Information and Coding Theory **Autumn 2014**

Lecture 3: Oct 7, 2014

Lecturer: Shi Li Scribe: Mrinalkanti Ghosh

In last lecture we have seen an use of entropy to give a tight upper bound in number of triangles in an extremal graph in certain family. In this lecture we explore few more combinatorial applications of entropy.

Before we get into the examples, let us recapitulate a few facts about entropy:

- $H(Y|X) \leq H(Y)$
- $H(X_1 X_1 \dots X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_1 \dots X_{i-1}) \leq \sum_{i=1}^{n} H(X_i)$
- $H(X) \leq \log |U|$ where U is the range of discrete random variable X with equality when X is distributed uniformly.

The problems we discuss in this lecture require us to count various structures in certain extremal objects. In the first look, these problems seem unrelated to information theory or entropy. Nevertheless, we can use properties of entropy profitably to approach these problems.

Some of the problems discussed in this lecture also appears in [Jai01]. Interested readers are rquested to refer to [Jai01] for details and for many other intersecting examples which are not covered here.

1 Counting repetition in a square matrix

Claim 1.1 *Let A be an $n \times n$ matrix such that each row and column has at most t distinct numbers. There is a number x such that x appears in A at least n^2/t^2 times.*

Proof: Let X, Y, Z be a triple of random variables where $X \sim_U [n]$, $Y \sim_U [n]$ and $Z = A_{XY}$ (X and Y are independent), i.e., we independently and uniformly randomly choose a row X, a column Y and set Z to the (X, Y)-th element of the matrix.

Since each row has at most t distinct numbers, for all $i \in [n]$, $H(Z|X = i) \leq \log t$ and hence $H(Z|X) = \mathbb{E}_i [H(Z|X = i)] \leq \log t$. Similarly, $H(Z|Y) \leq \log t$.

So we want to establish relationship among $H(Z|X), H(Z|Y)$ and $H(Z)$.

To that end, we observe that Z is completely determined by X, Y and hence $H(Z|XY) = 0$. Also, X, Y are independent: $H(XY) = H(X) + H(Y)$. Combining these two, we know that

$$H(XYZ) \;=\; H(XY) + H(Z|XY) \;=\; H(X) + H(Y)\,.$$

Applying the chain rule in a different way and using the fact that conditioning reduces entropy, we get that

$$\begin{aligned} H(Z) &= H(XY) - H(XY|Z) \\ &= H(X) + H(Y) - H(X|Z) - H(Y|XZ) \\ &\geq H(X) + H(Y) - H(X|Z) - H(Y|Z) \end{aligned}$$

Finally, noticing that $H(X) - H(X|Z) = H(Z) - H(Z|X)$ and $H(Y) - H(Y|Z) = H(Z) - H(Z|Y)$ gives that

$$H(Z) \geq 2H(Z) - H(Z|X) - H(Z|Y) \quad \Rightarrow \quad H(Z) \leq H(Z|X) + H(Z|Y) \leq 2 \log t\,.$$

Recall that $H(Z) = \sum_z p(z) \cdot \log\left(\frac{1}{p(z)}\right)$. Thus, from the above we get that *on average* $\log\left(\frac{1}{p(z)}\right)$ is at most $2\log t$. Thus, there exists an element z^* for which $\log\left(\frac{1}{p(z^*)}\right) \leq 2\log t$. In particluar, let z^* be the element which occurs with the highest probability. Then $\log\left(\frac{1}{p(z^*)}\right) \leq \log\left(\frac{1}{p(z)}\right) \forall z$, which gives

$$\log\left(\frac{1}{p(z^*)}\right) = \sum_z p(z) \cdot \log\left(\frac{1}{p(z^*)}\right) \leq \sum_z p(z) \cdot \log\left(\frac{1}{p(z)}\right) \leq 2\log t\,.$$

Thus, we get that $p(z^*) \geq 1/t^2$ which means that z^* occurs in at least n^2/t^2 times in the matrix. ■

Remark 1.2 *The bound shown above is asymptotically tight: Consider a $n \times n$ matrix and a t such that $t \mid n$. Divide the matrix in t^2 squares with sides n/t. Index the squares by $[1, t^2]$. Put the number i in the i-th square. Clearly, the requirements for the matrix is satisfied and each number appears n^2/t^2 times.*

2 Shearer's Lemma

We first start with an example similar to the one we saw in the last lecture. As we will see later, the bound here is easily implied by Shearer's lemma and the proof of Shearer's lemma is essentially a generalization of the proof here.

2.1 Areas under projection vs. Volume

Let $S = \{(x_i, y_i, z_i) : i \in [n]\}$ be a set of n (distinct) points in $\mathbb{R}^3$ and let Q_{xy}, Q_{yz}, Q_{xz} be the projections of S to XY, YZ and XZ plains respectively. To be explicit: $Q_{xy} = \{(x_i, y_i) : i \in [n]\}$ and so on. We want to show:

Claim 2.1 $\max\{|Q_{xy}|, |Q_{yz}|, |Q_{xz}|\} \geq n^{2/3}$

Proof: Let X, Y, Z be random variables defined as $(X, Y, Z) \sim_U S$.
Clearly, $H(XYZ) = \log n$. Now $(X, Y) \in Q_{xy}$ and so $H(XY) \leq \log|Q_{xy}|$. Similarly, $H(YZ) \leq \log|Q_{yz}|$ and $H(XZ) \leq \log|Q_{xz}|$.
Now we use an inequality proved in last class:

Fact 2.2 $H(XY) + H(YZ) + H(XZ) \geq 2H(XYZ)$

We include a proof for the sake of completeness:

Proof:

$$\begin{aligned} H(XY) + H(YZ) + H(XZ) &= H(Y) + H(X|Y) + H(YZ) + H(XZ) \\ &\geq H(Y|XZ) + H(X|YZ) + H(YZ) + H(XZ) \\ &= 2H(XYZ) \end{aligned}$$

■

So, $\log|Q_{xy}| + \log|Q_{yz}| + \log|Q_{xz}| \geq H(XY) + H(YZ) + H(XZ) \geq 2H(XYZ) = 2\log n$. Hence, $\max\{\log|Q_{xy}|, \log|Q_{yz}|, \log|Q_{xz}|\} \geq 2\log n/3$, or equivalently, $\max\{|Q_{xy}|, |Q_{yz}|, |Q_{xz}|\} \geq n^{2/3}$

■

Remark 2.3 *The bound shown above is asymptotically tight: Consider an axis-parallel solid cube with side lengths $m-1$ with one corner placed at the origin. Let S be the integral points of this cube. Clearly, $n = m^3$ and $|Q_{xy}| = |Q_{yz}| = |Q_{xz}| = m^2 = n^{2/3}$.*

2.2 Shearer's Lemma

Consider the fact 2.2 and that the entropy is sub-additive. The fact that entropy is subadditive gives that for three variables X, Y and Z

$$H(X) + H(Y) + H(Z) \geq H(XYZ).$$

In the above example, we needed to bound $H(XY)+H(YZ)+H(ZX)$, where each term is not the entropy of a single random variable, but the entropy of a *subset* of random variables. Since these subsets cover each variable X, Y and Z *twice*, we got

$$H(XY) + H(YZ) + H(ZX) \geq 2 \cdot H(XYZ).$$

To state Shearer's lemma, it is more convenient to think of the above in terms of *distributions* on the subsets of the random variables. Formally, the statement is as follows

Lemma 2.4 (Shearer's Lemma) *Let $\mathbf{X} = \{X_i : i \in [n]\}$ be a set of random variables. For any $S \subset [n]$, let us denote $X_S = \{X_i : i \in S\}$. Let D be an arbitrary distribution on $2^{[n]}$ (set of all subsets of $[n]$) and let μ be such that $\forall i \in [n]$ $\mathbb{P}_{S\sim D}[i \in S] \geq \mu$. Then*

$$\mathop{\mathbb{E}}_{S\sim D}[H(X_S)] \geq \mu \cdot H(\mathbf{X}).$$

Before we see the proof, let us see how the lemma implies the previous examples. Let us think of $X_1 = X$, $X_2 = Y$ and $X_3 = Z$. For subadditivity, we can consider D to be a uniform distribution on the sets $\{1\}, \{2\}, \{3\}$. Then, a random set from D contains a given element $i \in \{1, 2, 3\}$ with probabiltiy 1/3 and the lemma gives

$$\frac{1}{3} \cdot H(X) + \frac{1}{3} \cdot H(Y) + \frac{1}{3} \cdot H(Z) \geq \frac{1}{3} \cdot H(XYZ).$$

Similarly, taking D to be uniform over $\{1,2\},\{2,3\}$ and $\{3,1\}$ gives $\mu = 2/3$ and hence

$$\frac{1}{3} \cdot H(XY) + \frac{1}{3} \cdot H(YZ) + \frac{1}{3} \cdot H(ZX) \;\geq\; \frac{2}{3} \cdot H(XYZ)\,.$$

We now proceed to the proof of the lemma.

Proof:

$$\begin{aligned}
\mathop{\mathbb{E}}_{S\sim D}[H(X_S)] &= \mathop{\mathbb{E}}_{S\sim D}\left[\sum_{i\in S} H(X_i|X_{S\cap[i-1]})\right] && \text{by Chain rule}\\
&\geq \mathop{\mathbb{E}}_{S\sim D}\left[\sum_{i\in S} H(X_i|X_{[i-1]})\right] && H(X_i|X_A) \geq H(X_i|X_B) \text{ for } A \subset B\\
&= \mathop{\mathbb{E}}_{S\sim D}\left[\sum_{i\in[n]} \mathbb{1}_S(i)\, H(X_i|X_{[i-1]})\right] && \mathbb{1}_s \text{ is indicator function for set} S\\
&= \sum_{i\in[n]} \mathop{\mathbb{P}}_{S\sim D}[i\in S]\, H(X_i|X_{[i-1]})\\
&\geq \mu \sum_{i\in[n]} H(X_i|X_{[i-1]}) = \mu \cdot H(\mathbf{X})
\end{aligned}$$

■

3 Number of graphs in a triangle-intersecting family

In this section we consider families of graphs on vertex set $[n]$ for a fixed n.

Definition 3.1 *A family $\mathcal{G}$ of graphs is called intersecting if $\forall T, K \in \mathcal{G},\ T \cap K$ contains an edge.*

Fact 3.2 *For an intersecting family $\mathcal{G}$ with edges contained in a set $E_0 \subseteq \binom{[n]}{2}$, $|\mathcal{G}| \leq 2^{|E_0|}/2$.*

Proof: For any graph $G \in \mathcal{G}$, let G^c be a graph such that $E(G^c) = E_0 \setminus E(G)$. Now $G^c \notin \mathcal{G}$, because $G \cap G^c$ does not have any edge. Since there are $2^{|E_0|}$ possible graphs on $[n]$ with edges contained in E_0, we have $|\mathcal{G}| \leq 2^{|E_0|}/2$. ■

Note that when $\mathcal{E} = \binom{[n]}{2}$, we have any intersecting family $\mathcal{G}$ has at most $2^{\binom{n}{2}}/2$ many graphs.

Definition 3.3 *A family $\mathcal{G}$ of graphs is called $\triangle$-intersecting if $\forall T, K \in \mathcal{G},\ T \cap K$ contains a $\triangle$.*

Now we prove the main lemma of this section:

Lemma 3.4 *If $\mathcal{G}$ is $\triangle$-intersecting then $|\mathcal{G}| \leq 2^{\binom{n}{2}}/4$.*

Proof: For any $R \subset [n]$ let us define G_R to be the disjoint union of a clique on R and a clique on $[n] \setminus R$. Formally, $E(G_R) = \{\{i,j\} \;:\; i,j \in R \text{ or } i,j \in [n] \setminus R\}$.

Note that $G_R \cap \triangle$ must have one edge: since at least 2 vertices of a triangle is either in R or in $[n] \setminus R$ together and the edge connecting them is in G_R. Since $\mathcal{G}$ is a $\triangle$-intersecting, $\forall T, K \in \mathcal{G}$, we have $G_R \cap T \cap K$ has at least one edge. Hence $\mathcal{G}_R := \{G_R \cap T \;:\; T \in \mathcal{G}\}$ is an intersecting family with edges contained in $E(G_R)$.

Consider a collection of random variables $\mathbf{X} = \{X_e \;:\; e \in \binom{[n]}{2}\}$ which is sampled as follows: select a graph G uniformly and randomly form $\mathcal{G}$ and for each edge e set $X_e = \mathbb{1}_{E(G)}(e)$ i.e., $X_e = 1$ if $e \in E(G)$ and 0 otherwise. Since there is a bijection mapping between a graph G and the corresponding values of $\mathbf{X}$, we have $H(\mathbf{X}) = \log |\mathcal{G}|$. We will now apply Shearer's lemma by considering a distribution D over subsets of $\binom{[n]}{2}$.

We define the distribution D as follows: Pick a random $R \subset [n]$ of size $n/2$ (say n is even) and take the subset $S = E(G_R)$. As before, we define $X_S = \{X_e \mid e \in S\}$. Also, by symmetry, for each $e \in \binom{[n]}{2}$, $\mathbb{P}_S[e \in S] = \frac{|E(G_R)|}{\binom{n}{2}} = \frac{2\binom{n/2}{2}}{\binom{n}{2}} = \mu$ (say). Applying Shearer's lemma gives

$$\mathop{\mathbb{E}}_{S \sim D}[H(X_S)] \;\geq\; \mu \cdot H(\mathbf{X}) \;=\; \mu \cdot \log |\mathcal{G}| \;.$$

It remains to bound the LHS in the above expression. To do so, we note that for any $R \subset [n]$ and the corresponding S, a set of values for X_S corresponds to a graph in $\mathcal{G}_R$. By the above discussion, we know that $\mathcal{G}_R$ is an intersecting family with edges contained in $E(G_R)$ and by Fact 3.2, $|\mathcal{G}_R| \leq 2^{|E(G_R)|-1}$. This gives that $H(X_S) \leq |E(G_R)| - 1 = 2\binom{n/2}{2} - 1$. Combining this with the bound from Shearer's lemma gives

$$\log |\mathcal{G}| \;\leq\; \frac{1}{\mu} \cdot \left(2\binom{n/2}{2} - 1\right) \;=\; \frac{\binom{n}{2}}{2\binom{n/2}{2}} \cdot \left(2\binom{n/2}{2} - 1\right) \;=\; \binom{n}{2} - \frac{n-1}{n/2-1} \;\leq\; \binom{n}{2} - 2\,,$$

and hence $|\mathcal{G}| \leq 2^{\binom{n}{2}}/4$ ∎

Remark 3.5 *There exists a $\triangle$-intersecting family $\mathcal{G}$ with $|\mathcal{G}| \geq 2^{\binom{n}{2}}/8$: we fix a triangle and take any subset of remaining $\binom{n}{2} - 3$ edges to form a $\triangle$-intersecting family.*

Open Problem 3.6 *There is a gap between the upper and lower bound of the size of a $\triangle$-intersecting family. Are those bounds tight? Can the gap be improved?*

References

[Jai01] Jaikumar Radhakrishnan, "Entropy and counting", *IIT Kharagpur Golden Jubilee Volume*, November 2001,A link to the article.