(1)

LECTURE 8

Concentration of measure: general phenomenon that, if X_i weakly dependent and f not sensitive to each input, then

$$\mathbb{P}\left[|f(X_1, \dots, X_n) - \mathbb{E} f(X_1, \dots, X_n)| > t\right] \text{ small.}$$

One of main classical examples:

Thm: (Hoeffding ineq) $X_1, \dots, X_n$ independent, $X_i \in [a_i, b_i]$ a.s. $S_n := \sum_{i=1}^n X_i$ (i.e., $f(X_1, \dots, X_n) = \text{sum}$). Then,

$$\mathbb{P}\left[|S_n - \mathbb{E} S_n| > t\right] \leq 2 \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

$$\sigma^2 = \frac{1}{4} \sum_{i=1}^n (b_i - a_i)^2.$$

"S_n is σ^2-subgaussian", behaves like $N(0, \sigma^2)$

Ex: $X_i \overset{iid}{\sim} \text{Unif}(\{\pm 1\}) \rightsquigarrow \sigma^2 = n$. Can rewrite conclusion as

$$\mathbb{P}\left[\left|\frac{1}{\sqrt{n}} S_n\right| > t\right] \leq 2 \exp\left(-\frac{t^2}{2}\right)$$

$\leftarrow$ captures Gaussian tails in CLT, $\frac{1}{\sqrt{n}} S_n \approx N(0,1)$.

"Chernoff method" = exponential Markov ineq.:

$$\mathbb{P}[X \geq t] = \mathbb{P}[\exp(\lambda X) \geq \exp(\lambda t)] \leq \frac{\mathbb{E} \exp(\lambda X)}{\exp(\lambda t)}.$$

Conveniently, factorizes over independent sums: X, Y ind. $\rightsquigarrow$

$$\mathbb{E} \exp(\lambda(X + Y)) = \mathbb{E} \exp(\lambda X)\, \mathbb{E} \exp(\lambda Y).$$

$\to$ moment generating function.

Def: X is σ^2-subgaussian if, $\forall \lambda \in \mathbb{R}$, $\phi_X(\lambda) := \mathbb{E} \exp(\lambda(X - \mathbb{E}X)) \leq \exp\left(\frac{\sigma^2}{2} \lambda^2\right)$

Prop: $X \sim N(0, \sigma^2) \Longrightarrow \phi_X(\lambda) = \exp\left(\frac{\sigma^2}{2} \lambda^2\right)$.

Prop: If X, Y indep., respectively σ^2- and τ^2-subgaussian, then $X+Y$ is $(\sigma^2+\tau^2)$-subgaussian.

Prop: If X is σ^2-subgaussian, then $\mathbb{P}[|X-\mathbb{E}X| \geq t] \leq 2\exp(-t^2/2\sigma^2)$.

Pf: Chernoff method:

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

Optimal $\lambda = t/\sigma^2 \rightsquigarrow$ bound follows.

Lem: (Hoeffding) If $X \in [a,b]$ a.s., then X is $\frac{1}{4}(b-a)^2$-subgaussian.
↪ pf. of Thm immediate by combining above.

Def: $\psi_X(\lambda) := \log \phi_X(\lambda) = \log \mathbb{E}\exp(\lambda(X-\mathbb{E}X)) \leftarrow$ cumulant generating fn.

Prop: If ψ_X smooth, $\psi_X''(\lambda) \leq \sigma^2 \ \forall \lambda \in \mathbb{R}$, then X σ^2-subgaussian.

Pf: $\psi_X(0) = \log(1) = 0$.

$$\psi_X'(0) = \left.\frac{\mathbb{E}(X-\mathbb{E}X)\cancel{\exp(\lambda(X-\mathbb{E}X))}}{\mathbb{E}\cancel{\exp(\lambda(X-\mathbb{E}X))}}\right|_{\lambda=0} = 0$$

Taylor thm. w/ Lagrange remainder
~~Integrate ψ_X'' bound~~ $\rightsquigarrow \psi_X(\lambda) \leq \sigma^2 \cdot \frac{\lambda^2}{2} \rightarrow \phi_X(\lambda) \leq \exp\left(\frac{\sigma^2}{2}\lambda^2\right)$.

Pf: (of Lem) WLOG $\mathbb{E}X = 0$. Compute:

$$\psi_X'(\lambda) = \frac{\mathbb{E}X\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)} = \mathbb{E}X\rho(X) \qquad \psi_X''(\lambda) = \frac{\mathbb{E}X^2\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)} - \left(\frac{\mathbb{E}X\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}\right)^2 = \mathbb{E}X^2\rho(X) - (\mathbb{E}X\rho(X))^2$$

$\rho(X) := \frac{\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}$ is a density relative to Law(X).
↪ (cf. Radon-Nikodym)

$\rightarrow$ $\psi_X''(\lambda) = \mathrm{Var}[Y_\lambda]$, Y_λ a reweighted version of X.

In particular, $Y_\lambda \in [a,b]$ a.s.

$\rightsquigarrow$ $\mathrm{Var}[Y_\lambda] = \mathrm{Var}\Big[\underbrace{Y_\lambda - \frac{b-a}{2}}_{|\cdot| \le \frac{b-a}{2}}\Big] \le \mathbb{E}\left(Y_\lambda - \frac{b-a}{2}\right)^2 \le \frac{(b-a)^2}{4}$.

$\rightarrow$ proves Hoeffding Lemma $\rightarrow$ proves Hoeffding ineq.

//

Surprisingly, can be generalized to martingales!

Thm: (Azuma) $(M_n)_{n \ge 0}$ mgl, (A_n) and (B_n) predictable, $c_t > 0$:

$$A_t \le M_t - M_{t-1} \le B_t \quad \text{and} \quad B_t - A_t \le c_t \quad \text{, a.s.}$$

Then, $M_n - M_0$ is σ^2-subgaussian, $\sigma^2 = \frac{1}{4}\sum_{t=1}^{n} c_t^2$.

Pf: Tower rule: $\mathbb{E}\exp(\lambda M_n) = \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda \sum_1^n (M_t - M_{t-1})\right) \middle| \mathcal{F}_{n-1}\right]\right]$

$$= \mathbb{E}\left[\exp\left(\lambda \sum_1^{n-1} (M_t - M_{t-1})\right) \mathbb{E}\left[\exp\left(\lambda (M_n - M_{n-1})\right) \middle| \mathcal{F}_{n-1}\right]\right]$$

... some details to check, but can repeat argument conditionally, $\rightarrow$

$$\le \exp\left(\frac{\lambda^2}{2} \cdot \frac{c_n^2}{4}\right) \mathbb{E}\left[\exp\left(\lambda \sum_1^{n-1} (M_t - M_{t-1})\right)\right]$$

$$\le \cdots \le \exp\left(\frac{\lambda^2}{2} \cdot \frac{1}{4} \sum_1^n c_t^2\right).$$

Def: $f: \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$, $\delta_i(f) := \sup_{x, x_i'} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right|$.

Cor: (McDiarmid) If $X_1, \ldots, X_n$ any ind. r.v., then

$f(X_1, \ldots, X_n)$ is σ^2-subgaussian, $\sigma^2 = \sum \delta_i(f)^2$

④

Pf: Use Doob martingale: $M_i := \mathbb{E}[f(X_1, \dots, X_n) \mid X_1, \dots, X_i]$.
$M_0 = \mathbb{E}f(X)$, $M_n = f(X)$, so $f(X) - \mathbb{E}f(X) = M_n - M_0$.

$(X_1', \dots, X_n')$ ind. copy, $X^{(i)} := (X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots, X_n)$

$$\begin{aligned} M_i - M_{i-1} &= \mathbb{E}[f(X) \mid X_1, \dots, X_i] - \mathbb{E}[f(X) \mid X_1, \dots, X_{i-1}] \\ &= \mathbb{E}[f(X) \mid X_1, \dots, X_i] - \mathbb{E}[f(X^{(i)}) \mid X_1, \dots, X_{i-1}] \\ &= \mathbb{E}[f(X) - f(X^{(i)}) \mid X_1, \dots, X_i] \end{aligned}$$

$\Longrightarrow -\delta_i(f) \le M_i - M_{i-1} \le \delta_i(f)$. $\to$ use Azuma, $c_i = 2\delta_i(f)^2$.

Ex: (Balls + bins) Throw m balls into n bins unif at random.
$Z := \#\{\text{empty bins}\}$, $\mathbb{E}Z = \mathbb{E}\sum_i \mathbb{1}\{\text{bin } i \text{ empty}\} = n\left(1 - \frac{1}{n}\right)^m$.
$X_j :=$ index of bin of ball j, $j = 1, \dots, m$.
X_j indep., $Z = f(X_1, \dots, X_m)$, $\delta_j(f) \le 1$ (effect of moving one ball.)

$\Longrightarrow \mathbb{P}\left[\left|Z - n\left(1 - \frac{1}{n}\right)^m\right| > t\right] \le 2\exp\left(-\frac{t^2}{2m}\right)$ $\boxed{\text{(same as coupon collector)}}$

$\Longrightarrow Z = n\left(1 - \frac{1}{m}\right)^m \pm O(\sqrt{m})$ (very good if, e.g., $n \approx m$.)

Ex: (Chromatic number) $G =$ Erdos-Renyi random graph$\overset{n \text{ vertices}}{\vee}$ (each edge w/prob $1/2$).
$\omega(G) :=$ chromatic number $=$ min # vertex colors s.t. adjacent vertices have different colors. Combinatorial reasoning $\to$ $\mathbb{E}\,\omega(G) \sim \frac{n}{\log n}$.

$X_1 = \{\text{edges } \{1,i\}\}_{i>1}$ $X_2 = \{\text{edges } \{2,i\}\}_{i>2}$ etc. $\omega(G) = f(X_1, \dots, X_{n-1})$

$\delta_i(f) = 1$, since can always give vertex its own color. $\to \omega(G) = \frac{n}{\log n} \pm O(\sqrt{n})$.