

# **Scalable Learning of Probabilistic Circuits**

Renato Lui Geh

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: Computer Science

Advisor: Prof. Denis Deratani Mauá, PhD

This work was supported by CNPq grant #133787/2019-2,  
CAPES grant #88887.339583/2019-00 and EPECLIN FM-USP.

São Paulo  
November 1, 2021



# **Scalable Learning of Probabilistic Circuits**

Renato Lui Geh

This is the original version of the thesis  
prepared by candidate Renato Lui Geh, as  
submitted to the Examining Committee.

I hereby authorize the total or partial reproduction and publishing of this work for educational ou research purposes, as long as properly cited.

# Acknowledgements



# Resumo

Renato Lui Geh. **Aprendizado Escalável de Circuitos Probabilísticos**. Dissertação (Mestrado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2021.

**Palavras-chave:** Circuitos probabilísticos. Aprendizado de máquina. Modelos probabilísticos.





# Abstract

Renato Lui Geh. **Scalable Learning of Probabilistic Circuits**. Thesis (Master's). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2021.

**Keywords:** Probabilistic circuits. Machine learning. Probabilistic models.





# **Nomenclature**

## **List of Symbols**

$\sigma$  Sigmoid function

## **List of Figures**

## **List of Tables**

## **List of Algorithms**

## Notation

We use the following notation throughout the work. Random variables are written in upper case (e.g.  $X$ ,  $Y$ ) and their values in lower case (e.g.  $x$ ,  $y$ ). We identify propositional variables with 0/1-valued random variables, and use them interchangeably. Sets of variables and their joint values are written in boldface (e.g.  $\mathbf{X}$ ,  $\mathbf{x}$ ). Given a Boolean formula  $f$ , we write  $\langle f \rangle$  to denote its semantics, i.e. the Boolean function represented by  $f$ . For Boolean formulas  $f$  and  $g$ , we write  $f \equiv g$  if they are logically equivalent, that is, if  $\langle f \rangle = \langle g \rangle$ ; we abuse notation and write  $\phi \equiv f$  to indicate that  $\phi = \langle f \rangle$  for a Boolean function  $\phi$ .



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contributions and Dissertation Outline . . . . .	3

## Appendices

## Annexes

	<b>References</b>	<b>5</b>
--	-------------------	----------





# Chapter 1

## Introduction

When reasoning about the world, rarely can we find a realistic model that perfectly subsumes all of the needed relationships for flawless prediction. As such, the presence of a reliable uncertainty quantifier in intelligent systems is essential in developing performant yet diagnostible agents. This is made explicitly clear in the case of high-risk settings, such as autonomous vehicles or automated power plant systems, where a wrong prediction could cause disastrous consequences. A well known example in the case of the former is obstacle avoidance: while the agent should be capable of accurately identifying obstructions in its way during normal conditions, so should it be able to identify its own lack of confidence in high uncertainty situations like ones brought by environmental factors, such as severe blizzard or heavy rain, that could result in unreliable predictions, in which case calling out for human help may turn out to be the safest option. Other interesting applications of uncertainty quantification include out-of-distribution detection, which in our previous example could be visualized as the agent identifying a human’s driving as abnormally irregular (due to inebriation, infirmity, etc.) and appropriately taking control of the vehicle before a potential accident takes place.

One popular approach to quantifying uncertainty is through probability theory. By abstracting the world as a probability distribution with a finite number of observable random variables that encode a possibly incomplete knowledge of the environment, we are, in theory, able to answer a diverse set of complex queries as long as we have access to the (approximate) true data distribution. Practice is far different from theory, however, as most machine learning models either lie within a very limited range in the tractability spectrum in terms of inference (VERGARI *et al.*, 2020) or are too simplistic for complex real-world problems. Besides, although the majority of recent advances in deep learning claim some probabilistic meaning from the model’s output, they are often uncalibrated distributions, a result of focusing on maximizing predictive accuracy at the expense of predictive uncertainty (GUO *et al.*, 2017; OVADIA *et al.*, 2019; CHERNIKOVA *et al.*, 2019), ultimately producing overconfident and peculiar results (SZEGEDY *et al.*, 2013; WEI *et al.*, 2018; SU *et al.*, 2019; CHERNIKOVA *et al.*, 2019). Crucially, mainstream deep models (i.e. standard neural networks) usually optimize a conditional distribution over the to-be-predicted random variables – and are thus often called *discriminative* models – and do not model the actual joint distribution of the data, limiting inference capabilities and

uncertainty estimation.

In contrast, *generative* models seek to extract information from the joint (in varying capacities), and have lately seen a sharp rise in interest within deep learning. Despite this, most popular models do not admit either exact or tractable querying of key inference scenarios. For instance, although Generative Adversarial Networks (GANs) allow for efficient sampling (GOODFELLOW *et al.*, 2014), basic queries such as likelihood or marginals are outside of their capabilities. Similarly, Normalizing Flows (NF) also permit access to efficient sampling, with the added feature of computing likelihoods, but are severely limited by the base distribution when it comes to discrete data (REZENDE and MOHAMED, 2015; PAPAMAKARIOS *et al.*, 2021) albeit recent works on discretizing NFs have shown empirically good results (LIPPE and GAVVES, 2021; ZIEGLER and RUSH, 2019). Variational Auto-Encoders (VAEs) are (under certain conditions) a generalization of NFs (YU, 2020; GRITSENKO *et al.*, 2019) with known extensions for categorical data (ROLFE, 2017; VAHDAT, MACREADY, *et al.*, 2018; VAHDAT, ANDRIYASH, *et al.*, 2018), but only permit access to sampling and an upper-bound on the likelihood, with the latter available only after solving a complex optimization task (KINGMA and WELLING, 2014).

Despite the impressive achievements of the aforementioned generative models on realistically producing samples consistent with evidence, in none of the previous models are complex queries like structured prediction under partial observations, *maximum a posteriori* (MAP), conditional or marginal probabilities tractable. An obvious alternative would be Probabilistic Graphical Models (PGMs), although they too suffer from intractability when dealing with high treewidth networks (R. DECHTER, 1998; KOLLER and FRIEDMAN, 2009), severely limiting expressivity (cite here). Instead, we draw our attention to an expressive class of models that subsumes several families of probabilistic models with tractable inference capabilities.

Probabilistic Circuits (PCs) define a superclass of probabilistic models distinctly specified by recursive compositions of distributions through graphical formalisms. Vaguely speaking, PCs are computational graphs akin to neural networks, but whose network structure and computational units abide by special constraints. Within these specific conditions span a wide range of subclasses, each establishing a distinct set of restrictions on their structure in order to enable different segments within the tractability spectrum. As an example, Sum-Product Networks (SPNs, POON and DOMINGOS, 2011) are usually loosely defined over a couple of constraints: namely *smoothness* and *decomposability*, which in turn enables likelihood, marginal and conditional computations. Arithmetic Circuits (ACs, DARWICHE, 2003) add *determinism* to the mix, allowing for tractable computation of MAP probabilities. Similarly, Cutset Networks (CNets, RAHMAN *et al.*, 2014) employ the same constraints as ACs, but accept more expressive distributions as part of their computational units. Probabilistic Sentential Decision Diagrams (PSDDs, KISA *et al.*, 2014), Probabilistic Decision Graphs (PDGs, JAEGER, 2004) and And/Or-Graphs (AOGs, RINA DECHTER and MATEESCU, 2007) all require *smoothness* and *determinism*, but also call for a stronger version of *decomposability*, permitting all queries previously mentioned as well as computation of the Kullback-Leibler divergence and expectation between two circuits (CHOI *et al.*, 2020).

Interestingly, PCs are sufficiently expressive for generative *and* discriminative modeling

(KHOSRAVI *et al.*, 2019; RASHWAN *et al.*, 2018; ROOSHENAS and LOWD, 2016; GENS and DOMINGOS, 2012; SHAO *et al.*, 2020). In this dissertation though, we shall focus on the generative side of PCs.

## 1.1 Contributions and Dissertation Outline

referenced structural constraints in passing



# References

- [CHERNIKOVA *et al.* 2019] A. CHERNIKOVA, A. OPREA, C. NITA-ROTARU, and B. KIM. “Are self-driving cars secure? evasion attacks against deep neural networks for steering angle prediction”. In: *2019 IEEE Security and Privacy Workshops*. 2019, pp. 132–137 (cit. on p. 1).
- [CHOI *et al.* 2020] YooJung CHOI, Antonio VERGARI, and Guy Van den BROECK. “Probabilistic circuits: a unifying framework for tractable probabilistic models”. In: (2020). In preparation (cit. on p. 2).
- [DARWICHE 2003] Adnan DARWICHE. “A differential approach to inference in bayesian networks”. In: *Journal of the ACM* 50.3 (2003), pp. 280–305 (cit. on p. 2).
- [R. DECHTER 1998] R. DECHTER. “Bucket elimination: a unifying framework for probabilistic inference”. In: *Learning in Graphical Models*. Ed. by Michael I. JORDAN. Dordrecht: Springer Netherlands, 1998, pp. 75–104. ISBN: 978-94-011-5014-9. DOI: [10.1007/978-94-011-5014-9\\_4](https://doi.org/10.1007/978-94-011-5014-9_4). URL: [https://doi.org/10.1007/978-94-011-5014-9\\_4](https://doi.org/10.1007/978-94-011-5014-9_4) (cit. on p. 2).
- [Rina DECHTER and MATEESCU 2007] Rina DECHTER and Robert MATEESCU. “And/or search spaces for graphical models”. In: *Artificial Intelligence* 171.2 (2007), pp. 73–106. ISSN: 0004-3702. DOI: <https://doi.org/10.1016/j.artint.2006.11.003>. URL: <https://www.sciencedirect.com/science/article/pii/S000437020600138X> (cit. on p. 2).
- [GENS and DOMINGOS 2012] Robert GENS and Pedro DOMINGOS. “Discriminative learning of sum-product networks”. In: *Advances in Neural Information Processing Systems* 25. NIPS, 2012, pp. 3239–3247 (cit. on p. 3).
- [GOODFELLOW *et al.* 2014] Ian GOODFELLOW *et al.* “Generative adversarial nets”. In: *Advances in Neural Information Processing Systems* 27. Ed. by Z. GHAHRAMANI, M. WELLING, C. CORTES, N. D. LAWRENCE, and K. Q. WEINBERGER. Curran Associates, Inc., 2014, pp. 2672–2680. URL: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf> (cit. on p. 2).

- [GRITSENKO *et al.* 2019] Alexey A. GRITSENKO, Jasper SNOEK, and Tim SALIMANS. “On the relationship between normalising flows and variational- and denoising autoencoders”. In: *Deep Generative Models for Highly Structured Data, ICLR 2019 Workshop, New Orleans, Louisiana, United States, May 6, 2019*. OpenReview.net, 2019. URL: [https://openreview.net/forum?id=HklKEUUY%5C\\_E](https://openreview.net/forum?id=HklKEUUY%5C_E) (cit. on p. 2).
- [GUO *et al.* 2017] Chuan GUO, Geoff PLEISS, Yu SUN, and Kilian Q WEINBERGER. “On calibration of modern neural networks”. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 2017, pp. 1321–1330 (cit. on p. 1).
- [JAEGER 2004] Manfred JAEGER. “Probabilistic decision graphs-combining verification and ai techniques for probabilistic inference”. In: *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 12.1 supp (Jan. 2004), pp. 19–42. ISSN: 0218-4885. DOI: [10.1142/S0218488504002564](https://doi.org/10.1142/S0218488504002564). URL: <https://doi.org/10.1142/S0218488504002564> (cit. on p. 2).
- [KHOSRAVI *et al.* 2019] Pasha KHOSRAVI, Yitao LIANG, YooJung CHOI, and Guy VAN DEN BROECK. “What to expect of classifiers? reasoning about logistic regression with missing features”. In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, July 2019, pp. 2716–2724. DOI: [10.24963/ijcai.2019/377](https://doi.org/10.24963/ijcai.2019/377). URL: <https://doi.org/10.24963/ijcai.2019/377> (cit. on p. 3).
- [KINGMA and WELLING 2014] Diederik P. KINGMA and Max WELLING. “Auto-encoding variational bayes”. In: *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. Ed. by Yoshua BENGIO and Yann LECUN. 2014. URL: <http://arxiv.org/abs/1312.6114> (cit. on p. 2).
- [KISA *et al.* 2014] Doga KISA, Guy Van den BROECK, Arthur CHOI, and Adnan DARWICHE. “Probabilistic sentential decision diagrams”. In: *Knowledge Representation and Reasoning Conference (2014)* (cit. on p. 2).
- [KOLLER and FRIEDMAN 2009] Daphne KOLLER and Nir FRIEDMAN. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009. ISBN: 0262013193 (cit. on p. 2).
- [LIPPE and GAVVES 2021] Phillip LIPPE and Efstratios GAVVES. “Categorical normalizing flows via continuous transformations”. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: <https://openreview.net/forum?id=-GLNZeVDuik> (cit. on p. 2).
- [OVADIA *et al.* 2019] Yaniv OVADIA *et al.* “Can you trust your model’s uncertainty? evaluating predictive uncertainty under dataset shift”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019 (cit. on p. 1).

## REFERENCES

- [PAPAMAKARIOS *et al.* 2021] George PAPAMAKARIOS, Eric NALISNICK, Danilo Jimenez REZENDE, Shakir MOHAMED, and Balaji LAKSHMINARAYANAN. “Normalizing flows for probabilistic modeling and inference”. In: *Journal of Machine Learning Research* 22.57 (2021), pp. 1–64. URL: <http://jmlr.org/papers/v22/19-1028.html> (cit. on p. 2).
- [POON and DOMINGOS 2011] Hoifung POON and Pedro DOMINGOS. “Sum-product networks: a new deep architecture”. In: *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*. 2011, pp. 337–346 (cit. on p. 2).
- [RAHMAN *et al.* 2014] Tahrima RAHMAN, Prasanna KOTHALKAR, and Vibhav GOGATE. “Cutset networks: a simple, tractable, and scalable approach for improving the accuracy of chow-liu trees”. In: *Proceedings of the 2014th European Conference on Machine Learning and Knowledge Discovery in Databases*. 2014, pp. 630–645 (cit. on p. 2).
- [RASHWAN *et al.* 2018] Abdullah RASHWAN, Pascal POUPART, and Chen ZHITANG. “Discriminative training of sum-product networks by extended baum-welch”. In: *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*. Vol. 72. Proceedings of Machine Learning Research. 2018, pp. 356–367 (cit. on p. 3).
- [REZENDE and MOHAMED 2015] Danilo REZENDE and Shakir MOHAMED. “Variational inference with normalizing flows”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis BACH and David BLEI. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, July 2015, pp. 1530–1538. URL: <https://proceedings.mlr.press/v37/rezende15.html> (cit. on p. 2).
- [ROLFE 2017] Jason Tyler ROLFE. “Discrete variational autoencoders”. In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL: <https://openreview.net/forum?id=ryMxXPfex> (cit. on p. 2).
- [ROOSHENAS and LOWD 2016] Amirmohammad ROOSHENAS and Daniel LOWD. “Discriminative structure learning of arithmetic circuits”. In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. Ed. by Arthur GRETTON and Christian C. ROBERT. Vol. 51. Proceedings of Machine Learning Research. Cadiz, Spain: PMLR, Sept. 2016, pp. 1506–1514. URL: <https://proceedings.mlr.press/v51/rooshenas16.html> (cit. on p. 3).
- [SHAO *et al.* 2020] Xiaoting SHAO *et al.* “Conditional sum-product networks: imposing structure on deep probabilistic architectures”. In: *Proceedings of the 10th International Conference on Probabilistic Graphical Models*. Ed. by Manfred JAEGER and Thomas Dyhre NIELSEN. Vol. 138. Proceedings of Machine Learning Research. PMLR, Sept. 2020, pp. 401–412. URL: <https://proceedings.mlr.press/v138/shao20a.html> (cit. on p. 3).



- [SU *et al.* 2019] Jiawei SU, Danilo Vasconcellos VARGAS, and Kouichi SAKURAI. “One pixel attack for fooling deep neural networks”. In: *IEEE Transactions on Evolutionary Computation* 23.5 (2019), pp. 828–841 (cit. on p. 1).
- [SZEGEDY *et al.* 2013] Christian SZEGEDY *et al.* “Intriguing properties of neural networks”. In: *arXiv preprint arXiv:1312.6199* (2013) (cit. on p. 1).
- [VAHDAT, ANDRIYASH, *et al.* 2018] Arash VAHDAT, Evgeny ANDRIYASH, and William MACREADY. “Dvae#: discrete variational autoencoders with relaxed boltzmann priors”. In: *Advances in Neural Information Processing Systems*. Ed. by S. BENGIO *et al.* Vol. 31. Curran Associates, Inc., 2018. URL: <https://proceedings.neurips.cc/paper/2018/file/9f53d83ec0691550f7d2507d57f4f5a2-Paper.pdf> (cit. on p. 2).
- [VAHDAT, MACREADY, *et al.* 2018] Arash VAHDAT, William G. MACREADY, Zhengbing BIAN, Amir KHOSHAMAN, and Evgeny ANDRIYASH. “DVAE++: discrete variational autoencoders with overlapping transformations”. In: *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*. Ed. by Jennifer G. DY and Andreas KRAUSE. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 5042–5051. URL: <http://proceedings.mlr.press/v80/vahdat18a.html> (cit. on p. 2).
- [VERGARI *et al.* 2020] Antonio VERGARI, YooJung CHOI, Robert PEHARZ, and Guy Van den BROECK. *Probabilistic Circuits: Representations, Inference, Learning and Applications*. AAAI Tutorial. 2020 (cit. on p. 1).
- [WEI *et al.* 2018] Wenqi WEI *et al.* “Adversarial examples in deep learning: characterization and divergence”. In: *arXiv preprint arXiv:1807.00051* (2018) (cit. on p. 1).
- [YU 2020] Ronald YU. *A Tutorial on VAEs: From Bayes’ Rule to Lossless Compression*. 2020. arXiv: [2006.10273](https://arxiv.org/abs/2006.10273) [cs.LG] (cit. on p. 2).
- [ZIEGLER and RUSH 2019] Zachary M. ZIEGLER and Alexander M. RUSH. “Latent normalizing flows for discrete sequences”. In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Ed. by Kamalika CHAUDHURI and Ruslan SALAKHUTDINOV. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 7673–7682. URL: <http://proceedings.mlr.press/v97/ziegler19a.html> (cit. on p. 2).