TP5: Information Theory

Yi Zhang

November 25, 2019

Abstract

1 Quantifiers of Information

U, V and W are three binary random variables. Their joint probability mass function is depicted as below:

U	V	W	$p(u,v,w)$
0	0	0	$\frac{1}{4}$
0	0	1	0
0	1	0	$\frac{1}{4}$
0	1	1	$\frac{1}{8}$
1	0	0	0
1	0	1	$\frac{1}{8}$
1	1	0	0
1	1	1	$\frac{1}{4}$

1.1 Entropy

Entropy of discrete r.v. $X \in \mathcal{X}$

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log_2 p_X(x)$$

Given the joint probability mass function, we have the U, V and W probability mass functions each as below:

U	$p_U(u)$
0	$\frac{5}{8}$
1	$\frac{3}{8}$

V	$p_V(v)$
0	$\frac{3}{8}$
1	$\frac{5}{8}$

W	$p_W(w)$
0	$\frac{1}{2}$
1	$\frac{1}{2}$

Then we have

$$H(U) = -p_U(u=0)\log_2 p_U(u=0) - p_U(u=1)\log_2 p_U(u=1) = 0.9544$$

$$H(V) = -p_V(v=0)\log_2 p_V(v=0) - p_V(v=1)\log_2 p_V(v=1) = 0.9544$$

$$H(W) = -p_W(w=0)\log_2 p_W(w=0) - p_W(w=1)\log_2 p_W(w=1) = 1$$

1.2 Conditional Entropy

Conditional entropy of r.v. $X \in \mathcal{X}$ given $Y \in \mathcal{Y}$ is defined by:

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p_Y(y) H(X|Y=y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 p_{X|Y}(x|y)$$

where the joint probability of r.v. $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ is defined by:

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)\cdot p_X(x) = p_{X|Y}(x|y)\cdot p_Y(y)$$

Given the joint probability mass function, we have the joint and conditional probability mass functions each as below:

U	V	$p_{U,V}(u,v)$
0	0	$\frac{1}{4}$
0	1	$\frac{3}{8}$
1	0	$\frac{1}{8}$
1	1	$\frac{1}{4}$

V	U	$p_{V,U}(v,u)$
0	0	$\frac{1}{4}$
0	1	$\frac{1}{8}$
1	0	$\frac{3}{8}$
1	1	$\frac{1}{4}$

W	U	$p_{W,U}(w,u)$
0	0	$\frac{1}{2}$
0	1	0
1	0	$\frac{1}{8}$
1	1	$\frac{3}{8}$

U	V	$p_{U\|V}(u\|v)$
0	0	$\frac{2}{3}$
0	1	$\frac{3}{5}$
1	0	$\frac{1}{3}$
1	1	$\frac{2}{5}$

V	U	$p_{V\|U}(v\|u)$
0	0	$\frac{2}{5}$
0	1	$\frac{1}{3}$
1	0	$\frac{3}{5}$
1	1	$\frac{2}{3}$

W	U	$p_{W\|U}(w\|u)$
0	0	$\frac{4}{5}$
0	1	0
1	0	$\frac{1}{5}$
1	1	1

Then we have

$$\begin{aligned}
H(U|V) = & \\
& - p_{U,V}(u=0,v=0) log_2 p_{U|V}(u=0|v=0) \\
& - p_{U,V}(u=0,v=1) log_2 p_{U|V}(u=0|v=1) \\
& - p_{U,V}(u=1,v=0) log_2 p_{U|V}(u=1|v=0) \\
& - p_{U,V}(u=1,v=1) log_2 p_{U|V}(u=1|v=1) \\
& = 0.9512
\end{aligned}$$

$$
\begin{aligned}
H(V|U) = & \\
& - p_{V,U}(v=0,u=0)log_2 p_{V|U}(v=0|u=0) \\
& - p_{V,U}(v=0,u=1)log_2 p_{V|U}(v=0|u=1) \\
& - p_{V,U}(v=1,u=0)log_2 p_{V|U}(v=1|u=0) \\
& - p_{V,U}(v=1,u=1)log_2 p_{V|U}(v=1|u=1) \\
& = 0.9512
\end{aligned}
$$

$$
\begin{aligned}
H(W|U) = & \\
& - p_{W,U}(w=0,u=0)log_2 p_{W|U}(w=0|u=0) \\
& - p_{W,U}(w=0,u=1)log_2 p_{W|U}(w=0|u=1) \\
& - p_{W,U}(w=1,u=0)log_2 p_{W|U}(w=1|u=0) \\
& - p_{W,U}(w=1,u=1)log_2 p_{W|U}(w=1|u=1) \\
& = 0.4512
\end{aligned}
$$

1.3 Mutual Information

Mutual information between two r.v. $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ is defined by:

$$
\begin{aligned}
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \\
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 \frac{p_{X|Y}(x|y)}{p_X(x)}\frac{p_Y(y)}{p_Y(y)} \\
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)\log_2 \frac{p_{X|Y}(x|y)}{p_X(x)}
\end{aligned}
$$

or

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

So we have

$$
\begin{aligned}
I(U;V) &= H(U) - H(U|V) = 0.0032 \\
I(U;W) &= H(W) - H(W|U) = 0.5488
\end{aligned}
$$

Given the joint probability mass function, we also have the joint and conditional probability mass functions each as below:

V	W	$p_{V,W}(v,w)$
0	0	$\frac{1}{4}$
0	1	$\frac{1}{8}$
1	0	$\frac{1}{4}$
1	1	$\frac{3}{8}$

V	W	$p_{V|W}(v|w)$
0	0	$\frac{1}{2}$
0	1	$\frac{1}{4}$
1	0	$\frac{1}{2}$
1	1	$\frac{3}{4}$

Then we can compute $I(U;V,W)$ using

$$I(U;V,W) = \sum_{u \in \mathcal{U}} \sum_{v,w \in \mathcal{V},\mathcal{W}} p_{U,V,W}(u,v,w) \log_2 \frac{p_{U,V,W}(u,v,w)}{p_U(u) p_{V,W}(v,w)}$$

So

$$\begin{aligned}
I(U;V,W) = & \\
& p_{U,V,W}(u=0,v=0,w=0) log_2 \frac{p_{U,V,W}(u=0,v=0,w=0)}{p_U(u=0)p_{V,W}(v=0,w=0)} \\
& + p_{U,V,W}(u=0,v=0,w=1) log_2 \frac{p_{U,V,W}(u=0,v=0,w=1)}{p_U(u=0)p_{V,W}(v=0,w=1)} \\
& + p_{U,V,W}(u=0,v=1,w=0) log_2 \frac{p_{U,V,W}(u=0,v=1,w=0)}{p_U(u=0)p_{V,W}(v=1,w=0)} \\
& + p_{U,V,W}(u=0,v=1,w=1) log_2 \frac{p_{U,V,W}(u=0,v=1,w=1)}{p_U(u=0)p_{V,W}(v=1,w=1)} \\
& + p_{U,V,W}(u=1,v=0,w=0) log_2 \frac{p_{U,V,W}(u=1,v=0,w=0)}{p_U(u=1)p_{V,W}(v=0,w=0)} \\
& + p_{U,V,W}(u=1,v=0,w=1) log_2 \frac{p_{U,V,W}(u=1,v=0,w=1)}{p_U(u=1)p_{V,W}(v=0,w=1)} \\
& + p_{U,V,W}(u=1,v=1,w=0) log_2 \frac{p_{U,V,W}(u=1,v=1,w=0)}{p_U(u=1)p_{V,W}(v=1,w=0)} \\
& + p_{U,V,W}(u=1,v=1,w=1) log_2 \frac{p_{U,V,W}(u=1,v=1,w=1)}{p_U(u=1)p_{V,W}(v=1,w=1)} \\
& = 0.6101
\end{aligned}$$

Using the chain rule for the mutual information

$$\begin{aligned}
I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\
I(X;Y,Z) &= I(X;Y) + I(X;Z|Y)
\end{aligned}$$

where conditional mutual information $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$

So we have

$$\begin{aligned}
I(X;Y,Z) &= H(X) - H(X|Z) + H(X|Z) - H(X|Y,Z) \\
&= H(X) - H(X|Y,Z) \\
I(X;Y,Z) &= H(X) - H(X|Y) + H(X|Y) - H(X|Y,Z) \\
&= H(X) - H(X|Y,Z)
\end{aligned}$$

Then we have

$$I(U;V,W) = H(U) - H(U|V,W)$$

So $H(U|V,W) = H(U) - I(U;V,W) = 0.3443$

1.4 Joint Entropy

General chain rule for entropy,

$$H\left(X_1, X_2, \ldots, X_n\right) = \sum_{i=1}^{n} H\left(X_i | X_{i-1}, \ldots, X_1\right)$$

So we have

$$H(U, V, W) = H(W) + H(V|W) + H(U|V, W) = 2.25$$

where $H(W) = 1.0$, $H(U|V, W) = H(U) - I(U; V, W) = 0.3443$ and

$$\begin{aligned} H(V|W) = & \\ & - p_{V,W}(v=0, w=0) log_2 p_{V|W}(v=0|w=0) \\ & - p_{V,W}(v=0, w=1) log_2 p_{V|W}(v=0|w=1) \\ & - p_{V,W}(v=1, w=0) log_2 p_{V|W}(v=1|w=0) \\ & - p_{V,W}(v=1, w=1) log_2 p_{V|W}(v=1|w=1) \\ & = 0.9056 \end{aligned}$$

Or we directly use the joint probability mass function to compute the entropy,

$$\begin{aligned} H(U, V, W) = & \\ & - p_{U,V,W}(u=0, v=0, w=0) log_2 p_{U,V,W}(u=0, v=0, w=0) \\ & - p_{U,V,W}(u=0, v=0, w=1) log_2 p_{U,V,W}(u=0, v=0, w=1) \\ & - p_{U,V,W}(u=0, v=1, w=0) log_2 p_{U,V,W}(u=0, v=1, w=0) \\ & - p_{U,V,W}(u=0, v=1, w=1) log_2 p_{U,V,W}(u=0, v=1, w=1) \\ & - p_{U,V,W}(u=1, v=0, w=0) log_2 p_{U,V,W}(u=1, v=0, w=0) \\ & - p_{U,V,W}(u=1, v=0, w=1) log_2 p_{U,V,W}(u=1, v=0, w=1) \\ & - p_{U,V,W}(u=1, v=1, w=0) log_2 p_{U,V,W}(u=1, v=1, w=0) \\ & - p_{U,V,W}(u=1, v=1, w=1) log_2 p_{U,V,W}(u=1, v=1, w=1) \\ & = 2.25 \end{aligned}$$

2 Communication Through Noisy Channels

The results of the experiments are shown as below,

repetitions	the fraction of erroneous words to the total	the rate of communication
0	0.399	/
3	0.156	0.333
5	0.033	0.2
11	0	0.091
23	0	0.043

$$R = \frac{\log_2 \mathcal{M}}{n} = \frac{1}{repetitions}, (0 \leq R \leq 1)$$

It is clear that the rate of communication decreases as the decreasing corresponding error fraction. In this term, the rate of communication could be interpreted as the number of the word changes. As a result, when we increase the number of repetitions, the rate of communication decreases, the corresponding error fraction decreases until 0.

3 Conclusions