Lecture 06: Concentration Bounds (Recall, Applications, Generalizations)

Recall: Proof Outline of Chernoff-Bound

- Let $\mathbb{X}$ be a Bernoulli random variable that is 1 with probability p; otherwise it is 0 with probability $(1 - p)$.
- Let $\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(n)}$ are n-independent copies of the random variable $\mathbb{X}$
- We define $\mathbb{S}_n = \sum_{i=1}^{n} \mathbb{X}^{(i)}$
- Note that $\mathbb{E}[\mathbb{S}_n] = np$, using linearity of expectation
- We want to understand how small the following expression is:

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

- The first step is: For any $h > 0$, we have

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

  We use the fact that the function $\exp(hx)$ is monotonically increasing for $h > 0$

- We apply Markov to the right hand side to obtain

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

- Using the independence of $\mathbb{X}_1, \dots, \mathbb{X}_n$, we get

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \frac{\mathbb{E}\left[\exp(h\mathbb{S}_n)\right]}{\exp(hn(p+\varepsilon))} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[\exp(h\mathbb{X}_i)\right]}{\exp(hn(p+\varepsilon))}$$

Recall: Proof Outline of Chernoff-Bound III

- Note that $\mathbb{E}\left[\exp(h\mathbb{X})\right] = 1 - p + p\exp(h)$. So, we get the bound:

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

- The above statement is true for all $h > 0$. We set $h = h^*$ for which the right hand side expression is minimized. Note that the expression is minimized for

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

- For this setting, we get

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

- Using Taylor expansion around $\varepsilon_0 = 0$, we can show that $2^{-\mathrm{D}_{\mathrm{KL}}(p+\varepsilon,p)}$ expression is $\leqslant \exp(-2\varepsilon^2)$
- Overall, we get

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

Some Observations

- What is the probability $\mathbb{P}\left[\mathbb{S}_n \leqslant n(p-\varepsilon)\right]$? Hint: Apply the previous bound for $\mathbb{Y} = 1 - \mathbb{X}$ random variable.
- What is the probability $\mathbb{P}\left[\left|\mathbb{S}_n - pn\right| \geqslant n\varepsilon\right]$?

Application: Estimating Bias of a Coin

Suppose there is a coin with unknown bias p. How will you estimate p within an additive error ε with probability $1-\nu$?

- We perform n random tosses of the coin
- We output our estimate $\widetilde{p}$ to be the number of heads observed divided by n
- Choose n such that $2\exp(-2\varepsilon^2 n) \leqslant \nu$

Application: Soundness Amplification

Suppose $\mathcal{A}(x; r)$ is a randomized algorithm that decides whether $x \in L$ or not. Let $\mathbb{U}$ be the uniform distribution over the random tape for $\mathcal{A}(\cdot; \cdot)$. We are given the following guarantee:

- If $x \in L$, then $\mathbb{P}\left[\mathcal{A}(x; \mathbb{U}) = 1\right] \geqslant 0.7$, and
- If $x \notin L$, then $\mathbb{P}\left[\mathcal{A}(x; \mathbb{U}) = 0\right] \geqslant 0.7$

- Run $\mathcal{A}(x; r^{(i)})$, where $r^{(i)}$ is random tape independently drawn according to $\mathbb{U}$, and θ_i be the output of the algorithm
- Output the majority of the bits $\{\theta_1, \dots, \theta_n\}$

How to choose n such that the probability of our algorithm to correctly output $x \in L$ or not with probability $1 - \nu$?
Hint: Our algorithm is wrong when the set $\{\theta_1, \dots, \theta_n\}$ has $< n/$ correct answer bits.

Another Chernoff-Bound

- $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ are independent random variables
- $\mathbb{X}_i$ is a random variable in the range $[0,1]$ such that $\mathbb{E}\left[\mathbb{X}_i\right] = p_i$
- We define $np = \sum_{i=1}^{n} p_i$ and $\mathbb{S}_n = \sum_{i=1}^{n} \mathbb{X}_i$

Another Chernoff-Bound II

- We follow the previous proof's template

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \frac{\mathbb{E}\left[\exp(h\mathbb{S}_n)\right]}{\exp(hn(p+\varepsilon))} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[h\mathbb{X}_i\right]}{\exp(hn(p+\varepsilon))}$$

- Claim: $\mathbb{E}\left[h\mathbb{X}_i\right] \leqslant \mathbb{E}\left[B(1,p_i)\right]$. Recall that $B(1,p_i)$ is the random variable that outputs 1 with probability p_i; 0 otherwise. The above inequality is a consequence of Jensen's Inequality. Let f be a convex downward function, i.e. it looks like $\exp(x)$. The chord joining $(x_0, f(x_0))$ and $(x_1, f(x_1))$ is below the chord joining $(x_0', f(x_0'))$ and $(x_1', f(x_1'))$ when $x_0, x_1 \in [x_0', x_1']$.
- Intuition: $\mathbb{E}\left[h\mathbb{X}_i\right]$, when $\mathbb{E}\left[\mathbb{X}_i\right] = p_i$, is maximized when $\mathbb{X}_i = B(n, p_i)$

- So, we have:

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \frac{\prod_{i=1}^{n}(1-p_i+p_i\exp(h))}{\exp(hn(p+\varepsilon))}$$

- Now we apply AM-GM inequality to the right hand side

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \left(\frac{\frac{1}{n}\sum_{i=1}^{n} 1-p_i+p_i\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$
$$= \left(\frac{1-p+p\exp(h)}{\exp(h(p+\varepsilon))}\right)^n$$

- Now we bound as in the original Chernoff bound proof

Hoeffding's Bound

- Let $(\mathbb{X}_1, \ldots \mathbb{X}_n)$ be independent random variables
- Let $\mathbb{X}_i$ has range $[a_i, b_i]$ with $\mathbb{E}[\mathbb{X}_i] = p_i$
- We define $\mathbb{S}_n = \sum_{i=1}^n \mathbb{X}_i$ and $np = \sum_{i=1}^n p_i$

- As in the previous case, we can get:

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \frac{\prod_{i=1}^{n} \mathbb{E}\left[h\mathbb{X}_i\right]}{\exp(hn(p+\varepsilon))}$$

- By Jensen's Inequality we can bound the right hand side by consider a distribution that only puts probability on a_i and b_i such that it has expectation p_i. This distribution outputs a_i with probability $\frac{b_i-p_i}{b_i-a_i}$, and outputs b_i with probability $\frac{p_i-a_i}{b_i-a_i}$. We use this distribution to obtain the upper bound on the right hand side.

$$\mathbb{P}\left[\mathbb{S}_n \geqslant n(p+\varepsilon)\right] \leqslant \frac{\prod_{i=1}^{n} \left(\frac{b_i-p_i}{b_i-a_i}\exp(ha_i) + \frac{p_i-a_i}{b_i-a_i}\exp(hb_i)\right)}{\exp(hn(p+\varepsilon))}$$

Hoeffding's Bound

- How to proceed to get an upper bound?