Martingales + Concentration Inequalities

Concentration of measure: phenomenon that

$$\mathbb{P}\left\{ |f(X_1, \dots, X_n) - \mathbb{E} f(X_1, \dots, X_n)| > t \right\} \text{ "small", } \underline{\text{if}}$$

(1) f does not depend too much on any input

(2) X_i only weakly dependent.

Classical example:

Thm: (Hoeffding ineq) $X_1, \dots, X_n$ indep; $X_i \in [a_i, b_i]$ a.s.

$S_n := \sum_{i=1}^n X_i$ (i.e. $f(x_1, \dots, x_n) = \sum_i x_i$). Then,

$$\mathbb{P}\left\{ |S_n - \mathbb{E} S_n| > t \right\} \le 2 \exp\left(-\frac{t^2}{2\sigma^2}\right) \approx \mathbb{P}\left\{ |g| > t \right\}$$

$g \sim N(0, \sigma^2)$

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

Ex: SRW: $X_i \overset{iid}{\sim} \mathrm{Unif}(\{\pm 1\}) \rightsquigarrow \mathbb{E} S_n = 0$, $a_i = -1$, $b_i = +1$

$\rightsquigarrow \sigma^2 = n$

Thm $\Longrightarrow$ $\mathbb{P}\left\{ |S_n| > t \right\} \le 2 \exp\left(-\frac{t^2}{2n}\right)$.

i.e. $\mathbb{P}\left[\left|\frac{S_n}{\sqrt{n}}\right| > \underset{t/\sqrt{n}}{\underset{\shortparallel}{s}} \right\} \le 2 \exp\left(-\frac{s^2}{2}\right)$ "non-asymptotic CLT-like tail bound."

General pf technique: "Chernoff method", i.e. exponential Markov ineq.

$$\mathbb{P}[Y > t] = \mathbb{P}[\exp(\lambda Y) > \exp(\lambda t)]$$

$$\overset{\text{(Markov)}}{\leq} \frac{\mathbb{E}\exp(\lambda Y)}{\exp(\lambda t)} \leftarrow \phi_Y(\lambda)$$

, calculate / bound numerator, optimize param. λ.

Conveniently, numerator (mgf) factorizes over ind. sums.

$$Y, Z \text{ ind.} \Rightarrow \mathbb{E}\exp(\lambda(Y+Z)) = \mathbb{E}\exp(\lambda Y)\exp(\lambda Z)$$

$$= (\mathbb{E}\exp(\lambda Y))(\mathbb{E}\exp(\lambda Z)).$$

Def: (mgf) $\phi_X(\lambda) := \mathbb{E}\exp(\lambda(X - \mathbb{E}X))$

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

"variance proxy"

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

Prop 1: X, Y indep, X is σ^2-subgauss, Y is τ^2-subgauss

$\Rightarrow X + Y$ is $(\sigma^2 + \tau^2)$-subgauss

Prop 2: If X is σ^2-subgauss., then $\mathbb{P}[|X - \mathbb{E}X| > t] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$

Pf: Chernoff method: $\mathbb{P}[X - \mathbb{E}X > t] \leq \dfrac{\phi_X(\lambda)}{\exp(\lambda t)}$

$$\overset{\text{subgauss.}}{\leq} \exp\left(\frac{\sigma^2}{2}\lambda^2 - t\lambda\right) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right). \quad \blacksquare$$

Minimize over $\lambda \rightsquigarrow \lambda^* = \frac{t}{\sigma^2}$

$\boxed{\text{Lem}}$: (Hoeffding) $X \in [a, b]$ a.s. $\Rightarrow X$ is $\frac{1}{4}(b-a)^2$-subgauss

Pf of Thm: Lem $\Rightarrow X_i$ is $\frac{1}{4}(b_i - a_i)^2$ subgauss, Prop 1 $\Rightarrow S_n$ is σ^2-subgauss, Prop 2 $\Rightarrow$ result.

<u>Def</u>: (cgf) $\psi_X(\lambda) := \log \phi_X(\lambda) = \log \mathbb{E}\exp(\lambda(X-\mathbb{E}X))$.

<u>Prop</u>: ψ_X smooth, $\psi_X''(\lambda) \le \sigma^2 \ \forall \lambda \in \mathbb{R} \Rightarrow X$ is σ^2-subgauss.

<u>Pf</u>: $\psi_X(0) = \log \mathbb{E}\exp(0) = \log 1 = 0$.

$$\psi_X'(0) = \frac{\mathbb{E}(X-\mathbb{E}X)\cancelto{1}{\exp(\lambda(X-\mathbb{E}X))}}{\cancelto{1}{\mathbb{E}\exp(\lambda(X-\mathbb{E}X))}}\Bigg|_{\lambda=0} = 0.$$

Exc: Taylor thm w/ remainder $\Longrightarrow \psi_X(\lambda) \le \sigma^2 \cdot \frac{\lambda^2}{2}$

$$\Longrightarrow \phi_X(\lambda) \le \exp\left(\sigma^2 \cdot \frac{\lambda^2}{2}\right).$$

<u>Pf of H. lem</u>: WLOG $\mathbb{E}X = 0$. Derivatives of ψ_X:

$$\psi_X(\lambda) = \log \mathbb{E}\exp(\lambda X)$$

$$\psi_X'(\lambda) = \mathbb{E}X\underbrace{\frac{\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}}_{\rho_\lambda(X)} = \mathbb{E}X\rho_\lambda(X), \quad \rho_\lambda(X) = \frac{\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}$$

$$= \mathbb{E}_{X\sim\mu_\lambda} X$$

$\rho_\lambda(X)$ is a <u>density</u> w.r.t. Law(X), $\rho_\lambda = \frac{d\mu_\lambda}{d\,\mathrm{Law}(X)}$

(<u>Ex</u>: Suppose X has density $\pi(x)$.

$$\mathbb{E}X\rho_\lambda(X) = \int x \underbrace{\rho_\lambda(x)\pi(x)}_{\text{density of } \mu_\lambda} dx.)$$

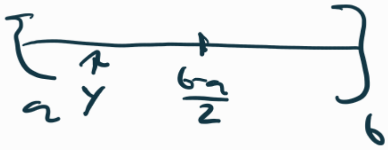
$$\psi_X''(\lambda) = \mathbb{E}X^2\underbrace{\frac{\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}}_{\rho_\lambda(X)} - \Bigg(\mathbb{E}X\underbrace{\frac{\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}}_{\rho_\lambda(X)}\Bigg)^2$$

$$= \mathbb{E}X^2\rho_\lambda(X) - (\mathbb{E}X\rho_\lambda(X))^2 = \mathbb{E}_{X\sim\mu_\lambda}X^2 - \left(\mathbb{E}_{X\sim\mu_\lambda}X\right)^2$$

$$= \mathrm{Var}_{X\sim\mu_\lambda}[X].$$

$\Rightarrow \psi_X''(\lambda) = \operatorname{Var} Y$ for $Y \sim \mu_\lambda \leftarrow$ "reweighted" law(X)

in particular $Y \in [a,b]$ a.s.



$$\operatorname{Var}[Y] = \operatorname{Var}\left[Y - \frac{b-a}{2}\right] \leq \mathbb{E}\left(Y - \frac{b-a}{2}\right)^2$$
$$\leq \left(\frac{b-a}{2}\right)^2 = \frac{1}{4}(b-a)^2.$$

$\Rightarrow \psi_X''(\lambda) \leq \sigma^2 := \frac{1}{4}(b-a)^2 \longrightarrow X$ is σ^2-subgauss. ∎

Surprisingly, can generalize to <u>martingales</u>!

<u>Thm</u>: (Azuma) $(M_n)_{n\geq 0}$ mtg. $(A_n), (B_n)$ predictable, $(c_n > 0)$ deterministic. If

$$\left.\begin{aligned} A_t &\leq M_t - M_{t-1} \leq B_t \quad \text{a.s.} \\ B_t - A_t &\leq c_t \quad \text{a.s.} \end{aligned}\right| \ \forall t = 1, \dots, n$$

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

(Hoeffding for mtg.)

<u>Pf</u>: WLOG $M_0 = 0$.

$$\phi_{M_n}(\lambda) = \mathbb{E}\exp(\lambda M_n)$$
$$= \mathbb{E}\exp\left(\lambda \sum_{1}^{n}(M_t - M_{t-1})\right)$$
$$\overset{\text{(tower)}}{=} \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda \sum_{1}^{n}(M_t - M_{t-1})\right) \Big| \mathcal{F}_{n-1}\right]\right]$$

$$\overset{(\text{fact}'_n)}{=} \mathbb{E}\left[\exp\left(\lambda \sum_{1}^{n-1}(M_t - M_{t-1})\right) \underbrace{\mathbb{E}\left[\exp\left(\lambda \underbrace{(M_n - M_{n-1})}_{\in [A_n, B_n) \text{ a.s.}}\right) \Big| \mathcal{F}_{n-1}\right]}_{\substack{\text{conditional Hoeffding Lem.} \\ \leq \exp\left(\frac{\lambda^2}{2} \cdot \frac{c_n^2}{4}\right)}}\right]$$

$$\leq \exp\left(\frac{\lambda^2}{2} \cdot \frac{c_n^2}{4}\right) \phi_{M_{n-1}}(\lambda)$$

-- Repeat n times: --

$$\leq \exp\left(\frac{\lambda^2}{2} \cdot \underbrace{\frac{1}{4} \sum_{1}^{n} c_n^2}_{\sigma^2}\right). \qquad \blacksquare$$

<u>Def</u>: $f : \mathcal{X}_1 \times \dots \times \mathcal{X}_n \to \mathbb{R} \rightsquigarrow$

$$\delta_i(f) := \sup_{\substack{x, x_i' \\ x = (x_1, \dots, x_n)}} \left| f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n) \right|$$

$\boxed{\underline{\text{Cor}}:}$ (McDiarmid ("bdd. differences")) If $X_1, \dots, X_n$ any indep. r.v. then $f(X_1, \dots, X_n)$ is σ^2-subgauss. where $\sigma^2 = \sum_{i=1}^{n} \delta_i(f)^2$ ($\approx$ weak version $\|\nabla f\|^2$)

<u>Pf</u>: Use <u>Doob martingale</u>:

$$M_i := \mathbb{E}[\underbrace{f(X_1, \dots, X_n)}_{f(X_1, \dots, X_n)} \mid X_1, \dots, X_i]$$

$M_0 = \mathbb{E} f(X) \qquad M_n = f(X) \qquad M_n - M_0 = f(X) - \mathbb{E} f(X).$

Towards using Azuma, need to control $M_i - M_{i-1}$.

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

(Introduce $X_1', \dots, X_n'$ iid copies of $X_1, \dots, X_n$

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

$$= \mathbb{E}[f(X) \mid X_1, \dots, X_i] - \underbrace{\mathbb{E}[f(X^{(i)}) \mid X_1, \dots, X_{i-1}]}_{= \mathbb{E}[f(X^{(i)}) \mid X_1, \dots, X_i]}$$

$$= \mathbb{E}[\underbrace{f(X) - f(X^{(i)})}_{\in [-\delta_i(f), \delta_i(f)]} \mid X_1, \dots, X_i]$$

$\Longrightarrow -\delta_i(f) \leq M_i - M_{i-1} \leq \delta_i(f)$

$\rightsquigarrow$ use Azuma where $c_i = 2\delta_i(f)$

$\rightarrow \sigma^2 = \frac{1}{4} \sum_1^n (2\delta_i(f))^2 = \sum_i \delta_i(f)^2.$ ∎

Ex: (Balls + bins (multinomial))

Throw m balls into n bins unif. at random.

$Z_m := \#\{\text{empty bins}\}$ (cf. "coupon collector")

$$\mathbb{E} Z = \mathbb{E} \sum_{i=1}^n \mathbb{1}\{\text{bin } i \text{ empty}\} = n \cdot \left(1 - \frac{1}{n}\right)^m$$

Goal: McDiarmid $\Longrightarrow$ Z concentrates around $\mathbb{E} Z$.

$X_j \in [n]$, $X_j :=$ index of bin ball j landed in, $j = 1, \dots, m$.

X_j indep. $Z = f(X_1, \ldots, X_m)$

$\left(= \#\{ i \in [n] : \text{no } X_j = i \}. \right)$

$\delta_j(f) \leq 1$ (# empty bins changes by ≤ 1
if you move one ball.)

$$\sigma^2 = \sum_{j=1}^{m} (\delta_j(f))^2 = m.$$

$$\Rightarrow \mathbb{P}\left\{ \left| Z - n\left(1 - \frac{1}{n}\right)^m \right| > t \right\} \leq 2\exp\left(-\frac{t^2}{2m}\right)$$

$$\Rightarrow Z = n\left(1 - \frac{1}{n}\right)^m \pm O(\sqrt{m})$$

(e.g. if $m = n$, $\to \infty$:

$$= \frac{n}{e} \pm O(\sqrt{n}).$$