Categorical Probability Theory

Kaue de Mello Nogueira Piza

Thesis presented to the
Institute of Mathematics and Statistics
of the University of São Paulo
in partial fulfillment
of the requirements
for the degree of
Master of Science

Program: Matemática
Advisor: Prof. Dr. Hugo Luiz Mariano

This study was financed in part by the CNPq

São Paulo
October, 2023

Categorical Probability Theory

Kaue de Mello Nogueira Piza

This is the original version of the
thesis prepared by candidate Kaue
de Mello Nogueira Piza, as submitted
to the Examining Committee.

Acknowledgements

Thanks to my advisor, Hugo Luiz Mariano, for the guidance and kind patience.

Thanks to my familly and friends, who suported me in the period of elaboration of this work.

Thanks to the CNPq (National Council for Scientific and Technological Development) for the funding in the first years of this work.

Resumo

Kaue de Mello Nogueira Piza. **Teoria Categorial das Probabilidades**. Dissertação (Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2023.

Este trabalho tem como objetivo apresentar alguns dos principais conceitos da abordagem categorial à teoria das probabilidades. Nele são apresentados alguns capítulos com conceitos que são requisitos para um melhor entendimento dos resultados finais (abordagem da teoria das probabilidades via teoria da medida, teorias das categorias). As duas abordagens aqui descritas são: Monadas de Giry e Categorias de Markov. No contexto da primeira abordagem são apresentadas as principais definições, no contexto da segunda é demontrado um teorema de composição análogo ao já conhecido no contexto clássico de teoria ergódica.

Palavras-chave: Probabilidade. Teoria das Categorias. Monadas. Categoriais Monoidais.

Abstract

Kaue de Mello Nogueira Piza. **Categorical Probability Theory**. Thesis (Master's).
Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2023.

This work aims to present some of the main concepts of the categorical approach to probability theory. We first lay down some prerequisites definitions and results from classical probability theory (measure theoretic) and category theory, then we define two main approaches to the subject: One, the first one historically and most classic, via Giry monads and then moves to the more recent concept of a Markov category. At the end, we show a recent result that is an analogue to classical decomposition theorems in ergodic theory.

Keywords: Probability. Category Theory. Monads. Monoidal Categories.

Contents

Introduction

Probability theory is usually developed within the realm of measure theory, heavily depending on analytical methods and, hence, local properties of the underlying spaces. As pointed out in Fritz, 2020, this approach of probability (and statistics) can be compared to programming in machine level code.

The first ideas in a more general framework on which one can develop probability theory are exposed in Lawvere, 1962, it starts with and axiomatic approach into “systems of Markov kernels", within this context we can define in a much more general way concepts from probability theory such as random variables, independence, statistics and so on. This approach was further developed in Giry, 1982, with the approach of monads on the category of measurable spaces, it shows that to develop those systems of Markov kernels we just need a suitable class of measurable spaces and one of probability measures on it, with those we can define a monad on this category, the Giry monad, and this system of Markov kernels arises naturally from the Kleisli construction.

The most recent approach to the categorical construction (as to the author's knowledge) is in the seminal paper Fritz, 2020, with the notion of a Markov category. A Markov category is a symmetric monoidal one with additional structure (such as “copying" and “discarding") that behaves like categories of Markov kernels. The idea of categories that behaves like those of Markov kernels comes from Lawvere, 1962 and Giry, 1982, they are axiomatized in a very similar way to their form as presented in this work in Cho and Jacobs, 2019, where the notions of “copying" and “discarding" are axiomatically introduced in this context (in Cho and Jacobs, 2019 they are called affine CD categories).

Within this categorical framework we get a few advantages comparing to the measure theoretic approach: One of them is greater generality, we are dealing with categories that satisfies some conditions, hence we don't need to restrain ourselves to probability theory, the results achieved are valid to every category with the structure as developed, hence some of the results can be translated into different contexts. Another benefit from this categorical approach is that once we get a result from a category that satisfies the axioms of Markov category, this result is automatically valid to any category that models them, hence, we get to prove only once results that are automatically valid to set-ups ranging from probability theory on finite sets to completely general Markov kernels between measurable spaces and also to (some) stochastic processes, since all of them are models do a Markov category.

Follows an outline of this work:

- We first start with basic definitions from measure theory and probability theory, following some canonical references such as Taylor, 1973, Varadhan, 2001, Billingsley, 1986 and some more focused on our context such as Panangaden, 2009.
- In the second chapter we give and overview of prerequisite concepts from category theory, the main definitions and results, here the reference is the canonical Lane, 1998 and the more focused on out context Perrone, 2021.
- In the chapter about Giry Monads we give the first categorical approach to the subject, as in Lawvere, 1962 and Giry, 1982, following Perrone, 2018. We finish with a result on the Kantorovich monad as detailed in Perrone, 2018.
- In the Markov categories chapter we detail the approach of Markov categories as in Fritz, 2020.
- We conclude with a Miscellaneous chapter containing two results, one with the categorical analogue of a theorem on ergodic decomposition due to Moss and Perrone, 2023 and other with an analogue of the Fischer-Neyman factorization theorem due to Fritz, 2020.

Chapter 1

An overview of probability theory

We start recollecting basic definitions up to the construction of the integral for a real-valued measure. Then we move to the analysis of real-valued random variables. Some of the results here (e.g. the Kolmogorov's $0-1$ law) will be seen in the general context of category theory. The idea is to build the intuitive background needed to grasp the ideas behind the categorical framework, for example, why choosing a monad seems like a reasonable setting to develop categorical probability theory? Why the Kleisli contruction on the Giry monad gives us the Markov kernels from the space of measures? And so on.

At the same time, all the definitions here are not strictly necessary for what will be developed in the next chapters. One can go on to a very deep understanding of the theory later developed thinking about "random maps" and "deterministic maps" in a very intuitive way without any lost, as will be pointed in future chapters.

The outline of the chapter is:

- Recollection of basic concepts of measure theory and integration, the main references are Panangaden, 2009 and Billingsley, 1986
- Definition of basic concepts of probability theory and random variables, as in Billingsley, 1986, Taylor, 1973 and Varadhan, 2001.
- Some classical results, as the law of large numbers and the Kolmogorov $0-1$ law as in Taylor, 1973.

1.1 Main definitions of Measure Theory

Definition 1.1.1. A measurable space is a pair (Ω, Σ) where Ω is a set and Σ is a family of subsets of Ω satisfying the following conditions:

1. $\emptyset \in \Sigma$

2. If $A \in \Sigma$ then $A^c \in \Sigma$
3. For a countable family $\{A_i\}_{i \in \mathbb{N}} \subset \Sigma$ we have $\cup_{i \in \mathbb{N}} A_i \in \Sigma$

The family Σ satisfying this conditions is called a σ-algebra on Ω, its elements are called measurable sets. If we relax the third condition to ask only for the closure under unions only for finite families then we have a σ-field. Also, if in the third condition we ask not for general countable union but for the countable union of pairwise disjoint sets then we have a λ-system over Ω, also called a Dynkin system over Ω.

Lemma 1.1. *For any family $\{\Sigma_\alpha\}_{\alpha \in \Lambda}$ of σ-algebras over some set Ω the intersection of this family is again a σ-algebra.*

The above lemma makes well defined the following:

Definition 1.1.2. For a set Ω and a subset $\Sigma \subset \mathcal{P}(\Omega)$ the σ-algebra generated by Σ is the intersection of all σ-algebras on Ω that contains Σ, it is denoted by $\sigma(\Sigma)$.

One important example is the σ-algebra of the borelians. Consider a topological space (Ω, τ), the σ-algebra generated by all the open sets of (Ω, τ) is called the borelian σ-algebra of Ω, denoted by $\mathcal{B}(\Omega, \tau)$, a set in $\mathcal{B}(\Omega, \tau)$ is called a Borel set, or a borelian set. When the topology is clear from the context we denote the borelians just by $\mathcal{B}(\Omega)$. Whenever we use a topological or metric space, unless otherwise stated, we are always assuming it endowed with the Borel σ-algebra.

Proposition 1.2. *Consider $f : \Omega \to \Omega'$ a function between two sets and $\Sigma_{\Omega'}$ a σ-algebra on Ω', then the set $\{f^{-1}(A) \mid A \in \Sigma_{\Omega'}\}$ is a σ-algebra of Ω. Which is called the σ-algebra on Ω induced by f.*

Lemma 1.3. *Consider $f : \Omega \to \Omega'$ a function between sets and $\mathcal{G} \subset \mathcal{P}(\Omega')$, then:*

$$\{f^{-1}(A) \mid A \in \mathcal{G}\} = \{f^{-1}(A) \mid A \in \sigma(\mathcal{G})\}$$

That is, is a σ-algebra is generated by some set, then the induced σ-algebra is the same as the set of pre-images of the generating set.

Usually, in the light of the above result (and this will be clearer shortly), we only know the elements of a σ-algebra that generated it, that is, we have some set of subsets and want to measure things for them (like volume), but to do measure theory we need to have a measure space, so we need a σ-algebra, one can take the σ-algebra generated by those subsets, which is usually what we do, like with the Borel σ-algebra. The problem is that with that approach we actually don't know who the elements of the σ-algebra generated are[1] and, depending on the context they can be quite complicated. The following two results, the monotone class theorem and the Dynkin's $\lambda - \pi$ theorem are quite handy in those contexts.

The notation $A_n \uparrow A$ is used when we have a nested family of sets $\{A_i\}_{i \in \mathbb{N}}$ where $A_i \subset A_{i+1}$ for all $i \in \mathbb{N}$ and with $\cup_{n \in \mathbb{N}} A_i = A$, analogously we write $A_n \downarrow A$ if $A_i \supset A_{i+1}$ for all $i \in \mathbb{N}$ and $\cap_{n \in N} A_i = A$.

[1] Remember of the definition of the σ-algebra generated by some set, is highly nonconstructive, is the intersection of **all** σ-algebras containing it.

Definition 1.1.3. A family of sets $\mathcal{M}$ is called monotone if it is closed under $\uparrow$ and $\downarrow$. That is, for a family $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{M}$ if $A_i \uparrow A$ then $A \in \mathcal{M}$ and if $A_i \downarrow A$ then $A \in \mathcal{M}$.

Like with σ-algebras arbitrarily intersection of monotone classes is again a monotone class, then we can define the monotone class generated by a set. Two quite straightforward results are that a σ-algebra is always monotone and that a monotone σ-field is a σ-algebra. Hence if we only have closedness under finite unions but we also have closedness for set-sequence convergence ($\uparrow$ and $\downarrow$) we get closedness under countable unions. One direct way to see this result is that if we take, for a σ-field $\mathcal{F}$ the monotone class generated by it, then we get a σ-algebra, and by definition, this σ-algebra contains the σ-algebra generated by $\mathcal{F}$, the monotone class theorems says that those are actually equal:

Theorem 1.4 (monotone class theorem)**.** *Consider a set Ω and $\mathcal{F}$ a σ-field over Ω. The monotone class generated by $\mathcal{F}$ and the σ-algebra generated by $\mathcal{F}$ are equal.*

Note that, if we have a σ-field, taking the monotone class generated by it is a much more intuitive process, we are just picking up, alongside the elements of the σ-field, the union of all nested increasing sequences and the intersection of all nested decreasing sequences of elements of the σ-field. The theorem says that in this context this very “constructive" process generates for us the σ-algebra at the same time, hence we have a way to describe the elements of the σ-algebra generated by the σ-field.

Definition 1.1.4. A π-system is a family of sets closed under finite intersections.

Definition 1.1.5. A λ-system over a set Ω is is a family of subsets of Ω, $\mathcal{F}$, where:

- $\Omega \in \mathcal{F}$
- $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
- For a family $\{A_i\}_{i\in\mathbb{N}}$ with $A_i \cap A_j = \emptyset$ if $i \neq j$ for $i, j \in \mathbb{N}$, we have that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$

It is quite straightforward from the definitions to check that something that is a π-system and a λ-system is a σ-algebra.

Theorem 1.5 (Dynkin $\pi - \lambda$ theorem)**.** *If $\mathcal{M}$ is a π-system and $\mathcal{L}$ is a λ-system then $\mathcal{M} \subset \mathcal{L}$ implies $\sigma(\mathcal{M}) \subset \mathcal{L}$.*

Again, this is a theorem that makes the generated σ-algebra more well behaved in some scenarios, if we know how the elements of some $\mathcal{L}$ are then we know how the elements of the generated σ-algebra looks like.

Definition 1.1.6. Consider (Ω, Σ_Ω) and $(\Omega', \Sigma_{\Omega'})$ two measurable spaces, a function $f : \Omega \to \Omega'$ is called a measurable function if $f^{-1}(A) \in \Sigma_\Omega$ for all $A \in \Sigma_{\Omega'}$. That is, if the inverse image or measurable sets are measurable sets.

Proposition 1.6. *Consider (Ω, Σ_Ω) a measurable space and (Ω', d) a metric space. If a family of measurable functions $\{f_n : \Omega \to \Omega' \mid n \in \mathbb{N}\}$ converges pointwise to a function f, then f is measurable.*

Definition 1.1.7. A measurable function with finite range is called a simple function.

Theorem 1.7. *Consider (Ω, Σ_Ω) a measurable space and a non-negative measurable function $f : \Omega \to \mathbb{R}$. Exists a family of simple functions $\{s_n\}_{n\in\mathbb{N}}$, with $s_i \leq f$ and $s_i \leq s_{i+1}$ for all $i \in \mathbb{N}$*

that converges pointwise to f.

Definition 1.1.8. Consider (Ω, Σ_Ω) a measurable space. A measure μ on (Ω, Σ_Ω) is a function $\mu : \Sigma_\Omega \to \mathbb{R}_{\geq 0}$ where:

- $\mu(\emptyset) = 0$
- For a collection $\{A_i\}_{i \in \mathbb{N}} \subset \Sigma_\Omega$ of pairwise disjoint elements of Σ_Ω, we have that

$$\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i)$$

A triple of a set Ω, a σ-algebra Σ_Ω and a measure μ, $(\Omega, \Sigma_\Omega, \mu)$, is called a measure space. A measure with image in $[0, 1]$ is called a probability measure.

One example that will be very recurrent for us of a measure is the Dirac measure: Consider a measure space $(\Omega, \Sigma_\Omega, \mu)$, define the Dirac measure for an element $x \in \Omega$ by:

$$\delta_x(A) = \begin{cases} 1\,, x \in A \\ 0\,, x \notin A \end{cases}$$

Note that δ_x is a probability measure and also, if we vary x and fix A, $\delta_{-}(A)$ is a measurable function on Ω. Sometimes, to make that two points of view evident we use the notation $\delta(x, A)$ instead of $\delta_x(A)$, hence δ becomes a function $\delta : \Omega \times \Sigma_\Omega \to [0, 1]$.

1.2 Integration

Definition 1.2.1. We say that a simple s function in a measure space $(\Omega, \Sigma_\Omega, \mu)$ is integrable if for every a in the range of s with $a \neq 0$ we have that $\mu(s^{-1}(a)) < \infty$.

Definition 1.2.2. Consider $(\Omega, \Sigma_\Omega, \mu)$ a measure space and $s : \Omega \to \mathbb{R}$ a simple integrable function, the integral of s over Ω with respect to the measure μ is defined as:

$$\int_\Omega s d\mu = \sum_{y \in s(\Omega)} y\mu(s^{-1}(y))$$

Remember that, by definition $s(\Omega)$ is finite, and by the characterization of simple integrable function this integral as defined above is well defined.

Definition 1.2.3. Consider f a non-negative real-valued measurable function of the measure space $(\Omega, \Sigma_\Omega, \mu)$, we say that f is integrable if the set Γ of all non-negative simple functions $s \leq f$ contains only integrable functions and we define its integral by:

$$\int_\Omega f d\mu = \sup_{s \in \Gamma} \int_\Omega s d\mu$$

For a function f we define $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$

Definition 1.2.4. Consider f a measurable function on $(\Omega, \Sigma_\Omega, \mu)$, we say that f is integrable if both f_+ and f_- are and define its integral by:

$$\int_\Omega f d\mu = \int_\Omega f_+ d\mu - \int_\Omega f_- d\mu$$

Lemma 1.8. *Consider f a non-negative measurable function in $(\Omega, \Sigma_\Omega, \delta_x)$, then*

$$\int_\Omega f d\delta_x = f(x)$$

Definition 1.2.5. For an integrable measurable function f in a measure space $(\Omega, \Sigma_\Omega, \mu)$ and a measurable set $A \in \Sigma_\Omega$, we denote by

$$\int_A f d\mu$$

the integral of f restricted to the measurable space $(A, \Sigma_\Omega \cap A, \mu_A)$, where μ_A is the restriction of the measure μ to the subspace $(A, \Sigma_\Omega \cap A)$.

Theorem 1.9 (Monotone convergence theorem). *Consider $\{f_n\}_{n \in \mathbb{N}}$ a sequence of measurable functions in a measurable space $(\Omega, \Sigma_\Omega, \mu)$ and*

1. *For all $x \in \Omega$ we have that $f_i(x) \leq f_{i+1}(x)$, and $0 \leq f_i(x) \leq \infty$ for all $i \in \mathbb{N}$.*
2. $\sup_{n \in \mathbb{N}} f_n(x) = f(x)$ *for all $x \in \Omega$.*

Then:

$$\int_\Omega f d\mu = \sup_{n \in \mathbb{N}} \int_\Omega f_n d\mu$$

The Riesz representation theorem states that every bounded linear functional T on the space of compactly supported continuous functions on a space $(\Omega, \Sigma_\Omega, \mu)$ is the same as integration over Ω with respect to the measure μ, that is, for any measurable f, we have

$$Tf = \int_\Omega f d\mu$$

Considering the Riemann integral we can apply this theorem to say that there is a measure on $\mathbb{R}$ that represents the Riemann integral, that is, exists a measure μ such that for any measurable f we have:

$$\int_{-\infty}^{\infty} f(x) dx = \int_{\mathbb{R}} f d\mu$$

Where the left-hand side represents the traditional Riemann integral and the right hand side of the equation the integral as we have just constructed. This measure μ is called the Lebesgue measure on $\mathbb{R}$.

For example, the measure that represents the functional that for every measurable function f evaluates it at the point $x \in \Omega$ is the Dirac measure δ_x.

1.3 Basic concepts of probability theory

Definition 1.3.1. Consider a probability space $(\Omega, \Sigma, \mathbb{P})$ and (Ω', Σ') a measurable space. A (Ω', Σ')-valued random variable is a measurable function $X : \Omega \to \Omega'$. When $(\Omega', \Sigma') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we will say that X is just a random variable, omitting the (Ω', Σ'). We will abbreviate “random variable" by r.v.

Whenever the space $(\Omega, \Sigma, \mathbb{P})$ is irrelevant to some property we are proving about a r.v. X, we will omit it from the definition. Hence, when the reader see just “a r.v. X" declared, we are assuming a general space $(\Omega, \Sigma, \mathbb{P})$ where this r.v. is defined on.

Remark. Consider X a r.v. in the space $(\Omega, \Sigma, \mathbb{P})$ and $f : \mathbb{R} \to \mathbb{R}$ a measurable function. Then fX is again a r.v.

Definition 1.3.2. Consider a probability space $(\Omega, \Sigma, \mathbb{P})$ and X a r.v. on it. We define the probability distribution measure on $\mathbb{R}$ associated with X the measure μ_X defined as

$$\mu_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}) = \mathbb{P}(X \in A)$$

Definition 1.3.3. A function $F : \mathbb{R} \to [0, 1]$ with the following properties:

- F is monotone.
- F is right-continuous[2].
- $\lim_{x \to +\infty} F(x) = 1$
- $\lim_{x \to -\infty} F(x) = 0$

Is called a distribution function.

Definition 1.3.4. Consider X a random variable, we define the probability distribution function of X, $F_X : \mathbb{R} \to [0, 1]$ as

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X^{-1}(-\infty, x))$$

Remark. It can be easily proved from the definitions that F_X is a distribution function for X a r.v.

Theorem 1.10. *There is a one-to-one correspondence between probability measures μ defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and distribution functions.*

Proof. Consider a measure μ and define $F_\mu(x) = \mu((-\infty, x])$ and for a distribution function F define $\widetilde{\mu_F}((a, b]) = F(b) - F(a)$ for all $-\infty \leq a \leq b \leq \infty$. To complete the proof use Carathéodory Extension Theorem to extend uniquely $\widetilde{\mu_F}$ to a measure μ_F. Then it suffices to show that $F_{\mu_F} = F$ and $\mu_{F_\mu} = \mu$. □

Definition 1.3.5. A r.v. on the probability space $(\Omega, \Sigma, \mathbb{P})$ is called a discrete random variable if there is a countable $B \subset \mathbb{R}$ such that $\mathbb{P}(X \in B) = 1$.

[2] And since it is monotone, it has left limits.

Definition 1.3.6. A distribution function F is is discrete if exists $\{x_j\}_{j\in\mathbb{N}} \subset \mathbb{R}$ and weights $\{p_j\}_{j\in\mathbb{N}} \subset \mathbb{R}$ with $\sum_{j\in\mathbb{N}} p_j = 1$ such that $F(x) = \sum_{\{j|x_j\leq x\}} F(x_j)p_j$

Definition 1.3.7. A distribution function F is called absolutely continuous if exists a non-negative measurable function $f : \mathbb{R} \to \mathbb{R}_+$ such that $F(b) - F(a) = \int_a^b f(t)dt$.

Remark. It follows from results of measure theory that if F is an absolutely continuous function then it is differentiable a.e. and $dF = f$ a.e.

Definition 1.3.8. A distribution function F is called singular if F is continuous and $dF = 0$ a.e.

Theorem 1.11. *Consider F a distribution function. F can be written as a convex combination of a discrete, an absolutely continuous and a singular distribution function. That is, there is $a, b, c \in [0,1]$ and F_a, F_d, F_s a absolutely continuous, discrete and singular distribution functions, respectively with $F = aF_a + bF_d + cF_d$.*

Definition 1.3.9. Consider a probability space $(\Omega, \Sigma, \mathbb{P})$ and X a r.v. on this space where $\int_\Omega |X|d\mathbb{P} < \infty$, then we define the expectation of X, denoted as $E[X]$ by

$$E[X] = \int_\Omega X d\mathbb{P}$$

And we define its variance, denoted by $Var[X]$ as $E[X^2] - E[X]^2$.

Lemma 1.12. *Consider a probability space $(\Omega, \Sigma, \mathbb{P})$ and X a r.v. on this space where $\int_\Omega |X|d\mathbb{P} < \infty$ and $X \geq 0$. Then*

$$E[X] = \int_\mathbb{R} x\mu_X(dx)$$

Lemma 1.13. *Consider X a r.v. and $f : \mathbb{R} \to \mathbb{R}$ a measurable function, and that $E[fX]$ exists. Then*

$$E[fX] = \int_\mathbb{R} f(x)\mu_X(dx)$$

Assuming that the second integral exists.

Proposition 1.14. *Consider X a r.v. with $X \geq 0$, then:*

$$\sum_{n\geq 1} \mathbb{P}(X \geq n) \leq E[X] \leq 1 + \sum_{n\geq 1} \mathbb{P}(X \geq n)$$

Corollary 1.14.1. *Consider X where $\mathbb{P}(X \in \mathbb{N}) = 1$, then*

$$E[X] = \sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n)$$

Proposition 1.15 (Jensen's Inequality). *Consider $\phi : \mathbb{R} \to \mathbb{R}$ a convex function and X an* *integrable r.v. with ϕX integrable, then:*

$$\phi(E[X]) \leq E[\phi X]$$

Proposition 1.16 (Chebyshev's Inequality)**.** *Consider $f : \mathbb{R}_+ \to \mathbb{R}$ a positive monotone non-decreasing function and X a positive r.v. Then*

$$\mathbb{P}(X \geq a) \leq \frac{1}{f(a)} E[fX]$$

Definition 1.3.10. Consider $(\Omega, \Sigma, \mathbb{P})$ a probability space and $A_1, \dots, A_n \in \Sigma$ events. We say that they are independent if for all $\{n_1, \dots, n_p\} \subset \{1, \dots, n\}$ with $n_i \neq n_j$ for $i \neq j$ we have that:

$$\mathbb{P}\left[\bigcap_{i=1}^{p} A_{n_i}\right] = \prod_{i=1}^{p} \mathbb{P}[A_{n_i}]$$

Definition 1.3.11. For $X_1, \dots, X_n$ r.v. we say that they are independent if for every family $\{B_i \subset \mathcal{B}(\mathbb{R}) \mid i = 1, \dots, n\}$ we have

$$\mathbb{P}\left[\bigcap_{i=1}^{n} \{X_i \in B_i\}\right] = \prod_{i=1}^{n} \mathbb{P}[X_i \in B_i]$$

For a non finite set $\{x_\alpha \mid \alpha \in \Gamma\}$ we say that the r.v. are independent if the are for every finite subset of Γ.

Proposition 1.17. *Consider X and Y two independent r.v. with $E[|X|] < \infty$ and $E[|Y|] < \infty$, then:*

$$E[XY] = E[X]E[Y]$$

1.4 First results

Definition 1.4.1. For a set of random variables $X_1, \dots, X_n$ we say that they are identically distributed if $F_{X_i} = F_{X_j}$ for all i and j in $\{1, \dots, n\}$.

Remark. When we have a set of r.v. $\{X_1, \dots, X_n\}$ we will denote the fact that they are independent and identically distributed by i.i.d.

Theorem 1.18 (Weak Law of Large Numbers)**.** *Consider $\{X_i\}_{i \in \mathbb{N}}$ i.i.d. r.v. with $E[X_1^2] < \infty$ and $E[|X|] < \infty$. Let us denote $E[X_1] = \mu$. Then*

$$\lim_{n \to \infty} E\left[\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right] = 0$$

1.5 Convergence

Definition 1.5.1. Consider $(\Omega, \Sigma, \mathbb{P})$ a probability space, X a r.v. on it and $\{X_i \mid i \in \mathbb{N}\}$ a sequence of r.v. defined on it. We say that $X_n \to X$ almost everywhere in $\mathbb{P}$ (denoted by a.e. in $\mathbb{P}$) if exists a set $A \in \Sigma$ with $\mathbb{P}[A] = 1$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. When the probability measure is clear from the context we will just say $X_n \to X$ a.e. in $\mathbb{P}$ by $X_n \to X$ a.e.

Proposition 1.19. *Consider* $(\Omega, \Sigma, \mathbb{P})$ *a probability space,* X *a r.v. on it and* $\{X_i \mid i \in \mathbb{N}\}$ *a sequence of r.v. defined on it, then* $X_n \to X$ *a.e. is equivalent to*

$$\lim_{n\to\infty} \mathbb{P}\left[\bigcap_{i\geq n}\{\omega \in \Omega \mid |X_i(\omega) - X(\omega)| \leq \epsilon\}\right] = 1$$

for all $\epsilon > 0$.

Definition 1.5.2. Consider $(\Omega, \Sigma, \mathbb{P})$ a probability space, X a r.v. and $\{X_i \mid i \in \mathbb{N}\}$ a sequence of r.v. defined on it. We say that X_n converges to X in probability, denoted by $X_n \xrightarrow{\mathbb{P}} X$ if $\mathbb{P}[|X_n - X| > \epsilon] \to 0$ for every $\epsilon > 0$.

Lemma 1.20. *If* $X_n \to X$ *a.e. then* $X_n \xrightarrow{\mathbb{P}} X$.

Lemma 1.21. *If* $X_n \xrightarrow{\mathbb{P}} X$ *then exists a subsequence* $X_{n_k} \to X$ *a.e.*

Definition 1.5.3. For a given $p \in \mathbb{R}_{>0}$ we say that $X_n \xrightarrow{L^p} X$ if

$$\lim_{n\to\infty} E[|X_n - X|^p] \to 0$$

Lemma 1.22. L^p *convergence for some* $p \in \mathbb{R}_{>0}$ *implies probability convergence.*

Theorem 1.23. *Suppose* $X_n \xrightarrow{\mathbb{P}} X$ *and exists* $p \in \mathbb{R}_{>0}$ *and r.v.* Y *with* $E[Y^p] < \infty$ *and* $x_n \leq Y$, *then* $X_n \xrightarrow{L^p} X$.

Theorem 1.24 (Borel-Cantelli)**.** *Consider* $(\Omega, \Sigma, \mathbb{P})$ *a probability space and* $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$, *then:*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[E_n] < \infty \Rightarrow \mathbb{P}\left[\limsup_{n\in\mathbb{N}} E_n\right] = 0$$

Remark. A point $\omega \in \Omega$ is in $\limsup_{n\in\mathbb{N}} E_n$ if and only if exists a sequence $\{n_j\}_{j\in\mathbb{N}}$, strictly increasing, such that $\omega \in E_{n_j}$ for every $j \in \mathbb{N}$. On other terms, a point is in the *limsup* if given any event E_n we can always find another one E_m with $m > n$ with $\omega \in E_m$, so, the point ω never "stops to appear", or appear infinitely often. This justify a frequent notation, in the context of probability theory, for the *limsup* of a sequence of events E_n as E_n i.o. (infinitely often).

Corollary 1.24.1. *Consider a sequence* $\{X_n\}_{n\in\mathbb{N}}$ *of r.v. i.i.d. where* $E[X_1^4] < \infty$*, then*

$$\sum_{i=1}^{n} \frac{X_i}{n} \to E[X_1] \text{ a.e.}$$

Theorem 1.25. *Consider* $(\Omega, \Sigma, \mathbb{P})$ *a probability space and* $\{E_n\}_{n\in\mathbb{N}} \subset \Sigma$ *independent events, then:*

$$\sum_{n\in\mathbb{N}} \mathbb{P}[E_n] = \infty \Rightarrow \mathbb{P}\left[\limsup_{n\in\mathbb{N}} E_n\right] = 1$$

Corollary 1.25.1. *Consider* $(\Omega, \Sigma, \mathbb{P})$ *a probability space and* $\{X_n\}_{n\in\mathbb{N}}$ *r.v. i.i.d., then:*

$$E[|X_1|] = \infty \Rightarrow \mathbb{P}\left[\limsup_{n\to\infty}(|X_n| \geq n)\right] = 1$$

and, for any $c \in \mathbb{R}$

$$E[|X_1|] = \infty \Rightarrow \mathbb{P}\left[\{\omega \mid (X_1 + \cdots + X_n)/n \to c\}\right] = 0$$

We now focus on a new type of convergence, convergence in distribution. To start we need to define weak measure convergence.

Definition 1.5.4. Consider $\mathbb{P}$, $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We say that $\mathbb{P}_n \to \mathbb{P}$ weakly, denoted by $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ if for all $a, b \in \mathbb{R}$ with $a < b$ we have $\mathbb{P}(a, b) = 0$ and

$$\mathbb{P}_n[(a, b]] \to \mathbb{P}[(a, b]]$$

Lemma 1.26. *Consider* F, $\{F_n\}_{n\in\mathbb{N}}$ *distribution functions and* μ_F, $\{\mu_{F_n}\}_{n\in\mathbb{N}}$ *its associated measures, then* $\mu_{F_n} \xrightarrow{w} \mu$ *if, and only if,* $F_n(x) \to F(x)$ *for all* $x \in \mathbb{R}$ *continuity points of* F*.*

Definition 1.5.5. Consider F, $\{F_n\}_{n\in\mathbb{N}}$ distribution functions and μ_F, $\{\mu_{F_n}\}_{n\in\mathbb{N}}$ its associated measures, then we say that $F_n \to F$ in distribution, denoted by $F_n \xrightarrow{d} F$ if $\mu_n \xrightarrow{w} \mu$ or, equivalently, if $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$ continuity points of F.

Definition 1.5.6. Consider $\{X_n\}_{n\in\mathbb{N}}$ and X r.v., we say that X_n converges to X in distribution, denoted by $X_n \xrightarrow{d} X$ if $F_{X_n} \xrightarrow{d} F_X$.

Remark. Note that, unlike the other convergences we have defined so far, the convergence $X_n \xrightarrow{d} X$ is defined in terms of their distribution functions, which are real functions, so the r.v. $\{X_n\}_{n\in\mathbb{N}}$ and X can be all defined in different probability spaces and we can still talk about their convergence.

Theorem 1.27 (Helly-Bray)**.** *Consider* $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ *a family of probability measures in* $\mathbb{R}$*. Then* $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ *is and only if for every continuous bounded real function* f *we have:*

$$\int f d\mathbb{P}_n \to \int f d\mathbb{P}$$

in other notation:

$$E[fX_n] \to E[fX]$$

Lemma 1.28. *For random variables, probability convergence implies distribution convergence.*

1.6 Characteristic Functions

Definition 1.6.1. Consider $(\Omega, \Sigma, \mathbb{P})$ a probability space and X a r.v. on it. Then we define its characteristic function $\phi_X : \mathbb{R} \to \mathbb{C}$ as:

$$\phi_X(t) = E[e^{itX}] = \int_\Omega e^{itX} d\mathbb{P}$$

Lemma 1.29. *Some properties of the characteristic functions:*

- $\phi_X(0) = 1$
- $|\phi_X(t)| \leq 1$
- ϕ_X *is positive defined*[3]
- $\phi_X(-t) = \overline{\phi_X(t)}$
- ϕ_X *is uniformly continuous.*

Definition 1.6.2. Consider μ a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then we can define the characteristic function associated with μ by:

$$\phi_\mu(t) = \int e^{itx} \mu(dx)$$

We saw that probability measures in $(\mathbb{R}, \mathbb{R})$ are in one to one correspondence with distribution functions, now we state that the same occurs with characteristic functions:

Theorem 1.30. *Consider μ, ν two probabilities measures in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then $\mu = \nu$ if, and only if, $\phi_\mu = \phi_\nu$.*

Theorem 1.31. *Exists a constant $k \in \mathbb{R}_{>0}$ such that for all $a > 0$ and μ probability measure in $(\mathbb{R}, \mathcal{B})$, we have*[4]

$$\mu\left(\left[-\frac{1}{a}, \frac{1}{a}\right]^c\right) \leq \frac{k}{a} \int_0^a (1 - \mathbf{Re}(\phi_\mu(t)))dt$$

Theorem 1.32 (Lévy Continuity Theorem). *For a given sequence of random variables X_n, if $\phi_{X_n}(t) \to \phi(t)$ for all $t \in \mathbb{R}$ and some function ϕ which is continuous at $t = 0$, then X_n converges in distribution to some random variable X with $\phi_X = \phi$.*

---

[3] Remember that for a complex valued real function $f : \mathbb{R} \to \mathbb{C}$ to be positive defined means that for all $n \in \mathbb{N}$, $z \in \mathbb{C}^n$ and $t \in \mathbb{R}^n$ we have that $\sum_{i,j=1}^n z_i f(t_i - t_j)\overline{z_j} \geq 0$.

[4] Where **Re** is the real part of a complex number.

Now a proposition that helps to find the functional form of the characteristic function.

Proposition 1.33. *Consider X a random variable with $E[|X|^k] \leq \infty$ for some $k \geq 0$, then ϕ_K admits an expansion of order k around zero, meaning:*

$$\phi_X(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!} E[X^j] + \frac{t^k}{k!} r_k(t)$$

with

- $|r_k| \leq 4E[|X|^k]$
- $\lim_{t \to 0} r_k(t) = 0$

Now an important result that follows from the Levy Continuity Theorem

Proposition 1.34. *Let $\{X_i\}$ be a family of i.i.d r.v. where $E[X_i] = \mu$, then define $Y_i = X_i - \mu$, we assume that $E[|X_i|^2] < \infty$ and $\sigma^2 = Var[Y_i]$, then*

$$\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2)$$

Theorem 1.35. *Consider X_i i.i.d. r.v., where $\mu = E[|X_i|] < \infty$, then:*

$$\frac{\sum_{j=1}^{n}}{n} \xrightarrow{\mathbb{P}} \mu$$

1.7 Series of random variables

Theorem 1.36. *Consider $\{X_i\}_{i \in \mathbb{N}}$ a series of random variables, we define $S_n = \sum_{i=1}^{n} X_i$, then S_n is again a random variable and we can define its associated measure μ_{S_n}, we have the following results:*

- *If exists some probability measure μ with $\mu_{S_n} \xrightarrow{w} \mu$, so $F_{S_n} \xrightarrow{d} F_\mu$, then exists a random variable S with $S_n \xrightarrow{\mathbb{P}} S$, and $\mu_S = \mu$.*
- *If $S_n \xrightarrow{\mathbb{P}} S$ then $S_n \to S$ a.e.*

So, in order to prove that a series of random variable converges almost everywhere it is enough to prove that the sequence of measures associated with the partial sums converges weakly.

Theorem 1.37 (Kolmogorov 1-Series Theorem)**.** *Consider $\{X_i\}_{n \in \mathbb{N}}$ a sequence of independent r.v. where $E[X_i] = 0$ and $E[X_i^2] < \infty$ for all $i \in \mathbb{N}$, then S_n converges a.e.*

Theorem 1.38 (Kolmogorov 2-Series Theorem)**.** *Consider $\{X_i\}_{n \in \mathbb{N}}$ a sequence of independent r.v. where $E[X_i] = a_i$, with $a_i \in \mathbb{R}$ and $E[X_i^2] < \infty$ for all $i \in \mathbb{N}$, then if $\sum_{i \in \mathbb{N}} a_i$ converges and $\sum_{i \in \mathbb{N}} Var[X_i]$ converges we have that S_n converges a.e.*

Theorem 1.39 (Kolmogorov 3-Series Theorem). *Consider $\{X_i\}_{n\in\mathbb{N}}$ a sequence of independent r.v. Consider a $c \in \mathbb{R}_{>0}$ and $Y_i = X_i$ if $|X_i| \leq c$ and 0 otherwise. If:*

- $\sum_{i\in\mathbb{N}} \mathbb{P}[X_i \neq Y_i] \leq \infty$
- *The sequence $S_n^Y = \sum_{i\in n} E[Y_i]$ converges.*
- $\sum_{i\in\mathbb{N}} Var[Y_i] \leq \infty$

Then S_n converges a.e.

1.8 Law of Large Numbers

Theorem 1.40 (The Strong Law of Large Numbers). *Consider $\{X_i\}_{i\in\mathbb{N}}$ a sequence of i.i.d. r.v. with $E[|x_i|] < \infty$ and $E[X_1] = \mu$, then:*

$$\frac{\sum_{i=1}^n X_i}{n} - \mu \to 0 \ a.e.$$

Definition 1.8.1.

Consider $(\Omega, \Sigma, \mathbb{P})$ a probability space. For a random variable X on it we call the σ-algebra generated by it the smallest one that makes X a measurable function, that denoted by $\sigma(X)$, for a sequence of independent random variables $\{X_i\}_{i\in\mathbb{N}}$ we define the tail σ-algebra generated by it, denoted by $\Sigma_T(\{X_i\}_{i\in\mathbb{N}})$, the intersection of all $\sigma(X_k, X_{k+1}, \ldots)$ for $k \in \mathbb{N}$, that is:

$$\Sigma_T(\{X_i\}_{i\in\mathbb{N}}) = \bigcap_{k\in\mathbb{N}} \sigma\left(\{\sigma(X_k, X_{k+1}, \ldots) \mid k \in \mathbb{N}\}\right)$$

Theorem 1.41 (Kolmogorov 0-1 Law). *Consider $\{X_i\}_{i\in\mathbb{N}}$ a family of random variables, then an event $A \in \Sigma_T(\{X_i\}_{i\in\mathbb{N}})$ if, and only if, $\mathbb{P}[A] = 0$ or $\mathbb{P}[A] = 1$.*

Theorem 1.42 (The Second Strong Law of Large Numbers). *Consider $\phi : \mathbb{R} \to \mathbb{R}_+$ a function with the following properties:*

- *ϕ is an even function.*
- *For $x > 0$ we have that $\frac{\phi(x)}{x}$ is an increasing function and $\frac{\phi(x)}{x^2}$ is a decreasing function.*

And $\{X_i\}_{i\in\mathbb{N}}$ a family of random variables with $E[X_i] = 0$ and $\{a_n\}_{n\in\mathbb{N}}$ a increasing divergent sequence of real numbers. Then if $\sum_{i\in\mathbb{N}} \frac{1}{\phi(a_i)} E[\phi(X_i)]$ converges we have that $\sum_{i\in\mathbb{N}} \frac{X_i}{a_i}$ converges a.e. and $\sum_{i=1}^n \frac{X_i}{n}$ converges to 0 a.e.

Chapter 2

Category Theory

Here we will develop the basic concepts of category theory, we start with the definition of categories and functors and then move on to representable functors, limits and colimits and adjoints. The idea is to give to the reader not familiarized with the subject a theoretical minimum, we will always move within the theory in parallel with examples from different fields (geometry, algebra, etc.). The notion that will be really relevant for the remaining of the text is that of an adjoint, so if the reader is familiarized with the concept and its characterization in terms of a unit and a counit this chapter can be skipped.

Follows an outline of the chapter:

- We first define categories and functors, as in Lane, 1998
- We define the notion of representability and the Yoneda Lemma, then move on to the concepts of limits and colimits, as in Perrone, 2021
- We finish with the notion of adjunction, following Lane, 1998 and Borceux, 1994.

2.1 Categories and functors

Definition 2.1.1. A category C consists of two collections, one of objects, denoted by $\mathbf{Ob}(C)$ and one of morphisms $\mathbf{Mor}(C)$ with the following properties:

- Each morphism has assigned two objects, the source and target (or domain and codomain), if a morphism f has source a and target b this will be denoted by $d : a \to b$.
- Each object $a \in \mathbf{Ob}(C)$ has a distinguished morphism $id_a : a \to a$ called the identity morphism.
- For each pair of morphism $f : a \to b$ and $g : b \to c$ exists a morphism $fg : a \to c$, called the composition

This structure must satisfy:

- For each morphism $f : a \to b$ we have:

$$fid_a = id_b f = f$$

- For $f : a \to b$, $g : b \to c$ and $h : c \to d$ we have

$$h(gf) = (hg)f$$

  That is, the composition is associative.

Morphisms are also called arrows, we will use this terms interchangeably. We denote the collection of morphisms between one object a and other b in a category C by $\hom_C(a, b)$.

Some common examples of category are:

- **Set** where objects are sets and moprhism set-functions.
- **Top** where objects are topological spaces and morphisms continuous functions.
- **Vect** where objects are vector spaces and morphisms linear transformations.

Definition 2.1.2. Consider a category C and two objects $a, b \in \mathbf{Ob}(C)$, we say that a and b are isomorphic, denoted by $a \cong b$ if exists morphisms $f : a \to b$ and $g : b \to a$ with $fg = id_b$ and $gf = id_a$. Is for a given f exists g as above we say that f is invertible or that f is an isomoprhism.

Definition 2.1.3. A grupoid is a category where every morphism is invertible.

Definition 2.1.4. Given a category C the opposite category, denoted by C^{op} is the one where the objects are the same and a morphism $f^{op} : a \to b$ in C^{op} is a morphism $f : b \to a$ in C.

Definition 2.1.5. A morphism $f : a \to b$ in a category C is called a monomorphism if for every object $c \in \mathbf{Ob}(C)$ and arrows $g_1, g_2 : c \to a$ we have that if $fg_1 = fg_2$ then $g_1 = g_2$.

Definition 2.1.6. A morphism $f : a \to b$ in a category C is called an epimorphism if for every object $c \in \mathbf{Ob}(C)$ and arrows $g_1, g_2 : b \to c$ we have that if $g_1 f = g_2 f$ then $g_1 = g_2$.

Definition 2.1.7. Consider C and D two categories, a functor $F : C \to D$ consists of:

- For each object $a \in \mathbf{Ob}(C)$ and object $Fa \in \mathbf{Ob}(D)$.
- For each morphism $f \in \hom_C(a, b)$ a morphism $Ff \in \hom_D(Fa, Fb)$.

Such that:

- For any $a \in \mathbf{Ob}(C)$, $Fid_a \cong id_{Fa}$.
- For every pair of composable morphisms $f : a \to b$ and $g : b \to c$ in C, $F(gf) = FgFf$ in D.

Definition 2.1.8. A functor $F : C^{op} \to D$ is also called a contravariant functor $F : C \to D$.

Definition 2.1.9. A presehaf on a category C is a functor $F : C^{op} \to \mathbf{Set}$.

For example, for the category **Set**, the hom functor hom : $A \mapsto \hom_{\mathbf{Set}}(A, \mathbb{R})$ is a presheaf.

Definition 2.1.10. Consider C and D categories, F and G functors $F, G : C \to D$, a natural transformation α between F and G, denoted by $\alpha : F \to G$ consists of:

- For each object $a \in \mathbf{Ob}(C)$, a morphism $\alpha_a : Fa \to Ga$, called the component of α at a.
- for each morphism $f \in \hom_C(a, b)$, the following diagram commutes:

$$\begin{array}{ccc} Fa & \xrightarrow{Ff} & Fb \\ {\scriptstyle \alpha_a}\downarrow & & \downarrow{\scriptstyle \alpha_b} \\ Ga & \xrightarrow[Gf]{} & Gb \end{array}$$

Definition 2.1.11. Consider C a category, a subcategory S of C consist of:

- A sub collection of objects of C, that is, $\mathbf{Ob}(S) \subset \mathbf{Ob}(C)$.
- For each pair of composable morphisms $f : a \to b$ and $g : b \to c$ in C with $f, g \in \mathbf{Mor}(S)$, then $gf \in \mathbf{Mor}(f)$.

Definition 2.1.12. A subcategory S of a category C is called wide if $\mathbf{Ob}(S) = \mathbf{Ob}(C)$.

Some common examples of wide subcategories are:

- The category of sets and injective functions is a wide subcategory **Set**.
- The category of metric spaces and Lipschitz functions is a wide subcategory of the category os metric spaces and continuous functions.

Definition 2.1.13. A subcategory S of a category C is called full if given any two objects $a, b \in \mathbf{Ob}(S)$, $\hom_S(a, b) = \hom_C(a, b)$.

Some common examples:

- The category of abelian groups is a full subcategory of the category of groups.
- The category of finite dimensional vector spaces is a full subcategory of the category of vector spaces. The category of compact Hausdorff spaces is a full subcategory of the category of topological spaces.

Definition 2.1.14. Consider a functor between categories $F : C \to D$, we say that F is:

- Faithful if for every pair of morphisms $f, g \in \mathbf{Mor}(C)$, if $Ff = Fg$ then $f = g$.
- Full if for every $g \in \hom_D(Fa, Fb)$ exists an arrow $f in \hom_C(a, b)$ with $Ff = g$.
- Essentially surjective if for every object $b \in \mathbf{Ob}(D)$ exists an object $a \in \mathbf{Ob}(C)$ with $Fa = b$.
- Fully faithful if it is full and faithful.

Definition 2.1.15. Consider C and D categories, we say that thwey are equivalent if existys functors $F : C \to D$, $G : D \to C$ and natural isomorphisms $\eta : GF \to id_C$ and $\mu : FG \to id_D$.

Definition 2.1.16. A functor $F : C \to D$ defines an equivalence of categories if, and only if, thes fully faithful and essentially surjective.

2.2 The Yoneda lemma

Definition 2.2.1. Consider a category C, we say that a functor $F : C \to \mathbf{Set}$ is representable if it is naturally isomorphic to the functor $\hom_C(c, -)$ for some object $c \in \mathbf{Ob}(C)$. We call c the representing object.

Theorem 2.1 (Yoneda embedding). *Consider C a category, and $b, c \in \mathbf{Ob}(C)$. There is a natura bijection*

$$\hom_C(c, d) \cong \hom_{\mathbf{PSh}(C)}(\hom_c(-, c), \hom_c(-, d))$$

Lemma 2.2 (Yoneda lemma). *Consider C a category, $c \in \mathbf{Ob}(C)$ and $F : C^{op} \to \mathbf{Set}$ a presheaf in C, consider the map*

$$\hom_{\mathbf{PSh}(C)}(\hom_C(-, c), F) \to Fc$$

defined as

$$\alpha \in \hom_{\mathbf{PSh}(C)}(\hom_C(-, c), F) \mapsto \alpha_x(id_c) \in Fc$$

This is a bijection and it is natural both in c and in F.

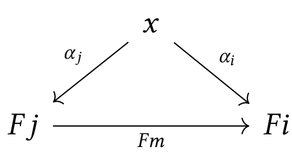
2.3 Limits and Colimits

Definition 2.3.1. Consider C a category and J a small category[1] and consider x and object of C. The constans diagram at x indexed by J is the functor $F_X : J \to C$ where:

- $F_X i = x$ for all $i \in \mathbf{Ob}(J)$.
- $F_x f = id_x$ for all $f \in \mathbf{Mor}(J)$.

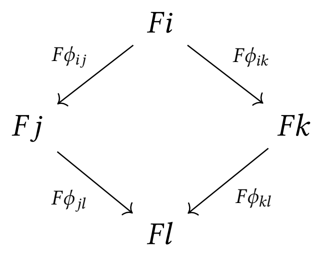
Definition 2.3.2. Consider C a category and J a small category, x and object of C and $F : J \to C$ a functor. A cone over F with tip x is a natural transformation from the constant diagram at x indexed by J to the functor F. A cone under F, or a cocone, with bottom x, is a natural transformation from the functor F to the constant diagram at x.

So, in a more descriptive way, a cone over F is, for every $j \in \mathbf{Ob}(J)$ a morphism $\alpha_j : x \to Fj$ such that for every morphism $m : j \to i$ the following triangle commutes:

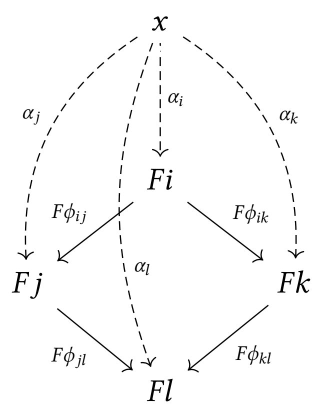


[1] One in which the objects as well as the morphisms forms a set.

Or, in a more case, consider the following diagram:



Then a cone over F with tip x can be seen as the following diagram:

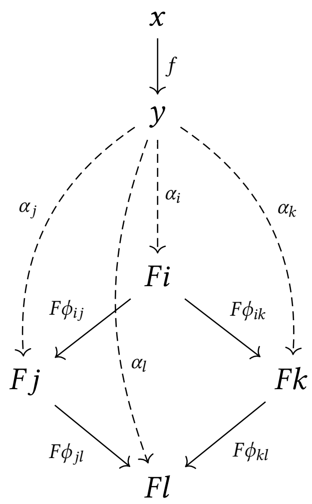


So, the intuition is that a cone with tip x over a diagram is an object that "sees" all the diagram. And a cocone, analogously, is one seen by the diagram, (all the dashed arrows reversed).

Definition 2.3.3. Consider J and C as in the above definition, and $F : J \to C$ a diagram, we define the presheaf $Cone(-, F) : C^{op} \to \mathbf{Set}$ as:

- For an object $x \in \mathbf{Ob}(C)$, $Cone(x, F)$ is the set of cones over F with tip x.

- For a morphism $f : x \to y$ we get a map $Cone(y, F) \to Cone(x, F)$ with the compositions $\alpha_i f$ as described in the diagram below (using the example above):



We also define the functor $Cone(F, -) : C \to \mathbf{Set}$ by:

- For an object of C, x, $Cone(F, x)$ is the set of cocones with tip x.
- For a morphism $f : x \to y$, we get the morphism $Cone(F, x) \to Cone(F, y)$ analogously with the one above, composing the arrows.

Definition 2.3.4. Consider $F : J \to C$ a diagram as in the above definition, a limit of F, if it exists, is an object $\lim F \in \mathbf{Ob}(C)$ representing the presheaf $Cone(-, F)$. A colimit of F, if it exists, is an object $\mathrm{colim} F \in \mathbf{Ob}(C)$ representing the functor $Cone(F, -)$.

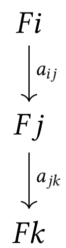
So, breaking into pieces the definition, the object $\lim F$ is equipped, by definition of representation, with a natural isomorphism

$$\hom_C(-, \lim F) \to Cone(-, F)$$

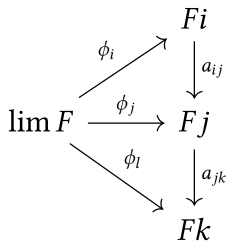
by the Yoneda lemma, the natural elements (or fibers) of this transformation are specified by universal elements of

$$Cone(\lim F, F)$$

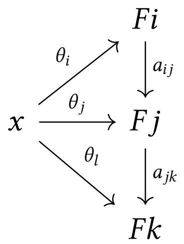
That is, given a diagram



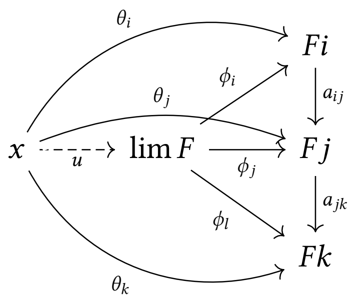
Its limit is an object $\lim F$ and arrows ϕ, such that the following diagram commutes:



And for every object $x \in \mathbf{Ob}(C)$ with arrows $\theta_{\_}$ such that the following diagram commutes:



Exists a unique morphism $u : x \to \lim F$ such that the following diagram commutes:



So, if we use the analogy that a cone over a diagram with tip x represents the fact that x “sees" the objects and morphisms in the diagram, the limit would be the closest one to see it, every other cone over it with another tip will see the diagram “trough" its limit.

The notion of a colimit has an analogous interpretation, but now it is “seen" by the diagram ans it sees every objects that is seen by the diagram.

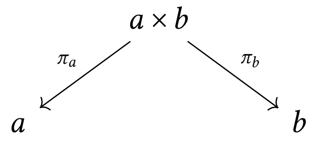
Definition 2.3.5. A category is called complete if it has all limits, that is, every diagram in the category has a limit. A category is called cocomplete if it has all colimits.

A lot of very important categorical concepts can be defined in terms of limits and colimits, such as the product, we will define some of these concepts but for each one we will also break down the definition to give a more constructive one.

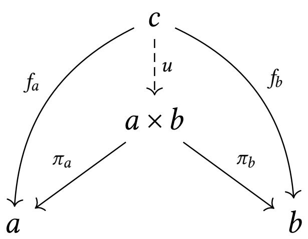
Definition 2.3.6. The limit of a discrete diagram (one with only the identity arrows) is called the product of the diagram. For a discrete diagram with objects $\{a_i\}_{i\in\mathbb{N}} \subset \mathbf{Ob}(C)$, for some category C, the product is usually denoted by $\prod_{i\in\mathbb{N}} a_i$. The colimit is called the coproduct and denoted by $\coprod_{i\in\mathbb{N}} a_i$.

Consider the discrete diagram with only two objects

in the category **Set**. By definition, the limit will be a set $a \times b$ with functions $\pi_a : a \times b \to a$ and $\pi_b : a \times b \to b$:

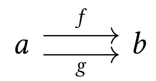


Such that for every set c and morphisms $f_a : c \to a$, $f_b : c \to b$ exists a unique $u : c \to a \times b$ such that the following diagram commutes:



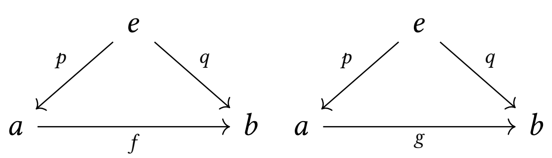
The $a \times b$ is the usual cartesian product, π_a is the projection in the first coordinate and π_b in the second, the u is the function defined as $u(x) = (f_a(x), f_b(x))$, hence $\pi_a u = f_a$ and $\pi_b u = f_b$.

Definition 2.3.7. Consider a category C and two morphisms $f, g \in \hom_C(a, b)$. The limit of the following diagram:



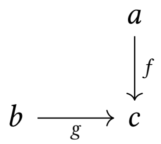
Is called the equalizer of the arrows f and g and the colimit the coequalizer.

From the definition, the limit would be and object e and morphisms $p, q : e \to a$ such that the following diagramms commutes:

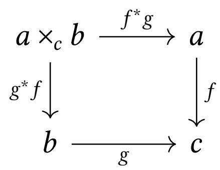


Which amounts to $q = fp$ and $q = gp$, hence $pf = pg$. So we can omit the arrow q, and find ourselves with the usual definition of equalizer: An object e and a morphism $p : e \to a$ such that $fp = gp$ (it “equalizes" the arrows), and for every object c and arrow $h : c \to a$ with $fh = gh$ exists a unique $u : c \to e$ such that $up = h$.

Definition 2.3.8. Consider three objects $a, b, c \in \mathbf{Ob}(C)$ of a category C, and morphisms $f \in \hom_C(a, c)$ and $g \in \hom_C(b, c)$, like the diagram below:

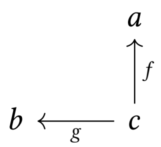


The limit of this diagram is called the pullback, or fibered product. The object lim is usually denoted by $a \times_c b$, the maps giving the universal cone are denoted by $f^*g \in \hom_C(a \times_c b, a)$ and $g^*f \in \hom_C(a \times_c b, b)$, as in the diagram below:

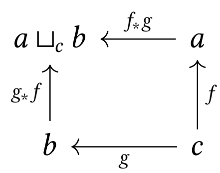


For example, if we are in the category of sets and the maps f and g are inclusions (both a and b are subsets of c), then the pullback is the intersection.

Definition 2.3.9. Consider three objects $a, b, c \in \mathbf{Ob}(C)$ of a category C, and morphisms $f \in \hom_C(c, a)$ and $g \in \hom_C(c, b)$, like the diagram below:



The colimit of this diagram is called the pushout. The object colim is usually denoted by $a \sqcup_c b$, the maps giving the universal cone are denoted by $f_*g \in \hom_C(a, a \sqcup_c b)$ and $g_*f \in \hom_C(b, a \sqcup_c b)$, as in the diagram below:



Definition 2.3.10. The limit of the empty diagram, if it exists, is called the terminal object of a category. The colimit of the empty diagram, if it exists, is called the initial object.

Definition 2.3.11. A functor that preserves limits is called continuous. A functor that preserves colimits is called cocontinuous.

Theorem 2.3. *Representable functors are continuous.*

2.4 Adjunctions

Definition 2.4.1. Consider two categories C and D and functors $F : C \to D$ and $G : D \to C$, an adjunction between F and G is a bijection

$$\hom_D(Fc, d) \cong \hom_C(c, Gd)$$

natural for each $c \in \mathbf{Ob}(C)$ and $d \in \mathbf{Ob}(D)$. We say that F is the left-adjoint and G is the right-adjoint, this relation is represented by $F \dashv G$. Two morphisms related by this bijection are called transposed to each other. We will use the symbols $\sharp$ and $flat$ to denote the transpose. That is, for a morphism $f \in \hom_D(Fc, d)$ its image under the bijection is denoted by $f^\flat \in \hom_C(c, Gd)$, and for a map $g \in \hom_C(c, Gd)$, its image under the bijection is denoted by $g^\sharp \in \hom_D(Fc, d)$, as the image below pictures:

$$\hom_D(Fc, d) \underset{\sharp}{\overset{\flat}{\rightleftarrows}} \hom_C(c, Gd)$$

Now for an alternative definition of adjunction:

Theorem 2.4. *Consider two categories C and D and functors $F : C \to D$ and $G : D \to C$. Suppose we have a pair of natural transformations $\eta : id_C \to GF$ and $\epsilon : FG \to id_D$ satisfying the following (known as the triangle inequalities)*

$$\begin{array}{ccc} F & \xrightarrow{F\eta} & FGF \\ & \searrow^{id_F} & \downarrow{\scriptstyle \epsilon F} \\ & & F \end{array} \qquad \begin{array}{ccc} G & \xrightarrow{\eta G} & GFG \\ & \searrow^{id_G} & \downarrow{\scriptstyle G\epsilon} \\ & & G \end{array}$$

Then $F \dashv G$. The transformations η and ϵ are called the unit and the counit, respectively, of the adjunction $F \dashv G$.

With the above theorem we get an adjunction from the unit and counit transformations. But from a given adjunction we can construct η and ϵ. Consider $F \dashv G$ as in the definition above. Now, for a given $c \in \mathbf{Ob}(C)$ we have a natural bijection $\hom_D(Fc, d) \cong \hom_C(c, Gd)$, which is the same as to say that we have an isomorphism of functors:

$$\hom_D(Fc, -) \cong \hom_C(c, G-) : D \to \mathbf{Set}$$

So, for any $c \in \mathbf{Ob}(C)$, the functor $\hom_C(c, G-)$ is representable, and represented by the object Fc. By the Yoneda lemma, that means that this isomorphism of functors is specified uniquely by an element of $\hom_C(c, GFc)$, and this morphisms has to be the image of id_{Fc} by the bijection, that is $(id_{Fc})^\flat$, this is the unit, for a given $c \in \mathbf{Ob}(C)$, $\eta_c = id_{Fc}^\flat$. Analogously, for a given object $d \in \mathbf{Ob}(D)$, we can define $(id_{Gd})^\sharp$, an that would be our counit. Hence, we can either construct the adjunction/define it by means of an unit and a counit of derive them from the adjunction itself.

Theorem 2.5. *Right-adjoint functors are continuous and left-adjoint functors are cocontinuous.*

2.5 Monads

Definition 2.5.1. Consider a category C, a monad on C consists of of a triple (T, η, μ):

- T is an endofunctor $T : C \to C$
- η is a natural transofrmation $\eta : \mathrm{id}_C \to T$
- μ is a natural transformation $\mu : T^2 \to T$

Such that all the following diagrams commutes

$$
\begin{array}{ccc}
T^3 & \xrightarrow{T\mu} & T^2 \\
{\scriptstyle \mu T}\downarrow & & \downarrow{\scriptstyle \mu} \\
T^2 & \xrightarrow[\mu]{} & T
\end{array}
\qquad
\begin{array}{ccccc}
T & \xrightarrow{\eta T} & T^2 & \xleftarrow{T\eta} & T \\
& {\scriptstyle id}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle id} & \\
& & T & &
\end{array}
$$

We call η the unit and μ the multiplication of the monad.

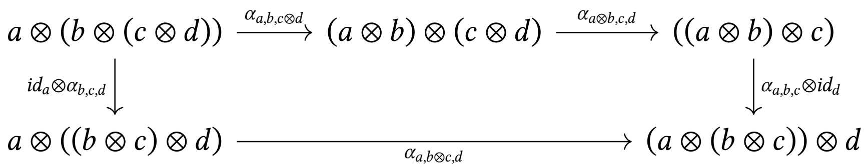
There are two intuitive ways to look at monads, one is seeing a monad over a category as a space of generalized elements, or the closure, and the other as the space where we can evaluate formal expressions, we will look into this interpretations in the following chapter.

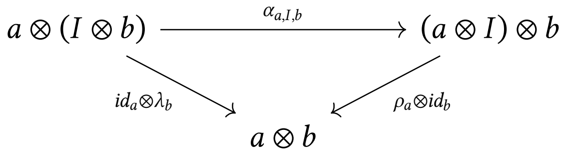
2.6 Monoidal categories

Definition 2.6.1. A monoidal category is a triple $(C, \otimes, I)$ where C is a category, $\otimes : C \times C \to C$ is a bifunctor, a distinguished object I called the unit of the monoidal category and natural isomorphism α, λ, ρ described as:

- $\alpha : (- \otimes -) \otimes - \to - \otimes (- \otimes -)$, called associator, where for each $c, c', c'' \in \mathrm{Ob}(C)$ we have a isomorphism:
$$\alpha_{c,c',c''} : (c \otimes c') \otimes c'' \to c \otimes (c' \otimes c'')$$
- $\lambda : (I \otimes -) \to -$, called left unitor, where for each c we have an isomorphism:
$$\lambda_c : I \otimes c \to c$$
- $\rho : (- \otimes I) \to -$, called right unitor where, again, we get the isomorphism for each object:
$$\rho_c : c \otimes I \to c$$

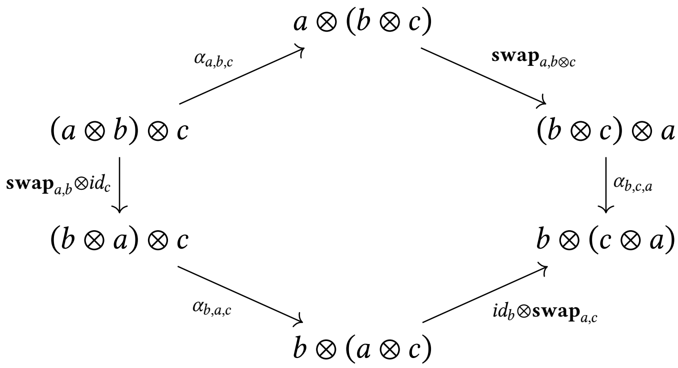
And the natural isomorphisms α, λ, ρ are subjected to coherence conditions: For objects $a, b, c, d \in \mathrm{Ob}(C)$ the following diagrams commutes



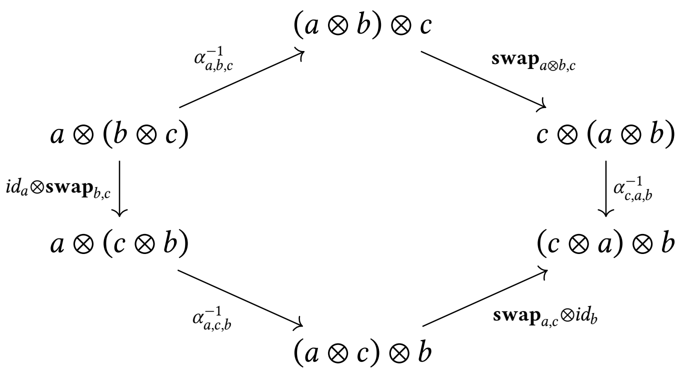


A monoidal category is called strict if the isomorphisms α, λ, ρ are identities, in this case the coherence conditions are trivial.

Definition 2.6.2. A braided monoidal category is a monoidal category $(C, \otimes, I)$ with a natural isomorphism $\mathbf{swap} : (- \otimes -) \to (- \otimes -)$, where $\mathbf{swap}_{a,b} : a \otimes b \to b \otimes a$. This isomorphism is called the braiding or the swap isomorphism, such that the following diagrams commutes for all objects in C:



and



Those are called the hexagon identities.

Braided monoidal categories are the ones with a notion of commutativity, but not in the sence we are used to, it just says to us that we have to have a way to swap elements in the tensors product and not that this squares to the identity, as we would expect in a notion of commutativity, this notion is better translated by the notion os a symmetic monoidal category, as defined below

Definition 2.6.3. A symmetric monoidal category is a braided monoidal category for which the swap map satisfies the additional condition:

$$\mathbf{swap}_{a,b}\mathbf{swap}_{b,a} \cong id_{b \otimes a}$$

Definition 2.6.4. A semicartesian category is a monoidal one $(C, \otimes, I)$ where the unit I is a terminal object.

Now, as we have seen monads are a sructure upon a category, we need to find a way to make this monad structure compatible with the monoidal structure in the cases where we have an underlying category with this structure, that is what the next definitions are for.

Definition 2.6.5. Consider $(C, \otimes_C, I_C)$ and $(D, \otimes_D, I_D)$ two monoidal categories. A functor $T : C \to D$ is called a lax monoidal functor (or just monoidal functor, in our context), if we have natural transformations $T_{a,b} : Ta \otimes_D Tb \to T(a \otimes_C b)$ and $T_0 : I_D \to TI_C$ such that the following diagrams commutes:

$$\begin{array}{ccc}
(Ta \otimes_D Tb) \otimes_D Tc & \xrightarrow{\alpha_D} & Ta \otimes_D (Tb \otimes_D Tc) \\
\downarrow{\scriptstyle T_{a,b} \otimes_D id_{Tc}} & & \downarrow{\scriptstyle id_{Ta} \otimes_D T_{b,c}} \\
T(a \otimes_C b) \otimes_D Tc & & Ta \otimes_D T(b \otimes_C c) \\
\downarrow{\scriptstyle T_{a \otimes_C b, c}} & & \downarrow{\scriptstyle T_{a, b \otimes_C c}} \\
T((a \otimes_C b) \otimes_C c) & \xrightarrow[T\alpha_C]{} & T(a \otimes_C (b \otimes_C c))
\end{array}$$

$$\begin{array}{ccc}
Ta \otimes_D I_D & \xrightarrow{id_{Ta} \otimes_D T_0} & Ta \otimes_D TI_c \\
\downarrow{\scriptstyle \rho_D} & & \downarrow{\scriptstyle T_{a,I_c}} \\
Ta & \xleftarrow[T\rho_C]{} & T(a \otimes_C I_c)
\end{array}
\quad \text{and} \quad
\begin{array}{ccc}
I_D \otimes_D Ta & \xrightarrow{T_0 \otimes_D id_{Ta}} & TI_C \otimes_D Ta \\
\downarrow{\scriptstyle \lambda_D} & & \downarrow{\scriptstyle T_{I_c,a}} \\
Ta & \xleftarrow[T\lambda_C]{} & T(I_C \otimes_C a)
\end{array}$$

Where the ρ_i, λ_i and α_i are the right unitor, left unitor and associator, respectively, for each monoid with $i \in \{C, D\}$. Sometimes, when the notation may cause confusion, a lax monoidal functor may be denoted by a pair (T, Φ), in those cases, Φ is denoting the natural transformation $T_{a,b}$ that we used in our definition. To keep the notation lighter we used the $T_{a,b}$ convention and T_0, but we will use sometime the pair notatio, distinguishing the natural transofrmation, in those cases it is also common to denote T_0 by just Φ, without any parameters. We will denote it by Φ_0.

Definition 2.6.6. Consider $(C, \otimes_C, I_C)$ and $(D, \otimes_D, I_D)$ monoidal categories and $(T, \Phi^T), (F, \Phi^F) : (C, \otimes_C, I_C) \to (D, \otimes_D, I_D)$ lax monoidal functors. We say that a natural transformation $\theta : (T, \Phi^T) \to (F, \Phi^F)$ is a monoidal natural transformation if the following diagrams commutes

$$
\begin{array}{ccc}
Ta \otimes_D Tb & \xrightarrow{\theta_a \otimes_D \theta_b} & Fa \otimes_D Fb \\
{\scriptstyle \Phi^T_{a,b}} \downarrow & & \downarrow {\scriptstyle \Phi^F_{a,b}} \\
T(a \otimes_C b) & \xrightarrow[\theta_{a \otimes_C b}]{} & F(a \otimes_C b)
\end{array}
\quad \text{and} \quad
\begin{array}{ccc}
 & I_D & \\
{\scriptstyle \Phi^T_0} \swarrow & & \searrow {\scriptstyle \Phi^F_0} \\
TI_c & \xrightarrow[\theta_{I_C}]{} & FI_C
\end{array}
$$

Definition 2.6.7. Consider a monoidal category $(C, \otimes, I)$ and a monad on it (T, η, μ). We say that T is a monoidal monad if T is a lax monoidal endofunctor and η, μ are monoidal natural transformations. Following the above notations, a monoidal monad will be sometimes denoted by $(T, \eta, \mu, \Phi^T, \Phi^T_0)$.

Chapter 3

Giry Monads

In this chapter we develop one of the most classical approaches of categorical probability, the one by means of a Giry monad, [Giry, 1982, Lawvere, 1962]. The fundamental idea is to begin with the category of measurable spaces, and find a convenient notion to connect a measurable space to the space of probability measures over it, demanding a few basic things, indispensable to the core ideas of probability theory, this connection must take place in the category of measurable spaces, after all, that is where we know how to do things in the first place, we mus have a way to evaluate our measures (the expected value), we must be able to get the notion of joint distribution and marginals. It turns out that the first two requirements are well modeled by the notion of a monad, the last two when we endow it with a monoidal structure.

Once we get the basic model, we can start to relax our premises, for example, we know that for a monad we have associated a Kleisli category, what this category would be in this context? And we see that from the category of measurable spaces and the monad we are gonna build upon it we can via the Kleisli construction get to the category of measurable spaces with Markov Kernels, and so on.

Follows a short outline of the chapter:

- The first two sections contains the main definitions to follow along, the categories **Meas** and **Stoch** and the notion of a monad. The main references are Lane, 1998 and Voevodsky, 2004.
- Then we go to explain the Kleisli construction and also how to see monads as a model to “generalized" elements of the underlying category. The main references are Perrone, 2018, Lane, 1998 and Marmolejo and Wood, 2010.
- We introduce the notion of a Giry monad. The references are Giry, 1982, Lawvere, 1962 and Perrone, 2018.
- A short categorical detour to show that we can view monads as adjunctions and vice-versa and with a description of monoidal categories. The principal references are Lane, 1998 and Fritz and Perrone, 2018.
- Two final sections with a further discussion in the Giry monad construction and

also one overview of the construction of the Kantorovich Monad, a Giry Monad over the category of complete metric spaces, the main reference for this part is PERRONE, 2018.

3.1 The categories Meas and Stoch

We start with the category **Meas** of measurable spaces (Ω, Σ_Ω) with measurable functions as the morphisms. The category **Stoch** has again as objects measurable spaces (Ω, Σ_Ω), but now a morphism $(\Omega, \Sigma_\Omega) \to (\Omega', \Sigma_{\Omega'})$ is given by a map

$$f : \Sigma_{\Omega'} \times \Omega \to [0, 1]$$

we denote $f(S, c)$ by $f(S \mid c)$ and the morphisms f have the following properties:

- $f(- \mid c) : \Sigma_{\Omega'} \to [0, 1]$ is a probability measure in $(\Omega', \Sigma_{\Omega'})$ for every $c \in \Omega$
- $f(S \mid -) : \Omega \to [0, 1]$ is measurable for every $S \subset \Sigma_{\Omega'}$.

The composition law with another $g : \Sigma_{\Omega''} \times \Omega' \to [0, 1]$ is given by

$$gf(S \mid c) = \int_{h \in \Omega'} g(S \mid h) f(dh \mid c)$$

known as the Chapman-Kolmogorov formula. In order to see that this new composite actually is a morphism in the category, that is, satisfy the two conditions of the morphisms we note:

- The σ-additivity in S follows from the same property in g and the dominated convergence theorem.
- The measurability in c for a fixed S, first assume that g is simple, then follows from the measurability assumption on f, the it follows for any g by the definition of the Lebesgue integral.

The identity morphisms are the indicator functions on the first parameter, that is $id(S \mid c) = 1$ if $c \in S$ and 0 otherwise.

To verify that **Stoch** is a category and see some of its properties, the reader may consult PANANGADEN, 2009, chapter 5, he actually constructs a more general category **SRel**, where $f(- \mid c)$ is a subprobability but the results are naturally applied to **Stoch** as we defined here (more in line with VOEVODSKY, 2004 and FRITZ, 2020). We will also treat in a lot of details the category **Stoch** in the next chapter.

3.2 The Kleisli category construction: monads as generalized elements

Let us take a closer look in the definition of a monad. We are given a category C, the first thing we have to have, an endofunctor $T : C \to C$, that is, to each element c we assign a new one, Tc, let us call Tc the extension of c by T in C. We also need, for each morphism

$f : c \to c'$, a morphism that takes our extension of c by T and led us to the extension of c' by T, hence, a coherent way to move from our base ground, c, to its extension (so far that is only the definition of a functor). We will call now T an extension, as explained. Now we also need the natural transformation η, which is, for each element c, a map $\eta_c : c \to Tc$, hence, an embedding of c into its extension and for each map $f : c \to c'$ we must have $\eta_{c'} f = \eta_c T f$, in diagrams:

$$\begin{array}{ccc} c & \xrightarrow{f} & c' \\ {\scriptstyle \eta_c}\downarrow & & \downarrow{\scriptstyle \eta_{c'}} \\ Tc & \xrightarrow[Tf]{} & Tc' \end{array}$$

Now η_c will be the base element c inside its extension Tc. One natural question is: If we are talking about "generalized elements" and T is the extension of one element into its generalized form, why do we need η? Why don't just say that the base element inside its extension is, well, just Tc? Let us see an example: Consider the category **Set**, and a set X, take T for the power endofunctor, that is, $T(X) = \mathcal{P}(X)$, we have a notion of element in X, of course, it is a set, but we also have a natural way, for every $x \in X$ to see it as an element of $\mathcal{P}(X)$, its singleton $\{x\} \in \mathcal{P}(X)$, and in this context, that is what the η_X would be, the map $\eta_X : X \to \mathcal{P}(X)$ that takes each element into its singleton. So the TX is not quite the right "depiction" for X inside its extension TX, but this is rather represented by η_X, that is the intuition behind the η transformation in the monad definition, the depiction of object in the base category inside their extensions.

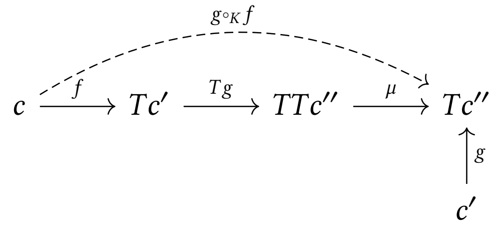
In the same way that η was our depiction of elements inside their extensions, μ will be our depiction of generalized elements of generalized elements, into generalized elements, we will explain it better below.

Consider a context where we would need this notion of a generalized element, like if we are applying something (a limit, le us say) and the result may not live within our original object c, then we can extend that object by T and find that this result lives within Tc, if we have a monad structure, η give us a way to embed our original object in a coherent way to live alongside this result, so our extension is well behaved, μ tells us that we only need to extend our object once, to make room for our result, if we have to extend it twice, that is, our result is actually in T^2c, we have a way to go back to our original extension Tc, by means of $\mu : T^2 \to T$. And this is consistent with our embedding η, since by the definition of a monad we have: $\mu T\eta \cong id \cong \mu\eta T$. For a formal approach of monads as generalized elements the reader can consult Marmolejo and Wood, 2010.

In our example, the way to go from subsets of subsets of a set to subsets of the same set is just to take the union, so the union is our $\mu : \mathcal{P}\mathcal{P} \to \mathcal{P}$.

Definition 3.2.1. Consider a category C and (T, η, μ) a monad on it. A morphism $k \in \hom_C(c, Tc')$ is called a Kleisli morphism of T from c to c'.

Definition 3.2.2. Consider a category C and (T, η, μ) a monad on it, also two Kleisli morphisms $f \in \hom_C(c, Tc')$ and $g \in \hom_C(c', Tc'')$, then the Kleisli composition of f anf g, denoted by $g \circ_K f$ is defined as $g \circ_K f = \mu(Tg)f$, or, in diagrams:



Definition 3.2.3. Given a category C and a monad (T, η, μ) on it. The **Kleisli** category of T, denoted by C_T is the category where objects are the objects of C and morphisms are the Kleisli morphisms of T. The composition is given by the Kleisli composition.

For one object in the Kleisli category, its identity $id_c^{C_T}$ is a map $id_c^{C_T} \in \hom_C(c, Tc)$, from the definitions above it follows that this identity must satisfy the following equation:

$$\mu_c T_c id_c^{C_T} = id_c^C$$

And this equality must hold for any $c \in \mathbf{Ob}(C)$, from the definition of monad we see that $id_c^{C_T} = \eta_c$, that is why η is called the unit of the monad, because in the Kleisli category of the monad its fibers are the identity maps of the elements.

Before focusing on the Eilenberg-Moore construction, let us see what the Kleisli construction would look like in our example of sets and the monad of subsets.

A Kleisli morphism from a set X to a set Y is a multi-valued function $f : X \to \mathcal{P}(Y)$, that for every element in X associates a subset in Y, for another multi-valued function $g : Y \to \mathcal{P}(Z)$, the Kleisli composition would be taking first the image under f, then applying the power set functor in the map g to go from $\mathcal{P}(Y)$ to $\mathcal{P}^2(Z)$, that is, consider $x \in X$, we take the image $f(x) \subset Y$, then we apply g for each element in $f(x)$ and apply the union (μ) in all of those, in other words

$$(g \circ_K f)(x) = \bigcup_{y \in f(x)} g(y) \subset Z$$

To summarize the example, the powerset functor with singleton inclusion as unit and union as multiplication/composition is a monad on the category of sets, and the Kleisli category of this nomad is the category of sets and multi-valued functions.

3.3 The first Giry monad

Consider the category **Meas** of measurable spaces and the endofunctor P that for each space $X \in \mathbf{Ob}(\mathbf{Meas})$ assigns the space PX of probability measures on X, as described in the begining of the chapter. For each space X, we have a map $\eta_X : X \to PX$ that evaluates the Dirac measure, that is, $\eta_X(x) = \delta_x$ for $x \in X$, and for each space X we get a map $\mu_X : PPX \to PX$ which maps a measure $\mathbb{P} \in PPX$ to the measure $\mu_{X_\mathbb{P}} \in PX$ defined by:

$$A \in \Sigma_X \mapsto \int_{PX} p(A) d\mathbb{P}(p)$$

Lemma 3.1. *(P, η, μ), as described above, is a monad over **Meas**.*

The proof can be consulted in PERRONE, 2018 .This monad is one of the class of monads called **Giry monads**. By the definition given in GIRY, 1982 a Giry monad is “the monad on a category of suitable spaces which sends each suitable space X to the space of suitable probability measures on X", hence there is a nonunique definition of a Giry monad, this one we've constructed is one of them, one of the main challenges of categorical probability theory, as we will see, is to find good classes of suitable spaces and suitable probability measures. For example one can ask for X to be a complete metric space, or compact Hausdorff spaces.

Proposition 3.2. *The Kleisi category of **Meas** is the category **Stoch**.*

The proof can be consulted in PERRONE, 2018.

3.4 The Eilenberg-Moore category and the adjuction equivalence

Definition 3.4.1. Consider (T, η, μ) a monad over a category C. A T-algebra is a pair (c, h), where $c \in \mathbf{Ob}(C)$ and $h \in \hom_C(Tc, c)$ where the following diagrams commutes:

$$\begin{array}{ccc} T^2c & \xrightarrow{Th} & Tc \\ {\scriptstyle \mu_c}\downarrow & & \downarrow{\scriptstyle h} \\ Tc & \xrightarrow[h]{} & c \end{array} \qquad \begin{array}{ccc} c & \xrightarrow{\eta_c} & Tc \\ & {\scriptstyle id_c}\searrow & \downarrow{\scriptstyle h} \\ & & c \end{array}$$

The object c is called underlying object of the algebra and the morphism h the structure map of the algebra. A morphism $f : (c, h) \to (c', h')$ between two T-algebras is an arrow $f \in \hom_C(c, c')$ where

$$\begin{array}{ccc} Tc & \xrightarrow{h} & c \\ {\scriptstyle Tf}\downarrow & & \downarrow{\scriptstyle f} \\ Tc' & \xrightarrow[h']{} & c' \end{array}$$

commutes.

Definition 3.4.2. For every monad (T, η, μ) on a category C and object $c \in \mathbf{Ob}(C)$ we get an algebra (Tc, μ_c), which is called the free algebra generated by c. A T-algebra is called free if it is of this form for some object c in C.

Definition 3.4.3. Consider C a category and (T, η, μ) a monad over it. With the definition above we get a category, the category of T-algebras, which is called the **Eilenberg-Moore** category of T and denoted by C^T.

Proposition 3.3. *Consider (T, η, μ) a monad on a category C, (c, h) a T-algebra and $a \in$ **Ob**(C), then there is a natural bijection:*

$$\hom_C(c, a) \cong \hom_{C^T}(Tc, a)$$

For a proof, the reader can consult Lane, 1998, as for the other proof of the results presented in this section. Both constructions, the Eilenberg-Moore and the Kleisli are "equivalent" to each other, in the following sense:

Lemma 3.4. *Consider (T, η, μ) a monad on a category C. There is an equivalence of categories between C_T (the Kleisli category of T) and the full subcategory of C_T of the free T-algebras.*

Now we focus on a standard result about monads, that the notion of a monad is dual to the notion of adjunction.

Lemma 3.5. *For an adjunction (F, G, η, ϵ), the triple $(GF, \eta, G\epsilon F)$ is a monad in the associated category.*

The proof is quite straightforward from the definitions, so this lemma shows us that every adjunction gives rise to a monad, and the following proposition shows us the converse:

Proposition 3.6. *Consider C a category and (T, η, μ) a monad over it. There are functors F^T, G^T and η^T, ϵ^T such that $(F^T, G^T, \eta^T, \epsilon^T) : C \to C^T$ is an adjunction and the monad defined over C by it is the starting monad (T, η, μ).*

Proof. Define G^T by the functor that forgets about the structure map of the algebra and F^T the ones that maps $c \in \mathbf{Ob}(C)$ into the algebra (Tc, μ_c) (the free T-algebra on c). Then it is a straightforward calculation to see that $(T, \eta, \mu) \cong (G^T F^T, \eta^T, G^T \epsilon^T F^T)$ where $\epsilon^T(c, h) = h$ and $\eta^T = \eta$. □

3.5 Further into the Giry monad

As we've seen in the first Giry moand section, we can construct a Giry monad by the following steps: Taking the category **Meas** and defining the endofunctor P and the one that assigns for each space X the space PX of probability measures on it. The unit of this monad is the Dirac measure natural transformation $\delta_X : X \to PX$ that maps each element $x \in X$ into the Dirac measure δ_x and the composition law of the monad is given by taking expectation $E : PP \to P$, that is for a given measure $\mathbb{P} \in PPX$ we assign $E\mathbb{P} \in PX$ by

$$E\mathbb{P} : A \mapsto \int_{PX} p(A) d\mathbb{P}(p)$$

Usually we are not just interest in uderstand the measures in the measurable spaces, but also the joint distributions in the product spaces. That is way the natural setting to a categorical probability theory should be not just a Giry monad on **Meas** but also a monoidal structure in this category. This structure is given by the canonical set-theoretical product, we will still use the notation $X \otimes Y$ to keep it consistent with the following sections. At the same time, we need to give a monoidal structure to the Giry monad P we have constructed, this amount to the transformations Φ^T and Φ_0^T, in the context of Giry monads we will call them ∇^P and ∇_0^P, and omit the P of the notation when the monad is clear from the context.

Now we will keep to work with our intuition on **Meas** but assume a more general context, that we are in a semicartesian monoidal category C, the I can be interpreted as a deterministic state, this notion can be justified by the fact that for every space X we have that $X \otimes I \cong X$, so considering the states of I alongside with the ones of X is the same of just looking for X, hence, there is no randomness associated with the object I.

Definition 3.5.1. A Giry moand is called affine if $P(1) \cong 1$.

In the context of affine Giry monads with a monoidal structure we can forget about the map ∇_0^P.

Remembering the definition, ∇ is such that

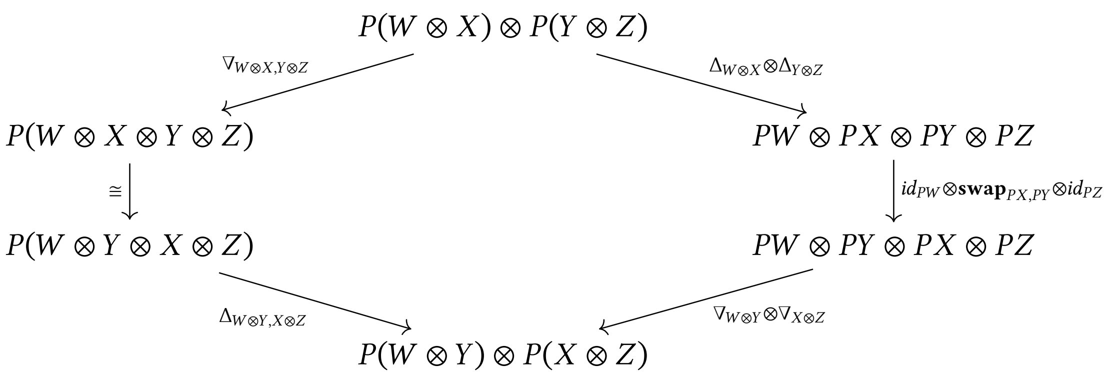
$$\nabla_{X,Y} : PX \otimes PY \to P(X \otimes Y)$$

Also, let us assume that this category is not only a monoidal one but a symmetric monoidal one, hence our set up is a symmetric semicartesian monoidal category $(C, \otimes, I)$ with an affine Giry monoidal monad (P, δ, E, ∇).

In this setting, given a measure $\mathbb{P}_1 \in PX$ and a measure $\mathbb{P}_2 \in PY$ there is a canonical way, via ∇, to assign a joint probability measure $\nabla(\mathbb{P}_1 \otimes \mathbb{P}_2) \in P(\mathbb{P}_1 \otimes \mathbb{P}_2)$.

In the same way we described a monoidal monad, we could define an opmonoidal monad, one where the endofunctor is an oplax monoidal endofunctor, that is, we have natural transformation $\Delta^P_{X,Y} : T(X \otimes Y) \to X \otimes Y$ (all the diagrams in the definitions are constructed in an analogous way), and within this context for a joint probability measure $p \in PX$ we get a pair or marginal distributions $\Delta(p) = (p_X, p_Y) \in PX \otimes PY$.

In this context of a monad with both structures, a monoidal and an opmonoidal structure, we can define a bimonoidal monad by the means of interactions of Δ and ∇ with the monoidal structure $\otimes$ and the braiding structure **swap**. Resuming, the following diagram must commute:



Ans the Giry monad we have defined actually has a bimonoidal structure (which is to say that all the operations of forgeting randomness $X \to I$, taking the expectation E and the Dirac measure δ are coherent with the bimonoidal structue ∇, Δ), and $\Delta\nabla \cong id_{-\otimes-}$. Hence, we have a categorigal notion of joints and marginals. For complete definitions the reader can consult the section 1.2.2 and the appendix A of Perrone, 2018, and Fritz and Perrone, 2018 for complete proofs.

3.6 The Kantorovich Monad

The idea of this section is to show an application of the Giry monad approach to probability theory. We start with a metric space X and construct a metric in PX and then define the Kantorovich Monad over the category of metric spaces. The reference for this section, including the detailed proofs of the results, is Perrone, 2018.

Definition 3.6.1. The category **Met** is the category where objects are metric spaces (X, d_X) and morphisms $f : X \to Y$ are short functions, that is, functions where, for all $x, x' \in X$ we have

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

We can endow the category with a monoidal structure by defining $X \otimes Y$ to be the product set with the metric

$$d_{X\otimes Y}((x, y), (x', y')) = d_X(x, x') + d_Y(y, y')$$

Definition 3.6.2. The category **CMet** is the category of complete metric spaces, with morphisms and monoidal structure just as in **Met**

Lemma 3.7. ***CMet*** *is a full subcategory of* ***Met****.*

Whenever we talk about a probabiolity measure in a metric space we are considering the σ-algebra of the borelians and a Radon probability measure.

Definition 3.6.3. Consider a metric space X, a probability measure $\mathbb{P}$ on X has finite first moment if the expected distance between two random points is finite, that is

$$\int_X d_X(x, y) d\mathbb{P}(x) d\mathbb{P}(y) < \infty$$

Definition 3.6.4. For X a metric space, PX will denote the set of probability measures of finite first moment on X.

Considering the push forward of measures it is clear that $P : \mathbf{Met} \to \mathbf{Set}$ is a functor, now we want to endow the set PX itself with a metric. That is what we will do, but rather for the category **CMet**

Definition 3.6.5. Consider $X \in \mathbf{Ob}(\mathbf{CMet})$, The Wasserstein space PX associated with X is the set of probability measures with finite first moment with metric defined as

$$d_{PX}(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\mathbb{P}\in\Gamma(\mathbb{P}_1,\mathbb{P}_2)} \int_{X\times X} d_X(x, y) d\mathbb{P}(x, y)$$

Where $\Gamma(\mathbb{P}_1, \mathbb{P}_2)$, as defined in the first chapter, is the set of probability measures on $X \times X$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$.

One interpretation of the metric in the Wasserstein space, is that for two distributions over X the distance measures the amount of work needed to change one into the other.

Theorem 3.8. *If X is a complete metric space, then its Wasserstein space is also complete.*

With the above theorem and the the technical lemma 2.1.14 in Perrone, 2018 we conclude that P : **CMet** → **CMet** is an endofunctor, called the Kantorovich functor.

Considering that finitely supported measures with rational coefficients are dense in PX, and those measures can be completely determined by powers of X up to permutations, one can see PX as the colimit of the diagrams of powers of X within the category $CMet$, using this we can define the Kantorovich functor as the colimit of power functors[1], within that framework we can characterize the map E (the integration of a measure on measures to a measure) by a universal property, taking δ to be the canonical Dirac transformation we endow P with the structure of a monad.

Here we just painted a very general picture of how the monad structure arises, bute one thing is worth noting, that already point out the power of the approach: As observed in the last paragraph, E is the usual integration on the space of measures, but it arises naturally from a colimit of power functors, that is, it is a way to integrate without integration.

To wrap upon this section, we mention a result that makes our point of the power of the categorical approach to the subject, consider **ConvBan** the category whose object are convex subsets of Banach spaces and morphisms short affine maps, with this definition we have a canonical functor **ConvBan** → $\mathbf{CMet}^P$ and this functor is actually an equivalence of categories, that is:

Theorem 3.9. ***ConvBan*** ≅ $\mathbf{CMet}^P$

Where $\mathbf{CMet}^P$ denotes the Eilenberg-Moore category, as described in the previsous sections.

[1] The construction is much more intricate than just taking powers of the space, and can be consulted in details in Perrone, 2018.

Chapter 4

Markov Categories

In this chapter we lay the basic notions of categorical probability theory in its most recent developments. We will follow closely the treatment in the seminal paper Fritz, 2020 and we refer to this paper for the reader interested in a more detailed exposition of the subject. Follows a layout of the chapter:

- We start giving detailed definitions on the categories **FinStoch** and **Stoch**, following the expositions in Fritz, 2020, Moss and Perrone, 2023 and Perrone, 2023.
- We move to the important notion of a deterministic morphism in a Markov category, showing that if we have only deterministic morphisms, our category is "trivial" (in the sense that it is cartesian). WE also state a strictification theorem, the mains reference for this and the next sessions is Fritz, 2020.
- The we define important notions within the context of Markov categories that extendes the classical ones, joint states, conditioning and the notion of almost surely.
- In the last section we give the main definitions from statistics as developed in the framework of Markov categories, preparing for the first result in the next chapter.

4.1 Building up from two examples

We will start by defining a Markov category in a constructive/intuitive way, then in the next section we give the formal definition alongside with the main concepts withn Markov Categories.

We will start with two examples, **Stoch** and **FinStoch**. The objects of **FinStoch** are finite sets and the objects of **Stoch** are measurable spaces. A morphism between two objects can be called a channel sometimes in the context of Markov categories.

In **FinStoch** morphisms between two sets X and Y, $f : X \to Y$ are stochastic matrices, that is a matrix with rows indexed by elements of X and columns indexed by elements of Y, with positive entries and where the sum at each column adds up to 1. Hence we can see a morphism $f : X \to Y$ as a function $f : X \times Y \to [0, 1]$, with $f(x, y) \mapsto f(y \mid x)$ such that for any $x \in X$ we have

$$\sum_{y \in Y} f(y \mid x) = 1$$

One can interpret $f(y \mid x)$ as the transition probability from state x to state y. But note that for each fixed $x \in X$ we get a function $f_x : Y \to [0,1]$ with $f(y \mid x)$ and, by definition, this is a probability measure, hence we can also see a morphism $f : X \to Y$ as a familly of probability measures on Y indexed (or parametrized) by X, the familly f_x. Hence we can see the morphisms f as a function from X to the space of probability measures over X, denoted by PX.

In the context of **Stoch** a morphism $f : X \to Y$ is a Markov kernel from the measurable space (X, Σ_X) to the measurable space (Y, Σ_Y), which is a map $f : X \times \Sigma_Y \to [0,1]$ with $f : (x, S) \mapsto f(S \mid x)$ that is measurable in the first argument and is a probability measure in the second argument. That is inspired by the morphisms we had in **FinStoch**, cince if we consider $f_x = f(- \mid x) : \Sigma_Y \to [0,1]$ we get a probability measure (in the context of **FinStoch** we could see it as only "acting" on the elements of Y, not on subsets of it).

From any measurable function $f : X \to Y$ we get a kernel K_f defined as

$$K_f(S \mid x) = \delta_{f(x)}(S)$$

that is, the dirac measure on the point $f(x)$. This is a function defined on $X \times \Sigma_Y$ with takes values in $[0,1]$, it is by definition measurable in the first coordinate and is a probability measure in the second coordinate.

Now suppose we have a distinguished object, called unit, denoted by I (in the context of **Stoch** and **FinStoch** I is the one-point space). A morphism $f : I \to X$, for a measurable space X, is called a source in **Stoch**, note that a source is a morphism $f : \{\cdot\} \times \Sigma_Y \to [0,1]$ which is a probability measure on the second coordinate, but that is just a probability measure, and any probability measure $\mathbb{P}$ on Y trivially defines such a morphism $I \to Y$, hence there is a correspondence between morphisms from the unit to a measurable space and probability measures on that space.

In order to **Stoch** and **FinStoch** to become a category we need to have identity morphisms $id_X : X \to X$ and composition gf with $f : X \to Y$ and $g : Y \to Z$. So let us look at those for each one of the categories: In the **FinStoch** one the identity morphism is just the identity matrix and the composition is matrix multiplication, tha can be written as:

$$gf(z \mid x) = \sum_{y \in Y} g(z \mid y) f(y \mid x)$$

for $z \in Z$ and $x \in X$. This formula is also known as the Chapman-Kolmogorov formula.

For the **Stoch** category, identity is the Dirac measure, in the following sense:

$$id_X : X \times \Sigma_X \to [0,1]$$

with

$$id_X(S \mid x) = \delta_x(S)$$

and composition is given by

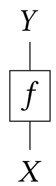
$$gf(S \mid x) = \int_Y g(S \mid y) f(dy \mid x)$$

whete the symbol $f(dy \mid x)$ means that integration is with respect to the measure $f_x = f(- \mid x)$ defined in Σ_Y, that is $f(dy \mid x) = df_x$, hence we can rewrite the above equation as

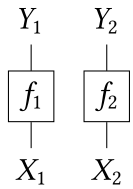
$$gf(S \mid x) = \int_Y g(S \mid y) df_x$$

It is straightforward to verify that those compositions satisfies the categorical requisites of associativity and coherence with identities.

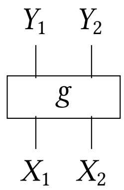
Within Markov categories we also have a notion of parallel composition. That is, for two objects X and Y we want a composite object $X \otimes Y$, which is going to be the space of joint distributions. Now is the time for us to introduce the notation that will be used throghout this chapter, string diagrams. Within some category as the ones we are building here (for now, with a “product" $\otimes$ and a unit I), we denote a morphism $f : X \to Y$ as the one below:



Joint distributions are represented by parallel lines, hence for $f_1 : X_1 \to Y_1$ and $f_2 : X_2 \to Y_2$, the product $f_1 \otimes f_2$ is denoted in the following diagram:

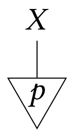


To make clear the difference from a morphism like $f_1 \otimes f_2$ to a generic morphism $g : X_1 \otimes X_2 \to Y_1 \otimes Y_2$ we denote the second by:



Note that diagrams are always readed from bottom up.

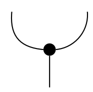
We also have a special notation for a morphism $p : I \to X$, as follows:



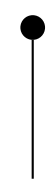
We also ask for a **swap** morphism, that is, for any X and Y objects an isomorphism $\mathbf{swap}_{X,Y} : X \otimes Y \to Y \otimes X$.

Finally, we impose two more things on out category: For each object X we cave assoated a map $\mathbf{copy}_X : X \to X \otimes X$, the copy map and a map $\mathbf{del}_X : X \to I$ called delete or discard. Those are represented in string diagrams by:

$\mathbf{copy}_X$ =



$\mathbf{del}_X$ =



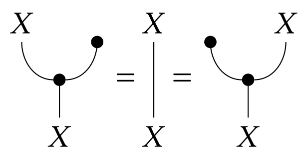
The copy, in the context of **Stoch** and **FinStoch** is the diagonal assignation, that is $\mathbf{copy}_X : X \to X \otimes X$ defined by $\mathbf{copy}_X(x) = (x, x)$.

Now for the $\mathbf{del}_X$ map, the unit I has only the trivial measure 1, hence the delete (or forget) map is sending to the trivial measure, by summing (in the context of **Stoch**) or integrating (in **FinStoch**) the probabilities.

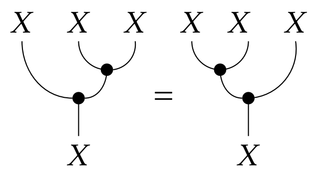
Given now a state (distribution) p on some $X \otimes Y$ we can get the marginal distribution on Y by composing the state with $id_X \otimes \mathbf{del}_Y$, that is, "forgetting" about Y, which gives us a distribution (state) p_X in X.

The maps **copy** and **del** are to satisfy the following axioms:

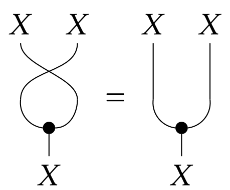
Copying and deleting eitheir one of the "coordinates" is the same as the identity:



Copying either one of the corrdinates after a copy leads to the same result:

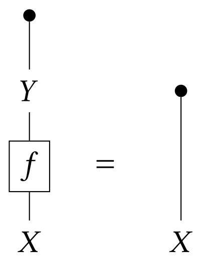


Copying and then swaping is equivalent to copying:



We also enforce this maps to be compatible with the the $(\otimes, I)$ structure we have so far.

The last thing we nees, to properly model our examples, is to ask a condition that is analogue to “the sum of all probabilities is one", called a normalization condition, this is reflected in the last axiom:



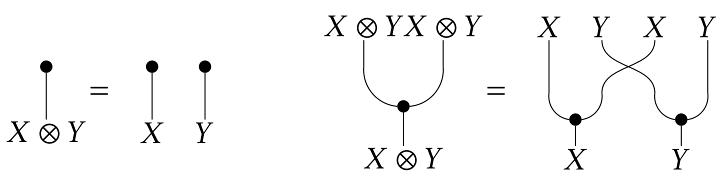
Morphisms that satisfies this axiom are also calle unital morphisms.

Definition 4.1.1. A symmetric monoidal category $(\mathbf{C}, \otimes, I)$ with a distinguished commumtative comonoid structure $(\mathbf{copy}_X, \mathbf{del}_X)$ for each object X, that is compatible with tensor product and for which all morphisms are unital is called a Markov Category.

Follows from our definitions that $\mathbf{del}_X = \mathbf{id}_I$, this together with our last condition (the normalization one, or “naturality of **del**”) adds up to the fact that I is a terminal object in the Markov category, or, put differently, that a Markov category is a semicartesian symmetric monoidal one. The details and some other insights on this equivalence can be consulted in remarks 2.2 - 2.5 in FRITZ, 2020.

In order to get the reader more comfortable with string diagramms, we’ll detail the description of being “compatible” with tensor product in the above definition both in diagramms and with string diagramms.

The compatibility ammounts to two conditions:



Where the first one can be described as the commutativity of the following diagram:

$$\begin{array}{ccccc} X\otimes Y & & \xrightarrow{\mathbf{del}_{X\otimes Y}} & & I \\ \mathbf{id}\downarrow & & & & \downarrow \\ X\otimes Y & \xrightarrow{\mathbf{del}_X\otimes \mathbf{id}} & I\otimes Y & \xrightarrow{\mathbf{id}\otimes\mathbf{del}_Y} & I\otimes I \end{array}$$

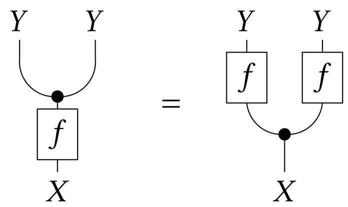
And the second one of the following:

$$\begin{array}{ccccc}
X \otimes Y & & \xrightarrow{\quad \mathbf{copy}_{X \otimes Y} \quad} & & (X \otimes Y) \otimes (X \otimes Y) \\
\Big\downarrow \mathbf{id} & & & & \Big\downarrow \cong \\
& & & & (X \otimes (Y \otimes X)) \otimes Y \\
& & & & \Big\downarrow (\mathbf{id} \otimes \mathbf{swap}) \otimes \mathbf{id} \\
& & & & (X \otimes (X \otimes Y)) \otimes Y \\
& & & & \Big\downarrow \cong \\
X \otimes Y & \xrightarrow{\mathbf{copy}_X \otimes \mathbf{id}} & (X \otimes X) \otimes Y & \xrightarrow{\mathbf{id} \otimes \mathbf{copy}_Y} & (X \otimes X) \otimes (Y \otimes Y)
\end{array}$$

4.2 Deterministic morphisms

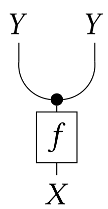
One of the main notions in Markov categories is the one of a "deterministic morphism", those will translate the idea of morphisms with a "randomnnes" on it, follows the formal definition:

Definition 4.2.1. In a markov category $(\mathbf{C}, \otimes, I)$, a morphism $f : X \to Y$ is called deterministic if it commutes with $\mathbf{copy}_X$, in the following sense:

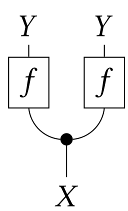


Let us see what a deterministic morphism means in the context of **Stoch** of measurable space and Markov kernels.

The diagram



Translates to the morphism $(\Sigma_Y \otimes \Sigma_Y) \times X \to [0, 1]$ $(S \times T, x) \mapsto f(S \mid x) f(T \mid x)$ and the diagram



translates to the morphism $(S \times T, x) \mapsto f(S \cap T \mid x)$. For those morphisms to be equal we must have $f(S \mid x)^2 = f(S \mid x)$ for all $S \in \Sigma_Y$, which means that the probability measure $f(- \mid x) : \Sigma_Y \to [0, 1]$ is a zero-one valued measure, and it can be seen that this condition is actually enough (see Fritz, 2020, lemma 4.2), hence the deterministic morphisms in

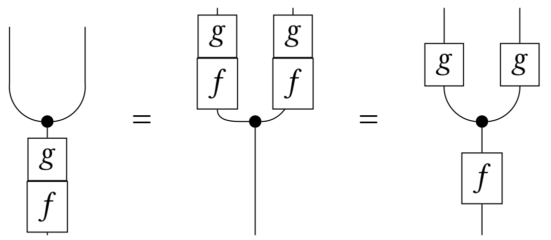
Stoch are exactly the kernels where the associated probability measure is a zero-one measure.

Now for a technical lemma on the behaviour of deterministic morphisms

Lemma 4.1. *Consider $f : X \to Y$ and $g : Y \to Z$ morphisms in a markov category.*

- *If f and g are deterministic then gf is also deterministic.*
- *If gf is detemrinistic and f is a deterministic epimorphisms, then g is deterministic as well.*
- *If f and g are inverse isomorphisms and if one of them is deterministic, the the other one. is also deterministic.*

Proof. The first and last one comes directly from the definition. Fro the second one we note the following computation:



□

The following is an important characterization of categories with only deterministic maps:

Proposition 4.2. *For a Markov category* **C**, *the following are equivalent:*

1. *Every morphism of* **C** *is deterministic.*
2. *The copy structure forms a natural transformation.*
3. *The monoidal structure of* **C** *is cartesian*

The first proof can be consulted in Fritz, 2020 and the second one at Moss and Perrone, 2023.

4.3 The 2-category Markov and a strictification theorem

As we have seen a monoidal category is called strict if the associator and the left and right unitors are both identity natural transformations.

Definition 4.3.1. A Markov category is called strict if its underlying monoidal structure is strict.

Definition 4.3.2. The 2-category **Markov** is the one where objects are Markov categories, 1-morphisms are functors $F : \mathbf{C} \to \mathbf{D}$ where for every $X \in \mathbf{Ob}(\mathbf{C})$ the following diagram commutes:

$$\begin{array}{ccc} & FX & \\ {}^{\mathbf{copy}_{FX}}\swarrow & & \searrow^{F(\mathbf{copy}_X)} \\ FX \otimes FX & \xrightarrow[\cong]{} & F(X \otimes X) \end{array}$$

those morphisms are called Markov functors. The 2-morphisms $\alpha : F \to G$ are monoidal natural transformations where the components are deterministic. An equivalence in this category **Markov** is called a comonoid equivalence. A Markov functor wich is faithfull is called a Markov embedding.

Here a technical characterization of comonoid equivalences, usefull to prove equivalence in some contexts:

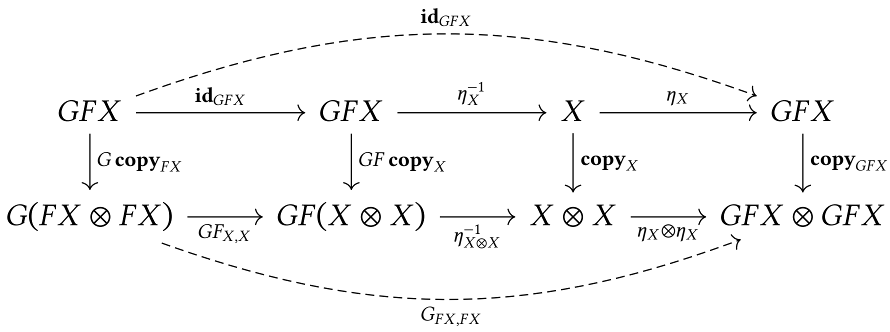
Proposition 4.3. *A symmetric monoidal functor $F : \mathbf{C} \to \mathbf{D}$ is a comonoid equivalence if, and only if, it is an equivalence of categories and for every object Y in the category $\mathbf{D}$ exists an object X in the category $\mathbf{C}$ and a deterministic isomorphism $i : Y \to FX$, thst is, where the following diagram commutes*

$$\begin{array}{ccc} Y & \xrightarrow{i} & FX \\ {\scriptstyle \mathbf{copy}_Y}\downarrow & & \downarrow{\scriptstyle \mathbf{copy}_{FX}} \\ Y \otimes Y & \xrightarrow[i \otimes i]{} & FX \otimes FX \end{array}$$

Proof. The first direction is immediate, we can just take $X = F^{-1}Y$, where F^{-1} is the the inverse equivalence $F^{-1} : \mathbf{D} \to \mathbf{C}$ of F. Now for the converse, for a given $Y \in \mathbf{D}$, we select a X_Y object of $\mathbf{C}$ and a isomoprhism $i : Y \to FX_Y$ that makes the diagram comutes, then those choices of X_Y and i determines an equivalence $G : \mathbf{D} \to \mathbf{C}$ that right adjoint to F via the counit $\epsilon : FG \to \mathbf{id}_{\mathbf{D}}$ and unit $\eta : \mathbf{id}_{\mathbf{C}} \to GF$. By definition η has deterministic components, then we just need to show the same thing for η, which is equivalent to show that, for every X object of $\mathbf{C}$, the following diagram commutes:

$$\begin{array}{ccc} FX & \xrightarrow{F\eta_X} & FGFX \\ {\scriptstyle \mathbf{copy}_{FX}}\downarrow & & \downarrow{\scriptstyle \mathbf{copy}_{FGFX}} \\ FX \otimes FX & \xrightarrow{F\eta_X \otimes F\eta_X} & FGFX \otimes FGFX \\ {\scriptstyle \cong}\downarrow & & \downarrow{\scriptstyle \cong} \\ F(X \otimes X) & \xrightarrow{F(\eta_X \otimes \eta_X)} & F(GFX \otimes GFX) \end{array}$$

The lowe square commuters by naturality and since $F\eta_X$ is the inverse of ϵ_{FX} , which is deterministic, it is also deterministic. Hence the only thing left is to prove that G preserves comonoids, considering an object of the form FX, for X object of $\mathbf{C}$, then in the following diagram



The left square commutes by F being a monoidal functor, the middle one by definition of η and the right one by the fact that η has deterministic components. A general case, considering any object Y of $\mathbf{D}$ can be reduced to the above via the map i. □

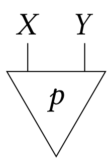
Now for a result that allows us to work only within strict Markov categories and hence prove things using string diagramms, a more detailed proof can be consulted in theorem 10.17 of Fritz, 2020.

Theorem 4.4. *Any Markov category is comonoid equivalent to a strict one.*

Proof. Consider a Markov catecory $\mathbf{C}$ and the subcategory of deterministic morphisms $\mathbf{C}_{\text{det}}$, this subcategory is monoidal cartesian, hence we can choose a strict monoidal category $\mathbf{C}^{'}_{\text{det}}$ and a symmetric monoidal equivalence $F_{\text{det}} : \mathbf{C}^{'}_{\text{det}} \to \mathbf{C}_{\text{det}}$, we extend $\mathbf{C}'_{\text{det}}$ to a category $\mathbf{C}'$ by considering the same objects but the morphisms from X to Y as the morphisms in $\mathbf{C}$ between $F_{\text{det}}X$ and $F_{\text{det}}Y$, hence we get a functor $F : \mathbf{C}' \to C$ extending F_{det}, since $\mathbf{C}^{'}_{\text{det}}$, it follows that $\mathbf{C}'$ inherit a Markov structure, from the fact that F_{det} is symmetric monoidal and F is its extension we see that the criterion of the last proposition is satisfied. □

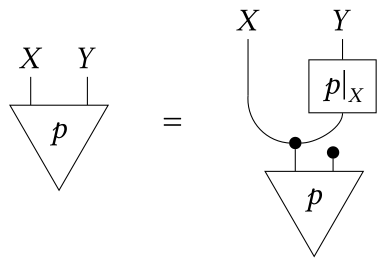
4.4 Joint States and Conditioning

A joint state is simply a state in the joint $X \otimes Y$, that is, a $p : I \to X \otimes Y$, as the diagram below:



Definition 4.4.1. Given a joint state $p : I \to X \otimes Y$ a conditional distribution of p given

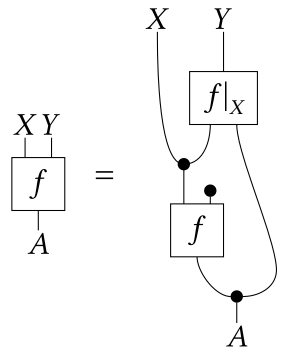
X is a morphism $p|_X : X \to Y$ such that:



Definition 4.4.2. A Markov category is said to have conditional distributions if every joint state $p : I \to X \otimes Y$ has a conditional distribution $p|_X$.

There is just one stepback in the above definition of conditioning, it is not "recursive" in the following sense: If we have a state $p : I \to X \otimes Y \otimes Z$ then we would like to get a conditional $p|_{X \otimes Y} : X \otimes Y \to Z$ with a "composition", first we condition on X and then on Y, but the second conditioning is not a conditioning in the sense of the above definition, since we are only conditioning distributions, hence the following definition:

Definition 4.4.3. Consider **C** a Markov category. We say that this category has conditionals if for every morphism $f : A \to X \otimes Y$ exists $f|_X : X \otimes A \to Y$ with:

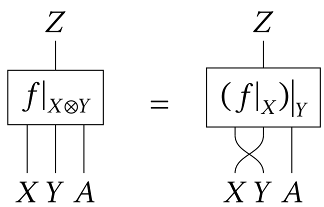


Now for a lemma that justify the above definition, as motivated before, the proof can be consulted in Fritz, 2020, lemma 11.11.

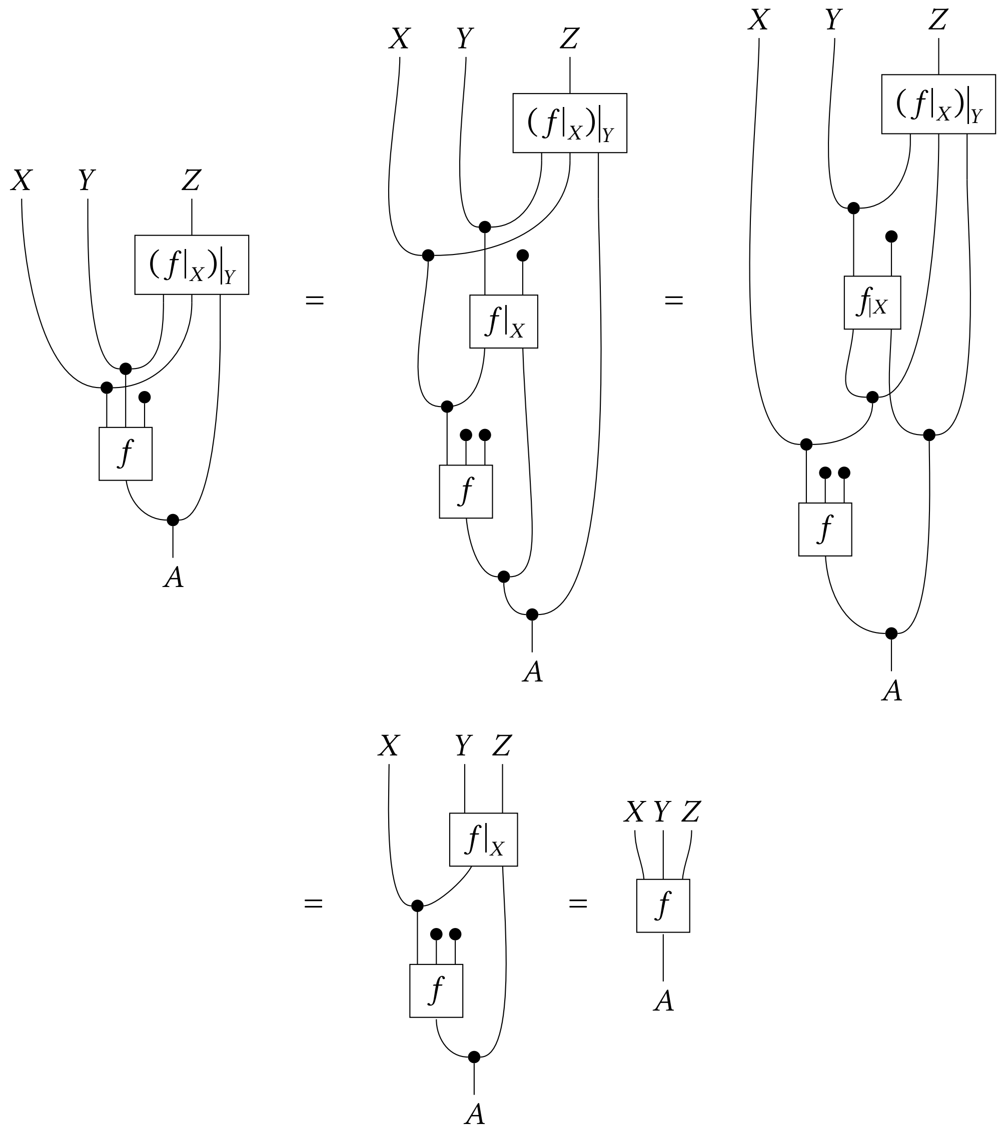
Lemma 4.5. *Consider a Markov category* **C** *with conditionals, then for any morphism* $f : A \to X \otimes Y \otimes Z$ *we have that*

$$f|_{X \otimes Y} = (f|_X)|_Y \circ (\boldsymbol{swap}_{X,Y}\ \boldsymbol{id}_A)$$

Proof. Note that the lemma can be transcribed in string diagrams as:



which can be verified from the diagrams:

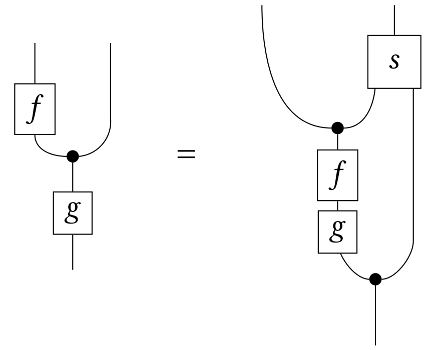


□

So $f|_{X\otimes Y}$ is, up to swapping coordinates, like conditioning first on X and then on Y. Now a usefull result on Markov categories with conditioning, it says that we gave a disintegration property on them, for a proof the reader may consult proposition 11.17 in Fritz, 2020.

Proposition 4.6. *Consider* **C** *a Markov category with conditional. For every* $g : A \to X$

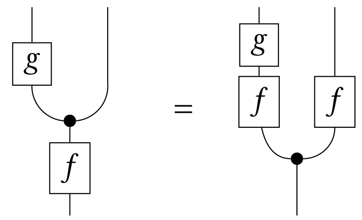
and $f : X \to Y$ exists a $s : A \otimes Y \to X$ such that:



Proof. We just define s to be the consitional of the composite morphism $(f \otimes \mathbf{id}) \circ \mathbf{copy} \circ g$, then the definition of conditionals implies the diagram on the right. □

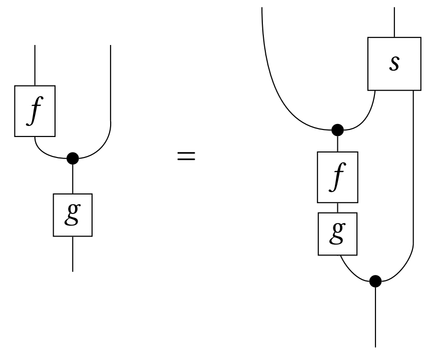
We now focus on the conpect of positivity.

Definition 4.4.4. A Markov category is positive if for all morphisms $f : X \to Y$ and $g : Y \to Z$ with gf deteministic we have the following identity:

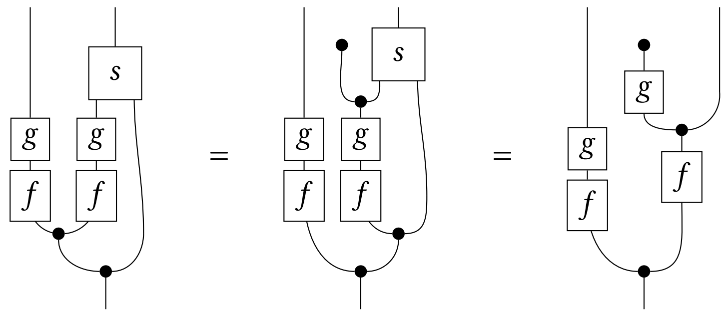


Lemma 4.7. *If a Markov category has conditionals then it is positive.*

Proof. We select s as in 4.6, then we get that:



And using determinism we get



□

For example, consider **Stoch**, to say it is positive is to say that given measurable spaces $(X, \Sigma_X), (Y, \Sigma_Y), (Z, \Sigma_Z)$ and morphisms $f : X \to Y$ and $g : Y \to Z$, then

$$\int_{y \in T} g(S \mid y) f(dy \mid x) = f(T \mid y) \int_{y \in Y} g(S \mid y) f(dy \mid x)$$

for all $S \in \Sigma_Z$, $T \in \Sigma_Y$ and $x \in X$. But since gf is deterministic and as we have observed, deterministic morphisms in **Stoch** satisfies

$$gf(U \mid x) = \int_{y \in Y} g(U \mid y) f(dy \mid x) \in \{0, 1\}$$

for $U \in \Sigma_Y$. We can consider, for simplicity, that gf is zero valued, then the integrand vanishes almost ssurely, hence $y \mapsto f(U \mid y)$ vanishes almost surely, with respect to the measure $f_x = f(- \mid x)$ on all Y, hence both sides of the first equation vanishes, then we get the equality. The we just proved that **Stoch** satisfies the positivity condition hence it is a positive Markov category.

Lemma 4.8. *If* $\mathbf{C}_1 \to \mathbf{C}_2$ *is a Markov embedding and* $\mathbf{C}_2$ *is positive, then* $\mathbf{C}_1$ *is positive.*

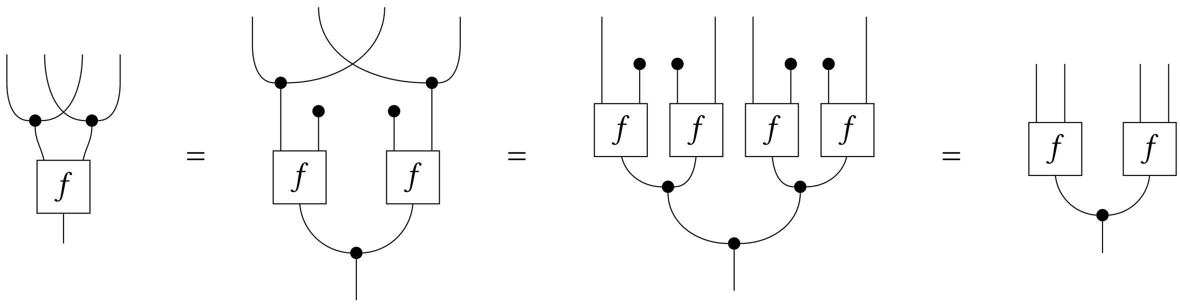
also

Lemma 4.9. *Consider* $\mathbf{C}$ *a positive Markov category, then for a morphism* $f : A \to X \otimes Y$ *the following are equivalent:*

1. *The morphism* f *is deterministic.*

2. *Both of its marginals are deterministic.*

Proof. The first implication, assuming f deterministic comes from the definition itself, since the marginalization morphism is deterministic. Now assuming that both marginal are deterministic, we can check the determinism of f by the following computation:



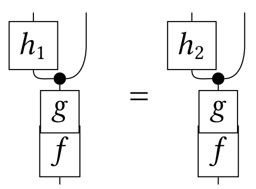
The middle step if from the assumption, the first and last ones comes from the notion of conditional independence and some of its technical features, that can be consulted in proposition 12.14 of Fritz, 2020. □

We know from lemma 4.1 that for two morphisms g and f with gf deterministic and f a deterministic epi we get that g is also deterministic. In the context of positive Markov categories, we can drop the deterministic condition from f. Hence if a composition is deterministic and the first factor is epi, then the second factor is also deterministic. To state the result:

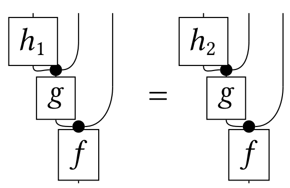
Lemma 4.10. *Consider* $\mathbf{C}$ *a positive Markov category.* $f : X \to Y$ *and* $g : Y \to Z$ *two morphisms in* $\mathbf{C}$*. If* gf *is deterministic and* f *is an epimorphism then* g *is also deterministic.*

Proof. The proof is analogous to the proof of 4.1. □

Definition 4.4.5. A Markov category $\mathbf{C}$ is called casual if: Whenever



holds for all morphisms involved, then

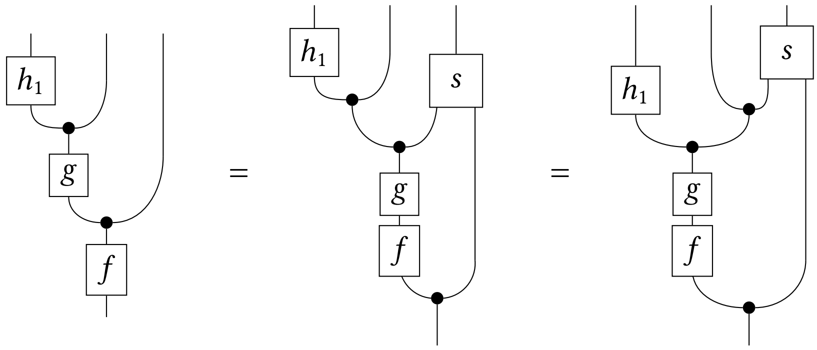


also holds.

Now for the relation between casuality and conditionals:

Proposition 4.11. *If a Markov category* $\mathbf{C}$ *has conditionals then it is casual.*

Proof. Consider the morphism s as in 4.6, then we get

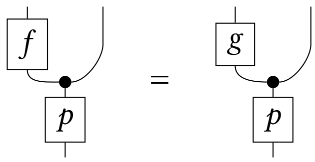


if we exchange, in this diagram, h_1 by h_2 we get that the left diagram is equal with h_2 or h_2 by assumption, hence the equality of the right diagram with h_2 and h_2, which is the definition of a casual Markov category. □

The reverse statement is not true. For example **Stoch** doesn't have conditionals, but it is casual (Fritz, 2020, example 11.35).

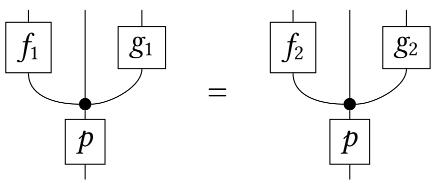
4.5 The notion of almost surely

Definition 4.5.1. Consider **C** a Markov category and a morphism $p : \Theta \to Y$, we define, for two morphisms $f, g : X \to Y$ the notion of p-almost surely equality, denoted by $f =_{p\text{–a.s.}} g$ if



In the category **Cat** p-a.s. equality is to day that $f(T \mid -)$ and $g(T \mid -)$ are almost surely equal with respect to the probability measure $p(- \mid a)$ for all fixed a, S and T.

Lemma 4.12. *Consider* **C** *a Markov category and* $p : \Theta \to X$*. If* $f_1, f_2 : X \to Y$ *and* $g_1, g_2 : X \to Z$ *such that* $f_1 =_{p\text{–a.s.}} f_2$ *and* $g_1 =_{p\text{–a.s.}} g_2$*, then:*



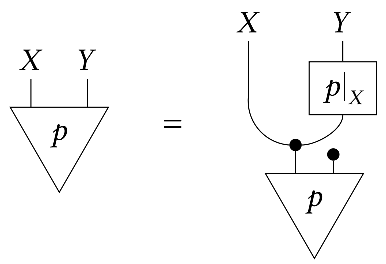
Proof. This comes directly from the associativity and commutativity of the comultiplication and the definition of almost surely above. □

To keep the notation consisend, since we are using $p|_X$ do denote conditioning, for a $p : I \to X \otimes Y$, we will denote the marginal distributions by $p|^m_X$ [1].

[1] In Fritz, 2020, the marginals are denoted by $p(x)$, with a lowercase of the object.

Proposition 4.13. *In a Markov category* **C***, consider a joint distribution* $p : I \to X \otimes Y$*, the conditional* $p|_X : X \to Y$ *is unique* $p|_X^m$*-a.s.*

Proof. It is a direct consequence from the identity



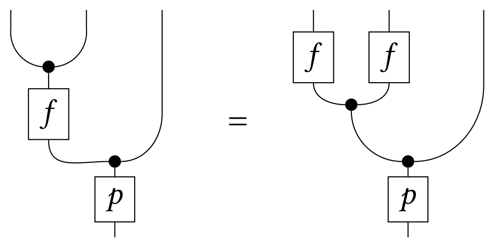
Given by the definition of conditioning. □

Definition 4.5.2. For a Markov category **C**, a **C**-probability space on it is a pair (X, p) where X is an object of **C** and $p : I \to X$ a distribution. When **C** is clear from the context or irrelevant for the result we drop the category name from the notation. For two probability spaces (X, p) and (Y, q) a morphism $f : X \to Y$ is measure preserving if $fp = q$.

Definition 4.5.3. Suppose that a Markov category **C** is casual. The category of probability spaces over **C** and Markov kernels, denoted by **ProbStoch**(**C**) has as objects probability spaces (X, p), (Y, q) and as morphisms maps $f : X \to Y$ which are measure preserving p-a.s. That is, where $fp = q$ modulo p-a.s. equality.

To see that his indeed defines a category the reader can consult Fritz, 2020, proposition 13.9. With the notion of almost surely we can generalize definitions such as what it means to a morphism to be deterministic, such as:

Definition 4.5.4. In a Markov category **C** and a distribution $p : \Theta \to X$, we say that $f : X \to Y$ is p-a.s. deterministic if:



In the catgory **Stoch** a Markov kernel f to be p-a.s. deterministic means that for all $\theta \in \Theta$ we have

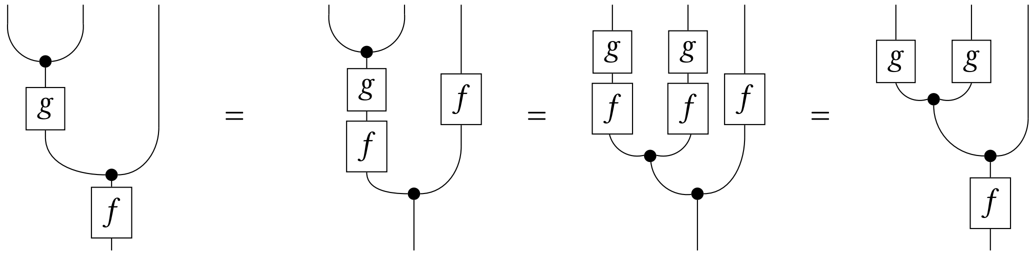
$$\int_{x \in R} f(S \mid x) f(T \mid x) p(dx \mid \theta) = \int_{x \in R} f(S \cap T) p(dx \mid \theta)$$

for all $R \in \Sigma_X$ and $S, T \in \Sigma_Y$. Analogously to the case of deterministic morphisms, Markov kernels satisfying this condition are exactly those where $f(S \mid x) \in \{0, 1\}$ for every $S \in \Sigma_X$ $p(- \mid \theta)$-a.s. for all θ.

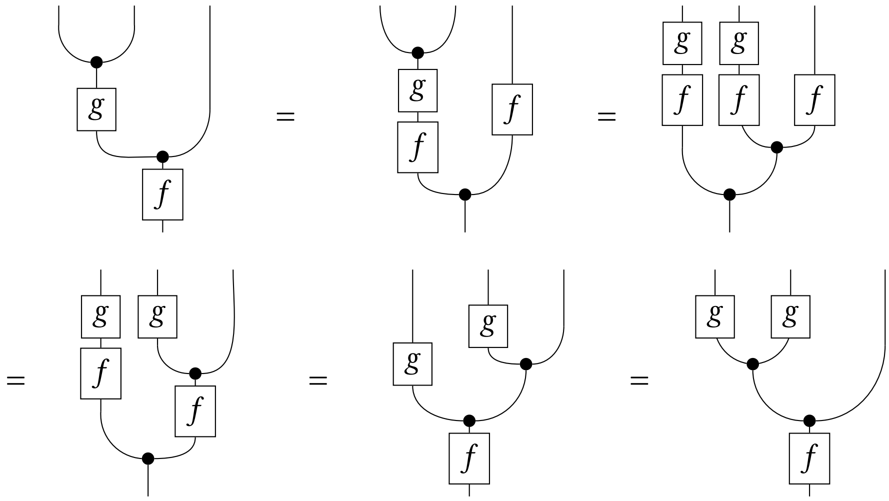
In the same way we have generalized the result from 4.1 for positive Markov categories, we have the same generalization considering almost surely deterministic maps

Proposition 4.14. *Consider, within a Markov category* **C**, *morphisms* $f : X \to Y$ *and* $g : Y \to Z$ *such that* gf *is deterministic. If* f *is deterministic or* **C** *is positive, then* g *is* f*-a.s. deterministic.*

Proof. For the first item, we can observe from the definitions



For the seccond observation we use the definition of positivity and associativity to get:



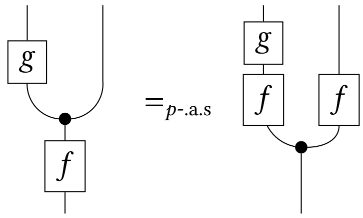
□

A proof can be consulted in Fritz, 2020, proposition 13.5. Note that if we take $gf = id$ and **C** positive, then g is f-a.s. deterministic. Hence, in a positive Markov category, reversable arrows are almost surely deterministic with respect to their inverse.

Definition 4.5.5. We say that a Markov category **C** is strictly positive if morphisms

$$\Theta \xrightarrow{p} X \xrightarrow{f} Y \xrightarrow{g} Z$$

such that gf are p-a.s. deterministic then it is also valid that:



Lemma 4.15. *If a Markov category* **C** *has conditionals then it is strictly positive.*

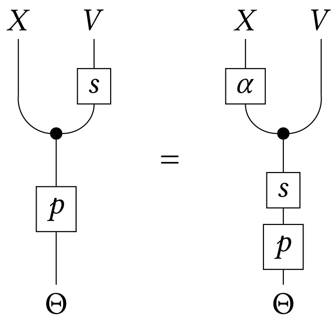
Proof. The proof goes exactly as in 4.7 but now with the aditional distribution p in the bottom of the diagram. □

4.6 Statistics in the categorical probability framework

Definition 4.6.1. Given an object X in a Markov category **C**, called a sample space, a statistical model with values in X consists of a pais (Θ, p) where Θ is an object in **C** and $p : \Theta \to X$.

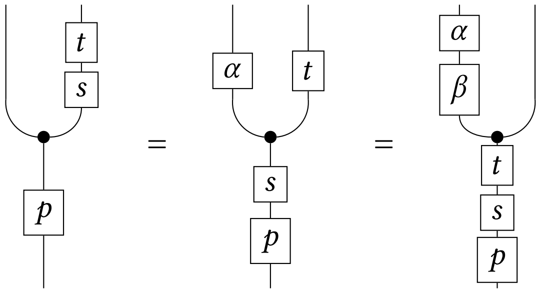
Definition 4.6.2. A statistic for a statistical model $p : \Theta \to X$ is a deterministic morphism s whose domain is s.

Definition 4.6.3. A statistic $s : X \to V$ for a statistical model $p : \Theta \to X$ is sufficient if there is a morphism $\alpha : V \to X$ such that:



Lemma 4.16. *Consider a statistical model $p : \Theta \to X$ in a Markov category ans $s : X \to V$, $t : V \to W$ statistics for it. If s is sufficient for p and t is sufficient for sp, then the composite is also sufficient for p.*

Proof. Consider α the "sufficiency witness" for s with respect to p and β the witness of sufficiency for t with respect to sp. We just need to show that $\alpha\beta$ is a sufficiency witness for ts with respect to p, and this comes from:



□

Chapter 5

Miscellaneous

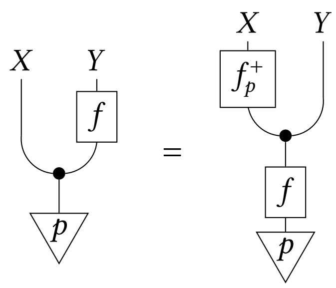
This chapter contains two applications of the theory developed so far, both analogues to result that we have in the classical context. One due to Fritz, 2020, of a version of the Fischer-Neyman factorization theorem, the other one due to Moss and Perrone, 2023, where the notions of a dynamical system and an ergodic decomposition are introduced in the framework of Markov categories and then an ergodic decomposition theorem within this framework is proved. The title Miscellaneous was choosen due to the potential of expanditure of this chapter, new results and interpretations of classical contexts within Markov categories are constantlly being published (such as Perrone, 2023).

5.1 The Fischer-Neyman factorization theorem

Theorem 5.1 (A synthetic Fischer-Neyman factorization theorem)**.** *Consider* $\mathbf{C}$ *a strictly positive Markov category. The a statistic* $s : X \to V$ *is sufficient for a statistical model* $p : \Theta \to X$ *if, and only if, there is a* $\alpha : V \to X$ *such that* $\alpha s p = p$ *and* $s\alpha =_{sp\text{-a.s.}} \boldsymbol{id}_V$

5.2 A synthetic ergodic decomposition theorem

Definition 5.2.1. In a Markov category C, let $p : I \to X$ be a state, and let $f : X \to Y$ be a morphism. A *disintegration of* p via f, or a *Bayesian inversion of* f *with respect to* p is a morphism $f_p^+ : Y \to X$ such that the following holds.



Definition 5.2.2. Let $p : I \to X$ be a state in a Markov category. A decomposition of p

consists is a pair (q, k) where:

- q is a state $q : I \to Y$,
- k is a morphism $k : Y \to X$

such that $k \circ q = p$.

Definition 5.2.3. Let $(q : I \to Y, k : Y \to X)$ be a decomposition of $p : I \to X$. We say that (q, k) is a *trivial* decomposition of p if and only if k is q-almost surely equal to

$$Y \xrightarrow{\textbf{del}} I \xrightarrow{p} X.$$

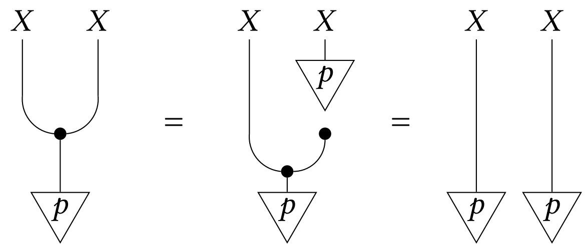
We call the state p indecomposable if all its decompositions are trivial.

Proposition 5.2. *Every indecomposable state is deterministic.*

Proof. Let $p : I \to X$ be an indecomposable state. We can decompose p as $p = id \circ p$, and by hypothesis this decomposition is trivial. Therefore, the identity id is p-almost surely equal to

$$X \xrightarrow{\textbf{del}} I \xrightarrow{p} X.$$

That is,



Hence p is deterministic. □

Definition 5.2.4. Given a monoid M, we will denote by $\mathbf{B}M$ the category associated with it (with a single arrow and morphisms given by M), for a category $\mathbf{C}$ a dynamical system is a functor $\mathbf{B}M \to \mathbf{C}$. We say that the image of the one object of $\mathbf{B}M$ in $\mathbf{C}$ is a dynamical system in $\mathbf{C}$ with monoid M.

For example, taking $M = \mathbb{N}$ and $\mathbf{C} = \mathbf{Stoch}$ we get that A dynamical system is a discrete time Morkov chain.

Definition 5.2.5. Let X be a dynamical system with monoid M in a Markov category $\mathbf{C}$. The Markov colimit or Markov quotient of X over the action of M is an object, which we denote by X_{inv} together with a deterministic map $r : X \to X_{inv}$, which is a colimit both in $\mathbf{C}$ and in the subcategory $\mathbf{C}_{det}$ of deterministic morphisms.

Definition 5.2.6. Consider X a dynamical system in **Stoch** with monoid M. A measurable set $A \in \Sigma_X$ is called invariant if for every $m \in M$, it is associated Markov kernel k_m is such that $m(A \mid x) = \delta_x(A)$.

Proposition 5.3. *Let X be a deterministic dynamical system*[1] *in* ***Stoch*** *with monoid M.*

[1] In the sense that it is a dynamical system in the subcategory $\mathbf{C}_{det}$.

Then the Markov quotient X_{inv} exists, and it is given by the same set X, equipped with the invariant σ-algebra.

Proof. We construct the kernel $r : X \to X_{inv}$ as follows,

$$r(A|x) = 1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

for every $x \in S$ and every measurable (invariant) set $A \in \Sigma_{X_inv}$. This is the kernel induced by the function $X \to X_{inv}$ considering the set-theoretic identity. This function is measurable, and so it induces a well-defined Markov kernel. For every $m \in M$, every $x \in X$ and every measurable (invariant) set $A \in \Sigma_{X_{inv}}$,

$$\int_X r(A|x')\, m(dx'|x) = \int_X 1_A(x')\, m(dx'|x) = m(A|x) = 1_A(x) = r(A|x),$$

where we used invariance of A. Hence, r is left-invariant.

Let $s : X \to S$ be a right-invariant Markov kernel. Define the kernel $\tilde{s} : X_{inv} \to S$ simply by

$$\tilde{s}(B|x) := s(B|x)$$

for all $x \in X_inv$ (equivalently, $x \in X$) and all measurable $B \subseteq S$. To see that $\tilde{s}$ is measurable in x, consider a Borel-generating interval $(r, 1] \subseteq [0, 1]$ for some $0 \leq r < 1$. We have to prove that the set

$$\tilde{s}^*(B, r) := \{x \in X_{inv} : s(B|x) > r\}$$

is measurable in X_{inv}, i.e., as a subset of X, is measurable and invariant. We know that it is measurable as a subset of X, since s is a Markov kernel. Let us prove the invarianc property: Using the fact that s is right-invariant, and that $\tilde{s}^*(B, r)$ is measurable as a subset of X,

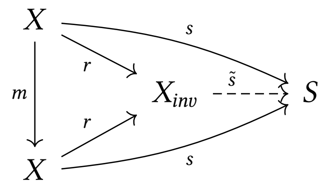
$$\begin{aligned} s(B|x) &= \int_X s(B|x')\, m(dx'|x) \\ &= \int_{\tilde{s}^*(B,r)} s(B|x')\, m(dx'|x) + \int_{X\setminus\tilde{s}^*(B,r)} s(B|x')\, m(dx'|x) \end{aligned}$$

Now let us use that m is deterministic, so that either we have $m(\tilde{s}^*(B, r)|x) = 1$, or $m(X \setminus \tilde{s}^*(B, r)|x) = 1$. In the first case, we have that $s(B|x') > r$ on a set of measure 1, therefore $s(B|x) > r$, i.e. $x \in \tilde{s}^*(B, r)$. In the latter case, $s(B|x') \leq r$ on a set of measure 1, and therefore $s(B|x) \leq r$, i.e. $x \notin \tilde{s}^*(B, r)$. Since m is deterministic, these are the only possibilities, and so we have that

$$m(\tilde{s}^*(B, r)|x) = \begin{cases} 1 & x \in \tilde{s}^*(B, r) \\ 0 & x \notin \tilde{s}^*(B, r). \end{cases}$$

This means precisely that $\tilde{s}^*(B, r)$ is invariant, and so $\tilde{s}$ is measurable. Therefore, $\tilde{s}$ is a well-defined Markov kernel $X_{inv} \to S$.

For uniqueness, note that $\tilde{s}$ is the only possible choice of kernel $X_{inv} \to S$ making



commute: let $k : X_{inv} \to S$ be another such kernel. Then for all $x \in X$ and every measurable $B \subseteq X$,

$$k(B|x) = \int_{X_{inv}} k(B|x')\,\delta(dx'|x) = \int_{X_{inv}} k(B|x')\,r(dx'|x) = s(B|x).$$

Moreover, by construction, $\tilde{s}$ is deterministic if and only if s is. Hence the derired universal property of X_{inv}. □

Definition 5.2.7. Let X be a dynamical system with monoid M in a Markov category $\mathbf{C}$. An invariant state $p : I \to X$ is *ergodic* if for every invariant deterministic observable $c : X \to R$, the composition $c \circ p$ is deterministic.

Definition 5.2.8. Let X be a dynamical system with monoid M in a Markov category $\mathbf{C}$ and Y an object of C with a state $q : I \to Y$ and a morphism $k : Y \to X$. We say that k is q-almost surely ergodic if

- k is q-almost surely invariant, and
- whenever $r : X \to R$ is invariant and deterministic (not just almost surely), then $r \circ k$ is q-almost surely deterministic.

Proposition 5.4. *Let X be a dynamical system with monoid M in a Markov category $\mathbf{C}$, and suppose that the Markov colimit X_{inv} of X exists. Let Y be an object of C with a state $q : I \to Y$ and a q-almost surely invariant morphism $k : Y \to X$. Then k is q-almost surely ergodic if and only if the composition with the universal cocone*

$$Y \xrightarrow{k} X \xrightarrow{r} X_{inv}$$

is q-almost surely deterministic.

Proof. We prove the following result first: An invariant state $p : I \to X$ is ergodic if and only if the composition with the universal cocone

$$I \xrightarrow{p} X \xrightarrow{r} X_{inv}$$

is deterministic.

First, suppose that the composite $r \circ p$ is deterministic. Let $c : X \to R$ be an invariant deterministic observable. By definition of Markov colimit, c factors (uniquely) as a composite $\tilde{c} \circ r$, where $\tilde{c} : X_{inv} \to R$ is deterministic. Therefore

$$c \circ p = \tilde{c} \circ r \circ p = \tilde{c} \circ (r \circ p)$$

is a composite of deterministic maps, and hence is deterministic. This is true for every invariant deterministic c, and so p is ergodic.

The converse follows by taking c in the definition of ergodicity to be $r : X \to X_{inv}$, which is deterministic and invariant.

For the result stated, it is a direct application for the one proved above. □

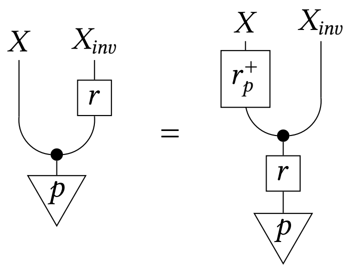
Theorem 5.5 (Synthetic ergodic decomposition theorem)**.** *Let* **C** *be a Markov category. Let X be a deterministic dynamical system in C with monoid M. Suppose that*

- *X has disintegrations;*
- *The Markov colimit X_{inv} of the dynamical system exists.*

Then every invariant state of X can be written as a composition $k \circ q$ such that k is q-almost surely ergodic.

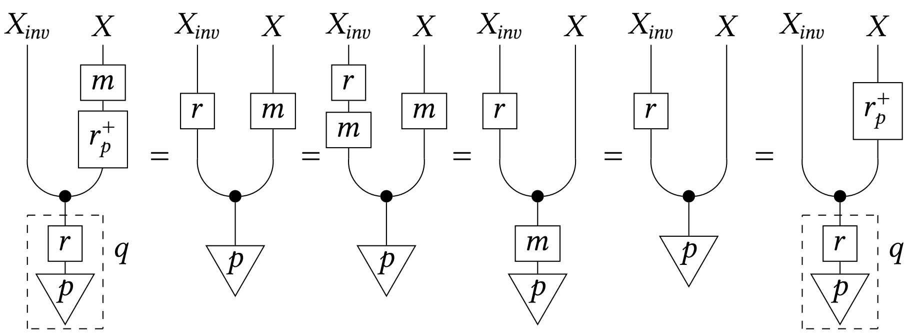
The proof goes exactly as in Moss and Perrone, 2023:

Proof. Let $p : I \to X$ be an invariant state. Consider the map $r : X \to X_{inv}$, and form the disintegration $r_p^+ : X_{inv} \to X$.



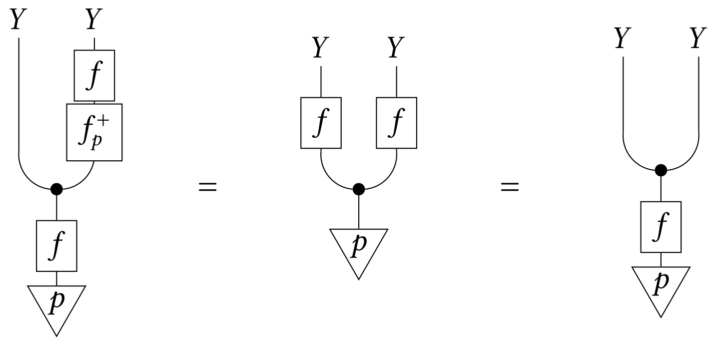
By marginalizing the equation above over X_{inv}, we see that $p = r_p^+ \circ r \circ p$, i.e. we are decomposing p into the composition of $r \circ p : I \to X_{inv}$ followed by r_p^+. Now denote $r \circ p$ by q.

Let us show that r_p^+ is q-almost surely ergodic. To see that r_p^+ is q-almost surely left-invariant, note that for all $m \in M$,



using, in order, the definition of r_p^+ as a disintegration, right-invariance of r, determinism of m, left-invariance of p, and again the definition of r_p^+ as a disintegration.

By 5.4, all that remains to be shown in order to prove q-almost sure ergodicity is that $r \circ r_p^+$ is q-almost surely deterministic. To see this, note that since r is deterministic, we can apply



with r in place of f. The proposition tells us that $r \circ r_p^+$ is q-almost surely equal to the identity, which is deterministic. □

Bibliography

[Billingsley 1986] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1986 (cit. on pp. 2, 3).

[Bogdan 1976] Victor M. Bogdan. “A new approach to the theory of probability via algebraic categories”. In: *Measure Theory*. Ed. by Alexandra Bellow and Dietrich Kölzow. Berlin, Heidelberg: Springer Berlin Heidelberg, 1976, pp. 345–367. isbn: 978-3-540-38107-5.

[Borceux 1994] F. Borceux. *Handbook of Categorical Algebra: Volume 1, Basic Category Theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1994 (cit. on p. 17).

[Cho and Jacobs 2019] Kenta Cho and Bart Jacobs. “Disintegration and bayesian inversion via string diagrams”. *Mathematical Structures in Computer Science* 29.7 (2019), pp. 938–971 (cit. on p. 1).

[Fritz 2020] Tobias Fritz. “A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics”. *Advances in Mathematics* 370 (2020), p. 107239 (cit. on pp. 1, 2, 32, 41, 45–47, 49–51, 54–57, 61).

[Fritz and Perrone 2018] Tobias Fritz and Paolo Perrone. “Bimonoidal structure of probability monads”. *Electronic Notes in Theoretical Computer Science* 341 (2018), pp. 121–149 (cit. on pp. 31, 37).

[Giry 1982] Michèle Giry. “A categorical approach to probability theory”. In: *Categorical Aspects of Topology and Analysis*. Ed. by B. Banaschewski. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 68–85. isbn: 978-3-540-39041-1 (cit. on pp. 1, 2, 31, 35).

[Johnstone 2014] P.T. Johnstone. *Topos Theory*. Dover Books on Mathematics. Dover Publications, 2014.

[Lane 1998] S.M. Lane. *Categories for the Working Mathematician*. Graduate Texts in Mathematics. Springer, 1998 (cit. on pp. 2, 17, 31, 36).

[Lawvere 1962] William Lawvere. "The category of probabilistic mappings". Available at https://ncatlab.org/nlab/files/lawvereprobability1962.pdf. 1962 (cit. on pp. 1, 2, 31).

[Marmolejo and Wood 2010] Francisco Marmolejo and R. J. Wood. "Monads as extension systems |no iteration is necessary". In: 2010 (cit. on pp. 31, 33).

[Moss and Perrone 2023] Sean Moss and Paolo Perrone. "A category-theoretic proof of the ergodic decomposition theorem". *Ergodic Theory and Dynamical Systems* (2023), pp. 1–27 (cit. on pp. 2, 41, 47, 61, 65).

[Panangaden 2009] Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009 (cit. on pp. 2, 3, 32).

[Perrone 2018] Paolo Perrone. "Categorical Probability and Stochastic Dominance in Metric Spaces". Available at https://paoloperrone.org/phdthesis.pdf. PhD thesis. Leipzig, Germany: University of Leipzig, 2018 (cit. on pp. 2, 31, 32, 35, 37–39).

[Perrone 2021] Paolo Perrone. *Notes on Category Theory with examples from basic mathematics*. Available at https://arxiv.org/abs/1912.10642. 2021 (cit. on pp. 2, 17).

[Perrone 2023] Paolo Perrone. *Markov Categories and Entropy*. 2023. arXiv: 2212.11719 (cit. on pp. 41, 61).

[Taylor 1973] S.J. Taylor. *Introduction to Measure and Integration*. Cambridge University Press, 1973 (cit. on pp. 2, 3).

[Varadhan 2001] S.R.S. Varadhan. *Probability Theory*. Vol. 7. Courant Lecture Notes. Courant Institute of Mathematical Sciences, 2001 (cit. on pp. 2, 3).

[Voevodsky 2004] Vladimir Voevodsky. "A categorical approach to the probability theory". Available at https://www.math.ias.edu/Voevodsky/files/. 2004 (cit. on pp. 31, 32).