Class 9: Agenda and Questions

1 Announcements

- HW4 due Friday!

2 Lecture Recap and Questions?

Questions from minilectures and pre-class quiz? (Compressed sensing; RIP; Gaussian matrices have the RIP with high probability.)

3 More practice with tail bounds; compare and contrast

For this part of the class, let's review some of the tail bounds we've seen. We'll explore when we'd like to use which.

We'll look at the following types of bounds:

1. Markov: For any non-negative random variable X, $\Pr\{X \geq t\} \leq \frac{\mathbb{E}X}{t}$.
2. Chebyshev: For any random variable X, $\Pr\{|X - \mathbb{E}X| \geq t\} \leq \frac{\mathrm{Var}(X)}{t^2}$.
3. Higher moment bounds: For any random variable X and any $c \geq 1$, $\Pr\{|X| \geq t\} \leq \frac{\mathbb{E}[|X|^c]}{t^c}$.
4. Multiplicative Chernoff bound: If $X_i \sim \mathrm{Ber}(p)$ are i.i.d. and $X = \sum_{i=1}^n X_i$, then for any $\varepsilon \in (0,1)$, $\Pr\{X \geq \mu(1+\varepsilon)\} \leq \exp(-C\varepsilon^2\mu)$ (for some constant C).
5. Additive Chernoff bound: If $X_i \sim \mathrm{Ber}(p)$ are i.i.d. and $X = \sum_{i=1}^n X_i$, then for any $\varepsilon > 0$, $\Pr\{X \geq \mu + t\} \leq \exp(-Ct^2/n)$ (for some constant C).

These are not all of the bounds we've seen (in particular, we've seen some other forms of Chernoff bounds), but hopefully this will be enough to get our intuition going.

3.1 Markov vs. Chebyshev vs. higher moments

Group Work

For this group work, define $\zeta(k) := \sum_{i=1}^{\infty} i^{-k}$. Recall that $\zeta(1) = \infty$ (that is, the sum diverges), while $\zeta(k) < \infty$ for all $k \geq 2$. (For example, $\zeta(2) = \pi^2/6$).

For this problem, it doesn't really matter what the $\zeta(k)$ are, just that they are constants (for $k \geq 2$) and that they are decreasing as $k \to \infty$.

1. Let X be a random variable so that, for each $i \geq 1$,

$$\Pr[X = i] = \frac{1}{\zeta(3)} \cdot \frac{1}{i^3}.$$

   - What is $\mathbb{E}X$?
   - Which of our tail bounds apply to an expression like $\Pr[X \geq t] \leq$ _____?
   - For all the tail bounds that apply, what do they say about how to fill in the blank? Use asymptotic notation as $t \to \infty$. (e.g., your answer should be like $O(1/t)$ or $O(1/t^2)$ or $e^{-\Omega(t)}$ or something like that).

2. Let X be a random variable so that, for each $i \geq 1$,

$$\Pr[X = i] = \frac{1}{\zeta(4)} \cdot \frac{1}{i^4}.$$

   Answer the same questions as in the previous part.

3. Let X be a random variable so that, for each $i \geq 1$,

$$\Pr[X = i] = \frac{1}{\zeta(k)} \cdot \frac{1}{i^k}.$$

   Answer the same questions as in the previous part.

4. What do you take away about Markov vs Chebyshev vs higher moment bounds? (e.g., when would you want to use one vs the others?)

Group Work: Solutions

1. $\mathbb{E}X = \sum_i i \cdot \frac{1}{i^3} \cdot \frac{1}{\zeta(3)} = \frac{\zeta(2)}{\zeta(3)}$.
2. Markov's inequality applies. None of the others do, since

$$\mathbb{E}X^2 = \sum_i i^2 \cdot \frac{1}{i^3} \cdot \frac{1}{\zeta(3)} = \zeta(1)/\zeta(3) = \infty.$$

   (and similarly $\mathbb{E}X^k = \infty$ for any $k \geq 2$). Thus, the variance and higher moments are all infinite, so Chebyshev and higher moment bounds tell us that $\Pr[X \geq t] \leq \infty$, which isn't very useful. Chernoff doesn't apply since X is not the sum of independent random variables.

If we apply Markov, we get

$$\Pr[X \geq t] \leq \frac{\mathbb{E}X}{t} = \frac{\zeta(2)/\zeta(3)}{t} = O(1/t).$$

3. $\mathbb{E}X = \frac{1}{\zeta(4)} \sum_i i \cdot \frac{1}{i^4} = \zeta(3)/\zeta(4)$. Now both Markov and Chebyshev (and the moment bound for $k = 2$, which is similar to Chebyshev apply, and none of the others do. Markov says

$$\Pr[X \geq t] \leq \frac{\zeta(3)/\zeta(4)}{t} = O(1/t).$$

Chebyshev says

$$\Pr[X \geq t] \leq \Pr[|X - \mu| \geq t - \mu] \leq \frac{\mathrm{Var}(X)}{(t-\mu)^2}$$

where $\mu = \mathbb{E}X = \zeta(3)/\zeta(4)$. We have

$$\mathbb{E}[X^2] = \frac{1}{\zeta(4)} \sum_i i^2 \frac{1}{i^4} = \frac{\zeta(2)}{\zeta(4)}.$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X] = \frac{\zeta(2) - \zeta(3)}{\zeta(4)}.$$

So Chebyshev says

$$\Pr[X \geq t] \leq \frac{(\zeta(2) - \zeta(3))/\zeta(4)}{(t - \zeta(3)/\zeta(4))^2} = O(1/t^2).$$

The second moment bound says that

$$\Pr[X \geq t] \leq \frac{\mathbb{E}X^2}{t^2} = \frac{\zeta(2)/\zeta(4)}{t^2} = O(1/t^2).$$

4. We have the same type of thing as in the previous problem: the moment bounds apply up to $k - 2$, and we have

$$\mathbb{E}[X^j] = \frac{1}{\zeta(k)} \sum_i i^{j-k} = \frac{\zeta(k-j)}{\zeta(k)}.$$

Thus, the j'th moment bound, for $j \leq k-2$, says

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X^j]}{t^j} = \frac{(\zeta(k-j))/\zeta(k)}{t^j} = O(1/t^j).$$

Thus, the best of these is when $j = k-2$, and we get

$$\Pr[X \geq t] = O(1/t^{k-2}).$$

5. Markov applies more generally than Chebyshev, which applies more generally than higher-moment bounds for $k \geq 2$. In the asymptotic regime we're thinking about here (e.g., $\mathbb{E}[X^k]$ is either some constant or ∞, and $t \to \infty$ is growing), higher moments will give you better results when they apply.

3.2 Chernoff vs Chebyshev

Group Work

For this group work, say that $X_i \sim \mathrm{Ber}(p)$ are i.i.d., and let $X = \sum_{i=1}^n X_i$.

1. Suppose that $p = 1/4$, and let $\varepsilon \in (0,1)$. What do each of the Chernoff bounds, and Chebyshev's inequality, say about $\Pr[X \geq \mu(1+\varepsilon)]$, where $\mu = \mathbb{E}X = pn$? Give your answer in big-O notation as $n \to \infty$ and $\varepsilon \to 0$. (e.g., an expression like $O(1/(n \cdot \varepsilon))$ or $e^{-\Omega(n/\varepsilon)}$).
2. Suppose that $p = 1/4$. How big does ε have to be (asymptotically, in terms of n) to prove a statement like

$$\Pr[X \geq \mu(1+\varepsilon)] \leq 0.001?$$

What do each {Chebyshev, multiplicative chernoff, additive chernoff} say? What if you want a statement like

$$\Pr[X \geq \mu(1+\varepsilon)] \leq \delta,$$

for some very small parameter $\delta \to 0$ that should show up in your asymptotic bound on ε?
3. Same two questions but now for $p = 1/\sqrt{n}$.
4. What take-aways do you get about when it's a good idea to use Chebyshev vs each type of Chernoff?

Group Work: Solutions

1.
$$\Pr[X \geq \mu(1+\varepsilon)] \leq \begin{cases} \exp(-C\varepsilon^2 n) & \text{mult. Chernoff} \\ \exp(-C(\mu^2\varepsilon^2)/n) = \exp(-C\varepsilon^2 n) & \text{add. Chernoff} \\ \frac{\mathrm{Var}(X)}{\varepsilon^2\mu^2} = \frac{n\cdot(3/16)}{\varepsilon^2 n^2/16} = O(\frac{1}{\varepsilon^2 n}). & \text{Chebyshev} \end{cases}$$

2. For 0.001, all of the bounds say the same thing, which is that we should take $\varepsilon = O(1/\sqrt{n})$.

   However, if we replace 0.001 with δ, Chebyshev says $\varepsilon = O(1/\sqrt{n\delta})$, and both Chernoff bounds say $\varepsilon = \sqrt{\log(1/\delta)/n}$. Thus, the Chernoff bounds have a much better dependence on δ (since $\log(1/\delta) = o(1/\delta)$ as $\delta \to 0$).

3. If $p = 1/\sqrt{n}$, then $\mu = \sqrt{n}$. We have:

$$\Pr[X \geq \mu(1+\varepsilon)] \leq \begin{cases} \exp(-C\varepsilon^2\sqrt{n}) & \text{mult. Chernoff} \\ \exp(-C(\mu^2\varepsilon^2)/n) = \exp(-C\varepsilon^2) & \text{add. Chernoff} \\ \frac{\mathrm{Var}(X)}{\varepsilon^2\mu^2} = \frac{n\cdot(1/\sqrt{n})(1-1/\sqrt{n})}{\varepsilon^2 n} = O(\frac{1}{\varepsilon^2\sqrt{n}}). & \text{Chebyshev} \end{cases}$$

   To get a probability bound of 0.001 on the right hand side, multiplicative Chebyshev and Chernoff say that we should take $\varepsilon = O(n^{-1/4})$. On the other hand, additive Chernoff either says nothing (if C happens to be large), or at best says that we should take ε to be a constant, which is way bigger than $O(n^{-1/4})$.

   To get δ, additive Chernoff is still not great; multiplicative chernoff says $\varepsilon = O(\sqrt{\log 1/\delta}/n^{1/4})$; chebyshev says $\varepsilon = \frac{1}{\sqrt{\delta}n^{1/4}}$. Again, Chernoff gets a much better dependence on δ.

It looks like Chernoff and Chebyshev do about the same if we only want a constant probability of failure and only care about the asymptotics. However, if we want a small probability of failure, Chrnoff will do better (when it applies). It also looks like multiplicative Chernoff is better than additive Chernoff when p is really small.

4 Practice with Poissonization (if time)

Group Work

Suppose you are dropping n balls into n bins.

For each of the following two bounds, explain at a very high level (e.g., a non-quantitative outline) how you would use Poissonization (and de-Poissonization) to fill in the blank. (In particular, what would be your strategy for picking the Poisson parameter? Should it be a bit bigger or a bit smaller than n?)

After you've done that for both bounds, try to fill in the quantitative details if you have time.

1. As above, you are dropping n balls into n bins. Let X_i be the indicator random variable that is 1 if there are at most two balls in bin i, and zero otherwise. Thus, $X = \sum_i X_i$ is the number of bins with at most two balls in them.

   Use Poissonization to fill in the blank as best as possible, using big-Oh notation as $n \to \infty$.

$$\Pr[X \geq 0.9999n] \leq \_\_\_\_\_\_$$

2. Now let $Y_i = 1 - X_i$, so Y_i is 1 iff there are *more* than 2 balls in bin i. Let $Y = \sum_i Y_i$ be the number of bins with more than 2 balls in them.

   Use Poissonization to fill in the blank as best as possible, using big-Oh notation as $n \to \infty$.

$$\Pr[Y \geq 0.3 \cdot n] \leq \_\_\_\_\_\_$$

Group Work: Solutions

1. We want to use Poissonization, so we should probably define some new random variables, $\tilde{X}_i$, to be the analog to X_i when we drop a Poisson number of balls rather than n balls. So let's say we're going to drop $k \sim \text{Poi}(\text{something})$ balls into n bins (we need to figure out what "something" is), and let $\tilde{X}_i$ be 1 if there are at most 2 balls in bin i and 0 otherwise. Let $\tilde{X} = \sum_i \tilde{X}_i$. How should we pick [something]? Well, we hope that we'll be able to bound

$$\Pr[\tilde{X} \geq 0.9999n] \leq \_\_\_\_\_$$

   by using the niceness of the Poisson distribution, so we should hope that

$$\Pr[X \geq 0.9999n] \leq \Pr[\tilde{X} \geq 0.9999n],$$

   since that means that we'll be able to chain together our inequalities to get an inequality in the right direction. Intuitively, we'll make the probability that $\tilde{X}$ is large *bigger* if we drop fewer balls. (If we drop fewer balls, we're more likely to end up with lots of bins with ≤ 2 balls in them). So we should take [something] to be *less* than n.

   This gives us our outline:

   - Let $k \sim Poi(n/2)$ (here, we choose $n/2$ as "something less than n" and we'll see if it works out. If it doesn't work out, maybe we'll have to choose $3n/4$ or $0.999n$ or something like that).
   - Define $\tilde{X}_i$ and $\tilde{X}$ as above. Then $\tilde{X}_i$ are independent, and their mean is $\mu := \Pr[Z \leq 2]$, where $Z \sim \text{Poi}(1/2)$. We'll hope that μ is some number that's less than 0.9999.

- We can bound

$$\Pr[\tilde{X} \geq 0.9999n] \leq \Pr[\tilde{X} \geq \mu n + (0.9999 - \mu)n] \leq \exp(-C(0.9999-\mu)^2 n) = \exp(-\Omega(n))$$

  by a Chernoff bound, where C is some constant. (This is assuming that $\mu < 0.9999$, which we'll need to check when we do the quantitative part).
- By our tail bound for Poisson random variables,

$$\Pr[|k - n/2| \geq n/2] \leq 2\exp\left(\frac{-(n/2)^2}{2(n/2 + n/2)}\right) = \exp(-\Omega(n)).$$

  Thus, except with probability $\exp(-\Omega(n))$, the number of balls dropped in the Poissonized version is less than the number dropped in the original version.
- We can now write:

$$\begin{aligned}
\Pr[X \geq 0.9999n] &\leq \Pr[\tilde{X} \geq 0.9999n | k \leq n] \\
&= \frac{\Pr[\tilde{X} \geq 0.9999n \text{AND} k \leq n]}{\Pr[k \leq n]} \\
&\leq \frac{\Pr[\tilde{X} \geq 0.9999n]}{\Pr[k \leq n]} \\
&\leq \frac{\Pr[\tilde{X} \geq 0.9999n]}{1 - \exp(-\Omega(n))} \\
&\leq \Pr[\tilde{X} \geq 0.9999n] + \exp(-\Omega(n)).
\end{aligned}$$

  In the last line, we used the fact that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \leq 1 + x,$$

  where $x \leftarrow \exp(-\Omega(n))$.
- Thus, assuming that $\mu < 0.9999$, we get

$$\Pr[X \geq 0.9999n] \leq \exp(-\Omega(n)) + \exp(-\Omega(n)) = \exp(-\Omega(n)).$$

Finally, we can check to see quantitatively if this works out. We need to compute μ and make sure it's less than 0.9999. We have

$$\mu = \frac{e^{-1/2}(1/2)^0}{0!} + \frac{e^{-1/2}(1/2)^1}{1!} + \frac{e^{-1/2}(1/2)^2}{2!} = \frac{1 + 1/2 + 1/8}{\sqrt{e}} = 0.9856,$$

which is thankfully smaller than 0.9999. (If it wasn't, we could replace 1/2 with 3/4, or 7/8, or 0.9999, or whatever it needed to be so that this number was smaller.)

Note: Instead of using the conditioning argument above, we could also use a union bound. That is, we could write

$$\Pr[\tilde{X} \geq 0.9999n] \leq \exp(-\Omega(n)),$$

and then imagine an experiment where we couple $\tilde{X}$ and X in the natural way: to decide how to drop balls in the Poissonized process, if $k \leq n$, we'll drop them exactly the same as the first k balls in the non-Poissonized process. If $k > n$, we'll do something independent from the non-Poissonized process. This way, each of X and $\tilde{X}$ on their own have the right distribution, but looking at them together we see that we always have $X \geq \tilde{X}$ whenever $k \leq n$. Now there are two bad things that can happen. Either $k > n$, or $\tilde{X} \geq 0.9999n$. If neither of those happen, $X < \tilde{X} < 0.9999n$. Thus,

$$\Pr[X \geq 0.9999n] \leq \Pr[\tilde{X} \geq 0.9999n] + \Pr[k > n] \leq 2\exp(-\Omega(n)) = \exp(-\Omega(n)).$$

2. Here the outline is similar, but now we need to choose k differently. As above, let $k \sim \text{Poi}([\textit{something}])$ where [something] is TBD. Let $\tilde{Y}_i$ be 1 iff there are more than 2 balls in bin i, and 0 otherwise, and let $\tilde{Y} = \sum_i \tilde{Y}_i$. We're going to want to show that $\Pr[Y \geq 0.3n] \leq \Pr[\tilde{Y} \geq 0.3n] \leq$ ____ to make the inequalities go the way we want, which means that we want $\tilde{Y}$ to generally be *larger* than Y. That means that we should be dropping *more* balls, since the more balls we drop, the more likely we are to have many bins with at least 2 balls. Thus, let's try taking $k \sim \text{Poi}(2n)$ and see what happens. Our outline would look like:
   - Let $k \sim Poi(2n)$
   - Define $\tilde{Y}_i$ and $\tilde{Y}$ as above. Then $\tilde{Y}_i$ are independent, and their mean is $\mu := \Pr[Z > 2]$, where $Z \sim \text{Poi}(2)$. We'll hope that μ is some number that's smaller than 0.3.
   - We can bound

     $$\Pr[\tilde{Y} \geq 0.3n] \leq \Pr[\tilde{Y} \geq \mu n + (0.3 - \mu)n] \leq \exp(-C(0.3-\mu)^2 n) = \exp(-\Omega(n))$$

     by a Chernoff bound, where C is some constant. (This is assuming that $\mu < 0.3$, which we'll need to check when we do the quantitative part).
   - By our tail bound for Poisson random variables,

     $$\Pr[|k - 2n| \geq n] \leq 2\exp\left(\frac{-(n)^2}{2(n+n)}\right) = \exp(-\Omega(n)).$$

     Thus, except with probability $\exp(-\Omega(n))$, the number of balls dropped in the Poissonized version is more than the number dropped in the original version.

- We can now write:

$$\begin{aligned}\Pr[Y \geq 0.3n] &\leq \Pr[\tilde{Y} \geq 0.3n | k \geq n] \\ &= \frac{\Pr[\tilde{Y} \geq 0.3n \text{AND} k \geq n]}{\Pr[k \geq n]} \\ &\leq \frac{\Pr[\tilde{Y} \geq 0.3n]}{\Pr[k \geq n]} \\ &\leq \frac{\Pr[\tilde{Y} \geq 0.3n]}{1 - \exp(-\Omega(n))} \\ &\leq \Pr[\tilde{Y} \geq 0.3n] + \exp(-\Omega(n)).\end{aligned}$$

- Thus, assuming that $\mu < 0.3$, we get

$$\Pr[X \geq 0.3n] \leq \exp(-\Omega(n)) + \exp(-\Omega(n)) = \exp(-\Omega(n)).$$

Finally, let's do the quantitative part...is $\mu < 0.3$? We have

$$\mu = 1 - (e^{-2}2^0/0!) - (e^{-2}2^1/1!) - (e^{-2}2^2/2!) = 1 - \frac{1}{e^2}(1 + 2 + 2) = 0.323323.$$

Rats! This isn't smaller than 0.3! What to do? Well, we had a lot of slack when we picked "$2n$" as "something larger than n". Let's try picking $n \cdot 1.1$ and see what happens. You can check that our outline should go through exactly the same (as far as the asymptotic notation is concerned). This time, we get

$$\mu = 1 - (e^{-1.1}(1.1)^0/0!) - (e^{-1.1}(1.1)^1/1!) - (e^{-1.1}(1.1)^2/2!) = 1 - \frac{1}{e^{1.1}}(1 + 1.1 + (1.1)^2/2) = 0.09958.$$

Now that's < 0.3. Hooray! If we were handing this in for HW, we should now go back through our outline but replace all the $2n$'s with $1.1n$'s, and verify that $\mu < 0.3$ when we use it instead of punting it until later.