2. Probabilistic method: First and second moments

2.1 Introduction

In the last lecture we saw how randomness is used in the design of efficient algorithms and got a brief glimpse about the application of the probabilist method. In this lecture, we will continue with the probabilistic method and introduce the first moment method.

2.2 MAX-CUT using the probabilistic method

Let $G(V, E)$ be a graph on n vertices and m edges. A *cut* of the graph is a partition of $V = V_1 \cup V_2$. The *size* of the cut is the number of edges that have one end point in V_1 and the other in V_2.

Problem 2.1 (MAX-CUT)**.** *Given a graph $G(V, E)$ find the cut with the maximum size.*

This problem is known to be NP-hard. In the next lecture we will see how to use the probabilistic method to show that a cut with a large size always exists.

Lemma 2.2. *A graph $G(V, E)$ on n vertices with m edges has a cut of size at least $m/2$.*

Proof. For each vertex $u \in V$, with probability $1/2$, u goes to V_1 and with probability $1/2$, u goes to V_2. For each edge $(u, v) \in E$, let X_{uv} denote the random variable that is 1 is (u, v) is a cut-edge, and 0 otherwise. Therefore, $\Pr[X_{uv} = 1] = 1/2$. Let X be the random variable denoting the sum $\sum_{(u,v)\in E} X_{uv}$. By linearity of expectation $E[X] = \sum_{(u,v)\in E} E[X_{uv}]$. But, $E[X_{uv}] = 1/2$. Therefore $E[X] = m/2$. If the expected cut-size is at least $m/2$, then G contains at least one cut of size $m/2$. □

A few things to keep in mind about the proof.

- It does not tell you how to find a cut of size $m/2$.
- Unlike in the example of PIT, this does not tell you that the random cut that you created will have size at least $m/2$ with high probability. (Can you convert this to such an algorithm?)
- Introduces the first moment method. If a random variable X has a finite expectation, then it takes some value $\geq E[X]$ with positive probability and takes some value $\leq E[X]$ with some positive probability.
- **Linearity of expectation** - Let X be the sum of the random variables $X_1, X_2, \ldots, X_n$. Then $E[X] = \sum_{i=1}^{n} E[X_i]$. This assumes nothing about the random variables. (Can you prove this statement?)

2.3 Markov's inequality

Now we move onto an inequality that shows that the probability of a variable being far from the mean is not high. The exact statement of the inequality is as follows.

Lemma 2.3 (Markov's inequality)**.** *Let X be a non-negative random variable with finite expectation. Then,* $\Pr[X \geq a] \leq E[X]/a$*.*

Proof. Let Ω be the underlying sample space for the random variable X, i.e. $X : \Omega \to \mathbb{R}$. Now let $\Omega' \subset \Omega$ be the set such that for $\omega \in \Omega'$, $X(\omega) \geq a$. Now we can write $E[X] = \sum_{\omega \in \Omega'} \Pr[X = \omega] X(\omega) + \sum_{\omega \in \Omega \setminus \Omega'} \Pr[X = \omega] X(\omega) \geq \sum_{\omega \in \Omega'} \Pr[X = \omega] \cdot a$, and the lemma follows. □

Exercise 2.4. *Let X be a non-negative integral random variable with finite expectation. Show that* $\Pr[X \neq 0] \leq E[X]$*. Can you use this fact to prove the union bound?*

2.4 Chebyshev's inequality

Now we look at the second moment, or the variance of a random variable. While the expectation of a random variable may not convey the actual distribution of the values, variance gives a measure of how much the random variable is distributed close to the expectation. The variance of a random variable X is given by $\mathrm{Var}[X] = E[(X - E[X])^2]$.

Lemma 2.5 (Chebyshev's inequality)**.** *Let X be a random variable with finite expectation $E[X]$ and non-zero variance* $\mathrm{Var}[X]$*. Then for any positive k, we have*

$$\Pr[|X - E[X]| \geq k] \leq \frac{\mathrm{Var}[X]}{k^2}$$

Proof. $\mathrm{Var}[X] = E[(X - E[X])^2] \geq k^2 \Pr[|X - E[X]| \geq k]$. □

Exercise 2.6. *Consider the random variable X in Lemma 2.2. What is the variance of the random variable X? What does Chebyshev's inequality tell us about the deviation from the expectation of the random variable X?*

2.5 Chernoff bounds

We now come to Chernoff bounds, which gives the strongest concentration around the mean. Naturally, the bound does not hold for any random variable unlike the Markov and Chebyshev inequality. So, let's state the inequality.

Lemma 2.7 (Chernoff bounds)**.** *Let $X_1, X_2, \ldots, X_n$ be* independent *random variables such that $X_i = 1$ w.p p, and $X_i = 0$ w.p $1 - p$. Let $X = \sum_{i=1}^{n} X_i$ be a random variable such that $E[X] = \mu$. Then the following inequalities hold.*

$$\Pr[X \geq (1+\delta)\mu] \leq e^{\frac{-\delta^2 \mu}{2+\delta}} \text{ for } \delta > 0$$
$$\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2} \text{ for } 0 < \delta < 1$$

Notice the asymmetry in the bounds. This results from some approximations of the log function. In the next lecture we will see a proof of this inequality and an application to error reduction of the algorithm.

Consider the following randomized algorithm $\mathcal{A}$ that takes an input $\mathcal{I}$ of length n, and outputs the current answer with probability $1/2 + 1/n$. I.e. $\Pr[\mathcal{A}(I) \text{ is correct }] \geq \frac{1}{2} + \frac{1}{n}$. We will see how to boost the accuracy to 99/100 (or even arbitrarily close to 1) using the Chernoff bounds (what happens if the accuracy is 1/2?).