Lecture 09: Hoeffding Bound Proof

Recall: Chernoff I

- Let us recall the Chernoff Bound
- Let $\mathbb{X}$ be a random variable over the samples space $\{0,1\}$ such that $\mathbb{P}[\mathbb{X}=1]=p$ and $\mathbb{P}[\mathbb{X}=0]=1-p$
- Consider n independent samples of the distribution $\mathbb{X}$. This is represented by the random variable $(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \dots, \mathbb{X}^{(n)})$.
- Our object of study is: $\mathbb{S}_{n,p} = \sum_{i=1}^{n} \mathbb{X}^{(i)}$.
- Note that $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = np$, by the linearity of expectation
- Chernoff bound states that $\mathbb{S}_{n,p}$ is significantly larger than the expected values only with an exponentially small probability

$$\mathbb{P}\left[\mathbb{S}_{n,p} - \mathbb{E}\left[\mathbb{S}_{n,p}\right] \geqslant \Delta\right] \leqslant \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\frac{\Delta}{n}, p\right)\right) \leqslant \exp(-2\Delta^2/n)$$

- Intuitively, if $\Delta = O(\sqrt{n})$, then it is highly likely that $\mathbb{P}\left[\mathbb{S}_{n,p} - \mathbb{E}\left[\mathbb{S}_{n,p}\right] \geqslant \Delta\right]$ is small (it can be any small constant). Let us call this the "radius of concentration."

Recall: Chernoff II

- Note that (1) this bound is independent of $\mathbb{E}\left[\mathbb{S}_{n,p}\right]$, and (2) the Chernoff bound hold even when p is a function of n itself.
  An Example. Suppose $p = n^{-1/3}$. Then, we have $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = np = n^{2/3}$. For this case, the radius of concentration is again $\Delta = O(\sqrt{n})$.
- We say that the Chernoff bound is "meaningful/useful" when the radius of concentration is a $o(\mathbb{E}\left[\mathbb{S}_{n,p}\right])$.
  An Example. Suppose $p = n^{-2/3}$. In this case, we have $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = np = n^{1/3}$. The radius of concentration is $O(\sqrt{n})$, which is not $o(\mathbb{E}\left[\mathbb{S}_{n,p}\right])$.

Deviation below the Expectation I

- Chernoff bound states that the probability that $\mathbb{S}_{n,p}$ exceeds $\mathbb{E}\left[\mathbb{S}_{n,p}\right]$ by Δ is at most $\exp(-2\Delta^2/n)$
- How can we state that it is also unlikely that $\mathbb{S}_{n,p}$ is lower than $\mathbb{E}\left[\mathbb{S}_{n,p}\right]$ by Δ is small?

Deviation below the Expectation II

- We are interested in

$$\mathbb{P}\left[\mathbb{S}_{n,p} - \mathbb{E}\left[\mathbb{S}_{n,p}\right] \leqslant -\Delta\right] \leqslant ?$$

- Let us introduce the random variable $\mathbb{Y} = 1 - \mathbb{X}$. Note that $\mathbb{P}\left[\mathbb{Y} = 1\right] = 1 - p$ and $\mathbb{P}\left[\mathbb{Y} = 0\right] = p$.
- Let $\mathbb{T}_{n,1-p} = \sum_{i=1}^{n} \mathbb{Y}^{(i)}$.
- Note that $\mathbb{E}\left[\mathbb{T}_{n,1-p}\right] = n(1-p)$
- We can now use Chernoff bound in the following manner

$$\begin{aligned}
\mathbb{P}\left[\mathbb{S}_{n,p} - \mathbb{E}\left[\mathbb{S}_{n,p}\right] \leqslant -\Delta\right] &= \mathbb{P}\left[(n - \mathbb{S}_{n,p}) - (n - \mathbb{E}\left[\mathbb{S}_{n,p}\right)\right] \geqslant \Delta\right] \\
&= \mathbb{P}\left[\mathbb{T}_{n,1-p} - \mathbb{E}\left[\mathbb{T}_{n,1-p}\right] \geqslant \Delta\right] \\
&\leqslant \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(1 - p + \frac{\Delta}{n}, 1 - p\right)\right) \\
&\leqslant \exp\left(-2\Delta^2/n\right)
\end{aligned}$$

Hoeffding Bound I

- Let $(\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_n)$ be independent random variables such that $\mathbb{X}_i$ is over the sample space $[a_1, b_i]$
- We study the random variable $\mathbb{S}_n = \sum_{i=1} \mathbb{X}_i$
- We are interested in the probability

$$\mathbb{P}\left[\mathbb{S}_n - \mathbb{E}\left[\mathbb{S}_n\right] \geqslant \Delta\right] \leqslant ?$$

- **Think:** Without loss of generality we can assume that $\mathbb{E}\left[\mathbb{X}_i\right] = 0$. **Why?**
- Hoeffding's bound states that

$$\mathbb{P}\left[\mathbb{S}_n - \mathbb{E}\left[\mathbb{S}_n\right] \geqslant \Delta\right] \leqslant \exp\left(-\frac{\Delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Hoeffding Bound II

- **Think:** The following two results suffice to prove the Hoeffding's bound using the technique that we used to prove the Chernoff bound.

Lemma

Let $\mathbb{X}$ be a random variable over the sample space $[a, b]$ such that $\mathbb{E}[\mathbb{X}] = 0$. For any $h > 0$, we have

$$\mathbb{E}\left[\exp(h\mathbb{X})\right] \leqslant \frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb)$$

Lemma (Hoeffding's Lemma)

For $a < 0 < b$, we have

$$\frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb) \leqslant \exp(h^2(b-a)^2/8)$$

Hoeffding Bound III

- Next, we prove these two lemmas

Proof of the First Lemma I

- **Goal.** Let $\mathbb{X}$ be a random variable over the sample space $[a, b]$ such that $\mathbb{E}[\mathbb{X}] = 0$. For any $h > 0$, we have

$$\mathbb{E}\left[\exp(h\mathbb{X})\right] \leqslant \frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb)$$

- In the lecture, we proved the underlying intuition for this result. Here, we discuss how to formalize that proof intuition.
- Consider $x \in [a, b]$ (remember a is a negative real number)
- We want to compute p and q such that $pa + qb = x$ and $p + q = 1$. Note that $p = \frac{b-x}{b-a}$ and $q = \frac{x-a}{b-a}$ is the solution.
- By Jensen's we have

$$p\exp(ha) + q\exp(hb) \geqslant \exp(p \cdot ha + q \cdot hb) = \exp(hx)$$

Proof of the First Lemma II

- Therefore, we can write the following inequality

$$\frac{b-\mathbb{X}}{b-a}\exp(ha)+\frac{\mathbb{X}-a}{b-a}\exp(hb)\geqslant\exp(h\mathbb{X})$$

- Taking expectations both sides, we get

$$\begin{aligned}
&\mathbb{E}\left[\frac{b-\mathbb{X}}{b-a}\exp(ha)+\frac{\mathbb{X}-a}{b-a}\exp(hb)\right]\geqslant\mathbb{E}\left[\exp(h\mathbb{X})\right]\\
\iff&\frac{b-\mathbb{E}\left[\mathbb{X}\right]}{b-a}\exp(ha)+\frac{\mathbb{E}\left[\mathbb{X}\right]-a}{b-a}\exp(hb)\geqslant\mathbb{E}\left[\exp(h\mathbb{X})\right]\\
\iff&\frac{b}{b-a}\exp(ha)-\frac{a}{b-a}\exp(hb)\geqslant\mathbb{E}\left[\exp(h\mathbb{X})\right]
\end{aligned}$$

And, we are done!

Proof of the Second Lemma (Hoeffding's Lemma) I

- **Goal.** For $a < 0 < b$, we have

$$\frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb) \leqslant \exp(h^2(b-a)^2/8)$$

Or, equivalently

$$\ln\left(\frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb)\right) \leqslant h^2(b-a)^2/8$$

- We shall use the following variable substitution $u = h(b-a)$
- Consider the following simplification

$$\begin{aligned}
&\frac{b}{b-a}\exp(ha) - \frac{a}{b-a}\exp(hb) \\
&= \exp(ha)\left(\frac{b}{b-a} - \frac{a}{b-a}\exp(h(b-a))\right) \\
&= \exp(ha)\left(1 + \frac{a}{b-a} - \frac{a}{b-a}\exp(h(b-a))\right)
\end{aligned}$$

Proof of the Second Lemma (Hoeffding's Lemma) II

- We use the following substitution: $\theta = \frac{-a}{b-a}$. Substituting the value of u, we get $\theta = (-a)/(u/h) \iff ah = -\theta u$.
- So, we get

$$\exp(ha)\left(1 + \frac{a}{b-a} - \frac{a}{b-a}\exp(h(b-a))\right) = \exp(-\theta u)(1 - \theta + \theta\exp(u))$$

- Taking ln, our **goal** is to prove the following statement

$$f_\theta(u) := -\theta u + \ln(1 - \theta + \theta\exp(u)) \leqslant u^2/8$$

- We shall use Taylor's remainder theorem on $f_\theta(u)$

Proof of the Second Lemma (Hoeffding's Lemma) III

- Note that

$$
\begin{aligned}
f_\theta(u) &= -\theta u + \ln(1-\theta+\theta\exp(u)) \\
f'_\theta(u) &= -\theta + \frac{\theta\exp(u)}{1-\theta+\theta\exp(u)} \\
f''_\theta(u) &= \frac{\theta\exp(u)}{1-\theta+\theta\exp(u)} - \frac{\theta^2\exp(2u)}{\left(1-\theta+\theta\exp(u)\right)^2} \\
&= t(1-t) \leqslant 1/4,
\end{aligned}
$$

where $t = \frac{\theta\exp(u)}{1-\theta+\theta\exp(u)}$.

Proof of the Second Lemma (Hoeffding's Lemma) IV

- So, we get

$$f_\theta(u) = f_\theta(0) + f'_\theta(0)u + f''_\theta(v)u^2/2,$$

for some $v \in [0, u]$. That is,

$$f_\theta(u) = 0 + 0u + f''_\theta(v)u^2/2 \leqslant u^2/8$$

This step completes the proof of the lemma.

Extra-credit Problem

We ended the lecture with a discussion of providing a alternate/tighter proof for Hoeffding's Lemma.