Lecture 05: Chernoff Bound: An Introduction

Introduction

- Let $\mathbb{X}$ represent the $\mathrm{Bern}(p)$ random variable
- Let $\mathbb{X}^{(1)}, \ldots, \mathbb{X}^{(n)}$ represent n independent and identical copies of the random variable $\mathbb{X}$
- Let $\mathbb{S}_n := \mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(n)}$ represent the sum of these n random variables
- Observe that $\mathbb{S}_n$ is a random variable over the sample space $\{0, 1, \ldots, n\}$ with mean $\mathbb{E}[\mathbb{S}_n] = np$
- For example, if $\mathbb{X}$ represents a coin-toss, then $\mathbb{S}_n$ is a random variable representing the number of observed Heads when n coin-tosses are performed
- How does the random variable $\mathbb{S}_n$ concentrate around its mean? What is the probability of $\mathbb{S}_n$ to be "far" from the expected value?

Analysis using Markov Bound

- One can use Markov bound to deduce

$$\mathbb{P}\left[\mathbb{S}_n \geqslant \lambda \cdot (np)\right] \leqslant \frac{1}{\lambda}.$$

- Can we do better?

Analysis using Chebyshev's Inequality I

- By Chebyshev's Inequality, we have

$$\mathbb{P}\left[|\mathbb{S}_n - np| \geqslant t\right] \leqslant \frac{\mathrm{Var}\left[\mathbb{S}_n\right]}{t^2}.$$

Analysis using Chebyshev's Inequality II

- What is $\mathrm{Var}[\mathbb{S}_n]$?

$$
\begin{aligned}
\mathrm{Var}[\mathbb{S}_n] &= \mathbb{E}\left[\mathbb{S}_n^2\right] - \left(\mathbb{E}[\mathbb{S}_n]\right)^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} \mathbb{X}^{(i)}\right)^2\right] - (np)^2 \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{X}^{(i)2} + \sum_{i \neq j} \mathbb{X}^{(i)}\mathbb{X}^{(j)}\right] - n^2p^2 \\
&= n \cdot \mathbb{E}\left[\left(\mathbb{X}^{(1)}\right)^2\right] + n(n-1) \cdot \mathbb{E}\left[\mathbb{X}^{(1)}\mathbb{X}^{(2)}\right] - n^2p^2 \\
&= n \cdot p + n(n-1) \cdot p^2 - n^2p^2 = n \cdot p(1-p).
\end{aligned}
$$

Analysis using Chebyshev's Inequality III

- Think: The probability of $\mathbb{S}_n$ being $\Theta\left(\sqrt{np(1-p)}\right)$ far from the mean is at most a constant.
- Think: Can we use higher moments to get better bounds?
- Think: Let $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ be a joint distribution and $\mathbb{S}_n = \sum_{i=1}^n \mathbb{X}_i$. Suppose the marginals $\mathbb{X}_i = \mathrm{Bern}\,(p)$ and the random variables $\mathbb{X}_i$ and $\mathbb{X}_j$ are *pair-wise independent* when $j \neq i$. Can we still apply this estimation technique?

A Large Deviation Bound

Observe that

$$\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] = \sum_{i=k}^{n} \binom{n}{i} \cdot p^i (1-p)^{n-i}.$$

Claim

$$\binom{n}{k} \cdot p^k (1-p)^{n-k} \leqslant \mathbb{P}\left[\mathbb{S}_n \geqslant k\right] \leqslant \binom{n}{k} \cdot p^k.$$

- Think: How to prove this claim?
- Think: For what values of p and k is the upper bound meaningful? Hint: Use Stirling's formula.
- Think: When $p = 1/2$, for what values of k is the upper bound < 1?

Using Stirling's Approximation I

- Our objective is to study the expression

$$\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] = \sum_{i=k}^{n} \binom{n}{i} \cdot p^i (1-p)^{n-i}.$$

- It is obvious that this expression is at least the term $\binom{n}{k} \cdot p^k (1-p)^{n-k}$.
- In homework, we lower-bounded this term by

$$\frac{1}{\sqrt{8np'(1-p')}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right),$$

where $p' = k/n$ and $\mathrm{D}_{\mathrm{KL}}(a, b) = a \ln\left(\frac{a}{b}\right) + (1-a)\ln\left(\frac{1-a}{1-b}\right)$ represents the Kullback–Leibler divergence

Using Stirling's Approximation II

- Therefore, to obtain a tight upper bound of the original expression, we should aim to obtain an upper bound that is in terms of $\exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right)$
- Towards this objective, we prove the following upper bound on the j-th term of the summation

Claim

For $j \geqslant 0$, we have

$$\binom{n}{k+j} \cdot p^{k+j}(1-p)^{n-k-j} \leqslant \rho^j \cdot \binom{n}{k} \cdot p^k(1-p)^{n-k},$$

where $\rho = \frac{1-p'}{p'} \cdot \frac{p}{1-p}$ and $p' = k/n$.

Think: Why is $\rho < 1$, if $p' = k/n > p$?

Using Stirling's Approximation III

- Therefore, we have

Theorem

$$\begin{aligned}\mathbb{P}\left[\mathbb{S}_n \geqslant k\right] &< \frac{1}{1-\rho} \cdot \binom{n}{k} \cdot p^k(1-p)^{n-k} \\ &\leqslant \frac{1}{1-\rho} \cdot \frac{1}{\sqrt{2\pi np'(1-p')}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right) \\ &= \frac{p'(1-p)}{(p'-p)} \cdot \frac{1}{\sqrt{2\pi np'(1-p')}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p', p\right)\right),\end{aligned}$$

where $\rho = \frac{1-p'}{p'} \cdot \frac{p}{1-p}$ *and* $p' = k/n$.

The first inequality follows from the claim above. The second inequality is an estimation using the Stirling's approximation, which we proved in homework.

Using Stirling's Approximation IV

- Observe that if p' is a constant $> p$, then
  1. The lower and the upper bounds are within a constant factor of each other!
  2. The probability is exponentially decreasing in n.
- The conclusions are summarized in the next result

Using Stirling's Approximation V

Lemma (Conclusions)

Let $\mathbb{S}_n = \mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(n)}$*, where* $\mathbb{X} = \mathrm{Bern}(p)$*.*

1. $$\mathbb{P}[\mathbb{S}_n \geqslant k] \leqslant \binom{n}{k} p^k.$$

2. $$1 \leqslant \frac{\mathbb{P}[\mathbb{S}_n \geqslant k]}{\binom{n}{k} p^k (1-p)^{n-k}} \leqslant \frac{1}{1-\rho},$$

   where $\rho = \frac{1-p'}{p'} \cdot \frac{p}{1-p}$ *and* $p' = k/n$*.*

3. $$\binom{n}{k} p^k (1-p)^{n-k} = \frac{\Theta(1)}{\sqrt{np'(1-p')}} \cdot \exp\left(-n \cdot \mathrm{D}_{\mathrm{KL}}(p', p)\right),$$

   where $\mathrm{D}_{\mathrm{KL}}(a, b) = a \ln\left(\frac{a}{b}\right) + (1-a) \ln\left(\frac{1-a}{1-b}\right)$ *and* $p' = k/n$*.*