Introduction to Markov Categories

Eigil Fjeldgren Rischel

University of Copenhagen

Categorical Probabiliy and Statistics, June 2020

TLDR

- Consider a category where the maps are “stochastic functions”, or “parameterized probability distributions”.
- This is a symmetric monoidal category
- Many important notions in probability/statistics are expressible as diagram equations in this category.
- We can axiomatize the structure of this category to do “synthetic probability”.
- Several theorems admit proofs in this purely synthetic setting.

Overview of talk

Introduction

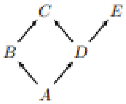
Diagrams for probability

Markov categories

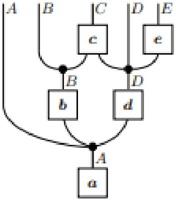
Kolmogorov's 0 to 1 law

Sufficient statistics

A graphical model



$$P(ABCDE) = P(A)P(B|A)P(D|A)P(C|BD)P(E|D)$$

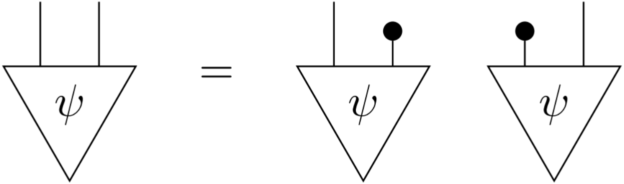


(Figure stolen from Kissinger-Jacobs-Zanasi: Causal Inference by String Diagram Surgery)

Independence

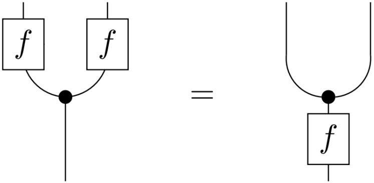
A map $I \to X \otimes Y$ is a "joint distribution". When are the two variables "independent"?

- If the distribution is the product of the marginals.
- If you can generate X and Y separately and get the same result.



Deterministic

What does it mean that $f : X \to Y$ is deterministic? "If you run it twice with the same input, you get the same output".



Markov categories

A *Markov category* (Fritz 2019) is a category with the structure to interpret these examples: a symmetric monoidal category with a terminal unit and a choice of comonoid on every object.

(These have been considered by several different authors)

Examples of Markov categories

- Stoch: measurable spaces and Markov kernels.
- FinStoch: finite sets and stochastic matrices.
- BorelStoch: *Standard Borel spaces* and Markov kernels.
- Gauss: Finite-dimensional real vector spaces and stochastic processes of the form "an affine map + Gaussian noise".
- SetMulti: Sets and *multivalued functions*.
- More exotic examples.

Kolmogorov's 0 to 1 law (classical)

Theorem(Kolmogorov)

Let $X_1, X_2 \ldots$ be an infinite family of independent random variables. Suppose $A \in \sigma(X_1, \ldots)$ (A is an event which depends "measurably" on these variables), and A is independent of any finite subset of the X_ns. Then $P(A) \in \{0, 1\}$.

Example: A is the event "the sequence X_i converges". The theorem says either the sequence converges almost surely, or it diverges almost surely.

Digression: Infinite tensor products

An "infinite tensor product" $X_{\mathbb{N}} := \bigotimes_{n\in\mathbb{N}} X_n$ is the cofiltered limit of the finite tensor products $\left(X_F := \bigotimes_{n\in F} X_n\right)_{F\subset\mathbb{N} \text{ finite}}$ if this limit exists and is preserved by tensor products $- \otimes Y$

An infinite tensor product is called a *Kolmogorov product* if all the projections to finite tensor products $\pi_F : X_{\mathbb{N}} \to X_F$ are deterministic.

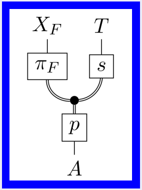
(This somewhat technical condition is necessary to fix the comonoid structure on $X_{\mathbb{N}}$)

Kolmogorov's 0 to 1 law (abstract)

With a suitable definition of infinite tensor products, we can prove:

Theorem(Fritz-R)

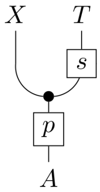
Let $p : A \to \bigotimes_{i\in\mathbb{N}} X_n$ and $s : \bigotimes_{i\in\mathbb{N}} X_i \to T$ be maps, with s deterministic and p presenting the indepenence of all the Xs. Suppose in each diagram



$\bigotimes_{i\in F} X_i$ is independent of T. Then $sp : A \to T$ is deterministic.

Applying this theorem to BorelStoch recovers the classical statement.

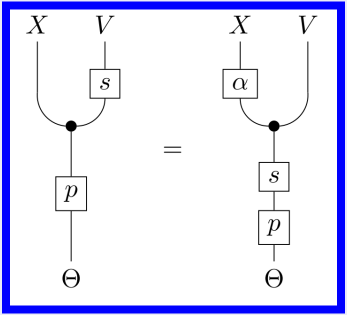
Proof(sketch)



- First, we see that T is independent of the whole infinite product $X_{\mathbb{N}}$ as well.
- This statement means that two maps $A \to X_{\mathbb{N}} \otimes T$ agree.
- By assumption the codomain is a limit, so it suffices to check that all the projections $A \to X_{\mathbb{N}} \otimes T \to X_F \otimes T$ agree.
- This is true by assumption.
- A diagram manipulation now shows that T, being both independent of $X_{\mathbb{N}}$ and a deterministic function of it, is a deterministic function of A.

Sufficient statistics

- A "statistical model" is simply a map $p : \Theta \to X$.
- A "statistic" is a deterministic map $s : X \to V$.
- A statistic is *sufficient* if $X \perp \Theta | V$ That means that we have α such that



Fisher-Neyman

Classically: Suppose we are in "a nice situation" (measures with density...)

Fisher-Neyman Theorem

A statistic $s(x)$ is sufficient if and only if the density $p_\theta(x)$ factors as $h(x)f_\theta(s(x))$

Abstract version: Suppose we are in "a nice Markov category". Then:

Abstract Fisher-Neyman (Fritz)

s is sufficient iff there is $\alpha : V \to X$ so that $\alpha s p = p$, and so that $s\alpha = 1_V$ almost surely.

Thank you for listening!

Some papers mentioned:

- Fritz(2019): A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics arxiv:1908.07021.
- Fritz-R(2020): Infinite products and zero-one laws in categorical probability arxiv:1912.02769
- Jacobs-Kissinger-Zanasi(2018): Causal inference by String Diagram Surgery arxiv:1811.08338