# Information Theory and Coding.

## Module 1 - Basics of Information Theory.

* Introduction

→ Information is the source of a communication system, whether it is analog or digital.

→ Information theory is a mathematical approach to the study of coding of information along with the quantification, storage and communication of information.

* Condition of Occurrence of Events

→ There are mainly 3 conditions of occurrence.

① Condition of Uncertainity :- When an event has not occurred, that condition is called condition of uncertainity.

② Condition of Surprise :- When an ~~condition~~ event has just occurred, that condition is called condition of surprise

③ Condition of Information :- When an event has occurred but a while back, that condition is called information.

→ These conditions occur at different times, and helps us in gaining knowledge about probability of events.

* Measure of Information

Let us consider the communication system which transmits messages $m_1, m_2, m_3 \dots$ with probability of occurrence $P_1, P_2, P_3 \dots$ then the amount of information transmitted through the message $m_K$ with probability $P_K$ is

$$\text{Amount of Information } (I_K) = \log_2 \left( \frac{1}{P_K} \right)$$

Unit of Information :- Information is measured in bits {an abbreviation of a binary digit}

* Properties of Information.

① More the uncertainity of the message, more the amount of information carried.

② If the reciever knows the message being transmitted, amount of information is 0.

③ If $I_1$ is the info. carried by $m_1$ and $I_2$ is the info carried by $m_2$, then the amount of info carried compositely is $m_1 + m_2 \rightarrow I_1 + I_2$.

(4) If there are $M = 2^N$ equally likely messages, the amount of info carried is N bits.

Example - Calculate $I_K$ for $P_K = \frac{1}{4}$.

By Formula $I_K = \log_2\left(\frac{1}{P_K}\right)$

$$\therefore \quad I_K = \log_2\left(\frac{1}{1/4}\right) = \log_2(4) = \log_2(2^2)$$

$$= 2$$

$$\therefore \quad I_K = 2 \text{ bits.}$$

* Entropy.

→ Entropy can be defined as the measure of the average information content per source symbol.

The formula for calculating entropy is given as:-

$$H(M) = P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{1}{P_2}\right) + P_3 \log_2\left(\frac{1}{P_3}\right) \cdots$$

$$\therefore \quad H = \sum_{i=1}^{n} P_i \log_2\left(\frac{1}{P_i}\right)$$

H is measured in bits/message.

classmate

Date
Page

* Information Rate

Information Rate, also Known as transmission rale, is the measure of how much information can be transmitted over a communication channel per unit time. It is measured in bits per second (bps) or multiple for {Mbps, Gbps}

$$R = \frac{\text{Number of bits transmitted}}{\text{time to transmit the bits}}$$

It depends on various factors.

① Bandwidth :- The maximum amount of information that can be transmitted through a channel.

② Signal -to Noise Ratio (SNR) - Amount of noise present in the communication channel. High SNR means more info transmitted.

③ Channel characteristics :- It may also depend on the characteristics of the channel, such as attenuation, interference etc.

④ ~~Rae~~ Hardware - The devices used for transmission and recieving ~~bar~~ signals also affect IR. High power transmitter transmits more data.

* Joint and conditional Entropy.

→ Joint Entropy is the measure of uncertainity or random variables in the system.

→ Also refferred to as the entropy of the joint probability distribution of two or more random variables.

The formula for joint entropy of two discrete random variables is:-

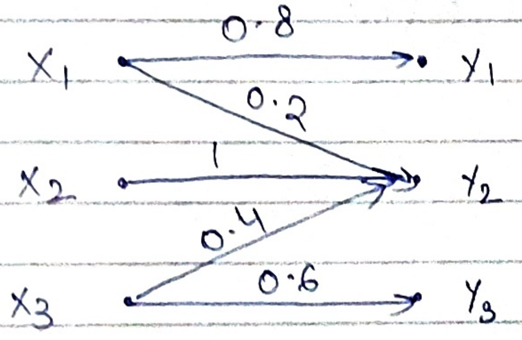
$$H(X,Y) = \sum \sum P(X,Y) \log_2 \left(\frac{1}{P(X,Y)}\right)$$

Where $P(X,Y)$ is the joint probability mass function of X and Y.

$$\boxed{H(X,Y) \leq H(X) + H(Y)}$$

Joint probability is usually given in two ways.

① Relational Probability Form.




This forms shows how X is contributing towards Y. It can be formed into a matrix by mapping.

② Matrix form.

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array} \begin{array}{c} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\ \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 1 & 0 \\ 0 & 0.4 & 0.6 \end{bmatrix} \end{array}$$

→ This value equals $P(Y/X)$

## * Conditional Entropy.

The average conditional self-information is called conditional entropy.

$$H(X/Y) = \sum_{j=1}^{m} \sum_{K=1}^{n} p(x_j, y_K) \log \left( \frac{1}{p(x_j/y_K)} \right)$$

$$H(Y/X) = \sum_{j=1}^{m} \sum_{K=1}^{n} p(x_j, y_K) \log \left( \frac{1}{p(y_K/x_j)} \right)$$

→ Relationship b/w Entropies.

$$H(XY) = H(X/Y) + H(X)$$

$$H(XY) = H(Y/X) + H(Y)$$

* Mutual Information for two discrete random variables.

Mutual Information is defined as the amount of information transferred where $x_i$ is transmitted and $y_j$ is recieved.

$$I(x_i, y_j) = \log \left[ \frac{P(x_i / y_j)}{P(x_i)} \right]$$

→ Average Mutual Information - Represented by $I(X;Y)$ and calculated in bits/symbol.

It is the amount of source information gained per recieved symbol.

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j)\, I(x_i, y_j)$$

* Channel Models:

→ A channel model is a mathematical model that describes how information is transmitted through a communication channel.

There are several types of channel models,

1. Binary Symmetric Channel (BSC) - In this model the channel can transmit two symbols {0,1} but there is a probability that each transmitted bit is flipped or corrupted at transmission

2. Memory channel. - It is a communication channel that has a memory or a history of previously transmitted symbols. It can affect the probability distribution of the current recieved symbols

3. Memory-less channels - The probability of the current recieved symbol depends only on the current transmitted symbols, not depending on symbols transmitted previously.

4. Finite-State Channels. - The probability of the current recieved symbols depend on a finite number of previously transmitted symbols.

Example - Gilbert-Elliot channel.

classmate

Date

Page

(4) Infinite - State channels - The probability of current recieved symbol depends on an Infinite number of previously transmitted symbols.

Example - AWGN model, Rayleight fading channel.

* Some additional models.

(1) Additive White Gaussian Noise Channel (AWGN)

In this model, the transmitted signal is corrupted by random Gaussian noise.

(2) Rayleigh Fading Channel. - Mainly used to describe Wireless communication channels. It takes into account that signals transmitted over wireless channels may be reflected, diffracted and scattered by various objects.

(3) Erasure Channel - Channel can transmit symbols from a given alphabet, but there is a probability p that each symbol will be lost (or erased) during transmission.

* Channel capacity.

→ Channel capacity refers to the maximum amount of information that can be transmitted over a channel under given conditions.

→ Channel capacity is an important metric for designing communication systems and determining the maximum data rate that can be acheived over a given channel.

* Shannon's Theorem.

→ Shannon's Theorem, also known as Shannon-Hartl Theory, is a fundamental result in information theory that describes the maximum possible data rate that can be reliably transmitted over a noisy communication channel.

This theorem provides a mathematical formula for calculating theoritical maximum capacity of a communication channel. in bits per second based on bandwidth, SNR {Signal to Noise Ratio}

The formula is given by:-

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

Where B is the badd width of the channel in Hz.
S is the signal power. (in Watts)
N is the Noise power (in Watts)
C is the maximum channel capacity.

* Some Applications of this theorem.

1. Optimization of transmission rates
2. Design for error - correction and detection.
3. Selection of appropriate modulation schemes.

---

* Priori and Post - Priori entropy.

→ A Priori refers to knowledge which is independent of experience or observation. It is knowledge which is gained through deduction, reason or intuition

→ A post - priori, refers to knowledge which was gained through personal experiences. It is gained through empirical evidence or sensory experience

---

* Equivocation.

→ Equivocation refers to the amount of uncertanity that remains in a message even after some part of it has been transmitted.

→ Equivocation occurs when the message being transmitted contains symbols or words with multiple meanings.

→ Depends on factors like :-

1. Noise in the communication channel.
2. Limitations in the encoding and decoding processes.

---

* Proofs.

1. Relation between Condition Entropy, Joint Entropy, and Marginal Entropy.

Given - $H(X,Y) = -\sum\sum p(x,y) \log p(x,y)$

To Prove - $H(X,Y) = H(X) + H(Y|X)$

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

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

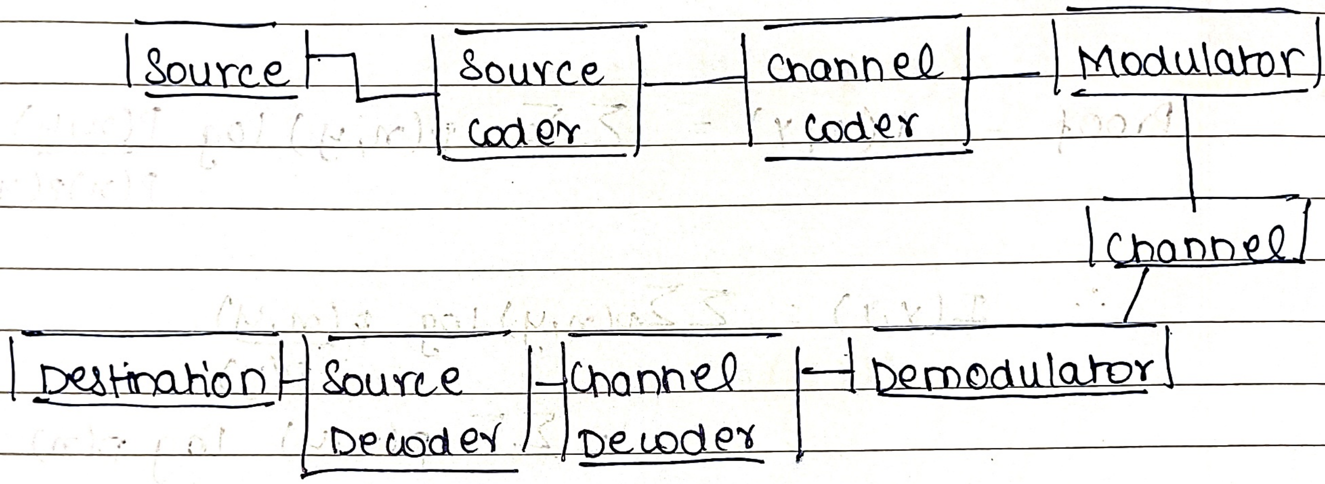
$$= -\sum_{x}\sum_{y} p(x,y) \log p(x)$$

$$-\sum_{x}\sum_{y} p(x,y) \log p(y|x)$$

$$- \sum_{x} p(x) \log p(x) - \sum_{x} \sum_{y} p(x,y) \log p(y|x)$$

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

* Transmitter - Reciever Diagram.




Channel capacity per symbol

$$C_s = \max_{\{P_{x_i}\}} I(X;Y) \text{ b/symb.}$$

Channel capacity per second (C)

$$C = r\, C_s \text{ b/s}$$

classmate

Date

Page

② Mutual Information and Entropy

Given - $p(x,y)$ is the Joint Probability for $x \in X$ and $y \in Y$.

To Prove - $I(X,Y) = H(X) - H(X|Y)$
$H(Y) - H(Y|X)$

Proof - $I(X,Y) = \sum\sum p(x,y) \log \frac{P(x,y)}{P(x)P(y)}$

$\therefore$ $I(X,Y) = \sum\sum p(x,y) \log \frac{p(x,y)}{p(y)}$
$- \sum\sum p(x,y) \log p(x)$

We know $P(x,y) = P(x|y)\, p(y)$

$\therefore$ $I(X,Y) = \sum_x \sum_y p(x|y)\, p(y) \log p(x|y)$
$- \sum\sum p(x,y) \log p(x)$

$\therefore$ $I(X,Y) = \sum_y p(y) \sum_x p(x|y) \log p(x|y)$

$- \sum_x \sum_y p(x,y) \log (p(x))$

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

$$\therefore \; H(X) = -\sum_{x \in X} p(x) \log(p(x))$$

$$H(X|Y) = \sum_{y \in Y} p(y) \, H(X|y = y)$$

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

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

→ Lossless channel - A matrix where there is atleast one non-zero element in each column.

→ Deterministic channel - A matrix where there is atleast one non-zero element in each row.

→ Noiseless channel - A matrix which is both lossless and deterministic.

→ Binary symmetric channel (BSC) - Has Two