Lecture 07: Chernoff Bound: Easy to Use Forms

Recall

- Recall that $1 \leqslant \mathbb{X}_i \leqslant 1$ are independent random variables, for $1 \leqslant i \leqslant n$. Let $p_i = \mathbb{E}[\mathbb{X}_i]$, for $1 \leqslant i \leqslant n$. Define $\mathbb{S}_{n,p} := \mathbb{X}_1 + \mathbb{X}_2 + \cdots + \mathbb{X}_n$, where $p := (p_1 + \cdots + p_n)/n$.
- Chernoff bound states that

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

- **Objective of this lecture.** We shall obtain easier to compute, albeit weaker, upper bounds on the probability

First Form I

- We shall prove the following bound

Theorem

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

- 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

$$\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right) \geqslant 2\varepsilon^2$$

First Form II

- We shall use the Lagrange form of the Taylor approximation theorem to the following function

$$f(\varepsilon) = \mathrm{D}_{\mathrm{KL}}(p+\varepsilon, p) = (p+\varepsilon)\log\frac{p+\varepsilon}{p} + (1-p-\varepsilon)\log\frac{1-p-\varepsilon}{1-p}$$

  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

$$f''(\varepsilon) = \frac{1}{p+\varepsilon} + \frac{1}{1-p-\varepsilon} = \frac{1}{(p+\varepsilon)(1-p-\varepsilon)}$$

First Form III

- By applying the Lagrange form of the Taylor's remainder theorem, we get the following result. For every ε, 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

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

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

$$f(\varepsilon) = f''(\theta\varepsilon) \cdot (\varepsilon^2/2) \geqslant 4 \cdot (\varepsilon^2/2) = 2\varepsilon^2$$

This completes the proof.

Second Form I

- In the previous bound, we consider the probability of $\mathbb{S}_{n,p}$ 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 $\mathbb{S}_{n,p}$ exceeding the expected value np by a multiplicative amount $\lambda(np)$. We shall prove the following result

Theorem

For $\lambda > 0$, we have

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant np(1+\lambda)\right] \leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p(1+\lambda), p\right)) \leqslant \exp\left(-\frac{\lambda^2}{2(1+\lambda/3)}np\right)$$

- Comment: Note that this bound depends on p.

Second Form II

- Our objective is to prove

$$\mathrm{D_{KL}}\left(p(1+\lambda), p\right) \geqslant \frac{\lambda^2}{2\left(1+\lambda/3\right)} \cdot p.$$

- Let us expand the left-hand side expression

$$\begin{aligned}&\mathrm{D_{KL}}\left(p(1+\lambda), p\right)\\ &= p(1+\lambda)\log(1+\lambda) + \underbrace{(1-p(1+\lambda))\log\left(\frac{1-p(1+\lambda)}{1-p}\right)}\end{aligned}$$

Second Form III

- We will approximate the expression with the underbrace. For brevity, let us substitute $p' = p + \lambda p$. The expression becomes

$$\begin{aligned}(1-p')\log\left(\frac{1-p'}{1-p}\right) &= -(1-p')\log\frac{1-p}{1-p'}\\ &= -\log\left(1+\frac{\lambda p}{1-p'}\right)^{1-p'}\\ &\geqslant -\lambda p.\end{aligned}$$

  The final inequality follows from the fact that $(1+x) \leqslant \exp(x)$.

Second Form IV

- Substituting, this simplification, we have

$$\mathrm{D}_{\mathrm{KL}}\left(p(1+\lambda), p\right) \geqslant (1+\lambda)p\log(1+\lambda) - \lambda p.$$

If we prove the following claim then we are done.

Claim

$$(1+\lambda)\log(1+\lambda) - \lambda \geqslant \frac{\lambda^2}{2\left(1+\lambda/3\right)}.$$

Proving this claim is left as an exercise.

Third Form I

- We have always been looking at the probability that the sum $\mathbb{S}_{n,p}$ significantly exceeds the expected value of the sum. We shall now consider the probability that the sum is $\mathbb{S}_{n,p}$ is significantly lower than the expected value of the sum.
- We can apply the Chernoff bound of the r.v. $1 - \mathbb{X}_i$ and get the following result

$$\begin{aligned}\mathbb{P}\left[\mathbb{S}_{n,p} \leqslant n(p-\varepsilon)\right] &= \mathbb{P}\left[n - S_{n,p} \geqslant n(1-p+\varepsilon)\right] \\ &\leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(1-p+\varepsilon, 1-p\right))\end{aligned}$$

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

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

Third Form II

- We are, however, interested in obtaining a bound where the deviation is multiplicative. That is,

$$\mathbb{P}\left[\mathbb{S}_{n,p} \leqslant np(1-\lambda)\right] \leqslant ??$$

  where $1 > \lambda > 0$.
- We shall prove the following bound

Theorem

For $1 > \lambda > 0$, we have

$$\begin{aligned}\mathbb{P}\left[\mathbb{S}_{n,p} \leqslant np(1-\lambda)\right] &\leqslant \exp(-n\mathrm{D}_{\mathrm{KL}}\left(1-p(1-\lambda), 1-p\right)) \\ &\leqslant \exp(-\lambda^2 np/2)\end{aligned}$$

Third Form III

- We shall proceed just like the proof of the "second form." It suffices to prove that

$$\mathrm{D}_{\mathrm{KL}}\left(1-p(1-\lambda), 1-p\right) \geqslant \lambda^2 p/2$$

- Let us expand and write $\mathrm{D}_{\mathrm{KL}}\left(1-p(1-\lambda), 1-p\right)$ as follows

$$(1-p(1-\lambda))\log\frac{1-p(1-\lambda)}{1-p} + p(1-\lambda)\log(1-\lambda)$$

Third Form IV

Note that

$$
\begin{aligned}
&(1-p(1-\lambda))\log\frac{1-p(1-\lambda)}{1-p}\\
&=-(1-p(1-\lambda))\log\frac{1-p}{1-p(1-\lambda)}\\
&=-(1-p(1-\lambda))\log\left(1-\frac{\lambda p}{1-p(1-\lambda)}\right)\\
&\geqslant-(1-p(1-\lambda))\cdot\left(-\frac{\lambda p}{1-p(1-\lambda)}\right)=\lambda p
\end{aligned}
$$

The last inequality is from the fact that $1-x\leqslant\exp(-x)$ for all $x\geqslant 0$. (Comment: Since there is a negative sign in front, the inequality is in the opposite direction when substituted)

Third Form V

- Substituting this result, we get that

$$\mathrm{D_{KL}}\left(1 - p(1-\lambda), 1 - p\right) \geqslant \lambda p + p(1-\lambda)\log(1-\lambda)$$

So, it suffices to prove that

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

Or, equivalently, we need to prove that

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

Third Form VI

- To prove this inequality, we will proceed by Lagrange form of the Taylor's remainder theorem on the function $f(x) = (1-x)\log(1-x)$.

$$f(x) = (1-x)\log(1-x), \qquad f'(x) = -\log(1-x) - 1$$
$$f''(x) = \frac{1}{1-x}, \qquad f'''(x) = \frac{1}{(1-x)^2} \geqslant 0.$$

Therefore, we have

$$\begin{aligned} f(\lambda) &= f(0) + f'(0)\lambda + f''(0)\lambda^2/2 + f'''(\theta\lambda)\lambda^3/6 \\ &\geqslant f(0) + f'(0)\lambda + f''(0)\lambda^2/2 \\ &= 0 - \lambda + \lambda^2/2, \end{aligned}$$

which completes the proof.

Conclusion

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

Theorem

The random variables $\mathbb{X}_1, \dots, \mathbb{X}_n$ *are independent and* $0 \leqslant \mathbb{X}_i \leqslant 1$. *Let* $\mathbb{S}_{n,p} := \mathbb{X}_1 + \cdots + \mathbb{X}_n$. *Furthermore, we define* $p := (\mathbb{E}[\mathbb{X}_1] + \cdots + \mathbb{E}[\mathbb{X}_n])/n$. *Then, the following results hold*

1. *For* $\varepsilon \geqslant 0$, *we have*

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \exp(-2n\varepsilon^2), \textit{ and}$$
$$\mathbb{P}\left[\mathbb{S}_{n,p} \leqslant n(p-\varepsilon)\right] \leqslant \exp(-2n\varepsilon^2)$$

2. *For* $\lambda \geqslant 0$, *we have*

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant np(1+\lambda)\right] \leqslant \exp(-\lambda^2 np/2(1+\lambda/3))$$

3. *For* $1 > \lambda \geqslant 0$, *we have*

$$\mathbb{P}\left[\mathbb{S}_{n,p} \leqslant np(1-\lambda)\right] \leqslant \exp(-\lambda^2 np/2)$$

Appendix: Chernoff for Sampling without Replacement I

- Recall the discuss in the class regarding sampling without replacement
- Our claim was "sampling without replacement" is more concentrated than "sampling with replacement"
- This appendix is intended to help you think more about this problem
- Let $\mathbb{X}$ be the first sample and $\mathbb{Y}$ represent the second sample
- Suppose $\mathbb{X} = \mathrm{Bern}(p)$
- Suppose $(\mathbb{Y}|\mathbb{X} = 0) = \mathrm{Bern}(p_0)$ and $(\mathbb{Y}|\mathbb{X} = 1) = \mathrm{Bern}(p_1)$ and we have $\mathbb{E}[\mathbb{Y}] = p$
- Recall that if the first sample is 0 then the expectation of the next sample increase. And, if the first sample is 1 then the expectation of the next sample reduces. Therefore, we must have $p_0 \geqslant p_1$. Suppose $p_0 = p + \delta$ and $p_1 = p - \varepsilon$.

Appendix: Chernoff for Sampling without Replacement II

- Since, $\mathbb{E}[\mathbb{Y}] = p$, we have the constraint that

$$(1-p)\delta = p\varepsilon.$$

- Now, our objective is the prove the claim

Claim

For any $H \geqslant 1$, we have

$$(1-p)\left((1-p_0)+p_0 H\right)+pH\left((1-p_1)+p_1 H\right) \leqslant \left((1-p)+pH\right)^2.$$

- Think: How to prove this result?
- Think: Does the inequality still hold if $p_0 < p_1$?
- Think: Does the inequality hold when $\mathbb{E}[\mathbb{Y}] \neq \mathbb{E}[\mathbb{X}]$?