An overview of categorical probability theory

Ruben Van Belle

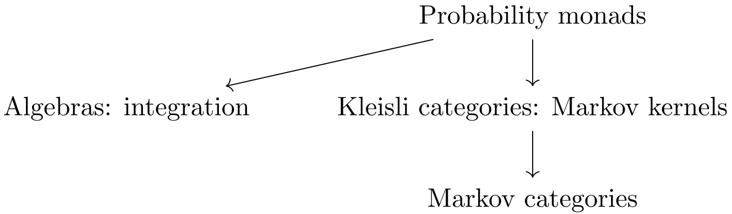
September 2022

Introduction

1. What are *the* objects and morphisms of probability theory?
2. (Incomplete) overview of the current literature.

Introduction

1. What are *the* objects and morphisms of probability theory?
2. (Incomplete) overview of the current literature.



What are *the* objects and morphisms of probability theory?

What are *the* objects and morphisms of probability theory?

1. Probability spaces?

What are *the* objects and morphisms of probability theory?

1. Probability spaces?
2. Random variables?

What are *the* objects and morphisms of probability theory?

1. Probability spaces?
2. Random variables?
3. Markov kernels?

1. Probability spaces

- The category **Prob**:
  - objects: probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$
  - morphisms: measure-preserving maps, i.e. measurable maps $f : \Omega_1 \to \Omega_2$ such that $\mathbb{P}_1 \circ f^{-1} = \mathbb{P}_2$.

1. Probability spaces

- The category **Prob**:
  - objects: probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$
  - morphisms: measure-preserving maps, i.e. measurable maps $f : \Omega_1 \to \Omega_2$ such that $\mathbb{P}_1 \circ f^{-1} = \mathbb{P}_2$.
- **Prob** does not have many (co)limits
  - In general no (co)product or equalizers.
  - There is no initial object (*unbounded randomness*).
  - There is a terminal objects and coequializers exist.

1. Probability spaces

- (*many couplings*) For $\mathbb{P}_1$ and $\mathbb{P}_2$ on Ω, there exists many $\mathbb{P}$ on Ω^2 such that:

$$\mathbb{P} \circ \pi_1^{-1} = \mathbb{P}_1 \quad \text{and} \quad \mathbb{P} \circ \pi_2^{-1} = \mathbb{P}_2.$$

1. Probability spaces

- (*many couplings*) For $\mathbb{P}_1$ and $\mathbb{P}_2$ on Ω, there exists many $\mathbb{P}$ on Ω^2 such that:

$$\mathbb{P} \circ \pi_1^{-1} = \mathbb{P}_1 \quad \text{and} \quad \mathbb{P} \circ \pi_2^{-1} = \mathbb{P}_2.$$

But we want to talk about the **interaction** between stochastic events.

1. Probability spaces

- (*many couplings*) For $\mathbb{P}_1$ and $\mathbb{P}_2$ on Ω, there exists many $\mathbb{P}$ on Ω^2 such that:

$$\mathbb{P} \circ \pi_1^{-1} = \mathbb{P}_1 \quad \text{and} \quad \mathbb{P} \circ \pi_2^{-1} = \mathbb{P}_2.$$

But we want to talk about the **interaction** between stochastic events. We can not do this using probability spaces (e.g. it does not make sense to say that two probability spaces are independent).

1. Probability spaces

- (*many couplings*) For $\mathbb{P}_1$ and $\mathbb{P}_2$ on Ω, there exists many $\mathbb{P}$ on Ω^2 such that:

$$\mathbb{P} \circ \pi_1^{-1} = \mathbb{P}_1 \quad \text{and} \quad \mathbb{P} \circ \pi_2^{-1} = \mathbb{P}_2.$$

  But we want to talk about the **interaction** between stochastic events. We can not do this using probability spaces (e.g. it does not make sense to say that two probability spaces are independent).
- Probability spaces are an important aspect of probability theory, but **not the main objects of interest**.

$\rightarrow$ Probability theory $\neq$ measure theory with measures of total mass 1

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let E be a Polish space, a **random variable** is a measurable map $X : \Omega \to E$.

- *bounded randomness*

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let E be a Polish space, a **random variable** is a measurable map $X : \Omega \to E$.

- *bounded randomness*
- Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$, determine a random variable that describes their interaction:

$$(X_1, X_2) : \Omega \to \mathbb{R}^2$$

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let E be a Polish space, a **random variable** is a measurable map $X : \Omega \to E$.

- *bounded randomness*
- Two random variables $X_1, X_2 : \Omega \to \mathbb{R}$, determine a random variable that describes their interaction:

$$(X_1, X_2) : \Omega \to \mathbb{R}^2$$

- This does not look very categorical:
  - The domain and codomain seem of a *different type*.
  - What are morphisms between random variables? $\to$ order, martingale relation?

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- A random variable $X : \Omega \to E$ induces a probability measure on E, namely $\mathbb{P} \circ X^{-1}$.

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- A random variable $X : \Omega \to E$ induces a probability measure on E, namely $\mathbb{P} \circ X^{-1}$. We are stricly losing information.
- A random variable $X : \Omega \to \mathbb{R}$ can be interpreted as a **density** function. It induces a measure on Ω by

$$A \mapsto \mathbb{E}[X 1_A].$$

2. Random variables

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- A random variable $X : \Omega \to E$ induces a probability measure on E, namely $\mathbb{P} \circ X^{-1}$. We are stricly losing information.
- A random variable $X : \Omega \to \mathbb{R}$ can be interpreted as a **density** function. It induces a measure on Ω by

  $$A \mapsto \mathbb{E}[X1_A].$$

  This measure is absolutely continuous with respect to $\mathbb{P}$. Every such measure is of this form (*Radon-Nikodym theorem*).

3. Markov kernels

Let Ω_1 and Ω_2 be a measurable spaces and let $\mathcal{G}\Omega_2$ be the space of probability measures on Ω_2. A **Markov Kernel** is a measurable map

$$f : \Omega_1 \to \mathcal{G}\Omega_2.$$

- For Markov kernels $f_1 : \Omega_1 \to \mathcal{G}\Omega_2$ and $f_2 : \Omega_2 \to \mathcal{G}\Omega_3$, there is a Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_3$ defined by

$$f(\omega)(A) = \int_{\Omega_2} f_2(\omega_2)(A) f_1(\omega_1)(\mathrm{d}\omega_2),$$

for all $\omega_1 \in \Omega_1$ and measurable $A \subseteq \Omega_3$.

3. Markov kernels

Let Ω_1 and Ω_2 be a measurable spaces and let $\mathcal{G}\Omega_2$ be the space of probability measures on Ω_2. A **Markov Kernel** is a measurable map

$$f : \Omega_1 \to \mathcal{G}\Omega_2.$$

- For Markov kernels $f_1 : \Omega_1 \to \mathcal{G}\Omega_2$ and $f_2 : \Omega_2 \to \mathcal{G}\Omega_3$, there is a Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_3$ defined by

$$f(\omega)(A) = \int_{\Omega_2} f_2(\omega_2)(A) f_1(\omega_1)(\mathrm{d}\omega_2),$$

  for all $\omega_1 \in \Omega_1$ and measurable $A \subseteq \Omega_3$.
- Describes interactions between different stochastic events.

3. Markov kernels

- For a Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_2$ and a probability measure $\mathbb{P}_1$ on Ω_1, there is a probability measure $\mathbb{P}$ on $\Omega_1 \times \Omega_2$ such that

$$\mathbb{P}(A \times B) = \int_A f(\omega)(B)\mathbb{P}_1(\mathrm{d}\omega),$$

for all measurable $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$.

3. Markov kernels

- For a Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_2$ and a probability measure $\mathbb{P}_1$ on Ω_1, there is a probability measure $\mathbb{P}$ on $\Omega_1 \times \Omega_2$ such that

$$\mathbb{P}(A \times B) = \int_A f(\omega)(B)\mathbb{P}_1(\mathrm{d}\omega),$$

for all measurable $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$. *In general*, not every probability measure on $\Omega_1 \times \Omega_2$ is of this form (*regular conditional probabilities*).

3. Markov kernels

- For a Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_2$ and a probability measure $\mathbb{P}_1$ on Ω_1, there is a probability measure $\mathbb{P}$ on $\Omega_1 \times \Omega_2$ such that

$$\mathbb{P}(A \times B) = \int_A f(\omega)(B)\mathbb{P}_1(\mathrm{d}\omega),$$

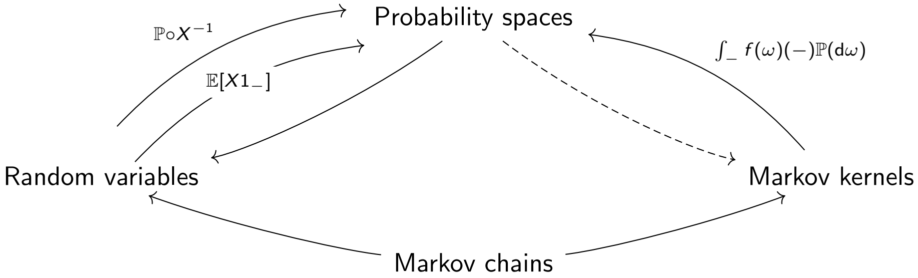
for all measurable $A \subseteq \Omega_1$ and $B \subseteq \Omega_2$. *In general*, not every probability measure on $\Omega_1 \times \Omega_2$ is of this form (*regular conditional probabilities*).

- Let $\mathbb{P}$ be a probability measure on $\Omega_1 \times \Omega_2$. The assignment

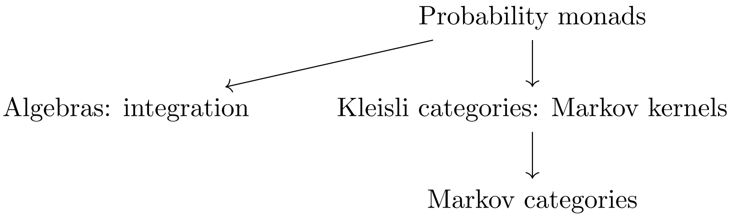
$$A \mapsto \mathbb{E}[1_{\Omega_1 \times A} \mid \pi^{-1}(\mathcal{F}_1)]$$

is $\mathbb{P} \circ \pi_1^{-1}$-*almost surely* σ-additive.

Conclusion


Probability spaces
$\mathbb{P} \circ X^{-1}$
$\mathbb{E}[X 1_{-}]$
$\int_{-} f(\omega)(-)\mathbb{P}(\mathrm{d}\omega)$
Random variables
Markov kernels
Markov chains

An overview of categorical probability theory



Where do probability monads come from?

Lawvere: probabilitstic mappings (1962)

Discusses a category of Markov kernels.

Lawvere: probabilitstic mappings (1962)

Discusses a category of Markov kernels.

- Lawvere introduces the **category of probabilistic mappings** $\mathcal{P}$:
  - objects: measurable spaces $(\Omega, \mathcal{F})$,
  - morphisms: Markov kernels ('probabilistic mappings') $f : \Omega_1 \to \mathcal{G}\Omega_2$,
  - composition: composition of Markov kernels

Lawvere: probabilitstic mappings (1962)

Discusses a category of Markov kernels.

- Lawvere introduces the **category of probabilistic mappings** $\mathcal{P}$:
  - objects: measurable spaces $(\Omega, \mathcal{F})$,
  - morphisms: Markov kernels ('probabilistic mappings') $f : \Omega_1 \to \mathcal{G}\Omega_2$,
  - composition: composition of Markov kernels
- Define $\Phi : \mathcal{P}^{\mathbb{N}} \to \mathcal{P}^{\mathbb{N}}$ by

$$\Phi((\Omega_n)_n)_m := \prod_{k<m} \Omega_k$$

A **discrete stochastic process** is a object Ω in $\mathcal{P}^{\mathbb{N}}$ together with a morphism $f : \Phi(\Omega) \to \Omega$ in $\mathcal{P}^{\mathbb{N}}$, i.e. a collection of Markov kernels

$$\left(f_m : \prod_{k<m} \Omega_k \to \mathcal{G}\Omega_m \right)_m .$$

- A morphism of discrete stochastic processes from (Ω^1, f^1) to (Ω^2, f^2) is a morphism $g : \Omega^1 \to \Omega^2$ in $\mathcal{P}^{\mathbb{N}}$ such that

$$g \circ f = f' \circ \Phi(g).$$

Let **Stoch** be the category of stochastic processes.

- A morphism of discrete stochastic processes from (Ω^1, f^1) to (Ω^2, f^2) is a morphism $g : \Omega^1 \to \Omega^2$ in $\mathcal{P}^{\mathbb{N}}$ such that

$$g \circ f = f' \circ \Phi(g).$$

  Let **Stoch** be the category of stochastic processes.

- Let N be the monoid of natural numbers, considered as a category. A **discrete Markov process** is a functor $N \to \mathcal{P}$, i.e. a measurable space Ω together with a Markov kernel $f : \Omega \to \mathcal{G}\Omega$. Let **Mark** be the category of disrete Markov processes, i.e. $[N, \mathcal{P}]$.

- A morphism of discrete stochastic processes from (Ω^1, f^1) to (Ω^2, f^2) is a morphism $g : \Omega^1 \to \Omega^2$ in $\mathcal{P}^{\mathbb{N}}$ such that

$$g \circ f = f' \circ \Phi(g).$$

  Let **Stoch** be the category of stochastic processes.
- Let N be the monoid of natural numbers, considered as a category. A **discrete Markov process** is a functor $N \to \mathcal{P}$, i.e. a measurable space Ω together with a Markov kernel $f : \Omega \to \mathcal{G}\Omega$. Let **Mark** be the category of disrete Markov processes, i.e. $[N, \mathcal{P}]$.
- Question: Does the inclusion **Mark** $\to$ **Stoch** have adjoints?

Giry: A categorical approach to probability theory (1982)

In this paper, Giry recognizes Lawvere's category of probabilistic mappings as the Kleisli category of a certain monad (*Giry monad*).

Giry: A categorical approach to probability theory (1982)

In this paper, Giry recognizes Lawvere's category of probabilistic mappings as the Kleisli category of a certain monad (*Giry monad*).

- For a measurable space Ω. Let $\mathcal{G}\Omega$ be the **set of probability measures** on Ω.

Giry: A categorical approach to probability theory (1982)

In this paper, Giry recognizes Lawvere's category of probabilistic mappings as the Kleisli category of a certain monad (*Giry monad*).

- For a measurable space Ω. Let $\mathcal{G}\Omega$ be the **set of probability measures** on Ω. For a measurable subset $A \subseteq \Omega$, we have an evaluation map

$$\mathrm{ev}_A : \mathcal{G}\Omega \to [0, 1].$$

  $\mathcal{G}\Omega$ becomes a measurable spaces by endowing it with the σ-algbra generated by the evaluation maps.

Giry: A categorical approach to probability theory (1982)

In this paper, Giry recognizes Lawvere's category of probabilistic mappings as the Kleisli category of a certain monad (*Giry monad*).

- For a measurable space Ω. Let $\mathcal{G}\Omega$ be the **set of probability measures** on Ω. For a measurable subset $A \subseteq \Omega$, we have an evaluation map

$$\mathrm{ev}_A : \mathcal{G}\Omega \to [0, 1].$$

  $\mathcal{G}\Omega$ becomes a measurable spaces by endowing it with the σ-algbra generated by the evaluation maps.

  For a measurable map $f : \Omega_1 \to \Omega_2$, **pushing forward** along f defines a measurable map $\mathcal{G}f : \mathcal{G}\Omega_1 \to \mathcal{G}\Omega_2$.

Giry: A categorical approach to probability theory (1982)

In this paper, Giry recognizes Lawvere's category of probabilistic mappings as the Kleisli category of a certain monad (*Giry monad*).

- For a measurable space Ω. Let $\mathcal{G}\Omega$ be the **set of probability measures** on Ω. For a measurable subset $A \subseteq \Omega$, we have an evaluation map

$$\mathrm{ev}_A : \mathcal{G}\Omega \to [0,1].$$

  $\mathcal{G}\Omega$ becomes a measurable spaces by endowing it with the σ-algbra generated by the evaluation maps.

  For a measurable map $f : \Omega_1 \to \Omega_2$, **pushing forward** along f defines a measurable map $\mathcal{G}f : \mathcal{G}\Omega_1 \to \mathcal{G}\Omega_2$.

  This gives an **endofunctor** $\mathcal{G}$: **Mble** $\to$ **Mble**.

- There is a measurable map $\mu_\Omega : \mathcal{G}\mathcal{G}\Omega \to \mathcal{G}\Omega$:

$$\mu_\Omega(\mathbf{P})(A) := \int_{\lambda \in \mathcal{G}\Omega} \lambda(A)\mathbf{P}(\mathrm{d}\lambda),$$

for all $\mathbf{P} \in \mathcal{G}\mathcal{G}\Omega$ and measurable subsets $A \subseteq \Omega$.

- There is a measurable map $\mu_\Omega : \mathcal{G}\mathcal{G}\Omega \to \mathcal{G}\Omega$:

$$\mu_\Omega(\mathbf{P})(A) := \int_{\lambda \in \mathcal{G}\Omega} \lambda(A)\mathbf{P}(\mathrm{d}\lambda),$$

  for all $\mathbf{P} \in \mathcal{G}\mathcal{G}\Omega$ and measurable subsets $A \subseteq \Omega$.
- We have a map $\eta_\Omega : \Omega \to \mathcal{G}\Omega$:

$$\eta_\Omega(\omega) := \delta_\omega,$$

  for all $\omega \in \Omega$.

- There is a measurable map $\mu_\Omega : \mathcal{G}\mathcal{G}\Omega \to \mathcal{G}\Omega$:

$$\mu_\Omega(\mathbf{P})(A) := \int_{\lambda \in \mathcal{G}\Omega} \lambda(A)\mathbf{P}(\mathrm{d}\lambda),$$

  for all $\mathbf{P} \in \mathcal{G}\mathcal{G}\Omega$ and measurable subsets $A \subseteq \Omega$.
- We have a map $\eta_\Omega : \Omega \to \mathcal{G}\Omega$:

$$\eta_\Omega(\omega) := \delta_\omega,$$

  for all $\omega \in \Omega$.
- These form natural transformation $\mu : \mathcal{G}\mathcal{G} \to \mathcal{G}$ and $\eta : 1_{\mathbf{Mble}} \to \mathcal{G}$ and $(\mathcal{G}, \mu, \eta)$ forms a monad, *the Giry monad*.

- There is a measurable map $\mu_\Omega : \mathcal{G}\mathcal{G}\Omega \to \mathcal{G}\Omega$:

$$\mu_\Omega(\mathbf{P})(A) := \int_{\lambda \in \mathcal{G}\Omega} \lambda(A)\mathbf{P}(\mathrm{d}\lambda),$$

  for all $\mathbf{P} \in \mathcal{G}\mathcal{G}\Omega$ and measurable subsets $A \subseteq \Omega$.
- We have a map $\eta_\Omega : \Omega \to \mathcal{G}\Omega$:

$$\eta_\Omega(\omega) := \delta_\omega,$$

  for all $\omega \in \Omega$.
- These form natural transformation $\mu : \mathcal{G}\mathcal{G} \to \mathcal{G}$ and $\eta : 1_{\mathbf{Mble}} \to \mathcal{G}$ and $(\mathcal{G}, \mu, \eta)$ forms a monad, *the Giry monad*.
- Lawvere's category of probabilistic mappings is the Kleisli category of the Giry monad.

- There is a measurable map $\mu_\Omega : \mathcal{G}\mathcal{G}\Omega \to \mathcal{G}\Omega$:

$$\mu_\Omega(\mathbf{P})(A) := \int_{\lambda \in \mathcal{G}\Omega} \lambda(A)\mathbf{P}(\mathrm{d}\lambda),$$

  for all $\mathbf{P} \in \mathcal{G}\mathcal{G}\Omega$ and measurable subsets $A \subseteq \Omega$.
- We have a map $\eta_\Omega : \Omega \to \mathcal{G}\Omega$:

$$\eta_\Omega(\omega) := \delta_\omega,$$

  for all $\omega \in \Omega$.
- These form natural transformation $\mu : \mathcal{G}\mathcal{G} \to \mathcal{G}$ and $\eta : 1_{\mathbf{Mble}} \to \mathcal{G}$ and $(\mathcal{G}, \mu, \eta)$ forms a monad, *the Giry monad*.
- Lawvere's category of probabilistic mappings is the Kleisli category of the Giry monad.
- Giry also introduces a monad on **Pol**, the category of Polish spaces and continuous functions.

- *The Kolmogorov extension problem*: Let $(\Omega_i)_{i \in I}$ be a collection of measurable spaces. Consider a probability measure $\mathbb{P}_J$ on $\prod_{j \in J} \Omega_j$ for every finite set $J \subseteq I$. Suppose that this collection is *consistent*.

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

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

  for all finite $J \subseteq I$?

  Answer: sometimes. (*Kolmorogov extension problem*)

- *The Kolmogorov extension problem*: Let $(\Omega_i)_{i \in I}$ be a collection of measurable spaces. Consider a probability measure $\mathbb{P}_J$ on $\prod_{j \in J} \Omega_j$ for every finite set $J \subseteq I$. Suppose that this collection is *consistent*.

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

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

  for all finite $J \subseteq I$?

  Answer: sometimes. (*Kolmorogov extension problem*)

  Giry translates this problem as follows:

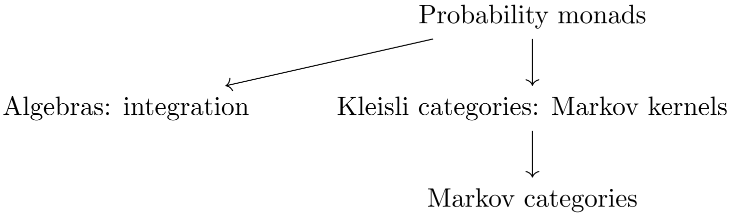
  'Does the functor **Mble** $\to$ **Mble**$_{\mathcal{G}}$ preserve cofiltered limits?'.

  Several (technical) conditions are given for which this is the case. As a corollary the *Ionescu-Tulcea theorem* is discussed.

- Lawvere expressed *discrete* Markov processes using $\mathcal{P}(= \mathbf{Mble}_{\mathcal{G}})$. In probability theory, we also want to talk about *continuous* stochastic processes.

- Lawvere expressed *discrete* Markov processes using $\mathcal{P}(= \mathbf{Mble}_{\mathcal{G}})$. In probability theory, we also want to talk about *continuous* stochastic processes.

  For this, Giry introduced **random topological actions**. Let C be a **category internal to Pol**. A random topological action is functor $C \to \mathbf{Pol}_{\mathcal{G}}$ satisfying certain (continuity) conditions.


Probability monads
Algebras: integration
Kleisli categories: Markov kernels
Markov categories

Swirszcz: Monadic functors and convexity (1973)

Studies which monads have the category of convex spaces as algebras.

Swirszcz: Monadic functors and convexity (1973)

Studies which monads have the category of convex spaces as algebras.

- The category **Conv**:
  - objects: convex subsets of vector spaces,
  - morphims: affine maps

Swirszcz: Monadic functors and convexity (1973)

Studies which monads have the category of convex spaces as algebras.

- The category **Conv**:
  - objects: convex subsets of vector spaces,
  - morphims: affine maps
- The forgetful functor **Conv** $\rightarrow$ **Set** is **not** monadic.

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Comp** *is* monadic.

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Comp** *is* monadic.

  The corresponding monad is the *Radon monad*, that sends a compact Hausdorff space X to the space $\mathcal{R}X$ of Radon probability measures on X.

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Comp** *is* monadic.

  The corresponding monad is the *Radon monad*, that sends a compact Hausdorff space X to the space $\mathcal{R}X$ of Radon probability measures on X.

  The algebra structure on a compact convex space X is given by the *barycenter map* $b : \mathcal{R}X \to X$,

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Comp** *is* monadic.

  The corresponding monad is the *Radon monad*, that sends a compact Hausdorff space X to the space $\mathcal{R}X$ of Radon probability measures on X.

  The algebra structure on a compact convex space X is given by the *barycenter map* $b : \mathcal{R}X \to X$,which sends a Radon probability measure $\mathbb{P}$ to the unique element $x \in X$ such that

$$f(x) = \int f \mathrm{d}\mathbb{P}$$

  for all continuous affine maps $f : X \to \mathbb{R}$.

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Comp** *is* monadic.

  The corresponding monad is the *Radon monad*, that sends a compact Hausdorff space X to the space $\mathcal{R}X$ of Radon probability measures on X.

  The algebra structure on a compact convex space X is given by the *barycenter map* $b : \mathcal{R}X \to X$,which sends a Radon probability measure $\mathbb{P}$ to the unique element $x \in X$ such that

$$f(x) = \int f \mathrm{d}\mathbb{P}$$

  for all continuous affine maps $f : X \to \mathbb{R}$.

  The results uses a monadicity theorem (Linton).

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Set** *is* monadic.

- The category **CompConv**:
  - objects: compact convex subsets of locally convex topological vector spaces,
  - morphisms: affine maps
- The forgetful functor **CompConv** $\to$ **Set** *is* monadic.

  The corresponding monad on **Set** sends a set X to the set of Radon measures on the Čech-Stone compactifcation of X.

Probability monads

Algebras: integration

Kleisli categories: Markov kernels

Markov categories

Probability monads

Since then, many variations on the Giry monad have been studied and are referred to as *probability monads*.

- *Distribution monad*: A probability monad on **Set**, sending every set X to

$$\{\text{finitely supported probability measures on } X\}.$$

Probability monads

Since then, many variations on the Giry monad have been studied and are referred to as *probability monads*.

- *Distribution monad*: A probability monad on **Set**, sending every set X to

$$\{\text{finitely supported probability measures on } X\}.$$

- Monads of valuations on topological spaces and locales (Fritz, Perrone, Vickers)

Probability monads

Since then, many variations on the Giry monad have been studied and are referred to as *probability monads*.

- *Distribution monad*: A probability monad on **Set**, sending every set X to

$$\{\text{finitely supported probability measures on } X\}.$$

- Monads of valuations on topological spaces and locales (Fritz, Perrone, Vickers)
- Monads of subprobabilities/stochastic relations (Panangaden)

- *Kantorovich monad* (van Breugel, Fritz, Perrone): A probabilty monad on $\mathbf{CMet}_1$, the category of complete metric spaces and 1-Lipschitz maps. A complete metric space is send to its *Kantorovich space*

- *Kantorovich monad* (van Breugel, Fritz, Perrone): A probabilty monad on $\mathbf{CMet}_1$, the category of complete metric spaces and 1-Lipschitz maps. A complete metric space is send to its *Kantorovich space*

  A Radon probability measure $\mathbb{P}$ **has finite moment** if

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

- *Kantorovich monad* (van Breugel, Fritz, Perrone): A probabilty monad on $\mathbf{CMet}_1$, the category of complete metric spaces and 1-Lipschitz maps. A complete metric space is send to its *Kantorovich space*

  A Radon probability measure $\mathbb{P}$ **has finite moment** if

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

  The **Kantorovich space** of a complete metric space X is the complete metric space of all Radon probability measure that have finite moment. The metric is given by the *Wasserstein distance*:

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

  for all Radon probability measure of finite moment $\mathbb{P}_1$ and $\mathbb{P}_2$.

Kantorovich spaces and Wasserstein distances are important in *transport theory* and have many useful properties.

Kantorovich spaces and Wasserstein distances are important in *transport theory* and have many useful properties.

E.g. the finitely supported probability measures are dense in the Kantorovich space.

Kantorovich spaces and Wasserstein distances are important in *transport theory* and have many useful properties.

E.g. the finitely supported probability measures are dense in the Kantorovich space.

- Many *other metric variations*: ordered metric spaces, compact metric spaces, general Lipschitz maps, ...

Kantorovich spaces and Wasserstein distances are important in *transport theory* and have many useful properties.

E.g. the finitely supported probability measures are dense in the Kantorovich space.

- Many *other metric variations*: ordered metric spaces, compact metric spaces, general Lipschitz maps, ...
- Many probability monads are *codensity monads* of functor of probability measures on *countable* spaces.

What are the algebras of probability monads?

The algebras of probability monads

From Swirszcz, we already know that algebras of probability monads should have a *convex structure* and the structure map should give a *barycenter* or *centre of mass*.

- *Example*: Define a map $\alpha : \mathcal{G}[0,\infty] \to [0,\infty]$ by

$$\mathbb{P} \mapsto \int_0^\infty x\mathbb{P}(\mathrm{d}x).$$

The algebras of probability monads

From Swirszcz, we already know that algebras of probability monads should have a *convex structure* and the structure map should give a *barycenter* or *centre of mass*.

- *Example*: Define a map $\alpha : \mathcal{G}[0,\infty] \to [0,\infty]$ by

$$\mathbb{P} \mapsto \int_0^\infty x\mathbb{P}(\mathrm{d}x).$$

  - $\alpha(\delta_{x_0}) = \int x\delta_{x_0}(\mathrm{d}x) = x_0$ for all $x_0 \in [0,\infty]$,

The algebras of probability monads

From Swirszcz, we already know that algebras of probability monads should have a *convex structure* and the structure map should give a *barycenter* or *centre of mass*.

- *Example*: Define a map $\alpha : \mathcal{G}[0, \infty] \to [0, \infty]$ by

$$\mathbb{P} \mapsto \int_0^\infty x\mathbb{P}(\mathrm{d}x).$$

  - $\alpha(\delta_{x_0}) = \int x\delta_{x_0}(\mathrm{d}x) = x_0$ for all $x_0 \in [0, \infty]$,
  - For $\mathbf{P} \in \mathcal{G}[0, \infty]$,

$$\alpha(\mu(\mathbf{P})) = \int_0^\infty x\mu(\mathbf{P})(\mathrm{d}x) = \int_\lambda \int_0^\infty x\lambda(\mathrm{d}x)\mathbf{P}(\mathrm{d}\lambda) = \int_\lambda \alpha(\lambda)\mathbf{P}(\mathrm{d}) = \alpha(\mathbf{P}\circ\alpha^{-1})$$

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\to$ Random variables should take values in *algebras of probability monads*.

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

  $$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

  $\to$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\to$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).
  - Given a convex space X, we can define $\alpha\left(\sum_{n=1}^{N} \alpha_n \delta_{x_n}\right) = \sum_{n=1} \alpha_n x_n$.

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\to$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).
  - Given a convex space X, we can define $\alpha\left(\sum_{n=1}^{N} \alpha_n \delta_{x_n}\right) = \sum_{n=1} \alpha_n x_n$.
  - Give an algebra A, we define a convex structure by $\lambda a + (1-\lambda)b := \alpha(\lambda \delta_a + (1-\lambda)\delta_b)$.

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\rightarrow$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).
  - Given a convex space X, we can define $\alpha\left(\sum_{n=1}^{N} \alpha_n \delta_{x_n}\right) = \sum_{n=1} \alpha_n x_n$.
  - Give an algebra A, we define a convex structure by $\lambda a + (1-\lambda)b := \alpha(\lambda \delta_a + (1-\lambda)\delta_b)$.
- *Radon monad*: The algebras are compact convex subsets of locally convex topological vector spaces. (Swirszcz)

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\rightarrow$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).
  - Given a convex space X, we can define $\alpha\left(\sum_{n=1}^{N} \alpha_n \delta_{x_n}\right) = \sum_{n=1} \alpha_n x_n$.
  - Give an algebra A, we define a convex structure by $\lambda a + (1-\lambda) b := \alpha(\lambda \delta_a + (1-\lambda)\delta_b)$.
- *Radon monad*: The algebras are compact convex subsets of locally convex topological vector spaces. (Swirszcz)
- *Kantorovich monad*: The algebras are closed convex subsets of Banach spaces. (Perrone, Fritz)

- Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to [0, \infty]$ be a random variable.

$$\mathbb{E}[X] = \alpha(\mathbb{P} \circ X^{-1}).$$

$\to$ Random variables should take values in *algebras of probability monads*.

- *Distribution monad*: The algebras are convex spaces (in the sense of Stone).
  - Given a convex space X, we can define $\alpha\left(\sum_{n=1}^{N} \alpha_n \delta_{x_n}\right) = \sum_{n=1} \alpha_n x_n$.
  - Give an algebra A, we define a convex structure by $\lambda a + (1 - \lambda)b := \alpha(\lambda\delta_a + (1 - \lambda)\delta_b)$.
- *Radon monad*: The algebras are compact convex subsets of locally convex topological vector spaces. (Swirszcz)
- *Kantorovich monad*: The algebras are closed convex subsets of Banach spaces. (Perrone, Fritz)
- *Giry monad*: more difficult! (Dobberkat, Sturtz)

Perrone: partial evaluations

- *Conditional expectation*: Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra.

Perrone: partial evaluations

- *Conditional expectation*: Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra.

$$\mathbb{E}[X \mid \mathcal{G}] : (\Omega, \mathcal{G}, \mathbb{P}\mid_{\mathcal{G}}) \to \mathbb{R}$$

is the $\mathbb{P}$-almost surely unique random variable such that

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] 1_A] = \mathbb{E}[X 1_A],$$

for all $A \in \mathcal{G}$.

Perrone: partial evaluations

- *Conditional expectation*: Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra.

$$\mathbb{E}[X \mid \mathcal{G}] : (\Omega, \mathcal{G}, \mathbb{P} \mid_{\mathcal{G}}) \to \mathbb{R}$$

  is the $\mathbb{P}$-almost surely unique random variable such that

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]1_A] = \mathbb{E}[X1_A],$$

  for all $A \in \mathcal{G}$.

- Let (A, α) be an algebra of a probability monad T and let $\mathbb{P}_1$ and $\mathbb{P}_2$ be in TA. Then $\mathbb{P}_2$ is **a partial evaluation** of $\mathbb{P}_1$ if there exists $\mathbf{P} \in TTA$ such that

$$\mathbf{P} \circ \alpha^{-1} = \mathbb{P}_1 \quad \text{and} \quad \mu_A(\mathbf{P}) = \mathbb{P}_2.$$

- *Example*: $\mathbb{P}|_{\mathcal{G}} \circ \mathbb{E}[X \mid \mathcal{G}]^{-1}$ is a partial evaluation of $\mathbb{P} \circ X^{-1}$.

- *Example*: $\mathbb{P}\mid_{\mathcal{G}} \circ \mathbb{E}[X \mid \mathcal{G}]^{-1}$ is a partial evaluation of $\mathbb{P} \circ X^{-1}$.

  Moreover, every $\mathbb{P}_1, \mathbb{P}_2$ in *TA* such that $\mathbb{P}_2$ is a partial evaluation of $\mathbb{P}_1$ are of this form for some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, sub-$\sigma$-algebra $\mathcal{G} \subseteq \mathcal{F}$ and random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$. (Perrone)

Kleisli categories of probability monads and Markov categories

Fritz: Markov categories

Markov categories are monoidal categories, similar to Kleisli categories of probability monads.

Fritz: Markov categories

Markov categories are monoidal categories, similar to Kleisli categories of probability monads.
For measurable spaces Ω_1 and Ω_2, we have morphims

$$\mathcal{G}\Omega_1 \times \mathcal{G}\Omega_2 \to \mathcal{G}(\Omega_1 \times \Omega_2)$$

by sending $(\mathbb{P}_1, \mathbb{P}_2)$ to $\mathbb{P}_1 \otimes \mathbb{P}_2$.

Fritz: Markov categories

Markov categories are monoidal categories, similar to Kleisli categories of probability monads.
For measurable spaces Ω_1 and Ω_2, we have morphims

$$\mathcal{G}\Omega_1 \times \mathcal{G}\Omega_2 \to \mathcal{G}(\Omega_1 \times \Omega_2)$$

by sending $(\mathbb{P}_1, \mathbb{P}_2)$ to $\mathbb{P}_1 \otimes \mathbb{P}_2$. This makes the Giry monad into a *commutative monad*.
Therefore the Kleisli category of the probability monad inherits the *symmetric (semicartesian) monoidal* structure on **Mble**.

Fritz: Markov categories

Markov categories are monoidal categories, similar to Kleisli categories of probability monads.
For measurable spaces Ω_1 and Ω_2, we have morphims

$$\mathcal{G}\Omega_1 \times \mathcal{G}\Omega_2 \to \mathcal{G}(\Omega_1 \times \Omega_2)$$

by sending $(\mathbb{P}_1, \mathbb{P}_2)$ to $\mathbb{P}_1 \otimes \mathbb{P}_2$. This makes the Giry monad into a *commutative monad*.
Therefore the Kleisli category of the probability monad inherits the *symmetric (semicartesian) monoidal* structure on **Mble**.
There are similar constructions for other probability monads.

Let Ω be a measurable space. Every objects has a canonical *commutative comonoid* structure:

- The *comultiplication* $\Omega \to \mathcal{G}(\Omega \times \Omega)$ is defined by

$$\omega \mapsto \delta_{(\omega,\omega)}.$$

- The *conuit* is the unique map $X \to \mathcal{G}\mathbf{1}$.

Let Ω be a measurable space. Every objects has a canonical *commutative comonoid* structure:

- The *comultiplication* $\Omega \to \mathcal{G}(\Omega \times \Omega)$ is defined by

$$\omega \mapsto \delta_{(\omega,\omega)}.$$

- The *conuit* is the unique map $X \to \mathcal{G}\mathbf{1}$.

A **Markov category** is a symmetric monoidal category with a comultiplication structure on every object,

Let Ω be a measurable space. Every objects has a canonical *commutative comonoid* structure:

- The *comultiplication* $\Omega \to \mathcal{G}(\Omega \times \Omega)$ is defined by

$$\omega \mapsto \delta_{(\omega,\omega)}.$$

- The *conuit* is the unique map $X \to \mathcal{G}\mathbf{1}$.

A **Markov category** is a symmetric monoidal category with a comultiplication structure on every object, such that the counits form a natural transformation.

A lot has been written about Markov categories, including versions of the

- de Finetti theorem,
- certain 0-1 laws,
- the ergodic decomposition theorem

in Markov categories (Fritz, Perrone,Moss, ...).

A lot has been written about Markov categories, including versions of the

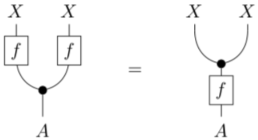
- de Finetti theorem,
- certain 0-1 laws,
- the ergodic decomposition theorem

in Markov categories (Fritz, Perrone,Moss, ...).
We will look at some important definitions in concepts in Markov categories:

- deterministic morphism,
- almost surely equal morphisms
- conditionals
- Kolmogorov products

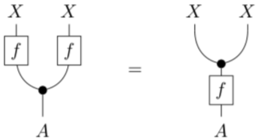
Let $\mathcal{C}$ be a Markov category:

- A morphism $f : A \to X$ is **deterministic** if



Let $\mathcal{C}$ be a Markov category:

- A morphism $f : A \to X$ is **deterministic** if



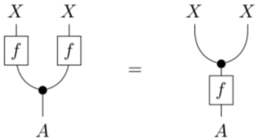
For $\mathbf{Mble}_{\mathcal{G}}$, this means

$$f(a)(A)f(a)(B) = f(a)(A \cap B)$$

for all $a \in A$ and measurable $A, B \subseteq X$.

Let $\mathcal{C}$ be a Markov category:

- A morphism $f : A \to X$ is **deterministic** if



For $\mathbf{Mble}_{\mathcal{G}}$, this means

$$f(a)(A)f(a)(B) = f(a)(A \cap B)$$

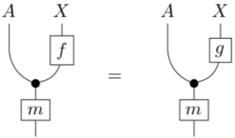
for all $a \in A$ and measurable $A, B \subseteq X$. In particular,

$$f(a)(A)^2 = f(a)(A)$$

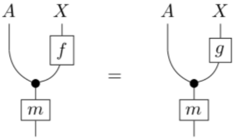
and therefore $f(a)(A) \in \{0, 1\}$.

- A **representable Markov category** is a Markov category that is the Kleisli category of some monad on $\mathcal{C}_{\mathrm{det}}$, the category with the same objects as $\mathcal{C}$ and *deterministic* morphisms

- Let $m : I \to X$ and $f, g : A \to X$ in $\mathcal{C}$, then f and g are m-**almost surely equal** if



- Let $m : I \to X$ and $f, g : A \to X$ in $\mathcal{C}$, then f and g are m-**almost surely equal** if



In $\mathbf{Mble}_{\mathcal{G}}$, this means that

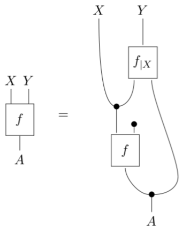
$$\int_B f(a)(A)m(\mathrm{d}a) = \int_B g(a)(A)m(\mathrm{d}a)$$

for all $a \in A$ and measurable $A, B \subseteq X$.
Therefore,

$$f(\cdot)(A) = g(\cdot)(A) \quad m - \text{almost surely.}$$

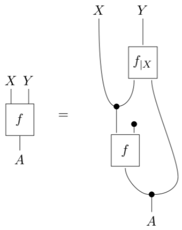
Conditionals in Markov categories

A Markov category **has conditionals** if for every $f : A \to X \otimes Y$, there exists $f\,|_X \colon X \otimes A \to Y$ such that



Conditionals in Markov categories

A Markov category **has conditionals** if for every $f : A \to X \otimes Y$, there exists $f|_X : X \otimes A \to Y$ such that

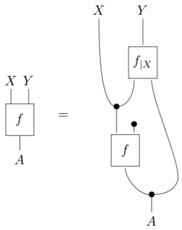


$\mathbf{Mble}_{\mathcal{G}}$ does **not** have conditionals, since not every $\mathbb{P}$ on $\Omega_1 \times \Omega_2$ is of the form

$$A \times B \mapsto \int_B f(\omega)(B)\mathbb{P}(\mathrm{d}\omega)$$

for some Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_2$.

A Markov category **has conditionals** if for every $f : A \to X \otimes Y$, there exists $f\mid_X : X \otimes A \to Y$ such that



$\mathbf{Mble}_{\mathcal{G}}$ does **not** have conditionals, since not every $\mathbb{P}$ on $\Omega_1 \times \Omega_2$ is of the form

$$A \times B \mapsto \int_B f(\omega)(B)\mathbb{P}(\mathrm{d}\omega)$$

for some Markov kernel $f : \Omega_1 \to \mathcal{G}\Omega_2$.
Standard Borel spaces do have conditionals.

Kolmogorov products in Markov categories

Let $\mathcal{C}$ be a Markov category and let $(X_i)_{i \in I}$ be a collection of objects.

Kolmogorov products in Markov categories

Let $\mathcal{C}$ be a Markov category and let $(X_i)_{i \in I}$ be a collection of objects. Suppose that

- The limit $\lim_{J \subseteq I} \bigotimes_{j \in J} X_j$ over all finite subsets J of I exists,

Kolmogorov products in Markov categories

Let $\mathcal{C}$ be a Markov category and let $(X_i)_{i \in I}$ be a collection of objects. Suppose that

- The limit $\lim_{J \subseteq I} \bigotimes_{j \in J} X_j$ over all finite subsets J of I exists,
- and that it is preseved by $-Y$ for all objects Y,

Kolmogorov products in Markov categories

Let $\mathcal{C}$ be a Markov category and let $(X_i)_{i\in I}$ be a collection of objects. Suppose that

- The limit $\lim_{J\subseteq I} \bigotimes_{j\in J} X_j$ over all finite subsets J of I exists,
- and that it is preseved by $-Y$ for all objects Y,
- and that

$$\pi_{J'\subseteq J} : \bigotimes_{j\in J} X_j \to \bigotimes_{j\in J'} X_j$$

is *deterministic* for all finites sets $J' \subseteq J$ of I.

Then $\lim_{J\subseteq I} \bigotimes_{j\in J} X_j$ is called the **Kolmogorov product** of $(X_i)_{i\in I}$.

Kolmogorov products in Markov categories

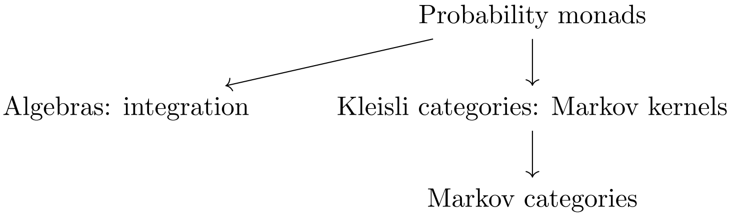
Let $\mathcal{C}$ be a Markov category and let $(X_i)_{i\in I}$ be a collection of objects. Suppose that

- The limit $\lim_{J\subseteq I} \bigotimes_{j\in J} X_j$ over all finite subsets J of I exists,
- and that it is preseved by $-Y$ for all objects Y,
- and that

$$\pi_{J'\subseteq J} : \bigotimes_{j\in J} X_j \to \bigotimes_{j\in J'} X_j$$

  is *deterministic* for all finites sets $J' \subseteq J$ of I.

Then $\lim_{J\subseteq I} \bigotimes_{j\in J} X_j$ is called the **Kolmogorov product** of $(X_i)_{i\in I}$. In $\mathbf{Mble}_{\mathcal{G}}$ Kolmogorov products don't exist in general, but countable Kolmogorov products of $\mathrm{Standard\ Borel\ spaces}$ do exist (*Kolmogorov extension theorem*).


Probability monads
Algebras: integration
Kleisli categories: Markov kernels
Markov categories