**Information and Coding Theory** **Autumn 2022**

# Lecture 3: October 4, 2022

Lecturer: Madhur Tulsiani

## 1 Source Coding Theorem

We begin by recalling the Shannon Code. We considered a random variable $X$ that took on values $a_1, a_2, \ldots, a_n$ with probabilities $p_1, p_2, \ldots, p_n$. We wanted to encode the values of $X$ such that the expected number of bits needed is small. If $\ell_1, \ell_2, \ldots, \ell_n$ are the number of bits needed to encode $a_1, a_2, \ldots, a_n$, then we saw that a prefix free code exists iff:

$$\sum_{i=1}^{n} 2^{\ell_i} \;\leq\; 1$$

Furthermore, we saw that the expected length of the encoding is lower bounded by $H(X)$ and upper bounded by $H(X) + 1$ (a code as specified as above, the Shannon code may be constructed by setting $\ell_i = \lceil \log(1/p_i) \rceil$).

We will now try to improve this upper bound and we will do so by considering multiple copies of $X$. The idea is that by amortizing the loss over many symbols, we can start to approach an expected length equal to the lower bound i.e. the entropy.

The design may be done as follows: Consider $m$ copies of the random variable $X$, $\{X_1, \ldots, X_m \in U\}$ and a code $C : \mathcal{X}^m \to \{0,1\}^*$. Let $|\mathcal{X}|^m = N$. Now, we know that:

$$H(X_1, \ldots, X_m) \;\leq\; \sum_{i=1}^{N} p_i \left\lceil \log \frac{1}{p_i} \right\rceil \;\leq\; H(X_1, \ldots, X_m) + 1$$

Let us also assume that the $m$ copies of $X$ are drawn i.i.d. Using this assumption we try to work out the quantity $H(X_1, \ldots, X_m)$. Which may be expanded using the chain rule and independence:

$$\begin{aligned} H(X_1, \ldots, X_m) &= H(X_1) + H(X_2|X_1) + \cdots + H(X_m|X_1, \ldots, X_{m-1}) \\ &= H(X_1) + H(X_2) + \cdots + H(X_m) \\ &= m \cdot H(X) \end{aligned}$$

Therefore, we get

$$\mathbb{E}\left[|C(X_1, \ldots, X_m)|\right] \;\leq\; m \cdot H(X) + 1\,.$$

Thus, we used $H(X) + \frac{1}{m}$ bits on average per copy of $X$. This leads us to the source coding theorem.

**Theorem 1.1** (Fundamental Source Coding Theorem (Shannon)). *For all $\varepsilon > 0$ there exists a $n_0$ such that for all $n \geq n_0$ and given $n$ copies of $X$, $X_1, \ldots, X_n$ sampled i.i.d., it is possible to communicate $(X_1, \ldots, x_n)$ using at most $H(X) + \varepsilon$ bits per copy on average.*

## 2 Bounding binomial sums

We use the subadditivity property to obtain an upper bound on the number of subsets of $[n] = \{1, \ldots, m\}$ of size at most $k$ i.e., we need to bound size of the following set

$$S \;=\; \{(x_1, \ldots, x_n) \in \{0,1\}^n \;\mid\; x_1 + \cdots + x_n \leq k\} \;.$$

Of course we can write the following expression for the size of $S$

$$|S| \;=\; \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k},$$

but how much *is* the value of the above sum? We will estimate it in terms of the binary entropy function defined as

$$H_2(p) \;:=\; p \cdot \log\left(\frac{1}{p}\right) + (1-p) \cdot \log\left(\frac{1}{1-p}\right)$$

Note that $h(p)$ is the entropy of a random variable $X$, which takes value 1 with probability $p$ and 0 with value $1-p$ (or vice-versa). This immediately tells us that the maximum possible value of $H_2(p)$ is 1, which is achieved at $p = 1/2$. The function $H_2(p)$ can also easily be shown to be concave. It is also increasing for $p \in (0, 1/2)$ and decreasing for $p \in (1/2, 1)$

**Exercise 2.1.** *Prove that the function $H_2(p)$ is concave, using $H(Y|X) \leq H(Y)$.*

**Exercise 2.2.** *Prove that the function $H_2(p)$ is increasing when $p \in (0, 1/2)$. You can also prove this using $H(Y|X) \leq H(Y)$ if you like.*

We now return to the estimation problem. Let $(X_1, \ldots, X_n)$ be a uniformly distributed over $S$. Thus, we have that $H(X_1, \ldots, X_n) \;=\; \log|S|$. We can also use sub-additivity to say that

$$\log|S| \;=\; H(X_1, \ldots, X_n) \;\leq\; H(X_1) + \cdots + H(X_n) \;=\; n \cdot H(X_1)\,,$$

where the last equality used the symmetry of the variables $X_1, \ldots, X_n$. Now since $X_1$ was an indicator variable, let us say that it takes value 1 with probability $p$ and value 0 with probability $1-p$. Then $H(X_1) = H_2(p)$. Also, we have that $X_1 + \cdots + X_n \leq k$, which gives by symmetry that $p = \mathbb{E}[x_1] \leq k/n$. Finally, we note that since the function $H_2(p)$ is increasing for $p \leq 1/2$, we get $H(X_1) = H_2(p) \leq H_2(k/n)$. This gives the bound

$$\log|S| \;\leq\; n \cdot H_2(k/n) \quad \Rightarrow \quad |S| \;\leq\; 2^{n \cdot h(k/n)}\,.$$

You can check that the upper bound obtained here is not too bad since the sum is approximately equal to $\frac{2^{n \cdot H_2(k/n)}}{\sqrt{2\pi \cdot k \cdot (1-k/n)}}$.

## 3 Shearer's Lemma and Combinatorial Applications

The sub-additivity property of entropy lets us bound the entropy of the tuple $(X_1, \ldots, X_n)$ in terms of the individual entropies $H(X_1), \ldots, H(X_n)$. Shearer's lemma can be viewed as a generalization of this statement which lets us obtain better bounds in case we can estimate the entropy of subsets of random variables containing more than one random variable.

**Lemma 3.1** (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 3.2** (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 $\mu$ be such that $\forall i \in [n]$ $\mathbb{P}_{S \sim D}\left[i \in S\right] \geq \mu$. Then*

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

**Exercise 3.3.** *Check that Lemma 3.2 implies Lemma 3.1. Also check that both these lemmas imply sub-additivity.*

We will discuss the proof in the next lecture. You are encouraged to try proving the lemma, and also look at surveys by Radhakrishnan [Rad03] and Galvin [Gal14]. A generalization of Shearer's lemma was also used in the paper by Friedgut [Fri04], which is also a great one to read for many interesting applications to analysis and inequalities. We sketch the proof of one such inequality below, for bounding volumes of convex bodies.

### 3.1 Bounding volumes using projections

Consider a set of points $S$ in (say) three dimensions, such that the projections in the $xy$, $yz$ and $zx$ plain contain $n_1$, $n_2$ and $n_3$ points respectively. How many points can there be in the set $S$? Note that since many points in $S$ can have the same projection on a plane, the numbers $n_1$, $n_2$ and $n_3$ can each be much smaller than $|S|$. However, since two different points cannot have the same projection in *all three* planes, we know that each triple of projections must determine a unique point. This gives

$$|S| \;\leq\; n_1 \cdot n_2 \cdot n_3\,.$$

It turns out that we can significantly improve this bound using Shearer's lemma. Let $(X, Y, Z)$ be a triple of random variables denoting the coordinates of a uniformly sampled point from $S$. Thus, we have that $H(X, Y, Z) = \log |P|$. Moreover, using Shearer's lemma, we also get that

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

since the family of pairs on the right includes each random variable twice. Also, since $(X, Y)$ denotes the projection of a random point from $S$ in the $xy$ plane, and total number of projections is $n_1$, we get that $H(X,Y) \leq \log n_1$. Similarly, $H(Y,Z) \leq \log n_2$ and $H(Z,X) \leq \log n_3$. Combining these estimates gives

$$2 \cdot \log|S| \;\leq\; \log n_1 + \log n_2 + \log n_3 \quad\Rightarrow\quad |S| \;\leq\; \sqrt{n_1 \cdot n_2 \cdot n_3}\,.$$

Note that there is nothing special about three dimensions. One can also prove the following $d$-dimensional analogue using the same argument.

**Proposition 3.4.** *Let $S \subseteq \mathbb{R}^d$ be a finite set of points in $d$ dimensions, and let $S_1, \ldots, S_d$ denote the set of projections orthogonal to each of the $d$ coordinate axes. Then we have*

$$|S| \;\leq\; \left(\prod_{i=1}^{d} |S_i|\right)^{1/(d-1)}\,.$$

This can also be used to bound the volume of a body $B$ in $d$ dimensions in terms of the $(d-1)$-dimensional volumes of its projections. One can consider the body to be a union of axis parallel cubes, with a point at the center of each cube. Then, a limiting argument combined with the above estimate gives the following result known as the Loomis-Whitney inequality.

**Proposition 3.5** (Loomis-Whitney inequality)**.** *Let $B \subseteq \mathbb{R}^d$ be a measurable body and let $B_1, \ldots, B_d$ denote its projections orthogonal to each of the coordinate axes. Then, we have*

$$\mathsf{Vol}_d(B) \;\leq\; \left(\prod_{i=1}^{d} \mathsf{Vol}_{d-1}(B_i)\right)^{1/(d-1)}\,.$$

## 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