CSC 2421H : Graphs, Matrices, and Optimization Lecture 7 : 29 Oct 2018

Concentration Bounds

Lecturer: Sushant Sachdeva *Scribe: Fengwei Sun*

1 Scalar Chernoff Bound

Definition 1.1. *Let* $X_1, \ldots, X_t$ *be independent random variables such that*

$$0 \leq X_i \leq R, \mathbb{E}\sum_i X_i = \sum_i \mathbb{E}X_i = \mu$$

Then for all $0 < \epsilon < 1$*, we have*

$$P[\sum_i X_i \geq (1+\epsilon)\mu] \leq e^{-\frac{\epsilon^2\mu}{3R}}, P[\sum_i X_i \leq (1-\epsilon)\mu] \leq e^{-\frac{\epsilon^2\mu}{2R}}$$

Example: Suppose we conduct t independent tosses of a fair coin. Let $X_i = \begin{cases} 1 & if\ heads \\ 0 & o/w \end{cases}$. Then the number of heads in this trial is $\sum_{i=1}^t X_i$, and $\mathbb{E}(\#\ of\ heads) = \mathbb{E}\sum_{i=1}^t X_i = \frac{t}{2}$.

To obtain a good estimate of the probability that we see at least 600 heads out of 1000 tosses, we can apply the Chernoff bound with the parameters $\epsilon = 0.2, R = 1, \mu = 500$, and get

$$P(at\ least\ 600\ heads\ out\ of\ 1000\ tosses) = P(\sum_{i=1}^{1000} X_i \geq (1+\epsilon)500) \leq e^{-\frac{0.2^2 * 500}{3}} = e^{-\frac{20}{3}} \approx e^{-7}$$

Question: You can a coin with a bias in $\{\frac{1}{2}+\alpha, \frac{1}{2}-\alpha\}$. How many tosses do you need to decide which bias with probability of at least $1-\delta$?

Algorithm:

1. Toss the coin t times independently;
2. If there are at least $\frac{t}{2}$ heads, output $\frac{1}{2}+\alpha$; otherwise output $\frac{1}{2}-\alpha$.

We would like to bound $P(failure) \leq \delta$.

Case 1: The coin has a bias $\frac{1}{2}+\alpha$. Then $P(failure) = P(\sum X_i \le \frac{t}{2})$.

To apply Chernoff, we find that $R = 1, \mu = t(\frac{1}{2}+\alpha)$. Since we want $(1-\epsilon)\mu = \frac{t}{2}$, we have $\epsilon = 1 - \frac{t}{2\mu} = 1 - \frac{1}{2\alpha}$. Therefore,

$$P(\sum X_i \le \frac{t}{2}) = P(\sum X_i \le (1-\epsilon)\mu) \le e^{-\frac{\epsilon^2\mu}{2}} \le \delta$$

Since $\epsilon^2\mu \approx \Theta(t\alpha^2)$, we have

$$t \ge \frac{\Theta(1)}{\alpha^2}\log\frac{1}{\delta}$$

Case 2: The coin has a bias $\frac{1}{2}-\alpha$. Then $P(failure) = P(\sum X_i \ge \frac{t}{2})$.

To apply Chernoff, we find that $R = 1, \mu = t(\frac{1}{2}-\alpha)$. Since we want $(1+\epsilon)\mu = \frac{t}{2}$, we have $\epsilon = \frac{t}{2\mu} - 1 = \frac{1}{2\alpha} - 1$. Therefore,

$$P(\sum X_i \ge \frac{t}{2}) = P(\sum X_i \ge (1+\epsilon)\mu) \le e^{-\frac{\epsilon^2\mu}{3}} \le \delta$$

Since $\epsilon^2\mu \approx \Theta(t\alpha^2)$, we have

$$t \ge \frac{\Theta(1)}{\alpha^2}\log\frac{1}{\delta}$$

Both cases indicate that the t is bounded by the logarithm of $\frac{1}{\delta}$, which means that t would be relatively small even for very small δ.

2 Matrix Chernoff Bound

Definition 2.1. *Let $X_1, \ldots, X_t \in \mathbb{R}^{d\times d}$ be symmetric independent random variables such that*

$$0 \preceq X_i \preceq RI,\ \mu_{min}I \preceq \mathbb{E}\sum X_i \preceq \mu_{max}I$$

Then we have

$$P[\lambda_{max}(\sum X_i) \ge (1+\epsilon)\mu_{max}] \le de^{-\frac{\epsilon^2\mu_{max}}{3R}},$$

$$P[\lambda_{min}(\sum X_i) \le (1-\epsilon)\mu_{min}] \le de^{-\frac{\epsilon^2\mu_{min}}{2R}}$$

Note:

1. The condition that $0 \preceq X_i \preceq RI$ is equivalent to $||X_i|| \le R$, or $\lambda_{max}(X_i) \le R$, where $||A|| = \max_{x\neq 0}\frac{||Ax||}{||x||}$, and when A is symmetric, $||A|| = max\{\lambda_{max}, -\lambda_{min}\}$.
2. $\mu_{min} = \lambda_{min}(\mathbb{E}\sum X_i), \mu_{max} = \lambda_{max}(\mathbb{E}\sum X_i)$.

Example: (Construction of random expander graphs) Suppose we would like to generate an expander graph with n vertices (assuming n is even).

Define a matching as a graph where $d_v = 1$ for all vertices v. Let $H = \frac{1}{t}$(union of t independent perfect matching). Notice that for all vertices u in the graph H, $d_u = 1$.

Using the matrix Chernoff bound, we can show that H is an expander.

The Laplacian of H is $L_H = \sum_i \frac{1}{t} L_i$, where L_i is the Laplacian of the i^{th} matching.

Let $X_i = \frac{1}{t} L_i$. We know that $X_i \succeq 0$ and $\lambda_{max}(X_i) = \frac{1}{t}\lambda_{max}(L_i) = \frac{2}{t}$.

Also, if we look at a specific vertex u in each matching, it is connected to all other vertices with equal probablity $\frac{1}{n-1}$. This indicates that $\mathbb{E}L_H = \mathbb{E}L_1 = \frac{1}{n-1} L_{K_n}$.

Before we apply Chernoff bound on $X_i's$, one issue we notice is that $\lambda_{min}(X_i) = 0$.

To fix that, we let $X_i = \frac{1}{t} L_i + \frac{1}{t(n-1)} \mathbb{1}\mathbb{1}^T$. Now we have $\mathbb{E}\sum X_i = \frac{1}{n-1} L_{K_n} + \frac{1}{n-1}\mathbb{1}\mathbb{1}^T = \frac{n}{n-1} I_n$, and thus $\mu_{max} = \mu_{min} = \frac{n}{n-1}$.

We can also show that, $\lambda_{max}(X_i) \le \frac{2}{t}$ after the change of variable.

Let $y = \hat{y} + c\frac{\mathbb{1}}{\sqrt{n}}$ where $\hat{y}^T \mathbb{1} = 0$. Then

$$y^T X_i y = \hat{y}^T(\frac{1}{t}L_i)\hat{y} + \frac{c^2 n}{t(n-1)} \le (\frac{2}{t})\hat{y}^T\hat{y} + \frac{n}{(n-1)t}c^2 \le \frac{2}{t}(\hat{y}^T\hat{y} + c^2) \le \frac{2}{t}||y||^2 = \frac{2}{t}$$

Now we apply the Chernoff bound, and get the following:

$$P[\lambda_{max}(\sum X_i) \ge (1+\epsilon)\frac{n}{n-1}] \le ne^{\frac{-\epsilon^2 \frac{n}{n-1}}{3\cdot\frac{2}{t}}} = ne^{-\frac{\epsilon^2 t}{6}(\frac{n}{n-1})}$$

If we pick $t \ge \frac{12}{\epsilon^2}\log n$, we have $P[\lambda_{max}(\sum X_i) \ge (1+\epsilon)\frac{n}{n-1}] \le n \cdot \frac{1}{n^2} = \frac{1}{n}$.

Similarly, we have $P[\lambda_{min}(\sum X_i) \le (1-\epsilon)\frac{n}{n-1}] \le \frac{1}{n}$.

Therefore, we can conclude that, with probability of at least $1 - \frac{2}{n}$,

$$\lambda_{max}(\sum X_i) \le (1+\epsilon)\frac{n}{n-1}, \lambda_{min}(\sum X_i) \ge (1-\epsilon)\frac{n}{n-1}$$

or

$$(1-\epsilon)\frac{n}{n-1} I \preceq \sum X_i \preceq (1+\epsilon)\frac{n}{n-1}$$

To see that H is a good approximation of a complete graph, let $\Pi = I - \frac{1}{n}\mathbb{1}\mathbb{1}^T = \frac{1}{n}L_{K_n}$. Notice that $\Pi^2 = \Pi$.

Consider $\Pi^T(\sum X_i)\Pi$. We have

$$(1-\epsilon)\frac{n}{n-1}\frac{1}{n}L_{K_n} \preceq \Pi^T(\sum X_i)\Pi \preceq (1+\epsilon)\frac{n}{n-1}\frac{1}{n}L_{K_n}$$

Since

$$\Pi^T(\sum X_i)\Pi = \Pi^T(L_H + \frac{1}{n-1}\mathbb{1}\mathbb{1}^T)(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T)$$

$$= \Pi^T(L_H + \frac{1}{n-1}\mathbb{1}\mathbb{1}^T - \frac{n}{n(n-1)}\mathbb{1}\mathbb{1}^T) = \Pi^T L_H = L_H$$

Therefore,

$$(1-\epsilon)\frac{1}{n-1}L_{K_n} \preceq L_H \preceq (1+\epsilon)\frac{1}{n-1}L_{K_n}$$

Now we get an ϵ-expander H with $t \cdot \frac{n}{2} = \Theta(\frac{n\log n}{\epsilon^2})$ edges, and $L_H \approx_\epsilon L_{K_n}$.

In general, we would like to write the Chernoff bound as the following:

With probability of at least $1 - 2de^{-\frac{\epsilon^2\mu_{min}}{2R}}$, $(1-\epsilon)\mu_{min}I \preceq \sum X_i \preceq (1+\epsilon)\mu_{max}I$.

Definition 2.2. *$H = (V, E')$ is an ϵ-spectral sparsifier of $G = (V, E)$ if $\frac{1}{1+\epsilon}L_G \preceq L_H \preceq (1+\epsilon)L_G$, denoted as $L_H \approx_\epsilon L_G$.*

Equivalently, $\forall x \in \mathbb{R}^V$, $\frac{1}{1+\epsilon}x^T L_G x \leq x_H^L x \leq (1+\epsilon)x^T L_G x$.

Note: Let $x = \mathbb{1}_S$ where $S \subset V$. Then $x^T L_H x = \sum_{(u,v)\in E} w(u,v)(x(u)-x(v))^2 = |E(S,\bar{S})|$.

Theorem: For all $G = (V, E)$, there exists $H = (V, E')$ such that $L_H \approx_\epsilon L_G$ and $|E'| \leq \Theta(\frac{n\log n}{\epsilon^2})$