Recap: Chernoff-Hoeffding Bounds

$Z = X_1 + \cdots + X_n$ $\quad$ $X_1 \dots X_n$ mutually independent

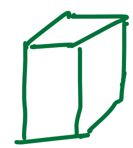
$X_i \sim \text{Bernoulli}(p_i)$

$\mu = \mathbb{E}[Z] = \sum_i \mathbb{E}[X_i] = \sum_i p_i$

- (Chernoff/Hoeffding)

$$\left.\begin{aligned} &- \forall \delta > 0 \quad \mathbb{P}[Z \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \\ &- \forall \delta \in (0,1) \quad \mathbb{P}[Z \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \end{aligned}\right\} \leq e^{-\delta^2\mu/3} \text{ for } \delta \in (0,1)$$

Application: Load Balancing



Assign jobs (randomly) so that none of the machines have high load.

$E_j \equiv$ Machine j has load $\geq t$

Bound $\mathbb{P}\left(\bigcup_j E_j\right)$

Ex: $\mathbb{P}[A \cup B] \leq \mathbb{P}[A] + \mathbb{P}[B]$

Analyzing for $m = n$

$$X_{ij} = \begin{cases} 1 & \text{if job } i \text{ goes to machine } j \\ 0 & \text{o.w.} \end{cases} \Big\} \text{ Bernoulli}(1/n)$$

$$Z_j = \#\text{ jobs on machine } j = \sum_i X_{ij}$$

$$\mathbb{E}[Z_j] = n \cdot \frac{1}{n} = 1$$

$$\mathbb{P}[Z_j \geq t] = \mathbb{P}[Z_j \geq (1 + (t-1)) \cdot 1] \leq \frac{e^{t-1}}{t^t}$$

$$\mathbb{P}[\exists j \; Z_j \geq t] \leq n \cdot \frac{1}{e} \cdot \left(\frac{e}{t}\right)^t$$

$$\leq n \cdot \frac{1}{e} \cdot \left(\frac{1}{n}\right)^2 \quad \text{for} \quad t = \frac{3 \ln n}{\ln \ln n}$$

Power of two choices [Mitzenmacher]: $t = O(\ln \ln n)$

Random 3-SAT

Max 3-SAT : $\varphi \equiv C_1, C_2, \ldots C_m$ clauses in n vars

$$C_i = (\ell_{i_1} \vee \ell_{i_2} \vee \ell_{i_3})$$

ℓ_{i_1} Can be a variable x_j or neg. $\bar{x}_j$.

e.g. $(x_1 \vee \bar{x}_9 \vee x_2), (x_5 \vee \bar{x}_2 \vee x_3), (\bar{x}_4 \vee \bar{x}_1 \vee \bar{x}_3)$

Goal : Assign $x_1 \ldots x_n \in \{\text{True}, \text{False}\}$

To satisfy maximum no. of $C_1 \ldots C_m$

- For a random 3-SAT instance φ

(with $m \geq C \cdot n$ clauses for $C \geq 10/\epsilon^2$)

$$P[\text{optimal assignment satisfies} \geq (\tfrac{7}{8} + \epsilon) \cdot m \text{ clauses}] \leq e^{-n}$$

Defining a random instance

Fix : $S_1, \ldots \ldots, S_m \subseteq [n]$ $\quad S_i = \{i_1, i_2, i_3\}$ (say)

$$\ell_{i_j} = \begin{cases} x_{i_j} & \text{w.p. } 1/2 \\ \bar{x}_{i_j} & \text{w.p. } 1/2 \end{cases} \qquad C_i = (\ell_{i_1} \vee \ell_{i_2} \vee \ell_{i_3})$$

Fix an assignment $\alpha \in \{0,1\}^n \equiv \{\text{true}, \text{false}\}^n$

$$Z_\alpha = \sum_{i=1}^{m} Y_{i,\alpha} \; \} \text{ assignment } \alpha \text{ satisfies } i^{th} \text{ clause}$$

$$\mathbb{E}[Z_\alpha] = \sum_i \mathbb{E}[Y_{i,\alpha}] = m \cdot \frac{7}{8}$$

Union bounds

$$Z_\alpha = \sum_{i=1}^{m} Y_{i,\alpha} \Big\}\ \text{assignment } \alpha \text{ satisfies } i^{th} \text{ clause}$$

$$P\left[Z_\alpha \geq \left(\frac{7}{8}+\epsilon\right)m\right] = P\left[Z_\alpha \geq \left(1+\epsilon\cdot\frac{8}{7}\right)\cdot\frac{7m}{8}\right]$$

$$\leq e^{-\epsilon^2\cdot\frac{8}{7}\cdot\frac{m}{3}} \leq e^{-2n} \quad \text{for } m \geq \frac{10}{\epsilon^2}$$

$$P\left[\exists \text{ assignment } \alpha \text{ s.t. } Z_\alpha \geq \left(\frac{7}{8}+\epsilon\right)m\right]$$

$$\leq 2^n \cdot e^{-2n} \leq e^{-n}$$

Interlude : 0/1 random variables vs ± 1 variables

$$Z = Y_1 + \cdots\cdots + Y_n \qquad \underbrace{Y_i = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}}_{\text{Rademacher r.v.s}} \quad \text{i.i.d.}$$

$$\mathbb{P}[\,|Z| \geq t\,] = ?$$

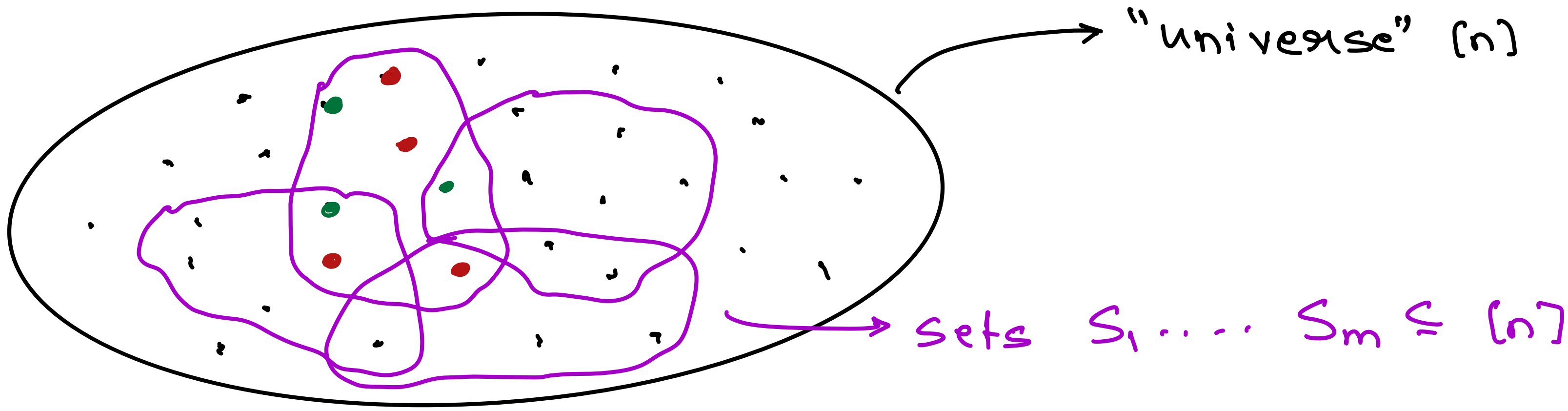
Consider $x_i = \frac{1-Y_i}{2}$ $\qquad x_i \sim \text{Bernoulli}(1/2)$

Say $Z' = x_1 + \cdots\cdots + x_n$

$$\mathbb{P}[\,|Z| \geq t\,] = \mathbb{P}\left[\,\left|Z' - \frac{n}{2}\right| \geq \frac{t}{n} \cdot \frac{n}{2}\right]$$

$$\leq 2 \cdot e^{-(t^2/n^2) \cdot \frac{n}{2} \cdot \frac{1}{3}} = 2 \cdot e^{-t^2/6n}$$

Balanced colorings (minimizing discrepancy)



Goal : Color points in $[n] = \{1, \dots, n\}$ by red/green

s.t. each S_i is as "balanced" as possible

$$x_j = \begin{cases} 1 & \text{if } j^{th} \text{ point is red} \\ -1 & \text{if } j^{th} \text{ point is green} \end{cases}$$

$$Z_i = \sum_{j \in S_i} x_j$$

$|Z_i| =$ "imbalance"

Balancing all sets

$$\mathbb{P}[|Z_i| \geq t] \leq 2 \cdot e^{-t^2/6 \cdot |S_i|} \leq 2 \cdot e^{-t^2/6n}$$

$$\mathbb{P}[\exists i \;\; |Z_i| \geq t] \leq n \cdot 2 \cdot e^{-t^2/6n}$$

$\therefore$ $\exists$ coloring with maximum imbalance $\leq \sqrt{12 n \ln n}$

Alternate forms : Hoeffding's inequality

- For $X_1 \dots X_n$ i.i.d. Rademacher variables

$$\mathbb{P}\left[\,\left|\sum a_i X_i\right| \geq t\,\right] \leq 2 \cdot e^{-t^2/2\|a\|^2}$$

<u>Proof:</u>

$$\mathbb{P}\left[\sum a_i X_i \geq t\right] = \mathbb{P}\left[e^{\lambda \sum a_i X_i} \geq e^{\lambda t}\right]$$

$$\leq \frac{\mathbb{E}\, e^{\lambda \sum a_i X_i}}{e^{\lambda t}} = \frac{\prod_i \mathbb{E}\, e^{\lambda a_i X_i}}{e^{\lambda t}}$$

$$\mathbb{E}\, e^{\lambda a_i X_i} = \frac{1}{2}\left(e^{\lambda a_i} + e^{-\lambda a_i}\right) \leq e^{\lambda^2 a_i^2/2}$$

$$\mathbb{P}\left[\sum a_i X_i \geq t\right] \leq \frac{\prod_i e^{\lambda^2 a_i^2/2}}{e^{\lambda t}} = \frac{e^{\lambda^2 \|a\|^2/2}}{e^{\lambda t}}$$

Optimal $\lambda = \dfrac{t}{\|a\|^2}$

Hoeffding for bounded random variables

▷ Let $X_1, \ldots, X_n$ be independent with $X_i \in [a_i, b_i]$

$$\mathbb{P}\left[\sum X_i - \mathbb{E}\left[\sum X_i\right] \geq t\right] \leq \exp\left(\frac{-2t^2}{\sum_i (b_i - a_i)^2}\right)$$

Ex: Prove above using $\mathbb{E}\left[e^{\lambda(X_i - \mathbb{E}[X_i])}\right] \leq \exp\left(\frac{\lambda^2 (b_i - a_i)^2}{8}\right)$