The Parameterized Complexity of Approximate Inference in Bayesian Networks *

Johan Kwisthout

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands j.kwisthout@donders.ru.nl

Abstract. Computing posterior and marginal probabilities constitutes the backbone of almost all inferences in Bayesian networks. These computations are known to be intractable in general [2]; moreover, it is known that *approximating* these computations is also NP-hard [3]. In the original paper we use *fixed-error randomized tractability analysis* [4], a recent randomized analogue of parameterized complexity analysis [5], to systematically address the complexity of (randomized) approximate inference in Bayesian networks. In this extended abstract we will give a brief introduction of the key concepts and results in this paper.

1 Approximate inference in Bayesian networks

As computing posterior probabilities in Bayesian networks is an NP-hard problem [2], one often resorts to approximate inferences. For example, rather than computing the exact value of a posterior probability distribution $\mathbf{P} = \Pr(H \mid E)$, one might settle for a distribution $\mathbf{Q}$ that is easier to compute and that is close to the target distribution $\mathbf{P}$. One such approximation approach is to *sample* from the distribution, yielding a randomized algorithm that (given sufficient samples) approximates the distribution. In the full paper [1] we investigated the properties of such randomized approximation problems. Given that Bayesian inference is NP-hard, an efficient *general* randomized algorithm can be ruled out, unless BPP = NP; we are thus focusing on *parameterizing* the problem [5] such that the expected running time of the algorithm is exponential only in the parameter, yet polynomial in the input. Formally, (parameterized) complexity theory is built on decision problems, hence, we are interested in parameters k for which the following[1] decision problem becomes feasible:

Additive-approximated Conditional Probability (AA-CPROB)
Input: A Bayesian network $\mathcal{B}$ with designated non-overlapping subsets of variables $\mathbf{H}$ and $\mathbf{E}$ and corresponding joint value assignments $\mathbf{h}$ to $\mathbf{H}$ and $\mathbf{e}$ to $\mathbf{E}$; in addition, error bound ϵ and rational number r.
Question: Is $\Pr(\mathbf{h} \mid \mathbf{e}) \pm \epsilon > r$?

* This is an extended abstract of [1].

[1] Among other variants such as relative approximations and marginal inferences.

2 Fixed-error randomized tractability

The complexity classes PP and BPP are defined as classes of decision problems that are decidable by a randomized algorithm in polynomial time with a particular probability of error; the difference between these two classes is in the bound on the error probability. In particular, a decision problem $\Pi \in \mathsf{PP}$ if and only if there exists a randomized algorithm accepting *Yes*-instances and rejecting *No*-instances of Π with probability strictly larger than $1/2$. In contrast, for BPP these probabilities are polynomially *bounded away* from $1/2$, allowing for effectively 'boosting up' the probability of acceptance while still taking only polynomial time. In order to *parameterize* the probability of acceptance—and thus making the time needed to boost the probability of acceptance close to 1 relative to a parameter—we introduce the fixed-error randomized tractability class FERT; informally, a problem in FERT is tractable if the parameter is small and intractable otherwise. This is formalized as follows:

Definition 1 (FERT). *Let Π be a decision problem and let k-Π be a parameterization of Π. We have that k-$\Pi \in \mathsf{FERT}$ if and only if there exists a randomized algorithm that accepts* Yes-*instances x of Π with probability at least $1/2 + \min(f(k), 1/|x|^c)$ for a constant c and arbitrary function $f : \mathbb{R} \to \langle 0, 1/2]$;* No-*instances are accepted with probability at most $1/2 - \min(f(k), 1/|x|^c)$.*

3 Highlighted results

Due to space constraints we just briefly list some interesting results as a teaser to the full paper, where d denotes the in-degree of the network, and D_e formalizes the maximal range of the parameters in the CPTs relative to the evidence:

$\{\Pr(\mathbf{h} \mid \mathbf{e}), \epsilon\}$-AA-CProb $\in \mathsf{FERT}$ but $\{d, \epsilon\}$-AA-CProb $\notin \mathsf{FERT}$ and also $\{\Pr(\mathbf{h} \mid \mathbf{e}), d\}$-AA-CProb $\notin \mathsf{FERT}$

$\{D_e\}$-AA-CProb $\notin \mathsf{FERT}$ but $\{D_e, \epsilon\}$-AA-CProb $\in \mathsf{FERT}$

In general, this approach allows us to explicate what *does* and what *doesn't* make approximate inference feasible in Bayesian networks. The full paper has a complete overview of known and new results cast into this complexity framework.

References

1. Kwisthout, J.: Approximate inference in Bayesian networks: Parameterized complexity results. International Journal of Approximate Reasoning (in press)
2. Cooper, G.F.: The computational complexity of probabilistic inference using Bayesian belief networks. Artificial Intelligence **42**(2) (1990) 393–405
3. Dagum, P., Luby, M.: Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial Intelligence **60**(1) (1993) 141–153
4. Kwisthout, J.: Tree-width and the computational complexity of MAP approximations in Bayesian networks. Journal of AI Research **53** (2015) 699–720
5. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Springer, Berlin (1999)