101001/EC600C

INFORMATION THEORY & CODING

Anila Kuriakose,
Assistant Professor, Department of ECE, RSET

Introduction

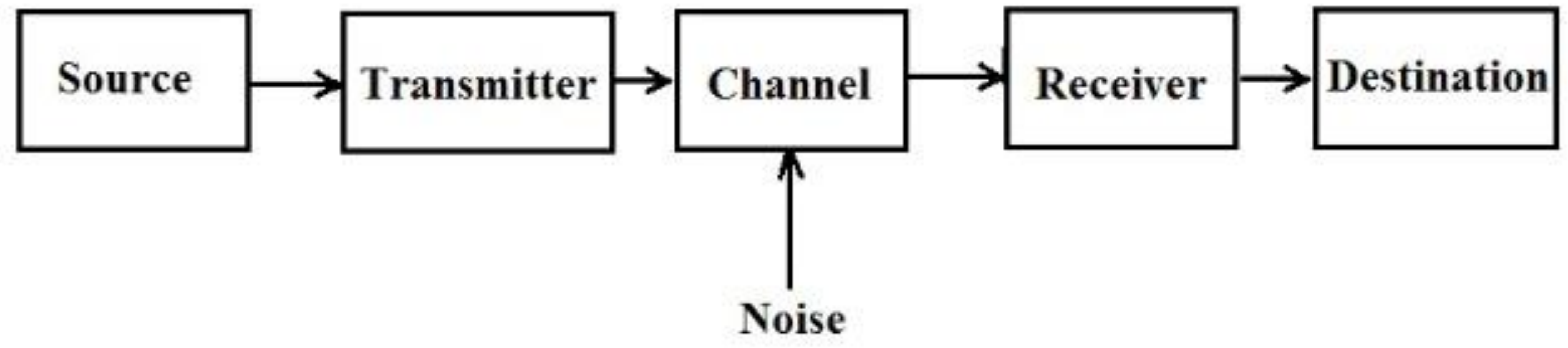
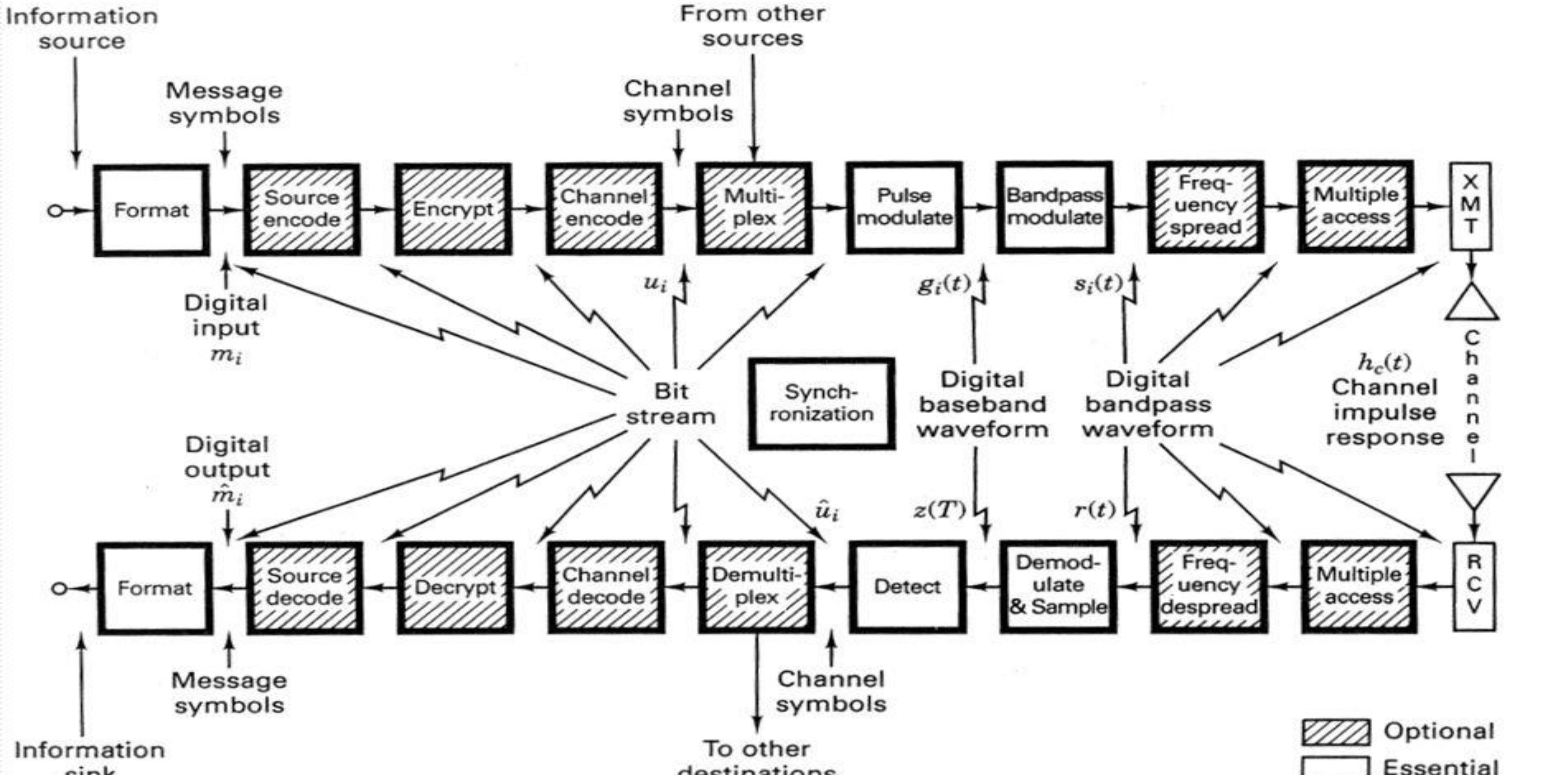


Fig. Basic Communication System Model

Digital communications system block diagram



About 101001/EC600C ITC

- No. of Credits – 4

Prerequisites

- 100908/MA100A Linear Algebra and Calculus
- 100902/MA400A Probability, Random Process and Numerical Methods
- 100902/EC400C Signals and Systems.

Text Books:

1. P S Satyanarayana “Concepts of Information Theory & Coding” , 2016
2. Joy A Thomas, Thomas M Cover *Elements of Information Theory*, Wiley-Interscience.
3. Shu Lin & Daniel J. Costello. Jr., *Error Control Coding : Fundamentals and Applications*, 2nd Edition.

Course Outcomes

After the completion of the course the student will be able to

CO1: Explain measures of information – entropy, conditional entropy, mutual information

CO 2: Apply Shannon's source coding theorem for data compression.

CO 3: Apply the concept of channel capacity for characterize limits of error-free transmission.

CO 4: Apply linear block codes for error detection and correction.

CO5: Apply algebraic codes with reduced structural complexity for error correction.

CO 6: Understand encoding and decoding of covolutional and LDPC codes.

Mark Distribution

Total	CIE				ESE
	Attendance	Internal Examination	Assignment/Quiz/ Course Project	Total	
150	10	25	15	50	100

Assignment (5 marks) + Course Project (10 marks)

Information Theory

- Information
- Entropy
- Properties of Entropy
- Joint and Conditional Entropy
- Mutual Information
- Properties of Mutual Information.

INTRODUCTION

- "Information Theory" is one of the most imp. application of 'Probability Theory' which has extensive potential application to 'Communication Systems'.

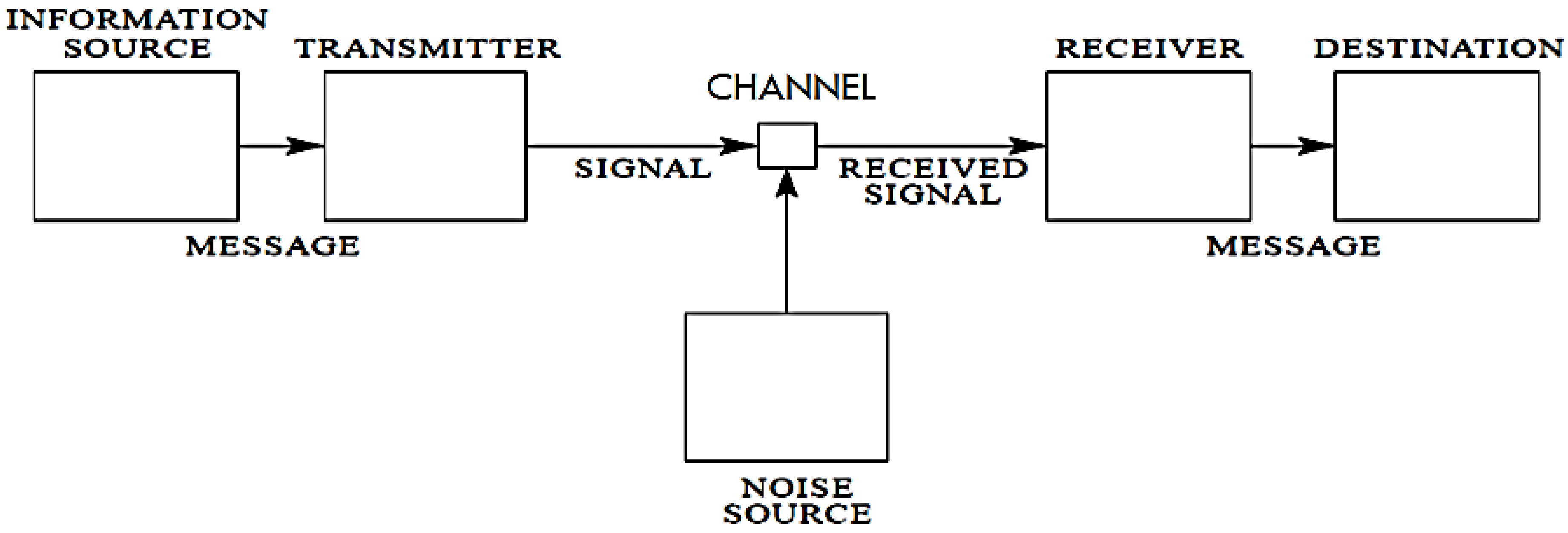


Fig: 1.1 General communication system from Shannon's "A Mathematical Theory of Communication"

Measure of Information

- The most significant feature of information → **uncertainity or unpredictability.**
- **Lets discuss with a few examples …**

- E.g. Weather forecast of Kerala
  - Sun will rise in the east (p=1) → Totally certain (least information)
  - It will rain (p=1/3) → uncertainty (fair amount of information)
  - There will be tornadoes (p=1/100000) → Total surprise (maximum information)

Conclusion: The amount of information increases when probability decreases.

Amount of Information (I)

- Consider a source that produces messages A and B with probabilities P(A) and P(B).
- The amount of information associated with the event A with probability P(A) is

$$I(A) = \Phi\{ P(A) \}$$

Where $\Phi\{.\}$ is the function to be determined.

Amount of Information (I)

3. More information will be conveyed by a less probable message.

If P(A) < P(B), then I(A) > I(B)

4.When the events A and B are statistically independent, the compound message C=A,B has the information

I(C) = I(A)+I(B)

P(C) = P(AB)=P(A)P(B)

There's only one function that satisfies all the above properties and that's the **Logarithmic function.**

$$I(A) = \log_b\left(\frac{1}{P_A}\right)$$

Where I(A) is **the amount of Self-Information/ Uncertainity**

Units of Information

- Depends on the logarithm base ‘b’.
- The unit of information with logarithmic base b=2 is called **Bi**nary Uni**t** (bit).
- The unit of information with logarithmic base b=3 is called **Triples.**
- For quarternary choice b = 4, the unit is **Quadruples.**
- For a decimal choice b =10, and unit is decits or **Hartleys.**
- For natural algorithm, I= ln(1/P), **Neipers** or **Nats** is used.

Conversion of Information Units

Unit	Bits (Base 2)	Nats (Base e)	Decits (Base 10)
Bits (Base 2)		1 bit $= \frac{1}{log_2 e}$ $= 0.6932\, nat$	1 bit $= \frac{1}{log_2 10}$ $= 0.3010$ decit
Nats (Base e)	1 nat $= \frac{1}{\ln 2}$ =1.4426 bits		1 nat $= \frac{1}{\ln 10}$ $= 0.4342$ decit
Decits (Base 10)	1 decit $= \frac{1}{log_{10} 2}$ $= 3.3219$ bits	1 decit $= \frac{1}{log_{10} e}$ $= 2.3026$ nats	

Problems 1.1

1) A source produces a binary symbol with a probability of 0.75. Determine the amount of information associated with the symbol in bits , nats and hartleys.

Equ.

I(A) = $\boldsymbol{log_2\left({}^{1}/_{P(A)}\right)}$ **bit**

P(A) = 0.75

∴ I(A) = $\log_2(1/0.75)$

For calculation, $I(A) = \dfrac{log_{10}\left({}^{1}/_{0.75}\right)}{log_{10}2} = 0.415\ bit$

I(A)=0.415*$\dfrac{1}{log_2\ e}$= 0.288 *nats*

I(A)=0.415*$\dfrac{1}{log_2 10}$= 0.125 *decits/hartleys*

Problem

2) A card is selected at random from a deck. If it is from the red suit, how much information have you gained? How much more information is needed to completely specify the card?

Solution:

26 red cards out of 52. $p=1/2$

$I= \log(1/1/2)= \log 2 =1$ bit

To completely specify the card $p'=1/52$

$I'= \log(1/1/52)=5.7$ bits

Additional information needed $I'-I= 4.7$ bits

Problems 1.5

- In a certain community 25% girls are blondes and 75% of all blondes have blue eyes. Also, 50% of all girls have blue eyes. If you are told that a girl has blue eyes how much information you have received? If you are further told that she is a blonde how much additional information you have received?

- Blondes (A) & Blue eye (B)

Problems 1.5

Solution:

- A ⟶ blonde, B ⟶ blue eye
- P(A) = 0.25 = ¼;   P(B) = 0.5=1/2;   I(B) = $\log_2$(1/P(B))=1 bit
- P(B|A)=0.75 =3/4
- P(AB) = P(A) P(B/A) = 3/16
- P(A|B) =P(AB)/P(B) = 3/8
- I(A|B) = log(8/3) = 1.415
- Additional information received = I(A/B) – I(B)= 0.415 bits

$$P(A|B) = \frac{P(A \quad B)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Entropy & its Derivation

- Consider a source emitting sequences of symbols from a fixed finite source alphabet, $S=\{s_1, s_2, \ldots. s_q\}$ → **Discrete source.**

Entropy

- The amount of information associated with each symbol is

$$I(s_k) = \log(1/p_k) , \quad k=1,2,\ldots,q \rightarrow \textbf{Self information.}$$

Entropy

We know that $n_1+n_2+\ldots+n_q=n$

And $p_k = \frac{n_k}{n}$ and replacing for $I(s_k)$,

$$\boxed{H(S) = \sum_{k=1}^{q} P_k \log\left(\frac{1}{P_k}\right) \text{ bits/symbol}}$$

H(S)→ Entropy (average information) of the source

Entropy

- $$H(S) = \sum_{k=1}^{q} P_k log_2(\frac{1}{P_k}) \text{ bits/symbol}$$

Example of Entropy

$H(S) = \sum_{k=1}^{q} p_k log_2(\frac{1}{p_k})$ bits/symbol

Consider the following:

- $S_1 = \{s_{11}, s_{12}\}$: $P_1 = \{1/256, 255/256\}$
- $S_2 = \{s_{21}, s_{22}\}$: $P_2 = \{1/2, 1/2\}$
- $S_3 = \{s_{31}, s_{32}\}$: $P_3 = \{7/16, 9/16\}$

Ans

- $H(S_1) = \sum_{k=1}^{2} p_k log_2(\frac{1}{p_k})$
- $H(S_1) = = (1/256).\log_2 256 + (255/256).\log_2(256/255)$
- $H(S_1) = 0.037$ bits/symbol

Example of Entropy

$$H(S) = \sum_{k=1}^{q} p_k log_2(\frac{1}{p_k}) \text{ bits/symbol}$$

Consider the following:

I. $S_1 = \{s_{11}, s_{12}\}$: $P_1 = \{1/256, 255/256\}$

II. $S_2 = \{s_{21}, s_{22}\}$: $P_2 = \{1/2, 1/2\}$

III. $S_3 = \{s_{31}, s_{32}\}$: $P_3 = \{7/16, 9/16\}$

Ans

- $H(S_2) = (½).\log_2 2 + (½).\log_2 2 = 1$ bits/symbol
- $H(S_3) = 0.0989$ bits/symbol

Information rate/ Entropy Rate

- If the source is emitting symbols at a fixed rate of ‘r_s’ symbols/sec, the average source information rate ‘R’ is defined as

$$\boxed{R = r_s \cdot H \text{ bits/sec}}$$

[(symbols/second)*(bits/symbol)].

- R → information rate, H → Entropy or average information, r_s →symbol rate
- Information rate (R) is represented in average number of bits of information per second.

Problems 1.7

Q) An analog signal is band limited to B Hz, sampled at the Nyquist rate, and the samples are quantized into 4-levels. The quantization levels Q_1, Q_2, Q_3, and Q_4 (messages) are assumed independent and occur with probs.

$P_1 = P_4 = 1/8$ and $P_2 = P_3 = 3/8$. Find the information rate of the source.

Solution:

$$\boldsymbol{H = \sum_{k=1}^{q} p_k log_2(\frac{1}{p_k})} \text{ bits/symbol}$$

$$H = p_1 \log_2(1/p_1) + p_2 \log(1/p_2) + p_3 \log_2(1/p_3) + p_4 \log(1/p_4)$$

Problems 1.7

Q) Band limited to B Hz, sampled at the Nyquist rate, $P_1 = P_4 = 1/8$ and $P_2 = P_3 = 3/8$. Find the information rate of the source.

- $H = p_1 \log_2(1/p_1) + p_2 \log(1/p_2) + p_3 \log_2(1/p_3) + p_4 \log(1/p_4)$
- $H = \frac{1}{8} log_2(8) + \frac{3}{8} log_2(8/3) + \frac{3}{8} log_2(8/3) + \frac{1}{8} log_2(8)$
- H = 1.8 bit/message
- Information rate, $R = r_s$ ***x*** H

R = (2B) x 1.8 = (3.6 B) bits/sec

Average Symbol Duration (τ) and Symbol Rate (r_s)

- Average symbol duration,

$$\tau = \sum_{k=1}^{q} p_k . \tau_k \text{ Secs/symbol}$$

- By $r_s = 1/\tau$ symbols/sec,

Problems 1.8

- Calculate the entropy rate of a conventional telegraph source with dash twice as long as a dot and half as probable. The dot lasts for 0.2ms and the same interval exists for the pause between the symbols.

Solution:

- $P_{dot} + P_{dash} = 1;$ $P_{dash} = 1/2\ P_{dot}\ ;$ $P_{dot} = 2/3;$ $P_{dash} = 1/3$
- $\tau_{dash} = 2\ \tau_{dot}$ $\tau_{dot} = \tau_{pause} = 0.2ms$ $\tau_{dash} = 0.4ms$

- $\tau = 2/3x0.2+1/3x0.4+0.2 = 0.467$ ms/symbol
- $r_s = 1/\ \tau = 2142.9$ symbols/sec
- $H = 2/3 \log 3/2 + 1/3 \log 3 = 0.92$bits/symbol
- $R = r_s H = 1971.43$ bits/sec

Homework

1Q) An analog signal is band limited to B Hz, sampled at the Nyquist rate, and the samples are quantized into 4-levels. The quantization levels Q_1, Q_2, Q_3, and Q_4 (messages) are assumed independent and occur with probs.

$P_1 = P_4 = 1/8$ and $P_2 = P_3 = 3/8$. Find the information rate of the source.

2Q) A certain source has eight symbols that are produced in blocks of 3 at a rate of 1000 blocks/sec. The first symbol in each block is always the same. The remaining 2 are filled by any of the eight symbols with equal probability. What is the entropy rate?

Problems 1.9

- A certain source has eight symbols that are produced in blocks of 3 at a rate of 1000 blocks/sec. The first symbol in each block is always the same. The remaining 2 are filled by any of the eight symbols with equal probability. What is the entropy rate?

Solution:

- All symbols are equiprobable, $H = H_{max} = \log 8 = 3$ bits/symbol
- Since the symbols are produced in blocks of 3, net entropy is

$$H_T = H_1 + H_2 + H_3$$

- $H_1 = 0$, first symbol always same.

Problems 1.9

- $H_2 = H_3 = 3$ bits/symbol second and third symbols can be any one of the 8 symbols with equal probability
- $H_T = H_1 + H_2 + H_3 = 6$ bits/block
- $R = r_s H_T = 1000 \times 6 = 6000$bits/sec

Properties of Entropy

1. The entropy function is continuous in each and every independent variable p_k in the interval (0,1).

Since p_k is continuous in the interval (0,1), logarithm of a continuous function is also continuous.

Properties of Entropy (contd.)

3.Extremal Property

Consider a zero memory information source with a q symbol alphabet S={$s_1, s_2, \ldots, s_q$} with probabilities P= ={$p_1, p_2, \ldots, p_q$}

Entropy of the source $H(S) = \sum_{k=1}^{q} p_k log(\frac{1}{P_k})$

Consider log q - H(S)

$$\log q - H(S) = \log q - \sum_{k=1}^{q} p_k log(\frac{1}{p_k})$$

$$= (\sum_{k=1}^{q} p_k). log\ q - \sum_{k=1}^{q} p_k log(\frac{1}{p_k})$$

$$= \sum_{k=1}^{q} p_k \left(log_2 q - log_2 \left(\frac{1}{p_k}\right)\right)$$

We know that $\quad log_2 q p_k = \frac{log_e q p_k}{log_e 2}$

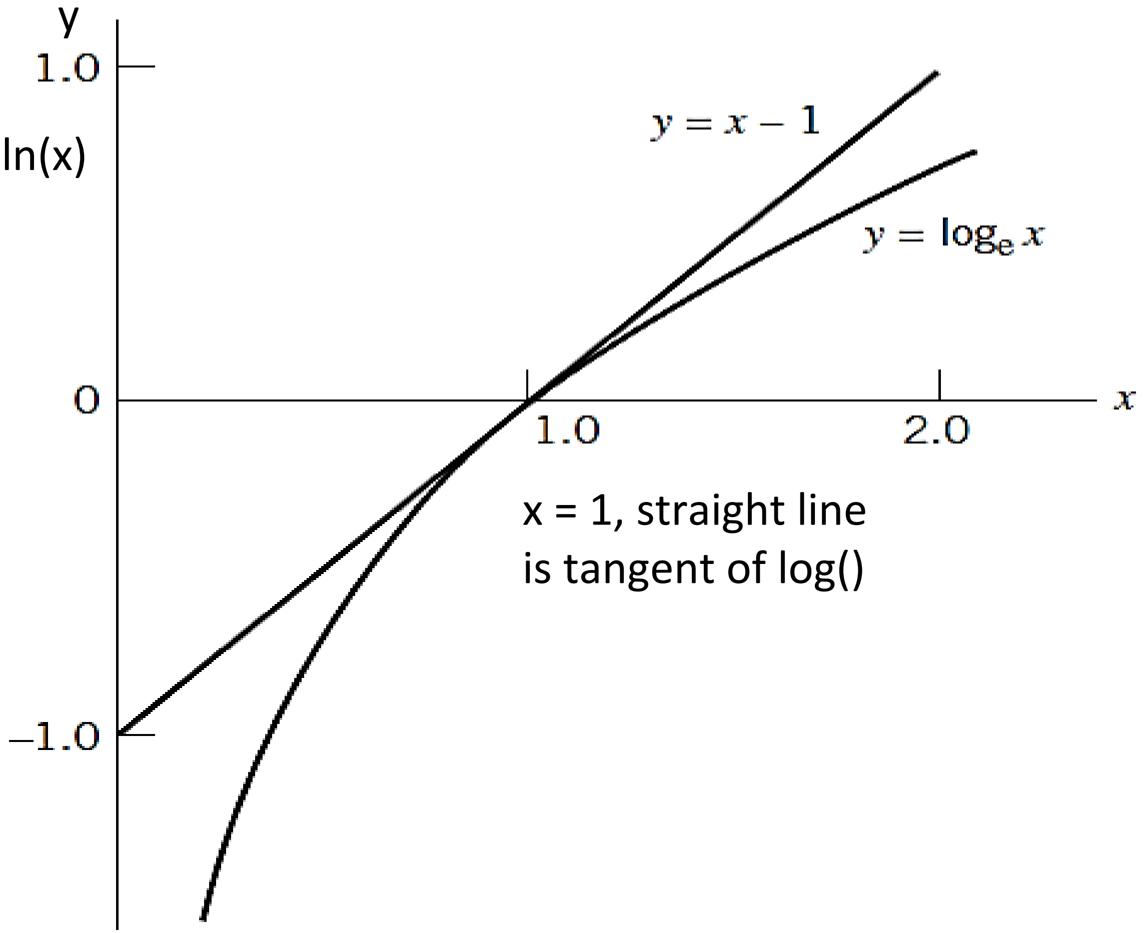
(This relation is used to convert the base from 2 to e so that the logarithmic inequality can be used in the next step)

Logarithmic Inequalities

- $\ln x \leq (x-1)$, equality iff x = 1
- Multiply by -1:

$$ln\left(\frac{1}{x}\right) \geq (1-x), \text{ equality iff } x = 1$$

- This property of the logarithmic function will be used in establishing the extremal property of the Entropy function (i.e. Maxima or minima property).



Graphs of the functions $x - 1$ and $\log x$ versus x.

Properties of Entropy

$* \; ln\left(\frac{1}{x}\right) \geq (1 - x)$, equality iff
x = 1

log q - H(S) $= \log_2 e \, . \sum_{k=1}^{q} p_k \; log_e(\, q \, p_k)$ ------------------------(1)

Applying *,

$$\log q - H(S) \geq \log e \sum_{k=1}^{q} p_k \left(1 - \frac{1}{q p_k}\right)$$ Equality iff qP_k=1

$$\geq \log e \sum_{k=1}^{q} p_k - \sum_{k=1}^{q} \frac{1}{q}$$ Equality iff P_k=1/q

$$\sum_{k=1}^{q} p_k = \sum_{k=1}^{q} \frac{1}{q} = 1$$

Hence ***H(S) ≤ log q***

For a zero memory source with a q symbol alphabet the entropy is maximum if and only if all the source symbols are equiprobable.

$$\mathbf{H(S)_{max} = \log q \quad iff \quad p_k = 1/q}$$

Maximum Entropy (Important)

- For a zero memory source with a ***q*** symbol alphabet the entropy is maximum iff all the source symbols are equiprobable.

$$\mathbf{H(S)_{max} = \log q \qquad iff \;\; p_k = 1/q \; ; \; k = 1, 2, \ldots, q}$$

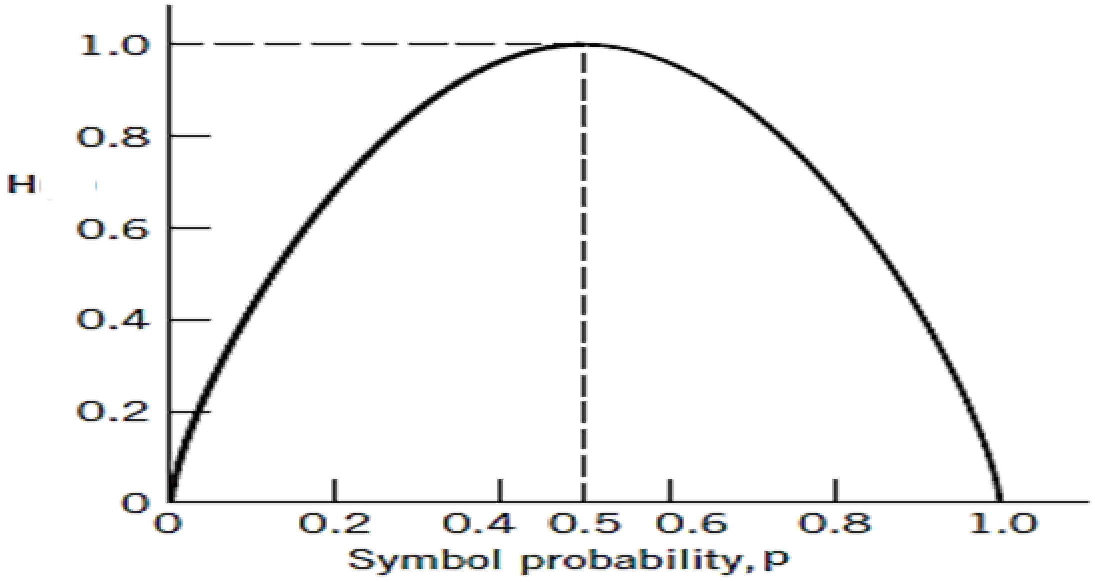
Properties of Entropy

- Zero Memory binary Source

S={0, 1}, P={q, p}, p(0) = q, p(1) = p and p + q = 1

$$H(S) = p\log_2\left(\frac{1}{p}\right) + q\log_2\left(\frac{1}{q}\right) = -p\log_2 p - (1-p)\log_2(1-p) \text{ bits}$$

a) When p = 0, the entropy H(S) = 0,
b) When p = 1, the entropy H(S) = 0,
c) The entropy H(S) attains its maximum value H_{max} = 1 bit when p = 1/2; i.e., when symbols 1 and 0 are equally probable.



Properties of Entropy – **Additive Property**

Suppose one of the symbol is divided into sub symbols :

- Then splitted entropy is

$$H'(S) = H(p_1, p_2, \ldots, p_{q\text{-}1}, p_{q1}, p_{q2}, \ldots, p_{qm})$$

- Entropy, $H(S) = \sum_{k=1}^{q} p_k log\left(\frac{1}{p_k}\right)$ bits/symbol

$$H'(S) = \sum_{k=1}^{q-1} p_k log\frac{1}{p_k} + \sum_{j=1}^{m} p_{qj} log\left(\frac{1}{p_{qj}}\right)$$

$$= \sum_{k=1}^{q} p_k log\frac{1}{p_k} - p_q log\frac{1}{p_q} + \sum_{j=1}^{m} p_{qj} log\left(\frac{1}{p_{qj}}\right)$$

Substitute $p_q = \sum_{j=1}^{m} p_{qj}$;

$$H'(S) = \sum_{k=1}^{q} p_k log\frac{1}{p_k} - \sum_{j=1}^{m} p_{qj}\, log\frac{1}{p_q} + \sum_{j=1}^{m} p_{qj} log\left(\frac{1}{p_{qj}}\right)$$

Properties of Entropy – Additive Property

- $H'(S) = \sum_{k=1}^{q} p_k log\frac{1}{p_k} - \sum_{j=1}^{m} p_{qj}\, log\frac{1}{p_q} + \sum_{j=1}^{m} p_{qj} log\left(\frac{1}{p_{qj}}\right)$

  $= \sum_{k=1}^{q} p_k log\frac{1}{p_k} + \sum_{j=1}^{m} p_{qj}[log\left(\frac{1}{p_{qj}}\right) - log\frac{1}{p_q}]$

- Second term multiply & divide by p_q also apply log rule,
- $H'(S) = H(p_1, p_2, \ldots, p_q) + p_q \sum_{j=1}^{m} \frac{P_{qj}}{P_q} [\, log\frac{p_q}{p_{qj}}]$
- $H'(S) = H(S) + K$
- $\therefore H'(S) = H(S) +$ a positive quantity
- Since $p_{qj} \leq p_q$ *for all j*

Properties of Entropy – Additive Property

- Since the entropy functions are non negative, we have

$$H(p_1, p_2, \ldots, p_{q-1}, p_{q1}, p_{q2}, \ldots., p_{qm}) \geq H(p_1, p_2, \ldots, p_{q-1}, p_q)$$

$$\therefore \mathbf{H'(S) \geq H(S)}$$

- Partitioning of the symbols into sub symbols cannot decrease the entropy.

Previous Problems 1.6

- Consider a discrete memoryless source with a source alphabet

  A = { s_o, s_1, s_2} with respective probs. p_0 = ¼, p_1 = ¼, p_2 = ½. Find the entropy of the source.

Solution:

$$\boldsymbol{H(S) = \sum_{k=1}^{q} p_k log_2(\frac{1}{p_k})} \text{ bits/symbol} \quad \boldsymbol{H(S) = \sum_{k=0}^{q-1} p_k log_2(\frac{1}{p_k})}$$

$$H(S) = \sum_{k=0}^{2} p_k log_2\left(\frac{1}{p_k}\right)$$

i.e., H(S) = $p_0 \log_2(1/p_0) + p_1 \log_2(1/p_1) + p_2 \log(1/p_2)$ = 1.5 bits/symbol

Understand

- Say $S = \{s_1, s_2, s_3\}$; $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$

  $H(S) = 1.5$ bits/symbol

- Say $S' = \{s_1, s_2, s_3, s_4\}$; $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$

  $H(S') = 1.75$ bits/symbol

- Say $S'' = \{s_1, s_2, s_3, s_4, s_5\}$; $P = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16})$

  $H(S'') = 1.875$ bits/symbol

- $H(S'') > H(S') > H(S)$; this going *to prove by additivity property*

Properties of Entropy – summary

1. **The entropy function is continuous in each and every independent variable p_k in the interval (0, 1).**
2. **The entropy function is a symmetrical function of its arguments**

   $$H(p_k, 1-p_k) = H(1-p_k, p_k) \qquad k=1, 2, \ldots, q$$

3. **Extremal Property (upper bound on entropy):**

   $$\log q - H(S) \geq 0; \text{ Hence } H(S) \leq \log_2 q$$

- For a zero memory source with a ***q*** symbol alphabet the entropy is maximum iff all the source symbols are equiprobable.

  $$H(S)_{max} = \log q \text{ bits/symbol}$$

4. **Additive Property**

   $$H'(S) = H(S) + \text{a positive quantity}$$

- **$H'(S) \geq H(S)$**
- Partitioning of the symbols into sub symbols cannot decrease the entropy.

Extension of a zero memory source (2nd order Extension)

- Consider a binary source emitting symbols S_1 and S_2 with probabilities P_1 and P_2 such that $P_1+P_2=1$.
- The second extension of this source will have 4 symbols given by
- $S_1 S_1$ with probability $P_1 P_1 = P_1^2$
- $S_1 S_2$ with probability $P_1 P_2$
- $S_2 S_1$ with probability $P_2 P_1$
- $S_2 S_2$ with probability $P_2 P_2 = P_2^2$

with sum of probabilities $P_1^2 + P_1P_2 + P_2P_1 + P_2^2 = (P_1 + P_2)^2 = 1$

Entropy of the second extension of the source is $\boxed{\mathbf{H(S^2)=2H(S)}}$

Extension of a zero memory source (3rd Extension)

The third extension of the source will have 8 symbols given by

$S_1\,S_1\,S_1$ with probability P_1^3 ; $S_1\,S_1\,S_2$ with probability $P_1^2\,P_2$

$S_1\,S_2\,S_1$ with probability $P_2\,P_1^2$; $S_1\,S_2\,S_2$ with probability $P_2^2\,P_1$

$S_2\,S_1\,S_1$ with probability $P_2\,P_1^2$; $S_2\,S_1\,S_2$ with probability $P_2^2\,P_1$

$S_2\,S_2\,S_1$ with probability $P_2^2\,P_1$; $S_2\,S_2\,S_2$ with probability P_2^3

with sum of probabilities $(P_1 + P_2)^3 = 1$

Entropy of the third extension of the source is $\boxed{H(S^3)=3H(S)}$

In general, Entropy of the nth extension of the source is $\boxed{\mathbf{H(S^n)=n.H(S)}}$

Problem 2.0

- A zero memory source has a source alphabet S = $\{s_1, s_2, s_3\}$, with P = {1/2, 1/4, 1/4}. Find the entropy of the source. Find the entropy of the second extension and verify.

Solution:

- The entropy of the source, $\boldsymbol{H(S) = \sum_{k=1}^{q} p_k log_2(\frac{1}{p_k})}$ bits/symbol
- H(S) = (1/2) x log(2) + 2 x (1/4) x log(4)

  =1.5 bits/symbol

Problem 2.0

Q) $S = \{s_1, s_2, s_3\}$, with $P = \{1/2, 1/4, 1/4\}$.

Solution:

- Second extension and the corresponding probabilities are tabulated as:
- $S^2 = \{s_1s_1, s_1s_2, s_1s_3, s_2s_1, s_2s_2, s_2s_3, s_3s_1, s_3s_2, s_3s_3\}$
- $P(S^2) = \{ 1/4, 1/8, 1/8, 1/8, 1/16, 1/16, 1/8, 1/16, 1/16 \}$
- Then, $H(S^2) = \{(1/4) \times \log 4\} + \{4\times (1/8) \log 8\} + \{4 \times (1/16) \times \log 16\}$

  $= 3.0$ bits / sym.

$\therefore \{H(S^2)\}/ \{H(S)\} = 3 / 1.5 = 2$; and indeed $H(S^2) = 2 \times H(S)$

Conditional probability

- **Conditional probability** is a measure of the *probability* of an *event occurring* given that another event has (by assumption, presumption, assertion or evidence) *occurred*.
- Given an event *B* with non-zero Probability, the Conditional probability of an event *A* assuming *B* has occurred is:

$$P(A|B) = \frac{P(AB)}{P(B)}\ ; P(B) > 0$$

- Read $P(A|B)$ as "*P of A assuming B*" or "*P of A given B*"

Total Probability & Bayes' Theorem

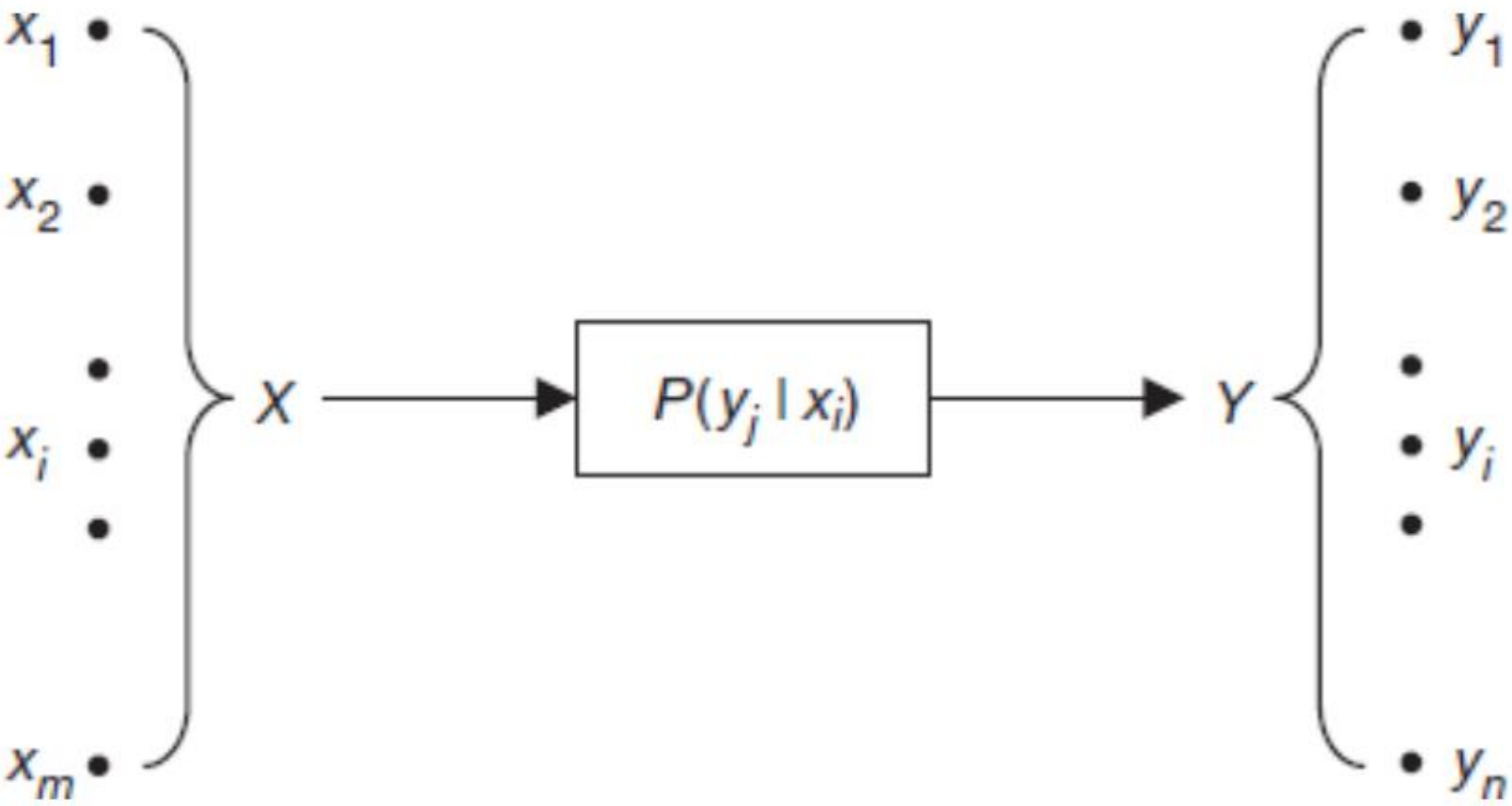
- The law of total probability is a fundamental rule relating marginal probabilities to conditional probabilities. It expresses the total probability of an outcome which can be realized via several distinct events.

$$\mathrm{P(B)} = \sum_{k=1}^{n} P(A_k).P(B|A_k)$$

- The Bayes theorem describes the probability of an event based on the prior knowledge of the conditions that might be related to the event.

$$P(A_j|B) = \frac{P(A_j).P(B|A_k)}{\sum_{k=1}^{n} P(A_k).P(B|A_k)}$$

Discrete Memoryless Channels



Discrete Memoryless Channels

- A statistical model with an input X and an output Y.
- During each unit of the time (signaling interval), the channel accepts an input symbol from X, and in response it generates an output symbol from Y.
- The channel is "discrete" when the alphabets of X and Y are both finite.
- It is "memoryless" when the current output depends on only the current input and not on any of the previous inputs

Joint Probability Matrix [JPM]

Fig: Simple Communication System

- This simple system may be uniquely characterized by the **'Joint probability matrix'** (JPM), P(X, Y) of the *probabilities* existent between the input and output ports.
- Consider a discrete memoryless channel:
  - Statistical model with an input X and an output Y that is a noisy version of X; both X and Y are random variables

Joint Probability Matrix [JPM]

$$P(X,Y)=\begin{bmatrix} p(x_1,y_1) & p(x_1,y_2) & P(x_1,y_3) & \cdots\cdots & P(x_1,y_n) \\ P(x_2,y_1) & P(x_2,y_2) & P(x_2,y_3) & \cdots\cdots & P(x_2,y_n) \\ P(x_3,y_1) & P(x_3,y_2) & P(x_3,y_3) & \cdots\cdots & P(x_3,y_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P(x_m,y_1) & P(x_m,y_2) & P(x_m,y_3) & \cdots\cdots & P(x_m,y_n) \end{bmatrix}$$

- It is said to be "**memoryless**" when the current output symbol depends *only* on the current input symbol and *not* any previous or future symbol.

Joint Probability Matrix [JPM]

- $\sum_{\forall k} \sum_{\forall j} P(x_k, y_j) = 1$ sum of all entries of JPM
- $\sum_j P(x_k, y_j) = P(x_k)$ sum of all entries of JPM in the k^{th} row
- $\sum_k P(x_k, y_j) = P(y_j)$ sum of all entries of JPM in the j^{th} column

And also

- $\sum_k P(x_k) = \sum_j P(y_j) = 1$
- Simple communication network there are five probability schemes of interest viz:
  - P(X), P(Y), P(X, Y), P(X|Y) and P(Y|X).

Entropy functions

- Accordingly there are five entropy functions:

1) **H(X):** Average information per character or symbol transmitted by the *source or the entropy of the source*.
2) **H(Y):** Average information received per character at the *receiver or the entropy of the receiver*.
3) **H(X,Y):** Average information per pair of transmitted and received characters or the average uncertainty of the communication system as a whole.

Entropy functions

- Accordingly there are five entropy functions:

4) **H(X|Y):** A specific character y_j being received.

   The average value of the Entropy associated with this scheme when y_j covers all the received symbols i.e., $E\{H(X|y_j)\}$ is the entropy $H(X|Y)$, called the 'Equivocation'

5) **H(Y|X):** Similar to H (X|Y), this is a measure of information about the receiver. Also called as noise entropy.

Joint and Conditional Entropies

$$P(X,Y)=\begin{bmatrix} p(x_1,y_1) & p(x_1,y_2) & P(x_1,y_3) & \cdots & P(x_1,y_n) \\ P(x_2,y_1) & P(x_2,y_2) & P(x_2,y_3) & \cdots & P(x_2,y_n) \\ P(x_3,y_1) & P(x_3,y_2) & P(x_3,y_3) & \cdots & P(x_3,y_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P(x_m,y_1) & P(x_m,y_2) & P(x_m,y_3) & \cdots & P(x_m,y_n) \end{bmatrix}$$

- We define the following entropies, which can be directly computed from the **JPM**.

$$H(X,Y)=p(x_1,y_1)\log\frac{1}{p(x_1,y_1)}+p(x_1,y_2)\log\frac{1}{p(x_1,y_2)}+\ldots+p(x_1,y_n)\log\frac{1}{p(x_1,y_n)}$$

$$+p(x_2,y_1)\log\frac{1}{p(x_2,y_1)}+p(x_2,y_2)\log\frac{1}{p(x_2,y_2)}+\ldots+p(x_2,y_n)\log\frac{1}{p(x_1,y_1)}$$

$$+\ldots p(x_m,y_1)\log\frac{1}{p(x_m,y_1)}+p(x_m,y_2)\log\frac{1}{p(x_m,y_2)}+\ldots p(x_m,y_n)\log\frac{1}{p(x_m,y_n)}$$

Joint and Conditional Entropies

- or $H(X,Y) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k, y_j)}\right)$
- $H(X) = \sum_{k=1}^{m} p(x_k) log_2 \left(\frac{1}{p(x_k)}\right)$
- From equ. Σ_j P(x_k ,y_j) = P(x_k) sum of all entries of JPM in the k^{th} row; only for the multiplication term, this equation can be re-written as:
- $H(X) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k)}\right)$
- III^{ly} $H(Y) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(y_j)}\right)$

Joint and Conditional Entropies

- Next, from the definition of the conditional probability we have:
- $P\{X = x_k|Y = y_j\} = \frac{P\{X=x_k, Y=y_j\}}{P\{Y=y_j\}}$
- $p\{x_k|y_j\} = {p(x_k, y_j)}/{p(y_j)}$
- Then, $\sum_{k=1}^{m} p(x_k|y_j) = \frac{1}{p(y_j)} \sum_{k=1}^{m} p(x_k, y_j) = \frac{1}{p(y_j)} . p(y_j)$=1
- $$\boxed{H(X|y_j) = \sum_{k=1}^{m} p(x_k|y_j) log_2 \left(\frac{1}{p(x_k|y_j)}\right)}$$

Joint and Conditional Entropies

- Taking the average of the above entropy function for all admissible characters received, we have the average "***Conditional Entropy***" or "***Equivocation***":
- $H(X|Y) = E\{H(X|y_j)\}_j$

$$= \sum_{j=1}^{n} p(y_j) H(X|y_j)$$

$$= \sum_{j=1}^{n} p(y_j) \sum_{k=1}^{m} p(x_k|y_j) log_2\left(\frac{1}{p(x_k|y_j)}\right)$$

Or $\boxed{H(X|Y) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2\left(\frac{1}{p(x_k|y_j)}\right)}$

Joint and Conditional Entropies

- Similarly the conditional entropy H(Y|X) by:

$$H(Y|X) = \sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j) log_2\left(\frac{1}{p(y_j|x_k)}\right)$$

- *'The entropy you want is simply the double summation of joint probability multiplied by logarithm of the reciprocal of the probability of interest'*.
- All the five entropies so defined are all inter-related.

Joint and Conditional Entropies

We have:

- $H(Y|X) = \sum_k \sum_j p(x_k, y_j) log_2 \left(\frac{1}{p(y_j|x_k)}\right)$

Since, $\frac{1}{p(y_j|x_k)} = \frac{p(x_k)}{p(x_k, y_j)}$

- We can straight away write:

$$H(Y|X) = \sum_k \sum_j p(x_k, y_j) log_2 \left(\frac{1}{p(x_k, y_j)}\right) - \sum_k \sum_j p(x_k, y_j) log_2 \left(\frac{1}{p(x_k)}\right)$$

- Or $H(Y|X) = H(X, Y) - H(X)$
- i.e., $H(X, Y) = H(X) + H(Y|X)$
- Also, $H(X, Y) = H(Y) + H(X|Y)$

Joint and Conditional Entropies

- Consider $H(X) - H(X|Y)$.
- $H(X) - H(X|Y) = \sum_k \sum_j p(x_k, y_j) log_2 \left\{\frac{1}{p(x_k)} - \frac{1}{p(x_k|y_j)}\right\}$

$$= \sum_k \sum_j p(x_k, y_j) log_2 \left\{\frac{p(x_k, y_j)}{p(x_k).p(y_j)}\right\}$$

- Using the logarithm inequality derived earlier, you can write the above equation as:
- $H(X) - H(X|Y) = \text{loge} \sum_k \sum_j p(x_k, y_j) ln \left\{\frac{p(x_k, y_j)}{p(x_k).p(y_j)}\right\}$

Joint and Conditional Entropies

$$\geq loge \sum_k \sum_j p(x_k, y_j) \left\{1 - \frac{p(x_k).p(y_j)}{p(x_k, y_j)}\right\}$$

$$\geq loge \left\{\sum_k \sum_j p(x_k, y_j) - \sum_k \sum_j p(x_k).p(y_j)\right\}$$

$$\geq loge \left\{\sum_k \sum_j p(x_k, y_j) - \sum_k p(x_k).\sum_j p(y_j)\right\} \geq 0$$

ln x ≤ (x-1), equality iff x = 1
Multiply by -1;
$ln\left(\frac{1}{x}\right) \geq (1 - x)$, equality iff x = 1

- Since $\sum_k \sum_j p(x_k, y_j) = \sum_k p(x_k) = \sum_j p(y_j) = 1$
- Thus, $H(X) \geq H(X|Y)$ & $H(Y) \geq H(Y|X)$

Equality will hold iff $P(x_k,y_j) = P(x_k)\ P(y_j)$
ie. Input symbols and output symbols are statistically independent of each other.

Problem 2.1

- Determine different entropies for the JPM given below and verify their relationships.

$$P(X, Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

Solution:

- $\Sigma_j \, P(x_k, y_j) = P(x_k)$
- $\Sigma_k P(x_k, y_j) = P(y_j)$

Problem 2.1

$$P(X, Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

- $\Sigma_j P(x_k, y_j) = P(x_k)$
- $P(X) = [P(x_1), P(x_2), P(x_3), P(x_4), P(x_5)]$

  $= [0.4, 0.13, 0.04, 0.15, 0.28]$
- $\Sigma_k P(x_k, y_j) = P(y_j)$
- $P(Y) = [P(y_1), P(y_2), P(y_3), P(y_4)]$

  $= [0.34, 0.13, 0.26, 0.27]$

Problem 2.1

$$P(X, Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

- $H(X) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k)}\right)$
- $H(Y) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(y_j)}\right)$
- i.e. $H(X) = \sum_{k=1}^{5} \sum_{j=1}^{4} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k)}\right)$
- $H(X) =$ 0.4 log(1/0.4) + 0.13 log(1/0.13) + 0.04 log(1/0.04) + 0.15 log(1/0.15) + 0.28 log(1/0.28)

P(X) = [0.4, 0.13, 0.04, 0.15, 0.28]

$H(X) = 2.021934821\ bits/sym$

- $H(Y) =$ 0.34 log(1/0.34) + 0.13 log(1/0.13) + 0.26 log(1/0.26) + 0.27 log(1/0.27) = 1.927127708 bits/sym

P(Y) = [0.34, 0.13, 0.26, 0.27]

Problem 2.1

$$P(X, Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

- $H(X,Y) = \sum_{k=1}^{m} \sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k, y_j)} \right)$
- i.e. $H(X,Y) = \sum_{k=1}^{5} \sum_{j=1}^{4} p(x_k, y_j) log_2 \left(\frac{1}{p(x_k, y_j)} \right)$

H(X,Y) = 3 x 0.2 x log(1/0.2) + 0.1 x log(1/0.1) + 4 x 0.01 x log(1/0.01) + 3 x 0.02 x log(1/0.02) + 2 x 0.04 x log(1/0.04) + 2 x 0.06 x log(1/0.06) = 3.188311023 bits/symbol

Problem 2.1

$$P(X,Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

Since, $\dfrac{1}{p(y_j \mid x_k)} = \dfrac{p(x_k)}{p(x_k, y_j)}$

- Divide the entries in the j^{th} column of the **JPM** of $p(y_j)$

P(Y) = [0.34, 0.13, 0.26, 0.27]

$$P(X|Y) = \begin{bmatrix} \frac{0.2}{0.34} & 0 & \frac{0.2}{0.26} & 0 \\ \frac{0.1}{0.34} & \frac{0.01}{0.13} & \frac{0.01}{0.26} & \frac{0.01}{0.27} \\ 0 & \frac{0.02}{0.13} & \frac{0.02}{0.26} & 0 \\ \frac{0.04}{0.34} & \frac{0.04}{0.13} & \frac{0.01}{0.26} & \frac{0.06}{0.27} \\ 0 & \frac{0.06}{0.13} & \frac{0.02}{0.26} & \frac{0.20}{0.27} \end{bmatrix}$$

Problem 2.1

- $H(X|Y) = \sum_{j=1}^{n}\sum_{k=1}^{m} p(x_k, y_j) log_2\left(\frac{1}{p(x_k|y_j)}\right)$

$$P(X|Y) = \begin{bmatrix} \frac{0.2}{0.34} & 0 & \frac{0.2}{0.26} & 0 \\ \frac{0.1}{0.34} & \frac{0.01}{0.13} & \frac{0.01}{0.26} & \frac{0.01}{0.27} \\ 0 & \frac{0.02}{0.13} & \frac{0.02}{0.26} & 0 \\ \frac{0.04}{0.34} & \frac{0.04}{0.13} & \frac{0.01}{0.26} & \frac{0.06}{0.27} \\ 0 & \frac{0.06}{0.13} & \frac{0.02}{0.26} & \frac{0.20}{0.27} \end{bmatrix}$$

$$\begin{aligned} \therefore H(X|Y) = {} & 0.2\log\frac{0.34}{0.2} + 0.2\log\frac{0.26}{0.2} + 0.1\log\frac{0.34}{0.1} \\ & + 0.01\log\frac{0.13}{0.01} + 0.01\log\frac{0.26}{0.01} + 0.01\log\frac{0.27}{0.01} \\ & + 0.02\log\frac{0.13}{0.02} + 0.02\log\frac{0.26}{0.02} + 0.04\log\frac{0.34}{0.04} \\ & + 0.04\log\frac{0.13}{0.04} + 0.01\log\frac{0.26}{0.01} + 0.06\log\frac{0.27}{0.06} \\ & + 0.06\log\frac{0.13}{0.06} + 0.02\log\frac{0.26}{0.02} + 0.2\log\frac{0.27}{0.2} \end{aligned}$$

$H(X|Y) = 1.261183315$ bits/sym

Problem 2.1

$$P(X, Y) = \begin{bmatrix} 0.2 & 0 & 0.2 & 0 \\ 0.1 & 0.01 & 0.01 & 0.01 \\ 0 & 0.02 & 0.02 & 0 \\ 0.04 & 0.04 & 0.01 & 0.06 \\ 0 & 0.06 & 0.02 & 0.2 \end{bmatrix}$$

Similarly divide the entries in the k^{th} row of **JPM** by $P(x_k)$ to obtain $P(Y/X)$

$$P(X) = [0.4, 0.13, 0.04, 0.15, 0.28]$$

$$P(Y|X) = \begin{bmatrix} \frac{0.2}{0.4} & 0 & \frac{0.2}{0.4} & 0 \\ \frac{0.1}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} \\ 0 & \frac{0.02}{0.04} & \frac{0.02}{0.04} & 0 \\ \frac{0.04}{0.15} & \frac{0.04}{0.15} & \frac{0.01}{0.15} & \frac{0.06}{0.15} \\ 0 & \frac{0.06}{0.28} & \frac{0.02}{0.28} & \frac{0.20}{0.28} \end{bmatrix}$$

Problem 2.1

- $H(Y|X) = \sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j) log_2 \left(\frac{1}{p(y_j|x_k)}\right)$

$$P(Y|X) = \begin{bmatrix} \frac{0.2}{0.4} & 0 & \frac{0.2}{0.4} & 0 \\ \frac{0.1}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} & \frac{0.01}{0.13} \\ 0 & \frac{0.02}{0.04} & \frac{0.02}{0.04} & 0 \\ \frac{0.04}{0.15} & \frac{0.04}{0.15} & \frac{0.01}{0.15} & \frac{0.06}{0.15} \\ 0 & \frac{0.06}{0.28} & \frac{0.02}{0.28} & \frac{0.20}{0.28} \end{bmatrix}$$

$$H(Y|X) = 2\times 0.2 log\frac{0.4}{0.2} + 0.1 log\frac{0.13}{0.1} + 3\times 0.01 log\frac{0.13}{0.01} + 2\times 0.02 log\frac{0.04}{0.02}$$
$$+ 2\times 0.04 log\frac{0.05}{0.04} + 0.01 log\frac{0.15}{0.01} + 0.06 log\frac{0.15}{0.06} + 0.06 log\frac{0.28}{0.06}$$
$$+ 2\times 0.02 log\frac{0.28}{0.02} = 1.166376202 \text{ bits / sym.}$$

$H(Y|X) = 1.166376202$ bits/sym

Problem 2.1

- Thus by actual computation we have:

$$H(X,Y) = 3.188311023 \text{ bits/symbol}$$

$$H(X) = 2.021934821 \; bits/sym$$

$$H(Y) = 1.927127708 \text{ bits/sym}$$

$$H(X|Y) = 1.261183315 \text{ bits/sym}$$

$$H(Y|X) = 1.166376202 \text{ bits/sym}$$

- Clearly, ***H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)***

  H(X) > H(X|Y) and H(Y) > H(Y|X)

6) If X & Y are discrete random source and P(X, Y) is their joint probability distribution and is given as

$$P(X, Y) = \begin{bmatrix} 0.08 & 0.05 & 0.02 & 0.05 \\ 0.15 & 0.07 & 0.01 & 0.12 \\ 0.10 & 0.06 & 0.05 & 0.04 \\ 0.01 & 0.12 & 0.01 & 0.06 \end{bmatrix}$$

Calculate H(X), H(Y), H(X|Y), H(Y|X), H(X, Y) & I(X, Y).

Also verify the formula H(X, Y) = H(X) + H(Y|X).

Bayes' Rule for RV

- From probability theory, Bayes' Rule
- $P(x_k, y_j) = P(x_k|y_j).P(y_j)$ --- (m1)
- $P(x_k, y_j) = P(y_j|x_k).P(x_k)$ --- (m2)

From equ. M1 & M2:

- $P(x_k|y_j).P(y_j) = P(y_j|x_k).P(x_k)$
- $\frac{P(x_k|y_j)}{P(x_k)} = \frac{P(y_j|x_k)}{P(y_j)}$ --- (m3)

Mutual Information

- Mutual information is a general *measure of the dependence between two random variables*.
- ***Measure of the amount of information that one RV contains about another RV***.
  - Entropy $H(X)$ is the *uncertainty* (“self-information”) of a single random variable.
  - Conditional entropy $H(X|Y)$ is the entropy of one random variable conditional upon knowledge of another.
  - The average amount of decrease of the randomness of X by observing Y is the average information that Y gives us about X.

Mutual Information

- The formal definition of the mutual information of two random variables ***X*** and ***Y***, whose joint distribution is defined by *P(X, Y)* is given by:

$$\boxed{I(X;Y) = \sum_k \sum_j p(x_k, y_j)\, log\left(\frac{p(x_k, y_j)}{p(x_k)p(y_j)}\right)}$$

- It is important in communication where it can be used to maximize the amount of information shared between sent and received signals.
- Relationship between entropy and mutual information:

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

Mutual Information

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

- $H(X) \rightarrow$ represents the uncertainty about the channel input before the channel output is observed,
- $H(X|Y) \rightarrow$ represents the uncertainty about the channel input after the channel output is observed,
- Mutual information $I(X;Y)$ represents the uncertainty about the channel input that is resolved by observing the channel output.
- The Entropy corresponding to mutual information [i.e. $\boldsymbol{I(X,Y)}$] indicates a measure of the information transmitted through a channel. Hence, it is called '*Transferred Information*'.

Mutual Information

- ***Mutual Information*** or ***Transinformation*** of the channel denoted by

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

- *The mutual information I(X;Y) is a measure of the uncertainty about the channel input, which is resolved by observing the channel output.*
- We may define it in another way, as:

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

- *The mutual information I(Y;X) is a measure of the uncertainty about the channel output that is resolved by sending the channel input.*

Properties of Mutual Information

$I(X;Y) = I(Y;X)$; **Symmetry**

$I(X;Y) \geq 0$; **Nonnegativity**

Relation to conditional and joint entropy

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

- The mutual information is always greater than or equal to zero, with equality iff X and Y are independent.
- It is lower than the entropy of either variable, and equality only occurs iff one variable is a deterministic function of the other.
- The higher the mutual information, the stronger the dependency between X and Y:

$$0 \leq I(X;Y) \leq \min(H[X], H[Y])$$

Properties of Mutual Information

- The difference between the initial uncertainty of the source symbol x_k, $\log \frac{1}{p(xk)}$ and
- The final uncertainty of the same source after receiving y_j, $\log \frac{1}{p(xk/yj)}$ is the information gained through the channel.
- This difference is known as the *mutual information between the symbols x_k and y_j*.

Properties of Mutual Information

- $I(x_k, y_j) = \log\left(\frac{1}{p(x_k)}\right) - \log\left(\frac{1}{p(x_k|y_j)}\right)$

$$= \log\left(\frac{p(x_k|y_j)}{p(x_k)}\right) \quad = \log\left(\frac{p(x_k|y_j)}{p(x_k)}\right) \quad \text{--- (1)}$$

Averaging over all admissible characters x_k and y_j, we obtain the average information gain of the receiver.

$$I(X,Y) = E\{I(x_k, y_j)\} = \sum_k \sum_j I(x_k\,, y_j) p(x_k, y_j) \qquad \text{--- (2)}$$

Substituting (1) in (2)

$$I(X,Y) = E\{I(x_k, y_j)\} = \sum_k \sum_j p(x_k, y_j) \log \frac{p(x_k, y_j)}{p(x_k)p(y_j)}$$

Properties of Mutual Information

- Now, $I(X,Y) = E\{I(x_k,y_j)\} = \sum_k \sum_j p(x_k,y_j) \log \frac{p(x_k,y_j)}{p(x_k)p(y_j)}$
- This equation can be written is different forms
- $I(X,Y) = E\{I(x_k,y_j)\} = \sum_k \sum_j p(x_k,y_j) \log \frac{p(x_k/y_j)p(y_j)}{p(x_k)p(y_j)}$
- $I(X,Y) = \sum_k \sum_j p(x_k,y_j) \left\{\log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k/y_j)}\right\}$ --- (3)

Properties of Mutual Information

- Another form
- $I(X,Y) = E\{I(x_k, y_j)\} = \sum_k \sum_j p(x_k, y_j) \log \frac{p(y_j/x_k)p(x_k)}{p(x_k)p(y_j)}$
- $I(X,Y) = \sum_k \sum_j p(x_k, y_j) \left\{\log \frac{1}{p(y_j)} - \log \frac{1}{p(y_j/x_k)}\right\}$ --- (4)
- Rearranging this equation,
- $I(X,Y) = \sum_k \sum_j p(x_k, y_j) \left\{\log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k/y_j)}\right\}$

$$= H(X) - H(X/Y)$$

Properties of Mutual Information

- Also
- $I(X,Y) = \sum_k \sum_j p(x_k, y_j) \left\{\log \frac{1}{p(y_j)} - \log \frac{1}{p(y_j/x_k)}\right\}$
  $= H(Y) - H(Y/X)$
- Also
- $I(X,Y) = \sum_k \sum_j p(x_k, y_j) \left\{\log \frac{1}{p(x_k)} + \log \frac{1}{p(y_j)} - \log \frac{1}{p(x_k, y_j)}\right\}$
- $I(X,Y) =$
  $\sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(x_k)} + \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(y_j)} - \sum_k \sum_j p(x_k, y_j) \log \frac{1}{p(x_k, y_j)}$
  $= H(X) + H(Y) - H(X,Y)$

Properties of Mutual Information

- Mutual information is symmetrical w.r.t. its arguments i.e., $I(x_k,y_j) = I(y_j,x_k)$.
- $I(x_k, y_j) = \log\frac{p(x_k,y_j)}{p(x_k)p(y_j)} = \log\frac{p(y_j/x_k)p(x_k)}{p(x_k)p(y_j)} = \log\frac{p(y_j/x_k)}{p(y_j)}$

$$= \log\frac{1}{p(y_j)} - \log\frac{1}{p(y_j/x_k)} = I(y_j) - I(y_j/x_k)$$

$$I(x_k, y_j) = I(y_j, x_k)$$

$$\frac{p(x_k,yj)}{p(xk)} = p(y_j/x_k)$$

Mutual Information

- This difference is known as the *mutual information between the symbols* x_k *and* y_j.

$$I(x_k, y_j) = \log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k|y_j)}$$

$$= \log \frac{p(x_k|y_j)}{p(x_k)} \qquad \text{---(m1)}$$

Or $I(x_k, y_j) = \log \dfrac{p(x_k, y_j)}{p(x_k)\, p(y_j)}$ ---(m2)

Also $I(x_k) = I(x_k, x_k) = log \dfrac{p(x_k|x_k)}{p(x_k)} = log \left(\dfrac{1}{p(x_k)}\right)$

Self Information

- From equ (m2), $\frac{p(x_k,y_j)}{p(x_k)} = P(y_j|x_k)$,
- $I(x_k, y_j) = \log \frac{P(y_j|x_k)}{p(y_j)} = log\left(\frac{1}{p(y_j)}\right) - log\left(\frac{1}{P(y_j|x_k)}\right)$
- Or $I(x_k, y_j) = I(y_j) - I(y_j|x_k)$ ---(m3)
- Simply means that “the Mutual information’ is symmetrical with respect to its arguments. i.e.
- $I(x_k, y_j) = I(y_j, x_k)$

- Averaging equ. (m2) over all admissible characters x_k and y_j, we obtain the average information gain of the receiver:
- $I(X, Y) = E\{I(x_k, y_j)\}$

$$= \sum_k \sum_j I(x_k, y_j) . p(x_k, y_j)$$

$$= \sum_k \sum_j p(x_k, y_j) . \log \frac{p(x_k, y_j)}{p(x_k)\, p(y_j)} \qquad \text{---(m4)}$$

From equ. (m4), we have:

$$I(x_k, y_j) = \sum_k \sum_j p(x_k, y_j) . \left\{ \log \frac{1}{p(x_k)} - \log \frac{1}{p(x_k|y_j)} \right\} = H(X) - H(X|Y)$$

$$I(X, Y) = \sum_k \sum_j p(x_k, y_j) \ [\log \frac{1}{p(y_j)} . - \log \frac{1}{p(y_j \mid x_k)}$$

$$= H(Y) - H(Y \mid X) \qquad \ldots\ldots\ldots\ldots$$

$$I(X,Y) = \sum_k \sum_j p(x_k, y_j) \ \log \frac{1}{p(x_k)} + .\sum_k \sum_j p(x_k, y_j) \ \log \frac{1}{p(y_j)} -$$

$$\sum_K \sum_J p(x_k, y_j) \ \log \frac{1}{p(x_k \ y_j)}$$

Or $I(X, Y) = H(X) + H(Y) - H(X, Y)$

Mutual Information

- H(X) bits are needed to represent one input symbol on an average.
- If we observe the respective output symbol, we need only $H(X|Y)$ bits to represent the input symbol.
- Observation of the output symbols provides us with $H(X)$ - $H(X|Y)$ bits of information, difference is called *Mutual Information* or *Trans-information* of the channel denoted by

  $I(X,Y)$. i.e. $I(X,Y) = H(X)-H(X|Y)$
- *The mutual information $I(X;Y)$ is a measure of the uncertainty about the channel input, which is resolved by observing the channel output.*

Mutual Information

- We may define it in another way, as:

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

- *The mutual information I(Y;X) is a measure of the uncertainty about the channel output that is resolved by sending the channel input.*
- For a symbol x_k, if $p(x_k)$ is known then $p(x_k/y_j)$ can be computed at the receiver end.
- The difference between the initial uncertainty of the source symbol x_k, $\log \frac{1}{p(xk)}$ and the final uncertainty of the same source after receiving y_j, $\log \frac{1}{p(xk/yj)}$ is the information gained through the channel.

Properties of Mutual Information

1) $I(X;Y) = I(Y;X)$
2) $I(X;Y) \geq 0$
3) $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
4) $I(X;Y) = H(X) + H(Y) - H(X,Y)$

- The Entropy corresponding to mutual information [i.e. I(X,Y)] indicates a measure of the information transmitted through a channel. Hence, it is called '*Transferred Information*'.