Lecture 07: Chernoff Bound: Easy to Use Forms
Consider a **negatively associated** joint distribution \((X_1, X_2, \ldots, X_n)\), such that the marginal distributions \(X_i\) are over \([0, 1]\)

OR

a **negatively correlated** joint distribution \((X_1, X_2, \ldots, X_n)\), such that the marginal distributions \(X_i\) are over \(\{0, 1\}\)

Define \(p_i = \mathbb{E}[X_i]\), for \(i \in \{1, 2, \ldots, n\}\). Define

\[
p = \left( p_1 + p_2 + \cdots + p_n \right) / n
\]

Define \(S_{n,p} := X_1 + X_2 + \cdots + X_n\). Chernoff bound states that

\[
P[S_{n,p} \geq (p + \varepsilon)n] \leq \exp \left( -n \cdot D_{KL}(p + \varepsilon, p) \right)
\]

**Objective of this lecture.** We shall obtain easier-to-compute, albeit weaker, upper bounds on the probability
Chernoff Bound Proof Template: Recall

Proof for the negatively associated case.

- Define $Y_i = \text{Bern}(p_i)$, where $p_i = \mathbb{E}[X_i]$ and $i \in \{1, 2, \ldots, n\}$

$$\mathbb{P}[S_{n,p} \geq (p + \varepsilon)n] = \mathbb{P}\left[H \sum_{i=1}^{n} X_i \geq H(p+\varepsilon)n\right]$$

(for any $H > 1$)

$$\leq \frac{\mathbb{E}\left[H \sum_{i=1}^{n} X_i\right]}{H(p+\varepsilon)n}$$

(Markov inequality)

$$\leq \frac{\prod_{i=1}^{n} \mathbb{E}\left[H^{X_i}\right]}{H(p+\varepsilon)n}$$

(Neg. association & increasing property of $H(\cdot)$)

$$\leq \frac{\prod_{i=1}^{n} \mathbb{E}\left[H^{Y_i}\right]}{H(p+\varepsilon)n}$$

(Convexity property of $H(\cdot)$)

$$= \frac{\prod_{i=1}^{n} ((1 - p_i) + p_i \cdot H)}{H(p+\varepsilon)n}$$

$$\leq \frac{(1 - p) + p \cdot H)^n}{H(p+\varepsilon)n}$$

(AM-GM inequality)

Rest of the analysis is identical to the previous lecture’s analysis
We shall prove the following bound

**Theorem**

\[
\mathbb{P} \left[ S_{n,p} \geq n(p + \varepsilon) \right] \leq \exp(-nD_{KL}(p + \varepsilon, p)) \leq \exp(-2n\varepsilon^2)
\]

Comment: The upper bound is easy to compute. However, this bound does not depend on \( p \) at all.

To prove this result, it suffices to prove that

\[
D_{KL}(p + \varepsilon, p) \geq 2\varepsilon^2
\]
We shall use the Lagrange form of the Taylor approximation theorem to the following function

\[ f(\varepsilon) = D_{KL} (p + \varepsilon, p) = (p + \varepsilon) \log \frac{p + \varepsilon}{p} + (1 - p - \varepsilon) \log \frac{1 - p - \varepsilon}{1 - p} \]

Observe that \( f(0) = 0 \)

Differentiating once, we have

\[ f'(\varepsilon) = \log \frac{p + \varepsilon}{p} - \log \frac{1 - p - \varepsilon}{1 - p} \]

Observe that \( f'(0) = 0 \)

Differentiating twice, we have

\[ f''(\varepsilon) = \frac{1}{p + \varepsilon} + \frac{1}{1 - p - \varepsilon} = \frac{1}{(p + \varepsilon)(1 - p - \varepsilon)} \]
By applying the Lagrange form of Taylor's remainder theorem, we get the following result. For every $\varepsilon$, there exists $\theta \in [0, 1]$ such that

$$f(\varepsilon) = f(0) + f'(0) \cdot \varepsilon + f''(\theta \varepsilon) \cdot \frac{\varepsilon^2}{2} = f''(\theta \varepsilon) \cdot \frac{\varepsilon^2}{2}$$

Note that $f(\theta \varepsilon) = \frac{1}{(p + \theta \varepsilon)(1 - p - \theta \varepsilon)}$. We can apply the AM-GM inequality to conclude that

$$(p + \theta \varepsilon)(1 - p - \theta \varepsilon) \leq \left( \frac{(p + \theta \varepsilon) + (1 - p - \theta \varepsilon)}{2} \right)^2 = \frac{1}{4}$$

Therefore, we get that $f''(\theta \varepsilon) \geq 4$. Substituting this bound, we get

$$f(\varepsilon) = f''(\theta \varepsilon) \cdot \frac{\varepsilon^2}{2} \geq 4 \cdot \frac{\varepsilon^2}{2} = 2\varepsilon^2$$

This completes the proof.
In the previous bound, we consider the probability of $S_{n,p}$ exceeding the expected value $np$ by an additive amount $n\varepsilon$. Now, we want to explore the case when the offset is multiplicative. That is, we want to consider the probability of $S_{n,p}$ exceeding the expected value $np$ by a multiplicative amount $\lambda(np)$. We shall prove the following result.

**Theorem**

For $\lambda > 0$, we have

$$\Pr [S_{n,p} \geq np(1 + \lambda)] \leq \exp(-nD_{KL}(p(1 + \lambda), p))$$

$$\leq \exp \left( -\frac{\lambda^2}{2(1 + \lambda/3)} np \right)$$

Comment: Note that this bound depends on $p$. 

**Concentration Bounds**
Our objective is to prove

\[ D_{KL} (p(1 + \lambda), p) \geq \frac{\lambda^2}{2 (1 + \lambda/3)} \cdot p. \]

Let us expand the left-hand side expression

\[
D_{KL} (p(1 + \lambda), p) = p(1 + \lambda) \log(1 + \lambda) + (1 - p(1 + \lambda)) \log \left( \frac{1 - p(1 + \lambda)}{1 - p} \right)
\]
We will approximate the expression with the underbrace. For brevity, let us substitute \( p' = p + \lambda p \). The expression becomes

\[
(1 - p') \log \left( \frac{1 - p'}{1 - p} \right) = -(1 - p') \log \frac{1 - p}{1 - p'}
\]

\[
= - \log \left( 1 + \frac{\lambda p}{1 - p'} \right)^{1-p'}
\]

\[
\geq -\lambda p.
\]

The final inequality follows from the fact that \((1 + x) \leq \exp(x)\).
Substituting, this simplification, we have

\[ D_{KL}(p(1 + \lambda), p) \geq (1 + \lambda)p \log(1 + \lambda) - \lambda p. \]

If we prove the following claim then we are done.

**Claim**

\[ (1 + \lambda) \log(1 + \lambda) - \lambda \geq \frac{\lambda^2}{2 (1 + \lambda/3)}. \]

Proving this claim is left as an exercise.
• We have always been looking at the probability that the sum \( S_{n,p} \) significantly exceeds the expected value of the sum. We shall now consider the probability that the sum is \( S_{n,p} \) is significantly lower than the expected value of the sum.

• We can apply the Chernoff bound of the r.v. \( 1 - X_i \) and get the following result

\[
P [ S_{n,p} \leq n(p - \varepsilon) ] = P [ n - S_{n,p} \geq n(1 - p + \varepsilon) ] 
\leq \exp(-nD_{KL}(1 - p + \varepsilon, 1 - p))
\]

By using the first form of our bounds that we studied today, we can conclude that

\[
P [ S_{n,p} \leq n(p - \varepsilon) ] \leq \exp(-nD_{KL}(1 - p + \varepsilon, 1 - p)) \leq \exp(-2n\varepsilon^2)
\]
We are, however, interested in obtaining a bound where the deviation is multiplicative. That is,

\[ P[S_{n,p} \leq np(1 - \lambda)] \leq ?? \]

where \( 1 > \lambda > 0 \).

We shall prove the following bound:

**Theorem**

*For \( 1 > \lambda > 0 \), we have*

\[ P[S_{n,p} \leq np(1 - \lambda)] \leq \exp(-nD_{KL}(1 - p(1 - \lambda), 1 - p)) \leq \exp(-\lambda^2 np/2) \]
We shall proceed just like the proof of the “second form.” It suffices to prove that

\[ D_{KL} (1 - p(1 - \lambda), 1 - p) \geq \lambda^2 p/2 \]

Let us expand and write \( D_{KL} (1 - p(1 - \lambda), 1 - p) \) as follows

\[
(1 - p(1 - \lambda)) \log \frac{1 - p(1 - \lambda)}{1 - p} + p(1 - \lambda) \log(1 - \lambda)
\]
Note that

\[(1 - p(1 - \lambda)) \log \frac{1 - p(1 - \lambda)}{1 - p} = - (1 - p(1 - \lambda)) \log \frac{1 - p}{1 - p(1 - \lambda)}\]

\[= - (1 - p(1 - \lambda)) \log \left(1 - \frac{\lambda p}{1 - p(1 - \lambda)}\right)\]

\[\geq - (1 - p(1 - \lambda)) \cdot \left(\frac{\lambda p}{1 - p(1 - \lambda)}\right) = \lambda p\]

The last inequality is from the fact that \(1 - x \leq \exp(-x)\) for all \(x \geq 0\). (Comment: Since there is a negative sign in front, the inequality is in the opposite direction when substituted)
Substituting this result, we get that

$$D_{KL}(1 - p(1 - \lambda), 1 - p) \geq \lambda p + p(1 - \lambda) \log(1 - \lambda)$$

So, it suffices to prove that

$$\lambda p + p(1 - \lambda) \log(1 - \lambda) \geq \lambda^2 p/2$$

Or, equivalently, we need to prove that

$$\lambda + (1 - \lambda) \log(1 - \lambda) \geq \lambda^2 / 2$$
To prove this inequality, we will proceed by Lagrange form of the Taylor’s remainder theorem on the function
\[ f(x) = (1 - x) \log(1 - x). \]

\[ f(x) = (1 - x) \log(1 - x), \quad f'(x) = - \log(1 - x) - 1 \]
\[ f''(x) = \frac{1}{1 - x}, \quad f'''(x) = \frac{1}{(1 - x)^2} \geq 0. \]

Therefore, we have
\[
f(\lambda) = f(0) + f'(0)\lambda + f''(0)\lambda^2/2 + f'''(\theta \lambda)\lambda^3/6 \\
\geq f(0) + f'(0)\lambda + f''(0)\lambda^2/2 \\
= 0 - \lambda + \lambda^2/2,
\]

which completes the proof.
To conclude, let us summarize the results that we derived today.

**Theorem**

The random variables $X_1, \ldots, X_n$ are negatively correlated and $0 \leq X_i \leq 1$. Let $S_{n,p} := X_1 + \cdots + X_n$, where $p := (\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]) / n$. Then, the following results hold

1. For $\varepsilon \geq 0$, we have
   \[
   P[S_{n,p} \geq n(p + \varepsilon)] \leq \exp(-2n\varepsilon^2), \quad \text{and} \quad P[S_{n,p} \leq n(p - \varepsilon)] \leq \exp(-2n\varepsilon^2)
   \]

2. For $\lambda \geq 0$, we have
   \[
   P[S_{n,p} \geq np(1 + \lambda)] \leq \exp(-\lambda^2 np/2(1 + \lambda/3))
   \]

3. For $1 > \lambda \geq 0$, we have
   \[
   P[S_{n,p} \leq np(1 - \lambda)] \leq \exp(-\lambda^2 np/2)
   \]

**Concentration Bounds**
Appendix: Chernoff for Sampling without Replacement

- Suppose there are $R$ red balls and $B$ blue balls in an urn.
- Draw $n$ samples without replacement.
- Let $(X_1, X_2, \ldots, X_n)$ be the random variables such that $X_i$ indicates whether the $i$-th draw is a red ball or not.
- Prove that $E[X_i] = E[X_1] = R/(R + B)$.
- Our objective is to show that the random variables are negatively correlated. That is, the covariance of $X_1$ and $X_2$ is negative. Toward that objective, prove

$$E[X_i \cdot X_j] = E[X_1 \cdot X_2] \leq E[X_1] \cdot E[X_2] = E[X_i] \cdot E[X_j]$$