COMPSCI 690T Coding Theory and Applications | Apr 10, 2017

Lecture 19

Instructor: Arya Mazumdar | *Scribe: Names Redacted*

1 Information Theory

1.1 Entropy

Definition 1.1.1 Given some random variable X, and possible outcomes $\Omega_X = \{1, 2, ..., M\}$ such that $p(X = i) = p_i$, the entropy $H(X)$ is:

$$H(X) = -\sum_{i\in\Omega_X} p_i \log p_i$$

Definition 1.1.2 The joint entropy of two random variables X, Y is:

$$H(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y)$$

If X and Y are independant it's easy to see that $H(X, Y) = H(X) + H(Y)$.

Definition 1.1.3 The conditional entropy of two random variables X, Y is:

$$H(X|Y = y) = -\sum_{x} p(x|y) \log p(x|y)$$

Summing over all $y \in \Omega_Y$, sample space of Y, we get,

$$\begin{aligned}
H(X|Y) &= -\sum_{x\in\Omega_X, y\in\Omega_Y} p(y)p(x|y) \log p(x|y) \\
&= -\sum_{x,y} p(x, y) \log p(x|y) \\
&= -\sum_{x\, y} p(x, y) \log \frac{p(x, y)}{p(y)} \\
&= -\sum_{x,y} p(x, y) \log p(x, y) - p(x, y) \log p(y) \\
&= H(X, Y) - \sum_{y\in\Omega_Y} \log p(y) \sum_{x\in\Omega_X} p(x, y) \\
&= H(X, Y) - H(Y)
\end{aligned}$$

Proposition 1.1.1

$$H(X, Y) = H(Y|X) + H(X) \tag{1}$$

Proof

$$\begin{aligned}
H(X|Y) &= -\sum_{x\in\Omega_X, y\in\Omega_Y} p(x, y) \log p(x, y) + p(x, y) \log p(y) \\
&= H(X, Y) + \sum_{y} \log p(y) \sum_{x} p(x, y) \\
&= H(X, Y) - H(Y) \\
H(X, Y) &= H(X|Y) + H(Y)
\end{aligned}$$

■

By symmetry, we get:

$$
\begin{aligned}
&H(X|Y) + H(Y) = H(Y|X) + H(X) \\
\Rightarrow &H(X) - H(X|Y) = H(Y) - H(Y|X)
\end{aligned}
$$

Looking at the above expression, we can consider the quantity to be the amount of information about X obtained from Y, and we denote that as the mutual information between X and Y.

Definition 1.1.4 The mutual information between X and Y is $I(X;Y) = I(Y;X) = H(X) - H(X|Y) = H(Y) - H(Y|X)$

1.2 Relative Entropy & Bounds

First, we will look at some intuitive bounds.

Knowing Y can never reduce the amount of information you have on X, and therefore:

$$H(X) \geq H(X|Y) \tag{2}$$

Moreover, the amount of information in two random variables would be more than that in one single random variable, so:

$$H(X,Y) \geq H(X) \tag{3}$$

Now Lets define the quantity of Relative Entropy $D(p||q)$ for 2 probability mass functions of equal indeces, $p\{p_1, p_2, ...p_m\}$ and $q\{q_1, q_2, ...q_m\}$

$$D(p||q) = \sum_i p_i \log \frac{p_i}{q_i}$$

Proposition: $D(p(x,y)||p(x)p(y)) = I(X;Y)$
Proof

$$
\begin{aligned}
D(p(x,y)||p(x)p(y)) &= \sum_{X\in\Omega_X, y\in\Omega_Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{X\in\Omega_X, y\in\Omega_Y} p(x|y)p(y) \log \frac{p(x|y)}{p(x)} \\
&= \sum_{X\in\Omega_X, y\in\Omega_Y} p(x|y)p(y) \log p(x|y) - \sum_{X\in\Omega_X, y\in\Omega_Y} p(x|y)p(y) \log p(x) \\
&= \sum_{X\in\Omega_X, y\in\Omega_Y} p(x,y) \log p(x|y) - \sum_{X\in\Omega_X, y\in\Omega_Y} p(x) \log p(x) \\
&= -H(X|Y) + H(X) \\
&= I(X;Y)
\end{aligned}
$$

■

Proposition: $D(p||q) \geq 0$
Proof

$$
\begin{aligned}
D(p||q) &= \sum_i p_i \log_2 \frac{p_i}{q_i} \\
&= -\log_2 e \sum_i p_i \log \frac{q_i}{p_i} \\
&\geq -\log_2 e \sum_i p_i (\frac{q_i}{p_i} - 1) \\
&= -\log_2 e \sum_i q_i - p_i \\
&= -\log_2 e (1-1) = 0 \\
\Rightarrow D(p||q) &\geq 0
\end{aligned}
$$

where the $\geq$ comes from the inequality, $\log y \leq y - 1$ ■

Proposition: $\log_2 M \geq H(X)$ where M is $|\Omega|$
Proof Assume a uniform distribution, so $\forall i \in \Omega, q_i = \frac{1}{M}$. So $H(X) = -\sum_i \frac{1}{M} \log_2 \frac{1}{M} = \log_2 M$

$$
\begin{aligned}
D(p||q) &= \sum_i p_i \log \frac{p_i}{q_i} \\
&= \sum_i p_i \log p_i - \sum_i p_i \log_2 q_i \\
&= \sum_i p_i \log p_i - \sum_i p_i \log_2 \frac{1}{M} \\
&= -H(X) + \log_2 M \geq 0 \\
\Rightarrow \log_2 M &\geq H(X)
\end{aligned}
$$

Note that the ≥ 0 comes from the earlier proposition regarding relative entropy ■

1.3 Fano's inequality

Theorem Let S be a random variable with finite outcomes $|\Omega| = M$. Let X be another random variable that is represents S transmitted through some channel. $\hat{S}$ is the estimator for S that we derive from X with a function f such that $\hat{S} = f(X)$.

$$S \longrightarrow X \longrightarrow \hat{S}$$

Under this setup, the probability of error in estimating S is lower bounded by:

$$P_e = P(\hat{S} \neq S) \leq \frac{H(S|\hat{S}) - 1}{\log(M-1)}$$

Proof Let E be an indicator variable such that:

$$E = \begin{cases} 1, & \text{if } S \neq \hat{S} \\ 0, & \text{if } S = \hat{S} \end{cases}$$

$$\begin{aligned}
H(S|\hat{S}) &\leq H(S,E|\hat{S}) && \text{According to (3)} \\
&= H(E|\hat{S}) + H(S|E,\hat{S}) && \text{According to (1)} \\
&\leq H(E) + H(S|E,\hat{S}) && \text{According to (2)} \\
&= H(E) + P(E=1)H(S|E=1,\hat{S}) + P(E=0)H(S|E=0,\hat{S}) \\
&= h(P_e) + P_e \log(M-1) + 0 \\
\Rightarrow P_e &\geq \frac{H(S|\hat{S}) - 1}{\log(M-1)}
\end{aligned}$$

■

2 Wiretap Channel (Wyner and Ozarow 1984)

2.1 Problem Description

Alice would like to encode K bits of information to N bits and transmit it to Bob, but there is a wiretapper Eve, who could see any μ bits of the transmission. We would like to design a channel so as to minimize the amount of information leakage to Eve.

K-bit S —encode→ N-bit X —decode→ $\hat{S}$

Alice — Eve sees μ-bit Z — Bob

More formally, we have a K bit file S, and obtain an N bit file X for transmission. Let $T \subset \{1, 2, ..., N\}$ and $|T| = \mu$, and the information available to the wiretapper is $Z = X_T$.

The information leakage of this channel is defined to be:

$$\Delta = \min_T H(S|X_T)$$

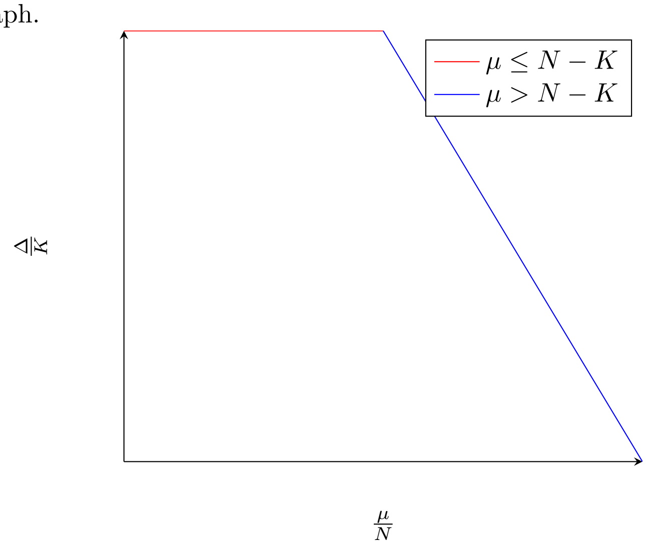
2.2 Parameters

The rate of this channel is $R = \frac{K}{N}$, the proportion available to the wiretapper is $\alpha = \frac{\mu}{N}$ and the relative information leakage is $\delta = \frac{\Delta}{K}$.

For these parameters, we have a $[K, N, \mu, \Delta]$, or a $[R, \alpha, \delta]$ scheme for the wiretap channel.

As an example, let's think about a $[N = 2, K = 1, \mu = 1]$ scheme. The way we encode S is that we get a random bit ζ and then encode $X = (\zeta, S + \zeta)$. We decode this by $\hat{S} = X_1 \oplus X_2$. This is correct because $\zeta \oplus \zeta + S = S$ no matter what the value of ζ is but the wiretapper gets no information if only allowed to view one bit.

Fix any $R = K/N$, the relations between these parameters could be characterized by the following graph.



Specifically, we know that:

$$\Delta \leq \begin{cases} K, & \text{if } 0 \leq \mu \leq N - K \\ N - \mu, & \text{if } \mu \geq N - K \end{cases}$$