# Kan extensions in probability theory

*Ruben Van Belle*

Doctor of Philosophy
University of Edinburgh
2023

# Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Ruben Van Belle*)

*To my grandfathers, Guy Van Belle and Chiel Dekkers*

# Abstract

In this thesis we will discuss results and ideas in probability theory from a categorical point of view. One categorical concept in particular will be of interest to us, namely that of Kan extensions. We will use Kan extensions of '*ordinary*' functors, *enriched* functors and *lax natural transformations* to give categorical proofs of some fundamental results in probability theory and measure theory. We use Kan extensions of 'ordinary' functors to represent probability monads as codensity monads. We consider a functor representing probability measures on countable spaces. By Kan extending this functor along itself, we obtain a codensity monad describing probability measures on all spaces. In this way we represent probability monads such as the Giry monad, the Radon monad and the Kantorovich monad.

Kan extensions of lax natural transformations are used to obtain a categorical proof of the Carathéodorody extensions theorem. The Carathéodory extension theorem is a fundamental theorem in measure theory that says that premeasures can be extended to measures. We first develop a framework for Kan extensions of lax natural transformations. We then represent outer and inner (pre)measures by certain lax and colax natural transformations. By applying the results on extensions of transformations a categorical proof of Carathéodory's extension theorem is obtained.

We also give a categorical view on the Radon–Nikodym theorem and martingales. For this we need Kan extensions of enriched functors. We start by observing that the finite version of the Radon–Nikodym theorem is trivial and that it can be interpreted as a natural isomorphism between certain functors, enriched over **CMet**, the category of complete metric spaces and 1-Lipschitz maps. We proceed by Kan extending these, to obtain the general version of the Radon–Nikodym theorem. Concepts such as conditional expectation and martingales naturally appear in this construction. By proving that these extended functors preserve certain cofiltered limits, we obtain categorical proofs of a weaker version of a martingale convergence theorem and the Kolmogorov extension theorem.

# Lay summary

Kan extensions are an important concept in category theory. This is an area in mathematics that itself can be viewed as an abstraction of mathematics. It provides a universal language to study different areas of mathematics. One of the advantages of doing this is that results in certain areas of mathematics can be obtained from abstract categorical principles. Another advantage is that results and concepts between different fields of mathematics can be compared to each other. Many parts of mathematics have been investigated from this point of view. However, probability theory has been studied less from a categorical perspective. The main objective of this thesis is to describe probabilistic concepts using the language of category theory and prove results in probability theory using categorical principles. One concept in particular will be used, which is that of Kan extensions. These are extensions that are optimal in some sense.

We will show that knowing how probabilities work on finite sets, determines how probabilities work on more general spaces; the Kan extension of finitely supported measures gives all probability measures. Furthermore, a proof of the Carathéodory extension theorem is given using Kan extensions of lax natural transformations. We will also give a categorical proof for the Radon–Nikodym theorem. We first observe that for finite sets the result is trivial and then proceed by Kan extending the trivial version of the theorem to the general version. Furthermore, this leads to a new proof of the martingale convergence theorem and categorical descriptions of several probabilistic concepts.

# Acknowledgements

I would like to thank my supervisor, Tom Leinster, for his advice and for many motivating and interesting conversations during my PhD. I would also like to thank my parents for their continuous support during my time in Edinburgh.

# Contents

# Chapter 1

# Introduction

In this section we will give some preliminaries on probability theory and category theory, in particular on Kan extensions. In Section 1.1 we will give an overview of several basic concepts and important results in probability theory. We do this from a categorical point of view. We start from the very basics of probability theory and end with some more advanced results. We do this from the perspective of a category theorist and spend plenty of time on discussing which concepts have straightforward categorical interpretations and which ideas in probability theory are more difficult to grasp using the language of category theory.

In Section 1.2, we give an overview of results of Kan extensions. We describe the classical Kan extensions of '*ordinary*' functors and their main properties. We then proceed by looking at Kan extensions of *enriched* functors and Kan extensions of *lax natural transformations*.

## 1.1 Probability theory

It is not obvious what *the* main objects of probability theory are. Several options come to mind:

1. Is probability theory about *probability spaces*?
2. Is probability theory about *random variables*?
3. Is probability theory about *Markov kernels*?

In Section 1.1.1, we discuss why probability spaces are important to describe stochastic events, but we also discuss why they are not sufficiently powerful to describe everything we want to model using probability theory. We will also describe some interesting categorical properties and some categorical limitations about probability spaces. In Section 1.1.2 and Section 1.1.3, we give an overview of the theory of random variables and the theory of Markov kernels. We see that both of these offer a more complete framework to mathematically formalize probability theory. However, we will also see that for certain situations the random variables approach is more useful than that of Markov kernels and vice versa. We will also see that the framework of Markov kernels has a more straightforward categorical interpretation than the one of random variables.

### 1.1.1 Probability theory is about probability spaces

An important first step in probability theory is to find a way to mathematically model randomness. If we want to describe a stochastic event with only a *countable* number of outcomes, there is a very intuitive way to model this. Namely, let $A$ be the countable set of outcomes and let $(p_a)_{a \in A} \in [0,1]^A$ be a collection of real numbers between 0 and 1 such that

$$\sum_{a \in A} p_a = 1.$$

For example to describe a fair coin mathematically, we just need to look at the set $\{H, T\}$ together with the probabilities $p_H := 1/2$ and $p_T := 1/2$. This model is the *Bernoulli distribution* with parameter 1/2.

Another example can be given using whiteboard markers. Suppose there is an infinite supply of used whiteboard markers, of which 25% still work. We arbitrarily choose a marker and test it. If the marker does not write well, we throw it away and choose another one. We are interested in how many attempts it takes to find a working marker. There is a *countable* collection of possible outcomes, namely $\mathbb{N} \setminus \{0\}$. The probability that we find a good marker on the $n$th attempt is given by

$$p_n := \left(\frac{3}{4}\right)^{n-1} \cdot \frac{1}{4}.$$

This model is the *geometric distribution* with the parameter 1/4 and it will take an average of 4 attempts to find a marker that works.

However, if the set of possible outcomes of our stochastic event is uncountable this definition does not work anymore. Indeed, if $A$ is uncountable and we have a collection $(p_a)_{a \in A}$ such that $\sum_{a \in A} p_a = 1$, then there can be at most a countable number of elements in $A$ with a strictly positive probability. This means that the model essentially reduces to the countable case, which can not be used to describe every stochastic event.

Suppose we are throwing darts at a regulation dart board (which has a diameter of 17³/4 inch). We assume that every dart lands on the dartboard and that we are not aiming at a specific point on the dartboard. There are an uncountable amount of points on the board the dart could land in and the probability of landing in one specific point is 0. However, the probability of landing in a region $A$ of the board with area $S$ square inch is given by

$$p_A := \frac{S}{(8.875)^2 \pi}.$$

In particular, the probability of landing the dart within a distance $r$ inch of the center of the board is equal to $\left(\frac{r}{8.875}\right)^2$. From this example it follows that it is not enough to assign a probability to every point of the dart board. We also need to assign a probability to every region (i.e. a subset of points on the board) of which we can measure the surface area.

This idea is formalized by Kolmogorov's axioms of probability, introduced in [41]. Let $\Omega$ be a set and let $\mathcal{F}$ be a **$\sigma$-algebra** of subsets of $\Omega$, i.e. a collection of subsets that is closed under countable unions and complements, containing the subsets $\emptyset$ and $\Omega$. The elements of $\mathcal{F}$ are called **measurable** subsets and the pair $(\Omega, \mathcal{F})$ a **measurable space**. In the darts example, $\Omega$ would be the collection of points on the board and $\mathcal{F}$ the collection of regions on the board, whose surface area we can measure. Then Kolmogorov's axioms of probability define a *probability measure* as follows.

**Definition 1.1.1** (Kolmogorov's axioms of probability). A **probability measure** on $(\Omega, \mathcal{F})$ is a map $\mathbb{P} : \mathcal{F} \to [0,1]$ such that:

1. $\mathbb{P}(\Omega) = 1$.

2. For every countable collection $(A_n)_{n=1}^{\infty}$ of pairwise disjoint measurable subsets,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right).$$

A triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**. In the case that $\Omega$ is countable, it is enough to assign to every element $\omega \in \Omega$ a number $p_\omega$ between 0 and 1 such that $\sum_{\omega \in \Omega} p_\omega = 1$, since this can be uniquely extended to a probability measure on $(\Omega, \mathcal{P}(\Omega))$. Therefore the intuitive model for stochastic events with countable possible outcomes is also captured by the probability spaces. In Chapter 3, we will show that Kolmogorov's generalization arises from a purely categorical construction.

Measurable spaces often come from more structured spaces, such as topological spaces or metric spaces. For example, let $(X, \mathcal{T})$ be a topological space. Then $(X, \sigma(\mathcal{T}))$ is a measurable space, where $\sigma(\mathcal{T})$ is the smallest $\sigma$-algebra containing $\mathcal{T}$. This $\sigma$-algebra is called the **Borel $\sigma$-algebra**. Many measurable spaces arise in the following way. Let $E$ be a set and let $\mathcal{B}$ be an algebra of subsets of $E$, i.e. a collection of subsets of $E$ that is closed under finite unions and complements, containing the subsets $\emptyset$ and $E$. Consider the smallest $\sigma$-algebra that contains $\mathcal{B}$, which is denoted as $\sigma(\mathcal{B})$. Then $(E, \sigma(\mathcal{B}))$ forms a measurable space. However, because a $\sigma$-algebra is closed under operations that do not distribute over each other, it is usually not possible to describe what a general element of $\sigma(\mathcal{B})$ looks like in terms of elements of $\mathcal{B}$. This makes it difficult to define a probability measure on $(E, \sigma(\mathcal{B}))$. A solution to this problem is given by *Carathéodory's extension theorem*, an important result in measure theory which we will discuss in detail from a categorical point of view in Chapter 4. The classical proof for this result can be found in [40] (Theorem 1.53).

**Theorem 1.1.2** (Carathéodory extension theorem)**.** *Let $E$ be a set and let $\mathcal{B}$ be an algebra of subsets. Let $p : \mathcal{B} \to [0, 1]$ be a map such that:*

1. $p(E) = 1$,

2. *For all countable collections $(A_n)_{n=1}^{\infty}$ of pairwise disjoint subsets in $\mathcal{B}$, such that their union $A := \bigcup_{n=1}^{\infty} A_n$ is also an element in $\mathcal{B}$, we have*

$$\sum_{n=1}^{\infty} p(A_n) = p(A).$$

*Then there exists a unique probability measure $\mathbb{P}$ on $(E, \sigma(\mathcal{B}))$ such that for all $A \in \mathcal{B}$,*

$$\mathbb{P}(A) = p(A).$$

The obvious structure-preserving maps between probability spaces are **measure-preserving maps**, i.e. for probability spaces $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, a measure-preserving map is a function $f : \Omega_1 \to \Omega_2$ such that $f^{-1}(\mathcal{F}_2) \subseteq \mathcal{F}_1$ and such that $\mathbb{P}_1(f^{-1}(A)) = \mathbb{P}_2(A)$ for all $A \in \mathcal{F}_2$. In other words, a map is measure-preserving if the pushforward of $\mathbb{P}_1$ along $f$, denoted by $\mathbb{P}_1 \circ f^{-1}$ or $f_*\mathbb{P}_1$, is equal to $\mathbb{P}_2$. Clearly, composing measure-preserving maps gives again a measure-preserving map. Therefore probability spaces together with measure-preserving maps form a category, which we will denote by **Prob**.

A first attempt towards a categorical view on probability theory is to investigate this category of probability spaces. We will do this by studying the limits and colimits in this category and its canonical symmetric monoidal structure.

In Proposition 1.1.5 we will show that **Prob** does not have products in general. However, it does a have a natural *tensor product*. Consider two probability space $\boldsymbol{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\boldsymbol{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Using the Carathéodory extension theorem, there exists a unique probability measure $\mathbb{P}_1 \otimes \mathbb{P}_2$ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that

$$\mathbb{P}_1 \otimes \mathbb{P}_2(E_1 \times E_2) = \mathbb{P}_1(E_1)\mathbb{P}_2(E_2),$$

for all $E_1 \in \mathcal{F}_1$ and $E_2 \in \mathcal{F}_2$.

We obtain a probability space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ which we denote by $\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Omega}_2$. This defines a functor

$$\otimes : \mathbf{Prob} \times \mathbf{Prob} \to \mathbf{Prob}.$$

This functor together with the one-point probability space gives **Prob** a *symmetric monoidal* structure.

More generally, given an indexed collection of probability spaces $(\boldsymbol{\Omega}_i := (\Omega_i, \mathcal{F}_i, \mathbb{P}_i))_{i \in I}$, we can use the Carathéodory extension to prove that there exists a unique probability measure on $\left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{F}_i\right)$ such that

$$\mathbb{P}\left(\prod_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i),$$

for $(A_i)_{i \in I}$ with $A_i \in \mathcal{F}_i$ and only a *finite* amount of the $A_i$'s different from $\Omega_i$. We denote this probability measure by $\bigotimes_{i \in I} \mathbb{P}_i$ and the obtained probability space as $\bigotimes_{i \in I} \boldsymbol{\Omega}_i$. This is explained in Section 10.6 in [11].

**Theorem 1.1.3.** *The category* **Prob** *has colimits of non-empty diagrams.*

*Proof.* We will first show that **Prob** has coproducts. Let $(\boldsymbol{\Omega}_i := (\Omega_i, \mathcal{F}_i, \mathbb{P}_i))_{i \in I}$ be a non-empty collection of probability spaces. Define $\boldsymbol{\Omega} := \coprod_{i \in I} \Omega_i$ and

$$\mathcal{F} := \left\{ \coprod_i A_i \mid A_i \in \mathcal{F}_i \text{ for all } i \in I, \mathbb{P}_{i_1}(A_{i_1}) = \mathbb{P}_{i_2}(A_{i_2}) \text{ for all } i_1, i_2 \in I \right\}.$$

The set $\mathcal{F}$ is closed under complements, countable increasing unions and finite disjoint unions, therefore it is a $\sigma$-algebra. Define $\mathbb{P} : \mathcal{F} \to [0,1]$ by sending $\coprod_{i \in I} A_i$ to $\mathbb{P}_i(A_i)$ for any $i \in I$. This defines a probability measure on $(\Omega, \mathcal{F})$, which forms a probability space $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$. For $i \in I$, there is a measure-preserving inclusion $\iota_i : \boldsymbol{\Omega}_i \to \boldsymbol{\Omega}$.

Suppose now that there is a probability space $\boldsymbol{\Xi} := (\Xi, \mathcal{G}, \mathbb{Q})$ together with measure-preserving maps $f_i : \boldsymbol{\Omega}_i \to \boldsymbol{\Xi}$ for every $i \in I$. There is a map $f : \Omega \to \Xi$ such that for every $i \in I$,

$$\begin{tikzcd}
\Omega_i \arrow[r, "\iota_i"] \arrow[dr, "f_i"'] & \Omega \arrow[d, "f"] \\
& \Xi
\end{tikzcd}$$

For $E \in \mathcal{G}$, we find that for every $i \in I$,

$$\mathbb{P}(f^{-1}(E)) = \mathbb{P}_i(f^{-1}(E) \cap \Omega_i) = \mathbb{Q}(E).$$

It follows now that $f^{-1}(E) \in \mathcal{F}$ and that $f$ is measure-preserving. This shows that $\boldsymbol{\Omega}$ together with the maps $(\iota_i)_i$ is the coproduct of $(\boldsymbol{\Omega}_i)_{i \in I}$.

We will now show that **Prob** also has coequalizers. Consider probability spaces $\boldsymbol{\Omega}_1 :=$

$(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\boldsymbol{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ and measure-preserving maps $f_1, f_2 : \boldsymbol{\Omega}_1 \to \boldsymbol{\Omega}_2$. Let $(\Omega, \mathcal{F})$ be the coequalizer of the measurable maps $f_1, f_2 : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ and let $e : (\Omega_2, \mathcal{F}_2) \to (\Omega, \mathcal{F})$ be the coequalizer map. Define $\mathbb{P}$ as $\mathbb{P}_2 \circ e^{-1}$, i.e. the pushforward of $\mathbb{P}$ along $e$. Then $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $e$ becomes a measure-preserving map. This forms the coequalizer of $f_1$ and $f_2$ in **Prob**.

Since **Prob** has coproducts and coequalizers, the claim follows. ☐

From the proof of Theorem 1.1.3, we can see that the underlying measurable space of the coproduct of probability spaces is, in general, not the same as the coproduct of the underlying measurable spaces of the probability spaces. In other words, the forgetful functor $U : \mathbf{Prob} \to \mathbf{Mble}$ does not preserve coproducts. It does however preserve coequalizers.

**Proposition 1.1.4.** *The category* **Prob** *does not have an initial object.*

*Proof.* Suppose there is an initial object $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ in **Prob**, then there is a unique map $\boldsymbol{\Omega} \to \boldsymbol{\Omega} \otimes \boldsymbol{\Omega}$. This has to be the diagonal map $\Delta : \Omega \to \Omega \times \Omega$. It follows that $\mathbb{P} \circ \Delta^{-1} = \mathbb{P} \otimes \mathbb{P}$.

Now consider any probability space $\boldsymbol{\Xi} := (\Xi, \mathcal{G}, \mathbb{Q})$ such that there exists a measurable subset $B \in \mathcal{G}$ such that $0 < \mathbb{Q}(B) < 1$. Because $\boldsymbol{\Omega}$ is an initial object, there exists a unique measure-preserving map $f : \boldsymbol{\Omega} \to \boldsymbol{\Xi}$. Therefore $A := f^{-1}(B)$ is a measurable subset such that $0 < \mathbb{P}(A) < 1$. We will write $A^{\mathrm{C}} := \Omega \setminus A$. It follows now that

$$0 < \mathbb{P}(A)\mathbb{P}(A^{\mathrm{C}}) = \mathbb{P} \otimes \mathbb{P}(A \times A^{\mathrm{C}}) = \mathbb{P} \circ \Delta^{-1}(A \times A^{\mathrm{C}}) = \mathbb{P}(A \cap A^{\mathrm{C}}) = 0.$$

This is a contradiction. Therefore, no initial object can exist in **Prob**. ☐

What Proposition 1.1.4 means probabilistically is that there is no upper bound on the randomness that probability spaces can express. There is no ultimate probability distribution from which every other probability distribution is the pushforward measure in a unique way.

We will now see that the situation for limits in **Prob** is more complicated.

**Proposition 1.1.5.** *The category* **Prob** *does not have all products.*

*Proof.* Let $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, such that there exists $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$. Suppose $\boldsymbol{\Xi} := (\Xi, \mathcal{G}, \mathbb{Q})$ is the product of $\boldsymbol{\Omega}$ with itself with measure-preserving projection maps $p_1, p_2 : \boldsymbol{\Xi} \to \boldsymbol{\Omega}$.

There is a measurable map $(\Xi, \mathcal{G}) \to (\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ such that the following diagram commutes:

$$\begin{tikzcd}
& (\Xi, \mathcal{G}) \arrow[dl, "p_1"'] \arrow[d, "p"] \arrow[dr, "p_2"] & \\
(\Omega, \mathcal{F}) & (\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F}) \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & (\Omega, \mathcal{F})
\end{tikzcd}$$

Furthermore, there is a measure-preserving map $p : \boldsymbol{\Omega} \to \boldsymbol{\Xi}$ such that we have a commutative diagram,

$$\begin{tikzcd}
& \boldsymbol{\Omega} \arrow[dl, "1_{\boldsymbol{\Omega}}"'] \arrow[d, "f"] \arrow[dr, "1_{\boldsymbol{\Omega}}"] & \\
\boldsymbol{\Omega} & \boldsymbol{\Xi} \arrow[l, "p_1"] \arrow[r, "p_2"'] & \boldsymbol{\Omega}
\end{tikzcd}$$

However, there is also a measure-preserving map $g : \mathbf{\Omega} \otimes \mathbf{\Omega} \to \Xi$ such that

$$\begin{tikzcd} & \mathbf{\Omega} \otimes \mathbf{\Omega} \arrow[dl, "\pi_1"'] \arrow[d, "g"] \arrow[dr, "\pi_2"] & \\ \mathbf{\Omega} & \mathbf{\Xi} \arrow[l, "p_1"] \arrow[r, "p_2"'] & \mathbf{\Omega} \end{tikzcd}$$

From this we can conclude that $\pi_1 pg = p_1 g = \pi_1$ and $\pi_2 pg = p_2 g = \pi_2$ and therefore

$$pg = 1_{\Omega \times \Omega}.$$

Similarly, we have that $\pi_1 pf = p_1 f = 1_\Omega$ and $\pi_2 pf = p_2 f = 1_\Omega$, which implies that

$$pf = \Delta.$$

Here $\Delta : \Omega \to \Omega \times \Omega$ is the diagonal map. Since $f$ and $g$ are measure-preserving, we see that $\mathbb{P} \circ f^{-1} = \mathbb{Q} = \mathbb{P} \otimes \mathbb{P} \circ g^{-1}$. It follows that

$$\mathbb{P} \circ \Delta^{-1} = \mathbb{P} \circ f^{-1} \circ p^{-1} = \mathbb{P} \otimes \mathbb{P} \circ g^{-1} \circ p^{-1} = \mathbb{P} \otimes \mathbb{P}.$$

Let $A$ be the measurable subset such that $0 < \mathbb{P}(A) < 1$, then

$$0 < \mathbb{P}(A)\mathbb{P}(A^{\mathrm{C}}) = \mathbb{P} \otimes \mathbb{P}(A \times A^{\mathrm{C}}) = \mathbb{P} \circ \Delta^{-1}(A \times A^{\mathrm{C}}) = \mathbb{P}(A \cap A^{\mathrm{C}}) = 0.$$

This is a contradiction and therefore $\mathbf{\Omega}$ has no product with itself in **Prob**. ☐

A more probabilistic way of explaining the result in Proposition 1.1.5 is as follows. Given two probability spaces $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, we can look for probability measures $\mathbb{P}$ on $(\Omega_1 \times \Omega_2, \mathcal{F} \otimes \mathcal{F})$ such that $\mathbb{P} \circ \pi_1^{-1} = \mathbb{P}_1$ and $\mathbb{P} \circ \pi_2^{-1} = \mathbb{P}_2$, where $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ and $\pi_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ are the projection maps. Such a probability measure is called a **coupling** of $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$. The product of the probability measures $\mathbb{P}_1 \otimes \mathbb{P}_2$ always forms a coupling, but in general there are many other couplings. In other words, the two probability spaces $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ do not determine a joint probability distribution on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. This means that if we model stochastic events by probability spaces, we do not have enough information to describe the interactions between the stochastic events.

In general we also do not have equalizers in **Prob**. Indeed, consider the probability space $\{H, T\}$ associated to a fair coin. The identity map and the map that switches heads and tails are both measure-preserving. If an equalizer existed, its underlying map should be the empty set, but this cannot be a probability space. Since subspace constructions often arise as equalizers, this indicates that we can not describe conditional probability measures using a categorical construction in **Prob**. Given a probability space $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ and a measurable subset $A$ with $\mathbb{P}(A) > 0$, we can try to *restrict* $\mathbb{P}$ to $A$. However, since we want to end up with another probability measure, we need to rescale. This gives us the **conditional probability measure** defined by

$$\mathbb{P}(B \mid A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

The category **Prob** does have a terminal object. This is the probability space, whose underlying set is the set of one element. This is not the only limit that exists in **Prob**. Some cofiltered limits also exist. This is described by the Kolmogorov extension theorem for which a proof can be found in Theorem 10.6.2 in [11]. For this we need the concept of *standard Borel spaces*. A **standard Borel space** is a measurable space that is induced by a **Polish space**, i.e.

a separable completely metrizable topological space.

**Theorem 1.1.6** (Kolmogorov extension theorem). *Let $I$ be a non-empty set and let $(\Omega_i)_{i\in I}$ be a collection of standard Borel spaces. For every finite subset $J$ of $I$, let $\mathbb{P}_J$ be a probability measure on $\prod_{i\in J}\Omega_i$ such that*

$$\mathbb{P}_{J_2}\circ\pi^{-1}_{J_1\subseteq J_2}=\mathbb{P}_{J_1},$$

*for finite subsets $J_1\subseteq J_2$, where $\pi_{J_1\subseteq J_2}:\prod_{i\in J_2}\Omega_i\to\prod_{i\in J_1}\Omega_i$ is the projection map.*

*Then there exist a unique probability measure $\mathbb{P}$ on $\prod_{i\in I}\Omega_i$ such that*

$$\mathbb{P}\circ\pi_J=\mathbb{P}_J,$$

*for all finite subsets $J\subseteq I$, where $\pi_J:\prod_{i\in I}\Omega_i\to\prod_{i\in J}\Omega_i$ is the projection map.*

Kolmogorov's extension theorem can be rephrased more categorically. Let $I$ be a non-empty set and let $(\Omega_i)_{i\in I}$ be a collection of standard measurable spaces. Let $\mathcal{P}_f(I)$ be the poset of finite subsets of $I$ ordered by inclusion. Consider a diagram

$$D:\mathcal{P}_f(I)^{\mathrm{op}}\to\mathbf{Prob}$$

such that the underlying measurable space of $D(J)$ is equal to $\prod_{i\in J}\Omega_i$ and the underlying measurable map of $D(J_1\subseteq J_2)$ is given by the projection map $\pi_{J_1\subseteq J_2}:\prod_{i\in J_2}\Omega_i\to\prod_{i\in J_1}\Omega_i$. The Kolmogorov extension theorem then states that $D$ has a limit.

The statement does not hold anymore for a collection of *arbitrary* probability spaces. A counterexample is described in [2], which shows that **Prob** does not have all cofiltered limits.

Moreover, it follows that a joint distribution $\mathbb{P}$ on $\prod_{i\in I}\Omega_i$ is completely determined by its *finite-dimensional* distributions $\{\mathbb{P}\circ\pi_J^{-1}\mid J\subseteq I\text{ finite}\}$. In fact, this does hold for arbitrary measurable spaces. For a collection of measurable space $(\Omega_i)_{i\in I}$ and probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $\prod_{i\in I}\Omega_i$ such that

$$\mathbb{P}_1\circ\pi_J^{-1}=\mathbb{P}_2\circ\pi_J^{-1},$$

for all finite subsets $J$ of $I$, we have that $\mathbb{P}_1=\mathbb{P}_2$. This is a consequence of Corollary 1.6.3 in [11] which follows from the $\pi$-$\lambda$-theorem

**Example 1.1.7.** Let $I=[0,\infty)$ and for $t_1<\ldots<t_n$ in $I$ define a probability measure on $\mathbb{R}^n$ by

$$\mathbb{P}_{t_1,\ldots,t_n}(B):=\int_B\prod_{i=1}^n\frac{\exp\frac{-(x_i-x_{i-1})^2}{2(t_i-t_{i-1})}}{\sqrt{2\pi(t_i-t_{i-1})}}\mathrm{d}(x_1,\ldots,x_n),$$

for all measurable subsets $B$ of $\mathbb{R}^n$.

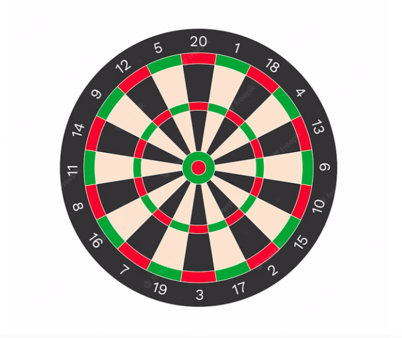
By the Kolmogorov extension theorem, there exists a unique probability measure $\mathbb{P}$ on $\mathbb{R}^I$ such that its finite-dimensional distributions are precisely given by $\mathbb{P}_{t_1,\ldots,t_n}$. This is an important first step in the proof that Brownian motion exists. The remaining part is about showing that $\mathbb{P}$ can be restricted to a probability measure on $\mathcal{C}([0,\infty),\mathbb{R})$. This is explained in Chapter 7 of [16].

We can conclude that the category **Prob** has some nice structures, but is not expressive enough to properly deal with interactions between stochastic events. The concept of probability spaces on its own does not provide enough information to study (in)dependence and interaction between different stochastic events.

### 1.1.2 Probability theory is about random variables

In Section 1.1.1, we saw that the concept of probability spaces on its own is not sufficient to describe *interactions* between stochastic events. A better way for modelling stochastic events is given by *random variables.* This solves the problem of describing interactions between stochastic events, but also allows us to talk about the *(conditionally) expected* outcome of a stochastic event. Furthermore, random variables also form a very convenient tool to describe random events that evolve over time, by using *stochastic processes.*

Suppose we want to model a stochastic event whose outcomes lie in a measurable space $E$. A typical way to categorically describe the elements of an object $A$ in a category $\mathcal{C}$ is to look at the maps $I \to A$, where $I$ is the terminal object of the category. By changing $I$ to an arbitrary fixed object $U$ in the category $\mathcal{C}$, we can talk about **generalized elements**. The concept of random variables follows a similar idea. Since we want to generate randomness in the space $E$, we fix a probability space $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$. A **random variable** is then defined as a measurable map $X : (\Omega, \mathcal{F}) \to E$. We will often just write $X : \Omega \to E$.



Let us look again at the example of throwing darts. Consider the subset of $\mathbb{R}^2$ given by

$$\Omega := \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1/\pi\right\}.$$

Let $\mathcal{F}$ be the restriction of the Borel $\sigma$-algebra on $\mathbb{R}^2$ to $\Omega$. The Lebesgue measure $\lambda$ on $\mathbb{R}^2$ restricted to the subset $\Omega$, becomes a probability measure $\mathbb{P}$. This gives us a probability space $\boldsymbol{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$. Every area on the dart board corresponds to a different score between 1 and 60. Therefore we can model the throw of a dart without aiming at a specific point on the board as a random variable

$$X : \Omega \to \{1, \dots, 60\}.$$

Suppose we are throwing three darts without aiming to any specific point on the board. Then these correspond to three *independent* random variables

$$X_1 : \Omega^{\otimes 3} \to \{1, \dots, 60\}, \quad X_2 : \Omega^{\otimes 3} \to \{1, \dots, 60\} \quad \text{and} \quad X_3 : \Omega^{\otimes 3} \to \{1, \dots, 60\}.$$

The total score of the game of darts is then given by a random variable $S : \Omega^{\otimes 3} \to \{1, \dots, 180\}$ defined by

$$S := X_1 + X_2 + X_3.$$

We can now ask questions such as:

- What is the probability that $S$ is equal to 180?
- Given that $X_1$ is equal to 60, what is the probability that $S$ equals 180?

- What is the expected outcome of $S$?
- Give that $X_1$ is equal to 60, what is the expected outcome of $S$?

From this, it is —at least intuitively— clear that the framework of random variables can describe interactions and dependence between different stochastic events.

The first two questions can be answered by looking at the **distribution of $X$**. This is the pushforward measure of $\mathbb{P}$ along $X$, i.e. the probability measure on $E$ defined by

$$A \mapsto \mathbb{P}(X^{-1}(A)).$$

Typical notation for this measure is given by $X_*(\mathbb{P})$ or $\mathbb{P} \circ X^{-1}$. We will use the latter notation. The probability that a random variable takes values in a measurable subset $A$ of $E$ is then given by $\mathbb{P} \circ X^{-1}(A)$, which is also often denoted as $\mathbb{P}(\{X \in A\})$. So in the particular case of the game of darts, the answer to the first question would be that the probability that $S$ equals 180 is

$$\mathbb{P}^{\otimes 3}(\{S = 180\}) = \mathbb{P}^{\otimes 3}(\{X_1 = 60\} \cap \{X_2 = 60\} \cap \{X_3 = 60\}) = \mathbb{P}(\{X = 60\})^3.$$

To answer the second question, we need to condition the distribution of $S$ with respect to the event that $X_1$ equals 60. Therefore we find that the answer is equal to

$$\mathbb{P}^{\otimes 3}(\{S = 180\} \mid \{X_1 = 60\}) = \frac{\mathbb{P}^{\otimes 3}(\{S = 180\} \cap \{X_1 = 60\})}{\mathbb{P}^{\otimes 3}(\{X_1 = 60\})} = \mathbb{P}(\{X = 60\})^2.$$

Given two random variables $X : \Omega \to E$ and $Y : \Omega \to E$, there is a random variable $(X, Y) : \Omega \to E^2$. The distribution of $(X, Y)$ is called the **joint distribution of $X$ and $Y$**. If the joint distribution of $X$ and $Y$ is equal to the product of the distribution of $X$ with the distribution of $Y$, i.e. $\mathbb{P} \circ (X, Y)^{-1} = \mathbb{P} \circ X^{-1} \otimes \mathbb{P} \circ Y^{-1}$, then we say that $X$ and $Y$ are **independent** random variables. We see that $X_1, X_2$ and $X_3$ are pairwise independent random variables, but $X_1$ and $S$ are not independent.

To answers the last two questions of the darts example, we need to make sure that we are working with random variables taking values in the *real line* (or a sufficiently nice Banach space). For a random variable $X : \Omega \to \mathbb{R}$, we define the **expectation of $X$** as

$$\mathbb{E}[X] := \int X(\omega)\mathbb{P}(\mathrm{d}\omega),$$

whenever it exists. We will also use the notation $\int X \mathrm{d}\mathbb{P}$. For the random variable $S$ of the darts game we find that $\mathbb{E}[S] = 3\mathbb{E}[X]$. To find the expectation of $S$ given that $X_1 = 60$ we calculate as follows:

$$\begin{aligned} \mathbb{E}[S \mid X_1 = 60] &= \frac{1}{\mathbb{P}^{\otimes 3}(\{X_1 = 60\})} \int_{\{X_1 = 60\}} S \mathrm{d}\mathbb{P}^{\otimes 3} \\ &= 60 + 2\mathbb{E}[X]. \end{aligned}$$

We see that the conditional expectation of a random variable $X$ with respect to an event $A$, such that $\mathbb{P}(A) > 0$, can therefore be defined as

$$\mathbb{E}[X \mid A] := \frac{1}{\mathbb{P}(A)} \int_A X \mathrm{d}\mathbb{P}.$$

However, there is a more general concept of *conditional expectation.* Instead of conditioning with respect to *one event*, we will condition with respect to a whole $\sigma$-subalgebra of $\mathcal{F}$. Let $X : \Omega \to \mathbb{R}$ be an integrable, random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a $\sigma$-subalgebra. Define the probability space $\mathbf{\Omega} \mid_{\mathcal{G}} := (\Omega, \mathcal{G}, \mathbb{P} \mid_{\mathcal{G}})$. Suppose there is a $\mathcal{G}$-measurable random variable $Y : \Omega \to \mathbb{R}$ such that

$$\int_A X \mathrm{d}\mathbb{P} = \int_A Y \mathrm{d}\mathbb{P} \mid_{\mathcal{G}}$$

for every $A$ in $\mathcal{G}$, then $Y$ is called the **conditional expectation of $X$ with respect to $\mathcal{G}$**. It follows immediately that if such a random variable exists, it has to be $\mathbb{P} \mid_{\mathcal{G}}$**-almost surely** unique. This means that if we have $Y_1$ and $Y_2$ satisfying the condition, we have that $\mathbb{P} \mid_{\mathcal{G}} (Y_1 = Y_2) = 1$. If such a random variable exists, we denote it as $\mathbb{E}[X \mid \mathcal{G}]$.

To motivate this definition, consider the $\sigma$-subalgebra $\mathcal{G}_A$ generated by one measurable subset $A$ of $\Omega$, i.e.

$$\mathcal{G}_A := \{\emptyset, A, A^{\mathrm{C}}, \Omega\}.$$

Then define $Y : \Omega \to \mathbb{R}$ by the assignment

$$\omega \mapsto \begin{cases} \mathbb{E}[X \mid A] & \text{if } \omega \in A \\ \mathbb{E}[X \mid A^{\mathrm{C}}] & \text{if } \omega \in A^{\mathrm{C}}. \end{cases}$$

Clearly, $Y$ is measurable with respect to $\mathcal{G}_A$. Furthermore,

$$\int_A Y \mathrm{d}\mathbb{P} \mid_{\mathcal{G}_A} = \mathbb{E}[X \mid A]\mathbb{P}(A) = \int_A X \mathrm{d}\mathbb{P}$$

and similar equalities for the other elements of $\mathcal{G}_A$. Therefore, we see that the concept of conditioning with respect to an event can be recovered from the general conditional expectation with respect to a $\sigma$-subalgebra.

**Example 1.1.8.** Let us look again at the darts example. A dartboard can be divided in five areas: the area of single points $S_1$, the double ring $S_2$, the triple ring $S_3$, the outer bullseye $S_4$ and the inner bullseye $S_5$. Let $\mathcal{G}$ be the $\sigma$-subalgebra of $\mathcal{F}$ generated by these areas. We can then define

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \mathbb{E}[X \mid S_i]$$

for $\omega \in S_i$. For a point $\omega$ in the triple ring we have

$$\begin{aligned} \mathbb{E}[X \mid \mathcal{G}](\omega) &= \frac{1}{\lambda(S_3)} \int_{S_3} X \mathrm{d}\lambda \\ & \frac{1}{\lambda(S_3)} \sum_{k=1}^{20} \frac{\lambda(S_3)}{20} 3k = \frac{63}{2} \end{aligned}$$

and for $\omega$ in the inner bullseye we have

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \frac{1}{\lambda(S_5)} \int_{S_5} X \mathrm{d}\lambda = 50.$$

Clearly, we also have for every $i \in \{1, \ldots, 5\}$ that

$$\int_{S_i} \mathbb{E}[X \mid \mathcal{G}] \mathrm{d}\lambda \mid_{\mathcal{G}} = \mathbb{E}[X \mid S_i]\lambda(S_i) = \int_{S_i} X \mathrm{d}\lambda$$

which shows that the defined random variable is indeed the conditional expectation of $X$ with respect to $\mathcal{G}$.

The conditional expectation of a real-valued random variable that has finite expectation exists. However, the proof for the existence of the conditional expectation is not obvious. One way to prove the existence relies on the Radon–Nikodym theorem. This is Theorem 4.2.4 in [11]. One of the conditions of this theorem requires the notion of *absolute continuity* of a measure with respect to another measure. A measure $\mu$ is **abosulely continuous** with respect to a probability measure $\mathbb{P}$ if $\mu(A) = 0$ for every $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$.

**Theorem 1.1.9** (Radon–Nikodym theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mu$ be a finite signed measure on $(\Omega, \mathcal{F})$ such that $\mu \ll \mathbb{P}$. Then there exists a $\mathbb{P}$-almost surely unique random variable $X : \Omega \to \mathbb{R}$ with finite expectation such that*

$$\mu(A) = \int_A X \, d\mathbb{P}$$

*for all $A \in \mathcal{F}$.*

The random variable we obtain from the Radon–Nikodym theorem is called the **Radon–Nikodym derivative of $\mu$ with respect to** $\mathbb{P}$ and is usually denoted as $\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}$.

**Proposition 1.1.10** (Existence of conditional expectation)**.** *Let $X$ be a random variable with finite expectation and let $\mathcal{G}$ be a $\sigma$-subalgebra of $\mathcal{F}$. Then the conditional expectation of $X$ with respect to $\mathcal{G}$ exists.*

*Proof.* We can define a measure $\mu$ on $(\Omega, \mathcal{G})$ by the assignment $\mu(A) := \int_A X \mathrm{d}\mathbb{P}$ for every $A$ in $\mathcal{G}$. It is clear that $\mu$ is absolutely continuous with respect to $\mathbb{P} \mid_{\mathbb{G}}$. It follows now from the Radon–Nikodym theorem that there exists a $\mathbb{P}_{\mathcal{G}}$-almost surely unique random variable $Y$ (which is $\mathcal{G}$-measurable), such that for all $A \in \mathcal{G}$

$$\int_A X \mathrm{d}\mathbb{P} = \mu(A) = \int_A Y \mathrm{d}\mathbb{P}_{\mathcal{G}}.$$

It follows that $Y$ is the conditional expecation of $X$ with respect to $\mathcal{G}$. ☐

Taking the conditional expectation with respect to $\sigma$-subalgebras has many nice properties, as described in the following proposition. This is Proposition 10.4.3 in [11]. Here we use the concept of *independent* $\sigma$-subalgebras. Two $\sigma$-subalgebras $\mathcal{G}_1$ and $\mathcal{G}_2$ are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

for every $A \in \mathcal{G}_1$ and $B \in \mathcal{G}_2$.

**Proposition 1.1.11.** *Let $X$ and $Y$ be real random variables with finite expectation and let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be $\sigma$-subalgebras. The following properties hold:*

- $\mathbb{E}[X + Y \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[Y \mid \mathcal{G}]$, $\mathbb{P}$-*almost surely.*
- *If $X \leq Y$, then $\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]$, $\mathbb{P}$-almost surely.*
- $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$, $\mathbb{P}$-*almost surely.*
- $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$.

- *If $X$ is $\mathcal{G}$-measurable, then $\mathbb{E}[X \mid \mathcal{G}] = X$, $\mathbb{P}$-almost surely.*
- *If $\sigma(X)$, i.e. the smallest $\sigma$-algebra making $X$ measurable, and $\mathcal{G}$ are independent, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$, $\mathbb{P}$-almost surely.*

So far we have seen that random variables form an ideal tool to describe dependence and interaction between stochastic events. But there are more advantages to working with random variables. They allow us to describe stochastic events that evolve over time, by considering *stochastic processes.* One of the key interests in studying stochastic processes is their long-term behaviour. We are interested in the convergence properties of stochastic processes when time goes to infinity.

Let $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ be a fixed probability space and let $E$ be a measurable space. A collection of random variables $(X_t : \Omega \to E \mid t \in [0, \infty))$ is called a **stochastic process**. One of the most important stochastic processes is *Brownian motion*, which is used to model the location of particles suspended in a liquid or gas. Mathematically, **Brownian motion** is axiomatised in the following way. It is a stochastic process $(B_t : \Omega \to \mathbb{R})_t$ such that:

1. For $t_1 < t_2$ and $0 < s$, $B_{t_2+s} - B_{t_2}$ is a Gaussian random variable with mean 0 and variance $s$ and is independent from $B_{t_1}$.
2. The assignment $t \mapsto B_t$ is $\mathbb{P}$-almost surely continuous, i.e.

$$\mathbb{P}(\{\omega \in \Omega \mid t \mapsto B_t(\omega) \text{ is continuous}\}) = 1.$$

Proving that such a stochastic process exists is again not trivial. Part of the proof of the existence is given in Example 1.1.7. To also prove the second axiom, saying that $\mathbb{P}$-almost surely every sample path is continuous requires Kolmogorov's continuity criterion. This is explained in Chapter 7 of [16].

For $t \in [0, \infty)$, let $\mathcal{F}_t$ be the smallest $\sigma$-algebra that makes $B_s$ measurable for all $s < t$. Note that $B_{t+s} - B_t$ is independent from the $\sigma$-algebra $\mathcal{F}_t$. By Proposition 1.1.11, it follows that

$$\mathbb{E}[B_{t+s} \mid \mathcal{F}_t] = \mathbb{E}[B_{t+s} - B_t \mid \mathcal{F}_t] + \mathbb{E}[B_t \mid \mathcal{F}_t] = 0 + X_t.$$

Stochastic process with a property like this form an important class of processes, namely *martingales.* These processes are usually defined in the context of *filtered probability spaces.*

Let $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $I \in \{\mathbb{N}, [0, \infty)\}$. An indexed collection $(\mathcal{F}_t \mid t \in I)$ of $\sigma$-subalgebras of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$ and such that

$$\sigma\left(\bigcup_t \mathcal{F}_t\right) = \mathcal{F}$$

is called a **filtration of $\mathbf{\Omega}$**. A quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ where $(\mathcal{F}_t)_t$ is a filtration of $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **filtered probability space**. In more categorical terms, we have a functor $D : [0, \infty)^{\mathrm{op}} \to \mathbf{Prob}$ with $D(t) = (\Omega, \mathcal{F}_t, \mathbb{P} \mid_{\mathcal{F}_t})$ such that the limit of $D$ is precisely $(\Omega, \mathcal{F}, \mathbb{P})$.

If we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$, then we say that a stochastic process $(X_t : \Omega \to \mathbb{R})_t$ is **adapted (to $(\mathcal{F}_t)_t$)** if $X_t$ is $\mathcal{F}_t$-measurable for every $t \in [0, \infty)$. Suppose now that we have an adapted stochastic process $(X_t)_t$ such that every random variable has finite expectation. Then we say that the process is a **martingale** if

$$\mathbb{E}[X_{t+s} \mid \mathcal{F}_t] = X_t \quad \mathbb{P}\text{-almost surely},$$

for all $t \in [0, \infty)$ and $0 < s$.

We already saw that a Brownian motion is a martingale. However not every stochastic process is a martingale as shown in the following example.

**Example 1.1.12.** Let $(B_t)_t$ be a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_t, \mathbb{P})$. Consider the adapted stochastic process $(B_t^2)_t$. For $s < t$ we find that

$$\begin{aligned}\mathbb{E}[B_t^2 \mid \mathcal{F}_s] &= \mathbb{E}[(B_t - B_s + B_s)^2 \mid \mathcal{F}_s] \\ &= \mathbb{E}[(B_t - B_s)^2 \mid \mathcal{F}_s] + \mathbb{E}[B_s^2 \mid \mathcal{F}_s] + \mathbb{E}[2B_s(B_t - B_s) \mid \mathcal{F}_s] \\ &= \mathbb{E}[(B_t - B_s)^2] + B_s^2 + 2B_s\mathbb{E}[B_t - B_s] \\ &= (t-s) + B_s^2\end{aligned}$$

Here we used that $B_s^2$ is $\mathcal{F}_s$-measurable and that $B_t - B_s$ is a Gaussian random variable with mean 0 and variance $t - s$ which is independent from $\mathcal{F}_s$ together with Proposition 1.1.11.

We say that a martingale is **continuous** if

$$\mathbb{P}(\{\omega \in \Omega \mid t \mapsto X_t(\omega) \text{ is continuous}\}) = 1.$$

In a similar way we can define **cadlag**[1] martingales. An important reason to consider martingales is because they have nice convergence properties, as described by Doob's martingale convergence theorems. Doob's orginial proof can be found in Section XI.14 of [13].

**Theorem 1.1.13** (Almost sure martingale convergence theorem)**.** *Let $(X_t)_t$ be a cadlag martingale such that*

$$\sup_t \mathbb{E}[|X_t|] < \infty.$$

*Then there exist a random variable $X : \Omega \to \mathbb{R}$ such that $X_t \to X$ $\mathbb{P}$-almost surely as $t \to \infty$.*

**Theorem 1.1.14** ($L^1$ martingale convergence theorem)**.** *Let $(X_t)_t$ be a cadlag martingale such that*

$$\lim_{\lambda \to \infty} \sup_t \mathbb{E}\left[|X_t| 1_{\{|X_t| > \lambda\}})\right] = 0.$$

*Then there exists a random variable $X$ such that $\mathbb{E}[|X|] < \infty$ and such that $X_t \to X$ in $L^1$ and $\mathbb{P}$-almost surely as $t \to \infty$. Moreover, for all $t \in [0, \infty)$,*

$$\mathbb{E}[X \mid \mathcal{F}_t] = X_t \quad \mathbb{P}\text{-almost surely.}$$

There are several variations on this. For example the *$L^p$-martingale convergence theorem* and the *backwards martingale convergence theorem.*

Another main result about the convergence of stochastic processes is the *strong law of large numbers.* Intuitively this result says that the average outcome of applying an experiment a large number of times is close to the true expectation. This result can be proved using a variation of the martingale convergence theorem (*the backward martingale convergence theorem*). This proof can be found in Section XI.19 of [13].

**Theorem 1.1.15** (Strong law of large numbers)**.** *Let $(X_n : \Omega \to \mathbb{R})_n$ be a sequence of identically distributed, pairwise independent, integrable random variables. Then*

$$\frac{1}{N}\sum_{n=1}^{N} X_n \to \mathbb{E}[X_1] \qquad \text{as } N \to \infty$$

---

[1] *Continue à droite, limite à gauche*, i.e. right continuous with left limits.

$\mathbb{P}$-*almost surely and in* $L^1$.

Moreover, this is another result in probability theory, where finite approximations determine the global result.

We can conclude that random variables provide a far better way to model stochastic events than probability spaces. Not only can we describe dependence between stochastic events, we can also say a lot about the expected outcome of a stochastic event. Moreover, the theory of random variables leads to a theory of stochastic processes, which can be used to model complex stochastic events that change over time.

On the other hand, the framework of random variables is harder to describe from a category theory point of view. There are no obvious objects and morphisms between them, that could be considered as a category. Furthermore, we are in the situation where we are looking at maps where the domain and codomain seem to live in different worlds. A random variable has a domain that is associated to a probability space and a codomain that is related to a more structured space (such as a Banach space or metric space). Furthermore, the concept of random variables requires us to fix a probability space. This gives a certain upper bound on the possible randomness that can be modelled using random variables. For example, if one would consider a probability space whose underlying probability measure is a Dirac delta, the random variables on the probability space could only model (deterministic) elements in the space of outcomes.

### 1.1.3 Probability theory is about Markov kernels

We saw that probability spaces were not sufficient to model all aspects of stochastic events and that random variables provide a framework that is powerful enough to solve this issue. One of the main problems that random variables solve is the modelling of interactions and dependence between stochastic events. Another approach to describe this is given by Markov kernels. The idea of dependency between stochastic events is baked into definition of a Markov kernel.

A Markov kernel describes the transition from an experiment with outcomes in a measurable space $X$ to an experiment, with outcomes in a measurable space $Y$, that *depends* on the outcome of the first experiment.

Let $\mathcal{G}Y$ be the space of all probability measures on $Y$ together with the $\sigma$-algebra generated by the evaluation maps

$$\begin{array}{rccc} \mathrm{ev}_A : & \mathcal{G}Y & \to & [0,1] \\ & \mathbb{P} & \mapsto & \mathbb{P}(A) \end{array}$$

for all measurable subsets $A$ of $Y$. A **Markov kernel** from $X$ to $Y$ is a measurable map $X \to \mathcal{G}Y$. We will denote this as $X \rightsquigarrow Y$. Suppose we have three experiments in a row, where the second one depends on the outcome of the first one and the third one depends on the outcome of the second one. We could ask how the last experiment depends on the outcome of the first one. This can be described mathematically by the *composition of Markov kernels*. For Markov kernels $f : X \rightsquigarrow Y$ and $g : Y \rightsquigarrow Z$, the **composite of Markov kernels** $f$ and $g$ is a Markov kernel $g \circ f : X \rightsquigarrow Z$ that sends an element $x$ to the probability measure on $Z$ defined by

$$A \mapsto \int [g(y)](A)[f(x)](\mathrm{d}y).$$

The following example is Example 1.1 in [46].

**Example 1.1.16.** Consider a frog that lives in a pond with two lily pads, *east* and *west*. Let $X := \{e, w\}$ be the space representing these lily pads. Every morning the frog tosses a coin. If the coin lands heads up, the frog stays on the lily pad where he is. If the coin lands on tails, the frog jumps to the other lily pad.

Let $p$ be the probability that the coin lands heads up. This stochastic event can be described by a Markov kernel $f : X \rightsquigarrow X$, defined by

$$[f(e)](\{e\}) = p, \qquad [f(e)](\{w\}) = 1 - p$$

$$[f(w)](\{e\}) = 1 - p \quad \text{and} \quad [f(w)](\{w\}) = p.$$

We can also describe the transition over a period of two days by composing $f$ with itself. This gives us a Markov kernel $f \circ f : X \rightsquigarrow X$.

$$\begin{aligned} [f \circ f](e)(\{e\}) &= \int [f(x)](\{e\})[f(e)](\mathrm{d}x) \\ &= ([f(e)](\{e\}))\,([f(e)](\{e\})) + ([f(w)](\{e\}))\,([f(e)](\{w\})) \\ &= p^2 + (1-p)^2 \end{aligned}$$

The fact that we have a meaningful composition indicates that there might be a category in which Markov kernels are morphisms. Since Markov kernels represent measurable maps of the form $X \to \mathcal{G}Y$, they remind us of Kleisli morphisms. In Chapter 2, we will see that this is indeed the case. Markov kernels are precisely morphisms in the Kleisli category of the Giry monad and the composition of Markov kernels corresponds to the Kleisli composition of these morphisms. Categories of this form were generalized by Fritz in [20] to *Markov categories*, which we will discuss in Chapter 2.

There are several connections between Markov kernels and random variables and probability spaces. Consider for example a Markov kernel $f : X \rightsquigarrow Y$ and a probability measure $\mathbb{P}$ on $X$. Then we can define a joint probability distribution on $X \times Y$ as follows

$$A \times B \mapsto \int_A [f(x)](B)\mathbb{P}(\mathrm{d}x)$$

for all measurable subsets $A$ of $X$ and $B$ of $Y$. It is not clear if this also works in the other direction. Suppose we have a joint probability distribution $\mathbb{Q}$ on $X \times Y$. We are now wondering if $\mathbb{Q}$ arises from a Markov kernel. Consider the projection map $\pi_X : X \times Y \to X$ and let $\mathcal{F}_X$ be the $\sigma$-algebra of $X$. We could *attempt* to define a Markov kernel $f : X \rightsquigarrow Y$ by the assignment

$$[f(x)](B) := \mathbb{E}[1_{X \times B} \mid \pi_X^{-1}(\mathcal{F}_X)](x).$$

Here $1_{X \times B}$ is the indicator function $X \times Y \to [0, 1]$ which is 1 on $X \times B$ and 0 otherwise. In this case,

$$\begin{aligned} \int_A [f(x)](B)[\mathbb{Q} \circ \pi_X^{-1}](\mathrm{d}x) &= \int_{\pi_X^{-1}(A)} 1_{X \times B} \mathrm{d}\mathbb{Q} \\ &= \int 1_{A \times B} \mathrm{d}\mathbb{Q} = \mathbb{Q}(A \times B). \end{aligned}$$

However, since conditional expectation is only defined $\mathbb{Q}$-almost surely, we need to choose a representative for every conditional expectation if we actually want to evaluate it. The problem now is that we need to choose these representatives in a way such that $f(x)$ is $\sigma$-additive for *every* $x \in X$ (not just $\mathbb{Q} \circ \pi_X^{-1}$-almost everywhere). This is very far from trivial, but under certain conditions it is possible. This leads to a very powerful theorem, the *disintegration theorem*. This is Corollary 452N in [18].

**Theorem 1.1.17** (Disintegration theorem)**.** *Suppose that $X$ and $Y$ are standard Borel spaces and let $\mathbb{Q}$ be a joint probability distribution on $X \times Y$. Then there exists a $\mathbb{Q} \circ \pi_X^{-1}$-almost surely unique Markov kernel $f : X \rightsquigarrow Y$ such that*

$$\mathbb{Q}(A \times B) = \int_A [f(x)](B)[\mathbb{Q} \circ \pi_X^{-1}](dx)$$

We just described the relation between Markov kernels and joint probability distributions. There is also a connection between Markov kernels and a special class of stochastic processes, namely *Markov chains*. These processes are *memoryless* in the sense that the future *only* depends on the present and not on the past. More formally we can define a Markov process as follows. Let $E$ be the standard Borel space of outcomes of the stochastic process and $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ be a filtered probability space. A stochastic process $(X_n : \Omega \to E \mid n \in \mathbb{N})$ is a **Markov chain** if

$$\mathbb{E}\left[f(X_{n+1}) \mid \mathcal{F}_n\right] = \mathbb{E}[f(X_{n+1}) \mid \sigma(X_n)]$$

for all bounded measurable $f : E \to \mathbb{R}$. Here $\sigma(X_n)$ is the smallest $\sigma$-algebra making $X_n$ measurable. In a similar style to the disintegration theorem, it can be shown that for every $n$ there exists a Markov kernel $f_n : E \rightsquigarrow \mathcal{G}E$ such that

$$f(-)(A) = \mathbb{E}[1_{X_{n+1}^{-1}(A)} \mid \sigma(X_n)] \quad [\mathbb{P} \circ X_{n+1}^{-1}]\text{-almost surely.}$$

Such a Markov kernel is called a **regular conditional probability**

We can also go the other way around. Suppose we have a collection of Markov kernels $(f_n : E \rightsquigarrow E)_n$ and an initial probability distribution $\mathbb{P}_0$. Then by the Ionescu-Tulcea theorem (Theorem 14.32 in [40]), there exists a probability distribution $\mathbb{P}$ on $E^{\mathbb{N}}$ such that $(\pi_n : E^{\mathbb{N}} \to E)_n$ is a Markov chain and its corresponding regular conditional probabilities are precisely $(f_n)_n$.

When studying stochastic processes, one is often interested in its behaviour when time tends to infinity. This is one of the reasons why the class of Markov chains is important, since they often have nice properties about convergence to the *stationary distribution* and about *mixing times*.

We can conclude that Markov kernels are a powerful way to model stochastic events, their interactions and dependence. Even when stochastic events change in time, Markov kernels often can still be used. Furthermore, we already indicated that Markov kernels have an interesting categorical interpretation, which we will discuss in detail in Chapter 2. We also do not need to fix a probability space, causing us to not have to put an a priori *bound* on the possible randomness of our model.

### 1.1.4 Conclusion

We conclude that probability theory is not about one central object. Different objects and concepts are important and we require all of them to truly understand probability theory. Probability spaces are the least complete model to describe randomness. Random variables and Markov kernels offer more complete approaches, however each approach has its advantages and disadvantages depending on the type of stochastic event we want to model. Moreover, the concept of random variables relies on the idea of probability spaces.

Furthermore, some of these concepts are easier to express categorically than others. Probability spaces and measure-preserving maps between them form a category in an obvious way and Markov kernels can be recognized as the Kleisli category of a probability monad. Random variables, however, are more difficult to treat categorically, since the domain and codomain seem to live in different categories.

## 1.2 Kan extensions

In this section we will give an overview of Kan extensions, an important concept in category theory. Kan extensions are ubiquitous in category theory and mathematics. The original concept of Kan extensions deals with the problem of extending '*ordinary*' functors between categories in a universal way. We study these in Section 1.2.1. However, the concept of Kan extension has been generalized to *enriched* functors between *enriched* categories, which we discuss in Section 1.2.2. A special case of this is given by enriching over **Cat** or **Pos**. In Section 1.2.3, we will briefly discuss a particular example of this, that will be important in Chapter 4. In this example we will focus on extensions of *lax natural transformations*.

### 1.2.1 Kan extensions of ordinary functors

In this section we will discuss the classical Kan extensions of functors between categories. Let $\mathcal{C}, \mathcal{D}$ and $\mathcal{E}$ be categories and let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{E}$ be functors. We are interested in functors $\mathcal{D} \to \mathcal{E}$ with a universal property, such that they can be interpreted as the *extreme* possible extensions of $G$ along $F$. These are referred to as *Kan extensions.* We will focus on the *right* Kan extension. However, everything can be dualized to *left* Kan extensions.

**Definition 1.2.1** (Local Kan extensions)**.** A **right Kan extension of** $G$ **along** $F$ is a functor $R : \mathcal{D} \to \mathcal{E}$ together with a natural isomorphism

$$[\mathcal{C}, \mathcal{E}](- \circ F, G) \cong [\mathcal{D}, \mathcal{E}](-, R).$$

Therefore, by the Yoneda lemma, a right Kan extension of $G$ along $F$ is a pair $(R, \epsilon)$, where $R : \mathcal{D} \to \mathcal{E}$ is a functor and $\epsilon : R \circ F \Rightarrow G$ is a natural transformation, such that for every functor $\tilde{R} : \mathcal{D} \to \mathcal{E}$ and natural transformation $\tilde{\epsilon} : \tilde{R} \circ F \Rightarrow G$, *there exists a unique* natural transformation $\delta : R \Rightarrow \tilde{R}$ such that

$$\begin{tikzcd}
\mathcal{C} \arrow[rr, "G"] \arrow[dr, "F"'] & {} \arrow[d, phantom, ""{name=A}] & \mathcal{E} \\
& \mathcal{D} \arrow[ur, dashed, "R"{name=R}] \arrow[ur, bend right=40, "\tilde{R}"'{name=T}] \arrow[Rightarrow, from=2-2, to=1-2, shorten=12pt, "\epsilon"] \arrow[Rightarrow, from=T, to=R, "\delta"'] &
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{C} \arrow[rr, "G"] \arrow[dr, "F"'] & {} & \mathcal{E} \\
& \mathcal{D} \arrow[ur, "\tilde{R}"'] \arrow[Rightarrow, from=2-2, to=1-2, shorten=12pt, "\tilde{\epsilon}"] &
\end{tikzcd}$$

This universal property determines the right Kan extension *up to natural isomorphism*, therefore we will usually talk about *the* right Kan extension of $G$ along $F$. The functor $R : \mathcal{D} \to \mathcal{E}$ is denoted as $\mathrm{Ran}_F G$.

**Definition 1.2.2.** Let $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{E}_1$ be functors and let $(R, \epsilon)$ be the right Kan extension of $G$ along $F$. A functor $H : \mathcal{E}_1 \to \mathcal{E}_2$ **preserves** the right Kan extension of $G$ along $F$, if $(HR, H\epsilon)$ is a right Kan extension of $HG$ along $R$.

Let $G_1, G_2 : \mathcal{C} \to \mathcal{E}$ be functors and let $\tau : G_1 \Rightarrow G_2$ be a natural transformation. This induces a natural transformation

$$[\mathcal{C}, \mathcal{E}](- \circ F, G_1) \Rightarrow [\mathcal{C}, \mathcal{D}](- \circ F, G_2).$$

If the right Kan extensions of $G_1$ and $G_2$ along $F$ exist, we obtain a natural transformation $[\mathcal{D}, \mathcal{E}](-, \mathrm{Ran}_F, G_1) \Rightarrow [\mathcal{D}, \mathcal{E}](-, \mathrm{Ran}_F, G_2)$. Because the Yoneda embedding is full and faithful, this correspond to a unique natural transformation $\mathrm{Ran}_F G_1 \Rightarrow \mathrm{Ran}_F G_2$, which is denoted by $\mathrm{Ran}_F \tau$.

Using a similar argument, we see that $\mathrm{Ran}_F \tau_2 \circ \mathrm{Ran}_F \tau_1 = \mathrm{Ran}_F(\tau_2 \circ \tau_1)$ for natural transformations $\tau_1 : G_1 \Rightarrow G_2$ and $\tau_2 : G_2 \Rightarrow G_3$ between functors $G_1, G_2, G_3 : \mathcal{C} \to \mathcal{E}$.

This indicates that, *if the right Kan extension of every functor along $F$ exists*, it defines a functor $\mathrm{Ran}_F - : [\mathcal{C}, \mathcal{E}] \to [\mathcal{D}, \mathcal{E}]$ that is *right* adjoint to precomposition with $F$.

**Proposition 1.2.3** (Global Kan extension)**.** *Suppose that the functor $- \circ F : [\mathcal{D}, \mathcal{E}] \to [\mathcal{C}, \mathcal{D}]$ has a right adjoint $R$. Let $G : \mathcal{C} \to \mathcal{E}$ be a functor. Then $R(G)$ together with the counit $\epsilon_G : R(G) \circ F \Rightarrow G$ is a right Kan extension of $G$ along $F$.*

Under certain assumptions on the categories $\mathcal{D}$ and $\mathcal{E}$, we can give concrete constructions of these Kan extensions. In this case, the Kan extensions are called **pointwise**.

The following result, which is Theorem X.3.1 in [49] (essentially verbatim), constructs the right Kan extension of a functor $G$ along $F$ as a limit.

**Theorem 1.2.4** (Right Kan extension as a pointwise limit)**.** *Given $F : \mathcal{C} \to \mathcal{D}$, let $G : \mathcal{C} \to \mathcal{E}$ be a functor such that the limit*

$$Rd := \lim \left( d \downarrow F \to \mathcal{C} \xrightarrow{G} \mathcal{E} \right)$$

*exists for all $d \in \mathcal{D}$, with limiting cone $(\lambda_f : Rd \to Gc)_{f : d \to Fc}$. Each morphism $f : d_1 \to d_2$ induces a unique arrow*

$$Rf : Rd_1 \to Rd_2,$$

*commuting with the limiting cones. These formulas define a functor $R : \mathcal{D} \to \mathcal{E}$, and for each $c \in \mathcal{C}$ the components of the limiting cones $\lambda_{1_{Fc}} =: \epsilon_c$ define a natural transformation $\epsilon : R \circ F \Rightarrow G$, and $(R, \epsilon)$ is a right Kan extension of $G$ along $F$.*

From this we have the following corollaries, which are Corollary X.3.2 and Corollary X.3.3 from [49].

**Corollary 1.2.5.** *If $\mathcal{C}$ is small and $\mathcal{E}$ is complete, any functor $G : \mathcal{C} \to \mathcal{E}$ has a right Kan extension along $F$.*

**Corollary 1.2.6.** *If the functor $F$ is full and faithful, then the universal arrow $\epsilon : \mathrm{R} \circ F \Rightarrow G$ is a natural isomorphism.*

The concept of Kan extensions is very useful. Every universal construction can be written in terms of Kan extensions, or —in Mac Lane's words—*all concepts are Kan extensions*"[49]. The following examples are Theorem 1 and Theorem 2 and Exercise 3 in Chapter 7 of [49] respectively.

**Example 1.2.7.** A functor $F : \mathcal{C} \to \mathcal{E}$ has a left adjoint if and only if the right Kan extension of $1_{\mathcal{C}} : \mathcal{C} \to \mathcal{C}$ along $F$ exists and is preserved by $F$.

**Example 1.2.8.** The right Kan extension of a functor $D : \mathcal{C} \to \mathcal{E}$ along $! : \mathcal{C} \to \mathbf{1}$ is the limit of $D$.

**Example 1.2.9.** A functor $i : \mathcal{C} \to \mathcal{D}$ is codense if and only if $1_{\mathcal{D}}$ is the right Kan extension of $i$ along itself.

Let $F : \mathcal{C} \to \mathcal{D}$, be a functor with a left adjoint $L : \mathcal{D} \to \mathcal{C}$. Then the endofunctor $FL : \mathcal{D} \to \mathcal{D}$ has a canonical monad structure induced by the unit $\eta$ and counit $\epsilon$ of the adjunction. The unit of the adjunction becomes the unit of the monad, and $F\epsilon_G$ becomes the multiplication of the monad.

This observation can be generalized to a functor $F : \mathcal{C} \to \mathcal{D}$, without a left adjoint. Suppose that the right Kan extension of $F$ along itself exists. Then the endofunctor $\mathrm{Ran}_F F : \mathcal{D} \to \mathcal{D}$ has a canonical monad structure. Monads that arise like this are called *codensity monads.*

Let $(T^F, \epsilon)$ be a right Kan extension of $F$ along $F$. There is a natural transformation $1_F : 1_{\mathcal{D}} \circ F \Rightarrow F$, and therefore by the universal property of right Kan extensions there is a natural transformation $\eta : 1_{\mathcal{D}} \to T^F$ such that

$$\begin{tikzcd}
\mathcal{C} \arrow[rr, "F"] \arrow[dr, "F"'] & & \mathcal{D} \\
& \mathcal{D} \arrow[ur, "1_{\mathcal{D}}"'] \arrow[u, Rightarrow, "1_F"'] &
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{C} \arrow[rr, "F"] \arrow[dr, "F"'] & & \mathcal{D} \\
& \mathcal{D} \arrow[ur, dashed, "T^F"{name=T}] \arrow[ur, bend right=70, "1_{\mathcal{D}}"'{name=I}] \arrow[u, Rightarrow, "\epsilon"'] \arrow[from=I, to=T, Rightarrow, "\eta"] &
\end{tikzcd}$$

Furthermore, there is a natural transformation

$$T^F T^F \circ F \xRightarrow{T^F \epsilon} T^F \circ F \overset{\epsilon}{\Rightarrow} F.$$

This induces a natural transformation $\mu : T^F T^F \Rightarrow T^F$ such that

$$\begin{tikzcd}
\mathcal{C} \arrow[rr, "F"] \arrow[dr, "F"'] & & \mathcal{D} \\
& \mathcal{D} \arrow[ur, "T^F T^F"'] \arrow[u, Rightarrow] &
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{C} \arrow[rr, "F"] \arrow[dr, "F"'] & & \mathcal{D} \\
& \mathcal{D} \arrow[ur, dashed, "T^F"{name=T}] \arrow[ur, bend right=70, "T^F T^F"'{name=I}] \arrow[u, Rightarrow, "\epsilon"'] \arrow[from=I, to=T, Rightarrow, "\mu"] &
\end{tikzcd}$$

The triple $(T^F, \nu, \mu)$ forms a monad on $\mathcal{D}$, the **codensity monad of** $F$. More on this can be found in Section 2 of [45].

### 1.2.2 Kan extensions in enriched category theory

The idea of Kan extensions of '*ordinary*' functors between categories can be generalized to Kan extensions of *enriched* functors between *enriched* categories. This will be the topic of this section.

Let $\mathcal{V}$ be a closed monoidal category. For $\mathcal{V}$-enriched categories $\mathcal{C}, \mathcal{D}$ and $\mathcal{E}$ and $\mathcal{V}$-enriched functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{E}$, we want to study $\mathcal{V}$-*enriched* extensions of $G$ along $F$. To define right Kan extensions of $\mathcal{V}$-enriched functors we can essentially copy Definition 1.2.1 from the non-enriched case.

**Definition 1.2.10** (Local Kan extensions)**.** A **right Kan extension of** $G$ **along** $F$ is a $\mathcal{V}$-enriched functor $R : \mathcal{D} \to \mathcal{E}$ together with a $\mathcal{V}$-enriched natural transformation

$$[\mathcal{C}, \mathcal{E}](- \circ F, G) \cong [\mathcal{D}, \mathcal{E}](-, R).$$

By the same argument as for non-enriched Kan extensions, a right Kan extension of $G$ along $F$ is a pair $(R, \epsilon)$, where $R : \mathcal{D} \to \mathcal{E}$ is an enriched functor and $\epsilon : R \circ F \Rightarrow G$ is an enriched natural transformation, satisfying a universal property. This is also explained in Theorem 4.43 in [37].

Similarly to Proposition 1.2.3, we obtain a right adjoint to the functor $[\mathcal{D},\mathcal{E}] \to [\mathcal{C},\mathcal{E}]$ defined by precomposing with $F$, *if the right Kan extension of any functor along $F$ exists*, generalizing the concept of *global right Kan extensions*.

However, for $\mathcal{V}$-enriched functors, we can not use the limit formula from Theorem 1.2.4 anymore. However, we can rewrite the limit formula in terms of weighted limits to make it make sense in the $\mathcal{V}$-enriched setting. Suppose that the weighted limit

$$Rd := \{D(d, F-), G\}$$

exists for all $d \in \mathcal{D}$. The following canonical morphisms in $\mathcal{V}$ induce a $\mathcal{V}$-enriched functor $R : \mathcal{D} \to \mathcal{E}$.

$$\begin{aligned}\mathcal{D}(d_1, d_2) &\to [\mathcal{E},\mathcal{V}]\Big([\mathcal{C},\mathcal{V}]\big(\mathcal{D}(d_1, F-), \mathcal{E}(\bullet, G-)\big), [\mathcal{C},\mathcal{V}]\big(\mathcal{D}(d_2, F-), \mathcal{E}(\bullet, G-)\big)\Big)\\ &\cong [\mathcal{E},\mathcal{V}]\big(\mathcal{E}(\bullet, Rd_1), \mathcal{E}(\bullet, Rd_2)\big)\\ &\cong \mathcal{E}(Rd_1, Rd_2)\end{aligned}$$

We can define a $\mathcal{V}$-enriched natural transformation $\epsilon : R \circ F \to G$ by defining $\epsilon_c$ as follows, for $c \in \mathcal{C}$:

$$\begin{aligned}I &\to [\mathcal{C},\mathcal{V}]\big(\mathcal{C}(c,-), \mathcal{D}(Fc, G-)\big)\\ &\to [\mathcal{E},\mathcal{V}]\Big([\mathcal{C},\mathcal{V}]\big(\mathcal{D}(Fc, F-), \mathcal{E}(\bullet, G-)\big), [\mathcal{C},\mathcal{V}]\big(\mathcal{C}(c,-), \mathcal{E}(\bullet, G-)\big)\Big)\\ &\cong [\mathcal{E},\mathcal{V}]\big(\mathcal{E}(\bullet, R \circ Fc), \mathcal{E}(\bullet, Gc)\big)\\ &\cong \mathcal{E}(R \circ Fc, Gc)\end{aligned}$$

Then $(R, \epsilon)$ is a right Kan extension of $G$ along $F$. Indeed, let $\tilde{R} : \mathcal{D} \to \mathcal{E}$ be an enriched functor. Then we have the following isomorphisms in $\mathcal{V}$.

$$\begin{aligned}[\mathcal{C},\mathcal{E}](\tilde{R} \circ F, G) &\cong \int_{c\in\mathcal{C}} \mathcal{E}(\tilde{R} \circ Fc, Gc)\\ &\cong \int_{c\in\mathcal{C}} [\mathcal{D},\mathcal{V}]\big(\mathcal{D}(-, Fc), \mathcal{E}(\tilde{R}-, Gc)]\big)\\ &\cong \int_{c\in\mathcal{C}} \int_{d\in\mathcal{D}} \big[\mathcal{D}(d, Fc), \mathcal{E}(\tilde{R}d, Gc)\big]\\ &\cong \int_{d\in\mathcal{D}} [\mathcal{C},\mathcal{V}]\big(\mathcal{D}(d, F-), \mathcal{E}(\tilde{R}d, G-)\big)\\ &\cong \int_{d\in\mathcal{D}} \mathcal{E}(\tilde{R}d, Rd) \cong [\mathcal{D},\mathcal{E}](\tilde{R}, R).\end{aligned}$$

This is essentially Theorem 4.38 in [37]. Here we used the Fubini theorem for ends, which can be found in Section 2.1 of [37]. This isomorphism is natural in $\tilde{R}$, which shows the claim. Right Kan extensions that can be expressed using the weighted limit formula are called **pointwise right Kan extensions**[2].

---

[2] Note that Kelly uses this as the defintion of *right Kan extensions* and refers to our Definition 1.2.10 as *weak*

Several results from the non-enriched setting still hold for the enriched context. In particular, the enriched analogues of Corollary 1.2.5, Corollary 1.2.6 and Example 1.2.7 are given by Proposition 4.33, Proposition 4.23, Theorem 4.81 in [37] respectively.

Also the concept of codensity monads generalizes to the enriched context. By the exact same argument as for non-enriched functors, the right Kan extensions of an enriched functor $F$ has a canonical (enriched) monad structure.

### 1.2.3 Kan extensions of lax natural transformations

The Kan extensions in Section 1.2.1 live in the 2-category **Cat**, while the Kan extensions from Section 1.2.2 live in the 2-category $\mathcal{V} - \mathbf{Cat}$. This leads to the concept of Kan extensions in an arbitrary strict 2-category. However, Kan extensions can be defined in an arbitrary 2-category. We will look at a particular example of lax transformations between **Pos**-enriched functors.

Let $\mathcal{C}$ be a **Pos**-enriched category. Consider **Pos**-enriched functors $F, G : \mathcal{C} \to \mathbf{Pos}$, i.e. functors $F, G : \mathcal{C}_0 \to \mathbf{Pos}$ such that $Ff_1 \leq Ff_2$ and $Gf_1 \leq Gf_2$ for $f_1, \in f_2 \in \mathcal{C}(A, B)$ with $f_1 \leq f_2$.

A lax natural transformation $\lambda : F \Rightarrow G$ is a collection of morphisms $(\lambda_A : FA \to GA)_{A \in \mathcal{C}}$ such that

$$\begin{tikzcd}
FA \arrow[r, "\lambda_A"] \arrow[d, "Ff"'] \arrow[dr, phantom, "\leq"] & GA \arrow[d, "Gf"] \\
FB \arrow[r, "\lambda_B"] & GB
\end{tikzcd}$$

for all morphisms $f \in \mathcal{C}(A, B)$.

Let $[\mathcal{C}, \mathbf{Pos}]_l$ be the category that has **Pos**-enriched functors as objects and lax natural transformations as morphisms.

For lax natural transformations $\lambda_1, \lambda_2 : F \Rightarrow G$, we write $\lambda_1 \leq \lambda_2$ if and only if

$$(\lambda_1)_A(x) \leq (\lambda_2)_A(x)$$

for all objects $A$ in $\mathcal{C}$ and all elements $x$ in $FA$. This turns the set of lax natural transformations from $F$ to $G$ into a poset. We can conclude that $[\mathcal{C}, \mathbf{Pos}]_l$ is naturally enriched over **Pos**. Therefore, we can interpret $[\mathcal{C}, \mathbf{Pos}]_l$ as a 2-category.

This concept of Kan extensions in this particular 2-category leads to a definition of right Kan extensions of lax natural transformations. Let $F, G, H : \mathcal{C} \to \mathbf{Pos}$ be **Pos**-enriched functors and let $\lambda : F \Rightarrow H$ and $\iota : F \Rightarrow G$ be lax natural transformations. Then a right Kan extension of $\lambda$ along $\iota$ is given by a lax natural transformation $\tau : G \to H$ such that $\tau \circ \iota \leq \lambda$ and such that for every other lax natural transformation $\tilde{\tau}$ with $\tilde{\tau} \circ \iota \leq \lambda$, we have that

$$\tilde{\tau} \leq \tau.$$

---

*right Kan extensions* [37].

## Chapter 2

# Background in categorical probability theory

In this chapter we will give an overview of some results in categorical probability theory. We will start with an historical overview, where we discuss papers by Lawvere, Swirsczc, Semadeni and Giry. In the rest of the chapter we will discuss some more recent work. We will describe probability monads and the probabilistic interpretation of their Eilenberg-Moore algebras. In the last section we will give a short overview of the theory of Markov categories. This an important and useful approach to view probability theory categorically and has led to many results in recent years.

## 2.1 Historical overview of categorical probability theory

In this section we will give a selection of some of the earliest research that looks at probability theory from a categorical point of view. A lot of the ideas described in these works form the basis of modern research in categorical probability theory. The earliest of these dates back to the 1960s by Lawvere [44]. We will discuss Lawvere's notes in Section 2.1.1. In the 1970s, Swirszcz and Semadeni published several papers on categorical constructions in functional analysis, in particular on convexity. Their work describes certain probability monads and their algebras. We will discuss two published works in Section 2.1.2. In the 1980s Giry wrote a paper discussing a certain probability monad and its properties [28]. This monad is now known as the *Giry monad* and we discuss this in Section 2.1.3.

In this section we will mostly use the same notation and terminology used by the authors in the original papers. These might sometimes be different from the terminology and notation used in the rest of this thesis.

### 2.1.1 Lawvere

One of the very first descriptions of probability theory from a categorical perspective can be found in seminar notes from 1962 by Lawvere with the title 'The category of probabilistic mappings' [44]. In this note Lawvere defines a category of Markov kernels and uses this to describe Markov chains and stochastic processes. Lawvere defines a category $\mathcal{P}$ as follows:

- Objects: measurable spaces,

- Morphims: Markov kernels,
- Composition: composition of Markov kernels,
- Identities: for a measurable space $X$, the identity morphism $X \to X$ is given by the Markov kernel $X \rightsquigarrow X$ that sends $x$ to the Dirac delta $\delta_x$.

It is shown that these data indeed form a category and it is called the **category of probabilistic mappings**. Lawvere proceeds by defining a functor $\mathcal{D} : \mathcal{P} \to \mathbf{Mble}$. The functor $\mathcal{D}$ sends a measurable space $\Omega$ to the collection

$$\mathcal{D}\Omega := \{\text{probability measures on } \Omega\},$$

together with the smallest $\sigma$-algebra that makes

$$\begin{array}{rccc} \mathrm{ev}_A : & \mathcal{D}\Omega & \to & [0,1] \\ & \mathbb{P} & \mapsto & \mathbb{P}(A) \end{array}$$

measurable for every measurable subset $A$ of $\Omega$. On morphisms the functor $\mathcal{D}$ is defined by sending a Markov kernel $T : \Omega \rightsquigarrow \Omega'$ to the measurable map $\mathcal{D}f : \mathcal{D}\Omega \to \mathcal{D}\Omega'$ which maps a probability measure $\mathbb{P}$ on $\Omega$ to the probability measure on $\Omega'$ defined by

$$A \mapsto \int [T(x)](A)\mathbb{P}(\mathrm{d}x).$$

Furthermore, it is shown that this functor has a left adjoint.

In the last part of the note, Markov chains and stochastic processes are discussed using the category of probabilistic mappings. The first step is done by defining a functor $\Phi : \mathcal{P}^{\mathbb{N}} \to \mathcal{P}^{\mathbb{N}}$. The functor sends an indexed collection of measurable spaces $(\Omega_n)_n$ to

$$\left( \prod_{k<m} \Omega_k \right)_{m \in \mathbb{N}}.$$

A **discrete stochastic process** is now defined as an object $\Omega$ in $\mathcal{P}^{\mathbb{N}}$ together with a morphism in $\mathcal{P}^{\mathbb{N}}$,

$$P : \Phi(\Omega) \to \Omega.$$

This means it is an indexed collection of measurable spaces $(\Omega_n)_n$ together with a collection of Markov kernels

$$\left( P_m : \prod_{k<m} \Omega_k \rightsquigarrow \Omega_m \right).$$

The Markov kernel $P_m$ should be interpreted as follows: given the past and the present $(\omega_1, \ldots, \omega_{m-1})$ we find a probability distribution of what might happen in the future.

A **morphism of discrete stochastic processes** from $(\Omega, P)$ to $(\Omega', P')$ is a morphism $f : \Omega \to \Omega'$ in $\mathcal{P}^{\mathbb{N}}$ such that the following diagram commutes:

$$\begin{tikzcd}
\Phi(\Omega)_n \arrow[r, rightsquigarrow, "\Phi(f)_n"] \arrow[d, rightsquigarrow, "P_n"'] & \Phi(\Omega')_n \arrow[d, rightsquigarrow, "P'_n"] \\
\Omega_n \arrow[r, rightsquigarrow, "f_n"'] & \Omega'_n
\end{tikzcd}$$

These form a category **Stoch** with discrete stochastic processes as objects.

To define Markov chains, Lawvere uses the monoid $N$ of natural numbers interpreted as a category with one object. A **discrete Markov process** is then defined as a functor $N \to \mathcal{P}$. This means that a discrete Markov process is determined by an object $\Xi$ in $\mathcal{P}$ together with a Markov kernel $f : \Xi \rightsquigarrow \Xi$. Moreover, every discrete Markov process is a discrete stochastic process. Indeed, define for every $n \in \mathbb{N}$, $\Omega_n := \Xi$. Then $\Omega := (\Omega_n)_n$ is an object in $\mathcal{P}^{\mathbb{N}}$. Define a morphism $P : \Phi(\Omega) \to \Omega$ in $\mathcal{P}^{\mathbb{N}}$ by

$$P_m(\xi_1, \ldots, \xi_{m-1}) := f(\xi_{m-1}).$$

We therefore see that a discrete Markov process is a discrete stochastic process where the future *only* depends on the present and not the past. The full subcategory of **Stoch** of discrete Markov processes is denoted by **Mark**.

Lawvere finishes the note with an open question: Does the inclusion **Mark** → **Stoch** have an adjoint functor?

### 2.1.2 Swirszcz and Semadeni

In the paper 'Monadic functors and convexity' from 1973, Swirszcz studies a category of compact convex spaces and shows that the inclusion in the category of all compact spaces is monadic [61]. The monad arising from this is a probability monad that is now known as the *Radon monad*.

A key result Swirszcz uses throughout the paper is a strengthened version of a monadicity theorem by Linton (Theorem 3 in [47] and Theorem 1.2 in [9]).

**Theorem 2.1.1** (Linton's monadicity theorem)**.** *Let $\mathcal{D}$ be a category that has kernel pairs of retractions, and let $\mathcal{C}$ be a category that has kernel pairs and coequalizers. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor. Suppose that*

- *$F$ has a left adjoint,*
- *$f$ is a regular epimorphism in $\mathcal{C}$ if and only if $Ff$ is,*
- *$(f, g)$ is a kernel pair in $\mathcal{C}$ if and only if $(Ff, Fg)$ is.*

*Then $F$ is monadic. In the case that $\mathcal{D} =$ **Set**, the conditions are also necessary.*

Swirszcz starts the paper by defining a category of convex spaces. This category has convex subsets of vector spaces as objects and affine maps between them, i.e. maps that preserve convex combinations. This category is denoted by **Conv**. It is shown that the forgetful functor **Conv** → **Set** is *not* monadic. Affine maps $f, g : A \to B$ between convex spaces are constructed such that $(Uf, Ug)$ is a kernel pair in **Set**, but $(f, g)$ is not a kernel pair in **Conv**. By Linton's monadicity theorem it follows that the forgetful functor can not be monadic.

In the second part of the paper, Swirszcz focusses on the category **CompConv** of *compact* convex spaces. The objects of this category are compact convex subsets of locally convex Hausdorff spaces and the morphisms are the affine continuous maps between them. Using Linton's monadicty theorem it is shown that the forgetful functors

$$\mathbf{CompConv} \to \mathbf{Comp} \quad \text{and} \quad \mathbf{CompConv} \to \mathbf{Set}$$

are monadic. Here **Comp** is the category of compact Hausdorff spaces and continuous maps between them.

Also in 1973, Semadeni writes a book with the title ‘Monads and their Eilenberg-Moore algebras in functional analysis’, continuing the work of Swirszcz [56]. In Chapter 7 of this book the monads associated with the monadic forgetful functors in Swirszcz’s paper are described.

A concrete description of the left adjoint of $U : \mathbf{CompConv} \to \mathbf{Comp}$ is given. The left adjoint $S : \mathbf{Comp} \to \mathbf{CompConv}$ is defined on objects by sending a compact Hausdorff space $X$ to the set of Radon probability measures on $X$ together with the topology that makes the evaluation map $\mathrm{ev}_f : S(X) \to \mathbb{R} : \mu \mapsto \int f \mathrm{d}\mu$ continuous for every (bounded) continuous function $f : X \to \mathbb{R}$. This is the topology of *weak convergence of probability measures*. A continuous map $f : X \to Y$ is sent to a continuous map $Sf : S(X) \to S(Y)$ that sends a Radon probability measure $\mu$ to the pushforward of $\mu$ along $f$.

The unit of this adjunction is given by the continuous maps

$$\begin{array}{rccc} \eta_X : & X & \to & US(X) \\ & x & \mapsto & \delta_x . \end{array}$$

The counit is defined by maps $\epsilon_K : SU(K) \to K$ that sends a Radon probability measure on a compact convex space to its centroid (center of gravity). By the Krein-Milman theorem the collection of finitely supported probability measures is dense in $SU(K)$. Because $SU(K)$ and $K$ are compact Hausdorff spaces it is enough to define $\epsilon_K$ on finitely supported probability measures. Using the convex structure of $K$, this can be done by the following assignment

$$\sum_{n=1}^{N} \lambda_n \delta_{x_n} \mapsto \sum_{n=1}^{N} \lambda_n x_n .$$

The induced monad has $T := US : \mathbf{Comp} \to \mathbf{Comp}$ as underlying endofunctor, which sends a compact Hausdorff space to the compact Hausdorff space of Radon probability measures on $X$. The unit of the monad is the same as the unit $\eta$ of the adjunction. The multiplication is given by $\mu := U\eta_S : TT \to T$, which sends a Radon probability measure on $US(X)$ to its centroid in $US(X)$. Concretely this means that for $\mathrm{M} \in TT(X)$ we have that

$$\mu_X(\mathrm{M})(A) = \int \lambda(A)\mathrm{M}(\mathrm{d}\lambda)$$

for every Borel subset $A$ of $X$. This monad is now usually referred to as the *Radon monad* [34].

It follows now from Swirszcz’s result that there is an equivalence $\mathbf{Comp}^T \simeq \mathbf{CompConv}$. Semadeni also gives a description of this equivalence. A compact convex space $K$ can be given an algebra structure by the structure map $\epsilon_K$. The proof that every $T$-algebra is isomorphic to a compact convex subset of a locally convex Hausdorff space such that the algebra’s structure map corresponds to taking the centroid operation is more involved and forms the largest part of Chapter 7 of Semadeni’s book.

The monad induced by the forgetful functor $\mathbf{CompConv} \to \mathbf{Set}$ sends a set $X$ to the set of all Radon probability measures on the Stone-Čech compactification of $X$.

### 2.1.3 Giry

In 1984 Giry wrote the paper ‘A categorical approach to probability theory’ [28]. In this paper Giry describes two probability monads, which are now both known as the *Giry monad*.

One of the monads that are introduced by Giry is a monad on $\mathbf{Mes}$, the category of measurable spaces and measurable maps between them. The underlying endofunctor $\Pi : \mathbf{Mes} \to \mathbf{Mes}$ sends a measurable space $\Omega$ to the set of all probability measures on $\Omega$ together with the smallest

$\sigma$-algebra that makes every evaluation map

$$\begin{array}{rccc} \mathrm{ev}_A : & \Pi(\Omega) & \to & [0,1] \\ & \mathbb{P} & \mapsto & \mathbb{P}(A) \end{array}$$

measurable for every measurable subset $A$ of $\Omega$. The functor sends a measurable map $f : \Omega \to \Omega'$ to the measurable map $\Pi(f) : \Pi(\Omega) \to \Pi(\Omega')$ which sends a probability measure $\mathbb{P}$ to $\mathbb{P} \circ f^{-1}$, the pushforward of $\mathbb{P}$ along $f$.

For a measurable space $\Omega$, there is a measurable map $\eta_\Omega : \Omega \to \Pi(\Omega)$ that sends an element $\omega \in \Omega$ to the Dirac delta $\delta_\omega$. For a measurable map $f : \Omega \to \Omega'$, we have that for every $\omega \in \Omega$,

$$\delta_\omega \circ f^{-1} = \delta_{f(\omega)}.$$

This means that the following diagram commutes:

$$\begin{tikzcd} \Omega \arrow[r, "f"] \arrow[d, "\eta_\Omega"'] & \Omega' \arrow[d, "\eta_{\Omega'}"] \\ \Pi(\Omega) \arrow[r, "\Pi(f)"'] & \Pi(\Omega') \end{tikzcd}$$

It follows that we have a natural transformation $\eta : 1_{\mathbf{Mes}} \to \Pi$.

Furthermore, we also have a measurable map $\mu_\Omega : \Pi^2(\Omega) \to \Pi(\Omega)$ which sends a probability measure $\mathbf{P}$ on $\Pi(\Omega)$ to the probability measure on $\Omega$ defined by

$$A \mapsto \int \mathrm{ev}_A \mathrm{d}\mathbf{P}.$$

The fact that $\mu_\Omega(\mathbf{P})$ is $\sigma$-additive follows from the linearity of integrals and the monotone convergence theorem. For a measurable map $f : \Omega \to \Omega'$ we have that

$$\int \mathrm{ev}_A \mathrm{d}[\mathbf{P} \circ \Pi(f)^{-1}] = \int \mathrm{ev}_A \circ \Pi(f) \mathrm{d}\mathbf{P} = \int \mathrm{ev}_{f^{-1}(A)} \mathrm{d}\mathbf{P}$$

for every $\mathbf{P} \in \Pi^2(\Omega)$ and measurable subset $A$ of $\Omega'$. It follows that the diagram

$$\begin{tikzcd} \Pi^2(\Omega) \arrow[r, "\Pi^2(f)"] \arrow[d, "\mu_\Omega"'] & \Pi^2(\Omega') \arrow[d, "\mu_{\Omega'}"] \\ \Pi(\Omega) \arrow[r, "\Pi(f)"'] & \Pi(\Omega') \end{tikzcd}$$

commutes. Therefore we a natural transformation $\mu : \Pi^2 \to \Pi$. Giry proceeds by proving that this is a monad. This is Theorem 1 in [28].

**Theorem 2.1.2** (Giry)**.** *The triple* $(\Pi, \eta, \mu)$ *forms a monad.*

*Proof.* Let $\Omega$ be measurable space and let $B$ be a measurable subset of $\Pi(\Omega)$. For a probability measure $\mathcal{P} \in \Pi^3(\Omega)$ we have that

$$\int 1_B \mathrm{d}\left[\mu_{\Pi(\Omega)}(\mathcal{P})\right] = \left[\mu_{\Pi(\Omega)}(\mathcal{P})\right](B) = \int \mathrm{ev}_B \mathrm{d}\mathcal{P} = \int \left[\int 1_B \mathrm{d}\mathbf{P}\right] \mathcal{P}(\mathrm{d}\mathbf{P}).$$

By the usual argument, it follows that we that for every measurable map $f : \Pi(\Omega) \to \mathbb{R}$

$$\int f \mathrm{d}\mu_{\Pi(\Omega)}(\mathcal{P}) = \int \left[\int f \mathrm{d}\mathbf{P}\right] \mathcal{P}(\mathrm{d}\mathbf{P}).$$

In particular for $\mathrm{ev}_A : \Pi(\Omega) \to \mathbb{R}$, where $A$ is a measurable subset of $\Omega$, we find that

$$\begin{aligned}\left[\left(\mu \circ \mu_{\Pi(\Omega)}\right)(\mathcal{P})\right](A) &= \int \mathrm{ev}_A \mathrm{d}\left[\mu_{\Pi(\Omega)}(\mathcal{P})\right] \\ &= \int \left[\int \mathrm{ev}_A \mathrm{d}\mathbf{P}\right] \mathcal{P}(\mathrm{d}\mathbf{P}) \\ &= \int \mathrm{ev}_A \circ \mu_\Omega \mathrm{d}\mathcal{P} \\ &= \int \mathrm{ev}_A \mathrm{d}\mathcal{P} \circ \mu_\Omega^{-1} \\ &= \mu_\Omega\left(\mathcal{P} \circ \mu_\Omega^{-1}\right) \\ &= \left[\left(\mu_\Omega \circ \Pi(\mu_\Omega)\right)(\mathcal{P})\right](A)\end{aligned}$$

Since this holds for any $\mathcal{P} \in \Pi^3(\Omega)$ and measurable subset $A$ of $\Omega$, we have the following commutative diagram:

$$\begin{tikzcd}
\Pi^3(\Omega) \arrow[r, "\Pi(\mu_\Omega)"] \arrow[d, "\mu_{\Pi(\Omega)}"'] & \Pi^2(\Omega) \arrow[d, "\mu_\Omega"] \\
\Pi^2(\Omega) \arrow[r, "\mu_\Omega"'] & \Pi(\Omega)
\end{tikzcd}$$

Furthermore, for a probability measure $\mathbb{P} \in \Pi(\Omega)$ and a measurable subset $A$ of $\Omega$, we see that

$$\int \mathrm{ev}_A \mathrm{d}[\mathbb{P} \circ \eta_\Omega^{-1}] = \int \mathrm{ev}_A \circ \eta_\Omega \mathrm{d}\mathbb{P} = \mathbb{P}(A)$$

and

$$\int \mathrm{ev}_A \mathrm{d}\delta_{\mathbb{P}} = \mathrm{ev}_A(\mathbb{P}) = \mathbb{P}(A).$$

We can therefore conclude that the following diagram commutes:

$$\begin{tikzcd}
\Pi(\Omega) \arrow[r, "\Pi(\eta_\Omega)"] \arrow[dr, "1_{\Pi(\Omega)}"'] & \Pi^2(\Omega) \arrow[d, "\mu_\Omega"'] & \Pi(\Omega) \arrow[l, "\eta_{\Pi(\Omega)}"'] \arrow[dl, "1_{\Pi(\Omega)}"] \\
& \Pi(\Omega) &
\end{tikzcd}.$$

This shows that the triple $(\Pi, \eta, \mu)$ forms a monad. □

Giry constructs a similar monad on **Pol**, the category of Polish spaces and continuous maps.

Moreover, Giry notes that the Kleisli category $\mathbf{Mes}_\Pi$ of $(\Pi, \eta, \mu)$ is precisely Lawvere's category of probabilistic mappings from [44]. Indeed, the objects of $\mathbf{Mes}_\Pi$ are measurable spaces and a Kleisli morphism $\Omega \to \Pi(\Omega')$ is precisely a Markov kernel $\Omega \rightsquigarrow \Omega'$. Furthermore, for Kleisli morphisms $f : \Omega \to \Pi(\Omega')$ and $g : \Omega' \to \Pi(\Omega'')$ the Kleisli composite is given by

$$\Omega \xrightarrow{f} \Pi(\Omega') \xrightarrow{\Pi(g)} \Pi^2(\Omega'') \xrightarrow{\mu_{\Omega''}} \Pi(\Omega'').$$

This means that an element $\omega \in \Omega$ is send to a probability measure that sends a measurable subset $A$ of $\Omega''$ to

$$\int \mathrm{ev}_A \mathrm{d}\left[f(\omega) \circ g^{-1}\right] = \int (\mathrm{ev}_A \circ g)(\omega'')[f(\omega)](\mathrm{d}\omega'') = \int [g(\omega'')](A)[f(\omega)](\mathrm{d}\omega'').$$

We can see now that the Kleisli composition of Kleisli morphisms corresponds precisely to the composition of Markov kernels.

In the second part of the paper different properties of the monad are studied. In particular properties related to cofiltered limits ('projective limits') are described. Several technical conditions are given on cofiltered diagrams such that the monad preserves the limits of these diagrams. Giry uses these results to give a categorical proof of the Ionescu-Tulcea theorem.

In the last part of the paper Giry gives a generalization of Lawvere's representation of discrete stochastic processes in [44]. For this Giry introduces *random topological actions* to represent *continuous* stochastic processes using the introduced monad.

## 2.2 Probability monads

It is clear that the monads introduced by Semadeni in [56] and the ones introduced by Giry in [28] follow the same pattern. Monads similar to these are today commonly referred to as *probability monads*. This is an informal term to mean monads that are *similar* to the *Giry monad* or the *Radon monad*.

Over the years many new monads have been introduced that follow Giry's and Semadeni's constructions. These include monads of Radon probability measures on spaces of complete or compact metric spaces [25, 34, 62], but also monads of continuous valuations on ordered spaces [1, 29], monads of Radon probability measures on ordered topological spaces [1, 27, 36] and monads of probability measures on quasi-Borel spaces [32]. Later in the section we will focus on one particular example, the *Kantorovich monad* on the category of complete metric spaces [25].

A more rudimentary example of probability monads is given by the *distribution monad* on **Set**. The underlying endofunctor $\mathcal{D} : \mathbf{Set} \to \mathbf{Set}$ sends a set $A$ to the set

$$\left\{ (p_a)_{a \in A} \in [0,1]^A \mid \sum_{a \in A} p_a = 1, p_a > 0 \text{ for only a finite amount of elements } a \in A \right\}.$$

The unit of the monad $\eta : 1_{\mathbf{Set}} \to \mathcal{D}$ consists of maps $\eta_A : A \to \mathcal{D}A$ for every set $A$, defined by

$$\eta_A(a)_b := \begin{cases} 1 \text{ if } a = b, \\ 0 \text{ otherwise} \end{cases},$$

for every $b \in A$. The monad's multiplication is defined by maps $\mu_A : \mathcal{D}\mathcal{D}A \to \mathcal{D}A$ for every set

$A$. This map sends $(p_\lambda)_{\lambda\in\mathcal{D}(A)} \in \mathcal{D}\mathcal{D}(A)$ to

$$\left(\sum_{\lambda\in\mathcal{D}(A)} \lambda_a p_\lambda\right)_{a\in A}.$$

The reason this monads deserves to be called a *probability monad*, is because $(p_a)_{a\in A}$ can be identified as a finitely supported probability measure on the measurable space $(A, \mathcal{P}(A))$. Under this identification, the unit and multiplication of the distribution monad behave in the same way as those of the Giry monad.

Another example of a probability monad is the *Kantorovich monad* introduced in [25] on the category **CMet** of complete metric spaces and 1-Lipschitz maps between them. A probability measure $\mathbb{P}$ on a metric space $(X, d)$ is called a **Radon probability measure** if

$$\mathbb{P}(A) = \sup\{\mathbb{P}(K) \mid K \subseteq A,\ \text{compact}\},$$

for every open subset $A$ of $X$. A probability measure is said to have finite first moment if

$$\int d(\cdot, x_0)\mathrm{d}\mathbb{P} < \infty$$

for some (and therefore for every) $x_0 \in X$. Let $\mathcal{K}(A)$ be the set of all Radon probability measures that have finite first moment. For elements $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathcal{K}(A)$, define

$$d_K(\mathbb{P}_1, \mathbb{P}_2) := \sup\left\{\left|\int f\mathrm{d}\mathbb{P}_1 - \int f\mathrm{d}\mathbb{P}_2\right| \mid f : X \to \mathbb{R}\ \text{1-Lipschitz}\right\}.$$

Because $\mathbb{P}_1$ and $\mathbb{P}_2$ have finite first moment, $d_K(\mathbb{P}_1, \mathbb{P}_2) \in [0, \infty)$. It is now not difficult to see that $d_K$ defines a metric on $\mathcal{K}(X)$, this is the **Kantorovich metric**. The obtained metric space $(\mathcal{K}(X), d_K)$ is called the **Kantorovich space**. In the case that $(X, d)$ is a complete metric space, then so is the corresponding Kantorovich space.

**Proposition 2.2.1.** *The pushforward measure of a Radon probability measure of finite first moment on a metric space $X$ along a 1-Lipschitz map $f : X \to Y$ is a Radon probability measure that has finite first moment on the metric space $Y$.*

*Proof.* Let $A$ be a measurable subset of $Y$ and let $\epsilon > 0$. There is a compact subset $K \subseteq f^{-1}(A)$ such that

$$\begin{aligned}\mathbb{P}(f^{-1}(A)) &\leq \mathbb{P}(K) + \epsilon\\ &\leq \mathbb{P}(f^{-1}f(K)) + \epsilon\\ &\leq \sup\{[\mathbb{P}\circ f^{-1}](C) \mid C \subseteq A\ \text{compact}\} + \epsilon.\end{aligned}$$

Taking $\epsilon \to 0$ gives us that

$$[\mathbb{P}\circ f^{-1}](A) \leq \sup\{[\mathbb{P}\circ f^{-1}](C) \mid C \subseteq A\ \text{compact}\}.$$

Clearly we also always have the other inequality, proving that $\mathbb{P}\circ f^{-1}$ is a Radon probability measure.

Moreover, we know that there exists an $x_0 \in X$ such that

$$\infty > \int d_X(\cdot, x_0)\mathrm{d}\mathbb{P} \geq \int d_Y(f(\cdot), f(x_0)\mathrm{d}\mathbb{P} = \int d(\cdot, f(x_0))\mathrm{d}[\mathbb{P} \circ f^{-1}].$$

It follows that $\mathbb{P} \circ f^{-1}$ has finite first moment. □

By the previous proposition, we know that for every 1-Lipschitz map $f : X \to Y$ there is a map $\mathcal{K}(f) : \mathcal{K}(X) \to \mathcal{K}(Y)$. This map sends a Radon probability measure $\mathbb{P}$ of finite first moment to its pushforward measure along $f$. Moreover, for elements $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathcal{K}(X)$ we see that

$$\begin{aligned} d_K(\mathbb{P}_1 \circ f^{-1}, \mathbb{P}_2 \circ f^{-1}) &= \sup\left\{\left|\int g\mathrm{d}[\mathbb{P}_1 \circ f^{-1}] - \int g\mathrm{d}[\mathbb{P}_2 \circ f^{-1}]\right| \mid g : Y \to \mathbb{R} \text{ 1-Lipschitz}\right\} \\ &= \sup\left\{\left|\int g \circ f\mathrm{d}\mathbb{P}_1 - \int g \circ f\mathrm{d}\mathbb{P}_2\right| \mid g : Y \to \mathbb{R} \text{ 1-Lipschitz}\right\} \\ &\leq \sup\left\{\left|\int g\mathrm{d}\mathbb{P}_1 - \int g\mathrm{d}\mathbb{P}_2\right| \mid g : X \to \mathbb{R} \text{ 1-Lipschitz}\right\} \end{aligned}$$

This shows that the map $\mathcal{K}(f)$ is 1-Lipschitz. The above induces an endofunctor $\mathcal{K} : \mathbf{CMet} \to \mathbf{CMet}$.

For a complete metric space $X$, the assignment $x \mapsto \delta_x$ defines a map $\eta_X : X \to \mathcal{K}(X)$. This map is also 1-Lipschitz, indeed for $x_1$ and $x_2$ in $X$, we find that

$$d_K(\delta_{x_1}, \delta_{x_2}) = \sup\{|f(x_1) - f(x_2)| \mid f : X \to \mathbb{R} \text{ 1-Lipschitz}\} \leq d_X(x_1, x_2).$$

There is also a map $\mu_X : \mathcal{K}\mathcal{K}(X) \to \mathcal{K}(X)$ sending $\mathbf{P}$ to the probability measure on $X$ defined by $\mu_X(\mathbf{P})(A) := \int \mathrm{ev}_A \mathrm{d}\mathbf{P}$.

**Proposition 2.2.2.** *Let $\mathbf{P} \in \mathcal{K}(X)$, then $\mu_X(\mathbf{P})$ is a Radon measure.*

*Proof.* Let $\epsilon > 0$. There exists a compact subset $\mathbf{K} \subseteq \mathcal{K}(X)$ such that $\mathbf{P}(\mathbf{K}^{\mathrm{C}}) < \epsilon$.

Let $A$ be an open subset of $X$. There exists an increasing sequence of continuous maps $(f_n : X \to [0,1])_n$ that converge pointwise to the indicator function $1_A$. It follows that the maps

$$\left(\mathrm{ev}_{f_n} : \mathcal{K}(X) \to [0,1] : \mathbb{P} \mapsto \int f_n \mathrm{d}\mathbb{P}\right)_n$$

form an increasing sequence of maps that converge to $\mathrm{ev}_A$. By the Portmanteau theorem, we know that $\mathrm{ev}_{f_n}$ is continuous for every $n$, and therefore it follows by Dini's theorem that $\mathrm{ev}_A$ is uniformly continuous on the compact subset $\mathbf{K}$. This means that there exists a $\delta > 0$ such that for every $\mathbb{P}_1, \mathbb{P}_2 \in \mathbf{K}$ such that $d_K(\mathbb{P}_1, \mathbb{P}_2) < \epsilon$,

$$|\mathbb{P}_1(A) - \mathbb{P}_2(A)| < \epsilon.$$

Since $\mathbf{K}$ is compact, there exist $\mathbb{P}_1, \dots, \mathbb{P}_n$ in $\mathbf{K}$ such that

$$\mathbf{K} \subseteq \bigcup_{k=1}^{n} B(\mathbb{P}_n, \delta).$$

Moreover, since $\mathbb{P}_k$ is Radon for every $1 \leq k \leq n$, there exists a compact subset $K \subseteq A$ such that

$$\mathbb{P}_k(A) - \mathbb{P}_k(K) < \epsilon$$

for every $1 \leq k \leq n$. This means that

$$\mathbb{P}(A) - \mathbb{P}(K) < 2\epsilon$$

for every $\mathbb{P} \in \mathbf{K}$. We conclude that

$$\begin{aligned} \mu_X(\mathbf{P})(A) &\leq \int_{\mathbf{K}} \mathbb{P}(A)\mathbf{P}(\mathrm{d}\mathbb{P}) + \epsilon \\ &\leq \int_{\mathbf{K}} \mathbb{P}(K)\mathbf{P}(\mathrm{d}\mathbb{P}) + 3\epsilon \\ &\leq \mu_X(\mathbf{P})(K) + 3\epsilon \\ &\leq \sup\{\mu_X(\mathbf{P})(K) \mid K \subseteq A\} + 3\epsilon. \end{aligned}$$

Taking $\epsilon \to 0$ proves the claim.

□

The map $\mu_X$ is 1-Lipschitz, as can be seen from the following inequalities for every $\mathbf{P}_1$ and $\mathbf{P}_2$ in $\mathcal{K}\mathcal{K}(X)$.

$$\begin{aligned} d_K(\mu_X(\mathbf{P}_1), \mu_X(\mathbf{P}_2)) &= \sup\left\{\left|\int f \mathrm{d}\mu_X(\mathbf{P}_1) - \int f \mathrm{d}\mu_X(\mathbf{P}_2)\right| \mid f : X \to \mathbb{R} \text{ 1-Lipschitz}\right\} \\ &= \sup\left\{\left|\int \left[\int f \mathrm{d}\mathbb{P}\right] \mathbf{P}_1(\mathrm{d}\mathbb{P}) - \int \left[\int f \mathrm{d}\mathbb{P}\right] \mathbf{P}_2(\mathrm{d}\mathbb{P})\right| \mid f : X \to \mathbb{R} \text{ 1-Lipschitz}\right\} \\ &\leq \sup\left\{\left|\int g \mathrm{d}\mathbf{P}_1 - \int g \mathrm{d}\mathbf{P}_2\right| \mid g : \mathcal{K}(X) \to \mathbb{R} \text{ 1-Lipschitz}\right\} \end{aligned}$$

Similar to Giry's proof of Theorem 2.1.2 it can be shown that the triple $(\mathcal{K}, \eta, \mu)$ is a monad. This monad is called the **Kantorovich monad**. It is clear that this monad belongs to the class of probability monads.

Another monad that could be considered as a probability monad is the *ultrafilter monad* on **Set** [45]. The monad's endofunctor sends a set $A$ to the set of all ultrafilters on $A$. An ultrafilter $\mathcal{U}$ on $A$ has the property that for every subset $B$ of $A$, either $B$ or $B^{\mathrm{C}}$ belongs to $\mathcal{U}$. We can think of the elements of $\mathcal{B}$ as the subsets of $A$ with *probability* 1 and all the other subsets as those with probability 0. In this way we can interpret an ultrafilter as a finitely additive probability measure taking values in $\{0, 1\}$.

Interestingly, the ultrafilter monad can be constructed as the codensity monad of the inclusion functor $\mathbf{Set}_f \to \mathbf{Set}$, as shown in [38]. In Chapter 3 we will prove that many other probability monads arise in a similar way.

### 2.2.1 Strength

A property that many probability monads seem to have in common is that they are *commutative monads*. Intuitively, a commutative monad is a monad on symmetric monoidal category that interacts well with its monoidal structure.

Let $(\mathcal{C}, \otimes, I)$ be a symmetric monoidal category and let $(T, \eta, \mu)$ be a monad on $\mathcal{C}$. There is a functor $\otimes(1_{\mathcal{C}}, T) : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ sending $(A, B)$ to $A \otimes TB$ and there is a functor $T \circ \otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ that sends a pair of objects $(A, B)$ to $T(A \otimes B)$.

A **strength** is a natural transformation

$$\tau : \otimes(1_{\mathcal{C}}, T) \to T \circ \otimes$$

that behaves appropriately with respect to the unitor, associator and braiding of the symmetric monoidal category and also commutes appropriately with respect to the unit and multiplication of the monad. A monad on a symmetric monoidal category together with a strength forms a **strong monad**.

In the case that $\mathcal{C}$ is a symmetric closed monoidal category, for objects $X$ and $Y$ the strength morphisms $\tau_{[X,Y],Y} : [X, Y] \otimes TY \to T([X, Y] \otimes Y)$ induces a morphism

$$[X, Y] \otimes TY \xrightarrow{\tau_{[X,Y],Y}} T([X, Y] \otimes Y) \xrightarrow{T\mathrm{ev}} TX,$$

which correspond to a morphism

$$[X, Y] \to [TX, TY].$$

Using the strength axioms it can be shown that this provides an enrichment of the monad $T$, where $\mathcal{C}$ is viewed as enriched over itself.

The category **Mble** is a Cartesian monoidal category. A strength for the Giry monad can be given by the maps

$$X \times \mathcal{G}Y \to \mathcal{G}(X \times Y) : (x, \mathbb{P}) \mapsto \delta_x \otimes \mathbb{P},$$

for all measurable spaces $X$ and $Y$. The category **Mble** is *not* Cartesian closed and therefore the strength of the Giry monad does not make it an enriched monad. To avoid this, one can consider the Cartesian closed category of quasi-Borel spaces, introduced in [32]. The category **Mble** can be given a canonical symmetric closed monoidal structure, however it is shown in [55] that the Giry monad is not strong with respect to this monoidal structure.

The category of complete metric spaces is a symmetric closed monoidal category for the tensor product $(X \times Y, d_{X \otimes Y})$ of complete metric spaces $(X, d_X)$ and $(Y, d_Y)$ defined by

$$d_{X \otimes Y}((x_1, y_1), (x_2, y_2)) := d_X(x_1, x_2) + d_Y(y_1, y_2)$$

for all $x_1, x_2 \in X$ and $y_1, y_2 \in Y$.

A strength for the Kantorovich monad can be defined in a similar way as for the Giry monad, namely by the assignment $(x, \mathbb{P}) \mapsto \delta_x \otimes \mathbb{P}$. This is Corollary 5.8 in [24]. Since **CMet** *is* closed monoidal, we can interpret the Kantorovich monad as a **CMet**-enriched monad.

### 2.2.2 Eilenberg-Moore algebras

An interesting category related to a monad is its category of (Eilenberg-Moore) algebras. The work of Swirszcz and Semadeni shows that the category of algebras of the Radon monad is equivalent to the category **CompConv** of compact convex spaces. This is the category of compact convex subsets of locally convex Hausdorff vector spaces with affine continuous maps between them. Moreover, Semadeni gave a concrete description of this equivalence. Indeed, Semadeni showed that every algebra for the Radon monad is isomorphic to a compact convex subset in such a way that the algebra's structure map —under the identification— corresponds to a barycenter operation.

**Example 2.2.3.** Consider the compact Hausdorff space $[0,1]$ and let $(\mathcal{R}, \eta, \mu)$ be the Radon monad as described by Semadeni in [56]. We can define a map $\alpha : \mathcal{R}([0,1]) \to [0,1]$ by the assignment

$$\mathbb{P} \mapsto \int \mathrm{Id}_{[0,1]} \mathrm{d}\mathbb{P},$$

where $\mathrm{Id}_{[0,1]}$ is the identity map on $[0,1]$. By definition of the topology on $\mathcal{R}([0,1])$, this map is continuous. For every $x \in [0,1]$,

$$\alpha(\delta_x) = \int \mathrm{Id}_{[0,1]} \mathrm{d}\delta_x = x.$$

Furthermore, for $\mathbf{P} \in \mathcal{R}\mathcal{R}([0,1])$ we see that

$$\begin{aligned}
\alpha(\mu(\mathbf{P})) &= \int \mathrm{Id}_{[0,1]} \mathrm{d}\mu(\mathbf{P}) \\
&= \int \left[ \int \mathrm{Id}_{[0,1]} \mathrm{d}\lambda \right] \mathbf{P}(\mathrm{d}\lambda) \\
&= \int \alpha(\lambda) \mathbf{P}(\mathrm{d}\lambda) \\
&= \int \mathrm{Id}_{[0,1]} \mathrm{d}[\mathbf{P} \circ \alpha^{-1}] \\
&= \alpha(\mathbf{P} \circ \alpha^{-1})
\end{aligned}$$

From this it follows that $([0,1], \alpha)$ is an algebra for the Radon monad and indeed it is also a compact convex subset of the locally convex Hausdorff vector space $\mathbb{R}$.

Today, the algebras of several other probability monads have been characterized and it should not come as a surprise that all these descriptions involve some type of convexity structure.

Abstract convex spaces were first studied by Stone in [59] and later by several others such as Fritz, Gudder and Swirszcz [19, 30, 60]. Fritz defines a convex space as follows.

**Definition 2.2.4** (Definition 3.1 in [19])**.** A **convex space** is a set $X$ together with an indexed collection of maps $(\alpha_\lambda : X \times X \to X \mid \lambda \in [0,1])$ such that

- $\alpha_0(x,y) = x$ for all $x, y \in X$,
- $\alpha_\lambda(x,x) = x$ for all $x \in X$ and $\lambda \in [0,1]$,
- $\alpha_\lambda(x,y) = \alpha_{1-\lambda}(y,x)$ for all $x, y \in X$ and $\lambda \in [0,1]$,
- for all $x, y, z \in X$ and $\lambda, \mu \in [0,1)$,

$$\alpha_\lambda(\alpha_\mu(x,y), z) = \alpha_{\lambda\mu}\left(x, \alpha_{\frac{\lambda(1-\mu)}{1-\lambda\mu}}(y,z)\right).$$

An **affine map** between convex spaces $(X, (\alpha_\lambda)_\lambda)$ and $(Y, (\beta_\lambda)_\lambda)$ is a map $f : X \to Y$ such that $\beta_\lambda(f(x), f(y)) = f(\alpha_\lambda(x,y))$ for all $x, y \in X$ and $\lambda \in [0,1]$.

The category of convex spaces and affine maps between them is denoted as **Conv**. The following result tells us that this category arises as the category of algebras of the distribution monad. This is Theorem 4 in [33].

**Theorem 2.2.5.** *The category* **Conv** *of convex spaces is isomorphic to the category of algebras of the distribution monad.*

The set $\mathbb{R}$ has a canonical convex space structure. Indeed, define for $\lambda \in [0,1]$ a map $\alpha_\lambda : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by the assignment

$$(x, y) \mapsto \lambda x + (1 - \lambda) y.$$

However, there also exist less intuitive examples of convex spaces. Take for example the set $X := \{0, \infty\}$ and define for $\lambda \in (0,1)$ a map $\alpha_\lambda : X \times X \to X$ by

$$(x, y) \mapsto \begin{cases} 0 \text{ if } x = y = 0 \\ \infty \text{ otherwise} \end{cases}$$

and maps $\alpha_1$ and $\alpha_0$ by $\alpha_1(x, y) := x$ and $\alpha_0(x, y) := y$ for all $x, y \in X$. Then this also forms a convex space.

One of the main differences between these two examples is that $\mathbb{R}$ satisfies the *cancellation property*, while $\{0, \infty\}$ does not. A convex space $(X, (\alpha_\lambda)\lambda)$ is said to satisfy **the cancellation property** if for $x, y \in X$ such that there exists a $z \in X$ and a $\lambda \in (0, 1)$ such that $\alpha_\lambda(x, z) = \alpha_\lambda(y, z)$, then $x = y$.

A convex space that satisfies the cancellation property is how Stone defines an **abstract convex set** [59]. Moreover, Stone gives a characterization of these abstract convex sets as the convex subsets of vector spaces. The following theorem is Theorem 2 in [59] and Theorem 4 in [10].

**Theorem 2.2.6** (Stone)**.** *An abstract convex set is isomorphic to a convex subset of a vector space.*

In [10] and [25], a characterization of the category of algebras of the Kantorovich monad is given. It is shown that this category is isomorphic to the category of closed convex subsets of Banach spaces and the affine 1-Lipschitz maps between them. This is Theorem 5.3.1 in [25] and Theorem 9 in [10]. A quite technical characterization of the algebras of the Giry monad on Polish spaces is given by Doberkat in [12]. However, a description of the algebras of the Giry monad on measurable spaces remains unknown.

It is clear from the above that algebras of probability monads have something to do with integration. We could loosely interpret this as follows: '*algebras of probability monads are precisely the spaces where integration and expectation makes sense*'. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an algebra $(A, \alpha)$ for the Giry monad on measurable spaces. Let $X : (\Omega, \mathcal{F}) \to A$ be a measurable map. Then we define the **expectation of** $X$ as

$$\alpha(\mathbb{P} \circ X^{-1}).$$

For the particular case that $A = [0, 1]$ and $\alpha : \mathcal{G}([0, 1]) \to [0, 1]$ is given by integrating the identity map $\mathrm{Id}_{[0,1]}$, this gives

$$\alpha(\mathbb{P} \circ X^{-1}) = \int \mathrm{Id}_{[0,1]} \mathrm{d}[\mathbb{P} \circ X^{-1}] = \int X \mathrm{d}\mathbb{P}.$$

It follows that the generalized definition of random variables taking values in an algebra $(A, \alpha)$ reduces to the usual definition of expectation of real-valued random variables.

A next natural step is to ask if we can also interpret *conditional* expectation in the framework of algebras of probability monads. This question is answered by Fritz and Perrone in [26, 51] using *partial evaluations*.

**Definition 2.2.7** (Definition 2.1 in [26])**.** Let $(T, \eta, \mu)$ be a monad on a category $\mathcal{C}$ and let $(A, \alpha)$ be an algebra for this monad and let $S$ be an object of $\mathcal{C}$. Consider maps $p, q : S \to T(A)$. A **partial evaluation of** $p$ **into** $q$ is a map $k : S \to TT(A)$ such that the following diagram commutes

$$\begin{tikzcd}
 & S \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\
TA & TTA \arrow[l, "\mu_A"] \arrow[r, "T\alpha"'] & TA
\end{tikzcd}$$

The following result says that partial evaluation relation for the Kantorovich monad is precisely the *conditional expectation in distribution* relation for Radon probability measures on an agebra of the Kantorovich monad. This is Theorem 4.2.14 in [51].

**Theorem 2.2.8** (Perrone)**.** *Let $(A, \alpha)$ be an algebra for the Kantorovich monad and let $p, q \in \mathcal{K}A$ (i.e. maps $p, q : 1 \to \mathcal{K}A$). The following are equivalent.*

- *There exists a partial evaluation of $p$ into $q$.*
- *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a $\sigma$-subalgebra $\mathcal{G}$ of $\mathcal{F}$ together with an $\mathcal{F}$-measurable map $f : \Omega \to A$ and a $\mathcal{G}$-measurable map $g : \Omega \to A$ such that*

  $$\mathbb{P} \circ f^{-1} = p \quad \text{and} \quad \mathbb{P} \circ g^{-1} = q$$

  *and*

  $$\alpha\left(\mathbb{P}(- \mid G) \circ f^{-1}\right) = \alpha\left(\mathbb{P}(- \mid G) \circ g^{-1}\right),$$

  *for all $G \in \mathcal{G}$ with $\mathbb{P}(G) > 0$.*

## 2.3 Markov categories

Another important category related to monads is the Kleisli category, the category of *free algebras*. In the work of Giry and Lawvere, we already saw that the Kleisli category of the Giry monad has a nice probabilistic interpretation. Indeed, Kleisli maps and Kleisli composition precisely correspond to Markov kernels and their composition in probability theory. Similarly, we can give a probabilistic interpretation to the maps and composition in Kleisli categories of other probability monads.

We already saw that many probability monads are commutative. The strength morphisms $(\sigma_{A,B} : A \otimes TB \to T(A \otimes B)$ of a commutative monad $(T, \mu, \eta)$ induce morphisms

$$TA \otimes TB \to T(A \otimes TB) \to TT(A \otimes B) \xrightarrow{\mu_{A \otimes B}} T(A \otimes B)$$

for all objects $A$ and $B$ in the symmetric monoidal category $\mathcal{C}$. Using this, a tensor product can be defined on the Kleisli category $\mathcal{C}_T$. Indeed, a functor $\otimes : \mathcal{C}_T \times \mathcal{C}_T \to \mathcal{C}_T$ can be defined by sending a pair objects $(A, B)$ to $A \otimes B$ and by sending a pair of morphisms $f : A_1 \to TA_2$ and $g : B_1 \to TB_2$ to the morphism

$$A_1 \otimes B_1 \xrightarrow{f \otimes g} TA_2 \otimes TB_2 \to T(A_2 \otimes B_2).$$

Together with the object $I$, this induces a symmetric monoidal structure on $\mathcal{C}_T$. We see that most probability monads are defined on a **semicartesian monoidal category**, i.e. a monoidal category whose unit is the terminal object. Furthermore, because all our examples of probability monads are **affine**, i.e. the terminal object is preserved, the induced monoidal structure on the Kleisli category $\mathcal{C}_T$ also becomes semicartesian.

The semicartesian monoidal categories of interest have also another interesting property: every object has a canonical comonoid structure.

**Example 2.3.1.** Consider the Cartesian monoidal category **Mble**. Then for every measurable space, there are maps

$$\begin{array}{ccccccc} X & \to & X \times X & \qquad & X & \to & 1 \\ x & \mapsto & (x,x) & \qquad & x & \mapsto & * \end{array}$$

These structure maps make $X$ a comonoid in **Mble**.

Consider an affine monad $(T, \mu, \eta)$ on a monoidal category $\mathcal{C}$, where every object comes with a comonoid structure. Then also in the induced monoidal Kleisli category $\mathcal{C}_T$, every object has a comonoid structure. These comonoids are induced by the ones in the category $\mathcal{C}$. Indeed, for an object $X$, we have an induced comultiplication $X \to X \otimes X \xrightarrow{\eta_{X \otimes X}} T(X \otimes X)$ and an induced counit $X \to 1 \xrightarrow{\eta_1} T1$.

Fritz observed in [20] that the above described structures on a Kleisli category of a probability monad suffice to describe several probabilistic concepts in these Kleisli categories. Fritz uses these as axioms for an *abstraction* of such a Kleisli category, a *Markov category.*

**Definition 2.3.2.** A **Markov category** is a symmetric semicartesian monoidal category $(\mathcal{C}, \otimes, I)$ in which every object $X$ is equipped with a commutative comonoid structure

$$\mathrm{copy}_X : X \to X \otimes X \quad \text{and} \quad \mathrm{del}_X : X \to I,$$

such that the following diagrams commute for every morphism $f : X \to Y$.

$$\begin{tikzcd}
 & (X \otimes Y) \otimes (X \otimes Y) \arrow[dd, dash, "\cong"] \\
X \otimes Y \arrow[ur, "\mathrm{copy}_{X \otimes Y}"] \arrow[dr, "\mathrm{copy}_X \otimes \mathrm{copy}_y"'] & \\
 & (X \otimes X) \otimes (Y \otimes Y)
\end{tikzcd}$$

The definition of a Markov category does *not* require that the copy maps $(\mathrm{copy}_X : X \to X \otimes X)_{X \in \mathcal{C}}$ form a natural transformation. Therefore, there might be morphisms $f : X \to Y$ in the Markov category such that the following diagram does *not* commute.

$$\begin{tikzcd}
X \arrow[r, "\mathrm{copy}_X"] \arrow[d, "f"'] & X \otimes X \arrow[d, "f \otimes f"] \\
Y \arrow[r, "\mathrm{copy}_Y"'] & Y \otimes Y
\end{tikzcd}$$

A morphism in a Markov category for which the above diagram *does* commute is called a **deterministic morphism**.

**Example 2.3.3.** The Kleisli category of the Giry monad forms a Markov category, as described above. A deterministic morphism $f : X \to \mathcal{G}Y$ in $\mathbf{Mble}_{\mathcal{G}}$ is a morphism such that

$$[f(x)](A)[f(x)](B) = [f(x)](A \cap B)$$

for every $x \in X$ and measurable subsets $A$ and $B$ in $Y$. We can now conclude that $f(x)$ is a probability measure that can only takes values 0 or 1, or in other words a *deterministic* probability measure.

The composite of two deterministic morphisms is again deterministic. Objects of $\mathcal{C}$ together with the deterministic morphisms form a subcategory of $\mathcal{C}$, which is denoted by $\mathcal{C}_{\text{det}}$. In the case that every morphism in a Markov category is deterministic, i.e. $\mathcal{C}_{\text{det}} = \mathcal{C}$, it follows by Fox's theorem [17] that the Markov category is *Cartesian.*

We have already discussed that probability monads give rise to Markov categories. However, not every Markov category arises as the Kleisli category of a probability monad. The following example is Example 2.5 in [20].

**Example 2.3.4.** The category **FinStoch** has finite sets as objects. A morphism from $A$ to $B$ is a matrix $(p_{ab})_{a,b}$ such that

$$\sum_{b \in B} p_{ab} = 1$$

for every $a \in A$. The composition is defined by matrix multiplication. A monoidal structure is defined by taking the Cartesian product of finite sets and the *Kronecker product* of matrices. For every finite set $A$, there is a matrix defined by $p_{a_1 a_2 a_3} = 1$ if and only if $a_1 = a_2 = a_3$. This defines a morphism

$$\text{copy}_A : A \to A \times A.$$

A morphism $\text{del}_A : A \to 1$ is defined by $p_{a*} = 1$ for every $a \in A$.

This defines a Markov category that does not arise as the Kleisli category of a (probability) monad on $\mathbf{Set}_f$. This is shown in [22]. The argument goes as follows. If there would be such a monad $\mathbf{T}$, then there would be a bijection

$$\mathbf{FinStoch}(A, B) \cong \mathbf{Set}_f(A, B),$$

for all finite sets $A$ and $B$. However, for non-empty finite sets $A$ and $B$, $\mathbf{FinStoch}(A, B)$ is infinite, while $\mathbf{Set}_f(A, B)$ is finite.

Example 2.3.4 gives rise to the following question: '*When is a Markov category the Kleisli category of a monad?*'. This problem has been discussed in [22]. In this paper conditions are given for when a Markov category $\mathcal{C}$ is isomorphic to the Kleisli category of a monad on $\mathcal{C}_{\text{det}}$.

Another important property that Markov categories can have is the existence of *conditionals.* In Section 1.1.3, we saw that a Markov kernel $f : X \rightsquigarrow Y$ between measurable spaces together with a probability measure $\mathbb{P}$ on $X$ always induce a joint probability distribution on $X \times Y$, defined by

$$A \times B \mapsto \int_A [f(x)](B)\mathbb{P}(\mathrm{d}x).$$

However, it is not always the case that every joint probability distribution on $X \times Y$ is induced by a Markov kernel. We can rewrite this problem in the following way. Given a joint probability distribution on $X \times Y$, i.e. a Markov kernel $I \rightsquigarrow X \times Y$, can we find a Markov kernel $f : X \rightsquigarrow Y$ such that following diagram commutes?

$$\begin{tikzcd}
I \arrow[r, squiggly] \arrow[drr, squiggly] & X \times Y \arrow[r, squiggly, "1_X \times \text{del}_X"] & X \arrow[r, squiggly, "\text{copy}_X"] & X \times X \arrow[dl, squiggly, "1_X \times f"] \\
 & & X \times Y &
\end{tikzcd}$$

This would mean that

$$\mathbb{P}(A \times B) = \int [\delta_x \otimes f(x)](A \times B)[\mathbb{P} \circ \pi_1^{-1}](\mathrm{d}x) = \int_A [f(x)](B)[\mathbb{P} \circ \pi_1^{-1}](\mathrm{d}x).$$

This motivates the following definition.

**Definition 2.3.5.** A Markov categories **has conditionals** if for every morphism $I \to X \otimes Y$ there exists a morphism $f : X \to Y$ such that the following diagram commutes.

$$\begin{tikzcd}
I \arrow[r] \arrow[drr] & X \otimes Y \arrow[r, "1_X \otimes \mathrm{del}_X"] & X \arrow[r, "\mathrm{copy}_X"] & X \otimes X \arrow[dl, "1_X \otimes f"] \\
& & X \otimes Y &
\end{tikzcd}$$

**Remark 2.3.6.** Definition 2.3.5 is not the most general way to describe conditionals in Markov categories. Some results require the stronger notion of *conditionals with parameters* which is introduced in [20].

**Example 2.3.7.** By the disintegration theorem (Theorem 1.1.17) it follows that the Markov category of standard Borel spaces and Markov kernels between them *has conditionals.*

The axiom of existence of conditionals is an important condition for several important results in the theory of Markov categories. The applications of these results in the particular example of Kleisli categories of probability monads has led to categorical proof of several probabilistic results. In [21] a categorical proof for the de Finetti theorem is given using Markov categories that have conditionals. Other examples of applications of Markov categories are the d-separation criterion and categorical proofs for 0-1 laws [53, 23]. Moreover, Markov categories have also been used to study other areas in mathematics, such as ergodic theory [50].

Another remarkable advantage obtained from using Markov categories is that morphisms can be represented as string diagrams. This allows for a very intuitive visual way to think about the morphisms and properties in an abstract Markov category.

# Chapter 3

# Probability monads as codensity monads

## 3.1 Introduction

A probability distribution on a countable set $A$ is defined as a function $p : A \to [0,1]$ such that

$$\sum_{a \in A} p(a) = 1.$$

For uncountable sets however, this definition is not good any more. Indeed, let $p : A \to [0,1]$ be a function on an uncountable set $A$ such that $\sum_{a \in A} p(a) = 1$; then the support of $p$ is countable. But this means that $p$ is essentially a probability distribution on a countable set. Therefore, to express probability distributions on sets such as $\mathbb{R}$ or $\mathcal{C}([0,\infty), \mathbb{R})$ we need a different definition. Using measure theory a definition of a *probability measure* can be given by Kolmogorov's axioms.

An endofunctor $\mathcal{G}$ on the category of measurable spaces **Mble** can be defined by sending a measurable space $X$ to the measurable space of all probability measures on $X$. Using Dirac delta probability measures and integration, this endofunctor can be given a monad structure. This monad is called the *Giry monad* [28]. Monads on categories of measurable and topological spaces that are similar to the Giry monad are referred to as *probability monads* [34].

The right Kan extension of a functor along itself, assuming that it exists, can be given a natural monad structure. The obtained monad is called the *codensity monad* of that functor [45]. We will show that certain probability monads can be constructed as the codensity monads of functors that send a countable set $A$ to the space of all probability measures on $A$. This shows that probability measures arise naturally as the categorical extension of the more intuitive probability measures on countable sets. We will discuss this construction for several monads of different kinds of probability measures on different kinds of measurable and topological spaces.

We begin this chapter by recalling the definition of a codensity monad (Section 3.2). The rest of the chapter can be divided in two parts. In the first part we will discuss probability monads on categories of measurable spaces. For completeness we start in Section 3.3 with a detailed overview of probability measures and integration, which will lead to an integral representation theorem for probability measures. After this we will construct the Giry monad and variations of this monad as codensity monads in Section 3.4. Here the integral representation theorem from before will play a key role. The second part is similar to the first part, only here we will talk about probability monads on categories of topological spaces. Again we start with an overview of probability measures and integration on topological spaces in Section 3.5 and give a slightly more general version of the Daniell–Stone representation theorem in Section 3.5.2. This result will allow us to construct the Radon monad, bounded Lipschitz monad and Baire monad as codensity monads in Section 3.6. In this section we will also briefly discuss another probability monad and the problems that arose in the attempt to construct this monad as a codensity monad.

An overview of the results in this paper is given in the following table.

| Category | Probability monad | Theorem |
|---|---|---|
| measurable spaces | Giry monad | 3.4.1 |
| measurable spaces | Giry monad of fin. add. probability measures | 3.4.4 |
| premeasurable spaces | Giry monad of probability premeasures | 3.4.6 |
| compact Hausdorff spaces | Radon monad | 3.6.4 |
| compact metric spaces | Bounded Lipschitz monad | 3.6.10 |
| Hausdorff spaces | Baire monad | 3.6.13 |

**Related work**: The Giry monad was introduced by Giry in [28]. This monad and the finitely additive variation of it already have been constructed as codensity monads of inclusion functors of certain convex spaces in the category of measurable spaces by Avery in [3]. The constructions we will present are different but related to Avery's constructions. Avery uses powers of the unit interval and affine maps, while we will use finite and countable simplices. The Radon monad originates from work by Semadeni [56] and Swirszcz [61] and is further discussed in [36] and [34]. The bounded Lipschitz monad is similar to the Kantorovich monad on the category of compact metric spaces, which was introduced by van Breugel in [62]. This monad was extended to a monad on complete metric spaces by Perrone and Fritz [25, 27].

## 3.2 Codensity monads

We start this section by recalling the definition of a codensity monad.

Every functor $G : \mathcal{D} \to \mathcal{C}$ that has a left adjoint $F : \mathcal{C} \to \mathcal{D}$ induces a monad on $\mathcal{C}$, namely the endofunctor $GF : \mathcal{C} \to \mathcal{C}$ together with the unit of the adjunction as unit of the monad and $G\epsilon_F$ as the multiplication of the monad. Suppose now that we have a functor $G : \mathcal{D} \to \mathcal{C}$ such that the right Kan extension of $G$ along $G$ exists; then $G$ still induces a monad on $\mathcal{C}$. Let $(T^G : \mathcal{C} \to \mathcal{C}, \gamma : T^G \circ G \to G)$ be the right Kan extension of $G$ along $G$.

There is a natural transformation $1_G : 1_{\mathcal{C}} \circ G \to G$ and therefore by the universal property of right Kan extensions, there exists a unique $\eta : 1_{\mathcal{C}} \to G$ such that the following diagrams are equal:

$$\begin{tikzcd}
\mathcal{D} \arrow[rr, "G"] \arrow[dr, "G"'] & & \mathcal{C} & & \mathcal{D} \arrow[rr, "G"] \arrow[dr, "G"'] & & \mathcal{C} \\
& \mathcal{C} \arrow[ur, "1_{\mathcal{C}}"'] \arrow[u, phantom, "1_G\Uparrow"] & & = & & \mathcal{C} \arrow[ur, dashed, "T^G"] \arrow[ur, bend right=60, "1_{\mathcal{C}}"'] \arrow[u, phantom, "\Uparrow\gamma"] \arrow[ur, phantom, bend right=30, "\eta"'] &
\end{tikzcd}$$

There is a natural transformation $T^GT^G \circ G \xrightarrow{T^G\gamma} T^G \circ G \xrightarrow{\gamma} G$ and therefore by the universal property of right Kan extensions there exists a unique natural transformation $\mu : T^GT^G \to T^G$ such that the following diagrams are equal:

$$\begin{tikzcd}
\mathcal{D} \arrow[rr, "G"] \arrow[dr, "G"'] & & \mathcal{C} & & \mathcal{D} \arrow[rr, "G"] \arrow[dr, "G"'] & & \mathcal{C} \\
& \mathcal{C} \arrow[ur, "T^GT^G"'] \arrow[u, phantom, "\Uparrow"] & & = & & \mathcal{C} \arrow[ur, dashed, "T^G"] \arrow[ur, bend right=60, "T^GT^G"'] \arrow[u, phantom, "\Uparrow\gamma"] \arrow[ur, phantom, bend right=30, "\mu"'] &
\end{tikzcd}$$

**Proposition 3.2.1.** *The triple* $(T^G, \eta, \mu)$ *is a monad.*

A proof for this result can be found in Section 2 of [45]. The monad in Proposition 3.2.1 is called the **codensity monad of** $G$.

**Example 3.2.2.** The codensity monad of a right adjoint functor $G : \mathcal{D} \to \mathcal{C}$ is the monad induced by the adjunction. This follows from the fact that $\mathrm{Ran}_G G = GF$, where $F$ is the left adjoint of $G$.

**Example 3.2.3** (Kennison and Gildenhuys)**.** Let $\mathbf{Set}_f$ be the category of finite sets and maps. The codensity monad of the inclusion functor $\mathbf{Set}_f \to \mathbf{Set}$ is the ultrafilter monad. This is a result from [38].

The following result, which follows from Theorem 3.7.2 in [8], will be useful to find codensity monads of certain functors.

**Proposition 3.2.4.** *Let* $G : \mathcal{D} \to \mathcal{C}$ *be a functor, where* $\mathcal{D}$ *is an essentially small category and* $\mathcal{C}$ *is a complete category. Then the codensity monad of* $G$ *exists and*

$$T^G(X) = \lim(X \downarrow G \xrightarrow{U} \mathcal{D} \xrightarrow{G} \mathcal{C})$$

*for every object* $X$ *in* $\mathcal{C}$*. Here* $U : X \downarrow G \to \mathcal{D}$ *is the forgetful functor.*

## 3.3 Premeasurable spaces

In this section we will discuss probability premeasures and their finitely additive analogues on premeasurable spaces. We will give an overview of results on integration with respect to a probability premeasure and end the section with an integral representation theorem. In the case of probability measures, these correspond to standard results in measure theory. Premeasures play an important role in extension theorems such as the Carathéodory extension theorem, which will be discussed from a categorical point of view in Chapter 4. Because integration with respect to a premeasure is far less common than the usual Lebesgue integral, we will give detailed proofs for all results.

We will call a set $X$ together with an algebra of subsets $\mathcal{B}_X$ (i.e. a collection of subsets of $X$ that is closed under complements and finite intersections that contain $\emptyset$ and $X$) a **premeasurable space**. We say that a map $f : X \to Y$ between premeasurable spaces is **premeasurable** if $f^{-1}(\mathcal{B}_Y) \subseteq \mathcal{B}_X$.

Let $\mathbb{P} : \mathcal{B}_X \to [0, 1]$ be a function such that $\mathbb{P}(X) = 1$. The function $\mathbb{P}$ is called a **probability premeasure** if $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ for every pairwise disjoint collection $(A_n)_{n=1}^{\infty}$ of subsets in $\mathcal{B}_X$ such that $\bigcup_{n=1}^{\infty} A_n$ is also in $\mathcal{B}_X$. If $\mathcal{B}_X$ is a $\sigma$-algebra, $\mathbb{P}$ is just a *probability measure*. If we only have that $\mathcal{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ for disjoint subsets $A$ and $B$ in $\mathcal{B}_X$, then we call $\mathbb{P}$ a **finitely additive probability premeasure**. This is also known as a **charge**. If $\mathcal{B}_X$ is a $\sigma$-algebra we say that $\mathbb{P}$ is a **finitely additive probability measure**. Note that every probability (pre)measure is a finitely additive probability (pre)measure. Let $(X, \mathcal{B}_X)$ be a premeasurable space. We will call a function $s : X \to [0, 1]$ a **simple function on** $X$ if there exists a natural number $n \geq 1$ and $a_1, a_2, \ldots, a_n \in [0, 1]$ and $A_1, A_2, \ldots, A_n \in \mathcal{B}_X$ such that

$$s = \sum_{k=1}^{n} a_k 1_{A_k}.$$

Note that a simple function is always premeasurable. The collection of simple functions on $X$ is denoted by $\mathrm{Simp}(X, [0,1])$. Given a finitely additive probability premeasure $\mathbb{P}$, we can define a map $J_{\mathbb{P}} : \mathrm{Simp}(X, [0,1]) \to [0,1]$ by the assignment

$$\sum_{k=1}^{n} a_k 1_{A_k} \mapsto \sum_{k=1}^{n} a_k \mathbb{P}(A_k).$$

Because $\mathcal{B}_X$ is closed under complements and finite unions, we can use a common refinement argument to show that the assignment is independent of the representation of the simple function.

Again using a common refinement argument, we obtain the following proposition.

**Proposition 3.3.1.** *For simple functions $s, t : X \to [0,1]$ such that $s + t \leq 1$ and $r \in [0,1]$, we have that also $s + t$ and $rs$ are simple functions. Furthermore, $J_{\mathbb{P}}(s+t) = J_{\mathbb{P}}(s) + J_{\mathbb{P}}(t)$ and $J_{\mathbb{P}}(rs) = rJ_{\mathbb{P}}(s)$. If $s \leq t$, then also $t - s$ is a simple function and $J_{\mathbb{P}}(s) \leq J_{\mathbb{P}}(t)$.*

The following useful lemma follows from the construction described in the proof of Corollary 4.5.9 in [4].

**Lemma 3.3.2.** *Let $f : X \to [0,1]$ be a premeasurable map. There exists an increasing sequence $(s_n)_n$ of simple functions that converges uniformly to $f$ and there exists a decreasing sequence $(t_n)_n$ of simple functions that converges uniformly to $f$.*

Let $\mathbf{PreMble}(X, [0,1])$ be the set of premeasurable maps from $X$ to $[0,1]$. We define a map $I_{\mathbb{P}} : \mathbf{PreMble}(X, [0,1]) \to [0,1]$ by the assignment

$$f \mapsto \sup \{ J_{\mathbb{P}}(s) \mid s \leq f \text{ and } s \in \mathrm{Simp}(X, [0,1]) \}.$$

This map is well-defined because $J_{\mathbb{P}}$ is order-preserving by Proposition 3.3.1.

The following proposition summarizes results about the additivity and continuity of $I_{\mathbb{P}}$. In the case of probability measures, these are classical results in measure theory. For finitely additive probability premeasures these are lesser-known. Similar results have been discussed in [63].

**Proposition 3.3.3.** *Let $X$ be a premeasurable space and let $\mathbb{P}$ be a finitely additive probability premeasure on $X$. We have the following properties:*

*(i) For a simple function $s : X \to [0,1]$, we have $I_{\mathbb{P}}(s) = J_{\mathbb{P}}(s)$. In particular $I_{\mathbb{P}}(1) = 1$.*

*(ii) For premeasurable functions $f, g : X \to [0,1]$ such that $f \leq g$, we have $I_{\mathbb{P}}(f) \leq I_{\mathbb{P}}(g)$.*

*(iii) For a premeasurable map $f : X \to [0,1]$,*

$$I_{\mathbb{P}}(f) = \inf \{ J_{\mathbb{P}}(s) \mid s \geq f \text{ and } s \in \mathrm{Simp}(X, [0,1]) \}.$$

*(iv) For premeasurable maps $f, g : X \to [0,1]$ such that also $f + g \in \mathbf{PreMble}(X, [0,1])$, we have $I_{\mathbb{P}}(f+g) = I_{\mathbb{P}}(f) + I_{\mathbb{P}}(g)$.*[1]

*(v) Suppose that $\mathbb{P}$ is a probability premeasure. For an increasing sequence of premeasurable maps $(f_n : X \to [0,1])_{n=1}^{\infty}$ such that $f := \lim_{n\to\infty} f_n$ is also premeasurable, $\lim_{n\to\infty} I_{\mathbb{P}}(f_n) = I_{\mathbb{P}}(f)$.*

---

[1] Note that the finite sum of premeasurable maps $X \to [0,1]$ is not necessarily premeasurable. Consider for example the set $[0,1/2]^2$ endowed with the Boolean algebra generated by open rectangles. The two projection maps $\pi_1, \pi_2 : [0,1/2]^2 \to [0,1]$ are premeasurable, but their sum is not.

*(vi) Let $(f_n : X \to [0,1])_{n=1}^{\infty}$ be a collection of premeasurable maps such that $f := \sum_{n=1}^{\infty} f_n$ is also an element of* **PreMble**$(X,[0,1])$*. If $\mathbb{P}$ is a probability premeasure, then*

$$I_{\mathbb{P}}(f) = \sum_{n=1}^{\infty} I_{\mathbb{P}}(f_n).$$

*Proof.* (i) Since $s \leq s$ it follows by the definition of $I$ that $J_{\mathbb{P}}(s) \leq I_{\mathbb{P}}(s)$. For every simple function $t$ such that $t \leq s$ we have by Proposition 3.3.1 that $J_{\mathbb{P}}(t) \leq J_{\mathbb{P}}(s)$ and therefore $I_{\mathbb{P}}(s) \leq J_{\mathbb{P}}(s)$.

(ii) This follows from the fact that for a simple function $s$ such that $s \leq f$ we also have $s \leq g$.

(iii) The '$\leq$' inequality is clear. Now consider an $\epsilon \in (0,1]$ and simple functions $s$ and $t$ such that $s \leq f \leq t$ and such that $\|t-s\|_{\infty} \leq \epsilon$, which exist by Lemma 3.3.2. We find the following inequalities:

$$J_{\mathbb{P}}(t) = J_{\mathbb{P}}(t-s+s) = J_{\mathbb{P}}(t-s) + J_{\mathbb{P}}(s) \leq J_{\mathbb{P}}(\epsilon) + J_{\mathbb{P}}(s) \leq \epsilon + I_{\mathbb{P}}(f).$$

Here we used Proposition 3.3.1 and the definition of $I_{\mathbb{P}}$. The other inequality now follows.

(iv) Let $\epsilon > 0$. By the definition of $J_{\mathbb{P}}$ and (iii), there exist simple functions $s_f, t_f, s_g$ and $t_g$ such that $s_f \leq f \leq t_f$ and $s_g \leq g \leq t_g$ and such that

$$J_{\mathbb{P}}(t_f) - \epsilon \leq I_{\mathbb{P}}(f) \leq J_{\mathbb{P}}(s_f) + \epsilon$$

and

$$J_{\mathbb{P}}(t_g) - \epsilon \leq I_{\mathbb{P}}(g) \leq J_{\mathbb{P}}(s_g) + \epsilon.$$

Since $s_f + s_g$ and $(t_f + t_g) \wedge 1$ are simple functions by Proposition 3.3.1 and $s_f + s_g \leq f + g \leq (t_f + t_g) \wedge 1$, we find that

$$I_{\mathbb{P}}(f+g) - 2\epsilon \leq J_{\mathbb{P}}((t_f + t_g) \wedge 1) - 2\epsilon \leq J_{\mathbb{P}}(t_f) + J_{\mathbb{P}}(t_g) - 2\epsilon \leq I_{\mathbb{P}}(f) + I_{\mathbb{P}}(g)$$

and

$$I_{\mathbb{P}}(f) + I_{\mathbb{P}}(g) \leq J_{\mathbb{P}}(s_f) + J_{\mathbb{P}}(s_g) + 2\epsilon = J_{\mathbb{P}}(s_f + s_g) + 2\epsilon \leq I_{\mathbb{P}}(f+g) + 2\epsilon.$$

Here we again used Proposition 3.3.1 and (iii) and the definition of $I_{\mathbb{P}}$. The result follows by letting $\epsilon \to 0$.

(v) Since $f_n \leq f$ we have by (ii) that $I_{\mathbb{P}}(f_n) \leq I_{\mathbb{P}}(f)$ and therefore $\lim_{n\to\infty} I_{\mathbb{P}}(f_n) \leq I_{\mathbb{P}}(f)$. Now consider a simple function $s = \sum_{k=1}^{m} a_k 1_{A_k}$ such that $s \leq f$. For $r \in [0,1)$ and a natural number $n$ define the set

$$E_{n,r} := \{x \in X \mid f_n(x) \geq rs(x)\}.$$

Note that $E_{n,r} = \bigcup_{k=1}^{m} (f_n^{-1}([ra_k, 1]) \cap A_k)$ and therefore it is a subset in $\mathcal{B}_X$.

The function $rs1_{E_{n,r}}$ is simple and satisfies $rs1_{E_{n,r}} \leq f_n$. It follows now that

$$I_{\mathbb{P}}(f_n) \geq J_{\mathbb{P}}(rs1_{E_{n,r}}) = \sum_{k=1}^{m} ra_k \mathbb{P}(A_k \cap E_{n,r}).$$

Taking the limit $n \to \infty$ on both sides of the inequality gives us that

$$\lim_n I_{\mathbb{P}}(f_n) \geq \sum_{k=1}^{m} r a_k \left( \lim_{n\to\infty} \mathbb{P}(A_k \cap E_{n,r}) \right). \tag{3.1}$$

It is easy to verify that $(E_{n,r})_{n\in\mathbb{N}}$ increases to $X$. Now define $F_{1,r} := E_{1,r}$ and $F_{n,r} := E_{n,r} \setminus F_{n-1,r}$ and note that $(F_{n,r} \cap A_k)_{n\in\mathbb{N}}$ is a collection of pairwise disjoint subsets in $\mathcal{B}_X$ such that their union is $A_k$. Using that $\mathbb{P}$ is a probability premeasure we find that

$$\mathbb{P}(A_k) = \sum_{n=1}^{\infty} \mathbb{P}(F_{n,r} \cap A_k) = \lim_{n\to\infty} \mathbb{P}\left( \bigcup_{l=1}^{n} F_{l,r} \cap A_k \right) = \lim_{n\to\infty} \mathbb{P}(E_{n,r} \cap A_k) \tag{3.2}$$

Combining (3.1) and (3.2) gives us $\lim_{n\to\infty} I_{\mathbb{P}}(f_n) \geq rJ(s)$ for all $r \in [0,1)$, which implies that

$$\lim_{n\to\infty} I_{\mathbb{P}}(f_n) \geq J_{\mathbb{P}}(s).$$

Since this holds for any simple function $s$ such that $s \leq f$, we can conclude that $\lim_{n\to\infty} I_{\mathbb{P}}(f_n) \geq I_{\mathbb{P}}(f)$.

(vi) We first show the result for the case that every $f_n$ is simple. Because the finite sum of simple functions is again a simple function we can use (iv) and (v) to show the following equalities:

$$I_{\mathbb{P}}(f) = I_{\mathbb{P}}\left( \lim_{n\to\infty} \sum_{k=1}^{n} f_k \right) = \lim_{n\to\infty} I_{\mathbb{P}}\left( \sum_{k=1}^{n} f_k \right) = \lim_{n\to\infty} \sum_{k=1}^{n} I_{\mathbb{P}}(f_k) = \sum_{n=1}^{\infty} I_{\mathbb{P}}(f_k).$$

For the general case we use Lemma 3.3.2 to write $f_n$ as $\sum_{k=1}^{\infty} s_{k,n}$, where $s_{k,n}$ is a simple function for every $k$ and $n$. This gives us that

$$f = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} s_{k,n}.$$

Let $\psi : \mathbb{N} \setminus \{0\} \to \mathbb{N} \setminus \{0\} \times \mathbb{N} \setminus \{0\}$ be a bijection. We can now rewrite $f$ as follows:

$$f = \sum_{m=1}^{\infty} s_{\psi(m)}.$$

By the above it follows now that

$$I_{\mathbb{P}}(f) = \sum_{m=1}^{\infty} I_{\mathbb{P}}(s_{\psi(m)}) = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} I_{\mathbb{P}}(s_{k,n}) = \sum_{n=1}^{\infty} I_{\mathbb{P}}(f_n).$$

□

**Remark 3.3.4.** In the case that $\mathcal{B}_X$ is a $\sigma$-algebra, $I_{\mathbb{P}}$ becomes the usual integration operation and Proposition 3.3.3(v) and (vi) become the usual Monotone Convergence Theorems. Therefore we will also for the other cases write $\int_X f \mathrm{d}\mathbb{P}$ or $\int_X f(x)\mathbb{P}(\mathrm{d}x)$ for $I_{\mathbb{P}}(f)$.

**Proposition 3.3.5.** *Let* $I : \mathbf{PreMble}(X,[0,1]) \to [0,1]$ *be a map such that* $I(1) = 1$. *If* $I(f+g) = I(f) + I(g)$ *for all premeasurable maps* $f, g : X \to [0,1]$ *such that also* $f + g \in \mathbf{PreMble}(X,[0,1])$, *then there exists a unique finitely additive probability premeasure* $\mathbb{P}$ *such that*

$$I = I_{\mathbb{P}}.$$

*Proof.* Define $\mathbb{P}(A) := I(1_A)$ for all $A$ in $\Sigma_X$. Clearly, $\mathbb{P}$ is a finitely additive probability premeasure.

We have $I(qf) = qI(f)$ for all $q \in [0,1] \cap \mathbb{Q}$ and for all premeasurable functions $f : X \to [0,1]$. For a simple function $s$ such that $s \leq f$ we have that $f - s$ is a premeasurable function and it follows that $I(s) \leq I(f)$.

For a premeasurable function $f : X \to [0,1]$ and for $\epsilon > 0$ there exists a simple function $s = \sum_{k=1}^{m} a_k 1_{A_k} \leq f$ such that $I_{\mathbb{P}}(f) \leq \sum_{k=1}^{m} a_k \mathbb{P}(A_k) + \epsilon$. We can assume that $a_k$ is an element of $[0,1] \cap \mathbb{Q}$ for every $k$. We now see that

$$I_{\mathbb{P}}(f) \leq \sum_{k=1}^{m} a_k I(1_{A_k}) + \epsilon = I(s) + \epsilon \leq I(f) + \epsilon.$$

Letting $\epsilon \to 0$ gives us that $I_{\mathbb{P}}(f) \leq I(f)$. Using Proposition 3.3.3(iii) we obtain the other inequality in a similar way.

Let $\mathbb{P}'$ be another finitely additive probability premeasure with this property; then

$$\mathbb{P}'(A) = I_{\mathbb{P}'}(1_A) = I(1_A) = I_{\mathbb{P}}(1_A) = \mathbb{P}(A)$$

for every $A$ in $\mathcal{B}_X$. This implies $\mathbb{P}' = \mathbb{P}$. □

**Proposition 3.3.6.** *Let* $I : \mathbf{PreMble}(X,[0,1]) \to [0,1]$ *be a map such that* $I(1) = 1$. *If* $I\left(\sum_{n=1}^{\infty} f_n\right) = \sum_{n=1}^{\infty} I(f_n)$ *for every collection of premeasurable maps* $(f_n : X \to [0,1])_{n=1}^{\infty}$ *such that* $f := \sum_{n \in A} f_n$ *is also an element of* $\mathbf{PreMble}(X,[0,1])$ *for every finite or cofinite subset* $A$ *of* $\mathbb{N} \setminus \{0\}$, *then there exists a unique probability premeasure* $\mathbb{P}$ *such that*

$$I = I_{\mathbb{P}}.$$

*Proof.* By Proposition 3.3.5 there exists a unique finitely additive probability measure $\mathbb{P}$ such that $I = I_{\mathbb{P}}$. For a pairwise disjoint collection $(A_n)_{n=1}^{\infty}$ in $\mathcal{B}_X$ such that $\bigcup_{n=1}^{\infty} A_n$ is also an element of $\mathcal{B}_X$, we see that $\sum_{n \in A} 1_{A_n}$ is premeasurable for every finite or cofinite subset $A$ of $\mathbb{N} \setminus \{0\}$. It follows now that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = I_{\mathbb{P}}\left(\sum_{n=1}^{\infty} 1_{A_n}\right) = \sum_{n=1}^{\infty} I_{\mathbb{P}}(1_{A_n}) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

□

**Corollary 3.3.7.** *Let* $X$ *be a measurable space and let* $I : \boldsymbol{Mble}(X,[0,1]) \to [0,1]$ *be a map such that* $I(1) = 1$. *If* $I\left(\sum_{n=1}^{\infty} f_n\right) = \sum_{n=1}^{\infty} I(f_n)$ *for every collection of measurable maps* $(f_n : X \to [0,1])_{n=1}^{\infty}$ *such that* $\sum_{n=1}^{\infty} f_n \leq 1$, *then there exists a unique probability measure* $\mathbb{P}$ *such that*

$$I = I_{\mathbb{P}}.$$

## 3.4 Probability monads on categories of premeasurable spaces

In this section we will present several monads of (finitely additive) probability (pre)measures on categories of (pre)measurable spaces. We explain how each of these monads can be constructed as a codensity monad.

We will denote the category of measurable spaces and measurable maps by **Mble**. We write $\mathbf{Set}_c$ for the category of countable sets and functions.

### 3.4.1 Giry monad of probability measures

Here we will discuss the monad of probability measures on the category of measurable spaces, which is known as the *Giry monad*. We show how this monad arises as the codensity monad of a functor $G : \mathbf{Set}_c \to \mathbf{Mble}$.

Let $(X, \Sigma_X)$ be a measurable space and let $\mathcal{G}X$ be the set of all probability measures on $X$. For a measurable subset $A$ of $X$ let $\mathrm{ev}_A : \mathcal{G}X \to [0, 1]$ denote the function defined by the assignment $\mathbb{P} \mapsto \mathbb{P}(A)$. The set $\mathcal{G}X$ becomes a measurable space by endowing it with the smallest $\sigma$-algebra that makes $\mathrm{ev}_A$ measurable for all $A$ in $\Sigma_X$. We will denote this measurable space also by $\mathcal{G}X$.

Let $f : X \to Y$ be a measurable map between measurable spaces $X$ and $Y$. Every probability measure $\mathbb{P}$ induces a probability measure $\mathbb{P} \circ f^{-1}$ on $Y$ which is defined by

$$\mathbb{P} \circ f^{-1}(B) := \mathbb{P}(f^{-1}(B))$$

for all $B$ in $\Sigma_Y$. The assignment $\mathbb{P} \mapsto \mathbb{P} \circ f^{-1}$ defines a map $\mathcal{G}X \to \mathcal{G}Y$, which we will denote by $\mathcal{G}f$. It can be checked that $\mathcal{G}f$ is measurable.

The assignments $X \mapsto \mathcal{G}X$ and $f \mapsto Gf$ define a functor $\mathcal{G} : \mathbf{Mble} \to \mathbf{Mble}$.

For every measurable space $X$ we have a measurable map $\eta_X : X \to \mathcal{G}X$ that sends an element $x \in X$ to the probability measure $\delta_x$ that is defined by

$$\delta_x(A) := \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

Moreover, these maps form a natural transformation $\eta : 1_{\mathbf{Mble}} \to \mathcal{G}$.

We also have a measurable map $\mu_X : \mathcal{G}\mathcal{G}X \to \mathcal{G}X$ that sends a probability measure $\mathbf{P}$ on $\mathcal{G}X$ to the probability measure $\mu_X(\mathbf{P})$ on $X$ that is defined by

$$\mu_X(\mathbf{P})(A) := \int_{\mathcal{G}X} \mathbb{P}(A)\mathbf{P}(\mathrm{d}\mathbb{P}) \tag{3.4}$$

for all $A \in \Sigma_X$. Also these maps form a natural transformation $\mu : \mathcal{G}\mathcal{G} \to \mathcal{G}$. As discussed in Theorem 2.1.2 and Theorem 1 in [28], the triple $(\mathcal{G}, \eta, \mu)$ is a monad, the *Giry monad*.

Every countable set $A$ can be turned into a measurable space, namely the set $A$ endowed with the whole powerset of $A$ as $\sigma$-algebra. Every function of countable sets becomes measurable with respect to these $\sigma$-algebras. This leads to a functor $j : \mathbf{Set}_c \to \mathbf{Mble}$. Define the functor $G$ as

$$\mathbf{Set}_c \xrightarrow{j} \mathbf{Mble} \xrightarrow{\mathcal{G}} \mathbf{Mble}$$

This means that for a countable set $A$ the underlying set of $GA$ is equal to

$$\left\{(p_a)_a \in [0,1]^A \mid \sum_{a\in A} p_a = 1\right\}.$$

For a map $f : X \to GA$ we will use the notation $f_a$ to mean $\mathrm{ev}_{\{a\}} \circ f$. Note that a map $f : X \to GA$ is measurable if and only if $f_a$ is measurable for every $a$. For a map of countable sets $f : A \to B$ the measurable map $Gf : GA \to GB$ is given by the assignment

$$(p_a)_{a\in A} \mapsto \left(\sum_{a\in f^{-1}(b)} p_a\right)_{b\in B}.$$

**Theorem 3.4.1.** *The Giry monad is the codensity monad of $G$.*

*Proof.* Proposition 3.2.4 tells us that the codensity monad of $G$ exists. We will now show that for all measurable spaces $X$,

$$\mathcal{G}(X) = \lim(X \downarrow G \xrightarrow{U} \mathbf{Set}_c \xrightarrow{G} \mathbf{Mble}),$$

which is natural in $X$. Proposition 3.2.4 then implies that $T^G \cong \mathcal{G}$ for all measurable spaces $X$. For a measurable map $f : X \to GA$ define a map $p_f$ as

$$\mathcal{G}X \xrightarrow{\mathcal{G}f} \mathcal{G}\mathcal{G}jA \xrightarrow{\mu_{jA}} \mathcal{G}jA = GA.$$

This means that for $\mathbb{P} \in \mathcal{G}X$,

$$p_f(\mathbb{P}) = \left(\int_X f_a \mathrm{d}\mathbb{P}\right)_{a\in A}.$$

Consider a commutative triangle

$$\begin{tikzcd} & X \arrow[dl, "f"'] \arrow[dr, "g"] & \\ GA \arrow[rr, "Gs"'] & & GB. \end{tikzcd}$$

We have the following equalities:

$$\begin{aligned} Gs \circ p_f &= Gs \circ \mu_{jA} \circ \mathcal{G}f \\ &= \mathcal{G}js \circ \mu_{jA} \circ \mathcal{G}f \\ &= \mu_{jB} \circ \mathcal{G}\mathcal{G}js \circ \mathcal{G}f \\ &= \mu_{jB} \circ \mathcal{G}g = p_g \end{aligned}$$

In more measure theoretic terms this means that for a probability measure $\mathbb{P}$ on $X$ and an element $b$ in $B$ we have the following:

$$(Gs \circ p_f)(\mathbb{P})_b = \sum_{a\in s^{-1}(b)} \int_X f_a \mathrm{d}\mathbb{P} = \int_X \sum_{a\in s^{-1}(b)} f_a \mathrm{d}\mathbb{P} = \int_X (Gs \circ f)_b \mathrm{d}\mathbb{P} = \int_X g_b \mathrm{d}\mathbb{P} = p_g(\mathbb{P})_b.$$

We can conclude that $(\mathcal{G}X, (p_f)_f)$ forms a cone over the diagram $X \downarrow G \xrightarrow{U} \mathbf{Set}_c \xrightarrow{G} \mathbf{Mble}$.

For a measurable map $g : X \to Y$ and a measurable map $f : X \to GA$, we have the following commutative diagram.

$$\begin{tikzcd}
\mathcal{G}X \arrow[rr, "\mathcal{G}g"] \arrow[dr, "p_{f\circ g}"'] & & \mathcal{G}Y \arrow[dl, "p_f"] \\
& GA &
\end{tikzcd}$$

It follows now that there is a natural transformation $\varphi : \mathcal{G} \to T^G$.

We will now show that $(\mathcal{G}X, (p_f)_f)$ is the limiting cone of the diagram $X \downarrow G \overset{U}{\Longrightarrow} \mathbf{Set}_c \xrightarrow{G} \mathbf{Mble}$. This will imply that $\varphi$ is in fact a natural isomorphism. To do this let us consider some cone $(Y, (q_f)_f)$ over the diagram. Let $\mathbf{2} := \{0, 1\}$ and $\mathbf{1} := \{0\}$. For a measurable map $f : X \to [0, 1]$ let $\widehat{f} : X \to G\mathbf{2}$ be the measurable map that sends an element $x$ in $X$ to $(1 - f(x), f(x))$.

For an element $y \in Y$ define a map $I_y : \mathbf{Mble}(X, [0, 1]) \to [0, 1]$ by

$$I_y(f) := q_{\widehat{f}}(y)_1.$$

Let $t : \mathbf{1} \to \mathbf{2}$ be the map that sends 0 to 1 and let $e$ be the unique measurable map $X \to G\mathbf{1}$. We find the following commutative triangle:

$$\begin{tikzcd}
& X \arrow[dl, "\widehat{1}"'] \arrow[dr, "e"] & \\
G\mathbf{2} & & G\mathbf{1}. \arrow[ll, "Gt"]
\end{tikzcd}$$

Because $(Y, (q_f)_f)$ is a cone over the diagram we also have that the following triangle commutes:

$$\begin{tikzcd}
& Y \arrow[dl, "q_{\widehat{1}}"'] \arrow[dr, "q_e"] & \\
G\mathbf{2} & & G\mathbf{1}. \arrow[ll, "Gt"]
\end{tikzcd}$$

It now follows that $I_y(1) = q_{\widehat{1}}(y)_1 = 1$.

For a collection $(f_n : X \to [0, 1])_{n=1}^{\infty}$ of measurable maps such that $f := \sum_{n=1}^{\infty} f_n$ is also an element of $\mathbf{Mble}(X, [0, 1])$, let $h : X \to G\mathbb{N}$ be the measurable map defined by $h(x)_n := f_n(x)$ for $n \geq 1$ and $h(x)_0 := 1 - f(x)$. For $n \geq 1$ let $s_n : \mathbb{N} \to \mathbf{2}$ be the map that sends $n$ to 1 and every other element to 0. Let $s : \mathbb{N} \to \mathbf{2}$ be the map that sends 0 to 0 and every other element to 1. We have the following commutative diagrams:

$$\begin{tikzcd}
& X \arrow[dl, "h"'] \arrow[dr, "\widehat{f_n}"] & \\
G\mathbb{N} \arrow[rr, "Gs_n"'] & & G\mathbf{2}
\end{tikzcd}
\qquad
\begin{tikzcd}
& X \arrow[dl, "h"'] \arrow[dr, "\widehat{f}_0"] & \\
G\mathbb{N} \arrow[rr, "Gs"'] & & G\mathbf{2}
\end{tikzcd}$$

Therefore also the following triangles commute:

$$\begin{tikzcd}
& Y \arrow[dl, "q_h"'] \arrow[dr, "q_{\widehat{f_n}}"] & \\
G\mathbb{N} \arrow[rr, "Gs_n"'] & & G\mathbf{2}
\end{tikzcd}
\qquad
\begin{tikzcd}
& Y \arrow[dl, "q_h"'] \arrow[dr, "q_{\widehat{f}}"] & \\
G\mathbb{N} \arrow[rr, "Gs"'] & & G\mathbf{2}
\end{tikzcd}$$

It follows now that

$$I_y(f_0) = q_{\widehat{f}}(y)_1 = (Gs \circ q_h(y))_1 = \sum_{n\in\mathbb{N}} q_h(y)_n \tag{3.5}$$

and that for every $n \geq 1$,

$$I_y(f_n) = q_{\widehat{f_n}}(y)_1 = (Gs_n \circ q_h(y))_1 = q_h(y)_n. \tag{3.6}$$

Combining (3.5) and (3.6) gives us that $I_y(f) = \sum_{n\in\mathbb{N}} I_y(f_n)$. By Corollary 3.3.7 it follows that there exists a unique probability measure $\mathbb{P}_y$ such that $I_y = I_{\mathbb{P}_y}$. The assignment $y \mapsto \mathbb{P}_y$ defines a map $q : Y \to \mathcal{G}X$. We have that $\mathrm{ev}_A \circ q = (q_{\widehat{1_A}})_1$ for all $A \in \Sigma_X$ and therefore $q$ is measurable.

Let $f : X \to GA$ be a measurable map and let $a$ be an element of $A$. Let $s_a : A \to \mathbf{2}$ be the map that sends $a$ to 1 and every other element to 0. Since we have that $Gs_a \circ f = \widehat{f_a}$ we also have that $Gs_a \circ q_f = q_{\widehat{f_a}}$. In particular we find for every $y \in Y$ that

$$I_y(f_a) = q_{\widehat{f_a}}(y)_1 = (Gs_a \circ q_f(y))_1 = q_f(y)_a.$$

Using this we obtain for every $y \in Y$ and for every $a \in A$ that

$$[p_f \circ q(y)]_a = I_{\mathbb{P}_y}(f_a) = I_y(f)_a = q_f(y)_a.$$

This shows that $q$ is a morphism of cones from $(\mathcal{G}X, (p_f)_f)$ to $(Y, (q_f)_f)$.

Let $\tilde{q} : Y \to \mathcal{G}X$ be another morphism of cones. Then for every measurable subset $A$ of $X$ and for every $y$ in $Y$ we have that

$$\tilde{q}(y)(A) = (p_{\widehat{1_A}} \circ \tilde{q}(y))_1 = q_{\widehat{1_A}}(y)_1 = q(y)(A).$$

This shows that $\tilde{q} = q$ and therefore $(\mathcal{G}X, (p_f)_f)$ is the limiting cone over the diagram. This implies that $\mathcal{G}(X) \cong T^G(X)$ for all measurable spaces $X$. Moreover this induces a natural isomorphism $\mathcal{G} \cong T^G$.

The counit $\epsilon$ of the codensity monad is given

$$\epsilon_A : \mathcal{G}GA \xrightarrow{p_{1_{GA}}} GA : \mathbb{P} \mapsto \left(\int_{GA} p_a \mathbb{P}(\mathrm{d}p)\right)_a.$$

It follows now that the multiplication of the Giry monad satisfies the universal property of the multiplication of the codensity monad, namely that it is the unique natural transformation $\mathcal{G}\mathcal{G} \to \mathcal{G}$ such that

$$\begin{tikzcd}
\mathbf{Set}_c \arrow[r, "G"] \arrow[d, "G"'] \arrow[d, phantom, shift left=6, "\overset{\epsilon\circ\mathcal{G}\epsilon}{\Rightarrow}"] & \mathbf{Mble} \\
\mathbf{Mble} \arrow[ur, "\mathcal{G}\mathcal{G}"'] &
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
\mathbf{Set}_c \arrow[r, "G"] \arrow[d, "G"'] \arrow[d, phantom, shift left=6, "\overset{\epsilon}{\Rightarrow}"] & \mathbf{Mble} \\
\mathbf{Mble} \arrow[ur, "\mathcal{G}"] \arrow[ur, bend right=40, "\mathcal{G}\mathcal{G}"'] \arrow[ur, phantom, bend right=20, "\overset{\mu}{\Leftarrow}"] &
\end{tikzcd}$$

In a similar way it can be shown that the unit of the codensity monad of $G$ is equal to the unit of the Giry monad. ☐

**Remark 3.4.2.** While Theorem 3.4.1 states that $\mathrm{Ran}_G G = \mathcal{G}$, it is also true that $\mathrm{Ran}_j G = \mathcal{G}$. This construction immediately gives probability measures as set functions, without using an integral representation theorem.

**Remark 3.4.3.** A different construction for the Giry monad as a codensity monad is given in [3]. Avery shows that the codensity monad of the inclusion of the category of powers of the unit interval and affine maps in **Mble** is isomorphic to the Giry monad.

### 3.4.2 Giry monad of finitely additive probability measures

In the same way as the Giry monad of probability measures was defined, we can define a monad $(\mathcal{G}_f, \eta, \mu)$ of finitely additive probability measures on **Mble**. We call this monad the **Giry monad of finitely additive probability measures**.

Let $\mathbf{Set}_f$ be the category of finite sets and maps and let $i : \mathbf{Set}_f \to \mathbf{Set}_c$ be the inclusion functor. Let $G : \mathbf{Set}_c \to \mathbf{Mble}$ be as in the previous subsection and define $G_f := G \circ i$.

**Theorem 3.4.4.** *The Giry monad of finitely additive probability measures is the codensity monad of $G_f$.*

The proof for Theorem 3.4.4 is similar to the proof of Theorem 3.4.1. There are two places where the proof is slightly different. First, instead of using a countable index set $\mathbb{N}$, it is now enough to use a finite index set. Second, for this proof we use the integral representation result Proposition 3.3.5 instead of Corollary 3.3.7.

### 3.4.3 Giry monad of probability premeasures

In this section we will discuss how the monad of probability premeasures on the category of premeasurable spaces arises as the codensity monad of a functor $G_p$. This functor $G_p$ is similar to the functor $G$ in Section 3.4.1, however here the domain needs to be restricted to *finite* maps, because we are working with premeasurable maps.

Let $(X, \mathcal{B}_X)$ be a premeasurable space and let $\mathcal{G}_pX$ be the set of all probability premeasures on $X$. For $A \in \mathcal{B}_X$, let $\mathrm{ev}_A : \mathcal{G}_p(X) \to [0,1]$ be the map that sends a probability premeasure $\mathbb{P}$ to $\mathbb{P}(A)$. The set $\mathcal{G}_pX$ becomes a premeasurable space by endowing it with the smallest algebra that makes $\mathrm{ev}_A$ premeasurable for all $A \in \mathcal{B}_X$. We will denote this premeasurable space also by $\mathcal{G}_pX$.

Every premeasurable map $f : X \to Y$ induces a premeasurable map $\mathcal{G}_p : \mathcal{G}_pX \to \mathcal{G}_pY$ by pushing forward probability premeasures along $f$. This defines a functor $\mathcal{G}_f : \mathbf{PreMble} \to \mathbf{PreMble}$.

We can define natural transformations $\eta_p : 1_{\mathbf{PreMble}} \to \mathcal{G}_p$ and $\mu_p : \mathcal{G}_p\mathcal{G}_p \to \mathcal{G}_p$ in a similar way as we defined the unit (3.3) and the multiplication (3.4) for the Giry monad of probability measures. Similarly to Giry monad this construction gives us a monad. This can be seen by combining the proof of Theorem 2.1.2 together with Proposition 3.3.3.

**Proposition 3.4.5.** *The triple $(\mathcal{G}_p, \eta_p, \mu_p)$ is a monad.*

We call the monad in Proposition 3.4.5 the **Giry monad of probability premeasures**.

For a countable set $A$ let $\mathcal{P}_f(A)$ be the set of all finite and cofinite subsets of $A$. A map $f : A \to B$ between countable sets is called a **finite map** if $f^{-1}(\mathcal{P}_f(B)) \subseteq \mathcal{P}_f(A)$. Note that the composite of finite maps is a finite map. Let $\mathbf{Set}_c^f$ be the category of countable sets and finite maps. Every countable set $A$ can be turned into a premeasurable space, namely the set $A$ together with the algebra $\mathcal{P}_f(A)$. Every finite map between countable sets becomes premeasurable with respect to these algebras. We obtain a functor $j_p : \mathbf{Set}_c^f \to \mathbf{PreMble}$.

Now define the functor $G_p$ as

$$\mathbf{Set}_c^f \xrightarrow{j_p} \mathbf{PreMble} \xrightarrow{\mathcal{G}_p} \mathbf{PreMble}.$$

For a countable set $A$ the space $G_pA$ is the set $\{(p_a)_a \in [0,1]^A \mid \sum_{a\in A} p_a = 1\}$ together with the smallest algebra that makes $\mathrm{ev}_{A'}$ premeasurable for every finite or cofinite subset $A'$ of $A$. A map $f : X \to G_pA$ is premeasurable if and only if $\mathrm{ev}_{A'} \circ f$ is premeasurable for every $A' \in F_A$.

For a finite map of countable sets $f : A \to B$ the premeasurable map $G_pf : G_pA \to G_pB$ is given by the assignment

$$(p_a)_{a\in A} \mapsto \left( \sum_{a\in f^{-1}(b)} p_a \right)_{b\in B}$$

**Theorem 3.4.6.** *The Giry monad of probability premeasures is the codensity monad of $G$.*

Because all the maps between countable sets in the proof of Theorem 3.4.1 were finite, we can use a similar argument to prove Theorem 3.4.6. However now we need to use Proposition 3.3.6 instead of its corollary (Corollary 3.3.7).

**Remark 3.4.7.** Let $i : \mathbf{Set}_c^f \to \mathbf{Set}_c$ be the inclusion functor. We have that the codensity monad of $Gi$ is the Giry monad of probability measures, since in the proof of Theorem 3.4.1 we only used finite maps.

## 3.5 Hausdorff spaces

In Section 3.3 we presented several results about probability measures on measurable spaces. If we assume that the measurable space has more structure, we can say more about the probability measures on that space. This will be the topic of the current section. We will assume extra topological structure on the space and study the consequences for the probability measures on that space. We end the section with a generalized Daniell–Stone theorem.

### 3.5.1 Probability measures on Hausdorff spaces

Let $X$ be a Hausdorff space. Recall that the smallest $\sigma$-algebra that contains all the open sets of $X$ is called the *Borel $\sigma$-algebra* and is denoted by $\mathrm{Bo}_X$. The smallest $\sigma$-algebra that makes all the continuous functions $f : X \to [0,1]$ measurable is called the **Baire $\sigma$-algbera** and is denoted by $\mathrm{Ba}_X$. The following result, which is Theorem 7.1.1 in [14], states that for metric spaces these $\sigma$-algebras are the same.

**Proposition 3.5.1.** *For every Hausdorff space $X$ we have that $Ba_X \subseteq Bo_X$. For a metric space $X$ we have that $Ba_X = Bo_X$.*

A probability measure on $(X, \mathrm{Bo}_X)$ is called a **Borel probability measure** and a probability measure on $(X, \mathrm{Ba}_X)$ a **Baire probability measure**. Recall that a Borel probability measure $\mathbb{P}$ is called a *Radon probability measure* if

$$\mathbb{P}(A) = \sup\{\mathbb{P}(K) \mid K \subseteq A \text{ and } K \text{ is compact}\}$$

for all $A$ in $\mathrm{Bo}_X$. The next result tells us that on compact Hausdorff spaces, Baire probability measures correspond to Radon probability measures.

**Proposition 3.5.2.** *Every Baire probability measure on a compact Hausdorff space can be extended uniquely to a Radon probability measure.*

A proof for Proposition 3.5.2 can be found in [14] (Theorem 7.1.5).

### 3.5.2 Representations of probability measures on Hausdorff spaces

For a collection $L$ of real-valued functions on a set $X$ let $\sigma(L)$ denote the smallest $\sigma$-algebra on $X$ such that every $f \in L$ is measurable. Note that this $\sigma$-algebra is generated by sets of the form $\{f > r\} := \{x \in X \mid f(x) > r\}$ for $f \in L$ and $r \in \mathbb{R}$.

**Definition 3.5.3.** Let $X$ be a set and let $L \subseteq [0,\infty)^X$ be a subset of non-negative valued functions on $X$ together with the pointwise ordering. Let $\mathbb{N}L := \{nf \mid f \in L\}$. We call $L$ a **weak integration lattice**[2] if:

- $1 \in L$,
- for all $f, g \in L$ we have $f \vee g$, $f \wedge g$, $f \vee g - f \wedge g \in \mathbb{N}L$,
- for all $f \in L$ and for all $n \in \mathbb{N}$, $nf \wedge 1 \in \mathbb{N}L$ and
- for all $f \in L$ and for all $r \in [0,1]$, $rf \in L$.

**Definition 3.5.4.** We call a map $I : L \to [0,\infty)$ a **weak integration operator** if $I(1) = 1$ and for every collection $(f_n)_{n\in\mathbb{N}}$ in $L$ such that $f := \sum_{n\in\mathbb{N}} f_n \in L$, we have $I(f) = \sum_{n\in\mathbb{N}} I(f_n)$.

The following result is the Daniell–Stone representation theorem with weaker conditions. A proof, based on [39], is given in Appendix A.

**Theorem 3.5.5.** *Let $I$ be a weak integration operator on a weak integration lattice $L$ on a set $X$. There exists a unique probability measure $\mathbb{P}$ on $(X, \sigma(L))$ such that*

$$I_{\mathbb{P}}(f) = I(f)$$

*for all $f \in L$.*

**Proposition 3.5.6.** *Let $X$ be a compact Hausdorff space and let $L$ be a weak integration lattice such that every $f \in L$ is continuous and such that $f + g \in \mathbb{N}L$ for all $f, g \in L$. Let $I : L \to [0,\infty)$ be a function such that $I(f+g) = I(f) + I(g)$ for $f, g \in L$ with $f + g \in L$. Suppose also that $I(1) = 1$. Then there exists a unique probability measure $\mathbb{P}$ on $(X, \sigma(L))$ such that $I_{\mathbb{P}}(f) = I(f)$ for $f \in L$.*

The proof of Proposition 3.5.6 of this is given in Appendix A. The proof relies on Dini's theorem. These results can be used to prove several integral representation theorems of probability measures on certain topological spaces.

**Example 3.5.7.** Let $X$ be a metric space and let $L$ be the set of all 1-Lipschitz functions on $X$ taking values in $[0,1]$. Then $L$ is a weak integration lattice and $\sigma(L) = \mathrm{Ba}_X = \mathrm{Bo}_X$. To see this, observe that for every closed set $A$, the function $f_A := d(\cdot, A) \wedge 1 : X \to [0,1]$ is 1-Lipschitz and $A = f_A^{-1}(0) \in \sigma(L)$. Therefore $\mathrm{Bo}_X \subseteq \sigma(L)$ and clearly $\sigma(L) \subseteq \mathrm{Ba}_X$.

It follows now by Theorem 3.5.5 that every weak integration operator $I$ on $L$ induces a unique Borel probability measure on $X$ such that $I_{\mathbb{P}} = I$.

**Example 3.5.8.** Let $(X, d)$ be a metric space. A Borel probability measure $\mathbb{P}$ on $X$ is said to **have finite moment** if for all $x \in X$ we have

$$\int_X d(x,y)\mathbb{P}(\mathrm{d}y) < \infty.$$

---

[2] We use the term *weak* to indicate that this is the weaker version of the integration lattices discussed in Appendix A. For example, the subset of all 1-Lipschitz functions on a metric space form a *weak* integration lattice, but not an integration lattice.

Let $X$ be a metric space and let $L$ be the set of all 1-Lipschitz functions on $X$ taking values in $[0, \infty)$. Then $L$ is a weak integration lattice and $\sigma(L) = \mathrm{Ba}_X = \mathrm{Bo}_X$. Theorem 3.5.5 now tells us that there exists a unique Borel probability measure $\mathbb{P}$ such that $I_{\mathbb{P}} = I$. Since for every $x \in X$ the function $d(x, \cdot) : X \to [0, \infty)$ is a 1-Lipschitz function, we find

$$\int_X d(x,y)\mathbb{P}(\mathrm{d}y) = I_{\mathbb{P}}(d(x,\cdot)) = I(d(x,\cdot)) < \infty$$

and therefore $\mathbb{P}$ has finite moment.

**Example 3.5.9.** In Example 3.5.7 and 3.5.8 we can also use all Lipschitz maps instead of only the 1-Lipschitz maps.

**Example 3.5.10** (Riesz–Markov representation theorem)**.** Let $X$ be a compact Hausdorff space and let $L$ be the set of all continuous functions on $X$ taking values in $[0, 1]$. Then $L$ is a weak integration lattice and $\sigma(L) = \mathrm{Ba}_X$.

Let $I : L \to [0, 1]$ be a function such that $I(1) = 1$ and such that $I$ preserves binary sums that exist in $L$. Then Proposition 3.5.6 gives us a unique Baire probability measure $\mathbb{P}$ on $X$ such that $I_{\mathbb{P}}(f) = I(f)$ for all $f \in L$. Note that by Proposition 3.5.2 the Baire probability measure $\mathbb{P}$ can be uniquely extended to a Radon probability measure.

**Example 3.5.11.** Let $X$ be a compact metric space and let $L$ be the set of all 1-Lipschitz functions on $X$ taking values in $[0, 1]$. Then $L$ is a weak integration lattice that satisfies the conditions of Proposition 3.5.6.

Therefore for a function $I : L \to [0, 1]$ such that $I(f + g) = I(f) + I(g)$ for all $f, g \in L$ with $f + g$ also in $L$ and such that $I(1) = 1$, there exists a unique Radon probability measure $\mathbb{P}$ on $X$ such that $I_{\mathbb{P}} = I$.

**Example 3.5.12.** Let $X$ be a Hausdorff space and let $L$ be the set of continuous functions $X \to [0, 1]$. We see that $L$ is a weak integration lattice and $\sigma(L) = \mathrm{Ba}_X$. Using Theorem 3.5.5 we see that for every weak integration operator $I$ on $L$ there is a unique Baire probability measure $\mathbb{P}$ such that $I_{\mathbb{P}} = I$.

## 3.6 Probability monads on categories of Hausdorff spaces

In this section we will present monads of Radon probability measures and Baire probability measures and explain how they arise as codensity monads. We will introduce two new probability monads: the *bounded Lipschitz monad* and the *Baire monad*.

We write **Haus** for the category of Hausdorff spaces and continuous maps. The full subcategory of compact Hausdorff spaces is denoted by **CH**. The category of compact metric spaces and 1-Lipschitz maps is denoted by $\mathbf{KMet}_1$. For the category of finite sets and functions we write $\mathbf{Set}_f$.

### 3.6.1 Radon monad

We will discuss a monad of Radon probability measures on the category of compact Hausdorff spaces, known as the *Radon monad*. This monad was first introduced in [56] and [61] and further discussed in [34] and [36]. We explain how this monad can be constructed as the codensity monad of a functor $R : \mathbf{Set}_f \to \mathbf{CH}$. Although Radon probability measures are *$\sigma$-additive*, the domain of $R$ is the category of *finite* sets and maps. Note that compact Hausdorff spaces arise as the algebras of the codensity monad of the inclusion $\mathbf{Set}_f \to \mathbf{Set}$ (i.e. the ultrafilter monad) as discussed in [38] and [45].

Let $X$ be a topological space and let $\mathcal{R}X$ be the set of all Radon probability measures on $X$. Endow $\mathcal{R}X$ with the smallest topology such that the evaluation map $\mathrm{ev}_f : \mathcal{R}X \to [0,1]$, that sends $\mathbb{P}$ to $\int_X f \mathrm{d}\mathbb{P}$, becomes continuous for every continuous function $f : X \to [0,1]$. This is the topology of *weak convergence of probability measures.* We will denote the obtained topological space also by $\mathcal{R}X$.

The following result follows from the Banach-Alaoglu theorem (Section 3.15 in [54]).

**Proposition 3.6.1.** *The topological space $\mathcal{R}X$ is a compact Hausdorff space.*

**Lemma 3.6.2.** *Let $f : X \to Y$ be a continuous function between compact Hausdorff spaces and let $\mathbb{P}$ be a Radon probability measure on $X$. Then also $\mathbb{P} \circ f^{-1}$ is a Radon probability measure. Furtermore, the assignment $\mathbb{P} \mapsto \mathbb{P} \circ f^{-1}$ defines a continuous function $\mathcal{R}f : \mathcal{R}X \to \mathcal{R}Y$.*

The proof of Lemma 3.6.2 is elementary.

Using Proposition 3.6.1 and Lemma 3.6.2 we can define a functor $\mathcal{R} : \mathbf{CH} \to \mathbf{CH}$.

For a compact Hausdorff $X$ we can define maps $\eta_X : X \to \mathcal{R}X$ and $\mu_X : \mathcal{R}\mathcal{R}X \to \mathcal{R}X$ in the same way as we did for the Giry monad. These maps are well-defined and continuous and they induce natural transformations $\eta : 1_{\mathbf{CH}} \to \mathcal{R}$ and $\mu : \mathcal{R}\mathcal{R} \to \mathcal{R}$.

As has often been observed [34, 36, 56, 61]:

**Proposition 3.6.3.** *The triple $(\mathcal{R}, \eta, \mu)$ is a monad.*

As discussed before in Section 2.1.2, the monad in Proposition 3.6.3 is called the *Radon monad.*

Every finite set $A$ endowed with the whole powerset as topology, forms a compact Hausdorff space. Every map of finite sets is continuous with respect to these topologies. This gives a functor $k : \mathbf{Set}_f \to \mathbf{CH}$.

We will now consider the functor $R$ which is defined as

$$\mathbf{Set}_f \xrightarrow{k} \mathbf{CH} \xrightarrow{\mathcal{R}} \mathbf{CH}$$

This means that for a finite set $A$, the space $RA$ is the subspace of $[0,1]^A$ of all families $(p_a)_a$ such that $\sum_{a \in A} p_a = 1$. A map of finite sets $f : A \to B$ induces a continuous map $Rf : RA \to RB$ that sends an element $(p_a)_{a \in A}$ to

$$\left( \sum_{a \in f^{-1}(b)} p_a \right)_{b \in B}.$$

**Theorem 3.6.4.** *The Radon monad is the codensity monad of $R$.*

*Proof.* By Proposition 3.2.4 it follows that the codensity monad of $R$ exists. Let $X$ be a compact Hausdorff space. We will now show that

$$\mathcal{R}(X) = \lim(X \downarrow R \xrightarrow{U} \mathbf{Set}_f \xrightarrow{R} \mathbf{CH}).$$

Proposition 3.2.4 then implies that $T^R(X) \cong \mathcal{R}(X)$ for all compact Hausdorff spaces $X$.

For a continuous function $f : X \to RA$, define a map $p_f : \mathcal{R}X \to RA$ by

$$\mathcal{R}X \xrightarrow{\mathcal{R}f} \mathcal{R}\mathcal{R}kA \xrightarrow{\mu_{kA}} \mathcal{R}kA = RA.$$

That means that

$$p_f(\mathbb{P}) := \left(\int_X f_a \mathrm{d}\mathbb{P}\right)_{a \in A}.$$

In the same way as in the proof of Theorem 3.4.1 we can show that $(\mathcal{R}X, (p_f)_f)$ is a cone over the diagram $X \downarrow R \xrightarrow{U} \mathbf{Set}_f \xrightarrow{R} \mathbf{CH}$. Let $(Y, (q_f)_f)$ be a cone over the diagram. In the same way as in the proof of Theorem 3.4.1, we can define a map $I_y : \mathbf{CH}(X, [0,1]) \to [0,1]$ and show that this map sends 1 to 1 and preserves binary sums that exist in $\mathbf{CH}(X, [0,1])$. Now by Example 3.5.10, we know that there exists a unique Radon probability measure $\mathbb{P}_y$ on $X$ such that $I_y(f) = \int_X f \mathrm{d}\mathbb{P}_y$ for all continuous functions $f : X \to [0,1]$. The map $q : Y \to \mathcal{R}X$ defined by the assignment $y \mapsto \mathbb{P}_y$ is continuous and is the only morphism of cones from $(Y, (q_f)_f)$ to $(\mathcal{R}X, (p_f)_f)$. This shows that $\mathcal{R}(X) \cong T^R(X)$ for all compact Hausdorff spaces $X$. Moreover, this induces a natural isomorphism $\mathcal{R} \cong T^R$. It can be shown in the same way as in the proof of Theorem 3.4.1 that the unit and multiplication of the Radon monad are the same as those of the codensity monad of $R$. ☐

**Remark 3.6.5.** The domain of $R$ is the category of finite sets and maps. We would expect that this would only yield *finitely additive* measures. However Theorem 3.6.4 tells us that the codensity construction gives us Radon probability measures, which are *$\sigma$-additive*.

**Remark 3.6.6.** Instead of using the category of all finite sets as the domain of $R$ it is enough to use the category of sets $A$ with $|A| \leq 3$.

### 3.6.2 Bounded Lipschitz monad

In this section we will consider the metric version of Section 3.6.1. We introduce a new monad of Radon probability measures on the category of compact metric spaces with 1-Lipschitz maps. We show that this monad is the codensity monad of a functor $L$. This functor is similar to the functor $R$ from Section 3.6.1.

Let $X$ be a compact metric space and define $\mathcal{L}X$ to be the metric space of all Radon probability measures on $X$ together with the metric $d_{\mathcal{L}X}$ defined by

$$d_{\mathcal{L}X}(\mathbb{P}_1, \mathbb{P}_2) := \sup\left\{\left|\int_X f \mathrm{d}\mathbb{P}_1 - \int_X f \mathrm{d}\mathbb{P}_2\right| \mid f : X \to [0,1] \text{ is a 1-Lipschitz function}\right\}$$

for all Radon probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on $X$. This metric is called the **bounded Lipschitz metric**[3].

By Theorem 8.3.2 in [7] Using that the composite of 1-Lipschitz functions is a 1-Lipschitz function we can show that pushing forward along a 1-Lipschitz function $f$ defines a 1-Lipschitz function $\mathcal{L}f$.

Let $\mathbf{KMet}_1$ be the category of compact metric spaces and 1-Lipschitz maps. The above defines a functor $\mathcal{L} : \mathbf{KMet}_1 \to \mathbf{KMet}_1$.

For a compact metric space $(X, d)$ let $\eta_X$ and $\mu_X$ be the maps as defined for the Radon monad. For $x$ and $y$ in $X$ we find

$$d_{\mathcal{L}X}(\eta_X(x), \eta_X(y)) = \sup\{|f(x) - f(y)| \mid f : X \to [0,1] \text{is a 1-Lipschitz function}\} \leq d(x,y),$$

because $|f(x) - f(y)| \leq d(x,y)$ for every 1-Lipschitz function $f : X \to [0,1]$. Therefore $\eta_X$ is 1-Lipschitz with respect to the bounded Lipschitz metric.

---

[3] The Kantorovich distance is the metric obtained by taking the supremum over all 1-Lipschitz function on $X$ taking values in $\mathbb{R}$ instead of $[0,1]$. The obtained metric space is called the Kantorovich space of $X$.

For Radon probability measures $\mathbf{P}_1$ and $\mathbf{P}_2$ on $\mathcal{L}X$ we find for every 1-Lipschitz function $f : X \to [0,1]$ the following inequality:

$$\begin{aligned}\left|\int_{\mathcal{L}X} f \mathrm{d}\mu_X(\mathbf{P}_1) - \int_{\mathcal{L}X} f \mathrm{d}\mu_X(\mathbf{P}_2)\right| &= \left|\int_{\mathcal{L}X} \mathrm{ev}_f \mathrm{d}\mathbf{P}_1 - \int_{\mathcal{L}X} \mathrm{ev}_f \mathrm{d}\mathbf{P}_2\right| \\ &\leq d_{\mathcal{L}\mathcal{L}X}(\mathbf{P}_1, \mathbf{P}_2)\end{aligned}$$

In the last step we used the fact that the map $\mathrm{ev}_f : \mathcal{L}X \to [0,1]$, which sends a Radon probability measure $\mathbb{P}$ to $\int_X f \mathrm{d}\mathbb{P}$, is 1-Lipschitz. Since $f$ was arbitrary we conclude that $\mu_X$ is 1-Lipschitz.

It follows now that we have natural transformations $1_{\mathbf{KMet}_1} \to \mathcal{L}$ and $\mathcal{L}\mathcal{L} \to \mathcal{L}$ which we will also denote by $\eta$ and $\mu$.

Similar to the Giry monad and Proposition 3.6.3 we can prove the following result.

**Proposition 3.6.7.** *The triple* $(\mathcal{L}, \eta, \mu)$ *is a monad.*

We call the monad from Proposition 3.6.7 the **bounded Lipschitz monad**. This monad is similar to the Kantorovich monad, which is dicussed in [25, 27, 62].

Let $A$ be a finite set. Define a metric on $A$ by

$$d_A(a_1, a_2) := \begin{cases} 1 & \text{if } a_1 \neq a_2 \\ 0 & \text{if } a_1 = a_2 \end{cases}$$

for all $a_1$ and $a_2$ in $A$. This makes $A$ a compact metric space. Every function of finite sets becomes 1-Lipschitz with respect to these metrics. This gives a functor $l : \mathbf{Set}_f \to \mathbf{KMet}_1$.

We will now consider the functor $L$ which we define as

$$\mathbf{Set}_f \xrightarrow{l} \mathbf{KMet}_1 \xrightarrow{L} \mathbf{KMet}_1.$$

Let $A$ be a finite set. The space $LA$ is the subset of $[0,1]^A$ of families $(p_a)_a$ such that $\sum_{a \in A} p_a = 1$ together with the bounded Lipschitz metric $d_{LA}$. The following proposition tells us that $d_{LA}$ is in fact the *total variation distance.*

**Proposition 3.6.8.** *Let $A$ be a finite set. For $p$ and $q$ in $LA$,*

$$d_{LA}(p, q) = \sup\left\{\left|\sum_{a \in A'} p_a - \sum_{a \in A'} q_a\right| \mid A' \subseteq A\right\}.$$

*Proof.* We view $A$ as a metric space by endowing it with the metric $d_A$. For a subset $A'$ of $A$, the function $1_{A'} : A \to [0,1]$ is a 1-Lipschitz function. We find

$$\left|\sum_{a \in A'} p_a - \sum_{a \in A'} q_a\right| = \left|\int_A 1_{A'} \mathrm{d}p - \int_A 1_{A'} \mathrm{d}q\right| \leq d_{LA}(p, q).$$

This shows one inequality.

Define the sets $A^+ = \{a \in A \mid p_a \geq q_a\}$ and $A^- = \{a \in A \mid p_a < q_a\}$. For a 1-Lipschitz function $f : A \to [0,1]$, we find

$$\begin{aligned}\int_A f \mathrm{d}p - \int_A f \mathrm{d}q &= \sum_{a \in A} f(a)(p_a - q_a) \\ &\leq \sum_{a \in A^+} f(a)(p_a - q_a) \\ &\leq \sum_{a \in A^+} p_a - q_a \\ &= \left| \sum_{a \in A^+} p_a - \sum_{a \in A^+} q_a \right|\end{aligned}$$

Similarly we find

$$-\left( \int_A f \mathrm{d}p - \int_A f \mathrm{d}q \right) \leq \left| \sum_{a \in A^-} p_a - \sum_{a \in A^-} q_a \right| .$$

We can conclude that

$$\left| \int_A f \mathrm{d}p - \int_A f \mathrm{d}q \right| \leq \sup \left\{ \left| \sum_{a \in A'} p_a - \sum_{a \in A'} q_a \right| \mid A' \subseteq A \right\},$$

which proves the other inequality. □

This gives us the following useful corollary.

**Corollary 3.6.9.** *Let $X$ be a compact metric space and let $A$ be a finite set. A map $f : X \to LA$ is 1-Lipschitz if and only if*

$$x \mapsto \sum_{a \in A'} f(x)_a$$

*defines a 1-Lipschitz map $X \to [0,1]$ for every $A' \subseteq A$.*

This leads to the main result of this section.

**Theorem 3.6.10.** *The bounded Lipschitz monad is the codensity monad of $L$.*

*Proof.* We will again use Proposition 3.2.4. Let $(X, d)$ be a compact metric space and let $D_X$ denote the diagram

$$X \downarrow L \xrightarrow{U} \mathbf{Set}_f \xrightarrow{L} \mathbf{KMet}_1.$$

For a 1-Lipschitz function $f : X \to LA$ define the map $p_f : \mathcal{L}X \to LA$ by

$$\mathcal{L}X \xrightarrow{\mathcal{L}f} \mathcal{L}\mathcal{L}lA \xrightarrow{\mu_{lA}} \mathcal{L}lA = LA$$

This means that for every $\mathbb{P} \in \mathcal{L}X$,

$$p_f(\mathbb{P}) = \left( \int_X f_a \mathrm{d}\mathbb{P} \right)_{a \in A} .$$

In the same way as in the proof of Theorem 3.4.1, it can be shown that $(\mathcal{L}X, (p_f)_f)$ forms a cone over the diagram $D_X$.

We will now show that this is the limiting cone over the diagram. Let $(Y, (q_f)_f)$ be a cone over the diagram $D_X$. For a 1-Lipschitz function $f : X \to [0,1]$ let $\widehat{f} : X \to L\mathbf{2}$ be the function defined by $\widehat{f}(x) := (1 - f(x), f(x))$. Note that this function is 1-Lipschitz.

For an element $y \in Y$ define the map $I_y : \mathbf{KMet}_1(X, [0,1]) \to [0,1]$ by

$$I_y(f) := q_{\widehat{f}}(y)_1.$$

For $n \geq 1$, let $\mathbf{n} := \{0, \dots, n-1\}$. Let $t : \mathbf{1} \to \mathbf{2}$ be the map that sends 0 to 1 and let $e$ be the unique 1-Lipschitz map $X \to L\mathbf{1}$. We have that $Lt \circ e = \widehat{1}$, i.e. $t$ is a morphism $\widehat{1} \to e$ in the arrow category $X \downarrow L$. Since $(Y, (q_f)_f)$ is a cone over the diagram $D_X$ it follows that $Lt \circ q_e = q_{\widehat{1}}$. This implies that $I_y(1) = 1$.

Consider 1-Lipschitz functions $f_1, f_2 : X \to [0,1]$ such that also $f := f_1 + f_2$ is a 1-Lipschitz function that takes values in $[0,1]$. Define the map $h : X \to L\mathbf{2}$ that sends an element $x$ to $(1 - f(x), f_1(x), f_2(x))$. Because $1, f_1, f_2, 1 - f_2, 1 - f_1, 1 - f$ and $f$ are all 1-Lipschitz, so is $h$ by Corollary 3.6.9. Let $s : \mathbf{3} \to \mathbf{2}$ be the map that sends 0 to 0 and the other elements to 1. For an element $k \in \mathbf{3}$, let $s_k : \mathbf{3} \to \mathbf{2}$ be the map that sends $k$ to 1 and the other elements to 0.

We have that

$$Ls \circ h = \widehat{f}$$

and for every $k \in \mathbf{3}$ that

$$Ls_k \circ h = \widehat{f_k}.$$

Because $(Y, (q_f)_f)$ is a cone over the diagram $D_X$, these equalities imply that

$$Ls \circ q_h = q_{\widehat{f}}$$

and for every $k \in \mathbf{3}$ that

$$Ls_k \circ q_h = q_{\widehat{f_k}}.$$

Using this we obtain the following equalities:

$$I_y(f) = (Ls \circ q_h(y))_1 = q_h(y)_1 + q_h(y)_2 = Ls_1 \circ q_h(y))_1 + Ls_2 \circ q_h(y))_2 + = I_y(f_1) + I_y(f_2).$$

By Example 3.5.11 there exists a unique Radon probability measure $\mathbb{P}_y$ such that $I_y(f) = I_{\mathbb{P}_y}(f)$. The assignment $y \mapsto \mathbb{P}_y$ defines a map $q : Y \to \mathcal{L}X$. For $y_1$ and $y_2$ in $Y$ and for a 1-Lipschitz function $f : X \to [0,1]$, we find that

$$\left| \int_X f \mathrm{d}\mathbb{P}_{y_1} - \int_X f \mathrm{d}\mathbb{P}_{y_2} \right| = |I_{y_1}(f) - I_{y_2}(f)| = |q_{\widehat{f}}(y_1)_1 - q_{\widehat{f}}(y_2)_1| \leq d_Y(y_1, y_2).$$

It follows now that $q$ is 1-Lipschitz.

Let $f : X \to LA$ be a 1-Lipschitz map and let $a$ be an element of $A$. Let $s_a : A \to \mathbf{2}$ be the map that sends $a$ to 1 and every other element to 0. Since we have that $Ls_a \circ f = \widehat{f_a}$ we also have that $Ls_a \circ q_f = q_{\widehat{f_a}}$. In particular we find for every $y \in Y$ that

$$I_y(f_a) = q_{\widehat{f_a}}(y)_1 = (Ls_a \circ q_f(y))_1 = q_f(y)_a.$$

Using this we obtain for every $y \in Y$ and for every $a \in A$ that

$$[p_f \circ q(y)]_a = I_{\mathbb{P}_y}(f_a) = I_y(f)_a = q_f(y)_a.$$

This shows that $q$ is a morphism of cones from $(\mathcal{L}X, (p_f)_f)$ to $(Y, (q_f)_f)$.

Let $\tilde{q} : Y \to \mathcal{L}X$ be another morphism of cones. Then for every 1-Lipschitz function $f : X \to [0,1]$ we have that

$$\int_X f \mathrm{d}\tilde{q}(y) = (p_{\widehat{f}} \circ \tilde{q}(y))_1 = q_{\widehat{f}}(y)_1 = I_y(f).$$

This show that $\tilde{q}(y) = q(y)$ for all $y \in Y$ and therefore $(\mathcal{L}X, (p_f)_f)$ is the limiting cone over the diagram $D_X$. This implies that $\mathcal{L}(X) \cong T^L(X)$ for all compact metric spaces. This induces a natural isomorphism $\mathcal{L} \cong T^L$.

It can be checked that the unit and multiplication of the codensity monad of $L$ are equal to the unit and multiplication of the bounded Lipschitz monad. □

**Remark 3.6.11.** Note that the bounded Lipschitz monad can be extended to a monad on the category **KMet** of compact metric spaces and all Lipschitz maps. The compact metric space $\mathcal{L}X$ is isomorphic to the Kantorovich space in this category. This monad can also be obtained as a codensity monad in a similar way. The proof of this is a simple adaptation of the proof Theorem 3.6.10.

**Remark 3.6.12.** The Kantorovich monad on compact metric spaces introduced by van Breugel in [62] was extended to a monad on complete metric spaces by Fritz and Perrone [25, 27]. The underlying endofunctor of this monad sends a complete metric space $(X, d)$ to the Kantorovich space of $(X, d)$. This is the space of all Radon probability measure of finite moment together with the Kantorovich distance.

We would like to use the integral representation theorem from Example 3.5.8 to construct this monad as a codensity monad, but we have been unable to find a way to do so.

### 3.6.3 Baire monad

In this section we will introduce a new monad of probability measures, which we will call the *Baire monad* and we explain how this monad arises as a codensity monad. We write **Haus** to mean the category of Hausdorff spaces and continuous maps.

For a Hausdorff space $X$ let $\mathcal{B}X$ be the space of Baire probability measures on $X$ with the smallest topology such that $\mathrm{ev}_f : \mathcal{B}X \to [0,1]$ is continuous for every continuous function $f : X \to [0,1]$. Much as in the previous sections this induces a functor $\mathcal{B} : \mathbf{Haus} \to \mathbf{Haus}$ where **Haus** is the category of Hausdorff spaces and continuous maps. We can define a monad structure on this endofunctor and we call this monad the **Baire monad**.

For a countable set $A$ let $BA$ be the subspace of $[0,1]^A$ of families $(p_a)_{a\in A}$ such that $\sum_{a\in A} p_a = 1$. Every finite map[4] $f : A \to C$ of countable sets induces a continuous map $Bf : BA \to BC$. Let $\mathbf{Set}_c^f$ be the category of countable sets and finite maps. We obtain a functor $B : \mathbf{Set}_c^f \to \mathbf{Haus}$.

**Theorem 3.6.13.** *The Baire monad is the codensity monad of $B$.*

The proof of this result is similar to the proof of Theorem 3.4.6 and Theorem 3.4.1. Therefore we will only give a short sketch of the proof. We will again use Proposition 3.2.4 and show that $\mathcal{B}X$ is the limit of the diagram

$$X \downarrow B \to \mathbf{Set}_c^f \xrightarrow{B} \mathbf{Haus}.$$

Every continuous map $f : X \to BA$ induces a map $p_f : \mathcal{B}X \to BA$ by sending $\mathbb{P}$ to $\int_X f \mathrm{d}\mathbb{P}$. Because we use finite maps and the finite sum of continuous functions is continuous, these maps form a cone over the diagram.

---

[4] Finite maps were introduced in Section 3.4.3

To show that this cone is universal we can use a similar argument as in the proof of Theorem 3.4.1, since all the maps introduced here were finite maps. Only now we use the representation theorem from Example 3.5.12 instead of Proposition 3.3.6.

**Remark 3.6.14.** The results in this section would also work for arbitrary topological spaces. Because it is not common in probability theory to study probability measures on non-Hausdorff spaces we chose to restrict everything to Hausdorff spaces.

**Remark 3.6.15.** The construction of the Baire monad as a codensity monad is slightly different from the other probability monads. Here the functor is not of the form

$$\mathbf{Set}_c^f \xrightarrow{i} \mathbf{Haus} \xrightarrow{\mathcal{B}} \mathbf{Haus}.$$

Therefore it does not completely follow the pattern of the constructions in the previous monads. We tried to construct this monad as the codensity monad of a functor of the of form $\mathcal{B}i$ for some (inclusion) functor $i : \mathbf{Set}_c^f \to \mathbf{Haus}$, but we have not been able to do this so far.

# Chapter 4

# A categorical proof of the Carathéodory extension theorem

## 4.1 Introduction

The Carathéodory extension theorem is an important result in measure theory. It guarantees the existence of the Lebesgue measure (or more generally the Lebesgue–Stieltjes measure) and of product measures. But it also is a key result in the proof of the Kolmogorov extension theorem, which guarantees the existence of Brownian motion and which is closely related to martingale convergence results.

If we want to define a measure on a measurable space $(X, \Sigma)$, we need to assign to every subset $A \in \Sigma$ an element of $[0, \infty]$. However, we often work with $\sigma$-algebras of the form $\sigma(\mathcal{B})$, where $\mathcal{B}$ is an algebra of subsets of $X$, i.e. closed under finite unions and complements. In this case it can be very difficult to know what a general measurable subset looks like and to assign real numbers to them in a $\sigma$-additive way. This problem is solved by the Carathéodory extension theorem. It states that a $\sigma$-additive map $\mathcal{B} \to [0, \infty]$ can be extended to a $\sigma$-additive map $\sigma(\mathcal{B}) \to [0, \infty]$. Here, $\sigma$-additive should be interpreted as '$\sigma$-additive whenever unions of countable pairwise disjoint collections exist'. Moreover, this extension is very often, but not always, unique.

$$\begin{tikzcd}
\mathcal{B} \arrow[r] \arrow[d] & {[0,\infty]} \\
\sigma(\mathcal{B}) \arrow[ur, dashed] &
\end{tikzcd}$$

The classical proof for this result is relatively long and technical (see for example Theorem 1.41 in [40]). It consists of different steps of extending and restricting back and requires several smart 'tricks' and constructions. In this paper we will give a categorical proof for the Carathéodory extension theorem using results on extensions of *lax, colax* and *strict transformations* between functors. Several parts of this proof look similar to steps in the classical proof. However, in our proof all constructions follow from the Kan extension formulas. Moreover, viewing the Carathéodory extension theorem in this categorical framework, allows us to compare it to extension results in other areas of mathematics. Furthermore, this technique allows us to easily generalize Carathéodory's result to measures taking values in other posets or spaces.

To do this we start by studying categories of certain transformations between functors and extensions of transformations along transformations. In Section 4.2 and Section 4.3 we do this on a fairly abstract level. We discuss *(co)laxification (co)monads* and are in particular interested in the case that their (co)algebras are strict transformations. We give several abstract conditions for extensions of certain transformations to exist.

In Section 4.4 and Section 4.5, we give concrete constructions for the operations and extensions discussed in Section 4.2 and Section 4.3. In this part *(co)lax coends* and *(co)lax morphism classifiers* play an important role.

In the last two sections of this chapter we apply the previous sections to measure theory. We start by representing *inner* and *outer premeasures* by certain transformations between certain functors. By applying the extension results to these functors, we obtain our categorical proof for Carathédory's extension result. We furthermore make a distinction between *left* and *right* Carathédory extensions. The right one always exists and is the same as the extension in the classical proof, the left one however does not always exist. This is also related to the fact that inner measures and outer measures can behave surprisingly different from each other. Moreover, we immediately obtain a way to characterize these extensions by a universal property, namely as a maximal or minimal extension. This is interesting when uniqueness is not guaranteed.

| | Posets of transformations | Extensions of transformations |
|---|---|---|
| Abstract theory | 4.2 | 4.3 |
| Concrete constructions | 4.4 | 4.5 |
| Applications to measures | 4.6 | 4.7 |

## 4.2 Posets of (co)lax transformations

In this section we will introduce and discuss posets of $\Sigma$-*natural (co)lax transformations* and strict transformations. In particular we will be interested in the embeddings between these and when these are (co)reflective. Furthermore, we will look at the (co)monads these adjunctions induce and study their (co)algebras.

This section focuses on abstract existence results of (co)reflection operations; in Section 4.4 we will give concrete constructions of these operations. In Section 4.6, we will represent inner (outer) premeasures as $\Sigma$-natural (co)lax transformations and strict transformations. The (co)reflection operations described in this section will then correspond to operations that turn inner (outer) premeasures into premeasures.

Let $\mathcal{C}$ be a small poset-enriched category. Let **Pos** be the category of posets and order-preserving maps, viewed as enriched over itself. Let $F$ and $G$ be enriched functors $\mathcal{C} \to \mathbf{Pos}$. Let $\Sigma$ be a collection of morphisms in $\mathcal{C}$.

We will study several generalizations of natural transformations. In the first variation we *only* have naturality squares for morphisms in the fixed collection $\Sigma$.

**Definition 4.2.1.** A **$\Sigma$-natural (general) transformation** $\tau : F \to G$ is a collection of order-preserving maps $(\tau_A : FA \to GA)_{A \in \mathrm{ob}\mathcal{C}}$ such that

$$\begin{tikzcd}
FA \arrow[r, "\tau_A"] \arrow[d, "Ff"'] & GA \arrow[d, "Gf"] \\
FB \arrow[r, "\tau_B"'] & GB
\end{tikzcd}$$

commutes for all $f \in \Sigma$.

In the other variations we will also allow *weaker* naturality squares.

**Definition 4.2.2.** A **$\Sigma$-natural lax transformation** $\tau : F \to G$ is a collection of order-preserving maps $(\tau_A : FA \to GA)_{A \in \mathrm{ob}\mathcal{C}}$ such that

$$\begin{tikzcd}
FA \arrow[r, "\tau_A"] \arrow[d, "Ff"'] \arrow[dr, phantom, "\leq"] & GA \arrow[d, "Gf"] \\
FB \arrow[r, "\tau_B"'] & GB
\end{tikzcd}$$

for all morphisms $f : A \to B$ in $\mathcal{C}$ and such that this is an equality whenever $f \in \Sigma$.

Dually, we can define **$\Sigma$-natural colax transformations**, by reversing the inequality sign in the above definition.

**Definition 4.2.3.** A **strict transformation** $\tau : F \to G$ is a collection of order-preserving maps $(\tau_A : FA \to GA)_{A \in \mathrm{ob}\mathcal{C}}$ such that

$$\begin{tikzcd}
FA \arrow[r, "\tau_A"] \arrow[d, "Ff"'] & GA \arrow[d, "Gf"] \\
FB \arrow[r, "\tau_B"'] & GB
\end{tikzcd}$$

commutes for *all* morphisms $f : A \to B$ in $\mathcal{C}$.

Note that a strict transformation is the same as a general $\mathrm{Mor}(\mathcal{C})$-natural transformation. In the case that $\Sigma = \emptyset$, we will omit '$\Sigma$' in the terminology and notation.

The set of $\Sigma$-natural transformations is partially ordered. The order is defined by

$$\tau^1 \leq \tau^2 :\Leftrightarrow \tau^1_A(x) \leq \tau^2_A(x)$$

for all $A \in \mathrm{ob}\,\mathcal{C}$ and $x \in FA$ and for $\Sigma$-natural transformations $\tau^1$ and $\tau^2$.

The poset of $\Sigma$-natural transformations is denoted by $[F, G]^{\Sigma}$; the subposets of $\Sigma$-natural lax transformations, $\Sigma$-natural colax transformations and strict transformations are denoted by $[F, G]^{\Sigma}_l, [F, G]^{\Sigma}_c$ and $[F, G]_s$ respectively. We clearly have the following pullback square of inclusions.

$$\begin{tikzcd}
& {[F,G]_s} \arrow[dl] \arrow[dr] \arrow[dd, phantom, "\vee", near start] & \\
{[F,G]^{\Sigma}_c} \arrow[dr, "U_c"'] & & {[F,G]^{\Sigma}_l} \arrow[dl, "U_l"] \\
& {[F,G]^{\Sigma}} &
\end{tikzcd}$$

We will now give conditions on the functors $F$ and $G$ such that the inclusions in the above diagram are (co)reflective and such that these posets of transformations are complete.

**Proposition 4.2.4.** *Suppose $GA$ is cocomplete*[1] *for all $A \in \mathrm{ob}\mathcal{C}$ and suppose $Gf$ preserves all joins for all $f \in \Sigma$. Then $[F, G]^{\Sigma}_l$ and $[F, G]^{\Sigma}$ are cocomplete and the inclusion $U_l : [F, G]^{\Sigma}_l \to [F, G]^{\Sigma}$ preserves all joins.*

*Proof.* Let $(\lambda^i)_{i \in I}$ be a collection of $\Sigma$-natural transformations. Define for all $A \in \mathrm{ob}\,\mathcal{C}$ and $x \in FA$,

$$\lambda_A(x) := \bigvee_{i \in I} \lambda^i_A(x).$$

---

[1] Since $GA$ is a poset for every object $A$ in $\mathcal{C}$, $GA$ is also complete.

For $f : A \to B \in \Sigma$ and $x \in FA$,

$$Gf(\lambda_A(x)) = Gf\left(\bigvee_{i\in I} \lambda^i_A(x)\right) = \bigvee_{i\in I} Gf(\lambda^i_A(x)) = \bigvee_{i\in I} \lambda^i_B(Ff(x)) = \lambda_B(Ff(x)).$$

Therefore $\lambda$ is a $\Sigma$-natural transformation and it straightforward to verify that this is the join of $(\lambda^i)_{i\in I}$ in $[F, G]$. If $\lambda^i$ is lax for every $i \in I$, then for a map $f : A \to B$ in $\mathcal{C}$ and $x \in FA$,

$$Gf(\lambda_A(x)) = Gf\left(\bigvee_{i\in I} \lambda^i_A(x)\right) \leq \bigvee_{i\in I} Gf(\lambda^i_A(x)) \leq \bigvee_{i\in I} \lambda^i_B(Ff(x)) = \lambda_B(Ff(x)).$$

Here we used that $Gf$ is order-preserving and that $\lambda^i$ is lax for all $i \in I$. It follows that $\lambda$ is a $\Sigma$-natural lax transformation. This is the join of $(\lambda^i)_{i\in I}$ and clearly $U_l$ preserves this join. $\square$

Using Proposition 4.2.4 and the Adjoint Functor theorem for posets, we immediately obtain the following corollary.

**Corollary 4.2.5.** *If $GA$ is cocomplete for all $A \in \operatorname{ob}\mathcal{C}$ and if $Gf$ preserves joins for all $f \in \Sigma$, then $U_l : [F, G]^\Sigma_l \to [F, G]^\Sigma$ has a right adjoint $R_l : [F, G]^\Sigma \to [F, G]^\Sigma_l$.*

Since $U_l$ is full and faithful, the unit is an equality, i.e. $\lambda = R_l U_l \lambda$ for a $\Sigma$-natural lax transformation $\lambda$. Clearly, we also have dual results for Proposition 4.2.4 and Corollary 4.2.5. In particular, if $GA$ is complete for all $A \in \operatorname{ob}\mathcal{C}$ and $Gf$ preserves meets for all $f \in \Sigma$, then $U_c : [F, G]^\Sigma_c \to [F, G]^\Sigma$ has a left adjoint $L_c : [F, G]^\Sigma \to [F, G]^\Sigma_c$ and $\sigma = L_c U_c \sigma$ for a $\Sigma$-natural colax transformation $\sigma$.

Assume from now on that $GA$ is (co)complete and that $Gf$ preserves all joins and meets for $f \in \Sigma$. Using these adjunctions we obtain operations that turn lax transformations into colax transformations and vice versa. For a $\Sigma$-natural lax transformation $\lambda$, we denote

$$\overline{\lambda} := L_c U_l(\lambda),$$

and we call $\overline{\lambda}$ the **colaxification of**$\lambda$. For a $\Sigma$-natural colax transformation $\sigma$, we write

$$\underline{\sigma} := R_l U_c(\sigma),$$

and we call $\underline{\sigma}$ the **laxification of** $\sigma$.

In what follows we will often omit the forgetful functors. Using this convention we find that for a strict transformation $\tau$,

$$\underline{\tau} = \tau = \overline{\tau}.$$

The map $\overline{(-)} : [F, G]^\Sigma_l \to [F, G]^\Sigma_c$ is left adjoint to $\underline{(-)} : [F, G]^\Sigma_c \to [F, G]^\Sigma_l$. The following diagram summarizes this:

$$\begin{tikzcd}
{[F,G]^\Sigma_l} \arrow[r, shift left, "U_l"] \arrow[r, phantom, "\perp"] \arrow[rr, bend left, "\overline{(-)}"] & {[F,G]^\Sigma} \arrow[l, shift left, "R_l"] \arrow[r, shift left, "L_c"] \arrow[r, phantom, "\perp"] & {[F,G]^\Sigma_c} \arrow[l, shift left, "U_c"] \arrow[ll, bend left, "\underline{(-)}"]
\end{tikzcd}$$

The monad or closure operator on $[F, G]^\Sigma_l$ induced by this adjunction is denoted by $T$ and is called the **colaxification monad**. The induced comonad or interior operator on $[F, G]^\Sigma_c$ is denoted by $S$ and is called the **laxification monad**.

The following proposition discusses the situation where (co)laxifications of $\Sigma$-natural co(lax) transformations are strict.

**Proposition 4.2.6.** *The following are equivalent:*

1. $\overline{\lambda}$ *is strict for all* $\lambda \in [F,G]_l^\Sigma$,
2. $([F,G]_l^\Sigma)^T = [F,G]_s$,
3. $\underline{\sigma}$ *is strict for all* $\sigma \in [F,G]_c^\Sigma$
4. $([F,G]_c^\Sigma)^S = [F,G]_s$

*Proof.* $\underline{1 \Rightarrow 2}$: For a strict transformation $\tau$, we have that $T\tau = \underline{(\overline{\tau})} = \underline{\tau} = \tau$, making it a $T$-algebra (or $T$-closed element). For a $T$-algebra (or $T$-closed element) $\lambda$, we have that $T\lambda = \lambda$. By the hypothesis, we also have that the laxifcation of $\overline{\lambda}$ is again $\overline{\lambda}$, i.e. $T\lambda = \overline{\lambda}$. Combining this gives that $\lambda = \overline{\lambda}$, showing that $\lambda$ is strict.

$\underline{2 \Rightarrow 3}$: Because $\underline{\sigma} = T\underline{\sigma}$, it follows that $\underline{\sigma}$ is a $T$-algebra and therefore strict, by the hypothesis.

$\underline{3 \Rightarrow 4}$: This is similar to the proof of the implication $1 \Rightarrow 2$. Every strict transformation $\tau$ is a coalgebra since $S\tau = \tau$. For an $S$-coalgebra (or $S$-open element) $\sigma$, we have $\underline{\sigma} = S\sigma = \sigma$. Therefore $\sigma$ is lax and thus strict.

$\underline{4 \Rightarrow 1}$: Similarly to the proof of the implication $2 \Rightarrow 3$, it follows from the fact that $\overline{\lambda} = S\overline{\lambda}$. ☐

We say that the triple $(F, G, \Sigma)$ satisfies the **strictness condition** if it satisfies the conditions in Proposition 4.2.6. In this case the inclusion $[F,G]_s \to [F,G]_l^\Sigma$ has a left adjoint $L : [F,G]_l^\Sigma \to [F,G]_s$ and the inclusion $[F,G]_s \to [F,G]_c^\Sigma$ has a right adjoint $R : [F,G]_c^\Sigma \to [F,G]_s$. These operations turn $\Sigma$-natural (co)lax transformations into strict transformations in a universal way.

If the strictness condition is satisfied, then the poset of strict transformations is also complete and we can describe what joins and meets look like. This is explained in the following corollary.

**Corollary 4.2.7.** *Suppose that* $GA$ *is complete for all* $A \in \operatorname{ob}\mathcal{C}$ *and that* $Gf$ *preserves joins for all* $f \in \Sigma$. *Suppose that* $(F, G, \Sigma)$ *satisfies the strictness condition. Let* $(\tau^i)_{i \in I}$ *be a collection in* $[F,G]_s^\Sigma$. *Then their join exists and is given by*

$$\bigvee_{i \in I} \tau^i = \overline{\lambda},$$

*where* $\lambda_A(x) := \bigvee_{i \in I} \tau_A^i(x)$ *for all* $A \in \operatorname{ob}\mathcal{C}$ *and* $x \in FA$.

In the case that $(F, G, \Sigma)$ satisfies the strictness condition, we can summarize this section by the following diagram.

$$\begin{tikzcd}
 & {[F,G]_s} \arrow[dl, shift left=2] \arrow[dr, shift left=2] & \\
{[F,G]_c^\Sigma} \arrow[ur, shift left=2, "R"] \arrow[ur, phantom, "\top"] \arrow[dr, shift left=2, "U_c"] & \vee & {[F,G]_l^\Sigma} \arrow[ul, shift left=2, "L"] \arrow[ul, phantom, "\top"] \arrow[dl, shift left=2, "U_l"] \\
 & {[F,G]^\Sigma} \arrow[ul, shift left=2, "L_c"] \arrow[ul, phantom, "\top"] \arrow[ur, shift left=2, "R_l"] \arrow[ur, phantom, "\top"] &
\end{tikzcd}$$

## 4.3 Extensions of transformations

In this section, we will focus on extensions of (co)lax and strict $\Sigma$-natural transformations. We will give results on their existence and their properties. In Section 4.5, we will give concrete constructions of these extensions in particular cases. We will apply this to extend premeasures to measures in Section 4.7, using the representation of inner and outer premeasures as $\Sigma$-natural transformations from Section 4.6.

Let $\mathcal{C}$ be a small poset-enriched category and fix a collection $\Sigma$ of morphisms in $\mathcal{C}$. Let $F, G, H$ be enriched functors $\mathcal{C} \to \mathbf{Pos}$ and assume that $GA$ is complete for all $A \in \operatorname{ob}\mathcal{C}$ and that $Gf$ preserves joins and meets for all $f \in \Sigma$. Furthermore, let $\iota : F \to H$ be a strict transformation.

We will recall the definition of extensions of lax transformations from Section 1.2.3. Let $\lambda : F \to G$ be a $\Sigma$-natural general transformation (resp. lax, colax, strict). The **left extension of $\lambda$ along $\iota$** is a $\Sigma$-natural general transformation (resp. lax, colax, strict) $\mathrm{Lan}_\iota\lambda : H \to G$ such that $\mathrm{Lan}_\iota\lambda \circ \iota \geq \lambda$ and such that for every other $\lambda' : H \to G$ with $\lambda' \circ \iota \geq \lambda$, we have $\lambda' \geq \mathrm{Lan}_\iota\lambda$. We will sometimes use the notations $\mathrm{Lan}_\iota^{\mathrm{gen}}, \mathrm{Lan}_\iota^{\mathrm{lax}}, \mathrm{Lan}_\iota^{\mathrm{colax}}$ and $\mathrm{Lan}_\iota^{\mathrm{strict}}$ to emphasize that we are working with left extensions along $\iota$ of general, lax, colax and strict $\Sigma$-natural transformations respectively. [2]

$$\begin{tikzcd}
F \arrow[rr, "\lambda" description] \arrow[dd, "\iota" description] & & G \\
& {} & \\
H \arrow[uurr, "\leq / \mathrm{Lan}_\iota\lambda" description] \arrow[uurr, bend right=50, "\lambda'" description] & & 
\end{tikzcd}$$

We say that the extension is **proper** if $\mathrm{Lan}_\iota\lambda \circ \iota = \lambda$ and we call the extension **objectwise** if for all $A \in \operatorname{ob}\mathcal{C}$,

$$(\mathrm{Lan}_\iota\lambda)_A = \mathrm{Lan}_{\iota_A}\lambda_A.$$

Dually, we can define the **right extension of $\lambda$ along $\iota$.**

If an extension is objectwise, then we can reduce everything to Kan extensions of order-preserving maps. This gives us for example the following lemma.

**Lemma 4.3.1.** *If the left extension of $\lambda$ along $\iota$ is objectwise and if $\iota_A$ is full and faithful for every $A \in \operatorname{ob}\mathcal{C}$, then the extension is proper.*

*Proof.* Because $\iota_A$ is full and faithful for all $A \in \operatorname{ob}\mathcal{C}$, we know by Corollary 6.3.9 in [52] that $(\mathrm{Lan}_{\iota_A}\lambda_A) \circ \iota_A = \lambda_A$ for all $A \in \operatorname{ob}\mathcal{C}$. Because the extension of $\lambda$ along $\iota$ is objectwise, we conclude that

$$(\mathrm{Lan}_\iota\lambda) \circ \iota = \lambda.$$

Therefore the extension is proper. $\square$

Just as for Kan extensions of functors, extending transformations is adjoint to restricting transformations. The proof is essentially the same as the result for Kan extensions of functors (see for example Proposition 6.1.5 in [52]).

**Lemma 4.3.2.** *If the restriction map $- \circ \iota : [H, G]_\bullet \to [F, G]_\bullet$ has a right (resp. left) adjoint, then the right (resp. left) extension of $\lambda$ along $\iota$ exists for all $\lambda \in [F, G]_\bullet$. Moreover, the right (resp. left) adjoint is given by $\mathrm{Ran}_\iota-$ (resp. $\mathrm{Lan}_\iota-$).*

---

[2] The universal property determines the transformation, therefore we can talk about *the* left extension.

In the rest of this section we will look at conditions for extensions to exist and for them to be *proper* or *objectwise*. Extensions of $\Sigma$-natural general transformations always exist and are the best behaved.

**Proposition 4.3.3.** *The right and left extension of $\tau$ along $\iota$ exists for every $\Sigma$-natural general transformation $\tau \in [F,G]^{\Sigma}$.*

*Proof.* It is clear by the proof of Proposition 4.2.4 and its dual, that $[F,G]^{\Sigma}$ and $[H,G]^{\Sigma}$ are complete and that $-\circ\iota$ preserves all joins and meets. It follows now that the restriction map $-\circ\iota$ has a left and a right adjoint. The claim now follows from Lemma 4.3.2. ☐

For $\Sigma$-natural lax transformations things become more difficult. However, we still have that right extensions exist.

**Proposition 4.3.4.** *The right extension of $\lambda$ along $\iota$ exists for all $\lambda \in [F,G]^{\Sigma}_{l}$. Moreover,*

$$R_l(\mathrm{Ran}_\iota^{gen}\lambda) = \mathrm{Ran}_\iota^{lax}$$

*Proof.* Because $[F,G]^{\Sigma}_{l}$ is cocomplete by Proposition 4.2.4 and because $-\circ\iota$ preserves all joins, the restriction map has a right adjoint by the Adjoint Functor Theorem for complete posets.

The left adjoints in the following diagram commute

$$\begin{tikzcd}
{[F_{\sigma(\mathcal{B})}, G]^{\Sigma}_{l}} \arrow[r, "-\circ\iota"] \arrow[d, "U_l"'] & {[F_{\mathcal{B}}, G]^{\Sigma}_{l}} \arrow[d, "U_l"] \\
{[F_{\sigma(\mathcal{B})}, G]^{\Sigma}} \arrow[r, "-\circ\iota"'] & {[F_{\mathcal{B}}, G]^{\Sigma}}
\end{tikzcd}$$

Therefore their right adjoints also commute, i.e.

$$\begin{tikzcd}
{[F_{\sigma(\mathcal{B})}, G]^{\Sigma}_{l}} & {[F_{\mathcal{B}}, G]^{\Sigma}_{l}} \arrow[l, "\mathrm{Ran}_\iota^{\mathrm{lax}}"'] \\
{[F_{\sigma(\mathcal{B})}, G]^{\Sigma}} \arrow[u, "R_l"] & {[F_{\mathcal{B}}, G]^{\Sigma}} \arrow[l, "\mathrm{Ran}_\iota^{\mathrm{gen}}"] \arrow[u, "R_l"']
\end{tikzcd}$$

Therefore we have,

$$\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda = R_l(\mathrm{Ran}_\iota^{\mathrm{gen}}\lambda).$$

☐

**Remark 4.3.5.** We have the following inequality

$$\begin{tikzcd}
{[H, G]^{\Sigma}_{l}} \arrow[d, "U_l"'] \arrow[dr, phantom, "\leq"] & {[F, G]^{\Sigma}_{l}} \arrow[l, "\mathrm{Ran}_\iota -"'] \arrow[d, "U_l"] \\
{[H, G]^{\Sigma}} & {[F, G]^{\Sigma}} \arrow[l, "\mathrm{Ran}_\iota -"']
\end{tikzcd}$$

If this is an equality and right extensions of $\Sigma$-general transformations are objectwise and proper, then so are right extensions of $\Sigma$-lax transformations.

We have the following useful property about the existence of extensions of $\Sigma$-natural lax transformations and when they inherit properties from extensions of $\Sigma$-natural general transformations.

**Proposition 4.3.6.** *We have the following inequality*

$$\begin{tikzcd}
[H,G]^{\Sigma}_{l} \arrow[r, "-\circ\iota"] & [F,G]^{\Sigma}_{l} \\
[H,G]^{\Sigma} \arrow[u, "R_l"] \arrow[r, "-\circ\iota"] \arrow[ur, phantom, "\leq"] & [F,G]^{\Sigma} \arrow[u, "R_l"]
\end{tikzcd}$$

*If this is an equality, i.e. if*

$$R_l(-\circ\iota) = (R_l-)\circ\iota,$$

*then left extensions of $\Sigma$-natural lax transformations along $\iota$ exist and inherit objectwiseness and properness from left extensions of $\Sigma$-natural general transformations.*

*Moreover, right extensions of $\Sigma$-natural lax transformations inherit properness from right extensions of $\Sigma$-natural general transformations.*

*Proof.* The map $[H,G]^{\Sigma} \xrightarrow{-\circ\iota} [F,G]^{\Sigma}$ has a left adjoint by Proposition 4.3.3 and the following diagram commutes

$$\begin{tikzcd}
[H,G]^{\Sigma}_{l} \arrow[r, "-\circ\iota"] \arrow[d, "U_l"] & [F,G]^{\Sigma}_{l} \arrow[d, "U_l"] \\
[H,G]^{\Sigma} \arrow[r, "-\circ\iota"] & [F,G]^{\Sigma}
\end{tikzcd}$$

Since $U_l$ is full and faithful, it follows from the Adjoint Lifting Theorem (Theorem 4 in [35]) that $[H,G]^{\Sigma}_{l} \xrightarrow{-\circ\iota} [F,G]^{\Sigma}_{l}$ has a left adjoint. By the hypothesis we have that $R_l(-\circ\iota) = (R_l-)\circ\iota$. These are all right adjoint, so therefore their left adjoints commute as well, this means that the following diagram commutes

$$\begin{tikzcd}
[H,G]^{\Sigma}_{l} \arrow[d, "U_l"] & [F,G]^{\Sigma}_{l} \arrow[l, "\mathrm{Lan}_{\iota}-"'] \arrow[d, "U_l"] \\
[H,G]^{\Sigma} & [F,G]^{\Sigma} \arrow[l, "\mathrm{Lan}_{\iota}-"']
\end{tikzcd}$$

It follows now easily that properness and objectwiseness are inherited from left extensions of $\Sigma$-natural general transformations.

Suppose now that right extensions of $\Sigma$-natural general transformations along $\iota$ are proper. Because $U_l(-\circ\iota) = (U_l-)\circ\iota$, their right adjoints commute as well. Together with the hypothesis, this leads to the following commutative square.

$$\begin{tikzcd}
{[F,G]_l^\Sigma} \arrow[r, "\mathrm{Ran}_\iota"] & {[H,G]_l^\Sigma} \arrow[r, "\circ\iota"] & {[F,G]_l^\Sigma} \\
{[F,G]^\Sigma} \arrow[u, "R_l"] \arrow[r, "\mathrm{Ran}_\iota"] \arrow[rr, bend right, "1_{[F,G]}"'] & {[H,G]^\Sigma} \arrow[u, "R_l"] \arrow[r, "-\circ\iota"] & {[F,G]^\Sigma} \arrow[u, "R_l"]
\end{tikzcd}$$

We now have for a $\Sigma$-natural lax transformation $\lambda : F \to G$,

$$\mathrm{Ran}_\iota\lambda \circ \iota = \mathrm{Ran}_\iota(R_l U_l \lambda) \circ \iota = R_l U_l \lambda = \lambda.$$

□

We have dual results for Proposition 4.3.4 and Proposition 4.3.6 for $\Sigma$-natural colax transformations.

Under even more conditions, we can guarantee the existence of extensions of strict transformations. Let $T_1$ be the colaxification monad on $[H,G]_l^\Sigma$ as described before Proposition 4.2.6 and let $T_2$ be the colaxification monad on $[F,G]_l^\Sigma$.

**Proposition 4.3.7.** *Suppose that the following diagram commutes*

$$\begin{tikzcd}
{[H,G]_l^\Sigma} \arrow[r, "-\circ\iota"] \arrow[d, "T_1"] & {[F,G]_l^\Sigma} \arrow[d, "T_2"] \\
{[H,G]_l^\Sigma} \arrow[r, "-\circ\iota"'] & {[F,G]_l^\Sigma}
\end{tikzcd}$$

*Then* $-\circ\iota$ *induces an order-preserving map* $([H,G]_l^\Sigma)^{T_1} \to ([F,G]_l^\Sigma)^{T_2}$ *and this map has a right adjoint* $\mathrm{Ran}_\iota^{alg}$.

*Moreover,* $\mathrm{Ran}_\iota^{alg}\lambda = \mathrm{Ran}_\iota^{lax}\lambda$ *for all algebras* $\lambda \in \left([F,G]_l^\Sigma\right)^{T_2}$.

*Proof.* It follows from the hypothesis that the following square commutes.

$$\begin{tikzcd}
{([H,G]_l^\Sigma)^{T_1}} \arrow[r, "-\circ\iota"] & {([F,G]_l^\Sigma)^{T_2}} \\
{[H,G]_l^\Sigma} \arrow[u] \arrow[r, "-\circ\iota"] & {[F,G]_l^\Sigma} \arrow[u]
\end{tikzcd}$$

For a $\Sigma$-natural lax transformation $\lambda$ such that $T_1\lambda = \lambda$, we have that

$$T_2(\lambda \circ \iota) = T_1\lambda \circ \iota = \lambda \circ \iota.$$

Therefore the assignment $\lambda \mapsto \lambda \circ \iota$ induces an order-preserving map $([H,G]_l^\Sigma)^{T_1} \to ([F,G]_l^\Sigma)^{T_2}$

and we have the following commuting square:

$$\begin{tikzcd}
([H,G]_l^\Sigma)^{T_1} \arrow[r, "-\circ\iota"] \arrow[d] & ([F,G]_l^\Sigma)^{T_1} \arrow[d] \\
{[H,G]_l^\Sigma} \arrow[r, "-\circ\iota"] & {[F,G]_l^\Sigma}
\end{tikzcd}$$

By the Adjoint Lifting Theorem (Theorem 4 in [35]), the right adjoint of $-\circ\iota : [H,G]_l^\Sigma \to [F,G]_l^\Sigma$ can be lifted to a right adjoint of the algebras. Because the left adjoints commute, so do the right adjoints. It follows that properness and objectwiseness are inherited from the extensions of $\Sigma$-natural lax transformations. $\square$

Applying this to the case that the *strictness condition* holds, immediately gives us the following corollary.

**Corollary 4.3.8.** *Suppose that $(H,G,\Sigma)$ satisfies the strictness condition and that $\overline{\lambda}\circ\iota = \overline{\lambda\circ\iota}$ for all $\lambda \in [H,G]_l$. Then, $[H,G]_s \xrightarrow{-\circ\iota} [F,G]_s$ has a right adjoint. Moreover,*

$$\mathrm{Ran}_\iota^{strict}\lambda = \mathrm{Ran}_\iota^{lax}\lambda$$

*for all $\lambda \in [F,G]_s$.*

*Proof.* For $\lambda \in [H,G]_l^\Sigma$, we have that

$$T_2(\lambda\circ\iota) = \underline{(\overline{\lambda\circ\iota})} = \underline{(\overline{\lambda}\circ\iota)} = \overline{\lambda}\circ\iota = (T_1\lambda)\circ\iota.$$

Here we used that $\overline{\lambda}$ is a strict transformation, and therefore so is $\overline{\lambda}\circ\iota$.

It follows now from Proposition 4.3.7, that $[H,G]_s \xrightarrow{-\circ\iota} \left([F,G]_l^\Sigma\right)^{T_2}$ has a right adjoint $\mathrm{Ran}_\iota^{\mathrm{alg}}$. We have that

$$[F,G]_s \to \left([F,G]_l^\Sigma\right)^{T_2} \xrightarrow{\mathrm{Ran}_\iota^{\mathrm{alg}}} [H,G]_s$$

is right adjoint to $[H,G]_s \xrightarrow{-\circ\iota} [F,G]_s$. The last claim now also follows from Proposition 4.3.7. $\square$

Again, there are dual results of Proposition 4.3.7 and Corollary 4.3.8 for left extensions of strict transformations.

In general, the existence and properties of left and right extensions are not connected. We can for example have that right extensions do not exist, but left extensions do and are well-behaved. However, we do have the following connection between left and right extensions of strict transformations.

**Proposition 4.3.9.** *Suppose that $[H,G]_s \xrightarrow{-\circ\iota} [F,G]_s$ has a left and a right adjoint. Then, left extensions along $\iota$ are proper if and only if right extensions along $\iota$ are proper.*

*Proof.* Let $\tau \in [F,G]_s$. By the universal property of extensions, we have $(\mathrm{Ran}_\iota\tau)_A \leq \mathrm{Ran}_{\iota_A}\tau_A$ and $\mathrm{Lan}_{\iota_A}\tau_A \leq (\mathrm{Lan}_\iota\tau)_A$ for all $A \in \mathrm{ob}\,\mathcal{C}$. We have the following inequalities for all $A \in \mathrm{ob}\,\mathcal{C}$:

$$(\mathrm{Ran}_\iota\tau\circ\iota)_A \leq \mathrm{Ran}_{\iota_A}\tau_A\circ\iota_A \leq \tau_A \leq \mathrm{Lan}_{\iota_A}\tau_A\circ\iota_A \leq (\mathrm{Lan}_\iota\tau\circ\iota)_A$$

Suppose that left extensions along $\iota$ are proper, i.e. $\mathrm{Lan}_\iota\tau\circ\iota = \tau$. By the universal property of right extensions, it follows that

$$\mathrm{Lan}_\iota\tau \leq \mathrm{Ran}_\iota\tau.$$

Therefore $\tau = \mathrm{Lan}_\iota \tau \circ \iota = \mathrm{Ran}_\iota \tau \circ \iota$.

Similarly, if right extensions along $\iota$ are proper, we find by the universal property of left extensions that $\mathrm{Lan}_\iota \tau \leq \mathrm{Ran}_\iota \tau$ and we can conclude that $\mathrm{Lan}_\iota \tau \circ \iota = \mathrm{Ran}_\iota \tau \circ \iota = \tau$. ☐

An overview of the maps we *always* have is given by the following diagram. We assume that $(F, G, \Sigma)$ and $(H, G, \Sigma)$ satisfy the strictness condition. In the diagram, squares of the same colour commute and the dashed arrows indicate that these form the back of the parallelepiped.

$$\begin{tikzcd}
 & {[H,G]_s} \arrow[rrr, "-\circ\iota" description] \arrow[dr] \arrow[ddl] & & & {[F,G]_s} \arrow[dr] \arrow[ddl, dashed] & \\
 & & {[H,G]^{\Sigma}_{l}} \arrow[ul, green] \arrow[rrr, shift left, "-\circ\iota" description] \arrow[ddl] & & & {[F,G]^{\Sigma}_{l}} \arrow[ul, green] \arrow[lll, shift left, red, "\mathrm{Ran}_\iota -"] \arrow[ddl] \\
{[H,G]^{\Sigma}_{c}} \arrow[uur, red] \arrow[rrr, dashed, shift right, "-\circ\iota" description] \arrow[dr] & & & {[F,G]^{\Sigma}_{c}} \arrow[uur, red, dashed] \arrow[lll, dashed, shift right, green, "\mathrm{Lan}_\iota -"'] \arrow[dr, dashed] & & \\
 & {[H,G]^{\Sigma}} \arrow[ul, green, "L_c"] \arrow[uur, red, "R_l"'] \arrow[rrr, "-\circ\iota" description] & & & {[F,G]^{\Sigma}} \arrow[ul, green, dashed, "L_c"'] \arrow[uur, red, "R_l"'] \arrow[lll, bend right=10, red, "\mathrm{Ran}_\iota -"'] \arrow[lll, bend left=10, green, "\mathrm{Lan}_\iota -"] &
\end{tikzcd}$$

## 4.4 The (co)laxification formula

In this section we will give concrete constructions for the operations $\overline{(-)}$ and $\underline{(-)}$ discussed in Section 4.2, in the case that the functors are well-behaved. Using these we can give explicit constructions of joins and meets in posets of $\Sigma$-natural transformations. These constructions will be used to construct (pre)measures from inner and outer (pre)measures and to describe joins and meets in posets of (pre)measures in Section 4.6.

Let $\mathcal{C}$ be a small poset-enriched category and let $\Sigma$ be a collection of morphisms in $\mathcal{C}$. Let $F$ and $G$ be enriched functors $\mathcal{C} \to \mathbf{Pos}$ such that $GA$ is a complete for all $A \in \mathrm{ob}\,\mathcal{C}$ and such that $Gf$ preserves meets and joins for all $f \in \Sigma$.

Intuitively it is clear that $(F, G, \emptyset)$ satisfying the strictness condition is stronger than $(F, G, \Sigma)$ satisfying the strictness condition. This is the content of the following technical result. This proposition motivates that it is enough to give constructions of these operations in the case that $\Sigma = \emptyset$.

**Proposition 4.4.1.** *If $(F, G, \emptyset)$ satisfies the strictness condition, then $\overline{\lambda}^{\Sigma} = \overline{\lambda}^{\emptyset}$ for all $\lambda \in [F, G]^{\Sigma}_{l}$. In particular, $(F, G, \Sigma)$ satisfies the strictness condition.*[3]

*Proof.* The inclusions $[F, G]^{\Sigma}_{c} \to [F, G]^{\emptyset}_{c}$ and $[F, G]^{\Sigma} \to [F, G]^{\emptyset}$ are meet-preserving maps between complete posets. Therefore, they have left adjoints $F^{\Sigma}_{c} : [F, G]^{\emptyset}_{c} \to [F, G]^{\Sigma}_{c}$ and $F^{\Sigma} : [F, G]^{\emptyset} \to [F, G]^{\Sigma}$ and we have that

$$F^{\Sigma}_{c}(\sigma) = \sigma,$$

for every general $\Sigma$-natural colax transformation $\sigma : F \to G$. Moreover, the following square

---

[3] Here $\overline{(-)}^{\Sigma} : [F, G]^{\Sigma}_{l} \to [F, G]^{\Sigma}_{c}$ and $\overline{(-)}^{\emptyset} : [F, G]^{\emptyset}_{l} \to [F, G]^{\emptyset}_{c}$ refer to the operations described in Section 4.2.

commutes, since their right adjoints do

$$\begin{tikzcd}
{[F,G]^{\Sigma}} \arrow[d, "L_c^{\Sigma}"'] & {[F,G]^{\emptyset}} \arrow[l, "F^{\Sigma}"'] \arrow[d, "L_c^{\emptyset}"] \\
{[F,G]_c^{\Sigma}} & {[F,G]_c^{\emptyset}} \arrow[l, "F_c^{\Sigma}"]
\end{tikzcd}$$

This gives the following commutative diagram

$$\begin{tikzcd}
{[F,G]_l^{\Sigma}} \arrow[r] \arrow[d, "U_l^{\Sigma}"] \arrow[dd, bend right=60, "\overline{(-)}^{\Sigma}"'] & {[F,G]_l^{\emptyset}} \arrow[d, "U_l^{\emptyset}"'] \arrow[dd, bend left=60, "\overline{(-)}^{\emptyset}"] \\
{[F,G]^{\Sigma}} \arrow[d, "L_c^{\Sigma}"] & {[F,G]^{\emptyset}} \arrow[l, "F^{\Sigma}" description] \arrow[d, "L_c^{\emptyset}"'] \\
{[F,G]_c^{\Sigma}} & {[F,G]_c^{\emptyset}} \arrow[l, "F_c^{\Sigma}"]
\end{tikzcd}$$

For a $\Sigma$-natural lax transformation $\lambda : F \to G$, we know by the hypothesis that $\overline{\lambda}^{\emptyset}$ is strict and therefore it is a $\Sigma$-natural transformation. Therefore $\overline{\lambda}^{\Sigma} = F_c^{\Sigma}\overline{\lambda}^{\emptyset} = \overline{\lambda}^{\emptyset}$. □

For the rest of this section, let $\Sigma := \emptyset$. We will write $[F,G]_{\bullet}$ to mean $[F,G]^{\emptyset}_{\bullet}$ and we will refer to $\emptyset$-natural general, lax and colax transformations as just *general, lax and colax transformations* respectively.

To obtain the constructions for $\overline{(-)}$ and $\underline{(-)}$, we will define new functors $F_{\#}, F^{\#}_{\#}$ and $F^{\#}$ and strict transformations

$$\begin{tikzcd}
 & F & \\
F_{\#} \arrow[ur, "c'"] & & F^{\#} \arrow[ul, "l'"'] \\
 & F^{\#}_{\#} \arrow[ul, "c"] \arrow[ur, "l"'] &
\end{tikzcd}$$

such that $[F_{\#}, G]_s \cong [F,G]_c$; $[F^{\#}_{\#}, G]_s \cong [F,G]$ and $[F^{\#}, G]_s \cong [F,G]_l$.

Using these isomorphisms, we can rewrite $\overline{(-)} : [F,G]_l \to [F,G]_c$ as

$$[F,G]_l \cong [F^{\#}, G]_s \xrightarrow{-\circ l} [F^{\#}_{\#}, G]_s \xrightarrow{\mathrm{Lan}_c -} [F_{\#}, G]_s \cong [F,G]_c$$

and similarly, we can write $\underline{(-)} : [F,G]_c \to [F,G]_l$ as

$$[F,G]_c \cong [F_{\#}, G]_s \xrightarrow{-\circ c} [F^{\#}_{\#}, G]_s \xrightarrow{\mathrm{Ran}_l -} [F^{\#}, G]_s \cong [F,G]_c.$$

If the extensions in these compositions are objectwise, we can give an explicit expression for $\overline{(-)}$ and $\underline{(-)}$. In this section we will discuss conditions for when this is the case.

The enriched functor $F^{\#} : \mathcal{C} \to \mathbf{Pos}$ is defined on objects by sending every object $A$ in $\mathcal{C}$ to the lax coend[4]

$$F^{\#}A := \oint^{B\in \operatorname{ob}\mathcal{C}} \mathcal{C}(B,A)\times FB.$$

The functor $F^{\#}$ is called the **lax morphism classifier** and has been discussed in [5, 42, 48]. The following proposition gives a concrete description of $F^{\#}A$ for $A \in \operatorname{ob}\mathcal{C}$. For this, first define the following preorder $\mathcal{P}_A$, for an object $A$ in $\mathcal{C}$:

- The elements of $\mathcal{P}_A$ are pairs $(B \xrightarrow{g} A, y)$, where $g$ is a morphism in $\mathcal{C}$ and $y$ is an element of $FB$,
- We write $(B_1 \xrightarrow{g_1} A, y_1) \preceq (B_2 \xrightarrow{g_2} A, y_2)$ if there exists a morphism $s : B_2 \to B_1$ in $\mathcal{C}$ such that $g_1 s \leq g_2$ and $y_1 \leq Fs(y_2)$.

**Proposition 4.4.2.** *Let $A$ be an object of $\mathcal{C}$. Then $F^{\#}A$ is the poset induced*[5] *by the preorder $\mathcal{P}_A$.*

*Proof.* Let $\hat{\mathcal{P}}_A$ denote the poset induced by the preorder $\mathcal{P}_A$. We will now show that $\hat{\mathcal{P}}_A$ satisfies the universal property of lax coends.

For $B \in \operatorname{ob}\mathcal{C}$, there clearly is an order-preserving map $e_B : \mathcal{C}(B,A)\times FB \to \hat{\mathcal{P}}_A$. A map $s : B_2 \to B_1$ induces

$$\begin{tikzcd}
\mathcal{C}(B_1,A)\times FB_2 \arrow[r, "\mathcal{C}(s,A)\times \mathrm{Id}"] \arrow[d, "\mathrm{Id}\times Fs"'] \arrow[dr, phantom, "\preceq"] & \mathcal{C}(B_2,A)\times FB_2 \arrow[d, "e_{B_2}"] \\
\mathcal{C}(B_1,A)\times FB_1 \arrow[r, "e_{B_1}"'] & \hat{\mathcal{P}}_A
\end{tikzcd}$$

Indeed, for $f : B_1 \to A$ and $y \in FB_2$, we have that

$$(f, Fs(y)) \preceq (fs, y).$$

This means that the poset $\hat{\mathcal{P}}_A$ together with the maps $(e_B)_{B\in\operatorname{ob}\mathcal{C}}$ form a cowedge. We will now show that they form a *universal* cowedge. To do this, consider another cowedge, i.e. a poset $\mathcal{R}$ toghether with order-preserving maps $(\tilde{e}_B : \mathcal{C}(B,A)\times FB \to \mathcal{R})_{B\in\operatorname{ob}\mathcal{C}}$ such that for every morphism $s : B_2 \to B_1$,

$$\begin{tikzcd}
\mathcal{C}(B_1,A)\times FB_2 \arrow[r, "\mathcal{C}(s,A)\times \mathrm{Id}"] \arrow[d, "\mathrm{Id}\times Fs"'] \arrow[dr, phantom, "\preceq"] & \mathcal{C}(B_2,A)\times FB_2 \arrow[d, "\tilde{e}_{B_2}"] \\
\mathcal{C}(B_1,A)\times FB_1 \arrow[r, "\tilde{e}_{B_1}"'] & \mathcal{R}
\end{tikzcd}$$

Let $e : \mathcal{P}_A \to \mathcal{R}$ be the map that sends $(B \xrightarrow{g} A, y)$ to $\tilde{e}_B(g,y)$. For $(g_1,y_1) \preceq (g_2,y_2)$ in $\mathcal{P}_A$, there exists a morphism $s : B_2 \to B_1$ such that $g_1 s \leq g_2$ and $y_1 \leq Fs(y_2)$. Because $\tilde{e}_{B_1}$ and $\tilde{e}_{B_2}$ are order-preserving and because the maps $(\tilde{e}_B)_{B\in\operatorname{ob}\mathcal{C}}$ form a cowedge, we find the following relations in $\mathcal{R}$:

---

[4] Lax coends are explained in the Appendix B

[5] A preorder $P$ induces a poset by identifying elements $a$ and $b$ in $P$ with each other if $a \leq b$ and $b \leq a$.

$$\tilde{e}_{B_1}(g_1, y_1) \leq \tilde{e}_{B_1}(g_1, Fs(y_2)) \leq \tilde{e}_{B_2}(g_1 s, y_2) \leq \tilde{e}_{B_2}(g_2, y_2).$$

This shows that $e$ is order-preserving, and therefore it induces an order-preserving map $\hat{e} : \hat{\mathcal{P}}_A \to \mathcal{R}$. This is the unique order-preserving map such that $e \circ e_B = \tilde{e}_B$ for all $B \in \mathrm{ob}\,\mathcal{C}$. This shows the universal property. ☐

Let $[\mathcal{C}, \mathbf{Pos}]_l$ be the category of enriched functors $\mathcal{C} \to \mathbf{Pos}$ and lax transformations and let $[\mathcal{C}, \mathbf{Pos}]_s$ be the subcategory of enriched functors and strict transformations. The following result is Theorem 3.16 in [5] together with Section 7.1.2 in [48]. In [5] the result follows from a more general theorem. Here we will also give a direct proof.

**Proposition 4.4.3.** *The inclusion $[\mathcal{C}, \mathbf{Pos}]_s \to [\mathcal{C}, \mathbf{Pos}]_l$ has a left adjoint, which is given by the assignment $F \mapsto F^{\#}$.*

*Proof.* It is enough to show that $[F^{\#}, G]_s \cong [F, G]_l$ for all functors $F$ and $G$ naturally in $G$.

Given a strict transformation $\tau : F^{\#} \to G$, we can define for every $A \in \mathrm{ob}\,\mathcal{C}$ a functor $\lambda_A : FA \to GA$ by sending $x$ to $\tau_A(1_A, x)$. These form a lax transformation.

Given a lax transformation $\lambda : F \to G$, we can define a strict transformation $\tau : F^{\#} \to G$, by the assignment

$$\tau_A(g : B \to A, y) := \lambda_B(Fg(y)).$$

These define an isomorphism of posets. ☐

The counit of the adjunction in Proposition 4.4.3 is a strict transformation

$$l' : F^{\#} \to F$$

and is defined by

$$l'_A : F^{\#}A \to FA : (g : B \to A, y) \mapsto Fg(y),$$

for all objects $A$ in $\mathcal{C}$.

Dually, we define the functor $F_{\#} : \mathcal{C} \to \mathbf{Pos}$ by sending every object $A$ in $\mathcal{C}$ to the colax coend

$$F_{\#}A := \oint^{B:\mathcal{C}} \mathcal{C}(B, A) \times FB.$$

A construction dual to the one from Proposition 4.4.2 can be given for this poset. We also have a dual version of Proposition 4.4.3, namely that the assignment $F \mapsto F_{\#}$ defines a left adjoint to the inclusion $[\mathcal{C}, \mathbf{Pos}]_s \to [\mathcal{C}, \mathbf{Pos}]_c$. The functor $F_{\#}$ is called the **colax morphism classifier** and has been studied in [5, 42, 48].

The counit is a strict transformation

$$c' : F_{\#} \to F$$

defined by

$$c'_A : F_{\#}A \to FA : (g, y) \mapsto Fg(y),$$

for all objects $A$ in $\mathcal{C}$.

Finally, consider the functor $F^{\#}_{\#} : \mathcal{C} \to \mathbf{Pos}$ that is defined by sending every object $A$ in $\mathcal{C}$ to

$$F^{\#}_{\#}A := \sum_{B \in \mathrm{ob}\,\mathcal{C}} \mathcal{C}(B, A) \times FB.$$

Similarly to Proposition 4.4.3, it can be shown that the assignment

$$F \mapsto F^{\#}_{\#}$$

defines a left adjoint to the inclusion $[\mathcal{C}, \mathbf{Pos}]_s \to [\mathcal{C}, \mathbf{Pos}]$. Here $[\mathcal{C}, \mathbf{Pos}]$ is the category of poset-enriched functors and general transformations between them.

The unit $F \to F^{\#}$ is an element of $[F, F^{\#}]_l \subseteq [F, F^{\#}]$ and therefore it corresponds to a strict transformation

$$l : F^{\#}_{\#} \to F^{\#}.$$

We have that

$$l_A(g, y) = (g, y)$$

for all $A \in \operatorname{ob}\mathcal{C}$ and $(g, y) \in F^{\#}_{\#}A$.

Similarly, the unit $F \to F_{\#}$ is an element of $[F, F^{\#}]_c \subseteq [F, F^{\#}]$ and also corresponds to a strict transformation

$$c : F^{\#}_{\#} \to F_{\#}$$

such that

$$c_A(g : B \to A, y) = (g, y)$$

for all $A$ in $\mathcal{C}$ and $(g, y)$ in $F^{\#}_{\#}A$.

This gives us the strict transformations that we will need in the rest of this section:

$$\begin{tikzcd}
& F & \\
F_{\#} \arrow[ur, "c'"] & & F^{\#} \arrow[ul, "l'"'] \\
& F^{\#}_{\#} \arrow[ul, "c"] \arrow[ur, "l"'] &
\end{tikzcd}$$

As explained above, writing the general, lax and colax transformations in terms of $F^{\#}$ and $F^{\#}_{\#}$ allows us to describe the operation $\overline{(-)}$ as a left extension. For a lax transformation $\lambda : F \to G$, let $\tilde{\lambda} : F^{\#} \to G$ be the corresponding strict transformation. For $A \in \operatorname{ob}\mathcal{C}$ and $x \in FA$, *assuming* that the left Kan extension exists, we can write

$$\overline{\lambda}_A(x) = \mathrm{Lan}_c(\tilde{\lambda} \circ l)_A(1_A, x).$$

This follows from the following commutative diagram:

$$\begin{tikzcd}
{[F, G]_l} \arrow[r, "U_l"] \arrow[d, phantom, "\wr\Vert"] & {[F, G]} \arrow[r, "L_c"] \arrow[d, phantom, "\wr\Vert"] & {[F, G]_c} \arrow[d, phantom, "\wr\Vert"] \\
{[F^{\#}, G]_l} \arrow[r, "- \circ l"'] & {[F^{\#}_{\#}, G]} \arrow[r, "\mathrm{Lan}_c -"'] & {[F_{\#}, G]}
\end{tikzcd}$$

*Suppose* now that this left extension is objectwise. Because $GA$ is cocomplete and $F^{\#}_{\#}A$ is small for every $A$, we then would have that $\overline{\lambda}_A(x)$ is equal to

$$\begin{aligned}
&\operatorname{colim}(c_A \downarrow (1_A, x) \to F^{\#}_{\#}A \xrightarrow{l_A} F^{\#}A \xrightarrow{\tilde{\lambda}_A} GA) \\
&= \bigvee \left\{ \tilde{\lambda}_A(g, y) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x \right\} \\
&= \bigvee \{ Gg(\lambda_C(y)) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x \}
\end{aligned}$$

In the following two results (Proposition 4.4.4 and Theorem 4.4.5), we will give conditions for when this is indeed the case. More specifically, we will give conditions on the functors $F$ and $G$ such that the **colaxification formula**,

$$\overline{\lambda}_A(x) = \bigvee \{ Gg(\lambda_C(y)) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x \},$$

holds for every lax transformation $\lambda : F \to G$ , object $A$ in $\mathcal{C}$ and $x \in FA$.

**Proposition 4.4.4.** *For $\lambda \in [F, G]^{\Sigma}$ and $A \in \operatorname{ob}\mathcal{C}$, define $\tau_A : FA \to GA$ by*

$$\tau_A(x) := \bigvee \{ Gg(\lambda_C(y)) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x \},$$

*for $x \in FA$. If $\tau$ is colax, then $\tau = L_c\lambda$. If moreover, $\lambda$ is lax, then $\tau = \overline{\lambda}$.*

*Proof.* Let $\tilde{\lambda}$ be the strict transformation $F^{\#}_{\#} \to G$, corresponding to $\lambda$ and let $\tilde{\tau}$ be the strict transformation $F_{\#} \to G$ corresponding to $\tau$. We want to show that $\tilde{\tau}$ is the left extension of $\tilde{\lambda}$ along $c$. We first show that $\tilde{\tau} \circ c \geq \tilde{\lambda}$, i.e.

$$\begin{tikzcd}
F^{\#}_{\#} \arrow[r, "\tilde{\lambda}"] \arrow[d, "c"'] \arrow[dr, phantom, "\geq"] & G \\
F_{\#} \arrow[ur, "\tilde{\tau}"'] &
\end{tikzcd}$$

For $(f : C \to A, y) \in F^{\#}_{\#}A$,

$$\begin{aligned}
(\tilde{\tau} \circ c)_A(f, y) = Gf(\tau_C(y)) &= Gf \bigvee \{ Gg(\lambda_{C'}(y')) \mid g : C' \to C; y' \in FC' \text{ such that } Fg(y') \leq y \} \\
&\geq \bigvee \{ Gf(Gg(\lambda_{C'}(y'))) \mid g : C' \to C; y' \in FC' \text{ such that } Fg(y') \leq y \} \\
&\geq Gf(\lambda_C(y)) = \tilde{\lambda}_A(f, y),
\end{aligned}$$

which means that $\tilde{\tau} \circ c \geq \tilde{\lambda}$.

Suppose now that $\nu : F_{\#} \to G$ is a strict transformation such that $\nu \circ c \geq \tilde{\lambda}$, i.e.

$$\begin{tikzcd}
F^{\#}_{\#} \arrow[r, "\tilde{\lambda}"] \arrow[d, "c"'] \arrow[dr, phantom, "\geq"] & G \\
F_{\#} \arrow[ur, "\nu"'] &
\end{tikzcd}$$

We will now show that $\tilde{\tau} \leq \nu$. Let $(f : C \to A, y) \in F_{\#}A$. Consider $g : C' \to C$ and $y' \in FC'$

such that $Fg(y') \leq y$, then $(g, y') \in F^{\#}_{\#}C$ and $c_C(g, y') \leq (1_C, y)$ in $F_{\#}C$ and therefore

$$\nu_C(1_C, y) \geq (\nu \circ c)_C(g, y') \geq \tilde{\lambda}_C(g, y') = Gg\lambda_{C'}(y').$$

Taking the supremum over all such $(g, y')$, we have

$$\nu_C(1_C, y) \geq \tau_C(y)$$

and by applying $Gf$ to both sides,

$$\nu_A(f, y) = Gf\nu_A(1_C, y) \geq Gf\tau_C(y) = \tilde{\tau}_A(f, y).$$

Here we used that $\nu$ is strict and the definition of $\tilde{\tau}$. This shows that $\tilde{\tau}$ is the left extension of $\tilde{\lambda} \circ l$ along $c$. ☐

Let $A \in \mathrm{ob}\,\mathcal{C}$ and $x \in FA$. Consider the subposets of $F^{\#}A$:

$$l'_A \downarrow x := \{(g : C \to A, y) \mid Fg(y) \leq x\}$$

and

$$(l'_A \downarrow x)_= := \{(g : C \to A, y) \mid Fg(y) = x\}.$$

**Theorem 4.4.5** (Colaxification formula)**.** *Let $\kappa$ be a regular cardinal. Suppose that*

*($G1$) $Gf$ preserves $\kappa$-directed joins for all morphisms $f$ in $\mathcal{C}$,*

*($F1$) $(l'_A \downarrow x)_= \subseteq l'_A \downarrow x$ is cofinal* [6]*,*

*($F2$) $\mathcal{C}$ has and $F$ preserves $\kappa$-wide pullbacks.*

*Then $\overline{\lambda}$ is a strict transformation for every lax transformation $\lambda : F \to G$, i.e. $(F, G, \emptyset)$ satisfies the strictness condition.*

*Furthermore, the colaxification formula holds, i.e. for $A \in \mathrm{ob}\,\mathcal{C}$ and $x \in FA$,*

$$\overline{\lambda}_A(x) = \bigvee \{Gg(\lambda_C(y)) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x\}.$$

*Proof.* For $A \in \mathrm{ob}\,\mathcal{C}$ and $x \in FA$ denote

$$S_{A,x} := \{Gg\lambda_C(y) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x\}$$

and write $\tau_A(x) := \bigvee S_{A,x}$. We will now show that $S_{A,x}$ is $\kappa$-directed.

Let $I$ be a set such that $|I| < \kappa$. And consider a collection $(g_i : C_i \to A, y_i)_{i \in I}$ in $l'_A \downarrow x$. From ($F1$) it follows that for every $i \in I$, there exists a $(\tilde{g}_i : \tilde{C}_i \to A, \tilde{y}_i) \in (l'_A \downarrow x)_=$ such that

$$(g_i, y_i) \preceq (\tilde{g}_i, \tilde{y}_i),$$

i.e. for every $i \in I$ there is a map $s_i : \tilde{C}_i \to C_i$ such that $g_i s_i \leq \tilde{g}_i$ and such that $y_i \leq Fs_i(\tilde{y}_i)$.[7]

---

[6] A subset $S$ of a poset $P$ is **cofinal** if for every $p \in P$ there exists an $s \in S$ such that $p \leq s$.

[7] Note that we need the axiom of choice here for $\kappa > \aleph_0$.

The maps $(\tilde{g}_i)_{i \in I}$ form a $\kappa$-wide pullback diagram. Let $C$ be the pullback of this diagram and let $(p_i : C \to \tilde{C}_i)_{i \in I}$ be the collection of projections.

$$\begin{tikzcd}
 & \vdots & \\
 & \tilde{C}_i \arrow[dr, "\tilde{g}_i"] & \\
C \arrow[ur, "p_i"] \arrow[dr, "p_j"'] \arrow[r, phantom, ">"] & \vdots & A \\
 & \tilde{C}_j \arrow[ur, "\tilde{g}_j"'] & \\
 & \vdots &
\end{tikzcd}$$

Because $F\tilde{g}_i(\tilde{y}_i) = x$ for all $i \in I$, we have that $(\tilde{y}_i)_{i \in I} \in \bigtimes_{FA} FC_i$. Since $F$ preserves this limit, there exists a unique $y \in FC$ such that $Fp_i(y) = \tilde{y}_i$. Let $g : C \to A$ be the morphism that is equal to $\tilde{g}_i p_i$ for all $i \in I$. We find for all $i \in I$ that

$$\begin{aligned}
Gg_i(\lambda_{C_i}(y_i)) &\leq Gg_i(\lambda_{C_i}(Fs_i(\tilde{y}_i))) \\
&= Gg_i(\lambda_{C_i}(Fs_i Fp_i(y))) \\
&\leq (Gg_i s_i p_i)(\lambda_C(y)) \\
&\leq (G\tilde{g}_i p_i)(\lambda_C(y)) = Gg\lambda_C(y)
\end{aligned}$$

We conclude that $S_{A,x}$ is $\kappa$-directed. Using $(G1)$, for a morphism $f : A \to B$, we see that

$$\begin{aligned}
Gf\tau_A(x) &= \bigvee \{Gf(Gg(\lambda_C(y))) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x\} \\
&\leq \bigvee S_{B,Ff(x)} = \tau_B(Ff(x))).
\end{aligned}$$

This shows that $\tau$ is a colax transformation. It follows from Proposition 4.4.4 that $\overline{\lambda} = \tau$. This shows the second claim.

To prove the first claim we need to show that for $f : A \to B$

$$L := \bigvee \{Gf(Gg(\lambda_C(y))) \mid g : C \to A; y \in FC \text{ such that } Fg(y) \leq x\} \geq \bigvee S_{B,Ff(x)}$$

Consider $(g : C \to B, y)$ in $l'_B \downarrow Ff(x)$. Then by $(F1)$ there exist $(\tilde{g} : \tilde{C} \to B, \tilde{y}) \in (l'_B \downarrow Ff(x))_=$ such that $(g, y) \preceq (\tilde{g}, \tilde{y})$, i.e. there is a map $s : \tilde{C} \to C$ such that $gs \leq \tilde{g}$ and $y \leq Fs(\tilde{y})$.

Let $P$ be the pullback of $f$ and $\tilde{g}$ and let $p_1 : P \to A$ and $p_2 : P \to \tilde{C}$ be the projection maps.

$$\begin{tikzcd}
P \arrow[r, "p_1"] \arrow[d, "p_2"'] \arrow[dr, phantom, "\lrcorner", very near start] & A \arrow[d, "f"] \\
\tilde{C} \arrow[r, "\tilde{g}"] & B
\end{tikzcd}$$

Clearly $(x, \tilde{y})$ is an element of $FA \times_{FB} F\tilde{C}$. Because $F$ preserves pullbacks, there exists a unique $z \in F(A \times_B \tilde{C})$ such that $Fp_1(z) = x$ and $Fp_2(z) = \tilde{y}$. We now have

$$\begin{aligned} Gg(\lambda_C(y)) &\leq Gg(\lambda_C(Fs(\tilde{y}))) \\ &= Gg(\lambda_C(FsFp_2(z))) \\ &\leq (Ggsp_2)(\lambda_P(z)) \\ &\leq (G\tilde{g}p_2)(\lambda_P(z)) \\ &= Gf(Gp_1(\lambda_P(z))) \leq L. \end{aligned}$$

Since $(g, y)$ was arbitrary, this shows that $\bigvee S_{B,Ff(x)} \leq L$. This proves that $\overline{\lambda}$ is a strict transformation. □

**Remark 4.4.6.** In the case of Theorem 4.4.5, $(F, G, \emptyset)$ satisfies the strictness conditions. Therefore, Proposition 4.4.1 can be applied.

## 4.5 The extension formula

In this section we will give sufficient conditions for extensions to be objectwise. This will allow us to give concrete formulas for the extensions of $\Sigma$-natural (co)lax transformations that were discussed in Section 4.3. In Section 4.7 we will use these constructions and results to extend premeasures to measures, using the representation of (pre)measures as strict transformations described in Section 4.6.

Let $\mathcal{C}$ be a small poset-enriched category and let $\Sigma$ be a collection of morphisms in $\mathcal{C}$. Let $F, G, H$ be enriched functors $\mathcal{C} \to \mathbf{Pos}$. Suppose that $GA$ is a complete for all $A \in \operatorname{ob} \mathcal{C}$ and that $Gf$ preserves meets and joins for all $f \in \Sigma$ and let $\iota : F \to H$ be a strict transformation.

We will start with a result on the objectwiseness of extensions of $\Sigma$-natural *general* transformations, this is the content of Theorem 4.5.1. In Theorem 4.5.2, we will give sufficient conditions for extensions of $\Sigma$-natural colax transformations to be objectwise.

**Theorem 4.5.1.** *Suppose that $Ff$ and $Hf$ have left adjoints $(Ff)_*$ and $(Hf)_*$ for all $f : A \to B$ in $\Sigma$, such that the following square commutes*

$$\begin{tikzcd} FA \arrow[d, "\iota_A"] & FB \arrow[l, "(Ff)_*"'] \arrow[d, "\iota_B"] \\ HA & HB \arrow[l, "(Hf)_*"] \end{tikzcd}$$

*Then left extensions of $\Sigma$-natural transformations along $\iota$ are objectwise.*

*Proof.* Let $\tau : F \to G$ be a $\Sigma$-natural transformation. It is enough to show that $(\operatorname{Lan}_{\iota_A} \tau_A)_{A \in \operatorname{ob} \mathcal{C}}$

is a $\Sigma$-natural general transformation. Let $f : A \to B$ be in $\Sigma$ and let $x \in HA$,

$$\begin{aligned} Gf\mathrm{Lan}_{\iota_A}\tau_A(x) &= Gf\mathrm{colim}(\iota_A \downarrow x \to FA \xrightarrow{\tau_A} GA) \\ &= \mathrm{colim}(\iota_A \downarrow x \to FA \xrightarrow{\tau_A} GA \xrightarrow{Gf} GB) \\ &= \mathrm{colim}(\iota_A \downarrow x \to FA \xrightarrow{Ff} FB \xrightarrow{\tau_B} GB) \\ &= \mathrm{colim}(\iota_A \downarrow x \to \iota_B \downarrow Hf(x) \to FB \xrightarrow{\tau_B} GB) \\ &= \mathrm{colim}(\iota_B \downarrow Hf(x) \to FB \xrightarrow{\tau_B} GB) = \mathrm{Lan}_{\iota_B}\tau_B(Hf(x)) \end{aligned}$$

The second equality follows from the fact that $Gf$ preserves joins, since $f \in \Sigma$. Because $Gf\tau_A = \tau_B Ff$, we have the third equality. Equality four, follows from the fact that $\iota$ is a strict transformation.

Using the hypothesis, it can be shown that

$$\iota_A \downarrow x \to \iota_B \downarrow Hf(x) \cong \iota_A(Ff)_* \downarrow x$$

is right adjoint and therefore final. The last equality now follows. ☐

**Theorem 4.5.2.** *Let $\kappa$ be a regular cardinal.*

($G1$) *Suppose that $Gf$ preserves $\kappa$-directed joins for all morphisms $f$ in $\mathcal{C}$.*

($\iota 1$) *Suppose that $\iota_A$ is $\kappa$-flat for all $A \in ob\mathcal{C}$, i.e. $\iota_A \downarrow x$ is $\kappa$-directed for all $x \in FA$.*

($\iota 1$) *$Ff$ and $Hf$ have left adjoints $(Ff)_*$ and $(Hf)_*$ for all $f : A \to B$ in $\Sigma$, such that the following square commutes*

$$\begin{tikzcd} FA & FB \arrow[l, "(Ff)_*"'] \arrow[d, "\iota_B"] \\ HA \arrow[from=u, "\iota_A"'] & HB \arrow[l, "(Hf)_*"'] \end{tikzcd}$$

*Then left extensions of $\Sigma$-natural colax transformations along $\iota$ are objectwise.*

*Proof.* Let $\sigma : F \to G$ be a $\Sigma$-natural colax transformation. It is enough to show that $(\mathrm{Lan}_{\iota_A}\sigma_A)_{A\in \mathrm{ob}\,\mathcal{C}}$ is a $\Sigma$-natural colax transformation. Let $f : A \to B$ be a morphism in $\mathrm{ob}\,\mathcal{C}$ and let $x \in HA$,

$$\begin{aligned} Gf\mathrm{Lan}_{\iota_A}\sigma_A(x) &= Gf\mathrm{colim}(\iota_A \downarrow x \to FA \xrightarrow{\sigma_A} GA) \\ &= \mathrm{colim}(\iota_A \downarrow x \to FA \xrightarrow{\sigma_A} GA \xrightarrow{Gf} GB) \\ &\leq \mathrm{colim}(\iota_A \downarrow x \to \iota_B \downarrow Hf(x) \to FB \xrightarrow{\sigma_B} GB) \\ &\leq \mathrm{colim}(\iota_B \downarrow Hf(x) \to FB \xrightarrow{\sigma_B} GB) = \mathrm{Lan}_{\iota_B}\sigma_B(Hf(x)) \end{aligned}$$

In the second equality we use ($G1$). The inequality in the third line follows from the fact that $\sigma$ is colax and the last inequality is induced by the inclusion $\iota_A \downarrow x \to \iota_B \downarrow Hf(x)$. This shows that it is a colax transformation.

To show that the colax transformation is a $\Sigma$-natural, suppose that $f \in \Sigma$. In this case the first inequality becomes an equality because $\sigma$ is a $\Sigma$-natural transformation and therefore $Gf\sigma_A = \sigma_B Ff$. From ($\iota 2$) it is straightforward to check that

$$\iota_A \downarrow x \to \iota_B \downarrow Hf(x) \cong \iota_A(Ff)_* \downarrow x$$

is right adjoint and therefore final. It follows that the second inequality is an equality in this case. ☐

## 4.6 Premeasures as transformations

In this section we will represent certain premeasures by certain transformations between functors. Roughly speaking, colax and lax transformation correspond to *inner* and *outer premeasures* and strict transformations correspond to *measures*. Using these representations, we can apply the results from Section 4.2 and Section 4.5 to obtain results and formulas for joins and meets of certain premeasures and for operations to turn a certain kind of premeasure in a different kind in a universal way.

In this section we will use the notation $\biguplus$ and $\uplus$ to emphasize that we are working with *disjoint* unions of subsets.

### 4.6.1 Premeasures on $\overline{\mathcal{B}}$

We will discuss a representation theorem for premeasures on a collection of subsets $\overline{\mathcal{B}}$ as transformations from a functor $F_{\mathcal{B}}$, describing collections of subsets of a premeasurable space $(X, \mathcal{B})$, to a functor $G$ describing real values (Theorem 4.6.6). This representation result is related to the ideas in Section 3 of [57]. The rest of this section is focused on proving that $(F_{\mathcal{B}}, G, \emptyset)$ satisfies the strictness condition and that the colaxification formula holds. For this we use Theorem 4.4.5.

We will use the poset-enriched category $\mathbf{Part}_c$ whose objects are countable sets and whose morphisms are partial maps between them. For partial maps $f, g : A \to B$ between countable sets $A$ and $B$, we write $f \leq g$ if $\mathrm{dom} f \subseteq \mathrm{dom} g$ and $f(a) = g(a)$ for all $a \in \mathrm{dom} f$. This defines a partial order on the set of partial maps from $A$ to $B$, making $\mathbf{Part}_c$ a poset-enriched category.

For a countable set $A$ and an element $a \in A$, let $s_a : A \to \mathbf{1}$ be the partial map that is only defined on the singleton $\{a\}$. Let $\Sigma$ be the collection of all partial maps with codomain $\mathbf{1}$ that are defined on at most one element.

Let $(X, \mathcal{B})$ be a premeasurable space. Define the collection of subsets

$$\overline{\mathcal{B}} := \left\{ \biguplus_{n=1}^{\infty} B_n \mid B_n \in \mathcal{B} \text{ for all } n \geq 1 \right\}.$$

For a countable set $A$, define $F_{\mathcal{B}}(A)$ as the poset

$$\left\{ (E_a)_{a \in A} \in \overline{\mathcal{B}}^A \mid (E_a)_{a \in A} \text{ is a pairwise disjoint collection} \right\},$$

where the order is defined pointwise. For a partial map $f : A \to B$ of countable sets, define a map $F_{\mathcal{B}} f : F_{\mathcal{B}} A \to F_{\mathcal{B}} B$ by the assignment

$$(E_a)_{a \in A} \mapsto \left( \biguplus_{f(a)=b} E_a \right)_{b \in B}.$$

This defines an enriched functor $F_{\mathcal{B}} : \mathbf{Part}_c \to \mathbf{Pos}$.

Let $A$ be a countable set. The set $[0, \infty]^A$ together with the pointwise order, is a poset which we will denote by $GA$. For a partial map $f : A \to B$ of countable sets, the assignment

$$(p_a)_{a\in A} \mapsto \left(\sum_{f(a)=b} p_a\right)_{b\in B}$$

defines an order-preserving map $Gf : GA \to GB$. We obtain an enriched functor $G : \mathbf{Part}_c \to \mathbf{Pos}$.

**Proposition 4.6.1.** *For a countable set $A$, the poset $GA$ is complete and for $s : A \to \mathbf{1}$ in $\Sigma$, $Gs$ preserves all meets and joins.*

*Proof.* It is clear that $GA = [0,\infty]^A$ is complete.

Suppose that $s$ is only defined on $\{a\}$. For a collection $(x_i)_{i\in I}$ in $GA$, we find that

$$Gs\left(\bigvee_{i\in I} x_i\right) = \left(\bigvee_{i\in I} x_i\right)_a = \left(\bigvee_{i\in I} (x_i)_a\right) = \bigvee_{i\in I} Gs(x_i).$$

In the case that $s$ is nowhere defined, we have that

$$Gs\left(\bigvee_{i\in I} x_i\right) = 0 = \bigvee_{i\in I} Gs(x_i).$$

The same argument can be used to prove that $Gs$ preserves meets. □

**Definition 4.6.2.** Let $\mu : \overline{\mathcal{B}} \to [0,\infty]$ be an order-preserving map such that $\mu(\emptyset) = 0$.

- $\mu$ is called a **outer premeasure** if $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n)$ for every pairwise disjoint collection $(E_n)_n$ in $\overline{\mathcal{B}}$.

- $\mu$ is called a **inner premeasure** if $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) \geq \sum_{n=1}^{\infty} \mu(E_n)$ for every pairwise disjoint collection $(E_n)_n$ in $\overline{\mathcal{B}}$.

In the case that $\mathcal{B}$ is a $\sigma$-algebra, these are called **outer measures** and **inner measures** respectively.

**Remark 4.6.3.** Every premeasure on $\mathcal{B}$ can be uniquely extended to a $\sigma$-additive map $\overline{\mu} : \overline{\mathcal{B}} \to [0,\infty]$. Indeed, define

$$\overline{\mu}\left(\dot{\bigcup_{n=1}^{\infty}} B_n\right) := \sum_{n=1}^{\infty} \mu(B_n).$$

To show that this map is well-defined, consider pairwise disjoint collections $(B_n)_n$ and $(C_m)_m$ in $\mathcal{B}$ such that

$$\dot{\bigcup_{n=1}^{\infty}} B_n = \dot{\bigcup_{m=1}^{\infty}} C_m.$$

We find the following equations:

$$\begin{aligned}\sum_{n=1}^{\infty}\mu(B_n) &= \sum_{n=1}^{\infty}\mu\left(B_n \cap \dot{\bigcup_{m=1}^{\infty}} C_m\right)\\ &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\mu(B_n \cap C_m)\\ &= \sum_{m=1}^{\infty}\mu\left(\dot{\bigcup_{n=1}^{\infty}} B_n \cap C_m\right)\\ &= \sum_{m=1}^{\infty}\mu(C_m).\end{aligned}$$

It is straightforward to see that this is the $\sigma$-additive map that uniquely extends $\mu$. Therefore premeasures can be identified with maps $\overline{\mathcal{B}} \to [0,\infty]$ that are both an inner *and* an outer premeasure.

The poset of order-preserving maps $\mu : \overline{\mathcal{B}} \to [0,\infty]$ such that $\mu(\emptyset) = 0$ with the pointwise order is denoted by $M_0(X,\mathcal{B})$. The subposet of outer (inner) premeasures is denoted by $M_{\leq}(X,\mathcal{B})$ ($M_{\geq}(X,\mathcal{B})$. We will write $M_{=}(X,\mathcal{B})$ for the subposet of premeasures, i.e.

$$M_{=}(X,\mathcal{B}) = M_{\leq}(X,\mathcal{B}) \cap M_{\geq}(X,\mathcal{B}).$$

For $\mu \in M_0(X,\mathcal{B})$ and a countable set $A$, define $\tau^{\mu}_A : F_{\mathcal{B}}A \to GA$ by the assignment

$$(E_a)_a \mapsto (\mu(E_a))_a.$$

**Theorem 4.6.4.** *Let $\mu \in M_0(X,\mathcal{B})$, then $\tau^{\mu} := (\tau^{\mu}_A)_A$ is a $\Sigma$-natural transformation. Moreover,*

- *if $\mu \in M_{\leq}(X,\mathcal{B})$, then $\tau$ is a $\Sigma$-natural* lax *transformation;*
- *if $\mu \in M_{\geq}(X,\mathcal{B})$, then $\tau$ is a $\Sigma$-natural* colax *transformation;*
- *if $\mu \in M_{=}(X,\mathcal{B})$, then $\tau$ is a strict transformation.*

*Proof.* Let $A$ be a countable set and let $a_0 \in A$. For $x := (E_a)_{a\in A} \in F_{\mathcal{B}}A$, we have that

$$Gs_{a_0}(\tau^{\mu}_A(x))_0 = \tau^{\mu}_A(x)_{a_0} = \mu(E_{a_0}) = \tau^{\mu}_{\mathbf{1}}(F_{\mathcal{B}}s_a(x)).$$

For a map $s : A \to \mathbf{1}$ that is nowhere defined, we see that

$$Gs(\tau^{\mu}_A(x))_0 = 0 = \mu(\emptyset) = \tau^{\mu}_{\mathbf{1}}(F_{\mathcal{B}}s(x))_0.$$

This shows that $\tau^{\mu}$ is $\Sigma$-natural. Suppose now that $\mu$ is an outer premeasure. Consider a partial map $f : A \to B$ of countable sets. For $x := (E_a)_a \in F_{\mathcal{B}}A$ and $b \in B$, we find that

$$Gf(\tau^{\mu}_A(x))_b = \sum_{f(a)=b}\mu(E_a) \geq \mu\left(\dot{\bigcup_{f(a)=b}} E_a\right) = \tau^{\mu}_B(F_{\mathcal{B}}f(x))_b.$$

It follows that $\mu$ is a $\Sigma$-natural lax transformation. The other claims follow in a similar way. □

The assignment $\mu \mapsto \tau^{\mu}$ defines order-preserving maps $\varphi_0, \varphi_l, \varphi_c, \varphi_s$ which form the following commutative diagram.

$$\begin{tikzcd}
 & M_0(X,\mathcal{B}) \arrow[rr, "\varphi_0"] & & {[F_{\mathcal{B}}, G]^{\Sigma}} & \\
 & \vee & M_{\geq}(X,\mathcal{B}) \arrow[ul, dashed] \arrow[rr, dashed, "\varphi_c"] & \vee & {[F_{\mathcal{B}}, G]^{\Sigma}_{c}} \arrow[ul] \\
M_{\leq}(X,\mathcal{B}) \arrow[uur] \arrow[rr, "\varphi_l"] & & {[F_{\mathcal{B}}, G]^{\Sigma}_{l}} \arrow[uur] & & \\
 & M_{=}(X,\mathcal{B}) \arrow[ul] \arrow[uur, dashed] \arrow[rr, "\varphi_s"] & & {[F_{\mathcal{B}}, G]s} \arrow[ul] \arrow[uur] &
\end{tikzcd}$$

For a $\Sigma$-natural transformation $\tau : F_{\mathcal{B}} \to G$, denote the order-preserving map $\tau_{\mathbf{1}} : \overline{\mathcal{B}} \to [0, \infty]$ by $\mu_{\tau}$.

**Theorem 4.6.5.** *Let $\tau \in [F_{\mathcal{B}}, G]^{\Sigma}$, then $\mu_{\tau} \in M_0(X, \mathcal{B})$. Moreover,*

- *if $\tau \in [F_{\mathcal{B}}, G]^{\Sigma}_{l}$, then $\mu_{\tau} \in M_{\leq}(X, \mathcal{B})$;*
- *if $\tau \in [F_{\mathcal{B}}, G]^{\Sigma}_{c}$, then $\mu_{\tau} \in M_{\geq}(X, \mathcal{B})$;*
- *if $\tau \in [F_{\mathcal{B}}, G]_{s}$, then $\mu_{\tau} \in M_{=}(X, \mathcal{B})$.*

*Proof.* Consider the map $s : \emptyset \to \mathbf{1}$ and let $*$ be the unique element in $F_{\mathcal{B}}(\emptyset)$. Since $s \in \Sigma$, it follows that

$$\mu_{\tau}(\emptyset) = \tau_{\mathbf{1}}(F_{\mathcal{B}}s(*)) = Gs(\tau_{\emptyset}(*)) = 0.$$

This shows that $\mu_{\tau} \in M_0(X, \mathcal{B})$. Suppose now that $\mu$ is a $\Sigma$-natural lax transformation. A pairwise disjoint collection $(E_n)_n$ in $\overline{\mathcal{B}}$ is an element in $F_{\mathcal{B}}(\mathbb{N})$ which we will denote by $x$. Consider the map $s : \mathbb{N} \to \mathbf{1}$ which is defined everywhere and for every $n \geq 1$, let $s_n : \mathbb{N} \to \mathbf{1}$ be the map that is only defined on $\{n\}$. Since $s_n$ is an element of $\Sigma$, we find that for every $n \in \mathbb{N}$ that

$$\tau_{\mathbb{N}}(x)_n = Gs_n(\tau_{\mathbb{N}}(x)) = \tau_{\mathbf{1}}(F_{\mathcal{B}}(x)) = \tau_{\mathbf{1}}(E_n) = \mu_{\tau}(E_n).$$

Because $\tau$ is lax, we also have that

$$\mu_{\tau}\left(\bigcupdot_{n=1}^{\infty} E_n\right) = \tau_{\mathbf{1}}(F_{\mathcal{B}}s(x)) \leq Gs(\tau_{\mathbb{N}}(x)) = \sum_{n=1}^{\infty} \tau_{\mathbb{N}}(x)_n.$$

Combining the above gives us that

$$\mu_{\tau}\left(\bigcupdot_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu_{\tau}(E_n)$$

and therefore $\mu_{\tau}$ is an outer premeasure. The other claims follow in a similar way. $\square$

The assignment $\tau \mapsto \mu_{\tau}$ defines order-preserving maps $\psi_0, \psi_l, \psi_c, \psi_s$ which form the following commutative diagram.

$$\begin{tikzcd}
 & M_0(X,\mathcal{B}) & & {[F_\mathcal{B},G]^\Sigma} \arrow[ll, "\psi_0" description] & \\
 & & M_{\geq}(X,\mathcal{B}) \arrow[ul, dashed] & & {[F_\mathcal{B},G]^\Sigma_c} \arrow[ul] \arrow[ll, dashed, "\psi_c" description] \\
M_{\leq}(X,\mathcal{B}) \arrow[uur] & & {[F_\mathcal{B},G]^\Sigma_l} \arrow[ll, "\psi_l" description] \arrow[uur] & & \\
 & M_{=}(X,\mathcal{B}) \arrow[ul] \arrow[uur, dashed] & & {[F_\mathcal{B},G]s} \arrow[ll, "\psi_s" description] \arrow[ul] \arrow[uur] &
\end{tikzcd}$$

**Theorem 4.6.6.** *The maps $\psi_0, \psi_l, \psi_c$ and $\psi_s$ are the inverses of the maps $\varphi_0, \varphi_l, \varphi_c$ and $\varphi_s$ respectively.*

*Proof.* It is enough to show that $\psi_0$ is the inverse of $\varphi_0$, since the other maps are restrictions of these maps. For an order-preserving map $\mu : \overline{\mathcal{B}} \to [0,\infty]$ such that $\mu(\emptyset) = 0$, we find that

$$\psi_0(\varphi_0(\mu))(E) = \varphi_0(\mu)_{\mathbf{1}}(E) = \mu(E)$$

for every $E \in \overline{\mathcal{B}}$. It follows that $\psi_0(\varphi_0(\mu)) = \mu$.

Let $\tau : F_\mathcal{B} \to G$ be a $\Sigma$-lax transformation. Let $A$ be a countable set and let $x := (E_a)_{a\in A}$ be an element of $F_\mathcal{B}A$. For $a_0 \in A$, let $s_{a_0} : A \to \mathbf{1}$ be the partial map that is only defined on $\{a_0\}$. Note that this map is an element of $\Sigma$. We find the following equalities:

$$\begin{aligned}
[\varphi_0(\psi_0(\tau))]_A(x)_{a_0} &= [\psi_0(\tau)](E_{a_0}) \\
&= \tau_{\mathbf{1}}(E_{a_0}) \\
&= \tau_{\mathbf{1}}(F_\mathcal{B}s_{a_0}(x)) \\
&= Gs_{a_0}(\tau_A(x)) = \tau_A(x)_{a_0}
\end{aligned}$$

Since this hold for every such $A, x$ and $a_0$, we can conclude that $\varphi_0(\psi_0(\tau)) = \tau$. □

### 4.6.2 The operation $\overline{(-)}$

The operation $\overline{(-)} : [F_\mathcal{B}, G]^\Sigma_l \to [F_\mathcal{B}, G]^\Sigma_c$ induces an operation

$$M_{\leq}(X,\mathcal{B}) \xrightarrow{\varphi_l} [F_\mathcal{B},G]^\Sigma_l \xrightarrow{\overline{(-)}} [F_\mathcal{B},G]^\Sigma_c \xrightarrow{\psi_c} M_{\geq}(X,\mathcal{B})$$

which turns an outer premeasure into an inner premeasure. We will denote this operation also by $\overline{(-)}$, this means that for an outer premeasure $\lambda$ we obtain the inner premeasure

$$\overline{\lambda} := \psi_c(\overline{\varphi_l(\lambda)}).$$

More specifically, we have that for every $E \in \overline{\mathcal{B}}$ that

$$\overline{\lambda}(E) := \overline{\varphi_l(\lambda)}_{\mathbf{1}}(E).$$

We want to apply Theorem 4.4.5 to the transformations that represent premeasures. The following two results (Proposition 4.6.7 and Proposition 4.6.8) will help us to verify the conditions of Theorem 4.4.5. This will lead to the proof of Proposition 4.6.9 that describes how to turn a lax transformation $F_{\mathcal{B}} \to G$ into a premeasure.

In a similar way we obtain an operation $\underline{(-)}$ that send outer premeasures to inner premeasures.

**Proposition 4.6.7.** *The category* $\mathbf{Part}_c$ *has pullbacks and* $F_{\mathcal{B}} : \mathbf{Part}_c \to \mathbf{Pos}$ *preserves them.*

*Proof.* Consider a pullback square in $\mathbf{Part}_c$,

$$\begin{tikzcd} A \times_C B \arrow[r, "p_A"] \arrow[d, "p_B"'] & A \arrow[d, "f"] \\ B \arrow[r, "g"'] & C \end{tikzcd} .$$

Let $(E_a)_a \in F_{\mathcal{B}}A$ and $(F_b)_b \in F_{\mathcal{B}}B$ such that

$$\mathop{\dot{\bigcup}}_{f(a)=c} E_a = \mathop{\dot{\bigcup}}_{g(b)=c} F_b$$

for every $c \in C$. For all $a \in A$ and $b \in B$, we can write $E_a = \dot{\bigcup}_{n=1}^{\infty} B_n^a$ and $F_b = \dot{\bigcup}_{n=1}^{\infty} B_n^b$, where $B_n^a$ and $B_n^b$ are elements of $\mathcal{B}$ for all $n \geq 1$. For $(a,b) \in A \times_C B$ define

$$G_{(a,b)} := \mathop{\dot{\bigcup}}_{n\geq 1, m\geq 1} B_n^a \cap B_m^b = E_a \cap F_b.$$

Then $\big(G_{(a,b)}\big)_{(a,b)}$ is an element of $F_{\mathcal{B}}(A \times_C B)$. This defines a map

$$\varphi : F_{\mathcal{B}}A \times_{F_{\mathcal{B}}C} F_{\mathcal{B}}B \to F_{\mathcal{B}}(A \times_C B).$$

For $a_0 \in A$, we find that

$$\begin{aligned} \mathop{\dot{\bigcup}}_{p_A((a,b))=a_0} G_{(a,b)} &= \mathop{\dot{\bigcup}}_{g(b)=f(a_0)} G_{(a_0,b)} \\ &= \mathop{\dot{\bigcup}}_{g(b)=f(a_0)} E_{a_0} \cap F_b \\ &= E_{a_0} \cap \mathop{\dot{\bigcup}}_{g(b)=f(a_0)} F_b = E_{a_0} \end{aligned}$$

Similarly for $b_0 \in B$, we have that $\dot{\bigcup}_{p_B((a,b))=b_0} G_{(a,b)} = F_{b_0}$. Therefore it follows that $\varphi$ is the inverse to the canonical map $F_{\mathcal{B}}(A \times_C B) \to F_{\mathcal{B}}A \times_{F_{\mathcal{B}}C} F_{\mathcal{B}}B$. ☐

**Proposition 4.6.8.** *For* $f : A \to B$ *in* $\mathbf{Part}_c$*,* $Gf$ *preserves directed joins.*

*Proof.* Let $(p^i)_{i\in I}$ be a directed collection of elements in $GA$. For $a \in A$, define $p_a := \sup_{i\in I} p_a^i$. For $b \in B$, we always have that

$$Gf((p_a)_a)_b = \sum_{f(a)=b} p_a \geq \sup_{i\in I} \sum_{f(a)=b} p_a^i = \sup_{i\in I} Gf(p^i).$$

To prove the other inequality, consider an element $b \in B$. Suppose first that there exists $a \in f^{-1}(\{b\})$ such that $p_a = \infty$. In that case, for every $N \geq 1$, there exists $i_N \in I$ such that $N \leq p_a^{i_N}$.

In that case

$$N \leq p_a^{i_N} \leq \sum_{f(a)=b} p_a^{i_N} \leq \sup_{i\in I} \sum_{f(a)=b} p_a^i.$$

By taking $N \to \infty$, we conclude that $\sup_{i\in I} \sum_{f(a)=b} p_a^i = \infty$, which shows the equality.

Suppose now that $p_a < \infty$ for all $a \in f^{-1}(\{b\})$ and consider a finite set $A' \subseteq f^{-1}(\{b\})$ and $\epsilon > 0$. For every $a \in A'$ there exists $i_a \in I$ such that

$$p_a - \frac{\epsilon}{|A'|} \leq p_a^{i_a}.$$

Because $(p^i)_{i\in I}$ is directed and $A'$ is finite, there exists an $i_0 \in I$ such that $p^{i_a} \leq p^{i_0}$ for all $a \in A'$. Summing over $A'$ leads to

$$\sum_{a\in A'} p_a - \epsilon \leq \sum_{a\in A'} p_a^{i_a} \leq \sum_{a\in A'} p_a^{i_0} \leq \sum_{f(a)=b} p_a^{i_0} \leq \sup_{i\in I} \sum_{f(a)=b} p_a^i.$$

The claim now follows by first taking $\epsilon \to 0$ and then taking the supremum over all finite subsets of $f^{-1}(\{b\})$. ☐

Using the previous results (Proposition 4.6.7 and Proposition 4.6.8) we can now obtain a formula to turn an outer measure into a measure in a universal way, by applying Theorem 4.4.5.

**Proposition 4.6.9.** *The triple* $(F_{\sigma(\mathcal{B})}, G, \Sigma)$ *satisfies the strictness condition. Moreover, the map* $M_{\leq}(X, \sigma(\mathcal{B})) \cong [F_{\sigma(\mathcal{B})}, G]_l^{\Sigma} \to [F_{\sigma(\mathcal{B})}, G]_s \cong M(X, \sigma(\mathcal{B}))$, *sends an outer measure* $\mu$ *to a measure* $\overline{\mu}$, *which is defined as*

$$\overline{\mu}(E) := \sup \left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid \biguplus_{n=1}^{\infty} E_n \subseteq E \right\},$$

*for all* $E \in \sigma(\mathcal{B})$.

*Proof.* By Proposition 4.6.8, we know that $Gf$ preserves directed joins for $f : A \to B$ in $\mathbf{Part}_c$ From Proposition 4.6.7 we know that $F_{\mathcal{B}}$ preserves pullbacks.

We will now prove that condition $(F1)$ from Theorem 4.4.5 holds. Let $A$ be a countable set and let $(X_a)_{a\in A} \in F_{\sigma(\mathcal{B})}A$. Consider $g : C \to A$ and $(Y_c)_{c\in C} \in F_{\sigma(\mathcal{B})}C$ such that $Z_a := \biguplus_{g(c)=a} Y_c \subseteq X_a$ for all $a \in A$. In other words

$$(g : C \to A, (Y_c)_{c\in C})$$

is an element of $l'_A \downarrow (X_a)_{a\in A}$.

For $d \in C \amalg A =: D$, define

$$\tilde{Y}_d := \begin{cases} Y_c \text{ if } d = c \\ X_a \setminus Z_a \text{ if } d = a \end{cases}.$$

It is clear that $(\tilde{Y})_{d\in D}$ is an element of $F_{\mathcal{B}}D$. Consider the partial map $g + 1 : D \to A$, which is the same as $g$ on $C$ and the identity on $A$. Let $s : D \to C$ be the finite partial map that is the identity on $C$ and is not defined on $A$.

We have the following inequality

$$\begin{tikzcd}
C \arrow[r, "g"] & A \\
D \arrow[u, "s"] \arrow[ur, "\leq", "g+1"'] &
\end{tikzcd}$$

and we have that $\biguplus_{s(d)=c} \tilde{Y}_d = Y_c$ for all $c \in C$ and

$$\biguplus_{(g+1)(d)=a} \tilde{Y}_d = Z_a \uplus (X_a \setminus Z_a) = X_a.$$

for all $a \in A$. This means that

$$(g+1 : D \to A, (\tilde{Y}_d)_{d \in D})$$

is an element of $(l'_A \downarrow (X_a)_{a \in A})_=$ and that

$$(g, (Y_c)_{c \in C}) \preceq (g+1, (\tilde{Y}_d)_{d \in D}).$$

Therefore it follows that $(l'_A \downarrow (X_a)_{a \in A})_= \subseteq (l'_A \downarrow (X_a)_{a \in A})$ is cofinal. Theorem 4.4.5 now says that $(F_{\sigma(\mathcal{B})}, G, \Sigma)$ satisfies the strictness condition and that for $\lambda \in [F_{\sigma(\mathcal{B})}, G]_l^\Sigma$ and $x \in F_{\sigma(\mathcal{B})}A$

$$\overline{\lambda}_A(x) = \bigvee \{Gg\lambda_C(y) \mid g : C \to A; y \in F_{\mathcal{B}}C; F_{\mathcal{B}}g \leq x\}.$$

Let $\mu$ be an outer premeasure, then by definition $\overline{\mu} := \mu_{\overline{\lambda^\mu}}$. For $E \in \sigma(\mathcal{B})$ this means,

$$\begin{aligned}
\overline{\mu}(E) &= \overline{\lambda^\mu}_{\mathbf{1}}(E) \\
&= \sup\left\{ Gg\lambda^\mu((E_c)_c) \mid g : C \to \mathbf{1};\ \biguplus_{c \in \mathrm{dom} g} E_c \subseteq E \right\} \\
&= \sup\left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid \biguplus_{n=1}^{\infty} E_n \subseteq E \right\}.
\end{aligned}$$

□

The following proposition now immediately follows from Corollary 4.2.7, giving us a formula for the joins of premeasures.

**Proposition 4.6.10.** *The poset $M_=(X, \sigma(\mathcal{B}))$ has all joins and for $(\mu_i)_i$ in $M_=(X, \sigma(\mathcal{B}))$,*

$$\left(\bigvee_{i \in I} \mu_i\right)(E) = \sup\left\{ \sum_{n=1}^{\infty} \sup_{i \in I} \mu_i(E_n) \mid \biguplus_{n=1}^{\infty} E_n \subseteq E \right\}$$

*for $E \in \sigma(\mathcal{B})$.*

### 4.6.3 The operation $R_l$ and $\underline{(-)}$

The operation $R_l : [F_{\mathcal{B}}, G]^\Sigma \to [F_{\mathcal{B}}, G]_l^\Sigma$ induces an operation

$$M_0(X, \mathcal{B}) \xrightarrow{\varphi_0} [F_{\mathcal{B}}, G]^\Sigma \xrightarrow{R_l} [F_{\mathcal{B}}, G]_l^\Sigma \xrightarrow{\psi_l} M_{\leq}(X, \mathcal{B})$$

which turns an order-preserving map $\mu : \overline{\mathcal{B}} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ into an outer premeasure.

We will denote this operation also by $R_l$, the means that for an order-preserving map $\mu : \overline{\mathcal{B}} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ we obtain an outer premeasure

$$R_l(\mu) := \psi_l(R_l(\varphi_0(\mu)).$$

Therefore, for every $E \in \overline{\mathcal{B}}$ we have that $R_l(\mu)(E) = R_l(\varphi_0(\mu))_{\mathbf{1}}(E)$.

**Proposition 4.6.11.** *For $\mu \in M_0(X, \mathcal{B})$ and for $E \in \overline{\mathcal{B}}$, we have that*

$$R_l\mu(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid E \subseteq \dot{\bigcup_{n=1}^{\infty}} E_n \right\}.$$

*Proof.* By the dual of Proposition 4.4.4 it is enough to show that

$$E \mapsto \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) \mid E \subseteq \dot{\bigcup_{n=1}^{\infty}} E_n \right\} =: \tilde{\mu}(E)$$

is an outer premeasure for every order-preserving $\mu : \overline{\mathcal{B}} \to [0, \infty]$ such that $\mu(\emptyset) = 0$.

Consider pairwise disjoint $(E_n)_n$ in $\overline{\mathcal{B}}$ and let $\epsilon > 0$. For every $n \geq 1$, there exists pairwise disjoint $(E^n_m)_m$ in $\overline{\mathcal{B}}$ such that $E_n \subseteq \dot{\bigcup}_{m=1}^{\infty} E^n_m$ and such that

$$\tilde{\mu}(E_n) \geq \sum_{n=1}^{\infty} \mu(E^n_m) - \frac{\epsilon}{2^n} \geq \sum_{n=1}^{\infty} \mu(E^n_m \cap E_n) - \frac{\epsilon}{2^n}.$$

Summing over all $n \geq 1$ gives us that

$$\sum_{n=1}^{\infty} \tilde{\mu}(E_n) \geq \sum_{n \geq 1, m \geq 1} \mu(E^n_m \cap E_n) - \epsilon \geq \tilde{\mu}\left( \dot{\bigcup_{n=1}^{\infty}} E_n \right) - \epsilon.$$

Taking $\epsilon \to 0$ gives us the result.

□

Since $\underline{(-)}$ is defined as the restriction of $R_l$ to inner premeasures, we have the following corollary.

**Corollary 4.6.12.** *Let $\sigma \in M_{\geq}(X, \mathcal{B})$ then for $E \in \overline{\mathcal{B}}$,*

$$\underline{\sigma}(E) = \inf \left\{ \sum_{n=1}^{\infty} \sigma(E_n) \mid E \subseteq \dot{\bigcup_{n=1}^{\infty}} E_n \right\}.$$

*Moreover, the poset $M_{=}(X, \mathcal{B})$ has all meets and for $(\sigma_i)_i \in M_{+}(X, \mathcal{B})$,*

$$\left( \bigwedge_{i \in I} \sigma_i \right)(E) = \inf \left\{ \sum_{n=1}^{\infty} \inf_{i \in I} \sigma_i(E_n) \mid E \subseteq \dot{\bigcup_{n=1}^{\infty}} E_n \right\}$$

*for every $E \in \overline{\mathcal{B}}$.*

## 4.7 The Carathéodory extension theorem

In this last section, we will apply the results of extensions of transformations from Section 4.3 and Section 4.5 to the transformations that represent measures and premeasures, as described in Section 4.6. This will lead to a proof of the Carathéodory extension theorem.

Let $(X, \mathcal{B})$ be a premeasurable space. Let $\sigma(\mathcal{B})$ be the $\sigma$-algebra generated by $\mathcal{B}$. Let $\Sigma, F_{\mathcal{B}}, F_{\sigma(\mathcal{B})}$ and $G$ as in Section 4.6. For a countable set $A$, we have an inclusion $\iota_A : F_{\mathcal{B}}A \to F_{\sigma(\mathcal{B})}A$. These define a strict transformation $\iota : F_{\mathcal{B}} \to F_{\sigma(\mathcal{B})}$. Our goal is to prove that we can extend strict transformations $F_{\mathcal{B}} \to G$ along $\iota$ and that these extensions are proper.

We will show that right extensions of strict transformations always exist, which will imply the Carathéodory extension theorem. We will refer to this extension as the **right Carathéodory extension**. Moreover, this characterizes the right Carathéodory extension by a universal property. We will also briefly discuss left extensions. However, we will show that the **left Carathéodory extension** does not exist in general.

The operation $\mathrm{Ran}_\iota^{\mathrm{lax}} : [F_{\mathcal{B}}, G]_l^\Sigma \to [F_{\sigma(\mathcal{B})}, G]_l^\Sigma$ induces an operation

$$M_{\leq}(X, \mathcal{B}) \xrightarrow{\psi_l} [F_{\mathcal{B}}, G]_l^\Sigma \xrightarrow{\mathrm{Ran}_\iota^{\mathrm{lax}}} [F_{\sigma(\mathcal{B})}, G]_l^\Sigma \xrightarrow{\varphi_l} M_{\leq}(X, \sigma(\mathcal{B})),$$

which turns an outer premeasure into an outer measure. We will denote this operation also by $\mathrm{Ran}_\iota^{\mathrm{lax}}$, i.e. for an outer premeasure $\lambda$ we obtain an outer measure

$$\mathrm{Ran}_\iota^{\mathrm{lax}} \lambda := \varphi_l(\mathrm{Ran}_\iota^{\mathrm{lax}}(\psi_l(\lambda))).$$

We use similar notational conventions for Kan extensions of premeasures, inner premeasures and order-preserving set functions that send $\emptyset$ to 0.

### 4.7.1 The right Carathéodory extension

**Lemma 4.7.1.** *For $\mu \in M_0(X, \mathcal{B})$, the right extension of $\mu$ along $\iota$ exists and for every $E \in \sigma(\mathcal{B})$,*

$$(\mathrm{Ran}_\iota^{\mathrm{gen}} \mu)(E) = \inf \left\{ \mu(B) \mid E \subseteq B, B \in \overline{\mathcal{B}} \right\}.$$

*Proof.* It follows by Proposition 4.3.3 that the right extension of $\mu$ along $\iota$ exists. Observe that the following diagram commutes.

$$\begin{tikzcd}
M_0(X, \mathcal{B}) \arrow[d, phantom, "\cong"] & M_0(X, \sigma(\mathcal{B})) \arrow[l, "(-)|_{\overline{\mathcal{B}}}"'] \arrow[d, phantom, "\cong"] \\
{[F_{\mathcal{B}}, G]^\Sigma} & {[F_{\sigma(\mathcal{B})}, G]^\Sigma} \arrow[l, "-\circ\iota"']
\end{tikzcd}$$

Hence the right Kan extensions corresponds to the usual right Kan extension of order-preserving maps. We only need to verify that $\inf\{\mu(E) \mid \emptyset \subseteq E, E \in \overline{\mathcal{B}}\} = 0$, but this is trivial. ☐

We will now describe what right extensions of outer premeasures look like.

**Proposition 4.7.2.** *Let $\lambda \in M_{\leq}(X, \mathcal{B})$. The right extension of $\lambda$ along $\iota$ exists. If $\lambda$ is a premeasure, the extension is proper.*

*Proof.* The existence follows from Proposition 4.3.4. Let $E \in \sigma(\mathcal{B})$. By Proposition 4.6.11 and

Lemma 4.7.1 we find that

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda](E) = \inf\left\{\sum_{n=1}^{\infty}(\mathrm{Ran}_\iota^{\mathrm{gen}}\lambda)(E_n) \mid E \subseteq \biguplus_{n=1}^{\infty} E_n, E_n \in \sigma(\mathcal{B})\right\}$$

For $B \in \overline{\mathcal{B}}$, we always have that

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda](B) \leq \lambda(B).$$

We will now show that the other inequality also holds in the case that $\lambda$ is a premeasure. For $\epsilon > 0$, there exists a collection $(E_n)_n$ of pairwise disjoint element in $\sigma(\mathcal{B})$ such that $B \subseteq \biguplus_{n=1}^{\infty} E_n$ and such that

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}](B) \geq \sum_{n=1}^{\infty}[\mathrm{Ran}_\iota^{\mathrm{gen}}\lambda](E_n) - \frac{\epsilon}{2}.$$

Using Lemma 4.7.1, it follows that for every $n \geq 1$, there exists $B_n \in \overline{\mathcal{B}}$ such that $E_n \subseteq B_n$ and

$$[\mathrm{Ran}_\iota^{\mathrm{gen}}\lambda](E_n) \geq \lambda(B_n) - \frac{\epsilon}{2^{n+1}}.$$

It follows now that

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda](B) \geq \sum_{n=1}^{\infty}\lambda(B_n) - \epsilon.$$

In the case that $\lambda$ is a measure, then

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda](B) \geq \lambda(B) - \epsilon,$$

because $B \subseteq \bigcup_{n=1}^{\infty} B_n$. By letting $\epsilon \to 0$, we can conclude that

$$[\mathrm{Ran}_\iota^{\mathrm{lax}}\lambda](B) = \lambda(B).$$

Because the following diagram commutes, it follows that the extension is proper in the case that $\lambda$ is a premeasure.

$$\begin{tikzcd}
M_{\leq}(X,\mathcal{B}) \arrow[d, phantom, "\cong"] & M_{\leq}(X,\sigma(\mathcal{B})) \arrow[l, "(-)|_{\overline{\mathcal{B}}}"'] \arrow[d, phantom, "\cong"] \\
{[F_{\mathcal{B}}, G]_l^{\Sigma}} & {[F_{\sigma(\mathcal{B})}, G]_l^{\Sigma}} \arrow[l, "-\circ\iota"']
\end{tikzcd}$$

□

**Theorem 4.7.3.** *For $\mu \in M_{\leq}(X,\sigma(\mathcal{B}))$, we have that*

$$\overline{\mu} \mid_{\mathcal{B}} = \overline{\mu \mid_{\mathcal{B}}}.$$

*Proof.* Since $\mu \leq \overline{\mu}$, we have that $\mu \mid_{\mathcal{B}} \leq \overline{\mu} \mid_{\mathcal{B}}$ and therefore $\overline{\mu \mid_{\mathcal{B}}} \leq \overline{\mu} \mid_{\mathcal{B}}$.

Let $C \in \mathcal{B}$. If $\overline{\mu \mid_{\mathcal{B}}}(C) = \infty$, then we have that $\infty = \overline{\mu \mid_{\mathcal{B}}}(C) = \overline{\mu}(C)$. Suppose now that $\mu \mid_{\mathcal{B}} (C)$ is finite. Define the set

$$S_C := \{A \in \sigma(\mathcal{B}) \mid \forall\epsilon > 0\ \exists B \in \mathcal{B} : \mu(A\Delta_C B) < \epsilon \text{ and } A \cup B \subseteq C\}$$

Here we use the notation $A\Delta_C B := ((C \setminus A) \cap B) \cup (A \cap (C\setminus))$, i.e. the symmetric difference of subsets of $C$.

If $A \in S_C$ and $\epsilon > 0$, then there exists $B$ such that $\mu(A\Delta B) < \epsilon$. Therefore $\mu((C\setminus A - \Delta_C B) = \mu(A\Delta_C(C \setminus B)) < \epsilon$. We conclude that $C \setminus A \in S_A$ and therefore $S_A$ is closed under relative complements within $C$.

Let $\epsilon > 0$. For $A_1, A_2 \in S_C$, there exists $B_1, B_2 \in \mathcal{B}$ such that $\mu(A_1\Delta_C B_1) < \frac{\epsilon}{3}$ and $\mu(A_2\Delta_C B_2) < \frac{\epsilon}{3}$. Since $(A_1 \cup A_2)\Delta_C(B_1 \cup B_2) \subseteq (A_1\Delta_C B_1) \cup (A_2\Delta_C B_2)$, it follows that $\mu((A_1 \cup A_2)\Delta_C(B_1 \cup B_2)) \le \epsilon$. Therefore $S_C$ is closed under finite unions.

Let $\epsilon > 0$ and let $(A_n)_{n=1}^{\infty}$ pairwise disjoint in $S_C$. For $N \ge 1$, there exist $(B_n)_{n=1}^{N}$ such that $\mu(A_n\Delta_C B_n) \le \frac{\epsilon}{2^{N+1}}$ for all $n \in \{1, \dots, N\}$. These can be refined to $2^N - 1$ disjoint $(\tilde{B}_n)_{n=1}^{2^N-1}$ elements in $\mathcal{B}$ of which $2^N - 1 - N$ are contained in the intersection of two distinct elements of $(B_n)_{n=1}^{N}$. Let $\tilde{B}_1, \dots, \tilde{B}_N$ be those subsets that are not contained in an intersection of any two distinct elements of $(B_n)_{n=1}^{N}$. For distinct $n_1, n_2 \le N$, we find that

$$\begin{aligned}\mu(B_{n_1} \cap B_{n_2}) &\le \mu(B_{n_1} \cap B_{n_2} \cap A_{n_1}) + \mu(B_{n_1} \cap B_{n_2} \cap C \setminus A_{n_1})\\ &\le \mu(B_{n_2} \cap C \setminus A_{n_2}) + \mu(B_{n_1} \cap C \setminus A_{n_1})\\ &\le 2\frac{\epsilon}{2^{N+1}} = \frac{\epsilon}{2^N}\end{aligned}$$

It follows that for $k \in N+1, \dots, 2^{N-1}$ we have that $\mu(\tilde{B}_k) \le \frac{\epsilon}{2^N}$. It follows now that

$$\begin{aligned}\sum_{n=1}^{N}\mu(A_n) &\le \sum_{n=1}^{N}(\mu(A_n \cap B_n) + \mu(A_n\Delta_C B_n))\\ &\le \sum_{n=1}^{N}\left(\mu(B_n) + \frac{\epsilon}{2^{N+1}}\right)\\ &\le \sum_{n=1}^{2^N-1}\mu(\tilde{B}_n) + \epsilon\\ &\le \sum_{n=1}^{N}\mu(\tilde{B}_n) + \sum_{n=N+1}^{2^N-1}\mu(\tilde{B}_n) + \epsilon\\ &\le \sum_{n=1}^{N}\mu(\tilde{B}_n) + 2\epsilon \le \overline{\mu\mid_{\mathcal{B}}}(C) + 2\epsilon\end{aligned}$$

By letting $N \to \infty$ and then letting $\epsilon \to 0$, we find for pairwise disjoint $(A_n)_{n=1}^{\infty}$ in $S_C$, that

$$\sum_{n=1}^{\infty}\mu(A_n) \le \overline{\mu\mid_{\mathcal{B}}}(C) < \infty \tag{4.1}$$

It follows that there exists $N \ge 1$ such that $\sum_{n=N+1}^{\infty}\mu(A_n) < \frac{\epsilon}{2}$. Because

$$\left(\dot{\bigcup_{n=1}^{\infty}} A_n\right)\Delta_C\left(\dot{\bigcup_{n=1}^{N}} B_n\right) \subseteq \left(\bigcup_{n=1}^{N} A_n\Delta_C B_n\right) \cup \dot{\bigcup_{n=N+1}^{\infty}} A_n,$$

we have that

$$\mu\left(\dot{\bigcup_{n=1}^{\infty}} A_n \Delta \bigcup_{n=1}^{N} B_n\right) \leq \sum_{n=1}^{N} \mu(A_n \Delta_C B_n) + \sum_{n=N+1}^{\infty} \mu(A_n) \leq \epsilon.$$

We conclude that $S_C$ is closed under countable disjoint unions. This means that $S$ is a $\sigma$-algebra of subset of $C$ that contains all elements in $\mathcal{B}$ that are contained in $C$. By Proposition 4.6.9 and (4.1) it follows that $\overline{\mu}(C) \leq \overline{\mu_{\mathcal{B}}}(C)$. □

**Theorem 4.7.4.** *Right extensions along $\iota$ of strict transformations exist and are proper.*

*Proof.* Theorem 4.7.3 states that $\overline{\mu\mid_{\mathcal{B}}} = \overline{\mu}\mid_{\mathcal{B}}$ for all outer measures $\mu$. This shows that $\overline{\lambda}\circ\iota = \overline{\lambda\circ\iota}$ for all $\lambda \in [F_{\sigma(\mathcal{B})}, G]_l^{\Sigma}$. Because $(F_{\sigma(\mathcal{B})}, G, \Sigma)$ satisfy the strictness condition, the result now follows from Corollary 4.3.8 and Proposition 4.7.2. □

**Theorem 4.7.5** (Carathéodory)**.** *Let $\rho$ be a premeasure on a premeasurable space $(X, \mathcal{B})$. There exists a measure $\mu$ on $(X, \sigma(\mathcal{B}))$, such that $\mu(B) = \rho(B)$ for all $B \in \mathcal{B}$.*

*Proof.* Let $\mu$ be the measure corresponding to the right extension along $\iota$ of the strict transformation $\tau^{\rho} : F_{\mathcal{B}} \to G$. Since the extension is proper, it follows that $\mu(B) = \rho(B)$ for all $B \in \mathcal{B}$. □

### 4.7.2 The left Carathéodory extension

We start by showing that extensions of $\Sigma$-natural colax transformations along $\iota$ are objectwise and proper. This is an application of Theorem 4.5.2.

**Proposition 4.7.6.** *Let $\sigma \in M_{\geq}(X, \mathcal{B})$, then $\mathrm{Lan}_{\iota}^{colax}\sigma$ exists and is proper.*

*Proof.* For $s : A \to \mathbf{1}$ in $\Sigma$ that is defined on $\{a_0\}$. We have that $F_{\mathcal{B}}s : F_{\mathcal{B}}A \to F_{\mathcal{B}}\mathbf{1}$ is defined by the assignment

$$(E_a)_{a\in A} \mapsto E_{a_0}.$$

Now define $(F_{\mathcal{B}}s)_* : F_{\mathcal{B}}\mathbf{1} \to F_{\mathcal{B}}A$ by

$$(F_{\mathcal{B}}s)_*(B)_a := \begin{cases} B \text{ if } a = a_0 \\ \emptyset \text{ otherwise} \end{cases},$$

for all $B \in F_{\mathcal{B}}\mathbf{1}$ and $a \in A$.

It is straightforward to verify that $(F_{\mathcal{B}}s_{a_0})_*$ is left adjoint to $F_{\mathcal{B}}s_{a_0}$ and these left adjoints satisfy condition $(\iota_2)$ in Theorem 4.5.1.

It follows now from Theorem 4.5.1 that $\tilde{\sigma} := \mathrm{Lan}_{\iota}^{\mathrm{gen}}\sigma$ exists and is objectwise, i.e.

$$[\mathrm{Lan}_{\iota}^{\mathrm{gen}}\sigma](E) = \sup\left\{\sigma(B) \mid B \subseteq E\right\}.$$

Since $\iota_A$ is full and faithful for all countable sets $A$, we know by Lemma 4.3.1 that the extension is proper.

We will now show that $\tilde{\sigma}$ is an inner measure. Let $(E_n)_n$ be a pairwise disjoint collection in $\overline{\mathcal{B}}$ and let $\epsilon > 0$. For every $n \geq 1$, there exists a $B_n \in \overline{\mathcal{B}}$ such that $B_n \subseteq E_n$ and such that

$$\tilde{\sigma}(E_n) \leq \sigma(B_n) + \frac{\epsilon}{2^n}.$$

Summing over all $n \geq 1$, gives us that

$$\sum_{n=1}^{\infty} \tilde{\sigma}(E_n) - \epsilon \leq \sum_{n=1} \sigma(B_n) \leq \sigma\left(\dot{\bigcup_{n=1}^{\infty}} B_n\right) \leq \tilde{\sigma}.$$

By letting $\epsilon \to 0$, we can see that $\mathrm{Lan}_{\iota}^{\mathrm{gen}}\sigma$ is an inner measure.

We know by Proposition 4.3.4 that $L_c\mathrm{Lan}_{\iota}^{\mathrm{gen}}\sigma = \mathrm{Lan}_{\iota}^{\mathrm{colax}}\sigma$. But since $\mathrm{Lan}_{\iota}^{\mathrm{gen}}$ is already colax, it follows that $[\mathrm{Lan}_{\iota}^{\mathrm{colax}}](E) = \sup\{\sigma(B) \mid B \subseteq E\}$ for all $E \in \sigma(\mathcal{E})$ and that $\mathrm{Lan}_{\iota}^{\mathrm{colax}}\sigma$ is proper. □

However, left extensions along $\iota$ of *strict* transformations might not exist in general. In other words, the left Carathéodory extension of a premeasure does not always exist, as can be seen from the following counterexample, based on Example 4.20 in [64].

**Example 4.7.7.** Let $\mathcal{B}$ be the algebra of subsets of $\mathbb{Q}$, generated by $\{(a, b] \cap \mathbb{Q} \mid a, b \in \mathbb{Q}\}$. Let $\rho : \mathcal{B} \to [0, \infty]$ be the premeasure that takes the value $\infty$ for all non-empty subsets in $\mathcal{B}$. Suppose that this premeasure has a left Carathéodory extension $\mu$. Note that since right Caratheodory extensions exist and are proper, we know by Proposition 4.3.9 that the left Carathédoroy extension has to be proper. For $r > 0$, there is a measure $\mu_r$ on $\mathbb{Q}$ that is defined by $\mu_r(\{q\}) = r$ for all $q \in \mathbb{Q}$. We clearly have that $\mu_r(B) = \rho(B)$ for all $B \in \mathcal{B}$. By the universal property of left extensions it follows that $\mu \leq \mu_r$, i.e. for all $r > 0$ and $q \in \mathbb{Q}$, we have

$$\mu(\{q\}) \leq r.$$

This implies that $\mu = 0$, which is clearly a contradiction.

**Remark 4.7.8.** It follows from the $\pi$-$\lambda$ theorem (Theorem 1.19 in [40]), that a *finite* measure is determined by its values on a generating algebra of subsets. This means that if there exist a proper extension, it has to be unique. By the Carathéodory extension theorem (Theorem 4.7.5), we know that such an extension exist.

It follows now that for *finite* premeasures, if the left Caratheéodory extension exist it is necessarily proper by Proposition 4.3.9 and therefore it has to be equal to the right extension.

# Chapter 5

# A categorical treatment of the Radon–Nikodym theorem and martingales

## 5.1 Introduction

The Radon–Nikodym theorem gives a correspondence between random variables and measures, two important concepts in probability theory and measure theory. If we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then we can look at the *random variables* on this probability space and the measures that are *absolutely continuous* with respect to $\mathbb{P}$. The Radon–Nikodym theorem tells us that there is a canonical correspondence between these two. The random variable associated to a measure $\mu$ that is absolutely continuous with respect to $\mathbb{P}$ is called the *Radon–Nikodym derivative with respect to* $\mathbb{P}$. The classical proof gives a concrete construction of these Radon–Nikodym derivatives and relies on the Hahn decomposition theorem, e.g. Theorem 31.B in [31], Theorem 4.2.2 in [11] and Theorem 3.2.2 in [6]. Important applications of this result are the existence of conditional expectations and the Girsanov theorem in stochastic calculus.

In Section 5.2 of this chapter, we will give a *categorical proof* for this result. Moreover, we will not only prove that there is a bijection between random variables and measures, but also that this bijection is an *isometry.* We will do this by starting with the trivial case, when $\Omega$ is finite. We translate this categorically as a natural isomorphism between certain functors. We will then proceed by *Kan extending* the trivial, finite version of the result to the general result. This happens in two parts. The first part (Section 5.2.2) is straightforward and purely categorical, not relying on any results in measure theory. The second part of the proof (Section 5.2.3) does require some measure theory, in particular it relies on the Riesz–Fischer theorem (Theorem 5.2.1). Furthermore, the concept of conditional expectation naturally arises from the Kan extension construction in Section 5.2.3.

In Section 5.3, we will focus on *martingales*, a special class of stochastic processes. Important examples of martingales are Brownian motion and unbiased random walks. Furthermore, martingales have nice convergence properties, which are described by *Doob's martingale convergence theorem.* The proof of this result relies on stopping times, the optional stopping theorem and several lemmas about upcrossings by stochastic processes. The original proof can be found in Section XI.14 in [13]. We will give a categorical proof of a weaker version of this result. We do this by showing that a certain class of functors preserve certain cofiltered limits (Theorem

5.3.15). By applying this to the functors representing random variables from Section 5.3.3, we immediately obtain a proof for a weaker martingale convergence theorem. Moreover, if we apply the same result to the functor representing measures from Section 5.3.3, we find a Kolmogorov extension-type theorem. In this section we use the results from Section 5.2. However, we first lift everything to the *enriched* setting. We consider everything to be enriched over the category of *complete metric spaces*. This part is crucial to obtain the main result in this section (Theorem 5.3.15).

In this chapter we will consider every distance function to be an **extended pseudometric**, i.e. a function $d : X \times X \to [0, \infty]$ on a set $X$ such that for all $x, y \in X$, $d(x, x) = 0$ and $d(x, y) = d(y, x)$, satisfying the triangle-inequality. However,we will refer to an extended pseudometric $d$ as just **metric** and to the pair $(X, d)$ as just **metric space**. The category of these metric spaces with the 1-Lipschitz maps between them is denoted by **Met** and the full subcategory of *complete* metric spaces by **CMet**.

For a measurable map $f : X \to Y$ and a measure $\mu$ on $X$, we write $\mu \circ f^{-1}$ to mean the **pushforward measure** of $\mu$ along $f$, i.e. for every measurable subset $E$ of $Y$,

$$\left(\mu \circ f^{-1}\right)(E) := \mu(f^{-1}(E)).$$

## 5.2 The Radon–Nikodym theorem

The Radon–Nikodym theorem is an important result in measure theory and probability theory. The goal of this section is to give a *categorical* proof of this result. We will start this section by recalling the result and explaining its applications in probability theory. We will then discuss the trivial *finite* version of the Radon–Nikodym theorem and explain how this translates categorically. After that, we focus on '*Kan extending*' the trivial finite result to the general result.

The Radon-Nikodym gives a connection between measures and integrable functions.

Let $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that two measurable functions $f, g : \Omega \to [0, \infty)$ are **$\mathbb{P}$-almost surely** equal if $\mathbb{P}(f = g) = 1$ and we write $f =_{\mathbb{P}} g$. This defines an equivalence relation on the set $\{f \in \mathbf{Mble}(\Omega, [0, \infty)) \mid \mathbb{E}(f) < \infty\}$ of measurable functions $X \to [0, \infty)$. The set

$$\{f \in \mathbf{Mble}(\Omega, [0, \infty)) \mid \mathbb{E}(f) < \infty\} / =_{\mathbb{P}}$$

becomes a metric space, by endowing it with the **$L^1$-metric** defined by

$$d_{L^1}(f, g) := \int |f - g| \mathrm{d}\mathbb{P},$$

for all $f, g \in \mathbf{Mble}(\Omega, [0, \infty)) / =_{\mathbb{P}}$. We will denote this space of random variables by $\mathrm{RV}(\mathbf{\Omega})$. For a real number $r > 0$, let $\mathrm{RV}_r(\mathbf{\Omega})$ be the subspace of random variables $f$ such that $\mathbb{P}(f \leq r) = 1$.

An important result about the space of random variables that we will need later is the Riesz–Fischer theorem.

**Theorem 5.2.1** (Riesz–Fischer)**.** *Let* $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $r > 0$ *be a real number. The metric spaces* $\mathrm{RV}(\mathbf{\Omega})$ *and* $\mathrm{RV}_r(\mathbf{\Omega})$ *are complete.*

*Proof.* This is Theorem 2.2 in [58]. □

For a measure $\mu$ on a measurable space $(\Omega, \mathcal{F})$, we say that $\mu$ is **absolutely continuous with respect to** $\mathbb{P}$ if $\mu(A) = 0$ for all measurable subsets $A \subseteq \Omega$ such that $\mathbb{P}(A) = 0$. This is denoted as $\mu \ll \mathbb{P}$. The set of measures on $(\Omega, \mathcal{F})$ that are absolutely continuous with respect to

$\mathbb{P}$ becomes a metric space, by endowing it with the **total variation metric** defined by

$$\begin{aligned} d_{\mathrm{TV}}(\mu,\nu) &:= |\mu-\nu|(\Omega) \\ &= \sup\left\{ \sum_{n=1}^{\infty} |\mu(A_n)-\nu(A_n)| \;\middle|\; \dot{\bigcup_{n=1}^{\infty}} A_n = \Omega \right\}, \end{aligned}$$

for all $\mu,\nu \ll \mathbb{P}$. We denote this space of measures by $\mathrm{M}(\mathbf{\Omega})$. For a real number $r>0$, we write $\mathrm{M}_r(\mathbf{\Omega})$ for the subspace of measures $\mu$ on $(\Omega,\mathcal{F})$ such that $\mu \le r\mathbb{P}$, i.e. $\mu(A) \le r\mathbb{P}(A)$ for all $A \in \mathcal{F}$.

**Remark 5.2.2.** Using the Hahn decomposition of the signed measure $\mu-\nu$, we can see that

$$d_{\mathrm{TV}}(\mu,\nu) = \sup\{|\mu(A)-\nu(A)| + |\mu(A^{\mathrm{C}})-\nu(A^{\mathrm{C}})| \mid A \in \mathcal{F}\}.$$

In the case that $\mu(\Omega)=\nu(\Omega)$, we have that $d_{\mathrm{TV}}(\mu,\nu) = 2\sup\{|\mu(A)-\nu(A)| \mid A \in \mathcal{F}\}$.

We have the following well-known result, for which we will give a short proof. A proof can also be found in Section III.7.4 in [15].

**Proposition 5.2.3.** *Let* $\mathbf{\Omega} := (\Omega,\mathcal{F},\mathbb{P})$ *be a probability space and let* $r>0$ *be a real number. The metric spaces* $\mathrm{M}(\mathbf{\Omega})$ *and* $\mathrm{M}_r(\mathbf{\Omega})$ *are complete.*

*Proof.* Let $(\mu_n)_n$ be a Cauchy sequence in $\mathrm{M}(\mathbf{\Omega})$. Since $|\mu_p(A)-\mu_q(A)| \le d_{\mathrm{TV}}(\mu_p,\mu_q)$ for all $A \in \mathcal{F}$, it follows that $(\mu_n(A))_n$ is a Cauchy sequence in $[0,\infty)$. Define a map $\mu : \mathcal{F} \to [0,\infty)$ by

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

It is clear that $\mu$ is finitely additive. Let $(A_n)_{n=1}^{\infty}$ be a decreasing sequence of measurable subsets such that $A_n \downarrow \emptyset$.

For $\epsilon>0$, there is an $N$ such that $d_{\mathrm{TV}}(\mu_p,\mu_q) < \frac{\epsilon}{2}$ for all $p,q \ge N$. Since $\mu_N$ is $\sigma$-additive, there is an $M$ such that for $m \ge M$, $\mu_N(A_m) < \frac{\epsilon}{2}$. Therefore, for $n \ge N$ and $m \ge M$, we have

$$\mu_n(A_m) \le \mu_N(A_m) + d_{\mathrm{TV}}(\mu_n,\mu_N) \le \epsilon.$$

Taking $n\to\infty$, shows that $\mu(A_m) < \epsilon$ and therefore $\lim_{m\to\infty}\mu(A_m)=0$. We conclude that $\mu$ is $\sigma$-additive.

We finish the proof by showing that $(\mu_n)_n$ converges to $\mu$ in $\mathrm{M}(\mathbf{\Omega})$. Let $K \ge 1$ be a natural number. Then for every countable partition $(A_k)_{k=1}^{\infty}$ and $p,q \ge N$,

$$\sum_{k=1}^{K} |\mu_p(A_k)-\mu_q(A_k)| \le d_{\mathrm{TV}}(\mu_p,\mu_q) < \frac{\epsilon}{2}.$$

Taking $p\to\infty$ first and then $K\to\infty$ gives that

$$\sum_{k=1}^{\infty} |\mu(A_k)-\mu_q(A_k)| < \frac{\epsilon}{2},$$

for every countable partition $(A_k)_{k=1}^{\infty}$ and $q \ge N$.

By taking the supremum over all countable partitions, we find that

$$d_{\mathrm{TV}}(\mu,\mu_q) < \frac{\epsilon}{2}$$

for all $q \geq N$. This shows that $\mu_n \to \mu$ in $\mathrm{M}(\mathbf{\Omega})$. $\square$

The space of random variables and the space of measures on $\mathbf{\Omega}$ are connected in the following way. For a random variable $f \in \mathrm{RV}(\mathbf{\Omega})$, define a measure $\varphi(f)$ on $(\Omega, \mathcal{F})$ by the assignment

$$A \mapsto \int_A f \mathrm{d}\mathbb{P}.$$

If $\mathbb{P}(A) = 0$, then $\varphi(f)(A) = 0$ and therefore $\varphi(f) \in \mathrm{M}(\mathbf{\Omega})$. Moreover for $f, g \in \mathrm{RV}(\mathbf{\Omega})$ and a measurable partition $(A_n)_{n=1}^{\infty}$ of $\Omega$,

$$\sum_{n=1}^{\infty} \left| \int_{A_n} f \mathrm{d}\mathbb{P} - \int_{A_n} g \mathrm{d}\mathbb{P} \right| \leq \sum_{n=1}^{\infty} \int_{A_n} |f - g| \mathrm{d}\mathbb{P} = \int |f - g| \mathrm{d}\mathbb{P}.$$

Taking the supremum over all such partitions gives that $d_{\mathrm{TV}}(\varphi(f), \varphi(g)) \leq d_{L^1}(f, g)$. Therefore the map $\varphi : \mathrm{RV}(\mathbf{\Omega}) \to \mathrm{M}(\mathbf{\Omega})$ is 1-Lipschitz.

**Theorem 5.2.4** (Radon–Nikodym)**.** *The map $\varphi$ is an isomorphism of metric spaces. In particular, for every measure $\mu$ such that $\mu \ll \mathbb{P}$ there exists a $\mathbb{P}$-almost surely unique measurable map $f : \Omega \to [0, \infty)$ such that for all $A \in \mathcal{F}$,*

$$\mu(A) = \int_A f \, d\mathbb{P}.$$

The $f$ in Theorem 5.2.4 is called the **Radon–Nikodym derivative of $\mu$ with respect to** $\mathbb{P}$ and is denoted as $\frac{\mathrm{d}\mu}{\mathrm{d}\mathbb{P}}$.

The following example is an application of the Radon–Nikodym theorem in probability theory. The existence of *conditional expectation* can be proven using this result. The concept of conditional expectation is important in *martingale* theory, which we will discuss further in Section 5.3. This is a slight generalization from the concept of conditional expectation with respect to a $\sigma$-algebra discussed in Section 1.1.2.

**Example 5.2.5** (Conditional expectation)**.** Consider two probability spaces $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ and let $g : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ be a measure-preserving map, i.e. $\mathbb{P}_1 \circ g^{-1} = \mathbb{P}_2$.

For a random variable $f \in \mathrm{RV}(\mathbf{\Omega}_1)$, we can define a measure $\mu$ on $\mathbf{\Omega}_2$ as

$$\mu(B) = \int_{g^{-1}(B)} f \mathrm{d}\mathbb{P}_1$$

for all $B \in \mathcal{F}_2$. If $\mathbb{P}_2(B) = 0$, then $\mathbb{P}_1(g^{-1}(B)) = 0$ and therefore $\mu \in \mathrm{M}(\mathbf{\Omega}_2)$.

By the Radon–Nikodym theorem (Theorem 5.2.4), there exists a unique $\tilde{f} \in \mathrm{RV}(\mathbf{\Omega}_2)$ such that

$$\int_{g^{-1}(B)} f \mathrm{d}\mathbb{P}_1 = \mu(B) = \int_B \tilde{f} \mathrm{d}\mathbb{P}_2. \tag{5.1}$$

The random variable $\tilde{f}$ is called the **conditional expectation of $f$ with respect to** $g$ and is denoted as $\mathbb{E}[f \mid g]$. Because of the uniqueness in the Radon–Nikodym theorem, equation (5.1) is the defining property for the random variable $\mathbb{E}[f \mid g]$.

### 5.2.1 The finite Radon–Nikodym theorem

In the case that we are working with a *finite* probability space, the Radon–Nikodym theorem becomes trivial. We will discuss this trivial version in this section and explain how this can be expressed categorically. To do this we will define two functors, one expressing random variables and one expressing measures. The finite version of the Radon–Nikodym theorem then corresponds to saying that these functors are isomorphic.

A **finite probability space** is a probability space whose underlying set $A$ is finite and whose $\sigma$-algebra is the whole powerset $\mathcal{P}(A)$. We will write $(A, p)$ instead of $(A, \mathcal{P}(A), p)$. For an element $a \in A$, we write $p_a$ or $p(a)$ to mean $p(\{a\})$.

We denote the category of probability spaces and measure-preserving maps by **Prob** and the full subcategory of finite probability spaces by $\mathbf{Prob}_f$. The inclusion functor $\mathbf{Prob}_f \to \mathbf{Prob}$ is denoted by $i$.

We start by defining a functor of measures $\mathrm{M}^f : \mathbf{Prob}_f \to \mathbf{CMet}$. This functor sends a finite probability space $(A, p)$ to $\mathrm{M}(A, p)$ and a measure-preserving map $s : (A, p) \to (B, q)$ of finite probability spaces to the 1-Lipschitz map $\mathrm{M}^f(s) : \mathrm{M}(A, p) \to \mathrm{M}(B, q)$, which is defined by the assignment

$$m \mapsto m \circ s^{-1}.$$

Similarly we can define a functor $\mathrm{M}^f_r : \mathbf{Prob}_f \to \mathbf{CMet}$ for every positive real number $r$.

We can also define a functor of random variables $\mathrm{RV}^f : \mathbf{Prob}_f \to \mathbf{CMet}$ in the following way. On objects this functor is defined by sending a finite probability space $(A, p)$ to the metric space $\mathrm{RV}(A, p)$. On morphisms this functor sends a measure-preserving map $s : (A, p) \to (B, q)$ of finite probability spaces to the 1-Lipschitz map $\mathrm{RV}^f(s) : \mathrm{RV}(A, p) \to \mathrm{RV}(B, q)$ which is defined for $g \in \mathrm{RV}(A, p)$ by

$$\mathrm{RV}^f(s)(g) : (B, q) \to [0, \infty) : b \mapsto \begin{cases} \frac{1}{q_b} \sum_{s(a)=b} p_q g(a) & \text{if } q_b \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The map $\mathrm{RV}^f(s)(g)$ does not depend on the representation of $g$ and therefore it is well-defined. In a similar way we define functors $\mathrm{RV}^f_r : \mathbf{Prob}_f \to \mathbf{CMet}$ for real numbers $r > 0$.

For a finite probability space $(A, p)$, we define the 1-Lipschitz map

$$\begin{array}{rccc} (\rho^f_r)_A : & \mathrm{RV}^f_r(A, p) & \to & \mathrm{M}^f_r(A, p) \\ & g & \mapsto & (g(a)p_a)_{a \in A}. \end{array}$$

The finite version of the Radon–Nikodym theorem can now be expressed in the following way.

**Proposition 5.2.6** (Finite bounded Radon–Nikodym)**.** *The maps* $((\rho^f_r)_A)_{(A,p)}$ *form a natural isomorphism* $\rho^f_r : \mathrm{RV}^f_r \to \mathrm{M}^f_r$.

*Proof.* It is easy to see that $(\rho^f_r)_A$ is well-defined and invertible for every finite probability space $(A, p)$. If follows now by Proposition 4.2 in [46] that this is an isomorphism of metric spaces. It is straightforward to check the naturality of $\rho^f_r$. □

Since $\mathrm{RV}^f_r$ and $\mathrm{M}^f_r$ are isomorphic by Proposition 5.2.6, so are their right Kan extensions along the inclusion $i : \mathbf{Prob}_f \to \mathbf{Prob}$.

$$\begin{tikzcd}
\mathbf{Prob}_f \arrow[r, bend left, "\mathrm{M}^f_r"{name=U}] \arrow[r, bend right, "\mathrm{RV}^f_r"'{name=D}] \arrow[d, "i"'] \arrow[r, phantom, "\cong"] & \mathbf{CMet} \\
\mathbf{Prob} \arrow[ur, dashed] \arrow[ur, dashed, bend right] \arrow[ur, phantom, bend right=15, "\cong"] &
\end{tikzcd}$$

In the following two sections, we will study the right Kan extensions of these functors. In Section 5.2.2 we will describe what the right Kan extension of the finite measures functor $\mathrm{M}_r^f : \mathbf{Prob}_f \to \mathbf{CMet}$ along $i : \mathbf{Prob}_f \to \mathbf{Prob}$ looks like and in Section 5.2.3 we will do the same for the finite random variables functor $\mathrm{RV}_r^f$.

This will lead to a categorical proof for the bounded Radon–Nikodym theorem in Section 5.2.4.

### 5.2.2 The measures functor M

In this section we will study the right Kan extension of the functor $\mathrm{M}_r^f : \mathbf{Prob}_f \to \mathbf{CMet}$ along the inclusion functor $i : \mathbf{Prob}_f \to \mathbf{Prob}$.

$$\begin{tikzcd}
\mathbf{Prob}_f \arrow[r, "\mathrm{M}_r^f"] \arrow[d, "i"'] & \mathbf{CMet} \\
\mathbf{Prob} \arrow[ur, dashed, "\mathrm{Ran}_i\mathrm{M}_r^f"'] &
\end{tikzcd}$$

We will first describe how $\mathrm{Ran}_i\mathrm{M}_r^f$ acts on objects in Theorem 5.2.7 and then how it acts on morphisms in Proposition 5.2.8. It will turn out that this right Kan extension expresses certain measures on arbitrary probability spaces. This will then motivate the notation $\mathrm{M}_r := \mathrm{Ran}_i\mathrm{M}_r^f$.

We will then show that these functors form a diagram $D_{\mathrm{M}}$:

$$\begin{tikzcd}
\dots \arrow[r] & \mathrm{M}_1 \arrow[r] & \dots \arrow[r] & \mathrm{M}_2 \arrow[r] & \dots \arrow[r] & \mathrm{M}_r \arrow[r] & \dots
\end{tikzcd}$$

In the second part of this section we will describe the colimit of this diagram. The obtained colimiting functor will express measures and therefore this will motivate the notation $\mathrm{M} : \mathbf{Prob} \to \mathbf{CMet}$ for the colimit of $D_{\mathrm{M}}$.

**Theorem 5.2.7.** *Let* $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then*

$$\mathrm{Ran}_i\mathrm{M}_r^f(\mathbf{\Omega}) = \mathrm{M}_r(\mathbf{\Omega}).$$

*Proof.* Let $U : \mathbf{\Omega} \downarrow i \to \mathbf{Prob}_f$ be the forgetful functor and let $D_{\mathbf{\Omega}}$ denote the diagram

$$\mathbf{\Omega} \downarrow i \xrightarrow{U} \mathbf{Prob}_f \xrightarrow{\mathrm{M}_r^f} \mathbf{CMet}.$$

We will now show that $\mathrm{M}_r(\mathbf{\Omega}) = \lim D_{\mathbf{\Omega}}$.

For a measure-preserving map $g : \mathbf{\Omega} \to \mathbf{A}$, where $\mathbf{A} := (A, p)$ is some finite probability space, define a map $p_g : \mathrm{M}_r(\mathbf{\Omega}) \to \mathrm{M}_r^f(\mathbf{A})$ by

$$p_g(\mu) := \mu \circ g^{-1}.$$

It can be checked that this map is well-defined and 1-Lipschitz. It is also straightforward to check that the metric space $\mathrm{M}_r(\mathbf{\Omega})$ together with the maps $(p_f)_f$ form a cone over the diagram $D_{\mathbf{\Omega}}$.

We will now show that this cone is universal. To do that, consider another cone $(Y, (q_f)_f))$ over the diagram $D_{\mathbf{\Omega}}$.

Let $E$ be a measurable subset of $\mathbf{\Omega}$. Let $\mathbf{2} := \{0, 1\}$ and let $p_E$ be the probability measure on $\mathbf{2}$ defined by

$$p_E(1) := \mathbb{P}(E).$$

The assignment

$$\omega \mapsto \begin{cases} 1 \text{ if } \omega \in E \\ 0 \text{ otherwise,} \end{cases}$$

defines a measure-preserving map $1_E : \boldsymbol{\Omega} \to (\mathbf{2}, p_E)$.

For $y \in Y$, define a map $\mu_y : \mathcal{F} \to [0, \infty)$ by

$$\mu_y(A) := q_{1_A}(y)(1).$$

since $q_{1_E}(y)$ is an element of $\mathrm{M}^f_r(\mathbf{2}, p_E)$,

$$\mu_y(E) = q_{1_E}(y)(1) \leq r p_E(1) = r\mathbb{P}(E)$$

and therefore $\mu_y \leq r\mathbb{P}$.

We will now show that $\mu_y$ is a measure. Consider disjoint measurable subsets $E_1$ and $E_2$ of $\Omega$. Let $\mathbf{3} := \{0, 1, 2\}$ and let $p_{E_1,E_2}$ be the probability measure on $\mathbf{3}$, defined by

$$p_{E_1,E_2}(1) := \mathbb{P}(E_1) \quad \text{and} \quad p_{E_1,E_2}(2) := \mathbb{P}(E_2).$$

The assignment

$$\omega \mapsto \begin{cases} 1 \text{ if } \omega \in E_1 \\ 2 \text{ if } \omega \in E_2 \\ 0 \text{ otherwise,} \end{cases}$$

defines a measure preserving map $1_{E_1,E_2} : \boldsymbol{\Omega} \to (\mathbf{3}, p_{E_1,E_2})$. Let $s : \mathbf{3} \to \mathbf{2}$ be the map that fixes 0. The map $\mathbf{3} \to \mathbf{2}$ that sends 1 to 1 and the other elements to 0 is denoted by $s_1$. The map $s_2 : \mathbf{3} \to \mathbf{2}$ is defined in a similar way.

We have the following commutative triangles:

$$\begin{tikzcd}
& (\Omega, \mathcal{F}, \mathbb{P}) \arrow[dl, "1_{E_2}"'] \arrow[d, "1_{E_1,E_2}"] \arrow[dr, "1_{E_1}"] & \\
(\mathbf{2}, p_{E_2}) & (\mathbf{3}, p_{E_1,E_2}) \arrow[l, "s_2"] \arrow[r, "s_1"'] & (\mathbf{2}, p_{E_1})
\end{tikzcd}
\qquad
\begin{tikzcd}
& (\Omega, \mathcal{F}, \mathbb{P}) \arrow[dl, "1_{E_1 \cup E_2}"'] \arrow[dr, "1_{E_1,E_2}"] & \\
(\mathbf{2}, p_{E_1 \cup E_2}) & & (\mathbf{3}, p_{E_1,E_2}) \arrow[ll, "s"]
\end{tikzcd}$$

Because $(Y, (q_f)_f)$ is a cone over the diagram $D_{\boldsymbol{\Omega}}$, we also have the following commutative triangles:

$$\begin{tikzcd}
& Y \arrow[dl, "q_{1_{E_2}}"'] \arrow[d, "q_{1_{E_1,E_2}}"] \arrow[dr, "q_{1_{E_1}}"] & \\
\mathrm{M}^f_r(\mathbf{2}, p_{E_2}) & \mathrm{M}^f_r(\mathbf{3}, p_{E_1,E_2}) \arrow[l, "\mathrm{M}^f_r(s_2)"] \arrow[r, "\mathrm{M}^f_r(s_1)"'] & \mathrm{M}^f_r(\mathbf{2}, p_{E_1})
\end{tikzcd}
\qquad
\begin{tikzcd}
& Y \arrow[dl, "q_{1_{E_1 \cup E_2}}"'] \arrow[dr, "q_{1_{E_1,E_2}}"] & \\
\mathrm{M}^f_r(\mathbf{2}, p_{E_1 \cup E_2}) & & \mathrm{M}^f_r(\mathbf{3}, p_{E_1,E_2}) \arrow[ll, "\mathrm{M}^f_r(s)"]
\end{tikzcd}$$

Using the above diagrams we find

$$\mu_y(E_1 \cup E_2) = q_{1_{E_1 \cup E_2}}(y)_1 = q_{1_{E_1,E_2}}(y)_1 + q_{1_{E_1,E_2}}(y)_2 = q_{1_{E_1}}(y)_1 + q_{1_{E_2}}(y)_1 = \mu_y(E_1) + \mu_y(E_2),$$

which shows that $\mu_y$ is finitely additive.

Let $(E_n)_n$ be a sequence of measurable subsets of $\mathbf{\Omega}$ that decreases to $\emptyset$. Because $0 \leq \mu_y(E_n) \leq r\mathbb{P}(E_n)$, also

$$0 \leq \lim_n \mu_y(E_n) \leq r \lim_n \mathbb{P}(E_n) = 0.$$

We can conclude that $\mu_y$ is an element of $\mathrm{M}_r(\mathbf{\Omega})$. The assignment $y \mapsto \mu_y$ defines a map $q : Y \to \mathrm{M}_r(\mathbf{\Omega})$. For $y_1, y_2$ in $Y$, we find for every measurable subset $A$ that

$$\begin{aligned} |\mu_{y_1}(A) - \mu_{y_2}(A)| + |\mu_{y_1}(A^{\mathrm{C}}) - \mu_{y_2}(A^{\mathrm{C}})| &= |q_{1_A}(y_1)(1) - q_{1_A}(y_2)(1)| + |q_{1_{A^{\mathrm{C}}}}(y_1)(0) - q_{1_{A^{\mathrm{C}}}}(y_2)(0)| \\ &\leq d_{\mathrm{TV}}(q_{1_A}(y_1), q_{1_A}(y_2)) \leq d_Y(y_1, y_2). \end{aligned}$$

Taking the supremum over all $A \in \mathcal{F}$, shows us that $q$ is 1-Lipschitz. Let $(A, p)$ be a finite probability space and let $g : \mathbf{\Omega} \to (A, p)$ be a measure-preserving map. It follows that

$$p_g(\mu_y)(a) = \mu_y \circ g^{-1}(a) = q_{1_{(\{a\} \circ g)}}(y)(1) = q_g(y) \circ 1^{-1}_{\{a\}}(1) = q_g(y)(a)$$

for all $a \in A$ and $y \in Y$. This means that $q$ is a morphism of cones. Furthermore, this morphism of cones is unique. We can now conclude that $\mathrm{M}_r(\mathbf{\Omega}) = \mathrm{Ran}_i \mathrm{M}_r(\mathbf{\Omega})$. ☐

We have just described how the functor $\mathrm{Ran}_i \mathrm{M}_r^f$ behaves on objects. In the following proposition we will study how it acts on morphisms.

**Proposition 5.2.8.** *For a measure-preserving map of probability spaces $f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$,*

$$\mathrm{M}_r(f)(\mu) = \mu \circ f^{-1},$$

*for all $\mu \in \mathrm{M}_r(\mathbf{\Omega}_1)$.*

*Proof.* By the universal property of right Kan extensions, we know that $M_r(f) : M_r(\mathbf{\Omega}_1) \to M_r(\mathbf{\Omega}_2)$ is the unique morphism such that

$$\begin{tikzcd} \mathrm{M}_r(\mathbf{\Omega}_1) \arrow[rr, "\mathrm{M}_r(f)"] \arrow[dr, "p^1_{hf}"'] & & \mathrm{M}_r(\mathbf{\Omega}_2) \arrow[dl, "p^2_f"] \\ & \mathrm{M}_r^f(A, p) & \end{tikzcd}$$

commutes for every measure-preserving map $h : \mathbf{\Omega}_2 \to (A, p)$, where $(A, p)$ is some finite probability space. Here $p^1_{hf}$ and $p^2_f$ are the projection maps defined in the proof of Theorem 5.2.7. It is clear that the map $\mathrm{M}_r(\mathbf{\Omega}_1) \to \mathrm{M}_r(\mathbf{\Omega}_2)$ defined by the assignment $\mu \mapsto \mu \circ f^{-1}$ satisfies this property and therefore the claim follows. ☐

Theorem 5.2.7 and Proposition 5.2.8 tell us that the functor $\mathrm{Ran}_i \mathrm{M}_r^f : \mathbf{Prob} \to \mathbf{CMet}$ expresses measures. We will therefore use the notation $\mathrm{M}_r := \mathrm{Ran}_i \mathrm{M}_r^f$ from now on.

Furthermore, note that in the proofs of Theorem 5.2.7 and Proposition 5.2.8 we have not used any non-trivial measure-theoretic results. The proofs in this section are straightforward categorical proofs.

For $r \leq s$, there is a natural transformation $\mathrm{M}_r^f \to \mathrm{M}_s^f$ and therefore a natural transformation $\mathrm{M}_r \to \mathrm{M}_s$. This natural transformation is given by the inclusion maps

$$\mathrm{M}_r(\mathbf{\Omega}) \to \mathrm{M}_s(\mathbf{\Omega}),$$

for all probability spaces $\mathbf{\Omega}$. This gives us a diagram

$$D_{\mathrm{M}} : (0, \infty) \to [\mathbf{Prob}, \mathbf{CMet}]$$

of functors and natural transformations

$$\ldots \longrightarrow M_1 \longrightarrow \ldots \longrightarrow M_2 \longrightarrow \ldots \longrightarrow M_r \longrightarrow \ldots .$$

In the rest of this section, we will study what the colimit of this diagram looks like. The next proposition tells us how the colimiting functor acts on objects.

**Proposition 5.2.9.** *Let $\mathbf{\Omega}$ be a probability space, then* $\mathrm{colim} D_{\mathrm{M}}(\mathbf{\Omega}) = \mathrm{M}(\mathbf{\Omega})$.

*Proof.* Consider the subset

$$S := \{\mu \in \mathrm{M}(\mathbf{\Omega}) \mid \exists r > 0 : \mu \leq r\mathbb{P}\}.$$

It is enough to show that $S$ is dense in $\mathrm{M}(\mathbf{\Omega})$. For $\mu \in \mathrm{M}(\mathbf{\Omega})$, define

$$\mu_n := \mu \wedge n\mathbb{P},$$

For every $n$, let $(P_n, N_n)$ be the Hahn decomposition of the signed measure $\mu - n\mathbb{P}$. It is clear that $(P_n)_n$ decreases and let $P := \bigcap_{n=1}^{\infty} P_n$. We have that

$$\mu_n(N_n) = \inf\left\{\sum_{m=1}^{\infty} \mu(E_m) \wedge n\mathbb{P}(E_m) \mid \dot{\bigcup_{m=1}^{\infty}} E_m = N_n\right\} = \mu(N_n).$$

It follows that for every $n \geq 1$

$$d_{\mathrm{TV}}(\mu, \mu_n) = \mu(\Omega) - \mu_n(\Omega) = \mu(P_n) - \mu_n(P_n) \leq \mu(P_n).$$

Furthermore, for every $n \geq 1$, we have that

$$\infty > \mu(P) \geq \lim_{n \to \infty} n\mathbb{P}(P_n) \geq \lim_{n \to \infty} n\mathbb{P}(P).$$

This is only possible when $\mathbb{P}(P) = 0$. Since $\mu \ll \mathbb{P}$, it follows that $\mu(P) = 0$. Therefore,

$$\lim_{m \to \infty} d_{\mathrm{TV}}(\mu, \mu_n) = \lim_{n \to \infty} \mu(P_n) = \mu(P) = 0.$$

Since $\mu_n \in S$ for every $S$ and $\mu_n \to \mu$ as $n \to \infty$, the claim follows. □

We will now discuss what the colimit of $D_{\mathrm{M}}$ does on morphisms.

**Proposition 5.2.10.** *Let $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces and let $f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ be a measure-preserving map. Then* $(\mathrm{colim} D_{\mathrm{M}})(f)$ *is the* 1*-Lipschitz map* $\mathrm{M}(\mathbf{\Omega}_1) \to \mathrm{M}(\mathbf{\Omega}_2)$ *defined by*

$$\mu \mapsto \mu \circ f^{-1}.$$

*Proof.* The map $(\mathrm{colim} D_{\mathrm{M}})(f)$ is the unique map $\mathrm{M}(\mathbf{\Omega}_1) \to \mathrm{M}(\mathbf{\Omega}_2)$ such that the following diagram commutes for every natural number $r > 0$

$$\begin{tikzcd}
\mathrm{M}(\mathbf{\Omega}_1) \arrow[r, "(\mathrm{colim} D_{\mathrm{M}})(f)"] & \mathrm{M}(\mathbf{\Omega}_2) \\
\mathrm{M}_r(\mathbf{\Omega}_1) \arrow[r, "\mathrm{M}_r(f)"'] \arrow[u] & \mathrm{M}_r(\mathbf{\Omega}_2) \arrow[u]
\end{tikzcd}$$

The 1-Lipschitz map $\mathrm{M}(\mathbf{\Omega}_1) \to \mathrm{M}(\mathbf{\Omega}_2) : \mu \mapsto \mu \circ f^{-1}$ satisfies this condition and therefore it has to be equal to $(\mathrm{colim} D_{\mathrm{M}})(f)$. ☐

Proposition 5.2.9 and Proposition 5.2.10 tell us that the functor $\mathrm{colim} D_{\mathrm{M}}$ describes measures. Therefore we will from now on use the notation

$$\mathrm{M} := \mathrm{colim} D_{\mathrm{M}}.$$

### 5.2.3 The random variables functor RV

We will start this section by describing what the right Kan extension of the functor $\mathrm{RV}_r^f : \mathbf{Prob}_f \to \mathbf{CMet}$ along the functor $i : \mathbf{Prob}_f \to \mathbf{Prob}$ looks like.

$$\begin{tikzcd}
\mathbf{Prob}_f \arrow[r, "\mathrm{RV}_r^f"] \arrow[d, "i"'] & \mathbf{CMet} \\
\mathbf{Prob} \arrow[ur, dashed, "\mathrm{Ran}_i \mathrm{RV}_r^f"'] &
\end{tikzcd}$$

We will do this by first showing how $\mathrm{Ran}_i \mathrm{RV}_r^f$ acts on objects in Theorem 5.2.11 and then how it acts on morphisms in Proposition 5.2.13. The conclusion will be that this right Kan extension describes bounded random variables on arbitrary probability spaces. We will therefore introduce the notation $\mathrm{RV}_r$ to mean the functor $\mathrm{Ran}_i \mathrm{RV}_r^f : \mathbf{Prob} \to \mathbf{CMet}$.

We will proceed the section by showing that these functors form a diagram $D_{\mathrm{RV}}$:

$$\ldots \longrightarrow \mathrm{RV}_1 \longrightarrow \ldots \longrightarrow \mathrm{RV}_2 \longrightarrow \ldots \longrightarrow \mathrm{RV}_n \longrightarrow \ldots$$

In the remaining part of the section we will study the colimit of $D_{\mathrm{RV}}$. We will show that this colimiting functor describes random variables and we will therefore denote this functor as RV.

**Theorem 5.2.11.** *Let* $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then*

$$\mathrm{Ran}_i \mathrm{RV}_r^f(\mathbf{\Omega}) = \mathrm{RV}_r(\mathbf{\Omega}).$$

*Proof.* Let $U : \mathbf{\Omega} \downarrow i \to \mathbf{Prob}_f$ be the forgetful functor and let $D_{\mathbf{\Omega}}$ denote the diagram

$$\mathbf{\Omega} \downarrow i \xrightarrow{U} \mathbf{Prob}_f \xrightarrow{\mathrm{RV}_r^f} \mathbf{CMet}.$$

We will now show that $\mathrm{RV}_r(\mathbf{\Omega}) = \lim D_{\mathbf{\Omega}}$.

For a measure-preserving map $f : \mathbf{\Omega} \to (A, p)$ where $(A, p)$ is some finite probability space, define a map $p_f : \mathrm{RV}_r(\mathbf{\Omega}) \to \mathrm{RV}_r^f(A, p)$ by

$$p_f(g)(a) := \begin{cases} \frac{1}{p_a} \int_{f^{-1}(a)} g \mathrm{d}\mathbb{P} & \text{if } p_a \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

for every $g$ in $\mathrm{RV}_r(\mathbf{\Omega})$ and $a$ in $A$. It is straightforward to check that the definition of $p_f$ is independent of the choice of representative. It can also be checked that $p_f$ is 1-Lipschitz. Consider a commutative diagram

$$\begin{tikzcd} \mathbf{\Omega} \arrow[r, "f_1"] \arrow[dr, "f_2"'] & (A, p) \arrow[d, "s"] \\ & (B, q) \end{tikzcd}$$

Let $g$ in $\mathrm{RV}_r(\mathbf{\Omega})$ and $b$ in $B$, we have

$$\begin{aligned} \left[ \left( \mathrm{RV}_r^f(s) \circ p_{f_1}(g) \right) (b) \right] q_b &= \sum_{a \in s^{-1}(b)} p_a \left( [p_{f_1}(g)](a) \right) \\ &= \sum_{a \in s^{-1}(b)} \int_{f^{-1}(a)} g \mathrm{d}\mathbb{P} \\ &= \int_{f_2^{-1}(b)} g \mathrm{d}\mathbb{P}. \end{aligned}$$

It now follows that $(\mathrm{RV}_r(\mathbf{\Omega}), (p_f)_f)$ is a cone over the diagram $D_{\mathbf{\Omega}}$. We will now show that this cone is universal. To do that, we consider another cone $(Y, (q_f)_f)$ over the diagram $D_{\mathbf{\Omega}}$.

For $y \in Y$ and measure-preserving map $\mathbf{\Omega} \xrightarrow{f} (A, p)$, we define a simple function $\Omega \to [0, \infty)$ as follows:

$$s_f^y = \sum_{a \in A} ([q_f(y)](a)) 1_{f^{-1}(a)}.$$

Consider a commutative diagram

$$\begin{tikzcd} \mathbf{\Omega} \arrow[r, "f"] \arrow[dr, "g"'] & (A, p) \arrow[d, "s"] \\ & (B, r) \end{tikzcd}$$

We find for every $b \in B$ that

$$\int_{g^{-1}(b)} s_f^y \mathrm{d}\mathbb{P} = \sum_{s(a)=b} ([q_f(y)](a)) p_a = q_g(y)(b) r_b$$

It follows that $p_g(s_f^y) = q_g(y)$. Note that $\mathbf{\Omega} \downarrow i$ is cofiltered and therefore $(s_f^y)_f$ forms a net in $\mathrm{RV}_r(\mathbf{\Omega})$. Suppose now that $(s_f^y)_f$ has a limit $s^y$ in $\mathrm{RV}_r(\mathbf{\Omega})$. Then it is easy to see that $p_g(s^y) = q_g(y)$ for all $g \in \mathbf{\Omega} \downarrow i$. By the Riesz–Fischer theorem (Theorem 5.2.1) we only need to show that $(s_f^y)_f$ is a Cauchy net.

For this we will use the following two inequalities, which we will prove in Appendix C. For a commutative diagram

$$\begin{tikzcd}
\mathbf{\Omega} \arrow[r, "f"] \arrow[dr, "g"'] & (A,p) \arrow[d, "s"] \\
& (B,r)
\end{tikzcd}$$

we have

$$\mathbb{E}[\left(s_g^y\right)^2] \leq \mathbb{E}[(s_f^y)^2]. \tag{5.2}$$

and

$$0 \leq d_{L^1}(s_f^y, s_g^y)^2 \leq \mathbb{E}[(s_f^y - s_g^y)^2] = \mathbb{E}[(s_f^y)^2] - \mathbb{E}[(s_g^y)^2] \tag{5.3}$$

Now by inequality (5.2) we conclude that $\left(\mathbb{E}\left[(s_f^y)^2\right]\right)_f$ is a bounded, monotone net and therefore it converges. Using (5.3) we can now conclude that $(s_f^y)_f$ is a Cauchy net. The Riesz–Fischer theorem (Theorem 5.2.1) tells us that the net $(s_f^y)_f$ converges to some $s^y$ in $\mathrm{RV}_r(\mathbf{\Omega})$. This defines a map $Y \to \mathrm{RV}_r(\mathbf{\Omega})$. It can be checked that this is a 1-Lipschitz map. Because $p_g(s^y) = q_g(y)$ for every $g$ in $\mathbf{\Omega} \downarrow i$ it follows that this map is in fact a morphism of cones. Finally, it is straightforward to check that this is the unique morphism of cones. We can now conclude that $\mathrm{RV}_r(\mathbf{\Omega}) = \mathrm{Ran}_i\mathrm{RV}_r^f(\mathbf{\Omega})$. □

**Remark 5.2.12.** Note that the net $(s_f^y)_f$ in the proof of Theorem 5.2.11 could be interpreted as a martingale. The argument we used to show that this net convergences is similar to the proof of the martingale convergence theorem in [43].

We now know what the right Kan extension $\mathrm{Ran}_i\mathrm{RV}_r^f$ does on objects, but not yet how it acts on morphisms. This is described in the following proposition.

**Proposition 5.2.13.** *Let $f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ be a measure-preserving map of probability spaces. Let $g \in \mathrm{RV}_r(\mathbf{\Omega}_1)$. Then*

$$\mathrm{RV}_r(f)(g) = \mathbb{E}[g \mid f].$$

*Proof.* By the defining property of conditional expectation it is enough to show that

$$\mathbb{E}_{\mathbf{\Omega}_1}[g1_{f^{-1}(B)}] = \mathbb{E}_{\mathbf{\Omega}_2}[\mathrm{RV}_r(f)(g)1_B].$$

for all measurable subsets $B$ of $\mathbf{\Omega}_2$. Since $\mathrm{RV}_r$ is defined as $\mathrm{Ran}_i\mathrm{RV}_r^f$, $\mathrm{RV}_r(f)$ is the unique map $\mathrm{RV}_r(X) \to \mathrm{RV}_r(Y)$ such that the following diagram commutes for every measure-preserving map $h : \mathbf{\Omega}_2 \to (A,p)$, where $(A,p)$ is some finite probability space.

$$\begin{tikzcd}
\mathrm{RV}_r(\mathbf{\Omega}_1) \arrow[rr, "\mathrm{RV}_r(f)"] \arrow[dr, "p^1_{hf}"'] & & \mathrm{RV}_r(\mathbf{\Omega}_2) \arrow[dl, "p^2_h"] \\
& \mathrm{RV}_r^f(A,p) &
\end{tikzcd}$$

Here $p^1_{hf}$ and $p^2_f$ are the projection maps defined in the proof of Theorem 5.2.11. Let $\mathbf{2} := \{0,1\}$ and let $q$ be the probability measure on $\mathbf{2}$ defined by $q_1 := \mathbb{P}_Y(B)$. Consider the measure-preserving map

$$h : \mathbf{\Omega}_2 \to (\mathbf{2}, q)$$

defined by the assignment

$$h(\omega) := \begin{cases} 1 \text{ if } \omega \in B \\ 0 \text{ otherwise.} \end{cases}$$

We now find

$$\mathbb{E}_{\mathbf{\Omega}_1}[g 1_{f^{-1}(B)}] = q_1[p^1_{hf}(g)](1) = q_1[p^2_h(\mathrm{RV}_r(f)(g))](1) = \mathbb{E}_{\mathbf{\Omega}_2}[\mathrm{RV}_r(f)(g) 1_B].$$

This proves the claim.

□

For positive real numbers $r \leq s$, there is a natural transformation $\mathrm{RV}^f_r \to \mathrm{RV}^f_s$ and therefore a natural transformation $\mathrm{RV}_r \to \mathrm{RV}_s$. This natural transformation is given by the inclusion maps

$$\mathrm{RV}_r(\mathbf{\Omega}) \to \mathrm{RV}_s(\mathbf{\Omega}),$$

for all probability spaces $\mathbf{\Omega}$. This gives a diagram $D_{\mathrm{RV}} : (0, \infty) \to [\mathbf{Prob}, \mathbf{CMet}]$ of functors and natural transformations

$$\ldots \longrightarrow \mathrm{RV}_1 \longrightarrow \ldots \longrightarrow \mathrm{RV}_2 \longrightarrow \ldots \longrightarrow \mathrm{RV}_r \longrightarrow \ldots.$$

In the rest of this section, we will describe what the colimit of this diagram looks like. We will describe the colimiting functor's behaviour on objects and morphisms in the following two propositions.

**Proposition 5.2.14.** *Let $\mathbf{\Omega}$ be a probability space. Then $(\mathrm{colim} D_{\mathrm{RV}})(\mathbf{\Omega}) = \mathrm{RV}(\mathbf{\Omega})$.*

*Proof.* Consider the subset

$$S := \{f \in \mathrm{RV}(\mathbf{\Omega}) \mid \exists r > 0 : \mathbb{P}(f \leq r) = 1\}.$$

We will show that $S$ is dense in $\mathrm{RV}(\Omega)$. For $f \in \mathrm{RV}(\Omega)$, define

$$f_n := f \wedge n.$$

Clearly, $f_n \in S$ and by the Monotone Convergence Theorem, we have that $f_n \to f$ in $\mathrm{RV}(\mathbf{\Omega})$.

It follows now that for any complete metric space $Y$ and 1-Lipschitz map $f : S \to Y$, there is a unique 1-Lipschitz map $\tilde{f} : \mathrm{RV}(\mathbf{\Omega}) \to Y$. From this it follows that $\mathrm{colim}(D_{\mathrm{RV}}(\mathbf{\Omega})) = \mathrm{RV}(\mathbf{\Omega})$ and thus $(\mathrm{colim} D_{\mathrm{RV}})(\mathbf{\Omega}) = \mathrm{RV}(\mathbf{\Omega})$. □

We will end this section by showing how $\mathrm{colim} D_{\mathrm{RV}}$ acts on morphisms.

**Proposition 5.2.15.** *Let $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces and let $g : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ be a measure-preserving map. Then $(\mathrm{colim} D_{\mathrm{RV}})(g)$ is the 1-Lipschitz map $\mathrm{RV}(\mathbf{\Omega}_1) \to \mathrm{RV}(\mathbf{\Omega}_2)$ defined by*

$$f \mapsto \mathbb{E}[f \mid g].$$

*Proof.* The map $(\mathrm{colim} D_{\mathrm{RV}})(g)$ is the unique map $\mathrm{RV}(\mathbf{\Omega}_1) \to \mathrm{RV}(\mathbf{\Omega}_2)$ such that the following diagram commutes for every natural number $r > 0$:

$$\begin{tikzcd}
\mathrm{RV}(\mathbf{\Omega}_1) \arrow[r, "(\mathrm{colim} D_{\mathrm{RV}})(g)"] & \mathrm{RV}(\mathbf{\Omega}_1) \\
\mathrm{RV}_r(\mathbf{\Omega}_1) \arrow[u] \arrow[r, "\mathrm{RV}_r(g)"'] & \mathrm{RV}_r(\mathbf{\Omega}_1) \arrow[u]
\end{tikzcd}$$

From Proposition 5.2.13 it follows that the 1-Lipschitz map $\mathrm{RV}(\mathbf{\Omega}_1) \to \mathrm{RV}(\mathbf{\Omega}_2) : f \mapsto \mathbb{E}[f \mid g]$ satisfies this condition and therefore it has to be equal to $(\mathrm{colim}D_{\mathrm{RV}})(g)$. ☐

Proposition 5.2.14 and Proposition 5.2.15 tell us that the functor $\mathrm{colim}D_{\mathrm{RV}}$ describes random variables. Therefore we will from now on use the notation

$$\mathrm{RV} := \mathrm{colim}D_{\mathrm{RV}}.$$

### 5.2.4 The Radon–Nikodym theorem

We will now conclude Section 5.2 by giving a categorical proof of the Radon–Nikodym theorem. We will first look at a weaker bounded version (Theorem 5.2.17) and then extend this to the general version (Theorem 5.2.19). In Proposition 5.2.16 and Proposition 5.2.18, we give the concrete construction of the correspondence between random variables and measures that we obtain from the categorical proofs.

For the weaker bounded version of the Radon–Nikodym theorem, we will use the functors $\mathrm{M}_r$ and $\mathrm{RV}_r$ as defined in Section 5.2.2 and Section 5.2.3.

Recall that we have a natural transformation $\rho_r^f : \mathrm{RV}_r^f \to \mathrm{M}_r^f$. This induces a natural transformation $\mathrm{Ran}_i\rho_r^f : \mathrm{Ran}_i\mathrm{RV}_r^f \to \mathrm{Ran}_i\mathrm{M}_r^f$. This is a natural transformation $\mathrm{RV}_r \to \mathrm{M}_r$, which we will denote by $\rho_r$. In the following proposition, we will describe this natural transformation.

**Proposition 5.2.16.** *Let $\mathbf{\Omega}$ be a probability space. For $g \in \mathrm{RV}_r(\mathbf{\Omega})$ and $B$ a measurable subset of $\mathbf{\Omega}$, then $(\rho_r)_{\mathbf{\Omega}}$ is the map $\mathrm{RV}_r(\mathbf{\Omega}) \to \mathrm{M}_r(\mathbf{\Omega})$ defined by the assignment*

$$g \mapsto \int_{(-)} g d\mathbb{P}.$$

*Proof.* Since $\rho_r$ is defined as $\mathrm{Ran}_i\rho_r^f$, the map $(\rho_r)_{\mathbf{\Omega}}$ is the unique map $\mathrm{RV}_r(\mathbf{\Omega}) \to \mathrm{M}_r(\mathbf{\Omega})$ such that the following diagram commutes for every measure-preserving map $h : \mathbf{\Omega} \to (A, p)$, where $(A, p)$ is some finite probability space.

$$\begin{tikzcd}
\mathrm{RV}_r(\mathbf{\Omega}) \arrow[r, "(\rho_r)_{\mathbf{\Omega}}"] \arrow[d, "p_h^{\mathrm{RV}}"'] & \mathrm{M}_r(\mathbf{\Omega}) \arrow[d, "p_h^{\mathrm{M}}"] \\
\mathrm{RV}_r^f(A, p) \arrow[r, "(\rho_r^f)_A"] & \mathrm{M}_r^f(A, p)
\end{tikzcd}$$

Here $p_h^{\mathrm{RV}}$ and $p_h^{\mathrm{M}}$ are the projection maps defined in the proofs of Theorem 5.2.7 and Theorem 5.2.11.

Let $\mathbf{2} := \{0, 1\}$ and let $p$ be the probability measure on $\mathbf{2}$ defined by $p_1 := \mathbb{P}(B)$. We have a measure preserving map

$$h : \mathbf{\Omega} \to (\mathbf{2}, p)$$

defined by the assignment

$$h(x) := \begin{cases} 1 \text{ if } x \in B \\ 0 \text{ otherwise.} \end{cases}$$

The commutative diagram for this measure preserving map gives us

$$[(\rho_r)_{\mathbf{\Omega}}(g)](B) = [p_h^{\mathrm{M}} \circ (\rho_r)_{\mathbf{\Omega}}(g)]_1 = [(\rho_r^f)_{\mathbf{2}} \circ p_h^{\mathrm{RV}}(g)]_1 = \mathbb{E}[g1_{h^{-1}(1)}] = \mathbb{E}[g1_B],$$

for all $g \in \mathrm{RV}_r(\mathbf{\Omega})$ and $B \in \mathcal{F}$. ☐

Because $\rho_r^f$ is an isomorphism, so is $\rho_r$. This gives us the bounded Radon–Nikodym theorem.

**Theorem 5.2.17** (Bounded Radon–Nikodym)**.** *The natural transformation* $\rho_r : \mathrm{RV}_r \to \mathrm{M}_r$ *is an isomorphism.*

*Proof.* By Proposition 5.2.6, we know that $\mathrm{RV}_r^f \xrightarrow{\rho_r^f} \mathrm{M}_r^f$ is an isomorphism. Using Theorem 5.2.7 and Theorem 5.2.11 and the fact that the Kan extension is functorial, we conclude that

$$\mathrm{RV}_r \xrightarrow{\mathrm{Ran}_i \rho_r^f} \mathrm{M}_r$$

is a natural isomorphism.

$$\begin{tikzcd}
\mathbf{Prob}_f \arrow[rr, bend left, "\mathrm{M}_r^f"] \arrow[rr, bend right, "\mathrm{RV}_r^f"'] \arrow[rr, phantom, "\cong"] \arrow[dr, "i"'] & & \mathbf{CMet} \\
& \mathbf{Prob} \arrow[ur, bend left, "\mathrm{M}_r"] \arrow[ur, bend right, "\mathrm{RV}_r"'] \arrow[ur, phantom, "\cong"] &
\end{tikzcd}$$

□

The natural transformations $\rho_r : \mathrm{RV}_r \to \mathrm{M}_r$ for every $r > 0$, induce a morphism of diagrams $\tilde{\rho} : D_{\mathrm{RV}} \to D_{\mathrm{M}}$. Therefore we obtain a natural transformation $\mathrm{colim}\tilde{\rho} : \mathrm{colim}D_{\mathrm{RV}} \to \mathrm{colim}D_{\mathrm{M}}$. This is a natural transformation $\mathrm{RV} \to \mathrm{M}$, which we will denote by $\rho$. The following proposition describes this natural transformation.

**Proposition 5.2.18.** *Let* $\mathbf{\Omega}$ *be a probability space, then* $\rho_{\mathbf{\Omega}}$ *is the map* $\mathrm{RV}(\mathbf{\Omega}) \to \mathrm{M}(\mathbf{\Omega})$ *defined by the assignment*

$$f \mapsto \int_{(-)} f d\mathbb{P}.$$

*Proof.* The map $\rho_{\mathbf{\Omega}}$ is the unique map that makes the following diagram commute for every $r > 0$.

$$\begin{tikzcd}
\mathrm{RV}(\mathbf{\Omega}) \arrow[r, "\rho_{\mathbf{\Omega}}"] & \mathrm{M}(\mathbf{\Omega}) \\
\mathrm{RV}_r(\mathbf{\Omega}) \arrow[u, "i_{r,\mathbf{\Omega}}"] \arrow[r, "(\rho_r)_{\mathbf{\Omega}}"] & \mathrm{M}_r(\mathbf{\Omega}) \arrow[u, "j_{r,\mathbf{\Omega}}"']
\end{tikzcd}$$

Where $i_{r,\mathbf{\Omega}}$ and $j_{r,\mathbf{\Omega}}$ are the inclusion maps. We have that for all $r > 0$,

$$\int_{(-)} i_{r,\mathbf{\Omega}}(f) \mathrm{d}\mathbb{P} = j_{r,\mathbf{\Omega}} \rho_r(f)$$

for all $f \in \mathrm{RV}_r(\mathbf{\Omega})$. The claim now follows. □

We are now ready to complete the categorical proof for the Radon–Nikodym theorem.

**Theorem 5.2.19** (Radon–Nikodym)**.** *The natural transformation* $\rho : \mathrm{RV} \to \mathrm{M}$ *is an isomorphism.*

*Proof.* Because $\rho_r : \mathrm{RV}_r \to \mathrm{M}_r$ is an isomorphism for every $r > 0$, so is $\tilde{\rho} : D_{\mathrm{RV}} \to D_{\mathrm{M}}$. We find that $\rho := \mathrm{colim}\tilde{\rho} : \mathrm{RV} \to \mathrm{M}$ is a natural isomorphism.

$$\begin{tikzcd}
\dots \arrow[r] & \mathrm{RV}_1 \arrow[r] \arrow[d, dashed, "\cong"] & \dots \arrow[r] & \mathrm{RV}_r \arrow[r] \arrow[d, dashed, "\cong"] & \cdots \arrow[r] & \mathrm{RV} \arrow[d, dashed, "\cong"] \\
\dots \arrow[r] & \mathrm{M}_1 \arrow[r] & \dots \arrow[r] & \mathrm{M}_r \arrow[r] & \cdots \arrow[r] & \mathrm{M}
\end{tikzcd}$$

□

## 5.3 The martingale convergence theorem

In this section we will focus on a special class of stochastic processes, namely *martingales*. These stochastic processes have nice convergence properties, of which we will prove one categorically later in this section. Important examples of martingales are Brownian motion and unbiased random walks.

Let $\mathbf{\Omega} := (\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $I$ be a directed poset. A **filtration** is an indexed collection $(\mathcal{F}_i)_{i \in I}$ of $\sigma$-subalgebras of $\mathcal{F}$ such that $\mathcal{F}_i \subseteq \mathcal{F}_j$ for $i \leq j$ and such that

$$\sigma\left(\bigcup_{i \in I} \mathcal{F}_i\right) = \mathcal{F}.$$

We say that $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{i \in I}, \mathbb{P})$ is a **filtered probability space**. The probability space $(\Omega, \mathcal{F}_i, \mathbb{P} \mid_{\mathcal{F}_i})$ is denoted by $\mathbf{\Omega}_i$. For $i \leq j$ in $I$, there is a measure-preserving map $f_{ij} : \mathbf{\Omega}_j \to \mathbf{\Omega}_j$ and for every $i \in I$ there is a measure-preserving map $f_i : \mathbf{\Omega} \to \mathbf{\Omega}_i$.

An indexed collection $(X_i)_{i \in I}$ of random variables such that $X_i \in \mathrm{RV}(\mathbf{\Omega}_i)$ is called a **martingale** if

$$\mathbb{E}[X_j \mid f_{ij}] = X_i$$

for all $i \leq j$ in $I$.

Martingales often have nice convergence properties. We will categorically prove a weaker version of the following *martingale convergence theorems* in Section 5.3.4. The proofs can be found in Section XI.14 in[13].

**Theorem 5.3.1** (Doob's $L^1$ martingale convergence theorem)**.** *Let $(X_n)_{n=1}^{\infty}$ be a martingale such that*

$$\lim_{\lambda \to \infty} \sup_n \mathbb{E}[X_n 1_{\{X_n > \lambda\}}] = 0,$$

*then $(X_n)_n$ converges to a random variable $X$ in $L^1$-norm and for all $n \geq 1$,*

$$\mathbb{E}[X \mid f_n] = X_n.$$

**Theorem 5.3.2** (Doob's $L^p$ martingale convergence theorem)**.** *Let $p > 1$ and let $(X_n)_{n=1}^{\infty}$ be a martingale such that*

$$\sup_n \mathbb{E}[X_n^p] < \infty,$$

*then $(X_n)_n$ converges to a random variable $X$ in $L^p$-norm and for all $n \geq 1$,*

$$\mathbb{E}[X \mid f_n] = X_n.$$

To give a categorical proof, the setting from Section 5.2 does not quite work. We need to change everything from Section 5.2 to the *enriched* setting. We will enrich everything over the closed monoidal category **CMet**, which we will discuss in Section 5.3.1. We then show in Section

5.3.2 and Section 5.3.3 that the results from Section 5.2 still work when everything is enriched over **CMet**. We then conclude Section 5.3 by giving a categorical proof for a weaker version of the martingale convergence theorems in Section 5.3.4.

### 5.3.1 The closed monoidal category CMet

In this section we will give an overview of well-known results about metric spaces. For completeness, we give proofs for all the results.

Let $i : \mathbf{CMet} \to \mathbf{Met}$ be the inclusion functor of the full subcategory of complete metric spaces in the category of metric spaces.

**Proposition 5.3.3.** *The category* **CMet** *is complete and* $i : \mathbf{CMet} \to \mathbf{Met}$ *preserves these limits.*

*Proof.* For a collection of complete metric spaces $(X_i, d_i)_{i \in I}$ let $X := \prod_{i \in I} X_i$ and define $d : X \times X \to [0, \infty]$ by

$$d((x_i)_i, (y_i)_i) := \sup_{i \in I} d_i(x_i, y_i).$$

It is clear that $d$ defines a metric[1] and that the projection maps $\pi_i : X \to X_i$ are 1-Lipschitz. Let $(x^n)_n$ be a Cauchy sequence in $(X, d)$. Clearly, $(x_i^n)_n$ is a Cauchy sequence in $(X_i, d_i)$ for all $i \in I$. It follows that $(x_i^n)_n$ converges to an element $x_i$ in $(X_i, d_i)$. Denote $x := (x_i)_i$. For $\epsilon > 0$, there exists an $N \geq 1$ such that for $n_1, n_2 \geq N$,

$$d((x_i^{n_1})_i, (x_i^{n_2})_i) \leq \epsilon.$$

For $i \in I$, there exists $M_i \geq N$ such that $d_i(x_i^{M_i}, x_i) \leq \epsilon$. It follows now that

$$d_i(x_i^N, x_i) \leq d_i(x_i^N, x_i^{M_i}) + d_i(x_i^{M_i}, x_i) \leq 2\epsilon.$$

Since $N$ does not depend on $i$, we can take the supremum over all $i \in I$ and conclude that

$$d(x^N, x) \leq 2\epsilon.$$

Therefore, $(x^n)_n$ converges to $x$ in $(X, d)$ and thus it is a complete metric space. The complete metric space $(X, d)$ is the product of $(X_i, d_i)_{i \in I}$.

For morphisms $f, g : (X, d_X) \to (Y, d_Y)$ in **CMet**. Let

$$E := \{x \in X \mid f(x) = g(x)\}$$

and let $d_E$ be the restriction of $d_X$ to $E \times E$. This forms a metric space $(E, d_E)$. For a Cauchy sequence $(e_n)_n$ in $(E, d_E)$ we know that $(e_n)_n$ converges to some $x$ in $(X, d_X)$. Because $f$ and $g$ are 1-Lipschitz, we see that

$$f(x) = \lim_n f(e_n) = \lim_n g(e_n) = g(x).$$

Therefore, $x \in E$ and $(E, d)$ is complete. The complete metric space $(E, d)$ is the equalizer of $f$ and $g$ in **CMet**.

It is clear that $i : \mathbf{CMet} \to \mathbf{Met}$ preserves these limits. □

---

[1] Recall that all our metrics are really *extended pseudometrics*. Here, $d$ can take the value $\infty$ even when $d_i$ is finite for every $i$ in $I$.

**Proposition 5.3.4** (Completion)**.** *The inclusion* $i : \mathbf{CMet} \to \mathbf{Met}$ *has a left adjoint.*

*Proof.* The completion functor $\overline{(-)} : \mathbf{Met} \to \mathbf{CMet}$ that sends a metric space $(X, d)$ to its completion $\overline{(X, d)}$ is left adjoint to the inclusion functor $i : \mathbf{CMet} \to \mathbf{Met}$. ☐

Proposition 5.3.4 tells us that **CMet** is a *reflective* subcategory of **Met**. We will use this in the following result about colimits in **Met** and **CMet**.

**Proposition 5.3.5.** *The categories* **CMet** *and* **Met** *are cocomplete.*

*Proof.* For a collection of metric spaces $(X_i, d_i)_{i\in I}$, let $X := \coprod_{i\in I} X_i$ and define $d : X \times X \to [0, \infty]$ by

$$d(x, y) := \begin{cases} d_i(x, y) \text{ if } x, y \in X_i \\ \infty \text{ otherwise.} \end{cases}$$

Then $(X, d)$ forms a metric spaces and the inclusion maps $\iota_i : X_i \to X$ are 1-Lipschitz maps. The metric space $(X, d)$ is the coproduct of $(X_i, d_i)_i$.

For morphisms $f, g : (X, d_X) \to (Y, d_Y)$ in **Met**, let $\sim$ be the smallest equivalence relation such that $y_1 \sim y_2$ if there exist an $x \in X$ such that $f(x) = y_1$ and $g(x) = y_2$. Denote $F := Y/\sim$. Define a map $d : F \times F \to [0, \infty]$ by

$$d(y_1, y_2) := \inf\{\sum_{k=1}^{n} d_Y(x_k, z_k) \mid y_1 \sim x_1, y_2 \sim z_n \text{ and } z_k \sim x_{k+1}\}.$$

This map is well-defined and is a metric. The quotient map $Y \to F$ is 1-Lipschitz and it is easy to verify that $F$ is the coequalizer of $f$ and $g$ in **Met**.

By Proposition 5.3.4, **CMet** is reflective in $\mathbf{Met}_1$ and therefore it is cocomplete. Colimits in **CMet** are constructed by reflecting colimits in **Met**. ☐

For (complete) metric spaces $(X_1, d_1)$ and $(X_2, d_2)$ let $(X_1, d_1) \otimes (X_2, d_2)$ be the (complete) metric space formed by the set $X_1 \times X_2$ with the metric

$$d((x_1, x_2), (y_1, y_2)) := d_1(x_1, y_1) + d(x_2, y_2).$$

For 1-Lipschitz maps $f_1 : (X_1, d_{X_1}) \to (Y_1, d_{Y_1})$ and $f_2 : (X_2, d_{X_2}) \to (Y_2, d_{Y_2})$, there is a 1-Lipschitz map

$$f_1 \otimes f_2 : (X_1, d_{X_1}) \otimes (X_2, d_{X_2}) \to (Y_1, d_{Y_1}) \otimes d(Y_2, d_{Y_2})$$

defined by

$$(x_1, x_2) \mapsto (f_1(x_1), f_2(x_2)).$$

This obtained functor $\otimes : \mathbf{CMet} \times \mathbf{CMet} \to \mathbf{CMet}$ together with the metric space $\mathbf{1}$ consisting of one element form a symmetric monoidal structure on **CMet**.

For metric spaces $(X, d_X)$ and $(Y, d_Y)$, let $[(X, d_X), (Y, d_Y)]$ be the set of 1-Lipschitz maps $(X, d_X) \to (Y, d_Y)$ together with the metric defined by

$$d(f, g) := \sup\{d_Y(f(x), g(x)) \mid x \in X\}.$$

**Proposition 5.3.6.** *If* $(Y, d_Y)$ *is complete, then so is* $[(X, d_X), (Y, d_Y)]$.

*Proof.* Let $(f_n)_n$ be a Cauchy sequence in $[(X, d_X), (Y, d_Y)]$, then it is clear that $(f_n(x))_n$ is a Cauchy sequence for every $x$. Therefore $(f_n(x))_n$ converges to an element $f_x \in Y$.

Define a map $f : X \to Y$ by sending $x$ to $f_x$. Consider $x_1$ and $x_2$ in $X$ and let $\epsilon > 0$. There exists an $n \geq 1$ such that

$$\begin{aligned} d_Y(f(x_1), f(x_2)) &\leq d_Y(f_{x_1}, f_n(x_1)) + d_Y(f_n(x_1), f_n(x_2)) + d_Y(f_n(x_2), f_{x_2}) \\ &\leq d_Y(f_n(x_1), f_n(x_2) + \epsilon \leq d_X(x_1, x_2) + \epsilon. \end{aligned}$$

Taking $\epsilon \to 0$ shows that $f$ is 1-Lipschitz.

For $\epsilon > 0$, there is an $N \geq 1$ such that for $n_1, n_2 \geq N$,

$$d(f_{n_1}, f_{n_2}) \leq \epsilon.$$

Let $n \geq N$. For $x \in X$, there exists $M_x \geq n$ such that $d_Y(f_x, f_{M_x}(x)) \leq \epsilon$ and therefore

$$d_Y(f_x, f_n(x)) \leq d_Y(f_x, f_{M_x}(x)) + d_Y(f_{M_x}(x), f_n(x)) \leq 2\epsilon.$$

Since $n$ is independent from $x$, we can conclude that $d(f, f_n) \leq 2\epsilon$ for all $n \geq N$, which means that $(f_n)_n$ converges to $f$ in $[(X, d_X), (Y, d_Y)]$. □

**Proposition 5.3.7.** *The monoidal category* **CMet** *is closed.*

*Proof.* Let $(X, d_X), (Y, d_Y)$ and $(Z, d_Z)$ be complete metric spaces. It is easy to verify that there is a natural bijection

$$\mathbf{CMet}(X \otimes Y, Z) \cong \mathbf{CMet}(X, [Y, Z]).$$

□

**Proposition 5.3.8.** *For metric spaces $X$ and $Y$, $\overline{X \otimes Y} \cong \overline{X} \otimes \overline{Y}$*

*Proof.* Let $Z$ be a complete metric space. We have the following natural bijections:

$$\begin{array}{c} X \otimes Y \to Z \\ \hline X \to [Y, Z] \\ \hline \overline{X} \to [Y, Z] \\ \hline \overline{X} \otimes Y \to Z \\ \hline Y \to [\overline{X}, Z] \\ \hline \overline{Y} \to [\overline{X}, Z] \\ \hline \overline{X} \otimes \overline{Y} \to Z \end{array}$$

Here we used that by Proposition 5.3.6, $[Y, Z]$ and $[\overline{X}, Z]$ are complete, since $Z$ is. Since $Z$ was chosen arbitrarily, the claim now follows. □

Proposition 5.3.7 says that **CMet** is a closed monoidal category. In what follows we will look at categories that are *enriched* over this closed monoidal category. The 2-category of **CMet**-enriched categories, enriched functors and enriched natural transformations is denoted as **CMet-Cat**.

The forgetful functor $U$ : **CMet** $\to$ **Set** induces a 2-functor $U_*$ : **CMet-Cat** $\to$ **Cat**. Therefore, for **CMet**-enriched categories $\mathcal{C}$ and $\mathcal{D}$, there is a functor

$$\mathbf{CMet\text{-}Cat}[\mathcal{C}, \mathcal{D}] \to \mathbf{Cat}[U_*\mathcal{C}, U_*\mathcal{D}].$$

The following lemmas will be used later to lift the results from Section 5.2 to the enriched setting.

**Lemma 5.3.9.** *For a* $\mathbf{CMet}$*-enriched categories* $\mathcal{C}$*, the functor*

$$\mathbf{CMet}{-}\mathbf{Cat}[\mathcal{C}, \mathbf{CMet}] \to \mathbf{Cat}[U_*\mathcal{C}, U_*\mathbf{CMet}]$$

*is full and faithful.*

*Proof.* Let $F$ and $G$ be enriched functors $\mathcal{C} \to \mathbf{CMet}$. There is a one-to-one correspondence between 1-Lipschitz maps $Fc \to Gc$ and $1 \to [Fc, Gc]$ for all objects $c$ in $\mathcal{C}$. It follows now that every natural transformation $U_*F \to U_*G$ can be uniquely lifted to an enriched natural transformation $F \to G$. □

The following corollary states that if the non-enriched right Kan extension of enriched functors is an enriched functor, then it is also the enriched right Kan extension.

**Corollary 5.3.10.** *Let* $\mathcal{C}, \mathcal{D}$ *be* $\mathbf{CMet}$*-enriched categories. Let* $F : \mathcal{C} \to \mathbf{CMet}$*,* $G : \mathcal{C} \to \mathcal{D}$ *and* $H : \mathcal{D} \to \mathbf{CMet}$ *be enriched functors and let* $\epsilon : U_*H \circ U_*G \to U_*F$ *be a (non-enriched) natural transformation such that*

$$\begin{tikzcd}
U_*\mathcal{C} \arrow[r, "U^*F"] \arrow[d, "U_*G"'] & U_*\mathbf{CMet} \\
U_*\mathcal{D} \arrow[ur, "U_*H"'] \arrow[ur, Rightarrow, shift left=3, shorten=3mm, "\epsilon"] &
\end{tikzcd}$$

$\epsilon$ *exhibits* $U_*H$ *as the right Kan extension of* $U_*F$ *along* $U_*G$*, then there exists a unique enriched natural transformation* $\tilde{\epsilon} : HG \to F$ *such that* $U_*\tilde{\epsilon} = \epsilon$ *and* $\tilde{\epsilon}$ *exhibits* $H$ *as the right Kan extension of* $F$ *along* $G$*.*

### 5.3.2 Prob is enriched over CMet

Let $\mathbf{\Omega}_1 := (\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $\mathbf{\Omega}_2 := (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces and $s \geq 1$. Let $\mathrm{prob}^s[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ be the set of measurable maps $\mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ such that $\mathbf{P}_1 \circ f^{-1} < s\mathbb{P}_2$ in the case that $s > 1$ and the set of measure-preserving maps in the case that $s = 1$ . For $f_1, f_2 \in \mathrm{prob}[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$, define

$$d_{\mathbf{\Omega}_1, \mathbf{\Omega}_2}(f_1, f_2) := 2 \sup\{\mathbb{P}(f_1^{-1}(A) \triangle f_2^{-1}(A)) \mid A \text{ measurable subset of } \mathbf{\Omega}_2\}.$$

This turns $\mathrm{prob}^s[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ into a metric space.

For probability spaces $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ let $\mathrm{Prob}^s_r[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ be the metric space we obtain by scaling the metric of $\mathrm{prob}^s[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ by a factor $r > 0$. Moreover, let $\mathbf{Prob}^s_r[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ be the completion of $\mathrm{Prob}^s_r[\mathbf{\Omega}_1, \mathbf{\Omega}_1]$.

For now, we will focus on the case that $s = 1$ and we will omit the index $s$ in this case. This means that $\mathrm{Prob}_r[\mathbf{\Omega}_1, \mathbf{\Omega}_2]$ consists of the measure-preserving maps. Later, in Corollary 5.3.14 we will have to use the case that $s > 1$.

Note that the limit of the diagram

$$\mathbf{Prob}_1[\mathbf{\Omega}_1, \mathbf{\Omega}_2] \longleftarrow \ldots \longleftarrow \mathbf{Prob}_r[\mathbf{\Omega}_1, \mathbf{\Omega}_2] \longleftarrow \ldots$$

is the discrete pseudometric space of measure-preserving maps $f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$.

Consider probability spaces $\mathbf{\Omega}_1, \mathbf{\Omega}_2$ and $\mathbf{\Omega}_3$. Sending a pair of measure-preserving maps $f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$ and $g : \mathbf{\Omega}_2 \to \mathbf{\Omega}_3$, to $g \circ f : \mathbf{\Omega}_1 \to \mathbf{\Omega}_3$, defines a 1-Lipschitz map

$$- \circ - : \mathrm{prob}[\mathbf{\Omega}_1, \mathbf{\Omega}_2] \otimes \mathrm{prob}[\mathbf{\Omega}_2, \mathbf{\Omega}_3] \to \mathrm{prob}[\mathbf{\Omega}_1, \mathbf{\Omega}_3]$$

Indeed, this follows from the fact that

$$\begin{aligned}2\mathbb{P}_1\left((g_1f_1)^{-1}(A)\triangle(g_2f_2)^{-1}(A)\right) &\leq 2\mathbb{P}_1\left((g_1f_1)^{-1}(A)\triangle(g_2f_1)^{-1}(A)\right)+2\mathbb{P}_1\left((g_2f_1)^{-1}(A)\triangle(g_2f_2)^{-1}(A)\right)\\ &= 2\mathbb{P}_2(g_1^{-1}(A)\triangle g_2^{-1}(A))+2\mathbb{P}_1(f_1^{-1}(g_2^{-1}(A))\triangle f_2^{-1}(g_2^{-1}(A)))\\ &\leq d_{\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_3}(g_1,g_2)+d_{\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(f_1,f_2)\end{aligned}$$

for all measurable subsets of $\boldsymbol{\Omega}_3$. Using Proposition 5.3.8, this induces a 1-Lipschitz map

$$\mathbf{Prob}_r[\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2]\otimes\mathbf{Prob}_r[\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_3]\to\mathbf{Prob}_r[\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_3]$$

The above describes a category enriched over **CMet**, whose objects are probability spaces and whose hom-objects are given by $\mathbf{Prob}_r[\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2]$ for probability spaces $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$. We denote this category by $\mathbf{Prob}_r$. The subcategory of finite probability spaces is denoted as $\mathbf{Prob}_r^f$ and clearly there is an enriched inclusion functor $i_r : \mathbf{Prob}_r^f \to \mathbf{Prob}_r$. Furthermore, note that $U_*\mathbf{Prob}_r$ is the (non-enriched) category **Prob** for all $r > 0$.

### 5.3.3 The enriched functors $\mathrm{M}_r$ and $\mathrm{RV}_r$

In this section we will show that everything proved in Section 5.2 still works in the enriched context.

The (non-enriched) functor $\mathrm{M}_r : \mathbf{Prob} \to \mathbf{CMet}$ from Section 5.3 induces a **CMet**-enriched functor $\mathbf{Prob}_r \to \mathbf{CMet}$. Indeed, the assignment $f \mapsto \mathrm{M}_r(f)$, induces a 1-*Lipschitz map*

$$\mathrm{prob}_r[\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2]\to[\mathrm{M}_r(\boldsymbol{\Omega}_1),\mathrm{M}_r(\boldsymbol{\Omega}_2)].$$

To see this, consider two measure-preserving maps $f_1, f_2 : \boldsymbol{\Omega}_1 \to \boldsymbol{\Omega}_2$. For $\mu \in \mathrm{M}_r(\boldsymbol{\Omega}_1)$ and a measurable subset $A$ of $\boldsymbol{\Omega}_2$, we find that

$$\begin{aligned}|\mathrm{M}_r(f_1)(\mu)(A)-\mathrm{M}_r(f_2)(\mu)(A)| &= |\mu(f_1^{-1}(A))-\mu(f_2^{-1}(A)))|\\ &= \left|\int 1_{f_1^{-1}(A)}-1_{f_2^{-1}(A)}\mathrm{d}\mu\right|\\ &\leq \mu\left(f_1^{-1}(A)\triangle f_2^{-1}(A)\right)\\ &\leq r\mathbb{P}_1(f_1^{-1}(A)\triangle f_2^{-1}(A))\end{aligned}$$

Therefore, we have that

$$|\mathrm{M}_r(f_1)(\mu)(A)-\mathrm{M}_r(f_2)(\mu)(A)|+|\mathrm{M}_r(f_1)(\mu)(A^{\mathrm{C}})-\mathrm{M}_r(f_2)(\mu)(A^{\mathrm{C}})|\leq rd_{\Omega_1,\Omega_2}(f_1,f_2).$$

Taking the supremum over all measurable subsets $A$ of $\boldsymbol{\Omega}_2$, gives us that

$$d_{\mathrm{M}_r(\boldsymbol{\Omega}_2)}(\mathrm{M}_r(f_1)(\mu),\mathrm{M}_r(f_2)(\mu))\leq rd_{\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(f_1,f_2).$$

Finally, by taking the supremum of over all $\mu \in \mathrm{M}_r(\boldsymbol{\Omega}_1)$, we see that the assignment $f \mapsto \mathrm{M}_r(f)$ defines a 1-Lipschitz map. This gives a 1-Lipschitz map $\mathbf{Prob}_r[\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2] \to [\mathrm{M}_r(\boldsymbol{\Omega}_1),\mathrm{M}_r(\boldsymbol{\Omega}_2)]$. The obtained enriched functor $\mathbf{Prob}_r \to \mathbf{CMet}$ is also denoted by $\mathrm{M}_r$.

The restriction to finite probability spaces is denoted by $\mathrm{M}_r^f$ and is an enriched functor since it is the composite of the enriched functors $\mathrm{M}_r : \mathbf{Prob}_r \to \mathbf{CMet}$ and $i_r : \mathbf{Prob}_r^f \to \mathbf{Prob}_r$.

**Proposition 5.3.11.** *The commutative triangle of enriched functors*

$$\begin{tikzcd}
\mathbf{Prob}_r^f \arrow[r, "\mathrm{M}_r^f"] \arrow[d, "i_r"'] & \mathbf{CMet} \\
\mathbf{Prob}_r \arrow[ur, "\mathrm{M}_r"'] &
\end{tikzcd}$$

*exhibits* $\mathrm{M}_r$ *as the right Kan extension of* $\mathrm{M}_r^f$ *along* $i_r$.

*Proof.* This follows from Theorem 5.2.7 together with Corollary 5.3.10. ☐

The (non-enriched) functor $\mathrm{RV}_r : \mathbf{Prob} \to \mathbf{CMet}$ from Section 5.2 induces a **CMet**-enriched functor $\mathbf{Prob}_r \to \mathbf{CMet}$. Indeed, the assignment $f \mapsto \mathrm{RV}_r(f)$ induces a 1-*Lipschitz map*

$$\mathrm{prob}(\mathbf{\Omega}_1, \mathbf{\Omega}_2) \to [\mathrm{RV}_r(\mathbf{\Omega}_1), \mathrm{RV}_r(\mathbf{\Omega}_2)].$$

To see this, consider two measure-preserving maps $f_1, f_2 : \mathbf{\Omega}_1 \to \mathbf{\Omega}_2$. For $X \in \mathrm{RV}_r(\mathbf{\Omega}_1)$, consider the measurable subset $A^+ := \{\mathbb{E}[X \mid f_1] \geq \mathbb{E}[X \mid f_2]\}$ and let $A^-$ be its complement.[2] We now find that

$$\begin{aligned}
d_{\mathrm{RV}_r(\mathbf{\Omega}_2)}(\mathrm{RV}_r(f_1)(X), \mathrm{RV}_r(f_2)(X)) &= \mathbb{E}[|\mathbb{E}[X \mid f_1] - \mathbb{E}[X \mid f_2]|] \\
&= \mathbb{E}[(\mathbb{E}[X \mid f_1] - \mathbb{E}[X \mid f_2])1_{A^+}] + \mathbb{E}[(\mathbb{E}[X \mid f_2] - \mathbb{E}[X \mid f_1])1_{A^-}] \\
&= \mathbb{E}[X(1_{f_1^{-1}(A^+)} - 1_{f_2^{-1}(A^+)} + 1_{f_2^{-1}(A^-)} - 1_{f_1^{-1}(A^-)})] \\
&\leq r(\mathbb{E}[|1_{f_1^{-1}(A^+)} - 1_{f_2^{-1}(A^+)}|] + \mathbb{E}[|1_{f_1^{-1}(A^-)} - 1_{f_2^{-1}(A^-)}|]) \\
&= r(\mathbb{P}(f_1^{-1}(A^+) \triangle f_2^{-1}(A^+)) + \mathbb{P}(f_1^{-1}(A^-) \triangle f_2^{-1}(A^-))) \\
&\leq r d_{\mathbf{\Omega}_1, \mathbf{\Omega}_2}(f_1, f_2)
\end{aligned}$$

By taking the supremum over all $X \in \mathrm{RV}_r(\mathbf{\Omega}_1)$, we see that the assignment $f \mapsto \mathrm{RV}_r(f)$ defines a 1-Lipschitz maps. This induces a 1-Lipschitz map $\mathbf{Prob}_r[\mathbf{\Omega}_1, \mathbf{\Omega}_2] \to [\mathrm{RV}_r(\mathbf{\Omega}_1, \mathrm{RV}_r(\mathbf{\Omega}_2)]$. The enriched functor $\mathbf{Prob}_r \to \mathbf{CMet}$ that we obtain will also be denoted by $\mathrm{RV}_r$.

The restriction to finite probability spaces is also denoted by $\mathrm{RV}_r^f$ and is an enriched functor since it is the composite of the enriched functors $\mathrm{RV}_r : \mathbf{Prob}_r \to \mathbf{CMet}$ and $i_r : \mathbf{Prob}_r^f \to \mathbf{Prob}_r$.

**Proposition 5.3.12.** *The commutative triangle of* enriched *functors*

$$\begin{tikzcd}
\mathbf{Prob}_r^f \arrow[r, "\mathrm{RV}_r^f"] \arrow[d, "i_r"'] & \mathbf{CMet} \\
\mathbf{Prob}_r \arrow[ur, "\mathrm{RV}_r"'] &
\end{tikzcd}$$

*exhibits* $\mathrm{RV}_r$ *as the right Kan extension of* $\mathrm{RV}_r^f$ *along* $i_r$.

*Proof.* This follows from Theorem 5.2.11 together with Corollary 5.3.10. ☐

---

[2] The subset $A^+$ should actually be defined as $\{g_1 \geq g_2\}$ for some measurable maps $g_1 : \Omega_2 \to [0, r]$ and $g_2 : \Omega_2 \to [0, r]$ representing $\mathbb{E}[X \mid f_1]$ and $\mathbb{E}[X \mid f_2]$ respectively. However, everything that follows is independent from the choice of $g_1$ and $g_2$ and therefore we just write $\{\mathbb{E}[X \mid f_1] \geq \mathbb{E}[X \mid f_2]\}$.

### 5.3.4 The martingale convergence theorem

Let $(\Omega, (\mathcal{F}_i)_{i\in I}, \mathcal{F}, \mathbb{P})$ be a filtered probability space. For $i \in I$, we write $\mathbf{\Omega}_i$ for $(\Omega, \mathcal{F}_i, \mathbb{P}\mid_{\mathcal{F}_i})$ and $\mathbf{\Omega}$ for $(\Omega, \mathcal{F}, \mathbb{P})$. For $i \leq j$ in $I$ there is a measure-preserving map $f_{ij} : \mathbf{\Omega}_j \to \mathbf{\Omega}_i$ and for $i \in I$ there is a measure-preserving map $f_i : \mathbf{\Omega} \to \mathbf{\Omega}_i$. This induces a diagram $D_{\mathbf{\Omega}}$ in the underlying category of $\mathbf{Prob}_r$ of which $\mathbf{\Omega}$ is the conical limit.

In the case where $I = \mathbb{N}$, we have the following diagram.

$$\begin{tikzcd}
\mathbf{\Omega}_1 & \mathbf{\Omega}_2 \arrow[l] & \mathbf{\Omega}_3 \arrow[l] & \dots \arrow[l] & \mathbf{\Omega} \arrow[l]
\end{tikzcd}$$

**Lemma 5.3.13.** *Let* $(\Omega, (\mathcal{F}_i)_{i\in I}, \mathcal{F}, \mathbb{P})$ *be a filtered probability space. Then for* $E \in \mathcal{F}$*, there exists a sequence* $(E_n)_n$ *in* $\bigcup_{i\in I} \mathcal{F}_i$ *such that*

$$\mathbb{P}(E \triangle E_n) \to 0.$$

*Proof.* Note that $\bigcup_{i\in I} \mathcal{F}_i$ is closed under complements and *finite* intersections and unions, because $I$ is directed. The claim now immediately follows from Theorem D in Section 13 in [31]. $\square$

**Corollary 5.3.14.** *Let* $\mathbf{A}$ *be a finite probability space and let* $s > 1$*. The enriched functor* $\mathbf{Prob}_r^s(-, \mathbf{A}) : \mathbf{Prob}_r \to \mathbf{CMet}^{\mathrm{op}}$ *preserves the limit of* $D_{\mathbf{\Omega}}$*.*

*Proof.* We will first show that

$$\bigcup_{i\in I} \mathrm{Prob}_r^s(\mathbf{\Omega}_i, \mathbf{A}) \subseteq \mathrm{Prob}_r^s(\mathbf{\Omega}, \mathbf{A})$$

is a dense subset. Let $f : \mathbf{\Omega} \to \mathbf{A}$ be a map such that $\mathbb{P} \circ f^{-1} < sp$, then there exists an $0 < r < s$ such that $\mathbb{P} \circ f^{-1} \leq (s - r)p$. We can assume without loss of generality that $A = \{1, 2, \dots, n\}$ and that $p_i > 0$ for $1 \leq i \leq m$ and $p_i = 0$ for $m < i \leq n$. Let $p_{\min} := \min\{p_1, \dots, p_m\}$. For $0 < \epsilon < \frac{r}{2} p_{\min}$, it follows from Proposition 5.3.13 that there exists an $i \in I$ and $(E_k)_{k=1}^n$ in $\mathcal{F}_i$ such that

$$\sup_{k=1}^{n} \mathbb{P}(E_k \triangle f^{-1}(k)) < \frac{\epsilon}{4n^2}.$$

Now define the following subsets in $\Sigma_i$:

$$\begin{aligned}
F_1 &:= E_1 \\
F_2 &:= E_2 \setminus F_1 \\
&\vdots \\
F_{m-1} &:= E_{m-1} \setminus F_{m-2} \\
F_m &:= \Omega \setminus \left( \bigcupdot_{k=1}^{m-1} F_k \right) \\
F_{m+1} &:= \emptyset \\
&\vdots \\
F_n &:= \emptyset
\end{aligned}$$

These define a map $f_i : \mathbf{\Omega}_i \to \mathbf{A}$ such that $f^{-1}(k) = F_k$ for every $1 \leq k \leq n$. We also have that

- $\mathbb{P}(F_1 \triangle f^{-1}(1)) < \frac{\epsilon}{4n^2} < \frac{\epsilon}{2n}$,

- For $2 \leq l \leq m-1$:

$$\begin{aligned}\mathbb{P}(F_k \triangle f^{-1}(k)) &\leq \mathbb{P}(E_k \triangle f^{-1}(k)) + \mathbb{P}(F_{k-1} \cap f^{-1}(k)) \\ &< \mathbb{P}(E_k \triangle f^{-1}(k)) + \mathbb{P}(E_{k-1} \triangle f^{-1}(k-1)) < 2\frac{\epsilon}{4n^2} \leq \frac{\epsilon}{2n}\end{aligned}$$

- $\mathbb{P}(F_m \triangle f^{-1}(m)) = \mathbb{P}\left(\dot{\bigcup}_{k=1}^{m-1} F_k \triangle \bigcup_{k=1}^{m-1} f^{-1}(k)\right) \leq \sum_{k=1}^{m-1} \mathbb{P}(F_k \triangle f^{-1}(k)) < (2m-3)\frac{\epsilon}{4n^2} \leq \frac{\epsilon}{2n}$

- For $m < k \leq n$: $\mathbb{P}(F_k \triangle f^{-1}(k)) = \mathbb{P}(f^{-1}(k)) = 0$.

It follows that for every $1 \leq k \leq n$, that

$$\mathbb{P} \circ f_i^{-1}(k) = \mathbb{P}(F_k) \leq \mathbb{P} \circ f^{-1}(k) + \epsilon \leq (s-r)p_k + \frac{r}{2}p_k < sp_k.$$

It follows that $f_i \in \mathrm{Prob}_r^s[\mathbf{\Omega}, \mathrm{A}]$. We also see that

$$d_{\mathbf{\Omega},\mathbf{A}}(f_i, f) < \epsilon.$$

It follows $\bigcup_{i \in I} \mathrm{Prob}_r^s(\mathbf{\Omega}_i, \mathbf{A})$ is dense in $\mathrm{Prob}_r^s(\mathbf{\Omega}, \mathbf{A})$ and therefore it is dense in $\mathbf{Prob}_r^s(\mathbf{\Omega}, \mathbf{A})$. From this we can conclude that

$$\bigcup_{i \in I} \mathbf{Prob}_r^s(\mathbf{\Omega}_i, \mathbf{A}) \subseteq \mathbf{Prob}_r^s(\mathbf{\Omega}, \mathbf{A})$$

is a dense subset, which completes the proof. □

**Theorem 5.3.15.** *The functor* $\mathrm{RV}_r : \mathbf{Prob}_r \to \mathbf{CMet}$ *preserves the limit of* $D_{\mathbf{\Omega}} : I \to \mathbf{Prob}$.

*Proof.* Since $\mathrm{RV}_r$ is the right Kan extension of $\mathrm{RV}_r \circ i_r$ along $i_r$, it can be represented as a weighted limit, as explained in Section 4.1 in [37].

$$\mathrm{RV}_r(\mathbf{\Omega}) \cong \{\mathbf{Prob}_r(\mathbf{\Omega}, i_r-), \mathrm{RV}_r \circ i_r\}.$$

Let $1 < s$. We will first show that there is a map

$$\mathrm{RV}_r(\mathbf{\Omega}) \to \{\mathbf{Prob}_r^s(\mathbf{\Omega}, i_r-), \mathrm{RV}_{sr}^f\}.$$

Let $Y$ be a complete metric space and let $f : Y \to \mathrm{RV}_r(\mathbf{\Omega})$ be a 1-Lipschitz map. We will now define a natural transformation

$$\mathbf{Prob}_r^s(\mathbf{\Omega}, i_r-) \to \mathbf{CMet}[Y, \mathrm{RV}_{sr}^f(-)].$$

Let $\mathbf{A} := (A, p)$ be a finite probability space and define a map

$$\varphi_{\mathbf{A}} : \mathbf{Prob}_r^s(\mathbf{\Omega}, \mathbf{A}) \to \mathbf{CMet}[Y, \mathrm{RV}_{sr}^f(\mathbf{A})]$$

by defining $\varphi_{\mathbf{A}}(g)(y) \in \mathrm{RV}^f_{sr}(\mathbf{A})$ by the assignment

$$a \mapsto \begin{cases} \frac{1}{p_a} \int_{g^{-1}(a)} f(y) \mathrm{d}\mathbb{P} & \text{if } p_a \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

for every $y \in Y$ and $g \in \mathbf{Prob}^s_r(\mathbf{\Omega}, \mathbf{A})$. Since $f \leq r$ almost surely and $\mathbb{P} \circ g^{-1}(a) \leq sp_a$, it follows that this map is well-defined. Furthermore, since $f$ is 1-Lipschitz, so is $\varphi_{\mathbf{A}}(g)$.

By the same argument that was used to prove that $\mathrm{RV}_r$ is an enriched functor in Section 5.3.3, it follows that $\varphi_{\mathbf{A}}$ itself is 1-Lipschitz.

To prove naturality in $\mathbf{A}$, consider a measure-preserving map $s : \mathbf{A} \to \mathbf{B}$ between finite probability spaces $\mathbf{A} := (A, p)$ and $\mathbf{B} := (B, q)$.

Since

$$\frac{1}{q_b} \sum_{s(a)=b} \int_{g^{-1}(a)} f(y) \mathrm{d}\mathbb{P} = \frac{1}{q_b} \int_{(sg)^{-1}(a)} f(y) \mathrm{d}\mathbb{P}$$

for every $b \in B$, we conclude that $(\varphi_{\mathbf{A}})_{\mathbf{A}}$ forms a natural transformation. This leads to a 1-Lipschitz map

$$\begin{aligned} \mathbf{CMet}(Y, \mathrm{RV}_r(\mathbf{\Omega})) &\to [\mathbf{Prob}^f_r, \mathbf{CMet}](\mathbf{Prob}^s_r(\mathbf{\Omega}, i_r-), \mathbf{CMet}(Y, \mathrm{RV}^f_{sr}(-))) \\ &= \mathbf{CMet}(Y, \left\{\mathbf{Prob}^s_r(\mathbf{\Omega}, i_r-), \mathrm{RV}^f_{sr}\right\}). \end{aligned}$$

This induces a map $\psi_s : \mathrm{RV}_r(\mathbf{\Omega}) \to \left\{\mathbf{Prob}^s_r(\mathbf{\Omega}, i_r-), \mathrm{RV}^f_{sr}\right\}$, which is natural in $\mathbf{\Omega}$. This leads to to following morphisms:

$$\begin{aligned} \lim_j \mathrm{RV}_r(\mathbf{\Omega}_j) &\to \lim_j \{\mathbf{Prob}^s_r(\mathbf{\Omega}_j, i_r-), \mathrm{RV}^f_{sr}\} \\ &= \{\mathrm{colim}_j \mathbf{Prob}^s_r(\mathbf{\Omega}_j, i_r-), \mathrm{RV}^f_{sr}\} \\ &= \{\mathbf{Prob}^s_r(\mathbf{\Omega}, i_r-), \mathrm{RV}^f_{sr}\} \\ &\to \{\mathbf{Prob}_r(\mathbf{\Omega}, i_r-), \mathrm{RV}^f_{sr}\} = \mathrm{RV}_{sr}(\mathbf{\Omega}). \end{aligned}$$

Here we used Corollary 5.3.14 for the second equality. Taking the limit for $s \to 1$, we obtain a map

$$\psi : \lim_j \mathrm{RV}_r(\mathbf{\Omega}_j) \to \mathrm{RV}_r(\mathbf{\Omega}).$$

For a finite probability space $\mathbf{A}$, we have the outer diagram commutes.

$$\begin{tikzcd}
\mathrm{RV}_r(\mathbf{\Omega}) \arrow[d] \arrow[dr, hook] & \\
\lim_j \mathrm{RV}_r(\mathbf{\Omega}_j) \arrow[r, "\psi_s" description] \arrow[d] & \mathrm{RV}_{sr}(\mathbf{\Omega}) \arrow[d] \\
{[\mathrm{colim}_j \mathbf{Prob}_{sr}(\mathbf{\Omega_j}, \mathbf{A}), \mathrm{RV}_{sr}(\mathbf{\Omega})]} \arrow[r] & {[\mathbf{Prob}_{sr}(\mathbf{\Omega}, \mathbf{A}), \mathrm{RV}_{sr}(\mathbf{\Omega})]}
\end{tikzcd}$$

Hence the top triangle commutes, by the universal property of $\mathrm{RV}_{rs}(\mathbf{\Omega})$ as a weighted limit.

Moreover, for every $j \in I$, we see that the following diagram commutes.

$$\begin{tikzcd}
\lim_j \mathrm{RV}_r(\mathbf{\Omega}) \arrow[rr, "\psi_s"] \arrow[dr, "\pi_j"'] & & \mathrm{RV}_{sr}(\mathbf{\Omega}) \arrow[dl, "\mathrm{RV}_{sr}(f_i)"] \\
& \mathrm{RV}_{sr}(\mathbf{\Omega}_j) &
\end{tikzcd}$$

Therefore, the map

$$\lim_j \mathrm{RV}_r(\mathbf{\Omega}_j) \xrightarrow{\psi_s} \mathrm{RV}_{sr}(\mathbf{\Omega}) \to \lim_j \mathrm{RV}_{sr}(\mathbf{\Omega}_j)$$

is the inclusion map. By taking limit for $s \to 1$ in both cases, we can conclude that the canonical map $\mathrm{RV}_r(\mathbf{\Omega}) \to \lim_j \mathrm{RV}(\mathbf{\Omega}_j)$ is an isomorphism. $\square$

By the previous theorem, the functor $\mathrm{RV}_r$ preserves the limit of $D_{\mathbf{\Omega}}$. The limit of $\mathrm{RV}_r D_{\mathbf{\Omega}}$ can be constructed in the usual way we construct cofiltered limits in **CMet**.

The underlying set of the limit of $\mathrm{RV}_r D_{\mathbf{\Omega}}$ is given by

$$\left\{ (X_i)_{i \in I} \in \prod_{i \in I} \mathrm{RV}_r(\mathbf{\Omega}_i) \mid \mathrm{RV}_r(f_{ij})(X_j) = X_i \text{ for all } i \leq j \right\}$$

which is equal to

$$\left\{ (X_i)_{i \in I} \in \prod_{i \in I} \mathrm{RV}_r(\mathbf{\Omega}_i) \mid \mathbb{E}[X_j \mid f_{ij}] = X_i \right\}.$$

This means that the underlying set of $\lim \mathrm{RV}_r D_{\mathbf{\Omega}}$ is precisely the collection of martingales, uniformly bounded by $r$, on the filtered probability space $(\Omega, (\mathcal{F}_i)_{i \in I}, \mathcal{F}, \mathbb{P})$. Theorem 5.3.15 now says that the map

$$\mathrm{RV}_r(\mathbf{\Omega}) \to \lim_i \mathrm{RV}_r(\mathbf{\Omega}_i)$$

defined by the assignment

$$X \mapsto (\mathrm{RV}_r(f_i)(X))_{i \in I} = (\mathbb{E}[X \mid f_i])_{i \in I}$$

is an isomorphism. In other words, for every martingale $(X_i)_i$ there is a $\mathbb{P}$-almost surely unique random variable $X \in \mathrm{RV}_r(\mathbf{\Omega})$ such that

$$\mathbb{E}[X \mid f_i] = X_i.$$

This proves, *categorically*, the following weaker martingale convergence theorem.

**Theorem 5.3.16.** *Let $(X_i)_{i \in I}$ be a martingale such that for all $i \in I$,*

$$\mathbb{P}(X_i \leq r) = 1.$$

*Then there exists a unique $X \in \mathrm{RV}_r(\mathbf{\Omega})$ such that for all $i \in I$,*

$$\mathbb{E}[X \mid f_i] = X_i.$$

Theorem 5.3.15 also implies that the functor $\mathrm{M}_r$ preserves the limit of $D_{\mathbf{\Omega}}$. Also for this functor we can construct the cofiltered limit $\lim \mathrm{M}_r D_{\mathbf{\Omega}}$ in the usual way; its underlying set is given by

$$\left\{(\mu_i)_{i\in I} \in \prod_{i\in I} \mathrm{M}_r(\mathbf{\Omega}_i) \mid \mathrm{M}_r(f_{ij})(\mu_j) = \mu_i \text{ for all } i \leq j\right\}$$

which is equal to

$$\left\{(\mu_i)_{i\in I} \in \prod_{i\in I} \mathrm{M}_r(\mathbf{\Omega}_i) \mid \mu_j \mid_{\mathcal{F}_i} = \mu_i \text{ for all } i \leq j\right\}$$

Theorem 5.3.15 says that the map

$$\mathrm{M}_r(\mathbf{\Omega}) \to \lim_i \mathrm{M}_r(\mathbf{\Omega}_i)$$

defined by the assignment

$$\mu \mapsto (\mathrm{M}_r(f_i)(\mu))_{i\in I} = (\mu \mid_{\mathcal{F}_i})_{i\in I}$$

is an isomorphism. Therefore, for every family $(\mu_i)_{i\in I}$ of measures, where $\mu_i \in \mathrm{M}_r(\mathbf{\Omega}_i)$ such that for all $i \leq j$,

$$\mu_i \mid_{\mathcal{F}_j} = \mu_j,$$

there exists a unique $\mu \in \mathrm{M}_r(\mathbf{\Omega})$ such that

$$\mu \mid_{\mathcal{F}_i} = \mu_i.$$

This gives a categorical proof for the following version of the Kolmogorov extension theorem.

**Theorem 5.3.17.** *Consider a family $(\mu_i)_{i\in I}$ such that $\mu_i$ is a measure on $\mathbf{\Omega}_i$ and $\mu_i \leq r\mathbb{P}$. Suppose that for all $i \leq j$,*

$$\mu_j \mid_{\mathcal{F}_i} = \mu_i.$$

*Then there exists a unique measure $\mu$ on $\mathbf{\Omega}$ with $\mu \leq r\mathbb{P}$ such that for all $i \in I$,*

$$\mu \mid_{\mathcal{F}_i} = \mu_i.$$

**Remark 5.3.18.** Theorem 5.3.15 implies an even stronger result than Theorem 5.3.16 and 5.3.17. It not only says that for every martingale $(X_i)_i$ there exists a random variable $X$ such that $\mathbb{E}[X \mid f_i] = X_i$, but it says that this happens in an *isometric* way. In other words, $\sup_i d(X_i, Y_i) = d(X, Y)$ for martingales $(X_i)_i$ and $(Y_i)_i$ and their corresponding limiting random variables $X$ and $Y$. In a similar way, we have that the Kolmogorov extension from Theorem 5.3.17 is isometric too.

Furthermore, for a *consistent* family of measures $(\mu_i)_i$ and its limiting measure $\mu$ such as in Theorem 5.3.16, the collection of Radon–Nikodym derivates $\left(\frac{\mathrm{d}\mu_i}{\mathrm{d}\mathbb{P}}\right)_i$ form a martingale and the limiting random variable of this martingale is the Radon–Nikodym derivative of $\mu$ with respect to $\mathbb{P}$.

**Remark 5.3.19.** An alternative approach we could have taken is to define the category **Prob** as the category of probability spaces and *equivalence classes* of almost surely equal measure-preserving maps. Using this approach, we would not have to deal with the *pseudo*metric spaces.

# Appendix A

# Integral representation theorems

## A.1 Carathéodory extension theorem

Let $X$ be a set.

**Definition A.1.1.** A family $S$ of subsets of $X$ is called a **semi-ring** if it contains $\emptyset$ and is closed under finite intersections and such that the relative complement of a set in $S$ can be written as a finite union of sets in $S$.

Let $\sigma(S)$ denote the smallest $\sigma$-algebra on $X$ that contains $S$.

**Theorem A.1.2** (Carathéodory). *Let $\mu : S \to [0, \infty)$ be a map such that $\mathbb{P}(\emptyset) = 0$. Suppose that for every collections $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint elements in $S$ such that their union $A$ is also in $S$ we have $\mu(A) = \sum_{n\in\mathbb{N}} \mu(A_n)$. Then $\mu$ extends uniquely to a $\sigma$-finite measure $\rho$ on $(X, \sigma(S))$.*

A proof of this result can be found in [14] (Proposition 3.2.4).

## A.2 Integration lattices

Let $X$ be a set and let $f$ and $g$ be real-valued functions on $X$. We use the notation $f \vee g$ $(f \wedge g)$ to mean the pointwise maximum (minimum) of the two functions.

Let $L \subseteq [0, \infty)^X$ be a subset of non-negative valued functions on $X$. A subset $L$ is called an **integration lattice** if $1 \in L$ and for all $f, g \in L$ and $r \in [0, \infty)$ also $f \wedge g, f \vee g, rf$ and $f \vee g - f \wedge g$ are elements of $L$.

A map $I : L \to [0, \infty)$ is called an **integration operator** if $I(1) = 1$ and for every collection $(f_n)_n$ in $L$ such that $f := \sum_{n\in\mathbb{N}} f_n$ is also in $L$ we have $I(f) = \sum_{n\in\mathbb{N}} I(f_n)$.

**Lemma A.2.1.** *Let $I$ be a integration operator on a integration lattice $L$ and let $r \in [0, \infty)$. Then $I(rf) = rI(f)$.*

*Proof.* For $n, m \geq 1$ and $f \in L$ we find the following:

$$I\left(\frac{m}{n}f\right) = mI\left(\frac{1}{n}f\right) = \frac{m}{n}nI\left(\frac{1}{n}f\right) = \frac{m}{n}I(f).$$

It follows that the statement holds for $r \in [0,\infty) \cap \mathbb{Q}$. Now take any $r \in [0,\infty)$ and let $(q_n)_n$ be an increasing sequence of rational non-negative numbers that converges to $r$. We find that $rf = q_1 f + \sum_{n\in\mathbb{N}}(q_{n+1} - q_n)f$ and therefore we have that

$$I(rf) = I(q_1 f) + \sum_{n\in\mathbb{N}} I((q_{n+1} - q_n)f) = q_1 + \sum_{n\in\mathbb{N}}(q_{n+1} - q_n)I(f) = rI(f).$$

□

**Remark A.2.2.** Lemma A.2.1 is also true in the case that $I$ is a *weak* integration operator on a *weak* integration lattice $L$ and $r$ is an element of $[0,1]$.

The following result is a variant of the Daniell–Stone representation theorem and the proof follows Kindler's proof in [39] closely.

**Theorem A.2.3.** *Let $I$ be an integration operator on an integration lattice $L$. There exists a unique probability measure $\mathbb{P}$ on $(X, \sigma(L))$ such that*

$$I_{\mathbb{P}}(f) = I(f)$$

*for all $f \in L$.*

*Proof.* For $f, g \in L$ such that $f \le g$ define

$$[f,g) := \{(x,t) \in X \times [0,\infty) \mid f(x) \le t < g(x)\}$$

and let $S$ be the family of all subsets of $X \times [0,\infty)$ of this form.

For $f_1 \le g_1$ and $f_2 \le g_2$ in $L$ we have:

$$[f_1, g_1) \cap [f_2, g_2) = [f_1 \vee f_2, g_1 \wedge g_2)$$

and

$$[f_1, g_1) \setminus [f_2, g_2) = [f_1, g_1 \wedge f_2) \cup [f_1 \vee g_2, g_1).$$

Since we also clearly have that $\emptyset \in S$ and $L$ is closed under taking finite minima and maxima, we conclude that $S$ is a semi-ring.

Now define $\mu : S \to [0,\infty)$ by

$$\mu([f,g)) := I(g - f)$$

for all $f \le g$. Note that this is well-defined since $g - f = g \vee f - g \wedge f$, which is an element of $L$. We clearly have that $\mu(\emptyset) = \mu([f,f)) = I(0) = 0$. For a collection $([f_n, g_n))_{n\in\mathbb{N}}$ of pairwise disjoint subsets in $S$ such that $\bigcup_{n\in\mathbb{N}}[f_n, g_n) = [f,g)$ for some $f \le g$ in $L$ we can show that

$$g - f = \sum_{n\in\mathbb{N}}(g_n - f_n).$$

It follows that

$$\mu([f,g)) = I(g-f) = \sum_{n\in\mathbb{N}} I(g_n - f_n) = \sum_{n\in\mathbb{N}} \mu([f_n, g_n)).$$

It follows now from Theorem A.1.2 that $\mu$ can be extended uniquely to a $\sigma$-finite measure $\rho$ on $(X \times [0,\infty), \sigma(S))$.

We will now show that $\sigma(L) \otimes \mathrm{Bo}_{[0,\infty)} \subseteq \sigma(S)$. For $f \in L$ and $r \in [0,\infty)$ define $f_n :=$

$n(f \vee r - f) \wedge 1$ and note that this an element of $L$. For $s \in [0,\infty)$ we have

$$\bigcup_{n=1}^{\infty} [0, sf_n) = \{f < r\} \times [0, s).$$

Therefore every subset of the form $\{f < r\} \times [0,s)$ is in $\sigma(S)$. This shows that $\sigma(L) \otimes \mathrm{Bo}_{[0,\infty)} \subseteq \sigma(S)$.

Now define $\mathbb{P} : \sigma(L) \to [0,1]$ by $\mathbb{P}(A) := \rho(A \times [0,1))$. Note that for $r$ and $s$ in $[0,\infty)$ we have

$$\begin{aligned}
\rho(\{f < r\} \times [0,s)) &= \rho\left(\bigcup_{n=1}^{\infty} [0, sf_n)\right) \\
&= \lim_{n\to\infty} \rho([0, sf_n)) \\
&= \lim_{n\to\infty} I(sf_n) \\
&= \lim_{n\to\infty} sI(f_n) \\
&= s \lim_{n\to\infty} \rho([0, f_n)) \\
&= s\rho(\{f < r\} \times [0,1)) = s\mathbb{P}(\{f < r\})
\end{aligned}$$

Here we used that $[0, sf_n) \subseteq [0, sf_{n+1})$ and Lemma A.2.1. It follows that

$$\rho(\{r_1 \le f < r_2\} \times [0,s)) = s\mathbb{P}(\{r_1 \le f < r_2\}).$$

We have

$$\mathbb{P}(X) = \rho(X \times [0,1)) = \rho([0,1)) = I(1) = 1$$

and the $\sigma$-additivity of $\mathbb{P}$ is inherited from $\rho$. Therefore $\mathbb{P}$ is a probability measure on $(X, \sigma(L))$.

For a measurable map $f : X \to [0,\infty)$ there is an increasing sequence of non-negative simple functions $(s_n := \sum_{k=1}^{m_n} a_k^n 1_{A_k^n})_n$, with $A_k^n$ of the form $\{r_1 \le f < r_2\}$, that converges to $f$. Define

$$B_n := \bigcup_{k=1}^{m_n} A_k^n \times [0, a_k^n)$$

and note that $\bigcup_{n=1}^{\infty} B_n = [0, f)$ and that $B_n \subseteq B_{n+1}$ for all $n$. We have the following equalities:

$$I(f) = \rho([0,f)) = \lim_{n\to\infty} \rho(B_n) \tag{A.1}$$

and

$$\rho(B_n) = \sum_{k=1}^{m_n} \rho(A_k^n \times [0, a_k^n)) = \sum_{k=1}^{m_n} a_k^n \mathbb{P}(A_k^n) = I_{\mathbb{P}}(s_n) \tag{A.2}$$

Combining (A.1) and (A.2) with the monotone convergence theorem we conclude that $I(f) = I_{\mathbb{P}}(f)$.

Let $\mathbb{P}'$ be another probability measure with this property. Then we have for all $f \in L$ and $r \in [0,\infty)$ that

$$\mathbb{P}'(\{f < r\}) = \lim_{n\to\infty} I'_{\mathbb{P}}(f_n) = \lim_{n\to\infty} I_{\mathbb{P}}(f_n) = \mathbb{P}(\{f < r\}).$$

Since the sets $\{f < r\}$ form a $\pi$-system that generates $\sigma(L)$ we can conclude that $\mathbb{P} = \mathbb{P}'$. ☐

## A.3 Proofs of the results in section 3.5.2

*Proof of Theorem 3.5.5.* It is easy to verify that $\mathbb{N}L$ is an integration lattice. For an element $mf \in \mathbb{N}L$ define $I'(mf) := mI(f)$. Suppose $mf = ng$ for $f, g \in L$. Then we have by Remark A.2.2 that

$$mI(f) = (m \vee n)\frac{m}{m \vee n}I(f) = (m \vee n)I\left(\frac{m}{m \vee n}f\right) = (m \vee n)I\left(\frac{n}{m \vee n}g\right) = nI(g).$$

This shows that $I'$ is well-defined and that $I'(f) = I(f)$ for all $f \in L$.

It is clear that $I'(1) = 1$. Let $(m_n f_n)_{n\in\mathbb{N}}$ be a collection of elements in $\mathbb{N}L$ such that $mf := \sum_{n\in\mathbb{N}} m_n f_n$ is also an element of $\mathbb{N}L$. We observe that $f = \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} a(k,n)\frac{f_n}{m}$ where $a(k,n) = 1$ if $k \leq m_n$ and 0 otherwise. Note that $a(k,n)\frac{f_n}{m}$ is an element of $L$ and therefore

$$I(f) = \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} I\left(\frac{a(k,n)}{m}f\right) = \frac{1}{m}\sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} a(k,n)I\left(f\right) = \frac{1}{m}\sum_{n\in\mathbb{N}} m_n I(f_n).$$

We can conclude that $I'$ is an integration operator on the integration lattice $\mathbb{N}L$ and therefore by Theorem A.2.3 there is a unique probability measure on $(X, \sigma(\mathbb{N}L))$ such that $\int_X f\mathrm{d}\mathbb{P} = I'(f)$. It is clear that $\sigma(L) \subseteq \sigma(\mathbb{N}L)$. For $mf \in \mathbb{N}L$ and $r \in [0,\infty)$ we find that $\{mf < r\} = \{f < \frac{r}{m}\} \in \sigma(L)$ and therefore also $\sigma(\mathbb{N}L) \subseteq \sigma(L)$.

Let $\mathbb{P}'$ be another probability measure on $(X, \sigma(L))$ such that $\int_X f\mathrm{d}\mathbb{P}' = I(f)$ for all $f \in L$, then we have that

$$I'(mf) = mI(f) = m\int_X f\mathrm{d}\mathbb{P}' = \int_X mf\mathrm{d}\mathbb{P}'.$$

This implies that $\mathbb{P}' = \mathbb{P}$. ☐

**Lemma A.3.1.** *Let $I : L \to [0,\infty)$ be an additive operator on a weak integration lattice $L$. For $f, g \in L$ such that $f \leq g$ we have $I(f) \leq I(g)$, and if $g - f \in L$ then $I(g - f) = I(g) - I(f)$.*

*Proof.* There exists an $n$ such that $\frac{g-f}{n} \in L$. We find

$$\frac{1}{n}I(g) = I\left(\frac{g}{n}\right) = I\left(\frac{g-f}{n} + \frac{f}{n}\right) = I\left(\frac{g-f}{n}\right) + \frac{1}{n}I(f) \geq \frac{1}{n}I(f).$$

Here we use Remark A.2.2. The first part of the statement now follows. The second part can be proven in a similar way. ☐

*Proof of Proposition 3.5.6.* First note that for $f \in L$ and $n \geq 1$, then

$$\begin{aligned} I(f) &= I\left(\frac{(n-1)f}{n} + \frac{f}{n}\right) \\ &= I\left(\frac{(n-1)f}{n}\right) + I\left(\frac{f}{n}\right) \\ &= I\left(\frac{(n-2)f}{n}\right) + I\left(\frac{f}{n}\right) + I\left(\frac{f}{n}\right) \\ &= \ldots \\ &= nI\left(\frac{f}{n}\right). \end{aligned}$$

By Theorem 3.5.5 it is enough to show that $I$ is a weak integration operator. Let $(f_n)_{n\in\mathbb{N}}$ be a collection of elements in $L$ such that also $f := \sum_{n\in\mathbb{N}} f_n$ is an element of $L$. Define for every $n$ the function $g_n := \sum_{k\leq n} f_k$. For every $n$ there exists an $m_n$ such that $\frac{g_k}{m_n} \in L$ for all $k \leq n$. Indeed, for $n = 1$ this is clear. Suppose the claim holds for a natural number $n$, then $\frac{g_n}{m_n}$ and $\frac{f_{n+1}}{m_n}$ are element of $L$. Because

$$\frac{g_{n+1}}{m_n} = \frac{g_n}{m_n} + \frac{f_{n+1}}{m_n}$$

is an element of $\mathbb{N}L$, there exists an $m'$ such that $\frac{1}{m'}\frac{g_{n+1}}{m_n} \in L$. Since $\frac{g_k}{m_n} \in L$ for $k \leq n$, we also have that $\frac{g_k}{m'm_n} \in L$. It follows that for $m_{n+1} := m'm_n$ we have that $\frac{g_k}{m_{n+1}} \in L$ for all $k \leq n+1$. The claim now follows by induction. The functions $(g_n)_n$ form a sequence of continuous functions that increases pointwise to the continuous function $f$. Because $X$ is a compact Hausdorff space, Dini's theorem tells us that $(g_n)_n$ converges uniformly to $f$. Let $\epsilon > 0$. There exists an $n$ such that

$$\|f - g_n\| \leq \epsilon. \tag{A.3}$$

By the above there exists an $m_n$ such that $\frac{1}{m_n}g_k \in L$ for all $k \leq n$. Because $L$ is a weak integration lattice, there exists an $m$ such that

$$h_n := \frac{1}{m}\left(\frac{f}{m_n} - \frac{g_n}{m_n}\right)$$

is also an element of $L$. By (A.3) and the first part of Lemma A.3.1 we find that

$$I(h_n) \leq I\left(\frac{\epsilon}{m_n m}\right) = \frac{\epsilon}{m_n m}. \tag{A.4}$$

By the second part of Lemma A.3.1 we have

$$I\left(\frac{f}{m_n m}\right) - I\left(\frac{g_n}{m_n m}\right) = I(h_n). \tag{A.5}$$

Using (A.4) and (A.5) and the hypothesis we find the following:

$$\begin{aligned}
I(f) - \sum_{k\leq n} I(f_k) &= I(f) - \sum_{k\leq n} m_n I\left(\frac{f_k}{m_n}\right)\\
&= I(f) - m_n\left(I\left(\frac{f_1}{m_n}\right) + I\left(\frac{f_2}{m_n}\right) + I\left(\frac{f_3}{m_n}\right) + \ldots + I\left(\frac{f_n}{m_n}\right)\right)\\
&= I(f) - m_n\left(I\left(\frac{g_2}{m_n}\right) + I\left(\frac{f_3}{m_n}\right) + \ldots + I\left(\frac{f_n}{m_n}\right)\right)\\
&= I(f) - m_n\left(I\left(\frac{g_3}{m_n}\right) + \ldots + I\left(\frac{f_n}{m_n}\right)\right)\\
&= \ldots\\
&= I(f) - m_n I\left(\frac{g_n}{m_n}\right)\\
&= m_n m\left(I\left(\frac{f}{m_n m}\right) - I\left(\frac{g_n}{m_n m}\right)\right)\\
&= m_n m I(h_n) < \epsilon
\end{aligned}$$

Because $(I(f) - \sum_{k \leq m} I(f_k))_m$ is a decreasing sequence of positive numbers, we can conclude that $I(f) = \sum_{n \in \mathbb{N}} I(f_n)$. □

# Appendix B

# Lax coends

In this section we will give the definition of lax coends. More details about lax coends can be found in Chapter 7 of [48].

Let $\mathcal{A}$ and $\mathcal{B}$ be poset-enriched categories and let $S : \mathcal{A}^{\mathrm{op}} \times \mathcal{A} \to \mathcal{B}$ be an enriched functor.

**Definition B.0.1.** A **lax cowedge** $\omega$ consists of

- an object $B$ in $\mathcal{B}$,
- a morphism $\omega_A : S(A, A) \to B$ for every object $A$ in $\mathcal{A}$ and
- for every morphism $f : A_1 \to A_2$,

$$\begin{tikzcd}
S(A_1, A_2) \arrow[r, "S(f\times 1_{A_2})"] \arrow[d, "S(1_{A_1}\times f)"'] \arrow[dr, phantom, "\leq"] & S(A_2, A_2) \arrow[d, "\omega_{A_2}"] \\
S(A_1, A_1) \arrow[r, "\omega_{A_1}"'] & B
\end{tikzcd} .$$

**Definition B.0.2.** Let $\omega^1 := (B^1, (\omega^1_A)_A, (\omega^1_f)_f)$ and $\omega^2 := (B^2, (\omega^2_A)_A, (\omega^2_f)_f)$ be lax cowedges of $S$. A **morphism of cowedges** from $\omega^1$ to $\omega^2$ is a morphism $s : B_1 \to B_2$ such that

$$\begin{tikzcd}
S(A_1, A_2) \arrow[r, "S(f\times 1_{A_2})"] \arrow[d, "S(1_{A_1}\times f)"'] \arrow[dr, phantom, "\leq"] & S(A_2, A_2) \arrow[d, "\omega_{A_2}"] \arrow[ddr, "\omega^2_{A_2}"] & & & S(A_1, A_2) \arrow[r, "S(f\times 1_{A_2})"] \arrow[d, "S(1_{A_1}\times f)"'] \arrow[dr, phantom, "\leq"] & S(A_2, A_2) \arrow[d, "\omega^2_{A_2}"] \\
S(A_1, A_1) \arrow[r, "\omega_{A_1}"] \arrow[drr, "\omega^2_{A_1}"'] & B_1 \arrow[dr, "s"] & = & = & S(A_1, A_1) \arrow[r, "\omega^2_{A_1}"] & B_2 \\
& = & B_2 & & &
\end{tikzcd}$$

for all morphism $f : A_1 \to A_2$.

The category of lax cowedges of $S$ and their morphims is denoted by $\mathbf{Cowedge}_l(S)$.

**Definition B.0.3.** The lax coend of $S$ is an initial object in $\mathbf{Cowedge}_l(S)$.

If $(B, (\omega_A)_A, (\omega_f)_f)$ is the lax coend of $S$, i.e. the initial object in $\mathbf{Cowedge}_l(S)$, then we also write

$$B =: \oint^{A \in \operatorname{ob} \mathcal{A}} S(A, A).$$

Dually, we can define a category of colax cowedges of $S$, $\mathbf{Cowedge}_c(S)$ and define colax coends as initial objects in this category. The notation we will use for the colax coend of $S$ is $\oint^{A \in \operatorname{ob} \mathcal{A}} S(A, A)$.

# Appendix C

# Inequalities in the proof of Theorem 5.2.11

**Lemma C.0.1.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(A, p)$ and $(B, q)$ be finite probability spaces. Let $f : (\Omega, \mathcal{F}, \mathbb{P}) \to (A, p)$ and $g : (\Omega, \mathcal{F}, \mathbb{P}) \to (B, q)$ be measure-preserving maps. Let $s : (A, p) \to (B, q)$ be a measure-preserving map such that $sf = g$. Let $s_g^y$ and $s_f^y$ be such as in Theorem 5.2.11. The following equalities hold.*

*(i)* $s_g^y s_f^y = \sum_{b \in B} [q_g(y)](b) \sum_{a \in s^{-1}(b)} [q_f(y)](a) 1_{f^{-1}(a)}$.

*(ii)* $\mathbb{E}[s_f^y s_g^y] = \sum_{b \in B} [q_g(y)](b)^2 q_b$.

*(iii)* $(s_g^y)^2 = \sum_{b \in B} [q_g(y)](b)^2 1_{g^{-1}(b)}$ *and* $(s_f^y)^2 = \sum_{a \in A} [q_f(y)](a)^2 1_{g^{-1}(a)}$.

*(iv)* $\mathbb{E}[(s_g^y)^2] = \sum_{b \in B} [q_g(y)](b)^2 q_b$ *and* $\mathbb{E}[(s_f^y)^2] = \sum_{a \in A} [q_f(y)](a)^2 p_a$.

*(v)* $\mathbb{E}[(s_g^y)^2] \leq \mathbb{E}[(s_f^y)^2]$.

*(vi)* $\mathbb{E}[(s_f^y)^2] - \mathbb{E}[(s_g^y)^2] = \mathbb{E}[(s_f^y - s_g^y)^2]$.

*Proof.* For (i),

$$\begin{aligned} s_g^y s_f^y &= \sum_{a \in A, b \in B} [q_f(y)](a)[q_g(y)](b) 1_{f^{-1}(a) \cap g^{-1}(b)} \\ &= \sum_{a \in A, b \in B} [q_f(y)](a)[q_g(y)](b) 1_{f^{-1}(s^{-1}(b) \cap \{a\})} \\ &= \sum_{b \in B} \sum_{a \in s^{-1}(b)} [q_f(y)](a)[q_g(y)](b) 1_{f^{-1}(a)} \\ &= \sum_{b \in B} [q_g(y)](b) \sum_{a \in s^{-1}(b)} [q_f(y)](a) 1_{f^{-1}(a)} \end{aligned}$$

Integration (i) gives us

$$\mathbb{E}[s_f^y s_g^y] = \sum_{b \in B} [q_g(y)](b) \sum_{a \in s^{-1}(b)} [q_f(y)](a) p_a = \sum_{b \in B} [q_g(y)](b) q_g(y)(b) q_b$$

This implies (ii). The results (iii) and (iv) follow from (i) an (ii). For $q_b \neq 0$,

$$[q_g(y)](b)^2 = \left( \sum_{a \in s^{-1}(b)} [q_f(y)](a) \frac{p_a}{q_b} \right)^2 \leq \sum_{a \in s^{-1}(b)} [q_f(y)](a)^2 \frac{p_a}{q_b}.$$

Here we used that $x \mapsto x^2$ defines a convex function. Multiplying both sides by $q_b$ and summing over $b \in B$, gives us (v). For (vi)

$$\begin{aligned} \mathbb{E}[(s_f^y)^2] &= \mathbb{E}[(s_f^y - s_g^y + s_g^y)^2] && \text{(C.1)} \\ &= \mathbb{E}[(s_f^y - s_g^y)^2] + \mathbb{E}[(s_g^f)^2] + 2(\mathbb{E}[s_f^y s_g^2] - \mathbb{E}[(s_g^y)^2]) \quad = \mathbb{E}[(s_f^y - s_g^y)^2] + \mathbb{E}[(s_g^f)^2]. && \text{(C.2)} \end{aligned}$$

In the fourth equality we used that $\mathbb{E}[s_f^y s_g^2] - \mathbb{E}[(s_g^y)^2] = 0$ by (ii) and (iv). □

The first inequality in the proof Theorem 5.2.11 that we needed to show is exactly Lemma C.0.1(v). For the second inequality, we combine Lemma C.0.1(vi) with Jensen's inequality.

# Bibliography

[1] M. Alvarez-Manilla, A. Jung, and K. Keimel. The probabilistic powerdomain for stably compact spaces. *Theoret. Comput. Sci.*, 328(3):221–244, 2004.

[2] E. S. Andersen and B. Jessen. On the introduction of measures in infinite product sets. *Danske Vid. Selsk. Mat.-Fys. Medd.*, 25(4):1–8, 1948.

[3] T. Avery. Codensity and the Giry monad. *J. Pure Appl. Algebra*, 220(3):1229–1251, 2016.

[4] K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges, A Study of Finitely Additive Measures*, volume 109 of *Pure and Applied Mathematics*. Academic Press, 1983.

[5] R. Blackwell, G. M. Kelly, and A. J. Power. Two-dimensional monad theory. *J. Pure Appl. Algebra*, 59(1):1–41, 1989.

[6] V. I. Bogachev. *Measure Theory. Vol. I.* Springer-Verlag, Berlin, 2007.

[7] V. I. Bogachev. *Measure Theory. Vol. II.* Springer-Verlag, Berlin, 2007.

[8] F. Borceux. *Handbook of Categorical Algebra. 1: Basic Category Theory*, volume 50 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1994.

[9] F. Borceux and B. J. Day. A monadicity theorem. *Bull. Austral. Math. Soc.*, 17(2):291–296, 1977.

[10] V. Capraro and T. Fritz. On the axiomatization of convex subsets of Banach spaces. *Proc. Amer. Math. Soc.*, 141(6):2127–2135, 2013.

[11] D. L. Cohn. *Measure Theory.* Birkhäuser Boston, Inc., Boston, MA, 1993. Reprint of the 1980 original.

[12] E.-E. Doberkat. Eilenberg-Moore algebras for stochastic relations. *Inform. and Comput.*, 204(12):1756–1781, 2006.

[13] J. L. Doob. *Measure Theory*, volume 143 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.

[14] R. M. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[15] N. Dunford and J. T. Schwartz. *Linear Operators. Part I.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988.

[16] R. Durrett. *Probability—Theory and Examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, 2019.

[17] T. Fox. Coalgebras and Cartesian categories. *Comm. Algebra*, 4(7):665–667, 1976.

[18] D. H. Fremlin. *Measure Theory. Vol. 4: Topological Measure Spaces.* Torres Fremlin, Colchester, 2006.

[19] T. Fritz. Convex spaces i: Definition and examples. Preprint, `https://arxiv.org/abs/0903.5522`, 2009.

[20] T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Adv. Math.*, 370:107239, 2020.

[21] T. Fritz, T. Gonda, and P. Perrone. de Finetti's theorem in categorical probability. *J. Stoch. Anal.*, 2(4):Art. 6, 26, 2021.

[22] T. Fritz, T. Gonda, P. Perrone, and E.F. Rischel. Representable Markov categories and comparison of statistical experiments in categorical probability. *Theoret. Comput. Sci.*, 961:113896, 2023.

[23] T. Fritz and A. Klingler. The *d*-separation criterion in categorical probability. *J. Mach. Learn. Res.*, 24:Paper No. [46], 49, 2023.

[24] T. Fritz and P. Perrone. Bimonoidal structure of probability monads. In *Proceedings of the Thirty-Fourth Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIV)*, volume 341 of *Electron. Notes Theor. Comput. Sci.*, pages 121–149. Elsevier Sci. B. V., Amsterdam, 2018.

[25] T. Fritz and P. Perrone. A probability monad as the colimit of spaces of finite samples. *Theory Appl. Categ.*, 34(7):170–220, 2019.

[26] T. Fritz and P. Perrone. Monads, partial evaluations, and rewriting. In *The 36th Mathematical Foundations of Programming Semantics Conference, 2020*, volume 352 of *Electron. Notes Theor. Comput. Sci.*, pages 129–148. Elsevier Sci. B. V., Amsterdam, 2020.

[27] T. Fritz and P. Perrone. Stochastic order on metric spaces and the ordered Kantorovich monad. *Adv. Math.*, 366:107081, 2020.

[28] M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology and Analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 68–85. Springer, Berlin-New York, 1982.

[29] J. Goubault-Larrecq and X. Jia. Algebras of the extended probabilistic powerdomain monad. In *The Proceedings of ISDT 2019, the 8th International Symposium on Domain Theory and Its Applications*, volume 345 of *Electron. Notes Theor. Comput. Sci.*, pages 37–61. Elsevier B. V., Amsterdam, 2019.

[30] S. P. Gudder. A general theory of convexity. *Rend. Sem. Mat. Fis. Milano*, 49:89–96, 1979.

[31] P. R. Halmos. *Measure Theory.* D. Van Nostrand Co., Inc., New York, N. Y., 1950.

[32] C. Heunen, O. Kammar, S. Staton, and H. Yang. A convenient category for higher-order probability theory. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, Piscataway, NJ, 2017.

[33] B. Jacobs. Convexity, duality and effects. In *Theoretical computer science*, volume 323 of *IFIP Adv. Inf. Commun. Technol.*, pages 1–19. Springer, Berlin, 2010.

[34] B. Jacobs. From probability monads to commutative effectuses. *J. Log. Algebr. Methods Program.*, 94:200–237, 2018.

[35] P. T. Johnstone. Adjoint lifting theorems for categories of algebras. *Bull. London Math. Soc.*, 7(3):294–297, 1975.

[36] K. Keimel. The monad of probability measures over compact ordered spaces and its Eilenberg-Moore algebras. *Topology Appl.*, 156(2):227–239, 2008.

[37] G. M. Kelly. Basic concepts of enriched category theory. *Repr. Theory Appl. Categ.*, (10):vi+137, 2005. Reprint of the 1982 original.

[38] J. F. Kennison and Dion Gildenhuys. Equational completion, model induced triples and pro-objects. *J. Pure Appl. Algebra*, 1(4):317–346, 1971.

[39] J. Kindler. A simple proof of the Daniell-Stone representation theorem. *Amer. Math. Monthly*, 90(6):396–397, 1983.

[40] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer, London, second edition, 2014.

[41] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Co., New York, 1956.

[42] S. Lack and M. Shulman. Enhanced 2-categories and limits for lax morphisms. *Adv. Math.*, 229(1):294–356, 2012.

[43] C. W. Lamb. A short proof of the martingale convergence theorem. *Proc. Amer. Math. Soc.*, 38:215–217, 1973.

[44] W. Lawvere. The category of probabilistic mappings. Preprint, `https://ncatlab.org/nlab/files/lawvereprobability1962.pdf`, 1962.

[45] T. Leinster. Codensity and the ultrafilter monad. *Theory Appl. Categ.*, 28(13):332–370, 2013.

[46] D. A. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2017.

[47] F. E. J. Linton. Applied functorial semantics. II. In *Sem. on Triples and Categorical Homology Theory (ETH, Zürich, 1966/67)*, pages 53–74. Springer, Berlin, 1969.

[48] F. Loregian. *(Co)end Calculus*, volume 468 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2021.

[49] S. Mac Lane. *Categories for the Working Mathematician*. Graduate Texts in Mathematics, Vol. 5. Springer-Verlag, New York-Berlin, 1971.

[50] S. Moss and P. Perrone. Preprint, `https://arxiv.org/abs/2207.07353`.

[51] P. Perrone. Categorical probability and stochastic dominance in met- ric spaces. Thesis, `www.paoloperrone.org/phdthesis.pdf`, 2018.

[52] E. Riehl. *Category Theory in Context*. Aurora: Dover Modern Math Originals. Dover Publications, 2017.

[53] E. F. Rischel and T. Fritz. Infinite products and zero-one laws in categorical probability. *Compositionality*, 2(3):20, 2020.

[54] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[55] T. Sato. The Giry monad is not strong for the canonical symmetric monoidal closed structure on meas. Preprint, https://arxiv.org/abs/1612.05939, 2016.

[56] Z. Semadeni. *Monads and their Eilenberg-Moore Algebras in Functional Analysis*. Queen's Papers in Pure and Applied Mathematics, No. 33. Queen's University, Kingston, Ont., 1973.

[57] S. Staton and S. Uijlen. Effect algebras, presheaves, non-locality and contextuality. *Inform. and Comput.*, 261(part 2):336–354, 2018.

[58] E. M. Stein and R. Shakarchi. *Real Analysis: Measure theory, Integration, and Hilbert Spaces*, volume 3 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2005.

[59] M. H. Stone. Postulates for the barycentric calculus. *Ann. Mat. Pura Appl. (4)*, 29:25–30, 1949.

[60] T. Swirszcz. Monadic functors and categories of convex sets. Preprint No 70.

[61] T. Swirszcz. Monadic functors and convexity. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 22:39–42, 1974.

[62] F. van Breugel. The metric monad for probabilistic nondeterminism. Preprint, http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf, 2005.

[63] J. Šipoš. Integral with respect to a pre-measure. *Math. Slovaca*, 29(2):141–155, 1979.

[64] G. L. Wise and E. B. Hall. *Counterexamples in Probability and Real Analysis*. The Clarendon Press, Oxford University Press, New York, 1993.