Markov categories

Probability and Statistics as a Theory of Information Flow

Tobias Fritz

November 2020

References

- Peter V. Golubtsov,
  Axiomatic description of categories of information converters. *Problemy Peredachi Informatsii* 35(3), 80–98 (1999).
  (And other similar papers by Golubtsov.)

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

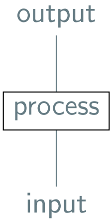
- Tobias Fritz,
  A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Adv. Math.* 370, 107239 (2020).
  arXiv:1908.07021.

- 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.

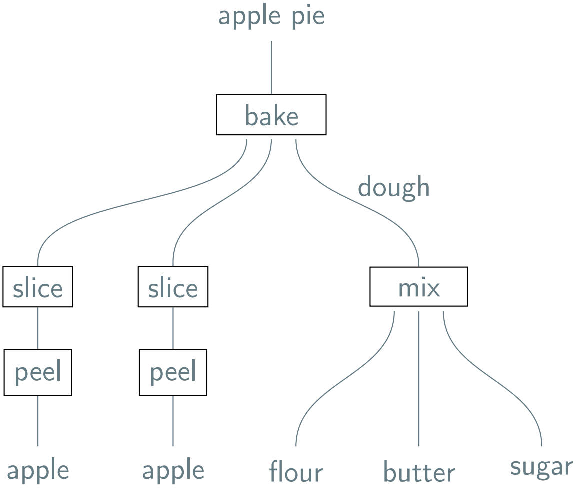
- Evan Patterson,
  The algebra and machine representation of statistical models, PhD thesis.
  arXiv:2006.08945.

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

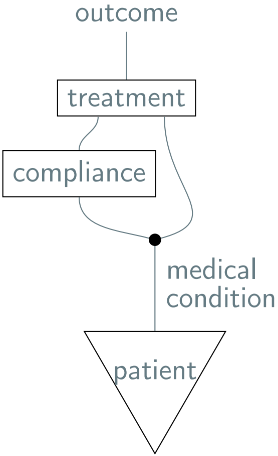
Many of us like to reason about processes in terms of string diagrams:



We can compose processes into networks:



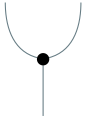
Suppose that we want to reason about **flow of information** in a medical trial. Then we seem to need diagrams like this:



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

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

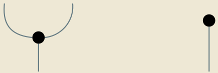
- **copying information**:



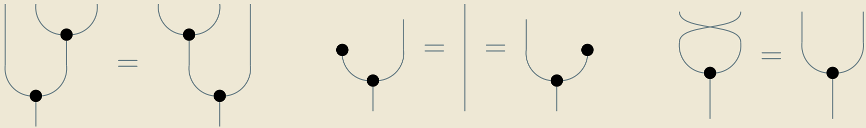
- **deleting information**:

Definition

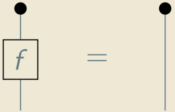
A **Markov category C** is a symmetric monoidal category supplied with **copying** and **deleting** operations on every object,



giving commutative comonoid structures



which interact well with the monoidal structure, and such that



(do we really want this?)

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

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

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

Outline

In the rest of this talk, I will sketch:

- How to develop (some) theorems of probability theory in terms of Markov categories. (In some cases, turning theorems into definitions.)
- There is a vast landscape of Markov categories, going much beyond probability theory.

In both respects, 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!

Also as in topos theory: there is a hierarchy of additional axioms of different strength.

Bold working hypothesis:

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

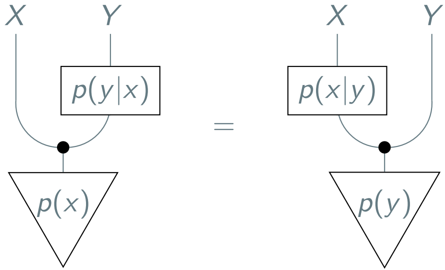
Indeed, for example **Bayesian networks** can be defined in Markov categories (Brendan Fong's MSc).

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

$\Rightarrow$ Somebody should use this to generalize **causal inference**!

A first theoretical development: Bayesian inversion

Bayes' rule takes the form:



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

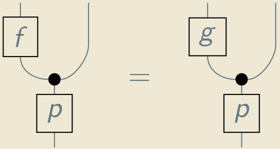
In any Markov category **with conditionals**, there is a sense in which the map $p(y|x) \mapsto p(x|y)$ is a **dagger functor**!

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

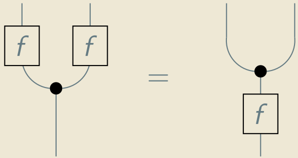


- **Intuition:** f and g behave the same on all inputs produced by p.
- Other concepts (besides equality) also relativize with respect to p-almost surely.

Determinism

Definition

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

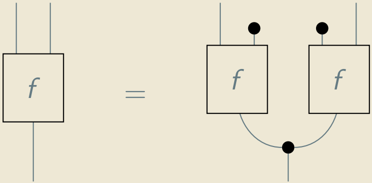


- **Intuition:** Applying f to copies of input = copying the output of f.
- The deterministic morphisms form a cartesian monoidal subcategory.

Conditional independence

Definition

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



A piece of probability theory

One of the fundamental theorems of probability is the **law of large numbers**:

$$\mathbf{P}\left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}[X]\right] = 1. \qquad (*)$$

I don't yet know how to state and prove this in terms of Markov categories. But we have proven a closely related classical result synthetically.

Hewitt–Savage zero-one law

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

Then $\mathbf{P}(A) \in \{0, 1\}$.

This implies that $(*)$ is 0 or 1, but we don't know which!

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!

Kleisli categories are Markov categories

Proposition

Let

- $\mathbf{C}$ be a category with finite products,
- P a commutative monad on $\mathbf{C}$ with $P(1) \cong 1$.

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

Examples:

- The Kleisli category of the **Giry monad** is a Markov category.
- Kleisli categories of other monads that capture measure-theoretic probability.
- The Kleisli category of the non-empty power set monad, which is (almost) **Rel**.

The proposition still holds when $\mathbf{C}$ is merely a Markov category itself!

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

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 all the way 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.

A Markov category for information theory?

There are well-known analogies between probability and information theory:

- Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- Conditional entropy: $H(A|B) = H(AB) - H(B)$.

Question

Is there a Markov category for information theory explaining these analogies?

Maybe like this:

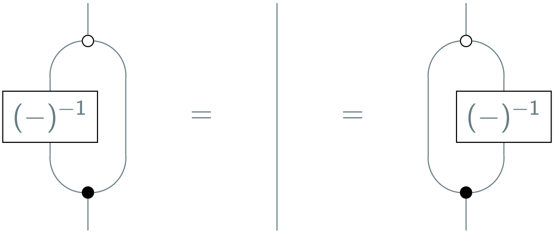
- Objects are finite sets,
- Morphisms $f : X \to Y$ are compatible families of stochastic maps

$$(f_n : X^{\times n} \to Y^{\times n})_{n \in \mathbb{N}}$$

modulo some suitable asymptotic equivalence as $n \to \infty$.

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**.

$\Rightarrow$ Peter Arndt's suggestion: Develop categorical algebra for hyperstructures!

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.

  $\Rightarrow$ **We need enriched Markov categories**!