TUTORIAL VIII

1 Average lower bound for sorting through entropy

We consider the problem of sorting a list of n distinct elements using K binary questions on the input. Let $a_1 < a_2 < \cdots < a_n$ be the sorted elements of our list. Our input is a permutation σ of the sorted list $[a_{\sigma(1)}, \ldots, a_{\sigma(n)}]$. The objective is to find σ (applying σ^{-1} sorts the list).

Let $X \sim \mu$ a random permutation of $\{1, \ldots, n\}$ (not necessarily uniform), $Y = (Y_i)_{1 \leq i \leq K} \in \{0,1\}^K$ the outputs of the questions asked by the algorithm $\mathcal{A}$ solving this problem (comparison results for instance), and $\hat{X} = \mathcal{A}(Y)$ the output of the algorithm $\mathcal{A}$. Formally, it means that $X \to Y \to \hat{X}$ is a Markov chain, since we suppose that $\mathcal{A}$ uses only the outputs Y to determine its guess $\hat{X}$ of X.

The goal is to find a lower bound on the average number of comparisons $\mathbb{E}[K]$ for correct algorithms $\mathcal{A}$ if we suppose that the input is distributed following μ.

1. Find a necessary condition for the correction of $\mathcal{A}$ in terms of entropies.
2. Show that $H(X) \leq \mathbb{E}[K] + H(K)$ if $\mathcal{A}$ is correct.
3. Prove that for any r.v. $Z \in \{0, \ldots, N\}$, $H(Z) \leq \mathbb{E}[Z] + 2$.
4. Conclude. What can we say if μ is the uniform distribution on permutations of $\{1, \ldots, n\}$? What bound would we get if we had allowed d-ary questions?

Remark. The correction assumption of the algorithm implies implicitly that it worked in finite time - the size of the input is fixed, there is $n!$ such instances, if A is deterministic then it works in finite time; if A is probabilistic there may be some ambiguity on what is correction, but what we mean is that on any particular choice of probability outputs, it works in finite time, so that would implies also what we want. The case where we allow infinite number of queries might still work (go check it!)

2 Number of independent sets in a graph

For counting objects, the properties of the Shannon entropy are sometimes useful. We will use the following lemma seen in last lecture:

Lemma 2.1 (Shearer, 1986)**.** *If $S_1, \ldots, S_m \subseteq [n]$ and for every $i \in [n]$, i appears at least k times, then*

$$k \cdot H(X_1 \ldots X_n) \leq \sum_{i=1}^{m} H(X_{S_i})$$

The objective of this exercise is to bound the number of independent sets of d-regular bipartite graphs, ie. bipartite graphs such that all their vertices have same degree d. More precisely, we want to show the following theorem using Shearer's lemma:

Theorem 2.2 (Kahn, 2001)**.** *If G is a d-regular bipartite graph with n vertices, then the number of independent sets of G is at most $(2^{d+1} - 1)^{\frac{n}{2d}}$.*

Remark. For simplicity of the proof, we restrict our attention to bipartite graphs, but the statement is true in general[1].

[1] Zhao, Y., 2010. The number of independent sets in a regular graph. Combinatorics, Probability and Computing, 19(2), pp.315-320, https://arxiv.org/pdf/0909.3354.pdf

2.1 Preliminaries

1. Using the chain rule[2] on $H(X_{S_j})$, prove Lemma 2.1.
2. How many independent sets can there be in a 1-regular generic graph?
3. What about a 2-regular generic graph?
4. Give a family of graphs where the bound of Theorem 2.2 is achieved.

2.2 Proof of Theorem 2.2

We assume the vertices of G are labelled by $[n] := \{1, \dots, n\}$ and let A and B two subsets of $[n]$ with edges only between A and B. We assume moreover that $|A| \geq |B|$. We write $\mathcal{I}(G)$ as the set of independent sets of G. Let I be a uniformly random independent set in $\mathcal{I}(G)$. We can specify I by specifying whether for each vertex v, $v \in I$ or not. For $v \in [n]$, let $X_v = \mathbf{1}_{v \in I}$. We decompose the entropy in two parts:

$$H(X_1 \dots X_n) = H(X_A) + H(X_B|X_A) \, .$$

1. What is the relation between $H(X_1 \dots X_n)$ and $|\mathcal{I}(G)|$?
2. Bound $H(X_B|X_A)$ in terms of $H(X_b|X_{N(b)})$, where $N(b) = \{a \in A : (a, b) \in E(G)\}$ is the neighbourhood of b.
3. Show that $H(X_b|X_{N(b)}) \leq P_{Q_b}(1) =: q_b$ where $Q_b = 1$ if $I \cap N(b) = \emptyset$ and $Q_b = 0$ otherwise.
4. Using Shearer's lemma, bound $H(X_A)$ in terms of $H(X_{N(b)})$
5. Why do we have $H(X_{N(b)}) = H(X_{N(b)}Q_b)$?
6. Conditioning on Q_b, prove that $H(X_{N(b)}) \leq h_2(q_b) + (1 - q_b)\log(2^d - 1)$
7. Deduce that $H(X_1 \dots X_n) \leq \frac{|B|}{d}\log(2^d - 1) + \frac{1}{d}\sum_{b \in B}\left(h_2(q_b) + q_b \log\frac{2^d}{2^d - 1}\right) =: f((q_b)_{b \in B})$
8. What is the maximum of f on $[0, 1]^B$?
9. Conclude.

[2] $H(X_1 \dots X_n) = \sum_{i=1}^n H(X_i|X_1 \dots X_{i-1}) = H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_1 \dots X_{n-1})$