

Lecture 11: November 3, 2016

Lecturer: Madhur Tulsiani

1 Basics of probability

We recall very briefly the basics of probability and random variables. For a much better and detailed introduction, please see the lecture notes by Terry Tao, linked from the course homepage.

1.1 The finite case

Let Ω be a finite set. Let $\mu : \Omega \rightarrow [0, 1]$ be a function such that

$$\sum_{\omega \in \Omega} \mu(\omega) = 1.$$

We often refer to Ω as a sample space and the function μ as a probability distribution on this space. An event can be thought of as a subset of outcomes i.e., any $E \subseteq \Omega$ defines an event, and we define its probability as

$$\mathbb{P}[E] = \sum_{\omega \in E} \mu(\omega).$$

Conditioning on an event E is equivalent to restricting the probability space to the set E . We then consider the conditional probability measure μ_E defined as

$$\mu_E(\omega) = \begin{cases} \frac{\mu(\omega)}{\mathbb{P}[E]} & \text{if } \omega \in E \\ 0 & \text{otherwise} \end{cases}.$$

Thus, one can define the conditional probability of an event F as

$$\mathbb{P}[F | E] = \sum_{\omega \in F} \mu_E(\omega) = \sum_{\omega \in E \cap F} \frac{\mu(\omega)}{\mathbb{P}[E]} = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[E]}.$$

A real-valued random variable over Ω is any function $X : \Omega \rightarrow \mathbb{R}$. We define

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \mu(\omega) \cdot X(\omega).$$

For a random variable X and an event E , we define $\mathbb{E}[X | E]$ as $\sum_{\omega} \mu_E(\omega) \cdot X(\omega)$, with μ_E as above. Verify the following identities.

Proposition 1.1 *Let Ω be a finite probability space with a corresponding function μ , let $E, F \subseteq \Omega$ be events, and let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then*

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

Everything defined above can also be extended to countable spaces but we need to be careful about the convergence of the above summations.

2 Randomized polynomial identity testing

We use the above 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 2.1 (Schwartz-Zippel lemma) *Let $f(x_1, x_2, \dots, x_n)$ be a non-zero polynomial of degree $d \geq 0$, i.e.,*

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

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:

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

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 $\deg(g(x_2, \dots, x_n)) \leq d - k$.

Now let us define two events.

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

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

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

Also, fixing the values of $x_2 = a_2, \dots, x_n = a_n$ such that $g(a_2, \dots, a_n) \neq 0$, $f(x_1, a_2, \dots, a_n)$ 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. ■

2.1 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 2.2 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.

3 Computing expectations

We will demonstrate the computation of expectations with some examples.

Let Z be a random variable of number of heads associated with n tosses of coins. Let X_i be the random variable associated with the i^{th} toss, defined as

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

Thus $\Omega = \{0, 1\}^n$. Let us assume that the coin tosses are independent of each other (though, as you will see, we will not need this assumption here).

Example 3.1 *With the assumption that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = 0] = 1/2$, we want to compute $\mathbb{E}[Z]$.*

Since $Z = \sum X_i$, we have, $\mathbb{E}[Z] = \sum \mathbb{E}[X_i]$. Now,

$$\begin{aligned} \mathbb{E}[X_i] &= 1 \cdot \mathbb{P}[X_i = 1] + 0 \cdot \mathbb{P}[X_i = 0] \\ &= \mathbb{P}[X_i = 1] = 1/2 \end{aligned}$$

Hence $\mathbb{E}[Z] = n/2$.

Note that if a random variable, X_e takes a binary value if an event, e occurs or not, then the expected value of X_e is $\mathbb{P}[e]$.

Example 3.2 *Instead of the uniform probability of heads and tails, if $\mathbb{P}[X_i = 1] = p$ and $\mathbb{P}[X_i = 0] = 1 - p$, then $\mathbb{E}[Z] = n \cdot p$.*

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