Mathematical Toolkit **Autumn 2019**

Lecture 14: November 14, 2019

Lecturer: Madhur Tulsiani

1 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 $\mathbb{E}[T] = n \ln n + \Theta(n)$.

2 A randomized algorithm for Max 3-SAT

Recall that a 3-SAT formula φ 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 φ with m clauses, one can 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 2.1 *Given access to the 3-SAT formula φ, 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

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

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.