Lecture 09: Hoeffding Bound Proof
Let us recall the Chernoff Bound

Let $X$ be a random variable over the samples space $\{0, 1\}$ such that $P[X = 1] = p$ and $P[X = 0] = 1 - p$

Consider $n$ independent samples of the distribution $X$. This is represented by the random variable $(X(1), X(2), \ldots, X(n))$.

Our object of study is: $S_{n,p} = \sum_{i=1}^{n} X(i)$.

Note that $E[S_{n,p}] = np$, by the linearity of expectation.

Chernoff bound states that $S_{n,p}$ is significantly larger than the expected values only with an exponentially small probability

$$P[S_{n,p} - E[S_{n,p}] \geq \Delta] \leq \exp(-nD_{KL}(p + \frac{\Delta}{n}, p)) \leq \exp(-2\Delta^2/n)$$

Intuitively, if $\Delta = O(\sqrt{n})$, then it is highly likely that $P[S_{n,p} - E[S_{n,p}] \geq \Delta]$ is small (it can be any small constant). Let us call this the “radius of concentration.”
Note that (1) this bound is independent of $\mathbb{E}[S_{n,p}]$, 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 $\mathbb{E}[S_{n,p}] = np = n^{2/3}$. 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 $o(\mathbb{E}[S_{n,p}])$.

**An Example.** Suppose $p = n^{-2/3}$. In this case, we have $\mathbb{E}[S_{n,p}] = np = n^{1/3}$. The radius of concentration is $O(\sqrt{n})$, which is not $o(\mathbb{E}[S_{n,p}])$. 

Concentration
Chernoff bound states that the probability that \( S_{n,p} \) exceeds \( \mathbb{E}[S_{n,p}] \) by \( \Delta \) is at most \( \exp(-2\Delta^2/n) \).

How can we state that it is also unlikely that \( S_{n,p} \) is lower than \( \mathbb{E}[S_{n,p}] \) by \( \Delta \) is small?
Deviation below the Expectation II

- We are interested in
  \[ P \left[ S_{n,p} - \mathbb{E} \left[ S_{n,p} \right] \leq -\Delta \right] \leq ? \]

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

\[
P \left[ S_{n,p} - \mathbb{E} \left[ S_{n,p} \right] \leq -\Delta \right] = P \left[ (n - S_{n,p}) - (n - \mathbb{E} [S_{n,p}]) \geq \Delta \right] \\
= P \left[ T_{n,1-p} - \mathbb{E} [T_{n,1-p}] \geq \Delta \right] \\
\leq \exp \left( -nD_{KL} \left( 1 - p + \frac{\Delta}{n}, 1 - p \right) \right) \\
\leq \exp \left( -2\Delta^2 / n \right)
\]
Let \((X_1, X_2, \ldots, X_n)\) be independent random variables such that \(X_i\) is over the sample space \([a_i, b_i]\).

We study the random variable \(S_n = \sum_{i=1}^{n} X_i\).

We are interested in the probability

\[
P[S_n - E[S_n] \geq \Delta] \leq ?
\]

**Think:** Without loss of generality we can assume that \(E[X_i] = 0\). **Why?**

Hoeffding's bound states that

\[
P[S_n - E[S_n] \geq \Delta] \leq \exp \left( -\frac{\Delta^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right)
\]
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 $X$ be a random variable over the sample space $[a, b]$ such that $E[X] = 0$. For any $h > 0$, we have

$$E[\exp(hX)] \leq \frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb)$$

**Lemma (Hoeffding’s Lemma)**

For $a < 0 < b$, we have

$$\frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb) \leq \exp(h^2(b-a)^2/8)$$
Next, we prove these two lemmas
Goal. Let $X$ be a random variable over the sample space $[a, b]$ such that $\mathbb{E}[X] = 0$. For any $h > 0$, we have

$$\mathbb{E}[\exp(hX)] \leq \frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb)$$

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

$$p \exp(ha) + q \exp(hb) \geq \exp(p \cdot ha + q \cdot hb) = \exp(hx)$$
Therefore, we can write the following inequality

\[
\frac{b - X}{b - a} \exp(ha) + \frac{X - a}{b - a} \exp(hb) \geq \exp(hX)
\]

Taking expectations both sides, we get

\[
\mathbb{E} \left[ \frac{b - X}{b - a} \exp(ha) + \frac{X - a}{b - a} \exp(hb) \right] \geq \mathbb{E} \left[ \exp(hX) \right]
\]

\[
\iff \quad \frac{b - \mathbb{E}[X]}{b - a} \exp(ha) + \frac{\mathbb{E}[X] - a}{b - a} \exp(hb) \geq \mathbb{E} \left[ \exp(hX) \right]
\]

\[
\iff \quad \frac{b}{b - a} \exp(ha) - \frac{a}{b - a} \exp(hb) \geq \mathbb{E} \left[ \exp(hX) \right]
\]

And, we are done!
Proof of the Second Lemma (Hoeffding’s Lemma)

**Goal.** For \( a < 0 < b \), we have

\[
\frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb) \leq \exp\left( h^2 \frac{(b-a)^2}{8} \right)
\]

Or, equivalently

\[
\ln \left( \frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb) \right) \leq h^2 \frac{(b-a)^2}{8}
\]

We shall use the following variable substitution \( u = h(b-a) \)

Consider the following simplification

\[
\frac{b}{b-a} \exp(ha) - \frac{a}{b-a} \exp(hb) = \exp(ha) \left( \frac{b}{b-a} - \frac{a}{b-a} \exp(h(b-a)) \right)
\]

\[
= \exp(ha) \left( 1 + \frac{a}{b-a} - \frac{a}{b-a} \exp(h(b-a)) \right)
\]
We use the following substitution: \( \theta = \frac{-a}{b-a} \). Substituting the value of \( u \), we get \( \theta = \frac{-a}{u/h} \iff ah = -\theta u \).

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

\[
f_\theta(u) := -\theta u + \ln(1 - \theta + \theta \exp(u)) \leq u^2/8
\]

We shall use Taylor’s remainder theorem on \( f_\theta(u) \)
Note that

\[
f_\theta(u) = -\theta u + \ln(1 - \theta + \theta \exp(u))
\]

\[
f'_\theta(u) = -\theta + \frac{\theta \exp(u)}{1 - \theta + \theta \exp(u)}
\]

\[
f''_\theta(u) = \frac{\theta \exp(u)}{1 - \theta + \theta \exp(u)} - \frac{\theta^2 \exp(2u)}{(1 - \theta + \theta \exp(u))^2}
\]

\[
= t(1 - t) \leq 1/4,
\]

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

\[ f_\theta(u) = 0 + 0u + f''_\theta(v)u^2/2 \leq u^2/8 \]

This step completes the proof of the lemma.
Extra-credit Problem

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