15-859(D) Randomized Algorithms
Notes for 9/24/98

```
* useful probabilistic inequalities
* Start on randomized rounding
```

Useful probabilistic inequalities

Say we have a random variable X. 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 inequality 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 r.v., and $\mu = \mathbf{E}[S]$. 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:

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

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

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

Here is a somewhat intuitive proof, for the case of a fair random walk. Very similar to proof in book, but with more intuitive interpretation.

Theorem 1 *Let* $X = X_1 + \ldots + X_n$ *with* $\mathbf{Pr}[X_i = 1] = \mathbf{Pr}[X_i = -1] = 1/2$*, and* X_i *mutually independent. Then*

$$\mathbf{Pr}[X > \lambda\sqrt{n}] < e^{-\lambda^2/2}$$

for $\lambda > 0$*.*

Proof. Let's look at a multiplicative version of the random walk. Let's say that we start at 1, and on a heads we multiply our current position by $(1 + \lambda/\sqrt{n})$ and on a tails we multiply our current position by $(1 - \lambda/\sqrt{n})$. This is not exactly fair since if you have equal numbers of heads and tails, you end up to the left. E.g., if you have one of each, then final position is $1 - \lambda^2/n$. On the other hand, what is the *expected* value of the final position?

Formally, can write the random variable Y for this walk as:

$$Y = 1 \cdot Y_1 \cdot Y_2 \cdots Y_n$$

where $\mathbf{Pr}[Y_i = 1 + \lambda/\sqrt{n}] = \mathbf{Pr}[Y_i = 1 - \lambda/\sqrt{n}] = 1/2$ and the Y_i are independent. Since we have full independence,

$$\mathbf{E}[Y] = \mathbf{E}[Y_1] \cdot \mathbf{E}[Y_2] \cdots = 1.$$

Let's now think about what Markov's inequality applied to Y tells us about our original (additive) version of the random walk. In particular, Markov tells us that

$$\mathbf{Pr}[Y > 1 \cdot e^{\lambda^2/2}] \leq e^{-\lambda^2/2}.$$

Here's what makes this interesting. On the one hand, we lose something by examining Y instead of X in that even the case $Y = 1$ corresponds to having seen more heads than tails. On the other hand, we gain something because it take only a few extra heads to pull Y up a lot since Y is multiplicative. So the question now is, how many more heads than tails does it take to pull Y above $e^{\lambda^2/2}$? (Roughly, each additional $\sqrt{n}/\lambda$ heads multiplies Y by e. So, once we have $Y = 1$ we only need $\frac{\lambda}{2}\sqrt{n}$ extra heads. The other $\frac{\lambda}{2}\sqrt{n}$ extra heads are needed to reach $Y = 1$ in the first place.)

The claim is that we've set things up so that everything works out. Specifically, say $X = \lambda\sqrt{n}$. I.e., we have $\frac{1}{2}(n - \lambda\sqrt{n})$ tails and $\frac{1}{2}(n + \lambda\sqrt{n})$ heads. Then (first pairing up the heads and tails, then looking at the extra heads) we have:

$$\begin{aligned}
Y &= (1 - \lambda^2/n)^{\frac{1}{2}(n - \lambda\sqrt{n})}(1 + \lambda/\sqrt{n})^{\lambda\sqrt{n}} \\
& \text{(and now using } (1+\epsilon)^n \approx e^{\epsilon n}\text{)} \\
&\approx e^{-(\lambda^2/n)\frac{1}{2}(n - \lambda\sqrt{n}) + (\lambda/\sqrt{n})(\lambda\sqrt{n})} \\
&\geq e^{-\lambda^2/2 + \lambda^2} \\
&= e^{\lambda^2/2}.
\end{aligned}$$

So, we're done. (Actually, the "≈" approximation is slightly off in the wrong direction above, so to be formal one would need to be more careful there.) ■

Randomized routing/rounding

Given an undirected 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 $1/poly(n)$.