# De Finetti's Theorem in Categorical Probability

Tobias Fritz
joint work with Tomáš Gonda and Paolo Perrone

July 2021

# References


- Kenta Cho and Bart Jacobs,
  **Disintegration and Bayesian inversion via string diagrams**. *Math. Struct. Comp. Sci.* 29, 938–971 (2019). arXiv:1709.00322.

- Tobias Fritz,
  **A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics**.
  *Adv. Math.* 370, 107239 (2020). arXiv:1908.07021.

- Tobias Fritz and Eigil Fjeldgren Rischel,
  **The zero-one laws of Kolmogorov and Hewitt–Savage in categorical probability**.
  *Compositionality* 2, 3 (2020). arXiv:1912.02769.

- Tobias Fritz, Tomáš Gonda, Paolo Perrone, Eigil Fjeldgren Rischel,
  **Representable Markov Categories and Comparison of Statistical Experiments in Categorical Probability**.
  arXiv:2010.07416.

- Bart Jacobs, Sam Staton,
  **De Finetti's construction as a categorical limit**.
  *Coalgebraic Methods in Computer Science* 2020. arXiv:2003.01964.

- Tobias Fritz, Tomáš Gonda, Paolo Perrone,
  **De Finetti's Theorem in Categorical Probability**.
  arXiv:2105.02639


For a broader perspective, see the videos from the online workshop Categorical Probability and Statistics!

# Overview

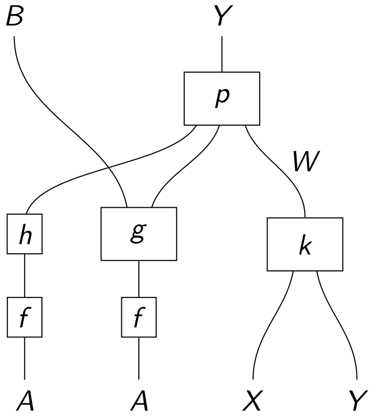
- Goal: state and prove a classical theorem of probability theory without talking about (numerical) probabilities.
- Based on a recent categorical approach to probability.
- The big picture:

| Traditional probability theory | Categorical probability theory |
|---|---|
| **Analytic:** says what probabilities are | **Synthetic:** says how probabilities behave |
| Analogous to number systems | Analogous to abstract algebra |

The basic primitives are morphisms in a symmetric monoidal category:


- **Intuitively**, a morphism is a probabilistic function: random output for any input.
- We impose axiom that (partly) formalize this intuition.

We can compose morphisms using string diagram calculus, like this:

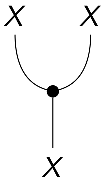



This defines an overall morphism

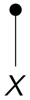
$$A \otimes A \otimes X \otimes Y \longrightarrow B \otimes Y.$$

Postulate additional pieces of structure:

- ▷ Every object $X$ has a **copying function**:

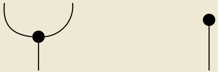



- ▷ Every object $X$ has a **deletion function**:

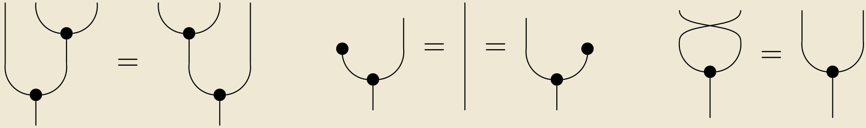


## Definition

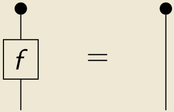
A **Markov category C** is a symmetric monoidal category supplied with **copying** and **deleting** operations on every object,



giving commutative comonoid structures



which interact well with the monoidal structure, and such that for all $f$,



# Semantics

**BorelStoch** is the category with:

- **Standard Borel spaces** as objects (finite sets, $\mathbb{N}$ and $[0,1]$).
- Measurable **Markov kernels** as morphisms.
- Products of measurable spaces for $\otimes$.

**BorelStoch** satisfies all of the axioms that I will mention.

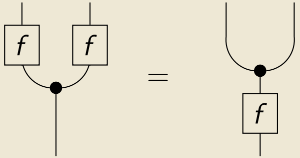
It is the Kleisli category of the **Giry monad**!

# Determinism

Throughout, we're in a Markov category $\mathbf{C}$.

**Definition**

A morphism $f : X \to Y$ is **deterministic** if it commutes with copying,




- **Intuition:** Applying $f$ to copies of input = copying the output of $f$.
- The deterministic morphisms form a cartesian monoidal subcategory $\mathbf{C}_{\mathrm{det}}$.

# Representability

**Definition**

A Markov category $\mathbf{C}$ is **representable** if for every $X \in \mathbf{C}$ there is $PX \in \mathbf{C}$ and a natural bijection

$$\mathbf{C}_{\mathrm{det}}(-, PX) \cong \mathbf{C}(-, X),$$

and **a.s.-compatibly representable** if this respects $p$-a.s. equality for every $p$.

- **Intuition:** $PX$ is space of probability measures on $X$.
- Under the bijection, the deterministic $\mathrm{id} : PX \to PX$ corresponds to

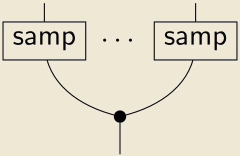
  $$\mathsf{samp}_X : PX \to X,$$

  the map that returns a random sample from a distribution.

**BorelStoch** is representable in a very particular way:

### Theorem (De Finetti, abstract version)

$PX$ is the equalizer of all the finite permutations on $X^{\mathbb{N}}$, with universal arrow given by




- Intuition:

$$\text{probability distribution} \;=\; \text{prescription of how to sample from it}$$

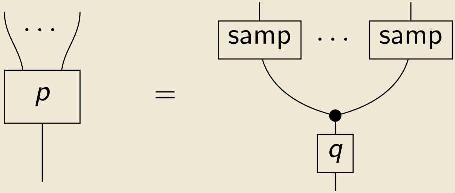
- Difficult to prove: existence part of universal property.

# The de Finetti theorem

**Theorem**

Let **C** be an a.s.-compatibly representable Markov category with conditionals and countable Kolmogorov products.

Then for every $p : A \to X^{\mathbb{N}}$ invariant under finite permutations, there is $q : A \to PX$ such that




- **BorelStoch** satisfies these assumptions.
- Mystery: we know of no other nontrivial Markov category which does!

# Detour: de Finetti and Bayesianism

- Suppose that I hand you a coin (which may be biased).
- How much would you bet on the outcome

$$\text{heads}, \quad \text{tails}, \quad \text{tails}$$

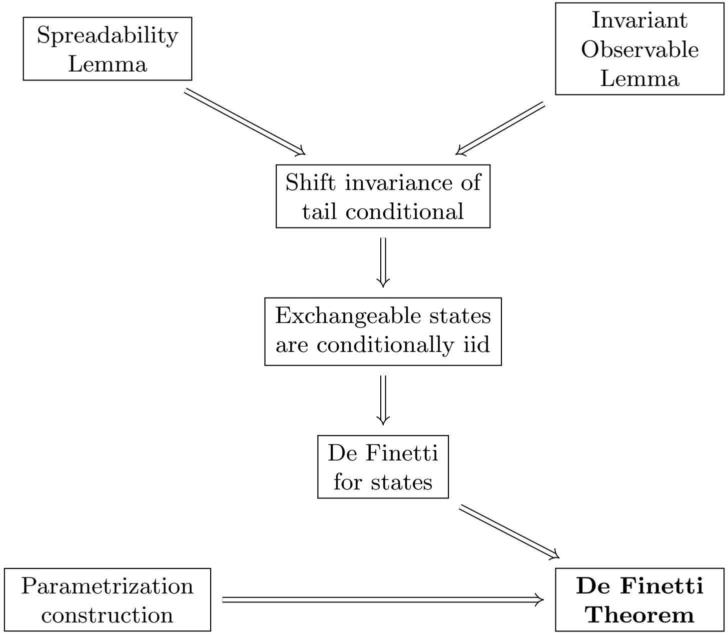
  when the coin is flipped 3 times?

  $\Rightarrow$ Surely the same as you would bet on

$$\text{tails}, \quad \text{tails}, \quad \text{heads}.$$

- Your bets satisfy **permutation invariance**.
  $\Rightarrow$ They correspond to a measure on $[0, 1]$, the space of biases.
- For a Bayesian, this is the **prior** over the biases.

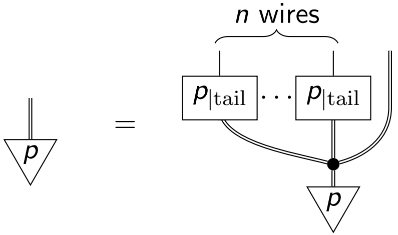
# Structure of proof



# Proof teaser

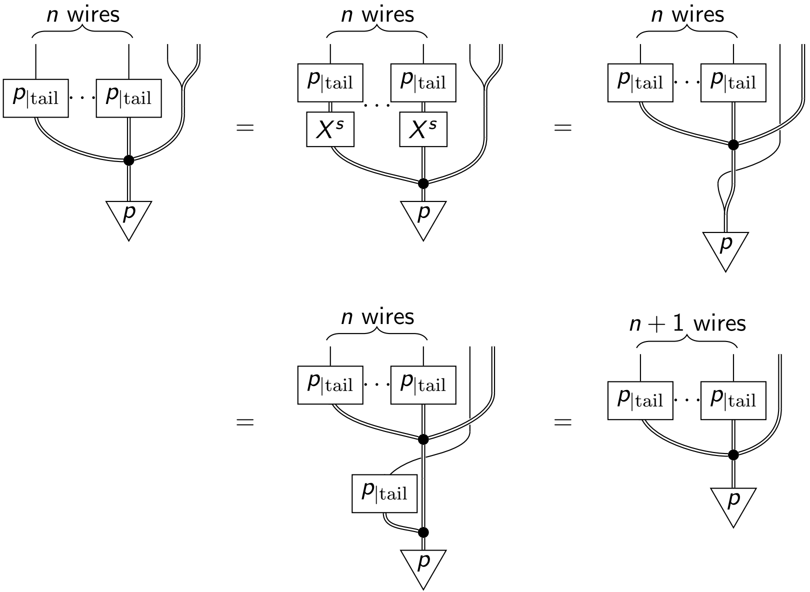
Suppose that $p$ is a state.

By universal property of Kolmogorov products, it is enough to show




for every finite $n$.

Using induction on $n$,




□

# Summary and Outlook

- Markov categories are an emerging framework for “synthetic” probability theory.
- We already have synthetic versions of several theorems of probability and statistics:
  - 0/1-laws of Kolmogorov and Hewitt-Savage,
  - Fisher factorization theorem on sufficient statistics,
  - Blackwell-Sherman-Stein theorem on informativeness of statistical experiments,
  - **De Finetti’s theorem** on permutation-invariant distribution.

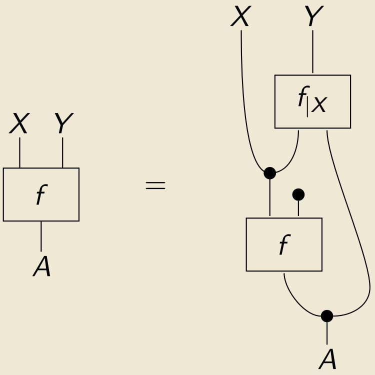
# Summary and Outlook

- Sometimes such developments require turning theorems into definitions.
- **Next:** a synthetic treatment of the law of large numbers.
- This has further tantalizing connections with ergodic theory.
- In parallel, we also aim at a better understanding of the semantics.
- Central question here: how common are Markov categories with conditionals?

# Bonus slides: Conditionals

**Definition**

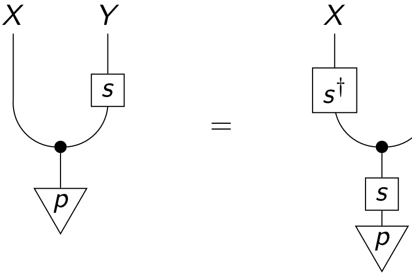
**C has conditionals** if for every $f : A \to X \otimes Y$ there is $f_{|X} : X \otimes A \to Y$ with




▷ **Intuition:** The outputs of $f$ can be generated one at a time.

# Bayesian inversion

Every $s : X \to Y$ has a **Bayesian adjoint** $s^\dagger : Y \to X$ satisfying:



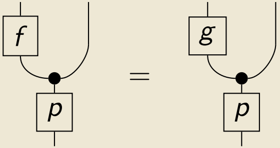

The Bayesian adjoint $s^\dagger$ depends on $p$.

# Almost sure equality

**Definition**

Let $p : A \to X$ and $f, g : X \to Y$.

$f$ and $g$ are **equal** $p$**-almost surely**, $f =_{p\text{-a.s.}} g$, if




- **Intuition:** $f$ and $g$ behave the same on all inputs produced by $p$.
- Other concepts (besides equality) also relativize with respect to $p$-almost surely.

# Infinite tensor products

Let $(X_i)_{i \in I}$ be a family of objects.

For finite $F \subseteq F' \subseteq I$, we have projection morphisms

$$\bigotimes_{i \in F'} X_i \longrightarrow \bigotimes_{i \in F} X_i$$

given by composing with deletion for all $i \in F' \setminus F$.

# Infinite tensor products

**Definition**

The **infinite tensor product**

$$X^I := \bigotimes_{i \in I} X_i$$

is the limit of the finite tensor products $X^F := \bigotimes_{i \in F} X_i$ if it exists and is preserved by every $- \otimes Y$.

- **Intuition:** To map into an infinite tensor product, one needs to map consistently into its finite subproducts.

# Kolmogorov products

**Definition**

An infinite tensor product $X^I$ is a **Kolmogorov product** if the limit projections $\pi^F : X^I \to X^F$ are deterministic.

- This additional condition fixes the comonoid structure on $X^I$.
- We need countable Kolmogorov products already in order to state the de Finetti theorem.

# Spreadability Lemma

**Lemma**

If $p : A \to X^{\mathbb{N}}$ is exchangeable, then $p$ is also invariant with respect to applying any injective map $\mathbb{N} \to \mathbb{N}$ to the tensor factors.

- **Intuition:** If random variables $X_1, X_2, \dots$ are permutation-invariant, then they have the same distribution as $X_2, X_3, \dots$

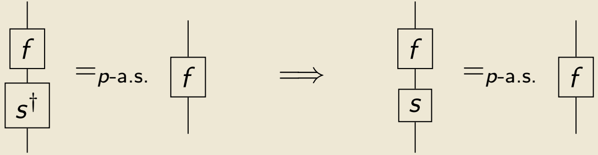
*Proof sketch.* On every finite $F \subseteq \mathbb{N}$, every injection $\mathbb{N} \to \mathbb{N}$ coincides with a suitable permutation.

# Invariant Observable Lemma

**Lemma**

Let $p\colon I \to X$ and $s\colon X \to X$ satisfy $sp = p$.

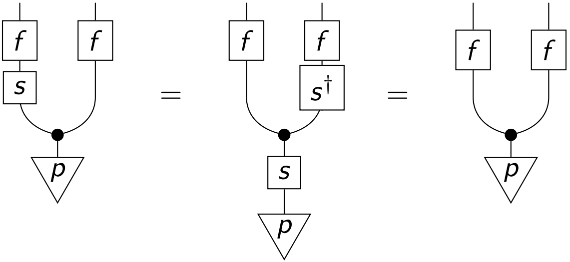
Then for deterministic $f\colon X \to Y$,




- **Intuition:** $s$ and $p$ make $X$ into a **measure-preserving dynamical system**, $f$ is an observable.
- If $f$ is invariant "backward in time", then it is also invariant "forward in time".

# Invariant Observable Lemma

*Proof sketch.*



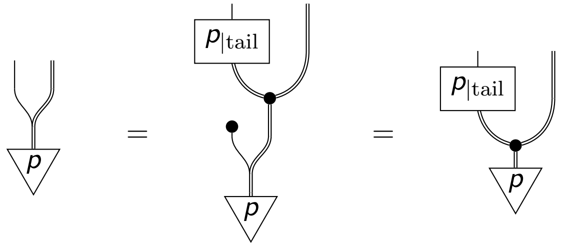

Like an equation between inner products in "$L^2(A, p)$".

$\Rightarrow$ The claim follows by "Cauchy-Schwarz". □

# The tail conditional

We use double wires to denote $X^{\mathbb{N}}$.

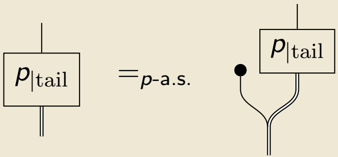
By the existence of conditionals, there is $p_{|\mathrm{tail}}$ such that




The second equation is by the Spreadability Lemma.

# Shift invariance of the tail conditional

**Lemma**




- **Intuition:** $p_{|\mathrm{tail}}$ is independent of any finite initial segment.

*Proof sketch.* An application of the Invariant Observable Lemma. Its assumption holds by the Spreadability Lemma. □

# Kleisli categories are Markov categories

**Proposition**

Let

- **D** be a category with finite products,
- $P$ a commutative monad on **D** with $P(1) \cong 1$.

Then the Kleisli category $\mathrm{Kl}(P)$ is a Markov category in the obvious way.

Examples:

- Kleisli category of the Giry monad, other related monads for measure-theoretic probability.
- Kleisli category of the non-empty power set monad, which is (almost) **Rel**.

The proposition still holds when **D** is merely a Markov category itself!

# Classical de Finetti theorem

A sequence $(x_n)_{n\in\mathbb{N}}$ of random variables on a space $X$ is **exchangeable** if their distribution is invariant under finite permutations $\sigma$,

$$\begin{aligned}&\mathbb{P}\big[x_1 \in S_{\sigma(1)}, \dots, x_n \in S_{\sigma(n)}\big]\\ =\ &\mathbb{P}[x_1 \in S_1, \dots, x_n \in S_n]\,.\end{aligned}$$

**Theorem**

If $(x_n)$ is exchangeable, then there is a measure $\mu$ on $PX$ such that

$$\mathbb{P}[x_1 \in S_1, \dots, x_n \in S_n] = \int p(x_1 \in S_1) \cdots p(x_n \in S_n)\, \mu(dp).$$

Idea: sequence of tosses of a coin with unknown bias!

# Categories of comonoids

**Proposition**

Let $\mathbf{C}$ be any symmetric monoidal category. Then the category with:

- Commutative comonoids in $\mathbf{C}$ as objects,
- Counital maps as morphisms,
- The specified comultiplications as copy maps,

is a Markov category.

A good example is $\mathbf{Vect}_k^{\mathrm{op}}$ for a field $k$:

- The comonoids correspond to commutative $k$-algebras of $k$-valued random variables.
- We obtain **algebraic probability theory** with "random variable transformers" as morphisms (formal opposites of Markov kernels).

# Diagram categories and ergodic theory

**Proposition**

Let $\mathbf{D}$ be any category and $\mathbf{C}$ a Markov category. The category in which

- Objects are functors $\mathbf{D} \to \mathbf{C}_{\mathrm{det}}$,
- Morphisms are natural transformations with components in $\mathbf{C}$.

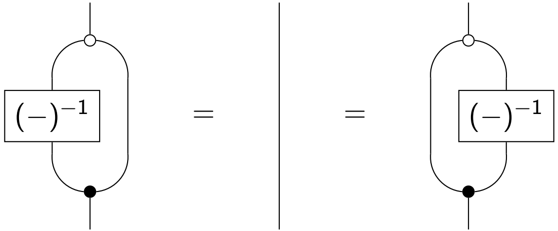
With the poset $\mathbf{D} = \mathbb{Z}$, we get a category of **discrete-time stochastic processes**.

This generalizes an observation going back to (Lawvere, 1962).

We can also take $\mathbf{D} = \mathbf{B}G$ for a group $G$, resulting in categories of dynamical systems with deterministic dynamics but stochastic morphisms.

# Hyperstructures: categorical algebra in Markov categories

A **group** $G$ is a monoid $G$ together with $(-)^{-1} : G \to G$ such that




This equation can be interpreted in any Markov category! (Together with the bialgebra law.)

- More generally, one can consider models of any algebraic theory in any Markov category.

- In Kleisli categories of probability-like monads, these are known as **hyperstructures**.

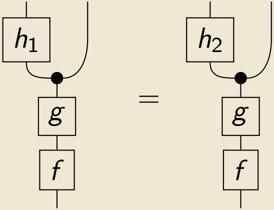
- Peter Arndt's suggestion:

  Develop categorical algebra for hyperstructures in terms of Markov categories!

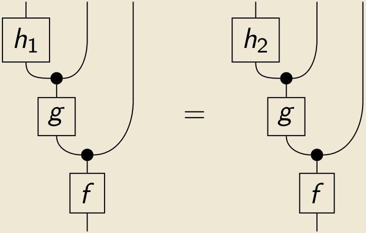
# The causality axiom

## Definition

**C** is **causal** if




implies



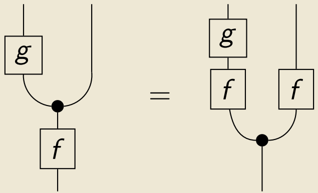

- **Intuition:** The choice between $h_1$ and $h_2$ in the “future” of $g$ does not influence the “past” of $g$.

- Not every Markov category is causal.

# The positivity axiom

## Definition

**C** is **positive** if whenever $gf$ is deterministic for composable $f$ and $g$, then also



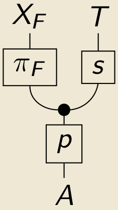

- **Intuition:** If a deterministic process has a random intermediate result, then that result can be computed independently from the process.
- Not every Markov category is positive.
- Dario Stein: every causal Markov category is positive!

### Theorem (Kolmogorov zero–one law)

Let $X_I$ be a Kolmogorov product of a family $(X_i)_{i \in I}$.

If

- $p : A \to X_I$ makes the $X_i$ independent and identically distributed, and
- $s : X_I \to T$ is such that




displays $X_F \perp T \,||\, A$ for every finite $F \subseteq I$,

then $ps$ is deterministic.

# The classical Hewitt–Savage zero-one law

**Theorem**

*Let* $(x_n)_{n\in\mathbb{N}}$ *be independent and identically distributed random variables, and* $S$ *any event depending only on the* $x_n$ *and invariant under finite permutations.*

*Then* $\mathrm{P}(S) \in \{0, 1\}$.

# The synthetic Hewitt–Savage zero-one law

**Theorem**

Let $J$ be an infinite set and $\mathbf{C}$ a causal Markov category. Suppose that:

- The Kolmogorov power $X^{\otimes J} := \lim_{F \subseteq J \text{ finite}} X^{\otimes F}$ exists.
- $p : A \to X^{\otimes J}$ displays the conditional independence $\perp_{i \in J} X_i \parallel A$.
- $s : X^J \to T$ is deterministic.
- For every finite permutation $\sigma : J \to J$, permuting the factors $\tilde{\sigma} : X^{\otimes J} \to X^{\otimes J}$ satisfies

$$\tilde{\sigma} p = p, \qquad s\tilde{\sigma} = s.$$

Then $sp$ is deterministic.

Proof is by string diagrams, but far from trivial!

# Why categorical probability?

In no particular order:

- Applications to probabilistic programming.
- Prove theorems in greater generality and with more intuitive proofs.
- Reverse mathematics: sort out interdependencies between theorems.
- Ultimately, prove theorems of higher complexity?
- Simpler teaching of probability theory. (String diagrams!)
- Different conceptual perspective on what probability is.

# Discrete probability theory as a Markov category

One of the paradigmatic Markov categories is **FinStoch**, the category of finite sets and **stochastic matrices**: a morphism $f : X \to Y$ is

$$(f(y|x))_{x \in X, y \in Y} \in \mathbb{R}^{X \times Y}$$

with

$$f(y|x) \geq 0, \qquad \sum_y f(y|x) = 1.$$

Composition is the **Chapman-Kolmogorov formula**,

$$(gf)(z|x) := \sum_y g(z|y)\, f(y|x).$$

A morphism $p : 1 \to X$ is a **probability distribution**.

A general morphism $X \to Y$ has many names: **Markov kernel**, probabilistic mapping, communication channel, ...

The monoidal structure implements **stochastic independence**,

$$(g \otimes f)(xy|ab) := g(x|a)\, f(y|b).$$

The copy maps are

$$\mathrm{copy}_X : X \longrightarrow X \times X, \qquad \mathrm{copy}_X(x_1, x_2|x) = \begin{cases} 1 & \text{if } x_1 = x_2 = x, \\ 0 & \text{otherwise.} \end{cases}$$

The deletion maps are the unique morphisms $X \to 1$.

- Works just the same with "probabilities" taking values in any **semiring** $R$.

- Taking $R$ to be the **Boolean semiring** $\mathbb{B} = \{0, 1\}$ with

$$1 + 1 = 1$$

  results in the Kleisli category of the nonempty finite powerset monad.

  $\Rightarrow$ We get a Markov category for non-determinism.

- Measure-theoretic probability: Kleisli category of the **Giry monad**.