UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Matematica Tullio Levi Civita

Scuola di dottorato di ricerca in matematica

Curriculum: Matematica computazionale

Ciclo: XXIII

Tesi redatta con il contributo finanziario della Fondazione Cariparo

Statistics and stochastic modeling of complex systems:

upscaling of biodiversity and random walks in random environment

Coordinatore: Prof. Martino Bardi

Supervisore: Prof. Marco Formentin

Co-supervisore: Prof. Alessandra Bianchi

Dottorando: Samuele Stivanello

Abstract

In this thesis we study two problems, of both theoretical and applicative nature, arising in the context of statistics and stochastic modeling of complex systems.

The first research line deals with the upscaling problem, which is a classical problem in statistics and have applications ranging from ecology to linguistics. The goal is to extract information at global scale starting from partial knowledge of the considered datasets. The proposed method exploits the form-invariance property of the Negative Binomial, the selected distribution to describe the frequency of frequencies.
The novelties presented in this part are essentially three. Firstly, we extend the parameters' range of the Negative Binomial, allowing the distribution to display a Power Law behavior of its tail. Secondly, we derive an estimator that, besides inferring the global number of types in a database given a small random sample of it, also estimates how the frequencies of the types change across scales. Lastly, we generalize our method to accommodate the cases where we dispose only of presence/absence information of the types, without knowing their abundances in the sample.

The second research line focuses on the asymptotic analysis of a random dynamics in a disordered environment. The model can be described as a random walk with i.i.d. heavy-tailed increments moving on an environment provided by a renewal point process on the real line with i.i.d. heavy-tailed inter-distances. The interest in the model is twofold: on one hand it extends the set of random walks on point processes to the case in which the law of the random walk and the law of the points of the environment have infinite variance or even infinite mean; on the other hand it provides a discrete-time generalization of the Lévy-Lorentz gas, a process that well describes the motion in inhomogeneous media with anomalous behavior. Our model displays a super-diffusive behavior that can be tuned suitably through the choice of the parameters of the two different heavy-tailed laws.
For all combinations of the parameters, we prove the annealed functional limit theorem for the suitably rescaled process, relative to the optimal Skorokhod topology in each case. When the limit process is not càdlàg, we prove convergence of the finite-dimensional distributions. When the limit process is deterministic, we also prove a limit theorem for the fluctuations, again relative to the optimal Skorokhod topology.

Sommario

In questa tesi studiamo due problemi, sia di natura teorica sia con risvolti applicativi, che emergono dal contesto statistico e di modellizzazione stocastica dei sistemi complessi.

La prima linea di ricerca riguarda il problema di upscaling, un problema classico in statistica con applicazioni che spaziano dall'ecologia all'analisi del linguaggio. L'obiettivo è quello di estrarre informazioni su scala globale a partire da una conoscenza parziale dei datasets a disposizione. Il metodo proposto sfrutta la proprietà di invarianza di forma della Binomiale Negativa, la distribuzione scelta per descrivere l'andamento delle frequenze di frequenze.
Le novità presentate in questa parte sono essenzialmente tre. Per iniziare, abbiamo esteso il range dei parametri della Binomiale Negativa; in questo modo le code della distribuzione decadono secondo una legge di potenza. Poi, abbiamo costruito uno stimatore che, oltre a predire il numero totale di specie diverse nel database a partire da un campione casuale, riesce anche a stimarne la variazione delle abbondanze in relazione alla scala. Per finire, abbiamo generalizzato il nostro metodo al caso in cui siano disponibili solo informazioni sulla presenza o assenza delle specie nei vari sotto-campioni, senza conoscere la loro abbondanza.

La seconda linea di ricerca si concentra sull'analisi asintotica di dinamiche stocastiche in ambiente disordinato. Il modello che consideriamo può essere descritto come una passeggiata aleatoria, i cui incrementi sono i.i.d. e presentano code pesanti, che si muove su punti della retta real generati da un processo di rinnovamento tale che le distanze tra punti vicini sono i.i.d. con code pensanti. L'interesse nel modello è duplice: da un lato estende l'insieme di passeggiate aleatorie su processi di punto al caso in cui sia la legge della passeggiata aleatoria sia quella del punti dell'ambiente hanno varianza infinita o anche media infinita; dall'altro si può considerare come una generalizzazione a tempo discreto del Lévy-Lorentz gas, un processo che ben descrive il moto in ambiente disomogeneo con comportamento anomalo.
Il nostro modello mostra un comportamento super-diffusivo che può essere opportunamente regolato attraverso la scelta dei parametri delle due leggi che lo governano. A seconda delle varie combinazioni dei parametri, abbiamo dimostrato teoremi limite per il processo opportunamente riscalato, relativi alla topologia di Skorokhod ottimale a seconda dei casi. Quando il processo limite non è càdlàg, abbiamo dimostrato la convergenza delle distribuzioni finito-dimensionali. Quando invece il processo limite è deterministico, abbiamo provato anche un teorema limite per le fluttuazioni, sempre rispetto alla topologia di Skorokhod ottimale.

Ackowledgements

First of all, I want to thank the Mathematics Department *Tullio Levi-Civita* of the University of Padova, where I have been studying for eight years, for giving me the opportunity of an important personal and intellectual development. Apart from the technical tools that I have learned, it has been an exceptional lesson on the importance of hard work and motivation to succeed in the professional environment.

I would sincerely like to express my great appreciation to my advisors Prof. Marco Formentin and Prof. Alessandra Bianchi, who have welcomed me in my research journey and involved me in interesting and important research projects. Their solid mentorship has been incredibly important for me, and it is impossible to describe the amount of time, effort and resources they dedicated to me. Their precious support and continuous encouragement have gone well beyond the academic standards, and I am particularly grateful for their copious and priceless advice. They really took care of my personal and professional growth.

Moreover, I want to thank all the coauthors for the great time spent working together. I really appreciated the chance to work with smart, talented and experienced researchers as they are, and the possibility to interact with different approaches and with scientists from a different background.

My special thanks go to all the friends and colleagues with whom I had the pleasure to share the office and the academic ride. The time spent together has been one of the best parts of the Ph.D. life. I thank also all the people I met during conferences, schools and workshops for the insightful discussions that were great opportunities to deepen my knowledge and constituted a source of inspiration in my work.

Finally, I would like to thank my family for having supported and encouraged me in every step I took. Last, but not least, a special thanks goes to Alice, who plays a fundamental role in my life.

Introduction

The last decades have been witnessing a great development in probability theory, mainly thanks to the incredible adaptability of this branch of mathematics to different fields of science. From a little subclass of measure theory, probability gained its own status, being the ideal bridge between pure mathematics and application to various areas, ranging from statistical physics, chemistry, and computer science to economics, finance, and biology. This interlacement of different research areas represented a stimulating boost that produced new probabilistic methods borrowed and adapted from other branches of mathematics. This mutual intertwining is responsible for increasing interest in the field. The mathematical models we developed and exploited in this thesis find inspiration from specific physical problems. Randomness is of course a natural feature of such models. It is fundamental to ensure theoretical properties of the system when dealing with random samples. On the other hand, when dealing with a scaling limit, it allows to capture the cumulative effect of microscopic mechanism.
Usually a system is not completely observable, and even when it is, the computation would be forbidding and hardly treatable if we consider all the variables that enter the description. Hence the mathematical analysis requires the development of new powerful probabilistic ideas, combined with inputs from other branches of mathematics. Some of the techniques we presented here are ready to use, finding direct application to a large class of datasets; whereas some other results, whilst bringing powerful insights, serve as a more general descriptions of complex systems and would require a more detailed characterization to take practical advantages in real world problems.

This thesis is about two different works both dealing with the analysis of complex systems using probabilistic methods. The aim of this framework is to study regular patterns emerging at large scales. The meaning of the word *large* is twofold.
As first, we can imagine a system consisting of a large number of coexisting particles, possibly interacting between each other. Hence the term *large* is referred to the cardinality of the ensemble. A usual line of investigation is on the microscopic interacting mechanism leading to the macroscopical behavior arising only for large system size. Here we focus on another crucial aspect: given a relatively small sample of the particles, we infer information on the global distribution, i.e. on the distribution at the large scale. In particular we aim at upscaling information concerning biodiversity. If we split the particles into different types, we want to predict the total of existing types living at the global scale, and the abundance of each of them.

On the other end, we can imagine a sequence of stochastic processes of scale index N. For each fixed N we are able to give a precise description of the underlying principles driving the motion. However, the resulting equations are often too detailed and hard to handle. The stochastic-process limit, obtained as an approximation of large scale N, strip away unessential details and reveals key features of the motion. The term *large* refers here to the scale index, and we focus on the scaling dynamics that allows regularities to emerge. The results presented here will focus on stochastic process limits with nonstandard scaling and nonstandard limit process. Analyzing these stochastic models, we aim at establishing rigorous results on questions of fundamental nature, like scaling limit and random fluctuations around a law of large numbers limit.
Hence, probability theory aims at explaining statistical regularity associated with a macroscopic view of uncertainty. We will analyze the microscopic scale, be it a local sample of data from a huge dataset or the key components that drive the dynamics of stochastic processes, to characterize laws acting on a large scale.

Upscaling statistical patterns from reduced partial information of big datasets in natural and social sciences

The first part of the thesis is devoted to present the results of two works both linked with the upscaling problem, i.e. how to infer total biodiversity from the observation of local samples.
The origin of this problem dates back to the early 1940s, when the British naturalist and chemist Alexander Steven Corbet spent a couple of years in Malaya studying and counting different species of butterflies [Cor41]. He wondered how many different species existed in the region, given the observations he had. In particular, he noted down both the number of different species and the number of butterflies of the same species he observed. Day by day, increasing the sample, he observed an increasing number of butterflies belonging to popular species, and also some single butterflies of new species. Back to England, its colleague Ronald Aylmer Fisher tried to estimate how many new species he would have observed in a new travel to Malaya. The famous father of statistics was only the first mathematician to tackle down this problem of species estimation at those days [FCW43].
Since then, it has found huge application in different scientific fields, from ecology [CC94; BF93; CB02] to bioscience [LL16; Hug+01; ILL09], leading to the development of a myriad of estimators that try to cover all the different faces of this huge framework [GT56; BMW03; CMC04; MC05; WL05; SH08; Kun+18; OSW16; CC16].
The challenge of biodiversity upscaling has attracted increasing interest in recent years, producing a wide range of competing approaches. Such methods, if successful, could have important applications to multi-scale biodiversity estimation and monitoring.

To start introducing the framework, we highlight from the beginning we are moving towards a parametric estimator, leaving out every Bayesian consideration. This means that we do not consider any a-priori knowledge on parameters of the distribution, even if

we will actually assume to have a very specific parametric distribution. The explanation is straightforward once we agree on the assumptions. We will see how some empirical quantities, like the Relative Species Abundance (RSA), show a similar behavior in diverse contexts. Hence we will impose that the object in question will follow a particular parametric distribution, able to accommodate well every empirically observed behavior, once the parameters are suitably tuned. The is no model selection, nor any a-posteriori distribution.

Mathematical preliminaries In Chapter 1 we will properly introduce the Negative Binomial distribution (NB) from a mathematical viewpoint. It will serve to capture the Relative Species Abundance (RSA), i.e. the abundances' distribution of the different items of a datasets. To be precise, the core of our framework is a truncated version of the Negative Binomial, obtained by removing the point $\{0\}$ from the codomain, thus preventing a species to be extinct (see Section 1.1). A truncated Negative Binomial with parameters $\xi \in (0,1)$ and $r \in (0,\infty)$ reads as follows

$$\mathbb{P}(n) = \frac{1}{1-(1-\xi)}\binom{n+r-1}{n}\xi^n(1-\xi)^r, \qquad n \geq 1 \tag{1}$$

In Section 1.2 we motivate the particular choice of the Negative Binomial, showing that it emerges naturally as the steady state distribution of a particular birth and death process. These kind of processes are able to describe the population dynamics, i.e. its growth and decline, as driven by the birth and death rates of the different species.
This construction embrace the Neutral Theory [Hub01], and following this approach we assume every species to be demographically equivalent. As a byproduct we obtain that equation (1) describes the probability density function of every species; thus different realizations of the same distribution (1) represents the different species' abundances forming the resulting empirical RSA at the global scale.
In Section 1.3 we show that the functional form of a Negative Binomial does not change when sampling different fractions of areas. The fact that the RSA maintains the same functional form at different scales will be central in our framework. We will refer to this property as *form-invariance*. We remark that form-invariance should not be confused with *scale-invariance*, where the latter holds for distributions f such that $f(px) = g(p)f(x)$ where $g(p)$ is a multiplicative scale-dependent constant. In contrast, with form-invariant we mean a distribution which maintains the same functional form under random sampling. Indeed, if we assume to randomly sample a fraction p of the population, the resulting empirical local RSA is again distributed according to a Negative Binomial, with same parameter r and rescaled parameter ξ_p given by

$$\xi_p = \frac{p\xi}{1-\xi(1-p)} \tag{2}$$

Section 1.4 is devoted to the extension of the domain of the parameter r, allowing it to take values also in the negative interval $(-1,0)$. The negative value of r reflects in a Power Law behavior of the RSA's tail described in (1), with an exponential cutoff. This

different trait permits to use the Negative Binomial to accommodate empirical RSA even when their tails are not exponentially distributed, being able to capture Power Laws with exponent $1 - r$.
Finally in Section 1.5 we explicit some relations between our truncated Negative Binomial and the famous Log-series distribution, named after the great statistician Ronald A.Fisher [FCW43], which has been for many years the reference distribution used to describe RSA patterns in ecological communities.

Statistical models for upscaling from Negative Binomially distributed data
In Chapter 2 we derive our upscaling method, declining it accordingly to the requirements of the different datasets. The key features of out method are explained in Section 2.1. Form-invariance under different sampling efforts is at the core of our approach, allowing for an analytical expression for how parameters of the distribution change across scales, as described in equation (2). In particular we can derive the analytical expression for a species not to be observed in the local sample, that is

$$P(\text{species not observed at scale } p) = 1 - \frac{1-(1-\xi_p)^r}{1-(1-\xi)^r}. \tag{3}$$

On the other hand, if we denote by S the total number of existing species and by S_p the number of different species observed in the local sample, we have that

$$\text{Fraction of species not observed at scale } p = \frac{S - S_p}{S}. \tag{4}$$

Hence, comparing 3 with 4 and arranging the terms we obtain an estimator for S that reads:

$$\hat{S} = S_p \frac{1 - \left(1 - \frac{\xi_p}{p+\xi_p(1-p)}\right)^r}{1-(1-\xi_p)^r} \tag{5}$$

In a similar way we can derive an estimator $\hat{S}_q$ for the number of different species existing at any scale $q \in [0,1]$, and hence we can obtain the Species Accumulation Curve (SAC). In Section 2.2 we derive again an estimator for S, we but modify our method to deal with the case of local samples of different nature. In particular we assume to dispose of a collection of n disjoint local samples each containing only occurrences information. This means that we know which species are present in every sample, but we do not access to the information of their abundance. For simplicity of presentation, we assume that the n samples are all of the same size and that, when combined together, they represent a fraction p of the entire datasets. Under this conditions, we empirically compute the number $\hat{S}_{p_k}$ of different species at scale $p_k = \frac{k}{n}p$ by averaging the number of observed species when gathering k of the n different subsamples. We already described how to express the theoretical SAC for values $q = p_k$ as a function of the two parameters (r, ξ_p). Hence, by fitting the parametrical curve $S_{p_k}(r, \xi_p)$ using the data points $\hat{S}_{p_k}$ we can obtain an estimation for the parameters r and ξ_p that describe the RSA at scale p. At this point, we can apply the same method derived in the previous Section.

In Section 2.3 we extend the upscaling method described in Section 2.1 to encompass finest questions. Namely we can locally access to more information than just the number of species S_p and their RSA. In this Section we analyze how to proper use the information of the local abundance of every species to predict their global abundance. In details, we derive an estimator $\hat{S}^{(p^c)}(K \mid L)$ to predict the number of species in the unseen fraction of the population with abundance in a certain range K, given that their observed local abundance fall in another certain range L. The key ingredient to construct this estimator is a simple application of the Bayes formula. We indicate by $n_j^{(q)}$ the abundance of the species j at the scale q, where $q = p$ is the local observed scale, $q = p^c$ is the unobserved scale $1 - p$ and $q = 1$ is the global scale. We obtain

$$\mathbb{P}\left[n_j^{(p^c)} = k \mid n_j^{(p)} = \ell\right] = \frac{\mathbb{P}\left[n_j^{(p)} = \ell \mid n_j^{(1)} = k + \ell\right] \mathbb{P}\left[n_j^{(1)} = k + \ell\right]}{\mathbb{P}\left[n_j^{(p)} = \ell\right]} \tag{6}$$

This formula will help us to find the estimator $\hat{S}^{(p^c)}(k \mid \ell)$ for the species with abundance k in the unobserved fraction $1 - p$ of the population, given that they have abundance ℓ in the sample at observed scale p. Then, a simple sum over $k \in K$ and over $\ell \in L$ allows us to find the desired expression for $\hat{S}^{(p^c)}(K \mid L)$.

Inferring macro-ecological patterns from local species' occurrences In Chapter 3 we present the results of an application of our novel rigorous statistical framework to upscale ecological biodiversity patterns from local information on species occurrence data, as described in Section 2.2.
Section 3.1 motivates the need of an estimator able to upscale biodiversity only from information about the presence or absence of a species in different surveyed plots, without specifying the number of individuals within them. In fact, many databases contain a highly imprecise the information on abundances, if available at all, whereas we usually dispose of the information on the presence/absence of species. Though different upscaling approaches have been proposed in ecological literature, to the best of our knowledge, most of them have not been generalized to the case of binary data.
In Section 3.2 we recall the key steps of our method and we linger on analogies and differences of this method with the one based on abundance described in Section 2.1. In particular we accurately describe the dataset assumption thus legitimating ourselves to correctly apply the method. Moreover, we introduce a new quantity, the Relative Species Occupancy (RSO), describing the distribution of the occurrences (number of occupied cells) across species. Our framework gives directly all parameters of the RSO by solely fitting the SAC curve, through which one can obtain the r and ξ parameters, which well describe both the RSA and the RSO distributions at all spatial scales of interest.
Finally in Section 3.3 we proper present the results of this application. Firstly we apply our method on four computer generated forests, varying the different datasets for the distribution of their RSA (truncated Negative Binomially or Log Normal) and for the the presence or absence of spacial correlation. Then we test our method on sub-samples taken from two empirical forest data (BCI and Pasoh) for which we have informations

on both species occurrence and abundances. We compare our results on species richness obtained only from presence-absence data with the most popular non-parametric indicators proposed in the literature, finding that our method outperforms all the others for both BCI and Pasoh forests.

Upscaling human activity data: a case study with a statistical ecology approach Chapter 4 is again an application of our upscaling method, this time to four datasets deriving from human activities. In Section 4.1 we give detailed description of these datasets, namely concerning Email activity, Twitter posts, Wikipedia articles and Gutenberg books. The interest on these datasets is twofold: firstly we are applying a method coming from ecology to datasets of artificial origin; secondly, the RSAs of these datasets display a Power Law behavior that differs from the usual exponential decay of forests' RSA. Hence we serve of the setting derived in Section 1.4, allowing the parameter r to take values in the negative interval $(-1, 0)$ thus reflecting in a Power Law behavior of the Negative Binomial that we use to accommodate the RSA.
Sections 4.2 and 4.3 are dedicated to present the results we obtained when testing our method on these datasets. For the first time we infer how the species' abundances change across scales, i.e. we applied the theoretical estimator derived in Section 2.3. Focusing on Twitter, we tested the estimator for the change in popularity of hashtags from a portion p of the observed tweets to the remaining $1 - p$ fraction of the unobserved tweets. By means of the popularity of a hashtag we naively count the number of posts containing it that come to circulate within the network thanks to other users' tweets. In particular we predict the the number $S^{(p^c)}(K^+ \mid L^+)$ of species having abundance at least K at the unobserved scale $1 - p$ given that they have abundance at least L at the sampled scale $p = 5\%$.
To conclude, we describe in Section 4.4 some possible application of our method, from resource management to collective monitoring and language learning process.

Random walks in random environment: limit theorems for heavy-tailed processes

The second part of the thesis will regard the study of a *Lévy flight on a one-dimensional Lévy random medium*. With this expression we indicate a discrete random walk Y that performs its jumps un the points of a random environment ω accordingly to an underlying random walk S that drives the dynamics of Y on the Lévy medium.
In our model, the expression 'Lévy random medium' indicates a stochastic point process ω on the real line, i.e. a collection of random points where the distances between nearby points have heavy-tailed distributions. Processes of this kind have been receiving a surge of attention, of late, both in the physical and mathematical literature; cf., respectively, [BFK00; Sch02; BCV10; Bur+12; ZDK15] and [Bia+16; BLP20; MS18]. They model a variety of situations that are of interest in the sciences. In particular, they are used as supports for various kinds of random walks, in order to study phenomena of anomalous transport and anomalous diffusion. An incomplete list of general or recent references on

this topic includes [SZF95; KRS08; Cri+14; ZDK15; Art+18; Rad+19].
The underlying random walk S has i.i.d. heavy tailed increments that are independent of ω. Our process of interest if Y, taht is defined as $Y_n := \omega_{S_n}$. This means that the process Y performs the same jumps as S, but on the marked points ω instead of $\mathbb{Z}$.
Term 'Lévy flight' usually indicates a discrete-time random walk with long-tailed instantaneous jumps. This is in contrast to a 'Lévy walk', which in general designates a *persistent*, continuous- or discrete-time, random walk with long-tailed trajectory segments that are run at constant finite speed [ZDK15]. A Lévy walk is often seen as an interpolation of a Lévy flight. Lévy walks are generalization of a system that first appeared in the physical literature 20 years ago with the name *Lévy-Lorentz gas* [BFK00] (more precisely, the Lévy-Lorentz gas is the case where the underlying random walk is simple and the interpolation is performed with unit-speed). It was devised as a one-dimensional toy model for the study of anomalous diffusion in porous media [Lev97; BFK00; BCV10].
There are several reasons to study our Lévy flight on random medium. The most self-serving, on the part of the present authors, is to build a basis to investigate the properties of the associated Lévy walk, as described above (see the proofs in [Bia+16; BLP20]). Also, Y can be thought of as the limit of a continuous-time random walk with resting times on the points ω_k, when the ratio between the speed of the walker and the typical resting time diverges. This can be used to model a variety of situations where a given dynamics is very fast compared to its decision times, e.g. electronic signal on a network whose nodes act as relatively slow processing stations, human mobility (assuming, as it is often the case, that resting times are substantially longer than travel times), etc.
This particular model aside, there is no lack of general motivation for the study of random walks on points processes, especially in light of the fact that the topic is regrettably less developed than others in the field of random walks, with the exception perhaps of random walks on percolation clusters *et similia*. For some interesting lines of research see, e.g., [CF+09; CFP13; Kub13; BR15; Zhu15; Rou14] and references therein. A recent paper which we extend with the present thesis is [MS18].

Ideas of weak convergence Chapter 5 is devoted to present the notion of weak convergence. Of course this argument has been extensively treated in many monographs like [Bil13; Whi02; Fel57; JS13] and the purpose of this Chapter is rather to highlight the importance of this notion and to recall some crucial aspects.
The basic notion are gathered in Section 5.1. In particular, Portmanteau Theorem can be considered as a formal definition of weak convergence. We give a small in depth analysis of the class of measures in $\mathcal{C}([0,1])$, i.e. the space of continuous functions defined on the unit interval, showing in particular that the finite-dimensional sets are a determining class, but not a convergence determining class. Finally we characterize weak convergence of measures by means of the Prohorov metric.
We then introduce the concept of random element in Section 5.2. In this way convergence in distribution of random element is completely equivalent to weak convergence of the laws induced by the random elements. For example, we say that the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(S, \mathcal{S})$ converges in distribution

to a limit random variable X if the law $\mathbb{P} \circ X_n^{-1}$ induced on the space $(S, \mathcal{S})$ converges weakly to $\mathbb{P} \circ X^{-1}$.
In Section 5.3 we give an insight on the compactness approach to show weak convergence. Provided that the sequence of measures is relatively compact, then convergence of the finite-dimensional distribution is enough to guarantee weak convergence. Relative compactness is not usually easy to exhibit, however it is usually tantamount to the concept of tightness. Another way to prove weak convergence is by using the mapping approach, as described in Section 5.4. Basically, if we assume that $X_n \xrightarrow{d} X$, then it holds that also $h(X_n) \xrightarrow{d} h(X)$ for any continuous mapping h. Moreover, convergence holds even if h is discontinuous but the probability that X takes values on discontinuity point of h is negligible, i.e. if $\mathbb{P}(X \in Disc[h]) = 0$. To conclude the chapter we recall in Section 5.5 the role of characteristic functions to determine weak convergence. Indeed it suffice to have that the characteristic functions of a sequence of measure $(\mathbb{P}_n)_{n\in\mathbb{N}}$ to converge pointwise to the characteristic function of $\mathbb{P}$.

The càdlàg space $\mathcal{D}$ In Chapter 6 we introduce the càdlàg space $\mathcal{D}$, i.e. the space of functions that are right-continuous and with finite left limits at any point of the domain. The basic notions of this space are listed in Section 6.1. In particular we show that the space $\mathcal{D}([0,1])$ consists of piecewise continuous functions, defined on the unit interval, with discontinuities of the first kind. Moreover, only finitely many jumps can exceed a given threshold.
In Section 6.2 we endow the space $\mathcal{D}([0,1])$ with the J_1-Skorokhod distance and we characterize convergence with respect to the topology induced by the metric d_{J_1}. In particular two functions $x, y \in \mathcal{D}$ are ε-close w.r.t. the metric J_1 if they are ε-uniformly close over $[0,1]$ after allowing continuous small perturbations of time (the function argument), i.e.

$$d_{J_1,[0,1]}(x,y) := \inf_{\lambda\in\Lambda_{[0,1]}} \max\left\{ \sup_{t\in[0,1]} |x\circ\lambda(t) - y(t)|\,,\ \sup_{t\in[0,1]} |\lambda(t) - t| \right\} < \varepsilon \qquad (7)$$

where $\Lambda_{[0,1]}$ is the set of strictly increasing functions λ mapping the domain $[0,1]$ onto itself, such that both λ and its inverse λ^{-1} are continuous. In Section 6.3 we will show that $\mathcal{D}([0,1])$ is complete under the metric d_{J_1}. Moreover we give a criterium to characterize relative compactness and, similarly to what done for $\mathcal{C}([0,1])$ we analyze the role of the finite-dimensional sets in $\mathcal{D}([0,1])$. Finally we give conditions under which the convergence of the sum $x_n + y_n \to x + y$ w.r.t. J_1 is guaranteed.
In Section 6.4 we extend the notion of càdlàg space to any bounded interval I and also to unbounded intervals. We also adapt the definition of the J_1 distance and we characterize convergence even for unbounded intervals. Finally we specify a less customary space with càglàd trajectories in $\mathbb{R}^-$, that we are going to adopt for the results of Chapter 9. In Section 6.5 we define processes with càdlàg trajectories and characterize some approaches to show convergence, namely the continuous mapping and the compactness one. The idea is to regard stochastic processes as random elements with values in the space of trajectories, i.e. we consider random function with value in $\mathcal{D}$. With this in

mind, convergence of a sequence of stochastic processes naturally becomes convergence of a sequence of probability measures on a space of functions. Our attention will then be oriented to the function space $\mathcal{D}$ containing the sample paths and to its topology.
Section 6.6 is very important since we derive three additional topologies which we can endow the space $\mathcal{D}$ with. In particular the J_2 topology is similar to J_1 with the only difference that the deformation of time λ can be a discontinuous bijection. For completeness of treatment we include the definitions of the M_1 and M_2 metrics, that evaluate distance of functions on the basis of their graphs. Finally Section 6.7 is devoted to the proof of the continuity of the addition map with respect to the J_2 topology. To the author's knowledge, it is the first time that such a proof is showed. It will be fundamental for our results of Chapter 9.

Limit theorems for random walks Chapter 7 is a condensate of famous results on the theory of random walks. A generic random walk $S = (S_n)_{n\in\mathbb{N}}$ is defined in the following way.

$$\begin{aligned} S_0 &= 0 \\ S_n &= \sum_{i=1}^{n} \xi_i, \qquad \text{for } n > 0. \end{aligned} \tag{8}$$

In general the random variables $(\xi_i)_{i\in\mathbb{N}}$ can follow any general distribution. In Section 7.1 we adopt the usual hypothesis that the random variables $(\xi_i)_{i\in\mathbb{N}}$ appearing in (8) are independent and identically distributed (i.i.d.) according to a distribution with finite moments at leas of order two. Under this assumption we can derive some of the most famous theorems in probability theory, namely the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The latter reads

$$\frac{S_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1); \tag{9}$$

where $\mu = \mathbb{E}[\xi]$, $\sigma^2 = \mathbb{V}[\xi]$ and $\mathcal{N}(0,1)$ is the standard normal distribution. Furthermore, we can build a continuous process, the normalized partial sum process, and show that is converges weakly to the Brownian Motion. This is the content of the Donsker's Theorem, a functional and more powerful version on the CLT.
In Section 7.2 we deal with the case where the increments $(\xi_i)_{i\in\mathbb{N}}$ are i.i.d. but their distribution have infinite second moment. In particular we focus our analysis on the case where the increments fall into the normal domain of attraction of a stable distribution of a certain index $\alpha \in (0,2)$. In this case, it is possible to prove that there exist a sequence $(m_n)_{n\in\mathbb{N}}$ and a constant c such that it holds $(S_n - m_n)/cn^{1/\alpha} \xrightarrow{d} S_\alpha$, where S_α is a stable random variable of index α. When $\alpha \in (1,2)$ we are able to prove a LLN, and the stable version of the CLT reads

$$\frac{S_n - \mu n}{n^{1/\alpha}} \xrightarrow{d} S_\alpha. \tag{10}$$

Comparing equations (9) and (10) we can appreciate the different rescaling when dealing with stable laws of infinite variance. In the case $\alpha \in (0,1)$ the LLN does not hold, but

the rescaled sum $S_n/n^{1/\alpha}$ converges weakly to S_α due to stability properties. Finally, we prove a stable version of Donsker's theorem, where the limit process is no longer the Brownian Motion, but a Lévy stable process.

Dynamics on a Lévy random environment In Chapter 8 we set the notation and prepare the field to describe our results. Here we no longer assume, as in Chapter 7, the increments of the random walk to be i.i.d., introducing dependence between the steps. We focus on a specific choice of dependence, considering the class of Random Walks in Random Environment (RWRE).
As first, we introduce in Section 8.1 the class of renewal point processes on the real line. In particular we set a point ω_0 on the origin, and the rest of points $(\omega_n)_{n\in\mathbb{Z}}$ are distributed on $\mathbb{R}$ so that the distance between neighboring points belong to a common distribution ξ. In this way $\omega_n := \pm\sum_{i=\pm1}^{n}\zeta_i$, where the sign is $+$ (resp. $-$) accordingly to $n > 0$ (resp. $n < 0$). We will be interested on a point process ω where the distances $\zeta_k := \omega_{k+1} - \omega_k$ between neighboring targets are i.i.d. variables in the normal domain of attraction of a β-stable distribution, with $\beta \in (0,1)\cup(1,2)$.
We then consider in Section 8.2 generalization of the Lévy-Lorentz gas, namely a process $(Y_n)_{n\in\mathbb{N}}$ constructed with the following rule. We consider a random walk $S = (S_n)_{n\in\mathbb{N}}$ on $\mathbb{Z}$ with $S_0 = 0$ and i.i.d. increments $(\xi_i)_{i\in\mathbb{N}}$ in the normal domain of attraction of an α-stable variable, with $\alpha \in (0,1)\cup(1,2)$. We refer to S as the underlying random walk and we define our process Y, starting at the origin, as

$$Y_n := \omega_{S_n}, \qquad n >\geq 0. \tag{11}$$

In other words, S drives the dynamics of Y on the point process ω. To conclude this Chapter we briefly recall a couple of previous results in Section 8.3, where the authors considered a simplified version of this setting.

Limit theorems for Lévy flights on a Lévy random medium Chapter 9 is devoted to the presentation of the results and of their proof . In Section 9.1 we define in details our model, adapting the general setting presented in Section 8.2 and specifying some preliminary convergence.
The theorems are stated in Section 9.2 and proved in Section 9.3. The different statements of the theorems depend on the values of α and β.

We give annealed limit theorems for Y in all cases $\alpha, \beta \in (0,1)\cup(1,2)$, identifying in each case both the scale n^γ, whereby

$$\left(\frac{Y_{\lfloor nt\rfloor}}{n^\gamma}, t \in [0,\infty)\right) \tag{12}$$

converges to a non-null limit, and the limit process. In all cases we prove the optimal, or at least morally optimal, functional limit theorem, meaning that we show distributional convergence of the process with respect to the strongest Skorokhod topology that applies there.

Af first we consider the cases with $\beta \in (0,1)$, i.e. when the distance between neighboring targets of the point process have infinite mean. If $\alpha \in (0,1)$, or $\alpha \in (1,2)$ but $\mu = \mathbb{E}[\xi] = 0$, then $\gamma = 1/\alpha\beta$ and we can only prove convergence of the finite-dimensional distributions of the suitably rescaled random walk in random environment. If, otherwise, $\alpha \in (1,2)$ and $\mu = \mathbb{E}[\xi] \neq 0$, then $\gamma = 1/\beta$ we are able to prove convergence of the suitable rescaled process with respect to the J_2-Skorokhod topology.
Next we move to the case $\beta \in (1,2)$. Since the distance between neighboring targets is a positive value, then its mean must be a positive number $\nu > 0$. If $\alpha \in (0,1)$, or $\alpha \in (1,2)$ but $\mu = \mathbb{E}[\xi] = 0$, then $\gamma = 1/\alpha$ and we can prove weak convergence of the suitable rescaled sequence of processes with respect to the J_1-Skorokhod topology. We can obtain a similar result also if $\alpha \in (1,2)$ and $\mu = \mathbb{E}[\xi] \neq 0$. In this latter case $\gamma = 1$ and the limit process is deterministic. Our last theorem states weak convergence of the fluctuations around the limit process. In particular they are or order $n^{max\{1/\alpha, 1/\beta\}}$.

Contents

Part I

Upscaling statistical patterns from reduced partial information of big datasets in natural and social sciences

Chapter 1

Mathematical preliminaries

The problem of inferring total biodiversity when only scattered samples are observed is a long-story problem. The origin of this problem dates back to an ecology question, namely one would like to infer total biodiversity after having observed local scattered samples.

We have already mention in the Introduction that the first person to ask himself such a question, or at least the first person to ask it publicly, was Alexander Steven Corbet, and consequently the first estimation was performed by Ronald Aylmer Fisher. Since then 8 decades have passed and the challenge of biodiversity upscaling has attracted increasing interest, producing a wide range of competing approaches.

Extrapolating species richness from the local samples is not straightforward as it may appear. We start this journey from the upscaling problem in ecology, where the main goal is to infer the global number of different species that populates a big forest, given the observation available at small local places. Tropical forests have long been recognized as one of the largest pools of biodiversity [Cro+15]. In fact, more than two-fifths of the number of worldwide trees can be found either in tropical or sub-tropical forests, though the surveyed and classified fraction is very small. Global patterns of empirical abundance distributions show that tropical forests vary in their absolute number of species but display surprising similarities in the distribution of individuals across species [Vol+05; McG+07].

One way to tackle the problem is by describing specific patterns of different samples with parametrical distributions. The class of distribution should be able to capture the general behavior, while allowing specific fit by varying the values of the parameters. A widespread statistical tool used to describe the commonness and rarity of species in an ecological community is the Relative Species Abundance (RSA), which is a list of the species present within a region along with the number of individuals per species (or, equivalently, the relative abundance of that species among the population) [Mac60; Mag13]. RSA is thus a key element to measure biodiversity, as it gives information about how common or rare a species is relative to other species in a given location or community. Note that RSA distributions are usually graphed as frequency histograms ("Preston plots";[Pre48]) or rank-abundance diagrams ("Whittaker Plots"; [Whi65]). Throughout this thesis we will display RSA using frequency histograms, using a Log-

Log scale in both axis to plot the results. This approach, where the mechanism driving abundances is the competition between species for available resources, is referred as a niche-based model. In this kind of model, the splitting of the resource "pie" among different species is responsible for the number each species's individuals populating the community. To resume, the competition for the consumption of similar resources has the direct consequence that species with access to abundant resources will have higher carrying capacities than those with little access. In order to survive, each species needs to provide itself a niche-space. Numerous niche apportionment models have been developed, each making different assumptions about how species carve up niche-space. For example it can be done by slicing up the resources' pie into smaller pieces or by moving into a vacant niche and exploiting newly available locations.

In this thesis we will not consider any niche-based model, rather our framework leans on neutral theory approach. The Unified Neutral Theory of Biodiversity and Biogeography (UNTB) [Hub01] is a special form of mechanistic model that takes an entirely different approach to community composition than the niche apportionment models. Instead of species populations reaching equilibrium within a community, the UNTB model is dynamic, allowing for continuing changes in relative species abundances through drift. The best way to show the basics of Neutral Theory is perhaps by visualizing a community of individuals as a grid with a certain number of spaces. Each space is occupied with individuals of different species. The main assumptions are that the number of spaces available in the grid are limited. Hence if a certain species increases its abundance, then this lead to a decrease of the number of individuals other species in the grid. Hence the distribution of abundances does not depend directly on apportionments of resources, i.e. the model does not make any assumption or description of this process, rather it takes into account birth, death, immigration, extinction and speciation to modify community composition over time. They are the key drivers of change leading to the distribution of biodiversity in the ecosystem.
We will describe in Section 1.2 the stochastic model leading to a steady state distribution of abundances that we pick to sample the distribution. We will come back to this point later. Now we want to give a closer look on how RSA data are empirically obtained. Speaking about forests, typically the RSA is measured at local scales (e.g. in quadrats or transects), in which the identities of all the individuals living in the area are known. The sampled RSA can be fit to a given functional form at that scale. However, that form may change at different spatial scales (see figure 1.1), thus hindering analytical treatment [Aza+15]. For practical reasons, biodiversity is typically measured or monitored at fine spatial scales. However, the key drivers of ecological change need to act at larger scales [Ber+16]. Conservation issues, for example, apply to diversity at global, national or regional scales. Hence we need to balance the opposite tendencies of focusing too much on the small world and of disregarding little variations in local samples.

It is therefore fundamental to be able to specifically upscale globally the patterns, like the RSA, that we are able to detect and measure locally. Many different biodiversity estimators have been proposed since the Pandora's box was open by Fisher almost 80 years ago. Even though they have been developed under different statistical frameworks,

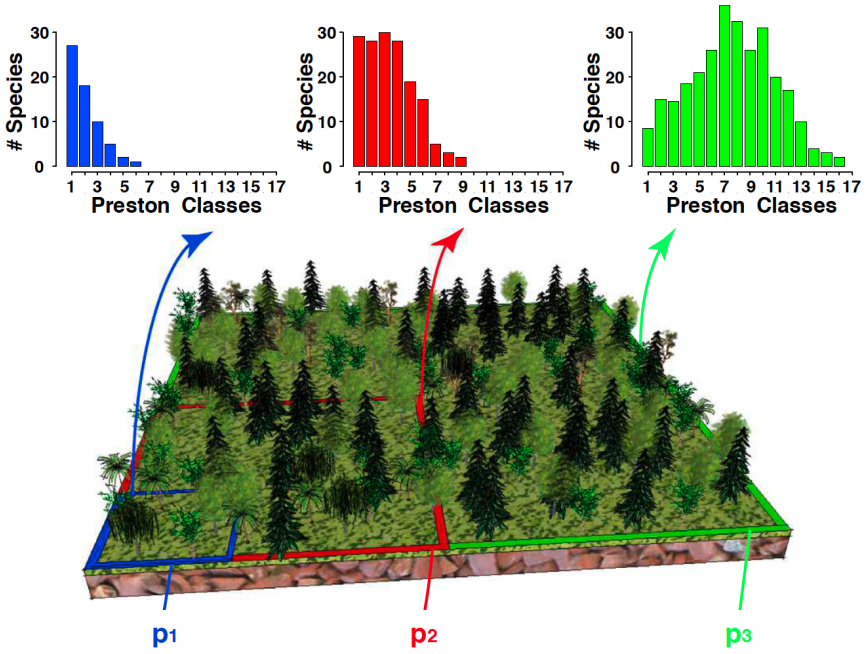

Figure 1.1: **RSA at different scales.** Relative Species Abundance at different spatial scales for Barro Colorado Island, in Panama. The form of the RSA changes with the scale.

most of them reflect the need of extrapolation from a local-regional scale extrapolation rather than assuming sparse random sampling. The majority of estimators are sensitive to the spatial distribution of different plants [Plo+00; Aza+16], to the sample coverage, and to the specific sampling method used [Cha+09]. Notice finally that usually only local scattered samples are available. For example, notice that Figure 1.2 displays the different locations where a sample of a relatively small area is available. When dealing with forests, it is common to have access to multiple separated areas, rather than a single big one. Sparse datasets prevent from sampling errors, such as labeling as common a single species only because it shows numerous specimens gathered in a small area.

To the point, the goal of this thesis appears very clear: we want to use a parametric model to upscale local information to the global scale. The fundamental hypothesis of our work is that the RSA is distributed as a Negative Binomial distribution. Throughout this and the following Chapters we will construct our framework step by step and then present our results.

1.1 The truncated Negative Binomial distribution

It's time to properly introduce the truncated Negative Binomial distribution, that will play a central role in this thesis. In this Section we give a brief excursus on the Negative

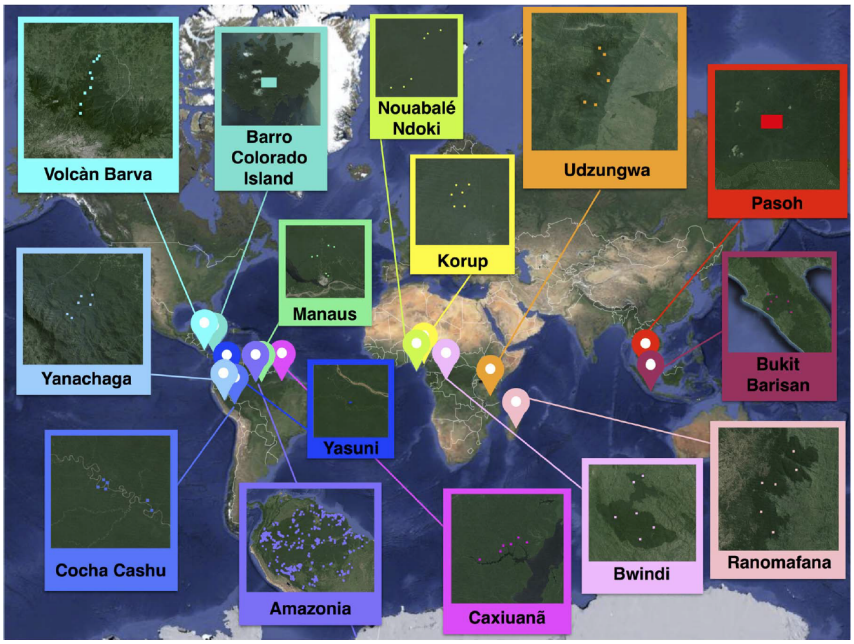

Figure 1.2: **The challenge of estimating global tropical species richness.** A map depicting 15 different forests for which the coordinates of each sampled subplot are known. The upscaling goal is to deduce the species richness and abundances of each entire forest on the basis of the very limited knowledge coming from the scattered samples in the marked dots.

Binomial distribution, in order to clarify what regime of parameters we are going to adopt and what are the implication of the truncation.
Originally this distribution was used to compute the number of expected failure before obtaining a certain number of successes in a sequence of independent and identically distributed Bernoulli trials. Indeed, the probability of obtaining n failures before having the r-th success is given by the following probability, where p represents the probability of failure of every independent copy of a Bernoulli random variable,

$$\mathbb{P}(n \text{ failures before } r\text{-th success}) = \binom{n+r-1}{n} p^n (1-p)^r \quad n \geq 0 \tag{1.1}$$

The previous formula holds for any $r \in \mathbb{N}^*$ and any $n \in \mathbb{N}$, where $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$.
Note that equation (1.1) actually defines a probability density function. In facts, suppose to define the following analytical function f by two equivalent expressions:

$$\frac{1}{1-x} =: f(x) := \sum_{k=0}^{\infty} x^k \qquad x \in (0,1). \tag{1.2}$$

Then, by differentiating $(r-1)$ times the function f, and repeating the calculus for both expressions in (1.2), we obtain

$$(r-1)!(1-x)^{-r} = f^{(r-1)}(x) = \sum_{k=r}^{\infty} \frac{(k-1)!}{(k-r)!} x^{k-r} = \sum_{n=0}^{\infty} \frac{(n+r-1)!}{n!} x^n. \tag{1.3}$$

Finally, dividing both sides by $(r-1)!(1-x)^{-r}$ we get

$$1 = \sum_{n=0}^{\infty} \frac{(n+r-1)!}{n!(r-1)!}(1-x)^r x^n = \sum_{n=0}^{\infty} \binom{n+r-1}{n} \xi^r (1-\xi)^n \tag{1.4}$$

where in the last passage we perform a change of variables $\xi = 1-x$.

We wish to extend the previous definition allowing r to take real positive values. To do this, we need to specify the meaning of the binomial coefficient $\binom{s}{m}$ when $s \in \mathbb{R}^+$. We have

$$\binom{s}{m} := \begin{cases} \frac{s(s-1)\cdots(s-m+1)}{m!} & \text{if } m \geq 0 \\ 0 & \text{if } m < 0 \end{cases} \tag{1.5}$$

Note that we re-obtain the classical binomial coefficient when $s \in \mathbb{N}$ and $m \in \{0, .., s\}$. However, when s is non-integer, then the expression above can not be interpreted combinatorially. It does not make sense to speak about choosing m things out of a set of s items. However, the equation above does have mathematical validity and we will need it to define our statistical framework. Note moreover that the definition above holds for any integer $m \in \mathbb{Z}$. The extension to negative integers is quite straightforward, since we simply impose the coefficient to be zero. Note however that the definition stands also for $m > s$, leading eventually to negative factors in the numerator.

The definition above suffices for our purposes. However, to conclude this excursus on binomial coefficients, we give the following definition of binomial coefficient, holding for any pair of complex numbers z, w.

$$\binom{z}{w} := \lim_{u \to z} \lim_{v \to w} \frac{\Gamma(u+1)}{\Gamma(v+1)\Gamma(u-v+1)} \tag{1.6}$$

where $\Gamma(\cdot)$ is the famous Gamma function. We briefly want to show that the latter definition reduces to the previous one when $z = s \in \mathbb{R}$ and $w = m \in \mathbb{Z}$. Indeed, notice that $\Gamma(y+1) = y\Gamma(y)$, hence $\Gamma(s+1)/\Gamma(s-m+1) = s(s-1)\cdots(s-m+1)$ and $\Gamma(m+1) = m!$, thus recovering exactly equation (1.5). Note that the limits in the equation above are necessary to properly handle the cases where z or w are singularities of the Γ function, i.e. points $0, -1, -2$, etcetera. Singularities of the distributions will play and important role, as we will see in Section 1.4.

We are now ready to define the Negative Binomial distribution.

Definition 1. We say that a random variable X with values on $\mathbb{N}$ is distributed like a Negative Binomial if

$$\mathcal{P}_{NB}(X = n | r, \xi) = \binom{n+r-1}{n} \xi^n (1-\xi)^r, \qquad n \geq 0 \tag{1.7}$$

for some parameters $\xi \in (0,1)$ and $r \in (0,\infty)$. The binomial coefficient appearing in the latter formula is defined as in (1.5).

Note that equation (1.7) actually represents a probability distribution, with discrete density summing to 1. In facts, we have

$$\begin{aligned}
\sum_{n=0}^{\infty} \mathcal{P}_{NB}(X=n|r,\xi) &= \sum_{n=0}^{\infty} \binom{n+r-1}{n} \xi^n (1-\xi)^r \qquad (1.8)\\
&= \sum_{n=0}^{\infty} \frac{\Gamma(n+r)}{n!\Gamma(r)} \xi^n (1-\xi)^r \\
&= \sum_{n=0}^{\infty} \frac{\int_0^\infty t^{n+r-1} e^{-t} dt}{n!\Gamma(r)} \xi^n (1-\xi)^r \\
&= \frac{1}{\Gamma(r)} (1-\xi)^r \int_0^\infty e^{-t} t^{r-1} \sum_{n=0}^{\infty} \frac{t^n \xi^n}{n!} dt \\
&= \frac{1}{\Gamma(r)} (1-\xi)^r \int_0^\infty e^{-t} t^{r-1} e^{t\xi} dt \\
&= \frac{1}{\Gamma(r)} \int_0^\infty (1-\xi)^r e^{-t(1-\xi)} t^{r-1} dt \\
&= \frac{1}{\Gamma(r)} \int_0^\infty (1-\xi)^r e^{-s} \frac{s^{r-1}}{(1-\xi)^{r-1}} \frac{ds}{1-\xi} \\
&= \frac{1}{\Gamma(r)} \int_0^\infty e^{-s} s^{r-1} ds \\
&= \frac{1}{\Gamma(r)} \Gamma(r) = 1
\end{aligned}$$

Hence the Negative Binomial defined in (1.7) is a probability distribution on $\mathbb{N}$.

We wish to slightly modify the previous definition to better accomodate our scopes. Mathematically speaking, we want to exclude $n = 0$ from the codomain of the Negative Binomial defined above. We give here a brief motivation of this choice. Suppose indeed there are $S \in \mathbb{N}$ different species in the population, each one labeled by an index $i \in \{1, .., S\}$, and denote by η_i the abundance of the i-th species, i.e. the number of individuals belonging to that species. In the case of forests' sample, η_i denotes the number of plants of the given species i. By assumption, any species has strictly positive abundance, since we are only considering existing species. Thus we want to impose the probability of an existing species to have null abundance to be equal to zero. However, we want here to hang on the mathematical point of view, leaving all the considerations regarding the framework to the next Chapter.
We define the following truncated Negative Binomial distribution, obtained from (1.7) by excluding $n = 0$ from the codomain and renormalizing the densities.

Definition 2 (truncated NB)**.** For any given choice of $\xi \in (0,1)$ and $r \in (0,\infty)$, we say that a random variable X with values on $\mathbb{N}^*$ is distributed like a truncated Negative

Binomial of parameters r and ξ if

$$\mathbb{P}(n) = c(r,\xi)\mathcal{P}_{NB}(n|r,\xi), \qquad n \geq 1 \tag{1.9}$$

where $\mathcal{P}_{NB}(n|r,\xi)$ is the classical Negative Binomial distribution already defined in (1.7) and the normalizing constant is given by

$$c(r,\xi) = \left(\sum_{n=1}^{\infty} \mathcal{P}_{NB}(n|r,\xi)\right)^{-1} = \frac{1}{1-(1-\xi)} \tag{1.10}$$

Equation (1.9) plays a key role throughout this thesis. The next sections are devoted to give some properties and extension of this distribution. In Section 1.2 we present how this distribution emerges naturally as a steady state of a particular birth and death process. In Section 1.3 we exhibit and prove the fundamental property of the Negative Binomial, i.e. form invariance. In Section 1.4 we allow the parameter ξ to take negative values in the interval $(-1, 0)$ and we analyze the consequences. Finally in Section 1.5 we relate our truncated NB with the famous Fisher Log-series.
We present these properties from a mathematical point of view, postponing the implication they have on our framework to Chapter 2.

1.2 Stochastic model leading to a Negative Binomial RSA

In this section, we want to tackle the Negative Binomial distribution from a different point of view. The aim of this Section is to show that a Negative Binomially distributed RSA emerges naturally as the steady state of a birth and death process. These kind of processes lean at the bottom ground of the principles of biological processes. Indeed, many frameworks model the growth and decline of a population by mean of such processes. Let us assume that our ecological community consists of S different species, each evolving independently from the others.

Let $\mathcal{P}_{n,s}(t)$ be the probability that, at time t, the species s consists of exactly n individuals. Note that $s \in \{1, .., S\}$ is just a label, and by independence the same holds analogously for the other species. We assume that the population dynamics is driven by a birth and death process, hence there will be constants $b_{n,s}$ and $d_{n,s}$ regulating the behavior of a population. Note that these constants depend on the particular species and on the actual abundance n. However, these models works with birth and death factors that are homogeneous over time. We can thus derive the master equation governing the evolution of $\mathcal{P}_{n,s}$ over time:

$$\frac{\partial}{\partial t}\mathcal{P}_{n,s}(t) = \mathcal{P}_{n-1,s}(t)b_{n-1,s} + \mathcal{P}_{n+1,s}(t)d_{n+1,s} - \mathcal{P}_{n,s}(t)b_{n,s} - \mathcal{P}_{n,s}(t)d_{n,s} \tag{1.11}$$

The above equation holds for any $n \geq 0$. Of course for $n = 0$ we must impose reflecting boundary conditions, i.e. $b_{-1,s} = d_{0,s} = 0$. It is possible to derive the steady state solution, that reads

$$\mathcal{P}_{n,s} = P_{0,s} \prod_{i=0}^{n-1} \frac{b_{i,s}}{d_{i+1,s}} \tag{1.12}$$

where the term $P_{0,s}$ is a normalizing constant which can by found by imposing that $\sum_{n=1}^{\infty} \mathcal{P}_{n,s} = 1$. Thus we solved the system but the solution is uninformative, since it depends on too many parameters. We want to relax the assumptions, imposing that the birth and death rates depend on density dependent rates b_s and d_s, which represent the pre-capita birth and death rates, and on the term r_s, taking into account immigration events and/or intraspecific interactions. Thus, our coefficients read

$$b_{n,s} = b_s(n + r_s) \tag{1.13}$$
$$d_{n,s} = d_s n \tag{1.14}$$

Setting $\xi_s = b_s/d_s$, and substituting the new terms into (1.12), we obtain

$$\mathcal{P}_{n,s} = P_{0,s}\binom{n + r_s - 1}{n}\xi_s^n \tag{1.15}$$

The normalizing constant $P_{0,s}$ can by found by imposing

$$1 = \sum_{n=1}^{\infty} \mathcal{P}_{n,s} = P_{0,s}\binom{n + r_s - 1}{n}\xi_s^n = P_{0,s}[1 - (1 - \xi_s)^{r_s}](1 - \xi_s)^{-r_s} \tag{1.16}$$

Therefore, the probability that the s-th species has n individual at equilibrium is given by a truncated Negative Binomial with parameters (r_s, ξ_s), where the normalization takes into account the fact that any existing species must have non zero abundance, i.e.

$$\mathcal{P}_{n,s} = \frac{1}{1 - (1 - \xi_s)^{r_s}}\binom{n + r_s - 1}{n}\xi_s^n(1 - \xi_s)^{r_s} \tag{1.17}$$

Note that the previous equation depends on the index s. If we join the neutral theory, then every species is demographically equivalent, meaning that the constants b_s, d_s, r_s lose their species dependence and become universal. Hence the abundance of every species at equilibrium is drawn from the same truncated Negative Binomial distribution.

1.3 Self Similarity Property of Negative Binomial

Suppose we have a population distributed in different species according to a truncated Negative Binomial distribution. This means that the abundance η_i of species i is negative binomially distributed according to (1.9) for every $i \in \{1, .., S\}$, with S being the total number of different species. Assume that the population is spatially independent distributed. Here the term spatially refers to the original application to forests, but for the moment we can elevate it to a more abstract form. In facts, spatiality is linked to the order in which individuals can be considered. If we assume that there is no prescript order, then spatial assumption holds. At this point, we do not enter into real world implication of this assumption. The reader only needs to keep in mind that there is no order or any sort of link between the individuals.

Suppose we want to sample a fraction $p \in (0,1)$ of the population. In real application, p is small since the goal of this procedure is to infer global estimation without analyzing all the dataset. However, taking p too small could affect sensitivity of the method to the sampling scheme adopted.
If a certain species have abundance n in the total population, then the probability that we observe exactly k individuals of that species at the local scale p is given by the classical binomial formula

$$\mathcal{P}_{binom}(k|n,p) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k \in \{0,...,n\} \\ 0 & k > n \end{cases} \tag{1.18}$$

Recall that we are assuming that the overall abundance of the species at the global scale follow the truncated Negative Binomial distribution (1.9). We strengthen the previous notation by indicating the global density by $\mathbb{P}(n|1)$, where the number 1 represent the fraction of the population we are considering. With the same convention, $\mathbb{P}(k|p)$ indicated the probability that a certain species has abundance k at the local scale p. The following proposition states that the local abundance of species is again distributed proportionally to a Negative Binomial.

We already introduce the meaning of the Relative Species Abundance (RSA). Sometimes the empirical of theoretical distribution of the species' abundances is also referred to as the Species Abundance Distribution. In the following, while considering these two terminologies to be equivalent, we will make frequently use of the first notation, taking its meaning for granted.

Proposition 1.3.1. *(Self Similarity for the NB)*
Let the Relative Species Abundance (RSA) at the global scale be distributed as

$$\mathbb{P}(n|1) = c(r,\xi)\mathcal{P}_{NB}(n|r,\xi), \tag{1.19}$$

with $c(r,\xi)$ and $\mathcal{P}_{NB}(n|r,\xi)$ defined as in (1.7) *and* (1.10)*. Denote by $\mathcal{P}_{binom}(k|n,p)$ the binomial sampling probability at a scale p, i.e. the conditional probability that a species has abundance k at the subscale p, given that it has abundance n at the global scale, as defined in* (1.18)*. Then, the local RSA $\mathbb{P}(k|p)$ at the sample scale p is again distributed proportionally to a Negative Binomial, with rescaled parameter ξ_p and same r:*

$$\mathbb{P}(k|p) = \begin{cases} c(r,\xi) \cdot \mathcal{P}_{NB}(k|r,\xi_p) & k \geq 1 \\ 1 - c(r,\xi)/c(r,\xi_p) & k = 0 \end{cases} \tag{1.20}$$

with

$$\xi_p = \frac{p\xi}{1-\xi(1-p)}. \tag{1.21}$$

Proof. To compute the probability that a species in the subpopulation has abundance k, for some strictly positive k, we need to condition on the fact that the species has abundance n in the whole population, and then we sum over n.

If $k \geq 1$, this reduce to compute

$$
\begin{aligned}
\mathbb{P}(k|p) &= \sum_{n\geq k} \mathcal{P}_{binom}(k|n,p)\cdot \mathbb{P}(n|1) \qquad (1.22)\\
&= \sum_{n\geq k} \binom{n}{k} p^k(1-p)^{n-k}\cdot c(\xi,r)\binom{n+r-1}{n}\xi^n(1-\xi)^r\\
&= c(\xi,r)\left(\frac{p}{1-p}\right)^k(1-\xi)^r\sum_{n\geq k}\binom{n}{k}\binom{n+r-1}{n}[\xi(1-p)]^n\\
&= c(\xi,r)\left(\frac{p}{1-p}\right)^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\sum_{n\geq k}\frac{\Gamma(n+r)}{(n-k)!}[\xi(1-p)]^n\\
&= c(\xi,r)\left(\frac{p}{1-p}\right)^k(1-\xi)^r\frac{1}{k!\Gamma(r)}[\xi(1-p)]^k\sum_{n\geq 0}\frac{\Gamma(k+n+r)}{n!}[\xi(1-p)]^n\\
&= c(\xi,r)p^k\xi^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\sum_{n\geq 0}\frac{\int_0^\infty e^{-t}t^{n+k+r-1}dt}{n!}[\xi(1-p)]^n\\
&= c(\xi,r)p^k\xi^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\int_0^\infty e^{-t}t^{k+r-1}\sum_{n\geq 0}\frac{[t\xi(1-p)]^n}{n!}dt\\
&= c(\xi,r)p^k\xi^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\int_0^\infty e^{-t}t^{k+r-1}e^{t\xi(1-p)}dt\\
&= c(\xi,r)p^k\xi^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\int_0^\infty e^{-s}s^{k+r-1}\frac{1}{[1-\xi(1-p)]^{k+r}}ds\\
&= c(\xi,r)p^k\xi^k(1-\xi)^r\frac{1}{k!\Gamma(r)}\frac{1}{[1-\xi(1-p)]^{k+r}}\Gamma(k+r)\\
&= c(\xi,r)\frac{\Gamma(k+r)}{k!\Gamma(r)}\left(\frac{\xi p}{1-\xi(1-p)}\right)^k\left(\frac{1-\xi}{1-\xi(1-p)}\right)^r\\
&= c(\xi,r)\binom{k+r-1}{k}\xi_p^k(1-\xi_p)^r\\
&= c(\xi,r)\cdot\mathcal{P}_{NB}(k|r,\xi_p)
\end{aligned}
$$

Finally, if $k=0$ we have

$$\mathbb{P}(0|p) = 1-\sum_{k\geq 1}\mathbb{P}(k|p) = 1-c(\xi,r)\sum_{k\geq 1}\mathcal{P}_{NB}(k|r,\xi_p) = 1-\frac{c(\xi,r)}{c(\xi_p,r)}. \qquad (1.23)$$

with ξ_p given in (1.21), thus concluding the proof. □

The fact that the RSA maintains the same functional form at different scales will be central in our framework. We denote this property by say the RSA is *form-invariant*, and mathematically it corresponds to the self-similarity property showed above.

Note that this form-invariance property holds also between two general subscales p and q. Assume $0 < p \leq q \leq 1$. We already analyzed the case $q = 1$ above, and the case $p = q$ reduce to identity. Applying Proposition 1.3.1 for both scales we obtain

$$\xi_p = \frac{p\xi}{1 - \xi(1-p)} \tag{1.24}$$

$$\xi_q = \frac{q\xi}{1 - \xi(1-q)} \tag{1.25}$$

By inverting equation (1.25) to explicit ξ as a function of ξ_q, and inserting the result in (1.24), we want to obtain a formula for ξ_p which depends only on the information at the larger scale q, but not on the global parameters we do not have access to.

$$\begin{aligned} \xi_p &= \frac{p\xi}{1 - \xi(1-p)} = \frac{p\frac{\xi_q}{q+\xi_q(1-q)}}{1 - \frac{\xi_q}{q+\xi_q(1-q)}(1-p)} = \frac{p\xi_q}{q + \xi_q(1-q) - \xi_q(1-p)} \\ &= \frac{p\xi_q}{q - \xi_q(q-p)} = \frac{\frac{p}{q}\xi_q}{1 - \xi_q(1-\frac{p}{q})} \end{aligned} \tag{1.26}$$

Note that equation (1.26) is a more general case of (1.24), that can be interpreted as the particular case $q = 1$. Hence what really matters is the relative ratio of the two scales.

1.4 Power Law tails of the Negative Binomial depending on parameters

Negative Binomial density function with parameters ξ and $r > 0$ results to capture very well empirical RSA patterns in tropical forests [Tov+17; Tov+19a]. However, their characteristic behavior displays exponential tails. When moving to different datasets, especially with the human activity databases analyzed in the following, the RSAs are characterized by heavy tails (see Figure 4.3). We postpone to Chapter 4 the discussion on the origin of this different trait; however, it is a matter of fact that the observed RSAs of some datasets can not be captured by a standard Negative Binomial distribution with $r \in \mathbb{R}^+$. Nevertheless, power law behaviors can be accommodated when allowing the clustering parameter r to take negative values, $r \in (-1, 0)$, thus enabling us to adapt and generalize the theoretical framework we will describe in Chapter 2 to portray regular statistics for human activities and to use activity information on local scale to predict hidden features of the human dynamics at the global scale.

The goal of this Section is to show the cruciality of this extension of the parameter region, i.e. that negative values of $r \in (-1, 0)$ reflect in a Power Law behavior of the RSA's tail with an exponential cutoff, which well describes the observed patterns in human activities datasets. We point our that both parameters intervene in the shape of the RSA, being r responsible for the power law tail with exponent $\alpha = 1 - r$ and ξ for the position of the exponential truncation of the distribution. Note that this section is

purely theoretical, nevertheless the predicted exponent $\alpha = 1 - r$ matches very well our findings when we empirically fit the data.

We start by recalling our truncated negative binomial distribution, already defined in (1.9):

$$\mathbb{P}(n) = c(r, \xi)\binom{n + r - 1}{n}\xi^n(1 - \xi)^r \tag{1.27}$$

The following Theorem belongs to a the class of the so called Tauberian Theorems [FS08]. It relates the tail's decay of a probability density function to the behavior of its generating function near singularities.

Theorem 1.4.1. *Let $Y(z)$ be the generating function of a discrete random variable with density P, and denote by R_Y its dominant singularity. If*

$$Y(z) \sim c_Y \left(1 - z/R_Y\right)^{\beta}, \qquad \text{for } z \to R_Y \tag{1.28}$$

for some $\beta \in \mathbb{R} \setminus \{0, 1, 2, ...\}$, then the distribution $P(n)$ satisfies

$$P(n) \sim \frac{c_Y n^{-\beta-1} R_Y^{-n}}{\Gamma(-\beta)}, \qquad \text{as } n \to \infty \tag{1.29}$$

where c_Y is a constant independent on n, and Γ is the Gamma function.

A proof of the previous Theorem can be found in [Wal+12] and [FS08].

We wish to apply this theorem to our truncated negative binomial distribution (1.27). We start by examining the probability generating function:

$$Y(z) = \sum_{n=0}^{\infty} P(n) z^n \tag{1.30}$$

Taking a closer look to (1.27), observe that the normalizing factor $c(r, \xi)$ does not play any significant role. This is due to the fact that we want to investigate the singularities of $Y(z)$ and the prefactor $c(r, \xi)$ does not affect the result. Moreover, the tail of a truncated Negative Binomial is exactly the same of a standard Negative Binomial, hence we simply disregard the normalizing constant and conduct the analysis for a standard Negative Binomial.

Since we aim at finding the lowest-norm singularity of the probability generating function $Y(z)$, we proceed with the computation replacing $P(n)$ with the standard Negative Binomial $\mathcal{P}_{NB}(n|r, \xi)$:

$$\begin{aligned} Y(z) &= \sum_{n=0}^{\infty} \binom{n + r - 1}{n} \xi^n (1 - \xi)^r z^n \\ &= \sum_{n=0}^{\infty} \binom{n + r - 1}{n} [z\xi]^n (1 - z\xi)^r \cdot \frac{(1 - \xi)^r}{(1 - z\xi)^r} \\ &= \frac{(1 - \xi)^r}{(1 - z\xi)^r} \cdot \sum_{n=0}^{\infty} \binom{n + r - 1}{n} [z\xi]^n (1 - z\xi)^r \end{aligned} \tag{1.31}$$

If now $z\xi < 1$, i.e. for $z < \frac{1}{\xi}$ the sum converges to 1 as we are summing over $\mathbb{N}$ the marginals of a Negative Binomial with parameters r and $z\xi$.

Thus we are left with

$$Y(z) = \frac{(1-\xi)^r}{(1-z\xi)^r} = c_Y(1-z\xi)^{-r} \tag{1.32}$$

It turns out that $Y(z)$ has a singularity at $z = 1/\xi$.

A singularity of a complex function is a point in the complex plane where the function is not analytic. A pole is the prime example, but a square-root branch point and a branch cut are also singularities.

We now want to express $Y(z)$ as in (1.28) to apply the theorem. But in our case

$$Y(z) = c_Y(1-z\xi)^{-r} = c_Y(1-\frac{z}{R_Y})^{\beta} \tag{1.33}$$

with $\beta = -r$ and $R_Y = \frac{1}{\xi}$. The theorem above provides a characterization of the tails of the (truncated) Negative Binomial. The distribution, when n is very large, satisfies

$$P(n) \sim \frac{c_Y n^{r-1}\xi^n}{\Gamma(-\beta)} = \frac{c_Y n^{r-1}e^{n\log(\xi)}}{\Gamma(-\beta)}. \tag{1.34}$$

Note that $\xi < 1$ and $r - 1 < -1$ so we have the multiplication of exponential and power law both approaching zero when n increase to infinity. Hence the distribution resembles a Power Law until n is such that $n \sim -1/\ln(\xi)$. Note that this cutoff depends both on r and on ξ. In particular, the Power Law range is greater for sharper slopes, i.e. for bigger absolute values of $r - 1$, and for values of ξ approaching 1.

Note finally that we justified our decision to remove the prefactor $c(r, \xi)$ from the computation since the tails of a truncated Negative Binomial are not affected by the truncation. However an explicit computation can be done directly for a truncated Negative Binomial in a similar way. We would obtain:

$$\begin{aligned} Y(z) &= c(r,\xi)\sum_{n=1}^{\infty}\binom{n+r-1}{n}\xi^n(1-\xi)^r z^n \\ &= \frac{1}{1-(1-\xi)^r}\frac{(1-\xi)^r}{(1-z\xi)^r}\cdot\sum_{n=1}^{\infty}\binom{n+r-1}{n}[z\xi]^n(1-z\xi)^r \\ &= c_Y\frac{1-(1-z\xi)^r}{(1-z\xi)^r} \\ &= c_Y(1-z\xi)^{-r} - c_Y \end{aligned} \tag{1.35}$$

Now, with the convention $R_Y = 1/\xi$ and $\beta = -r$, we have that

$$Y(z) \sim c_Y(1-z/R_Y)^{\beta}, \qquad \text{for } z \to R_Y \tag{1.36}$$

that again correspond to the requirement (1.28) allowing us to conclude exactly in the same way we did for the standard Negative Binomial.

1.5 Relation of NB with Log-series

Log-series is probably the most widespread distribution in ecology. In this brief Section we want to relate it with our Negative Binomial. The Log-series distribution dates back to 1943 when Fisher discovered it empirically while studying Corbet's tables of butterflies. As done for the Negative Binomial, it can be derived at the stationary solution of the master equation of a birth and death process, with a different choice for the birth rate. Assume now that the population dynamics in the community are governed by ecological drift and random speciation instead of migration from metacommunity. Then one can set the birth rate equal to

$$b_{n,s} = b_s n \tag{1.37}$$

$$b_{0,s} = \nu \tag{1.38}$$

In this case, the reflecting boundary condition $b_{0,s} = \nu$ ensure that, whenever the species s goes extinct, the community do not die forever, with a birth rate that is forced to remain positive. This constraint models the speciation phenomenon, thus it is an acceptable hypothesis even though it lose the interpretation of an extinct species giving birth to a new individual.

Similarly to what done for the Negative Binomial, we want to substitute the specific birth and death rates to the general solution of the master equation of a birth and death process. We obtain

$$\mathcal{P}_{n,s} = P_{0,s} \frac{\nu}{b_s} \frac{x_s^n}{n} \tag{1.39}$$

where, as always, the normalizing constant $P_{0,s}$ can be determined by imposing that

$$1 = \sum_{n=1}^{\infty} \mathcal{P}_{n,s} = P_{0,s} \frac{\nu}{b_s} \sum_{n=1}^{\infty} \frac{x_s^n}{n} = P_{0,s} \frac{\nu}{b_s} [-\log(1 - x_s)] \tag{1.40}$$

which leads to

$$\mathcal{P}_{n,s} = -\frac{1}{\log(1 - x_s)} \frac{x_s^n}{n}. \tag{1.41}$$

If again we assume to join the neutral theory, which corresponds to consider all species as demographically identical, we can drop the index s from all the terms of the above derivation. We obtained that the global RSA is such that every species is such that the probability to have abundance n is given by a (Fisher) Log-Series distribution with parameter $x = b/d$, the odd between the birth and the death rates.

We want now to focus on the relation between the Log-Series distribution and the Negative Binomial. The crucial point is that we can think of the Log-Series as a special case of the Negative Binomial that can be obtained as a limit case when the parameter

r approaches zero.

$$\begin{aligned}\lim_{r\to 0} c(r,\xi)\mathcal{P}_{NB}(n|r,\xi) &= \lim_{r\to 0} \frac{(1-\xi)^r}{1-(1-\xi)^r}\binom{n+r-1}{n}\xi^n \\ &\approx \lim_{r\to 0} \frac{1+r\log(1-\xi)}{-r\log(1-\xi)}\frac{r}{n}\xi^n \\ &= \frac{1}{-\log(1-\xi)}\frac{\xi^n}{n} = \mathcal{P}_{LS}(n|\xi)\end{aligned} \tag{1.42}$$

Hence we can view our Negative Binomial parametrization as an expansion of the great original work of Fisher.

Chapter 2

Statistical models for upscaling from Negative Binomially distributed data

In this Chapter we want to describe the Upscaling method we will use on our Datasets. This Section will only concern the mathematical derivation of the framework we will use to perform our analysis. In the following Chapter we try to convince the reader that the choice of using this particular framework is actually valid.

First of all we want to set up some basic notation. Let N denote the number of individuals at the global scale. The local scale consists of a fraction p of the population, drawn completely at random. The parameter S indicates the number of different species existing globally. We assume the Relative Species Abundance at the global scale to be distributed as a truncated Negative Binomial distribution of parameters r and ξ, i.e.

$$\mathbb{P}(n|1) = c(r,\xi)\mathcal{P}_{NB}(n|r,\xi) = \frac{1}{1-(1-\xi)^r}\binom{n+r-1}{n}\xi^n(1-\xi)^r \tag{2.1}$$

for any $n > 0$. Note that, if we embrace the neutral theory assumption of demographic equivalence between species, then the abundance η_s of any species $s \in \{1,..,S\}$ can be thought as a different realization of the same truncated Negative Binomial distribution of parameters $\xi \in (0,1)$ and $r \in (-1,0)\cup(0,+\infty)$, and hence equation (2.1) holds. We want to emphasize the fact that by allowing r to take values in $(-1,0)$ we are able to extend an existing framework to datasets which display a Power Law tail in their RSAs. Hence this framework can accomodate different RSA behaviors. Moreover, we underline that we are assuming independence between species. This means that, if such a population has to be generated in silico, we can draw the species' abundance from distribution (2.1), leading to a total population of $N = \eta_1 + \cdots + \eta_S$ individuals. This means that we are not imposing any a-priori parameter N, rather it emerges naturally from this procedure.

The second step is to preform a sample at scale p. Intuitively, this means that we are choosing at random a fraction p of individuals. This can be done in many ways. Built-in algorithm solve the problem straightforwardly. What matters is that the individuals at

the local scale remain independent. For example this can be done by selecting or leaving out from the sample each individual with probability p. This leads to a local sample of size n close to pN, but we do not impose perfect odds. In facts, we assume to compute the sample fraction $p' = n/N$ after the sample is realized, and of course we expect to have $p' \approx p$, that is actually what happens. In real application, one desires to obtain a sample of size close to pN, and the parameter actually used for upscaling is the computed p'. We point out this fact here at the beginning, to prevent incomprehension. In order to avoid an abuse of notation, in what follows we assume without loss of generality that the desired fraction p we claim for the sample actually corresponds to p'.

Definition 3. For every $j = 1, ..., S$, we indicate with $\eta_j^{(p)}$, $\eta_j^{(p^c)}$ the abundance of the species j in the observed (resp. unobserved) fraction p (resp. $1-p$) of the population.

At the local scale p we have a total number of $S_p \leq S$ observed species. Note that at scale p we can have $\eta_s^{(p)} = 0$ for some species s. By a simple relabeling, we assume that $\eta_s^{(p)} > 0$ for $s \in \{1, .., S_p\}$ and $\eta_s^{(p)} = 0$ for $s \in \{S_p + 1, .., S\}$. Note finally that at subscale p we only access information to $\eta_1^{(p)}, ..., \eta_{S_p}^{(p)}$, with the parameter S remaining unknown.

2.1 Estimator for the total number of species and SAC

The focus of this Section is to derive an estimator for the total number S of different species in the global population, given the observations at the local scale p. The first step consists on determining the relationship between the total number of species S in the entire population and the number S_p of observed species at subscale p.

We denote by $S_p(k)$ the number of species having exactly k individuals at the local scale p, and analogously we use $S(k)$ to indicate the number of species having exactly k individuals at the global scale. The following simple relation indicates that the ratio of the number of observed species with abundance k is clearly related to the probability of a species to have abundance k at the local scale.

$$\frac{S_p(k)}{S_p} \approx \mathbb{P}(k|p), \qquad k \geq 1 \tag{2.2}$$

In a similar way, the empirical probability that a species of the existing S has null abundance at scale p corresponds to the fraction of unsurveyed species, i.e.

$$\frac{S_p(0)}{S} = \frac{S - S_p}{S} \tag{2.3}$$

On the other hand, we formally obtained the theoretical probability of not observing a species at the local scale, thus we can write that

$$\mathbb{P}(k = 0|p) \simeq \frac{S - S_p}{S} \tag{2.4}$$

where $\mathbb{P}(k = 0|p)$, that we abbreviate in $\mathbb{P}(0|p)$ in the following, was derived in (1.20) and is given by

$$\mathbb{P}(0|p) = 1 - \frac{c(r, \xi)}{c(r, \xi_p)} \tag{2.5}$$

with ξ_p given by

$$\xi_p = \frac{p\xi}{1 - \xi(1 - p)}. \tag{2.6}$$

Hence, rearranging suitably relation (2.4) in order to isolate S, we obtain the relation

$$S \simeq \frac{S_p}{1 - \mathbb{P}(0|p)} \tag{2.7}$$

Note that we are assuming the global RSA to be distributed as a truncated Negative Binomial of parameters r and ξ, whose value is unknown. However, in such a framework, we derived in (1.20) the the abundances of the species observed at the local scale p follows again a truncated Negative Binomial distribution with parameters r and ξ_p. Hence, we are able to fit the parameters $\hat{r}$ and $\hat{\xi}_p$ from the observed local sample. We can also derive an estimation of $\hat{\xi}$ by inverting (2.6), thus having a value for every parameter. An estimator for S can be derived by plugging this procedure into equation (2.7), i.e.

$$\begin{aligned} \hat{S} &\overset{(2.7)}{=} \frac{S_p}{1 - \mathbb{P}(0|p)} \\ &\overset{(2.5)}{=} S_p \frac{1 - (1 - \hat{\xi})^{\hat{r}}}{1 - (1 - \hat{\xi}_p)^{\hat{r}}} \\ &\overset{(2.6)}{=} S_p \frac{1 - \left(1 - \frac{\hat{\xi}_p}{p + \hat{\xi}_p(1-p)}\right)^{\hat{r}}}{1 - (1 - \hat{\xi}_p)^{\hat{r}}} \end{aligned} \tag{2.8}$$

Proposition 2.1.1. *The estimator $\hat{S}$ derived in (2.8) it theoretically unbiased and its variance goes to zero as $p \to 1$.*

Proof. Observe that

$$\begin{aligned} \mathbb{E}\left[\hat{S}\right] &= \frac{1 - \left(1 - \frac{\hat{\xi}_p}{p + \hat{\xi}_p(1-p)}\right)^{r}}{1 - (1 - \hat{\xi}_p)^{r}} \mathbb{E}\left[S_p\right] \\ &= \frac{1 - \left(1 - \frac{\hat{\xi}_p}{p + \hat{\xi}_p(1-p)}\right)^{r}}{1 - (1 - \hat{\xi}_p)^{r}} \mathbb{E}\left[\sum_{i=1}^{S} \mathbb{1}_{\{\eta_i^{(p)} > 0)\}}\right] \\ &= \frac{1 - \left(1 - \frac{\hat{\xi}_p}{p + \hat{\xi}_p(1-p)}\right)^{r}}{1 - (1 - \hat{\xi}_p)^{r}} \sum_{i=1}^{S} \left[1 - \mathbb{P}(\eta_i^{(p)} > 0)\right] \end{aligned} \tag{2.9}$$

$$= \frac{1-\left(1-\dfrac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}{1-(1-\hat{\xi}_p)^r} S \frac{1-(1-\hat{\xi}_p)^r}{1-(1-\hat{\xi})^r} = S$$

hence $\hat{S}$ is an unbiased estimator of S. Note that the distribution of $\eta_i^{(p)}$ does not depend on i because, at the global scale $p = 1$, $\eta_i^{(1)}$, $i = 1, ...S$ are i.i.d. according to $\mathbb{P}(\cdot|1)$. This estimator turns out to be unbiased if $\hat{\xi}_p = \xi_p$ for every subsample, i.e. if for any subsample we were able to recover the "real" local parameter ξ_p. This is not true in our case, where $\hat{\xi}_p$ exhibits a dependence on the finite size subsample. However the dependence on the sample is negligible.
To conclude note that the variance of $\hat{S}$ can expressed as

$$\begin{aligned}
\mathbb{V}ar\left[\hat{S}\right] &= \left[\frac{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}{1-(1-\hat{\xi}_p)^r}\right]^2 \mathbb{V}ar\left[S_p\right] \qquad (2.10)\\
&= \left[\frac{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}{1-(1-\hat{\xi}_p)^r}\right]^2 \mathbb{V}ar\left[\sum_{i=1}^{S} \mathbb{1}_{\{n_i^{(p)}>0)\}}\right]\\
&= \left[\frac{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}{1-(1-\hat{\xi}_p)^r}\right]^2 S\left(\frac{1-(1-\hat{\xi}_p)^r}{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}\right)\left(1-\frac{1-(1-\hat{\xi}_p)^r}{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}\right)\\
&= S\left[\frac{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}{1-(1-\hat{\xi}_p)^r}\right]\left(1-\frac{1-(1-\hat{\xi}_p)^r}{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^r}\right)
\end{aligned}$$

and it is easy to show that the expression in the last line goes tp 0 as $p \to 1$. □

Thus we can use equation (2.8) to estimate the total number of species given a subsample at scale p.
Note that we can do more. For any $q \in (p, 1)$ we can apply the same chain of equations with some slight modifications to estimate $\hat{S}_q$. To be precise, for any $q \geq p$ we would obtain

$$\hat{S}_q = S_p \frac{1-\left(1-\dfrac{\xi_p}{\frac{p}{q}+\xi_p(1-\frac{p}{q})}\right)^{\hat{r}}}{1-(1-\hat{\xi}_p)^{\hat{r}}} = S_p \frac{1-\left(\dfrac{p\left(1-\hat{\xi}_p\right)}{p+\hat{\xi}_p\,(q-p)}\right)^{\hat{r}}}{1-(1-\hat{\xi}_p)^{\hat{r}}} \qquad (2.11)$$

Hence, since for $q \in (0, p)$ we can measure directly S_q by subsampling, we obtained an explicit formula describing the behavior of the Species-Area-Curve (SAC) for every $q \leq 1$. This parametric curve describes the total number of different species that we can observe when varying the sample size.

Moreover we can recover the RSA at the global scale by plugging the estimated parameters $\hat{\xi}$ and $\hat{r}$ into (2.1). i.e.

$$\begin{aligned}\mathbb{P}(n|1) &= c\left(\hat{r}, \hat{\xi}(p, \hat{\xi}_p)\right) \mathcal{P}_{NB}\left(n|\hat{r}, \hat{\xi}(p, \hat{\xi}_p)\right) \\ &= \frac{1}{1-\left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^{\hat{r}}} \binom{n+r-1}{n} \left(\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^n \left(1-\frac{\hat{\xi}_p}{p+\hat{\xi}_p(1-p)}\right)^{\hat{r}}\end{aligned} \tag{2.12}$$

2.2 Upscaling from Occurrences Data

In this Section we modify our method in order to accomplish the same upscaling goal when disposing of a different type of datas. In facts, many databases give no information about the abundances of a species within an area, but only its occurrence in each of the surveyed plots. We derive a method able to infer species richness and abundances at large spatial scales in biodiversity-rich ecosystems when species presence/absence information is available on various scattered samples, by using the same framework introduced above.

Recall once again that we assume the RSA to have a Negative Binomial functional form, i.e.

$$\mathbb{P}(n|1) = c(r,\xi)\mathcal{P}_{NB}(r,\xi) = \frac{1}{1-(1-\xi)^r}\binom{n+r-1}{n}\xi^n(1-\xi)^r. \tag{2.13}$$

where $c(r,\xi)$ is the normalizing constant, taking into account that every of the S existing species has a strictly positive abundance at the global scale. If we assume a certain species to have abundance n at the global scale, and we sample globally at random a fraction p of individuals, then, the probability to have k individuals of that species in the sample is distributed according to a binomial distribution:

$$\mathcal{P}_{binom}(k|n,p) = \binom{n}{k}p^k(1-p)^{n-k}, \qquad k = 0, ..., n. \tag{2.14}$$

As shown in Proposition 1.3.1, the local RSA at the sample scale p is again distributed proportionally to a Negative Binomial, with rescaled parameter ξ_p and same r:

$$\mathbb{P}(k|p) = \begin{cases} c(r,\xi)\cdot\mathcal{P}_{NB}(k|r,\xi_p) & k \geq 1 \\ 1 - c(r,\xi)/c(r,\xi_p) & k = 0 \end{cases} \tag{2.15}$$

with

$$\xi_p = \frac{p\xi}{1-\xi(1-p)}. \tag{2.16}$$

Note that, as we showed in (1.26), we can derive the relation for any two different scales, only thing that matters being their ratio.

We assume now to dispose of multiple small samples, where we can only access to the presence/absence information for any species in any area. We indicate by $p^\star$ the sample

size, i.e. the fraction of individual we surveyed. In the following we will use $S^\star$ as an abbreviation for $S_{p^\star}$, the number of different species at the sample size $p^\star$. As always, we label the different species with a different number $s \in \{1, ..., S^\star\}$.
Next, we split our sample into n subsamples, each of them consisting of the same number of individuals. We assume to have presence/absence information for every species in any of the n subsamples. Note that, in practice, the procedure works the other way round. It happens to dispose of n different samples with presence/absence information, which in turn will be gathered to form a single bigger sample. That is the reason lying behind our assumption to actually have access to the presence/absence information in any of the subsamples.

We build the binary matrix $\Omega = \left(\omega_s^i\right)_{s=1,..,S^\star}^{i=1,..,n}$, where the entry ω_s^i gives information on the presence/absence of the species s in the i-th subsample, i.e. $\omega_s^i = 1$ if and only if the species s is present in the i-th subsample. At this point we define the subscales $p_k := \frac{k}{n} p^\star$ for $k = 1, ..., n$; and we empirically compute the number of species at the subscale p_k by averaging the number of observed species when gathering k of the n different subsamples. We obtain:

$$\hat{S}_{p_k} := \frac{1}{\binom{n}{k}} \sum_{\substack{I \subseteq \{1,..,n\} \\ |I| = k}} \sum_{s=1}^{S^\star} \mathbb{1}\left(\sum_{i \in I} \omega_s^i \geq 1\right), \tag{2.17}$$

where $\mathbb{1}(X)$ is the indicator function, which equals one when the random event X happens and it is zero otherwise.
Note that the RSA at scale p_k is again proportional to a Negative Binomial, with same r and rescaled ξ_{p_k} that we can express as a function of $\xi_{p^\star}$ as shown in (1.26), i.e.

$$\xi_{p_k} = \frac{\frac{p_k}{p^\star}\xi_{p^\star}}{1 - \xi_{p^\star}(1 - \frac{p_k}{p^\star})} =: U(p_k \mid p^\star, \xi_{p^\star}) \tag{2.18}$$

From our choice of p_k, it turns out that $p_k/p^\star = k/n$, but the function U defined in (2.18) is well defined for general choice of $p^\star$ and p_k.

From our general framework we derived in (2.11) a relation linking the number of existing species at different scales. Expressing this relation for our case, we obtain:

$$S_{p_k} = S^\star \frac{1 - (1 - U(p_k \mid p^\star, \xi_{p^\star}))^r}{1 - (1 - \xi_{p^\star})^r}. \tag{2.19}$$

Note that the only unknown parameters at the right end side of the latter equation are $\xi_{p^\star}$ and r, hence we can think of S_{p_k} as a parametric curve depending on the parameters r and $\xi_{p^\star}$. The idea is to fit this curve with the empirical values in (2.17). Once obtained the empirical values for $\hat{\xi}_{p^\star}$ and $\hat{r}$, we can use them to depict the RSA at the local scale $p^\star$. At this point we have all the ingredients to apply the general upscaling method described at the beginning of this Chapter.

The key feature of the method is the possibility, given only presence/absence data, to connect and infer different biodiversity patterns at the global scale. Indeed, we can

predict, in addition to the SAC and the RSA, also the Relative Species Occurrence (RSO). The RSO is another important pattern which gives the probability that a species occupies a given number of cells at the global scale, given that the forest can be tiled into a given number of M equal sized cells of a given area. This is a bit different to what stated before, since we are introducing the concept of species occupying a given area. In ecological application, it is very common to refer to areas surveyed rather than number of individuals, since we are assuming, in homogeneous forests, the fraction of sampled area to be proportional to the number of individuals observed. Moving away from this ecological example, we can think of a cell as a subset of individuals of a given cardinality.

Assume now to split the global dataset into M different cells of equal size a. For any $v \leq M$, the RSO returns the probability that a species occupies v cells at the global scale. In order to find an expression for it, we firstly need the probability $Q_{occ}(v \mid n, M, 1)$ that a species occupies v over M cells at the global scale $p = 1$, given that is has abundance n. Under the mean field hypothesis, this is given by an hyper-geometric distribution

$$Q_{occ}(v \mid n, M, 1) = \frac{\binom{M}{v}\binom{n-1}{v-1}}{\binom{n+M-1}{M-1}} \tag{2.20}$$

The RSO distribution $Q(v \mid M, 1)$ can thus be obtained by marginalizing with respect to the abundance n:

$$Q(v \mid M, 1) = \sum_{n=1}^{\infty} Q_{occ}(v \mid n, M, 1)\mathbb{P}(n|1), \tag{2.21}$$

where $\mathbb{P}(n|1)$ is the global RSA given by equation (2.13).

2.3 Variation of popularity across scales

The second innovation that we are going to introduce in our work is a method to estimate the variation of popularity. Until now we studied the distribution of the abundances of the observed species at the local scale, but we estimated only the number of unseen species, disregarding of their abundances. We can describe our upscaling method with the following steps. As first we need to introduce a statistics, indicating the observed species at sample scale p:

$$S_p = \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)} > 0\}} \tag{2.22}$$

Next, we compute the mean of the statistics:

$$\mathbb{E}\left[S_p\right] = \mathbb{E}\left[\sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)} > 0\}}\right] = \sum_{j=1}^{S} \mathbb{E}\left[\mathbb{1}_{\{\eta_j^{(p)} > 0\}}\right] \tag{2.23}$$

$$= \sum_{j=1}^{S} \mathbb{P}\left(\eta_j^{(p)} > 0\right) = S \cdot \mathbb{P}\left(k > 0|p\right) = S \cdot \left[1 - \mathbb{P}\left(k = 0|p\right)\right]$$

Arranging the latter equation we isolated the quantity we are interested to estimate:

$$S = \frac{\mathbb{E}\left[S_p\right]}{1 - \mathbb{P}\left(0|p\right)} \tag{2.24}$$

Finally, an estimator of S is obtained replacing the mean $\mathbb{E}\left[S_p\right]$ by the observable $\hat{S}_p$:

$$\hat{S} = \frac{\hat{S}_p}{1 - \mathbb{P}\left(0|p\right)} = \hat{S}_p \frac{1 - \left(1 - \frac{\hat{\xi}_p}{p + \hat{\xi}_p(1-p)}\right)^{\hat{r}}}{1 - (1 - \hat{\xi}_p)^{\hat{r}}} \tag{2.25}$$

where we adopted the notation $\hat{S}_p$ to emphasize that the number of different species observed locally depends on the specific sample. With no surprise, we recover the same result as in (2.8). We want to stress that this new formulation allows us to push further our investigation, when applying the same procedure to different statistics.

Note that (2.24) and (2.25) share the same term $\mathbb{P}(0|p)$. However the two probability are different. To be precise, in (2.24) $\mathbb{P}(0|p)$ is given by formula (2.5) where ξ and r are the real global parameters, and ξ_p is given by (2.6). On the other hand, in (2.25) we assume to know no other information than what observed at the sample scale. Hence the analytical form of $\mathbb{P}(0|p)$ is again given by formula (2.5), but the parameters are given by $\hat{\xi}_p$ and $\hat{r}$ (fitted from the local data), and $\hat{\xi}$ (obtained by inverting (2.6)). Hence in (2.25) the parameters $\hat{\xi}_p$ and $\hat{r}$ that appear in the denominator correlate with $\hat{S}_p$. Nevertheless, $\hat{\xi}_p$ and $\hat{r}$ turn out to be really close to the real value ξ_p and r, for any random subsample.

2.3.1 Estimator for the "new" species

Recall that we are sampling S_p species at scale p from a pool consisting of N individuals spread into S different species. If a species j is not observed in the sample at scale p, we say that j is a "new" species. The meaning of this definition can be easily explained. If you imagine to further sample your population, you can pick individuals belonging to species already observed or you can discover indeed "new" species.
Consider then the following statistics for the new species:

$$S_{new}^{(p^c)} = \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=0, \eta_j^{(p^c)}>0\}} \tag{2.26}$$

The following chain of equality turns out to be meaningful in the following:

$$\begin{aligned} S_{new}^{(p^c)} &= \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=0, \eta_j^{(p^c)}>0\}} = \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=0, \eta_j^{(1)}>0\}} \\ &= \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=0\}} = \sum_{j=1}^{S} \left(1 - \mathbb{1}_{\{\eta_j^{(p)}>0\}}\right) = S - S_p \end{aligned} \tag{2.27}$$

We can recover an estimator for the "new" species from the known estimator for S. This remark seems trivial, and the chain of equation above appears redundant. Nevertheless it is crucial for the development of our work. We stress that the statistics $S_{new}^{(p^c)}$ in (2.26) uses both the information at the sample scale and the information contained in the unseen fraction of the population, whereas the statistics for S_p in (2.22) considers only the observed individuals.

Consider again the statistics (2.26) that represents the number of unobserved species in the sample of size p, that are present in the remaining population of size $1-p$.
We want to recover an estimator for the new species. We then compute the expected value of the statistics:

$$\begin{aligned}\mathbb{E}\left[S_{new}^{(p^c)}\right] &= \mathbb{E}\sum_{j=1}^{S}\mathbb{1}_{\{\eta_j^{(p)}=0,\eta_j^{(p^c)}>0\}} = S\cdot\mathbb{P}\left(\eta_j^{(p)}=0,\eta_j^{(p^c)}>0\right) \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=0,\eta_j^{(1)}>0\right) = S\cdot\mathbb{P}\left(\eta_j^{(p)}=0\right) = S\cdot\mathbb{P}(0|p)\end{aligned} \tag{2.28}$$

The expected value turns out to be a product of two factors: $\mathbb{P}(0|p)$ is a probability that we know from (1.20) and hence we can compute, while S is an unknown quantity, but again we can estimate it with $\hat{S}$ as derived in (2.25). Combining pieces together, we derived the following estimator:

$$\hat{S}_{new}^{(p^c)} = \frac{S_p}{1-\mathbb{P}(0|p)}\cdot\mathbb{P}(0|p) \tag{2.29}$$

This procedure capture the techniques that we want to use to derive more estimators. This turning point leads us to new statistics that consider also the popularity.

2.3.2 Estimator for the new species with popularity

We start from the statistics:

$$S^{(p^c)}(k) = \sum_{j=1}^{S}\mathbb{1}_{\{\eta_j^{(p)}=0,\eta_j^{(p^c)}=k\}} \tag{2.30}$$

representing the species that are not present in the local sampled population at scale p, and having abundance k in the remaining part of the population. Note that if we obtain this, then we can easily extend the result to the statistics that consider the local unseen species with abundance at least k in the remaining part of the population, namely

$$S^{(p^c)}(k^+) = \sum_{j\geq k} S^{(p^c)}(j) \tag{2.31}$$

thus the previous section can be included here, simply noticing that:

$$S_{new}^{(p^c)} = S^{(1-p)}(1^+) = \sum_{k\geq 1} S^{(p^c)}(k) \tag{2.32}$$

We proceed, as did in (2.28), by computing the expected value of (2.30):

$$\begin{aligned}\mathbb{E}\left[S^{(p^c)}(k)\right] &= \mathbb{E}\left[\sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=0,\eta_j^{(p^c)}=k\}}\right] \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=0,\eta_j^{(p^c)}=k\right) \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=0,\eta_j^{(1)}=k\right) \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=0\mid\eta_j^{(1)}=k\right)\cdot\mathbb{P}\left(\eta_j^{(1)}=k\right)\end{aligned} \tag{2.33}$$

where we used that $\mathbb{P}\left(\eta_j^{(p)}=x,\eta_j^{(p^c)}=y\right)=\mathbb{P}\left(\eta_j^{(p)}=x,\eta_j^{(1)}=x+y\right)$.
Equation (2.33) reduces to a product of three factors: $\mathbb{P}\left(\eta_j^{(p)}=0\mid\eta_j^{(1)}=k\right)=(1-p)^k$, from the binomial distribution (1.18); $\mathbb{P}\left(\eta_j^{(1)}=k\right)=\mathbb{P}(k|1)$ is known and given by (2.1); S is unknown, and we need an estimator for it.
We can use again the previous results (2.25) to define $\hat{S}$, thus obtaining

$$\hat{S}^{(p^c)}(k)=\hat{S}\cdot(1-p)^k\cdot\mathbb{P}(k|1)=\frac{S_p}{1-\mathbb{P}\left(0|p\right)}\cdot(1-p)^k\cdot\mathbb{P}(k|1) \tag{2.34}$$

This is the estimator for the new species with abundance k.

2.3.3 Estimator for popularity change

In (2.34) we obtained an estimator for the popularity of the new species. This represent a partial result, predicting the abundance of unseen species, i.e. species with abundance zero at the sample scale. We want to extend this reasoning, aiming to predict the abundance of species in the remaining part of the population, also for those species with positive abundance $\ell>0$ in the observed sample.
We start from the statistics:

$$S^{(p^c)}(\ell\to k)=\sum_{j=1}^{S}\mathbb{1}_{\{\eta_j^{(p)}=\ell,\eta_j^{(p^c)}=k\}} \tag{2.35}$$

In this way, we are monitoring all the possible changes in popularity. Note that we can compute also intervals of abundances by summing on different values of l and k. We proceed by computing the expected value:

$$\begin{aligned}\mathbb{E}\left[S^{(p^c)}(\ell\to k)\right] &= \mathbb{E}\left[\sum_{j=1}^{S}\mathbb{1}_{\{\eta_j^{(p)}=\ell,\eta_j^{(p^c)}=k\}}\right] \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=\ell,\eta_j^{(p^c)}=k\right) \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=\ell,\eta_j^{(1)}=k+\ell\right) \\ &= S\cdot\mathbb{P}\left(\eta_j^{(p)}=\ell\mid\eta_j^{(1)}=k+\ell\right)\cdot\mathbb{P}\left(\eta_j^{(1)}=k+\ell\right).\end{aligned} \tag{2.36}$$

As in (2.33), we reduce ourselves to a product of three factors: $\mathbb{P}\left(\eta_j^{(p)} = \ell \mid \eta_j^{(1)} = k+\ell\right) = \binom{k+\ell}{\ell} p^\ell (1-p)^k$, from the binomial distribution (1.18); $\mathbb{P}\left(\eta_j^{(1)} = k+\ell\right) = \mathbb{P}(k+\ell|1)$, from (1.9); S is unknown, we need an estimator for it and we can use $\hat{S}$ from (2.25). Hence we obtained

$$\begin{aligned}\hat{S}^{(p^c)}(\ell \to k) &= \hat{S} \cdot \mathbb{P}\left(\eta_j^{(p)} = l | \eta_j^{(1)} = k+l\right) \cdot \mathbb{P}(k+\ell|1) \\ &= \frac{S_p}{1-\mathbb{P}(0|p)} \cdot \binom{k+\ell}{\ell} p^\ell (1-p)^k \cdot c(r,\xi) \binom{k+\ell+r-1}{k+\ell} \xi^{k+\ell}(1-\xi)^r\end{aligned} \tag{2.37}$$

The estimator $\hat{S}^{(p^c)}(\ell \to k)$ in (2.37) predicts the number of species with abundance k in the unobserved fraction $1-p$ of the population and abundance ℓ in the sample at scale p. Note that this estimator is independent of the effective number of species with abundance ℓ at scale p; indeed we are using the sample at scale p only to estimate the parameters $\hat{\xi}_p$ and $\hat{r}$, that we need to predict $\hat{S}$ and to derive $\hat{\xi}$. Hence we are using only partial information of the local scale.

2.3.4 Conditional Estimator for popularity change

Since we can observe the sample at scale p, we want to take this information into account. We want to build an estimator for the species with abundance k in the unobserved fraction $1-p$ of the population, given that they have abundance ℓ in the sample at observed scale p. We want to take into account the information on the number of specie we observe with abundance ℓ at local observed scale p. We define

$$S_p(\ell) := \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)} = \ell\}}. \tag{2.38}$$

that indicates the number of species with abundance ℓ at scale p. The Bayes' formula allows us to obtain explicit expressions of quantities that we will need in the following.

$$\begin{aligned}\mathbb{P}\left[\eta_j^{(p^c)} = k \mid \eta_j^{(p)} = \ell\right] &= \mathbb{P}\left[\eta_j^{(1)} - \eta_j^{(p)} = k \mid \eta_j^{(p)} = \ell\right] \\ &= \mathbb{P}\left[\eta_j^{(1)} - \ell = k \mid \eta_j^{(p)} = \ell\right] \\ &= \mathbb{P}\left[\eta_j^{(1)} = k+\ell \mid \eta_j^{(p)} = \ell\right] \\ &= \frac{\mathbb{P}\left[\eta_j^{(p)} = \ell \mid \eta_j^{(1)} = k+\ell\right] \mathbb{P}\left[\eta_j^{(1)} = k+\ell\right]}{\mathbb{P}\left[\eta_j^{(p)} = \ell\right]}\end{aligned} \tag{2.39}$$

Note that we know the probabilities that appear in the latter formula, since:

- $\mathbb{P}\left[\eta_j^{(p)} = \ell \mid \eta_j^{(1)} = k+\ell\right] = \binom{k+\ell}{\ell} p^\ell (1-p)^k$ is the binomial distribution we already made repeatedly use of.

- $\mathbb{P}\left[\eta_j^{(1)} = k+\ell\right] = \mathbb{P}(k+\ell|1) = c(r,\xi)\binom{k+\ell+r-1}{k+\ell}\xi^{k+\ell}(1-\xi)^r$ is the global truncated Negative Binomial distribution as in (1.9).
- $\mathbb{P}\left[\eta_j^{(p)} = \ell\right] = \mathbb{P}(\ell|p) = c(r,\xi)\binom{\ell+r-1}{\ell}\xi_p^{\ell}(1-\xi_p)^r$ is again a truncated Negative Binomial with rescaled parameter ξ_p as in (1.20).

We retrace the same steps used for the conditional estimator. We start from the statistics

$$S^{(p^c)}(k \mid \ell) = \sum_{j=1}^{S} \mathbb{1}_{\{\eta_j^{(p)}=\ell\}} \mathbb{1}_{\{\eta_j^{(p^c)}=k, \eta_j^{(p)}=\ell\}} = \sum_{j=1}^{S_p(\ell)} \mathbb{1}_{\{\eta_j^{(p^c)}=k|\eta_j^{(p)}=\ell\}} \tag{2.40}$$

where we assume that the species with abundance ℓ at scale p are exactly those with label $j \in \{1, ..., S_p(\ell)\}$. We then proceed by computing the expected value

$$\begin{aligned}\mathbb{E}\left[S^{(p^c)}(k \mid \ell)\right] &= \mathbb{E}\left[\sum_{j=1}^{S_p(\ell)} \mathbb{1}_{\{\eta_j^{(p^c)}=k|\eta_j^{(p)}=\ell\}}\right] = S_p(l) \cdot \mathbb{P}\left(\eta_j^{(p^c)} = k \mid \eta_j^{(p)} = \ell\right) \\ &= S_p(\ell) \cdot \frac{\mathbb{P}\left(\eta_j^{(p)} = \ell \mid \eta_j^{(1)} = k+\ell\right)\mathbb{P}\left(\eta_j^{(1)} = k+\ell\right)}{\mathbb{P}\left(\eta_j^{(p)} = \ell\right)}\end{aligned} \tag{2.41}$$

Note that empirically $\mathbb{P}\left(\eta_j^{(p)} = \ell\right) = S_p(\ell)/S$, and we recover $\mathbb{E}\left[S^{(1-p)}(\ell \to k)\right]$.
To obtain an estimator for $S^{(p^c)}(k \mid \ell)$, we replace the right hand side of the latter formula (2.41) with the probabilities computed by using the parameters obtained from the fitting, i.e.

$$\hat{S}^{(p^c)}(k \mid \ell) = S_p(\ell) \cdot \frac{\binom{k+\ell}{\ell}p^{\ell}(1-p)^k \cdot c(r,\hat{\xi})\binom{k+\ell+r-1}{k+\ell}\hat{\xi}^{k+\ell}(1-\hat{\xi})^r}{c(r,\hat{\xi})\binom{\ell+r-1}{\ell}\hat{\xi}_p^{\ell}(1-\hat{\xi}_p)^r} \tag{2.42}$$

Proposition 2.3.1. *The estimator $\hat{S}^{(p^c)}(k|\ell)$ derived in (2.42) it theoretically unbiased.*

Proof. Observe that

$$\hat{S}^{(p^c)}(k \mid \ell) = S_p(\ell) \cdot \frac{\binom{k+\ell}{\ell}p^{\ell}(1-p)^k \cdot c(r,\hat{\xi})\binom{k+\ell+r-1}{k+\ell}\hat{\xi}^{k+\ell}(1-\hat{\xi})^r}{c(r,\hat{\xi})\binom{\ell+r-1}{\ell}\hat{\xi}_p^{\ell}(1-\hat{\xi}_p)^r} \tag{2.43}$$

We need to verify that:

$$\mathbb{E}\left[\hat{S}^{(p^c)}(k|\ell)\right] = \mathbb{E}\left[S^{(p^c)}(k|\ell)\right] \tag{2.44}$$

Note that since the object $\hat{S}^{(p^c)}(k|\ell)$ is a random variable, we are testing if mean of our estimator reflects the mean of $S^{(p^c)}(k|\ell)$.
From the computation above, we know that:

$$\mathbb{E}\left[\hat{S}^{(p^c)}(k|\ell)\right] = \mathbb{E}\left[S_p(\ell)\cdot \frac{\binom{k+\ell}{\ell}p^\ell(1-p)^k\cdot c(r,\hat{\xi})\binom{k+\ell+r-1}{k+\ell}\hat{\xi}^{k+\ell}(1-\hat{\xi})^r}{c(r,\hat{\xi})\binom{\ell+r-1}{\ell}\hat{\xi}_p^\ell(1-\hat{\xi}_p)^r}\right] \tag{2.45}$$

$$\mathbb{E}\left[S^{(p^c)}(k|\ell)\right] = S\cdot\binom{k+l}{l}p^l(1-p)^k\cdot c(r,\xi)\binom{k+l+r-1}{k+l}\xi^{k+l}(1-\xi)^r \tag{2.46}$$

So we are left to prove that the following equality holds true:

$$\mathbb{E}\left[\frac{S_p(\ell)}{c(r,\hat{\xi})\binom{\ell+r-1}{\ell}\hat{\xi}_p^\ell(1-\hat{\xi}_p)^r}\right] = S. \tag{2.47}$$

But this is true, indeed we have that

$$\begin{aligned}\mathbb{E}\left[S_p(\ell)\right] &= \mathbb{E}\left[\sum_{j=1}^{S}\mathbb{1}_{\{n_j^{(p)}=\ell\}}\right] = \sum_{j=1}^{S}\mathbb{P}\left(n_j^{(p)}=\ell\right) \\ &= S\cdot\mathbb{P}(\ell|p) = S\cdot c(r,\hat{\xi})\binom{\ell+r-1}{\ell}\hat{\xi}_p^\ell(1-\hat{\xi}_p)^r\end{aligned} \tag{2.48}$$

□

Note that, again, we can pass from punctual estimation to cumulative ones. In order to estimate the number of species with abundance in a certain range $[k_{min}, k_{max}]$, given that they have abundance in the range $[\ell_{min}, \ell_{max}]$ in the observed sampled scale, we only need to sum over $k \in [k_{min}, k_{max}]$ and over $\ell \in [\ell_{min}, \ell_{max}]$ the terms on the right hand side of (2.42). For example we can define

$$\hat{S}^{(p^c)}(K^+ \mid L^+) = \sum_{\ell\geq L}\sum_{k\geq K}\hat{S}^{(p^c)}(k \mid \ell) \tag{2.49}$$

and this is exactly the estimator we are going to test on our databases in Chapter 4.

Chapter 3

Inferring macro-ecological patterns from local species' occurrences

3.1 Biodiversity estimators from presence/absence local data

The problem of inferring total biodiversity when only scattered samples are observed is a long-standing problem. Recently, for instance, a semi-analytical method has been proposed to upscale species richness assuming a Fisher Log-series as the species-abundance distribution (RSA) [Sli+15]. The Log-series distribution is often used to describe RSA patterns in many different ecological communities, characterized by high biodiversity [Aza+16]. Thanks to the availability and reliability of the species abundance data in forests (given by systematic and periodic field campaigns and high detectability of species), this method has been typically applied to tropical forests. The robustness of the upscaling method relies on the stability property of Fisher's α parameter, usually named the diversity index, which it ought not to depend on the forest sample size.

The method proposed in [Sli+15] is composed of three main steps: 1) Fisher's α is calculated assuming that the species have a Log-series distribution, and using as input the observed species S_p and number of trees N_p. 2) The total number of stems N for the whole area of interest is extrapolated (this is not a trivial task and there is no consensus on the best methods to implement it. Generally, constant average stem density is assumed [Sli+15], so that $N_p = Np$). 3) Estimate the number of species at the largest scale using the formula $S = \alpha \ln(1 + N/\alpha)$ [FCW43].

On the basis of theoretical and computational analysis as well as using data from 15 tropical forests located all over the globe, it has been shown in [Tov+17] that the Log-series method described above suffers from important limitations. The RSA, especially at large scales or with increasing sampling effort [Chi07], often displays an interior mode [Aza+16], which a Log-series cannot capture. Indeed, the Fisher's distribution is not flexible enough [Aza+15] to describe different RSA patterns found in tropical forests [Mag05; Vol+07; Mag13; Aza+16].

The observed RSA at the sample scale is a common input for many analytical methods have been proposed to upscale species richness. These methods have been proved

to typically perform better than non-parametric estimators of biodiversity [Cha05]. In contrast with the former, non-parametric approaches do not assume a specific family of probability distributions. In particular, non-parametric methods do not make any assumption on the RSA distribution and they thus perform no fit of empirical patterns, rather they only take into account rare species, which are intuitively assumed to carry all the needed information on the undetected species in a sample.
Nevertheless, all the aforementioned methods need abundance data in order to infer biodiversity at larger scale. However, in many open-access databases (e.g. species abundance data obtained from metagenomics) this information is highly imprecise, if available at all. Indeed, there are lots of datasets which give only information about the presence or absence of a species in different surveyed plots, without specifying the number of individuals within them. Some non-parametric approaches have been generalized to infer species richness from this presence-absence data [Cha05; CC16].

Table 3.1 summarizes the most popular estimators and for each one details the predicted biodiversity as a function of the input data. However, most of them have the strong limitation that they do not have an explicit dependence of the observation scale, leading to poor estimates of the number of species at the global scales. Moreover, with the exception of $Chao_{\text{wor}}$, in all the methods listed in Table 3.1 the prediction for $\hat{S}$, given the observed sample at scale p, does not converge to the actual value of S as $p \to 1$. This is due to the fact that the methods do not have an explicit dependence on the surveyed area, rather they give an upscaled biodiversity estimate only based on the number of singletons or doubletons (see Figure 3.2).

The only estimator which takes into account the ratio between the surveyed area and the global one is the one introduced by Chao [Cha05; Cha+09; CC16] and denoted here as $Chao_{\text{wor}}$ (see Table 3.1). This method takes into account the number of species detected in one sample only and those detected in exactly two samples observed at the sample scale to infer the total species richness at the whole forest scale. However, it has been shown that Chao's method, although giving reliable species estimates, it does not properly capture the empirical Species Accumulation Curve (SAC) [Tov+17], which describes how the number of species changes across spatial scales.
Moreover, both parametric and non-parametric methods proposed in the literature do not give any insights on the species abundance at both local or larger scales. Indeed the problem of relating occupancy data with information on species abundance is a relevant issue in theoretical ecology [RN03; Eli+06]. In particular, given the information on the presence or absence of a species in different scattered plots, one would like to infer its population size or, more generally, the RSA distribution of the forest.

The first line of Table 3.1 is referred to our method based on the Negative Binomial parametrization for the RSA. This flexible analytical method provides, from local presence/absence information, robust estimates of species richness and important macroecological patterns of biodiversity (SAC, RSA, RSO), as tested in both in-silico generated and two rainforests. The method is not specific for forests, and it may be applied to any database in the form of a binary matrix, where presence/absence features (tree species in our case) are detected across different samples.

Estimator	**Predicted S**	**Details**
NB	$S^* \dfrac{1-(1-\xi)^r}{1-(1-\hat{\xi}_{p^*})^r}$	$(\xi,\ r)$ NB parameters at $p=1$ $(\hat{\xi}_{p^*},\ r)$ NB parameters at p^*
Chao_2	$S^* + \begin{cases} \dfrac{M^*-1}{M^*}\dfrac{Q_1^2}{2Q_2} & Q_2>0 \\ \dfrac{M^*-1}{M^*}\dfrac{Q_1(Q_1-1)}{2} & Q_2=0 \end{cases}$	Q_i = number of species detected in i plot at the scale p^*
iChao_{i2}	$S_{Chao_2} + \dfrac{M^*\text{-}3}{4M^*}\dfrac{Q_3}{Q_4}\left(Q_1\text{-}\dfrac{(M^*\text{-}3)Q_2Q_3}{2(M^*\text{-}1)Q_4}\right)^+$	$S_{Chao_2} = S$ predicted by $Chao_2$ method
Chao_{wor}	$S^* + \dfrac{Q_1^2}{\frac{M^*}{M^*-1}2Q_2 + \frac{p^*}{1-p^*}Q_1}$	
Jacknife_1	$S^* + \dfrac{M^*-1}{M^*}Q_1$	
Jacknife_2	$S^* + \dfrac{2M^*-3}{M^*}Q_1 - \dfrac{(M^*-2)^2}{M^*(M^*-1)}Q_2$	
Turing	$S^*_{abun} + \dfrac{S^*_{rare}}{\hat{C}_{rare}}$	$S^*_{abun} = \sum_{n>10} Q_n$ $S^*_{rare} = \sum_{n=1}^{10} Q_n$ $\hat{C}_{rare} = 1 - Q_1/\sum_{n=1}^{10} nQ_n$
ICE	$S_{Turing} + \dfrac{Q_1}{\hat{C}_{rare}}\hat{\gamma}^2_{rare}$	$\hat{\gamma}^2_{rare} = (\gamma-1)^+$, where $\gamma = \dfrac{S^*_{rare}C_{rare}}{\hat{C}_{rare}(C_{rare}-1)}Q$ $Q = \dfrac{\sum_{n=1}^{10} n(n-1)Q_n}{(\sum_{n=1}^{10} nQ_n)(\sum_{n=1}^{10} nQ_n\text{-}1)}$ C_{rare} = # of samples with at least 1 rare species

Table 3.1: Summary table of the most popular biodiversity estimators for presence/absence data. In formulas, M^* is the total number of sampled cells. See [Cha05; CC16] for more details about non-parametric methods.

3.2 Implementation of the framework

In this work we proposed and tested a novel rigorous statistical framework to upscale ecological biodiversity patterns from local information on species occurrence data. Indeed, many databases give no information about the abundances of a species within an area, but only its occurrence in each of the surveyed plots. We introduced in Section 2.2 the analytical framework which we will exploit to infer species richness and abundances at large spatial scales in biodiversity-rich ecosystems when species presence/absence information is available on various scattered samples.
The underlying hypotheses that we need in order to perform these estimates is that the RSA at a given scale is distributed according to a Negative Binomial, a simple and versatile distribution that depending on its parameters can display an interior mode or Log-series like behavior, i.e. it can accommodate different RSA shapes. Therefore we can use the same RSA function to reproduce different ecosystems' RSA, as those typically observed in real ecosystems [Mag05; Vol+07; Chi07; Mag13; Aza+16]. Even more generally, by using mixtures of Negative Binomials, a case for which our framework still works, we could fit more complex RSA shapes [Tov+17].

The upscaling method is based on the form-invariance property of the Negative Binomial, as described in Section 2.2. Our approach allows to infer and link within a unique framework important and well-known biodiversity patterns of ecological theory, such as the Species Accumulation Curve (SAC) and the Relative Species Abundance (RSA) as well as a new emergent pattern, which is the Relative Species Occupancy (RSO). In facts, by solely fitting the SAC curve, we can directly obtain the r and ξ parameters, which well describe both the RSA and the RSO distributions at all spatial scales of interest.
The RSO is a new descriptor/measure of biodiversity within an ecological community which describes the distribution of species occurrences in scattered plots. Its distribution displays a fat tail, indicating that many species typically occupies only few scattered plots, while only very few species are pervasive and are found in most of the plot. Our prediction is that this property is not particular for the dataset here considered, rather it is another emergent patterns [Suw+13; Aza+16] pervasive in highly biodiverse ecosystems.

By using our framework, we are able to generate accurate and robust predictions for computer-generated forests and for 2 empirical tropical forests. We demonstrate the accuracy of our predictions using data from two well-studied forest stands. Moreover, we compared our results with other popular methods proposed in the literature to infer species richness from presence-absence data and we showed that our framework gives better estimates. Our framework is also able to give a quantitative estimate of the sampling effort needed for achieving species richness predictions with error bars below approximately 5% (this percentage was arbitrarily chosen as an illustration and our approach can be straightforwardly used for any other percentage of error). These estimates have been obtained through Monte Carlo simulations that test the self-consistency of the Negative Binomial method and allow us to infer these critical sampling thresholds.

Before illustrating the details, we want to highlight differences and similarities be-

tween the present work and the method derived in Section 2.1. Both approaches are based on the form-invariance property of the Negative Binomial distribution but, instead of using population estimates at local scales, here we require only the knowledge of species' occurrences at multiple local scales. In other words, the loss of information at one local scale (i.e. for each sample we know if a species is present, but not the number of its individuals) is balanced by the presence-absence information on multiple local scales. Such a generalization is useful when empirical datasets provide information only on the presence/absence of species. We will show that this will be enough to infer population's distribution as well.

The proposed method is, under the 'well mixed' hypothesis, general and not limited to tropical forests. We recall here a key feature of the theoretical framework we introduced in Section 2.2. If we assume that the RSA at the global scale is distributed according to a truncated Negative Binomial, then, the form-invariance property ensure that any local sample RSA is again proportional to a Negative Binomial with same parameter r and rescaled ξ. In details, if the RSA at a given scale p^* is distributed according to a (truncated) NB of parameters r and ξ_{p^*}, i.e.

$$\mathbb{P}(n|p^*) = c(r, \xi_{p^*}) \binom{n+r-1}{n} \xi_{p^*}^n (1-\xi_{p^*})^r \tag{3.1}$$

then the RSA at a different scale p is again proportional to a Negative Binomial, with same parameter r and rescaled parameter ξ_p, given by

$$\xi_p = \frac{\frac{p}{p^*}\xi_{p^*}}{1-\xi_{p^*}(1-\frac{p}{p^*})} =: U(p \mid p^*, \xi_{p^*}) \tag{3.2}$$

Remark 1. For forests, spatial independence reduce to assume that a certain species of trees is uniformly distributed among all the surface of the forests. This is usually not true, since plants likely arise in proximity of another plant of the same species. However, we need to keep in mind that we are not sampling too small areas, so that an high presence of a certain species is likely linked to a high abundance on the whole forest. Moreover, we do not analyze a single sample, rather we take several scatter samples of multiple small areas distributed all over the forest, leading to a reduction of spatial correlation in observed data.

From our general framework derived in Section 2.2, the following relation links the number of existing species at different scales.

$$S_p = S_{p^*} \frac{1-(1-U(p \mid p^*, \xi_{p^*}))^r}{1-(1-\xi_{p^*})^r}. \tag{3.3}$$

Note that by taking $p > p^*$ we can predict biodiversity at a larger scale than the observed one. However, the parameters of the distribution are still unknown. The basic idea is to use the previous equation to compare theoretical biodiversity at subscales $p \leq p^*$ with empirical values obtained from subsampling. This procedure allows us to better describe the information we gather at the sampled level. Hence, starting from the empirical presence/absence data from scattered local samples we will be able to:

1. infer species richness at larger scales, thus the SAC up to the whole forest scale;
2. obtain information on species abundances in order to construct the RSA at both local and global scales;
3. introduce and infer the Relative Species Occupancy (RSO), i.e. the distribution of the occurrences (number of occupied cells) across species, at both local and global scales. This biodiversity pattern is a prediction of our modelling framework. It can be measured empirically and may be of ecological relevance as it proxies the distribution of species ranges (the area where a particular species can be found) in the ecosystem.

The key feature of the method is the possibility, given only presence/absence data, to connect and infer different biodiversity patterns at the global scale. We tested our framework on in-silico generated forests and on the two well-studied tropical forests of Barro Colorado Island and Pasoh. We finally compared the global estimates with the abundance-based method proposed in Section 2.1.

Operatively speaking, we performed the following steps (see Figure 3.1):

- First, given a set of scattered samples, list the species in it.
  In formulae, sample $C = \{c_1, \dots, c_{M^*}\}$, $M^* \geq 2$, cells covering a fraction p^* of the whole forest in which S^* species are observed. To each cell c_i, associate a vector $\Omega(c_i) = \{\omega_1^i, \dots, \omega_{S^*}^i\}$, with $\omega_s^i \in \{0, 1\}$, $s \in \{1, \dots, S^*\}$, $i \in \{1, \dots, M^*\}$. The entry ω_s^i of vector $\Omega(c_i)$ gives information on the presence/absence of the species s in the cell c_i, i.e. $\omega_s^i = 1$ if species s is present in cell c_i, $\omega_s^i = 0$ otherwise.

- Compute the empirical species-area curve as follows. From now on, let us suppose that all the M^* cells are of equal size a. This assumption does not affect the general framework but it simplifies the computation of the SAC. Call A the area of the whole forest, so that $p^* = M^* a/A$. At each sub-sampling scale $p_k = ka/A$, with $k \in \{1, \dots, M^*\}$, compute the average number of observed species as

$$\hat{S}_{p_k} = \frac{1}{\binom{M^*}{k}} \sum_{\substack{I \subseteq \{1,\dots,M^*\} \\ |I|=k}} \sum_{s=1}^{S^*} \mathbb{1}\left(\sum_{i \in I} \omega_s^i \geq 1 \right), \tag{3.4}$$

  where $\mathbb{1}(X)$ is the indicator function, which equals one when the random event X happens and it is zero otherwise.
  By equation (3.4) we are computing, for every scale $p_k \leq p^*$, the empirical average of number of the species observed in all subsets of k cells.
  Since computing all subsets of k cells among M^* is numerically expensive for large M^*, in the analyses we computed the average among 100 randomly chosen subsets.
  We underline here that, computing the species accumulation curve through the empirical average of the number of species in k random selected cells, we are neglecting any spatial information. Let us stress once again that null or small spatial correlation is required for a rigorous derivation of our estimates.

- Fit the empirical species accumulation curve data (3.4) with the theoretical equation (3.3), evaluated in $p = p_k$, i.e.

$$\hat{S}_{p_k} \approx S_{p_k} = S_{p^*} \frac{1 - (1 - U(\tilde{p}_k|1, \xi_{p^*}))^r}{1 - (1 - \xi_{p^*})^r} \tag{3.5}$$

  and obtain the parameters $(\hat{r}, \hat{\xi}_{p^*})$ which best describe the empirical curve $\hat{S}_{p_k}$. These are the parameters of the NB relative species abundance distribution at the sample scale p^*. This protocol allows us to capture some spatial effects in the effective parameters.

- As showed in Section 2.1, under the hypotheses of absence of strong spatial correlations due to both inter-specific or intra-specific interactions, strong environmental gradients and abundances distributed according to a Negative Binomial at the whole forest scale, the RSA distributions at different scales have the same functional form of the RSA at the scale p^*, and only the values of the parameter ξ_p changes as a function of the scale p. Thus we obtain an analytical form of the upscaled RSA at any scale p given we know it at scale p^* in term of the equation (3.2), relating $\xi_p = \xi(p)$ to p, p^* and $\xi_{p^*} = \xi(p^*)$. Therefore we can predict the total number of species, S, at the whole forest scale, $p = 1$.

3.3 Results

3.3.1 Tests on *in-silico* databases

We test our presence/absence upscaling method on four computer generated forests without and with spatial correlations. Indeed, we expect that in the first case our framework will give more accurate estimates, and we wish to test how the introduction of correlations affect the reliability of our results.
As RSA we choose a Negative Binomial (NB forest) of parameters $r = 0.8$ and $\xi = 0.999$ and a Log-Normal (LN forest) with parameters $\mu = 5$ and $\sigma = 1$. Once generated the abundance of every species ($S = 4974$ for the NB forest and 5000 for the LN forest), we distribute the individuals within the forest area, here set equal to a square of 4900×4900 units, according to two different processes: at random or according to a modified Thomas process [Tov+16] with a clustering radius of 15 units.
We then divide each forest generated as described above into $M = 98 \times 98$ units cells and compute the $M \times S$ presence/absence matrix, thus forgetting the information about the species distribution. Finally, we sub-sample the 5% of the cells (corresponding to a fraction $p = 0.05$ of the total forest area) and apply our method to infer the total number of species in each of the four *in-silico* forests.

We also compared our results to those obtained by accounting for the data on species populations with an abundance-based upscaling framework developed and tested in Section 2.1. In the case of the NB forest, the two methods performed very well for both the random and the clumped distribution (i.e. individuals distributed on the space

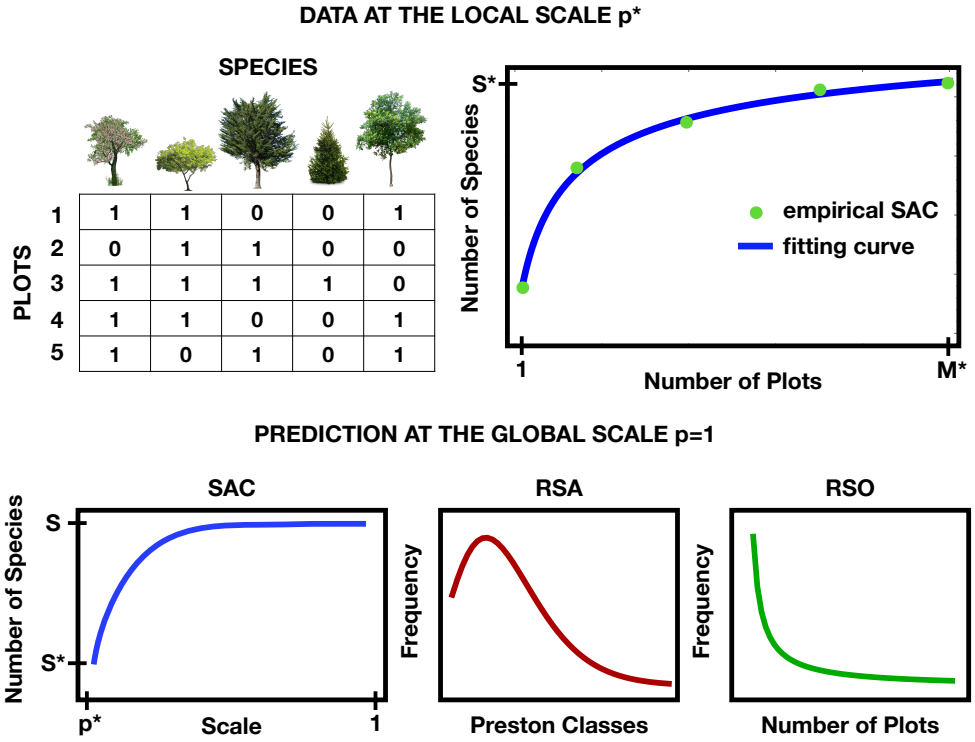

Figure 3.1: **Schematic presentation of our theoretical upscaling framework.** It consists of three steps. (A) We start from a dataset in the form of a binary matrix giving information on the presence or absence of S^* species within each of the M^* surveyed plots. (B) We perform the best fit of the empirically SAC computed via (3.4). (C) Using the best-fit parameters obtained in (B) and using our upscaling equations (2.11), (2.12) and (2.21), we predict the species richness S of the whole forest and three important macro-ecological patterns: the SAC, the RSA and the RSO.

according to a Thomas point process) with an average prediction error below 1% in absolute value (see Table3.2). In the Thomas distributed forests, the error increased, although remaining around 3% for the presence/absence method and around 7.5% for the abundance-based one (using maximum likelihood methods. The latter percentage error can be improved using calibrated statistical method for the single fit).

Thus, with respect to the degree of individuals' clustering, the new framework seems to give more robust estimates than the second one. This is due to two main reasons, Firstly, for the presence-absence case, we fit the empirical SAC, which has a very smooth functional shape, and it is easy to describe through our analytical SAC. On the other hand, the RSA displays a more complex and variable shape and thus fitting it with the NB is a more delicate task (indeed we find sensible differences on the accuracy using different statistical methods for the fit). Secondly, binary data on which the empirical

Forest RSA	Spatial Distribution			
	Random		Thomas	
	P/A	RSA	P/A	RSA
Log-normal	3.1 ± 0.51	7.6 ± 0.52	2.5 ± 1.8	7.2 ± 3.1
Negative binomial	-0.50 ± 0.34	-0.52 ± 0.28	-0.81 ± 1.6	-0.60 ± 1.7

Table 3.2: Predictive error for three generated forest (characterized by a log-normal and a negative binomial RSA) having individuals distributed according to a high clustering Thomas process and at random. Tests were performed by sampling a fraction $p = 0.05$ of each forest and by applying our framework (P/A columns) and the abundance-based method (RSA columns) to predict the true number of species S (5000 for the LN forest and 4974 for the NB forest). For each estimated S_{pred}, the average relative percentage error $(S_{\text{pred}} - S)/S \cdot 100$ between the true number of species and the predicted one is shown together with the corresponding standard deviation. Results are relative to 100 iterations.

SAC is based are less sensitive to sampling fluctuations.

3.3.2 Tests on natural forests

We finally test our method on sub-samples taken from two empirical forest data for which we have informations on both species occurrence and abundances. In particular we extract abundances of tree species observed in 50ha of rainforests from Pasoh (Malaysia) and Barro Colorado Island (Panama) together with the spatial locations of each of their individual. The Pasoh and Barro Colorado Island datasets are provided by the Center of Tropical Research Science of the Smithsonian Tropical Research Institute (https://stri.si.edu/).
Firstly, we divide both forest data into a grid consisting of $M = 800$ equal-sized cells of area 625 m^2 and we derive the $M \times S^*$ presence/absence matrix for the S^* observed species ($S^* = 927$ for Pasoh forest and 301 for BCI). We then sub-sample species occurrence for different fractions $0 < p < p^*$ of the cells and apply our framework to infer the number of species and other biodiversity patterns (RSA, RSO and SAC) at the corresponding largest empirically-observable scale p^*, for which we know the ground truth.

We compared our results on species richness obtained only from presence-absence data with the most popular non-parametric indicators proposed in the literature [Cha05; CC16], which are summarized in Table3.1. We found that our method outperforms all the others for both BCI and Pasoh forests, as shown in Figure 3.2. We also remark that all these methods have the further limitation that they can only infer the total species richness, without allowing for an estimate of the abundances' and occurrences'

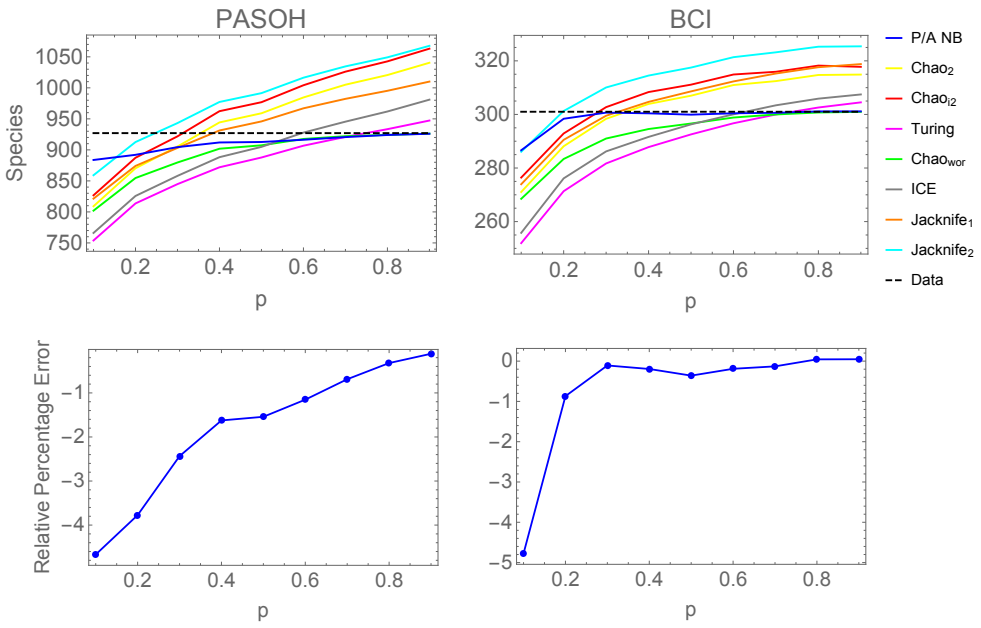

Figure 3.2: **Test from different scales for Pasoh and BCI.** For each forest we sub-sample a fraction p of p^* of the available spatial cells and apply different popular upscaling methods based on presence/absence data (see Table 3.1) and our method to predict the true number of species, S^* (dashed line) , observed in our data. While our (P/A NB) and Chao$_{wor}$ methods do converge at S_{p^*} as p goes to p^*, all the others have a monotonically increasing behaviour due to the independence, in their predictions, of the scale p^*. We can see that for both rainforests, our method outperforms all the others. Bottom panels show the relative percentage error $(S_{\text{pred}} - S^*)/S^* \cdot 100$ obtained with our framework between the predicted number of species S_{pred} and S^*. We find that the method underestimated the true number of species of at most 5%. The larger the sample area, the smaller the relative error.

distributions, i.e the shape of the RSA and the RSO. Indeed, as shown in Figure 3.4, from the local presence-absence data, we can reconstruct, among the SAC, the RSA at the whole tropical forest scale, thus relating species occurrence data with information on the abundances. In particular we can see that the inferred RSA are statistically comparable with the empirical ones obtained by using all the information on species' abundances which we deleted before applying our method.

Another biodiversity pattern that we can infer from our framework is the RSO. As shown in Figure 3.4, we find that, as for the RSA, this pattern seems to have a universal form which can be well described and correctly inferred through our neutral approach. Also, our finding suggests that, when spatial effects are negligible, the RSO distribution has a wide range of values in which it is well approximated by a universal power law, regardless of the details of the populations' dynamics. One may assume that this latter is

driven by a simple stochastic process with constant per capita birth and death rates. Such a slow decay of $Q(v|M,1)$ indicates that species in real systems exhibit huge variations in their occurrences, which may be weakly correlated to species' habitat preferences or environmental heterogeneities. We should expect strong asymmetries among their occurrences: for instance, if we tile up a landscape into $M = 1,000$ elementary cells, then about a third of all species should live in less than 1% of them; whereas about 1.5% of the species should be found in more than 90% of the total cells (see Figure 3.4).

We highlight that the SAC (green line), the RSA (red line) and RSO (blue line) predicted patterns in Figure 3.4 have not been obtained through the fit of some parameters, but they have been analytically predicted through our upscaling equations (2.11), (2.12) and (2.21). The only fitting occurs at the scale $p = 0.1p^*$ using the empirical SAC to parametrize eq. (3.4). In other words, by fitting species occurrence data at the sample scale, our framework allows to estimate: 1) The RSA at the sample scale; 2) The SAC, the RSA and the RSO at larger scales. We provide an open source R code that performs the above estimates giving as input only the presence-absence matrix data.
After testing our model on controlled computer generated data and real forest subsamples, we apply our framework to predict the species richness of the two tropical forests. Moreover, we compare our results to those obtained with the upscaling framework based on RSA pattern previously developed and tested in [Tov+17] using the method described in Section 2.1.

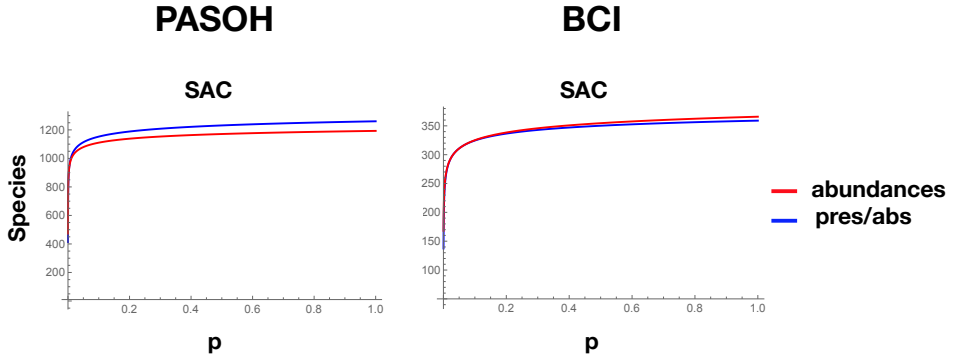

Figure 3.3: **SAC predicted for Pasoh and BCI using abundance method versus presence/absence method.** Using all the available data for both tropical forests, we compare the prediction for the SAC curve obtained by the abundance method with the results obtained with the presence/absence framework presented here. At the whole forests' scale $p = 1$, the two predictions are 3σ compatible ($S^{\text{abund}}_{\text{Pasoh}} = 1193 \pm 36$, $S^{\text{p/a}}_{\text{Pasoh}} = 1260 \pm 22$, $S^{\text{abund}}_{\text{BCI}} = 366 \pm 15$, $S^{\text{p/a}}_{\text{BCI}} = 359 \pm 2$).

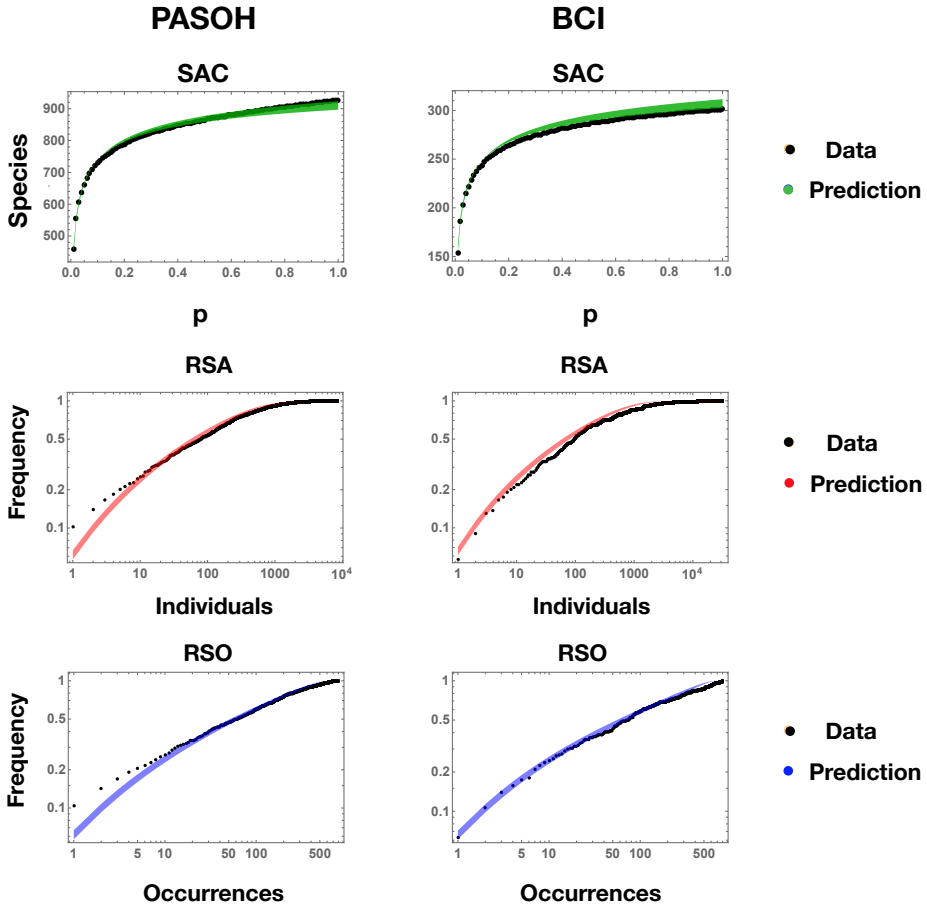

Figure 3.4: **Test on ecological macro-patterns for Pasoh and BCI.** For each forest we sub-sample a fraction $p = 0.1$ of the available spatial cells and apply our framework to predict three important ecological pattern at the largest scale at which we have information, p^*. In the first row we see the prediction for the SAC curve, which describes how the number of observed species increases with the sampled area, from $p = 0.1$ to 1, corresponding to p^* in these tests. In the second row we plot the cumulative empirical RSA, the distribution of abundances across species against the framework prediction in logarithmic scale. Finally, we test the ability of the model to capture the empirical RSO, i.e. the distribution of the occurrences (number of occupied cells) across species, with the third row panel showing, in logarithmic scale, the comparison between the predicted and observed cumulative distributions. For both forests, all the three patterns result to be well described by our framework.

We therefore predict, through the presence/absence method, the species richness at the whole forest scale ($p = 1$) for BCI and Pasoh tropical forests. Figure 3.3) shows the prediction of the overall (and unknown) SAC for a scale ranging from 50 to 14000 hectares for the Pasoh ($p^* \approx 0.0036$) and to 1560 for the BCI ($p^* \approx 0.032$). The blue

curves represent the prediction obtained only using presence-absence data whereas red curves are the SAC inferred by exploiting also the information about species' population through the abundance-based method.
We find that the two methods give comparable results for both the databases, a confirm of the robustness of the theoretical framework.

3.4 Conclusion

Our key goal was to expand the ability to upscale species richness and obtain abundance distributions from presence-absence data. We proposed and tested a parametric method on two forests' datasets, where we also disposed of data on abundances. Since the results are comparable (see Figure 3.3), we think that such a method could be exploited in many contexts, where abundance information are not available or trustable.
This is particularly true for microbial or marine (e.g. plankton) ecological data obtained from metagenomics [MNK16] and 16S ribosomal gene sequences [Soe+12]. The use of sequence-based taxonomic classification of environmental microbes has exploded in recent years and these approaches are becoming a standard method for characterizing the biodiversity of both prokaryotes and eukaryotes. Thanks to advance in high throughput sequencing we begin to be able quantifying the vast number of microbes in our environments, expanding our knowledge on microbial diversity. However, large fractions of the sequence reads remain unclassified and also species abundance estimated have a very high uncertainty. Thus, being able to estimated species richness and abundance distributions from species occurrence data may lead to a big step-forward in the taxonomic classification of microbial ecosystems.

Chapter 4

Upscaling human activity data: a case study with a statistical ecology approach

4.1 Datasets deriving from human activities

In ecology one of the most studied emerging patterns is the Relative Species Abundance (RSA), that gives the fraction of species with the same number of individuals. To determine large scale RSA features from the distribution of species abundances within a small random sample is a major challenge in ecology and we already mentioned that through years plenty of methods have been proposed [GT56; HSS09; CC14; Sli+15; OSW16]. The success of such methods depends on the following notable fact: different ecosystems like tropical forests or coral reefs [Vol+03; Vol+07; Sli+15; Tov+17], despite their disparate locations and different evolutionary history, share a common log-series shape of the empirical RSA which implies that the number of different species grows as the logarithm of the population size (see Figure 4.1). In the present work we adopted and extended a statistical framework which was firstly designed in ecology to get new insights into human activities databases with the aim of inferring global statistics of a dataset from a random sample of it.

Indeed, we consider four databases concerning human activities: emails, Twitter, Wikipedia and Gutenberg. Here we give a brief description of the data and we set the correspondence between species and individuals within each dataset. For further details, see [For+14] for email dataset and [Mon+17] for Twitter, Wikipedia and Gutenberg data.

1. *Email communication*: This dataset is a collection of almost 7 millions emails, that corresponds to the activity of a Department of the Università degli Studi di Padova during two years: 2012 and 2013. The collected data are in the form {sender, receiver, timestamp}. For our analysis, we selected the first column of the table. Here we define a sender identity to label a species and the number of sent

emails to be the number of individuals pertaining to a species;

2. *Twitter posts*: Our dataset consists of a table where each row is of the form {timestamp, hashtag, user}. For our purposes, we selected the second column of the table. Here hashtags play the role of species and the number of different tweets containing a certain hashtag represents its population size;

3. For *Wikipedia articles*: Our data represents all words contained in a collection of Wikipedia pages. We label each different word with a different number. Note that the same word always maintain its correspondence to the same number, regardless of the Wikipedia page it belongs. Each word is a different species while its abundance is given by the number of occurrences of the word in the dataset;

4. *Gutenberg books*: Our data represents all words contained in a collection of book from the Gutenberg archive. we used the same setting as for Wikipedia.

Once defined, as we did above, what correspond to species and individuals, the RSA of each dataset displays a Zipf tail [Mon+17] and the rate at which new elements appear shows a sublinear power-law growth, signature of the Heaps law (see also Figure 4.1). Statistical regularities in human dynamics have been widely observed in many different contexts and a variety of models have been proposed to understand such recurrent patterns [Bar05; Mal+09; Lor+11; LMT12; YSK12; Tör+13; Dev+16; YHM17; KJK18; Maz+18]. In particular, Zipf's laws have been observed since decades in computational linguistic and many models generating such laws have been proposed (see [Baa02] and [Kor07] for a review). However, in the present work we focus on inference, not modeling. We adopt the statistical framework described in Section 2.3 that gives reliable estimates for the number of users, hashtags, and words from a random sample of mails, posts and word occurrences (see Table 4.1).
Moreover, our framework predicts how the number of users/hashtags/words grows with the recorded activity (mails/posts/pages/books) (see Figure 4.2). Hence we infer how the abundance of a species may change across scales (see Table 4.2). This for example means that, observing a small portion of tweets and the popularity of a given hashtag among them, we can predict whether it will remain popular or not in the unseen part of the network.

In our statistical model we make the hypothesis the RSA distributions of the four human activity datasets can be described by Negative Binomials, each with a clustering coefficient r in the range $(-1, 0)$ (see Section 1.4). This choice, justified by the heavy tail of the observed RSAs (see Figure 4.3), has the major consequence that the RSA is *form-invariant*, as shown in Section 1.3. Form-invariance should not be confused with scale-invariance, a property only satisfied by power-laws. We recall that by form-invariance we mean that, when a portion of individuals are randomly sampled, the resulting RSA is still Negative-Binomially distributed with a heavy tail showing the same exponent as of the whole dataset (see Figure 4.3 and Section 1.3). Form-invariance property of the RSA allows us to build reliable estimators for the number of new features (new email users, new hashtags, new words) at each scale of the dataset starting from random samples of

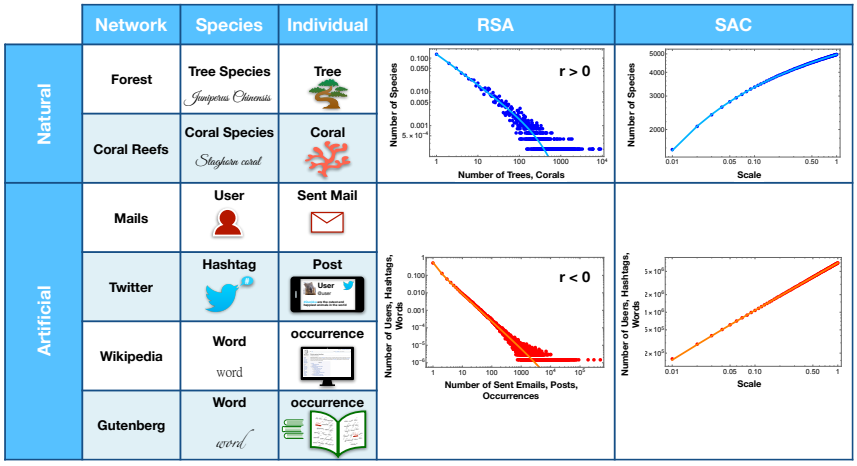

Figure 4.1: **From Ecology to Human Activities.** The figure depicts the correspondence between species/individuals in a natural ecosystem and users/sent emails, hashtags/posts, words/occurrences in each one of the four datasets considered in the paper. Once the proper correspondence is established, it turns out that both natural and artificial RSAs can be well described by a negative binomial distribution. As exemplified in the fourth column, all human activity RSA curves can be accommodated by with a negative value of r in the interval $(-1, 0)$, whereas natural ecosystems prefer $r > 0$. In the last column, typical shapes of the SAC curves in log-log scale are shown for both kind of systems.

the whole databases. Our framework encompasses both the natural and artificial worlds in terms of frequencies of abundances: the relative species abundances in the natural world (ecology) and the popularity of human activities (social sciences) such as email, hashtags and words.

Our approach brings two main novelties/advantages. First, the choice to model the frequency of frequencies distribution according to a Negative Binomial distribution. In particular, the idea of exploiting its form-invariance property to obtain an effective yet simple estimator which explicitly depends on the scale is new in the field of distributions with Power Law behavior. Actually, to our knowledge, upscaling has never been investigated for email communication and Twitter datasets whereas in linguistic different parametric and non parametric statistical models has been used to infer how the number of types grows as new samples are added [Baa02].
Second, within our framework we also derive an estimator for how the type abundances change across scales (see Section 2.3). This problem, as far as we know, has not been previously investigated although it could be of interest when interpreting abundances of types as a measure of popularity in social network data.

4.1.1 Application of our upscaling method

We start by illustrating our approach, its potentiality and the kind of results it can provide as applied to e-mail communication. We consider the senders activity network where each node is a user and a directed link from node A to node B represents an email issued from user A to user B. We set the identity of a sender to label the species and the number of sent emails to be the individuals pertaining to a species. Thus, for instance, if user A has sent n emails we say that species A has n individuals. Suppose an observer have access to a small sample of sent emails, or, equivalently, to partial information on links and nodes of the email communication network. Our approach allows to infer the number of nodes (i.e. the number of users) and the link statistics of the whole network, thus revealing features previously unknown to the observer (see Figure 4.2).
Correspondence between species/individuals and human activities can be set similarly for the remaining datasets (see Figure 4.1).

Our statistical ecology approach gives the following results:

- **RSA universality and form-invariance.** In each activity the RSA of the whole dataset turns out to be heavy tailed with an exponent between -1.8 and -1.4 (see Figure 4.3). Moreover, this exponent is maintained at different scales (see also Figures 4.4), supporting our choice of modeling the RSA by means of a form-invariant distribution that keeps fixed the tail exponent through scales.

- **Inference of unseen human activities.** On the scale invariance property of the RSA we build a statistical framework which gives robust and accurate estimates for the number of email senders, Twitter hashtags, words of Wikipedia pages and Gutenberg books from a random sample of sent mails, posts and word occurrences (see Table 4.1). Moreover, our framework allows to reconstruct the growth of the number of users/hashtags/words with the recorded activity (mails/posts/pages/books), which represents another well known pattern in ecological theory called the *Species-Accumulation Curve* (SAC) (see Figure 4.2).

- **Popularity in social networks.** In Twitter and in social networks in general, popularity is known to be relevant, for instance, to manipulate mass opinion or to share information. One naive way to measure the popularity of a hashtag is to count the number of times it appears in other users' tweets. In our ecological interpretation, a hashtag represents a species, while the number of posts associated to it, gives the species' abundance. Within our framework, we can infer how the abundance of a species changes across scales (see Table 4.2), thus allowing to monitor whether a locally popular hashtag will remain popular also in the undetected part of the network or not.

4.2 Upscaling the number of different types

In the following we recall the key steps of our upscaling framework. Denote with N the population size and with S the number of species (i.e. senders, hashtags, words) of the

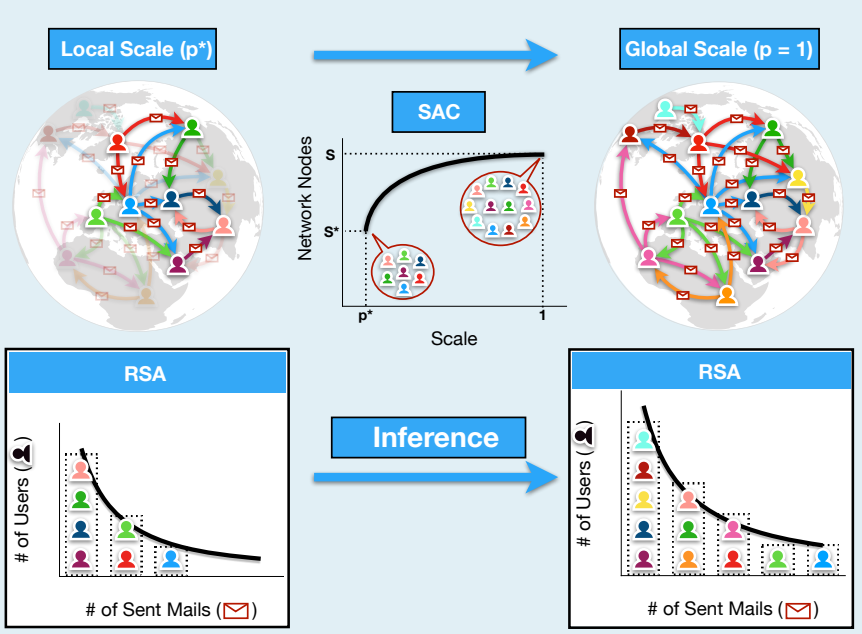

Figure 4.2: **Sketch of our theoretical framework.** Consider the email senders' network where each node is a sender and a directed link from node A to node B is an email issued from user A to user B. We set the identity of a sender to be the species and a sent email to be an individual of that species. For instance, if the user A has sent n emails, then the species A has n individuals. If an observer has access to a fraction p^* of the sent emails, s/he can partially recover the network (top-left) and the RSA curve at the local scale p (bottom-left). Within our framework, this information suffices to infer the number of species and the RSA curve at the global scale $p = 1$ (bottom-right). In terms of the network, the number of species corresponds to the number of users or nodes and the RSA gives the degree statistics. In this sense, our method reveals network features pertaining to the whole community activity initially unknown to the observer (top-right). Moreover, we can predict how the number of users increases with the number of links recorded, (i.e. the SAC curve in ecology), an information that may be used to optimize network design forecasting its growth.

whole database. Given a scale $p^* \in (0, 1)$, consider a random sample of size p^*N in which we recover $S_{p^*} \leq S$ species. Denote by $S_{p^*}(n)/S_{p^*}$ the fraction of species with n individuals at scale p^*, i.e. the sampled RSA. We assume that, at the global scale $p = 1$, the theoretical RSA $\mathbb{P}(n|1)$ is proportional to a truncated Negative Binomial distribution, $\mathcal{P}_{NB}(n|r, \xi)$, with parameters $r \in (-1, 0)$ and $\xi \in (0, 1)$:

$$\mathbb{P}(n|1) = c(r, \xi)\mathcal{P}_{NB}(n|r, \xi) \qquad \text{for } n \geq 1 \tag{4.1}$$

where the normalizing factor $c(r, \xi) = 1/(1 - (1 - \xi)^r)$ takes into account that each of the S species consists of at least one individual at the global scale.

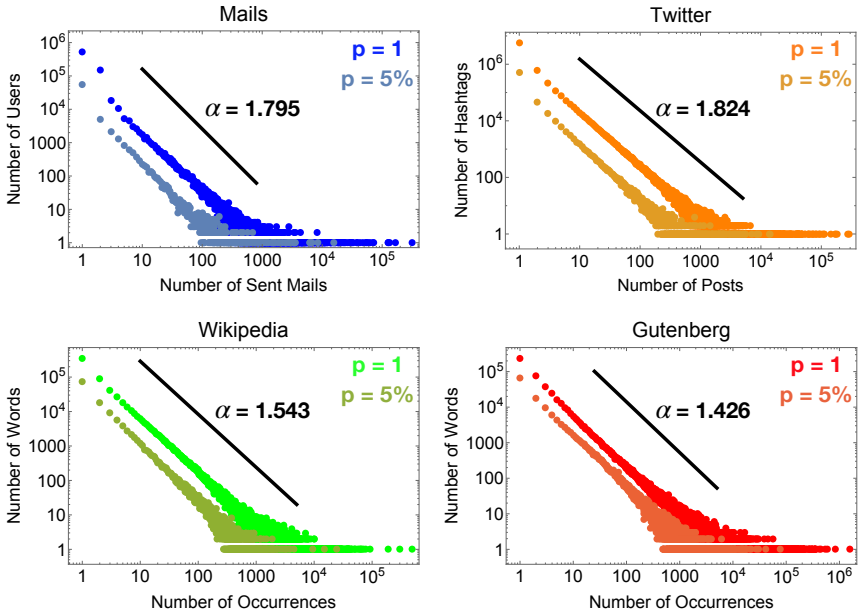

Figure 4.3: **Universality and form-invariance of the empirical RSAs.** Empirical RSA curves at the global scale ($p = 1$) and the local scale ($p = 5\%$) are shown. The patterns result scale-free in all the analyzed datasets, with a heavy-tailed form maintained through the different human activities and scales. This scale-invariance property of the RSAs allows for a successful implementation of our theoretical framework. In particular, our model predicts that the heavy-tail exponent α is related to the clustering parameter r of the RSA negative binomial distribution via the relation $\alpha = 1 - r$ (see Materials and Methods and Supplementary Section S1.3). In each plot, for a visual inspection, we inserted a black line with slope $-\alpha = -1 + \hat{r}$, where $\hat{r}$ have been obtained by fitting the local patterns at $p = 5\%$ through a negative binomial (see also Table 4.1). We can see that such lines describe very well the heavy-tail regime of the empirical RSAs at both local and global scale in all four cases. For the RSA fitting curves and predicted patterns, see Figure 4.4.

RSAs given in (4.1) have the following features:

1. values of $r \in (-1, 0)$ reflect in a heavy-tailed behavior of the RSAs. More precisely, the right tail of (4.1) has the form $n^{r-1} \exp(n \log \xi)$ (see Section 1.4), where the exponential cut-off disappears in the limit $\xi \to 1$. In this latter case (4.1) describes a pure power-law tail behavior. Such heavy-tailed behavior well describes the observed RSA patterns in human activities (see 4.4). Moreover, the exponent $\alpha = 1 - r$ matches very well with the empirical data.

2. Distribution (4.1) is *form-invariant*, meaning that the RSA $\mathbb{P}(n|p)$ maintains the same functional form at different scales p (see Section 1.3), a property observed

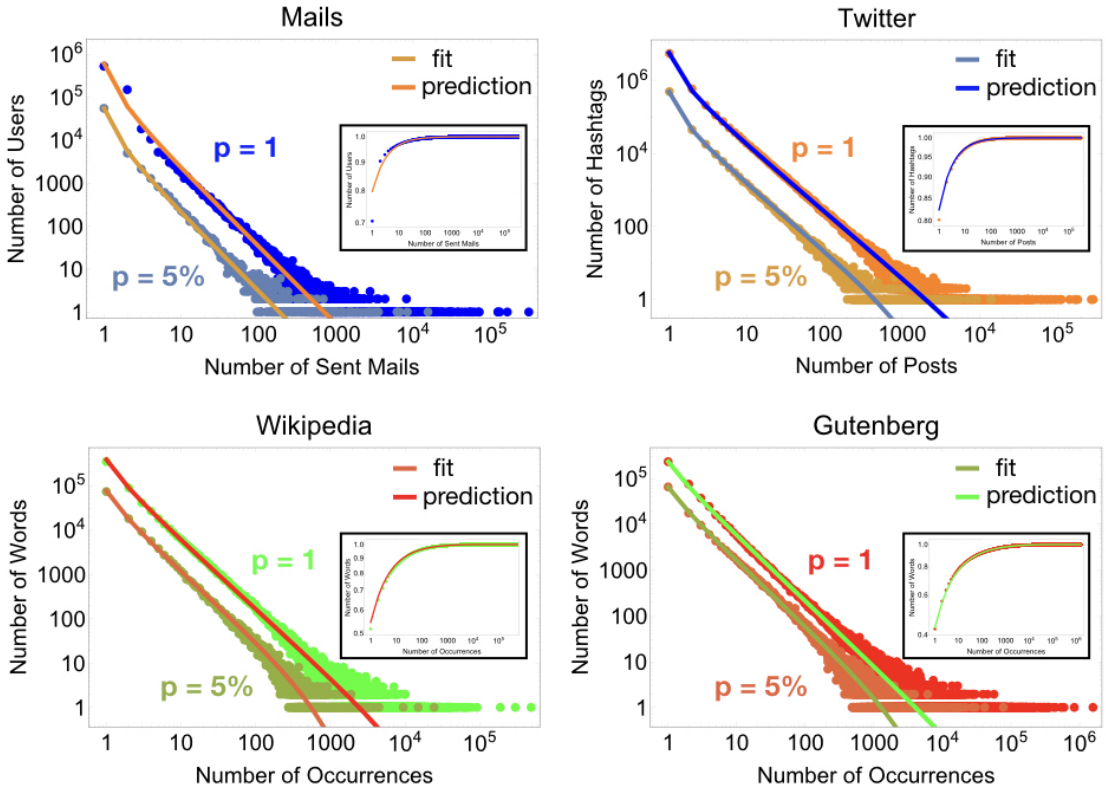

Figure 4.4: **Best-fit and predicted patterns from a local sample scale $\mathbf{p^* = 5\%}$.** Empirical RSA curves at global scale ($p = 1$) and local scale ($p^* = 5\%$) are shown. In each panel, coloured lines over the local RSAs represent the distribution obtained via a best-fit of the empirical pattern with a negative binomial having $r \in (-1, 0)$. Lines over the global RSA distributions represent our prediction for the RSAs at the global scales obtained via our upscaling equations for both the parameters and the biodiversity. In each panel, insets showing the corresponding global cumulative RSA (both empirical and predicted) are added.

in the empirical RSAs of all the four databases (see Figure 4.3). In mathematical terms, the RSA at any scale p is again proportional to a Negative Binomial distribution with the same r and a rescaled parameter

$$\xi_p = p\xi/(1 - \xi(1 - p)). \tag{4.2}$$

Properties 1) and 2) are the building blocks of our predictive statistical framework.
Our goal is to infer the total amount of species S (senders, hashtags, words) present in the entire database given the number of species S_{p^*} observed in a sample at the local scale p^* and their corresponding abundance (number of mails, posts, occurrences). From this limited information, we can construct the empirical values of the RSA at scale p^*,

and fit it to obtain the estimates $\hat{r}$ and $\hat{\xi}_{p^*}$ of the parameters that best capture the behavior of the sampled data. Finally, thanks to the form-invariance property, one can obtain the value of the global parameter $\hat{\xi}$ via eq. (4.2) (henceforth we will denote with $\hat{\cdot}$ our estimation of any quantity $\cdot$).

Our upscaling method, derived in Section 2.1 leads to the following formula to estimate the total number of different species present at the global scale

$$\hat{S} \simeq S_{p^*} \frac{1 - \left(1 - \frac{\hat{\xi}_{p^*}}{p^* + \hat{\xi}_{p^*}(1-p^*)}\right)^{\hat{r}}}{1 - (1 - \hat{\xi}_{p^*})^{\hat{r}}} \tag{4.3}$$

To test the reliability of estimator (4.3), we extracted, from each dataset, ten subsamples each covering a fraction $p^* = 5\%$ of the databases' individuals (sent emails, posted hashtags, occurrences of words). We then inferred the total number of species (email senders, posted hashtags in Twitter data and different words in Wikipedia pages and Gutenberg books) from the empirical RSA constructed at $p^* = 5\%$. The average relative upscaling error is small in all four cases: about 0.1% for sent emails, 3% for Twitter hashtags, 6% for Wikipedia words and -2% for Gutenberg words. In Table 4.1 we report the average values of the fitted parameters together with the average relative percentage error between the predicted number of species, $\hat{S}$, and the true one, S (mean and standard deviation are displayed for all datasets). See also the Supplementary Material of our article [Tov+19b] for the results obtained by considering a different fraction $p^* = 3\%$ of the four datasets as starting information.

	Emails	Twitter	Wikipedia	Gutenberg
Species	$752,299$	$6,972,453$	$673,872$	$554,193$
Individuals	$6,914,872$	$34,696,973$	$29,606,116$	$126,289,661$
$\mathbf{r}$	-0.795	-0.824	-0.543	-0.426
$\xi_{\mathbf{P}^*}$	0.9999	0.9991	0.9985	0.9997
Relative Error	$0.112 \pm 0.385\%$	$3.33 \pm 0.17\%$	$6.11 \pm 0.118\%$	$-2.30 \pm 0.23\%$

Table 4.1: **Predicted relative errors.** Upscaling results for the number of species of the four analyzed datasets from local samples covering a fraction $p^* = 5\%$ of the corresponding global dataset. For each human activity, we display the number of species (users, hashtags, words) and individuals (sent mails, posts, occurrences) at the global scale together with the fitted RSA distribution parameters at the sampled scale and the relative percentage error (mean and standard deviation) between the true number of species and the one predicted by our framework. See Figure 4.4 for the corresponding fitting curves and predicted global RSA patterns.

4.2.1 Local Analysis

Before using our method to infer information at the global scale, we conduct some investigations on its properties, as some kind of empirical justification in order to proceed and use it.
As first we consider our sample and observe the plot of the its Relative Species Abundance. Despite the different origin of data, all the four datasets' samples display a similar behavior. The Power Law decay of the RSAs in the different cases suggest us to use our upscaling method with Negative clustering parameter $r \in (-1, 0)$. Before performing it, we explore a bit our samples. For each of the four human activity samples, we consider further subsamples at subscales $q_j = \frac{j}{10}$, $j = 1, ..., 9$. Hence at the subscale q_j (also generically called p^{**} to rely on the usual notation) we are considering a fraction q_j of the original sample at scale $p^* = 5\%$. Once we get this subsamples we apply our method to estimate, at each subscale q_j, the number of species at the "global" scale p^*.
Clearly this work leads to an estimate of a quantity that we already knew from observation. So why is that important? Our method applies if our random sample is indeed random. It can happen that the way we sample is not actually random, due to hidden correlations which we are not aware of. Hence, testing our method on our sample inform us that there is no evidence of errors in the sampling procedure. Thus we rely more on the fact that the final estimate succeeds to be close to the real value.
In Figure 4.5 we show the error we commit when testing the method at subscales q_j to predict the number of species present at the "global" scale p. Note that q_j is not an absolute value, but a fraction of p.
In the bottom panels, we displayed the relative percentage error graphs between the true number of species, S^*, and the one predicted from the local information at the different sub-scales p^{**}. We see that, for all datasets and sub-scales, our method always led to an error below 5%. Moreover, it displays an intuitive decreasing behavior as the available information increases, a desirable property for an estimator. We performed the same analysis also starting from a sample at the scale $p^* = 3\%$, obtaining comparable results.

4.3 Upscaling hashtags' popularity in Twitter database

The second novelty that we introduce in our work is a method to estimate the variation of popularity, a fundamental concept arising naturally when investigating human dynamics [She+14; Zha+15; Sin+16]. Indeed, until now we exploited the information on the abundance of the observed species at the local scale only to estimate the number of unseen species, disregarding of their abundances. Instead, abundance information can be used to predict, for example, the most active users of the email network, the commonest words in a book or, having in mind more topical applications, the popularity of a hashtag in Twitter database. In particular, focusing on Twitter, various sophisticated measures of popularity based on semantic analyses have been proposed. Here, by mean of the popularity of a hashtag we naively count the number of posts containing it that

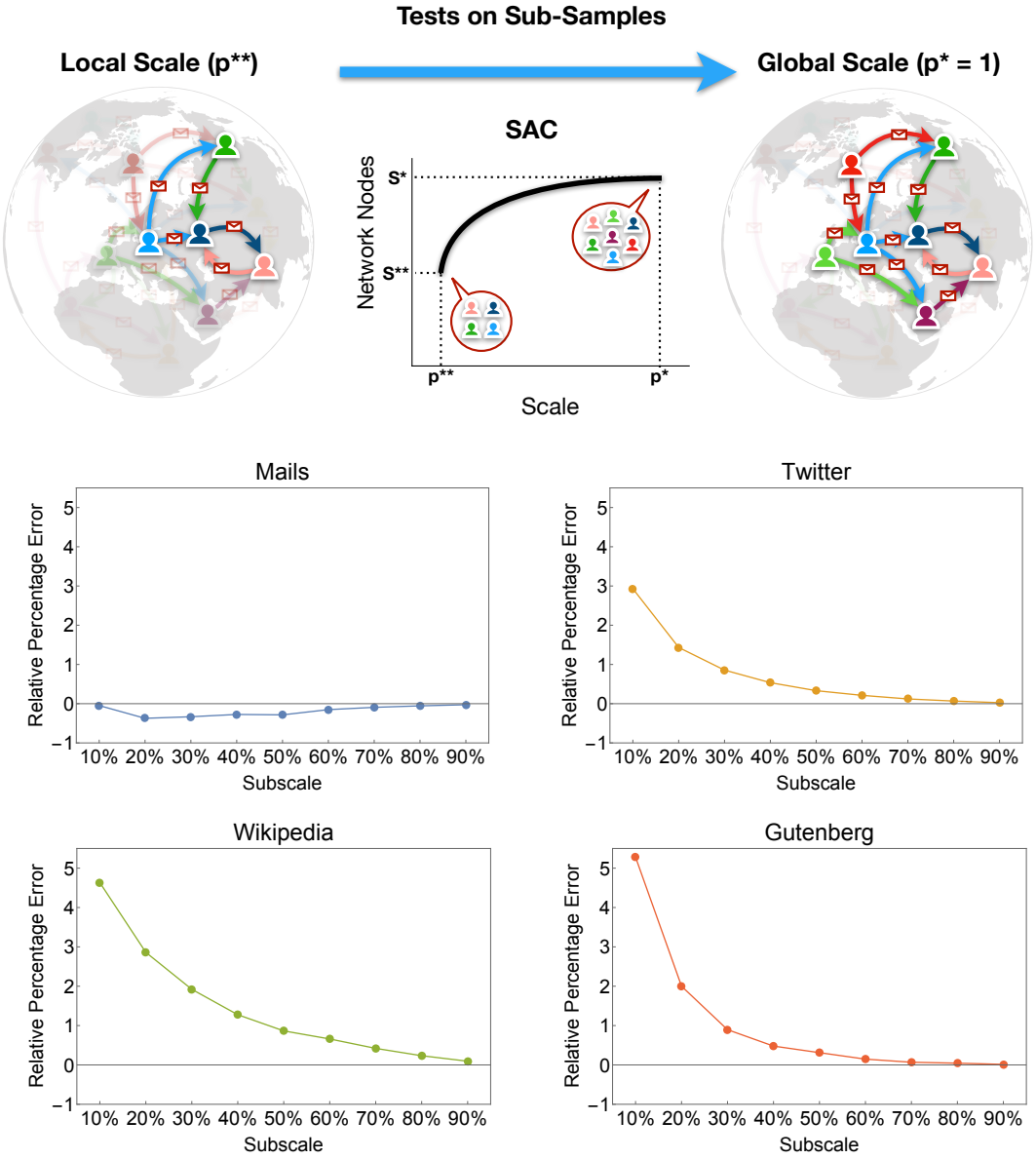

Figure 4.5: **Relative percentage errors at different sub-scales from $\mathbf{p^* = 5\%}$.** Starting from a sample at $p^* = 5\%$ of each human activity database, we sub-sampled it at different spatial sub-scales $p^{**} \in \{10\%, \ldots, 90\%\}$ of p^* and computed the relative percentage error between the number of predicted species, $\hat{S}^*$, and the true number of species, S^*, observed in the sample at p^*, here considered as the global scale ($p^* = 1$).

come to circulate within the network thanks to other users' tweets. This information is encompassed within the RSA pattern. Indeed, hashtags posted a low number of times are those positioned in the left side of the curve, whereas hashtags with high popularity

are located in its right tail. Our goal now is to derive an estimator for the change in popularity of hashtags from a portion p^* of the observed tweets to the remaining $1-p^*$ fraction of the unobserved tweets.
Let us thus denote by L a fixed threshold of posts above which we consider a hashtag popular at the sampled scale p^* and let us indicate with $S_{p^*}(L^+)$ the number of different hashtags having abundance at least L in the surveyed collection of posts. We wish to check whether these (locally) popular species result to be popular also in the unseen fraction of the network, $1-p^*$. Let us then denote by K the fixed popularity threshold at the unsurveyed scale. We are looking for an estimator of the number of species having popularity at least K in the $1-p^*$ unseen part of the tweets, given that they have popularity at least L at scale p^*. From our theoretical framework (equation (2.49) in Section 2.3), we derive an estimator $\hat{S}^{(p^{*c})}(K^+|L^+)$ of such a quantity. These species are therefore globally popular within the network.

We first derived the estimator $\hat{S}^{(p^{*c})}(k|\ell)$ for the number of species having popularity exactly k at scale $1-p^*$ given that they have popularity exactly l at scale p^*. (see Section 2.3 for details).

$$\hat{S}^{(p^{*c})}(k|\ell) = S_{p^*}(l) \cdot \frac{\binom{k+l}{l} p^{*l}(1-p^*)^k \binom{k+l+\hat{r}-1}{k+l} \hat{\xi}^{k+l}(1-\hat{\xi})^{\hat{r}}}{\binom{l+\hat{r}-1}{l} \hat{\xi}_{p^*}^{l}(1-\hat{\xi}_{p^*})^{\hat{r}}} \tag{4.4}$$

The estimator for $\hat{S}^{(p^{*c})}(K^+|L^+)$ can thus be obtained by summing up (4.4) for all $k \geq K$ and for all $\ell \geq L$ as did in (2.49) thus obtaining

$$\hat{S}^{(p^{*c})}(K^+ \mid L^+) = \sum_{\ell \geq L} \sum_{k \geq K} \hat{S}^{(p^{*c})}(k \mid \ell) \tag{4.5}$$

We tested the above estimator by fixing the (arbitrary) value of the threshold L equal to $10, 25, 40, 55$ and varying the value of K in a arbitrary range for ten sub-samples of Twitter database. The values of K are selected so that $\bar{k} = argmax_{k \in \mathbb{N}} \hat{S}^{(p^{*c})}(k|L)$ and then $k_i = \frac{i+1}{4}\bar{k}$ for $i = 1, ..., 5$. The average errors obtained in the predictions are displayed in Table 4.2. For all the considered cases, we achieved very good estimates, with an average relative percentage error below 0.2% in absolute value.

4.4 Discussions

To conclude, we showed how our statistical ecology framework could be successfully applied to human activities. We tested our method in four databases: email senders activity, Twitter hashtags, words in Wikipedia pages and Gutenberg books. Once set the correspondence to what we consider species and individuals of a species, our approach reveals that the RSA is scale-free in each mentioned dataset with a heavy-tailed form maintained at different scales - with roughly the same exponent - through the different human activities considered (see Figure 4.3). This form-invariant property allows for a successful implementation of our predictive statistical framework. However, the heavy

L	K	$S^{(p^{*c})}(K^+\|L^+)$	$\hat{S}^{(p^{*c})}(K^+\|L^+)$	Avg % Error	Variance
10	77	14266	14274	-0.0029	0.0012
10	115	14113	14106	0.0534	0.0151
10	154	13551	13545	0.2457	0.0428
10	192	12509	12584	0.4679	0.0731
10	231	11305	11367	0.5372	0.0965
25	219	5977	5976	0.0018	≈ 0
25	329	5943	5950	0.0448	0.0110
25	439	5667	5689	0.0896	0.0610
25	548	5064	5056	-0.1793	0.0878
25	658	4295	4321	0.2010	0.1853
40	362	3749	3749	-0.0001	≈ 0
40	543	3742	3742	0.0393	0.0058
40	724	3591	3579	-0.0715	0.0668
40	905	3096	3091	0.0368	0.0660
40	1086	2600	2583	-0.5634	0.0370
55	504	2673	2673	≈ 0	≈ 0
55	756	2672	2671	-0.0141	0.0013
55	1008	2569	2568	-0.0978	0.0565
55	1260	2195	2199	0.0023	0.0557
55	1512	1806	1820	0.1286	0.2070

Table 4.2: **Percentage errors for popularity change predictions in Twitter database.** For $L = 10,\ 25,\ 40,\ 55$ (first column) and different values of K(second column), we estimated, from ten different Twitter sub-samples ($p^* = 5\%$), the number of species having abundance at least K at the unobserved scale $1 - p^* = 95\%$ given that they have abundance at least L at the sampled scale p^* via estimator (4.5). The average, among the ten sub-samples, of the true number of species, $S^{(p^{*c})}(K^+|L^+)$, and predicted one by our method, $\hat{S}^{(p^{*c})}(K^+|L^+)$, are displayed in the third and fourth columns, respectively. Finally, in the last two columns, we inserted the mean and the variance of the relative error obtained in the ten predictions.

tail of the observed RSAs cannot be captured by a standard negative binomial distribution with $r \in \mathbb{R}^+$. Nevertheless, such behaviors can be accommodated when allowing the clustering parameter r to take negative values, $r \in (-1, 0)$ (see Section 1.4 and Figure 4.4). This allows us to exploit the form-invariance property of the negative binomial distribution to propose an estimator for the statistics of the unseen human activity from small random samples. In particular, from the activity (sent emails per senders, posts per hashtags, word occurrences) in a small random samples, we infer the number of species (senders, hashtags, words) at the global scale.

Moreover, we predict how the popularity of species changes with the scale, an issue of evident importance when thinking of social networks like Twitter. Finally, we compare

our estimates with the true known values and in all the considered databases the relative error is small (see Table 4.1 and Table 4.2). This result confirms the ability of our theoretical method to capture hidden quantities of the human dynamics when only random samples are available. Our results pave the way for new applications in upscaling problems beyond statistical ecology.

Indeed, our findings may have applications in different situations, spreading from resource management in emails to collective attention monitoring in Twitter and to language learning process in word databases. Let us see one example for each aforementioned context of how our framework could help in decision making processes related to different aspects of social activity networks.
Let us start from the resource managing application. Suppose an internet/email provider starts a campaign to increase customers; for instance the provider wishes to double the number of subscribers. Now, in order to predict if more resources (e.g. number of servers in the email example) are necessary to supply the newly entered subscribers, the provider needs to infer the total amount of activity bursting thanks to these new users. Our method provides a possible solution to this inference problem. Indeed, by inverting equation (4.3), which represents the well-known species-accumulation curve in theoretical ecology, one obtains an analytical link between the total amount of activity (e.g. number of sent emails) and the number of users. In particular, the activity does not grow linearly with the users, as one may naively guess.
Thus, the information our framework provides on the species-accumulation curve may help the provider to decide how many further resources are needed for the expected number of new users. Clearly, this knowledge is useful either to avoid money waste in case no further resources are required, or to provide new structures/servers in advance in order to safely support the user activity and not to loose unsatisfied customers. Moreover, being aware of how many new structures are needed also helps balance their costs of installation, managing and maintenance with the profit coming from subscriptions.

A second application regards attention monitoring and information spreading. Nowadays social networks constitutes a fundamental source for spreading information and disinformation as well. They have being exploited to influence the mass opinion and attention in many different social contexts, from politics to economy. It is enough to think about the influencer phenomenon arising in almost all social networks. In Twitter, popularity of a user may be read from the number of times a hashtag s/he initiated appears in other users' tweets.
In our ecological interpretation, a hashtag represents a species, while the number of posts associated to it gives the species' abundance. Therefore, if the species s/he represents comes to be part of the right tail of the RSA distribution, it constitutes one of the community dominant species and thus we can say s/he is popular, whereas if it comes to fall at the left tail of the RSA, it is a hyper-rare species, thus not having received the desired attention. Therefore, in order to control someone's position within the global network, it is necessary to have access to the RSA at the whole community scale.
However, this datum is usually not provided by the social network managing organization. Twitter, for example, only releases information on the total number of tweets

posted across time. Nevertheless, there are other services as the Sample Tweets APIs or the Decahose stream service which provide the clients with real-time random samples covering small percentages (up to 10%) of the total tweets. With this information, our framework offers the possibility to fully reconstruct the global RSA as well as to monitor how the number of popular hashtags scales from the monitored sample up to the whole activity network.
This latter information may also be useful for governments or public administrations in general to communicate important news (health information, emergency procedures, elections etc...) to the citizens. In particular, our method allows to know the number of further tweets one eventually needs to effectively spread the information, thus allowing to undertake the proper measures (a bigger publicity campaign to obtain more followers, the development of bot applications, etc.) to achieve the goal.

Finally, our theoretical framework may also be exploited in language learning process monitoring. For example, let us suppose one is learning a foreign speech. S/he may then be interested in the number of books that are needed s/he needs to read in order to be sure to expand her/his own vocabulary in order for it to cover a fixed percentage of all the speech words. The species-accumulation curve emerging in this context thanks to our ecological correspondence between words/species and occurrences/abundances can thus be interpreted in a broader sense as a learning curve, with the total number of words encountered during the learning process (by dialogue experience, frontal lectures or personal readings) in the x-axis and the number of different words s/he manages to properly exploit in her/his speech in the y-axis.

Part II

Random walks in random environment: limit theorems for heavy-tailed processes

Chapter 5

Ideas of weak convergence

The notion of weak convergence is the building block of a large class of limit theorems in probability theory. We reserve this chapter for a proper introduction to the notion of weak convergence. The motivation for such a general treatment is dual. The first reason s that weak convergence is the building block of a large class of limit theorems in probability theory, and as such it can't be neglected or relegated to a minor section. Secondly, every novel result of this thesis is based on weak convergence, and since we will make extensively use of it, we agreed that this chapter is essential.

Let's start with some motivation and consider a context in which weak convergence arises naturally. Take a sequence $(F_n)_{n\in\mathbb{N}}$ of distribution functions on the real line. Recall that a distribution function is a non decreasing, right continuous function $F : \mathbb{R} \mapsto [0, 1]$ that satisfies $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$. We say that the sequence $(F_n)_{n\in\mathbb{N}}$ converge weakly to some limit distribution function F, and we denote it by $F_n \xrightarrow{w} F$, if the following limit holds for every continuity point of F.

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{5.1}$$

Consider now the probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$, defined on the class od Borel subsets of the real line, uniquely determined by the following requirements

$$\mathbb{P}_n(-\infty, x] = F_n(x) \qquad\qquad \mathbb{P}(-\infty, x] = F(x) \tag{5.2}$$

Note that the condition that x is a continuity point for F translate to the set $\{x\}$ having $\mathbb{P}$-measure 0. Hence we can think to extend the notion of weak convergence to probability measures by expressing condition (5.1) in the language of probability measures. We are tempted to say that the sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ converges weakly to $\mathbb{P}$ if the following limit holds for every point $\{x\}$ of $\mathbb{P}$-measure 0.

$$\lim_{n\to\infty} \mathbb{P}_n(-\infty, x] = \mathbb{P}(-\infty, x] \tag{5.3}$$

It turns out that actually convergence on particular sets of the form $(-\infty, x]$ is enough to guarantee weak convergence of the probability measures on all $\mathcal{B}(\mathbb{R})$. We skip this critical aspect for the moment, taking (5.3) as a definition.

Let's proceed now to illustrate how to extend the latter relation to a generic probability space S. Consider the probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$ defined on $\mathcal{B}(S)$, the class of the Borel subsets of S. Inspired by (5.3), we are lead to say that the sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ converge weakly to $\mathbb{P}$ if the following limit holds for every Borel set $\mathcal{A} \subset \mathcal{B}(S)$ such that $\mathbb{P}(\partial\mathcal{A}) = 0$.

$$\lim_{n\to\infty} \mathbb{P}_n(\mathcal{A}) = \mathbb{P}(\mathcal{A}) \tag{5.4}$$

The previous equation established the meaning of weak convergence for probability measures in a very general sense. This may appear very linear, but, when facing real problems, an assertion like (5.4) is very hard to prove directly. Hence it emerges a gap between the concept of weak convergence and its application. Actually, most of the times is not even possible to show that (5.4) holds. This is very disappointing and that's why we need a deep look into the concept of weak convergence.

5.1 Definitions and basic properties

As first we want to set the notation. Let S indicate a generic metric space and $\mathcal{S} = \mathcal{B}(S)$ the Borel σ-algebra on S, i.e. the σ-algebra generated by the open subsets of S. Note that we are assuming that S is a topological space endowed with a topology $\mathcal{T}$. The topology $\mathcal{T}$ is a collection of subsets of S, called open sets, satisfying certain axioms. In particular, a topology is any collection of subsets, including the whole set S and the empty set $\emptyset$, that is closed under arbitrary unions and finite intersections. Every metric d determines a topology generated by the open balls $\mathcal{B}_r^{(d)}(x) = \{y \in S : d(x, y) < r\}$, for $x \in S$ and $r > 0$, but not every topology can be induced by a metric. We will primarily be concerned with metrizable topologies, the ones that can be induced by a metric. We usually drop the index (d) on open balls when it is clear which metric we are dealing with. For metric spaces, the notation (S, d) often replaces $(S, \mathcal{S})$, meaning that $\mathcal{S}$ is the σ-algebra on S generated by the topology induced by the metric d.
In this introductory section we don't need to specify which topology we are using, but this will be crucial in the following. Unless specified, every measure treated in the following will be defined on the general space $(S, \mathcal{S})$. We are interested in probability measures, i.e. in the subclass of measures on $\mathcal{S}$ that are non-negative, countably additive and such that they give measure 1 to the whole space S. The elements of $\mathcal{S} = \mathcal{B}(S)$ are called measurable sets. It's essential to mention that measurability is guaranteed on every element of $\mathcal{S}$, whereas in general is not true that every subset of S id measurable. The notion of tightness is the relation we need to link compactness and probability measures.

Definition 4. A probability measure $\mathbb{P}$ on $(S, \mathcal{S})$ is tight if, for every $\varepsilon > 0$ the exists a compact set K such that $\mathbb{P}(K) > 1 - \varepsilon$.

The notion of tightness is one of the cornerstones of weak convergence, but the previous definition is not very ready to be used. However, there exists some characterization for tightness. Before moving on, note that until now we only considered metrizable topological space. Often, one would like to impose two additional regularity properties:

separability and completeness. Recall that a topological space is separable if it has a countable dense subset. We say that a metric space (S, d) is complete if every Cauchy sequence converge to a limit in S, where a sequence $(x_n)_{n\in\mathbb{N}}$ satisfies the Cauchy requirement if, for every $\varepsilon > 0$, the exists $\bar{n} \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for every m, n greater than $\bar{n}$. Completeness is useful for characterizing compactness, because a closed subset of a complete metric space is compact if and only if it is totally bounded. We will often work with metrics that are not complete, nevertheless it is usually possible to construct a topologically equivalent metric under which the space becomes complete. We will refer to this property as topologically completeness, meaning that even if the space is not complete under the metric that we are considering, we can make the space complete using a different metric that generates the same topology.

Proposition 5.1.1. *If the space $(S, \mathcal{S})$ is separable and complete, then every probability measure $\mathbb{P}$ on $(S, \mathcal{S})$ is tight.*

We will come back later on to the concept of tightness. We only want to highlight that one usually works with Polish spaces, i.e. complete separable metric spaces, hence the assumptions of the previous theorem are usually guaranteed. We will be interested in tightness of family of probability measure, and again separability and completeness will play a fundamental role.

We want now to focus our attention on probability measures. In particular, one may asks whether there exists a characterization to uniquely determine a measure, that is, if there exists a certain condition satisfied by two measure, then the two measure are indeed the same measure. This characterization involves determining class.

Definition 5. We say that $\mathcal{V} \subset \mathcal{S}$ if a *determining class* for $(S, \mathcal{S})$ if $\mathcal{V}$ generates $\mathcal{S}$ and, whenever two probability measures P_1 and P_2 take equal, finite values on every element of $\mathcal{V}$, then $P_1 \equiv P_2$.

Of course there are conditions that allows us to show that a given $\mathcal{V}$ is a determining class for $(S, \mathcal{S})$. For example, a sufficient conditions for $\mathcal{V}$ to be a determining class are that S is a countable union of elements of $\mathcal{V}$ and that $\mathcal{V}$ is a semiring: that is, $\emptyset \in \mathcal{V}$ and for any pair of sets $A, B \in \mathcal{V}$, we have that $A \cap B \in \mathcal{V}$ and $A \setminus B$ is a finite union of disjoint elements of $\mathcal{V}$. Using these conditions it can be shown that a measure defined on the finite-length half-open intervals of $\mathbb{R}$ has a unique extension to the Borel sets of $\mathbb{R}$. Let's now define the set of real bounded and continuous functions on S

$$\mathcal{C}_b(S) := \{f : S \mapsto \mathbb{R} \mid f \text{ bounded and continuous }\}. \tag{5.5}$$

This set will be crucial to establish many properties of probabilities on $\mathcal{S}$. The next theorem shows how to use $\mathcal{C}_b(S)$ to uniquely determine a probability measure on $\mathcal{S}$. It states that the values of $\int_S f d\mathbb{P}$ for $f \in \mathcal{C}_b(S)$ completely determine the values of $\mathbb{P}$.

Theorem 5.1.1. *Let P_1 and P_2 be two probability measure on $(S, \mathcal{S})$. Then, $P_1 \equiv P_2$ if and only if*

$$\int_S f dP_1 = \int_S f dP_2 \qquad \forall f \in \mathcal{C}_b(S). \tag{5.6}$$

The idea of the proof consists of two steps. Firstly, we need to show that the class of closed set is a determining class, meaning that every probability $\mathbb{P}$ is determined by the values of $\mathbb{P}(C)$ for closed sets $C \in \mathcal{S}$. (This comes as a consequence of the fact that every probability measure $\mathbb{P}$ on $\mathcal{S}$ is regular, where by regular we mean that, given $\mathcal{A} \in \mathcal{S}$, for every $\varepsilon > 0$ there exist a closed set $C \subseteq \mathcal{A}$ and an open set $O \supseteq \mathcal{A}$ such that $\mathbb{P}(O \setminus C) < \varepsilon$.) Secondly, we need to show that every indicator function $\mathbb{1}_C$ of closed set $C \in \mathcal{S}$ can be approximated by elements of $\mathcal{C}_b(S)$.

From this theorem, the possibility to define weak convergence by means of integrals of functions in $\mathcal{C}_b(S)$ should appear more clear. If $(P_n)_{n\in\mathbb{N}}$ is a sequence of probability measures on $(S, \mathcal{S})$, we are led to say that the sequence converge weakly to a limit probability P if the following limit holds for all $f \in \mathcal{C}_b(S)$

$$\lim_{n\to\infty} \int_S f dP_n = \int_S f dP. \tag{5.7}$$

We are tempted to require more. For example that the previous relation holds for a different class of measurable function consisting of all indicator functions. The previous equation would become

$$\lim_{n\to\infty} P_n(A) = P(A) \qquad \forall A \in \mathcal{S}. \tag{5.8}$$

But the latter condition is too strong. Consider, for example, the probability measures $(P_n)_{n\in\mathbb{N}}$,P that give unit mass to the points $(x_n)_{n\in\mathbb{N}}$,x in S. If $x_n \to x$ with $x_n \neq x$ for all n, we have $P(\{x\}) = 1$ by definition, but $P_n(\{x\}) = 0$ for all n, hence the required convergence fails. The reason of this failure lies behind the nature of the set $A = \{x\}$. We need to avoid set A with positive P-measure on boundaries. The equivalence of an integral definition to the more natural relation (5.4) is exactly the statement of Portmanteau Theorem, as stated in the following theorem.

Theorem 5.1.2. *[Portmanteau] Let $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$ be probability measures on $(S, \mathcal{S})$. The following conditions are equivalent:*

1. $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, *i.e.* $\mathbb{P}_n$ *converges weakly to* $\mathbb{P}$;
2. $\lim_{n\to\infty} \int_S f d\mathbb{P}_n = \int_S f d\mathbb{P} \quad \forall f \in \mathcal{C}_b(S)$;
3. $\limsup_n \mathbb{P}_n(C) \leq \mathbb{P}(C)$, *for every closed set* $C \in \mathcal{S}$;
4. $\liminf_n \mathbb{P}_n(O) \geq \mathbb{P}(O)$, *for every open set* $O \in \mathcal{S}$;
5. $\lim_n \mathbb{P}_n(\mathcal{A}) = \mathbb{P}(\mathcal{A})$, *for every set* $\mathcal{A} \in \mathcal{S}$ *such that* $\mathbb{P}(\partial\mathcal{A}) = 0$.

The previous theorem provides useful conditions equivalent to weak convergence. Every condition could be taken as definition. We emphasize in particular the second and the last conditions, commonly used as definition of weak convergence of probability measures. Note that from Theorem 5.1.1 we have that the integrals $\int_S f d\mathbb{P}$ completely determine $\mathbb{P}$. Hence, since the sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ cannot converge to different limits at the same time, this uniquely determines the limit probability measure $\mathbb{P}$.

Using Theorem 5.1.2 as a definition of weak convergence, we can easily obtain as a by product the notion of weak convergence in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On this space, the general probability $\mathbb{P}$ has distribution function F defined by

$$F(x) := \mathbb{P}(y : y \leq x) \qquad x \in \mathbb{R} \tag{5.9}$$

It follows from the definition above that F is non decreasing. The function F is continuous from the right since, for every $x \in \mathbb{R}$, we have

$$F(x^+) = \lim_{\delta \to 0^+} F(x+\delta) = \bigcap_{\delta > 0} \mathbb{P}(y : y \leq x + \delta) = \mathbb{P}(y : y \leq x) = F(x) \tag{5.10}$$

Hence F is continuous at x if and only if it is continuous from below, i.e. if $F(x^-) = F(x)$. But we have, for every $x \in \mathbb{R}$, that

$$F(x^-) = \lim_{\delta \to 0^+} F(x-\delta) = \bigcup_{\delta > 0} \mathbb{P}(y : y \leq x - \delta) = \mathbb{P}(y : y < x) \tag{5.11}$$

To conclude, we have that F is continuous at x if and only if $\mathbb{P}(y : y \leq x) = \mathbb{P}(y : y < x)$, i.e. if and only if the interval $(-\infty, x]$ is a $\mathbb{P}$-continuity set. Thus weak convergence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ implies the ordinary convergence of distribution functions expressed in (5.1). Also the converse is true, but we need to introduce the notion of convergence determining class. It results to be very useful to prove weak convergence by showing that $\mathbb{P}_n(\mathcal{A}) \to \mathbb{P}(\mathcal{A})$ for some special class of sets $\mathcal{A}$.

Definition 6. We say that $\mathcal{U} \subset \mathcal{S}$ is a convergence determining class for the sequence $(\mathbb{P}_n)_{n \in \mathbb{N}}$ if the following convergence

$$\mathbb{P}_n(\mathcal{A}) \to \mathbb{P}(\mathcal{A}) \qquad \forall A \in \mathcal{U} \tag{5.12}$$

implies that $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

The following Proposition allows us to characterize a convergence determining class

Proposition 5.1.2. *If $\mathcal{U} \subset \mathcal{S}$ is such that:*

- *$\mathcal{U}$ is closed under finite intersections;*
- *each open set of $\mathcal{S}$ can be obtained as a finite or countable union of elements of $\mathcal{U}$;*

then $\mathcal{U}$ is a convergence determining class for $(\mathbb{P}_n)_{n \in \mathbb{N}}$.

Remark 2. If S is separable, then the second condition of the previous proposition can be replaced by requiring that, for every $x \in S$ and every $\varepsilon > 0$, there exists $\mathcal{A} \in \mathcal{U}$ such that $x \in \mathcal{A} \setminus \partial\mathcal{A}$ and $\mathcal{A} \subset B_\varepsilon(x)$, where $B_\varepsilon(x)$ denotes the open ball of radius ε centered at x.

Going back to our distributions functions, let $\mathcal{U}$ be the class of intervals $(a, b]$ for which the extremes points have $\mathbb{P}$-measure 0. We have that $\mathcal{U}$ is a convergence determining class. In facts, $\mathcal{U}$ is closed under finite intersections and, since only countably many points can have positive $\mathbb{P}$-measure, it verifies the condition expressed in remark 2. Note that every extreme point of an interval in $\mathcal{U}$ is a continuity point for F. Therefore, since $F_n \xrightarrow{w} F$ implies that $\mathbb{P}_n((a, b]) \to \mathbb{P}((a, b])$ holds for every $(a, b] \in \mathcal{U}$, and $\mathcal{U}$ is a convergence determining class, we showed that, in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $F_n \xrightarrow{w} F$ is equivalent to $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

5.1.1 Measures on $[0, 1]$

This serves only as a simple example. The next case we want to describe is $S = \mathcal{C} = \mathcal{C}([0, 1])$, the space of continuous functions on the unit interval, endowed with the uniform metric $d(x, y) = \sup_{t\in[0,1]} |x(t) - y(t)|$. As first we want to define the class of projection maps. For any $k \in \mathbb{N}$ and any $\{t_1, ..., t_k\} \in [0, 1]^k$ we define the map

$$\begin{aligned} \pi_{t_1,..,t_k} : \mathcal{C}([0, 1]) &\mapsto \mathbb{R}^k \\ x &\to (x(t_1), ..., x(t_k)) \end{aligned} \tag{5.13}$$

Varying $k \in \mathbb{N}$ and the points $\{t_1, ..., t_k\} \in [0, 1]^k$, the sets of the form $\pi^{-1}_{t_1,..,t_k}(H)$ for $H \in \mathcal{B}(\mathbb{R}^k)$ are called finite-dimensional sets. Note that $\pi_{t_1,..,t_k}$ is continuous and the finite-dimensional sets belong to $\mathcal{B}(\mathcal{C})$, the Borel σ-algebra of $\mathcal{C}([0, 1])$, and they form a finitely additive field. We want to show that the finite-dimensional sets generates the σ-algebra $\mathcal{B}(\mathcal{C})$. Note that

$$\mathcal{B}_r(x) = \bigcap_{t\in\mathbb{Q}\cap[0,1]} \{y \in \mathcal{C} : |y(t) - x(t)| \le r\} \tag{5.14}$$

hence we can write the closed ball of radius $r > 0$ and center $x \in \mathcal{C}$ as a result of a countable intersection of finite-dimensional sets. Then, we can generate open balls through countable union of closed balls, and open sets as countable union of open balls. Hence the finite-dimensional sets generates $\mathcal{B}(\mathcal{C})$. Since they form a field, the finite-dimensional sets are a determining class.
The same result can be formalized by mean of the Kolmogorov extension theorem. Indeed, the finite-dimensional distributions uniquely determine a probability distribution on $\mathcal{C}$, endowed with the Kolmogorov σ-algebra generated by the coordinate projections. Moreover, the Borel σ-algebra on $\mathcal{C}$ with the uniform norm coincides with the Kolmogorov σ-algebra generated by the coordinate projections. Hence the finite-dimensional distributions also determine the distribution of a stochastic process with sample paths in $\mathcal{C}$.

Now, let's consider the functions $x \equiv 0$ and x_n given by

$$x_n(t) := \begin{cases} nt & \text{if } t \in [0, 1/n] \\ 2 - nt & \text{if } t \in [1/n, 2/n] \\ 0 & \text{if } t \in [2/n, 1] \end{cases} \tag{5.15}$$

The piecewise linear trajectory of x_n starts in 0, then reach the peak value of 1 at time $1/n$, then at time $2/n$ it comes back to 0 and here it stays. Let P be the probability

measure that assign unit mass to x, meaning that $P(A)$ is 1 or 0 depending on x lying in A or not. In a similar way, define, for $n \in \mathbb{N}$, P_n to be the unit mass at point x_n. We can have weak convergence $P_n \xrightarrow{w} P$ if and only if we have uniform convergence of x_n to x, i.e. if we have convergence in the topology of $\mathcal{C}$. But we have that

$$\sup_{t\in[0,1]} |x_n(t) - x(t)| = 1 \qquad \forall n \in \mathbb{N}. \tag{5.16}$$

Hence weak convergence of $(P_n)_{n\in\mathbb{N}}$ is not possible. However, for any finite-dimensional set of the form $A = \pi^{-1}_{t_1,..,t_k}(H)$ we have that $P_n(A) = P(A)$ whenever n is big enough so that $2/n < \inf\{t_1, .., t_k\}$. Hence the finite-dimensional sets are a determining class, but not a convergence determining class.

For any given probability $\mathbb{P}$ on $\mathcal{C}$, The various measures $\mathbb{P} \circ \pi^{-1}_{t_1,..,t_k}$ on the spaces $\mathbb{R}^k$ are called the finite-dimensional distributions of $\mathbb{P}$. Since finite-dimensional sets are a determining class, then the finite-dimensional distributions uniquely determine a measure. On the other hand, we showed that the finite-dimensional sets are not a convergence determining class, hence convergence of finite-dimensional distributions is not enough to exhibit weak convergence. The difficulty, interest and usefulness of weak convergence in $\mathcal{C}$ all spring from the fact that it involves considerations that goes beyond the convergence finite-dimensional distributions. We will come back to the space $\mathcal{C}$ of continuous function in the following.

5.1.2 Prohorov convergence

Recall the (S, d) is our reference metric space, and we defined probability measures on it. We change now our perspective by considering the space $\mathcal{P} = \mathcal{P}(S)$ of all probability measures on the metric space S. We define the Prohorov metric π on $\mathcal{P}$ as the one induced by the metric d on the space S. For any $P_1, P_2 \in \mathcal{P}(S)$ we have

$$\pi(P_1, P_2) := \inf\{\varepsilon > 0 : P_1(\mathcal{A}) \le P_2(\mathcal{A}^\varepsilon) + \varepsilon, P_2(\mathcal{A}) \le P_1(\mathcal{A}^\varepsilon) + \varepsilon \quad \forall \mathcal{A} \in \mathcal{B}(S)\} \tag{5.17}$$

where we denoted by $\mathcal{A}^\varepsilon := \{y \in S : d(y, \mathcal{A}) < \varepsilon\}$ the open ε-neighborhood of a set $\mathcal{A} \in \mathcal{S}$. It can be shown that π is a metric, being non-negative by definition, satisfying the triangular inequality and the symmetry property. Thus $(\mathcal{P}, \pi)$ is a metric space, and it turns out that is separable and complete.

A sequence of probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$ in the metric space $(\mathcal{P}, \pi)$ converges to the limit $\mathbb{P} \in \mathcal{P}$ if, for all $\varepsilon > 0$, there exists an integer $\bar{n}$ such that $\pi(\mathbb{P}_n, \mathbb{P}) < \varepsilon$ for all $n \ge \bar{n}$. This is the classical notion of convergence in metric spaces. The next theorem claim equivalence of this latter definition with the weak convergence.

Theorem 5.1.3 (Prohorov convergence)**.** *For any separable metric space (S, d), the function π on $\mathcal{P}(S)$, defined as in (5.17), is a separable metric. If $(\mathbb{P}_n)_{n\in\mathbb{N}}$ is a sequence of probability measures on the space $(S, \mathcal{S})$, convergence $\pi(\mathbb{P}_n, \mathbb{P}) \to 0$ on the space $\mathcal{P}(S)$ is equivalent to weak convergence $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ as stated in Theorem 5.1.2.*

The topology induced by the Prohorov metric is quite complicated, but we can bypass this problem. This is done by means of random elements. In facts, we are primarily interested in weak convergence $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, and random elements manage to relate convergence of the probability measure the problem to another convergence easier to handle, as we will see further on.

We now want to give other characterization of weak convergence. A sequence $(x_n)_{n\in\mathbb{N}}$ of points in a metric space converge to a limit x if and only if every subsequence x_{n^k} has a further subsequence $x_{n^{k_i}}$ converging to x. The same can be said for weak convergence, as stated in the following proposition.

Proposition 5.1.3. *We have weak convergence $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ if and only if every subsequence $\mathbb{P}_{n^k}$ has a further subsequence $\mathbb{P}_{n^{k_i}}$ that converges weakly to $\mathbb{P}$.*

Hence we can characterize weak convergence through converging subsequence. This should suggest that compactness enter the game. In a metric space, a subset K is compact if every sequence $(x_n)_{n\in\mathbb{N}}$ of elements of K has a convergent subsequence with limit in K. It is possible to define compactness even in topological spaces, without using any metric. Recall that an open cover of a set K is a collection of open sets such that K is contained in the union of those sets. A subset K of a topological space is compact if each open cover of K has a finite subcover. It turns out that compactness plays a crucial role for proving weak convergence, as we will see in Section 5.3.

5.2 Convergence in distribution

In this section we want to relate weak convergence to the more known theory of convergence in distribution. It turns out that it is the same concept stated in different terms. Thus, many results can be borrowed from distributional convergence theory without any effort. We introduce for the first time the concept of random elements, which we specialize in the following to be random variables or random functions.

Definition 7. A random element X is a measurable mapping from a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ to a metric space $(S, \mathcal{S})$, where $\mathcal{S}$ is the topology induced by the metric of S.

In the underlying probability space, Ω is a set, $\mathcal{B}$ is a σ-algebra and $\mathbb{P}$ is a probability measure. With reference to the spaces introduced in the previous section, if $S = \mathbb{R}$ we call X a random variable, whereas if $S = \mathcal{C}$ we refer to X a stochastic process. We define the distribution P of the random element X on the space $(S, \mathcal{S})$ a the push-forward of the probability measure $\mathbb{P}$ through X, i.e. $P := \mathbb{P} \circ X^{-1}$. Hence, we have

$$P(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\omega \in \Omega : X(\omega) \in A) = \mathbb{P}(X \in A), \qquad \forall A \in \mathcal{S} \tag{5.18}$$

When dealing with random elements, we are primarily interested on their distributions, hence the underlying probability $\mathbb{P}$ is usually left unspecified. The probability space $(\Omega, \mathcal{B}, \mathbb{P})$ is indeed arbitrary, whereas the probability P is always defined on a metric

space. In the case $S = \mathbb{R}$ we can recover the distribution function defined in (5.9), and we obtain the following relation

$$F(x) = P(y : y \leq x) = \mathbb{P}(X \leq x) \qquad x \in \mathbb{R}. \tag{5.19}$$

The distribution induced by a random element allows us to easily compute certain quantities like expectations. Computation of integrals with respect to a general probability measure reduce, thank to a simple change of variables, to a more treatable form:

$$\mathbb{E}[h(X)] := \int_\Omega h(X(\omega))d\mathbb{P}(\omega) = \int_S h(x)dP(x). \tag{5.20}$$

Note that each probability measure on a metric space can be obtained as the distribution if some random element on some probability space, just take the random element to be the identity function on the space itself. Hence the class of distributions coincide with the class of probability measures on metric spaces. Nevertheless, we reserve to call a probability measure on a metric space a distribution only when it is properly a distribution of some random element already under discussion.

Some results can be stated in the new language of convergence in distribution. For example, we say that a sequence of random elements $(X_n)_{n\in\mathbb{N}}$ convergence in distribution to a random element X if and only if the sequence od distributions $(P_n)_{n\in\mathbb{N}}$ induced by $(X_n)_{n\in\mathbb{N}}$ converges weakly to the distribution P induced by X. In formula:

$$X_n \xrightarrow{d} X \quad \iff \quad P_n \xrightarrow{w} P \tag{5.21}$$

As the reader probably already understood, we can also rewrite Portmanteau Theorem with the language of convergence in distribution. We call a set $A \in \mathcal{S}$ an $X-$continuity set if $\mathbb{P}(X \in \partial A) = 0$. We have the equivalence of the following conditions:

1. $X_n \xrightarrow{d} X$, i.e. X_n converges in distribution to X;
2. $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad \forall f \in \mathcal{C}_b(S)$;
3. $\limsup_n \mathbb{P}_n(X_n \in C) \leq \mathbb{P}(X \in C)$, for every closed set $C \in \mathcal{S}$;
4. $\liminf_n \mathbb{P}_n(X_n \in O) \geq \mathbb{P}(X \in O)$, for every open set $O \in \mathcal{S}$;
5. $\lim_n \mathbb{P}_n(X_n \in \mathcal{A}) = \mathbb{P}(X \in \mathcal{A})$, for every X- continuity set $\mathcal{A} \in \mathcal{S}$.

Note that, in order to make sense, we need that $(X_n)_{n\in\mathbb{N}}$ and X all take values in the same metric space S, whereas it is not important on which particular probability space they are defined. In facts, the only things that matter are the distributions P_n and P induced on $(S, \mathcal{S})$. To be formal, the correct notation would say that X is defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ and each of the X_n is defined on a different probability space $(\Omega_n, \mathcal{B}_n, \mathbb{P}_n)$. Thus we should write $\mathbb{P}_n(X_n \in \mathcal{A})$ or $\mathbb{E}_n[f(X_n)]$ meaning that the integral $\int_{\Omega_n} f(X_n)d\mathbb{P}_n$ is with respect to the probability measure $\mathbb{P}_n$ on $\mathcal{B}_n$. Nevertheless, thanks to (5.18) and (5.20) we can always refer to and compute this quantities using the induced

distributions P_n defined on $(S, \mathcal{S})$. Hence, the spaces $(\Omega_n, \mathcal{B}_n, \mathbb{P}_n)$ don't play any role and we are allowed to omit their specification. Hence, since we make no specific mention to the structure of these probability spaces, we can suppress some notation and write $\mathbb{P}(X_n \in \mathcal{A})$ for $\mathbb{P}_n(X_n \in \mathcal{A})$ and $\mathbb{E}[f(X_n)]$ for $\mathbb{E}_n[f(X_n)]$, unless we need to specify the actual spaces.

In subsection 5.1.2 we established equivalence between weak convergence $P_n \xrightarrow{w} P$ in a metric space (S, d) and convergence in the Prohorov space $(\mathcal{P}(\mathcal{S}), \pi)$. Now, if we think the weakly converging measures $P, (P_n)_{n\in\mathbb{N}}$ to be the laws of random elements $X, (X_n)_{n\in\mathbb{N}}$, we can state that we have convergence in distribution $X_n \xrightarrow{d} X$. The following Theorem says something more. We write $\stackrel{d}{=}$ to indicate that two objects have the same distribution.

Theorem 5.2.1. *[Skorokhod representation theorem] Suppose that $X_n \xrightarrow{d} X$ in a separable metric space S. Then there exist other random elements $(\tilde{X}_n)_{n\in\mathbb{N}}, \tilde{X}$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with values in S, such that*

$$\tilde{X} \stackrel{d}{=} X, \qquad \tilde{X}_n \stackrel{d}{=} X_n \quad \text{for } n \in \mathbb{N} \tag{5.22}$$

and

$$\mathbb{P}\left(\lim_{n\to\infty} \tilde{X}_n = \tilde{X}\right) = 1. \tag{5.23}$$

Hence from a sequence of weak converging laws in a metric space we are able to build a sequence of random elements, with the same laws, that converges almost surely to the same limit. Note that w can also prove sure convergence rather than merely almost sure convergence. It is not surprising that this is possible, since changing the value of all the X_n on a set with measure zero does not affect the conditions for convergence in distribution.

5.3 Compactness Approach

We have seen how to prove weak convergence using mappings between spaces. We want now to proceed and investigate other techniques that help us to show weak convergence. The notion of relative compactness turns out to be essential.

Definition 8. A family Π of probability measures on $(S, \mathcal{S})$ is relatively compact if every sequence of elements of Π contains a weakly convergent subsequence.

Thus, if Π is relatively compact, then every sequence $(P_n)_{n\in\mathbb{N}}$ has a subsequence $(P'_n)_{n\in\mathbb{N}}$ weakly convergent to some limit Q. Note that Q is a probability measure on S but not necessarily an element of Π. Before moving on, we want to highlight the importance of relatively compactness. Suppose that we have a sequence of probability measure $(P_n)_{n\in\mathbb{N}}$ on the space $\mathcal{C}$ of continuous functions such that their finite-dimensional distributions converge weakly to those of P. We know that this condition is not sufficient to imply weak convergence of the sequence itself, since the finite-dimensional distributions on $\mathcal{C}$ are a determining class, but not a convergence determining class. Assume also that

the sequence $(P_n)_{n\in\mathbb{N}}$ is relatively compact. We want to show that, under this additional assumption, the actual sequence $(P_n)_{n\in\mathbb{N}}$ converge weakly to P.

Starting from the definition, we have that every subsequence $(P'_n)_{n\in\mathbb{N}}$ has a further subsequence $(P''_n)_{n\in\mathbb{N}}$ converging weakly to some limit Q. Since $P''_n \xrightarrow{w} Q$, then also the finite-dimensional distributions of P''_n converge weakly to those of Q. Then, the finite-dimensional distributions of Q must coincide to those of P, since we know that the finite-dimensional distributions of P_n converge weakly to those of P, and any weakly convergent subsequence $(P''_n)_{n\in\mathbb{N}}$ must converge to the same weak limit of the original sequence $(P_n)_{n\in\mathbb{N}}$. Using the fact that a probability measure on $\mathcal{C}$ is completely determined by its finite-dimensional distributions, we conclude that we must have $P \equiv Q$. Hence, we showed that any subsequence $(P'_n)_{n\in\mathbb{N}}$ contains a further subsequence $(P''_n)_{n\in\mathbb{N}}$ that is weakly convergent to P. It follows from Proposition 5.1.3 that $P_n \xrightarrow{w} P$.
This idea is a powerful technique to prove weak convergence in $\mathcal{C}$ and other functional spaces. As first we have to prove that the finite-dimensional distributions converge weakly to some limit and we have to identify the limit. Next we have to prove that the sequence $(P_n)_{n\in\mathbb{N}}$ is relatively compact. Note that we only require the sequence $(P_n)_{n\in\mathbb{N}}$ to be relatively compact and, for any $k \in \mathbb{N}$ and any $(t_1, ..., t_k)$, we ask $P_n \circ \pi^{-1}_{t_1,..,t_k}$ to converge weakly to some probability measure $\mu_{t_1,..,t_k}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. We are not assuming a priori that $\mu_{t_1,..,t_k}$ are the finite-dimensional distributions of a probability measure P. However, it still follows that each subsequence $(P'_n)_{n\in\mathbb{N}}$ has a further subsequence $(P''_n)_{n\in\mathbb{N}}$ that converges weakly to some limit. The limit is unique, since every limit must have $\mu_{t_1,..,t_k}$ as its finite-dimensional distributions, and the finite-dimensional distributions completely characterize a probability measure. Hence, using this procedure to prove weak convergence of the sequence $(P_n)_{n\in\mathbb{N}}$, we obtain as a byproduct the limit P to which it converges.
To use this method we need an effective criterion for relative compactness. The notion of tightness plays a central role.

Definition 9. A family of probability measure Π on the metric space $(S, \mathcal{S})$ is said to be tight if, for every $\varepsilon > 0$, there exists a compact set K such that $\mathbb{P}(K) > 1 - \varepsilon$ for every $\mathbb{P} \in \Pi$.

Note that if the class Π consists of single measure $\mathbb{P}$, this definition reduces to Definition 4. The following theorem, due to Prohorov, will relate tightness to relative compactness for a family Π of probability measures.

Theorem 5.3.1 (Prohorov)**.** *Let* Π *be a family of probability measures on the metric space* $(S, \mathcal{S})$*.*

- *If* Π *is tight, then it is relatively compact.*
- *Assume* S *to be separable and complete. If* Π *is relatively compact, then it is tight.*

Hence, provided the space S to be separable and complete, tightness is a sufficient condition to show relative compactness. We do not treat here any criteria to show tightness, but the reader can find reference for example in [Whi02, sec 11.6].

5.3.1 Compactness approach in $\mathcal{C}$

We already introduced the compactness approach for probability measure on a metric space. In this section we want to analyze the specific case of the space $\mathcal{C}$ of continuous functions on $\mathbb{R}$. In the thesis we deal with the space $\mathcal{D}$, functions whose trajectories are càdlàg . The techniques we need are adjustments from the same techniques one uses in $\mathcal{C}$, hence this section should be regarded as a guided commented example, where the ideas already introduced are carried out using the same stochastic process notation we will use for $\mathcal{D}$. In particular, we want to describe the standard approach to establish stochastic process limits based on compactness.

Let's start from some consideration on the space $\mathcal{C}([0,1))$ of the real-valued continuous functions defined on the unit interval. This serves as a motivation, and the considerations done here can be extended to more general domains-codomains. As first we want to introduce the modulus of continuity, for a generic function $x \in \mathcal{C}([0,1))$. There are different notations in literature, and here we basically refer to a simplified version, the least we need in order to make some consideration on the nature of $\mathcal{D}$.

For any given function $x \in \mathcal{C}([0,1])$ and any $\delta > 0$, we define the modulus of continuity w as follows

$$w(x,\delta) := \sup_{|t-s|<\delta} |x(t) - x(s)| \tag{5.24}$$

A necessary and sufficient condition for x to be uniformly continuous over $[0,1]$ is

$$\lim_{\delta \to 0} w(x,\delta) = 0 \tag{5.25}$$

The latter condition is satisfied by every $x \in \mathcal{C}([0,1))$. The Ascoli-Arzelá theorem uses the modulus of continuity w to completely characterize relative compactness in $\mathcal{C}([0,1))$.

Theorem 5.3.2. *(Ascoli-Arzelá in $\mathcal{C}$) The set $A \subset \mathcal{C}([0,1])$ is relatively compact with respect to the uniform topology if and only if the two following conditions are satisfied:*

$$\sup_{x \in A} |x(0)| \leq \infty \tag{5.26}$$

$$\lim_{\delta \to 0} \sup_{x \in A} w(x,\delta) = 0 \tag{5.27}$$

From this theorem it is possible to establish conditions to determine tightness for a sequence of probability measures on $\mathcal{C}$.

Proposition 5.3.1. *The sequence $(P_n)_{n \in \mathbb{N}}$ of probability measures on $\mathcal{C}$ is tight if and only if the following two conditions hols:*

1. *For each positive ε and η, the exists M and $\bar{n}$ such that*

$$P_n(x : |x(0)| > M) \leq \eta, \qquad \forall n \geq \bar{n} \tag{5.28}$$

2. *For each positive ε and η, the exists $\delta \in (0,1)$ and $\bar{n}$ such that*

$$P_n(x : w(x,\delta) \geq \varepsilon) \leq \eta, \qquad \forall n \geq \bar{n} \tag{5.29}$$

Finally, note the relation between the uniform norm (6.3) and the modulus of continuity defined in (5.24).

5.4 Weak convergence and mappings

In the previous sections we gave definitions and characterizations of weak convergence. However, they are difficult to use directly in order to effectively prove weak convergence of a sequence. In various cases, we will use some strategies that allow us to ensure weak convergence without showing it directly. The continuous-mapping approach exploits previously established weak convergence and the continuous-mapping theorem to obtain weak convergence of new quantities of interest. Alternative approaches are the already mentioned compactness approach and various stochastic approaches which exploit special structures, such as Markov and martingale structure, which we do not mention in this thesis. Let h be a mapping between two metric spaces

$$h : (S, \mathcal{S}) \mapsto (S', \mathcal{S}') \tag{5.30}$$

Suppose that, on the space $(S, \mathcal{S})$ the sequence of probability measure $(P_n)_{n\in\mathbb{N}}$ converges weakly to the probability measure P. We know that the mapping h induces the distributions $(P_n \circ h^{-1})_{n\in\mathbb{N}}$ and $P \circ h^{-1}$ on the space $(S', \mathcal{S}')$. We would like to state that $P_n \circ h^{-1} \xrightarrow{w} P \circ h^{-1}$. However this is not verified in general, unless the mapping h satisfies certain constraints. For example everything works fine if h is continuous. In facts, a necessary and sufficient condition to show weak convergence is that

$$\int_{S'} f(y) P_n \circ h^{-1}(dy) \to \int_{S'} f(y) P \circ h^{-1}(dy) \qquad \forall f \in \mathcal{C}_b(S') \tag{5.31}$$

Note that the previous equation, upon a change of variables, is equivalent to

$$\int_{S} f \circ h(x) P_n(dx) \to \int_{S} f \circ h(x) P(dx) \qquad \forall f \in \mathcal{C}_b(S) \tag{5.32}$$

and this is true since $P_n \xrightarrow{w} P$ and $f \circ h \in \mathcal{C}_b(S)$. As an example, take h to be the projection $\pi_{t_1,..,t_k}$ from $\mathcal{C}([0,1])$ to $\mathbb{R}^k$ al ready defined in (6.24) for any choice of $k \in \mathbb{N}$ and $\{t_1, ..., t_k\} \in [0,1]^k$. The function is continuous, hence weak convergence $P_n \xrightarrow{w} P$ implies weak convergence of the finite-dimensional distributions $P_n \circ \pi^{-1}_{t_1,..,t_k} \xrightarrow{w} P \circ \pi^{-1}_{t_1,..,t_k}$. Recall that the converse fails because, as shown by counterexample, the class on finite-dimensional set is not convergence determining. Of course we can state the same result also in the language of convergence in distribution

Theorem 5.4.1. *(continuous mapping theorem) If $X_n \xrightarrow{d} X$ in $(S, \mathcal{S})$ and $h : (S, \mathcal{S}) \mapsto (S', \mathcal{S}')$ is continuous, then*

$$h(X_n) \xrightarrow{d} h(X) \quad \text{in } (S', \mathcal{S}') \tag{5.33}$$

It's possible to weaken the continuity assumption, provided that the discontinuities are under control. Let $h : (S, \mathcal{S}) \mapsto (S', \mathcal{S}')$ be a measurable function. We do not require continuity for h, and we call $Disc[h]$ the set of its discontinuities. Suppose we have sequence $(P_n)_{n\in\mathbb{N}}, P$ of probability measures on $(S, \mathcal{S})$ converging weakly to P.

We would like to know under which conditions on h, convergence $P_n \xrightarrow{w} P$ imply weak convergence of $P_n \circ h^{-1} \xrightarrow{w} P \circ h^{-1}$. It turns out that it suffices h to have discontinuities occurring on a set of P-measure 0. The power of this result is immediately clear when stated with the language of convergence in distribution.

Theorem 5.4.2. *Let $(X_n)_{n\in\mathbb{N}}, X$ be random elements taking values in S, and suppose that $X_n \xrightarrow{d} X$. Let the function $h : (S, \mathcal{S}) \mapsto (S', \mathcal{S}')$ be measurable, and denote by $Disc[h]$ the set of its discontinuity points. If $P(X \in Disc[h]) = 0$, then $h(X_n) \xrightarrow{d} h(X)$.*

Hence, if we have a sequence of random elements $(Y_n)_{n\in\mathbb{N}}$ obtained applying a measurable transformations h to another converging sequence $(X_n)_{n\in\mathbb{N}}$, then proving convergence in distribution of $Y_n = h(X_n)$ simply reduce to check whether h satisfies the discontinuity condition of the previous theorem. In particular, Theorem 5.4.2 extends the continuous mapping theorem to settings where we dispose of measurable functions that are only continuous almost everywhere.
To conclude, we give another version the the continuous mapping theorem, extending the result to the situation where we have a sequence of such measurable functions converging to a limit function and we ask the conditions under which $h_n(X_n) \xrightarrow{d} h(X)$. We claim that it suffices to have $h_n(x_n) \to h(x)$ whenever $x_n tox$ for a subset $E \subseteq S$ of limit points such that $P(X \in E) = 1$.

Theorem 5.4.3. *[generalized continuous-mapping theorem] Let h and $(h_n)_{n\in\mathbb{N}}$ be measurable functions mapping the separable metric space $(S, \mathcal{S})$ into the metric space $(S', \mathcal{S}')$. Let E be the set of $x \in S$ such that $h_n(x_n) \to h(x)$ fails for some sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \to x$. If $X_n \xrightarrow{d} X$ and $\mathbb{P}(X \in E) = 0$, then $h_n(X_n) \xrightarrow{d} h(X)$.*

Proof. Note that, if $g_n \equiv g$ for all n, then $E = Disc[h]$, so that Theorem 5.4.3 contains both Theorems 5.4.1 and 5.4.2 as special cases.

This generalization is an application of the Skorokhod representation Theorem 5.2.1. Starting with the convergence in distribution $X_n \xrightarrow{d} X$, we use Skorokhod representation theorem to obtain the copies $(\tilde{X}_n)_{n\in\mathbb{N}}$ and $\tilde{X}$ with the same distribution of $(X_n)_{n\in\mathbb{N}}$ and such that X and $\tilde{X}_n \to \tilde{X}$ almost surely. We have that $\mathbb{P}(\tilde{X} \in E) = 0$, since $\tilde{X} \overset{d}{=} X$ and $\mathbb{P}(X \in E) = 0$. It follows that also $h_n(\tilde{X}_n) \to h(\tilde{X})$ almost surely. It is well know that almost sure convergence implies convergence in distribution, hence $h_n(\tilde{X}_n) \xrightarrow{d} h(\tilde{X})$. Finally, observe that $h_n(\tilde{X}_n) \overset{d}{=} h_n(X_n)$ for all $n \in \mathbb{N}$, and analogously $h(\tilde{X}) \overset{d}{=} h(X)$. Thus, the desired convergence in distribution is proven. □

5.5 Characteristic functions

Suppose now we have a probability $\mathbb{P}$ defined on the space $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Its characteristic function p is defined as follows

$$p(t) = \mathbb{E}\left[e^{it\cdot x}\right] = \int_{\mathbb{R}^k} e^{it\cdot x}\mathbb{P}(dx) \qquad t \in \mathbb{R}^k. \tag{5.34}$$

The characteristic function of a probability measure is a really important quantity, since it defines the measure. We have the following result:

Theorem 5.5.1. *Let P, Q be two probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, and let p, q be the respective characteristic functions. If $p(t) = q(t)$ for all $t \in \mathbb{R}^k$, then $P \equiv Q$.*

Note that the function $f(x) = e^{it\cdot x} \in \mathcal{C}_b(\mathbb{R}^k)$. Hence the previous theorem refines Theorem 5.1.1, that asserts that a probability measure P is determined by the values of $\int f dP$ for $f \in \mathcal{C}_b(S)$. Characteristic functions are also useful to deal with weak convergence. Suppose that $(P_n)_{n\in\mathbb{N}}$ and P are probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ with characteristic functions $(p_n)_{n\in\mathbb{N}}$ and p. The following theorem states that convergence of the characteristic functions is equivalent to weak convergence of the probability measure.

Theorem 5.5.2. *The sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ converges weakly to P if and only if, for every $t \in \mathbb{R}^k$ the sequence $(p_n(t))_{n\in\mathbb{N}}$ converge pointwise to $p(t)$.*

Of course necessity follows from the fact that $f(x) = e^{it\cdot x} \in \mathcal{C}_b(\mathbb{R}^k)$. We want to convince the reader that convergence of $p_n(t) \to p(t)$ for all t is indeed sufficient to prove $P_n \xrightarrow{w} P$. Suppose for a moment that the sequence $(P_n)_{n\in\mathbb{N}}$ is tight. Let $g(t) := \lim_{n\to\infty} p_n(t)$. Prohorov Theorem says that if $(P_n)_{n\in\mathbb{N}}$ is tight, then it is relatively compact. It means that every subsequence $(P'_n)_{n\in\mathbb{N}}$ contains a further subsequence $(P''_n)_{n\in\mathbb{N}}$ weakly convergent to some limit, whose characteristic function must be the limit $\lim_{n\to\infty} p_n(t) = g(t)$. This limit is unique, and Theorem 5.5.1 allows us to identify the probability measure P has the only one that has g for characteristic function. This argument is really similar to the one we used to show that a sequence of probability measure in $\mathcal{C}$ converges weakly if and only if it is tight and the finite-dimensional distributions converge weakly. However, the theorem above do not require tightness for the sequence $(P_n)_{n\in\mathbb{N}}$. It is possible to show that continuity of the limit function $g(t) = \lim_{n\to\infty} p_n(t)$ at $t = 0$ is a sufficient condition to obtain tightness for the family $(P_n)_{n\in\mathbb{N}}$. Finally, note that any characteristic function g of a probability measure P is continuous by construction.

To conclude this section we want to highlight analogies and differences between finite-dimensional distributions on $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$ and characteristic functions on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Note that they are difficult type of objects, but they play a similar role when speaking about weak convergence. Let $(P_n)_{n\in\mathbb{N}}$ and P be probability measures defined on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, and let $(p_n)_{n\in\mathbb{N}}$ and p be the respective characteristic functions. We know that the characteristic function p completely determine the probability measure P, and the same holds for p_n. Let $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$ be probability measures on the space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, and, for any choice of $k \in \mathbb{N}$ and $t_1, ..., t_k$ positive real numbers, let $(\mathbb{P}_n \circ \pi^{-1}_{t_1,..,t_k})_{n\in\mathbb{N}}$ and $\mathbb{P} \circ \pi^{-1}_{t_1,..,t_k}$ be the respective finite-dimensional distributions. Analogously to what seen for the characteristic functions, the finite-dimensional distributions $\mathbb{P} \circ \pi^{-1}_{t_1,..,t_k}$ completely determine the measure $\mathbb{P}$. We saw through a counterexample that weak convergence of finite-dimensional distribution does not necessarily imply weak convergence of the sequence of probability measures $(\mathbb{P}_n)_{n\in\mathbb{N}}$. However, if $(\mathbb{P}_n)_{n\in\mathbb{N}}$ is tight and $(\mathbb{P}_n \circ \pi^{-1}_{t_1,..,t_k})_{n\in\mathbb{N}}$ weakly converges to some $\mu_{t_1,..,t_k}$, then also $(\mathbb{P}_n)_{n\in\mathbb{N}}$ converges weakly to some $\mathbb{P}$. The

same kind of results hold for characteristic functions. In facts, if $(p_n(t))_{n\in\mathbb{N}}$ converges weakly to some $g(t)$ for each t and if $(P_n)_{n\in\mathbb{N}}$ is tight, then $(P_n)_{n\in\mathbb{N}}$ converges weakly to some P, but the previous statement fails without tightness. However, tightness is ensured whenever $g(t) := \lim_{n\to\infty} p_n(t)$ is continuous at $t = 0$. If we assume now that $g(t) = p(t)$ is the characteristic function of P, then we can state that pointwise convergence $\lim_{n\to\infty} p_n(t) = p(t)$ for all t implies weak convergence $P_n \xrightarrow{w} P$. On the other hand, assuming that, for any $k \in \mathbb{N}$ and any $t_1, ..., t_k$, $\mu_{t_1,..,t_k} = \mathbb{P} \circ \pi^{-1}_{t_1,..,t_k}$ are the finite-dimensional distributions of $\mathbb{P}$, we have that weak convergence of the finite-dimensional distributions $\mathbb{P}_n \circ \pi^{-1}_{t_1,..,t_k} \xrightarrow{w} \mathbb{P} \circ \pi^{-1}_{t_1,..,t_k}$, together with tightness of $(\mathbb{P}_n)_{n\in\mathbb{N}}$, implies weak convergence of the probability measures $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

Chapter 6

The càdlàg space $\mathcal{D}$

The Brownian motion (see Definition 16 below) is a well known process worth for many applications in different fields, spanning from finance to evolutionary biology, from molecular diffusion to current flow. Due to its independent gaussian increments, it suits a variety of different problems, accommodating the need to model uncertainty with relatively easy computations. There are a vast class of processes that can be obtained as modification of an underlying Brownian motion. However, many other processes do not belong to this class. It is the case, for example, of counting processes, like the Poisson process (see Section 8.1.1). The space $\mathcal{C}$ is unsuitable for the description of processes containing jumps. In this chapter we want to describe the natural space where processes with jumps live, and we want to study weak convergence in this space that contains certain discontinuous functions.
We now focus on the underlying function space of possible sample paths for the stochastic processes we want to study. We deal with stochastic processes with discontinuous trajectories, but not too irregular, allowing only discontinuities of the first kind, usually called jumps.

We consider the space $\mathcal{D}$ of all right-continuous $\mathbb{R}^d$-valued functions with left limits, defined on a sub interval of the real line. To be precise, we will denote by $\mathcal{D}(I, \mathbb{R}^d)$ the spaces of $\mathbb{R}^d$-valued functions defined on a subinterval $I \subseteq \mathbb{R}$ such that, for every $f \in \mathcal{D}(I, \mathbb{R}^d)$ we have

$$\lim_{x \to x_0^+} f(x) = f(x_0) \quad \forall x_0 \in I \tag{6.1}$$

$$\lim_{x \to x_0^-} f(x) = f(x_0^-) < \infty \quad \forall x_0 \in I \tag{6.2}$$

Usually the domain I is the unit interval $[0, 1]$, or the open/closed interval from 0 to T, or the positive semiaxis $[0, \infty)$. When $d = 1$ we will usually drop the codomain, shortening $\mathcal{D}(I, \mathbb{R})$ by $\mathcal{D}(I)$. Finally, we will use the simple notation $\mathcal{D}$ when the domain is clear from the context, or when doing general considerations holding for any interval domain $I \subseteq \mathbb{R}$. With very little change, the theory can be extended to functions taking values in metric spaces other than $\mathbb{R}^d$. The space $\mathcal{D}$ is also known as the space of càdlàg functions,

an acronym for the French *continue à droite, limitée à gauche*. The space $\mathcal{D}$ includes all continuous functions and the discontinuous functions with jump-discontinuity. Hence functions in $\mathcal{D}$ are bounded in any compact set $K \subseteq I$. If $f \in \mathcal{D}$, we denote by $f(x)$ the value of f at point x. Note that the domain $I \subseteq \mathbb{R}$ represents the time of our process, hence we often use letters s, t to indicate time points, while we reserve letters x, y for elements of $\mathcal{D}$. We denote by $x(t^-)$ the left hand limit at time t, and by $\Delta x(t)$ the vales of the jump $x(t) - x(t^-)$.
We also use the abbreviations $\mathcal{D}^+ \equiv \mathcal{D}(\mathbb{R}^+)$ and $\mathcal{D} \equiv \mathcal{D}(\mathbb{R})$. Lastly, we denote by $\mathcal{D}_0$ and $\mathcal{D}_0^+$ the subspaces of nondecreasing functions of $\mathcal{D}$ and $\mathcal{D}^+$, respectively.

6.1 Definitions

We start by considering the space $\mathcal{D}([0,1])$, i.e. by assuming that the domain is the unit interval $[0,1]$ and the range is $\mathbb{R}$. In Section 6.4 we will see how to adapt the results for a generic interval I rather then $[0,1]$, and for codomain $\mathbb{R}^d$ with $d > 1$.
The reference metric is the uniform metric $||\cdot||$, defined in terms of the uniform norm

$$||x|| := \sup_{t\in[0,1]} |x(t)|. \tag{6.3}$$

On the subspace $\mathcal{C}$ the uniform metric works well, but it does not on $\mathcal{D}$. Using the uniform metric, a necessary condition for two functions to be close is that corresponding jumps occur exactly at the same times. We want to relax this requirement, allowing corresponding jumps to occur at close points. Appropriate topologies were introduced by Skorokhod in [Sko56]. The general idea is to define a new metric that consider two functions close if they are uniformly close over $[0,1]$ after allowing small perturbations of time. This is the basic intuition undergoing the J_1-Skorokhod topology on $\mathcal{D}([0,1])$. It brings enormous advantages, but it does not tackle all the problems. That's why one need to endow the space $\mathcal{D}$ with one of the four topologies introduced by Skorokhod . We will define all of them, and explicit the relation between them. It turns out that some of them are comparable, being stronger or weaker if they contain or they are contained in another Skorokhod topology, but a general classification fails. Moreover, we will introduce a fifth topology, a primer up the the writer's knowledge, that is needed to prove one of our results.

We want now to equip the space $\mathcal{D}$ with a topology, such that: the space is Polish (so we can apply classical limit theorems on Polish spaces); the Borel σ-algebra is exactly the σ-algebra generated by all evaluation maps of functions in $\mathcal{D}$ (because the "law" of a process is precisely a measure on this σ-algebra).

We would like to repeat the considerations done for the space $\mathcal{C}$. However, any function in $\mathcal{D} \setminus \mathcal{C}$ does not satisfies condition (5.25) if we take as modulus of continuity the one defined in (5.24). Hence we need to modify the definition for càdlàg functions. For any $x \in \mathcal{D}([0,1])$ and any $\delta > 0$ we define

$$w'(x,\delta) := \inf_{t_1,..,t_r} \left\{ \max_{i\le r} w(x,[t_{i-1},t_i)) : \inf_{i\le r} |t_i - t_{i-1}| \ge \delta \right\} \tag{6.4}$$

where $w(x, I) = \sup_{s,t \in I} |x(t) - x(s)|$ and the points are such that $0 = t_0 < \cdots < t_r = 1$. The following Lemma gives the corresponding uniformity idea of (5.25) for càdlàg functions.

Lemma 6.1.1. *For any $x \in \mathcal{D}([0,1])$ and any $\varepsilon > 0$, there exist points $t_0, .., t_r$ such that*

$$0 = t_0 < t_1 < \cdots < t_r = 1 \tag{6.5}$$

$$w(x, [t_{i-1}, t_i)) < \varepsilon \quad \forall i = 1, ..., r \tag{6.6}$$

The proof contains no bright ideas. However, we decided to include it here, since it suggests that the extreme points $\{0, 1\}$ of the domain do not play any specific role. Hence the theorem can be adapted to a generic interval and, passing to the limit, to unbounded intervals. Therefore, we will make use of this results even when dealing with other domains than the unit interval.

Proof. Let $\bar{t}$ be the supremum of those $t \in [0, 1]$ for which we can decompose the interval $[0, \bar{t})$ into finitely many subintervals $[t_{i-1}, t_i)$ satisfying (6.6). We want to show that $\bar{t}$ must be equal to 1 (or, in general, to the right extreme point). In facts, it must be $\bar{t} > 0$ since any function $x \in \mathcal{D}([0, 1])$ is continuous from the right at 0. Since $x(\bar{t}^-)$ exists, also $[0, \bar{t})$ can be decomposed. Suppose $\bar{t} < 1$. Since $x \in \mathcal{D}([0, 1])$, we would have $x(\bar{t}) = x(\bar{t}^+)$, and hence there would exists $\eta > 0$ such that $w(x, [\bar{t}, \bar{t} + \eta)) < \varepsilon$. But this would be in contrast with our hypothesis that $\bar{t}$ is the supremum, hence it must be $\bar{t} = 1$. □

This Lemma is important not only to guarantee the existence of a set of points $\{t_0, ..., t_r\}$, but also for a sequence of consequences that will help us to better understand the natura of $\mathcal{D}$. It follows that there can be at most finitely many points t_i at which the jump $|x(t_i) - x(t_i^-)|$ exceeds a given positive number; therefore, any $x \in \mathcal{D}$ has at most countably many discontinuities. Moreover, it follows that x is bounded and it can be uniformly approximated by simple functions constant over intervals, so that it is Borel measurable.

Consider our modulus $w'(x, \delta)$ defined in (6.4). Similarly to (5.25) we can state, by Lemma 6.1.1, that, for every $x \in \mathcal{D}$,

$$\lim_{\delta \to 0} w'(x, \delta) = 0. \tag{6.7}$$

Note that the definition of $w'(x, \delta)$ makes sense even if x does not lie in $\mathcal{D}$. Just as (5.25) is necessary and sufficient for an arbitrary function x on $[0, 1]$ to lie in $\mathcal{C}$, (6.7) is necessary and sufficient for x to lie in $\mathcal{D}$. We want to compare $w(x, \delta)$ with $w'(x, \delta)$. The first, obvious relation is

$$w'(x, \delta) < w(x, 2\delta) \tag{6.8}$$

since $[0, 1)$ can be split into subintervals $[t_{i-1}, t_i)$ satisfying $\delta < t_i - t_{i-1} < 2\delta$. There is no general inequality in the opposite direction since, for any $x \in \mathcal{D} \setminus \mathcal{C}$, we have that $w(x, \delta)$ does not go to 0 for $\delta \to 0$, whereas $w'(x, \delta)$ do. However, even if $x \in \mathcal{D}$, the supremum

$\sup_{t\in(0,1]} |x(t) - x(t^-)|$ is achieved because only finitely many jumps can exceed a given positive number. Hence we have

$$w(x,\delta) \leq 2w'(x,\delta) + \sup_{t\in(0,1]} |x(t) - x(t^-)| \tag{6.9}$$

We don't focus on the proof of the previous equation, but simply observe that, when $x \in \mathcal{C}$, the relation above reduce to

$$w(x,\delta) \leq 2w'(x,\delta) \qquad \text{if } x \in \mathcal{C}. \tag{6.10}$$

Hence, the moduli $w(x,\delta)$ and $w'(x,\delta)$ are essentially the same for continuous functions x. The crucial role played by w' is to characterize compact subsets of $\mathcal{D}$, and we can derive an analogous of Theorem 5.3.2 for the space $\mathcal{D}$, as we will see with Theorem 6.3.1.

6.2 The J_1-Skorokhod topology

We are now ready to introduce a primer topology on the space $\mathcal{D}$. Note that the uniform topology can be selected also for the space $\mathcal{D}$, but it brings many problems naturally connected with its nature. When functions have discontinuities, we do not want to insist that corresponding jumps occur exactly at the same times in order for the functions to be close. But this is inevitable when dealing with the uniform topology. Two functions x and y are near one another with respect to the uniform topology if the graph of z can be carried onto the graph of y by a uniformly small perturbation of the ordinates, with the abscissas kept fixed. We want now to allow also a uniformly small deformation of the time scale. Pausing for a moment from this abstract mathematical derivation, the physical interpretation and motivation reflects the recognition that we cannot measure time with perfect accuracy any more than we can position.

Let Λ denote the class of continuous, strictly increasing mappings from $[0,1]$ into itself. Note that this implies that every mapping in this class keeps fixed the extreme points 0 and 1. Moreover, if $\lambda \in \Lambda$, then also $\lambda^{-1} \in \Lambda$, hence Λ can also be thought as the class of homeomorphism in $[0,1]$. To be precise we should denote Λ by $\Lambda_{[0,1]}$ and in general adopt the notation Λ_I. However, we assume that the domain is clear from the context, hence we always drop the subscript of the interval in the following.
We will refer to a function $\lambda \in \Lambda$ as a change of time, as it can be see as an acceleration/slow down of the usual time scale. Note that every change of time λ is continuous and increasing, starting at the origin and such that $\lambda(1) = 1$. Let id be the identity map, in this case on $[0,1]$, i.e. $\mathrm{id}(t) = t$ for $t \in [0,1]$. Then, the standard J_1 metric on $\mathcal{D} = \mathcal{D}([0,1],\mathbb{R})$ is given by the following definition.

Definition 10. For any $x, y \in \mathcal{D}([0,1])$, we denote

$$d_{J_1,[0,1]}(x,y) := \inf_{\lambda\in\Lambda_{[0,1]}} \max\left\{ \sup_{t\in[0,1]} |x\circ\lambda(t) - y(t)|\,,\ \sup_{t\in[0,1]} |\lambda(t) - t| \right\}, \tag{6.11}$$

This defines a distance on $\mathcal{D}([0,1])$, which we refer to as the J_1 or $J_1([0,1])$ distance.

Note the d_{J_1} is actually a metric. To see that it's finite just consider $\lambda = \mathrm{id}$. Then, it's clearly positive by definition of $||\cdot||$-norm. If $d_{J_1}(x, y) = 0$ then we must have that $||\lambda - \mathrm{id}|| = 0$ and $||x \circ \lambda - y|| = 0$. This means that it must happens that, for every $t \in [0, 1]$, $x(t) = y(t)$ or $x(t) = y(t^-)$, which in turns lead to the conclusion that $x = y$. The other direction is straightforward. Observe that

$$\sup_{t\in[0,1]} |\lambda t - t| = \sup_{t\in[0,1]} |t - \lambda^{-1}t| \tag{6.12}$$

$$\sup_{t\in[0,1]} |x(\lambda t) - y(t)| = \sup_{t\in[0,1]} |x(t) - y(\lambda^{-1}t)| \tag{6.13}$$

hence symmetry follows easily. We showed above that the inverse of any function in the class Λ still belong to Λ. The same holds for the composition of any $\lambda_1.\lambda_2 \in \Lambda$. The triangular inequality follows from

$$||\lambda_1\lambda_2 - \mathrm{id}|| \leq ||\lambda_1\lambda_2 - \lambda_2|| + ||\lambda_2 - \mathrm{id}|| = ||\lambda_1 - \mathrm{id}|| + ||\lambda_2 - \mathrm{id}|| \tag{6.14}$$

$$||x \circ \lambda_1\lambda_2 - y|| \leq ||x \circ \lambda_1\lambda_2 - z \circ \lambda_2|| + ||z \circ \lambda_2 - y|| = ||x \circ \lambda_1 - z|| + ||z \circ \lambda_2 - y|| \tag{6.15}$$

This metric d_{J_1} defines the J_1-Skorokhod topology, which renders $\mathcal{D}$ a Polish space.

6.2.1 Characterization of J_1 Convergence

This metric induces a topology and a notion of limit in $\mathcal{D}([0, 1])$. The following proposition characterize the mood of convergence for elements x_n of $\mathcal{D}$ to a limit x in the J_1-Skorokhod topology.

Proposition 6.2.1. *A sequence $(x_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{D}$ converges to $x \in \mathcal{D}$ w.r.t. the J_1-Skorokhod topology, and we write*

$$f_n \to f \qquad \text{in } (\mathcal{D}([0, 1]), J_1), \tag{6.16}$$

as $n \to \infty$, if and only if there exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \subset \Lambda$ of increasing homeomorphisms such that

$$\lim_{n\to\infty} \sup_{t\in[0,1]} |x_n \circ \lambda_n(t) - x(t)| = 0, \tag{6.17}$$

$$\lim_{n\to\infty} \sup_{t\in[0,1]} |\lambda_n(t) - t| = 0. \tag{6.18}$$

Note that J_1-Skorokhod convergence of the proposition above implies that $x_n(t) \to x(t)$ holds for continuity points t of x and hence for all but countably many t. We can say more. The J_1-Skorokhod topology restricted to the space $\mathcal{C}$ coincides with the uniform topology there. In facts we have that

$$|x_n(t) - x(t)| \leq |x_n(t) - x(\lambda_n t)| + |x(\lambda_n t) - x(t)| \tag{6.19}$$

thus it follows that, if x is uniformly continuous on all $t \in [0, 1]$, the J_1-Skorokhod convergence implies uniform convergence.

Proposition 6.2.2. *The J_1-Skorokhod topology is weaker than the (local) uniform topology. Moreover, if x is a continuous function, a sequence $(x_n)_{n\in\mathbb{N}}$ converges to x with respect to the J_1-Skorokhod topology if and only if it converges to x (locally) uniformly.*

The local uniform topology coincides with the uniform topology in any bounded domain. We are dealing with functions defined in $[0, 1]$, hence the adjective local may appear redundant. However, every consideration made here can be extended to unbounded domain, as we will see later on, and that's why we explicitly include the local information at this point.
Note that uniform convergence in $\mathcal{D}$ implies that there exists a number $\bar{n}$ such that the discontinuity points of f_n coincide with the discontinuity points of f for every $n > \bar{n}$. This requirement is hard to satisfy. However, the J_1-Skorokhod topology relax this constraint, allowing the jumps in a limit function to be unmatched in the converging functions.

Remark 3. It is a known and easy-to-prove fact that, if $I_0 \subseteq [0, 1]$ is an interval at positive distance from the discontinuities of f, and $f_n \to f$ in $(\mathcal{D}([0, 1]), J_1)$, then $\sup_{t\in I_0} |f_n(t) - f(t)| \to 0$.

6.3 Some properties

6.3.1 Completeness of $(\mathcal{D}, J_1)$

The space $\mathcal{D}$ is not complete under the metric d_{J_1} defined in (6.11). This is more a technical issue. The space $\mathcal{D}$ is Polish under the J_1-Skorokhod topology, hence there exists another metric $d^\star_{J_1}$, equivalent to d_{J_1}, under which the space $\mathcal{D}$ is complete. Equivalence between the two metrics means that they both generates the same topology. The additional idea of $d^\star_{J_1}$ is to require not only the homeomorphism λ to be close to the identity as in (6.12), but also the slope of λ to be close to 1, thus avoiding any growth close to the coordinate directions. We define

$$||\lambda||^\star = \sup_{s<t} \left\| \log \frac{\lambda t - \lambda s}{t - s} \right\|. \tag{6.20}$$

If $||\lambda||^\star$ is bounded, we have that the actual slope of λ is bounded away from 0 and $+\infty$. If λ is such that $\lambda(0) = 0$, $\lambda(1) = 1$, and $||\lambda||^\star$ is finite, then the slope of the chords of λ are bounded away from 0 and infinity and therefore it is both continuous and strictly increasing and hence is a member of Λ. The converse is not true, however, since there exists functions in Λ that do not satisfy boundedness of (6.20). We want to define the equivalent metric $d^\star_{J_1}$ by replacing the quantity in (6.12) with $||\lambda||^\star$. In other words

$$d^\star_{J_1}(x, y) = \inf_{\lambda\in\lambda} \{||x \circ \lambda - y|| \vee ||\lambda||^\star\}. \tag{6.21}$$

We showed that the metric d_{J_1} is incomplete, but, since the topology is topologically complete, there exists a topologically equivalent metric $d^\star_{J_1}$ that is complete. Completeness facilitates characterizing the compact sets. From now on, we drop the superscript $\star$ in the notation. We are interested in topology, hence the use of the metric (6.11) or

(6.21) is interchangeable since they are topologically equivalent. We simply write d_{J_1} and assume that the space is complete.
Hence there is equivalence between the fact that $d_{J_1}(x_n, x) \to 0$ and the existence of a sequence $(\lambda_n)_{n\in\mathbb{N}} \subset \Lambda$ such that $||\lambda_n||^\star$ is bounded for every n and $||x_n \circ \lambda_n - x|| \to 0$.

The space $(\mathcal{D}, J_1)$ is separable with respect to the J_1-Skorokhod topology; for example the set consisting of the piecewise constant, rational valued, functions with only finitely many discontinuities, all at rational time points of the domain, is countable and dense. Thus the space $(\mathcal{D}, J_1)$ is Polish.

6.3.2 Compactness in $\mathcal{D}$

The following theorem helps us to characterize relative compactness in $\mathcal{D}$.

Theorem 6.3.1. *(Ascoli-Arzelá in $\mathcal{D}$) The set $A \subset \mathcal{D}([0,1])$ is relatively compact for the J_1-Skorokhod topology if and only if the two following conditions are satisfied:*

$$\sup_{x\in A} \sup_t |x(t)| \leq \infty \tag{6.22}$$

$$\lim_{\delta\to 0} \sup_{x\in A} w'(x,\delta) = 0 \tag{6.23}$$

The important part of the theorem is the sufficiency of the two condition to have relative compactness, an essential requirement when dealing with weak convergence.

6.3.3 Finite-dimensional sets

Finite-dimensional sets play in $\mathcal{D}$ the same role they do in $\mathcal{C}$. Recall that we want to use convergence of the finite-dimensional sets to determine convergence of the entire path. When dealing with processes, we wish to use convergence of finite-dimensional distribution as a key ingredient for proving convergence of processes. For the moment, we stick on finite-dimensional sets without involving any stochasticity.
For $k \in \mathbb{N}$, and $0 \leq t_1 \leq \cdots \leq t_k \leq 1$, we can define the natural projection as usual:

$$\begin{aligned} &\pi_{t_1,..,t_k} : \mathcal{D} \mapsto \mathbb{R}^k \\ &\pi_{t_1,..,t_k}(x) = (x(t_1), ..., x(t_k)) \end{aligned} \tag{6.24}$$

The finite-dimensional sets are defined as sets of the form $\pi^{-1}_{t_1,..,t_k} H$, for $H \subset \mathbb{R}^k$.

We want to analyze measurability and continuity of the projection (6.24). We can reduce to a single time point t, since a mapping into $\mathbb{R}^k$ is measurable if each component of the mapping is measurable. Similarly we investigate continuity.
As first, notice that π_0 and π_1 are continuous. If facts, every $y \in \mathcal{D}$ is ε-close to $x \in \mathcal{D}$ with respect to the J_1-Skorokhod topology if and only if the exists $\lambda \in \Lambda$ such that $||\lambda||^\star < \varepsilon$ and $||x\circ\lambda - y|| < \varepsilon$. But every $\lambda \in \Lambda$ fixes 0 and 1, hence $|\pi_0(x) - \pi_0(y)| = |x(0) - y(0)| < \varepsilon$ and the same holds for π_1.
Consider now a time point $t \in (0,1)$. If points x_n converge to x in the J_1-Skorokhod topology and x is continuous at t, then $x_n(t) \to x(t)$, as shown in (6.19). Suppose now

that x is discontinuous at t. Consider the situation in which $x_n \equiv x \circ \lambda_n$, where λ_n is the element of Λ that carries t to $t - 1/n$ and is linear on $[0, t)$ and $(t, 1]$. Then we have convergence of x_n to x with respect to the J_1-Skorokhod topology, but $x_n(t) \nrightarrow x(t)$. Therefore, we can conclude that π_t is continuous at x if and only if x is continuous at t.

As for measurability, we consider for any fixed t the function $h_\varepsilon(x) = \varepsilon^{-1} \int_t^{t+\varepsilon} x(s)ds$. We have that h_ε is continuous in the J_1-Skorokhod topology, since $\lim_n h_\varepsilon(x_n) = h_\varepsilon(x)$ whenever $x_n \to x$ with respect to the J_1-Skorokhod topology. In facts, every function in $\mathcal{D}$ is bounded, and $x_n(s) \to x(s)$ for continuity points of x and hence for points s outside a set o Lebesgue measure 0. Now, since x is right continuous we have that $h_{m^{-1}}(x) \to x(t) = \pi_t(x)$ for each x as $m \to \infty$. Therefore each π_t is measurable.
We can say more on the finite-dimensional sets.

Proposition 6.3.1. *Suppose that $T \subset [0, 1]$ is a dense subset of time points containing 1. Let $\sigma[\pi_t : t \in T]$ be the σ-algebra generated by real functions π_t for $t \in T$; and let $p[\pi_t : t \in T]$ be the class of sets $\pi^{-1}_{t_1,..,t_k} H$, where k is arbitrary, $t_i \in T$, $H \subset \mathbb{R}^k$.*
Then $\sigma[\pi_t : t \in T] = \mathcal{D}$ and $p[\pi_t : t \in T]$ is a separating class.

6.3.4 Convergence of the sum

Note that linear combinations $ax + by$ of function x, y in $\mathcal{D}$ are again càdlàg , hence $\mathcal{D}$ is a vector space. However, $\mathcal{D}$ is not a topological vector space (and thus not a Banach space) because those structures require addition to be continuous.
Indeed we may have that $x_n \to x$ and $y_n \to y$ with respect to the J_1-Skorokhod topology, and yet $x_n + y_n$ not converging to $x + y$ with respect to the same topology.

In applications of the continuous-mapping approach to establish stochastic-process limits, we will often want to add or subtract two functions. Thus it is very important that addition can be made to preserve convergence. It turns out that addition on $\mathcal{D} \times \mathcal{D}$ is measurable and it is continuous at limits in a large subset of $\mathcal{D} \times \mathcal{D}$.

We need to require something. For example continuity of y is sufficient to guarantee convergence of the sum. However, this is usually a strong requirement, and we usually deal with proper discontinuous trajectories. The next proposition ensures convergence of the sum, conditional of the existence of a common time sequence for convergence.

Proposition 6.3.2. *Let $x_n \to x$ and $y_n \to y$ with respect to the J_1-Skorokhod topology. Assume that for each $t > 0$ there exists a sequence $(t_n)_{n\in\mathbb{N}}$ converging to t, such that $\Delta x_n(t_n) \to \Delta x(t)$ and $\Delta y_n(t_n) \to \Delta y(t)$. Then $x_n + y_n \to x + y$ with respect to J_1.*

Note that if t is a continuity point both for x and for y, i.e. $\Delta x(t) = 0$ and $\Delta y(t) = 0$, then it suffices to take $t_n \equiv t$ for n great enough to obtain the desired sequence $(t_n)_{n\in\mathbb{N}}$. The problem of finding a sequence is engaging when t is a discontinuity point. Suppose that t is a discontinuity point for x. From the definition of J_1-Skorokhod convergence there must exists a sequence $(t_n)_{n\in\mathbb{N}}$ so that $t_n \to t$ and $\Delta x_n(t_n) \to \Delta x(t)$. If t is a continuity point for y, then convergence $\Delta y_n(t_n) \to \Delta y(t)$ holds naturally, otherwise it is not guaranteed. This fact suggests that the previous Proposition can be reformulated

in terms of discontinuity points. This is exactly the object of the following Theorem, which provide an easier to check condition to ensure convergence of the sum.

Theorem 6.3.2. *Let $x, (x_n)_{n\in\mathbb{N}}$ and $y, (y_n)_{n\in\mathbb{N}}$ be such that $x_n \to x$ in J_1 and $y_n \to y$ in J_1. If $Disc[x] \cap Disc[y] = \emptyset$ then $x_n + y_n \to x + y$ in with respect to the J_1-Skorokhod topology.*

We point out that the previous theorem can be extended to the M-Skorokhod topologies (see Section 6.6). We extended it also to the J_2-Skorokhod topology (See Section 6.7). Up to the author's knowledge, this is the first time that an explicit proof of the result appears in the literature.

6.4 Extension of $\mathcal{D}$

We want now to modify the space $\mathcal{D}([0,1],\mathbb{R})$ in two ways. Firstly we want to allow the domain to be any interval $I \subseteq \mathbb{R}$ rather than just the uni interval $[0,1]$. Secondly we want to extend the range of the functions from $\mathbb{R}$ to $\mathbb{R}^d$ for any $d \geq 1$. Finally we want to describe a slight modification of the trajectory that we will adopt in this thesis.

6.4.1 Domain extension

Note that in all our definitions, the unit interval $[0,1]$ can be replaced by any bounded interval I with no particular effort.
For example we can take the interval $I = [0,t]$ and everything works exactly the same. Then, if we consider unbounded intervals like $[0,+\infty)$, it is natural to characterize convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{D}([0,\infty)) = \mathcal{D}^+$ through the convergence of the restrictions of x_n to the subintervals $[0,t]$ in the space $\mathcal{D}([0,t])$.
However, we could encounter some troubles. For example, if we consider the function $x = \mathbb{1}_{[t,\infty)}$ as the limit function of the sequence $(x_n)_{n\in\mathbb{N}}$ where $x_n = \mathbb{1}_{[t_n,\infty)}$ with $t_n = t + 1/n$, we would like to state that $x_n \to x$ in $\mathcal{D}^+$. However, for every $n \in \mathbb{N}$ the restriction of x_n to $[0,t]$ is the zero function, whereas the restriction of x is not. Nevertheless, the restriction of $\mathbb{1}_{[t+1/n,\infty)}$ converges to $\mathbb{1}_{[t,\infty)}$ in $\mathcal{D}([0,s])$ for any $s \neq t$.
Thus, we say that a sequence $(x_n)_{n\in\mathbb{N}}$ converges to x in $\mathcal{D}^+$ if the correspondent restrictions to $[0,t]$ converge in $\mathcal{D}([0,t])$ for all $t > 0$ that are continuity points of x.

Definition 11. If $I \subset \mathbb{R}$ is a half-line, say $I = [a,+\infty)$, and $(x_n)_{n\in\mathbb{N}}$, x are functions of $\mathcal{D}(I)$, we say that $x_n \to x$ in $(\mathcal{D}(x), J_1)$, for $n \to \infty$, if, for all $T > a$ such that x is continuous at T,

$$x_n \to x \qquad \text{in } (\mathcal{D}([a,T]), J_1). \tag{6.25}$$

The analogous definition is given for $I = (a,+\infty)$ or $I = (-\infty,a]$, etc. If $I = \mathbb{R}$, we say that $f_n \to f$ in $(\mathcal{D}, J_1)$ if, for all $T > 0$ such that f is continuous at T and $-T$,

$$f_n \to f \qquad \text{in } (\mathcal{D}([-T,T]), J_1). \tag{6.26}$$

The above definition defines a J_1 topology on $\mathcal{D}(I)$, in all cases where I is a half-line or the entire $\mathbb{R}$. It is easy to write a metric that generates the $J_1(I)$ topology (see [Whi80, Section 2]).
For example we can define directly a metric d^∞ on $\mathcal{D}([0,\infty),\mathbb{R})$ by

$$d^\infty(x,y) := \int_0^\infty e^{-t}\left[d^{(t)}(x,y)\wedge 1\right]dt \tag{6.27}$$

Remark 4. The definition of limit in $(\mathcal{D}([a,+\infty)),J_i)$ $(i=1,2)$ amounts to checking that $f_n \to f$ in $(\mathcal{D}([a,T]),J_i)$, for all $T>a$ such that f is continuous at T, see (6.25). It is easy to see that this is tantamount to checking that $f_n \to f$ in $(\mathcal{D}([a,T)),J_i)$, for all $T>a$ such that f is continuous at T. In the remainder (see for example Section 9.3.3) we will liberally switch between the two conditions, as is more convenient.

6.4.2 Trajectories in $\mathbb{R}^d$

In order to go from dimension 1 to d, the ideas is, as always, to define a metric on the components and then upscale it to all the dimensions. We have two principal ways to do it. The first way is to replace the norm $|\cdot|$ on $\mathbb{R}$ with a corresponding norm on $\mathbb{R}^d$ such as the maximum norm:

$$||x|| := \max_{i\in\{1,..,d\}} |x_i| \tag{6.28}$$

In this way, we obtained the standard (strong) topology on $\mathcal{D}([0,1],\mathbb{R}^d)$. Another possibility is to use the product topology on $\mathcal{D}^d$, under which we have convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ to x if every coordinate variables is converging, i.e. $x_n^1 \to x^i$ for all $i\in\{1,...,d\}$. The product topology on $\mathcal{D}^d$ is induced by the metric

$$d_p(x,y) := \sum_{i=1}^d d(x^i,y^i) \tag{6.29}$$

where d is the metric on $\mathcal{D}$, for example $d=d_{J_1}$. Since convergence in the strong topology implies convergence in the product topology, we also call the product topology the weak topology. As abstract spaces, one may of course identify $\mathcal{D}(\mathbb{R}^d)$ with the cartesian product $\mathcal{D}(\mathbb{R})^d$. But as topological spaces, for the Skorokhod topology, the topology of $\mathcal{D}(\mathbb{R}^d)$ is strictly finer than the product topology of $\mathcal{D}(\mathbb{R})^d$.

6.4.3 Càglàd trajectories

If I is an interval or a half-line intersecting $(-\infty,0)$, or $I=\mathbb{R}$, we consider a less customary function space: $\mathcal{D}(I)$ is the space of all functions $x: I \longrightarrow \mathbb{R}$ such that $s\mapsto x(s)$ is càdlàg for $s\geq 0$ and $s\mapsto x(-s)$ is càdlàg for $s\geq 0$. In other words, the restriction of x to $I\cap(-\infty,0]$ is càglàd). Notice that this implies that x is continuous at 0.

Remark 5. Although there are good reasons of convenience for using spaces of functions that are càdlàg on $\mathbb{R}^+$ and càglàd on $\mathbb{R}^-$ (see Section 9.3), some readers may find this choice odd and prefer to always work with càdlàg functions, even for domains I intersecting $(-\infty, 0)$. Clearly, to any $x \in \mathcal{D}(I)$ as defined earlier there corresponds a unique càdlàg version

$$x_{\text{cadlag}}(t) := \lim_{s \to t^+} x(s). \tag{6.30}$$

For any bounded I, it is easy to see that $d_{J_1,I}(x_{\text{cadlag}}, y_{\text{cadlag}}) = d_{J_1,I}(x, y)$.

Remark 6. In this paper the only two cases in which we work with I intersecting $(-\infty, 0)$ are $I = [-M, M]$ and $I = \mathbb{R}$. In both cases we only deal with functions f such that $x(0) = 0$. It is easy to see that, under such additional condition, it is no loss of generality to require that the homeomorphism λ fixes 0, i.e., $\lambda(0) = 0$. This makes it clear that, in such cases, $x_n \to x$ in $(\mathcal{D}([-M, M]), J_1)$ if and only if both $x_n \to x$ and $x_n(-\cdot) \to x(-\cdot)$ in $(\mathcal{D}([0, M]), J_1)$. Here $x(-\cdot)$ denotes the function $t \mapsto x(-t)$.

6.5 Processes on $\mathcal{D}$

Throughout this thesis we will focus on convergence of processes with càdlàg trajectories. The starting point is to define random functions on $\mathcal{D}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, a random element of $\mathcal{D}$ is a $\mathcal{F}$-measurable function X mapping Ω into $\mathcal{D}$.
Note that $X_t(\omega) = \pi_t(X(\omega))$ defines a random variable, that can be interpreted as the position of the process at time t.
We want to observe the converse relation. We already saw how the projection map (6.24) generates the finite-dimensional sets of the form $\pi^{-1}_{t_1,..,t_k} H$. Given a probability measure P on $\mathcal{D}$, the finite-dimensional distributions are defined as the pull back measures $P \circ \pi^{-1}_{t_1,..,t_k}$. We have seen in 6.3.3 that the projections are not everywhere continuous on $\mathcal{D}$, hence weak convergence $P_n \xrightarrow{w} P$ does not always imply that $P_n \circ \pi^{-1}_{t_1,..,t_k} \xrightarrow{w} P \circ \pi^{-1}_{t_1,..,t_k}$ weakly.

We want now to consider $\mathbb{R}$ valued càdlàg processes. Let X be such a process, defined on a triple $(\Omega, \mathcal{F}, \mathbb{P})$. Then it may be considered as a random variable taking its values in the Polish space $\mathcal{D}$, supposedly equipped with the J_1-Skorokhod topology.

Let us define

$$J(X) := \{t : \mathbb{P}(\Delta X_t \neq 0) > 0\} \tag{6.31}$$

$$U(X) := \{u > 0 : \mathbb{P}(|\Delta X_t| = u \text{ for some } t) > 0\} \tag{6.32}$$

It is possible to show that the sets $J(X)$ and $U(X)$ are at most countable.
With no surprise, we can relate the notion of convergence in distribution of stochastic processes to the weak convergence of their law.

Definition 12. Consider now a sequence $(X_n)_{n \in \mathbb{N}}$ of $\mathbb{R}^d$-valued càdlàg processes, each X_n defined on some space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$. We say that X_n converge in distribution to X, and denote it by

$$X_n \xrightarrow{d} X \tag{6.33}$$

if the law of X_n converge weakly to the law of X, ad we denote it by

$$\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X), \tag{6.34}$$

where the law of a stochastic process is the distribution induced on the arrival metric space by the probability measure on the domain, i.e. $\mathcal{L}(X) = \mathbb{P} \circ X^{-1}$.

As for processes with continuous trajectories, there are different approach to prove convergence in distribution (6.33) of a sequence of processes with trajectories in $\mathcal{D}$. The continuous mapping approach is applicable also in the space $\mathcal{D}$, and Theorem 5.4.2 applies also in this case. It reads

Theorem 6.5.1. *[Countinuous Mapping] Let $(X_n)_{n\in\mathbb{N}}, X$ be stochastic processes with values in $\mathcal{D}^d$, and suppose that $X_n \xrightarrow{d} X$ w.r.t. the J_1-Skorokhod topology. Let the function $h : (\mathcal{D}^d, J_1) \mapsto (\mathcal{D}^k, J_1)$ be measurable, and denote by $Disc[h]$ the set of its discontinuity points. If $P(X \in Disc[h]) = 0$, then $h(X_n) \xrightarrow{d} h(X)$.*

6.5.1 Compactness approach in $\mathcal{D}$

We wish to relate now convergence in distribution of processes with the convergence of its finite dimensional distributions. In the space $\mathcal{C}$ we showed that weak convergence of the sequence of processes imply convergence of the finite-dimensional distribution. The converse holds only if we provide tightness of for the sequence of processes, i.e. if the sequence of the laws of the processes is tight. We wish to apply the same compactness approach also in the space $\mathcal{D}$. However, things are a bit different, due to the lack of continuity of the trajectories.

Proposition 6.5.1. *If $X_n \xrightarrow{d} X$, then we have convergence of the finite-dimensional distributions only ad continuity points, i.e. for any $k \in \mathbb{N}$ and $t_1, .., t_k \in [0,1] \setminus J(X)$ we have*

$$(X_n(t_1), ..., X_n(t_k)) \xrightarrow{d} (X(t_1), ..., X_n(t)) \tag{6.35}$$

We want now to consider the question of proving that $X_n \xrightarrow{d} X$. The most common method proceeds through the following procedure:

1. Prove that $(X_n)_{n\in\mathbb{N}}$ is tight.
2. Prove that $\mathcal{L}(X)$ is the only possible limit for the sequence $\mathcal{L}(X_n)_{n\in\mathbb{N}}$

Note that this two points are enough to ensure weak convergence of the sequence of processes $(X_n)_{n\in\mathbb{N}}$. Condition 1. is equivalent to require that the sequence $\mathcal{L}(X_n)_{n\in\mathbb{N}}$ is relatively compact. The usual way to satisfy condition 2. is by mean of the finite-dimensional distributions.

Lemma 6.5.1. *Let T be a dense subset of $[0,1]$, and X, Y be two càdlàg processes satisfying $(X_{t_1}, ..., X_{t_k}) \overset{d}{=} (Y_{t_1}, ..., Y_{t_k})$ for all $k \in \mathbb{N}$, $t_i \in T$. Then $\mathcal{L}(X) = \mathcal{L}(Y)$.*

Hence condition 2. can be replaced by proving that we have convergence of the finite-dimensional distributions at time points in a dense subset $T \subset [0,1]$. The idea is to relate this dense subset to the set of points where the process is continuous. Let

$$T_X := \{t > 0 : \mathbb{P}(t \in Disc[X]) = 0\} \cup \{1\}. \tag{6.36}$$

Note that the set T_X contains 0 and 1, and its complement in $[0,1]$ is at most countable.

Theorem 6.5.2. *There is convergence in distribution $X_n \xrightarrow{d} X$ in $\mathcal{D}$ with respect to J_1, if*

$$(X_n(t_1), ..., X_n(t_k)) \xrightarrow{d} (X(t_1), ..., X(t_k)) \tag{6.37}$$

holds for all $t_i \in T_X$ and $(X_n)_{n\in\mathbb{N}}$ is tight with respect to J_1.

This approach is similar to the one in $\mathcal{C}$, except that we want to require convergence of the finite-dimensional distributions only for time points t that are almost surely continuity points of the limit process X. Also the proof of the previous theorem follows similarly. The idea is to show that, whenever a subsequence $(X_{n_k})_{k\in\mathbb{N}}$ converges in distribution to some Y, then Y must coincide with X. By assumption of the theorem we have that $(X_{n_k}(t_1), ..., X_{n_k}(t_m)) \xrightarrow{d} (X(t_1), ..., X(t_m))$ whenever $t_1, .., t_m$ lie in T_X; also by Proposition 6.5.1 er have that $(X_{n_k}(t_1), ..., X_{n_k}(t_m)) \xrightarrow{d} (Y(t_1), ..., Y(t_m))$ whenever $t_1, .., t_m$ lie in T_Y. Hence $(X(t_1), ..., X(t_m)) = (Y(t_1), ..., Y(t_m))$ for any $t_1, .., t_m \in T_X \cap T_Y$. But $T_X \cap T_Y$ contains 0 and 1 and its complement is at most countable, therefore $Y \equiv X$.

To conclude this section, one should explicit some criteria to show tightness. Since we don't make explicit use of any of them in this thesis, we completely skip this part. The reader can find good references in [Bil13].

6.6 The four Skorokhod topologies

In this section we introduce three additional notions of distance/topology that turn out to be crucial in the following. A complete treatment of these topologies can be found, e.g., in [Whi02, Sections 3.3. & 11.5].

We already stressed the importance of the space $\mathcal{D}$ to represent trajectory with jump discontinuities. We endowed the space $\mathcal{D}$ with the J_1-Skorokhod topology, that allows us to consider close two functions whose corresponding jumps occurs at close times, even if no exactly the same. We introduced a continuous perturbation of the time, uniformly close to the identity. This solution solves many problems, but there are other cases that require something different. To be more clear, consider the function $x = \mathbb{1}_{[1/2,1]}$. This trajectory has a jump of size 1 at time $t = 1/2$, hence $x \in \mathcal{D}$. We want to consider different sequences of functions in $\mathcal{D}$ and investigate over its convergence to x. The sequence $(x_n)_{n\in\mathbb{N}}$ given by

$$x_n(t) = \begin{cases} 0 & t \in \left[0, \frac{1}{2}\right) \\ 1 + \frac{1}{n} & t \in \left[\frac{1}{2}, 1\right] \end{cases} \tag{6.38}$$

converge uniformly to x, since $\lim_{n\to\infty} \sup_{t\in[0,1]} |x_n(t) - x(t)| = 0$. In this case the uniform metric works fine, and we do not even need a time change.
Suppose now that the sequence of functions $(x_n)_{n\in\mathbb{N}}$ are now defined as follows

$$x_n(t) = \begin{cases} 0 & t \in \left[0, \frac{1}{2} - \frac{1}{n}\right) \\ 1 + \frac{1}{n} & t \in \left[\frac{1}{2} - \frac{1}{n}, 1\right] \end{cases} \tag{6.39}$$

The novelty here is that the jump of x_n occurs at time $t = 1/2 - 1/n$, which is different for any $n \in \mathbb{N}$. However, Using the J_1-Skorokhod topology and the notion of convergence expressed in Proposition 6.2.1 we can affirm that $x_n \to x$ with respect to the J_1-Skorokhod topology.

We want now to illustrate other similar cases in which this type of convergence fails, even if our intuition suggests that the trajectories are close. In the next example we consider trajectories that coincide with $x = \mathbb{1}_{[1/2,1]}$ except in a small left neighborhood of the jump time $t = 1/2$. Consider the sequence of functions $(x_n)_{n\in\mathbb{N}}$ are now defined as follows

$$x_n(t) = \begin{cases} 0 & t \in \left[0, \frac{1}{2} - \frac{1}{n}\right) \\ n \cdot \left(t - \frac{1}{2} + \frac{1}{n}\right) & t \in \left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2}\right) \\ 1 & t \in \left[\frac{1}{2}, 1\right] \end{cases} \tag{6.40}$$

Note that x_n is continuous for every $n \in \mathbb{N}$. It's evident that $||x_n - x|| = \sup_t |x_n(t) - x(t)| = 1$. A bit less straightforward is that also $d_{J_1}(x_n, x) \nrightarrow 0$ as $n \to \infty$. In facts, for any $n \in \mathbb{N}$, we have that any $\lambda_n \in \Lambda$ is such that there exists t_{λ_n} such that $x_n(\lambda_n(t_{\lambda_n})) = 1/2$, whereas $x(t_{\lambda_n})$ equals 0 or 1. Thus $\sup_t |x_n \circ \lambda_n(t) - x(t)| \geq 1/2$ for every λ_n, allowing us to conclude that $d_{J_1}(x_n, x) \geq 1/2$ for every $n \in \mathbb{N}$. The problem here is that x_n assumes values that are far from the values of x. This occurs in a small time interval shrinking to 0 as $n \to \infty$, leading to pointwise convergence $x_n(t) \to x(t)$ for every t. This example exhibits the need of a new topology with respect to which the sequence $(x_n)_{n\in\mathbb{N}}$ defined in (6.40) converges to the limit function $x = \mathbb{1}_{[1/2,1]}$. The M_1-Skorokhod topology will solve this problem, by considering two functions close if their respective complete graph is close. We want now to show another example, which will require the introduction of a third topology in order to converge.

Consider the sequence of functions $(x_n)_{n\in\mathbb{N}}$ are now defined as follows

$$x_n(t) = \begin{cases} 0 & t \in \left[0, \frac{1}{2} - \frac{2}{n}\right) \\ 1 & t \in \left[\frac{1}{2} - \frac{2}{n}, \frac{1}{2} - \frac{1}{n}\right) \\ 0 & t \in \left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2}\right) \\ 1 & t \in \left[\frac{1}{2}, 1\right] \end{cases} \tag{6.41}$$

The function x_n defined in (6.41) consists of multiple jumps going back and forth from 0 to 1 near the jump point $t = 1/2$. Contrary to (6.41) this time the values taken by x_n are close to those of x, actually they are the same, and again we have pointwise convergence $x_n(t) \to x(t)$ for every t. However, we still have that $||x - x_n|| = 1$ for every $n \in \mathbb{N}$, and the J_1-Skorokhod distance is again ineffective since $d_{J_1}(x_n, x) = 1$ for every $n \in \mathbb{N}$. This motivates the introduction of a third Skorokhod topology, where we enlarge the class Λ of homeomorphism to include also bijections of the unitary interval that are close to the identity. This is the goal of the J_2-Skorokhod topology, as we will see later on.

Note that both topologies J_2 and M_1 are weaker than J_1. Convergence with respect to the J_1-Skorokhod topology implies convergence with respect to both J_2 and M_1. A direct comparison between J_2 and M_1 fails. As an example, notice that the sequence in (6.40) is not converging with respect to the J_2-Skorokhod topology, and conversely the sequence in (6.41) is not converging with respect to the M_1-Skorokhod topology.

To conclude this tracking shot of examples, we consider another sequence of functions $(x_n)_{n\in\mathbb{N}}$ defined as

$$x_n(t) = \begin{cases} 0 & t \in \left[0, \frac{1}{2} - \frac{3}{n}\right) \\ n \cdot \left(t - \frac{1}{2} + \frac{3}{n}\right) & t \in \left[\frac{1}{2} - \frac{3}{n}, \frac{1}{2} - \frac{2}{n}\right) \\ n \cdot \left(\frac{1}{2} - \frac{2}{n} - t\right) & t \in \left[\frac{1}{2} - \frac{2}{n}, \frac{1}{2} - \frac{1}{n}\right) \\ n \cdot \left(t - \frac{1}{2} + \frac{1}{n}\right) & t \in \left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2}\right) \\ 1 & t \in \left[\frac{1}{2}, 1\right] \end{cases} \tag{6.42}$$

The function x_n defined in (6.42) is constantly 0 until $t = 1/2 - 3/n$, then it is piecewise linear touching the value 1 at $t = 1/2 - 2/n$, the coming back to 0 at $t = 1/2 - 1/n$ and growing to 1 at $t = 1/2$, where it stays. Neither J_2-topology nor M_1-topology can accomodate convergence of (6.42). To accomplish this goal we need to introduce a fourth topology on $\mathcal{D}$, the M_2-Skorokhod topology, which is weaker both than J_2 and M_1. This examples serve as a motivation to the introduction of different topologies on the space $\mathcal{D}$. When proving convergence, one wants to select the strongest topology with respect to

which convergence holds. Through this example we hope to give familiarity to concept we are rigorously introducing.

We want now to give the proper definitions of all the different topologies we just mentioned. To start with, we need to definition of the complete graph of a function.

Definition 13. The complete graph of a function $x \in \mathcal{D}([0,1],\mathbb{R})$ is the following set

$$\Gamma_x := \{(z,t) \in \mathbb{R} \times [0,1] \,:\, z = \alpha x(t^-) + (1-\alpha)x(t) \text{ for some } \alpha \in [0,1]\} \tag{6.43}$$

In other words, the complete graph of a function x is given by the collection of all convex combinations of $x(t^-)$ and $x(t)$ for any t in the domain of x. Note that if x is continuous at t, then $x(t^-) = x(t)$. Hence the complete graph of a function x consists of all the points of the original graph, together with the vertical segments connecting the jumps. The completed graph is thus a connected subset of the plane $\mathbb{R}^2$ containing the graph and the line segments joining $(t, x(t^-))$ and $(t, x(t))$ for all discontinuity points t. The second step concerns giving a parametric representation of the complete graphs. As first we need to establish an order between points. Given two points (t_1, z_1) and (t_2, z_2) belonging to Γ_x, we say that $(t_1, z_1) \leq (t_2, z_2)$ if either (i) $t_1 < t_2$ or (ii) $t_1 = t_2$ and $|x(t_1^-) - z_1| \leq |x(t_2^-) - z_2|$. It means that we are ordering the points with the same logic if we imagine to draw the complete graph from left to right with a continuous line. With this order in mind, we can build parametric representations of Γ_x. The set $\Pi(x)$ of all parametric representations of Γ_x consists of the functions $t \mapsto (s(t), z(t))$ from $[0,1]$ to Γ_x that are nondecreasing with respect to the order we defined above on the complete graph Γ_x.
The M_1 metric is defined as follows

$$d_{M_1}(x_1, x_2) = \inf_{\substack{(s_1,z_1)\in\Pi(x_1)\\(s_2,z_2)\in\Pi(x_2)}} \{||s_1 - s_2|| \vee ||z_1 - z_2||\} \tag{6.44}$$

The metric d_{M_1} is complete and generate the M_1-Skorokhod topology. Note that, as soon as we take a limit continuous limit function x, then $d_{M_1}(x_n, x) \to 0$ is completely equivalente to $||x_n - x|| \to 0$, similarly to what happen in J_1. Consider the example (6.40) above. We want to descrive explicit parametric representations of x and x_n in order to prove convergence with respect to the M_1-Skorokhod topology. We build $t \mapsto z_n(t)$ and $t \mapsto z(t)$ in such a way that $z_n(t) \equiv z(t)$. (Note that this is possible since every parametric representation of Γ_{x_n} and Γ_x in increasing in the spatial component and is such that $z_n(0) = z(0) = 0$ and $z_n(1) = z(1) = 1$.) As for the time representation we select, for example, the following one, for any fixed $0 < a < b < 1$:

$$s_n(t) = \begin{cases} \frac{1}{a}t \cdot (\frac{1}{2} - \frac{1}{n}) \\ (\frac{1}{2} - \frac{1}{n}) + \frac{1}{b-a}(t-a) \cdot \frac{1}{n} \\ \frac{1}{2} + \frac{1}{1-b}(t-b) \cdot \frac{1}{2} \end{cases} \qquad s(t) = \begin{cases} \frac{1}{a}t \cdot \frac{1}{2} & t \in [0,a) \\ \frac{1}{2} & t \in [a,b) \\ \frac{1}{2} + \frac{1}{1-b}(t-b) \cdot \frac{1}{2} & t \in [b,1] \end{cases} \tag{6.45}$$

In words, in the interval $[0, a)$ the time parametrization s_n grows linearly from 0 to $1/2 - 1/n$ whereas s grows linearly from 0 to $1/2$; then in the interval $[a, b)$ s_n grows linearly till $1/2$ while s stays constant; finally in the interval $[b, 1]$ they both grows linearly from $1/2$ to 1. Using this parametrization, we have that $||s_n - s|| = 1/n$. Hence

$$d_{M_1}(x_n, x) \leq \{||s_n - s|| \vee ||z_n - z||\} = 1/n \to 0 \quad \text{as } n \to \infty. \tag{6.46}$$

Unlike J_1, the M_1-Skorokhod topology, the one induced by the metric (6.44), can not be extended by allowing the range to be an arbitrary Polish space, because the completed graphs require linear structure.

We continue the discussion on the topologies in the functions space $\mathcal{D}$ by introducing the J_2-Skorokhod topology. For the discussion here, we assume again that the functions are real-valued and that the function domain is $[0, 1]$. We will see later on how to extend the same definitions and proof to a general bounded or unbounded interval domain I. The idea behind the J_2-Skorokhod topology is to define a metric in a similar way to what done for J_1. The main difference with respect to d_{J_1} is that we replace the set Λ of increasing homeomorphism of $[0, 1]$ with the larger class Λ' consisting of all bijections on the set $[0, 1]$, without requiring any continuity.

Definition 14. For any $x, y \in \mathcal{D}([0, 1])$, the J_2 or $J_2(I)$ distance $d_{J_2}(x, y)$ is defined as in the r.h.s. of (6.11), but with the infimum taken over all bijections $\lambda \in \Lambda'$, i.e.

$$d_{J_2}(x, y) := \inf_{\lambda \in \Lambda'} \{||x \circ \lambda - y|| \vee ||\lambda - \mathrm{id}||\}. \tag{6.47}$$

Clearly we have $\Lambda \subset \Lambda'$ and thus $d_{J_2}(x, y) \leq d_{J_1}(x, y)$. The notions of J_2-convergence in all cases of I are derived as seen earlier for J_1. The basic innovation brought by the J_2 topology is that we can consider close functions that reach close discontinuity values in a different order, provided that the close values are attained at close points.

We now want to introduce the last of the four topologies introduced by Skorokhod in [Sko56]. We saw before the centrality of the complete graph Γ_x of a function $x \in \mathcal{D}$ to define the M_1-distance. With no surprise, the M_2-Skorokhod topology will still make use of the complete graph, but relaxing the conditions on the parametric representation that was crucial in the definition of d_{M_1}. As first, we need to introduce the Hausdorff metric between compact subsets of a metric space. We define the Hausdorff distance between two compact subsets K_1 and K_2 of the metric space $\mathbb{R}^k$ as follows

$$d_H(K_1, K_2) := \sup_{x_1 \in K_1} d(x_1, K_2) \vee \sup_{x_2 \in K_2} d(x_2, K_1) \tag{6.48}$$

The Hausdorff distance indicates the supremum over all distances between a point belonging to one of two sets and the other set. Thus, even in the situation of a compact set properly contained in another compact set, the result is a positive Hausdorff distance. Note now that, if $x \in \mathcal{D}([0, 1], \mathbb{R}^k)$ then its complete graph Γ_x is a compact subset of $\mathbb{R}^{k+1}$. For any $x_1, x_2 \in \mathcal{D}$, we defined the M_2 metric on $\mathcal{D}$ as

$$d_{M_2}(x_1, x_2) := d_H(\Gamma_{x_1}, \Gamma_{x_2}). \tag{6.49}$$

The metric d_{M_2} induces the M_2-Skorokhod topology on $\mathcal{D}$, which is weaker than M_1 because it does not require any order due to the parametric representation. To be precise, the M_2-Skorokhod topology can be derived similarly to what done for M_1 after changing the definition of a parametric representation. Instead of requiring that $t \to (s(t), z(t))$ be nondecreasing, using the order on the completed graphs, we only require the time component function z to be nondecreasing. With that definition, it is evident that the M_1 topology is stronger than the M_2 topology, i.e. M_1 convergence implies M_2 convergence.

6.7 Continuity of the addition map in J_2

In this Section we want to extend Theorem 6.3.2 to the J_2-Skorokhod topology. This result is important on its own since, up to the author's knowledge, this is the first time that an explicit proof of the result appears in the literature. Moreover, we need it for proving our findings (see Chapter 9).

Theorem 6.7.1. *Let I be a (closed, open or half-open) bounded interval. The addition map $\ell : \mathcal{D}(I) \times \mathcal{D}(I) \longrightarrow \mathcal{D}(I)$ defined by $\ell(x, y) := x + y$ is measurable and it is J_2-continuous at all pairs (x, y) such that*

$$\mathrm{Disc}(x) \cap \mathrm{Disc}(y) = \emptyset. \tag{6.50}$$

The latter assertion amounts to the claim that $d_{J_2,I}(x_n, x) \to 0$ and $d_{J_2,I}(y_n, y) \to 0$, as $n \to \infty$, imply $d_{J_2,I}(x_n + y_n, x + y) \to 0$.

Proof. We follow the same line of arguments as in the proof of [Whi80, Theorem 4.1]. In fact, the measurability of ℓ is proved exactly as in the referenced theorem. As for the continuity claim, we fix $I := [a, b)$, which is the case needed in Section 9.3.4. The other three cases, $I = [a, b]$, $I = (a, b]$ or $I = (a, b)$, are proved exactly in the same way. Without loss of generality, we also assume to work with càdlàg functions (as opposed to functions that have a càdlàg and a càglàd restriction). This case happens, e.g., if $0 \leq a < b$.

For a fixed $\varepsilon > 0$, we must show that exists $\bar{n} \in \mathbb{Z}^+$ and, for all $n \geq \bar{n}$, bijections $\lambda_n : [a, b) \longrightarrow [a, b)$ such that

$$\sup_{t \in [a,b)} |\lambda_n(t) - t| < \varepsilon, \tag{6.51}$$

$$\sup_{t \in [a,b)} |(x_n + y_n) \circ \lambda_n(t) - (x + y)(t)| < \varepsilon. \tag{6.52}$$

Since x, y are càdlàg, by [Bil13, Chapter 3, Lemma 1] there exist two finite sets of points

$$\mathcal{P}_x = \{a = t_0, t_1, ..., t_n, t_{n+1} = b\} \tag{6.53}$$

$$\mathcal{P}_y = \{a = s_0, s_1, ..., s_m, s_{m+1} = b\} \tag{6.54}$$

such that, for all $i = 1, \dots, n+1$ and $j = 1, \dots, m+1$,

$$\sup_{q_1,q_2 \in [t_{i-1},t_i)} |x(q_1) - x(q_2)| < \frac{\varepsilon}{8}, \tag{6.55}$$

$$\sup_{q_1,q_2 \in [s_{i-1},s_i)} |y(q_1) - y(q_2)| < \frac{\varepsilon}{8}. \tag{6.56}$$

From this construction we have that the discontinuity points of x (respectively y) with jump size bigger than $\varepsilon/8$ are contained in $\mathcal{P}_x$ (respectively $\mathcal{P}_y$). By hypothesis these two sets of points are disjoint. Moreover, we can select the other points of $\mathcal{P}_x$ and $\mathcal{P}_y$ so that $\mathcal{P}_x \cap \mathcal{P}_y = \{a, b\}$. Let 4δ be the distance between the closest pair of points of $\mathcal{P} := \mathcal{P}_x \cup \mathcal{P}_y$. For $i = 1, \dots, n$ and $j = 1, \dots, m$, we construct closed intervals $\mathcal{J}_i^{(x)}$ and $\mathcal{J}_j^{(y)}$ such that

$$[t_i - \delta, t_i + \delta] \subset \operatorname{int}\big(\mathcal{J}_i^{(x)}\big) \subset \mathcal{J}_j^{(y)} \subset (t_i - 2\delta, t_i + 2\delta), \tag{6.57}$$

$$[s_j - \delta, s_j + \delta] \subset \operatorname{int}\big(\mathcal{J}_i^{(x)}\big) \subset \mathcal{J}_j^{(y)} \subset (s_j - 2\delta, s_j + 2\delta). \tag{6.58}$$

This implies in particular that these intervals are pairwise disjoint.

Now let us assume that there exist $\bar{n} \in \mathbb{Z}^+$ and bijections $\mu_n, \nu_n : [a, b) \longrightarrow [a, b)$ so that, for all $n \geq \bar{n}$,

$$\sup_{t\in[a,b)} |\mu_n(t) - t| < \min\{\varepsilon, \delta\}, \quad \sup_{t\in[a,b)} |x_n \circ \mu_n(t) - x(t)| < \frac{\varepsilon}{4}, \quad \mu_n(\mathcal{J}_i^{(x)}) = \mathcal{J}_i^{(x)} \tag{6.59}$$

$$\sup_{t\in[a,b)} |\nu_n(t) - t| < \min\{\varepsilon, \delta\}, \quad \sup_{t\in[a,b)} |y_n \circ \nu_n(t) - y(t)| < \frac{\varepsilon}{4}, \quad \nu_n(\mathcal{J}_j^{(y)}) = \mathcal{J}_j^{(y)} \tag{6.60}$$

for all $i = 1, \dots, n$ and $j = 1, \dots, m$. The first and second conditions in both (6.59) and (6.60) can be satisfied by the hypotheses $x_n \to x$, $y_n \to y$ in $(\mathcal{D}([a, b)), J_2)$. We postpone for a moment the proof that μ_n, ν_n can be found to satisfy the third conditions as well. Let us construct the bijection $\lambda_n : [a, b) \longrightarrow [a, b)$ as follows:

$$\lambda_n(t) := \begin{cases} \mu_n(t) & \text{for } t \in \mathcal{J}_i^{(x)} \text{ with } i = 1, \dots, n, \\ \nu_n(t) & \text{for } t \in \mathcal{J}_j^{(y)} \text{ with } j = 1, \dots, m, \\ t & \text{otherwise.} \end{cases} \tag{6.61}$$

We have the following estimates:

$$\sup_{t\in\mathcal{J}_i^{(x)}} |x_n \circ \lambda_n(t) - x(t)| = \sup_{t\in\mathcal{J}_i^{(x)}} |x_n \circ \mu_n(t) - x(t)| < \frac{\varepsilon}{4}, \tag{6.62}$$

by (6.59), and

$$\begin{aligned}
&\sup_{t\in\mathcal{J}_j^{(y)}} |x_n\circ\lambda_n(t)-x(t)| \qquad (6.63)\\
&\quad= \sup_{t\in\mathcal{J}_j^{(y)}} |x_n\circ\nu_n(t)-x(t)|\\
&\quad\le \sup_{t\in\mathcal{J}_j^{(y)}} |x_n\circ\nu_n(t)-x\circ\mu_n^{-1}\circ\nu_n(t)| + \sup_{t\in\mathcal{J}_j^{(y)}} |x\circ\mu_n^{-1}\circ\nu_n(t)-x(t)|\\
&\quad\le \sup_{u\in[a,b)} |x_n\circ\mu_n(u)-x(u)| + \sup_{t\in\mathcal{J}_j^{(y)}} |x(\mu_n^{-1}\circ\nu_n(t))-x(t)|\\
&\quad< \frac{\varepsilon}{4}+\frac{\varepsilon}{8}=\frac{3\varepsilon}{8}.
\end{aligned}$$

In the first term of the final estimate of (6.63) we have renamed $u := \mu_n^{-1}\circ\nu_n(t)$ and used (6.59). For the second term we have observed that, by (6.59)-(6.60), the bijection $\mu_n^{-1}\circ\nu_n$ is closer to the identity than 2δ. Since $t\in\mathcal{J}_j^{(y)}$ this implies, by (6.58), that t and $\mu_n^{-1}\circ\nu_n(t)$ belong to the same interval $[t_{i-1},t_i)$, for some i. Thus we have used (6.55). Lastly, if we denote $\mathcal{J} := \big(\bigsqcup_{i=1}^n \mathcal{J}_i^{(x)}\big)\sqcup\big(\bigsqcup_{j=1}^m \mathcal{J}_j^{(y)}\big)$,

$$\begin{aligned}
&\sup_{t\in[a,b)\setminus\mathcal{J}} |x_n\circ\lambda_n(t)-x(t)| \qquad (6.64)\\
&\quad= \sup_{t\in[a,b)\setminus\mathcal{J}} |x_n(t)-x(t)|\\
&\quad\le \sup_{t\in[a,b)\setminus\mathcal{J}} |x_n(t)-x\circ\mu_n^{-1}(t)| + \sup_{t\in[a,b)\setminus\mathcal{J}} |x\circ\mu_n^{-1}(t)-x(t)|\\
&\quad\le \sup_{u\in[a,b)} |x_n\circ\mu_n(u)-x(u)| + \sup_{t\in[a,b)\setminus\mathcal{J}} |x(\mu_n^{-1}(t))-x(t)|\\
&\quad< \frac{\varepsilon}{4}+\frac{\varepsilon}{8}=\frac{3\varepsilon}{8},
\end{aligned}$$

where we have used the same arguments as for (6.63): observe in fact that if $t\notin\mathcal{J}$ then, by (6.57)-(6.58), t is at distance larger than δ from $\mathcal{P}\setminus\{a,b\}$. By (6.59) $|\mu_n^{-1}(t)-t|<\delta$ and so t and $\mu_n^{-1}(t)$ belong to the same interval $[t_{i-1},t_i)$, for some i, triggering (6.55).

From (6.62)-(6.64) we have that $\sup_{t\in[a,b)}|x_n\circ\lambda_n(t)-x(t)|<\varepsilon/2$, and the same obviously holds for y, whence

$$\begin{aligned}
&\sup_{t\in[a,b)} |(x_n+y_n)\circ\lambda_n(t)-(x+y)(t)| \qquad (6.65)\\
&\qquad\le \sup_{t\in[a,b)} |x_n\circ\lambda_n(t)-x(t)| + \sup_{t\in[a,b)} |y_n\circ\lambda_n(t)-y(t)|<\varepsilon,
\end{aligned}$$

giving (6.52). The inequality (6.51) follows from definition (6.61) and (6.59)-(6.60) and so Theorem 6.7.1 is proved.

It remains to show that the bijections μ_n,ν_n can be chosen to satisfy the rightmost conditions of (6.59)-(6.60), for a suitable choice of the intervals $\{\mathcal{J}_i^{(x)}\}_{i=1}^n$, $\{\mathcal{J}_j^{(y)}\}_{j=1}^m$. We proceed by explicitly constructing μ_n, as the construction of ν_n is completely analogous.

The hypothesis $d_{J_2,[a,b)}(x_n, x) \to 0$ amounts to the existence of bijections $\rho_n : [a,b) \longrightarrow [a,b)$ such that, for $n \geq \bar{n}$,

$$\sup_{t\in[a,b)} |\rho_n(t) - t| < \frac{1}{2}\min\{\varepsilon, \delta\}, \tag{6.66}$$

$$\sup_{t\in[a,b)} |x_n \circ \rho_n(t) - x(t)| < \frac{\varepsilon}{8}. \tag{6.67}$$

For $i = 1, \ldots, n$, set $[a_i, b_i] := [t_i - \delta, t_i + \delta]$ and

$$\mathcal{J}_i^{(x)} := [a_i'', b_i''] := \left[\inf\big(\{a_i\} \cup \rho_n([a_i, b_i])\big) - \frac{\delta}{2}\,,\, \sup\big(\{b_i\} \cup \rho_n([a_i, b_i])\big) + \frac{\delta}{2}\right]. \tag{6.68}$$

By construction $|a_i'' - t_i| < 2\delta$ and $|b_i'' - t_i| < 2\delta$. Hence the intervals $\mathcal{J}_i^{(x)}$ satisfy (6.57).

The bijection $\mu_n : [a,b) \longrightarrow [a,b)$ is defined with the following structure:

$$\mu_n(t) := \begin{cases} \mu_n^{(i)}(t) & \text{for } t \in \mathcal{J}_i^{(x)} \text{ with } i = 1, \ldots, n, \\ t & \text{otherwise,} \end{cases} \tag{6.69}$$

where $\mu_n^{(i)} : \mathcal{J}_i^{(x)} \longrightarrow \mathcal{J}_i^{(x)}$ are bijections that we construct in several steps as follows.

First, on $[a_i, b_i] \subset \mathcal{J}_i^{(x)}$, we define $\mu_n^{(i)}|_{[a_i,b_i]} := \rho_n|_{[a_i,b_i]}$. In light of (6.68) and applying (6.66) to $t \in [a_i, b_i]$, we see that

$$\mu_n^{(i)}([a_i, b_i]) = \rho_n([a_i, b_i]) \subset (a_i'', b_i'') \subset \mathcal{J}_i^{(x)}. \tag{6.70}$$

Denote $\mathcal{A}_i := [a_i'', t_i) \setminus \mu_n^{(i)}([a_i, b_i])$ and $\mathcal{B}_i := [t_i, b_i'] \setminus \mu_n^{(i)}([a_i, b_i])$. These are, respectively, the lower and upper parts of $[a_i'', b_i''] = \mathcal{J}_i^{(x)}$ that have not yet been assigned as image points of $\mu_n^{(i)}$ (which is only partially defined at this stage). Using the fact that the inequality (6.66) is strict, we can find η with $0 < \eta < \min\{\varepsilon, \delta\}/2$ such that

$$\inf \mu_n^{(i)}([a_i, b_i]) > a_i - \eta\,, \qquad \sup \mathcal{A}_i < a_i + \eta\,, \tag{6.71}$$

$$\sup \mu_n^{(i)}([a_i, b_i]) < b_i + \eta\,, \qquad \inf \mathcal{B}_i > b_i - \eta. \tag{6.72}$$

Set $a_i' := a_i - \eta$ and $b_i' := b_i + \eta$. We have $a_i'' < a_i' < a_i < b_i < b_i' < b_i''$. The inequalities (6.71) show that the yet-to-be-assigned image set $\mathcal{A}_i$ can we written as

$$\mathcal{A}_i = [a_i'', a_i'] \sqcup (\mathcal{A}_i \cap (a_i', a_i + \eta)), \tag{6.73}$$

where $\mathcal{A}_i \cap (a_i', a_i + \eta)$ has the cardinality of the continuum because, by the first inequality of (6.71), there exists $\sigma > 0$ such that $(a_i', a_i' + \sigma) \subset \mathcal{A}_i \cap (a_i', a_i + \eta)$. By reasons of cardinality, then, there exists a bijection $\phi_i^- : (a_i', a_i) \longrightarrow \mathcal{A}_i \cap (a_i', a_i + \eta)$. By construction, since $a_i' = a_i - \eta$,

$$\sup_{t\in(a_i',a_i)} |\phi_i^-(t) - t| \leq 2\eta < \min\{\varepsilon, \delta\}. \tag{6.74}$$

We define $\mu_n^{(i)}|_{(a_i',a_i)} := \phi_i^-$ and $\mu_n^{(i)}|_{[a_i'',a_i']} := \mathrm{id}$. Analogously, the inequalities (6.72) give

$$\mathcal{B}_i = (\mathcal{B}_i \cap (b_i - \eta, b_i')) \sqcup [b_i', b_i''] \tag{6.75}$$

and there exists a bijection $\phi_i^+ : (b_i, b_i') \longrightarrow \mathcal{B}_i \cap (b_i - \eta, b_i')$ for which the analogue of estimate (6.74) holds. Finally, we define $\mu_n^{(i)}|_{(b_i,b_i')} := \phi_i^+$ and $\mu_n^{(i)}|_{[b_i',b_i'']} := \mathrm{id}$. This completes the definition of $\mu_n^{(i)}$ as a bijection of $\mathcal{J}_i^{(x)}$.

By (6.66), (6.74) and its analogue for ϕ_i^+, we see that

$$\sup_{t\in\mathcal{J}_i^{(x)}} |\mu_n^{(i)}(t) - t| < \min\{\varepsilon, \delta\}. \tag{6.76}$$

Also, by the definition of $\mu_n^{(i)}|_{[a_i,b_i]}$ and (6.67),

$$\sup_{t\in[a_i,b_i]} |x_n \circ \mu_n^{(i)}(t) - x(t)| < \frac{\varepsilon}{8}. \tag{6.77}$$

Furthermore,

$$\begin{aligned} &\sup_{t\in\mathcal{J}_i^{(x)}\setminus[a_i,b_i]} |x_n \circ \mu_n^{(i)}(t) - x(t)| \qquad (6.78)\\ &\quad\le \sup_{t\in\mathcal{J}_i^{(x)}\setminus[a_i,b_i]} |x_n \circ \mu_n^{(i)}(t) - x \circ \rho_n^{-1} \circ \mu_n^{(i)}(t)| + \sup_{t\in\mathcal{J}_i^{(x)}\setminus[a_i,b_i]} |x \circ \rho_n^{-1} \circ \mu_n^{(i)}(t) - x(t)|\\ &\quad\le \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \frac{\varepsilon}{4}. \end{aligned}$$

The above estimates are derived in a way similar to that used in (6.63): for the first term we use (6.67) after the change of variable $u := \rho_n^{-1} \circ \mu_n^{(i)}(t)$; for the second term we use (6.55) and the fact that t and $\rho_n^{-1} \circ \mu_n^{(i)}(t)$ belong to the same interval $[t_{k-1}, t_k)$, for some k (this is because, due to (6.66) and (6.76), $t \in \mathcal{J}_i^{(x)}$ and $|\rho_n^{-1} \circ \mu_n^{(i)}(t) - t| < 3\delta/2$). Moreover, denoting $\mathcal{J}^{(x)} := \bigsqcup_{i=1}^n \mathcal{J}_i^{(x)}$, it is now easy to use (6.67), (6.57) and (6.55) to estimate

$$\begin{aligned} &\sup_{t\in[a,b)\setminus\mathcal{J}^{(x)}} |x_n \circ \mu_n(t) - x(t)| \qquad (6.79)\\ &\quad= \sup_{t\in[a,b)\setminus\mathcal{J}^{(x)}} |x_n(t) - x(t)|\\ &\quad\le \sup_{t\in[a,b)\setminus\mathcal{J}^{(x)}} |x_n(t) - x \circ \rho_n^{-1}(t)| + \sup_{t\in[a,b)\setminus\mathcal{J}^{(x)}} |x \circ \rho_n^{-1}(t) - x(t)|\\ &\quad< \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \frac{\varepsilon}{4}. \end{aligned}$$

Finally, the definition (6.69) of μ_n and the inequalities (6.76)-(6.79) yield (6.59) and conclude the proof of Theorem 6.7.1. □

Remark 7. One can define a new topology of the Skorokhod type in the same way as Definitions 10 and 14 but taking the infimum in (6.47) over all *piecewise increasing and continuous* (PIC) bijections $\lambda : I \longrightarrow I$. A PIC bijection $\lambda : I \longrightarrow I$ is one such that I can be partitioned into a finite number of intervals, on each of which λ is increasing and continuous. Observe that in this case λ^{-1} is also a PIC bijection. For want of a better name, let us call this topology $J_{3/2}$. Evidently, $J_{3/2}$ is weaker than J_1 and stronger than J_2. It is not hard to see that Theorem 6.7.1 can be proved as well with the $J_{3/2}$-distance in place of the J_2-distance. Furthermore, in the proofs of Theorems 9.2.2 and 9.2.4, every time we needed to construct a sequence of bijections in order to prove a J_2-convergence, we have indeed produced a sequence of PIC bijections. Therefore, all assertions in this paper that are stated for the topology J_2, see (9.22), (9.30), (9.32), hold for the topology $J_{3/2}$ as well.

Chapter 7

Limit theorems for random walks

Introduction Random Walks are used very often in probability theory and in everyday's life. Suppose that we are repeatedly tossing a fair coin. We start with zero points and, after every toss, we gain a point if the outcome is Head, whereas we lose a point in case of Tail. Many questions arise in this very simple framework. How many points will we have after n tosses? If we fix a winning value $W > 0$, how long does it take before we reach (if ever) W? If we fix also a losing value $L < 0$, what is the probability to reach the winning values W before the losing value L? These are some basic questions that we might ask during this tossing game. It turns out that there exists an object, called Random Walk, that is perfectly suitable to model this situation. Moreover, the mathematical description we acquaint by using random walk is so powerful that we can answer to many more questions than the simple ones pointed above. The centrality of random walks in probability theory relies on the abundance of applications we can deal with by a model focused on a random walk. We start now by giving proper definition.

Definition 15. Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of random variables with values in the same metric space. The random walk $S = (S_n)_{n\in\mathbb{N}}$ is defined as follows

$$\begin{aligned} S_0 &= 0 \\ S_n &= \sum_{i=1}^{n} \xi_i \quad n > 0. \end{aligned} \tag{7.1}$$

Note that in general the random variables $(\xi_i)_{i\in\mathbb{N}}$ can follow any general distribution. However, in many cases, we require some assumption in order obtain a model that is close to the situation we want to describe. We refer to the increments ξ_i as the steps of our random walk. Hence, suppose that a walker is moving according to the steps $(\xi_i)_{i\in\mathbb{N}}$. A possible assumption is that every step is independent from the others. This is reasonable in the case of multiple coin tossing, or in the case where the back-forth direction of each step is based on a fair dice roll. We can decide ξ_i to equals 1 if the dice displays a number smaller than 5, otherwise $\xi_1 = -1$. From this example it appears clearly that the positive direction is privileged, even though the steps are independent. Suppose now that we want to model the motion of a random driver in a city. Every time he encounters

a road crossing, he choses one of the ways with uniform probability. This situation can be modeled again as a random walk. Note, however, that the possible directions depend strictly on the road map of the particular city we are considering. To conclude this series of examples, think about the day-by-day trend of a stock price in stock markets. In this situation we can reach any level and there are no physical constraints for the price. However, some stocks display a strong correlation of the steps, i.e. some stocks are likely maintain the trend the day after. This is a different situation with respect to the one with the dice explained above since here the probability of a particular direction is not fixed, but depends on the previous steps. We just gave a glance of some possibilities among the vast amount of scenarios we can face by using a random walk.

7.1 Increments with finite variance

We start this section by analyzing some properties of the random walk in a simple scenario: we assume now that the increments $(\xi_i)_{i\in\mathbb{N}}$ are independent and identically distributed (i.i.d.), with finite second moment.

7.1.1 Law of Large Numbers

One of the main results in probability theory is the Law of Large Numbers. We use this terminology to indicate any theorem stating that $\frac{1}{n}S_n$ converges in some sense. There are many results of this kind, and the convergence can be in probability or almost sure. In this latter case, the theorem is denoted by the Strong Law of Large Numbers.

Theorem 7.1.1 (Strong Law of Large Numbers)**.** *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random variables defined on the same space $(\Omega, \mathcal{F}, \mathbb{P})$. Let*

$$\mu = \mathbb{E}[\xi_i] < \infty \qquad \text{and } \sigma^2 = \mathbb{V}ar[\xi_i] < \infty \tag{7.2}$$

Define the random walk $(S_n)_{n\in\mathbb{N}}$ as in (7.1). Then

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \xi_i = \mu \tag{7.3}$$

where convergence is intended almost surely and in L^2.

Suppose that we are observing a realization $\omega \in \Omega$ of the sequence of n random variables, then we refer to $\frac{1}{n}\sum_{i=1}^{n}\xi_i(\omega)$ as the empirical mean, and we denote it by $\bar{\mu}$. This is linked to experiments and statistics. Suppose that we are performing an experiment repeatedly and independently, then $\bar{\mu}$ represents the average value of the outcome of the experiments. Then $\bar{\mu}$ is called an estimator for the mean μ of the distribution of the random variables. Usually we are interested to study when $\bar{\mu}$ converges to μ as $n \to \infty$.

The utility of this theorem is enormous. Jut to give an examples, think about Montecarlo approximation, where the Law of Large Numbers (LLN) is used to estimate the quantity $\int f(x)dx$ that can be difficult to compute analytically. Making the story short,

the idea lies behind the fact that integrals can be thought as expectations, thus we can approximate $\mathbb{E}[f(\xi)]$ by the empirical average $\frac{1}{n}\sum_{i=1}^{n} f(\xi_i)$.

Moreover, one may ask how large n needs to be in order to obtain high levels of confidence when estimating the usually unknown true value μ using the empirical average $\bar{\mu}$. The Central Limit Theorem is the right answer to this and other questions.

7.1.2 Central Limit Theorem

With the LLN in mind, a new question arises naturally. How fast $\bar{\mu}$ converges to μ? How big should n be to sure that $\mid \bar{\mu} - \mu \mid < \varepsilon_n$? In other words, we know that $\frac{1}{n}S_n \to \mu$ and we want to say something more on its distribution. The CLT states that $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$, where $\mu = \mathbb{E}[\xi]$ and $\sigma^2 = \mathbb{V}ar[\xi]$. The Central Limit Theorem is one of the most impressive achievements of probability theory. Note that the hypothesis required are really minimal. In facts, we only require the random variables to have finite variance σ^2, without any specific constraint on the distribution. Central limit theorems have played a paramount role in probability theory. There are many different and more complicated versions of this theorem, as we want to relax the i.i.d. hypothesis of the random variables. We cite here the classical and easier version.

Theorem 7.1.2 (Central Limit Theorem)**.** *Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\xi_i] = \mu$ and $\mathbb{V}ar[\xi_i] = \sigma^2 < \infty$. Let $(S_n)_{n\in\mathbb{N}}$ be defined as in* (7.1)*. Then*

$$Y_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1). \tag{7.4}$$

As you may guess from the theorem, the term "central" refers to the pervasive, although non unique, role of the normal distribution as a limit of the normalized sum of (classically independent, like in Theorem 7.1.2) random variables.

Note that in (7.4) we defined a sequence $(Y_n)_{n\in\mathbb{N}}$ of random variables converging to a standard normal. We highlight once again that every Y_n is a random variable, obtained from a suitable rescaling of the sum of the n random variables $\xi_1, .., \xi_n$.

7.1.3 Invariance principle

We want now to say something more about the random walk $S = (S_n)_{n\in\mathbb{N}}$. The idea is to build a stochastic process whose trajectories are obtained by interpolating the discrete steps of the random walk S. After a suitable rescaling, we want to prove convergence of this process. This is the goal of the Donsker's Theorem (also known as invariance principle), which we want to state paying attention on the notation, recalling the notions of weak convergence stated in Chapter 5.

In a few words, Donsker's theorem is a Functional Central Limit Theorem (FCLT) generalizing the CLT above. Theorem 7.1.2 is a limit theorem for S_n, the n-th partial sum, as $n \to \infty$; instead the FCLT is a limit theorem for the entire sequence of partial sums. We express it via the normalized partial-sum process

$$S^{(n)}(t) := \frac{S_{\lfloor nt \rfloor} - \mu nt}{\sigma\sqrt{n}}, \quad t \in [0,1], \tag{7.5}$$

where $\lfloor \cdot \rfloor$ denotes the integer part of a number, i.e. $\lfloor nt \rfloor = m$ if m is the biggest integer such that $m \leq nt$. Note that $S^{(n)}$ is a process with value in $\mathcal{D}$, where $t \in [0,1]$ is the time index, whereas (n) is an index for the sequence of processes.

Theorem 7.1.3 (Donsker's Theorem)**.** *Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\xi_i] = \mu$ and $\mathbb{V}ar[\xi_i] = \sigma^2 < \infty$. Let $(S_n)_{n\in\mathbb{N}}$ be as in (7.1), and $S^{(n)}$ be the normalized partial-sum processes defined in (7.5). Then*

$$S^{(n)} \xrightarrow{d} B \qquad \text{in } (\mathcal{D}, J_1), \tag{7.6}$$

where $B = (B(t), t \in [0,1])$ is the standard Brownian motion.

The Brownian Motion is a process with a lot of nice properties. Among them, a non trivial one is that it really exists. In facts, the usual way to introduce the Brownian Motion is to give properties that we want to be satisfied, and the difficult part is to show that there actually exists a process with those properties. However, existence is a consequence of Donsker's Theorem, meaning that the Brownian Motion can be defined as the limit process obtained in (7.6), once we checked that the limit for the normalized partial sums (7.5) exists. We give here a definition of the standard univariate Brownian Motion.

Definition 16. The Brownian Motion is a process $(B(t), t \geq 0)$ that satisfies the following properties

1. $B(0) = 0$ with probability 1.
2. Trajectories are almost surely continuous.
3. It has independent increments, i.e. for any $0 \leq s < t < \infty$, $B(t) - B(s)$ is independent from the σ-algebra generated by $\{B(s'), s' \leq s\}$.
4. It has gaussian increments, i.e. for any $0 \leq s < t < \infty$, $B(t) - B(s) \sim \mathcal{N}(0, t-s)$.

Note that both the Brownian Motion and the sequence of processes $(S^{(n)})_{n\in\mathbb{N}}$ have continuous trajectories, hence Donsker's Theorem holds in the space $\mathcal{C}$ of continuous function. However, the statement in $\mathcal{D}$ we gave above allows us to define the sequence of processes $(S^{(n)})_{n\in\mathbb{N}}$ without introducing any interpolating term. However, Donsker's theorem is usually formulated in $\mathcal{C}$, with convergence of the processes

$$S^{(n)}(t,\omega) := \frac{1}{\sigma\sqrt{n}} \left(S_{\lfloor nt \rfloor} - \mu \lfloor nt \rfloor + \frac{1}{n}(nt - \lfloor nt \rfloor)(\xi_{\lfloor nt \rfloor + 1}(\omega) - \mu) \right), \quad t \geq 0 \tag{7.7}$$

to the Brownian Motion with respect to the uniform topology.
Stochastic processes with stationary and independent increments are called Lévy processes; hence the Brownian Motion is a particular case of a Lévy process with continuous sample path.

From Definition 16 we can infer an important property of self-similarity, namely that the distribution of the process $(B(ct), t \geq 0)$ is the same of the one of $(\sqrt{c}B(t), t \geq 0)$,

for any constant $c > 0$. In the case of Brownian Motion, taking out the square root of a positive constant c does not affect the distribution of the process. This is a special case of self-similarity, whose general description is the content of the next definition.

Definition 17. We say that an $\mathbb{R}^d$-values stochastic process $\{Z(t), t \geq 0\}$ is self-similar with index $H > 0$ if, for all constants $c \in \mathbb{R}^+$, we have

$$\{Z(ct), t \geq 0\} \stackrel{d}{=} \{c^H Z(t), t \geq 0\}. \tag{7.8}$$

The scaling exponent H is often called the Hurst parameter.

Note that we necessarily have $Z(0) = 0$ with probability. According to the previous definition, the Brownian Motion is self similar with index $H = 1/2$. Graphically, self similar processes are such that the plots are identical (same shape, but different scale) at any time scale.

We want to spend some more comment on Donsker's theorem. Note that, if we assume as in Theorem 7.1.3 that $(\xi_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with $\mathbb{E}[\xi_i] = \mu$ and $\mathbb{V}ar[\xi_i] = \sigma^2 < \infty$, and defining $S_n = \sum_{i=1}^n \xi_i$, then using the CLT Theorem 7.1.2 we have that

$$\frac{S_n - \mu n}{\sigma\sqrt{n}} \stackrel{d}{\to} \mathcal{N}(0, 1) \tag{7.9}$$

The idea of Theorem 7.1.3 is to formulate a refinement of the CLT by proving weak convergence in distribution of a certain functional of the partial sum S_n. Equation (7.6) can be reformulated in terms of weak convergence in the following way

$$P_n \stackrel{w}{\to} W \quad \text{as } n \to \infty, \tag{7.10}$$

where, for each $n \in \mathbb{N}$, P_n is the law of the process $S^{(n)}$, i.e. $P_n(A) = \mathbb{P}(\omega : S^{(n)}(\omega) \in A)$, whereas W is the Wiener measure, the law on the space $\mathcal{C}$ induced by the Brownian Motion. We want to highlight once more that (7.6) is more powerful than the classical CLT. This should be clear from Chapter 5, still we want to mark this point. If we denote by $A = \{x \in \mathcal{C} : x(1) \leq M\}$ for some $M \in \mathbb{R}$ then we obtain that

$$P_n(A) = \mathbb{P}\left[\omega : \frac{S_n(\omega) - \mu n}{\sigma\sqrt{n}} \leq M\right], \tag{7.11}$$

Since $S^{(n)}(t = 1) = (S_n(\omega) - \mu n)/\sigma\sqrt{n}$. Moreover, we have that $W(\partial A) = 0$, hence (7.10) implies that

$$\mathbb{P}\left[\omega : \frac{S_n(\omega) - \mu n}{\sigma\sqrt{n}} \leq M\right] \to W\left[x : x(1) \leq M\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du, \tag{7.12}$$

so the CLT stated in (7.4) is a simple consequence of the Donsker's Functional CLT. In facts, Theorem 7.1.3 says that the entire path of the continuously interpolated random walk $S^{(n)}$ during the first n steps is, for n large, distributed approximately as the path up to time $t = 1$ of a Brownian Motion, whereas the CLT only states that the distribution of the final position of the normalized partial-sum process (7.5) is close to those of a Brownian Motion at time $t = 1$.

7.2 Increments with infinite variance

During this Chapter we recalled some classical results of probability theory under the usual assumption that the increments $(\xi_i)_{i\in\mathbb{N}}$ of the random walk follow a distribution with finite mean and variance. Without going deep into this derivation, we can say that the random variables $(\xi_i)_{i\in\mathbb{N}}$ have finite q-th moment, for some index $q \geq 2$. We want now to investigate how the world change if we drop the finite second moment condition. Note that, when the second moments are infinite, there still may be limits, but the limits are very different. The existence of a limit depends critically on the regular behavior of the tails of the underlying probability distribution.
This is a very broad topic, and we seriously risk to spend too many pages without even be comprehensive of the minimal aspects of this field, hence we decided from the beginning to make a specific choice of random variables. However, regularity of the tails is very natural to assume. Throughout this thesis we will use random variables that are stable according to the definitions we give in this section. We will be focused on the case where the new limit processes have discontinuous sample path, so the space $\mathcal{D}$ become of central importance, and convergence can not be showed in $\mathcal{C}$ like in Theorem 7.1.3. To start with, assume here to deal with a sequence $(\zeta_i)_{i\in\mathbb{N}}$ id i.i.d. random variables with infinite second moment. Again, we are interested in the partial sum $S_n = \zeta_1 + \cdots \zeta_n$ for $n \geq 1$, with $S_0 = 0$.

We may wonder which particular infinite variance distribution is the best to choose for our random walk. Definition 18 and Theorem 7.2.1 clarify why our choice of stable distribution is natural.

Definition 18. Let μ be a nontrivial distribution on $\mathbb{R}$. The domain of attraction of μ is the set of all the random variables X with the property that there exists a sequence of numbers $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ such that

$$\frac{S_n - m_n}{c_n} \xrightarrow{\mathrm{d}} \mu. \tag{7.13}$$

Theorem 7.2.1. *Let μ be a nontrivial distribution on $\mathbb{R}$. The domain of attraction of μ is non-empty if and only if μ is a stable distribution.*

A proof of the previous Theorem can be found in [Kle08, Theorem 16.23].
Paraphrasing the previous Theorem, if we want our partial sum, eventually centered and rescaled, to converge to a non trivial limit, we need the increments to fall in a domain of attraction of a stable law. Therefore, we first dedicate a few pages introducing this class of random variables, and then we state the analogous of the results of the previous Section when dealing with stable distributions.

7.2.1 Stable Laws

In what follows we will deal with stable laws. Before moving on, we give the precise definition and some basic properties.

Definition 19. A random variable X is said to have a stable law if, for any $a_1, a_2 \in \mathbb{R}$ there exists $b \in \mathbb{R}$ and $c \in \mathbb{R}^+$ such that

$$a_1 x_1 + a_2 X_2 \stackrel{d}{=} b + cX, \tag{7.14}$$

where X_1, X_2 are independent copies of X. A stable law is strictly stable if $b \equiv 0$.

A very important property of stable laws is that the constant $c = c(a_1, a_2)$ is such that it holds

$$c^\alpha = a_1^\alpha + a_2^\alpha \tag{7.15}$$

for some constant α that is called the index of the stable law. The relation above implies that, if we take $a_1 \equiv 1$ we obtain, for any $n \in \mathbb{N}$,

$$X_1 + X_2 + \cdots X_n \stackrel{d}{=} n^{1/\alpha} X + b_n \tag{7.16}$$

where again $X_1, .., X_n$ are independent copies of X. Definition 19 is very clear, but still difficult to handle.

Property (7.14) is simple to check in order to state if a particular variable is stable, however it does not give significant insight of the behavior of a stable random variable. We want to characterize a stable distribution by giving some explicit form of its distribution.

The parametrization we give now comes from the work of [ST94] and it gives us an explicit parametrization of α-stable random variables on the real line. It depends on four parameters. The stability index $\alpha \in (0, 2]$ is the main index of the distribution; in the following we will always refer to this index-parameter, except when stated differently. The other three parameters give information on other characteristics of the distribution: the scale parameter $\sigma > 0$; the skewness parameter $\beta \in [-1, 1]$; and the location parameter $\mu \in \mathbb{R}$. Note that, while the scale and skewness parameters have connection with their role of characterizing amplitude and asymmetry, the location (or shift) parameter μ can be interpreted as the mean of the distribution if and only if $\alpha \geq 1$. In facts, if $\alpha < 1$, the mean is infinite. Denoting a generic α-stable distribution by $S_\alpha(\sigma, \beta, \mu)$, then the logarithm of its characteristic function is

$$\log \mathbb{E}\left[e^{i\theta S_\alpha(\sigma,\beta,\mu)}\right] = \begin{cases} -\sigma^\alpha |\theta|^\alpha (1 - i\beta sgn(\theta) \tan(\frac{\pi\alpha}{2})) + i\mu\theta, & \alpha \neq 1 \\ -\sigma|\theta|(1 + i\beta\frac{2}{\theta} sgn(\theta) \log|\theta|) + i\mu\theta, & \alpha = 1 \end{cases} \tag{7.17}$$

We point out again that the case $\alpha = 1$ is special, with special properties and particular formulas. This happens because it is a boundary case. In this thesis we deal with $\alpha \in (0, 1)$ and $\alpha \in (1, 2)$, without investigating the singular case $\alpha = 1$. Note that we exclude from our treatment also $\alpha = 2$, the gaussian case. If $\alpha = 2$ then μ is the mean, $2\sigma^2$ the variance, and the value of β is irrelevant since $\tan(\pi) = 0$. If we combine expression (7.17) with property (7.16) we obtain that, if $X_1, .., X_n$ are i.i.d. random variables distributed as $S_\alpha(\sigma, \beta, \mu)$,

$$X_1 + \cdots + X_n \stackrel{d}{=} \begin{cases} n^{1/\alpha} X_1 + \mu(n - n^{1/\alpha}), & \alpha \neq 1 \\ nX_1 + \frac{2}{\pi}\sigma\beta n \log(n), & \alpha = 1. \end{cases} \tag{7.18}$$

The previous equation will be central when dealing with random walks, since we can use (7.18) to approximate the behavior of the partial sum S_n for n large. To conclude this introduction on stable random variables, we want to highlight that the nonGaussian (i.e. $\alpha < 2$, that are the variables we are interested in) stable laws have power law tails. For $\alpha \in (0, 2)$ we have that

$$\mathbb{P}(S_\alpha(\sigma, \beta, \mu) > x) \sim x^{-\alpha} C_\alpha \frac{1+\beta}{2} \sigma^\alpha, \quad \text{as } x \to \infty \tag{7.19}$$

It is possible to show that an α-stable random variable have finite p-moments for every $p < \alpha$, i.e.

$$\mathbb{E}\left[|S_\alpha(\sigma, \beta, \mu)|^p\right] < \infty \iff p \in (0, \alpha) \tag{7.20}$$

In light of this fact we can view equation (7.18) under a new light. When $\alpha \in (1, 2)$ then the rescaled partial sum S_n/n converges to μ, where μ is the mean of X_1 that exists finite due to (7.20). In this case, it is possible to derive a LLN kind of result, as we will see. However, when $\alpha \in (0, 1)$, such a rescaling is no longer possible. We obtain a new kind of limit that characterizes stable laws, leading the rescaled sum $S_n/n^{1/\alpha}$ to converge in distribution to an α-stable random variable distributed as $S_\alpha(\sigma, \beta, \mu)$. Finally, back to the case $\alpha \in (1, 2)$, we will derive a CLT type of result, i.e. observing that the centered sum $S_n - \mu$, rescaled by $n^{1/\alpha}$, converges in the limit to a stable random variable. We want now to make these statements clear.

We saw above the characterization of stable laws. Similar results holds for random variables that fall in the domain of attraction of a stable law $S_\alpha(\sigma, \beta, \mu)$. With respect to Definition 18, we specify that a random variable X is said to fall in the normal domain of attraction of a stable random variable $S_\alpha(\sigma, \beta, \mu)$ if we can chose $c_n = c \cdot n^{1\alpha}$ and m_n such that the limit stable law has $\mu = 0$ and $\sigma = 1$, i.e. if

$$\frac{S_n - m_n}{c n^{1/\alpha}} \xrightarrow{\text{d}} S_\alpha(1, \beta, 0), \tag{7.21}$$

The main difference with respect to a proper stable law is that equation (7.21) is true only in the limit, whereas for a proper stable law the equality in distribution holds for any fixed $n \in \mathbb{N}$.

If X is in the normal domain of attraction on an α-stable law, then

$$\mathbb{P}(|X| > x) \sim A x^{-\alpha} \quad \text{as } x \to \infty \tag{7.22}$$

In the following we will consider a sequence $(X_n)_{n\in\mathbb{N}}$ of i.i.d. random variables whose distribution fall in the normal domain of attraction of a stable random variable $S_\alpha(\sigma, \beta, \mu)$. Then we define the random walk $S = (S_n)_{n\in\mathbb{N}}$ as follows:

$$\begin{aligned} S_0 &= 0 \\ S_n &= \sum_{i=1}^{n} \xi_i \quad n > 0. \end{aligned} \tag{7.23}$$

7.2.2 Law of Large Numbers

If $\alpha \in (1,2)$, then equation (7.20) ensure that the first moment is finite, and the mean too. Hence, if $\mu = \mathbb{E}[X] \neq 0$, we have that $S_n \overset{d}{\approx} n\mathbb{E}[X]$, and we can derive a Law of Large Numbers kind of convergence.

Theorem 7.2.2. *Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables whose distribution is a stable random variable $S_\alpha(\sigma,\beta,\mu)$ with stable index α in the range $(1,2)$. Let $S = (S_n)_{n\in\mathbb{N}}$ be defined as in (7.23). Then the following limit holds.*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mathbb{E}[X] \quad \text{as } n \to \infty. \tag{7.24}$$

Note that, if $\mathbb{E}[X] \neq 0$, then the limit is actually $\mu = \mathbb{E}[X]$. On the other hand, if $\mathbb{E}[X] = 0$, then the limit above is meaningless and we need to analyze a different mode of convergence. The correct leading order to proper rescale S_n becomes $n^{1/\alpha}$ (note that $1/\alpha < 1$ since $\alpha \in (1,2)$), driving to

$$\frac{S_n}{n^{1/\alpha}} \xrightarrow{\mathrm{d}} cS_\alpha(\sigma = 1, \beta, \mu = 0) \quad \text{as } n \to \infty. \tag{7.25}$$

A similar result can be obtained in the case $\alpha \in (0,1)$. In this case the increments $(X_i)_{i\in\mathbb{N}}$ have infinite mean, hence clearly $S_n/n \to \infty$ and we need to use a power higher than 1 to obtain a non trivial convergence. Again, the suitable exponent turns out to be $1/\alpha$ (note that this time $1/\alpha > 1$ since $\alpha \in (0,1)$), leading us again to a result that reads exactly like (7.25), with the exception that this time $S_\alpha(\sigma = 1, \beta, \mu = 0)$ is a stable distribution with a different value of α.

7.2.3 Stable Central Limit Theorem

Note that equation (7.21) can be interpreted as a stable version of the Central Limit Theorem. The parameter m_n is responsible for the centering, and $n^{1/\alpha}$ is the suitable rescaling. We will always deal with variables belonging to the normal domain of attraction of stable laws (note that proper stable laws are in their own domain of attraction), hence the stable CLT reads as stated in the following Theorem.

Theorem 7.2.3. *Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables whose distribution falls in a normal domain of attraction of a stable random variable $S_\alpha(\sigma,\beta,\mu)$ with stable index α in the range $(1,2)$. Let $S = (S_n)_{n\in\mathbb{N}}$ be defined as in (7.23). Then the following limit holds.*

$$\frac{S_n - \mu n}{n^{1/\alpha}} \xrightarrow{d} S_\alpha(\sigma, \beta, \mu = 0) \quad \text{as } n \to \infty. \tag{7.26}$$

Theorem 7.2.3 allows us to appreciate the difference between $n^{1/\alpha}$, the usual rescaling of stable laws, and $\sqrt{n}$, the rescaling used in the classical Central Limit Theorem.

7.2.4 Stable invariance principle

In the same spirit to what done for "classical" random variables, i.e. for random variables with finite second moment, we want to state a functional central limit theorem, involving the continuous process associated with the partial sum. Hence we form the normalized process by the following rule

$$S^{(n)}(t) := c_n^{-1} \left(S_{\lfloor nt \rfloor} - m_n nt \right), \quad t \geq 0 \tag{7.27}$$

where $(m_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ are deterministic sequences with $c_n \to \infty$ as $n \to \infty$, so that $S^{(n)}$ takes values in $\mathcal{D}$. Usually we have $m_n \equiv \mu$, and in all cases contained in this thesis we will assume it for simplicity, and $c_n = cn^{1/\alpha}$ if we assume that the steps of the discrete random walk fall in the normal domain of attraction of a α-stable distribution. The only case in which the translation constants $(m_n)_{n\in\mathbb{N}}$ need to properly depend on n is when the stable index is $\alpha = 1$. We will introduce α-stability soon, giving the basic definitions also for the special case $\alpha = 1$; however, when deriving results, we will always deal with $\alpha \neq 1$, hence we skip every consideration involving this spacial case. Back to our sequence (7.27), we will see how the common case is $c_n = n^{1/\alpha}$ for $\alpha \in (0, 2)$, where the index α depends on the asymptotic behavior of the tail probability $\mathbb{P}(|\xi| > t)$ as $t \to \infty$. If we assume every natural regular condition, then the normalized sum $S^{(N)}$ in (7.27) will converge in $(\mathcal{D}, J_1)$ to a stable Lévy Motion.
A Stable Process is a stochastic process whoso finite-dimensional distributions are stable laws. For example the Gaussian processes are stable since their finite-dimensional distributions are gaussian. However, Gaussian distributions are stable but they have finite second moment. We will be interested in stable processes whose finite-dimensional distributions are stable with infinite second moment. These nonGaussian stable distributions have heavy tails, meaning that exceptionally large increments are likely to happen.

We want now to derive a Functional (Stable) CLT when the increments of the underlying random walk are given by stable laws (or variables in the normal domain of attraction of stable laws). The limit process will be a stable Lévy motion, a particular case in the general class of Lévy processes. A Lévy process is a stochastic process $L \equiv (L(t), t \geq 0)$ with trajectories in the càdlàg space $\mathcal{D}$, such that $L(0) = 0$ almost surely, and it has stationary and independent increments. An α-stable Lévy motion is a Lévy process $S \equiv (S(t), t \geq 0)$ such that the increments have stable laws, i.e.

$$S(t + s) - S(s) \stackrel{d}{=} S_\alpha(t^{1/\alpha}, \beta, 0) \stackrel{d}{=} t^{1/\alpha} S_\alpha(1, \beta, 0) \tag{7.28}$$

for any $s \geq 0$ and $t > 0$. The last formula implies that the stable Lévy motion has stationary increments. Note that if $\alpha = 2$, then we recover the Brownian Motion, i.e. this definition comprehends as a particular case the Brownian Motion. However, we will deal with proper stable Lévy motions, being interested in the cases $\alpha in(0, 1)$ and $\alpha \in (1.2)$. In those scenarios, stable Lévy motion is self-similar with index $H = 1/\alpha$, i.e.

$$(S(ct), t \geq 0) \stackrel{d}{=} (c^{1/\alpha} S(t), t \geq 0). \tag{7.29}$$

Again, note that for the Brownian Motion we obtained a similar self-similarity property with Hurst-parameter $H = 1/2$. Still, for $\alpha \in (1,2)$, the stable Lévy Motion, like the Brownian Motion, has trajectories with unbounded total variation in each bounded interval.

Though there are similarities, there are also many differences between the Brownian Motion and the stable Lévy Motion with $\alpha < 2$. For example the Lévy Motion, except for its deterministic drift, is a pure-jump process. It has infinitely many discontinuities in any finite interval with probability 1. However, there exists a version of the stable Lévy Motion with trajectories in $\mathcal{D}$ and we will consider that version. We consider for a moment the case $\beta = 1$: for $\alpha < 1$, the stable Lévy Motion has nondecreasing trajectories, whereas for $\alpha \geq 1$ the jumps of the stable Lévy Motion are only positive, but the process has a negative drift.

We give now the statement of the functional CLT in the case where the summands are i.i.d. stable random variables. Note the differences with the correspondent classical FCLT in the case of finite variance random variables. Stability of the increments is reflected in the limit Lévy Motion, and the infinite-second moment is responsible for discontinuity of trajectories.

Theorem 7.2.4. *(stable invariance principle) Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of real-values random variables whose distribution belongs to the normal domain of attraction of $S_\alpha(\sigma,\beta,\mu)$ for $\alpha \in (0,1)\cup(1,2)$. Let $S = (S_n)_{n\in\mathbb{N}}$ be defined as in (7.23). Then, if we define, for any $n \in \mathbb{N}$, the associated normalized process*

$$S^{(n)}(t) := \frac{S_{\lfloor nt \rfloor} - m_n t}{c n^{1/\alpha}}, \quad t \geq 0, \tag{7.30}$$

then the following convergence in distribution holds

$$S^{(n)} \xrightarrow{d} S \quad \text{in } (\mathcal{D}, J_1), \tag{7.31}$$

where the limit process S is the standard α-stable Lévy Motion such that

$$S(t) \stackrel{d}{=} t^{1/\alpha} S_\alpha(1,\beta,0). \tag{7.32}$$

Note that m_n can be selected so that

$$m_n = \begin{cases} 0 & \text{if } \alpha \in (0,1) \\ n\mathbb{E}[X] & \text{if } \alpha \in (1,2) \end{cases} \tag{7.33}$$

In this Chapter we start by considering sums of i.i.d. random variables with finite variance. We derived classical results for this situation, showing in particular LLN, CLT, and Donsker's theorem. Then we relaxed some assumptions, restricting our analysis to the class of α-stable random variables, giving explicit expression of their distribution and focusing on the cases $\alpha \in (0,1)$ and $\alpha \in (1,2)$. We considered sums of i.i.d. stable random variables, analyzing how similar results can be obtained, focusing in particular on the differences. We recover a Law of Large Numbers for $\alpha \in (1,2)$, and a Central

Limit Theorem where the rescaling is $n^{1/\alpha}$. Finally, we obtained an invariance principle, the analogous of Donsker's theorem for stable laws, where the limit process is a stable Lévy Motion, and convergence holds in the càdlàg space $\mathcal{D}$.

One may asks what happens if we drop the independence requirement of the random variables generating the sum. The situation we obtain is quite general and difficult to analyze. If we admit dependence between the steps of the sum, then we do not obtain general results holding for any kind of dependence. In Chapter 8 we introduce some preliminary studies that will help us to understand the motivation for the particular type of dependence we will analyze. Then we give detailed description of the framework where our results take place. Namely, we will concentrate to the case where the environment in which the motion takes place is responsible for the correlation between the steps.

Chapter 8

Dynamics on a Lévy random environment

In this Chapter we want to deepen what happens when we do not require independence between the steps of the random walk. There are many ways to introduce dependence and we will focus on the particular case of random walks in random environment (RWRE), where the environment plays a crucial role in the correlation of the steps. In particular we deal with the specific case where the environment is a Point Process on $\mathbb{R}$. The sake for this model is twofold:

- the analysis of a Random Walk whose steps are correlated, due to the presence of a random environment. This particular model falls in the large reign of RWRE which consists of a great literature that goes beyond our purposes. There are many left over problems, which we do not even mention. We simply look at RWRE as a source of inspiration.

- the motion we are going to describe offers a simple model for a process in inhomogeneous environment, and this simple hypothesis is able to explain its anomalous behavior. Hence we exhibit mathematically treatable features of a framework that leads to non diffusive regime.

The study of anomalous processes is far more recent with respect to processes with diffusive rescaling. We are therefore interested on these anomalous diffusive processes, where the variance of a moving particle has a super- or sub-linear growth in time.

In the physical literature, such anomalous behavior has been observed in many systems: Lorentz gases with infinite horizon, rotating flows, intermittent dynamical systems, etc. Several models have been put forth to describe such situations. These models need to balance the amount of ingredients that allow anomalous diffusion to arise, but they usually brea heavy drawbacks as their lack of mathematical tractability, which prevent the flourish of deep analysis. Among them, the simplest are the homogeneous random walks whose transition probabilities have an infinite second moment (and possibly an infinite first moment too) [GK68]. Especially in the physical literature, they are sometimes

dubbed Lévy flights. To define them properly, Lévy flights are discrete random walks with i.i.d steps whose distribution is an infinite variance stable law, i.e.

$$X_n := \sum_{i=1}^{n} \xi_i \qquad \xi_i \sim S_\alpha(\sigma, \beta, \mu) \tag{8.1}$$

It is a well known fact that Lévy flights easily break normal diffusion. However their defining feature is also their most serious limitation, meaning that the variance of the walker's position is infinite at all times, failing to reproduce the superlinear time dependence that is typical of many systems of interest, such as those mentioned earlier.

The analysis of anomalous diffusion proceed by considering other more realistic models, such as the Lévy walks, that can be thought as the continuous interpolation of Lévy walks. The jumps are still picked from a long-tailed distribution, as in the Lévy flights, but the walker needs a certain time to complete a jump; typically constant speed is assumed, hence the time required to perform the jump turns out to be proportional to the length of the jump [Cri+14; ZDK15]. The lack of mathematical structure in Lévy Walks is perhaps the main reason to explain their side role in the field of anomalous transport. Their naive model results unsatisfactory to depict the features that are responsible for anomalous diffusion. It seems natural that this particular rescaling should depend on the environment, whereas Lévy walks incorporate it as an intrinsic particle feature.
Hence, we came to the conclusion that our model of interest should heavily rely on the environment, the latter being responsible for anomalous diffusion. Indeed, in many real-world applications (in statistical physics, optics, epidemics, etc.), the long inertial segments, or ballistic flights, that cause superdiffusion are not due to an anomalous mechanism that governs the dynamics of the agent (a particle, a photon, an animal, etc.) but rather to the complexity of the medium in which the motion occurs. An example is molecular diffusion in porous media (see the reference lists of [Lev97; BFK00; BCV10] for more examples).

The importance of correctly modeling the environment where the motion takes place appears clear. In many interesting situations the walker moves in a complex and/or disordered environment and in this framework the theory of homogeneous random walk, though phenomenal for modeling transport in regular media, fails to represent the dynamics. In facts, spatial inhomogeneities induce strong correlations that can have a strong impact on the transport properties, which cannot be simulated by a simple homogenous model [Gra80; Bei82; BS83].

The centrality of this aspect led to the definition of a class of processes called random walks in random environment (RWRE), where the transition probabilities are themselves random functions of space (see [Zei06] for a review).
This rich class of walks is typically studied from two different viewpoints: that of the quenched processes, where one focuses on the dynamics for a typical fixed environment, and that of the annealed (or averaged) processes, where the interest is on the effect of averaging over the environments.

A primer example of RWRE is represented by random conductance models (RCM). Basically they can be represented as a random walk in $\mathbb{Z}^d$, where the conductance C_{xy}

can be interpreted as the "cost" to pay if we want to move from a certain point x to another specific point y. In details, let E_d be the set of non-oriented nearest neighbor bonds, $\{C_e\}_{e\in E_d}$ be a sequence of i.i.d. non-negative random variables, and $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space that governs the randomness of the conductance. For each $\omega \in \Omega$, let $\{X_n^\omega\}_{n\geq 0}$ be a discrete time random walk whose transition probabilities are given by $P_\omega(X_{n+1} = y \mid X_n = x) = C_{xy}/Z_x$, where $Z_x := \sum_{y\sim x} C_{xy}$, and $y \sim x$ if and only if $\{x, y\} \in E_d$. Conductance $\{\mu_x\}_{x\in\mathbb{Z}^d}$ plays a fundamental role when dealing to continuous time random walks (CTRW) associated to the discrete one described above. Depending on time parametrization, two natural CTRW arise. In both of them, the walker waits a proper holding time on the site before jumping instantaneously to the next site, according to the discrete random walk.

1. Constant speed random walk, where the holding time at time x is i.i.d. exponentially distributed with mean 1 for all $x \in \mathbb{Z}^d$;

2. Variable speed random walk, where the holding time at x is exponentially distributed with mean $1/Z_x$.

Random conductance models are an example of random walks where the environment in which we are moving plays a significant role. Without further dwelling on details of RCM, we want to highlight the fact that, when random walk is influenced by the environment, its steps can not be independent any more. There are many other ways to introduce dependence between the steps of a random walk. For example by letting the jump distribution depend in some way on the previous steps. Here we want to deal with a different case. The previous steps matters because they brought the walker on a particular position, but the jump distribution does not depend on the previous steps, rather on the particular position the walker is occupying. In other words, the lattice where the random walk is performed is not independent under space translations.

Recently two models [BFK00; Sch02], though rather different from one another, both proposed a system where standard dynamics are performed on an anomalous environment, which forces long jumps and is therefore responsible for the anomalous behavior. This new approach is motivated by many physical situation such as human mobility, epidemics or network routing. Even more, a surge of interest in this topic has lately come from the physics of materials, since a new glassy material has been devised, through which light exhibits anomalous properties that can be experimentally controlled [BBW08]. Here the complexity of the underlying network is responsible for anomalous diffusion. The starting point of our investigation is the Lévy-Lorentz Gas, which was firstly introduced in [BFK00] as a representative model for this kind of dynamics. In particular, the design of this so-called Lévy glass suggests an interpretation of the motion of light in it by way of a Lévy walk in a disordered environment, as studied in [BGA09; GAB12; BBV11; BCV10; Bur+12] (with varying degrees of approximation).

Taking enormous inspiration from this papers, where no rigorous proof is given, we expand their work deriving some analytical properties.
In the following we will then focus on a modified version of the Lévy-Lorentz gas, i.e. a specific case of RWRE where the random environment is given by a Point Process on the

real line. The idea consists on having a Point Process, whose realization determines the positions of the scattered targets, and then perform a Random Walk on these targets.

8.1 Point Processes on $\mathbb{R}$

A point process on a set $\mathcal{S}$ is a random element whose value is a countable subset of points $x \in \mathcal{S}$ that has no accumulation points.
In the following, we will take $\mathbb{R}$ as the reference measurable space $\mathcal{S}$, endowed with the usual Borel σ-algebra $\mathcal{B} = \mathcal{B}(\mathbb{R})$. The extension to higher dimensions is straightforward for some definitions, but the difficulty grows very quickly when proving some results. Since our goal is to define the framework of our research, we restrict this introduction to the case $d = 1$.

A point process on $\mathbb{R}$ is thus a collection of points that follow a specific pattern. Hence we need a description of the distribution of the points in the real line.
Consider a probability space $(\Omega_{en}, \mathcal{F}, P)$, and a mapping $p : \Omega_{en} \times \mathcal{B}(\mathbb{R}) \mapsto \mathbb{N}$ such that

1. For every $\omega \in \Omega_{en}$, $p(\omega, \cdot)$ is a locally finite measure on $\mathcal{B}(\mathbb{R})$;

2. For every $B \in \mathcal{B}(\mathbb{R})$, $p(\cdot, B)$ is a random variable over $\mathbb{N}$.

We can think of p as defining a mapping which maps $\omega \in \Omega_{en}$ into a measure $\mathcal{M}(\mathcal{B}(\mathbb{R}))$, where $\mathcal{M}(\mathcal{B}(\mathbb{R}))$ is the set of all locally finite measures on $\mathcal{B}(\mathbb{R})$. If we endow $\mathcal{M}(\mathcal{B}(\mathbb{R}))$ with the minimal σ-algebra so that all evaluation maps of the form $\pi_B : \mu \mapsto \mu(B)$, where $B \in \mathcal{B}(\mathbb{R})$ is relatively compact, are measurable. Then, p defined above is a random element and, for every $\omega \in \Omega_{en}$, p_ω is a locally finite measure over $\mathcal{B}(\mathbb{R})$.
This construction allows us to define an event of a point process as

$$p = \sum_{i=1}^{n} \delta_{X_i} \tag{8.2}$$

where δ denotes the Dirac measure, n is a random variable with support in $\mathbb{N} \cup +\{\infty\}$, and X_i are random elements of $\mathbb{R}$. Moreover, we can easily derive the following counting measure over any interval $(t_1, t_2]$

$$N(t_1, t_2) = \int_{t_1}^{t_2} p(t)dt. \tag{8.3}$$

8.1.1 Renewal Point Processes

We interrupt here the description of a general point process on $\mathbb{R}$. From the plenty of possible choices available, we restrict ourselves to a very intuitive one. In the following we consider the particular class of renewal point processes, that can be described by simply given the common distribution of the distance between neighboring points. To describe a generic renewal point process, we need first to introduce the counting processes.

Definition 20. A stochastic process $\{N(t), t \geq 0\}$ is said to be a counting process if $N(t)$ represents the total number of "events" that occur by time t.

Hence $N(t) \in \mathbb{N}$ and the trajectories of N are piecewise linear. From the previous definition, it also follows that a counting process $\{N(t), t \geq 0\}$ is nondecreasing, i.e. $N(s) \leq N(t)$ for every $s < t$. Moreover, the difference $N(t) - N(s)$ indicated the number of events that occur in the interval $(s, t]$. If the number of events that occur in disjoint intervals are independent, then the counting process is said to have independent increments, whereas if the distribution of the number of events that occur in any interval depends only on the length of the interval, then the counting process is said to possess stationary increments.
For example, the counting process $\{N(t), t \geq 0\}$ is said to be a Poisson process with rate $\lambda > 0$, if it satisfies the following properties

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

(ii) The process has independent increments.

(iii) For any $t \geq 0$, the number of events in any interval of length t is Poisson distributed with mean λT. That is, for all $s, t, \geq 0$,

$$\mathbb{P}[N(t+s) - N(s) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n \in \mathbb{N}. \tag{8.4}$$

Note that condition (iii) implies that a Poisson process has stationary increments and also that $\mathbb{E}[N(t)] = \lambda t$. Finally, observe that the times between successive events are independent and identically distributed exponential random variables. This brings us to the following Definition

Definition 21. Consider a sequence $(X_n)_{n\geq 1}$ of random variables taking values in $\mathbb{R}^+$. Let $\{N(t), t \geq 0\}$ be a counting process such that the time between the $(n-1)$st and the nth event is given by X_n. If the random variables $(X_n)_{n\geq 1}$ are independent and identically distributed, then he counting process $\{N(t), t \geq 0\}$ is said to be a renewal process.

For a renewal process with inter-arrival times given by $(X_n)_{n\geq 1}$, let

$$S_0 = 0, \qquad S_n = \sum_{i=1}^{n} X_i, \quad n \geq 1; \tag{8.5}$$

so that S_n denotes the time of the nth renewal. Using equation (8.5), $N(t)$ can be written as

$$N(t) = \max\{n : \, S_n \leq t\}. \tag{8.6}$$

Assume that $\mathbb{E}[X_i] = \mu$ for some $\mu > 0$. Then, by the strong law of large numbers it follows that, with probability 1,

$$\frac{S_n}{n} \to \mu \qquad \text{as } n \to \infty. \tag{8.7}$$

Hence S_n must be going to infinity as n goes to infinity. Thus, for any fixed $t > 0$, S_n can be less than or equal to t for at most a finite number of values of n and hence by equation (8.6) $N(t)$ must be finite. This is a direct confirmation that the renewal point process do not contain any accumulation point.
Note that we could obtain the same result even dropping the assumption of the mean of the X_n to be finite. Moreover it is possible to show that $N(t) \to \infty$ as $t \to \infty$, since the event $\{X_n = \infty\}$ has probability 0.

A more interesting question that arises is to determine the rate at which $N(t)$ goes to infinity. The answer is given by the next proposition.

Proposition 8.1.1. *Let $\{N(t), t \geq 0\}$ be a renewal point process with inter-arrival times give by the sequence $(X_n)_{n\geq 1}$. Assume $\mathbb{E}[X_n] = \mu$, with $\mu \in \mathbb{R}^+ \cup \infty$. Then, with probability 1,*

$$\frac{N(t)}{t} \to \frac{1}{\mu} \qquad \text{as } t \to \infty. \tag{8.8}$$

Proof. Let us consider the random variable $S_{N(t)}$, representing the time of last renewal prior to or at time t. Similarly $S_{N(t)+1}$ indicates the time of the first renewal after time t. We have the following chain of inequalities

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)} \tag{8.9}$$

However, from the strong law of large numbers, it follows that $S_{N(t)}/N(t) = \sum_{i=1}^{N(t)} X_i/N(t)$ is the average of $N(t)$ i.i.d. random variables, and hence $S_{N(t)}/N(t) \to \mu$ as $N(t) \to \infty$. But since $N(t) \to \infty$ when $t \to \infty$, we obtain

$$\frac{S_{N(t)}}{N(t)} \to \mu \qquad \text{as } t \to \infty. \tag{8.10}$$

It is not difficult to prove that also $S_{N(t)+1}/N(t) \to \infty$ as $t \to \infty$, and hence the result follows by equation (8.9). □

Note that if $\mu = \infty$, then $N(t)$ goes to infinity at a superlinear rate.
Until now we implicitly assume that $\mathbb{P}(X_n = 0) < 1$, otherwise we would obtain a collection of coinciding points of any practical interest. From now on, we will assume the previous distribution to have probability 0, i.e. we assume the support of the distribution for the distance between neighboring points to be $(0, +\infty)$, since we want to consider two coinciding points as one. As a historical remark, we point out that the first point processes had the real half line $[0, \infty)$ as their state space. In this situation, the support was usually interpreted as time, and the points represented the time when determined events occurred.
To conclude the list of constraints of the point process we are considering in this work, we assume that a point is always present at the origin. In this way, we are not analyzing a stationary point process.

Although for us the time interpretation fails, we are actually building our point process on $\mathbb{R}$ by concatenating two independent renewal processes on the positive and negative half real line, and by imposing a point to be always present at the origin.
The point process obtained in such way is not strictly stationary, i.e. for any $t > 0$ the distribution of $N(s+t) - N(s)$ is independent on s provided that the interval $(s, s+t]$ does not intersect the origin.

8.2 Random walk in a Lévy random medium

The expression *Lévy random medium* indicates a stochastic point process, in some space, where the distances between nearby points have heavy-tailed distributions. Processes of this kind have been receiving a surge of attention, of late, both in the physical and mathematical literature; cf., respectively, [BFK00; Sch02; BCV10; Bur+12; ZDK15] and [Bia+16; BLP20; MS18]. They model a variety of physical nature situations of scientific interest. In particular, they are used as supports for various kinds of random walks, in order to study phenomena of anomalous transport and anomalous diffusion. An incomplete list of general or recent references on this topic includes [SZF95; KRS08; Cri+14; ZDK15; Art+18; Rad+19; VBB19].

The random medium that we consider in this paper is perhaps the most natural choice for a Lévy random medium in the real line: a sequence of random points $\omega = (\omega_k,\, k \in \mathbb{Z})$, where $\omega_0 = 0$ and the nearest-neighbor distances $\zeta_k = \omega_{k+1} - \omega_k$ are i.i.d. variables in the normal domain of attraction of a β-stable variable, with $\beta \in (0,1) \cup (1,2)$. Here β is the index of the stable distribution, not the skewness parameter, which equals 1 because $\zeta_k > 0$.

A random walk $Y = (Y_n,\, n \in \mathbb{N})$ takes place on ω according to the following rule. Independently of ω, there exists a random walk $S = (S_n,\, n \in \mathbb{N})$ on $\mathbb{Z}$ with $S_0 = 0$ and i.i.d. increments in the normal domain of attraction of an α-stable variable, with $\alpha \in (0,1) \cup (1,2)$. We define $Y_n := \omega_{S_n}$. This means that the process Y performs the same jumps as S, but on the marked points ω_k instead of $\mathbb{Z}$. For example, if a realization of S is $(0, 3, -1, \ldots)$, the process Y starts at the origin of $\mathbb{R}$, then jumps to the third marked point to the right of 0, then to the first marked point to the left of 0, and so on. In other words, S *drives* the dynamics of Y on the Lévy medium. For this reason we call it the *underlying random walk*.

Our process of interest is Y. We may describe it as a *Lévy flight on a one-dimensional Lévy random medium*. This phrase is borrowed from the physical literature, where the term 'Lévy flight' usually indicates a discrete-time random walk with long-tailed instantaneous jumps. This is in contrast to a 'Lévy walk', which in general designates a *persistent*, continuous- or discrete-time, random walk with long-tailed trajectory segments that are run at constant finite speed [ZDK15]. A Lévy walk is often seen as an interpolation of a Lévy flight. For example, an important process from the standpoint of applications is $X := (X(t),\, t \in [0, +\infty))$, the unit-speed interpolation of Y. This means that, for any realization of Y, a trajectory of X starts at the origin and visits all the points Y_n in the given order, traveling between them with velocity 1 or -1, depending on Y_{n+1} being to

the right or to the left of Y_n, respectively.
The walk X is a generalization of a system that first appeared in the physical literature 2 decades ago with the name *Lévy-Lorentz gas* [BFK00] (more precisely, the Lévy-Lorentz gas is the case where the underlying random walk is simple).

There are several reasons to study our Lévy flight on random medium. The most self-serving, on the part of the present authors, is to build a basis to investigate the properties of the associated Lévy walk, as described above (see the proofs in [Bia+16; BLP20]). Also, Y can be thought of as the limit of a continuous-time random walk with resting times on the points ω_k, when the ratio between the speed of the walker and the typical resting time diverges. This can be used to model a variety of situations where a given dynamics is very fast compared to its "decision times", e.g., electronic signal on a network whose nodes act as relatively slow processing stations; human mobility (assuming, as it is often the case, that resting times are substantially longer than travel times); etc. This particular model aside, there is no lack of general motivation for the study of random walks on points processes, especially in light of the fact that the topic is regrettably less developed than others in the field of random walks, with the exception perhaps of random walks on percolation clusters *et similia*. For some interesting lines of research see, e.g., [BR15; CF+09; CFP13; Kub13; Zhu15; Rou14] and references therein. A recent paper which we extend with the present work is [MS18].

In this paper we give *annealed* limit theorems for Y in all cases $\alpha, \beta \in (0,1) \cup (1,2)$, identifying in each case both the scale n^γ whereby

$$\left(\frac{Y_{\lfloor nt \rfloor}}{n^\gamma} \,,\, t \in [0, +\infty) \right) \tag{8.11}$$

converges to a non-null limit, and the limit process. In all cases we prove the optimal, or at least morally optimal, functional limit theorem, meaning that we show distributional convergence of the process with respect to (w.r.t.) the strongest Skorokhod topology that applies there. There are cases in which there can be no convergence in the J_1 or M_1 topologies: in such cases we prove convergence w.r.t. J_2. When the limit process is not càdlàg (or càglàd) we show convergence of the finite-dimensional distributions. Finally, in the cases where the limit of (8.11) is deterministic, we prove a functional limit theorem for the corresponding fluctuations, again relative to the optimal topology.

8.2.1 Notation

In this Section we fix the notation holding also for Chapter 9, where we present the results of our work. Moreover, we mention here a couple of previous results that motivated and fueled our research.

Random medium. The random medium that we consider in this paper is perhaps the most natural choice for a Lévy random medium in the real line: a sequence of random points $\omega = (\omega_k,\, k \in \mathbb{Z})$, where $\omega_0 = 0$ and the nearest-neighbor distances $\zeta_k = \omega_{k+1} - \omega_k$ are i.i.d. variables in the normal domain of attraction of a β-stable variable, with $\beta \in$

$(0,1) \cup (1,2)$. We will denote a generic random variable with distribution ζ_i as ζ. Here β is the index of the stable distribution, not the skewness parameter, which equals 1 because $\zeta > 0$.

Consider now the sequence $(\zeta_i)_{i\in\mathbb{Z}\setminus 0}$ of i.i.d. positive random variables describing the inter-distances of our point process. In the case $\beta \in (0,1)$, this means that, as $n \to +\infty$,

$$\frac{1}{n^{1/\beta}} \sum_{i=1}^{n} \zeta_i \xrightarrow{\mathrm{d}} Z_1^{(\beta)}, \tag{8.12}$$

where $Z_1^{(\beta)}$ is a stable variable of index β and skewness parameter 1.
In the case $\beta \in (1,2)$ we have instead

$$\frac{1}{n^{1/\beta}} \sum_{i=1}^{n} (\zeta_i - \nu) \xrightarrow{\mathrm{d}} \tilde{Z}_1^{(\beta)}, \tag{8.13}$$

for a stable variable $\tilde{Z}_1^{(\beta)}$ of index β. In this case, necessarily, ν is the expectation of ζ. The random medium associated to $(\zeta_i)_{i\in\mathbb{Z}\setminus 0}$ is defined to be:

$$\omega_0 = 0, \qquad \omega_k = \begin{cases} \sum_{i=1}^{k} \zeta_i & \text{if } k > 0, \\ 0 & \text{if } k = 0, \\ -\sum_{i=k}^{-1} \zeta_i & \text{if } k < 0. \end{cases} \tag{8.14}$$

This determines a point process $\omega := (\omega_k,\, k \in \mathbb{Z})$ on $\mathbb{R}$ that we call *Lévy random medium* to emphasize the fact that the distribution of ζ has a heavy tail. Each point ω_k will be called a *target*. In other words, the distances between neighboring targets are drawn according to independent random variables ζ_i. We denote the set of all possible environments by Ω_{en}, and the law on it by P.

Underlying random walk. Independently of ω, there exists a $\mathbb{Z}$-valued random walk $S := (S_n)_{n\in\mathbb{N}}$ with $S_0 = 0$ and i.i.d. increments $\xi_i := S_i - S_{i-1}$ that are independent of ζ (and thus of ω). In other words, S is given by

$$S_0 = 0, \qquad S_n = \sum_{i=1}^{n} \xi_i \quad \text{for } n \in \mathbb{Z}^+. \tag{8.15}$$

We will refer to S as the *underlying* random walk and denote by Q the distribution on the space Ω_{RW} of all its realizations.

Random walk on the random medium. The *random walk on the random medium* $Y := (Y_n)_{n\in\mathbb{N}}$ takes place on ω according to the following rule:

$$Y_n := \omega_{S_n}, \quad n \in \mathbb{N}. \tag{8.16}$$

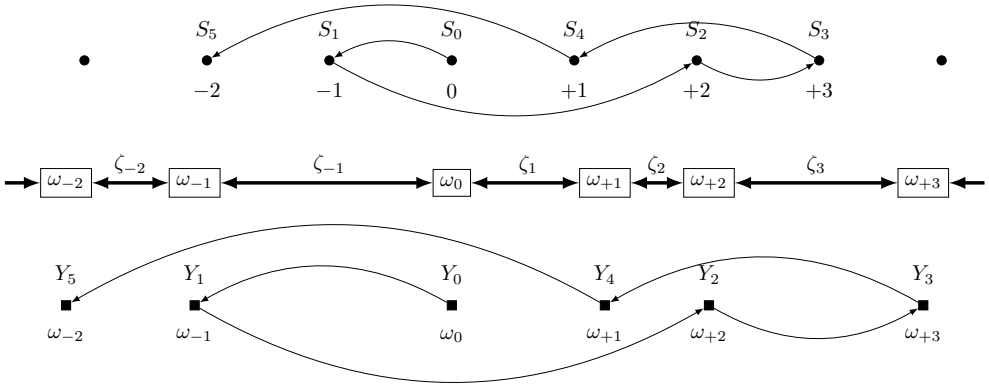

Figure 8.1: Top: A realization of the underlying random walk S on $\mathbb{Z}$. Middle: A realization of the random medium ω, with inter-distances given by ζ_i. Bottom: The corresponding process Y jumps between the targets ω according to the walk S.

In other words, Y performs the same jumps as S, but on the points of ω; see Figure 8.1 for a hands-on explanation. With reference to the example in Figure 8.1, if $S = (0, -1, 2, 3, 1, -2, ...)$ is a realization of the underlying random walk, the corresponding random walk on the point process will be $Y = (\omega_0, \omega_{-1}, \omega_2, \omega_3, \omega_1, \omega_{-2}, \ldots)$.

In the following we will focus on a particular case of underlying random walk, where the i.i.d increments $(\xi_i, i \in \mathbb{N})$ belongs to the normal domain of attraction of an α-stable variable, with $\alpha \in (0, 1) \cup (1, 2)$. Our main results consist on the derivation of the asymptotic law of Y, under suitable scaling, with respect to the probability measure $\mathbb{P} := P \times Q$ governing the entire system (medium and dynamics). This is sometimes referred to as the *annealed* or *averaged* law of Y. This and other results are the core of Chapter 9.

To conclude this Section, we want to analyze a couple of previous results that inspired our work. Indeed, we generalized the results contained in [Bia+16] and [BLP20], where they considered a modified version of the Lévy-Lorentz Gas, namely a Random Walk on a Point Process on the real line, where the distance between targets was given by a stable distribution of index β.

8.3 Previous Results

Inspired by the Lévy-Lorentz Gas, in [Bia+16] the system under examination is a random walk on a Point Process. A random array of points, called targets, were given on the real line, such that the distances between two neighboring targets are i.i.d. with finite mean; they are, however, allowed to have infinite variance, which is the reason why the game gained interest. This long-tailed choice includes all random variables with distribution in the basin of attraction of a β-stable distribution, when the stable index β falls in the

range $(1, 2)$. A particle moves with unit speed between the targets, driven by a random walk that is independent of all the rest. The underlying random walk S was assumed to be symmetric and to have finite second moment.
More in detail, they assumed the origin to be always a target and the starting point of the motion at initial time $t = 0$. According to the sequence of steps $\xi_1, \xi_2, \xi_3, ...$ of the underlying random walk, the particle, after reaching the ξ_i-th target, starts to travel towards the ξ_{i+1}-th target with unit speed. This is therefore a continuous-time random walk $(X(t), t \geq 0)$, whose trajectories have long inertial segments due to a random environment, which is why they speak of a random walk in a Lévy random environment.
The results in [Bia+16] regards the continuous-time process X and the discrete-time random walk on the point process Y. As first, the quenched random walk on the point process was studied, proving a quenched CLT and the convergence of all the accordingly rescaled moments. For the latter discrete-time process Y, a quenched CLT and the convergence of all the normally rescaled moments to those of a suitable Gaussian were proven. Then the attention was directed to the continuous-time motion, the constant-speed interpolation of a symmetric random walk on the marked points, leading, by comparison, to the derivation of the quenched CLT for the continuous-time process. These results imply the annealed normal CLT for both the continuous- and discrete-time walks.
Hence, for the regime $\beta \in (1, 2)$ of the β-stable Point Process where we move, for a.e. realization $\omega \in \Omega_{en}$ of the medium, the continuous-time process $t^{-1/2}X^{\omega}(t)$ converges, as $t \to \infty$, to a Gaussian variable independent of ω. This confirmed and added on some of the predictions of [BCV10]for the Lévy-Lorentz gas.

In this second article [BLP20], the authors considered a similar setting to the one analyzed above. Again, the process under observation is a Random Walk on a random medium on the real line. The random medium is given by a marked point process in which the distances between consecutive points are i.i.d. variables taken in the basin of attraction of a β-stable distribution, with $\beta \in (0, 1)$. The underlying random walk is assumed to be have finite moment of order $\gamma > 2/\beta$, hence it has at list finite variance. As before, the objects of interests are the discrete time random walk Y of the point process, and its constant speed interpolation X.
Since the distances between targets are fat-tailed variables (their first moment is infinite), the trajectories of the process occasionally experience extremely long inertial segments, leading one to believe that the process is superdiffusive. By that we mean that it scales like a power of time with exponent bigger than $1/2$. The purpose of the paper is to show this in a very specific sense. More precisely, if $X(t) \equiv X^{\omega}(t)$ denotes the continuous-time process, with the label ω representing the random medium, they proved that the finite-dimensional distributions of $(n^{-1/(\alpha+1)}X(nt), t \geq 0)$, w.r.t. both the random medium and the random dynamics, converge to those of a certain process which they explicitly identify (Theorem 2.1). In other words, an annealed generalized CLT for the finite-dimensional distributions of $(X(t), t \geq 0)$ is proven.

The importance of [BLP20] lies in the fact that superdiffusion for the limit process was explicitly shown, representing a primer for random walks in Lévy random media.

The obtained scaling is also in agreement with the one that has been identified recently for a related model [Art+18].

Chapter 9 is completely devoted to the presentation of our results. We exploit the powerful framework of RWRE, designing a Lévy-Lorentz Gas type of motion, where not only the random medium is set according to a stable point process as in [Bia+16; BLP20], but also the jump distribution of the underlying random walk consists of heavy tailed stable Lévy steps.

Chapter 9

Limit theorems for Lévy flights on a Lévy random medium

In this Chapter we want to present our results, contained in [Sti+20]. The general framework where we move is a generalization of the one described in Section 8.2. Our journey started in Chapter 7 where we illustrated some basic feature of random walks and more general processes with i.i.d. increments. Then, in Chapter 8 we introduced dependence and in particular we stick to the case in which dependence is a direct consequence of the fact that the medium where the process lives is not homogeneous. Taking inspiration from the Lévy-Lorentz gas, we analyze the case of random walk on a point process in the real line. The marked point process consists of scattered points whose distance between neighboring targets is drawn from a β-stable distribution. Firstly, the case with $\beta \in (1, 2)$, i.e. with infinite variance but finite mean, was studied [Bia+16], then the results were extended to the case $\beta \in (0, 1)$, i.e. the distribution of the distance between neighboring targets has infinite mean [BLP20].
Not much work has been done on systems that combine long-tailed jumps and disordered media. To the authors' knowledge, the first such examples are the Lévy flights perturbed by random drift fields introduced in [Fog94]. In this case the cause of the anomalous diffusion is the distribution of the jumps. The only examples of long-tailed random walks in random environment these authors have found in the rigorous mathematical literature are the long-range walks on point processes studied in [BR15; CF+09; CFG+09; Rou15].

Now we are ready to combine the effects of heavy-tailed distribution of the jumps of the underlying random walk, with the heavy-tailed distribution of the distance between neighboring targets of the marked point process where the motion takes place. In particular, for the random environment we keep the same choice of β-stable distribution adopted in Section 8.2, whilst adding the hypothesis that the underlying random walk are drawn from a stable distribution of index $\alpha \in (0, 1) \cup (1, 2)$. This combination of heavy-tailed stable distributions open further scenarios for the limit process we want to analyze.

9.1 The Model

9.1.1 Setup

As mentioned in the previous Chapter, the Lévy flight on random medium that we consider is a random walk performed over the points of a certain random point process. We proceed to define all the necessary constructions.

Random medium. Let $\zeta := (\zeta_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. positive random variables. We assume that the law of ζ belongs to the normal basin of attraction of a β-stable distribution, with $\beta \in (0,1)\cup(1,2)$. We refer to Section 8.2 for a complete treatment of the random environment.

We recall that the associated point process $\omega := (\omega_k,\, k\in\mathbb{Z})$ on $\mathbb{R}$, that we call *Lévy random medium* to emphasize the fact that the distribution of ζ has a heavy tail, is determined by the following relation:

$$\omega_0 = 0, \qquad \omega_k = \begin{cases} \sum_{i=1}^{k} \zeta_i & \text{if } k>0,\\ 0 & \text{if } k=0,\\ -\sum_{i=k}^{-1} \zeta_i & \text{if } k<0. \end{cases} \tag{9.1}$$

Underlying random walk. We consider a $\mathbb{Z}$-valued random walk $S := (S_n)_{n\in\mathbb{N}}$, with $S_0 = 0$ and i.i.d. increments $\xi_i := S_i - S_{i-1}$ that are independent of ζ (and thus of ω). We will denote a generic random variable with distribution ξ_i as ξ. In other words, S is given by

$$S_0 = 0, \qquad S_n = \sum_{i=1}^{n} \xi_i \quad \text{for } n\in\mathbb{Z}^+. \tag{9.2}$$

The law of ξ belongs to the normal basin of attraction of an α-stable distribution, with $\alpha\in(0,1)\cup(1,2)$. This is the main difference of our framework with respect to those of Section 8.3. Analogous convergences to those given in (8.12) and (8.13) apply to the ξ_i. Hence, the *underlying* random walk S obeys the following stable limits. In the case $\alpha\in(0,1)$, we have that, as $n\to+\infty$,

$$\frac{1}{n^{1/\alpha}} \sum_{i=1}^{n} \xi_i \xrightarrow{\mathrm{d}} W_1^{(\alpha)}, \tag{9.3}$$

where $W_1^{(\alpha)}$ is a stable variable of index α. In the case $\alpha\in(1,2)$ we have instead

$$\frac{1}{n^{1/\alpha}} \sum_{i=1}^{n} (\xi_i - \mu) \xrightarrow{\mathrm{d}} \widetilde{W}_1^{(\alpha)}, \tag{9.4}$$

for another stable variable $\widetilde{W}_1^{(\alpha)}$ of index α. In this case, necessarily, μ is the expectation of ξ_i and the skewness parameter of $\widetilde{W}_1^{(\alpha)}$ is 0.

Note that, when the mean of the steps μ equals zero, the centering of the previous equation seems redundant. However, when $\mu \neq 0$, the non-centered sum leads to a LLN type of limit, with rescaling $1/n$ and almost sure convergence to the mean μ itself (see Section 7.2).

Random walk on the random medium. As in the previous Chapter, the *random walk on the random medium* $Y := (Y_n,\, n \in \mathbb{N})$ is defined to be:

$$Y_n := \omega_{S_n}, \quad n \in \mathbb{N}. \tag{9.5}$$

In the following we will focus on the derivation of the asymptotic law of Y, under suitable scaling, with respect to the probability measure $\mathbb{P}$ governing the entire system (medium and dynamics). This is sometimes referred to as the *annealed* or *averaged* law of Y.

9.1.2 Limit processes for ω and S

We now recall some elementary functional limit theorems for suitable rescalings of the processes ω and S, cf. (9.1) and (9.2).
By definition, for all $k \in \mathbb{Z}$, ω_k is a sum of $|k|$ i.i.d. random variables ζ_i in the normal domain of attraction of a β-stable distribution. We first deal with the case $\beta \in (0,1)$. For every $s \in \mathbb{R}$ we define

$$\hat{\omega}^{(n)}(s) := \begin{cases} \dfrac{\omega_{\lfloor ns \rfloor}}{n^{1/\beta}} & \text{if } s \geq 0, \\ \dfrac{\omega_{\lceil ns \rceil}}{n^{1/\beta}} & \text{if } s < 0. \end{cases} \tag{9.6}$$

Let $(Z_{\pm}^{(\beta)}(s),\, s \geq 0)$ be two i.i.d. càdlàg Lévy β-stable processes such that $Z_{\pm}^{(\beta)}(0) = 0$ and $Z_{\pm}^{(\beta)}(1)$ is distributed like $Z_1^{(\beta)}$, as introduced in (8.12) (these two conditions uniquely determine the common distribution of the processes). Set

$$Z^{(\beta)}(s) := \begin{cases} Z_+^{(\beta)}(s) & \text{if } s \geq 0, \\ -Z_-^{(\beta)}(-s) & \text{if } s < 0. \end{cases} \tag{9.7}$$

Then (see, e.g., [Whi02, Section 4.5.3]), as $n \to \infty$,

$$\hat{\omega}^{(n)} \xrightarrow{\mathrm{d}} Z^{(\beta)} \qquad \text{in } (\mathcal{D}, J_1). \tag{9.8}$$

When $\beta \in (1,2)$, the average distance $\nu := \mathbb{E}[\zeta_i]$ between successive targets is finite and positive by assumptions. So, at first order, a Strong Law of Large Numbers holds. More in detail, setting

$$\bar{\omega}^{(n)}(s) := \begin{cases} \dfrac{\omega_{\lfloor ns \rfloor}}{n} & \text{if } s \geq 0, \\ \dfrac{\omega_{\lceil ns \rceil}}{n} & \text{if } s < 0, \end{cases} \tag{9.9}$$

we have

$$\bar{\omega}^{(n)} \xrightarrow{\text{a.s.}} \nu\,\mathrm{id} \qquad \text{in } (\mathcal{D}, J_1), \tag{9.10}$$

as $n \to \infty$. Here and in the rest of the paper id denotes the identity function, on whatever domain it is defined. Furthermore, a functional convergence similar to (9.8) holds for the fluctuations around this Law of Large Numbers. More explicitly, for $s \in \mathbb{R}$, define

$$\tilde{\omega}^{(n)}(s) := \begin{cases} \dfrac{\sum_{i=1}^{\lfloor ns \rfloor}(\zeta_i - \nu)}{n^{1/\beta}} & \text{if } s \geq 0, \\ \dfrac{-\sum_{i=\lceil ns \rceil}^{-1}(\zeta_i - \nu)}{n^{1/\beta}} & \text{if } s < 0. \end{cases} \tag{9.11}$$

Then, as $n \to \infty$,

$$\tilde{\omega}^{(n)} \xrightarrow{\text{d}} \tilde{Z}^{(\beta)} \qquad \text{in } (\mathcal{D}, J_1), \tag{9.12}$$

where the process $\tilde{Z}^{(\beta)}$ is defined similarly to $Z^{(\beta)}$, cf. (9.7), but with $\tilde{Z}^{(\beta)}_{\pm}(1)$ distributed like $\tilde{Z}^{(\beta)}_1$, introduced in (8.13).

Analogous limit theorems hold for the continuous-time rescaled versions of the underlying random walk S. By definition, S_n is a sum of n i.i.d. random variables ξ_i in the normal domain of attraction of an α-stable distribution. We distinguish two regimes, depending on the values of α and $\mu := \mathbb{E}[\xi_i]$ (when applicable).

The first regime corresponds to the cases $\alpha \in (0,1)$, or $\alpha \in (1,2)$ with $\mu = 0$. In these situations, the drift of the underlying random walk is either undefined or null. In either case, it does not affect the convergence of the process

$$\hat{S}^{(n)}(t) := \frac{S_{\lfloor nt \rfloor}}{n^{1/\alpha}}, \tag{9.13}$$

which we define for $t \geq 0$. In fact, let $W^{(\alpha)}$ denote a Lévy α-stable process with $W^{(\alpha)}(0) = 0$ and $W^{(\alpha)}(1)$ distributed like $W^{(\alpha)}_1$, where the latter has been defined after (9.3). Then, as $n \to \infty$,

$$\hat{S}^{(n)} \xrightarrow{\text{d}} W^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1). \tag{9.14}$$

When $\alpha \in (1,2)$ and $\mu \neq 0$, set, for $t \geq 0$,

$$\bar{S}^{(n)}(t) := \frac{S_{\lfloor nt \rfloor}}{n}. \tag{9.15}$$

By the functional version of the Strong Law of Large Numbers,

$$\bar{S}^{(n)} \xrightarrow{\text{a.s.}} \mu\,\mathrm{id} \qquad \text{in } (\mathcal{D}^+, J_1), \tag{9.16}$$

as $n \to \infty$. As for the fluctuations, defining

$$\tilde{S}^{(n)}(t) := \frac{\sum_{i=1}^{\lfloor nt \rfloor}(\xi_i - \mu)}{n^{1/\alpha}}, \tag{9.17}$$

we get

$$\tilde{S}^{(n)} \xrightarrow{\mathrm{d}} \widetilde{W}^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1). \tag{9.18}$$

where $\widetilde{W}^{(\alpha)}$ is a Lévy α-stable process with $\widetilde{W}^{(\alpha)}(0) = 0$ and $\widetilde{W}^{(\alpha)}(1)$ distributed like $\widetilde{W}_1^{(\alpha)}$ (defined as in (9.4)).

9.2 Results

We now present our convergence results for the Lévy flight Y which, as we shall see, strongly depend on the values of α and β. All theorems are stated using the notation established in the previous section.

We first analyze the case $\beta \in (0,1)$, corresponding to an infinite expected distance between the targets of the random medium

Theorem 9.2.1. *Let $\beta \in (0,1)$ and assume that either $\alpha \in (0,1)$ or $\alpha \in (1,2)$ with $\mu = 0$. For $t \in \mathbb{R}^+$ define*

$$\hat{Y}^{(n)}(t) := \hat{\omega}^{(n)} \circ \hat{S}^{(n)}(t) = \frac{Y_{\lfloor nt \rfloor}}{n^{1/\alpha\beta}}, \tag{9.19}$$

where $\hat{\omega}^{(n)}$ and $\hat{S}^{(n)}$ have been introduced, respectively, in (9.6) and (9.13). Then the finitedimensional distributions of $\hat{Y}^{(n)}$ converge to those of $Z^{(\beta)} \circ W^{(\alpha)}$, i.e., for any $m \in \mathbb{Z}^+$ and $t_1, \ldots, t_m \in \mathbb{R}^+$,

$$\left(\hat{Y}^{(n)}(t_1), \ldots, \hat{Y}^{(n)}(t_m)\right) \xrightarrow{d} \left(Z^{(\beta)}(W^{(\alpha)}(t_1)), \ldots, Z^{(\beta)}(W^{(\alpha)}(t_m))\right), \tag{9.20}$$

as $n \to \infty$.

Theorem 9.2.1 is rather weak, in that it only proves convergence of the finite-dimensional distributions of the process $\hat{Y}^{(n)}$ defined in (9.19). Observe, however, that the limit process $Z^{(\beta)} \circ W^{(\alpha)}$ has trajectories that are not càdlàg with positive probability (see for example the explanation around (2.9) of [BLP20]). Therefore, a functional limit theorem w.r.t. a Skorokhod topology is not the natural result to expect. On the other hand, when $\alpha \in (1,2)$ and $\mu \neq 0$, the assertion can be strengthened as follows.

Theorem 9.2.2. *Let $\beta \in (0,1)$ and $\alpha \in (1,2)$ with $\mu \neq 0$. For $t \in \mathbb{R}^+$ define*

$$\hat{Y}^{(n)}(t) := \hat{\omega}^{(n)} \circ \bar{S}^{(n)}(t) = \frac{Y_{\lfloor nt \rfloor}}{n^{1/\beta}}, \tag{9.21}$$

cf. (9.6) and (9.15). Then, as $n \to \infty$,

$$\hat{Y}^{(n)} \xrightarrow{d} \operatorname{sgn}(\mu)\, |\mu|^{1/\beta}\, Z_+^{(\beta)} \qquad \text{in } (\mathcal{D}^+, J_2). \tag{9.22}$$

Remark 8. Since $Z_+^{(\beta)} \overset{\mathrm{d}}{=} Z_-^{(\beta)}$, one could put either process in the r.h.s. of (9.22), irrespectively of the sign of μ.

Remark 9. The convergence (9.22) fails in the topology J_1, or even M_1 [Whi02, Section 3.3]. The topology J_2 is thus the strongest among the classical Skorokhod topologies with respect to which the convergence holds.

To justify the claim, observe that, in general, $\bar{S}^{(n)}$ is a wildly oscillating function around $\mu\mathrm{id}$, and $Z^{(\beta)}$ is almost surely discontinuous. More in detail, assume that $\mu > 0$ and let $s \in \mathbb{R}$ be a discontinuity point of $Z_+^{(\beta)}$ with a jump, say, of order 1 in n. Since, for $n \to \infty$, $\hat{\omega}^{(n)}$ is very close to $Z_+^{(\beta)}$ in J_1, there exists a discontinuity point s_n of $\hat{\omega}^{(n)}$, very close to s, with a jump of order 1. Now, if we exclude the case where the underlying random walk S is deterministic, $\bar{S}^{(n)}(t)$ is a non-monotonic function of $t \in I$, for every interval $I \subset \mathbb{R}^+$ and n large enough, depending on I (this is an elementary Brownian-bridge result). So one can find a small interval I such that, as t runs through I, $\bar{S}^{(n)}(t)$ oscillates many times around s_n. Therefore $\hat{\omega}^{(n)} \circ \bar{S}^{(n)}(t)$ has many back-and-forth jumps of order 1. This prevents convergence both in J_1 and in M_1, cf. [Whi02, Figure 11.2]. What allows for J_2-convergence is that the fluctuations of $\bar{S}^{(n)}$ around $\mu\mathrm{id}$ vanish, as $n \to \infty$. This means that the oscillations of $\bar{S}^{(n)}(t)$ around s_n, and therefore the large oscillations of $\hat{\omega}^{(n)} \circ \bar{S}^{(n)}(t)$, occur only in a vanishing interval $I_n \subset I$. Therefore one can find a non-continuous change of the coordinate t, say $\rho_n : [0, T) \longrightarrow [0, T)$, which is globally close to the identity and "reorders" the points in I_n in the sense that $\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n$ only has one jump of order 1. The problem thus reduces to the much easier problem of showing the J_1-convergence of the latter process. See the proof of Theorem 9.2.2 for the rigorous arguments. Lastly, we observe that all the results presented in this paper involving the J_2 topology could in fact be stated for a stronger Skorokhod -type topology. We refer the interested reader to Remark 7 in Section 6.7.

Next we consider the case $\beta \in (1, 2)$, where the inter-distances of the random medium have finite mean.

Theorem 9.2.3. *Let $\beta \in (1, 2)$ and recall the notation* (9.9), (9.13) *and* (9.15).

1. *Assume $\alpha \in (0, 1)$, or $\alpha \in (1, 2)$ with $\mu = 0$. For $t \in \mathbb{R}^+$ set*

$$\hat{Y}^{(n)}(t) := \bar{\omega}^{(n)} \circ \hat{S}^{(n)}(t) = \frac{Y_{\lfloor nt \rfloor}}{n^{1/\alpha}}. \tag{9.23}$$

Then, as $n \to \infty$,

$$\hat{Y}^{(n)} \xrightarrow{d} \nu W^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1). \tag{9.24}$$

2. *Assume $\alpha \in (1, 2)$ and $\mu \neq 0$. Setting*

$$\bar{Y}^{(n)}(t) := \bar{\omega}^{(n)} \circ \bar{S}^{(n)}(t) = \frac{Y_{\lfloor nt \rfloor}}{n} \tag{9.25}$$

one has

$$\bar{Y}^{(n)} \xrightarrow{d} \nu\mu\mathrm{id} \qquad \text{in } (\mathcal{D}^+, J_1). \tag{9.26}$$

As stated in point *2.* above, when $\alpha \in (1, 2)$ and $\mu \neq 0$, the sequence of processes $\bar{Y}^{(n)}$ converges to a multiple of the identity function. The next theorem gives the explicit asymptotics of the fluctuations of $\bar{Y}^{(n)}$ around its deterministic limit.

Theorem 9.2.4. *Let $\alpha, \beta \in (1,2)$ with $\mu \neq 0$, and let $\bar{Y}^{(n)}$ be the process defined in* (9.25).

1. *If $\alpha < \beta$ define*

$$\tilde{Y}^{(n)}(t) := \frac{n(\bar{Y}^{(n)}(t) - \nu\mu t)}{n^{1/\alpha}}. \tag{9.27}$$

Then, when $n \to \infty$,

$$\tilde{Y}^{(n)} \xrightarrow{d} \nu\widetilde{W}^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1), \tag{9.28}$$

where $\widetilde{W}^{(\alpha)}$ has been defined after (9.18).

2. *If $\alpha > \beta$ define*

$$\tilde{Y}^{(n)}(t) := \frac{n(\bar{Y}^{(n)}(t) - \nu\mu t)}{n^{1/\beta}}. \tag{9.29}$$

Then, when $n \to \infty$,

$$\tilde{Y}^{(n)} \xrightarrow{d} \operatorname{sgn}(\mu)\,|\mu|^{1/\beta}\,\tilde{Z}_+^{(\beta)} \qquad \text{in } (\mathcal{D}^+, J_2), \tag{9.30}$$

where $\tilde{Z}_+^{(\beta)}$ has been defined after (9.12).

3. *If $\alpha = \beta$ define*

$$\tilde{Y}^{(n)}(t) := \frac{n(\bar{Y}^{(n)}(t) - \nu\mu t)}{n^{1/\alpha}}. \tag{9.31}$$

Let $\tilde{Z}_+^{(\alpha)}$ and $\widetilde{W}^{(\alpha)}$ be two independent α-stable processes, as previously defined. As $n \to \infty$,

$$\tilde{Y}^{(n)} \xrightarrow{d} \operatorname{sgn}(\mu)\,|\mu|^{1/\beta}\,\tilde{Z}_+^{(\alpha)} + \nu\widetilde{W}^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_2). \tag{9.32}$$

Remark 10. The same considerations as in Remark 9 apply to the optimality of the J_2 topology in the limits (9.30) and (9.32).

9.3 Proofs

9.3.1 Proof of Theorem 9.2.1: Convergence of finite-dimensional distributions

We establish the assertion by extending the proof of [BLP20, Theorem 2.2]. We first prove the following:

Lemma 9.3.1. *Let $\hat{\omega}^{(n)}$ and $\hat{S}^{(n)}$ be the processes defined in* (9.6) *and* (9.13)*, respectively. Then, when $n \to \infty$,*

$$(\hat{\omega}^{(n)}, \hat{S}^{(n)}) \xrightarrow{d} (Z^{(\beta)}, W^{(\alpha)}) \qquad \text{in } (\mathcal{D} \times \mathcal{D}^+,\, J_1 \otimes J_1)\,, \tag{9.33}$$

where $J_1 \otimes J_1$ denotes the product topology on the product space $\mathcal{D} \times \mathcal{D}^+$.

Proof. From (9.8) and (9.14) we have that $\hat{\omega}^{(n)} \xrightarrow{\mathrm{d}} Z^{(\beta)}$ in $(\mathcal{D}, J_1)$ and $\hat{S}^{(n)} \xrightarrow{\mathrm{d}} W^{(\alpha)}$ in $(\mathcal{D}^+, J_1)$. Since $\hat{\omega}^{(n)}$ and $\hat{S}^{(n)}$ are independent, the result follows from [Whi02, Theorem 11.4.4]. □

By virtue of the Skorokhod Representation Theorem (see Theorem 5.2.1), we may assume that the convergence in the statement of Lemma 9.3.1 holds almost everywhere. If this is not the case, there exists a probability space where it does, and since the specifics of the probability space are irrelevant for the next discussion, we avoid here to change the notation for the processes in the new space. Notice also that since $Z^{(\beta)}$ is a β-stable process, it is almost surely continuous at s, for any $s \in \mathbb{R}$, and similarly $W^{(\alpha)}$ is almost surely continuous at t, for any $t \in \mathbb{R}^+$. In particular, by the independence of the two processes, the event that $W^{(\alpha)}$ is continuous at t and $Z^{(\beta)}$ is continuous at $W^{(\alpha)}(t)$ has probability 1, for any $t \in \mathbb{R}^+$. Therefore the hypotheses of the next lemma hold almost surely.

Lemma 9.3.2. *Fix $t > 0$ and consider a realization (ω, S) of the random medium and of the underlying random walk such that $W^{(\alpha)}$ is continuous in t and $Z^{(\beta)}$ is continuous at $W^{(\alpha)}(t)$. Then we have*

$$\lim_{n\to\infty} \hat{\omega}^{(n)}(\hat{S}^{(n)}(t)) = Z^{(\beta)}(W^{(\alpha)}(t)). \tag{9.34}$$

Proof. Let $\varepsilon \in (0,1)$ and $\eta \in (0, \varepsilon)$ be such that

$$\sup_{s:|s-W^{(\alpha)}(t)|<2\eta} |Z^{(\beta)}(W^{(\alpha)}(t)) - Z^{(\beta)}(s)| < \frac{\varepsilon}{2}. \tag{9.35}$$

Also choose $\varsigma \in (0, \eta)$ so that

$$\sup_{u:|u-t|<\varsigma} |W^{(\alpha)}(t) - W^{(\alpha)}(u)| < \frac{\eta}{2}. \tag{9.36}$$

Let n be large enough so that $d_{J_1,[0,t+1]}(\hat{S}^{(n)}, W^{(\alpha)}) < \varsigma/2$, see (6.11). In other words, there exists an increasing homeomorphism φ_n of $[0, t+1]$ such that, for all $u \in [0, t+1]$,

$$|u - \varphi_n(u)| < \frac{\varsigma}{2}, \tag{9.37}$$

$$|\hat{S}^{(n)}(u) - W^{(\alpha)}(\varphi_n(u))| < \frac{\varsigma}{2}. \tag{9.38}$$

Hence, using (9.38) and (9.36) we get

$$\begin{aligned} &|\hat{S}^{(n)}(t) - W^{(\alpha)}(t)| \\ &\quad \le |\hat{S}^{(n)}(t) - W^{(\alpha)}(\varphi_n(t))| + |W^{(\alpha)}(\varphi_n(t)) - W^{(\alpha)}(t)| \\ &\quad < \frac{\varsigma}{2} + \frac{\eta}{2} < \eta, \end{aligned} \tag{9.39}$$

since $|\varphi_n(t) - t| < \varsigma/2$. Assume moreover that n is large enough so that

$$d_{J_1,[0,|W^{(\alpha)}(t)|+1]}\left(\hat{\omega}^{(n)}, Z^{(\beta)}\right) < \frac{\eta}{2}, \tag{9.40}$$

$$d_{J_1,[0,|W^{(\alpha)}(t)|+1]}\left(\hat{\omega}^{(n)}(-\cdot), Z^{(\beta)}(-\cdot)\right) < \frac{\eta}{2}, \tag{9.41}$$

where the notation in the l.h.s. of (9.41) was introduced in Remark 6. Then there exists an increasing homeomorphism ψ_n of $[-|W^{(\alpha)}(t)|-1, |W^{(\alpha)}(t)|+1]$, with $\psi_n(0) = 0$, such that, for all $s \in [-|W^{(\alpha)}(t)| - 1, |W^{(\alpha)}(t)| + 1]$,

$$|s - \psi_n(s)| < \frac{\eta}{2}, \tag{9.42}$$

$$|\hat{\omega}^{(n)}(s) - Z^{(\beta)}(\psi_n(s))| < \frac{\eta}{2}. \tag{9.43}$$

Note also that (9.39) ensures that $\hat{S}^{(n)}(t) \in [-|W^{(\alpha)}(t)| - 1, |W^{(\alpha)}(t)| + 1]$, so that by (9.42) and (9.39),

$$\begin{aligned} |\psi_n(\hat{S}^{(n)}(t)) - W^{(\alpha)}(t)| &\le |\psi_n(\hat{S}^{(n)}(t)) - \hat{S}^{(n)}(t)| + |\hat{S}^{(n)}(t) - W^{(\alpha)}(t)| \\ &< \frac{\eta}{2} + \eta < 2\eta\,, \end{aligned} \tag{9.44}$$

and from (9.43) we obtain

$$|\hat{\omega}^{(n)}(\hat{S}^{(n)}(t)) - Z^{(\beta)}(\psi_n(\hat{S}^{(n)}(t)))| < \eta/2. \tag{9.45}$$

Finally, using (9.45), (9.35) and (9.44), we obtain:

$$\begin{aligned} &|\hat{\omega}^{(n)}(\hat{S}^{(n)}(t)) - Z^{(\beta)}(W^{(\alpha)}(t))| \\ &\quad \le |\hat{\omega}^{(n)}(\hat{S}^{(n)}(t)) - Z^{(\beta)}(\psi_n(\hat{S}^{(n)}(t)))| + |Z^{(\beta)}(\psi_n(\hat{S}^{(n)}(t))) - Z^{(\beta)}(W^{(\alpha)}(t))| \\ &\quad \le \frac{\eta}{2} + \frac{\varepsilon}{2} < \varepsilon. \end{aligned} \tag{9.46}$$

This shows (9.34). □

Proof of Theorem 9.2.1. Let $m \in \mathbb{N}^+$ and $t_1, \ldots, t_m \in [0, +\infty)$. With probability one $W^{(\alpha)}$ is continuous at $t_1, \ldots, t_m$ and $Z^{(\beta)}$ is continuous at $W^{(\alpha)}(t_1), \ldots, W^{(\alpha)}(t_m)$. When restricting to such realizations, using Lemma 9.3.2 with $t = t_i$, we have that $n^{-1/\alpha\beta}\, Y_{\lfloor n\, t_i \rfloor} = \hat{\omega}^{(n)}(\hat{S}^{(n)}(t_i))$ converges almost surely to $Z^{(\beta)}(W^{(\alpha)}(t_i))$, for all $i \in \{1, \ldots, m\}$. On the intersection of these events of probability one, the joint convergence for all $i \in \{1, \ldots, m\}$ holds. This implies the desired distributional convergence. □

9.3.2 Proof of Theorem 9.2.3: Limit theorems for $\beta \in (1,2)$

Although Theorem 9.2.3 was stated after Theorem 9.2.2, we give the proof of the former first, because it is simpler and somehow preliminary to the proof of the latter. As a matter of fact, we only prove assertion *1.* Assertion *2* is carried out similarly with no additional effort.

Lemma 9.3.3. *The composition map $h : \mathcal{D}_0 \times \mathcal{D}^+ \longrightarrow \mathcal{D}^+$ (see Chapter 6 for the definitions of $\mathcal{D}_0$ and $\mathcal{D}^+$) defined by*

$$h(w,s) := w \circ s \tag{9.47}$$

is measurable and it is J_1-continuous on $(\mathcal{C} \cap \mathcal{D}_0) \times \mathcal{D}^+$, where $\mathcal{C}$ is the space of continuous functions on $\mathbb{R}$. More precisely, the continuity is intended w.r.t. the topology $J_1 \otimes J_1$ on the domain of h and J_1 on its target space.

Proof of Lemma 9.3.3. As the measurability of h is easy, we concentrate on the continuity statement. Assume that, as $n \to \infty$, $(w_n, s_n) \to (w, s)$ in $(\mathcal{D}_0 \times \mathcal{D}^+, J_1 \otimes J_1)$, with $w \in \mathcal{C} \cap \mathcal{D}_0$ and $s \in \mathcal{D}^+$. This means that, for all $M, T > 0$,

$$w_n \to w \qquad \text{in } (\mathcal{D}([-M,M]), J_1), \tag{9.48}$$

$$s_n \to s \qquad \text{in } (\mathcal{D}([0,T]), J_1). \tag{9.49}$$

In particular, if we fix $T > 0$, there exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ of homeomorphisms of $[0,T]$ such that

$$\sup_{t\in[0,T]} |\lambda_n(t) - t| \to 0, \tag{9.50}$$

$$\sup_{t\in[0,T]} |s_n \circ \lambda_n(t) - s(t)| \to 0. \tag{9.51}$$

We have

$$\begin{aligned} &\sup_{t\in[0,T]} |w_n \circ s_n \circ \lambda_n(t) - w \circ s(t)| \qquad (9.52)\\ &\quad \le \sup_{t\in[0,T]} |w_n \circ s_n \circ \lambda_n(t) - w \circ s_n \circ \lambda_n(t)| + \sup_{t\in[0,T]} |w \circ s_n \circ \lambda_n(t) - w \circ s(t)| \\ &\quad \le \sup_{u\in[-M,M]} |w_n(u) - w(u)| + \sup_{t\in[0,T]} |w \circ s_n \circ \lambda_n(t) - w \circ s(t)|, \end{aligned}$$

where $M = M(T) := \sup_n \sup_{v\in[0,T]} |s_n(v)|$. This quantity is finite because $s \in \mathcal{D}^+$, and thus it is bounded on $[0,T]$, and from (9.49).

Now, the first of the last two terms of (9.52) vanishes by (9.48) and Remark 3. The second term vanishes by (9.51) and the uniform continuity of w on $[-M,M]$. Finally, (9.50), (9.52) and the arbitrariness of T show that $h(w_n, s_n) = w_n \circ s_n \to w \circ s = h(w,s)$, which is what we sought to prove. □

Proof of assertion 1 of Theorem 9.2.3. As defined in (9.23), $\hat{Y}^{(n)} := \bar{\omega}^{(n)} \circ \hat{S}^{(n)}$. Denoting by h the composition map as in the previous lemma, we set out to prove that

$$\hat{Y}^{(n)} = h(\bar{\omega}^{(n)}, \hat{S}^{(n)}) \xrightarrow{\mathrm{d}} h(\nu \mathrm{id}, W^{(\alpha)}) = \nu W^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1), \tag{9.53}$$

as $n \to \infty$. We do so by applying the following extension of the Continuous Mapping Theorem, see e.g. [Bil13, Theorem 5.1]: if $h : \mathcal{S} \longrightarrow \mathcal{S}'$ is a measurable function between two metric spaces, which are also regarded as measure spaces w.r.t. the respective Borel σ-algebras, $(X_n)_{n\in\mathbb{N}}$ and X are $\mathcal{S}$-valued random variables with $X_n \xrightarrow{\mathrm{d}} X$, and $\mathbb{P}(X \in \mathrm{Disc}(h)) = 0$, where $\mathrm{Disc}(h) \subset \mathcal{S}$ denotes the set of discontinuities of h, then $h(X_n) \xrightarrow{\mathrm{d}} h(X)$.

Note that the topological spaces $\mathcal{D}_0$ and $\mathcal{D}^+$, and thus $\mathcal{D}_0 \times \mathcal{D}^+$, are metrizable; see the comment after (6.26). From the independence of $\bar{\omega}^{(n)}$ and $\hat{S}^{(n)}$, and by (9.10), (9.14), and the definition of product topology,

$$(\bar{\omega}^{(n)}, \hat{S}^{(n)}) \xrightarrow{\mathrm{d}} (\nu \mathrm{id}, W^{(\alpha)}) \qquad \text{in } (\mathcal{D}_0 \times \mathcal{D}^+, J_1 \otimes J_1). \tag{9.54}$$

To apply the theorem and obtain (9.53) it remains to prove that the probability that $(\nu \mathrm{id}, W^{(\alpha)})$ hits a discontinuity of h is zero. But $(\nu \mathrm{id}, W^{(\alpha)}) \in (\mathcal{C} \cap \mathcal{D}_0) \times \mathcal{D}^+$, where h is continuous by Lemma 9.3.3. □

9.3.3 Proof of Theorem 9.2.2: Limit theorems for $\beta \in (0, 1)$

In comparison with the proof of Theorem 9.2.3, the main technical hurdle here is that in the composition $\hat{Y}^{(n)} = \hat{\omega}^{(n)} \circ \bar{S}^{(n)}$, cf. (9.21), the inner function (also referred to as *random time change*) is not increasing and one cannot use [Bil13, Theorem 5.1]. We shall only prove Theorem 9.2.2 in the case $\mu > 0$, as the other case is all but identical.

In view of Remark 4, we need to show that, for any $T > 0$, which we consider fixed throughout this proof, the restriction of $\hat{Y}^{(n)}$ to $[0, T)$ converges in $(\mathcal{D}([0, T)), J_2)$ to the restriction of $W^{(\alpha)} \circ \mu \mathrm{id}$ to $[0, T)$. By a double use of the Skorokhod Representation Theorem, there exist two probability spaces $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$, and processes

$$\begin{aligned}
&\hat{\boldsymbol{\omega}}^{(\boldsymbol{n})} : \Omega_1 \to \mathcal{D}, \\
&\boldsymbol{Z}^{(\boldsymbol{\beta})} : \Omega_1 \to \mathcal{D}, \\
&\bar{\boldsymbol{S}}^{(\boldsymbol{n})} : \Omega_2 \to \mathcal{D}^+, \\
&\tilde{\boldsymbol{S}}^{(\boldsymbol{n})} : \Omega_2 \to \mathcal{D}^+, \\
&\widetilde{\boldsymbol{W}}^{(\boldsymbol{\alpha})} : \Omega_2 \to \mathcal{D}^+,
\end{aligned} \tag{9.55}$$

with, respectively, the same distributions as $\tilde{\omega}^{(n)}, Z^{(\beta)}, \bar{S}^{(n)}, \tilde{S}^{(n)}, \widetilde{W}^{(\alpha)}$, such that the distributional convergences (9.8), (9.16) and (9.18) become almost sure convergences in the suitable spaces:

$$\begin{aligned}
&\hat{\boldsymbol{\omega}}^{(\boldsymbol{n})} \xrightarrow{\text{a.s.}} \boldsymbol{Z}^{(\boldsymbol{\beta})} && \text{on } (\Omega_1, \mathbb{P}_1) && \text{in } (\mathcal{D}, J_1) \\
&\bar{\boldsymbol{S}}^{(\boldsymbol{n})} \xrightarrow{\text{a.s.}} \mu \mathrm{id} && \text{on } (\Omega_2, \mathbb{P}_2) && \text{in } (\mathcal{D}^+, J_1), \\
&\tilde{\boldsymbol{S}}^{(\boldsymbol{n})} \xrightarrow{\text{a.s.}} \widetilde{\boldsymbol{W}}^{(\boldsymbol{\alpha})} && \text{on } (\Omega_2, \mathbb{P}_2) && \text{in } (\mathcal{D}^+, J_1).
\end{aligned} \tag{9.56}$$

Since $\hat{\omega}^{(n)}$ and $\bar{S}^{(n)}$ are independent, we regard the processes (9.55) as defined on $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \times \mathbb{P}_2)$, so all the joint distributions of processes in boldface type are the same as for the corresponding processes in regular type. Also, in the interest of readability and confident there will be no confusion, we slightly abuse the notation and write the boldface processes in regular type.

Let us denote by Ω_1' the set of realizations $\gamma_1 \in \Omega_1$ such that $\hat{\omega}^{(n)}[\gamma_1] \to Z^{(\beta)}[\gamma_1]$, as $n \to \infty$, in $(\mathcal{D}, J_1)$. In particular $\mathbb{P}_1(\Omega_1') = 1$. Similarly, let us denote by Ω_2' the set of realizations $\gamma_2 \in \Omega_2$ such that $\tilde{S}^{(n)}[\gamma_2] \to \widetilde{W}^{(\alpha)}[\gamma_2]$ in $(\mathcal{D}^+, J_1)$. Again $\mathbb{P}_2(\Omega_2') = 1$. Since $\tilde{S}^{(n)} = (n\bar{S}^{(n)} - \mu\lfloor n\,\cdot\rfloor)/n^{1/\alpha}$ converges almost surely to a Lévy stable process, whose trajectories are bounded when restricted to $[0,T)$ (though not uniformly bounded in γ_2), it is easy to see that, for any $\eta \in (0,1)$, there exist $C_\eta > 0$ and $\bar{n}_\eta \in \mathbb{N}$ such that the event

$$B_\eta := \left\{\gamma_2 \in \Omega_2' : \sup_{t\in[0,T)} \frac{n|\bar{S}^{(n)}[\gamma_2](t) - \mu t|}{n^{1/\alpha}} \le C_\eta \text{ for } n \ge \bar{n}_\eta\right\} \tag{9.57}$$

has probability $\mathbb{P}_2(B_\eta) > 1 - \eta$. Observe that C_η and $\bar{n}_\eta$ depend on T as well. Our goal for the rest of the proof will be to show that, for all realizations $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$,

$$\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \longrightarrow Z^{(\beta)} \circ \mu\,\mathrm{id} \qquad \text{in } (\mathcal{D}([0,T)), J_2), \tag{9.58}$$

as $n \to \infty$. This easily implies that the above convergence holds almost surely in $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \times \mathbb{P}_2)$. In fact, it holds on $\Omega_1' \times \bigcup_{k\in\mathbb{N}} B_{\eta_k}$, where $(\eta_k)_{k\in\mathbb{N}}$ is some vanishing sequence of numbers in $(0,1)$, and

$$\mathbb{P}_2\left(\bigcup_{k\in\mathbb{N}} B_{\eta_k}\right) \ge \limsup_{k\to\infty} \mathbb{P}_2(B_{\eta_k}) \ge \limsup_{k\to\infty} (1 - \eta_k) = 1. \tag{9.59}$$

Now, almost sure converges implies distributional converge (for the boldface processes and thus for the original processes). Finally, since $Z_\pm^{(\beta)}$ are β-stable and having assumed that $\mu > 0$, we observe that $Z^{(\beta)} \circ \mu\,\mathrm{id} = \mu^{1/\beta} Z_+^{(\beta)}$, thus proving (9.22).

So we are left with proving (9.58). We first give a plan of the proof, warning the reader that all the involved quantities depend in general on $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$, but we often omit this dependence. As per Definition 14, we will need to construct bijections $\tau_n : [0,T) \longrightarrow [0,T)$ such that, when $n \to \infty$,

$$\sup_{t\in[0,T)} |\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \tau_n(t) - Z^{(\beta)}(\mu t)| \to 0, \tag{9.60}$$

$$\sup_{t\in[0,T)} |\tau_n(t) - t| \to 0. \tag{9.61}$$

1. As a first step to obtain $(\tau_n)_{n\in\mathbb{N}}$, we construct bijections $\rho_n : [0,T) \longrightarrow [0,T)$ such that

$$\sup_{t\in[0,T)} |\rho_n(t) - t| \longrightarrow 0 \tag{9.62}$$

$$\bar{S}^{(n)} \circ \rho_n \in \mathcal{D}_{\mathrm{pc}}([0,T)) \cap \mathcal{D}_0([0,T)), \tag{9.63}$$

$$\bar{S}^{(n)} \circ \rho_n \longrightarrow \mu\,\mathrm{id} \qquad \text{in } (\mathcal{D}([0,T)), J_1), \tag{9.64}$$

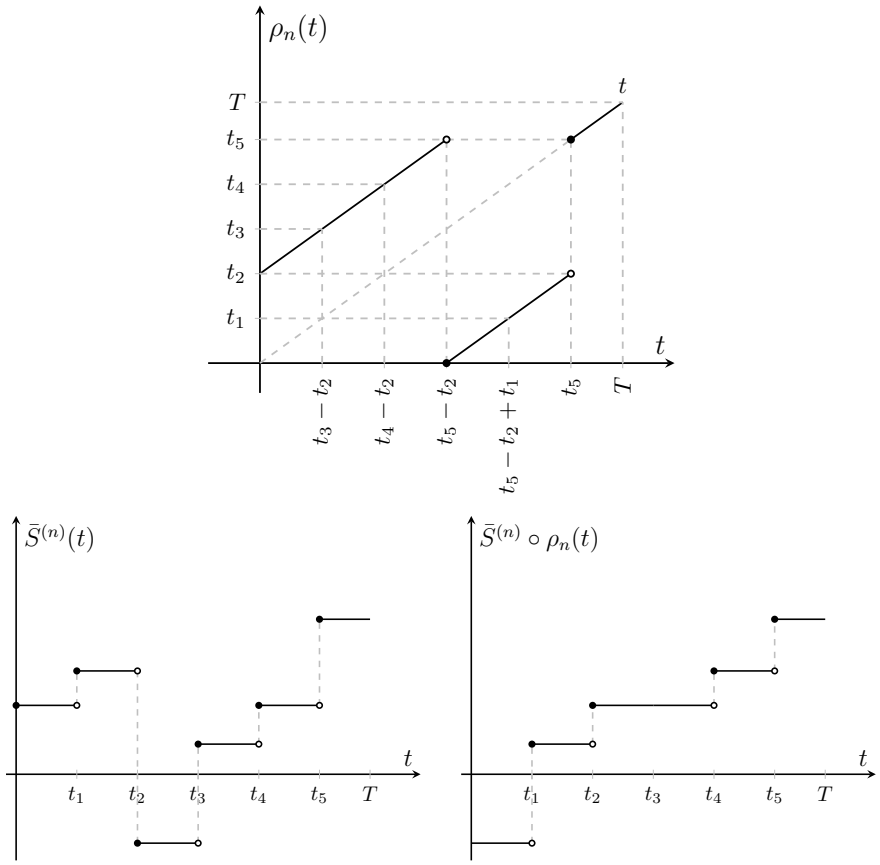

Figure 9.1: Upper panel: representation of ρ_n in $[0, T)$. Lower panel: sample path of $\bar{S}^{(n)}$ in $[0, T)$ (left) and the nondecreasing composition $\bar{S}^{(n)} \circ \rho_n$ (right).

where $\mathcal{D}_{\mathrm{pc}}([0, T))$ denotes the set of càdlàg piecewise constant functions of $[0, T)$ and $\mathcal{D}_0([0, T))$ denotes the set of càdlàg nondecreasing functions of $[0, T)$. See Figure 9.1 (upper panel) for an example of ρ_n associated to a given realization of $\bar{S}^{(n)}$.

2. Since $\hat{\omega}^{(n)} \longrightarrow Z^{(\beta)}$ in $(\mathcal{D}, J_1)$ for any $\gamma_1 \in \Omega_1'$, we can apply [Whi80, Theorem 3.1], which gives sufficient conditions for the composition of two càdlàg functions to be continuous in the J_1 topology. Using (9.63)-(9.64) we will get

$$\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n \longrightarrow Z^{(\beta)} \circ \mu\mathrm{id} \qquad \text{in } (\mathcal{D}([0, T)), J_1), \tag{9.65}$$

for any $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$. By definition of J_1-convergence, there exists a sequence

of homeomorphisms λ_n of $[0,T)$, such that, as $n \to \infty$,

$$\sup_{t\in[0,T)} |\lambda_n(t) - t| \to 0, \tag{9.66}$$

$$\sup_{t\in[0,T)} |\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n \circ \lambda_n(t) - Z^{(\beta)}(\mu t)| \to 0. \tag{9.67}$$

3. Note that (9.67) is exactly (9.60) for the bijection $\tau_n := \rho_n \circ \lambda_n$. Therefore, it will remain to establish (9.61), that is,

$$\sup_{t\in[0,T)} |\rho_n \circ \lambda_n(t) - t| \to 0. \tag{9.68}$$

We now fill the gaps in the steps above.

Construction of ρ_n. We begin by constructing the bijection $\rho_n : [0,T) \longrightarrow [0,T)$. Since $\bar{S}^{(n)}(t) = S_{\lfloor nt \rfloor}/n$ for $t \in [0,T)$ and $\gamma_2 \in B_\eta$, we have that

$$|S_{\lfloor u \rfloor} - \mu u| \leq C_\eta\, n^{1/\alpha}, \tag{9.69}$$

uniformly for $u \in [0, nT)$. Without loss of generality, we assume that $nT \in \mathbb{N}$. If not, one can take T' slightly larger than T, with $nT' \in \mathbb{N}$, and work in $[0,T']$. Let $p(\cdot)$ be a permutation of $\{0,1,\ldots,nT-1\}$ such that $S_{p(0)} \leq S_{p(1)} \leq \cdots \leq S_{p(nT-1)}$. We define ρ_n as follows:

$$\rho_n(t) := t - \frac{i}{n} + \frac{p(i)}{n}, \qquad t \in \Big[\frac{i}{n}, \frac{i+1}{n}\Big),\ i \in \{0,\ldots,nT-1\}. \tag{9.70}$$

Clearly, ρ_n is a bijection that maps $[i/n,(i+1)/n)$ affinely onto $[p(i)/n,(p(i)+1)/n)$. By construction of $p(\cdot)$, $\bar{S}^{(n)} \circ \rho_n$ is nondecreasing. The next proposition shows that ρ_n is uniformly close to the identity.

Proposition 9.3.1. *For all $\gamma_2 \in B_\eta$,*

$$\sup_{i\in\{0,\ldots,nT-1\}} |i - p(i)| \leq \frac{2C_\eta}{\mu}\, n^{1/\alpha} + 1. \tag{9.71}$$

Proof. To better explain the proof we refer to Figure 9.2, which shows a sample path of $u \mapsto S_{\lfloor u \rfloor}$ and the corresponding upper and lower bounds given by (9.69).

We establish (9.71) by estimating from below the cardinality of the sets

$$L_i := \{j \in \{0,\ldots,nT-1\} : S_{p(i)} \geq S_j\}, \tag{9.72}$$

$$U_i := \{j \in \{0,\ldots,nT-1\} : S_{p(i)} \leq S_j\}. \tag{9.73}$$

Let us begin with a lower bound for $|L_i|$. To this end, consider the interval $\mathcal{B}_i := [0, b_{i,n}] \subseteq [0,nT)$, where $b_{i,n} := p(i) - 2C_\eta n^{1/\alpha}/\mu$. This set was introduced because it

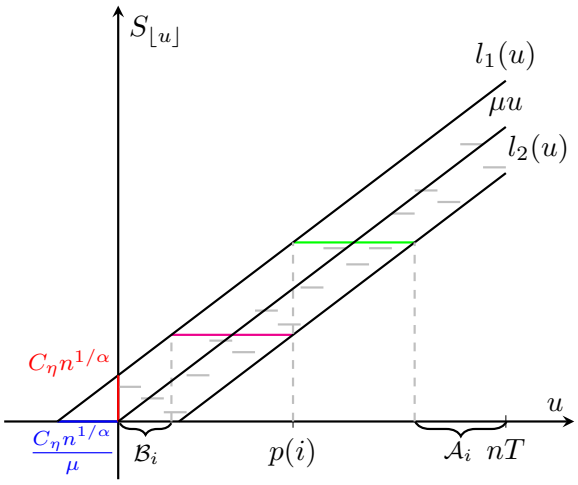

Figure 9.2: A sample path of $S_{\lfloor u \rfloor}$ and the corresponding upper and lower bounds given by (9.69) and represented by the graphs of the functions $l_1(u) = \mu u + C_\eta n^{1/\alpha}$ and $l_2(u) = \mu u - C_\eta n^{1/\alpha}$. The horizontal and verticals sizes of the strip are, respectively, $2C_\eta n^{1/\alpha}/\mu$ and $2C_\eta n^{1/\alpha}$. The sets $\mathcal{B}_i, \mathcal{A}_i$ are defined in the body of the text.

has the property that, for all $u \in \mathcal{B}_i$, $S_{\lfloor u \rfloor} \leq S_{p(i)}$, see Figure 9.2. Observe that, for small values of $p(i)$, $\mathcal{B}_i$ might be empty. These considerations show that

$$|L_i| \geq |\mathcal{B}_i \cap L_i| = |\mathcal{B}_i \cap \mathbb{Z}| = \max\{\lfloor b_{i,n} \rfloor + 1, 0\}. \tag{9.74}$$

On the other hand, since $S_{p(i)}$ is the $(i+1)$-th smallest value of the set $\{S_j\}_{j=0}^{nT-1}$, we know that $|L_i| = i + 1$. From the above inequality, then,

$$i \geq \lfloor b_{i,n} \rfloor = \left\lfloor p(i) - 2\frac{C_\eta n^{1/\alpha}}{\mu} \right\rfloor \geq p(i) - 2\frac{C_\eta n^{1/\alpha}}{\mu} - 1. \tag{9.75}$$

We proceed analogously to produce a lower bound for $|U_i|$. Set $\mathcal{A}_i := [a_{i,n}, nT)$, where $a_{i,n} = p(i) + 2C_\eta n^{1/\alpha}/\mu$. Figure 9.2 shows that $S_{\lfloor t \rfloor} \geq S_{p(i)}$ for all $t \in \mathcal{A}_i$, whence $|\mathcal{A}_i \cap U_i| = \lfloor nT - a_{i,n} \rfloor + 1$. On the other hand, $|U_i| = nT - i + 1$. Since $|U_i| \geq |\mathcal{A}_i \cap U_i|$ we get

$$nT - i \geq \lfloor nT - a_{i,n} \rfloor \geq nT - a_{i,n} - 1, \tag{9.76}$$

and so

$$i \leq a_{i,n} + 1 = p(i) + 2\frac{C_\eta n^{1/\alpha}}{\mu} + 1, \tag{9.77}$$

concluding the proof. □

J_2-convergence on $\Omega_1' \times B_\eta$. To derive (9.65) we first notice that, for any $\gamma_2 \in B_\eta$,

$$\bar{S}^{(n)} \circ \rho_n \to \mu \mathrm{id} \qquad \text{in } (\mathcal{D}([0,T)), J_1), \tag{9.78}$$

as $n \to \infty$. This is in fact a consequence of the following uniform convergence:

$$\begin{aligned} &\sup_{t\in[0,T)} |\bar{S}^{(n)} \circ \rho_n(t) - \mu t| \qquad (9.79)\\ &\quad = \sup_{i\in\{0,\dots,nT-1\}} \sup_{t\in[i/n,(i+1)/n)} \Big|\bar{S}^{(n)}\Big(t - \frac{i-p(i)}{n}\Big) - \mu t\Big| \\ &\quad \le \sup_{i\in\{0,\dots,nT-1\}} \sup_{t\in[i/n,(i+1)/n)} \Big|\bar{S}^{(n)}\Big(t - \frac{i-p(i)}{n}\Big) - \mu\Big(t - \frac{i-p(i)}{n}\Big)\Big| \\ &\qquad + \sup_{i\in\{0,\dots,nT-1\}} \Big|\frac{i-p(i)}{n}\Big| \\ &\quad = C_\eta \frac{n^{1/\alpha}}{n} + \frac{2C_\eta}{\mu}\frac{n^{1/\alpha}}{n} + \frac{1}{n}, \end{aligned}$$

which vanishes for $n \to \infty$. Now the plan is to once again apply [Whi80, Theorem 3.1] to show that (9.78) and the limit $\hat{\omega}^{(n)} \to Z^{(\beta)}$ in $(\mathcal{D}, J_1)$, which holds because $\gamma_1 \in \Omega_1'$, imply (9.65).

There is a problem, however. The theorem cannot be applied *tout court* because neither $\hat{\omega}^{(n)}$ nor $Z^{(\beta)}$ are càdlàg functions, cf (9.6)-(9.7). On the other hand, we can use the considerations of Remark 5 to show that $\hat{\omega}^{(n)}_{\mathrm{cadlag}} \to Z^{(\beta)}_{\mathrm{cadlag}}$ in $(\mathcal{D}, J_1)$ and thus, by [Whi80, Theorem 3.1],

$$\hat{\omega}^{(n)}_{\mathrm{cadlag}} \circ \bar{S}^{(n)} \circ \rho_n \longrightarrow Z^{(\beta)}_{\mathrm{cadlag}} \circ \mu \mathrm{id} \qquad \text{in } (\mathcal{D}([0,T)), J_1). \tag{9.80}$$

But the restrictions of $Z^{(\beta)}_{\mathrm{cadlag}}$ and $Z^{(\beta)}$ to $[0, \mu T)$ coincide, so will obtain (9.65) when we prove that $\hat{\omega}^{(n)}_{\mathrm{cadlag}} \circ \bar{S}^{(n)} \circ \rho_n$ is asymptotic to $\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n$ in $(\mathcal{D}([0,T)), J_1)$, as $n \to \infty$. We will show more, namely that, for some $C > 0$,

$$\sup_{t\in\mathbb{R}^+} \Big|\hat{\omega}^{(n)}_{\mathrm{cadlag}} \circ \bar{S}^{(n)} \circ \rho_n(t) - \hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n(t)\Big| \le \frac{C}{n^{1/\beta}}. \tag{9.81}$$

In fact, with the help of (9.1) and (9.6) observe that

$$\hat{\omega}^{(n)}_{\mathrm{cadlag}}(s) - \hat{\omega}^{(n)}(s) = \begin{cases} \dfrac{\zeta_j}{n^{1/\beta}} & \text{if } s = \dfrac{j}{n},\ j \in \mathbb{Z}^- \\ 0 & \text{otherwise} \end{cases}, \tag{9.82}$$

where the numbers ζ_j, for $j \in \mathbb{Z}^-$, are fixed, as the realization $\gamma_1 \in \Omega_1'$ of the medium is fixed. Now, the realization $\gamma_2 \in B_\eta$ of the underlying random walk is also fixed. Since the drift μ is positive, $S_n < 0$ occurs only for a finite number of times n. The values

of these excursions below zero and their times are contained in this chain of inequalities $S_{p(0)} \le S_{p(1)} \le \ldots \le S_{p(m-1)} < 0 \le S_{p(m)}$, for some $m \in \mathbb{N}$. Thus

$$\sup_{t\in[0,p(m)/n)} \left|\hat{\omega}^{(n)}_{\mathrm{cadlag}} \circ \bar{S}^{(n)} \circ \rho_n(t) - \hat{\omega}^{(n)} \circ \bar{S}^{(n)} \circ \rho_n(t)\right| \le \frac{\max\{\zeta_j\}_{j=S_{p(0)}}^{-1}}{n^{1/\beta}}. \tag{9.83}$$

Since the expression on the above l.h.s. is identically 0 for $t \ge p(m)/n$, we have proved (9.81), thus (9.65), thus (9.66)-(9.67).

We are left to prove that $\tau_n = \rho_n \circ \lambda_n$ satisfies (9.68). We do so with the help of Proposition 9.3.1:

$$\begin{aligned}\sup_{t\in[0,T)} |\rho_n \circ \lambda_n(t) - t| &= \sup_{s\in[0,T)} |\rho_n(s) - \lambda_n^{-1}(s)| \qquad (9.84)\\ &= \sup_{i\in\{0,\ldots,nT-1\}} \sup_{s\in[i/n,(i+1)/n)} \left|s - \frac{i-p(i)}{n} - \lambda_n^{-1}(s)\right| \\ &\le \sup_{s\in[0,T)} |s - \lambda_n^{-1}(s)| + \sup_{i\in\{0,\ldots,nT-1\}} \left|\frac{p(i)-i}{n}\right|,\end{aligned}$$

which converges to 0 as $n \to \infty$ by (9.66) and (9.71). This finally shows that $\hat{\omega}^{(n)} \circ \bar{S}^{(n)} \to Z^{(\beta)} \circ \mu\mathrm{id}$ in $(\mathcal{D}([0,T)), J_2)$ for all $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$, concluding the proof of Theorem 9.2.2.

9.3.4 Proof of Theorem 9.2.4: Limit theorems for the fluctuations

Once again, we only prove the theorem in the case $\mu > 0$. Using definitions (9.11) and (9.17) we write

$$n(\bar{Y}^{(n)} - \mu\nu\mathrm{id}) = n^{1/\beta}\tilde{\omega}^{(n)} \circ \bar{S}^{(n)} + n^{1/\alpha}\nu\tilde{S}^{(n)} + \nu\mu(\lfloor n\mathrm{id}\rfloor - n\mathrm{id}). \tag{9.85}$$

The asymptotic behavior of (9.85) depends crucially on the ratio α/β, hence we distinguish three cases. Observe that since $(\lfloor n\mathrm{id}\rfloor - n\mathrm{id})$ is bounded, the last term of the above r.h.s. vanishes in the limit, whether we divide it by $n^{1/\alpha}$ or by $n^{1/\beta}$.

Case $\alpha < \beta$. Substituting $\hat{\omega}^{(n)}$ with $\tilde{\omega}^{(n)}$ and $Z^{(\beta)}$ with $\tilde{Z}^{(\beta)}$, cf. (9.12), in the proof of Theorem 9.2.2 shows that, for $\beta \in (1,2)$ and $n \to \infty$,

$$\tilde{\omega}^{(n)} \circ \bar{S}^{(n)} \xrightarrow{\mathrm{d}} \mu^{1/\beta}\tilde{Z}^{(\beta)}_+ \qquad \text{in } (\mathcal{D}^+, J_2). \tag{9.86}$$

Since in this case $1/\alpha > 1/\beta$, we obtain

$$\frac{\tilde{\omega}^{(n)} \circ \bar{S}^{(n)}}{n^{1/\alpha - 1/\beta}} \xrightarrow{\mathrm{d}} 0 \qquad \text{in } (\mathcal{D}^+, J_1), \tag{9.87}$$

where we have observed that we can pass from J_2-convergence to J_1-convergence because both convergences reduce to uniform convergence when the limit function is continuous, cf. Remark 3. Finally, by (9.85), (9.87) and (9.18),

$$\frac{n(\bar{Y}^{(n)} - \mu\nu\mathrm{id})}{n^{1/\alpha}} \xrightarrow{\mathrm{d}} \nu\widetilde{W}^{(\alpha)} \qquad \text{in } (\mathcal{D}^+, J_1), \tag{9.88}$$

which amounts to (9.28), as desired.

Case $\alpha > \beta$. Since in this case $1/\beta > 1/\alpha$, the leading order term in (9.85) is the first term, whose limit has been identified in (9.86). We conclude that

$$\frac{n(\bar{Y}^{(n)} - \mu\nu\,\mathrm{id})}{n^{1/\beta}} \xrightarrow{\mathrm{d}} \mu^{1/\beta}\,\tilde{Z}_+^{(\beta)} \qquad \text{in } (\mathcal{D}^+, J_2), \tag{9.89}$$

i.e., (9.30) holds for the case $\mu > 0$.

Case $\alpha = \beta$. Except for certain complications, the proof of this case will follow the same ideas as that of Theorem 9.2.2 in Section 9.3.3. We will detail the parts that need a new argument and describe quickly those that are proved exactly as done earlier.

In view of (9.85), we rewrite our process of interest as

$$\tilde{Y}^{(n)} = \tilde{\omega}^{(n)} \circ \bar{S}^{(n)} + \nu\tilde{S}^{(n)} + \delta_n = \ell\big(h\big(\tilde{\omega}^{(n)}, \bar{S}^{(n)}\big), \nu\tilde{S}^{(n)}\big) + \delta_n, \tag{9.90}$$

where $h(x,y) := x \circ y$ is the composition map from $\mathcal{D} \times \mathcal{D}^+$ to $\mathcal{D}^+$ and $\ell(x,y) := x + y$ is the addition map from $\mathcal{D}^+ \times \mathcal{D}^+$ to $\mathcal{D}^+$. Also $\delta_n := \nu\mu(\lfloor n\,\mathrm{id}\rfloor - n\,\mathrm{id})/n^{1/\alpha}$ is a negligible term, as $n \to \infty$, in any relevant distance. As was done in Section 9.3.3, we use the Skorokhod Representation Theorem twice to obtain two probability spaces $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$, and processes

$$\begin{aligned}
&\tilde{\boldsymbol{\omega}}^{(\boldsymbol{n})} : \Omega_1 \to \mathcal{D},\\
&\tilde{\boldsymbol{Z}}^{(\boldsymbol{\alpha})} : \Omega_1 \to \mathcal{D},\\
&\bar{\boldsymbol{S}}^{(\boldsymbol{n})} : \Omega_2 \to \mathcal{D}^+,\\
&\tilde{\boldsymbol{S}}^{(\boldsymbol{n})} : \Omega_2 \to \mathcal{D}^+,\\
&\widetilde{\boldsymbol{W}}^{(\boldsymbol{\alpha})} : \Omega_2 \to \mathcal{D}^+,
\end{aligned} \tag{9.91}$$

with respectively the same distribution as $\tilde{\omega}^{(n)}, \tilde{Z}^{(\alpha)}, \bar{S}^{(n)}, \tilde{S}^{(n)}$ and $\widetilde{W}^{(\alpha)}$, and such that

$$\begin{aligned}
&\tilde{\boldsymbol{\omega}}^{(n)} \xrightarrow{\text{a.s.}} \tilde{\boldsymbol{Z}}^{(\boldsymbol{\alpha})} && \text{on } (\Omega_1, \mathbb{P}_1) && \text{in } (\mathcal{D}, J_1)\\
&\bar{\boldsymbol{S}}^{(\boldsymbol{n})} \xrightarrow{\text{a.s.}} \mu\,\mathrm{id} && \text{on } (\Omega_2, \mathbb{P}_2) && \text{in } (\mathcal{D}^+, J_1)\\
&\tilde{\boldsymbol{S}}^{(\boldsymbol{n})} \xrightarrow{\text{a.s.}} \widetilde{\boldsymbol{W}}^{(\boldsymbol{\alpha})} && \text{on } (\Omega_2, \mathbb{P}_2) && \text{in } (\mathcal{D}^+, J_1)
\end{aligned} \tag{9.92}$$

Once again, since the processes relative to the medium and those relative to the dynamics are independent, it is correct to regard all boldface processes as defined on $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \times \mathbb{P}_2)$. Again we simplify the notation and use the regular typeset for all processes (9.91). Let us define

$$\Omega_1' := \{\gamma_1 \in \Omega_1 \,:\, \tilde{\omega}^{(n)}[\gamma_1] \to \tilde{Z}^{(\alpha)}[\gamma_1] \text{ in } (\mathcal{D}, J_1)\}, \tag{9.93}$$

$$\Omega_2' := \{\gamma_2 \in \Omega_2 \,:\, \tilde{S}^{(n)}[\gamma_2] \to \widetilde{W}^{(\alpha)}[\gamma_2] \text{ in } (\mathcal{D}^+, J_1)\}. \tag{9.94}$$

These are full-measure sets in their respective spaces. Notice that (essentially by the definition of $\tilde{S}^{(n)}$) $\bar{S}^{(n)}[\gamma_2] \to \mu\,\mathrm{id}$, for all $\gamma_2 \in \Omega_2'$. Now let us fix $T > 0$. We already

know that for any $\eta \in (0,1)$, there exist $C_\eta > 0$ and $\bar{n}_\eta \in \mathbb{N}$ (both numbers depending on T as well) such that the set

$$B_\eta := \Big\{ \gamma_2 \in \Omega_2' : \sup_{t\in[0,T)} \frac{n|\bar{S}^{(n)}[\gamma_2](t) - \mu t|}{n^{1/\alpha}} \leq C_\eta \text{ for } n \geq \bar{n}_\eta \Big\} \tag{9.95}$$

has measure $\mathbb{P}_2(B_\eta) > 1 - \eta$. Now one proceeds as in the proof of Theorem 9.2.2, using $\tilde{\omega}^{(n)}$ and $\tilde{Z}^{(\alpha)}$ in place of $\hat{\omega}^{(n)}$ and $Z^{(\beta)}$, respectively. The fact that now $\beta = \alpha \in (1,2)$ causes no breaks in the proof. One obtains that, for all $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$,

$$h(\tilde{\omega}^{(n)}, \bar{S}^{(n)}) = \tilde{\omega}^{(n)} \circ \bar{S}^{(n)} \to \tilde{Z}^{(\alpha)} \circ \mu\,\mathrm{id} \qquad \text{in } (\mathcal{D}([0,T)), J_2). \tag{9.96}$$

By construction of B_η, $\nu\tilde{S}^{(n)} \to \nu\widetilde{W}^{(\alpha)}$ in $(\mathcal{D}([0,T)), J_2)$ for all $\gamma_2 \in B_\eta$. If we are able to prove that, for all $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$, the addition map ℓ is continuous at $(\tilde{Z}^{(\alpha)} \circ \mu\,\mathrm{id}, \nu\widetilde{W}^{(\alpha)})$ in the space $\mathcal{D}([0,T)), J_2)$, we obtain by (9.90) that

$$\lim_{n\to\infty} \tilde{Y}^{(n)} = \lim_{n\to\infty} \ell\big(h\big(\tilde{\omega}^{(n)}, \bar{S}^{(n)}\big), \nu\tilde{S}^{(n)}\big) = \tilde{Z}^{(\alpha)} \circ \mu\,\mathrm{id} + \nu\widetilde{W}^{(\alpha)}, \tag{9.97}$$

for all realizations $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$. Since $\eta \in (0,1)$ is arbitrary, the above limit extends to an almost sure limit in $(\Omega_1 \times \Omega_2, \mathbb{P}_1 \times \mathbb{P}_2)$. Passing to distributional convergence and (freely) varying the choice of T, we finally achieve (9.32) for the case $\mu > 0$. This ends the proof of Theorem 9.2.4.

It remains to produce a continuity result for the addition map of two càdlàg functions w.r.t. the J_2 topology. For a general space of càdlàg functions, Whitt proves that ℓ is continuous, relative to the topologies J_1, M_1 or M_2, at all pairs (x, y) such that $\mathrm{Disc}(x) \cap \mathrm{Disc}(y) = \emptyset$ (see [Whi80, Theorem 4.1] and [Whi02, Corollaries 12.7.1 & 12.11.5], respectively). In Theorem 6.7.1 of the Appendix we extend this statement to the topology J_2. As for its applicability to our case, notice that we can indeed assume that, for all $(\gamma_1, \gamma_2) \in \Omega_1' \times B_\eta$,

$$\mathrm{Disc}(\widetilde{Z}^{(\alpha)} \circ \mu\,\mathrm{id}) \cap \mathrm{Disc}(\nu\tilde{W}^{(\alpha)}) = \emptyset, \tag{9.98}$$

because $\widetilde{Z}^{(\alpha)}$ and $\tilde{W}^{(\alpha)}$ are independent α-stable processes and one can always remove from $\Omega_1' \times \Omega_2'$ the null set of realizations for which (9.98) does not hold.

Conclusions

In this thesis we faced two main problems, dealing with statistics and stochastic modeling of complex systems. In both cases, we analyzed the microscopic scale, be it a local sample of data from huge datasets or the key components that drive the dynamics of stochastic processes, to characterize laws acting on a large scale.

The first part of the thesis dealt with the biodiversity problem, i.e. estimating the species richness of a large area from scattered local surveys within it. Biodiversity provides support for life, vital provisions, regulating services and it has positive cultural impacts. It is therefore important to have accurate methods to measure biodiversity, in order to safeguard it when we discover it to be threatened. For practical reasons, biodiversity is usually measured at fine scales whereas diversity issues (e.g. conservation) interest regional or global scales. Moreover, biodiversity may change across spatial scales. It is therefore a key challenge to be able to translate local information on biodiversity into global patterns.
Our framework (see Chapter 2) is grounded on the form-invariance property of the Negative Binomial distribution. The Negative Binomial is a simple and versatile distribution that depending on its parameters can display an interior mode or Log-series like behavior or a Power Law behavior, i.e. it can accommodate different RSA shapes. Therefore we used the same RSA function to reproduce different ecosystems' RSA, as those typically observed in real ecosystems and in datasets derived from human activities.
Expanding the ability to upscale species richness and obtain abundance distributions from presence/absence data is of fundamental importance in many contexts, where abundance information are not available or trustable.

In Chapter 3, we presented a general analytical framework to extrapolate species richness and other relevant biodiversity patterns at the whole forest scale from local information on species presence/absence.
Furthermore, we introduce a new descriptor/measure of biodiversity within an ecological community, the RSO, which describes the distribution of species occurrences in scattered plots. The RSO distribution displays a fat tail, indicating that many species typically occupies only few scattered plots, while only very few species are pervasive and are found in most of the plot. Our prediction is that this property is not particular for the dataset here considered, rather it is another emergent pattern pervasive in highly biodiverse ecosystems.
To summarize, this flexible analytical method provides, from local presence/absence in-

formation, robust estimates of species richness and important macro-ecological patterns of biodiversity (SAC, RSA, RSO), as tested in both *in-silico* generated and two rainforests. The method may be applied to any database in the form of a binary matrix, where presence/absence features (tree species in our case) are detected across different samples.

Our second work (see Chapter 4) consists on an application of our upscaling method on four different "artificial" datasets coming from human dynamics (email communication, Twitter posts, Wikipedia articles and Gutenberg books). Big data require new techniques to handle the information they come with, and we think that our upscaling method can bring a different insight in this field.
More precisely, we infer the number of senders, hashtags and words of the whole dataset and how their abundances (i.e. the popularity of a hashtag) change through scales from a small sample of sent emails per sender, posts per hashtag and word occurrences. Our approach is grounded on statistical ecology as we map inference of human activities into the unseen species problem in biodiversity. Our findings may have applications to resource management in emails, collective attention monitoring in Twitter and language learning process in word databases.
Note that this work goes beyond the application of our method to datasets of different source. In facts, as far as we know, the problem of estimating how abundances vary through scales has not been previously investigated and it could be of interest when interpreting abundances of types as a measure of popularity in social network data.

The second part of the thesis deal with limit theorems for a specific class of random walk in random environment. Indeed, although classical results in the theory of random walks are widely used to model a wide range of transport processes in diverse fields as physics, chemistry and biology, in many interesting situations the walker moves in a complex and/or disordered environment, and the correlations induced by spatial inhomogeneities can have a strong impact on the transport properties.
We thus described a process that belongs to the class of random walks in random environment (RWRE), and in particular we described a Lévy flight on a one-dimensional Lévy random medium. The interest of this process is twofold. As first, our work deepened the knowledge random walks whose steps are correlated, due to the presence of a random environment. Secondly, our process is a simple model for a process in inhomogeneous environment, and this simple hypothesis is able to explain its anomalous behavior. Indeed, recent years have witnessed a growing interest around anomalous diffusive processes, where the variance of a moving particle has a super- or sub-linear growth in time. In the physical literature, such anomalous behavior has been observed in many systems, such as Lorentz gases with infinite horizon, rotating flows or intermittent dynamical systems. Though several models have been put forth to describe such situations, not much work has been done on systems that combine long-tailed jumps and disordered media. Such kind of models are representative of the many physical situations (human mobility, epidemics, network routing, etc.) in which anomalous diffusion is caused by the complexity of an underlying network.

With this work we hope to push forward the research in this field. We proved annealed limit theorems for Y. The scaling limit is a power of n that depends of the range of both parameters $\alpha, \beta \in (0,1) \cup (1,2)$, the stable indexes of the distribution of the steps of underlying random walk and of the distances between targets of the random environment respectively.
To the authors' knowledge, this is the first time that heavy tailed distribution are chosen both for S and for ω. Future lines of investigations must consider the Lévy walk X associated to the the process Y, i.e. a continuous interpolation performed with constant speed.

Currently, there is an ongoing work aiming at characterizing the asymptotic behavior of the first passage time (FPT) in the positive semiaxis. The FPT plays a key role when dealing with many fundamental questions in the theory of stochastic processes and finds applications among different fields of the natural sciences and economics, from the meeting time of two molecules in diffusion-controlled chemical reactions to the price of a stock reaching a specific value and the ruin problem of actuarial science.

In the framework of random walks in one dimension, a remarkable result is the Sparre-Andersen theorem [And54; Fel71] that shows the universality of the ubiquitous first-passage-time distribution. In particular, it states that the probability distribution of the number of steps needed to land on the positive semiaxis for a walker starting at the origin does not depend on the details of the jump-distribution of the walker (provided this distribution to be symmetric and continuous) and decays as $n^{-3/2}$.
However, we want to focus on the time needed for the FTP, rather than on the number of steps. To deal with such cases, a modification of the Sparre-Andersen proof has been made in [Art+14], assuming rather general situations in which the time variable correlates with the step variable. Still, their findings do not apply to our case, since they maintain the i.i.d. steps' assumption for the random walk, whereas in our model the length of the steps is correlated due to the presence of the Lévy random environment.

Currently we manage to characterize the asymptotics of the FPT only for a simple symmetric underlying random walk in the usual Lévy random medium ω of stable index β. To be precise, we accomplished it only when $\beta \in (1,2)$, whereas for $\beta \in (0,1)$ we provide unmatching estimates. When dealing with more general cases, the difficulty of the treatment increases exponentially. We can derive some relation, but not a closed formula. However, this research looks promising a possibly helpful for the development of this field.

Bibliography Part I

[Aza+15] Sandro Azaele et al. "Towards a unified descriptive theory for spatial ecology: predicting biodiversity patterns across spatial scales". In: *Methods in Ecology and Evolution* 6.3 (2015), pp. 324–332.

[Aza+16] Sandro Azaele et al. "Statistical mechanics of ecological systems: Neutral theory and beyond". In: *Rev. Mod. Phys.* 88 (3 July 2016), p. 035003.

[Baa02] R Harald Baayen. *Word frequency distributions*. Vol. 18. Springer Science & Business Media, 2002.

[Bar05] Albert-László Barabási. "The origin of bursts and heavy tails in human dynamics". In: *Nature* 435.7039 (2005), p. 207.

[Ber+16] Enrico Bertuzzo et al. "Geomorphic controls on elevational gradients of species richness". In: *Proceedings of the National Academy of Sciences* 113.7 (2016), pp. 1737–1742.

[BMW03] Ulrich Brose, Neo D. Martinez, and Richard J. Williams. "Estimating species richness: sensitivity to sample coverage and insensitivity to spatial patterns". In: *Ecology* 84.9 (2003), pp. 2364–2377.

[BF93] John Bunge and M. Fitzpatrick. "Estimating the number of species: a review". In: *Journal of the American Statistical Association* 88.421 (1993), pp. 364–373.

[Cha05] Anne Chao. "Species estimation and applications". In: *Encyclopedia of statistical sciences* (2005).

[CB02] Anne Chao and John Bunge. "Estimating the number of species in a stochastic abundance model". In: *Biometrics* 58.3 (2002), pp. 531–539.

[CC14] Anne Chao and Chun-Huo Chiu. "Species richness: estimation and comparison". In: *Wiley StatsRef: Statistics Reference Online* (2014), pp. 1–26.

[CC16] Anne Chao and Chun-Huo Chiu. "Species richness: estimation and comparison". In: *Wiley StatsRef: Statistics Reference Online* (2016).

[Cha+09] Anne Chao et al. "Sufficient sampling for asymptotic minimum species richness estimators". In: *Ecology* 90.4 (2009), pp. 1125–1133.

[Chi07] Ryan A. Chisholm. "Sampling species abundance distributions: resolving the veil-line debate". In: *Journal of theoretical biology* 247.4 (2007), pp. 600–607.

[CMC04] Robert K Colwell, Chang Xuan Mao, and Jing Chang. "Interpolating, extrapolating, and comparing incidence-based species accumulation curves". In: *Ecology* 85.10 (2004), pp. 2717–2727.

[CC94] Robert K. Colwell and Jonathan A. Coddington. "Estimating terrestrial biodiversity through extrapolation". In: *Philosophical Transactions: Biological Sciences* (1994), pp. 101–118.

[Cor41] A. Steven Corbet. "The distribution of butterflies in the Malay Peninsula (Lepid.)" In: *Physiological Entomology* 16.10-12 (1941), pp. 101–116.

[Cro+15] Thomas W. Crowther et al. "Mapping tree density at a global scale". In: *Nature* 525.7568 (2015), pp. 201–205.

[Dev+16] Pierre Deville et al. "Scaling identity connects human mobility and social interactions". In: *Proceedings of the National Academy of Sciences* 113.26 (2016), pp. 7047–7052.

[Eli+06] Jane Elith et al. "Novel methods improve prediction of species' distributions from occurrence data". In: *Ecography* 29.2 (2006), pp. 129–151.

[FCW43] Ronald A. Fisher, A. Steven Corbet, and Carrington B. Williams. "The relation between the number of species and the number of individuals in a random sample of an animal population". In: *The Journal of Animal Ecology* (1943), pp. 42–58.

[FS08] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2008.

[For+14] Marco Formentin et al. "Hidden scaling patterns and universality in written communication". In: *Physical Review E* 90.1 (2014), p. 012817.

[GT56] I. J. Good and G. H. Toulmin. "The number of new species, and the increase in population coverage, when a sample is increased". In: *Biometrika* 43.1-2 (1956), pp. 45–63.

[HSS09] John Harte, Adam B Smith, and David Storch. "Biodiversity scales from plots to biomes with a universal species–area curve". In: *Ecology letters* 12.8 (2009), pp. 789–797.

[Hub01] Stephen P Hubbell. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Vol. 32. Princeton University Press, 2001.

[Hug+01] Jennifer B. Hughes et al. "Counting the uncountable: statistical approaches to estimating microbial diversity". In: *Applied and environmental microbiology* 67.10 (2001), pp. 4399–4406.

[ILL09] Iuliana Ionita-Laza, Christoph Lange, and Nan M. Laird. "Estimating the number of unseen variants in the human genome". In: *Proceedings of the National Academy of Sciences* 106.13 (2009), pp. 5008–5013.

[KJK18] Márton Karsai, Hang-Hyun Jo, and Kimmo Kaski. *Bursty human dynamics*. Springer, 2018.

[Kor07] András Kornai. *Mathematical linguistics*. Springer Science & Business Media, 2007.

[Kun+18] William E Kunin et al. "Upscaling biodiversity: estimating the species–area relationship from small samples". In: *Ecological Monographs* 88.2 (2018), pp. 170–187.

[LL16] Kenneth J. Locey and Jay T. Lennon. "Scaling laws predict global microbial diversity". In: *Proceedings of the National Academy of Sciences* 113.21 (2016), pp. 5970–5975.

[LMT12] Vittorio Loreto, Animesh Mukherjee, and Francesca Tria. "On the origin of the hierarchy of color names". In: *Proceedings of the National Academy of Sciences* 109.18 (2012), pp. 6819–6824.

[Lor+11] Vittorio Loreto et al. "Statistical physics of language dynamics". In: *Journal of Statistical Mechanics: Theory and Experiment* 2011.04 (2011), P04006.

[Mac60] Robert H. MacArthur. "On the relative abundance of species". In: *The American Naturalist* 94.874 (1960), pp. 25–36.

[Mag05] Anne E. Magurran. "Species abundance distributions: pattern or process?" In: *Functional Ecology* 19.1 (2005), pp. 177–181.

[Mag13] Anne E. Magurran. *Measuring biological diversity*. John Wiley & Sons, 2013.

[Mal+09] R Dean Malmgren et al. "On universality in human correspondence activity". In: *science* 325.5948 (2009), pp. 1696–1700.

[MC05] Chang Xuan Mao and Robert K. Colwell. "Estimation of species richness: mixture models, the role of rare species, and inferential challenges". In: *Ecology* 86.5 (2005), pp. 1143–1153.

[Maz+18] Andrea Mazzolini et al. "Statistics of shared components in complex component systems". In: *Physical Review X* 8.2 (2018), p. 021023.

[McG+07] Brian J. McGill et al. "Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework". In: *Ecology letters* 10.10 (2007), pp. 995–1015.

[MNK16] Peter Menzel, Kim Lee Ng, and Anders Krogh. "Fast and sensitive taxonomic classification for metagenomics with Kaiju". In: *Nature communications* 7 (2016).

[Mon+17] Bernardo Monechi et al. "Waves of novelties in the expansion into the adjacent possible". In: *PloS one* 12.6 (2017).

[OSW16] Alon Orlitsky, Ananda Theertha Suresh, and Yihong Wu. "Optimal prediction of the number of unseen species". In: *Proceedings of the National Academy of Sciences* 113.47 (2016), pp. 13283–13288.

[Plo+00] Joshua B. Plotkin et al. "Species-area curves, spatial aggregation, and habitat specialization in tropical forests". In: *Journal of theoretical biology* 207.1 (2000), pp. 81–99.

[Pre48] Frank W. Preston. "The commonness, and rarity, of species". In: *Ecology* 29.3 (1948), pp. 254–283.

[RN03] J. Andrew Royle and James D. Nichols. "Estimating abundance from repeated presence–absence data or point counts". In: *Ecology* 84.3 (2003), pp. 777–790.

[She+14] Huawei Shen et al. "Modeling and predicting popularity dynamics via reinforced poisson processes". In: *Twenty-eighth AAAI conference on artificial intelligence*. 2014.

[SH08] Tsung-Jen Shen and Fangliang He. "An incidence-based richness estimator for quadrats sampled without replacement". In: *Ecology* 89.7 (2008), pp. 2052–2060.

[Sin+16] Roberta Sinatra et al. "Quantifying the evolution of individual scientific impact". In: *Science* 354.6312 (2016), aaf5239.

[Sli+15] JW Ferry Slik et al. "An estimate of the number of tropical tree species". In: *Proceedings of the National Academy of Sciences* 112.24 (2015), pp. 7472–7477.

[Soe+12] David AW Soergel et al. "Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences". In: *The ISME journal* 6.7 (2012), p. 1440.

[Suw+13] Samir Suweis et al. "Emergence of structural and dynamical properties of ecological mutualistic networks". In: *Nature* 500.7463 (2013), p. 449.

[Tör+13] János Török et al. "Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment". In: *Physical review letters* 110.8 (2013), p. 088701.

[Tov+16] Anna Tovo et al. "Application of optimal data-based binning method to spatial analysis of ecological datasets". In: *Spatial Statistics* 16 (2016), pp. 137–151.

[Tov+17] Anna Tovo et al. "Upscaling species richness and abundances in tropical forests". In: *Science advances* 3.10 (2017), e1701438.

[Tov+19a] Anna Tovo et al. "Inferring macro-ecological patterns from local species' occurrences". In: *Oikos* (2019).

[Tov+19b] Anna Tovo et al. "Upscaling human activity data: an ecological perspective". In: *arXiv preprint arXiv:1912.03023* (2019).

[Vol+03] Igor Volkov et al. "Neutral theory and relative species abundance in ecology". In: *Nature* 424.6952 (2003), p. 1035.

[Vol+05] Igor Volkov et al. "Density dependence explains tree species abundance and diversity in tropical forests." In: *Nature* 438.7068 (Dec. 2005), pp. 658–61. ISSN: 1476-4687.

[Vol+07] Igor Volkov et al. "Patterns of relative species abundance in rainforests and coral reefs". In: *Nature* 450.7166 (2007), pp. 45–49.

[Wal+12] Joris Walraevens et al. "Stochastic queueing-theory approach to human dynamics". In: *Physical Review E* 85.2 (2012), p. 021139.

[WL05] Ji-Ping Z. Wang and Bruce G. Lindsay. "A penalized nonparametric maximum likelihood approach to species richness estimation". In: *Journal of the American Statistical Association* 100.471 (2005), pp. 942–959.

[Whi65] Robert H Whittaker. "Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution". In: *Science* 147.3655 (1965), pp. 250–260.

[YHM17] Taha Yasseri, Scott A Hale, and Helen Z Margetts. "Rapid rise and decay in petition signing". In: *EPJ Data Science* 6.1 (2017), p. 20.

[YSK12] Taha Yasseri, Robert Sumi, and János Kertész. "Circadian patterns of wikipedia editorial activity: A demographic analysis". In: *PloS one* 7.1 (2012), e30091.

[Zha+15] Qingyuan Zhao et al. "Seismic: A self-exciting point process model for predicting tweet popularity". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2015, pp. 1513–1522.

Bibliography Part II

[And54] Erik Sparre Andersen. "On the fluctuations of sums of random variables". In: *Mathematica Scandinavica* (1954), pp. 263–285.

[Art+14] Roberto Artuso et al. "Sparre-Andersen theorem with spatiotemporal correlations". In: *Physical Review E* 89.5 (2014), p. 052111.

[Art+18] Roberto Artuso et al. "Non-homogeneous persistent random walks and Lévy–Lorentz gas". In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.8 (2018), p. 083209.

[BFK00] E Barkai, V Fleurov, and J Klafter. "One-dimensional stochastic Lévy-Lorentz gas". In: *Physical Review E* 61.2 (2000), p. 1164.

[BBW08] Pierre Barthelemy, Jacopo Bertolotti, and Diederik S Wiersma. "A Lévy flight for light". In: *Nature* 453.7194 (2008), pp. 495–498.

[BGA09] CWJ Beenakker, CW Groth, and AR Akhmerov. "Nonalgebraic length dependence of transmission through a chain of barriers with a Lévy spacing distribution". In: *Physical Review B* 79.2 (2009), p. 024204.

[Bei82] Henk van Beijeren. "Transport properties of stochastic Lorentz models". In: *Reviews of Modern Physics* 54.1 (1982), p. 195.

[BS83] Henk van Beijeren and Herbert Spohn. "Transport properties of the one-dimensional stochastic Lorentz Model: I. Velocity autocorrelation function". In: *Journal of Statistical Physics* 31.2 (1983), pp. 231–254.

[BR15] Noam Berger and Ron Rosenthal. "Random walks on discrete point processes". In: *Annales de l'IHP Probabilités et statistiques*. Vol. 51. 2. 2015, pp. 727–755.

[BLP20] Alessandra Bianchi, Marco Lenci, and Françoise Pène. "Continuous-time random walk between Lévy-spaced targets in the real line". In: *Stochastic Processes and their Applications* 130.2 (2020), pp. 708–732.

[Bia+16] Alessandra Bianchi et al. "Random walks in a one-dimensional Lévy random environment". In: *Journal of Statistical Physics* 163.1 (2016), pp. 22–40.

[Bil13] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[BBV11] P Buonsante, R Burioni, and A Vezzani. "Transport and scaling in quenched two-and three-dimensional Lévy quasicrystals". In: *Physical Review E* 84.2 (2011), p. 021105.

[BCV10] Raffaella Burioni, Luca Caniparoli, and Alessandro Vezzani. "Lévy walks and scaling in quenched disordered media". In: *Physical Review E* 81.6 (2010), p. 060101.

[Bur+12] Raffaella Burioni et al. "Scattering lengths and universality in superdiffusive Lévy materials". In: *Physical Review E* 86.3 (2012), p. 031125.

[CF+09] P Caputo, A Faggionato, et al. "Diffusivity in one-dimensional generalized Mott variable-range hopping models". In: *The Annals of Applied Probability* 19.4 (2009), pp. 1459–1494.

[CFP13] P Caputo, A Faggionato, and T Prescott. "Invariance principle for Mott variable range hopping and other walks on point processes". In: *Annales de l'IHP Probabilités et statistiques*. Vol. 49. 3. 2013, pp. 654–697.

[CFG+09] Pietro Caputo, Alessandra Faggionato, Alexandre Gaudillière, et al. "Recurrence and transience for long-range reversible random walks on a random point process". In: *Electronic Journal of Probability* 14 (2009), pp. 2580–2616.

[Cri+14] Giampaolo Cristadoro et al. "Transport properties of Lévy walks: an analysis in terms of multistate processes". In: *EPL (Europhysics Letters)* 108.5 (2014), p. 50002.

[Fel57] William Feller. "An introduction to probability theory and its applications". In: (1957).

[Fel71] William Feller. *An Introduction to Probability theory and its application Vol II.* John Wiley and Sons, 1971.

[Fog94] Hans C Fogedby. "Lévy flights in random environments". In: *Physical review letters* 73.19 (1994), p. 2517.

[GK68] BV Gnedenko and AN Kolmogorov. *Limit Distributions for Sums of Independent Random Variables. Series in Statistics, revised edn.* 1968.

[Gra80] Peter Grassberger. "Velocity autocorrelations in a simple model". In: *Physica A: Statistical Mechanics and its Applications* 103.3 (1980), pp. 558–572.

[GAB12] CW Groth, AR Akhmerov, and CWJ Beenakker. "Transmission probability through a Lévy glass and comparison with a Lévy walk". In: *Physical Review E* 85.2 (2012), p. 021138.

[JS13] Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media, 2013.

[KRS08] Rainer Klages, Günter Radons, and Igor M Sokolov. *Anomalous transport: foundations and applications*. John Wiley & Sons, 2008.

[Kle08] Achim Klenke. *Probability theory. Universitext*. 2008.

[Kub13] Naoki Kubota. "Quenched invariance principle for simple random walk on discrete point processes". In: *Stochastic Processes and their Applications* 123.10 (2013), pp. 3737–3752.

[Lev97] P Levitz. "From Knudsen diffusion to Levy walks". In: *EPL (Europhysics Letters)* 39.6 (1997), p. 593.

[MS18] Marcin Magdziarz and Wladyslaw Szczotka. "Diffusion limit of Lévy-Lorentz gas is Brownian motion". In: *Communications in Nonlinear Science and Numerical Simulation* 60 (2018), pp. 100–106.

[Rad+19] Mattia Radice et al. "Transport properties and ageing for the averaged Lévy–Lorentz gas". In: *Journal of Physics A: Mathematical and Theoretical* 53.2 (2019), p. 025701.

[Rou14] Arnaud Rousselle. "Annealed invariance principle for random walks on random graphs generated by point processes in Rd". In: *Preprint* (2014).

[Rou15] Arnaud Rousselle. "Recurrence or transience of random walks on random graphs generated by point processes in Rd". In: *Stochastic Processes and their Applications* 125.12 (2015), pp. 4351–4374.

[ST94] G Samorodnitsky and MS Taqqu. "Stable Non Gaussian Random Processes (Chapman & Hall, New York)". In: *Stable non-Gaussian random processes. Chapman and Hall, New York.* (1994).

[Sch02] Michael Schulz. "Lévy flights in a quenched jump length field: a real space renormalization group approach". In: *Physics Letters A* 298.2-3 (2002), pp. 105–108.

[SZF95] Micheal F Shlesinger, George M Zaslavsky, and Uriel Frisch. "Lévy flights and related topics in physics". In: (1995).

[Sko56] Anatoly V Skorokhod. "Limit theorems for stochastic processes". In: *Theory of Probability & Its Applications* 1.3 (1956), pp. 261–290.

[Sti+20] Samuele Stivanello et al. "Limit theorems for Lévy flights on a 1D Lévy random medium". In: *arXiv preprint arXiv:2007.03384* (2020).

[VBB19] Alessandro Vezzani, Eli Barkai, and Raffaella Burioni. "Single-big-jump principle in physical modeling". In: *Physical Review E* 100.1 (2019), p. 012108.

[Whi80] Ward Whitt. "Some useful functions for functional limit theorems". In: *Mathematics of operations research* 5.1 (1980), pp. 67–85.

[Whi02] Ward Whitt. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues.* Springer Science & Business Media, 2002.

[ZDK15] V Zaburdaev, S Denisov, and J Klafter. "Lévy walks". In: *Reviews of Modern Physics* 87.2 (2015), pp. 483–530.

[Zei06] Ofer Zeitouni. "Random walks in random environments". In: *Journal of Physics A: Mathematical and General* 39.40 (2006), R433.

[Zhu15] Lingjiong Zhu. "Large deviations for one-dimensional random walks on discrete point processes". In: *Statistics & Probability Letters* 97 (2015), pp. 69–75.