Probabilistic estimates for the generalized maximum satisfiability problem

M. Cochand

Institut de Mathématiques, Université de Lausanne, CH-1015 Lausanne, Suisse

Received 25 June 1991; revised 6 February 1992

Abstract

The generalized maximum satisfiability problem (GMAXSAT) deals with variables taking their values in a finite set. A set of logical clauses is given and the goal is to find an assignment of values to the variables, minimizing the number $\beta$ of unsatisfied clauses. For randomly generated instances of a uniform type we study the distribution of $\beta$, as well as the distribution of the maximal size $\varkappa$ of a satisfiable subproblem, by means of the first and second moment method (Spencer, 1987). Numerical estimates for the distribution of $\varkappa$ and $\beta$ are given for some instances. In relation with the asymptotic behavior, we show that $\varkappa$ has almost surely three possible values only. Furthermore, in the spirit of Burkard and Fincke (1985), we show that for some sequences of random instances, the size of which tends to $\infty$, the relative error of any algorithm for GMAXSAT tends almost surely towards zero.

1. Introduction

The generalized maximum satisfiability problem (GMAXSAT) can be described as follows: Find a mapping $f: N := \{1, \ldots, n\} \to R := \{1, \ldots, r\}$ such that the graph of $f$ contains a minimum number of elements of an a priori given family $\mathcal{C}$ of subsets of $N \times R$. Clearly, only those subsets $C \in \mathcal{C}$, no two elements of which have the same first component, are relevant here, and we shall assume that all $C \in \mathcal{C}$ have this property. This problem can be thought of as assigning to $n$ variables $x_1, \ldots, x_n$ a value $f(x_i) \in R$, in order to satisfy a maximal number of clauses of the form $\neg [(f(x_{i_1}) = j_1) \land \cdots \land (f(x_{i_u}) = j_u)]$ (corresponding to the subsets $\{(x_{i_1}, j_1), \ldots, (x_{i_u}, j_u)\} \in \mathcal{C}$). The maximum satisfiability problem (MAXSAT) occurs by taking $r = 2$. In the sequel, we consider exclusively the restriction of GMAXSAT where all elements of $\mathcal{C}$ have the same cardinality $k$, and denote it GMAXSAT($k$). Given an instance of GMAXSAT (i.e. a triple $(N, R, \mathcal{C})$), we shall say that an assignment satisfying all clauses is valid and that an instance is satisfiable if it has a valid assignment. Algorithmic results obtained by heuristics of radically different nature (for instance, SAMD [5] and SKBLZ [3]) are qualitatively very similar, and
particularly stable as far as the parameters of the random generator producing the instances are kept constant. More precisely, given $0 < p < 1$ and \( k \in \mathbb{N} \), instances obtained by choosing the clauses independently with probability \( p \) among all possible clauses of cardinality \( k \) seem to indicate that the minimum number of unsatisfied clauses is a random variable with a well-located peak. The present work is an attempt to analyze this phenomenon. We shall also study the distribution of the maximum cardinality \( \alpha \) of a subset \( N' \) of \( N \) such that the instance restricted to \( N' \) is satisfiable, a question very similar to the problem of studying the maximum stable set in a random graph.

Matula [6] noticed that for a family of random graphs, the edges of which occur independently with the same probability \( p \) (on a given node set), the maximum cardinality of a stable set is a random variable which presents a well-located peak (see [1, p. 251] for various extensions).

Observe that any instance of GMAXSAT can be reduced to finding a stable set of maximum cardinality in an appropriate graph. In fact, given an instance of GMAXSAT, consider the graph \( G(V, E) \), where:

\[
V = \{(C, T)\} \text{ is the set of all ordered pairs } (C, T) \text{ where } C \in \mathcal{C} \text{ is a clause and } T \text{ an assignment of values for those variables involved in } C \text{ and making } C \text{ a satisfied clause.}
\]

\[
E = \{\{(C_1, T_1), (C_2, T_2)\}\} \text{ is the set of pairs of elements of } V \text{ such that } C_1 \text{ and } C_2 \text{ involve a common variable which is given by } T_1 \text{ a value which is not the one given by } T_2.
\]

It is straightforward to check that every stable set of \( G \) corresponds to a subset of clauses which can be simultaneously satisfied and vice versa. Unfortunately, this reduction does not allow us to make use of those results related to the stable sets: random instances of GMAXSAT with clauses occurring independently with probability \( p \) do not result in a set of random graphs, the edges of which occur independently with the same probability. Consequently, we shall rely directly on the so-called \textit{first and second moment method} (to be recalled in Section 2) to study the distribution of \( \alpha \) (Section 3) and \( \beta \) (Section 4). Asymptotically, we show that \( \alpha \) is almost surely one of the three well-determined values if \( k \geq 3 \), and we derive an asymptotic lower bound for \( \beta \). In the spirit of a result of Burkard and Fincke [2], we shall show in Section 5 that for some sequences of instances, the size of which tends to \( \infty \), the relative error of any algorithm converges almost surely towards 0. This result will be used to show that the first and second moment method fails to establish a threshold phenomenon for GMAXSAT\((k) \) [1, p. 252], [7, p. 18].

One motivation of this work was to localize the optimum of some random instances for the sake of algorithm testing. Appreciable results have been obtained, but unfortunately for instances of rather big size for our purpose: they will be discussed in Section 6.

As a first step, we translate GMAXSAT in the setting of hypergraphs, which is more adequate for our purpose. Given an instance \((N, R, \mathcal{C})\) of GMAXSAT, we associate to it a hypergraph \( H(V, E) \), where \( V := N \times R \) and \( E := \mathcal{C} \). By our assumption on \( \mathcal{C} \), \( H \) is \( n \)-partite (i.e. \( V := V_1 \cup \cdots \cup V_n \) where \( V_i := i \times R \)).

In order to avoid possible confusions, we shall use \( \text{Exp}(X) \) to denote the expectation of a random variable \( X \).
2. Random hypergraphs and related random variables

In the sequel we shall exclusively consider the class \( H_{nr}^k \) of those doubly uniform \( n \)-partite hypergraphs \( H(V = V_1 \cup \cdots \cup V_n, E) \) characterized by

\[
|V_i| = r, \quad 1 \leq i \leq n,
\]

\[
|e| = k, \quad \forall e \in E.
\]

We denote by \( V(H) \) (\( E(H) \)) the node set (the edge set) of \( H \) and for \( T \subset V \)

\[
E_H(T) := \{ e \in E(H) \mid e \subset T \} \quad \text{(or simply } E(T) \).
\]

A subset \( T \subset V \) is said to be a snake if \( |T \cap V_i| \leq 1, \quad 1 \leq i \leq n \) (more precisely, a \( d \)-snake if \( |T| = d \)). A snake is said to be stable in \( H \) if \( |E_H(T)| = 0 \).

Given \( H \in H_{nr}^k \), we are interested in the following problems:

**Problem P1.** Determine \( \alpha(H) := \max \{|T|; T \subset V, \ T \text{ is a stable snake in } H \} \).

**Problem P2.** Determine \( \beta(H) := \min \{|(E_H(T))|; T \subset V, \ T \text{ is an } n \text{-snake in } H \} \).

In the spirit of GMAXSAT, \( V \) is the set of possible values for \( x_i \) and assignments correspond to \( n \)-snakes. Furthermore, the assignment corresponding to a given \( n \)-snake \( T \) does not satisfy a given clause iff the corresponding edge \( e \) is in \( E_H(T) \).

In a random context, i.e. when hypergraphs \( H \in H_{nr}^k \) have a given probability \( P(H) \) of occurrence, \( \alpha \) and \( \beta \) are random variables, the distribution of which is our main concern. For our random setting, we choose to formalise the random generators adopted in [5], i.e. each of the \( \binom{n}{r}^k \) possible edges occurs, independently, with probability \( p \).

It follows that

\[
P(H) = p^d q^{\binom{n}{r}^k - d}, \quad \text{where } q = 1 - p
\]

and \( d := \text{Card}(E(H)) \).

The resulting probability space will be denoted by \( H_{nr}^{kp} \).

The distribution of \( \alpha \) and \( \beta \) will be studied indirectly, by means of the family of random variables \( X_{dl}: H_{nr}^k \rightarrow \mathbb{N} \) \((d, L \in \mathbb{N})\), defined as follows:

\[
X_{dl}(H) := \text{Card} \{ T \subset V \mid T \text{ is a } d \text{-snake}, \text{Card}(E_H(T)) \leq L \}.
\]

Random variables \( \alpha \) and \( X_{d0} \) are related by

\[
[\alpha(H) \geq d] \Leftrightarrow [X_{d0}(H) > 0]
\]

and variables \( \beta \) and \( X_{nl} \) by

\[
[\beta(H) \leq L] \Leftrightarrow [X_{nl}(H) > 0].
\]

The methodology applied to study \( \alpha \) and \( \beta \) is the so-called *first and second moment method* [7, p. 19], which works as follows. Let \( X := X_{d0} \) be the number of stable snakes of cardinality \( d \) in a random hypergraph \( H \). \( X \) is then a nonnegative integer valued random variable; hence, we know that

\[
\text{Exp}(X^2)/\text{Exp}(X^2) \leq \text{Pr}(X > 0) \leq \text{Exp}(X).
\]
Now, let $S_1, \ldots, S_m$ be the $m$ snakes of cardinality $d$ in $H$ ($m = \binom{n}{d}$). Define the following indicator variables: $Y_i = 1$ if $S_i$ is stable and $Y_i = 0$ otherwise. The choice of the probabilistic model clearly implies that the $Y_i$'s are identically distributed.

We then have

$$
\text{Exp}(X^2) = \sum_{i=1}^{m} \text{Exp}(XY_i) - m \text{Exp}(XY_i | Y_i = 1) \Pr(Y_i = 1)
$$

$$
= \text{Exp}(X | Y_i = 1) \text{Exp}(X).
$$

In conclusion, we obtain

$$
\frac{\text{Exp}(X)}{\text{Exp}(X | Y_i = 1)} \leq \Pr(X > 0) \leq \frac{\text{Exp}(X)}{\text{Exp}(X | Y_i = 1)}.
$$

(Of course, this argument remains valid if $X$ denotes the number of snakes of cardinality $n$ having at most $L$ edges ($S_i$ then denotes a snake of cardinality $n$, so $m = n^r$, and $Y_i = 1$ if $S_i$ has at most $L$ edges). Our goal is now to derive $\text{Exp}(X)$ and $\text{Exp}(X | Y_i = 1)$ for both cases, and to determine the behavior of the resulting bounds for $n \to \infty$. In order to discuss asymptotic matters, we shall consider the space $H^k_{nr}$, natural extension of $H^k_{nr}$ for $n \to \infty$. The elements $H$ of the underlying set $H^k_{nr}$ are those infinite doubly uniform hypergraphs with node set $\mathbb{N} \times R$ and hyperedges of the form $\{(n_1, r_1), (n_2, r_2), \ldots, (n_k, r_k)\}$, $n_1 < n_2 < \cdots < n_k$.

Given $H$ in $H^k_{nr}$, let $H_n$ stand for the subhypergraph of $H$ induced by the node set $\{1, \ldots, n\} \times R$. We shall say that almost every (a.e.) hypergraph of $H^k_{nr}$ has a property $Q$ if $\Pr(H$ has $Q) = 1$. In our context, $Q$ will have the form: There exists $n_0(H) \in \mathbb{N}$ such that, for $n > n_0$, $H_n$ has a property $Q(n)$. Saying that $Q$ holds for a.e. hypergraph $H$ amounts then to saying that those $H$ such that $Q(n)$ does not hold for infinitely many values of $n$ form a set of measure zero. By the Borel–Cantelli lemma, $Q$ holds for a.e. hypergraph if

$$
\sum_n \Pr(H_n \text{ has not } Q(n)) < \infty.
$$

The weaker property $\Pr(H_n \text{ has } Q(n)) \to 1$ for $n \to \infty$ shall also be considered.

### 3. Bounds for Problem P1

In order to study the distribution of $\alpha(H)$, the maximum cardinality of a stable snake in a hypergraph $H \in H^k_{nr}$, it will be convenient to set $X_d := X_{d,0}$, i.e. the number of stable $d$-snakes of a hypergraph $H$ in $H^k_{nr}$.

**Proposition 3.1.** (a) The expectation of the number of stable $d$-snakes in $H^k_{nr}$ is given by

$$
\text{Exp}(X_d) = \binom{n}{d} r^d q^d.
$$

(b) The expectation of the number of maximal stable $d$-snakes in $H^k_{nr}$ is given by

$$
\binom{n}{d} r^d q^d \left(1 - q^{(\varepsilon+1)n^r}\right)^{(n-d)}.
$$

Notice that (b) appears here as a byproduct and not as part of our approach.
Proof. (a) The probability of a $d$-snake $T$ being stable is the probability that none of the $\binom{d}{2}$ potential edges occurs, i.e. $q(d)$. Moreover, there are $\binom{n}{d}$ $d$-snakes of cardinality $d$.

(b) For $T$ being maximal, it is in addition necessary that all of the remaining $r(n-d)$ nodes forming a snake with $T$ have at least an edge with $k-1$ nodes of $T$. The probability that one of these $r(n-d)$ nodes has no edges with $k-1$ nodes of $T$ is $q(k-1)$. Hence, the probability that all of them have at least an edge with $k-1$ nodes of $T$ is $(1 - q(k-1))^r(n-d)$. \hfill $\square$

Proposition 3.2. The expectation of the number of stable $d$-snakes in $H^{kp}_n$, assuming that a particular $d$-snake is stable, is given by

$$\text{Exp}(X_d | Y_1 = 1) = \sum_{j=0}^{d} q(d) - (\tbinom{d}{j} - \tbinom{d}{d-j}) \left( \prod_{e=f(d)}^{d-j} \sum_{e=f(d)}^{d-j} \binom{n-d}{d-j-e} \right) (r-1)^{e-r-j-e}$$

where $f(d) = \max\{0, 2d - j - n\}$.

Proof. Let $T$ be a particular $d$-snake. For $0 \leq j \leq d$, the probability of a $d$-snake $T'$, with $|T \cap T'| = j$, being stable is $q(d) - (\tbinom{d}{j})$. The result follows by noting that the main parenthesis is the number of $d$-snakes $T'$ with $|T \cap T'| = j$ (i.e. represents the number of nodes in $T'$ being in the same class of the partition of $V$ as a node of $T$, but not being identical to it). $f(d)$ provides enough space for the $d - j - e$ nodes of $T'$ to be placed in those classes of the partition of $V$ not intersected by $T$. \hfill $\square$

Numerical estimates obtained on the basis of the above results are presented in Section 6.

In order to discuss the asymptotic behavior of $\pi(H_n)$, we shall need the following lemma.

Lemma 3.3. Let $k$ be a fixed integer. For $d \geq k$ and $1 \geq \varepsilon \geq k/d$, we have

$$(d - (1 + \varepsilon)k/2)^{k-1} \leq (d - 1)(d - 2) \cdots (d - k + 1) \leq (d - k/2)^{k-1}.$$

Proof. Note that

$$\left( \prod_{i=1}^{k-1} (d - i) \right)^2 = \prod_{i=1}^{k-1} (d - i) (d - k + i).$$

Let $D \in \mathbb{N}$ and $\varepsilon \leq 1$. By the above lemma it follows that

$$\text{Exp}(X_d) \leq \frac{1}{d!} (nrq^{d - (1 + \varepsilon)k/2})^d$$

for all $d \geq k/\varepsilon$,

$$\text{Exp}(X_d) \geq \frac{1}{d!} ((n - D)rq^{d - k/2})^d$$

for all $d \leq D$. 

\hfill $\square$
Note that in the context of asymptotic results inequalities must be considered for $n \geq 1$.

Define

$$f(d, n, \varepsilon) := nrq^{d - (1 + \varepsilon)k/2^{k-1}k!},$$

$$g(d, n, D) := (n - D)rq^{d - k/2^{k-1}k!},$$

$$e(n) := k/(\log_{1,q}(nr))^{1/k-1},$$

$$d^*(n) := \min\{d \in \mathbb{N} | f(d, n, e(n)) \leq 1\} - \lceil (1 + e(n))k/2 + (k! \log_{1,q}(nr))^{1/(k-1)} \rceil$$

$$d(n) := \max\{d \in \mathbb{N} | g(d, n, d^*(n)) \geq n^{(k + 1/2)/(k - 1)d^*(n)d^*(n)k/2} \}$$

It is easy to check that $d(n) \leq d^*(n)$ and $d^*(n) \geq k/e(n)$; hence, $\exp(X_{d^*(n)}) \to 0$ and $\exp(X_{d(n)}) \to \infty$ for $n \to \infty$. It follows that

$$\Pr(\alpha(H_n) < d^*(n)) \to 1 \text{ for } n \to \infty.$$

**Lemma 3.4.** For a.e. $H \in H_{k}^{k=}$, $k \geq 3$, there is a constant $m_0 = m_0(H)$ such that if $n > m_0$

then

$$\alpha(H_n) \leq d^*(n).$$

**Proof.** We show that $\sum_n \Pr(X_{d^*(n)} + 1 > 0) < \infty$. Let $n$ be fixed. Note first that

$$\exp(X_{d^*(n)}) \leq 1 \text{ implies } q^{(e^*(n))} \leq \left( \frac{d^*(n)}{r(n - d^*(n))} \right)^{d^*(n)}.$$

It follows that

$$\Pr(X_{d^*(n)} + 1 > 0) \leq \exp(X_{d^*(n)} + 1) \leq \frac{\exp(X_{d^*(n)} + 1)}{\exp(X_{d^*(n)})} = \frac{n - d^*(n)}{d^*(n) + 1}rq(e^*(n))$$

$$\leq n - d^*(n) \left( q^{(e^*(n))} \right)^{k(d^*(n) - 1)}$$

$$\leq \left( \frac{d^*(n)}{r(n - d^*(n))} \right)^{k(d^*(n) - 1)}$$

$$\leq C_0 \left( \frac{\log_{1,q}(rn)}{(rn)^{k-1}} \right)^{d^*(n)/(d^*(n) - 1)}$$

for some constant $C_0$. Hence $\sum_n \Pr(X_{d^*(n)} + 1 > 0) < \infty$ for $k \geq 3$. \(\square\)

**Lemma 3.5.** For a.e. $H \in H_{n_0}^{k}$, $k \geq 3$, there is a constant $n_0 = n_0(H)$ such that if $n > n_0$

then

$$\alpha(H_n) \geq d(n).$$
Proof. We show that \( \sum_{n} \Pr(X_{d(n)} = 0) < \infty \).

For \( j \geq k \), let \( n(j) := \min \{ n \in \mathbb{N} \mid d(n) \geq j \} \). Moreover, let \( 0 < \tau < 1, M > 1 \) and \( n \geq M \) be fixed.

Note that a \( d \)-snake \( S_i \) contributes \( q(n) \) to \( \text{Exp}(X_d) \) but only \( q(n) - q(n) \) to \( \text{Exp}(X_d \mid Y_1 = 1) \) if \( |S_i \cap S_1| = j \). Moreover, there are at most \( \binom{d(n)}{j} \) such snakes.

It follows that

\[
\Pr(X_{d(n)} = 0) \leq \frac{\text{Exp}(X_{d(n)} \mid Y_1 = 1) - \text{Exp}(X_{d(n)})}{\text{Exp}(X_{d(n)})}
\]

\[
\leq \frac{\sum_{j=k}^{d(n)} \binom{d(n)}{j} (\frac{n-j}{d(n)-j}) r_{d(n)-j} q(n) - q(n)}{\binom{d(n)}{j} r_{d(n)} q(n)}
\]

\[
\leq \sum_{j=k}^{d(n)} \frac{(\binom{d(n)}{j})^2}{r(j) q(n)} = \sum_{j=k}^{d(n)} \phi(n,j).
\]

We show that for \( k < j \leq d(n), \phi(n,j) \leq \sigma(j) n^{-1+\tau} \), where \( \sigma(j) \) is independent of \( n \).

We have

\[
\phi(n,j) \leq \frac{d^*(n)^2}{j!(n-j+1) r_q(j-k/2)^{k-1}/k! j}
\]

\[
= \frac{1}{j!} \left( \frac{d^*(n)^2}{(n-j+1) r_q(j-k/2)^{k-1}/k! j} \right)^{k-1} \left( n-j+1 \right)^{(k-1)/k}
\]

\[
\times \left( \frac{1}{(n-j+1) r_q(j-k/2)^{k-1}/k! (k-1)/k} \right)^j
\]

\[
\leq \frac{1}{j!} \left( \frac{d^*(n)^2}{(n-d^*(n)+1) r_q(d(n)-k/2)^{k-1}/k!} \right)^{k-1} \left( n-j+1 \right)^{(k-1)/k}
\]

\[
\times \left( \frac{1}{(n-j-d^*(n(j)) + 1) r_q(d(n(j))-k/2)^{k-1}/k! (k-1)/k} \right)^j
\]

(recall that \( j \leq d(n(j)) \leq d(n) \))

\[
\leq \frac{1}{j!} \left( \frac{d^*(n)^2}{n^{(k+1)/2}/(k(k-1)d^*(n))} \right)^{(k-1)/k}
\]

\[
\times \left( \frac{d^*(n)^2}{n} \right)^{(k-1)/k}
\]

\[
\times \left( \frac{1}{n(j)^{(k+1)/2}/(kd^*(n(j)))} \right)^j
\]
(recall the definition of $d(n)$ and of $d(n(j))$)

\[
\begin{align*}
= & \frac{1}{j!} \frac{1}{n^{1+\tau}} \left( \frac{n(j)}{n} \right)^j (k^2 + 1) j((k - 1)/k (d^*(n(j))) - (k - 1)/k - (1 + \tau) \\
\times & \frac{n(j)^{(k - 1)/k} - (1 + \tau)}{d^*(n(j))^{2j/k - 1}} \\
\leq & \frac{1}{j!} \frac{1}{n^{1+\tau}} \left( \frac{n(j)}{n} \right)^j (k - 1)/k - (1 + \tau) n(j)^{(k - 1)/k (d^*(n(j)))} + (1 + \tau) \\
\leq & \frac{1}{j!} \frac{1}{n^{1+\tau}} \sigma(j)
\end{align*}
\]

(notice that the exponent of the third factor is positive for $j \geq k \geq 3$), where

\[\sigma(j) := n(j)^{(k - 1)/k (d^*(n(j)))} + (1 + \tau)\]

Note that there exists $J_0$ such that $\sigma(j) < 1$ for $j > J_0$, since $j = d(n(j))$ for $j \to \infty$ and $d(n)/d^*(n) \to 1$ for $n \to \infty$ (see proof of Theorem 3.6).

Our results can be summarized as follows:

**Theorem 3.6.** For a.e. $H \in H^{kp}_{d^*}$, $k \geq 3$, there is a constant $n(H)$ such that if $n > n(H)$ then

\[
\left[ \frac{1}{2} k - \frac{k + 1}{(k - 1)^2} + (k! \log_{1/\alpha} (nr))^{1/(k - 1)} \right] \leq \alpha(H_n) \leq \left[ \frac{1}{2} k + (k! \log_{1/\alpha} (nr))^{1/(k - 1)} \right].
\]

Moreover, with probability tending to unity $\alpha(H_n)$ does not take the upper value.

**Proof.** Define

\[
A := (k! \log_{1/\alpha} (nr))^{1/(k - 1)},
\]

\[
\delta := (k! \log_{1/\alpha} ((n - d^*(n)))^{1/(n(k - 1))} (n - d^*(n)))^{1/(k - 1)}.
\]
Starting with $\Delta - \delta \leq (A^{k-1} - \delta^{k-1})/A^{k-2}$ we get $\Delta - \delta \leq (k + 1)/(k - 1) \leq 2$ for $n \gg 1$, and use this fact with $\Delta - \delta = (A^{k-1} - \delta^{k-1})(A^{k-2} + A^{k-3} \delta + \cdots + \delta^{k-2})^{-1}$ to obtain $\Delta - \delta \to (k + 1)/(k - 1)^2$ for $n \to \infty$. Finally, $d^*(n) \leq d(n) + 2$ for $n \gg 1$.

4. Bounds for Problem P2

In order to localize the optimum of GMAXSAT, we study for $L \in \mathbb{N}$ the probability that an $n$-snake $T$ occurs with $|E(T)| \leq L$ (i.e. the distribution of $X_{nL}$). In this context, the random variable $X_L := X_{nL}$ for $H \in H^p_n$ gives the number of $n$-snakes having at most $L$ edges.

**Proposition 4.1.** For $H \in H^p_n$, the expectation of the number of $n$-snakes having $L$ edges at most is given by

$$\text{Exp}(X_L) = r^n \sum_{m=0}^{L} \binom{\ell}{m} p^{m} q^{(\ell) - m}.$$  

**Proof.** The maximal possible number of edges in a particular $n$-snake is $(\ell)$ and each of them occurs independently with probability $p$. The number of edges for an $n$-snake is consequently a binomial random variable: $\sim \mathcal{B}((\ell), p)$. The result follows multiplying the probability of such a variable being not more than $L$ by the number $r^n$ of $n$-snakes. 

**Proposition 4.2.** The conditional expectation of the number of $n$-snakes $T'$ with $|E(T')| \leq L$ in $H^p_n$, given that a particular $n$-snake $T$ has $|E(T)| \leq L$, is equal to

$$\text{Exp}(X_L | Y_1 = 1) = \sum_{d=0}^{n} \binom{n}{d} (r - 1)^{n-d} \sum_{a=0}^{L} \binom{f(d, a)}{g(d, a)} \frac{P(B_d \leq h(d, t))P(H_{ad} = t)}{P(B' \leq L)},$$

with

$$B_d \sim \mathcal{B}((\ell), p), \quad B' \sim \mathcal{B}((\ell), p), \quad H_{ad} \sim \mathcal{H}((\ell), a, (\ell)).$$

($\mathcal{B}$ and $\mathcal{H}$ denoting binomial and hypergeometric laws, respectively),

$$f(d, a) = \max \{0, a - (\ell) + (\ell)\}, \quad g(d, a) = \min \{a, (\ell)\},$$

$$h(d, t) = \min \{L - t, (\ell) - (\ell)\}.$$

**Proof.** Let $D = T \cap T'$. In order to keep the exposition reasonable, we shall use the notation $P_d(A) := P(A \mid |D| = d)$ with $0 \leq d \leq n$ for the conditional probability of an event $A$ given that $|D| = d$. Assume now that $d$ is fixed. Partitioning the cases according to $a = |E(T)|$ we get

$$P_d(|E(T')| \leq L \mid |E(T)| \leq L) = \sum_{a=0}^{L} P_d(|E(T')| \leq L \mid |E(T)| = a)P(|E(T)| = a \mid |E(T)| \leq L). \quad (4.1)$$
The first factor under the summation sign in (4.1) is equal to
\[
\sum_{t = 0}^{f(d, a)} P_d(|E(T')| \leq L | |E(D)| = t)P_d(|E(D)| = t | |E(T)| = a),
\] (4.2)
where \( t \) is the number of edges of \( D \).

The first factor under the summation sign in (4.2) is equal to
\[
\sum_{b = 0}^{h(d, t)} P_d(|E(T') \setminus E(D)| = b),
\] (4.3)
where the equality is obtained by partitioning the cases according to \( b = |E(T') \setminus E(D)| \) and setting \( h(d, t) = \min \{L - t, \binom{n}{a} - \binom{a}{b}\} \).

This last summation is equal to \( P(B_d \leq h(d, t)) \) where \( B_d \) is a binomial variable with parameters \( (\binom{n}{a} - \binom{a}{b}, p) \).

The second factor under the summation in (4.2) is equal to \( P(H_{ad} = t) \) where \( H_{ad} \) follows a hypergeometric law with parameters: \( \binom{a}{d} \), the size of the set having \( a \) distinguished elements and in which one chooses a subset of cardinality \( \binom{a}{d} \).

The second factor under the summation in (4.1) is equal to \( P(B' = a)/P(B' \leq L) \) where \( B' \) follows a binomial law with parameters \( (\binom{n}{a}, p) \).

The conclusion follows by noting that for fixed \( d \), there are \( \binom{a}{d}(r - 1)^{n-d} \) possible choices for the nodes of \( T' \). \( \square \)

Numerical estimates obtained on the basis of the above results are presented in Section 6.

We now study the asymptotic behavior of \( \beta(H_n) \). Let \( N := \binom{n}{a} \) and note that for \( L > 1 \) and \( n \gg 1 \):
\[
\Exp(X_L) \leq r^n q^N \sum_{m=0}^{L} \frac{1}{m!} \left( \frac{np}{q} \right)^m \leq r^n q^N \left( \frac{np}{q} \right)^L = \left( \frac{r^n q^N}{L^L} \left( \frac{np}{q} \right) \right)^L.
\]
As an immediate consequence, defining
\[
L(n) = L(n, q) := \frac{\ln(r^n q^N)}{\ln(q(1 - c)/(Np))},
\]
we have \( \Exp(X_{L(n)}) \to 0 \) for \( n \to \infty \); furthermore, we have the following lemma.

**Lemma 4.3.** For a.e. \( H \in H_{H_n}^{kp} \), there is a constant \( m_0 = m_0(H) \) such that if \( n > m_0 \) then
\[
L(n, q) \leq \beta(H_n).
\]
**Proof.** \( \Exp(X_{L(n-1)}) < q/(pN) \) by the above majorization. \( \square \)

For \( L_0 \geq L \) we have
\[
\Exp(X_L) \geq r^n q^N \sum_{m=0}^{L_0} \frac{1}{m!} \left( (N - L_0) \frac{p}{q} \right)^m \geq r^n q^N \frac{1}{L!} \left( (N - L_0) \frac{p}{q} \right)^L
\]
\[
= \frac{1}{L!} \left( r^n q^N \right)^{1/L} \left( (N - L_0) \frac{p}{q} \right)^L.
\]
As an immediate consequence, defining
\[
L_0(n) := \left\lceil \frac{-2 \ln(q^N)}{\ln(N)} \right\rceil \quad \text{and} \quad \bar{L}(n) := \left\lfloor \frac{\ln(r^n q^N)}{\ln(q(1 + \varepsilon)/(N - L_0(n))p)} \right\rfloor,
\]
we have \(\text{Exp}(X_{L_0(n)}) \to \infty\) for \(n \to \infty\). To see that \(\bar{L}(n) \leq L_0(n)\) (for \(n \geq 1\)), note that for \(n \to \infty\) we have
\[
L(n) = \left\lfloor \frac{\ln(r^n q^N)}{-\ln(N)} \right\rfloor = \bar{L}(n).
\]
It follows that \(L_1(n)\), the smallest integer such that \(\text{Exp}(X_{L_1(n)}) \geq 1\), lies in the interval \([L(n), \bar{L}(n)]\). Note that despite the convergence of \(L(n)\) and \(\bar{L}(n)\) to the same limit, this interval does not collapse! The interest of \(L_1(n)\) is that it is the point around which the first and second moment method usually builds its success, when the so-called threshold phenomenon occurs [1, p. 252], [7, p. 17]. In our problem, the threshold, if it exists, does not show up around \(L_1(n)\). We shall prove in the next section that no linear function of \(L_1(n)\) exists which asymptotically is an upper bound for \(\beta\).

5. Asymptotic behavior

Given \(n, k\) and \(p\), it is easy to see that for any \(\varepsilon > 0\) there exists \(r = r(n, k, p, \varepsilon)\) such that
\[
\text{Prob}\{3 \text{ stable n-snake}\} = \text{Prob}\{\beta = 0\} > 1 - \varepsilon. \tag{5.1}
\]
For fixed \(k, p\) and \(\varepsilon\), we are interested (as \(n \to \infty\)) in the behavior of the smallest \(r = r(n, k, p, \varepsilon)\) such that (5.1) holds.

In order to deal with this problem, we shall consider a sequence of random variables on the probability space \(H_{nk}^{kp}\), natural extension of \(H_{nr}^{kp}\) for \(n \to \infty\).

Given a mapping \(r: \mathbb{N} \to \mathbb{N}\), we define on \(H_{nk}^{kp}\) the sequence of random variables \(\text{Min}_n\) as follows, denoting \(r(n)\) by \(r_n\): For fixed \(n\), \(\text{Min}_n\) is equal to \(\beta\) on \(H_{nr}^{kp}\) which by definition is the restriction of \(H_{nk}^{kp}\) to its \(n\) first variables and to their \(r_n\) first values. In the same way, we define \(\text{Max}_n\) considering the maximum number of edges in an \(n\)-snake instead of the minimum.

We shall exhibit a mapping \(r(n)\) such that (5.1) does not hold for small \(\varepsilon\). More precisely, \(r(n)\) is such that for the sequences of random variables associated to \(r_n = r(n)\), the ratio \(\text{Max}_n/\text{Min}_n\) converges almost surely to 1 for \(n \to \infty\).

An essential step in this direction is given by a theorem of Burkard and Fincke [2]. For sequences of random combinatorial problems, the size of which tends to \(\infty\), this result gives sufficient conditions for the ratio between worst and best solution to be in any neighborhood of 1 with probability tending to 1.

We follow [2] to introduce the necessary background: Let \((P_n), n \in \mathbb{N}\), denote a family of combinatorial optimization problems defined on finite ground sets \(E_n\). The feasible solutions of a problem \(P_n\) are defined by a nonempty class \(T_n\) of subsets of \(E_n\). A feasible solution \(S\) is therefore a subset of \(E_n\). We denote by \(|T_n|\) the cardinality of \(T_n\) and by \(|S|\) the number of elements \(e \in E_n\) belonging to \(S\).
Further, let $c_n : E_n \rightarrow \mathbb{R}_+$ be a weight function, which maps $E_n$ into the nonnegative reals.

Problem $P_n$ consists in finding:

$$\min_{S \in T_n} \sum_{e \in S} c_n(e). \quad (5.2)$$

**Theorem 5.1** (Burkhard and Fincke [2]). Let $c_n(e), n \in \mathbb{N}, e \in S \in T_n$, be identically distributed random variables in $[0, 1]$ with expected value $E := \Exp(c_n(e))$ and variance $\sigma^2 := \sigma^2(c_n(e)) > 0$. For given $\varepsilon > 0$, let $\varepsilon_0$ fulfill

$$0 < \varepsilon_0 \leq \sigma^2 \quad \text{and} \quad 0 < \frac{E + \varepsilon_0}{E - \varepsilon_0} \leq 1 + \varepsilon$$

and define $\lambda_0 := 2(\varepsilon_0 \sigma/(\varepsilon_0 + 2\sigma^2))^2$. Furthermore, let the two following conditions be satisfied:

(a) $c_n(e), e \in S$, are independently distributed for fixed $S \in T_n, n \in \mathbb{N}$.

(b) $|S| = |\hat{S}|$ for all $S, \hat{S} \in T_n, n$ fixed.

Then

(i) \begin{align*}
\text{Prob} \left\{ \forall S \in T_n : \left| \sum_{e \in S} (c_n(e) - E) \right| < \varepsilon_0 |S| \right\} \geq 1 - 2|T_n| \exp(-|S|\lambda_0).
\end{align*}

(ii) Moreover, if

(c) $\lambda_0 |S| - \ln |T_n| \to \infty$ as $n \to \infty$ then

\begin{align*}
\text{Prob} \left\{ \frac{\max_{S \in T_n} \sum_{e \in S} c_n(e)}{\min_{S \in T_n} \sum_{e \in S} c_n(e)} < 1 + \varepsilon \right\} \geq 1 - 2|T_n| \exp(-|S|\lambda_0) \to 1 \quad \text{for} \; n \to \infty.
\end{align*}

In the random context of GMAXSAT($k$) as defined by $H_{nN}$, the application of this theorem rests on the following correspondences.

An instance of problem $P_n$ is an instance of GMAXSAT($k$) (resulting from $H_{nN}^{kp}$) considered as follows: $E_n$ is in our case the set whose elements are the \binom{n}{k} $k$-snakes which can be defined on $V = \cup V_i$. The set $T_n$ of feasible solutions is $\{S_n(D) \subseteq E_n : D$ is an $n$-snake and $S_n(D)$ is the set of those \binom{n}{k} $k$-snakes contained in $D\}$. Random variable $c_n(e)$ is a Bernoulli variable taking value 1 with probability $p$.

Problem (5.2) is equivalent to GMAXSAT($k$) considering that edges $e$ are present in $H \in H_{nN}^{kp}$ iff $c_n(e) = 1$. Although nothing has been said about $r_n$, it should be clear that all hypotheses of the above theorem, but (c), are fulfilled.

Subsuming we have

$$|T_n| = (r_n)^n, \quad |S_n| = \binom{n}{k},$$

and we are concerned with the evolution of

$$x_n = \lambda_0 |S_n| - \ln |T_n| = \lambda_0 \binom{n}{k} - n \ln(r_n)$$
when \( n \to \infty \). Clearly, only a mapping \( r_n := r(n) \) with a monstrous increase could avoid \( x_n \to +\infty \). We shall show that the same holds for the almost sure convergence of \( \text{Max}_n / \text{Min}_n \) to 1.

We shall need the two following lemmas, the first one closely related to the Borel–Cantelli lemma.

**Lemma 5.2** (see Gnedenko [4, p. 238]). Let \( \xi_n, n \in \mathbb{N} \), be a sequence of random variables on a probability space \( H \). Then if \( \forall L \in \mathbb{N} \) we have

\[
\sum_{n=1}^{\infty} \text{Prob}\left\{ |\xi_n - \xi| \geq \frac{1}{L} \right\} < \infty,
\]

the sequence \( \xi_n \) converges almost surely to \( \xi \).

**Lemma 5.3.** Let \( 2 \leq k \in \mathbb{N} \) and \( 0 < C \in \mathbb{R} \). Then there exists \( n_0 \in \mathbb{N} \) such that for \( n > n_0 \)

\[
\begin{align*}
(a) & \quad n^{k-2} < C(n - 1)^{k-1}, \\
(b) & \quad \frac{(e((n + 1)^{k-2})^{n+1}}{(e(n^{k-2})^n} < \frac{(e(Cn^{k-1})^{n+1}}{(e(C(n - 1)^{k-1})^n}.
\end{align*}
\]

The proof is given in the appendix.

**Theorem 5.4.** Let \( k \in \mathbb{N} \) and \( 0 < p < 1 \) be given. Then for \( r_n \leq \exp(n^{k-2}) \), the sequence of random variables \( \xi_n = \text{Max}_n / \text{Min}_n \) defined on \( H_{\mathbb{N}}^p \) converges almost surely to \( \xi = 1 \).

Before giving the proof, we make some comments about this result.

The distribution of the number \( \omega \) of edges in a randomly chosen \( n \)-snake of a random hypergraph \( H_{\mathbb{N}}^p \) becomes highly concentrated in the sense that

\[
\frac{\text{Max}_n - \text{Min}_n}{N} \to 0 \quad \text{for} \quad n \to \infty,
\]

where \( N = \left( \begin{array}{c} \omega \\ n \end{array} \right) \) is the maximum possible value for \( \omega \).

Moreover, \( Np \in [\text{Min}_n, \text{Max}_n] \) implies \( 0 \notin [\text{Min}_n, \text{Max}_n] \). In relation with the form of a mapping \( r(n, k, p, \varepsilon) \) such that (5.1) holds, Theorem 5.4 says that for small \( \varepsilon \), it must grow faster than \( \exp(n^{k-2}) \).

From the algorithmic viewpoint, Theorem 5.4 says that for big instances of \( \text{GMAXSAT}(k) \) any algorithm yields a good solution.

**Proof of Theorem 5.4.** Let \( 2 \leq L \in \mathbb{N} \) be fixed. We want to show that

\[
\sum_{n=1}^{\infty} \text{Prob}\{ |\xi_n - 1| \geq 1/L \} < \infty.
\]

Let

\[
\begin{align*}
\varepsilon_L := & \min\{ L/(3L), \sigma^2 \}, \\
\lambda_L := & \frac{2(\varepsilon_L\sigma/\varepsilon_L + 2\sigma^2))^2, \\
C_L := & \frac{\lambda_L}{2k\left( \frac{1}{k-1} \right)^{k-1}}.
\end{align*}
\]
Let \( n_0 \in \mathbb{N} \) be such that for \( n \geq n_0 \) we have \( n^{k-2} < C_L(n-1)^{k-1} \). By Lemma 5.2 it is enough to see that \( \sum_{n=n_0}^{\infty} \text{Prob}\{|\xi_n - 1| \geq 1/L\} < \infty \). For \( n \geq n_0 \), let us consider the three following events \( A_n, B_n \) and \( C_n \) defined by

\[
A_n := \left\{ \forall S_n \in T_n, \left| \sum_{e \in S_n} (\xi_n(e) - E) \right| < \varepsilon_L |S_n| \right\},
\]

\[
B_n := \{ \xi_n \leq (E + \varepsilon_L)/(E - \varepsilon_L) \},
\]

\[
C_n := \{ |\xi_n - 1| < 1/L \}.
\]

Note that

\[
B_n \Rightarrow C_n \quad \text{since} \quad \varepsilon_L \leq E/(3L) < E/(2L + 1) \Rightarrow (2L + 1)\varepsilon_L < E \Rightarrow (E + \varepsilon_L)/(E - \varepsilon_L) - 1 < 1/L.
\]

\( A_n \Rightarrow B_n \) for \( A_n \) holds in particular for \( S_n = S_{\min} \) and \( S_n = S_{\max} \) giving the values of \( \min \) and \( \max \).

Consequently, \( |S_n|E - |S_n|\varepsilon_L \leq \min \leq \max \leq |S_n|E + |S_n|\varepsilon_L \).

It follows that \( A_n < B_n \subset C_n \) and by Theorem 5.1(i),

\[
\text{Prob}(C_n) \geq \text{Prob}(A_n) \geq 1 - 2|T_n|\exp(-|S_n|\lambda_L).
\]

Hence,

\[
\text{Prob}\{|\xi_n - 1| \geq 1/L\} \leq P(C_n) < 2|T_n|\exp(-|S_n|\lambda_L) =: a_n.
\]

We shall establish the convergence of \( \sum_{n=n_0}^{\infty} a_n \) by applying d'Alembert's criteria (i.e. \( \lim_{n \to \infty} a_{n+1}/a_n < 1 \)). Recall that \( |S_n| = \binom{n}{k}, |T_n| = r_n^n \), and take \( r_n = \exp(n^{k-2}) \).

Moreover, for those values of \( n \) which are of interest \( n^{k-2} < C_L(n-1)^{k-1} \) where \( C_L := (\lambda_L/(2k))((k-1))^{k-1} \).

Let

\[
\theta_n := \frac{a_{n+1}}{a_n} = \frac{(r_{n+1})^{n+1} e^{\lambda_L |S_n|}}{(r_n)^n} = \frac{(r_{n+1})^{n+1} e^{\lambda_L k}}{(r_n)^n} = \frac{(r_{n+1})^{n+1} e^{\lambda_L k}}{(r_n)^n} = \frac{1}{e^{\lambda_L n^{k-2}}},
\]

Furthermore (Lemma 5.3)

\[
\frac{(r_{n+1})^{n+1}}{(r_n)^n} < \frac{(e^{(C_L(n^{k-1})^{n+1}}) e^{(C_L(n^{k-1})^{n+1}}}{(e^{(C_L(n^{k-1})^{n+1}})^n}.
\]

Taking the logarithm we get

\[
\ln(\theta_n) < C_L[(n+1)n^{k-1} - (n)(n-1)^{k-1}] - \lambda_L \binom{n}{k-1}.
\]

We majorize \( \binom{n}{k} \) in the first term noting that \( f(x) = x(x - 1)^{k-1} \) is convex for \( k \geq 1 \).

Hence, \( \binom{n}{k} = f(n+1) - f(n) \equiv f'(n+1) \).

\[
C_L[ \binom{n}{k} ] \leq C_L(n^{k-1} + (n+1)(k-1)n^{k-2}) < \lambda_L \binom{n}{k-1}^{k-1}.
\]
This inequality is obtained by substituting the actual value of $C_L$ and after some algebra given in the appendix.

Finally,

$$\ln(\theta_n) < \lambda_L\left(\frac{n}{k-1}\right)^{k-1} - \binom{n}{k-1} \to -\infty \text{ for } n \to \infty.$$ 

**Remarks.** (1) From Stirling’s formula, it follows that for $k \geq 4$ we have $\ln(n!) \leq n^2 \leq n^{k-2} - \ln(R(n))$, i.e. $R(n) > n!$.

(2) Our proof is valid with the hypothesis of Theorem 5.1 concerning variables $c_n(e)$, i.e. in a context more general than GMAXSAT.

A nice application of Theorem 5.4 is related to the determination of an upper bound for $\beta(H_n)$. Recall that $\bar{L}(n)$ is an upper bound for $L_1(n)$, the smallest integer such that $\text{Exp}(L_1(n)) > 1$. For $\alpha(H_n)$ and other similar problems [7, p. 17], bounds are located close to the value where the expectation passes above 1. We show now that this is not the case for $\beta(H_n)$. In order to stress the dependency, we shall write $\bar{L}(n, p)$ for $\bar{L}(n)$.

**Corollary 5.5.** There exists no linear function $F$ such that for a.e. $H \in H_{\mathbb{N}_{\mathbb{P}}}$,

$$\beta(H_n) \leq F(L_1(n, p)).$$

**Proof.** We assume that the result is false, i.e. there exists a linear function $F$ such that for a.e. $H \in H_{\mathbb{N}_{\mathbb{P}}}$,

$$\beta(H_n) \leq F(L_1(n, p)). \tag{5.4}$$

This implies

$$\beta(H_n) \leq F(\bar{L}(n, p)) \text{ for a.e. } H \in H_{\mathbb{N}_{\mathbb{P}}}. $$

Let $\gamma(H_n)$ denote the maximum number of edges in an $n$-snake. In fact, $\gamma(H_n) = N - \sigma(H_n)$ where $\sigma(H_n)$ stands for the minimum number of edges not taken in an $n$-snake (let us call them anti-edges, they occur with probability $q$). Note now that $\sigma(H_n)$ which counts the minimum number of anti-edges in an $n$-snake is a random variable of the same type as $\beta(H_n)$ and that consequently (5.4) must hold for $\sigma(H_n)$ too. It follows that

$$\sigma(H_n) \leq F(\bar{L}(n, q)) \text{ for a.e. } H \in H_{\mathbb{N}_{\mathbb{P}}},$$

and consequently

$$\gamma(H_n) \geq N - F(\bar{L}(n, q)) \text{ for a.e. } H \in H_{\mathbb{N}_{\mathbb{P}}}. $$

Observe now that we have

$$\frac{N - F(\bar{L}(n, q))}{F(\bar{L}(n, p))} \to \infty \text{ for } n \to \infty,$$

so that

$$\gamma(H_n)/\beta(H_n) \to 1$$

is impossible, a contradiction with Theorem 5.4. \qed
Table 1

<table>
<thead>
<tr>
<th>$L$</th>
<th>$LB$</th>
<th>$UB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.00000004</td>
<td>0.00000004</td>
</tr>
<tr>
<td>6</td>
<td>0.00000048</td>
<td>0.00000048</td>
</tr>
<tr>
<td>7</td>
<td>0.00000524</td>
<td>0.00000524</td>
</tr>
<tr>
<td>8</td>
<td>0.00004976</td>
<td>0.00005043</td>
</tr>
<tr>
<td>9</td>
<td>0.00041451</td>
<td>0.00042509</td>
</tr>
<tr>
<td>10</td>
<td>0.00305095</td>
<td>0.00319754</td>
</tr>
<tr>
<td>11</td>
<td>0.01971207</td>
<td>0.02168053</td>
</tr>
<tr>
<td>12</td>
<td>0.10528084</td>
<td>0.13360529</td>
</tr>
<tr>
<td>13</td>
<td>0.37791665</td>
<td>0.75349630</td>
</tr>
<tr>
<td>14</td>
<td>0.72844337</td>
<td>1.00000000</td>
</tr>
<tr>
<td>15</td>
<td>0.90406051</td>
<td>1.00000000</td>
</tr>
<tr>
<td>16</td>
<td>0.95975679</td>
<td>1.00000000</td>
</tr>
<tr>
<td>28</td>
<td>0.99507002</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

$n = 10$, $r = 15$, $k = 3$, $p = 0.4$.
Expected number of edges for $H \in H_{15}^W$: $\approx 162000$.
Expected number of edges induced by an $n$-snake: $\approx 48$.

14 is the smallest integer $L$ such that the expected number of $n$-snakes inducing at most $L$ edges is greater than 1, it is $\approx 3.9$.

Table 2

<table>
<thead>
<tr>
<th>$L$</th>
<th>$LB$</th>
<th>$UB$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00000001</td>
<td>0.00000001</td>
</tr>
<tr>
<td>2</td>
<td>0.00000139</td>
<td>0.00000172</td>
</tr>
<tr>
<td>3</td>
<td>0.00014245</td>
<td>0.00018588</td>
</tr>
<tr>
<td>4</td>
<td>0.01068334</td>
<td>0.01510481</td>
</tr>
<tr>
<td>5</td>
<td>0.38651031</td>
<td>0.98197117</td>
</tr>
<tr>
<td>6</td>
<td>0.93177029</td>
<td>1.00000000</td>
</tr>
<tr>
<td>7</td>
<td>0.95942475</td>
<td>1.00000000</td>
</tr>
<tr>
<td>20</td>
<td>0.96341124</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

$n = 100$, $r = 20$, $k = 3$, $p = 0.002$.
Expected number of edges for $H \in H_{20}^W$: $\approx 2587200$.
Expected number of edges induced by an $n$-snake: $\approx 323$.

6 is the smallest integer $L$ such that the expected number of $n$-snakes inducing at most $L$ edges is greater than 1, it is $\approx 53.2$.

6. Numerical results

Numerical estimates for the distribution of $\alpha$ and $\beta$ have been computed on the basis of (2.1), applying the formulas of Propositions 3.1, 3.2, 4.1 and 4.2.

Computations related to the above formulas (especially in relation with Proposition 4.2) are rather delicate due to round-off errors in computer arithmetic. In order to support our calculations, we made an experiment based on the following remark.
Table 3

<table>
<thead>
<tr>
<th>L</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.00000002</td>
<td>0.00000006</td>
</tr>
<tr>
<td>4</td>
<td>0.00000191</td>
<td>0.00000787</td>
</tr>
<tr>
<td>5</td>
<td>0.00014304</td>
<td>0.00081848</td>
</tr>
<tr>
<td>6</td>
<td>0.00845863</td>
<td>0.07094821</td>
</tr>
<tr>
<td>7</td>
<td>0.28954463</td>
<td>1.00000000</td>
</tr>
<tr>
<td>8</td>
<td>0.89588256</td>
<td>1.00000000</td>
</tr>
<tr>
<td>9</td>
<td>0.94414397</td>
<td>1.00000000</td>
</tr>
<tr>
<td>21</td>
<td>0.94867623</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

n = 100, r = 100, k = 2, p = 0.095.

Expected number of edges for $H \in H_{nr}^k$: $\approx 4702.501$.

Expected number of edges induced by an n-snake: $\approx 470$.

7 is the smallest integer $L$ such that the expected number of n-snakes inducing at most $L$ edges is greater than 1, it is $\approx 5.27$.

The formula in Proposition 4.2 gives the conditional expected value of the number of n-snakes having not more than $L$ edges, given that a particular n-snake has not more than $I$ edges. Choosing $I = (\frac{L}{2})$, conditionality is irrelevant, and for this value of $L$ we should have $\text{Exp}(X_L | Y_1 = 1) = \text{Exp}(X_L)$. Calculations for $n = 10$, $k = 3$, $L = 120$ and various densities always gave inequalities of the form

\[1 - 10^{-7} < \text{computed value of } \frac{\text{Exp}(X_L)}{\text{Exp}(X_L | Y_1 = 1)} < 1.\]

Tables 1–3 give estimates for instances of GMAXSAT(k) of 3 different sizes (problem P2).

In the spirit of (2.1) we give for various $L \in \mathbb{N}$ real numbers $LB$ and $UB$ such that

The probability $P(L)$ of having an n-snake with at most $L$ edges satisfies

\[LB \leq P(L) \leq UB.\]

**Remark.** Let us notice first that no result of interest has been obtained for MAXSAT (binary variables). This is in fact not surprising given that our lower bound LB is the ratio between expectation and conditional expectation, and that in the binary case almost all snakes intersect significantly the one supporting the conditionality. It follows that this conditionality is strong and lets few hopes of having $\frac{\text{Exp}(X_L)}{\text{Exp}(X_L | Y_1 = 1)}$ close to 1 for values of $L$ below the expected number of edges in an n-snake (for the given value of $p$). On the basis of the same argument, good results were expected for variables having many possible values. For instance, with

\[n = 100, \quad r = 20, \quad k = 3, \quad p = 0.002,\]

we got for $L = 3$ an upper bound $UB \approx 0.00018$ and for $L = 6$ a lower bound $LB \approx 0.93$. Said differently, the probability of the minimum of GMAXSAT being not more than 6 is estimated bigger than 0.93, whereas the one that this minimum is 3 or less is bounded by $10^{-4}$. Consequently, the minimum lies with high probability in the interval $[4, 6]$, when, for the given problem, the expected number of edges in an
Table 4
GMAXSAT, binary constraints

<table>
<thead>
<tr>
<th>n, r, k</th>
<th>p</th>
<th>U'999</th>
<th>U'950</th>
<th>U'900</th>
<th>ZESP</th>
<th>L100</th>
<th>L050</th>
<th>L001</th>
</tr>
</thead>
<tbody>
<tr>
<td>50, 50, 2</td>
<td>0.4</td>
<td>12</td>
<td>21</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>10</td>
<td>16</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>9</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>200, 200, 2</td>
<td>0.4</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>500, 500, 2</td>
<td>0.4</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>17</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 5
MAXSAT, binary constraints

<table>
<thead>
<tr>
<th>n, r, k</th>
<th>p</th>
<th>U'999</th>
<th>U'950</th>
<th>U'900</th>
<th>ZESP</th>
<th>L100</th>
<th>L050</th>
<th>L001</th>
</tr>
</thead>
<tbody>
<tr>
<td>50, 2, 2</td>
<td>0.4</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>200, 2, 2</td>
<td>0.4</td>
<td>5</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>500, 2, 2</td>
<td>0.4</td>
<td>7</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>7</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>6</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
n-snake is about 323. Furthermore, 6 is the smallest integer \( Z \) such that the expectation \( E(Z) \) of the number of \( n \)-snakes having \( Z \) edges or less is greater than 1; more precisely, \( E(6) \approx 53 \). This indicates in some sense how rare those \( n \)-snakes are having a number of edges in the interval \([4, 6]\). With such characteristics, problems randomly generated with those parameters \( n, r, k \) and \( p \) seem to be ideal instances for testing algorithms. Unfortunately, the expected number of edges in such hypergraphs is about 2587200, which is beyond the possibilities of most computers at hand.

Tables 4–6 summarize results related to the maximum cardinality of a stable snake (problem \( P1 \)) and should be interpreted as follows:

- \( U999 \) (resp. \( U950, U900 \)) is the biggest integer \( D \) such that the probability of having a stable snake of cardinality greater than or equal to \( D \) is bigger than 0.999 (resp. 0.950, 0.900).

- \( ZESP \) is the biggest integer \( D \) such that the expected number of stable snakes of cardinality greater than or equal to \( D \) is greater than 1.

- \( L100 \) (resp. \( L050, L001 \)) is the smallest integer \( D \) such that the probability of having a stable snake of cardinality greater than or equal to \( D \) is less than 0.100 (resp. 0.050, 0.001).

Good localizations could be obtained for edge density \( p \geq 0.4 \) (the greater \( p \), the better the results). As an example, for

\[
n = 100, \quad r = 100, \quad k = 2, \quad p = 0.4,
\]

---

Table 6
MAXSAT, ternary constraints

<table>
<thead>
<tr>
<th>( n, r, k )</th>
<th>( p )</th>
<th>( U999 )</th>
<th>( U950 )</th>
<th>( U900 )</th>
<th>( ZESP )</th>
<th>( L100 )</th>
<th>( L050 )</th>
<th>( L001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50, 2, 3</td>
<td>0.4</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>200, 2, 3</td>
<td>0.4</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>500, 2, 3</td>
<td>0.4</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>
we got

\[ U^{999} = 23, \quad U^{950} = 26, \quad U^{900} = 26, \quad L^{100} = 28, \quad L^{050} = 28, \quad L^{001} = 28. \]

For such instances of GMAXSAT, with estimated probability at least 0.998, one may find a subset of variables of cardinality in the interval \([23, 27]\) such that the problem restricted to this subset is satisfiable. This interval reduces to \([26, 27]\) for a probability of 0.9.

Appendix

Proof of Lemma 5.3. (a) For \( n_0 \geq 2^{k-2}/C \) and \( n > n_0 \) we have

\[
\frac{C(n - 1)^{k-1}}{n^{k-2}} = \frac{C(n - 1)^{k-1}}{(n - 1 + 1)^{k-2}} \geq \frac{C(n - 1)^{k-1}}{(2(n - 1))^{k-2}} = \frac{C(n - 1)}{2^{k-2}} > 1.
\]

(b) Let

\[ l = k - 1, \quad a = (n + 1)^{k-2}, \quad b = n^{k-2}, \quad c = Cn^{k-1}, \quad d = C(n - 1)^{k-1}. \]

Note that

\[
\frac{d}{b} < \frac{c}{a},
\]

i.e.

\[
\frac{C(n - 1)^{l}}{n^{l-1}} < \frac{Cn^{l}}{(n + 1)^{l-1}},
\]

i.e.

\[(n + 1)^{l-1}(n - 1)^{l} < n^{l}n^{l-1}.
\]

In fact,

\[(n + 1)^{l-1}(n - 1)^{l} = (n^2 - 1)^{l-1}(n - 1) < (n^2)^{l-1}n = n^{l}n^{l-1}.
\]

Consequently, for \( n > n_0 \) we have by (a): \( a, b, c, d > 1, \ a < c, \ b < d \) fulfilling

\[
\frac{a}{b} < \frac{c}{d}.
\]

There exists \( \tau < 1 \) such that \( b = \tau d (\Rightarrow a < \tau c) \). It then follows that

\[
\frac{(e^a)^{n+1}}{(e^b)^n} = e^a\left(\frac{e^a}{e^b}\right)^n < e^a\left(\frac{e^c}{e^d}\right)^n = e^a(e^{(c-d)n})^n < e^c\left(\frac{e^c}{e^d}\right)^n = \frac{(e^c)^{n+1}}{(e^d)^n}.
\]

Proof of inequality (5.3), Theorem 5.4.

\[
C_L \left[ \frac{1}{k-1} \right] \leq C_L(n^{k-1} + (n + 1)(k - 1)n^{k-2})
\]

\[
= \lambda_L \frac{1}{2k(k - 1)} (n^{k-1} + (n + 1)(k - 1)n^{k-2})
\]
\[
= \lambda_L \frac{1}{2k} \left( \left( \frac{n}{k-1} \right)^{k-1} + \frac{(n+1)(k-1)n^{k-2}}{(k-1)^{k-1}} \right)
= \lambda_L \frac{1}{2k} \left( \left( \frac{n}{k-1} \right)^{k-1} + \frac{n^{k-1}}{(k-1)} + \frac{n^{k-2}}{(k-1)^2} \right)
= \lambda_L \frac{1}{2k} \left( k\left( \frac{n}{k-1} \right)^{k-1} + \frac{n^{k-2}}{k-1} \right)
= \lambda_L \left( \frac{1}{2} \left( \frac{n}{k-1} \right)^{k-1} + \frac{1}{2k} \left( \frac{n}{k-1} \right)^{k-2} \right)
< \lambda_L \left( \frac{n}{k-1} \right)^{k-1}.
\]

Acknowledgement

The author wishes to thank R. Burkard, J.-P. Gabriel and C. Mazza for fruitful discussions. He is also grateful to A. Gaillard, P. Jaillet and to an anonymous referee for valuable contributions which led to several improvements in the content and in the exposition of this paper.

References