universität wien

MASTERARBEIT

Titel der Masterarbeit

Characterizing weak convergence of partial sum processes to Lévy processes without Gaussian part by random measures.

Verfasser

Fabian Pühringer, BSc BA

angestrebter akademischer Grad

Master of Science (MSc.)

Wien, September 2014

Studienkennzahl lt. Studienblatt:	A 066 821
Studienrichtung lt. Studienblatt:	Mathematik
Betreuer:	Assoz. Prof. Dr. Roland Zweimüller

To Lilly and my beloved parents

Table of Contents

Abstract

Following a paper of Marta Tyran-Kamińska we provide necessary and sufficient conditions for partial sum processes to converge to Lévy processes without Gaussian part in terms of random measures [20], Theorem 3.1. In this context, we give a short introduction to the theory of the space $\mathcal{D}$ of càdlàg functions with Skorokhod's $\mathcal{J}_1$-topology and vague convergence on the space of random measures/point processes. A proof of the Lévy-Itō decomposition using the Lévy-Khintchine formula, as well as Kallenberg's Theorem are presented.

1. Introduction

A main purpose of this work is to also make this subject "accessible" to readers who aren't necessarily probability theorists by heart and to provide a proper reading for them as well. It is thus an attempt to fill the gap between introductory literature, which doesn't bother the reader with too many technical details, and further literature, where one is already supposed to know "these things".
Consequently this text is kept rather simple, in the sense that anyone with basic knowledge of measure and probability theory (and ideally some topology) should be able to read it without any issues. For convenience, the reader will find references and proofs of some – in a subjective point of view – *well-known* results and concepts, which are not covered by basic literature, in the Appendix.

1.1. Notational Issues

The purpose of this small section is to try to avoid difficulties/ambiguities caused by notation as far as possible. Yet it is not extensive but tries to cover some notation which is neither explicitly defined in the course nor "common knowledge".

On basic probabilistic notation: For the rest of this text, we fix some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ for an arbitrary (uncountable) set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure $\mathbb{P}$ on $\mathcal{A}$. For some topological space X, the corresponding Borel σ-algebra is denoted by $\mathcal{B}_X$.

On sequences and convergence: A sequence $(x_n)_{n\geq 1}$ with values in X, i.e. $x_n \in X$ for all $n \geq 1$ is abbreviated by (x_n). This is a quite comfortable way to avoid indices, while still indicating that we are talking about a sequence and not just the element x_n. Furthermore, $x_n \nearrow x$ additionally indicates that $x_1 \leq x_2 \leq \dots$ and $x_n \uparrow x$ means $x_1 < x_2 < \dots$ (similarly $x_n \searrow x$ and $x_n \downarrow x$).

On convergence types: A likely thing to happen in texts on probability theory is the appearance of many different modes of convergence. To keep notation as simple as possible, we indicate them by labeling the arrows. For a sequence (X_n) of random elements and a random element X_0 (all of which taking values in the same measurable space) we write $X_n \overset{a.s.}{\to} X_0$ for almost sure convergence, $X_n \overset{\mathbb{P}}{\to} X_0$ for convergence in probability and $X_n \overset{\mathcal{L}^p}{\to} X_0$ for convergence in $\mathcal{L}^p$. Any new mode of convergence introduced in the text is denoted in the same fashion, with just one important exception: weak convergence. It seems to me that the most common way it is abbreviated in the literature is with a double-arrow, i.e. writing $X_n \Rightarrow X_0$ instead of $X_n \xrightarrow{d} X_0$.

On some special notation: The set of real-valued bounded continuous functions on X is denoted by $\mathcal{C}_b(X)$ and the set of non-negative real-valued continuous functions on X having compact support $supp(f)$ by $\mathcal{C}_c^+(X)$. In metric spaces (S, d), the open ball of radius $r > 0$ around $x \in S$ is denoted by $B_r(x)$ and its closure by $B_r[x]$. The abbreviation λ stands for Lebesgue's measure on $\mathbb{R}$.

Some general remarks: Often we will regard a (real-valued) function on some set as an element or "point" in its corresponding function space, it is thus certainly a good idea to indicate the difference to a point in the underlying space. Therefore I use bold letters for "function points" (and stochastic processes), for instance $\mathbf{x} \in \mathcal{C}(S, \mathbb{R})$.

On multidimensional notation: For $x, y \in \mathbb{R}^d$, $a \in \mathbb{R}$ we don't dwell too much on the fact that actually $x = (x_1, \ldots, x_d)$ and any laborious notation like $\langle x, y\rangle$ or $a \cdot x$ is avoided by simply writing xy and ax in those cases.

1.2. Motivation

Let us start at the beginning. One – if not the – main part of probability theory is to explain why repeated random experiments (given they are somehow nice) tend to show regularity. The most basic results in this context are certainly the Laws of Large Numbers and – even more interesting for our purposes – the Central Limit Theorem. A basic version of the latter reads as follows.

Theorem (Central Limit Theorem, CLT). *Let (X_n) be a sequence of real-valued iid random variables with finite mean $m := \mathbb{E}[X_1]$ and finite variance $\sigma^2 > 0$. Then*

$$\begin{aligned} S_n := \frac{X_1 + \cdots + X_n}{\sqrt{n}} &\Rightarrow N \sim \mathcal{N}(m, \sigma^2) \quad \textit{or, equivalently,} \\ \frac{X_1 + \cdots + X_n - n\mathbb{E}[X_1]}{\sqrt{n\sigma^2}} &\Rightarrow \tilde{N} \sim \mathcal{N}(0,1). \end{aligned} \tag{1.1}$$

Here "$N \sim \mathcal{N}(m, \sigma^2)$" abbreviates "$N$ is normally distributed with mean m and variance σ^2" and "iid" stands for "independent and identically distributed". As already mentioned, the arrow "$\Rightarrow$" indicates convergence in distribution/weak convergence of the sequence of normalized partial sums (S_n) or, alternatively the sequence of its distributions. That is,

$$F_{S_n}(x) := \mathbb{P}[S_n \leq x] \to \mathbb{P}[N \leq x] =: F_N(x) \text{ as } n \to \infty,$$

for every continuity point x of F_N.

Although the CLT is very well-known and may even cause boredom for some, it still *is* a fascinating result! Whatever (let's say so for the moment to emphasize the importance) iid sequence of random events one takes, summing them up and scaling in the right way, leads to a normal distribution in the long run. Repeating a random experiment again and again and looking at the output thus gives *statistic regularity in* *a macroscopic view*, – a phrase taken from Whitt's great introductory book on heavy-tail phenomenas [23]. The challenge of formulating more such theorems has most likely been the core task and motivation of probability theory. Indeed, mathematicians of the preceding centuries have put quite much effort into extending this result. Assuming some more regularity one can for instance drop the assumption that the sequence has to be identically distributed (Lyapunov's CLT, see [6], Theorem 27.3), there are versions of the CLT in higher dimensions (cf. [12], Theorem 15.56) and versions using more general objects than sequences (Lindeberg-Feller CLT, see [12], Theorem 15.43).

We now move to the next level. Instead of just looking at S_n, we could ask ourselves if there is some kind of regularity in the whole path? For sure we cannot expect regularity of the terms in $((S_k)_{k=1}^n)_{n\geq 1}$, since they live on different spaces. But we can try to fit those paths into a common area, like the interval $[0,1]$, to talk about them approaching in a reasonable way. This is achieved by defining stochastic processes $\mathbf{S}^{(n)} := \big(S_t^{(n)}\big)_{t\in[0,1]}$, $n \geq 1$ via

$$S_t^{(n)} := \tfrac{1}{\sqrt{n}}\big(X_1 + \cdots + X_{\lfloor nt\rfloor}\big) \text{ for } t \in [0,1]. \tag{1.2}$$

Here $\lfloor x\rfloor$ denotes the biggest integer not exceeding x. Now the processes $\mathbf{S}^{(n)}$ are made up in such a way, that their paths are functions in $[0,1]$, being constant between the "jumps".
So far the classic Central Limit Theorem only provides "pointwise" convergence, i.e. $S_t^{(n)} \Rightarrow N_t \sim \mathcal{N}(mt, (\sigma t)^2)$ for every *fixed* $t \in [0,1]$, but what about convergence of the whole sequence of *processes* $(\mathbf{S}^{(n)})$? A very intuitive approach to this question derived from looking at clever plots of such paths, which indicate when one can expect such a convergence, can be found in ([23], Chapter 1). In particular, it is evident that we need to center the paths so they don't run away to infinity. By setting $\tilde{X}_n := \frac{X_n - \mathbb{E}[X_1]}{\sigma}$, $n \geq 1$ we can wlog assume that this is indeed the case. Moreover, this normalizes the variance and we can define processes $\tilde{\mathbf{S}}^{(n)} = (\tilde{S}_t^{(n)})_{t\in[0,1]}$ via

$$\tilde{S}_t^{(n)} := \tfrac{1}{\sqrt{n}\sigma}\big(X_1 + \cdots + X_{\lfloor nt\rfloor} - tnm\big). \tag{1.3}$$

The following result is undoubtedly the most famous stochastic process limit theorem.

Theorem (Donsker's Theorem). *Let* $\tilde{\mathbf{S}}^{(n)}$ *be as in* (1.3)*. Then*

$$\tilde{\mathbf{S}}^{(n)} \Rightarrow \mathbf{B},$$

where $\mathbf{B} = (B_t)_{t\in[0,1]}$ *is a (standard) Brownian Motion on* $[0,1]$.

A (standard) Brownian Motion, which we will generically denote by $\mathbf{B}$, is a *continuous* stochastic process having stationary and independent increments (what the French call a PAIS - "processus à accroissement indépendents et stationnaires", see Definition 1.1 below for details), marginal distribution $B_t \sim \mathcal{N}(0,t)$ for $t > 0$ and which satisfies $B_0 = 0$ almost surely (a.s.). Evidently Donsker's Theorem implies the Central Limit Theorem, but it is however a far stronger result and a milestone in the theory of stochastic process limits. Indeed, the Continuous Mapping Theorem allows to use it as basis for many other such results.

Theorem (Continuous Mapping Theorem, CMT). *Let* (X_n) *be a sequence of random elements on a metric space* (S, d)*, s.t.* $X_n \Rightarrow X$ *for some* X *and let* $F : (S, d) \to (S', d')$ *be a map to another metric space which is continuous w.r.t. the distribution of* X*, i.e.* $\mathbb{P}[X \in \{x \in S : F \text{ is discontinuous at } x\}] = 0$*. Then*

$$F(X_n) \Rightarrow F(X).$$

A standard application of this theorem is to take a well-known (stochastic process) limit result, like Donsker's Theorem, and any continuous function F. Then one immediately gets a new limit result $F(X_n) \Rightarrow F(X)$. Another advantage is that by this means one can even carry given results to different spaces.

Again, we make a (this time smaller) step forward and look at more general situations. Although we haven't really emphasized on that point yet, the CLT and Donsker's Theorem fail if the sequence (X_n) is not "nice" enough. Especially they required existence of the first two moments. Thus the terms X_n must have some kind of "light tails", meaning that one can neglect large values, because their odds are extremely small. A classic example is the normal distribution, for which one readily checks that its tails are exponentially bounded. Other examples are exponentially and Gamma distributed random variables.
However, if X_n is for instance Pareto distributed, $X_n \sim \mathit{Pareto}(p)$, i.e. X_n satisfies $\mathbb{P}[X_n > x] = x^{-p}$ for some $p \in [1, 2]$, the Central Limit Theorem and hence Donsker's Theorem both fail. The reason is that the variance is not finite anymore (for $p = 1$ even the mean doesn't exist); they have "heavy tails". In this case the behaviour of the paths is in the long run determined by very few, extraordinarily large values what makes the averaging effect neglectable. Consequently the sequence of partial sum processes tends to be dominated by a few jumps which happen with a *relatively high* probability. Assuming that the sequence (X_n) is not too displeasing, there still is some kind of regularity, although this time the limit process is not continuous any more (for such a result see for instance [16], Theorem 7.1). In fact the limit is a so-called Lévy process.

Definition 1.1. A stochastic process $\mathbf{X} = (X_t)_{t \geq 0}$ is called LÉVY PROCESS if

(i) $X_0 = 0$ a.s.

(ii) $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for $0 \leq t_1 \leq \cdots \leq t_n$, $n \geq 1$ (*independent increments*)

(iii) $X_{t+s} - X_s \stackrel{d}{=} X_t$ for all $s, t \geq 0$ (*stationary increments*)

(iv) $\mathbf{X}$ a.s. has right-continuous paths with left limits.

Here "$\stackrel{d}{=}$" denotes equality in distribution. Functions satisfying condition (*iv*) are discussed in Chapter 2 and we talk more about Lévy processes in general in Chapter 5. For the moment it may help to think of a Lévy process as a Brownian Motion having some jumps. Indeed, we will see that this intuition is not too far away from reality (cf. the Lévy-Itō decomposition, Corollary 5.20).
This new results, which take "heavy-tailed" distributions into account, are in fact very useful in practice. Insurance companies and in particular reinsurances naturally have to

deal with events that occur relatively seldom but have a huge impact on the companies performance. Devastating natural disasters are good examples. Also huge financial crises have the potential to vastly impact the whole financial industry or even economies and states, while being rare events.

1.3. Organization and Relevance

As mentioned before, the Continuous Mapping Theorem is a nice tool in weak convergence theory used to create new limit theorems from existing ones. Another famous approach to prove stochastic limit results are so-called tightness methods, which are mainly due to Prohorov. They use the fact that tightness of a sequence $(\mathbf{X}^{(n)})$ and convergence on a convergence determining class – a notion borrowed from Billingsley ([7], p. 18) – together imply weak convergence. A sequence of random elements (X_n) in a separable metric space (S, d) is hereby called TIGHT if for every $\varepsilon > 0$ there is a compact set K_ε s.t. $\mathbb{P}[X_n \in K_\varepsilon] < \varepsilon$ *uniformly* in n. For random variables $X, X_1, X_2, \ldots$ a class of sets $\mathcal{S}$ is called CONVERGENCE DETERMINING if convergence of the distributions along those sets, i.e. $\mathbb{P}_{X_n}[S] \to \mathbb{P}_X[S]$ for all $S \in \mathcal{S}$, implies $X_n \Rightarrow X$. In the context of stochastic processes, finite dimensional sets often act as a nice convergence determining class and establishing convergence along them is in general far easier than proving a stochastic process limit directly. If one can additionally show tightness of the sequence $(\mathbf{X}^{(n)})$, weak convergence of the whole process is achieved. A key result in this approach is Prohorov's Theorem, characterizing tightness by (weak) relative compactness of the sequence in Polish spaces (Theorem B.5).

The main goal of this text is to serve as an introduction to another approach, which allows proving and disproving weak convergence of partial sum processes to Lévy processes. It is provided by a theorem of Marta Tyran-Kamińska presented in her paper "*Convergence to Lévy stable processes under some weak dependence conditions*" ([20], Theorem 3.1). The main idea behind this approach is to collect jumps of a process in another object based on a two-dimensional plot. Those "new objects" are (Poisson) random measures and our goal is to characterize weak convergence of partial sum processes, similar as the ones in (1.3), to a Lévy process by (weak) convergence of corresponding random measures. An introduction and more details on random measures and Poisson random measures are provided in Chapters 3 and 4.
To establish such a characterization, we inevitably have to specify what it means for random measures to converge, i.e. we have to find a suitable topology. Moreover, we want to switch between processes and measures via the CLT, making it necessary to deal with topologies on both the space of functions satisfying condition (*iv*) in Definition 1.1 and spaces of measures. As already indicated, making those spaces Polish would be a very welcome extra, building bridges to many more results of weak convergence theory (see [7]).

Chapters 2–5 have an introductory character and provide basic properties and some well-known theorems of the corresponding fields. The proof of Marta Tyran-Kamińska's theorem then presented in Chapter 6.

For further reading and applications of the main result, two papers of Marta Tyran-

Kamińska [20] and [21] are highly recommended. The first identifies necessary and sufficient conditions for convergence of partial sum processes of a strictly stationary sequence of random vectors to stable Lévy processes, while the second provides applications to dynamical systems.

2. The Space $\mathcal{D}$ and Skorokhod's Topology

This chapter is devoted to the space $\mathcal{D}$ of càdlàg functions and consists of four sections. The first introduces the space $\mathcal{D}$ and provides some basic results. It was written after some proper reading of Sections 12 and 16 in [7] and the main course is derived from there.

Section 2.2 gives an insight into the richness of $\mathcal{D}$. We create some rather unpleasant càdlàg functions and prove a decomposition result. The main course was created single-handedly but guided by comments and suggestions of my supervisor. Although I know, that those results are all well-known, I unfortunately didn't manage to find any proper references.

The next section motivates the use of Skorokhod's topology on $\mathcal{D}$ to describe convergence of sequences of càdlàg functions. Basic properties of the topology are shown and a proof of the fact that it makes $\mathcal{D}$ a Polish space is partly provided. The results are mostely taken from [7] (especially Chapters 12 and 16) but presented in a slightly different way. Some details to comments appearing in the beginner-friendly book of [23] (especially from Section 3.3) were added.

The main impulse for the last part came from [9] and especially the beginning is fashioned after VI.2, p. 337ff of this book. On the contrary, the very end of this chapter was written in a close collaboration with Prof. Zweimüller and contains "elementary" proofs of some prominent continuity results for maps on $\mathcal{D}$ in the sense that they do not use any tightness criteria in comparison to corresponding results in [9].

2.1. Basic Properties of Càdlàg Functions

As indicated in the introduction a (nice) sequence of partial sum processes whose underlying sequence has "heavy-tails" tends to converge to a process having jumps. Unfortunately dropping the continuity assumption is a rather dangerous thing to do, since highly irregular functions may arise. Luckily limits of partial sum processes as in (1.3) tend to only have "good" discontinuity points. In this context "good" means that they are of the same type as the ones of paths of Lévy processes, see condition (iv) of Definition 1.1. Thus reassured, we start the study of such functions.

Definition 2.1. Let $\mathbf{x} : [0,\infty) \to \mathbb{R}^d$. Assuming they exist, denote by $\mathbf{x}(t+) := \lim_{s \searrow t} \mathbf{x}(s)$ the RIGHT LIMIT OF $\mathbf{x}$ IN t, by $\mathbf{x}(t-) := \lim_{s \nearrow t} \mathbf{x}(s)$ the LEFT LIMIT OF $\mathbf{x}$ IN t and if both exist for a fixed t, write $\Delta\mathbf{x}(t) := \mathbf{x}(t+) - \mathbf{x}(t-)$. We say that $\mathbf{x}$ HAS LEFT or RIGHT LIMITS respectively, given right or left limits exist for all t. If additionally $\mathbf{x}(t+) = \mathbf{x}(t)$ or $\mathbf{x}(t-) = \mathbf{x}(t)$ for all $t \geq 0$ the function $\mathbf{x}$ is RIGHT-CONTINUOUS or LEFT-CONTINUOUS. The set

$$\begin{aligned}
\mathcal{D} := \mathcal{D}_\infty &:= \mathcal{D}([0,\infty), \mathbb{R}^d) \\
&:= \{\mathbf{x} : [0,\infty) \to \mathbb{R}^d : \mathbf{x} \text{ is right-continuous and has left limits}\} \\
&= \{\mathbf{x} : [0,\infty) \to \mathbb{R}^d : \mathbf{x}(t+) \text{ exists, } \mathbf{x}(t+) = \mathbf{x}(t) \text{ for } t \in [0,\infty) \text{ and} \\
&\qquad\qquad \mathbf{x}(t-) \text{ exists for } t \in (0,\infty)\},
\end{aligned}$$

is the space of all $\mathbb{R}^d$-VALUED CÀDLÀG FUNCTIONS ON $[0,\infty)$.

Remark 2.2. (i) The word càdlàg is a French acronym for *continue à droite, limite à gauche*, which is actually more common than its English counterpart *RCLL* standing for "right continuous with left limits".

(ii) $\mathcal{D}$ is a real vector space since for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$ we have $(\mathbf{x}+\mathbf{y})(t-) = \mathbf{x}(t-) + \mathbf{y}(t-)$, $(\mathbf{x}+\mathbf{y})(t+) = \mathbf{x}(t+) + \mathbf{y}(t+)$ and evidently $\gamma\mathbf{x} \in \mathcal{D}$ for all $\gamma \in \mathbb{R}$.

(iii) We generically denote any càdlàg function by $\mathbf{x}$, and as indicated by notation, $\mathbf{x}$ is a point in $\mathcal{D}$. Besides, we sometimes refer to $\mathbf{x}(t)$ as the evaluation of $\mathbf{x}$ at time t in an intuitive way without implying any deeper interpretation.

(iv) The notation extends naturally: $\mathcal{D}([s,t], \mathbb{R}^d)$ is the set of all càdlàg functions defined on $[s,t]$ for $0 \le s < t < \infty$ with values in $\mathbb{R}^d$. Note that we neither require existence of a left limit in s nor right-continuity in t. In particular we denote by $\mathcal{D}_T := \mathcal{D}([0,T], \mathbb{R}^d)$ the $\mathbb{R}^d$-VALUED CÀDLÀG FUNCTIONS ON $[0,T]$. Most of the time we write $\mathcal{D}$ instead of $\mathcal{D}_\infty$ and when we talk about "càdlàg functions", we mean elements of $\mathcal{D}_\infty$ unless stated otherwise.

Example 2.3. (i) Any continuous function is evidently càdlàg, i.e. $\mathcal{C}([0,\infty), \mathbb{R}^d) \subseteq \mathcal{D}([0,\infty), \mathbb{R}^d) = \mathcal{D}$. Especially any continuous $\mathbf{x}$ satisfies $\Delta\mathbf{x}(t) = 0$ for all t and the converse holds as well; if $\Delta\mathbf{x}(t) = 0$ for all $t \ge 0$, then $\mathbf{x}$ is continuous.

(ii) A first prototype of a non-continuous càdlàg function is

$$\mathbf{x} : [0,\infty) \to \mathbb{R}, \qquad s \mapsto \mathbb{1}_{[t,\infty)}(s) = \begin{cases} 1 & \text{if } s \ge t \\ 0 & \text{if } s < t \end{cases} \tag{2.1}$$

for some $t > 0$. Let us check that $\mathbf{x} \in \mathcal{D}$. Obviously $\mathbf{x}$ is continuous everywhere apart from t and the jump in t is of the desired form, since $\lim_{s \searrow 0} \mathbf{x}(t+s) = \lim_{s \searrow 0} 1 = 1 = \mathbf{x}(t)$ and $\lim_{s \searrow 0} \mathbf{x}(t-s) = \lim_{s \searrow 0} 0 = 0$.

As we have seen, càdlàg functions are not continuous in general which makes proving seemingly simple results in many situations far more tedious than in the continuous case. However, the main difference are the jumps and fortunately there are not too many of them; a result we prove using the following, very useful Lemma from Billingsley ([7], Lemma 1, p. 122).

Lemma 2.4. *Let $\varepsilon > 0$, $T \ge 0$ and pick some $\mathbf{x} \in \mathcal{D}_T$. Then there are finitely many points $0 = t_0 < t_1 < \cdots < t_k = T$ s.t.*

$$\sup_{r,s \in [t_{i-1}, t_i)} |\mathbf{x}(s) - \mathbf{x}(r)| \le \varepsilon \text{ for all } 1 \le i \le k. \tag{2.2}$$

Proof. Let $\tilde{t}$ be the supremum of all $t \in [0,T]$ s.t. the interval $[0,t]$ can be decomposed into finitely many subintervals $[t_{i-1}, t_i)$ satisfying (2.2). By right-continuity in 0 we know that $\mathbf{x}(0) = \mathbf{x}(0+)$, so there is some $s_0 > 0$ s.t. $|\mathbf{x}(s) - \mathbf{x}(0)| < \varepsilon$ for all $s \le s_0$ and thus $\tilde{t} \ge s_0 > 0$ (pick for instance $t_0 = 0, t_1 = s_0$ for a decomposition of $[0, s_0]$ satisfying (2.2)).
Since $\tilde{t}$ is the supremum, there is a sequence $\tilde{t}_n \nearrow \tilde{t}$ s.t. the interval $[0, \tilde{t}_n]$ can be decomposed in the desired way for every n. By positivity of $\tilde{t}$, its left limit $\mathbf{x}(\tilde{t}-)$

exists and hence there is some n_0 s.t. $|\mathbf{x}(\tilde{t}_n) - \mathbf{x}(\tilde{t}-)| \leq \varepsilon$ for all $n \geq n_0$. Consequently $[0, \tilde{t}]$ can be decomposed in the desired way (pick $t_0 = 0, t_1, \dots, t_k = \tilde{t}_n$ given by the decomposition of $\tilde{t}_n$ and add $t_{k+1} = \tilde{t}$). Assuming $\tilde{t} = T_0 < T$, right-continuity gives $\mathbf{x}(\tilde{t}) = \mathbf{x}(\tilde{t}+)$ and we can apply the same argument we used to show that $\tilde{t} > 0$ to conclude that $\tilde{t} > T_0$, a contradiction. □

Definition 2.5. Let $\mathbf{x} \in \mathcal{D}$. Then $\mathcal{J}(\mathbf{x}) := \{t \geq 0 : |\Delta\mathbf{x}(t)| > 0\}$ denotes the SET OF POSITIVE JUMPS OF $\mathbf{x}$, $\mathcal{J}_\varepsilon(\mathbf{x}) := \{t \geq 0 : |\Delta\mathbf{x}(t)| > \varepsilon\}$ the SET OF JUMPS OF $\mathbf{x}$ OF SIZE $> \varepsilon$ and $\mathcal{J}_{\varepsilon,T}(\mathbf{x}) := \{t \leq T : |\Delta\mathbf{x}(t)| > \varepsilon\}$ the SET OF JUMPS OF $\mathbf{x}$ IN $[0, T]$ OF SIZE $> \varepsilon$.

Corollary 2.6 (Cardinality of Jumps). *Let $\varepsilon > 0$, $T \geq 0$ and $\mathbf{x} \in \mathcal{D}$. Then $\mathcal{J}_{\varepsilon,T}(\mathbf{x})$ is finite, $\mathcal{J}_\varepsilon(\mathbf{x})$ is discrete and $\mathcal{J}(\mathbf{x})$ is at most countable.*

Proof. The first statement is an immediate consequence of Lemma 2.4 and implies the others, since $\mathcal{J}_\varepsilon(\mathbf{x}) \cap [0, T] = \mathcal{J}_{\varepsilon,T}(\mathbf{x})$ for all $T \geq 0$ and $\mathcal{J}(\mathbf{x}) = \bigcup_{N \geq 1} \bigcup_{n \geq 1} \mathcal{J}_{\frac{1}{n},N}(\mathbf{x})$ is at most countable as a countable union of finite sets. □

Whenever we write $\mathcal{J}_\varepsilon(\mathbf{x}) = \{t_1, t_2, \dots\}$ we usually assume that $t_1 < t_2 < \dots$, i.e. that the points $t_1, t_2, \dots$ are ordered.

Corollary 2.7. *Let $\mathbf{x} \in \mathcal{D}$. Then $\sup_{s \leq T} |\mathbf{x}(s)| < \infty$ for all $T \geq 0$. Càdlàg functions are thus bounded on compact sets.*

Proof. Again, this is a direct consequence of Lemma 2.4. Indeed, if $t_0, \dots, t_k$ is a decomposition of $[0, T]$ satisfying (2.2), then $\|\mathbf{x}\|_{T,\infty} \leq \sum_{i=1}^k |\Delta\mathbf{x}(t_i)| + k\varepsilon < \infty$. □

Definition 2.8. We call $\mathbf{x} \in \mathcal{D}$ a JUMP FUNCTION if it is of the form $\mathbf{x} = \sum_{n \geq 1} a_n \mathbb{1}_{[t_n,\infty)}$ for some $t_n \in (0, \infty)$, $a_n \in \mathbb{R}^d$, $n \geq 1$. If $a_n = 1$ for all n we say that $\mathbf{x}$ is a PURE JUMP FUNCTION. Note that the set $\{t_n, n \geq 1\}$ is necessarily discrete by Corollary 2.6 and we can thus assume wlog that it is ordered.

The following lemma is a useful tool to construct interesting càdlàg functions.

Lemma 2.9 (Uniform limit of càdlàg functions). *If $(\mathbf{x}_n)$ is a sequence in $\mathcal{D}$ and $\mathbf{x}_n \to \mathbf{x}$ uniformly, in symbols $\mathbf{x}_n \overset{u}{\to} \mathbf{x}$, then $\mathbf{x} \in \mathcal{D}$.*

Proof. We claim that $\lim_{s \searrow t} \mathbf{x}(s) = \mathbf{x}(t)$ for all fixed $t \geq 0$. To obtain this, note that $\mathbf{x}_n$ converges uniformly to $\mathbf{x}$ and $\lim_{s \searrow t} \mathbf{x}_n(s)$ exists for all n. The double-limit $\lim_{s \searrow t} \lim_{n \to \infty} \mathbf{x}_n(s)$ is well-defined by the Moore-Osgood Theorem (Theorem E.1) and moreover exchanging its order is permited. Thus

$$\begin{aligned}\mathbf{x}(t+) &= \lim_{s \searrow t} \mathbf{x}(s) = \lim_{s \searrow t} \lim_{n \to \infty} \mathbf{x}_n(s) = \lim_{n \to \infty} \lim_{s \searrow t} \mathbf{x}_n(s) = \lim_{n \to \infty} \mathbf{x}_n(t+) = \lim_{n \to \infty} \mathbf{x}_n(t) \\ &= \mathbf{x}(t).\end{aligned}$$

A very similar argument shows that $\lim_{s \nearrow t} \mathbf{x}(s)$ exists, hence $\mathbf{x} \in \mathcal{D}$. □

2.2. Decomposing Càdlàg Functions

By now, we only had some basic examples of càdlàg maps like jump functions which only have finitely many discontinuities in compact intervals. They are rather easy to handle analytically hence it would be pleasant to decompose any $\mathbf{x} \in \mathcal{D}$ into a continuous part $\mathbf{x}^{(c)} \in \mathcal{C}([0,\infty), \mathbb{R}^d)$ and a jump function $\mathbf{x}^{(j)} \in \mathcal{D}$. But càdlàg functions are not that simple in general, indeed the following shows that such a decomposition is not always possible.

Example 2.10 ($\mathbf{x} = \mathbf{x}^{(c)} + \mathbf{x}^{(j)}$ fails). For $n \geq 1$ let $\mathbf{x}_n := \sum_{k=1}^{2^{n-1}} \frac{(-1)^{k+1}}{2^{n-1}} \mathbb{1}_{[\frac{2k-1}{2^n},1]} \in \mathcal{D}_1$, that is a jump function which has 2^{n-1} jumps of size $\frac{1}{2^{n-1}}$ with altering sign in $[0,1]$ (see Figure 2.1). It is easy to see that all $\mathcal{J}(\mathbf{x}_n)$ are pairwise disjoint and $\|\mathbf{x}_n\|_\infty \leq \frac{1}{2^{n-1}}$. Now $\mathbf{x} := \sum_{n\geq 1} \mathbf{x}_n \in \mathcal{D}_1$ by Lemma 2.9, since for $\mathbf{x}^{(N)} := \sum_{n=1}^{N} \mathbf{x}_n \in \mathcal{D}_1$, we have $\|\mathbf{x} - \mathbf{x}^{(N)}\|_\infty = \|\sum_{n\geq N+1} \mathbf{x}_n\|_\infty \leq \sum_{n\geq N+1} \|\mathbf{x}_n\|_\infty \leq \frac{1}{2^N} \to 0$, thus $\mathbf{x}^{(N)} \xrightarrow{u} \mathbf{x}$.
If we try to write $\mathbf{x} = \mathbf{x}^{(c)} + \mathbf{x}^{(j)}$ for $\mathbf{x}^{(c)} \in \mathcal{C}([0,1], \mathbb{R})$ and a jump function $\mathbf{x}^{(j)}$, then we need to collect all discontinuities in the latter, i.e. $\mathbf{x}^{(j)}(\cdot) = \sum_{s\leq \cdot} \Delta\mathbf{x}(s) \mathbb{1}_{[\,\cdot\,,\infty)}$. But this series doesn't converge absolutely! Indeed, for any interval $\emptyset \neq I = [a,b] \subseteq [0,1]$ we have $\sum_{s\in I} |\Delta\mathbf{x}(s)| = \sum_{s\in I}\sum_{n\geq 1} |\Delta\mathbf{x}_n(s)| = \sum_{n\geq 1}\sum_{s\in I} |\Delta\mathbf{x}_n(s)| \approx \sum_{n\geq 1} (b-a)2^{n-1}\frac{1}{2^{n-1}} = \infty$. Here we used that the $\mathcal{J}(\mathbf{x}_n)$ are pairwise disjoint and that $\mathbf{x}_n$ (asymptotically) has $(b-a)2^{n-1}$ jumps of size 2^{n-1} in I.

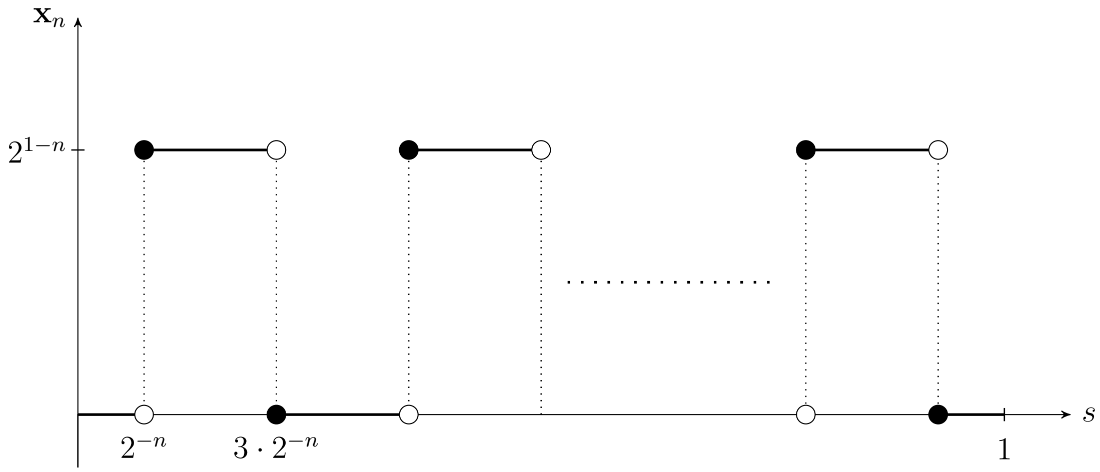

Figure 2.1: A plot of the jump function $\mathbf{x}^{(n)}$.

Since our first attempt failed, let's try to find a more clever decomposition. We may suggest that it was crucial that the jumps of $\mathbf{x}_n$ were of alternating sign, which allowed the function $\mathbf{x}$ to have a large variation while staying bounded. A new way of decomposing $\mathbf{x}$ could thus be $\mathbf{x} = \mathbf{x}^{(c)} + \sum_{n\geq 1} \mathbf{x}_n^{(j)}$ for a continuous $\mathbf{x}^{(c)}$ and a sequence of jump functions $\mathbf{x}_n^{(j)}$. This evidently works in the above example, but unfortunately is not true in general.

Example 2.11 ($\mathbf{x} = \mathbf{x}^{(c)} + \sum_{n\geq 1} \mathbf{x}_n^{(j)}$ fails). We choose a similar function as above, but instead of alternating the sign of the jumps, we let them be positive and compensate them in an appropriate way. To achieve this, let $\mathbf{x}_n := \sum_{k=1}^{2^{n-1}} \frac{1}{2^{n-1}} \mathbb{1}_{[\frac{2k-1}{2^n},1]} \in \mathcal{D}_1$.

Again the sets $\mathcal{J}(\mathbf{x}_n)$ are pairwise disjoint but now $\|\mathbf{x}_n\|_\infty = \mathbf{x}_n(1) = 1$. A clever continuous modification assures that the sequence still is small in the uniform norm. Define $\tilde{\mathbf{x}}_n(s) := \mathbf{x}_n(s) - s$ (see Figure 2.2), then one readily checks that $\|\tilde{\mathbf{x}}_n\|_\infty \leq \frac{1}{2^n}$ and we can use the same argument as in the previous example to conclude that $\mathbf{x} := \sum_{n\geq 1} \tilde{\mathbf{x}}_n \in \mathcal{D}_1$. Again, for I as above we know that $\tilde{\mathbf{x}}_n$ asymptotically has $(b-a)2^{n-1}$ positive jumps of size $\frac{1}{2^{n-1}}$ in I. To collect all jumps of $\mathbf{x}$, we define $\mathbf{x}_n^{(j)} := \mathbf{x}_n$. Then the increment $|\sum_{n\geq 1} \mathbf{x}_n^{(j)}(b) - \sum_{n\geq 1} \mathbf{x}_n^{(j)}(a)| = \sum_{s\in I} \Delta \mathbf{x}_n^{(j)}(s) \approx \sum_{n\geq 1}(b-a)2^{n-1}\frac{1}{2^{n-1}} = \infty$ and the decomposition breaks down.

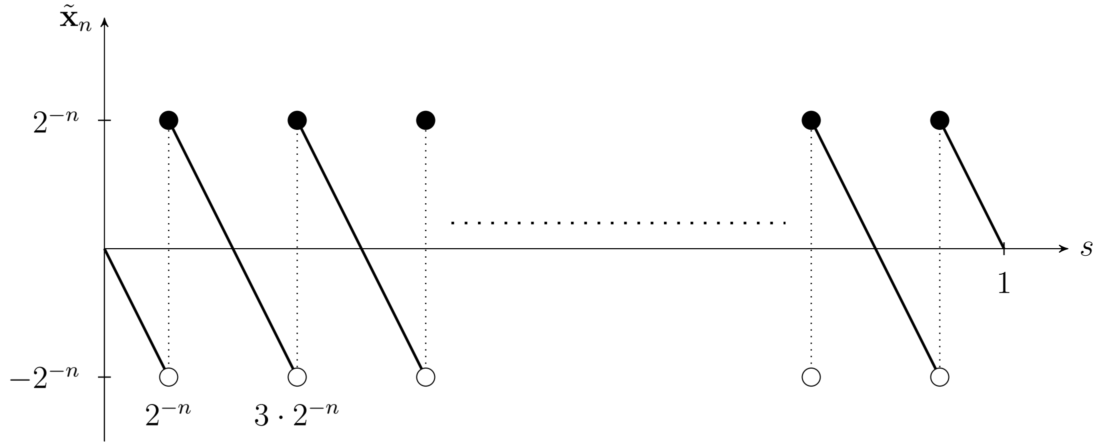

Figure 2.2: A plot of the jump function $\tilde{\mathbf{x}}_n$ which has only positive jumps.

The fault in the last attempt was to focus too much on the jumps and ignore the impact of the continuous part. In the preceding example, it is chosen in such a way that it avoids an accumulation of discontinuities and allows $\mathbf{x}$ to have have solely positive jumps and an infinite variation while staying bounded. If we take this thoughts into account, we finally get a possible decomposition.

Proposition 2.12 (Decomposition of càdlàg functions). *For any $\mathbf{x} \in \mathcal{D}$ there are continuous functions $\mathbf{x}^{(c)}$, $\mathbf{x}_n^{(c)}$ and jump functions $\mathbf{x}_n^{(j)}$ for $n \geq 0$ s.t.*

$$\mathbf{x} = \mathbf{x}^{(c)} + \sum_{n\geq 0} (\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)}) \text{ uniformly.} \tag{2.3}$$

Proof. Let $\mathbf{x} \in \mathcal{D}$, fix $N \geq 0$ and restrict $\mathbf{x}$ to $[0, N]$, where N is s.t. $\Delta\mathbf{x}(N) = 0$. Note that one can use almost all N due to Corollary 2.6.
First put all jumps of size > 1 in a jump function $\mathbf{x}_0^{(j)} := \sum_{t\in\mathcal{J}_1(\mathbf{x})} \Delta\mathbf{x}(t)\mathbb{1}_{[t,\infty)}$ in $\mathcal{D}_N$ and set $\mathbf{x}_0^{(c)} := 0$. The function $\mathbf{x}_0^{(j)}$ is bounded on $[0, N]$ and thus there is some M s.t. $\|\mathbf{x}_0^{(j)}\|_\infty \leq M$. To collect all small jumps, define $\mathcal{I}_n := \mathcal{J}_{\frac{1}{2^n}} \setminus \mathcal{J}_{\frac{1}{2^{n-1}}}$ and set $\mathbf{x}_n^{(j)} := \sum_{t\in\mathcal{I}_n(\mathbf{x})} \Delta\mathbf{x}(t)\mathbb{1}_{[t,\infty)}$ in $\mathcal{D}_N$ for $n \geq 1$. Note that the sets $\mathcal{I}_n$ are pairwise pairwise disjoint and that $\mathbf{x}_n^{(j)}$ contains all jumps of $\mathbf{x}$ of size in $(\frac{1}{2^n}, \frac{1}{2^{n-1}}]$.
We have to compensate the jumps in an appropriate way to avoid a situation as in Example 2.11. Let us therefore assume for the moment that we have a sequence of continuous functions $(\mathbf{x}_n^{(c)})$ such that $\|\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)}\|_\infty \leq \frac{1}{2^{n-1}}$ for all $n \geq 1$. We claim that

$\mathbf{x}^{(c)} := \mathbf{x} - \sum_{n\geq 0}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})$ is a valid decomposition. Why is this true? Using Lemma 2.9 one readily sees that $\tilde{\mathbf{x}} := \sum_{n\geq 0}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})$ is càdlàg, hence $\mathbf{x}^{(c)} = \mathbf{x} - \tilde{\mathbf{x}}$ is càdlàg as well. Now

$$\begin{aligned}\Big\|\mathbf{x} - \mathbf{x}^{(c)} - \sum_{n=0}^{m}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})\Big\|_\infty &\overset{\text{Def. } \mathbf{x}^{(c)}}{=} \Big\| \sum_{n\geq m+1}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})\Big\|_\infty \\ &\leq \sum_{n\geq m+1} \underbrace{\|(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})\|_\infty}_{\leq \frac{1}{2^{n+1}}} \to 0,\end{aligned}$$

as $m \to \infty$, since the last term is the tail of a convergent sequence. This shows (2.3) provided $\mathbf{x}^{(c)}$ is continuous. Assuming the contrary, there is $t \in [0, N]$ s.t. $|\Delta\mathbf{x}^{(c)}(t)| > 0$. Then $|\Delta\mathbf{x}^{(c)}(t)| = |\Delta(\mathbf{x} - \tilde{\mathbf{x}})(t)| = |\Delta\mathbf{x}(t) - \Delta\tilde{\mathbf{x}}(t)| > 0$, a contradiction. Indeed $\Delta\mathbf{x}(s) = \Delta\tilde{\mathbf{x}}(s)$ for all $s \in [0, N]$ by construction, since whenever $\Delta\mathbf{x}(s) = 0$, so is $\Delta\mathbf{x}_n^{(j)}(s)$ for all n (i.e. $s \notin \bigcup_{n\geq 1}\mathcal{I}_n$) and whenever $|\Delta\mathbf{x}(s)| > 0$, there is exactly one $m \in \mathbb{N}$ s.t. $|\Delta\mathbf{x}(s)| \in \mathcal{I}_m$, thus $\Delta\mathbf{x}(s) = \Delta\mathbf{x}_m^{(j)}(s)$ and $\Delta\mathbf{x}_n^{(j)}(s) = 0$ for all $n \neq m$ by definition.
Moreover, problems as in the first two counterexamples cannot occur, since on any interval $I = [a, b]$ as above, $\mathbf{x}^{(c)}$ is bounded and $|\tilde{\mathbf{x}}(b) - \tilde{\mathbf{x}}(a)| = |\sum_{n\geq 0}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})(b) - \sum_{n\geq 0}(\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)})(a)| \leq \sum_{n\geq 0}\|\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)}\|_\infty \leq 2 + M < \infty$.
Now to the construction of $\mathbf{x}_n^{(c)}$. Note that the set $\mathcal{I}_n \cap [0, N] := \{t_n^{(1)}, \ldots, t_n^{(m_n)}\}$ is finite and consequently there is a positive distance $\delta := \delta(n, N)$ between the jumps. Interpolate the points

$$\begin{gathered}\mathbf{x}_n^{(c)}(0) = 0, \quad \mathbf{x}_n^{(c)}\big(\tfrac{t_n^{(1)}}{2}\big) = 0, \quad \mathbf{x}_n^{(c)}\big(\tfrac{t_n^{(2)} - t_n^{(1)}}{2}\big) = \mathbf{x}_n^{(j)}\big(t_n^{(1)}\big), \;\cdots \\ \mathbf{x}_n^{(c)}\big(\tfrac{t_n^{(m_n)} - t_n^{(m_n-1)}}{2}\big) = \mathbf{x}_n^{(j)}\big(t_n^{(m_n-1)}\big), \quad \mathbf{x}_n^{(c)}(N) = \mathbf{x}_n^{(j)}\big(t_n^{(m_n-1)}\big)\end{gathered}$$

linearly. This is just a linear connection of the half points of the steps of $\mathbf{x}_n^{(j)}$ (see Figure 2.3) and thus their difference cannot exceed the size of the jumps, i.e. $\|\mathbf{x}_n^{(j)} - \mathbf{x}_n^{(c)}\|_\infty \leq \frac{1}{2^{n-1}}$ on $[0, N]$. Since N was (almost) arbitrary we are done. $\square$

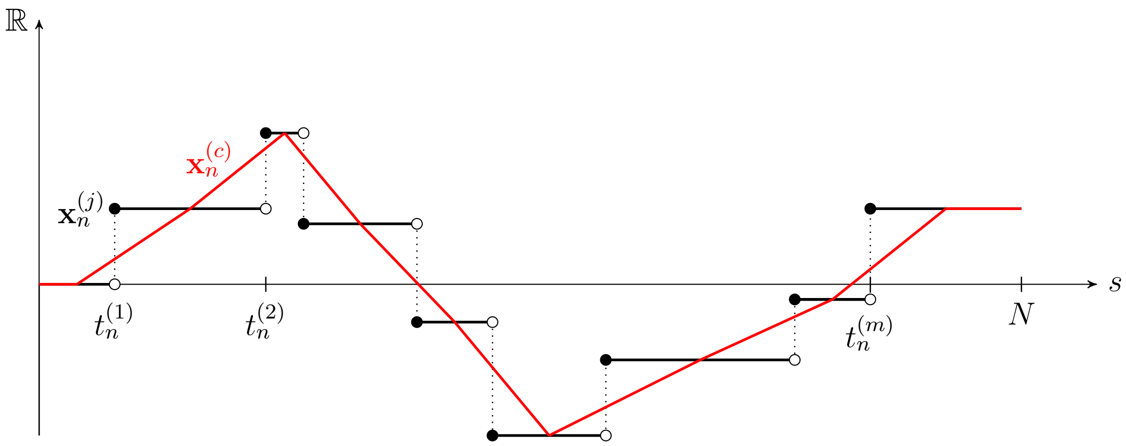

Figure 2.3: Construction of the continuous approximation $\mathbf{x}_n^{(c)}$ of $\mathbf{x}_n^{(j)}$.

2.3. The ABC of Skorokhodology

Our main goal is to talk about weak convergence of partial sum processes to Lévy processes in terms of a corresponding convergence in another space. To "translate" those two concepts with a function we want to use the Continuous Mapping Theorem. Therefore we need to be able to talk about continuity and make $\mathcal{D}$ a topological space. However, there are many different topologies on $\mathcal{D}$ to choose from, some of which are well-known from basic analysis, like the uniform topology, while others, which are mainly due to Skorokhod, are more specific. The one we are going to use is the so-called Skorokhod J_1-topology. Since we won't consider any other, we drop the prefix J_1 and rather talk about *the* Skorokhod topology. Before we actually introduce it, let's see why the uniform topology on $\mathcal{D}$ is not useful for our purposes.

Example 2.13 (Shortcoming of the uniform topology on $\mathcal{D}$). Fix some $t \in (1,\infty)$. In the setting of càdlàg functions, we most certainly want $(\mathbf{x}_n) := (\mathbb{1}_{[t_n,\infty)})$ to converge to $\mathbf{x} := \mathbb{1}_{[t,\infty)}$, given $t_n \to t$. If for example $t_n := t + \frac{1}{n} \to t$ we see that the sequence $(\mathbf{x}_n)$ somehow approaches $\mathbf{x}$, however $\|\mathbf{x}_n - \mathbf{x}\|_\infty \geq |\mathbf{x}_n(t) - \mathbf{x}(t)| = |0-1| = 1$ for all n.

Apparently we need a better way to describe convergence in $\mathcal{D}$. But how could we do that? As we have seen, the major problem is that whenever there is a small gap between discontinuity points (of the same height), the uniform distance exceeds the size of this jump. Even if the jumping-times converge together (in a non-trivial way) there is no convergence in the uniform topology. A possible solution are small "time-shifts", which allow moving a jump of $\mathbf{x}_n$ to a close but different place, like a discontinuity of $\mathbf{x}$.
In the preceding example it is easy to find such a sequence $(\tilde{\lambda}_n)$ which maps the jump of $\mathbf{x}_n$ onto the respective jump in $\mathbf{x}$. Indeed, we can just interpolate the points $\tilde{\lambda}_n(0) = 0$ and $\tilde{\lambda}_n(t) = t_n$ linearly and set $\tilde{\lambda}_n(s) = s + (t_n - t)$ for $s \geq t$. Then $\mathbf{x}_n \circ \tilde{\lambda}_n = \mathbb{1}_{[t_n,\infty)} \circ \tilde{\lambda}_n = \mathbb{1}_{[t,\infty)} = \mathbf{x}$ for all n. This motivates the following.

Definition 2.14. Let $\Lambda := \Lambda_\infty := \{\lambda : \lambda \text{ is a homeomorphism on } [0,\infty)\} = \{\lambda \in \mathcal{C}([0,\infty),[0,\infty)) : \lambda \text{ is strictly increasing, } \lambda(0) = 0 \text{ and } \lambda(n) \to \infty, \text{ as } n \to \infty\}$ be the SET OF TIME-SHIFTS. A sequence $(\mathbf{x}_n)$ in $\mathcal{D}$ CONVERGES IN THE SKOROKHOD TOPOLOGY TO $\mathbf{x}$ (or simply CONVERGES TO $\mathbf{x}$ IN $\mathcal{D}$), in symbols $\mathbf{x}_n \rightsquigarrow \mathbf{x}$, for some $\mathbf{x} \in \mathcal{D}$, if there is a sequence (λ_n) in Λ, s.t.

$$\|\lambda_n - \mathrm{Id}\,\|_\infty = \sup_{s\geq 0} |\lambda_n(s) - s| \to 0 \text{ and} \tag{2.4}$$

$$\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{N,\infty} := \sup_{s\in[0,N]} |\mathbf{x}_n \circ \lambda_n(s) - \mathbf{x}(s)| \to 0 \text{ for all } N \geq 0. \tag{2.5}$$

To simplify notation, write $\|\mathbf{x}\|_{N,\infty} = \sup_{s\in[0,N]} |x(s)|$ for general $\mathbf{x} \in \mathcal{D}$ and $\overset{ucs}{\to}$ if uniform convergence holds on compact sets. We can thus rewrite (2.4) as $\lambda_n \overset{u}{\to} \mathrm{Id}$ and (2.5) becomes $\mathbf{x}_n \circ \lambda_n \overset{ucs}{\to} \mathbf{x}$. If we want to emphasize that convergence holds along a specific sequence (λ_n), we say $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n).

Remark 2.15. (i) This mode of convergence naturally extends to a sequence in $\mathcal{D}_T$ by using time-shifts in $\Lambda_T := \{\lambda : \lambda \text{ is a homeomorphism on } [0,T]\}$ instead. It is worth mentioning, that necessarily $\lambda(T) = T$ for all $\lambda \in \mathcal{D}_T$.

(ii) If $\lambda_1, \lambda_2 \in \Lambda$, then $\lambda_1 \circ \lambda_2 \in \Lambda$ and $\lambda \in \Lambda$ iff $\lambda^{-1} \in \Lambda$.

(iii) In Example 2.13 we had $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via $(\tilde{\lambda}_n)$ but not via $(\lambda_n) = (Id)$, although both converge to the identity in the uniform topology. Hence we cannot just take any sequence being close to the identity to get convergence of the shifted functions, condition (2.5) is crucial.

(iv) The mode of convergence $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ defines a topology, the SKOROKHOD TOPOLOGY ON $\mathcal{D}$. It is weaker than the uniform topology, since $\mathbf{x}_n \xrightarrow{u} \mathbf{x}$ implies $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (Id).

So far this concept seems to be rather "natural". It is certainly reasonable to not only measure differences of the values of a function, but to allow a (uniformly) small deformation of the time scale to suit the structure of $\mathcal{D}$. Yet, the Skorokhod topology has some properties which complicate many seemingly simple things. We are going to list a few inconveniences in the following according to [23], Section 3.3.

Example 2.16 (No approximation by continuous functions). Let $\mathbf{x} := \mathbb{1}_{[\frac{1}{2},1]} \in \mathcal{D}_1$ and $\mathbf{x}_n$ have value 0 on $[0, \frac{1}{2} - \frac{1}{n}]$ and 1 on $[\frac{1}{2}, 1]$. In between we interpolate linearly (i.e. $\mathbf{x}_n(s) := n(s - \frac{1}{2} + \frac{1}{n})\mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2}]} + \mathbb{1}_{[\frac{1}{2},1]}$, see Figure 2.4). Then $\mathbf{x}_n \not\rightsquigarrow \mathbf{x}$, because for any $\lambda_n \in \Lambda_1$ the function $\mathbf{x}_n \circ \lambda_n$ is continuous and thus assumes all values between 0 and 1 for all n. Thus $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_\infty \geq \frac{1}{2}$ for any $\lambda_n \in \Lambda_1$, $n \geq 1$.

Example 2.17 (No approximation by double-jumps). Above, convergence in the Skorokhod topology failed, due to a lack of jumps in the approaching sequence. But also double-jumps can spoil convergence. Let $\mathbf{x}$ be as above and look at the sequence $\mathbf{x}_n := \frac{1}{2}\big(\mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2})} + \mathbb{1}_{[\frac{1}{2},1]}\big)$, $n \geq 1$ (see Figure 2.4). Then $\mathbf{x}_n \in \mathcal{D}_1$ for all n and it approaches $\mathbf{x}$ with a double-jump. But $\mathbf{x}_n \not\rightsquigarrow \mathbf{x}$, since the discontinuity of the terms $\mathbf{x}_n$ at the time $t_n := \frac{1}{2} - \frac{1}{n}$ requires $\mathbf{x}$ to have a jump of size $\frac{1}{2}$ at time $t = \frac{1}{2}$.
More properly, for all $n \geq 1$ the function $\mathbf{x}_n$ has value $\frac{1}{2}$ on the interval $[\frac{1}{2} - \frac{1}{n}, \frac{1}{2})$, hence $\mathbf{x}_n \circ \lambda_n(s) = \frac{1}{2}$ for $s \in \lambda_n^{-1}([\frac{1}{2} - \frac{1}{n}, \frac{1}{2})) \neq \emptyset$. Consequently $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_\infty \geq \frac{1}{2}$ for all $\lambda_n \in \Lambda_1$, since $\mathbf{x}$ only assumes the values 0 and 1.

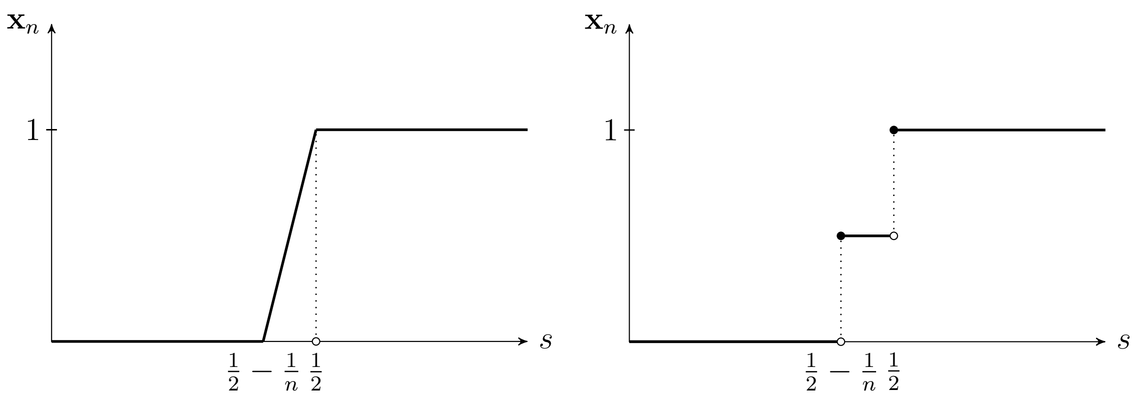

Figure 2.4: Both approaches fail due to a lack/an abundance of jumps.

What happened in the previous two examples is a result of the "matching jump" property of the Skorokhod topology, i.e. whenever $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) and $|\Delta\mathbf{x}(t)| > 0$, there must be a sequence $t_n \to t$ s.t. $\Delta\mathbf{x}_n(t_n) \to \Delta\mathbf{x}(t)$. In fact $t_n := \lambda_n(t)$ does the job (see Proposition 2.31 (*i*)). The converse is also true: if there is a sequence $t_n \to t$ s.t. $|\Delta\mathbf{x}_n(t_n)| \geq \varepsilon > 0$ and $\mathbf{x}_n \rightsquigarrow \mathbf{x}$, then $\mathbf{x}$ must have a jump of size $\geq \varepsilon$ at time t (see Proposition 2.31 (*ii*)).

Example 2.18 (Failing convergence of restricted functions). For $(\mathbf{x}_n)$ and $\mathbf{x}$ from 2.13 we have seen that $\mathbf{x}_n \rightsquigarrow \mathbf{x}$. But $\tilde{\mathbf{x}}_n \not\rightsquigarrow \tilde{\mathbf{x}}$ for $\tilde{\mathbf{x}}_n := \mathbf{x}_n\big|_{[0,t]}$ and $\tilde{\mathbf{x}} := \mathbf{x}\big|_{[0,t]}$, since $\tilde{\mathbf{x}}_n \equiv 0$ for all n (what actually makes time-shifting an odd thing to do), while $\tilde{\mathbf{x}}(t) = 1$. Now since any $\lambda_n \in \Lambda_t$ has to satisfy $\lambda(t) = t$, we have $\|\tilde{\mathbf{x}}_n \circ \lambda_n - \tilde{\mathbf{x}}\|_{N,\infty} \geq |\tilde{\mathbf{x}}_n(\lambda_n(t)) - \tilde{\mathbf{x}}(t)| = |\tilde{\mathbf{x}}_n(t) - \tilde{\mathbf{x}}(t)| = |0-1| = 1$ for all n and some $N \geq t$. Thus $\tilde{\mathbf{x}}_n \not\rightsquigarrow \tilde{\mathbf{x}}$. Here the only problem is, that $\mathbf{x}$ (or $\tilde{\mathbf{x}}$) is discontinuous at t and we couldn't get rid of the jump at the edge.

Luckily those things are not too dramatic. On the one hand we could avoid the "matching jumps" property by using another Skorokhod topology (like the M_1-topology). We will soon explain why we won't do that though. On the other hand convergence of càdlàg functions restricted to $[0,T]$ only fails if $|\Delta\mathbf{x}(T)| > 0$.

Lemma 2.19. *If* $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ *via* (λ_n)*, then* $\mathbf{x}_n\big|_{[0,T]} \rightsquigarrow \mathbf{x}\big|_{[0,T]}$ *iff* $\mathbf{x}$ *is continuous in* T.

Proof. ($\Rightarrow$) One only has to slightly adapt the argument of the preceding example to show that convergence fails if $|\Delta\mathbf{x}(T)| > 0$.

($\Leftarrow$) In the example above we have seen that the main problem is, that even though $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) it is possible that $\lambda_n\big|_{[0,T]} \notin \Lambda_T$ because $\lambda_n(T) = T$ may fail for all n. If $|\Delta\mathbf{x}(T)| = 0$ we claim that the sequence (λ_n) can be used to find $(\tilde{\lambda}_n)$ in Λ_T s.t. convergence of the restricted sequence holds via $(\tilde{\lambda}_n)$.
Fix some $\varepsilon > 0$. By continuity of $\mathbf{x}$ in T there is $\delta > 0$ s.t. $\sup_{r,s\in[T-3\delta,T+3\delta]} |\mathbf{x}(s) - \mathbf{x}(r)| \leq \frac{\varepsilon}{4}$. Assume wlog that $\|\lambda_n - \mathrm{Id}\,\|_\infty \leq \frac{\delta}{2}$ and $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{T+3\delta,\infty} < \frac{\varepsilon}{4}$ for all n, let $\tilde{\lambda}_n(s) = \lambda_n(s)$ for $s \in [0, T-\delta]$ and interpolate $\tilde{\lambda}_n(T-\delta) = \lambda_n(T-\delta) \leq (T-\delta+\frac{\delta}{2}) < T$ linearly with $\tilde{\lambda}_n(T) = T$. This assures that $\tilde{\lambda}_n \in \Lambda_T$ and $\tilde{\lambda}_n \xrightarrow{u} \mathrm{Id}$, since $\|\tilde{\lambda}_n - \mathrm{Id}\,\|_\infty \leq \|\lambda_n - \mathrm{Id}\,\|_\infty$. Particularly we have $\|\tilde{\lambda}_n - \mathrm{Id}\,\|_\infty \leq \frac{\delta}{2}$ for all n.
We claim that $\sup_{r,s\in[T-2\delta,T+2\delta]} |\mathbf{x}_n(s) - \mathbf{x}_n(r)| \leq \frac{\varepsilon}{2}$ for all n. Assuming the contrary implies that there are some $r_n, s_n \in [T-3\delta, T+3\delta]$ s.t. $|\mathbf{x}_n \circ \lambda_n(r_n) - \mathbf{x}_n \circ \lambda_n(s_n)| > \frac{\varepsilon}{2}$. Then $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{T+3\delta,\infty} \geq \sup_{t\in[T-3\delta,T+3\delta]} |\mathbf{x}_n \circ \lambda_n - \mathbf{x}| \geq |\mathbf{x}_n \circ \lambda_n(r_n) - \mathbf{x}(r_n)| \vee |\mathbf{x}_n \circ \lambda_n(s_n) - \mathbf{x}(s_n)| \geq \frac{\varepsilon}{4}$, contradicting our assumption. Finally,

$$
\begin{aligned}
\|\mathbf{x}_n\big|_{[0,T]} \circ \tilde{\lambda}_n - \mathbf{x}\big|_{[0,T]}\|_{T,\infty} &= \|\mathbf{x}_n \circ \tilde{\lambda}_n - \mathbf{x}\|_{T,\infty} \\
&\leq \|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{T-\delta,\infty} + \sup_{s\in[T-\delta,T]} |\mathbf{x}_n \circ \tilde{\lambda}_n(s) - \mathbf{x}(s)| \\
&\leq \tfrac{\varepsilon}{4} + \sup_{s\in[T-\delta,T]} |\mathbf{x}_n \circ \tilde{\lambda}_n(s) - \mathbf{x}_n \circ \lambda_n(s)| \\
&\qquad + \sup_{s\in[T-\delta,T]} |\mathbf{x}_n \circ \lambda_n(s) - \mathbf{x}(s)| \\
&\leq \tfrac{\varepsilon}{2} + \sup_{r,s\in[T-2\delta,T+2\delta]} |\mathbf{x}_n(s) - \mathbf{x}_n(r)| \leq \varepsilon
\end{aligned}
$$

for all n and hence $\mathbf{x}_n\big|_{[0,T]} \rightsquigarrow \mathbf{x}\big|_{[0,T]}$ via $(\tilde{\lambda}_n)$. $\square$

Remark 2.20. Rephrase the above statement in terms of the map $Rest_T : \mathcal{D} \to \mathcal{D}_T$, $\mathbf{x} \mapsto \mathbf{x}\big|_{[0,T]}$. Then Lemma 2.19 becomes: $Rest_T$ is continuous in $\mathbf{x}$ iff T is a continuity point of $\mathbf{x}$.

Proposition 2.21. *The Skorokhod topology is metrizable.*

Proof. The proof of this result is fashioned after [7], p. 123ff and Theorem 16.2.
(i) Fix $T > 0$ and let us first show that we can metricize $\mathcal{D}_T$. For $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{D}_T$ we can use the definition of convergence, to define a designated distance function

$$d_T(\mathbf{x}, \mathbf{y}) := \inf_{\lambda \in \Lambda_T} \{\|\lambda - \operatorname{Id}\|_\infty \vee \|\mathbf{x} - \mathbf{y} \circ \lambda\|_\infty\}.$$

This is most certainly a natural first guess. Later we will see, that it is in fact not the cleverest way to metricize $\mathcal{D}$. However, let us now check that this is indeed a metric.
Evidently $d_T(\mathbf{x}, \mathbf{y})$ is non-negative and if we take $\lambda = \operatorname{Id}$, the fact that càdlàg functions on $[0, T]$ are bounded (Corollary 2.7), implies $d_T(\mathbf{x}, \mathbf{y}) \in [0, \infty)$. Again taking $\lambda = \operatorname{Id}$ shows that $d_T(\mathbf{x}, \mathbf{x}) = 0$ and in turn $d_T(\mathbf{x}, \mathbf{y}) = 0$ for $\mathbf{x}, \mathbf{y} \in \mathcal{D}_T$ means that both terms in the infimum have to vanish. Hence there is a sequence (λ_n) in Λ_T s.t. $\lambda_n \stackrel{u}{\to} \operatorname{Id}$ and $\mathbf{y} \circ \lambda_n \stackrel{u}{\to} \mathbf{x}$. For any fixed $t \geq 0$ we then know that $\mathbf{y} \circ \lambda_n(t) \to \mathbf{x}(t)$. Since there is a subsequence (n_k) s.t. $\lambda_{n_k} \nearrow \operatorname{Id}$ or $\lambda_{n_k} \searrow \operatorname{Id}$, we either get $\mathbf{y}(t-) = \mathbf{x}(t)$ or $\mathbf{y}(t) = \mathbf{x}(t)$. In the second case $\mathbf{x} = \mathbf{y}$ is evident. In the first the map $\mathbf{x}$ must be continuous at t, since otherwise $\mathbf{y} \circ \lambda_{n_k}(t-) \not\to \mathbf{x}(t-)$ contradicting uniform convergence. The same argument shows that also $\mathbf{y}$ has to be continuous at t thus $\mathbf{x} = \mathbf{y}$ holds again.
Symmetry follows from the fact that $\lambda \in \Lambda_T$ iff $\lambda^{-1} \in \Lambda_T$ and

$$\begin{aligned} d_T(\mathbf{x}, \mathbf{y}) &= \inf_{\lambda \in \Lambda_T} \{\|\lambda - \operatorname{Id}\|_\infty \vee \|\mathbf{x} - \mathbf{y} \circ \lambda\|_\infty\} \\ &= \inf_{\lambda \in \Lambda_T} \{\|\operatorname{Id} - \lambda^{-1}\|_\infty \vee \|\mathbf{x} \circ \lambda^{-1} - \mathbf{y}\|_\infty\} = d_T(\mathbf{y}, \mathbf{x}). \end{aligned}$$

It remains to prove the triangle inequality. Therefore we pick $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathcal{D}_T$ and recall that the composition of $\lambda_1, \lambda_2 \in \Lambda_T$ is again a homeomorphism on $[0, T]$. We obtain $\|\lambda_1 \circ \lambda_2 - \operatorname{Id}\|_\infty \leq \|\lambda_1 \circ \lambda_2 - \lambda_2\|_\infty + \|\lambda_2 - \operatorname{Id}\|_\infty = \|\lambda_1 - \operatorname{Id}\|_\infty + \|\lambda_2 - \operatorname{Id}\|_\infty$, $\|\mathbf{x} - \mathbf{z} \circ \lambda_1 \circ \lambda_2\|_\infty \leq \|\mathbf{x} - \mathbf{y} \circ \lambda_2\|_\infty + \|\mathbf{y} \circ \lambda_2 - \mathbf{z} \circ \lambda_1 \circ \lambda_2\|_\infty = \|\mathbf{x} - \mathbf{y} \circ \lambda_2\|_\infty + \|\mathbf{y} - \mathbf{z} \circ \lambda_1\|_\infty$ and consequently

$$\begin{aligned} d_T(\mathbf{x}, \mathbf{z}) &\leq \inf_{\lambda_1, \lambda_2 \in \Lambda_T} \{\|\lambda_1 \circ \lambda_2 - \operatorname{Id}\|_\infty \vee \|\mathbf{x} - \mathbf{z} \circ \lambda_1 \circ \lambda_2\|_\infty\} \\ &\leq \inf_{\lambda_1, \lambda_2 \in \Lambda_T} \{(\|\lambda_1 - \operatorname{Id}\|_\infty + \|\lambda_2 - \operatorname{Id}\|_\infty) \vee (\|\mathbf{x} - \mathbf{y} \circ \lambda_2\|_\infty + \|\mathbf{y} - \mathbf{z} \circ \lambda_1\|_\infty)\} \\ &= \inf_{\lambda_2 \in \Lambda_T} \{\|\lambda_2 - \operatorname{Id}\|_\infty \vee \|\mathbf{x} - \mathbf{y} \circ \lambda_2\|_\infty)\} \\ &\qquad + \inf_{\lambda_1 \in \Lambda_T} \{\|\lambda_1 - \operatorname{Id}\|_\infty \vee \|\mathbf{y} - \mathbf{z} \circ \lambda_1\|_\infty)\} \\ &= d_T(\mathbf{x}, \mathbf{y}) + d_T(\mathbf{y}, \mathbf{z}). \end{aligned}$$

Hence d_T is a metric on $\mathcal{D}_T$ and $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ iff $d_T(\mathbf{x}_n, \mathbf{x}) \to 0$ is evident by definition. Therefore d_T metricizes the Skorokohod topology on $\mathcal{D}_T$.

(ii) The extension to $\mathcal{D}$ is now standard. We define (see [23], p. 83)

$$d(\mathbf{x}, \mathbf{y}) := \int_0^\infty e^{-s}(d_s(\mathbf{x}, \mathbf{y}) \wedge 1)\mathrm{d}s. \tag{2.6}$$

Here we need to be a bit careful, because d is only well-defined if the map $s \mapsto d_s(\mathbf{x},\mathbf{y})$ is measurable. A proof of this result can be found in [15], Lemma 4.16.
Note that $d(\mathbf{x},\mathbf{y}) \in [0, \int_0^\infty e^{-s}\mathrm{d}s] = [0,1]$. Now if $d(\mathbf{x},\mathbf{y}) = 0$ then $d_s(\mathbf{x},\mathbf{y}) = 0$ for (Lebesgue-)almost every s, i.e. the restrictions of $\mathbf{x}$ and $\mathbf{y}$ to $[0,s]$ coincide for almost all s by results of (i), what implies $\mathbf{x} = \mathbf{y}$ (take for instance a sequence $s_n \uparrow \infty$ along which they coincide). Symmetry and the triangle inequality are immediate consequences of the respective properties of d_s and thus d is a metric on $\mathcal{D}$.

(iii) It remains to check $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ iff $d(\mathbf{x}_n,\mathbf{x}) \to 0$. The prove of this result is a slightly adapted variant of the proof of Theorem 16.2 in [7].
For sufficiency, pick a sequence (t_m) of continuity points of $\mathbf{x}$, s.t. $t_m \uparrow \infty$. By assumption $d_s(\mathbf{x}_n,\mathbf{x}) \to 0$ for almost all s, wlog we can thus assume $d_{t_m}(\mathbf{x}_n,\mathbf{x}) \to 0$ as $n \to \infty$ for all $m \geq 1$. By (i) there are sequences $(\lambda_n^{(t_m)})$ in Λ_{t_m} s.t. $\mathbf{x}_n\big|_{[0,t_m]} \rightsquigarrow \mathbf{x}\big|_{[0,t_m]}$ via $(\lambda_n^{(t_m)})_{n\geq 1}$ for all $m \geq 1$. Now we can find a sequence of numbers (m_n) s.t. $m_n \uparrow \infty$ and $d_{t_m}(\mathbf{x}_n,\mathbf{x}) < \frac{1}{m_n}$. This can be achieved by picking numbers $l_1 < l_2 < \ldots$ s.t. $n \geq l_m$ implies $d_{s_m}(\mathbf{x}_n,\mathbf{x}) \leq \frac{1}{m}$ for all m and defining $m_n = m$ for $l_m \leq n < l_{m+1}$. Then $l_{m_n} \leq m < l_{m_{n+1}}$ and the sequence (m_n) has the desired properties. We claim that $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) for $\lambda_n(t) := \begin{cases} \lambda_n^{(t_{m_n})}(s) & \text{if } s \leq t_{m_n} \\ s & \text{if } s > t_{m_n} \end{cases}$. That $\lambda_n \xrightarrow{u} \mathrm{Id}$ is obvious, since for any $s \geq 0$ we have $|\lambda_n(s) - s| \leq \frac{1}{m_n} \to 0$. On the other hand for any $T \geq 0$ we can choose n s.t. $T \leq m_n$ and get $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{T,\infty} \leq \|\mathbf{x}_n \circ \lambda_n^{(t_{m_n})} - \mathbf{x}\|_{m_n,\infty} \leq \frac{1}{m_n} \to 0$.
Necessity is also a bit tricky, since $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ doesn't imply $\mathbf{x}_n\big|_{[0,s]} \rightsquigarrow \mathbf{x}_n\big|_{[0,s]}$ for all $s \geq 0$ as we have seen in Example 2.18. Luckily this only fails if $|\Delta\mathbf{x}(s)| > 0$ by Lemma 2.19 and since $\mathcal{J}(\mathbf{x})$ is at most countable, $d_s(\mathbf{x}_n,\mathbf{x}) \to 0$ for Lebesgue-almost all s. Hence

$$\lim_{n\to\infty} d(\mathbf{x}_n,\mathbf{x}) = \lim_{n\to\infty} \int_0^\infty e^{-s}(d_s(\mathbf{x}_n,\mathbf{x}) \wedge 1)\mathrm{d}s \overset{DCT}{=} \int_0^\infty e^{-s}(\lim_{n\to\infty} d_s(\mathbf{x}_n,\mathbf{x}) \wedge 1)\mathrm{d}s = 0,$$

by the Dominated Convergence Theorem (DCT). □

Corollary 2.22. *There is convergence* $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ *iff* $\mathbf{x}_n\big|_{[0,T]} \rightsquigarrow \mathbf{x}\big|_{[0,T]}$ *for almost all* T*. Furthermore we can replace "almost all* T*" by "continuity points of* $\mathbf{x}$*" in the above statement.*

Proof. We start with the first claim. Necessity in both cases is just Lemma 2.19 (and Corollary 2.6), while sufficiency for the first statement follows from dominated convergence and the fact that d given by (2.6) is a metric for the topology. More properly, the assumption that $\mathbf{x}_n\big|_{[0,T]} \rightsquigarrow \mathbf{x}\big|_{[0,T]}$ for almost all T, implies that $d_T(\mathbf{x}_n,\mathbf{x}) \to 0$ almost everywhere and consequently $d(\mathbf{x}_n,\mathbf{x}) \to 0$ by dominated convergence.
A direct proof of sufficiency in the second part is provided by (iii) above. □

We already mentioned that Skorokhod came up with other topologies on $\mathcal{D}$ having the property that "matching jumps" aren't needed. For more information on one of them, which is called M_1-topology, we refer to Whitt [23]. However, we don't dig deeper into this topic and continue working with the J_1-topology. The reason is, that the M_1-topology is coarser anyway, in general more complicated to work with and also has the following main problem.

Example 2.23 (The addition is not continuous). Let $\mathbf{x}_n := \mathbb{1}_{[\frac{1}{2}-\frac{1}{n},1]}$, $\mathbf{y}_n := -\mathbb{1}_{[\frac{1}{2}+\frac{1}{n},1]}$, then $\mathbf{x}_n \rightsquigarrow \mathbf{x} := \mathbb{1}_{[\frac{1}{2},1]}$ and $\mathbf{y}_n \rightsquigarrow \mathbf{y} = -\mathbf{x}$, but $\mathbf{x}_n + \mathbf{y}_n = \mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n})} \not\rightsquigarrow \mathbf{x} + \mathbf{y} = 0$. Again, whatever homeomorphism $\lambda \in \Lambda_1$ we apply, the shifted sequence always assumes the values 0 and 1.

Remark 2.24. This has a dramatic consequence: $\mathcal{D}$ equipped with the Skorokhod topology is no topological vector space and is therefore somewhat outside traditional functional analysis!

Many important results in the theory of weak convergence require the underlying spaces to be Polish, thus we are interested in completeness of $\mathcal{D}$. Unfortunately $\mathcal{D}$ is not complete if we use the metric d defined in (2.6).

Example 2.25 ($(\mathcal{D}, d)$ is not complete). ([7], Example 12.2)
It suffices to check this statement for $\mathcal{D}_1$. If we let $\mathbf{x}_n := \mathbb{1}_{[0,\frac{1}{2^n}]}$ for $n \geq 1$ and $\mathbf{x} \equiv 0$ on $[0,1]$, we see by a similar argument as in Example 2.16 or 2.17 that $\mathbf{x}_n \not\rightsquigarrow \mathbf{x}$. Let us now find some time-shifts making the sequence $(\mathbf{x}_n)$ Cauchy w.r.t. d. Interpolate $\lambda_n(0) = 0$, $\lambda_n(\frac{1}{2^n}) = \frac{1}{2^{n+1}}$ and $\lambda_n(1) = 1$ linearly. Then $\lambda_n \in \Lambda_1$ for all n, $\|\lambda_n - \mathrm{Id}\,\|_\infty \leq \frac{1}{2^{n+1}}$ and $\|\mathbf{x}_n - \mathbf{x}_{n+1} \circ \lambda_n\|_\infty = 0$. Thus $d(\mathbf{x}_n, \mathbf{x}_{n+1}) \leq \frac{1}{2^{n+1}}$ for all n and consequently we have $d(\mathbf{x}_n, \mathbf{x}_m) \leq d(\mathbf{x}_n, \mathbf{x}_{n+1}) + \cdots + d(\mathbf{x}_{m-1}, \mathbf{x}_m) \leq \sum_{k=n+1}^{m} \frac{1}{2^k} \leq \frac{1}{2^n}$ for all $m \geq n$.

Although one may now have the impression that too much goes wrong in this space, we now show that is not as bad as it may seem at the moment.
First to the problem with the addition. An easy condition for its continuity is that both sequences converge via the same time-shifts. We will put some more effort in proving a similar but more general statement at the end of this chapter (cf. Proposition 2.43).

Lemma 2.26. *If $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) and $\mathbf{y}_n \rightsquigarrow \mathbf{y}$ via (λ_n), then $\mathbf{x}_n + \mathbf{y}_n \rightsquigarrow \mathbf{x} + \mathbf{y}$ via (λ_n).*

Proof. Let us check the two needed conditions for convergence in $\mathcal{D}$. That $\|\lambda_n - \mathrm{Id}\,\|_\infty \to 0$ is evident and we have $(\mathbf{x}_n + \mathbf{y}_n) \circ \lambda_n = \mathbf{x}_n \circ \lambda_n + \mathbf{y}_n \circ \lambda_n \overset{ucs}{\to} \mathbf{x} + \mathbf{y}$ by assumption. □

To address the completeness problem we use a famous result from Billingsley [7]. He provides us with an alternative metric d^o, which is equivalent to d, but makes $\mathcal{D}$ a complete and even separable space.

Theorem 2.27 ($(\mathcal{D}, d^o)$ is Polish). *There exists a metric d^o, which is equivalent to d, s.t. $(\mathcal{D}, d^o)$ is a Polish space, i.e. completely metrizable and separable. Moreover, there is convergence $d^o(\mathbf{x}_n, \mathbf{x}) \to 0$ in $\mathcal{D}$ iff $d^o_T(\mathbf{x}_n, \mathbf{x}) := d\big(\mathbf{x}_n\big|_{[0,T]}, \mathbf{x}\big|_{[0,T]}\big) \to 0$ for any continuity point T of $\mathbf{x}$.*

Proof. We omit the proof. For the definition of the metric d^o on $\mathcal{D}_1$ see [7], p. 125 and for its extension to $\mathcal{D}_\infty$ see p. 168. The remaining parts follow from the Theorems 16.2 and 16.3 on p. 169 and 170ff. □

The main difference between the metric d^o and d is that the latter allows a time-shift $\lambda \in \Lambda$ to have arbitrarily high slopes, which was essential in Example 2.25, while d^o requires the shifts not only to be asymptotically close to the identity, but to also have *slopes* close to 1.

Remark 2.28. An easy but nice consequence of Theorem 2.27 is that if a sequence $(\mathbf{x}_n)$ converges in the Skorokhod topology, i.e. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ for some $\mathbf{x}$, then $\mathbf{x}$ is càdlàg. This result extends Lemma 2.9 which required uniform convergence.

2.4. Continuity and Skorokhod's Topology

The following continuity results (especially Lemma 2.34, Corollary 2.37 and Proposition 2.43) are a product of an attempt to find "elementary" proofs of those well-known results without using tightness criteria as Jacod and Shirjaev did in [9].

To begin this section, we finally prove the "matched jump" property of the Skorokhod topology, which is a simple consequence of the following rather technical, but very useful result.

Lemma 2.29. *Let $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n), $t \geq 0$, set $t_n := \lambda_n(t)$ for $n \geq 1$ and let (t'_n) be a sequence s.t. $t'_n \to t$. If*

(i) *$t'_n \leq t_n$ for all n, then $\mathbf{x}_n(t'_n-) \to \mathbf{x}(t-)$*

(ii) *$t'_n < t_n$ for all n, then $\mathbf{x}_n(t'_n) \to \mathbf{x}(t-)$*

(iii) *$t'_n \geq t_n$ for all n, then $\mathbf{x}_n(t'_n) \to \mathbf{x}(t)$*

(iv) *$t'_n > t_n$ for all n, then $\mathbf{x}_n(t'_n-) \to \mathbf{x}(t)$.*

Proof. Cf. [9], Proposition 2.1, p. 337ff.
(i) Let $t = 0$ first, then $t_n = \lambda_n(0) = 0$ for all n and thus any sequence (t'_n) coincides with (t_n) for trivial reasons. For $t > 0$ assume wlog $t'_n > 0$ for all n. Then $\mathbf{x}_n(t'_n-)$ exists and $\mathbf{x}_n(t'_n - \frac{1}{k}) \to \mathbf{x}_n(t'_n-)$ as $k \to \infty$. Hence for all n we can find $k_0(n)$, s.t. $|\mathbf{x}_n(t'_n - \frac{1}{k}) - \mathbf{x}_n(t'_n-)| \leq \frac{1}{n}$ for all $k \geq k_0(n)$ and consequently there is a sequence (s_n) s.t. $s_n < t'_n$ and $|\mathbf{x}_n(t'_n-) - \mathbf{x}_n(s_n)| \to 0$ (take for instance $s_n = t'_n - \frac{1}{k_0(n)}$). Define $r_n := \lambda_n^{-1}(s_n)$, then $r_n \uparrow t$ and thus $|\mathbf{x}(r_n) - \mathbf{x}(t-)| \to 0$. To verify this, note first that $s_n < t'_n \leq t_n = \lambda_n(t)$ by assumption and therefore $r_n = \lambda_n^{-1}(s_n) < t$ for all n and second that $|r_n - t| \leq |\lambda_n^{-1}(s_n) - s_n| + |s_n - t'_n| + |t'_n - t| \to 0$. Here the first term on the right-hand side converges to 0, since $\lambda_n^{-1} \xrightarrow{u} \mathrm{Id}$ and the remaining two are consequences of $s_n \to t'_n$ and $t'_n \to t$. Moreover, the choice of (r_n) assures that $|\mathbf{x}_n(s_n) - \mathbf{x}(r_n)| = |\mathbf{x}_n \circ \lambda_n(r_n) - \mathbf{x}(r_n)| \to 0$, since $r_n < t \leq N$ for some N and $\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{N,\infty} \to 0$ by assumption. Plugging all that together gives

$$|\mathbf{x}_n(t'_n-) - \mathbf{x}(t-)| \leq |\mathbf{x}_n(t'_n-) - \mathbf{x}_n(s_n)| + |\mathbf{x}_n(s_n) - \mathbf{x}(r_n)| + |\mathbf{x}(r_n) - \mathbf{x}(t-)| \to 0.$$

(ii) The proof is just a shorter version of (i), where we can use $(s_n) = (t'_n)$. For $r_n := \lambda_n^{-1}(t'_n)$ we easily get $r_n \uparrow t$ and $|\mathbf{x}_n(t'_n) - \mathbf{x}(r_n)| = |\mathbf{x}_n \circ \lambda_n(r_n) - \mathbf{x}(r_n)| \to 0$ as above, so $|\mathbf{x}_n(t'_n) - \mathbf{x}(t)| \leq |\mathbf{x}_n(t'_n) - \mathbf{x}(r_n)| + |\mathbf{x}(r_n) - \mathbf{x}(t-)| \to 0$.

(iii) Similar to (i).

(iv) Similar to (ii). □

Corollary 2.30. *Let $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n), $t \geq 0$ and (t'_n) be any sequence converging to t.*

(i) *If $\Delta\mathbf{x}(t) = 0$, then $\mathbf{x}_n(t'_n) \to \mathbf{x}(t)$, $\mathbf{x}_n(t'_n-) \to \mathbf{x}(t-)$ and $\Delta\mathbf{x}_n(t'_n) \to \Delta\mathbf{x}(t)$.*

(ii) *If $\Delta\mathbf{x}_n(t'_n) \to \Delta\mathbf{x}(t) > 0$, then $t'_n = \lambda_n(t)$ for all n large.*

Proof. Cf. [9], Proposition 2.1, p. 337ff.
(i) Recall that in a metric space a sequence $x_n \to x$ iff every subsequence contains a further subsequence which converges to x. Here any subsequence of (t'_n) contains a further subsequence (t'_{l_n}) s.t. all terms either lie to the left or the right of t. In both cases $(\mathbf{x}_{l_n}(t'_{l_n}-))$ and $(\mathbf{x}_{l_n}(t'_{l_n}))$ converge to $\mathbf{x}(t) = \mathbf{x}(t-)$ by Lemma 2.29. Hence the limit points coincide with the desired limit and thus $\mathbf{x}_n(t'_n) \to \mathbf{x}(t)$ and $\mathbf{x}_n(t'_n-) \to \mathbf{x}(t-)$.

(ii) Assume the contrary. Then we can find a subsequence (n_k) s.t. either $t'_{n_k} < t_{n_k}$ or $t'_{n_k} > t_{n_k}$. In the first case we apply (i), (ii) and in the second (iii), (iv) of Lemma 2.29 to conclude that $\Delta\mathbf{x}_{n_k}(t'_{n_k}) = \mathbf{x}_{n_k}(t'_{n_k}) - \mathbf{x}_{n_k}(t'_{n_k}-) \to 0$. This is a contradiction to the assumption $|\Delta\mathbf{x}_n(t'_n)| \to |\Delta\mathbf{x}(t)| > 0$. □

Proposition 2.31 ("Matched-jump" property). *Let $\varepsilon > 0$, $t \geq 0$ and $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n).*

(i) *Then $t_n := \lambda_n(t) \to t$, $\mathbf{x}_n(t_n) \to \mathbf{x}(t)$, $\mathbf{x}_n(t_n-) \to \mathbf{x}(t-)$ and $\Delta\mathbf{x}_n(t_n) \to \Delta\mathbf{x}(t)$.*

(ii) *If there is a sequence $t'_n \to t$ s.t. $|\Delta\mathbf{x}_n(t'_n)| \geq \varepsilon$, then $|\Delta\mathbf{x}(t)| \geq \varepsilon$.*

Proof. (i) That $t_n \to t$ is evident, since $\lambda_n \overset{u}{\to} \mathrm{Id}$. This implies that the sequence (t_n) is contained in a compact set and thus $\mathbf{x}_n(t_n) = \mathbf{x}_n \circ \lambda_n(t) \to \mathbf{x}(t)$, since $\mathbf{x}_n \circ \lambda_n \overset{ucs}{\to} \mathbf{x}$ by assumption. That also $\mathbf{x}_n(t_n-) \to \mathbf{x}(t-)$ is just Lemma 2.29 (i) applied to $(t'_n) = (t_n)$ and finally $\Delta\mathbf{x}_n(t_n) = \mathbf{x}_n(t_n) - \mathbf{x}_n(t_n-) \to \mathbf{x}(t) - \mathbf{x}(t-) = \Delta\mathbf{x}(t)$.

(ii) For any $\eta > 0$ we have

$$|t'_n - t| < \tfrac{\eta}{2} \qquad \text{and} \qquad \|\lambda_n - \mathrm{Id}\|_\infty < \tfrac{\eta}{2},$$

given n is large enough. For those n the function $\mathbf{x}_n$ has a jump of absolute value $\geq \varepsilon$ in t'_n and thus $\mathbf{x}_n \circ \lambda_n$ must have such a jump at $t''_n \in [t-\eta, t+\eta]$. Let us assume that $\mathbf{x}$ has no jump of size $\geq \varepsilon$ in $[t-\eta, t+\eta]$, that is $\varepsilon - |\Delta\mathbf{x}(s)| \geq \delta$ for all $s \in [t-\eta, t+\eta]$ and some $\delta > 0$. Then

$$\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{N,\infty} \geq |\mathbf{x}_n \circ \lambda_n(t''_n) - \mathbf{x}(t''_n)| \vee |\mathbf{x}_n \circ \lambda_n(t''_n-) - \mathbf{x}(t''_n-)| \geq \tfrac{\delta}{2},$$

for all large n, a contradiction to $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n). Finally, since $\eta > 0$ was arbitrary we get $|\Delta\mathbf{x}(t)| \geq \varepsilon$. □

We have seen that if $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) and the limit function has a jump of positive size at time t, any sequence (t'_n) "matching" this jump, finally has to coincide with $(\lambda_n(t))$.
Later, we want to cleverly modify càdlàg functions and for instance eliminate some jumps. Therefore the following notation is very useful.

Definition 2.32. Let $\mathbf{x} \in \mathcal{D}$ and $\mathcal{T} := \{t_i, i \geq 1\}$ be an ordered, discrete subset of $[0, \infty)$, i.e. $t_1 < t_2 < \dots$ and $\mathcal{T} \cap K$ is finite for any compact set K. Then we define two operations "$@\mathcal{T}$" and "$\backslash\mathcal{T}$" on $\mathcal{D}$ via

$$\mathbf{x}^{@\mathcal{T}} := \sum_{i\geq 1} \Delta\mathbf{x}(t_i)\mathbb{1}_{[t_i,\infty)} \qquad \text{and} \qquad \mathbf{x}^{\backslash\mathcal{T}} := \mathbf{x} - \mathbf{x}^{@\mathcal{T}}.$$

If $\mathcal{T} = \{t\}$ for some $t \geq 0$, we write $\mathbf{x}^{@t}$ and $\mathbf{x}^{\backslash t}$ instead of $\mathbf{x}^{@\{t\}}$ and $\mathbf{x}^{\backslash\{t\}}$.

Remark 2.33. (i) $\mathbf{x}^{@\mathcal{T}}$ is a jump function, thus càdlàg and hence $\mathbf{x}^{\backslash\mathcal{T}}$ is càdlàg too.

(ii) The function $\mathbf{x}^{\backslash\mathcal{T}}$ is continuous at all points in $\mathcal{T}$.

(iii) For $\varepsilon > 0$ and a given $\mathbf{x} \in \mathcal{D}$ the set $\mathcal{J}_\varepsilon(\mathbf{x})$ is discrete by Corollary 2.6. Consequently $\mathbf{x}^{\backslash\mathcal{J}_\varepsilon(\mathbf{x})} \in \mathcal{D}$ has no jump of size $> \varepsilon$.

(iv) One easily checks that $(\gamma\mathbf{x} + \mathbf{y})^{@\mathcal{T}} = \gamma\mathbf{x}^{@\mathcal{T}} + \mathbf{y}^{@\mathcal{T}}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, $\gamma \in \mathbb{R}$ and $\mathbf{x}^{@(\mathcal{T}_1 \uplus \mathcal{T}_2)} = \mathbf{x}^{@\mathcal{T}_1} + \mathbf{x}^{@\mathcal{T}_2}$ for disjoint $\mathcal{T}_1, \mathcal{T}_2$. Evidently this statements are also true for $\mathbf{x}^{\backslash\mathcal{T}}$ and $\mathbf{y}^{\backslash\mathcal{T}}$.

We are naturally interested in continuity of those two operations and – as suspected – they are not. To see that, let $t \geq 0$, $\mathbf{x}_n := \mathbb{1}_{[t+\frac{1}{n},\infty)}$ and $\mathbf{x} := \mathbb{1}_{[t,\infty)}$. Then $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ but $\mathbf{x}_n^{@t} \equiv 0 \not\rightsquigarrow \mathbf{x}^{@t} = \mathbf{x}$. This is again quite inconvenient, so we should try to at least save some kind of continuity property. Therefore we adapt to the structure of the Skorokhod topology and allow small deformations of the set $\mathcal{T}$.

Lemma 2.34. *Let $\mathcal{T} = \{t^{(i)}, i \geq 1\}$ be a discrete, ordered subset of $[0,\infty)$ and assume $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n). For each i, take a sequence $\big(t_n^{(i)}\big)$ converging to $t^{(i)}$ s.t. $\Delta\mathbf{x}_n\big(t_n^{(i)}\big) \to \Delta\mathbf{x}\big(t^{(i)}\big)$ and let $\mathcal{T}_n := \{t_n^{(i)}, i \geq 1\}$. Then we can assume that $\mathcal{T}_n$ is ordered and there is some $m \in \mathbb{N}_0$ s.t. $t_n^{(i)} \in \mathcal{T}_n \cap [0,N]$ iff $i \leq m$ for almost all $N \geq 0$ and n sufficiently large. Moreover,*

$$\mathbf{x}_n^{@\mathcal{T}_n} \rightsquigarrow \mathbf{x}^{@\mathcal{T}} \text{ via } (\lambda_n) \qquad \text{and} \qquad \mathbf{x}_n^{\backslash\mathcal{T}_n} \rightsquigarrow \mathbf{x}^{\backslash\mathcal{T}} \text{ via } (\lambda_n). \tag{2.7}$$

Proof. (i) Since $t^{(2)} > t^{(1)}$ we only need to change $t_n^{(2)}$ finitely many times to assure that $t_n^{(2)} > t_n^{(1)}$ for all n. Again, since $t^{(3)} > t^{(2)}$ we only have to change $t_n^{(3)}$ finitely many times to have $t_n^{(3)} > t_n^{(2)}$ for all n and so on.
Pick $N \notin \mathcal{T} \cup \big(\bigcup_{n\geq 1} \mathcal{T}_n\big)$, what is possible since this is just a countable set (although this may eventually mean that N is no integer). The compact set $[0,N]$ only contains finitely many elements of $\mathcal{T}$ and $\mathcal{T}_n$. Define $m := \max\{i \geq 1 : t^{(i)} < N\}$, where $\max\{\emptyset\} := 0$. Now we can find some n_0 s.t. $t_n^{(i)} \in [0,N]$ for all $n \geq n_0$ iff $i \leq m$ because $t_n^{(i)} \to t^{(i)} \neq N$ for all $i \geq 1$ and $\mathcal{T}$ and $\mathcal{T}_n$ are ordered.

(ii) For the remaining part it suffices to prove the left-hand side of (2.7) due to Lemma 2.26. We start with the special case when $\mathcal{T} = \{t\}$ for some $t \geq 0$. Since $\lambda_n \xrightarrow{u} \mathrm{Id}$ it remains to show that $\mathbf{x}_n^{@t_n} \circ \lambda_n \xrightarrow{ucs} \mathbf{x}^{@t}$. If $\Delta\mathbf{x}(t) = 0$, i.e. $\mathbf{x}^{@t} = 0$, this is easy, because $\|\mathbf{x}_n^{@t_n} \circ \lambda_n\|_\infty = |\Delta\mathbf{x}_n(t_n)| \to |\Delta\mathbf{x}(t)| = 0$ by Corollary 2.30 (i). If $\Delta\mathbf{x}(t) > 0$, Corollary 2.30 (ii) implies $t_n = \lambda_n(t)$, thus $\mathbb{1}_{[t_n,\infty)} \circ \lambda_n = \mathbb{1}_{[t,\infty)}$ for large n. For those n and a sufficiently big N we get

$$\begin{aligned}\|\mathbf{x}_n^{@t_n} \circ \lambda_n - \mathbf{x}^{@t}\|_{N,\infty} &= \|\Delta\mathbf{x}_n(t_n)\mathbb{1}_{[t_n,\infty)} \circ \lambda_n - \Delta\mathbf{x}(t)\mathbb{1}_{[t,\infty)}\|_{N,\infty} \\ &= \|\Delta\mathbf{x}_n(t_n)\mathbb{1}_{[t,\infty)} - \Delta\mathbf{x}(t)\mathbb{1}_{[t,\infty)}\|_{N,\infty} \leq |\Delta\mathbf{x}_n(t_n) - \Delta\mathbf{x}(t)| \to 0.\end{aligned}$$

(iii) To prove the general case, fix $N \geq 0$ as above and verify that $\|\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda_n - \mathbf{x}^{@\mathcal{T}}\|_{N,\infty} \to 0$. By (i) we can identify $\mathbf{x}_n^{@\mathcal{T}_n} = \mathbf{x}_n^{@\{t_n^{(1)},\ldots,t_n^{(m)}\}}$ and $\mathbf{x}^{@\mathcal{T}} = \mathbf{x}^{@\{t^{(1)},\ldots,t^{(m)}\}}$ on $[0,N]$ for $n \geq n_0$. This allows us to attribute this problem to (ii). Indeed $\mathbf{x}_n^{@t_n^{(i)}} \rightsquigarrow \mathbf{x}^{@t^{(i)}}$

via (λ_n) holds for $1 \le i \le m$ and thus $\mathbf{x}_n^{@\{t_n^{(1)},\dots,t_n^{(m)}\}} \rightsquigarrow \mathbf{x}^{@\{t^{(1)},\dots,t^{(m)}\}}$ via (λ_n) by Lemma 2.26. In particular we get

$$\|\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda_n - \mathbf{x}^{@\mathcal{T}}\|_{N,\infty} = \|\mathbf{x}_n^{@\{t_n^{(1)},\dots,t_n^{(m)}\}} \circ \lambda_n - \mathbf{x}^{@\{t^{(1)},\dots,t^{(m)}\}}\|_{N,\infty} \to 0$$

as desired. $\square$

Let us now look at a slightly different version of the above statement. It will be very useful later, since it allows us to collect/eliminate jumps of a càdlàg function in a continuous way.

Definition 2.35. For $\mathbf{x} \in \mathcal{D}$ define the points where its jumps exceed a certain size $\varepsilon > 0$ iteratively via $t_\varepsilon^{(0)}(\mathbf{x}) := 0$, $t_\varepsilon^{(i+1)}(\mathbf{x}) := \inf\{t > t_\varepsilon^{(i)}(\mathbf{x}) : |\Delta\mathbf{x}(t)| > \varepsilon\}$ for $i \ge 1$ $(\inf(\emptyset) := \infty)$ and let $\mathcal{T}^\varepsilon(\mathbf{x}) := \{t_\varepsilon^{(i)}(\mathbf{x}), i \ge 1\}$. We collect the values of all jumps of positive size of $\mathbf{x}$ in the set $U(\mathbf{x}) := \{u > 0 : |\Delta\mathbf{x}(t)| = u \text{ for some } t > 0\}$.

Remark 2.36. The set $U(\mathbf{x})$ is at most countable for any $\mathbf{x} \in \mathcal{D}$, since càdlàg functions have at most countably many jumps.

Corollary 2.37 (Collecting/Eliminating jumps). *If $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n), $\varepsilon > 0$ and $\varepsilon \notin U(\mathbf{x})$, then both $\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon} \rightsquigarrow \mathbf{x}^{@\mathcal{T}^\varepsilon}$ and $\mathbf{x}_n^{\setminus\mathcal{T}_n^\varepsilon} \rightsquigarrow \mathbf{x}^{\setminus\mathcal{T}^\varepsilon}$ via (λ_n) for $\mathcal{T}^\varepsilon := \mathcal{T}^\varepsilon(\mathbf{x})$ and $\mathcal{T}_n^\varepsilon := \mathcal{T}^\varepsilon(\mathbf{x}_n)$.*

Proof. By Lemma 2.26 it suffices to show the first statement. Obviously $\lambda_n \xrightarrow{u} \mathrm{Id}$ already holds. For convenience we drop the index ε of the jumping times in the following and write $\mathcal{T}^\varepsilon := \{t^{(1)}, t^{(2)}, \dots\}$ as an ordered set. For a fixed $N \ge 1$ as above there is a m s.t. $\mathcal{T}^\varepsilon \cap [0,N] = \{t^{(1)},\dots,t^{(m)}\}$ and $T_n^\varepsilon \cap [0,N] = \{t_n^{(1)},\dots,t_n^{(l_n)}\}$ for some $l_n \ge 0$. To apply Lemma 2.34 we need to assure that $\mathcal{T}_n^\varepsilon \cap [0,N] = \{\tilde{t}_n^{(1)},\dots,\tilde{t}_n^{(m)}\} := \{\lambda_n(t^{(1)}),\dots,\lambda_n(t^{(m)})\}$ for large n, i.e. it only consists of "matching-jump" terms.
On the one hand the sequences $(\tilde{t}_n^{(i)})_{n\ge1}$ for $i \ge 1$ pose no problem, since $\varepsilon \notin U(\mathbf{x})$ by assumption and thus the jumps of $\mathbf{x}_n$ along the sequences $(\lambda_n(t^{(i)}))$ finally have to exceed ε for all $1 \le i \le m$. We can then apply the first part of Lemma 2.34 to $\tilde{\mathcal{T}}_n^\varepsilon := \{\tilde{t}_n^{(i)}, i \ge 1\}$ to get $\tilde{t}_n^{(i)} \in \mathcal{T}_n \cap [0,N]$ iff $i \le m$.
On the other hand "non-matching" sequences are also no difficulty. Since $\mathbf{x}_n$ can have at most finitely many jumps of size $> \varepsilon$ in $[0,N]$, there are $s_n^{(1)},\dots,s_n^{(k_n)} \in \mathcal{T}_n^\varepsilon \setminus \{\tilde{t}_n^{(1)},\dots,\tilde{t}_n^{(m)}\}$ for a *bounded* sequence (k_n) (assuming (k_n) is not bounded, there exists a subsequence (n') s.t. $k_{n'} \to \infty$ and $s_{n'}^{(k_{n'})} \to \tilde{t} \in [0,N]$, what causes an accumulation of large jumps and hence contradicts Corollary 2.6), i.e. $\mathcal{T}_n^\varepsilon$ contains at most finitely many points which are no "matching-jumps". Now apply Proposition 2.31 (ii) to ensure that $|\Delta\mathbf{x}_n(s_n^{(i)})| < \varepsilon$ for all $1 \le i \le k_n$ and n sufficiently large.

Altogether we have $\mathcal{T}_n^\varepsilon \cap [0,N] = \{\tilde{t}_n^{(1)},\dots,\tilde{t}_n^{(m)}\}$, with $\Delta\mathbf{x}(\tilde{t}_n^{(i)}) \to \Delta\mathbf{x}(\tilde{t}^{(i)})$ for all $i \le m$ if n is large and we get $\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon} \rightsquigarrow \mathbf{x}^{@\mathcal{T}^\varepsilon}$ by a similar argument as in the proof of Lemma 2.34. $\square$

Example 2.38. In the above corollary the assumption that $\varepsilon \notin U(\mathbf{x})$ is crucial. For a simple counterexample let $\mathbf{x}_n := (1 - \frac{1}{n})\mathbb{1}_{[t,\infty)} \rightsquigarrow \mathbf{x} := \mathbb{1}_{[t,\infty)}$, while $0 \equiv \mathbf{x}_n^{@\mathcal{T}_n^1} \not\rightsquigarrow \mathbf{x}^{@\mathcal{T}^1} = \mathbb{1}_{[t,\infty)}$. In many applications this is not really a problem, since $U(\mathbf{x})$ is at most countable.

The remaining part of this section provides a condition for the continuity of the addition to hold.

Definition 2.39. For $\mathbf{x} \in \mathcal{D}$ let $\Delta\mathbf{x}$ denote the SIZE OF THE BIGGEST JUMP OF $\mathbf{x}$, i.e. $\Delta\mathbf{x} := \sup_{t\in\mathcal{J}(\mathbf{x})}\{|\Delta\mathbf{x}(t)|\}$ and write $|\mathbf{x}|$ as an abbreviation for the function $s \mapsto |\mathbf{x}(s)|$.

Proposition 2.40. *Let $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ and $\Delta\mathbf{x} \leq \eta$ for some $\eta \geq 0$. If we pick any two sequences (λ_n) and (λ'_n) s.t. $\lambda_n \stackrel{u}{\to} \mathrm{Id}$ and $\lambda'_n \stackrel{u}{\to} \mathrm{Id}$, then $|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}_n \circ \lambda_n| \vee \eta \stackrel{ucs}{\to} \eta$. In particular, we have $|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}| \vee \eta \stackrel{ucs}{\to} \eta$ if $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n).*

Proof. Assume the contrary. Then there is some $\varepsilon > 0$, $N \geq 0$, a sequence (t_k) in $[0, N]$ and a subsequence (n_k) s.t. $|\mathbf{x}_{n_k} \circ \lambda_{n_k}(t_k) - \mathbf{x}_{n_k} \circ \lambda'_{n_k}(t_k)| \geq \eta + \varepsilon$ for all k. By looking at appropriate subsequences, suppose that the bounded sequence (t_k) converges to some $\tilde{t} \in [0, N]$. The assumption on (λ_n) and (λ'_n) implies that both $s_k := \lambda_{n_k}(t_k) \to \tilde{t}$ and $s'_k := \lambda'_{n_k}(t_k) \to \tilde{t}$. By looking at suitable further subsequences we can apply Lemma 2.29 to conclude that all limit points of $(\mathbf{x}_{n_k} \circ \lambda_{n_k}(t_k))$ and $(\mathbf{x}_{n_k} \circ \lambda'_{n_k}(t_k))$ are either $\mathbf{x}(\tilde{t})$ or $\mathbf{x}(\tilde{t}-)$ (recall the proof of Corollary 2.30 (i)). But their distance is $|\mathbf{x}(\tilde{t}) - \mathbf{x}(\tilde{t}-)| \leq \Delta\mathbf{x} \leq \eta$, a contradiction.
If $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n), then $|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}| \vee \eta \leq (|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}_n \circ \lambda_n| + |\mathbf{x}_n \circ \lambda_n - \mathbf{x}|) \vee \eta \stackrel{ucs}{\to} \eta$, by the result above and $\mathbf{x}_n \circ \lambda_n \stackrel{ucs}{\to} \mathbf{x}$. □

Proposition 2.41. *Let $(\mathbf{x}_n)$ be a sequence in $\mathcal{D}$. Assume that for any $k \geq 1$ there is a sequence $\big(\lambda_n^{(k)}\big)_{n\geq 1}$ in Λ, s.t.*

$$\lambda_n^{(k)} \stackrel{u}{\to} \mathrm{Id} \qquad \text{and} \qquad |\mathbf{x}_n \circ \lambda_n^{(k)} - \mathbf{x}| \vee \tfrac{1}{k} \stackrel{ucs}{\to} \tfrac{1}{k} \quad \text{as } n \to \infty$$

for some $\mathbf{x} \in \mathcal{D}$. Then $\mathbf{x}_n \rightsquigarrow \mathbf{x}$.

Proof. We define a sequence (λ_n) s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) iteratively using a diagonalization argument. Begin with $n_1 := 1$ and assume that some numbers $n_2, \ldots, n_{k-1}$ are already given. By assumption we can find $n_k > n_{k-1}$ s.t. $\|\lambda_n^{(k)} - \mathrm{Id}\,\|_\infty \leq \frac{1}{k}$ and $\|\mathbf{x}_n \circ \lambda_n^{(k)} - \mathbf{x}\|_{k,\infty} < \frac{2}{k}$ for all $n \geq n_k$.
Now let $\lambda_n := \lambda_n^{(k)}$ for $n_k \leq n < n_{k+1}$ and $k \geq 1$. We claim that $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n). Indeed, $\lambda_n \stackrel{u}{\to} \mathrm{Id}$ is evident. To show uniform convergence of the shifted sequence on compact sets, we fix some $N \geq 0$ and $\varepsilon > 0$. Now for any $n \geq n_{k'}$, where $k' \geq 0$ is chosen s.t. $[0, N] \subseteq [0, k']$ and $\varepsilon \geq \frac{2}{k'}$, we have

$$\|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{N,\infty} \leq \|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{k',\infty} \leq \|\mathbf{x}_n \circ \lambda_n^{(k')} - \mathbf{x}\|_{k',\infty} \leq \tfrac{2}{k'} \leq \varepsilon$$

and we are done. □

Lemma 2.42. (i) *Let $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n) and $t_n \to t > 0$ s.t. $\Delta\mathbf{x}_n(t_n) \to \Delta\mathbf{x}(t)$. Then for any $\delta > 0$ we can find $n' \geq 0$ and a sequence (λ'_n) s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ'_n) and $\lambda'_n = \lambda_n$ on $[t-\delta, t+\delta]^c$ with $\lambda'_n(t) = t_n$ for all $n \geq n'$.*
(ii) *Let $\mathcal{T} := \{t^{(i)}, i \geq 1\}$ be an ordered discrete subset of $[0, \infty)$. For each i let $\big(t_n^{(i)}\big)$ be a sequence s.t. $t_n^{(i)} \to t^{(i)} > 0$ and $\Delta\mathbf{x}_n(t_n^{(i)}) \to \Delta\mathbf{x}(t^{(i)})$. Then there are $n'(i)$, a sequence (λ^*_n) in Λ s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ^*_n) and for each $i \geq 1$ we have $\lambda^*_n(t^{(i)}) = t_n^{(i)}$ for all $n \geq n'(i)$.*

Proof. (i) If $\Delta\mathbf{x}(t) > 0$ we can take $(\lambda'_n) = (\lambda_n)$ by Corollary 2.30 (ii). Hence assume $\Delta\mathbf{x}(t) = 0$ and fix $\delta \geq \varepsilon > 0$ and $N \geq 1$. Let us rewrite the desired properties of (λ'_n) first. To assure that $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ'_n) we need $\|\lambda'_n - \operatorname{Id}\|_\infty < \varepsilon$ and $\|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}\|_{N,\infty} < \varepsilon$. Since $\|\lambda'_n - \operatorname{Id}\|_\infty \leq \|\lambda'_n - \lambda_n\|_\infty + \|\lambda_n - \operatorname{Id}\|_\infty$ and $\|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}\|_{N,\infty} \leq \|\mathbf{x}_n \circ \lambda'_n - \mathbf{x}_n \circ \lambda_n\|_{N,\infty} + \|\mathbf{x}_n \circ \lambda_n - \mathbf{x}\|_{N,\infty}$ it suffices to find (λ'_n) s.t. the first terms on the right-hand sides do not exceed $\frac{\varepsilon}{2}$.
We split up the conditions once more. Therefore let $\mathcal{T} := \{t^{(0)} := 0, t^{(1)}, \ldots, t^{(m)}\}$ be the ordered times, where $\mathbf{x}$ has jumps of size $\geq \frac{\varepsilon}{4}$ in $[0, N+1]$ and $\{0\}$. By defining sequences $\big(t_n^{(j)}\big) := \big(\lambda_n(t^{(j)})\big)$, we know that $t_n^{(j)} \to t^{(j)}$ s.t. $\Delta\mathbf{x}_n\big(t_n^{(j)}\big) \to \Delta\mathbf{x}(t^{(j)})$ for $0 \leq j \leq m$ (Proposition 2.31 (i)). By Lemma 2.34 we have $\mathbf{x}_n^{@\mathcal{T}_n} \rightsquigarrow \mathbf{x}^{@\mathcal{T}}$ and $\mathbf{x}_n^{\setminus\mathcal{T}_n} \rightsquigarrow \mathbf{x}^{\setminus\mathcal{T}}$ both via (λ_n) for the set $\mathcal{T}_n := \{t_n^{(0)} := 0, t_n^{(1)}, \ldots, t_n^{(m)}\}$.
This allows us to write $\mathbf{x}_n = \mathbf{x}_n^{@\mathcal{T}_n} + \mathbf{x}_n^{\setminus\mathcal{T}_n}$, $\mathbf{x} = \mathbf{x}^{@\mathcal{T}} + \mathbf{x}^{\setminus\mathcal{T}}$ on $[0, N]$ and use the triangle-inequality once more to rewrite the required properties of (λ'_n). For convenience we summarize them now, we need:

$$(i)\ \lambda'_n = \lambda_n \text{ on } [t-\delta, t+\delta]^c, \qquad (ii)\ \lambda'_n(t) = t_n, \qquad (iii)\ \|\lambda'_n - \lambda_n\|_\infty \leq \tfrac{\varepsilon}{2},$$
$$(iv)\ \|\mathbf{x}_n^{\setminus\mathcal{T}_n} \circ \lambda'_n - \mathbf{x}_n^{\setminus\mathcal{T}_n} \circ \lambda_n\|_{N,\infty} \leq \tfrac{\varepsilon}{2} \qquad \text{and} \qquad (v)\ \|\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda'_n - \mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda_n\|_{N,\infty} = 0.$$

There is an index $0 \leq i \leq m-1$, s.t. $t \in (t^{(i)}, t^{(i+1)})$ (the point t cannot equal $t^{(i)}$ or $t^{(i+1)}$, since we assumed $\Delta\mathbf{x}(t) = 0$ and $t > 0$). Let $\eta > 0$ be s.t. $|t^{(j+1)} - t^{(j)}| < 2\eta$, for all $0 \leq j \leq k-1$, $|t^{(i)} - t| < 5\eta$, $|t - t^{(i+1)}| < 5\eta$ and $4\eta < \varepsilon$. Moreover, there is n_1 s.t. for all $n \geq n_1$ we have $|t_n^{(j)} - t^{(j)}| < \eta$ for all $0 \leq j \leq m$, what separates the tails of the sequences.

The clever choice of $\mathcal{T}$ ensures that $\Delta\mathbf{x}^{\setminus\mathcal{T}} < \frac{\varepsilon}{4}$ on $[0, N+1]$ and by Proposition 2.40 condition (iv) is satisfied if condition (iii) holds.
Now to the rest: Let us first find out how much space we have to vary λ_n without violating condition (v). If we assume wlog that $n' \geq n_1 \vee n_2$, where n_2 is such that $|t_n - t| < \eta$ for all $n \geq n_2$, then $|t_n^{(i)} - t_n| \geq 3\eta$ and $|t_n^{(i+1)} - t_n| \geq 3\eta$ for all $n \geq n'$. Now to satisfy condition (v) we set $\lambda'_n(s) = \lambda_n(s)$ for all $s \notin [t-\eta, t+\eta]$ (hence (i) is true, since we assumed that $4\eta < \varepsilon \leq \delta$) and assure that

$$(v)^* \sup_{s \in [t-\eta, t+\eta]} |\lambda'_n(s) - \lambda_n(s)| < 2\eta$$

holds. Indeed, this gives (v) as we will see soon. Besides it also implies (iii) because $4\eta < \varepsilon$ and thus (iv). To verify that $(v)^* \Rightarrow (v)$ we only need to look at the points where the sequences (λ_n) and (λ'_n) do not have to coincide a priori. On $[t-\eta, t+\eta]$ we know that $\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda_n \equiv \mathbf{x}_n^{@\{t_n^{(0)}, \ldots, t_n^{(i)}\}}(t_n^{(i)}) =: C$ is constant, since λ_n varies less than η from Id and the next jump $t_n^{(i)}$ or $t_n^{(i+1)}$ is more than 3η away from t. By condition $(v)^*$ the new time-shift λ'_n varies less than 2η from Id on $[t-\eta, t+\eta]$ and therefore cannot reach those jumps either. Consequently also $\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda'_n \equiv C$ on $[t-\eta, t+\eta]$ and hence (v) holds.
It only remains to satisfy condition (ii) without spoiling condition $(v)^*$. This is now rather simple, since we are allowed to vary λ_n by 2η on $[t-\eta, t+\eta]$. Now for $n \geq n'$ we assumed that $\|\lambda_n - \operatorname{Id}\|_\infty < \eta$ and $|t_n - t| < \eta$, thus setting $\lambda'_n(t) := t_n$ and interpolating

linearly on $[t-\eta, t+\eta]$ implies (ii), while $(v)^*$ still holds.

(ii) Assume first that for each $N \geq 1$ we have a sequence $\left(\lambda_n^{(N)}\right)$ s.t. $\lambda_n^{(N)} \xrightarrow{u} \mathrm{Id}$, $\|\mathbf{x}_n \circ \lambda_n^{(N)} - \mathbf{x}\|_{N,\infty} \to 0$ and $\lambda_n^{(N)}(t^{(i)}) = t_n^{(i)}$ for all $i \geq 1$ s.t. $t^{(i)} \in [0, N]$. Then $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n^*) by Proposition 2.41, where (λ_n^*) is the sequence we constructed in the proof of Proposition 2.41 with a diagonalization argument.
It remains to show that such sequences $(\lambda_n^{(N)})$ exist. This is now easy, since for every fixed N there are only finitely many elements $t^{(i)}, \ldots, t^{(m)}$ of $\mathcal{T}$ in the interval $[0, N]$, hence they have a positive distance exceeding $2\delta > 0$ and we can apply (i) iteratively. The fact that at each step we only change the sequence inside of a δ-neighbourhood of the $t^{(i)}$, ensures that any change happens in disjoint areas, having no influence on any preceding modification. □

Proposition 2.43. *Let $(\mathbf{x}_n)$ and $(\mathbf{y}_n)$ be sequences in $\mathcal{D}$ s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ and $\mathbf{y}_n \rightsquigarrow \mathbf{y}$ with the property that for every $t > 0$ there is a common sequence (t_n) satisfying $t_n \to t$ s.t. $\Delta\mathbf{x}_n(t_n) \to \Delta\mathbf{x}(t)$ and $\Delta\mathbf{y}_n(t_n) \to \Delta\mathbf{y}(t)$. Then $\mathbf{x}_n + \mathbf{y}_n \rightsquigarrow \mathbf{x} + \mathbf{y}$.*

Proof. We want to apply Proposition 2.41. Therefore we fix some $\eta > 0$ and have to find a sequence (λ_n) in Λ, s.t.

$$\lambda_n \xrightarrow{u} \mathrm{Id} \qquad \text{and} \qquad |(\mathbf{x}_n + \mathbf{y}_n) \circ \lambda_n - (\mathbf{x} + \mathbf{y})| \vee \eta \xrightarrow{ucs} \eta. \tag{2.8}$$

First, eliminate all jumps of size $\geq \frac{\eta}{4}$ from both $\mathbf{x}$ and $\mathbf{y}$. Accordingly, let $\mathcal{T} := \{t \geq 0 : |\Delta\mathbf{x}(t)| \vee |\Delta\mathbf{y}(t)| \geq \frac{\eta}{4}\} =: \{t^{(1)}, t^{(2)}, \ldots\}$ and $\mathcal{T}_n := \{t_n^{(i)}, i \geq 1\}$ for the sequences $(t_n^{(i)})$ from the assumption. By Lemma 2.34 the set $\mathcal{T}_n$ is ordered for all n. Moreover it implies that there are $m \geq 1$ and n_0 s.t. $\mathcal{T} \cap [0, N] = \{t^{(1)}, \ldots, t^{(m)}\}$ and $\mathcal{T}_n \cap [0, N] = \{t_n^{(1)}, \ldots, t_n^{(m)}\}$ for all $n \geq n_0$.
By assumption we can find sequences $(\lambda_n^{\mathbf{x}})$ and $(\lambda_n^{\mathbf{y}})$ in Λ s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via $(\lambda_n^{\mathbf{x}})$ and $\mathbf{y}_n \rightsquigarrow \mathbf{y}$ via $(\lambda_n^{\mathbf{y}})$. Part (ii) of Lemma 2.42 shows, that we can replace those sequences by $(\lambda_n^{\mathbf{x}*})$ and $(\lambda_n^{\mathbf{y}*})$ s.t. convergence still holds, but $\lambda_n^{\mathbf{x}*}(t^{(i)}) = t_n^{(i)}$ if $n \geq n_{\mathbf{x}}(i)$ and $\lambda_n^{\mathbf{y}*}(t^{(i)}) = t_n^{(i)}$ if $n \geq n_{\mathbf{y}}(i)$ for all $i \geq 1$.
Define $(\lambda_n) := (\lambda_n^{\mathbf{x}*})$ and prove that $\mathbf{x}_n + \mathbf{y}_n \rightsquigarrow \mathbf{x} + \mathbf{y}$ via (λ_n) by assuring (2.8). The first part of (2.8) is evident, so let us check the second

$$\begin{aligned} \|(\mathbf{x}_n + \mathbf{y}_n) \circ \lambda_n - (\mathbf{x} + \mathbf{y})\|_{N,\infty} \leq {} & \|\mathbf{x}_n^{\setminus \mathcal{T}_n} \circ \lambda_n - \mathbf{x}^{\setminus \mathcal{T}}\|_{N,\infty} + \|\mathbf{x}_n^{@\mathcal{T}_n} \circ \lambda_n - \mathbf{x}^{@T}\|_{N,\infty} \\ & + \|\mathbf{y}_n^{\setminus \mathcal{T}_n} \circ \lambda_n - \mathbf{y}^{\setminus \mathcal{T}}\|_{N,\infty} + \|\mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n - \mathbf{y}^{@T}\|_{N,\infty}. \end{aligned} \tag{2.9}$$

Lemma 2.34 shows that the first and the second term on the right-hand side tend to 0. Note that this is not the case for the last two terms of (2.9), since in general $(\lambda_n^{\mathbf{x}*}) \neq (\lambda_n^{\mathbf{y}*})$. Fortunately Proposition 2.40 doesn't care about what a time-shifting sequence does, as long as it converges uniformly to the identity. Here this is indeed the case and since $\mathbf{y}_n^{\setminus \mathcal{T}_n} \rightsquigarrow \mathbf{y}^{\setminus \mathcal{T}}$ we can apply it to find some n_1 s.t. $\|\mathbf{y}_n^{\setminus \mathcal{T}_n} \circ \lambda_n - \mathbf{y}^{\setminus \mathcal{T}}\|_{N,\infty} \leq \frac{\eta}{2}$ for all $n \geq n_1$.
For $n \geq n' := \max\{n_{\mathbf{x}}(i) \vee n_{\mathbf{y}}(i) \vee n_0 \vee n_1, 1 \leq i \leq m\}$ we have $\lambda_n^{\mathbf{x}*}\left(t^{(i)}\right) = \lambda_n^{\mathbf{y}*}\left(t_n^{(i)}\right) = t_n^{(i)}$

for all $1 \leq i \leq m$. Consequently

$$\begin{aligned}\mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n = \mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n^{\mathbf{x}*} &= \sum_{i=1}^{m} \Delta\mathbf{y}_n(t_n^{(i)}) \mathbb{1}_{\left[t_n^{(i)},\infty\right)} \circ \lambda_n^{\mathbf{x}*} = \sum_{i=1}^{m} \Delta\mathbf{y}_n(t^{(i)}) \mathbb{1}_{[t^{(i)},\infty)} \\ &= \sum_{i=1}^{m} \Delta\mathbf{y}_n(t_n^{(i)}) \mathbb{1}_{\left[t_n^{(i)},\infty\right)} \circ \lambda_n^{\mathbf{y}*} = \mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n^{\mathbf{y}*}\end{aligned}$$

for $n \geq n'$ and we can thus control the last term in (2.9) via

$$\|\mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n - \mathbf{y}^{@\mathcal{T}}\|_{N,\infty} = \|\mathbf{y}_n^{@\mathcal{T}_n} \circ \lambda_n^{\mathbf{y}*} - \mathbf{y}^{@\mathcal{T}}\|_{N,\infty} \to 0.$$

Plugging all results together, the left-hand side of (2.9) converges to 0, hence we know that also the non-trivial part of (2.8) holds and we are done. $\square$

3. Random Measures and Point Processes

This chapter contains basic results of the theory of random measures and is strongly fashioned after [15], although some results on point processes were generalized to random measures. Moreover, the texts [10] and [16] had a notable influence on this part of the text and also the book on infinitely divisible point processes from Matthes [13] gave some essential input.

The chapter as a whole has an introductory character. All results presented count to the basics of the theory of random measures and the captions read as a book on probability theory. We talk about moments (intensity), distributions, Laplace functionals and independence of random measures. The last section contains some results on independent increments which is a central property of a special class of random measures which we study in Chapter 4.

3.1. Motivation

Random measures are an essential ingredient of this text, since they will allow us to characterize weak convergence of partial sum processes to (some types of) Lévy processes. Now, before we fully devote ourselves to the theory, let us look at an easy example, giving a first insight into the relation between Lévy processes and an intuitive concept of a "random measure".

Example 3.1. Let $\mathbf{X} = (X_t)_{t\geq 0}$ be a Poisson process with parameter $\lambda \in (0,\infty)$. This is a Lévy process whose paths are a.s. pure jump functions and which has marginals $X_t \sim Poi(\lambda t)$. Such a path is completely characterized by its points of discontinuities $T_1(\omega), T_2(\omega), \ldots$ via $X_t(\omega) = \sum_{n\geq 1} \mathbb{1}_{[T_1(\omega),\infty)}(t)$. By putting unit-mass onto those times we get a measure $\mathbf{N}(\omega) := \sum_{k\geq 1} \delta_{T_k(\omega)}$ for any *fixed* $\omega \in \Omega$. Interpreting ω as an argument, we may suspect some relation between the random process $\mathbf{X}$ and the object $\mathbf{N}$. In fact "$X_t(\omega) = \mathbf{N}(\omega)[0,t]$ in distribution" as we will see later. This is the heart of the idea of describing weak convergence of partial sum processes in terms of corresponding weak convergence of "random measures" $\mathbf{N}^{(n)}$ and done rigorously in Chapter 6.

3.2. Measure Spaces and Definitions

3.2.1. Random Measures

We have to set some notation first. The underlying space, on which a "random measure" operates, is denoted by E. For most of the theory E could be arbitrary, but it is convenient to assume basic some properties. We let E be a *locally compact* space with a *countable base*, hence E is Hausdorff and second-countable. Furthermore we equip E with the corresponding Borel σ-algebra $\mathcal{B}_E$, although it should be noted that one can actually replace it by any other σ-algebra without much additional effort.

Definition 3.2. Let $(E, \mathcal{B}_E)$ be as above. Then $\mathcal{M}(E) := \{\nu : \mathcal{B}_E \to [0,+\infty] : \nu$ is a Radon measure$\}$ is the set of all (Radon) measures on $(E, \mathcal{B}_E)$. Accordingly, we call $\mathcal{M}_p(E) := \{\mu \in \mathcal{M}(E) : \mu \text{ is } \overline{\mathbb{N}}_0\text{-valued}\}$ the SET OF ALL (RADON) POINT MEASURES ON $(E, \mathcal{B}_E)$.

Remark 3.3. Unfortunately there is no consistent definition of Radon measures in the literature. We call a measure ν on $\mathcal{B}_E$ RADON if $\nu(A) < \infty$ for any relatively compact $A \in \mathcal{B}_E$. A measurable set A is RELATIVELY COMPACT if its closure $\overline{A}$ is compact.

It seems reasonable to define a random measure as a random element with values in $\mathcal{M}(E)$. So far this wouldn't make sense, since we haven't yet taken care of measurability, i.e. equipped $\mathcal{M}(E)$ with a decent σ-algebra $\mathscr{M}(E)$. The most natural thing to do with a measure ν is using it, i.e. calculating $\nu(A)$ for $A \in \mathcal{B}_E$. Consequently we choose the σ-algebra in such a way s.t. the maps $m_A : \mathcal{M}(E) \to [0, \infty]$, $\nu \mapsto \nu(A)$ are measurable. A natural choice is to take the smallest one having this property: the initial σ-algebra

$$\mathscr{M}(E) := \sigma(m_A, A \in \mathcal{B}_E). \tag{3.1}$$

Definition 3.4. A random element $\mathbf{N} := \mathbf{N}(\omega)$ with values in $\mathcal{M}(E)$ is called RANDOM MEASURE. The map $\mathbf{N}$ is thus a measurable function from the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathcal{M}(E), \mathscr{M}(E))$. The map $N_A := m_A \circ \mathbf{N} : (\Omega, \mathcal{A}) \to ([0, \infty], \mathcal{B}_{[0,\infty]})$ for $A \in \mathcal{B}_E$ is called RANDOM MEASURE OF A.

Remark 3.5. (i) A random measure $\mathbf{N}$ is finite/σ-finite/without atoms/... if the respective property holds for almost all $\mathbf{N}(\omega)$.

(ii) Note that on the one hand $\mathbf{N}(\omega, \cdot)$ is a measure on $\mathcal{B}_E$ for any fixed $\omega \in \Omega$ and on the other hand $\mathbf{N}(\cdot, A)$ is a random variable with values in $([0, \infty], \mathcal{B}_{[0,\infty]})$ for any fixed $A \in \mathcal{B}_E$ (see Proposition 3.6). So apart from the fact that $\mathbf{N}$ doesn't have to take values in $[0, 1]$, it is a stochastic kernel with source $(\Omega, \mathcal{A})$ and target $(\mathcal{M}(E), \mathscr{M}(E))$.

(iii) One can view $N_A = m_A \circ \mathbf{N}$ as some kind of „coordinate“ of $\mathbf{N}$ and m_A as some kind of „projection“.

(iv) As for random variables, we drop the argument ω for convenience, whenever it is not essential.

Proposition 3.6. *A map $\mathbf{N} : (\Omega, \mathcal{A}) \to (\mathcal{M}(E), \mathscr{M}(E))$ is a random measure iff $N_A : (\Omega, \mathcal{A}) \to ([0, \infty], \mathcal{B}_{[0,\infty]})$, $\omega \mapsto \mathbf{N}(\omega, A)$ is measurable for all $A \in \mathcal{B}_E$.*

Proof. ($\Rightarrow$) The map $m_A : (\mathcal{M}(E), \mathscr{M}(E)) \to ([0, \infty], \mathcal{B}_{[0,\infty]})$, $\nu \mapsto \nu(A)$ is measurable due to the definition of $\mathscr{M}(E)$ and therefore the composition $N_A = m_A \circ \mathbf{N}$ is measurable by assumption.

($\Leftarrow$) We check measurability of $\mathbf{N}$ using the *good sets principle*. Informally that is proving that all sets have the desired property and are thus "good", formally we want to prove that $\mathcal{G} := \{G \in \mathscr{M}(E) : \mathbf{N}^{-1}(G) \in \mathcal{A}\} = \mathscr{M}(E)$.
First note that $\mathscr{M}(E) = \sigma(m_A, A \in \mathcal{B}_E) = \sigma(m_A^{-1}(B), A \in \mathcal{B}_E, B \in \mathcal{B}_{[0,\infty]})$ by the definition of the initial σ-algebra and $m_A^{-1}(B) \in \mathcal{G}$ for any $B \in \mathcal{B}_{[0,\infty]}$ and $A \in \mathcal{B}_E$, since

$$\mathbf{N}^{-1}(m_A^{-1}(B)) = \mathbf{N}^{-1}(\{\nu \in \mathcal{M}(E) : \nu(A) \in B\}) = \{\omega \in \Omega : N_A(\omega) \in B\} \in \mathcal{A}$$

by assumption. It remains to show that $\mathcal{G}$ is a σ-algebra, because then $\mathscr{M}(E) = \sigma(m_A^{-1}(B) : A \in \mathcal{B}_E, B \in \mathcal{B}_{[0,\infty]}) \subseteq \mathcal{G} \subseteq \mathscr{M}(E)$. This follows easily from properties of the preimage and the fact that $\mathcal{A}$ is a σ-algebra. Indeed $\mathbf{N}^{-1}(\emptyset) = \emptyset \in \mathcal{A}$, for $G \in \mathcal{G}$ also $\mathbf{N}^{-1}(\mathcal{M}(E) \setminus G) = \mathbf{N}^{-1}(\mathcal{M}(E)) \setminus \mathbf{N}^{-1}(G) = \Omega \setminus \mathbf{N}^{-1}(G) \in \mathcal{A}$ and $\mathbf{N}^{-1}(\bigcup_{k\geq 1} G_k) = \bigcup_{k\geq 1} \mathbf{N}^{-1}(G_k) \in \mathcal{A}$ for $G_1, G_2, \dots \in \mathcal{G}$. □

A quite simple example of a random measure is induced by Lebesgue's measure λ. The idea is to first pick a real number and then manipulate the standard Lebesgue measure of a set by this value. Let's look at the details for one possible example.

Example 3.7. Take $\left((0,1], \mathcal{B}_{(0,1]}, \lambda\big|_{(0,1]}\right)$ as a probability space, $((0,1], \mathcal{B}_{(0,1]})$ as a state space and set $\mathbf{N}(x, A) := \frac{\lambda(A)}{x}$ for $x \in (0,1]$ and $A \in \mathcal{B}_{(0,1]}$.
Let us verify that $\mathbf{N}$ is a random measure. By Proposition 3.6 it suffices to check that $N_A^{-1}(B) \in \mathcal{B}_{(0,1]}$ for all $A \in \mathcal{B}_{(0,1]}$ and $B \in \mathcal{B}_{(0,\infty]}$. To obtain this claim, observe that

$$N_A^{-1}(B) = \{x \in (0,1] : \tfrac{\lambda(A)}{x} \in B\} = (0,1] \cap f^{-1}(B) \in \mathcal{B}_{(0,1]},$$

since $f : (0,1] \to [0,\infty]$, $f(x) := \frac{\lambda(A)}{x}$ is a continuous, hence (Borel-)measurable map.

From basic measure theory we know that in similar situations it suffices to check measurability for an advantageous subsystem having some weaker closedness properties than σ-algebras.

Definition 3.8. A family of *relatively compact* sets $\mathcal{E} \subseteq \mathcal{B}_E$ is called DETERMINING CLASS FOR $\mathcal{B}_E$ if $\mathcal{E}$ is a generating π-system (i.e. $\emptyset \neq \mathcal{E}$ is closed under intersections, $\sigma(\mathcal{E}) = \mathcal{B}_E$, see Definition A.1) and either one of the following conditions hold

(i) there is an EXHAUSTING SEQUENCE (E_n), i.e. $E_n \nearrow E$, $E_n \in \mathcal{E}$ for all n

(ii) there is a MEASURABLE PARTITION (E_n), i.e. a family of pairwise disjoint sets in $\mathcal{E}$ s.t. $\biguplus_{n\geq 1} E_n = E$.

Remark 3.9. In $\mathbb{R}/\mathbb{R}^d$ one can for instance take open/half-open/closed intervals/rectangles as simple and convenient examples.

Proposition 3.10. *Let $\mathcal{E}$ be a determining class for $\mathcal{B}_E$. Then $\mathbf{N}$ is a random measure iff N_A is measurable for all $A \in \mathcal{E}$. Moreover, $\sigma(m_A, A \in \mathcal{E}) = \mathscr{M}(E)$.*

Proof. Proposition 3.6 shows that necessity is trivial and that it suffices to proof measurability of N_A for all $A \in \mathcal{B}_E$. Suppose first that $\mathcal{E}$ contains an exhausting sequence (E_n) and use the good sets principle again.
Fix $n \geq 1$ and define $\mathcal{G} := \{G \in \mathcal{B}_E : N_{G\cap E_n} \text{ is measurable}\}$. By assumption $\mathcal{G}$ contains the generating π-system $\mathcal{E}$, so by Dynkin's λ-π Theorem, it suffices to show that $\mathcal{G}$ is a λ-system to conclude that $\mathcal{G} = \mathcal{B}_E$ (see Definition A.1 and Theorem A.2 for more details).
Since $E_n \in \mathcal{E}$ the map N_{E_n} is measurable and hence $E \in \mathcal{G}$. For a pairwise disjoint sequence (G_n) in $\mathcal{G}$ we know that $N_{(\biguplus_{k\geq 1} G_k)\cap E_n} = N_{\biguplus_{k\geq 1}(G_k\cap E_n)} = \sum_{k\geq 1} N_{G_k\cap E_n}$ is – as a sum of measurable functions – measurable. If $G \in \mathcal{G}$, then $N_{(G^c)\cap E_n} = N_{E_n} - N_{G\cap E_n}$ is a difference of two *finite*, non-negative random variables and thus measurable. Applying

Dynkin's λ-π-Theorem yields $\mathcal{G} = \mathcal{B}_E$ for all n. Finally the map $N_A = \lim_{n\to\infty} N_{A\cap E_n}$ is – as a limit of measurable functions – measurable for all $A \in \mathcal{G} = \mathcal{B}_E$.
If $\mathcal{E}$ has a measurable partition (E_n), apply the same argument to conclude that $\mathcal{G} = \mathcal{B}_E$. Observe that $N_A = \sum_{n\geq 1} N_{A\cap E_n}$ and the fact that measurability is closed under countable summation together imply the result in this case.
One readily checks that $\sigma(m_A, A \in \mathcal{E}) = \sigma(m_A, A \in \mathcal{B}_E) = \mathscr{M}(E)$. □

3.2.2. Point Processes

The most important subset of $\mathcal{M}(E)$ is the one of integer-valued measures, since their random counterparts will help us describe discontinuities of Lévy processes.

Definition 3.11. Let $\mathscr{M}_p(E) := \mathscr{M}(E)\big|_{\mathcal{M}_p(E)} = \{A \cap \mathcal{M}_p(E) : A \in \mathscr{M}(E)\}$ be the trace σ-algebra of $\mathcal{M}_p(E)$ in $\mathscr{M}(E)$. We call a random element $\mathbf{N}$ in $(\mathcal{M}_p(E), \mathscr{M}_p(E))$ a POINT PROCESS.

A short remark before we show that $\mathcal{M}_p(E)$ is actually well-defined: if $\nu \in \mathcal{M}_p(E)$, then one can write $\nu = \sum_{k\geq 1} c_k \delta_{x_k}$, where the points $x_1, x_2, \dots \in E$ are the atoms of ν having multiplicity $c_k = \nu(\{x_k\}) \in \overline{\mathbb{N}}_0$. On the other hand any measure having such a representation is evidently a point process again.

Proposition 3.12. *The σ-algebra $\mathcal{M}_p(E)$ is well-defined since $\mathcal{M}_p(E) \in \mathscr{M}(E)$.*

Proof. This proof uses some topological arguments, whose details are (partly) presented in the Appendix.
Pick a countable base (G_n) of relatively compact sets in E. This is possible since E is locally compact and second-countable (see Proposition C.2). Then we claim that $\mathcal{M}_p(E) \overset{(!)}{=} \bigcap_{n\geq 1}\{\nu \in \mathcal{M}(E) : \nu(G_n) \in \overline{\mathbb{N}_0}\} = \bigcap_{n\geq 1} m_{G_n}^{-1}(\overline{\mathbb{N}}_0)$. Since the right-hand side is a countable intersection of measurable sets, the result follows. Let us now verify this claim.
($\supseteq$) This is evident, since $\nu \in \mathcal{M}_p(E)$ is $\overline{\mathbb{N}}_0$-valued on any measurable set.

($\subseteq$) (i) Assume that ν is a measure having integer values on all base sets G_n and note that since it is Radon, any G_n has finite measure. We show first that all singletons have integer values as well. Due to our assumptions on E, any $x \in E$ satisfies the equality $\bigcap\{O \subseteq E \text{ open: } x \in U\} = \{x\}$. In fact this even holds for all first-countable spaces (see [14], p. 194). Since (G_n) is a base, there is some $m \geq 0$ s.t. $G_m \subseteq O$ for any open set $O \subseteq E$, thus $\bigcap_{n\geq 1}\{G_n : x \in G_n\} = \{x\}$. Moreover for any $m \geq 0$, the set $\bigcap_{n=0}^{m}\{G_n : x \in G_n\}$ is open and thus there is some $H_m \in \{G_n, n \geq 1\}$ s.t. $x \in H_m \subseteq \bigcap_{n=0}^{m}\{G_n : x \in G_n\}$. This yields an – wlog ordered – sequence of base sets $H_1 \supseteq H_2 \supseteq \dots$ s.t. $\lim_{n\to\infty} H_n = \{x\}$. By continuity of measures $\lim_{n\to\infty} \nu(H_n) = \nu(\{x\})$ and since the terms on the left-hand side are integer valued, so must be the right-hand side (a converging sequence of integers is finally constant). Consequently there is an element $H_x \in \{G_n, n \geq 1\}$ s.t. $\nu(\{x\}) = \nu(H_x) \in \overline{\mathbb{N}}_0$ for any $x \in E$.

(ii) For any compact set $K \subseteq E$ the open cover $\{H_x, x \in K\}$ contains a finite subcover, ergo there is a finite set S_K in K s.t. $\nu\big|_K = \sum_{x\in S_K} \nu(\{x\})\delta_x$ and ν is thus a point process on K. We can write $E = \lim_{m\to\infty} K_m$ for some compact sets K_m (e.g. $K_m :=$

$\overline{\bigcup_{j=1}^{m} G_j}$; a finite union of relatively compact sets is again relatively compact). Then $\nu = \sum_{x \in S} \nu(\{x\})\delta_x$ for $S = \bigcup_{m \geq 1} S_{K_m}$ is a point process. □

Remark 3.13. (i) The term point process is probably a bit misleading, since it is possible that it assigns mass greater 1 to singletons. Take $\mathbf{N} := 2\delta_0$ for a trivial example.

(ii) It is easy to see that the adapted results of Propositions 3.6 and 3.10 for $N_A : (\Omega, \mathcal{A}) \to (\overline{\mathbb{N}}_0, \mathcal{P}(\overline{\mathbb{N}}_0))$ ($\mathcal{P}$ denotes the power set function) are still valid for point processes. In fact, we can just replace $\mathcal{M}(E)$ and $\mathscr{M}(E)$ in the corresponding proofs by $\mathcal{M}_p(E)$ and $\mathscr{M}_p(E)$. Therefore we only have to ensure that $\mathscr{M}_p(E) = \sigma\big(m_A\big|_{M_p(E)}, A \in \mathcal{B}_E\big)$, which is an immediate consequence of Proposition A.3.

(iii) Note that for any $A \in \mathcal{B}_E$ the map N_A is measurable if $N_A^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{E}$, where $\mathcal{E}$ is such that $\sigma(\mathcal{E}) = \mathcal{B}_{[0,\infty]}$ for random measures or $\sigma(\mathcal{E}) = \mathcal{P}(\overline{\mathbb{N}}_0)$ for point processes.

Example 3.14 (n random points in E). Let $X := (X_1, \dots, X_n)$ be a random vector with values in E^n, $n \geq 1$. It serves as a model of a random choice of finitely many elements in a space. Equivalently, one can regard it as a point process via

$$\mathbf{N} = \mathbf{N}(\omega, A) := \sum_{k=1}^{n} \mathbb{1}_A(X_k(\omega)) = \sum_{k=1}^{n} \delta_{X_k(\omega)}(A).$$

Here measurability of $\mathbf{N}$ follows from Proposition 3.6, because the map $f(x_1, \dots, x_n) := \sum_{k=1}^{n} \mathbb{1}_A(x_k)$ is measurable and thus $N_A^{-1}(B) = (f \circ X)^{-1}(B) \in \mathcal{A}$ for any $A \in \mathcal{B}_E$, $B \in \mathcal{P}(\overline{\mathbb{N}}_0)$ (or $B \in \overline{\mathbb{N}}_0$ by Remark 3.13 *(iii)*).

Example 3.15 (N random points in E). We extend the preceding example to a random number of points. Therefore let (X_n) be a sequence of random elements in E and N be a $\overline{\mathbb{N}}_0$-valued random variable. Then we claim that

$$\mathbf{N}(\omega, A) := \sum_{k=1}^{N(\omega)} \mathbb{1}_A(X_k(\omega)) = \sum_{k=1}^{N(\omega)} \delta_{X_k(\omega)}(A)$$

is a random measure, which serves as a model for a choice of a random number of random points in E. The empty sum is hereby set to 0, i.e. $\mathbf{N} = 0$ on $\{N = 0\}$.
To ensure that $\mathbf{N}$ really is a point process, measurability is required. Recalling Proposition 3.10 and Remark 3.13 (iii), pick $A \in \mathcal{B}_E$ and note that for $m \geq 0$

$$\{N_A = m\} = \bigcup_{k \geq m} \bigg(\Big\{ \sum_{i=1}^{k} \mathbb{1}_A(X_i) = m \Big\} \cap \{N = k\} \bigg) \in \mathcal{A},$$

since both N and $\sum_{i=1}^{k} \mathbb{1}_A(X_i)$ are measurable.

Example 3.16 (Counting and point processes). Let $\mathbf{C} := (C_t)_{t \geq 0}$ be a COUNTING PROCESS, that is a $\overline{\mathbb{N}}_0$-valued stochastic process with $C_0 = 0$ which a.s. has jumps of size 1 and right-continuous paths. One may think of the Poisson process as an example.

Then there is a point process $\mathbf{N}$ s.t. $N_{[0,t]}(\omega) = C_t(\omega)$ for all $t \geq 0, \omega \in \Omega$.
To see this, let $\mathbf{N}$ assign unit-mass to the points where $\mathbf{C}$ jumps. Therefore define

$$X_k(\omega) := \inf\{t \geq 0 : C_t(\omega) \geq k\} \text{ for } k \geq 0 \text{ and } \inf\{\emptyset\} := \infty.$$

Each X_k is now a non-negative random variable, since C_t is measurable and $X_k^{-1}([0,t]) = \{X_k \leq t\} = \{C_t \geq k\} \in \mathcal{A}$, for all $t \geq 0$. Here the last equality holds due to the assumed right-continuity of $\mathbf{C}$. The point process $\mathbf{N} = \sum_{k\geq 1} \delta_{X_k}$ now satisfies $N_{[0,t]} = C_t$.

3.3. Intensity and Mean Measure

As for random variables we are interested in its moments, especially in the expected amount of mass a random measure $\mathbf{N}$ associates to certain sets.

Definition 3.17. For a given random measure $\mathbf{N}$ we call the map $\mu := \mu_{\mathbf{N}} : (E, \mathcal{B}_E) \to ([0,\infty], \mathcal{B}_{[0,\infty]}),\ A \mapsto \mathbb{E}[\mathbf{N}(A)] = \mathbb{E}[N_A] = \int_\Omega N_A(\omega)\mathrm{d}\mathbb{P}(\omega)$ the INTENSITY or MEAN MEASURE OF $\mathbf{N}$.

Proposition 3.18. *The map μ is indeed a measure on $\mathcal{B}_E$.*

Proof. For every $\omega \in \Omega$ we know that $\mathbf{N}(\omega, \cdot)$ is a measure, so $\mu(\emptyset) = \mathbb{E}[\mathbf{N}(\emptyset)] = 0$ since $\mathbf{N}(\omega, \emptyset) = 0$ for all ω. To prove σ-additivity pick pairwise disjoint $A_1, A_2, \ldots$ in $\mathcal{B}_E$ and note that

$$\mu\Big(\biguplus_{k\geq 1} A_k\Big) = \mathbb{E}\Big[\mathbf{N}\Big(\biguplus_{k\geq 1} A_k\Big)\Big] = \mathbb{E}\Big[\sum_{k\geq 1} N_{A_k}\Big] \overset{MCT}{=} \sum_{k\geq 1} \mathbb{E}\left[N_{A_k}\right] = \sum_{k\geq 1} \mu(A_k),$$

by the Monotone Convergence Theorem (MCT). $\square$

Naturally we are interested in the intensity of our two examples of point processes.

Example 3.19 (n random points in E). The mean measure of $\mathbf{N}$ given in Example 3.14 of $A \in \mathcal{B}_E$ is

$$\mu(A) = \mathbb{E}[N_A] = \mathbb{E}\left[\sum_{k=1}^n \mathbb{1}_A(X_k)\right] = \sum_{k=1}^n \mathbb{P}[X_k \in A].$$

In the special case that the sequence (X_k) is iid, the last term equals $n\mathbb{P}_{X_1}[A]$, where $\mathbb{P}_{X_1} := \mathbb{P} \circ \mathbf{X}_1^{-1}$ is the distribution of X_1.

Example 3.20 (N random points in E). If we assume that the $X_1, X_2, \ldots$ are independent of N (what is certainly reasonable in some situations), the corresponding mean measure of $\mathbf{N}$ given in Example 3.15 is

$$\begin{aligned}\mu(A) = \mathbb{E}[N_A] = \mathbb{E}\Big[\sum_{k=1}^N \mathbb{1}_A(X_k)\Big] &= \sum_{n\geq 0} \mathbb{E}\Big[\sum_{k=1}^N \mathbb{1}_A(X_k)\Big| N = n\Big]\mathbb{P}[N = n] \\ &\overset{ind.}{=} \sum_{n\geq 0} \mathbb{P}[N = n]\mathbb{E}\Big[\sum_{k=1}^n \mathbb{1}_A(X_k)\Big] \overset{\text{Ex. 3.14}}{=} \sum_{n\geq 0} \mathbb{P}[N = n] \sum_{k=1}^n \mathbb{P}[X_k \in A].\end{aligned}$$

In the special case that the sequence (X_k) is iid, we thus get

$$\sum_{n\geq 0} n\mathbb{P}[N = n]\mathbb{P}_{X_1}[A] = \mathbb{E}[N]\mathbb{P}_{X_1}[A].$$

If $\mathbf{N}$ is a random measure and its distribution is denoted by $\mathbb{P}_{\mathbf{N}} := \mathbb{P} \circ \mathbf{N}^{-1}$, we can write $\mu(A) = \int_{\mathcal{M}(E)} \nu(A) \mathrm{d}\mathbb{P}_{\mathbf{N}}(\nu)$. Consequently we can expect similar relations between μ and $\mathbf{N}$ as we have for random variables and its expectation.

Proposition 3.21. *Let* $\mathbf{N}$ *be a random measure with intensity* μ *and* $A \in \mathcal{B}_E$. *Then* $\mu(A) < \infty$ *implies* $N_A < \infty$ *a.s. and if* μ *is* (σ)*-finite, so is* $\mathbf{N}$ *almost surely. Moreover,* $N_A = 0$ *a.s. if* $\mu(A) = 0$.

Proof. Let $\mu(A) = \mathbb{E}[N_A] < \infty$. Assuming that $\mathbb{P}[N_A = \infty] > 0$ immediately gives a contradiction, since $\infty = \infty\mathbb{P}[N_A = \infty] \leq \mu(A)$, similarly one gets the last statement. The result on finite intensities follows by setting $A = E$. If μ is σ-finite but $\mathbb{P}[\mathbf{N}$ is not σ-finite$] > 0$, then any sequence (A_n) of measurable sets, s.t. $\mu(A_n) < \infty$ and $\mu(E) = \lim_{n\to\infty} A_n$ finally has to satisfy $\mathbb{P}[N_{A_n} = \infty] > 0$, contradicting the first result. $\square$

Remark 3.22. Note, that as for random variables, the converse statements are not true. In Example 3.7 we had $\mathbf{N}(x, A) \leq \frac{1}{x} < \infty$ for any $x \in (0, 1]$ and $A \in \mathcal{B}_{(0,1]}$, hence $\mathbf{N}$ was finite a.s., while its intensity μ is not finite, since

$$\mu((0, 1]) = \int_0^1 \mathbf{N}(x, (0, 1]) \mathrm{d}\lambda(x) = \int_0^1 \frac{1}{x} \mathrm{d}x = \infty.$$

This also happens for point processes as the following example shows.

Example 3.23. Set $N := \lceil X \rceil$ for $X \sim \mathit{Cauchy}$ in Example 3.15, where $\lceil \cdot \rceil$ denotes the "ceiling" function, that is the smallest integer greater than X, and Cauchy denotes the standard Cauchy distribution. In addition, the sequence (X_n) shall be iid with $X_1 \sim \mathit{Unif}([0, 1])$. Then $\mu(A) = \infty$ for any $A \in \mathcal{B}_E$, if A has positive Lebesgue measure due to the result in Example 3.20 and $\mathbb{E}[N] \geq \mathbb{E}[X] = \infty$. But N and therefore also $\mathbf{N}$ is finite (a.s.).

3.4. Distributions of Random Measures

By definition a random measure $\mathbf{N}$ is just a random element in the well-chosen measurable space $(\mathcal{M}(E), \mathscr{M}(E))$. Naturally we are interested in its distribution $\mathbb{P}_{\mathbf{N}}$ as the most important characteristic of a random object. Note that $\mathbb{P}_{\mathbf{N}}$ is a measure on $\mathscr{M}(E)$, i.e. a measure on the space of measures, what may seem a bit unhandy. Thus we very much appreciate the following result.

Proposition 3.24 (Characterizing $\mathbb{P}_{\mathbf{N}}$ via marginals). *Let* $\mathbf{N}$ *be a random measure on* $(E, \mathcal{B}_E)$ *and* $\mathcal{E}$ *be a determining class of* $\mathcal{B}_E$. *Then* $\mathbb{P}_{\mathbf{N}}$ *is completely determined by its (pairwise disjoint) finite dimensional distributions on* $\mathcal{E}$, *i.e. the distribution of the random vectors* $(N_{A_1}, \ldots, N_{A_n})$, $n \geq 1$ *for pairwise disjoint* $A_1, \ldots, A_n$ *in* $\mathcal{E}$.

Proof. Cf. [12], Theorem 24.5.
(i) By Proposition 3.10 the family $\{m_A^{-1}(B), A \in \mathcal{E}, B \in \mathcal{B}_{[0,\infty]}\}$ generates $\mathscr{M}(E)$ and so does $\{\bigcap_{k=1}^n m_{A_k}^{-1}(B_k) : A_k \in \mathcal{E}, B_k \in \mathcal{B}_{[0,\infty]}, 1 \leq k \leq n, n \geq 1\}$. Since the latter is

even a π-system the Uniqueness Theorem for measures (Theorem A.5) implies that the distribution $\mathbb{P}_{\mathbf{N}} = \mathbb{P} \circ \mathbf{N}^{-1}$ is determined by the values on the sets

$$\begin{aligned}\mathbf{N}^{-1}\Big(\bigcap_{k=1}^{n} m_{A_k}^{-1}(B_k)\Big) &= \bigcap_{k=1}^{n}\{\omega : \mathbf{N} \circ m_{A_k} \in B_k\} = \bigcap_{k=1}^{n}\{\omega : N_{A_k} \in B_k\} \\ &= \bigcap_{k=1}^{n} N_{A_k}^{-1}(B_k)\end{aligned}$$

for $n \geq 1$. But those values are determined by the random vectors $(N_{A_1}, \dots, N_{A_n})$ for $A_1, \dots, A_n$ in $\mathcal{E}$.

(ii) It remains to show is that it suffices to consider pairwise disjoint sets. Therefore fix $n \geq 1$ and sets $A_1, \dots, A_n \in \mathcal{E}$. Let $\mathcal{G} = \{G_1, \dots, G_m\} := \{C_1 \cap \dots \cap C_n : C_k \in \{A_k, A_k^c\}, 1 \leq k \leq n\}$ be the induced partition of E. Then $(N_{A_1}, \dots, N_{A_n}) = F(N_{G_1}, \dots, N_{G_m})$ for

$$F : [0,\infty]^m \to [0,\infty]^n, \qquad (x_1, \dots, x_m) \mapsto \Big(\sum_{j:G_j \in A_1} x_j, \dots, \sum_{j:G_j \in A_n} x_j\Big).$$

The map F is evidently measurable, consequently the distribution of $(N_{A_1}, \dots, N_{A_n})$ is given by the one of $(N_{G_1}, \dots, N_{G_m})$ via

$$\begin{aligned}\mathbb{P}_{(N_{A_1},\dots,N_{A_n})} &= \mathbb{P} \circ (N_{A_1}, \dots, N_{A_n})^{-1} = \mathbb{P} \circ (F(N_{G_1}, \dots, N_{G_m}))^{-1} \\ &= \mathbb{P} \circ (N_{G_1}, \dots, N_{G_m})^{-1} \circ F^{-1} = \mathbb{P}_{(N_{G_1},\dots,N_{G_m})} \circ F^{-1}\end{aligned} \tag{3.2}$$

and the result is proved. □

Remark 3.25. There is a similar simplification as in Remark 3.13 (iii). The distribution of $(N_{A_1}, \dots, N_{A_n})$ for $A_1, \dots, A_n \in \mathcal{B}_E$ is determined by its values on $N_A^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{G}$, s.t. $\sigma(\mathcal{G}) = \mathcal{B}_{[0,\infty]}$ for random measures and $\sigma(\mathcal{G}) = \mathcal{P}_{\overline{\mathbb{N}}_0}$ for point processes. We can for instance take rectangles in the first and singletons in the second case.

3.5. Random Integrals

This section widely agrees with the corresponding parts in [15], p. 126ff.

Since we can integrate w.r.t. measures, the same should be possible for random measures, although the integral will likely carry some randomness. That such a "random integral", as we define in the following, is indeed a random variable, will be checked soon. Moreover, we will see that it can be used to determine the law of the random measure.

Definition 3.26. We denote the SET OF ALL NON-NEGATIVE, MEASURABLE FUNCTIONS ON $(E, \mathcal{B}_E)$ by $\mathcal{B}_E^+ := \{f : (E, \mathcal{B}_E) \to ([0,\infty], \mathcal{B}_{[0,\infty]})$ measurable$\}$. For a given random measure $\mathbf{N}$ and $f \in \mathcal{B}_E^+$ the RANDOM INTEGRAL OF f W.R.T. $\mathbf{N}$ is defined as

$$N(f) : (\Omega, \mathcal{A}) \to ([0,\infty], B_{[0,\infty]}), \qquad \omega \mapsto \int_E f(y) \mathrm{d}\mathbf{N}(\omega, y) := \int_E f(y) \mathbf{N}(\omega, \mathrm{d}y).$$

The same handy notation $\nu(f) := \int_E f(y) \mathrm{d}\nu(y)$ is used for deterministic measures ν.

Remark 3.27. (i) As N_A for $A \in \mathcal{B}_E$, the map $N(f)$ can be interpreted as some kind of "coordinate" of $\mathbf{N}$.

(ii) Since $N(\omega, f)$ is a deterministic integral for every fixed $\omega \in \Omega$, $N(f)$ is (a.s.) linear and $\mathbb{R}^+$-homogeneous ($N(\lambda f + g) = \lambda N(f) + N(g)$ for $\lambda \geq 0$), monotone ($f \leq g \Rightarrow N(f) \leq N(g)$), and monotonly convergent ($f_n \nearrow f \Rightarrow N(f_n) \nearrow N(f)$).

Proposition 3.28. *Let* $\mathbf{N}$ *be a random measure with intensity* μ *and let* $f \in \mathcal{B}_E^+$. *Then* $N(f)$ *is a non-negative random variable and* $\mathbb{E}[N(f)] = \mu(f)$.

Proof. Let $f \in \mathcal{B}_E^+$. A classic result from measure theory says that there exist simple, measurable functions $f_n := \sum_{k=1}^{m_n} c_{k,n} \mathbb{1}_{A_{k,n}}$, for $A_{k,n} \in \mathcal{B}_E$ and $c_{k,n} \geq 0$ for all $k, n \geq 1$, s.t. $f_n \nearrow f$. Then $N(f_n) = \sum_{k=1}^{m_n} c_{k,n} N_{A_{k,n}}$ is – as a finite sum of random variables – measurable. Using the preceding remark we know that $N(f_n) \nearrow N(f)$ and therefore $N(f)$ is measurable as well.
Moreover,

$$
\begin{aligned}
\mathbb{E}[N(f)] &= \mathbb{E}\left[\lim_{n\to\infty} N(f_n)\right] \overset{MCT}{=} \lim_{n\to\infty} \mathbb{E}\left[\sum_{k=1}^{m_n} c_{k,n} N_{A_{k,n}}\right] = \lim_{n\to\infty} \sum_{k=1}^{m_n} c_{k,n}\mu(A_{k,n}) \\
&= \lim_{n\to\infty} \int_E f_n(y)\mathrm{d}\mu(y) \overset{MCT}{=} \int_E f(y)\mathrm{d}\mu(y) = \mu(f)
\end{aligned}
$$

by the Monotone Convergence Theorem. □

The extension of the random integral to not necessarily non-negative functions is standard. By looking at the positive and negative part $f^\pm = \max\{\pm f, 0\}$ of a measurable real-valued function, we write $f = f^+ - f^-$ and define $N(f) := N(f^+) - N(f^-)$. To guarantee that this difference avoids undefined terms like $\infty - \infty$, we need to assure that at least one of them is a.s. finite. Especially we can look at random integrals of $\mathcal{L}^1(\mu)$-functions, since Proposition 3.21 implies that $N(f^\pm) < \infty$ a.s. if $\mu(f^\pm) \leq \mu(|f|) = \|f\|_1 < \infty$.
Before we continue with the theory, let's take a quick look at our guiding examples.

Example 3.29 (n, N random points in E). For $\mathbf{N}$ from Example 3.15 and $f \in \mathcal{B}_E^+$ we have

$$
N(f) = \int_E f(y)\mathrm{d}\mathbf{N}(y) = \int_E f(y)\mathrm{d}\Big(\sum_{k=1}^{N} \delta_{X_k}\Big)(y) = \sum_{k=1}^{N} f(X_k). \tag{3.3}
$$

Taking $N \equiv n$ yields the result in the special case of n random points in E.

Another characterization of the distribution of a random measure in terms of random integrals is now easily established.

Corollary 3.30 (Characterizing $\mathbb{P}_\mathbf{N}$ via integrals). *Let* $\mathbf{N}$ *be a random measure and* $\mathcal{E}$ *a determining class of* $\mathcal{B}_E$. *Then the distribution of the vectors* $(N_{A_1}, \ldots, N_{A_n})$ *for* $n \geq 1$ *and pairwise disjoint* $A_1, \ldots, A_n \in \mathcal{E}$ *is determined by the distribution of the vectors* $(N(f_1), \ldots, N(f_m))$, *where* $f_k \in \mathcal{B}_E^+$ *for all* $k = 1, \ldots, m$, $m \geq 1$ *and vice versa. Consequently* $\mathbb{P}_\mathbf{N}$ *is also determined by the latter.*

Proof. The first statement is an immediate consequence of the fact that $f_k = \mathbb{1}_{A_k} \in \mathcal{B}_E^+$ and therefore $(N(f_1), \dots, N(f_n)) = (N_{A_1}, \dots, N_{A_n})$ for this special choice of functions. That $\mathbb{P}_{\mathbf{N}}$ is determined by $(N(f_1), \dots, N(f_n))$ follows easily from Proposition 3.24.
On the other hand we can approximate any $f \in \mathcal{B}_E^+$ by simple functions as in the proof of Proposition 3.28, hence the law of $(N(f_1), \dots, N(f_n))$ is given by the law of $(N_{A_1}, \dots, N_{A_m})$. □

3.6. Laplace Functionals

We quickly recall some of the basic properties of Laplace transforms and characteristic functions for vanilla random variables.

Definition 3.31. For a $\mathbb{R}^d$-valued random variable X (or its distribution $\mathbb{P}_X$) the CHARACTERISTIC FUNCTION φ_X is given by

$$\varphi_X : \mathbb{R}^d \to [0,1], \qquad u \mapsto \mathbb{E}[e^{iuX}] = \int_\Omega e^{iuX(\omega)} \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} e^{iuy} \mathrm{d}\mathbb{P}_X(y),$$

if X is non-negative, the map

$$\psi_X : [0,\infty)^d \to [0,1], \qquad u \mapsto \mathbb{E}[e^{-uX}] = \int_\Omega e^{uX(\omega)} \mathrm{d}\mathbb{P}(\omega) = \int_0^\infty e^{uy} \mathrm{d}\mathbb{P}_X(y)$$

is the LAPLACE TRANSFORM OF X (or its distribution $\mathbb{P}_X$).

Remark 3.32. (i) Formally we have $\varphi_X(u) = \psi_X(iu)$.

(ii) If $X \le 0$ a.s., one can apply the Laplace transform to $-X$, the point being that one just needs X to have constant sign and that non-negativity is not necessary but it suffices to look at this case.

Theorem 3.33 (Properties of characteristic functions). *Let $X, Y, X_1, X_2, \dots$ be some $\mathbb{R}^d$-valued random variables. Then*

(i) *X and Y are independent iff $\varphi_{(X,Y)}(u) = \varphi_X(u_1)\varphi_Y(u_2)$ for any $u = (u_1, u_2) \in \mathbb{R}^{d \times d}$*

(ii) *φ_X determines the distribution of X*

(iii) *If φ is $2n$ times differentiable at 0 for some $n \ge 1$, then*

$$\mathbb{E}[X^{2n}] = (-1)^n \Big[\frac{\mathrm{d}^{2n}}{\mathrm{d}u^{2n}} \varphi_X(u)\Big]_{t=0} < \infty$$

(iv) *Let in $X \in \mathcal{L}^m$. Then the characteristic function is n times differentiable at 0 for $n \le m$ and*

$$\mathbb{E}[X^n] = i^{-n} \Big[\frac{\mathrm{d}^n}{\mathrm{d}u^n} \varphi_X(u)\Big]_{t=0} \tag{3.4}$$

(v) *$\varphi_{X_n}(u) \to \varphi_X(u)$ for all $u \ge 0$ iff $X_n \Rightarrow X$.*

Proof. For the proof of (i) see [3], Theorem 2.1. For $(ii)/(iii)$ look at proofs [12], Theorem 15.8 and 15.34. The statements (iv) and (v) are provided in [6], section 26 (especially equation (26.9)) and Theorem 26.3 respectively. $\square$

Assuming additionally that X is either non-negative (or non-positive), (i), (ii) and (v) still hold, when using the Laplace transform instead. One can carry the concept of a Laplace transform over to (random) measures.

Definition 3.34. Let Q be a finite measure on $(\mathcal{M}(E), \mathscr{M}(E))$. Then the map

$$\Psi_Q : \mathcal{B}_E^+ \to [0,\infty), \qquad f \mapsto \int_{\mathcal{M}(E)} e^{-\nu(f)} \mathrm{d}Q(\nu).$$

is called LAPLACE TRANSFORM OF Q. It is a somewhat natural extension of a Laplace transform of the distribution of a random vector.
For a random measure $\mathbf{N}$, the map

$$\Psi_{\mathbf{N}} : \mathcal{B}_E^+ \to [0,1], \qquad f \mapsto \mathbb{E}[e^{-N(f)}] = \int_\Omega e^{-N(f)(\omega)} \mathrm{d}\mathbb{P}(\omega) = \int_{\mathcal{M}(E)} e^{-\nu(f)} \mathrm{d}\mathbb{P}_{\mathbf{N}}(\nu)$$

is the LAPLACE FUNCTIONAL OF $\mathbf{N}$. Note that $\Psi_{\mathbf{N}}$ is just the Laplace transform of $\mathbb{P}_{\mathbf{N}}$.

Corollary 3.35 (Characterizing $\mathbb{P}_{\mathbf{N}}$ via its Laplace functional). *The distribution of a random measure $\mathbf{N}$ is determined by its Laplace transform $\Psi_{\mathbf{N}}$.*

Proof. By Corollary 3.30 the measure $\mathbb{P}_{\mathbf{N}}$ is determined by the distribution of the non-negative vectors $(N(f_1), \dots, N(f_n))$, where $f_k \in \mathcal{B}_E^+$ for all $k = 1, \dots, n$ and $n \geq 0$. The distribution of any such vector is determined by its Laplace transform $\Psi_{(N(f_1),\dots,N(f_n))}(t_1, \dots, t_n) = \mathbb{E}\left[e^{-(t_1 N(f_1)+\dots+t_n N(f_n))}\right] = \Psi_{\mathbf{N}}(t_1 f_1 + \dots + t_n f_n)$ (here we used the linearity of the random integral). Ergo $\mathbb{P}_{\mathbf{N}}$ is determined by $\Psi_{\mathbf{N}}$. $\square$

Example 3.36 (n, N random points in E). Let $f \in \mathcal{B}_E^+$ and look at the point process of Example 3.14. Then for the special case where the sequence (X_k) is iid and independent from N we have

$$\Psi_{\mathbf{N}}(f) = \mathbb{E}[e^{-N(f)}] \overset{(3.3)}{=} \mathbb{E}\Big[exp\Big\{-\sum_{k=1}^n f(X_k)\Big\}\Big] \overset{ind.}{=} \prod_{k=1}^n \mathbb{E}[e^{-f(X_k)}] \overset{id.}{=} \Big(\mathbb{E}[e^{-f(X_1)}]\Big)^n.$$

The same but for Example 3.15 gives

$$\begin{aligned}\Psi_{\mathbf{N}}(f) &\overset{(3.3)}{=} \mathbb{E}\Big[exp\Big\{-\sum_{k=1}^N f(X_k)\Big\}\Big] = \sum_{n\geq 0} \mathbb{E}\Big[exp\Big\{-\sum_{k=1}^N f(X_k)\Big\}\Big| N = n\Big]\mathbb{P}[N=n] \\ &= \sum_{n\geq 0} \mathbb{E}\Big[exp\Big\{-\sum_{k=1}^n f(X_k)\Big\}\Big]\mathbb{P}[N=n] = \sum_{n\geq 0} \Big(\mathbb{E}[e^{-f(X_1)}]\Big)^n \mathbb{P}[N=n],\end{aligned}$$

using the preceding result in the last step. If additionally $N \overset{d}{=} Poi(\kappa)$ for some $\kappa \geq 0$, then

$$\begin{aligned}\Psi_{\mathbf{N}}(f) &= e^{-\kappa} \sum_{n\geq 0} \frac{\big(\kappa\mathbb{E}[e^{-f(X_1)}]\big)^n}{n!} = exp\{-\kappa + \kappa\mathbb{E}[e^{-f(X_1)}]\} = exp\{-\kappa\mathbb{E}[1 - e^{-f(X_1)}]\} \\ &= exp\left\{-\kappa \int_E 1 - e^{-f(y)} \mathrm{d}\mathbb{P}_{X_1}(y)\right\} = exp\left\{-\int_E 1 - e^{-f(y)} \mathrm{d}(\kappa\mathbb{P}_{X_1})(y)\right\}.\end{aligned}$$

The last case where the sequence (X_k) is iid and independent from $N \sim Poi(\kappa)$ will be of importance later.
We end this subsection with a simple, but useful property of the Laplace functional.

Proposition 3.37. *Let (f_n) be a sequence in $\mathcal{B}_E^+$ s.t. $f_n \nearrow f$ for some $f \in \mathcal{B}_E^+$ and let $\mathbf{N}$ be a random measure. Then $\Psi_{\mathbf{N}}(f_n) \searrow \Psi_{\mathbf{N}}(f)$.*

Proof. Random integrals are monotonically convergent, hence $N(f_n) \nearrow N(f)$ and thus $\Psi_{\mathbf{N}}(f_n) = \mathbb{E}[e^{-N(f_n)}] \overset{MCT}{\searrow} \mathbb{E}[e^{-N(f)}] = \Psi_{\mathbf{N}}(f)$, by monotone convergence. □

3.7. Independence of Random Measures

Since any family $\mathscr{N}$ of random measures consists of measurable members, we can talk about their independence, i.e. the independence of its finite subfamilies. We start this section with a rather technical result.

Proposition 3.38 (Independence of random measures). *Let $\mathscr{N}$ be a family of random measures and $\mathcal{E}$ a determining class of $\mathcal{B}_E$. Then $\mathscr{N}$ is independent iff for each finite choice of pairwise disjoint sets $A_1, \dots, A_n \in \mathcal{E}$ and $\mathbf{N}^{(1)}, \dots, \mathbf{N}^{(m)} \in \mathscr{N}$ the random vectors $\left(N_{A_1}^{(k)}, \dots, N_{A_n}^{(k)}\right)$, $1 \le k \le m$ are independent:*

$$\mathbb{P}_{\left(N_{A_1}^{(1)},\dots,N_{A_n}^{(1)},\dots,N_{A_1}^{(m)},\dots,N_{A_n}^{(m)}\right)} = \bigotimes_{k=1}^{m} \mathbb{P}_{\left(N_{A_1}^{(k)},\dots,N_{A_n}^{(k)}\right)}. \tag{3.5}$$

Proof. A family of random elements is independent iff all finite subfamilies are independent. Ergo we have to show that the independence of any fixed $\mathbf{N}^{(1)}, \dots, \mathbf{N}^{(m)} \in \mathscr{N}$ is equivalent to (3.5).
Expressing independence of the subfamily in terms of the random tupel $(\mathbf{N}^{(1)}, \dots, \mathbf{N}^{(m)})$ gives

$$\mathbb{P}_{(\mathbf{N}^{(1)},\dots,\mathbf{N}^{(m)})} = \bigotimes_{k=1}^{m} \mathbb{P}_{\mathbf{N}^{(k)}}. \tag{3.6}$$

The Uniqueness Theorem for measures (Theorem A.5) shows that equality in (3.6) holds if they agree on an appropriate subfamily. Consider the π-system $\mathscr{M}'(E) := \{\bigcap_{k=1}^n m_{A_k}^{-1}(B_k) : A_k \in \mathcal{E}, B_k \in \mathcal{B}_{[0,\infty]}, k = 1, \dots, n, n \ge 1\}$ which we already used in the proof of Proposition 3.24 to determine the law of one random measure. Now we have not one, but m random measures, hence we need equality of the measures in (3.6) on the generating π-system $(\mathscr{M}'(E))^m$. This means that for any $d_1, \dots, d_m \in \mathbb{N}$ and $A_j^{(i)} \in \mathcal{B}_E, 1 \le i \le m, 1 \le j \le d_k, 1 \le k \le m$ the vectors

$$\left(N_{A_1^{(1)}}^{(1)}, \dots, N_{A_{d_1}^{(1)}}^{(1)}\right), \dots, \left(N_{A_1^{(m)}}^{(m)}, \dots, N_{A_{d_m}^{(m)}}^{(m)}\right) \text{ are independent.} \tag{3.7}$$

Thus necessity is trivial, since (3.7) implies (3.5).
To get sufficiency, we borrow some more ideas from the proof of Proposition 3.24. Let $\mathcal{G} := \{G_1, \dots, G_l\}$ be the partition generated by all the sets $A_j^{(i)}$. Then the distribution

of the tupel $\left(N_{G_1}^{(1)}, \ldots, N_{G_l}^{(1)}, \ldots, N_{G_1}^{(m)}, \ldots, N_{G_l}^{(m)}\right)$ is known, since it is determined by (3.5). Moreover, $N_{A_j^{(i)}} = \sum_{k: G_k \subseteq \mathcal{A}_j^{(i)}} N_{G_k}$ for all i, j and hence

$$\left(N_{A_1^{(1)}}^{(1)}, \ldots, N_{A_{d_1}^{(1)}}^{(1)}, \ldots, N_{A_1^{(m)}}^{(m)}, \ldots, N_{A_{d_m}^{(m)}}^{(m)}\right) = F\left(N_{G_1}^{(1)}, \ldots, N_{G_l}^{(1)}, \ldots, N_{G_1}^{(m)}, \ldots, N_{G_l}^{(m)}\right), \tag{3.8}$$

for the measurable map

$$\begin{aligned} F : [0,\infty]^{lm} &\to [0,\infty]^{d_1+\cdots+d_m} \\ (x_{1,1}, \ldots, x_{l,1}, \ldots, x_{1,m}, \ldots, x_{l,m}) &\mapsto \Big(\sum_{k: G_k \in A_j^{(i)}} G_k \Big)_{1 \le j \le d_k, 1 \le i \le m}. \end{aligned}$$

The distribution of the left-hand side in (3.8) is now determined by the left-hand side of (3.5) and consequently the vectors in (3.7) are independent. □

The preceding independence condition is rather hard to check in general. We give another one in terms of the Laplace functional.

Proposition 3.39 (Independence and Laplace functional). *Two random measures* $\mathbf{M}$ *and* $\mathbf{N}$ *are independent iff*

$$\begin{aligned} \Psi_{(\mathbf{M},\mathbf{N})}(f,g) &= \mathbb{E}\left[e^{-(M(f)+N(g))}\right] = \mathbb{E}\left[e^{-M(f)}\right]\mathbb{E}\left[e^{-N(g)}\right] \\ &= \Psi_{\mathbf{M}}(f)\Psi_{\mathbf{N}}(g), \end{aligned} \tag{3.9}$$

for every $f, g \in \mathcal{B}_E^+$.

Proof. Necessity is evident. For sufficiency it suffices to show that (3.9) is equivalent to the independence of the vectors $(M_{A_1}, \ldots, M_{A_m})$ and $(N_{B_1}, \ldots, N_{B_n})$ for pairwise disjoint $A_i, B_j \in \mathcal{B}_E, 1 \le i \le m, 1 \le j \le n$ and $m, n \ge 1$ due to Proposition 3.38. By Corollary 3.30 this is equivalent to the independence of the vectors $(M(f_1), \ldots, M(f_m))$ and $(N(g_1), \ldots, N(g_n))$ for $f_1, \ldots, f_m, g_1, \ldots, g_n \in \mathcal{B}_E^+$. Recalling the proof of Corollary 3.35 we see that this is again equivalent to (3.9) for $f := s_1 f_1 + \cdots + s_m f_m$ and $g := t_1 g_1 + \cdots + t_n g_n$. □

Remark 3.40. With a basic induction argument we can extend the result of Proposition 3.39 to finitely many random measures. Thus we have found another way to check the independence of a family of random measures. Often this is more convenient than trying to verify (3.5).

3.8. Independent Increments

In the very beginning of this chapter we described a Poisson Process $\mathbf{X}$ via a random measure $\mathbf{N}$ putting unit-mass on the times when $\mathbf{X}$ jumps. In the long run we want to do the same for general Lévy processes. Since increments of Lévy processes have nice properties, one may suspect that an “associated random measure” inherits them in a way. It is therefore reasonable to look at classes of random measures having some special properties.

Definition 3.41. A random measure $\mathbf{N}$ is said to have INDEPENDENT INCREMENTS if $N_{A_1}, \dots, N_{A_n}$ are independent for pairwise disjoint sets $A_1, \dots, A_n$ in $\mathcal{B}_E$ and all $n \geq 1$.

The following argument shows that it suffices to check these conditions on a smaller subfamily of sets. This time we need some more closedness properties than before, simply being $\cap$-stable doesn't suffice any more.

Definition 3.42. A family of relatively compact sets $\mathcal{H} \subseteq \mathcal{B}_E$ is called SEMIRING ON E if

(i) $\emptyset \in \mathcal{H}$

(ii) $H_1, H_2 \in \mathcal{H} \implies H_1 \cap H_2 \in \mathcal{H}$

(iii) For $H_2 \subseteq H_1 \in \mathcal{H}$ there are pairwise disjoint $C_1, \dots, C_n \in \mathcal{H}$ s.t. $H_1 \setminus H_2 = \biguplus_{k=1}^n C_k$.

If $\mathcal{H}$ additionally satisfies either condition (i) or (ii) of Definition 3.8 it is called DETERMINING SEMIRING FOR $\mathcal{B}_E$.

Remark 3.43. One can easily extend condition (iii) in the above definition to get $H_1 \setminus (H_2 \cup \dots \cup H_m) = \biguplus_{k=1}^{n_m} C_k$ for pairwise disjoint semiring elements $H_l \subseteq H_1$, $1 \leq l \leq m$ and $C_k \in \mathcal{H}$, $1 \leq k \leq n_m$.

Lemma 3.44. *Let $\mathbf{N}$ be a random measure and $\mathcal{H}$ a determining semiring of $\mathcal{B}_E$. If for all $n \geq 1$ the random variables $N_{A_1}, \dots, N_{A_n}$ for pairwise disjoint $A_1, \dots, A_n$ in $\mathcal{H}$ are independent, then $\mathbf{N}$ has independent increments.*

Proof. It suffices to show that $N_{A \setminus \biguplus_{k=2}^n A_k}, N_{A_2}, \dots, N_{A_n}$ are independent for any $A \in \mathcal{B}_E$ and pairwise disjoint $A_2, \dots, A_n \in \mathcal{B}_E$. Now if $\mathcal{H}$ contains a countable partition (E_k), this is equivalent to the independence of $N_{(A \cap E_m) \setminus \biguplus_{k=2}^n A_k}, N_{A_2}, \dots, N_{A_n}$ for all m, since $N_{A \setminus \biguplus_{k=2}^n A_k} = \sum_{m \geq 1} N_{(A \cap E_m) \setminus \biguplus_{k=2}^n A_k}$. If $\mathcal{H}$ contains an exhausting sequence the argument is similar, one just has to replace the sum by a limit. To prove independence fix m and use the good sets principle once more. Define

$$\begin{aligned} \mathcal{G}_{1,m} := \{A_1 \in \mathcal{B}_E : {} & N_{(A_1 \cap E_m) \setminus (\biguplus_{k=2}^n B_k)}, N_{B_2}, \dots, N_{B_n} \text{ are independent for all} \\ & \text{pairwise pairwise disjoint } B_2, \dots, B_n \in \mathcal{H}\}. \end{aligned}$$

By Remark 3.43 we can write $(A_1 \cap E_m) \setminus (\biguplus_{k=2}^n B_k)$ as a pairwise disjoint union of sets in $\mathcal{H}$. Together with the fact that summing up is a measurable transformation this yields $\mathcal{H} \subseteq \mathcal{G}_{1,m}$.
Let us show that indeed $G_{1,m} = \mathcal{B}_E$ with Dynkin's λ-π Theorem (Theorem A.2). That $E \in \mathcal{G}_{1,m}$ is evident, since $E \cap E_m = E_m \in \mathcal{H}$. To show that $\mathcal{G}_{1,m}$ is closed under complements, pick $A \in \mathcal{G}_{1,m}$ and note that

$$N_{(A^c \cap E_m) \setminus (\biguplus_{k=2}^n B_k)} = N_{(E \cap E_m) \setminus (\biguplus_{k=2}^n B_k)} - N_{(A \cap E_m) \setminus (\biguplus_{k=2}^n B_k)},$$

is – as a difference of two *finite* random variables being independent from $N_{B_2}, \dots, N_{B_k}$ – independent from them. For pairwise disjoint $A_1^{(1)}, A_1^{(2)}, \dots$ in $\mathcal{G}_{1,m}$, set $A_1 := \biguplus_{l \geq 1} A_1^{(l)}$. Then

$$N_{(A_1 \cap E_m) \setminus (\biguplus_{k=2}^n B_k)} = \sum_{l \geq 1} N_{\left(A_1^{(l)} \cap E_m\right) \setminus (\biguplus_{k=2}^n B_k)},$$

which is again independent of $B_2, \dots, B_n$, since all terms in the sum on the right-hand side are. Dynkin's λ-π Theorem thus implies $\mathcal{G}_{1,m} = B_E$.

(ii) We slightly alter the above argument. Therefore fix some $\tilde{m}$ and set

$$\mathcal{G}_{2,\tilde{m}} := \{A_2 \in \mathcal{B}_E : N_{A_1}, N_{(A_2\cap E_{\tilde{m}})\setminus(A_1\uplus(\biguplus_{k=3}^n B_k))}, N_{B_3}, \dots, N_{B_n} \text{ are independent for pairwise disjoint } B_3, \dots, B_m \in \mathcal{H},\ A_1 \in \mathcal{B}_E\}.$$

Since m was arbitrary in the first part, $N_{A_1\setminus(\biguplus_{k=2}^n B_k)}, N_{B_2}, \dots, N_{B_n}$ are independent for pairwise disjoint $B_2, \dots, B_m \in \mathcal{H}$ and *any* $A_1 \in \mathcal{B}_E$, thus $\mathcal{H} \in \mathcal{G}_{2,\tilde{m}}$. The exact same procedure as above shows that $\mathcal{G}_{2,\tilde{m}} = \mathcal{B}_E$. Iterating this process gives

$$\mathcal{G}_{n,m} := \{A_n \in \mathcal{B}_E : N_{A_1}, \dots, N_{A_{n-1}}, N_{(A_n\cap E_m)\setminus(\biguplus_{k=1}^{n-1} A_k)} \text{ are independent for disjoint } A_1 \dots, A_{n-1} \in \mathcal{B}_E\} = \mathcal{B}_E.$$

for all m and we are done. $\square$

Example 3.45 (n, N random points in E and independent increments). Compare to [15], Proposition 3.6.
For Example 3.14 independent increments are not possible (apart from trivial cases). Let's consider for instance the special case where $\mathbf{N} = \sum_{k=1}^n \delta_{X_k}$ for iid $X_1, \dots, X_n$. As soon as there is a set $A \in \mathcal{B}_E$ which is non-trivial, meaning that $\mathbb{P}[X_1 \in A] \in (0,1)$, both $\mathbb{P}[N_A = n] > 0$ and $\mathbb{P}[N_{A^c} = n] > 0$ but $\mathbb{P}[\{N_A = n\} \cap \{N_{A^c} = n\}] = 0$. Consequently $\mathbf{N}$ cannot have independent increments.

Fortunately it is possible to achieve independent increments for the model of N random points in space. To see that, look at $\mathbf{N} := \sum_{k=1}^N X_k$ in the special case where (X_n) is an iid sequence with distribution $Q := \mathbb{P}_{X_1}$ and $N \sim Poi(\kappa)$ for some parameter $\kappa \in [0, \infty)$. Additionally, N shall be independent of the sequence.
To begin we show that $N_A \sim Poi(\kappa Q(A))$ for any $A \in \mathcal{B}_E$. For $m \geq 0$

$$\begin{aligned}
\mathbb{P}[N_A = m] &= \mathbb{P}\Big[\biguplus_{l\geq m}\Big(\Big\{\sum_{j=1}^l \delta_{X_j}(A) = m\Big\} \cap \{N = l\}\Big)\Big] \\
&\overset{ind.}{=} \sum_{l\geq m} \mathbb{P}\Big[\sum_{j=1}^l \delta_{X_j}(A) = m\Big]\mathbb{P}[N = l].
\end{aligned}$$

The first probability in the second line is binomial (the sequence (X_k) is iid) and has parameter $Q(A)$, thus

$$\begin{aligned}
\mathbb{P}[N_A = m] &= \sum_{l\geq m} \binom{l}{m} Q(A)^m (1 - Q(A))^{l-m} \frac{e^{-\kappa}\kappa^l}{l!} \\
&= \sum_{l\geq m} \frac{(\kappa(1 - Q(A)))^{l-m}}{(l-m)!} \frac{e^{-\kappa}(\kappa Q(A))^m}{m!} \\
&= \frac{e^{-\kappa}(\kappa Q(A))^m}{m!} \sum_{l\geq 0} \frac{(\kappa(1 - Q(A)))^l}{l!} = e^{-\kappa Q(A)} \frac{(\kappa Q(A))^m}{m!}. \qquad (3.10)
\end{aligned}$$

The next step is to prove that increments are independent for measurable partitions $E_0, \dots, E_n$ of E. By Proposition 3.44 it suffices to show that $\mathbb{P}[N_{E_0} = m_0, \dots, N_{E_n} = m_n]$ factorizes for all $m_i \geq 0$. Define $m := \sum_{i=0}^n m_i$ and note that for $m \geq 0$

$$\begin{aligned}\mathbb{P}[N_{E_0} = m_0, \dots, N_{E_n} = m_n] &= \mathbb{P}\Big[\bigcap_{j=0}^n \{N_{E_j} = m_j\} \cap \{N = m\}\Big] \\ &= \mathbb{P}\Big[\sum_{k=1}^m \delta_{X_k}(E_0) = m_0, \dots, \sum_{k=1}^m \delta_{X_k}(E_n) = m_n\Big]\mathbb{P}[N = m].\end{aligned}$$

We assumed that (X_k) is iid, hence the first probability is multinomial and equals

$$\binom{m}{m_0 \dots m_n} \prod_{i=0}^n Q(E_i)^{m_i} = \frac{m!}{m_0! \dots m_n!} \prod_{i=0}^n Q(E_i)^{m_i} = m! \prod_{i=0}^n \frac{Q(E_i)^{m_i}}{m_i!}.$$

Furthermore $E_0, \dots, E_n$ is a partition, hence $\sum_{i=0}^n Q(E_i) = 1$ and therefore $e^{-\kappa} = \prod_{i=0}^n e^{-\kappa Q(E_i)}$. This gives

$$\begin{aligned}\mathbb{P}[N_{E_0} = m_0, \dots, N_{E_n} = m_n] &= e^{-\kappa}\kappa^m \prod_{i=0}^n \frac{Q(E_i)^{m_i}}{m_i!} = \prod_{i=0}^n e^{-\kappa Q(E_i)} \frac{(\kappa Q(E_i))^{m_i}}{m_i!} \\ &\overset{(3.10)}{=} \prod_{i=0}^n \mathbb{P}[N_{E_i} = m_i]. \end{aligned} \tag{3.11}$$

Now to the general case. For any pairwise disjoint sets $E_1, \dots, E_n \in \mathcal{B}_E$ set $E_0 := E \setminus \biguplus_{j=1}^n E_j$ to get a measurable partition. Then

$$\begin{aligned}\mathbb{P}\Big[\bigcap_{j=1}^n \{N_{E_j} = m_j\}\Big] &= \sum_{m_0 \geq 0} \mathbb{P}\Big[\bigcap_{j=0}^n \{N_{E_j} = m_j\}\Big] \overset{(3.11)}{=} \sum_{m_0 \geq 0} \mathbb{P}[N_{E_0} = m_0] \prod_{j=1}^n \mathbb{P}[N_{E_j} = m_j] \\ &= \prod_{j=1}^n \mathbb{P}[N_{E_j} = m_j] \underbrace{\sum_{m_0 \geq 0} \mathbb{P}[N_{E_0} = m_0]}_{=1} = \prod_{j=1}^n \mathbb{P}[N_{E_j} = m_j]\end{aligned}$$

for any $m_1, \dots, m_n \geq 0$ and thus $\mathbf{N}$ has independent increments.

Theorem 3.46 (Characterization of $\mathbb{P}_{\mathbf{N}}$). *Let $\mathbf{N}$ be a random measure with independent increments. Then its distribution $\mathbb{P}_{\mathbf{N}}$ is determined by either*

(i) *its „marginals“, i.e. the distribution of N_A for all $A \in \mathcal{E}$, where $\mathcal{E}$ is a determining class of $\mathcal{B}_E$, or*

(ii) *its “function-coordinates”, i.e. the distribution of $N(f)$ for all $f \in \mathcal{B}_E^+$, or*

(iii) *its Laplace functional, i.e. $\Psi_{\mathbf{N}}(f)$ for $f \in \mathcal{B}_E^+$.*

Proof. (i) By Proposition 3.24 the distribution $\mathbb{P}_{\mathbf{N}}$ is determined by the knowledge of $(N_{A_1}, \dots, N_{A_n})$ for pairwise disjoint $A_1, \dots, A_n \in \mathcal{E}$ and $n \geq 1$. Due to the independence of the increments of $\mathbf{N}$ we get

$$\mathbb{P}_{(N_{A_1}, \dots, N_{A_n})} = \bigotimes_{k=1}^n \mathbb{P}_{N_{A_k}},$$

hence the left-hand side is determined by knowledge of the right-hand side.

(ii) This result follows from (i), since $f = \mathbb{1}_A \in \mathcal{B}_E^+$ for all $A \in \mathcal{B}_E$ (recall the proof of Corollary 3.30).

(iii) An immediate consequence of (ii) (recall the proof of Corollary 3.35). $\square$

Proposition 3.47. *Let* $(\mathbf{N}^{(k)})$ *be a sequence of independent random measures with independent increments. Then the sum* $\mathbf{N} := \sum_{k\geq 1} \mathbf{N}^{(k)}$ *is a random measure with independent increments.*

Proof. First of all, $\mathbf{N}$ is a random measure since measurability is stable under countable summation. Let $A_1, \dots, A_n \in \mathcal{B}_E$ be pairwise disjoint sets. We have to show that the components of

$$\big(N_{A_1}, \dots, N_{A_n}\big) = \Big(\Big(\sum_{k\geq 1} N^{(k)}\Big)_{A_1}, \dots, \Big(\sum_{k\geq 1} N^{(k)}\Big)_{A_n}\Big) = \Big(\sum_{k\geq 1} N^{(k)}_{A_1}, \dots, \sum_{k\geq 1} N^{(k)}_{A_n}\Big)$$

are independent. The random measures $(\mathbf{N}^{(k)})$ are independent and so are the vectors $\big(N^{(1)}_{A_n} \dots, N^{(1)}_{A_1}\big)$, $\big(N^{(2)}_{A_n}, \dots, N^{(2)}_{A_1}\big), \dots$ by Proposition 3.38. Moreover, each vector consists of independent random variables, since each $\mathbf{N}^{(k)}$ has independent increments. Consequently, the whole family $\{N^{(k)}_{A_j}, 1 \leq j \leq n, k \geq 1\}$ is independent and so are the components of their image under the measurable function $F : \{1, \dots, n\} \times \mathbb{N} \to \overline{\mathbb{R}}^n$, $\big(x^{(k)}_i, 1 \leq i \leq n, k \geq 1\big) \mapsto \big(\sum_{k\geq 1} x^{(k)}_1, \dots, \sum_{k\geq 1} x^{(k)}_n\big)$. $\square$

4. Poisson Random Measures and Vague Topology

This chapter is roughly divided into three parts. The first consists of Sections 4.1 and 4.2 and has a rather introductory character. It provides some basic results on PRM, as well as an analytical characterization in terms of Laplace functionals. The second part (Sections 4.3 and 4.4) is crucial for our purposes and partly based on a lecture of Prof. Zweimüller held at the University of Vienna, winter term 13/14. It provides first connections between special classes of point processes and stochastic processes (notably Compound Poisson processes and martingales), which will play an important role in the next chapter.
The last part follows the course of Chapters 3.4 and 3.5 in [15] and is devoted to the theory of vague and weak convergence. Section 4.5 deals with the vague topology and contains a proof of the well-known result that the measure spaces $\mathcal{M}(E)$ and $\mathcal{M}_p(E)$ endowed with the vague topology are Polish. The very last section then provides some useful conditions for weak convergence of random measures/point processes, in particular a proof of Kallenberg's Theorem is presented.

4.1. Definition, Construction and Basic Properties

Definition 4.1. A measure ν is called Σ-FINITE if there exist finite measures (ν_k) s.t. $\nu = \sum_{k\geq 1} \nu_k$. A POISSON RANDOM MEASURE WITH INTENSITY μ is a point process $\mathbf{N}$ on E with Σ-finite mean measure μ, s.t.

(i) $\mathbf{N}$ has independent increments

(ii) $N_A \sim Poi(\mu(A))$ for all $A \in \mathcal{B}_E$.

We use the abbreviations PRM or $PRM(\mu)$ and write $\mathbf{N} \sim PRM(\mu)$.

Remark 4.2. (i) By Theorem 3.46 the law of $\mathbf{N}$ is completely determined by (ii), in fact it even suffices to check it on a determining class $\mathcal{E}$ for $\mathcal{B}_E$.

(ii) It is not yet clear that PRM exists! So far we have only shown existence of PRM having finite mean measure in Example 3.45. There we even gave an explicit construction, justifying the importance of our guiding point process examples.

(iii) The reason why we introduced Σ-finite measures is purely technical. On the one hand we want a countable sum of PRM to be a PRM again so we need this new property, since a sum of σ-finite measures doesn't have to be σ-finite again. On the other hand defining a PRM for arbitrary intensity measures, wouldn't enable us to show existence.

Before we give a construction in the case where μ is Σ-finite (and therewith show that PRM exists), we prove some basic properties.

Proposition 4.3 (Restrictions and Sums of PRM).

(i) *Sums of independent PRM are again PRM. More precisely, if* $(\mathbf{N}^{(k)})$ *with* $\mathbf{N}^{(k)} \sim PRM(\mu_k)$ *is an independent sequence, then*

$$\mathbf{N} := \sum_{k\geq 1} \mathbf{N}^{(k)} \sim PRM\Big(\sum_{k\geq 1} \mu_k\Big).$$

(ii) *Let* $\mathbf{N} \sim PRM(\mu)$. *For any* $F \in \mathcal{B}_E$ *the restriction* $\mathbf{N}\big|_F$ *is again a PRM, with restricted intensity, i.e.* $\mathbf{N}\big|_F := \mathbf{N}\big|_{\Omega\times(\mathcal{B}_E\cap F)} \sim PRM(\mu\big|_F)$.

(iii) *If* $E_1, E_2, \dots$ *in* $\mathcal{B}_E$ *are pairwise disjoint, the family* $\{\mathbf{N}\big|_{E_k}, k \geq 1\}$ *is independent. Moreover, the family* $\{N\big|_{E_k}(f) = N(\mathbb{1}_{E_k} f), k \geq 1\}$ *is independent for any* $f \in \mathcal{B}_E^+$.

Proof. (i) The sum $\mathbf{N}$ is again a point process by Proposition 3.47 and its intensity μ is given by $\mu = \sum_{k\geq 1} \mu_k$. Indeed, for any $A \in \mathcal{B}_E$

$$\mathbb{E}[N_A] = \mathbb{E}\Big[\sum_{k\geq 1} N_A^{(k)}\Big] \overset{MCT}{=} \sum_{k\geq 1} \mathbb{E}[N_A^{(k)}] = \sum_{k\geq 1} \mu_k(A) = \mu(A).$$

It is easy to see that μ is again Σ-finite and thus it remains to show that $N_A \sim Poi(\mu(A))$ for any $A \in \mathcal{B}_E$, what is, however, a well-known property of the Poisson distribution.

(ii) The restricted PRM is evidently a point process again since F is measurable and furthermore $\mathbb{E}[\mathbf{N}\big|_F(A)] = \mathbb{E}\big[N_A \mathbb{1}_F\big] = \mathbb{E}[N_{A\cap F}] = \mu(A \cap F) = \mu\big|_F(A)$ for any $A \in \mathcal{B}_E$.

(iii) Wlog we have to show independence of the subfamily $\mathbf{N}\big|_{E_1}, \dots, \mathbf{N}\big|_{E_n}$ by means of Proposition 3.38. For a fixed $m \geq 1$ and all $1 \leq k \leq n$ pick pairwise disjoint $A_1^{(k)}, \dots, A_m^{(k)} \in \mathcal{B}_E \cap E_k$ and show that $(N_{A_1^{(1)}}, \dots, N_{A_m^{(1)}}), \dots, (N_{A_1^{(n)}}, \dots, N_{A_m^{(n)}})$ are independent random vectors. Actually, we can even do more. Since the sets E_k, $k \geq 1$ are pairwise disjoint, the whole family $\big\{A_j^{(k)}, 1 \leq j \leq m, 1 \leq k \leq n\big\}$ is pairwise disjoint and hence $\{N_{A_j^{(k)}}, 1 \leq j \leq m, 1 \leq k \leq n\}$ is a system of independent random variables. Especially this implies independence of the vectors.
The statement about the family of random integrals is now a immediate consequence since we can approximate them by simple functions. □

This proposition already indicates how to construct a PRM if the intensity is Σ-finite.

Proposition 4.4 (Existence of PRM). *Poisson random measures exist.*

Proof. Cf. [12], Theorem 24.12.
For Σ-finite μ there is a sequence of finite measures (μ_k) s.t. $\mu = \sum_{k\geq 1} \mu_k$. Then $\mathbf{N}^{(k)} := \sum_{i\geq 1}^{N_k} \delta_{X_i^{(k)}} \sim PRM(\mu_k)$ for an iid sequence $X_1^{(k)}, X_2^{(k)} \dots$ with distribution $\mathbb{P}_{X_1^{(k)}} = \mu_k \frac{1}{\mu_k(E)}$ for all $k \geq 1$ by Example 3.45 and Remark 4.2. Moreover, we can assume that the $\mathbf{N}^{(k)}$ are independent and applying Proposition 4.3 gives $\mathbf{N} := \sum_{k\geq 1} \mathbf{N}^{(k)} \sim PRM(\sum_{k\geq 1} \mu_k) = PRM(\mu)$. □

4.2. Analytical Characterization of PRM

The following is a classic result in the theory of Poisson random measures. The proof follows arguments in [15], Proposition 3.6 and [16], Theorem 5.1.

Theorem 4.5 (Analytical characterization). *A point process* $\mathbf{N}$ *with* Σ*-finite intensity* μ *is a* $PRM(\mu)$ *iff*

$$\Psi_{\mathbf{N}}(f) = \mathbb{E}\big[e^{-N(f)}\big] = exp\Big\{ - \int_E 1 - e^{-f(y)} \mathrm{d}\mu(y) \Big\} \tag{4.1}$$

for all $f \in \mathcal{B}_E^+$.

Proof. ($\Rightarrow$) Let $\mathbf{N} \sim PRM(\mu)$ and $f = u\mathbb{1}_A \in \mathcal{B}_E^+$ for $A \in \mathcal{B}_E$, $u \in [0,\infty]$. Then $N(f) = uN_A$ and $N_A \sim Poi(\mu(A))$ by assumption. Consequently,

$$\begin{aligned} \Psi_{\mathbf{N}}(f) &= \mathbb{E}[e^{-N(f)}] = \mathbb{E}[e^{-uN_A}] = \Psi_{N_A}(u) = exp\{-(1-e^{-u})\mu(A)\} \\ &= exp\Big\{ - \int_E (1 - e^f)\mathrm{d}\mu \Big\}. \end{aligned} \tag{4.2}$$

If $f = \sum_{k\geq 1}^n u_k \mathbb{1}_{A_k}$ for pairwise disjoint $A_1, \dots, A_n \in \mathcal{B}_E$ and $u_k \in [0,\infty]$, then (4.1) still holds, since

$$\begin{aligned} \Psi_{\mathbf{N}}(f) &= \mathbb{E}[e^{-N(f)}] = \mathbb{E}\Big[exp\Big\{ - \sum_{k=1}^n u_k N_{A_k} \Big\}\Big] \overset{\text{ind. incr.}}{=} \prod_{k=1}^n \mathbb{E}[e^{u_k N_{A_k}}] \\ &\overset{(4.2)}{=} \prod_{k=1}^n exp\Big\{ - \int_E 1 - e^{u_k \mathbb{1}_{A_k}} \mathrm{d}\mu \Big\} = exp\Big\{ - \int_E 1 - e^{-f}\mathrm{d}\mu \Big\}. \end{aligned} \tag{4.3}$$

Now for any $f \in \mathcal{B}_E^+$ there are simple functions f_n s.t. $f_n \nearrow f$, hence $\Psi_{\mathbf{N}}(f_n) \searrow \Psi_{\mathbf{N}}(f)$ by Proposition 3.37 and $1 - e^{f_n} \nearrow 1 - e^f$. Monotone convergence gives

$$\Psi_{\mathbf{N}}(f) = \lim_{n\to\infty} \Psi_{\mathbf{N}}(f_n) \overset{(4.3)}{=} \lim_{n\to\infty} exp\Big\{ - \int_E 1 - e^{-f_n}\mathrm{d}\mu \Big\} \overset{MCT}{=} exp\Big\{ - \int_E 1 - e^{-f}\mathrm{d}\mu \Big\},$$

what implies necessity.

($\Leftarrow$) Assume that the point process $\mathbf{N}$ satisfies (4.1). Then for $f := u\mathbb{1}_A$ for any $A \in \mathcal{B}_E$ we have $\Psi_{N_A}(u) = \mathbb{E}[e^{-uN_A}] = \mathbb{E}[e^{-N(f)}] = \Psi_{\mathbf{N}}(f) \overset{(4.1)}{=} exp\Big\{ - \int_E (1 - e^f)\mathrm{d}\mu \Big\} = exp\{-(1-e^{-u})\mu(A)\}$ and thus $N_A \sim Poi(\mu(A))$.
To prove independence of the increments pick pairwise disjoint $A_1, \dots, A_n \in \mathcal{B}_E$ and set $f := u_1\mathbb{1}_{A_1} + \dots + u_n\mathbb{1}_{A_n}$. Then

$$\begin{aligned} \Psi_{(N_{A_1},\dots,N_{A_n})}(u_1,\dots,u_n) &= \Psi_{\mathbf{N}}(f) \overset{(4.1)}{=} exp\Big\{ - \int_E 1 - e^{-f}\mathrm{d}\mu \Big\} \\ &= \prod_{k=1}^n exp\{-(1-e^{-u_k})\mu(A_k)\} = \prod_{k=1}^n \Psi_{N_{A_k}}(u_k), \end{aligned}$$

hence $N_{A_1}, \dots, N_{A_n}$ are independent by Theorem 3.46. □

Corollary 4.6. *Let* $\mathbf{N} \sim PRM(\mu)$ *with finite* μ. *Then there there is an iid sequence* (X_n) *and an independent,* $\overline{\mathbb{N}}_0$*-valued random variable* N *s.t.* $\mathbf{N} \overset{d}{=} \sum_{k=1}^N \delta_{X_k}$.

Proof. If $\mu(E) < \infty$ we can write $\mu(E) = \kappa Q$ for a probability measure Q and $\kappa \in [0,\infty)$. Let $N \sim Poi(\kappa)$ and define an iid sequence (X_n) independent from N s.t. $X_1 \overset{d}{=} Q$. Then we know that the Laplace transform of $\tilde{\mathbf{N}} := \sum_{k=1}^N \delta_{X_k}$ it is the same as the Laplace transform of $\mathbf{N}$ (see Example 3.36). Therefore $\tilde{\mathbf{N}} \overset{d}{=} \mathbf{N}$ by Theorem 3.46. □

4.3. PRM and $CPoi$

At the beginning of the preceding chapter we have seen that starting from a counting process $\mathbf{C} = (C_t)_{t\geq 0}$ with $C_0 = 0$ we can construct a point process assigning mass C_t to the intervals $[0, t]$ for all $t \geq 0$. Especially we did that for a Poisson process. In this case the resulting point process is a PRM with finite intensity as we will see subsequently.

Example 4.7 (PRM and Poisson processes). Let $\mathbf{X} = (X_t)_{t\geq 0}$ be a Poisson process on $\mathbb{R}$ (for convenience) with parameter $\kappa \geq 0$ and let $\mathbf{N}$ be a point process satisfying $N_{[0,t]}(\omega) = X_t(\omega)$ for all $t \geq 0$ (see Example 3.16). We claim that $\mathbf{N} \sim PRM\big(\kappa\lambda\big|_{[0,\infty)}\big)$. Applying Lemma 3.44, it suffices to show that $\mathbf{N}$ has independent increments on the family $\mathcal{H} := \{[s, t), 0 \leq s \leq t\}$. (It is easy to check that this is actually a determining semiring.) Moreover, we have $N_{[0,t)} = N_{[0,t]} = X_t$, since $N_{\{t\}} = X_t - X_{t-} = 0$ a.s. (a simple result for Lévy processes, see Corollary 5.7).
If we pick $0 \leq s_1 < t_1 < \cdots < s_n < t_n$ then $N_{[s_k,t_k)} = X_{t_k} - X_{s_k}$ for $k = 1, \ldots, n$ and since Poisson processes have independent increments, the random variables $(X_{t_1} - X_{s_1}), \ldots, (X_{t_n} - X_{s_n})$ are independent. Hence $\mathbf{N}$ has independent increments.
It remains to show that $N_A \sim Poi(\kappa\lambda(A))$ for all $A \in \mathcal{B}_{[0,\infty)}$. Using Theorem 3.46 it suffices to check this condition on $\mathcal{H}$, since it evidently forms a determining class for $[0, \infty)$. This is now easy, since $N_{[s,t)} = X_t - X_s \stackrel{d}{=} X_{t-s} \sim Poi(\kappa(t-s))$ by construction and the stationarity of the increments of $\mathbf{X}$.

Our next goal is to describe the connection between a more general class of Lévy processes – the so-called "Compound Poisson processes" – and PRM.

Definition 4.8. Let ν be a finite measure on $\mathcal{B}_{\mathbb{R}^d}$. A random variable Z with values in $\mathbb{R}^d$ is COMPOUND POISSON DISTRIBUTED WITH MEASURE ν, if $\varphi_Z(u) = \mathbb{E}[e^{iuZ}] = exp\big\{-\int_{\mathbb{R}^d} 1 - e^{-iuy}\mathrm{d}\nu(y)\big\}$, in symbols $Z \sim CPoi(\nu)$.

Proposition 4.9. *Let $Z \sim CPoi(\nu)$, then there is a probability measure Q and a parameter $\kappa \geq 0$ s.t. $Z \stackrel{d}{=} \tilde{Z} := \sum_{k=1}^{N} X_k$ for an iid sequence (X_k) with $\mathbb{P}_{X_1} = Q$ and $N \sim Poi(\kappa)$ for $\kappa \geq 0$, which are mutually independent.*

Proof. Before we start, let us recall Example 3.36 where we already calculated the Laplace functional of a similar point process. But here we have to be careful, since we allow Z to change sign. Thus we have to use characteristic functions.
Now to the proof. Write $\nu = \kappa Q$ for $\kappa \geq 0$ and a probability measure Q. Then

$$\begin{aligned}\varphi_{\tilde{Z}}(u) = \mathbb{E}[e^{iu\tilde{Z}}] &= \sum_{n\geq 0} \mathbb{P}[N = n]\mathbb{E}[e^{it\sum_{k=1}^{n} X_k}] \stackrel{iid}{=} \sum_{n\geq 0} \mathbb{P}[N = n]\varphi_{X_1}(u)^n \\ &= \sum_{n\geq 0} \frac{(\kappa\varphi_{X_1}(u))^n}{n!} = exp\{-(1 - \varphi_{X_1}(u))\kappa\} = exp\Big\{ - \int_{\mathbb{R}} 1 - e^{iuy}\mathrm{d}(\kappa Q)(u)\Big\}.\end{aligned}$$

Hence $Z \stackrel{d}{=} \tilde{Z}$, since they have the same characteristic function. □

Definition 4.10. A COMPOUND POISSON PROCESS WITH MEASURE ν is a Lévy process $\mathbf{Z} = (Z_t)_{t\geq 0}$ s.t. $Z_t \sim CPoi(t\nu)$ for all $t \geq 0$ and finite ν. We abbreviate that by $CPoi$ or more precisely by $CPoi(\nu)$. In abuse of notation we also use $\mathbf{Z} \sim CPoi(\nu)$ for a process $\mathbf{Z}$, to indicate that it is a Compound Poisson process with measure ν.

Proposition 4.11. *Let $\mathbf{Z} = (Z_t)_{t\geq 0}$ be a stochastic process. If $\mathbf{Z} \sim CPoi(\nu)$, then there is a probability measure Q and a parameter $\kappa \geq 0$ s.t. $\mathbf{Z} \stackrel{d}{=} \tilde{\mathbf{Z}} = (\tilde{Z}_t)_{t\geq 0} := \left(\sum_{k=1}^{X_t} Y_k\right)_{t\geq 0} \sim CPoi(\nu)$ for an iid sequence (Y_n) with $\mathbb{P}_{Y_1} = Q$ and a Poisson process $\mathbf{X} = (X_t)_{t\geq 0}$ with parameter κ, which are mutually independent. Moreover, Z_t and $\tilde{Z}_t$ have the same characteristic function $\varphi_{Z_t}(u) = \varphi_{\tilde{Z}_t}(u) = exp\left\{-t\int_{\mathbb{R}^d} 1 - e^{-iuy}\mathrm{d}\nu(y)\right\}$ for all $t \geq 0$.*

Proof. Let ν, Q be such that $\nu = \kappa Q$ for $\kappa \geq 0$ and a probability measure Q. Since $X_t \sim Poi(t\kappa)$ for all $t \geq 0$, the statement about the characteristic function is an immediate consequence of Proposition 4.9. To show that $\tilde{\mathbf{Z}} \sim CPoi(\nu)$, thus $\mathbf{Z} \stackrel{d}{=} \tilde{\mathbf{Z}}$, it remains to prove that $\tilde{\mathbf{Z}}$ is Lévy. That $\tilde{Z}_0 = 0$ a.s. follows from the fact that $X_0 = 0$ a.s. and stationarity/independence of the increments are inherited by the respective properties of $\mathbf{X}$. For example $Z_t - Z_s = \sum_{k=1}^{X_t} Y_k - \sum_{k=1}^{X_s} Y_k = \sum_{k=1}^{X_t - X_s} Y_k \stackrel{d}{=} \sum_{k=1}^{X_{t-s}} Y_k = Z_{t-s}$. That the paths are a.s. càdlàg is an easy consequence of the definition of $\tilde{\mathbf{Z}}$ and the fact that it only jumps when $\mathbf{X}$ has a discontinuity. □

The preceding proposition showed that Compound Poisson processes exist and are – roughly speaking – Poisson processes where the jumps have random height. Especially we get a Poisson process for $\mathbb{P}_{X_1} = \delta_{\{1\}}$, i.e. $X_1 \equiv 1$. To construct such a process using random measures, look at a Poisson random measure $\mathbf{N}$ on $E = [0,\infty) \times \mathbb{R}^d$. For a fixed ω the points (t, x) to which $\mathbf{N}$ assigns unit mass, then determine a path of a process. Those jumping points are hereby given by the "time-component" t, while the height of the corresponding jump is given by the "space-component" x.
As for Poisson processes there are no double-jumps in a $CPoi$ (a trivial consequence of Proposition 4.11). Consequently we want the probability that $\mathbf{N}$ assigns mass to more that one point at the same time t to be 0 as well.

Proposition 4.12. *Let $\mathbf{N} \sim PRM\big(\lambda\big|_{[0,\infty)} \otimes \nu\big)$ on $E = [0,\infty) \times \mathbb{R}^d$, where ν is a σ-finite measure. Then $\mathbb{P}[\exists t \geq 0 : N_{\{t\}\times\mathbb{R}^d} \geq 2] = 0$.*

Proof. Cf. [12], Theorem 24.13.
Since ν is σ-finite there is a sequence (E_n) s.t. $E_n \nearrow E$ and $\nu(E_n) < \infty$ for all n. Then $\mathbb{P}[\exists t \geq 0 : N_{\{t\}\times\mathbb{R}^d} \geq 2] = \lim_{n\to\infty} \mathbb{P}[\exists t \geq 0 : N_{\{t\}\times E_n} \geq 2\}]$ by continuity of measures. Hence it suffices to check that all terms on the right-hand side vanish. Moreover $\mathbb{P}[\exists t \geq 0 : N_{\{t\}\times E_n} \geq 2] = \sum_{k\geq 1} \mathbb{P}[\exists t \in [k-1, k) : N_{\{t\}\times E_n} \geq 2]$ and since we equipped the time component with Lebesgue's measure, all those probabilities coincide. Ergo it even suffices to check that $\mathbb{P}[\exists t \in [0,1) : N_{\{t\}\times E_n} \geq 2] = 0$.
Let $m \geq 0$, then

$$\begin{aligned}
1 - \mathbb{P}[\exists t \in [0,1) : N_{\{t\}\times E_n} \geq 2] &= \mathbb{P}\Big[\bigcap_{l=1}^{m} \big\{N_{[\frac{l-1}{m},\frac{l}{m})\times E_n} \leq 1\big\}\Big] \\
&\stackrel{ind.}{=} \prod_{l=1}^{m} \mathbb{P}\big[N_{[\frac{l-1}{m},\frac{l}{m})\times E_n} \leq 1\big] \\
&= \prod_{l=1}^{m} \big(\mathbb{P}\big[N_{[\frac{l-1}{m},\frac{l}{m})\times E_n} = 0\big] + \mathbb{P}\big[N_{[\frac{l-1}{m},\frac{l}{m})\times E_n} = 1\big]\big).
\end{aligned}$$

Now $N_{[\frac{l-1}{m},\frac{l}{m})\times E_n} \sim Poi\Big((\lambda\otimes\nu)\big([\frac{l-1}{m},\frac{l}{m})\times E_n\big)\Big) = Poi\big(\frac{1}{m}\nu(E_n)\big)$, so

$$\begin{aligned}1-\mathbb{P}[\exists t\in[0,1): N_{\{t\}\times E_n}\geq 2] &= \prod_{l=1}^{m}\Big(e^{-\frac{\nu(E_n)}{m}} + \frac{\nu(E_n)}{m}e^{-\frac{\nu(E_n)}{m}}\Big)\\ &= e^{-\nu(E_n)}\Big(1+\frac{\nu(E_n)}{m}\Big)^m.\end{aligned}$$

Finally the last term converges to 1 as $m\to\infty$. □

Theorem 4.13 (*PRM* and *CPoi*). *Let $E=\mathbb{R}^d$, ν be a measure on $\mathcal{B}_E$ with $\nu(\{0\}) = 0$ and $\mathbf{N}\sim PRM\big(\lambda\big|_{[0,\infty)}\otimes\nu\big)$. Then for any $B\in\mathcal{B}_{\mathbb{R}^d}$ s.t. $\nu(B)<\infty$ the process $\mathbf{Z}=(Z_t)_{t\geq 0} := \big(\int_{[0,t]\times B} y\mathrm{d}\mathbf{N}(s,y)\big)_{t\geq 0} = (N(\mathrm{Id}_E\,\mathbb{1}_{[0,t]\times B}))_{t\geq 0}$ is a Compound Poisson process with measure $\nu\big|_B$.*

Proof. We show $Z_t\sim CPoi(t\nu)$ for any $t\geq 0$ using characteristic functions

$$\begin{aligned}\varphi_{Z_t}(u) &= \varphi_{N(\mathrm{Id}_E\,\mathbb{1}_{[0,t]\times B})}(u) = \mathbb{E}\big[e^{iuN(y\mathbb{1}_{[0,t]\times B}(s,y))}\big] = \mathbb{E}\big[e^{-N(-iuy\mathbb{1}_{[0,t]\times B}(s,y))}\big]\\ &\overset{\text{Thm. 4.5}}{=} exp\Big\{-\int_{[0,\infty)\times E} 1-e^{iuy\mathbb{1}_{[0,t]\times B}(s,y)}\mathrm{d}(\lambda\otimes\nu)(s,y)\Big\}\\ &= exp\Big\{-t\int_B 1-e^{iuy}\mathrm{d}\nu(y)\Big\} = exp\Big\{-\int_E 1-e^{iuy}\mathrm{d}(t\nu\big|_B)(y)\Big\}.\end{aligned}$$

Let us now show that $\mathbf{Z}$ is Lévy. It has independent increments, since for all $0=t_0<t_1<\cdots<t_n$ the random variables

$$Z_{t_k}-Z_{t_{k-1}} = N(\mathrm{Id}_E\,\mathbb{1}_{(t_{k-1},t_k]\times B}) = N\big|_{(t_{k-1},t_k]\times B}(\mathrm{Id}_E), \qquad 1\leq k\leq n$$

are independent by Proposition 4.3 (*ii*) and (*iii*).
We pick $0\leq s<t$ and do the same calculations as before to get

$$\varphi_{Z_t-Z_s}(u) = \varphi_{N(\mathrm{Id}\,\mathbb{1}_{(s,t]\times B})}(u) = exp\Big\{-\int_B 1-e^{iuy}\mathrm{d}((t-s)\nu)(y)\Big\} = \varphi_{Z_{t-s}}(u).$$

Hence $\mathbf{Z}$ has stationary increments by Theorem 3.33.
Since $\mu(\{0\}\times B)=\lambda(\{0\})\nu(B)=0$ by the definition of the product measure, Proposition 3.21 implies that $N_{\{0\}\times B}=0$ and thus $Z_0=0$ (both a.s.). It remains to prove that the paths are a.s. càdlàg. This is now easy since $\mu([0,t]\times B) = (\lambda\otimes\nu)([0,t]\times B) = t\nu(B)<\infty$, thus $\mathbf{N}$ has a.s. at most finitely many atoms in $[0,t]\times B$. For a fixed ω denote the atoms by $(s_1,y_1),\ldots,(s_k,y_k)$, then $Z_s=\sum_{j=1}^k y_j\mathbb{1}_{\{s_j\leq s\}}$ for $s\in[0,t]$. Consequently $\mathbf{Z}$ has a.s. càdlàg paths on $[0,t]$ and since t was arbitrary, the result follows. □

4.4. *PRM* and Martingales

We end the part on *PRM* with a short glimpse at the relation between Poisson random measures and martingales.

Definition 4.14. A family of σ-algebras $(\mathcal{F}_t)_{t\geq 0}$ is called FILTRATION if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $0 \leq s \leq t$. If $\mathcal{F}_t \subseteq \mathcal{A}$ for all t, the space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called FILTERED PROBABILITY SPACE. A filtration is said to SATISFY THE USUAL CONDITIONS if it is right-continuous, i.e. $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$ for all $t \geq 0$, and complete, i.e. $A \in \mathcal{F}_0$ for all $A \in \mathcal{A}$ s.t. $\mathbb{P}[A] = 0$.
A stochastic process $\mathbf{M} := (M_t)_{t\geq 0}$ is a (CONTINUOUS-TIME) MARTINGALE W.R.T. THE FILTRATION $(\mathcal{F}_t)_{t\geq 0}$ if for all $t \geq 0$

(i) M_t is $\mathcal{F}_t$ measurable, i.e. M is ADAPTED TO $(\mathcal{F}_t)_{t\geq 0}$

(ii) $\mathbb{E}[|M_t|] < \infty$

(iii) $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ for all $0 \leq s \leq t$.

Remark 4.15. For any given filtration we always assume wlog that it satisfies the usual conditions by passing to the so-called "augmented filtration". We don't provide any more details, the curious reader is referred to [2], p. 84ff.

Definition 4.16. A measure π on $\mathcal{B}_{\mathbb{R}_0^d}$ (where $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{0\}$) (or on $B_{\overline{\mathbb{R}}_0^d}$, with $\pi(\overline{\mathbb{R}}_0^d \setminus \mathbb{R}_0^d = 0))$ is a LÉVY MEASURE if it is σ-finite and satisfies

$$\int_{\mathbb{R}_0^d} 1 \wedge |y|^2 \mathrm{d}\pi(y) < \infty. \tag{4.4}$$

Remark 4.17. (i). The reason why we avoid mass on $\{0\}$ is that we want a *PRM* to describe jumps of a stochastic process and jumps of height $\{0\}$ don't really deserve that name.

(ii). The σ-algebras $\mathcal{B}_{\overline{\mathbb{R}}_0^d}$ and $\mathcal{B}_{\overline{\mathbb{R}}^d}$ are always chosen in such a way that $\mathcal{B}_{\overline{\mathbb{R}}^d} = \mathbb{R}^d \cap \mathcal{B}_{\overline{\mathbb{R}}^d}$ and $\mathcal{B}_{\overline{\mathbb{R}}_0^d} = \mathbb{R}_0^d \cap \mathcal{B}_{\mathbb{R}_0^d}$.

(iii). Note that a set in $\overline{\mathbb{R}}_0^d$ is relatively compact iff it is bounded away from 0 ($\overline{\mathbb{R}}^d$ is compact).

Theorem 4.18 (*PRM* **and martingales).** *Let* $\mathbf{N} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \pi)$*, where* π *is a Lévy measure on* $\mathbb{R}^d$*. Fix* $B \in \mathcal{B}_{\mathbb{R}^d}$ *satisfying* $\pi(B) < \infty$*,* $\int_B |y| \mathrm{d}\pi(y) < \infty$ *and define a process* $\mathbf{Z}^{(B)} = \big(Z_t^{(B)}\big)_{t\geq 0}$ *via*

$$Z_t^{(B)} := \int_{[0,t]\times B} y \mathrm{d}\mathbf{N}(s,y) - t \int_B y \mathrm{d}\pi(y).$$

Then $\mathbf{Z}^{(B)}$ *is a martingale w.r.t. the filtration* $(\mathcal{F}_t)_{t\geq 0} := \big(\sigma(\mathbf{N}\big|_{[0,s]\times\mathbb{R}_0^d}, 0 \leq s \leq t)\big)_{t\geq 0}$*. If we assume additionally that* $\int_B |y|^2 \mathrm{d}\pi(y) < \infty$*, then* $\mathbb{E}\big[\big|Z_t^{(B)}\big|^2\big] = t\int_B |y|^2 \mathrm{d}\pi(y) < \infty$*. Hence* $\mathbf{Z}^{(B)}$ *is a* $\mathcal{L}^2$*-martingale, that is* $Z_t^{(B)} \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$ *for all* $t \geq 0$*.*

Proof. (*i*) Let us first show that $\mathbf{Z}^{(B)}$ is a martingale and fix $t \geq 0$. To prove that $\mathbf{Z}^{(B)}$ is adapted, it suffices to check that $\int_{[0,t]\times B} y \mathrm{d}\mathbf{N}(s,y) = N\big|_{[0,t]\times B}(f(s,y))$ for $f(s,y) = y$ is $\mathcal{F}_t$ measurable, since the "drift" $-t\int_B y\mathrm{d}\pi(y)$ is constant (for fixed t). To verify this, pick simple functions f_n on $[0,t] \times B$ s.t. $f_n \nearrow f$. Then $N\big|_{[0,t]\times B}(f_n(s,y))$ is (by

definition) $\mathcal{F}_t$-measurable, and so is their limit $N\big|_{[0,t]\times B}(f(s,y))$.
Now to the integrability condition

$$\begin{aligned}\mathbb{E}\big[|Z_t^{(B)}|\big] &\overset{\Delta-\text{ineq.}}{\leq} \mathbb{E}\Big[\int_{[0,t]\times B}|y|\mathrm{d}\mathbf{N}(s,y)+t\int_B|y|\mathrm{d}\pi(y)\Big]\\ &=\int_{[0,t]\times B}|y|\mathrm{d}\lambda\otimes\pi(y)+t\int_B|y|\mathrm{d}\pi(y)=2t\int_B|y|\mathrm{d}\pi(y)<\infty.\end{aligned}$$

It is easy to see that $\mathbf{Z}^{(B)}$ has independent and stationary increments, in fact it is just a *CPoi with drift* (see Theorem 4.13). Using this we conclude the first part by

$$\begin{aligned}\mathbb{E}\big[Z_t^{(B)}-Z_s^{(B)}|\mathcal{F}_s\big] &\overset{\text{ind. inc.}}{=} \mathbb{E}\big[Z_t^{(B)}-Z_s^{(B)}\big] \overset{\text{stat. inc.}}{=} \mathbb{E}\big[Z_{t-s}^{(B)}\big]\\ &=\mathbb{E}\Big[\int_{[0,t-s]\times B}y\mathrm{d}\mathbf{N}(u,y)\Big]-(t-s)\int_B y\mathrm{d}\pi(y)=0.\end{aligned}$$

(*ii*) The second part of this theorem is certainly well-known. Since I unfortunately couldn't find any references, it is proven single-handedly.
It remains to calculate $\mathbb{E}\big[\big|Z_t^{(B)}\big|^2\big]$ under the stronger assumptions. Looking at the components $Z_t^{(B)}=(Z_{t,1}^{(B)},\dots,Z_{t,d}^{(B)})$ we get $\big|Z_t^{(B)}\big|^2=\big(Z_{t,1}^{(B)}\big)^2+\dots+\big(Z_{t,d}^{(B)}\big)^2$. For $y=(y_1,\dots,y_d)$ it thus suffices to show that $\mathbb{E}\big[\big(Z_{t_j}^{(B)}\big)^2\big]=t\int_B(y_j)^2\mathrm{d}\pi(y)$. For convenience, we drop the index j.
We calculate the second moment with characteristic functions using formula (3.4).

$$\begin{aligned}\mathbb{E}\big[\big(Z_t^{(B)}\big)^2\big] &= i^{-2}\left[\frac{\mathrm{d}^2}{\mathrm{d}u^2}\varphi_{Z_t^{(B)}(u)}\right]_{u=0}\\ &=-\left[\frac{\mathrm{d}^2}{\mathrm{d}u^2}\varphi_{\int_{[0,t]\times B}y\mathrm{d}\mathbf{N}(s,y)}(u)\varphi_{t\int_B y\mathrm{d}\pi(y)}(u)\right]_{u=0}\end{aligned}\tag{4.5}$$

The last equality is true since the second term is constant for fixed t and therefore independent from the first. Now the product of those characteristic functions is given by

$$\begin{aligned}&\varphi_{\int_{[0,t]\times B}y\mathrm{d}\mathbf{N}(s,y)}(u)\varphi_{t\int_B y\mathrm{d}\pi(y)}(u)\\ &\quad\overset{\text{Thm. 4.13}}{=} exp\Big\{-t\Big(\int_B 1-e^{iuy}\mathrm{d}\pi(y)\Big)\Big\}exp\Big\{-iut\int_B y\mathrm{d}\pi(y)\Big\}\\ &\quad= exp\Big\{-t\int_B 1-e^{iuy}+iuy\mathrm{d}\pi(y)\Big\}.\end{aligned}\tag{4.6}$$

Assuming we can interchange the derivative and the integral, we get

$$\frac{\mathrm{d}}{\mathrm{d}u}exp\Big\{-t\int_B 1-e^{iuy}+iuy\mathrm{d}\pi(y)\Big\}=exp\{\dots\}\Big(-t\int_B(-iye^{iuy}+iy\mathrm{d}\pi(y)\Big),$$

and

$$\begin{aligned}&\frac{\mathrm{d}^2}{\mathrm{d}u^2}exp\Big\{-t\int_B 1-e^{iuy}+iuy\mathrm{d}\pi(y)\Big\}\\ &\quad=exp\{\dots\}\Big(-t\int_B(-iye^{iuy}+iy\mathrm{d}\pi(y)\Big)^2+exp\{\dots\}\Big(-t\int_B y^2e^{iuy}\mathrm{d}\pi(y)\Big).\end{aligned}\tag{4.7}$$

Plugging in $u = 0$ makes the first term in (4.7) vanish and furthermore $\exp\{...\} = 1$. Hence the left-hand side of (4.5) equals $\int_B y^2 \mathrm{d}\pi(y)$.

Now we verify that exchanging the integral and the derivative in the last two lines is admissible. Let us first prove that $\frac{\mathrm{d}}{\mathrm{d}u} \int_B 1 - e^{iuy} + iuy \mathrm{d}\pi(y) = \int_B \frac{\partial}{\partial u}(1 - e^{iuy} + iuy)\mathrm{d}\pi(y)$. Therefore we need to show that we can exchange limit and integral when looking at the differential quotient, i.e.

$$\begin{aligned}
&\lim_{h\to 0} \int_B \tfrac{1}{h}\big((1 - e^{iy(u+h)} + iy(u+h)) - (1 - e^{iyu} + iyu)\big)\mathrm{d}\pi(y) \\
&\qquad = \lim_{h\to 0} \int_B \tfrac{1}{h}\big(e^{iyu} - e^{iy(u+h)} + ihy\big)\mathrm{d}\pi(y) \\
&\qquad \overset{(!)}{=} \int_B \lim_{h\to 0} \tfrac{1}{h}\big(e^{iyu} - e^{iy(u+h)} + ihy\big)\mathrm{d}\pi(y) = \int_B \frac{\partial}{\partial u}(1 - e^{iuy} + iuy)\mathrm{d}\pi(y).
\end{aligned}$$

To obtain the marked equality $\overset{(!)}{=}$ we use dominated convergence. Note that

$$\begin{aligned}
\tfrac{1}{h}\big(e^{iyu} - e^{iy(u+h)} + ihy\big) = \tfrac{1}{h}\big(e^{iyu}(1 - e^{ihy}) + ihy\big) &= \tfrac{1}{h}\big(e^{iyu}(-ihy + O(h^2)) + ihy\big) \\
&\to iy(1 - e^{iuy}) \in \mathbb{C}.
\end{aligned}$$

Hence for small h the absolute value of the left-hand side is bounded by $g(y) := |iy(1 - e^{iuy})| + 1$ which is a π-integrable function on B by assumption.
Finally we need to show that exchanging derivative and integral in (4.7) poses no problem. Here it remains to prove that $\frac{\mathrm{d}}{\mathrm{d}u} \int_B iy(1 - e^{iuy})\mathrm{d}\pi(y) = \int_B \frac{\partial}{\partial u}(iy(1 - e^{iuy}))\mathrm{d}\pi(y)$. Using the same argument as above, we look at the differential quotient of the left-hand side and show that it is bounded by an integrable function:

$$\begin{aligned}
\tfrac{1}{h}\big(iy(1 - e^{iy(u+h)} - 1 + e^{iuy})\big) &= \tfrac{iy}{h}\big(iy(u+h) + O(h^2) - iyu + O(h^1)\big) \\
&= \tfrac{iy}{h}(iyh + O(h^2)) \to -y^2,
\end{aligned}$$

and for small h the absolute value of the left-hand side is thus bounded on B by $g(y) := y^2 + 1$. Moreover, g is a π-integrable function, since we assumed that $\int_B |y|^2 \mathrm{d}\pi(y) < \infty$ and $\pi(B) < \infty$. □

4.5. Vague Convergence and Vague Topology

It is now time to finally introduce a topology on the measure spaces: the vague topology. We do that by specifying a mode of convergence on $\mathcal{M}(E)$ and $\mathcal{M}_p(E)$. This will then enable us to talk about continuity of maps $S : \mathcal{D} \to \mathcal{M}(E)$ and $T : \mathcal{M}(E) \to \mathcal{D}$! The topology is "nice" in a way that it makes $\mathcal{M}(E)$ and $\mathcal{M}_p(E)$ metrizable as complete, separable spaces. Since we cannot really expect that $\mathcal{M}(E)$ has those properties if E hasn't, we additionally assume that E is an uncountable Polish space with metric d_E.

Definition 4.19. Let (ν_n) be a sequence in $\mathcal{M}(E)$. Then ν_n CONVERGES VAGUELY TO ν for some $\nu \in \mathcal{M}(E)$, in symbols $\nu_n \xrightarrow{v} \nu$, if

$$\nu_n(f) = \int_E f \mathrm{d}\nu_n \to \int_E f \mathrm{d}\nu = \nu(f)$$

for all $f \in \mathcal{C}_c^+(E) := \{f : E \to \mathbb{R} : f \text{ is continuous with compact support}\}$.

Remark 4.20 (Comparing Vague and Weak Convergence). The distribution of $\mathbf{N}$ is a measure on $(\mathcal{M}(E), \mathscr{M}(E))$. Since we equipped this space with a topology, we can compare weak convergence of random measures to vague convergence of its corresponding distributions. If $(\mathbf{N}^{(n)})$ is a sequence of random measures, then $\mathbf{N}^{(n)} \Rightarrow \mathbf{N}$ means, that $\mathbb{P}_{\mathbf{N}^{(n)}}(f) = \int_{\mathcal{M}(E)} f \mathrm{d}\mathbb{P}_{\mathbf{N}^{(n)}} \to \int_{\mathcal{M}(E)} f \mathrm{d}\mathbb{P}_{\mathbf{N}} = \mathbb{P}_{\mathbf{N}}(f)$ for all $f \in \mathcal{C}_b(\mathcal{M}(E))$. Recall that $\mathcal{C}_b(\mathcal{M}(E))$ is the set of all bounded continuous functions on $\mathcal{M}(E)$.
Hence weak convergence of random measures $\mathbf{N}^{(n)}$ and vague convergence of their distributions $\mathbb{P}_{\mathbf{N}^{(n)}}$ is *nearly* the same. The only difference is that vague convergence uses a smaller class of test functions. But this difference is essential: in the vague topology it is possible that mass "escapes" compact sets and so it may happen that vague convergence holds in situations when weak convergence fails. However, if the sequence $(\mathbf{N}^{(n)})$ is tight, hence nearly all mass stays in some compact set, those two concepts evidently coincide.

This mode of convergence induces a topology, which is defined as follows.

Definition 4.21. The VAGUE TOPOLOGY ON $\mathcal{M}(E)$ is the coarsest topology on $\mathcal{M}(E)$ s.t. the maps $I_f : \mathcal{M}(E) \to \mathbb{R}$, $\nu \mapsto \nu(f) = \int_E f \mathrm{d}\nu$ for $f \in \mathcal{C}_c^+(E)$ are continuous.

Remark 4.22 (Vague Topology). (see [4], p. 188ff and [15], p. 140, 146-147)

(i) Convergence in the vague topology is equivalent to vague convergence.

(ii) The vague topology can be embedded in an appropriate product space. Therefore look at the family I_f, $f \in \mathcal{C}_c^+(E)$ and define $I : \mathcal{M}(E) \to \prod_{f \in \mathcal{C}_c^+(E)} [0, \infty)$, $\nu \mapsto (I_f(\nu))_{f \in \mathcal{C}_c^+(E)} := (\nu(f))_{f \in \mathcal{C}_c^+(E)}$. Using this map, we can characterize vague convergence by convergence in $\prod_{f \in \mathcal{C}_c^+(E)} [0, \infty)$ equipped with the product topology, i.e. the initial topology w.r.t. all projections $\pi_f : \prod_{f \in \mathcal{C}_c^+(E)} [0, \infty) \to [0, \infty)$. This makes I continuous in a canonical way and also justifies regarding $N(f)$ as a "coordinate". Moreover, injectivity of I is a consequence of the Riesz Representation Theorem. We just state this as a fact, the curious reader is referred to Theorems 29.1/29.3 and p. 192 in [4] for more details. Using this, the map I becomes bijective on its image $I(\mathcal{M}(E))$ and so is its inverse I^{-1}. Due to the product topology, I^{-1} is also continuous and I is thus an isomorphism between $\mathcal{M}(E)$ and $I(\mathcal{M}(E))$.

(iii) The family $\{\{\nu \in \mathcal{M}(E) : \nu(f) \in (s, t)\}, f \in \mathcal{C}_c^+(E), 0 \leq s < t \leq \infty\}$ forms a subbase of the vague topology, since the system $\{\pi_f^{-1}((s, t)), f \in \mathcal{C}_c^+(E), 0 \leq s < t \leq \infty\}$ is a subbase of the product topology. Their finite intersections are base sets, while open sets are in turn just unions of base sets.

(iv) Furthermore, if a locally compact space E is Polish, then $\mathcal{M}(E)$ equipped with the vague topology has a countable basis. This fact is non-trivial and a proof can be found in [4] (Theorem 31.5). Especially we can subtract a countable subbase from the one we had in *(iii)* by looking at a suitable sequence (f_n) of functions in $\mathcal{C}_c^+(E)$ and points (t_n) in E.

At this point it is natural to ask about the relation between the induced Borel σ-algebra $\mathcal{B}_{\mathcal{M}(E)}$ and the σ-algebra $\mathscr{M}(E)$.

Proposition 4.23. *$\mathcal{B}_{\mathcal{M}(E)} = \mathscr{M}(E)$ and $\mathcal{B}_{\mathcal{M}_p(E)} = \mathscr{M}_p(E)$.*

Proof. We start with the first statement.
($\subseteq$) By the preceding remark (especially point (iii)) it suffices to check that subbase elements belong to $\mathscr{M}(E)$. We have seen that those are of the form $\{\nu \in \mathcal{M}(E) : \nu(f) \in (s,t)\} = I_f^{-1}((s,t))$ for countably many $f \in \mathcal{C}_c^+(E)$, $0 \le s < t \le \infty$ and I_f defined as above. The claim follows from the identity $\mathscr{M}(E) = \sigma(m_A, A \in \mathcal{B}_E) = \sigma(I_f, f \in \mathcal{C}_c^+(E))$ which one can easily verify.

($\supseteq$) Since $\mathscr{M}(E) = \sigma(I_f, f \in \mathcal{C}_c^+(E))$ it remains to check that $I_f^{-1}(B) \in \mathcal{B}_{\mathcal{M}(E)}$ for all open $B \in \mathcal{B}_{[0,\infty]}$. We can write B as a countable union of open intervals and thus $I_f^{-1}(B)$ as a countable union of open sets in $\mathcal{M}(E)$ due to the existence of a countable base. Hence $I_f^{-1}(B) \in \mathcal{B}_{\mathcal{M}(E)}$.

The second statement is an immediate consequence of the first since we have $\mathcal{B}_{\mathcal{M}_p(E)} = \mathcal{B}_{\mathcal{M}(E)}\big|_{\mathcal{M}_p(E)} = \mathscr{M}(E)\big|_{\mathcal{M}_p(E)} = \mathscr{M}_p(E)$. $\square$

A variant of Urysohn's Lemma (cf. Lemma C.5) which guarantees that we can approximate compact or open and relatively compact sets by continuous functions having compact support, we can give a characterization of vague convergence, similar to Portmanteaus Theorem for weak convergence.

Theorem 4.24 (Characterizing vague convergence). *Let $\nu, \nu_1, \nu_2, \dots$ in $\mathcal{M}(E)$. Then the following are equivalent*

(i) $\nu_n \xrightarrow{v} \nu$

(ii) $\limsup_{n\to\infty} \nu_n(K) \le \nu(K)$ *and* $\liminf_{n\to\infty} \nu_n(A) \ge \nu(A)$ *for all compact K and relatively compact, open A in $\mathcal{B}_E$*

(iii) $\nu_n(A) \to \nu(A)$ *for all relatively compact $A \in \mathcal{B}_E$ s.t. $\nu(\partial A) = 0$.*

Proof. $(i) \Rightarrow (ii)$ Let K be a compact set. Then there is a sequence (f_n) in $\mathcal{C}_c^+(E)$ and compact sets (K_n) in E s.t. $K_n \searrow K$ and $\mathbb{1}_K \le f_n \le \mathbb{1}_{K_n}$ by Lemma C.5. Now $\nu_n(f_m) \to \nu(f_m)$ for all $m \ge 1$ by assumption and thus $\limsup_{n\to\infty} \nu_n(K) \le \limsup_{n\to\infty} \nu_n(f_m) = \nu(f_m)$ for all $m \ge 1$. But $\lim_{m\to\infty} \nu(f_m) = \lim_{m\to\infty} \int_E f_m \mathrm{d}\nu \overset{DCT}{=} \int_E \mathbb{1}_K \mathrm{d}\nu = \nu(K)$. Here the Dominated Convergence Theorem applied since $\nu(f_m) \le \nu(\mathbb{1}_{K_1}) < \infty$ for all m. The remaining result on open and relatively compact sets follows from the same argument using the second part of Lemma C.5.

$(ii) \Rightarrow (iii)$ Pick a relatively compact set $A \in \mathcal{B}_E$ s.t. $\nu(\partial A) = 0$. The sets $\overline{A}$ and A^o satisfy the requirements of (ii) and get the same mass under ν. Therefore

$$\nu(\overline{A}) = \nu(A^o) \overset{(ii)}{\le} \liminf_{n\to\infty} \nu_n(A^o) \le \liminf_{n\to\infty} \nu_n(A) \le \limsup_{n\to\infty} \nu_n(A) \le \limsup_{n\to\infty} \nu_n(\overline{A}) \overset{(ii)}{\le} \nu(\overline{A}).$$

Ergo the inequalities turn into equalities and $\liminf_{n\to\infty} \nu_n(A) = \limsup_{n\to\infty} \nu_n(A) = \nu(A)$.

$(iii) \Rightarrow (i)$ Let $f \in \mathcal{C}_c^+(E)$ have compact support $K \subseteq E$ and show $\nu_n(f) \to \nu(f)$.

Therefore we use a "sandwich argument" and try to find suitable upper and lower bounds of $\nu(f)$ in terms of simple functions (i.e. with finite range). Later on we want to apply (iii) to those simple functions, so we need to assure that the boundaries of the "steps" don't contain any atoms. Since f vanishes outside K we can equivalently look at $\tilde{f} := f\big|_K$. The set of atoms of $\nu \circ \tilde{f}^{-1}$ (apart from 0) is given by $\mathcal{S} := \{x > 0 : (\nu \circ \tilde{f}^{-1})(\{x\}) > 0\}$ and is at most countable, since $\mathcal{S} = \bigcup_{n\geq 1} \mathcal{S}_n$, where $\mathcal{S}_n := \{x > 0 : (\nu \circ \tilde{f}^{-1})(\{x\}) > \frac{1}{n}\}$ is a finite set for all n.
For any given $\varepsilon > 0$ we can intersect the range $0 = c_0 < c_1 < \cdots < c_k = \|\tilde{f}\|_\infty < \infty$, s.t. $c_j \in \mathcal{S}^c$ and $|c_{j-1} - c_j| \leq \varepsilon$ for all $1 \leq j \leq k$. Now we have a "sandwich"

$$\sum_{j=1}^k c_{j-1}\mathbb{1}_{(c_{j-1},c_j]}(\tilde{f}(y)) \leq \tilde{f}(y) \leq \sum_{j=1}^k c_j \mathbb{1}_{(c_{j-1},c_j]}(\tilde{f}(x)) \text{ and} \tag{4.8}$$

$$\sum_{j=1}^k c_{j-1}(\nu \circ \tilde{f}^{-1})((c_{j-1}, c_j]) \leq \nu(\tilde{f}(y)) \leq \sum_{j=1}^k c_j(\nu \circ \tilde{f}^{-1})((c_{j-1}, c_j]). \tag{4.9}$$

The choice of the $c_j \in \mathcal{S}^c$ and Proposition C.1 ensure that

$$\begin{aligned}\nu(\partial(\tilde{f}^{-1}((c_{j-1}, c_j])) &\leq \nu(\tilde{f}^{-1}(\partial(c_{j-1}, c_j])) \\ &= (\nu \circ \tilde{f})^{-1}(\{c_{j-1}\}) + (\nu \circ \tilde{f})^{-1}(\{c_j\}) = 0\end{aligned}$$

The difference between the left- and the right-hand side of (4.8) is $\leq \varepsilon$ for all $y \in K$ by construction and consequently the difference of the respective terms in (4.9) is $\leq \varepsilon\nu(K)$. Therefore

$$\begin{aligned}\limsup_{n\to\infty} \nu_n(f) &= \limsup_{n\to\infty} \int_K f \mathrm{d}\nu_n \overset{(4.8)}{\leq} \limsup_{n\to\infty} \int_K \sum_{j=1}^k c_j \mathbb{1}_{(c_{j-1},c_j]}(f(\tilde{y}))\mathrm{d}\nu_n(y) \\ &= \limsup_{n\to\infty} \sum_{j=1}^k c_j(\nu_n \circ \tilde{f}^{-1})((c_{j-1}, c_j]) \overset{(iii)}{=} \sum_{j=1}^k c_j(\nu \circ \tilde{f}^{-1})((c_{j-1}, c_j]) \\ &\leq \nu(f) + \varepsilon\nu(K).\end{aligned}$$

Using the lower bounds in (4.8) and (4.9), we get $\liminf_{n\to\infty} \nu_n(f) \geq \nu(f) - \varepsilon\nu(K)$ and hence $\nu_n(f) \to \nu(f)$ as desired. □

This allows us to proof a nice geometric interpretation of convergence of point processes.

Proposition 4.25 (Vague convergence of point processes). *Let $\nu, \nu_1, \nu_2, \dots$ in $\mathcal{M}_p(E)$, K compact s.t. $\nu(\partial K) = 0$ and $\nu_n \overset{v}{\to} \nu$. Then there is some $n_0(K) < \infty$ s.t. the measures ν_n and ν have the same number $m \in \mathbb{N}_0$ of (not necessarily distinct) points getting unit-mass in K for $n \geq n_0 := n_0(K)$. If we denote them by $x_1^{(n)}, \dots, x_m^{(n)} \in E$ and $x_1, \dots, x_m \in E$ respectively, then*

$$\nu_n\big|_K = \sum_{k=1}^m \delta_{x_k^{(n)}} \text{ and } \nu\big|_K = \sum_{k=1}^m \delta_{x_k}. \tag{4.10}$$

If the points are labeled properly, then $x_k^{(n)} \to x_k$ for all $1 \leq k \leq m$.

Proof. Since K is a compact set s.t. $\nu(\partial K) = 0$ we know that $\nu(K) < \infty$, $\nu_n(K) < \infty$ for all n and $\nu_n(K) \to \nu(K) =: m \in \mathbb{N}_0$ by Theorem 4.24. Ergo there is some $n_0 := n_0(K)$ s.t. the sequence is constant for all $n \geq n_0$ and the first statement follows. Using Remark 3.13 (iv) write $\nu\big|_K = \sum_{k=1}^{l} c_k \delta_{y_k}$ for some $c_1, \dots, c_l$ and *distinct* $y_1, \dots, y_l$ in K^o, $l \leq m$. There are only finitely many atoms, so for any neighbourhoods $U_1, \dots, U_l$ in K of the points $y_1, \dots, y_l$ there are pairwise disjoint neighbourhoods $\tilde{U}_1, \dots, \tilde{U}_l$ s.t. $y_k \in \tilde{U}_k \subseteq U_k$ and $\nu(\partial \tilde{U}_k) = 0$ (E is metrizable). The sets $\tilde{U}_k$ are evidently relatively compact, so we can apply Theorem 4.24 once more to conclude that $\mu_n(\tilde{U}_k) \to \mu(\tilde{U}_k)$ for all $1 \leq k \leq l$. But all those terms are again integer valued, hence $\mu_n(\tilde{U}_k) = \mu(\tilde{U}_k)$ for all k and $n \geq n_0$ (wlog).
This implies that the atoms of μ_n (if labeled properly) converge to the corresponding atoms of μ. □

The goal of the remaining part of this section is to make the measure spaces $\mathcal{M}(E)$ and $\mathcal{M}_p(E)$ Polish. The following result is very useful for this purpose, since a closed subspace of a completely metrizable and separable space inherits all those properties (cf. Proposition C.4). Said otherwise, due to the following it suffices to only show that $\mathcal{M}(E)$ is Polish.

Proposition 4.26. *$\mathcal{M}_p(E)$ is vaguely closed in $\mathcal{M}(E)$.*

Proof. Cf. [15], Proposition 3.14.
(i) The idea is to pick a sequence (ν_n) of point measures s.t. $\nu_n \xrightarrow{v} \nu$ for $\nu \in \mathcal{M}(E)$ and use the good sets principle to prove that ν is integer valued on all measurable sets. Therefore we need a decent determining class $\mathcal{E}$ of B_E. Recalling Theorem 4.24, the family $\mathcal{E} := \{A \in \mathcal{B}_E : A \text{ is relatively compact and } \nu(\partial A) = 0\}$ is certainly a good candidate. In fact it is already a π-system, since $\partial(A_1 \cap A_2) \subseteq \partial A_1 \cup \partial A_2$. To find an exhausting sequence in $\mathcal{E}$, note that E is locally compact and second-countable, so there is a countable base of relatively compact sets (Proposition C.2). Especially we can find relatively compact sets E_n, $n \geq 1$ in $\mathscr{M}(E)$ s.t. $E_n \nearrow E$. Unfortunately $E_n \in \mathcal{E}$ is not evident a priori, so we better assure that $\nu(\partial E_n) = 0$. Fix n and note that the boundary of E_n is the same as the one of the compact set $\overline{E_n}$. If we can show that there is a set $\tilde{E}_n$ in the dwelling $\overline{E}_n^{\delta}$ (which is compact if $\delta > 0$ is small enough, see Lemma C.5) s.t. $\nu(\partial \tilde{E}_n) = 0$, we can replace E by $\tilde{E}$ and assume wlog that the boundaries of E_n are ν-null sets. We proceed as in the proof of Theorem 4.24 and set $\mathcal{S}_n := \{\eta \in [0, \delta] : \nu(\partial E_n^{\eta}) > \frac{1}{n}\}$. Again, since $\nu\big(\overline{E_n}^{\delta}\big)$ is finite and the boundaries of E_n^{η} are pairwise disjoint for distinct values of η, the set $\mathcal{S}_n$ is finite for all n. Consequently the union $\mathcal{S} := \{\eta \in [0, \delta] : \nu(\partial E_n^{\eta}) > 0\}$ is at most countable and hence there is some $\tilde{\eta} \notin \mathcal{S}$ s.t. $\nu(\partial E_n^{\tilde{\eta}}) = 0$.
It remains to check that $\mathcal{E}$ generates $\mathcal{B}_E$. This is now evident, since the preceding argument even shows, that for any compact set K there are sets K_n, $n \geq 1$ in $\mathcal{E}$ s.t. $K_n \searrow K$, thus $\sigma(\mathcal{E}) \supseteq \sigma(\text{compact sets}) = \mathcal{B}_E$.

(ii) Let us collect the good sets and fix some n. We claim that the family $\mathcal{G}_n := \{G \in \mathcal{B}_E : \nu(G \cap E_n) \in \mathbb{N}_0\}$ contains the determining class $\mathcal{E}$. Indeed, Theorem 4.24 implies that $\nu_n(A) \to \nu(A) \in \mathbb{N}_0$ for any set $A \in \mathcal{E}$ because $\nu_n(\partial(A)) = 0$, $\nu(A) < \infty$ and $\nu_n(A) \in \mathbb{N}_0$ for all n by assumption. Moreover, $\mathcal{E}$ is a π-system, so $G \cap E_n \in \mathcal{E}$ and

hence $\mathcal{E} \in \mathcal{G}_n$ for all n.
To apply Dynkin's λ-π Theorem, verify that $\mathcal{G}_n$ is a λ-system. That $E \in \mathcal{G}_n$ follows from the fact that $E_n = E \cap E_n \in \mathcal{E}$. If $G \in \mathcal{G}_n$, then $\nu(G^c \cap E_n) = \nu(E \cap E_n) - \nu(G \cap E_n) \in \mathbb{N}_0$ and since the sum of integers is again an integer, also $\nu\Big(E_n \cap \biguplus_{n\geq 1} G_n\Big) = \nu\Big(\biguplus_{n\geq 1}(G_n \cap E_n)\Big) = \sum_{n\geq 1} \nu(G_n \cap E_n) \in \mathbb{N}_0$ for pairwise disjoint $G_1, G_2, \dots$ in $\mathcal{G}$. Note that the sum cannot diverge, since it is bounded by $\nu(E_n) < \infty$. Altogether $\nu\big|_{E_n}$ is a point measure for all n, $E_n \nearrow E$ and hence $\nu \in \mathcal{M}_p(E)$. $\square$

Proposition 4.27 (Characterizing vague relative compactness). *Let $M \subseteq \mathcal{M}(E)$ or $M \subseteq \mathcal{M}_p(E)$ be a set of measures. Then the following are equivalent*

(i) *M is vaguely relatively compact*

(ii) $\sup\limits_{\nu\in M} \nu(f) < \infty$ *for all $f \in \mathcal{C}_c^+(E)$*

(iii) $\sup\limits_{\nu\in M} \nu(A) < \infty$ *for all $A \in \mathcal{B}_E$ relatively compact.*

Proof. It suffices to show the result for random measures, since $\mathcal{M}_p(E)$ is vaguely closed in $\mathcal{M}(E)$.
$(i) \Rightarrow (ii)$ For a fixed $f \in \mathcal{C}_c^+(E)$ show (i) by proving that $\{\nu(f) : \nu \in \overline{M}\}$ is compact in $[0,\infty)$. Since the map $I_f : \mathcal{M}(E) \to [0,\infty)$, $\nu \mapsto \nu(f)$ is continuous for each $f \in \mathcal{C}_c^+(E)$, we obtain $\sup_{\nu\in M} \nu(f) = \sup_{\nu\in M} I_f(\nu) = \sup_{\nu\in\overline{M}} I_f(\nu) = \sup_{\nu\in\overline{M}} \nu(f)$. Now $\overline{M}$ is compact and so is its image under I_f, i.e. $\{I_f(\nu), \nu \in \overline{M}\} = \{\nu(f), \nu \in \overline{M}\}$ is compact.

$(ii) \Rightarrow (i)$ We have to show that $\overline{M}$ is compact for any relatively compact set $M \subseteq \mathcal{M}(E)$. Therefore we use the map $I : \overline{M} \to R$, $\nu \mapsto (\nu(f))_{f\in\mathcal{C}_c^+(E)}$. Here $R := \prod_{f\in\mathcal{C}_c^+(E)}[0, \sup_{\nu\in M} \nu(f)] = \prod_{f\in\mathcal{C}_c^+(E)} I_f(\overline{M})$ is a compact subset of $\prod_{f\in\mathcal{C}_c^+(E)}[0,\infty)$ by Tychonoff's Theorem (see Theorem C.3), since $I_f(\overline{M}) := [0, \sup_{\nu\in M} \nu(f)]$ is compact in $[0,\infty)$ for each $f \in \mathcal{C}_c^+(E)$ by assumption. Recalling the arguments in Remark 4.22, the spaces $\overline{M}$ and R are homeomorphic via I and consequently $\overline{M}$ is compact since $R = I(\overline{M})$ is compact.

$(iii) \Rightarrow (ii)$ This is trivial, since we can approximate any $f \in \mathcal{C}_c^+(E)$ with simple functions.

$(ii) \Rightarrow (iii)$ We only sketch the proof of the second result. Pick any relatively compact set $A \in \mathcal{B}_E$, then (ii) is satisfied for $\overline{A}$, since we can approximate $\mathbb{1}_{\overline{A}}$ with functions (f_n) in $\mathcal{C}_c^+(E)$ due to Lemma C.5. It then remains to control what happens on ∂A. The boundary is compact, so it contains at most finitely many atoms $S := \{x_1, \dots, x_n\}$. We are now going to alter the sequence (f_n) a bit. We have to assure that it remains in $\mathcal{C}_c^+(E)$ and vanishes on S to guarantee that $\nu(f_n) \to \nu(A) = \nu(\overline{A}) - \nu(S)$. This can be done by eliminating values $\neq 0$ on S in a "smooth" way. For a fixed k let $\tilde{f}_n$ equal f_n everywhere but on the ball of radius $\frac{1}{n}$ around x_k, define $\tilde{f}_n(x_k) = 0$, connect them linearly and use a mollifier to smoothen the edges. Then each $\tilde{f}_n \in \mathcal{C}_c^+(E)$ has value 0 on S as needed and still $\tilde{f}_n \searrow \mathbb{1}_K$. $\square$

Definition 4.28. A family $\mathcal{H}$ of subsets of E s.t.

(i) $\mathcal{H}$ is a base of the topology E,

(ii) $\mathcal{H}$ consists of relatively compact sets and

(iii) $\mathcal{H}$ is closed under intersections and unions, i.e. $H_1, H_2 \in \mathcal{H}$ implies $H_1 \cap H_2 \in \mathcal{H}$ and $H_1 \cup H_2 \in \mathcal{H}$,

is called $\cap\cup$-BASE OF E. In Euclidean spaces finite unions of open rectangles with rational vertices are an example of a countable $\cap\cup$-base.

Remark 4.29. Any (countable) base of relatively compact sets of E can be extended to a (countable) $\cap\cup$-base by adding all finite unions of finite intersections of base sets. Especially this shows that a countable $\cap\cup$-base exists for E. In particular, $\mathcal{H}$ is a π-system and $\sigma(\mathcal{H}) = \mathcal{B}_E$.

Theorem 4.30 ($\mathcal{M}(E)$ and $\mathcal{M}_p(E)$ are Polish). *The spaces $\mathcal{M}(E)$ and $\mathcal{M}_p(E)$ endowed with the vague topology are Polish.*

Proof. It suffices to show the result for random measures, since the set of point measures is a closed subspace and inherits all properties from $\mathcal{M}(E)$ (cf. Proposition 4.26 and Proposition C.4).
Recall that the space $\mathcal{M}(E)$ is isomorphic to $I(\mathcal{M}(E))$, a subspace of the product space $\prod_{f\in\mathcal{C}_c^+(E)}[0,\infty)$. Consequently it suffices to show that the product is Polish but unfortunately this product is uncountable. As a workaround, we can try to find countably many components (h_k) in $\mathcal{C}_c^+(E)$ s.t. vague convergence $\nu_n \xrightarrow{v} \nu$ is characterized by $\nu_n(h_k) \to \nu(h_k)$ for all k.

(i) To achieve this, pick a $\cap\cup$-base $\mathcal{H}$ of E. Lemma C.5 shows that there are sequences of functions $(f_{j,i}), (g_{j,i})$, $j, i \geq 1$ in $\mathcal{C}_c^+(E)$ s.t. $f_{j,i} \searrow \mathbb{1}_{\overline{H_i}}$ and $g_{j,i} \nearrow \mathbb{1}_{H_i}$ as $j \to \infty$. Let $\{h_1, h_2, \dots\} := \{f_{j,i}, g_{j,i}, j, i \geq 1\}$. Note that any $\nu \in \mathcal{M}(E)$ is determined by the values $\{\nu(h_k), k \geq 1\}$, since for any $H_i \in \mathcal{H}$ we have $\nu(H_i) = \lim_{j\to\infty} \nu(g_{j,i})$ and thus ν is determined on $\mathcal{H}$ and moreover on its generated σ-algebra $\mathcal{B}_{\mathcal{M}(E)} \overset{\text{Prop. 4.23}}{=} \mathscr{M}(E)$.

(ii) Next we claim that $\nu_n \xrightarrow{v} \nu$ for some $\nu \in \mathcal{M}(E)$ iff

$$\text{for each } k, \text{ there is } c_k < \infty \text{ s.t. } \nu_n(h_k) \to c_k \text{ and in this case } c_k = \nu(h_k). \tag{4.11}$$

$(\Rightarrow)$ Evident.

$(\Leftarrow)$ First we show that $\{\nu_n, n \geq 1\}$ is relatively compact. Therefore fix $f \in \mathcal{C}_c^+(E)$, let $K := supp(f)$ and use Proposition 4.27. The set K is compact so it can be covered by finitely many sets from the $\cap\cup$-base $\mathcal{H}$. Moreover, the union of those sets is again an element of $\mathcal{H}$, thus there is H_{i_0}, s.t. $K \subseteq H_{i_0}$. Now

$$f \leq \|f\|_\infty \mathbb{1}_K \leq \|f\|_\infty \mathbb{1}_{G_{i_0}} \leq \|f\|_\infty \mathbb{1}_{\overline{G_{i_0}}} \leq \|f\|_\infty f_{j,i_0},$$

thus $\sup_{n\geq 1} \nu_n(f) \leq \sup_{n\geq 1} \|f\|_\infty \nu_n(f_{j,i_0}) < \infty$, since for all k the sequence $(\nu_n(h_k))$ converges as $n \to \infty$ by (4.11).
The obtained (vague) relative compactness shows that there is a convergent subsequence (l_n) s.t. $\nu_{l_n} \xrightarrow{v} \nu$ for some $\nu \in \mathcal{M}(E)$. Property (4.11) implies that $\nu_{l_n}(h_k) \to$

$c_k = \nu(h_k)$, so all limit points coincide and hence $\nu_n(h_k) \to \nu(h_k)$ for all k. Since this determines the measure ν we have $\nu_n \xrightarrow{v} \nu$.

(iii) The *countable* product space $\prod_{k\geq 1} I_{h_k}(\overline{M})$ is a closed subspace of $\prod_{k\geq 1}[0,\infty)$. Both spaces are Polish by Proposition C.4 and the fact that $[0,\infty)$ is Polish. Consequently the vague topology is Polish. □

4.6. Weak Convergence of Random Measures

Although no result of this section is needed as a preparation for the proof of the main theorem, it provides some useful conditions for weak convergence of random measures to hold. An essential statement in this context is the famous Kallenberg's Theorem (the curious reader is referred to Theorem 4.39) providing a "workable" condition for weak convergence in $\mathcal{M}_p(E)$.
For the proof we use a tightness approach leading to subsequence arguments due to Prohorov's Theorem and the fact that $\mathcal{M}(E)$ is Polish (cf. Theorem B.5 and Theorem 4.30). Preceding this procedure we prove that – as in the case of random vectors (see Theorem 3.33) – weak convergence of random measures is the same as convergence of the respective Laplace functionals using tightness arguments. Proofs in this section are taken from [15], Propositions 3.19–3.23.

For convenience we recall some basic results of the theory of weak convergence.

Definition 4.31. Let (S,d) be a metric space. A sequence of probability measures (Q_n) on $\mathcal{B}_S$ is said to CONVERGE WEAKLY TO Q for some probability measure Q on $\mathcal{B}_S$, in symbols $Q_n \Rightarrow Q$, if

$$Q_n(f) = \int_S f \mathrm{d}Q_n \to \int_S f \mathrm{d}Q = Q(f) \text{ for all } f \in \mathcal{C}_b(S), \tag{4.12}$$

where $\mathcal{C}_b(S)$ denotes the set of all bounded, continuous functions on S.
In particular, we say that a sequence of random elements (X_n) in S CONVERGES WEAKLY (IN DISTRIBUTION) TO SOME RANDOM ELEMENT X in S if $\mathbb{P}_{X_n} \Rightarrow \mathbb{P}_X$. Note that (4.12) then becomes $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

In fact the class of test functions $\mathcal{C}_b(S)$ is a bit bigger than necessary, some equivalent definitions of weak convergence are given by the following famous result.

Theorem 4.32 (Portmanteau Theorem). *Let $X, X_1, X_2, \ldots$ some random elements on a metric space (S,d). Then the following are equivalent*

(i) $X_n \Rightarrow X$

(ii) $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded Lipschitz-continuous functions f on S*

(iii) $\liminf_{n\to\infty} \mathbb{P}[X_n \in F] \geq \mathbb{P}[X \in F]$ *for all closed $F \subseteq S$*

(iv) $\limsup_{n\to\infty} \mathbb{P}[X_n \in G] \leq \mathbb{P}[X \in G]$ *for all open $G \subseteq S$*

(v) $\lim_{n\to\infty} \mathbb{P}[X_n \in A] \to \mathbb{P}[X \in A]$ *for all $A \in \mathcal{B}_S$ satisfying $\mathbb{P}[X \in \partial A] = 0$.*

Proof. We don't prove this well-known result here, for an excellent reference see [12], Theorem 13.16. □

Remark 4.33. Compare this result to Theorem 4.24!

Remark 4.34. A set M in the Polish space $\mathcal{M}(E)$ is tight iff it is relatively compact iff sequences in M contain convergent subsequences. This is most certainly the case if M is a convergent sequence.

A more interesting result about tightness of a sequence of random measures is the following.

Lemma 4.35 (Characterizing tightness). *Let $(\mathbf{N}^{(n)})$ be a sequence of random measures. Then the following are equivalent*

(i) *the family $\{\mathbf{N}^{(n)}, n \geq 0\}$ is tight in $\mathcal{M}(E)$*

(ii) *the family $\{N^{(n)}(f), n \geq 0\}$ is tight in $\mathbb{R}$ for all $f \in \mathcal{C}_c^+(E)$, i.e.*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\big[N^{(n)}(f) > t\big] = 0$$

(iii) *the family $\{N_A^{(n)}, n \geq 0\}$ is tight in $\mathbb{R}$ for all relatively compact $A \in \mathcal{B}_E$, i.e.*

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\big[N_A^{(n)} > t\big] = 0.$$

Proof. *(i)* ⇒ *(ii)* Let $\{\mathbf{N}^{(n)}, n \geq 1\}$ be a tight family, $\varepsilon > 0$ and fix $f \in \mathcal{C}_c^+(E)$. Then there is a compact set $\mathcal{K} := \mathcal{K}(\varepsilon) \in \mathscr{M}(E)$ s.t. $\mathbb{P}[\mathbf{N}^{(n)} \in \mathcal{K}] > 1 - \varepsilon$ for all n. We have to find a compact $K \subseteq \mathbb{R}$ s.t. $\mathbb{P}[N^{(n)}(f) \in K] > 1 - \varepsilon$ for all n. Since $\mathcal{K}$ is compact in the vague topology, it is relatively compact and hence $M := \sup_{\nu\in\mathcal{K}} \nu(f) < \infty$ by Proposition 4.27, thus $K := [0, M]$ is compact. To verify that K really does the job, note that $I_f^{-1}(K) \supseteq \mathcal{K}$, so we obtain

$$\mathbb{P}[N^{(n)}(f) \in K] = \mathbb{P}[I_f(\mathbf{N}^{(n)}) \in K] = \mathbb{P}[\mathbf{N}^{(n)} \in I_f^{-1}(K)] \geq \mathbb{P}[\mathbf{N}^{(n)} \in \mathcal{K}] > 1 - \varepsilon.$$

(ii) ⇒ *(i)* For any fixed $\varepsilon > 0$ we have to find a compact set $\mathcal{K} \in \mathcal{M}(E)$ s.t. $\mathbb{P}[\mathbf{N}^{(n)} \notin \mathcal{K}] \leq \varepsilon$ for all n. We cannot take the intersection of all $I_f^{-1}(K_f)$ since there are uncountably many and compactness or even measurability could get lost. But we can try to control $(\mathbf{N}^{(n)})$ along a well chosen sequence of functions. Therefore we pick $(f_k) \in \mathcal{C}_c^+(E)$ s.t. $f_k \nearrow \mathbb{1}_E$. For each k the family $\{N^{(n)}(f_k), n \geq 1\}$ is tight by assumption. Hence there are compact sets K_k s.t. $\mathbb{P}[N^{(n)}(f_k) \notin K_k] < \frac{\varepsilon}{2^{k+1}}$ for all n. In particular, there are $c_k \geq 0$ s.t. $\mathbb{P}[N^{(n)}(f_k) > c_k] < \frac{\varepsilon}{2^{k+1}}$. We claim that $\mathcal{K} := \overline{M}$ does the job, where $M := \bigcap_{k\geq1}\{\nu \in \mathcal{M}(E) : \nu(f_k) \leq c_k\}$. Let us check that $\mathcal{K}$ is indeed compact, i.e. M is relatively compact. Therefore it suffices to show $\sup_{\nu\in\mathcal{M}} \nu(f) < \infty$ for all $f \in \mathcal{C}_c^+(E)$ by Proposition 4.27. We can dominate every $f \in \mathcal{C}_c^+(E)$ by $\|f\|_\infty f_{k_0}$ for some k_0 large enough. Consequently it suffices to show $\sup_{\nu\in M} \nu(f_k) < \infty$ for all n. But this is evident, since $\sup_{\nu\in M} \nu(f_k) \leq c_k$ for all k by definition of M. Now

$$\begin{aligned}\mathbb{P}[\mathbf{N}^{(n)} \notin \mathcal{K}] &\leq \mathbb{P}[\mathbf{N}^{(n)} \notin M] = \mathbb{P}\Big[\bigcup_{k\geq1}\{N^{(n)}(f_k) > c_k\}\Big] \leq \sum_{k\geq1} \mathbb{P}[N^{(n)}(f_k) > c_k] \\ &\leq \sum_{k\geq1} \frac{\varepsilon}{2^{k+1}} = \varepsilon,\end{aligned}$$

and hence $\{\mathbf{N}^{(n)}, n \geq 1\}$ is tight.
The remaining equivalent between (i) and (iii) can be established in the exact same way, using the other characterization of relative compactness in Proposition 4.27. $\square$

Our first important result is that weak convergence of random measures (or of their distributions) can be characterized by convergence of the corresponding Laplace functionals. Compare the following to Theorem 3.33 (v).

Proposition 4.36 (Weak convergence and Laplace functionals). *Let* $(\mathbf{N}^{(n)})$ *be a sequence of random measures. Then* $\mathbf{N}^{(n)} \Rightarrow \mathbf{N}$ *for some random measure* $\mathbf{N}$ *iff* $\Psi_{\mathbf{N}^{(n)}}(f) \to \Psi_{\mathbf{N}}(f)$ *for all* $f \in \mathcal{C}_c^+(E)$.

Proof. $(\Rightarrow)$ This direction is simple, since the Continuous Mapping Theorem and the continuity of I_f imply that $N^{(n)}(f) = I_f(\mathbf{N}^{(n)}) \Rightarrow I_f(\mathbf{N}) = N(f)$, so for any fixed $f \in \mathcal{C}_c^+(E)$

$$\lim_{n\to\infty} \Psi_{\mathbf{N}^{(n)}}(f) = \lim_{n\to\infty} \mathbb{E}\big[e^{-N^{(n)}(f)}\big] = \mathbb{E}\big[e^{-N(f)}\big] = \Psi_{\mathbf{N}}(f).$$

$(\Leftarrow)$ Replacing a fixed $f \in \mathcal{C}_c^+(E)$ by $tf \in \mathcal{C}_c^+(E)$ in the above line for any $t \geq 0$ implies convergence of the Laplace transforms of $N^{(n)}(f)$ to the Laplace transform of $N(f)$ by the linearity of the integral. Ergo we have $N^{(n)}(f) \Rightarrow N(f)$ for all $f \in \mathcal{C}_c^+(E)$ by Theorem 3.33 (v). Since weakly convergent sequences are tight (see Remark 4.34), so is $\{N^{(n)}(f), n \geq 0\}$ for all $f \in \mathcal{C}_c^+(E)$ and we can apply Lemma 4.35 to get tightness of $\{\mathbf{N}^{(n)}, n \geq 0\}$. By Prohorov's Theorem we know that any subsequence (k_n) contains a further subsequence $(l_n) := (k_{m_n})$ s.t. $\mathbf{N}^{(l_n)} \Rightarrow \tilde{\mathbf{N}}$ for some random measure $\tilde{\mathbf{N}}$. It remains to assure that $\tilde{\mathbf{N}} \stackrel{d}{=} \mathbf{N}$ for all such convergent subsequences (l_n). To obtain this use that $\Psi_{\mathbf{N}^{(l_n)}}(f) \to \Psi_{\tilde{\mathbf{N}}}(f)$ and $\Psi_{\mathbf{N}^{(n)}}(f) \to \Psi_{\mathbf{N}}(f)$ for all $f \in \mathcal{C}_c^+(E)$ by assumption. This implies that $\Psi_{\tilde{\mathbf{N}}} = \Psi_{\mathbf{N}}$ and hence $\tilde{\mathbf{N}} \stackrel{d}{=} \mathbf{N}$ by Corollary 3.35. $\square$

We have just seen that it is possible to characterize weak convergence of random measures by Laplace functionals. Although this is in most cases far more convenient that working with the distributions $\mathbb{P}_{\mathbf{N}^{(n)}}$ itself, the Laplace functional can be quite hard to calculate. In the following we provide a very useful result – known as Kallenberg's Theorem – which shows that in the special case of point processes one can prove weak convergence by controlling the behaviour of some basic parameters. Especially this is an important tool to prove weak convergence to a PRM.

Before we can prove this fundamental result some more preparations are required.

Definition 4.37. For $\nu \in \mathcal{M}_p(E)$ the set $S(\nu) := \{x \in E : \nu(\{x\}) > 0\}$ is called SUPPORT OF ν. The map $S^* : \mathcal{M}_p(E) \to \mathcal{M}_p(E)$, $\nu = \sum_{x_k \in S(\nu)} c_k \delta_{x_k} \mapsto \sum_{x_k \in S(\nu)} \delta_{x_k} =: \nu^*$ is called SIMPLIFIER and ν^* is the SIMPLE VERSION OF ν or the SIMPLIFIED ν.

Since we are interested in point processes and not only in point measures, we need to apply the map S^* to random objects. For a point process $\mathbf{N}$ we especially want $S^*(\mathbf{N})$ to be a random element of $\mathcal{M}_p(E)$ which requires measurability. However, S^* only takes care about the underlying support and ignores multiplicities, so one may suspect that it is even measurable w.r.t. a coarser σ-algebra than $\mathcal{B}_E$. This is the heart of the following uniqueness result, which is itself an essential part of Kallenberg's Theorem.

Proposition 4.38 (Uniqueness of point processes). *Let $\mathbf{M}, \mathbf{N}$ be two point processes and $\mathcal{H}$ a $\cap\cup$-base of E. If $\mathbb{P}[M_H = 0] = \mathbb{P}[N_H = 0]$ for all $H \in \mathcal{H}$, then $\mathbf{M}^* \stackrel{d}{=} \mathbf{N}^*$. Moreover, $S^* : (\mathcal{M}_p(E), \sigma(\{\{\mu \in \mathcal{M}_p(E) : \mu(H) = 0\}, H \in \mathcal{H}\})) \to (\mathcal{M}_p(E), \mathscr{M}_p(E))$ is measurable.*

Proof. (i) By assumption the distributions $\mathbb{P}_{\mathbf{M}}$ and $\mathbb{P}_{\mathbf{N}}$ coincide on the family $\mathcal{G} := \{\{\nu \in \mathcal{M}_p(E) : \nu(H) = 0\} : H \in \mathcal{H}\}$. Note that $\mathcal{G}$ is a π-system, since for $G_1, G_2 \in \mathcal{G}$, there are some respective sets $H_1, H_2 \in \mathcal{H}$ s.t. $\mathcal{G}_i = \{\nu \in \mathcal{M}_p(E) : \nu(H_i) = 0\}$, $i = 1, 2$, hence $G_1 \cap G_2 = \{\nu \in \mathcal{M}_p(E) : \nu(H_1) = 0\} \cap \{\nu \in \mathcal{M}_p(E) : \nu(H_2) = 0\} = \{\nu \in \mathcal{M}_p(E) : \nu(H_1 \cup H_2) = 0\} \in \mathcal{G}$ since $H_1 \cup H_2 \in \mathcal{H}$. The Uniqueness Theorem for measures (Theorem A.5) now guarantees that $\mathbb{P}_{\mathbf{M}} = \mathbb{P}_{\mathbf{N}}$ on $\sigma(\mathcal{G})$. Let us assume for the moment, that we have already proven measurability of S^*. Then $(S^*)^{-1}(A) \in \sigma(\mathcal{G})$ for any $A \in \mathcal{M}_p(E)$ and consequently

$$\begin{aligned}\mathbb{P}_{\mathbf{M}^*}[A] &= \mathbb{P}[\mathbf{M}^* \in A] = \mathbb{P}[S^* \circ \mathbf{M} \in A] = \mathbb{P}[\mathbf{M} \in (S^*)^{-1}(A)] = \mathbb{P}_{\mathbf{M}}[(S^*)^{-1}(A)] \\ &= \mathbb{P}_{\mathbf{N}}[(S^*)^{-1}(A)] = \mathbb{P}_{\mathbf{N}^*}[A],\end{aligned}$$

i.e. $\mathbf{M}^* \stackrel{d}{=} \mathbf{N}^*$.

(ii) It remains to show that S^* is measurable from $(\mathcal{M}_p(E), \sigma(\mathcal{G}))$ to $(\mathcal{M}_p(E), \mathscr{M}_p(E))$. Therefore note that $\mathcal{H}$ is a determining class for $\mathcal{B}_E$, since it contains a base of the topology. Consequently $\mathscr{M}_p(E) = \sigma(m_H, H \in \mathcal{H})$ by Proposition 3.10 and the second part of Remark 3.13. Consider the following chain of functions

$$(\mathcal{M}_p(E), \sigma(\mathcal{G})) \stackrel{S^*}{\to} (M_p(E), \sigma(m_H, H \in \mathcal{H})) \stackrel{m_H}{\to} (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0)).$$

By the universal property of the initial σ-algebra, the map S^* is measurable iff $S^*_H := S^* \circ m_H$ is measurable for all $H \in \mathcal{H}$. Fix H for the moment and cover it by finitely many "small" sets in the following way. $\mathcal{H}$ contains a base of the topology, hence for every $x \in H$ there are open sets $x \in H_{x,n} \in \mathcal{H}$, s.t. $\sup_{x,y \in H_{x,n}} d_E(x, y) \leq \frac{1}{n}$. Now H is relatively compact, so for each n the open cover $\{H_{x,n}, x \in H\}$ contains a finite subcover $\{H_{k,n}, 1 \leq k \leq k_n\}$. By looking at the induced partition we can additionally assume that all those elements all pairwise disjoint. Wlog the family $\tilde{\mathcal{H}} := \{\{H_{k,n}, 1 \leq k \leq n\}, n \geq 1\}$ is nested, i.e. for all $1 \leq k \leq n$, there is some $1 \leq l \leq n$ s.t. $H_{l,n+1} \subseteq H_{k,n}$ by intersecting properly.
Now on any $H \in \mathcal{H}$ a point measure ν has finitely many atoms of positive distance, which can be separated by the family $\tilde{\mathcal{H}}$ if n is large enough. Thus

$$S^*_H(\nu) = \nu^*(H) = \lim_{n \to \infty} \sum_{k=1}^{k_n} \nu(H_{k,n}) \wedge 1. \tag{4.13}$$

Due to (4.13) it remains to show that $T^* : (\mathcal{M}_p(E), \sigma(\mathcal{G})) \to (\{0,1\}, \mathcal{P}(\{0,1\}), \nu \mapsto \nu(H_{k,n}) \wedge 1$ is measurable, which is now evident because $(T^*)^{-1}(\{0\}) = \{\nu \in \mathcal{M}_p(E) : \nu(H_{k,n}) = 0\} \in \sigma(\mathcal{G})$ for all $k, n \geq 1$. □

Theorem 4.39 (Kallenberg's Theorem). *Let $\mathbf{N}$ be a simple point process and $\mathcal{H}$ a $\cap\cup$-base of E, s.t. additionally $\mathbb{P}[N_{\partial H} = 0] = 1$ for all $H \in \mathcal{H}$. If $(\mathbf{N}^{(n)})$ is a sequence* *of point processes s.t.*

$$(i)\ \lim_{n\to\infty} \mathbb{P}\big[N_H^{(n)} = 0\big] = \mathbb{P}[N_H = 0] \text{ and} \tag{4.14}$$

$$(ii)\ \lim_{n\to\infty} \mathbb{E}\big[N_H^{(n)}\big] = \mathbb{E}[N_H] < \infty \tag{4.15}$$

for all $H \in \mathcal{H}$. Then $\mathbf{N}^{(n)} \Rightarrow \mathbf{N}$.

Remark 4.40. Before we prove this result, a short remark about the class of sets $\mathcal{H}$ one may use. In the case where E is a subset of $\mathbb{R}^d$ or corresponding compactifications/subsets, we can simply take the family of finite unions of open rectangles. They form a base of the topology on E (induced by the metric) and also a $\cap\cup$-base as already mentioned. The condition that additionally $\mathbb{P}[N_{\partial H} = 0] = 1$ seems to be quite tricky, but poses no real problem. In fact one readily checks that it is possible to just throw away the rectangles whose boundaries get positive mass with positive probability. Hint: The usual cardinality argument shows that there are not more than countably many such "bad" rectangles and hence we can find a sequence of "good" ones approaching them.

Proof. (i) Apply Lemma 4.35 to get tightness of the family $\{\mathbf{N}^{(n)}, n \geq 0\}$:

$$\lim_{t\to\infty} \limsup_{n\to\infty} \mathbb{P}\big[N_H^{(n)} > t\big] \leq \lim_{t\to\infty} \limsup_{n\to\infty} \frac{\mathbb{E}\big[N_H^{(n)}\big]}{t} \overset{(4.15)}{=} \lim_{t\to\infty} \frac{\mathbb{E}[N_H]}{t} \overset{(4.15)}{=} 0.$$

Here we used Markov's inequality in the first step. So far we have only shown that the condition (iii) in Lemma 4.35 holds for all $H \in \mathcal{H}$. But this suffices, since we assumed that $\mathcal{H}$ is a base so we can cover any (relatively) compact set $A \in \mathcal{B}_E$ with finitely many (open) elements of $\mathcal{H}$ and hence $\{\mathbf{N}^{(n)}, n \geq 0\}$ is tight.
For any subsequence (k_n) there is a further subsequence $(l_n) = (k_{m_n})$ s.t. $\mathbf{N}^{(l_n)} \Rightarrow \tilde{\mathbf{N}}$ for some $\tilde{\mathbf{N}}$. Since weak convergence of $\mathbf{N}^{(l_n)}$ implies vague convergence of $\mathbb{P}_{\mathbf{N}^{(l_n)}}$ and $\mathcal{M}_p(E)$ is vaguely closed by Proposition 4.26, $\tilde{\mathbf{N}}$ is point process. It remains to show $\tilde{\mathbf{N}} \overset{d}{=} \mathbf{N}$.

(ii) We claim that

$$\mathbb{P}[\tilde{N}_K = 0] \geq \mathbb{P}[N_K = 0] \text{ for all } K \text{ compact.} \tag{4.16}$$

By Lemma C.5 there is a sequence (f_k) in $\mathcal{C}_c^+(E)$ and compact sets (K_k) s.t. $\mathbb{1}_K \leq f_k \leq \mathbb{1}_{K_k} \searrow \mathbb{1}_K$. Hence $\mathbb{P}[\tilde{N}_K = 0] \geq \mathbb{P}[\tilde{N}(f_k) = 0]$. The set $\{0\}$ is evidently closed so we can use Portmanteau's Theorem and the fact that $\mathbf{N}^{(l_n)} \Rightarrow \tilde{\mathbf{N}}$ to see that the last term is not smaller than $\limsup_{n\to\infty} \mathbb{P}[N^{(l_n)}(f_k) = 0]$. In addition, there is a sequence (H_k) in $\mathcal{H}$ s.t. $K_k \subseteq H_k \searrow K$ since $\mathcal{H}$ contains a base. Summing up, we have

$$\begin{aligned} \mathbb{P}[\tilde{N}_K = 0] &\geq \limsup_{n\to\infty} \mathbb{P}[N^{(l_n)}(f_k) = 0] \geq \limsup_{n\to\infty} \mathbb{P}[N_{K_k}^{(l_n)} = 0] \\ &\geq \limsup_{n\to\infty} \mathbb{P}[N_{H_k}^{(l_n)} = 0] \overset{(4.14)}{=} \mathbb{P}[N_{H_k} = 0], \end{aligned}$$

for all k and as $k \to \infty$ we get $\mathbb{P}[\tilde{N}_K = 0] \geq \mathbb{P}[N_K = 0]$ for all compact K, since $H_k \searrow K$.

(*iii*) We finally prove $\mathbf{N} \stackrel{d}{=} \tilde{\mathbf{N}}^* = \tilde{\mathbf{N}}$. For the second equality use that $\mathcal{H}$ contains a base so it suffices to show that $\tilde{\mathbf{N}}$ has a.s. no multiple point in any $H \in \mathcal{H}$, i.e. $\mathbb{P}[\tilde{N}_H > \tilde{N}_H^*] = 0$. Property (4.16) implies that the map $m_H : \mathcal{M}(E) \to [0,\infty], \nu \mapsto \nu(H)$ is a.s. continuous w.r.t. $\mathbb{P}_{\tilde{\mathbf{N}}}$ by Theorem 4.24 (*iii*) and the fact that $\mathbb{P}[N_{\partial H} = 0] = 1$. Now the Continuous Mapping Theorem yields $N_H^{(l_n)} = m_H\big(\mathbf{N}^{(l_n)}\big) \Rightarrow m_H(\tilde{\mathbf{N}}) = \tilde{N}_H$ for all $H \in \mathcal{H}$. Consequently $\mathbb{P}\big[N_H^{(l_n)} = 0\big] \to \mathbb{P}[\tilde{N}_H = 0]$. But $\mathbb{P}\big[N_H^{(l_n)} = 0\big] \to \mathbb{P}[N_H = 0]$, hence $\mathbb{P}[N_H = 0] = \mathbb{P}[\tilde{N}_H = 0]$ for all $H \in \mathcal{H}$ by (4.14). Proposition 4.38 tells us that $\mathbf{N} = \mathbf{N}^* \stackrel{d}{=} \tilde{\mathbf{N}}^*$ and consequently

$$\mathbb{E}[N_H] = \mathbb{E}[\tilde{N}_H^*] \le \mathbb{E}[\tilde{N}_H] \stackrel{\text{Fatou}}{\le} \liminf_{n\to\infty} \mathbb{E}\big[N_H^{(l_n)}\big] \le \limsup_{n\to\infty} \mathbb{E}\big[N_H^{(l_n)}\big] \stackrel{(4.15)}{=} \mathbb{E}[N_H]$$

for all $H \in \mathcal{H}$. Hence $\mathbb{E}[\tilde{N}_H] = \mathbb{E}[\tilde{N}_H^*]$ for all $H \in \mathcal{H}$, implying $\mathbb{P}[\tilde{N}_H > \tilde{N}_H^*] = 0$ for all $H \in \mathcal{H}$, thus $\tilde{\mathbf{N}}^* = \tilde{\mathbf{N}}$ and $\tilde{\mathbf{N}} \stackrel{d}{=} \mathbf{N}$ as desired. $\square$

5. Introduction to Lévy Processes

So far we have indicated that Lévy processes appear as limits of suitable partial sum processes and are – next to PRM – the central objects in this text. Therefore we provide a (very) short introduction to the theory of Lévy processes in this chapter. A milestone will be the famous Lévy-Itō decomposition, which is proved by the Lévy-Khintchine formula and a constructive argument using random measures (see [5], Chapter 1, Theorem 1).
The results in this part are a combination of Chapter 4 in [18] and the alredy mentioned lecture held by Prof. Zweimüller.

5.1. Examples and Basic Results

Before digging into the theory, we recall the definition and some well-known examples of Lévy processes.

Definition 5.1. A stochastic process $\mathbf{X} = (X_t)_{t\geq 0}$ is called LÉVY PROCESS if

(i) $X_0 = 0$ a.s.

(ii) $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for $0 \leq t_1 \leq \cdots \leq t_n$, $n \geq 1$ (*independent increments*)

(iii) $X_{t+s} - X_s \overset{d}{=} X_t$ for all $s, t \geq 0$ (*stationary increments*)

(iv) $\mathbf{X}$ a.s. has right-continuous paths with left limits. (*càdlàg paths*)

Example 5.2 ((Compound) Poisson process). Any Poisson process $\mathbf{X} = (X_t)_{t\geq 0}$ with parameter $\kappa \geq 0$ is a Lévy process satisfying the additional assumption that $X_t \sim Poi(t\kappa)$ for all $t \geq 0$. More generally, any Compound Poisson process $\mathbf{Z} = (Z_t)_{t\geq 0}$ with measure ν is a Lévy process with $Z_t \sim CPoi(t\nu)$ for all $t \geq 0$. Its characteristic function is given by (see Proposition 4.11)

$$\varphi_{Z_t}(u) = exp\left\{-\int_{\mathbb{R}^d} 1 - e^{-uy}\mathrm{d}(t\nu)(y)\right\} = exp\left\{-t\int_{\mathbb{R}^d} 1 - e^{-uy}\mathrm{d}\nu(y)\right\}. \tag{5.1}$$

Example 5.3 ((Standard) Brownian Motion). This famous stochastic process, which we generically denote by $\mathbf{B} = (B_t)_{t\geq 0}$, is a *continuous* Lévy process which satisfies $B_t \sim \mathcal{N}(0, t)$ (actually $\mathcal{N}(0, t \cdot \mathrm{Id})$) for all $t \geq 0$. More generally one can look at correlated components and add a drift. The characteristic function of a Brownian motion with covariance matrix $Q \in \mathbb{R}^{d\times d}$ and drift $b \in \mathbb{R}^d$ is given by (see Proposition D.1)

$$\varphi_{B_t}(u) = exp\big\{-t(-iub + \tfrac{1}{2}Qu \cdot u)\big\}. \tag{5.2}$$

We prove some elementary results as a warm-up for the following section.

Definition 5.4. A stochastic process $\mathbf{X}$ is STOCHASTICALLY CONTINUOUS if

$$\lim_{s\to t} \mathbb{P}[|X_t - X_s| > \varepsilon] = 0, \text{ i.e. } X_s \xrightarrow{\mathbb{P}} X_t \text{ as } s \to t \tag{5.3}$$

for all $t \geq 0$ and $\varepsilon > 0$ fixed.

Proposition 5.5. *Lévy processes are stochastically continuous.*

Proof. Let (s_k) be a sequence in $[0,\infty)$ s.t. $s_k \to t \geq 0$. By stationarity of the increments $|X_t - X_{s_k}| = |X_{s_k} - X_t| \stackrel{d}{=} |X_{|t-s_k|}|$ for all k and hence

$$\mathbb{P}[|X_t - X_{s_k}| > \varepsilon] = \mathbb{P}[|X_{|t-s_k|}| > \varepsilon] \to \mathbb{P}[|X_0| > \varepsilon] = 0$$

for any $\varepsilon > 0$, since right-continuity implies $X_{|t-s_k|} \stackrel{a.s.}{\to} X_0 = 0$. □

Remark 5.6. As for the Brownian motion, different defining properties can be found in the literature. One of the most common alternative definitions of Lévy processes replaces condition (*iv*) by (5.3). Equivalence follows from the fact that assuming stochastic continuity, one can always find a càdlàg modification of $\mathbf{X}$. For a proof of this regularization result, see [11], Theorem 15.1.

Corollary 5.7. *Lévy processes have almost surely no jumps at fixed times, i.e.* $\mathbb{P}[|\Delta X_t| > 0] = 0$ *for all* $t \geq 0$.

Proof. Proposition 5.5 and the existence of left limits imply both $X_s \stackrel{\mathbb{P}}{\to} X_t$ and $X_s \stackrel{\mathbb{P}}{\to} X_{t-}$ as $s \nearrow t$. Uniqueness of limits gives $X_t = X_{t-}$ almost surely. □

Due to our results on càdlàg functions, a Lévy process a.s. has at most finitely many "large" jumps in bounded intervals and at most countably many discontinuities (see Proposition 2.6).

Proposition 5.8. *The sum of finitely many independent Lévy processes is again a Lévy process.*

Proof. By induction it suffices to show that $\mathbf{Z} := \mathbf{X} + \mathbf{Y}$ is Lévy if $\mathbf{X}$ and $\mathbf{Y}$ are two independent Lévy processes. For stationarity and independence of the increments let $0 \leq q \leq r \leq s \leq t$, $u = (u_1, u_2) \in \mathbb{R}^{d \times d}$, then

$$\begin{aligned}\varphi_{(Z_r - Z_q, Z_t - Z_s)}(u) &= \varphi_{((X_r - X_q)+(Y_r - Y_q),(X_t - X_s)+(Y_t - Y_s))}(u) \\ &= \varphi_{X_r - X_q + Y_r - Y_q}(u_1)\varphi_{X_t - X_s + Y_t - Y_s}(u_2) = \varphi_{Z_{r-q}}(u_1)\varphi_{Z_{t-s}}(u_2).\end{aligned}$$

This already proves that increments of $\mathbf{Z}$ are stationary and a simple generalization of this argument to $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$ for $n \geq 1$ shows independence. That $Z_0 = X_0 + Y_0 = 0$ is obvious and $\mathbf{Z}$ has a.s. càdlàg paths, since the sum of two càdlàg functions is again càdlàg ($\mathcal{D}$ is a vector space, see Remark 2.2). □

Remark 5.9. The preceding proposition shows that for any $t \geq 0$ the characteristic function of $X_t := Z_t + B_t$, where $\mathbf{Z} = (Z_t)_{t\geq 0} \sim CPoi(\nu)$ for some finite measure ν on $\mathbb{R}^d$ and $\mathbf{B} = (B_t)_{t \geq 0}$ is a Brownian motion with covariance matrix Q and drift b, is given by

$$\varphi_{X_t}(u) = exp\Big\{ -t\Big(-iub + \tfrac{1}{2}Qu \cdot u + \int_{\mathbb{R}^d} 1 - e^{iuy}\mathrm{d}\nu(y)\Big)\Big\} \tag{5.4}$$

5.2. The Lévy-Itō Decomposition

In the following we exploit the close relation between Lévy processes and infinitely divisible random variables. Results about the latter and techniques involving random measures are used to prove the Lévy-Itō decomposition.

Definition 5.10. A random variable Z (or its distribution) is called INFINITELY DIVISIBLE if there is an iid vector $(Z^{(m,1)}, \dots, Z^{(m,m)})$ s.t.

$$Z^{(m,1)} + \cdots + Z^{(m,m)} \overset{d}{=} Z$$

for all $m \geq 0$.

Theorem 5.11 (Lévy-Khintchine formula). *Let Z be an infinitely divisible $\mathbb{R}^d$-valued random variable. Then there exist a unique $b \in \mathbb{R}^d$, a positive-semidefinite and symmetric matrix $Q \in \mathbb{R}^{d\times d}$ and a Lévy measure π s.t.*

$$\varphi_Z(u) = \exp\Big\{ - \Big(- iub + \tfrac{1}{2}Qu \cdot u + \int_{\mathbb{R}_0^d} 1 - e^{iuy} + iuy\mathbb{1}_{\{|y|<1\}} \mathrm{d}\pi(y) \Big) \Big\}. \tag{5.5}$$

Proof. For a proof of this result see [2], Theorem 1.2.14 and Theorem 2.4.16. □

Remark 5.12. (i) The fact that π is a Lévy measure guarantees that the integral in (5.5) is well-defined. Indeed, by Taylor's Theorem there is a constant $c \in [0, \infty)$ s.t.

$$\begin{aligned}
&\int_{\mathbb{R}_0^d} |1 - e^{iuy} + iuy\mathbb{1}_{\{y:|y|<1\}}| \mathrm{d}\pi(y) \\
&\qquad \leq \int_{B_1(0)^c} |1 - e^{iuy} + iuy\mathbb{1}_{\{y:|y|<1\}}| \mathrm{d}\pi(y) + \int_{B_1(0)\setminus\{0\}} c|y|^2 \mathrm{d}\pi(y) \\
&\qquad \leq \int_{B_1(0)^c} |1 - e^{iuy}| \mathrm{d}\pi(y) + \int_{B_1(0)\setminus\{0\}} c(1 \wedge |y|^2) \mathrm{d}\pi(y) \\
&\qquad \leq 2 \int_{B_1(0)^c} 1 \wedge |y|^2 \mathrm{d}\pi(y) + c \int_{B_1(0)\setminus\{0\}} 1 \wedge |y|^2 \mathrm{d}\pi(y) < \infty
\end{aligned}$$

(ii) Compare (5.5) and (5.4), the characteristic function of the Lévy process $\mathbf{X} = \mathbf{B} + \mathbf{Z}$ of Remark 5.9.

Definition 5.13. Let Z be an infinitely divisible random variable. Then we call (b, Q, π) satisfying (5.5) the CANONICAL TRIPLE OF Z.

Lemma 5.14. *Let $\mathbf{X}$ be a Lévy process. Then X_t is infinitely divisible for any $t \geq 0$ and*

$$\varphi_{X_t}(u) = \exp\Big\{ - t\Big(- iub + \tfrac{1}{2}Qu \cdot u + \int_{\mathbb{R}_0^d} 1 - e^{iuy} + iuy\mathbb{1}_{\{|y|<1\}} \mathrm{d}\pi(y) \Big) \Big\}, \tag{5.6}$$

where (b, Q, π) is the canonical triple of X_1.

Proof. Since Lévy processes have independent and stationary increments, infinite divisibility follows from

$$X_t = (X_t - X_{t-\frac{t}{n}}) + (X_{t-\frac{t}{n}} - X_{t-\frac{2t}{n}}) + \cdots + (X_{\frac{t}{n}} - X_0)$$

for any $n \in \mathbb{N}$ and $t \geq 0$. Consequently, Theorem 5.11 applies to X_t and comparing (5.5) and (5.6) shows that the only thing left is to prove $\varphi_{X_t} = (\varphi_{X_1})^t$ for any $t \geq 0$.
Let us start with integers. Again stationarity and independence of the increments give $\varphi_{X_n} = (\varphi_{X_1})^n$ via $X_n = (X_n - X_{n-1}) + \cdots + (X_1 - X_0)$. Moreover, since $(\varphi_{X_1})^m = \varphi_{X_m} = (\varphi_{X_{\frac{m}{n}}})^n$ equality even holds for all rationals.
Take $t \geq 0$ and a rational sequence (q_n) s.t. $q_n \searrow t$. Then by dominated convergence

$$\begin{aligned}(\varphi_{X_1}(u))^t &= \lim_{n\to\infty} (\varphi_{X_1}(u))^{q_n} = \lim_{n\to\infty} \varphi_{X_{q_n}}(u) = \lim_{n\to\infty} \mathbb{E}[e^{iuX_{q_n}}] \overset{DCT}{=} \mathbb{E}\big[\lim_{n\to\infty} e^{iuX_{q_n}}\big] \\ &= \mathbb{E}[e^{iuX_t}] = \varphi_{X_t}(u),\end{aligned}$$

for any $u \in \mathbb{R}^d$, since Lévy processes are a.s. right-continuous. □

Definition 5.15. The preceding results allow us to abuse notation and call (b, Q, π) the CANONICAL TRIPLE OF $\mathbf{X}$. Moreover, if $b = 0$ and $Q = 0$ we say that $\mathbf{X}$ is a LÉVY PROCESS WITHOUT GAUSSIAN PART.

Definition 5.16. Let $\mathbf{X}$ be a stochastic process s.t. for all $t \geq 0$

$$\varphi_{X_t}(u) = e^{-t\eta_{\mathbf{X}}(u)}$$

for some $\eta_{\mathbf{X}} \in \mathcal{C}(\mathbb{R}^d, \mathbb{C})$. Then $\eta_{\mathbf{X}}$ is called the CHARACTERISTIC EXPONENT OF $\mathbf{X}$.

Remark 5.17. By Lemma 5.14 and equation (5.6) a Lévy process has a characteristic exponent given by

$$\eta_{\mathbf{X}}(u) = -iub + \tfrac{1}{2}Qu \cdot u + \int_{\mathbb{R}_0^d} 1 - e^{iuy} + iuy\mathbb{1}_{\{|y|<1\}} \mathrm{d}\pi(y), \tag{5.7}$$

where (b, Q, π) is the canonical triple of $\mathbf{X}$.

So far we have seen that we can associate a characteristic exponent – determined by a vector, a matrix and a measure – to any Lévy process. The next step is to construct a corresponding process to such a given triple. As a consequence the canonical triple would completely determine a Lévy process.
To achieve this, we add some more Compound Poisson processes to manipulate (5.4) in such a way that it becomes (5.6). Unfortunately a general Lévy measure π doesn't have to have finite mass, hence summing up finitely many cleverly chosen $CPoi$ won't suffice to collect all discontinuities. Combining countably many, each of them covering a proper height-area is certainly a good idea, but getting from a finite to an infinite sum of Compound Poisson processes poses some problems. For instance it can lead to an accumulation of small jumps if not done properly. Recall what happened in Section 2.2. Yet, this section also indicates a possible solution via the decomposition result (Proposition 2.12). The trick is to compensate the "drift" of the involved processes!

Theorem 5.18. *For any vector $b \in \mathbb{R}^d$, positive-semidefinite and symmetric matrix $Q \in \mathbb{R}^{d\times d}$ and Lévy measure π there exists a Lévy process $\mathbf{X}$ with canonical triple (b, Q, π).*

Proof. Cf. [5], Theorem 1 in Chapter I.
The goal is to construct a Lévy process $\mathbf{X}$ s.t. the characteristic exponent of $\mathbf{X}$ is given by (5.7). We proceed in three steps.

(i) Continuous part
Let $\mathbf{B}$ be a Brownian motion with covariance matrix Q and drift b, then

$$\eta_{\mathbf{B}}(u) = -ibu + \tfrac{1}{2}Qu \cdot u \tag{5.8}$$

by (5.2) and we have already taken care of the first two terms in (5.7).

Now we need to add some proper jumps. Since the measure π contains the information about discontinuities of the process, we try to describe them with a PRM $\mathbf{N}$ on $E := ([0,\infty) \times \mathbb{R}^d, \mathcal{B}_{[0,\infty)\times\mathbb{R}^d})$ with intensity $\lambda \otimes \pi$ which is independent of $\mathbf{B}$.

(ii) Large jumps
Define $\mathbf{Z} = (Z_t)_{t\geq 0}$ via

$$Z_t := \int_{[0,t]\times\{y:|y|>1\}} y \mathrm{d}\mathbf{N}(s,y) = \int_{[0,t]\times\mathbb{R}^d} y \mathrm{d}\mathbf{N}\big|_{\{|y|>1\}}(s,y). \tag{5.9}$$

By Proposition 4.3 we know that $\mathbf{N}\big|_{\{|y|>1\}} \sim PRM\big(\pi\big|_{\{|y|>1\}}\big)$ and since $\pi(\{y : |y| > 1\}) = \int_{\{y:|y|>1\}} 1 \wedge |y|^2 \mathrm{d}\pi(y)$ is finite by (4.4), $\mathbf{Z} \sim CPoi(\pi\big|_{\{|y|>1\}})$ by Theorem 4.13. The characteristic exponent of $\mathbf{Z}$ is given by (see (5.1))

$$\eta_Z(u) = \int_{\mathbb{R}_0^d} (1 - e^{iuy}) \mathrm{d}\pi\big|_{\{y:|y|>1\}}(y) = \int_{\mathbb{R}_0^d} (1 - e^{iuy}) \mathbb{1}_{\{y:|y|>1\}} \mathrm{d}\pi(y). \tag{5.10}$$

(iii) Small jumps
For $n \geq 0$ let $I_n := \big\{y \in \mathbb{R}^d : \frac{1}{2^{n+1}} < |y| \leq \frac{1}{2^n}\big\}$ and define $\mathbf{Z}^{(n)} = (Z_t^{(n)})_{t\geq 0}$ via

$$Z_t^{(n)} := \int_{[0,t]\times I_n} y \mathrm{d}N(s,y) - t\int_{I_n} y \mathrm{d}\pi(y). \tag{5.11}$$

Using the same arguments as before, $\mathbf{Z}^{(n)}$ is a Compound Poisson process having intensity $\pi\big|_{I_n}$ for all n but with an additional drift. Note that the drift is *compensating* in the sense that for all t we have $\mathbb{E}\big[Z_t^{(n)}\big] = 0$ due to Proposition 3.28. To collect the largest jumps of height ≤ 1 define $\mathbf{M}^{(k)} = (M_t^{(k)})_{t\geq 0}$ via

$$M_t^{(k)} := \sum_{n=0}^{k} Z_t^{(n)} = \int_{[0,t]\times\left\{y:\frac{1}{2^{k+1}}<|y|\leq 1\right\}} y \mathrm{d}N(s,y) - t\int_{\left\{y:\frac{1}{2^{k+1}}<|y|\leq 1\right\}} y \mathrm{d}\pi(y). \tag{5.12}$$

Proposition 4.3 tells us that $\mathbf{M}^{(k)}$ is a finite sum of independent Lévy processes (the $I_1, \ldots, I_k$ are pairwise disjoint) and is therefore itself Lévy by Proposition 5.8. Moreover,

$$\varphi_{M_t^{(k)}}(u) = exp\Big\{ - t\Big(\int_{\mathbb{R}_0^d} \Big(1 - e^{iuy} + iuy\Big) \sum_{i=0}^{k} \mathbb{1}_{I_i}(y) \mathrm{d}\pi(y)\Big)\Big\}, \text{ and therefore}$$

$$\eta_{M^{(k)}}(u) = \int_{\mathbb{R}_0^d} \left(1 - e^{iuy} + iuy\right) \mathbb{1}_{\left\{\frac{1}{2^{k+1}} < |y| \leq 1\right\}} \mathrm{d}\pi(y). \tag{5.13}$$

$\mathbf{M}^{(k)}$ is a finite sum of processes which are independent from $\mathbf{B}$ and $\mathbf{Z}$ and is hence itself independent from $\mathbf{B}$ and $\mathbf{Z}$ for all k. Now when we look at the characteristic exponent of $\tilde{\mathbf{X}} := \mathbf{B} + \mathbf{Z} + \mathbf{M}^{(k)}$ – the sum of those in (5.8), (5.10) and (5.13) – and compare it to the desired result (5.7), we see that the „only" thing left is $k \to \infty$. But this step is non-trivial since a priori it is not clear that the limit of $\mathbf{M}^{(k)}$, so it exists, is again Lévy, independent from $\mathbf{B}$ and $\mathbf{Z}$, or has the desired characteristic exponent. The following theorem, which we proof later, addresses these problems.

Theorem 5.19. *There is a Lévy $\mathcal{L}^2$-martingale $\mathbf{M} = (M_t)_{t\geq 0}$ w.r.t. the filtration $(\mathcal{F}_t)_{t\geq 0} := \left(\sigma\left(\mathbf{N}\big|_{[0,s]\times\mathbb{R}^d}, 0 \leq s \leq t\right)\right)_{t\geq 0}$ which is independent from $\mathbf{B}$ and $\mathbf{Z}$ s.t. $\left\|M_t^{(k)} - M_t\right\|_2 \to 0$ as $k \to \infty$ for all $t \geq 0$.*
Furthermore, for all $T \geq 0$ there is a subsequence (k_j) s.t. almost surely $M_t^{(k_j)} \to M_t$ uniformly on $[0, T]$, that is

$$\mathbb{P}\left[\left\|\mathbf{M}^{(k_j)} - \mathbf{M}\right\|_{T,\infty} \to 0 \text{ as } j \to \infty\right] = 1.$$

We continue the proof of Theorem 5.18. Using the above result and dominated convergence along the almost surely converging subsequence (k_j), we get

$$\begin{aligned}
\varphi_{M_t}(u) = \mathbb{E}[e^{iuM_t}] &= \mathbb{E}\Big[\lim_{j\to\infty} e^{iuM_t^{(k_j)}}\Big] \overset{DCT}{=} \lim_{j\to\infty} \mathbb{E}\Big[e^{iuM_t^{(k_j)}}\Big] \\
&\overset{(5.13)}{=} \lim_{j\to\infty} exp\Big\{-t\Big(\int_{\mathbb{R}_0^d} (1 - e^{iuy})\mathbb{1}_{\{\frac{1}{2^{k_j+1}} < |y| \leq 1\}} \mathrm{d}\pi(y)\Big)\Big\} \\
&\overset{DCT}{=} exp\Big\{-t\Big(\int_{\mathbb{R}_0^d} (1 - e^{iuy})\mathbb{1}_{\{0<|y|\leq 1\}} \mathrm{d}\pi(y)\Big)\Big\} \text{ and therefore}
\end{aligned}$$

$$\eta_{\mathbf{M}}(u) = \int_{\mathbb{R}_0^d} (1 - e^{iuy})\mathbb{1}_{\{0<|y|\leq 1\}} \mathrm{d}\pi(y). \tag{5.14}$$

Combining the characteristic exponents of the independent processes $\mathbf{B}$, $\mathbf{Z}$ and $\mathbf{M}$ shows that $\mathbf{X} := \mathbf{B} + \mathbf{Z} + \mathbf{M}$ satisfies (5.7). Hence we have successfully constructed a Lévy process to a given characteristic exponent/canonical triple. □

A closer look at the proof of Theorem 5.18 together with Theorem 5.11 gives the following famous result.

Corollary 5.20 (Lévy-Itō decomposition). *Let $\mathbf{X}$ be a Lévy process. Then there is a canonical triple (b, Q, π), s.t. its characteristic exponent is given by (5.7). Moreover, there are a Brownian motion $\mathbf{B}$ with covariance matrix Q and drift b, a Compound Poisson process $\mathbf{Z}$ and a $\mathcal{L}^2$-martingale $\mathbf{M}$, all mutually independent, s.t. $\mathbf{X} \overset{d}{=} \mathbf{B} + \mathbf{Z} + \mathbf{M}$.* □

Proof of Theorem 5.19. We proceed in several steps.
(i) Existence of $\mathbf{M}$: Fix $k \geq 1$ and $T > 0$. Applying Theorem 4.18 to $B_k := \biguplus_{n=0}^k I_n$ assures that the process $\mathbf{M}^{(k)}$ is a martingale w.r.t. the given filtration. Since $B_k \subseteq (0, 1]$

and therefore $\int_{B_k} |y|^2 \mathrm{d}\pi(y) = \int_{\mathbb{R}_0^d} 1 \wedge |y|^2 \mathrm{d}\pi(y) < \infty$ it is even a $\mathcal{L}^2$-martingale, i.e. $M_t^{(k)} \in \mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, for all $t \geq 0$.
Let $l \geq k \geq 1$ and apply Theorem 4.18 to $B_{k,l} := I_{k+1} \cup \cdots \cup I_l$

$$\mathbb{E}\Big[\big|M_T^{(l)} - M_T^{(k)}\big|^2\Big] = \mathbb{E}\Big[\big|Z_T^{(B_{k,l})}\big|^2\Big] \overset{\text{Thm. 4.18}}{=} T^2 \int_{\mathbb{R}_0^d} |y|^2 \mathbb{1}_{B_{k,l}} \mathrm{d}\pi(y) \to 0, \text{ as } k \to \infty,$$

by dominated convergence ($|y|^2 \mathbb{1}_{B_{k,l}} \leq 1 \wedge |y|^2$ and the latter is integrable, since π is a Lévy measure).
Therefore the sequence $(M_T^{(k)})$ is Cauchy in $\mathcal{L}^2$, thus there exists a limit $M_T \in \mathcal{L}^2$ s.t. $M_T^{(k)} \overset{\mathcal{L}^2}{\to} M_T$. Define a process $\mathbf{M} = (M_t)_{t \in [0,T]}$ via $M_t := \mathbb{E}[M_T|\mathcal{F}_t]$. It is easy to see that $\mathbf{M}$ is a $\mathcal{L}^2$-martingale. Moreover, the conditional expectation is a contraction on $\mathcal{L}^2$ (see Proposition B.2), hence

$$M_t^{(k)} = \mathbb{E}\big[M_T^{(k)}|\mathcal{F}_t\big] \overset{\mathcal{L}^2}{\to} \mathbb{E}[M_T|\mathcal{F}_t] = M_t \text{ for all } t \in [0,T].$$

(ii) $\mathbf{M}$ *has a.s càdlàg paths:* Applying Doob's inequality (Theorem B.3) to $\mathbf{M}^{(k)} - \mathbf{M}$ yields

$$\mathbb{E}[\|\mathbf{M}^{(k)} - \mathbf{M}\|_{T,\infty}^2] = \mathbb{E}\big[\|(\mathbf{M}^{(k)} - \mathbf{M})^2\|_{T,\infty}\big] \overset{Doob}{\leq} 4\mathbb{E}\big[|M_T^{(k)} - M_T|^2\big] \to 0.$$

This means that $\|\mathbf{M}^{(k)} - \mathbf{M}\|_{T,\infty} \overset{\mathcal{L}^2}{\to} 0$ and by Proposition A.4 there is a subsequence (k_j) s.t. almost surely

$$\big(M_t^{(k_j)}\big)_{t \in [0,T]} \to (M_t)_{t \in [0,T]} \text{ uniformly.} \tag{5.15}$$

Consequently $\mathbf{M}$ has a.s. càdlàg paths by Lemma 2.9 or Remark 2.28. Moreover, $M_0 = 0$, since $M_0^{(k_j)} = 0$ for all j.

(iii) $\mathbf{M}$ *has independent and stationary increments:* The following standard argument shows stationarity and independence of two disjoint increments, it can easily be extended (by induction).

$$\begin{aligned}
\varphi_{(M_t - M_s, M_r - M_q)}((u_1, u_2)) &= \mathbb{E}\left[e^{i(M_t - M_s)u_1 + i(M_r - M_q)u_2}\right] \\
&\overset{DCT}{=} \lim_{j \to \infty} \mathbb{E}\left[e^{i\left(M_t^{(k_j)} - M_s^{(k_j)}\right)u_1 + i\left(M_r^{(k_j)} - M_q^{(k_j)}\right)u_2}\right] \\
&= \lim_{j \to \infty} \mathbb{E}\left[e^{iM_{t-s}^{(k_j)}u_1}\right] \mathbb{E}\left[e^{iM_{r-q}^{(k_j)}u_2}\right] \\
&\overset{DCT}{=} \varphi_{M_{t-s}}(u_1)\varphi_{M_{r-q}}(u_2).
\end{aligned}$$

(iv) $\mathbf{M}$ *is independent from* $\mathbf{B}$ *and* $\mathbf{Z}$*:* We only show independence of $\mathbf{M}$ and $\mathbf{Z}$, the exact same argument applies to $\mathbf{M}$ and $\mathbf{B}$.
First note that $\mathbf{M}$ and $\mathbf{Z}$ are both Lévy, hence the required independence of the vectors $(M_{t_0}, M_{t_1} - M_{t_0}, \ldots, M_{t_n} - M_{t_{n-1}}) \overset{d}{=} (M_{t_0}, M_{t_1 - t_0}, \ldots, M_{t_n - t_{n-1}})$ and $(Z_{t_0}, Z_{t_1} - Z_{t_0}, \ldots, Z_{t_n} - Z_{t_{n-1}}) \overset{d}{=} (Z_{t_0}, Z_{t_1 - t_0}, \ldots, Z_{t_n - t_{n-1}})$ for $0 \leq t_0 < \cdots < t_n$, $n \geq 1$ reduces to

the independence of M_t and Z_t for all $t \geq 0$. Recalling that $\mathbf{M}^{(k)}$ and $\mathbf{Z}$ are independent for all k, we conclude with a similar argument as in *(iii)*:

$$\begin{aligned}
\varphi_{(M_t,Z_t)}((u_1,u_2)) = \mathbb{E}\left[e^{iM_tu_1+iZ_tu_2}\right] &\overset{DCT}{=} \lim_{j\to\infty} \mathbb{E}\left[e^{iM_t^{(k_j)}u_1+iZ_tu_2}\right] \\
&= \lim_{j\to\infty} \mathbb{E}\left[e^{iM_t^{(k_j)}u_1}\right]\mathbb{E}\left[e^{iZ_tu_2}\right] \\
&\overset{DCT}{=} \varphi_{M_t}(u_1)\varphi_{Z_t}(u_2).
\end{aligned}$$

and we are (finally) done. □

Theorem 5.21. *For $\varepsilon > 0$ define $\mathbf{M}^{(\varepsilon)} = (M_t^{(\varepsilon)})_{t\geq 0}$ via*

$$M_t^{(\varepsilon)} := \int_{[0,t]\times\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\mathbf{N}(s,y) - t\int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y). \tag{5.16}$$

Then a.s. $\mathbf{M}^{(\varepsilon)} \to \mathbf{M}$ locally uniformly, i.e. for all $T \geq 0 : \|\mathbf{M}^{(\varepsilon)} - \mathbf{M}\|_{T,\infty} \overset{a.s.}{\to} 0$.

Proof. The following consists of (slightly) adapted proofs of Lemma 20.2 and 20.3 from [18].
The above argument especially showed that for any sequence $\varepsilon_n \to 0$ local uniform convergence holds almost surely along a suitable subsequence $(\tilde{\varepsilon}_n)$ (recall (5.12) and (5.15)). The remaining task is thus to show that this is also true for the *whole* sequence, i.e. $\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}\|_{T,\infty} \overset{a.s.}{\to} 0$ for any $T > 0$.

(i) Fix some $T > 0$ and assume wlog that $1 > \varepsilon_n \downarrow 0$ and set $\varepsilon_0 := 1$. This is possible, since only the way the sequence approaches 0 is important for the convergence behaviour.
Following (5.11) set $\tilde{I}_n := \{y \in \mathbb{R}^d : \varepsilon_n < |y| \leq \varepsilon_{n-1}\}$ and define $\tilde{\mathbf{Z}}^{(n)} := (\tilde{Z}_t^{(n)})_{t\geq 0}$ for $n \geq 1$ via

$$\tilde{Z}_t^{(n)} := \int_{[0,t]\times\tilde{I}_n} y\mathrm{d}N(s,y) - t\int_{\tilde{I}_n} y\mathrm{d}\pi(y).$$

As before, $(\tilde{\mathbf{Z}}^{(n)})$ is a sequence of independent Lévy process and we can write $\mathbf{M}^{(\varepsilon_n)} = \sum_{k=0}^n \tilde{\mathbf{Z}}^{(k)}$ for all n.

(ii) We claim that for any n and $\delta > 0$

$$\mathbb{P}\Big[\max_{1\leq j\leq n} \|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > 3\delta\Big] \leq 3 \max_{1\leq j\leq n} \mathbb{P}\Big[\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > \delta\Big] \tag{5.17}$$

This statement is known as Etemadi's inequality and we follow the proof of lemma 20.2 in [18].
Abbreviate $\mathbf{S}^{(n)} := \max_{1\leq j\leq n} \|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty}$ for $n \geq 1$ and let $a, b > 0$. The events $A_j := \{\mathbf{S}^{(j-1)} \leq a + b < \|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty}\}$ are evidently pairwise disjoint and $\{\mathbf{S}^{(n)} > a + b\} =$

$\biguplus_{j=1}^n A_j$. We calculate

$$\begin{aligned}
\mathbb{P}\big[\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > a\big] \geq \mathbb{P}\Big[\{\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > a\} \cap \biguplus_{j=1}^{n} A_j\Big] &= \mathbb{P}\Big[\biguplus_{j=1}^{n} (A_j \cap \{\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > a\})\Big] \\
&= \sum_{j=1}^{n} \mathbb{P}[A_j \cap \{\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > a\}] \\
&\geq \sum_{j=1}^{n} \mathbb{P}[A_j \cap \{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq b\}].
\end{aligned}$$

The last inequality holds, since given A_j we know that $\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > a + b$, hence $\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > a$ is especially satisfied, if the increment $\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty}$ doesn't exceed b. Moreover, A_j only depends on $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(j)}$, while $\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}$ depends on $\mathbf{Z}^{(j+1)}, \dots, \mathbf{Z}^{(n)}$, hence they are independent and

$$\begin{aligned}
&\sum_{j=1}^{n} \mathbb{P}[A_j \cap \{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq b\}] \\
&\qquad = \sum_{j=1}^{n} \mathbb{P}[A_j]\mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq b\}] \\
&\qquad \geq \min_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq b\}] \sum_{j=1}^{n} \mathbb{P}[A_j] \\
&\qquad = \min_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq b\}]\mathbb{P}[\mathbf{S}^{(n)} > a + b].
\end{aligned}$$

Let $a = \delta$ and $b = 2\delta$, then

$$\begin{aligned}
\mathbb{P}\big[\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > \delta\big] &\geq \mathbb{P}[\mathbf{S}^{(n)} > 3\delta] \min_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} \leq 2\delta\}] \\
&= \mathbb{P}[\mathbf{S}^{(n)} > 3\delta]\Big(1 - \max_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > 2\delta\}]\Big) \\
&\geq \mathbb{P}[\mathbf{S}^{(n)} > 3\delta]\Big(1 - 2\max_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > \delta\}]\Big). \qquad (5.18)
\end{aligned}$$

For the last step, we used that

$$\begin{aligned}
\max_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > 2\delta] &\leq \max_{1\leq j\leq n} \mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > \delta\} \cup \{\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > \delta\}] \\
&\leq \max_{1\leq j\leq n} \big(\mathbb{P}[\{\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > \delta] + \mathbb{P}[\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > \delta]\big) \\
&\leq 2\max_{1\leq j\leq n} \mathbb{P}[\|\mathbf{M}^{(\varepsilon_j)}\|_{T,\infty} > \delta].
\end{aligned}$$

Note that if on the one hand $\max_{1\leq k\leq n} \mathbb{P}\Big[\|\mathbf{M}^{(\varepsilon_k)}\|_{T,\infty} > \delta\Big] \geq \frac{1}{3}$, equation (5.17) is satisfied for trivial reasons. On the other hand, if $\max_{1\leq k\leq n} \mathbb{P}\Big[\|\mathbf{M}^{(\varepsilon_k)}\|_{T,\infty} > \delta\Big] < \frac{1}{3}$, then (5.18) shows that

$$\mathbb{P}\big[\|\mathbf{M}^{(\varepsilon_n)}\|_{T,\infty} > \delta\big] \geq \tfrac{1}{3}\mathbb{P}[\mathbf{S}^{(n)} > 3\delta]$$

and (5.17) holds as well.

(ii) We finally prove $\|\mathbf{M}^{(\varepsilon)} - \mathbf{M}\|_{T,\infty} \overset{a.s.}{\to} 0$ by using the (slightly) adapted argument from Lemma 20.3 of [18].

First note that $\|\mathbf{M}^{(\varepsilon)_j} - \mathbf{M}\|_{T,\infty} \overset{\mathbb{P}}{\to} 0$ by part *(ii)* of the proof of Corollary 5.20, since convergence in $\mathcal{L}^1$ implies convergence in probability. Applying the first part to $\tilde{\mathbf{Z}}^{(n+1)}, \tilde{\mathbf{Z}}^{(n+2)}, \dots$ we get for any $m > n$

$$\mathbb{P}\Big[\max_{n \le j \le m} \|\mathbf{M}^{(\varepsilon_j)} - \mathbf{M}\|_{T,\infty} > 3\delta\Big] \overset{(5.17)}{\le} 3 \max_{n \le j \le m} \mathbb{P}\Big[\|\mathbf{M}^{(\varepsilon_j)} - \mathbf{M}\|_{T,\infty} > \delta\Big]. \tag{5.19}$$

Thus for $m \to \infty$ (and the continuity of measures)

$$\mathbb{P}\Big[\sup_{j \ge n} \|\mathbf{M}^{(\varepsilon_j)} - \mathbf{M}\|_{T,\infty} > 3\delta\Big] \le 3 \sup_{j \ge n} \mathbb{P}\Big[\|\mathbf{M}^{(\varepsilon_j)} - \mathbf{M}\|_{T,\infty} > \delta\Big] \to 0 \text{ as } n \to \infty,$$

since $\|\mathbf{M}^{(\varepsilon)} - \mathbf{M}\|_{T,\infty} \overset{\mathbb{P}}{\to} 0$. Consequently the left-hand side tends to 0 and because $\delta > 0$ was arbitrary,

$$\lim_{n\to\infty} \sup_{j \ge n} \|\mathbf{M}^{(\varepsilon_j)} - \mathbf{M}\|_{T,\infty} = \limsup_{n\to\infty} \|\mathbf{M}^{(\varepsilon_n)} - \mathbf{M}\|_{T,\infty} = 0 \text{ almost surely}$$

as desired. □

5.3. Weak Convergence in $\mathcal{D}$

As for random measures, we discuss weak convergence of random elements in $\mathcal{D}$, although actually we just prove a little Lemma which will be useful in the next chapter.

Lemma 5.22. *Let $\mathbf{X}$ be a Lévy process and $(\mathbf{X}^{(n)})$ a sequence of stochastic processes. Then $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ iff $\mathbf{X}^{(n)}\big|_{[0,T]} \Rightarrow \mathbf{X}\big|_{[0,T]}$ for all $T > 0$, i.e. $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ in $\mathcal{D}$ iff $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ in $\mathcal{D}_T$.*

Proof. ($\Rightarrow$) Fix $T \ge 0$. To apply the Continuous Mapping Theorem, we need to check that the map $Rest_T : \mathcal{D} \to \mathcal{D}_T$, $\mathbf{x} \mapsto \mathbf{x}\big|_{[0,T]}$ is a.s. continuous w.r.t. the distribution of $\mathbf{X}$. Since $Rest_T$ is continuous iff $|\Delta\mathbf{x}(T)| = 0$ by Theorem 2.19/Remark 2.20 this is indeed true, because Corollary 5.7 implies that $\mathbb{P}_{\mathbf{X}}[\{\mathbf{x} \in \mathcal{D} : |\Delta\mathbf{x}(T)| > 0\}] = \mathbb{P}[|\Delta X_T| > 0] = 0$. Now the CMT implies that

$$\mathbf{X}^{(n)}\big|_{[0,T]} = Rest_T(\mathbf{X}^{(n)}) \Rightarrow Rest_T(\mathbf{X}) = \mathbf{X}\big|_{[0,T]}.$$

($\Leftarrow$) This is a simple application of the Dominated Convergence Theorem. For all $f \in \mathcal{C}_b(\mathcal{D}, \mathbb{R})$

$$\begin{aligned}\mathbb{E}[f(\mathbf{X}^{(n)})] = \mathbb{E}\Big[\lim_{T\to\infty} f\Big(\mathbf{X}^{(n)}\big|_{[0,T]}\Big)\Big] &\overset{DCT}{=} \lim_{T\to\infty} \mathbb{E}\Big[f\Big(\mathbf{X}^{(n)}\big|_{[0,T]}\Big)\Big] \\ &\overset{\text{assum.}}{\to} \lim_{T\to\infty} \mathbb{E}\big[f(\mathbf{X}\big|_{[0,T]})\big] \overset{DCT}{=} \mathbb{E}\Big[\lim_{T\to\infty} f\big(\mathbf{X}\big|_{[0,T]}\big)\Big] = \mathbb{E}[f(\mathbf{X})],\end{aligned}$$

thus $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ in $\mathcal{D}$. □

6. Characterizing Weak Convergence in $\mathcal{D}$

The basic structure of this chapter is roughly given by three parts. After a general motivation and some preparations in the first, we prove two important continuity results in Section 6.3. The proof of the first one – although well-known – was created single-handedly using results presented in Section 2.4. The other statement is presented according to [16], Section 7.2.3, but generalized to the d-dimensional case. In the remainder of this section a short argument is provided which allows an slight generalization of the original result in [20].
In the very last part the proof of the central theorem from Marta Tyran-Kamińska is presented.

6.1. Motivation

In the introduction we recalled some of the most famous weak convergence results in probability theory, the Central Limit Theorem and Donsker's Theorem. Let us take a step back to get a more abstract view on the problems those theorems faced. In fact, when searching for some *statistic regularity*, that is regularity shown by repeating some experiment over and over again, the following situation arises. If (X_k) is an iid sequence of random variables, are there scaling and translating sequences (b_k) and (c_k) s.t.

$$X^{(n)} := \tfrac{1}{b_n}(X_1 + \cdots + X_n - c_n) \Rightarrow X, \tag{6.1}$$

for some random variable X as $n \to \infty$? For the classic CLT, we need X_1 to have finite and positive variance σ^2. Then (6.1) holds for $b_n = \sqrt{n\sigma^2}$, $c_n = n\mathbb{E}[X_1]$ and $X \sim \mathcal{N}(0,1)$ (see (1.1)). Donsker's Theorem tells us that (6.1) holds for $X_n := \mathbf{S}^{(n)} \in \mathcal{D}_1$ as in (1.2), $b_n = \sqrt{n\sigma^2}$ and $c_n = nt\mathbb{E}\big[S_1^{(1)}\big] \in \mathcal{D}_1$.
Equivalently, we can embed this result in a functional generalization of (6.1)

$$(X_t^{(n)})_{t\geq 0} = \mathbf{X}^{(n)} \Rightarrow \mathbf{X} \text{ in } \mathcal{D} \qquad \text{for} \qquad X_t^{(n)} := \frac{1}{b_n}\Big(\sum_{1\leq k\leq nt} X_k - tc_n\Big). \tag{6.2}$$

In fact Donsker says that (6.2) holds on $[0,1]$ for a real-valued iid sequence (X_k) having finite positive variance σ^2, $b_n = \sqrt{n\sigma^2}$, $c_n = n\mathbb{E}[X_1]$ and $\mathbf{X} = \mathbf{B}$.

It turns out that some other famous results, notably Poisson Limit Theorems, need a slightly more general setting since we cannot describe them in terms of (6.1) or (6.2). Therefore we not only look at a sequence (X_k) of random variables, but a *scheme*, which is a family of random variables of the form $(X_{k,n})_{k,n\geq 1}$, and redefine $\mathbf{X}^{(n)} = (X_t^{(n)})_{t\geq 0}$ in (6.2) via

$$X_t^{(n)} := \Big(\sum_{1\leq k\leq nt} X_{k,n} - tc_n\Big). \tag{6.3}$$

Note that to simplify notation one can add the factor $\frac{1}{b_n}$ into the scheme $(X_{k,n})$ due to its dependence on n. Evidently both the CLT and Donsker's Theorem can be embedded in this more general version (assuming some iid-conditions on the scheme). This in now the setting we are interested in. We want to characterize

$$\mathbf{X}^{(n)} \Rightarrow \mathbf{X} \text{ in } \mathcal{D} \tag{6.4}$$

for some $\mathbf{X}^{(n)}$ given through (6.3), by weak convergence of a cleverly chosen sequence $(\mathbf{N}^{(n)})$ of random measures.
Since all jumps of $\mathbf{X}^{(n)}$ are given by $X_{k,n}$ and happen at times $\frac{k}{n}$ the sequence nearly suggests itself. We collect this information in

$$\mathbf{N}^{(n)} := \sum_{k\geq 1} \delta_{\left(\frac{k}{n}, X_{k,n}\right)} \text{ for } n \geq 1. \tag{6.5}$$

Now does the sequence $(\mathbf{N}^{(n)})$ converges weakly if $(\mathbf{X}^{(n)})$ does? How does the potential limit look like and how are $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ and $\mathbf{N}^{(n)} \Rightarrow \mathbf{N}$ related? We address these questions in the following and give an answer in Theorem 6.15.

6.2. Preparations

This small section contains some easy technical properties needed later on. It purely exists for completeness reasons, so the reader may skim through this part quickly if desired.

Lemma 6.1. *Let ν be a σ-finite measure on $E := \overline{\mathbb{R}}^d_0$. Then there are at most countably many $u > 0$ s.t. $\nu(\{x : |x| = u\}) > 0$.*

Proof. Assume there are uncountably many $u > 0$ s.t. $\nu(\partial B_u(0)) > 0$, where $B_u(0)$ denotes the ball of radius u around 0. Since ν is σ-finite, there are sets A_n s.t. $E = \bigcup_{n\geq 1} A_n$ and $\nu(A_n) < \infty$ for all n. This is a countable union, hence there is n_0 s.t. A_{n_0} contains uncountably many hyper-surfaces $\partial B_u(0)$ having positive ν-mass. Uniting the first sets, we can wlog assume that $n_0 = 1$.
For $k \geq 1$ let $R_k := \{u > 0 : \nu(\{x : |x| = u\}) \geq \frac{1}{k}\}$ and $R_0 := \{u > 0 : \nu(\{x : |x| = u\}) > 0\}$, then $R_0 = \bigcup_{k\geq 1} R_k$ and hence $R_0 \cap A_1 = \bigcup_{k\geq 1}(R_k \cap A_1)$.
Again this is a countable union, while $R_0 \cap A_1$ is assumed to be uncountable. Consequently there is an index k_1 s.t. $R_{k_1} \cap A_1$ contains uncountably many elements, what implies that there are infinitely many distinct $u_i \in A_1$ satisfying $\nu(\{x : |x| = u_i\}) \geq \frac{1}{k_1}$ in A_1, contradicting $\nu(A_1) < \infty$. ☐

Definition 6.2. Let $\mathbf{X}$ be a stochastic process. A measurable function $\tau : (\Omega, \mathcal{A}) \to ([0,\infty], \mathcal{B}_{[0,\infty]})$ is a RANDOM TIME. In particular, τ is called STOPPING TIME W.R.T. THE (wlog augmented) CANONICAL FILTRATION $\mathcal{F}_t(\mathbf{X}) := \sigma(X_s, s \leq t)$, $t \geq 0$ if $\{\tau > t\} \in \mathcal{F}_t(\mathbf{X})$ for all $t \geq 0$.
For a random time τ the map

$$X_\tau : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}),\ \omega \mapsto \begin{cases} \lim\limits_{n\to\infty} X_{\tau(\omega)\wedge n}(\omega) & \text{if the limits exists and is finite} \\ 0 & \text{otherwise.} \end{cases}$$

is the VALUE OF $\mathbf{X}$ AT τ.

Definition 6.3. For a fixed $\varepsilon > 0$ define the times $T_\varepsilon^{(i)}(\mathbf{X})$, $i \geq 1$ when $\mathbf{X}$ has a jump of size $> \varepsilon$ inductively by $T_\varepsilon^{(0)}(\mathbf{X}) := 0$ and $T_\varepsilon^{(i+1)}(\mathbf{X}) := \inf\{t > T_\varepsilon^{(i)}(\mathbf{X}) : |\Delta X_t| > \varepsilon\}$.
Let $\mathbf{X}$ be a Lévy process and set $U(\mathbf{X}) := \{\varepsilon > 0 : \mathbb{P}[|\Delta X_t| = \varepsilon \text{ for some } t > 0] > 0\}$.

Remark 6.4. Note that the objects defined above are "random" versions of their corresponding counterparts in Definition 2.35

Before we can actually work with those new objects, we better assure that they are all measurable and well-defined.

Proposition 6.5 (Measurability). *Let* $\mathbf{X}$ *be a Lévy process and let* τ *be a random time. Then*

(i) *the map* $|\Delta_t| : (\Omega, \mathcal{A}) \to ([0,\infty), \mathcal{B}_{[0,\infty)})$, $\omega \mapsto |\Delta X_t(\omega)|$ *is a random variable for any* $t > 0$. *Moreover, the set* $\{|\Delta X_t| = \varepsilon\}$ *is measurable for any* $t, \varepsilon > 0$

(ii) X_τ *is measurable*

(iii) *for any* $\varepsilon > 0$ *the maps* $T_\varepsilon^{(i)}(\mathbf{X}) := 0$, $i \geq 0$ *given by* $T_\varepsilon^{(0)}(\mathbf{X}) := 0$, $T_\varepsilon^{(i+1)}(\mathbf{X}) := \inf\{t > T_\varepsilon^{(i)}(\mathbf{X}) : |\Delta X_t| > \varepsilon\}$ *are stopping times w.r.t. the (augmented) canonical filtration*

(iv) *the set* $\{|\Delta X_t| = \varepsilon \text{ for some } t \geq 0\}$ *is measurable.*

Proof. (i) Fix some $t, \varepsilon > 0$. For measurability of $|\Delta_t|$ it suffices to show that the set $\{|\Delta X_t| > \varepsilon\}$ is measurable. Note that if ω is such that $|\Delta X_t(\omega)| > \varepsilon$, then there are arbitrarily close rationals $p < t < q \in \mathbb{Q}$ s.t. $|X_q(\omega) - X_p(\omega)| > \varepsilon$. Hence we can let them approach t, i.e. find sequences (p_n), (q_n) in $\mathbb{Q}$ s.t.

$$t - \tfrac{1}{n} \leq p_n < t < q_n \leq t + \tfrac{1}{n} \text{ and } |X_{q_n}(\omega) - X_{p_n}(\omega)| > \varepsilon \tag{6.6}$$

for all $n \geq 1$. Thus the following is true

$$\{|\Delta X_t| > \varepsilon\} = \bigcup_{m \geq 1} \bigcap_{n \geq m} \bigcup_{\substack{p_n, q_n \in \mathbb{Q} \\ \text{as in (6.6)}}} \{|X_{q_n} - X_{p_n}| > \varepsilon\}. \tag{6.7}$$

Indeed, "$\subseteq$" is satisfied due to what we had above and since $\mathbf{X}(\omega)$ has at most countably many jumps for a fixed ω, we also get "$\supseteq$". Measurability is now an easy consequence of (6.7).
Since

$$\begin{aligned}\{|\Delta X_t| = \varepsilon\} &= \bigcap_{m,n \geq 1} \left(\{|\Delta X_t| \leq \varepsilon + \tfrac{1}{m}\} \cap \{|\Delta X_t| > \varepsilon - \tfrac{1}{n}\}\right) \\ &= \bigcap_{m,n \geq 1} \left(\{|\Delta X_t| > \varepsilon + \tfrac{1}{m}\}^c \cap \{|\Delta X_t| > \varepsilon - \tfrac{1}{n}\}\right)\end{aligned}$$

the second statement follows.

(ii) If τ assumes at most countably many different values, then the statement is rather easy to verify. Indeed, by the definition of X_τ and the fact that $\{\tau < \infty\}$ is measurable it suffices to show that $\{\tau \leq n\} \cap \{|X_\tau| \leq t\}$ is a measurable set. Fix some n and denote the values of τ which are $\leq n$ by $\{s_1, s_2, \dots\}$. Observe that

$$\{\tau \leq n\} \cap \{|X_\tau| \leq t\} = \bigcup_{n \geq 1} \{\tau = s_n\} \cap \{|X_\tau| \leq t\} = \bigcup_{n \geq 1} \{\tau = s_n\} \cap \{|X_{s_n}| \leq t\}$$

is measurable, since τ is a random time assuming only countably many values and $\mathbf{X}$ is a stochastic process.
For an arbitrary random time τ there is a sequence (τ_k) of random times only assuming values in $\{\frac{j}{k}, j \geq 0\}$ s.t. $\tau_k \searrow \tau$ pointwise (e.g. $\tau_k(\omega) := \frac{\lceil k\tau(\omega)\rceil}{k}$). By right-continuity of $\mathbf{X}$ we know that $X_{\tau_k}(\omega) \to X_\tau(\omega)$ holds for a.e. ω and since pointwise limits of measurable functions are measurable, we are done.

(iii) Lévy processes have independent and stationary increments thus it suffices to show the statement for $T_\varepsilon^{(1)}(\mathbf{X})$ (observe that $T_\varepsilon^{(2)}(\mathbf{X}) \overset{d}{=} T_\varepsilon^{(1)}(\mathbf{X}) + T_\varepsilon^{(1)}(X_{T_\varepsilon^{(1)}(\mathbf{X})} + \mathbf{X} \circ \theta_{T_\varepsilon^{(1)}(\mathbf{X})})$ for the shift operator $\theta_t((X_s)_{s\geq 0}) := (X_{s+t})_{s\geq 0})$.
That $T_\varepsilon^{(1)}(\mathbf{X})$ is measurable w.r.t. $\mathcal{F}_{t+}(\mathbf{X}) = \mathcal{F}_t(\mathbf{X})$ follows from (6.7) since

$$\{T_\varepsilon^{(1)}(\mathbf{X}) \leq t\} = \bigcup_{m\geq 1} \bigcap_{n\geq m} \bigcup_{\substack{p_n, q_n \in \mathbb{Q} \\ |q_n - p_n| \leq \frac{2}{n} \\ p_n < q_n < t + \frac{1}{n}}} \{|X_{q_n} - X_{p_n}| > \varepsilon\}$$

for all $t > 0$.

(iv) This is a straight-forward consequence of the preceding results, since

$$\{|\Delta X_t| = \varepsilon \text{ for some } t \geq 0\} = \bigcup_{i\geq 0} \big\{\varepsilon > 0 : \mathbb{P}\big[\big|\Delta X_{T_{1/n}^{(i)}(\mathbf{X})}\big| = \varepsilon, T_{1/n}^{(i)}(\mathbf{X}) < \infty\big] > 0\big\}$$

is measurable. □

Proposition 6.6. *Let $\mathbf{X}$ be a Lévy process. Then $U(\mathbf{X})$ is at most countable.*

Proof. Cf. [9], Lemma 3.12, p. 149.
The claim follows from

$$\begin{aligned} U(\mathbf{X}) &= \bigcup_{i,n\geq 1} \big\{\varepsilon > 0 : \mathbb{P}\big[\big|\Delta X_{T_{1/n}^{(i)}(\mathbf{X})}\big| = \varepsilon, T_{1/n}^{(i)}(\mathbf{X}) < \infty\big] > 0\big\} \\ &\subseteq \bigcup_{i,n\geq 1} \{\varepsilon > 0 : \mathbb{P}[|Z_{n,i}| = \varepsilon] > 0\}, \end{aligned}$$

where $|Z_{n,i}| := |\Delta X_{T_{1/n}^{(i)}(\mathbf{X})}|$ is a random variable (see Proposition 6.5). This is a countable union of at most countable sets by Lemma 6.1. □

6.3. Building Bridges between $\mathcal{D}$ and $\mathcal{M}_p(E)$

We are getting closer to the essence of this text, the connection between càdlàg functions and measures, or weak convergence to Lévy processes and to PRM respectively. For the latter the Continuous Mapping Theorem is an indispensable tool, so we should rather find some nice continuous transformations between $\mathcal{D}$ and $\mathcal{M}(E)$ and vice versa. The following two lemmata provide such "bridges" between those two spaces.

Lemma 6.7 (Continuity of $F_{T,\varepsilon}$). *For all $\varepsilon > 0$ and $T \geq 0$ let $E_{T,\varepsilon} := [0, T] \times \{x : |x| > \varepsilon\} \subseteq [0, \infty) \times \mathbb{R}_0^d$. Then the map*

$$F_{T,\varepsilon} : \mathcal{D} \to \mathcal{M}_p(E_{T,\varepsilon}), \qquad \mathbf{x} \mapsto \sum_{t\leq T : |\Delta\mathbf{x}(t)| > \varepsilon} \delta_{(t,\Delta\mathbf{x}(t))}$$

is continuous at all points $\mathbf{x}$ *s.t.* $\varepsilon \notin U(\mathbf{x})$ *and* $|\Delta\mathbf{x}(T)| < \varepsilon$.
In particular for a Lévy process $\mathbf{X}$ *and* $\varepsilon \notin U(\mathbf{X})$ *this map is a.s. continuous w.r.t. to its distribution* $\mathbb{P}_{\mathbf{X}}$.

Proof. Fix $\mathbf{x} \in \mathcal{D}$, $\varepsilon > 0$ and $T > 0$ s.t $\varepsilon \notin U(\mathbf{x})$ and $|\Delta\mathbf{x}(T)| < \varepsilon$. We have to prove that $F_{T,\varepsilon}(\mathbf{x}_n) \stackrel{v}{\to} F_{T,\varepsilon}(\mathbf{x})$ for any sequence $(\mathbf{x}_n)$ s.t. $\mathbf{x}_n \rightsquigarrow \mathbf{x}$.

(i) First we reduce the problem by using already proven results. The continuity statement in Corollary 2.37 implies

$$\sum_{t\geq 0:|\Delta\mathbf{x}_n(t)|>\varepsilon} \Delta\mathbf{x}_n(t)\mathbb{1}_{[t,\infty)} = \sum_{i\geq 0} \Delta\mathbf{x}_n(t_{\varepsilon,n}^{(i)})\mathbb{1}_{\left[t_{\varepsilon,n}^{(i)},\infty\right)} = \mathbf{x}^{@\mathcal{T}_n^\varepsilon} \stackrel{\text{Cor. 2.37}}{\rightsquigarrow} \mathbf{x}^{@\mathcal{T}^\varepsilon}$$
$$= \sum_{i\geq 0} \Delta\mathbf{x}(t_{\varepsilon}^{(i)})\mathbb{1}_{\left[t_{\varepsilon}^{(i)},\infty\right)} = \sum_{t\geq 0:|\Delta\mathbf{x}(t)|>\varepsilon} \Delta\mathbf{x}(t)\mathbb{1}_{[t,\infty)}.$$

Moreover, we assumed that $|\Delta\mathbf{x}(T)| < \varepsilon$ hence $\mathbf{x}^{@\mathcal{T}^\varepsilon}$ is continuous at T. We we can thus use Lemma 2.19 to conclude that $Rest_T\big(\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon}\big) \rightsquigarrow Rest_T\big(\mathbf{x}^{@\mathcal{T}^\varepsilon}\big)$ for $Rest_T : \mathcal{D} \to \mathcal{D}_T$, $\mathbf{x} \mapsto \mathbf{x}\big|_{[0,T]}$.
This greatly reduced the remaining work. For the rest define

$$\mathcal{S}_{T,\varepsilon} := \{\mathbf{x} \in \mathcal{D}_T : \mathbf{x}(s) = \sum_{i=1}^{m} \Delta\mathbf{x}(t_i)\mathbb{1}_{\{t_i\leq s\}} \text{ for an ordered set } 0 \leq t_1 < t_2 < \cdots < t_m$$
$$\text{s.t. } |\Delta\mathbf{x}(t_i)| > \varepsilon, 1 \leq i \leq m, m \geq 1\}.$$

This is the subspace of jump functions in $\mathcal{D}_T$ having discontinuities of size $> \varepsilon$. Note that $Rest_T(\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon}) \in \mathcal{S}_{T,\varepsilon}$ for all n, $Rest_T(\mathbf{x}^{@\mathcal{T}^\varepsilon}) \in \mathcal{S}_{T,\varepsilon}$ and $F_{T,\varepsilon}(\mathbf{x}) = \tilde{F}_{T,\varepsilon}(Rest_T(\mathbf{x}^{@\mathcal{T}^\varepsilon}))$ for

$$\tilde{F}_{T,\varepsilon} : \mathcal{S}_{T,\varepsilon} \to \mathcal{M}_p(E_{T,\varepsilon}), \qquad \mathbf{x} \mapsto \sum_{i=1}^{m} \delta_{(t_i,\Delta\mathbf{x}(t_i))}.$$

Ergo it suffices to show that this composition is continuous. This is easier than proving continuity of $F_{T,\varepsilon}$ directly, since $\tilde{F}_{T,\varepsilon}$ operates on a far simpler space and we already know much about the behaviour of $Rest_T(\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon})$ and $Rest_T(\mathbf{x}^{@\mathcal{T}^\varepsilon})$.

(ii) Let $\mathbf{x}_n \rightsquigarrow \mathbf{x}$ via (λ_n), then the proof of Corollary 2.37 shows that there is some m s.t. $\mathcal{T}^\varepsilon \cap [0,T] = \{t_\varepsilon^{(i)}, 1 \leq i \leq m\}$ and $\mathcal{T}_n^\varepsilon \cap [0,T] = \{\lambda_n(t_\varepsilon^{(i)}), 1 \leq i \leq m\}$ for all large n (note that $\Delta\mathbf{x}(T) = 0$ thus $t_\varepsilon^{(i)} \neq T$ for all i), i.e.

$$\tilde{\mathbf{x}}_n := Rest_T\big(\mathbf{x}_n^{@\mathcal{T}_n^\varepsilon}\big) = \sum_{i=1}^{m} \Delta(\mathbf{x}_n \circ \lambda_n)(t_\varepsilon^{(i)})\mathbb{1}_{\{[\lambda_n(t_\varepsilon^{(i)}),\infty)\}} \text{ and}$$
$$\tilde{\mathbf{x}} := Rest_T\big(\mathbf{x}^{@\mathcal{T}^\varepsilon}\big) = \sum_{i=1}^{m} \Delta\mathbf{x}(t_\varepsilon^{(i)})\mathbb{1}_{\{[(t_\varepsilon^{(i)}),\infty)\}}.$$

For those n and for any $f \in \mathcal{C}_c^+(E_{T,\varepsilon})$

$$\tilde{F}_{T,\varepsilon}(\tilde{\mathbf{x}}_n)(f) = \int_{E_{T,\varepsilon}} f\mathrm{d}\left(\sum_{i=1}^{m} \delta_{\left(\lambda_n(t_\varepsilon^{(i)}),\Delta\mathbf{x}_n(\lambda_n(t_\varepsilon^{(i)})\right)}\right) = \sum_{i=1}^{m} f\big(\lambda_n(t_\varepsilon^{(i)}), \Delta\mathbf{x}_n(\lambda_n(t_\varepsilon^{(i)}))\big)$$
$$\to \sum_{k=1}^{m} f\big(t_\varepsilon^{(i)}, \Delta\mathbf{x}(t_\varepsilon^{(i)})\big) = \tilde{F}_{T,\varepsilon}(\tilde{\mathbf{x}})(f),$$

since f was assumed to be continuous and $\big(\lambda_n(t_\varepsilon^{(i)}), \Delta\mathbf{x}(\lambda_n(t_\varepsilon^{(i)}))\big) \to \big(t_\varepsilon^{(i)}, \Delta\mathbf{x}(t_\varepsilon^{(i)})\big)$ for all $1 \le i \le m$ as $n \to \infty$. Hence $\tilde{F}_{T,\varepsilon}(\tilde{\mathbf{x}}_n) \xrightarrow{v} \tilde{F}_{T,\varepsilon}(\tilde{\mathbf{x}})$ and therefore the map $F_{T,\varepsilon}$ is continuous at $\mathbf{x}$.

(iii) Let $\mathbf{X}$ be a Lévy process. For $\varepsilon \notin U(\mathbf{X})$ we need to prove that $\mathbb{P}_{\mathbf{X}}[\{\mathbf{x} \in \mathcal{D} : \varepsilon \in U(\mathbf{x})\} \cup \{\mathbf{x} \in \mathcal{D} : |\Delta\mathbf{x}(T)| > 0\}] = 0$. By Corollary 5.7 we know that $\mathbb{P}_{\mathbf{X}}[\{\mathbf{x} \in \mathcal{D} : |\Delta\mathbf{x}(T)| > 0\}] = 0$ and it remains to show that $\mathbb{P}[\mathbf{X} \in \{\mathbf{x} \in \mathcal{D} : \varepsilon \in U(\mathbf{x})\}] = 0$. This is a consequence of

$$\begin{aligned}\mathbb{P}[\mathbf{X} \in \{\mathbf{x} \in \mathcal{D} : \varepsilon \in U(\mathbf{x})\}] &= \mathbb{P}[\varepsilon \in \{u > 0 : |\Delta X_t| = u \text{ for some } t \ge 0\}] \\ &= \mathbb{P}[|\Delta X_t| = \varepsilon \text{ for some } t \ge 0] = 0,\end{aligned}$$

since $\varepsilon \notin U(\mathbf{X})$. Note that this also shows measurability of the involved terms by Proposition 6.5. $\square$

To prove the next continuity result, we follow [16], Section 7.2.3.

Definition 6.8. Let $\Gamma := \{\nu \in \mathcal{M}_p([0,\infty) \times \overline{\mathbb{R}}_0^d) : \nu\big(\partial([0,\infty) \times \{y : |y| > \varepsilon\})\big) = 0, \nu(\{t\} \times \overline{\mathbb{R}}_0^d) \in \{0,1\} \text{ for all } t \ge 0\}$.

Remark 6.9. Alternatively, we can write $\Gamma = \{\nu \in \mathcal{M}_p([0,\infty)\times\overline{\mathbb{R}}_0^d) : \nu\big(\partial([0,\infty)\times\{y : |y| > \varepsilon\})\big) = 0$ and at most one unit-atom of ν lies on any "vertical" hyperplane$\}$. For $d = 1$ the last conditions means, that ν has at most one unit-atom on any vertical line in $[0,\infty) \times \overline{\mathbb{R}}$, that is a line of the form $\{t\} \times \overline{\mathbb{R}}$ for some t.

Lemma 6.10 (Continuity of G_ε). *The map*

$$\begin{aligned}&G_\varepsilon : \mathcal{M}_p([0,\infty) \times \overline{\mathbb{R}}_0^d) \to \mathcal{D}, \\ &\qquad \nu := \sum_{k\ge1} \delta_{(s_k,y_k)} \mapsto \Big(\int_{[0,t]\times\{y:|y|>\varepsilon\}} x \mathrm{d}\nu(s,y)\Big)_{t\ge0} = \Big(\sum_{s_k\le t} y_k \mathbb{1}_{\{|y_k|>\varepsilon\}}\Big)_{t\ge0}\end{aligned}$$

is continuous at all $\nu \in \Gamma$.
In particular G_ε is a.s. continuous w.r.t. the distribution of $\mathbf{N} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \mu)$ for $\mu \in \mathcal{M}(\overline{\mathbb{R}}_0^d)$ s.t. $\mu(\overline{\mathbb{R}}_0^d \setminus \mathbb{R}_0^d) = 0$ and $\mu(\partial B_\varepsilon(0)) = 0$.

Proof. Cf. [16], Section 7.2.3.
Since $[0,t] \times \{y \in \overline{\mathbb{R}}_0^d : |y| > \varepsilon\}$ is a relatively compact subset of $[0,\infty) \times \overline{\mathbb{R}}_0^d$, the above sum on the right-hand side is finite for all $t \ge 0$ and G_ε is thus well-defined.

(i) Let $\nu \in \Gamma$ and (ν_n) be a sequence in $\mathcal{M}_p([0,\infty) \times \overline{\mathbb{R}}_0^d)$ s.t. $\nu_n \xrightarrow{v} \nu$. If we achieve $G_\varepsilon(\nu_n) \rightsquigarrow G_\varepsilon(\nu)$ in $\mathcal{D}_T$ for almost all T, the first statement is obtained by Corollary 2.22 or Theorem 2.27. Fix $T \ge 0$ s.t. $|\Delta G_\varepsilon(\nu)(T)| = 0$. Since $G_\varepsilon(\nu) \in \mathcal{D}$, any but countably many $T \in [0,\infty)$ are permitted.
The set $\overline{E_{T,\varepsilon}}$ is compact and $\nu(\partial E_{T,\varepsilon}) = 0$, because $\nu \in \Gamma$. Now there are numbers m and n_0 s.t. $\nu_n(\overline{E_{T,\varepsilon}}) = \nu_n(E_{T,\varepsilon}) = \nu(E_{T,\varepsilon}) = \nu(\overline{E_{T,\varepsilon}}) = m$ for all $n \ge n_0$ by Proposition 4.25. This means that ν_n and ν finally have the same number of points getting unit-mass in $E_{T,\varepsilon}$. Moreover,

$$\nu_n\big|_{E_{T,\varepsilon}} = \sum_{k=1}^m \delta_{\left(s_k^{(n)}, y_k^{(n)}\right)} \qquad \text{and} \qquad \nu\big|_{E_{T,\varepsilon}} = \sum_{k=1}^m \delta_{(s_k,y_k)}, \tag{6.8}$$

for some $s_k, s_k^{(1)}, \ldots, s_k^{(m)}$ in $[0,T]$, $y_k, y_k^{(1)}, \ldots, y_k^{(1)}$ in $B_\varepsilon[0]^c$, $1 \le k \le m$. In particular, we know that $s_0 := 0 < s_1 < \cdots < s_m < s_{m+1} := T$, since $\nu \in \Gamma$, hence non of the atoms can either occur at the same time or on a border. Let $\varepsilon > 0$ and choose $\delta > 0$ so small that $s_{k+1} - s_k \ge 2\delta$ for $0 \le k \le m$. Wlog we can assume that $m\delta \le \varepsilon$. The second part of Proposition 4.25 implies that (given proper labeling) there is some $n_1 \ge n_0$ s.t. $\|(s_k^{(n)}, y_k^{(n)}) - (s_k, y_k)\|_\infty \le \delta$ for all $n \ge n_1$. All $(s_k^{(n)}, y_k^{(n)})$ are thus contained in δ-boxes around (s_k, y_k), which especially implies that all $(s_k^{(n)}, y_k^{(n)})$ are distinct for all $0 \le k \le m+1$ given $n \ge n_1$.

(ii) To ensure convergence in the Skorokhod topology pick a sequence of time-shifts (λ_n) in Λ_T mapping the atoms of ν_n at time $s_k^{(n)}$ onto the atoms of ν at time s_k. Therefore we fix n and let $\lambda_n \in \Lambda_T$ be given by a linear interpolation between $\lambda_n(0) = 0, \lambda_n(T) = T$ and $\lambda_n(s_k) = s_k^{(n)}$ for all $0 \le k \le m+1$. Then

$$\begin{aligned}
\|G_\varepsilon(\nu_n) \circ \lambda_n - G_\varepsilon(\nu)\|_{T,\infty} &= \|G_\varepsilon\big(\nu_n\big|_{E_{T,\varepsilon}}\big) \circ \lambda_n - G_\varepsilon\big(\nu\big|_{E_{T,\varepsilon}}\big)\|_{T,\infty} \\
&= \sup_{t\in[0,T]} \Big| \sum_{s_k^{(n)} \le \lambda_n(t)} y_k^{(n)} - \sum_{s_k \le t} y_k \Big| \\
&= \sup_{t\in[0,T]} \Big| \sum_{\lambda_n^{-1}(s_k^{(n)}) \le t} y_k^{(n)} - \sum_{s_k \le t} y_k \Big| \\
&= \sup_{t\in[0,T]} \Big| \sum_{s_k \le t} \big(y_k^{(n)} - y_k\big) \Big| \le \sum_{s_k \le T} \|y_k^{(n)} - y_k\|_{T,\infty} \le m\delta.
\end{aligned}$$

Finally, at every point s_k where we interpolate, we make an error of at most δ each time, thus $\|\lambda_n - \mathrm{Id}\,\|_\infty \le m\delta$. This implies that $d_T(G_\varepsilon(\nu_n), G_\varepsilon(\nu)) \to 0$, i.e $G_\varepsilon(\nu_n) \rightsquigarrow G_\varepsilon(\nu)$ in $\mathcal{D}_T$.

(iii) It remains to prove that $\mathbb{P}_\mathbf{N}[\Gamma] = 1$ for $\mathbf{N} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \mu)$. Let $\tilde{\mu} := \lambda \otimes \mu$ be the intensity of $\mathbf{N}$. The definition of the product measure implies that on the one hand $\tilde{\mu}(\{s\} \times \{y : |y| > \varepsilon\}) = \lambda(\{s\})\mu(\{y : |y| > \varepsilon\}) = 0$ for $s \in \{0, \pm\infty\}$ and on the other hand $\tilde{\mu}([0,\infty) \times \{y : |y| = z\}) = \lambda([0,\infty))\mu(\{y : |y| = z\}) = 0$, for $z \in \{\varepsilon, \infty\}$ by assumption. Thus $\tilde{\mu}(\partial([0,\infty) \times \{x : |x| > \varepsilon\})) = 0$ and consequently $\mathbb{P}_\mathbf{N}[\{\nu \in \mathcal{M}_p([0,\infty) \times \overline{\mathbb{R}}_0^d) : \nu(\partial([0,\infty) \times \{x : |x| > \varepsilon\})) = 0\}] = 1$ by Proposition 3.21. Since PRM have a.s. no double jumps by Proposition 4.12 we are done. □

Now we can progress to the stochastic level. Intuitively we want to assign a point process to a given Lévy process by collecting the times and the corresponding heights of its jumps.

Definition 6.11. Let $\mathbf{X}$ be a stochastic process having a.s. càdlàg paths. Then

$$\mathbf{N_X} := \sum_{t \ge 0 : |\Delta X_t| > 0} \delta_{(t, \Delta X_t)}.$$

is called ASSOCIATED JUMP PROCESS OF $\mathbf{X}$.

Before we prove that $\mathbf{N_X}$ actually is a point process let us get some more intuition. Assume $d = 1$ for the moment. If the path $\mathbf{X}(\omega)$ has a discontinuity at time t, then the graph shows some jump. By putting unit mass onto the time and the size (including the sign) of the jump, we collect all necessary information in a measure $\mathbf{N_X}(\omega)$ on $[0,\infty)\times\mathbb{R}$. An example for a jump function is given in the following figure.

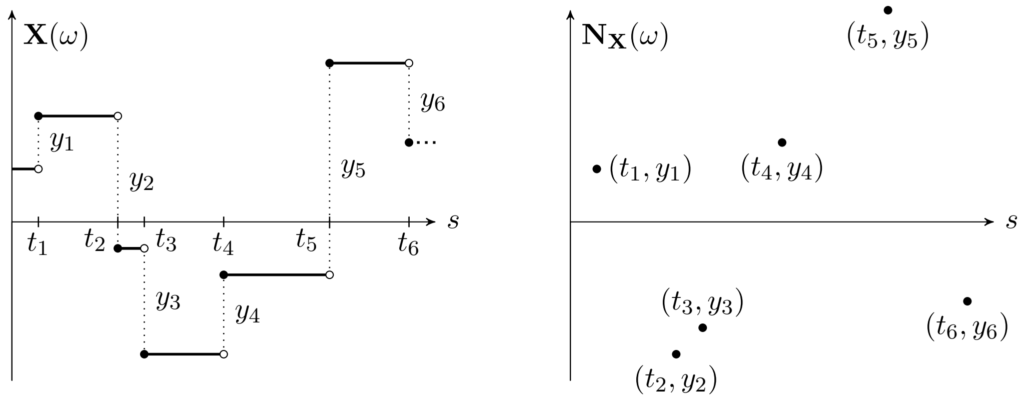

Figure 6.1: A path of a simple càdlàg process and a realization of its associated jump process

Recalling the construction of Lévy processes for a given canonical triple (Theorem 5.8), we see that a Lévy process without Gaussian part is "constant between jumps". We may thus assume that we can very well describe this process in terms of the information contained in its associated jump process.

Proposition 6.12. *Let* $\mathbf{X}$ *be a stochastic process having a.s. càdlàg paths. Then* $\mathbf{N_X}$ *is a point process on* $E=[0,\infty)\times\overline{\mathbb{R}}_0^d$.

Proof. We have to show that $\mathbf{N_X}:(\Omega,\mathcal{A})\to(\mathcal{M}_p(E),\mathscr{M}_p(E))$ is measurable. Observe that $\mathbf{N_X}=\Xi\circ\mathbf{X}$ for

$$\Xi:\mathcal{D}\to\mathcal{M}_p(E),\qquad \mathbf{x}\to\sum_{t\geq 0:|\Delta\mathbf{x}(t)|>0}\delta_{(t,\Delta\mathbf{x}(t))}. \tag{6.9}$$

This map is well-defined, because càdlàg functions have at most countably many jumps. Since $\mathbf{X}:(\Omega,\mathcal{A})\to(\mathcal{D},\mathcal{B}_{\mathcal{D}})$ is a stochastic process, it is measurable. Consequently it suffices to show measurability of $\Xi:(\mathcal{D},\mathcal{B}_{\mathcal{D}})\to(\mathcal{M}_p(E),\mathscr{M}_p(E))$.
Note that $\mathscr{M}_p(E)=\mathcal{B}_{\mathcal{M}_p(E)}$ by Lemma 4.23, so the map $F_{T,\varepsilon}$ is measurable due to Lemma 6.7 for suitable $T,\varepsilon>0$. The aim is to write Ξ as a pointwise limit of such functions, to conclude its measurability. The fact that pointwise limits of measurable functions are again measurable – which is well-known if the range is $\mathbb{R}$ – also holds in Polish spaces, see [1], Lemma 4.29.
Fix $\mathbf{x}\in\mathcal{D}$ and pick some sequences (T_n) and (ε_n) satisfying $T_n\to\infty$ and $\varepsilon_n\to 0$ s.t. $\Delta\mathbf{x}(T_n)=0$ and $\varepsilon_n\notin U(\mathbf{x})$ for all n. This is possible, since the set where those conditions are not satisfied are at most countable (see Corollary 2.6 and Remark 2.36). We claim that in this case $F_{T_n,\varepsilon_n}(\mathbf{x})\xrightarrow{v}\Xi(\mathbf{x})$ for all $\mathbf{x}$, what consequently implies measurability of Ξ and thus the result. This is now easy, since any $f\in\mathcal{C}_c^+(E)$ has compact support, hence there are some $\tilde{T},\tilde{\varepsilon}>0$ s.t. $supp(f)\subseteq[0,\tilde{T}]\times\{y\in\overline{\mathbb{R}}_0^d:|y|\geq\tilde{\varepsilon}\}$ (see Proposition E.2). Consequently $\Xi(\mathbf{x})(f)=F_{T_n,\varepsilon_n}(\mathbf{x})(f)$ for sufficiently large n. $\square$

Lemma 6.13. *Let* $\mathbf{X}$ *be a Lévy process with canonical triple* $(0, 0, \pi)$*. Then* $\mathbf{N_X} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \pi)$*.*

Proof. Let $\mathbf{N} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \pi)$. By the Lévy-Itō decomposition (Theorem 5.20) and the proof of Theorem 5.18, there is a $\mathcal{L}^2$-martingale $\mathbf{M} = \sum_{n\geq 1} \mathbf{Z}^{(n)}$ for $\mathbf{Z}^{(n)}$ given by (5.11) and a Compound Poisson process $\mathbf{Z}$ given by (5.9), s.t.

$$\mathbf{X} \overset{d}{=} \tilde{\mathbf{X}} := \mathbf{M} + \mathbf{Z} = \Big(\int_{[0,t]\times\mathbb{R}_0^d} y \mathrm{d}\mathbf{N}(s, y) \Big)_{t\geq 0}.$$

The following equalities hold a.s. for $x \neq 0$

$$\begin{aligned} \Delta\tilde{X}_t = x &\iff \Delta(M_t + Z_t) = x \iff \Delta\Big(\int_{[0,t]\times\mathbb{R}_0^d} y \mathrm{d}\mathbf{N}(s, y) \Big) = x \\ &\overset{\text{Prop. 4.12}}{\iff} \mathbf{N}(\{(t, x)\}) = 1. \end{aligned}$$

Here we used in the next to the last step that $\mathbf{N}$ has a.s. no double jumps. Consequently $\{(t, x) \in [0, \infty) \times \mathbb{R}_0^d : \mathbf{N}(t, x) = 1\} = \{(t, x) \in [0, \infty) \times \mathbb{R}_0^d : \Delta\tilde{X}_t = x \neq 0\}$. Thus $\mathbf{N} = \mathbf{N}_{\tilde{\mathbf{X}}} = \sum_{t\geq 0:\Delta\tilde{X}_t\neq 0} \delta_{(t,\Delta\tilde{X}_t)}$ and

$$\mathbf{N_X} \overset{d}{=} \mathbf{N}_{\tilde{\mathbf{X}}} = \mathbf{N} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \pi),$$

hence $\mathbf{N_X} \sim PRM(\lambda\big|_{[0,\infty)} \otimes \pi)$ as desired. Note that $\mathbf{N_X} \overset{d}{=} \mathbf{N}_{\tilde{\mathbf{X}}}$ is a consequence of $\mathbf{X} \overset{d}{=} \tilde{\mathbf{X}}$, since

$$\mathbb{P}_{\mathbf{N_X}} = \mathbb{P} \circ \mathbf{N_X}^{-1} = \mathbb{P} \circ (\Xi \circ \mathbf{X})^{-1} = \mathbb{P} \circ \mathbf{X}^{-1} \circ \Xi^{-1} = \mathbb{P}_{\mathbf{X}} \circ \Xi^{-1} = \mathbb{P}_{\tilde{\mathbf{X}}} \circ \Xi^{-1} = \mathbb{P}_{\mathbf{N}_{\tilde{\mathbf{X}}}}$$

using the map Ξ from (6.9). □

6.4. The Theorem

Definition 6.14. A $\mathbb{R}^d$-valued random variable is NON-DEGENERATE if $\mathbb{P}[0 < |X| < \infty] = 1$.

Theorem 6.15. *Let* $\mathbf{X} = (X_t)_{t\geq 0}$ *be a Lévy process with canonical triple* $(0, 0, \pi)$ *and* $\mathbf{N} \sim PRM\big(\lambda\big|_{[0,\infty)} \otimes \pi\big)$*. Let* $(\mathbf{X}^{(n)})$ *be the sequence of partial sum processes given via*

$$X_t^{(n)} = \sum_{1\leq k\leq nt} X_{k,n} - tc_n, \; t \geq 0 \tag{6.10}$$

for non-degenerate $(X_{k,n})$ *and let* $(\mathbf{N}^{(n)})$ *be a sequence of point processes given by*

$$\mathbf{N}^{(n)} := \mathbf{N}_{\mathbf{X}^{(n)}} = \sum_{k\geq 1} \delta_{(\frac{k}{n}, X_{k,n})}. \tag{6.11}$$

Then $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ *in* $\mathcal{D}$ *iff the following two conditions hold*

(i) $\mathbf{N}^{(n)} \Rightarrow \mathbf{N}$ *in* $\mathcal{M}_p([0, \infty) \times \overline{\mathbb{R}}_0^d)$

(ii) $\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\mathbb{P}\Big[\Big\|\sum_{k\le nt}X_{k,n}\mathbb{1}_{\{|X_{k,n}|\le\varepsilon\}}-t\Big(c_n-\int_{\{x:\varepsilon<|x|\le 1\}}x\mathrm{d}\pi(x)\Big)\Big\|_{T,\infty}\ge\delta\Big]=0$
for every $\delta>0$ *and* $T>0$, *where the limit is taken over all* $\varepsilon\notin U(\mathbf{X})$.

Before we actually prove this theorem, let us get some intuition on how to interprete the regularity condition (ii) above.

Remark 6.16. Recalling Section 2.2. we observe that it is always possible to uniformly "compensate" small jumps of a càdlàg function with a continuous function. The integral in the above condition is now the attempt to compensate the drift of the small jumps of the partial sum process. If condition (ii) fails, then the a.s. uniform compensation fails and we cannot expect the limit to be a.s. càdlàg.

Proof. Cf. [20], Theorem 3.1.
Note that $\mathbf{N}\overset{d}{=}\mathbf{N_X}$ by Lemma 6.13, hence we can wlog assume $\mathbf{N}=\mathbf{N_X}$, since the condition (i) is equivalent for both.

($\Leftarrow$) (i) The Lévy-Itō decomposition implies that there is a Compound Poisson process $\mathbf{Z}=(Z_t)_{t\ge0}$, $Z_t:=\int_{[0,t]\times\{y:|y|>1\}}y\mathrm{d}\mathbf{N}(s,y)$ and a $\mathcal{L}^2$-martingale $\mathbf{M}$ s.t. $\mathbf{X}=\mathbf{M}+\mathbf{Z}$. For $\varepsilon>0$ define $\mathbf{M}^{(\varepsilon)}$ as in (5.16) and fix some $T>0$. Then $\|\mathbf{M}^{(\varepsilon)}-\mathbf{M}\|_{T,\infty}\overset{a.s.}{\to}0$ by Theorem 5.21 and thus

$$\|\mathbf{Z}+\mathbf{M}^{(\varepsilon)}-\mathbf{X}\|_{T,\infty}\overset{a.s.}{\to}0\text{ as }\varepsilon\searrow 0.\tag{6.12}$$

Almost sure uniform convergence on compact sets implies almost sure convergence in the Skorokhod topology. Evidently this implies convergence in distribution on $\mathcal{D}_T$ for all T and thus

$$\mathbf{M}^{(\varepsilon)}+\mathbf{Z}\Rightarrow\mathbf{X}\text{ in }\mathcal{D}\text{ as }\varepsilon\searrow 0.\tag{6.13}$$

by Lemma 5.22.

(ii) Define $\mathbf{X}^{(n,\varepsilon)}:=(X_t^{(n,\varepsilon)})_{t\ge0}$, via $X_t^{(n,\varepsilon)}:=\sum_{k\le nt}X_{k,n}\mathbb{1}_{\{|X_{k,n}|>\varepsilon\}}$. We claim that $\mathbf{X}^{(n,\varepsilon)}=G_\varepsilon(\mathbf{N}^{(n)})$. To verify this, note that for each $t\ge0$

$$\begin{aligned}(G_\varepsilon(\mathbf{N}^{(n)}))_t&=\int_{[0,t]\times\{y:|y|>\varepsilon\}}y\mathrm{d}\Big(\sum_{j\ge1}\delta_{\left(\frac{k}{n},X_{k,n}\right)}\Big)=\int_{\{y:|y|>\varepsilon\}}y\mathrm{d}\Big(\sum_{k=1}^{nt}\delta_{X_{k,n}}\Big)\\&=\sum_{k\le nt}X_{k,n}\mathbb{1}_{\{|X_{k,n}|>\varepsilon\}}=X_t^{(n,\varepsilon)}.\end{aligned}$$

Define $\mathbf{X}^{(\varepsilon)}:=G_\varepsilon(\mathbf{N})=\int_{[0,t]\times\{y:|y|>\varepsilon\}}y\mathrm{d}\mathbf{N}(s,y)$. Since G_ε is continuous w.r.t. to $\mathbb{P}_\mathbf{N}$ by Lemma 6.10, we can apply the Continuous Mapping Theorem to get

$$\mathbf{X}^{(n,\varepsilon)}=G_\varepsilon(\mathbf{N}^{(n)})\Rightarrow G_\varepsilon(\mathbf{N})=\mathbf{X}^{(\varepsilon)}\text{ in }\mathcal{D},\tag{6.14}$$

since $\mathbf{N}^{(n)}\Rightarrow\mathbf{N}$ by assumption.

(iii) Define $\tilde{\mathbf{X}}^{(n,\varepsilon)} := (\tilde{X}_t^{(n,\varepsilon)})_{t\geq 0}$ via $\tilde{X}_t^{(n,\varepsilon)} := X_t^{(n,\varepsilon)} - t\int_{\{y:|y|>\varepsilon\}} y\mathrm{d}\pi(y)$. The deterministic drift doesn't spoil weak convergence, hence

$$\begin{aligned}\tilde{\mathbf{X}}^{(n,\varepsilon)} &= \mathbf{X}^{(n,\varepsilon)} - \int_{[0,\,\cdot\,]\times\{y:|y|>\varepsilon\}} y\mathrm{d}(\lambda\otimes\pi)(y) \overset{(6.14)}{\Rightarrow} \mathbf{X}^{(\varepsilon)} - \int_{[0,\,\cdot\,]\times\{y:|y|>\varepsilon\}} y\mathrm{d}(\lambda\otimes\pi)(y) = \\ &= \underbrace{\mathbf{X}^{(1)}}_{=\mathbf{Z}} + \underbrace{\Big(\mathbf{X}^{(\varepsilon)} - \mathbf{X}^{(1)} - \int_{[0,\,\cdot\,]\times\{y:|y|>\varepsilon\}} y\mathrm{d}(\lambda\otimes\pi)(y)\Big)}_{=\mathbf{M}^{(\varepsilon)}} \\ &= \mathbf{Z} + \mathbf{M}^{(\varepsilon)} \text{ in } \mathcal{D}.\end{aligned}$$

Now we are in the situation that $\tilde{\mathbf{X}}^{(n,\varepsilon)} \Rightarrow \mathbf{Z}+\mathbf{M}^{(\varepsilon)}$ as $n\to\infty$, $\mathbf{Z}+\mathbf{M}^{(\varepsilon)} \Rightarrow \mathbf{X}$ as $\varepsilon\to 0$ and both convergence results hold in $\mathcal{D}$. We apply the Converging Together Theorem (Theorem B.6), which provides a condition for $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ on $\mathcal{D}_T$ in terms of $(\mathbf{X}^{(n,\varepsilon)})$ in this case. Therefore pick (ε_m) s.t. $\varepsilon_m\in(0,1)$, $\pi(\{x:|x|=\varepsilon_m\})=0$ for all m and $\varepsilon_m \searrow 0$ (this is possible by Lemma 6.1). It remains to show (B1) for $X_{m,n} = \tilde{\mathbf{X}}^{(n,\varepsilon_m)}$ and $Z_m := \mathbf{Z}+\mathbf{M}^{(\varepsilon_m)}$, that is

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}\big[d_T^o\big(\mathbf{X}^{(n)},\tilde{\mathbf{X}}^{(n,\varepsilon_m)}\big)\geq\delta\big] = 0$$

for any $\delta>0$.
Wlog we have $d_T^o\big(\mathbf{X}^{(n)},\tilde{\mathbf{X}}^{(n,\varepsilon_m)}\big) \leq \|\mathbf{X}^{(n)}-\tilde{\mathbf{X}}^{(n,\varepsilon_m)}\|_{T,\infty}$ since this estimate is evidently true for the metric d_T we constructed in Proposition 2.21, which is equivalent to d_T^o by Theorem 2.27. Hence it suffices to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathbb{P}\big[\|\mathbf{X}^{(n)}-\tilde{\mathbf{X}}^{(n,\varepsilon_m)}\|_{T,\infty}\geq\delta\big] = 0. \tag{6.15}$$

Now

$$\begin{aligned}X_t^{(n)} - \tilde{X}_t^{(n,\varepsilon_m)} &= \sum_{k\leq nt} X_{k,n} - tc_n - \sum_{k\leq nt} X_{k,n}\mathbb{1}_{\{|X_{k,n}|>\varepsilon_m\}} - t\int_{\{x:\varepsilon_m<|y|\leq 1\}} y\mathrm{d}\pi(y) \\ &= \sum_{k\leq nt} X_{k,n}\mathbb{1}_{\{|X_{k,n}|\leq\varepsilon_m\}} - t\Big(c_n - \int_{\{x:\varepsilon_m<|y|\leq 1\}} y\mathrm{d}\pi(y)\Big)\end{aligned}$$

and thus (6.15) is just condition (ii), since $\pi(\{x:|x|=\varepsilon_m\})=0$ implies that $\varepsilon_m\notin U(\mathbf{X})$ for all m. Consequently $\mathbf{X}^{(n)}\Rightarrow\mathbf{X}$ in $\mathcal{D}_T$ for (almost) all $T\geq 0$ and hence $\mathbf{X}^{(n)}\Rightarrow\mathbf{X}$ in $\mathcal{D}$ by Lemma 5.22.

$(\Rightarrow)$ (i) We first show that $\mathbf{N}^{(n)}\Rightarrow\mathbf{N}$. By Proposition 4.36 it suffices to fix $f\in\mathcal{C}_c^+([0,\infty)\times\mathbb{R}^d)$ and show that the corresponding Laplace functionals converge, i.e. $\Psi_{\mathbf{N}^{(n)}}(f)\to\Psi_{\mathbf{N}}(f)$.
For $\varepsilon\notin U(\mathbf{X})$, the map $F_{T,\varepsilon}:\mathcal{D}\to\mathcal{M}_p(E_{T,\varepsilon})$ from Lemma 6.7 is continuous w.r.t. $\mathbb{P}_{\mathbf{X}}$. The Continuous Mapping Theorem implies $F_{T,\varepsilon}(\mathbf{X}^{(n)})\Rightarrow F_{T,\varepsilon}(\mathbf{X})$ in $\mathcal{M}_p(E_{T,\varepsilon})$ and Proposition 4.36 gives

$$\mathbb{E}[e^{-F_{T,\varepsilon}(\mathbf{X}^{(n)})(f)}]\to\mathbb{E}[e^{-F_{T,\varepsilon}(\mathbf{X})(f)}] \text{ for all } f\in\mathcal{C}_c^+(E_{T,\varepsilon}). \tag{6.16}$$

Now for a fixed $f \in \mathcal{C}_c^+([0,\infty) \times \mathbb{R}^d)$ there is an $\varepsilon > 0$ and $T > 0$ s.t. $supp(f) \subseteq E_{T,\varepsilon}$ (see Proposition E.2). Thus

$$\begin{aligned} N^{(n)}(f) &= \int_{supp(f)} f(y)\mathrm{d}\mathbf{N}^{(n)}(y) = \int_{E_{T,\varepsilon}} f(y)\mathrm{d}\mathbf{N}^{(n)}(y) \\ &= \int_{[0,\infty)\times\mathbb{R}^d} f(y)\mathrm{d}\Big(\sum_{k\leq nT:|X_{k,n}|>\varepsilon} \delta_{\left(\frac{k}{n},X_{k,n}\right)}\Big)(y) = F_{T,\varepsilon}(\mathbf{X}^{(n)})(f). \end{aligned} \tag{6.17}$$

Similarly, one verifies $F_{T,\varepsilon}(\mathbf{X})(f) = N(f)$ and we conclude this part by

$$\begin{aligned} \Psi_{\mathbf{N}^{(n)}}(f) &= \mathbb{E}\big[e^{-N^{(n)}(f)}\big] \overset{(6.17)}{=} \mathbb{E}\big[e^{-F_{T,\varepsilon}(\mathbf{X}^{(n)})(f)}\big] \overset{(6.16)}{\to} \mathbb{E}\big[e^{-F_{T,\varepsilon}(\mathbf{X})(f)}\big] = \mathbb{E}\big[e^{-N(f)}\big] \\ &= \Psi_{\mathbf{N}}(f). \end{aligned}$$

(ii) It remains to check that $\mathbf{X}^{(n)} \Rightarrow \mathbf{X}$ in $\mathcal{D}$ implies condition *(ii)*.
Let $\varepsilon > 0$ be such that $\varepsilon \notin U(\mathbf{X})$ and $H_\varepsilon : \mathcal{D} \to \mathcal{D}$, $\mathbf{x} \mapsto \mathbf{x}^{\setminus \mathcal{T}^\varepsilon}$. Corollary 2.37 implies that this map is continuous at all $\mathbf{x} \in \mathcal{D}$ s.t. $\varepsilon \notin U(\mathbf{x})$. A similar argument as in Lemma 6.7 shows that is also a.s. continuous w.r.t. $\mathbb{P}_{\mathbf{X}}$ given $\varepsilon \notin U(\mathbf{X})$ (what is satisfied by assumption). Thus

$$\begin{aligned} \Big(\sum_{j\leq nt} X_{j,n}\mathbb{1}_{\{|X_{j,n}|\leq\varepsilon\}} - tc_n\Big)_{t\geq 0} &= H_\varepsilon(\mathbf{X}^{(n)}) \overset{CMT}{\Rightarrow} H_\varepsilon(\mathbf{X}) \\ &= \Big(X_t - \sum_{s\leq t:|\Delta X_s|>\varepsilon} \Delta X_s \mathbb{1}_{[s,\infty)}(t)\Big)_{t\geq 0}. \end{aligned} \tag{6.18}$$

As in the proof of the first part, adding a deterministic drift doesn't spoil weak convergence. Let $\tilde{H}_\varepsilon(\mathbf{x})(t) = H_\varepsilon(\mathbf{x})(t) + t\int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y)$, then

$$\Big(\sum_{k\leq nt} X_{k,n}\mathbb{1}_{\{|X_{k,n}|\leq\varepsilon\}} - t\Big(c_n - \int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y)\Big)\Big)_{t\geq 0} = \tilde{H}_\varepsilon(\mathbf{X}^{(n)}) \Rightarrow \tilde{H}_\varepsilon(\mathbf{X}) \text{ in } \mathcal{D}.$$

Moreover, recalling the Lévy-Itō decomposition

$$\begin{aligned} (\tilde{H}_\varepsilon(\mathbf{X}))_t &= X_t - \sum_{s\leq t:|\Delta X_s|>\varepsilon} \Delta X_s \mathbb{1}_{[s,\infty)}(t) + t\int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y) \\ &= X_t - \int_{[0,t]\times\{y:|y|>\varepsilon\}} y\mathrm{d}\mathbf{N}(s,y) + t\int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y) \\ &= X_t - \int_{[0,t]\times\{y:|y|>1\}} y\mathrm{d}\mathbf{N}(s,y) \\ &\qquad - \Big(\int_{[0,t]\times\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\mathbf{N}(s,y) - t\int_{\{y:\varepsilon<|y|\leq 1\}} y\mathrm{d}\pi(y)\Big) \\ &= X_t - Z_t - M_t^{(\varepsilon)}, \end{aligned}$$

hence $\tilde{H}_\varepsilon(\mathbf{X}^{(n)}) \Rightarrow \mathbf{X} - \mathbf{Z} - \mathbf{M}^{(\varepsilon)}$ in $\mathcal{D}$.

(iii) The set $F_\delta := \{\mathbf{x} \in \mathcal{D} : \|\mathbf{x}\|_{T,\infty} \geq \delta\}$ is closed in $\mathcal{D}$ (see Proposition E.3) so the

Portmanteau Theorem yields

$$\begin{aligned}\limsup_{n\to\infty}\mathbb{P}\Big[\Big\|\sum_{k\le nt}X_{k,n}\mathbb{1}_{\{|X_{k,n}|\le\varepsilon\}}-t\Big(c_n-\int_{\{y:\varepsilon<|y|\le 1\}}y\mathrm{d}\pi(y)\Big)\Big\|_{T,\infty}\ge\delta\Big]\\ =\limsup_{n\to\infty}\mathbb{P}[\tilde H_\varepsilon(\mathbf{X}^{(n)})\in F_\delta]\overset{\text{Thm. 4.32}}{\le}\mathbb{P}[\tilde H_\varepsilon(\mathbf{X})\in F_\varepsilon]\\ =\mathbb{P}[\|\mathbf{X}-\mathbf{Z}-\mathbf{M}^{(\varepsilon)}\|_{T,\infty}\ge\delta]\end{aligned}\tag{6.19}$$

Finally, the last term in (6.19) converges to 0 as $\varepsilon\searrow 0$, since $\|\mathbf{X}-\mathbf{M}^{(\varepsilon)}-\mathbf{Z}\|_{T,\infty}\overset{a.s.}{\to}0$ by (6.12), what implies condition (ii). □

Appendix

A. Measure Theory

Definition A.1 (λ-system, π-system). Let E be a set. A family $\mathcal{E} \in \mathcal{P}(E)$ is a λ-SYSTEM ON E if it satisfies the following conditions

(i) $E \in \mathcal{E}$

(ii) if $A \in \mathcal{E}$ then $A^c = E \setminus A \in \mathcal{E}$

(iii) if $A_1, A_2, \dots$ in $\mathcal{E}$ are pairwise disjoint, then $\biguplus_{n\geq 1} A_n \in \mathcal{E}$.

If $\mathcal{E}$ satisfies

(i) $\mathcal{E} \neq \emptyset$

(ii) if $A, B \in \mathcal{E}$ then $A \cap B \in \mathcal{E}$.

then it is a π-SYSTEM ON E.

Theorem A.2 (Dynkin's λ-π Theorem). Let $\mathcal{E}$ be a π-system on E and $\mathcal{G}$ a λ-system on E s.t. $\mathcal{E} \subseteq \mathcal{G}$. Then $\sigma(\mathcal{E}) \subseteq \mathcal{G}$.

Proof. Cf. [8], Theorem 6.7 (in German) or [19], Theorem 5.5. □

Proposition A.3. Let $(E, \mathcal{E}), (F_i, \mathcal{F}_i)$ be some measurable spaces, $i \in \mathcal{I}$. Let $\mathcal{F} := \{f_i : (E, \mathcal{E}) \to (F_i, \mathcal{F}_i), i \in \mathcal{I}\}$ and $E' \in \mathcal{E}$. Then

$$\sigma(f, f \in \mathcal{F}) \cap E' = \sigma(f\big|_{E'}, f \in \mathcal{F}).$$

Proof. The definition of the initial σ-algebra implies

$$\begin{aligned}\sigma(f\big|_{E'}, f \in \mathcal{F}) &= \sigma((f_i\big|_{E'})^{-1}(B), B_i \in \mathcal{F}_i, i \in \mathcal{I}) = \sigma(f_i^{-1}(B) \cap E', B_i \in \mathcal{F}_i, i \in \mathcal{I}) \\ &= \sigma(f_i^{-1}(B), B_i \in \mathcal{F}_i, i \in \mathcal{I}) \cap E' = \sigma(f, f \in \mathcal{F}) \cap E'.\end{aligned}$$

as desired. □

Proposition A.4. Let (f_n) be a sequence of functions in $\mathcal{L}^p$ for $p \in [1, \infty)$. If $f_n \stackrel{\mathcal{L}^p}{\to} f$ for some $f \in \mathcal{L}^p$, then there is a subsequence (n_k) s.t. $f_{n_k} \stackrel{a.s.}{\to} f$.

Proof. Cf. [19], Corollary 12.8. □

Theorem A.5 (Uniqueness Theorem for Measures). Let ν_1, ν_2 be two measures on a σ-algebra $\mathcal{G}$ and let $\mathcal{E} \subseteq \mathcal{G}$ be a π-system s.t. $\sigma(\mathcal{E}) = \mathcal{G}$. If ν_1 is σ-finite and $\nu_1(E) = \nu_2(E)$ for all $E \in \mathcal{E}$, then $\nu_1 = \nu_2$.

Proof. Cf. [22], Section 3.5, p. 81ff or [4], Theorem 5.4. □

B. Probability Theory

Theorem B.1 (Jensen's inequality). Let $I \subseteq \mathbb{R}$ be an interval and $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ be a random variable taking values in I. For a convex function $\varphi : I \to \mathbb{R}$ and a σ-algebra $\mathcal{F} \in \mathcal{A}$ the following inequality holds

$$\mathbb{E}[\varphi(X)|\mathcal{F}] \geq \varphi(\mathbb{E}[X|\mathcal{F}]).$$

Proof. Cf. [12], Theorem 8.20. □

Theorem B.2 ($\mathbb{E}[\,\cdot\,|\mathcal{F}]$ is a contraction). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $F \subseteq \mathcal{A}$ a sub σ-algebra and $p \in [1, \infty]$. Then the map

$$\mathbb{E}[\,\cdot\,|\mathcal{F}] : \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) \qquad X \mapsto \mathbb{E}[X|\mathcal{F}]$$

is a contraction on $\mathcal{L}^p$, i.e. $\|\mathbb{E}[X|\mathcal{F}]\|_p \leq \|X\|_p$

Proof. For $p < \infty$ this is an immediate consequence of Jensen's inequality. If $p = \infty$ note that $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X||\mathcal{F}] \leq \mathbb{E}[\|X\|_\infty|\mathcal{F}] = \|X\|_\infty$. □

Theorem B.3 (Doob's inequality). Let $p > 1$ and $\mathbf{X}$ be a right-continuous martingale in continuous time, s.t. $X_t \in \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ for all $t \geq 0$. Then

$$\mathbb{E}[\|\mathbf{X}\|_{T,\infty}^p] = \mathbb{E}\Big[\Big(\sup_{0\leq t\leq T} |X_t|\Big)^p\Big] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|X_T|^p]$$

for q s.t. $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Cf. [12], Theorem 11.2 (*ii*). □

Theorem B.4 (Continuous Mapping Theorem). Let (X_n) be a sequence of random elements on a metric space (S, d), s.t. $X_n \Rightarrow X$ for some X and let $F : (S, d) \to (S', d')$ be a map to another metric space that is continuous w.r.t. the distribution of X, i.e. $\mathbb{P}[X \in \{x \in S : F \text{ is discontinuous at } x\}] = 0$. Then

$$F(X_n) \Rightarrow F(X).$$

Proof. See [12], Theorem 13.25. □

Theorem B.5 (Prohorov's Theorem). Let (S, d) be a metric space and $\mathcal{N} \subseteq \mathcal{M}(S)$ be some family of probability measures on $\mathcal{B}_S$. Then

$\mathcal{N}$ is tight $\implies$ $\mathcal{N}$ is weakly relatively sequentially compact, that is every sequence $(\mathbf{N}^{(n)})$ in $\overline{\mathcal{N}}$ contains a subsequence (n_k) s.t. $\mathbf{N}^{(n_k)} \Rightarrow \mathbf{N}$ for some $\mathbf{N} \in \mathcal{N}$.

If (S, d) is Polish, the converse holds as well.

Proof. Cf. [12], Theorem 13.29. □

Theorem B.6 (Converging Together Theorem). Let (S, d) be a metric space. Suppose that $(X_{m,n}, X_n) \in S \times S$ are some random elements. If $X_{m,n} \Rightarrow Z_m$ as $n \to \infty$ for all m, $Z_m \Rightarrow X$ as $m \to \infty$ and

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left[d(X_{m,n}, X_n) \geq \varepsilon\right] = 0 \tag{B1}$$

for all $\varepsilon > 0$. Then $X_n \Rightarrow X$.

Proof. Cf. [7], Theorem 3.2, p. 28.
Let $F \subseteq S$ be any closed set and $F_\varepsilon := \{x \in S : d(x, F) \leq \varepsilon\}$, then

$$\mathbb{P}[X_n \in F] \leq \mathbb{P}[X_{m,n} \in F_\varepsilon] + \mathbb{P}[d(X_{m,n}, X_n) \geq \varepsilon].$$

Using $X_{m,n} \Rightarrow Z_m$ and Portmanteau's Theorem (Theorem 4.32), we can apply the limes superior to the preceding line (note that F_ε is closed)

$$\limsup_{n\to\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[Z_m \in F_\varepsilon] + \limsup_{n\to\infty} \mathbb{P}[d(X_{m,n}, X_n) \geq \varepsilon].$$

The same argument applied to $Z_m \Rightarrow X$ gives

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}[X_n \in F] &\leq \limsup_{m\to\infty} \mathbb{P}[Z_m \in F_\varepsilon] + \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}[d(X_{m,n}, X_n) \geq \varepsilon] \\ &\overset{\text{(B1)}}{\leq} \mathbb{P}[X \in F_\varepsilon] + 0.\end{aligned}$$

For the last inequality we used Portmanteau's Theorem again. Finally $\varepsilon \searrow 0$ implies $F_\varepsilon \searrow F$ since F is closed. $\square$

C. Topology and Functional Analysis

Proposition C.1. Let $f : X \to Y$ be a continuous map between topological spaces. Then $\partial(f^{-1}A) \subseteq f^{-1}(\partial A)$ for all $A \subseteq Y$.

Proof. Let $x \in \partial(f^{-1}(A))$. Then there are sequences $x_n \to x$ and $\tilde{x}_n \to x$ s.t. $x_n \notin f^{-1}(A)$ and $\tilde{x}_n \in f^{-1}(A)$ for all n. Hence $f(x_n) \to f(x)$ and $f(\tilde{x}_n) \to f(x)$ s.t. $f(x_n) \notin A$ and $f(\tilde{x}_n) \in A$ for all n. Consequently $f(x) \in \partial A$ and $x \in f^{-1}(\partial A)$. $\square$

Proposition C.2. Let E be a locally compact, second-countable space. Then there is a countable base of relatively compact sets.

Proof. The result follows if we can show that given a base $\mathcal{B}$, the set $\tilde{\mathcal{B}} := \{A \in \mathcal{B} : \overline{A}$ is compact$\}$ is a base as well. It even suffices to show that for any $x \in E$ and any open set $x \in O$, there is an element $B_O \in \tilde{\mathcal{B}}$ s.t. $x \in B_O \subseteq O$. Since the space is locally compact, there is a compact set K_x s.t. $x \in K_x \subseteq O$ and since $\mathcal{B}$ is a base, there is an element $x \in B_O \subseteq K_x^o \subseteq O$. Note that E is Hausdorff, hence K_x is closed and thus $B_O \in \tilde{\mathcal{B}}$ as desired. $\square$

Theorem C.3 (Tychonoff's Theorem). Let $\mathcal{I}$ be any index set and let $\{E_i, i \in \mathcal{I}\}$ be a family of compact spaces. Then $E = \prod_{i\in\mathcal{I}} E_i$ is compact in the product topology.

Proof. Cf. [14], Theorem 37.3. □

Proposition C.4 (about Polish Spaces). Let (E_n) be sequence of Polish spaces. Then the product $E = \prod_{n \geq 1} E_n$ is Polish. Moreover closed subspaces of Polish spaces are Polish.

Proof. This is an easy consequence of the well known facts, that a countable product of metric spaces/separable spaces/complete spaces is again metrizable/separable/complete, when endowed with the product topology. The same holds for closed subspaces equipped with the trace topology. □

Lemma C.5. Let K be a compact and O be an open, relatively compact set in a locally compact and second-countable space E. Then

(i) there are compact sets $K_n \searrow K$ and a non-increasing sequence (f_n) with $f_n \in \mathcal{C}_c^+(E)$ s.t. $\mathbb{1}_K \leq f_n \leq \mathbb{1}_{K_n}$. Moreover, if K is compact so is $K^\delta := K + \delta$ for $\delta > 0$ sufficiently small

(ii) there is a sequence of open and relatively compact sets (O_n) and a non-decreasing sequence (f_n) with $f_n \in \mathcal{C}_c^+(E)$ s.t. $\mathbb{1}_O \geq f_n \geq \mathbb{1}_{O_n}$.

Proof. Cf. [15], Lemma 3.11. □

D. Stochastic Processes

Proposition D.1 (Brownian Motion). A stochastic process $\mathbf{B} = (B_t)_{t \geq 0}$ with values in $\mathbb{R}$ is a (STANDARD) BROWNIAN MOTION if

(i) $B_0 = 0$ a.s.

(ii) $\mathcal{B}$ has independent and stationary increments

(iii) $B_t \sim \mathcal{N}(0, t)$ for all $t \geq 0$

(iv) $\mathbf{B}$ has a.s. continuous paths.

To get a Brownian motion on $\mathbb{R}^d$, let $\mathbf{B}^{(1)} = (B_t^{(1)})_{t \geq 0}, \dots, \mathbf{B}^{(d)} = (B_t^{(d)})_{t \geq 0}$ be some Brownian motions on $\mathbb{R}$ and set $\mathbf{B} := (\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(d)})$. The COVARIANCE MATRIX of $\mathbf{B}$ is $Q := (q_{i,j})_{1 \leq i,j \leq d}$ given by $q_{i,j} = \mathbb{E}\big[B_1^{(i)} B_1^{(j)}\big]$. It is always symmetric, positive semidefinite and determines the characteristic function of B_t via

$$\varphi_{B_t}(u) = e^{-t \frac{1}{2} Q u \cdot u}. \tag{D1}$$

Proof. That the matrix is always symmetric and positive semidefinite is easy to check. Equation (D1) is not hard either, since evidently $B_t \sim \mathcal{N}(0, tQ)$ and a multidimensional normal distribution with covariance matrix Q has the above characteristic function, see [12], Theorem 15.53. □

E. Miscellaneous

Theorem E.1 (Moore-Osgood Theorem). Let (S, d) be a metric space with a limit point $x \in S$ and let $S' \subseteq S$. Suppose that $f, f_1, f_2, \dots$ are some real-valued functions on S and (c_n) are some real numbers. If

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ uniformly on } S \qquad \text{and} \qquad \lim_{y\to x} f_n(y) = c_n \text{ pointwise on } \mathbb{N},$$

then the following double-limit exists and

$$\lim_{y\to x}\lim_{n\to\infty} f_n(y) = \lim_{n\to\infty}\lim_{y\to x} f_n(y).$$

Proof. Cf. [17], Theorem 7.11. □

Proposition E.2. Let $f \in \mathcal{C}_c^+(E)$, where $E := [0,\infty) \times \mathbb{R}_0^d$. Then there are numbers $\varepsilon > 0$ and $T > 0$ s.t. $supp(f) \subseteq [0,T] \times (\{y : |y| > \varepsilon\}) =: E_{T,\varepsilon}$.

Proof. By assumption $supp(f)$ is compact, hence there is some $T > 0$ s.t. $supp(f) \subseteq [0,T] \times \mathbb{R}_0^d$.
Assume by contrary, that there is no $\varepsilon > 0$ s.t. $supp(f) \subseteq E_{T,\varepsilon}$. Then we can find a sequence (t_n, y_n) in $[0,T] \times \bar{\mathbb{R}}_0^d$ s.t. $|y_n| \to 0$ and $f((t_n, y_n)) > 0$ for all n. Moreover, $t_n \in [0,T]$ for all n, so there is a limiting point $t \in [0,T]$ and a subsequence (n_k), s.t. $t_{n_k} \to t$. This implies $(t_{n_k}, y_{n_k}) \to (t, 0)$ and since the support of f is compact, hence closed, this would imply $(t, 0) \in supp(f)$, a contradiction. □

Proposition E.3. For $\delta > 0$ the set $F_\delta := \{\mathbf{x} \in \mathcal{D} : \|\mathbf{x}\|_{T,\infty} \geq \delta\}$ is closed in the Skorokhod topology.

Proof. Since $F_\delta^c = \{\mathbf{x} \in \mathcal{D} : \|\mathbf{x}\|_{T,\infty} < \delta\}$ is evidently open in the coarser uniform topology, the statement follows. Indeed, if we pick any $\mathbf{x} \in F_\delta^c$, then $\delta - \|\mathbf{x}\|_{T,\infty} =: \varepsilon > 0$. Hence by the definition of ε we have $\mathbf{x} + B_{\frac{\varepsilon}{2}}(0) \subseteq F_\delta^c$ (where the ball $B_{\frac{\varepsilon}{2}}(0)$ is either taken w.r.t. the Skorokhod or the uniform topology). □

Bibliography

[1] Aliprantis, C. D., and Border, K. C. *Infinite Dimensional Analysis. A Hitchhiker's Guide*, 3 ed. Springer, New York [i.a], 2006.

[2] Applebaum, D. *Lévy Proesses and Stochastic Calculus.* Cambridge University Press, Cambridge [i.a], 2004.

[3] Applebaum, D., et al. *Quantum Independent Increment Processes I.* Springer, New York [i.a], 2005.

[4] Bauer, H. *Measure and Integration Theory.* Walter de Gruyter GmbH, Berlin, 2001.

[5] Bertoin, J. *Lévy Processes.* Cambridge University Press, Cambridge [i.a], 1996.

[6] Billingsley, P. *Probability and Measure*, 3 ed. John Wiley and Sons, Inc., New York [i.a], 1995.

[7] Billingsley, P. *Convergence of Probability Measures.* John Wiley and Sons, Inc., New York [i.a], 1999.

[8] Elstrodt, J. *Mass und Integrationstheorie. Sechste, korrigierte Auflage*, 6 ed. Springer-Verlag, Berlin, Heidelberg, 2009.

[9] Jacod, J., and Shiryaev, A. N. *Limit Theorems for Stochastic Processes*, 7 ed. Springer, New York [i.a], 2003.

[10] Kallenberg, O. *Random Measures.* Akademie-Verlag, Berlin, 1975/76.

[11] Kallenberg, O. *Foundations of Modern Probabilty Theory. Second Edition*, 2 ed. Springer, New York [i.a], 2002.

[12] Klenke, A. *Probability Theory. A Comprehensive Course.* Springer Science+Business Media LLC, New York [i.a], 2007.

[13] Matthes, K., Kerstan, J., and Mecke, J. *Infinitely Divisible Point Processes.* Akademie-Verlag Berlin, John Wiley and Sons Ltd, Berlin, Chichester, 1978. English translation.

[14] Munkres, J. R. *Topology*, 2 ed. Prentice Hall, Inc., Upper Saddle River, US, 2000.

[15] Resnick, S. I. *Extreme Values, Regular Variation, and Point Processes.* Springer, New York [i.a], 1987.

[16] Resnick, S. I. *Heavy-Tail Phenomena.* Springer Science+Business Media LLC, New York, [i.a], 2007.

[17] Rudin, W. *Principles of Mathematical Analysis*, 2 ed. McGraw-Hill, Inc, United States, 1976.

[18] Sato, K.-I. *Lévy processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge [i.a], 1999. first English translation.

[19] Schelling., R. L. *Measures, Integrals and Martingales*. Cambridge University Press, Cambridge i.a., 2005.

[20] Tyran-Kamnińska, M. Convergence to Lévy stable processes under some weak dependence conditions. *Stochastic Processes and their Applications 120* (2010), 1629–1650.

[21] Tyran-Kamnińska, M. Weak convergence to Lévy stable processes in dynamical systems. *Stochastics and Dynamics 10* (2010), 263–289.

[22] Vestrup, E. M. *The Theory of Measures and Integration*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2003.

[23] Whitt, W. *Stochastic-Process Limits. An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer, New York [i.a], 2002.

Supplements

Deutsche Zusammenfassung

Stochastische Grenzwertsätze stellen eine natürliche Erweiterung der grundlegenden Fragestellungen der Wahrscheinlichkeitstheorie nach statistischer Regularität dar, indem sie sich nicht nur für ein Endresultat, sondern für den gesamten "Weg" bis dorthin interessieren. Als einfaches Beispiel dient der (faire) Münzwurf. Fundamentale und wohlbekannte Resultate der Wahrscheinlichkeitstheorie besagen, dass durch Wiederholen des Wurfes Regularität eintritt, in dem Sinn, als dass sich das Verhältnis der Anzahl an *Kopf* und *Zahl* fast sicher dem Wert 1 annähert. Interessiert man sich nun, nicht nur ausschliesslich für das Verhältnis nach n Würfen, sondern dafür *auf welchem Weg* dieses Verhältnis zu Stande gekommen ist, so benötigt man stochastische Konvergenz von Funktionen. Ähnliche Fragestellungen wie diese führten die Wahrscheinlichkeitstheorie auf die Spur stochastischer Grenzwertsätze.

Ausgehend von Donskers Theorem, dem wohl berühmtesten stochastischen Grenzwertsatz, wurden im Laufe der Jahre viele weitere bewiesen. Klassische Werkzeuge, um solche Aussagen zu erhalten bzw. aus schon bestehenden Resultaten neue zu kreieren, sind das *Continuous Mapping Theorem* und *Straffheitskriterien.*

Das Ziel dieser Arbeit ist nun eine neue Methode vorzustellen, mit deren Hilfe es möglich ist Konvergenz bzw. auch ein Versagen von Konvergenz nachzuprüfen. Etwas präziser gesagt, stellt diese Arbeit eine Methode vor, wodurch schwache Konvergenz (Konvergenz in Verteilung) von Partialsummenprozessen zu (speziellen) Lévyprozessen durch Konvergenz geeignet gewählter Objekte *charakterisiert* werden kann. Die Pendants sind dabei sogenannte *zufällige Maße* bzw. *Punktprozesse*, welche die Unstetigkeitsstellen der Partialsummenprozesse beschreiben. Das erklärte Ziel ist also, schwache Konvergenz von zufälligen Maßen zu erklären und dann mit jener der Prozesse in Verbindung zu setzen.

Für diese Verbindung ist es nötig, schwache Konvergenz von Folgen in unterschiedlichen Räumen zu verknüpfen. Eine naheliegende Möglichkeit dafür ist das Continuous Mapping Theorem. Dieses erfordert jedoch (im Wesentlichen) Stetigkeit entsprechender Abbildungen, wodurch es nötig wird, eine geeignete Topologie auf den entsprechenden Räumen zu erklären. In Kapitel 2 geben wir daher eine kurze Einführung in die Skorokhod'sche $\mathcal{J}_1$-Topologie auf dem Raum der càdlàg Funktionen (welcher den Lévyprozessen zu Grunde liegt) und beweisen erste Stetigkeitsresultate, sowie die Tatsache, dass $\mathcal{D}$ vollständig metrisierbar ist. Selbstredend ist auch eine Topologie auf Maßräumen nötig, welche – nach einem einführenden Teil über zufällige Maße und Punktprozesse – in Kapitel 4 eingeführt wird. Wie wir zeigen werden, besitzt auch die vage Topologie die schöne Eigenschaft, die Maßräume vollständig metrisierbar und separabel zu machen. Ein weiteres Kernresultat dieses Kapites ist das Theorem von Kallenberg, welches schwache Konvergenz von Punktprozessen durch einige wenige Parameter charakterisiert. In Kapitel 4 werden weiters erste (einfache) Verbindungen zwischen Poisson'schen zufälligen Maßen und zusammengesetzen Poissonprozessen bzw. Martingalen aufgezeigt.

Kapitel 5 ist eine (sehr) kurze Einführung in die Theorie der Lévyprozesse, in dem die Lévy-Itō Zerlegung mit Hilfe der Lévy-Khintchine Formel und zufälligen Maßen bewiesen wird. Kaptiel 6 stellt das Herz der Arbeit dar, in dem die zuvor erlangten Resultate zusammengetragen werden und zentrale Stetigkeitsaussagen für Abbildungen zwischen dem Raum $\mathcal{D}$ und dem der Punktprozesse $\mathcal{M}_p(E)$ bewiesen werden. Die Arbeit endet mit dem Beweis des zentralen Theorems von Marta Tyran-Kamińska.

Curriculum Vitae

Personal Data

Name	Fabian Pühringer
Date of birth	June 9^{th}, 1990
Place of birth	Wels, Austria
Nationality	Austria

Education

09/00–06/08	**Secondary School**, *BG/BRG Anton-Brucknerstrasse Wels*, Graduation with distinction.
10/08–08/12	**Bachelor Mathematics**, *University of Vienna*, Graduation with distinction.
10/08–08/13	**Bachelor Political Science**, *University of Vienna*, Graduation with distinction.
09/13–01/14	**ERASMUS Outgoing semester**, *Université Pierre et Marie Curie (Paris 6)*.
since 10/12	**Master Mathematics**, *University of Vienna*.

Scholarships

05/14-09/14	Scholarship for master students, FWF
2010-2013	Academic Excellence Scholarships, University of Vienna

Teaching Experience

10/12–01/13	**Teaching Assistant**, *University of Natural Resources and Life Sciences, Vienna (BOKU Wien)*, class name: "Mathematik für LBT" (Mathematics for Life Science Students).
03/13–06/13	**Teaching Assistant**, *University of Technology, Vienna (TU Wien)*, class name: "Analysis für Informatik" (Calculus for Computer Scientists).