GROUP MEMBERS

1.	OKOLIE CHIGOZIE EMMANUEL	-	192424
2.	OLALERE TAOFEEK	-	192439
3.	OLANIYAN TOLULOPE CHRISTOPHER.	-	192443
4.	OLANIYI SHEKINAT TOLANI.	-	192444
5.	OLOPADE OLUSEGUN DAMILOLA	-	192456
6.	ONIOSUN AYODELE ABRAHAM	-	192477
7.	ORIMOLADE IFEOLUWA STEPHEN	-	192487
8.	OSAGIE MOSES UYI	-	192489

QUESTIONS

1. Provide an information with the derived proof concept of entropy to the information theory.
2. Provide the chain rule for conditional entropy and relative entropy.
3. Provide the derived formula for mutual information using useful expressions that can be used in information theory.
4. Provide the chain rule for mutual information.

(1) PROOF CONCEPT OF ENTROPY

Entropy of a discrete RV can be defined as:

- a measure of uncertainty of a random variable
- X a discrete random variable

$X \sim \binom{x_i}{p_i}_{i \in I}$, $\mathcal{X}$ alphabet of X, $p(x) = P(X = x)$, mass function of X

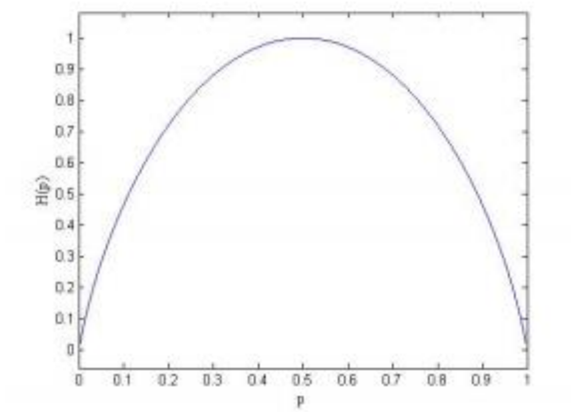

Figure 1: Graph of $H(p)$

Definition 1. The *entropy* of the discrete random variable X

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{1}$$

$$H(X) = E_p\left(\log \frac{1}{p(x)}\right) \qquad \text{equivalent expression} \tag{2}$$

- measured in *bits*!
- base 2! $H_b(X)$ entropy in base b; for $b = e$, measured in *nats*!
- convention $0 \log 0 = 0$, since $\lim_{x \searrow 0} x \log x = 0$

Entropy - Properties

Lemma 2. $H(X) \geq 0$

Lemma 3. $H_b(X) = \log_b a H_a(X)$

Example 4. Let the RV

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

$$H(X) = -p \log p - (1-p) \log(1-p) =: H(p) \tag{3}$$

$H(X) = 1$ bit when $p = \frac{1}{2}$. Graph in Figure 1

(2)

A. CHAIN RULE FOR CONDITIONAL ENTROPY

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

Proof.

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

Equivalently (shorter proof): we can write

$$\log p(X,Y) = \log p(X) + \log p(Y|X)$$

and apply E to both sides.

B. CHAIN RULE FOR RELATIVE ENTROPY

$$D\left(p(x,y)\parallel q(x,y)\right) = D(p(x)\parallel q(x)) + D\left(p(y|x)\parallel q(y|x)\right)$$

Proof.

$$\begin{aligned}
&D(p(x,y)||q(x,y)) \\
&= \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{q(x,y)} \\
&= \sum_x\sum_y p(x,y)\log\frac{p(x)p(y|x)}{q(x)q(y|x)} \\
&= \sum_x\sum_y p(x,y)\log\frac{p(x)}{q(x)} + \sum_x\sum_y p(x,y)\log\frac{p(y|x)}{q(y|x)} \\
&= D(p(x)\parallel q(x)) + D\left(p(y|x)\parallel q(y|x)\right).
\end{aligned}$$

(3) DERIVED FORMULA FOR MUTUAL INFORMATION

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

(4) CHAIN RULE FOR MUTUAL INFORMATION

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I\left(X_i; Y | X_{i-1}, X_{i-2}, \ldots, X_1\right).$$

Proof.

$$\begin{aligned}
&I(X_1, X_2, \ldots, X_n; Y) \\
&= H\left(X_1, X_2, \ldots, X_n\right) - H\left(X_1, X_2, \ldots, X_n | Y\right) \\
&= \sum_{i=1}^{n} H\left(X_i | X_{i-1}, \ldots, X_1\right) - \sum_{i=1}^{n} H\left(X_i | X_{i-1}, \ldots, X_1, Y\right) \\
&= \sum_{i=1}^{n} I\left(X_i; Y | X_1, X_2, \ldots, X_{i-1}\right)
\end{aligned}$$