Algorithmic Combinatorics — Wi 2016–17

Shagnik Das — Tibor Szabó

Solution to Exercise Sheet 15, Exercise 3

Exercise 3 A *Boolean variable* can take one of two values — either *true* or *false*. Given a variable x, its negation $\neg x$ takes the opposite value. A *literal* is either a variable or its negation. A *k-clause* is the 'or' of k literals corresponding to distinct variables, and is true if and only if at least one of its literals evaluates to true. Finally, a *k-SAT formula* is the 'and' of a number of k-clauses, and is true if and only if all of its clauses are true. For example, for the 3-SAT formula

$$f(x_1, x_2, x_3, x_4) = (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee \neg x_4) \wedge (x_2 \vee \neg x_3 \vee x_4),$$

we have $f(T,T,T,T) = T$ and $f(F,F,F,F) = T$, but $f(T,F,T,T) = F$, where T represents true and F represents false. In general, we say that a k-SAT formula is *satisfiable* if there is some input for which it evaluates to being true, and so our example f is satisfiable.

Prove that every k-SAT formula where no variable appears in more than $\frac{2^k}{ek}$ clauses is satisfiable.

Solution: Suppose we are given a k-SAT formula f where every variable appears in at most $\frac{2^k}{ek}$ clauses. We shall use the Lovász Local Lemma to show that f is satisfiable.

We shall show that when evaluated for random values of the variables, f is true with positive probability, which in particular implies it is satisfiable. Having no reason to do otherwise, we consider a uniformly random input, where each variable x_i is true or false with probability $\frac{1}{2}$ each, independently of all other variables.[1]

In order for f to be true, each of its k-clauses must be true. Hence, for each clause C, let E_C be the event that C is false. This happens if and only if each of the literals in C is false. By the uniformity, each of the k literals is false with probability $\frac{1}{2}$, and since they correspond to distinct variables, they are independent. Hence $\mathbb{P}(E_C) = 2^{-k}$ for each clause C, and we may take $p = 2^{-k}$.

The event E_C is determined solely by the values of the variables that appear in the clause C, and hence is mutually independent of the clauses that do not use any of the variables in C. The clause C has k variables, each of which appears in at most $\frac{2^k}{ek}$ clauses, and hence there are at most $\frac{2^k}{e}$ clauses, including C, that share a variable with C. Thus E_C is mutually independent of a set of all but at most $\frac{2^k}{e}$ events, and we may take this value to be $d+1$.

Thus $ep(d+1) = e \cdot 2^{-k} \cdot \frac{2^k}{e} = 1$, and so by the Lovász Local Lemma, with positive probability none of the events E_C occur, and hence f is indeed satisfiable.

[1] Actually, there might well be reason to do otherwise. The formula need not be symmetric with respect to the variable x_i: perhaps as literals x_i appears more often than $\neg x_i$. For instance, if x_i only appears in its positive form (that is, the literal $\neg x_i$ is never used), then it makes sense to always set x_i to be true. Some more involved arguments, using generalisations of the Local Lemma, indeed adjust the probabilities to account for the relative frequencies of x_i and $\neg x_i$, although counterintuitively, one should bias the variables towards the *less* frequent literal!