Tail Bounds

1 Markov/Chebyshev Bounds

1.1 Markov's Inequality

Recall Markov's Inequality: if X is a non-negative random variable, then

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

We'll use this to prove Adelman's Theorem (again). Choose as before the randomness vectors i.i.d. uniformly. Define the random variable X_i as the number of inputs in whose row we have not yet picked a 1 after i randomness strings have been chosen. Then $X_0 \leq 2^n$ (all rows). Since the new string $\boldsymbol{r}_{i+1}$ has, for each remaining row, probability $\geq 1/2$ of hitting that row,

$$\mathbb{E}[X_{i+1} \mid X_i] \leq \frac{1}{2} X_i.$$

Then by the law of total expectation,

$$\begin{aligned}
\mathbb{E}[X_i] &= \mathbb{E}[\mathbb{E}[X_i \mid X_{i-1}]] \\
&= \dots \\
&\leq 2^{-i} X_0 \\
&\leq 2^{n-i}.
\end{aligned}$$

For $i \geq n+1$, $\mathbb{E}[X_i \leq 1/2]$. Applying Markov's Inequality,

$$\begin{aligned}
\Pr[\text{all rows have been hit}] &= 1 - \Pr[\text{at least one row remains}] \\
&= 1 - \Pr[X_i \geq 1] \\
&\geq 1 - \frac{\mathbb{E}[X_i]}{1} \\
&\geq 1 - 2^{n-i}
\end{aligned}$$

This value decreases exponentially, and $i = n+1$ gives probability $\geq 1/2$, which by the Probabilistic Method shows that there is a set of $n+1$ randomness strings that cover all inputs.

1.2 Chebyshev's Inequality

Stronger tail bounds can be obtained if we know more about our random variable behavior. For a random variable X,

$$\begin{aligned}
Var[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
\sigma_X &= \sqrt{Var[X]}.
\end{aligned}$$

Then Chebyshev's Inequality is as follows: for every random variable X,

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{Var[X]}{t^2}.$$

This is often useful if X is the sum of pairwise independent random variables, or if $Var[X]$ is easy enough to analyze (and not too large).

Proof. Clearly

$$\Pr[|X - \mathbb{E}[X]| \geq t] = \Pr[(X - \mathbb{E}[X])^2 \geq t^2]$$

Let $Y = (X - \mathbb{E}[X])^2$. Then $Y \geq 0$. So by Markov's Inequality,

$$\begin{aligned} \Pr[Y \geq t^2] &\leq \frac{\mathbb{E}[Y]}{t^2} \\ &= \frac{\mathbb{E}[X - \mathbb{E}[X])^2]}{t^2} \\ &= \frac{Var[X]}{t^2}. \end{aligned}$$

$\square$

2 Finding a Median

2.1 Randomized Algorithm

Given a set S, represented as an unsorted array, find a median x such that exactly half of the elements are $\leq x$. The standard algorithm uses a divide-and-conquer approach with a randomly chosen pivot to run in linear time. Here, we will use a very different sampling-based algorithm which illustrates that random samples are representative, and uses a typical tail bound-style analysis.

Algorithm 1 Median-Finding Algorithm

Sample $n^{3/4}$ elements from S, pairwise independently and uniformly at random with replacement. Call this set R.
Sort R in time $\mathcal{O}(n^{3/4} \log n)$.
Let $\ell = n^{3/4}/2 - \sqrt{n}$ and $h = n^{3/4}/2 + \sqrt{n}$. Let $a = R_{(\ell)}$ and $b = R_{(h)}$ in sorted order of R.
Compare all elements of S to a and b to determine

$$P = \{x \in S : a \leq x \leq b\}.$$

Use this comparison to find the positions of a and b in the sorted set S.
if $n/2$ is not between those positions, or $|P| > 4n^{3/4} + 2$ **then**
    Start over.
else
    Sort P and determine the median from sorted P and the rank of a.
end if

This algorithm tries to ensure that the median is in P, and P is not too large. Note that the algorithm is always correct, but the runtime varies, so it is a Las Vegas algorithm.

2.2 Runtime Analysis

Lines 1, 2, and 9 run in time $\mathcal{O}(n^{3/4} \log n) = o(n)$. Lines 3 and 7 take time $\mathcal{O}(1)$. Line 4 makes $2n$ comparisons. So the total runtime is

$$(2n + o(n)) \cdot \# \text{ of restarts.}$$

The # of restarts is a geometric random variable, so the expected number of restarts is the inverse of the success probability.

What could go wrong in one iteration of the algorithm?

1. $a >$ median.
2. $b <$ median.
3. P is too large (a too small and/or b too large).

Cases 1 and 2 are symmetric, so we will analyze cases 1 and 3.

2.2.1 Case 1

Case 1 occurs when we undersample to the left of the median – that is, we had fewer than $n^{3/4}/2 - \sqrt{n}$ samples to the left of the median. The expected number of samples to the left of the median is $|R|/2 = n^{3/4}/2$. This means the number of samples deviated by at least $\sqrt{n}$ from its expectation.

Let X be the number of samples in R that are $\leq$ median. Let X_i be the indicator random variable representing whether sample i is $\leq$ median. Then $X = \sum_i X_i$, and

$$\mathbb{E}[X_i] = \Pr[X_i = 1] = 1/2,$$

so

$$\mathbb{E}[X] = |R|/2 = n^{3/4}/2.$$

Our goal is to bound $\Pr[|X - n^{3/4}/2| < \sqrt{n}]$. We'd like to apply Chebyshev, but we need the variance. Here,

$$Var[X] = \sum_i Var[X_i] + \sum_{i<j} Cov[X_i, X_j].$$

But all the covariance terms are 0 because of pairwise independence. Because the X_i are Bernoulli random variables,

$$\begin{aligned} Var[X] &= \sum_i Var[X_i] \\ &= \sum_i \Pr[X_i = 1] \cdot \Pr[X_i = 0] \\ &= \sum_i \frac{1}{4} \\ &= n^{3/4}/4. \end{aligned}$$

Now we can apply Chebyshev:

$$\begin{aligned} \Pr[|X - n^{3/4}/2| \geq \sqrt{n}] &\leq \frac{n^{3/4}/4}{\sqrt{n}^2} \\ &= n^{-1/4}/4. \end{aligned}$$

2.2.2 Case 3

We'd like to analyze the case when $|P| \geq 4n^{3/4}$, which occurs when a is too small and/or b is too large. For this to happen, we must have oversampled left of $n/2 - 2n^{3/4}$ and/or oversampled right of $n/2 + 2n^{3/4}$. We'll analyze the first case in detail.

Assume we randomly picked more than $n^{3/4}/2 - \sqrt{n}$ elements left of $n/2 - 2n^{3/4}$. Let X be the number of samples in R that are $\leq n/2 - 2n^{3/4}$. Then

$$\begin{aligned}\mathbb{E}[X] &= n^{3/4} \cdot (1/2 - 2n^{-1/4}) \\ &= n^{3/4}/2 - 2\sqrt{n}.\end{aligned}$$

Note that $\mathbb{E}[X]$ is lesser than our assumption by $\sqrt{n}$. Also,

$$\begin{aligned}Var[X] &= n^{3/4} \cdot (1/2 - 2n^{-1/4})(1/2 + 2n^{-1/4}) \\ &\leq n^{3/4}/4.\end{aligned}$$

Applying Chebyshev,

$$\begin{aligned}\Pr[a \text{ too small}] &\leq \Pr[|X - \mathbb{E}[X]| \geq \sqrt{n}] \\ &\leq \frac{n^{3/4}/4}{\sqrt{n}^2} \\ &= n^{-1/4}/4.\end{aligned}$$

2.2.3 Summary

We showed that the failure probabilities for all four cases are $n^{-1/4}/4$. Using the union bound, the overall failure probability is at most $n^{-1/4}$. So the success probability of any one iteration is at most $1 - n^{-1/4}$. Thus, the expected number of restarts is at most

$$\begin{aligned}\frac{1}{1 - n^{-1/4}} &\leq 1 + 2n^{-1/4} \\ &= 1 + o(1).\end{aligned}$$

Putting it all together, the total amount of work in expectation is

$$2(1 + o(1))n + o(n) = 2n + o(n).$$

The best known deterministic algorithm takes $3n$ comparisons, with a lower bound of $2n$ for all deterministic algorithms.

3 Chernoff/Hoeffding Bounds

These are tail bounds that give stronger guarantees on random variables X that can be written as $\sum_i X_i$ where each X_i is bounded, and all the X_i are mutually independent. This applies frequently in computer science-related scenarios. They are similar to a quantitative version of the "Law of Large Numbers": that the mean of enough i.i.d. samples converges to the population mean.

For Chernoff Bounds, assume we have $X = \sum_i X_i$ with X_i Bernoulli and $\Pr[X_i = 1] = p_i$. Let $\mu = \mathbb{E}[X] = \sum_i p_i$.

Theorem 1. *[Chernoff].*

1. *For any* $\delta > 0$,
$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

2. *For any* $0 < \delta < 1$,
$$\Pr[X < (1-\delta)\mu] < (e^{-\delta^2/2})^{\mu}.$$

A special case is when $p_i = p$*, so* X *is a sum of i.i.d random variables. Then* $\mu = np$*, so the bounds decrease exponentially in* n*.*

For Hoeffding Bounds, assume we have $X = \sum_i X_i$ with $a_i \leq X_i \leq b_i$ for all i deterministically.

Theorem 2. *[Hoeffding]. From survey on concentration inequalities by Mcdiarmid. For all* $\Delta \geq 0$,

$$\Pr[|X - \mathbb{E}[X]| \geq \Delta] \leq 2\exp\left(\frac{-2\Delta^2}{\sum_i (b_i - a_i)^2}\right).$$

The standard tail bounds are for independent random variables. Most have counterparts if the X_i are negatively correlated. The proof idea is that

$$\exp(\alpha \sum_i X_i) = \prod \exp(\alpha X_i),$$

and since the $\exp(\alpha X_i)$ are independent, the expectation is the product of expectations, so we can apply Markov's Inequality.

The main workflow using these bounds is as follows:

1. Decompose random variable into sum of independent Bernoulli's.
2. Use linearity of expectation to find the mean.
3. Use Chernoff bounds to find concentration.
4. Take a union bound over failure cases.
5. Choose δ big enough based on the result of the union bound.

Next class, we will see examples of this method.