HIDDEN MARKOV MODELS AND THE BAYES FILTER IN CATEGORICAL PROBABILITY

TOBIAS FRITZ, ANDREAS KLINGLER, DREW MCNEELY, AREEB SHAH-MOHAMMED, AND YUWEN WANG

Consider the problem of navigating a spacecraft through the vastness of space. Ensuring that it follows its intended trajectory can require frequent corrections by appropriate application of thrust. Knowing how to correct the trajectory necessitates first of all a good understanding of the spacecraft's current position and velocity. Since such *orbit determination* comes with inherent inaccuracies [TSB04], it is essential to have a robust method to estimate the spacecraft's true position based on the available noisy observations. This is the type of problem which is addressed and optimally solved by the mathematical framework of **hidden Markov models** and the **Bayes filter**. A hidden Markov model describes the evolution of a spacecraft's position and velocity as a Markov chain of "hidden states" together with a sequence of noisy observations at each time step; the Bayes filter describes the optimal inferences that can be made about the hidden states from the noisy observations [CMR05, S1̈3].

In the present work, we develop the theory of hidden Markov models and the Bayes filter for Markov categories with conditionals in general. This provides two primary benefits over the traditional formulation:

- It provides a unifying account of various special cases like discrete probability, Gaussian probability (the Kalman filter), measure-theoretic probability and possibilistic nondeterminism.
- The string-diagrammatic formulation also sheds new light on the theory by providing a visually intuitive picture of the information flow involved.

Moreover, we hope that our current results will also be able to provide a new approach to *Blackwell's unique ergodicity problem* [Cv10], which currently still remains open in the general case where both the hidden states and the observed variables are continuous. However, this will still require us to develop more ergodic theory in categorical probability first.

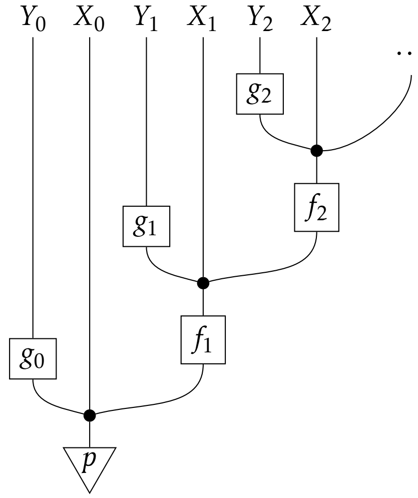
Let us now turn a more technical overview of our results. Throughout, we work in a Markov category $\mathcal{C}$ with conditionals [Fri20], such as the category of standard Borel measurable spaces and Markov kernels, or the category of sets and total relations.

Department of Mathematics and Institute for Theoretical Physics, University of Innsbruck, Austria

Department of Aerospace Engineering, University of Texas at Austin, USA

TF and ASM acknowledge funding from the Austrian Science Fund (FWF) through project P 35992-N. AK acknowledges support through project P 33122-N of the Austrian Science Fund (FWF) and funding of the Austrian Academy of Sciences (ÖAW) through the DOC scholarship 26547. YW's research is supported by the Austrian Science Fund (FWF) project P 34129 as well as by the University of Innsbruck Early Stage Funding Program.

Hidden Markov Models. A hidden Markov model in a Markov category is any state that can be written in this form:



(1)

Here, the objects $(X_i)_{i\in\mathbb{N}}$ correspond to the type of the hidden state, which may vary in time; the objects $(Y_i)_{i\in\mathbb{N}}$ are likewise the types of the observations. The diagram can either have infinitely many outputs, in which case its codomain is the Kolmogorov product [FR20]

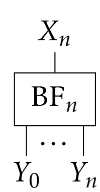
$$\bigotimes_{i\in\mathbb{N}}(Y_i\otimes X_i),$$

provided that that exists, or it can terminate after a finite number of steps. Our first main result is a characterization of such hidden Markov models in terms of conditional independence, where the proof makes important use of our earlier results on Bayesian networks and the d-separation criterion [FK23]. It in particular recovers the classical definition of hidden Markov models in terms of conditional independence, as in:

Theorem 1. *A state $p : I \to \bigotimes_{i\in\mathbb{N}}(Y_i\otimes X_i)$ is a hidden Markov model if and only if it satisfies the conditional independences*

$$X_{n+1} \perp X_0,\ldots,X_{n-1},Y_0,\ldots,Y_n \mid X_n, \qquad Y_{n+1} \perp X_0,\ldots,X_{n-1},Y_0,\ldots,Y_{n-1} \mid X_n \qquad \forall n.$$

Our next considerations concern the Bayes filter. This gadget answers the following question: given observations $Y_0,\ldots,Y_n$, what can we infer about the hidden state X_n? This is captured by marginalizing (1) on $X_0,\ldots,X_{n-1}$ and forming the **conditional** on $Y_0,\ldots,Y_n$, which results in a morphism



If the Markov category $\mathcal{C}$ is **representable** in the sense of [FGPR23], then the deterministic counterpart $\mathrm{BF}_n^\sharp$, which takes values in the distribution object PX_n, can be thought of as outputting our belief about the hidden state given a sequence of observations.

We have developed the following results about the Bayes filter entirely in terms of string diagrams.

Theorem 2. *There is a recursive construction of* BF_n *in terms of* BF_{n-1}.

Plugging in a sequence of (deterministic) inputs into BF_n results in a state on X_n which we call the *instantiated Bayes filter*. In the Markov category Gauss, this construction specializes to the well-known *Kalman filter*.

Theorem 3. *The joint distribution of the filter posteriors* $\mathrm{BF}_n^\sharp$ *is itself a Markov chain.*

This surprising result generalizes [Cv10, Lemma 2.4] to the setting of Markov categories.

Final note: We are currently also working on a computer implementation of all of this, and of the Bayes filter in particular, and we expect to be able to present it as part of an ACT talk.

References

[CMR05] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005.

[Cv10] Pavel Chigansky and Ramon van Handel. A Complete Solution to Blackwell's Unique Ergodicity Problem for Hidden Markov Chains. *The Annals of Applied Probability*, 20(6):2318–2345, 2010. arXiv:0910.3603.

[FGPR23] Tobias Fritz, Tomáš Gonda, Paolo Perrone, and Eigil Fjeldgren Rischel. Representable Markov categories and comparison of statistical experiments in categorical probability. *Theoretical Computer Science*, page 113896, 2023. arXiv:2010.07416.

[FK23] Tobias Fritz and Andreas Klingler. The d-separation criterion in categorical probability. *J. Mach. Learn. Res.*, 24(46):1–49, 2023. http://jmlr.org/papers/v24/22-0916.html.

[FR20] Tobias Fritz and Eigil Fjeldgren Rischel. Infinite products and zero-one laws in categorical probability. *Compositionality*, 2:3, 2020. compositionality-journal.org/papers/compositionality-2-3.

[Fri20] Tobias Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Adv. Math.*, 370:107239, 2020. arXiv:1908.07021.

[SÏ3] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, Cambridge, 2013.

[TSB04] Byron D. Tapley, Bob E. Schutz, and George H. Born. *Statistical Orbit Determination*. Elsevier, 2004.