**Mathematical Toolkit** **Autumn 2025**

Lecture 12: November 6, 2025

Lecturer: Madhur Tulsiani

# 1 Randomized polynomial identity testing

We use our knowledge of events and conditioning, to prove the following lemma, which gives an algorithm for testing if a polynomial $f$ in $n$ variables $x_1, \ldots, x_n$ over a field $\mathbb{F}$ is identically zero. While this is usually referred to as the Schwartz-Zippel lemma, or the DeMillo-Lipton- Schwartz-Zippel lemma, it actually has a longer history as described in (Section 3.1 of) this article by Arvind et al. [AJMR19]. We refer to it as the polynomial identity lemma.

**Lemma 1.1 (Polynomial identity lemma)** *Let $f(x_1, x_2, \ldots, x_n)$ be a non-zero polynomial of degree $d \geq 0$, i.e.,*

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \sum c_{i_1 i_2 \ldots i_n} \cdot x_1^{i_1} \cdot x_2^{i_2} \cdots x_n^{i_n} \\ s.t., &\quad i_1 + i_2 + \ldots + i_n \leq d \end{aligned}$$

*over a field, $\mathbb{F}$. Let $S \subseteq \mathbb{F}$, be a finite subset and let $x_1, x_2, \ldots, x_n$ be selected uniformly at random from $S$, independently. Then,*

$$\mathbb{P}\left[f(x_1, x_2, \ldots, x_n) = 0\right] \leq \frac{d}{|S|}.$$

**Proof:** We will prove this lemma by induction on $n$. This lemma can be proved simply by using conditional probability.

*Base Case*: n = 1

A non zero polynomial, $f(x_1)$ can have at most $d$ roots. Hence, $\mathbb{P}\left[f(x_1) = 0\right] \leq \frac{d}{|S|}$.

*Induction Step*

Assume that the lemma holds for any polynomial in $n-1$ variables. We need to prove that it holds true for $f(x_1, x_2, \ldots, x_n)$. We can write $f$ as:

$$f(x_1, x_2, \ldots, x_n) = x_1^k \cdot g(x_2, \ldots, x_n) + h(x_1, x_2, \ldots, x_n)$$

where, $k$ is largest degree of $x_1$. Thus we have $0 < k \leq d$ (if $k = 0$ then we are already done). We also have that $\deg(g(x_2, \ldots, x_n)) \leq d - k$.

Now let us define two events.

$$E \equiv \{f(x_1, x_2, \ldots, x_n) = 0\} \quad \text{and} \quad F \equiv \{g(x_2, \ldots, x_n) = 0\}$$

We can then write,

$$\mathbb{P}[E] = \mathbb{P}[F] \cdot \mathbb{P}[E|F] + \mathbb{P}[\neg F] \cdot \mathbb{P}[E|\neg F] \,.$$

We now analyze each of the terms. By the induction hypothesis, we have

$$\mathbb{P}[F] = \mathbb{P}[g(x_2, \ldots, x_n) = 0] = \frac{d-k}{|S|} \,.$$

Also, fixing the values of $x_2 = a_2, \ldots, x_n = a_n$ such that $g(a_2, \ldots, a_n) \neq 0$, $f(x_1, a_2, \ldots, a_n)$ is a degree-$k$ polynomial in $x_1$. Thus, using the base case, we get that

$$\mathbb{P}[E|\neg F] \leq \frac{k}{|S|} \,.$$

Bounding the other two probabilities by 1, we get that

$$\mathbb{P}[E] \leq \frac{d-k}{|S|} \cdot 1 + 1 \cdot \frac{k}{|S|} = \frac{d}{|S|}$$

as desired. ■

### 1.1 An application: bipartite perfect matching

Consider the following example which applied the Schwartz-Zippel lemma for testing if a given bipartite graph has a perfect matching. Given a bipartite graph, $G = (U, V, E)$ with $|U| = |V| = n$, we say that the graph has a perfect matching, if there exists a set $E' \subseteq E$ of $n$ edges, with exactly one edge in $E'$ being incident on every vertex of $G$.

Let us define the Tutte matrix $A$ as

$$A_{ij} = \begin{cases} x_{ij} & \text{if (i, j)} \in E \\ 0 & else \end{cases}$$

Note that $A$ is not necessarily symmetric. The determinant of $A$ can be written as,

$$\mathrm{Det}(A) = \sum_{\pi:[n]\to[n]} \mathrm{sign}(\pi) \prod_{i=1}^{n} A_{i,\pi(i)}$$

where $\pi$ defines the permutation from rows to columns. Note that the determinant is a degree-$n$ polynomial in the variables $x_{ij}$. Verify the follwing:

**Exercise 1.2** *$G$ has a perfect matching if and only if $Det(A) \not\equiv 0$.*

In this case, computing the determinant is expensive with $n!$ terms. But if we are given the values of the variables $x_{ij}$, we can simply compute the determinant using the Gaussian elimination method. The degree of the polynomial above is $n$. Thus, if we assign all variables randomly from a set of $2n$ real values, if $\mathrm{Det}(A) \not\equiv 0$, we will detect it with probability at least $1/2$.

The randomized algorithm given by the polynomial identity lemma can be used to parallelize the checking as well. There is no known deterministic algorithm for this problem which can be parallelized efficiently.

## 2 The probabilistic method

We now come to very powerful method for proving the existence of several interesting combinatorial objects. The general framework, known as the "probabilistic method" has many variants explored in the beautiful (and highly recommended!) book on the subject by Alon and Spencer [AS08].

We will explore vanilla version of the method, known as the first moment method, which only requires computing expectations. At the heart of it is the simple idea captured by the following proposition.

**Proposition 2.1** *Let $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}[X] \geq c$ for some $c \in \mathbb{R}$. Then, there exists $\omega \in \Omega$ (with probability measure $\nu(\omega) > 0$) such that $X(\omega) \geq c$.*

**Proof:** Suppose that for all $\omega \in \Omega$ with $\nu(\omega) > 0$, we have $X(\omega) < c$. Then,

$$\mathbb{E}[X] \;=\; \sum_{\omega\in\Omega} \nu(\omega)\cdot X(\omega) \;<\; \sum_{\omega\in\Omega} \nu(\omega)\cdot c \;=\; c\,,$$

which contradicts the fact that $\mathbb{E}[X] \geq c$. ■

**Exercise 2.2** *Prove that if $\mathbb{E}[X] \leq c$, then there exists $\omega \in \Omega$ (with $\nu(\omega) > 0$) such that $X(\omega) \leq c$.*

**Exercise 2.3** *Is it true that if $\mathbb{E}[X] = c$, then there exists $\omega \in \Omega$ with $X(\omega) = c$?*

The above simple proposition can yield very interesting results, when the random variable $X$ is set-up properly. In particular, when we want $X$ to measure some property of a combinatorial object, and we set up the distribution such that $\mathbb{E}[X]$ is close to some bound we are interested in, we get that there exists a combinatorial object achieving those bounds. We will see a few examples of this principle.

### 2.1 A randomized algorithm for Max 3-SAT

Recall that a 3-SAT formula $\varphi$ is of the form

$$\varphi \equiv C_1 \wedge \cdots \wedge C_m ,$$

where each $C_i$ is a clause of the form $C_i = (l_{i_1} \vee l_{i_2} \vee l_{i_3})$ and each $l_{i_j}$ is in turn $x_{i_j}$ or its negation $\overline{x}_{i_j}$. We assume that each clause contains three *distinct* variables.

In the problem Max 3-SAT, the goal is not necessarily to satisfy all the clauses, but rather find an assignment to the variables which satisfies as many clauses as possible. We show that for any formula $\varphi$ with $m$ clauses, there exists an assignment satisfying $7m/8$ clauses. Moreover, this can be turned into an algorithm, and one can efficiently find an assignment satisfying $7m/8$ clauses.

Consider assigning each of the variables $x_1, \ldots, x_n$ a value in $\{0,1\}$ independently at random. Let $Z$ be a random variable equal to the number of clauses satisfied by the random assignment. We can write

$$Z = Y_1 + \cdots + Y_m ,$$

where $Y_i$ if the clause $C_i$ is satisfied and 0 otherwise. By linearity of expectation $\mathbb{E}[Z] = \sum_{i=1}^{m} \mathbb{E}[Y_i]$. Note $C_i = (l_{i_1} \vee l_{i_2} \vee l_{i_3})$ is not satisfied if and only if $l_{i_1} = l_{i_2} = l_{i_3} = 0$ which happens with probability $1/8$ since the three literals correspond to three distict variables, which are assigned values 0 and 1 independently with probability $1/2$ each. Thus, $\mathbb{P}[Y_i = 0] = 1/8$, which gives

$$\mathbb{E}[Z] = \sum_{i=1}^{m} \mathbb{E}[Y_i] = \sum_{i=1}^{m} \left(1 - \frac{1}{8}\right) = \frac{7m}{8} .$$

Thus, there *exists* an assignment which satisfies at least $7m/8$ clauses. We now argue that it can be found efficiently. Note that

$$\mathbb{E}[Z] = \frac{1}{2} \cdot \mathbb{E}[Z \mid x_1 = 0] + \frac{1}{2} \cdot \mathbb{E}[Z \mid x_1 = 1] .$$

Thus, at least one of the expectations on the right hand side must be at least $7m/8$. We now need the fact that each of these expectations can be computed efficiently.

**Exercise 2.4** *Given access to the 3-SAT formula $\varphi$, the expectations $\mathbb{E}[Z \mid x_1 = 0]$ and $\mathbb{E}[Z \mid x_1 = 1]$ can both be computed in time $O(m)$ where $m$ is the number of clauses. Actually, it is also possible to do this in time $O(t)$ if $x_1$ appears in only $t$ clauses and we are given the list of these clauses.*

Using the above, we can find a value $b_1 \in \{0,1\}$ such that

$$\mathbb{E}[Z \mid x_1 = b_1] \geq \frac{7m}{8} .$$

Continuing similarly by induction, we can find $b_1, \ldots, b_n$ such that

$$\mathbb{E}\left[Z \mid x_1 = b_1, \ldots, x_n = b_n\right] \;\geq\; \frac{7m}{8}\,.$$

Since $Z$ is fixed given the values of all the variables, we get that the assignment $(b_1, \ldots, b_n)$ satisfies at least $7m/8$ clauses.

### 2.2 Independent Sets

Let us consider one more application of the *Probabilistic Method*, which is a powerful tool show the existence of objects with certain properties without necessarily constructing them. In the previous lecture we used probabilistic reasoning to show that there exists an assignment to a 3-SAT formula with $m$ clauses satisfying $7m/8$ clauses, and then also gave an algorithm to find such an assignment. We will now use the method to show the existence of large independent sets in graphs.

Consider a graph $G = (V, E)$. A set $S \subseteq V$ is said to be an independent set if no edge lies completely within the set $S$. That is, $\forall e = \{i, j\}$, either $i \notin S$ or $j \notin S$. We are interested in finding a large independent set.

Let $N(i)$ denote the set of all neighbors of $i$ i.e., $N(i) = \{j \mid \{i,j\}\} \in E$ and let $\deg(i) = |N(i)|$. Let us first consider a weaker statement which can be proved without any probabilistic reasoning at all.

**Proposition 2.5** *Let $G = (V, E)$ be a graph with $n$ vertices and let $d$ be such that $deg(i) \leq d$ for all $i \in [n]$. Then there exists an independent set $S$ of size $|S| \geq \frac{n}{d+1}$.*

**Proof:** Start with $S = \emptyset$ and consider the vertices of the graph in the order $1, \ldots, n$. When considering vertex $i$, if none of the neighbors of $i$ (vertices in $N(i)$)) are already included in $S$, then include $i$ in $S$. At any step in this process, including a vertex in $S$ removes at most $d$ vertices from being included later. Since at the end, we finish processing all the $n$ vertices, we must have $|S| \geq \frac{n}{d+1}$. ■

The above bound is good in some cases, but the degrees of vertices in the graph might vary a lot and in particular asking for a uniform bound $d$ which holds for all vertices might be too lossy (consider a "star" graph with one vertex connected to $n-1$ others, and no other edges). The following result gives a much better bound.

**Theorem 2.6** *Let $G = (V, E)$ be a graph with $n$ vertices. Then there exists an independent set $S$ such that*

$$|S| \;\geq\; \sum_{i=1}^{n} \frac{1}{deg(i) + 1} \;\geq\; \frac{n}{\max_i\{deg(i)\} + 1}\,.$$

The main trick in such kind of problems is to set up the right kind of probabilistic experiment, the analysis is usually quite easy. In this question, we can't do everything independently unlike in some previous questions. Suppose that we do - and hence pursue the following idea: Put each $v_i$ in $S$ with probability $p$. We can't guarantee that we would not pick up both the endpoints of an edge to keep in $S$. However, this idea can also be made to work and is very useful in some settings. For now, we will prove the theorem using the observation that we can run the greedy algorithm starting with a *random* ordering of the vertices, instead of the fixed ordering $1, \ldots, n$. If we have an example where we have a single high-degree vertex surrounded by low-degree vertices, then in a random ordering we are much more likely to process one of the low-degree neighbors first (which are all good for the analysis).

**Proof:** Pick a random permutation $\pi$ of the vertices $\{1, 2, \ldots n\}$. We define the set $S$ as the set of all vertices which appear before all their neighbors in the ordering given by the permutation $\pi$.

$$S = \{i \mid \pi(i) < \pi(j) \ \ \forall j \in N(i)\} .$$

This is clearly an independent set since if $i \in S$, then for all $j \in N(i)$, we have $\pi(j) > \pi(i)$ and hence $j \notin S$. We now analyze the size of this independent set. We have $|S| = \sum_i X_i$, where

$$X_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$$

Thus, $\mathbb{E}\left[|S|\right] = \sum_i \mathbb{E}\left[X_i\right]$. To compute $\mathbb{E}\left[X_i\right]$, we notice that a random permutation of $[n]$ also induces a random ordering of the set $\{i\} \cup N(i)$. The probability that $i$ appears before any of its neighbors in the ordering is $1/(\deg(i) + 1)$. Thus,

$$\mathbb{E}\left[X_i\right] = \frac{1}{\deg(i) + 1} ,$$

which gives

$$\mathbb{E}\left[|S|\right] \ = \ \sum_{i=1}^{n} \frac{1}{\deg(i) + 1} ,$$

and hence there must exist an independent set $S$ with the above size. ■

## 3 Tail Inequalities

We will develop some inequalities which let us bound the probability of a random variable taking a value very far from its expectation.

### 3.1 Markov's Inequality

This is the most basic inequality we will use. This is useful if the only thing we know about a random variable is its expectation. It will also be useful to derive other inequalities later.

**Lemma 3.1 (Markov's Inequality)** *Let $Z$ be non-negative variable. Then,*

$$\mathbb{P}\left[Z \geq t\right] \;\leq\; \frac{\mathbb{E}\left[Z\right]}{t}. \tag{1}$$

**Proof:** We start by considering the event $E \equiv \{Z \geq t\}$. We can then write,

$$\mathbb{E}\left[Z\right] \;=\; \mathbb{P}\left[E\right] \cdot \mathbb{E}\left[Z \mid E\right] + \mathbb{P}\left[E^{c}\right] \cdot \mathbb{E}\left[Z \mid E^{c}\right].$$

Using non-negativity of $Z$, we get

$$\mathbb{E}\left[Z\right] \;\geq\; \mathbb{P}\left[E\right] \cdot \mathbb{E}\left[Z \mid E\right] \;\geq\; \mathbb{P}\left[E\right] \cdot t \;=\; \mathbb{P}\left[Z \geq t\right] \cdot t\,,$$

which completes the proof. ■

### 3.2 Chebyshev's Inequality

The variance of a random variable $X$ is defined as

$$\mathsf{Var}\left[X\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2$$

Also, for two random variables $X$ and $Y$, we define the covariance as

$$\mathsf{Cov}\left[X, Y\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right] = \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right] \cdot \mathbb{E}\left[Y\right].$$

**Lemma 3.2 (Chebyshev's inequality)** *Let $Z$ be a random variable and let $\mu = \mathbb{E}\left[Z\right]$. Then,*

$$\mathbb{P}\left[|Z - \mu| \geq t\right] \;\leq\; \frac{\mathsf{Var}\left[Z\right]}{t^2} \;=\; \frac{\mathbb{E}\left[(Z - \mu)^2\right]}{t^2}. \tag{2}$$

**Proof:** Consider the non-negative random variable $(Z - \mu)^2$. Applying Markov's inequality we have

$$\mathbb{P}\left[|Z - \mu| \geq t\right] \;=\; \mathbb{P}\left[(Z - \mu)^2 \geq t^2\right] \;\leq\; \frac{\mathbb{E}\left[(Z - \mu)^2\right]}{t^2}.$$

■

### 3.3 Coin tosses revisited

An unbiased coin is tossed $n$ times. Probability that head shows up in each toss is $\frac{1}{2}$. Let $Z$ be a random variable for the number of heads that have showed up after $n$ tosses. We also have random variables $X$ for $i^{th}$ coin toss, where $X_i = 1$ if head shows up in $i^{th}$ toss and 0 otherwise.

So we have

$$Z = \sum_{i=1}^{n} X_i \quad \text{and} \quad \mathbb{E}[Z] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2}.$$

Let us now compare the kind of bounds we get using Markov's and Chebyshev's inequalities.

**Application of Markov's inequality** . Using Markov's inequality we have,

$$\mathbb{P}\left[Z \geq \frac{3n}{4}\right] \leq \frac{\mathbb{E}[Z]}{(3n/4)} \;\Rightarrow\; \mathbb{P}\left[Z \geq \frac{3n}{4}\right] \leq \frac{2}{3} \;\Rightarrow\; \mathbb{P}\left[Z - \frac{n}{2} \geq \frac{n}{4}\right] \leq \frac{2}{3}.$$

**Application of Chebyshev's inequality** . We will show that Chebyshev's inequality gives a stronger bound on probability. Since $Z$ is a Binomial random variable, we have that

$$\mathsf{Var}[Z] \;=\; n \cdot \frac{1}{2} \cdot \left(1 - \frac{1}{2}\right) \;=\; \frac{n}{4}.$$

Applying Chebyshev's inequality we have,

$$\mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq t\right] \;\leq\; \frac{n}{4t^2}.$$

Setting $t = n/4$ and $t = \sqrt{n}$, gives the following bounds

$$\mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq \frac{n}{4}\right] \;\leq\; \frac{4}{n} \quad \text{and} \quad \mathbb{P}\left[\left|Z - \frac{n}{2}\right| \geq \sqrt{n}\right] \;\leq\; \frac{1}{4}$$

Thus, Chebyshev's inequality gives a much stronger bound on a deviation of $n/4$ from the mean, and can also bound the probability of deviations as small as $\sqrt{n}$. In particular, it gives a non-trivial bound whenever the deviation is larger than $\sqrt{\mathsf{Var}[Z]}$, a quantity which is referred to as the *standard deviation* of the random variable $Z$.

## References

[AJMR19] Vikraman Arvind, Pushkar S. Joglekar, Partha Mukhopadhyay, and S. Raja, *Randomized polynomial-time identity testing for noncommutative circuits*, Theory of Computing **15** (2019), no. 7, 1–36. 1

[AS08] Noga Alon and Joel Spencer, *The probabilistic method*, Wiley-Interscience Series, 2008. 3