Lecture 4: Chernoff bounds and randomized rounding CSE 525, Spring 2019
Instructor: James R. Lee

1 Large deviations

We have seen how knowledge of the variance of a random variable X can be used to control deviation of X from its mean. This is the heart of the second moment method. But often we can control even higher moments, and this allows us to obtain much stronger concentration properties.

A prototypical example is when $X_1, X_2, \ldots, X_n$ is a family of independent (but not necessarily identically distributed) $\{0,1\}$ random variables and $X = X_1 + X_2 + \cdots + X_n$. Let $p_i = \mathbb{E}[X_i]$ and define $\mu = \mathbb{E}[X] = p_1 + p_2 + \cdots + p_n$. In that case, we have the following multiplicative form of the "Chernoff bound."

Theorem 1.1 (Chernoff bound, mutiplicative error)**.** *For every $\beta \geqslant 1$, it holds that*

$$\mathbb{P}\left(X \geqslant \beta\mu\right) \leqslant \left(\frac{e^{\beta-1}}{\beta^\beta}\right)^\mu, \tag{1.1}$$

and

$$\mathbb{P}\left(X \leqslant \frac{\mu}{\beta}\right) \leqslant \left(\frac{e^{1/\beta-1}}{\beta^\beta}\right)^\mu. \tag{1.2}$$

It's easy to use these formulae, but it sometimes helps to employ the slightly weaker bounds: If we put $\beta = 1 + \delta$ for $0 < \delta < 1$,

$$\mathbb{P}\left(X \geqslant (1+\delta)\mu\right) \leqslant e^{-\frac{\delta^2\mu}{3}},$$
$$\mathbb{P}\left(X \leqslant (1-\delta)\mu\right) \leqslant e^{-\frac{\delta^2\mu}{2}}.$$

The main point is that these tail bounds go down *exponentially* in the mean μ and the multiplicative deviation β, as opposed to the previous tail bounds we've seen (Markov and Chebyshev) that only go down polynomially.

Proof of Theorem 1.1. Much as we proved Chebyshev's inequality by applying Markov's inequality to the random variable $|X - \mathbb{E} X|^2$, the Chernoff bound is proved by applying a function to the underlying random variable X and then applying Markov's inequality.

Let $t \geqslant 0$ be a parameter we will choose later and write

$$\mathbb{P}[X \geqslant \beta\mu] = \mathbb{P}\left[e^{tX} \geqslant e^{t\beta\mu}\right] \leqslant \frac{\mathbb{E}[e^{tX}]}{e^{t\beta\mu}}. \tag{1.3}$$

The point of applying the function $X \mapsto e^{tX}$ is that we can exploit independence:

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(X_1+\cdots+X_n)}\right] = \mathbb{E}\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{tX_i}\right]. \tag{1.4}$$

Now write:

$$\mathbb{E}[e^{tX_i}] = (1-p_i) + p_i e^t = 1 + p_i(e^t - 1) \leqslant e^{p_i(e^t-1)},$$

where the last inequality uses $1 + x \leqslant e^x$ which is valid for all $x \in \mathbb{R}$.

Plugging this into (1.4) yields

$$\mathbb{E}\left[e^{tX}\right] \leqslant \prod_{i=1}^{n} e^{p_i(e^t-1)} = e^{\mu\left(e^t-1\right)}$$

Now recalling (1.3), we have

$$\mathbb{P}[X \geqslant \beta\mu] \leqslant e^{\mu(e^t-1-\beta t)}\,.$$

Choosing $t = \ln\beta$ yields (1.1). One can prove (1.2) similarly. □

1.1 Randomized rounding

A classical technique in the field of approximation algorithms is to write down a linear programming relaxation of a combinatorial problem. The linear program (LP) is then solved in polynomial time, and one *rounds* the fractional solution to an integral solution that is, hopefully, not too much worse than the optimal solution. A classical example goes back to Raghavan and Thompson.

Let $D = (V, A)$ be a directed network, and suppose that we are given a sequence of *terminal pairs* $(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)$ where $\{s_i\}, \{t_i\} \subseteq V$. We use $\mathfrak{I} = (D, \{(s_i, t_i)\})$ denote this instance of the *min-congestion disjoint paths problem*. The goal is to choose, for every i, a directed s_i-t_i path γ_i in D so as to minimize the maximum *congestion* of an arc $e \in A$:

$$\mathrm{opt}(\mathfrak{I}) = \text{minimize}\left\{\max_{e\in A} \#\{i : e \in \gamma_i\}\right\}.$$

This problem is NP-hard. Our goal will be an *approximation algorithm* that outputs a solution $\{\gamma_i\}$ so that the congestion of every edge is at most $\alpha \cdot \mathrm{opt}$. The number α is called the *approximation factor* of our algorithm.

A fractional relaxation. Our approach will be to compute first a fractional solution that sends 1 unit of flow from s_i to t_i for every i. A flow can be thought of in the following way. Let $\mathcal{P}_i$ denote the set of simple, directed s_i-t_i paths in D, and let $\mathcal{P}$ denote the set of all simple directed paths in D. Here, simple means that no arc is repeated.

A *multi-flow* F is a mapping $F : \mathcal{P} \to \mathbb{R}_+$ of paths to nonnegative real numbers. The multi-flow F *routes the demands* $\{(s_i, t_i)\}$ if, for every $i = 1, \ldots, k$, we have $\sum_{\gamma\in\mathcal{P}_i} F(\gamma) = 1$, i.e. we send at least one unit of flow from s_i to t_i for every i. Finally, the *congestion* of an arc $e \in A$ under the flow F is the value $\mathrm{con}_F(e) = \sum_{\gamma\in\mathcal{P}:e\in\gamma} F(\gamma)$, i.e. the amount of flow passing through the arc e.

We make the definition:

$$\mathrm{LP}(\mathfrak{I}) = \text{minimize}_F \left\{\max_{e\in A} \mathrm{con}_F(e)\right\},$$

where the minimum is over all multi-flows F that route the demands $\{(s_i, t_i)\}$. It should be clear that $\mathrm{LP}(\mathfrak{I}) \leqslant \mathrm{opt}(\mathfrak{I})$. The reason we write $\mathrm{LP}(\mathfrak{I})$ is that this value can be computed by a linear program of polynomial-size. This is not precisely clear from our formulation because there are possibly exponentially many paths in $\mathcal{P}$, but there is a compact formulation of the LP using standard techniques (see the remark at the end of this section).

Given a multi-flow F, we will round it to an *integral* multi-flow F', where an integral flow is one such that, for every $i = 1, \ldots, k$, we have $F'(\gamma) = 1$ for *exactly* one $\gamma \in \mathcal{P}_i$. Note that an integral flow represents a solution to the initial disjoint paths problem. Furthermore, we will now show that for some $\alpha \geqslant 1$, we have

$$\max_{e\in A} \mathrm{con}_{F'}(e) \leqslant \alpha \cdot (1 + \max_{e\in A} \mathrm{con}_F(e))\,.$$

In particular, if we apply this to the optimal fractional flow F^*, we arrive at

$$\max_{e \in A} \mathrm{con}_F(e) \leqslant \alpha \cdot (1 + \max_{e \in A} \mathrm{con}_{F^*}(e)) = \alpha(1 + \mathrm{LP}(\mathfrak{I})) \leqslant \alpha(1 + \mathrm{opt}(\mathfrak{I})) \leqslant 2\alpha \cdot \mathrm{opt}(\mathfrak{I}) ,$$

implying that we have achieved an 2α-approximation to the optimal solution. (Note that we have used the trivial bound opt $\geqslant 1$.)

Theorem 1.2. *Let $n = |V|$ and suppose that $n \geqslant 4$. If there is a multi-flow F that routes the demands $(s_1, t_1), \ldots, (s_k, t_k)$, then there exists an integral multi-flow F' that routes the demands, and furthermore*

$$\max_{e \in A} F'(e) \leqslant C \frac{\log n}{\log \log n} \left(1 + \max_{e \in A} F(e)\right) , \tag{1.5}$$

where $C > 0$ is a universal constant.

Proof. We will produce a *random* integral multi-flow F' that routes the demands $\{(s_i, t_i)\}$ and argue that it satisfies the conditions of the theorem with high probability.

For every $i = 1, \ldots, k$, we do the following independently. We know that $\sum_{\gamma \in \mathcal{P}_i} F(\gamma) = 1$. Thus we can think of F as providing a probability distribution over s_i-t_i paths. We let γ_i denote a random s_i-t_i path chosen with probability $F(\gamma)$ for $\gamma \in \mathcal{P}_i$.

The set of paths $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ gives us an integral multi-flow F'. We are left to bound the maximum congestion of an edge. To this end, fix an edge $e \in A$. For every $\gamma \in \mathcal{P}$ such that $e \in \gamma$, let X_γ be the indicator random variable that is 1 when the path γ is chosen in the rounding. Then the number of edges going through the edge e after rounding is given by the random variable

$$\mathrm{con}_{F'}(e) = \sum_{\gamma : e \in \gamma} X_\gamma . \tag{1.6}$$

We may assume that $\mathrm{con}_F(e) \geqslant 1$ because we are comparing $\mathrm{con}_{F'}(e)$ to $1 + \mathrm{con}_F(e)$ in (1.5).

First, we have

$$\mathbb{E}[\mathrm{con}_{F'}(e)] = \sum_{\gamma : e \in \gamma} \mathbb{E}[X_\gamma] = \sum_{\gamma : e \in \gamma} F(\gamma) = \mathrm{con}_F(e) .$$

So at least in expectation, the congestion does not increase. If the $\{X_\gamma\}$ random variables were independent, then we could apply the Chernoff bound. Unfortunately, this not necessarily the case. For instance, if $\gamma, \gamma' \in \mathcal{P}_i$ both contain the edge e, then X_γ and $X_{\gamma'}$ are not independent; in fact, at most one of them can be equal to 1. Thus we will first rewrite $\mathrm{con}_{F'}(e)$ as a sum of independent $\{0, 1\}$ random variables.

Let Y_i be the indicator variable that equals 1 if the unique s_i-t_i path in F' uses the edge e, i.e. if $e \in \gamma_i$. Then the $\{Y_i\}$ are independent (since we round each s_i-t_i pair independently). Moreover, we have $Y_i = \sum_{\gamma \in \mathcal{P}_i : e \in \gamma} X_\gamma$, so $\mathrm{con}_{F'}(e) = \sum_{i=1}^k Y_i$.

Since $\mathrm{con}_{F'}(e)$ is a sum of independent $\{0, 1\}$ random variables, we can apply the Chernoff bound (Theorem 1.1) to conclude that

$$\mathbb{P}\left[\mathrm{con}_{F'}(e) \geqslant \beta \cdot \mathrm{con}_F(e)\right] \leqslant \left(\frac{e^{\beta - 1}}{\beta^\beta}\right)^{\mathrm{con}_F(e)} \leqslant \frac{e^{\beta-1}}{\beta^\beta} ,$$

where in the last inequality we have used our assumption that $\mathrm{con}_F(e) \geqslant 1$. We would like to choose the latter bound to be at most n^{-3}. To do this, we need to choose $\beta = C \frac{\log n}{\log \log n}$ for some constant C. (You should check that this is the right choice of β.)

Setting β like this, we have

$$\mathbb{P}\left[\mathrm{con}_{F'}(e) \geqslant \beta \cdot \mathrm{con}_F(e)\right] \leqslant \frac{1}{n^3},$$

and thus by a union bound over the n^2 possible edges,

$$\mathbb{P}\left[\exists e \in A \text{ such that } \mathrm{con}_{F'}(e) \geqslant \beta \cdot \mathrm{con}_F(e)\right] \leqslant n^2 \cdot \frac{1}{n^3} \leqslant \frac{1}{n}.$$

Thus with probability at least $1 - \frac{1}{n}$, our integral flow F' satisfies the claim of the theorem. □

Remark 1.3. Note that if we knew $\mathrm{con}_F(e) \geqslant C' \log n$ for some constant C' and every $e \in A$, then we could actually choose $\beta = O(1)$ and still achieve a bound of n^{-3} on the probability of an over congested edge. This means that if all the fractional congestions are $\Omega(\log n)$, we can get an $O(1)$ approximation.

Remark 1.4. To compute the optimal fractional multi-flow, we write a linear program with variables $\{F_e : e \in A\}$. Our program should minimize the value λ such that $F_e \leqslant \lambda$ for every $e \in A$. Moreover, to make sure that the variables $\{F_e : e \in A\}$ correspond to an optimal flow, we should add the flow constraints at every non-terminal vertex: The flow in should be equal to the flow out of the vertex. At terminals, we have to allow there to be a surplus or deficit based on whether we are at a source or a sink. This program has $O(m)$ variables and $O(m + n)$ linear constraints, where $m = |A|$ and $n = |V|$.