The law of large numbers in categorical probability

Tobias Fritz

based on work with Tomáš Gonda, Paolo Perrone and Eigil Rischel

May 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,
  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!

The law of large numbers

Theorem

Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real-valued independent random variables with identical distribution and $\mathbb{E}[|x_1|] < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}[x_1]$$

with probability 1.

- **Example:** Upon repeatedly tossing a fair coin, the relative frequency of heads approaches $\frac{1}{2}$ with probability 1.

- But where are the categories?

Teaser

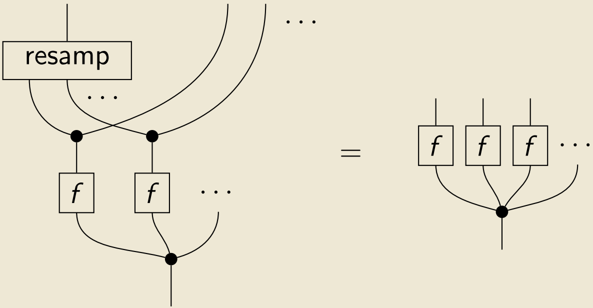
My goal is to explain this form of the law of large numbers:

Theorem/Definition

For every object X there is a partial morphism

$$\text{resamp} : X^{\mathbb{N}} \to X$$

such that for every $f : A \to X$,



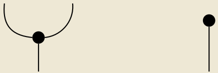
The big picture

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

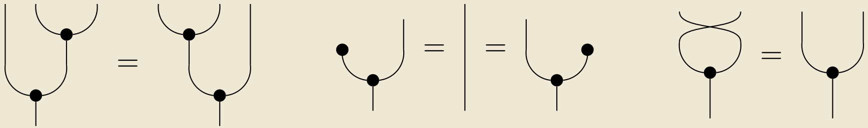
- There will be no numerical probabilities!
- I can say more about the motivations and scope of categorical probability if there is need for discussion.

Definition

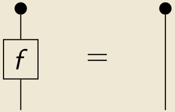
A **Markov category** is a symmetric monoidal category supplied with **copying** and **deleting** operations on every object,



giving commutative comonoid structures



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



Semantics

There are many different (and interesting) Markov categories.

But for today, I have one particular intended semantics in mind:

Definition

BorelStoch is the category with:

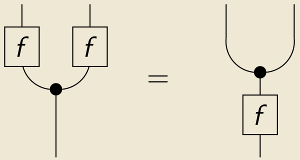
- **Standard Borel spaces** as objects (finite sets, $\mathbb{N}$ and $[0, 1]$).
- Measurable **Markov kernels** as morphisms.
- Products of measurable spaces for $\otimes$.

BorelStoch encodes standard measure-theoretic probability.

Determinism

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.
- Deterministic morphisms form a cartesian monoidal subcategory $\mathbf{C}_{\mathrm{det}}$.
- $\mathbf{BorelStoch}_{\mathrm{det}}$ is the category of measurable functions between standard Borel spaces.

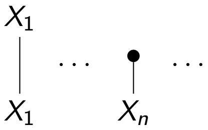
Infinite tensor products

Let $(X_n)_{n\in\mathbb{N}}$ be a family of objects.

For finite $F \subseteq F' \subseteq \mathbb{N}$, we have projection morphisms

$$\bigotimes_{n\in F'} X_n \longrightarrow \bigotimes_{n\in F} X_n$$

given by composing with deletion for all $n \in F' \setminus F$, like this:



Infinite tensor products

Definition

The **infinite tensor product**

$$X^{\mathbb{N}} = \bigotimes_{n \in \mathbb{N}} X_n$$

is the limit of the finite tensor products $X^F := \bigotimes_{n \in F} X_n$ 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

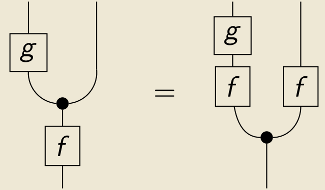
An infinite tensor product $X^{\mathbb{N}}$ is a **Kolmogorov product** if the limit projections $\pi^F : X^{\mathbb{N}} \to X^F$ are deterministic.

- This additional condition fixes the comonoid structure on $X^{\mathbb{N}}$.
- From now on: assume Markov category with countable Kolmogorov products.
- Satisfied by **BorelStoch** (Kolmogorov extension theorem).

Positivity

Definition

C is **positive** if the following holds: if a composite gf is deterministic, then also



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

Partial morphisms

- In the law of large numbers, the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i$$

does not always exist.

- This suggests the need for **partial morphisms** in categorical probability.

- Under certain "partializability" conditions we indeed get a **monoidal restriction category**.

Partial morphisms

Definition

A positive Markov category is **partializable** if deterministic monos are closed under

- pullbacks,
- tensor products.

- In **BorelStoch**, the deterministic subobjects of X are the **measurable sets** $S \subseteq X$.
- This is a deep fact of descriptive set theory!
- Thanks to it, one can show that **BorelStoch** is partializable.

Resampling

- In statistics, **resampling** is a set of methods to estimate generalization (e.g. cross-validation).
- Here, I mean something different but closely related:

Resampling	=	Pick an element from an infinite sequence uniformly at random

- This doesn't make literal sense since **there is no uniform distribution on** $\mathbb{N}$.
- But it can be made to work for *some* sequences (x_n).
- Thus we get a partial morphism

$$\mathrm{resamp} : X^{\mathbb{N}} \to X.$$

Resampling

- In **BorelStoch**, we would want to define

$$\mathrm{resamp}(S \mid (x_n)) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1_S(x_i)$$

  whenever this limit exists for all measurable S.

- For finite n, this indeed corresponds to choosing an element from a finite sequence uniformly at random.

- The problem is that the resulting

$$\mathrm{resamp}(- \mid (x_n))$$

  would not always be a probability measure.

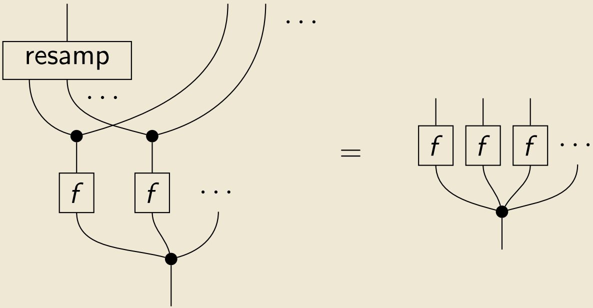
- Can be fixed by imposing **uniform** existence of the limits.

Back to the law of large numbers

We can now understand the following:

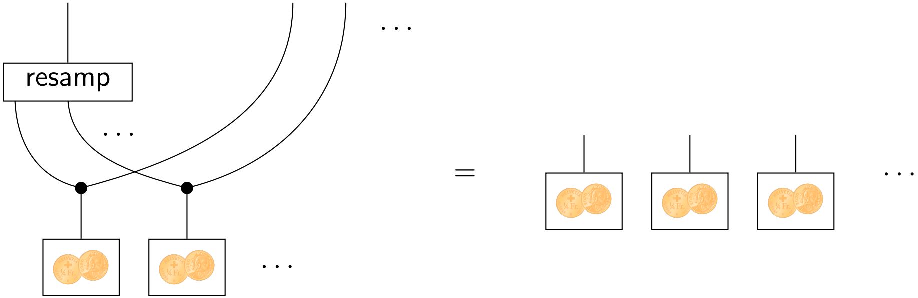
Theorem

In **BorelStoch**, every $f : A \to X$ satisfies



This statement encodes the **Glivenko–Cantelli theorem** on the convergence of the empirical distribution, a strong law of large numbers.

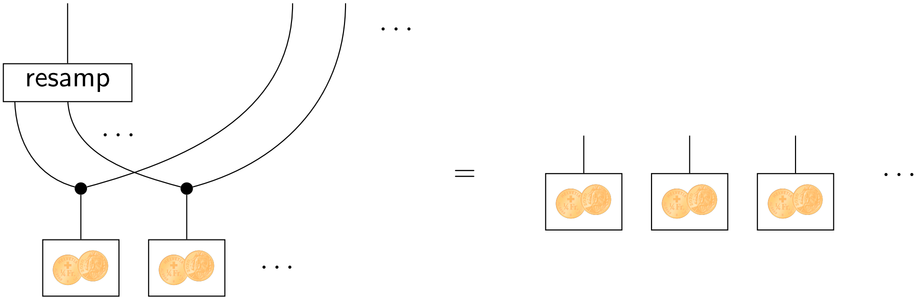
Coin example



- **Interpretation**: flipping infinitely many fair coins and then picking a random one makes the latter
  - independent of, and
  - identically distributed as

  the others.

Coin example



▷ In terms of random outcomes $c_i \in \{$, $\}$, this equation says

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1$$$$(c_i) =_{\mathrm{a.s.}} \mathbb{P}[$$$$] = \frac{1}{2},$$

an instance of the law of large numbers.

Summary

- Markov categories = emerging framework for synthetic probability.
- We have abstract versions of some 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 theorem on exchangeable distributions.
- We should try to add the law of large numbers to this list!
- There are further hints of connections with ergodic theory.

Bonus slides: 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

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

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

- 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.

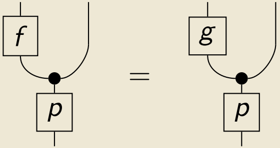
- Measure-theoretic probability: Kleisli category of the **Giry monad**.

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

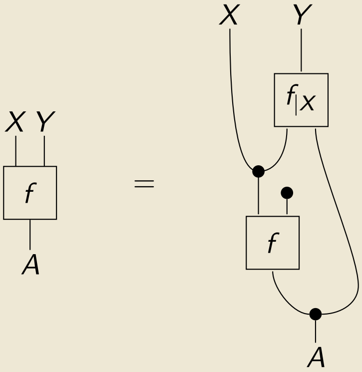


- **Intuition:** f and g behave the same on all inputs produced by p.
- In **BorelStoch**, coincides with the standard notion of a.s. equality.
- Other concepts relativize similarly with respect to p-almost surely.

Conditionals

Definition

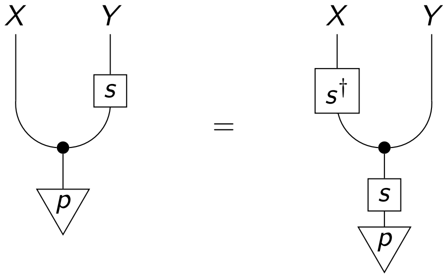
A Markov category **has conditionals** if for every $f : A \to X \otimes Y$ there is $f_{|X} : X \otimes A \to Y$ with



▷ **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:

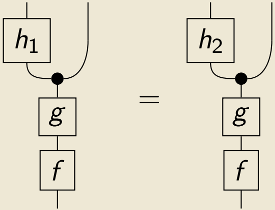


The Bayesian adjoint $s^\dagger$ depends on p.

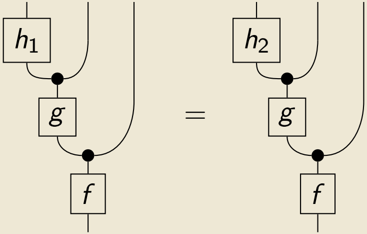
The causality axiom

Definition

C is **causal** if



implies

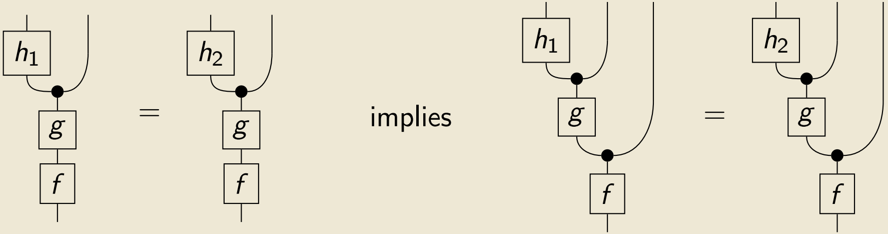


- **Intuition:** The choice between h_1 and h_2 in the "future" of g does not influence the "past" of g.
- Not every Markov category is causal.

The causality axiom

Definition

C is **causal** if



- **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.

Representability

Definition

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

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

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.

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!

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 $\mathbf{C}$.

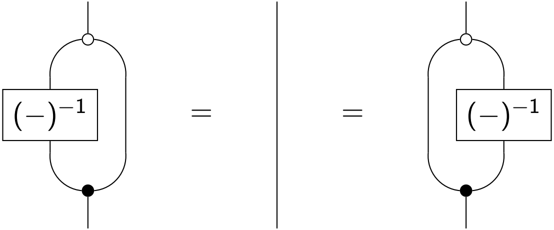
With the poset $\mathbf{D} = \mathbb{Z}$, we get a category of **discrete-time stochastic processes**.

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!

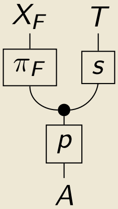
The synthetic Kolmogorov zero–one law

Theorem

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!

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

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

- Your bets satisfy **permutation invariance**. Can we say more?

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,

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

Theorem

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

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

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

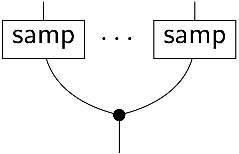
The de Finetti theorem

Assumption: All three axioms above hold. (True for **BorelStoch**.)

Definition

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

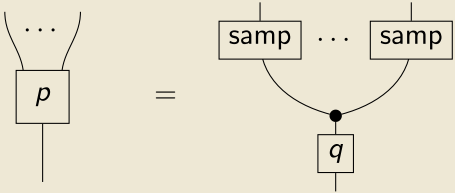
Sampling $\mathbb{N}$ times gives a morphism $PX \to X^{\mathbb{N}}$ given by



The de Finetti theorem

Theorem

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



- **Intuition:** The probabilities associated to your bets arise from **sampling from a random distribution**.