

Lecture 12: November 13, 2018

Lecturer: Madhur Tulsiani

1 Randomized polynomial identity testing

We will use our knowledge of conditional probability to prove the following lemma, which gives an algorithm for testing if a polynomial f in n variables x_1, \dots, x_n over a field \mathbb{F} is identically zero.

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

$$f(x_1, x_2, \dots, x_n) = \sum c_{i_1 i_2 \dots i_n} \cdot x_1^{i_1} \cdot x_2^{i_2} \dots x_n^{i_n}$$

$$\text{s.t., } i_1 + i_2 + \dots + i_n \leq d$$

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

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

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

Induction Step

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

$$f(x_1, x_2, \dots, x_n) = x_n^k \cdot g(x_1, \dots, x_{n-1}) + h(x_1, x_2, \dots, x_n)$$

where, k is largest degree of x_n . Thus we have $0 < k \leq d$ (if $k = 0$ then we are already done). We also have that $\deg(g(x_1, \dots, x_{n-1})) \leq d - k$.

Now let us define two events.

$$E \equiv \{f(x_1, x_2, \dots, x_n) = 0\} \quad \text{and} \quad F \equiv \{g(x_1, \dots, x_{n-1}) = 0\}$$

We can then write,

$$\mathbb{P}[E] = \mathbb{P}[F] \cdot \mathbb{P}[E|F] + \mathbb{P}[F^c] \cdot \mathbb{P}[E|F^c].$$

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

$$\mathbb{P}[F] = \mathbb{P}[g(x_1, \dots, x_{n-1}) = 0] \leq \frac{d-k}{|S|}.$$

Also, fixing the values of $x_1 = a_1, \dots, x_{n-1} = a_{n-1}$ such that $g(a_1, \dots, a_{n-1}) \neq 0$, $f(x_1, a_2, \dots, a_n)$ is a degree- k polynomial in x_n . Thus, using the base case, we get that

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

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

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

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

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

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

Exercise 1.2 G has a perfect matching if and only if $\text{Det}(A) \neq 0$.

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 $\text{Det}(A) \neq 0$, we will detect it with probability at least $1/2$.

The randomized algorithm by Schwartz-Zippel 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 Random Variables and Expectations

We defined expectations in the previous lecture. Check the following facts about expectations:

Exercise 2.1 For any two random variables $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ and $c_1, c_2 \in \mathbb{R}$, we have

$$\mathbb{E}[c_1 \cdot X_1 + c_2 \cdot X_2] = c_1 \cdot \mathbb{E}[X_1] + c_2 \cdot \mathbb{E}[X_2].$$

Thus, expectation defines a linear transformation on the space of real-valued random variables (for a given Ω).

Exercise 2.2 Let $X, Y : \Omega \rightarrow \mathbb{R}$ be two independent random variables. Then show that

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Also, show that the above is not a sufficient condition for independence: give an example of two random variables X and Y such that $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ but X and Y are not independent.

We will not see some very useful random variables.

Bernoulli random variables A Bernoulli(p) random variable X is defined as taking the value 1 with probability p and the value 0 with probability $1 - p$. We can write this as $\mathbb{P}[X = x] = p^x(1 - p)^{1-x}$. One may intuitively think of a Bernoulli random variable as the indicator function of “heads” in an outcome space $\Omega = \{\text{tails}, \text{heads}\}$ of a biased coin toss. Alternatively, we simply take the outcome space to be $\Omega = \{0, 1\}$. More generally, indicator functions of events are Bernoulli random variables.

Finite Bernoulli i.i.d. sequence We can also think of a sequence of coin tosses, with

$$X_i = \begin{cases} 1 & \text{if toss } i \text{ is heads} \\ 0 & \text{if toss } i \text{ is tails} \end{cases} .$$

being n Bernoulli random variables in the probability space $\Omega_n = \{0, 1\}^n$, i.e., $X_i(\omega) = \omega_i$. Define the product probability measure on this finite space using:

$$\mu_n(\omega) = \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i} .$$

Note that if $p = \frac{1}{2}$, we have $\mu_n(\omega) = \frac{1}{2^n}$, i.e., \mathbb{P}_n is the uniform distribution over the outcome space, as all outcomes are equally likely.

Exercise 2.3 For the outcome space defined above, verify that:

- For any fixed i , X_i is indeed a Bernoulli(p) random variable, and
- If $I \subset [n]$ and $J \subset [n]$ are disjoint, then any function of X_I and any function of X_J are independent random variables.

As noted in the previous lecture, when the latter point holds, we simply say that X_1, \dots, X_n are independent. Furthermore since all the X_i have the same distribution, we call the sequence i.i.d., meaning independent and identically distributed.

Binomial random variables Let Z_n be a random variable counting the number of heads associated with n independent biased coin tosses. We can model this in Ω_n above as $Z_n = \sum X_i$.

Let us calculate the expectation of Z . By linearity we have $\mathbb{E}[Z_n] = \sum \mathbb{E}[X_i]$. Since $Z_n = \sum X_i$, we have, $\mathbb{E}[Z_n] = \sum \mathbb{E}[X_i]$. Now,

$$\begin{aligned} \mathbb{E}[X_i] &= 1 \cdot \mathbb{P}[X_i = 1] + 0 \cdot \mathbb{P}[X_i = 0] \\ &= \mathbb{P}[X_i = 1] = p \end{aligned}$$

Hence $\mathbb{E}[Z_n] = np$. Note that we did not use independence in the above calculations. We just needed that for each i , $\mathbb{E}[X_i] = p$.

We do need independence, and namely the product probability measure, to calculate $\mathbb{P}(Z_n = k)$ for $k \in [n]$ (this is often called the probability mass function. First note that the shorthand $(Z_n = k)$ simply means $\{\omega \in \Omega : Z_n(\omega) = k\}$. Since all ω that have the same number (in this case k) of 1's have the same probability, we simply need to count how many such ω 's there are, and multiply by this individual probability.

Exercise 2.4 Verify that $\mathbb{P}_n(Z_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Z is called a Binomial(n, p) random variable.

Infinite Bernoulli i.i.d. sequence and Geometric random variables We would like to generalize the Bernoulli sequence probability space to an infinite sequence. We would like to choose $\Omega = \{0, 1\}^{\mathbb{N}}$ as our outcome space, but this is not a countable set. We will come back to the issue of properly defining the probability space with this uncountable Ω .

For now, if we still consider the mental experiment of infinite i.i.d. Bernoulli(p) sequence of random variables X_1, X_2, \dots , which we interpret once more as coin tosses. We define Y be the number of tosses till the first heads. If we are just interested in Y (the first heads rather than all outcomes of all tosses), we can take Ω to be \mathbb{N} .

Exercise 2.5 *Although we cannot define a countable probability space for the infinite i.i.d. Bernoulli sequence, show that if we just want define a space for Y , we can take $\Omega = \mathbb{N}$ and $\mathbb{P}(i) = (1 - p)^{i-1} \cdot p$ for $i \geq 1$.*

Y is known as a Geometric(p) random variable.

Let us calculate $\mathbb{E}[Y]$, in a somewhat creative way. Let E be the event that the first toss is heads. Then by total expectation we have,

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[Y|E] \cdot \mathbb{P}[E] + \mathbb{E}[Y|E^c] \cdot \mathbb{P}[E^c] \\ &= 1 \cdot \mathbb{P}[E] + (1 + \mathbb{E}[Y]) \cdot (1 - p)\end{aligned}$$

Thus we have, $\mathbb{E}[Y] = \frac{1}{p}$. The main observation that we used here is that, thanks to independence, when the first toss is *not* heads, then the problem resets (with the hindsight of one consumed toss).

3 Coupon Collection

Consider the following problem: There are n kinds of items/coupons and at each time step we get one coupon chosen to be from one of the n types at random. All types are equally likely at each step and the choices at different time steps are independent. We define a random variable, T which is the time when we first have all the n types of coupons. Find $\mathbb{E}[T]$.

We can make the following claim:

$$T = \sum_{i=1}^n X_i,$$

where X_i is the time to get from the $i - 1$ to the i types of coupons. Thus we have,

$$\mathbb{E}[T] = \sum_i \mathbb{E}[X_i]$$

Note that X_i is a geometric random variable with parameter $\frac{n-i+1}{n}$, since if we have $i - 1$ type of coupons, X_i represents the time till we receive a coupon belonging to any one of the remaining $n - i + 1$ types. Thus,

$$\mathbb{E}[X_i] = \frac{n}{n - i + 1}.$$

Therefore,

$$\mathbb{E}[T] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} = n \cdot H(n)$$

where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is the n^{th} harmonic number. It is known (see Wikipedia for example) that $H_n = \ln n + \Theta(1)$. Thus, we have that $\mathbb{E}[T] = n \ln n + \Theta(n)$.