21-737 Probabilistic Combinatorics

Notes on the 2-SAT phase transition

Let F be a random 2-SAT formula on a set of n variables with $m = cn$ clauses where the clauses are chosen uniformly at random from the set of

$$4\binom{n}{2}$$

possible clauses without replacement. So our probability space is the uniform distribution on the collection $\binom{4\binom{n}{2}}{m}$ 2-SAT formulae on n variables. Our goal in this note is to show that if $c > 1$ then the probability that F is satisfiable goes to 0 as n tends to infinity. We follow the arguement of Chvatál and Reed.

Note 1. *If A is a set of $o(n^{1/2})$ distinct clauses then*

$$\mathbb{P}(A \subset F) \sim \left(\frac{m}{4\binom{n}{2}}\right)^{|A|} \sim \left(\frac{c}{2n}\right)^{|A|}$$

Recall that for each 2-SAT formula F we define a digraph $D(F)$ on the vertex set given by the set of $2n$ literals. If $u \vee w$ is a clause in F then we add the arcs $\overline{u} \to w$ and $\overline{w} \to u$ in the digraph. Note that we can view the arcs in $D(F)$ as implications.

For the purposes of this note, we define a bicycle in a formula F to be a cycle in $D(F)$ of the form

$$x, w_1, w_2, \ldots, w_{t-1}, \overline{x}, w_{t+1}, \ldots, w_{2t-1}, x$$

where $x, w_1, w_2, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{2t-1}$ are literals on distinct variables and

$$t = \log^2 n.$$

Note that if $D(F)$ contains such a bicycle then F is not satisfiable. Note further that each bicycle in $D(F)$ corresponds to a set of clauses in F; we will refer to these clauses as the 'clauses in F' without further comment. Let X be the number of bicycles of this form that appear in $D(F)$. We show that $\mathbb{P}(X > 0) \to 1$, and thus the probability of satisfiability goes to 0.

We apply the second moment method. We begin with the expected value.

$$\mathbb{E}[X] \sim (2n)(n-1)_{2t-2}2^{2t-2} \cdot \left(\frac{c}{2n}\right)^{2t} \sim (2n)^{2t-1} \cdot \left(\frac{c}{2n}\right)^{2t} = \frac{c^{2t}}{2n}.$$

As this expected value goes to infinity with n, it suffices to show $Var[X] = o(\mathbb{E}[X]^2)$.

For each potential bicycle A let X_A be the indicator random variable for the event that A appears in $D(F)$. For a fixed bicycle A let N_0 be the number of bicycles B that share no clauses with A. And for $1 \le i \le j$ let $N(i,j)$ be the number of bicycles B that intersect A

in i clauses that span j literals. We have

$$\begin{aligned}
Var[X] &= \sum_{A,B} \mathbb{E}[X_A X_B] - \mathbb{E}[X_A]\mathbb{E}[X_B] \\
&\leq \sum_{A} N_0 \left[\frac{(m)_{4t}}{\left(4\binom{n}{2}\right)_{4t}} - \left(\frac{(m)_{2t}}{\left(4\binom{n}{2}\right)_{2t}} \right)^2 \right] + \sum_{1 \leq i < j} N(i,j) 2 \left(\frac{c}{2n}\right)^{4t-i} \qquad (1) \\
&\leq 2 \sum_{A} \sum_{1 \leq i < j} N(i,j) \left(\frac{c}{2n}\right)^{4t-i}
\end{aligned}$$

Now, we bound $N(i,j)$. Let ℓ be the number of maximal paths spanned by the arcs in the intersection of bicycles A and B. If these bicycles intersect in i clauses that span j vertices along the cycle that defines A then there $j-i$ such paths. So set $\ell = j - i$. We have

$$N(i,j) \leq 2\binom{2t}{2\ell} \cdot (2t)^{\ell} \ell! 2^{\ell} \cdot 4 \cdot (2n)^{2t-j} \leq 8(2t)^{3\ell}(2n)^{2t-j}.$$

The explanation for this bound is as follows. We begin by specifying ℓ disjoint paths in A. We do this by specifying the start and end points as a collection of 2ℓ vertices in A. Once these vertices are selected there are two ways to specify the paths (one is the complement of the other). Our next step is to place these paths in B. There are at most $(2t)^{\ell}$ ways to specify the starting points. Once the starting points are placed we determine which path to start at which point in $\ell!$ ways. Next we note that there are two choices of the direction for each path: One goes around the 'outside' of B, the other traverses the negations. We account for this with the 2^{ℓ}. Note that there may be a choice to switch from the 'inside' to 'outside' (or vice versa) if the path cross the special points x and $\overline{x}$; this is why we multiply by 4. Finally, we choose the literals in B that do not appear in A.

Now that we are ready to bound the variance. We do this in 3 cases.

Case 1. $\ell \geq 2$.

The contribution to (1) in this case can be bounded as follows:

$$(2n)^{2t-1} \cdot (2t)^2 \cdot 8(2t)^{3\ell}(2n)^{2t-j} \cdot \left(\frac{c}{2n}\right)^{4t-i}.$$

The first term is the number of choice for the bicycle A. The second term is a bound on the number of choices of i and j. This is followed by an upper bound on $N(i,j)$, and the final term comes from the probability that both A and B appear. Noting that $t = \log^2 n$ and recalling $\mathbb{E}[X] \sim c^{2t}/2n$ and $c > 1$ we conclude that this contribution to the variance is at most

$$\mathbb{E}[X]^2(2n)^{1+i-j}(2t)^{3\ell}O((\log n)^4) = \mathbb{E}[X]^2 2n \left(\frac{(2t)^3}{2n}\right)^{\ell} O(\log n)^4).$$

As $\ell \geq 2$, we see that this contribution is $o(\mathbb{E}[X]^2)$.

Case 2. $\ell = 1$ and $j \geq t$

We have the following on bound the contribution to (1) for this case:

$$(2n)^{2t-1} \cdot t \cdot 8(2t)^3(2n)^{2t-j} \cdot \left(\frac{c}{2n}\right)^{4t-i}.$$

As in the previous case, the first term is the number of choices for the bicycle A, the second term is the number of choices for j, this is followed by an estimate for $N(i,j)$, and finally the probability term. This contribution is at most

$$\mathbb{E}[X]^2(2n)^{1-\ell} \cdot 2^7(\log n)^6 \cdot \left(\frac{1}{c}\right)^i.$$

As $i \geq t-1$ and $t = \log^2 n$, the c^i in the denominator dominates the $\log^6 n$ in the numerator. Therefore, the total contribution to the variance in this case is $o(\mathbb{E}[X^2])$.

Case 3. $\ell = 1$ and $j < t$

Here we follow the calculation from the previous case, but we are not able to use the c^i in the denominator as it may be too small relative to $\log^8 n$. We make a subtle observation that improves the bound. Lets look again at our estimate for $N(i,j)$. We bound the number of choices for B by first identifying the path in A and the placement of that path in B. There are $O(t^3)$ ways to make these choices. We then set the remaining literals in B one at a time. However, since $\ell = 1$ and $j < t$, when we make these choices we will need to make a choice for x or $\overline{x}$ after the antipodal element has already been set (either in the path or as one of the previous choices of the $2t-j$ choices we make when we set the elements of B that do not appear in A. In both cases, this element is prescibed. We conclude that in this case we have

$$N(i,j) = O(t^3 n^{2t-j-1}).$$

Plugging this improved estimate into the bounds for the previous case, we see that the contribution to variance here is $o(\mathbb{E}[X^2])$.