Martingales + Concentration of Measure:

Concentration: general phenomenon that if $X_1, \dots, X_n$ only "weakly dependent", and $f(x_1, \dots, x_n)$ "weakly sensitive" to each input, then

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

One of main classical examples:

Thm: (Hoeffding ineq.) $X_1, \dots, X_n$ independent, $X_i \in [a_i, b_i]$ a.s. $\forall i$, $S_n := \sum_{i=1}^n X_i$ (i.e. $f(x_1, \dots, x_n) := \sum_{i=1}^n x_i$). Then,

$$\mathbb{P}\left[|S_n - \mathbb{E}S_n| > t\right] \leq \underbrace{2\exp\left(-\frac{t^2}{2\sigma^2}\right)}_{\approx \text{ tail of } N(0,\sigma^2)}, \quad \sigma^2 = \frac{1}{4}\sum_{i=1}^n (b_i - a_i)^2.$$

Ex: (SRW) $X_i \overset{iid}{\sim} \text{Unif}(\{\pm 1\}) \longrightarrow \sigma^2 = n$, $\mathbb{E}S_n = 0$, can rewrite Hoeffding:

$$\mathbb{P}\left[\left|\frac{1}{\sqrt{n}} S_n\right| > t\right] \leq \underbrace{2\exp\left(-\frac{t^2}{2}\right)}_{\approx \text{ tail of } N(0,1)} \quad \longrightarrow \quad \text{captures } S_n/\sqrt{n} \approx N(0,1) \text{ (in tails)}, \ \text{Var}(S_n/\sqrt{n}) = 1$$

Proof technique: "Chernoff method": exponential Markov ineq.

$$\mathbb{P}[S \geq t] = \mathbb{P}[\exp(\lambda S) \geq \exp(\lambda t)] \leq \frac{\boxed{\mathbb{E}\exp(\lambda S)}}{\exp(\lambda t)} = \frac{\phi_S(\lambda)}{\exp(\lambda t)} = \exp(\psi_S(\lambda) - \lambda t)$$

Def: $\phi_X(\lambda) := \mathbb{E}\exp(\lambda X)$ (moment generating function)
$\psi_X(\lambda) := \log \phi_X(\lambda)$ (cumulant generating function)

Prop: X, Y indep $\Rightarrow \phi_{X+Y}(\lambda) = \phi_X(\lambda)\phi_Y(\lambda)$, $\psi_{X+Y}(\lambda) = \psi_X(\lambda) + \psi_Y(\lambda)$.

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

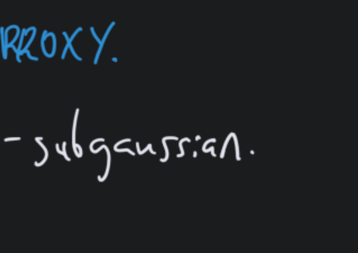
Prop: $X \sim N(0,\sigma^2) \Rightarrow \phi_X(\lambda) = \exp\left(\frac{\sigma^2\lambda^2}{2}\right)$.

Prop: X, Y indep., X σ^2-subgauss., Y τ^2-subgauss. $\Rightarrow X+Y$ is $(\sigma^2+\tau^2)$-subgauss.

Lem: If X σ^2-subgaussian $\Rightarrow \mathbb{P}[|X - \mathbb{E}X| > t] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right)$.

Pf: Chernoff method:

$$\mathbb{P}[X - \mathbb{E}X > t] \leq \frac{\phi_{X-\mathbb{E}X}(\lambda)}{\exp(\lambda t)} \leq \exp\Big(\underbrace{\frac{\sigma^2\lambda^2}{2} - \lambda t}_{\text{quadratic in } \lambda}\Big)$$



Optimal $\lambda^* = \frac{t}{\sigma^2} \rightsquigarrow \quad \leq \exp\left(\frac{-t^2}{2\sigma^2}\right)$.

Remains: show that $X \in [a,b]$ then X σ^2-subgaussian.

Lem: (Hoeffding) If $X \in [a,b]$ a.s., then X is $\underbrace{\frac{1}{4}(b-a)^2}_{\text{VARIANCE PROXY.}}$-subgaussian.

Prop: If $X \in [a,b]$ a.s., then $\text{Var}(X) \leq \frac{1}{4}(b-a)^2$.

Pf: $\text{Var}(X) = \text{Var}\left(X - \frac{a+b}{2}\right) \leq \mathbb{E}\left(X - \frac{a+b}{2}\right)^2 \leq \left(\frac{b-a}{2}\right)^2 = \frac{1}{4}(b-a)^2$.

Prop: If $\psi_{X-\mathbb{E}X}(\lambda)$ smooth, $\psi''_{X-\mathbb{E}X}(\lambda) \leq \sigma^2 \ \forall \lambda \in \mathbb{R} \Rightarrow X$ σ^2-subgaussian.

Pf: Idea: Taylor @ $\lambda = 0$. Recall: $\psi_{X-\mathbb{E}X}(\lambda) = \log \mathbb{E}\exp(\lambda(X - \mathbb{E}X))$.

$\psi_{X-\mathbb{E}X}(0) = \log(1) = 0$. $\psi'_{X-\mathbb{E}X}(0) = \left.\frac{\mathbb{E}(X-\mathbb{E}X)\cancel{\exp(\lambda(X-\mathbb{E}X))}}{\cancel{\mathbb{E}\exp(\lambda(X-\mathbb{E}X))}}\right|_{\lambda=0} = 0$.

Taylor (w/ Lagrange remainder) $\Rightarrow \forall \lambda \ \exists \theta$ s.t. $\psi_{X-\mathbb{E}X}(\lambda) = \cancel{\psi(0)} + \lambda\cancel{\psi'(0)} + \frac{\lambda^2}{2}\psi''(\theta) \leq \frac{\sigma^2\lambda^2}{2}$.

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

$$\psi'_X(\lambda) = \frac{\mathbb{E}X\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}, \quad \psi''_X(\lambda) = \frac{\mathbb{E}X^2\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)} - \left(\frac{\mathbb{E}X\exp(\lambda X)}{\mathbb{E}\exp(\lambda X)}\right)^2$$

$\rho(x) := \frac{\exp(\lambda x)}{\mathbb{E}\exp(\lambda X)}$ $\quad = \mathbb{E}X\rho(X) = \mathbb{E}Y$ $\quad = \mathbb{E}X^2\rho(X) - (\mathbb{E}X\rho(X))^2 = \text{Var}(Y)$

(✪) ρ is density w.r.t. $\text{Law}(X) =: \mu$ (i.e. $\rho \geq 0$, $\int \rho \, d\mu = 1 = \mathbb{E}\rho(X)$.)

Y = "tilting" or "reweighting" of X. $Y \in [a,b]$ a.s.

$\psi''_X(\lambda) = \text{Var}(Y) \leq \frac{1}{4}(b-a)^2$. ∎

Pf: (of H. ineq): $X_i \in [a_i, b_i]$ is $\frac{1}{4}(b_i - a_i)^2$-subgauss $\Rightarrow S := \sum X_i$ is $\left(\sum \frac{1}{4}(b_i - a_i)^2\right)$-subgaussian.

Generalization to martingales:

Thm: (Azuma) $(M_n)_{n\geq 0}$ martingale, (A_n), (B_n) predictable, $c_n > 0$ s.t.:

$$A_i \leq M_i - M_{i-1} \leq B_i \quad \text{and} \quad B_i - A_i \leq c_i \text{ a.s.} \quad (\text{w.r.t. filtration } (F_n)_{n\geq 0}.)$$

Then, $M_n - M_0$ is σ^2-subgauss, $\sigma^2 = \frac{1}{4}\sum_{i=1}^n c_i^2$.

Pf: Tower rule:

$$\mathbb{E}\exp(\lambda(M_n - M_0)) = \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda\sum_1^n (M_i - M_{i-1})\right) \Big| F_{n-1}\right]\right]$$

$$= \mathbb{E}\Big[\exp\Big(\lambda\underbrace{\sum_1^{n-1}(M_i - M_{i-1})}_{M_{n-1} - M_0}\Big)\underbrace{\mathbb{E}\left[\exp(\lambda(M_n - M_{n-1})) | F_{n-1}\right]}_{\text{conditional Hoeffding lemma.}}\Big]$$

$$\leq \dots \leq \exp\left(\frac{\lambda^2}{2} \cdot \frac{c_n^2}{4}\right)\mathbb{E}\left[\exp(\lambda(M_{n-1} - M_0))\right]$$

$$\leq \text{-- repeat } (n-1) \text{ times --} \leq \exp\left(\frac{\lambda^2}{2} \cdot \sigma^2\right). \ ∎$$

Ex: Enough to have $|M_i - M_{i-1}| \leq b_i$ a.s. $\rightsquigarrow A_i = -b_i$, $B_i = b_i \longrightarrow c_i = 2b_i$ $\longrightarrow \sigma^2 = \sum b_i^2$.

Concrete version of general concentration principle:

Def: $f : \mathcal{X}_1 \times \dots \times \mathcal{X}_n \longrightarrow \mathbb{R}$ ($\mathcal{X}_i$ ↳ sets), $\delta_i(f) := \sup\{|f(x) - f(x')| : x, x' \text{ only differ in coord } i\}$.

Cor: (Slightly weak McDiarmid) $X_1, \dots, X_n$ any indep. r.v. (not necessarily $\in \mathbb{R}$) f is s.t. $\delta_i(f) < \infty$, $\mathbb{E}|f(X_1, \dots, X_n)| < \infty$, then

$f(X_1, \dots, X_n)$ is σ^2-subgauss, where $\sigma^2 = \sum_{i=1}^n \delta_i(f)^2$.

$$\mathbb{P}\left[|f(X_1, \dots, X_n) - \mathbb{E}f(X_1, \dots, X_n)| > t\right] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Pf: Use Doob martingale: $M_i := \mathbb{E}[f(X_1, \dots, X_n) | X_1, \dots, X_i]$ (i.e., $F_i = \sigma(X_1, \dots, X_i)$.)

$M_0 = \mathbb{E}f(X_1, \dots, X_n)$, $M_n = f(X_1, \dots, X_n)$
$\Rightarrow M_n - M_0 = f - \mathbb{E}f$.

Just need: $|M_i - M_{i-1}| \leq \delta_i(f)$ a.s.

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

Rk: $\text{Law}(X) = \text{Law}(X^{(i)})$.

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

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

Use Azuma w/ $c_i = 2b_i = 2\delta_i(f)$.

Ex: (Balls + bins (coupon collector)) Throw m balls into n bins unif at random independently. $Z := \#\{\text{empty bins}\} = \sum_{i=1}^n \mathbb{1}\{\text{bin } i \text{ empty}\}$

$$\mathbb{E}Z = n\mathbb{P}[\text{bin 1 empty}] = n\left(\frac{n-1}{n}\right)^m = n\left(1 - \frac{1}{n}\right)^m$$

e.g. $n = cm$, $m \to \infty$, $\mathbb{E}Z \sim cm\exp\left(-\frac{1}{c}\right) = \frac{c}{e^{1/c}} \cdot (m)$.

X_j = index of bin ball j lands in ($\in \{1, \dots, n\}$, for $j = 1, \dots, m$).

$Z = f(X_1, \dots, X_m)$ $\quad \delta_j(f) = 1 \rightsquigarrow \sigma^2 = m$.

$$\Rightarrow \mathbb{P}[|Z - \mathbb{E}Z| \geq t] \leq 2\exp\left(-\frac{t^2}{2m}\right).$$

Rk: No model-specific thinking!

e.g. w/prob ≥ 0.999, $|Z - \underbrace{\mathbb{E}Z}_{\Theta(m)}| \leq C\sqrt{m}$.