Lecture Outlines :

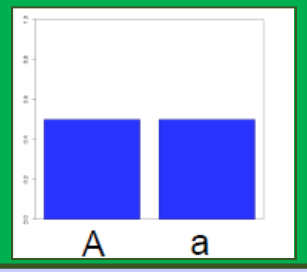
- Random Variables
- Discrete Probability Distribution
- Distribution Functions for Random Variables
- Expectation of a Discrete Random Variables
- Variance of a Discrete Random Variable
- Continuous Random Variables
- Distribution Functions for Continuous Random Variables
- Variance of a Discrete Random Variable
- The Binomial Distribution
- Normal distribution (Gaussian)
- Poisson Distributions

Random Variables: Suppose that to each point of a sample space we assign a number. We then have a function defined on the sample space. This function is called a random variable (or stochastic variable). It is usually denoted by a capital letter such as X or Y.

Example 1: X is the variable for the number of heads for a coin tossed three times
Solution: X = 0, 1, 2, 3

Discrete random variable has accountable number of possible values

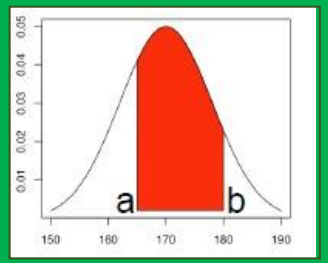
- Number of sales, Number of calls, People in line, Mistakes per page, dice.



A continuous random variable takes all values in an interval of numbers

- electrical current, pressure, temperature, time, voltage, blood pressure,

the speed of a car,
the real numbers from 1 to 6.



The probability distribution of a discrete random variable is a graph, table or formula that specifies the probability associated with each possible outcome the random variable can assume.

- $f(x) \geq 0$ **for all values of x**
- $\sum_x f(x) = 1$ $f(x)$ **is the probability function**

Example 2: Suppose that a fair coin is tossed twice. Let *X* represent the number of heads that can come up. Find the probability function corresponding to the random variable X.

Solution: For each sample point we can associate a number for X as follows:

Sample Point	HH	HT	TH	TT
X	2	1	1	0

P(HH) = 1/4, P(HT) = 1/4, P(TH) = 1/4, P(TT) = 1/4
P(X=0) = P(TT) = 1/4
P(X=1) = P(HT ∪ TH) = P(HT)+P(TH)= 1/4+1/4 =1/2
P(X=2) = P(HH) = 1/4

The probability function is given by

x	0	1	2
f(x)	1/4	1/2	1/4

Cumulative Distribution Function (CDF) F(x) **- is a function that returns the probability that a random variable X is less than or equal to a value. The CDF is also sometimes called the distribution function (DF).**

$$F(x) = P(X \leq x)$$

Requirements for CDFs

(1) $F(x) \geq 0$ **everywhere the distribution is defined**
(2) $F(x)$ **non-decreasing everywhere the distribution is defined.**
(3) $F(x) \rightarrow 1$ **as** $x \rightarrow \infty$

Example 3: Consider the probability distribution of the number of rewards you will get this semester

x	*f(x)*	*F(x)*
0	*0.05*	*0.05*
1	*0.15*	*0.20*
2	*0.20*	*0.40*
3	*0.60*	*1.00*

Expectation : **The expected value, or mean, of a random variable is a measure of central location.**

$$E(X) = \Upsilon = \sum x.f(x)$$

Example 4: **Suppose that a game is to be played with a single die assumed fair. In this game a player wins 20$ if a 2 turns up; 40$ if a 4 turns up; loses 30$ if a 6 turns up; while the player neither wins nor loses if any other face turns up. Find the expected sum of money to be won.**

Solution:

$f(x_1) = f(x_2) = f(x_3) = f(x_4) = f(x_5) = f(x_6) = 1/6$

x	0	+20	0	+40	0	-30
f(x)	1/6	1/6	1/6	1/6	1/6	1/6

The expected value, or expectation, is

$$E(X) = (0)\left(\frac{1}{6}\right) + (20)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (40)\left(\frac{1}{6}\right) + (0)\left(\frac{1}{6}\right) + (-30)\left(\frac{1}{6}\right) = 5$$

The player should expect to pay 5 $ in order to play the game.

Variance : *A positive quantity that measures the spread of the distribution of the random variable about its mean value. Larger values of the variance indicate that the distribution is more spread out.*

$$\sigma_{XX}^{22} = E[(XX-\mu)^{22}] = \sum_{jj=11}^{n}(x_{jj}-\mu)^{22}f(x_{jj}) = \sum_{x}(x-\mu)^{22}f(x)$$

The *standard deviation* is the positive square root of the variance

$$\sqrt{\sigma^{22}} = \sqrt{E[(XX-\mu)^{22}]} = \sqrt{\sum(x-\mu)^{22}f(x)}$$

Example 5: Find the variance and standard deviation for the game played in previous Example

Solution:

$E(XX) = \Upsilon = 5$

x_{jj}	0	+20	0	+40	0	-30
$f(x_{jj})$	1/6	1/6	1/6	1/6	1/6	1/6

$$\sigma_{11}^{22} = (00-55)^{22}\left(\frac{11}{66}\right) + (2200-55)^{22}\left(\frac{11}{66}\right) + (00-55)^{22}\left(\frac{11}{66}\right) + (4400-55)^{22}\left(\frac{11}{66}\right) + (00-55)^{22}\left(\frac{11}{66}\right) + (-3300-55)^{22}\left(\frac{11}{66}\right) = \frac{22225500}{66}$$

$$\sigma_{XX}^{22} = \sqrt{\frac{22225500}{66}}$$

Some properties of expected values :

- $E(a \cdot x) = a \cdot E(x) = a \cdot \mu$

where a is a constant

- $E(a \cdot x + b) = a \cdot E(x) + b = a \cdot \mu + b$ where a and b are constants
- $E(x + y) = E(x) + E(y)$

Some properties of variance :

- $\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$ where $\mu = E(X)$
- $Var(cX) = c^2 Var(X)$ where c is any constant
- The quantity $E[(X - a)^2]$ is a minimum when $a = \mu = E(X)$
- If X and Y are independent random variables,

$Var(X + Y) = Var(X) + Var(Y)$ or $\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$

$Var(X - Y) = Var(X) - Var(Y)$ or $\sigma^2_{X-Y} = \sigma^2_X - \sigma^2_Y$

Exercise 1: compute the expected value and variance of the number of rewards in Example 3 in this Lecture.

$E(x) = 0*.05+1*0.15+2*0.20+3*0.60 = 2.35$

$Var\ (x) = (0-2.35)^2 * 0.05+ (1-2.35)^2*0.15+(2-2.35)^2*0.20+(3-2.35)^2*0.60 = 0.8275$

Homework 1: Let X and Y be the random independent events of rolling a fair die. Compute the expected value of X + Y, and the variance of X + Y.

Homework 2: The number of e-mail messages received per hour has the following distribution. compute the expected value and variance.

x = number of message	10	11	12	13	14	15
f(x)	0.08	0.15	0.30	0.20	0.20	0.07

Continuous Random Variables :

A nondiscrete random variable *X* is said to be *absolutely continuous*, or simply *continuous*, if its distribution function may be represented as

$$F(x) = P(XX \le x) = \int_{-\infty}^{x} f(u)du$$

where the function *f(x)* has the properties

1. $f(x) \ge 00$
2. $\int_{-\infty}^{\infty} f(x) = 11$

The function *f(x)* is called the probability density function (p.d.f.).

Example 6: **Find the constant c such that the function**

$$f(x) = \begin{cases} cx^2 & 0 < x < 3 \\ 00 & otheroomse \end{cases}$$

is a density function, and then find *P*(1 < *X* < 2).

Solution:

Notice that if $c \ge 0$, then Property 1 is satisfied. So *f* (*x*) must satisfy Property 2 in order for it to be a density function.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{3} cx^2\,dx = \left|\frac{cx^3}{3}\right|_{0}^{3} = 9c$$

and since this must equal 1, c = $\frac{1}{9}$, and our density function is

$$f(x) = \begin{cases} \frac{1}{9}x^2 & 0 < x < 3 \\ 0 & otherwise \end{cases}$$

Next,

$$P(1 < X < 2) = \int_{1}^{2} \frac{1}{9}x^2\,dx = \frac{x^3}{27}\Big|_{1}^{2} = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

Cumulative Distribution Function (c.d.f.)

The c.d.f. of a continuous random variables is defined exactly the same as for discrete random variables

$$F(x) = P(X \leq x)$$

where *x* is any real number, i.e., $-\infty \leq x \leq \infty$. So,

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

Example 7: **Find the distribution function for example 6.**

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{99} x^2 dx = \frac{x^3}{2222}$$ ***where $x \leq 3$.***

Expectation :

If *X* is a continuous random variable having probability density function *f(x)*, then it can be shown that

$$\mu_x = E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Example 8: **The density function of a random *X* is given by**

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

The expected value of *X* is then

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^2 x\left(\frac{1}{2}x\right) dx = \int_0^2 \frac{x^2}{2}\, dx = \left.\frac{x^3}{6}\right|_0^2 = \frac{4}{3}$$

Variance :

If *X* is a continuous random variable having probability density function *f(x)*, then the variance is given by

$$\sigma_{XX}^{22} = E\left[(XX-\mu)^{22}\right] = \int_{-\infty}^{\infty} (x-\mu)^{22} f(x)dx$$

Example 9: **Find the variance and standard deviation of the random variable from Example 8, using the fact that the mean was found**

To be $\mu_x = E(X) = \frac{44}{33}$

$$\sigma^{22} = E\left[(XX-\frac{44}{33})^{22}\right] = \int_{-\infty}^{\infty}\left(x-\frac{44}{33}\right)^{22} f(x)dx = \int_{-\infty}^{\infty}\left(x-\frac{4}{33}\right)^{22}\left(\frac{11}{22}x\right)dx = \frac{22}{99}$$

And so the standard deviation is $\sigma = \sqrt{\frac{22}{99}} = \frac{\sqrt{22}}{33}$.

Binomial: **An experiment for which the following four conditions are satisfied is called a *binomial experiment*.**

1. **The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.**
2. **The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success (S) or failure (F).**
3. **The trials are independent.**
4. **The probability of success is constant from trial to trial: denoted by p.**

The probability mass function (PMF) of X is

for x = 0; 1; 2; . . . ; n

***p* = probability of success**

***q* = *(1 – p)* probability of failure**

***n* = number of trials**

***x* = number of successes**

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Example 10: In a digital communication system, the number of bits in error in a packet depicts a Binomial discrete random variable

Example 11: **The probability of getting exactly 2 heads in 6 tosses of a fair coin is**

Solution:

$$P(XX = 22) = \binom{66}{22}\left(\frac{11}{22}\right)^{22}\left(\frac{11}{22}\right)^{66-22} = \frac{66!}{22!\ 44!}\left(\frac{11}{22}\right)^{22}\left(\frac{11}{22}\right)^{44} = \frac{1155}{6644}$$

Exercise 2: **If I toss a coin 20 times, what's the probability of getting 2 or fewer heads?**

Solution:

$$\binom{20}{0}(.5)^0(.5)^{20} = \frac{20!}{20!0!}(.5)^{20} = 9.5x10^{-7} +$$

$$\binom{20}{1}(.5)^1(.5)^{19} = \frac{20!}{19!1!}(.5)^{20} = 20x9.5x10^{-7} = 1.9x10^{-5} +$$

$$\binom{20}{2}(.5)^2(.5)^{18} = \frac{20!}{18!2!}(.5)^{20} = 190x9.5x10^{-7} = 1.8x10^{-4}$$

$$= 1.8x10^{-4}$$

(Mean and Variance for Binomial Distribution) :

If X is a binomial random variable with parameters p and n, then

Mean $\mu = np$

Variance $\sigma^2 = npq$

Standard Deviation $\sigma = \sqrt{npq}$

Exercise 3: **Digital Channel the chance that a bit transmitted through a digital transmission channel is received in error is 0.1. Also, assume that the transmission trials are independent. Let X = the number of bits in error in the next four bits transmitted. Determine P(X = 2), the mean, variance, and standard deviation of this experiment.**

Solution:

$$P(XX = 22) = \binom{44}{22} (00.11)^{22} (00.99)^{44-22}$$

$$E(X) = \mu_x = 44(00.11) = 00.44 \qquad \sigma^{22} = 44(00.11)(00.99) = 00.3366$$

Outcome	x	Outcome	x
OOOO	0	*EOOO*	1
OOOE	1	*EOOE*	2
OOEO	1	*EOEO*	2
OOEE	2	*EOEE*	3
OEOO	1	*EEOO*	2
OEOE	2	*EEOE*	3
OEEO	2	*EEEO*	3
OOOO	3	*EEEE*	4

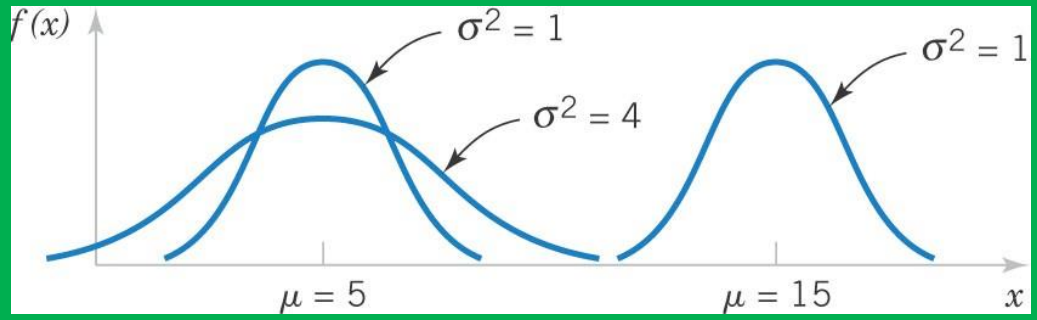
Homework 3: Suppose that three telephone users use the same number and that we are interested in estimating the probability that more than one will use it at the same time. If independence of telephone habit is assumed, the probability of exactly k persons requiring use of the telephone at the same time is given by the mass function $p_X(k)$ associated with the binomial distribution. Let it be given that, on average, a telephone user is on the phone 5 minutes per hour.

Homework 4: In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable. The probability that a bit is received in error is 10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur? Let X denote the number of errors.

Note : Clearly, this probability is difficult to compute. Fortunately, the normal distribution can be used to provide an excellent approximation in this example.

Homework 5: Consider the problem of missile firing. Enumerate the possible outcomes of trail firing of missiles. Lets S denote the success of each trail with probability *p* and *F* denote the failure with probability *q=1-p*, then there are, 2^4 possible outcomes s listed below as set U, the complete sample space.

- One of the most important examples of a continuous probability distribution is the ***normal distribution***, sometimes called the ***Gaussian distribution***. most popular to communication engineers is … AWGN Channels.
- Random variation of many physical measurements are normally distributed.
- The location and spread of the normal are independently determined by mean (μ) and standard deviation (σ).

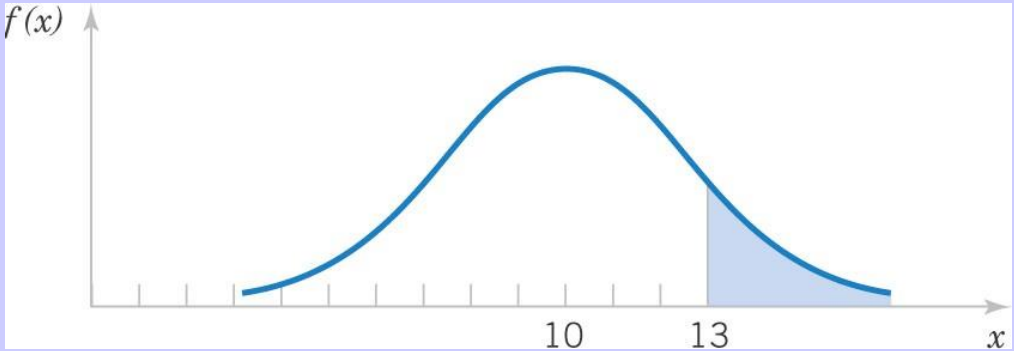


The density function for this distribution is given by

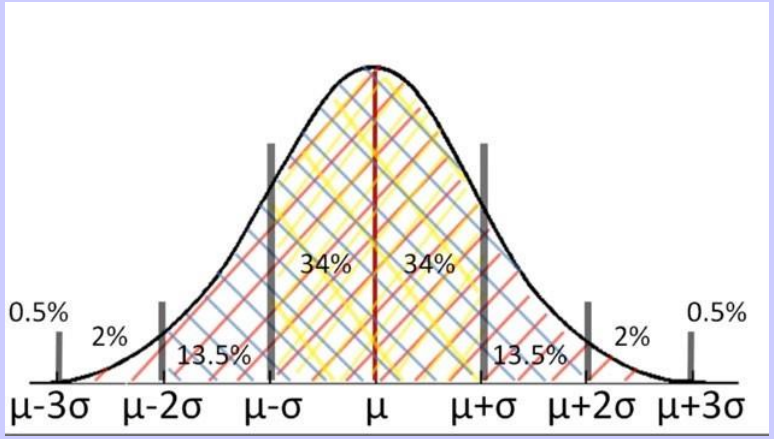
$$f(x) = \frac{11}{\sqrt{22\pi\sigma}} e^{\frac{-(x-\mu)^{22}}{22\sigma^{22}}} \qquad -\infty < x < \infty$$

where μ and σ are the mean and standard deviation, respectively. The corresponding distribution function is given by $F(x) = P(XX \le x) = \frac{11}{\sigma\sqrt{22\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^{22}/22\sigma^{22}} dv$

Example 12: **Assume that the current measurements in a strip of wire follows a normal distribution with a mean of 10 mA & a variance of 4 mA2. Let X denote the current in mA. What is the probability that a measurement exceeds 13 mA?**



Graphical probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$.



A normal random variable with

$$\mu = 0 \text{ and } \sigma^{22} = 1$$

Is called a standard normal random variable and is denoted as Z.

$$ZZ = \frac{XX - \mu}{\sigma}$$

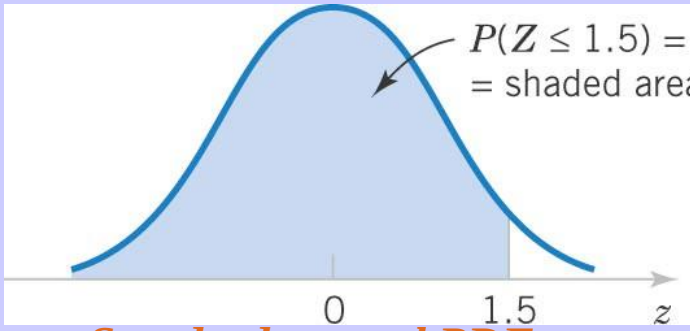
The cumulative distribution function of a standard normal random variable is denoted as:

$$\Phi(z) = P(Z \leq z) = F(z)$$

Values are found in Z Table.

Example 13: **Assume Z is a standard normal random variable. Find P(Z ≤ 1.50).**

Answer: **0.93319**



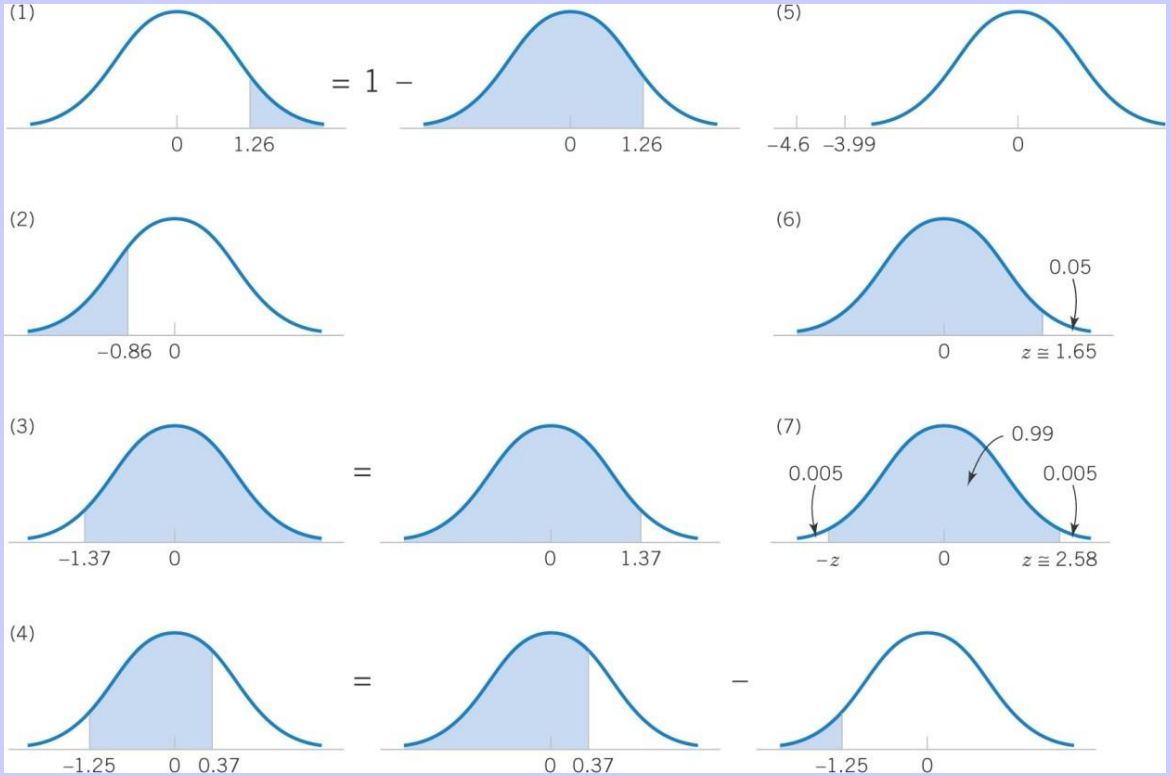
Standard normal PDF

z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Exercise 4: $P(Z \leq 1.53)$. **Answer:** 0.93699
Exercise 5: $P(Z \leq 0.01)$. **Answer:** 0.50398

Example 14:

1. P(Z > 1.26) = **0.1038**
2. P(Z < -0.86) = **0.195**
3. P(Z > -1.37) = **0.915**
4. P(-1.25 < 0.37) = **0.5387**
5. P(Z ≤ -4.6) ≈ **0**
6. Find z for P(Z ≤ z) = **0.05, z = -1.65**
7. Find z for (-z < Z < z) = **0.99, z = 2.58**



Graphical displays for standard normal distributions.

Example 15: **From a previous example with μ = 10 and σ = 2 mA, what is the probability that the current measurement is between 9 and 11 mA?**

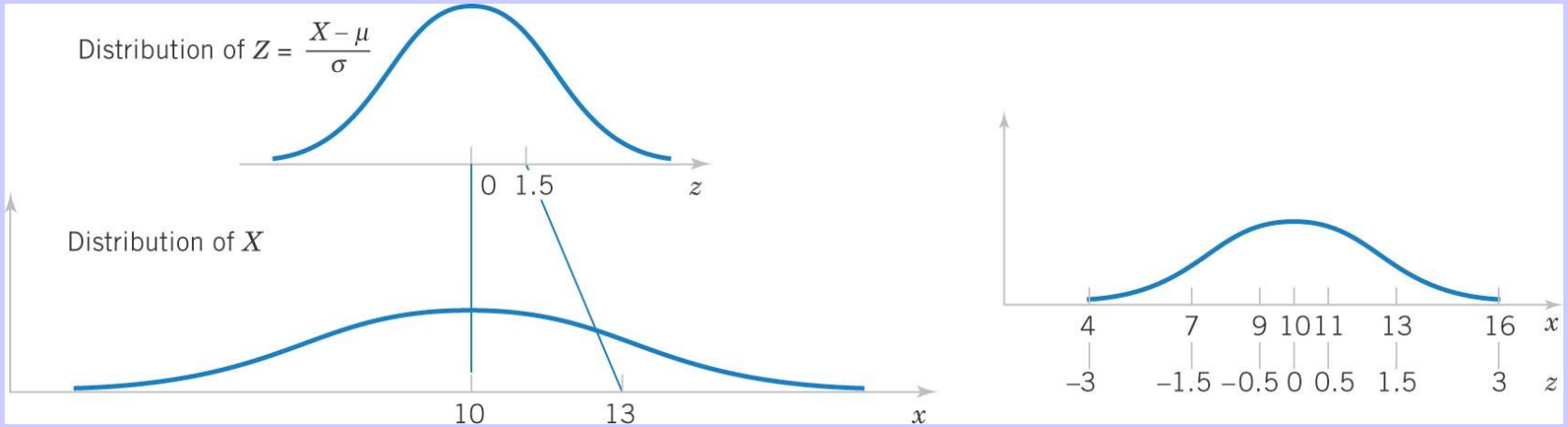
Answer:

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2}\right)$$

$$= P(-0.5 < z < 0.5)$$

$$= P(z < 0.5) - P(z < -0.5)$$

$$= 0.69146 - 0.30854 = 0.38292$$



Example 16: **Determine the value for which the probability that a current measurement is below this value is 0.98.**

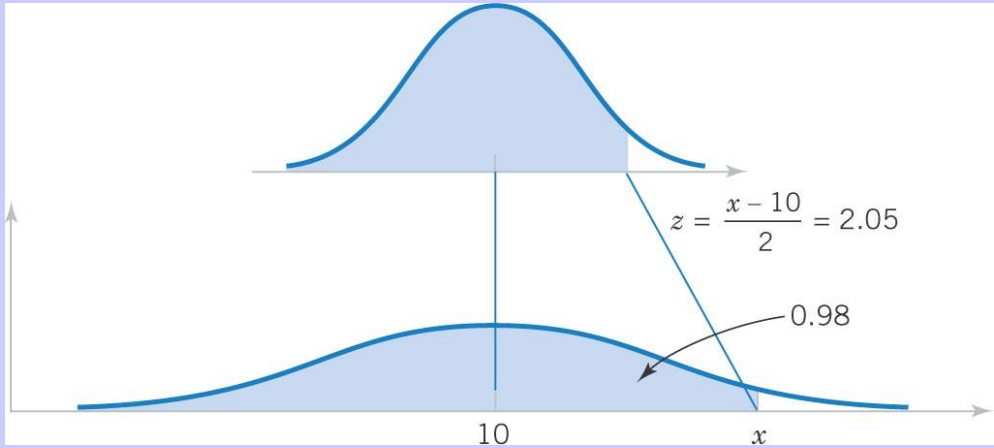
Answer:

$$P(X < x) = P\left(\frac{X-10}{2} < \frac{x-10}{2}\right)$$

$$= P\left(Z < \frac{x-10}{2}\right) = 0.98$$

$z = 2.05$ is the closest value.

$$z = 2(2.05) + 10 = 14.1 \text{ mA}.$$



Determining the value of x to meet a specified probability.

Homework 6: **Assume that in the detection of a digital signal, the background noise follows a normal distribution with μ = 0 volt and σ = 0.45 volt. The system assumes a digital 1 has been transmitted when the voltage exceeds 0.9.**

1. **What is the probability of detecting a digital 1 when none was sent? Let the random variable N denote the voltage of noise.**
2. **Determine the symmetric bounds about 0 that include 99% of all noise readings.**
3. **Suppose that when a digital 1 signal is transmitted, the mean of the noise distribution shifts to 1.8 volts. What is the probability that a digital 1 is not detected? Let S denote the voltage when a digital 1 is transmitted.**

Homework 7: Repeat solve Homework 4 by using Gaussian distribution.

Homework 8: A television cable company receives numerous phone calls throughout the day from customers reporting service troubles and from would-be subscribers to the cable network. Most of these callers are put "on hold" until a company operator is free to help them. The company has determined that the length of time a caller is on hold is normally distributed with a mean of 3.1 minutes and a standard deviation 0.9 minutes. Company experts have decided that if as many as 5% of the callers are put on hold for 4.8 minutes or longer, more operators should be hired.

a) What proportion of the company's callers are put on hold for more than 4.8 minutes? Should the company hire more operators?

b) At another cable company (length of time a caller is on hold follows the same distribution as before), 2.5% of the callers are put on hold for longer than x minutes. Find the value of x

Homework 9: Suppose that a binary message either 0 or 1 must be transmitted by wire from location *A* to location *B*. However, the data sent over the wire are subject to a channel noise disturbance, so to reduce the possibility of error, the value 2 is sent over the wire when the message is 1 and the value - 2 is sent when the message is 0. If x, $x = \pm 2$, is the value sent from location *A*, then *R*, the value received at location *B*, is given by $R = x + N$, where *N* is the channel noise disturbance. When the message is received at location B the receiver decodes it according to the following rule:

If $R \geq 0.5$, then 1 is concluded

If $R < 0.5$, then 0 is concluded

If the channel noise follows the standard normal distribution compute the probability that the message will be wrong when decoded.

Let X be a discrete random variable that can take on the values 0, 1, 2,… such that the probability function of X is given by

$$f(x) = P(X = x) = \frac{\lambda^{x} e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \ldots \qquad \lambda = np$$

where λ is a given positive constant. This distribution is called the Poisson distribution

Example 17: If the probability that an individual will suffer a bad reaction from injection of a given serum is 0.001, determine the probability that out of 2000 individuals, (a) exactly 3, (b) more than 2, individuals.

Solution:

X is Bernoulli distributed, but since bad reactions are assumed to be rare events, we can suppose that X is Poisson distributed

a) $P(X = x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$ $\lambda = np = (2000)(0.001) = 2$

$P(X = 3) = \frac{\lambda^{3} e^{-2}}{3!} = 0.18$

b) $P(X>2) = 1 - [P(X=0)+P(X=1)+P(X=2)]$

$$= 11 - \left[\frac{2^{00}e^{-2}}{00!} + \frac{2^{11}e^{-2}}{11!} + \frac{2^{22}e^{-2}}{22!}\right.$$

$$= 1 - 5e^{-2} = 0.323$$

Exercise 6: **Let's say you want to send a bit string of length $n = 10^{44}$ where each bit is independently corrupted with $p = 10^{-66}$. What is the probability that the message will arrive uncorrupted?**

Solution:

$\lambda = np = 10^{44}10^{-66} = 0.01$.

$$P(X = 00) = \frac{0.01^{00}e^{-0.01}}{00!} = 00.9999$$

Homework 10: **Are we could have modelled X as a binomial distribution. That would have been computationally harder to compute but would have resulted in the same number (up to the millionth decimal)**

Example 18: **The number of visitors to a webserver per minute follows a Poisson distribution. If the average number of visitors per minute is 4, what is the probability that:**

a) There are two or fewer visitors in one minute?

b) There are exactly two visitors in 30 seconds?

Solution:

a) The average number of visitors in a minute. In this case the parameter λ = 4. So the probability of two or fewer visitors in a minute is

$$P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X = 0) = \frac{e^{-4}4^{0}}{0!} = e^{-4}$$

$$P(X = 1) = \frac{e^{-4}4^{1}}{1!} = 4e^{-4}$$

$$P(X = 2) = \frac{e^{-4}4^{2}}{2!} = 8e^{-4}$$

The probability of two or fewer visitors in a minute is $e^{-4} + 4e^{-4} + 8e^{-4} = 0.238$

b) If the average number of visitors in 1 minute is 4, the average in 30 seconds is 2.

$$P(X = 2) = \frac{e^{-2}2^{2}}{2!} = 2e^{-2} = 0.271.$$

Homework 12: **Consider a computer system with Poisson job-arrival stream at an average of 2 per minute. Determine the probability that in any one-minute interval there will be**

a) 0 jobs;
b) exactly 2 jobs;
c) at most 3 arrivals.

Uniform Distribution

A random variable X is said to be uniformly distributed in a ≤ x ≤ b if its density function is flat over a region.

$$f(x) = \frac{11}{b-a}, \qquad for\ a \le x \le b$$

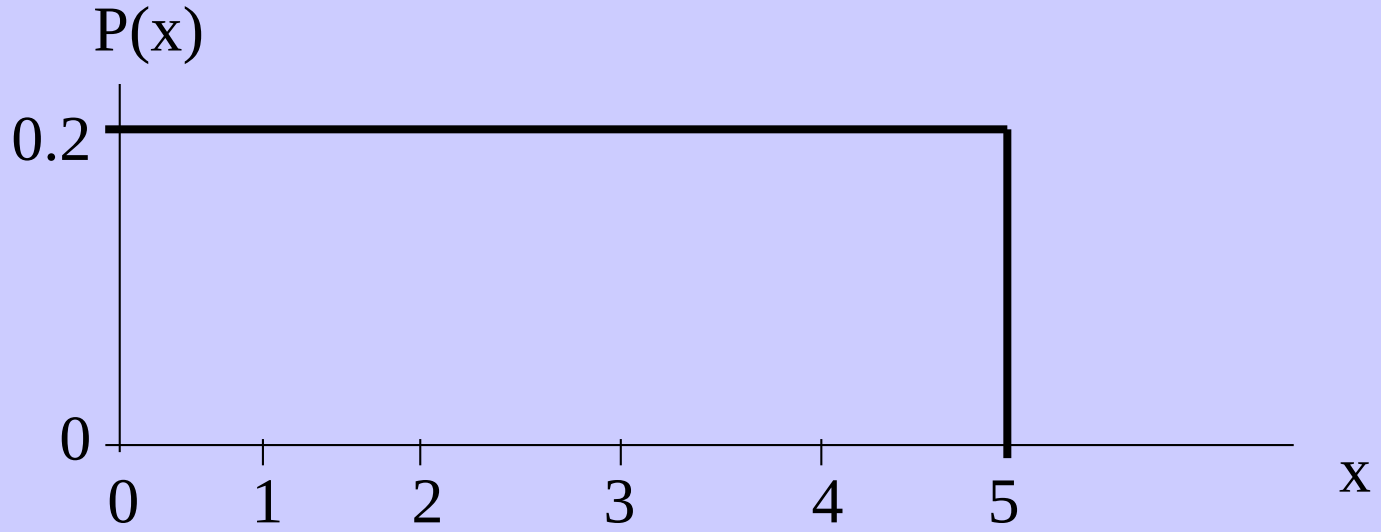
The distribution function is given by

$$F(x)=P(X \le x) = \begin{cases} 00 & xx < a \\ (x-a)/(b-a), & a \le x \le b \\ 11 & xx \ge b \end{cases}$$

The mean and variance are, respectively

$$\mu_x = \frac{11}{22}(a+b), \qquad \sigma^{22} = \frac{11}{1122}(b-a)^{22}$$

Example 19: The figure below depicts the probability distribution for temperatures in a manufacturing process. The temperatures are controlled so that they range between 0 and 5 degrees Celsius, and every possible temperature is equally likely.



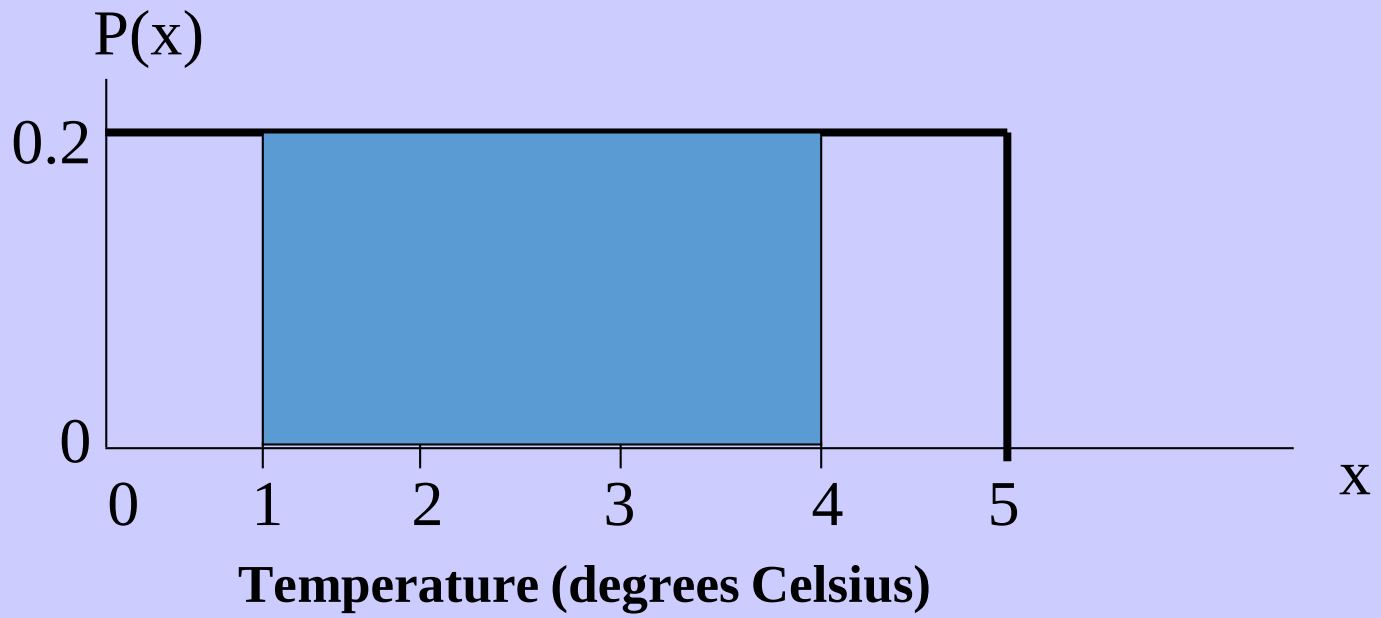
Temperature (degrees Celsius)

1. What is the Probability that the temperature is exactly 4 degrees? **Answer:** 0

Since we have a continuous random variable there are an infinite number of possible outcomes between 0 and 5, the probability of one number out of an infinite set of numbers is 0.

2. What is the probability the temperature is between 1C and 4C?

Answer:



The total area of the rectangle is 1, and we can see that the part of the rectangle between 1 and 4 is 3/5 of the total, so P(1 ≤ x ≤ 4) = 3/5*(1) = 0.6.

Homework 13: **The current (in mA) measured in a piece of copper wire is known to follow a uniform distribution over the interval [0, 25]. Write down the formula for the probability density function f(x) of the random variable X representing the current. Calculate the mean and variance of the distribution and find the cumulative distribution function F(x).**

Ministry of Higher Education and Scientific Research

Al-Furat Al- Awsat Technical University

Engineering Technical college / Najaf

Communication Engineering Department

Information Technology (CE231)

2nd Class 2018/2019

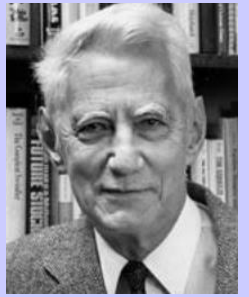
Lecturer Ali M. Alsahlany

Lecture Outlines :

- **Introduction**
- **General model of communication system**
- **Information Source**
- **Self Information**
- **Entropy**
- **Information Rate**
- **Joint Entropy**
- **Conditional Entropy**
- **Mutual Information**

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

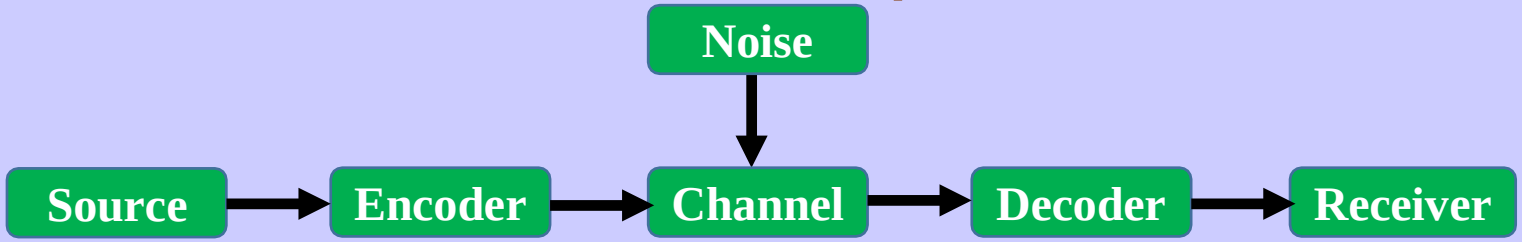
(Claude Shannon 1948)

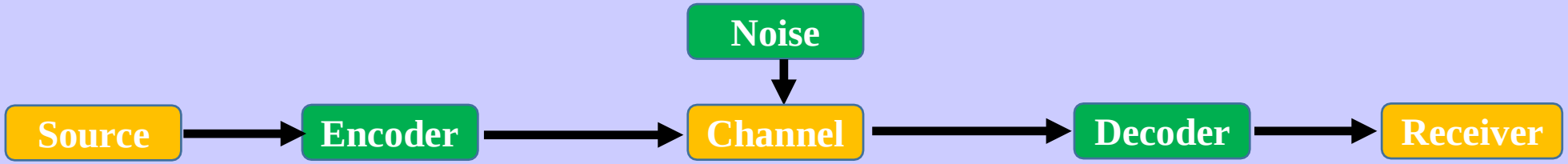


Information Theory is concerned with the theoretical limitations and potentials of systems that communicate. E.g., \What is the best compression or communications rate we can achieve"

Communication is sending information from one place and/or time to another place and/or time over a medium that might have errors.

General model of communication





Source

- *Voice, Words, Pictures, Music.*

Channel

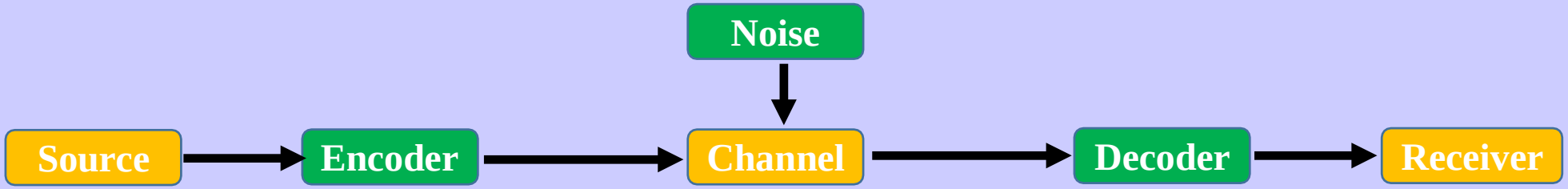
- *Telephone line, High frequency radio link, Space communication link, Biological organism (send message from brain to foot, or from ear to brain)*

Noise

- *Some signal with time-varying frequency response, cross-talk, thermal noise, impulsive switch noise, etc.*
- *Represents our imperfect understanding of the universe. Thus, we treat it as random, often however obeying some rules, such as that of a probability distribution.*

Receiver

- *The destination of the information transmitted, Person, Computer, Disk, Analog Radio or TV internet*



Encoder

- *processing done before placing info into channel*
- *First stage: data reduction (keep only important bits or remove source redundancy),*
- *Followed by redundancy insertion catered to channel.*
- *A code = a mechanism for representation of information of one signal by another.*
- *An encoding = representation of information in another form.*

Decoder

- *exploit and then remove redundancy*
- *remove and fix any transmission errors*
- *restore the information in original form*

Consider Discrete Information Source

Assuming:

- *Information source generates a group of symbols from a given alphabet* $S = \{s_o, s_1, \ldots, s_{K-1}\}$
- *Each symbol has a probability* P_k
- *Symbols are independent*

Information Source → $S = \{s_o, s_1, \ldots, s_{K-1}\}$

$P(s_k) = P_k, \; k = 0,1,\ldots,K-1$, $\sum_{k=0}^{K-1} P_k = 1$

The amount of information gained from knowing that the source produces the symbols is s_k *is* related with as p_k follows :

- *If* $P_k = 1$

Then there is no uncertainty of occurrence of the event ; no gain of information i.e., there is no need for communications because the receiver knows everything.

- *As* P_k *decreases*

Then the uncertainty increases; the reception of s_k *corresponds to some gain of information*

But How Much?

Self Information: Is a function which measures the amount of information after observing the symbol s_k

$$I(s_k) = \log_b \frac{1}{P(s_k)}$$

2, bits e=2.718, nats 10, Hartleys

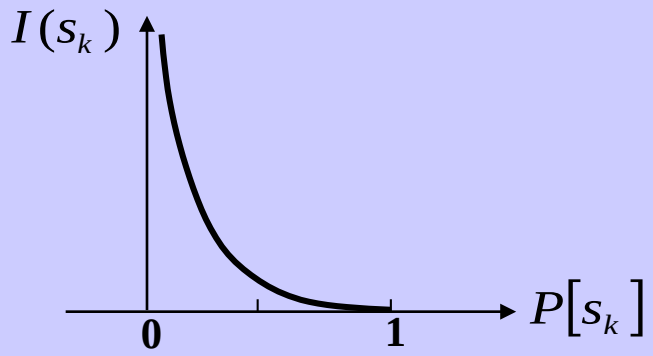
$$\log_a(p) = \frac{Ln\ p}{Ln\ a}$$

The unit of information depends on the base of the log

The amount of information in bits about a symbol is closely related to its probability of occurrence

A low probability event contains a lot of information and vice versa

Properties of Self Information



$I(S_i) \geq 0$, a real nonnegative measure

$I(S_i)$ is a continuous function of p

$I(s_k) > I(s_i)$ if $P_k < P_i$

-The information obtained from the occurrence of two independent events is the sum of the information obtained from the occurrence of the individual events

$$I(AB) = \log_b \frac{1}{P(AB)}$$

$$= \log_b \frac{1}{P(A)P(B)}$$

$$= \log_b \frac{1}{P(A)} + \log_b \frac{1}{P(B)}$$

$$= I(A) + I(B)$$

$$\therefore I(AB) = I(A) + I(B)$$

Example 1: *Let H and T are the outcomes of a flipping coin, calculate the self information for the following cases:*

(a) Fair coin with P (H) = P (T) = 0.5
(b) Unfair coin with P(H) = 1/8, P(T) = 7/8

(a) I (H) =I (T))= 1 bit
(b) I (H) = 3 bits I (T) = 0.193 bits

Example 2: *A source puts out one of five possible messages during each message interval. The probability of these messages are {$m_1,\ldots,m_5$} :P1 = 1/2, P2 = 1/4, P3 = 1/8, P4 = 1/16 and P5 = 1/16 What is the information content of these messages in bit?*

$$I(m_1)=\log_2\frac{1}{P(m_1)}=-\log_2 P(m_1)=-\log_2(\frac{1}{2})=1bit$$

$$I(m_2)=\log_2\frac{1}{P(m_2)}=-\log_2 P(m_2)=-\log_2(\frac{1}{4})=2bits$$

$$I(m_3)=\log_2\frac{1}{P(m_3)}=-\log_2 P(m_3)=-\log_2(\frac{1}{8})=3bits$$

$$I(m_4)=\log_2\frac{1}{P(m_4)}=-\log_2 P(m_4)=-\log_2(\frac{1}{16})=4bits$$

$$I(m_5)=\log_2\frac{1}{P(m_5)}=-\log_2 P(m_5)=-\log_2(\frac{1}{16})=4bits$$

Exercise 1: For 128 equally likely and independents messages find the information content (in bits) in each of the messages.

Solution:

$$I(m) = \log_2 \frac{1}{P(m)} = -\log_2 P(m) = \log_2(128) = 7 bits$$

Homework 1: Suppose in sizing up the data storage requirements for a word processing system to be used in production of a book, it is required to calculate the information capacity. The book consist of 450 pages, 500 words per page, each containing 5 symbols are chosen at random from a 37- ary alphabet (26 letters, 10 numerical digits and one blank space). Calculate the information capacity of the book.

Entropy: ***It is the average number of bits per symbol required to describe a source***

For a source containing N independent symbols, its entropy is defined as

$$H = \sum_{i=1}^{N} P_i \, I(s_i)$$

$$H = \sum_{i=1}^{N} P_i \log_b \frac{1}{P(s_i)}$$

Example 3 : ***Calculate the entropy of the outcomes of a fair flipping coin***

$$H = P(H)\log_2 \frac{1}{P(H)} + P(T)\log_2 \frac{1}{P(T)}$$

$$H = 0.5\log_2 \frac{1}{0.5} + 0.5\log_2 \frac{1}{0.5} = 0.5 + 0.5 = 1 bit/symbol$$

Properties of Entropy

H is a positive quantity $H \geq 0$

*If all a priori probabilities are equally likely ($P_i = 1/N$ for all N symbols) then **the entropy is maximum** and given by:*

$$H = \log_b N$$

The proof:

If all a priori probabilities are equally likely ($P_i = 1/N$ for all N symbols)

$$H = -\sum_{i=1}^{N} P_i \log_2 (P_i) = -(1/N)\sum_{i=1}^{N} \log_2 (1/N)$$

$$= -(1/N)[N \log_b (1/N)]$$

$$= -\log_2 (1/N) = \log_2 N$$

$$\therefore 0 \leq H \leq \log_b N$$

i.e., you need $\log_2 N$ bits to represent a variable that can take one of N values if N is a power of 2
If these values are equally probable, the entropy is equal to certain number of bits

If one of the events is more probable than others, observation of that event is less informative

Conversely, rarer events provide more information when observed. Since observation of less probable events occurs more rarely, the net effect is that the entropy received from non uniformly distributed data is less than $\log_2 N$

Entropy is zero when one outcome is certain, so entropy refers to disorder or uncertainty of a message

***According to Shannon**, the entropy is the average of the information contained in each message of the source, irrespective the meaning of the message*

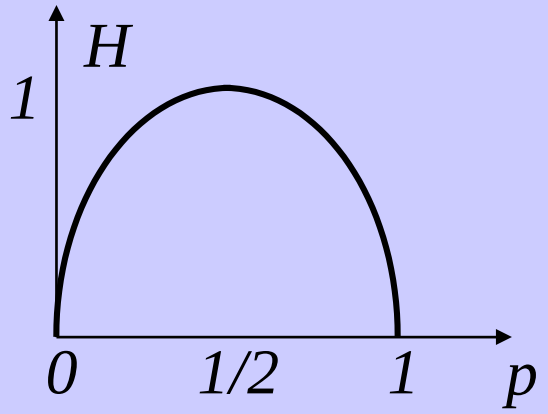
Example 4: *Find and plot the entropy of the binary code in which the probability of occurrence for the symbol 1 is P and for the symbol 0 is 1-P*

$$H = -\sum_{i=1}^{2} P_i \log_2 P_i = -P\log_2 P - (1-P)\log_2 (1-P)$$

$v \log v \rightarrow 0 \rightarrow$ as $v \rightarrow 0$

$$P = 0 \Rightarrow H = 0 \text{ bit/symbol}$$

$$P = 1 \Rightarrow H = 0 \text{ bit/symbol}$$



$$P = \frac{1}{2} \Rightarrow H = -\frac{1}{2}\log_2 \frac{1}{2} - \frac{1}{2}\log_2 \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \text{ bit/symbol}$$

$$P = \frac{1}{4} \Rightarrow H = -\frac{1}{4}\log_2 \frac{1}{4} - \frac{3}{4}\log_2 \frac{3}{4} = 0.8113 \text{ bits/symbol}$$

Example 5: *Calculate the average information in bits/character in English assuming each letter is equally likely*

$$H = \sum_{i=1}^{26} \log_2 N = \log_2 26 = 4.7 \text{ bits/character}$$

Exercise 2: For Uniform distribution P(xi)= 1/N find the average information

Solutions:

$$H(X) = \sum_{i=1}^{N} \frac{1}{N} I(x_i) = \frac{1}{N} \sum_{i=1}^{N} -\log_2 \frac{1}{N} I(x_i) = \frac{1}{N} * N * \log_2 N$$

$$= \log_2 N$$

Homework 2: Consider source transmitting six symbols with probability as given :

A	(1/2)	D	(1/16)
B	(1/32)	E	(1/4)
C	(1/8)	F	(1/32)

Find the average information or entropy.

Homework 3: A source sending 2 symbols. A and B if the $(H x_i) = 00.66$ and self information is 0.3 find the probability of A and B.

The information rate is represented by R and it is represented in average number of bits of information per second. And is given as,

Information Rate : R = rH

Information rate R it is calculated as follows:

$$R = \left(r \text{ in } \frac{symbol}{second} \right) * \left(H \text{ in } \frac{bits}{symbol} \right) = \frac{bits}{second}$$

Example 6: **A PCM source transmits four samples (messages) with a rate 2 samples / second. The probabilities of occurrence of these 4 samples (messages) are p1 = p4 = 1/8 and p2 = p3 = 3/8. Find out the information rate of the source.**

Solution:

$$H = p_{11} \log_{22}\left(\frac{11}{p_{11}}\right) + p_{22} \log_{22}\left(\frac{11}{p_{22}}\right) + p_{33} \log_{22}\left(\frac{11}{p_{33}}\right) + p_{44} \log_{22}\left(\frac{11}{p_{44}}\right)$$

$$H = \frac{11}{88} \log_{22}(88) + \frac{33}{88} \log_{2}\left(\frac{33}{88}\right) + \frac{33}{88} \log_{2}\left(\frac{33}{88}\right) + \frac{11}{88} \log_{22}(88) = 11.88 \frac{bits}{message}$$

$$R = rH = 22B \frac{message}{second} * 11.88 \frac{bits}{message} = 33.66 \frac{bits}{second}$$

In the example we discussed above, there are four samples (levels). Those four levels can be coded using binary PCM as shown down in the table:

Message or level	Probability	Binary digits
Q1	1/8	00
Q2	3/8	01
Q3	3/8	10
Q4	1/8	11

Since one bit is capable of conveying 1 bit of information, the above coding scheme is capable of conveying 4 bits of information per second. But in example, we have obtained that we are transmitting 3.6 bits of information per second. This shows that the information carrying ability of binary PCM is not completely utilized by the transmission scheme discussed in example . This situation is improved in the next example.

Example 7: *In the transmission scheme of example 6, calculate the information rate if all messages are equally likely.*

Solution: *Since they are equally likely, their probabilities p1 = p2 = p3 = p4 = 1/4*

$$H = log_{22}(44) = 22\ bits/message$$

$$R = rH = 22\frac{message}{second} * 22\ \frac{bits}{message} = 44\frac{bits}{second}$$

Just before this example we have seen that a binary coded PCM with 2 bits per message is capable of conveying 4 bits of information per second. This has been made possible since all the messages are equally likely. Thus with binary PCM coding the maximum information rate is achieved if all messages are equally likely.

The joint entropy represents the amount of information needed on average to specify the value of two discrete random variables

The entropy of the pairing (X,Y)

$$H(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} p(x_i, y_j) \log_2 p(x_i, y_j)$$

***Example 8**: Let X represent whether it is sunny or rainy in a particular town on a given day. Let Y represent whether it is above 70 degrees or below seventy degrees. Compute the entropy of the joint distribution P(X, Y) given by*

P(sunny, hot) = 1/2
P(sunny, cool) = 1/4
P(rainy, hot) = 1/4
P(rainy, cool) = 0

Answer:

$H(X, Y) = -[1/2 \log 1/2 + 1/4 \log 1/4 + 1/4 \log 1/4 + 0 \log 0]$

$= -[-1/2 + -1/2 + -1/2] = \frac{33}{22}$ *bits/symbol*

Homework 4 xxxx *For discrete memory channel the joint probability is tabulated as :.*

$$P(x,y) = \begin{bmatrix} 00.22 & 00.11 & 00.33 \\ 00.0011 & 00.0033 & 00.0000 \end{bmatrix}$$

Find H(X), H(Y), and H(X,Y)

Homework 5: *Two random variables have joint probability distribution p(x, y) given*

p(x, y)		y		
		0	1	2
x	0	3/24	2/24	1/24
	1	2/24	5/24	2/24
	2	6/24	1/24	2/24

Find E(X), E(Y), H(X), H(Y), and H(X,Y)

Given a pair of random variables (X, Y), the conditional entropy H(X/Y) is defined as

$$= \quad H(XX/Y) = \sum_{jj=11}^{n} \sum_{ii=11}^{n} p\,(x_{ii}, y_{jj})\, log_{22} p(x_{ii}/y_{jj})$$

$$= \quad H(Y/XX) = \sum_{jj=11}^{n} \sum_{ii=11}^{n} p\,(x_{ii}, y_{jj})\, log_{22} p(y_{jj}/x_{ii})$$

Example 9: *For discrete memory channel the joint probability is tabulated as:.*

p(x, y)	y=0	y=1
x=0	1/2	1/4
x=1	0	1/4

Find joint and conditional entropy $H(Y \mid X)$

Answer:

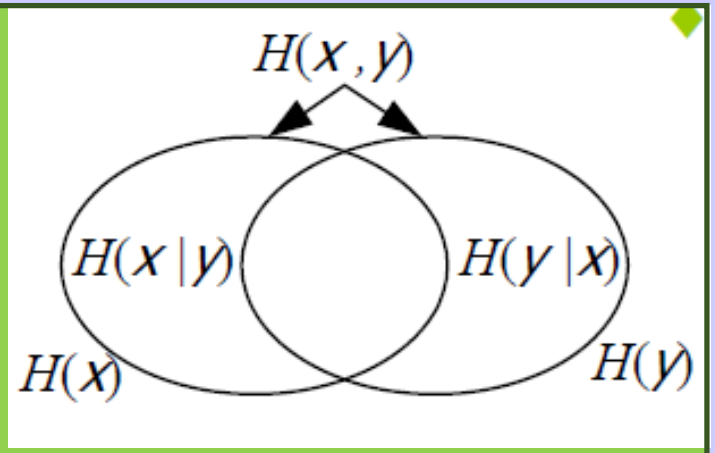
H(X,Y)= -1/2log(1/2) – 1/4log(1/4) – 0log(0) – 1/4log(1/4) = 1.5 bits/symbol

H(Y/X)= -1/2log2/3-1/4log1/3-0log0-1/4log1 = 0.689 bits/symbol

Chain Rule:

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

H (Y/X) is the average additional information in Y when you know X



Example 10: Find joint and conditional entropy

p(x, y)	y=0	y=1
x=0	1/2	1/4
x=1	0	1/4

Answer:

$H(X/Y) = H(X,Y) - H(X) = H(1/2,1/4,0,1/4) - H(3/4,1/4) = 0.6989$

Homework 6: *A transmitter produces three symbols A, B, C which are related with joint probabilities as shown in the table below.*

k	P(k)
A	11/30
B	7/12
C	1/20

P(j/k)		J		
		A	B	C
k	A	0	4/5	1/5
	B	1/2	1/2	0
	C	1/2	2/5	1/10

Calculated the joint probabilities, the average entropy of a given symbols..

Mutual Information $I(X;Y)$:consider the set of symbols $x_{11}, x_{22},.., x_n$. The source may produce $y_{11}, y_{22},.., y_n$. Theoretically, if the noise and jamming is zero then the set x = set y and n = m. however , due to noise and jamming there will be a conditional probability of $p(y_{jj}/x_{jj})$.

This the statistical average of all the pairs $I(x_{ii}, y_{jj})$, i= 1,2,…, n, j= 1,2, …,m .

$$I(X,Y) = \sum_{jj=11}^{m}\sum_{ii=11}^{n} p(x_{ii}, y_{jj})\, I(x_{ii}, y_{jj})$$

$$= \sum_{jj=11}^{m}\sum_{ii=11}^{n} p(x_{ii}, y_{jj})\, log_{22} \frac{p(x_{ii}/y_{jj})}{p(x_{ii})}$$

$$= \sum_{ii=11}^{m}\sum_{ii=11}^{n} p(x_{ii}, y_{jj})\, log_{22} \frac{p(y_{jj}/x_{ii})}{p(y_{ii})}$$ **bits**

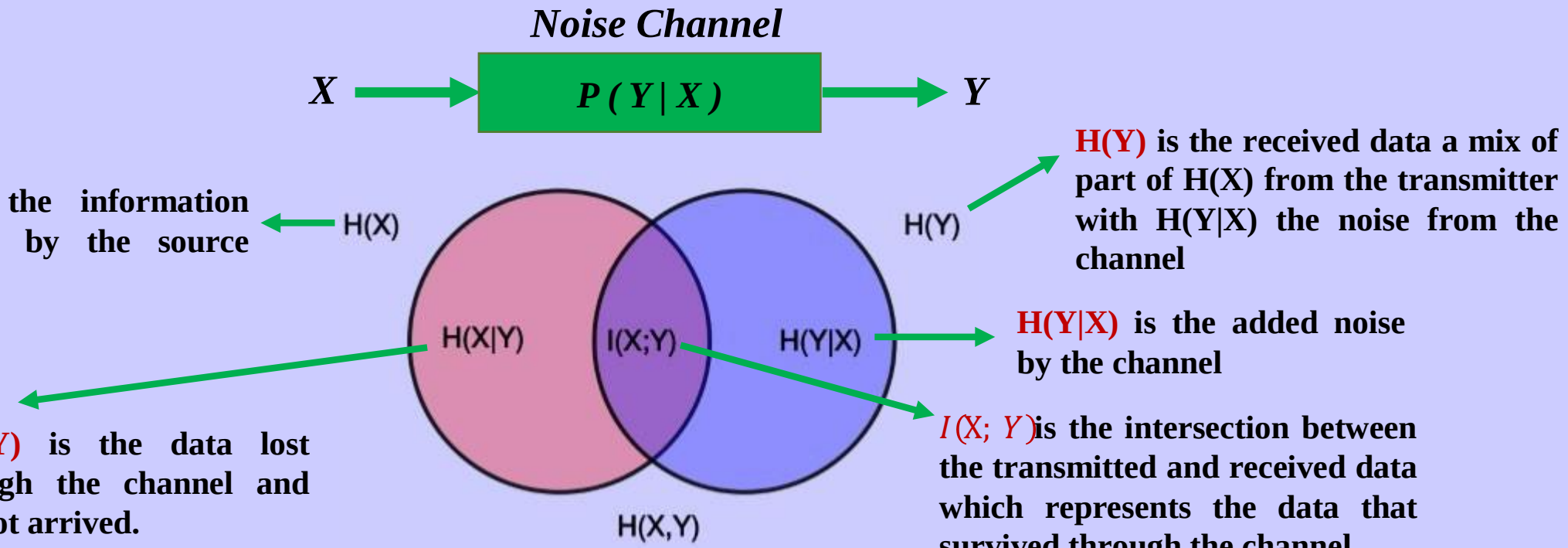
T_x $\begin{bmatrix} x_{11} \\ x_{22} \\ . \\ . \\ . \\ x_n \end{bmatrix}$ **Noise channel** $\begin{bmatrix} y_{11} \\ y_{22} \\ . \\ . \\ . \\ y_n \end{bmatrix}$ T_y

Homework 7: Prove that

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

Mutual Information $I(X;Y)$: *It is a measure of mutual dependence between variables; it quantifies the amount of information obtain about random variable from the other.*

Example 11: For X data transmitted and received as Y after passing through the channel.



H(X) is the information generated by the source then sent

H(Y) is the received data a mix of part of H(X) from the transmitter with H(Y|X) the noise from the channel

H(Y|X) is the added noise by the channel

H(X|Y) is the data lost through the channel and did not arrived.

$I(X;Y)$ is the intersection between the transmitted and received data which represents the data that survived through the channel.

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

Example 12: *of joint entropy. Let p(x, y) be given by*

$$\begin{array}{c} \\ X_{11} \\ X_{22} \\ X_{33} \end{array} \begin{array}{c} \begin{array}{cc} Y_{11} & Y_{22} \end{array} \\ \begin{bmatrix} 00.00 & 00.2200 \\ 00 & 00.112200 \\ 00.00662200 & 00.00662200 \end{bmatrix} \end{array}$$

Find

(a) The marginal entropies [H(X), H(Y)].

(b) The system entropy H(X, Y).

(c)The noise and losses entropies H(Y | X).

(d) The mutually information between ($x_{11} and y_{22}$).

(e) The transformation.

(f) Draw a channel model.

Solution:

a. From p(X, Y)

P(x) = [0.75 0.125 0.125]

P(y) = [0.5625 0.4375]

$$H(X) = -\sum_{ii=11}^{N} p(x_{ii}) \log_{22} p(x_{(i)}) \quad , \quad H(Y) = -\sum_{jj=11}^{M} p(Y_{jj}) \log_{22} p(Y_{(jj)})$$

H(X) = -[0.75$\log_{22}$ 0.75+ 0.125 $\log_{22}$ 0.125+0.125$\log_{22}$0.125] = 1.06127 bits/ symbol

H(Y) = -[0.5662200$\log_{22}$ 0.5662200 + 0.44334400$\log_{22}$ 0.44334400] = 0.9887 bits/ symbol

b. $H(X,Y) = -\sum_{jj=11}^{m}\sum_{ii=11}^{n} p(X_{ii}, Y_{jj}) \log_{22} p(X_{ii}, Y_{jj})$

$H(X,Y) = 1.875$ bits/symbol

c. $H(Y/X) = H(X,Y) - H(X) = 1.875 - 1.06127 = 0.81373$ bits/symbol

$H(X/Y) = H(X,Y) - H(Y) = 1.875 - 0.9887 = 0.8863$

d. $I(X_{11}, Y_{2}) = \log_{22} \frac{p(X_{ii},Y_{jj})}{p(X_{11})}$, $I(X_{11}, Y_{2}) = \log_{22} \frac{p(X_{11}/Y_{22})}{p(X_{ii})}$ but $p(X_{11}/Y_{22}) = \frac{p(X_{11},Y_{22})}{p(Y_{22})}$

Then $I(X_{11}, Y_{22}) = \log_{22} \frac{p(X_{11},Y_{22})}{p(X_{11}).p(Y_{22})} = \log_{22} \frac{0.25}{0.75*0.4375} = -0.3923$ bits, that means Y_{22} gives ambiguity about X_{11}

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

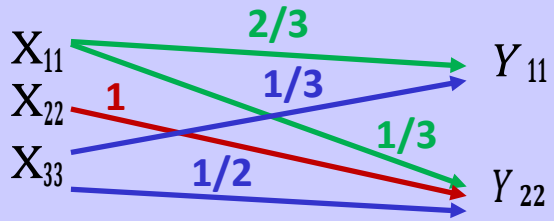
$= 1.06127 - 0.8863 = 0.1749$ bits/ symbol

f. To draw channel model we must find p(y/x) matrix from p(x, y)

$$p(Y/X) = -\sum_{jj=11}^{N} (X_{ii}, Y) / p(X_{ii})$$

$$\begin{array}{c} \\ X_{11} \\ X_{22} \\ X_{33} \end{array} \begin{array}{c} \begin{array}{cc} Y_{11} & Y_{22} \end{array} \\ \begin{bmatrix} (00.00)/(00.4400) & (00.2200)/(00.4400) \\ 00 & (00.112200)/(00.112200) \\ (00.00662200)/(00.112200 & (00.00662200)/(00.112200) \end{bmatrix} \end{array} = \begin{bmatrix} 22 & 11 \\ \overline{33} & \overline{33} \\ 00 & 11 \\ 11 & 11 \\ \overline{22} & \overline{22} \end{bmatrix}$$

Unit row summation

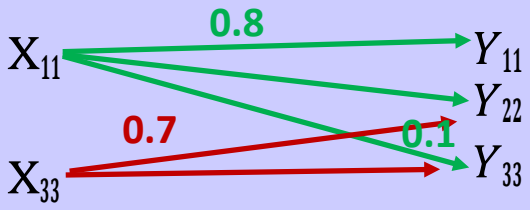


Homework 8: **For the following channel model**

p(X11)=0.6

Find :

1. ***H(X)***
2. ***H(Y)***
3. ***Noise and losses Entropy***



***Exercise 3**: of joint entropy. Let p(x, y) be given by*

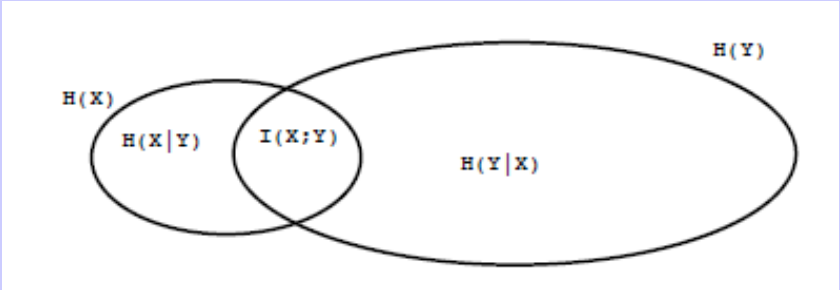
p(X, Y)		y	
		0	1
x	0	1/3	1/3
	1	0	1/3

Find

(a) H(X),H(Y).
(b) H(X | Y),H(Y | X).
(c) H(X, Y).
(d) H(Y) − H(Y | X).
(e) I(X; Y) .
(f) Draw a Venn diagram for the quantities in (a) through (e).

Solution:

(a) H(X) =2/3 log 3/2 +1/3 log 3= 0.918 bits = H(Y).
(b) H(X|Y)=1/3H(X|Y=0)+2/3H(X|Y=1)=0.667 bits = H (Y|X).
*(c) H(X,Y) = 3*1/3 log 3 = 1.5858 bits.*
(d) H(Y)-H(Y|X)=0.251 bits.
(e) I(X;Y)=H(Y)-H(Y|X)=0.251 bits
(f) Venn diagram to illustrate the relationships of entropy and relative entropy



Ministry of Higher Education and Scientific Research

Al-Furat Al- Awsat Technical University

Engineering Technical college / Najaf

Communication Engineering Department

Information Technology (CE231)

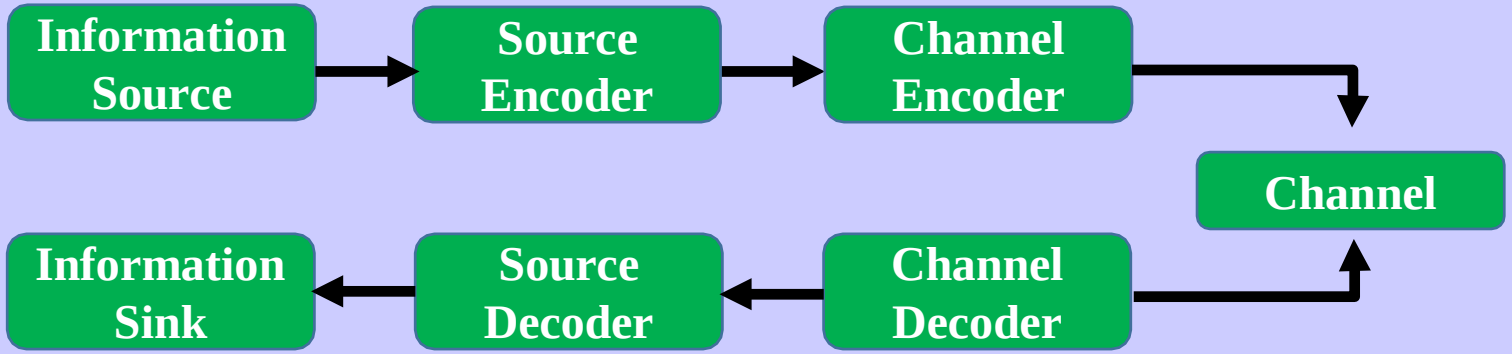
2nd Class 2018/2019

Lecturer Ali M. Alsahlany

Lecture Outlines :

- **Introduction**
- **Source coding of discrete sources**
- **Fixed Length Code**
- **Variable Length Code**
- **Minimum Code Length**
- **Code Efficiency**
- **Average Code Length**
- **Requirements for a useful symbol code**

Introduction: A general block diagram of a point-to-point digital communication system given in Figure below. The source encoder converts the sequence of symbols from the source to a sequence of binary digits, preferably using as few binary digits per symbol as possible. The source decoder performs the inverse operation.



- *Discrete sources*

The output of a discrete source is a sequence of symbols from a known discrete alphabet X. This alphabet could be the alphanumeric characters, the characters on a computer keyboard, English letters, Chinese characters, the symbols in sheet music (arranged in some systematic fashion), binary digits, etc.

- *Analog waveform sources*

The output of an analog source, in the simplest case, is an analog real waveform, representing, for example, a speech waveform. The word analog is used to emphasize that the waveform can be arbitrary and is not restricted to taking on amplitudes from some discrete set of values.

- *Simple Model of a source (Called a DISCRETE MEMORYLESS SOURCE OR DMS)*

- *Alphabet size of 4 (A, B, C, D)*
- $P(A) = p_{11}$, $P(B) = p_{22}$, $P(C) = p_{33}$, $P(D) = p_{44}$

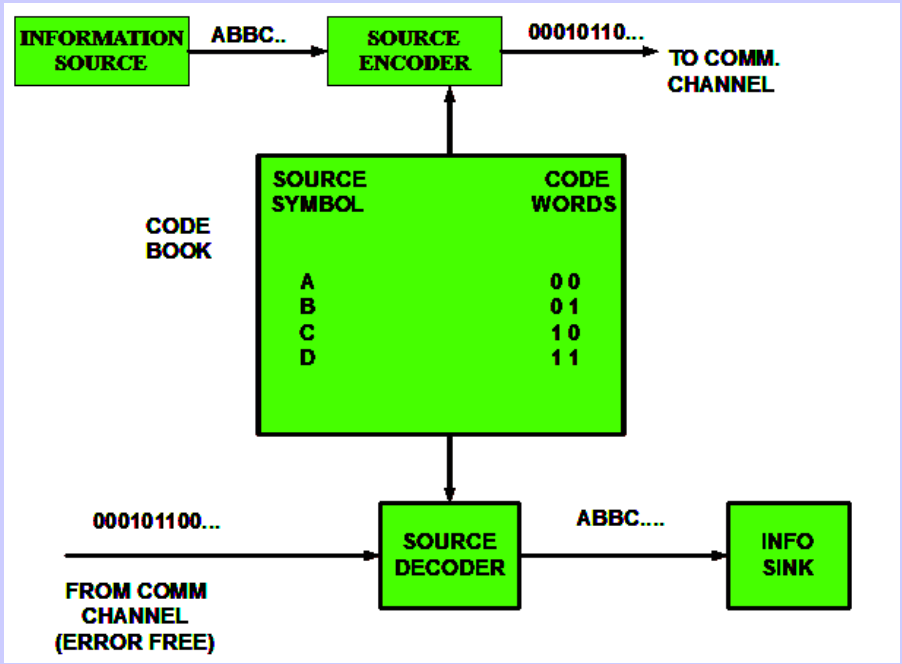
Self Information: *Is a function which measures the amount of information after observing the symbol A* $I(A) = \log_2 \frac{1}{P(A)}, bit$

Entropy: *It is the average number of bits per symbol required to describe a source* $H = \sum_{i=1}^{N} P_i \, I(s_i) \quad , bit / symbol$

Simplest Code

A → 00
B → 01
C → 10
D → 11

Symbol (SS_{kk}): A
Symbol probability: P(A)
Codeword (c_k): 00
Codeword length (l): 2



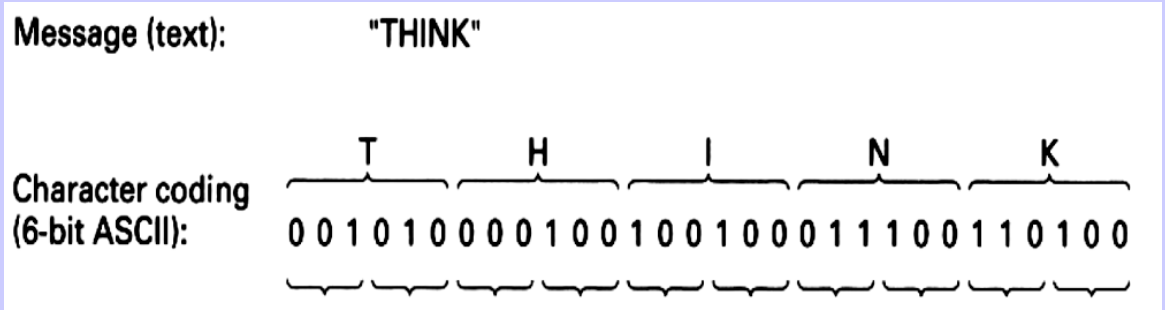
Coding:

- *Basic idea is to use as few binary digits as possible and still be able to recover the information exactly*

- *Sometimes the output of the source decoder must be an exact {replica of the information (e.g. computer data) — called NOISELESS CODING*
- *Other times the output of the source decoder can be approximately equal to the information (e.g. music, tv, speech) — called CODING WITH DISTORTION*

Fixed Length Code

- *It assigns a unique binary sequence to each input alphabet character such as 5, 6, or 7 bits*



Assume N is the number of symbols

- *If N is a power of 2, the fixed codeword length* $R = l = \log_2 N$
- *If N is not a power of 2, the fixed codeword length* $R = l = \log_2 N \uparrow$

$$H \le log_{22} N \ , \qquad l \ge H$$

$l = H$ *(If the symbols have the same probability)*

Variable Length Code

- When the source symbols are not equally probable, a more efficient encoding method is to use variable length code words
- A significant amount of data compression can be realized when there is a wide differences in probabilities of the symbols. To achieve this compression, there must also be a sufficiently large number of symbols

Ex: Morse Code

The codewords of letters that occur more frequently are shorter than those for letters that occur less frequently

A	.-	M	--	Y	-.--	6	-....
B	-...	N	-.	Z	--..	7	--...
C	-.-.	O	---	Ä	.-.-	8	---..
D	-..	P	.--.	Ö	---.	9	----.
E	.	Q	--.-	Ü	..--	.	.-.-.-
F	..-.	R	.-.	Ch	----	,	--..--
G	--.	S	...	0	-----	?	..--..
H	....	T	-	1	.----	!	..--.

- The variable length codewords should depend on the probability

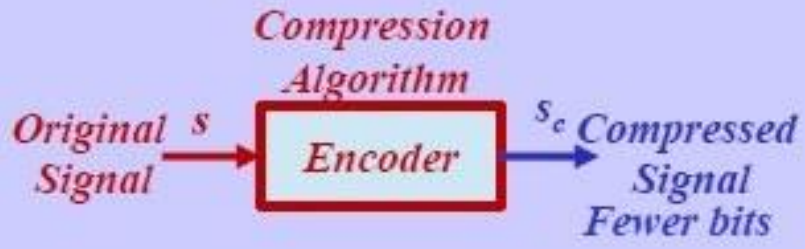
Minimum Code Length

To be efficient one using knowledge of the statistics of the source such that:

- *Frequent source symbols should be assigned SHORT code words*
- *Rare source symbols should be assigned LONGER code words*

Average Code Length

- *Consider a source has K symbols, each symbol s_k has probability P_k, and represented by a code word C_k of length l_k bits, then*



Average code word length

$$\bar{l} = \sum_{k=1}^{K} P_k \, l_k$$

Variance of the code word length

$$Variance = \sum_{k=1}^{K} P_k \, (l_k - \bar{l})^2$$

Example 1: **Simple discrete source, Alphabet size of 4 (A,B,C,D),** $P(A) = p_{11} = P(B) = p_{22} = P(C) = p_{33} = P(D) = p_{44}$, **Calculate the average and variance of the code**

Solution:

$\bar{l} = 2(p_{11}+p_{22} + p_{33} + p_{44}) = 2$

Example 2: Calculate the average code word length for the following symbols

Symbol S	P(s)	Code
A	0.25	11
B	0.30	00
C	0.12	010
D	0.15	011
E	0.18	10

$$\bar{L} = \sum_{k=11}^{K} P_k \, l_k$$

$$\bar{L} = 0.25(2) + 0.30(2) + 0.12(3) + 0.15(3) + 0.18(2) = 2.27 \text{ bits}$$

It does not mean that we have to find a way to transmit a noninteger number of bits. Rather, it means that on average the length of the code is 2.27 bits

Minimum Possible Code Word Length

$$L_{min} = H(S)$$

The outputs of an information source cannot be represented by a source code whose average length less than the source entropy

$$L_{min} \geq H(S)$$ ***Shannon's First Theorem***

Code Efficiency

$$\eta = \frac{L_{min}}{\bar{L}} = \frac{H(S)}{\bar{L}} = \frac{Entropy}{Average\ code\ length}$$

An efficient code means $\eta \longrightarrow 1$

Compression Ratio

$$CR = \frac{Number\ of\ bits\ of\ the\ fixed\ code\ that\ represents\ the\ symbols}{Average\ code\ length\ of\ the\ variable\ length\ code}$$

Example 3: **Source has an alphabet of 26 letters (English Alphabet) with independent and identically distributed random variables: { Xi , i=1,...,26} each occurring with the same probability. Evaluate the efficiency of a fixed length binary code in which:**

a. Each letter is encoded separately into a binary sequence

b. Two letters at a time are encoded into a binary sequence

a. Each letter is encoded separately into a binary sequence

$$H = \sum_{W=11}^{2222} \log_{22} N = \sum_{W=11}^{2222} \log_{22} 2222 = 44.77 \frac{bVVts}{symbol}$$

$$R = L = \sum_{W=11}^{2222} \log_{22} N \; t = 55 \frac{bVVts}{symbole}$$

$$\eta = \frac{H(S)}{\bar{L}} = 9944.00000077\%$$

b. Two letters at a time are encoded into a binary sequence

N=26*26=676 possible sequences

$R = log_{22}$**227722 t= 10bits/2 symbol =5bits/symbol** $\eta = \frac{H(S)}{\bar{L}} = 9944.00000077$

Example 4: ***Calculate the entropy, minimum length, average code word length, and the compression ratio for the flowing source and the given two different codes for the following symbols***

Source Symbol SS_{kk}	Symbol Probability P_{kk}	Code I		Code II	
		Symbol Code Word cc_{kk}	Code Word Length l_{kk}	Symbol Code Word cc_{kk}	Code Word Length l_{kk}
SS_{00}	1/2	00	2	0	1
SS_{11}	1/4	01	2	10	2
SS_{22}	1/8	10	2	110	3
SS_{33}	1/8	11	2	1111	4

$H(s) = \sum_{k=11}^{K} P \log_{22}\left(\frac{11}{P_{\text{W}}}\right)$ = 1/2log 22 (2)+1/4log 22 (4)+ 1/8log 22 (8)+1/8log 22 (8)= 1.75 bits/symbol=$L_{m\text{VV}n}$

$\eta = \frac{H(S)}{L}$

$L = \sum_{k=11}^{K} P \; l$ k k

$\eta = \frac{11.7755}{11.7755} = 00.007755$

22

$L = \left(11 \times \frac{11}{22} + 22 \times \frac{11}{44} + 33 \times \frac{11}{00} + 44 \times \frac{11}{00}\right) = 11.007755$

$\eta = \frac{}{11.007755} = 00.99333333$

CR= 2/2=1

CR= 2/1.875=1.0666

Uniquely Decodable

A code is not uniquely decodable if two symbols have the same codeword, i.e., if $Cs_{(i)} = Cs_{(j)}$ for any $i \neq j$ or the combination of two codewords gives a third one

It allows the user to invert mapping to the original sequence of elements of the source to (decoding)

Example 5:

Source symbol	Code 1	Code 2
	Symbol Codeword	Symbol Codeword
A	00	0
B	1	1
C	00	00
D	11	11

CODE1, *the symbols A &C are assigned to the same codeword. Thus, the first requirement of a useful code is that each symbol be assigned to a unique binary sequence IS NOT VALID.*

CODE2 , *It is confusing to detect the code, Why? C is decoded as combination of (A, A), also D (B,B).*

Example 6:

$S_{(i)}$	P	Code 1	Code 2	Code 3
$S_{(11)}$	0.5	0	0	0
$S_{(22)}$	0.25	0	10	01
$S_{(33)}$	0.125	1	110	011
$S_{(44)}$	0.125	10	111	0111
	1	1.125	1.75	1.875

CODE1, the symbols S_1 &S_2 are assigned to the same codeword. S_4 is a combination of (S_3,S_1) or (S_3,S_2). It is non uniquely detectable code

CODE2 , all code words end with 0 except the last one is three 1s. The decoding rule is simple, accumulate bits until you get 0 or until you have three 1s. There is no ambiguity in this rule.

CODE3 , each codeword starts with 0, and the only time we see a 0 is in the beginning of a codeword. Therefore, the decoding rule is accumulate bits until you see 0. The bit before the 0 is the last bit of the previous codeword.

- *Based on the average length, code 1 appears to be the best code However, it is not useful because it is non detectable*

Example 6:

$S_{(i)}$	P	Code 1	Code 2	Code 3
$S_{(1)}$	0.5	0	0	0
$S_{(2)}$	0.25	0	10	01
$S_{(3)}$	0.125	1	110	011
$S_{(4)}$	0.125	10	111	0111
$\bar{L}$	1	1.125	1.75	1.875

CODE2 , is called INSTANTENOUS CODE because the decoder knows the moment a codeword is complete.

CODE3 , is called NOT INSTANTENOUS CODE because we have to wait till the beginning of the next codeword before we know that the current codeword is complete

This property (Instantaneous) is not a requirement for unique decodability, because it depends

How does the channel terminate the transmission?
-e.g. it could explicitly mark the end
-it could send only 0s after the end
-it could send random garbage after the end,...

How soon do we require a decoded symbol to be known? - e.g. "instantaneously" as soon as the codeword for the symbol is received.
-within a fixed delay of when its codeword is received
-not until the entire message has been received

Example 7:

Decode the string 01111111111111111 by code1

0101010101010101 by code2

S_i	Code 1	Code 2
S_1	0	0
S_2	01	01
S_3	11	10

Using code1

The string could be decoded either $S_1 S_3 S_3 S_3 S_3 S_3 S_3 S_3 S_3$ and 1 is left which is not a codeword (Invalid decoding)

OR the string could be decoded as $S_2 S_3 S_3 S_3 S_3 S_3 S_3 S_3$ and nothing left (Valid decoding) IT IS UNIQUE DECODABLE

Using code2

The string could be decoded either $S_1 S_3 S_3 S_3 S_3 S_3 S_3 S_3 S_3$ (Valid decoding)

OR the string could be decoded as $S_2 S_2 S_2 S_2 S_2 S_2 S_2 S_2 S_1$ and left 0 which is a codeword S_1 (Valid decoding)

IT IS NOT UNIQUE DECODABLE

Test for Unique Decodability

Suppose we have two binary codewords a and b, where, a is k bits long, b is n bits long, and k < n If the first k bits of b are identical to a, then a is called a prefix of b. The last n−k bits of b are called the dangling suffix

Example 8: *if a = 010 and b = 01011, then a is a prefix of b and the dangling suffix is 11*

Steps of Unique Decodability

1-Construct a list of all the codewords.
2-Examine all pairs of codewords to see if any codeword is a prefix of another codeword .
3-Whenever you find such a pair, add the dangling suffix to the list unless you have added the same dangling suffix to the list in a previous iteration.
4-Now repeat the procedure using this larger list.
5-Continue in this fashion until one of the following two things happens:

You get a dangling suffix that is a codeword, not uniquely decodable	There are no more unique dangling , uniquely decodable

Example 9:

Si	Code 1	Code 2
S1	0	0
S2	01	01
S3	11	10

Code1: S_1 is a prefix of S_2 ⟶ code 1 will be [0,01,11,1]
1 is a prefix of 11 but it is already added then the last version is [0,01,11,1]
the codeword 1 is not a codeword. IT IS UNIQUELY DECODABLE

Code2: S_1 is a prefix of S_2 ⟶ code 2 will be [0,01,10, 1]
In new list, 1 is a prefix for 10, the dangling suffix 0 , which is the codeword. then the last version is [0,01,10,1]
IT IS NOT UNIQUELY DECODABLE

- *The variable length codewords should DEPEND ON PROBABILITY*
- *Code should be UNIQUELY DECODABLE*

Prefix Code

- Is a code in which no codeword is the beginning of another codeword
- Is a code in which no codeword is a prefix to another codeword

Since no codeword is a prefix of the other, we will never face the possibility of a dangling suffix being a codeword. In this case, the set of dangling suffixes is the null set, and we do not have to worry about finding a dangling suffix that is identical to a codeword

A prefix code is uniquely decodable but the converse is not true

Example 10:

$S_{(i)}$	Code 1	Code 2	Code 3
$S_{(11)}$	0	0	0
$S_{(22)}$	1	10	01
$S_{(33)}$	00	110	011
$S_{(44)}$	11	111	0111
	Not Uniquely Decodable Nor Prefix Code	*Uniquely Decodable Codes*	

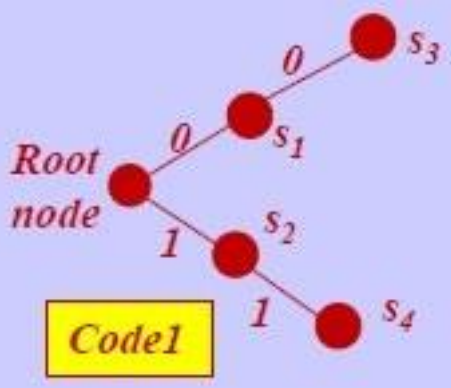
Homework 1: *Determine the prefix codes and the uniquely decodable codes of the following codes*

Symbol	Prob.	Code 1	Code 2	Code 3	Code 4
A	P[A]=1/2	1	1	0	00
B	P[B]=1/4	01	10	10	01
C	P[C]=1/8	001	100	110	10
D	P[D]=1/16	0001	1000	1110	11
E	P[E]=1/16	00001	10000	1111	110
Average Length		31/16	31/16	30/16	33/16

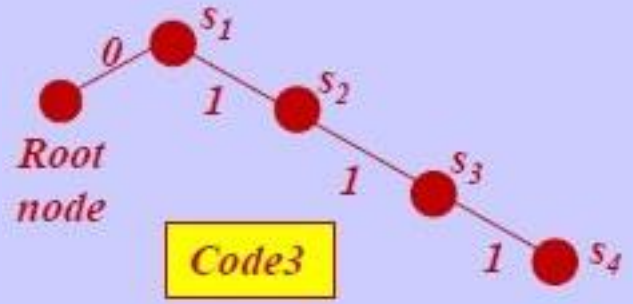
Test for Prefix Code

1. Draw the rooted binary tree corresponding to the code (Decoding Tree) Starts from a single node (the root node) and has a maximum of two possible branches at each node. One of these branches corresponds to a 1 and the other branch corresponds to a 0
2. The code for any symbol can be obtained by traversing the tree from the root to the external node corresponding to that symbol .

Example 11: Draw the rooted binary tree for the following codes



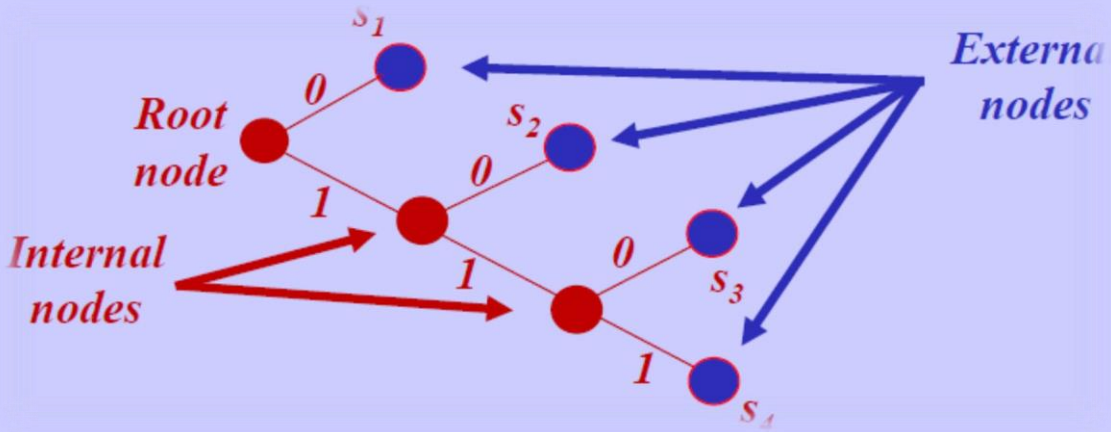
S_i	Code 1	Code 2	Code 3
S_1	0	0	0
S_2	1	10	01
S_3	00	110	011
S_4	11	111	0111



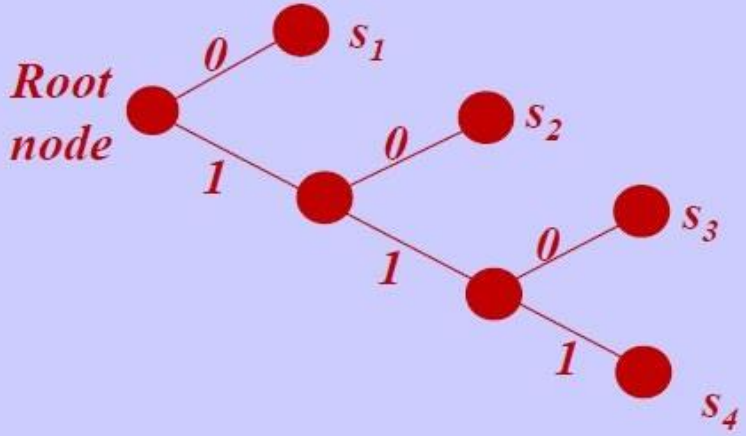
Determine the type of nodes apart from the root node :

Internal nodes that give rise to other nodes

External nodes or leaves that terminated



The prefix code has the codewords are only associated with the external nodes



Kraft-McMillan Inequality

$$\sum_{k=1}^{K} 2^{-l_k} \leq 1$$

A prefix code must satisfy the Kraft McMillan's inequality
But a code satisfies Kraft McMillan's inequality is not necessarily be a prefix code

Example 12:

s_k	Code	
	C_k	l_k
s_0	0	1
s_1	10	2
s_2	110	3
s_3	11	2

For this code $2^{-1} + 2^{-2} + 2^{-3} + 2^{-2} = 9/8$
which means that the code IS NOT A PREFIX CODE

i.e., Kraft-McMillan Inequality can determine that a given code IS NOT A PREFIX CODE

Example 13: *For this code*

s_k	Code	
	c_k	l_k
s_0	0	1
s_1	100	3
s_2	110	3
s_3	11	2

$$2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$$

Is the code a PREFIX code?
NO WHY?
s_3 is a beginning of s_2

Hence, if we have a uniquely decodable code that its codeword lengths satisfy the Kraft- McMillan inequality , then we can always find a prefix code with those codeword lengths

Thus, by restricting ourselves to prefix codes, no need for nonprefix uniquely decodable codes that have a shorter average length

Code Redundancy

Is the difference between the average length and the entropy

$$\rho = \bar{L} - H = \sum_{k=1}^{K} P_k (l_k - \log P_k(s_k))$$

Example 6: The code C_{11} = {0; 101} is a prefix code because 0 is not a prefix of 101, nor is 101 a prefix of 0.

Exercise 1: Is the code C_{22} = {0; 10; 110; 111} prefix or not?

Exercise 2: Is the code C_{33} = {00; 01; 10; 11} prefix or not?

Exercise 3: Is the code C_{44} = {1; 101} prefix or not?

Homework 2: : *Calculate the entropy, minimum length, average code word length, and the compression ratio for the flowing source*

	SS_{kk}	P_{kk}	C_{kk}	l_{kk}
C	A	1/2	0	1
	B	1/4	10	2
	C	1/8	110	3
	D	1/8	111	3

Ministry of Higher Education and Scientific Research

Al-Furat Al- Awsat Technical University

Engineering Technical college / Najaf

Communication Engineering Department

Information Technology (CE231)

2nd Class 2018/2019

Lecturer Ali M. Alsahlany

Lecture Outlines :

- **Shannon-Fano code**
- **Huffman Code**
- **Huffman vs. Shannon**

It is a technique for constructing a prefix code based on a set of symbols and their probabilities (estimated or measured)

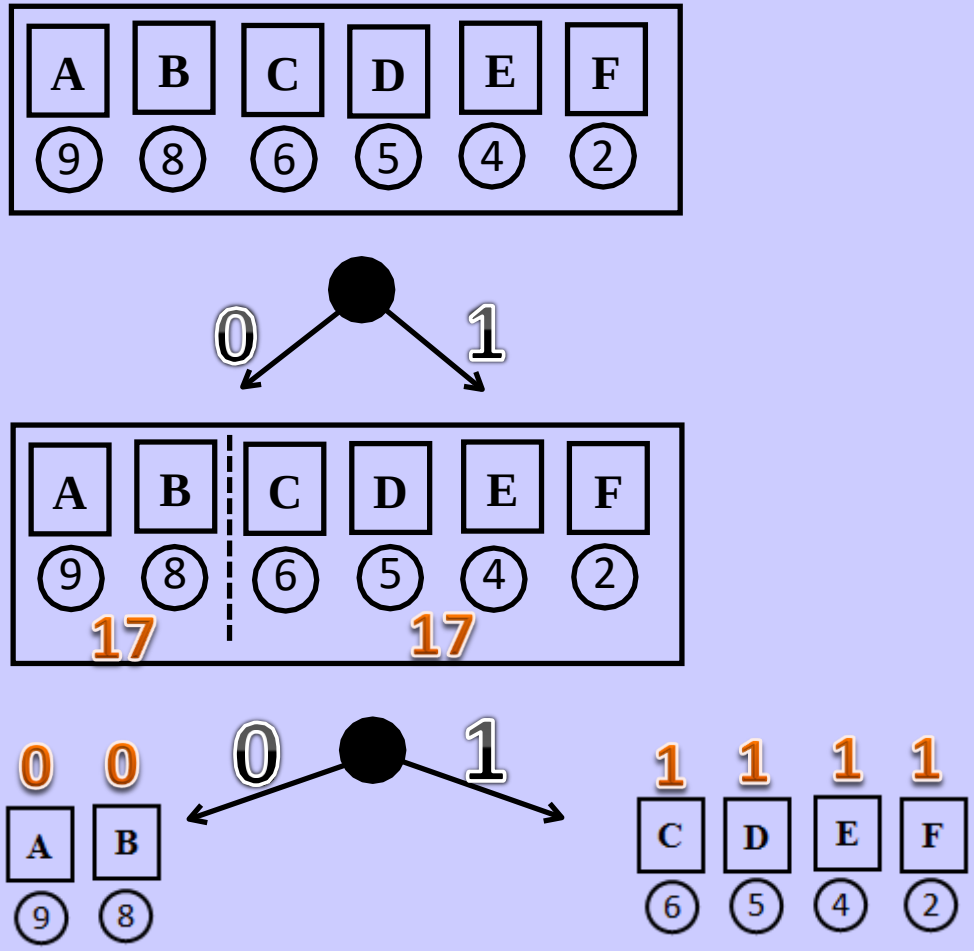
It is suboptimal in the sense that it does not achieve the lowest possible expected code word length

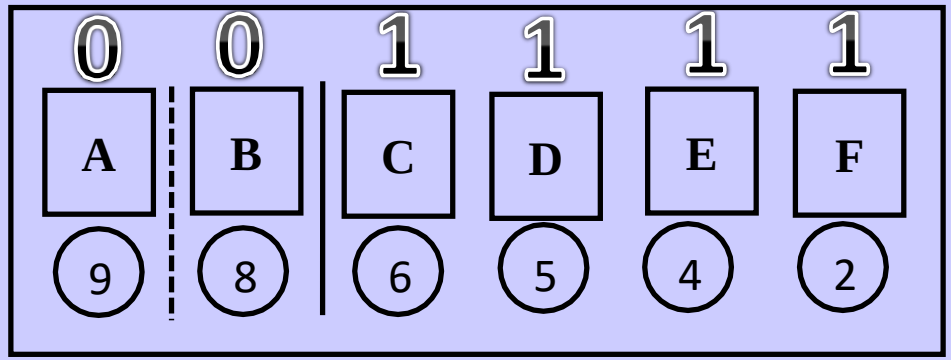
Algorithm

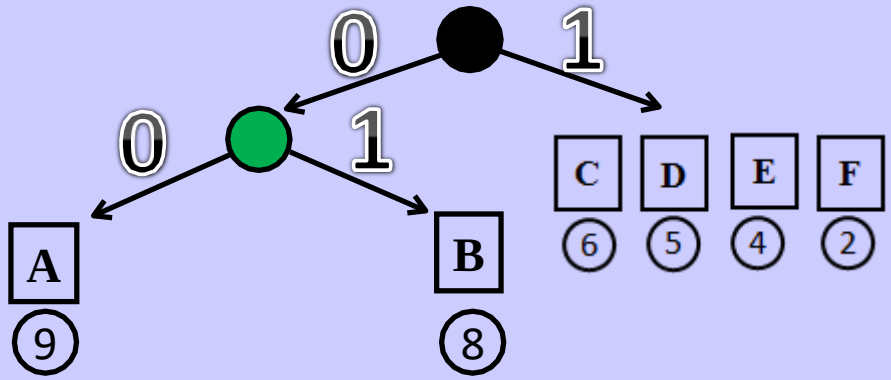
1. *Source symbols are listed in order of decreasing probability from left to right.*
2. *Divide the list into two parts, with the total probability (or frequency of occurrence) of the left part being as close to the total of the right as possible.*
3. *The left part of the list is assigned the binary digit 0, and the right part is assigned the digit 1. This means that the codes for the symbols in the first part will all start with 0, and the codes in the second part will all start with 1.*
4. *Recursively apply the steps 2 and 3 to each of the two halves, subdividing groups and adding bits to the codes until each symbol has become a corresponding code leaf on the tree.*

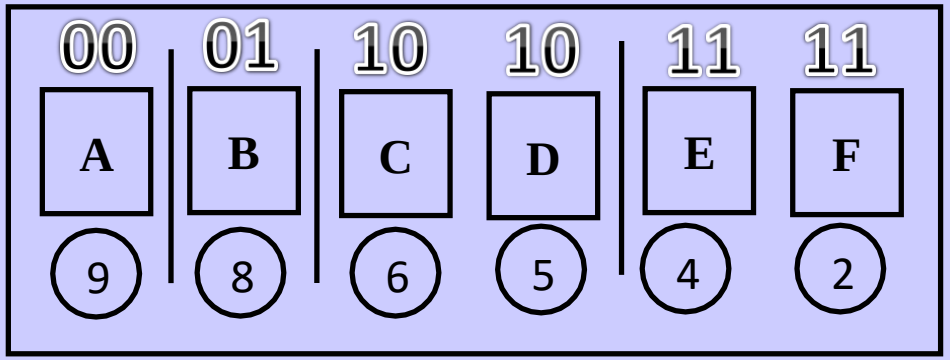
Example 1: *Assume a sequence of alphabet S={ A , B , C , D , E , F } with the following occurrence weights, {9, 8, 6, 5, 4, 2} respectively. Apply Shannon Fano coding and discuss the suboptimality.*

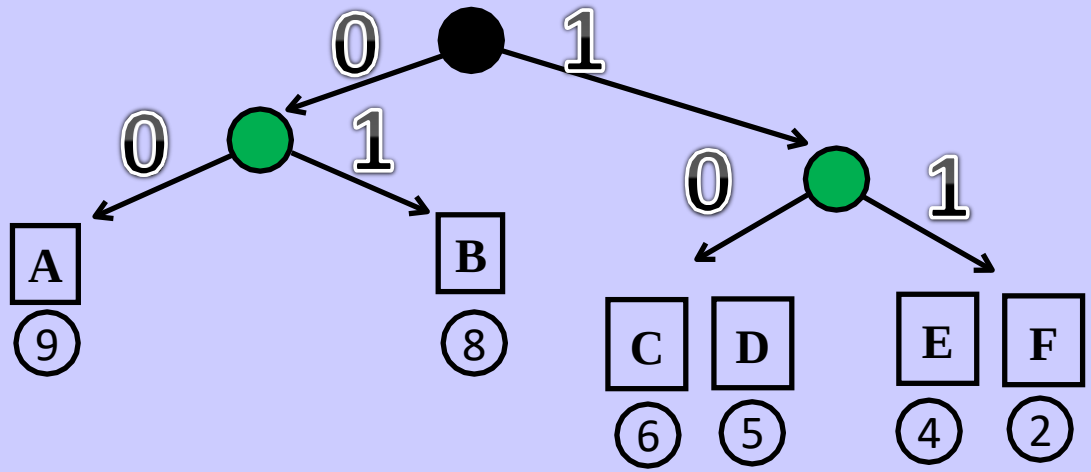
Solution:

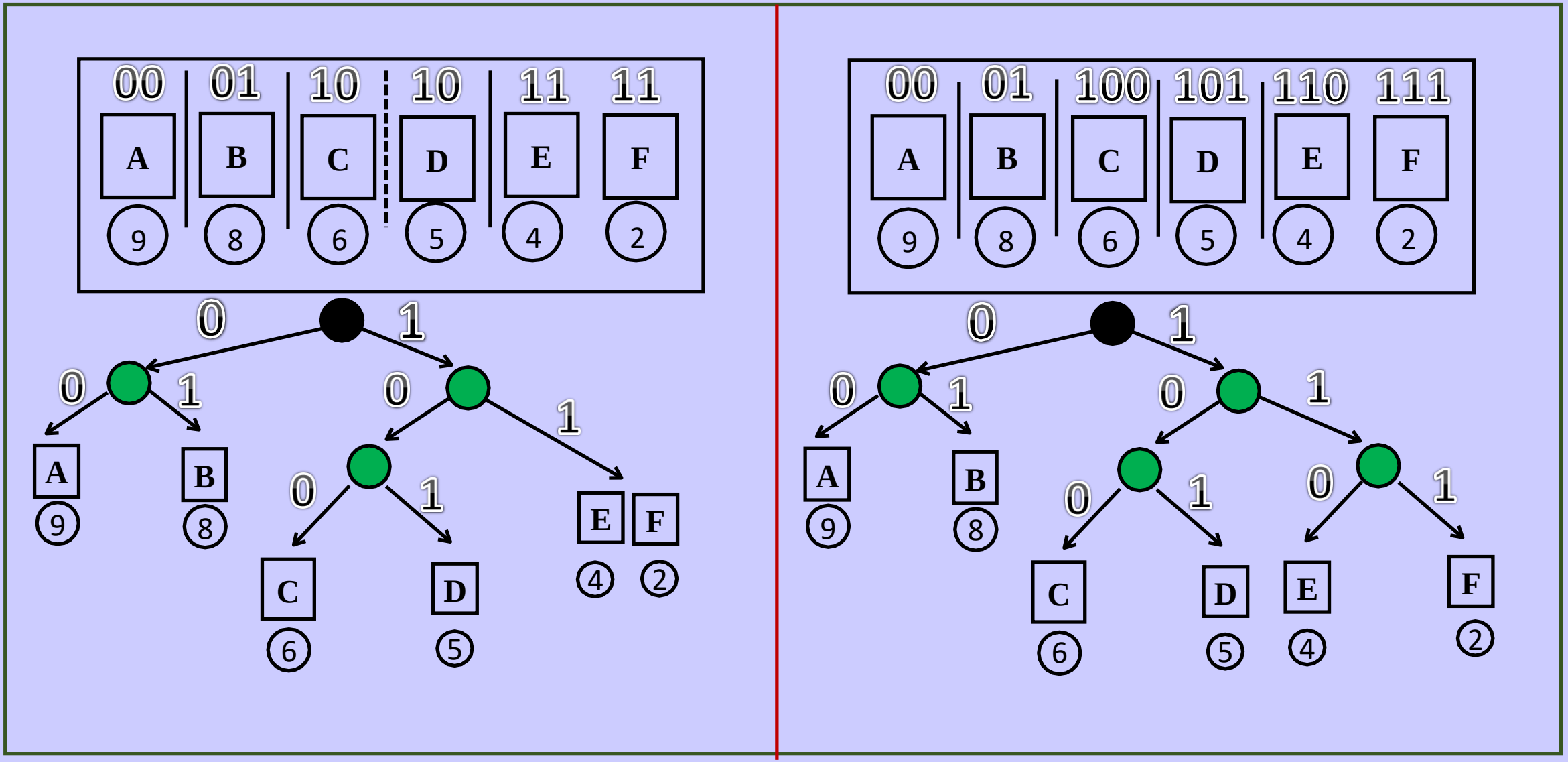



0 0 1 1 1 1
A B C D E F
9 8 6 5 4 2


0 1
0 1
A B C D E F
9 8 6 5 4 2


00 01 10 10 11 11
A B C D E F
9 8 6 5 4 2


0 1
0 1
0 1
A B C D E F
9 8 6 5 4 2


00 01 10 10 11 11
A B C D E F
9 8 6 5 4 2
0 1
0 1 0 1
A B 0 1 E F
9 8 C D 4 2
6 5
00 01 100 101 110 111
A B C D E F
9 8 6 5 4 2
0 1
0 1 0 1
A B 0 1 0 1
9 8 C D E F
6 5 4 2

Symbol	Frequency	P	CODE
A	9	0.265	00
B	8	0.235	01
C	6	0.176	100
D	5	0.147	101
E	4	0.118	110
F	2	0.059	111

$$L = \sum_{k=11}^{K} P_k \, l_k$$

L=2x(0.235+0.265)+3x(0.176+0.147+0.118+0.058)= 2.5 bits/symbol

Exercise 1: ***Construct the Shannon-Fano code and the corresponding efficiency for DMS with the following probability.***

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_{ii})$	0.5	0.15	0.15	0.08	0.08	0.02	0.01	0.01

Solution:

x	$P(x_{ii})$								Final	length
x_1	0.5	0							0	1
x_2	0.15	1	0	0					100	3
x_3	0.15	1	0	1					101	3
x_4	0.08	1	1		0				110	3
x_5	0.08	1	1		1	0			1110	4
x_6	0.02	1	1		1	1	0		11110	5
x_7	0.01	1	1		1	1	1	0	111110	6
x_8	0.01	1	1		1	1	1	1	111111	6

$$H(X) = -\sum_{ii=11}^{N} x_{ii} \log_{22} p(x_{ii}) \qquad = \qquad H = \sum_{i=1}^{N} P_i \, I(s_i)$$

$$H(X) = 00.55 \times -\log_{22}(00.55) + 22 \times (00.1155) \times -\log_{22}(00.1155) + 22 \times 00.0000 \times -\log_{22}(00.0000) +$$
$$00.0022 \times -\log_{22}(00.0022) + 22 \times 00.0011 \times \log_{22}(00.0011) = \mathbf{2.149 \approx 2.15\ bit/symbol}$$

$$\bar{L} = 00.55 \times 11 + 00.1155 \times 33 + 00.1155 \times 33 + 00.0000 \times 33 + 00.0000 \times 44 + 00.0022 \times 55 + 00.0011 \times 66 + 00.0011 \times 66$$
$$= 22.1100\ bit/symbol$$

$$\eta = \frac{L_{min}}{\bar{L}} = \frac{H(S)}{\bar{L}} = \frac{22.1155}{22.1100} = 9900\ \%$$

Homework 1: *Apply the Shannon-Fano code procedure for DMS with the following probability. Find the corresponding efficiency and redundancy.*

x	x_1	x_2	x_3	x_4	x_5	x_6
$P(x_{ii})$	0.3	0.25	0.15	0.12	0.08	0.1

Homework 2: *Apply the Shannon-Fano code procedure for DMS with the following probability. Find the corresponding efficiency and redundancy.*

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$P(x_{ii})$	1/4	1/8	1/16	1/16	1/16	1/4	1/16	1/8

Homework 3: *The text [AAAABBBCDC]*

1. *Find min code word used Shaanon-Fano.*

***Huffman procedure is based on** two observations regarding optimum prefix codes:*

1- Symbols that occur more frequently (have a higher probability of occurrence) will have shorter codewords than symbols that occur less frequently
2- The two symbols that occur least frequently will have the same length. It is commonly used for lossless compressions

- ***Huffman code** is a prefix, variable length code that can achieve the shortest average code length for a given input alphabet with probability mass function (pmf).*
- *In general, Huffman coding is a form of statistical coding as not all characters occur with the same frequency (Probability).*
- *The process of finding the optimal code was algorithmized by Huffman.*
- *A code is an optimal prefix code if it has the same average codeword length as a Huffman code for the given pmf.*

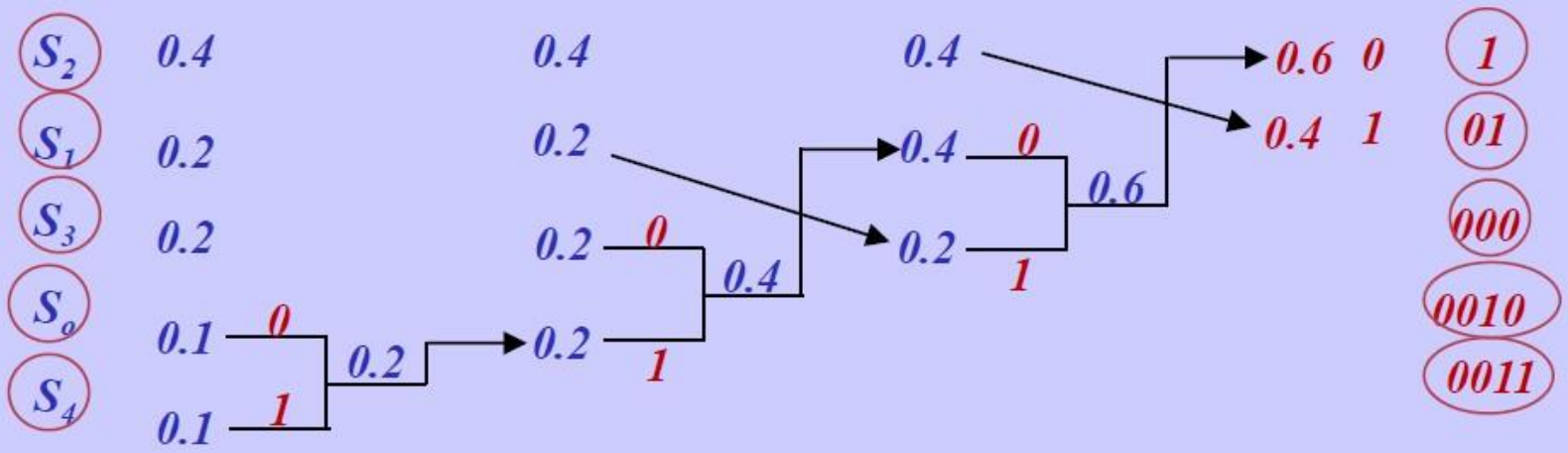
Huffman Algorithm

1. Source symbols are listed in order of decreasing probability (frequency).
2. The two source symbols of lowest probability are assigned a 0 and 1(splitting stage).
3. These two source symbols are combined into a new source symbol with probability equal to the sum of the two original probabilities (The list of source symbols and therefore source statistics is thereby reduced in size by one).
4. The probability of the new symbol is placed in the list in accordance with its value

The steps are repeated until we are left with a final list of source statistics of only two for which a 0 and a 1 are assigned

Example 2: ***Find the Huffman code for the following source given the corresponding probabilities;***

S	S_0	S_1	S_2	S_3	S_4
$P(x_{ii})$	1/10	2/10	4/10	2/10	1/10



$$H(S) = (00.44) \times log_{22}(11/00.44) + 22\text{x}(00.22)log_{22}(11/00.22) + 22\text{x}(00.11)log_{22}(11/00.11) = 22.1122119933$$

$$\text{L} = 00.44\text{x}11 + 00.22\text{x}22 + 00.22\text{x}33 + 00.11\text{x}44 + 00.11\text{x}44 = 22.22 > H(S)$$

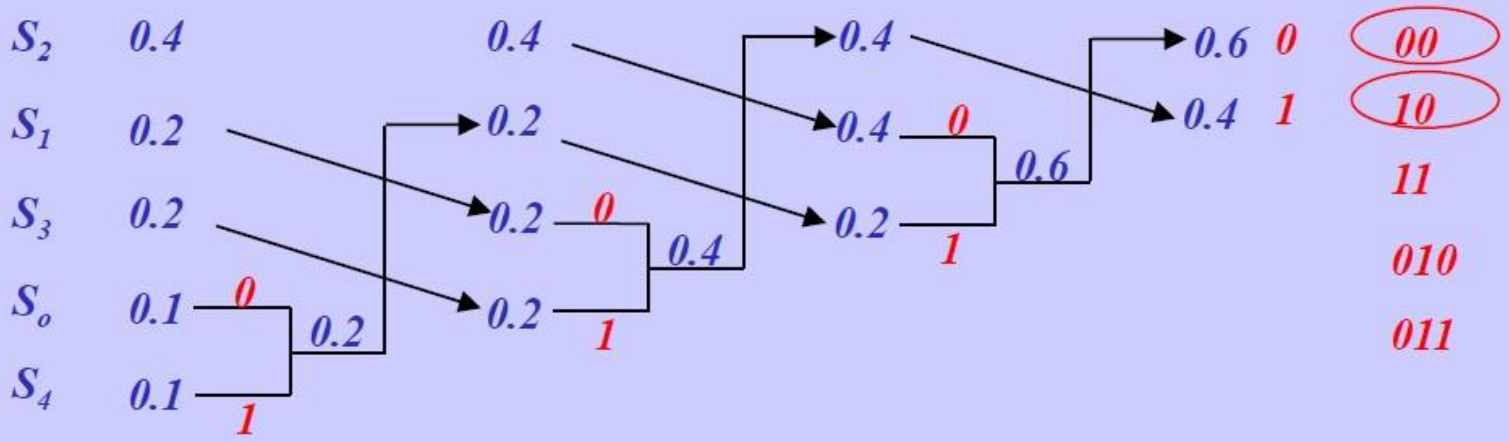
$$CCCC = \frac{33}{22.22} = 11.336644 \qquad \eta = \frac{L\_miin}{L^-} = \frac{H(S)}{L^-} = \textbf{2.12193/2.2 = 96.45\%}$$

$$\sigma^{22} = \sum_{k=11}^{K} P_k\,(l_k - \text{L})^{22} = 11.3366$$

Alternative solution:

Example 3:

S	$P(x_{ii})$
S_0	0.1
S_1	0.2
S_2	0.4
S_3	0.2
S_4	0.1



$$H(S) = (00.44) \times log_{22}(11/00.44) + 22\times(00.22)log_{22}(11/00.22) + 22\times(00.11)log_{22}(11/00.11) = 22.1122119933$$

$$\bar{L} = 00.44\times22 + 00.22\times22 + 00.22\times22 + 00.11\times33 + 00.11\times33 = 22.22 > H(S)$$

$$CCCC = \frac{33}{22.22} = 11.336644 \qquad \eta = \frac{L\_min}{\bar{L}} = \frac{H(S)}{\bar{L}} = \textbf{2.12193/2.2 = 96.45\%}$$

$$\sigma^{22} = \sum_{k=11}^{K} P_k\,(l_k - \bar{L})^{22} = 00.1166$$

- *Hence, to obtain a minimum variance Huffman code, we always put the combined symbol as high in the list as possible*

Homework 4: *Consider the following short text*

Eerie eyes seen near lake

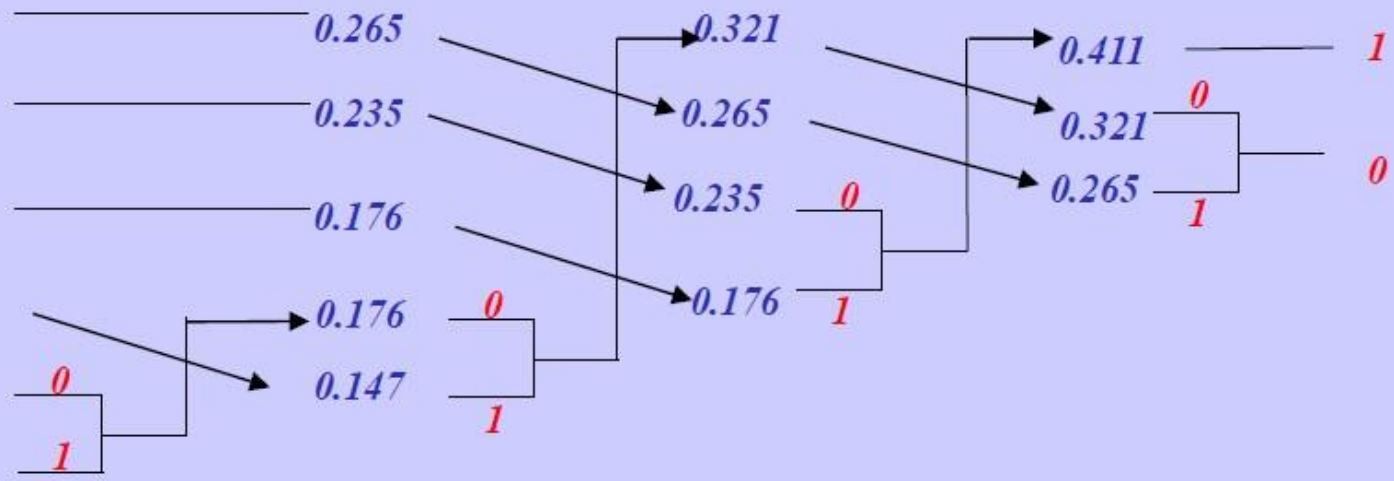
Build the Huffman code to encode the text

Homework 5: *Source transmitted symbols with probability as shown in table below. Find corresponding efficiency.*

S	$P(x_{ii})$
S_1	0.3
S_2	0.25
S_3	0.2
S_4	0.12
S_5	0.08
S_5	0.05

Huffman vs. Shannon

S	Freq.	P
A	9	0.265
B	8	0.235
C	6	0.176
D	5	0.147
E	4	0.118
F	2	0.059



Huffman Code	Shannon Code
01	00
10	01
11	100
001	101
0000	110
0001	111

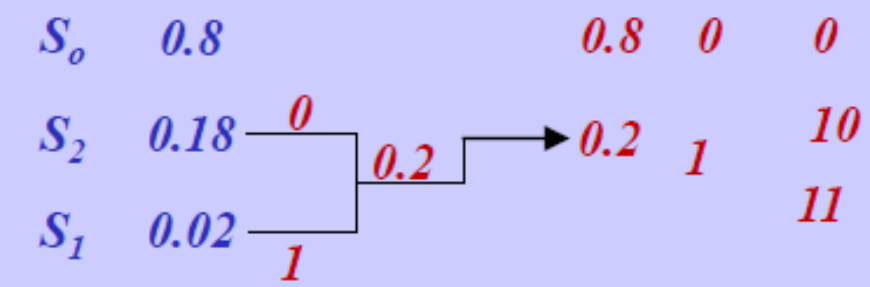
$\bar{l} = \sum_{k=11}^{K} P_k\, l_k$, $\bar{l} = 2x(0.235+0.265+0.176)+3x(0.147)+4x\,(0.118+0.058) = 2.47$ bits/symbol

The average code length is less than that of Shannon Fano code, Hence Huffman is optimum, but Shannon is suboptimal

In applications where the alphabet size is large; P_{max} is generally quite small, and the amount of deviation of the entropy from the average code length (or in terms of a percentage of the rate) is quite small. However, in cases where the alphabet is small and the probability of occurrence of the different letters is skewed, the value of P_{max} can be quite large and the Huffman code can become rather inefficient when compared to the entropy.

Ex: Find the Huffman code for the following source given the corresponding probabilities

S_o	0.8
S_1	0.02
S_2	0.18



$$H(S) = (00.00) \times log_{22}(11/00.00) + (00.0022) log_{22}(11/00.0022) + (00.1100) log_{22}(11/00.1100) = 00.001166\ biits/symbol$$

$$\bar{L} = 00.00 \times 11 + 00.0022 \times 22 + 00.1100 \times 22 = 11.22\ biits/symbol$$

There is a big difference between the average code length and the entropy (Code Redundancy),

$$\rho = \bar{L} - H(S) = 0.384\ bits/symbol$$

We can reduce the rate (average length) by grouping (blocking)symbols together which is called

Extended (Block) Huffman Code

-Consider a source S that emits a sequence of letters [$s_1, s_2, s_3, \ldots s_m$]

-Each element of the sequence is generated independently

-The entropy for this source is given by $H(S) = \sum_{ii=11}^{N} p_{(s_{ii})} log_{22} \frac{11}{p(s_{ii})}$

-One can generate a Huffman code for this source with ***rate R*** *(In data compression it is bits per symbol like the average length not bits per second as in communication) such that* $H(S) \leq R \leq H(S) + 11$

-Encode the sequence by generating one codeword for every n symbols, hence there are m_n *combinations of n symbols*

-Denote the rate for the new source as $R^{(n)}$ *, hence* $H(S^{(n)}) \leq R^{(n)} \leq H(S^{(n)})$ **+1**

$R^{(n)}$ *is the number of bits required to code n symbols together. Therefore, the number of bits required per symbol, R, is given by*

$$R \leq R^{(n)}/n \qquad \frac{H(S^{(n)})}{n} \leq \frac{R^{(n)}}{n} \leq \frac{H(S^{(n)})}{n} + \frac{1}{n}$$

Example 4: *Find the extended Huffman code for the following source given the corresponding probabilities by blocking every two symbols.*

S_1	0.8
S_2	0.02
S_3	0.18

Solution:

-m=3, n=2, the number of possible symbol pairs or the extended symbols are $33^{22}=9$

-The probability of each extended symbol is calculated by multiplying the probabilities of the original single items together

-Follow the steps of getting Huffman code, you get the code in that table

Extended Symbol	Probability Model	Huffman Code
S_1S_1	0.64	0
S_1S_2	0.016	10101
S_1S_3	0.144	11
S_2S_1	0.016	101000
S_2S_2	0.0004	10100101
S_2S_3	0.0036	1010011
S_3S_1	0.144	100
S_3S_2	0.0036	10100100
S_3S_3	0.0324	1011

$$H(S) = (00.00) \times log_{22}(11/00.00) + (00.0022)log_{22}(11/00.0022) + (00.1100)log_{22}(11/00.1100) = 00.001166\ biits/symbol$$

$$L_{ext} = 11.77222200\ \ biits/symbol$$

Each symbol in the extended alphabet corresponds to two symbols from the original alphabet

$$L_{ext} = \frac{11.77222200}{22} = 00.006611\ biits/symbol$$

Redundancy = 0.0045 bits/symbol

There is no big difference between the average code length and the entropy,

Hence, one can see that by encoding the output of the source in longer blocks of symbols can guarantee a closer rate to the entropy

But as we block more and more symbols together, the size of the alphabet grows exponentially, and the Huffman coding scheme becomes impractical

Ministry of Higher Education and Scientific Research

Al-Furat Al- Awsat Technical University

Engineering Technical college / Najaf

Communication Engineering Department

Information Technology (CE231)

2nd Class 2018/2019

Lecturer Ali M. Alsahlany

Lecture Outlines :

- Introduction
- Error detection codes
  - Error detection codes / parity codes
- Error correction codes
  - Error correction codes / Basic definitions
  - Error Correction Codes / Hamming Codes
  - Error Correction Codes / Cyclic Codes

Channel Coding: *To ensure reliable communications, techniques have been developed that allow bit errors to be detected and corrected. The process of error detection and correction involves adding extra redundant bits to the data to be transmitted. This process is generally referred to as channel coding.*

Channel coding methods fall into to two separate categories:

- ***Error detection codes:*** *only have the ability to confirm that bit error(s) has occurred, however they cannot tell you which bit was in error. To fix the error, the receiver must request a retransmission.*
- ***Error correcting codes or forward error correction (FEC) codes:*** *have the ability to detect some bit errors and fix them without requiring a retransmission.*

Block Codes **In block codes, the encoder takes in *k* information or message bits and produces a codeword of length *n* =*k* + *r*. Since *k* + *r* > *k*, the quantity r represents the number of extra bits (redundancy) added. This would be referred to as an (*n*, *k*) linear code.**

$n = k + r$

k: bbbt bbn $data$

r: $rudandanc$ bbbt

$n = codeccord$ bbbt

***Example 1**: Linear block code (7,4) with k= 4 and k + r = 7*

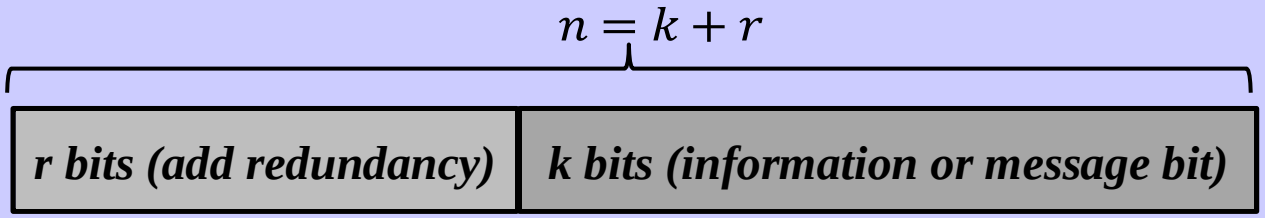
Message	Code words
(0 0 0 0)	(0 0 0 0 0 0 0)
(0 0 0 1)	(1 0 1 0 0 0 1)
(0 0 1 0)	(1 1 1 0 0 1 0)

A key point about channel coding is that there is a cost to be paid of increasing reliability. The extra n bits added by encoding result in:

- ***Larger file sizes for storage.***
- ***Higher required transmission data rates.***

This cost is represented by the code rate. The code rate R_c is the ratio of the number of information bits k to the number of bits in the codeword k + r.

$$R_c = k/n$$



2/24/2019

***Example 2*: code (8,7)**
That is mean n = 8bit, k = 7bit, and r = 8 – 7 = 1bit
$R_c = k/n = 77/88$

Simple Error Detection Codes:
Parity codes The simplest kind of error-detection code is the *parity code*. To construct an even-parity code, add a parity bit such that the total number of 1's is even.

Information bits	Even parity code	Odd parity code
000	000 **0**	000 **1**
001	001 **1**	001 **0**
010	010 **1**	010 **0**
011	011 **0**	011 **1**

Parity check code (n , k)
$R_c = k/(k + 11)$

***1-bit parity codes* can *detect* single bit errors, but they *do not* detect *2 bit errors*.**
Probability (detecting errors)= Probability (odd number of errors) and
Probability (undetecting errors)= Probability (even number of errors).

Example 3: Parity check code of k=7 bit. Calculate P (undetected error) and P (detected error)

Solution:

$$P\ (undetect\ error) = \sum_{k} C_k^n p^k (11 - p^k) = C_{22}^{88} p^{22} (11-p)^{66} + C_{44}^{88} p^{44} (11-p)^{44} + C_{66}^{88} p^{66} (11-p)^{22} + C_{88}^{88} p^{88} \approx$$

$$28x1111^{-44}$$

The probability of detected errors will be:

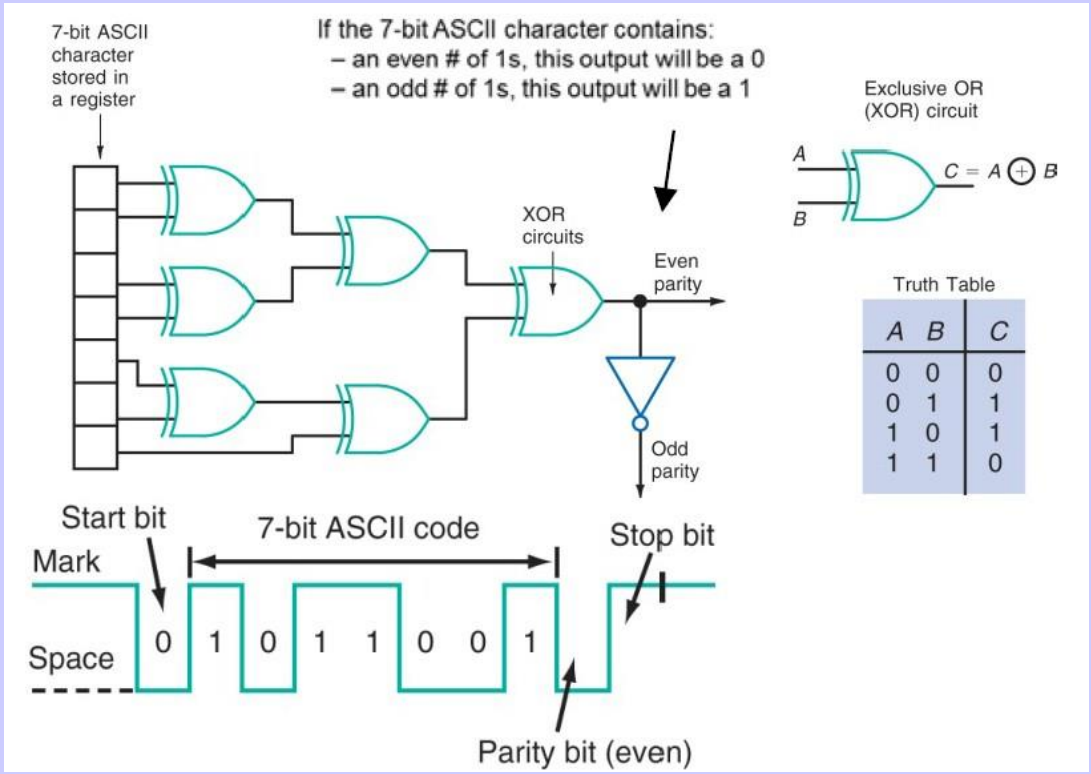
$$P\ (detect\ error) = \sum_{k} C_k^n p^k (11 - p^k) = C_{11}^{88} p^{11} (11-p)^{77} + C_{33}^{88} p^{33} (11-p)^{55} + C_{55}^{88} p^{55} (11-p)^{33} + C_{77}^{88} p^{77} (11-p) \approx$$

$$8\times1111^{-33}$$

Homework1: *assume there are n=12bits in a codeword (packet). Probability of error in a single bit transmission* $p_b = 1111^{-33}$*. Find the probability of error-detection failure.*

To implement these parity generators, simple Ex-OR gates are used at TX and RX as shown below

Information bits	Even parity code
1010111	1
0110101	0



***Example 4*: Parity cheak code (8,7) find code word if data** $I_{11} = [11111111111111], I_{22} = [11111111111111], and\ I_{33} = [11111111111111]$

Solution:

$C_{bb} = [I_{bb}: parbbty\ bbbt]$

$C_{11} = [11111111111111111]$ $C_{22} = [11111111111111111]$ $C_{22} = [11111111111111]$

***Excersise 1*: detect error in the received code word (use parity check code (8,7) then find data words.**

$C_{11} = [11111111111111111]$ $C_{22} = [11111111111111111]$ $C_{33} = [11111111111111111]$

Solution:

- ***In code*** C_{11} ***number of ones is even there is no error, data is 0101010***
- ***In code*** C_{22} ***number of ones is odd there is error***
- ***In code*** C_{33} ***number of ones is even there is no error, data is 1111011***

Systematic and nonsystematic codes: *If information bits (a 's) are unchanged in their values and position at the transmitted codeword, then this code is said to be systematic.*

Input data [D]=$[a_{11}\ a_{22}\ a_{33}\ \dots\dots\dots a_k]$,

Output systematic (n,k) codeword is [C] = $[a_{11}\ a_{22}\ a_{33}\ \dots\dots\dots a_k\ c_{11}\ c_{22}\ c_{33}\ \dots\dots\dots c_r]$

However if data bits are spread or changed at the output codeword then, the code is said to be nonsystematic:

Output nonsystematic (7,4) codeword is [C] = $[c_{22}\ a_{11}\ c_{33}\ a_{22}\ c_{11}\ a_{44}\ a_{33}]$

Hamming distance: *The ability of error detection and correction codes depends on this parameter. The Hamming distance between two codewords* c_{bb} *and* c_{jj} *is denoted by* d_{bbjj} *which is the number of bits that differ. For a binary (n,k) code with* 22^k *possible codeword then the minimum Hamming distance (HD) is the min)* (d_{bbjj}) *of course* $n \geq d_{bbjj} \geq 11$

Example 5: *Find the Hamming distance between the two codewords:*
$C_{11} = [11111111111111]$ and $C_{22} = [11111111111111]$.
Solution: *Here, the number of bits that differ is 2, hence* $d_{1122} = 22$

Homework 2: *Find the minimum Hamming distance for the 3 codewords.*
$C_{11} = [11111111111111]$, $C_{22} = [11111111111111]$, $C_{22} = [11111111111111]$

Hamming weight: *This is the number of 1's in the non zero codeword* c_{bb}. *It is denoted by* cc_{bb}. *As will be shown later, and for linear codes,* $cc_{mbbn} = HD = mbbn(d_{bbjj})$. *This simplifies the calculation of HD. As an example 5,*
Example 6: *If* $C_{11} = [11111111111111]$ *, then* $cc_{bb} = 3$, *for* $C_{22} = [11111111111111]$, *then* $cc_{bb} = 2$ *and so on.*

Linear and non Linear codes: *when the parity bits are obtained from a linear function of the k information bits then the code is said to be linear, otherwise it is a nonlinear code.*

Hamming Bound: **The purpose of Hamming bound is either**
1) to choose the number of parity bits (r) so that a certain error correction capability is obtained. Or
2) to find the error correction capability (t) if the number of parity bits (r) is known for binary codes, this is given by:

$$22^{n-k} = 22r \geq \sum_{jj=11}^{t} C_{jj}^{n}$$

where t is the number of corrected bits.

$$t = \mathbf{bb}nt[\frac{HD - 11}{22}]$$

$$no.of\ detected\ error\ bbbt = HD - 11$$

Example 7: **for a single correction code with k=4 find the no. of parity bits that should be added.**
Solution:

$22r \geq \sum_{jj=11}^{11} C_{jj}^{44+r} = C_{11}^{44+r} + C_{11}^{44+r}$ **This gives $2^r \geq 1+(4+r)$ and the minimum r is r=3**

(take minimum r to have max code efficiency). This is the (7,4) code. the code is said to be perfect code.

Homework 3: **if k=5 and up to 3 errors are to be corrected, find the no. of check bits that should be added.**

Note: If the (n,k) codewords are trans. through a channel having error prob=p_e, then prob. of decoding a correct word at the Rx for t-error correcting code will be:
P(correct words)=p(no error)+p(1 error)+……..p(t errors)
and prob(erroneous word)=1-P(correct word).

Excersise 2: *If Code of* $t = 22, n = 3311$ *find* k

Solution: $n = 3311 \quad k = ? \quad r = ?$

$$22^r \geq \sum_{jj=11}^{t} C_{jj}^{n}$$

$$22^r \geq C_{11}^{3311} + C_{11}^{3311} + C_{22}^{3311}$$

$$22^r \geq 11 + 3311 + \frac{3311 * 3311}{22}$$

$$22^r \geq 444477 \Rightarrow\Rightarrow \quad r \geq 88.4455$$

n=3311 r =44 k=22bit

Code (31,22)

$$t = 22 \Rightarrow\Rightarrow t = bbnt\left[\frac{HD-11}{22}\right] \Rightarrow\Rightarrow HD = 55$$

No. of detected error= $(HD - 11) = 44bbbt$

$$P(\ correct\ error) = \sum_{bb=11}^{t} C_{bb}^{n} P^{bb} (11 - P)^{n-bb}$$

$$P_{(\ correct\ error\ )} = C_{11}^{3311} P^{11} (11 - P)^{3311} + C_{11}^{3311} P^{11} (11 - P)^{3311} + C_{22}^{3311} P^{22} (11 - P)^{2244}$$

Example 8: Block code (7, 4) calculate no. detect and correct bit Hamming Distance (HD) and P(correct error).

Solution: $n = 77 \quad k = 44 \quad r = 77 - 44 = 33$ bbbt

$$22^{r} \geq \sum_{ii=11}^{t} C_{ii}^{n}$$

$$22^{33} \geq C_{11}^{77} + C_{11}^{77} + C_{22}^{77} \ldots\ldots$$

$$88 \geq 11 + 77 + \frac{66 * 77}{22} \ldots\ldots$$

Taken two element that mean $t = 11$

$$t = bbnt\left[\frac{HD - 11}{22}\right] \Rightarrow\Rightarrow HD = 33$$

No. of detected error $= (HD - 11) = 22$

$$P(\ correct\ error) = \sum_{bb=11}^{t} C_{bb}^{n} P^{bb} (11 - P)^{n-bb}$$

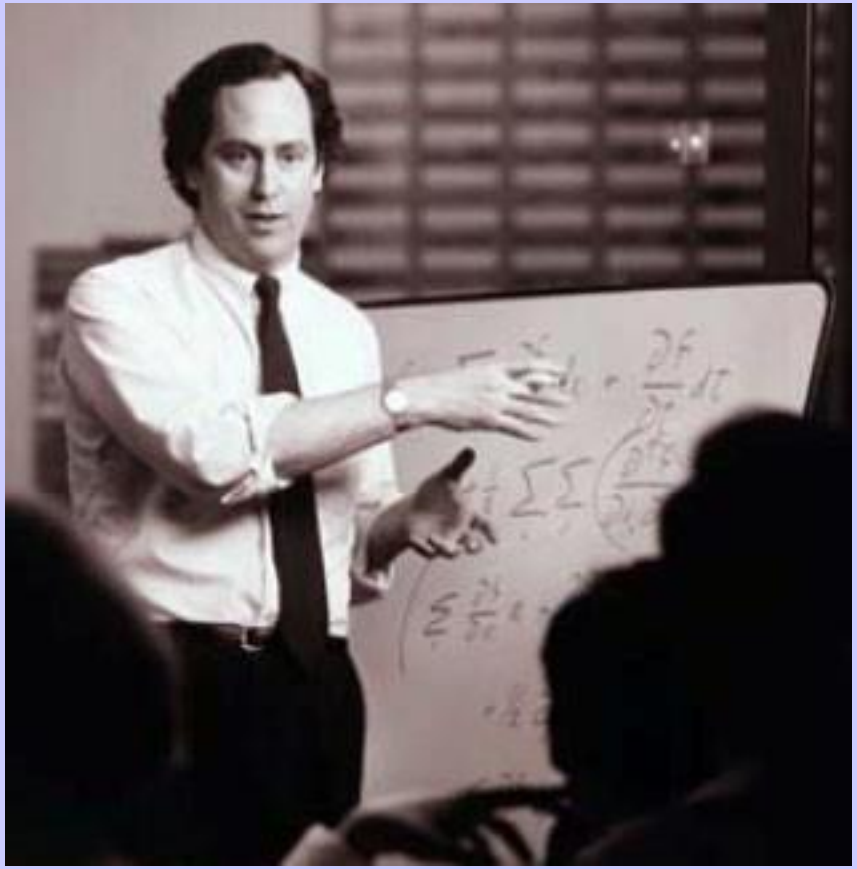
$$P(\ correct\ error\ ) = C_{11}^{77} P^{11} (11 - P)^{77} + C_{11}^{77} P^{11} (11 - P)^{66}$$

$$= 11 * 10(11 - P)^{77} + 7P(11 - P)^{66}$$

$$\text{Code rate} = \frac{k}{n} = \frac{44}{77}$$

One way to detect and correct errors is to add parity checks to the codewords:

- **If we add a parity check bit at the end of each codeword we can detect one (but not more) error per codeword.**
- **By clever use of more than one parity bits, we can actually identify where the error occurred and thus also correct errors.**
- **Designing ways to add as few parity bits as possible to correct and detect errors is a really hard problem.**



Richard W. Hamming (11.2.1915-7.1.1998)

Linear Block Codes:

Only systematic binary codes will be described. The r parity bits are obtained using a linear function of the a's data. Mathematically, this can be described by the set of equations:

$$C_1=h_{11}a_1+h_{12}a_2+h_{13}a_3+\ldots\ldots\ldots+h_{1k}a_k$$
$$C_2=h_{21}a_1+h_{22}a_2+h_{23}a_3+\ldots\ldots\ldots+h_{2k}a_k$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$
$$C_r=h_{r1}a_1+h_{r2}a_2+h_{r3}a_3+\ldots\ldots\ldots+h_{rk}a_k$$

Where + is mod-2 addition (EX-OR), product is the AND multiplication and h_{ij} coefficients are binary variables for a binary coding. The complete output codeword can be written in matrix form as:

$[C]= [D][G]$.……(1) , where:

$$[G]=\begin{bmatrix} 1 & 0 & 0 & 0 & h_{11} & h_{21} & h_{31} & . & h_{r1} \\ 0 & 1 & 0 & 0 & h_{12} & h_{22} & h_{32} & . & h_{r2} \\ 0 & 0 & 1 & 0 & . & . & . & . & . \\ 0 & 0 & 0 & 1 & h_{1k} & h_{2k} & h_{3k} & . & h_{rk} \end{bmatrix}$$

$G= [I_k : P_{kxr}]$ which is kxn matrix.

This matrix is called the generator matrix of the linear block code (LBC). Equation(1) can also be written in matrix form as:

$[H][C]^T=[0]$ ……(2)

where: $[C]=[a_1\ a_2\ a_3\ \ldots\ldots a_k\ c_1\ c_2\ c_3\ \ldots\ldots c_r]$ and [H] matrix is in fact related with [G] matrix by:

$[H]=[-P^T : I_r]$, and for binary coding this – sign drops out. This rxn [H] matrix is called the parity check matrix. As will be shown, encoding can be done either using eq(1) ([G] matrix) or eq(2) ([H] matrix), but decoding is done using [H] matrix only.

Example 9: a given binary (7,4) Hamming code with a parity check matrix:

$$G=\begin{bmatrix}1000011\\0100101\\0010110\\0001111\end{bmatrix}$$

$$G=[I_{kxk}:P_{kxr}]$$

Find: 1) no. of error correction and detect capability 2) Code rate 3) encoder circuit 4) if data [1011] find codewords.

Solution: n=7 k=4 r=3

$$22^{r} \geq \sum_{jj=11}^{t} C_{jj}^{77}$$

$$t = 1\,, HD = 3$$
$$No.ooo\ detect\ error = HD - 1 = 2$$

From example 8

$$C = D \,.\, G$$

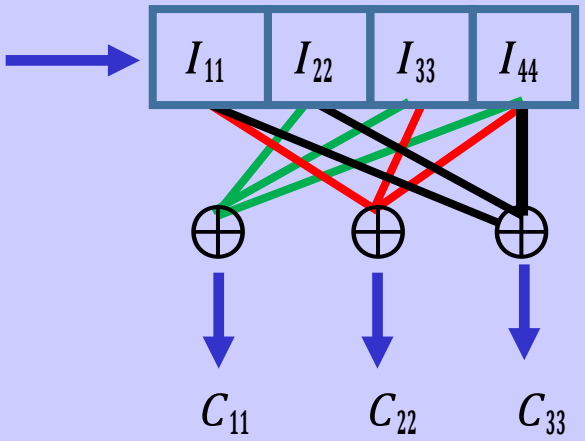
$$C = [11111111].\begin{bmatrix} 1111111111111 \\ 1111111111111 \\ 1111111111111 \\ 1111111111111 \end{bmatrix}$$

$$C = [1011010\,]$$

$$C = [I_{11} I_{22} I_{33} I_{44} C_{11} C_{22} C_{33}]$$

$$C_{11} = I_{22} \oplus I_{33} \oplus I_{44}$$
$$C_{22} = I_{11} \oplus I_{33} \oplus I_{44}$$
$$C_{33} = I_{11} \oplus I_{22} \oplus I_{44}$$

$C_{11}=I_{22} \oplus I_{33} \oplus I_{44}$
$C_{22}=I_{11} \oplus I_{33} \oplus I_{44}$
$C_{33}=I_{11} \oplus I_{22} \oplus I_{44}$



Codeword Truth Table:

Data 4 bits $I_{11}I_{22}I_{33}I_{44}$	Codeword 7 bits $I_{11}I_{22}I_{33}I_{44}C_{11}C_{22}C_{33}$
0000	0000000
0001	0001111
0010	0010
0011	0011
….	….
….	….
1111	1111

If codeword [0011001] find received signal

$$H = \begin{bmatrix} P^T & : I \\ rxk & rxr \end{bmatrix} \qquad H = \begin{bmatrix} 0111100 \\ 1011010 \\ 1101001 \end{bmatrix}$$

$$H.C^T = \begin{bmatrix} 0111100 \\ 1011010 \\ 1101001 \end{bmatrix} . \begin{vmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{vmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It is correct codeword datat is [0011]

If codeword [0010001] find received signal

$$H.C^T = \begin{bmatrix} 0111100 \\ 1011010 \\ 1101001 \end{bmatrix} . \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

There is error [0010001] correct codeword [0011001]

Homework 4: **The generator matrix of a LBC is given by:**

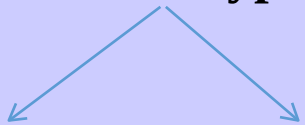
$$[G]=\begin{bmatrix}1&0&0&1&1&0&1&1&0&0\\0&1&0&0&1&1&0&1&1&0\\0&0&1&0&0&1&1&1&0&1\end{bmatrix}$$

a-Use Hamming bound to find error correction capability. b-Find the parity check matrix. c-find the code table, Hamming weight and the error correction capability then compare with part(a). d-If the received word is [R]=[1011110011], find the corrected word at the Rx.

Cyclic Codes

These are subclass from the linear block codes. The name cyclic comes from the fact that any cyclic shift of a codeword is another codeword. i.e, if [C1]=[0011010] is a codeword then [C2]=[0001101] is another codeword obtained from [C1] by a right circular shift.

Cyclic codes can be classified into two types:



Nonsystematic cyclic codes **Systematic cyclic codes**

Generation of cyclic codes:
A) nonsystematic Cyclic Codes: (Multiplicative):

Procedure:

(1) For [D]=[a_1 a_2 ….. a_k] data word, write the data word in terms of a power of a dummy variable x with a_1 weighted as MSB (Most Significant Bit) and a_k as LSB(Least Significant Bit).

x^{k-11} x^{k-2} x^{22} x^{11} x^{11}

a_1 a_2 ..... a_{k-2} a_{k-1} a_k

MSB LSB

$D(x)=a_k+a_{k-1}\,x+a_{k-2}\,x^2+\ldots\ldots+a_2\,x^{k-2}+a_1\,x^{k-1}$ where "+" sign is mod-2 addition(Ex-OR)

For example if [D]=[1 1 1 0 1],

then $D(x)=1+x^2+x^3+x^4$

and if $D(x)=x^6+x^2+1$ then [D]=[1000101]

2) Multiply D(x) by what is called generator polynomial g(x) of order r=n-k.

(3) The output codeword polynomial will be:

C(x)=D(x) g(x) from which we can find the output codeword [C]

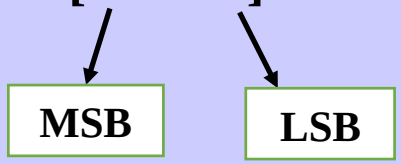
Example: $g(x) = x^{33} + x + 11$

A.

1. $xx^{rr} = xx^{33}$
2. **have x^r , 1**

B.

D=[1 0 1 1]



$$D(x) = 11 + 11x + 11x^{22} + 11x^{33}$$
$$D(x) = 11 + 11x + 11x^{33}$$

D=[1 0 0 0]

$$D(x) = x^{33}$$

Example 10: **Find codeword using nonsystematic cyclic code, if** $g(x) = x^{33} + x + 11$ **and** $D = [11\ 11\ 11\ 11]$

Solution:

1. $G(x) = x^{33} + x + 11$

   r=3

$N = k + r = 44 + 33 = 77$

Code (7,4) , code rate = 4/7

2. $D = [11\ 11\ 11\ 11]$

   $D(xx) = 11 + x$ ===> k=4

$$C(x) = G(x)D(x)$$

$$= (x + 11)(x^{33} + x + 11)$$

$$= [x^{44} + x^{22} + x + x^{33} + x + 11]$$

$$C(x) = [1111111111111111]$$

Homework 5: **Find codeword using nonsystematic cyclic code, if** $g(x) = x^{44} + x^{22} + 11$ **and** $D = [11111111]$