Lecture 13: Information Theory and Counting: Shearer's Lemma

Entropy

- Entropy = "Surprise"
- X is a random variable
- $p_X(x)$ is the probability of x according to the distribution X
- Surprise of seeing x is, $h_X(x) := -logp_X(x)$
- Entropy of X is the expected surprise, $H(X) := \mathbb{E}_{x \sim X}[h_X(x)]$
- Alternately, $H(X) = \sum_x -p_X(x) \log p_X(x)$

Conditional Entropy

- (X, Y) is a joint distribution
- $(X|Y = y)$ is the conditional distribution of X conditioned on the fact that $Y = y$
- Entropy of $(X|Y = y)$ is defined by $H(X|Y = y)$
- Conditional entropy $H(X|Y) := \mathbb{E}_{y \sim Y}[H(X|Y = y)]$

Entropy

- Chain Rule: $H(XY) = H(X) + H(X|Y)$
- Conditional Chain Rule: $H(XY|Z) = H(X|Z) + H(Y|XZ)$
- Inequalities:
  - $0 \leqslant H(X) \leqslant |\text{range}(X)|$
  - $H(X) \geqslant H(X|f(Y)) \geqslant H(X|Y)$
- Binary Entropy Function:
  $h(p) := -p \log p - (1-p) \log(1-p)$

Binomial Coefficient Tail

Theorem

$$\sum_{i \leqslant \alpha n} \binom{n}{i} \leqslant 2^{h(\alpha)n}$$

- Let $\mathcal{C}$ be the set of all subsets of $[n]$ of size at most αn
- Let X be a uniform distribution over $\mathcal{C}$
- Let $(X_1, \dots, X_n)$ be the characteristic vector corresponding to the subset sampled by X
- $\log |\mathcal{C}| = H(X) = H(X_1, \dots, X_n) = \sum_{i \in [n]} H(X_i | X_{<i}) \leqslant \sum_{i \in [n]} H(X_i)$
- Since all indices are symmetric, $H(X_i) = H(X_1)$
- Note that $H(X_1 | |X| = i) = h(i/n) \leqslant h(\alpha)$, for $i \leqslant \alpha n$
- Therefore, $H(X_1) \leqslant h(\alpha)$
- Overall $\log |\mathcal{C}| \leqslant nh(\alpha)$

Identifying Bad Balls

Consider the task of designing a set $\mathcal{D} = \{D_1, \ldots, D_\ell\}$ such that $D_i \subseteq [n]$, for $i \in [\ell]$ such that:

- Consider n ordered balls
- Let B be the set of positions with bad balls
- Suppose we are given an oracle that on input $D \subseteq [n]$ outputs $|B \cap D|$
- Using each set in $\mathcal{D}$ to query the oracle, output B

Theorem

$$\ell \geqslant n / \log(n+1)$$

- Note that $B \mapsto (|B \cap D_1|, \ldots, |B \cap D_\ell|)$ is a bijection (two different B and B' cannot have the same sequence, otherwise we cannot distinguish B from B')
- Let X be a uniform random variable over $2^{[n]}$ (i.e., the set of all subsets of $[n]$)
- $n = H(X) \leqslant \sum_{i \in [\ell]} H(|X \cap D_i|) \leqslant \ell \log(n+1)$

Number of Matchings

Let $G = (A, B, E)$ be a bipartite graph

Theorem (Brégman's Theorem)

Number of perfect matchings in G is at most $\prod_{v \in A} (d(v)!)^{1/d(v)}$

- Let Σ be the set of all perfect matchings
- Let σ be a uniform random variable over Σ
- $\log |\Sigma| = H(\sigma) = \sum_{v \in A} H(\sigma(v) | \sigma(u)_{u<v})$
- Trivial upper bound by $H(\sigma(v)) \leqslant \log d(v)$
- Idea: Expand according to a random permutation τ of vertices in A
- Think: How to get $\frac{1}{d(v)} (1 + \cdots + d(v))$ as upper bound to get the result

Introduction to Shearer's Lemma

- Consider a set S of n points in 3-dimensions
- Let n_1 be number of unique points by projecting S on $X = 0$ plane, n_2 be the number of unique points by projecting S on $Y = 0$ plane and n_3 be the number of unique points by projecting S on $Z = 0$ plane

Theorem

$$n \leqslant (n_1 n_2 n_3)^{1/2}$$

- Let (X, Y, Z) represent the coordinates of uniformly chosen point in S
- $\log n = H(X, Y, Z) = H(X) + H(Y|X) + H(Z|XY)$
- $\log n_1 \geqslant H(Y, Z) = H(Y) + H(Z|Y) \geqslant H(Y|X) + H(Z|XY)$
- $\log n_2 \geqslant H(X, Z) = H(X) + H(Z|X) \geqslant H(X) + H(Z|XY)$
- $\log n_3 \geqslant H(X, Y) = H(X) + H(X|Y)$
- $\log n \leqslant \frac{1}{2} \log(n_1 n_2 n_3)$

- Let $\mathcal{F}$ be a set of subsets of $[n]$
- For every $i \in [n]$, there are at least t subsets in $\mathcal{F}$ that contain i

Theorem (Shearer's Lemma)

$$H(X_1, \ldots, X_n) \leqslant \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_F)$$

- "Sub-additivity of Entropy" is obtained by considering $\mathcal{F}$ as the set of all singleton sets
- "Volume computed by projections" is obtained by considering $\mathcal{F}$ as the set of all subsets of size $(n-1)$

Theorem (Loomis-Whitney Theorem)

Let B be a measurable body in $\mathbb{R}^d$ and $|\cdot|$ represent the volume. Let B_j, for $j \in [d]$, represent the body when B is projected along the j-th coordinate axis. Then:

Proof of Shearer's Lemma

- The i-th smallest index in F is represented by F_i
- Consider the manipulation similar to the "volume argument"

$$\sum_{F \in \mathcal{F}} H(X_F) = \sum_{F \in \mathcal{F}} \sum_{i \leqslant |F|} H(X_{F_i} | X_{\{j < F_i\} \cap F})$$

$$\geqslant \sum_{F \in \mathcal{F}} \sum_{i \leqslant |F|} H(X_{F_i} | X_{\{j < F_i\}})$$

$$\geqslant t \sum_{i \in [n]} H(X_i | X_{<i}) = tH(X_1, \dots, X_n)$$