Introduction

In this set of notes, we are going to give a brief introduction to important concepts in communication theory. These notes are not meant to replace any textbook. They are given essentially to clarify concepts that are believed to be important. The course will give a great emphasis on digital communication. In this course also, emphasis will be given to randomness in communication. The concept of randomness is fundamental in communication.

We will introduce the concept of random signals. These signals are always present in any communication system, usually in the shape of noise. However, there is a more profound reason for studying randomness. The concept of information is itself random in nature. The reason is in the use of information. We say that a message brings some information if its content is not known in advance by the receiver. If the receiver knows in advance the message that will be transmitted, there is no point in transmitting that massage.

Chapter 1

Review on probability and random variables

1.1 Random experiment, sample space and events

An experiment is determined by its outcomes. If the outcome of the experiment can be predicted exactly, the experiment is said to be deterministic. If the outcome cannot be predicted, the experiment is said to be random. For example, it is possible to predict exactly the position of a rocket if we know exactly the initial conditions. However, it is impossible to predict the result of tossing a coin. The first experiment is deterministic while the second one is random.

Let us consider a random experiment. The set of all possible outcomes is called the "sample space".

Ex: The experiment of tossing a coin produces a sample space with two outcomes: H = "head" and T = "tail". Using set notation, we can write:

$S = \{H, T\}$.

The experiment of rolling a die produces a sample space with six possible outcomes: $S = \{f_1, f_2, f_3, f_4, f_5, f_6\}$ where f_i indicates that the face with i dots will show one top. When we roll a die, the outcome of the experiment will be one element from the set S.

The previous two examples are described by a finite sample space. However, the sample space can be infinite. For example, let us consider the experiment consisting on tossing a coin several times until a head shows up. If we are really unlucky, the head may never show up. So, in this case, the sample space will consist of all possible natural integers. Another infinite space can be used to describe the set of all possible initial phases from a sinewave generator. It is evident that the initial phase can be any real number in the interval $[0, 2\pi]$.

When we carry out a random experiment, we may be interested in "event" more complex than the simple appearance of an outcome. For example, in the die rolling experiment, we may be interested in the event: "the die shows an even number on its top face". Clearly, this event can be described by a subset of the sample space S. The subset is: "even" = $\{f_2, f_4, f_6\}$. However, not all subsets of S qualify to be events. Before defining formally the notion of event, we have to describe the algebra of subsets.

Inclusion (implication). A set $\mathcal{A}$ is included in a set $\mathcal{B}$ if the following property applies:

$$\forall \zeta \in \mathcal{A} \Rightarrow \zeta \in \mathcal{B} \tag{1.1}$$

We can also say that $\mathcal{A}$ implies $\mathcal{B}$. The inclusion is a relation of order between subsets of a set *S*. However, this relation is partial. Not all subsets can be related by this relation. We can use a graphical technique to display the different relations and operations that can exist between subsets of a sample space *S*. This graphical technique uses "Venn diagrams". The set *S* is represented by a rectangle and its different subsets are represented by closed figures inside the rectangle.

Ex:

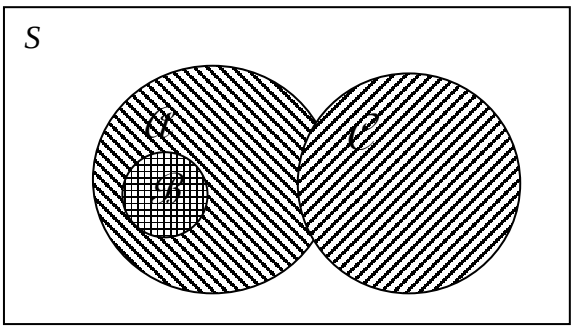


Fig 1- 1 Venn Diagram

The above Venn diagram shows a sample space *S* and three of its subsets. We clearly see that $\mathcal{B} \subset \mathcal{A} \subset S$. However, it is also apparent that $\mathcal{C}$ is not related to the other two subsets $\mathcal{A}$ and $\mathcal{B}$.

We can also define operations on subsets of a given sample space. A unary operation is the **"complementation".**

The complement of a subspace $\mathcal{A}$ is the subset of *S* that contains all the elements of *S* that are not in $\mathcal{A}$. More formally, it is defined as:

$$\overline{\mathcal{A}} = \left\{ \zeta \in S \mid \zeta \notin \mathcal{A} \right\} \tag{1.2}$$

We read "not $\mathcal{A}$" for $\overline{\mathcal{A}}$

We define also two binary operations.

The **"intersection"** of two subsets is defined as the set that contains the elements belonging to both subsets.

$$\mathcal{A} \cap \mathcal{B} = \left\{ \zeta \in S \mid \zeta \in \mathcal{A} \text{ and } \zeta \in \mathcal{B} \right\} \tag{1.3}$$

We should read the above as $\mathcal{A}$ and $\mathcal{B}$.

The **"union"** of two subsets is defined as the set that contains elements belonging to either the first set or the second set or to both.

$$\mathcal{A} \cup \mathcal{B} = \{\zeta \in S \mid \zeta \in \mathcal{A} \text{ or } \zeta \in \mathcal{B} \text{ or } \zeta \in \text{both}\} \tag{1.4}$$

We should read the above as $\mathcal{A}$ or $\mathcal{B}$.

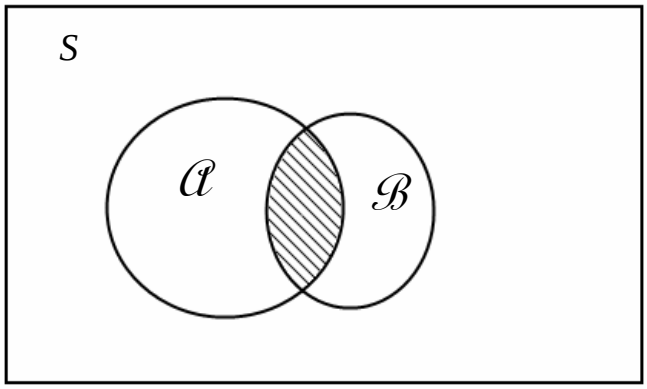


Fig 1- 2 Intersection

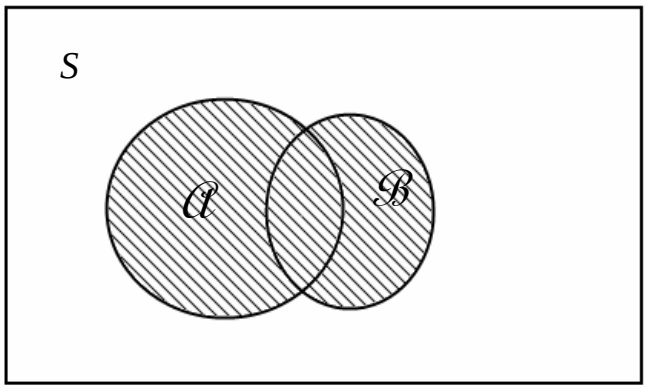


Fig 1- 3 Union

The above two figures illustrate the union and intersection of subsets. The three operations are related by the following laws: The De Morgan laws:

$$\overline{\mathcal{A} \cup \mathcal{B}} = \overline{\mathcal{A}} \cap \overline{\mathcal{B}} \text{ and } \overline{\mathcal{A} \cap \mathcal{B}} = \overline{\mathcal{A}} \cup \overline{\mathcal{B}}$$

Among the different subsets of *S*, there are two which are quite important: The sample space *S* itself, called the "certain event" and the empty set ∅, also called the "impossible event".

When we consider infinite sets, we can encounter sequences of subsets. A sequence of subsets of *S* is an infinite collection of subsets indexed by an integer. For example, let $S = [0, 1]$. We consider the following sequence $\mathcal{A}_n = \left\{\zeta \in S \mid 0 \le \zeta < \frac{1}{n}; n = 1, 2, \ldots\right\}$. This sequence has a limit (find it).

When we deal with these infinite sequences, we define infinite unions and infinite intersections: $\bigcup_{n=1}^{\infty}\mathcal{A}_n$ and $\bigcap_{n=1}^{\infty}\mathcal{A}_n$.

With the above rules of combinations between subsets, we can now define the notion of Boolean algebra of field.

Boolean algebra or Field

Consider a sample space *S* and a set $\mathcal{F}$ of subsets of *S*. $\mathcal{F}$ is called a Field or Boolean algebra if the following rules are satisfied:

1. If $\mathcal{A}\in\mathcal{F}$ then $\overline{\mathcal{A}}\in\mathcal{F}$, i.e. $\mathcal{F}$ is closed under complementation.
2. If $\mathcal{A}\in\mathcal{F}$ and $\mathcal{B}\in\mathcal{F}$ then $\mathcal{A}\cup\mathcal{B}\in\mathcal{F}$, i.e. $\mathcal{F}$ is closed under the operation of union.

If we consider infinite sets, then we can add a third axiom and define σ-fields or σ-Boolean algebra.

3. If the sequence $\mathcal{A}_n\in\mathcal{F}$; $n=1,2,\ldots$ then $\bigcup_{n=1}^{\infty}\mathcal{A}_n\in\mathcal{F}$, i.e. $\mathcal{F}$ is closed under infinite unions.

It is not hard to show that the above properties imply also closure for the intersection (Use the De Morgan's laws to show it).

To qualify as "event" a subset must be an element of a field. It is only on elements of fields that we can define a measure. We will see that probability is nothing but a specific measure.

Example:

Consider the sample space *S* ={a, b, c}. We can define several fields on this set. We can first define to set of all subsets. This set contains 2^3 elements. Another field is the one containing only *S* and ∅. Let us consider the field $\mathcal{F}$= {∅, {a, b}, {c}, *S*}. This set contains only 4 events. If they are the only significant ones, this field is enough to describe our experiment. In finite spaces, this notion of field and the distinction between events and subsets is not important and we can use as a field the set of all possible subsets. It is only in infinite spaces that we have to reduce our view to only significant sets.

Terminology: we say that the event $\mathcal{A}$ has been realized if as a result of the random experiment one of its elements is the result of the experiment.

Using the above definition of fields, we can now define set functions. They are mapping from sets to numbers. So, a set function is defined as:

Given a field $\mathcal{F}$, the set function *f* is defined by $f:\mathcal{A}\in\mathcal{F}\mapsto f\left(\mathcal{A}\right)\in\mathbb{R}$.

1.2 Probability definition

The probability is a set function. It is a measure and as such it shares the same properties as typical measures for example length of intervals or areas in two dimensional spaces.

In order to define a probability measure, we must perform a random experiment that will generate a sample space *S*. Furthermore, being a set function, the domain of the function must be a field (or a σ-field) of events (subsets of *S*). So, given a random experiment, a sample space *S* and a field of events $\mathcal{F}$, the probability of events in $\mathcal{F}$ is defined as a set function *P* satisfying the following three axioms:

1. $\forall \mathcal{A} \in \mathcal{F} \quad ; \quad P(\mathcal{A}) \geq 0$
2. $P(S) = 1$
3. $\forall \mathcal{A}, \mathcal{B} \in \mathcal{F}$ such that $\mathcal{A} \cap \mathcal{B} = \varnothing \; ; \; P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$

If the set *S* is infinite and we can define a σ-field, we have to add a fourth axiom:

4. $\forall \mathcal{A}_n \; ; n = 1, 2, \ldots$ such that $\mathcal{A}_i \cap \mathcal{A}_j = \varnothing \quad i \neq j \; ; \quad P\left(\bigcup_{n=1}^{\infty} \mathcal{A}_n\right) = \sum_{n=1}^{\infty} P(\mathcal{A}_n)$

When two events are such that $\mathcal{A} \cap \mathcal{B} = \varnothing$, we say that they are mutually exclusive. This means that the occurrence of both of them at the same time is impossible.

A probability space is then defined by the triplet: $(S, \mathcal{F}, P)$.

Using the above axioms, we can deduce several properties.

$P(\overline{\mathcal{A}}) = 1 - P(\mathcal{A})$

$P(\varnothing) = 0$

$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B})$

An event containing only one element is called an elementary event. In finite sets, it common to assume that all the elementary events are equiprobable (if there is no reason to favor any one).

1.3 Conditional Probability and Independence

In many occasions, we have some previous knowledge about the result of the random experiment, i.e. we know that the outcome will belong to some event $\mathcal{B}$. This restricts the set of all possibilities from the set *S* to the subset $\mathcal{B}$. So, if

some event $\mathcal{A}$ will occur, the outcome of the experiment will belong to both $\mathcal{A}$ and $\mathcal{B}$.

Example: Consider the experiment of rolling a die $S=\{f_1, f_2, f_3, f_4, f_5, f_6\}$. If we know that the result of the experiment is even, this defines the event $\mathcal{B}=\{f_2, f_4, f_6\}$. Consider the event $\mathcal{A}$ that the outcome is a number smaller than 5, i.e. $\mathcal{A}=\{f_1, f_2, f_3, f_4\}$, then the result of the random experiment will be either f_2 or f_4. This is nothing but the set $\mathcal{A}\cap\mathcal{B}$.

Definition

Given a probability space defined by the triplet $(S,\mathcal{F},P)$ and an event $\mathcal{B}\in\mathcal{F}$ such that $P(\mathcal{B})>0$, we define the conditional probability of an event $\mathcal{A}$ given the event $\mathcal{B}$ by:

$$P(\mathcal{A}|\mathcal{B})=\frac{P(\mathcal{A}\cap\mathcal{B})}{P(\mathcal{B})} \tag{1.5}$$

If we consider the previous example, $S=\{f_1, f_2, f_3, f_4, f_5, f_6\}$, $\mathcal{B}=\{f_2, f_4, f_6\}$ and $\mathcal{A}=\{f_1, f_2, f_3, f_4\}$. Then $\mathcal{A}\cap\mathcal{B}=\{f_2, f_4\}$. We can compute the different probabilities: $P(\mathcal{B})=\frac{3}{6}=\frac{1}{2}$, $P(\mathcal{A}\cap\mathcal{B})=\frac{2}{6}=\frac{1}{3}$. Finally,

$$P(\mathcal{A}|\mathcal{B})=\frac{P(\mathcal{A}\cap\mathcal{B})}{P(\mathcal{B})}=\frac{\frac{1}{3}}{\frac{1}{2}}=\frac{2}{3}.$$

A concept that is very important in probability theory is the concept of "independence".

Definition

The two events $\mathcal{A}$ and $\mathcal{B}$ are said to be independent if $P(\mathcal{A}\cap\mathcal{B})=P(\mathcal{A})P(\mathcal{B})$.

When two events are independent, the occurrence of one does not have any influence on the occurrence of the other. This can be seen from the definition of conditional probability: $P(\mathcal{A}|\mathcal{B})=\frac{P(\mathcal{A}\cap\mathcal{B})}{P(\mathcal{B})}=\frac{P(\mathcal{A})P(\mathcal{B})}{P(\mathcal{B})}=P(\mathcal{A})$. So, we see that the knowledge of $\mathcal{B}$ does not modify the probability of $\mathcal{A}$.

1.4 Theorem of total probability and Bayes rule

The theorem of total probability is a result commonly used in communication theory. We first have to define the notion of a partition.

$\mathcal{A}_i$, $i = 1, \ldots, n$ is a partition of S if:

1. It forms a cover of S, i.e. $\bigcup_{i=1}^{n} \mathcal{A}_i = S$
2. The subsets are two by two mutually exclusive: $A_i \cap A_j = \varnothing \ ; \ i \neq j$

Theorem of total probability

Given an event $\mathcal{B}$ and a partition $\mathcal{A}_i$, $i = 1, \ldots, n$ of S

$$P(\mathcal{B}) = P(\mathcal{B}|\mathcal{A}_1)P(\mathcal{A}_1) + P(\mathcal{B}|\mathcal{A}_2)P(\mathcal{A}_2) + \cdots + P(\mathcal{B}|\mathcal{A}_n)P(\mathcal{A}_n) \quad (1.6)$$

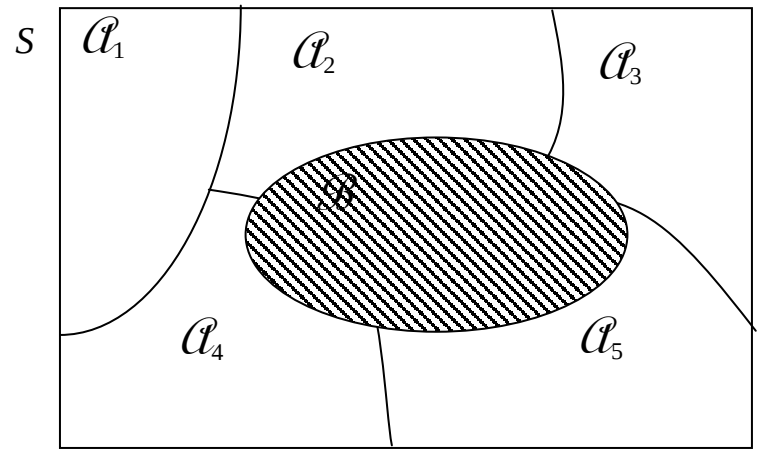


Fig 1- 4 Theorem of total probability

Fig 1- 4 illustrates the theorem of total probability. One application of the theorem is the calculation of the probability of error in communication. Suppose we sent two different messages m_1 and m_2 with probability $P(m_1)$ and $P(m_2)$. When we transmit the message m_1, we have a probability of error $P(error|m_1)$ and when we transmit m_2, the probability of error becomes $P(error|m_2)$. The average probability of error is then:

$$P(error) = P(error \mid m_1)P(m_1) + P(error \mid m_2)P(m_2)$$

Bayes rule is also frequently used in communication. It is based on the definition of the conditional probability.

Bayes rule

Given a probability space $(S, \mathcal{F}, P)$ and two events $\mathcal{A}$ and $\mathcal{B}$ such that $P(\mathcal{A}) > 0$ and $P(\mathcal{B}) > 0$, we can write:

$$P(\mathcal{A}|\mathcal{B})P(\mathcal{B}) = P(\mathcal{B}|\mathcal{A})P(\mathcal{A}) \quad (1.7)$$

Bayes rule is often given between an event $\mathcal{B}$ and a partition $\mathcal{A}_i$, i=1, …, n of S. In this case, it is written as:

$$P\left(\mathcal{A}_i \middle| \mathcal{B}\right) = \frac{P\left(\mathcal{B} \middle| \mathcal{A}_i\right) P\left(\mathcal{A}_i\right)}{\sum_{k=1}^{n} P\left(\mathcal{B} \middle| \mathcal{A}_k\right) P\left(\mathcal{A}_k\right)} \tag{1.8}$$

Example: The binary symmetrical channel (BSC)

Assume a binary information source that can generate one out of two symbols "0" and "1". At the output of the channel of communication, we observe a received symbol. So, in this example, we can describe two events at the transmitter: "transmit a zero", represented by $\{t = 0\}$ with probability α and "transmit a one", represented by $\{t = 1\}$ with probability $\beta = 1 - \alpha$. At the receiver, we observe either a one or a zero. So, here also, we have two events: "receive a zero", represented by $\{r = 0\}$ and "receive a one", represented by $\{r = 1\}$. In this channel, there two possibilities for errors: we transmit a zero and we receive a one and vice versa. We can measure two conditional probabilities: $P\left(r = 1 \middle| t = 0\right)$ and $P\left(r = 0 \middle| t = 1\right)$. If these two probabilities are equal, the channel is said to be symmetrical. At that time, we can write $P\left(r = 1 \middle| t = 0\right) = P\left(r = 0 \middle| t = 1\right) = p$ and $P\left(r = 0 \middle| t = 0\right) = P\left(r = 1 \middle| t = 1\right) = q = 1 - p$. The probability p is the channel probability of error. In all communication systems, the observer is placed at the output of the channel and he wants to guess what symbol has been transmitted. Assume we observe a one at the output, using Bayes rule, we can compute the probability that a one or a zero has been transmitted.

$$P\left(t = 1 \middle| r = 1\right) = \frac{P\left(r = 1 \middle| t = 1\right) P\left(t = 1\right)}{P\left(r = 1\right)} = \frac{P\left(r = 1 \middle| t = 1\right) P\left(t = 1\right)}{P\left(r = 1 \middle| t = 0\right) P\left(t = 0\right) + P\left(r = 1 \middle| t = 1\right) P\left(t = 1\right)}$$

$$= \frac{(1 - p)(1 - \alpha)}{\alpha p + (1 - p)(1 - \alpha)}$$

$$P\left(t = 0 \middle| r = 1\right) = \frac{P\left(r = 1 \middle| t = 0\right) P\left(t = 0\right)}{P\left(r = 1\right)} = \frac{P\left(r = 1 \middle| t = 0\right) P\left(t = 0\right)}{P\left(r = 1 \middle| t = 0\right) P\left(t = 0\right) + P\left(r = 1 \middle| t = 1\right) P\left(t = 1\right)}$$

$$= \frac{\alpha p}{\alpha p + (1 - p)(1 - \alpha)}$$

If the two symbols are equiprobable ($\alpha = \beta = ½$), the two probabilities simplify to: $P\left(t = 1 \middle| r = 1\right) = 1 - p$ and $P\left(t = 0 \middle| r = 1\right) = p$. So, if the probability of error p is less than ½ and if we receive a one, it is more probable that a one has

been transmitted. If p is larger than ½, we are surely using the channel in reverse and we should invert our decisions.

1.5 Random Variables

In many occasions, the result of the random experiment is a number. It can be a voltage level, a light intensity, a phase angle, etc. For this purpose, we should extend the previously defined theory from events to subsets of $\mathbb{R}$. So, given a probability space $\left(S,\mathcal{F},P\right)$, we define a mapping from the sample space S to the set of real numbers $\mathbb{R}$ such that the set of outcomes that map to an interval of real numbers is an event in $\mathcal{F}$. We add the restriction that the probability of an event that has elements that map to infinity is zero. This mapping defines a random variable. So, using a more formal language, a random variable X is defined as follows.

Definition

Given a probability space $\left(S,\mathcal{F},P\right)$, a random variable X is a mapping from the sample space S to the set of real numbers such that:

$X:\zeta\in S\mapsto X(\zeta)\in\mathbb{R}$ such that $\{\zeta\in S\mid X(\zeta)\le x\in\mathbb{R}\}\in\mathcal{F}$ (it is an event) and $P\left(\{\zeta\in S\mid X(\zeta)=+\infty\}\right)=P\left(\{\zeta\in S\mid X(\zeta)=-\infty\}\right)=0$

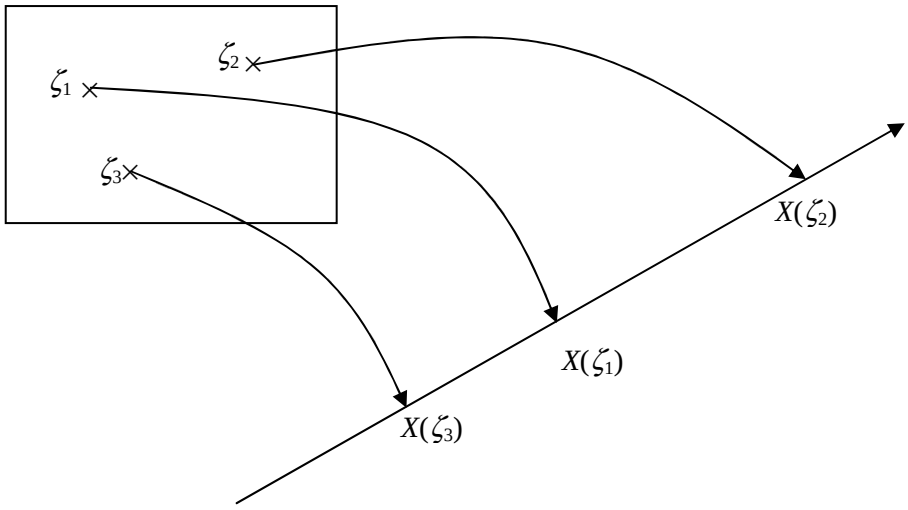


Fig 1- 5 Random Variable

Example: Consider the die rolling experiment and the mapping $X:f_k\mapsto X(f_k)=10k$. The sentence "X produces a number between 5 and 35" corresponds to the event in S equal to $\{f_1, f_2, f_3\} = \{\zeta\in S\mid 5 < X(\zeta) < 35\}$. This

event can be denoted directly as $\{5 < X < 35\}$ or $X \in]5, 35[$. We can now compute the probability $P\left(\{5 < X < 35\}\right) = P\left(\{f_1, f_2, f_3\}\right) = \frac{1}{2}$ and it is the probability that the function $X(\,.\,)$ produces a number inside the open interval $]5, 35[$.

We see that in order to compute probabilities, we have to go back to the original probability space $\left(S, \mathcal{F}, P\right)$. Fortunately, the following two functions allow us to do these computations directly in $\mathbb{R}$.

Cumulative Distribution Function (cdf)

The cumulative distribution function (cdf) of a random variable X is the following function:

$$F_X(x) = P\left(\{X \le x\}\right) \tag{1.9}$$

Using the cdf, we can compute the probability that the random variable produces a number inside an interval.

$$P\left(\{a < X \le b\}\right) = F_X(b) - F_X(a) \tag{1.10}$$

Properties of the cdf:

1. $F_X(-\infty) = 0$
2. $F_X(+\infty) = 1$
3. The function F_X is a monotonically increasing function. In other words: if $x_1 < x_2$ then $F_X(x_1) \le F_X(x_2)$
4. The function F_X is continuous from the right. $\lim\limits_{\substack{x \to x_0 \\ x > x_0}} F_X(x) = F_X(x_0)$

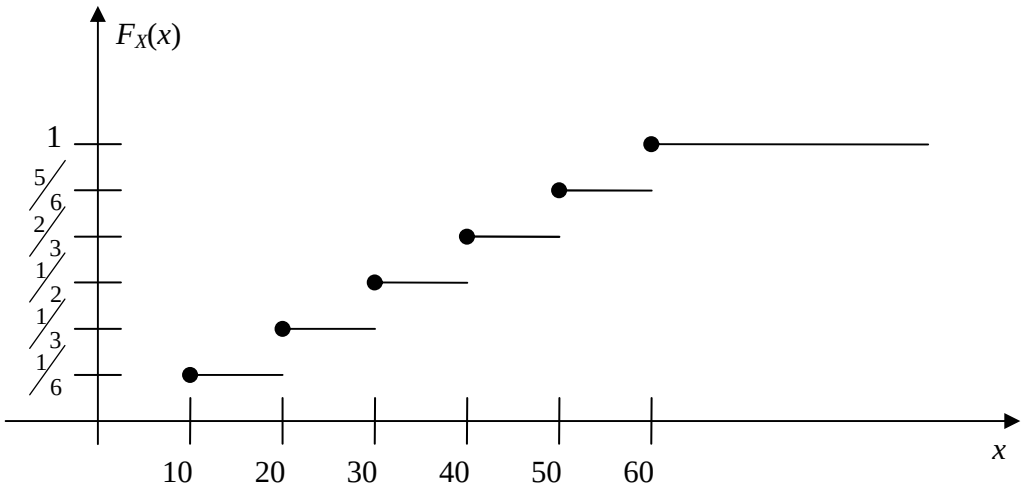


Fig 1- 6 Cumulative distribution function

Fig 1- 6 is the graph of the cdf of the random variable defined in the previous example. We can observe that it is a staircase function. In this example, the random variable will be realized on a finite set of values. We can remark that

the jump at each discontinuity is equal to the probability that the random variable takes the value at that point.

Some random variables can take values on a discrete set (finite or infinite) of numbers. They are simply called discrete random variables. They are best described by "probability masses". The set of possible values is: $\{x_1, x_2, \ldots\}$ and the probability masses are $P(\{X = x_1\})$, $P(\{X = x_2\})$, …

For these discrete random variables, the cdf is always a staircase function and can be expressed as a weighted sum of unit step functions:

$$F_X(x) = \sum_{k=1} P(\{X = x_k\}) u(x - x_k)$$

The summation is left without upper bound in order to take into account the infinite case.

If the cdf is a continuous function, we can show that $P(\{X = x_0\}) = 0$ at any point x_0. At that time, we say that we have a continuous random variable. There exist also mixed random variables that have a cdf formed by the sum of a continuous function and a staircase one.

When the random variable is of continuous type, it is better described by the probability density function.

Probability density function (pdf)

The probability density function (pdf) is defined as the derivative of the cdf of a given random variable:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{1.11}$$

The inverse relation is of course:

$$F_X(x) = \int_{-\infty}^{x} f_X(u) du$$

We can derive the different properties of the pdf from those of the cdf.

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f_X(x) dx = 1$
3. $P(\{a < X \leq b\}) = \int_a^b f_X(x) dx$; in this formula, the upper bound is included in the interval of integration while the lower one is excluded.
4. If the random variable is discrete, the pdf is given by: $f_X(x) = \sum_{k=1} P(\{X = x_k\}) \delta(x - x_k)$, where δ is the Dirac impulse

function. This means that the random variable is concentrated at the discrete set of values x_k, $k = 1, 2, \ldots$

1.6 Multivariate distributions

In many communication problems, we have to consider random vectors. The definition is essentially the same as the one-dimensional case. Given a probability space $\left(S, \mathcal{F}, P\right)$, a random vector is defined as a mapping from the sample space S to the n dimensional space $\mathbb{R}^n$.

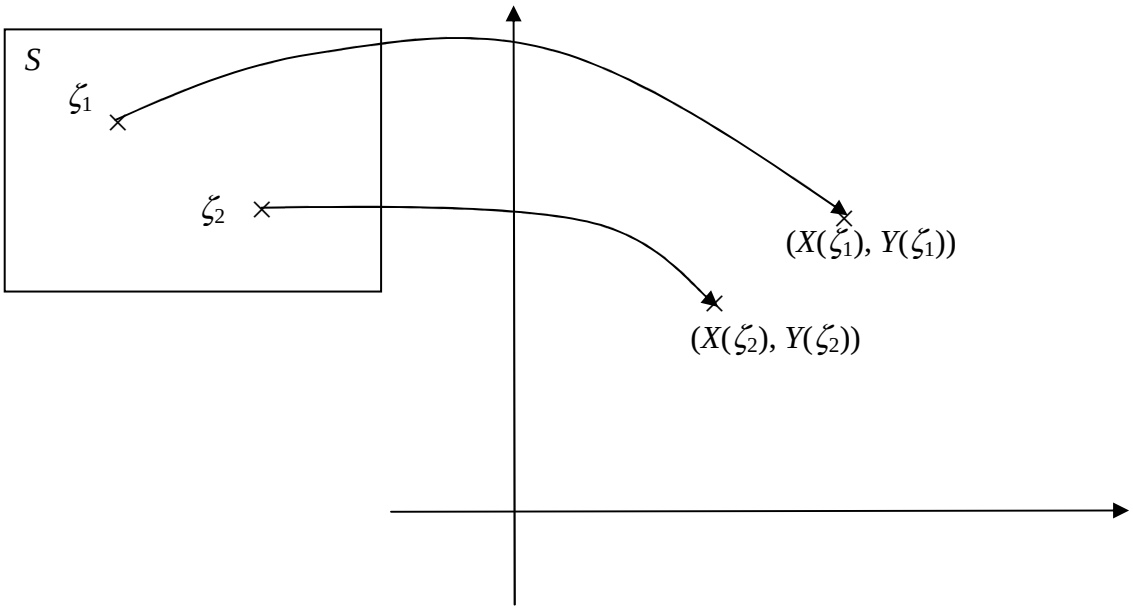


Fig 1- 7 Two dimensional random variable

For random vectors, we define joint cdf's and pdf's. We provide the basic definitions for the two dimensional case. However, these definitions are valid for any number of variables.

Joint cumulative distribution function

Given two random variables X and Y, the joint cdf is defined as:

$$F_{XY}(x, y) = P\left(\{X \le x\} \text{and} \{Y \le y\}\right) \tag{1.12}$$

It is the probability that the pair X(ζ), Y(ζ) falls in the region displayed in Fig 1- 8.

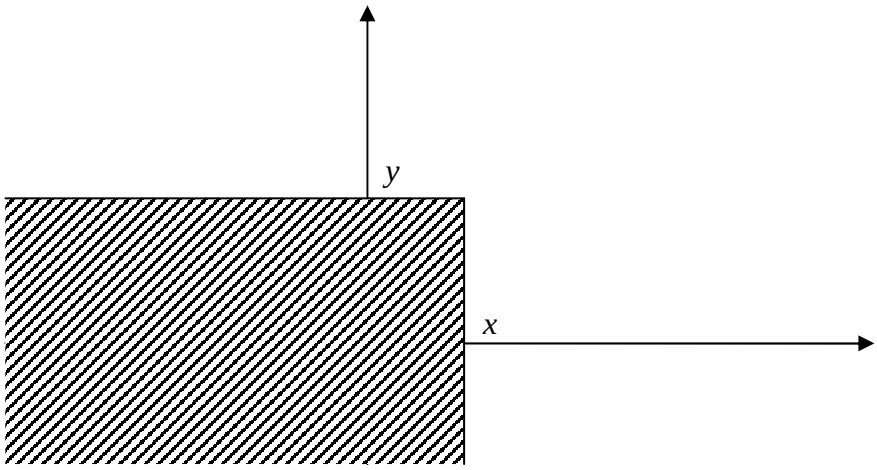


Fig 1- 8 CDF region

Joint probability density function

As in the one dimensional case, the pdf is defined as the derivative of the cdf:

$$f_{XY}(x,y)=\frac{\partial F_{XY}(x,y)}{\partial x\partial y} \tag{1.13}$$

The reverse relation is:

$$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)dudv \tag{1.14}$$

An importance concept in multivariate distributions is the one of the marginal distributions. For the two dimensional case, there exist two marginal distributions:

The marginal distribution of the random variable X is the distribution of the random variable X irrespective on the value of Y.

$$F_X(x)=P\left(\{X\le x\}and\{-\infty<Y<+\infty\}\right)=F_{XY}(x,+\infty) \tag{1.15}$$

We have the same definition for the other variable:

$$F_Y(y)=P\left(\{Y\le y\}and\{-\infty<X<+\infty\}\right)=F_{XY}(+\infty,y) \tag{1.16}$$

If we compute the derivatives:

$f_X(x)=\frac{dF_X(x)}{dx}=\frac{d}{dx}F_{XY}(x,+\infty)=\frac{d}{dx}\int_{-\infty}^{x}\int_{-\infty}^{+\infty} f_{XY}(u,v)dudv$, we obtain the following relations for the marginal densities:

$$f_X(x)=\int_{-\infty}^{+\infty} f_{XY}(x,y)dy \tag{1.17}$$

and

$$f_Y(y)=\int_{-\infty}^{+\infty} f_{XY}(x,y)dx \tag{1.18}$$

In general, the multivariate pdf is more useful than the multivariate distribution. With the pdf, we can compute probabilities for events that are more general than the ones described by cdf's.

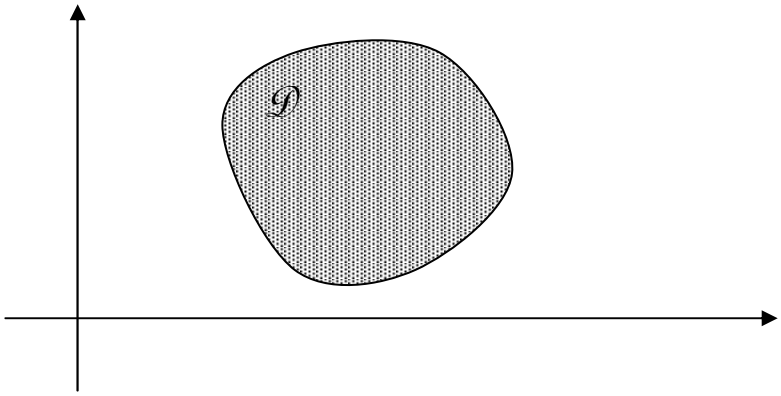


Fig 1- 9 Domaine of integration

For example, the probability that the pair *X*, *Y* produces a point inside the domain $\mathcal{D}$shown in Fig 1- 9 is given by the following integral:

$$P\left(\left\{(X,Y)\in\mathcal{D}\right\}\right)=\iint_{\mathcal{D}} f_{XY}(x,y)dxdy \tag{1.19}$$

When we deal with more than one random variable, we must take into account the dependences that exist between them.

Complex random variable

A complex random variable is nothing but a pair of real random variable and as such it should be described by the joint pdf of the real and imaginary part.

Independent random variables

Two random variables are independent if their joint pdf is equal to the product of the marginal pdf's.

$$f_{XY}(x,y)=f_X(x)\,f_Y(y) \tag{1.20}$$

Conditional pdf's

If the two random variables are not independent, then the realization of one (the event $\{X=x\}$) influences the realization of the other. We can define conditional pdf's on many events.

Conditioning on the realization of another random variable

$$f_{X|Y}\left(x\middle|\{Y=y\}\right)=\frac{f_{XY}(x,y)}{f_Y(y)} \tag{1.21}$$

The pdf $f_{X|Y}\left(x\middle|\{Y=y\}\right)$ is often written as $f_{X|Y}\left(x\middle|y\right)$.

Conditioning on the realization of a more general event

Consider a general event $\mathcal{A}$, we can define the following conditional cdf:

$$F_{X|\mathcal{A}}\left(x|\mathcal{A}\right)=\frac{P\left(\{X\le x\}\,and\,\mathcal{A}\right)}{P\left(\mathcal{A}\right)} \tag{1.22}$$

The pdf is of course the derivative:

$$f_{X|\mathcal{A}}\left(x|\mathcal{A}\right)=\frac{dF_{X|\mathcal{A}}(x|\mathcal{A})}{dx}$$

Using these definitions, we can define several Bayes rules:

Between probabilities:

Given two events $\mathcal{A}$ and $\mathcal{B}$:

$$P\left(\mathcal{A}\middle|\mathcal{B}\right)P\left(\mathcal{B}\right)=P\left(\mathcal{B}\middle|\mathcal{A}\right)P\left(\mathcal{A}\right) \tag{1.23}$$

Between pdf's:

$$f_{X|Y}(x|Y=y)\,f_Y(y)=f_{Y|X}(y|X=x)\,f_X(x) \tag{1.24}$$

Mixed Bayes rule:

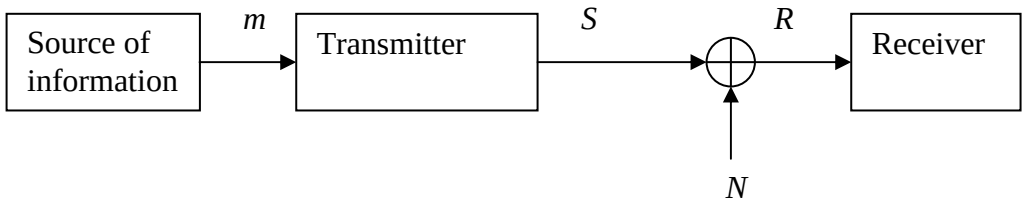
Given an event $\mathcal{A}$ and a random variable X

$$P\left(\mathcal{A}\middle|\{X=x\}\right)f_X(x)=f_{X|\mathcal{A}}\left(x|\mathcal{A}\right)P\left(\mathcal{A}\right) \tag{1.25}$$

In communication, it is the last rule that is mostly used. The following example is an important illustration of the use of the above rule.

Communication example:

Consider the following system:



The source of information produces one message out of an alphabet of two messages $\{m_0, m_1\}$ with an "a priori" probability $P(m_k)$, $k = 1,2$. The transmitter "maps" this message to a voltage level. To m_0 corresponds s_0 and to m_1 corresponds s_1. In what follows, we will assume that $s_1 > s_0$. During the transmission, the signal is added to a "noise" which is modeled as a zero mean random variable completely independent on the signal level S. The receiver observes the sum R and should decide which of the two messages has been transmitted.

Essentially, the receiver is a decision device. If there is no observation, the only information on the messages is the a priori probability. So, if we cannot observe the random variable *R*, we should decide on the most probable message. However, the observation of the variable *R* will provide us with some information on the transmitted message and is going to modify our "a priori" knowledge to an "a posteriori" one. So, we should compute the "a posteriori" probabilities $P\left(m_k \middle| \{R = r\}\right)$, $k = 0,1$. The a posteriori probability is the conditional probability of the message m_k given that we have observed the variable *R* and it has taken the value *r*. We have also knowledge about the added noise and it has the following pdf:

$$f_N(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n^2}{2\sigma^2}}$$

The decision rule becomes now: decide on the message that has the highest a posteriori probability. We say that we are applying the "maximum of a posteriori probability" (MAP) rule. Based on the data in our possession, we can compute the conditional pdf's $f_{R|m_k}(r|m_k)$, $k = 0,1$.

If we are given that m_0 is transmitted, the received random variable becomes $R = s_0 + N$. So, $N = R - s_0$. So, we can conclude that $f_{R|m_0}(r|m_0) = f_{N|m_0}(r - s_0 \,|\, m_0) = f_N(r - s_0)$. The last equality results from the fact that *N* does not depend on the transmitted message. For the other message, we obtain the same result: $f_{R|m_1}(r|m_1) = f_N(r - s_1)$.

To be able to decide, the receiver must compute the a posteriori probabilities. The decision rule is:

Decide m_0 if $P\left(m_0 \middle| \{R = r\}\right) > P\left(m_1 \middle| \{R = r\}\right)$; decide m_1 otherwise.

Using the mixed Bayes rule, the inequality is transformed to:

$$\frac{f_{R|m_0}(r \,|\, m_0)P(m_0)}{f_R(r)} > \frac{f_{R|m_1}(r \,|\, m_1)P(m_1)}{f_R(r)}$$

The denominator is common to both sides. It can be eliminated from the decision rule.

$$f_{R|m_0}(r \,|\, m_0)P(m_0) > f_{R|m_1}(r \,|\, m_1)P(m_1)$$

The conditional probabilities $f_{R|m_k}(r|m_k)$ are called "likelihood" functions. If the two messages are equiprobable, the decision rule simplifies to:

$$f_{R|m_0}(r \,|\, m_0) > f_{R|m_1}(r \,|\, m_1)$$

and the decision rule is now called "maximum likelihood" (ML). We can use the ML rule even if the a priori probabilities are not equal. However, we will see that such decision is not optimum. Fig 1- 10 shows that the decision rule divides the observation space (the variable r) into two regions: I_0 and I_1.

$I_0 =]-\infty, a]$ and $I_1 = [a, +\infty[$. The bound a is a threshold that divides the observation space into two regions. It corresponds the value of r for which the a posteriori probabilities are equal.

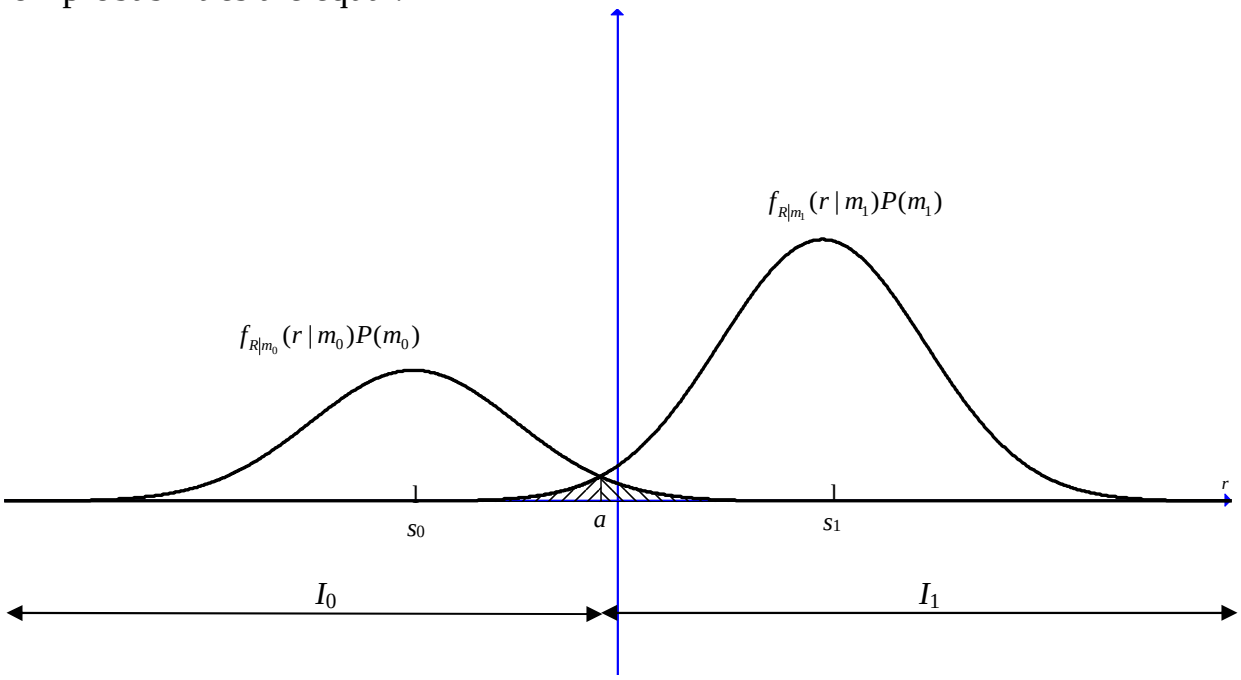


Fig 1- 10 decision regions

The value of the threshold for the given noise density is:

$$a = \frac{s_0 + s_1}{2} - \frac{\sigma^2}{s_1 - s_0} \ln \frac{P(m_1)}{P(m_0)}$$

The decision rule can be rephrased now as:

Decide m_0 if $r < a$, decide m_1 otherwise. In terms of the intervals I_0 and I_1, it is: decide m_0 if $r \in I_0$ and decide m_1 if $r \in I_1$. The receiver can be implemented using a comparator.

If we want to evaluate the quality of the receiver, we must evaluate the probability of error. There exist two possibilities of error: "decide m_1 when m_0 is transmitted" and "decide m_0 when m_1 is transmitted". So, we can compute two conditional probabilities of error: we call E the event "error".

$$P(E \mid m_0) = P(\{r \in I_1\} \mid m_0) = \int_a^{+\infty} f_{R|m_0}(r \mid m_0) dr$$

$$P(E \mid m_1) = P(\{r \in I_0\} \mid m_1) = \int_{-\infty}^{a} f_{R|m_1}(r \mid m_1) dr$$

and the total probability of error is computed using the theorem of total probability:

$$P(E) = P(m_0)P\left(E \mid m_0\right) + P(m_1)P(E \mid m_1)$$

$$P(E) = \int_a^{+\infty} P(m_0)\, f_{R|m_0}(r \mid m_0)dr + \int_{-\infty}^{a} P(m_1)\, f_{R|m_1}(r \mid m_1)dr$$

This probability corresponds to the hatched area in Fig 1- 10 can now show that the MAP rule corresponds to the minimum probability of error. Let compute the derivative with respect to *a* of the probability of error and set it to zero:

$$\frac{dP(E)}{da} = -P(m_0)\, f_{R|m_0}(a \mid m_0) + P(m_1)\, f_{R|m_1}(a \mid m_1) = 0$$

We see that the optimum threshold satisfies the MAP rule. So, the MAP receiver is the best receiver from a probability of error. If the threshold is chosen using another rule (the ML for example), the probability of error will be greater. Yet, there is a case where the ML decision rule is equivalent to the MAP rule. It corresponds to the equiprobable message case.

1.7 Averages

Averages are numbers that can provide a partial characterization of one or may random variables.

Mean

The mean or expected value of a random variable is defined as:

$$E\left[X\right] = m_X = \overline{X} = \int_{-\infty}^{+\infty} x f_X(x)dx \qquad (1.26)$$

If the random variable is discrete, the integration is replaced by a summation:

$$E\left[X\right] = \sum_{k=1} x_k P\left(\left\{X = x_k\right\}\right) \qquad (1.27)$$

If the random variable is function of another random variable, i.e. $Y = g(X)$, the expectation is given by:

$$E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(x)\, f_X(x)dx \qquad (1.28)$$

One function that is commonly used is $Y = X^n$. This function defines the moments of a random variable.

Moments

The n[th] order moment of a random variable is:

$$E\left[X^n\right] = \int_{-\infty}^{+\infty} x^n f_X(x)dx \qquad (1.29)$$

It is not hard to show that $E\left[X^0\right] = 1$ and that $E\left[X^1\right] = m_X$.

Another class of moments that is also very useful is the class of central moments:

Central moments

The central moment of order n is

$$E\left[\left(X-m_X\right)^n\right]=\int_{-\infty}^{+\infty}(x-m_X)^n f_X(x)dx \tag{1.30}$$

The central moment of order 1 is zero. The most important central moment is the second order one.

Variance

The central moment of order two is called the variance. It gives us an indication about the spread of the random variable.

$$v_X=\sigma_X^2=E\left[\left(X-m_X\right)^2\right] \tag{1.31}$$

Its square root σ_X is called the standard deviation. (We will see that for some random signals, it is the ac rms value).

By developing the square and taking the average, we obtain:

$$E\left[X^2\right]=\sigma_X^2+m_X^2 \tag{1.32}$$

For "ergodic[1]" signals, this relation states that the total power is the sum of the dc power plus the ac power.

The fact that the variance is an indication of the spread of a random variable around the mean is shown by the Chebyshev inequality:

Chebyshev inequality

Given a random variable X with a mean m_X and a finite variance σ_X^2, we have:

$$P\left(\left|X-m_X\right|\le k\sigma_X\right)\ge 1-\frac{1}{k^2};\ k>0 \tag{1.33}$$

For example, the probability to have a realization of the random variable inside an interval of width $6\sigma_X$ ($k = 3$) centered on the mean is about 0.89. The Chebyshev inequality provides us with a bound on the probability without the need of the actual pdf.

Multivariate case

For the multivariate case, we have to use vector notation. A random vector is denoted $\mathbf{X}=\left(X_1,X_2,\cdots,X_n\right)^T$. The superscript T stands for transpose. A vector is a one column matrix and is written as a bold Latin letter.

[1] This notion will be defined in the next chapter.

The joint pdf of X is denoted $f_{\mathbf{X}}(\mathbf{x}) = f_X(x_1, x_2, \ldots, x_n)$. So, the mean of a random vector is:

$$\mathbf{m}_{\mathbf{X}} = E[\mathbf{X}] = \int \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \iint \cdots \int \mathbf{x} f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n \quad (1.34)$$

All the integrals are taken over the whole space. Each one of its components is given by:

$$m_{\mathbf{X}k} = \int x_k f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \iint \cdots \int x_k f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

We can see that:

$$m_{\mathbf{X}k} = \iint \cdots \int x_k f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = \int x_k \left[\int \ldots \int f_{\mathbf{X}}(x_1, \ldots, x_k, \ldots x_n) dx_1 dx_2 \ldots dx_{k-1} dx_{k+1} \ldots dx_n \right] dx_k$$
$$= \int x_k f_{X_k}(x_k) dx_k$$

Each component of the mean vector is computed using its marginal pdf.

The same definition applies for variances. For the vector case, there are *n* variances, one for each component. They are computed using the marginal pdf's.

$$\sigma^2_{\mathbf{X}k} = E\left[\left(X_k - m_{\mathbf{X}k} \right)^2 \right]$$

Equality between random variables:

When dealing with random variables, it is quite hard to define equality between two random variables. We must not forget that a random variable is a mapping. In our course, when we say that two random variables are equal, we mean that they are equal in the mean square sense. We say that two random variables *X* and *Y* are equal in the mean square sense if $E\left[\left(X - Y \right)^2 \right] = 0$

There exist other definitions for the equality but the above one is the more convenient. Using the mean square definition of equality, we can find random variables that have realizations that differ on a discrete set. However, the probability that the random variables take values from this set is zero.

However, we have to take into account the interactions between the different components of the random vector. In the two dimensional case, we can define the covariance between two random variables.

Covariance:

Consider two random variables *X* and *Y*, with a joint pdf $f_{XY}(x, y)$. The covariance between the two variables is:

$$C(X, Y) = E[(X - m_X)(Y - m_Y)] = \iint (x - m_X)(y - m_Y) f_{XY}(x, y) dx dy \quad (1.35)$$

We can also remark that, if the two random variables are independent, the mean of the product is the product of means.

$$E[XY] = E[X]E[Y] \tag{1.36}$$

Furthermore, the covariance of two independent random variables is zero. Using this result, it is easy to show that the variance of the sum of independent random variables is the sum of the variances.

The covariance of two random variables satisfies the following inequality:

$$|C(X,Y)| \le \sigma_X \sigma_Y \tag{1.37}$$

with equality if and only if *X* and *Y* are linearly related (in the mean square sense).

Proof:

The proof is not very difficult and it is quite instructive. Consider the following average:

$$E\left[\{(X - m_X) - \lambda(Y - m_Y)\}^2\right]$$

It is a positive number and it is equal to zero if and only if there exists a value of λ for which $(X - m_X) = \lambda(Y - m_Y)$ in the mean square sense. This means that we are dealing with only one random variable, the other one can be deduced from the first one.

For other values of λ, the average is strictly positive. We can develop the square and we obtain the following quadratic polynomial:

$$\lambda^2 E\left[(Y - m_Y)^2\right] - 2\lambda E\left[(X - m_X)(Y - m_Y)\right] + E\left[(X - m_X)^2\right]$$

$$= \sigma_Y^2 \lambda^2 - 2C(X,Y) + \sigma_X^2 \ge 0$$

We know that a quadratic polynomial will have a constant sign (no real roots) if its discriminant is negative. Since the coefficient of λ^2 is positive, the polynomial will remain positive. The discriminant is:

$\Delta = 4[C(X,Y)]^2 - 4\sigma_X^2 \sigma_Y^2$ and it must be negative. So, $[C(X,Y)]^2 \le \sigma_X^2 \sigma_Y^2$.

(q.e.d.)

in many cases, one would like to know the degree of linear relationship between the variables. This degree is indicated by the correlation coefficient.

Correlation coefficient

$$\rho_{XY} = \frac{C(X,Y)}{\sigma_X \sigma_Y} \tag{1.38}$$

This number is bounded: $-1 \le \rho_{XY} \le 1$.

If $\rho_{XY} = 0$, we say that the two random variables are uncorrelated. We have already seen that independent random variables are uncorrelated. However, the

converse is not true. Uncorrelated random variables are not necessarily independent. When the correlation coefficient is unity (+ or − 1), the two random variables are linearly related (in the mean square sense).

If we have more than two random variables, we can resume the correlation between the different variables in a matrix:

The covariance matrix

Given a random vector $\mathbf{X} = (X_1, X_2, \cdots, X_n)^T$ having a mean $\mathbf{m_X}$, the covariance matrix is defined as:

$$\mathbf{\Sigma_X} = E\left[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\right] \tag{1.39}$$

Show that the covariance matrix is given by:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

where ρ_{ij} is the correlation coefficient between the random variable X_i and X_j and σ_i is the standard deviation of random variable X_i.

Characteristic function

An important average is the characteristic function. It is defined as:

$$\Phi_X(jv) = E\left[e^{jvX}\right] = \int_{-\infty}^{+\infty} e^{jvx} f_X(x)dx \tag{1.40}$$

It is apparent from the definition that the characteristic function is the Fourier transform of the pdf evaluated at the "frequency" $\omega = -v$. So, given the characteristic function $\Phi_X(jv)$, we can recover the pdf by inverse Fourier transform:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_X(jv)e^{-jvx}dv \tag{1.41}$$

Using the characteristic function, we can compute the pdf of a sum of two independent random variables. Let X and Y be two independent random variables. Their joint pdf is $f_{XY}(x,y) = f_X(x)\,f_Y(y)$. Let $Z = X + Y$. Its characteristic function is:

$$\Phi_Z(jv) = E\left[e^{jvZ}\right] = E\left[e^{jv(X+Y)}\right] = E\left[e^{jvX}e^{jvY}\right]$$

Since the average of a product of independent random variables is the product of averages, we obtain finally:

$$\Phi_Z(jv) = E\left[e^{jvX}\right]E\left[e^{jvY}\right] = \Phi_X(jv)\Phi_Y(jv)$$

So, the characteristic function of the sum of two independent random variables is the product of the characteristic functions of each variable. We know that the inverse Fourier transform of the product is the convolution of the inverse transforms. So, the pdf of the sum of two independent random variables is given by:

$$f_Z(z) = f_X(z) * f_Y(z) = \int_{-\infty}^{+\infty} f_X(u) f_Y(z-u) du$$

1.8 Some probability models

Discrete random variables

Binomial distribution

The binomial distribution is encountered in problems of repeated independent trials where we have only two alternatives for each trial. For example, we toss a coin *N* times and we are interested in the probability to have *k* heads in the trials. In digital communication, the repetition of trial consists of transmitting messages of length *N* symbols. In this case, we might be interested in the number of errors in the transmission. So, to describe the binomial distribution, let us consider a probability space $\left(S, \mathcal{F}, P\right)$ and an event $\mathcal{A} \in \mathcal{F}$. let $P(\mathcal{A}) = p$ and $P(\overline{\mathcal{A}}) = q$. Of course, we have $p + q = 1$. We repeat the experiment *N* times and we are interested in the sequence of events that contain *k* times the event $\mathcal{A}$. One possible sequence is: $\underbrace{\mathcal{A}\mathcal{A}\ldots\mathcal{A}}_{k \text{ times}}\underbrace{\overline{\mathcal{A}}\cdots\overline{\mathcal{A}}}_{N-k \text{ times}}$. The probability of this sequence is $p^k q^{N-k}$ because of the independence of the trials. However, this sequence is not the only one that has *k* times the event $\mathcal{A}$. There exist $\binom{N}{k} = \frac{N!}{(N-k)!k!}$ such sequences. So, if we define as a random variable the number of occurrences of the event $\mathcal{A}$ in *N* trials, We define a discrete random variable *K* that can take values in the finite set of integers {0, 1, …,*N*} with probability mass given by

$$P\left(\{K = k\}\right) = \binom{N}{k} p^k q^{N-k} \tag{1.42}$$

The mean of this random variable is: $m_K = Np$

The variance is: $\sigma_K^2 = Npq$

The name of the distribution comes for the binomial development $(p+q)^N$. If we add all possible probabilities, the sum must equal one. So, we must have:

$$1 = \sum_{k=0}^{N} P\left(\{K = k\}\right) = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}$$

However, $(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k}$. Since $p + q = 1$, the previous result is proved.

Example: What is the probability to have 3 heads if we toss a fair coin 5 times. Here, $p = q = 1/2$, $N = 5$ and $k = 3$. So, the probability is:

$$P = \binom{5}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2 = 0.3125$$

If the number of trials becomes very large, the use of the defining formula is quite inconvenient. If the numbers p and q are not too small, the following approximation can be used.

De Moivre-Laplace approximation:

$$P\left(\{K = k\}\right) \approx \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(k - Np)^2}{2Npq}\right] \tag{1.43}$$

However, for typical communication problems, we usually have p very small and N very large, while the product Np remains within reasonable range. A typical example is: We transmit blocks of data of one mega symbols (10^6) and each symbol has a probability of error of 10^{-6}. In this case, the $Np = 1$ and the use of the direct formula is impossible (try to compute 10^6!) while the De Moivre-Laplace approximation will give erroneous results. For this particular case, another approximation is used. It is called the rare event approximation or the Poisson approximation.

Poisson Approximation

If N is very large and p very small, while the product is neither too large nor too small, the Poisson approximation can be used:

Define $m_K = Np$.

$$P\left(\{K = k\}\right) \approx \frac{(m_K)^k}{k!} e^{-m_K} \tag{1.44}$$

In this case, we assume that the range of values that the random variable can take is infinite. In the case of the Poisson approximation, the sum of all probabilities must add to one.

$$\sum_{k=0}^{+\infty} P\left(\{K = k\}\right) = \sum_{k=0}^{+\infty} \frac{(m_K)^k}{k!} e^{-m_K} = e^{-m_K} \sum_{k=0}^{+\infty} \frac{(m_K)^k}{k!} = 1$$

We can also remark that the mean and the variance are equal.

Example:

Let $p = 10^{-6}$ and $N = 10^6$. The probability to have one error in the previously cited problem is:

$$P(\{K=1\}) = e^{-1}$$

The Poisson distribution is also used to model problems arising in Queuing theory. Assume a node in a data communication network. We know that the average rate of arrival of blocks of data at the node is λ blocks/second. We want to compute the probability that k blocks arrive in T seconds. This problem is solved using the Poisson distribution with a mean $m_K = \lambda T$.

$$P(\{K=k\}) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

Continuous distributions

Uniform distribution

We are often confronted to variables that can take values in a bounded range. The initial phase of a sinewave produced by a generator is one example. It can have any value in the interval [0, 2π]. If we cannot favor any portion of the interval, the best model is the uniform one. The random variable X is uniform in the interval [a, b] if its pdf is:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases} \qquad (1.45)$$

The mean is: $m_X = \dfrac{a+b}{2}$

The variance is: $\sigma_X^2 = \dfrac{(b-a)^2}{12}$

Gaussian distribution

This distribution is the most common distribution in probability theory. We will use it to model many physical random phenomena such as noise, scattering phenomenon, etc.

The random variable X has a Gaussian distribution if its pdf is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2} \right] \qquad (1.46)$$

The mean of the distribution is $E[X] = m$ and the variance is $E\left[(X-m)^2 \right] = \sigma^2$.

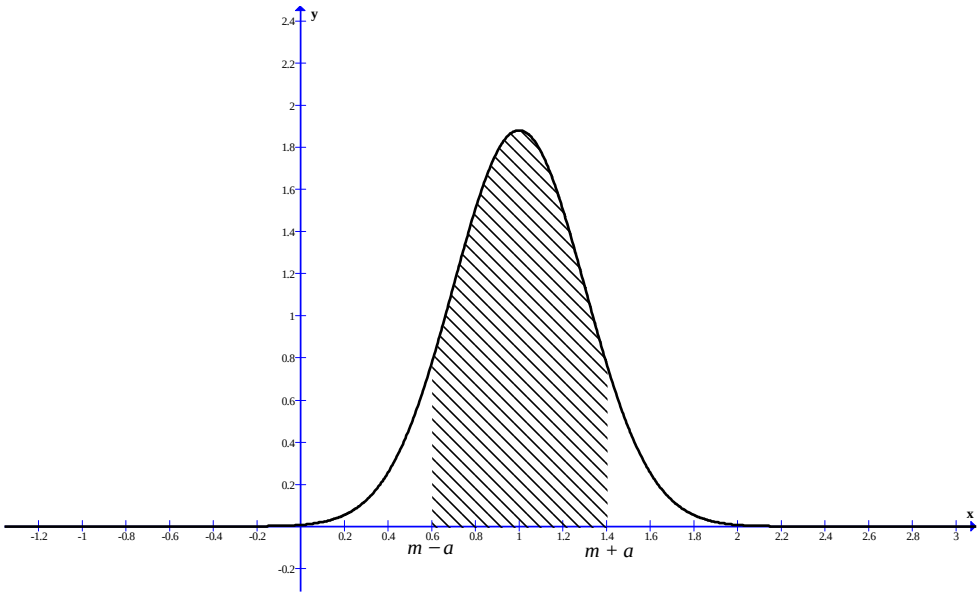


Fig 1- 11 Gaussian pdf with mean 1 and standard deviation 0.3

If we want to compute probabilities using the Gaussian pdf, we have to use tabulated functions. In our course, we will use the "error" function.

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \qquad (1.47)$$

This function can be used to compute the probability $P\left(\left\{\left|X - m\right| \le a\right\}\right)$, i.e. the probability that the random variable will produce a value inside an interval centered on the mean m and of width $2a$. It is the area shown in Fig 1- 11.

$$P\left(\left\{\left|X - m\right| \le a\right\}\right) = \int_{m-a}^{m+a} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] dx$$

Making the change of variable $y = \dfrac{x-m}{\sqrt{2}\sigma}$, we obtain;

$$P\left(\left\{\left|X - m\right| \le a\right\}\right) = \frac{1}{\sqrt{\pi}} \int_{-\frac{a}{\sqrt{2}\sigma}}^{+\frac{a}{\sqrt{2}\sigma}} \exp\left(-y^2\right) dy = \frac{2}{\sqrt{\pi}} \int_0^{+\frac{a}{\sqrt{2}\sigma}} e^{-y^2} dy$$

The last equality comes from the fact that the integrand is even. So, finally, the probability is given by: $P\left(\left\{\left|X - m\right| \le a\right\}\right) = \mathrm{erf}\left(\dfrac{a}{\sqrt{2}\sigma}\right)$.

If we want to compute the probability of the complementary event, we can use the "complementary error" function: $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$. So, the probability that a Gaussian random variable will produce a value away from the mean by a

distance greater than a is: $\operatorname{erfc}\left(\frac{a}{\sqrt{2}\sigma}\right)$. This is the area of the two tails of the graph, i.e. the area not hatched in Fig 1- 11.

When computing probability or error in communication (as in the example), we are interested in the area of just one tail, i.e. the area from a threshold located at a distance a from the mean and infinity. At that time, the probability is:

$$P(E) = \frac{1}{2}\operatorname{erfc}\left(\frac{a}{\sqrt{2}\sigma}\right) \tag{1.48}$$

Most of the tables giving values of the error function are limited to an argument of at most three. For larger values of the argument, you can use the following asymptotic expression:

$$\operatorname{erfc}(x) \approx \frac{\exp\left(-x^2\right)}{x\sqrt{\pi}} \qquad for \qquad x > 3$$

In the communication literature, you will encounter another function used for the evaluation of the probability of error. It is the Q-function. The Q-function is defined as:

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} \exp\left(-\frac{y^2}{2}\right) dy \tag{1.49}$$

This is the area of the tail of a zero mean unit variance Gaussian. It is related to the complementary error function by:

$$Q(x) = \frac{1}{2}\operatorname{erfc}\left(\frac{x}{2}\right) \tag{1.50}$$

and

$$\operatorname{erfc}(x) = 2Q\left(\sqrt{2}x\right) \tag{1.51}$$

So, given an expression using the Q-function, you can use (1.51) to obtain the equivalent expression using the complementary error function and if you are given an expression using the complementary error function, you can use (1.50) to obtain the expression using the Q-function.

We can also define Gaussian random vectors. They are defined by the multivariate Gaussian pdf:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right] \tag{1.52}$$

In equation(1.52), the random vector $\mathbf{X} = \left(X_1, X_2, \cdots, X_N\right)^T$ is N dimensional. Its mean is the vector **m** and the covariance matrix is the $N \times N$ matrix $\mathbf{\Sigma}$.

When the random variables are uncorrelated, the covariance matrix becomes diagonal.

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

In this case, the joint pdf can be written as:

$$f_X(x_1, x_2, \ldots, x_N) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma_1 \sigma_2 \ldots \sigma_N} \exp\left[-\frac{1}{2} \left\{ \frac{(x_1 - m_1)^2}{\sigma_1^2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} + \cdots + \frac{(x_N - m_N)^2}{\sigma_N^2} \right\} \right]$$

We see that the joint pdf is the product of one-dimensional pdf's. So, for Gaussian random variables, uncorrelation implies independence.

Another very important property of Gaussian random variables is that any linear combination of Gaussian random variables is a Gaussian random variable.

1.9 The central limit theorem

The fact that many random phenomena can be modeled by Gaussian variables is due to the fact that the sum of many bounded random variables tends toward Gaussian. This behavior is explained by the Central Limit Theorem (CLT). There are many different statements of the theorem. We are going to state one of them without giving the proof.

Theorem:

Given N independent random variables X_1, X_2, …, X_N with means m_1, m_2, …, m_N and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. Let

$$Z = \sum_{k=1}^{N} X_k \,,\; m = \sum_{k=1}^{N} m_k \text{ and } \sigma^2 = \sum_{k=1}^{N} \sigma_k^2$$

If all the variances are bounded, then for large N, we have:

$$P\left[a \le Z \le b\right] \approx \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-m)^2}{2\sigma^2} \right]$$

and if the random variables are of continuous type:

$$f_Z(z) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z-m)^2}{2\sigma^2} \right]$$

♦

The approximation becomes more precise as N increases. It is also more accurate around the center of the distribution rather than near the tail.

It is the central limit theorem that allows us to conclude that the thermal noise produced by metallic resistors is Gaussian. This voltage is due to the superposition of the small voltages produced by the vibration of ions in metallic crystals.

We can also use the central limit theorem to analyze slow fading channels.

In this section, we are going to analyze Rayleigh and Rice fading channels.

Rayleigh fading

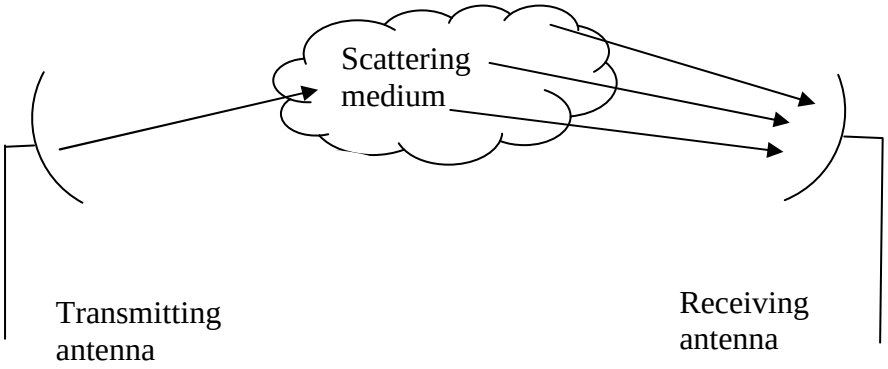


Fig 1- 12 Rayleigh Fading

Fig 1- 12 is an illustration of the mechanism of Rayleigh fading. The waveform transmitted passes through a scattering medium (usually an ionized cloud). The received waveform is a sum of a very large number of waveforms. This is due to the fact that every particle in the cloud receives the waveform from the transmitter and radiates it back. So, every component of the received waveform is a small fraction of the transmitted one and every one of them has a random delay.

We assume that the received waveform would be a sinewave with unity amplitude if there is no scattering medium in the path. We denote this waveform as: $s_i(t) = \cos \omega_0 t$.

The scattered waveform, on the other hand, is the following summation:

$s_r(t) = \sum_k A_k \cos\left(\omega_0 t + \Phi_k\right)$. A_k and Φ_k are random variables. It is quite reasonable to assume that the different random variables are independent. Furthermore, the variances of the A_k must be bounded since each reradiated wave is a small fraction of the incident one. Another reasonable assumption is that the

phase shifts are uniformly distributed in the interval [0, 2π]. Developing the cosine term in the summation, we obtain:

$$s_r(t) = X\cos\omega_0 t - Y\sin\omega_0 t \text{ where:}$$

$$X = \sum_k A_k \cos\Phi_k \text{ and } Y = \sum_k A_k \sin\Phi_k$$

The random variables *X* and *Y* are the sum of a large number of bounded independent random variables. The central limit theorem applies and we can conclude that they are Gaussian. The mean of *X* is:

$$E[X] = E\left[\sum_k A_k \cos\Phi_k\right] = \sum_k E[A_k \cos\Phi_k] = \sum_k E[A_k]E[\cos\Phi_k]$$

$$E[\cos\Phi_k] = \int_0^{2\pi} \cos\phi_k \frac{d\phi_k}{2\pi} = 0$$

So $E[X] = 0$; The same demonstration shows that $E[Y] = 0$.

The variance of *X* is given by:

$$E[X^2] = E\left[\sum_k A_k \cos\Phi_k \sum_j A_j \cos\Phi_j\right] = \sum_k \sum_j E[A_k A_j \cos\Phi_k \cos\Phi_j]$$

If $k \neq j$, the different random variables are independent and the average of a product is the product of averages. However, we have seen that $E[\cos\Phi_k] = 0$. So, the double summation reduces to a summation of squares.

$$E[X^2] = \sum_k E[A_k^2]E[\cos^2\Phi_k] = \sum_k \frac{E[A_k^2]}{2} = \sigma^2$$

We obtain the same result for the variance of obtain the same result for the variance of *Y*. $E[X^2] = E[Y^2] = \sigma^2$

We have now to compute the covariance between *X* and *Y*.

$$E[XY] = E\left[\sum_k A_k \cos\Phi_k \sum_j A_j \sin\Phi_j\right] = \sum_k \sum_j E[A_k A_j]E[\cos\Phi_k \sin\Phi_j]$$

As for the variance, the terms with $k \neq j$ are all zero.

$$E[XY] = \sum_k E[A_k^2]E[\cos\Phi_k \sin\Phi_k] = \sum E[A_k^2]E\left[\frac{1}{2}\sin 2\Phi_k\right]$$

So, finally, *E*[*XY*] = 0.

So, the received waveform has quadrature components that are zero mean independent Gaussian random variable with same variance σ^2.

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$$

We can also represent the received waveform using the amplitude-phase representation.

$$s_r(t) = R\cos\left(\omega_0 t + \Theta\right)$$

where $R = \sqrt{X^2 + Y^2}$ and $\Theta = \tan^{-1}\dfrac{Y}{X}$

The inverse relations are: $X = R\cos\Theta$ and $Y = R\sin\Theta$

In the amplitude-phase representation, the variables have a restricted range: $R \geq 0$ and $-\pi \leq \Theta \leq \pi$. The *R*, Θ variables correspond to polar coordinates. The jacobian of the transformation is *r*. So, the pdf expressed in this new set of coordinates is:

$$f_{R\Theta}(r,\theta) = \begin{cases} \dfrac{r}{2\pi\sigma^2}\exp\left(-\dfrac{r^2}{2\sigma^2}\right) & r \geq 0 \ ; \ -\pi \leq \theta \leq \pi \\ 0 & \text{elsewhere} \end{cases} \tag{1.53}$$

The marginal densities are:

$$f_R(r) = \begin{cases} \dfrac{r}{\sigma^2}\exp\left(-\dfrac{r^2}{2\sigma^2}\right) & r \geq 0 \\ 0 & r < 0 \end{cases} \tag{1.54}$$

$$f_\Theta(\theta) = \begin{cases} \dfrac{1}{2\pi} & -\pi \leq \theta \leq \pi \\ 0 & \text{elsewhere} \end{cases} \tag{1.55}$$

We remark that the two variables *R* and Θ are independent. The added phase is uniformly distributed in a 2π wide interval. This random phase precludes the use of modulations that must have a precise phase reference. The amplitude of the received wave is multiplied by a random number distributed according to the law described by equation(1.54). This distribution is called the Rayleigh distribution.

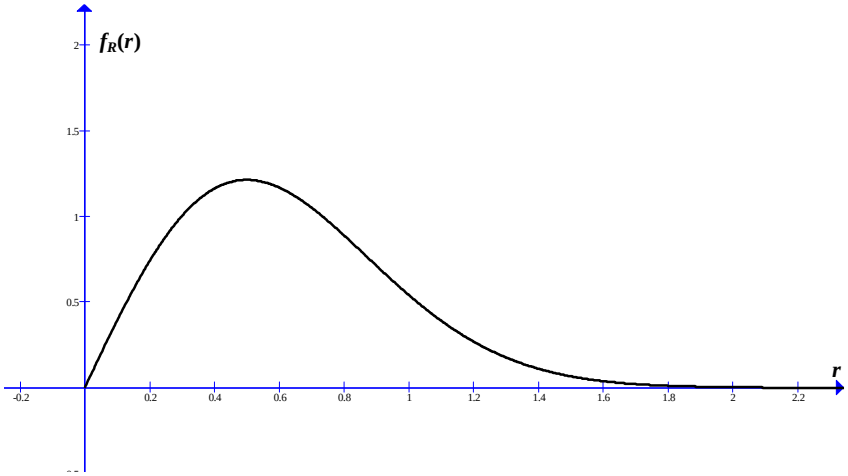


Fig 1- 13 Rayleigh pdf

The Rayleigh pdf is a unimodal pdf (it has a single local maximum).

Its mode (maximum) is located at $r = \sigma$.

Its mean is $E[R] = \sqrt{\frac{\pi}{2}}\sigma$, the variance is: $\left(2 - \frac{\pi}{2}\right)\sigma^2$

Rice fading

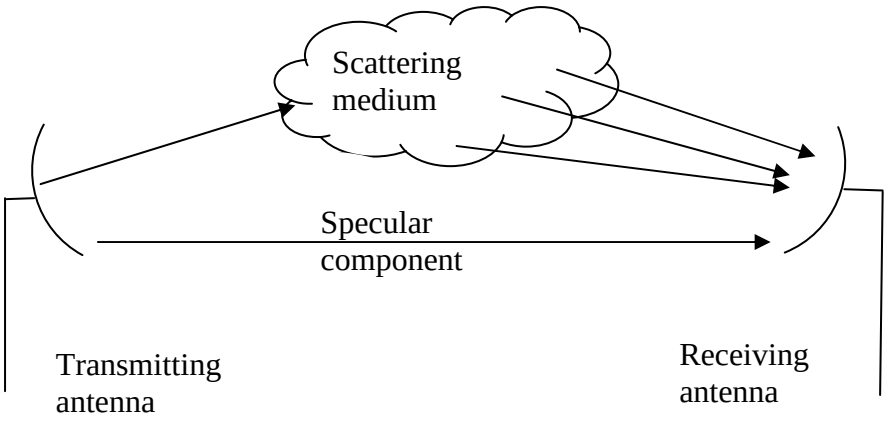


Fig 1- 14 Rice fading

In the Rice fading, the scattering medium does not cover completely the propagation path. Part of the signal propagates directly from the transmitter to the receiver. This component is called the "specular" component.

The specular component will produce a signal $A_0 \cos \omega_0 t$ at the receiver and we still have a scattered signal arriving at this receiver. So, the total signal at the receiver is: $s_r(t) = A_0 \cos \omega_0 t + X \cos \omega_0 t - Y \sin \omega_0 t$, where X and Y are independent zero mean Gaussian random variable with variance σ^2.

We can re-express the received signal as: $s_r(t) = X_1 \cos \omega_0 t - Y \sin \omega_0 t$.

$X_1 = X + A_0$ and Y are also independent Gaussian random variable with a variance σ^2. However, the mean of X_1 is not zero but it is equal to A_0. The joint pdf of X_1 and Y is:

$$f_{X_1Y}(x_1, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(x_1 - A_0)^2 + y^2}{2\sigma^2} \right] \tag{1.56}$$

In this case also, it is interesting to express the received signal in the amplitude phase representation.

$$s_r(t) = R\cos(\omega_0 t + \Theta)$$

$$R = \sqrt{(A_0 + X)^2 + Y^2} = \sqrt{X_1^2 + Y^2} \text{ and } \Theta = \tan^{-1}\frac{Y}{X + A_0} = \tan^{-1}\frac{Y}{X_1}$$

In this case also, we have a Cartesian to polar transformation. So, the pdf for the amplitude and phase is given by:

$$f_{R\Theta}(r, \theta) = r f_{X_1Y}(r\cos\theta, r\sin\theta)$$

$$f_{R\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} \exp\left[-\frac{(r\cos\theta)^2 + A_0^2 - 2rA_0\cos\theta + (r\sin\theta)^2}{2\sigma^2} \right]$$

$$f_{R\Theta}(r, \theta) = \frac{r}{2\pi\sigma^2} \exp\left[-\frac{r^2 + A_0^2}{2\sigma^2} \right] \exp\left[\frac{rA_0}{\sigma^2}\cos\theta \right] \tag{1.57}$$

Equation (1.57) is the expression of the joint pdf of R and Θ. We clearly see that it is impossible to express this pdf as a product of two functions, one being only a function of r, the other only a function of θ. So, it is evident that the two variables are dependent.

The marginal pdf of the envelop is given by:

$$f_R(r) = \int_{-\pi}^{\pi} f_{R\Theta}(r, \theta) d\theta = \frac{r}{\sigma^2} \exp\left[-\frac{r^2 + A_0^2}{2\sigma^2} \right] \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{rA_0}{\sigma^2}\cos\theta \right) d\theta$$

Using the fact that $I_0(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(z\cos\theta)\, d\theta$, where $I_0(z)$ is the zero order modified Bessel function of the first kind, we obtain finally:

$$f_R(r) = \begin{cases} \frac{r}{\sigma^2} \exp\left[-\frac{r^2 + A_0^2}{2\sigma^2} \right] I_0\left(\frac{rA_0}{\sigma^2} \right) & r \geq 0 \\ 0 & r < 0 \end{cases} \tag{1.58}$$

Equation (1.58) is the expression of the Rice distribution. When the ratio $\frac{A_0}{\sigma}$ is large, we can show that this pdf can be approximated very closely by a Gaussian:

$$f_R(r) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(r - A_0)^2}{2\sigma^2} \right] \tag{1.59}$$

If $\frac{A_0}{\sigma}$ is very small, the Rice pdf becomes identical to the Rayleigh pdf.

Chapter 2

Introduction to Stochastic Processes

2.1 Basic definitions

In communication, all the different waveforms that we will encounter have some degree of randomness. We have already stated in the introduction some of the reasons for studying random signals. In probability theory, these signals are called "stochastic processes". A stochastic process is a generalization of the concept of random variable.

So, given a probability space $\left(S,\mathcal{F},P\right)$, a stochastic process x is a mapping from the sample space S to a function space such that to every outcome ζ in S, it associates a function of time $x(t, \zeta)$. The mapping must be so that, for any time t_0, $x(t_0, \zeta)$ is a random variable.

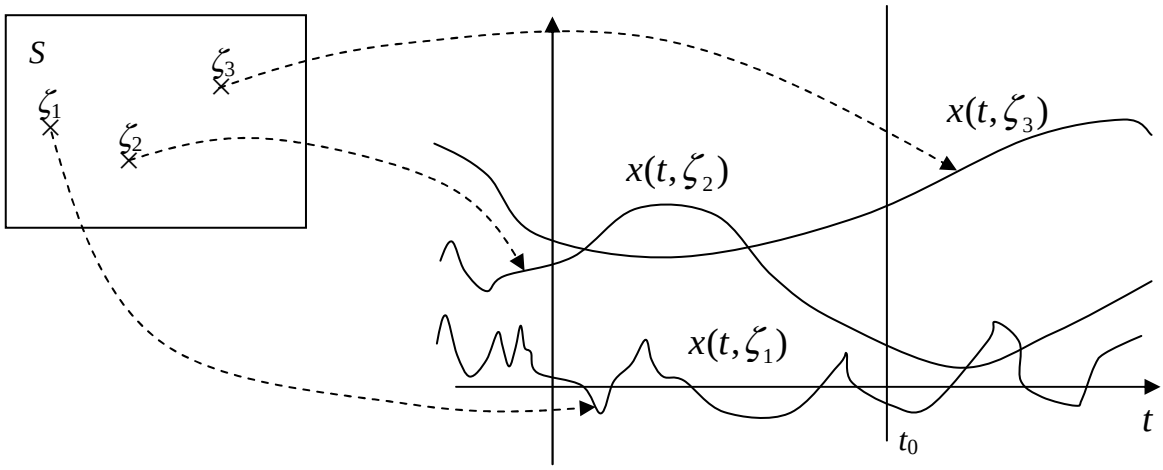


Fig.2- 1 Illustration of the mapping

The set of all functions that can be generated by the stochastic process is called the "ensemble" of functions.

Example: assume we toss an infinite number of times a fair coin. The outcome of the experiment is an infinite sequence of heads and tails in any order. A possible outcome can be:

$$\zeta = \ldots\text{HTHHTTTHT}\ldots$$

To the k^{th} toss, we associate a voltage as follows:

$$x(t,\zeta)=\begin{cases}+A & (k-1)T<t\le kT \quad \text{if } k^{th} \text{ result is head}\\ -A & (k-1)T<t\le kT \quad \text{if } k^{th} \text{ result is tail}\end{cases}$$

For the sequence given above, the corresponding waveform is:

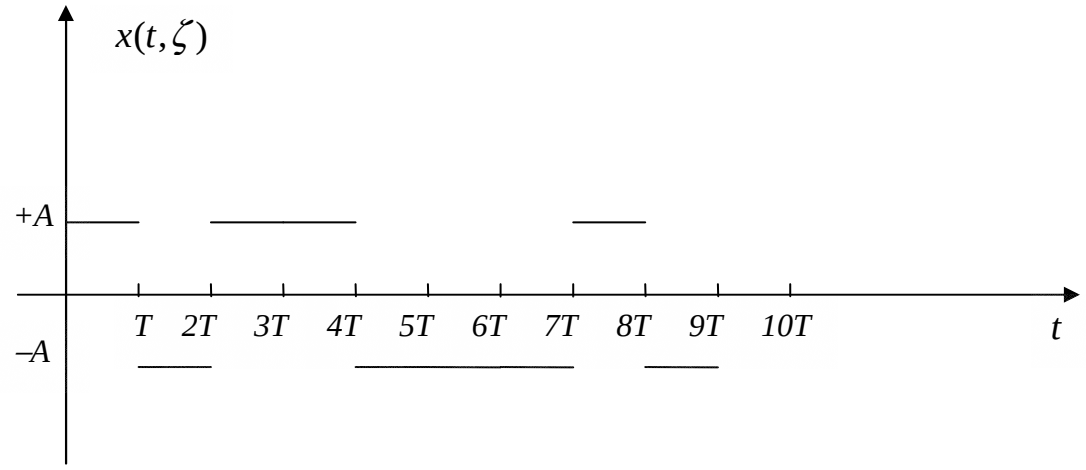


Fig.2- 2 Sample function from the process

The waveform corresponding to the outcome is called a sample function. The above process can represent the waveform generated by a synchronous serial interface between a computer and a terminal. This type of waveform is called "non return to zero" (NRZ) in data communication terminology. It is evident that the waveform is going to depend on the message to be transmitted.

If we want to characterize a stochastic process from a probabilistic point of view, we must remember that the process seen at a fixed instant t_0 is a random variable and as such it can be characterized by the pdf of this random variable. So, at a time t_0, the process $x(t,\zeta)$ is characterized by its "first order" pdf: $f_x(x_0;t_0)$ such that $f_x(x_0;t_0)dx_0 = P\left[x_0 \le x(t_0,\zeta) \le x_0 + dx_0\right]$, i.e. the probability to find the process having values in an interval around x_0 of width dx_0.

If now we consider the process $x(t,\zeta)$ at two different instants t_1 and t_2, we are in the presence of two random variables $x(t_1,\zeta)$ and $x(t_2,\zeta)$. So, we should characterize them by their joint pdf ("second order" pdf). This joint pdf is such that:

$$f_x(x_1,x_2;t_1,t_2)dx_1dx_2 = P\left[\left\{x_1 \le x(t_1) \le x_1 + dx_1\right\} and \left\{x_2 \le x(t_2) \le x_2 + dx_2\right\}\right]$$

At this point, as we did with random variable, we simplify the notation and write simply $x(t)$ instead of $x(t,\zeta)$ to indicate the random variable created by the process at the instant t. We can continue this process of sampling the random process and we can define the n^{th} order pdf.

Given n instants t_1, t_2, …, t_n, the n^{th} order pdf is such that

$$f_x(x_1,x_2,...,x_n;t_1,t_2,...,t_n)dx_1dx_2...dx_n = P\left[\left\{x_1 \le x(t_1) \le x_1 + dx_1\right\} and \left\{x_2 \le x(t_2) \le x_2 + dx_2\right\} and ... \left\{x_n \le x(t_n) \le x_n + dx_n\right\}\right]$$

In order to characterize completely a stochastic process, we must know the pdf's of all orders, i.e. the n^{th} order pdf for all n.

We can also define a pair of stochastic processes. The pair of random processes is characterized by joint pdf's, starting from the second order one. So, the second order pdf of the pair of stochastic processes $x(t)$ and $y(t)$ is the joint pdf of the pair of random variables $x(t_1)$ and $y(t_2)$ seen respectively at time t_1 and at time t_2. Of course, we can also define third order and even higher order pdf's for a pair of random processes.

One particular pair of stochastic processes is the one that defines a complex process $z(t) = x(t) + jy(t)$ where x and y are real processes.

2.2 Averages

It is quite infrequent to have the complete characterization of a stochastic process. In many cases, we have only access to averages. The averages that are commonly used are the "first order" averages (using the first order pdf) and the "second order" averages (using the second order pdf). The definitions are given for complex processes. However, the definitions using pdf's will be given for real processes.

First order averages

The first order averages that we can define are: the mean, the second order moment and the variance.

Mean

Given the process $x(t)$, we define its mean as:

$$m_x(t) = E\left[x(t)\right] = \int_{-\infty}^{+\infty} x f_x(x;t)dx \tag{2.1}$$

In general, the mean is a function of time. It is computed at a given value of time and it is an ensemble average. The variable of integration is not time.

The second order moment is defined as:

$$E\left[\left|x(t)\right|^2\right] = \int_{-\infty}^{+\infty} \left|x\right|^2 f_x(x;t)dx \tag{2.2}$$

and the variance is:

$$\sigma_x^2(t) = E\left[\left|x(t) - m_x(t)\right|^2\right] = \int_{-\infty}^{+\infty} \left|x - m_x(t)\right|^2 f_x(x;t)dx \tag{2.3}$$

If we develop the square, we obtain:

$$E\left[\left|x(t)\right|^2\right] = \sigma_x^2(t) + \left|m_x(t)\right|^2$$

Second order averages

The second order averages are going to be functions of two time variables. They consist on the different correlation and covariance functions.

If we are given a complex process $x(t)$, the autocorrelation function is the following function:

$$R_x(t_1,t_2) = E\left[x(t_1)x^*(t_2) \right] \tag{2.4}$$

For a real process, we have: $R_x(t_1,t_2) = \iint x_1 x_2 f_x(x_1,x_2;t_1,t_2)dx_1dx_2$. The domain of integration is the whole plane.

The autocovariance of the process is defined as:

$$C_x(t_1,t_2) = E\left[\left(x(t_1) - m_x(t_1) \right)\left(x(t_2) - m_x(t_2) \right)^* \right] \tag{2.5}$$

The autocovariance and the autocorrelation are related.

$$R_x(t_1,t_2) = C_x(t_1,t_2) + m_x(t_1)m_x^*(t_2)$$

and if we evaluate these functions with $t_1 = t_2 = t$, we obtain:

$R_x(t,t) = E\left[\left| x(t) \right|^2 \right]$, the second order moment

$C_x(t,t) = \sigma_x^2(t)$, the variance.

If we consider now two different processes $x(t)$ and $y(t)$, we can define the cross-correlation function:

$$R_{xy}(t_1,t_2) = E\left[x(t_1)y^*(t_2) \right] \tag{2.6}$$

and the cross-covariance function:

$$C_{xy}(t_1,t_2) = E\left[\left(x(t_1) - m_x(t_2) \right)\left(y(t_1) - m_y(t_2) \right)^* \right] \tag{2.7}$$

We also have the same type of relationship between cross-covariance and cross-correlation as we had between autocovariance and autocorrelation:

$$R_{xy}(t_1,t_2) = C_{xy}(t_1,t_2) + m_x(t_1)m_y^*(t_2)$$

When we consider two different processes $x(t)$ and $y(t)$, we say that x and y are orthogonal if $R_{xy}(t_1,t_2) = 0$ for all values of t_1 and t_2.

We say that the two processes are uncorrelated if $C_{xy}(t_1,t_2) = 0$ for all values of t_1 and t_2.

Consider now the sum of the two processes $z(t) = x(t) + y(t)$. If x and y are orthogonal, then $R_z(t_1,t_2) = R_x(t_1,t_2) + R_y(t_1,t_2)$

If they are uncorrelated, then $C_z(t_1,t_2) = C_x(t_1,t_2) + C_y(t_1,t_2)$

2.3 Stationarity

In many communication problems, we can be interested in processes that have statistical properties that do not change with time. We define the class of stationary random processes.

Definition:

A "strict sense" stationary random process is a process that has statistical properties that are independent on the time origin.

This means that all the pdf's do not change if we add a constant time shift.

$$\forall \lambda \therefore f_x(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = f_x(x_1, x_2, ..., x_n; t_1 + \lambda, t_2 + \lambda, ..., t_n + \lambda)$$

In particular, the first order pdf satisfies:

$$\forall \lambda \therefore f_x(x;t) = f_x(x;t+\lambda)$$

This simply means that the first order pdf does not depend on time. And this implies that all the first order averages are constant (time independent).

So, $m_x(t) = m_x$ = constant and $\sigma_x^2(t) = \sigma_x^2$ = constant.

The second order pdf satisfies also the same relation:

$\forall \lambda \therefore f_x(x_1, x_2; t_1, t_2) = f_x(x_1, x_2; t_1 + \lambda, t_2 + \lambda)$

The time shift λ being arbitrary, we can set it at any value. Let $\lambda = -t_2$. We obtain: $f_x(x_1, x_2; t_1, t_2) = f_x(x_1, x_2; t_1 - t_2, 0)$. So the second order pdf depends only on the time difference $t_1 - t_2$ and not on their values. This implies that the second order statistics do not depend on two variables but only on the time difference between the two instants. So, if the process $x(t)$ is stationary, we can define the autocorrelation and autocovariance as:

$$R_x(\tau) = R_x(t, t-\tau) = E\left[x(t)x^*(t-\tau) \right] \tag{2.8}$$

and

$$C_x(\tau) = C_x(t, t-\tau) = E\left[\left(x(t) - m_x \right) \left(x(t-\tau) - m_x \right)^* \right] \tag{2.9}$$

So, we see that the autocorrelation and autocovariance functions depend on only one variable. The variable is the time delay τ (called 'time lag") between two different instants of the signal. They are thus function of the size of the interval (τ) but not on its position (t).

Of course, the properties derived for non stationary signals apply to stationary ones.

$$R_x(\tau) = C_x(\tau) + \left| m_x \right|^2$$

Evaluating these functions at zero lag gives:

$$R_x(0) = E\left[\left| x(t) \right|^2 \right] \text{ and } C_x(0) = \sigma_x^2$$

Joint stationarity of two processes:

Two processes x(t) and y(t) are jointly stationary if they are stationary and their joint statistics are independent on the position of the time origin.

This implies that the cross correlation function is:

$$R_{xy}(\tau) = R_{xy}(t, t-\tau) = E\left[x(t) y^*(t-\tau) \right] \tag{2.10}$$

In order to be able to state that a given process is stationary in the strict sense (sss), we have to show that the pdf's of all orders are invariant to time translation. For general processes, this operation is usually very hard to do and in many case impossible.

However, we can test or compute the first and second order statistics for many processes. A wider definition of stationarity can be given:

Wide sense stationarity

A process is said to be wide sense stationary (wss) if the mean of the process is constant and if the autocorrelation function depends on the time lag and not on the position of the time window.

$E\left[x(t) \right] = m_x$ =constant

$E\left[x(t) x^*(t-\tau) \right] = R_x(t, t-\tau) = R_x(\tau)$ independent on *t*.

It is evident that a strict sense stationary process is also wide sense stationary. But the converse is not true. However, if the process is Gaussian, wide sense stationarity implies strict sense. We define also joint wide sense stationarity.

Symmetry properties of correlation functions

Consider a general process $x(t)$, interchanging the two times t_1 and t_2 produces:

$$R_x(t_2, t_1) = E\left[x(t_2) x^*(t_1) \right] = \left\{ E\left[x(t_1) x^*(t_2) \right] \right\}^*$$

$$R_x(t_1, t_2) = R_x^*(t_2, t_1) \tag{2.11}$$

If the process is stationary, then $R_x(\tau) = R_x^*(-\tau)$

If we consider two processes $x(t)$ and $y(t)$, the interchange produces;

$$R_{xy}(t_2, t_1) = R_{yx}^*(t_1, t_2) \tag{2.12}$$

If the two processes are jointly stationary, then $R_{xy}(\tau) = R_{yx}^*(-\tau)$

2.4 Ergodicity

In order to compute the different averages, we must either have an analytic formulation of the process or we must have a large collection of sample functions.

At that time, in order to evaluate the mean, for example, we must observe the different sample functions at a given time *t* and perform a statistical averaging of the different values at that particular time. This procedure is usually impossible to conduct. In many cases, we have at our disposal only one sample of the signal.

Hopefully, there exist a class of stochastic processes that have the property that one sample of the process represents all the ensemble. They are the "ergodic" processes. They have the property that the time average of one sample function is equal to the ensemble average. Of course, these signals are strict sense stationary.

The time average of a signal is:

$$< x(t) >= \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)dt \qquad (2.13)$$

So, an ergodic process is such that all ensemble averages can be computed using time averages of one observed sample.

For an ergodic process, the mean of the process is:

$m_x = E\left[x(t)\right] =< x(t,\zeta_k) >= \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t,\zeta_k)dt$ and the autocorrelation is:

$$R_x(\tau) = E\left[x(t)x^*(t-\tau) \right] = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t,\zeta_k)x^*(t-\tau,\zeta_k)dt$$

where $x(t,\zeta_k)$ is the observed sample function. If the process where not ergodic, these time averages would be random variables. The advantage of having this property of ergodicity is that these averages are measurable and they can be estimated from observation.

Another point worth noting for ergodic processes is the physical meaning of the different processes. We have already seen that, for any stationary process, we have:

$R_x(0) = \sigma_x^2 + \left|m_x\right|^2$ and since $m_x = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)dt$, m_x is the dc value of the signal.

$R_x(0) = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left|x(t)\right|^2 dt$ is the definition of the total power of the signal *x*(*t*). So, the previous relation can be interpreted as "the total power of the signal is the sum of the ac power and the dc power" if we understand that σ_x^2 is the ac power of the signal.

The following circuit can be used to evaluate the autocorrelation function for an ergodic process:

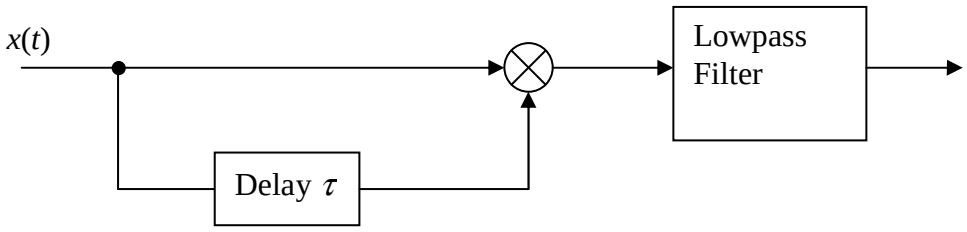


Fig.2- 3 Autocorrelation function evaluation

The use of a lowpass filter is justified from the fact that evaluating the time average of a signal amounts to evaluating the dc value of the signal. The lowpass filter is used to evaluate the time average must have a very small cut-off frequency.

2.5 Periodic process

A process is periodic if it is equal (in mean square sense) to a shifted value of itself. So, the process $x(t)$ is periodic if:

$$\exists T > 0 \therefore \forall t \in \mathbb{R} \quad ; \quad E\left[\left|x(t+T) - x(t)\right|^2\right] = 0 \tag{2.14}$$

If the process is real and stationary, we can show (exercise) that the autocorrelation function is also periodic with the same period.

$$R_x(t+T) = R_x(t)$$

2.6 Frequency analysis of random processes

In this section, we are going to consider only stationary (wss) stationary processes. These processes have the property to have the same statistics at all times (even at infinity). So, it is evident that the probability for an energy type of signal to be a member of the ensemble must be zero, since the energy signal must decay to zero at infinity and must have then a time varying variance. So, most probably, stationary random processes samples belong to the class of at least power type of signals. In fact, in the previous paragraph, we have seen that the ergodic signals that have a finite autocorrelation function at a zero lag belong to the class of power signals (on average).

In order to be able to analyze the distribution of this power in the frequency domain, we introduce the concept of "power spectrum" or "power spectral density" (psd). The power spectrum is defined as:

Given a stationary process $x(t)$, the power spectrum of x is defined as:

$$S_x(f) = \int_{-\infty}^{+\infty} R_x(\tau) \exp\left(-j2\pi f \tau\right) d\tau \tag{2.15}$$

and

$$R_x(\tau) = \int_{-\infty}^{+\infty} S_x(f) \exp(j2\pi f\tau)\, df \tag{2.16}$$

We can observe that $S_x(f)$ is a density of power and not a power by evaluating equation (2.16) for $\tau = 0$.

$$R_x(0) = \int_{-\infty}^{+\infty} S_x(f)\, df \tag{2.17}$$

The above relation shows clearly that the power spectrum is a density of power.

There is an alternate definition for the power spectrum. We have already said that a sample from a stationary random process has a probability zero to be an energy signal and as such, it is hard to define a Fourier transform for it. Assume we are given a stationary random process $x(t)$. Let $x(t,\zeta)$ be a sample function from it and consider a windowed version of it.

$$x_T(t,\zeta) = \begin{cases} x(t,\zeta) & -\frac{T}{2} \le t \le \frac{T}{2} \\ 0 & \text{elsewhere} \end{cases}$$

The windowed signal has a finite duration T and is an energy signal. We can compute its Fourier transform.

$$X_T(f,\zeta) = \int_{-\infty}^{+\infty} x_T(t,\zeta) \exp(-j2\pi ft)\, dt$$

The energy of this sample is:

$$E_T(\zeta) = \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t,\zeta)|^2\, dt = \int_{-\infty}^{+\infty} |X_T(f,\zeta)|^2\, df$$

So, $|X_T(f,\zeta)|^2$ is a density of energy. But this density is a random variable. The average energy spectral density is then $E\left[|X_T(f,\zeta)|^2\right]$ and the average power density for the duration T is: $\frac{1}{T}E\left[|X_T(f,\zeta)|^2\right]$. If we let the duration T become infinite, we obtain the power spectral density or power spectrum:

$$S_x(f) = \lim_{T\to\infty} \frac{1}{T} E\left[|X_T(f,\zeta)|^2\right] \tag{2.18}$$

This alternate definition is useful when we want to compute the power spectrum of some given signals. The definition given by equation (2.18) shows also that the power spectrum is a real positive number.

When we deal with two different processes, we can define cross power spectra.

$$S_{xy}(f) = \int_{-\infty}^{+\infty} R_{xy}(\tau) \exp(-j2\pi f \tau) d\tau \qquad (2.19)$$

Properties of power spectra and correlation functions

1. The autocorrelation function is maximum at the origin:
$$R_x(0) \geq |R_x(\tau)|$$
2. The power spectrum is real and positive.
3. If the process $x(t)$ does not contain any periodic component, we have:
$$\lim_{\tau \to \infty} R_x(\tau) = |m_x|^2$$
4. If the process is real and periodic, its autocorrelation is also periodic with the same fundamental period.
5. If we consider two processes x and y: $S_{xy}(f) = S_{yx}^*(f)$

Example:

To illustrate property 3 and 4, let us consider the measurement of hidden periodicity. We measure a process $z(t) = x(t) + n(t)$. The process $x(t)$ is periodic, with period T_0 and the process $n(t)$ is a zero mean, non periodic process, independent on $x(t)$. We also assume that all processes are ergodic. So, we can use DSP techniques to evaluate the autocorrelation function of $z(t)$. Due to the stated assumptions, we have:

$$R_z(\tau) = R_x(\tau) + R_n(\tau)$$

If we plot this autocorrelation function for large values of lag, property 3 states that $R_n(\tau) \approx 0$ for large τ. So, we obtain $R_z(\tau) \approx R_x(\tau)$ for large τ.

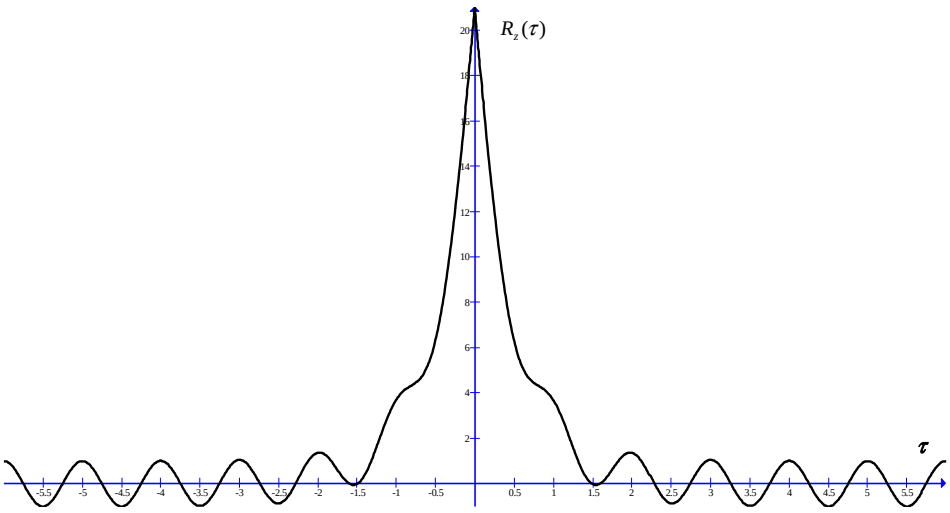


Fig.2- 4 Autocorrelation of $z(t)$

For the example shown in Fig.2- 4, we clearly see that after the lag value $\tau = 2$, the autocorrelation function is periodic.

2.7 Gaussian process

We have seen that the Gaussian distribution occurs quite often and can be used to describe many random phenomena. A Gaussian process is a process that has as nth order pdf a multivariate Gaussian. So, if the process $x(t)$ is Gaussian, then:

Let us consider the random vector $\mathbf{X} = (X_1, X_2, ..., X_n)^T$ corresponding to the observation of the process at times $\mathbf{t} = (t_1, t_2, \ldots, t_n)^T$. The nth order pdf is:

$$f_{\mathbf{X}}(\mathbf{x};\mathbf{t}) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{\det(\boldsymbol{\Sigma})}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{m})\right]$$

where $\mathbf{m} = (m_x(t_1), m_x(t_2), ..., m_x(t_n))^T$ is the mean vector and the matrix $\boldsymbol{\Sigma}$ is the covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} C_x(t_1,t_1) & C_x(t_1,t_2) & \cdots & C_x(t_1,t_n) \\ C_x(t_2,t_1) & C_x(t_2,t_2) & \cdots & C_x(t_2,t_n) \\ \vdots & \vdots & \ddots & \vdots \\ C_x(t_n,t_1) & C_x(t_n,t_2) & \cdots & C_x(t_n,t_n) \end{pmatrix}$$

If the process is wide sense stationary, the mean is a constant and the autocovariance function will be such that: $C_x(t_1,t_2) = C_x(t_1 - t_2)$. This implies that the process is also strict sense stationary. As a conclusion, a wide sense stationary Gaussian process is also strict sense stationary.

Another important property of Gaussian processes is that fact that any linear combination of Gaussian processes is Gaussian.

White noise

A process $n(t)$ that has a constant power spectrum at all frequencies is called "white noise" by analogy with white light. White light is a wave that contains the same amount of power at all visible colors. So, if the process is stationary, the power spectrum is:

$$S_n(f) = \frac{N_0}{2} \tag{2.20}$$

By computing the inverse Fourier transform, we find:

$$R_n(\tau) = \frac{N_0}{2}\delta(\tau) \tag{2.21}$$

The power (variance) of white noise is infinite and the dc value is zero (why?). This means that white noise is a mathematical abstraction and is not a process that exists in nature. However, it is a very useful concept. In general, we

work with bandlimited signals and systems. So, if the considered signal has a power spectrum that is constant over the frequency band of interest, then we can consider it as white.

The constant N_0 is called the single sided power spectrum. A typical white noise is the thermal noise. A metallic resistor produces a Gaussian process that can be considered as white up to frequencies of many hundreds of GHz. Nyquist has shown that a metallic resistor is equivalent to a noiseless resistor in series with a voltage source that produces a white Gaussian noise with $N_0 = 4kTR$, where $k = 1.38\ 10^{-23}$ J/°K is the Boltzmann constant and R is the value of the resistance in Ω.

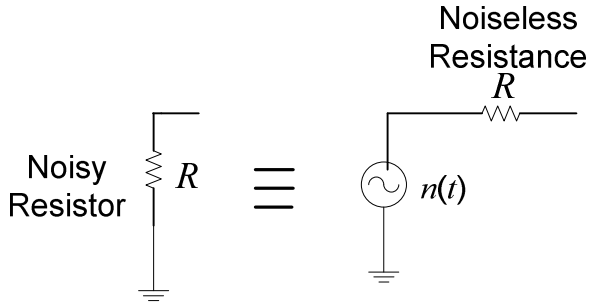


Fig.2- 5 Noisy metallic resistor

The fact that the autocorrelation function is an impulse function means that any two samples of the white noise taken at instants that are arbitrary close will be uncorrelated. If the noise is Gaussian, they will be independent.

Bandlimited white noise

When we want to simulate white noise on digital computers, we have to consider sampled signals. Shannon sampling theorem states that we can do so if the signal is bandlimited and if we sample it at twice the highest frequency of the signal. A bandlimited white noise $n(t)$ is a noise that is white over a finite bandwidth.

$$S_n(f) = \begin{cases} \frac{N_0}{2} & |f| \le B \\ 0 & |f| > B \end{cases} \tag{2.22}$$

The autocorrelation function is no more an impulse function.

$$R_n(\tau) = N_0 B \operatorname{sinc}(2B\tau) \tag{2.23}$$

This function is represented in Fig.2- 6. We can remark that if we sample a bandlimited white noise at a frequency of 2*B*, the samples will be uncorrelated. If the process is Gaussian, these samples will be independent.

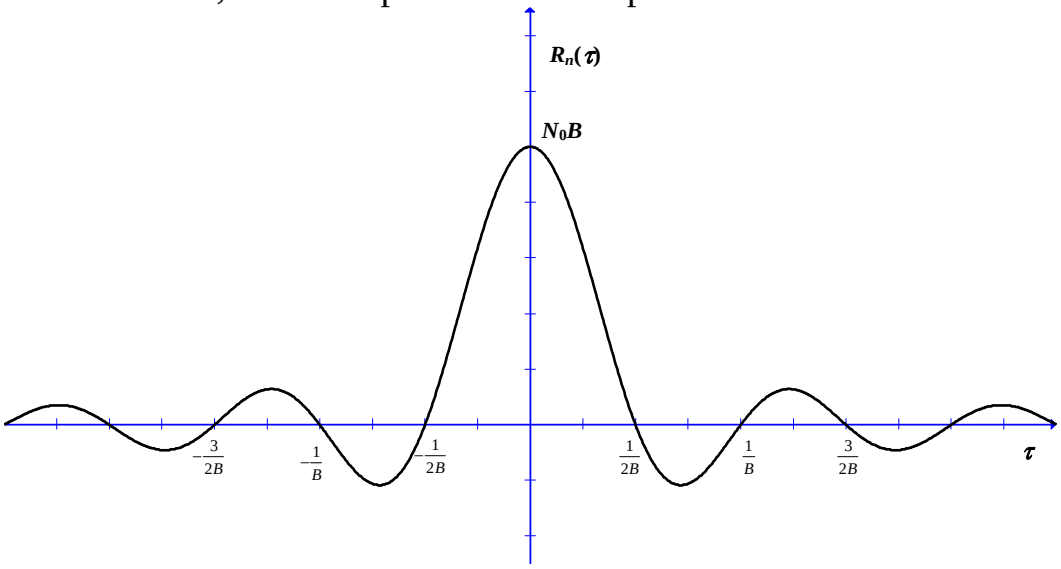


Fig.2- 6 Autocorrelation function of bandlimited white moise

The bandlimited white noise signal is zero mean and its power is $R_n(0) = N_0 B$.

2.8 Stochastic processes through linear time invariant systems

The LTI systems that we will consider are assumed to be stable. In the time domain, an LTI system is characterized by its impulse response. So, the input-output relationship is

$$y(t) = h(t) * x(t) = \int_{-\infty}^{+\infty} h(\lambda) x(t-\lambda) d\lambda \qquad (2.24)$$

The input is $x(t)$, the output is $y(t)$ and the impulse response is $h(t)$. For signals that possess Fourier transforms, the input-output relationship is:

$$Y(f) = H(f) X(f) \qquad (2.25)$$

However, we have seen that we cannot define Fourier transforms for stationary random processes. In the analysis that follows, the input process is assumed to be stationary. We can show that under the stated assumption of stability of the LTI system, the output process will also be stationary.

Mean of the output:

Applying equation (2.24) and averaging, we obtain:

$$m_y = E\left[\int_{-\infty}^{+\infty} h(\lambda) x(t-\lambda) d\lambda\right] = m_x \int_{-\infty}^{+\infty} h(\lambda) d\lambda = m_x H(0) \qquad (2.26)$$

Power spectrum of the output

If we use the alternate definition of the power spectrum (equation (2.18)), we can express the power spectrum of the output as a function of the one of the input as:

$$S_y(f) = |H(f)|^2 S_x(f) \tag{2.27}$$

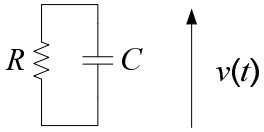
In general, the LTI system has a real impulse response. This implies that $|H(f)|^2 = H(f)H^*(f)$ is the Fourier transform of $h(t) * h(-t)$. So, the autocorrelation function of the output is related to the one of the input by the inverse Fourier transform of (2.27).

$$R_y(\tau) = h(\tau) * h(-\tau) * R_x(\tau) \tag{2.28}$$

If the input process $x(t)$ is Gaussian, the output $y(t)$ will also be Gaussian. This is due to the fact that the output is obtained by convolving the input with the impulse response. The convolution is a linear operation and we have seen that any linear combination of Gaussian processes is a Gaussian process.

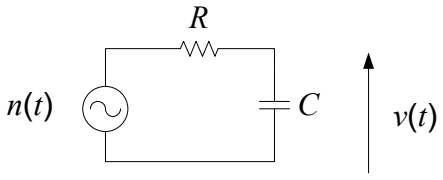
Example:

Consider the following circuit:



The noisy resistor is connected with a capacitor in parallel. We assume that the capacitor is noiseless. We want to compute the statistics of the noise signal $v(t)$.

Using the equivalence shown in Fig.2- 5, we can replace the circuit by:



This means that the white Gaussian noise $n(t)$ is filtered by the first order lowpass filter composed of the noiseless resistance and the capacitance. Its transfer function is:

$$H(f) = \frac{1}{1 + j2\pi RCf}$$

The noise $n(t)$ is white with a power spectrum $S_n(f)=\frac{N_0}{2}$. So, the power spectrum of the output process is:

$$S_v(f)=|H(f)|^2\frac{N_0}{2}=\frac{N_0}{2}\frac{1}{1+(2\pi RCf)^2}$$

The autocorrelation function is obtained by inverse Fourier transform.

$$R_v(\tau)=\frac{N_0}{4RC}\exp\left(-\frac{|\tau|}{RC}\right)=\frac{kT}{C}\exp\left(-\frac{|\tau|}{RC}\right)$$

So, $v(t)$ is a zero mean Gaussian process having the above autocorrelation. Since the dc value (mean) is zero, the total power of the signal is equal to the ac power (variance). The noise power across the circuit is:

$$\sigma_v^2=R_v(0)=\frac{kT}{C}$$

We remark that the noise power for this circuit depends on the capacitance value and it does not depend on the value of the resistance, even though the noise is produced by the resistor and not by the capacitor. Try to find an explanation.

Noise bandwidth

When a random signal is filtered, we may only be interested in the power of the signal at the output of the filter. If at the input we have a white noise with psd $N_0/2$ and the transfer function $H(f)$, the noise power at the output is:

$$N=\int_{-\infty}^{+\infty}\frac{N_0}{2}|H(f)|^2\,df=N_0\int_0^{+\infty}|H(f)|^2\,df$$

The last equality comes from the fact that the integrand is even for a real system.

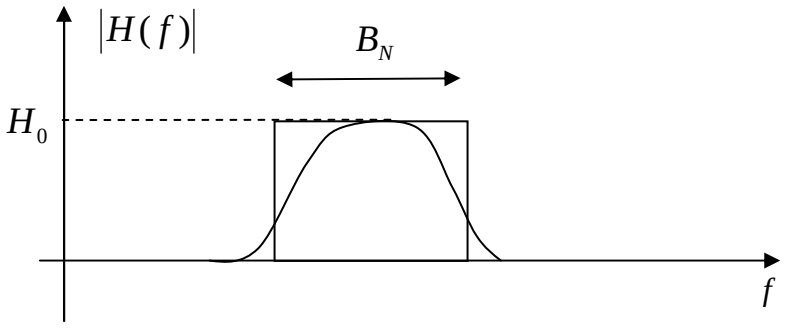


Fig.2- 7 Transfer function

Since we are not interested in the exact shape of the power spectrum of the output, but just on the power, we can compute this power using the concept of noise equivalent bandwidth.

The noise equivalent bandwidth is the bandwidth B_N of an ideal bandpass filter having an amplitude $H_0 = |H(f)|_{\max}$ and a power equal to N.

The power at the output of the ideal filter is: $N = H_0^2 N_0 B_N$. We obtain the following value for the noise equivalent bandwidth:

$$B_N = \frac{1}{H_0^2}\int_0^{+\infty} |H(f)|^2 \, df \tag{2.29}$$

If the amplitude maximum is unity, the power out of the filter is simply $N = N_0 B_N$

Example:

Consider the RC lowpass filter seen before. It transfer function is:

$H(f) = \dfrac{1}{1 + j2\pi RCf}$ and $H_0 = 1$. So, its noise bandwidth is:

$$B_N = \int_0^{+\infty} |H(f)|^2 \, df = \int_0^{+\infty} \frac{1}{1 + (2\pi RCf)^2} \, df$$

Given that $\dfrac{d}{dx}\tan^{-1} x = \dfrac{1}{1+x^2}$, we obtain:

$$B_N = \left[\frac{1}{2\pi RC} \tan^{-1}(2\pi RCf) \right]_0^{+\infty} = \frac{1}{4RC}$$

We can remark that the noise equivalent bandwidth is not very different from the 3 dB bandwidth $\dfrac{1}{2\pi RC}$. When we use high order filters, the noise equivalent bandwidth becomes practically equal to the 3 dB bandwidth. So, in general, if we don't have the shape of the filter but just its 3 dB bandwidth, we can use it for noise power computation.

2.9 Examples of autocorrelation and power spectrum calculation

Random phase sinewave

Consider the following signal:

$x(t) = A\cos(\omega_0 t + \Phi)$ where Φ is a random variable uniformly distributed over $[0, 2\pi]$.

The mean is:

$$m_x(t) = E\left[A\cos(\omega_0 t + \Phi) \right] = \int_0^{2\pi} A\cos(\omega_0 t + \phi) \times \frac{1}{2\pi} d\phi = 0$$

The autocorrelation function is:

$R_x(t, t-\tau) = E\left[A\cos(\omega_0 t + \Phi) A\cos(\omega_0 (t-\tau) + \Phi) \right]$

$$R_x(t,t-\tau)=E\left[\frac{A^2}{2}\cos(\omega_0\tau)\right]+E\left[\frac{A^2}{2}\cos(2\omega_0 t-\omega_0\tau+2\Phi)\right]$$

The second term of the above sum is

$$E\left[\frac{A^2}{2}\cos(2\omega_0 t-\omega_0\tau+2\Phi)\right]=\int_0^{2\pi}\frac{A^2}{2}\cos(2\omega_0 t-\omega_0\tau+2\phi)\times\frac{1}{2\pi}d\phi=0$$

So, finally, the autocorrelation function is:

$$R_x(t,t-\tau)=R_x(\tau)=\frac{A^2}{2}\cos\omega_0\tau$$

The mean is a constant and the autocorrelation is a function of only the lag τ and not the time *t*. The process is wide sense stationary.

The power spectrum is:

$$S_x(f)=\frac{A^2}{4}\left[\delta(f-f_0)+\delta(f+f_0)\right]$$

Random binary wave

Semi-random case

Consider the following signal:

$$z(t)=\sum_{k=-\infty}^{+\infty}a_k\,p(t-kT)$$

where

$$p(t)=\begin{cases}1 & 0\le t<T\\ 0 & \text{elsewhere}\end{cases}$$

The sequence of numbers a_k is a sequence of independent random numbers such that $a_k = +A$ or $a_k = -A$ with the same probability.

It is evident that the above signal cannot be stationary. It changes value at multiples of *T*. So, the position of the time origin is important.

$$m_z(t)=E\left[\sum_{k=-\infty}^{+\infty}a_k\,p(t-kT)\right]=\sum_{k=-\infty}^{+\infty}E[a_k]\,p(t-kT)$$

and $E[a_k]=A\times\frac{1}{2}-A\times\frac{1}{2}=0$

The autocorrelation function is $R_z(t_1,t_2)=E[z(t_1)z(t_2)]$

$z(t_1)=a_{k1}$ where a_{k1} is the amplitude in the time interval of width *T* containing t_1 and $z(t_2)=a_{k2}$ where a_{k2} is the amplitude in the time interval of width *T* containing t_2.

If t_1 and t_2 fall in different time intervals, $E[a_{k1}a_{k2}]=E[a_{k1}]E[a_{k2}]=0$ and if t_1 and t_2 fall in the same time interval $E[a_{k1}a_{k2}]=E[A^2]=A^2$.

Random case

In order to obtain a stationary signal, we must make sure that if we shift the signal by an arbitrary amount of time, we should obtain a sample function from the ensemble. The following signal is stationary:

$x(t) = z(t - T_d)$ where T_d is a random variable uniformly distributed over [0, T] and independent on the sequence a_k. The mean of the process is zero. In this example, we are going to compute the power spectrum first (we assume known the fact that the process is stationary). To do so, we use the alternate definition (2.18).

We have $x(t) = \sum_{k=-\infty}^{+\infty} a_k p(t - kT - T_d)$. Consider a finite time realization of the process $x_T(t) = \sum_{k=-n}^{+n} a_k p(t - kT - T_d)$ of duration $2nT$. The Fourier transform of $p(t - kT - T_d)$ is: $\exp\left(-j2\pi f\left(kT + \frac{T}{2} + T_d\right)\right) T \operatorname{sinc}(Tf)$

So, $X_T(f) = \sum_{k=-n}^{n} a_k \exp\left(-j2\pi f\left(kT + \frac{T}{2} + T_d\right)\right) T \operatorname{sinc}(Tf)$

The energy spectral density is:

$$X_T(f) X_T^*(f) = \sum_{k=-n}^{n} \sum_{i=-n}^{n} a_k a_i \exp\left(-j2\pi f (k-i) T\right) T^2 \operatorname{sinc}^2(Tf)$$

The average energy is then:

$$E\left[X_T(f) X_T^*(f)\right] = \sum_{k=-n}^{n} \sum_{i=-n}^{n} E\left[a_k a_i\right] \exp\left(-j2\pi f (k-i) T\right) T^2 \operatorname{sinc}^2(Tf)$$

$$E\left[a_k a_j\right] = \begin{cases} A^2 & k = i \\ 0 & k \neq i \end{cases}$$

So, $S_x(f) = \lim_{n\to\infty} \frac{1}{2nT} X_T(f) X_T^*(f) = \lim_{n\to\infty} \frac{(2n+1)}{2nT} A^2 T^2 \operatorname{sinc}(Tf)$

The power spectrum is then:

$$S_x(f) = A^2 T \operatorname{sinc}^2(fT)$$

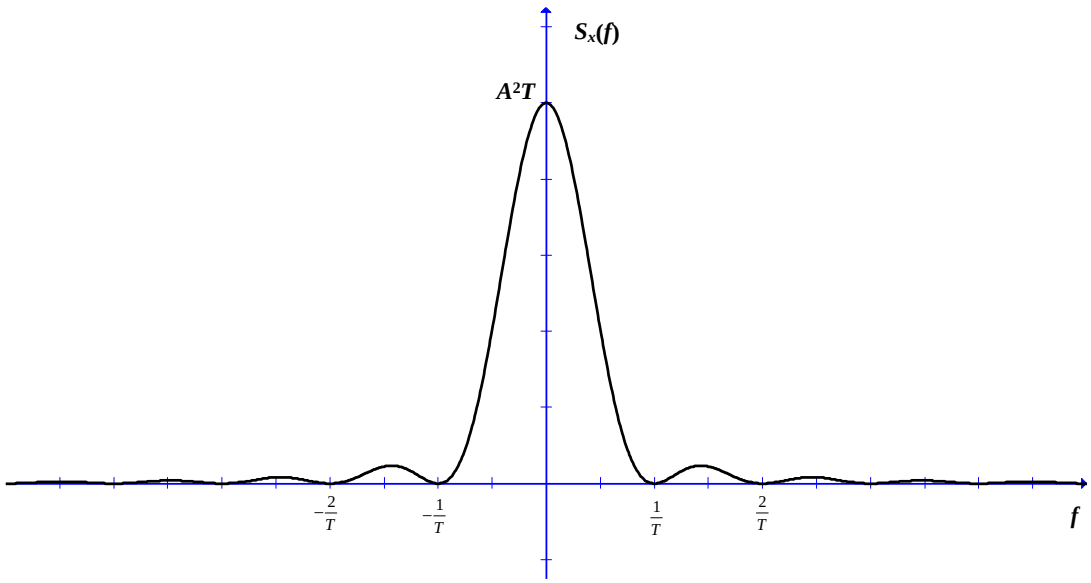


Fig.2- 8 Power spectrum of the random binary wave

The autocorrelation is the inverse Fourier transform of the above power spectrum.

$$R_x(\tau)=\begin{cases} A^2\left(1-\dfrac{|\tau|}{T}\right) & |\tau|\le T \\ 0 & |\tau|>T \end{cases}$$

Modulation

Consider the signal $x(t)=Az(t)\cos\left(\omega_0 t+\Phi\right)$ where $z(t)$ is a stationary random process with autocorrelation $R_z(\tau)$ and Φ is a random variable uniformly distributed on $[0, 2\pi]$ and independent on $z(t)$.

The mean is:

$$m_x(t)=E\left[Az(t)\cos\left(\omega_0 t+\Phi\right)\right]=E\left[z(t)\right]\int_0^{2\pi} A\cos\left(\omega_0 t+\phi\right)\times\frac{1}{2\pi}d\phi=0$$

The autocorrelation function is:

$$R_x(t,t-\tau)=E\left[A^2 z(t)z(t-\tau)\cos\left(\omega_0 t+\Phi\right)\cos\left(\omega_0\left(t-\tau\right)+\Phi\right)\right]$$

Repeating the same calculations as in the random phase sinewave, we obtain:

$$R_x(t,t-\tau)=R_x(\tau)=\frac{A^2}{2}E\left[z(t)z(t-\tau)\right]\cos\omega_0\tau=\frac{A^2}{2}R_z(\tau)\cos\omega_0\tau$$

The power spectrum is:

$$S_x(f)=\frac{A^2}{4}\left[S_z(f-f_0)+S_z(f+f_0)\right]$$

We find that the signal is also wide sense stationary and that the spectrum is translated the same way as for deterministic signals.

2.10 Narrowband noise

Using the concept of power spectrum, we can define lowpass and bandpass random signals. In this section, we are going to consider bandpass signals. We have already encountered bandpass deterministic signals in a previous course. Consider a random process with a bandpass power spectral density.

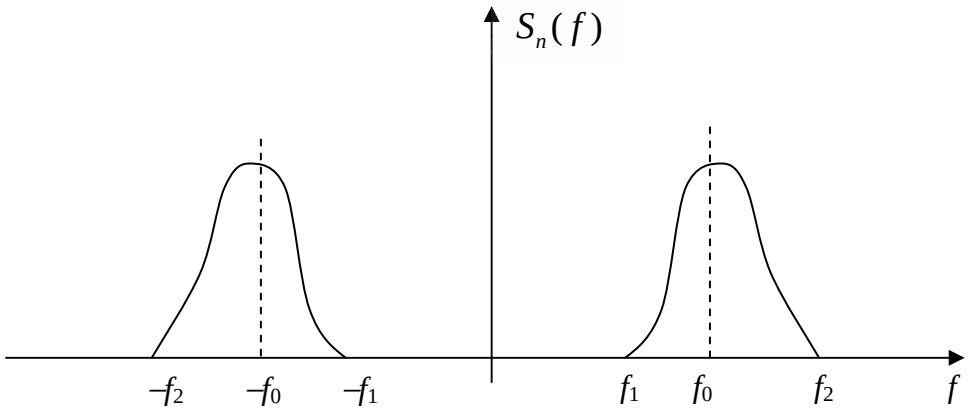


Fig.2- 9 Bandpass psd

The signal $n(t)$ having the power spectrum shown in Fig.2- 9 has all its power concentrated in the band between f_1 and f_2. This signal can be represented in quadrature form, just like the deterministic signal. Consider a frequency f_0 located between f_1 and f_2.The signal $n(t)$ can be written as:

$$n(t) = n_I(t)\cos\omega_0 t - n_Q(t)\sin\omega_0 t \qquad (2.30)$$

The above equality should be understood in the mean square sense. The signal $n_I(t)$ is called the "in-phase component" and the signal $n_Q(t)$ is called the "quadrature component". These two signals are baseband signals.

Let $B = f_2 - f_1$. If $f_0 >> B$, then the signal is called narrowband signal.

From the fact that the power spectrum does not contain an impulse at zero frequency, we can conclude that the mean of the signal $n(t)$ is zero.

Properties of the signals:

1. $E[n(t)] = E[n_I(t)] = E[n_Q(t)] = 0$
2. If $n(t)$ is Gaussian, the two components $n_I(t)$ and $n_Q(t)$ are also Gaussian.
3. The power spectra of the two components are equal and they are given by: $S_{n_I}(f) = S_{n_Q}(f) = Lp[S_n(f - f_0) + S_n(f + f_0)]$. The notation $Lp[\,.\,]$ stands for the lowpass part, i.e. the part of the spectrum around zero and the part shifted around $2f_0$ is eliminated.

4. From the above relation, we can deduce that $\sigma_n^2 = \sigma_{n_I}^2 = \sigma_{n_Q}^2$, i.e. the components have the same power as the signal $n(t)$.
5. The cross power spectrum of the two components is given by: $S_{n_I n_Q}(f) = jLp[S(f + f_0) - S(f - f_0)]$

If the original signal $n(t)$ has a power spectrum that has an axis of symmetry for its positive half around f_0, then the two quadrature components will be uncorrelated. If the signal is narrowband, the spectrum will have this symmetry and we can safely assume that the two components are uncorrelated. If the signal is Gaussian, the two components will be independent.

We can also have an envelop and phase representation of the process.

$$n(t) = r(t)\cos(\omega_0 t + \theta(t))$$

If $n(t)$ is narrowband and Gaussian, we have seen (Rayleigh fading) that the envelop and the phase will be respectively Rayleigh and uniform.

If we filter a sinewave corrupted with an additive Gaussian noise with a narrow bandpass filter, the output of the filter will be a sinewave added to narrowband noise. In this case, the envelop of the output will be Rician.