Discrepancy

Universe $[n]$

m sets $S_1, S_2, \dots, S_m \subseteq [n]$.

Goal: find a coloring $\chi : [n] \to \{-1, 1\}$ s.t.

$$\max_i |\chi(S_i)| \to \min \qquad \chi(S_i) = \sum_{j \in S_i} \chi(j).$$

Always can get n. $\qquad \|Ax\|_\infty \to \min \quad A \in \{0,1\}^{m \times n}$

Will be (mostly) considering $\qquad x \in \{-1,1\}^n$.

the case $m = n$.

Th Can get discrepancy $O(\sqrt{n \log n})$.

Pf random coloring.

Th $\exists$ set systems whose discrepancy is $\Omega(\sqrt{n})$.

Pf Consider a fixed coloring.

Assume $\frac{n}{2}$ $+1$'s, $\frac{n}{2}$ -1's, since this is the worst case.

$S_1, \dots, S_n$ - independent uniformly random.

$$\Pr\left[\forall i \ |\chi(S_i)| \le \frac{\sqrt{n}}{1000}\right] \le \left(\frac{1}{3}\right)^n.$$

$$\Pr_{S_i}\Big[\exists \chi\ \forall i\ \underbrace{|\chi(S_i)|}_{\le \frac{\sqrt{n}}{1000}}\Big] \ll 1.$$

whp discrepancy $\Omega(\sqrt{n})$.

Which bound is correct?

Can we do better than a random coloring?

Are there worse set systems than random?

Turns out we can improve an upper bound

Th Discrepancy is always $O(\sqrt{n})$.

$O\left(\sqrt{n \log \frac{m}{n}}\right)$ for the general case

(random gives $O(\sqrt{n \log m})$).

Today: non-constructive proof.

Thursday: algorithmic proof (lecture by Sami).

Partial coloring:

Enough to find $\chi: [n] \to \{-1, 0, 1\}$ with $|\{i : \chi(i) \neq 0\}| = \Omega(n)$.

s.t. $\max_i |\chi(S_i)| \le O\left(\sqrt{n \log \frac{m}{n}}\right)$.

Pf Can iterate.

$$1+\sqrt{\alpha\log\frac{1}{\alpha}}+\sqrt{2\alpha^2\log\frac{1}{\alpha}}+\sqrt{3\alpha^3\log\frac{1}{\alpha}}+\ldots\le O(1)$$

$\forall\, \alpha<1.$

We will use entropy of a discrete random variable.

Def $H(X)=\sum_i \Pr[X=i]\log_2\frac{1}{\Pr[X=i]}$.

Some facts: 1) $0\le H(X)\le \log_2|\operatorname{supp} X|$.

2) $H(X_1,\ldots,X_k)\le H(X_1)+\ldots+H(X_k)$

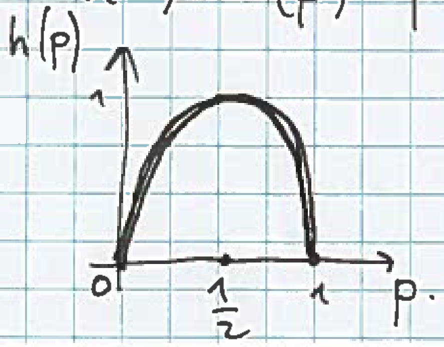
$=$ iff X_i's are independent.

3) If $\forall i$ $\Pr[X=i]\le\delta$, then $H(X)\ge\log_2\frac{1}{\delta}$.

Th To transmit n iid samples of X, necessary and sufficient $(H(X)+\bar{o}(1))\cdot n$ bits.

Example unfair coin

$$H(X)=h(p)=p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}$$



$h(0)=0$ $h(1)=0$ $h\left(\frac{1}{2}\right)=1$

$h\left(\frac{1}{2}+\eta\right)=1-\Theta(\eta^2)$.

Volume of a ball on the hypercube.

$\{-1,1\}^n$

$$|B(x,\alpha n)| \approx 2^{h(\alpha)n} \qquad h(\alpha) = \alpha \log\frac{1}{\alpha} + (1-\alpha)\log\frac{1}{1-\alpha}.$$

Follows from the Stirling formula.

Back to partial colorings.

Let $\mathcal{X}$ be a random coloring.
We will cover $m=n$.
b_i - nearest integer to $\dfrac{\mathcal{X}(S_i)}{100\sqrt{n}}$.

$\left(-\dfrac{\sqrt{n}}{100} \le b_i \le \dfrac{\sqrt{n}}{100}\right)$. $b_i = 0$ iff $-50\sqrt{n} \le \mathcal{X}(S_i) \le$
$\le 50\sqrt{n}$.

<u>Cl</u> $H(b_i) \le \varepsilon$ for a small constant ε.

<u>Pf</u>

$$\Pr[b_i = 0] \approx 1 - 2e^{-50^2/2}$$

$$\Pr[b_i=0]\log\frac{1}{\Pr[b_i=0]} \approx e^{-50^2/2}$$

$$\Pr[b_i=1] \le e^{-50^2/2}$$
$$\ge e^{-50^2/2} - e^{-150^2/2}$$

$$\Pr[b_i = 1] \log \frac{1}{\Pr[b_i = 1]} \leq e^{-50^2/2} \cdot \frac{50^2}{2}.$$

$$\Pr[b_i = 2] \log \frac{1}{\Pr[b_i = 2]} \approx e^{-150^2/2}$$

$\vdots$

$H(b_i)$ is small

$H(b_1, \dots, b_n) \leq \varepsilon n$

$\exists\, b_1^*, \dots, b_n^* \in \{\pm 1\}$ s.t.

$$\Pr_{x}[\forall i \quad b_i = b_i^*] \geq 2^{-\varepsilon n}.$$

$\#$ of x's that give b_i^* is $\geq 2^{(1-\varepsilon)n}$.

<u>Cl</u> $\exists\, x_1, x_2$ that give b_i^* s.t.

$$|\{i : x_1(i) \neq x_2(i)\}| \geq (1 - \delta(\varepsilon)) \cdot \frac{n}{2}.$$

<u>Pf</u> Follows from the volume bound

$$2^{H\left(\frac{1-\delta}{2}\right)n} << 2^{(1-\varepsilon)n} \quad \text{if } \delta \sim \sqrt{\varepsilon}.$$

$$\chi = \frac{\chi_1 - \chi_2}{2}$$

$$|\chi(S_i)| = \left| \frac{\chi_1(S_i) - \chi_2(S_i)}{2} \right| \leq O(\sqrt{n}) \quad \forall i.$$

$$\chi : [n] \longrightarrow \{-1, 0, 1\}$$

$$|\{i : \chi(i) \neq 0\}| = \Omega(n) \qquad \left(\geq (1-\delta) \cdot \frac{n}{2}\right).$$

Next time: how to find a good (partial) coloring efficiently.

Lots of open problems:

$\forall i \quad |\{j \mid i \in S_j\}| \leq t.$

<u>Th</u> discrepancy $\leq 2t-1$. $\Omega(\sqrt{t})$.

<u>Conj</u> $O(\sqrt{t})$.