# Information and Coding Theory

## MA41024/ MA60020/ MA60262

Bibhas Adhikari

Spring 2022-23, IIT Kharagpur

Lecture 4
January 16, 2023

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect?

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

Joint entropy Let $Z = (X, Y)$ be a pair of random variables with joint distribution $p(x, y)$. Then

$$H(Z) = H(X, Y) = \sum_{x,y} p(x, y) \log(1/p(x, y))$$

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

Joint entropy Let $Z = (X, Y)$ be a pair of random variables with joint distribution $p(x, y)$. Then

$$\begin{aligned} H(Z) &= H(X, Y) = \sum_{x,y} p(x, y) \log(1/p(x, y)) \\ &= \sum_{x,y} p(x)p(y|x) \log \frac{1}{p(x)} + \sum_{x,y} p(x)p(y|x) \log \frac{1}{p(y|x)} \end{aligned}$$

# Entropy

**Communication** Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

**Joint entropy** Let $Z = (X, Y)$ be a pair of random variables with joint distribution $p(x, y)$. Then

$$
\begin{aligned}
H(Z) &= H(X,Y) = \sum_{x,y} p(x,y)\log(1/p(x,y)) \\
&= \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(x)} + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)} \\
&= \sum_{x} p(x)\log\frac{1}{p(x)}\sum_{y} p(y|x) + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)}
\end{aligned}
$$

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

Joint entropy Let $Z = (X, Y)$ be a pair of random variables with joint distribution $p(x, y)$. Then

$$\begin{aligned} H(Z) &= H(X,Y) = \sum_{x,y} p(x,y)\log(1/p(x,y)) \\ &= \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(x)} + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)} \\ &= \sum_{x} p(x)\log\frac{1}{p(x)}\sum_{y} p(y|x) + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)} \\ &= H(X) + \sum_{x} p(x)H(Y|X=x) \end{aligned}$$

# Entropy

Communication Suppose we have a source rv $X$ and at the receiver end an output rv $Y$. The source letters are being transmitted through the channel. What do we expect? What is a channel?

Joint entropy Let $Z = (X, Y)$ be a pair of random variables with joint distribution $p(x, y)$. Then

$$\begin{aligned}
H(Z) &= H(X,Y) = \sum_{x,y} p(x,y)\log(1/p(x,y)) \\
&= \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(x)} + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)} \\
&= \sum_{x} p(x)\log\frac{1}{p(x)}\sum_{y} p(y|x) + \sum_{x,y} p(x)p(y|x)\log\frac{1}{p(y|x)} \\
&= H(X) + \sum_{x} p(x)H(Y|X=x) \\
&= H(X) + \mathbb{E}_x[H(Y|X=x)]
\end{aligned}$$

# Entropy

## Chain rule of entropy

Set $H(Y|X) = \mathbb{E}_x[H(Y|X = x)]$. Then we have

$$H(X, Y) = H(X) + H(Y|X)$$

# Entropy

## Chain rule of entropy

Set $H(Y|X) = \mathbb{E}_x[H(Y|X = x)]$. Then we have

$$H(X, Y) = H(X) + H(Y|X)$$

Similarly, we can obtain

$$H(X, Y) = H(Y) + H(X|Y)$$

# Entropy

## Chain rule of entropy

Set $H(Y|X) = \mathbb{E}_x[H(Y|X = x)]$. Then we have

$$H(X, Y) = H(X) + H(Y|X)$$

Similarly, we can obtain

$$H(X, Y) = H(Y) + H(X|Y)$$

**Homework** Let $(X, Y)$ be a joint random variable with $X \vee Y = 1$, $X \in \{0, 1\}$ and $Y \in \{0, 1\}$ such that $p(0, 1) = p(1, 0) = p(1, 1) = 1/3$. Then calculate $H(X)$, $H(Y)$, $H(Y|X = 0)$, $H(Y|X = 1)$, $H(Y|X)$, $H(X, Y)$

# Entropy

Proposition $H(Y) \geq H(Y|X)$

# Entropy

Proposition $H(Y) \geq H(Y|X)$

Proof

$$H(Y|X) - H(Y) \;=\; \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} - \sum_y p(y) \log \frac{1}{p(y)}$$

# Entropy

**Proposition** $H(Y) \geq H(Y|X)$

**Proof**

$$\begin{aligned} H(Y|X) - H(Y) &= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} - \sum_y p(y) \log \frac{1}{p(y)} \\ &= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} \\ &\quad - \sum_y p(y) \log \frac{1}{p(y)} \sum_x p(x|y) \end{aligned}$$

# Entropy

**Proposition** $H(Y) \geq H(Y|X)$

**Proof**

$$\begin{aligned}
H(Y|X) - H(Y) &= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} - \sum_y p(y) \log \frac{1}{p(y)} \\
&= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} \\
&\quad - \sum_y p(y) \log \frac{1}{p(y)} \sum_x p(x|y) \\
&= \sum_{x,y} p(x,y) \left( \log \frac{1}{p(y|x)} - \log \frac{1}{p(y)} \right)
\end{aligned}$$

# Entropy

**Proposition** $H(Y) \geq H(Y|X)$

**Proof**

$$\begin{aligned}
H(Y|X) - H(Y) &= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} - \sum_y p(y) \log \frac{1}{p(y)} \\
&= \sum_x p(x) \sum_y p(y|x) \log \frac{1}{p(y|x)} \\
&\quad - \sum_y p(y) \log \frac{1}{p(y)} \sum_x p(x|y) \\
&= \sum_{x,y} p(x,y) \left( \log \frac{1}{p(y|x)} - \log \frac{1}{p(y)} \right) \\
&= \sum_{x,y} p(x,y) \left( \log \frac{p(x)p(y)}{p(x,y)} \right)
\end{aligned}$$

# Entropy

Now let $W$ be a rv that takes the value $\frac{p(x)p(y)}{p(x,y)}$ with probability $p(x,y)$. Then using jensen's inequality

$$\sum_{x,y} p(x,y)\left(\log \frac{p(x)p(y)}{p(x,y)}\right) \leq \log\left(\sum_{x,y} \frac{p(x)p(y)}{p(x,y)} p(x,y)\right) = \log(1) = 0$$

# Entropy

Now let $W$ be a rv that takes the value $\frac{p(x)p(y)}{p(x,y)}$ with probability $p(x,y)$. Then using jensen's inequality

$$\sum_{x,y} p(x,y)\left(\log \frac{p(x)p(y)}{p(x,y)}\right) \leq \log\left(\sum_{x,y} \frac{p(x)p(y)}{p(x,y)} p(x,y)\right) = \log(1) = 0$$

Question What do you conclude ?

# Entropy

Now let $W$ be a rv that takes the value $\frac{p(x)p(y)}{p(x,y)}$ with probability $p(x, y)$. Then using jensen's inequality

$$\sum_{x,y} p(x,y) \left( \log \frac{p(x)p(y)}{p(x,y)} \right) \leq \log \left( \sum_{x,y} \frac{p(x)p(y)}{p(x,y)} p(x,y) \right) = \log(1) = 0$$

Question What do you conclude ?

Conditioning reduces entropy on average!!

# Entropy

Now let $W$ be a rv that takes the value $\frac{p(x)p(y)}{p(x,y)}$ with probability $p(x, y)$. Then using jensen's inequality

$$\sum_{x,y} p(x, y) \left( \log \frac{p(x)p(y)}{p(x, y)} \right) \leq \log \left( \sum_{x,y} \frac{p(x)p(y)}{p(x, y)} p(x, y) \right) = \log(1) = 0$$

Question What do you conclude ?

Conditioning reduces entropy on average!!
Homework $H(Y) = H(Y|X)$ if and only if $X$ and $Y$ are independent

# Entropy

Now let $W$ be a rv that takes the value $\frac{p(x)p(y)}{p(x,y)}$ with probability $p(x, y)$. Then using jensen's inequality

$$\sum_{x,y} p(x, y) \left( \log \frac{p(x)p(y)}{p(x, y)} \right) \leq \log \left( \sum_{x,y} \frac{p(x)p(y)}{p(x, y)} p(x, y) \right) = \log(1) = 0$$

Question What do you conclude ?

Conditioning reduces entropy on average!!
Homework $H(Y) = H(Y|X)$ if and only if $X$ and $Y$ are independent

Homework $H(Y|X, Z) \leq H(Y|Z)$

# Entropy

General case Suppose $\overline{X} = (X_1, X_2, \ldots, X_m)$.

# Entropy

General case Suppose $\overline{X} = (X_1, X_2, \ldots, X_m)$.

Homework Show (by induction) that

$$\begin{aligned} H(X_1, \ldots, X_m) &= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots \\ &\quad + H(X_m|X_1, \ldots, X_{m-1}) \end{aligned} \tag{1}$$

# Entropy

General case Suppose $\overline{X} = (X_1, X_2, \ldots, X_m)$.

Homework Show (by induction) that

$$H(X_1, \ldots, X_m) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots + H(X_m|X_1, \ldots, X_{m-1}) \quad (1)$$

Sub-additive property of entropy

$$H(X_1, \ldots, X_m) \leq H(X_1) + H(X_2) + \ldots + H(X_m)$$

# Entropy

General case Suppose $\overline{X} = (X_1, X_2, \ldots, X_m)$.

Homework Show (by induction) that

$$\begin{aligned} H(X_1, \ldots, X_m) &= H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots \\ &\quad + H(X_m|X_1, \ldots, X_{m-1}) \end{aligned} \tag{1}$$

Sub-additive property of entropy

$$H(X_1, \ldots, X_m) \leq H(X_1) + H(X_2) + \ldots + H(X_m)$$

Question Can the upper bound for expected code length of $H(X) + 1$ be improved?

# Entropy

## Recall

- Let $X$ be a rv with range set $\{a_1, \ldots, a_n\}$ and $p(a_i) = p_i$

# Entropy

## Recall

- Let $X$ be a rv with range set $\{a_1, \ldots, a_n\}$ and $p(a_i) = p_i$
- We want to encode $a_i$s with expected code length small i.e. expected number of bits needed is small

# Entropy

## Recall

- Let $X$ be a rv with range set $\{a_1, \ldots, a_n\}$ and $p(a_i) = p_i$
- We want to encode $a_i$s with expected code length small i.e. expected number of bits needed is small
- If $l_1, l_2, \ldots, l_n$ are the codeword lengths for $a_1, \ldots, a_n$ respectively then

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

# Entropy

## Recall

- Let $X$ be a rv with range set $\{a_1, \ldots, a_n\}$ and $p(a_i) = p_i$
- We want to encode $a_i$s with expected code length small i.e. expected number of bits needed is small
- If $l_1, l_2, \ldots, l_n$ are the codeword lengths for $a_1, \ldots, a_n$ respectively then

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

- We proved that the expected length is bounded below by $H(X)$ and bounded above by $H(X) + 1$ (Shannon code)

# Entropy

**Recall**

- Let $X$ be a rv with range set $\{a_1, \ldots, a_n\}$ and $p(a_i) = p_i$
- We want to encode $a_i$s with expected code length small i.e. expected number of bits needed is small
- If $l_1, l_2, \ldots, l_n$ are the codeword lengths for $a_1, \ldots, a_n$ respectively then

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

- We proved that the expected length is bounded below by $H(X)$ and bounded above by $H(X) + 1$ (Shannon code)

**Question** Can we improve the upper bound?

## The idea - Source Coding Theorem

## The idea - Source Coding Theorem

- Consider $m$ copies of the rv $X$, $X_1, \ldots, X_m$ and a code $C : \mathcal{X}^m \to \{0, 1\}^*$

## The idea - Source Coding Theorem

- Consider $m$ copies of the rv $X$, $X_1, \ldots, X_m$ and a code $C : \mathcal{X}^m \rightarrow \{0,1\}^*$
- Let $|\mathcal{X}|^m = N$

## The idea - Source Coding Theorem

- Consider $m$ copies of the rv $X$, $X_1, \ldots, X_m$ and a code $C : \mathcal{X}^m \to \{0,1\}^*$
- Let $|\mathcal{X}|^m = N$
- We know that

$$\mathbb{E}[|C(X_1, \ldots, X_m)|] \leq \sum_{i=1}^{N} p_i \lceil \log \frac{1}{p_i} \rceil \leq H(X_1, \ldots, X_m) + 1$$

## The idea - Source Coding Theorem

- Consider $m$ copies of the rv $X$, $X_1, \dots, X_m$ and a code $C : \mathcal{X}^m \to \{0,1\}^*$
- Let $|\mathcal{X}|^m = N$
- We know that

$$\mathbb{E}[|C(X_1,\dots,X_m)|] \le \sum_{i=1}^{N} p_i \lceil \log \frac{1}{p_i} \rceil \le H(X_1,\dots,X_m) + 1$$

- Assume that $m$ copies of $X$ are iid

## The idea - Source Coding Theorem

- Consider $m$ copies of the rv $X$, $X_1, \ldots, X_m$ and a code $C : \mathcal{X}^m \to \{0,1\}^*$
- Let $|\mathcal{X}|^m = N$
- We know that

$$\mathbb{E}[|C(X_1, \ldots, X_m)|] \leq \sum_{i=1}^{N} p_i \lceil \log \frac{1}{p_i} \rceil \leq H(X_1, \ldots, X_m) + 1$$

- Assume that $m$ copies of $X$ are iid
- Then

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

# Entropy

Thus we have

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

# Entropy

Thus we have

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

Thus we conclude that we can use $H(X) + \frac{1}{m}$ bits on average per copy of $X$

# Entropy

Thus we have

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

Thus we conclude that we can use $H(X) + \frac{1}{m}$ bits on average per copy of $X$

Theorem (Fundamental Source Coding Theorem (Shannon)). For any $\epsilon > 0$ 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 $H(X) + \epsilon$ bits per copy on average.