Lecture 5: Random k-SAT III

In the previous lecture, we studied satisfiability of random k-SAT for fixed k: specifically $k = 2$ and $k = 3$. The probability a random formula will be satisfiable is determined by the number of variables n it has in relationship to the number of clauses m. Let $n := \alpha m$. We sought to find $\alpha^* = \alpha^*(k)$ such that, for a randomly sampled formula F,

- $\alpha > \alpha^* \Rightarrow F$ is unsatisfiable whp.
- $\alpha < \alpha^* \Rightarrow F$ is satisfiable whp.

We now wish to estimate α^* for the case of large k.

1 First moment method for general k

We begin by applying the first moment method (1MM) to the random variable $Z = \#\{x \in \{0,1\}^n : F(x) = 1\}$, which gives the expected number of solutions to a random formula F. Recall that the 1MM tells us that if $\mathbb{E}Z$ vanishes as $n \to \infty$, then F is unsatisfiable whp. We have

$$\mathbb{E}Z = 2^n \left(1 - \frac{1}{2^k}\right)^m = \left[2\left(1 - \frac{1}{2^k}\right)^\alpha\right]^n$$

This inner expression equals 1 when $\alpha^*(k) = \log(\frac{1}{2})/\log(1 - \frac{1}{2^k})$. Using Taylor expansion we get that

$$\alpha^*(k) \approx \frac{-\log(2)}{0 - \frac{1}{2^k}} = \log(2) \cdot 2^k$$

This gives us a lower bound on α^*: we know that for smaller α the expectation vanishes, however, for larger α we cannot rule out the case that random formula are mostly unsatisfiable, but occasionally have a huge number of solutions. To do this, we can bound Z's variance using the second moment method (2MM).

2 Second moment method for general k

Recall that using the Cauchy-Schwarz inequality we can derive the following inequality:

$$\Pr[Z > 0] \geq \frac{(\mathbb{E}Z)^2}{\mathbb{E}Z^2}$$

Thus we can show there is likely a satisfying assignment whenever the second moment of Z is not much larger than the square of its expectation. We will try to find when this is the case. We begin by obtaining

$$\begin{aligned}
\mathbb{E}Z^2 &= \mathbb{E}_F \left(\sum_{x \in \{0,1\}^n} \mathbb{1}\{x \text{ satisfies } F\} \right)^2 \\
&= \sum_{x,y \in \{0,1\}} \mathbb{P}_F \left[x, y \text{ both satisfy } F \right] \\
&= \sum_{x,y \in \{0,1\}} \left(\mathbb{P}_C \left[x, y \text{ both satisfy } C \right] \right)^m
\end{aligned}$$

where the last equality follows from the fact each clause is sampled iid. We further break down the probability that x and y, both sampled iid from $\mathsf{Unif}(\{0,1\}^n)$, satisfy an arbitrary clause C as follows:

$$\begin{aligned}
\mathbb{P}_C[x, y \text{ both satisfy } C] &= 1 - \mathbb{P}[x, y \text{ don't both satisfy } C] \\
&= 1 - \mathbb{P}[x \text{ unsat}] - \mathbb{P}[y \text{ unsat}] + \mathbb{P}[x, y \text{ both unsat}] \\
&= 1 - \frac{1}{2^k} - \frac{1}{2^k} + \mathbb{P}[x, y \text{ both unsat}]
\end{aligned}$$

This last probability term can be expressed

$$\mathbb{P}[x, y \text{ both unsat}] = \frac{\#\{C \text{ that both violate}\}}{\#\{\text{all } C\}}$$

We next define an "overlap" function $r := r(x, y) = \#\{i : x_i = y_i\}$, which we use to state

$$\frac{\#\{C \text{ that both violate}\}}{\#\{\text{all } C\}} = \frac{\binom{r}{k}}{2^k \binom{n}{k}} \approx \left(\frac{r}{2n} \right)^k$$

where we use the fact $\binom{n}{k} \approx \frac{n^k}{k!}$ as shown in the previous lecture. We can now write

$$\begin{aligned}
\sum_{x,y \in \{0,1\}} \left(\mathbb{P}_C \left[x, y \text{ both satisfy } C \right] \right)^m &\approx \sum_{x \in \{0,1\}^n} \sum_{r=0}^{n} \#\{y : r(x,y) = r\} \left(1 - \frac{1}{2^{k-1}} + \left(\frac{r}{2n} \right)^k \right)^m \\
&= 2^n \sum_{r=0}^{n} \binom{n}{r} \left(1 - \frac{1}{2^{k-1}} + \left(\frac{r}{2n} \right)^k \right)^m \\
&:= 2^n \sum_{r=0}^{n} \binom{n}{r} f \left(\frac{r}{n} \right)^m
\end{aligned}$$

where we define a function f for closer analysis. We first observe that when $\frac{r}{n} = \frac{1}{2}$ we have

$$f \left(\frac{r}{n} \right) = \left(1 - \frac{1}{2^{k-1}} + \frac{1}{4^k} \right)^m = \left(1 - \frac{1}{2^k} \right)^m$$

Consider that the contribution to f near $\frac{r}{n} = \frac{1}{2}$ is approximately

$$2^n \cdot \sum_{r=n/2-c\sqrt{n}}^{n/2+c\sqrt{n}} \binom{n}{r} f \left(\frac{r}{n} \right)^m \approx f \left(\frac{1}{2} \right)^m \cdot 2^n \cdot 2^n = \left[2 \left(1 - \frac{1}{2^k} \right)^\alpha \right]^{2n} = (\mathbb{E}Z)^2$$

The 2MM then tells us that for α for which the majority contribution of f comes from its value near $\frac{n}{2}$, we can say F is satisfiable whp. Equivalently, we must show that x, y sampled iid from the uniform distribution over all satisfying assignments to F satisfies $r(x, y) \approx \frac{n}{2}$. As it turns out, this is not the case: the set of satisfying solutions tends to have bias towards either True or False for each of the variables, and this bias grows stronger as α grows. We proceed to show that the majority contribution of f does not come from its value near $\frac{n}{2}$, except for the vacuous case of $\alpha = 0$.

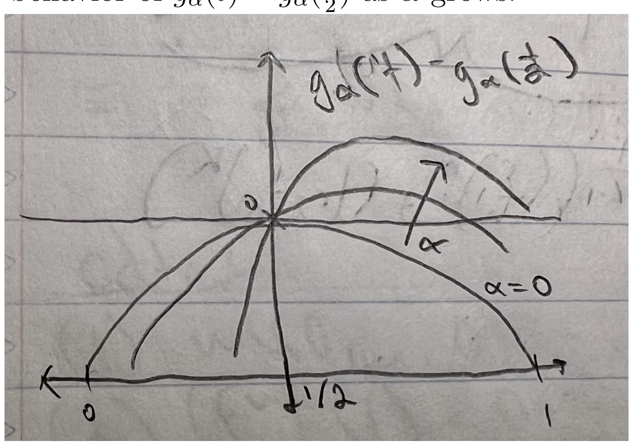
We write $r = tn$ for $t \in [0, 1]$ and apply Stirling's approximation to $\binom{n}{tn}$:

$$\begin{aligned}\binom{n}{tn} &= \frac{n!}{tn!((1-t)n)!} \\ &\approx \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{tn}{e}\right)^{tn}\left(\frac{(1-t)n}{e}\right)^{(1-t)n}} \\ &= \left[\frac{1}{t^t\,(1-t)^{1-t}}\right]^n \\ &= \mathsf{exp}(n[-t\log t - (1-t)\log(1-t)]) =: \mathsf{exp}(n \cdot H(t))\end{aligned}$$

Using this newly defined function $H(t)$ we can then write

$$\binom{n}{tn} f(t)^m \approx \left(e^{H(t)} f(t)^\alpha\right)^n = \mathsf{exp}(n[H(t) + \alpha \log f(t)]) =: \mathsf{exp}(n \cdot g_\alpha(t))$$

Finally, we look at the behavior of $g_\alpha(t) - g_\alpha(\frac{1}{2})$ as α grows.



As claimed, we see that as $\alpha > 0$ grows that increasingly more of the contribution of g_α (and thus f) comes from its value at inputs greater than $\frac{n}{2}$. In the next lecture we'll see how to address the "assignment drift" of satisfying solutions through a restricted counting argument.