

6.541/18.405 Final Project, Progress Report #2

Subject: The Computational Complexity of Probabilistic Inference

Proposed project title: On the Complexity of Inference in Probabilistic Graphical Models on Typical-Case Observations

George Matheos, April 29, 2024

1 My plan for the final project

1.1 High-level plan and input needed

My current plan is to write a research project, with a few new results I believe I know how to prove. These results build on analysis from Dagum and Luby [1, 2], and provide typical-case analysis counterparts to two theorems they prove relying on a worst-case analysis. I could alternatively write a literature report, as I have read a range of papers surveying the study of the complexity of inference. I could use input on whether the new results I am planning to include would be sufficient for a class project report, or if it would be better to do a literature report. I will go to Prof. William's office hours on Wed May 1 to ask him about this.

1.2 My plan

My plan is to study the complexity of probabilistic inference in probabilistic graphical models, on typical-case observations. To clearly articulate what I plan to study, I will introduce some notation.

A **probabilistic graphical model** is a tuple (V, E, P) , where V is a finite set of variables, E is a set of directed edges among the variable, and P describes a probability distribution over the value of each variable $v \in V$, given an assignment to its parent variables $\text{Pa}(v) \subseteq V$. $\text{Pa}(v)$ is the set of vertices $u \in V$ such that there is a directed edge $u \mapsto v \in E$. For any $v \in V$ and any assignment $a_{\text{Pa}(v)}$ to $\text{Pa}(v)$, I write $P(v = \cdot; a_{\text{Pa}(v)})$ to denote the probability distribution on the domain of v induced by this assignment and this graphical model. For our purposes, it is sufficient to take each $v \in V$ to be a boolean variable, so this domain can be taken as $\{0, 1\}$. It is required that (V, E) be a DAG. Then, P defines a joint distribution on all the variables in V , in a natural way.

For this report, I define an **inference problem** to consist of (1) a graphical model (V, E, P) , (2) a set of *observed variables* $Y \subseteq V$, (3) a set of *query variables* $X \subseteq V$ and (4) an assignment $y \in \{0, 1\}^{|Y|}$ to the collection of variables Y . I will write I to denote the tuple (P, E, V, X, Y) , the full inference problem, except excluding the assignment y . I will study several different types of inference problems, which will require an algorithm to return different types of information about the posterior distribution $P(X|Y = y)$.

On worst and typical case complexity. Cooper [3] and Dagum and Luby [1] have proven that if there exists an $\epsilon < 1/2$ and a polynomial time algorithm A such that for any inference problem (I, y) , $|A(I, y, x) - P(X = x|Y = y)| < \epsilon$, then **P = NP**. Their constructions involve constructing inference problems I , and then selecting an observed assignment y to a set Y of one variable ($|Y| = 1$). This assignment may have extremely low marginal probability: $P(Y = y) \ll 1$. As probabilistic graphical models are typically used to model aspects of the world, we ought expect that in practice, the observation values which we must condition our inferences upon will not be worst-case values, but typical-case values. Thus, in this project, I propose to study the complexity of inference under arbitrary (worst case) graphical models (P, V, E) , arbitrary latent variable subsets X , and arbitrary observation variable subsets Y , but typical-case observation values y .

My plan is to prove the following results suggesting that typical-case inference can be difficult.

1. Say that there exists $\epsilon, \delta < 1/2$ and a polynomial-time algorithm A such that for any I and any assignment x ,

$$\Pr_{y \sim P}[|A(I, y, X) - P(X = x|Y = y)| > \epsilon] < 1 - \delta$$

Then one-way functions do not exist.

2. The above result also holds if A is a probabilistic rather than deterministic algorithm. (The proof will involve a Chernoff style bound.)
3. Corollary: if one-way functions exist, then approximate posterior sampling given typical case observations, can be hard. There does not exist a probabilistic polynomial time algorithm A where $A(I, y)$ outputs an assignment x , such that its output distribution O has bounded total variation distance from the posterior: for any $\epsilon < 1/2$, we cannot have $\|O - P(X|Y = y)\|_{TV} < \epsilon$ for all inference problems.

These results will use similar constructions to Dagum and Luby's [1], though some extensions and modifications are needed to reduce from one-way functions rather than from SAT solving.

My plan is to additionally prove the following results about several cases in which typical-case inference *is* possible.

1. Say $|Y| = 1$. Then algorithms A as in item (2) and (3) in the previous list exists. (This is of interest because it shows that the constructions in [3] and [1] do not directly carry through to showing the hardness of inference with typical-case observations.)
2. Say we have either an upper bound B_I on the mutual information $I(X; Y)$, or an upper bound B_χ on the χ^2 -information $I_{\chi^2}(X; Y)$ [4]. Then probabilistic polynomial time algorithms A as in items (2) and (3) on the previous list exist, which work for all graphical models satisfying these upper bounds. (That is, the parametrized complexity of inference, on typical-case observations, with either of these information quantities as parameters, is polynomial.) Given a χ^2 -information upper bound B_χ , we can bound the algorithm runtime by a polynomial in B_χ . (We cannot get a polynomial bound based on the Shannon mutual information bound.)
3. (Tentative.) Inference in graphical models with upper-bounded mutual information $I(X; Y)$ on worst-case observations cannot be done in polynomial time unless $\mathbf{P} = \mathbf{NP}$. (The ability to state bounds like this therefore requires typical-case analysis.)

These upper bounds on the complexity of information in information-limited settings

I am planning to also include a discussion briefly surveying all of the papers about complexity of inference which I read during this project, to give you a sense of what I have studied (much of which ended up not being directly informative about these results).

2 Papers I have read

Here is a list of most of the papers I have read as a part of this course project: [5, 3, 1, 2, 6, 7, 8, 9, 10]. I also looked at [11, 12], but I have not read them in detail. My plan at this stage is to halt my literature review and try to write a project report proving the statements I described above.

I really enjoyed the papers about Majority K-SAT, by the way! I watched Shyan's video about it in addition to reading the paper, and thought it did a great job making clear the structural decomposition of the space of CNF formulae that informs the result. I was not able to figure out any new results related to this I thought I could prove, so for this project I thought I would write a report on the topics outlined above. But I'm quite interested in the line of research studying the combinatorial structure in concrete classes of probabilistic models, and understanding how this affects the complexity of probabilistic operations in them. (In the k-sat work, the probabilistic class of models is "sample n binary variables uniformly at random, then compute these CNF formulae on them", and the probabilistic operation under study is to compute a threshold on the marginal probability of observing all the CNF formulae.)

References

- [1] Paul Dagum and Michael Luby. “Approximating probabilistic inference in Bayesian belief networks is NP-hard”. In: *Artificial intelligence* 60.1 (1993), pp. 141–153.
- [2] Paul Dagum and Michael Luby. “An optimal approximation algorithm for Bayesian inference”. In: *Artificial Intelligence* 93.1-2 (1997), pp. 1–27.
- [3] Gregory F Cooper. “The computational complexity of probabilistic inference using Bayesian belief networks”. In: *Artificial intelligence* 42.2-3 (1990), pp. 393–405.
- [4] Yury Polyanskiy. *f-divergences*. Lecture notes. URL: https://people.lids.mit.edu/yp/homepage/data/LN_fdiv.pdf.
- [5] Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. “Random generation of combinatorial structures from a uniform distribution”. In: *Theoretical computer science* 43 (1986), pp. 169–188.
- [6] Nathanael L Ackerman, Cameron E Freer, and Daniel M Roy. “On the computability of conditional probability”. In: *Journal of the ACM (JACM)* 66.3 (2019), pp. 1–40.
- [7] Johan Kwisthout. “Approximate inference in Bayesian networks: Parameterized complexity results”. In: *International Journal of Approximate Reasoning* 93 (2018), pp. 119–131.
- [8] Shyan Akmal and Ryan Williams. “majority-3sat (and related problems) in polynomial time”. In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 1033–1043.
- [9] Till Tantau. “On the Satisfaction Probability of k -CNF Formulas”. In: *arXiv preprint arXiv:2201.08895* (2022).
- [10] Scott Aaronson. “The equivalence of sampling and searching”. In: *Theory of Computing Systems* 55.2 (2014), pp. 281–298.
- [11] Uriel Feige. “Relations between average case complexity and approximation complexity”. In: *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*. 2002, pp. 534–543.
- [12] Ankur Moitra. “Approximate counting, the Lovász local lemma, and inference in graphical models”. In: *Journal of the ACM (JACM)* 66.2 (2019), pp. 1–25.