Mathematical Toolkit **Spring 2023**

Lecture 12: May 1, 2023

Lecturer: Ali Vakilian (notes based on notes from Madhur Tulsiani)

1 Chernoff/Hoeffding Bounds

Consider n *independent* Boolean random variables $X_1, ..., X_n$, where X_i takes value 1 with probability p_i and 0 otherwise. Let $X = \sum_{i=1}^n X_i$. We set $\mu = \mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p_i$.

We will now derive a bound on the probability $\mathbb{P}[X \geq t]$ for $t = (1+\delta)\mu$ that is much stronger than what we were able to achieve from Chebyshev's inequaity by using the full power of mutual independence (rather than just pairwise independence).

Here is some intuition for the argument we will give. Define Y_i to be $e^{\lambda X_i}$ for some small $\lambda > 0$. So when $X_i = 1$ we have that $Y_i \approx 1 + \lambda$ and when $X_i = 0$ we have $Y_i = 1$, and $\mathbb{E}[Y_i] \approx 1 + p_i\lambda \approx e^{p_i\lambda}$. Consider now the *product* Y of the Y_i's. Since the Y_i are mutually independent, we have $\mathbb{E}[Y] = \prod_i \mathbb{E}[Y_i] \approx e^{\lambda \sum_i p_i} = e^{\lambda \mathbb{E}[X]}$. We can now apply Markov's inequality to say that there is at most a $1/k$ chance that $Y \geq k\mathbb{E}[Y]$. But notice that since $X = \frac{1}{\lambda}\ln(Y)$, this means that X is larger than $\frac{1}{\lambda}\ln(\mathbb{E}[Y])$ by at most an *additive* $\frac{\ln k}{\lambda}$. So, even if k is very large (so the probability of the event is very small), X is only larger than $\frac{1}{\lambda}\ln(\mathbb{E}[Y])$ by a small amount. Also, for small λ we have $\frac{1}{\lambda}\ln(\mathbb{E}[Y]) \approx \mathbb{E}[X]$. We are cheating here though: these approximations are not exact and become worse (and the true quantities go in the wrong direction) as λ gets large, so when we do this for real we will need to be careful. But this is the intuition. Let's now do the actual argument.

Using the fact that the function e^x is strictly increasing, we get that for $\lambda > 0$

$$\mathbb{P}[X \geq (1+\delta)\mu] \;=\; \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \;\overset{\text{(Markov)}}{\leq}\; \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}}.$$

We now have:

$$\begin{aligned}\mathbb{E}\left[e^{\lambda X}\right] \;=\; \mathbb{E}\left[e^{\lambda(X_1+\ldots+X_n)}\right] \;=\; \mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i}\right] &\overset{\text{(independence)}}{=} \prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] \\ &= \prod_{i=1}^n \left[p_i e^{\lambda} + (1-p_i)\right] \\ &= \prod_{i=1}^n \left[1 + p_i(e^{\lambda}-1)\right].\end{aligned}$$

At this point, we utilize the simple but very useful inequality:

$$\forall x \in R, \;\; 1 + x \leq e^x.$$

Since all the quantities in the previous calculation are non-negative, we can plug the above inequality in the previous calculation and we get:

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X}\right] &\leq \prod_{i=1}^{n} e^{\left(p_i(e^\lambda - 1)\right)} \\
&= e^{\sum_i p_i(e^\lambda - 1)} \\
&= e^{(e^\lambda - 1)\mu}
\end{aligned}$$

Thus, we get

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \;\leq\; \exp\Big((e^\lambda - 1)\mu - \lambda(1+\delta)\mu\Big).$$

We now want to minimize the right hand-side of the above inequality, with respect to λ. Setting the derivative of the exponent to zero, we get

$$e^\lambda \mu - (1+\delta)\mu = 0 \quad \Rightarrow \quad \lambda = \ln(1+\delta)\,.$$

Using this value for λ, we get

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \;\leq\; \frac{\exp\Big((e^\lambda - 1)\mu\Big)}{\exp\Big(\lambda(1+\delta)\mu\Big)} \;=\; \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \;=\; \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Exercise 1.1 *Prove similarly that*

$$\mathbb{P}\left[X \leq (1-\delta)\mu\right] \;\leq\; \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu.$$

(Note that $\mathbb{P}\left[X \leq (1-\delta)\mu\right] = \mathbb{P}\left[e^{-\lambda X} \geq e^{-\lambda(1-\delta)\mu}\right]$.) When $\delta \in (0,1)$, the above expressions can be simplified further. It is easy to check that

$$\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \leq e^{-\delta^2\mu/3}, \;\; 0 < \delta < 1,$$

and

$$\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu \leq e^{-\delta^2\mu/2}, \;\; 0 < \delta < 1.$$

So we get:

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1.$$

and

$$\mathbb{P}\left[X \leq (1-\delta)\mu\right] \leq e^{-\delta^2\mu/2}, \quad \text{for } 0 < \delta < 1.$$

1.1 Coin tosses once more

We will now compare the above bound with what we can get from Chebyshev's inequality. Let's assume that $X_1, ..., X_n$ are independent coin tosses, with $\mathbb{P}[X_i = 1] = \frac{1}{2}$. We want to get a bound on the value of $X = \sum_{i=1}^n X_i$. Using Chebyshev's inequality, we have

$$\mathbb{P}[|X - \mu| \geq \delta\mu] \;\leq\; \frac{\mathsf{Var}[X]}{\delta^2 \mu^2}.$$

And since in this particular case we have that $\mathsf{Var}[X] = n/4$ and $\mu = n/2$, we have

$$\mathbb{P}[|X - \mu| \geq \delta\mu] \;\leq\; \frac{1}{\delta^2 n}.$$

The above bound is only inversely polynomial in n, while the Chernoff-Hoeffding bound gives

$$\mathbb{P}[|X - \mu| \geq \delta\mu] \;\leq\; 2 \cdot \exp\left(-\delta^2 n/6\right),$$

which is exponentially small in n. This fact will prove very useful when taking a union bound over a large collection of events, each of which may be bounded using a Chernoff-Hoeffding bound. For example, consider the case where for m sets $S_1, \ldots, S_m \subseteq [n]$, we define

$$Z_{S_i} \;=\; \sum_{j \in S_i} X_j\,.$$

While the variables $Z_{S_1}, \ldots, Z_{S_m}$ are *not* necessarily independent, each of these is a sum of X_j variables, which are independent. Thus, we can say that for any S_i,

$$\mathbb{P}\left[\left|Z_{S_i} - \frac{|S_i|}{2}\right| \;\geq\; t\right] \;\leq\; 2 \cdot \exp\left(-2t^2/(3\,|S_i|)\right) \;\leq\; 2 \cdot \exp\left(-2t^2/(3n)\right)\,,$$

where we choose $\delta = 2t/\,|S_i|$ so that $\delta\,|S_i|\,/2 = t$. Thus, by a union bound over all $i \in [m]$, we get that

$$\mathbb{P}\left[\exists i \in [m].\; \left|Z_{S_i} - \frac{|S_i|}{2}\right| \;\geq\; t\right] \;\leq\; 2m \cdot \exp\left(-2t^2/(3n)\right)\,.$$

Thus, when $t = \sqrt{3n \cdot \ln m}$, the probability of the above event is at most $2/m$. Check that just using Chebyshev's inequality does not allow for such a strong bound on the probability of the above event.

Note that the above calculation used the following union bound

Exercise 1.2 *Let $E_1, \ldots, E_k$ be events on the same outcome space Ω. Then*

$$\mathbb{P}[E_1 \cup \cdots \cup E_k] \;\leq\; \sum_{i=1}^{k} \mathbb{P}[E_i]\,.$$

2 Random Vectors

Here is another interesting fact we can get using Chernoff/Hoeffding bounds. Suppose we pick m random vectors $v_1, \ldots, v_m$ in $\{-1,1\}^n$. Each of these vectors v_i will have the property that $\langle v_i, v_i \rangle = n$. But, it turns out that with high probability, for all $i \neq j$ we will have $|\langle v_i, v_j \rangle| \leq c\sqrt{n \log m}$ for some constant $c > 0$. So, even though we can have at most n orthogonal vectors in an n-dimensional space, we can have a much larger number of *nearly*-orthogonal vectors.

This fact comes immediately from Chernoff/Hoeffding bounds and the union bound. Fix some pair $i \neq j$, and for each $k \in \{1, 2, \ldots, n\}$ define indicator random variable X_k for the event that the kth coordinates of v_i and v_j are equal. Notice that $X_1, \ldots, X_n$ are independent with $\mathbb{P}[X_k = 1] = 1/2$. Let $X = \sum_k X_k$. By Chernoff/Hoeffding bounds, $\mathbb{P}[|X - n/2| \geq \delta n/2] \leq 2e^{-\delta^2 n/6}$. Notice that $|\langle v_i, v_j \rangle| = 2|X - n/2|$. So, using $\delta = 6\sqrt{\frac{\ln m}{n}}$ we have $\mathbb{P}\left[|\langle v_i, v_j \rangle| \geq 6\sqrt{n \ln m}\right] \leq 2e^{-6 \ln m} = 2/m^6$. So, by the union bound over all $O(m^2)$ pairs i, j we have that with high probability $|\langle v_i, v_j \rangle| = O(\sqrt{n \log m})$ for all $i \neq j$.

3 Balls and Bins revisited

We saw earlier that if we toss balls uniformly at random into n bins, then the expected number of balls we need to use until each bin has at least one ball in it is $\Theta(n \log n)$. Let's now consider some other statistics.

First, if we toss n balls into n bins, what is the expected fraction of empty bins? This is an easy direct calculation. Let X_i be the indicator random variable for the event that bin i remains empty. We have $\mathbb{E}[X_i] = \mathbb{P}[\text{no balls fall in bin } i] = (1 - 1/n)^n \approx 1/e$. So, the expected fraction of empty bins is $\approx 1/e$.

Next, if we toss n balls into n bins, how loaded will the most-loaded bin be? We can use Chernoff/Hoeffding bounds to argue that with high probability, no bin will have more than $t = \frac{3 \ln n}{\ln \ln n}$ balls in it.

Specifically, define Z_i = number of balls in bin i. We can write

$$Z_i = \sum_j X_{ij}, \quad \text{where} \quad X_{ij} = \begin{cases} 1 & \text{if ball } j \text{ is thrown in bin } i \\ 0 & \text{otherwise} \end{cases} .$$

Then, we have that each Z_i is a sum of n independent random variables with $\mathbb{E}[Z_i] = 1$. By Chernoff/Hoeffding bounds, we have that for each i,

$$\mathbb{P}[Z_i \geq t] \leq \frac{e^{t-1}}{t^t} \leq \left(\frac{e}{t}\right)^t .$$

To bound the maximum load in across all bins, we use a union bound to say that

$$\mathbb{P}\left[\exists i \in [n].\, Z_i \geq t\right] \;\leq\; \sum_{i=1}^{n} \mathbb{P}\left[Z_i \geq t\right] \;\leq\; n \cdot \left(\frac{e}{t}\right)^t ,$$

which is at most $\frac{1}{n}$ for the above value of t. Hence, with probability at least $1 - \frac{1}{n}$, the maximum number of balls in a bin is at most $\frac{3 \ln n}{\ln \ln n}$.