

Random Models of Very Hard 2QBF and Disjunctive Programs: An Overview*

Giovanni Amendola¹, Francesco Ricca¹, and Miroslaw Truszczyński²

¹ University of Calabria, Italy
{ricca,amendola}@mat.unical.it
² University of Kentucky, USA
mirek@cs.uky.edu

Abstract. We present an overview of models of random quantified boolean formulas and their natural random disjunctive ASP program counterparts that we have recently proposed. The models have a simple structure but also theoretical and empirical properties that make them useful for further advancement of the SAT, QBF and ASP solvers.

Keywords: QBF, ASP, Random models

1 Introduction

The study of theoretical and practical properties of random models has received substantial attention in Computer Science. The results obtained in these studies have had major practical impact especially in boolean satisfiability (SAT) [1, 13] and constraint satisfaction problems (CP) [10]. Random models are characterized by intriguing phase-transition phenomena that are often associated with the inherent hardness of instances. Inherently hard instances for SAT and QBF solvers are essential for designing and testing search methods employed by solvers [1]. Thus it is important to study models of random propositional formulas and QBFs that can reliably generate instances of a desired hardness [9].

Previous work in this area mainly focused on random formulas in the conjunctive normal form (CNF), and random prenex-form QBFs with the matrix in CNF or disjunctive normal form (DNF) (depending on the quantifier sequence). The fixed-length clause model of k -CNF formulas and its 2QBF extension have been the subjects of extensive studies. Formulas in the fixed-length clause model consist of m clauses over a (fixed) set of n variables, each clause with k non-complementary literals. All formulas are assumed to be equally likely. For that model it is known that there are reals $\rho_l(k)$ and $\rho_u(k)$ such that if $m/n < \rho_l(k)$, a formula from the model is almost surely satisfiable (SAT), and if $m/n > \rho_u(k)$, almost surely unsatisfiable (UNSAT). It is conjectured that $\rho_l(k) = \rho_u(k)$. That conjecture is still open. However, it holds asymptotically, i.e., the two bounds converge to each other with $k \rightarrow \infty$ [2]. An important empirical property of these models is called the *easy-hard-easy* pattern [11]. Basically, instances from

* Results originally published in [4]

the phase transition region are difficult to be solved (*hard* in jargon), and so they are useful for assessing solver performance; whereas those from regions on both sides of the phase transition are *easy* to solve and thus, they are well suited for solver testing and similar purposes [1].

The fixed-length clause model was extended to QBFs by Chen and Interian [6]. In addition to n and m (understood as above), their model includes parameters controlling the structure of formulas. Once these parameters are fixed, similar properties as in the case of the k -CNF model emerge. There is a phase transition region associated with a specific value of the ratio m/n (that does not depend on n) and the easy-hard-easy pattern can be experimentally verified. Both the fixed-length clause model for SAT and the Chen and Interian model for QBF are based on formulas in normal forms. However, many applications give rise to formulas in non-normal forms motivating studies of solvers of non-normal form formulas and QBFs, and raising the need of models of random non-normal form formulas. A first response to that challenge was provided by Navarro and Voronkov [12, 7] with the *fixed-shape* model. The model is similar to the extension of the k -CNF one to QBFs, but *fixed shape* (and size) non-normal form formulas are used in place of k -clauses as the key building blocks. Motivated by the work on random SAT and QBF models, researchers proposed models of random logic programs, and obtained empirical and theoretical results concerning their properties [16, 14, 15]. Those results could be of substantial interest to answer set programming (ASP) [5], a popular computational formalism based on disjunctive logic programs. However, the first results in this area were limited to *non-disjunctive* logic programs.

In [4] we have presented new models of random *non-normal* form formulas and 2QBFs. Specifically, we considered disjunctions of t k -CNF formulas (in the case of QBFs, using them as matrices). We called models generating such formulas *multi-component*. Instances obtained according to our models are different from the ones originating from the fixed-shape model of Navarro and Voronkov, as their building blocks (k -CNF formulas) do not have a fixed size. QBFs from our model have a natural representation as disjunctive logic programs, obtained by extending the well-known encoding of [8]. Thus, a model of random *disjunctive* logic programs originates from the multi-component models of QBFs.

In this paper we give an overview of the random models presented in [4], and in particular we focus on the theoretical properties of our models by recalling the theoretical bounds on the location of the phase transition. Experimental results reported in [4] confirm that our models exhibit the easy-hard-easy pattern in correspondence of the phase transition. Thus, our results provide new ways to generate hard and easy instances of SAT, and QBF formulas and ASP programs.

2 Preliminaries

In this section, we recall the *fixed-length clause model* for random CNF formulas and the *Chen-Interian model* for random QBF formulas. We then describe the models of QBFs and programs proposed by [4]. By $C(k, n, m)$ we denote the

set of all k -CNF formulas consisting of m clauses over (some fixed) set of n propositional variables. Similarly, $D(k, n, m)$ stands for the set of all k -DNF formulas of m *products* (conjunctions of non-complementary literals) over an n -element set of atoms.

The fixed-length clause model. The model is given by the set $C(k, n, m)$ of CNF formulas, with all formulas assumed equally likely. Formulas from the model can be generated by selecting m k -literal clauses over a set of n variables uniformly, independently and with replacement. Let us denote by $p(k, n, m)$ the probability that a random formula in $C(k, n, m)$ is SAT. We define $\rho_l(k)$ to be the supremum over all real numbers x such that for every $\rho < x$, it holds that $\lim_{n \rightarrow \infty} p(k, n, \lfloor \rho n \rfloor) = 1$. Similarly, we define $\rho_u(k)$ to be the infimum over all real numbers x such that for every $\rho > x$, $\lim_{n \rightarrow \infty} p(k, n, \lfloor \rho n \rfloor) = 0$. It is known that $\rho_l(k)$ and $\rho_u(k)$ are well defined. Moreover, $\rho_l(k) \leq \rho_u(k)$ and, it is conjectured that $\rho_l(k) = \rho_u(k)$. The conjecture holds asymptotically, that is, $\lim_{k \rightarrow \infty} \rho_l(k) = \lim_{k \rightarrow \infty} \rho_u(k)$.

The Chen-Interian model. The model generates QBFs of the form $\forall X \exists Y F$, where sets X and Y are disjoint, $|X| = A$, $|Y| = E$, and F is a CNF formula with m clauses, each containing a literals with variables in X and e literals with variables in Y . We write $Q(a, e; A, E; m)$ for the set of all such QBFs. The Chen-Interian model generates QBFs from $Q(a, e; A, E; m)$, with all formulas equally likely. Let $q(a, e; A, E; m)$ be the probability that a random QBF from $Q(a, e; A, E; m)$ is true. Let $r > 0$ be a fixed real. We set $\nu_l(a, e; r)$ to be the supremum over all real numbers x such that for every $\nu < x$, $\lim_{n \rightarrow \infty} q(a, e; A, E; \lfloor \nu n \rfloor) = 1$, where $A = \lfloor rE \rfloor$ and $n = A + E$. Similarly, we set $\nu_u(a, e; r)$ to be the infimum over all real numbers x such that for every $\nu > x$, $\lim_{n \rightarrow \infty} q(a, e; A, E; \lfloor \nu n \rfloor) = 0$, again with $A = \lfloor rE \rfloor$ and $n = A + E$. Chen and Interian proved that $\nu_l(a, e; r)$ and $\nu_u(a, e; r)$ are well defined. Clearly, $\nu_l(a, e; r) \leq \nu_u(a, e; r)$. Whether $\nu_l(a, e; r) = \nu_u(a, e; r)$ is an open problem. The quantities $\nu_l(a, e; r)$ and $\nu_u(a, e; r)$ delineate the phase-transition region. For QBFs generated from the model $Q(a, e; \lfloor rE \rfloor, E; \lfloor \nu n \rfloor)$ (with fixed r), Chen and Interian experimentally observed the easy-hard-easy pattern as ν grows, showed that the hard region is aligned with the phase transition, and that the same behavior emerges no matter what concrete r is fixed as the ratio A/E .

3 Multi-component models

Random SAT and QBF. Let \mathcal{F} be a class of propositional formulas (or a model of random formula). By t - \mathcal{F} we denote the class of all disjunctions of t formulas from \mathcal{F} (or a model generating disjunctions of random formulas from \mathcal{F}). Similarly, if \mathcal{Q} is a class (model) of QBFs of the form $\forall X \exists Y F$, where $F \in \mathcal{F}$, we write t - \mathcal{Q} for the class (model) of all QBFs of the form $\forall X \exists Y F$, where $F \in t$ - \mathcal{F} . We refer to models t - \mathcal{F} and t - \mathcal{Q} as *multi-component*. For QBFs we also consider the dual model to t - \mathcal{Q} , based on conjunctions of t DNF formulas. It gives rise to the multi-component model of disjunctive logic programs via the Eiter-Gottlob

translation [8]. In all cases, we assume that formulas (QBFs, respectively) are equally likely.

In our work we focus on models based on the model $C(k, n, m)$. First, we note that the corresponding multi-component model $t\text{-}C(k, n, m)$ has similar satisfiability properties and that the phase transition regions in the two models are closely related. Let $p_t(k, n, m)$ be the probability that a random formula in $t\text{-}C(k, n, m)$ is SAT.

Theorem 1 *Let $t \geq 1$ be a fixed integer. Then, for every $\rho < \rho_l(k)$, it holds that $\lim_{n \rightarrow \infty} p_t(k, n, \lfloor \rho n \rfloor) = 1$, and for every $\rho > \rho_u(k)$, $\lim_{n \rightarrow \infty} p_t(k, n, \lfloor \rho n \rfloor) = 0$.*

The proof follows from the identity $p_t(k, n, m) = 1 - (1 - p(k, n, m))^t$. Thus, if the phase transition conjecture holds for the single component model $C(k, n, m)$, it also holds for the multi-component model $t\text{-}C(k, n, m)$, and the threshold value is the same for every t .

We also considered the multi-component model $t\text{-}Q(a, e; A, E; m)$ of QBFs, with the Chen-Interian model as its single-component. Let $q_t(a, e; A, E; m)$ be the probability that a random QBF from $t\text{-}Q(a, e; A, E; m)$ is true (in particular, $q_1(a, e; A, E; m) = q(a, e; A, E; m)$). Extending Chen and Interian's work, we can prove that the phase transition for different values of t coincide (and coincide with the phase transition in the Chen-Interian model).

Theorem 2 *For every integer $t \geq 1$ and real number $r > 0$: if $\nu < \nu_l(a, e; r)$, it holds that $\lim_{n \rightarrow \infty} q_t(a, e; A, E; \lfloor \nu n \rfloor) = 1$, and if $\nu > \nu_u(a, e; r)$, it holds that $\lim_{n \rightarrow \infty} q_t(a, e; A, E; \lfloor \nu n \rfloor) = 0$ (where $A = \lfloor rE \rfloor$ and $n = A + E$).*

The theorems above describe the situation when t is fixed and n is large. When n is fixed and t grows, the analysis of $p_t(k, n, m)$ and $q_t(a, e; A, E; m)$ shows that the region of the transition from SAT to UNSAT shifts to the right. Of course, once we stop growing t and start increasing n again, the phase transition region will move back to the left.

Random Disjunctive Programs. Our model of random disjunctive programs is based on the translation from QBFs to programs due to Eiter and Gottlob [8]. It works on the model dual to the one we discussed above. The model consists of QBFs $\Phi = \exists X \forall Y F$, where $F \in D(e, a; E, A; m)$, the set of formulas dual to $C(e, a; E, A; m)$. The model clearly has the same properties (modulo a switch between true and false).

The Eiter-Gottlob translation has a simple extension to QBFs obtained from the multi-component models. We write $t\text{-}D_{dlp}(e, a; E, A; m)$ for the set of programs consisting of the following fixed part: $z \vee z'$, for each $z \in Z$; $y \leftarrow w$, and $y' \leftarrow w$, for each $y \in Y$; $w \leftarrow w_1, \dots, w_t$, and $w \leftarrow \text{not } w$. And the core of mt Horn rules of the form $w_h \leftarrow z_1, \dots, z_\ell$, where $h = 1, \dots, t$, each w_h is the head of exactly m rules, $\ell = e + a$, and the body of each rule has e atoms z and z' with $z \in X$ (and so, also a atoms of the form z and z' with $z \in Y$). Note that programs in $1\text{-}D_{dlp}(e, a; E, A; m)$ coincide with those in $D_{dlp}(e, a; E, A; m)$ modulo a rewriting, where the rule $w \leftarrow w_1$ is removed and w_1 is replaced by w .

Programs in $t\text{-}D_{dlp}(e, a; E, A; m)$ correspond to $\exists\forall$ QBFs whose matrix belongs to $t\text{-}D(e, a; E, A; m)$. They can be seen as the results of translating QBFs Φ into logic programs Φ_{dlp} , where we use variables w_i to represent component DNF formulas in the matrix of Φ . The correspondence $\Phi \mapsto \Phi_{dlp}$ preserves the semantics in the following sense.

Theorem 3 *Let $\Phi = \exists X\forall YF$, for $F \in t\text{-}D(e, a; E, A; m)$. Then Φ is true if and only if Φ_{dlp} has an answer set, where Φ_{dlp} is the disjunctive logic program in $t\text{-}D_{dlp}(e, a; E, A; m)$ corresponding to Φ .*

Empirical properties and conclusion. The experimental results on satisfiability [4, 3], agree with the above-reported theoretical analysis, that is formulas from our models show the easy-hard-easy pattern and a strong dependence of hardness on t . Thus, despite their simple structure, our models have theoretical and empirical properties that make them important for further advancement of solvers.

References

1. Achlioptas, D.: Random satisfiability. In: Handbook of Satisfiability, pp. 245–270 (2009)
2. Achlioptas, D., Moore, C.: The asymptotic order of the random k-sat threshold. In: FOCS 2002. pp. 779–788 (2002)
3. Amendola, G., Ricca, F., Truszczynski, M.: A generator of hard 2qbf formulas and ASP programs. In: KR. pp. 52–56. AAAI Press (2018)
4. Amendola, G., Ricca, F., Truszczynski, M.: Generating hard random boolean formulas and disjunctive logic programs. In: IJCAI. pp. 532–538. ijcai.org (2017)
5. Brewka, G., Eiter, T., Truszczynski, M.: Answer set programming at a glance. Commun. ACM **54**(12), 92–103 (2011)
6. Chen, H., Interian, Y.: A model for generating random quantified boolean formulas. In: IJCAI 2005. pp. 66–71 (2005)
7. Creignou, N., Egly, U., Seidl, M.: A framework for the specification of random SAT and QSAT formulas. In: TAP 2012. LNCS, vol. 7305, pp. 163–168 (2012)
8. Eiter, T., Gottlob, G.: On the computational cost of disjunctive logic programming: Propositional case. AMAI **15**(3-4), 289–323 (1995)
9. Gent, I.P., Walsh, T.: Beyond NP: the QSAT phase transition. In: AAAI. pp. 648–653 (1999)
10. Mitchell, D.G.: Resolution complexity of random constraints. In: CP 2002. LNCS, vol. 2470, pp. 295–309 (2002)
11. Mitchell, D.G., Selman, B., Levesque, H.J.: Hard and easy distributions of SAT problems. In: AAAI 1992. pp. 459–465 (1992)
12. Navarro, J., Voronkov, A.: Generation of hard non-clausal random satisfiability problems. In: AAAI 2005. pp. 436–442 (2005)
13. Selman, B., Mitchell, D.G., Levesque, H.J.: Generating hard satisfiability problems. Artif. Intell. **81**(1-2), 17–29 (1996)
14. Wang, K., Wen, L., Mu, K.: Random logic programs: linear model. TPLP **15**(6), 818–853 (2015)
15. Wen, L., Wang, K., Shen, Y., Lin, F.: A model for phase transition of random answer-set programs. ACM Trans. Comput. Log. **17**(3), 22 (2016)
16. Zhao, Y., Lin, F.: Answer set programming phase transition: A study on randomly generated programs. In: ICLP 2003. LNCS, vol. 2916, pp. 239–253 (2003)