Lecture 09: Chernoff Bound

Problem Introduction I

- Let $\mathbb{X}$ be a coin that outputs 1 with probability p, and outputs 0 with probability $1 - p$. The exact probability p is not known. Our objective is to estimate the probability p.
- Informally, our strategy is to toss this coin (independently) n times and report the fraction of outcomes that were heads. We want to understand the probability that this estimate is far from the real value of p.
- Let $\mathbb{X}^{(1)}, \mathbb{X}^{(2)}, \dots, \mathbb{X}^{(n)}$ be n independent coin tosses that are identically distributed as the random variable $\mathbb{X}$
- We are interested in studying the random variable

$$\mathbb{S}_{n,p} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} + \cdots + \mathbb{X}^{(n)}$$

This random variable $\mathbb{S}_{n,p}$ represents the total number of heads in the n coin tosses.

Problem Introduction II

- Formally, given $\varepsilon > 0$, we are interested in computing the probability that

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

That is, we are interested to prove that the probability of our estimate being "much larger" than p is small.

Approach using Stirling's Approximation I

- Suppose we have seen i heads. We can explicitly compute the probability that $\mathbb{S}_{n,p} = i$. as follows There are $\binom{n}{i}$ ways to choose the coins that turn up heads. The probability that these coins turn up heads is p^i. Moreover, the probability that the remaining coins turn up tails is $(1-p)^{n-i}$. So, we can claim the following

$$\mathbb{P}\left[\mathbb{S}_{n,p} = i\right] = \binom{n}{i} p^i (1-p)^{n-i}$$

- We can use this result to compute our desired probability as follows

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] = \sum_{i \geqslant n(p+\varepsilon)} \binom{n}{i} p^i (1-p)^{n-i}$$

Approach using Stirling's Approximation II

- For simplicity, let us assume that $n(p+\varepsilon)=k$ is an integer
- **Upper-bound.** We can prove that the maximum element $\binom{n}{i}p^i(1-p)^{n-i}$, where $i \geqslant k$, is achieved at $i=k$. We can use this observation to upper-bound the probability expression.

Approach using Stirling's Approximation III

$$
\begin{aligned}
\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] &= \sum_{i \geqslant k} \binom{n}{i} p^i (1-p)^{n-i} \\
&\leqslant \sum_{i \geqslant k} \binom{n}{k} p^k (1-p)^{n-k} \\
&= (n-k) \binom{n}{k} p^k (1-p)^{n-k} \\
&\leqslant \frac{n-k}{\sqrt{2\pi n (p+\varepsilon)(1-p-\varepsilon)}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right)\right) \\
&= \sqrt{\frac{n-k}{2\pi(p+\varepsilon)}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right)\right)
\end{aligned}
$$

Basically, this bound proves that

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

Approach using Stirling's Approximation IV

- **Lower-bound.** We can prove a lower bound by using the fact that "the probability of observing $\geqslant k$ heads" is more than "the probability of observing k heads."

$$\begin{aligned}
\mathbb{P}\left[\mathbb{S}_{n,p} = n(p+\varepsilon)\right] &> \mathbb{P}\left[\mathbb{S}_{n,p} = k\right] \\
&= \binom{n}{k} p^k (1-p)^{n-k} \\
&\geqslant \frac{1}{\sqrt{8\pi n(p+\varepsilon)(1-p-\varepsilon)}} \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right)\right)
\end{aligned}$$

  Basically, this bound proves that

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] = \Omega(1/\sqrt{n}) \exp\left(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right)\right)$$

- **Conclusion.** The upper and the lower-bounds can be combined to conclude that $\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right]$ is $\mathrm{poly}(n) \exp(-n\mathrm{D}_{\mathrm{KL}}\left(p+\varepsilon, p\right))$.

Chernoff Bound: Proof I

- Let us now upper bound the probability $\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right]$ using the Chernoff bound. The upper-bound will be slightly better than what we obtained using the Stirling approximation.
- Recall that $\mathbb{X}$ is a r.v. over the sample space $\{0,1\}$. Moreover, we have $\mathbb{P}[\mathbb{X}=1]=p$ and $\mathbb{P}[\mathbb{X}=0]=1-p$. Note that we have $\mathbb{E}[\mathbb{X}]=p$.
- We are studying the r.v.

$$\mathbb{S}_{n,p} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} + \cdots + \mathbb{X}^{(n)}$$

  Each random variable $\mathbb{X}^{(i)}$ is an independent copy of the random variable $\mathbb{X}$.
- Note that we have $\mathbb{E}\left[\mathbb{S}_{n,p}\right] = n\mathbb{E}[\mathbb{X}] = np$, by linearity of expectation

Chernoff Bound: Proof II

Theorem (Chernoff Bound)

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

Before we proceed to proving this result, let us interpret this theorem statement. Suppose $p = 1/2$ and $t = 1/4$. Then, it is exponentially unlikely that $\mathbb{S}_{n,p}$ surpasses $n(1/2 + 1/4) = 3n/4$

Chernoff Bound: Proof III

Let us begin with the proof.

- We are interested in upper-bounding the probability

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

- Note that, for any positive h, we have

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

  The exact value of h will be determined later. The intuition of using the $\exp(\cdot)$ function is to consider all the moments of $\mathbb{S}_{n,p}$

- Now, we apply Markov inequality to obtain

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

- Now, we need an observation. Suppose $\mathbb{A}$ and $\mathbb{B}$ are two independent random variables. Then, we have $\mathbb{E}\left[\exp(\mathbb{A}+\mathbb{B})\right] = \mathbb{E}\left[\exp(\mathbb{A})\right] \cdot \mathbb{E}\left[\exp(\mathbb{B})\right]$. We emphasize that $\mathbb{A}$ and $\mathbb{B}$ have to be independent to apply this result.
- Note that we have $\mathbb{S}_{n,p} = \sum_{i=1}^{n} \mathbb{X}^{(i)}$. So, we can apply the previous observation iteratively to obtain the following result.

$$\frac{\mathbb{E}\left[\exp(h\mathbb{S}_{n,p})\right]}{\exp(hn(p+\varepsilon))} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[\exp(h\mathbb{X}^{(i)})\right]}{\exp(hn(p+\varepsilon))} = \left(\frac{\mathbb{E}\left[\exp(h\mathbb{X})\right]}{\exp(h(p+\varepsilon))}\right)^{n}$$

- Recall that $\mathbb{X}$ is a random variable such that $\mathbb{P}\left[\mathbb{X}=0\right] = 1-p$ and $\mathbb{P}\left[\mathbb{X}=1\right] = p$. So, the random variable $\exp(h\mathbb{X})$ is such that $\mathbb{P}\left[\exp(h\mathbb{X})=1\right] = 1-p$ and $\mathbb{P}\left[\exp(h\mathbb{X})=\exp(h)\right] = p$. Therefore, we can conclude that

$$\mathbb{E}\left[\exp(h\mathbb{X})\right] = (1-p)\cdot 1 + p\cdot\exp(h) = 1-p+p\exp(h)$$

Chernoff Bound: Proof V

- Substituting this value, we get

$$\left(\frac{\mathbb{E}\left[\exp(h\mathbb{X})\right]}{\exp(h(p+\varepsilon))}\right)^n = \left(\frac{1-p+p\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$

- So, let us take a pause at this point and recall that what we have proven thus far. We have shown that, for all positive h, the following bound holds

$$\mathbb{P}\left[\mathbb{S}_{n,p} \geqslant n(p+\varepsilon)\right] \leqslant \left(\frac{1-p+p\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$

Chernoff Bound: Proof VI

- To obtain the tightest upper-bound we should use the value of $h = h^*$ that minimizes the right-hand size expression. For simplicity let us make a variable substitution $H = \exp(h)$. Let us define

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

Our objective is to find $H = H^*$ that minimizes $f(H)$.

- Let us compute $f'(H)$ and solve for $f'(H^*) = 0$. Note that we have

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

The solution $f'(H^*) = 0$ is given by

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

We can check that, for $\varepsilon > 0$, we have $H^* > 1$, that is, $h > 0$. We can consider the second derivative $f''(H)$ to prove that this extremum is a minima.
Instead of computing $f''(H)$, we can use a shortcut technique. We know that at H^*, the function $f(H)$ either has a maximum or a minimum. Moreover, there is only one extremum of the function $f(H)$. Note that $\lim_{H\to\infty} f(H) = \infty$, so $f(H^*)$ must be a minimum.

Chernoff Bound: Proof VIII

- Now, let us substitute the value of h^* to obtain

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