Theoretical Informatics and Applications
Theoret. Informatics Appl. **37** (2003) 127–147
DOI: 10.1051/ita:2003014

SMOOTH AND SHARP THRESHOLDS FOR RANDOM k-XOR-CNF SATISFIABILITY

Nadia Creignou[1] and Hervé Daudé[2]

Abstract. The aim of this paper is to study the threshold behavior for the satisfiability property of a random k-XOR-CNF formula or equivalently for the consistency of a random Boolean linear system with k variables per equation. For $k \geq 3$ we show the existence of a sharp threshold for the satisfiability of a random k-XOR-CNF formula, whereas there are smooth thresholds for $k = 1$ and $k = 2$.

Mathematics Subject Classification. 05C80, 68R05, 60C05.

1. Introduction

Threshold phenomena were first observed for random graphs by Erdös and Rényi [8]. They observed that for many interesting properties P the probability of P appearing in a random graph exhibits a sharp increase at a certain critical value of the edge probability. This threshold behavior occurs in various settings which arise in combinatoric and computer science. This behavior is of interest from a practical point of view since it has been observed [11, 20] that the hard instances of NP-hard problems (which provide a challenging test material in order to study average complexity of algorithms) are often associated with a phase transition.

Among all the problems for which phase transition has been studied, the satisfiability of k-CNF formulas, k-SAT for short, has sparked a lot of interest. Experiments on solving k-CNF random formulas (see for instance [15]) have provided evidence of the existence of a satisfiability threshold phenomenon with respect to

Keywords and phrases. Threshold phenomenon, satisfiability, phase transition, random Boolean linear systems.

[1] LIF, UMR 6166 du CNRS, Université de la Méditerranée, 163, avenue de Luminy, 13288 Marseille, France; e-mail: creignou@lif.univ-mrs.fr

[2] LATP, UMR 6632 du CNRS, Université de Provence, 39 rue Joliot-Curie, 13453 Marseille, France; e-mail: daude@gyptis.univ-mrs.fr

the ratio, c_k, of the number of clauses to the number of variables of formulas. Most of the papers investigating the existence of a phase transition for this problem are directed towards obtaining approximate estimates of its location. For instance for 3-SAT, for an observed sharp threshold of about $c_3 = 4.25$ the best lower bound is 3.003 [10] and the best upper bound is 4.506 [7], so there still exists a large gap between lower and upper bounds. A sharp threshold has been established for 2-SAT [5, 12], the critical value is $c_2 = 1$. Observing that 2-SAT is a special case of satisfiability which is solvable in polynomial time [2], in [6] we suggested to investigate all the tractable cases identified by Schaefer [22]. We studied the phase transition for the XOR-SAT problem, in which the "exclusive or" is used instead of the "usual or". We established a sharp threshold phenomenon with critical value $c = 1$ and more precisely described the probability distribution of the phase transition [6]. Our result concerned formulas with clauses of arbitrary length. So, a natural question arises: is there a comparable result for XOR-SAT with fixed clause-length formulas, namely k-XOR-SAT?

In this paper we study the threshold for the satisfiability of k-XOR-CNF formulas. Section 2 is a preliminary work devoted to two simple conditions assuring a sharp threshold. Based on correlation inequalities, this is a synthetic version of Friedgut's and Bourgain's results [9] in the general context of monotone properties of the hypercube. Then, we first prove (Sect. 3.2) that there are smooth threshold phenomena for $k = 1$ and $k = 2$ with a well-described probability distribution. The tools are well-known and come from classical random graphs' theory. In Section 4 we prove that there is a sharp threshold for $k \geq 3$ in using our preliminary work. This last result will need the most effort, we have to verify two conditions. The second condition is hard to study, we describe the strategy that can be used in order to tackle this point and we state three lemmas that concentrate the combinatorial and the probabilistic difficulties. We postpone the proof of these technical key lemmas in Section 5.

On one hand this complete exposure of new results on random k-XOR-SAT will serve for a didactical presentation of sharp threshold investigations. On the other hand let us observe that our results on k-XOR-SAT analyze the threshold behavior of the consistency of linear systems over the finite field $GF(2)$, a topic of independent interest that has been widely studied (see for example [16, 18, 19]).

2. Thresholds of monotone properties

2.1. Terminology

Let us introduce some terminology on thresholds of monotone properties.

Let N be a positive integer. Let s and s' be two vectors in $\{0,1\}^N$. The sum $s \oplus s'$ denotes the componentwise XOR-operation on s and s'. One says that s contains s', $s \geq s'$, if $s_i \geq s'_i$ holds for each coordinate $i = 1, \ldots, N$. Vectors from $\{0,1\}^N$ can be interpreted as characteristic functions over a set of cardinality N. Thus if s contains s', then the intended meaning of $s \oplus s'$ is the difference set $s \setminus s'$. A nonempty subset E of the set of vectors $\{0,1\}^N$ is called monotone increasing,

or simply monotone, if it satisfies the following condition: if $s' \in E$ and $s \in \{0,1\}^N$, such that $s \geq s'$, then $s \in E$. A monotone subset of $\{0,1\}^N$ that contains $\vec{0}$ is equal to $\{0,1\}^N$ and is called trivial.

For $0 \leq p \leq 1$, let $\mu_p \colon \{0,1\}^N \to \{0,1\}$ be the function which satisfies the equality

$$\mu_p(\{s\}) = (1-p)^{N-w(s)} p^{w(s)}$$

for each vector $s \in \{0,1\}^N$, where $w(s)$ denotes the Hamming weight of s. This function defines the so-called product measure μ_p on $\{0,1\}^N$, thus for any set E of 0-1 vectors:

$$\mu_p(E) = \sum_{s \in E} \mu_p(\{s\}).$$

The question of understanding how $\mu_p(E)$ varies with p is of principal interest. For instance the well-known random graph model $G(n,p)$ is the probability space over the set of graphs on n vertices where each edge appears independently with probability $p = p(n)$. In this case $N = \binom{n}{2}$ is the number of edges of the complete graph K_n on n vertices, and any subgraph G of K_n is encoded by some $x \in \{0,1\}^N$ ($x_i = 1$ if an only if the edge number i is in G). Given any graph theoretic property P the probability that $G(n,p)$ satifies P is nothing else but $\mu_p(P_N)$, where P_N assembles the subgraphs of K_n having property P. For many properties P of importance, there is a threshold effect, in the sense that $\mu_p(P_N)$ jumps from 0 to 1 in a small interval.

More generally, let $A \subseteq \{0,1\}^*$ be a monotone property, *i.e.* for every N, $A_N := A \cap \{0,1\}^N$ is monotone. (It is said to be nontrivial if for any sufficiently large N, A_N is nontrivial.) Set $f_N(p) = \mu_p(A_N)$. A quite natural result is:

Proposition 2.1 [3]**.** *If A is a nontrivial monotone increasing property, then for every N and for $0 \leq p_1 < p_2 \leq 1$ we have $f_N(p_1) < f_N(p_2)$.*

Thus, each $f_N(p)$ defines an increasing one-to-one correspondence from $[0,1]$ onto $[0,1]$. Let us write $p_c(N) = f_N^{-1}(c)$. For $c = 1/2$, $p_{1/2}(N)$ is called the *critical probability* of A. Let $\varepsilon \in (0,1/2]$, the interval $\tau(N) = [p_\varepsilon(N), p_{1-\varepsilon}(N)]$ is called the *threshold interval*. Let $\delta_\varepsilon(N)$ denote its length, $\delta_\varepsilon(N) = p_{1-\varepsilon}(N) - p_\varepsilon(N)$. One says that A has a *sharp threshold* if for every $\varepsilon \in (0,1/2]$ the ratio $\frac{\delta_\varepsilon(N)}{p_{1/2}(N)}$ tends to 0 as N tends to infinity. (Intuitively it means that $f_N(p)$ jumps from near 0 to near 1 in an interval which is small with respect to the critical probability when N tends to infinity.) If for some $\varepsilon > 0$ and for all N, the ratio $\frac{\delta_\varepsilon(N)}{p_{1/2}(N)}$ is bounded away from 0, then one says that A has a *smooth threshold*.

2.2. A general sharp threshold criterion

In a remarkable paper [9] Friedgut and Bourgain developed a general sharp threshold criterion for monotone subsets of the hypercube. Roughly speaking their result says that if a monotone property A is not influenced by elements of bounded weight, whatever these elements are in A or not, then A has a sharp threshold. Thus two conditions appear in their criterion. Taking into account the fact that A

is monotone the first condition says that elements from A of bounded weight have a negligible probability to appear. The second one says that the probability of being in A is not significantly modified when conditioning on the appearance of a given element, s_0, not in A and of bounded weight. Evidently these conditions have to be verified asymptotically and in the scaling window, namely for any $p = p_c$ where $\mu_{p_c}(s \in A) = c$ for any parameter c in $(0, 1)$.

Theorem 2.2 [9]**.** *Let A be a monotone property such that $p_{1/2}(N) = o(1)$. If the two following conditions are verified, then A has a sharp threshold.*

($C1$) *For each parameter $c \in (0, 1)$ and all positive integers K,*

$$\mu_{p_c(N)}(s \geq s' \text{ where } s' \in A \text{ and } w(s') \leq K) \xrightarrow[N\to+\infty]{} 0.$$

($C2$) *For each parameter $c \in (0, 1)$, for all positive integers K and all $s_0 \notin A$ with $w(s_0) = K$,*

$$\mu_{p_c(N)}(s \in A \mid s \geq s_0) \xrightarrow[N\to+\infty]{} c.$$

The question is how can we verify these two conditions for a specific property A. We propose a new criterion that is easier to verify and thus that is useful, as we will see for k-XOR-SAT, for investigating the phase transition for many monotone properties. The following result shows first that the key in order to verify ($C1$) is to study minimal elements of such properties: an element s from A is *minimal* if for all s' contained in s and different from s, $s' \notin A$. Second, in using two well-known correlation inequalities on monotone properties we render the second condition easier to handle.

Theorem 2.3. *Let A be a monotone property such that $p_{1/2}(N) = o(1)$. If the two following conditions are verified, then A has a sharp threshold.*

($C'1$) *For each parameter $c \in (0, 1)$ and all positive integers K,*

$$\mu_{p_c(n)}(s \geq m,\ m \text{ is minimal for } A \text{ and } w(m) \leq K) \xrightarrow[N\to+\infty]{} 0.$$

($C'2$) *For each parameter $c \in (0, 1)$, for all positive integers K and all $s_0 \notin A$ with $w(s_0) = K$,*

$$\mu_{p_c(N)}(s \in A, s \oplus s_0 \notin A \mid s \geq s_0) \xrightarrow[N\to+\infty]{} 0.$$

Proof. Observe that a vector s contains a vector s' from A of Hamming weight bounded by K if and only if s contains such a minimal element. Thus ($C'1$) is equivalent to ($C1$).

The following inequalities

$$c \leq \mu_{p_c(N)}(s \in A \mid s \geq s_0) \leq c + \mu_{p_c(N)}(s \in A, s \oplus s_0 \notin A \mid s \geq s_0),$$

show that $(C'2)$ implies $(C2)$. As stated in the following lemma these inequalities are consequences of two well-known correlation inequalities on monotone properties (see [13]). □

Lemma 2.4. *Let A and B be two monotone increasing properties, $p \in [0,1]$ and N be any positive integer, we have*

$$\mu_p(A_N) \leq \mu_p(A_N \mid B_N) \leq \mu_p(A_N) + \mu_p((A_N \cap B_N) \setminus (A_N \circ B_N) \mid B_N)$$

where $A_N \circ B_N = \{s,$ there exists $s' \in A_N$ with $s \geq s'$ and $s \oplus s' \in B_N\}$.

Proof.

$$\mu_p(A_N \mid B_N) = \frac{\mu_p(A_N \cap B_N)}{\mu_p(B_N)}.$$

The first inequality is nothing else but the F.K.G. inequality for increasing events (due to Fortuin, Kasteleyn and Ginibre):

$$\mu_p(A_N \cap B_N) \geq \mu_p(A_N)\mu_p(B_N).$$

Now, since for increasing events, $A_N \circ B_N \subset A_N \cap B_N$, we get:

$$\mu_p(A_N \cap B_N) = \mu_p(A_N \circ B_N) + \mu_p((A_N \cap B_N) \setminus (A_N \circ B_N));$$

and the second inequality comes from the B.K. inequality for increasing events (due to Berg and Kesten):

$$\mu_p(A_N \circ B_N) \leq \mu_p(A_N)\mu_p(B_N).$$ □

3. k-XOR-SAT: A challenging monotone property

In this section, we study the threshold phenomenon associated to the satisfiability of k-XOR-CNF formulas, or equivalently to the consistency of linear systems over $GF(2)$, k being the fixed number of variables per equations in such systems. First, in the natural background of linear algebra, we make precise specific probabilistic and combinatorial tools needed to investigate the phase transition associated to the monotone property k-XOR-SAT. We completely describe this phase transition for the particular cases $k = 1$ and $k = 2$ (in this last case the natural translation of 2-XOR-SAT into a graph property will be the key fact). Such a complete description is far from being accessible when $k \geq 3$ for the underlying combinatorial structure is that of hypergraphs. However, we will prove the sharpness of the threshold for k-XOR-SAT ($k \geq 3$) in the last sections.

3.1. Probabilistic model

A *k-XOR-clause* (or shortly a *k-equation*), C, is a linear equation over the finite field $GF(2)$ using exactly k variables, $C = ((x_1 \oplus \ldots \oplus x_k) = \varepsilon)$ where $\varepsilon = 0$ or 1. A *k-XOR-formula* (or shortly a *k-system*) is a conjunction of distinct k-XOR-clauses. A *truth assignment* I is a mapping that assigns 0 or 1 to each variable in its domain, it satisfies an XOR-clause $C = ((x_1 \oplus \ldots \oplus x_k) = \varepsilon)$ if and only if $I(C) := \sum_{i=1}^{p} I(x_i) \bmod 2 = \varepsilon$, and it satisfies a formula F iff it satisfies every clause in F.

We will denote by k-XOR-SAT (or shortly SAT) the property for a k-XOR-formula of being satisfiable (or equivalently the property for a k-system of being consistent) and by UNSAT the property of being unsatisfiable. The property UNSAT is monotone increasing.

Throughout the paper we reserve n for the number of variables ($\{x_1, \ldots, x_n\}$ denotes the set of variables). There are $\binom{n}{k}$ ways to choose a subset of k variables from the given set of variables $\{x_1, \ldots, x_n\}$. Each subset determines a sum which is a left-hand side of an affine equation. This sum can be put equal to 0 or to 1. Hence each subset of k variables implies two possible affine equations. Therefore there are together $N_k = 2\binom{n}{k}$ different k-XOR-clauses over n variables. Let $S^k(n,p) \in \{0,1\}^{N_k}$ denote the random formula on n variables where each k-equation appears independently with probability p. Thus a random formula is represented by a vector $s \in \{0,1\}^{N_k}$, such that for all coordinate $i = 1, \ldots, N$, the value of s_i is 1 if the ith k-XOR-equation appears in s and 0 otherwise. The Hamming weight of s, $w(s)$, represents the number of equations occurring in s.

We will denote by $SAT_n^{(k)}(p)$ the probability that the random formula $S^k(n,p)$ is satisfiable:

$$SAT_n^{(k)}(p) = \mu_p(k\text{-XOR-SAT}).$$

We are interested in studying the asymptotic behavior of this probability, when N_k (or equivalently n) tends to infinity. By abuse of notation we write $p_c(n)$ instead of $p_c(N_k)$. Throughout the paper we will assume, whenever it is needed, that the number n of variables we have is sufficiently large.

This probabilistic model is analogous to the random graph model $G(n,p)$ and is very handy. For instance the weight of the random k-system, $w(S^k(n,p))$, has binomial distribution of parameters N_k and p. Hence, one may keep in mind that $N_k.p$ is the average number of equations (*i.e.* the average weight) of a random k-system. It is easy to prove that a large number of equations is needed to observe the transition as shown by the following estimate, which is far from being tight but that will be sufficient for our purpose.

Lemma 3.1 [Lower bound lemma]**.** *If the number of equations is smaller than the square root of the number of variables, then the random linear system is almost surely satisfiable:*

$$\mu_p(UNSAT) = o(1) \text{ as soon as } N_k.p = o(\sqrt{n}).$$

Proof. Let us denote by S_1 the part of $S^k(n,p)$ corresponding to the equations having a fixed variable, say x_1. Then, $w(S_1)$ has binomial distribution of parameters N_{k-1} and p. Let U_i denotes the event that x_i appears at most once in the k-system, then $\mu_p(\cap_i U_i) \leq \mu_p(SAT)$ and by symmetry we get: $\mu_p(SAT) \geq 1-n(1-\mu_p(U_1))$. Observe now that $\mu_p(U_1) = \mu_p((S_1) \geq 2)$. Since $w(S_1)$ has binomial distribution, it is easy to check that $(1-\mu_p(U_1)) = o(\frac{1}{n})$ as soon as $N_k.p = o(\sqrt{n})$, thus in this case $\mu_p(UNSAT) = o(1)$. □

3.2. Smooth thresholds for 1-XOR-SAT and 2-XOR-SAT

Theorem 3.2. *The critical probability for the 1-XOR-SAT property verifies $p_c(n) = \theta(1/\sqrt{n})$ and 1-XOR-SAT has a smooth threshold. More precisely, taking $p(n) = t/\sqrt{n}$ for any positive constant t, we get*

$$\lim_{n\to+\infty} SAT_n^{(1)}\left(\frac{t}{\sqrt{n}}\right) = \mathrm{e}^{-t^2}.$$

Proof. Let $S^1(n,p)$ be the random 1-system. It is satisfiable if and only if for every variable x the two equations $(x=0)$ and $(x=1)$ do not both appear in the formula. Since for every x the probability of this event is $(1-p^2)$ and since the equations are drawn independently we have $SAT_n^{(1)}(p) = (1-p^2)^n$. Therefore, if $p = t/\sqrt{n}$ then $SAT_n^{(1)}(p) = \left(1-\frac{t^2}{n}\right)^n$ and $\lim_{n\to+\infty} SAT_n^{(1)}\left(\frac{t}{\sqrt{n}}\right) = \mathrm{e}^{-t^2}$. □

The satisfiability of 2-systems is strongly related to the existence of cycles in graphs. Indeed, suppose we are given a 2-system s in $\{0,1\}^{N_2}$. We construct a graph $G(s)$ with n vertices and $w(s)$ weighted edges. For each variable x_i we have a vertex in $G(s)$. For each equation $x_i \oplus x_j = \varepsilon$ we add the edge $\{x_i, x_j\}$ to $G(s)$ with the weight ε.

Lemma 3.3. *The 2-system s is satisfiable if and only if $G(s)$ does not contain any elementary cycle with odd weight.*

Proof. Clearly, every elementary cycle with odd weight in $G(s)$ corresponds to an unsatisfiable subsystem in s. Conversely, we can use a depth-first-forest of $G(s)$ to find a solution of s. Indeed, for each depth-first-tree choosing a Boolean value for the root determines exactly the Boolean value for every variable in the tree. Since $G(s)$ does not contain any elementary cycle with odd weight the assignment so obtained satisfies all the equations of s. □

The asymptotic behavior of the number of cycles in random graphs has been first investigated by Erdös and Rényi [8], and made precise by Janson [14] and Takacs [23]. This number converges in distribution to a Poisson law of parameter $\lambda = \sum_{l\geq 3} \lambda_l$ where λ_l is the limit of the average number of cycles of length l. The proofs of these authors can easily be used in our context (with only a slight

modification to take into account the weight) to estimate the probability that there is no cycle of odd weight in the random graph $G(s)$ associated with the random 2-system s.

Theorem 3.4. *The critical probability for the 2-XOR-SAT property verifies $p_c(n) = \theta(1/n)$ and 2-XOR-SAT has a smooth threshold. More precisely, taking $p(n) = t/n$ for any positive constant t, we get*

$$\lim_{n\to+\infty} SAT_n^{(2)}\left(\frac{t}{n}\right) = \mathrm{e}^{t/2}(1-2t)^{1/4} \text{ if } 0 < t < 1/2, \text{ and } 0 \text{ otherwise.}$$

Proof. Let X_c be the Bernoulli random variable that indicates whether $G(s)$ contains the cycle c of odd weight. The expectation of X_c is $E[X_c] = p^l$, where l is the length of c. Let us denote by Y_l the random variable that counts the number of cycles of odd weight and of length l:

$$Y_l = \sum_{c \text{ of length } l} X_c.$$

By means of the linearity of expectation

$$E[Y_l] = \sum_{c \text{ of length } l} E[X_c].$$

Since there are $\dfrac{n(n-1)\dots(n-l+1)2^{l-1}}{2l}$ possible cycles of odd weight and fixed length l, we get

$$E[Y_l] = \frac{n(n-1)\dots(n-l+1)2^{l-1}}{2l}p^l.$$

When $p = t/n$ (and $0 < t < 1/2$), $E[Y_l] \sim \dfrac{(2t)^l}{4l}$. The random variable Y_l is the sum of the rare events $X_c = 1$ (for such an event appears with probability p^l), thus as in [14] and [23] the asymptotic behavior of Y_l is given by a Poisson law whose parameter is $\dfrac{(2t)^l}{4l}$. Let us denote by Y the number of cycles of odd weight

$$Y = \sum_{l\geq 2} Y_l.$$

The asymptotic behavior of Y is also given (see [14] and [23]) by a Poisson law whose parameter is $\lambda_t = \sum_{l\geq 2} \frac{(2t)^l}{4l} = \frac{1}{4}\ln(1-2t) + \frac{t}{2}$. More formally,

$$Y \xrightarrow{d} P_0(\lambda_t).$$

In particular, the probability that there is no cycle of odd weight is $\exp(-\lambda_t)$, that is $\mathrm{e}^{t/2}(1-2t)^{1/4}$. □

4. A sharp threshold for k-XOR-SAT, $k \geq 3$

This section is devoted to the statement and the proof of our main result, namely the existence of a sharp threshold phenomenon for k-XOR-SAT, $k \geq 3$. Using the first moment method we will prove that $p_{1/2}(n) = o(1)$, thus allowing us to apply Theorem 2.3.

4.1. An upper bound for the threshold for k-XOR-SAT

Proposition 4.1. *The critical probability for the k-XOR-SAT property verifies $p_{1/2}(n) = O(n^{1-k}) = o(1)$. More precisely, for each parameter $c \in (0,1)$ and all values $\alpha > 1$, the probability $p_c(n)$ satisfies the relation*

$$p_c(n) \leq \frac{\alpha.n}{N_k}.$$

Proof. Let us decompose $\mu_p(s \in SAT)$ according to $w(s)$:

$$\mu_p(s \in SAT) = \sum_l \mu_p(s \in SAT \mid w(s) = l).\mu_p(w(s) = l). \tag{1}$$

Let us recall that $N_k.p$ is the average weight of s. Since $w(s)$ has a binomial distribution the systems s for which $w(s)$ is far from the average weight do not appear with significant probability. Thus let us decompose the right-hand side of (1) into two parts, Σ_1 and Σ_2, where

$$\Sigma_1 = \sum_{|l - N_k.p| > (N_k.p)^{2/3}} \mu_p(s \in SAT \mid w(s) = l).\mu_p(w(s) = l),$$
$$\Sigma_2 = \sum_{|l - N_k.p| \leq (N_k.p)^{2/3}} \mu_p(s \in SAT \mid w(s) = l).\mu_p(w(s) = l).$$

Since $w(s)$ has a binomial distribution and according to Bienaymé-Chebyshev's inequality the sum Σ_1 verifies

$$\Sigma_1 \leq \sum_{|l - N_k.p| > (N_k.p)^{2/3}} \mu_p(w(s) = l) \leq (N_k p)^{-1/3}(1-p). \tag{2}$$

To estimate Σ_2 we use the so-called first moment method. With each assignment $I : \{x_1, \ldots, x_n\} \longrightarrow \{0,1\}$ we associate the random variable X_I defined on k-XOR-systems by $X_I(s) = 1$ if and only if I satisfies s. Since $s \in SAT$ if and only if $\sum_I X_I(s) \geq 1$, we have

$$\mu_p(s \in SAT \mid w(s) = l) = \mu_p\left(\sum_I X_I(s) \geq 1 \mid w(s) = l\right). \tag{3}$$

Observe that any assignment I satisfies exactly the half of the N_k different k-XOR-equations. Therefore the probability that a random system s is satisfied by I only depends on the number l of equations occurring in s:

$$\mu_p(X_I(s) = 1 \mid w(s) = l) = \frac{\binom{N_k/2}{l}}{\binom{N_k}{l}} \leq 2^{-l}. \tag{4}$$

Thus, as there are 2^n possible assignments I, from (3) and (4) we get:

$$\mu_p(s \in SAT \mid w(s) = l) \leq 2^{n-l}.$$

Now, observe that Σ_2 deals with system s of weight at least $N_k.p - (N_k.p)^{2/3}$, thus according to the previous inequality

$$\Sigma_2 \leq 2^{n-N_k.p+(N_k.p)^{2/3}}. \sum_{|l-N_k.p|\leq (N_k.p)^{2/3}} \mu_p(s \in SAT).$$

The sum in the right-hand side being lower than 1 we obtain:

$$\Sigma_2 \leq 2^{n-N_k.p+(N_k.p)^{2/3}}. \tag{5}$$

Therefore, according to the relations (5) and (2), for $N_k.p \leq \alpha n$ with $\alpha > 1$ we have

$$\lim_{n \longrightarrow +\infty} \mu_p(SAT) = 0. \qquad \square$$

According to Theorem 2.3 and Proposition 4.1, proving the sharpness of the threshold for k-XOR-SAT amounts to verify the two conditions $C'1$ and $C'2$.

4.2. Condition $C'1$ for k-XOR-SAT

Proposition 4.2. *For each parameter $c \in (0,1)$ and all positive integers K,*

$$\mu_{p_c(n)}(s \geq m,\ m \text{ minimal for UNSAT and } w(m) \leq K) = o(1)$$

Proof. Observe that in a minimal unsatisfiable k-system every variable appears at least twice. Therefore, a minimal unsatisfiable k-system of weight t involves at most $\lfloor kt/2 \rfloor$ variables. Therefore, in order to get a minimal unsatisfiable k-XOR-system of weight t, one has first to choose a subset of $\lfloor kt/2 \rfloor$ variables, and second to choose t equations among the $2\binom{\lfloor kt/2 \rfloor}{k}$ equations one can construct from these variables. Thus, we can derive a bound on the number of minimal unsatisfiable k-XOR-systems of weight t:

$$\#\{m \text{ minimal for UNSAT and } w(m) = t\} \leq \binom{n}{\lfloor kt/2 \rfloor}\binom{2\binom{\lfloor kt/2 \rfloor}{k}}{t} = O\left(n^{kt/2}\right). \tag{6}$$

We will prove that $a(n) = o(1)$ where

$$a(n) = \mu_{p_c(n)}(s \geq m,\ m \text{ minimal for UNSAT and } w(m) \leq K).$$

We have:

$$a(n) \leq \sum_{t=1}^{K} \sum_{\substack{m \text{ minimal} \\ w(m)=t}} \mu_{p_c(n)}(s \geq m).$$

If m is of weight t, then $\mu_p(s \geq m) = p^t$, thus we obtain:

$$a(n) \leq \sum_{t=1}^{K} \#\{m \text{ minimal and } w(m) = t\}.p_c(n)^t. \tag{7}$$

Finally, from Proposition 4.1 we know that $p_c(n) = O(n^{1-k})$, therefore the inequalities (6) and (7) show that $a(n) = 0(n^{-\frac{1}{2}})$ as soon as k is greater than or equal to 3. □

According to Theorem 2.3, Propositions 4.1 and 4.2, proving the sharpness of the threshold for k-XOR-SAT amounts to verify $C'2$.

4.3. Condition $C'2$ for k-XOR-SAT

In our context let us recall that if a system s contains s_0 as a subsystem ($s \geq s_0$), then $s \oplus s_0$ denotes the system obtained from s in removing all the equations occurring in s_0. Here we will prove that, conditioned on the appearance of s_0 (a specific satisfiable subsystem of fixed weight), the probability that a random system s is unsatisfiable whereas $s \oplus s_0$ is satisfiable tends to be negligible as n goes to infinity:

Proposition 4.3. *For each parameter $c \in (0,1)$, for all positive integers K and all $s_0 \in SAT$ with $w(s_0) = K$,*

$$\mu_{p_c(n)}(s \in UNSAT, s \oplus s_0 \in SAT |\ s \geq s_0) \xrightarrow[n \to +\infty]{} 0.$$

Proof. The proof is divided into three lemmas. We have to measure the influence of s_0 (a fixed satisfiable system) on every s that contains it. Note that in the above conditional probability the random part is $s \oplus s_0 \in \{0,1\}^{N_k - w(s_0)}$, a satisfiable system. As we will see in the next section, it turns out that we can control the influence of s_0 in considering systems of the form (u, v) where u is a k-system and v a $(k-1)$-system. The induced measure, $\mu_p^{(t)}$, on such systems is obtained from the measure μ_p and depends on t, the number of variables occurring in s_0 (note that if $w(s_0) = K$, then $t \leq k.K$). Then we will consider $\mathcal{B}(u, v)$ the set of k-systems of weight $(k-1)$ that are inconsistent with (u, v) and we will show the

following:

Lemma 4.4. *Let M_k be the number of k-systems of weight $(k-1)$, $M_k = \binom{2\binom{n}{k}}{k-1}$. For each parameter $c \in (0,1)$, for all positive integers t and all $s_0 \in SAT$ in which t distinct variables occur, there exist absolute positive constants C and D such that for every $\gamma > 0$ and for every integer n sufficiently large,*

$$\mu_{p_c(n)}(s \in UNSAT, s \oplus s_0 \in SAT |\ s \geq s_0) \leq C.\mu^{(t)}_{p_c(n)}[(u,v) \in SAT, \#\mathcal{B}(u,v) \geq \gamma.M_k] + \gamma.D,$$

where

$$\mu^{(t)}_p(u,v) = t^{w(v)} p^{w(u)+w(v)} (1-p)^{N_k - w(s_0) - (w(u)+w(v))}.$$

In order to evaluate the right-hand side of the above inequality, the idea is to consider a slight modification of the initial probabilistic model where the influence of the condition on $\mathcal{B}(u,v)$ becomes easier to evaluate.

Consider the following construction: first choose a $(k, k-1)$-system, say $s = (u,v)$, with measure $\mu^{(t)}_p$, second draw uniformly one of the M_k k-system of size $(k-1)$, say a. The system (s,a) so obtained can be considered as a point in the product space $\{0,1\}^{2\binom{n-t}{k}+2\binom{n-t}{k-1}} \times \{1, \cdots, M_k\}$. The associated measure, let us call it $\nu^{(t)}_p$, can easily be expressed in terms of the measure $\mu^{(t)}_p$. Indeed we have:

$$\nu^{(t)}_p(s,a) = \frac{\mu^{(t)}_p(s)}{M_k}.$$

Using this measure we establish the second lemma (see Sect. 5 for the proof):

Lemma 4.5. *For each $p \in [0,1]$, for all positive integers t and for every $\gamma > 0$,*

$$\mu^{(t)}_p\left[s \in SAT, \#\mathcal{B}(s) \geq \gamma.M_k\right] \leq \frac{\nu^{(t)}_p((s,a) \in UNSAT) - \mu^{(t)}_p(s \in UNSAT)}{\gamma}.$$

Now, we will prove that the addition of a random k-system of weight $(k-1)$, a, has almost surely no effect on the satisfiability of a random system. This is exactly what is stated in the following last lemma (see Sect. 5 for the proof):

Lemma 4.6. *For each parameter $c \in (0,1)$, for all positive integers t*

$$|\ \nu^{(t)}_{p_c}((s,a) \in UNSAT) - \mu^{(t)}_{p_c}(s \in UNSAT)| \xrightarrow[n \to +\infty]{} 0.$$

The three lemmas above prove Proposition 4.3.

Last but not least our main result follows from Theorem 2.3 and Propositions 4.1, 4.2 and 4.3. □

Theorem 4.7. *For $k \geq 3$, k-XOR-SAT has a sharp threshold.*

5. Proofs of the technical lemmas

5.1. Proof of Lemma 4.4

Recall that M_k is the number of k-systems of weight $(k-1)$ and that $\mathcal{B}(u,v)$ the set of such systems that are inconsistent with (u,v). For a fixed satisfiable system s_0 with t variables, we have to prove that there exit absolute positive constants C and D such that for each parameter $c \in (0,1)$, for every $\gamma > 0$ and for every integer n sufficiently large,

$$\mu_{p_c(n)}(s \in UNSAT, s \oplus s_0 \in SAT |\ s \geq s_0) \leq C.\mu^{(t)}_{p_c(n)}\left[(u,v) \in SAT, \#\mathcal{B}(u,v) \geq \gamma.M_k\right] + \gamma.D.$$

The proof will be divided into 5 steps.

Step 1. In the above conditional probability the random part is $s \oplus s_0$, thus we work on $\{0,1\}^{N_k - w(s_0)}$. The system s becomes UNSAT because of s_0 if the system $s \oplus s_0$ constrains the variables from s_0 in a way which is inconsistent with the equations of s_0. Therefore it is natural to distinguish in s the equations containing variables occurring in s_0 from the others. Let $x_1, \ldots, x_t$ be the variables occurring in s_0. Recall that $t \leq k.K$ since $w(s_0)$ is bounded by K. Let $\tilde{s}$ be the system of equations from s that do not contain any variable from $\{x_1, \ldots, x_t\}$. Observe that if we replace $x_1, \ldots, x_t$ by some truth values, then the equations from $\tilde{s}$ have still k variables each, whereas the equations from $s \oplus s_0 \oplus \tilde{s}$ (that are the equations from $s \oplus s_0$ having each at least one variable from $\{x_1, \ldots, x_t\}$) become of size $< k$. Let us denote by s^* these equations of size $< k$ so obtained.

We know that s_0 is in SAT. Therefore, if we consider an assignment I on $\{x_1, \ldots, x_t\}$ satisfying s_0, then in order to get $s \in UNSAT$ and $s \oplus s_0 \in SAT$, the system $(\tilde{s}, s^*)$ obtained from $s \oplus s_0$ by replacing the variables $x_1, \ldots, x_t$ by their truth values $I(x_1), \ldots, I(x_t)$ must be unsatisfiable. Therefore,

$$\mu_p(s \in UNSAT, s \oplus s_0 \in SAT |\ s \geq s_0) \leq \mu_p(\tilde{s}, s^*) \in UNSAT, \tilde{s} \in SAT |\ s \geq s_0).$$

From now on we fix I an assignment on $\{x_1, \ldots, x_t\}$ satisfying s_0. Our goal is to delimit the influence of this assignment, that is the influence of s_0, on s in working on such systems $(\tilde{s}, s^*)$.

Let $E = \{s \text{ such that } / \ (\tilde{s}, s^*) \in UNSAT, \tilde{s} \in SAT\}$. We have obtained

$$\mu_p(s \in UNSAT, s \oplus s_0 \in SAT |\ s \geq s_0) \leq \mu_p(E \mid s \geq s_0). \tag{8}$$

Step 2. In this step we will show that we can restrict our investigation to systems s such that every equation in s^* has exactly $k-1$ variables, and such that the left-hand side of all equations in s^* are pairwise disjoint.

Let us call V_1 the set of systems s such that s^* contains equations of size $< k-1$. A system s is in V_1 if it contains at least one equation which has $i \geq 2$ variables in

$\{x_1, \ldots, x_t\}$ and $(k-i)$ variables in $\{x_{t+1}, \ldots, x_n\}$. Since there are $\binom{t}{i}\binom{n-t}{k-i}$ ways of choosing such a subset of variables which determines the left-hand side of an affine equation,

$$\mu_p(V_1) \leq p \sum_{i=0}^{k-2} 2\binom{t}{i}\binom{n-t}{k-i}.$$

Now, since the critical probability p_c is $O(n^{1-k})$ the previous inequality shows that

$$\mu_{p_c}(V_1) = o(1). \tag{9}$$

Thus **we can suppose that s^* only contains equations of size $k-1$**, in this way we can restrict our attention to systems s such that $(\tilde{s}, s^*)$ is a $(k, k-1)$-system.

However a technical difficulty appears for the Hamming weight of $(\tilde{s}, s^*)$ is not necessarily the same as that of $s \oplus s_0$. For instance, suppose that

$$s_0 = \begin{cases} x_1 & +x_2 & +x_3 & = 1 \\ x_1 & +x_3 & +x_4 & = 0, \end{cases}$$

$I(x_1) = I(x_3) = I(x_4) = 0$, $I(x_2) = 1$ and

$$s = \begin{cases} x_1 & +x_2 & +x_3 & = 1 \\ x_1 & +x_3 & +x_4 & = 0 \\ x_1 & +x_5 & +x_6 & = 1 \\ x_3 & +x_5 & +x_6 & = 1. \end{cases}$$

Then,

$$s \oplus s_0 = \begin{cases} x_1 & +x_5 & +x_6 & = 1 \\ x_3 & +x_5 & +x_6 & = 1, \end{cases}$$

whereas $(\tilde{s}, s^*)$ is reduced to one equation, $x_5 + x_6 = 1$, repeated twice.

Let us call V_2 the set of systems s that contain two equations having both the same $(k-1)$ variables in $\{x_{t+1}, \ldots, x_n\}$ (only such a pair of equations can provides an equation repeated twice in s^*). We have,

$$\mu_p(V_2) \leq p^2 \binom{t}{1}\binom{n-t}{k-1}.$$

Now, since the critical probability p_c is $O(n^{1-k})$ this last inequality shows that

$$\mu_{p_c}(V_2) = o(1). \tag{10}$$

Therefore **we can suppose that s^* does not contain twice the same equation**.

Finally, if $V = \{0,1\}^{2\binom{n}{k}} \setminus (V_1 \cup V_2)$, then from the relations (9) and (10) we get:

$$\mu_{p_c}(E \mid s \geq s_0) = \mu_{p_c}(E \cap V \mid s \geq s_0) + o(1). \tag{11}$$

We have shown that we can restrict our attention to systems (u, v) formed by a k-system, u, on $\{x_{t+1}, \dots, x_n\}$, coupled with a $(k-1)$-system, v, on $\{x_{t+1}, \dots, x_n\}$. Such a system can be encoded as a point of $\{0,1\}^{2\binom{n-t}{k}} \times \{0,1\}^{2\binom{n-t}{k-1}}$ and in the next step we will make precise the measure induced by on this new product space.

Step 3. Consider the following mapping $\Psi : s \mapsto (u, v) = (\tilde{s}, s^*)$ from systems s that contain s_0 to systems of the form (u, v), where equations from u are k-equations and equations from v are equations of size $\leq k-1$. Let us denote by $\mu_p^{(t)}$ the measure induced by Ψ on $\{0,1\}^{2\binom{n-t}{k}} \times \{0,1\}^{2\binom{n-t}{k-1}}$

$$\mu_p^{(t)}(u, v) = \mu_p\big(s \in V,\ \Psi(\tilde{s}, s^*) = (u, v) |\ s \geq s_0\big).$$

In order to evaluate $\mu_p^{(t)}(u, v)$, observe that the mapping $s \mapsto (\tilde{s}, s^*)$ is not one-to-one. For instance, if

$$s_0 = \begin{cases} x_1 & +x_2 & +x_3 & = 1 \\ x_1 & +x_3 & +x_4 & = 0 \end{cases}$$

and $I(x_1) = I(x_3) = I(x_4) = 0$, $I(x_2) = 1$, then

$$s_1 = \begin{cases} x_1 & +x_2 & +x_3 & = 1 \\ x_1 & +x_3 & +x_4 & = 0 \\ x_1 & +x_5 & +x_6 & = 1 \\ x_5 & +x_6 & +x_7 & = 1 \end{cases}$$

and

$$s_2 = \begin{cases} x_1 & +x_2 & +x_3 & = 1 \\ x_1 & +x_3 & +x_4 & = 0 \\ x_4 & +x_5 & +x_6 & = 1 \\ x_5 & +x_6 & +x_7 & = 1 \end{cases}$$

are mapped into the same system

$$\begin{cases} & x_5 & +x_6 & = 1 \\ x_5 & +x_6 & +x_7 & = 1. \end{cases}$$

Actually, given any $(k, k-1)$-system (u, v), there are $t^{w(v)}$ systems s in V such that $(\tilde{s}, s^*) = (u, v)$, moreover in the above conditional probability the random part is $s \oplus s_0$. Therefore,

$$\mu_p^{(t)}(u, v) = t^{w(v)} p^{w(u)+w(v)} (1-p)^{N_k - w(s_0) - (w(u)+w(v))},$$

and:

$$\mu_{p_c}(E \cup V |\ s \geq s_0) = \mu_{p_c}^{(t)}[(u, v) \in UNSAT, u \in SAT] + o(1). \tag{12}$$

Step 4. We will show that the right-hand side in (12) can be controlled in considering $\mathcal{A}(u, v)$ the set of $(k-1)$-equations that are inconsistent with the system (u, v). Indeed adding a $(k-1)$-system can be seen as a dynamical process in which

we add one equation at a time. Moreover if $(u,v) \in UNSAT$ and $u \in SAT$ then there exist v' and v'' such that $(u,v') \in SAT, (u,v',v'') \in UNSAT$, $w(v'') = 1$ and $v \geq v' + v''$. Thus, $\mu_p^{(t)}[(u,v) \in UNSAT, u \in SAT]$ is lower than or equal to

$$\sum_{(u,v')\in SAT} \mu_p^{(t)}(u,v') \frac{tp}{(1-p)} \#\mathcal{A}(u,v') \sum_{j=0}^{2\binom{n-t}{k-1}-|v'|-1} \binom{2\binom{n-t}{k-1}-|v'|-1}{j} \left(\frac{tp}{1-p}\right)^j$$

Therefore,

$$\mu_p^{(t)}[(u,v) \in UNSAT, u \in SAT] \leq Y_t(n)\Big(1+\frac{tp}{1-p}\Big)^{2\binom{n-t}{k-1}},$$

where $Y_t(n) = \sum_{(u,v')\in SAT} \mu_p^{(t)}(u,v') \frac{tp}{(1-p)} \#\mathcal{A}(u,v')$.

Now, for any $\delta > 0$ we have $Y_t(n) \leq S_1 + S_2$ where

$$S_1 = \sum_{(u,v')\in SAT, \#\mathcal{A}(u,v') \geq \delta\binom{n-t}{k-1}} \mu_p^{(t)}(u,v') \frac{tp}{(1-p)} \#\mathcal{A}(u,v'),$$

$$S_2 = \sum_{(u,v')\in SAT, \#\mathcal{A}(u,v') < \delta\binom{n-t}{k-1}} \mu_p^{(t)}(u,v') \frac{tp}{(1-p)} \#\mathcal{A}(u,v').$$

But $\frac{\binom{n-t}{k-1}tp_c}{(1-p_c)}$, $\mu_{p_c}^{(t)}\Big(\{0,1\}^{2\binom{n-t}{k}} \times \{0,1\}^{2\binom{n-t}{k-1}}\Big)$, and $\Big(1+\frac{tp_c}{1-p_c}\Big)^{2\binom{n-t}{k-1}}$ are bounded.

Hence

$$S_1 = O\left(\mu_{p_c}^{(t)}\left[(u,v') \in SAT, \#\mathcal{A}((u,v')) \geq \delta \binom{n-t}{k-1}\right]\right), S_2 = \delta.O(1).$$

Thus we have proved that there exit absolute positive constants A and B such that for every $\delta > 0$ and for every integer n sufficiently large,

$$\mu_{p_c}^{(t)}[s = (u,v),\ (u,v) \in UNSAT, u \in SAT] \leq A.\mu_{p_c}^{(t)}\left[s = (u,v) \in SAT, \#\mathcal{A}((u,v)) \geq \delta \binom{n-t}{k-1}\right] + B.\delta. \quad (13)$$

Step 5. In this last step, we will show that, in some sense, the size of $\mathcal{A}((u,v))$ is bounded from above by those of $\mathcal{B}((u,v))$. More precisely, we are going to prove that

$$\mu_p^{(t)}\left[(u,v) \in SAT, \#\mathcal{A}((u,v)) \geq \delta. \binom{n-t}{k-1}\right] \leq \mu_p^{(t)}\left[(u,v) \in SAT, \#\mathcal{B}((u,v)) \geq \phi(\delta).M_k\right], \quad (14)$$

where $\Phi(\delta) = D_k\delta^k$ with D_k a constant depending on k only.

The proof is based on the following trick. From k equations of $\mathcal{A}(s)$ one can build a system in $\mathcal{B}(s)$. Indeed, if each of the following equations belongs to $\mathcal{A}(s)$

$$\begin{array}{llll} x_1 & +\cdots & +x_{k-1} & =\varepsilon_1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{(k-1)^2+1} & +\cdots & +x_{k(k-1)} & =\varepsilon_k \end{array}$$

then a solution $(x_1, \cdots, x_n)$ of s must satisfy $\sum_{i=1}^{k(k-1)} x_i = \sum_{i=1}^{k}(\varepsilon_i + 1)$, hence the following k-system is in $\mathcal{B}(s)$

$$\left\{ \begin{array}{llll} x_1 & +\cdots & +x_k & =0 \\ \vdots & \vdots & \vdots & \vdots \\ x_{k(k-3)+1} & +\cdots & +x_{k(k-2)} & =0 \\ x_{k(k-2)+1} & +\cdots & +x_{k(k-1)} & =1+\sum_{i=1}^{k}(\varepsilon_i+1). \end{array} \right.$$

At first sight, this trick furnishes $\binom{\#\mathcal{A}(s)}{k}$ k-systems in $\mathcal{B}(s)$, and thus the conclusion is pure routine. However, one can raise two objections to this reasoning. Let us choose $k = 3$ for the exposition.

First, observe that the three following 2-equations: $x_1 + x_2 = 0, x_2 + x_3 = 0, x_3 + x_1 = 0$ do not produce a system with 3-equations. We have to guarantee that the trick actually furnishes a system in $\mathcal{B}(s)$.

Second, note that the two different systems of 2-equations: $x_1+x_2 = 0, x_3+x_4 = 0, x_5+x_6 = 0$ and $x_1+x_3 = 0, x_2+x_5 = 0, x_4+x_6 = 0$, lead to the same 3-system. Hence, one system in $\mathcal{B}(s)$ can be counted twice or more.

To make rigorous our initial reasoning we can first suppose that we start with k equations of $\mathcal{A}(s)$ whose sets of variables are pairwise disjoint. In order to see that the above calculus remains valid, it suffices to note that the number of systems of $(k-1)$-systems of weight k whose sets of variables are pairwise disjoint is asymptotically equivalent to the total number of $(k-1)$-systems of weight k. Second, observe that for a fixed k, at most $(k(k-1))!$ $(k-1)$-systems of weight k in which any variable appears at most once, can produce the same system k-system of weight $(k-1)$. Therefore the conclusion is still valid.

According to the relations (8, 11–13) and (14), Lemma 4.4 is proved.

5.2. Proof of Lemma 4.5

We have to prove that for each $p \in [0,1]$, for every $\gamma > 0$,

$$\mu_p^{(t)}\left[s \in SAT, \#\mathcal{B}(s) \geq \gamma.M_k\right] \leq \frac{\nu_p^{(t)}((s,a) \in UNSAT) - \mu_p^{(t)}(s \in UNSAT)}{\gamma},$$

where

$$\nu_p^{(t)}(s,a) = \frac{\mu_p^{(t)}(s)}{M_k}.$$

Observe that if s is UNSAT, then (s,a) is UNSAT. So,

$$\nu_p^{(t)}((s,a) \in UNSAT) = \nu_p^{(t)}((s,a) \in UNSAT, s \in SAT) + \nu_p^{(t)}(s \in UNSAT).$$

Moreover, since there are M_k ways to choose a, then by definition of $\nu_p^{(t)}$:

$$\nu_p^{(t)}(s \in UNSAT) = \mu_p^{(t)}(s \in UNSAT).$$

Therefore, the right-hand side of the above inequality is nothing else but

$$\frac{1}{\gamma}\nu_p^{(t)}((s,a) \in UNSAT, s \in SAT),$$

which is equal to

$$\frac{1}{\gamma}\nu_p^{(t)}(s \in SAT,\ a \in \mathcal{B}(s)),$$

and thus greater than

$$\frac{1}{\gamma}\nu_p^{(t)}(s \in SAT\ ,\ a \in \mathcal{B}(s)\ ,\ \#\mathcal{B}(s) \geq \gamma.M_k).$$

The above relation between $\nu_p^{(t)}$ and $\mu_p^{(t)}$ concludes the proof.

5.3. Proof of Lemma 4.6

We have to prove that for each parameter $c \in (0,1)$,

$$\mid \nu_{p_c}^{(t)}((s,a) \in UNSAT) - \mu_{p_c}^{(t)}(s \in UNSAT)\mid \xrightarrow[n\to+\infty]{} 0.$$

In order to make clear the first term: $\nu_{p_c}^{(t)}((s,a) \in UNSAT)$ let us observe that a fixed k-equation may appear both in s and a. We will now make precise the measure induced by $\nu_p^{(t)}$ on sets of different equations occuring in (s,a).

Let us consider $W(s,a)$ the system formed by set of equations in (s,a). If we denote by $D_s(a)$ the number of k-equations in a that do not already appear in s, then for a fixed $(k,k-1)$-system $w=(u,v)$ such that $|u| \geq k-1$ there are $\binom{|u|}{k-1}\binom{k-1}{j}$ ways to choose (s,a) such that $W(s,a)=w$ and $D_s(a)=j$. Thus, the induced probability on such systems w, let us call it $\rho_p^{(t)}$, can be expressed as:

$$\rho_p^{(t)}(w) = \sum_{j=0}^{k-1} \nu_p^{(t)}(W(s,a)=w\ ,\ D_s(a)=j).$$

$$\rho_p^{(t)}(u,v) = \sum_{j=0}^{k-1} \binom{|u|}{k-1}\binom{k-1}{j} p^{|u|-j}(1-p)^{N_k-|u|+j+w(s_0)} \frac{1}{M_k}\Big(\frac{tp}{(1-p)}\Big)^{|v|}$$
$$= \frac{p^{|u|}(1-p)^{N_k-|u|+w(s_0)}}{M_k}\binom{|u|}{k-1}\Big(\frac{tp}{(1-p)}\Big)^{|v|}\sum_{j=0}^{k-1}\binom{k-1}{j}\left(\frac{1-p}{p}\right)^j.$$

Therefore,

$$\rho_p^{(t)}(u,v) = \mu_p^{(t)}(u,v)\frac{\binom{|u|}{k-1}}{M_k}\left(\frac{1}{p}\right)^{k-1} \tag{15}$$

and, $\nu_p^{(t)}((s,a)\in UNSAT) = \rho_p^{(t)}(w\in UNSAT)$.

We are now in a position to prove the lemma. Recalling that the weight of a random k-system has binomial distribution of parameter N_k and p, we introduce R, the following subset of $\{1,\dots,N_k\}$:

$$R = \Big\{ r\in\{1,\dots,N_k\} \;/\; |r-N_kp| \geq (N_kp)^{2/3}\Big\}.$$

Let R^c denote the complement of R in $\{1,\dots,N_k\}$ and $U_r = \{w=(u,v)\in UNSAT, |u|=r\}$, we have:

$$\mu_p^{(t)}(UNSAT) = \sum_{r\in R}\mu_p^{(t)}(U_r) \;+\; \sum_{r\in R^c}\mu_p(U_r),$$
$$\rho_p^{(t)}(UNSAT) = \sum_{r\in R}\rho_p^{(t)}(U_r) \;+\; \sum_{r\in R^c}\rho_p^{(t)}(U_r).$$

Hence, $\nu_{p_c}^{(t)}((s,a)\in UNSAT) - \mu_{p_c}^{(t)}(s\in UNSAT)| \leq T_1+T_2+T_3$ where

$$T_1 = \mu_p^{(t)}(w=(u,v)\,/\,|u|\in R),\quad T_2 = \rho_p^{(t)}(w=(u,v)\,/\,|u|\in R)$$
$$\text{and}\quad T_3 = \sum_{r\in R^c}|\mu_p^{(t)}(U_r)-\rho_p^{(t)}(U_r)|. \tag{16}$$

Let us prove that T_1, T_2 and T_3 are $o(1)$ when $p=p_c=O(n^{1-k})$.

First, observe that

$$T_1 = \mu_p(u\,/\,|u|\in R)\sum_v\left(\frac{tp}{(1-p)}\right)^{|v|}.$$

When $p=p_c$ the second term of this product is bounded and as in Proposition 4.1, Chebyshev's inequality shows that the first term is $o(1)$. Therefore, $T_1=o(1)$.

In view of (15) similar arguments show that $T_2=o(1)$ and that

$$T_3 = O\left(\sup_{r\in R^c}\Big|\frac{\binom{r}{k-1}}{M_k}\left(\frac{1}{p}\right)^{k-1}-1\Big|\right).$$

According to Lemma 3.1, $N_k.p_c \longrightarrow +\infty$, hence it is not difficult to verify that this shows that $T_3 = o(1)$, thus completing the proof of Lemma 4.6.

6. Conclusion

We have proved the existence of a sharp threshold phenomenon for k-XOR-SAT, $k \geq 3$. Let us observe that, from another point of view, our result analyses the threshold behavior of the consistency of linear systems over the finite field $GF(2)$. Some related results on the rank of such systems or on the expectation of their number of solutions have been obtained by [16, 18, 19] in a slightly different model (in which repetitions of variables in the same equation are allowed), but the sharpness of the phase transition for the property of consistency of a random k-system was not proved. Hence, our result illustrates the interest of directing a lot of work towards obtaining general conditions for sharpness of a phase transition as Friedgut and Bourgain did.

Due to its connections to linear algebra, k-XOR-SAT is a well-known and well-studied problem. We are convinced that the precise and accurate study of its threshold behavior will have a great didactical impact in the scope of phase transitions. Our feeling is that k-XOR-SAT is a natural candidate to understand the probabilistic behavior of sharp phase transitions for random SAT type problems and to fill the gap between rigorous results and statistical physics calculations in such studies (see [21]).

References

[1] R. Aharoni and N. Linial, Minimal non 2-colorable hypergraphs and minimal unsatisfiable formulas. *J. Combin. Theory Ser. A* **43** (1986).
[2] B. Aspvall, M.F. Plass and R.E. Tarjan, A linear-time algorithm for testing the truth of certain quantified Boolean formulas. *Inform. Process. Lett.* **8** (1979) 121-123.
[3] B. Bollobás, *Random graphs*. Academic Press (1985).
[4] V. Chvátal, Almost all graphs with 1.44n edges are 3-colorable. *Random Struct. Algorithms* **2** (1991) 11-28.
[5] V. Chvátal and B. Reed, Mick gets some (the odds are on his side), in *Proc. of the 33rd Annual Symposium on Foundations of Computer Science*. IEEE (1992) 620-627.
[6] N. Creignou and H. Daudé, Satisfiability threshold for random XOR-CNF formulæ. *Discrete Appl. Math.* **96-97** (1999) 41-53.
[7] O. Dubois, Y. Boufkhad and J. Mandler, Typical random 3-SAT formulae and the satisfiability threshold, in *Proc. of the 11th ACM-SIAM Symposium on Discrete Algorithms*, SODA'2000 (2000) 124-126.
[8] P. Erdös and A. Rényi, On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **7** (1960) 17-61.
[9] E. Friedgut and an Appendix by J. Bourgain, Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.* **12** (1999) 1017-1054.
[10] A. Frieze and S. Suen, Analysis of two simple heuristics on a random instance of k-SAT. *J. Algorithms* **20** (1996) 312-355.
[11] I.P. Gent and T. Walsh, The SAT phase transition, in *Proc. of the 11th European Conference on Artificial Intelligence* (1994) 105-109.
[12] A. Goerdt, A threshold for unsatisfiability. *J. Comput. System Sci.* **53** (1996) 469-486.

[13] G. Grimmet, *Percolation*. Springer Verlag (1989).
[14] S. Janson, Poisson convergence and Poisson processes with applications to random graphs. *Stochastic Process. Appl.* **26** (1987) 1-30.
[15] S. Kirkpatrick and B. Selman, Critical behavior in the satisfiability of random Boolean expressions. *Science* **264** (1994) 1297-1301.
[16] V.F. Kolchin, Random graphs and systems of linear equations in finite fields. *Random Struct. Algorithms* **5** (1995) 425-436.
[17] V.F. Kolchin, *Random graphs*. Cambridge University Press (1999).
[18] V.F. Kolchin and V.I. Khokhlov, A threshold effect for systems of random equations of a special form. *Discrete Math. Appl.* **2** (1992) 563-570.
[19] I.N. Kovalenko, On the limit distribution of the number of solutions of a random system of linear equations in the class of boolean functions. *Theory Probab. Appl.* **12** (1967) 47-56.
[20] D. Mitchell, B. Selman and H. Levesque, Hard and easy distributions of SAT problems, in *Proc. of the 10th National Conference on Artificial Intelligence* (1992) 459-465.
[21] R. Monasson and R. Zecchina, Statistical mechanics of the random K-sat model. *Phys. Rev. E* **56** (1997) 1357.
[22] T.J. Schaefer, The complexity of satisfiability problems, in *Proceedings 10th STOC, San Diego (CA, USA)*. Association for Computing Machinery (1978) 216-226.
[23] L. Takács, On the limit distribution of the number of cycles in a random graph. *J. Appl. Probab.* **25** (1988) 359-376.

Communicated by J.M. Steyaert.
Received September, 2001. Accepted March, 2003.