**PROBABILITY MONADS**

ANDREAS KLINGLER

## 1. Introduction

These notes introduce a functorial description of probability distributions on arbitrary measurable spaces by the so-called *Giry monad* and the *Radon monad*. These monads generalize the definition of a probability functor for finite probability distributions, which we have already introduced in the lecture (see [8, 1.3.16]). As the name already reveals, these functors are equipped with an additional monad structure, i.e. a unit and a multiplication. This structure is useful to define categories like Kleisli categories. Going back to probability distributions, we will see that Kleisli categories are then a categorical way to defining stochastic maps between measurable spaces.

These notes are organized as follows: In Section 2, we will review monads and Kleisli categories and discuss both in the context of the finite probability monad. Especially, we show that its Kleisli category can be understood as the category of finite stochastic processes between sets. In the second part, we will consider different ways of defining a probability monad that captures more than just finite probability distributions. Especially, we will introduce the *Giry monad* (Section 3) and the *Radon monad* (Section 4).

## 2. Monads, Kleisli categories and its application to the finite probability monad

In the following, we will briefly review the concept of monads and Kleisli categories and introduce the example of the finite probability monad. For a more detailed exposition of this topic we refer to [8, Section 5.1].

### 2.1. **Monads and the Kleisli category.**

**Definition 1.** *Let* $\mathbf{C}$ *be a category. A* monad *on* $\mathbf{C}$ *consists of*

*(i) a functor* $T : \mathbf{C} \to \mathbf{C}$*,*
*(ii) a natural transformation* $\eta : \mathrm{id}_{\mathbf{C}} \Longrightarrow T$*, called* unit*,*
*(iii) a natural transformation* $\mu : TT \Longrightarrow T$*, called* multiplication *or* composition*.*

*such that the following three diagrams commute:*

$$(1)\qquad \begin{tikzcd} T \arrow[r, Rightarrow, "\eta T"] \arrow[dr, Rightarrow, "\mathrm{id}"'] & TT \arrow[d, Rightarrow, "\mu"] \\ & T \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, Rightarrow, "T\eta"] \arrow[dr, Rightarrow, "\mathrm{id}"'] & TT \arrow[d, Rightarrow, "\mu"] \\ & T \end{tikzcd} \qquad \begin{tikzcd} TTT \arrow[r, Rightarrow, "T\mu"] \arrow[d, Rightarrow, "\mu T"'] & TT \arrow[d, Rightarrow, "\mu"] \\ TT \arrow[r, Rightarrow, "\mu"] & T \end{tikzcd}$$

*E-mail address*: `andreas.klingler@uibk.ac.at`.

Given a monad $T$ on $\mathbf{C}$, the Kleisli category $\mathbf{Kl}(T)$ is the category consisting of the same objects as $\mathbf{C}$ and morphisms $f : X \to TY$, $g : Y \to TZ$ with composition $g \circ_{kl} f : X \to TZ$ given by

$$X \xrightarrow{f} TY \xrightarrow{Tg} TTZ \xrightarrow{\mu} TZ.$$

Assuming that $f : X \to Y$ being a morphism in $\mathbf{C}$, then the construction of Kleisli categories allows to construct a more generalized notion of function which contains functions like $f$ by associating them with $\eta_Y \circ f$. In the case of probability distributions, we will see that if $X, Y$ are finite sets, then $f : X \to TY$ is a map to the set of all probability distributions in $Y$ which can be understood as a stochastic process.

2.2. **The finite probability monad.** In the following, we want to highlight the use of the monad structure for the finite probability monad.

We define the finite probability functor as

$$P : \mathbf{Set} \to \mathbf{Set}$$

where $PX = \{p : X \to [0,1] : \operatorname{supp}(p) \text{ finite}\}$ and for $f : X \to Y$, $Pf := f_*$ is the pushforward, defined by

$$(f_* p)(y) = \sum_{\substack{x \in X: \\ f(x) = y}} p(x).$$

The unit is given by $\eta_X : X \to PX : x \mapsto \delta_x$, where $\delta_x$ is the Dirac distribution on $x$. Further, the multiplication is defined as

$$\mu_X : PPX \to PX : \mathfrak{p} \mapsto \sum_{q \in PX} q \cdot \mathfrak{p}(q) = \mathbb{E}_{\mathfrak{p}}[q].$$

Therefore, for a given (finite) distribution $\mathfrak{p}$ over the set of all probability distributions $PX$, $\mu_X(\mathfrak{p})$ computes the mean distribution over all probability distributions with respect to $\mathfrak{p}$.

A nice interpretation of this map can be found in [8, Example 5.1.2]: Assume your space of distributions $PX$ models two different coins: a fair one which faces on one side head and on the other tail, and an unfair coin facing on both sides tail. Now you pick a coin randomly and flip it. What is the distribution of heads and tails? This distribution can be computed as $\mathbb{E}_{\mathfrak{p}}[p]$ where $\mathfrak{p}$ is the uniform distribution over the distributions of the first and second coin.

**Proposition 2.** *$(P, \eta, \mu)$ forms a monad.*

*Proof.* We start by showing that $\eta$ is a natural transformation. This is indeed the case since

$$\delta_{f(x)}(y) = \sum_{\substack{\tilde{x} \in X \\ f(\tilde{x}) = y}} \delta_x(\tilde{x}) = (f_* \delta_x)(y).$$

To show that $\mu$ is a natural transformation, we have to verify that

$$\begin{tikzcd} PPX \arrow[r, "f_{**}"] \arrow[d, "\mu_X"'] & PPY \arrow[d, "\mu_Y"] \\ P(X, \Sigma_X) \arrow[r, "f_*"'] & P(Y, \Sigma_Y) \end{tikzcd}$$

commutes for every $f : X \to Y$ where $f_{**} := (f_*)_* = PPf$. In other words, we have to show for every $\mathcal{M} \in PPX$ that $f_*\mu_X(\mathcal{M}) = \mu_X(f_{**}\mathcal{M})$. For $y \in Y$ we have

$$\begin{aligned}
\mu_X(f_{**}\mathcal{M})(y) &= \sum_{q\in PY} \mathcal{M}(f_*q) \cdot q(y) = \sum_{q\in PY} \sum_{\substack{p\in PX \\ f_*p=q}} \mathcal{M}(p) \cdot q(y) \\
&= \sum_{p\in PX} (f_*p)(y) \cdot \mathcal{M}(p) = \sum_{\substack{x\in X \\ f(x)=y}} \sum_{p\in PX} p(x) \cdot \mathcal{M}(p) \\
&= \sum_{\substack{x\in X \\ f(x)=y}} \mu_X(\mathcal{M})(x) = (f_*\mu_X)(\mathcal{M})(y)
\end{aligned}$$

where we merged both sums in the third equality and used the definition of $f_*, f_{**}$ and $\mu_X$ in the remaining steps.

It remains to show that the diagrams in (1) commute. The first diagram commutes since for every $p \in PX$ we have that

$$\mu_X(\delta_p) = \sum_{q\in PX} q \cdot \delta_p(q) = p.$$

For the commutativity of the second diagram we observe that for $p, q \in PX$

$$\big((\eta_X)_*q\big)(p) = \begin{cases} q(x) & : \text{if } p = \delta_x \\ 0 & : \text{else.} \end{cases}$$

Therefore,

$$\mu_X\big((\eta_X)_*q\big) = \sum_{p\in PX} p \cdot \big((\eta_X)_*q\big)(p) = \sum_{x\in X} \delta_x q(x) = q$$

which shows the commutativity.

For the last diagram we have to check that $\mu_X \circ P\mu_X = \mu_X \circ \mu_{PX}$. Observe that for $\mathcal{M} \in PPPX$ and $q \in PX$ we have that

$$\big((\mu_X)_*\mathcal{M}\big)(q) = \sum_{\substack{\mathfrak{p}\in PPX \\ \mu_X(\mathfrak{p})=q}} \mathcal{M}(\mathfrak{p})$$

Therefore, we have for every $\mathcal{M} \in PPPX$

$$\begin{aligned}
(\mu_X \circ P\mu_X)(\mathcal{M}) &= \sum_{q\in PX} q \cdot \big((\mu_X)_*\mathcal{M}\big)(q) = \sum_{\mathfrak{p}\in PPX} \mathcal{M}(\mathfrak{p}) \cdot \mu_X(\mathfrak{p}) \\
&= \sum_{q\in PX} \sum_{\mathfrak{p}\in PPX} q \cdot \mathfrak{p}(q) \cdot \mathcal{M}(\mathfrak{p}) = \sum_{q\in PX} q \cdot \mu_{PX}(\mathcal{M}) = \mu_X \circ \mu_{PX}(\mathcal{M})
\end{aligned}$$

which shows the statement. □

2.3. **The Kleisli category of $P$.** We are now able to define the Kleisli category of the finite probability monad. By definition, a Kleisli morphism is given $f : X \to PY$ mapping each $x \in X$ to a finite probability distribution $p_x$ on $Y$. Therefore, $f$ can be seen as a stochastic map

$$X \to PY : x \mapsto f(-|x).$$

Note that for a given morphism $f : X \to PY$, the induced morphism $Pf$ is given by

$$f_* := Pf : PX \to PPY : (p \mapsto f_*p).$$

We will now compute the Kleisli-composition of morphisms $f : X \to PY$ and $g : Y \to PZ$. First, the composition $Pg \circ f$ is given by

$$X \to PPZ : x \mapsto \Big( q \mapsto \sum_{\substack{y \in Y \\ g(-|y)=q}} q \cdot f(y|x) \Big)$$

This finally leads to the Kleisli-composition

$$g \circ_{kl} f : X \to PZ : x \mapsto \Big( z \mapsto \sum_{y \in Y} g(z|y) f(y|x) \Big)$$

which is precisely the composition of finite stochastic processes, also called *Chapman-Kolmogorov formula*. Therefore, we can interpret the cateogry $\mathbf{Kl}(P)$ as a category of sets with finite stochastic processes acting as morphisms between these sets.

## 3. The Giry Monad

In the last section, we have restricted to finite probability distributions. In this situation, it was possible to specify probability distributions by their probability taking a specific point. In the following, we want to consider probability measures taking infinitely (even uncountably) many values. In this situation, it is no longer possible to specify each point's probability separately. We have to define a function, which assigns to each subset a "volume" and behaves in a reasonable way (for example, being additive under the union of disjoint subsets).

Take, for example, the uniform distribution on the interval $[0,1]$. Then, we know that $p([a,b]) = b - a$ for all real numbers $b > a$ in $[0,1]$. It would be convenient to extend this condition to a function $p : \mathcal{P}([0,1]) \to [0,1]$ satisfying that this function is countably additive and translational invariant. Surprisingly, using the axiom of choice, it can be shown that such an extension does not exist (see for example [1, Section 1.7]). Therefore, it is, in general, a too strong requirement asking a measure to be defined on all subsets. A probability measure has to be defined on a convenient collection of subset $\Sigma_{[0,1]} \subseteq \mathcal{P}([0,1])$ which are called $\sigma$-algebras. The collection $(X, \Sigma_X)$, where $X$ is any set and $\Sigma_X$ a $\sigma$-algebra is called a measurable space. We refer to [5, Chapter I - IV] for the discussion of basic notions in measure theory like $\sigma$-algebras, measurable functions, integrals over measurable spaces and basic results like the dominated convergence and the monotone convergence theorem which will be extensively used throughout this section.

In the following, we will introduce the *Giry monad*, which precisely captures all probability distributions over arbitrary measurable spaces. A detailed discussion of this monad can be found in several works, for example [7, Section 5.2] or [3, Section 4], among others. This monad will act on the category of *measurable spaces* **Meas** specified in the following way:

(i) The objects are measurable spaces $(X, \Sigma_X)$ where $\Sigma_X$ is a $\sigma$-algebra over $X$.
(ii) The morphisms $f : X \to Y$ are measurable maps, i.e. $f^{-1}(\Sigma_Y) \subseteq \Sigma_X$.

We define the Giry monad as a functor

$$P : \mathbf{Meas} \to \mathbf{Meas} \quad \text{with} \quad P(X, \Sigma_X) = \Big( P(X, \Sigma_X), \Sigma_{PX} \Big)$$

where

$$P(X, \Sigma_X) := \Big\{ p : \Sigma_X \to [0,1] : p \text{ is a prob. measure} \Big\}$$

and $\Sigma_{PX}$ is the smallest $\sigma$-algebra over $P(X, \Sigma_X)$ which makes the evaluation map

$$\mathsf{ev}_A : P(X, \Sigma_X) \to [0,1] : p \mapsto p(A)$$

measurable for every $A \in \Sigma_X$. In other words:

$$\Sigma_{PX} = \sigma\Big(\Big\{\mathsf{ev}_A^{-1}([0,r]) : 0 \le r \le 1, A \in \Sigma_X\Big\}\Big)$$

where

$$\sigma(U) := \bigcap_{\substack{U \subseteq \Sigma \\ \sigma\text{-algebra}}} \Sigma.$$

For a morphism $f : X \to Y$ (i.e. a measurable function), we define

$$Pf := f_* : P(X, \Sigma_X) \to P(Y, \Sigma_Y) : p \mapsto f_* p := p \circ f^{-1}$$

This assignment is well-defined, i.e. $f_* p$ is indeed a probability measure since $f$ is measurable. $Pf$ is also a valid morphism, as the following lemma shows.

**Lemma 3.** *Let $f : X \to Y$ be $\Sigma_X$ - $\Sigma_Y$ measurable. Then,*

$$f_* : P(X, \Sigma_X) \to P(Y, \Sigma_Y) : p \mapsto f_* p$$

*is $\Sigma_{PX}$ - $\Sigma_{PY}$ measurable.*

*Proof.* Since the pre-images of the evaluation maps generate $\Sigma_{PY}$, we have to show that for every $A \in \Sigma_Y$ and every $0 \le r \le 1$

$$(f_*)^{-1}\left(\mathsf{ev}_A^{-1}([0,r])\right) \in \Sigma_{PX}.$$

Note that for every $p \in P(X, \Sigma_X)$ and $A \in \Sigma_Y$ we have that

$$(\mathsf{ev}_A \circ f_*)(p) = \mathsf{ev}_A(f_* p) = p(f^{-1}(A)) = \mathsf{ev}_{f^{-1}(A)}(p).$$

Therefore,

$$(f_*)^{-1}\left(\mathsf{ev}_A^{-1}([0,r])\right) = (\mathsf{ev}_A \circ f_*)^{-1}([0,r]) = \mathsf{ev}_{f^{-1}(A)}^{-1}([0,r]) \in \Sigma_{PX}. \qquad \square$$

Since unitality and functoriality of $P$ are immediate we can conclude that $P$ is a functor.

Before starting with the construction of the monad structure, we present a lemma which will be very useful for further discussion. It relates integrals over different measures connected via a pushforward $f_*$ and can therefore be understood as a generalization of the change-of-variables formula for integrals over arbitrary measurable spaces. For a proof we refer to [1, Theorem 3.6.1].

**Lemma 4.** *Let $f : X \to Y$ and $g : Y \to \mathbb{R}$ be two measurable functions. $g$ is integrable with respect to $f_* p \in P(Y, \Sigma_Y)$ if and only if $g \circ f$ is integrable with respect to $p \in P(X, \Sigma_X)$. Furthermore, we have*

$$\int_X (g \circ f)(x)\, \mathrm{d}p(x) = \int_Y g(y)\, \mathrm{d}(f_* p)(y).$$

We are now able to introduce the monad structure on $P$. In a nutshell, the idea behind the construction of the two morphisms $\eta, \mu$ is very similar to the finite probability monad. Again, the unit morphism $\eta_X$ embeds every element $x \in X$ into the probability space $P(X, \Sigma_X)$ by associating it with the the Dirac measure $\delta_x$. Furthermore, the multiplication component $\mu_X : PPX \to PX$ maps a distribution on the space $P(X, \Sigma_X)$ to the mean distribution with respect to the given distribution.

We first show that $\eta_X$ is indeed a morphism, i.e. it is measurable. Let $0 \le r \le 1$ and $A \in \Sigma_X$. Then,

$$\eta_X^{-1}\left((\mathsf{ev}_A)^{-1}([0,r])\right) = \left\{ \begin{array}{ll} A^c & r < 1 \\ X & r = 1 \end{array} \right. \in \Sigma_X$$

since $\delta_x(A) \in \{0, 1\}$.

Further, $\eta$ is a natural transformation, since

$$\delta_{f(x)}(A) = \delta_x(f^{-1}(A)) = (f_*\delta_x)(A)$$

for every $x \in X$, morphism $f : X \to Y$ and $A \in \Sigma_Y$. Therefore, the naturality square

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\eta_X"'] & Y \arrow[d, "\eta_Y"] \\
P(X, \Sigma_X) \arrow[r, "f_*"'] & P(Y, \Sigma_Y)
\end{tikzcd}$$

commutes.

We define the multiplication $\mu : PP \Rightarrow P$ via

$$\mu_X : PPX \to PX : \mathcal{M} \mapsto \left( A \mapsto \int_{P(X,\Sigma_X)} p(A) \, \mathrm{d}\mathcal{M}(p) \right)$$

We will prove that $\mu$ is a natural transformation in three steps.

(i) $\mu_X(\mathcal{M})$ is a measure for all $\mathcal{M} \in PPX$: Obviously, $\mu_X(\mathcal{M}) : \Sigma_X \to [0,1]$ and

$$\mu_X(\mathcal{M})(X) = \int_{P(X,\Sigma_X)} p(X) \, \mathrm{d}\mathcal{M}(p) = \int_{P(X,\Sigma_X)} 1 \, \mathrm{d}\mathcal{M}(p) = 1$$

since $\mathcal{M}$ is a probability measure. Furthermore, let $\{A_i\}_{i \in \mathbb{N}}$ be a countable set of disjoint sets in $\Sigma_X$. Then

$$\begin{aligned}
\mu_X(\mathcal{M}) \left( \bigcup_{i=1}^{\infty} A_i \right) &= \int_{P(X,\Sigma_X)} p \left( \bigcup_{i=1}^{\infty} A_i \right) \, \mathrm{d}\mathcal{M}(p) \\
&= \int_{P(X,\Sigma_X)} \sum_{i=1}^{\infty} p(A_i) \, \mathrm{d}\mathcal{M}(p) \\
&= \sum_{i=1}^{\infty} \left( \int_{P(X,\Sigma_X)} p(A_i) \, \mathrm{d}\mathcal{M}(p) \right) = \sum_{i=1}^{\infty} \mu_X(\mathcal{M})(A_i)
\end{aligned}$$

where the second equality uses the fact that $p$ is a measure and the third equality holds by the dominated convergence theorem, since $p \mapsto \sum_{i=1}^{n} p(A_i)$ is measurable by definition of $\Sigma_{PX}$ and bounded by 1 for every $n \in \mathbb{N}$.

(ii) $\mu_X$ is a morphism, i.e. it is $\Sigma_{PPX}$ - $\Sigma_{PX}$ measurable: By definition of $\Sigma_{PX}$ it suffices to show that for every $A \in \Sigma_X$ and $0 \leq r \leq 1$

$$\mu_X^{-1}\left(\mathsf{ev}_A^{-1}([0,r])\right) \in \Sigma_{PPX}$$

In other words, we have to show that the map

$$\mathsf{ev}_A \circ \mu_X : PPX \to [0,1] : \mathcal{M} \mapsto \int_{P(X,\Sigma_X)} p(A) \, \mathrm{d}\mathcal{M}(p)$$

is measurable for every $A \in \Sigma_X$. Note that for every $\mathcal{A} \in \Sigma_{PX}$ we have that

$$\mathcal{M} \mapsto \mathsf{ev}_{\mathcal{A}}(\mathcal{M}) = \mathcal{M}(\mathcal{A}) = \int_{P(X,\Sigma_X)} F(p) \, \mathrm{d}\mathcal{M}(p)$$

where $F = 1_{\mathcal{A}}$ is measurable by definition of the $\sigma$-algebra of $PPX$. This implies that extending $F$ to any simple function $P(X, \Sigma_X) \to [0,1]$ keeps the mapping still measurable. Since $\mu_X$ can be represented by choosing $F := p \mapsto p(A)$ we can deduce that $\mu_X$ is measurable by applying the dominated convergence theorem and using the fact that the limit of measurable functions is still measurable.

(iii) $\mu$ is natural: For every morphism $f : X \to Y$, we define the induced morphism $f_{**} := PPf$ which reads

$$f_{**} : PPX \to PPY : \mathcal{M} \mapsto \mathcal{M} \circ f_*^{-1}$$

Then, for every $A \in \Sigma_Y$, every $\mathcal{M} \in PPX$ and every morphism $f : X \to Y$ we have

$$\begin{aligned}(f_* \mu_X(\mathcal{M}))(A) &= \int_{P(X,\Sigma_X)} p(f^{-1}(A))\, \mathrm{d}\mathcal{M}(p) \\ &= \int_{P(X,\Sigma_X)} \mathsf{ev}_A \circ f_*\, \mathrm{d}\mathcal{M} \\ &= \int_{P(Y,\Sigma_Y)} \mathsf{ev}_A\, \mathrm{d}\,(f_{**}\mathcal{M}) \\ &= \int_{P(Y,\Sigma_Y)} q(A)\, \mathrm{d}\,(f_{**}\mathcal{M})\,(q) = \mu_X(f_{**}\mathcal{M})(A)\end{aligned}$$

where we have used Lemma 4 on the measurable spaces $P(X, \Sigma_X)$ and $P(Y, \Sigma_Y)$. This shows that the naturality square

$$\begin{tikzcd} PPX \arrow[r, "f_{**}"] \arrow[d, "\mu_X"'] & PPY \arrow[d, "\mu_Y"] \\ P(X,\Sigma_X) \arrow[r, "f_*"'] & P(Y,\Sigma_Y) \end{tikzcd}$$

commutes for every $f : X \to Y$.

We are left to show that $(P, \eta, \mu)$ forms a monad, i.e. it satisfies Eq. (1). Showing this statement we need again a technical lemma.

**Lemma 5.** *Let $f : PX \to \mathbb{R}$ be bounded and measurable.*[1] *Then,*

$$\int_{PX} f\, \mathrm{d}(\mu_{PX}(\mathfrak{M})) = \int_{PPX} \left( \int_{PX} f\, \mathrm{d}\mathcal{M} \right) \mathrm{d}\mathfrak{M}(\mathcal{M})$$

*Proof.* To show this statement, we use a commonly used proof technique in measure theory which consists of the following steps:

(i) First, we show the statement for the characteristic function $f = 1_{\mathcal{A}}$ on every measurable set $\mathcal{A} \in \Sigma_{PX}$
(ii) By linearity, this implies that the statement holds for all simple functions (i.e. finite linear combinations of characteristic functions).
(iii) To proof the statement for arbitrary measurable $f$, we use the fact that every measurable function can be approximated arbitrarily well by simple functions.

We start by showing the statement for simple functions. Let $f = 1_{\mathcal{A}}$, where $\mathcal{A} \in \Sigma_{PX}$. Then,

$$\int_{PX} 1_{\mathcal{A}}\, \mathrm{d}(\mu_{PX}(\mathfrak{M})) = \big(\mu_{PX}(\mathfrak{M})\big)(\mathcal{A}) = \int_{PPX} \mathcal{M}(\mathcal{A})\, \mathrm{d}\mathfrak{M}(\mathcal{M}) = \int_{PPX} \left( \int_{PX} 1_{\mathcal{A}}\, \mathrm{d}\mathcal{M} \right) \mathrm{d}\mathfrak{M}(\mathcal{M}).$$

By linearity, this shows the statement for any simple function. Now, let $s_i \nearrow f$ a monotone increasing sequence of simple functions converging to $f$. Then, by monotone convergence theorem, we have

$$S_i(\mathcal{M}) := \int_{PX} s_i\, \mathrm{d}\mathcal{M} \nearrow S(\mathcal{M}) = \int_{PX} f\, \mathrm{d}\mathcal{M} \tag{2}$$

[1] This is a sufficient criterion being integrable with respect to all probability measures on $PX$

Therefore,

$$\begin{aligned}\int_{PX} f\,\mathrm{d}(\mu_{PX}(\mathfrak{M})) &= \lim_{i\to\infty}\int_{PX} s_i\,\mathrm{d}(\mu_{PX}(\mathfrak{M}))\\ &= \lim_{i\to\infty}\int_{PPX} S_i(\mathcal{M})\,\mathrm{d}\mathfrak{M}(\mathcal{M})\\ &= \int_{PPX} S(\mathcal{M})\,\mathrm{d}\mathfrak{M}(\mathcal{M}) = \int_{PPX}\left(\int_{PX} f\,\mathrm{d}\mathcal{M}\right)\mathrm{d}\mathfrak{M}(\mathcal{M})\end{aligned}$$

where we have used Eq. (2) in the first equality, the statement for simple functions in the second equality and monotone convergence together with Eq. (2) in the third equality. □

We are now able to proof the main statement.

**Theorem 6.** *$(P,\eta,\mu)$ forms a monad.*

*Proof.* Since we have already shown that $P$ is a functor and $\eta,\mu$ are natural transformations, it remains to show that the diagrams in Eq. (1) commute.

(i) We check $\mu_X\circ\eta_{PX} = \mathrm{id}_{PX}$. Let $p\in P(X,\Sigma_X)$. Then, for every $A\in\Sigma_X$ we have

$$(\mu_X\circ\eta_{PX}(p))(A) = \mu_X(\delta_p)(A) = \int_{P(X,\Sigma)} q(A)\,\mathrm{d}\delta_p(q) = p(A).$$

(ii) We check that $\mu_X\circ P\eta_X = \mathrm{id}_{PX}$. Let $x\in X$. Then, for every $A\in\Sigma_X$ we have

$$\begin{aligned}(\mu_X\circ P\eta_X(p))\,(A) &= \int_{P(X,\Sigma_X)} \mathrm{ev}_A\,\mathrm{d}((\eta_X)_*p)\\ &= \int_X \mathrm{ev}_A\circ\eta_X\,\mathrm{d}p\\ &= \int_X \delta_x(A)\,\mathrm{d}p(x) = p(A)\end{aligned}$$

where we have used in the second equality Lemma 4.

(iii) We have to check that $\mu_X\circ P\mu_X = \mu_X\circ\mu_{PX}$. We start by manipulating the left-hand side. Let $\mathfrak{M}\in PPPX$ and $A\in\Sigma_X$. Then

$$\begin{aligned}(\mu_X\circ P\mu_X)(\mathfrak{M})(A) &= \int_{P(X,\Sigma_X)} \mathrm{ev}_A\,\mathrm{d}((\mu_X)_*\mathfrak{M})\\ &= \int_{PPX} \mu_X(\mathcal{M})(A)\,\mathrm{d}\mathfrak{M}(\mathcal{M})\\ &= \int_{PPX}\left(\int_{P(X,\Sigma_X)} p(A)\,\mathrm{d}\mathcal{M}(p)\right)\mathrm{d}\mathfrak{M}(\mathcal{M})\end{aligned}$$

On the other hand, the right-hand side yields

$$\begin{aligned}(\mu_X\circ\mu_{PX})(\mathfrak{M})(A) &= \int_{P(X,\Sigma)} \mathrm{ev}_A\,\mathrm{d}(\mu_{PX}(\mathfrak{M}))\\ &= \int_{PPX}\left(\int_{P(X,\Sigma_X)} \mathrm{ev}_A\,\mathrm{d}\mathcal{M}\right)\mathrm{d}\mathfrak{M}(\mathcal{M})\\ &= \int_{PPX}\left(\int_{P(X,\Sigma_X)} p(A)\,\mathrm{d}\mathcal{M}(p)\right)\mathrm{d}\mathfrak{M}(\mathcal{M})\end{aligned}$$

Comparing both manipulations shows the statement. □

This concludes the construction of the Giry monad. Similar to the finite probability monad we would now be able to again construct the Kleisli category $\mathbf{Kl}(P)$ which we will skip here. In a nutshell, it can be shown that the category $\mathbf{Kl}(P)$ captures all possible Markov kernels between measurable spaces $(X, \Sigma_X)$ and $(Y, \Sigma_Y)$. Again, composition of morphisms is given by the Chapman-Kolmogorov formula. We refer to [3, Section 4] for further details.

## 4. The Radon Monad

Although we are now already equipped with a monad structure capturing all probability distributions on measurable spaces there are other approaches describing particular classes of probability distributions. In this section, we will construct the Radon monad $\mathcal{R}$ which captures all Radon probability distributions on compact Hausdorff spaces. We will explicitly show that $\mathcal{R}$ is a functor on the category of compact Hausdorff spaces **CHaus** (consisting of compact Hausdorff spaces as objects and continuous functions as maps) and introduce unit and multiplication, which are of a similar flavor to the discussed monadic structure in the Giry monad. Similar to the other probability monads, the corresponding Kleisli category $\mathbf{Kl}(\mathcal{R})$ captures a certain class of stochastic maps. In this case the construction would characterize the class of continuous Markov kernels, i.e. conditional probability distributions which are continuous in the condition variable. For a detailed exposition of topological Hausdorff spaces and Radon measures, we refer for example to [2, Chapter 1] and [10, Chapter 2]. Different (extended) constructions of the Radon monad can also be found in [6, Section 2.1], [4] or [3, Section 5].

4.1. **Borel and Radon measures.** We are now considering a subclass of measurable spaces and measures on them, namely measures on space of the form $(X, \mathcal{B}(X))$ where $X$ is a compact topological Hausdorff space and $\mathcal{B}(X)$ being the Borel $\sigma$-algebra generated by the topology of $X$, i.e.

$$\mathcal{B}(X) = \sigma\Big(\{U \subseteq X : U \text{ open}\}\Big).$$

Measures

$$\mu : \mathcal{B}(X) \to [0, \infty]$$

are called *Borel measures*.

Here, we restrict to *Radon measures* which are Borel measures behaving “continuous” under changing the measurable subsets $U \in \mathcal{B}(X)$. In other words, $\mu$ is called a Radon measure, if it is

(i) inner regular, i.e. for any open set $U \subseteq X$, $\mu(U) = \sup\{\mu(K) : K \subseteq U \text{ compact}\}$.
(ii) finite, i.e. $\mu(X) < \infty$.

Consider in the following the integral of a continuous function $f \in \mathcal{C}(X, \mathbb{R})$ with respect to the Radon measure $\mu$. This defines a function

$$I_\mu : \mathcal{C}(X, \mathbb{R}) \to \mathbb{R} : f \mapsto \int_X f \, \mathrm{d}\mu.$$

Since $\mathcal{C}(X, \mathbb{R})$ is a Banach space (using the supremum-norm), $I_\mu$ can even be understood as a linear operator which is in addition positive (i.e. if $f(x) \geq 0$ for all $x \in X$, then $I_\mu(f) \geq 0$). Further, $I_\mu$ is continuous, i.e. $\exists C \geq 0$ such that $I_\mu(f) \leq C\|f\|_{\sup}$. This follows directly from the fact that $X$ is a compact space and the fact that $\mu$ is a Radon measure. In a nutshell, $I_\mu$ is an element of the space $\mathcal{C}(X, \mathbb{R})^*$, the space of linear bounded functionals on $\mathcal{C}(X, \mathbb{R})$.

Surprisingly, this statement can also be reversed. More precisely, *every* bounded linear functional $\mathcal{C}(X, \mathbb{R}) \to \mathbb{R}$ which is in addition positive arises exactly in this way. This is the statement of the following theorem by Riesz (see for example [10, Theorem 2.14]).

**Theorem 7.** *Let $X$ be a compact Hausdorff space and $A : \mathcal{C}(X, \mathbb{R}) \to \mathbb{R}$ a continuous functional which is positive. Then, there exists a* unique *Radon measure $\mu$ such that $A = I_\mu$.*

Therefore, this theorem establishes a one-to-one correspondence between Radon measures on $X$ and a subset of $\mathcal{C}(X, \mathbb{R})^*$. Equipping $\mathcal{C}(X, \mathbb{R})^*$ with the operator norm we obtain the unit ball

$$\mathcal{A}_X := \{A \in \mathcal{C}(X, \mathbb{R})^* : |A[f]| \leq \|f\|_{\sup}\}.$$

In the following we define

$$\mathcal{R}X := \mathcal{A}_X \cap \{A \in \mathcal{C}(X, \mathbb{R})^* : A \text{ is a positive functional}\}.$$

Therefore we leverage the correspondence to Radon measures on $X$ which are subnormalized (i.e. $\mu(X) \leq 1$).[2]

It remains to show that $\mathcal{R}X$ is a compact Hausdorff space. Equipping $\mathcal{C}(X, \mathbb{R})^*$ with the weak $*$-topology[3] makes $\mathcal{A}_X$ to a compact set by the Banach-Alaoglu theorem (see for example [9, Thm. 3.15]). Since the space of positive functionals is closed this shows that $\mathcal{R}X$ is compact and Hausdorff.

This motivates the functor

$$\mathcal{R} : \mathbf{CHaus} \to \mathbf{CHaus} : X \mapsto \mathcal{R}X$$

where for every morphism $f : X \to Y$ we have

$$f_* := \mathcal{R}f : \mathcal{R}X \to \mathcal{U}_Y, \mu \mapsto \mu \circ f^{-1}$$

is well-defined. Note that $\mathcal{R}f$ is indeed a morphism since for every function $g \in \mathcal{C}(Y, \mathbb{R})$ we have that

$$I_{f_*\mu}(g) = I_\mu(g \circ f).$$

We will now again extend $\mathcal{R}$ to a monad by defining $\eta_X : X \to \mathcal{R}X : x \mapsto \delta_x$ the Dirac measure and

$$\mu_X : \mathcal{R}\mathcal{R}X \to \mathcal{R}X$$

defined as

$$\mu_X(\mathcal{M}) := f \mapsto \int_{\mathcal{R}X} \left( \int_X f \,\mathrm{d}\mu \right) \mathrm{d}\mathcal{M}(\mu)$$

where $\mathcal{M}$ is a Radon measure on $\mathcal{R}X$.

**Theorem 8.** *$(\mathcal{R}, \eta, \mu)$ forms a monad.*

*Proof.* We have already proven that $\mathcal{R}$ is a functor.

(i) We start by showing that $\eta$ is a natural transformation. First $\eta_X$ is well-defined, i.e. $\delta_x \in \mathcal{R}X$ since
$$|\delta_x(f)| = |f(x)| \leq \|f\|_{\sup}.$$
Further, $\eta_X$ is continuous, since $\mathsf{ev}_f \circ \delta_x = f$. Therefore, by definition of the weak $*$-topology, $\eta_X$ is continuous. Moreover, $\eta$ is natural since for every morphism $g : X \to Y$ and $f \in \mathcal{C}(Y, \mathbb{R})$ we have that $\delta_{g(x)}(f) = \delta_x(f \circ g)$.

[2] In fact it is also possible to restrict just to normalized measures which correspond to the positive unit sphere in $\mathcal{C}(X, \mathbb{R})^*$.

[3] the coarsest topology such that every linear map $\mathcal{C}(X, \mathbb{R})^* \to \mathbb{R}$ is continuous, i.e. the integration functional $I_\mu : f \mapsto \int_X f \,\mathrm{d}\mu$ is continuous for every Radon measure $\mu$.

(ii) We continue by showing that $\mu$ is a natural transformation. First $\mu_X$ is well defined, i.e. for $\mathcal{M} \in \mathcal{R}\mathcal{R}X$ we have $\mu_X(\mathcal{M}) \in \mathcal{R}X$ since

$$|\mu_X(\mathcal{M})(f)| \leq \int_{\mathcal{R}X} \left( \int_X |f| \, \mathrm{d}\mu \right) \mathrm{d}\mathcal{M}(\mu) \leq \int_{\mathcal{R}X} \|f\|_{\sup} \, \mathrm{d}\mathcal{M} \leq \|f\|_{\sup}$$

where we have used the triangle inequality in the first inequality and the fact that $\mu$ and $\mathcal{M}$ are subnormalized measures in the second inequality.

To show that $\mu_X : \mathcal{R}\mathcal{R}X \to \mathcal{R}X$ is continuous, we use again the definition of the weak $*$-topology and show that $\mathsf{ev}_f \circ \mu_X$ is continuous for every $f \in \mathcal{C}(X, \mathbb{R})$.

Therefore, let $f \in \mathcal{C}(X, \mathbb{R})$ be fixed. Note that $\mathcal{R}X \to \mathbb{R}, \mu \to I_\mu(f)$ is continuous by definition. Therefore, for $\mathcal{M} \in \mathcal{R}\mathcal{R}X$ we have

$$\mathsf{ev}_f \circ \mu_X(\mathcal{M}) = \int_{\mathcal{R}X} I_\mu(f) \, \mathrm{d}\mathcal{M}(\mu) = I_{\mathcal{M}}(G)$$

where $G := \mu \mapsto I_\mu(f) \in \mathcal{C}(\mathcal{R}X, \mathbb{R})^*$. Therefore, $\mathsf{ev}_f \circ \mu_X = \mathsf{ev}_G$ which is continuous by the definition of the topology on $\mathcal{R}\mathcal{R}X$.

(iii) It remains to show that the diagrams in Eq. (1) commute. But this is very similar to the proof of Theorem 6 and therefore we skip its verification. □

Concluding this section, we note that similar to the case of the finitary probability monad and the Giry monad, it is again possible to construct the Kleisli category $\mathbf{Kl}(\mathcal{R})$. We refer to [3, Section 5] for details. In a nutshell, $\mathbf{Kl}(\mathcal{R})$ captures now continuous Markov processes which map to Radon measures.

## 5. Conclusion

We have considered different ways of constructing categories of probability distributions. Starting with finite probability distributions we have extended the probability functor to arbitrary measurable spaces via the Giry monad and to compact Hausdorff spaces via the Radon monad. Furthermore, we have discussed an interpretation of the corresponding Kleisli categories as categories of Markov kernels.

This formulation gives a unified way of axiomatizing probability distributions independent of finite or infinite support and independent of using measure theoretic foundations. In addition to the presented constructions, the corresponding Kleisli categories can be equipped with additional structure (for example to be symmetric monoidal). Therefore this new level of abstraction allows to treat probability theory on arbitrary measurable spaces using a similar reasoning as finite probability distributions and highlights possible differences in both frameworks.

## References

[1] V. I. Bogachev. *Measure Theory: Volume I.* Springer, 2007.
[2] N. Bourbaki. *Elements of Mathematics. General Topology*, volume Part 1. Addison-Wesley Publishing, Paris, 1966.
[3] T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:1–98, 2020.
[4] R. Furber and B. Jacobs. From Kleisli Categories to Commutative C*-algebras: Probabilistic Gelfand Duality. *Logical Methods in Computer Science*, 11(2):1–28, 2015.
[5] P. R. Halmos. *Measure Theory*. Graduate texts in mathematics. Springer, New York, 1976.
[6] M. Mislove. Probabilistic Monads, Domains and Classical Information. *Electronic Proceedings in Theoretical Computer Science*, 88:87–100, 2012.
[7] P. Panangaden. *Labelled Markov Processes*. World Scientific, 2009.
[8] P. Perrone. Notes on Category Theory with examples from basic mathematics. *Lecture notes*, 2019.
[9] W. Rudin. *Functional Analysis*. McGraw-Hill, New York, 2nd edition, 1978.
[10] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, New York, 3rd edition, 1987.