3. Chernoff bound and error reduction

3.1 Introduction

In the last lecture we saw the first moment method, and Chebyshev's inequality. In today's lecture we look at the Chernoff bounds, and some applications.

3.2 Chernoff bounds

Let's recall the statement of Chernoff bounds.

Lemma 3.1 (Chernoff bounds)**.** *Let $X_1, X_2, \ldots, X_n$ be* independent *random variables such that $X_i = 1$ w.p p_i, and $X_i = 0$ w.p $1 - p_i$. Let $X = \sum_{i=1}^{n} X_i$ be a random variable such that $E[X] = \mu$. Then the following inequalities hold.*

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

Proof. Let $a \in \mathbb{R}$ be positive. For any $s > 0$, $\Pr[X \geq a] = \Pr[e^{sX} \geq e^{sa}]$. And by Markov's inequality $\Pr[X \geq a] \leq E[e^{sX}]/e^{sa}$. The function $E[e^{sX}]$ is known as the *moment generating function* (why? Just try expanding the function). The following claim is true. Now, $E[e^{sX}] = E[e^{s\sum_i X_i}] = \prod_i E[e^{sX_i}]$ (why is this true?). Now let's bound each of the $E[e^{sX_i}]$.

$$\begin{aligned} E[e^{sX_i}] &= e^s \cdot p_i + 1 \cdot (1 - p_i) \\ &= 1 + p_i(e^s - 1) \leq e^{p_i(e^s-1)} \end{aligned}$$

Therefore, $E[e^{sX}] \leq e^{(e^s-1)\sum_i p_i} = e^{\mu(e^s-1)}$[1]. Choosing $a = (1+\delta)\mu$, we get $\Pr[X \geq (1+\delta)\mu] \leq e^{\mu(e^s-1)}/e^{s\mu(1+\delta)}$. The value of s that minimizes $e^{\mu(e^s-1)}/e^{s\mu(1+\delta)}$ is $\ln(1+\delta)$ (how do you find that out?). Therefore, $\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu \leq e^{\frac{-\delta^2\mu}{2+\delta}}$ (why?). To prove the second bound, we write $\Pr[X \leq a] = \Pr[e^{-sX} \geq e^{-sa}]$ for $s > 0$, and follow the same steps. (Complete the proof as an exercise. In this case, you will have to take $s = \ln(\frac{1}{1-\delta})$ to minimize the probability and approximating this will give you the asymmetric bound that we mentioned). □

[1] You can avoid this inequality and obtain a bound that is the strongest form of the inequality. Moreover, it gives you a lot more intuition about the Chernoff bound. We won't discuss it here, but if interested you can refer the first chapter of the book "Concentration of measure for the analysis of randomized algorithms" by Panconesi and Dubhashi.

Lemma 3.2 (Hoeffding's extension)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in the interval $[0,1]$, and let $\bar{X} = \frac{1}{n}\sum_i X_i$. Then,*

$$\Pr[|\bar{X} - E[\bar{X}]| > \varepsilon] \leq 2e^{-2n\varepsilon^2}.$$

3.2.1 Application to error reduction

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 $2/3$ (or even arbitrarily close to 1) using the Chernoff bounds (what happens if the accuracy is $1/2$?).

The algorithm is simple. Repeat $\mathcal{A}$ t many times (we will fix the value of t later), and take the majority answer. Let $X_1, X_2, \ldots X_t$ denote the random variables such that X_i is 1 if at the i^{th} iteration of $\mathcal{A}$ it answers correctly. Therefore, $\Pr[X_i = 1] = \frac{1}{2} + \frac{1}{n}$. Let $X = \sum_i X_i$. Then, $E[X] = \frac{t}{2} + \frac{t}{n}$. We are interested in the probability that $X > t/2$. From Lemma 3.1, we know that $\Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$. Now, for $t = 10n^2$ and $\delta = 1/2n$ substituted in the Chernoff bound will show you that with probability at least $2/3$, the majority answer will be the right answer.

A few points to ponder here.

- How has the time complexity of the algorithm been affected?
- What about the randomness used by the algorithm?
- Can we bring about error reduction with more efficient use of randomness?

Exercise 3.3. *Suppose $\mathcal{A}$ is an algorithm for computing a function f such that for an input $\mathcal{I}$ of length n the following conditions hold:*

- *If $f(\mathcal{I}) = 1$, then $\Pr[\mathcal{A}(\mathcal{I}) = 1] = 1$.*
- *If $f(\mathcal{I}) = 0$, then $\Pr[\mathcal{A}(\mathcal{I}) = 0] \geq \frac{1}{n}$.*

Describe a new algorithm $\mathcal{A}'$ which does the following.

- *If $f(\mathcal{I}) = 1$, then $\Pr[\mathcal{A}'(\mathcal{I}) = 1] = 1$.*
- *If $f(\mathcal{I}) = 0$, then $\Pr[\mathcal{A}'(\mathcal{I}) = 0] \geq 1 - e^{-n}$.*

If $\mathcal{A}$ had a running time $t(n)$, what is the running time of your new algorithm? If $\mathcal{A}$ uses $r(n)$ random bits, how many random bits does $\mathcal{A}'$ use?