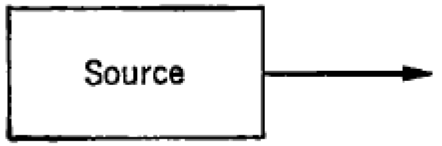
# Chapter 5: Information source

# 5.1. What is information source?

- Information is abstract. To talk on information, information theory presents each information by a source symbol.
- Set of source symbol (source alphabet) normally finite S = {s1, s2, …, sq}
- Source: emitting a sequence of source symbols (message) from alphabet m = {si1, si2, …} while sij is a symbol si ϵ S, j is time for creating symbol si
- Each symbol: selected according to some fixed probability law
- Refer to the source itself as S




si1, si2, …

- At any given time, emitted symbol is mapped to a value of a random variable (e.g. X)
  - Probability of random variable value = probability of symbol
- Source is a random variable

# 5.2. Type of information sources

- Discrete source
  - Output has an alphabet of distinct letters (source symbols)
  - The size of the alphabet is usually finite
  - Different types of discrete sources:
    - *Discrete memoryless source* (DMS): successive symbols emitted from the source are statistically independent.
      - Its output at a certain time does not depend on its output at any earlier time.
      - Random variable that represents DMS propertied by
        - X = {x1, x2…xn}
        - P(X) = {P(x1), P(x2),…P(Xn)}
    - *Discrete source with memory* (DSM) has the property that its output at a certain time may depend on its outputs at a number of earlier times
      - DSMs are usually modeled by means of Markov chains; they are then called *Markov sources*.
    - *Ergodic source* has the property that its output at any time has the same statistical properties as its output at any other time. Memoryless sources are, trivially, always ergodic; a source with memory is ergodic only if it is modeled by an ergodic Markov chain.

# 5.2. Type of information sources (cont.)

- Continuous source:
  - Output is set to be continuous time and continuous value
    - Normally called waveform
    - Random variable that represent continuous propertied by
      - X = {x}   xmin <x < xmax
      - $P_X$(x):  probability density distribution function

# 5.2. Type of information sources (Cont.)

- Binary source:
  - Is a discrete source
  - Alphabet set has only two values
  - Example: X = {0,1} and P(X)= {0.5, 0.5}
- Markov source:
  - Each symbol depends on the previous one.

$$p(x_{i_n}|x_{j_{n-1}}, x_{k_{n-2}}\dots) = p(x_{i_n}|x_{j_{n-1}})$$

  - At time n, output of source can be $x_i$ with the probability $p_i$ = p($x_{i_n}$ | $x_{j_{n-1}}$) when at time (n-1) output of source is $x_j$
  - $\sum_{j=1}^{L} p_{ij} = 1$ L: number of alphabet letters

# 5.2. Type of information sources (Cont.)

- Markov source:
- A Markov source: each symbol depends upon a finite number m of preceding symbols
  - m-th order Markov source
- A Markov source consists:
  - Alphabet
  - Set of states
  - Set of transitions between states
  - Set of labels for the transitions
  - Two sets of probabilities
    - Initial probability distribution on the set of states, which determines the probabilities of sequences starting with each symbol in the alphabet.
    - Set of transition probabilities. For each pair of states, $x_i$ and $x_j$, the probability of a transition from i to j is P(j|i).
      - The labels on the transitions are symbols from the alphabet

# 5.2. Type of information sources (Cont.)

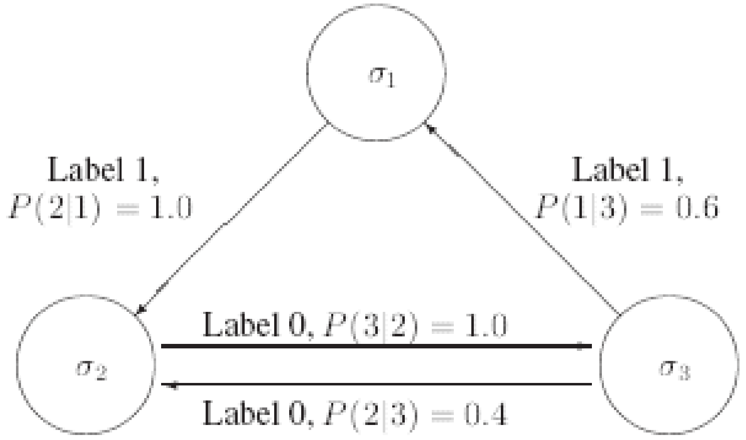
- Example Markov source:
- Alphabet {0,1} and set of states $\{\sigma_1, \sigma_2, \sigma_3\}$
- Suppose there are four transitions:

$\sigma_1 \rightarrow \sigma_2$ with label 1 and P(2|1) = 1

$\sigma_2 \rightarrow \sigma_3$ with label 0 and P(3|2) = 1

$\sigma_3 \rightarrow \sigma_1$ with label 1 and P(1|3) = 0.6

$\sigma_3 \rightarrow \sigma_2$ with label 0 and P(2|3) = 0.4

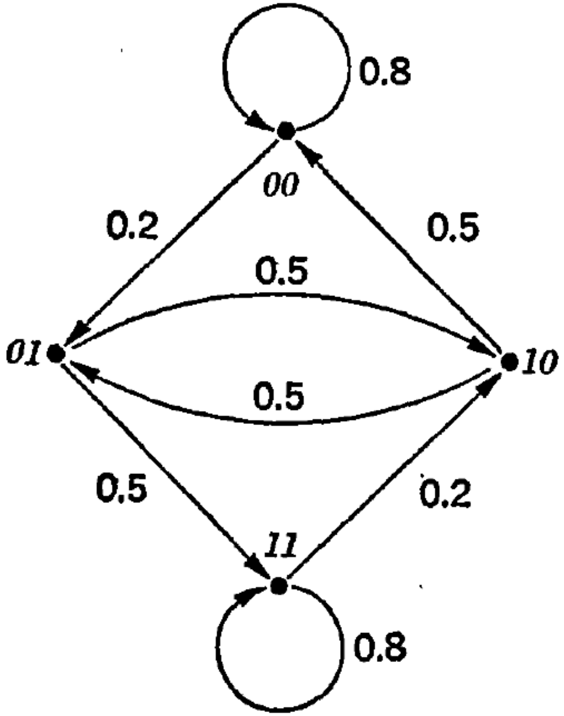



- Initial probability distribution: $P(\sigma_1) = 1/3$, $P(\sigma_2) = 1/3$, $P(\sigma_3) = 1/3$

# 5.2. Type of information sources (Cont.)

- A Markov source whose states are sequences of m symbols from the alphabet is called an mth-order Markov source.
- Example: second-order Markov source

  {0,1}
  P(0|00) = P(1|11) = 0.8
  P(1|00) = P(0|11) = 0.2
  P(0|01) = P(0|10) = P(1|01) = P(1|10) = 0.5

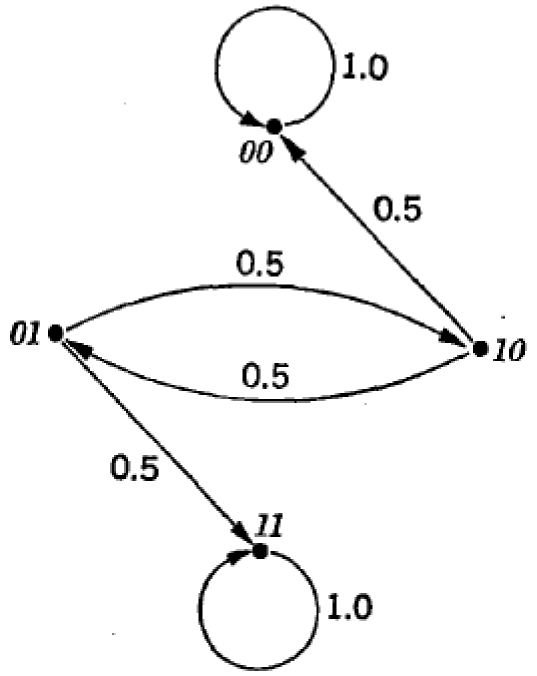



Transition probability from 01 to 10, which would be represented by P(10|01), would be represented instead by the probability of emission of 0 when in the state 01, that is P(0|01)

# 5.2. Type of information sources (Cont.)

- Non ergodic Markov source

{0,1}
P(0|00) = P(1|11) =1.0
P(1|00) = P(0|11) = 0
P(0|01) = P(0|10) = P(1|01) = P(1|10) = 0.5



# 5.2. Type of information sources (Cont.)

- n states $\{\sigma_1, \sigma_2 \dots \sigma_n\}$ has
  - Transition matrix:

$$\Pi = \begin{bmatrix} P(1|1) & P(1|2) & \cdots & P(1|N) \\ P(2|1) & P(2|2) & \cdots & P(2|N) \\ \vdots & \vdots & \ddots & \vdots \\ P(N|1) & P(N|2) & \cdots & P(N|N) \end{bmatrix}$$

  - $w_i^t$ is probability of source at state $\sigma_i$ at time t

$$W^t = \begin{bmatrix} w_1^t \\ w_2^t \\ \vdots \\ w_N^t \end{bmatrix}$$

  - Then: $W^{t+1} = \Pi W^t$

$$W^t = \Pi^t W^0$$

# 5.2. Type of information sources (Cont.)

- Stationary distribution: A probability distribution W over the states of a Markov source with transition matrix $\Pi$ that satisfies the equation $\Pi$ W= W is a stationary distribution
  - $\sum w_i$ = 1
- E.g. W = $\{w_1, w_2, w_3\}$

$$\Pi = \begin{bmatrix} 0.25 & 0.50 & 0.00 \\ 0.50 & 0.00 & 0.25 \\ 0.25 & 0.50 & 0.75 \end{bmatrix}$$

# 5.3. Calculate amount of information

- Discrete memoryless source
- Amount of information of symbol $s_i$

$$I(s_i) = \log \frac{1}{P(s_i)}$$

- Average amount of information per symbol in the source

$$\sum_S P(s_i) I(s_i)$$

- Entropy is defined as the average amount of information

$$H(S) \triangleq \sum_S P(s_i) \log \frac{1}{P(s_i)}$$

- H(S) max = log |S| when S has uniform distribution

# 5.3. Calculate amount of information (Cont.)

- Example:
  - Source S = {s1, s2, s3} with P(s1) = 1/2 and P(s2) = P(s3) = 1/4.
  - Then:

$$H(S) = -\frac{1}{2}\,log_2\frac{1}{2} - \frac{1}{4}\,log_2\frac{1}{4} - \frac{1}{4}\,log_2\frac{1}{4}$$

$$= \frac{1}{2}log_2\,2 + \frac{1}{4}log_2\,4 + \frac{1}{4}log_2\,4$$

$$= \frac{3}{2}\text{ bits/information}$$

# 5.3. Calculate amount of information (Cont.)

- Markov source:
  - $P_i$: probability distribution on the set of transitions from the $i^{th}$ state
  - $H(P_i)$: entropy of the $i^{th}$ state
  - M:

$$H(P_i) = -\sum_{j=1}^{N} P(j|i) \log(P(j|i)$$

$$H(M) = \sum_{i=1}^{N} w_i H(P_i) = -\sum_{i=1}^{N} \sum_{j=1}^{N} w_i P(j|i) \log(P(j|i))$$

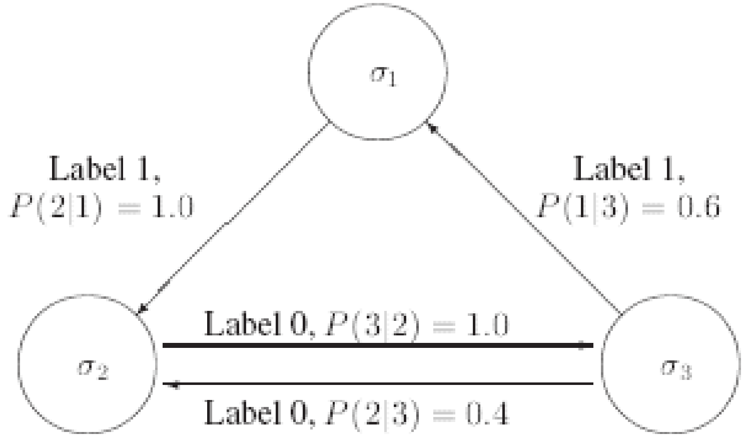
# 5.3. Calculate amount of information (Cont.)

$\sigma_1 \rightarrow \sigma_2$ with label 1 and P(2|1) = 1

$\sigma_2 \rightarrow \sigma_3$ with label 0 and P(3|2) = 1

$\sigma_3 \rightarrow \sigma_1$ with label 1 and P(1|3) = 0.6

$\sigma_3 \rightarrow \sigma_2$ with label 0 and P(2|3) = 0.4




H($P_i$) =? $w_i$ = ? H(M)=?

# 5.3. Calculate amount of information (Cont.)

- Continuous source:
  - Entropy of stationary source:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx$$

  - H(X) max:
    - Source has limited peak power: Pmax, Pmin are limited values
      - xmax = $\sqrt{Pmax}$ ; xmin = $\sqrt{Pmin}$
      - H(X) max = log (xmax – xmin) when source has uniform distribution ( P(x) = 1/(xmax- xmin)) for all x)
    - Source has limited average power: Pav is limited value
      - H(X) max = log $\sqrt{2\Pi e Pav}$ when source has Gaussian distribution
        - e: natural base

# 5.3. Calculate amount of information (Cont.)

- Continuous source:
  - Entropy of stationary source:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \log p(x) dx$$

  - H(X) max:
    - Source has limited peak power: Pmax is limited value
      - xmax = $\sqrt{Pmax}$ ; xmin = -$\sqrt{Pmax}$
      - H(X) max = log (2xmax) when source has uniform distribution ( P(x) = 1/(2xmax) for all x)
    - Source has limited average power: Pav is limited value
      - H(X) max = ln$\sqrt{2\Pi e Pav}$ when source has Gaussian distribution $\int_{\infty}^{\infty} x^2 p(x) dx = P_{av}^2$
        - e: natural base

# 5.4. Redundancy of source

- Source has H(X)max:
  - Amount of information carried by a source symbol is max
- Source has H(X) < H(X)max:
  - Amount of information carried by a source symbol is not max
- Source sequence of H(X)max is min to carry a given amount of information
  - Generate given amount of information: Source has H(X) < H(X)max need more source symbol than source has H(X) = H(X)max
  - Source has H(X) < H(X)max has several redundancy
- Redundancy of source defined by H(X)max – H(X)
  - Domains of sources have H(X)max and H(X) are identical
- Source has redundancy = 0: each symbol carries maximum amount of information
- Source has redundancy >0: need to be compressed to reduce the needed symbols
  - Best compression: compressed source has H(X) = H(X)max

# 5.4. Redundancy of source (Cont.)

- Example:
  - Source S1 = {0,1} with P(S1) = {1/2,1/2}
    - H(S1)max = $-\frac{1}{2}log_2\frac{1}{2}-\frac{1}{2}log_2\frac{1}{2}$ = $log_2 2$ = 1 bit/ symbol
  - Source S2 ={0,1} with P(S2) = {3/4,1/4}
    - H(S2) = $-\frac{3}{4}log_2\frac{3}{4}-\frac{1}{4}log_2\frac{1}{4}= log_2 4-\frac{3}{4}log_2 3$ ≈ 2 – 1.19 ≈ 0.81 bits/symbol

  → To create an amount of information 810 bits
    - S1 needs to generate 810 symbols
    - S2 needs to generate 1000 symbols
  - S2 has redundancy: H(X)max – H(X) = 1 – 0.81 = 0.19 bits/symbol

# 5.5. Extension source

- Extension source of a source S:
  - $S^n$ is a source so that its symbols $s_i^n$ is sequence of n symbols of source ($s_{ij}$)
    - $s_i^n$ = $s_{i1}s_{i2}s_{i3} \dots s_{in}$
  - The symbols in the sequence $s_i^n$ are independent
    - P($s_i^n$) = P($s_{i1}$) P($s_{i2}$)… P($s_{in}$)
- Entropy of $S^n$:
  - H($S^n$) = n H(S)

# 5.5. Extension source

- Memoryless source S{0,1}
- $P_0$ = 0.2, $P_1$ = 0.8
- Extension source?

E.g: $P_{00}$, $P_{001}$? H($S^2$)?

# 5.6. Information rate

- Information rate (R): Average amount of information that source can generate in a unit of time
- R = $n_o \times H(X)$
  - $n_o$: number of symbol that source can generate in a unit of time
  - H(X): average amount of information per symbol (entropy)
- In many cases of information theory, $n_o$ is physical parameter so that $n_o$ is assigned to unit value $(n_o = 1)$
- In case of discrete
  - $n_o$ = F
    - F: number of generated symbol in unit of time (frequency)
    - R = F $\times$ *H(X)*
    - *Rmax = F $\times$ log |X|*

# 5.6. Information rate (Cont.)

- Source transmits 9.6 kbaud: = no

  (baud = symbol/ second)

| $X_i$ | $P(X_i)$ | BCD word |
|---|---|---|
| A | 0.30 | 000 |
| B | 0.10 | 001 |
| C | 0.02 | 010 |
| D | 0.15 | 011 |
| E | 0.40 | 100 |
| F | 0.03 | 101 |

- Information rate =?

# 5.6. Information rate (Cont.)

$$
\begin{aligned}
H &= -\sum_{i=1}^{6} P(X_i) \cdot \log_2 P(X_i) = -0.30 \cdot \log_2 0.30 - 0.10 \cdot \log_2 0.10 - 0.02 \cdot \log_2 0.02 \\
&\quad -0.15 \cdot \log_2 0.15 - 0.40 \cdot \log_2 0.40 - 0.03 \cdot \log_2 0.03 \\
&= 0.52109 + 0.33219 + 0.11288 + 0.41054 + 0.52877 + 0.15177 \\
&= 2.05724 \text{ bits/symbol}
\end{aligned}
$$

Information rate: $R = H \cdot R_s = 2.05724 \text{ [bits/symbol]} \cdot 9600 \text{ [symbols/s]} = 19750 \text{ [bits/s]}$

# 5.6. Information rate (Cont.)

- In case of continuous
- $n_o$ is number of samples of corresponding discreted source
  - $n_o$ = 2 Fmax
    - Fmax: maximum frequency of the continuous source
    - R = 2 Fmax x H(x)
    - R = 2 Fmax x log (xmax – xmin) when source has limited peak power
    - R = 2 Fmax x log $\sqrt{2\Pi e P a v}$ when source has limited average power

# 5.6. Information rate (Cont.)

- In case of continuous
- $n_o$ is number of samples of corresponding discreted source
  - $n_o$ = 2 Fmax
    - Fmax: maximum frequency of the continuous source
    - R = 2 Fmax x H(x)
    - R = 2 Fmax x log (2xmax ) when source has limited peak power
    - R = 2 Fmax x ln $\sqrt{2\Pi e P a v}$ when source has limited average power