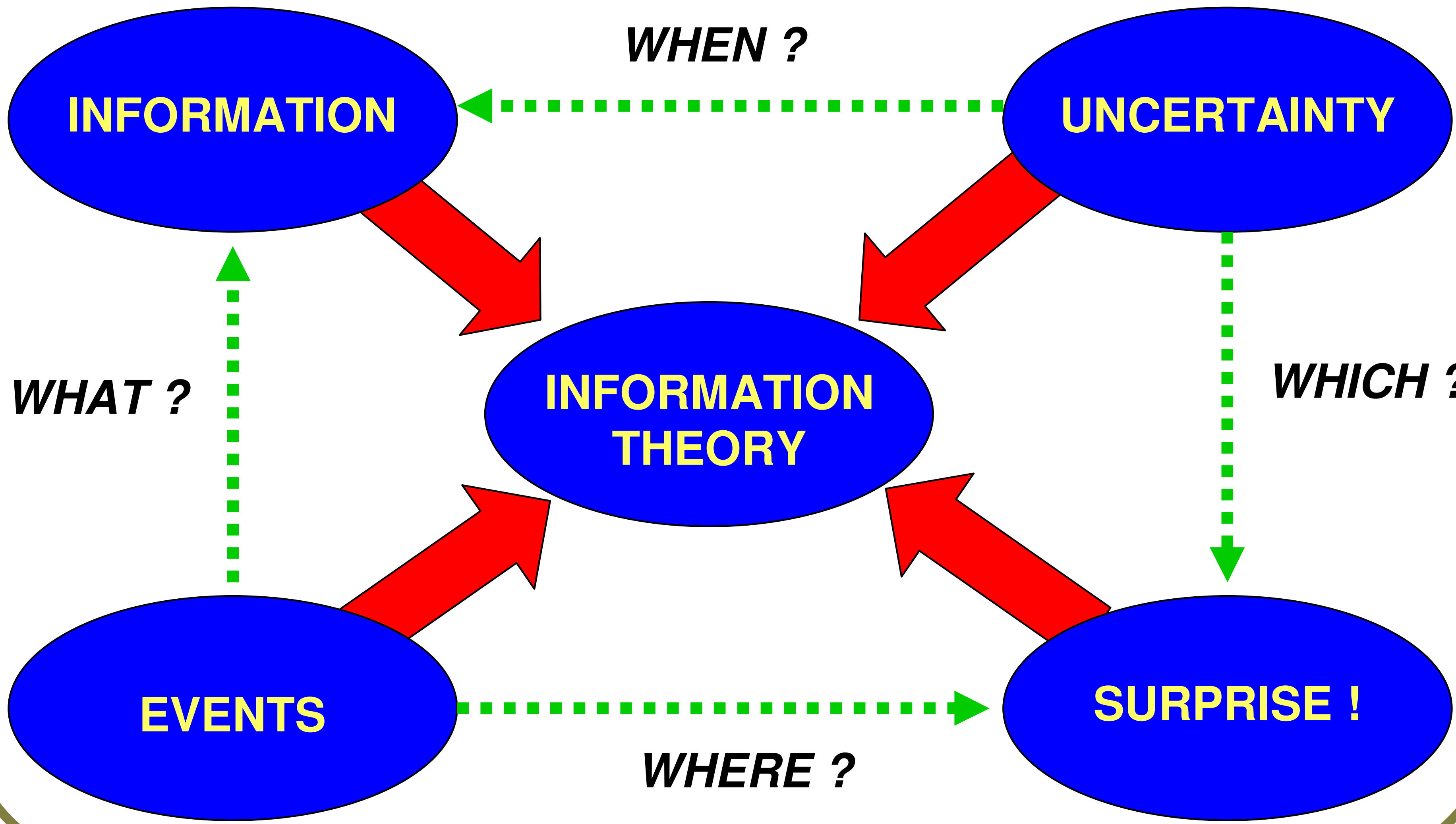
**FR. CONCEICAO RODRIGUES COLLEGE OF ENGINEERING**
Fr. Agnel Ashram, Bandstand, Bandra (W), Mumbai – 400050

# Module 6.0
# Information Theory

**Jayen Modi**
**Assistant Professor**
**Electronics Engineering Department**

**Electronic Circuits & Communication Fundamentals**
**ECCF (CSC 304) for S.E. (Computer Engg.) – Semester III**

# Introduction to Information Theory



# Introduction to Information Theory

**EVENT :-** It is regarded as the result or outcome of any certain experiment or process (E)

**UNCERTAINTY :-** It is a state in which the outcome or a result (event) isn't available till that event has occurred

**INFORMATION :-** It is conclusion drawn, data collected or knowledge obtained when a particular event occurs

**SURPRISE :-** It is the reaction towards the information obtained when a particular event has occurred

***There's definitely a direct connection between events, uncertainty, information & surprise !***

# Introduction to Information Theory

**We'll study the following main modules which are associated with information theory :-**

1. **Information (Concepts & Properties)**
2. **Entropy (Definition & Properties)**
3. **Information Rate (R)**
4. **Channel Capacity (C)** → ***Shannon's Theorem for Channel Capacity of a Gaussian Channel***

***Channel Capacity Theorem***

# Concept of Information

**Consider an experiment or process (E) that generates following possible events or outcomes as shown :-**

$$\boxed{E = \{e_1, e_2, e_3 \ldots\ldots e_N\}}$$

**If each independent event or outcome ($x_i$) has its own probability of $p(e_1)$, $p(e_2)$, $p(e_3)$ ….. $p(e_N)$ satisfying :-**

$$\boxed{p(E) = p(e_1) + p(e_2) + \ldots\ldots + p(e_N) = 1}$$

$$p(E) = \sum_{n=1}^{N} p(e_n) = 1$$

***sum of all probabilities should be equal to unity***

# 1. Tossing (Flipping) of Coin

**Experiment or Process (E) = Tossing of Coin**

***Here tossing (flipping) of an unbiased coin will result only in two fixed (definite) possible outcomes – <u>Heads or Tails</u>***


**Obverse** **Reverse**

# 1. Tossing (Flipping) of Coin


**Heads ($e_1$)** **Tails ($e_2$)**

$$\underbrace{\hspace{3cm}}$$

**Experiment or Process (E)**

$$\boxed{E = \{e_1, e_2\}}$$

# 2. Rolling of Dice (Game)

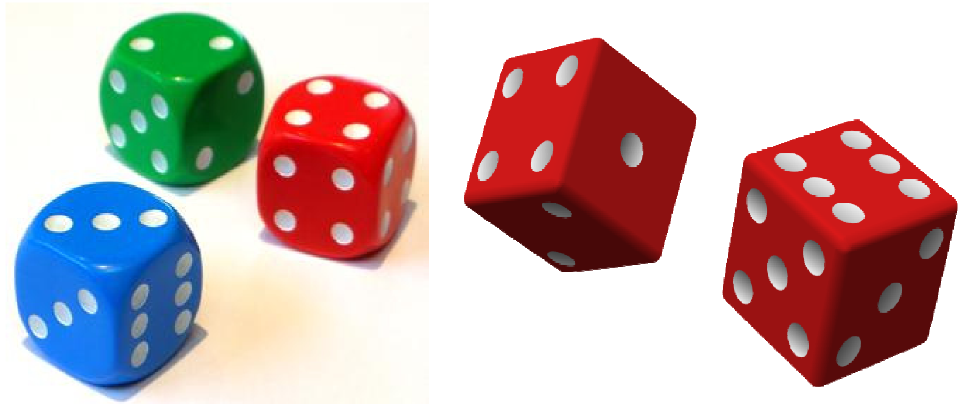
**Experiment or Process (E) = Rolling of Dice**

***Here rolling of unbiased dice 'N' times will result in one of the top surface numbers to be displayed at least one (1 to 6)***

***This implies that all numbers have an equal probability of appearing atleast once (1/6) for 'N' different rolling times***

# 2. Rolling of Dice (Game)

Experiment or Process (E)



***Outcome results in each number (1-6) being displayed on the dice surface at least once for 'N' rolls***

$$E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$$

# Concept of Information

Total amount of information obtained from occurrence of each individual event ($e_n$) is characterized by :-

$$I(e_n) = \log_a\left(\frac{1}{p(e_n)}\right)$$

*thus information obtained is inversely proportional to the probability of its occurrence*

Unit of this information $I(e_n)$ depend upon the logarithm base (a) & may be one of the following types :-

1. Bit/s if the base (a) = 2
2. Nat/s if the base (a) = e
3. Decit/s if the base (a) = 10

*base 2 is mainly used in information theory*

# Concept of Information

**Unit of this information $I(x_i)$ depend upon the logarithm base (a) & may be one of the following types :-**

$$I(e_n) = log_2\left(\frac{1}{p(e_n)}\right) \Rightarrow$$ **unit is in Bit/s for base (a) = 2**

$$I(e_n) = log_e\left(\frac{1}{p(e_n)}\right) \Rightarrow$$ **unit is in Nat/s for base (a) = e**

$$I(e_n) = log_{10}\left(\frac{1}{p(e_n)}\right) \Rightarrow$$ **unit is in Decit/s for base (a) = 10**

# Concept of Information

- Information is the measure of an outcome of a certain event ($e_n$) in an experiment (E)
- It is proportional to the inverse of probability of that event, $p(e_n)$ of that experiment (E)
- Mathematically it is represented as negative logarithm of probability of that event $p(e_n)$
- If an event is more likely to occur, then very little information can be obtained from it
- But if a less likely event occurs, information obtained from it will be very high

# Properties of Information

**The concept of information has the following statistical & mathematical properties :-**

- An event which is likely to occur, $p(e_n) = 1$ will carry no amount of information, $I(e_n) = 0$
- If an event $E = e_n$ occurs, it carries little or no information, but information is never lost
- The less probable an event is $p(e_1) < p(e_2)$ so more information is obtained i.e. $I(e_1) > I(e_2)$
- For two independent events $e_1$ & $e_2$ the total information obtained is their sum $I(e_1) + I(e_2)$

# Properties of Information

**The concept of information has the following statistical & mathematical properties :-**

- When $p(e_n) = 1$, then $I(e_n) = 0$
- Always $I(e_n) \geq 0$ for $0 \leq p(e_n) \leq 1$
- For $p(e_1) > p(e_2)$ thus $I(e_1) < I(e_2)$
- $I(e_1 \cdot e_2) = I(e_1) + I(e_2)$

*please refer to your class notes for entire description of properties*

# Average Information (Entropy)

- **Average information of an experiment (E) is of more interest than the information of each event ($e_n$)**
- **The average information associated with outcome of experiment (E) is called as the entropy**
- **Mathematically depends only upon the probabilities of sum of all events or outcomes of experiment (E)**

$$H(E) = \sum_{n=1}^{N} p(e_n) \log\left(\frac{1}{p(e_n)}\right)$$

***average information content of outcome of event in particular experiment (E)***

# Properties of Entropy – H(E)

For any system having N possible events, the entropy H(E) is given by the following equation :-

$$\boxed{0 \leq H(E) \leq \log_2 N}$$

where N is the total number of events of experiment (E) whose entropy H(E) has the following properties :-

1. H(E) = 0 if probability of some event $e_n$ is $p(e_n) = 1$ & all other events have a zero (0) probability

2. $H(E) = \log_2 N$ if probability of all events are equal for the entire experiment where $p(x_i) = 1/N$ for all $e_n$

# Information Rate (R)

- Information rate is defined as an average no. of bits of information per second (sec.)
- Assume a message source generates an ‘r’ amount of messages (events) per second
- Entropy H(E) is defined as the average no. of information bits per message
- Hence from the above, information rate (R) is mathematically expressed by :-

$$\boxed{R = r \cdot H}$$

***depends upon entropy (H) & also rate of messages per second (r)***

# Information Rate (R)

$R = r \times H$ → $r$ = No. of messages/sec.
$H$ = Information/messages

$$R = \left( r = \frac{\text{messages}}{\text{sec.}} \right) X \left( H = \frac{\text{information}}{\text{messages}} \right)$$

$$R = \left( r = \frac{\cancel{\text{messages}}}{\text{sec.}} \right) X \left( H = \frac{\text{information}}{\cancel{\text{messages}}} \right)$$

$$\boxed{R = \frac{\text{information}}{\text{sec.}} = \frac{\text{bits}}{\text{sec.}}}$$

# Channel Capacity (C)

**Transmission efficiency or channel efficiency defined as ratio of the actual information transmitted to overall maximum information transmitted over the channel**

$$\eta_c = \frac{\text{Actual Information Transmitted}}{\text{Maximum Information Transmitted}}$$

***ideally the channel efficiency should be unity i.e. there should be no error in maximum information transmitted through channel (actual information)***

# The Channel Capacity Theorem

- Channel capacity theorem deals with rate of information transmission over a channel
- It takes into account the channel capacity ‘C’ & source with positive information rate (R)
- For error free transmission over the channel it is imperative that we should have $R \leq C$
- Conversely if information rate (R) exceeds a channel capacity (C), then an error occurs

$$\boxed{R \leq C}$$

*called Shannon’s Theorem for channel capacity (C)*

# The Channel Capacity Theorem

**Statement of Shannon's Theorem for channel capacity can be stated as shown below :-**

***Given source of 'N' likely (same probability) messages with N >> 1 which generates information at rate 'R' for channel having capacity 'C' then if $R \leq C$, there exists a coding technique such that output of a source may be transmitted over the channel with a probability of error of receiving the message may be arbitrarily small.***

**Important feature of this theorem is that it indicates when $R \leq C$, error free transmission is possible in presence of noise, in the communication channel**

# Capacity of Gaussian Channel

**Channel capacity of a (white noise) bandlimited Gaussian channel is given by the equation :-**

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

***Shannon – Hartley theorem for Gaussian channel capacity***

**B = Channel Bandwidth (Hz)**
**S = Signal Power (W)**
**N = Noise Power (W)**
**η = Power Spectral Density (W/Hz)**

***η = N/B (average unit power per unit bandwidth)***

# Capacity of Gaussian Channel

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

*for noiseless channel, N = 0 hence $C \to \infty$ meaning noiseless channel has infinite channel capacity (C)*

**However for a infinite bandwidth as B → ∞ the channel capacity does not become ∞ because as bandwidth (B) increases, the noise (N) also increases**

$$\lim_{B \to \infty} C = 1.44\left(\frac{S}{\eta}\right)$$

*channel capacity (C) saturates & it becomes constant at a value called as the Shannon's Limit (as shown)*

***Hence even as channel bandwidth (B) appears to be infinite (∞) channel capacity (C) becomes constant***

# Capacity of Gaussian Channel

- **Ideally as bandwidth (B) increases, channel capacity (C) also increases initially (linearly)**
- **With increase in bandwidth (B) the effect of noise (N) also increases in the channel**
- **Since B $\rightarrow \infty$ then also C $\rightarrow \infty$ as given by the Shannon – Hartley theorem equation**
- **However as channel noise (N) also increases channel capacity (C) reaches constant value**
- **This constant value for the channel capacity (C) is called as the Shannon's Limit**