Lecture 9. Entropy function

Errata of Alon-Linial theorem: asymmetric dependency.

Original proof of algorithmic local lemma.

Moser's fix-it algorithm for k-SAT model.

Solve(φ).

randomly initialize $x_1, \dots, x_n$.

while $\exists$ unsatisfied clause C.

Fix(C)

Fix(C)

randomly resample vbl(C). (including C itself.)

while $\exists$ unsatisfied clause D overlapping with C

Fix(D)

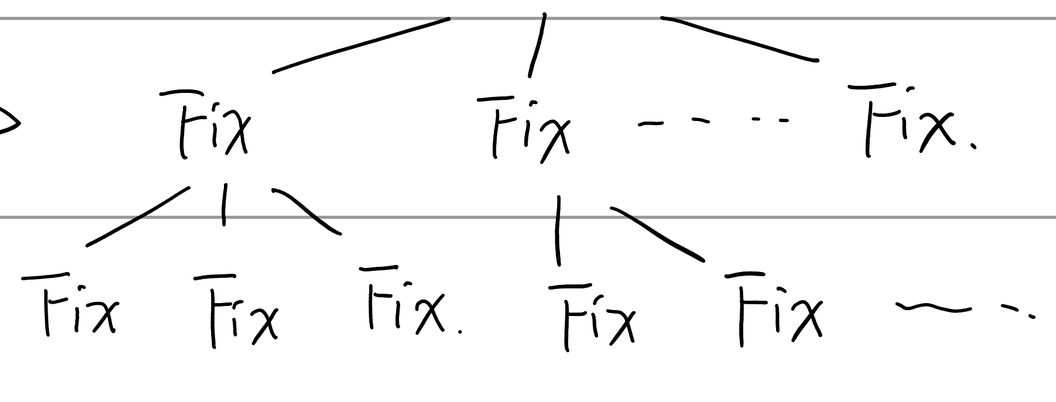
Theorem (Robin Moser, 2009 STOC best paper award).

Let d be the maximum degree of clauses. i.e. each clause intersects with $\le d$ other clauses (including itself). Then Solve(φ) returns a satisfying assignment in poly-time whp. as long as $d \le 2^{k-3}$.

Proof. Consider the recursion tree.

level l has $\leq m$ calls $\rightarrow$

$\forall$ Fix calls $\leq d$ children $\rightarrow$



The key fact is that if Fix(C) calls for some clause C.

the assignment for vbl(C) before resampling is determined.

So if the final assignment and the recursion tree is known.

all random bits used in the algorithm is determined.

Suppose there are T times of Fix calls before terminating.

clearly Solve() used $n + kT$ random bits in total.

We now argue that if T is sufficiently large.

the recursion tree used less bits than kT.

Claim: After any sub-tree rooted at level l terminating.

any satisfied clause cannot become unsatisfied.

So level l clauses are distinct. Denote them by m 0/1 bits.

For any other nodes in recursion tree. we need to record it using its index in its parent node. Denote each of them by $\lceil \log_2 d \rceil$ bits.

To construct the recursion tree. denote the DFS sequence.

For each node, use "1" + "node index" to push it in stack, and

use "0" to record the outting.

Overall. we need $\leq m + (\lceil \log_2 d \rceil + 2)T \leq m + (k-1)T$ bits.

Finally we use n bits to record the final assignment of x.

Again. note that every random bit used in Moser's algorithm.

is determined uniquely by the final assignment and recursion tree.

So we use $m+(k-1)T+n$ bits to record $n+kT$ random bits.

Theorem (Kolmogrov complexity. incompressibility theorem).

N uniform random bits cannot be encoded to $\leq N-l$ bits

with probability $\geq 1-O(2^{-l})$.

Theorem (Shannon's source coding theorem. simplified version).

For a discrete random variable X with finite support $\mathcal{X}$. let

$f: \mathcal{X} \to \{0,1\}^*$ be a uniquely decodable code. Then

$$\mathbb{E}[|f(X)|] \geq H(X) = -\sum_x \Pr[X=x] \log(\Pr[X=x]).$$

In particular. $X \sim \{0,1\}^N$ uniformly. $H(X) = -\log(1/2^N) = N$.

Back to Moser's proof.

$n+kT$ random bits encoded to $n+m+(k-1)T$ bits.

T is random. and may go to infinity. Fix $t = m + \log n$

Only $n+kt$ random bits. force to terminate if used up.

Algorithm succeeds: $n+m+(k-1)T + k(t-T)$ bits.

Algorithm fails: $\leq n+m+(k-1)t$ bit. with $p \leq 2^{m-t}$. $\square$

Definition. (Entropy). Given a discrete random variable $X \in \Omega$.

define the entropy $H(X) = \sum_{\omega \in \Omega} - \Pr[X=\omega] \log_2 \Pr[X=\omega]$.

(by convention. $0 \log 0 = 0$). Usually denote $p_\omega = \Pr[X=\omega]$

Basic properties.

Proposition 1. $H(X) \geq 0$.

Proposition 2 (Uniform bound). Let Ω be the support of X.

$H(X) \leq \log_2 |\Omega|$. with equality iff X is uniformly distributed.

Proof. Let $f(x) = -x \log_2 x$, $x \in [0,1]$. Since f is concave. we have

$$H(X) = \sum_\omega f(\Pr[X=\omega]) \leq |\Omega| \cdot f(1/\Omega) = |\Omega| \cdot \frac{1}{|\Omega|} \cdot \log |\Omega|.$$

by Jensen's inequality. $\square$

Proposition 3. (Concentration). If $H(X) \le t$. then $\exists\, \omega \in \Omega$. s.t.

$$P_\omega = \Pr[X = \omega] \ge 2^{-t} \qquad (H(X) = \text{expectation of } -\log_2 P_\omega)$$

Definition (Conditional entropy). Let E be an event.

$$H(X \mid E) = \sum_x -\Pr[X = x \mid E] \cdot \log_2 \Pr[X = x \mid E].$$

If Y is another random variable.

$$H(X \mid Y) = \sum_y H(X \mid Y = y) \cdot \Pr[Y = y] = \mathbb{E}_y[H(X \mid Y = y)]$$

Lemma (Chain rule). Let $H(X, Y)$ be the entropy of joint r.v.s

(X, Y). i.e. $H(X, Y) = \sum_{x,y} -\Pr[X = x, Y = y] \log_2 \Pr[X = x, Y = y]$

Then $H(X, Y) = H(X) + H(Y \mid X)$.

In particular, $H(X, Y) = H(X) + H(Y)$ if independent.

Proof. Let $p(x, y)$ denote $\Pr[X = x \mid Y = y]$. etc. Then

by Bayes' rule. $p(x, y) = p(y)\, p(x \mid y)$. So.

$$H(X \mid Y) = \mathbb{E}_y[H(X \mid Y = y)] = \sum_y -p(y) \sum_x p(x \mid y) \log_2 p(x \mid y)$$

$$= \sum_{x,y} -p(x, y) \cdot \log_2 \frac{p(x, y)}{p(y)}$$

$$= \sum_{x,y} -p(x, y) \log_2 p(x, y) + \sum_{x,y} p(x, y) \log_2 p(y)$$

$$= H(X, Y) + \sum_y p(y) \log_2 p(y) = H(X, Y) - H(Y). \qquad \square$$

Intuitively. $H(X|Y)$ measures the amount of additional information in X, given a particular value of Y. averaged over all values of Y.

Some important special cases:

① $H(X|Y) = 0$ if $Y = f(X)$ for some function f.

② $H(X|Y) = H(X)$ iff X. Y are independent.

③ $H(X|Y, Z) \leq H(X|Y)$. (dropping condition).

Subadditivity.

Theorem. $H(X, Y) \leq H(X) + H(Y)$.

Proof. $H(X) + H(Y) - H(X, Y) = -\sum_{x,y} p(x, y) \log_2 \frac{p(x)p(y)}{p(x, y)}$

by convexity of $-\log_2 x$ $\geq \log_2 \left(\sum_{x,y} p(x, y) \cdot \frac{p(x)p(y)}{p(x, y)} \right)$

$= \log_2 \left(\sum_{x,y} p(x)p(y) \right) = 0$. □

More generally. $H(X_1, \ldots X_n) \leq H(X_1 \ldots X_{n-1}) + H(X_n)$

$\leq \cdots \leq H(X_1) + \cdots + H(X_n)$.

Proof of chopping condition: $H(X|Y) = H(X, Y) - H(Y) \leq H(X)$.

$\forall z \in \Omega_Z$. $H(X|Y, Z = z) \leq H(X|Z = z)$

Taking expectation of z. it yields $H(X|Y, Z) \leq H(X|Z)$. □

Theorem (Shearer's lemma. special case).

$$2H(X,Y,Z) \leq H(X,Y) + H(X,Z) + H(Y,Z).$$

Proof. $H(X,Y) = H(X) + H(Y|X)$.

$$H(X,Z) = H(X) + H(Z|X)$$

$$H(Y,Z) \geq H(Y|X) + H(Z|X).$$

$\Rightarrow$ lhs $\geq 2H(X) + 2H(Y|X) + 2H(Z|X,Y) = 2H(X,Y,Z)$. □

Theorem (Chung. Frankl. Graham and Shearer, 1986).

Let $S_1 \ldots S_m$ be subsets of $[n]$ such that every $i \in [n]$ belongs to at least k of $S_1 \ldots S_m$. Then $kH(X_1 \ldots X_n) \leq \sum_{i=1}^{m} H(X_{S_i})$.

Corollary. Let $\mathcal{F}$ be a family of subsets of $[n]$. and p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then $\log_2 |\mathcal{F}| \leq \sum_{i=1}^{n} H_b(p_i)$. where $H_b(p) = -p \log_2 p - (1-p)\log_2(1-p)$.

$H_b(p) = H(\text{Bernoulli}(p))$.

Proof. For each $S \in \mathcal{F}$ let v_S be its indicator vector of length n.

Let $X = (X_1 \ldots X_n)$ be a random variable taking values in $\{0,1\}^n$.

and set $\Pr[X = v_S] = 1/|\mathcal{F}|$ for all $S \in \mathcal{F}$.

Clearly. $H(X) = |\mathcal{F}| \cdot (-1/|\mathcal{F}| \cdot \log_2 1/|\mathcal{F}|) = \log_2 |\mathcal{F}|$. and

$H(X_i) = H_b(p_i)$. So the result follows from subadditivity. □

Corollary. $\sum_{i \le np} \binom{n}{i} \le 2^{nH_b(p)}$ if $0 < p \le 1/2$.

Proof (by P. Frankl). Let $\tilde{\mathcal{F}} = \binom{[n]}{\le pn}$ be the family of all subsets of $[n]$ of size $\le pn$. and p_i be the fraction of containing i.

clearly $p_1 = p_2 = \cdots = p_n$. and $\sum_i p_i |\tilde{\mathcal{F}}| = \sum_{S \in \tilde{\mathcal{F}}} |S| \le pn|\tilde{\mathcal{F}}|$.

So $p_i \le p$. and thus $H_b(p_i) \le H_b(p)$. Apply above results. □.

Corollary. (Product Theorem). Let $S_1 \cdots S_m \subseteq \Omega$. where each $i \in \Omega$ appears in at least k subsets S_j. Then for every $\tilde{\mathcal{F}} \subseteq 2^{\Omega}$.

$$|\tilde{\mathcal{F}}|^k \le \prod_{i=1}^{m} |\tilde{\mathcal{F}}|_{S_i}|. \quad \text{where } \tilde{\mathcal{F}}|_S = \{F \cap S : F \in \tilde{\mathcal{F}}\}.$$

Proof. Let $n = |\Omega|$. Each subset of Ω corresponds to a $\{0,1\}$ vector of length n. Then, applying Shearer's lemma we have.

$$k \log_2 |\tilde{\mathcal{F}}| = k H(X_1 \ldots X_n) \le \sum_{i=1}^{m} H(X_{S_i}) \le \sum_{i=1}^{m} \log_2 |\tilde{\mathcal{F}}|_{S_i}|. \quad \square$$

Theorem. Let $\tilde{\mathcal{F}} \subseteq 2^{[n]}$ be a family of subsets of $[n]$. Suppose that $\forall S, T \in \tilde{\mathcal{F}}$. $S \cap T$ contains a pair of consecutive numbers.

Then $|\tilde{\mathcal{F}}| \le 2^{n-2}$.

Remark. Note that the bound is optimal. Consider $\tilde{\mathcal{F}} = 2^{[n-2]} + \{n-1, n\}$.

Proof. Let N_0 and N_1 be the set of all even numbers and odd numbers in $[n]$ respectively. Consider the projections : $\widetilde{\mathcal{F}}_i = \{ F \cap N_i : F \in \mathcal{F} \}$.

Note that. $\forall S, T \in \widetilde{\mathcal{F}}$. $\exists k$. s.t. $\{k, k+1\} \subseteq S \cap T$. So

$S \cap T \cap N_0 \neq \phi$ $S \cap T \cap N_1 \neq \phi$. Namely. $\forall S_i, T_i \in \widetilde{\mathcal{F}}_i$. $S_i \cap T_i \neq \phi$.

It implies that $|\widetilde{\mathcal{F}}_i| \leq 2^{|N_i|-1}$. Applying product theorem with $k=1$.

$$|\widetilde{\mathcal{F}}| \leq |\widetilde{\mathcal{F}}_0| \cdot |\widetilde{\mathcal{F}}_1| \leq 2^{|N_0|-1} \cdot 2^{|N_1|-1} = 2^{n-2}. \qquad \square$$

Theorem (Chung. Graham, Frankl and Shearer, 1986).

Every triangle intersecting family of subgraphs of K_n has size $< 2^{\binom{n}{2}-2}$.

Proof. Let $\mathcal{G}$ be a triangle intersecting family. $\forall S \subseteq [n]$ with $|S| = \lfloor \frac{n}{2} \rfloor$

let $E_S = \binom{S}{2} \cup \binom{[n] \setminus S}{2}$ (clique on S + clique on complement) and

$$r = |E_S| = \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} \leq \frac{1}{2}\binom{n}{2}.$$

$\forall S$. each triangle has ≥ 1 edge in E_S. Thus $\mathcal{G}$ restricted to

E_S is an intersecting family. So $|\mathcal{G}|_{E_S}| \leq 2^{|E_S|-1} = 2^{r-1}$

Each edge of K_n appears in $\geq k = \frac{r}{\binom{n}{2}}\binom{n}{\lfloor n/2 \rfloor}$ different E_S's.

Applying product theorem. it gives that $|\mathcal{G}|^k \leq (2^{r-1})^{\binom{n}{\lfloor n/2 \rfloor}}$.

Therefore $|\mathcal{G}| \leq 2^{\binom{n}{2} \cdot \frac{r-1}{r}} < 2^{\binom{n}{2}-2}$. $\qquad \square$