6.541/18.405 Final Project, Progress Report #1
Subject: The Computational Complexity of Probabilistic Inference

George Matheos, April 18, 2024

1 Progress Report

Thank you for the feedback on my project proposal! I will make sure to read Akmal and Williams 2022 and Tantau 2022, and I'll try to be careful with the results in this literature (especially about the complexity of majority counting).

1.1 Current status & plan for next steps

My plan for next week is to try to read the remaining papers on my list.

I've made progress on a couple of "new research" directions,[1] which I think could lead to a project report on the typical-case complexity of approximate inference in probabilistic graphical models. This could include one section showing that computing $P(X|Y)$ up to additive approximation error is hard in general, if there exist invertible one-way functions which are hard to invert on typical-case random inputs. It could also show that conversely, if we have upper bounds on the mutual information $I(X;Y)$, there are upper bounds on the typical-case complexity of approximating $P(X|Y)$. (This is only one situation where inference is tractable, but it seems potentially worth articulating because it rephrases Dagum and Luby's 1997 result in terms of information theoretic quantities.)

1.2 A couple pieces of input that could be useful at this stage

If you have any feedback on the potential final report direction in the last paragraph, please let me know. I want to finish reading the papers on my list before committing to it, but if you have thoughts at this stage about whether it seems like a reasonable project report, that would be useful to know.

Also, do you have any pointers to good introductory reading on invertible one-way functions which are hard to invert on random inputs?

1.3 Note on the remainder of this doc

I am including an appendix with details on my progress in my literature review, and on the new research directions I am considering. No need to read this.

A Literature review update

A.1 Papers read

1. 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

2. Gregory F Cooper. "The computational complexity of probabilistic inference using Bayesian belief networks". In: *Artificial intelligence* 42.2-3 (1990), pp. 393–405

3. Paul Dagum and Michael Luby. "Approximating probabilistic inference in Bayesian belief networks is NP-hard". In: *Artificial intelligence* 60.1 (1993), pp. 141–153

[1] By "new research", I mean deriving results I haven't seen in any papers yet (though it is possible I just haven't found the right ones).

4. Paul Dagum and Michael Luby. "An optimal approximation algorithm for Bayesian inference". In: *Artificial Intelligence* 93.1-2 (1997), pp. 1–27

5. 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

6. Johan Kwisthout. "Approximate inference in Bayesian networks: Parameterized complexity results". In: *International Journal of Approximate Reasoning* 93 (2018), pp. 119–131

A.2 Papers remaining to read

1. 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

2. Till Tantau. "On the Satisfaction Probability of k-CNF Formulas". In: *arXiv preprint arXiv:2201.08895* (2022)

3. 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

4. Scott Aaronson. "The equivalence of sampling and searching". In: *Theory of Computing Systems* 55.2 (2014), pp. 281–298

5. Tomoyuki Yamakami. "Polynomial time samplable distributions". In: *journal of complexity* 15.4 (1999), pp. 557–574

6. Uriel Feige. "Relations between average case complexity and approximation complexity". In: *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*. 2002, pp. 534–543

A.2.1 Other papers to potentially read

1. Chunxiao Li et al. "On the hierarchical community structure of practical Boolean formulas". In: *Theory and Applications of Satisfiability Testing–SAT 2021: 24th International Conference, Barcelona, Spain, July 5-9, 2021, Proceedings 24*. Springer. 2021, pp. 359–376

2. I also want to find a good introduction to one-way functions which are hard to invert for random inputs. I've realized this is very relevant. Any pointers?

B Update on possible directions for new research

B.1 Analyzing the complexity of approximating $P(X|Y)$ for average-case Y

TLDR: I have realized this is closely related to the existence of injective one-way functions which are hard to invert for average-case input. One project direction would be to read up on such one-way functions, and spell out this connection (which I have not seen mentioned in the literature) in my project report.

B.1.1 The question to consider

Consider a Bayesian network (probabilistic graphical model) B on a set $\mathcal{V}$ of boolean variables. This defines a distribution P on $\{0,1\}^{\mathcal{V}}$. Let X, Y be disjoint subsets of $\mathcal{V}$ and let $x \in \{0,1\}^X$ and $y \in \{0,1\}^Y$. I am interested in the complexity of computing $P(X = x|Y = y)$ up to ϵ additive error.

It is known (Dagum and Luby 1993) that if the following condition holds, **NP** = **BPP**:

$$\exists \epsilon \in (0, 1/2), \exists \text{ ptime probabilistic Turing Machine } M \text{ such that} \forall B \text{ a description of a Bayesian network,}$$
$$\forall X, Y \text{ disjoint subsets of the variables in } B, \forall x \in \{0,1\}^X, \forall y \in \{0,1\}^Y,$$
$$\mathbb{P}_r[|M(B, X, Y, x, y, r) - P(X = x|Y = y)| \leq \epsilon] \geq 2/3 \quad (1)$$

Here, r is uniformly sampled random bits for the probabilistic Turing machine M.
I have not, however, seen results casting doubt on the following proposition:

$$\exists \epsilon << 1/2, \exists \text{ ptime probabilistic Turing Machine } M \text{ such that} \forall B \text{ a description of a Bayesian network,}$$
$$\forall X, Y \text{ disjoint subsets of the variables in } B, \forall x \in \{0,1\}^X$$
$$\mathbb{P}_{r;y\sim P}[|M(B, X, Y, x, y, r) - P(X = x|Y = y)| \leq \epsilon] >> 2/3 \quad (2)$$

The difference is that in the former proposition, y is an arbitrary observation value. In the second proposition, the probabilistic model B is worst-case, but y is sampled from the marginal of P on Y. (To the extent that probabilistic models describe the actual distributions with which things occur in the real world, we expect the inference problems presented by the real world to be on average-case y.)

B.1.2 Progress toward results

Observation 1: if $|Y| = 1$, (2) holds. I have realized that if we restrict to cases where Y contains one boolean variable ($|Y| = 1$), then (2) holds. The reason is that if Y contains one variable, either (A) $P(Y = y) > 0.001$, in which case approximate inference is easy to do via rejection sampling, or $P(Y = \neg y) \geq 0.999$, which means that with high probability, $Y \neq y$ and it is okay if it is hard to approximate $P(X|y)$.

Observation 2: if we allow large $|Y|$, (2) implies invertible one-way functions cannot be hard to invert on most random inputs. Say there exists a family of injective functions $(f_i)_{i=1}^{\infty}$ with $f_i : \{0,1\}^{m_i} \to \{0,1\}^{n_i}$ for some natural numbers m_i, n_i. Say these functions have the property that there is no probabilistic Turing machine M which can invert f_i on a random input x with probability $> 2/3$ for all i. That is, $\forall M$,

$$\exists i \text{ s.t. } \mathbb{P}_{x\sim\text{Uniform}(\{0,1\}^{n_i}),r}(M(f_i, y, r) = x) < 2/3 \quad \text{where } y = f_i(x)$$

Then for any i, we can encode a Bayesian network B which (1) first samples variables $X_1, \ldots, X_{m_i}$ uniformly at random, and then (2) computes $Y = f_i(X_1, \ldots, X_{m_i})$ by encoding a deterministic circuit for f_i. If with high probability we could approximate $\mathbb{P}(X_1 = 0)$, we could apply an iterative algorithm to construct an assignment to $X_1, \ldots, X_{m_i}$, and then check if $Y = f_i(X_1, \ldots, X_{m_i})$. If we could do this with high probability, we could invert f_i with high probability.

B.2 Complexity upper bounds parametrized by information theoretic quantities

TLDR: I think I know how to derive upper bounds on the typical-case complexity of approximating $P(X|Y)$, given upper bounds on the mutual information $I(X;Y)$ (or, better, the "Chi squared information" I_χ) between X and Y. I'm not sure how significant these bounds are, though it is kind of nice that they reframe Dagum and Luby's classic 1997 upper bounds using somewhat less restrictive conditions than Dagum and Luby's (with the result that we can only get average-case upper bounds, not worst-case upper bounds). These could possibly appear in a final project report on the average-case complexity of inference in graphical models, which shows (1) a reduction from 1-way function inversion, as described in the previous section, to demonstrate that typical-case inference can be hard in general, and then provides (2) upper bounds on the typical-case complexity of inference in graphical models, given upper bounds on the mutual information between variables.

B.2.1 A couple results I think I've derived

Dagum and Luby 1997 shows that a sufficient condition for the approximation of $P(X|Y)$ in a graphical model to be tractable is that for every variable v in the model, and every assignment to the parents $Pa(v)$ in the model, $P(v|Pa(v)) > \epsilon$ for some ϵ. They show that if $\epsilon \in \theta(1/n^c)$ for some c, where n is the number of variables in the model, and we have a uniform upper bound on the number of variables in Y, then approximate inference is tractable in polynomial time.

I spent a while thinking about what happens if we have the less restrictive condition that the mutual information $I(X;Y) < B$ for some bound B. My WIP math suggests that we can say something like the following: if y is generated from the model, then with probability $1 - \delta$, y has the property that we can compute an ϵ additive approximation to $P(X|Y)$ in

$$\text{Poly}(\exp(\frac{I(X;Y)}{\epsilon\delta}))$$

time. (I need to check this math more carefully.)

If we can instead bound the "Chi squared information" $I_\chi := \mathbb{E}_{y\sim P}[\chi^2(P(X|Y=y)||P(X))]$, then I think we can make the same guarantee about inference, but with runtime bound

$$\text{Poly}(\frac{I_\chi}{\epsilon\delta})$$

The reason this is better is because the $\frac{1}{\epsilon\delta}$ term is not exponentiated.