Chernoff bounds and AMS

Last time :-

We assumed access to sk 'rows' of $8M$ bits each, drawn uniformly at random from $\{-1, +1\}$.

$u_{11} \ldots u_{1,8M} \leftarrow$ row

$u_{21} \ldots u_{2,8M}$

$\vdots$

$u_{sk,1} \ldots u_{sk,8M}$

We showed that if $s \geq \frac{16}{\varepsilon^2}$, then the estimator $Y_i = \left(\sum_{j=1}^{8M} f_j u_{ij}\right)^2$ satisfies

$$\Pr\left[|Y_i - F_2| \geq \varepsilon F_2\right] \leq \frac{1}{8}.$$

We now wish to choose k in such a way that the median of Y of $Y_1 \ldots Y_k$ satisfies

$$\Pr\left[|Y - F_2| \geq \varepsilon F_2\right] \leq \delta.$$

To do this, we defined X_i the random variables X_i

$$X_i = \begin{cases} 1 & \text{if } |Y_i - F_2| \geq \varepsilon F_2 \\ 0 & \text{otherwise.} \end{cases}$$

Now we have:

(1) $\Pr[X_i] \leq \frac{1}{8}$

(2) X_i are iid

(3) $|Y - F_2| \geq \varepsilon F_2 \Rightarrow \frac{1}{k}\sum_{i=1}^{k} X_i \geq \frac{1}{2}$.

Chernoff bounds

Let X_i be $\text{Ber}(p)$ r.v.'s, then

$$\Pr\left[\left|\sum_{i=1}^{k} X_i - pk\right| \geq \varepsilon n\right] \leq 2\exp(-\Theta(\varepsilon^2 k))$$

In particular.

$$\Pr\left(\sum_{i=1}^{k} X_i - pk\right) \geq \varepsilon n) \leq \exp(-\Theta(\varepsilon^2 k))$$

[Assume: $p + \varepsilon \leq 1$]

Pf: We try to use all moments at once!

$$\Pr\left[\sum_{i=1}^{k} X_i \geq (p+\varepsilon)k\right]$$

$$= \Pr\left[\exp\left(\lambda \sum_{i=k}^{k} X_i\right) \geq \exp\left(\lambda k (p+\varepsilon)\right)\right]$$

$$\forall\ \lambda > 0.$$

$$\leq \exp(-\lambda k t)\, E\left[\exp\left(\lambda \sum_{i=1}^{k} X_i\right)\right]$$

$$= \exp(-\lambda k t)\, E[\exp(\lambda X_1)]^k \quad \text{by independence and identical distribution.}$$

$$= \exp\left(-k\left(\lambda t - \log E[\exp(\lambda X_1)]\right)\right)$$

$$\leq \exp(-k\, \varphi_X(t))$$

where

$$\varphi_X(t) = \sup_{\lambda \geq 0} \left(\lambda t - \log E[\exp(\lambda X_1)]\right).$$

Need to compute this.

$E[\exp(\lambda X)]$, seen as a function of λ, is similar to the 'Laplace transform'.

Now, for the Bernoulli case.

$$E[\exp(\lambda X)] \leftarrow \cancel{E[1-X+Xe^{\lambda}]}$$

$$= 1-p+pe^{\lambda}$$

So, $\psi_X(t) = \sup_{\lambda \geq 0} (\lambda t - \log(1-p+pe^{\lambda}))$

$$f(\lambda) = \lambda t - \log(1-p+pe^{\lambda})$$

$$f'(\lambda) = t - \frac{pe^{\lambda}}{1-p+pe^{\lambda}}$$

$\Rightarrow f''(\lambda) < 0$ everwhere, so zero of $f'(\lambda)$ is a maximum. This is given by

$$t(1-p+pe^{\lambda}) = pe^{\lambda}$$

$$\text{or, } \frac{t(1-p)}{p(1-t)} = e^{\lambda}$$

$$\Rightarrow \lambda = \log\left(\frac{t(1-p)}{p(1-t)}\right).$$

So, we get

$$\varphi_X(t) = t \log\left(\frac{t}{p}\frac{1-p}{1-t}\right) - \log\left(1-p+\frac{t(1-p)}{1-t}\right)$$

$$= t\log\left(\frac{t(1-p)}{p(1-t)}\right) - \log\left(\frac{1-p}{1-t}\right).$$

$$= t\log\left(\frac{t}{p}\right) + (1-t)\log\left(\frac{1-t}{1-p}\right)$$

$$= D(t\|p).$$

So $\Pr\left(\sum X_i - pk > \varepsilon k\right) \leq \exp(-k D(t\|p))$

where $t = p+\varepsilon$.

Now $D(p+\varepsilon \| p)$

$$= (p+\varepsilon)\log\left(1+\frac{\varepsilon}{p}\right) + (1-p-\varepsilon)\log\left(\frac{1-p-\varepsilon}{1-p}\right)$$

$$= (p+\varepsilon)\log\left(1+\frac{\varepsilon}{p}\right) + (1-p-\varepsilon)\log\left(1-\frac{\varepsilon}{1-p}\right).$$

$$\left(\leq (p+\varepsilon)\left(\frac{\varepsilon}{p}\right) \quad (1-p-\varepsilon)\left(-\frac{\varepsilon}{1-p}\right)\right.$$

$$\left. 1 + \frac{\varepsilon}{p} - \varepsilon \quad \frac{\varepsilon^2}{p} + \frac{\varepsilon^2}{1-p} \ldots\right)$$

$$f(\varepsilon) = (p+\varepsilon)\log(p+\varepsilon) + (1-p-\varepsilon)\log(1-p-\varepsilon) + (p+\varepsilon)\log p - (1-p-\varepsilon)\log(1-p)$$

$f(0) = 0$

$$f'(\varepsilon) = \log\left(\frac{p+\varepsilon}{1-p-\varepsilon}\right) - \log\left(\frac{p}{1-p}\right)$$

$f'(0) = 0.$

$$f''(\varepsilon) = \frac{1}{p+\varepsilon} + \frac{1}{1-p-\varepsilon} \geq 4.$$

$$\Rightarrow D(p+\varepsilon \| p) \geq 2\varepsilon^2.$$

$\Rightarrow$ So we get the Chernoff bdd., and hence the correct tale tail: We have to choose k so that

~~$\exp(-2k\varepsilon^2)$~~

Need $\exp(-2k(\frac{\varepsilon}{\delta})^2) \leq \delta$

So, $k \geq 5\log(1/\delta)$ suffices.

Reducing storage :- Note that we used that the rows are mutually independent, but within each row, we only needed four-wise independence.

Thm :- There is a polynomial time algorithm which takes as input $O(k \log n)$ uniform random bits (denoted as R) and an index $i \in [n]$, and outputs in polynomial time ($\text{poly}(k \log n)$) a bit $A(R, i)$ s.t. $\forall\ i_1, i_2, i_3, \ldots i_k$ distinct and in $[n]$

$$A(R, i_1), A(R, i_2) \ldots A(R, i_n)$$

are k wise independent.

$$\Pr\left[\bigwedge_{l=1}^{k} A(R, i_l) = b_l\right] = \prod_{l=1}^{k} \Pr\left[A(R, i_l) = b_l\right]$$

$$\forall\ (b_1, \ldots b_k) \in \{0,1\}^l.$$

Thus we can generate each row with only $O(\log m)$ [rather than $\Theta(m)$] stored random bits.

Consider first the case $k=2$. Let $r_1, r_2 \ldots r_t$ be uniformly random bits. Then

Claim :- The $\left\{ \bigoplus_{i \in S} r_i \mid S \subseteq [t], S \neq \{\phi\} \right\}$ is a set of ~~pair~~ $2^t - 1$ pairwise independent random bits.

Pf :- Pr (Let $r_S = \bigoplus_{i \in S} r_i$

Then, clearly, when $S \neq \phi \{\}$,

$\Pr[r_S = 0] = \frac{1}{2}$.

Now Suppose $S \neq T$ are distinct non-empty sets. ~~Let~~ Wlog suppose $\exists\ a \in S \setminus T$ (if not ~~repl~~ ~~make I into S~~. ~~relable~~ relabel S & T as each other.

Now

$$\Pr[r_S = \alpha \wedge r_T = \beta] = \Pr[r_S = \alpha \mid r_T = \beta] \cdot \frac{1}{2}$$

$$= \frac{1}{2} \cdot \frac{1}{2} = \Pr[r_S = \alpha] \Pr[r_T = \beta]$$

k-wise independence:

Consider another construction. Let $F = \mathbb{F}_p$ be a large enough finite field. Let $a_0, a_1, \ldots a_{k-1}$ be a.a.r. chosen random elements from F.

Then the elements

$$\left\{ \alpha_x = \sum_{i=0}^{k-1} a_i x^i \;\middle|\; x \in \mathbb{F} \right\}$$ are k-wise independent. To see this, consider the probability

$$\Pr\left[\bigwedge_{i=1}^{k} \alpha_{x_i} = \beta_{x_i}\right]$$ where $x_1 \ldots x_k$ are distinct and $\beta_{x_1} \ldots \beta_{x_k}$ are arbitrary. We have.

$$\Pr\left[\bigwedge_{i=1}^{k} \alpha_{x_i} = \beta_{x_i}\right] = \frac{1}{p^k},$$ which proves k-wise independence.

For bits one does this by taking $p = 2^{\lceil \lg n \rceil}$, and then using a finite field structure on 2^d $[d = \lceil \lg n \rceil]$.