Notes on the monad of measures and the category of measure kernels

Shing Hin Ho

Abstract

In mathematical folklore, the space of all measures over a measurable space forms a monad on the category of measurable spaces. These notes formalize this result and prove that the monad is not strong. Furthermore, we show that the family of functions mapping measures to their (maximal) product measure is not measurable.

1 Introduction

In categorical approaches for probability theory, Lawvere famously introduced a probabilistic functor that maps a measurable space X to its space of probability measure $\mathcal{G}(X)$ [5]. Giry proved that $\mathcal{G}$ has the structure of a monad in [4] and therefore $\mathcal{G}$ is commonly known as the *Giry monad*. Giry defined the monad on two categories – the category of measurable spaces **Meas**, and the category of Polish spaces **Pol**, with the space of measures metrized by the Prokhorov metric. In **Meas**, the σ-algebra of $\mathcal{G}$ is the least σ-algebra that makes the family of evaluations $\{\mathrm{ev}_E : \mathcal{G}X \to \mathbb{R}\}_{E\in\Sigma_X}$ measurable, where

$$\mathrm{ev}_E : \mathbb{P} \longmapsto \mathbb{P}(E).$$

There are different variations of probability/measure monads since Giry's discovery and they extend/restrict to certain desirable spaces of measures such as restricting the measures to be Radon or finite.

We are interested in the general setting that does not impose any constraints on the measurable space or the set of measures, i.e. the space of all measures. Note that the monad is known to exist in category theory folklore – these notes simply formalize the result and explain why the monad $\mathcal{M}$ lacks desirable properties – namely the lack of *monadic strength* and a family of measurable functions that maps measures to their (maximal) product measure.

Our motivation to investigate this monad stems from denotational semantics. In denotational semantics, Moggi's monadic metalanguage allows the encoding of computational effects with a strong monad in a category with finite products [6]. In particular, there are approaches that use a monad of (all) measures to give semantics to statistical languages (e.g. [7]). However, the monad of measures, while it is known to exist (e.g. see Section 2.3 of [1]), has never been studied on its own/formalized. Also, as suggested in [8], the monad of measures may not be strong (a necessary condition for the Moggi's categorical semantics to be well-defined).

2 Preliminaries

Recall a *measurable space* X is a pair $(|X|, \Sigma_X)$ where $|X|$ is a set and $\Sigma_X \subseteq \mathcal{P}(X)$ is a *σ-algebra*, a set of subsets that contains $\varnothing$, and is closed under complements and countable unions. Similarly, a *topological space* X is a pair $(|X|, \mathcal{O}(X))$ where $|X|$ is a set and $\mathcal{O}(X)$ is a *topology*, a set of subsets that contains $\varnothing$ and X, and is closed under arbitrary unions and finite intersections.

Definition 2.1 (Borel space)**.** The *Borel space* of a topological space X is a measurable space $(|X|, \mathcal{B}(X))$ where $\mathcal{B}(X)$ is the least σ-algebra consisting $\mathcal{O}(X)$.

Example 2.2. The Borel space of the standard topology over $\mathbb{R}^n$ is the measurable space generated by the open n-balls of $\mathbb{R}^n$.

Definition 2.3 (measurable function)**.** Let X and Y be measurable spaces. A (Σ_X, Σ_Y)*-measurable function* is a function $f : X \to Y$ satisfying $f^{-1}(E) \in \Sigma_X$ for all $E \in \Sigma_Y$.

Definition 2.4 (measure)**.** A *measure* over the measurable space X is a function $\mu : \Sigma_X \to \overline{\mathbb{R}}_{\geq 0}$ [1] satisfying $\mu(\varnothing) = 0$ and $\mu(\biguplus_{i\in\mathbb{N}} E_i) = \sum_{i\in\mathbb{N}} \mu(E_i)$ for a sequence of pairwise disjoint measurable sets $\{E_i \in \Sigma_X\}_{i\in\mathbb{N}}$.

Example 2.5. The (standard) Gaussian measure $\mathcal{N}(- \mid 0, 1) : \Sigma_{\mathcal{O}(X)} \to \overline{\mathbb{R}}_{\geq 0}$ is a measure defined by

$$\mathcal{N}(- \mid 0, 1) : U \longmapsto \int_U \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)\,\mathrm{d}x.$$

For instance, $\mathcal{N}(\mathbb{R}^+ \mid 0, 1) = 1/2$, which reflects the fact that the probability of drawing a positive number from the standard normal distribution is $1/2$.

Definition 2.6 (measure kernel)**.** Let X and Y be measurable spaces. A *measure kernel* between X and Y is a function $k : X \times \Sigma_Y \to \overline{\mathbb{R}}_{\geq 0}$ such that for every $x \in X$, $k(x, -)$ is a measure, and for all $E \in \Sigma_Y$, $k(-, E)$ is $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable. We write $k : X \rightsquigarrow Y$ for the kernel $k : X \times \Sigma_Y \to \overline{\mathbb{R}}_{\geq 0}$.

Example 2.7. The function $(a, U) \longmapsto \mathcal{N}(U \mid a, 1)$ that maps a number $a \in \mathbb{R}$ and $U \in \mathcal{B}(\mathbb{R})$ to a normal distribution with mean a and variance 1 is a measure kernel from the Borel space of $\mathbb{R}$ to itself.

Remark 2.8. The category of measurable spaces **Meas** consists of measurable spaces as objects and measurable functions as morphisms. **Meas** is both complete and cocomplete. In particular, this means **Meas** has all small products and consequently and $(\mathbf{Meas}, \times, \mathbf{1})$ is a cartesian category, where $\mathbf{1}$ is the singleton measurable space.

[1] In this article, we write $\overline{\mathbb{R}}_{\geq 0}$ for the set (or when applicable, the measurable space) of extended, non-negative reals, i.e. $\overline{\mathbb{R}}_{\geq 0} = [0, \infty]$. We define $x + \infty = \infty$, $x \cdot \infty = \infty$ when $x \neq 0$, and $0 \cdot \infty = 0$ (standard in measure theory).

3 The category of measure kernels

Throughout this article, we fix three measurable spaces X, Y and Z. We now establish that the collection of measure kernels over all measurable spaces forms a category (Proposition 3.2).

Lemma 3.1. *Let $\mu : \Sigma_X \to \overline{\mathbb{R}}_{\geq 0}$ be a measure and $k : X \times \Sigma_Y \to \overline{\mathbb{R}}_{\geq 0}$ a measure kernel. Define a measure $\nu : \Sigma_Y \to \overline{\mathbb{R}}_{\geq 0}$ by $\nu : E \longmapsto \int_X k(x, E)\, \mu(\mathrm{d}x)$. Then for all $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable $f : Y \to \overline{\mathbb{R}}_{\geq 0}$,*

$$\int_Y f(y)\, \nu(\mathrm{d}y) = \int_X \int_Y f(y)\, k(x, \mathrm{d}y)\, \mu(\mathrm{d}x).$$

Proof. By the simple function approximation theorem, every f is of the form $f = \lim_{n\to\infty} f_n$ for a sequence of real-valued measurable simple functions $\{f_n = \sum_{i=1}^{m_n} c_{n,i} \mathbf{1}_{E_{n,i}}\}_{n\in\mathbb{N}}$. Notice

$$\begin{aligned}
\int_Y f(y)\, \nu(\mathrm{d}y) &= \lim_{n\to\infty} \sum_{i=1}^{m_n} c_{n,i} \int_X k(x, E_{n,i})\, \mu(\mathrm{d}x) && \text{(see below)} \\
&= \lim_{n\to\infty} \sum_{i=1}^{m_n} c_{n,i} \int_X \int_{E_{n,i}} k(x, \mathrm{d}y)\, \mu(\mathrm{d}x) && \text{(simple function)} \\
&= \int_X \int_Y f(y)\, k(x, \mathrm{d}y)\, \mu(\mathrm{d}z), && \text{(see below)}
\end{aligned}$$

where the first and third equalities hold due to the monotone convergence theorem and linearity of integrals (which always holds for non-negative extended real valued functions). ☐

Proposition 3.2. *The collection of measure kernels with measurable spaces as objects and measure kernels as morphisms form category* **MeasKrn**, *with the composition operator defined by*

$$(l \circ k)(x, E) = \int_Y l(y, E)\, k(x, \mathrm{d}y)$$

for $k : X \rightsquigarrow Y$ and $l : Y \rightsquigarrow Z$.

Proof. Notice the composition operator is well-defined – given $k : X \rightsquigarrow Y$, $l : Y \rightsquigarrow Z$ and $m : Z \rightsquigarrow A$, the composition is associative:

$$\begin{aligned}
(m \circ (l \circ k))(x, E) &= \int_Z m(z, E)\, (l \circ k)(x, \mathrm{d}z) \\
&= \int_Y \int_Z m(z, E)\, l(y, \mathrm{d}z)\, k(x, \mathrm{d}y) && \text{(Lemma 3.1)} \\
&= ((m \circ l) \circ k)(x, E).
\end{aligned}$$

For $\mathrm{id}_X : X \rightsquigarrow X$, it is defined by the point-mass kernel $\mathrm{id}_X(x, E) = \mathbf{1}_E(x)$, and $(\mathrm{id}_Y \circ k)(x, E) = \int_X \mathrm{id}_Y(y, E)\, k(x, \mathrm{d}y) = k(x, E) = (k \circ \mathrm{id}_X)(x, E)$. ☐

4 A measurable space of measures

The Giry monad $\mathcal{G}$ has two desirable properties: 1. the Kleisli category is the category of probability kernels **ProbKrn**, and 2. it is a strong monad . We show that the space of all measures satisfies the first condition (for **MeasKrn**) (Theorem 4.12), but it is not a strong monad (Theorem 5.4).

Definition 4.1 (space of measures)**.** Let X be a measurable space. The *measurable space of measures* $\mathcal{M}X$ is a set M_X consisting the set of measures over X, endowed with the least σ-algebra Σ_{M_X} that makes the family of evaluations $\mathrm{ev}_E : \mu \mapsto \mu(E)$ $\mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$-measurable for every $E \in \Sigma_X$.

Definition 4.2 (pushforward)**.** Let $\mu \in \mathcal{M}X$ and $f : X \to Y$ measurable. The *pushforward* of f along μ is a measure $f_*(\mu)$ defined by $f_*(\mu) : E \mapsto \mu(f^{-1}(E))$.

Lemma 4.3. *Let $\mathcal{E} \subseteq \mathcal{P}(Y)$ such that Σ_Y is the least σ-algebra containing $\mathcal{E}$. Then a function $f : X \to Y$ is (Σ_X, Σ_Y)-measurable if and only if for all $E \in \mathcal{E}$, $f^{-1}(E) \in \Sigma_X$.*

Proof. From left to right, the statement follows from the definition of measurability. From right to left, we proceed via the monotone class argument – suppose for all $E \in \mathcal{E}$, $f^{-1}(E) \in \Sigma_X$, we define a set of subsets Σ'_Y by

$$\Sigma'_Y \coloneqq \{E \in \Sigma_Y \mid f^{-1}(E) \in \Sigma_X\}.$$

Notice Σ'_Y is a σ-algebra – we know $\varnothing, Y \in \Sigma'_Y$ because $f^{-1}(\varnothing) = \varnothing \in \Sigma_X$ and $f^{-1}(Y) = X \in \Sigma_X$. Also, Σ'_Y is closed under complements and countable unions – suppose $E \in \Sigma'_Y$, then $f^{-1}(E) \in \Sigma_X$ by definition, which implies $X \setminus f^{-1}(E) = f^{-1}(Y \setminus E) \in \Sigma_X$ and consequently, $Y \setminus E \in \Sigma'_Y$. For countable union, suppose $\{E_i \in \Sigma'_Y\}_{i\in\mathbb{N}}$, then $f^{-1}(E_i) \in \Sigma_X$ for all $i \in \mathbb{N}$, which implies

$$\bigcup_{i\in\mathbb{N}} f^{-1}(E_i) = f^{-1}\Big(\bigcup_{i\in\mathbb{N}} E_i\Big) \in \Sigma_X,$$

and consequently, $\bigcup_{i\in\mathbb{N}} E_i \in \Sigma'_Y$. To show $\Sigma'_Y = \Sigma_Y$, we prove $\Sigma'_Y \subseteq \Sigma_Y$ and $\Sigma_Y \subseteq \Sigma'_Y$. Notice $\Sigma'_Y \subseteq \Sigma_Y$ is true by definition. To show $\Sigma_Y \subseteq \Sigma'_Y$, notice $\sigma(\mathcal{E}) = \Sigma_Y$ and $\mathcal{E} \subseteq \Sigma'_Y$ by assumption, since Σ'_Y is a σ-algebra and it contains $\mathcal{E}$, it contains the least σ-algebra generated by $\mathcal{E}$, implying $\Sigma_Y \subseteq \Sigma'_Y$ and $\Sigma'_Y = \Sigma_Y$. This implies for all $E \in \Sigma_Y$, $f^{-1}(E) \in \Sigma_X$. □

Lemma 4.4. *A function $f : X \to \mathcal{M}(Y)$ is $(\Sigma_X, \Sigma_{\mathcal{M}Y})$-measurable if and only if for all $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$ and $E \in \Sigma_Y$, we have $\{x \in X \mid f(x)(E) \in U\} \in \Sigma_X$.*

Proof. From the definition of generated σ-algebra, $\Sigma_{\mathcal{M}X}$ is equivalently

$$\begin{aligned}\Sigma_{\mathcal{M}X} &= \sigma\{\mathrm{ev}_E : M_X \to \overline{\mathbb{R}}_{\geq 0} \mid E \in \Sigma_X, U \in \overline{\mathbb{R}}_{\geq 0}\}\\ &= \sigma\{\{\mu \in M_X \mid \mu(E) \in U\} \mid E \in \Sigma_X, U \in \overline{\mathbb{R}}_{\geq 0}\}.\end{aligned}$$

This means the family $\{\mu \in M_X | \mu(E) \in U\}_{E\in\Sigma_X, U\in\overline{\mathbb{R}}_{\geq 0}}$ generates $\Sigma_{\mathcal{M}X}$. Recall that a function is measurable if and only if the pre-image of a generating class of the codomain are measurable sets (proven via a monotone class argument), this means f is $(\Sigma_X, \Sigma_{\mathcal{M}Y})$-measurable if and only if for all $E \in \Sigma_X$ and $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$, $\Sigma_X \ni f^{-1}\{\mu \in M_X \mid \mu(E) \in U\} = \{x \in X \mid f(x)(E) \in U\}$. □

Lemma 4.5. *Let $f : X \to Y$ be (Σ_X, Σ_Y)-measurable. Then the pushforward $f_* : \mathcal{M}X \to \mathcal{M}Y$ is $(\Sigma_{\mathcal{M}X}, \Sigma_{\mathcal{M}Y})$-measurable.*

Proof. By Lemma 4.4, it suffices to show $\{\mu \in \mathcal{M}X \mid f_*(\mu)(E) \in U\} \in \Sigma_{\mathcal{M}X}$ for all $E \in \Sigma_Y$ and $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$. Notice $\{\mu \in \mathcal{M}X \mid f_*(\mu)(E) \in U\} = \{\mu \in \mathcal{M}X \mid \mu(f^{-1}(E)) \in U\}$ and $f^{-1}(E) \in \Sigma_X$ by assumption, which is a subset that generates $\Sigma_{\mathcal{M}X}$ by Definition 4.1. $\square$

Lemma 4.6 (functoriality)**.** *Let $\mathcal{M} : \mathbf{Meas} \to \mathbf{Meas}$ be a map that sends X to $\mathcal{M}X$ and $f : X \to Y$ to $f_* : \mathcal{M}X \to \mathcal{M}Y$. Then $\mathcal{M}$ is a functor.*

Proof. By computation. Notice $\mathcal{M}(\mathrm{id}_X)(\mu)(E) = \mu(E) = \mathrm{id}_{\mathcal{M}X}(\mu)(E)$ and $\mathcal{M}(g \circ f)(\mu)(E) = g_*(f_*(\mu))(E) = (\mathcal{M}(g) \circ \mathcal{M}(f))(\mu)(E)$. $\square$

Lemma 4.7. *Let $f : X \to \overline{\mathbb{R}}_{\geq 0}$ be $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable. Then the integration map $I_f : \mathcal{M}X \to \overline{\mathbb{R}}_{\geq 0}$ defined below is $(\Sigma_{\mathcal{M}X}, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable:*

$$I_f : \mu \longmapsto \int_X f \,\mathrm{d}\mu.$$

Proof. By the simple function approximation theorem, every measurable function $f : X \to \overline{\mathbb{R}}_{\geq 0}$ is the pointwise limit of a sequence of $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable simple functions $\{f_n : X \to \overline{\mathbb{R}}_{\geq 0}\}_{n \in \mathbb{N}}$ with $f_n = \sum_{i=1}^{m_n} c_{n,i} \cdot \mathbf{1}_{U_{n,i}}$. This implies

$$I_f(\mu) = \lim_{n \to \infty} \int_X \sum_{i=1}^{m_n} c_{n,i} \cdot \mathbf{1}_{U_{n,i}} \,\mathrm{d}\mu = \lim_{n \to \infty} \sum_{i=1}^{m_n} c_{n,i} \cdot \mathrm{ev}_{U_{n,i}}(\mu),$$

where the first equality follows from the monotone convergence theorem. Since the sum, multiplication, and pointwise limit of a $\mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$-measurable function are measurable, I_f is measurable. $\square$

Proposition 4.8 (induced measure)**.** *Let $f : X \to \overline{\mathbb{R}}_{\geq 0}$ and $\mu \in \mathcal{M}X$. The* induced measure *of μ and f is a measure $f \cdot \mu \in \mathcal{M}X$ defined by*

$$(f \cdot \mu) : E \longmapsto \int_E f \,\mathrm{d}\mu.$$

For all $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable $f : X \to \overline{\mathbb{R}}_{\geq 0}$, the function $(f \cdot -) : \mathcal{M}X \to \mathcal{M}X$ is $(\Sigma_{\mathcal{M}X}, \Sigma_{\mathcal{M}X})$-measurable.

Proof. By Lemma 4.4, it suffices to show that for all $E \in \Sigma_X$, $(f \cdot -)(E) : \mathcal{M}X \to \overline{\mathbb{R}}_{\geq 0}$ is measurable. Suppose $E \in \Sigma_X$ and $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$, notice $(f \cdot \mu)(E) = I_{\mathbf{1}_E \cdot f}(\mu)$, which is $(\Sigma_{\mathcal{M}X}, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable by Lemma 4.7. $\square$

Lemma 4.9. *The following maps $\delta : X \to \mathcal{M}X$ and $\xi : \mathcal{M}^2X \to \mathcal{M}X$ are measurable:*

$$\delta(x) : U \longmapsto \mathbf{1}_U(x) \qquad \xi(\nu) : U \longmapsto \int_{\mathcal{M}X} \pi(U) \,\mathrm{d}\nu(\pi).$$

Proof. By Lemma 4.4, δ is $(\Sigma_X, \Sigma_{\mathcal{M}X})$-measurable if $\{x \in X \mid \delta(x)(E) \in U\} \in \Sigma_X$ for all $E \in \Sigma_X$ and $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$. Consider the four cases of $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$:

$$\{x \in X \mid \delta(x)(E) \in U\} = \begin{cases} \{x \in X \mid \delta(x)(E) \notin \{0,1\}\} = \varnothing & \text{if } 0 \notin U, 1 \notin U \\ \{x \in X \mid \delta(x)(E) = 1\} = E & \text{if } 0 \notin U, 1 \in U \\ \{x \in X \mid \delta(x)(E) = 0\} = X \setminus E & \text{if } 0 \in U, 1 \notin U \\ \{x \in X \mid \delta(x)(E) \in \{0,1\}\} = X & \text{if } 0 \in U, 1 \in U. \end{cases}$$

Since $\varnothing, E, X \setminus E$ and X are Σ_X-measurable sets, δ is $(\Sigma_X, \Sigma_{\mathcal{M}X})$-measurable. For $\xi : \mathcal{M}^2 X \to \mathcal{M}X$, notice $\xi(\nu)(U) = I_{\mathrm{ev}_U}(\nu)$. Since ev_U is $(\Sigma_{\mathcal{M}X}, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable by definition of the σ-algebra, I_{ev_U} is $(\Sigma_{\mathcal{M}^2 X}, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable by Lemma 4.7. This implies $\{\nu \in \mathcal{M}^2 X \,|\, \xi(\nu)(U) \in V\} = \{\nu \in \mathcal{M}^2 X \,|\, I_{\mathrm{ev}_U}(\nu) \in V\} \in \Sigma_{\mathcal{M}^2 X}$. By Lemma 4.4, ξ is $(\Sigma_{\mathcal{M}^2 X}, \Sigma_{\mathcal{M}X})$-measurable. ☐

Lemma 4.10. *The maps $\delta_X : X \to \mathcal{M}X$ and $\xi_X : \mathcal{M}^2 X \to \mathcal{M}X$ are natural.*

Proof. By diagram chasing on the naturality squares:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\delta_X"'] & Y \arrow[d, "\delta_Y"] \\ \mathcal{M}X \arrow[r, "f_*"] & \mathcal{M}Y \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}^2 X \arrow[r, "f_{**}"] \arrow[d, "\xi_X"'] & \mathcal{M}^2 Y \arrow[d, "\xi_Y"] \\ \mathcal{M}X \arrow[r, "f_*"] & \mathcal{M}Y \end{tikzcd}$$

The left square commutes because $\delta_Y(f(x)) = f_*(\delta_X)(x)$. The right square commutes because

$$(\xi_Y \circ f_{**})(\nu)(E) = \int_{\mathcal{M}X} \mathrm{ev}_{f^{-1}(E)} \,\mathrm{d}\nu = (\xi_X \nu)(f^{-1}(E)) = (f_* \circ \xi_X)(\nu)(E),$$

where the first equality holds due to the change-of-variable formula. ☐

Lemma 4.11. *Let $\nu \in \mathcal{M}^2 X$ and $f : X \to \overline{\mathbb{R}}_{\geq 0}$ be $(\Sigma_X, \mathcal{B}(\overline{\mathbb{R}}_{\geq 0}))$-measurable. Then*

$$\int_X f \,\mathrm{d}(\xi_X \nu) = \int_{\mathcal{M}X} \int_X f \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu).$$

Proof. By the simple function approximation theorem, every measurable $f : X \to \overline{\mathbb{R}}_{\geq 0}$ is the pointwise limit of a sequence of $(\Sigma_X, \mathcal{B}(\mathbb{R}_{\geq 0}))$-measurable simple functions $\{f_n : X \to \mathbb{R}_{\geq 0}\}_{n \in \mathbb{N}}$ with $f_n = \sum_{i=1}^{m_n} c_i \cdot \mathbf{1}_{U_i}$. This implies

$$\begin{aligned} \int_X f \,\mathrm{d}(\xi_X \nu) &= \lim_{n\to\infty} \int_X f_n \,\mathrm{d}(\xi_X \nu) && \text{(monotone convergence)} \\ &= \lim_{n\to\infty} \sum_{i=1}^{m_n} c_{n,i} \cdot (\xi_X \nu)(U_{n,i}) && \text{(simple function)} \\ &= \lim_{n\to\infty} \sum_{i=1}^{m_n} c_{n,i} \int_{\mathcal{M}X} \int_{U_{n,i}} \mathrm{d}\mu \,\mathrm{d}\nu(\mu) && \text{(definition of } \xi_X\text{)} \\ &= \lim_{n\to\infty} \int_{\mathcal{M}X} \int_X \sum_{i=1}^{m_n} c_{n,i} \cdot \mathbf{1}_{U_{n,i}} \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu) && \text{(linearity)} \\ &= \int_{\mathcal{M}X} \int_X f \,\mathrm{d}\mu \,\mathrm{d}\nu(\mu), && \text{(monotone convergence)} \end{aligned}$$

which completes the proof. ☐

Theorem 4.12 (existence)**.** *There is a monad $\mathcal{M} : \mathbf{Meas} \to \mathbf{Meas}$ such that the underlying set of $\mathcal{M}(X)$ is the set of measures over X and* $\mathrm{Kl}(\mathcal{M}) \cong$ **MeasKrn***.*

Proof. To show that $(\mathcal{M}, \xi, \delta)$ is a monad of measurable spaces, it suffices to show that the following squares commute:

$$\begin{tikzcd}
\mathcal{M}^3X \arrow[r, "(\xi_X)_*"] \arrow[d, "\xi_{\mathcal{M}X}"'] & \mathcal{M}^2X \arrow[d, "\xi_X"] \\
\mathcal{M}^2X \arrow[r, "\xi_X"] & \mathcal{M}X
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathcal{M}X \arrow[r, "\delta_{\mathcal{M}X}"] \arrow[d, "(\delta_X)_*"'] \arrow[dr, "\mathrm{id}_{\mathcal{M}X}"] & \mathcal{M}^2X \arrow[d, "\xi_X"] \\
\mathcal{M}^2X \arrow[r, "\xi_X"] & \mathcal{M}X
\end{tikzcd}$$

For the associativity square, suppose $\nu \in \mathcal{M}^3X$ and $E \in \Sigma_X$, then

$$\begin{aligned}
(\xi_X \circ (\xi_X)_*)(\nu)(E) &= \int_{\mathcal{M}X} \mathrm{ev}_E \, \mathrm{d}((\xi_X)_*\nu) \\
&= \int_{\mathcal{M}^2X} \mathrm{ev}_E \circ \xi_X \, \mathrm{d}\nu && \text{(change of variable)} \\
&= \int_{\mathcal{M}^2X} \int_{\mathcal{M}X} \mathrm{ev}_E(\pi) \, \mathrm{d}\mu(\pi) \, \mathrm{d}\nu(\mu) \\
&= \int_{\mathcal{M}X} \mathrm{ev}_E \, \mathrm{d}(\xi_{\mathcal{M}X}\nu) && \text{(Lemma 4.11)} \\
&= (\xi_X \circ \xi_{\mathcal{M}X})(\nu)(E).
\end{aligned}$$

For the unitality square, suppose $\mu \in \mathcal{M}X$ and $E \in \Sigma_X$, then

$$\begin{aligned}
(\xi_X \circ \delta_{\mathcal{M}X})(\mu)(E) &= \xi_X(\delta_{\mathcal{M}X}(\mu))(E) \\
&= \int_{\mathcal{M}X} \mathrm{ev}_E \, \mathrm{d}\delta_{\mathcal{M}X}(\mu)(\pi) \\
&= \mu(E) \\
&= \int_X \mathrm{ev}_E \circ \delta_{\mathcal{M}X} \, \mathrm{d}\mu && \text{(indicator function)} \\
&= \int_{\mathcal{M}X} \mathrm{ev}_E \, \mathrm{d}(\delta_{\mathcal{M}X})_*\mu && \text{(change of variable)} \\
&= (\xi_X \circ (\delta_{\mathcal{M}X})_*)(\mu)(E),
\end{aligned}$$

which completes the proof of existence. Next, to show that the Kleisli category $\mathrm{Kl}(\mathcal{M})$ is trivially isomorphic to the category **MeasKrn** of measure kernels via a bijection that sends $f : X \to \mathcal{M}Y$ to $(x, U) \mapsto f(x)(U)$ and a kernel $k : X \rightsquigarrow Y$ to $x \mapsto (U \mapsto k(x, U))$. ☐

5 Non-measurability of strength

Recall a monad T is *strong* with respect to the cartesian category $(\mathbf{C}, \times, \mathbf{1})$ if there is a natural transformation $\{\theta_{X,Y} : X \times T(Y) \to T(X \times Y)\}_{X,Y \in \mathbf{C}}$ such that certain diagrams commute (omitted for space, see Definition 3.2 in [6]). We demonstrate that $\mathcal{M}$ is not strong in four steps:

1. assume there is a strength $\{\theta_{X,Y}\}_{X,Y \in \mathbf{Meas}}$ (Definition 5.1),

2. show that the definition is flawed – there are $X, Y \in \mathbf{Meas}$ such that $\theta_{X,Y}$ is not measurable (Proposition 5.2),

3. recall every monad has a canonical strength in **Meas** when it exists (Proposition 5.3), and

4. show that any strength of $\mathcal{M}$ must be equal to the flawed definition, hence contradiction (Theorem 5.4).

Definition 5.1 (non-measurable strength)**.** Let X and Y be measurable spaces and $p_x : Y \to X \times Y$ the measurable function $y \mapsto (x, y)$. The *non-measurable strength* of X and Y is a function $\theta_{X,Y} : X \times \mathcal{M}Y \to \mathcal{M}(X \times Y)$ defined by

$$\theta_{X,Y} : (x, \mu) \longmapsto (p_x)_* \mu.$$

Proposition 5.2 (non-measurability)**.** *There exists $X, Y \in \mathbf{Meas}$ such that $\theta_{X,Y} : X \times \mathcal{M}Y \to \mathcal{M}(X \times Y)$ is not $(\Sigma_{X \times \mathcal{M}Y}, \Sigma_{\mathcal{M}(X \times Y)})$-measurable.*

Proof. Let X, Y be the Borel measurable space over $\mathbb{R}$. To show that $\theta_{\mathbb{R},\mathbb{R}}$ is not measurable. We show that $\theta_{\mathbb{R},\mathbb{R}}$ is not $(\Sigma_{\mathbb{R} \times \mathcal{M}\mathbb{R}}, \Sigma_{\mathcal{M}(\mathbb{R} \times \mathbb{R})})$-measurable by showing there is a $\mu \in \mathcal{M}(\mathbb{R})$ such that $\theta_{\mathbb{R},\mathbb{R}}(-, \mu) : \mathbb{R} \to \mathcal{M}(\mathbb{R})$ is not measurable. Consider a measure $\#_V : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\#_V(U) = \begin{cases} |U \cap V| & \text{if } U \cap V \text{ is finite} \\ \infty & \text{if } U \cap V \text{ is infinite,} \end{cases}$$

where V is a Borel-non-measurable set (e.g. the Vitali set). Notice the measure is well-defined (it is the counting measure on $\mathbb{R}$ restricted to V) and it satisfies $\#_V(\{x\}) = \mathbf{1}_V(x)$. We proceed by showing $\theta_{\mathbb{R},\mathbb{R}}(-, \#_V) : \mathbb{R} \to \mathcal{M}(\mathbb{R} \times \mathbb{R})$ is not $(\mathcal{B}(\mathbb{R}), \Sigma_{\mathcal{M}(\mathbb{R} \times \mathbb{R})})$-measurable. By Lemma 4.4, it suffices to show that there is a $U \in \mathcal{B}(\overline{\mathbb{R}}_{\geq 0})$ and $E \in \mathcal{B}(\mathbb{R})^2$ such that $\{x \in \mathbb{R} \mid \theta_{X,Y}(x, \#_V)(E) \in U\} \notin \mathcal{B}(\mathbb{R})$. Let $U = \{1\}$ and $E = \{(x, x) \mid x \in \mathbb{R}\}$. Then calculate

$$\begin{aligned} \{x \in \mathbb{R} \mid \theta_{X,Y}(x, \#_V)(E) \in U\} &= \{x \in \mathbb{R} \mid (p_x)_* \#_V(E) = 1\} \\ &= \{x \in \mathbb{R} \mid \#_V(p_x^{-1}(\{(y, y) \mid y \in \mathbb{R}\})) = 1\} \\ &= \{x \in \mathbb{R} \mid \#_V(\{x\}) = 1\} \\ &= V \notin \mathcal{B}(\mathbb{R}). \end{aligned}$$

This implies $\theta_{\mathbb{R},\mathbb{R}}(-, \#_V)$ is not $(\mathcal{B}(\mathbb{R}), \Sigma_{\mathcal{M}(\mathbb{R} \times \mathbb{R})})$-measurable, which then implies, $\theta_{\mathbb{R},\mathbb{R}}$ is not $(\Sigma_{\mathbb{R} \times \mathcal{M}\mathbb{R}}, \Sigma_{\mathcal{M}(\mathbb{R} \times \mathbb{R})})$-measurable. $\square$

Proposition 5.3 (Proposition 3.4, [6])**.** *Let* **C** *be well-pointed category and (T, ξ, δ, θ) a strong monad over* **C**. *Then $\{\theta_{X,Y}\}_{X,Y \in \mathbf{C}}$ is the unique family of morphisms such that the following diagram commutes for all $x : \mathbf{1} \to X$ and $m : \mathbf{1} \to TY$:*

$$\begin{array}{ccc} \mathbf{1} & \xrightarrow{\langle x\,,\,m \rangle} & X \times TY \\ {\scriptstyle m}\downarrow & & \downarrow{\scriptstyle \theta_{X,Y}} \\ TY & \xrightarrow{T(\langle x \circ !\,,\,\mathrm{id}_Y \rangle)} & T(X \times Y) \end{array}$$

Theorem 5.4 (non-strength)**.** *$\mathcal{M}$ is not strong with respect to the cartesian category* $(\mathbf{Meas}, \times, \mathbf{1})$.

Proof. Assume by contradiction that $\mathcal{M}$ is strong and has strength $\{\theta'_{X,Y} : X \times \mathcal{M}Y \to \mathcal{M}(X \times Y)\}_{X,Y}$. Then by Proposition 5.3, we know by diagram chasing that, for all $x \in X$, $\mu \in \mathcal{M}Y$ and $E \in \Sigma_{X \times Y}$,

$$\begin{aligned}\theta'_{X,Y}(x,\mu)(E) &= \mathcal{M}(\langle x \circ !, \mathrm{id}_Y \rangle)(\mu)(E) \\ &= \mu(\{y \in Y \mid (x,y) \in E\}) \\ &= \theta_{X,Y}(x,\mu)(E).\end{aligned}$$

However, by Proposition 5.2, $\theta'_{X,Y}$ is not always measurable, a contradiction. $\square$

Remark 5.5. In categorical probability, a canonical way of generating Markov categories is via the Kleisli category of a commutative, affine (meaning $T(\mathbf{1}) \cong \mathbf{1}$), symmetric monoidal monad [3]. Since $\mathcal{M}$ is neither commutative nor affine with respect to $(\mathbf{Meas}, \times, \mathbf{1})$, it does not enjoy such property (unlike the Giry monad).

6 Non-measurability of products

Recall that a *product measure* of $\mu \in \mathcal{M}X$ and $\nu \in \mathcal{M}Y$ is a measure $\rho \in \mathcal{M}(X \times Y)$ satisfying $\rho(E_1 \times E_2) = \mu(E_1)\nu(E_2)$ for all $E_1 \in \Sigma_X$ and $E_2 \in \Sigma_Y$. When μ and ν are σ-finite, ρ is guaranteed to be unique via Carathéodory's extension theorem. For any arbitrary measures μ and ν, there is always a *maximal* of such product measures $\mu \otimes^{\max} \nu$ known as the *maximal product measure.* The maximal product enjoys convenient properties such as the Fubini-Tonelli theorem even for arbitrary measures (see 252G of [2]) and is defined by

$$(\mu \otimes^{\max} \nu)(E) = \inf_{U \in \Sigma_X^{\mathbb{N}}, V \in \Sigma_Y^{\mathbb{N}}} \left\{ \sum_{i \in \mathbb{N}} \mu(U_i) \cdot \nu(V_i) \;\middle|\; E \subseteq \bigcup_{i \in \mathbb{N}} U_i \times V_i \right\},$$

where $\Sigma_X^{\mathbb{N}}$ and $\Sigma_Y^{\mathbb{N}}$ are the sets of sequences of Σ_X/Σ_Y-measurable sets. As a consequence of Proposition 5.2, we know there cannot be a family of measurable functions mapping measures to their maximal product:

Lemma 6.1. *There is no family of $(\Sigma_{\mathcal{M}X \times \mathcal{M}Y}, \Sigma_{\mathcal{M}(X \times Y)})$-measurable functions $\{\otimes^{\max}_{X,Y} : \mathcal{M}X \times \mathcal{M}Y \to \mathcal{M}(X \times Y)\}$ such that $\otimes^{\max}_{X,Y}(\mu,\nu)$ is the maximal product measure.*

Proof. Notice $\theta_{X,Y}(x,\mu)$ maps (x,μ) to the maximal product of $\delta_X(x)$ and μ:

$$\begin{aligned}(\delta_X(x) \otimes^{\max} \mu)(E) &= \inf_{U \in \Sigma_X^{\mathbb{N}}, V \in \Sigma_Y^{\mathbb{N}}} \left\{ \sum_{i \in \mathbb{N}} \delta_X(x)(U_i) \cdot \mu(V_i) \;\middle|\; E \subseteq \bigcup_{i \in \mathbb{N}} U_i \times V_i \right\} \\ &= \inf_{V \in \Sigma_Y^{\mathbb{N}}} \{\mu(V) \mid E_x \subseteq V\} \\ &= \mu(E_x) \\ &= \theta_{X,Y}(x,E).\end{aligned}$$

Suppose, by contradiction, the family of functions exists. Then the following diagram commutes and $\theta_{X,Y}$ is measurable for all $X, Y \in \mathbf{Meas}$:

$$\begin{tikzcd} X \times \mathcal{M}Y \arrow[r, "\delta_X \times \mathrm{id}_{\mathcal{M}Y}"] \arrow[dr, "\theta_{X,Y}"'] & \mathcal{M}X \times \mathcal{M}Y \arrow[d, "\otimes^{\max}_{X,Y}"] \\ & \mathcal{M}(X \times Y) \end{tikzcd}$$

which contradicts Proposition 5.2. $\square$

Proposition 6.2 (non-measurability of product measures)**.** *There is no family of* $(\Sigma_{\mathcal{M}X\times\mathcal{M}Y}, \Sigma_{\mathcal{M}(X\times Y)})$*-measurable functions* $\{\otimes_{X,Y} : \mathcal{M}X\times\mathcal{M}Y \to \mathcal{M}(X\times Y)\}$ *such that* $\mu \otimes_{X,Y} \nu$ *is a product measure for all* $\mu \in \mathcal{M}X$ *and* $\nu \in \mathcal{M}Y$*.*

Proof. Suppose the family of functions $\{\otimes_{X,Y}\}_{X,Y\in\mathbf{Meas}}$ exists and X is a measurable space that contains singleton measurable sets, i.e. $\{x\} \in \Sigma_X$ for all $x \in X$. Then for any $x \in X$ and $\rho \in \mathcal{M}(X\times Y)$, the following inequalities hold if ρ is a product measure of $\delta_X(x)$ and ν:

$$\rho(E) \leq \rho(X \times \{y \in Y \mid (x,y) \in E\}) = \delta_X(x)(X)\cdot\nu(E_x) = \theta_{X,Y}(x,E)$$
$$\rho(E) \geq \rho(\{x\} \times \{y \in Y \mid (x,y) \in E\}) = \delta_X(x)(\{x\})\cdot\nu(E_x) = \theta_{X,Y}(x,E).$$

This means $\rho = \theta_{X,Y}$ for any X that contain singletons. The problem is then reduced to the existence of a family of maximal product measures, and it does not exist by Lemma 6.1. $\square$

References

[1] S. Dash and S. Staton. A monad for probabilistic point processes. *Electronic Proceedings in Theoretical Computer Science*, 333:19–32, Feb. 2021.

[2] D. Fremlin. *Measure Theory*, volume 2. Torres Fremlin, 2000.

[3] T. Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, Aug. 2020.

[4] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, volume 915 of *Lecture Notes in Mathematics*, pages 68–85. Springer, 1982.

[5] E. W. Lawvere. The category of probabilistic mappings: With applications to stochastic processes, statistics, and pattern recognition. Manuscript, 1962.

[6] E. Moggi. Notions of computation and monads. *Inf. Comput.*, 93(1):55–92, July 1991.

[7] C. C. Shan and N. Ramsey. Exact Bayesian inference by symbolic disintegration. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages*, page 130–144. Association for Computing Machinery, 2017.

[8] S. Staton. Commutative semantics for probabilistic programming. In H. Yang, editor, *Programming Languages and Systems*, pages 855–879. Springer Berlin Heidelberg, 2017.