

Probability Theory, Random Variables and Random Processes

Difference between Deterministic and Random signals is pretty much apparent. Deterministic signals are one in which the value of the signals is known at every given point in time and thus it is possible to give such signals an accurate mathematical modelling without having any uncertainty about their behaviour. For eg. a fixed frequency, phase and amplitude sine wave.

Random signals, on the other hand, maintain a certain degree of uncertainty till they actually occur. Such signals are highly complicated to model mathematically but after sufficient observation over a period of time, it is possible to develop a probabilistic and statistical model for such signals.

Some examples are:

- ① Human speech
- ② Noise signals of various kind etc.

Let us try and develop first as to how a degree of randomness can be accounted for using probability assessments and then understand how this analysis can help us in analysing random signals as a whole.

Random Experiments

Any experiment for which the output, i.e. unpredictable, called a random experiment. Every output or outcome of a random experiment is a sample point. The set of all such sample points is called a sample space.

\Rightarrow Sample Space $\Omega \cong$ Set of all possible outcomes

$$= \{s_1, s_2, \dots, s_n\}$$

↑
sample point

Event: can be defined as a single sample point or a group of sample points. Hence any subset of sample space is an Event.

* For n samples in Ω , number of possible events is 2^n (power set)

Sure Event $S \cong$ Always occurs

Null Event $\emptyset \cong$ Never occurs

Both $S, \emptyset \subset \Omega$

(Ω, E, P) : Probability System / space

↑ ↑ ↑
Sample space Event space Probability measure

P : Probability measure assigned to each event in the field ' E '.

Properties

$$\textcircled{1} \quad P(S) = P(\Omega) = 1$$

$$\textcircled{2} \quad 0 \leq P(A) \leq 1$$

$$\textcircled{3} \quad \text{If } A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$$

These are called axioms of probability.

Let's discuss another implication

$$\Omega = \{s_1, s_2, \dots, s_N\}$$

$$P_i = P(s_i)$$

$$\textcircled{1} \quad P_i \leq 0 \quad \textcircled{2} \quad \sum_{s_i \in \Omega} P_i = 1 \quad \textcircled{3} \quad \text{If } A \subset \Omega$$

$$\Rightarrow P[A] = \sum_{s_i \in A} P_i$$

Probability Assignment (modelled on the physics of the situation)

Must reflect all the information available on the phenomenon.

Eg. Assuming all outcomes N are equally likely, then the probability of each is $1/N$.

Independent events

$$\text{2 experiments : } \Omega_1 = \{s'_1, s'_2, \dots, s'_m\} \rightarrow P_i$$

$$\Omega_2 = \{s''_1, s''_2, \dots, s''_n\} \rightarrow P_j$$

Joint experiment

$$\Omega = \{(s_i, s_j)\}$$

So the outcomes now will be these pairs. Thus instead of standard probabilities, we now assign joint probabilities.

In case sample points in Ω_2 occur without any information about the occurrence of any sample points in Ω_1 and vice versa, we can call Ω_1 & Ω_2 as independent experiments.

Mathematically

$$P(s'_i, s''_j) = P_i P_j \quad \text{then } s'_i \text{ & } s''_j \text{ are independent events}$$

Say in case 2 events $A_1, A_2 \in \Omega$

(145)

$$\text{then } P[A_1, A_2] = P[A_1] \cdot P[A_2] \quad \text{then } A_1 \text{ & } A_2 \text{ are independent events.}$$

These are also sometimes called exclusive events.

Conditional Probability

Mathematically denoted by $P[A|B]$, conditional prob. means the probability of A given that B has occurred.

Mathematical definition

$$P[A|B] = \frac{P[A, B]}{P[B]}$$

* If $P[A, B] = P[A] \cdot P[B]$ (independent) then

$$P[A|B] = P[A]$$

* In a way this is basically an attempt to have a better a priori knowledge of the occurrence of the event A. Thus the conditional probability will have a theoretically minimum value of $P[A]$ (in case of A & B being independent). In all practical cases where event A depends on B,

$$P[A|B] > P[A]$$

Another way of representing $P[A, B]$ is thus

$$P[A, B] = P[A|B] \cdot P[B].$$

* Baye's theorem:

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

* Total Probability theorem:

If $A_i \in \mathcal{S}$ $i=1, 2, \dots, N$ are all independent and B is some arbitrary event, then

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_N)P(A_N)$$

* Rewriting Baye's theorem

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^N P(B|A_i)P(A_i)}$$

Example 1: Coin toss experiment

$$\Omega = \{H, T\}$$

$$E = \{\emptyset, S, H, T\}$$

where $\emptyset = H$ and T both occur = $H \cap T$

$S = H$ or T one occurs = $H \cup T$

$H = H$ occurs

$T = T$ occurs.

Then let us define the mapping

$$x(H) \triangleq 1 \quad \& \quad x(T) \triangleq 0$$

$\Rightarrow H$ maps to 1 and T maps to 0.

Example of $\emptyset \triangleq x \leq -5$

Example of $S \triangleq x \leq 2$

Example 2: URN contains 3 balls: W, B, R

$$\Omega = \{W, B, R\}$$

The x may be defined as

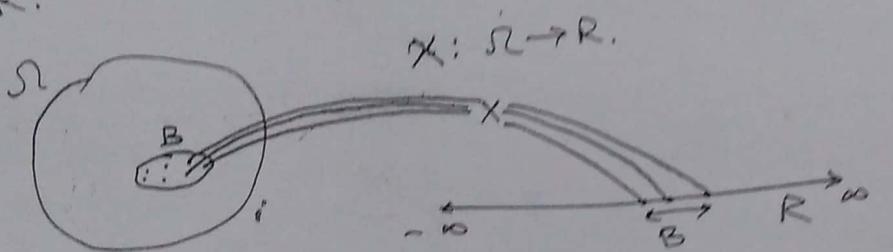
$$x(s) = \begin{cases} 1 & s = W \text{ or } B \\ 0 & s = R. \end{cases}$$

Once we have established the mapping of the random variable to the real value set, it is thus better to study the probability of the occurrence of the events on the basis of their real valued maps as well. An effective tool for this is Probability distribution function.

Random Variables

Let us assume that Ω is the sample space. Then all points in this set Ω will be sample points or possible outcomes. Thus there might be a subset of Ω , say B , which can be understood as an event.

Random variable X in this case can thus be defined as the variable which maps the events in Ω to real values in the set of all real values R .



- * There is no randomness in the mapping. Each event will uniquely map to different set of values in the real set R . But the values associated to each random occurrence or events will be randomly received. Hence the variable essentially has a random behaviour and hence the name random variable.

- * X is basically a rule to establish correspondence between sample points in Ω and points on real axis R .
In some ways it is a misnomer to call it a random variable firstly because it is actually a function, and secondly because there is nothing random about the mapping. It is called a random variable because of the randomness of the quantities or entities being mapped.

Probability Distribution Function

we cannot go about defining the probability $P[\cdot]$ for every point or subset of \mathbb{R} as it will be highly awkward (and difficult).

Thus it is convenient to introduce a pointwise probability function: PDF:

Formal definition

$$\text{PDF} : \triangleq P[X \leq x]$$

(1)

x = Random variable

x = numerical value that defines the range

$$\Rightarrow P[-\infty < X \leq x] \Rightarrow \xrightarrow{\text{---}} \mathbb{R} \quad x$$

$= P_X(x) \text{ or } F_X(x)$

Properties of $P_X(x)$

$$1. 0 \leq P_X(x) \leq 1$$

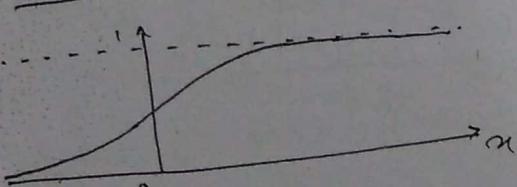
$$2. P_X(-\infty) = 0, P_X(\infty) = 1$$

3. $P_X(x)$ is always a non decreasing function of x .

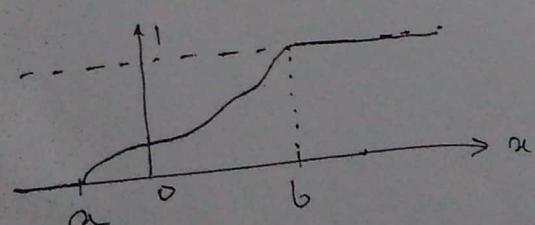
$$4. P_X[x_1 < X \leq x_2] = P_X(x_2) - P_X(x_1)$$

Books generally define this function as the Cumulative distribution function.

Some Examples

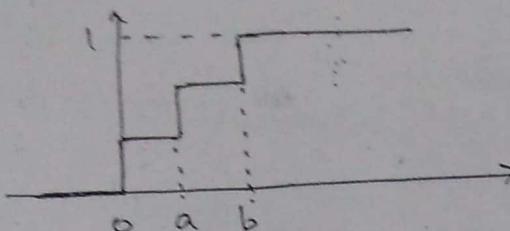


Continuous random variable
with values between $(-\infty, \infty)$



Continuous random variable
with values between (a, b)

(149)



Discrete RV with 3 possible values
 $\{0, a, b\}$

$$\therefore P[X < a] = P[0]$$

$$P[X \leq a] = P[0] + P[a]$$

* In case of a continuous RV, $P[X = a] = 0$ as it will be
 $P[X = a] = P(a) - P(a)$

Probability density function

$$p_x(x) \triangleq \frac{d P_x(x)}{dx}$$

This basically implies that

$$p_x(x) dx = \text{Prob}[x < X \leq x + dx]$$

As we know that $P_x(x)$ is a non decreasing function, we can see that $p_x(x)$ is never going to have a negative value. Thus $p_x(x)$ can have any non-negative value (it may be zero sometimes).

Properties

$$1. p(x) \geq 0 \quad -\infty < x < \infty$$

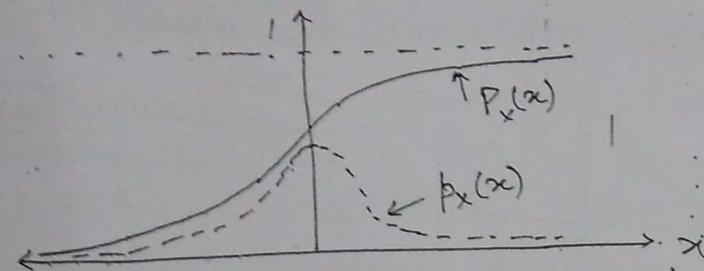
$$2. \int_{-\infty}^{\infty} p(x) dx = 1$$

$$3. P_x(x) = \int_{-\infty}^x p_x(u) du$$

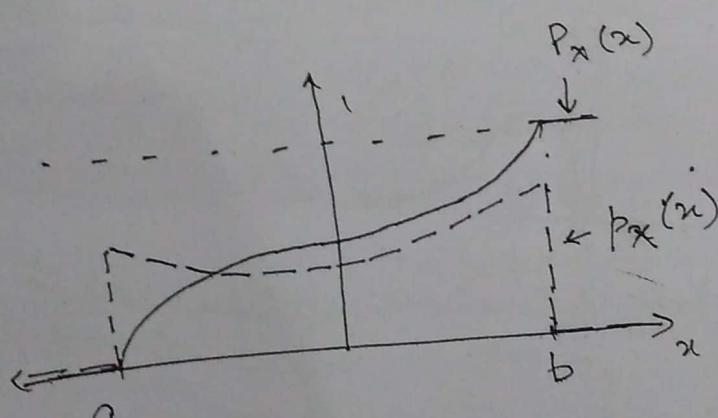
$$4. \int_{x_1}^{x_2} p_x(u) du = \text{Prob}[x_1 \leq X \leq x_2]$$

* Please note here that PDF / CDF are notations for the distribution function while 'pdf' is the notation for the density function.

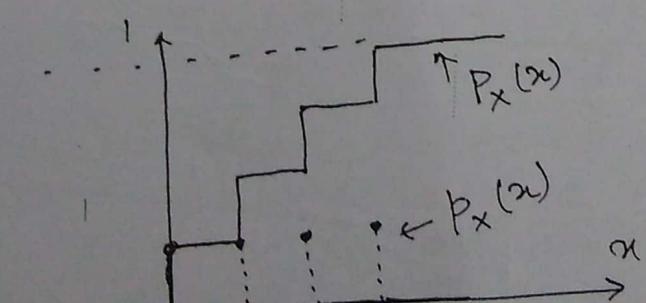
Some typical examples



Continuous Random Variables with values between $(-\infty, \infty)$



Continuous RV with values between (a, b)

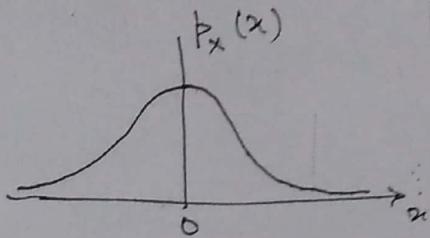


Discrete RV with 4 possible values at $\{0, a, b, c\}$

151

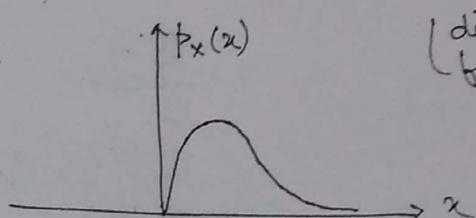
Most commonly used pdf's:

① Gaussian: $p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$



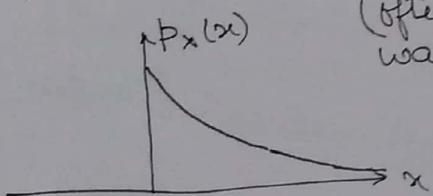
(most common
naturally occurring
pdf)
(* noise in commun
sys)

② Rayleigh: $p_x(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ (defined only for $x \geq 0$)



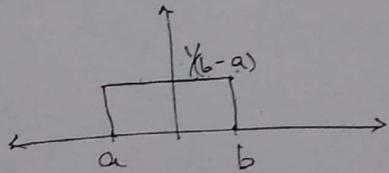
(directly derivable
from Rayleigh)

③ Exponential: $p_x(x) = \lambda e^{-\lambda x}$ (often used to model waiting time in queues)



④ Uniform: $p_x(x) = \frac{1}{b-a}$

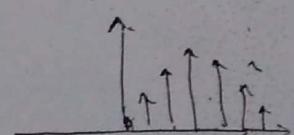
* If we have no apriori knowledge of a pdf of a RV, the safest assumption is to use a Uniform pdf



⑤ Binomial: $p_x(k) = {}^n C_k P^k (1-P)^{n-k}$

* Discrete RV.
Used to model random draw experiments with replacement.

where $n = \text{No of yes/no cases}$
 $k = \text{Kth case}$
 $P = \text{Probability of yes}$



Given the density or distribution functions, we have the complete description of the working or behaviour of the Random Variable.

But generally, the complete description is not always available and one is forced to work with more or a gross parameterization. Hence we define the Expected (or average) Values and moments of an RV.

Expected Values & Moments

$$\text{Expected Value } E[x] \stackrel{\text{(mean value)}}{\equiv} \sum_{\text{for discrete RV}} x_i p_x(x_i) = \int_{-\infty}^{\infty} x p_x(x) \text{ for continuous RV}$$

This is the quantity that is most likely to occur if an experiment is conducted a large number of times.

This definition is in concurrence with the standard definition of the sample average for a set of samples

$$x_1, x_2, \dots, x_N.$$

$$\text{i.e. } \lim_{N \rightarrow \infty} \frac{x_1 + x_2 + \dots + x_N}{N} = E[x]$$

* If we have finite number of possible samples taken over the entire range of x , the ratio of the total no. of occurrences and the total no. of samples under consideration gives the probability of occurrence of that sample. Hence we converge to the above mentioned equation describing the mean value.

More generally, expected value of any function $f(x)$ can be defined as

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

(153)

One function of special importance is

$$f(x) = x^n$$

$$\Rightarrow E[x^n] = \int_{-\infty}^{\infty} x^n p_x(x) dx \triangleq \bar{x}^n : n\text{th moment of } x.$$

* $E[x^n]$ and \bar{x}^n are used as interchangeable notations

Now, \bar{x} = Mean Value

\bar{x}^2 = Mean Squared Value

$\sqrt{\bar{x}^2}$ = Root Mean Squared Value
or Effective Value

* This leads us to an implication that it might be sufficient for us to know the mean values of the Random Variable to get a broad characterization of its behaviour.

Central Moments

Normal Moments are the moments about the origin, irrespective of the type of distribution.

Central Moments, on the other hand, are moments calculated about the mean value. Thus, to calculate the n th central moment of x , we take

$$f(x) = (x - \bar{x})^n$$

$$\Rightarrow E[(x - \bar{x})^n] = \overline{(x - \bar{x})^n} = \int_{-\infty}^{\infty} (x - \bar{x})^n p_x(x) dx$$

What follows is

$$\textcircled{1} \quad n=1 : \overline{(x - \bar{x})} = 0 \quad (E(x - \bar{x}) = E(x) - \bar{x} = \bar{x} - \bar{x} = 0)$$

$$\textcircled{2} \quad n=2 : \overline{(x - \bar{x})^2} = \sigma^2 = \text{Variance}$$

(measure of the average spread of the random variable over the mean value)

* A gross description of a random variables is possible with the knowledge of the mean value and the mean squared value or the variance.

For a detailed description we need the knowledge of the density function or distribution function. Note that for gross description, we don't really require the pdf. The sample point averages can be used to calculate the same.

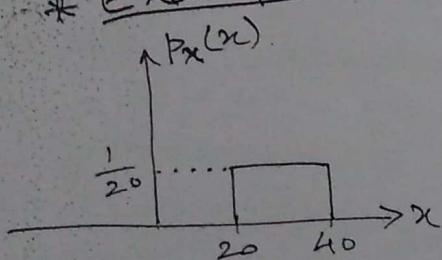
The nature of the Expectation operator is linear.

$$E[x_1 + x_2 + x_3 + \dots + x_m] = E[x_1] + E[x_2] + E[x_3] + \dots + E[x_m]$$

Similarly

$$\begin{aligned} \sigma^2 &= E[(x - \bar{x})^2] = E[x^2 - 2x\bar{x} + \bar{x}^2] \\ &= E[x^2] - 2\bar{x}E[x] + \bar{x}^2 \quad (\bar{x} \text{ is a constant}) \\ &= E[x^2] - \bar{x}^2 \\ &= \bar{x}^2 - \bar{x}^2 \\ &= \text{Mean squared value} - (\text{mean})^2 \end{aligned}$$

* Example



$$p_x(x) = \begin{cases} 0 & -\infty < x \leq 20 \\ \frac{1}{20} & 20 < x \leq 40 \\ 0 & 40 < x < \infty \end{cases} \quad \checkmark \quad (3)$$

$$\bar{x} = \int_{20}^{40} x \cdot \frac{1}{20} dx = \frac{1}{20} \left[\frac{x^2}{2} \right]_{20}^{40} = 30.$$

$$\bar{x}^2 = \int_{20}^{40} x^2 \cdot \frac{1}{20} dx = \frac{1}{20} \left[\frac{x^3}{3} \right]_{20}^{40} = 933.3 \quad (155)$$

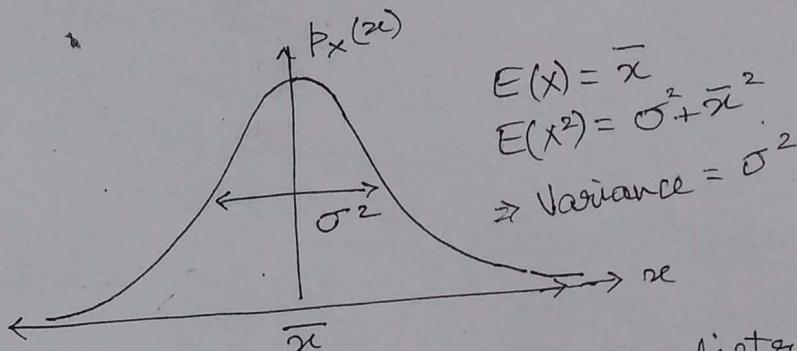
$$\Rightarrow \text{Variance} = \sigma^2 = \bar{x}^2 - \bar{x}^2 = 33.3$$

$$\text{Standard deviation} = \sigma = \sqrt{33.3} = 5.77$$

Gaussian Random Variable

A Random Variable is said to be Gaussian if its pdf takes the value of a Gaussian function.

$$p_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\bar{x})^2/2\sigma^2}$$



\bar{x} indicates the mean value of the distribution.
 σ^2 describes the spread about the mean value.
If σ^2 is small, the distribution is a narrow function. If σ^2 is large, the distribution is broad.

Gaussian model is the most common naturally occurring random variable. The reason for that is attributed to the central limit theorem. Central limit theorem states that if you add up a large number of random variables, then the resulting random variable tends to have a Normal Distribution which is another name for Gaussian Distribution.

What must be noted here is that the commonly observed natural phenomena are a combination of a large number of experiments, possibly independent of each other, so the net effect of these random experiments usually takes the Gaussian distribution as proved by the central limit theorem.

Let us take into consideration the noise in a communication system. Examining the cause of this noise, we may say that it is due to the random movement of electrons somewhere. The sum total of all these electrons (which is observed as the net current) will naturally have a Gaussian distribution.

Whenever current is observed there is always some variation in the value which exists because of the amount of electrons that are flowing. This randomness of variation is what we call the noise. Hence it can be understood that most of the noise encountered in such communication systems has a Gaussian distribution.

Besides the point discussed above, other advantages of using a Gaussian distribution are

① Mathematically, it is much more convenient to use this model, especially when dealing with a large number of random variables.

② Linear combination of Gaussian Random variables will also be a Gaussian RV.

③ Gaussian distribution can be completely described by the first two moments of the RV.

* For a general case, the knowledge of infinite moments will be required to correctly derive the exact form of distribution function.

Handling two or more random variables

Many times we encounter a situation where we are needed to handle more than one random variables. In such a case we can extend the notions of single random variables to multiple Random Variables by defining Joint Distribution Functions and Joint density functions, i.e. joint PDF and joint pdf.

Definitions:

$$\text{Joint PDF: } P_{xy}(x, y) = P\{X \leq x, Y \leq y\}$$

Properties:

(1) $0 \leq P_{xy}(x, y) \leq 1 \quad -\infty < x, y < \infty$

(2) $P(-\infty, y) = P(x, -\infty) = P(-\infty, -\infty) = 0$

(All such cases are impossible events)

(3) $P(\infty, \infty) = 1$

(4) $P(x, y)$ is a non decreasing function in both x and y

* Essentially, both the PDF and pdf in case of joint RVs take a 3 dimensional plot.

Joint pdf:

$$p_{xy}(x, y) = \frac{\partial^2 P(x, y)}{\partial x \partial y}$$

Properties:

(1) $p(x, y) \geq 0 \quad -\infty < x, y < \infty$

(2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1$

(3) $P_{xy}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(x, y) dx dy$

Some important Properties of Joint PDFs & pdfs

Joint PDF

$$P(\infty, y) = P_y(y); P(x, \infty) = P_x(x)$$

Marginal distribution fns.

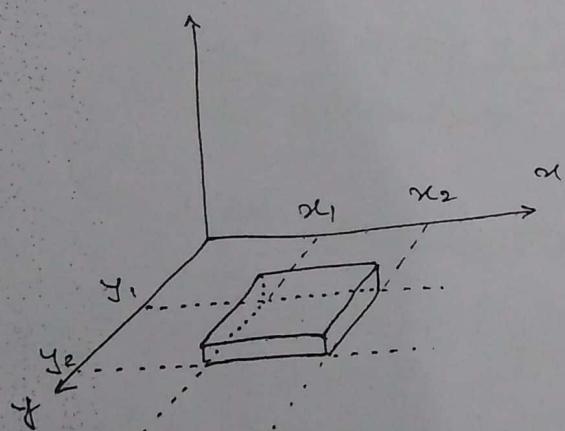
$$\text{Joint pdf} \\ P_X(x) = \int_{-\infty}^{\infty} p(x, y) dy$$

$$p_Y(y) = \int_{-\infty}^{\infty} p(x, y) dx$$

Marginal density fns.

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) dx dy = \text{Prob } \{x_1 \leq x \leq x_2; y_1 \leq y \leq y_2\}$$

Example: Uniform Joint pdfs. ✓



$$p(x, y) = \begin{cases} \frac{1}{(x_2 - x_1)(y_2 - y_1)} & x_1 \leq x \leq x_2, y_1 \leq y \leq y_2 \\ 0 & \text{otherwise} \end{cases}$$

*Additional Definitions for single Variables

Median of an RV

m is the median of an RV if $P[-\infty \leq x \leq m] = P[m \leq x \leq \infty]$

$$\Rightarrow \int_{-\infty}^m f_x(x) dx = \frac{1}{2}$$

(159)

2. Mode of an RV is the value of RV that occurs with maximum probability. That means x_m is mode of an RV ' x ' if $f_x(x) \leq f_x(x_m)$ where $-\infty \leq x \leq \infty$.

PTO

Statistical Independence (SI)

- x & y are statistically independent when knowledge of one gives no information about the other;

or $P_{xy}(x,y) = P_x(x) P_y(y)$: necessary and sufficient condition for SI

Example 1: 2D uniform density function

$$p_{xy}(x,y) = \begin{cases} \frac{1}{(x_2-x_1)(y_2-y_1)} & \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{array}{l} x_1 \leq x \leq x_2 \\ y_1 \leq y \leq y_2 \end{array}$$

otherwise.

$$= p_x(x) \cdot p_y(y)$$

$$\text{where } p_x(x) = \begin{cases} \frac{1}{x_2-x_1} & x_1 \leq x \leq x_2 \\ 0 & \text{otherwise} \end{cases}$$

$$p_y(y) = \begin{cases} \frac{1}{y_2-y_1} & y_1 \leq y \leq y_2 \\ 0 & \text{otherwise.} \end{cases}$$

Example 2: Statistically independent Gaussian RV's

$$p_{xy}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\frac{(x-\bar{x})^2}{\sigma_x^2} + \frac{(y-\bar{y})^2}{\sigma_y^2} \right] \right\}$$

Expected Values of Fns of 2 RV's

$$E[F(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) p_{xy}(x,y) dx dy.$$

Important special case

$$E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p_{xy}(x,y) dx dy$$

\triangleq Correlation between X and Y.

For example: Uniform $p_{xy}(x,y)$:

$$\begin{aligned} E[XY] &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} xy \frac{dx dy}{(x_2-x_1)(y_2-y_1)} \\ &= \frac{(x_2^2 - x_1^2)(y_2^2 - y_1^2)}{(x_2 - x_1)(y_2 - y_1) \cdot 4} = \left(\frac{x_2 + x_1}{2}\right) \left(\frac{y_2 + y_1}{2}\right) \\ &= E[X] \cdot E[Y] \end{aligned}$$

(basic consequence of SI)

More generally for SI cases

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p_{xy}(x,y) dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} xy p_x(x) p_y(y) dy \\ &= \int_{-\infty}^{\infty} x p_x(x) dx \cdot \int_{-\infty}^{\infty} y p_y(y) dy = \bar{x} \cdot \bar{y} \end{aligned}$$

= Product of the two means.

If any of these two means is zero, the correlation is also zero.

(161)

Covariance

If we consider correlation as a joint moment about origin, covariance may be defined as the central moment.

$$E[(x-\bar{x})(y-\bar{y})] \triangleq \text{Covariance}$$

$$\Rightarrow f(x,y) = (x-\bar{x})(y-\bar{y}).$$

Let us first define a few terms.

* $E[x^i y^j] \triangleq (i+j)$ th order joint moment of the two random variables x & y .

\Rightarrow Correlation is a second order joint moment.

Similarly

$$E[(x-\bar{x})(y-\bar{y})^i] \triangleq (i+j)$$
 th order joint central moment

In case we define x & y as uncorrelated R.V's then their covariance is found to be zero
(NOT when the correlation is zero)

* If x, y are SI then they must be uncorrelated, but if they are uncorrelated, they might or might not be statistically independent.

$$\begin{aligned} \sigma_{xy} &= E[(x-\bar{x})(y-\bar{y})] = E[x\bar{y} - \bar{x}\bar{y} - \bar{y}\bar{x} + \bar{x}\bar{y}] \\ &= E[x\bar{y}] - \bar{x}E[\bar{y}] - \bar{y}E[\bar{x}] + \bar{x}\bar{y} \\ &= E[x\bar{y}] - \bar{x}\bar{y} \end{aligned}$$

\therefore It can be seen mathematically as well, that if $E[x\bar{y}] = \bar{x}\bar{y}$ i.e. SI, then $\sigma_{xy} = 0$. which indicates that x and y are uncorrelated.

Correlation Coefficient

It is defined by the ratio

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

where $|\sigma_{xy}| \leq \sigma_x \sigma_y$

or $|r_{xy}| \leq 1$

It is simple enough to prove as follows:

$$\begin{aligned} E[a(x - \bar{x}) + (y - \bar{y})]^2 &= E[a^2(x - \bar{x})^2 + (y - \bar{y})^2 + 2a(x - \bar{x})(y - \bar{y})] \\ &= E[a^2(x^2 + \bar{x}^2 - 2x\bar{x}) + (y^2 + \bar{y}^2 - 2y\bar{y})] \\ &\quad + 2a(xy - \bar{x}\bar{y} - x\bar{y} + \bar{x}\bar{y}) \\ &= a^2 E[x^2] + a^2 \bar{x}^2 - 2a^2 \bar{x}^2 + E(y^2) + \bar{y}^2 - 2\bar{y}^2 \\ &\quad + 2aE(xy) - 2a\bar{x}\bar{y} - 2a\bar{x}\bar{y} + 2a\bar{x}\bar{y} \\ &= a^2 \sigma_x^2 + \sigma_y^2 + 2a\sigma_{xy} \end{aligned}$$

For all values of a , the equation is positive.
Hence the determinant of this quadratic must
be negative.

$$\begin{aligned} &\Rightarrow (\sigma_{xy}^2 - \sigma_x^2 \sigma_y^2) 4a^2 \leq 0 \\ &\Rightarrow \sigma_{xy}^2 - \sigma_x^2 \sigma_y^2 \leq 0 \\ &\Rightarrow \sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2 \end{aligned}$$

or $|\sigma_{xy}| \leq \sigma_x \sigma_y$

* If x and y are considered as vectors in an abstract space, correlation defines the inner product for these vectors. $E\{x^2\}$ and $E\{y^2\}$ represent the square of the length of the vector from origin.

$$\therefore \cos \theta = \frac{E\{xy\}}{\sqrt{E(x^2)E(y^2)}} \text{ where } \theta \text{ is the angle between the two vectors.}$$

In case x & y have zero mean, this ratio also defines the correlation coefficient.

(163)

Now $\rho_{xy} = 1$ simply means that

$$\sigma_{xy} = \sigma_x \sigma_y$$

A small analysis will bring us to the observation that this is only possible when y and x correspond with each other linearly i.e.

$$y = ax \text{ for some } a.$$

A similar observation is made for $\rho_{xy} = -1$ which indicates that x and y are corresponding RVs in length but maintain opposite directions. Such RVs are said to be antipodal in nature.

For $\rho_{xy} = 0$, we may conclude

$$\sigma_{xy} = 0$$

$$\Rightarrow E(xy) = E(x)E(y)$$

This shows that x and y are independent RVs and hence are uncorrelated.

Furthermore, if even one of these RVs has a zero mean, $E(xy) = 0$. In vector space, this can be observed as a case of the two RVs having orthogonality (or angle of $\frac{\pi}{2}$ b/w them).

* Note: (1) If x and y are uncorrelated, then we may state that $x - \bar{x}$ and $y - \bar{y}$ are orthogonal.

(2) If x and y are uncorrelated and $\bar{x} = 0$ or $\bar{y} = 0$ then we observe that x & y are orthogonal.

Before concluding our discussion on Random variables, let us briefly look into some of the other types of RV's.

Rayleigh Distribution

The random variable x is said to be Rayleigh distributed if its density function is given as

$$f_x(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Mean of a Rayleigh distributed Random variable can be calculated as follows : ③

$$E(x) = \int_0^\infty x \cdot \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} dx \quad \checkmark$$

Using the general expression

$$\int_0^\infty x^n e^{-ax^2} dx = \begin{cases} \frac{(n-1)!}{2^{n/2+1}} \cdot \frac{\sqrt{\pi}}{a^{n/2}} & n = \text{even} \\ \frac{[\frac{1}{2}(n-1)]!}{2a^{(n+1)/2}} & n = \text{odd.} \end{cases}$$

$$\Rightarrow E(x) = \frac{1}{\sigma^2} \cdot \frac{1}{4 \cdot 1/2\sigma^2} \cdot \sqrt{\frac{\pi}{1/2\sigma^2}}$$

$E(x) = \sqrt{\frac{\pi}{2}} \sigma$

(165)

Similarly

$$E[X^2] = \int_0^\infty x^2 e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma^2} \cdot \frac{\left[\frac{1}{2} (3-i) \right]!}{2 \cdot \left(\frac{1}{2\sigma^2} \right)^{(3+i)/2}}$$

$$= \frac{1}{\sigma^2} \cdot \frac{1}{2 \cdot (1/2\sigma^2)^2}$$

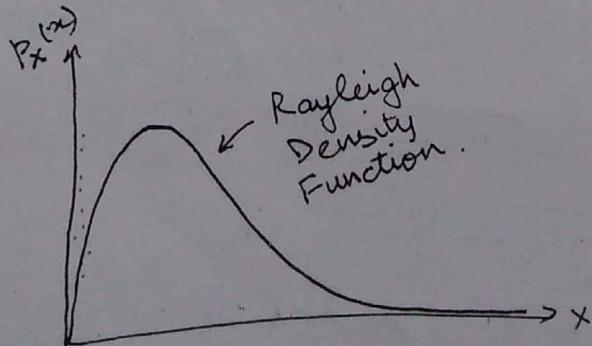
$$\boxed{E(X^2) = 2\sigma^2}$$

$$\Rightarrow \text{Variance} = 2\sigma^2 - \frac{\pi}{2}\sigma^2$$

$$\boxed{\sigma_x^2 = \left(2 - \frac{\pi}{2}\right)\sigma^2}$$

In communication, the signal amplitude values of a randomly received signal can usually be modelled as a Rayleigh distribution.

Mathematically, if x & y are two zero mean independent random variables of equal variance with Gaussian distributions, then the Random Variable $Z = \sqrt{x^2 + y^2}$ will have a Rayleigh distribution.



Exponential Distribution

A random variable x is exponentially distributed if its density function is given by

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

given that $\lambda > 0$.

The mean of this function is given by

$$E(x) = \int_0^\infty x \lambda e^{-\lambda x} dx.$$

$$\text{Say } \lambda x = t.$$

$$\Rightarrow dx = \frac{1}{\lambda} dt.$$

$$\Rightarrow E(x) = \frac{1}{\lambda} \int_0^\infty t e^{-t} dt$$

$$= \frac{1}{\lambda} \left[-t e^{-t} - e^{-t} \right]_0^\infty$$

$$= \frac{1}{\lambda} [0 - (0 - 1)]$$

$$\boxed{E(x) = \frac{1}{\lambda}}$$

Also, mean squared value can be evaluated as

$$E(x^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$$

$$\text{say } \lambda x = t.$$

$$\Rightarrow dx = \frac{dt}{\lambda}$$

$$\Rightarrow E(x^2) = \frac{1}{\lambda^2} \int_0^\infty t^2 e^{-t} dt$$

(167)

$$\begin{aligned} \Rightarrow E(X^2) &= \frac{1}{\lambda^2} \int_0^\infty t^2 e^{-t} dt \\ &= \frac{1}{\lambda^2} \left[-t^2 e^{-t} + 2[-te^{-t} - e^{-t}] \right]_0^\infty \\ &= \frac{1}{\lambda^2} [0 - [0+0-2]] \end{aligned}$$

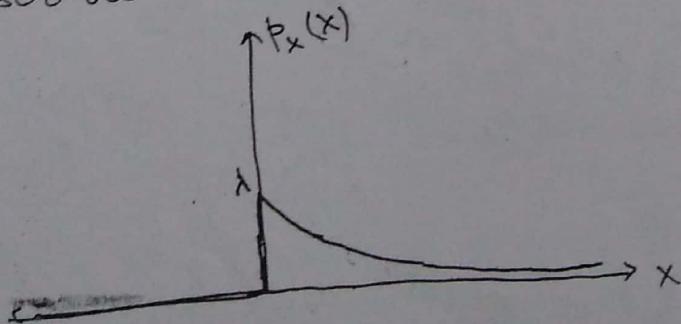
$$E(X^2) = \frac{2}{\lambda^2}$$

$$\Rightarrow \text{Variance} = E(X^2) - \bar{x}^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2}$$

$$\checkmark \quad \sigma_x^2 = \frac{1}{\lambda^2}$$

Normally, occurrence of events ~~happening~~ over nonoverlapping intervals are independent, such as arrival times of telephone calls or bus arrival times at a bus stop, then the waiting time distribution of these events can be shown to be exponential. Again if $x+y$ are independent Gaussian RVs with zero mean and same variance, then the RV described by $z = x^2 + y^2$ will be an exponential distribution with $\lambda = 2\sigma^2$.



Let us further discuss a few discrete Random Variables.

Binomial Distribution

This distribution describes the probability of having x successes in a series of n independent trials.

Provided that the probability of success in any one trial is ' p ', then the probability density function is given by

$$f_x(x) = {}^n C_x P^x (1-p)^{n-x}$$

where $x = 0, 1, 2, \dots, n$.

Thus, the expected value of x may be evaluated as

$$\begin{aligned} E(x) &= \sum_{x=0}^n x f_x(x) \\ &= \sum_{x=0}^n x {}^n C_x P^x (1-p)^{n-x} \\ &= \sum_{x=0}^n x \cdot \frac{n!}{x!(n-x)!} P^x (1-p)^{n-x} \end{aligned}$$

Here $x=0$ term may be excluded from the summation as it has to be 0.

$$\Rightarrow E(x) = \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} P^x (1-p)^{n-x}$$

Substitute $x = y+1$ and $n = m+1$

$$\begin{aligned} \Rightarrow E(x) &= \sum_{y=0}^m \frac{(m+1)!}{y!(m-y)!} P^{y+1} (1-p)^{m-y} \\ &= (m+1)P \sum_{y=0}^m \frac{m!}{y!(m-y)!} P \cdot (1-p)^{m-y} \end{aligned}$$

(169)

$$E(x) = np$$

\therefore the summation simply becomes unity.

$$\begin{aligned}
 E(X^2) &= \sum_{x=0}^n x^2 \cdot {}^n C_x P^x (1-P)^{n-x} \\
 &= \sum_{x=0}^n x^2 \cdot \frac{n!}{x!(n-x)!} P^x (1-P)^{n-x} \\
 &= \sum_{x=1}^n x \cdot \frac{n!}{(x-1)!(n-x)!} P^x (1-P)^{n-x} \\
 &= \sum_{x=2}^n (x-1+1) \frac{n!}{(x-1)!(n-x)!} \cdot P^x (1-P)^{n-x} \\
 &= \sum_{x=1}^n (x-1) \cdot \frac{n!}{(x-1)!(n-x)!} P^x (1-P)^{n-x} \\
 &\quad + \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} P^x (1-P)^{n-x} \\
 &= \sum_{x=2}^n \frac{n!}{(x-2)!(n-x)!} P^x (1-P)^{n-x} \\
 &\quad + np.
 \end{aligned}$$

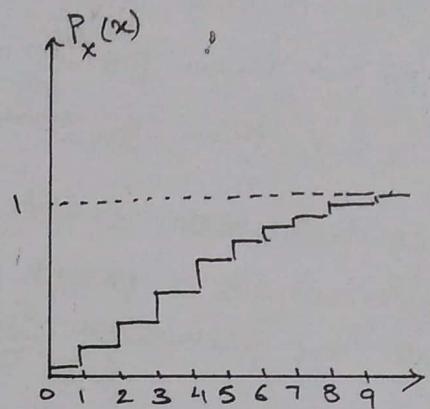
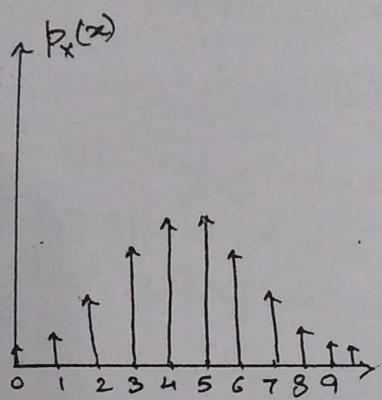
Assume $n = m+2$ & $x = y+2$.

$$\Rightarrow E(X^2) = (m+2)(m+1)P^2 \sum_{y=0}^m \frac{m!}{y!(m-y)!} \cdot P^y (1-P)^{m-y} + np$$

$$E(X^2) = n \cdot (n-1) \cdot P^2 + np$$

$$\Rightarrow \text{Variance} = n(n-1)P^2 + np - (np)^2$$

$$\sigma_x^2 = np(1-P)$$



Binomial Distribution.

Another distribution that is closely connected to the binomial distribution is the Poisson distribution which represents the number of occurrences of a rare event in a large number of trials. Some typical examples include the number of telephone calls at an exchange over a fixed duration, the number of winning tickets among those purchased in a large lottery and the number of printing errors in a book.

Poisson Distribution.

A random variable x is said to be poisson distributed with parameter λ if the probability density function is given by

$$p_x(x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

(6)

where $x = 0, 1, 2, \dots, \infty$

& $\lambda > 0$.

One may interestingly note that

$$\frac{p_x(x-1)}{p_x(x)} = \frac{e^{-\lambda} \cdot \lambda^{x-1}}{e^{-\lambda} \cdot \lambda^x} / \frac{(x-1)!}{x!} = \frac{x}{\lambda}$$

(171)

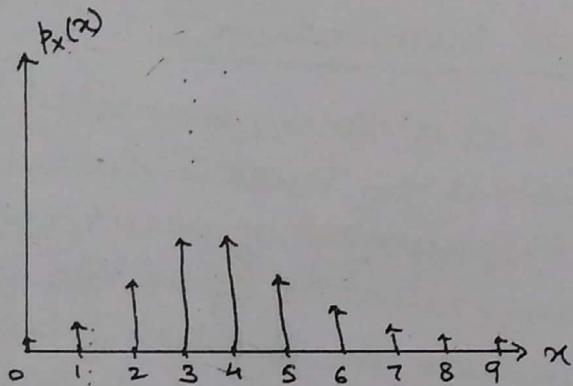
We observe that:

① If $x < \lambda$ then $p_x(x-1) < p_x(x)$ for any x .

② If $x > \lambda$ then $p_x(x-1) > p_x(x)$ for any x .

③ If $x = \lambda$ then $p_x(x-1) = p_x(x)$

Hence between $0 & \lambda$, the density function rises and reaches a peak value at λ . Beyond this value the function consistently falls till $x = \infty$.



Poisson Distribution.

The mathematical relationship between the Binomial and Poisson distributions can be mathematically established by what is called the Poisson Approximation theorem, which states: In a Binomial Distribution

If $n \rightarrow \infty$, $P \rightarrow 0$ such that $np \rightarrow \lambda$.

$$\text{then } n C_k P^k (1-P)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad k=0,1,2,\dots$$

Hence the mean of Poisson distribution

$$E[x] = \lambda (= np)$$

and the variance

$$\sigma_{x_i}^2 = \lambda (= np(1-p))$$

Some basic proofs

① Mean of Gaussian RV is \bar{x} .

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x \cdot \frac{e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx \\ &= \int_{-\infty}^{\infty} \frac{(x-\bar{x})}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} \frac{\bar{x}}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx. \end{aligned}$$

Assume
 $\frac{x-\bar{x}}{\sqrt{2}\sigma} = t$

$$\Rightarrow dx = \sqrt{2}\sigma dt.$$

$$\Rightarrow E(x) = \int_{-\infty}^{\infty} \frac{t}{\sqrt{\pi}} e^{-t^2} \cdot \sqrt{2}\sigma dt + \int_{-\infty}^{\infty} \frac{\bar{x}}{\sqrt{\pi}} e^{-t^2} dt.$$

The first integral is odd, and thus will
 simply be zero.
 over $-\infty$ to ∞

$$\Rightarrow E(x) = \frac{\bar{x}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt$$

$$= \frac{\bar{x}}{\sqrt{\pi}} \cdot \sqrt{\pi}$$

or $E(x) = \bar{x}$

(173)

④ Variance of Gaussian RV is σ^2

$$\begin{aligned}
 E[X^2] &= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} [(x-\bar{x}+\bar{x})^2 \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}] dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \left[\int_{-\infty}^{\infty} (x-\bar{x})^2 e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} \bar{x}^2 e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx + 2\bar{x} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \right] \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \left[\int_{-\infty}^{\infty} (x-\bar{x})^2 e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \right] - \frac{\bar{x}^2}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \\
 &\quad + \frac{2\bar{x}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\bar{x})^2 e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx - \bar{x}^2 + 2\bar{x}^2 \\
 &\quad \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx = 1 \right. \\
 &\quad \text{and } \left. \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} x e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx = \bar{x} \right) \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\bar{x})^2 e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx + \bar{x}^2
 \end{aligned}$$

Now assume $\frac{x-\bar{x}}{\sigma} = t$

$$\Rightarrow dx = \sqrt{2}\sigma dt$$

$$\begin{aligned} \Rightarrow E(x^2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} 2\sqrt{2}\sigma^3 t^2 e^{-t^2} dt + \bar{x}^2 \\ &= \frac{1}{\sqrt{\pi}} \frac{2\sqrt{2}\sigma^3}{\sqrt{2\sigma^2}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt + \bar{x}^2 \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \bar{x}^2 + \int_0^{\infty} t^2 e^{-t^2} dt + \bar{x}^2 \end{aligned}$$

Then by the general formula discussed earlier,

$$\begin{aligned} E(x^2) &= \frac{4\sigma^2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{4} + \bar{x}^2 \\ &= \sigma^2 + \bar{x}^2 \end{aligned}$$

Hence variance = $\sigma^2 + \bar{x}^2 - \bar{x}^2$

$$\text{or } \boxed{\sigma_x^2 = \sigma^2}$$

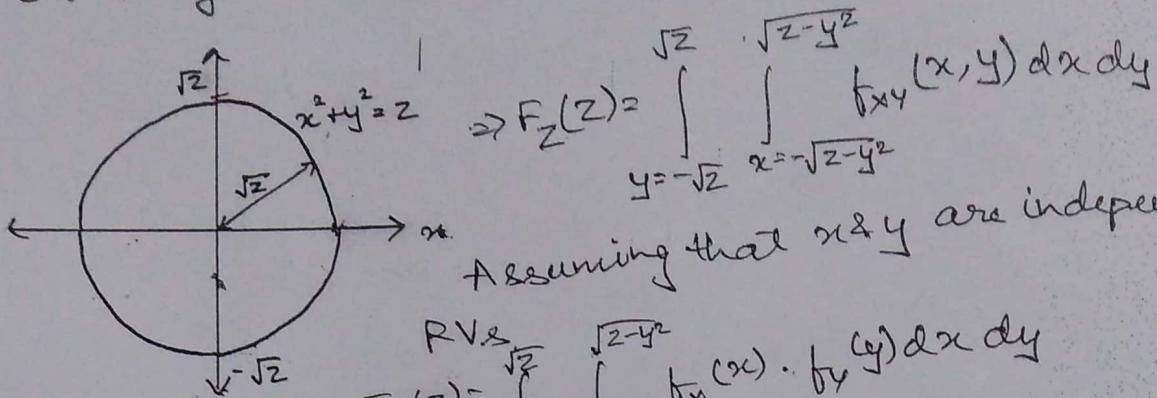
③ $Z = x^2 + y^2$ is an exponential distribution if x and y are independent Gaussian RVs with zero mean and common variance.

Ans The distribution function will be limited by the plane defined by x and y .

$$\Rightarrow F_Z(z) = P[x^2 + y^2 \leq z] = \iint_{x^2 + y^2 \leq z} f_{xy}(x, y) dx dy$$

(175)

Here $x^2 + y^2 \leq z$ represents a circle with radius \sqrt{z}



Assuming that x & y are independent

$$F_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \int_{x=-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_x(x) \cdot f_y(y) dx dy$$

The variables x and y are zero mean Gaussian with same frequency, thus

$$F_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \int_{x=-\sqrt{z-y^2}}^{\sqrt{z-y^2}} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy$$

Converting this expression into a density function, we have

$$f_Z(z) = \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\pi\sigma^2} \left[e^{-\frac{z}{2\sigma^2}} \right] \frac{1}{\sqrt{z-y^2}} dy$$

$$= \frac{e^{-z/2\sigma^2}}{2\pi\sigma^2} \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{z-y^2}} dy$$

$$= \frac{e^{-z/2\sigma^2}}{2\sigma^2} u(z)$$

$$\int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{z-y^2}} dy = \begin{cases} \frac{\pi}{2} \\ -\frac{\pi}{2} \end{cases} d\theta$$

Here $u(z)$ is a unit step function indicating the fact that z can never be negative

$$\Rightarrow f_Z(z) = \begin{cases} \frac{1}{2\sigma^2} e^{-z/2\sigma^2} & z \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

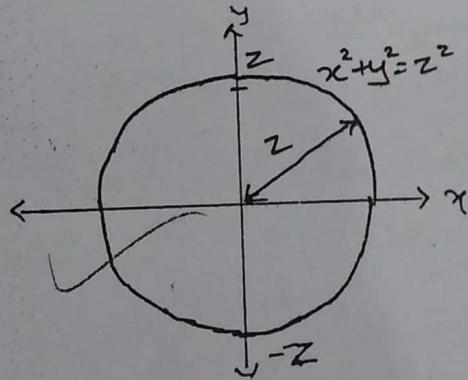
which is an exponential distribution with parameter $1/2\sigma^2$

④ For x and y being independent Gaussian RVs with zero mean and common variance, $z = \sqrt{x^2+y^2}$ will be a Rayleigh distribution.

Ans Again the distribution function of z will be

$$F_z(z) = \iint_{x^2+y^2 \leq z^2} f_{xy}(x, y) dx dy.$$

Here $x^2+y^2=z^2$ is a circle with radius z



This function can thus also be written as

$$F_z(z) = \int_{y=-z}^{z} \int_{x=\sqrt{z^2-y^2}}^{\sqrt{z^2-y^2}} f_{xy}(x, y) dx dy$$

The density function can be calculated as

$$f_z(z) = \frac{d F_z(z)}{dz} = \frac{d F_z(z)}{dx} \cdot \frac{dx}{dz}$$

$$= \int_{y=-z}^{z} \frac{z}{\sqrt{z^2-y^2}} [f_{xy}(\sqrt{z^2-y^2}, y) + f_{xy}(-\sqrt{z^2-y^2}, y)] dy$$

* Same argument can be used in the previous problem.

Also for two independent Gaussian RVs with zero mean and same variance

$$f_{xy}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}$$

$$\begin{aligned} \Rightarrow f_z(z) &= \int_{-z}^z \frac{1}{\sqrt{z^2-y^2}} \left[\frac{1}{2\pi\sigma^2} (2 \cdot e^{-\frac{y^2}{2\sigma^2}}) \right] dy \\ &= \frac{2z \cdot e^{-\frac{z^2}{2\sigma^2}}}{2\pi\sigma^2} \int_{-z}^z \frac{1}{\sqrt{z^2-y^2}} dy \\ &= \frac{2z \cdot e^{-\frac{z^2}{2\sigma^2}}}{\pi\sigma^2} \int_0^z \frac{1}{\sqrt{z^2-y^2}} dy \end{aligned}$$

Assuming $y = z \sin \theta$

$$dy = z \cos \theta d\theta$$

$$\Rightarrow f_z(z) = \frac{2z}{\pi\sigma^2} e^{-\frac{z^2}{2\sigma^2}} \int_0^{\frac{\pi}{2}} d\theta$$

$$= \frac{dz}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} \frac{1}{2} u(z)$$

$$= \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} u(z)$$

$$\Rightarrow f_z(z) = \begin{cases} \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}} & z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

which is a Rayleigh distribution with parameter σ^2 .

Random Process

Random variables were defined as a mapping between the sample space and the real number set \mathbb{R} .

In a similar fashion, Random processes are defined as a mapping from sample space to a set of functions over a variable. Most commonly used assumption is that this variable is time.

(* Another example of variable may be x by coordinates for an image)

Example: Coin toss

Random Variable

$$X: \begin{cases} H \rightarrow 0 \\ T \rightarrow 1 \end{cases}$$

Random Process

$$X: \begin{cases} t \mapsto \cos \omega_1 t & 0 \leq t \leq T \\ t \mapsto \cos \omega_2 t & 0 \leq t \leq T \end{cases}$$

Thus $X(t)$ is a collection of waveforms associated with $P = (\mathcal{R}, E, P)$

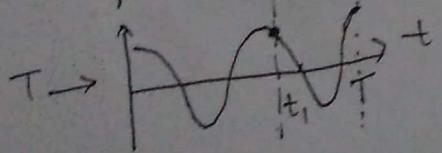
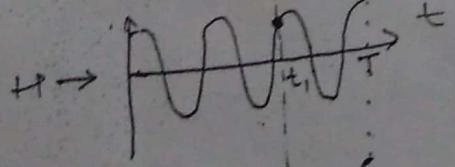
\mathcal{R} = Sample Space

E = Set of events

P = Probability set.

Another name for an RP is: Ensemble of Waveforms. Although this definition is convenient to observe a Random process as an extension of Random variable, it is not so convenient to model them mathematically this way. For that we need to define RPs in a different form.

Consider the example shown earlier.



Here at a given time t_1 , H and T will map to specific values i.e. $\cos \omega_1 t_1$ and $\cos \omega_2 t_1$. This mapping simply means that at time t_1 , X is a random variable. Similarly at any given instant of time, the value of a Random Process is a Random Variable.

(179)

Thus any Random process $X(t)$ can be defined as an indexed sequence of random variables.

Characterization of $X(t)$ (completely)
A Random Process can be characterized by the N pt. joint pdf of its amplitude at time instants $t_1, t_2, t_3, \dots, t_N$. This is clearly a horrendous job as for this we need to sample the RP at all possible joint instants t_1, t_2, \dots, t_N . Also we need to define pdfs for N variables where $N \in (0, \infty)$.

But what must be noted is that, for different values of N , the pdf must remain consistent.

This means that

(x_1, x_2) must have marginal pdfs as $P_{X(t_1), X(t_2)}(x_1, x_2)$. Similarly a pdf for 3 variables $P_{X(t_1)}(x_1), P_{X(t_2)}(x_2)$ must lead to the pdf for 2 variables as well as pdf for one variable.

Thus this is not a very convenient method unless we make some assumptions. Let us discuss some cases.

Stationarity:

Assuming that it is convenient to say that irrespective of where you sample the process, the pdf remains the same.

Thus instead of defining ∞ no. of first order pdfs we need only one.

Similarly, any second order pdf at instances t_1 and t_2 i.e. $P_{X(t_1), X(t_2)}(x_1, x_2)$ will depend only on the time difference and not on the time origin. This makes characterization very much simpler.

Formal definition of stationarity:

If the joint pdf of the N RV's is invariant under a shift of the time origin, the process is said to be statistically stationary, in the strict sense. Mathematically it can be stated as

$$(x_1, x_2, \dots, x_k)$$

$$P_{x(t_1), x(t_2), \dots, x(t_k)}(x_1, x_2, \dots, x_k)$$

$$= P_{x(t_1+T), x(t_2+T), \dots, x(t_k+T)}$$

Now, just like we could characterize a RV on the basis of their moments, it is similarly possible to characterize a Random Process in a similar way. Note here that for many RPs, it is sufficient to work with the first and second order density functions.

Let us examine the possibility of characterizing an RP on the basis of its moments:

Say $E[x(t)]$ is a mean value for $P_{x(t)}(x)$

Now if this process is stationary, $P_{x(t)}(x)$ must be independent of time t . Hence $E(x(t))$ must also have a value independent of t .

For a second order moment, $P_{x(t) x(t_2)}(x_1, x_2)$

is a complete characterization.

The possible second order moments are thus

$E[x(t_1) x(t_2)]$, $E[x^2(t_1)]$, $E[x^2(t_2)]$. Here $E[x(t_1)]$,

$E[x^2(t_2)]$ will be same for a stationary process.

$E[x^2(t_2)]$ will be time independent.

and both will be time independent.

$E[x(t_1) x(t_2)]$ is basically the correlation of $x(t_1)$ and $x(t_2)$. As these are the RVs of the same RP, this is called auto correlation

Thus we may define

$$R_x(t_1, t_2) = E[x(t_1)x(t_2)]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{x_1 x_2}(x_1, x_2; t_1, t_2) dx_1 dx_2$$

Here, if $p_{x_1 x_2}(x_1, x_2; t_1, t_2)$ is stationary, it will depend only on the difference of t_1, t_2 and not on the variable t .

$\therefore R_x(t_1, t_2)$ will depend only on $t_1 - t_2$ for a stationary process.

Mathematically

$$R_x(t_1, t_2) = R_x(t_1 - \tau, t_2) = R(\tau)$$

where τ is a time difference variable.
i.e. $\tau \equiv t_1 - t_2$

Now, what happens when $\tau = 0$

$$R_x(0) = E[x^2(t)] \text{ for any time instant } t.$$

= Mean square value of the process.

* Note that this holds only for stationary process. For a non stationary process

$$R_x(t, \tau) = E[x^2(t)] \text{ which is a function of time.}$$

For a stationary process

$$R_x(t, \tau) = R_x(0) = E[x^2(t)] \text{ which is a constant (independent of } t)$$

Properties of Auto correlation function

① $R_x(t_1, t_2) = R_x(t_2, t_1)$ as it basically only involves only multiplication of the two RVE and their pdf (joint).

In case the process is stationary, we observe,

$$R_x(\tau) = R_x(-\tau) \text{ where } \tau = t_1 - t_2.$$

② Correlation of two random variables is the maximum when the two are identical. Thus,

$$|R_x(t)| \leq R_x(0)$$

As discussed earlier, most of the processes can be sufficiently characterized simply by defining the first order pdfs and the second order pdfs. Defining pdfs of any higher order is rarely required. But as engineers, even this is too much to evaluate. This condition is called second order stationary process but it still requires infinite evaluations. Thus, evaluation of the first order and second order moments is sufficient to characterize a RVE as stationary in the widest sense. Hence such a process is called wide sense stationary.

Formal definition: Wide sense stationary Process

If mean, Variance are independent of time and if the Auto correlation function is a function only of $t_2 - t_1 \triangleq \tau$, the process is WSS.

$$\Rightarrow M_x^{(+)}) = E[x(t)] = \int_{-\infty}^{\infty} x \cdot p_{x(t)}(x, t) dx$$

$$\sigma_x^2(t) = E[(x(t) - \bar{x}(t))^2] = \int_{-\infty}^{\infty} (x - \bar{x}(t))^2 p_x(x; t) dx$$

$$R_x(t_1, t_2) = R_x(t_2 - t_1) \text{ all are independent of time.}$$

(183)

It is obvious to note that strict sense stationarity always implies wide sense stationarity. The reverse is very rarely true. Ex: For a Gaussian process, wide sense stationarity implies strict sense stationarity.

Prob: $x(t) = A \cos(\omega_0 t + \theta)$

This process has 3 variables, viz. A, ω_0 and θ .

Let us assume that θ is a random variable uniformly distributed between 0 and 2π .

$$\Rightarrow p(\theta) = \begin{cases} \frac{1}{2\pi} & 0 \leq \theta \leq 2\pi \\ 0 & \text{otherwise.} \end{cases}$$

* Here essentially we do not always require the definition of PDF to find the mean.

Now $E(x(t)) = \int_{-\infty}^{\infty} A \cos(\omega_0 t + \theta) \cdot \frac{1}{2\pi} d\theta$

$\therefore E(x(t))$ is independent of t .

$$E[x(t)x(t_2)] = E[x(t)x(t+\tau)]$$

$$= \int_{-\infty}^{\infty} A^2 \cos(\omega_0 t + \theta) \cos(\omega_0(t+\tau) + \theta) \cdot \frac{1}{2\pi} d\theta$$

$$= \frac{A^2}{4\pi} \left[\int_{-\infty}^{\infty} [\cos(\omega_0(2t+\tau+2\theta)) + \cos(\omega_0\tau)] d\theta \right]$$

$$= \frac{A^2}{4\pi} \left[2\pi \cos \omega_0 \tau + 0 \right]$$

$$= \boxed{\frac{A^2}{2} \cos \omega_0 \tau}$$

Hence $E[x(t)x(t_2)]$ is dependent only on the difference of t_1 & t_2 .

Frequency domain analysis of Random Process

As communication engineers, it is important for us to understand the behaviour of a process in frequency domain once its time domain analysis is understood. For a random process, analyzing spectral behaviour is important to understand some parameters that may be used to characterize them.

Let us recall the notion of Fourier transform.

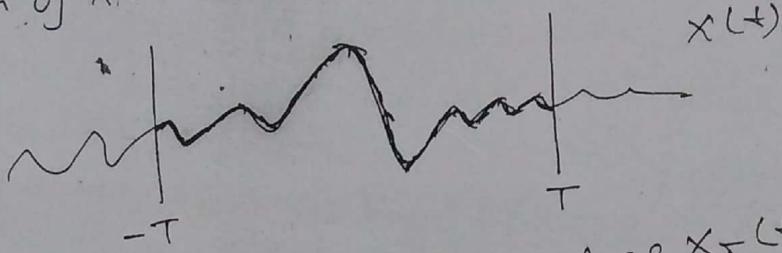
$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

The first condition for $x(t)$ is to be absolutely integrable, or $x(t)$ should be an energy signal. So the first difficulty is that we do not know whether our Random signal is an energy signal or a power signal. This difficulty is also faced for deterministic signals, though, and we can intuitively get around it by introducing impulse signals in the frequency domain.
A bigger problem here is that $x(t)$ being a random process, we have infinite set of possible waveforms in $x(t)$ and each will have a different Fourier transform. Thus it does not make any sense to talk about frequency domain representation in the general sense for Random processes. Instead, we may define the Energy (or Power, as applicable) as a function of frequency to understand the behaviour of a RP over the spectrum. Thus the important frequency domain function for random processes is the Power Spectral density (PSD).

(185)

Power Spectral Density

Let's say we have a random process $x(t)$. If we take out the portion of this $x(t)$ between $-T$ to T , this signal will have to be an energy signal. Such a truncated portion of $x(t)$ is called a sample function of $x(t)$.



This sample function is termed as $x_T(t)$. It may be mathematically defined as

$$x_T(t) = \begin{cases} x(t) & -T \leq t \leq T \\ 0 & \text{otherwise.} \end{cases}$$

Thus this will be a finite energy signal and hence its Fourier Transform will be defined. The Fourier transform may be denoted as

$$\boxed{F[x_T(t)]}$$

The Energy value of this representation can be found by evaluating the square of this quantity i.e. $|F[x_T(t)]|^2$. Note that this value is only for one sample function. An average value may be evaluated as the expected value of $|F[x_T(t)]|^2$ i.e. $E[|F[x_T(t)]|^2]$. This average energy can further be denoted as the power as

$\frac{1}{2T} E[|F[x_T(t)]|^2]$. Further more, the original function can be taken into account by evaluating this expression with $\lim_{T \rightarrow \infty}$

Hence, the expression for power spectral density is defined as

$$S_x(f) \triangleq \lim_{T \rightarrow \infty} \frac{1}{2\pi} E \left[|F(x_T(t))|^2 \right]$$

If we observe this calculation carefully, we may note that $S_x(f)$ is basically a second order moment of $x_T(t)$ in frequency domain. Thus it is natural to expect a relation between PSD and autocorrelation function which is also a second order moment. Upon inspection, it may be derived that $S_x(f)$ and $R_x(\tau)$ are Fourier transform pairs. This is known as Wiener-Khinchin's theorem and it states that

$$S_x(f) \xleftrightarrow{\text{FT}} R_x(\tau)$$

Ergodicity

A random process is ergodic if its statistical averages can be replaced with (equal to) its time average.

\Rightarrow Statistical Averages = Time Average for any sample function.

Let us first discuss the motivation behind the need of this concept. As is the case with many practical systems, let us assume that we have no knowledge of the distribution function of the RP. Then for the one sample function that we may observe, how can we evaluate the mean or covariance or PSD? In such a case if the process is ergodic, then the time average of a single sample function can be treated as (187) statistical averages.

Fortunately for us, most practical processes are found to be ergodic in nature.

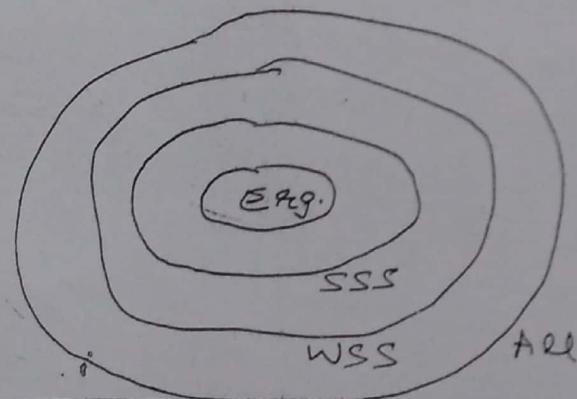
A time domain average may be defined as

$$\bar{x}(t) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt$$

Then by ergodicity property

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt = \int_{-\infty}^{\infty} x P_x(x; t) dt$$

* If we represent the class of Random Process in the form of a Venn diagram, then



Let us examine the properties of an Autocorrelation function in a little more detail.

$R_x(\tau)$: Properties.

For a real $x(t)$

① $R_x(\tau)$ is real valued.

② $R_x(\tau) = R_x(-\tau)$ i.e. even function

For a complex $x(t)$

$R_x(\tau) = R_x^*(\tau)$ i.e. conjugate symmetry

③ $R_x(\tau) \leq R_x(0)$ i.e. max at $\tau=0$

④ For periodic $x(t)$, $R_x(\tau)$ is also periodic.

⑤ $F[R_x(\tau)]$ will always be real valued and non-negative for all frequencies.

* This may serve as one of the tests to check if a function is an autocorrelation function

⑥ $R_x(0) = \sigma^2 = \text{Total Average Power.}$

$$\text{Proof: } R_x(\tau) = \int_{-\infty}^{\infty} S_x(f) e^{j2\pi f\tau} df.$$

$\Rightarrow R_x(0) = \int_{-\infty}^{\infty} S_x(f) df = \text{Net power of the signal.}$

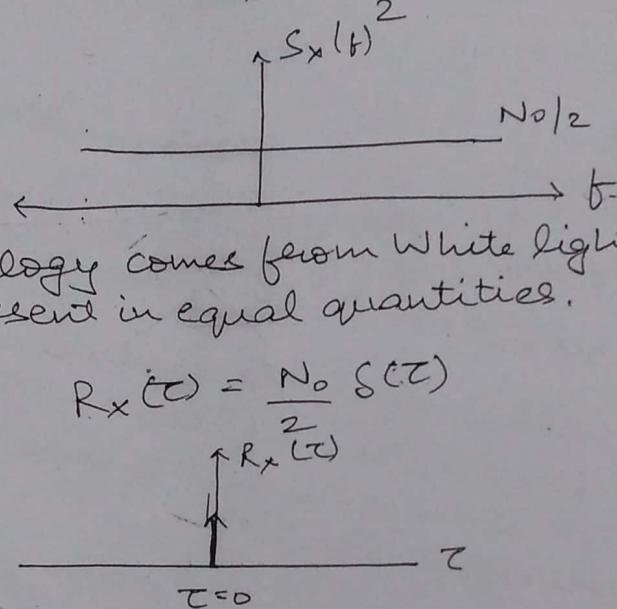
⑦ $\lim_{\tau \rightarrow \pm\infty} R_x(\tau) = \mu_x^2$ (Note: This property is basically only for non-periodic processes)

Let us further discuss some concepts which are of typical relevance to us in context of communication.

White Noise Process

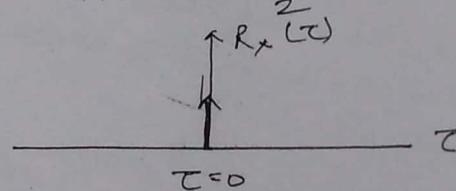
White Noise Process is one for which the PSD is constant for all f .

$$\text{i.e. } S_x(f) = \frac{N_0}{2} \text{ for all } f.$$



* The analogy comes from White light as all frequencies are present in equal quantities.

$$R_x(\tau) = \frac{N_0}{2} S(\tau)$$

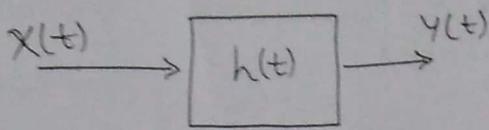


This is an idealization, though, and it can never practically exist. Reasons being:

- (1) Not power of this process from PSD is seen to be ∞ which is not practically possible.
- (2) Impulse function representation of $R_x(\tau)$ is also impractical due to lack of flexibility.

Hence, there can never be an ideal White process. But as seen in communication, the thermal noise which is present in almost all systems is seen to have a constant power value over a very large range of frequencies. Thus for all practical purposes, such a noise may be termed as a White Noise process.

Transmission of a Random Process through a linear
 $X(t)$ is a random process. Let us examine the properties of $y(t)$



$$\begin{aligned} E[Y(t)] &= E \left[\int_{-\infty}^{\infty} h(\tau) x(t-\tau) d\tau \right] \\ &= \int_{-\infty}^{\infty} h(\tau) E[x(t-\tau)] d\tau \\ &= \int_{-\infty}^{\infty} h(\tau) \mu_x(t-\tau) d\tau \end{aligned}$$

Normally the discussion of a random signal in communication is done for WSS processes. Hence, μ_x will be a constant and it can replace $\mu_x(t-\tau)$.

$$\Rightarrow E[Y(t)] = \int_{-\infty}^{\infty} h(\tau) \mu_x d\tau \\ = \mu_y \int_{-\infty}^{\infty} h(\tau) d\tau.$$

$$\Rightarrow \mu_y = \mu_x \int_{-\infty}^{\infty} h(\tau) d\tau$$

or $\boxed{\mu_y = \mu_x H(0)}$ (value of $H(f)$ at $f=0 \equiv H(0)$.
 $H(0)$ is also called the DC component of $H(f)$)

\Rightarrow If μ_x is time independent, then μ_y is also time independent.

* By Fourier transform

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi f t} dt$$

$$\therefore \boxed{H(0) = \int_{-\infty}^{\infty} h(t) dt}$$

(191)

Autocorrelation function

For $y(t)$, the ACF may be defined as

$$R_y(t, u) = E[y(t) y(u)]$$

$$\Rightarrow R_y(t, u) = E \left[\int_{-\infty}^{\infty} h(\tau_1) x(t - \tau_1) d\tau_1 \cdot \int_{-\infty}^{\infty} h(\tau_2) x(u - \tau_2) d\tau_2 \right]$$

$$= \int_{-\infty}^{\infty} h(\tau_1) \left[\int_{-\infty}^{\infty} h(\tau_2) E[x(t - \tau_1) x(u - \tau_2)] d\tau_2 \right] d\tau_1$$

If we have a WSS $x(t)$, then

$$E[x(t - \tau_1) x(t - \tau_2)] = R_x(t - \tau_1, u - \tau_2) = R_x(t - u + \tau_2 - \tau_1)$$

$$\Rightarrow R_y(t, u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\tau_1) h(\tau_2) R_x(\tau - \tau_1 + \tau_2) d\tau_1 d\tau_2 \quad \text{where } \tau = t - u$$

Upon integration, τ_1 & τ_2 components will disappear from the expression, leaving $R_x(t, u)$ to be a function only of $t - u$.

Therefore, a wide sense stationary signal when passed through an LTI filter, the output is again WSS.

Exercise: If cross correlation between two random processes is defined as

$$R_{xy}(\tau) = R_{xy}(t, t+\tau) = E[x(t)y(t+\tau)],$$

then verify:

$$\textcircled{1} R_{yx}(\tau) = R_x(\tau) * h(-\tau)$$

$$\textcircled{2} R_{xy}(-\tau) = R_x(\tau) * h(\tau)$$

which implies that

$$R_{xy}(\tau) = R_{yx}(-\tau)$$

Also verify that

$$\textcircled{3} R_y(\tau) = R_{yx}(\tau) * h(\tau)$$

$$\underline{\text{Ans}} \quad R_{yx}(\tau) = R_{yx}(t, t+\tau) = E[y(t+\tau)x(t+\tau)]$$

$$= E\left[\left(\int_{-\infty}^{\infty} h(\tau_1)x(t-\tau_1)d\tau_1\right)x(t+\tau)\right]$$

$$= E\left[\left(\int_{-\infty}^{\infty} h(\tau_1)x(t-\tau_1)x(t+\tau)d\tau_1\right)\right]$$

$$= \int_{-\infty}^{\infty} h(\tau_1) E\{x(t-\tau_1)x(t+\tau)\} d\tau_1$$

$$= \int_{-\infty}^{\infty} h(\tau_1) R_x(-\tau_1 - \tau) d\tau_1$$

$$= \int_{-\infty}^{\infty} h(-\tau) R_x(\tau_1 - \tau) d\tau.$$

$$\Rightarrow \boxed{R_{yx}(\tau) = R_x(\tau) * h(-\tau)} \quad \textcircled{1}$$

(193)

Similarly

$$R_{xy}(\tau) = R_{xy}(t, t+\tau) = E[x(t) y(t+\tau)]$$

$$= E[x(t) h(t+\tau) x(t-\tau_1) d\tau_1]$$

$$= E \left[x(t) \int_{-\infty}^{\infty} h(\tau_1) x(t+\tau-\tau_1) d\tau_1 \right]$$

$$= E \left[\int_{-\infty}^{\infty} h(\tau_1) x(t) x(t+\tau-\tau_1) d\tau_1 \right]$$

$$= \int_{-\infty}^{\infty} h(\tau_1) E[x(t+\tau-\tau_1) \cdot x(t)] d\tau_1$$

$$= \int_{-\infty}^{\infty} h(\tau_1) R_x(\tau-\tau_1) d\tau_1$$

$$\boxed{R_{xy}(\tau) = R_x(\tau) * h(\tau)} \quad \textcircled{2}$$

It can be observed that.

$$R_{yx}(\tau) = R_{xy}(-\tau)$$

Proof: $R_{xy}(-\tau) = R_x(-\tau) * h(-\tau)$

As $R_x(\tau)$ is an even function

$$R_x(\tau) = R_x(-\tau)$$

$$\Rightarrow \boxed{R_{yx}(\tau) = R_{xy}(-\tau)}$$

Now, as was observed earlier,

$$R_y(t, t+\tau) = \iint_{-\infty}^{\infty} h(\tau_1) h(\tau_2) R_x(\tau - \tau_1 + \tau_2) d\tau_1 d\tau_2$$

$$= \int_{-\infty}^{\infty} h(\tau_2) \left[\int_{-\infty}^{\infty} h(\tau) R_x((\tau + \tau_2) - \tau_1) d\tau_1 \right] d\tau_2$$

$$= \int_{-\infty}^{\infty} h(\tau_2) [h(\tau + \tau_2) * R_x(\tau + \tau_2)] d\tau_2$$

$$\Rightarrow R_y(\tau) = \int_{-\infty}^{\infty} [h(\tau') \otimes R_x(\tau')] h(\tau - \tau') d\tau'$$

$$= R_x(\tau) * h(\tau) + h(-\tau)$$

$$= R_x(\tau) * h(-\tau) * h(\tau)$$

$$R_y(\tau) = R_{yx}(\tau) * h(\tau) \quad \text{--- (3)}$$

These relations are very useful practically.

$R_{xy}(\tau) = R_x(\tau) * h(\tau)$

means that if we cross correlate the output and input signals, we get the convolution of the autocorrelation of input and the impulse response of the system. In certain systems where the $h(t)$ is not known, for eg. effect of channel on a signal, this relation can be used to make a good approximation.

Here if we use white noise as the input, we may observe that the cross correlation of the input and output will give us the value of $h(t)$ scaled by a factor of $\frac{N_o}{2}$. This is because white noise has an

ACF of $\frac{N_o}{2} \delta(t)$.

This means $R_{xy}(\tau) = \frac{N_o}{2} S_x(\tau) * h(\tau)$

$$= \frac{N_o}{2} h(\tau)$$

$$R_x(\tau) = R_x(\tau) * h(-\tau) * h(\tau)$$

Fourier transform of this equation leads to

$$S_y(f) = S_x(f) \cdot H^*(f) \cdot H(f)$$

$$S_y(f) = S_x(f) \cdot |H(f)|^2$$

(195)

Gaussian Random Process

Recall, that a Gaussian RV is one for which pdf $P_x(x)$ is

$$P_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

Also recall that a linear combination of Gaussian RV is again a Gaussian RV. This may be defined in another way if X_1, X_2, \dots, X_n are jointly Gaussian, if

RVs X_1, X_2, \dots, X_n are jointly Gaussian, if $\{g_i\}$

$$Y = \sum_{i=1}^n g_i X_i$$

Then we may generalize the above expression to define a Gaussian Process

$X(t)$ is Gaussian if

$y \triangleq \int_0^t g(t) X(t) dt$ is a Gaussian RV for every choice of $g(t)$ such that $\int_0^\infty |g(t)|^2 dt < \infty$ (Finite energy).

Recall: Many physical phenomena are modelled as Gaussian RPs because of a result obtained from the central limit theorem.

Central limit theorem

Let X_i be statistically independent RVs identical in nature (i.e. having identical pdfs) i.e. Mean = μ_x & variance = σ_x^2

For the purpose of normalization we may define

$$y_i = \frac{x_i - \mu_x}{\sigma_x} \quad (\text{For 0 mean & unit variance } y_i = x_i)$$

Also, we may define

$$V_N = \frac{1}{\sqrt{N}} \sum_{i=1}^N y_i$$

Then CLT states that as $N \rightarrow \infty$, N_N approaches a Gaussian RV with pdf: $N(0, 1)$

Small Exercise:

If X_1 and X_2 are 2 s.i. RVs, then for an RV

$$Y = X_1 + X_2$$

the pdf is defined as

$$p_Y(y) = p_{X_1}(y) * p_{X_2}(y).$$

Prove this.

When operating on 2 RVs, the pdf used is the joint pdf of the 2.

$$\Rightarrow \cancel{p_Y(y)} \approx p_{X_1, X_2}(x_1, x_2)$$

The distribution function is further defined as

$$P_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_2} p_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

As X_1 & X_2 are s.i. RVs \therefore

$$p_{X_1, X_2}(x_1, x_2) = p_{X_1}(x_1) p_{X_2}(x_2)$$

$$\Rightarrow P_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_2} p_{X_1}(x_1) p_{X_2}(x_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{y-x_2} p_{X_1}(x_1) p_{X_2}(x_2) dx_1 dx_2$$

~~We may use the assumption~~

$$x_2 = y - x_1$$

$$= \int_{-\infty}^{y-x_2} \left[\int_{-\infty}^{\infty} p_{X_1}(x_1) p(y-x_1) dx_1 \right] dy$$

The pdf may be defined as

$$p_Y(y) = \frac{d}{dy} \int_{-\infty}^{y-x_2} [p_{X_1}(x_1) * p_{X_2}(x_2)] dx_1$$

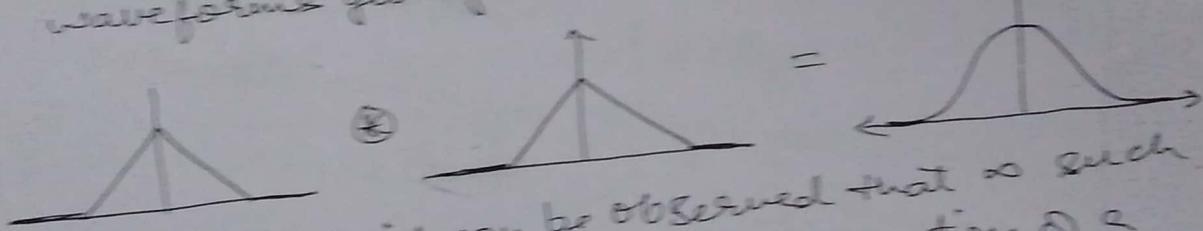
(197)

$$\Rightarrow p_y(y) = p_{x_1}(y) \oplus p_{x_2}(y)$$

Using this result, say we have x_1, x_2 as uniform RVS. \therefore their pdfs are rectangular functions. Hence the convolution of the two will be a triangular function.



Similarly, if there are 4 such variables, the net effect will be the convolution of the 2 triangular waveforms giving rise to a parabolic shape.



Thus intuitively it can be observed that such convolutions will lead to the formation of a Gaussian PDF.

Properties of Gaussian Processes

If $x(t)$ is passed through a stable linear filter then the output $y(t)$ is also Gaussian.

Proof: Let us say that the impulse response of the filter is $h(t)$.

$$\Rightarrow y(t) = \int_{-\infty}^t h(t-\tau) x(\tau) d\tau \quad 0 \leq t \leq \infty$$

Now, to check whether $y(t)$ is a Gaussian process, we use the basic definition of Gaussian process, i.e. $y(t)$ is Gaussian RP if

$$Z = \int g(t) y(t) dt \text{ is a GRP.}$$

$$\begin{aligned} Z &= \int_0^T g_y(t) \int_{-\infty}^{\infty} h(t-\tau) x(\tau) d\tau dt \\ &= \int_{-\infty}^T x(\tau) \int_{-\infty}^{\infty} g_y(t) h(t-\tau) dt d\tau \\ &= \int_0^T g(t) x(t) dt \end{aligned}$$

where $g(t) = \int_{-\infty}^{\infty} g_y(t) h(t-\tau) d\tau$
 \Rightarrow If $x(\tau)$ is a Gaussian Random Process, then Z is a GRV.
 \Rightarrow If $x(\tau)$ is a Random Process, then $y(t)$ is a GRP.
 Which simply implies that $y(t)$ is a GRP.

(2) Consider the set of random variables or samples $x(t_1), x(t_2), x(t_3), \dots, x(t_n)$ obtained by observing a Random Process $x(t)$ at times $t_1, t_2, t_3, \dots, t_n$. If the process $x(t)$ is Gaussian, then the set of RV's is jointly Gaussian for any n , with their n -fold joint probability density function being completely determined by specifying the set of means

$$m_{x(t_i)} = E(x(t_i)) \quad i=1, 2, \dots, n$$

and the set of covariance functions

$$C_{x(t_k, t_i)} = E((x(t_k) - m_{x(t_k)})(x(t_i) - m_{x(t_i)})) \quad k \neq i = 1, 2, \dots, n$$

199

③ If a Gaussian Process is stationary, then the process is also strictly stationary

④ If the random variables $x(t_1), x(t_2) \dots x(t_n)$ obtained by sampling a Gaussian Process $x(t)$ at times t_1, t_2, \dots, t_n are uncorrelated, that is

$$E[(x(t_k) - \mu_{x(t_k)})(x(t_i) - \mu_{x(t_i)})] = 0$$

then these RVs are statistically independent.

Proof: Say there are N jointly Gaussian RVs, x_1, x_2, \dots, x_N . Assume initially that they are s.i. Corresponding means & variances are

$$\mu_i, \sigma_{x_i} \quad i=1, 2, \dots, N.$$

The joint pdf is defined as

$$P_{x_1 x_2 x_3 \dots x_N}(x_1, x_2, \dots, x_N)$$

$$= \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma_{x_i}}} e^{-\frac{(x_i - \mu_{x_i})^2}{2\sigma_{x_i}^2}}$$

Let us use vector notations to describe this.

First of all the covariance b/w $(x_i - \mu_{x_i})$ &

~~$(x_k - \mu_{x_k})$~~

$$C_{x(i,j)} = E[\del{(x_k - \mu_{x_k})(x_i - \mu_{x_i})}]$$

$$\text{But for } i \neq k, E[(x_k - \mu_{x_k})(x_i - \mu_{x_i})]$$

$$= E[(x_k - \mu_{x_k})] \cdot E[x_i - \mu_{x_i}] \\ = 0 \quad (\text{first central moment})$$

and for $i=k$

$$C_{x(i,i)} = E[(x_i - \mu_{x_i})^2] = \sigma_i^2$$

Hence for covariance of 2 non identical RVs, ~~to be~~ to be ~~0~~, the 2 RVs are found to be statistically independent. Hence Property 4 is proved.

Let us denote the covariance function in terms of a matrix.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_N^2 \end{bmatrix}$$

In case of N RVs

The N RVs can be defined as the column vector

$$x = [x_1 \ x_2 \ \dots \ x_N]^T$$

Set of all means is again a column vector

$$\mu_i = [\mu_{x1} \ \mu_{x2} \ \dots \ \mu_{xN}]^T$$

Hence, the joint pdf can be defined as

$$P_{x_1 x_2 x_3 \dots x_N} (x_1, x_2, x_3, \dots, x_N) = e^{-\frac{1}{2} \sum_{i=1}^N \frac{(x_i - \mu_i)^2}{\sigma_i^2}}$$

$$= \frac{1}{(2\pi)^{N/2} (\sigma_1^2 \sigma_2^2 \dots \sigma_N^2)^{1/2}}$$

The exponential summation can be written as

$$(x - \mu_i)^T \sum^{-1} (x - \mu_i)$$

$$(x - \mu_i) = [x_1 - \mu_{x1} \ x_2 - \mu_{x2} \ \dots \ x_N - \mu_{xN}]^T$$

$$\sum^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & & \\ 0 & \frac{1}{\sigma_2^2} & \dots & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_N^2} \end{bmatrix}$$

$$\text{Also } \Delta = \det(\Sigma) = \sigma_1^2 \sigma_2^2 \dots \sigma_N^2$$

(201)

$$\Rightarrow P_{x_1 x_2 \dots x_N} (x_1, x_2, \dots, x_N) \\ = \frac{1}{(2\pi)^{N/2} |\Delta|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu_i)^T \Sigma^{-1} (x - \mu_i)\right)$$

Here, this equation is the general form of the joint pdf. The covariance matrix Σ will be a diagonal matrix only for s.i. RVs. In other cases it will be a filled matrix. Even in such a case the exponent will be a scalar that has all possible components of the joint RV.

Thus it is possible for the joint RV to be defined completely by defining the set of means μ_i and the set of covariance values Σ . Hence property 2 is proved.

Note that if these two sets are found to be stationary, the complete pdf defining the process is stationary. This means that all averages will be stationary. Therefore if the Gaussian process is stationary, the process is also strict sense stationary.

Wiener Process

Wiener Process is one of the most well known continuous time Random processes and is used to model a number of applications in pure as well as applied mathematics. In communication theory, Wiener Process is used to represent the integral of a Gaussian White ^{noise} Process making it an important model of noise in electronic engineering.

Wiener Process is characterized by the following properties. Let us say $W(t)$ is the Wiener process.

then:

- ① $W(0) = 0$
 - ② $W(t)$ is continuous everywhere.
 - ③ $W(t)$ has independent increments with $W(t) - W(s)$ defined as $N(0, t-s)$ distributed.
- This last property defines any arbitrary increment as a zero mean, $t-s$ variance, Normal or Gaussian distribution. Also, if we consider $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2$, then $W(t_1) - W(s_1)$ and $W(t_2) - W(s_2)$ are independent Random Variables.

Using the above properties, the probability density function of $W(t)$ can be nearly approximated as

$$f_{W(t)-W(s)}(x) = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{x^2}{2(t-s)}}$$

or
$$f_{W(t)}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

This clearly shows that

$$E[W(t)] = 0$$

$$\text{and } \text{Var}(W(t)) = t$$

$$\text{or } E[W^2(t)] = t$$

Thus it is evident that the moments of this random process are time dependents and hence Wiener process is Non-Stationary.

It can further be deduced that covariance of two Wiener random variables derived from a Wiener Random process sampled at time instances s & t is given by

$$\text{cov}(W(s), W(t)) = \min(s, t)$$

As the process is zero mean, using the equations derived earlier we may deduce that

$$\text{cov}(W(s), W(t)) = \text{corr}(W(s), W(t)).$$

Similarly, the correlation coefficient is given

by

$$f_{st} = \frac{\text{cov}(W(t), W(s))}{\sigma_{W(t)} \cdot \sigma_{W(s)}}$$

$$f_{st} = \frac{\min(s, t)}{\sqrt{st}}$$

* Wiener Process in communication may be characterized as a definite integral, from 0 to t , of a zero mean, unit variance ($N(0, 1)$) white Gaussian Process

Poisson Process

The poisson distribution is commonly used to describe an experiment called the Poisson Points Experiment.

In this experiment, the number of points n in the interval (t_1, t_2) is a poisson Random variable where

$$P[n(t_1, t_2) = k] = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$

Here $k = 0, 1, 2, \dots$

λ = rate of the process

$t = t_2 - t_1$ = Length of interval.

Also if the intervals (t_1, t_2) and (t_3, t_4) are non overlapping, then the random variables $n(t_1, t_2)$ and $n(t_3, t_4)$ are ~~indepe~~ independent.

Thus using these points we may define a Random Process, $x(t) = n(0, t)$

$$\Rightarrow x(t) = e^{-\lambda t} \cdot \frac{(\lambda t)^t}{t!}$$

The mean value of this process is therefore

$$E[x(t)] = \lambda t$$

and variance $\sigma_{x(t)}^2 = \lambda t$.

As both are time dependent, this process is non stationary.

(205)

The autocorrelation function of this process,

$$R_x(t_1, t_2) = E[x(t_1)x(t_2)]$$

$$= E[x(t_1)[x(t_2) - x(t_1)] + x(t_1)]$$

$$\begin{aligned} & \text{No of points between } t_2 \text{ & } t_1 \\ & = \text{No of points from } 0 \text{ to } t_2 \\ & - \text{No of points from } 0 \text{ to } t_1 \end{aligned}$$

$$R_x(t_1, t_2)$$

$$\Rightarrow R_x(t_1, t_2) = E[R_x(t_1)x(t_2) - E[x(t_1)]]$$

$$\Rightarrow R_x(t_1, t_2) = E[x^2(t_1)] + E[x(t_1)(x(t_2) - x(t_1))]$$

Assuming $t_2 > t_1$, $x(t_1)$ & $x(t_2) - x(t_1)$ are independent,

$$R_x(t_1, t_2) = E[x^2(t_1)] + E[x(t_1) \cdot x(t_2 - t_1)]$$

$$\begin{aligned} & \text{No of points between } t_2 \text{ & } t_1 \\ & = \text{No of points from } (0, t_2) \\ & - \text{No of points from } (0, t_2) \end{aligned}$$

$$\Rightarrow R_x(t_1, t_2) = E[x^2(t_1)] + E[x(t_1)] \cdot E[x(t_2 - t_1)]$$

$$= \lambda^2 t_1^2 + \lambda t_1 + \lambda(t_1)(\lambda(t_2 - t_1))$$

$$R_x(t_1, t_2) = \lambda t_1 + \lambda^2 t_1 t_2$$

In case $t_1 > t_2$,

$$R_x(t_1, t_2) = R_x(t_2, t_1) = \lambda t_2 + \lambda^2 t_1 t_2$$

$$\Rightarrow R_x(t_1, t_2) = \begin{cases} \lambda t_2 + \lambda^2 t_1 t_2 & t_1 > t_2 \\ \lambda t_1 + \lambda^2 t_1 t_2 & t_1 \leq t_2 \end{cases}$$

Appendix B

Mathematical Models for Communications Channels

Published Date: May 07, 2014

Overview

This tutorial is part of the National Instruments Signal Generator Tutorial series. Each tutorial in this series, will teach you a specific topic of common measurement applications, by explaining the theory and giving practical examples. This tutorial covers the mathematical models for communication channels. For additional signal generator concepts, refer to the Signal Generator Fundamentals main page.

Table of Contents

1. The Additive Noise Channel
2. The Linear Filter Channel
3. The Linear Time-Variant Filter Channel
4. Relevant NI products

1. The Additive Noise Channel

In the design of communication systems for transmitting information through physical channels, we find it convenient to construct mathematical models that reflect the most important characteristics of the transmission medium. Then, the mathematical model for the channel is used in the design of the channel encoder and modulator at the transmitter and the demodulator and channel decoder at the receiver. Next, we provide a brief description of the channel models that are frequently used to characterize many of the physical channels that we encounter in practice.

The Additive Noise Channel.—The simplest mathematical model for a communication channel is the additive noise channel, illustrated in Figure 1.8. In this model the transmitted signal $s(t)$ is corrupted by an additive random noise process

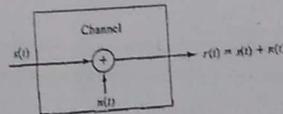


Figure 1.8—The additive noise channel.

Physically, the additive noise process may arise from electronic components and amplifiers at the receiver of the communication system, or from interference encountered in transmission, as in the case of radio signal transmission.

If the noise is introduced primarily by electronic components and amplifiers at the receiver, it may be characterized as thermal noise. This type of noise is characterized statistically as a Gaussian noise process. Hence, the resulting mathematical model for the channel is usually called the additive Gaussian noise channel. Because this channel model applies to a broad class of physical communication channels and because of its mathematical tractability, this is the predominant channel model used in our communication system analysis and design. Channel attenuation is easily incorporated into the model. When the signal undergoes attenuation in transmission through the channel, the received signal is

$$r(t) = as(t) + n(t) \quad (1.4.1)$$

where a represents the attenuation factor.

2. The Linear Filter Channel

In some physical channels such as wireline telephone channels, filters are used to ensure that the transmitted signals do not exceed specified bandwidth limitations and, thus, do not interfere with one another. Such channels are generally characterized mathematically as linear filter channels with additive noise, as illustrated in Figure 1.9. Hence, if the channel input is the signal $s(t)$, the

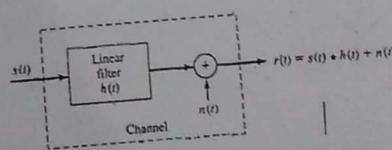


Figure 1.9—The linear filter channel with additive noise.

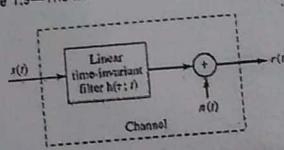


Figure 1.10—Linear time-variant filter channel with additive noise.

channel output is the signal

where $h(t)$ is the impulse response of the linear filter and $*$ denotes convolution.

3. The Linear Time-Variant Filter Channel

$$\begin{aligned} r(t) &= s(t) * h(t) + n(t) \\ &= \int_{-\infty}^{+\infty} h(\tau) s(t - \tau) d\tau + n(t) \end{aligned} \quad (1.4.2)$$

1/2

Physical channels such as underwater acoustic channels and ionospheric radio channels which result in time-variant multipath propagation of the transmitted signal may be characterized mathematically as time-variant linear filters. Such linear filters are characterized by time-variant channel impulse response $h(t; \tau)$ where $h(t; \tau)$ is the response of the channel at time t , due to an impulse applied at time τ . Thus, τ represents the "age" (elapsed time) variable. The linear time-variant filter channel with additive noise is illustrated Figure 1.11. For an input signal $s(t)$, the channel output signal is

$$\begin{aligned} r(t) &= s(t) * h(t; \tau) + n(t) \\ &= \int_{-\infty}^{+\infty} h(t; \tau)s(t - \tau) d\tau + n(t) \end{aligned} \quad (1.4.3)$$

A good model for multipath signal propagation through physical channels, such as the ionosphere (at frequencies below 30 MHz) and mobile cellular radio channels, is a special case of Equation (1.4.3) in which the time-variant impulse response has the form

$$h(t; \tau) = \sum_{k=1}^L a_k(\tau) \delta(t - \tau_k) \quad (1.4.4)$$

where the $\{a_k(\tau)\}$ represent the possibly time-variant attenuation factors for the L multipath propagation paths. If Equation (1.4.4) is substituted into Equation (1.4.3), the received signal has the form

$$r(t) = \sum_{k=1}^L a_k(\tau_k) s(t - \tau_k) + n(t) \quad (1.4.5)$$

Hence, the received signal consists of L multipath components, where each component is attenuated by $\{a_k\}$ and delayed by $\{\tau_k\}$.

The three mathematical models described above adequately characterize a large majority of physical channels encountered in practice. These three channel models are used in this text for the analysis and design of communication systems.

4. Relevant NI products

Customers interested in this topic were also interested in the following NI products:

- Function, Arbitrary, and RF Signal Generators
- Other Modular Instruments (digital multimeters, digitizers, switching, etc...)
- LabVIEW Graphical Programming Environment
- SignalExpress Interactive Software Environment

For the complete list of tutorials, return to the NI Signal Generator Fundamentals Main page.

Publication Information

Author: John G. Proakis and Masoud Salehi

Book: Communications Systems Engineering

Copyright: 2002

ISBN: 0-13-061793-8

Legal Notes

Excerpt from the book published by Prentice Hall Professional (<http://www.pphc.com>). Copyright Prentice Hall Inc. 2002. All rights reserved.