Research Statement: **Partiality and Stochastics in Categorical Probability**

Areeb Shah Mohammed

1 Introduction

Probability theory is a branch of mathematics that has had a profound impact on many fields of study, including statistics, computer science, and physics. At a technical level, modern probability is typically constructed through measure theory. This is rigorous and powerful, allowing one to apply the tools of analysis (to great success), but this analytic back-end can obscure the underlying intuition.

My research in categorical probability applies the tools of category theory to probability theory, with the goal of providing intuitive and flexible tools for understanding probabilistic systems. Category theory is a branch of mathematics that focuses on the relationships between mathematical structures, rather than the structures themselves. For something like a statistical model, this means taking as primary the information/causal flow between variables, rather than just the final probabilities of events. This is closer to how practitioners reason about their models, and the categorical setting makes this precise. Additionally, the categorical setting allows us to treat other probability-like structures (such as nondeterminism, or "possibility") in a unified way, by considering different examples of the same kind of category

I am to develop categorical tools to aid both application of probability to statistical modelling, as well as to deal with probabilistic constructions categorically. The thrust of my work has followed three main directions:

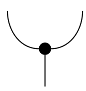
- I have worked on formalizing stochastic processes such as Markov chains and hidden Markov models in a categorical setting, as well as developing tools to perform Bayesian filtering and smoothing in a categorical framework [FKM$^+$25].
- I have worked on developing a categorical formalism for partially defined stochastic maps, which can be used to describe processes with nondeterministic outcomes that may also fail to return a result on certain inputs [SM25].
- I have worked on developing a categorical formalism for the process of constructing "relative frequencies" (typically called empirical measures) from a sequence of data points. Additionally, one can prove a categorical version of the law of large numbers (and associated results like the Glivenko–Cantelli theorem), a central result in probability theory which is key to recovering a distribution from data distributed according to it [FGL$^+$25].

Categorical probability has of late seen growing interest from researchers and practitioners of a variety of backgrounds, and I have been fortunate to have been able to collaborate with mathematicians, computer scientists, and physicists on various aspects of this work so far. A particular focus of mine has been to unify phenomena from different domains, and aid the transfer of techniques and understanding between them.

1.1 The categories of interest, informally

Markov categories (and the more general CD categories) are a framework for probability theory in a categorical setting. Much of the foundational material on categorical probability via Markov categories revolves around the idea that the main concepts of probability theory, such as statistical (in)dependence, determinism, conditioning, etc., can be meaningfully extended from categories of Markov kernels to more general Markov or CD categories.

Definition 1.1 ([CJ19, Gad96])**.** A **CD category** is a symmetric monoidal category C in which every object X is equipped with a distinguished commutative comonoid structure $\mathrm{copy}_X \colon X \to X \otimes X$, and $\mathrm{del}_X \colon X \to \mathcal{I}$, denoted in string diagram notation by

 and 

that is suitably compatible with the tensor product, and such that $\mathrm{copy}_{\mathcal{I}} = \mathrm{del}_{\mathcal{I}} = \mathsf{id}_{\mathcal{I}}$.

1. A map $f \colon X \to Y$ is **total** if it commutes with deletion; that is, if $\mathrm{del}_X\ f = \mathrm{del}_Y$. These are intuitively the maps that are defined everywhere.

2. A **Markov category** is a CD category in which every map is total.

3. A map $f \colon X \to Y$ is **copyable** if it commutes with copying; that is, if $\mathrm{copy}_Y\ f = (f \otimes f)\ \mathrm{copy}_X$. These are intuitively the maps that have no randomness.

4. A map is **deterministic** if it is both total and copyable. In practice, these tend to correspond to some kind of (measurable) function.

Typical examples of Markov categories include categories of measurable spaces and Markov kernels, such as the category $\mathsf{BorelStoch}$ of standard Borel spaces and Markov kernels, the category $\mathsf{FinStoch}$ of finite sets and stochastic matrices, and the category Gauss of Euclidean spaces and affine maps with Gaussian noise. Additionally, there are categories where the maps are "possibilistic", such as the category $\mathsf{SetMulti}$ of sets and multi-valued functions.

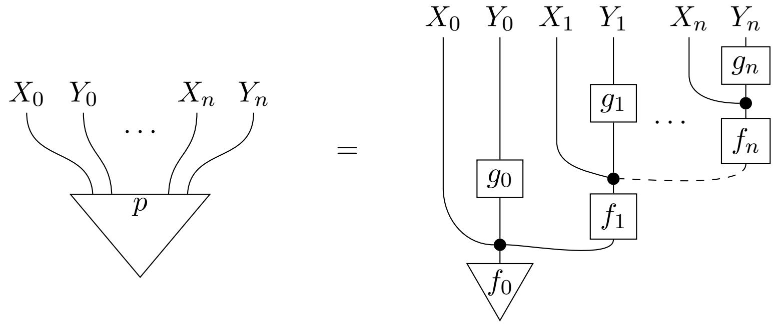
2 My work so far

2.1 What structure makes filtering algorithms work?

Hidden Markov models are statistical models used widely in various fields such as sensor estimation, optimal control, and decision theory. A hidden Markov model can be seen as modelling the evolution of a system in terms of a Markov chain of "hidden states", with only noisy observations being available at each time step. A natural problem when dealing with hidden Markov models is that of "Bayesian estimation" or "filtering", which is to estimate the current state of the "hidden" Markov chain given (the history of) the observations.

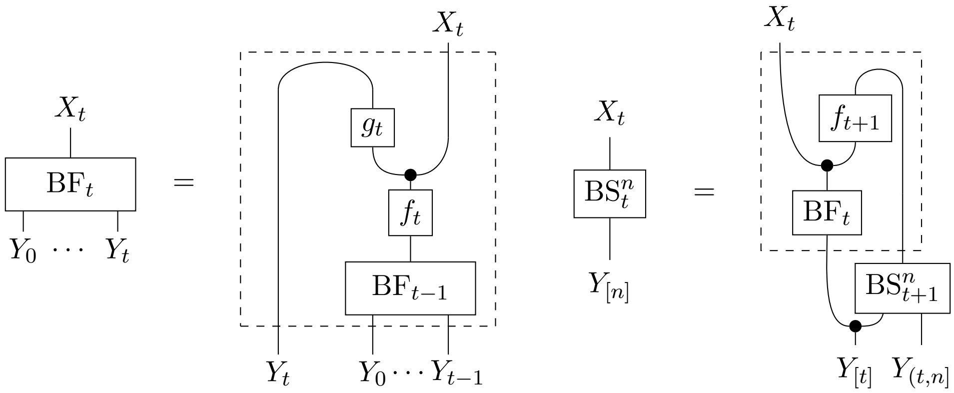
In [FKM+25], we developed the theory of hidden Markov models and the Bayes filter for Markov categories with conditionals in general. This has the advantage of simultaneously applying to discrete, continuous, and Gaussian probability. In addition, it also handles "nonstandard" settings such as possibilistic uncertainty, or probability with an additional measurable dependence on an external parameter. While in the classical measure-theoretic setting, our definition corresponds to the familiar notion of hidden Markov model, in the possibilistic setting it specializes to a form of non-deterministic state machine. Furthermore it natively integrates the two main ways of thinking about such models, namely as graphical models or as joint distributions satisfying certain conditional independence conditions.

Definition 2.1. A state $p\colon \mathcal{I} \to \bigotimes_{t\in[n]}(X_t \otimes Y_t)$ is an n-**step hidden Markov model** if there exists a state $f_0\colon \mathcal{I} \to X_0$ and sequences of maps $f_t\colon X_{t-1} \to X_t$ (the Markov chain transitions) and $g_t\colon X_t \to Y_t$ (the observation/emission kernels) such that



Bayesian filtering and smoothing is about computing certain conditionals.

Theorem 2.2. *One can derive recurrence relations for the Bayes filter* BF_t *and smoother* BS^n_t*, such as: (if a dashed box denotes the conditional on the wires bent inside it)*



Examples 2.3. Our work provides a unified approach to these problems, and our categorical results specialize to classic constructions that enjoy widespread use. In particular:

- Our recursive algorithm for filtering specializes in BorelStoch to the classical prediction–update or "forward" algorithm for Bayesian filtering.
- In Gauss, one gets the fast Kálmán filter.
- Our smoothing algorithms reduce to the classical forward–backward algorithm in BorelStoch; and to the faster Rauch–Tung–Striebel smoother in Gauss.
- Additionally, we also recover the fixed-interval smoothing algorithm, which is of more use practically when smoothing at times far from the end of the data stream.

This was implemented in [SW24] as a proof of concept in C++, formalizing Markov categories as abstract classes and implementing our constructions and algorithms in terms of those.

2.2 Why do relative frequencies work?

Many operations one performs in probability theory involve constructions such as limits or integrals that do not always exist. Even taking the average of a sequence (X_i) of random variables or taking an expectation is not defined in general.

A class of cornerstone results in probability theory involving partially defined operations are the laws of large numbers. Conceptually, they provide self-consistency to the idea of a probability measure: Although the actual expectation of a function under an unknown distribution cannot be inferred from a finite sequence of samples from it, the average of its values on the first n-samples converges with probability 1 to this expectation as the number of observations grows. This is why in practice, given a "large" (but finite!) number of samples, one can reasonably approximate its expectation by the limit of the average values.

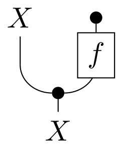
In particular, let f be the indicator function of a measurable set $T \subseteq X$. We would then have that the relative frequency of the event T in a sequence $(x_i)_{i\in\mathbb{N}}$ converges almost surely to its probability:

$$\lim_{n\to\infty} \frac{|\{i \leq n \mid x_i \in T\}|}{n} =_{\mu\text{-a.s.}} \mu(T).$$

As T varies, one can interpret the fraction on the left as the probability of T under the **empirical measure** of the first n elements of the sequence. However the conditions of the convergence of these empirical measures to μ are subtle and in particular can only be expected to hold for some of the sequences (x_i).

In [FGL$^+$25], we formalized this in a categorical setting by developing a framework for an **empirical sampling** *map*, which would intuitively take a sequence of points and return a (random) sample from the empirical measure of the sequence. However due to the lack of convergence in general, this map is only "partially" defined, in the sense of being a map in a CD category that is not total.

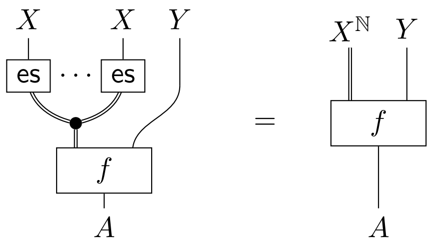
To do this, we introduced the notion of **quasi-Markov categories**, which are intuitively the CD categories where each map is total on its "domain", and simply undefined outside it. Formally, a quasi-Markov category is a CD category such that every map satisfies $f\,\mathrm{dom}(f) = f$, where the domain $\mathrm{dom}(f)$ of a map $f\colon X \to Y$ is the map



Definition 2.4. An **empirical sampling map** for X is a map $\mathsf{es}\colon X^{\mathbb{N}} \to X$ satisfying:

1. **Permutation invariance**: For every finite permutation σ of $\mathbb{N}$, es is invariant under pre-composition with the corresponding permutation (acting on the indices) $X^\sigma\colon X^{\mathbb{N}} \to X^{\mathbb{N}}$; that is, $\mathsf{es}\,X^\sigma = \mathsf{es}$.

2. **Empirical adequacy**: If a map $f\colon A \to X^{\mathbb{N}} \otimes Y$ is exchangeable in the first factor (that is, $(X^\sigma \otimes \mathsf{id}_Y)\, f = f$ holds for every finite permutation σ), then we have[1]



We also constructed such empirical sampling maps in a suitable category of standard Borel spaces and partially defined Markov kernels between them. Given empirical sampling maps and a few other assumptions on a quasi-Markov category:

- A **representable** Markov category has a coherent assignment of objects PX of "distributions on X", as well as "sampling maps" $\mathsf{samp}_X\colon PX \to X$ that intuitively draw random samples from a distribution. We constructed distribution objects such that repeated sampling distinguishes distributions.

- From this, we deduced a version of the **de Finetti theorem**, showing in this setting that exchangeable sequences are precisely the mixtures of IID measures.

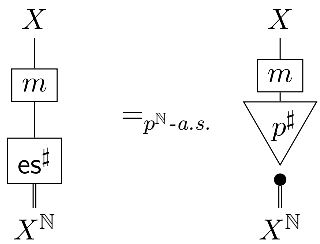
We then showed a categorical version of the Glivenko–Cantelli theorem, which states that the empirical measures of an IID sequence converge almost surely to the underlying distribution.

[1] We use doubled wires to indicate that the corresponding object is an infinite tensor product, as in the first output of f.

Theorem 2.5 (Synthetic Glivenko–Cantelli theorem)**.** *Consider an arbitrary exchangeable map $f\colon A \to X^{\mathbb{N}}$ and an arbitrary map $p\colon A \to X$. Then,*

From this, we derived the law of large numbers in the following form, which for a suitable choice of es reduces to the usual strong law of large numbers.

Corollary 2.6 (Synthetic strong law of large numbers)**.** *For arbitrary maps $p\colon I \to X$ and $m\colon PX \to X$,*



2.3 How do we compose partially defined stochastic operations?

To produce categories with the properties required by [FGL+25], I developed in the companion work [SM25] a **construction of a quasi-Markov category** Partial(C) of "partially defined stochastic maps" from a suitable Markov category C. This adapts the construction of partial maps in terms of spans with monic left legs. The novelty here is that the monoidal structures classically considered are (some variant of) cartesian products, while in the context of categorical probability this cartesian behavior corresponds to determinism (and cannot be assumed to hold for "stochastic" maps). Furthermore, I showed that many of the properties of use in the theory of Markov categories transfer from a partializable Markov category C to its partialization Partial(C).

Definition 2.7. We call a Markov category C **partializable** if:

1. It is positive (which also implies that all isomorphisms are deterministic);
2. Pullbacks of deterministic monomorphisms exist and are themselves deterministic;
3. Deterministic monomorphisms are closed under tensoring.

Definition 2.8. Given a partializable Markov category C, the quasi-Markov category of "partial maps in C", the **partialization** Partial(C) has:

1. Objects those of the original category $\mathcal{C}$;
2. Maps $X \to Y$ equivalence classes of spans

   $$X \overset{i}{\hookleftarrow} D \overset{f}{\to} Y$$

   with i a *deterministic* monomorphism;
3. Composition is done by pullback: For maps represented by spans $X \overset{i}{\hookleftarrow} D_f \overset{f}{\to} Y$ and $Y \overset{j}{\hookleftarrow} D_g \overset{g}{\to} Z$,

   $$\begin{array}{ccccccc} & & & E & & & \\ & & {\scriptstyle u}\swarrow & \lrcorner & \searrow{\scriptstyle v} & & \\ & D_f & & & & D_g & \\ {\scriptstyle i}\swarrow & & \searrow{\scriptstyle f} & & \swarrow{\scriptstyle j} & & \searrow{\scriptstyle g} \\ X & & & Y & & & Z \end{array}$$

   the composite is represented by $X \overset{iu}{\hookleftarrow} E \overset{gv}{\longrightarrow} Z$;
4. Tensoring is done componentwise: For maps $X \overset{i}{\hookleftarrow} D_f \overset{f}{\to} Y$ and $X' \overset{j}{\hookleftarrow} D_g \overset{g}{\to} Y'$, the tensor is represented by

   $$X \otimes X' \overset{i\otimes j}{\hookleftarrow} D_f \otimes D_g \overset{f\otimes g}{\longrightarrow} Y \otimes Y'$$
5. The CD structure is inherited from C.

The main example of a partializable Markov category is the category $\mathsf{BorelStoch}$ of standard Borel spaces and stochastic maps. The maps $X \to Y$ in $\mathsf{Partial}(\mathsf{BorelStoch})$ can be identified with stochastic maps $D \to Y$ for a measurable $D \subseteq X$, capturing the intuition of "partially defined stochastic maps". Other examples include variants such as the category $\mathsf{FinStoch}$ of finite probability spaces and stochastic maps, Dist and its variants of discrete measurable spaces and finitely supported stochastic maps in a suitable semiring, or the category $\mathsf{SetMulti}$ of sets and multivalued maps.

Theorems 2.9. *For a partializable Markov category* C*:*

1. *The category* $\mathsf{Partial}(\mathsf{C})$ *is a well defined CD category, and is quasi-Markov.*
2. $\mathsf{Partial}(\mathsf{C})$ *is a restriction category, and the poset enrichment of* $\mathsf{Partial}(\mathsf{C})$ *given by maps of spans has an equivalent description solely in terms of the CD structure.*
3. $\mathsf{Partial}(\mathsf{C})$ *is positive.*
4. *A map* $X \hookleftarrow D \overset{f}{\to} Y$ *of* $\mathsf{Partial}(\mathsf{C})$ *is copyable if and only if* f *is deterministic.*
5. $\mathsf{Partial}(\mathsf{C})$ *has conditionals if* C *does, and they are in a sense defined on the maximal possible domain.*

6. $\mathsf{Partial}(\mathsf{C})$ *has Kolmogorov products of size* K *if* C *does.*[2]

7. *Idempotents in* $\mathsf{Partial}(\mathsf{C})$ *correspond to idempotents on their domains in* C. *This correspondence restricts to a correspondence between split, balanced, strong, and static idempotents [FGL⁺23, Definition 4.1.1].*

In particular, the partialization $\mathsf{Partial}(\mathsf{BorelStoch})$ *of the category of standard Borel spaces (with the empirical sampling maps constructed in [FGL⁺25]) satisfies the requirements of the categorical Glivenko–Cantelli theorem and law of large numbers (Theorem 2.5 and Corollary 2.6).*

When C is representable, I showed that the distribution objects and sampling maps of C serve as distribution objects and sampling maps for $\mathsf{Partial}(\mathsf{C})$ as well. Consequently, the distribution functor defines a monad on the subcategory of copyable maps of $\mathsf{Partial}(\mathsf{C})$, and thus we can make sense of **partial algebras** for this monad. For instance, on standard Borel spaces, the "expectation map" assigning to a distribution p on $\mathbb{R}_{\geq 0}$ its expectation $\int x\, p(dx)$ is a partial algebra for the Giry monad, with domain the distributions p such that $\int x\, p(dx)$ is finite.

3 Research Directions

3.1 Extensions and refinements

3.1.1 Tractable filtering algorithms

The algorithms we developed in [FKM⁺25] for filtering and smoothing are exact, that is they compute the relevant conditionals precisely. However, in practice, exact computation is often infeasible, and one resorts to approximate methods. This includes methods such as particle filters and the Baum–Welch algorithm.

Their ubiquity in practice prompts the question of whether these methods can be treated categorically as well. In addition to the theoretical interest of understanding these methods categorically, such a treatment could also help in clarifying and formalizing the implementations of these algorithms in practice. Indeed, these methods involve a number of heuristics and design choices, an successful implementation is a matter of experience as much as theory. A categorical treatment could help in clarifying these choices, and in particular in understanding the trade-offs involved and guiding the design of new algorithms.

3.1.2 Partial stochastic map categories

The work [SM25] is merely the first introduction of a general construction of a probabilistically relevant class of quasi-Markov categories, and many further questions arise:

[2] This is a notion of infinite tensor product particularly suited to categorical probability, and in this case will be an instance of a restriction limit.

- Many Markov categories of interest are representable, meaning that they are Kleisli categories of a suitable monad. This suggests the question of whether one can give sufficient conditions on a commutative affine monad on a symmetric monoidal category to ensure that the Kleisli category is partializable. In light of the proofs of the partializability axioms for the main examples, this would likely involve conditions on the interaction of the monad with pullbacks along monomorphisms, similar to properties such as tautness.

- On the other hand, many of the results of [SM25] could be extended from the case where the domain inclusions is the class of deterministic monomorphisms to general well behaved ones (i.e. suitable stable systems of monomorphisms $\mathcal{M}$). In the theory of restriction categories, the *split* restriction categories are essentially equivalent to the categories of spans $\mathsf{Partial}(\mathsf{C}, \mathcal{M})$ for a restriction category C and a stable class of monomorphisms $\mathcal{M}$. This suggests that a sort of classification result of positive quasi-Markov categories with split domain idempotents may be possible, along the lines of every positive quasi-Markov category with split domain idempotents being equivalent to a partialization $\mathsf{Partial}(\mathsf{C}, \mathcal{M})$. This would additionally approach a complete characterization of the categories where the treatment of empirical sampling in [FGL$^+$25] would apply.

3.1.3 Partial algebras and empirical averaging

Our approach to empirical sampling maps and the categorical Glivenko–Cantelli theorem is just the first categorically, and many extensions and refinements are worth pursuing:

- Even in a measure-theoretic setting, the literature on empirical measures has traditionally focused on finite sequences, only varying the length as a parameter (with exceptions, such as [AP14]). This suggests our empirical sampling to be a tool worthy of study in its own right. Developing the theory of such maps would be interest to those working with empirical measures of infinite sequences, even in the measure-theoretic setting.

- On the other hand, [FGL$^+$25] constructs empirical sampling maps explicitly in $\mathsf{Partial}(\mathsf{BorelStoch})$, and the precise construction is technically subtle. Furthermore, the deduction of the standard law of large numbers from the categorical one involves carefully choosing the right empirical sampling map. This suggests the investigation of general constructions of classes of empirical sampling maps, and attempts to characterize categorically which ones lead to the standard law of large numbers. Alternatively, one could attempt to directly abstracting the notion of taking the composite of empirical sampling with the expectation, leading to an "empirical averaging" map from which a categorical strong law of large numbers could be derived directly.

3.2 New directions

3.2.1 Approximate methods, categorically

Several techniques such as Markov chain Monte Carlo methods, expectation–maximization algorithms, and variational inference are of practical use in a variety of settings, independent of the specific model details. A categorical treatment of such methods would enable the categorical framework to describe a wider range of algorithms using these techniques as building blocks. There is plenty of work on Bayesian updating in a categorical setting, but a categorical treatment of tractable methods for sampling and approximate inference is still lacking.

3.2.2 Expectations as lax algebras

On standard Borel spaces, the "expectation map" assigning to a distribution p on $\mathbb{R}$ its expectation does *not* define a partial algebra, even with domain the distributions with finite first moment. It is only a **lax partial algebra**, with the multiplication square only holding up to restriction. To be precise, one composite along the square is defined on the random distributions with finite first moment such that first moment of the expectations is finite, while the other is defined on those where the average of the first moments is finite. And the two are not equal in general.

The ubiquity of expectations in probability theory suggests the question of developing a general theory of lax partial algebras for the distribution monad. In fact, with expectations defined in terms of **Bochner integrals** [Coh13, Appendix E], expectations can be shown to define lax partial algebras for arbitrary separable Banach spaces. Furthermore, morphisms of these algebras would also have to be lax in a similar sense. Thus expectations would be objects is a category of lax partial algebras for a distribution monad and lax morphisms between them, in analogy to the phenomenon in the compact case with total algebras [Św74].

3.2.3 Categorical approaches to bootstrapping and (re-)sampling

Laws of large numbers are foundational to developing a wide variety of results in probability theory. They have proved elusive in categorical probability for some time, but with versions of them now available, the door is opened to developing further results relying on them, such as ergodic theorems. Alternatively, this could lead to more practical results such as a formal treatment of bootstrapping and more general (re-)sampling methods in Bayesian inference and/or randomized algorithms. Furthermore, issues of convergence regarding empirical measures are subtle, and the conceptual simplification afforded by the abstraction of the notion of empirical sampling map would lead to a clearer understanding of results derived in explicitly in terms of these.

References

[AP14] T. Austin and D. Panchenko. A hierarchical version of the de Finetti and Aldous–Hoover representations. *Probability Theory and Related Fields*, 159(3-4):809–823, 2014. doi: 10.1007/s00440-013-0521-0.

[CJ19] K. Cho and B. Jacobs. Disintegration and Bayesian Inversion via String Diagrams. *Math. Structures Comput. Sci.*, 29:938–971, 2019. doi: 10.1017/s0960129518000488.

[Coh13] D. L. Cohn. *Measure theory.* Birkhäuser, 2013.

[FGL+23] T. Fritz, T. Gonda, A. Lorenzin, P. Perrone, and D. Stein. Absolute continuity, supports and idempotent splitting in categorical probability. 2023. arXiv:2308.00651.

[FGL+25] T. Fritz, T. Gonda, A. Lorenzin, P. Perrone, and A. Shah Mohammed. Empirical Measures and Strong Laws of Large Numbers in Categorical Probability. *arXiv:2503.21576*, 2025. To appear in *Logical Methods in Computer Science.*

[FKM+25] T. Fritz, A. Klingler, D. McNeely, A. Shah Mohammed, and Y. Wang. Hidden Markov Models and the Bayes Filter in Categorical Probability. *IEEE Transactions on Information Theory*, 71(9):7052–7075, September 2025. doi: 10.1109/TIT.2025.3584695.

[Gad96] F. Gadducci. *On The Algebraic Approach To Concurrent Term Rewriting.* PhD thesis, University of Pisa, 1996.

[SM25] A. Shah Mohammed. Partializations of Markov Categories, September 2025. doi: 10.48550/arXiv.2509.05094.

[SW24] M. Soloviev and Y. Wang. Github project: Markov Categories in C++, 2024. url: github.com/blackhole89/markovcats.

[Św74] T. Świrszcz. Monadic functors and convexity. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.*, 22, 1974.