CS271 Randomness & Computation **Fall 2022**

Lecture 6: September 13

Instructor: Alistair Sinclair

The probabilistic method is non-constructive, in that it merely proves the existence of objects with certain properties rather than explicitly constructing them. In some cases, there is an easy way to make the method algorithmic, which we now illustrate with a simple example. (More sophisticated tools for making the method constructive will be discussed later in the class.)

6.1 MAX3SAT

The MAX3SAT problem is defined as follows. Given a boolean formula φ in 3CNF form on variables $\{x_1, \cdots, x_n\}$ and clauses $\{C_1, \cdots, C_m\}$, find the maximum number of clauses that can be satisfied by some truth assignment to the variables. (For the purposes of this section, we assume that an instance of MAX3SAT has exactly three literals per clause, all of whose variables are distinct.)

This is the optimization version of the 3SAT problem, which is NP-complete. Therefore, MAX3SAT is an NP-hard optimization problem. However, a simple probabilistic argument yields a surprisingly good lower bound on the optimum value for this problem.

Claim 6.1 *For every such φ, there exists an assignment satisfying at least $\frac{7m}{8}$ clauses.*

Proof: Pick a truth assignment to $\{x_1, \cdots, x_n\}$ uniformly at random. Let the random variable X denote the number of clauses satisfied, and write $X = \sum_i X_i$, where each X_i is an indicator r.v. defined by

$$X_i = \begin{cases} 1 & \text{if } C_i \text{ is satisfied;} \\ 0 & \text{otherwise.} \end{cases}$$

Notice now that

$$\mathrm{E}[X_i] = \Pr[C_i \text{ satisfied}] = \frac{7}{8}$$

as there exist 8 equiprobable truth assignments to the variables of C_i and for only one of these do all the literals of C_i have value false. By linearity of expectation,

$$\mathrm{E}[X] = \sum_{i=1}^{m} \mathrm{E}[X_i] = \frac{7}{8}m.$$

Finally, note that there must exist a point in the sample space at which X takes value at least $\mathrm{E}[X]$. Hence there exists an assignment satisfying at least $\frac{7m}{8}$ clauses. ■

Note that we can apply exactly the same argument to CNF formulas with clauses of varying lengths (and indeed for general *constraint satisfaction problems*, defined by a conjunction of more general boolean constraints). In general, the number of clauses we can satisfy is at least $\sum_i p_i$, where p_i is the probability that the ith clause is satisfied by a random assignment. The values p_i can easily be computed by inspection.

6.1.1 Finding a good assignment

Having established that there exists an assignment that satisfies at least $\frac{7}{8}m$ clauses, can we actually find one?

An obvious approach is to directly apply the above randomized construction, i.e., simply pick a random assignment and hope that it satisfies (close to) the expected number of clauses. This approach can be analyzed by an easy application of Markov's inequality, which says that, for a non-negative random variable Z with expectation $\mathrm{E}[Z]$, and any $\alpha > 0$,

$$\Pr\left[Z \geq \alpha \mathrm{E}[Z]\right] \leq \frac{1}{\alpha}.$$

Exercise: Let X be the random variable above, denoting the number of satisfied clauses in a random assignment. For any $1 < \alpha < 8$, show that

$$\Pr\left[X \leq \left(1 - \frac{\alpha}{8}\right) m\right] \leq \frac{1}{\alpha}.$$

[Hint: Apply Markov's inequality to the random variable $Z = m - X$.] Hence deduce that a random assignment satisfies at least a $\frac{3}{4}$ fraction of the clauses with probability at least $\frac{1}{2}$.

Exercise: Use Markov's inequality and the fact that X is integer-valued to show that we actually get

$$\Pr\left[X \geq \frac{7}{8}m\right] \geq \frac{1}{1+m}.$$

Deduce that we can actually achieve the expected value of $\frac{7}{8}m$ in polynomial time with high probability.

6.1.2 Method of conditional probabilities

In many examples, it is possible to efficiently *derandomize* the randomized construction used in the probabilistic method, and thus achieve the expected value (or better) *deterministically*. In its simplest incarnation, this technique is usually called the "method of conditional probabilities." We illustrate it using our MAX3SAT example.

Consider a 3CNF formula φ. We can think of the random construction of a truth assignment as sequentially assigning truth values: first pick a T/F value for x_1, then for x_2, and so on. This process can be pictured as a tree (see Figure 6.1).

We index each node of the tree with a formula Ψ: the formula at the root is φ, and to get the formula at a node at level i we simply replace the variables $x_1, \ldots, x_i$ in φ with their appropriate T/F values. Note that the expression associated with each node is similar to the original expression φ, except that some clauses may have less than three literals in them.

Also, for each node Ψ in the tree, we define the random variable X_Ψ to be the number of clauses that are satisfied in the tree below Ψ (i.e., when the assignments to the remaining variables are made randomly). Now consider the expectation $\mathrm{E}[X_\Psi]$, where Ψ is a node at level i (so that the next variable to be assigned is x_{i+1}). Using conditional expectation,

$$\mathrm{E}[X_\Psi] = \Pr[x_{i+1} = T] \cdot \mathrm{E}[X_{\Psi|_{x_{i+1}=T}}] + \Pr[x_{i+1} = F] \cdot \mathrm{E}[X_{\Psi|_{x_{i+1}=F}}] = \frac{1}{2}(\mathrm{E}[X_{\Psi_1}] + \mathrm{E}[X_{\Psi_2}]),$$

where Ψ_1 and Ψ_2 are the children of Ψ. Thus at least one of the children must have an expectation at least as large as that of Ψ.

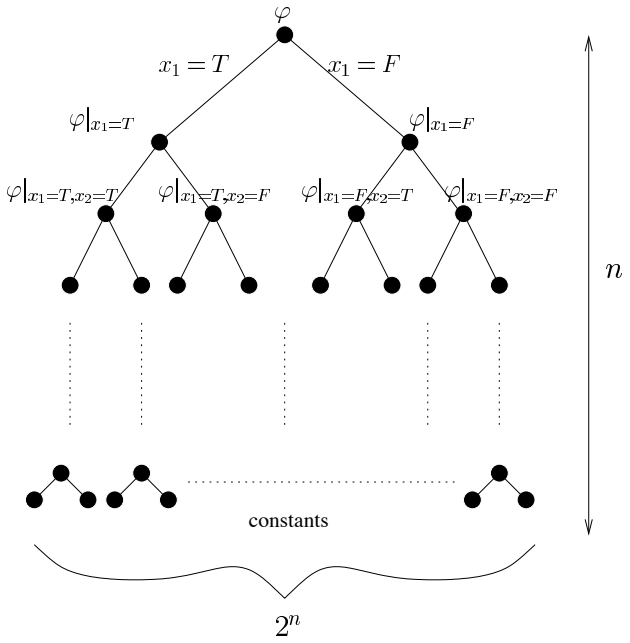

Figure 6.1: The self-reducibility tree of a formula φ with variables $x_1, x_2, \ldots, x_n$.

Now, note that at the root, $\mathrm{E}[X_\Psi] \geq \frac{7}{8}m$, when $\Psi = \varphi$. Hence, at each node we can choose a child such the expectation remains greater than $\frac{7m}{8}$. Moreover, note that for any partially assigned formula, it is possible to compute *exactly* the probability that each of its clauses is satisfied, and hence the expectation $\mathrm{E}[X_\Psi]$, in linear time. So we can in linear time determine the child of Ψ with the larger expectation. We choose this child and proceed down the tree, while maintaining the invariant that $\mathrm{E}[X_\Psi] \geq \frac{7}{8}m$. When we hit a leaf of the tree (after n levels) we will have a complete assignment, and since the invariant still holds that this assignment must in fact satisfy at least $\frac{7m}{8}$ clauses. ■

The above approach can be made to work in many of our examples. (**Exercise:** which ones?) The key ingredient is the ability to compute the expectation when some of the random choices have already been made. Even when this is not possible exactly, it is sometimes possible to compute the expectation approximately and thus to deterministically achieve a final result that is close to the expectation.

6.2 Variance and the second moment method

Markov's Inequality gives the best tail bound when the expectation is all we know about a non-negative random variable. In this lecture we will explore how this can be improved if additional information is available, such as higher moments. In particular we are interested in using the variance/second moment to obtain a tighter bound.

Definition 6.2 *The* variance *of a random variable X is defined as* $\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}X)^2] = \mathrm{E}X^2 - (\mathrm{E}X)^2$.

Intuitively variance measures how far the random variable is likely to be from its expectation. The following standard inequality makes this intuition quantitative.

Lemma 6.3 (Chebyshev's Inequality) $\Pr[|X - \mathrm{E}X| \geq \alpha] \leq \frac{\mathrm{Var}(X)}{\alpha^2}$.

Proof: Let $Y = (X - \mathrm{E}X)^2 \geq 0$. Applying Markov's Inequality to Y gives

$$\Pr[|X - \mathrm{E}X| \geq \alpha] = \Pr[Y \geq \alpha^2] \leq \frac{\mathrm{E}Y}{\alpha^2} = \frac{\mathrm{Var}(X)}{\alpha^2}. \tag{6.1}$$

■

Corollary 6.4 $\Pr[|X - \mathrm{E}X| \geq \beta\sqrt{\mathrm{Var}(X)}] \leq \frac{1}{\beta^2}$. [*Note that* $\sqrt{\mathrm{Var}(X)}$ *is the* standard deviation *of* X.]

Corollary 6.5 $\Pr[|X - \mathrm{E}X| \geq \beta\mathrm{E}X] \leq \frac{1}{\beta^2}\frac{\mathrm{Var}(X)}{(\mathrm{E}X)^2}$.

Exercise: Show that $\Pr[|X - \mathrm{E}X| \geq \beta\sqrt{\mathrm{Var}(X)}] \leq \frac{2}{1+\beta^2}$. Hence deduce that Chebyshev's inequality is not necessarily tight (i.e., it does not provide the tightest bound possible given $\mathrm{E}X$ and $\mathrm{Var}(X)$).

Notwithstanding the previous exercise, we cannot do much better than Chebyshev's inequality if all we know are the mean and variance of a random variable. However, in many cases we have much more information, which enables us to prove much tighter tail bounds. For example, if the random variable X is normally distributed with mean μ and variance σ^2, i.e., $X \sim N(\mu, \sigma^2)$, then

$$\Pr\left[|X - \mu| \geq \beta\sigma\right] \approx \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-\frac{\beta^2}{2}}}{\beta} \ll \frac{1}{\beta^2},$$

which is much sharper than Corollary 6.4. Similarly tight bounds hold when X is the sum of a large number of *independent* random variables with bounded range. We shall discuss this at greater length in a later lecture.

6.3 Thresholds in random graphs

As a first example, we will apply the variance/second moment method to find thresholds in random graphs.

In the $\mathcal{G}_{n,p}$ model of random graphs, a graph G with n vertices is constructed by including each of the $\binom{n}{2}$ possible edges independently with probability p. Thus E[number of edges] $= p\binom{n}{2}$ and E[degree of any vertex] $= p(n-1)$.

Typical Questions:

- Is $G \in \mathcal{G}_{n,p}$ connected?
- Does $G \in \mathcal{G}_{n,p}$ contain a 4-clique?
- Does $G \in \mathcal{G}_{n,p}$ contain a Hamilton cycle?

We will now attack the second of these questions, i.e., we'll look at the property that G contains a 4-clique. This will serve as an illustration of some general ideas.

Let the random variable X denote the number of 4-cliques in G. We look first at the expectation $\mathrm{E}X$ (i.e., the *first moment*). As usual we write $X = \sum_C X_C$, where C ranges over all subsets of four vertices and

$$X_C = \begin{cases} 1, & \text{if } C \text{ is a 4-clique;} \\ 0, & \text{otherwise.} \end{cases}$$

Now $\mathrm{E}X_C$ is the probability that C is a clique, which is just p^6 (as there must be six edges within C). So by linearity of expectation we have

$$\mathrm{E}X = \sum_C \mathrm{E}X_C = \binom{n}{4} p^6 = \theta(n^4 p^6).$$

Thus we see that:

- If $p = o(n^{-2/3})$, then $\mathrm{E}X \to 0$;
- If $p = \omega(n^{-2/3})$, then $\mathrm{E}X \to \infty$.

(Recall that the notation $f(n) = o(g(n))$ means that $f(n)/g(n) \to 0$ as $n \to \infty$, and $f(n) = \omega(g(n))$ means that $f(n)/g(n) \to \infty$ as $n \to \infty$.)

Can we translate this sharp jump in the expectation into a stronger statement about probabilities? It turns out we can. We call $p(n)$ a *threshold* for a property Q if

$$\begin{aligned} p = \omega(p(n)) &\Rightarrow \Pr[G \in \mathcal{G}_{n,p} \text{ has } Q] \to 1 \text{ as } n \to \infty, \text{ and} \\ p = o(p(n)) &\Rightarrow \Pr[G \in \mathcal{G}_{n,p} \text{ has } Q] \to 0 \text{ as } n \to \infty. \end{aligned}$$

Theorem 6.6 *The value $p(n) = n^{-2/3}$ is a threshold for containing a 4-clique.*

Proof: As above, let the random variable X denote the number of 4-cliques in G.

One direction is easy. Since X is an integer-valued random variable, $\Pr[X > 0] = \Pr[X \geq 1] \leq \mathrm{E}X$, by Markov's inequality. Since $\mathrm{E}X \to 0$ for $p = o(n^{-\frac{2}{3}})$, we deduce that $\Pr[G \text{ contains a 4-clique}] \to 0$, as desired.

The other direction is less trivial: notice that $\mathrm{E}X \to \infty$ does not immediately imply a useful lower bound on $\Pr[X > 0]$, since X could be 0 most of the time and very large with small probability. We therefore need to use the *second moment* of X. Specifically, we will show the following:

Claim 6.7 *When $p = \omega(n^{-2/3})$, we have $\frac{\mathrm{Var}(X)}{(\mathrm{E}X)^2} \to 0$ as $n \to \infty$.*

Note that the claim is equivalent to saying that $\mathrm{E}(X^2) = (1 + o(1))(\mathrm{E}X)^2$, i.e., that the second moment is asymptotically no larger than the square of the mean.

Before proving the claim, let us see how it implies the second direction of our theorem. By Chebyshev's Inequality, $\Pr[X = 0] \leq \Pr[|X - \mathrm{E}X] \geq \mathrm{E}X] \leq \frac{\mathrm{Var}(X)}{(\mathrm{E}X)^2}$. Therefore, by the claim, this latter ratio tends to zero as desired.

Proof of Claim 6.7: Expanding the variance of $X = \sum_C X_C$ we have

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}X^2 - (\mathrm{E}X)^2 && (6.2) \\ &= \sum_C \mathrm{E}X_C^2 - \sum_C (\mathrm{E}X_C)^2 + \sum_{C \neq D} \mathrm{E}(X_C X_D) - \sum_{C \neq D} \mathrm{E}X_C \mathrm{E}X_D && (6.3) \\ &= \sum_C \mathrm{Var}(X_C) + \sum_{C \neq D} \mathrm{Cov}(X_C, X_D), && (6.4) \end{aligned}$$

where the covariance of two random variables Y, Z is defined as $\mathrm{Cov}(Y, Z) = \mathrm{E}(YZ) - \mathrm{E}Y\mathrm{E}Z$. Note that $\mathrm{Cov}(Y, Z) = 0$ if Y, Z are independent, and otherwise can be positive, negative or zero.

Since the X_C are {0,1} random variables, we have $\sum_C \mathrm{Var}(X_C) = \sum_C (\mathrm{E}X_C - (\mathrm{E}X_C)^2) = \binom{n}{4}(p^6 - p^{12}) = \Theta(n^4p^6)$. To analyze the covariance terms, we consider three cases:

- Case 1: $|C \cap D| \leq 1$. In this case C, D share zero or one vertex, so X_C, X_D are independent and $\mathrm{Cov}(X_C, X_D) = 0$.
- Case 2: $|C \cap D| = 2$. Here C, D have two vertices in common, and $\mathrm{Cov}(X_C, X_D) \leq \mathrm{E}(X_C X_D) = \Pr[C, D \text{ are both cliques}]$. For each choice of C, D, this latter event requires that 11 specific edges are present, so the probability is p^{11}. Since there are $\binom{n}{6}\binom{6}{2} = \Theta(n^6)$ such pairs C, D, the total contribution of this covariance term is $\Theta(n^6p^{11})$.
- Case 3: $|C \cap D| = 3$. By similar reasoning to Case 2, the contribution from the covariance of such pairs C, D is $\Theta(n^5p^9)$.

Plugging these into (6.4) gives

$$\mathrm{Var}(X) = O(n^4p^6) + O(n^6p^{11}) + O(n^5p^9).$$

Now using the fact that $\mathrm{E}X = \Theta(n^4p^6)$ we get

$$\frac{\mathrm{Var}(X)}{(\mathrm{E}X)^2} = O(\frac{1}{n^4p^6}) + O(\frac{1}{n^2p}) + O(\frac{1}{n^3p^3}),$$

and each of these terms tends to 0 as $n \to \infty$ assuming that $p = \omega(n^{-2/3})$. This completes the proof of the claim and of the theorem. ■

We can generalize the above calculation to establish a threshold for containment of *any* fixed graph H (not just 4-cliques). The expected number of copies of H in a random graph $G \in \mathcal{G}_{n,p}$ is $\Theta(n^v p^e)$, where v, e are the numbers of vertices and edges respectively in H (**Exercise:** Why? Note that the constant in the Θ-notation conceals a factor that depends on the number of automorphisms of H.) Thus the natural candidate for a threshold is $p = n^{-v/e}$. [Note that the ratio v/e is just $2/d$, where d is the average degree of H.]

Call H *balanced* if the average degree of H is greater than or equal to the average degree of any induced subgraph of H. [Thus the 4-clique $H = K_4$ is balanced; but if we add a fifth vertex connected to just one of the original four vertices, then the resulting graph is not balanced.]

Theorem 6.8 *If H is balanced, then $p = n^{-v/e}$ is a threshold for containment of H.*

Exercise: Prove this theorem by mimicking the proof of Theorem 6.6.

For a general graph H, the theorem still holds but the ratio v/e now has to be minimized over all induced subgraphs of H. (Proof: another **exercise**.)

Theorem 6.8 was one of several threshold results proved in the seminal paper of Erdős and Rényi [ER60], which laid the foundations of the theory of random graphs. The more general version (for non-balanced H) is due to Bollobás [B81].

6.4 Behavior at the threshold

What happens when $p = c \cdot p(n)$ for some constant c ? The definition of threshold doesn't say anything about this. We will now briefly discuss the behavior at the threshold for some particular properties.

6.4.1 4-Cliques

Fact 6.9 *For $p = cn^{-2/3}$, let X be the number of 4-cliques. Then X is asymptotically Poisson with parameter $c^6/24$.*

(As a somewhat technically involved **exercise**, you are invited to prove this Fact. You need to show that, for each fixed k, $\Pr[X = k]$ approaches $e^{-\lambda}\frac{\lambda^k}{k!}$, where $\lambda = \frac{c^6}{24}$.) This implies that $\Pr[X > 0] \to 1 - e^{-c^6/24}$. Observe that as c varies, the probability that G contains a 4-clique varies smoothly. This is therefore called a "coarse threshold".

6.4.2 Connected components

Fact 6.10 *$p(n) = n^{-1}$ is a threshold for the property $Q \equiv$ "G has a connected component of size $\theta(n)$". Moreover, for $p(n) = c/n$, the size of the largest component in G is almost surely*

$$\begin{cases} \Theta(\log n) & \text{if } c < 1; \\ \Theta(n^{2/3}) & \text{if } c = 1; \\ \Theta(n) & \text{if } c > 1. \end{cases}$$

(By "almost surely" we mean "with probability tending to 1 as $n \to \infty$".) Since the behavior depends in detail on the value of the constant c there must be some activity in the lower order terms. It turns out that, if we set $p(n) = n^{-1} + c' \cdot n^{-4/3}$ then we get smooth behavior as c' varies. The "width of the transition" is thus $\theta(n^{-4/3})$. This is an example of a "sharp threshold."

6.4.3 Monotone properties

A graph property is said to be *monotone increasing* if adding edges to G cannot destroy the property (i.e., whenever G has the property, so does any graph obtained by adding edges to G). *Monotone decreasing* properties are defined analogously. A host of natural graph properties (including all the ones discussed above) are monotone. The following theorem of Bollobás and Thomason [BT86] confirms that the threshold phenomenon is ubiquitous:

Theorem 6.11 *Every monotone increasing (or decreasing) graph property has a threshold.*

6.5 Random graphs with constant edge probability

In our examples so far, the threshold values $p(n)$ were pretty small (e.g., $n^{-2/3}$, n^{-1}). It is natural to look at the case $p = 1/2$, i.e., the random graph model $\mathcal{G}_{n,1/2}$, in which every graph on n vertices has equal probability. For example, we know that almost every graph in this model has a k-clique for any fixed k, since the threshold is $n^{-k/\binom{k}{2}} = n^{-2/(k-1)} \ll 1/2$. This is a special case of the following more general theorem (which we will not prove) about properties of almost every graph:

Theorem 6.12 (Fagin [F76]) *For any* first order *property Q, and for any constant $p \in (0,1)$, either almost every $G \in \mathcal{G}_{n,p}$ has Q or almost every $G \in \mathcal{G}_{n,p}$ does not have Q.*

(The term "almost every" here means "with probability tending to 1 as $n \to \infty$".) Informally, we can define a first order property as one expressible as a finite sentence, using $\forall, \exists, \vee, \wedge, \neg$, variables denoting vertices of the graph, and the relation $\sim$ denoting adjacency in the graph.

Example 6.13 *"G has diameter 2" is a first order property as it can be expressed as*

$$\forall x \forall y \exists z((x = y) \vee (x \sim y) \vee ((x \sim z) \wedge (z \sim y))).$$

Fact 6.14 *The following properties hold for almost every G in $\mathcal{G}_{n,1/2}$:*

- *G has diameter 2.*
- *G does not have diameter 1.*
- *G is Hamiltonian.*
- *G is connected.*

The first two of these follow from the above theorem about first-order properties. The last two are not first-order properties and require separate proofs.

6.6 The clique number of a random graph

The clique number of a graph G is the size of a largest clique in G. Determining the clique number of a graph is a famous NP-hard problem. However, the following theorem says that the clique number of a *random* graph is known rather precisely:

Theorem 6.15 *For $G \in \mathcal{G}_{n,p}$ for any constant $p \in (0,1)$, the clique number of G is almost surely $\sim 2\log_{1/p}(n)$.*

In particular, when $p = 1/2$ the maximum clique size is almost surely $\sim 2\log_2 n$. We will see how to prove this theorem in the next lecture.

In fact, even more is known:

Theorem 6.16 (Bollobás and Erdős [BE76], Matula [M76]) *For $p = 1/2$, the maximum clique in almost every graph G has size either $k(n)$ or $k(n)+1$, for some integer $k(n)$.*

In other words, the clique number of almost every graph is specified within two adjacent integers!

As an example, for $n = 1000$ and $p = 1/2$, the largest clique size is 15 or 16 with high probability.

Challenge: Find a polynomial-time algorithm that outputs a clique of size $\geq (1+\epsilon)\log_2 n$ with high probability in almost every graph G. This is a major open problem in average-case complexity. The folklore says that all known algorithms find a clique of size no larger than (and usually as large as) $(1-\epsilon)\log_2 n$. Even though we know that G almost certainly contains a clique of size $2\log_2 n$, we don't know how to find one more than half this size!

References

[B81] B. Bollobás, "Random graphs," in *Combinatorics*, Proceedings, Swansea 1981, London Mathematical Society Lecture Note Series **52**, Cambridge University Press, pp. 80–102.

[BE76] B. Bollobás and P. Erdős, "Cliques in random graphs," *Mathematical Proceedings of the Cambridge Philosophical Society* **80** (1976), pp. 419–427.

[BT86] B. Bollobás and A. Thomason, "Threshold functions," *Combinatorica* **7** (1986), pp. 35–38.

[ER60] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), pp. 17–61.

[F76] R. Fagin, "Probabilities in finite models," *Journal of Symbolic Logic* **41** (1976), pp. 50–58.

[M76] D.W. Matula, *The largest clique size in a random graph*, Technical Report, Southern Methodist University, Dallas, 1976.