5.1 The Majority Problem and the Probabilistic Method

Recall the Majority Problem: given $n$ bits $x_1, \ldots, x_n$, we want to compute the $\text{MAJ}$, such that

$$\text{MAJ}(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if more } x_i \text{'s are 1's than 0's;} \\ 0 & \text{otherwise.} \end{cases}$$

The goal is to compute $\text{MAJ}(x_1, \ldots, x_n)$ using monotone circuits.

**Fact:** There exists an explicit linear-sized $O(\log n)$-depth circuit to compute $\text{MAJ}$. Use a binary tree where leaves are the inputs and each internal node returns the number of 1's among the leaves of its rooted subtree. Note that this approach requires addition, which is not monotone.

**Claim 5.1** There exists a monotone circuit of polynomial size and $O(\log n)$ depth which computes $\text{MAJ}$.

**Proof:** Define the $\text{MAJ}_3$ gate to take 3 inputs. (We can implement $\text{MAJ}_3$ with a depth 4, size 16 circuit using only 2-input monotone gates.)

Our circuit $C$ will involve a complete 3-ary tree of $\text{MAJ}_3$ gates, where the root is at level $D$, where $D$ is the depth of the tree. At level 1, each gate chooses 3 input bits uniformly at random (with replacement). We will say that the input bits $x_i$'s are at level 0 of this circuit.

We want to show that with probability $> 0$, this randomized (monotone, poly-sized, $O(\log n)$ depth) circuit computes $\text{MAJ}$ correctly. Then, there must exist a circuit that correctly computes $\text{MAJ}$.

Two observations:

- $D = k \log n$, for some $k$ to be chosen later. (This means that the number of gates at level 1 is $3^D = 3^k \log n$.)
- $C$ has size $S \leq 3^{D+1}$, which is polynomial in $n$, since $D = O(\log n)$

If we can show that for all inputs $x \in \{0, 1\}^n$, $\Pr(C(x) \neq \text{MAJ}(x)) \leq 2^{-(n+1)}$, then by a union bound, we can argue that $\Pr(\exists x \text{ s.t. } C(x) \neq \text{MAJ}(x)) \geq \frac{1}{2}$.

Fix our input $x \in \{0, 1\}^n$. Define $p_t$ as the probability that a gate at level $t$ outputs a 1. The inputs themselves are not probabilistic, so for $t = 0$, let this probability refer to a uniformly random choice of $x_i$. We want to show that for $k$ large enough,

1. if $\text{MAJ}(x_1, \ldots, x_n) = 1$, then $p_D \geq 1 - 2^{-(n+1)}$,
2. if $MAJ(x_1, \ldots, x_n) = 0$, then $p_D \leq 2^{-(n+1)}$.

We will only prove the first claim. The second is analogous and is left as an exercise. (It may be convenient to assume $n$ odd for this second case.) If $MAJ(x_1, \ldots, x_n) = 1$, then there is a slight bias toward 1 in the input $x_i$’s, i.e. $p_0 \geq \frac{1}{2} + \frac{1}{2n}$. Since there are 3 ways to have two inputs being 1 and one input being 0 (and similarly, there is 1 way to have three inputs as 1),

$$p_t + 1 = 3(p_t)^2(1 - p_t) + (p_t)^3$$

Then,

$$p_{t+1} - p_t = p_t(1 - p_t)(2p_t - 1)$$

We split the analysis of the circuit into 2 phases, showing that it takes $O(\log n)$ steps to amplify the probability of correctness (our bias) to $\frac{3}{4}$, and then another $O(\log n)$ steps to amplify to $1 - 2^{-(n+1)}$.

**Phase 1.** By assumption, $\frac{3}{4} \geq p_0 \geq \frac{1}{2} + \frac{1}{2n}$. Define $\epsilon_t = p_t - \frac{1}{2}$, i.e. the distance of the bias from $\frac{1}{2}$. We need to show that $\epsilon_t$ grows to $\frac{1}{4}$ exponentially so that only $O(\log n)$ steps are needed. We can show that:

$$\epsilon_{t+1} - \epsilon_t = p_{t+1} - p_t$$

$$= p_t(1 - p_t)(2p_t - 1)$$

$$= \left(\epsilon_t + \frac{1}{2}\right)\left(\frac{1}{2} - \epsilon_t\right)2\epsilon_t$$

$$= \left(\frac{1}{2} - 2\epsilon_t^2\right)\epsilon_t$$

Since $\epsilon_t \leq \frac{1}{4}$, $\epsilon_{t+1} - \epsilon_t \geq \frac{3}{8} \epsilon_t$ and $\epsilon_{t+1} \geq \frac{11}{8} \epsilon_t$. This recurrence implies that there is a $t_0 = \Theta(\log n)$ s.t. $\epsilon_{t_0} \geq \frac{1}{4}$.

**Phase 2.** By assumption, $p_{t_0} \geq \frac{3}{4}$. We will need faster growth than achieved in Phase 1. Define $\delta_t = 1 - p_t$, i.e. distance away from 1.

$$\delta_t - \delta_{t+1} = p_{t+1} - p_t$$

$$= p_t(1 - p_t)(2p_t - 1)$$

$$= (1 - \delta_t)\delta_t(1 - 2\delta_t)$$

$$= \delta_t(1 - 3\delta_t + 2\delta_t^2)$$

Then, $\delta_{t+1} = \delta_t(3\delta_t - 2\delta_t^2) \leq 3\delta_t^2$. Starting from the end of Phase 1, if $\delta_{t_0} \leq \frac{1}{4}$, then $\delta_{t_0 + t} \leq 3^{2^t - 1} \left(\frac{1}{4}\right)^{2^t} \leq \left(\frac{3}{4}\right)^{2^t}$. Since our speed up is doubly exponential, there is some $t = \Theta(\log n)$ for which $\delta_{t_0 + t} \leq 2^{-(n+1)}$. Then, $p_{t_0 + t} \geq 1 - 2^{-(n+1)}$.

**5.2 Chebyshev’s Inequality**

Markov’s inequality uses only $E(X)$ (i.e. $X$’s first moment). In this sense, Markov’s inequality is weak, since $X$ can achieve a value very far from $E(X)$. We introduce the following notion:
**Definition 5.2** For a random variable $X$, variance $\text{Var}(X)$ is $E((X - E(X))^2)$.

**Definition 5.3** For any two random variables $X$ and $Y$, the covariance $\text{cov}(X, Y)$ is $E(XY) - E(X)E(Y)$.

Note that if $X$ and $Y$ are completely independent, then $\text{cov}(X, Y) = 0$.

**Theorem 5.4** (Chebyshev’s Inequality) $\Pr(|X - E(X)| \geq \alpha) \leq \frac{\text{Var}(X)}{\alpha^2}$.

**Proof:** Let $Y = (X - E(X))^2$. $Y$ is a non-negative random variable, so by applying Markov’s Inequality,

$$\Pr(|X - E(X)| \geq \alpha) = \Pr(Y \geq \alpha^2) \leq \frac{E(Y)}{\alpha^2} = \frac{\text{Var}(X)}{\alpha^2}$$

\[\blacksquare\]

### 5.2.1 Example: Random Graphs

Let $G_{n,p}$ be the distribution over all $n$-vertex graphs where each of the possible $\binom{n}{2}$ edges appear in the graph independently with probability $p$. We say $G \sim G_{n,p}$ to mean that $G$ is a random variable (graph) taken from the distribution $G_{n,p}$.

Natural questions that one might consider include: Is $G$ expected to be connected? What is the chromatic number $\chi(G)$? Today we consider the following question: does $G$ contain a 4-clique?

As before, define a random variable $X = \text{number of 4-cliques in } G$. Define for every set $C$ of 4 vertices in $G$, an indicator variable $X_C$:

$$X_C = \begin{cases} 1 & \text{if } C \text{ is a clique;} \\ 0 & \text{otherwise.} \end{cases}$$

There are $\binom{n}{4}$ such indicator variables. Note that $X = \sum_{C \in \binom{[n]}{4}} X_C$. There are six edges in a 4-clique, and each is chosen independently, hence $E(X_C) = \Pr(X_C = 1) = p^6$. This implies, by linearity of expectation, that $E(X) = \binom{n}{4}p^6 = \Theta(n^4p^6)$. Thus

- if $p \ll n^{-2/3}$, then $E(X) \to 0$ as $n \to \infty$
- if $p \gg n^{-2/3}$, then $E(X) \to \infty$ as $n \to \infty$

These observations are not as strong as:

- if $p \ll n^{-2/3}$, then $\Pr(X > 0) \to 0$
- if $p \gg n^{-2/3}$, then $\Pr(X > 0) \to 1$

The first claim follows easily since $\Pr(X > 0) = \Pr(X \geq 1)$ which must approach 0 if $E(X) \to 0$. The second claim is implied from just the first moment, i.e. $(E(X) \to \infty)$ does not imply $(\Pr(X > 0) \to 1)$.

**Note:** local properties like this usually have “coarse thresholds”, i.e. seeing a clique does not drastically affect the probability of seeing another. Global properties (e.g. the existence of a Hamiltonian circuit) have “sharp thresholds”, i.e. if there is one Hamiltonian cycle in the graph, then the probability of there being more is very high.
Claim 5.5 If \( p \gg n^{-2/3} \), then \( \Pr(X > 0) \to 1 \).

**Proof:**

Note that \( \Var(X) = E((X - E(X))^2) \) can be rewritten as \( E(X^2) - (E(X))^2 \).

If \( X = 0 \), then \( |X - E(X)| \geq E(X) \). Therefore,

\[
Pr(X = 0) \leq Pr(|X - E(X)| \geq E(X)) \leq \frac{\Var(X)}{(E(X))^2} = \frac{E(X^2) - (E(X))^2}{(E(X))^2}
\]

We want to show that \( E(X^2) - (E(X))^2 \) is small compared to \( (E(X))^2 \).

\[
\Var(X) = E(X^2) - (E(X))^2
\]

\[
= E((\sum_C X_C)^2) - (\sum_C E(X_C))^2
\]

\[
= E(\sum_C (X_C^2) + \sum_{C \neq D} (X_C X_D)) - \sum_C (E(X_C))^2 - \sum_{C \neq D} E(X_C)E(X_D)
\]

\[
= \sum_C (E(X_C^2) - (E(X_C))^2) + \sum_{C \neq D} (E(X_C X_D) - E(X_C)E(X_D))
\]

\[
= \sum_C \Var(X_C) + \sum_{C \neq D} \cov(X_C, X_D)
\]

We want to show that the second (covariance) term is negligible compared to the first term, and then finally show that the first term is small compared to \( (E(X))^2 \).

For any \( C, D \subset V, C \neq D \), we compute \( \cov(X_C, X_D) \) in cases:

1. \( |C \cap D| \leq 1 \): No pairs of vertices are shared, so \( X_C \) and \( X_D \) are independent. \( \cov(X_C, X_D) = 0 \).

2. \( |C \cap D| = 2 \): One pair of vertices is shared, so one fewer total edge need be present. \( \cov(X_C, X_D) = E(X_C X_D) - p^{12} = p^{11} - p^{12} \leq p^{11} \). This can happen \( \binom{n}{6} \) times, so the total contribution to the sum of covariances from this case is \( \leq \binom{n}{6}p^{11} = \Theta(n^6p^{11}) \).

3. \( |C \cap D| = 3 \): \( \binom{4}{2} = 3 \) pairs of vertices are shared, so three fewer edges are necessary. \( \cov(X_C, X_D) = E(X_C X_D) - p^{12} = p^9 - p^{12} \leq p^9 \). This can happen \( \binom{n}{5} \) times, so the total contribution to the sum of the covariances is \( \leq \binom{n}{5}p^9 = \Theta(n^5p^9) \).

Notice that \( \Var(X_c) = E(X_c^2) - E(X_c)^2 = p^6 - p^{12} = \Theta(p^6) \). Then,

\[
\Var(X) = \sum_C \Var(X_C) + \sum_{C \neq D} \cov(X_C, X_D)
\]

\[
\leq \Theta(n^4p^6) + \Theta(n^6p^{11}) + \Theta(n^5p^9)
\]

\[
= \Theta(n^4p^6) + \Theta(n^6n^{-22/3}) + \Theta(n^5n^{-6}) \text{ for } p \approx n^{-2/3}
\]

\[
= \Theta(n^4p^6)
\]

And,

\[
\frac{\Var(X)}{(E(X))^2} = \frac{\Theta(n^4p^6)}{(\Theta(n^4p^6))^2} = \frac{1}{\Theta(n^4p^6)} \to 0 \text{ as } n \to \infty \text{ for } p \gg n^{-2/3}
\]

\[ \blacksquare \]