15-859(M) Randomized Algorithms
Notes for 01/24/11

```
* useful probabilistic inequalities: Markov, Chebyshev, Chernoff
* Proof of Chernoff bounds
* Application: Randomized rounding for randomized routing
```

Useful probabilistic inequalities

Say we have a random variable X. We often want to bound the probability that X is too far away from its expectation. [In first class, we went in other direction, saying that with reasonable probability, a random walk on n steps reached at least $\sqrt{n}$ distance away from its expectation]
Here are some useful inequalities for showing this:

Markov's inequality: Let X be a non-negative r.v. Then for any positive k:

$$\mathbf{Pr}[X \geq k\mathbf{E}[X]] \ \leq \ 1/k.$$

(No need for k to be integer.) Equivalently, we can write this as:

$$\mathbf{Pr}[X \geq t] \ \leq \ \mathbf{E}[X]/t.$$

Proof. $\mathbf{E}[X] \geq \mathbf{Pr}[X \geq t] \cdot t + \mathbf{Pr}[X < t] \cdot 0 = t \cdot \mathbf{Pr}[X \geq t]$.

Defn of Variance: $\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$. Standard deviation is square root of variance. Can multiply out variance definition to get:

$$\mathbf{var}[X] = \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2.$$

Chebyshev's inequality: Let X be a r.v. with mean μ and standard deviation σ. Then for any positive t, have:

$$\mathbf{Pr}[|X - \mu| > t\sigma] \ \leq \ 1/t^2.$$

Proof. Equivalently asking what is the probability that $(X - \mu)^2 > t^2\mathbf{var}[X]$. Now, just think of l.h.s. as a new non-negative random variable Y. What is its expectation? So, just apply Markov's inequality.

Let's suppose that our random variable $X = X_1 + \ldots + X_n$ where the X_i are simpler things that we can understand. Suppose there is not necessarily any independence. Then we can still compute the expectation

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \ldots + \mathbf{E}[X_n]$$

and use Markov. (i.e., expectation is same as if they were independent)

Suppose we have pairwise independence. Then, $\mathbf{var}[X]$ is same as if the X_i were fully independent. In fact, $\mathbf{var}[X] = \sum_i \mathbf{var}[X_i]$.

Proof.

$$\begin{aligned} \mathbf{E}[X^2] - (\mathbf{E}[X])^2 &= \sum_i \sum_j \mathbf{E}[X_i X_j] - \sum_i \sum_j \mathbf{E}[X_i]\mathbf{E}[X_j] \\ &= \sum_i E[X_i^2] - \sum_i E[X_i]^2 \end{aligned}$$

where the last equality holds because $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$ for independent random variables, and all pairs here are independent except when $i = j$. So, can apply Chebyshev easily.

Chernoff and Hoeffding bounds

What if the X_i's are fully independent? Let's say X is the result of a fair, n-step $\{-1, +1\}$ random walk (i.e., $\mathbf{Pr}[X_i = -1] = \mathbf{Pr}[X_i = +1] = 1/2$ and the X_i are mutually independent.) In this case, $\mathbf{var}[X_i] = 1$ so $\mathbf{var}[X] = n$ and $\sigma(X) = \sqrt{n}$. So, Chebyshev says:

$$\mathbf{Pr}[|X| \geq t\sqrt{n}] \leq 1/t^2.$$

But, in fact, because we have full independence, we can use the stronger *Chernoff* and *Hoeffding* bounds that in this case tell us:

$$\mathbf{Pr}[X \geq t\sqrt{n}] \leq e^{-t^2/2}.$$

The book contains some forms of these bounds. Here are some forms of them that I have found to be especially convenient.

Let $X_1, \ldots, X_n$ be a sequence of n independent $\{0, 1\}$ random variables with $\mathbf{Pr}[X_i = 1] = p_i$ not necessarily the same. Let S be the sum of the RVs, and let $\mu = \mathbf{E}[S]$. Then, for $0 \leq \delta \leq 1$, the following inequalities hold:

- $\mathbf{Pr}[S > (1+\delta)\mu] \leq e^{-\delta^2\mu/3}$,
- $\mathbf{Pr}[S < (1-\delta)\mu] \leq e^{-\delta^2\mu/2}$.

Additive bounds:

- $\mathbf{Pr}[S \geq \mu + \delta n] \leq e^{-2n\delta^2}$.
- $\mathbf{Pr}[S \leq \mu - \delta n] \leq e^{-2n\delta^2}$.

Also, for any $k > 1$, we get:

- $\mathbf{Pr}[S > k\mu] < \left(\frac{e^{k-1}}{k^k}\right)^{\mu}$.

We're going to prove the additive form of the bounds. The two directions are symmetric (the same proof works) so we will just focus on the upper end. To prove this, we are going to make heavy use of the fact that for independent random variables, the expected value of the product is the product of the expectations.

Proof. Consider the random variable $Z = e^{\lambda S}$, where λ is a quantity we will optimize for later. Using Markov's inequality, we can now say:

$$\mathbf{Pr}[S \geq \mu + \delta n] \;=\; \mathbf{Pr}[Z \geq e^{\lambda(\mu+\delta n)}] \;\leq\; \mathbf{E}[Z]/e^{\lambda(\mu+\delta n)}.$$

To figure out the RHS, we need to get a handle on $\mathbf{E}[Z]$. Using the fact that the X_i are independent, we get:

$$\mathbf{E}[Z] \;=\; \mathbf{E}[e^{\lambda(X_1+\ldots+X_n)}] \;=\; \mathbf{E}[e^{\lambda X_1}]\mathbf{E}[e^{\lambda X_2}]\cdots\mathbf{E}[e^{\lambda X_n}]$$

Now, if we were really lucky, we would have $\mathbf{E}[e^{\lambda X_i}] = e^{\lambda \mathbf{E}[X_i]}$. In that case, we would get $\mathbf{E}[Z] = e^{\lambda \mathbf{E}[S]} = e^{\lambda\mu}$. Plugging this in, we would get an overall bound of $e^{-\lambda\delta n}$, which would be great, especially since λ hasn't been fixed yet so we could make it as large as we want! But really, the two quantities are not equal, and in fact the gap depends on λ. One intuitive way to think of what is going on is like this. Let μ^* be the value of S under which Z equals its expectation. It is very unlikely that S will be much larger than μ^* because that will cause Z to be a *lot* larger than its expectation (since Z is exponential in S), especially when λ is large. On the other hand, the larger λ is, the larger μ^* is compared to μ. So we will then solve for the optimal tradeoff.

Specifically, let $R_i = \mathbf{E}[e^{\lambda X_i}]/e^{\lambda p_i}$. We can solve for the numerator: with probability p_i we get e^λ and with probability $1 - p_i$ we get $e^0 = 1$. So, we have:

$$R_i = \frac{p_i e^\lambda + (1 - p_i)}{e^{\lambda p_i}}.$$

Now, the claim is that for all values of p_i and λ we have $R_i \leq e^{\lambda^2/8}$. In particular, $-\lambda p_i + \ln(p_i e^\lambda + 1 - p_i) \leq \lambda^2/8$, which one can prove by Taylor expansion around 0. Let's not do that here....

Assuming this, we now get an overall bound of $e^{n\lambda^2/8-\lambda\delta n}$. Setting $\lambda = 4\delta$ to minimize this, we get $e^{n2\delta^2-4\delta^2 n} = e^{-2\delta^2 n}$. ∎

Randomized routing/rounding

Given a directed graph and a set of pairs $\{(s_i, t_i)\}$ we want to route these pairs to minimize the maximum congestion. This problem is NP-hard. Can we find an approximate solution?

Idea: (Raghavan & Thompson)

1. Solve fractionally. Think of as multi-commodity flow (e.g., allow s_i to route to t_i by sending 1/2 down one path, 1/4 down another path, and 1/4 down another). Can solve with linear programming: for each (directed) edge e, and each commodity i, have variable X_{ei}. Constraints for inflow = outflow. Constraints $\forall e, \sum_i X_{ie} \leq C$, and minimize C.

2. For each pair (s_i, t_i) we have a flow. Now what we do is view these fractional values as probabilities and select a path such that the probability we pick edge e is equal to the flow of this commodity on e. How can we do this algorithmically? (Give proof that greedy approach works.)

Analysis: fix some edge. Let f_i be the flow of commodity i on this edge. This also means that f_i is the probability that we picked this edge for routing (s_i, t_i). So, for a given edge, can think of $\{0, 1\}$ random variables X_i corresponding to event that we picked this edge for commodity i, where $\mathbf{Pr}[X_i = 1] = f_i$. For a given edge, these X_i are all INDEPENDENT. (Not independent for the same i across different edges, but that's OK). Expected value of sum is at most C. Now apply Chernoff.

$$\mathbf{Pr}[total > (1 + \epsilon)C] < e^{-\epsilon^2 C/3}$$

The point now is if this is small enough (e.g., $o(1/n^2)$) then the probability that *there exists* an edge whose congestion exceeds this bound is also small ($o(1)$).

So, if $C \gg \log(n)$, then w.h.p., maximum is only $1 + \epsilon$ times larger than the expectation.

What if $C = 1$, or C is constant? In this case, we can apply the bound:

$$Pr[total > kC] < (e^{k-1}/k^k)^C$$

So, set k to be $O(\log(n)/\log\log(n))$, and then get $1/poly(n)$.