TTIC/CMSC 31150 Mathematical Toolkit **Spring 2013**

Lecture 1: April 1, 2013

Madhur Tulsiani Scribe: Kaustav Kundu

The first topic that we are going to delve into is Discrete Probability. The following are the reference books for this topic:

- **Probability and Computing: Randomized Algorithms and Probabilistic Analysis** by *Michael Mitzenmacher, Eli Upfal*
- **The Probabilistic Method** by *Noga Alon, Joel H. Spencer*

In this part of the course, we will develop some basic tools to analyze randomized algorithms and random phenomena. Randomness is a convenient tool in the design of algorithms as it can often conceptually simplify the algorithm. Also, as we will see, for some problems we do have efficient randomized algorithms but do not yet have efficient deterministic algorithms.

The following is an example of a problem and a randomized algorithm, which we hope to be able to analyze by the end of this part of the course.

Example 0.1 (Balanced Allocation) *Given n servers and an incoming stream of requests, we would like to assign requests in a way to minimize load on the most loaded server.*

Let us first discuss about the different strategies to solve this problem. We consider the performance of each strategy when we have serviced m requests (think of m as $O(n)$).

a) At every step, use the least loaded server. The maximum load of any server will be $O\left(\frac{m}{n}\right)$. This is not a very good solution as we are going to use either $O(n)$ space (to store all the loads and find the minimum when a request comes in) or $O(n)$ time complexity (to check all loads when a request comes in).

b) At every step, use a random server. We can (and we will later) prove that the maximum load of any server is $O\left(\frac{m}{n}\log\ n\right)$. Note we only lose a $\log n$ factor in the maximum load and the time and space complexity of each step is only $O(1)$. $\log\ n$.

c) At every step, use the least loaded out of two random servers. The time and space complexity still remains $O(1)$, but the maximum load now drops to $O\left(\frac{m}{n}\text{loglog}\ n\right)$. Several applications of this idea were explored by Mitzenmacher in his PhD Thesis [Mitz96].

1 Introduction to Discrete Probability

Let us first consider a random experiment with only a finite number of outcomes. This gives a *probability space* Ω with each outcome $\omega \in \Omega$ represented as a cell in the grid below.

ω_1	ω_2		
	$\cdots$		
			ω_N

If the number of outcomes is finite, then simply assigning a non-negative probability to each outcome (which we call $\mathbb{P}[\omega]$) with the following properties, defines a valid probability space.

- $\mathbb{P}[\omega] \geq 0, \forall \omega \in \Omega$
- $\sum\limits_{\omega \in \Omega} \mathbb{P}[\omega] = 1$

An *event* is a subset of outcomes. Each outcome is called a *simple event* or *basic event*. In an infinite probability space it might not be possible to assign a meaningful probability to each outcome or basic event (we will see examples later). In that case we will choose a collection of events $\mathcal{E}$ which is closed under union, intersection and complementation, such that the null event $\emptyset \in \mathcal{E}$.

- If two events E_1 and E_2 are independent, we have $\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\,\mathbb{P}[E_2]$ (actually this property defines independence).
- For any valid events, E_1 and E_2, we have, $\mathbb{P}[E_1 \cup E_2] \leq \mathbb{P}[E_1] + \mathbb{P}[E_2]$.

For any two valid events, E_1 and E_2, we define the probability of E_1 *conditioned on* E_2 as

$$\mathbb{P}[E_1|E_2] = \frac{\mathbb{P}[E_1 \cap E_2]}{\mathbb{P}[E_2]}$$

The following is an easy observation ($\neg E_2$ denotes the complement of E):

$$\mathbb{P}[E_1] = \mathbb{P}[E_2] \cdot \mathbb{P}[E_1|E_2] + \mathbb{P}[\neg E_2] \cdot \mathbb{P}[E_1|\neg E_2]$$

1.1 An Application: Randomized Identity-Testing

We use the above to prove the following lemma, which gives an algorithm for testing if a polynomial f in n variables $x_1, \ldots, x_n$ over a field $\mathbb{F}$ is identically zero.

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

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

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 randomly from S independently. Then,

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

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

<u>*Base Case*: n = 1</u>
A non zero polynomial, $f(x_1)$ will have at most d roots to the equation $f(x_1) = 0$. Hence, $\mathbb{P}[f(x_1) = 0] \leq \frac{d}{|S|}$. The inequality condition is because the number of roots might be less than d, or all of them might not be in S.

<u>*Induction Step*</u>
Let us say that this hold true any polynomial $g(x_2, \ldots, x_n)$ 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 the maximum power of x_1. Thus we have $0 < k \leq d$ (If $k = 0$ then we are already done). We also have the $\deg(g(x_2, \ldots, x_n)) \leq d - k$.

Now let us define two events.
$E_1 : f(x_1, x_2, \ldots, x_n) = 0$
$E_2 : g(x_2, \ldots, x_n) = 0$

We can hence write,

$$\mathbb{P}[E_1] \;=\; \mathbb{P}[E_2] \cdot \mathbb{P}[E_1|E_2] + \mathbb{P}[\neg E_2] \cdot \mathbb{P}[E_1|\neg E_2] \;.$$

Without loss of generality, let us choose variables $x_2, \ldots, x_n$ first and then choose the value of x_1. Then, we have,

$$\begin{array}{rcll}
\mathbb{P}[E_2] & \leq & \dfrac{d-k}{|S|} & \because\ \deg(g(x_2, \ldots, x_n)) \leq d - k \\
\mathbb{P}[E_1|E_2] & \leq & 1 & \because\ \text{from the definition of probability} \\
prob\neg E_2 & \leq & 1 & \because\ \text{from the definition of probability} \\
\mathbb{P}[E_1|\neg E_2] & \leq & \dfrac{k}{|S|} & \because\ \text{fixing } x_2, \ldots, x_n \text{ s.t. } g(x_2, \ldots, x_n) \\
 & & & \quad \text{we get a degree k non-zero polynomial in } x_1
\end{array}$$

Hence $\mathbb{P}[f(x_1, x_2, \ldots, x_n) = 0] \leq \dfrac{d}{|S|}$. ■

Consider the following example which applied the Schwartz-Zippel lemma for testing if a bi-partite graph has a perfect matching.

Example 1.2 (Bi-Partite Matching) *Given a bi-partite graph, $G = (U, V, E)$ with $|U| = |V| = n$, check whether the graph has a perfect matching.*

Let us define a *Tutte matrix*, $\mathbf{A}_{n \times n}$,

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

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

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

where π defines the permutation from rows to columns.

Observe that G has a perfect matching iff $\text{Det}(\mathbf{A}) \neq 0$. If there is no perfect matching then the term corresponding to each permutation involves a zero entry of A and hence $\text{Det}(\mathbf{A}) = 0$. On the other hand, if there is perfect matching, then it corresponds to a non-zero term for a permutation, say π_0, in the above summation. Also, no other permutation will involve all the variables $x_{i,\pi_0(i)}$, and hence this term will not cancel. Thus, $\text{Det}(\mathbf{A})$ is a non-zero polynomial when there exists a perfect matching.

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 in $O\left(n^3\right)$. The degree of the polynomial above is n. Thus, if we assign all variables randomly from a set of $2n$ real values, if $\text{Det}(\mathbf{A}) = 0$, 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.

1.2 Random Variables and Expectation

A random variable can be defined as a function $X : \Omega \to \mathbb{R}$. Note that a random variable is just a fixed function. The randomness is simply in outcome ω.

For a random variable, its *expectation* is defined as,

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

For e.g., if $X : \Omega \to \mathbb{N}$,

$$\mathbb{E}[X] = \sum_{i} \mathbb{P}[X = i] \cdot i$$

The following properties of expectation will be quite useful.

- **Constants** : If c is a constant, $\mathbb{E}[c] = c$
- **Linearity** : If a, b, c are constants and X, Y are random variables, then $\mathbb{E}[aX + bY + c] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y] + c$

1.2.1 Computation of Expected Values

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

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

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 1.3 *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 $\mathbb{E}\mathbf{X} = n/2$.

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 1.4 *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 what is $\mathbb{E}[Z]$?*

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

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

Example 1.5 *What is $\mathbb{E}[\#occurrences\ of\ pattern\ 100]$ (assuming uniform probability)? Here* 1 *denotes a heads and* 0 *denotes tails.*

Let,

$$X_i = \begin{cases} 1 & \text{if } (i, i+1, i+2) \equiv 100 \\ 0 & \text{else} \end{cases}$$

Again, denoting the number of occurrences by Z, we have

$$\mathbb{E}[Z] = \sum_{i=1}^{n-2} \mathbb{E}X_i = \frac{n-2}{8}$$

Similarly, $\mathbb{E}[\#\text{occurrences of pattern } 101] = \dfrac{n-2}{8}$

For the next example, we consider an *infinite* sequence of independent coin tosses, with $\mathbb{P}[heads] = p$ for each coin.

Example 1.6 *Given, that $\mathbb{P}[heads] = p$, what is $\mathbb{E}[\#tosses\ till\ the\ first\ heads]$?*

We define Z as the number of tosses till the first heads. We know that $\mathbb{P}[Z = i] = (1-p)^{i-1}p$. Thus,

$$\begin{aligned}
\mathbb{E}\mathbf{X} &= \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \cdot p \\
&= p \cdot \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \\
&= p \cdot \frac{1}{(1-(1-p))^2} \\
&= \frac{1}{p}
\end{aligned}$$

Here, we used the fact that for $|x| \leq 1$, $\sum_{i=}^{\infty} i \cdot x^{i-1} = \frac{1}{(1-x)^2}$.

We can also compute this expectation in another way. Let E be the event that the first toss is heads. Then we have,

$$\begin{aligned}
\mathbb{E}[\mathbf{X}] &= \mathbb{E}[Z|E] \cdot \mathbb{P}[E] + \mathbb{E}[Z|\neg E] \cdot \mathbb{P}[\neg E] \\
&= 1 \cdot \mathbb{P}[E] + (1 + \mathbb{E}[\mathbf{X}]) \cdot (1-p)
\end{aligned}$$

Thus we have, $\mathbb{E}[Z] = \dfrac{1}{p}$.

The above is known as a *geometric random variable* with parameter p.

Exercise 1.7 *Consider an infinite sequence of independent tosses of a fair coin. Define the following random variables:*

$$\begin{aligned}
Z_1 &= \textit{Number of tosses after which the pattern HTT first appears} \\
Z_2 &= \textit{Number of tosses after which the pattern HTH first appears}
\end{aligned}$$

Compute $\mathbb{E}[Z_1]$ and $\mathbb{E}[Z_2]$ and verify that they are not *equal. Why is one patten more likely to occur first even though they are both occur an equal number of times (in expectation) in a given number of tosses?*

References

[Mitz96] M. Mitzenmacher, "The Power of Two Choices in Randomized Load Balancing", PhD thesis, University of California Berkeley, 1996.