

# Lecture 2 - A theorist's toolkit

## 1 Overview of this lecture

In this lecture we show three different examples where we use the probabilistic method to show properties of random objects. In the first example the random object is a *random 3-SAT formula*:

**Definition 1.** We let  $F(n, m)$  denote the probability distribution of 3-SAT formulae consisting of  $n$  variables and  $m$  out of the  $2^3 \binom{n}{3}$  possible clauses.

**Remark 2.** We get the number  $2^3 \binom{n}{3}$  because choosing 3 variables from a set of  $n$  variables can be done in  $\binom{n}{3}$  different ways and there are  $2^3$  different bitmasks that can be applied to each 3-tuple of variables.

In the two other examples the random object in question is *random graphs*.

**Definition 3.** We let  $G(n, p)$  denote the probability distribution over graphs of  $n$  nodes and where a pair of nodes have an edge between them with probability  $p$ .

But first we give an open problem relating to the previous lecture.

## 2 A problem to think about?

In the previous lecture it was shown that for a graph from  $G(n, 1/2)$  there is with high probability no clique of size  $2 \log n + 3$ , but asymptotically there is with high probability a clique of size  $c \log n$  for all constants  $c < 2$ . This motivates the following open problem:

**Open Problem 4.** How can we find a clique of expected size  $c \log n$ , for any constant  $c > 1$ , in a graph from  $G(n, 1/2)$ ?

**Remark 5.** The following algorithm finds a clique of expected size  $\log n$ :

1. Choose a random node  $v$  and put it in the clique.

2. Remove  $v$  and all nodes that are not neighbors to  $v$  from the graph.
3. If the graph is not empty, goto Step 1.
4. Return the clique.

### 3 Example 1: Satisfiability of random 3-SAT formulae

In this example we derive a result about satisfiability of formulae in  $F(n, m)$  depending on the values of  $n$  in relation to  $m$ .

Let  $X$  be the number of satisfying assignments to a formulae in  $F(n, m)$ . Furthermore, let  $\alpha$  be an assignment to the variables and define

$$X_\alpha = \begin{cases} 1 & \text{if } \alpha \text{ is a satisfying assignment} \\ 0 & \text{otherwise} \end{cases}.$$

We can calculate the expected number of satisfying assignments:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{\alpha} X_\alpha\right] = \sum_{\alpha} \mathbb{E}[X_\alpha] = \sum_{\alpha} \left(\frac{7}{8}\right)^m = 2^n \left(\frac{7}{8}\right)^m.$$

If  $\mathbb{E}[X] \leq 1/2$ , then we can conclude that the probability that a random formula is satisfiable is at most one half. For what range of values on  $n$  and  $m$  can we come to this conclusion?

$$1/2 \geq 2^n \left(\frac{7}{8}\right)^m \Rightarrow m \geq (n+1) \log_{8/7} 2$$

We can conclude that if  $m \geq (n+1) \log_{8/7} 2 \approx 5.191n$  then the formula has not a satisfying assignment with probability at least one half.

**Open Problem 6.** *What is the value of  $c$  such that a formula from  $F(n, c'n)$ , when  $n \rightarrow \infty$ , is satisfiable with high probability if  $c' < c$  and it is not satisfiable with high probability if  $c' > c$ .*

**Remark 7.** *It is not even known that there exists such a constant  $c$ , but few people doubt its existence. Experiments suggest that  $c \approx 4.2$ .*

### 4 Example 2: Cycle free graph

We want to find a graph with many edges but with no cycle of length  $r$ , where  $r$  is a constant. We do this by picking a graph from  $G(n, p)$ , where we choose  $p$  such that the expected number of cycles of length  $r$  is less than one half. We will then have a graph with roughly  $p \binom{n}{2}$  edges and probability at least one half that it is cycle free. What is the value of  $p$ ?

Let  $X$  be the number of cycles of length  $r$ . We calculate its expected value:

$$E[X] = \frac{n(n-1)(n-2)\dots(n+1-r)}{r} p^r \approx \frac{n^r p^r}{r} .$$

In order for  $E[X] = 1/2$  we need  $p \sim 1/n$ . This yields a very sparse graph. However, we can modify this approach somewhat in order to generate a more dense  $r$ -cycle free graph.

The idea is to create a random graph with some cycles, but after this we destroy all these cycles by removing an edge in each cycle. A lower bound on the expected number of edges left in the cycle free graph is then  $p \binom{n}{2} - \frac{p^r n^r}{r}$ . In order to find the value of  $p$  that maximizes this value we see for what value of  $p$  the derivative is equal to zero:

$$\begin{aligned} \frac{\partial}{\partial p} \left( p \binom{n}{2} - \frac{p^r n^r}{r} \right) &= 0 \\ p^{r-1} n^r &= \binom{n}{2} \\ p &\approx n^{\frac{2-r}{r-1}} = \frac{1}{n} n^{\frac{1}{r-1}} . \end{aligned}$$

By choosing  $p = n^{\frac{2-r}{r-1}}$  we get that the expected number of edges left in the cycle free graph is at least

$$\begin{aligned} p \binom{n}{2} - \frac{p^r n^r}{r} &= n^{\frac{2-r}{r-1}} \left( \binom{n}{2} - \frac{n^{2-r} n^r}{r} \right) \\ &\in \Omega \left( n^2 n^{\frac{2-r}{r-1}} \right) \\ &= \Omega \left( n^{1+\frac{1}{r-1}} \right) . \end{aligned}$$

**Remark 8.** *If  $r$  is odd then a bipartite graph is  $r$ -cycle free. If  $r$  is even I (Gustav) don't know if there is a known construction that produces denser graphs than the method described above.*

## 5 Example 3: Local and global colorability of graphs

We show that for every  $k$  there exists a constant  $\epsilon > 0$  and a graph  $G$  such that:

1.  $G$  is not  $k$ -colorable.
2. Every subgraph of  $G$  with at most  $\epsilon n$  nodes is 3-colorable.

Instead of looking at the complicated property of colorability we try to use graph properties that are easier to investigate. The following two claims let us do this.

**Claim 9.** *If  $G$  is a graph that can be colored with  $k$  colors then there exists an independent set of size  $n/k$  in  $G$ .*

*Proof.* Assume that  $G$  is colored with  $k$  colors. All nodes that are colored with the same color is an independent set. The average number of nodes colored with the same number is  $n/k$ , thus there exists an independent size of at least size  $n/k$ .  $\square$

**Claim 10.** *A minimal not 3-colorable subgraph that has  $s \geq 4$  nodes has at least  $3s/2$  edges.*

*Proof.* Assume that there is a minimal not 3-colorable subgraph that has less than  $3s/2$  edges. Then there exists a node  $v$  of degree at most 2. Exclude  $v$  and 3-color the rest of the graph. This can always be done because the subgraph was minimal not 3-colorable. But  $v$  can always be colored according to the coloration, because  $v$  only has at most two neighbors and thus only at most two colors are forbidden. Thus, the original subgraph is 3-colorable as well and we have come to a contradiction which completes the proof.  $\square$

We show that a graph from  $G(n, c/n)$  with high probability is not  $k$ -colorable but every subgraph with at least  $\epsilon n$  nodes is with at least probability one half 3-colorable, where  $c$  and  $\epsilon$  are constants that are to be defined.

The expected number of independent sets of size  $n/k$  in a graph from  $G(n, c)$  is

$$\binom{n}{n/k} \left(1 - \frac{c}{n}\right)^{\binom{n/k}{2}} \leq 2^n \left(1 - \frac{c}{n}\right)^{\frac{n}{c} \frac{c}{2k} (n/k-1)} \approx 2^n e^{-\frac{c(n/k-1)}{2k}} .$$

Thus, if we choose  $c > 2k^2$  then there is no independent set of size  $n/k$  with high probability and using Claim 9 we conclude that such a graph is not  $k$ -colorable.

We calculate the expected number of subgraphs that have  $s \leq \epsilon n$  nodes and at least  $3s/2$  edges:

$$\begin{aligned} \sum_{s=4}^{\epsilon n} \binom{n}{s} \binom{\binom{s}{2}}{\binom{3s}{2}} \left(\frac{c}{n}\right)^{3s/2} &\approx \sum_{s=4}^{\epsilon n} \left(\frac{ne}{s}\right)^s \left(\frac{se}{3}\right)^{3s/2} \left(\frac{c}{n}\right)^{3s/2} \\ &= \sum_{s=4}^{\epsilon n} \left(e^{5/2} 3^{-3/2} c^{3/2} (s/n)^{1/2}\right)^s \\ &< \sum_{s=4}^{\epsilon n} \left(\frac{1}{2}\right)^s \\ &< 1/8 . \end{aligned}$$

The second last inequality is valid as long as we choose  $\epsilon \leq 2^{-2}e^{-5}3^3c^{-3}$ . When evaluating the above expression we also used the following approximation:

$$\binom{n}{k} \approx \left(\frac{ne}{k}\right)^k .$$

Thus, we know that with probability larger than  $7/8$  the graph has no subgraphs with  $s \leq \epsilon n$  nodes and at least  $3s/2$  edges. Using Claim 10 we can conclude that with probability larger than  $7/8$  there exists no subgraph of size  $\epsilon n$  that cannot be colored using 3 colors. As almost all graphs from  $G(n, c/n)$  could not be colored with  $k$  colors we conclude that if we choose a graph from  $G(n, c/n)$ , then the graph is with probability at least  $7/8$  not  $k$ -colorable but every subgraph of size at most  $\epsilon n$  is 3-colorable.