TTIC 31200 – Information and Coding Theory – Discussion 1

Kavya Ravichandran*

10 January 2025

1 Jensen's Inequality

Recall the statement of Jensen's Inequality:

Theorem 1.1. *Let $S \subseteq \mathbb{R}^n$ be a convex set and X be a random variable over S. Then for convex $f : S \to \mathbb{R}$, we have:* $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$ *. If f is concave, we instead have* $\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$ *.*

Exercise 1.2. *Prove Jensen's inequality when X has discrete support.*

Solution: W e prove it for convex f. The same argument holds for concave. Let k be the support size. We prove this by induction on k. Let α_i be the probabilities associated with support elements $x_i \in S$. Our first step when trying to prove anything is to write down what it is we want to prove:

$$\text{Want To Show:} \qquad \forall\{\alpha_1, \dots \alpha_k : \sum \alpha_i = 1\}, x_1, \dots, x_k \qquad \sum_{i=1}^{k} \alpha_i f(x_i) \geq f\left(\sum_{i=1}^{k} x_i\right).$$

Base case: $k = 2$ Suppose the size of the support is 2. Then:

$$\mathbb{E}[f(X)] = \alpha_1 f(x_1) + (1-\alpha_1) f(x_2) \tag{1}$$

$$\text{by convexity of } f \qquad \geq f(\alpha\, x_1 + (1-\alpha)\, x_2)\,. \tag{2}$$

Inductive Assumption Next, we make the inductive assumption. Let us assume that $\forall \alpha_1, \dots, \alpha_k$ such that $\sum_{i=1}^{k} \alpha_i = 1$ and $\forall\{x_1, \dots, x_k\} \subset S^k$, we have $\sum_{i=1}^{k} \alpha_i f(x_i) \geq f\left(\sum_{i=1}^{k} x_i\right)$.

Induction We now perform the induction. We consider $\sum_{i=1}^{k+1} \alpha_i f(x_i)$. Let us define $\alpha' = \alpha_k + \alpha_{k+1}$. Then, since $\sum_{i=1}^{k+1} \alpha_i = 1$, we have that $\sum_{i=1}^{k-1} \alpha_i + \alpha' = 1$. Further, define $x' = \frac{\alpha_k x_k + \alpha_{k+1} x_{k+1}}{\alpha_k + \alpha_{k+1}}$. Now, we have:

$$\sum_{i=1}^{k+1} \alpha_i f(x_i) = \alpha_1 f(x_1) + \cdots + \alpha_{k-1} f(x_{k-1}) + \alpha_k f(x_k) + \alpha_{k+1} f(x_{k+1}) \tag{3}$$

$$\text{by convexity of f} \qquad \geq \alpha_1 f(x_1) + \cdots + \alpha_{k-1} f(x_{k-1}) + \alpha' f(x') \tag{4}$$

$$\text{by inductive assumption} \qquad \geq f(\alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \alpha' x') \tag{5}$$

$$= f(\alpha_1 x_1 + \cdots + \alpha_{k-1} x_{k-1} + \alpha_k x_k + \alpha_{k+1} x_{k+1}) \tag{6}$$

Thus, we have shown this for all discrete supports.

Next, we use Jensen's Inequality to prove two fundmental inequalities.

Exercise 1.3. *(AM-GM Inequality) For $x_1, \dots, x_n \geq 0$,*

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n) \geq (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{1/n}$$

*This document draws heavily on recitation notes developed by Max Ovsiankin when TAing a previous offering of this course.

Solution: C ase 1, if anything is 0, this is trivially true. Case 2, $x_i > 0 \forall i$. In this case we can safely take the log of the left-hand side and use Jensen's inequality:

$$\log\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right) \geq \frac{1}{n}\left(\log(x_1) + \log(x_2) + \cdots + \log(x_n)\right) \tag{7}$$

$$= \frac{1}{n}\log\left(\Pi_{i=1}^{n}\right) \tag{8}$$

From here, we can exponentiate both sides to get the original inequality.

Exercise 1.4. *(Young's Inequality) For $a, b \geq 0, \forall p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$ab < \frac{a^p}{p} + \frac{b^q}{q}$$

Solution: S ince $1/p$ and $1/q$ sum to 1, if we interpret them as probabilities, the right hand side looks like the expectation of the following random variable:

$$Y = \begin{cases} a^p & \text{w.p. } \frac{1}{p} \\ b^q & \text{w.p. } \frac{1}{q} \end{cases} .$$

Then, consider another random variable $X := \log Y$. We have that:

$$X = \begin{cases} p\log a & \text{w.p. } \frac{1}{p} \\ q\log b & \text{w.p. } \frac{1}{q} \end{cases} .$$

Putting these together, we have that $\log \mathbb{E}[Y] = \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right)$ and $\mathbb{E}[\log Y] = \mathbb{E}[X] = p\log a \cdot \frac{1}{p} + q\log b\frac{1}{q} = \log a + \log b$. Applying Jensen's inequality, we have that $\mathbb{E}[\log Y] \leq \log \mathbb{E}[Y]$ and so $\log ab \leq \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right)$. As before, exponentiating gives us the original inequality.

2 Conditional and Joint Entropy

Recall from lecture the chain rule, $H(X,Y) = H(X) + H(Y|X)$, where $H(Y|X) = \mathbb{E}_x[H(Y|X = x)]$ and the fact that conditioning reduces entropy on average, i.e., $H(Y) \geq H(Y|X)$. Note that the latter fact only holds *on average*; indeed, it is possible that for a particular value of x , the entropy actually goes up, but on average it must go down.

For a picture of how conditional and joint entropy relate, we refer to Figure 2.2 in Cover and Thomas (reproduced here in Figure 1).

This is a useful picture to have in mind regarding the relationships between various quantities of interest. We have not yet defined $I(X;Y)$, mutual information, in lecture but at a high level, it measures how much we lose by representing a joint distribution over two variables by the product of the marginals.

Exercise 2.1. *Express $H(X,Y)$ in terms of $H(X), H(Y), and I(X;Y)$.*

Solution: T his gives rise to a principle of inclusion-exclusion!

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

Finally, we will get some practice with using the relationships between these quantities to prove statements we'd like to prove. Recall the binary entropy function:

$$H_2(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p} .$$

Exercise 2.2. *Show that $H_2(p)$ is concave in p using $H(Y|X) \leq H(Y)$.*

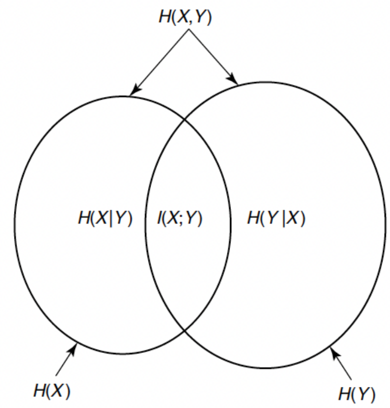

FIGURE 2.2. Relationship between entropy and mutual information.

Figure 1: Figure 2.2 from Cover and Thomas

Solution: A s always the best first step is to write down what we wish to prove:

$$\text{Want To Show:} \qquad H_2(\alpha p_1 + (1-\alpha)p_2) \geq \alpha H_2(p_1) + (1-\alpha)H_2(p_2)\,. \tag{9}$$

The right hand side looks like the expectation of a random variable, and the left hand side looks like the entropy of a Bernoulli random variable that comes up heads with probability $\alpha p_1 + (1-\alpha)p_2$. Let us try to define random variables accordingly. Let us start with:

$$X = \begin{cases} 1 & \text{w.p. } \alpha \\ 2 & \text{w.p, } 1-\alpha \end{cases}.$$

Now, we define another variable whose distribution depends on the outcome of X :

$$Y = \begin{cases} 0 & \text{w.p. } p_X \\ 1 & \text{w.p. } 1-p_X \end{cases}.$$

Let us compute $H(Y)$ and $H(Y|X)$. First, we have:

$$\begin{aligned} H(Y|X) &= \mathbb{P}\left[X=1\right] H(Y|X=1) + \mathbb{P}\left[X=2\right] H(Y|X=2) \qquad (10) \\ &= \alpha H_2(p_1) + (1-\alpha)H_2(p_2) \qquad (11) \end{aligned}$$

This is the right hand side of Eqn. 9. Without conditioning, we have that Y takes value 0 if $X=1$ and a coin of bias p_1 comes up heads or if $X=2$ and a coin of bias p_2 comes up heads. Similarly, Y takes value 1 if $X=1$ and a coin of bias p_1 comes up tails or if $X=2$ and a coin of bias p_2 comes up tails. Thus:

$$Y = \begin{cases} 0 & \text{w.p. } \alpha\, p_1 + (1-\alpha)\, p_2 \\ 1 & \text{w.p. } \alpha\,(1-p_1) + (1-\alpha)\,(1-p_2) \end{cases}.$$

This variable has entropy $H(Y) = H_2(\alpha\, p_1 + (1-\alpha)\, p_2)$.

Since we know that $H(Y) \geq H(Y|X)$, we have that $H_2(\alpha\, p_1 + (1-\alpha)\, p_2) \geq \alpha H_2(p_1) + (1-\alpha)H_2(p_2)$, implying that $H_2(p)$ is concave.