Digression — Inequalities — towards Concentration

Markov Inequality:- $X \geq 0 \quad c > 0 \quad E[X] = \mu < \infty$

$$\mathbb{P}(X \geq c) \leq \frac{\mu}{c}$$

[To capture $X : \mathbb{1}(X \geq c)$]

Tschebyschev Inequality X – R.v. $\quad \mu = E[X] < \infty \quad \sigma^2 = Var[X] < \infty$

Fix $k = 1, 2, 3, \ldots$

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$X_n = \sum_{i=1}^{n} \xi_i \qquad \xi_i \overset{d}{=} \text{Bernoulli}(p) \qquad \hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$[T\text{-}I] \quad \varepsilon > 0, \quad \mathbb{P}(|\hat{p}_n - p| \geq \varepsilon) \leq \frac{Var\, \hat{p}_n}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \quad \text{—} (*)$$

C.L.T.

$$\mathbb{P}\left(\sqrt{\frac{n}{p(1-p)}}\,(\hat{p}_n - p) \geq \alpha\right) \longrightarrow 1 - \int_{-\infty}^{\alpha} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \leq \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-\alpha^2/2}}{\alpha}$$

$$\Rightarrow \quad \mathbb{P}(\hat{p}_n - p \geq \varepsilon) \approx e^{-\frac{n\varepsilon^2}{2p(1-p)}} \quad \text{—} (**)$$

$X \geq 0$ Random variable

$t>0$: $\mathbb{P}(X \geq t) = \mathbb{P}(e^{sX} \geq e^{st}) \leq \frac{Ee^{sX}}{e^{st}} = e^{-st} E[e^{sX}]$

$s>0$

Markov

$\phi(x) = e^{x}$ - m.I.

[Chernoff] $\mathbb{P}(X \geq t) \leq \inf_{s \geq 0} e^{-st} E[e^{sX}]$

good bound on moment generating function

Lemma [Hoeffding]: X - random variable $EX=0$ and $\mathbb{P}(a \leq X \leq b) = 1$ for $a,b \in \mathbb{R}$

$$\forall s>0 \qquad E[e^{sX}] \leq e^{s^2(b-a)^2/8}$$

Proof:- [sketch]

$$x = \frac{x-a}{b-a} b + \frac{b-x}{b-a} a$$

$a \leq x \leq b$, $s>0$

$$e^{sx} \leq \frac{x-a}{b-a} e^{sb} + \frac{b-x}{b-a} e^{sa}$$

$$\Rightarrow E[e^{sX}] \leq \left[\frac{b}{b-a} - \frac{a}{b-a} e^{s(b-a)}\right] e^{sa}$$

$\cdot \; \lambda = -\frac{a}{b-a}$

$$E[e^{sX}] \leq \left[1-\lambda - \lambda e^{s(b-a)}\right] e^{-\lambda s(b-a)}$$

Take: $\phi(u) = \log(1-\lambda+\lambda e^u) - \lambda u$

use calculus :- $\phi(u) \leq u^2/8$ $u \geq 0$

[Taylor's Theorem]

$$E[e^{sX}] \leq e^{\phi(s(b-a))} \leq e^{s^2(b-a)^2/8}$$

$\square$

Theorem [Hoeffding's Inequality]

$X_1, \dots X_n$ be independent random variables

$X_i \in [a_i, b_i]$ w.p. 1 $\qquad S_n = \sum_{i=1} X_i$

$$\mathbb{P}(S_n - ES_n \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right)$$

$$\mathbb{P}(S_n - ES_n \leq -t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right)$$

$$\Rightarrow \quad \mathbb{P}(|S_n - ES_n| \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right)$$

Proof :- w.l.o.g $E(X_i) = 0$.

[Chernoff]. $\quad \mathbb{P}(S_n \geq t) \leq e^{-st} Ee^{sS_n}$

$$\leq e^{-st} \prod_{i=1}^n [Ee^{sX_i}]$$

$X_i \in [a_i, b_i)$ $\quad \underline{H.L.}$ $\quad E[e^{sX_i}] \leq e^{\frac{s^2(b_i-a_i)^2}{8}}$

$$\Rightarrow \quad \mathbb{P}(S_n \geq t) \leq e^{-st} \prod_{i=1}^{n} e^{\frac{s^2(b_i-a_i)^2}{8}}$$

$$= \exp\left(-st + \sum_{i=1}^{n} \frac{(b_i-a_i)^2}{8} s^2\right)$$

<u>minimum</u>

$$s_0 = \frac{4t}{\sum_{i=1}^{n}(b_i-a_i)^2} \quad \Rightarrow \quad \mathbb{P}(S_n \geq t) \leq e^{-2t^2/\sum_{i=1}^{n}(b_i-a_i)^2}$$

<u>Ex:</u> Rest of the proof is similar $\square$

<u>From Bounded random variables to Bounded differences</u>

<u>Theorem [McDiarmid's Inequality]</u>: $(X_1 .. X_n) \in \mathcal{X}^n$ be n-independent random variables. $g: \mathcal{X}^n \to \mathbb{R}$ has bounded differences i.e

$$\sup_{\substack{x_1 .. x_n \\ x_i' \in \mathcal{X}}} |g(x_1, .., x_i, .., x_n) - g(x_1, .. x_{i-1}, x_i', x_{i+1} .. x_n)| \leq c_i$$

$$\Rightarrow \quad \mathbb{P}\left(g(X_1 .. X_n) - Eg(X_1, .., X_n) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

$$\Rightarrow \quad \mathbb{P}(g(X_1 .. X_n) - Eg(X_1 .. X_n) \leq -t) \leq \exp\left(-2t^2/\sum_{i=1}^{n} c_i^2\right)$$

<u>Proof</u>:- $V_n = g(X_1 .. X_n) - E(g(X_1 .. X_n)) \equiv \sum_{i=1}^{n} V_i$

with V_i - independent

<u>Take</u> -> $V_i \equiv$ depending on $X_1, .., X_i$

-> $\psi_i : \mathcal{X}^{i-1} \to \mathbb{R}$ such that conditionally on X_{i-1};

$$\psi_i(X_1, \ldots, X_{i-1}) \le V_i \le \psi_i(X_1, X_{i-1}) + c_i$$

$$(Ex) \implies E\left(e^{sV_i} \mid \mathcal{X}^{i-1}\right) \le e^{s^2 c_i/8}$$

$$\mathbb{P}(V \geq t) \le e^{-st} E[e^{sV}]$$

$$= e^{-st} E[e^{s\sum_{i=1}^{n} V_i}]$$

[Condition application] $\le e^{-st} e^{\frac{s^2 c_n^2}{8}} E[e^{s\sum_{i=1}^{n-1} V_i}]$

[Repeat] n $.. \le \exp\left(-st + \frac{s^2}{8}\sum_{i=1}^{n} c_i^2\right)$

minimize over s to conclude

- $Z = g(X_1 \ldots X_n) \qquad H_i((X_1 \ldots X_i)) = E[Z \mid X_i]$

Set: $\Rightarrow V_i = H_i(X_1, \ldots, X_i) - H_{i-1}(X_1, \ldots, X_{i-1})$

- $\sum_{i=1}^{n} V_i = Z - EZ = \sigma$

- $\varphi_i(X_1, \ldots X_{i-1}) = \inf_{x \in \mathcal{X}} H_i(X_1 \ldots X_{i-1}, x) - H_{i-1}(X_1 \ldots X_{i-1})]$

– Rest to be done as assignment in Hw2. $\square$

Application in: { Dense graphs / Sparse graph

$a_i \leq X_i \leq b_i$

$\rightarrow g(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ – McDiarmids – reduces to Hoeffding

Uniform Deviations

$X_1, \ldots, X_n$ i.i.d $\mathcal{X}$ – random variables with law $\mathbb{P}$

$$\mathbb{P}_n^X(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \in A)$$

– $\mathbb{P}(A)$ = True probability of A.

Hoeffding :-   Take $A \in \mathcal{A} \equiv \sigma$-algebra

Ex $\Rightarrow$ $\quad \mathbb{P}\left(|\mathbb{P}_n(A) - \mathbb{P}(A)| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$

Now:

$$g(x_1, \ldots, x_n) = \sup_{A \in \mathcal{A}} |\mathbb{P}_n^{x}(A) - \mathbb{P}(A)|$$

set $x^{(n)} = (x_1 \ldots x_n)$ $\qquad x_{(i)}^{(n)} = (x_1, \ldots x_{i-1}, x_i', x_{i+1} \ldots x_n)$

$$g(x^{(n)}) - g(x_{(i)}^{(n)}) = \sup_{A \in \mathcal{A}} |\mathbb{P}_n^{x^{(n)}}(A) - \mathbb{P}(A)| - \sup_{A' \in \mathcal{A}} |\mathbb{P}_n^{x_{(i)}^{(n)}}(A') - \mathbb{P}(A')|$$

$$= \sup_{A \in \mathcal{A}} \inf_{A' \in \mathcal{A}} \left\{ |\mathbb{P}_n^{x^{(n)}}(A) - \mathbb{P}(A)| - |\mathbb{P}_n^{x_{(i)}^{(n)}}(A') - \mathbb{P}(A')| \right\}$$

$$\leq \sup_{A \in \mathcal{A}} \left\{ |\mathbb{P}_n^{x^{(n)}}(A) - \mathbb{P}(A)| - |\mathbb{P}_n^{x_{(i)}^{n}}(A) - \mathbb{P}(A)| \right\}$$

$$\leq \sup_{A \in \mathcal{A}} \left\{ |\mathbb{P}_n^{x^{(n)}}(A) - \mathbb{P}_n^{x_{(i)}^{n}}(A)| \right.$$

$$= \frac{1}{n} \sup_{A \in \mathcal{A}} \left(\mathbb{1}(x_i \in A) - \mathbb{1}(x_i' \in A)\right) \leq \frac{1}{n}$$

Remark:
$$g(x^{(n)}_{(i)}) - g(x^{(n)}) \leq \frac{1}{n}$$

$$\implies |g(x^{(n)}) - g(x^{(n)}_{(i)})| \leq \frac{1}{n}$$

Apply mc Diarmid with $c_i = \frac{1}{n}$ $\forall i = 1 .. n$

$$\mathbb{P}\left(|g(X_1 .. X_n) - E g(X_1 .. X_n)| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$$

$$g(X_1 .. X_n) = \sup_{A \in \mathcal{A}} |\mathbb{P}_n(A) - \mathbb{P}(A)|$$ concentrates around its mean $E[g(X_1 .. X_n)]$

Hoeffding:- $E_n(f) = \int f \, d\mathbb{P}_n$, $E(f) = \int f \, d\mathbb{P}$

$$\mathbb{P}\left(|E_n(f) - E(f)| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2}$$

$$g(X_1 .. X_n) = \sup_{f \in \mathcal{F}} | E_n(f) - E(f) |$$

show $C_i = 1/n$ holds for g as in previous slide

$$\mathbb{P}(| g(X_1 .., X_n) - E\, g(X_1 .. X_n)| > \varepsilon) \quad 2 e^{-2n\varepsilon^2}.$$

- Kernel density estimation : - Extension is possible
  and McDiarmid's inequality D