Lecture 3: Averages of RVs

Isabella Zhu

11 February 2025

§1 TLDR

The main result we will show today is as follows: let $X_1, X_2, \dots X_n$ be i.i.d samples and $\bar{X}_n$ denote their mean. Then,

$$\mathbb{P}(|X| > t) \leq 2e^{-cnt^2/\sigma^2}.$$

This is basically an exercise in Chernoff bounds. Will do this twice: once for *Hoeffding's inequality* and second time for *Bernstein's inequality*.

§2 Hoeffding's Inequality

Theorem 2.1

(Hoeffding) Let $X_1, X_2, \dots X_n$ i.i.d random variables such that $X_i \sim \text{subG}(\sigma_i^2)$ and $\mathbb{E}[X_i] = 0$. Then,

$$\mathbb{P}(|\bar{X}_n| > t) \leq 2e^{-\frac{n^2t^2}{2\sum_{i=1}^n \sigma_i^2}}.$$

In particular, if $X_i \in [a, b]$ then

$$\mathbb{P}(|\bar{X}_n| > t) \leq 2e^{-\frac{2nt^2}{(b-a)^2}}.$$

Proof. We will use Chernoff bounds. We know

$$\mathbb{E}[e^{sX_i}] \leq e^{s^2\sigma^2/2},$$

by definition of subgaussian. We will do a one sided tail bound first.

$$\begin{aligned}
\mathbb{P}(\bar{X}_n > t) &\leq \mathbb{E}[e^{s\bar{X}_n}]e^{-st} \\
&= \mathbb{E}[e^{\frac{s}{n}\sum_{i=1}^n X_i}]e^{-st} \\
&= e^{-st}\prod_{i=1}^n \mathbb{E}[e^{sX_i/n}] \\
&\leq \exp(\frac{s^2}{2n^2}\sum_{i=1}^n \sigma_i^2 - st.)
\end{aligned}$$

We optimize this in s. Taking derivative of expression and setting to 0, we get

$$s = \frac{n^2 t}{2\sum \sigma_i^2}.$$

Then we plug in s and get a lower bound of

$$\mathbb{P}(\bar{X}_n > t) \leq \exp(-\frac{n^2 t^2}{2\sum_{i=1}^n \sigma_i^2}).$$

We can compute a similar bound for $-X_i$ and then we're done by union bound.

For the part with the fixed interval, we use Hoeffding's lemma from lecture 2 and the rest is straightforward.

§2.1 Applications to Bernoulli RV

Let $X_1, X_2, \ldots X_n$ be i.i.d rvs drawn from $\text{Ber}(p)$. We remember that $\mathbb{E}[X] = p$ and $var(X_i) = p(1-p)$. In particular, $X_i \in [0, 1]$. Hoeffding tells us that

$$\mathbb{P}(|\bar{X}_n - p| > t) \leq 2e^{-2nt^2}.$$

This is like a Gaussian tail with order $O(1)$. But note that this tail bound is bad, since it doesn't take p into account (i.e. this bound works best when $p = 1/2$ and worst when $p = 0$ or $p = 1$).

Instead, we consider $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$. We have

$$\mathbb{P}(n\bar{X}_n \geq k) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}.$$

If we use Hoeffding, we get

$$\mathbb{P}(\bar{X}_n - p \geq \frac{k}{n} - p) \leq e^{-2(k-np)^2/n}.$$

This bound is still not good, as it doesn't capture the dependence on p. We can improve this by taking into account smaller variance (use Bernstein's inequality).

§3 Bernstein's Inequality

Theorem 3.1

$X_1, X_2, \ldots X_n$ i.i.d. such that $\mathbb{E}[X_i] = 0$, $|X_i| \leq b$, and $var(X_i) \leq \sigma^2$. Then, we have

$$\mathbb{P}(|X_i| > t) \leq 2e^{-\frac{nt^2}{2(\sigma^2 + tb)}}.$$

Remark 3.2. The bound can actually be improved slightly to

$$\mathbb{P}(|X_i| > t) \leq 2e^{-\frac{nt^2}{2(\sigma^2 + tb/3)}}.$$

Proof. Intuitively, we have extra info now because we also have variance. So we want to factor this into the MGF. We have

$$e^{sX} = 1 + sX + \frac{s^2X^2}{2} + \sum_{k\geq 3} \frac{s^kX^k}{k!}.$$

We have

$$\mathbb{E}[e^{sX}] \leq 1 + \frac{s^2\sigma^2}{2} + \frac{s^2\sigma^2}{2} \sum_{k\geq 3} \frac{2s^{k-2}|X|^{k-2}}{k!} \leq 1 + \frac{s^2\sigma^2}{2} + \frac{s^2\sigma^2}{2} \sum_{k\geq 1} \frac{2s^kb^k}{(k+2)!}.$$

We basically just pulled out a variance term and replaced X with b.

Simplifying this, we get

$$\mathbb{E}[e^{sX}] \leq 1 + \frac{s^2\sigma^2}{2} e^{sb}.$$

Note that $e^x \leq \frac{1}{1-x}$ for all $x < 1$. Therefore, we can further bound

$$\mathbb{E}[e^{sX}] \leq 1 + \frac{s^2\sigma^2}{2} e^{sb} \leq 1 + \frac{s^2\sigma^2}{2} \frac{1}{1-sb} \leq e^{s^2\sigma^2/2(1-sb)}, \quad \forall s < \frac{1}{b}.$$

where the last step is just Taylor ($1 + x \leq e^x$).

Finally, we plug this into a Chernoff bound

$$\begin{aligned}\mathbb{P}(\bar{X}_n > t) &\leq e^{\frac{ns^2\sigma^2}{2(1-sb)} - nst} \\ &= \exp\left(n\left[\tfrac{s^2\sigma^2}{2(1-sb)} - st\right]\right)\end{aligned}$$

Take $s = \frac{t}{\sigma^2+tb} < \frac{1}{b}$. Plugging in, we get

$$\mathbb{P}(\bar{X}_n > t) \leq \exp\left(\tfrac{-nt^2}{2(\sigma^2+tb)}\right).$$

Doing the other tail bound and union bound gives the final result.

§4 Bernstein vs. Hoeffding Tails

As a recap, we have Hoeffding

$$e^{-nt^2}$$

and Bernstein

$$e^{-nt^2/(\sigma^2+t)}$$

We are assuming $b = 1$ for both. For Bernstein to be good, want $t << \sigma^2$. If t is large, then Hoeffding is better.

§4.1 Bernoulli Revisited

Let $p = 1 - \epsilon$. Bernstein tells us

$$\mathbb{P}(\bar{X_n} - p > t) \leq e^{-nt^2/2(\epsilon+t)}.$$

Plugging in $t = \frac{\epsilon}{2}$, we get

$$\mathbb{P}(\bar{X_n} > 1 - \frac{\epsilon}{2}) \leq e^{-n\epsilon/12}$$

versus Hoeffding gives us something on the order of $e^{-n\epsilon^2}$, which is a worse bound.