International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Approximate inference in Bayesian networks: Parameterized complexity results ☆

Johan Kwisthout

Radboud University, Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525HR Nijmegen, The Netherlands

A R T I C L E I N F O

Article history:
Received 6 December 2016
Received in revised form 18 October 2017
Accepted 23 October 2017
Available online xxxx

Keywords:
Approximate inference
Sampling
Bayesian networks
Parameterized complexity
Stochastic algorithms
De-randomization

A B S T R A C T

Computing posterior and marginal probabilities constitutes the backbone of almost all inferences in Bayesian networks. These computations are known to be intractable in general, both to compute exactly and to approximate (e.g., by sampling algorithms). While it is well known under what constraints *exact* computation can be rendered tractable (viz., bounding tree-width of the moralized network and bounding the cardinality of the variables) it is less known under what constraints *approximate* Bayesian inference can be tractable. Here, we extend the existing formal framework of *fixed-error randomized tractability* (a randomized analogue of fixed-parameter tractability), and use it to address this problem, both by re-interpreting known results from the literature and by providing some additional new results, including results on fixed parameter tractable de-randomization of approximate inference.

1. Introduction

Computing posterior and marginal probabilities constitutes the backbone of almost all inferences in Bayesian networks. These computations are known to be intractable in general [6]; moreover, it is known that *approximating* these computations is also intractable. To be precise, deterministic approximation is proven to be NP-hard by itself [8]; tractable *randomized* approximation is also ruled out unless NP $\subseteq$ BPP. To render exact computation tractable, bounding the tree-width of the moralized network is both necessary (under the assumption of the Exponential Time Hypothesis) [23] and (with bounded cardinality) sufficient [24]. For *approximate* inference, the picture is less clear, in part because there are multiple approximation strategies that all have different properties and characteristics. First of all, it matters whether we approximate marginal, respectively conditional probabilities. The approximation error can be measured either absolutely (also called additive approximation), i.e., independent of the probability that is to be approximated, or relative (also called multiplicative approximation) to this probability. Finally, the approximation algorithm can be deterministic (always guaranteeing a bound on the error) or randomized (guaranteeing a bounded error with high probability). In this broad array there are a few (somewhat isolated) tractability results [7–9,14,29,30], but an overview of what can and cannot render approximate inference tractable is still lacking.

In this paper we extend and apply *fixed-error randomized tractability analysis* [21], a recent randomized analogue of parameterized complexity analysis [10], to systematically address this issue. We consider both absolute and relative approximation,

☆ This paper is part of the Virtual special issue on the Eighth International Conference on Probabilistic Graphical Models, Edited by Giorgio Corani, Alessandro Antonucci, Cassio De Campos.

E-mail address: j.kwisthout@donders.ru.nl.

https://doi.org/10.1016/j.ijar.2017.10.029
0888-613X/

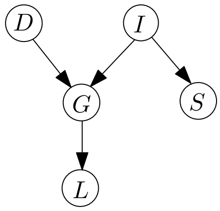

Fig. 1. The *Student* network, adapted from Figure 3.3 in [18].

using both randomized and deterministic algorithms, for the approximation of both marginal and conditional probabilities. We re-interpret old results and provide new results in terms of fixed-parameter or fixed-error tractability and intractability. In addition to identifying a number of corollaries from known results, some particular new contributions in this paper are de-randomization results of randomized approximations for fixed degree networks.

The remainder of this paper is structured as follows. After introducing the necessary preliminaries on Bayesian networks, approximation strategies, and parameterized computational complexity in Section 2, we introduce and extend fixed-error randomized tractability analysis in Section 3. We give an overview of results from the literature in Section 4.1 and some new results in Section 4.2. The paper is concluded in Section 5.

2. Preliminaries

In this section we introduce notation and provide for some preliminaries and our notational conventions in Bayesian networks, approximation algorithms, and complexity theory.

2.1. Bayesian networks

A (discrete) Bayesian network $\mathcal{B}$ is a graphical structure that models a set of discrete random variables, a joint probability distribution over these variables, and the conditional independences in this distribution [27]. $\mathcal{B}$ includes a directed acyclic graph $\mathbf{G}_{\mathcal{B}} = (\mathbf{V}, \mathbf{A})$, modeling the variables and conditional independences in the network, and a set of parameter probabilities Pr in the form of conditional probability tables (CPTs), capturing the strengths of the relationships between the variables. The network thus describes a joint probability distribution $\Pr(\mathbf{V}) = \prod_{i=1}^{n} \Pr(V_i \mid \pi(V_i))$ over its variables, where $\pi(V_i)$ denotes the parents of V_i in $\mathbf{G}_{\mathcal{B}}$. We define the *size* $|\mathcal{B}|$ of the network to be the number of bits needed to represent both $\mathbf{G}_{\mathcal{B}}$ and Pr. Our notational convention is to use upper case letters to denote individual nodes in the network, upper case bold letters to denote sets of nodes, lower case letters to denote value assignments to nodes, and lower case bold letters to denote joint value assignments to sets of nodes. The set of values v_i that a variable V can take is denoted as $\Omega(V)$.

Fig. 1 presents the running example we will use in this paper, the *Student network* introduced in [18]. Here, we model five random variables: The difficulty of a course D, with values difficult (d) and easy ($\neg d$); the intelligence of the student I, with values intelligent (i) and unintelligent ($\neg i$); the grade G of the course, where g_h, g_m, and g_l represent high, middle, and low grades, respectively; the SAT score S of the student (s_h = low, s_l = high); and finally whether one is willing to write a strong reference letter ($L = l$ if yes, $L = \neg l$ if no).

In the context of this paper we are particularly interested in the computation of marginal and conditional probabilities from a Bayesian network, defined as computational problems as follows:

MProb
Input: A Bayesian network $\mathcal{B}$ with designated subset of variables $\mathbf{H}$ and a corresponding joint value assignment $\mathbf{h}$ to $\mathbf{H}$.
Output: $\Pr(\mathbf{h})$.

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 $\mathbf{E}$.
Output: $\Pr(\mathbf{h} \mid \mathbf{e})$.

For example, in the student network it can be computed that $\Pr(S = s_h) = 0.275$ (i.e., the prior probability of scoring high at a SAT test is 0.275) and that $\Pr(L = l \mid I = \neg i) = 0.389$ (i.e., the probability that you write a strong recommendation letter for an unintelligent student is 0.389). It is well known that both MProb and CProb are intractable (NP-hard) problems to compute exactly, that is, there is strong evidence that there cannot exist a worst-case polynomial time algorithm for either of them.

Please cite this article in press as: J. Kwisthout, Approximate inference in Bayesian networks: Parameterized complexity results, Int. J. Approx. Reason. (2017), https://doi.org/10.1016/j.ijar.2017.10.029