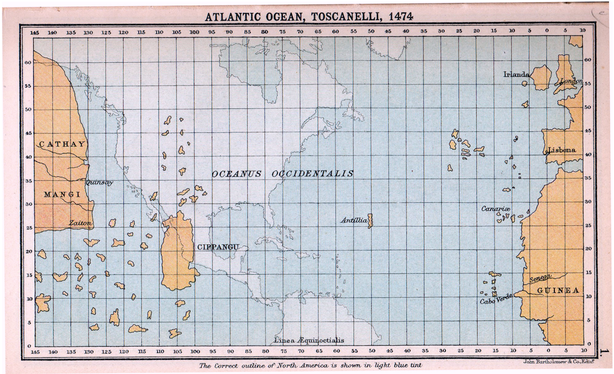
Categorical Probability: Results and Challenges

Tobias Fritz

May 2019

What this talk is (not)



Categorical probability is like finding the sea route to India:

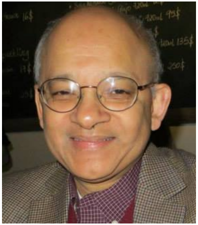
- Many possible routes to be explored without a coherent overall map.
- We may end up discovering something totally different than India!

A (not so) random sample of contributors

Bill Lawvere

?

Michèle Giry



Prakash Panangaden

Bart Jacobs

Paolo Perrone

Sharwin Rezagholi

David Spivak

Motivation

- Category theory has been hugely successful in algebraic geometry, algebraic topology, and theoretical computer science.
- Contemporary research in these fields can hardly even be conceived of without categorical machinery.
- Can and should we expect similar success in other areas?
- A case in point: **probability theory**!

Motivation

A structural treatment can help us achieve:

- Improved conceptual clarity.
- Greater generality due to higher abstraction.
- Therefore applicability in a range of contexts instead of only one.

For example, let $\mathtt{Sh}(\mathrm{I}\mathbb{R})$ be the category of sheaves on the poset of compact intervals in $\mathbb{R}$.

Conjecture (with David Spivak)

A probability space internal to $\mathtt{Sh}(\mathrm{I}\mathbb{R})$ is the same thing as an external stochastic process.

Suitably structural results on probability would therefore immediately give results on stochastic processes.

But first, **what is probability theory**?

- The study of randomness.
- Fundamental insight: probability is volume! $\Rightarrow$ **Measure theory**.
- Central themes:
  - Random variables and their distributions.
  - Theorems involving infinitely many variables.
  - ~~Conditioning and Bayes' rule.~~

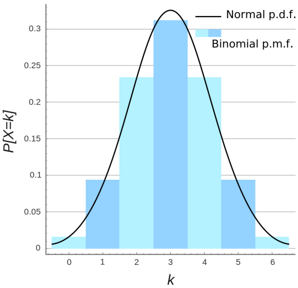
An example statement:

Central limit theorem

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. random variables with $\mathbf{E}[X_n] = \mu$ and $\mathrm{V}[X_n] = \sigma$. Then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \quad \xrightarrow{n\to\infty} \quad N(0,\sigma).$$

converges in distribution.



(Wikipedia, Cflm001)

Structures in categorical probability

Probability monad:

- probability measures
- pushforward of measures
- point measures δ_x
- averaging of measures

Eilenberg–Moore category:

- integration
- stochastic dominance
- martingales

Kleisli category:

- stochastic maps
- (conditional) independence
- statistics

- ▷ A probability monad lives on a category of sets or spaces.
- ▷ Most basic: the **convex combinations monad** on `Set`, where

$$DX := \left\{ \sum_i c_i \delta_{x_i} \;\middle|\; c_i \geq 0, \sum_i c_i = 1 \right\}$$

is the set of finitely supported probability measures on X.

- ▷ $p \in DX$ is a "random element" of X. For example a fair coin,

$$\frac{1}{2}\delta_{\mathrm{heads}} + \frac{1}{2}\delta_{\mathrm{tails}} \quad \in \quad D\left(\{\mathrm{heads}, \mathrm{tails}\}\right)$$

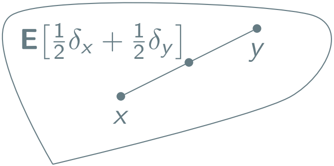
- ▷ Functoriality $Df : DX \to DY$ takes **pushforward measures**: applying a function to a random element of X produces a random element of Y.

- ▷ The unit $X \to DX$ assigns to every $x \in X$ the point mass δ_x at x.

- ▷ The multiplication $DDX \to DX$ computes the expected distribution,

$$\sum_i c_i \left(\sum_j d_{ij} \delta_{x_{ij}} \right) \longmapsto \sum_{i,j} c_i d_{ij} \delta_{x_{ij}}$$

- ▷ Algebras $\mathbf{E} : DA \to A$ are “convex spaces” in which every $p \in DA$ has a designated **barycenter** or **expectation value** $\mathbf{E}[p] \in A$.



Integration: the Eilenberg–Moore side

- Let A be an Eilenberg–Moore algebra, e.g. $A = \mathbb{R}$.
- Then for $p \in DX$ and a **random variable** $f : X \to A$,

$$\int_X f \, dp := \mathbf{E}[(Df)(p)].$$

- For $g : Y \to X$ and $q \in DY$, the **change of variables** formula

$$\int_Y (f \circ g) \, dq = \int_X f \, d(Dg)(q)$$

then holds by functoriality, $D(f \circ g) = D(f) \circ D(g)$.

Measure theory without measure theory

Basic idea

A probability measure on X is an idealized version of a **finite sample**: elements $(x_1, \ldots, x_n)$ of X representing the uniform distribution $\frac{1}{n}\sum_i \delta_{x_i}$.

All constructions and proofs with probability measures should be reducible to constructions and proofs with finite samples.

We construct a probability monad which implements this idea and makes it precise.

Let $\mathtt{CMet}$ be the category where

- objects (X, d_X) are complete metric spaces,
- morphisms $f : (X, d_X) \to (Y, d_Y)$ are **short maps**,

$$d_Y(f(x), f(x')) \leq d_X(x, x').$$

- For $S \in \mathtt{FinSet}$, we have the **power functor**

$$\mathtt{CMet} \longrightarrow \mathtt{CMet}, \qquad X \longmapsto X^S.$$

- We have isomorphisms $X^1 \cong X$ and $X^{S \times T} \cong (X^S)^T$.

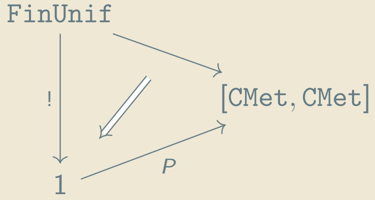
- These make the power functors into a **graded monad** on $\mathtt{CMet}$, which is a lax monoidal functor

$$\mathtt{FinUnif} \longrightarrow [\mathtt{CMet}, \mathtt{CMet}].$$

- Here, $\mathtt{FinUnif} \subseteq \mathtt{FinSet}$ is the subcategory of nonempty sets and functions with uniform fibres.

Theorem (with Paolo Perrone, arXiv:1712.05363)

There is a left Kan extension



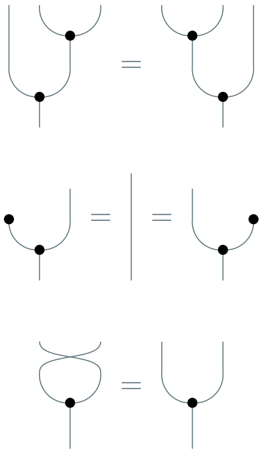
in the 2-category of symmetric monoidal categories and lax monoidal functors, where P is a probability monad such that

$$PX = \{\text{Radon measures on } X \text{ with finite first moment}\}.$$

This reduces (parts of) measure and probability to combinatorics!

Categories of stochastic maps: the Kleisli side

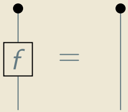
Let C be a symmetric strict monoidal category where each object carries a distinguished **commutative comonoid**:



We think of this structure as providing **copy** and **delete** operations.

Definition

$\mathtt{C}$ is a **category with comonoids** if these comonoids are compatible with the monoidal structure, and deletion is natural,



This makes $\mathtt{C}$ into a **semicartesian** monoidal category: we have natural maps

$$X \otimes Y \longrightarrow X, \qquad X \otimes Y \longrightarrow Y$$

which are abstract versions of **marginalization**, when composed with $p : I \to X \otimes Y$.

Example

Let `FinStoch` be the category of finite sets, where morphisms $f : X \to Y$ are **stochastic matrices** $(f_{xy})_{x \in X,\, y \in Y}$,

$$f_{xy} \geq 0, \qquad \sum_y f_{xy} = 1,$$

- f_{xy} is the probability that the output is y given the input x.
- We also write $f(y|x)$.
- Composition of morphisms is given by the Chapman–Kolomogorov equation,

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

- The monoidal structure is

$$(g \otimes f)(y, z|w, x) := g(y|w)f(z|x),$$

with canonical symmetry isomorphism.

- The copying operation is just copying,

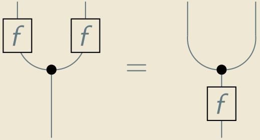
$$\delta(x_1, x_2|x) = \begin{cases} 1 & \text{if } x_1 = x_2 = x, \\ 0 & \text{otherwise.} \end{cases}$$

- With this, `FinStoch` is a category with comonoids.

Deterministic morphisms

Definition

A morphism $f : X \to Y$ is **deterministic** if the comonoids are natural with respect to f,



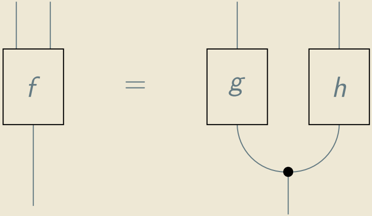
- ▷ The deterministic morphisms form a cartesian monoidal subcategory.
- ▷ In $\texttt{FinStoch}$, the deterministic morphisms are the stochastic matrices with entries in $\{0, 1\}$, i.e. the actual functions. They form a copy of $\texttt{FinSet}$.

Conditional independence

Categories with comonoids support several notions of conditional independence, including:

Definition

A morphism $f : A \to X \otimes Y$ **displays the conditional independence** $X \perp Y \,||\, A$ if there are $g : A \to X$ and $h : A \to Y$ such that

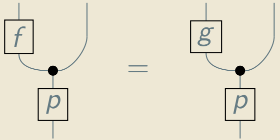


One can derive the usual properties of conditional independence purely formally.

Almost surely

Definition

Given $p : \Theta \to X$, morphisms $f, g : X \to Y$ are **equal p-almost surely** if



- Other concepts relativize similarly to almost surely concepts.

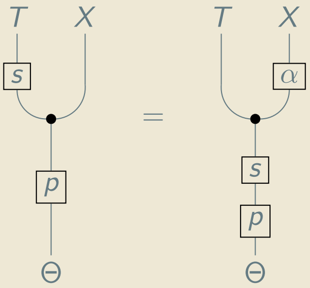
Proposition

If $gf = \mathrm{id}$, then g is f-almost surely deterministic.

Sufficient statistics

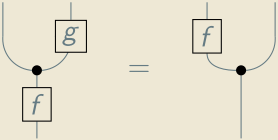
Definition

- A **statistical model** is a morphism $p : \Theta \to X$.
- A **statistic** for p is a deterministic split epimorphism $s : X \to T$.
- A statistic is **sufficient** if there is a splitting $\alpha : T \to X$ such that



Axiom

Suppose that $gf = \mathrm{id}$. Then



- This holds in `FinStoch`.
- Now there is a completely formal version of a classical result of statistics:

Fisher–Neyman factorization theorem (preliminary)

If the axiom holds, a statistic $s : X \to T$ is sufficient for $p : \Theta \to X$ if and only if there is a splitting $\alpha : T \to X$ with $\alpha s p = p$.

Other preliminary results

Let $p : \Theta \to X$ be a statistical model.

We have abstract versions of other classical theorems of statistics:

Basu's theorem

A complete sufficient statistic for p is independent of any ancillary statistic.

Bahadur's theorem

If a minimal sufficient statistic exists, then a complete sufficient statistic is minimal sufficient.

A challenge: zero-one laws

Kolmogorov's and Hewitt–Savage's zero-one law

Let

- $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables,
- A an event which is a function of the (X_n), and
- independent of $(X_n)_{n \in F}$ for any finite $F \subseteq \mathbb{N}$ (Kolmogorov), **or**
- invariant under finite permutation of the (X_n) (Hewitt–Savage).

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

- A categorical reformulation and proof in a suitable class of categories with colimits may now be within reach.

A challenge: concentration of measure

Concentration of measure is the phenomenon that

- ▷ if A is a set with $p(A) \geq 1/2$ in a metric probability space,
- ▷ then the ε-neighbourhood A_ε satisfies $p(A_\varepsilon) \approx 1$.

Theorem (Lévy)

On the n-sphere S^n,

$$p(A) \geq 1 - \sqrt{\frac{\pi}{8}}\, e^{-\frac{\varepsilon^2 n}{2}} \approx 1.$$

Law of large numbers

Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence with $\mathbf{E}[X_n] = \mu$. Then

$$\lim_{n\to\infty} \mathbf{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right] = 0.$$

Summary

- Categorical probability is currently like finding the sea route to India: several approaches with unclear relation.

- This talk has sketched a biased sample of approaches.

- It seems useful to distinguish:
  - Eilenberg–Moore category $\Rightarrow$ integration and its properties.
  - Kleisli category $\Rightarrow$ conditional independence, statistics.

- A clearer overall picture may emerge once we have further concrete results.

- The biggest challenge is to recover the specific analytical theorems of probability, such as the central limit theorem.