Lecture 08: Chernoff and Hoeffding Bound + Hypergeometric Distribution

Recall I

- Last lecture we were deriving the Chernoff Bound
- Let $\mathbb{X}$ be a Bernoulli distribution with mean p. That is $\mathbb{X}$ is a random variable over the sample space $\{0, 1\}$ such that the probability $\mathbb{P}[\mathbb{X} = 1] = p$ and $\mathbb{P}[\mathbb{X} = 0] = (1 - p)$.
- Let $(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \dots, \mathbb{X}^{(n)})$ be n independent and identical samples of the random variable $\mathbb{X}$
- Our object of study if the random variable

$$\mathbb{S}_{n,p} := \mathbb{X}^{(1)} + \cdots + \mathbb{X}^{(n)}$$

This random variable is over the sample space $\{0, 1, \dots, n\}$ and we have $\mathbb{P}\left[\mathbb{S}_{n,p} = i\right] = \binom{n}{i} p^i (1 - p)^{n-i}$. This distribution is also referred to as the binomial distribution $B_{n,p}$.

Recall II

- The expected value of $\mathbb{S}_{n,p}$ is $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = np$ by linearity of expectation
- We are interested in finding whether it is possible for the random variable $\mathbb{S}_{n,p}$ to deviate far from the expected value or not.
- Chernoff bound states that

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+t)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+t, p\right))$$

  That is, the Chernoff bound states that the probability of exceeding the mean by nt, for constant t, is exponentially small in n

Recall III

- Let us now recall the steps that were involved in deriving the Chernoff Bound
- The first observation was that, for any $h > 0$, we have

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+t)\right] = \mathbb{P}\left[\exp\left(h\mathbb{S}_{n,p}\right) \geqslant \exp(hn(p+t))\right]$$

The goal is to consider "all moments of the random variable $\mathbb{S}_{n,p}$ suitably weighted." The identity is a result of the fact that $\exp(h\cdot)$ is a monotonically increasing function for positive h.

- Then, we applied Markov to obtain the upper bound

$$\leqslant \frac{\mathbb{E}\left[\exp\left(h\mathbb{S}_{n,p}\right)\right]}{\exp(hn(p+t))}$$

We emphasize that this is the only place we shall apply an inequality. The tightness of the final bound is solely dependent on the tightness of this inequality!

Recall IV

- Next, we observe that the expectation

$$\mathbb{E}\left[\exp(hS_{n,p})\right] = \mathbb{E}\left[\exp(h\mathbb{X}^{(1)})\right] \cdots \mathbb{E}\left[\exp(h\mathbb{X}^{(n)})\right] = \left(\mathbb{E}\left[\exp(h\mathbb{X})\right]\right)^n$$

The first equality relies on the fact that the random variables $\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)}$ are independent. The final equality relies on the fact that the random variables $\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)}$ are identical to $\mathbb{X}$. Based on this observation, the upper-bound evaluates to the following

$$= \left(\frac{(1-p) + p\exp(h)}{\exp(h(p+t))}\right)^n$$

Recall V

- This bound holds for all positive h. We choose $h = h^*$ that minimizes the quantity $\left(\frac{(1-p)+p\exp(h)}{\exp(h(p+t))}\right)$. By basic calculus, we obtain

$$\exp(h^*) = \frac{(1-p)(p+t)}{p(1-p-t)}$$

- Substituting this value of $h = h^*$ in the upper-bound, we get

$$= \left(\left(\frac{p+t}{p}\right)^{p+t}\left(\frac{1-p-t}{1-p}\right)^{1-p-t}\right)^{-n} = \exp(-n\mathrm{D}_{\mathrm{KL}}(p+t, p))$$

- This completes the proof that

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+t)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}(p+t, p))$$

Recall VI

- This bound is still not easy to work. We shall derive bounds that are easier to calculate

- **Problem 1.** Suppose we are given a coin that outputs heads with probability p, and outputs tails with probability $(1-p)$. Can we estimate p accurately?
- Our algorithm is the following. We toss the coin n times and count the number of heads $\widetilde{n}$. Then, we output $\widetilde{p} = \frac{\widetilde{n}}{n}$ as an estimate of the quantity p.
- What is the probability that we are accurate? Chernoff bound states that

$$\mathbb{P}\left[S_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right))$$

So, the probability that $\widetilde{p} \geqslant p+\varepsilon$ is

$$\leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right))$$

Applications II

- **Problem 2.** Suppose we have a randomized algorithm that correctly decides whether $x \in L$ or not, for some language L, with probability 0.75. Can we construct another algorithm that correctly decides whether $x \in L$ or not with probability $1 - 2^{-k}$, for any $k \geqslant 2$?
- Hint: Run the algorithm a large number of times and take a majority of the outcome. Use Chernoff bound to analyze the algorithm.

Goal of today's Lecture I

- We will obtain a more "easy-to-evaluate" upper-bound for Chernoff Bound
- We shall generalize the Chernoff bound in two orthogonal directions to obtain two different bounds
  - Concentration of the Hypergeometric Distribution, and
  - Hoeffding Bound

Goal of today's Lecture II

- **Hypergeometric Distribution.** Let us establish a new way to interpreting the random variables $(\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)})$. Suppose we have a box of N balls. Among them pN are red and $(1-p)N$ are blue. The random variable $\mathbb{X}^{(1)}$ is a the random variable corresponding to the experiment of drawing a random ball from this box and checking whether the balls is red or not. Then, we replace the ball back in the box. Now, the random variable $\mathbb{X}^{(2)}$ corresponds to the drawing a random balls from the box and checking whether it is red or not.
  And, so on...
- In the hypergeometric distribution, we have $N \geqslant n$ and we perform the same experiment as above except that we do not replace the balls back into the bin!

Goal of today's Lecture III

- **Hoeffding Bound.** Instead of considering $(\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)})$ we consider independent random variables $(\mathbb{X}_1, \dots, \mathbb{X}_n)$ such that each $\mathbb{X}_i$ is a distribution over the sample space $[a_i, b_i]$ and, overall, we have $\mathbb{E}\left[\mathbb{X}_1 + \cdots + \mathbb{X}_n\right] = np$.

Easy to Calculate Chernoff Bound I

- Our goal is to find an easy to evaluate upper bound of

$$\exp(-n\mathrm{D}_{\mathrm{KL}}(p+t,p))$$

- This goal is equivalent to finding an easy to evaluate lower-bound of

$$\mathrm{D}_{\mathrm{KL}}(p+t,p) = (p+t)\ln\frac{p+t}{p} + (1-p-t)\ln\frac{1-p-t}{1-p} =: g(t)$$

- We do this by using Taylor expansion of $g(t)$ around $t=0$. Let us start with this process.

Easy to Calculate Chernoff Bound II

- Note that $g(0) = 0$
- Let us differentiate and obtain $g^{(1)}(t)$

$$g^{(1)}(t) = \ln \frac{p+t}{p} + 1 - \ln \frac{1-p-t}{1-p} - 1$$
$$= \ln \frac{p+t}{p} - \ln \frac{1-p-t}{1-p}$$

- Note that $g^{(1)}(0) = 0$
- Let us differentiate once more and obtain $g^{(2)}(t)$

$$g^{(2)}(t) = \frac{1}{p+t} + \frac{1}{1-p-t}$$

- So, we get $g^{(2)}(0) = \frac{1}{p(1-p)}$. Okay, so we got something non-negative. We shall truncate at the next term in the Taylor's Expansion.

Easy to Calculate Chernoff Bound III

- Let us differentiate once more and obtain $g^{(3)}(t)$

$$g^{(3)}(t) = -\frac{1}{(p+t)^2} + \frac{1}{(1-p-t)^2}$$

Easy to Calculate Chernoff Bound IV

- We shall see a few bounds. Let us expand till 2-terms. There exists $c \in [0, t]$ such that

$$
\begin{aligned}
g(t) &= g(0) + g^{(1)}(0)t + g^{(2)}(c)\frac{t^2}{2} \\
&= \frac{t^2}{2(p+c)(1-p-c)} \\
&\geqslant 2t^2, && \text{(by AM-GM Inequality)}
\end{aligned}
$$

- This bound is not sensitive to p. Let us get a bound sensitive to p. We consider 3-terms in the expansion now. For some $c \in [0, t]$, we have.

$$
\begin{aligned}
g(t) &= g(0) + g^{(1)}(0)t + g^{(2)}(0)\frac{t^2}{2} + g^{(3)}(c)\frac{t^2}{6} \\
&= \frac{t^2}{2p(1-p)} + g^{(3)}(c)\frac{t^2}{6}
\end{aligned}
$$

Easy to Calculate Chernoff Bound V

If $p \geqslant 1/2$, then we have $g^{(3)}(c) \geqslant 0$ and, consequently,

$$g(t) \geqslant \frac{t^2}{2p(1-p)}$$

- Let us summarize

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

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+t)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}(p+t,p)) \leqslant \exp\left(-n\frac{t^2}{2p(1-p)}\right)$$

when $p \geqslant 1/2$

Take a look at the graph at desmos
For a bound for all p, go one more term in the Taylor expansion.

Hypergeometric Distribution

- Recall the hypergeometric distribution. We are given N balls in a bin. There are pN red balls and $(1-p)N$ blue balls. We sample n balls from the bin without replacement. Let the samples be $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$. We are interested in the probability that we see $n(p+t)$ red balls.
- **Crucial Observation.** Pause just after time j (i.e., just after picking j balls).
  1. If you have p red balls in $(\mathbb{X}_1, \ldots, \mathbb{X}_j)$ then the probability that $\mathbb{X}_{j+1}$ is a red ball is p.
  2. If you have $< p$ red balls in $(\mathbb{X}_1, \ldots, \mathbb{X}_j)$ then the probability that $\mathbb{X}_{j+1}$ is a red ball is $> p$.
  3. If you have $> p$ red balls in $(\mathbb{X}_1, \ldots, \mathbb{X}_j)$ then the probability that $\mathbb{X}_{j+1}$ is a red ball is $< p$.
- **Conclusion.** The hypergeometric series pushes the "sum" towards the mean. So, it is more concentrated than the binomial distribution $B(n,p)$!
- Using coupling argument this intuition can be formalized.

Hoeffding Bound I

- **First Change.** Let us assume that the mean of each $\mathbb{X}_i$ is 0. This assumption is justified because we can consider the random variable $\mathbb{Y}_i = \mathbb{X}_i - \mathbb{E}[\mathbb{X}_i]$ instead. This simplification will make several of the mathematical expressions less cumbersome.
  Now, given the assumption that $\mathbb{E}[\mathbb{X}_i] = 0$ for all $i \in \{1, \ldots, n\}$, we have $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = 0$ as well. So, we are interested in bounding the probability

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant nt\right]$$

- **Second Change.** We have to upper-bound $\mathbb{E}\left[\exp(h\mathbb{X}_i)\right]$ given the fact that $\mathbb{X}_i$ is over the sample space $[a_i, b_i]$ and $\mathbb{E}\left[\mathbb{X}_i\right] = 0$. How do we proceed further?
  Let us use Hoeffding's Lemma

Lemma (Hoeffding's Lemma)

Let $\mathbb{X}_i$ be a r.v. over the sample space $[a_i, b_i]$ with $\mathbb{E}\left[\mathbb{X}_i\right] = 0$. Then, the following holds

$$\mathbb{E}\left[\exp(h\mathbb{X}_i)\right] \leqslant \exp\left(\frac{h^2(b_i - a_i)^2}{8}\right)$$

Hoeffding Bound III

- Then you should be able to complete the proof of Hoeffding Bound.

$$\mathbb{P}\left[\mathbb{S}_{n_p} \geqslant n(p+t)\right] \leqslant \exp\left(-\frac{2t^2n^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$