Information and Coding Theory **Autumn 2022**

Lecture 4: October 6, 2022

Lecturer: Madhur Tulsiani

1 Shearer's Lemma and Combinatorial Applications

Let us begin by restating Shearer's lemma, which we mentioned and used (but did not prove) in the previous lecture.

Lemma 1.1 (Shearer's Lemma). *Let $\{X_1, \ldots, X_n\}$ be a set of random variables. For any $S \subset [n]$, let us denote $X_S = \{X_i : i \in S\}$. Let $\mathcal{F} \subseteq 2^{[n]}$ be a collection of subsets of $[n]$ with the property that for all $i \in [n]$, we have that $|\{S \in \mathcal{F} \mid S \ni i\}| \geq t$. Then*

$$t \cdot H(X_1, \ldots, X_n) \leq \sum_{S \in \mathcal{F}} H(X_S).$$

We will actually prove a more general version of the lemma which can be stated in terms of a distribution over subsets of $[m]$ such that for each $i \in [n]$, we have a lower bound on the probability that a random subset from the distribution includes i. The lemma below can easily be seen to imply the version above, by using the uniform distribution on the collection $\mathcal{F}$.

Lemma 1.2 (Shearer's Lemma: distribution version). *Let $\{X_1, \ldots, X_n\}$ 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*

$$\mu \cdot H(X_1, \ldots, X_n) \leq \mathop{\mathbb{E}}_{S \sim D} [H(X_S)] .$$

Exercise 1.3. *Check that Lemma 1.2 implies Lemma 1.1. Also check that both these lemmas imply sub-additivity.*

We now prove Lemma 1.2

Proof: The proof of the lemma follows simply from the chain rule for entropy and the fact that conditioning reduces entropy (on average).

$$\begin{aligned}
\underset{S\sim D}{\mathbb{E}}[H(X_S)] &= \underset{S\sim D}{\mathbb{E}}\left[\sum_{i\in S} H\left(X_i \mid X_{S\cap[i-1]}\right)\right] && \text{by Chain rule} \\
&\geq \underset{S\sim D}{\mathbb{E}}\left[\sum_{i\in S} H\left(X_i \mid X_{[i-1]}\right)\right] && H(X_i|X_A) \geq H(X_i|X_B) \text{ for } A\subset B \\
&= \underset{S\sim D}{\mathbb{E}}\left[\sum_{i\in[n]} \mathbb{1}_{\{i\in S\}}\cdot H\left(X_i \mid X_{[i-1]}\right)\right] \\
&= \sum_{i\in[n]} \underset{S\sim D}{\mathbb{P}}[i\in S]\cdot H\left(X_i \mid X_{[i-1]}\right) \\
&\geq \mu\cdot\sum_{i\in[n]} H\left(X_i \mid X_{[i-1]}\right) \\
&= \mu\cdot H(X_1,\ldots,X_m)
\end{aligned}$$

■

We now consider some simple combinatorial applications of Shearer's lemma.

1.1 Counting graph homomorphisms

Shearer's lemma can be used to give an estimate of the number of ways of "embedding" a small graph G into a large graph H. For two graphs $G : (V_G, E_G)$ and $H = (V_H, E_H)$, an embedding (also called a homomorphism) of G in H is defined as a function $f : V_G \to V_H$ such that for all $(u, v) \in E_G$, we have $(f(u), f(v)) \in V_H$. Note that the definition does not prevent the image of non-edge pairs in E_G from being edges in E_H.

We will show an upper bound on the maximum number of embeddings for a graph G into any H with at most m edges. For now, let us take G to be the 5-cycle with vertex set $\{1, 2, 3, 4, 5\}$. Consider any graph H with at most m edges and let $F = (F(1), \ldots, F(5))$ be a collection of random variables denoting an embedding of G chosen uniformly from the set of all embeddings. Using Shearer's lemma, we can write

$$2\cdot H(F(1),\ldots,F(5)) \leq H(F(1),F(2)) + H(F(2),F(3)) + \cdots + H(F(5),F(1)).$$

Since $\{1, 2\}$ is an edge in G, the pair $(F(1), F(2))$ must correspond to an (ordered) edge in H. Since the number of edges in H is at most m, we get that $H(F(1), F(2)) \leq \log(2m)$. Using the same bound for all terms on the right, we get

$$H(F(1),\ldots,F(5)) \leq \frac{5}{2}\cdot\log(2m),$$

which gives a bound of $(2m)^{5/2}$ on the number of embeddings.

Exercise 1.4. *Check that the exponent of* $5/2$ *in the above bound is tight.*

The above method can also be used to give a tight estimate for any graph G (of constant size). In general, the exponent depends on a parameter known as the *fractional independent set number* of G. I will divide this proof in a few parts and add this as an extra problem in the homework. The solution to this problem need not be submitted.

The proof, along with many other combinatorial applications can also be found in the surveys by Radhakrishnan [Rad03] and [Gal14]. A generalization of Shearer's lemma was also used in the paper by Friedgut [Fri04] that we discussed in the previous lecture.

2 Mutual Information

The mutual information is a quantity which measures the amount of dependence between two random variables. Unlike correlation, which defines the random variables to take values in the same space, the mutual information can be defined for any two random variables. The mutual information between two random variables X and Y is defined by the formula

$$I(X;Y) \;=\; H(X) - H(X|Y)$$

Using the chain rule for entropy, we can see that

$$I(X;Y) \;=\; H(X) - H(X|Y) \;=\; H(Y) - H(Y|X) \;=\; H(X) + H(Y) - H(X,Y)\,.$$

We can use the first two expressions to observe that $I(X;Y) \geq 0$ and the last one to observe that $I(X;Y) = I(Y;X)$.

Example 2.1. *Consider the random variable* (X,Y) *with* $X \vee Y = 1$, $X \in \{0,1\}$ *and* $Y \in \{0,1\}$ *such that:*

$$(X,Y) = \begin{cases} 10 & \text{w.p } 1/3 \\ 01 & \text{w.p } 1/3 \\ 11 & \text{w.p } 1/3 \end{cases}$$

Then, we can calculate the entropy and mutual information as follows:

$$\begin{aligned} H(X) \;=\; H(Y) \;&=\; \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} \;=\; \log 3 - \frac{2}{3} \\ H(X,Y) \;&=\; \log 3 \\ I(X;Y) \;=\; H(X) + H(Y) - H(X,Y) \;&=\; \log 3 - \frac{4}{3} \end{aligned}$$

Conditioning on a third random variable Z, we can also define the conditional mutual information $I(X;Y|Z)$ as

$$\begin{aligned} I(X;Y|Z) &:= \mathop{\mathbb{E}}_{z}\left[I(X|Z=z;Y|Z=z)\right] \\ &= \mathop{\mathbb{E}}_{z}\left[H(X|Z=z) - H(X|Y,Z=z)\right] \\ &= H(X|Z) - H(X|Y,Z)\,. \end{aligned}$$

Consider the following example of three random variables.

Example 2.2. *Consider the random variable (X,Y,Z), $X \in \{0,1\}$, $Y \in \{0,1\}$ and $Z = X \oplus Y$ such that:*

$$(X,Y,Z) = \begin{cases} 000 & w.p\ 1/4 \\ 011 & w.p\ 1/4 \\ 101 & w.p\ 1/4 \\ 110 & w.p\ 1/4 \end{cases}$$

We can check that in this case, X,Y are independent and thus $I(X;Y) = 0$. However,

$$\begin{aligned} I(X:Y|Z) &= \mathop{\mathbb{E}}_{z}\left[I(X|Z=z;Y|Z=z)\right] \\ &= \frac{1}{2}I(X|Z=0;Y|Z=0) + \frac{1}{2}I(X|Z=1;Y|Z=1) \\ &= \frac{1}{2}\log 2 + \frac{1}{2}\log 2 \;=\; 1 \end{aligned}$$

The above example illustrates that unlike entropy, it is not true that conditioning (on average) decreases the mutual information. In the above example, while $I(X;Y) = 0$, we have $I(X;Y|Z) = 1$ which is in fact the maximum possible.

Recall that entropy provides theoretical limits on source coding, where the goal is to compress information when transmitting in a way such that whatever we send is received without any error. The concept of mutual information provides limits on transmission, when the transmission "channel" is noisy. We will discuss this in detail when we consider error-correcting codes, but it is instructive to consider the following example known as the "Binary Symmetric Chhannel".

Exercise 2.3. *Let X be a random variable supported on $\{0,1\}$, and let Y be a "noisy" copy of X, which is equal to X with probability $1-p$, and has the opposite value (0 is X is 1, and 1 if X is 0) with probability p. Calculate the maximum possible value of $I(X;Y)$ over all possible distributions for X. This is known as the* capacity *of the binary symmetric channel.*

As in the case of entropy, mutual information also obeys a chain rule.

Lemma 2.4. $I((X_1,\ldots,X_m);Y) = \sum_{i=1}^{m} I(X_i;Y|X_1,\ldots,X_{i-1})$

Proof: The chain rule for mutual information is a simple consequence of the chain rule for entropy. We have

$$\begin{aligned} I((X_1,\dots,X_m);Y) &= H(X_1,\dots,X_m) - H(X_1,\dots,X_m|Y) \\ &= \sum_{i=1}^{m} H(X_i|X_1,\dots,X_{i-1}) - \sum_{i=1}^{m} H(X_i|Y,X_1,\dots,X_{i-1}) \\ &= \sum_{i=1}^{m} [H(X_i|X_1,\dots,X_{i-1}) - H(X_i|Y,X_1,\dots,X_{i-1})] \\ &= \sum_{i=1}^{m} I(X_i;Y|X_1,\dots,X_{i-1}) \end{aligned}$$

■

References

[Fri04] Ehud Friedgut, *Hypergraphs, entropy, and inequalities*, The American Mathematical Monthly **111** (2004), no. 9, 749–760. 3

[Gal14] David Galvin, *Three tutorial lectures on entropy and counting*, arXiv preprint arXiv:1406.7872 (2014). 3

[Rad03] Jaikumar Radhakrishnan, *Entropy and counting*, Computational mathematics, modelling and algorithms **146** (2003). 3