HDP Session 6

Recall

- 1. Doob's Ineq: If $(M_n)_n$ is subM, then

$$\mathbb{P}\left(\sup_{0\le s\le t} M_s \ge a\right) \le \frac{\mathbb{E} M_t}{a} \quad \forall\, a>0$$

- 2. Kolmogorov's Ineq: If $(M_n)_n$ Martingale and $\mathbb{E} M_n^2 < \infty$ and $\mathbb{E} M_n = 0$

$$\mathbb{P}\left(\sup_{0\le s\le t} |M_s| \ge a\right) \le \frac{\mathrm{Var}(M_t)}{a^2}.$$

3. Hoeffding's lemma: $\mathrm{Ran}(X) \subseteq [a,b]$, then $X \in sG\left(\frac{1}{4}(b-a)^2\right)$

---

1. Azuma-Hoeffding Inequality:

- Let $(M_n)_n$ be a Martingale and $\exists\, (A_n)_n, (B_n)_n$ predictable processes s.t. $A_n \le M_n - M_{n-1} \le B_n$ and $\exists$ constants $c_n > 0$ s.t. $B_n - A_n \le c_n$ a.s. Then

$$\mathbb{P}\left(\sup_{0\le i\le t}(M_i - M_0) \ge \beta\right) \le \exp\left(-\frac{2\beta^2}{\sum_{n=1}^{t} c_n^2}\right)$$

Proof:

$$\mathbb{P}\left(\sup_{0\le i\le t}(M_i - M_0) \ge \beta\right) = \mathbb{P}\left(\sup_{0\le i\le t} e^{\lambda(M_i - M_0)} \ge e^{\lambda\beta}\right)$$

$$\le e^{-\lambda\beta}\, \mathbb{E}\left[e^{\lambda(M_t - M_0)}\right]$$

$\mathcal{F}_i = X_1, X_2, \dots, X_i$

$$= e^{-\lambda\beta}\, \mathbb{E}\left[\mathbb{E}\left[\underbrace{e^{\lambda(M_{t-1}-M_0)}}_{\text{const.}} e^{\lambda(M_t - M_{t-1})} \,\Big|\, \mathcal{F}_{t-1}\right]\right]$$

$$= e^{-\lambda\beta}\, \mathbb{E}\left[e^{\lambda(M_{t-1}-M_0)}\, \underbrace{\mathbb{E}\left[e^{\lambda(M_t - \underset{\mathbb{E}(M_i|\mathcal{F}_{i-1})}{M_{t-1}})} \,\Big|\, \mathcal{F}_{t-1}\right]}_{\le e^{\frac{\lambda^2 c_t^2}{8}}}\right]$$

(Hoeffding's lemma)

$$\le e^{-\lambda\beta}\, e^{\frac{\lambda^2 c_t^2}{8}}\, \mathbb{E}\left[e^{\lambda(M_{t-1} - M_0)}\right]$$

(By Induction) $\le e^{-\lambda\beta}\, e^{\lambda^2 \frac{\sum c_n^2}{8}}$ ✓

Optimising: $\lambda = \dfrac{4\beta}{\sum c_n^2}$, we have

$$\mathbb{P}(M_t - M_0 \geq \beta) \leq \mathbb{P}\left(\sup_{0\le i\le t} M_i - M_0 \geq \beta\right) \leq \exp\left(-\frac{2\beta^2}{\sum c_n^2}\right) \quad ▨$$

(Range $(X_i) = \mathcal{X}_i$)

Our purpose: $\mathbb{P}\left(f(\underset{\sim}{X}) - \mathbb{E} f(\underset{\sim}{X}) \geq \beta\right) \leq o(1)$.

$X_1, X_2, \dots, X_n$ are independent

2. Method of Bounded Difference:

Let $M_i = \mathbb{E}\left[f(\underset{\sim}{X}) \mid \mathcal{F}_i\right]$ – martingale $\qquad \underset{\sim}{X} = (X_1, \dots, X_n)$

$$\rightarrow f(X) - \mathbb{E} f(X) = \sum_{i=1}^{n} \underline{\underline{(M_i - M_{i-1})}} \qquad \begin{cases} M_n = f(X) \\ M_0 = \mathbb{E} f(X) \end{cases}$$

⓪ Def$^{\underline{n}}$: $D_i f$ is a random variable

$$D_i f(x_1, \dots, \hat{x}_i, \dots, x_n) = \operatorname*{ess\,sup}_{y\in\mathcal{X}_i} f(x_1, \dots, y, \dots, x_n) - \operatorname*{ess\,inf}_{y\in\mathcal{X}_i} f(x_1, \dots, y, \dots, x_n)$$

Random var. $\dashv \{X_1, X_2, \dots, X_n\} \setminus X_i$ $\qquad$ [ignore $\underline{\underline{\text{measure}}}$ zero events]

• $$M_i - M_{i-1} = \mathbb{E}[f(\underset{\sim}{X}) \mid \mathcal{F}_i] - \mathbb{E}[\underbrace{f(\underset{\sim}{X}) \mid \mathcal{F}_{i-1}}]$$

$$= \mathbb{E}[f(\underset{\sim}{X}) \mid \mathcal{F}_i] - \mathbb{E}[f(X^{(i)} \mid \mathcal{F}_i)]$$

$$\longrightarrow \quad = \underline{\mathbb{E}\left[f(X) - f(X^{(i)}) \mid \mathcal{F}_i\right]}$$

$\mathbb{E}\left[f(X_1, \dots, X_{i-1}, X_i, X_{i+1}, \dots) \mid X_1, \dots, X_{i-1}\right]$

$\qquad \parallel$

$\mathbb{E}\left[f(\underbrace{X_1, \dots, X_{i-1}, X_i', X_{i+1}, \dots}_{X^{(i)}}) \mid \underbrace{X_1, \dots, X_n}_{\mathcal{F}_i}\right]$

$X_i' \overset{d}{=} X_i$ and $X_i \perp\!\!\!\perp X_i'$.

• $$|M_i - M_{i-1}| \leq \mathbb{E}\left[|f(X) - f(X^{(i)})| \mid \mathcal{F}_i\right]$$

$$\leq \operatorname{ess\,sup} |D_i f(x_1, \hat{x}_i, x_n)| = \| D_i f\|_\infty$$

$$-\|D_i f\|_\infty \leq M_i - M_{i-1} \leq \|D_i f\|_\infty$$

[H.W.: Improve by a factor of 4]

$$\Rightarrow \boxed{\mathbb{P}\left(f(X) - \mathbb{E} f(X) \geq \beta\right) \leq \exp\left(- \frac{2\beta^2}{\cancel{2}\sum \|D_i f\|_\infty^2}\right)} \quad ▨$$

McDiarmid's Inequality

Barely any Improvement.

- $\mathrm{Var}\left(f(X) - \mathbb{E}f(X)\right) \doteq \sum_{i=1}^{n} \mathbb{E}\left(\underline{M_i - M_{i-1}}\right)^2$

$$\boxed{\mathrm{Var}\, f(X) \leq \sum_{i=1}^{n} \| D_i f \|_\infty^2} \longrightarrow \text{Can be Improved.}$$

—

☆ Improving the variance bound:

$$\mathrm{Var}\left(f(X,Y)\right) \overset{(\text{ANOVA})}{=} \mathbb{E}\left[\mathrm{Var}\left(f(X,Y) \mid X\right)\right] + \mathrm{Var}\left[\mathbb{E}\left[f(X,Y) \mid X\right]\right].$$

☆ **Lemma:** $\mathbb{E}\Big[\mathbb{E}\big[f(\underset{\sim}{X}) \mid \underbrace{X_1, \dots, \hat{X}_i, \dots, X_n}_{\mathcal{G}_i}\big] \,\Big|\, \underbrace{X_1, \dots, X_i}_{\mathcal{F}_i}\Big] = \mathbb{E}\left[f(\underset{\sim}{X}) \mid X_1, \dots, X_{i-1}\right]$

$$\therefore (M_i - M_{i-1})^2 = \left(\mathbb{E}[f(\underset{\sim}{X}) \mid \mathcal{F}_i] - \mathbb{E}[f(X) \mid \mathcal{F}_{i-1}]\right)^2$$

$$= \left(\mathbb{E}[f(X) \mid \mathcal{F}_i] - \mathbb{E}\left[\mathbb{E}[f(X) \mid \mathcal{G}_i] \mid \mathcal{F}_i\right]\right)^2$$

$$\leq \mathbb{E}\left(\left(f(X) - \mathbb{E}(f(X) \mid \mathcal{G}_i)\right)^2 \Big| \mathcal{F}_i\right)$$

$$\mathrm{Var}(f(X)) = \sum_{i=1}^{n} \mathbb{E}(M_i - M_{i-1})^2 \leq \sum_{i=1}^{n} \mathbb{E}\left(f(X) - \mathbb{E}[f(X) \mid \mathcal{G}_i]\right)^2 \;//$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\left(f(X) - \mathbb{E}(f(X) \mid \mathcal{G}_i)\right)^2 \Big| \mathcal{G}_i\right]\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(f(X) \mid \mathcal{G}_i\right)\right]$$

$$\mathrm{Var}(Y \mid X) := \mathbb{E}\left[\left(Y - \mathbb{E}(Y \mid X)\right)^2 \Big| X\right]$$

$$\boxed{\mathrm{Var}\, f(X) \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(f(X) \mid \mathcal{G}_i\right)\right]} \longrightarrow \text{Efron Stein Ineq.} \;\; \boxed{\text{ES1}}$$

Lemma: $\mathrm{Var}\, X = \frac{1}{2}\, \underline{\mathbb{E}(X - X')^2}$, X' is an indep copy of X

Proof:

$$\mathbb{E}(X - X')^2 = \mathbb{E}\left((X - \mathbb{E}X) - (X' - \mathbb{E}X')\right)^2$$

$$= \mathbb{E}(X - \mathbb{E}X)^2 + \mathbb{E}(X' - \mathbb{E}X')^2$$

$$= 2\,\mathrm{Var}(X).$$

$$\mathrm{Var}(f(X) \mid \mathcal{G}_i) = \frac{1}{2} \mathbb{E}\left((f(X) - f(X)^{(i)})^2 \mid \mathcal{G}_i \right)$$

$$\therefore \quad \mathrm{Var}\, f(X) \leq \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[(f(X) - f(x^{(i)}))^2 \mid \mathcal{G}_i \right] \right]$$

$$= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[(f(X) \underset{=}{-} f(x^{(i)}))^2 \right] \qquad [ES2]$$

<u>Bounded Difference Inequality:</u> $\quad |D_i f| < \infty$

$$\mathrm{Var}\, f(X) \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}(f(\underline{X}) \mid \mathcal{G}_i) \right]$$

$$\leq \frac{1}{4} \sum_{i=1}^{n} \mathbb{E}\left[(D_i f(X))^2 \right] \qquad \square$$

✪ <u>Quick Summary:</u>

exp bounds {

1. Azuma-Hoeffding: $\quad \mathbb{P}\left(\sup_{1 \leq i \leq t} M_i - M_0 \geq \beta \right) \leq \exp\left(- \frac{2\beta^2}{\sum c_n^2} \right)$

2. McDiarmid's Ineq: $\quad \mathbb{P}\left(f(X) - \mathbb{E} f(X) \geq \beta \right) \leq \exp\left(- \frac{2\beta^2}{\underbrace{\sum_{i=1}^{n} \|D_i f\|_\infty^2}} \right)$

3. Obvious: $\quad \mathrm{Var}\, f(X) \leq \sum_{i=1}^{n} \|D_i f\|_\infty^2$

ES: $\quad \mathrm{Var}\, f(X) \leq \sum_{i=1}^{n} \underbrace{\mathbb{E}\left[\overline{\mathrm{Var}(f(X) \mid \mathcal{G}_i)} \right]} = \frac{1}{2} \sum_{i=1}^{n} \mathrm{Var}\left((f(X) - f(x^{(i)}))^2 \mid \mathcal{G}_i \right)$

→ Bounded diff: $\quad \mathrm{Var}\, f(X) \leq \frac{1}{4} \sum_{i=1}^{n} \underset{=}{\mathbb{E}}\left[\underline{(D_i f(X))^2} \right].$ ← ✪

★ Examples

1\. **Balls and Bins** — Suppose we have m - balls, n - bins

$$Z_{n,m} = \#\text{ empty bins}$$

$$\underbrace{\mathbb{E}\, Z_{n,m}} = \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}_{B_i}\right) = \sum_{i=1}^{n}\left(1-\frac{1}{n}\right)^m = n\left(1-\frac{1}{n}\right)^m$$

Generating r.v.s: $X_i = j$ if i^{th} ball lands in B_j

$$Z_{n,m} = Z_{n,m}(X_1, X_2, \dots, X_m)$$

- $\| D_i Z_{n,m}(X_1,\dots,X_m)\|_\infty \le 2 \longrightarrow$ ($Z_{n,m}$ does not depend too much of any of the X_i's)

$$\mathbb{P}\left(\, |Z_{n,m} - \mathbb{E}\, Z_{n,m}| \geq \beta \right) \leq 2\exp\left(-\frac{2\beta^2}{\sum_{i=1}^{m}\|D_i f\|_\infty^2}\right)$$

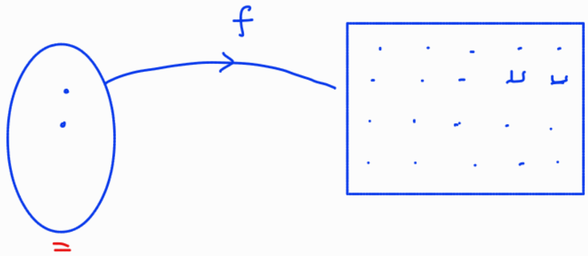
$$\Rightarrow \mathbb{P}\left(\, \Big| Z_{n,m} - \underbrace{n\left(1-\frac{1}{n}\right)^m}_{\to\, n\cdot\frac{1}{e}} \Big| \geq \underbrace{\beta\sqrt{m}}_{\sqrt{n}} \right) \leq 2\exp\left(-\frac{\beta^2}{2}\right).$$

- $m \asymp n$

$\underline{m = n^2}$

Hashing:

$z_1, z_2, \dots$



H.W. Find $\mathbb{P}(\text{collisions})$ and $\forall\, m \in \mathbb{N}$, find n s.t. $\mathbb{P}(\text{collisions}) < \frac{1}{2}$.

$|S| = s$

2. $\underbrace{X_1, X_2}_{\underset{\sim}{a}}, X_3, \underbrace{- - - - -}_{\underset{\sim}{a}}, X_n$ — random in a set S, $(a_1, a_2, \ldots, a_k)$ — given string $\in S^k$

S — Set of alphabets

- $\mathbb{P}(\underset{\sim}{X} \text{ contains } \underset{\sim}{a}) \longleftarrow$ Conc. of # occurences of $\underset{\sim}{a}$ in $\underset{\sim}{X}$.

- E_i = Event that $(X_i \ldots, X_{i+k-1}) = \underset{\sim}{a}$

$$N = \sum_{i=1}^{n-k+1} \mathbb{1}_{E_i} \longleftarrow \text{Could be highly correlated}$$

$$\mathbb{E}N = (n-k+1)\left(\frac{1}{s}\right)^k$$

Consider $\| D_i N(X_1, \ldots, X_n) \|_\infty \leq k$ (local corr)

$$\mathbb{P}\left((N - \mathbb{E}N) \geq \beta k \sqrt{n} \right) \leq e^{-2\beta^2}. \qquad \square$$

[H.W.] Q: We have a random string of length n. For what k is a given seq. $\mathbb{P}\left((a_1, \ldots, a_k) \in \underset{\sim}{X} \right) \geq \varepsilon$.

$\varnothing$ Conc. of measure of $\{0,1\}^n$.