proof of Theorem 3.5.

3.1.1 Proof of tightness

To prove tightness we will use the Billingsley criterion for tightness (Theorem 12.3 in [2]). That is the following result.

Theorem 3.7 (Billingsley Criterion)**.** *Let* $(\Omega, \mathscr{B}, \mathcal{P})$ *be a probability space. Let* $\{X_n\}$ *be a sequence of stochastic processes in* $\mathcal{C} = \mathcal{C}([0,1])$*. The sequence* $\{X_n\}$ *is tight if it satisfies these two conditions:*

(i) The sequence $\{X_n(0)\}$ *is tight.*

(ii) There exist $\gamma \geq 0$ *and* $\alpha > 1$ *constants and* F *a non-decreasing, continuous function on* $[0,1]$ *such that*

$$\mathcal{P}\left[|X_n(t_2) - X_n(t_1)| \geq \lambda\right] \leq \frac{1}{\lambda^\gamma} |F(t_2) - F(t_1)|^\alpha \tag{3.7}$$

holds for all t_1, t_2*, for all* n *and for all positive* λ*.*

Remark 3.8. *Notice that, by the Markov's inequality* [8]*, the condition in (3.7) is satisfied if*

$$\mathbb{E}\left(|X_n(t_2) - X_n(t_1)|^\gamma\right) \leq |F(t_2) - F(t_1)|^\alpha .$$

Then, using the Billingsley Criterion, we are going to prove tightness.

Because of the first condition in the criterion, notice that

$$x_n(0) = \frac{1}{\sqrt{n}} \int_0^0 (-1)^{N(u)} du = 0.$$

Therefore, the sequence $\{x_n(0)\}_n$ is constantly 0 and, hence, this sequence is tight.

On the other hand, by the second condition in the criterion, we just have to prove

$$\left(\mathbb{E}\left(x_n(t) - x_n(s)\right)\right)^4 \leq C(t-s)^2,$$

for C a constant.

Let us prove it. First of all, we have that

$$\begin{aligned} \left(\mathbb{E}\left(x_n(t) - x_n(s)\right)\right)^4 &= \left(\mathbb{E}\left(\frac{1}{\sqrt{n}}\int_0^{nt} (-1)^{N(u)} du - \frac{1}{\sqrt{n}}\int_0^{ns} (-1)^{N(u)} du\right)\right)^4 \\ &= \frac{1}{n^2}\left(\mathbb{E}\left(\int_{ns}^{nt} (-1)^{N(u)} du\right)\right)^4 . \end{aligned}$$

[8]**Markov's Inequality:** If X is a non-negative random variable and $a > 0$, then, using higher moments of X supported on values larger than 0, $\mathcal{P}\left[|X| \geq a\right] \leq \dfrac{\mathbb{E}\left(|X|^n\right)}{a^n}$.

Now, if we work with this, we can obtain the following.

$$\begin{aligned}
&\frac{1}{n^2}\left(\mathbb{E}\left(\int_{ns}^{nt}(-1)^{N(u)}du\right)\right)^4 \\
=\;& \frac{1}{n^2}\mathbb{E}\left(\int_{ns}^{nt}\int_{ns}^{nt}\int_{ns}^{nt}\int_{ns}^{nt}(-1)^{N(u_1)+N(u_2)+N(u_3)+N(u_4)}du_1du_2du_3du_4\right) \\
=\;& \frac{24}{n^2}\mathbb{E}\left(\int_{ns}^{nt}\int_{ns}^{u_4}\int_{ns}^{u_3}\int_{ns}^{u_2}(-1)^{N(u_4)-N(u_3)+N(u_2)-N(u_1)}du_1du_2du_3du_4\right) \\
=\;& \frac{24}{n^2}\Bigg(\int_{ns}^{nt}\int_{ns}^{u_4}\int_{ns}^{u_3}\int_{ns}^{u_2}\mathbb{E}\left((-1)^{N(u_4)-N(u_3)}\right)\mathbb{E}\left((-1)^{N(u_2)-N(u_1)}\right)\times \\
&\qquad \times du_1du_2du_3du_4\Bigg).
\end{aligned}$$

Then, using that if $X \sim Poiss(\lambda)$, then

$$\mathbb{E}\left((-1)^X\right) = e^{-2\lambda},$$

we can obtain the following:

$$\begin{aligned}
\left(\mathbb{E}\left(x_n(t)-x_n(s)\right)\right)^4 &\leq \frac{24}{n^2}\left(\int_{ns}^{nt}\int_{ns}^{u_2}\mathbb{E}\left((-1)^{N(u_2)-N(u_1)}\right)du_1du_2\right)^2 \\
&= \frac{24}{n^2}\left(\int_{ns}^{nt}\int_{ns}^{u_2}e^{-2(u_2-u_1)}du_1du_2\right)^2.
\end{aligned}$$

To continue, let us compute these integrals. First of all,

$$\begin{aligned}
\int_{ns}^{u_2}e^{-2(u_2-u_1)}du_1 &= \frac{1}{2}\int_{ns}^{u_2}2e^{2u_1-2u_2}du_1 \\
&= \frac{1}{2}\left[e^{2u_1-2u_2}\right]_{u_1=ns}^{u_2} \\
&= \frac{1}{2}\left(1-e^{-2(u_2-ns)}\right).
\end{aligned}$$

Then,

$$\begin{aligned}
\int_{ns}^{nt}\int_{ns}^{u_2}e^{-2(u_2-u_1)}du_1du_2 &= \int_{ns}^{nt}\frac{1}{2}\left(1-e^{-2(u_2-ns)}\right)du_2 \\
&= \int_{ns}^{nt}\left(\frac{1}{2}-\frac{1}{2}e^{-2(u_2-ns)}\right)du_2 \\
&= \frac{n(t-s)}{2}-\frac{1}{2}\int_{ns}^{nt}e^{-2(u_2-ns)}du_2.
\end{aligned}$$

Implementing the change of variables $v = u_2 - ns$ we obtain that

$$\begin{aligned}
\frac{n(t-s)}{2} - \frac{1}{2}\int_{ns}^{nt} e^{-2(u_2-ns)}du_2 &= \frac{n(t-s)}{2} - \frac{1}{2}\int_0^{nt-ns} e^{-2v}dv \\
&= \frac{n(t-s)}{2} + \frac{1}{4}\int_0^{nt-ns} -2e^{-2v}dv \\
&= \frac{n(t-s)}{2} + \frac{1}{4}\left[e^{-2v}\right]_{v=0}^{nt-ns} \\
&= \frac{n(t-s)}{2} + \frac{1}{4}\left(e^{-2n(t-s)} - 1\right) \\
&= \frac{1}{2}\left[n(t-s) + \frac{1}{2}\left(e^{-2n(t-s)} - 1\right)\right].
\end{aligned}$$

Then, doing the square of the integral,

$$\begin{aligned}
\left(\int_{ns}^{nt}\int_{ns}^{u_2} e^{-2(u_2-u_1)}du_1du_2\right)^2 &= \frac{1}{4}\left[n(t-s) + \frac{1}{2}\left(e^{-2n(t-s)} - 1\right)\right]^2 \\
&= \frac{1}{4}\left(n^2(t-s)^2 + n(t-s)\left(e^{-2n(t-s)} - 1\right)\right. \\
&\quad \left. + \frac{1}{4}\left(e^{-2n(t-s)} - 1\right)^2\right).
\end{aligned}$$

Finally, if we multiply by the term $\frac{24}{n^2}$, we obtain that

$$\begin{aligned}
\frac{24}{n^2}\left(\int_{ns}^{nt}\int_{ns}^{u_2} e^{-2(u_2-u_1)}du_1du_2\right)^2 &= \frac{24}{4n^2}\left(n^2(t-s)^2 + n(t-s)\left(e^{-2n(t-s)} - 1\right)\right. \\
&\quad \left. + \frac{1}{4}\left(e^{-2n(t-s)} - 1\right)^2\right) \\
&= \frac{6}{n^2}\left(n^2(t-s)^2 + n(t-s)\left(e^{-2n(t-s)} - 1\right)\right. \\
&\quad \left. + \frac{1}{4}\left(e^{-2n(t-s)} - 1\right)^2\right) \\
&= 6(t-s)^2 + \frac{6(t-s)}{ne^{2n(t-s)}} - \frac{6(t-s)}{n} \\
&\quad + \frac{3}{2n^2}\left(\left(e^{-2n(t-s)}\right)^2 - 2e^{-2n(t-s)} + 1\right).
\end{aligned}$$

Note that

$$6(t-s)^2 + \frac{6(t-s)}{ne^{2n(t-s)}} - \frac{6(t-s)}{n} + \frac{3}{2n^2}\left(\left(e^{-2n(t-s)}\right)^2 - 2e^{-2n(t-s)} + 1\right) \xrightarrow[n\to\infty]{} 6(t-s)^2.$$

Hence, we can say that

$$\left(\mathbb{E}\left(x_n(t) - x_n(s)\right)\right)^4 \leq 6(t-s)^2.$$

Therefore, we have seen that both conditions in the Billingsley criterion are satisfied by the processes $x_n(t)$. Therefore, we can ensure that the sequence is tight.

With tightness being proved it just remains to see that the law of the processes $\{x_n(t)\}$ converge weakly towards the law of a Brownian motion.

3.1.2 Identification of the limit law

In order to prove that the sequence of distributions $\{\mathbb{P}_n\}$ converges in distribution to the law of a Brownian motion we will use the Paul Levy's Theorem. This theorem is a characterization of the Brownian motion. Lévy discussed it in [9]. We can also find a more recent discussion by Le Gall in [8]. The theorem states the following.

Theorem 3.9 (Levy's Theorem). *Let $X = \{X_t,\ t \geq 0\}$ be a continuous stochastic process adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ taking values in $\mathbb{R}$ such that the process*

$$M_t = X_t - X_0, \quad t \geq 0,$$

is a continuous local martingale relative to $(\mathcal{F}_t)_{t\geq 0}$, and whose quadratic variation is given by

$$\langle M, M \rangle_t = t.$$

Then $X = \{X_t,\ t \geq 0\}$ is a Brownian motion.

Let us assume that the distributions of the processes $\{x_n(t)\}$ converge to the distribution of the process $Y = \{Y(t),\ t \geq 0\}$. Then, we just have to check, using Theorem 3.9, that Y satisfies the martingale condition and the quadratic variation condition.

3.1.2.1 Proof of the Martingale Condition

In order to prove that Y is a continuous martingale it is sufficient to prove that for $s_1 \leq s_2 \leq \cdots \leq s_m \leq s < t$ and for any bounded continuous function $f\colon \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\left(f(Y(s_1), \ldots, Y(s_m))(Y(t) - Y(s))\right)) = 0.$$

Since $\{x_n(t)\}$ converges in distribution towards Y, we have the following:

$$\mathbb{E}\left(f(Y(s_1), \ldots, Y(s_m))(Y(t) - Y(s))\right)) =$$

$$= \lim_{n\to\infty} \mathbb{E}\left(f(x_n(s_1), \ldots, x_n(s_m))(x_n(t) - x_n(s))\right)).$$

Therefore, the condition that we want to prove can be translated to

$$\lim_{n\to\infty} \mathbb{E}\left(f(x_n(s_1), \ldots, x_n(s_m))(x_n(t) - x_n(s))\right)) = 0.$$

Let us see this.

$$\begin{aligned}
&\mathbb{E}\left(f\left(x_n(s_1),\dots,x_n(s_m)\right)\left(x_n(t)-x_n(s)\right)\right) = \\
= \quad & \frac{1}{\sqrt{n}}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))\int_{ns}^{nt}(-1)^{N(u)}du\right) \\
= \quad & \frac{1}{\sqrt{n}}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))(-1)^{N(ns)}\int_{ns}^{nt}(-1)^{N(u)-N(ns)}du\right) \\
= \quad & \frac{1}{\sqrt{n}}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))(-1)^{N(ns)}\right)\int_{ns}^{nt}\mathbb{E}\left((-1)^{N(u)-N(ns)}\right)du \\
= \quad & \frac{1}{\sqrt{n}}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))(-1)^{N(ns)}\right)\int_{ns}^{nt}e^{-2(u-ns)}du.
\end{aligned}$$

Since f is bounded we can bound the expectation and obtain

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))(-1)^{N(ns)}\right)\int_{ns}^{nt}e^{-2(u-ns)}du \leq \\
\leq \quad & \frac{C}{\sqrt{n}}\int_{ns}^{nt}e^{-2(u-ns)}du \\
= \quad & \frac{C}{-2\sqrt{n}}\int_{ns}^{nt}-2e^{-2(u-ns)}du \\
= \quad & \frac{C}{-2\sqrt{n}}\left[e^{-2(u-ns)}\right]_{u=ns}^{nt} \\
= \quad & \frac{C}{-2\sqrt{n}}\left(e^{-2n(t-s)}-1\right) \\
= \quad & \frac{C}{2\sqrt{n}}\left(1-e^{-2n(t-s)}\right)\xrightarrow[n\to\infty]{}0.
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\mathbb{E}\left(f(x_n(s_1),\dots,x_n(s_m))(x_n(t)-x_n(s))\right)=0.$$

Thus, we have the martingale condition proved. Now, it just remains to prove the other condition, the one involving the quadratic covariation.

3.1.2.2 Proof of the Quadratic Variation Condition

In this case, we want to see that Y follows the distribution of a Brownian motion. To see so it is enough to prove that for $s_1 \leq s_2 \leq \cdots \leq s_m \leq s < t$ and for any bounded continuous function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}\left(f(Y(s_1),\dots,Y(s_m))(Y(t)-Y(s))^2-(t-s))\right)=0.$$

Notice that we can translate this condition, as we have done previously, using the convergence of the $x_n(t)$'s to the distribution of the $Y(t)$'s. Then, it is sufficient to prove the following.

$$\lim_{n\to\infty} \mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\left(x_n(t)-x_n(s)\right)^2-(t-s)\right)=0.$$

Let us prove it.

$$\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\left(x_n(t)-x_n(s)\right)^2\right)=$$

$$=\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\left(\frac{1}{n}\int_{ns}^{nt}(-1)^{N(u)}du\right)^2\right)$$

$$=\frac{2}{n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\int_{ns}^{nt}\int_{ns}^{u_2}(-1)^{N(u_2)-N(u_1)}du_1du_2\right)$$

$$=\frac{2}{n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\int_{ns}^{nt}\int_{ns}^{u_2}e^{-2(u_2-u_1)}du_1du_2$$

$$=\frac{1}{n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\int_{ns}^{nt}\int_{ns}^{u_2}2e^{-2(u_2-u_1)}du_1du_2$$

$$=\frac{1}{n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\int_{ns}^{nt}\left[e^{-2(u_2-u_1)}\right]_{u_1=ns}^{u_2}du_2$$

$$=\frac{1}{n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\int_{ns}^{nt}\left(1-e^{-2(u_2-ns)}\right)du_2$$

$$=\frac{1}{2n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\int_{ns}^{nt}\left(2-2e^{-2(u_2-ns)}\right)du_2$$

$$=\frac{1}{2n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\left[2u_2-e^{-2(u_2-ns)}\right]_{u_2=ns}^{nt}$$

$$=\frac{1}{2n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\left(2nt-e^{-2n(t-s)}-2ns+1\right)$$

$$=\frac{1}{2n}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\left(2nt-e^{-2n(t-s)}-2ns+1\right)$$

$$=\frac{1}{2n}2n\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\left(t-s+\frac{1}{2n}-\frac{1}{2n}e^{-2n(t-s)}\right)$$

$$=\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))\right)\left(t-s+o\left(\frac{1}{n}\right)\right).$$

Therefore,

$$\lim_{n\to\infty}\mathbb{E}\left(f(x_n(s_1),\ldots,x_n(s_m))(x_n(t)-x_n(s))^2-(t-s))\right)=0.$$

Thus, we have proved the condition involving the quadratic variation.

With all that, we have seen both conditions in the Lévy's Theorem (Theorem 3.9) and, therefore, we have that the processes

$$x_n(t) = \frac{1}{\sqrt{n}} \int_0^{nt} (-1)^{N(u)} du$$

converge in distribution to the Brownian motion.

3.2 Almost Sure Convergence

In this final part of the thesis we want to prove a stronger result than the one that we have just proved. To do so, we will define a generalization of the processes defined in (3.4).

Notice that, since the Poisson process always starts being $N(0) = 0$, the process

$$\frac{1}{\sqrt{n}} \int_0^{nt} (-1)^{N(u)} du$$

always starts increasing. Therefore, if we want to get a more general result, we have to solve this issue.

What we will do is what we have seen briefly in the beginning of this section. We will consider a random variable, A, following a Bernoulli distribution.

A Bernoulli distribution is a discrete probability distribution of a random variable which takes only two values: the random variable is either 1 with probability p or is 0 with probability $q = 1 - p$.

In our case, we will choose $A \sim Bernoulli(\frac{1}{2})$. Therefore, A takes the values 1 and 0 with the same probability.

Then, if we add the term $(-1)^A$ multiplying the integral in $x_n(t)$, we obtain the processes defined at (3.5),

$$\tilde{x}_n(t) := \frac{1}{\sqrt{n}} (-1)^A \int_0^{nt} (-1)^{N_\lambda(u)} du,$$

where $N_\lambda(t)$ is a Poisson process of intensity $\lambda > 0$.

Note that this processes start increasing or decreasing with the same probability, so we have solved the issue that we had with the processes considered by Stroock. These new processes are what we call **uniform transport processes**.

The aim of this section is to prove that these uniform transport processes converge almost surely to the Brownian motion. To see that, we will study the paper written by Richard J. Griego, David Heath and Alberto Ruiz-Moncayo, "*Almost Sure Convergence of Uniform Transport Processes to Brownian Motion*" [5].

In this paper, they present uniform transport processes using a Markov chain. In the **Appendix E** we can find an introduction of Markov chains and Markov processes. There, we use [11] to introduce these processes.

In this section of the **Appendices** we can see that the Poisson process is a Markov process. This will be key to understand the way that Griego, Heath and Ruiz-Moncayo define the uniform transport processes using a Markov process, because they base their definition on the fact that the Poisson process can be pictured as a Markov process.

Thus, let us focus on proving the result of almost sure convergence of the uniform transport processes towards the Brownian motion.

3.2.1 Almost Sure Convergence of Uniform Transport Processes

Finally, we will analyse the paper written by Griego, Heath and Ruiz-Moncayo and see how they prove the result of almost surely convergence towards the Brownian motion.

The first thing that we have to do is to present how they define uniform transport processes using a Markov process.

We just have to consider $v(t)$ a Markov process such that it has stationary transition probabilities (so we are working with a homogeneous Markov process), it has states $E = \{1, -1\}$ and that it has the following infinitesimal generator matrix:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then, we can define a sequence of uniform transport processes by

$$y_n(t) := n \int_0^t v(n^2 s) ds \tag{3.8}$$

for $n = 1, 2, \ldots$.

Note that these processes represent the position of a particle at time $t \geq 0$ in one dimension that switches between the uniform velocities n and $-n$ at the jump times of a Poisson process with intensity n^2, let us say $N_n(t)$.

To simplify the proof, as we did in the previous section, we will take the Poisson process with parameter 1.

To continue, we will proceed to prove the result of convergence. In order to do so,we need a result of Skorokhod that we can find in [12] (page 163). The theorem we need states the following:

Theorem 3.10 (Skorokhod Theorem)**.** *Suppose that $\xi_1, \xi_2, \ldots, \xi_n$ are random variables such that $\mathbb{E}(\xi_i) = 0$ and $Var(\xi_i) < \infty$ for every i. Suppose also that $B(t)$ is a Brownian motion. Then, there exist non-negative independent random variables $\tau_1, \tau_2, \ldots, \tau_n$ for which the variables*

$$B(\tau_1), B(\tau_1 + \tau_2) - B(\tau_1), \ldots, B(\tau_1 + \tau_2 + \cdots + \tau_n) - B(\tau_1 + \tau_2 + \cdots + \tau_{n-1}),$$

have the same joint distribution as do $\xi_1, \xi_2, \ldots, \xi_n$. Moreover, the following properties are satisfied:

(i) $\mathbb{E}(\tau_k) = Var(\xi_k)$ *for every k.*

(ii) There exists L_m a constant such that

$$(\mathbb{E}(\tau_k))^m \leq L_m (\mathbb{E}(\xi_k))^{2m},$$

for every k.

(iii) If $|\xi_i| \leq h$, then

$$\left| B(s) - B\left(\sum_{i=1}^{k} \tau_i\right)\right| \leq h \quad \text{for} \quad s \in \left[\sum_{i=1}^{k} \tau_i, \sum_{i=1}^{k+1} \tau_i\right].$$

We will not enter into the proof of this theorem.

Another result that we are going to use is the Borel-Cantelli lemma, that we can find in [10]. This lemma states the following.

Lemma 3.11. *Let $\{A_n,\ n \geq 1\}$ be a sequence of sets in the measurable space $(\Omega, \mathscr{B})$. Then, the next properties are satisfied:*

(i) $\mathcal{P}\left[\limsup_n A_n\right] \geq \limsup_n \mathcal{P}[A_n]$.

(ii) $\mathcal{P}\left[\liminf_n A_n\right] \leq \liminf_n \mathcal{P}[A_n]$.

Remark 3.12. *Notice that we can define the limit superior and the limit inferior as it follows. Let $\{x_n\}$ be a sequence. Then*

- $\limsup_{n\to\infty} x_n :=\lim_{n\to\infty}\left(\sup_{m\geq n} x_m\right)$
- $\liminf_{n\to\infty} x_n :=\lim_{n\to\infty}\left(\inf_{m\geq n} x_m\right)$

Let us continue by writing the main result that we want to prove in this section.

Theorem 3.13. *There exist realizations $\{y_n(t),\ t \geq 0\}$ of the previous uniform transport processes defined by (3.8) on the same probability space, $(\Omega, \mathscr{B}, \mathcal{P})$, as a standard Brownian motion $\{B(t),\ t \geq 0\}$, with $B(0) \equiv 0$, so that we have*

$$\lim_{n\to\infty} \max_{0\leq t\leq 1} |y_n(t) - B(t)| = 0, \quad \textit{almost surely.}$$

3.2.1.1 Proof of Theorem 3.13

3.2.1.1.1 Construction of Uniform Transport Processes

The first thing that we need to do is to construct, in an appropriate way, the uniform transport processes.

Therefore, let $(\Omega, \mathscr{B}, \mathcal{P})$ be the probability space for a standard Brownian motion $\{B(t),\ t \geq 0\}$ with $B(0) \equiv 0$.

In this same probability space we define, for each $n = 1, 2, \ldots$, a sequence of independent random variables $\left\{\xi_i^{(n)}\right\}_{i\geq 1}$, where each $\xi_i^{(n)}$ follows an exponential distribution with parameter $2n$, $\xi_i^{(n)} \sim Exp(2n)$. That is

$$\mathcal{P}\left[\xi_i^{(n)} > \lambda\right] = e^{-2n\lambda} \quad \text{for} \quad \lambda \geq 0.$$

We also assume that these random variables are independent of the Brownian motion $B(t)$.

Moreover, let $\{k_j\}_{j\geq 1}$ be a sequence of independent random variables that satisfy the following property:

$$\mathcal{P}\left[k_i = 1\right] = \mathcal{P}\left[k_i = -1\right] = \frac{1}{2}$$

for each i. Again, we assume that these random variables are independent of the Brownian motion and also that they are independent of the previous random variables we have defined, $\xi_i^{(n)}$'s.

Notice that these k_i's have a Bernoulli distribution with parameter $\frac{1}{2}$. Therefore, they will develop the role of the random variable A that we have defined at the beginning of this section in order to randomise the start of the processes considered by Stroock in the previous section. Hence, with these k_i's we will get this "starting increasing" or "starting decreasing" with the same probability.

Also, notice that, in order to achieve all this independence between the different random variables, we need to introduce a product space. For convenience, we will still call this new space $(\Omega, \mathscr{B}, \mathcal{P})$.

To continue, and involving all the random variables that we have defined, we consider the sequence of independent identically distributed random variables $\left\{k_i\xi_i^{(n)}\right\}_{i\geq 1}$ for each $n \geq 1$.

Notice that

$$\mathbb{E}\left(k_i\xi_i^{(n)}\right) = \mathbb{E}\left(k_i\right)\mathbb{E}\left(\xi_i^{(n)}\right) = \frac{\left(\mathcal{P}\left[k_i = 1\right]\cdot 1 + \mathcal{P}\left[k_i = -1\right]\cdot(-1)\right)}{2n} = \frac{\left(\frac{1}{2} - \frac{1}{2}\right)}{2n} = 0.$$

Moreover,

$$Var\left(k_i\xi_i^{(n)}\right) = \mathbb{E}\left(\left(k_i\xi_i^{(n)}\right)^2\right) - \left(\mathbb{E}\left(k_i\xi_i^{(n)}\right)\right)^2 = \mathbb{E}\left(k_i^2\left(\xi_i^{(n)}\right)^2\right).$$

Since k_i is 1 or -1, we have that $k_i^2 = 1$. Hence,

$$Var\left(k_i\xi_i^{(n)}\right) = \mathbb{E}\left(\left(\xi_i^{(n)}\right)^2\right).$$

Using the formula of the moments for an exponential distribution $Exp(\lambda)$,

$$\mathbb{E}\left(X^n\right)=\frac{n!}{\lambda^n},$$

we obtain that

$$Var\left(k_i\xi_i^{(n)}\right)=\mathbb{E}\left(\left(\xi_i^{(n)}\right)^2\right)=\frac{2!}{(2n)^2}=\frac{2}{4n^2}=\frac{1}{2n^2},$$

since $\xi_i^{(n)} \sim Exp(2n)$. Therefore, the random variables $k_i\xi_i^{(n)}$ have zero mean and variance $\dfrac{1}{2n^2}$.

Now, by Theorem 3.10, we have that for each $n \geq 1$ there exists a sequence of non-negative independent identically distributed random variables on $(\Omega, \mathscr{B}, \mathcal{P})$, let us say $\left\{\sigma_i^{(n)}\right\}_{i\geq 1}$, such that the following is satisfied.

If we define the sequence of partial sums $S_i=\sigma_1^{(n)}+\cdots+\sigma_i^{(n)}$, then the sequence $\{B\left(S_i\right)\}_{i\geq 1}$ has the same distribution as the sequence of partial sums of the $k_i\xi_i^{(n)}$'s, let us say $\left\{\tilde{S}_i\right\}_{i\geq 1}$ with $\tilde{S}_i=k_1\xi_1^{(n)}+\cdots+k_i\xi_i^{(n)}$, and, furthermore, by the first condition in Theorem 3.10,

$$\mathbb{E}\left(\sigma_i^{(n)}\right)=Var\left(k_i\xi_i^{(n)}\right)=\frac{1}{2n^2}.$$

Now, we define, for a fixed n and for $i=1,2,\ldots$,

$$\gamma_i^{(n)}:=\frac{\left|B\left(S_i\right)-B\left(S_{i-1}\right)\right|}{n}=\frac{\left|B\left(\sum_{j=0}^{i}\sigma_j^{(n)}\right)-B\left(\sum_{j=0}^{i-1}\sigma_j^{(n)}\right)\right|}{n}, \tag{3.9}$$

where we consider $\sigma_0^{(n)}\equiv 0$.

Notice that the random variables $\left\{\gamma_i^{(n)}\right\}_{i\geq 1}$ are independent and identically distributed, all following an exponential distribution with parameter $2n^2$. Therefore,

$$\mathbb{E}\left(\gamma_i^{(n)}\right)=\frac{1}{2n^2}.$$

To continue, we define a new stochastic process $\left\{B^{(n)}(t),\ t\geq 0\right\}$. We want it to be polygonal in the same sense as the processes that we have defined in the first section for the Donsker's theorem. Furthermore, we want them to satisfy

$$B^{(n)}\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)=B\left(\sum_{j=1}^{i}\sigma_j^{(n)}\right), \tag{3.10}$$

and $B^{(n)}(0)\equiv 0$.

Hence, with this definition, we have that $B^{(n)}(\cdot)$ has slope n or $-n$.

Also, let $\tau_i^{(n)}$ be the time of the i-th discontinuity of the right-hand derivative of $B^{(n)}(\cdot)$. This means that the process $B^{(n)}(\cdot)$ alternates linearly increasing and linearly decreasing and the $\tau_i^{(n)}$'s are the points where it changes from increasing to decreasing or vice versa. Since in each piece where it increases the process is linear, and the same for every piece where it decreases, the derivative on the $\tau_i^{(n)}$'s is discontinuous.

Now, what we do is claim that this $B^{(n)}(t)$ is a realization of the n-th uniform transport process.

3.2.1.1.2 Independent Increments Exponentially Distributed

To show that this $B^{(n)}(t)$ is, indeed, a realization of the n-th uniform transport process we need to check that the increments $\tau_i^{(n)} - \tau_{i-1}^{(n)}$, for $i = 1, 2, \ldots$, are independent and they all have an exponential distribution with parameter n^2. Let us check this.

First of all, we are going to see that the $\tau_1^{(n)}$ follows an exponential distribution with parameter n^2.

Notice first that the probability of

$$B\left(\sum_{j=0}^{i} \sigma_j^{(n)}\right) - B\left(\sum_{j=0}^{i-1} \sigma_j^{(n)}\right)$$

being positive is $\frac{1}{2}$, independent of the past up to time $\sum_{j=0}^{i-1} \sigma_j^{(n)}$. Therefore,

$$\tau_1^{(n)} = \gamma_1^{(n)} + \cdots + \gamma_N^{(n)},$$

where we have, for $i = 1, 2, \ldots$,

$$\mathcal{P}\left[N = i\right] = \frac{1}{2^i}.$$

Then, to see the distribution of $\tau_1^{(n)}$, we will follow what William Feller does in [3] (page 54, (5.6)).

In this example, William Feller discusses about the distribution of a random sum of independent identically distributed random variables. Let us see what does he say.

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with a common density f. Then, he defines the density of the random variable $S_n = X_1 + \cdots + X_n$ using the operation of convolution. Let us recall the definition of this operation.

Definition 3.14. *Let X, Y be two independent random variables with densities f and g respectively. The density of the random variable $X + Y$ is the* ***convolution*** *of the two densities:*

$$f * g(s) := \int_{-\infty}^{\infty} f(s-y)g(y)dy = \int_{-\infty}^{\infty} f(y)g(s-y)dy. \tag{3.11}$$

Moreover, if X, Y are positive random variables, then f and g are concentrated on $[0, +\infty]$ and the convolution $f * g$ reduces to:

$$f * g(s) := \int_{0}^{s} f(s-y)g(y)dy = \int_{0}^{s} f(x)g(s-x)dx. \tag{3.12}$$

Notice that, if we have k independent random variables with common density f, the density of the random sum $S_k = X_1 + \cdots + X_k$ is the k-fold convolution of f with itself, $f^{k*} = f * \overset{k}{\overbrace{\cdots}} * f$.

Then, what Feller does is randomize this number of terms k by doing

$$\mathcal{P}\left[N = k\right] = p_k.$$

Therefore, the density of the random sum S_N with the random number of terms N is

$$w = \sum_{k=1}^{\infty} p_k f^{k*}.$$

Feller gives a particular example considering $\{p_k\}_{k\geq 1}$ geometric distributed (i.e. $p_k = qp^{k-1}$) and considering f being an exponential density with parameter α. Then,

$$f^{k*}(x) = \frac{1}{(k-1)!}\alpha^k x^{k-1} e^{-\alpha x}.$$

With this, the density of S_N is

$$w(x) = q\alpha^{-\alpha x} \sum_{i=1} p^{i-1} \frac{(\alpha x)^{i-1}}{(i-1)!} = q\alpha e^{-\alpha q x},$$

i.e., is exponentially distributed.

In our case, we are working with the random sum

$$\tau_1^{(n)} = \gamma_1^{(n)} + \cdots + \gamma_N^{(n)}$$

where the $\gamma_i^{(n)}$'s are exponentially distributed with parameter $2n^2$ and the p_k follow a geometric distribution with parameter $\frac{1}{2}$, i.e. $p_k = 2^{-k}$.

Therefore, applying this example presented by Feller to our case, we obtain that the density of $\tau_1^{(n)}$ is the following one.

$$
\begin{aligned}
w(x) &= \sum_{k=1}^{\infty} p_k f^{k*}(x) \\
&= \sum_{k=1}^{\infty} 2^{-k} \frac{1}{(k-1)!} (2n^2)^k x^{k-1} e^{-2n^2 x} \\
&= n^2 e^{-2n^2 x} \sum_{k=1}^{\infty} \frac{(n^2 x)^{k-1}}{(k-1)!} \\
&= n^2 e^{-2n^2 x} e^{n^2 x} \\
&= n^2 e^{-n^2 x}.
\end{aligned}
$$

This means that $\tau_1^{(n)}$ is exponentially distributed with parameter n^2, as we wanted to see. Notice that the parameter of this exponential distribution is half the parameter of the exponential distribution corresponding to the $\gamma_i^{(n)}$'s.

Therefore, we have that the $\tau_i^{(n)}$'s have an exponential distribution with parameter n^2 and, hence, the increments $\tau_i^{(n)} - \tau_{i-1}^{(n)}$, for $i = 1, 2, \ldots$, have such distribution too.

Furthermore, since these increments are sums of disjoint blocks of the $\gamma_i^{(n)}$'s, we have that they are independent.

3.2.1.1.3 Proof of the Almost Sure Convergence

To finish, what we want to do is prove the almost sure convergence towards the Brownian motion.

What we are going to do is apply the *Kolmogorov inequality* ([10] page 92), which states the following.

Definition 3.15. *Let $X_1, \ldots, X_n$ be independent random variable with zero mean and satisfying $\mathbb{E}\left(X_i^2\right) < \infty$ for every $i = 1, \ldots, n$. Let $S_k = X_1 + \cdots + X_k$ for $1 \leq k \leq n$. Then, for all $\varepsilon > 0$ the following is satisfied:*

$$
\mathcal{P}\left[\max_{1 \leq k \leq n} |S_k| > \varepsilon\right] \leq \frac{Var\left(S_n\right)}{\varepsilon^2}. \tag{3.13}
$$

Applying this inequality to the random variables we are working with we obtain that for all $\varepsilon > 0$ we have

$$
\mathcal{P}\left[\max_{1 \leq i \leq 2n^2} \left|\gamma_1^{(n)} + \cdots + \gamma_i^{(n)}\right| \geq \varepsilon\right] \leq \frac{Var\left(\gamma_1^{(n)} + \cdots + \gamma_{2n^2}^{(n)}\right)}{\varepsilon^2}. \tag{3.14}
$$

If we work on this a little bit, we can obtain the following.

$$\begin{aligned}
\mathcal{P}\left[\max_{1\le i\le 2n^2}\left|\gamma_1^{(n)}+\cdots+\gamma_i^{(n)}\right|\ge\varepsilon\right] &\le \frac{Var\left(\gamma_1^{(n)}+\cdots+\gamma_{2n^2}^{(n)}\right)}{\varepsilon^2}\\
&\equiv \frac{\sum_{i=1}^{2n^2} Var\left(\gamma_i^{(n)}\right)}{\varepsilon^2}\\
&= \frac{\sum_{i=1}^{2n^2}\frac{1}{4n^4}}{\varepsilon^2}=\frac{\frac{2n^2}{4n^4}}{\varepsilon^2}=\frac{1}{2n^2\varepsilon^2}.
\end{aligned}$$

Now, by applying the Borel-Cantelli lemma (Lemma 3.11), we can obtain that

$$\lim_{n\to\infty}\max_{1\le i\le 2n^2}\left|\gamma_1^{(n)}+\cdots+\gamma_i^{(n)}-\frac{i}{2n^2}\right|=0,\quad\text{almost surely.}\tag{3.15}$$

Similarly, we can say that

$$\lim_{n\to\infty}\max_{1\le i\le 2n^2}\left|\sigma_1^{(n)}+\cdots+\sigma_i^{(n)}-\frac{i}{2n^2}\right|=0,\quad\text{almost surely.}\tag{3.16}$$

To finish, let $\gamma_0^{(n)}=\sigma_0^{(n)}\equiv 0$. Notice that we have

$$\begin{aligned}
&\left|B^{(n)}(t)-B(t)\right| =\\
=&\max_{1\le i\le 2n^2}\max_{t\in I_i}\left|B^{(n)}(t)-B^{(n)}\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)+\right.\\
&\left.+B^{(n)}\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)-B\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)+B\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)-B(t)\right|\\
\le&\max_{1\le i\le 2n^2}\max_{t\in I_i}\left|B^{(n)}(t)-B^{(n)}\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)\right|+\\
&+\left|B^{(n)}\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)-B\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)\right|+\left|B\left(\sum_{j=1}^{i}\gamma_j^{(n)}\right)-B(t)\right|,
\end{aligned}$$

where $I_i=\left[\frac{i-1}{2n^2},\frac{i}{2n^2}\right]$.

Since the Brownian motion is uniformly continuous, we have that

$$\left| B\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) - B(t) \right| \xrightarrow[i\to\infty]{} 0.$$

Similarly, we can assume that the realization of the uniform transport processes that we have constructed in $B^{(n)}(\cdot)$ is also uniformly continuous. Therefore, we also have that

$$\left| B^{(n)}(t) - B^{(n)}\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) \right| \xrightarrow[i\to\infty]{} 0.$$

Therefore, from the previous inequality, we obtain that

$$\begin{aligned}
\lim_{n\to\infty} \left| B^{(n)}(t) - B(t) \right| &\leq \lim_{n\to\infty} \max_{1\leq i\leq 2n^2} \max_{t\in I_i} \left| B^{(n)}\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) - B\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) \right| \\
&= \lim_{n\to\infty} \max_{1\leq i\leq 2n^2} \left| B^{(n)}\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) - B\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) \right|.
\end{aligned}$$

Then, using (3.10), we can obtain that

$$\lim_{n\to\infty} \max_{1\leq i\leq 2n^2} \left| B^{(n)}\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) - B\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) \right| =$$

$$= \lim_{n\to\infty} \max_{1\leq i\leq 2n^2} \left| B\left(\sum_{j=1}^{i} \sigma_j^{(n)}\right) - B\left(\sum_{j=1}^{i} \gamma_j^{(n)}\right) \right|$$

Finally, using (3.15) and (3.16), we obtain that

$$\lim_{n\to\infty} \max_{0\leq i\leq 2n^2} \left| B\left(\sum_{j=0}^{i} \sigma_j^{(n)}\right) - B\left(\sum_{j=0}^{i} \gamma_j^{(n)}\right) \right| =$$

$$\begin{aligned}
&= \lim_{n\to\infty} \max_{0\leq i\leq 2n^2} \left| B\left(\frac{i}{2n^2}\right) - B\left(\frac{i}{2n^2}\right) \right| \\
&= 0, \quad \text{almost surely.}
\end{aligned}$$

Therefore, we have that

$$\lim_{n\to\infty} \max_{0\leq t\leq 1} \left| B^{(n)}(t) - B(t) \right| = 0, \quad \text{almost surely.}$$

Thus, we have proved the almost sure convergence of the uniform transport processes towards the Brownian motion.

3.2.2 Further Results

During this section we have been proving the almost sure convergence of the uniform transport processes. Our purpose now is to complement this result and see some other results that extend this notion.

To do so, we will comment two results that we can find in the paper *Rate of Convergence of Uniform Transport Processes to Brownian Motion and Application to Stochastic Integrals* by Gorostiza and Griego [4].

The first result is a theorem which states the almost sure convergence of the uniform transport process (which we have already seen) but, furthermore, it adds a result on the rate of convergence of these uniform transport processes towards the Brownian motion.

Therefore, with this, we not only have the notion of convergence but also, in some sense, how fast do these realizations of uniform transport processes, that we have constructed before, converge. The theorem says the following.

Theorem 3.16. *There exist realizations* $\left\{B^{(n)}(t),\ t \geq 0\right\}$ *of the previous uniform transport processes on the same probability space,* $(\Omega, \mathscr{B}, \mathcal{P})$*, as a Brownian motion* $\{B(t),\ t \geq 0\}$*, with* $B(0) \equiv 0$*, so that the following conditions are satisfied:*

(i)
$$\lim_{n\to\infty} \max_{0\leq t\leq 1} \left|B^{(n)}(t) - B(t)\right| = 0, \quad \textit{almost surely.}$$

(ii) For all $q > 0$,
$$\mathcal{P}\left[\max_{0\leq t\leq 1} \left|B^{(n}(t) - B(t)\right| > \frac{\alpha\sqrt{\log^5(n)}}{\sqrt{n}}\right] = o\left(\frac{1}{n^q}\right) \quad \textit{as} \quad n \to \infty,$$

where α *is a positive constant depending on* q*.*

Note that the first condition is the one that we have already proved and the second condition is this rate of convergence, which we are not going to prove.

This is an interesting result but the other one that we want to comment is a more notorious extension of the almost convergence that we have seen. This time, instead of working with a stochastic process and see that it converges towards the Borwnian motion, we are going to work with stochastic integrals and we are going to see that a particular type of stochastic integrals converge almost surely towards an Itô integral.

Before stating and proving this result, let us introduce briefly the Itô calculus, the concept of stochastic integral and, in particular, the Itô integral.

Itô calculus started to be developed in the 1950s by Kyoshi Itô in an attempt to give rigorous meaning to some differential equations driven by the Brownian motion. Roughly speaking, Itô started to develop the analogous of the classical Leibniz and Newton's calculus for stochastic processes.

Despite there had been different generalizations of the Riemann integral in the classical analysis, it was not until Itô's development that a theory of integration of random mappings with respect to nowhere differentiable random integrators was developed.

The stochastic generalization of the Riemann integral that he developed is what we know nowadays as the Itô integral. This is an integral where the integrators and the integrands are stochastic processes:

$$\int_0^t H_s dB_s,$$

where H_s is a locally square-integrable stochastic process adapted to a filtration generated by the Brownian motion, B_s.

Now that we have presented the stochastic integrals we can state the last result that we were talking about.

As we have said, this theorem states the notion of almost sure convergence but, instead of working with uniform transport processes and the Brownian motion, we work with stochastic integrals and the Itô integral. Furthermore, as in Theorem 3.16, we again have a rate of convergence which gives us information regarding the velocity of convergence. The theorem is the following one.

Theorem 3.17. *Let* $\left\{B^{(n)}(t),\ t \geq 0\right\}$ *be a realization of the uniform transport processes we have seen before on the same probability space,* $(\Omega, \mathscr{B}, \mathcal{P})$*, as a Brownian motion* $\{B(t),\ t \geq 0\}$*. Let* $\psi(x,t)$ *be a real-valued function with partial derivatives* $\psi_x(x,t)$ *and* $\psi_t(x,t)$ *for* $-\infty < x < \infty$ *and* $0 \leq t \leq 1$*. Define*

$$I_n \equiv \int_0^1 \psi\left(B^{(n)}(t), t\right) dB^{(n)}(t), \tag{3.17}$$

and

$$I \equiv \int_0^1 \psi\left(B(t), t\right) dB(t) + \frac{1}{2}\int_0^1 \psi_x\left(B(t), t\right) dt, \tag{3.18}$$

where the stochastic integral I *is an Itô integral. Then, the following conditions are satisfied:*

(i)

$$\lim_{n\to\infty} I_n = I, \quad \textit{almost surely.} \tag{3.19}$$

(ii) For all $q > 0$*,*

$$\mathcal{P}\left[|I_n - I| > R_n \alpha_n\right] = o\left(\frac{1}{n^q}\right) \quad \textit{as} \quad n \to \infty,$$

where $\alpha_n = \dfrac{\alpha\sqrt{\log^5(n)}}{\sqrt{n}}$*, with* α *a positive constant depending on* q*, and*

$$R_n = \max_{0\leq t\leq 1} \max_{\left|x - B^{(n)}(t)\right| \leq \alpha_n} |\psi(x,t)| + \max_{0\leq t\leq 1} \max_{\left|x - B^{(n)}(t)\right| \leq \alpha_n} |\psi_t(x,t)|.$$

As in Theorem 3.16, we will not enter into proving the rate of convergence. What we will prove will be the first condition. To do so, we will base our proof on the paper of Wong and Zakai, *On the convergence of ordinary integrals to stochastic integrals* [14].

Proof. Let $F(\lambda,t)=\int_0^\lambda \psi(x,t)dx$. Then, by Taylor's expansion, we have

$$F\left(B^{(n)}(1),1\right)-F\left(B^{(n)}(0),0\right)=\int_0^1 \psi\left(B^{(n)}(t),t\right)dB^{(n)}(t)+\int_0^1 \psi_t\left(B^{(n)}(t),t\right)dt.$$

We have seen that $B^{(n)}(t)$ converges almost surely towards $B(t)$ for all $t\in[0,1]$. We also have seen, while we were constructing this realization of the uniform transport processes, that $B^{(n)}(t)$ is continuous and has bounded variation. Moreover, we have seen, by construction, that $B^{(n)}(t)$ is bounded. Then, using all these properties and using that $\psi(x,t)$ and $\psi_t(x,t)$ are continuous, we obtain that

$$\begin{aligned}&\lim_{n\to\infty}\int_0^1 \psi\left(B^{(n)}(t),t\right)dB^{(n)}(t)=\\ =\ &F\left(B(1),1\right)-F\left(B(0),0\right)-\int_0^1 \psi_t\left(B(t),t\right)dt,\ \text{almost surely.}\end{aligned} \tag{3.20}$$

A result of Itô in [6] states that, if the function $G(\xi,t)$ has a continuous first partial derivative with respect to t and a continuous second partial derivative with respect to ξ, for $-\infty<\xi<\infty$ and $a\le t\le b$, and if the random functions $f(t)$ and $g(t)$, for $a\le t\le b$, are independent of the aggregate differences $B(s)-B(t)$, $t\le s\le b$, then $f(t)\in L^2[0,1]$, $f(t)\in L^1[0,1]$ *a.s.* and

$$dz(t)=g(t)dt+f(t)dB(t).$$

Then, almost surely,

$$\begin{aligned}&G(z(b),b)-G(z(a),a)=\\ =\ &\int_a^b G_t(z(t),t)f(t)dB(t)+\int_a^b G_z(z(t),t)g(t)+G_t(z(t),t)+\frac{1}{2}f^2(t)G_{z^2}(z(t),t)\ dt.\end{aligned}$$

This result is what we call the Itô's Formula or Itô's Lemma. Applying this result to F, and taking $0\le t\le 1$, we obtain:

$$\begin{aligned}&F\left(B(1),1\right)-F\left(B(0),0\right)=\\ =\ &\int_0^1 \psi\left(B(t),t\right)dB(t)+\int_0^1 F_t\left(B(t),t\right)+\frac{1}{2}\psi_B\left(B(t),t\right)\ dt,\quad \text{almost surely.}\end{aligned} \tag{3.21}$$

Then, (3.19) just follows by substituting (3.21) into (3.20).

□

Appendices

Appendix A Other results on Weak Convergence

In this section we will see different results on weak convergence which consist in proving weak convergence by showing that $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for some special class of sets A.

The theorem that we want to prove is the following one.

Theorem A.1. *Let $\mathscr{U}$ be a subclass of $\mathscr{P}$ such that*

(i) $\mathscr{U}$ is closed under the formation of finite intersections.

(ii) Each open set in S is a finite or countable union of elements of $\mathscr{U}$.

Thus, if $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for every A in $\mathscr{U}$, then $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Proof. If $A_1, \ldots, A_m$ lie in $\mathscr{U}$, then their intersections also lie in $\mathscr{U}$. Therefore, by the inclusion-exclusion formula,

$$\begin{aligned}\mathbb{P}_n\left(\bigcup_{i=1}^m A_i\right) &= \sum_i \mathbb{P}_n(A_i) - \sum_{i,j} \mathbb{P}_n(A_iA_j) + \sum_{i,j,k} \mathbb{P}_n(A_iA_jA_k) - \cdots \\ &\to \sum_i \mathbb{P}(A_i) - \sum_{i,j} \mathbb{P}(A_iA_j) + \sum_{i,j,k} \mathbb{P}(A_iA_jA_k) - \cdots \\ &= \mathbb{P}\left(\bigcup_{i=1}^m A_i\right).\end{aligned}$$

Now, if G is open, we can express it as $G = \bigcup_i A_i$ for some sequence $\{A_i\}$ of elements of $\mathscr{U}$.

Then, given ε, we choose m so that $\mathbb{P}\left(\bigcup_{i \le m} A_i\right) > \mathbb{P}(G) - \varepsilon$.

Hence, because of the relation we have just seen,

$$\mathbb{P}(G) - \varepsilon < \mathbb{P}\left(\bigcup_{i\le m} A_i\right) = \limsup_n \mathbb{P}_n\left(\bigcup_{i\le m} A_i\right) \le \liminf_n \mathbb{P}_n(G).$$

Since ε was taken arbitrary, the condition (iv) of Theorem 2.4 is satisfied and hence we have weak convergence.

□

To continue, let $S(x, \varepsilon)$ denote the ε-sphere (open) with x as centre. Then, we have the following corollary of the previous theorem.

Corollary A.2. *Let $\mathscr{U}$ be a class of sets such that*

(i) $\mathscr{U}$ is closed under formation of finite intersections.

(ii) For every x in S and every $\varepsilon > 0$ there is an A in $\mathscr{U}$ such that

$$\mathring{A} \subset A \subset S(x, \varepsilon).$$

Thus,if S is separable and $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for every A in $\mathscr{U}$, then $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Proof. The condition (ii) implies that, for each $x \in G$, with G open, we have that $x \in \mathring{A} \subset A \subset G$ for some A in $\mathscr{U}$.

Now, since S is separable, there exists a sequence (finite or infinite) $\{A_i\}$ in $\mathscr{U}$ such that $G \subset \bigcup_i \mathring{A}_i$ and $A_i \subset G$. This implies that $G = \bigcup_i A_i$.

Therefore, $\mathscr{U}$ satisfies the hypothesis in the previous theorem and we obtain the result we were looking for.

□

Appendix B Compactness

In this section we are going to prove a technical lemma involving totally bounded sets, compactness and completeness. Let S be a metric space with the metric defined by (2.1). The lemma is the next one.

Theorem B.1 (**Theorem 2.7**). *For an arbitrary set A in S these three conditions are equivalent:*

(i) *$\bar{A}$ is compact.*

(ii) *Each sequence in A has a convergent subsequence, the limit of which necessarily lies in $\bar{A}$.*

(iii) *A is totally bounded*[9] *and $\bar{A}$ is complete.*

Proof. Notice that (ii) holds if and only if each sequence in $\bar{A}$ has a subsequence converging to a point in $\bar{A}$.

Also, A is totally bounded if and only if $\bar{A}$ is totally bounded.

Therefore, we can assume that $A = \bar{A}$, i.e. A is closed.

To make the proof clearer we introduce three more properties between (i) and (ii):

$(i1)$ Each countable open cover of A has a finite subcover.

$(i2)$ If $A \subset \bigcup_n U_n$ where the $U_n' s$ are open and $U_1 \subset U_2 \subset \cdots$, then $A \subset U_n$ for some n.

$(i3)$ If $A \supset C_1 \supset C_2 \supset \cdots$, where the C_n's are closed and nonempty, then $\bigcap_n C_n$ is nonempty.

We prove first that $(i1)$, $(i2)$, $(i3)$, (ii), (iii) are all equivalent.

That $(i1)$ implies $(i2)$ is clear. For the inverse implication we can assume $(i2)$ and assume that $\{U_n\}$ covers A. Replacing the U_n by $\bigcup_{k\leq n} U_k$ we have that $A \subset \bigcup_{k\leq n} U_k$ for some n and hence, $(i1)$.

Now, the equivalence between $(i2)$ and $(i3)$.

First, $(i2)$ says that $A \cap U_n \uparrow A$ implies that $A \cap U_n = A$ for some n.

On the other hand, $(i3)$ says that $A \cap C_n \downarrow \emptyset$ implies that $A \cap C_n = \emptyset$ for some n (C_n need not be contained in A).

If we take $C_n = U_n^c$, both statements say the same, so they are equivalent.

Then, we want to prove that $(i3)$ and (ii) are equivalent.

[9] A set A in a metric space S is **totally bounded** if there is a finite collection of open spheres with radius ε such that its centers are in A and the union of these balls contains A.

Assume first $(i3)$. If $\{x_n\}$ is a sequence in A, take $B_n = \{x_n, x_{n+1}, \ldots\}$ and $C_n = \bar{B}_n$. Each C_n is nonempty. Hence, since $(i3)$ holds by assumption, $\bigcap_n C_n$ contains some x. Since $x \in \bar{B}_n$, there is an $i_n \geq n$ such that $\rho(x, x_{i_n}) < \frac{1}{n}$. We choose these i_n's inductively so that $i_1 < i_2 < \cdots$. Then, $\lim_n \rho(x, x_{i_n}) = 0$, i.e., (ii) holds.

On the other hand, if C_n are decreasing, nonempty closed sets and (ii) holds, we take $x_n \in C_n$ and let x be the limit of a subsequence. Then, $x \in \bigcap_n C_n$. Hence, $(i3)$ holds.

Finally, we prove the equivalence between (ii) and (iii).

Assume first (ii).

If A is not totally bounded, then there exists a positive ε and an infinite sequence $\{x_n\}$ in A such that $\rho(x_m, x_n) \geq \varepsilon$ for $m \neq n$. But then $\{x_n\}$ contains no convergent subsequence, and therefore (ii) implies total boundedness. Also (ii) implies completeness because, if $\{x_n\}$ is Cauchy and has a subsequence converging to x, then the entire sequence converges to x.

On the other hand, assume (iii). Since A is totally bounded, it can be covered, for each n, by finitely many open balls $B_{n1}, \ldots, B_{nk_n}$ of radius $\frac{1}{n}$.

Let $\{x_m\}$ be a sequence in A. We choose $m_{11}, m_{12}, \ldots$ integers in such a way that $x_{m_{11}}, x_{m_{12}}, \ldots$ lie in the same B_{1k}. This choice is possible because there are finitely many of these balls.

Then, we choose $m_{21}, m_{22}, \ldots$ a subsequence of $m_{11}, m_{12}, \ldots$ in such a way that $x_{m_{21}}, x_{m_{22}}, \ldots$ all lie in the same B_{2k}. We keep doing this process.

Thus, if we call $r_i = m_{ii}$, then $x_{r_n}, x_{r_{n+1}}, \ldots$ all lie in the same B_{nk}.

Hence, it follows that $x_{r_1}, x_{r_2}, \ldots$ is Cauchy and, therefore, by completeness, it converges to some point of A.

We have seen $(i1), (i2), (i3), (ii), (iii)$ are equivalent. Now we just want to see that (i) and (iii) are equivalent.

Clearly (i) implies $(i1)$ because $\bar{A}$ is compact if each open cover has a finite subcover and, hence, each countable open cover too.

Therefore, we have that (i) implies (iii). It just remains to prove the other implication. We will see that $(i1)$ and (iii) imply (i).

If A is totally bounded, then it is clearly separable, and it follows by the Lindelöf property[10] that an arbitrary open cover of A has a countable subcover. Then, by $(i1)$, there is a further subcover that is finite.

□

[10] **Lindelöf property:** A Lindelöf space is a topological space in which every open cover has a countable subcover.

Appendix C Product Spaces

In this section we will prove a couple of results of weak convergence on product spaces that we use in the section on Donsker's Theorem.

First of all, let $S = S' \times S''$ be the product of two metric spaces S' and S''.

If S is separable, notice that this requires S' and S'' to be separable, then $\mathscr{P}$, $\mathscr{P}'$, $\mathscr{P}''$, which are the σ-fields of Borel sets of S, S', S'' respectively, are related by

$$\mathscr{P} = \mathscr{P}' \times \mathscr{P}\text{''}.$$

If we get a probability measure $\mathbb{P}$ on $(S, \mathscr{P})$, its two marginal distributions are defined by

$$\begin{cases} \mathbb{P}'(A') = \mathbb{P}(A' \times S''), & A' \in \mathscr{P}' \\ \\ \mathbb{P}''(A'') = \mathbb{P}(S' \times A''), & A'' \in \mathscr{P}''. \end{cases}$$

The first result that we are going to see is the following theorem.

Theorem C.1. *If S is separable then a necessary and sufficient condition for weak convergence, $\mathbb{P}_n \Rightarrow \mathbb{P}$, is that*

$$\mathbb{P}_n(A' \times A'') \longrightarrow \mathbb{P}(A' \times A'')$$

for each $\mathbb{P}'$-continuity set A' and each $\mathbb{P}''$-continuity set A'', where $\mathbb{P}$ and $\mathbb{P}_n$ are probability measures on $(S, \mathscr{P})$ and $\mathbb{P}'$, $\mathbb{P}''$ are the marginal distributions of $\mathbb{P}$.

Proof. First of all, let us denote by $\partial, \partial', \partial''$ the boundary operators of S, S', S'' respectively.

Since

$$\partial(A' \times A'') \subset ((\partial' A') \times S'') \cup (S' \times (\partial'' A'')),$$

the condition is necessary.

To prove sufficiency we use the previous corollary on weak convergence to the class $\mathscr{U}$ of sets $A' \times A''$ with A' a $\mathbb{P}'$-continuity set and A'' a $\mathbb{P}''$-continuity set.

$\mathscr{U}$ is closed under the formation of finite intersections and, by hypothesis, we have that $\mathbb{P}_n(A) \to \mathbb{P}(A)$ for A in $\mathscr{U}$.

Let $(x', x'') \in S$ and $\varepsilon > 0$. Let us consider the sets

$$A_\delta = \{y' \colon \rho'(x', y') < \delta\} \times \{y'' \colon \rho''(x'', y'') < \delta\}.$$

Notice that, for distinct δ, the sets $\partial' \{y' \colon \rho'(x', y') < \delta\}$ are disjoint and the sets $\partial'' \{y'' \colon \rho''(x'', y'') < \delta\}$ are also disjoint. Therefore, A_δ lies in $\mathscr{U}$ for some $0 < \delta < \varepsilon$.

Moreover, if we define the following metric for S,

$$\rho\left((x', x''), (y', y'')\right) = \max\left\{\rho'(x', y'), \rho''(x'', y'')\right\},$$

then we have that A_δ is the sphere with centre (x', x'') and radius δ.

Thus, the hypothesis of the corollary are satisfied and we obtain the sufficiency.

$\square$

Now, for $\mathbb{P}', \mathbb{P}''$ probability measures on $(S', \mathscr{P}')$ and $(S'', \mathscr{P}'')$ respectively, the product measure $\mathbb{P}' \times \mathbb{P}''$ is a probability measure on $\mathscr{P}' \times \mathscr{P}''$. Hence, if S is separable, it is a probability measure on $\mathscr{P}$ too. With this, the following theorem is consequence of the previous one.

Theorem C.2 (**Theorem 2.30**). *Let $\mathbb{P}'_n$ and $\mathbb{P}'$ be probability measures on $(S', \mathscr{P}')$ and $\mathbb{P}''_n$ and $\mathbb{P}''$ be probability measures on $(S'', \mathscr{P}'')$. If $S = S' \times S''$ is separable, then $\mathbb{P}'_n \times \mathbb{P}''_n \Rightarrow \mathbb{P}' \times \mathbb{P}''$ if and only if $\mathbb{P}'_n \Rightarrow \mathbb{P}'$ and $\mathbb{P}''_n \Rightarrow \mathbb{P}''$.*

Appendix D Prokhorov's Theorem

Let Π be a family of probability measures on $(S, \mathscr{P})$.

We will say that Π is **relatively compact** if every sequence of elements of Π contains a weakly convergent sequence in the sense that for every sequence $\{\mathbb{P}_n\}$ in Π there is a subsequence $\{\mathbb{P}_{n_k}\}$ and a probability measure $\mathcal{Q}$ on $(S, \mathscr{P})$ (but not necessarily in Π) such that $\mathbb{P}_n \Rightarrow \mathcal{Q}$.

We say that the family Π of probability measures on the metric space S is **tight** if for every $\varepsilon > 0$ there exists a compact set K such that $\mathbb{P}(K) > 1 - \varepsilon$ for all $\mathbb{P} \in \Pi$.

We will refer to the two following theorems as **Prokhorov's Theorem**.

Theorem D.1. *If Π is tight then it is relatively compact.*

Theorem D.2. *Suppose S is separable and complete. If Π is relatively compact then it is tight.*

Appendix E Markov Chains and Markov Processes

In this section we are going to introduce briefly Markov chains and Markov processes.

A Markov chain is a discrete time stochastic process, let us say $\{X_n\}_{n\geq 1}$, that has two particularities. The first one is that it has no "memory", in the sense of the future not depending on what the process did in the past but only on where the process is at the present. The second characteristic is that such processes can only have a finite or numerable number of values.

Note that we will call I the *set of states* of the process and each $i \in I$ will be a *state* of the process.

These processes can be represented using diagrams. For example, let us represent a process with $I = \{a, b, c\}$ and that follows the next rules:

- If it is on b it moves from b to a with probability 1.
- If it is on a, it moves to c with probability $\frac{1}{3}$ or it stays in a with probability $\frac{2}{3}$.
- If it is on c, it moves to a or to b with probability $\frac{1}{2}$.

The diagram corresponding to this process is the following one.

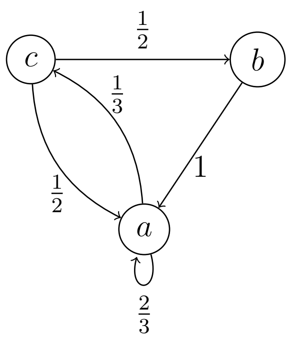

Figure 6: Example of Markov Chain.

To continue this introduction to Markov chains we define what a stochastic matrix is.

Definition E.1. *Let I be the set of states of a Markov chain and $P = (p_{ij})_{(i,j)\in I\times I}$ be a matrix. We will say that P is stochastic if and only if,*

(i) $p_{ij} \in [0, 1]$ *for all* i, j.

(ii) $\sum_{j\in I} p_{ij} = 1$ *for every* $i \in I$.

Notice that we have a bijection between the diagrams of the processes and the stochastic matrices. Therefore, we can represent those diagrams using matrices. For example, if we take the diagram in Figure 6 but setting $a = 0$, $b = 1$ and $c = 2$, we will obtain the following stochastic matrix:

$$\begin{pmatrix} \frac{2}{3} & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 \end{pmatrix}$$

With that being said, we can define what homogeneous Markov chains are.

Definition E.2. *A stochastic process* $\{X_n,\ n \geq 0\}$ *which takes values in a set of states* I *is an* ***homogeneous Markov chain*** *with initial distribution* ν *and with transition probabilities matrix* $P = (p_{ij})_{(i,j)\in I\times I}$ *if*

(i) X_0 *has distribution* ν.

(ii) For every $i_0, \ldots, i_{n+1} \in I$ *and* $n \geq 0$,

$$\mathbb{P}\left[X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n\right] = \mathbb{P}\left[X_{n+1} = i_{n+1} | X_n = i_n\right] = p_{i_n, i_{n+1}}.$$

Notice that this first equality on the second condition,

$$\mathbb{P}\left[X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n\right] = \mathbb{P}\left[X_{n+1} = i_{n+1} | X_n = i_n\right],$$

is what we call the **Markov property**. Basically it says that the process just depends on the present and not on the past. Note also that the second equality shows homogeneity in the sense that the probability does not depend on n.

This was for a discrete-time process, but we can consider processes on continuous time. Then, to study them, we define **Markov processes**.

Notice that, in this case, instead of taking values on a set of states, we take values on a *space of states*, E.

Definition E.3. *A stochastic process* $\{Y(t),\ t \in [0, T]\}$ *taking values in a space of states* E*, numerable, is a* ***Markov process*** *if for every* $s, t \geq 0$ *and* $j \in E$,

$$\mathbb{P}\left[Y(t+s) = j \,|Y(u), u \leq t\right] = \mathbb{P}\left[Y(t+s) = j \,|Y(t)\right].$$

Furthermore, if for every $i, j \in E$ *and* $s, t \geq 0$,

$$p_s(i,j) \coloneqq \mathbb{P}\left[Y(t+s) = j \,|Y(t) = i\right]$$

is independent of $t \geq 0$*, we will say that* Y *is a time-****homogeneous Markov process****.*

The function

$$t \longmapsto p_t(i,j),$$

for fixed $i, j \in E$ is called a **transition probability** and the family of matrices

$$P_t = \left(p_t(i,j)\right)_{(i,j)\in E\times E}$$

is the **transition function** of the Markov process.

Note that, when we have a homogeneous Markov process, we can say that the transition probabilities related to the process are *stationary*.

The homogeneous Markov processes satisfy the following properties.

Proposition E.4. *Let* $\{Y(t),\ t \geq 0\}$ *be a Markov process homogeneous on time. Then, for all* $i, j \in E$ *and for all* $s, t \geq 0$*, the following properties are satisfied:*

(i) $p_t(i,j) \geq 0$.

(ii) $\sum_{k \in E} p_t(i,k) = 1$.

(iii) $\sum_{k \in E} p_t(i,k) p_s(k,j) = p_{t+s}(i,j)$.

Remark E.5. *The last condition is known as the Chapman-Kolmogorov equation and, if we write it on matrix notation, we obtain* $P_{t+s} = P_t P_s$.

Proof. The first and the second property are obvious because of the way we have defined $p_s(i,j) := \mathbb{P}[Y(t+s) = j \,|Y(t) = i]$. Therefore, we just have to prove the third one. Using the Markov property we obtain:

$$\begin{aligned}
\sum_{k \in E} p_t(i,k) p_s(k,j) &= \sum_{k \in E} \mathbb{P}[Y(t) = k \,|Y(0) = i]\, \mathbb{P}[Y(t+s) = j \,|Y(t) = k] \\
&= \sum_{k \in E} \Big(\mathbb{P}[Y(t+s) = j \,|Y(t) = k, Y(0) = i] \times \\
&\qquad \times \mathbb{P}[Y(t) = k \,|Y(0) = i] \Big) \\
&= \sum_{k \in E} \left(\frac{\mathbb{P}[Y(t+s) = j, Y(t) = k, Y(0) = i]}{\mathbb{P}[Y(t) = k, Y(0) = i]} \times \right. \\
&\qquad \left. \times \frac{\mathbb{P}[Y(t) = k \,|Y(0) = i]}{\mathbb{P}[Y(0) = i]} \right) \\
&= \sum_{k \in E} \frac{\mathbb{P}[Y(t+s) = j, Y(t) = k, Y(0) = i]}{\mathbb{P}[Y(0) = i]} \\
&= \sum_{k \in E} \mathbb{P}[Y(t+s) = j, Y(t) = k \,|Y(0) = i] \\
&= \mathbb{P}[Y(t+s) = j \,|Y(0) = i] = p_{t+s}(i,j).
\end{aligned}$$

□

Furthermore, since the trajectories of the Markov processes are really irregular, we will consider processes satisfying one more property. This property is a condition of continuity on the transition function:

$$\lim_{t \to 0} p_t(i,j) = \delta_{i,j},$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$.

To finish this introduction, the last thing that we want to define on Markov processes is the *infinitesimal generator*. To do so, we first have to state the following proposition.

Proposition E.6.

(i) For all $i \in E$,

$$\lim_{t \downarrow 0} \frac{1 - p_t(i,i)}{t} =: q_i \in [0, \infty].$$

(ii) For all $i, j \in E$,

$$\lim_{t \downarrow 0} \frac{p_t(i,j)}{t} =: q_{i,j} < \infty.$$

We will not enter into proving this result, we can find a proof for this and all the reasoning behind in [11].

What we will pay attention to is to the fact that, since $p_0(i,i) = 1$, the condition (i) in Proposition E.6 tells us that the following limit exists:

$$\lim_{t \downarrow 0} \frac{p_t(i,i) - p_0(i,i)}{t} =: -q_i,$$

i.e., $p_t(i,i)$ has right derivative at $t = 0$. On the other hand, since $p_0(i,j) = 0$, the condition (ii) in Proposition E.6 tells us that $p_t(i,j)$ has right derivative at $t = 0$ and this derivative is $q_{i,j}$.

Therefore, we can compute the right-hand derivative of the matrix P_t, which lets us define the following concept.

Definition E.7. *The matrix*

$$A := \lim_{t \downarrow 0} \frac{P_t - I}{t} = \frac{d}{dt}(P_t)\bigg|_{t=0^+}$$

is called the ***infinitesimal generator*** *of the Markov process on continuous time.*

Remark E.8. *If we assume* $E = \mathbb{N}$*, the infinitesimal generator is of the form*

$$A = \begin{pmatrix} -q_0 & q_{0,1} & q_{0,2} & \cdot & \cdot & \cdot & \cdot \\ q_{1,0} & -q_1 & q_{1,2} & \cdot & \cdot & \cdot & \cdot \\ q_{2,0} & q_{2,1} & -q_2 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Finally, let us see how a Markov process looks like. We consider $E = \{a, b, c, d, e\}$. Then, the trajectory of a Markov process can be:

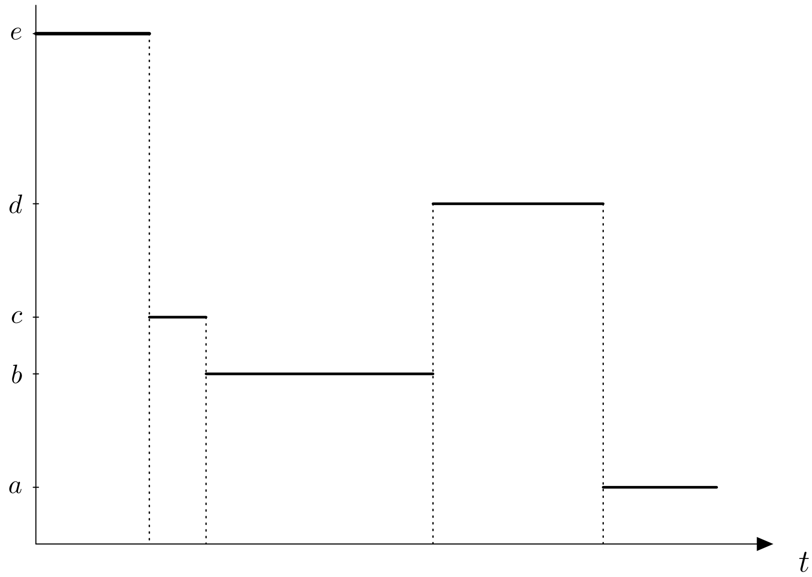

Figure 7: Trajectory of a Markov process.

Looking at this trajectory we can see that, despite the fact that this process is not increasing, it looks kind of like a Poisson process. In fact, we can see that the Poisson is a Markov process. That is because, since the Poisson process ($\{N(t),\ t \geq 0\}$) has independent increments, we can check the Markov property:

$$\mathbb{P}\left[N(t+s) = j \,|N(u), u \leq t\right] = \mathbb{P}\left[N(t+s) = j \,|N(t)\right],$$

and therefore, it is a Markov process.

References

[1] Xavier Bardina, Giulia Binotto, and Carles Rovira. "The Complex Brownian Motion as a Strong Limit of Processes Constructed from a Poisson Process". eng. In: *Journal of mathematical analysis and applications* 444.1 (2016), pp. 700–720. ISSN: 0022-247X.

[2] Patrick Billingsley. *Convergence of Probability Measures*. eng. Wiley series in probability and mathematical statistics. New York [etc: Wiley, 1968. ISBN: 0471072427.

[3] William Feller. *An Introduction to Probability Theory and its Applications*. eng. Wiley publications in statistics. New York: Wiley, 1952.

[4] Luis G. Gorostiza and Richard J. Griego. "Rate of Convergence of Uniform Transport Processes to Brownian Motion and Application to Stochastic Integrals". In: *Stochastics* 3.1-4 (1980), pp. 291–303.

[5] Richard J. Griego, David Heath, and Alberto Ruiz-Moncayo. "Almost Sure Convergence of Uniform Transport Processes to Brownian Motion". eng. In: *The Annals of mathematical statistics* 42.3 (1971), pp. 1129–1131. ISSN: 0003-4851.

[6] Kiyosi Itô. "On a Formula Concerning Stochastic Differentials". eng. In: *Nagoya mathematical journal* 3 (1951), pp. 55–65. ISSN: 0027-7630.

[7] Mark Kac. "A Stochastic Model Related to the Telegrapher's Equation". eng. In: *The Rocky Mountain journal of mathematics* 4.3 (1974), pp. 497–509. ISSN: 0035-7596.

[8] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, 2016.

[9] Paul Lévy. "Processus Stochastiques et Mouvement Brownien". In: *MR29120* 25 (1948).

[10] David Nualart and Marta Sanz-Solé. *Curs de Probabilitats*. cat. Estadística y análisis de datos ; 5. Barcelona: PPU, 1990. ISBN: 8476657188.

[11] Carles Rovira Escofet. *Processos Estocàstics: Un Curs Bàsic*. cat. 2021-06-09T11:17:54Z.

[12] Anatoliy V. Skorokhod. *Studies in the Theory of Random Processes*. eng. Adiwes international series in mathematics ; 7021. Reading (Mass.): Addison-Wesley, 1965.

[13] Daniel W. Stroock. "Lectures on Topics in Stochastic Differential Equations : lectures delivered at the Indian Institute of Science, Bangalore under the T.I.F.R.-I.I.Sc. Programme in applications of mathematics". eng. In: Lectures on mathematics and physics. Mathematics ; 68 (1982).

[14] Eugene Wong and Moshe Zakai. "On the Convergence of Ordinary Integrals to Stochastic Integrals". eng. In: *The Annals of mathematical statistics* 36.5 (1965), pp. 1560–1564. ISSN: 0003-4851.