Discrete Mathematics and Theoretical Computer Science **6**, 2004, 425–436

Generating functions and the satisfiability threshold

Vincent Puyhaubert

Algorithm Project, INRIA Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex

received Apr 20, 2004, *revised Sep 28, 2004*, *accepted Oct 12, 2004*.

The 3-SAT problem consists in determining if a boolean formula with 3 literals per clause is satisfiable. When the ratio between the number of clauses and the number of variables increases, a threshold phenomenon is observed: the probability of satisfiability appears in simulations to decrease sharply from 1 to 0 in the neighbourghood of a threshold value, conjectured to be close to 4.25. Although the threshold has been proved to exist for the 2-SAT formulæ and for closely related problems like 3-XORSAT, there is still no precise mathematical characterization of the threshold phenomena involved in the 3-SAT problem.

Recent works have provided so far upper and lower bounds for the threshold's potential location. We present here a unified approach to upper bounds that is based on urn models, generating functions, and saddle-point bounds. No new upper bound is presented here but instead, we show that several existing results (with long proofs) can be organized in a simpler and uniform manner.

Keywords: First moment method, exponential generating functions, saddle-point bounds

1 Introduction

We consider boolean formulæ over a set of variables $x_1,\dots,x_n$ (where the x_j range over $\{0,1\}$ or, equivalently, $\{\text{true},\text{false}\}$). A literal is either a variable x_j or a negated variable $\neg x_j$. It is known that each boolean formula admits a conjunctive normal form, being a conjunction of clauses, themselves disjunctions of literals. A 3-SAT formula is then such a formula with exactly 3 literals per clause. A typical formula is then for example:

$$\Phi = (x_1 \vee \neg x_2 \vee x_4) \wedge (\neg x_2 \vee \neg x_3 \vee x_5) \wedge (x_1 \vee \neg x_4 \vee \neg x_5) \wedge (x_3 \vee \neg x_4 \vee \neg x_5).$$

We will choose the standard probabilistic model where each clause is composed of a set of three literals composed of distinct variables. There are then $8\binom{n}{3}$ distinct clauses and $(8\binom{n}{3})^m$ formulæ with m clauses. The quantities n,m are the fundamental parameters of the model. From previous studies, we know that the "interesting" region is when n and m are linearly related $m = rn$. (Alternative models have occasionally been used for convenience in calculations, for example, the three literals may be ordered and not necessarily distinct so that there would be $8n^3$ clauses. All such models are easily proved to be equivalent with respect to the asymptotic probability of satisfiability).

1365–8050

In Figure 1, taken from Ludovic Meunier's master's thesis [19], a *phase transition phenomenon* can be observed regarding the satisfiability of these formulæ when they are drawn at random. As the ratio $r = m/n$ of the number m of clauses to the number n of variables increases, the probability of satisfiability drops abruptly from nearly 1 to nearly 0.

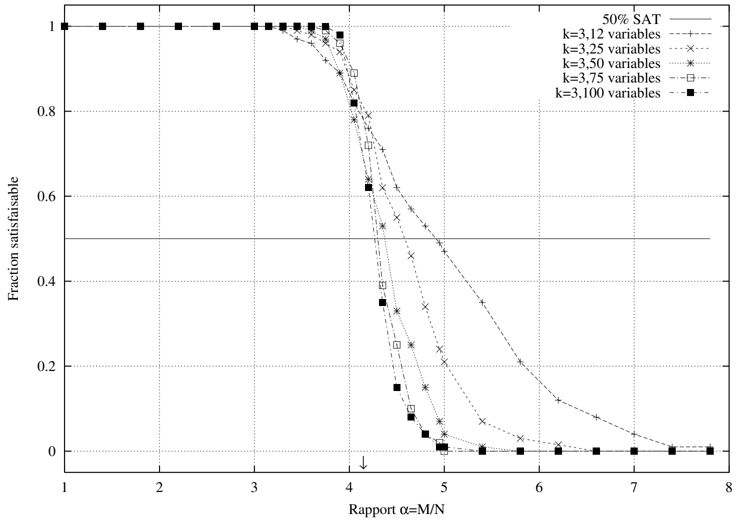

Fig. 1: Ratio of satisfiable formulæ *wrt* the parameter m/n.

From these experiments, it is believed that there exists a critical value r_3 such that for any $\varepsilon > 0$, the probability of satisfiability tends to 1 for $r < r_3 - \varepsilon$ (as m and n tend to infinity), and tends to 0 for $r > r_3 + \varepsilon$. Experiments suggest for r_3 the value 4.25 ± 0.05. However, so far, only successive upper and lower bounds on the threshold's potential location have been obtained. The table below lists the bounds successively established for the 3-SAT threshold. The bounds marked with a star admit an extension to k-SAT for any k. A survey by Olivier Dubois explains the techniques used to obtain them [7]. For the lower bounds, one can refer to the surveys by Franco [11] and Achlioptas [2].

Lower bounds for 3-SAT threshold		Upper bounds for 3-SAT threshold	
2.9^*	Chao and Franco (1986,1990) [6]	5.191^*	Franco and Paull (1983) [12]
$2/3^*$	Chvátal and Reed (1992) [20]	5.081	El Mafthoui and Fernandez de la Vega (1993) [10]
1.63	Broder et al. (1993) [5]	4.883^*	Dubois and Boufkhad (1999) [4]
3.003^*	Frieze and Suen (1996) [14]	4.762^*	Kamath et al. (1995) [16]
3.145	Achlioptas (2000) [1]	4.643^*	Dubois and Boufkhad (1997) [8]
3.26	Achlioptas and Sorkin (2000) [3]	4.602	Kirousis et al. (1998) [18]
3.42	Kaporis, Kirousis and Lalas (2002) [17]	4.596	Janson et al. (1999) [15]
		4.506	Dubois et al. (2003) [9]

Apart from these works, Friedgut [13] also proved that there exists a sequence γ_n such that for any $\varepsilon > 0$, if $m/n < \gamma_n - \varepsilon$ then, the probability of satisfiability tends to 1 as m and n increase while it tends to 0 if $m/n > \gamma_n + \varepsilon$. But it has not been proved that the sequence γ_n converges to some limit value γ, that would then be equal to the threshold r_3.

The aim of the present paper is to present some of the most significant upper bounds on the satisfiability threshold and organize them within a unified framework. We will specifically focus on enumerative proofs, based on generating functions.

2 Expectation of the number of solutions

The first bound for the 3-SAT threshold has been obtained by several authors as a direct application of the first moment method to the random variable giving the number of solutions of a random formula. Under an enumerative perspective, it can be regarded as a direct application of the following simple remark: since positive integer are ≥ 1, one has the inequality:

$$|\{\ \Phi \text{ satisfiable }\}| \leq |\{\ (\Phi, S) \text{ such that } \Phi \text{ is satisfied by } S\ \}| \tag{1}$$

where $|E|$ denotes the cardinality of the set E. There, S represents an assignment of the n variables to values in $\{0,1\}$. Let $C = \pm x_i \vee \pm x_j \vee \pm x_k$ be a clause and fix some assignment C (we identify negation $\neg$ with the minus sign). There is only one way to choose the signs of the three literals in order to have the value of C be false under S: each literal must have the opposite sign of its assignment. Then, there are 7 ways to choose the signs in order to render C true. The number of clauses satisfied by any given S is then $7\binom{n}{3}$. Since S is a solution of a 3-SAT formula Φ iff all clauses of Φ are satisfied by S, for any assignment, there are exactly $(7\binom{n}{3})^m$ formulas with m clauses which admit S as a solution.

The cardinality of the pairs (Φ, S) such that S is a solution of Φ is then given by $2^n(7\binom{n}{3})^m$. Dividing each term of (1) by the total number of formulæ $(8\binom{n}{3})^m$ gives (with $r = m/n$):

$$\Pr(\ \Phi \text{ satisfiable }) \leq \left(2\left(\frac{7}{8}\right)^r\right)^n. \tag{2}$$

For $r > \ln(2)/\ln(8/7) \approx 5.191$, the right side of (2) tends to 0 as n tends to infinity, and so does the probability of satisfiability. This gives the first upper bound obtained by Franco and Paull as early as 1983.

This argument can easily be extended to the problem k-SAT for any k. For any fixed assignment, the number of clauses it satisfies which are disjunctions of k literals is given this time by $(2^k - 1)\binom{n}{k}$. The number of satisfied formulæ is hence $\left((2^k-1)\binom{n}{k}\right)^m$ and since the total number of formulæ is $\left(2^k\binom{n}{k}\right)^m$, one has

$$\Pr(\ \Phi\ k\text{-SAT satisfiable }) \leq \left(2\left(\frac{(2^k-1)}{2^k}\right)^r\right)^n. \tag{3}$$

The probability of satisfiability of k-SAT formulæ tends to 0 when $r > -\ln(2)/\ln(1-2^{-k})$ and the following proposition holds

Proposition 1 (Franco and Paull [12]) *Let m,n be integers and $\Omega_{m,n}$ the set of k-SAT formulae with m clauses constructed over n variables. Let $m,n \to \infty$, while $m/n \geq r$. Then, for any $r > -\ln(2)/\ln(1-2^{-k})$, the expectation of the number of solutions of a random formula tends to 0 and hence*

$$Pr\{\phi \in \Omega_{m,n} \text{is satisfiable}\} \longrightarrow 0.$$

For $k = 3$, the bound is 5.191.

Note that this majorization of the threshold also holds for $k = 2$, but the value $-\ln(2)/\ln(3/4) \approx 2.4$ is appreciably greater than the exact value of the 2-SAT threshold, which has been proved to be 1. It also suggests $2^k \ln(2)$ as an approximate value of the threshold for large k of the potential threshold location.

3 Prime implicants

In the previous section, we have bounded the number of satisfiable formulæ by the number of formula-solution pairs. Since a satisfiable formula may have from 1 to almost 2^n solutions, the upper bound obtained may be very coarse. The next idea is to group some of the solutions which look very close to each other, and enumerate only these groups for each formula. In this way, it may be possible to get an improved upper bound on the satisfiability threshold.

This leads to the idea of using partial assignments and prime implicants, first introduced in this context by Olivier Dubois and Yacine Boufkhad [4]. A *partial assignment* A is simply an assignment to a subset of the n variables (possibly all, so that complete assignment are also considered partial assignments). Let us say that A satisfies a formula Φ iff all complete assignments A' extending A are solutions of Φ. A necessary and sufficient condition for this is that in each clause of Φ, at least one of the three literals is true under A. Note that this implies that at least one variable is assigned. If there are i missing variables in a partial assignment A, then A "groups" 2^i solutions together.

A natural order may be placed on partial assignments. We say that A is smaller than B if we can remove some assigned variables from B to get A. A *prime implicant* of a formula ϕ is then defined as a partial assignment which is minimal with respect to this order. Any satisfiable formula has then at least one prime implicant since it has at least one solution and the set of partial assignments is then non empty. As in the previous section, we get from there the inequality (see (1)):

$$|\{\ \Phi \text{ satisfiable }\}| \leq |\{\ (\Phi, I) \text{ such that } I \text{ is a prime implicant of } \Phi\ \}|. \tag{4}$$

Note that the sets of solutions grouped together by two different partial assignments are not necessarily disjoint and some formulæ may even have more prime implicants than solutions. But in fact, on average, the number of prime implicants of a random formula appears to be smaller by an exponential factor than the number of solutions.

At this point, we start an original re-derivation of the upper bound value 4.883 obtained by O. Dubois and Y. Boufkhad for the 3-SAT threshold, with the help of generating functions, which provides a much shorter proof of their estimate.

Let us consider the partial assignment I which assigns all variables $x_1, \ldots, x_\ell$ to 0. Then, all clauses in a formula Φ that admits I as a prime implicant must contain at least one literal satisfied by I. Any such clause is hence of the form $\neg x_i \vee a \vee b$ with $1 \leq i \leq \ell$. Let $A_{n,\ell}$ be the set of such clauses and $\alpha_{n,\ell}$ their number. Since the complement of $A_{n,\ell}$ is the set of clauses which do not contain any literal $(\neg x_i)_{i\in[\![1\,;\ell]\!]}$, it follows

$$\alpha_{n,\ell} = 8\binom{n}{3} - \left[\binom{\ell}{3} + 2\binom{\ell}{2}(n-\ell) + 4\ell\binom{n-\ell}{2} + 8\binom{n-\ell}{3}\right].$$

Recall that I has to be minimal with respect to the order defined earlier. Remove the variable x_i from the set of assigned variables. Then at least one clause in Φ no longer contains a positive literal. Hence Φ must contains a clause of the form $\neg x_i \vee a \vee b$ with a and b either of the form $(x_i)_{i\in[\![1\,;\ell]\!]}$ or $(\pm x_i)_{i\in[\![\ell+1\,;n]\!]}$. Let $B_{n,\ell,i}$ be the set of these clauses and $\beta_{n,\ell,i}$ their number. It is easy to see that these sets are mutually disjoint and have the same cardinality

$$\beta_{n,\ell,i} = \beta_{n,\ell} = \binom{\ell-1}{2} + 2(\ell-1)(n-\ell) + 4\binom{n-\ell}{2}.$$

Finally, to build a formula which admits I as a prime implicant, one must draw m elements in the set $A_{n,\ell}$ so that in the subsets $(B_{n,\ell,i})_{i\in[\![1\,;\ell]\!]}$, at least one element is taken. The number of such draws is then the number $I_{m,n,\ell}$ whose generating function is given by

$$I_{n,\ell}(z) = \sum_{m\geq 0} I_{n,\ell,m}\frac{z^m}{m!} = e^{z\,(\alpha_{n,\ell}-\ell\beta_{n,\ell})}\left(e^{z\,\beta_{n,\ell}}-1\right)^{\ell}. \tag{5}$$

The proof of this statement is given in the first part of the appendix. It is based on generating functions and the enumeration of throws in urns.

Now, it is clear that for any other partial assignment of ℓ variables, the number of formulæ of length m for which it is a prime implicant is the same as the one where the ℓ first variables are assigned to 0 and thus, it only depends on m, n and ℓ. Finally, since there are $\binom{n}{\ell}2^\ell$ possible partial assignments of ℓ variables, the total number of pairs (Φ, I) such that I is a prime implicant for Φ is given by:

$$|\{\ (\Phi, I \text{ prime implicant })\ \}| = \sum_{\ell=1}^{n}\binom{n}{\ell}2^\ell\, I_{m,n,\ell}. \tag{6}$$

The next step depends on the following general remark: if (f_k) is a sequence of non negative reals and $f(z) = \sum f_k\, z^k$ then for all $s > 0$ within the domain of convergence of $f(z)$:

$$f_k \leq \min_s\ \frac{f(s)}{s^k}.$$

This is a very classical majorization called the saddle-point bound (the minimum of the function $s \mapsto f(s)s^{-k}$ is called the saddle point and the shape of the complex function $\left|f(s)s^{-k}\right|$ in its neighbourhood looks like a horse's saddle). The majorization may seem coarse since

$$f(s)s^{-k} = f_0 s^{-k} + \cdots + f_k + f_{k+1}s + \cdots,$$

but in many cases it gives the exact exponential rate of growth of the coefficient f_k knowing the generating function $f(z)$.

From now on, we set $\ell = \alpha n$, $m = rn$ and make use of the majorizations $\alpha_{n,\ell} - \ell\beta_{n,\ell} \leq \frac{1}{3}\ell^2(3n-\ell)$ and $\beta_{n,\ell} \leq \frac{1}{2}(2n-\ell)^2$. From the previous remark, it follows that for any u_α

$$I_{rn,n,\alpha n} \leq (rn)!\, n^{2m}\left(e^{\alpha^2(1-\alpha/3)u_\alpha}\left(e^{2(1-\alpha/2)^2 u_\alpha}-1\right)^{\alpha} u_\alpha^{-r}\right)^{n}.$$

The same remark proves that for any integer s, one has $\frac{1}{s!} \leq \frac{e^s}{s^s}$. Combining this inequality (for the denominator) with Stirling's formula (for the numerator) ensures that there exists a constant C such that, for each integer n and each α such that αn is an integer, one has

$$\binom{n}{\alpha n} \leq C\sqrt{n}\left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right)^{-n}. \tag{7}$$

Naturally, stronger asymptotic estimates are possible, e.g. $\binom{n}{\alpha n} \sim (2\pi n\alpha(1-\alpha))^{-1/2}\left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right)^{-n}$ follows obviously from three applications of Stirling formula. However, equation (7) wholy suffices for our purposes while making subsequent argumentations simpler.

Finally, since $\frac{(rn)!n^{2rn}}{\left(8\binom{n}{3}\right)^{rn}} \sim \sqrt{2\pi r}e^{3r}n^{1/2}\left(\frac{3r}{4e}\right)^{rn}$, (4) and (6) give

$$\Pr(\ \Phi \text{ satisfiable }) \leq C'n \sum_{\alpha\in\left\{\frac{1}{n},\frac{2}{n},\ldots,\frac{n-1}{n},1\right\}} f_r(\alpha,\delta_\alpha)^n \tag{8}$$

with $$f_r(\alpha,\delta_\alpha) = \left(\frac{3r}{4e}\right)^r \frac{2^{\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\left(e^{\frac{\alpha^2(1-\alpha/3)}{2(1-\alpha/2)^2}\delta_\alpha}\left(e^{\delta_\alpha}-1\right)^{\alpha}\delta_\alpha^{-r}\,2^r\,(1-\alpha/2)^{2r}\right)$$

and where δ_α can be any strictly positive real. In order to get the best majorization, δ_α has to minimize the expression between parentheses. If $0 < \alpha < r$, this expression has a unique minimum δ_α given by

$$\frac{\alpha^2(1-\alpha/3)}{2(1-\alpha/2)^2}\delta_\alpha + \alpha\frac{\delta_\alpha}{1-e^{-\delta_\alpha}} = r.$$

Assume that $r > 1$. Then one gets

$$\Pr(\ \Phi \text{ satisfiable }) \leq C'n^2\left(\max_{\alpha\in]0,1]} g_r(\alpha)\right)^n \qquad \text{with} \qquad g_r(\alpha) = f_r(\alpha,\delta_\alpha).$$

Any real $r > 1$ such that the maximum of g_r over $]0,1]$ is strictly smaller than 1 is hence an upper bound on the satisfiability threshold. The value $r = 4.883$ is then the best value computed with Maple which has this property (details are given in the second part of appendix).

Proposition 2 (Dubois and Boufkhad [4]) *Let m,n be integers and $\Omega_{m,n}$ the set of* 3-SAT *formulae with m clauses constructed over n variables. Let $m,n \to \infty$, while $m/n \geq r$. Then, for any $r > 4.883$, the expectation of the number of prime implicants of a random formula tends to* 0 *and*

$$Pr\{\phi \in \Omega_{m,n} \mathit{is\ satisfiable}\} \longrightarrow 0.$$

The same analysis can easily be extended to k-SAT for any integer k.

4 Negatively prime solutions

The next idea is to introduce a partial order on the set of solutions. Define B to be an assignment smaller than A if one can change the values of some of B's variables from 0 to 1 to get A. We could envisage to enumerate only pairs (Φ,S) where S is a maximal solution with respect to this order; however, it is very difficult to find for any given assignment a simple characterization of formulæ for which it is a maximal solution. In order to remedy this situation, Olivier Dubois and Yacine Boufkhad introduced a weaker definition of *local maximal solution* (also called *negatively prime solution* or NPSs). This is a solution for which changing the value of any variable from 0 to 1 no longer gives a solution. This amounts to considering solutions which do not admit a greater solution that differs in exactly one variable. Once more, there holds the inequality

$$|\{\ \Phi \text{ satisfiable }\}| \leq |\{\ (\Phi, S) \text{ such that } S \text{ is an } \textsc{nps} \text{ of } \Phi\ \}| \tag{9}$$

and the probability of satisfiability is smaller than the expectation of the number of NPSs. What follows now is a new calculation of this expectation.

Let A be an assignment setting ℓ variables to the value 0 and Φ a formula for which A is an NPS . Then, all clauses of Φ must belong to the set A_n of all $7\binom{n}{3}$ clauses satisfied by A (as seen in Section 2). Now, if any variable x_i assigned to 0 is changed to 1, there must be at least one clause in Φ that is no longer satisfied by this new assignment: at least one clause must be thus of the form $\neg x_i \vee a \vee b$, where a and b are false under A. If we denote by C_{x_i} this last set of clauses (for each variable assigned to 0), then all these sets have the same number $\binom{n-1}{2}$ of elements and are mutually disjoint. As in the previous section, since there are $\binom{n}{\ell}$ solutions with ℓ variables assigned to 0, we get:

$$|\{\ (\Phi, A \text{ NPS }) \ \}| = \sum_{\ell=0}^{n} \binom{n}{\ell} m!\ [z^m] e^{z\left(7\binom{n}{3} - \ell\binom{n-1}{2}\right)} \left(e^{z\binom{n-1}{2}} - 1\right)^{\ell}$$

where $[z^m]f(z)$ represents the coefficient f_m in the Taylor expansion $f(z) = \sum f_i z^i$. The linearity of coefficient extraction gives a closed-form expression:

$$|\{\ (\Phi, A \text{ NPS }) \ \}| = m!\ [z^m] e^{z\, 7\binom{n}{3}} \left(2 - e^{-z\binom{n-1}{2}}\right)^{n}. \tag{10}$$

The same use of saddle-bound as in the previous section, the same majorization provided by Stirling's formula, and the change of variable $\delta = z\binom{n-1}{2}$ provide that, for any $\delta > 0$ with $m = rn$,

$$\Pr(\Phi \text{ satisfiable }) \leq C\sqrt{2\pi rn} \left(\left(\frac{3r}{8e}\right)^{r} \frac{e^{\frac{7}{3}\delta}\left(2 - e^{-\delta}\right)}{\delta^{r}} \right)^{n}. \tag{11}$$

This expression is minimized by $\delta\left(\frac{7}{3} - \frac{1}{2e^{\delta}-1}\right) = r$ and, with such a δ, is strictly smaller than 1 as soon as $r > 4.643$. Hence

Proposition 3 (Dubois and Boufkhad [8]) *Let m, n be integers and $\Omega_{m,n}$ the set of 3-SAT formulae with m clauses constructed over n variables. Let $m, n \to \infty$, while $m/n \geq r$. Then, for any $r > 4.643$, the expectation of the number of NPSs of a random formula tends to 0 and*

$$Pr\{\phi \in \Omega_{m,n} \textit{is satisfiable}\} \longrightarrow 0.$$

Once more, this upper bound can be extended to k-SAT for any k. For any assignment giving the value 0 to ℓ variables, the number of satisfied clauses which are disjunction of k literals is this time $(2^k - 1)\binom{n}{k}$. Once more, for all variable x_i assigned to 0, define the set C_{x_i} of clauses which forbid the flip of x_i from 0 to 1. Then, all these sets are mutually disjoint and have the same cardinality $\binom{n-1}{k-1}$. The same arguments then provide the exact formula

$$|\{\ (\Phi\ k\text{-SAT}, A\ \text{NPS}\ )\ \}| = \sum_{\ell=0}^{n} \binom{n}{\ell}\ m!\ [z^m] e^{z\left((2^k-1)\binom{n}{k}-\ell\binom{n-1}{k-1}\right)} \left(e^{z\binom{n-1}{k-1}}-1\right)^{\ell} \tag{12}$$

which gives the general majorization for any $\delta > 0$, considering k-SAT formulæ of length $m = rn$,

$$\Pr(\Phi\ k\text{-SAT satisfiable}\ ) \le C\sqrt{2\pi rn} \left(\left(\frac{kr}{2^k e}\right)^r \frac{e^{\frac{2^k-1}{k}\delta}\left(2-e^{-\delta}\right)}{\delta^r} \right)^n . \tag{13}$$

With $\delta\left(\frac{2^k-1}{k} - \frac{1}{2e^{\delta}-1}\right) = r$, one determines that the probability of satisfiability of k-SAT formulæ of length $m = rn$ tends to 0 as soon as $r > 10.217$ for $k = 4$ and $r > 43.508$ for $k = 6$ for instance. This general upper bound has first been obtained by Olivier Dubois and is still so far the best one for general k. However, the initial proof took 25 pages where ours only needs about 2.

5 Appendix

5.1 Throws into urns and generating functions

Let us consider a general problem. Let $M_1, \dots, M_p$ be subsets of $\mathbb{N}$. One has a collection of p urns and wants to throw in these urns m distinguishable balls (say by integer labels $1, 2, 3, \dots$), such that in the end, the number of balls in urn i is in M_i. For instance, if $M_1 = \{0, 1, 2\}$ and $M_2 = \{2, 3, \dots\}$, one want to throw at most two balls in the first urn and at least two balls in the second one. The problem is to enumerate the number of such throws.

Let F_m be this number. Then clearly, one has

$$F_m = \sum_{n_1+\cdots+n_p=m} \binom{m}{n_1;\cdots;n_p} [\![n_1 \in M_1]\!] \times \cdots \times [\![n_p \in M_p]\!] \tag{14}$$

where $[\![\mathcal{P}]\!]$ equals 1 if $\mathcal{P}$ is true and equals 0 otherwise. The binomial coefficient enumerates all ways to put n_1 balls in urn 1, n_2 balls in urn 2 and so on. The product is then equal to 1 if and only if all conditions concerning the final number of balls in each urn are satisfied.

Introduce the exponential generating function of the sequence (F_m). Then, (14) immediately implies

$$F(z) = \sum_{m\in\mathbb{N}} F_m \frac{z^m}{m!} = \left(\sum_{i\in M_1} \frac{z^i}{i!}\right) \times \cdots \times \left(\sum_{i\in M_p} \frac{z^i}{i!}\right). \tag{15}$$

For instance, the generating function of the example above is $F(z) = (\frac{z^2}{2} + z + 1)(e^z - z - 1)$.

Let us now consider a slightly more complex problem. We now have p subsets of urns, say $E_1, \dots, E_p$ of cardinality $e_1, \dots, e_p$, and p subsets of $\mathbb{N}$ $M_1, \dots, M_p$. This time, we want to throw m distinguishable balls in all these urns such that in the end, for any i, the total number of balls in the subset E_i of urns is an element of M_i. The formula (14) now becomes

$$F_m = \sum_{n_1+\cdots+n_p=m} \binom{m}{n_1;\cdots;n_p} [\![n_1 \in M_1]\!]\ (e_1)^{n_1}\ \cdots [\![n_p \in E_p]\!]\ (e_p)^{n_p}. \tag{16}$$

since there is now $(e_i)^{n_i}$ ways to throw n_i balls in the subset E_i of urns, and the translation of this formula into generating function gives

$$F(z) = \sum_{m\in\mathbb{N}} F_m \frac{z^m}{m!} = \left(\sum_{i\in M_1} \frac{(e_1 z)^i}{i!}\right) \times \cdots \times \left(\sum_{i\in M_p} \frac{(e_p z)^i}{i!}\right). \tag{17}$$

In section 3, we had to pick elements in one set $A_{n,\ell}$, so that in ℓ subsets $B_{n,\ell,i}$, at least 1 element is picked. Let $\tilde{A}_{n,\ell} = A_{n,\ell} - (\cup_{1\le i\le \ell} B_{n,\ell,i})$. This is equivalent to picking any number of elements in $\tilde{A}_{n,\ell}$ and at least one in each of the $B_{n,\ell,i}$.

"Replace" now each clause by an urn. Picking m clauses with replacement to construct a formula is then equivalent to throwing m distinguishable balls in the set of urns. The enumeration of such draws can thus be regarded as an urn problem just defined, where we dispose of $\ell+1$ sets of urns. The first set contains $\tilde{\alpha}_{n,\ell} = \alpha_{n,\ell} - \ell\beta_{n,\ell}$ urns and we may throw any number of balls in it (hence $M_1 = \mathbb{N}$). The other sets each contain $\beta_{n\ell}$ urns and we must throw at least one ball in one urn of the set (hence $M_2 = \cdots = M_{\ell+1} = \mathbb{N}^*$). The equation (17) finally gives the desired result.

5.2 Maple computations

In this section, we prove with the help of Maple that for $r = 4.883$, the maximum of g_r defined in section 3 is strictly smaller than 1. The idea is to get a coarse majorization of g_r' and then to ask Maple to evaluate g_r at sufficiently many points $x_0 < \cdots < x_p$. With the mean value theorem, the condition

$$1 - \max_{i\in[\![0\,;\,p]\!]} \{g_r(x_i)\} > \frac{||g_r'||}{2} \max_{i\in[\![0\,;\,p-1]\!]} |x_{i+1} - x_i| \tag{18}$$

ensures that for any $t \not\in \{x_0, \ldots, x_p\}$, it also holds that $g_r(t) < 1$. The purpose of this section is hence to replace many pages of analysis by Dubois and Boufkhad by a combination of straightforward inequalities and trivial computation.

A minor difficulty arises from the fact that the function $\alpha \mapsto \frac{2^\alpha}{\alpha^\alpha(1-\alpha)^{1-\alpha}}$ has infinite derivatives at 0 and 1. Hence, a rough majorization of g_r on intervals of the form $]0,\varepsilon]$ and $[1-\eta, 1]$ must be obtained first.

Monotonicity properties : The function g_r can be written as a product of one constant $\left(\frac{3r}{2e}\right)^r$ and two functions which are:

$$f_1 : \alpha \mapsto \frac{2^\alpha}{\alpha^\alpha(1-\alpha)^{1-\alpha}}\left(1-\frac{\alpha}{2}\right)^{2r} \qquad f_2 : \alpha \mapsto \exp\left(\frac{\alpha^2(1-\alpha/3)}{2(1-\alpha/2)^2}\delta_\alpha + \alpha \ln\left(e^{\delta_\alpha} - 1\right) - r \ln\, \delta_\alpha\right).$$

First of all, one can prove that δ_α and f_2 are both monotonic functions. Let us define

$$h(x,y) = \frac{x^2(1-x/3)}{2(1-x/2)^2} y + x\frac{y}{1-e^{-y}}.$$

Since $h(\alpha, \delta_\alpha) = r$, differentiating this expression with respect to α yields

$$\frac{\partial h}{\partial x}(\alpha,\delta_\alpha) + \delta_\alpha' \frac{\partial h}{\partial y}(\alpha,\delta_\alpha) = 0 \qquad \Rightarrow \qquad \delta_\alpha' = \left(-\frac{\frac{\partial h}{\partial x}}{\frac{\partial h}{\partial y}}\right)(\alpha,\delta_\alpha).$$

It now easy to verify that both $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$ are positive function on $[0,1]\times]0,+\infty[$ and hence, δ_α is a decreasing function of α. Concerning the function f_2, the condition $h(\alpha,\delta_\alpha)=r$ also implies

$$f_2'(\alpha) = \left(\frac{\alpha(12-6\alpha+\alpha^2)}{3(2-\alpha)^3}\delta_\alpha + \ln(e^{\delta_\alpha}-1)\right) f_2(\alpha).$$

The function $\alpha \mapsto \delta_\alpha$ being decreasing, for any α, $\ln(e^{\delta_\alpha}-1) \geq \ln(e^{\delta_1}-1) \approx 1.8017 > 0$. Hence, f_2' is positive and f_2 is an increasing function of α.

Finally, the function f_1 satisfies

$$f_1'(\alpha) = \left(\ln(2)-\ln(\alpha)+\ln(1-\alpha)-\frac{2r}{2-\alpha}\right) f_1(\alpha).$$

The expression between parentheses is smaller than $\ln(2)-\ln(\alpha)+\ln(1-\alpha)-r$, which is negative as soon as $\alpha > \frac{2}{e^r+2} \sim 0.0149$. The function f_1 is thus decreasing on $[0.015,1]$.

Majorization of $\mathbf{g_r}$ on $]0,0.897]$ and $[0.957,1]$: First of all, we make use of the majorization

$$\frac{2^\alpha}{\alpha^\alpha(1-\alpha)^{1-\alpha}}\left(1-\frac{\alpha}{2}\right)^{2r} \leq 3 \qquad \forall\, \alpha \in [0,1]$$

which implies that for any $\alpha \in]0,0.414]$, $g_r(\alpha) \leq 3\left(\frac{3r}{2e}\right)^r f_2(0.414) \approx 0.992$. Then we use the fact that f_1 is decreasing on $[0.015,1]$ to conclude that on any interval $[a,b]$ with $a > 0.015$,

$$\text{for any } \alpha \in [a,b], \qquad g_r(\alpha) \leq \left(\frac{3r}{2e}\right)^r f_1(a) f_2(b).$$

By subdividing $[0.015,1]$ into many small intervals, it is shown with this majorization that g_r is smaller than 1 on both intervals $]0,0.897]$ and $[0.957,1]$.

Majorization of $\mathbf{g_r'}$ on $[0.897,0.957]$: For any $\alpha \in [0.897,0.957]$, one has $\delta_{0.957} \leq \delta_\alpha \leq \delta_{0.897}$. Differentiating g_r with respect to α gives

$$g_r'(\alpha) = \frac{\partial f_r}{\partial \alpha} + \delta_\alpha' \frac{\partial f_r}{\partial \delta_\alpha} = \left(\frac{\partial f_r}{\partial \alpha} - \frac{\frac{\partial h}{\partial x}}{\frac{\partial h}{\partial y}}\frac{\partial f_r}{\partial \delta_\alpha}\right)(\alpha,\delta_\alpha).$$

It is easy to see that both $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$ are positive increasing functions of x and y so their respective greatest and lowest values are $\frac{\partial h}{\partial x}(0.957,\delta_{0.897})$ and $\frac{\partial h}{\partial y}(0.897,\delta_{0.957})$. To majorize $\frac{\partial f_r}{\partial \alpha}$ and $\frac{\partial f_r}{\partial \delta_\alpha}$, we introduce a_r such that $f_r = \exp(a_r)$. Then

$$\begin{aligned} a_r(\alpha,\delta_\alpha) &= C'' + \alpha\ln(2) - \alpha\ln(\alpha) - (1-\alpha)\ln(1-\alpha) + \frac{\alpha^2(1-\alpha/3)}{2(1-\alpha/2)^2}\delta_\alpha \\ &\quad + \alpha\ln(e^{\delta_\alpha}-1) - r\ln(\delta_\alpha) + 2r\ln(1-\alpha/2), \end{aligned}$$

and we have both $\frac{\partial f_r}{\partial x} = f_r \frac{\partial a_r}{\partial x}$ and $\frac{\partial f_r}{\partial y} = f_r \frac{\partial a_r}{\partial y}$. The derivatives of a_r are simple functions, and finally, a coarse majorization of f_r, $\frac{\partial a_r}{\partial x}$ and $\frac{\partial a_r}{\partial x}$ gives the bound:

$$\text{for any } \alpha \in [0.897,0.957], \qquad \left|g_r'(\alpha)\right| \leq 60.$$

Numerical evaluations: We set $r = 4.883$. We got Maple to compute the greatest value of g_r with an accuracy of 15 digits for points in the interval $[0.897, 0.957]$ at distance at most $0.6 \cdot 10^{-7}$ of each other (10^6 values). The greatest value returned was $1 - 0.2721 \cdot 10^{-5}$. Hence, the condition (18) is verified and the value 4.883 is proved to be a valid upper bound of the satisfiability threshold.

Acknowledgements

I want to thank Olivier Dubois for all the time he took to explain his articles to me, and Philippe Flajolet for pointing to generating functions as a useful tool for enumeration in threshold phenomena. Thanks also to an anonymous referee for many constructive and encouraging remarks.

References

[1] D. Achlioptas. Setting 2 variables at a time yields a new lower bound for random 3-sat. In *Proceeding of the 32nd ACM Symposium on Theory of Computing, Association for Computing Machinery*, pages 28–37, 2000.

[2] D. Achlioptas. Lower bounds for random 3-SAT via differential equations. *Theoret. Comput. Sci.*, 265(1-2):159–185, 2001. Phase transitions in combinatorial problems (Trieste, 1999).

[3] D. Achlioptas and G. B. Sorkin. Optimal myopic algorithms for random 3-SAT. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pages 590–600. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.

[4] Y. Boufkhad and O. Dubois. Length of prime implicants and number of solutions of random CNF formulae. *Theoret. Comput. Sci.*, 215(1-2):1–30, 1999.

[5] A. Z. Broder, A. M. Frieze, and E. Upfal. On the satisfiability and maximum satisfiability of random 3-CNF formulas. In *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, TX, 1993)*, pages 322–330, New York, 1993. ACM.

[6] M.-T. Chao and J. Franco. Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k-satisfiability problem. *Information Sciences. An International Journal*, 51(3):289–314, 1990.

[7] O. Dubois. Upper bounds on the satisfiability threshold. *Theoret. Comput. Sci.*, 265(1-2):187–197, 2001. Phase transitions in combinatorial problems (Trieste, 1999).

[8] O. Dubois and Y. Boufkhad. A general upper bound for the satisfiability threshold of random r-SAT formulae. *Journal of Algorithms*, 24(2):395–420, 1997.

[9] O. Dubois, M. Jacques, and Y. Boufkhad. Typical 3-sat formulae and the satisfiability threshold. In *Electronic Colloquium on computational complexity*, 2003.

[10] A. El Maftouhi and W. Fernandez de la Vega. On random 3-sat. *Combinatorics, Probability and Computing*, 4(3):189–195, 1995.

[11] J. Franco. Results related to threshold phenomena research in satisfiability: lower bounds. *Theoret. Comput. Sci.*, 265(1-2):147–157, 2001. Phase transitions in combinatorial problems (Trieste, 1999).

[12] J. Franco and M. Paull. Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5(1):77–87, 1983.

[13] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[14] A. Frieze and S. Suen. Analysis of two simple heuristics on a random instance of k-SAT. *Journal of Algorithms*, 20(2):312–355, 1996.

[15] S. Janson, Y. C. Stamatiou, and M. Vamvakari. Erratum to: "Bounding the unsatisfiability threshold of random 3-SAT" [Random Structures Algorithms **17** (2000), no. 2, 103–116; MR 2001c:68065]. *Random Structures and Algorithms*, 18(1):99–100, 2001.

[16] A. Kamath, R. Motwani, K. Palem, and P. Spirakis. Tail bounds for occupancy and the satisfiability threshold conjecture. *Random Structures and Algorithms*, 7(1):59–80, 1995.

[17] A. C. Kaporis, L. M. Kirousis, and E. G. Lalas. The probabilistic analysis of a greedy satisfiability algorithm. In M. R. and R. R., editors, *ESA 2002: 10th annual european symposium*, volume 2461 of *Lecture Notes in Computer Science*, pages 574–585, Rome, Italy, 2002. Springer.

[18] L. M. Kirousis, E. Kranakis, D. Krizanc, and Y. C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures and Algorithms*, 12(3):253–269, 1998.

[19] L. Meunier. Étude des transitions de phases en K-satisfaisabilité. Master's thesis, École polytechnique, 2000.

[20] C. V. and R. B. Mick gets some (the odds are on his side). In *33rd Annual Symposium on Foundations of Computer Science*, pages 620–627, Pittsburgh, Pennsylvania, October 1992. IEEE.