Lecture 10: Chernoff Bound: Generalizations

Overview I

Our objective is to generalize the Chernoff Bound that we proved in the previous lecture. Let us first recall the Chernoff bound result that we proved.

- Let $\mathbb{X}$ be a r.v. over $\{0,1\}$ such that $\mathbb{P}[\mathbb{X}=0]=1-p$ and $\mathbb{P}[\mathbb{X}=1]=p$
- Let $\mathbb{S}_{n,p}=\mathbb{X}^{(1)}+\mathbb{X}^{(2)}+\cdots+\mathbb{X}^{(n)}$
- Chernoff bound states that

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

Overview II

We shall generalize this result in two ways

1. For $1 \leqslant i \leqslant n$, let $\mathbb{X}_i$ be a r.v. over $\{0,1\}$ such that $\mathbb{P}[\mathbb{X}_i = 0] = 1 - p_i$ and $\mathbb{P}[\mathbb{X}_i = 1] = p_i$. Each $\mathbb{X}_i$ is independent of the other $\mathbb{X}_j$s. Let $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p = (p_1 + \cdots + p_n)/n$.
2. For $1 \leqslant i \leqslant n$, let $\mathbb{X}_i$ be a r.v. over $[0,1]$ such that $\mathbb{E}[\mathbb{X}_i] = p_i$.

Despite these two generalizations, the following bound continues to hold true.

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

First Generalization I

- Let $X_1, X_2, \ldots \mathbb{X}_n$ be independent random variables such that $\mathbb{P}[\mathbb{X}_i = 0] = 1 - p_i$ and $\mathbb{P}[\mathbb{X}_i = 1] = p_i$, for $1 \leqslant i \leqslant n$.
- Let $p := (p_1 + p_2 + \cdots + p_n)/n$
- Define $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$
- We bound the following probability. For any $H > 1$, we have

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] = \mathbb{P}\left[H^{\mathbb{S}_{n,p}} \geqslant H^{n(p+\varepsilon)}\right]$$

- Now, we apply the Markov inequality

$$\mathbb{P}\left[H^{\mathbb{S}_{n,p}} \geqslant H^{n(p+\varepsilon)}\right] \leqslant \frac{\mathbb{E}\left[H^{\mathbb{S}_{n,p}}\right]}{H^{n(p+\varepsilon)}} = \frac{\mathbb{E}\left[H^{\sum_{i=1}^{n} \mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}}$$

First Generalization II

- Since, each $\mathbb{X}_i$ are independent of other $\mathbb{X}_j$s, we have

$$\frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} = \frac{\prod_{i=1}^{n} 1 - p_i + p_i H}{H^{n(p+\varepsilon)}}$$

- We apply the AM-GM inequality to conclude that

$$\prod_{i=1}^{n} 1 - p_i + p_i H \leqslant \left(\frac{\sum_{i=1}^{n} 1 - p_i + p_i H}{n}\right)^n$$

Equality holds if and only if all $p_i = p$. This bound can now be substituted to conclude

$$\frac{\mathbb{E}\left[\prod_{i=1}^{n} H^{\mathbb{X}_i}\right]}{H^{n(p+\varepsilon)}} \leqslant \left(\frac{1 - p + pH}{H^{p+\varepsilon}}\right)^n$$

First Generalization III

- This is identical to the bound that we had in the Chernoff bound proof. We can use the following choice of H in the bound above to obtain the tightest possible bound

$$H^* = \frac{(p+\varepsilon)(1-p)}{p(1-p-\varepsilon)}$$

So, we get the bound

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

Second Generalization I

- Let $1 \leqslant \mathbb{X}_i \leqslant 1$ be a r.v. such that $\mathbb{E}[\mathbb{X}_i] = p_i$ and each $\mathbb{X}_i$ is independent of other $\mathbb{X}_j$s
- Just like the previous setting, we have $\mathbb{S}_{n,p} = \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p = (p_1 + p_2 + \cdots + p_n)/n$
- Note that if we prove the following bound, then we shall be done

$$\mathbb{E}\left[H^{\mathbb{X}_i}\right] \leqslant 1 - p_i + p_i H$$

We can use this bound in the previous proof and arrive at the identical upper-bound.

Second Generalization II

The proof follows from the following

$$
\begin{aligned}
\mathbb{E}\left[H^{\mathbb{X}_i}\right] &= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot H^x \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot H^{(1-x)\cdot 0 + x\cdot 1} \\
&\leqslant \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot \Big((1-x)\cdot H^0 + x\cdot H^1\Big) \qquad \text{, By Jensen's} \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot (1 - x + xH) \\
&= \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right] - \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot x + H \sum_{x\in[0,1]} \mathbb{P}\left[\mathbb{X}_i = x\right]\cdot x \\
&= 1 - p_i + p_i H \qquad \text{, Because } \mathbb{E}\left[\mathbb{X}_i\right] = p_i
\end{aligned}
$$

The appendix provides additional intuition for this analysis.

Appendix: Intuition for the Analysis I

- Let $\mathbb{X}$ be an r.v. over $[a, b]$ such that $\mathbb{E}[\mathbb{X}] = \mu$
- Let $f : \mathbb{R} \to \mathbb{R}$ be a concave upwards function (that is, it looks like $f(x) = x^2$)
- Jensen's inequality states that $f(\mathbb{E}[\mathbb{X}]) \leqslant \mathbb{E}\left[f(\mathbb{X})\right]$, and equality holds if and only if $\mathbb{X}$ has its entire probability mass at μ. Therefore, we can conclude that $f(\mu) \leqslant \mathbb{E}\left[f(\mathbb{X})\right]$
- So, we have a lower-bound on $\mathbb{E}\left[f(\mathbb{X})\right]$. Now, we are interested in obtaining an upper-bound on $\mathbb{E}\left[f(\mathbb{X})\right]$
- For the upper-bound note that is $\mathbb{X}$ deposits more probability mass away from μ, then $\mathbb{E}\left[f(\mathbb{X})\right]$ increases. In fact, increasing the mass further away increases $\mathbb{E}\left[f(\mathbb{X})\right]$ more. So, the maximum value of $\mathbb{E}\left[f(\mathbb{X})\right]$ is achieved when $\mathbb{X}$ deposits the entire probability mass either at a or b only. Let us find such a probability distribution under the constraint that $\mathbb{E}[\mathbb{X}] = \mu$

- Suppose $\mathbb{P}[\mathbb{X}^* = a] = p$. Then, we have $\mathbb{P}[\mathbb{X}^* = b] = 1 - p$. Further, the constraint $\mathbb{E}[\mathbb{X}^*] = \mu$ becomes $pa + (1-p)b = \mu$. Solving, we get

$$p = \frac{b-\mu}{b-a}$$

Therefore, we get $1 - p = \frac{\mu-a}{b-a}$. For this probability, we get

$$\mathbb{E}\left[f(\mathbb{X}^*)\right] = \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)$$

So, we expect the following bound to hold for a general r.v. $\mathbb{X}$

$$\mathbb{E}\left[f(\mathbb{X})\right] \leqslant \mathbb{E}\left[f(\mathbb{X}^*)\right] = \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)$$

This is not a formal proof. Let us prove this intuition formally.

- Let $\mathbb{X}$ be an r.v. over $[a, b]$ with $\mathbb{E}[\mathbb{X}] = \mu$. Note that by Jensen's inequality, we have

$$f(x) = f\left(\frac{b-x}{b-a}a + \frac{x-a}{b-a}b\right) \leqslant \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b)$$

  Now, we take expectation on both sides to conclude that

$$\begin{aligned}\mathbb{E}\left[f(\mathbb{X})\right] &\leqslant \mathbb{E}\left[\frac{b-\mathbb{X}}{b-a}f(a) + \frac{\mathbb{X}-a}{b-a}f(b)\right] \\ &= \frac{b-\mathbb{E}[\mathbb{X}]}{b-a}f(a) + \frac{\mathbb{E}[\mathbb{X}]-a}{b-a}f(b) \\ &= \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)\end{aligned}$$

- To conclude, we have the following bound.

$$f(\mu) \leqslant \mathbb{E}\left[f(\mathbb{X})\right] \leqslant \frac{b-\mu}{b-a}f(a) + \frac{\mu-a}{b-a}f(b)$$