| | |
|---|---|
| **TTIC/CMSC 31150 Mathematical Toolkit** | **Spring 2013** |

Lecture 4: April 10, 2013

Madhur Tulsiani — Scribe: Haris Angelidakis

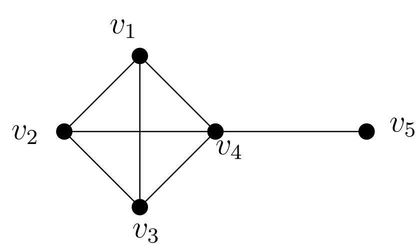
# 1 Chebyshev's Inequality recap

In the previous lecture, we used Chebyshev's inequality to get a bound on the probability that a random variable $X$ deviates from its expected value $\mu$ to some extent. More precisely, if we denote the variance of $X$ with $\sigma^2$ (and assume it's finite and non-zero), then Chebyshev's inequality is the following:

$$\mathbb{P}\left[|X-\mu| \geq k\sigma\right] \leq \frac{\mathsf{Var}\left[X\right]}{k^2\sigma^2} = \frac{1}{k^2}. \tag{1}$$

This is the so-called Second Moment Method, since it uses the second moment, i.e. the variance of $X$. As an application of the above inequality, we presented some thresholds in the Erdős-Rényi $G_{n,p}$ model. We extend that example here to show something that at first looks like a paradox.

So, we consider a random graph $G \sim G_{n,p}$ and let $Z_1$ be the number of copies of $K_4$ in $G$, and $Z_2$ be the number of copies of the following 5-node graph $G_0$ in $G$:

$G_0 \equiv$ 


Observe that the subgraph induced by the nodes $v_1, v_2, v_3, v_4$ is exactly $K_4$. Calculating the expectations of $Z_1$ and $Z_2$, we observe the following:

1. - $p \gg n^{-2/3} \Rightarrow \mathbb{E}\left[Z_1\right] \to \infty$
   - $p \ll n^{-2/3} \Rightarrow \mathbb{E}\left[Z_1\right] \to 0$
2. - $p \gg n^{-5/7} \Rightarrow \mathbb{E}\left[Z_2\right] \to \infty$
   - $p \ll n^{-5/7} \Rightarrow \mathbb{E}\left[Z_2\right] \to 0$

We now consider the case $n^{-5/7} \ll p \ll n^{-2/3}$. In this case, $\mathbb{E}\left[Z_1\right] \to 0$ and $\mathbb{E}\left[Z_2\right] \to \infty$. But, $K_4$ is a subgraph of $G_0$, so in every "appearance" of $G_0$ we obviously have an "appearance" of $K_4$. So, did we calculate something wrong? Observing things more carefully, we can see that, given a fixed $K_4$, there can be many copies of $G_0$ made by this copy of $K_4$, simply by connecting each vertex of $K_4$ with as many different vertices as we want. For each such vertex, we get a distinct copy of $G_0$.

This idea can be made precise so as to formally explain the above paradox, but we will skip the details.

In today's lecture, we will consider even sharper concentration bounds, which come in the form of the Chernoff/Hoeffding bounds. But before presenting the Chernoff bounds, we will also state the very useful Jensen's inequality, as is used in the context of probability theory. Consider the definition of the variance of a random variable $X$. We have $\mathsf{Var}[X] = \mathbb{E}\left[(X-\mu)^2\right]$. We obviously have $\mathsf{Var}[X] \geq 0$. Thus:

$$\mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \geq 0 \quad \Rightarrow \quad \mathbb{E}\left[X^2\right] \geq (\mathbb{E}[X])^2. \tag{2}$$

The above inequality is just a special case of the well-known Jensen's inequality.

## 2 Jensen's Inequality

At first, we need to define what a convex real-valued function is.

**Definition 2.1 (Convex function)** *Let $f : \mathbb{R} \to \mathbb{R}$. We say that $f$ is convex if for any $\lambda \in [0, 1]$ we have that*

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

The above property simply means that if we consider any two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$, the line segment connecting these two points on the plane lies above the graph of function $f$.

We can now state, without proof, Jensen's inequality.

**Theorem 2.2 (Jensen's Inequality)** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then for any random variable $X$, we have that*

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

We can now see that (2) is just a special case of Jensen's inequality, where we plug-in the convex function $f(x) = x^2$.

## 3 Chernoff/Hoeffding Bounds

We are now ready to get some sharper concentration bounds. We start by considering $n$ independent boolean random variables $X_1, ..., X_n$, each having value 1 with probability $p_i$. Let $Z = \sum_{i=1}^n X_i$. We set $\mu = \mathbb{E}[Z] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p_i$, and $p = \frac{\mu}{n} = \frac{\sum_{i=1}^n p_i}{n}$. So, we now want to get a bound on $\mathbb{P}[Z \geq (1+\delta)\mu]$.

At first, we use the fact that $e^x$ is strictly increasing, and so

$$\begin{aligned} \mathbb{P}[Z \geq (1+\delta)\mu] &= \mathbb{P}\left[e^{tZ} \geq e^{t(1+\delta)\mu}\right], \quad t > 0 \\ &\overset{(Markov)}{\leq} \frac{\mathbb{E}\left[e^{tZ}\right]}{e^{t(1+\delta)\mu}}. \end{aligned} \tag{3}$$

We now have:

$$
\begin{aligned}
\mathbb{E}\left[e^{tZ}\right] &= \mathbb{E}\left[e^{t(X_1+\ldots X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \\
&\overset{(indep)}{=} \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] = \prod_{i=1}^{n} \left[p_i e^t + (1-p_i)\right] \\
&= \prod_{i=1}^{n} \left[1 + p_i(e^t - 1)\right].
\end{aligned}
$$

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:

$$
\begin{aligned}
\mathbb{E}\left[e^{tZ}\right] &\leq \prod_{i=1}^{n} \exp\Big((e^t - 1)p_i\Big) \\
&= \exp\Big((e^t - 1)\mu\Big)
\end{aligned}
\tag{4}
$$

From (3) and (4) we get

$$\mathbb{P}\left[Z \geq (1+\delta)\mu\right] \leq \exp\Big((e^t - 1)\mu - t(1+\delta)\mu\Big). \tag{5}$$

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

$$
\begin{aligned}
e^t\mu - (1+\delta)\mu = 0 \Rightarrow \\
t = \ln(1+\delta).
\end{aligned}
$$

Using this value for $t$ in (5), we get

$$
\begin{aligned}
\mathbb{P}\left[Z \geq (1+\delta)\mu\right] &\leq \frac{\exp\Big((e^t - 1)\mu\Big)}{\exp\Big(t(1+\delta)\mu\Big)} \\
&\leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \\
&= \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}.
\end{aligned}
$$

Similarly, we can get that

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

(Note that $\mathbb{P}[Z \le (1-\delta)\mu] = \mathbb{P}\left[e^{-tZ} \ge e^{-t(1-\delta)\mu}\right]$.)

But we would like some simpler expression for the bound. It can be easily proved that

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

and so we finally get:

$$\mathbb{P}[Z \ge (1+\delta)\mu] \le e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1. \tag{6}$$

Similarly:

$$\mathbb{P}[Z \le (1-\delta)\mu] \le e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1. \tag{7}$$

From (6) and (7) we get

$$\mathbb{P}[|Z-\mu| \ge \delta\mu] \le 2e^{-\delta^2\mu/3}, \quad \text{for } 0 < \delta < 1. \tag{8}$$

This last inequality is the version of the Chernoff/Hoeffding bounds that we are going to use the most in the following. We now move to some applications.

## 4 Coin Tosses

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 $Z = \sum_{i=1}^{n} X_i$. Using Chebyshev's inequality as stated in (1), we get that

$$\mathbb{P}[|Z-\mu| \ge \delta\mu] \le \frac{\mathsf{Var}[Z]}{\delta^2\mu^2}.$$

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

$$\mathbb{P}[|Z-\mu| \ge \delta\mu] \le \frac{1}{\delta^2 n}.$$

The above bound is only inversely polynomial in $n$, while the one in (8) is exponentially small in $n$. This fact will prove very useful when we want to use a union bound in a large collection of event, as we will see in the application that follows.

## 5 Max-Cut in the Erdős-Rényi Model

Consider a graph $G \sim G_{n,p}$. Let $G = (V, E)$, with $|V| = n$ and $|E| = Z$ be the number of edges in $G$. We want to prove that the size of a Max-Cut of $G$ is with high probability roughly equal to $|E|/2$, which is equal to the expected size of a random cut.

At first, we can get a concentration bound on the number of edges in the graph, which will help us in the proof of the above statement. We have that $Z = \sum_{\{u,v\}:u,v\in V, u\neq v} X_{\{u,v\}}$, where $X_{\{u,v\}}$ is a random variable indicating if there exists an edge between $u$ and $v$. So we get

$$\mathbb{E}[Z] = \binom{n}{2} p.$$

Using Chernoff bounds, we get that

$$\mathbb{P}\left[|Z - \mathbb{E}[Z]| \geq \varepsilon \binom{n}{2} p\right] \leq 2\exp\left(\frac{-\varepsilon^2 \binom{n}{2} p}{3}\right).$$

So, if $\frac{-\varepsilon^2 \binom{n}{2} p}{3} \gg 1$, then w.h.p the number of edges in $G$ is some number in the interval

$$\left((1-\varepsilon)\binom{n}{2} p, (1+\varepsilon)\binom{n}{2} p\right).$$

To simplify the calculations that follow, we can set $\binom{n}{2} \simeq \frac{n^2}{2}$, as we can always adjust $\varepsilon$ to match the exact values. We also set $m = \frac{n^2 p}{2}$, which is roughly equal to $\mathbb{E}[Z]$. So, in order to prove our statement about the size of a max-cut, it is sufficient to show that the number of edges crossing any cut belongs in the interval $\left((1-\delta)\frac{m}{2}, (1+\delta)\frac{m}{2}\right)$.

In order to do so, we fix a cut $(S, \overline{S})$, with $|S| = k$ and $1 \leq k \leq \frac{n}{2}$. Let $Z_S$ be the number of edges crossing $(S, \overline{S})$. Observe that $Z_S$ can be written as a sum of indicator $0-1$ variables, and so we can use the Chernoff Bounds that we have already proved. We have that $\mu_S = \mathbb{E}[Z_S] = pk(n-k)$, which means that the expected size of the cut only depends on the size of $S$ and not on the specific elements of $S$, which makes sense if we think of how we generate our graph. A useful fact here is that $k(n-k) \leq \frac{n^2}{4}$, and so $\mu_S \leq \frac{pn^2}{4}$. So, we have

$$\begin{aligned}
\mathbb{P}[Z_S \geq (1+\delta)m/2] &\leq \mathbb{P}[Z_S \geq (1+\delta)\mu_S] \\
&\leq \exp\left(-\frac{\delta^2 \mu_S}{3}\right) \\
&= \exp\left(-\frac{\delta^2}{3} pk(n-k)\right) \\
&\leq \exp\left(-\frac{\delta^2}{6} pkn\right) \qquad \left(\text{since } k \leq \frac{n}{2}\right)
\end{aligned}$$

Taking a Union Bound now over all possible cuts, we have

$$\begin{aligned}
\mathbb{P}\left[\text{Max-Cut} \geq (1+\delta)\frac{m}{2}\right] &\leq \sum_{(S,\overline{S})} \mathbb{P}\left[Z_S \geq (1+\delta)\frac{m}{2}\right] \\
&= \sum_{k=1}^{n/2} \sum_{(S,\overline{S}):|S|=k} \mathbb{P}\left[Z_S \geq (1+\delta)\frac{m}{2}\right] \\
&\leq \sum_{k=1}^{n/2} \binom{n}{k} \exp\left(-\frac{\delta^2}{6} pkn\right) \\
&\leq \sum_{k=1}^{n/2} n^k \exp\left(-\frac{\delta^2}{6} pkn\right) \\
&= \sum_{k=1}^{n/2} \exp\left(k \ln n - \frac{\delta^2}{6} pkn\right).
\end{aligned}$$

Suppose now that $\frac{\delta^2 pn}{6} \geq 2\ln n \Rightarrow p \geq \frac{12\ln n}{\delta^2 n}$. Using this fact in the inequality above we get

$$\mathbb{P}\Big[\text{Max-Cut} \geq (1+\delta)\frac{m}{2}\Big] \leq \sum_{k=1}^{n/2} \exp\Big(-k\ln n\Big) = O\Big(\frac{1}{n}\Big).$$

The above inequality shows that if the probability $p$ is sufficiently large (to be more precise, if it such that the resulting graph has $\Omega(n\ln n)$ edges in expectation), then the size of a Max-Cut in this graph is w.h.p. very close to $\frac{|E|}{2}$. Observe that we cannot get this guarantee by using Chebyshev's inequality, as the number of events we are applying the union bound on is large, and the bound we can get from Chebyshev is only $O\big(\frac{1}{n}\big)$ for a single cut.

## 6 Randomized Routing in Networks

We will now consider a different application. Let's assume we have a network of nodes, placed on the vertices of an $n$-dimensional hypercube. In other words, we have a graph with node set $V = \{0,1\}^n$, and edge set $E = \{(x,y) : d_H(x,y) = 1\}$, where $d_H(x,y)$ is the Hamming distance of two $n$-bit strings, i.e. the number of bits in which $x$ and $y$ differ. We are given a *permutation* $\pi : \{0,1\}^n \to \{0,1\}^n$, which simply translates to the fact that vertex $x$ wants to send a packet to vertex $\pi(x)$. So, we want to find a routing strategy, such that each packet arrives to its destination at the minimum possible time. We will use a synchronous model, i.e., the routing occurs in discrete time steps, and in each time step, one packet is allowed to travel along each edge.

We are interested in *oblivious* strategies, i.e. for each $x$, the path which the packet going from $x$ to $\pi(x)$ will use only depends on $x$ and $\pi(x)$ and no other vertices. Observe that a packet from $x$ to $y$ takes time at least $d_H(x,y)$, and since $d_H(00...0, 11...1) = n$, we have a worst-case lower bound of $O(n)$ for any strategy.

For the above problem we have the following results:

**Theorem 6.1** *[KKT90] For any deterministic, oblivious routing strategy on the hypercube, there exists a permutation that requires* $\Omega\Big(\sqrt{\frac{2^n}{n}}\Big)$ *time steps.*

This is a bad lower bound for the worst-case scenario. Fortunately, randomization can give a significant improvement.

**Theorem 6.2** *[VB81] There exists a randomized, oblivious routing strategy that terminates in* $O(n)$ *time steps w.h.p.*

The randomized strategy of the above theorem consists of two phases:

- In phase 1, each packet is routed to an intermediate node $r(i)$, where $\sigma(i)$ is chosen uniformly at random.
- In phase 2, each packet is routed from $\sigma(i)$ to $\pi(i)$.

In both phases, we use the "bit-fixing" paths strategy to route the packets. In the "bit-fixing" strategy, we move from a vertex $x$ to a vertex $y$ by checking all bits of the two nodes from left to right, and at each bit in which $x$ differs from $y$, we make the corresponding change so as to reduce the Hamming distance between the two strings. Observe that the paths that we obtain from this strategy are always shortest paths. Also, note that $\sigma$ is not required to be a permutation of $\{0,1\}^n$, so this strategy is oblivious, in the sense that each node doesn't care about which intermediate node is chosen by other nodes. This strategy breaks the symmetry in the problem by simply choosing a random intermediate destination for each packet. This makes it impossible for an adversary to select a "bad" permutation.

The analysis of the above strategy will follow in the next lecture.

## References


[KKT90] C. Kaklamanis, D. Krizanc and A. Tsantilas. "Tight Bounds for Oblivious Routing in the Hypercube", in *Proceedings of the Symposium on Parallel Algorithms and Architecture*, 1990.

[VBL81] G. Valiant and G. J. Brebner. "Universal schemes for parallel communication", in *Proceedings of the 13th annual ACM Symposium on Theory of Computing*, 1981.