THE NUMBER OF 3-SAT FUNCTIONS

BY

L. Ilinca

Department of Mathematics, Indiana University
Bloomington IN 47405, USA
e-mail: ilinca@indiana.edu

AND

J. Kahn*

Department of Mathematics, Rutgers University
Piscataway NJ 08854, USA
e-mail: jkahn@math.rutgers.edu

ABSTRACT

With \( G_k(n) \) denoting the number of functions of \( n \) Boolean variables definable by \( k \)-SAT formulas, we prove that \( G_3(n) \) is asymptotic to \( 2^{n + \left( \frac{n}{3} \right)} \).

This is a strong form of the case \( k = 3 \) of a conjecture of Bollobás, Brightwell and Leader stating that for fixed \( k \), \( \log G_k(n) \sim \left( \frac{n}{k} \right) \).

1. Introduction

Let \( X_n = \{x_1, \ldots, x_n\} \) be a collection of Boolean variables. Each variable \( x \) is associated with a positive literal, \( x \), and a negative literal, \( \bar{x} \). Recall that a \( k \)-SAT formula (in disjunctive normal form) is an expression \( \mathcal{C} \) of the form

\[
C_1 \lor \cdots \lor C_t,
\]

with \( t \) a positive integer and each \( C_i \) a \( k \)-clause, that is, an expression \( y_1 \land \cdots \land y_k \), with \( y_1, \ldots, y_k \) literals corresponding to different variables. A formula (1) defines a Boolean function of \( x_1, \ldots, x_n \) in the obvious way; we will

* Supported by NSF grant DMS0701175.

Received May 11, 2010 and in revised form July 1, 2011

869
call such a function a $k$-SAT function. Though we will be concerned here almost exclusively with the case $k = 3$, we leave the discussion general for the moment.

Following Bollobás, Brightwell and Leader [2], we write $G_k(n)$ for the number of $k$-SAT functions of $n$ variables. (Note that the bulk of the $k$-SAT literature—including [2]—works with formulas in conjunctive normal form. Of course a function $f$ is representable by a CNF $k$-SAT formula precisely when $1 - f$ is a $k$-SAT function in the above sense; so our switch to disjunctive normal form has no effect on $G_k(n)$.)

Of course $G_k(n)$ is at most $\exp_2[2^k(n\choose k)]$, the number of $k$-SAT formulas; on the other hand it’s easy to see that

$$(2) \quad G_k(n) > 2^n(2^k - n2^{(n-1)/k}) \sim 2^{n+(n\choose k)}$$

(all formulas obtained by choosing $y_i \in \{x_i, \bar{x}_i\}$ for each $i$ and a set of clauses using precisely the literals $y_1, \ldots, y_n$ give different functions).

The problem of estimating $G_3(n)$ was suggested in [2] (and also, according to [2], by U. Martin). They showed

$$(3) \quad G_k(n) \leq \exp_2[(2\sqrt{\pi})^{(n\choose k)}],$$

for $k < n/2$, and conjectured that

$$(4) \quad \log_2 G_k(n) < (1 + o(1))(n\choose k)$$

for any fixed $k$. Even $k = 2$ is not easy; here (4) was proved in [2], and the precise asymptotics—

$$(5) \quad G_2(n) \sim \exp_2[n + (n\choose 2)]$$

—conjectured in [2] were proved by P. Allen in [1] and (later) in [8]. As is often the case, nothing from this earlier work seems to be of much help in treating larger $k$.

Here, for $k = 3$, we prove (4) and more, again showing (as in (5)) that (2) gives the asymptotics not just of $\log G_3(n)$, but of $G_3(n)$ itself:

**Theorem 1.1:** $G_3(n) \sim 2^{n+(n\choose 3)}$.

For a formula $\mathcal{C}$ as in (1) we may identify the associated function, say $f_{\mathcal{C}}$, with the set (henceforth also referred to as a “$k$-SAT function”) $F(\mathcal{C}) \subseteq \{0, 1\}^n$ of satisfying assignments for $\mathcal{C}$ (that is, $F(\mathcal{C}) = f_{\mathcal{C}}^{-1}(1)$). For our purposes it will also usually be convenient to think of $\mathcal{C}$ as the set $\{C_1, \ldots, C_t\}$ of clauses.