On the Approximability of Weighted Model Integration on DNF Structures

Ralph Abboud , **İsmail İlkan Ceylan** and **Radoslav Dimitrov**
Department of Computer Science, University of Oxford
{firstname.lastname}@cs.ox.ac.uk

Abstract

Weighted model counting (WMC) consists of computing the weighted sum of all satisfying assignments of a propositional formula. WMC is well-known to be #P-hard for exact solving, but admits a fully polynomial randomized approximation scheme (FPRAS) when restricted to DNF structures. In this work, we study *weighted model integration*, a generalization of weighted model counting which involves real variables in addition to propositional variables, and pose the following question: Does weighted model integration on DNF structures admit an FPRAS? Building on classical results from approximate volume computation and approximate weighted model counting, we show that weighted model integration on DNF structures can indeed be approximated for a class of weight functions. Our approximation algorithm is based on three subroutines, each of which can be a *weak* (i.e., approximate), or a *strong* (i.e., exact) oracle, and in all cases, comes along with accuracy guarantees. We experimentally verify our approach over randomly generated DNF instances of varying sizes, and show that our algorithm scales to large problem instances, involving up to 1K variables, which are currently out of reach for existing, general-purpose weighted model integration solvers.

1 Introduction

Weighted model counting (WMC) has been introduced as a unifying approach for encoding probabilistic inference problems that arise in various formalisms. Informally, given a propositional formula, and a weight function that assigns every truth assignment a weight, WMC amounts to computing the weighted sum of all the satisfying assignments (Gomes, Sabharwal, and Selman 2009). Many probabilistic inference problems in *probabilistic graphical models* (Koller and Friedman 2009), *probabilistic planning* (Domshlak and Hoffmann 2007), *probabilistic logic programming* (De Raedt, Kimmig, and Toivonen 2007), *probabilistic databases* (Suciu et al. 2011), and *probabilistic knowledge bases* (Borgwardt, Ceylan, and Lukasiewicz 2017) can be reduced to a form of WMC.

Despite its wide applicability, WMC is limited to discrete domains and thus cannot be applied to domains involving real variables, and this motivated the study of *weighted model integration* (WMI) (Belle, Passerini, and Van Den Broeck 2015), as a generalization of WMC.

Building on the foundations of *satisfiability modulo theories (SMT)* (Barrett et al. 2009) , WMI can capture *hybrid domains* with mixtures of Boolean and continuous variables. Briefly, the input to WMI is a *hybrid* propositional formula that additionally involves arithmetic constraints (e.g., linear constraints over real, or integer variables), and a weight function that defines a *density* for every truth assignment of the formula. WMI is then the task of computing the sum of *integrals* over the densities of all the satisfying assignments of the given hybrid propositional formula (Belle, Passerini, and Van Den Broeck 2015; Morettin, Passerini, and Sebastiani 2019).

The standard formulation of WMI assumes a formula in *conjunctive normal form* (CNF) as an input, and to date, there is no study of WMI which is specifically tailored to formulas in *disjunctive normal form* (DNF). This is surprising, as both variants are widely investigated for WMC. We write WMI(CNF) and WMI(DNF) in the sequel to distinguish between these cases. These problems are clearly #P-hard for *exact solving*, as are their respective special cases WMC(CNF) and WMC(DNF) (Valiant 1979). For *approximate solving*, however, there is a strong contrast in computational complexity between variants of weighted model counting problems: WMC(DNF) has a *fully polynomial randomized algorithm scheme (FPRAS)* (Karp, Luby, and Madras 1989), producing polynomial-time approximations with guarantees, whereas WMC(CNF) is NP-hard to approximate. The latter polynomial-time inapproximability result immediately propagates to WMI(CNF), while the approximability status of WMI(DNF) remains *open*. In this paper, we pose the following question: Does WMI(DNF) admit an FPRAS?

We answer this question in the affirmative, and provide a polynomial-time algorithm for WMI(DNF) with probabilistic accuracy guarantees. The intuition behind our result is based on two observations. First, the special case of WMI(DNF) without any arithmetic constraints corresponds to WMC(DNF) which has an FPRAS (Karp, Luby, and Madras 1989). Second, the special case of WMI(DNF) with constant weight functions, and without any Booleans, corresponds to computing the volume of *unions of convex bodies*, which also has an FPRAS (Bringmann and Friedrich 2010). Our result builds on these results, and extends them, by allowing extra constructs essential for WMI, while preserving the approximation guarantees. Our main contributions can be summarized as follows:

– We propose an efficient approximation algorithm for WMI(DNF), called APPROXWMI, extending the algorithm given in (Bringmann and Friedrich 2010).

– We prove that APPROXWMI is an FPRAS provided that the weight functions are *concave*, and can be factorized into products of weights of literals. We provide asymptotic bounds for the running time of the algorithm.

– We extend APPROXWMI to the case where the products of weights assumption is relaxed, and provide asymptotic bounds for the running time of the algorithm.

– We experimentally verify our approach, using a strong oracle for computing the volume of a body. Our experiments suggest that APPROXWMI solves large problem instances, including up to 1K variables, which are out of reach for any existing, general-purpose WMI solver.

The full paper is to appear at KR 2020, and is available at https://arxiv.org/pdf/2002.06726.pdf

References

Barrett, C.; Sebastiani, R.; Seshia, S.; and Tinelli, C. 2009. *Satisfiability modulo theories*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. 1 edition. 825–885.

Belle, V.; Passerini, A.; and Van Den Broeck, G. 2015. Probabilistic inference in hybrid domains by weighted model integration. In *Proc. of IJCAI*.

Borgwardt, S.; Ceylan, İ. İ.; and Lukasiewicz, T. 2017. Ontology-mediated queries for probabilistic databases. In *Proc. of AAAI*.

Bringmann, K., and Friedrich, T. 2010. Approximating the volume of unions and intersections of high-dimensional geometric objects. *Comput. Geom.* 43(6-7):601–610.

De Raedt, L.; Kimmig, A.; and Toivonen, H. 2007. ProbLog: A probabilistic prolog and its application in link discovery. In *Proc. of IJCAI*.

Domshlak, C., and Hoffmann, J. 2007. Probabilistic planning via heuristic forward search and weighted model counting. *JAIR* 30(1).

Gomes, C. P.; Sabharwal, A.; and Selman, B. 2009. Model counting. In *Handbook of Satisfiability*. IOS Press.

Karp, R. M.; Luby, M.; and Madras, N. 1989. Monte-Carlo approximation algorithms for enumeration problems. *J. Algorithms* 10(3).

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Morettin, P.; Passerini, A.; and Sebastiani, R. 2019. Advanced SMT techniques for weighted model integration. *AIJ* 275:1 – 27.

Suciu, D.; Olteanu, D.; Ré, C.; and Koch, C. 2011. *Probabilistic Databases*. Morgan & Claypool.

Valiant, L. G. 1979. The complexity of computing the permanent. *TCS* 8(2).