RECITATION - 1 :

Concentration Ineq.

By : Apoorv Vikram Singh

Motivation

- Functions of random variables are very useful.

  e.g. • $f(X_1, \ldots, X_n) = X_1 + \cdots + X_n$,

  • $f(X_1, \ldots, X_n) = \max_{i \in [n]} X_i$, etc.

- Expectation is relatively easy to compute/bound

- Concentration ineq. gives conditions under which

$$f(X_1, \ldots, X_n) \approx \mathbb{E}[f(X_1, \ldots, X_n)].$$

- $\mathbb{P}\left[\ \left|f(X_1,\dots,X_n) - \mathbb{E}[f(X_1,\dots,X_n)]\right| \geq \varepsilon\right] \leq \delta.$

- In AMLDS,

  $\mathbb{P}[$ of a bad event $] \leq \delta.$

  - generally we are interested in various regimes of ε & δ.

    Ideally small ε & small δ.

  - Small $\varepsilon \Rightarrow$ $f(X_1,\dots,X_n)$ NOT too far away from $\mathbb{E}[f(X_1,\dots,X_n)]$

  - Small $\delta \Rightarrow$ $f(X_1,\dots,X_n)$ is close to $\mathbb{E}[f(X_{1:n})]$ most of time.

e.g. In AMLDS, the smaller the δ, the union bound can be taken over more bad events.

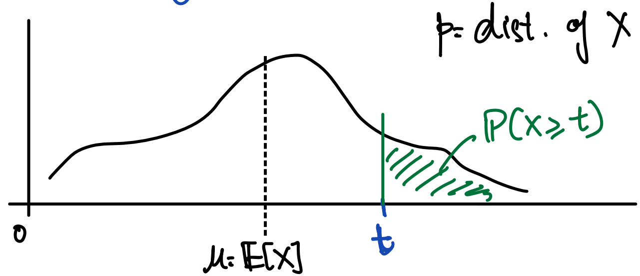
(for a fixed failure probability).

* Markov's Ineq.

• Thm: r.v. X, non-negative valued. For any $t > 0$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

Pf: Done in class.

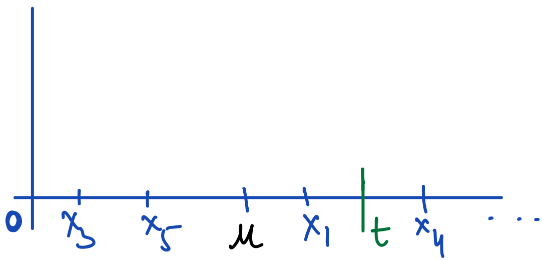


$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{s} \mathbb{P}(X=s)\cdot s \\
&= \sum_{s<t} \mathbb{P}(X=s)\cdot s + \sum_{s\geq t} \mathbb{P}(X=s)\, s \\
&\geq 0 + t \cdot \sum_{s\geq t} \mathbb{P}(X=s) \\
&= t \cdot \mathbb{P}(X \geq t)
\end{aligned}
$$

∎

- Let $X_1, \dots, X_n \overset{iid}{\sim}$ dist. of X (think of n as HUGE)

- $$\frac{X_1 + \cdots + X_n}{n} = \mathbb{E}X =: \mu$$



$$\frac{\sum X_i}{n} = \mathbb{E}X = \mu$$

$X_i \geq 0, \quad \forall i$

- What is the maximum number of X_i that are $\geq t$ & avg. is μ?

Worst case.

$\left\{ \begin{array}{l} \text{To compute that, suppose all } X_i\text{'s} < t \text{ are at } 0. \\ \text{All } X_i\text{'s} \geq t \text{ are at } t \end{array} \right.$

$$\therefore \quad \frac{0 + (\# X_i\text{'s} \geq t) \cdot t}{n} = \mathbb{E}X$$ → in worst case

$$= \frac{(\# X_i\text{'s} \geq t)}{n} \cdot t = \mathbb{E}X$$ → in worst case

$$\therefore \quad \frac{(\# X_i\text{'s} \geq t)}{n} \leq \frac{\mathbb{E}X}{t}$$ → in reality, we get $\leq$.

$$\Rightarrow \quad \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}.$$

$\blacksquare$

- What if r.v. X takes negative values?

Idea: Apply Markov's, but make your r.v. somehow non-negative.

Chebyshev's: Applied Markov's to $(X-\mathbb{E}X)^2$, instead of $X-\mathbb{E}X$

i.e.

$$\mathbb{P}\left((X-\mathbb{E}X)^2 \geq t^2\right) \leq \frac{\mathbb{E}\left((X-\mathbb{E}X)^2\right)}{t^2}$$

$$= \frac{Var(X)}{t^2}$$

$$? \Rightarrow \quad \mathbb{P}\left(|X-\mathbb{E}X| \geq t\right) \leq \frac{Var(X)}{t^2}.$$

We are interested in the event:

$$\mathbb{P}\Big(\underbrace{\{|X-\mathbb{E}X|\geq t\}}_{A}\Big)$$

So, what we did was, we looked at

$$\mathbb{P}\Big(\underbrace{\{(X-\mathbb{E}X)^2\geq t^2\}}_{B}\Big)$$

We need to show:

$$\underbrace{\{(X-\mathbb{E}X)^2\geq t^2\}}_{B} \underset{\Rightarrow}{\Rightarrow} \underbrace{\{|X-\mathbb{E}X|\geq t\}}_{A}$$

bound $\mathbb{P}(B)$.

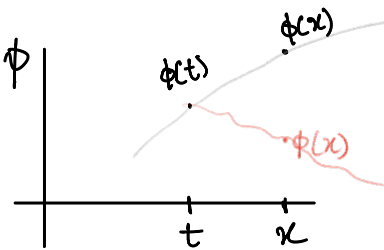
- <u>In general</u>:

Let ϕ be a function s.t. $\phi(x) \geq 0, \forall x$

$$\phi(x) \geq \phi(t) \implies x \geq t.$$

Then

$$\mathbb{P}(X \geq t) \leq \mathbb{P}(\phi(X) \geq \phi(t)) \leq \frac{\mathbb{E}\phi(X)}{\phi(t)}$$



$\therefore$ Choose a <u>non-decreasing</u> <u>non-negative</u> function.

- What was ϕ in Chebyshev's?

* <u>Goal</u> : Choose ϕ s.t. $\phi(t)$ is large

& $\mathbb{E}\phi(X)$ is easy to bound

(& small).

* Suppose $X \sim \mathcal{N}(0,1)$: Gaussian with mean 0

& variance $\sigma^2 = 1$

<u>Chebyshev's</u>:

$$\mathbb{P}(|X-\mathbb{E}X| \geq t) \leq \frac{1}{t^2}$$

<u>Actual</u>:

$$\mathbb{P}(|X-\mathbb{E}X| \geq t) \leq 2e^{-t^2/2} \qquad \simeq \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

* We see a HUGE difference.
  Let's fix error probability δ.

→ In the Gaussian case, Chebyshev's ineq. gives:

$$\mathbb{P}(|X-\mathbb{E}X| \geq t) \leq \frac{1}{t^2} = \delta \quad \Rightarrow \quad t = \sqrt{\frac{1}{\delta}}$$

$$\Rightarrow \quad \boxed{\mathbb{P}\left(|X-\mathbb{E}X| \geq \sqrt{\tfrac{1}{\delta}}\right) \leq \delta}$$

→ Actually:

$$\mathbb{P}(|X-\mathbb{E}X| \geq t) \leq 2e^{-t^2/2} = \delta$$

$$\Rightarrow \quad t = \sqrt{\log\left(\frac{2}{\delta}\right)}$$

$$\Rightarrow \quad \boxed{\mathbb{P}\left(X-\mathbb{E}X \geq \sqrt{\log(1/\delta)} \right) \leq \delta}.$$

$\therefore$ X is MUCH closer to its expectation when X is a Gaussian r.v.

COMPARE $\sqrt{\log \frac{1}{\delta}}$ v/s $\sqrt{\frac{1}{\delta}}$.

- Maybe a more clever ϕ could help us.
- Let's try $\phi(x) = x^4$.

Let $X \sim \mathcal{N}(0,1)$.

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \mathbb{P}\left((X - \mathbb{E}X)^4 \geq t^4\right) \leq \frac{\mathbb{E}\left((X - \mathbb{E}X)^4\right)}{t^4} \leq \frac{3\sigma^4}{t^4}$$

$$\Rightarrow \mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \frac{3}{t^4} = \delta \quad \Rightarrow \quad t = \sqrt[4]{\frac{3}{\delta}}$$

$$\Rightarrow \mathbb{P}\left(|X - \mathbb{E}X| \geq \sqrt[4]{\frac{3}{\delta}}\right) \leq \delta.$$

With magic of Wikipedia, we can get that

$$\mathbb{P}(|X-\mathbb{E}X| \geq \sigma t) \leq \mathbb{P}(|X-\mathbb{E}X|^{2m} \geq t^{2m}\sigma^{2m}) \leq \frac{2^{-m} \cdot \frac{(2m)!}{m!}}{t^{2m}}$$

$$\leq \frac{2^{m} \cdot m^{m}}{t^{2m}} = \delta$$

$$\Rightarrow \quad t^{2} = \frac{2^{m} \cdot m^{m}}{\delta^{1/m}} \Rightarrow t = \sqrt{\frac{2 \cdot m}{\delta^{1/m}}}$$

$$\mathbb{P}\left(|X-\mathbb{E}X| \geq \sqrt{\frac{2m}{\delta^{1/m}}}\right) \leq \delta$$

After so much hard work still didn't get $\sqrt{\log\left(\frac{1}{\delta}\right)}$.

- Good attempt, but $\mathbb{E}(X-\mathbb{E}X)^m$ is difficult to calculate in general.

- Maybe, I can apply a more clever function.

Let's try $\phi(x) = e^{\lambda x}$ (for some λ)

$$\mathbb{P}((X-\mathbb{E}X) \geq t) \leq \mathbb{P}\left(e^{\lambda(X-\mathbb{E}X)} \geq e^{\lambda t}\right)$$

$$\leq \frac{\mathbb{E}\left(e^{\lambda(X-\mathbb{E}X)}\right)}{e^{\lambda t}}$$

$$= \frac{\exp\left(\frac{\lambda^2 \sigma^2}{2}\right)}{e^{\lambda t}}$$

* **Recall the goal:** Make $\mathbb{E}\,\phi(X)$ small
$\phi(t)$ big.

λ is in our control, so I will choose the best λ.

$$\min_{\lambda} \exp\left(\frac{\lambda^2\sigma^2}{2} - \lambda t\right)$$

$$: \quad \exp\left(\frac{\lambda^2\sigma^2}{2} - \lambda t\right) \cdot \left(\frac{2\lambda\sigma^2}{2} - t\right) = 0$$

$$\therefore \quad \lambda = t/\sigma^2$$

$$\Rightarrow \exp\left(\frac{\lambda^2\sigma^2}{2} - \lambda t\right) = \exp\left(\frac{t^2}{\sigma^4}\cdot\frac{\sigma^2}{2} - \frac{t\,t}{\sigma^2}\right) = \exp\left(-t^2/2\sigma^2\right)$$

$\therefore$ We get

$$\mathbb{P}(X - \mathbb{E}X > t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) = \delta$$

$$\therefore \quad t = \sqrt{2\log\left(\frac{1}{\epsilon}\right)}$$

$$\Rightarrow \mathbb{P}\left(X - \mathbb{E}X > \sqrt{2\log\frac{1}{\epsilon}}\right) \leq \delta$$

$\Rightarrow$ We get the desired bound in order.

$\phi(x) = \exp(\lambda x)$, awesome!!!

* $\phi(x) = e^{\lambda x}$ is a nice idea for Gaussian.
BUT, is $\mathbb{E}[\exp(\lambda X)]$ easy to compute?

:. Lets try it for our fav. example.

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p)$$

$$\mathbb{E} \sum X_i = np \quad , \quad S = \sum X_i$$

· <u>Markov's</u> $\quad \mathbb{P}(S > t) \leq \frac{np}{t} = \delta$

$t = \frac{p}{\delta} \quad \Rightarrow \quad \mathbb{P}\left(S > \frac{np}{\delta}\right) \leq \delta .$

Chebyshov's

$$\mathbb{P}\left(\left| \sum x_i - np \right| \geq t \right) \leq \frac{\mathbb{E}\left(\sum x_i - np\right)^2}{t^2}$$

$$= \operatorname{var}\left(x_1 + \cdots + x_n \right) = \sum_i \operatorname{var}(x_i)$$

$$= np(1-p)$$

$$\therefore \ \mathbb{P}\left(|S-p| \geq t \right) \leq \frac{np(1-p)}{t^2} = \delta \Rightarrow t = \sqrt{\frac{np}{\delta}}$$

$$\Rightarrow \quad \mathbb{P}\left(\left| \cancel{S_n} - p \right| \geq \sqrt{\frac{np}{\delta}} \right) \leq \delta .$$

· Let's try with $\phi(x) = e^{\lambda x}$.

$$\mathbb{P}\left(\sum X_i - np \geq t\right) \leq \mathbb{P}\left(\exp\left(\lambda\left(\sum X_i - np\right)\right) \geq \exp(\lambda t)\right)$$

$$\leq \frac{\mathbb{E}\exp\left(\lambda\left(\sum X_i - np\right)\right)}{\exp(\lambda t)}$$

$$= \frac{\exp(-\lambda np)}{\exp(\lambda t)}\,\mathbb{E}\left[\exp\left(\lambda \sum X_i\right)\right]$$

Let's calculate $\mathbb{E}\left[\exp\left(\lambda \sum X_i\right)\right]$

$$\Rightarrow \mathbb{E}\left[\exp(\lambda X_1)\cdot\exp(\lambda X_2)\cdot\ldots\cdot\exp(\lambda X_n)\right]$$

$$= \mathbb{E}\left[\exp(\lambda X_1)\right]^n \quad \Rightarrow \text{why?}$$

$$\mathbb{E}\left[\exp(\lambda X_1)\right] = p\cdot\exp(\lambda) + (1-p)\exp(0)$$

$$= pe^{\lambda} + (1-p).$$

∴ We get that

$$\mathbb{P}\left(\left|\sum X_i - np\right| \geq t\right) \leq \frac{\left(pe^{\lambda} + (1-p)\right)^n \exp(-\lambda np)}{\exp(\lambda t)}$$

GOAL: Make RHS small

$$\Rightarrow \min \frac{\left(pe^{\lambda} + (1-p)\right)^n \exp(-\lambda np)}{\exp(\lambda t)}$$

$$n\left(pe^{\lambda}+(1-p)\right)^{n-1}\cdot pe^{\lambda}\exp(-\lambda np-\lambda t)$$

$$+\left(pe^{\lambda}+1-p\right)^{n}\exp(-\lambda np-\lambda t)\cdot(-np-t)=0$$

$\Rightarrow$

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

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

$\Rightarrow$ $$\lambda = \log\left(\frac{\left(\frac{1-p}{p}\right)\left(\frac{np+t}{n}\right)}{\left(1-\frac{np+t}{n}\right)}\right)$$

Too Difficult to evaluate.

Let's try to simplify.

Chernoff calculations

- $\left(pe^{\lambda}+(1-p)\right)^{n} \exp(-\lambda np - \lambda t)$
- Let $t = (1+\varepsilon)np$.

$$\Rightarrow \left(pe^{\lambda}+(1-p)\right)^{n} \exp\left(-\lambda np(2+\varepsilon)\right)$$

$$e^{\lambda x} \leq 1+(e^{\lambda}-1)x \quad , \quad 1+x \leq e^{x}$$

$$\begin{aligned}
\mathbb{E}\, e^{\lambda X_i} &\leq \mathbb{E}\left[1+(e^{\lambda}-1)X_i\right] \\
&= 1+(e^{\lambda}-1)p \\
&\leq \exp\left((e^{\lambda}-1)p\right)
\end{aligned}$$

$$\therefore \mathbb{P}\left(\sum X_i \geq (1+\varepsilon)\mu \right)$$

$$= \frac{\exp((e^\lambda - 1)p)^n}{\exp(\lambda(1+\varepsilon)\mu)}$$

$$\therefore \mathbb{P}\left(X \geq (1+\varepsilon)\mu \right) \leq \exp\left((e^\lambda - 1)np - \lambda(1+\varepsilon)np\right)$$

$$\min_\lambda (e^\lambda - 1)np - \lambda(1+\varepsilon)np$$

$$\Rightarrow e^\lambda \cancel{np} = (1+\varepsilon)\cancel{np} \qquad \lambda = \log(1+\varepsilon)$$

$$\Rightarrow \mathbb{P}\left(X \geq (1+\varepsilon)\mu \right) \leq$$

$$\exp\left(\varepsilon np - (1+\varepsilon)\log(1+\varepsilon)np\right)$$

$$\Rightarrow$$

$$\mathbb{P}\left((X-\mu) \geq \varepsilon\mu\right) \leq \exp\left(-\mu\left((1+\varepsilon)\log(1+\varepsilon)-\varepsilon\right)\right)$$

$$\because \varepsilon \in [0,1]$$

$$\mathbb{P}\left(X \geq (1+\varepsilon)\mu\right) \leq \exp\left(-\mu\varepsilon^2/3\right)$$

* **Chernoff**: Final Bound

$$\exp\left(-\mu \frac{\varepsilon^2}{3}\right) = \delta$$

$$\Rightarrow \quad \frac{\mu \varepsilon^2}{3} = \log\left(\frac{1}{\delta}\right) \Rightarrow \quad \varepsilon = \sqrt{\frac{3\log\left(\frac{1}{\delta}\right)}{\mu}}$$

$\therefore$

$$\mathbb{P}\left(\frac{X-\mu}{\mu} \geq \sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{np}}\right) \leq \delta$$

very similar

for gaussians

$$\mathbb{P}\left(\frac{X-\mu}{\mu} \geq \overset{\approx\sqrt{p}}{\sigma}\sqrt{\frac{2\log\left(\frac{1}{\delta}\right)}{n}}\right) \leq \delta.$$

* ~~In~~ In general, , If we know a good bound on

$$\mathbb{E}(e^{\lambda X_i})$$, that suffices.

$\Rightarrow$ If $X_i \in [a, b]$, then,

$a \leq 0 \leq b$ wlog

HW
$$\mathbb{E}(e^{\lambda X_i}) \leq \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right)$$

Subgaussian r.v.

$$\mathbb{E}(e^{\lambda X_i}) \leq \exp\left(\frac{\lambda^2 \sigma^2}{2}\right), \forall \lambda$$

* <u>Hoeffdings</u>

$$\mathbb{P}\left(\sum X_i \geq t\right) \leq \frac{\exp\left(\frac{\lambda^2 (b-a)^2 \cdot n}{8}\right)}{\exp(\lambda t)}$$

$$\exp\left(\frac{\lambda^2 (b-a)^2 n}{8} - \lambda t\right)$$ min wrt λ

$$2\lambda \frac{(b-a)^2 n}{8} = t \Rightarrow \lambda = \frac{4t}{(b-a)^2 n}$$

$$\Rightarrow$$

$$\mathbb{P}\left(\sum X_i \geq t\right) \leq \exp\left(\frac{2t^2}{(b-a)^2\, n} - \frac{4t^2}{(b-a)^2 n}\right)$$

$$\leq \exp\left(\frac{-2t^2}{(b-a)^2 \cdot n}\right) \qquad \blacksquare$$

McDiarmid's Inequality[1] — Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfy the bounded differences property with bounds $c_1, c_2, \ldots, c_n$.

Consider independent random variables $X_1, X_2, \ldots, X_n$ where $X_i \in \mathcal{X}_i$ for all i. Then, for any $\varepsilon > 0$,

$$\mathrm{P}\left(f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)] \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right),$$

$$\mathrm{P}(f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)] \leq -t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right),$$

and as an immediate consequence,

$$\mathrm{P}(|f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)]| \geq t) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Bounded difference

$$\left|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)\right| \leq c_i, \quad \forall i \in [n]$$

HW :

1. Try Chernoff's proof on your own.
2. Prove Hoeffding's lemma, i.e., $(a \le 0 \le b)$

   If $X_i \in [a, b]$ & $\mathbb{E}X_i = 0$, then

$$\mathbb{E}\left[e^{\lambda X_i} \right] \le \exp\left(\frac{\lambda (b-a)^2}{8} \right), \quad \forall \lambda \in \mathbb{R}$$

3. Work out proof of Chernoff using Hoeffding's lemma.
4. Google Subgaussian r.v.