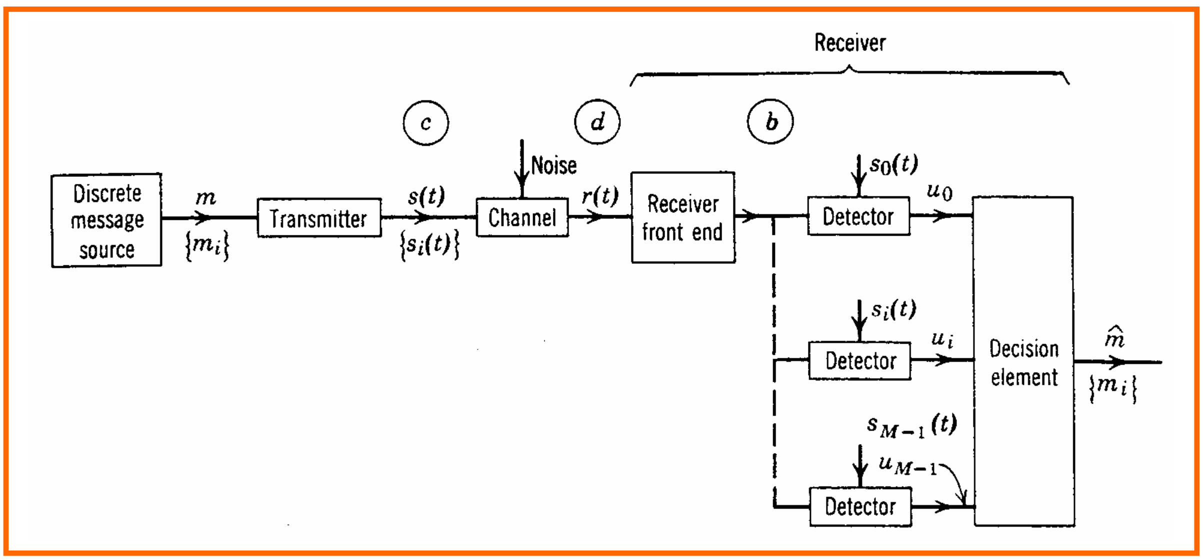
1. INTRODUCTION

- Volume of data transmission for the users of personal communications and Personal computer systems is growing unbounded, even to the extend of bringing the service providers to a complete stall.
- World-wide-web facilities have opened the eyes of millions of users to be more demanding on everything they interface with in their daily lives.
- Increasing demand for
  - high quality digital telephony,
  - digital TV, and
  - multi-media communications over advanced networks prompted numerous studies in the area of digital communications.
- Primary objective of this course is to broaden general and working knowledge of engineers and scientist in the area of modern techniques on emerging digital communication systems and in their implementations.
- This course is designed for engineers and researchers involved in all fields of digital communications and signal processing including: signal processing and data compression, video coding in multimedia applications, telecommunication systems and services, defense and manufacturing.

1.1 Communication System Models

- Generic Digital Communication Systems and Associated Signals:



- Messages are normally discrete and finite; but they could be very large in quantity.

$$m = \{m_i; i = 1,2,...\} \tag{1.1}$$

- There is a particular signal waveform generated and transmitted for each message m_I:

$$s(t) = \{s_i(t); i = 1,2,...\} \tag{1.2}$$

- White or bandlimited channel noise, device noise, distortion due to digitization, compression, etc., intersymbol interference, near and far end cross-talk, jammers are all generically lumped into "NOISE": *n(t).*
- Received signal is commonly represented as the sum of the transmitted signal (attenuated and delayed in the channel but reconstructed back to the original in the front end) plus the additive noise from all the ills mentioned above:

$$r(t) = s(t) + n(t) \tag{1.3}$$

- It can be one of many possible signal waveforms:

$$r(t) = \{r_j(t);\ j = 1,2,\dots\} \tag{1.4}$$

Some signals might be lost in the channel as in the case of an ***erasure channel***, or foreign signals might be picked up as in the case of a ***jammed channel*****.** Therefore, it is not necessary that $i=j$.

- Detected and decoded signal:

$$u(t) = \{u_j(t);\ j = 1,2,\dots\} \tag{1.5}$$

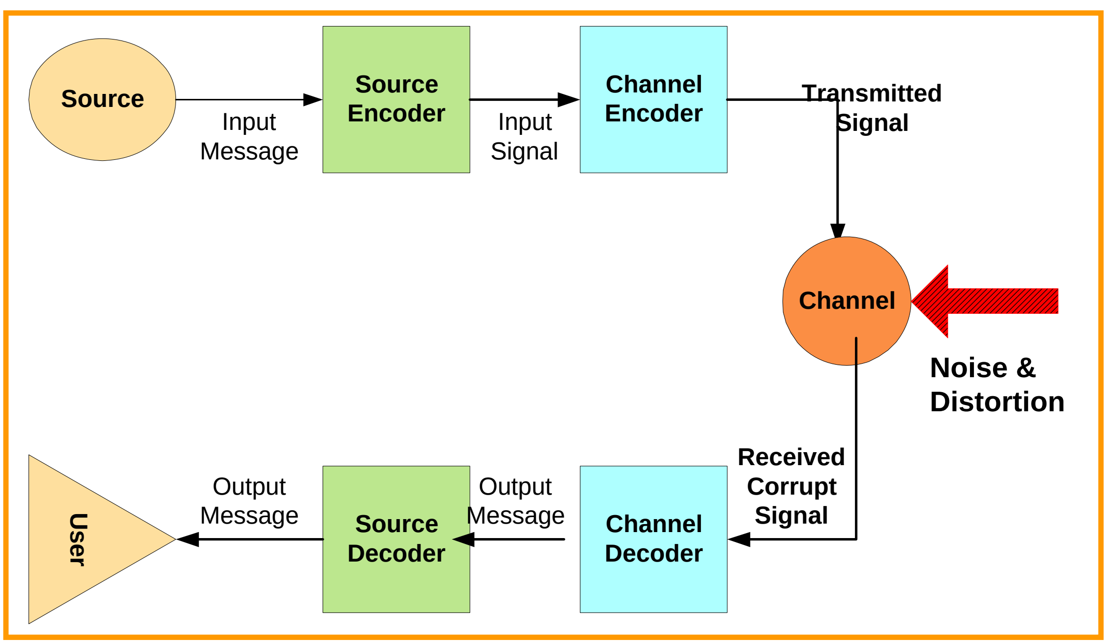
- Reconstructed message:

$$\hat{m} = \{\hat{m}_j;\ j = 1,2,\dots\} \tag{1.6}$$

TASK: Design a communication system with the property that transmitted signal m_i is received without error:

$$\hat{m}_j = m_i \tag{1.7}$$

- Shannon's Point-to-Point Digital Communication Systems Model:



Source: Set of Symbols generated by a person or a system to be sent over a transmission medium to a user. Examples: Speech/audio, Image/video, Telemetry and other sensor data, Computer data, Bio-sensor readings, etc.

Source Encoder: Messages from users are highly redundant. Compression of redundancy in a systematic manner is called source encoding. Examples:

- CELP coding for speech/audio signals.
- JPEG coding for still images.
- Lempel-Ziv universal lossless coding for text compression.

Channel Encoder: Coding for improved transmission over physical medium. Examples:

- Run-length line coding
- Convolutional codes
- QAM, FSK, DPSK, QPSK, and other codes for data transmission.

In many cases two are combined and called ***Encoder*** and described by: What designer gets to do to signal before sending it over the channel. It can include:

- Preprocessing,
- Sampling and A/D conversion,
- Signal decompositions,
- Modulation, and
- Compression.

Goal: Prepare signal for channel in a way decoder can recover good reproduction.

Channel: Physical medium for communication process. This portion of communication system is out of designer's control. It is often described in terms of a conditional probability distribution and a linear filtering operation. It could be: Deterministic or Random. Examples:

- On-line media:
  - Null (transparent) channel
  - Air/deep space
  - Telephone lines, twisted-pair/coaxial cable (POTS).
  - Ethernet
  - Fiber-optic link
- Off-line media:
  - CD
  - Magnetic tape/Magnetic disk
  - Computer memory

Channel and Source Decoders: They attempt to perform inverse operations of the source encoder and the channel encoder, respectively. The combination is called as the ***Decoder*** and described by: What decoder gets to do to channel output in order to reconstruct or render a version of the signal for the user. It can include inverses or approximate inverses of encoder operations, or other stuff to enhance reproduction.

Distortion and Noise: When the continuous or analog signals are digitized and compressed there is always a cost associated with the process. In digitization of band-limited signals, we employ Nyquist Theorem to guarantee exact reconstruction. However, any other source compression is realized at a cost of varying degree of imperfect representation. This is called distortion and it is NOT recoverable. In addition, signals in the communication link are faced with number of ills. They are loosely called noise.

The presence of noise on a signal changes it shape and characteristics and it limits the ability of the intended receiver to make correct symbol decisions, and thereby the effects the rate of reliable communication. Examples:

- Additive Gaussian White Noise.
- Device noise.
- Atmospheric noise in the microwave channels
- Intersymbol interference in data communication systems
- Interspeaker interference in voice communications
- Near-end and Far-end crosstalk and Echoes in Link and chamber
- Friendly and unfriendly jammers, etc.

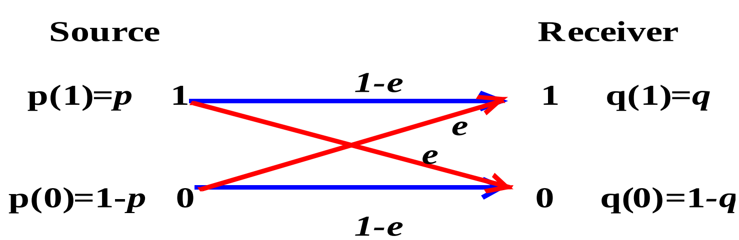
User: The intended user of the input information-bearing messages, usually a replica of the original input messages. The messages coming to user may not need an identical replica of the sender's information symbols. A good example would be access control of a safe room by the voice print of intended user.

1.2 Review of Probability Theory

Underlying assumption in communication**:** ***If the intended user knew what the source message was there is no need to communicate.***

Real-life scenario: Transmitter is connected to a random source; channel is corrupted also in a random manner; and the receiver cannot predict a transmitted message with certainty.

Example 1.1: Consider a Binary Symmetric Channel (BSC) with a bit-error-rate (BER) of *e*, where a binary source sends {1,0} with probability *{p, 1-p}:*



If we transmit ASCII characters (7-bit+parity bit) over this BSC:

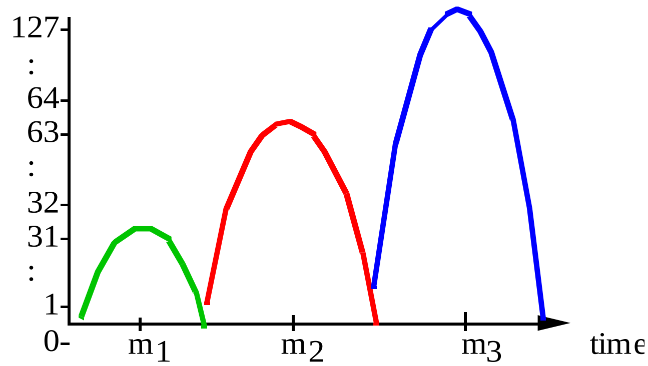
$$m = \{m_i \varepsilon (1,0); i = 0,1,2,...,127\} \quad (1.8)$$

$m_i = d_1 d_2 d_3 d_4 d_5 d_6 d_7$ and each digit is either 0 or 1.

Then we have a digit-by-digit binary transmission. Alternatively, if we rewrite (1.8) using

$$m_i = d_1.2^{-1} + d_2.2^{-2} + ... + d_7.2^{-7} \quad (1.9)$$

We can use m-ary signaling levels as shown below and transmit only one specific waveform for each message. This principle is now actively pursued in emerging communication systems with multi-level signaling.



Example 1.2: Radio Communication problem.

Signals from the transmitting antenna are reflected from (and refracted by) various layers of the ionosphere and scattered. This constitutes a "Diversity Channel," and the signals picked up by the receiving antenna are multi-path scattered signals from the genuine source as well as all other sources operating at the same frequency band, which are labeled as friendly and unfriendly jammers. Reception possibility and the quality are governed by a set of statistical measurements performed on incoming signals.

- An experiment is random when the conditions of measurements are not predetermined with sufficient accuracy and completeness to permit a "precise" prediction of a random trial, such as, digit-by-digit transmission of binary symbols over a BSC.
- Outcomes: Measured quantities from an experiment, such as amplitude and phase characteristics of incoming signals in a radio receiver.
- Results: Set of outcomes between which we choose to distinguish, such as AM signals only in a narrow-band of interest.
- Sample Space: **S:** Set of all possible outcomes of an experiment.
- Event: **A** = {*w*: Outcomes such that some condition on w is satisfied} (1.10)

Kolmogorov's Probability Axioms:

1. To every event A_i, a unique number $P(A_i)$ is assigned such that:
$$0 \le P(A_i) \le 1 \qquad (1.11)$$
2. $P(S) = 1$ (1.12)
3. If A and B are mutually exclusive; that is, $A \cap B = \phi$ then
$$P(A \cup B) = P(A) + P(B) \qquad (1.13)$$

3.A. If A and B are not mutually exclusive; that is, $A \cap B \neq \phi$ then
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (1.14)$$

- Compliment: $\overline{A}$: Compliment of A.
- Union event: $D = A \cup B$
- Total Probability: If all A_i are mutually disjoint then

$$P(\bigcup_i A_i) = \sum_i P(A_i) \tag{1.15}$$

- Joint Probability: If the sample space is partitioned into events of to different sets, such as, the amplitude and phase ranges of an incoming signal coming to a radio receiver, where

$$\{A_{i=}; i = 1,2,\ldots,n\} \quad \text{and} \quad \{B_j = j = 1,2,\ldots,m\}$$

$$0 \le P(A_i, B_j) \le 1 \tag{1.16}$$

- Marginal Probability: If B_j are mutually exclusive or disjoint then:

$$\sum_{j=1}^{m} P(A_i, B_j) = P(A_i) \tag{1.17}$$

and

$$\sum_{i=1}^{n} \{\sum_{j=1}^{m} P(A_i, B_j)\} = 1 \tag{1.18}$$

- Conditional Probability:

If $P(B) > 0$ then $P(A \mid B) = \dfrac{P(A,B)}{P(B)}$ (1.19)

If both $P(A) > 0$ and $P(B) > 0$ then:

$$P(A,B) = P(A \mid B).P(B) = P(B \mid A).P(A) \tag{1.20}$$

Which is known as the Bayes Rule.

Example 1.3: Die Tossing Experiment.

$$S = \{1,2,3,4,5,6\}$$

Let $A = \{Odd\_outcomes\} = \{1,3,5\}$. $P(A) = 1/2$

a) $\bar{A} = \{Even\_outcomes\} = \{2,4,6\}$ $P(\bar{A}) = 1/2$

b) If a new event is defined as small face values: $B = \{1,2,3\}$ and $D = A \bigcup B$ then $P(D) = P(\{1,2,3,5\}) = 4/6$.

Because:

$P(D) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/2 - P(\{1,3\}) = 1 - 1/3 = 2/3$

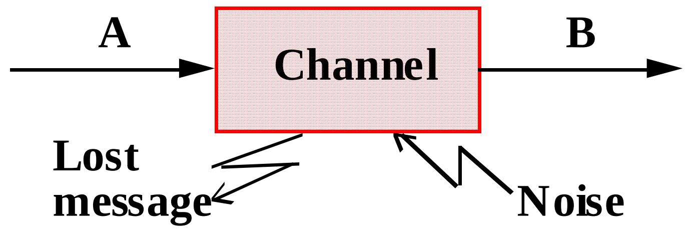
c) Given that the outcome was small what is the probability that it was odd?

$$P(A \mid B) = \frac{P(Odd, Small)}{P(Small)} = \frac{2/6}{1/2} = 2/3$$

Bayes Theorem: If $\{A_i = i = 1,2,\ldots,n\}$ are disjoint then $\bigcup_{i=1}^{n} A_i = S$ and $P(B) > 0$ then

$$P(A_i \mid B) = \frac{P(A_i, B)}{P(B)} = \frac{P(B \mid A_i).P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i).P(A_i)} \tag{1.21}$$

Example 1.4: Communication over a noisy channel.



$P(A_i)$: *Apriori* probabilities of input events (Designer's Problem.)

$P(A_i \mid B)$: *Aposteriori* probability of A_i conditioned on having observed the received signal B to be a particular outcome. (Optimum communication task of a receiver!).

Statistical Independence (SI): If occurrences of A does not depend on occurrences of B then

$$P(A \mid B) = P(A) \quad \text{and} \quad P(A,B) = P(A).P(B) \qquad (1.22)$$

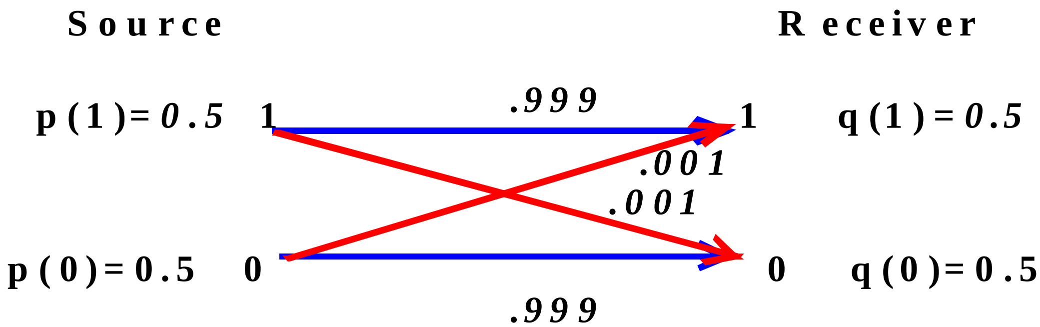
and *A,B* are called statistically independent.

Example 1.5: Die tossing experiment with: **Odd:** $P(A) = 1/2$ and **Small**: $P(B) = 1/2$

Probability of Odd & Small: $P(A,B) = P(\{1,3\}) = 1/3$

$P(A).P(B) = (1/2).(1/2) = 1/4 \neq 1/3$. Therefore, *A,B* are not SI.

Example 1.6: Because of silences, voice communication is very inefficient on dedicated POTS lines. Measurements show that "1/3" of the time actual speech goes through the line. This excess capacity can be used (a) to increase the service by TASI or DSI type services to better utilize the bandwidth or (b) to improve the performance in noisy channels. Given that each bit has a BER of 0.1% over a BSC and consecutive bits are SI what would be the bit rate if three bit-long sequences are sent for each bit? Assume that a majority rule decoder is used in the receiver to decide each triplet.



Solution:

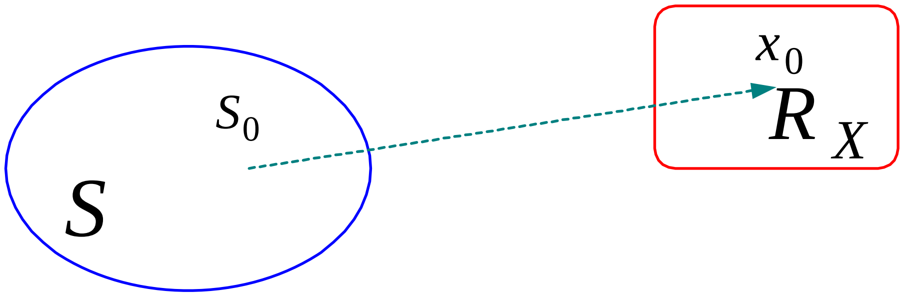
Symbols are incorrect when 2 or 3 out of three bits are in error. If we assume *p=1/2* then

$Total\ BER = 2\{(1/2).[3.(0.001)^2.(0.999) + (0.001)^3.(0.999)^0]\} \approx 3.10^{-6}$

We have traded-off rate (1/3) for a significantly reduced BER. This simple rule has been used in classical computer memory storage algorithms.

1.3 Random Variables

Means of associating real variables with random experiments and their outcomes.



Example 1.7: Coin tossing experiment with a Sample Space: $S = \{H, T\}$
Let us define a binary assignment function to the outcomes of each toss:

$$X(s) = \begin{cases} 1 & if \;\; s = H \\ 0 & if \;\; s = T \end{cases}$$

Example 1.8: Voltage reading across a load:
Sample Space: $S = \Re$ and $X(s) = V$

- Probability Notations: Given a continuous random variable: X, consider the event $X \le x$, we can represent the probability of this event in many ways:
$$P(X \le x) = \Pr(X \le x) = \Pr ob(X \le x) \qquad (1.23)$$

- Cumulative Distribution Function (cdf):
$$F(x) = P(X \le x) \quad for \quad -\infty < x < \infty \qquad (1.24)$$

Properties of cdf:

1. $0 \le F(x) \le 1$ (1.25)
2. $F(-\infty) = 0 \quad and \quad F(\infty) = 1$ (1.26)

- Probability Density Function (pdf):
$$f_X(x) = p(x) = \frac{dF(x)}{dx} \quad and \quad F(x_0) = \int_{-\infty}^{x_0} p(x)dx \qquad (1.27)$$

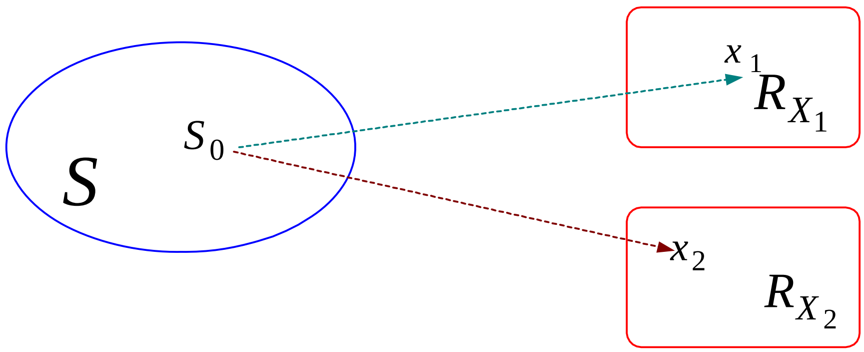
- If a r.v. is discrete or mixed type, the discrete portion is normally represented by probability mass functions (pmf):
$$p(x) = \sum_{i=1}^{n} P(X \le x).\delta(x - x_i) \qquad (1.28)$$

- If $-\infty < x_1 < x_2 < \infty$
then $P(X \le x_2) = P(X \le x_1) + P(x_1 < X \le x_2)$ (1.29)
or equivalently:
$$F(x_2) = F(x_1) + \Pr ob(x_1 < X \le x_2) \qquad (1.30)$$
and

$$\Pr ob(x_1 < X \le x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p_x(x)dx \tag{1.31}$$

Multiple Random Variables

Consider an experiment with a sample space S and the outcomes are mapped by two different functions, such as an AM receiver, where each received waveform has random amplitude and a random phase value.



- Joint Cumulative distribution function:

$$F(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} P(u_1, u_2)du_1du_2 \tag{1.32}$$

- Joint probability density function:

$$p(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} \tag{1.33}$$

- Marginal pdfs:

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2)dx_2 \tag{1.34}$$

- Properties:

1. $F(\infty,\infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x_1, x_2)dx_1dx_2 = 1$ (1.35)
2. $F(-\infty,-\infty) = F(-\infty, x_2) = F(x_1,-\infty) = 0$ (1.36)

- Conditional probabilities:

$$P(X_1 \le x_1 \mid X_2 = x_2) \equiv F(x_1 \mid x_2) = \frac{\dfrac{\partial F(x_1, x_2)}{\partial x_2}}{\dfrac{\partial F(x_2)}{\partial x_2}} = \frac{\int_{-\infty}^{x_1} p(u, x_2)du}{p(x_2)} \tag{1.37}$$

and $p(x_1 \mid x_2) = \dfrac{p(x_1, x_2)}{p(x_2)}$ if $p(x_2) > 0$ *for all values of* x_2. (1.38)

- Revisit Bayes Theorem:

$$p(x_1, x_2) = P(x_1 \mid x_2).p(x_2) = p(x_2 \mid x_1).p(x_1) \quad (1.39)$$

- For multiple dimensions of order n:

$$p(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n} \quad (1.40)$$

then:

$$p(x_1, x_2, ..., x_n) = P(x_1, .., x_k \mid x_{k+1}, ..., x_n).p(x_{k+1}, ..., x_n) \quad (1.41)$$

If the multiple r.v. are statistical independent:

$$F(x_1, x_2, ..., x_n) = F(x_1).F(x_2)...F(x_n) \quad (1.42)$$

$$p(x_1, x_2, ..., x_n) = p(x_1).p(x_2)...p(x_n) \quad (1.43)$$

Functions of Random Variables

Typical Question: If $Y = g(X)$ is a new r.v. obtained operating on *X*, can we compute statistics of *Y* from those of *X*?

Example 1.9: Linear functions, such as linear filtering or prediction operations.
If *a>0* and constant and *Y=aX+b* then

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P(X \le \frac{y-b}{a}) \quad (1.44)$$

Example 1.10: Quadratic functions:

$$F_Y(y) = P(Y \le y) = P((aX^2 + b) \le y) = P(|X| \le \sqrt{\frac{y-b}{a}}) \quad (1.45)$$

$$F_Y(y) = F_X(\sqrt{\frac{y-b}{a}}) - F_X(-\sqrt{\frac{y-b}{a}}) \quad (1.46)$$

and since $x_{1,2} = \pm\sqrt{\frac{y-b}{a}}$ are solutions of our quadratic equation we have:

$$p_Y(y) = \frac{P_X(x_1)}{2ax_1} + \frac{P_X(x_2)}{2a(-x_2)} \quad (1.47)$$

- In general, if there are *n* real roots of $Y = g(X)$ then the fundamental theorem of pdf transformations states that:

$$p_Y(y) = \sum_{i=1}^{n} \frac{p_X(x_i)}{|g'(x)|_{x=x_i}}$$ where $|g'(x)|_{x=x_i}$ is the derivative of $g(x)$ at $x=x_I$ (1.48)

and

$$p_Y(y_1, y_2, ..., y_n) = p_X(x_1 = g_1^{-1}, x_2 = g_2^{-1}, ..., x_n = g_n^{-1})|J| \quad (1.49)$$

where $x_i = g_i^{-1}$ is the i^{th} solution of $Y = g(X)$ equation and the Jacobian is given by:

$$J = \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial y_1} & \cdots & \frac{\partial g_n^{-1}}{\partial y_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial g_1^{-1}}{\partial y_n} & \cdots & \frac{\partial g_n^{-1}}{\partial y_n} \end{bmatrix} \tag{1.50}$$

Example 1.11: Linear Predictive Coding (LPC) with constant coefficients:

$$Y = AX \quad \text{or} \quad y = \sum_{j=1}^{n} a_{ij} x_j \quad \mathit{for} \quad i = 1,2,\dots,n \tag{1.51}$$

Assume A is non-singular then we have $X = A^{-1}Y$ with solutions:

$x_i = \sum_{j=1}^{n} b_{ij} y_j \quad \mathit{for} \quad i = 1,2,\dots,n \quad \mathit{and} \quad \{b_{ij}\}$ are the elements of A^{-1}. Since the system is linear we have:

$$J = \frac{1}{\det(A)} = \det(A^{-1}) \tag{1.52}$$

and the pdf of the output is:

$$p_Y(y_1, y_2, \dots, y_n) = p_X(x_1, x_2, \dots, x_n).\frac{1}{|\det(A)|} \tag{1.53}$$

1.4 Statistical Averages & Characteristic Functions

The generic formula for the computation of statistical (ensemble) averages is given by:

$$E\{X^n\} = \int_{-\infty}^{\infty} x^n p_X(x) dx \tag{1.54}$$

1. Mean or Ensemble Average (not time average) : $n=1$

$$E\{X\} = m_x = \mu_x \tag{1.55}$$

2. If $Y = g(X)$ then $E\{Y\} = E\{g(X)\} = \int_{-\infty}^{\infty} g(x) p_X(x) dx$ (1.56)

3. Mean Square Value: If $Y = X^2$ then $E\{X^2\} = \int_{-\infty}^{\infty} x^2 . p_X(x) dx$ (1.57)

4. Central Moments: It is the case for: $Y = (X - m_x)^n$

$$E\{Y\} = E\{(X - m_x)^n\} = \int_{-\infty}^{\infty} (x - m_x)^n . p_X(x) dx \tag{1.58}$$

4. Variance is simply the case for $n=2$:

$$\sigma_x^2 = E\{(X - m_x)^2\} = \int_{-\infty}^{\infty} (x - m_x)^2 p_X(x) dx \tag{1.59}$$

It is worth noting that the computation of variance in (1.59) requires a two-pass procedure to compute the mean first and then the variance. However, it can be shown from the superposition theorem that:

$$\sigma_x^2 = E\{X^2\} - (E\{X\})^2 = E\{X^2\} - m_x^2 \tag{1.60}$$

which is a single pass operation.

5. Other moments:

$$E\{(X_1 - m_1)^k.(X_2 - m_2)^l\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X_1 - m_1)^k.(X_2 - m_2)^l.p(x_1, x_2)dx_1dx_2 \tag{1.61}$$

6. Covariance: *k=l=1*

$$\mu_{ij} = E\{(X_i - m_i).(X_j - m_j)\} = E\{x_i x_j\} - m_i m_j \tag{1.62}$$

7. Characteristic Function (Moment Generating Function):

$$\Phi_X(s) = E\{e^{sx}\} = \int_{-\infty}^{\infty} e^{sx}.p_X(x)dx \quad \text{where } s \text{ is a complex variable.} \tag{1.63}$$

8. Properties:

$$\left.\frac{\partial \Phi_X(s)}{\partial s}\right|_{s=0} = E\{X\} = m_x \qquad \text{and} \qquad \left.\frac{\partial^2 \Phi_X(s)}{\partial s^2}\right|_{s=0} = E\{X^2\} \tag{1.64}$$

Example 1.13: Uniform Random Variables:

$$p_X(x) = \begin{cases} \dfrac{1}{b-a} & if \;\; a \le x \le b \\ 0 & otherwise \end{cases} \tag{1.65}$$

$$E\{X\} = \frac{1}{b-a}\int_a^b t dt = \frac{1}{b-a}.\left.\frac{t^2}{2}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$E\{X^2\} = \frac{1}{b-a}\int_a^b t^2 dt = \frac{1}{b-a}.\left.\frac{t^3}{3}\right|_a^b = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

and the variance is simply:

$$\sigma_x^2 = E\{X^2\} - (E\{X\})^2 = \frac{(b-a)^2}{12}$$

Finally, it can be shown that:

$$\Phi_X(s) = \frac{e^{jwb} - e^{jwa}}{jw(b-a)}$$

Example 1.14: Gaussian Random Variables:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m_x)^2/2\sigma^2} \tag{1.66}$$

where: $\mu = m_X = mean$ and $\sigma^2 = variance$ and the cdf is given by:

$$F_X(x) = \Pr ob(X \le x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(u-m_x)^2/2\sigma^2}.du \quad (1.67)$$

9. If the r.v. has zero-mean, $m_x = 0$ and unity variance, $\sigma^2 = 1$, we have the Standard Gaussian (or normal) distribution of N(0,1) and

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (1.68)$$

and $$Q(x) = \Pr ob(X > x) = 1 - F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-u^2/2}.du \quad (1.69)$$

which is both tabulated and plotted in many communication texts and handbooks. It is worth noting that the above expression holds for a generic Gaussian r.v.

$$\Pr ob(X > x) = 1 - F_X(x) = Q(\frac{x-m}{\sigma}) \quad (1.70)$$

$$\log_e \Phi_X(s) = ms + \frac{\sigma^2 s^2}{2}$$

with properties:

1. $Q(x) \equiv \frac{1}{2}[1 - erf(\frac{x}{\sqrt{2}})]$

2. $Q(x) \le e^{-x^2/2}$

Central Limit Theorem:

- Let $\{x_i;\ for\ 1 \le i \le N\}$ be a set of N independent identically distributed r.v. (i.i.d.) with pdf: $p_{x_i}(x) = p_X(x)$ and a finite variance σ^2.
- $$Z \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \quad (1.71)$$
- It can be shown that cdf of Z approaches to a cdf of a Gaussian r.v. as $N \to \infty$.
- Equivalently, each r.v. x_i represents a random event and Z is the cumulative effect of these events then cdf of this new r.v. approaches that of a Gaussian r.v. regardless of the cdf of each x_i.

(a) It can be shown that if x_i are i.i.d. Gaussian with the same mean and variance then

$$Z = a_1 x_1 + a_2 x_2 + ... + a_N x_N \quad (1.72)$$

is a Gaussian r.v. with zero mean $E\{Z\} = 0$ and a variance:

$$Var(Z) = \sigma_z^2 = \sigma^2 (a_1^2 + a_2^2 + ... + a_N^2) \quad (1.73)$$

(b) Two r.v. with $E\{X\} = E(Y) = 0$ and $\sigma_x^2 = \sigma_y^2 = \sigma^2$ are jointly Gaussian with a pdf:

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\{-\frac{x^2 - 2\rho xy + y^2}{2\sigma^2(1-\rho^2)}\} \quad (1.74)$$

where the correlation coefficient is given by: $\rho \equiv E\{XY\}/\sigma^2$ and $-1 \le \rho \le 1$.

(c) If X and Y have $E\{XY\} = 0$ then they are ***Uncorrelated***.

- If a vector $\underline{X}$ has elements which are zero-mean, independent and jointly Gaussian with the same variance then their joint pdf is given by the general Gaussian formula:

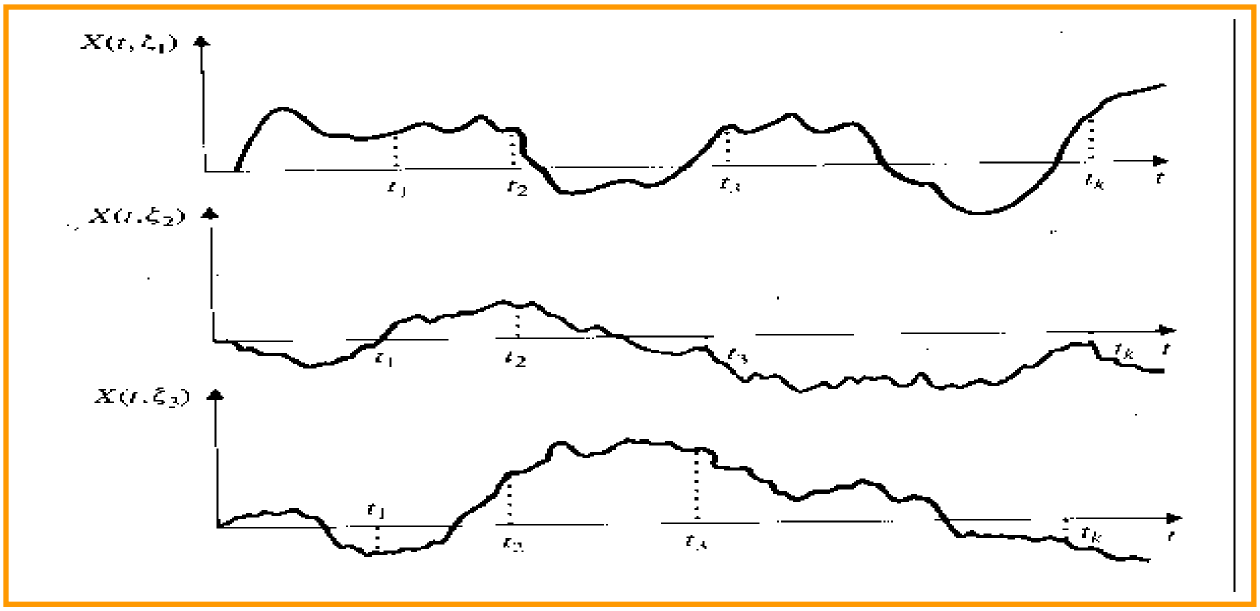
$$p_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{M/2}\sigma^M} \exp\{-\frac{\|\underline{x}\|^2}{2\sigma^2}\} \qquad (1.75)$$

where M is the dimension of the random vector $\underline{X}$.

1.5 Stochastic (Random) Processes

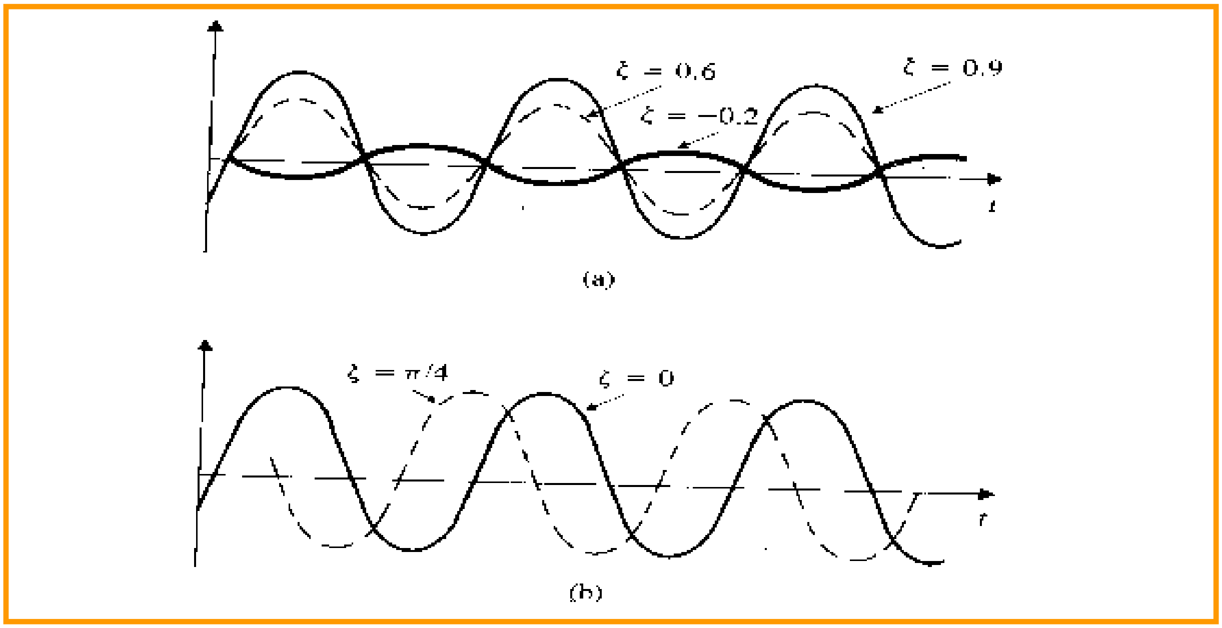
If the random variable is a time-dependent function, that is, it is changing with time then we have a random (stochastic) process defined by a random function: $x(t,\xi)$, where t is the ordinary time variable.

1. The graph of $x(t,\xi)$ for a fixed experimental outcome $\Theta = \xi$ is called a realization, such as, the received signal envelope in a binary PAM system:



2. For each fixed time t_k, the set of values $x(t_k,\xi)$ is a r.v. Indexed family of r.v. form the stochastic process $X(t)$.
3. If the time index set is discrete then we have a discrete-time random process. If the index is continuous then we have a continuous-time random process.

Example 1.15: Amplitude and phase modulated signals:
(a) AM: If $\xi \in [-V,V]$ Volts is changing wrt a pdf and $x(t,\xi) = \xi Cos(2\pi t)$.
(b) PM: If $\xi \in [-\pi,\pi]$ is normally uniformly distributed and $x(t,\xi) = Cos(2\pi t + \xi)$.



- Mean Function: In general, it is a function of time defined by:

$$m_x(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x p_{X(t)}(x) dx \qquad (1.76)$$

- Autocorrelation function:

$$R_x(t_1,t_2) = E\{X(t_1).X(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_{X(t_1)X(t_2)}(x_1,x_2)dx_1 dx_2 \qquad (1.77)$$

- Autocovariance function:

$$C_X(t_1,t_2) = E\{[X(t_1) - m_x(t_1)].[X(t_2) - m_x(t_2)]\} = R_x(t_1,t_2) - m_x(t_1).m_x(t_2) \qquad (1.78)$$

- Variance function:

$$\sigma^2_{X(t)} = E\{[X(t) - m_x(t)]^2\} = C_X(t,t) \qquad (1.79)$$

Gaussian Process:

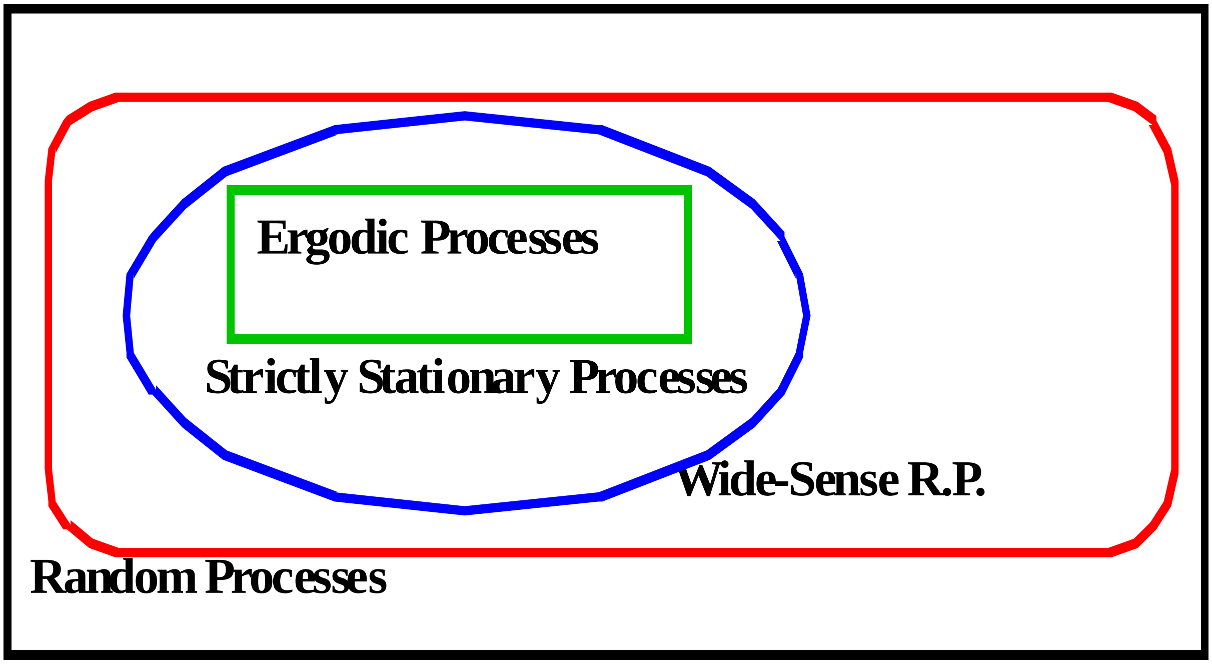
Let $\underline{X}(t) = \{X(t_1), X(t_2),..., X(t_n)\}$ be a process, where the elements of $\underline{X}(t)$ are jointly Gaussian for every finite set of time indices, and then $\underline{X}(t)$ is a Gaussian Process.

1. The joint pdf depends on the means:

$$m_X = E\{X\} = [\overline{x_1}, \overline{x_2},..., \overline{x_n}]$$

2. and the set covariances:

$$\lambda = E\{(x_i - \overline{x_i}).(x_j - \overline{x_j})\} = E\{x_i x_j\} - \overline{x_i}.\overline{x_j} \qquad (1.80)$$



- A process is ***strictly stationary*** if its nature of randomness stays unchanged with time (independent of time origin.) In other words, its joint pdf is independent of time shifts. This constraint is fairly difficult to meet in many real-life applications.
- A process is ***wide-sense (weakly) stationary*** if its mean and autocorrelations are independent of the actual time index.

1. $m_x(t) = m$
2. $R_x(\tau) = E\{X_{k+\tau}.X_k^*\}$ for discrete processes

   and $R_x(\tau) = E\{X(t+\tau).X^*(t)\}$ for continuous processes.

Properties:

1. Power of a zero-mean process:

   $R_x(0) = E\{|X_k|^2\}$ for discrete and $R_x(0) = E\{|X(t)|^2\}$ (1.81)

2. Power Spectral Density (PSD) is the Fourier transform of the autocorrelation function: $R_x(\tau)$

   $$S_x(w) = F\{R_x(\tau)\} = \int_{-\infty}^{\infty} R_x(\tau)e^{-jw\tau}d\tau \qquad (1.82)$$

3. Power is the total area under the power density curve:

   $$R_x(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(w)dw \qquad (1.83)$$

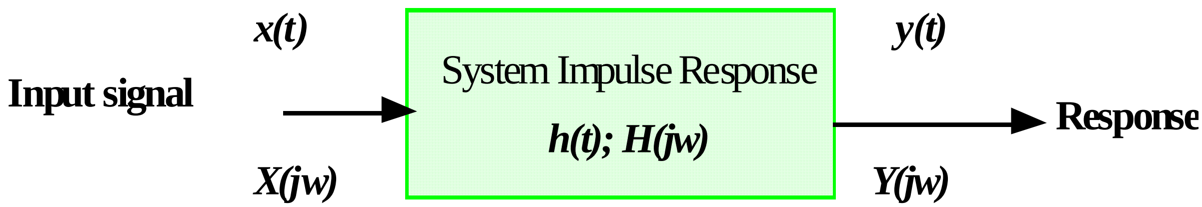
4. All strictly stationary processes are also wide-sense stationary. But the converse is not always true.
5. If the signal is Gaussian distributed then wide-sense stationarity implies strict-sense stationarity.

- If the ensemble (statistical) averages of all orders of a stationary process are equal to the time-averages of the same process then this process is *ergodic*.

$$\overline{x(t)} = \lim_{T\to\aleph}[\frac{1}{T}\int_T x(t)dt] = E\{x(t)\} \text{ and } \overline{x^P} = E\{x^P\} \quad (1.84)$$

where P is the order of the statistics.

1.6 Statistics of Linear System Responses



Consider a linear system with an impulse response $h(t)$ and a frequency response $H(jw)$:

Fundamental Relationship: Convolution in time-domain:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} h(\tau).x(t-\tau)d\tau = \int_{-\infty}^{\infty} x(\tau).h(t+\tau)d\tau \quad (1.85)$$

Multiplication in frequency-domain: $Y(jw) = X(jw).H(jw)$ (1.86)

If $x(t)$ is stationary then:

1. $m_y = E\{y\} = \int_{-\infty}^{\infty} h(\tau)E\{x(t-\tau)\}d\tau = m_x \int_{-\infty}^{\infty} h(\tau)d\tau = m_x H(0)$ (1.87)

   Output mean is the input mean times the frequency response at D.C.

2. $S_y(jw) = S_x(jw).|H(jw)|^2$ (1.88)

   Output power spectral density is the input times the magnitude-square of the system frequency response. Thus, no need to compute the output statistics at all; they are available from input statistics.

1.6 Additive Gaussian White Noise and Bandlimited White Noise

As discussed earlier, additive noise channel model is normally used as the channel model in communication systems because of its simplicity, maybe more critically, its mathematical tractability.

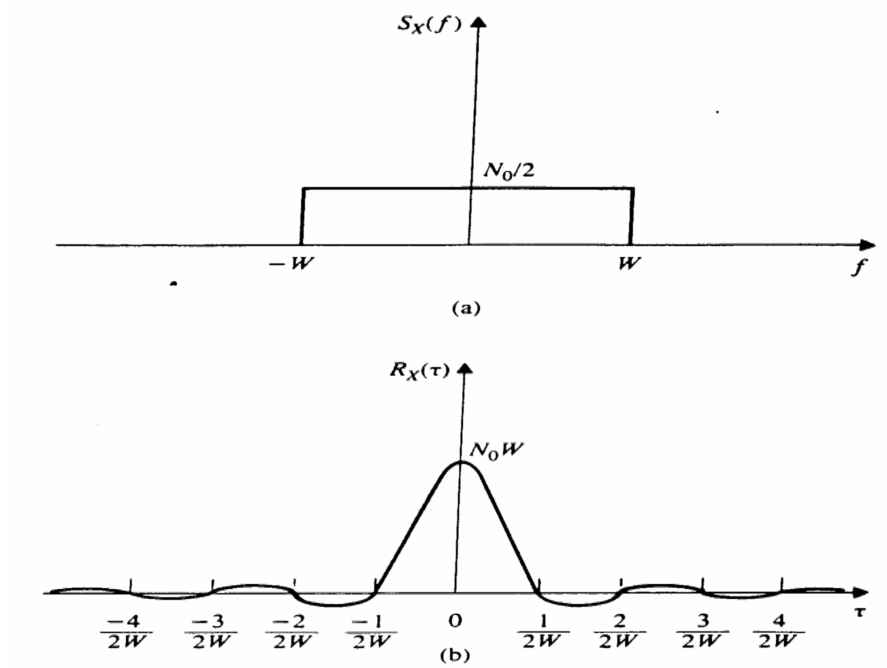
$$y(t) = x(t) + n_w(t) \quad (1.89)$$

where the noise term $n_w(t)$ is a zero-mean Gaussian process, with a constant power spectral density and an impulsive autocorrelation function:

$$S_n(jw) = N_0 \quad \text{for all } w \quad \text{and} \quad R_n(\tau) = \delta(\tau) \quad (1.90)$$

If the PSD is constant over a finite bandwidth then it is called a colored (pink) or bandlimited white noise as in the case of *Thermal Noise*.

Example 1.16: Thermal Noise with bandwidth: $W = 10^{12}\ Hz$.



It is worth noting that:

$$S_n(jw) = \begin{cases} N_0/2 & if\ -2\pi W \le w \le 2\pi W \\ 0 & otherwise \end{cases} \quad \text{and} \quad R_x(\tau) = \frac{B}{\pi} \cdot \frac{Sin(Bt)}{Bt}$$

- If x(t) is sampled at a rate π / B then the samples will have Z.C. at these sampling points and the sampled (discrete-time) white Gaussian noise will have "UNCORRELATED" samples:

$$R_y(\pi / B) = R_y(2\pi / B) = ... = 0 \qquad (1.91)$$

- If the noise is bandlimited as in thermal noise then the samples will be given by:

$$y(t) = \sum_{m=-\infty}^{\infty} x_m h(t - mT) \qquad (1.92)$$

where sampling period: $T = \pi / B$ and x_m are the samples of white noise. Last equation is also known as the Pulse Amplitude Modulation (PAM) formula and hence, this type of sampling is called PAM sampling.