Categorical Probability and the de Finetti theorem

Tobias Fritz

joint work with Tomáš Gonda, Paolo Perrone and Eigil Rischel

March 2021

Categorical Probability and the de Finetti theorem

now with proof!

Tobias Fritz

joint work with Tomáš Gonda, Paolo Perrone and Eigil Rischel

March 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.

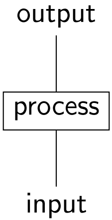
For a broader perspective, see the videos from the online workshop Categorical Probability and Statistics!

Why categorical probability?

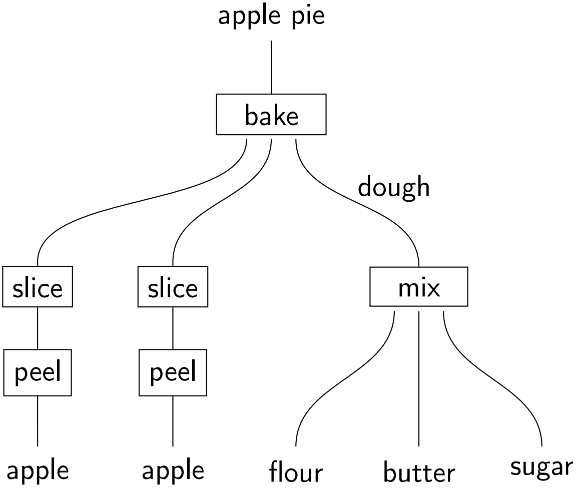
In no particular order:

- 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.

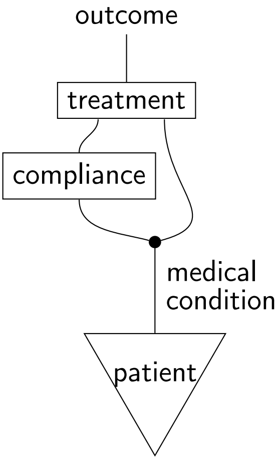
Many of us like to reason about processes in terms of string diagrams:



We can compose processes into networks:



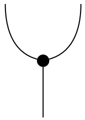
Suppose that we want to reason about **flow of information** in a medical trial. Then we seem to need diagrams like this:



→ Medical condition has an influence on **both** trial compliance and on treatment outcome!

Hence a theory of information flow needs additional pieces of structure:

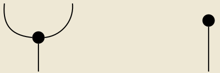
- ▷ **copying information**:



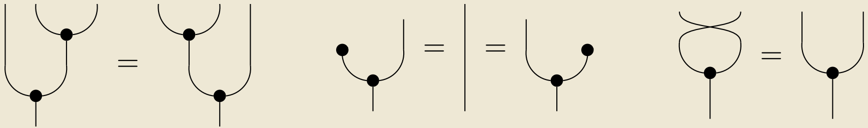
- ▷ **deleting information**:

Definition

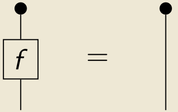
A **Markov category C** is a symmetric monoidal category supplied with **copying** and **deleting** operations on every object,



giving commutative comonoid structures



which interact well with the monoidal structure, and such that



(much can be done without)

A basic example

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**.

Outline

In the rest of this talk, I will sketch:

- How to develop (some) theorems of probability theory in terms of Markov categories.
- This includes work in progress on the **de Finetti theorem** characterizing permutation-invariant probability distributions.
- There is a vast landscape of Markov categories, going much beyond probability theory.

We're just at the beginning!

Analogy with topos theory

- Every constructive piece of mathematics holds in every topos, but there still are many toposes!

  Similarly, every theorem of probability proven in terms of Markov categories holds quite generally.

- There is a hierarchy of additional axioms of different strength.

Bold working hypothesis:

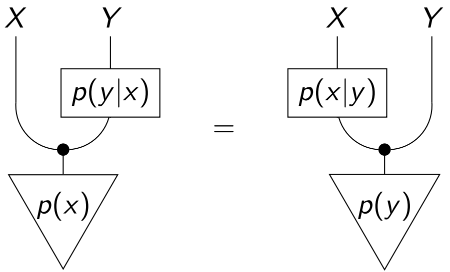
$$
\begin{aligned}
& \text{Theory of Markov categories} \\
\cong\ & \text{General theory of information flow} \\
\cong\ & \text{Generalized probability theory and statistics.}
\end{aligned}
$$

Indeed, for example **Bayesian networks** can be defined in Markov categories (Brendan Fong's 2013 MSc).

In other words, Markov categories are a general setting for talking about cause and effect.

A first theoretical development: Bayesian inversion

Bayes' rule takes the form:



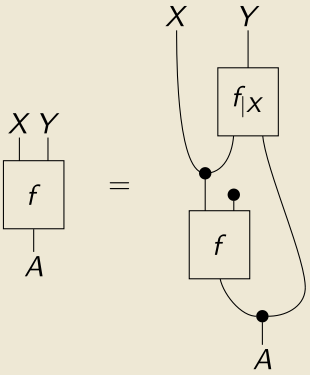
$$p(y|x)\,p(x) \quad = \quad p(x|y)\,p(y)$$

In any Markov category **with conditionals**, there is a sense in which the map $p(y|x) \mapsto p(x|y)$ is a **dagger functor**!

More generally:

Definition

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

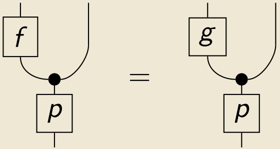


Almost sure equality

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

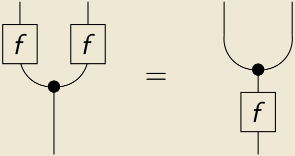


- **Intuition:** f and g behave the same on all inputs produced by p.
- Other concepts (besides equality) also relativize with respect to p-almost surely.

Determinism

Definition

In a Markov category, a morphism $f : X \to Y$ is **deterministic** if it commutes with copying,

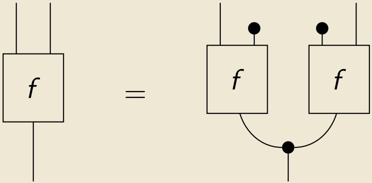


- **Intuition:** Applying f to copies of input = copying the output of f.
- The deterministic morphisms form a cartesian monoidal subcategory.

Conditional independence

Definition

In a Markov category, $f : A \to X \otimes Y$ **displays the conditional independence** $X \perp Y \parallel A$ if



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!

Markov categories as Kleisli categories

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

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

Then P turns out to extend to a commutative monad on $\mathbf{C}_{\mathrm{det}}$, and $\mathbf{C}$ is its Kleisli category!

The counit of the induced comonad is

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

the map that returns a random sample from a distribution.

Detour: random measures

- 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

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

- If μ is the distribution of your belief on the coin's bias, you would go for odds better than

$$\int p(\text{heads})\, p(\text{tails})^2\, \mu(dp).$$

So p is a **random measure** with distribution μ.

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 σ,

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

Theorem

If (x_n) is exchangeable, then there is a measure μ on PX such that

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

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

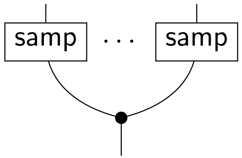
Synthetic de Finetti theorem

Assumption: **C** is an a.s. compatibly representable Markov category with conditionals and countable Kolmogorov products.

Definition

$f : A \to X^{\mathbb{N}}$ is **exchangeable** if it is invariant under composing with finite permutations.

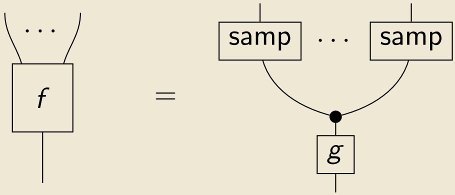
Sampling $\mathbb{N}$ times gives a morphism $PX \to X^{\mathbb{N}}$ given by



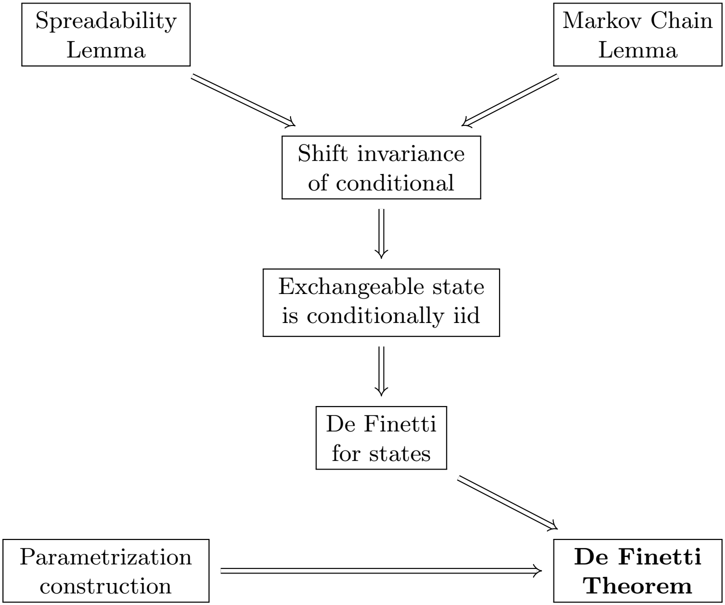
Synthetic de Finetti theorem

Theorem

For every exchangeable $f : A \to X^{\mathbb{N}}$ there is $g : A \to PX$ such that



Structure of proof



Categories of comonoids

Proposition

Let **C** be any symmetric monoidal category. Then the category with:

- Commutative comonoids in **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 $\mathbf{C}$.

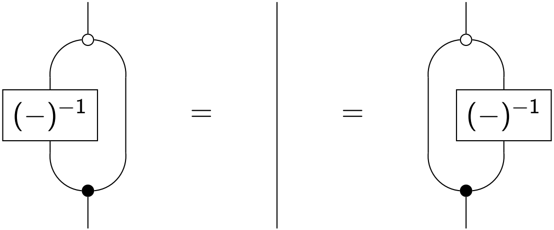
With the poset $\mathbf{D} = \mathbb{Z}$, we get a category of **discrete-time stochastic processes**.

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!

Summary

- Markov categories are an emerging formalism providing a general theory of information theory.
- Many qualitative results of probability theory generalize to Markov categories.
- These usually require additional axioms (of various degrees of strength).
- There is a vast unexplored landscape of Markov categories in which these results can be instantiated.
- This is similar to topos theory: a lot of mathematics can be developed constructively and then instantiated in an unexpectedly large number of contexts.

Some further directions

- Is there a "most convenient" Markov category **C** for measure-theoretic probability?

  Some desiderata:

  - **C** has conditionals.
  - **C** has Kolmogorov products.
  - **C** has supports.
  - $\mathbf{C}_{\mathrm{det}}$ is cartesian closed.

  I don't know of *any* non-cartesian Markov category with these properties!

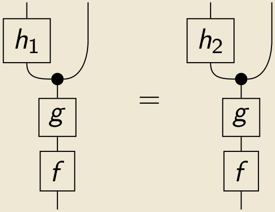
- Many results in probability theory are quantitative.

  $\Rightarrow$ **Do we need enriched Markov categories?**

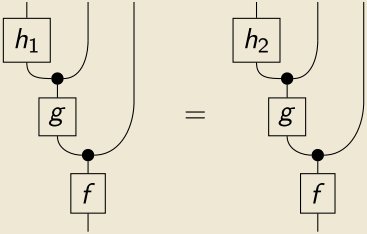
The causality axiom

Definition

C is **causal** if



implies

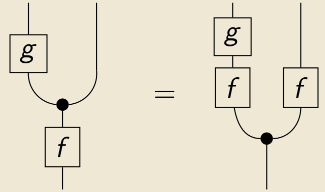


- **Intuition:** The choice between h_1 and h_2 in the "future" of g does not influence the "past" of g.
- Not every Markov category is causal.

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!

Definition

Let $(X_i)_{i \in I}$ be a family of objects. The **infinite tensor product**

$$X_I := \bigotimes_{i \in I} X_i$$

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

Definition

An infinite tensor product X_I is a **Kolmogorov product** if the limit projections $\pi_F : X_I \to X_F$ are deterministic.

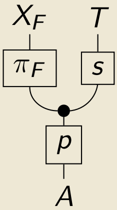
- This additional condition fixes the comonoid structure on X_I.

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 $\boldsymbol{P}(S) \in \{0, 1\}$.

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!