Lecture 05: Concentration Bounds

Theorem (Markov Inequality)

Let $\mathbb{X}$ be a positive random variable. Then the following holds:

$$\mathbb{P}[\mathbb{X} \geqslant t] \leqslant \frac{\mathbb{E}[\mathbb{X}]}{t}$$

- We only present the proof when the sample space Ω is discrete
- Suppose this statement is false, i.e.,

$$\mathbb{P}[\mathbb{X} \geqslant t] > \frac{\mathbb{E}[\mathbb{X}]}{t}$$

- Then we can perform the following analysis

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i\in\Omega} i\cdot\mathbb{P}[\mathbb{X}=i]\\ &= \sum_{i\in\Omega:\ i<t} i\cdot\mathbb{P}[\mathbb{X}=i] + \sum_{i\in\Omega:\ i\geqslant t} i\cdot\mathbb{P}[\mathbb{X}=i]\\ &\geqslant \sum_{i\in\Omega:\ i<t} 0\cdot\mathbb{P}[\mathbb{X}=i] + \sum_{i\in\Omega:\ i\geqslant t} t\cdot\mathbb{P}[\mathbb{X}=i]\\ &= t\cdot\mathbb{P}[\mathbb{X}\geqslant t] > \mathbb{E}[\mathbb{X}]\end{aligned}$$

- Hence, contradiction. This proves the theorem
- The probability distribution that shows that this inequality is tight is:

$$\mathbb{P}[\mathbb{X}=0] = 1 - 1/t \text{ and } \mathbb{P}[\mathbb{X}=t] = 1/t$$

Chebyshev's Inequality

Theorem (Chebyshev's inequality)

$$\mathbb{P}\left[\left|\mathbb{X} - \mathbb{E}\left[X\right]\right| \geqslant t\right] \leqslant \frac{\mathrm{Var}\left[\mathbb{X}\right]}{t^2}$$

- Note that $\mathbb{P}\left[\left|\mathbb{X} - \mathbb{E}\left[X\right]\right| \geqslant t\right] = \mathbb{P}\left[\left(\mathbb{X} - \mathbb{E}\left[X\right]\right)^2 \geqslant t^2\right]$
- Apply Markov Inequality

Some Exercises

- Compute the $\mathbb{E}[\mathbb{X}]$ and $\mathrm{Var}[\mathbb{X}]$ for the following probability distributions
  1. $\mathbb{P}[\mathbb{X} = 0] = 1 - 1/t$ and $\mathbb{P}[\mathbb{X} = t] = 1/t$
  2. For positive constant p, $\mathbb{P}[\mathbb{X} = 0] = 1 - p$ and $\mathbb{P}[\mathbb{X} = 1] = p$.
- Prove the following properties for independent probability distributions $\mathbb{X}$ and $\mathbb{Y}$
  1. Prove that $\mathbb{E}\left[\exp(\mathbb{X} + \mathbb{Y})\right] = \mathbb{E}\left[\exp(\mathbb{X})\right] \cdot \mathbb{E}\left[\exp(\mathbb{Y})\right]$
  2. Prove that $\mathrm{Var}[\mathbb{X} + \mathbb{Y}] = \mathrm{Var}[\mathbb{X}] + \mathrm{Var}[\mathbb{Y}]$

Chernoff Bound

- Let $\mathbb{X}$ be the probability distribution such that $\mathbb{P}[\mathbb{X}=0] = 1-p$ and $\mathbb{P}[\mathbb{X}=1] = p$ (Bernoulli variable)
- Let $B(n,p)$ be the random variable $\sum_{i=1}^{n} \mathbb{X}^{(i)}$, where $\mathbb{X}^{(i)}$ is the i-th independent sample according to the distribution $\mathbb{X}$
- $B(n,p)$ (sum of n independent Bernoulli variables)

Theorem (Chernoff Bound)

For $0 < t < 1-p$, we have

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

- $\mathrm{D}_{\mathrm{KL}}(\cdot,\cdot)$ is the Kullback-Leibler divergence and is defined as follows

$$\mathrm{D}_{\mathrm{KL}}(\alpha,\beta) := \alpha \lg\left(\frac{\alpha}{\beta}\right) + (1-\alpha)\lg\left(\frac{1-\alpha}{1-\beta}\right)$$

It is always $\geqslant 0$. Why?

Comments on the Bound

- Substituting $t = 1/\sqrt{n}$, we get that the probability $\mathbb{P}\left[B(n,p) \geqslant np + \sqrt{n}\right] \leqslant \text{const.}$, $B(n,p)$ is very strongly concentrated around its mean!
- Chernoff Bound is also very tight. That is, we have

$$\mathbb{P}\left[B(n,p) - np \geqslant nt\right] \geqslant \frac{\exp\left(-2t^2 n\right)}{\mathrm{poly}(n)}$$

  This can be proven using "Stirling Approximation" or "The Method of Types"
- Even using limited independence, one can get the same bound as Chernoff Bound: "Chernoff-Hoeffding Bounds for Applications with Limited Independence, " by Jeanette P. Schmidth, Alan Siegel, and Aravind Srinivasan.

Proof of Chernoff Bound

- We are interested in bounding

$$\mathbb{P}\left[B(n,p) \geqslant n(p+t)\right]$$

- For any $h > 0$, this probability is identical to

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

- By Markov Inequality, we get

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

$$= \frac{\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{X}^{(i)}\right)\right]}{\exp\left(hn(p+t)\right)}$$

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

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

Proof of Chernoff Bound III

- This expression is an upper-bound for all $h > 0$.
- So, we choose h that minimizes the upper bound expression.
- Using basic calculus, it can be shown that the expression

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

is minimized for

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

- Note that here we are using the fact that $(1 - p - t) > 0$

- Now, let us calculate the expression E

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

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

- Note that the expression E we obtained satisfies

$$-\lg E = \mathrm{D}_{\mathrm{KL}}\left(p+t, p\right)$$

- Therefore we get the first part of the Chernoff bound:

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

- For the final part of the inequality, we need to show that

$$\left(\frac{1-p}{1-p-t}\right)^{1-p-t} \cdot \left(\frac{p}{p+t}\right)^{p+t} \leqslant \exp\left(-2t^2\right)$$

- First Attempt (Which will fail)

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

Using the fact that $1+x \leqslant \exp(x)$

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

Using the fact that $1-x \leqslant \exp(-x)$

Therefore,

$$E = \left(\frac{1-p}{1-p-t}\right)^{1-p-t} \cdot \left(\frac{p}{p+t}\right)^{p+t} \leqslant \exp(t) \cdot \exp(-t) = 1$$

- Although this is true, but we do not get anything nontrivial.
- Our target is to show that $E \leqslant \exp(-2t^2)$.

- Second Attempt
- We use $\overline{p} = 1 - p$ for brevity
- $f(t) := \ln E = (p+t)\ln\left(\frac{p}{p+t}\right) + (\overline{p}-t)\ln\left(\frac{\overline{p}}{\overline{p}-t}\right)$
- Observe that $f(0) = 0$
- Note that

$$\begin{aligned} f'(t) &= \ln\left(\frac{p}{p+t}\right) - 1 - \ln\left(\frac{\overline{p}}{\overline{p}-t}\right) + 1 \\ &= \ln\left(\frac{p}{p+t}\right) - \ln\left(\frac{\overline{p}}{\overline{p}-t}\right) \end{aligned}$$

- Observe that $f'(0) = 0$
- Note that

$$f''(t) = -\frac{1}{p+t} - \frac{1}{\overline{p}-t} = -\frac{1}{(p+t)(\overline{p}-t)}$$

- Now, we use Taylor Series expansion around $x_0 = 0$. There exists a positive constant $c \in (0, t]$ such that:

$$\begin{aligned} f(t) &= f(0) + f'(0) \cdot t + f''(c) \cdot \frac{t^2}{2} \\ &= -\frac{1}{(p+c)(\overline{p}-c)} \frac{t^2}{2} \\ &\leqslant -2t^2 \end{aligned}$$

- For the last step, we used AM-GM inequality:

$$\sqrt{(p+c)(\overline{p}-c)} \leqslant \frac{(p+c)+(\overline{p}-c)}{2} = \frac{1}{2}$$

- Recall that we had defined $f(t) = \ln E$, so $f(t) \leqslant -2t^2$ implies $E \leqslant \exp(-2t^2)$. This completes the proof of the last part of Chernoff Bound.