STOCHASTIC PROCESSES

Dr Ahmad Khonsari

ECE Dept.

The University of Tehran

STOCHASTIC PROCESSES

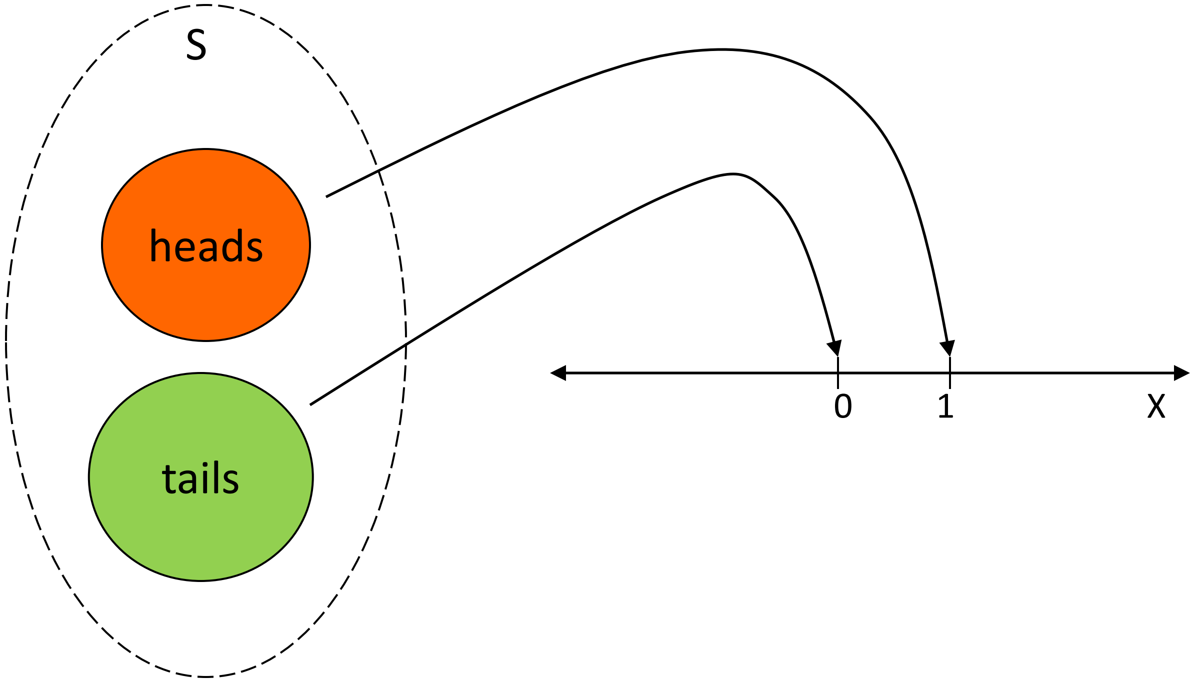
Basic concepts

Often the systems we consider evolve in time and we are interested in their dynamic behavior, usually involving some randomness.

- the length of a queue
- the number of students passing the course “computer networks” each year
- the temperature outside
- the number of data packets in a network

Random Variables

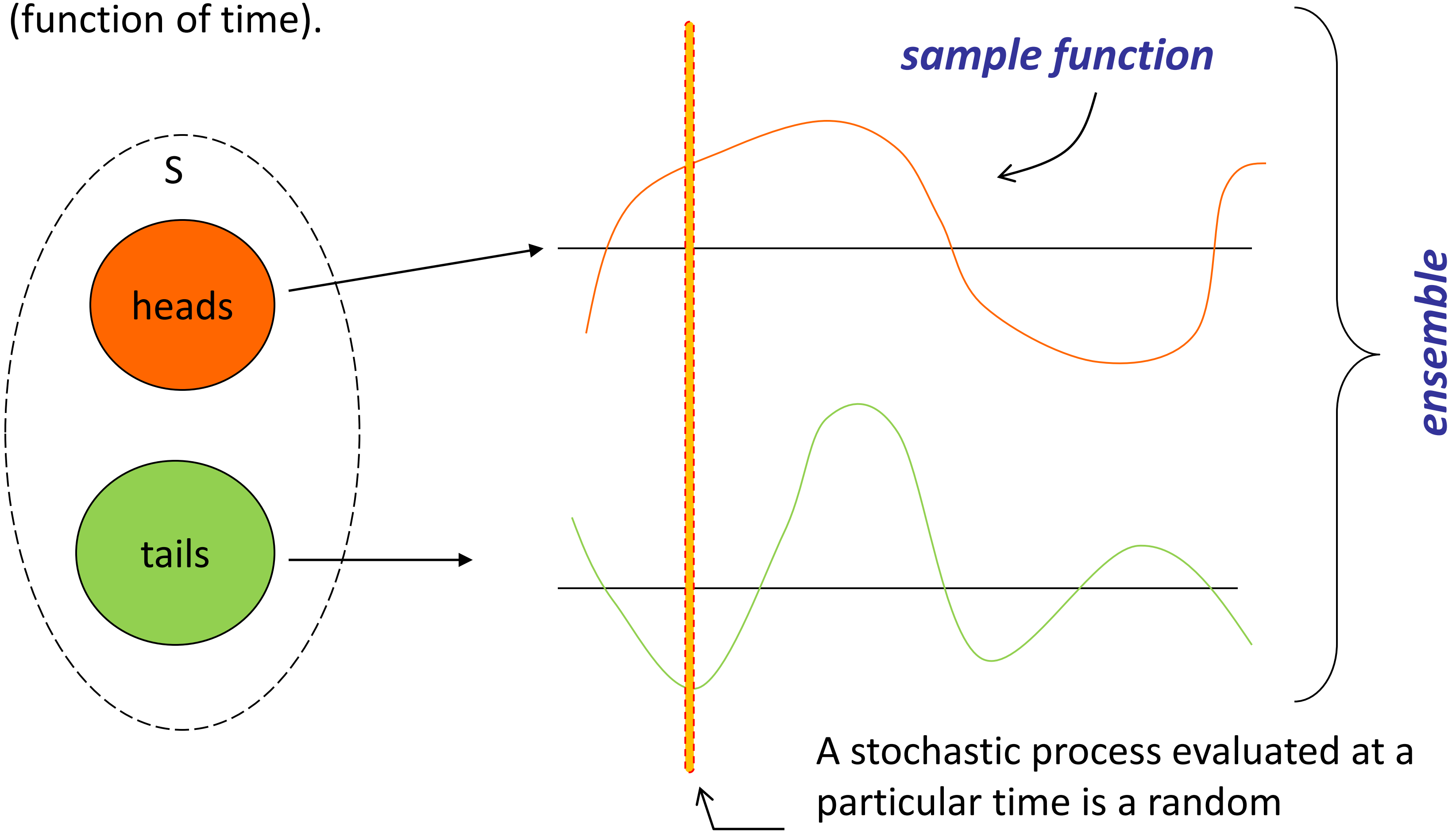
- Random variables map the outcome of a random experiment to a ***number***.



Stochastic (Random) Processes (movement)

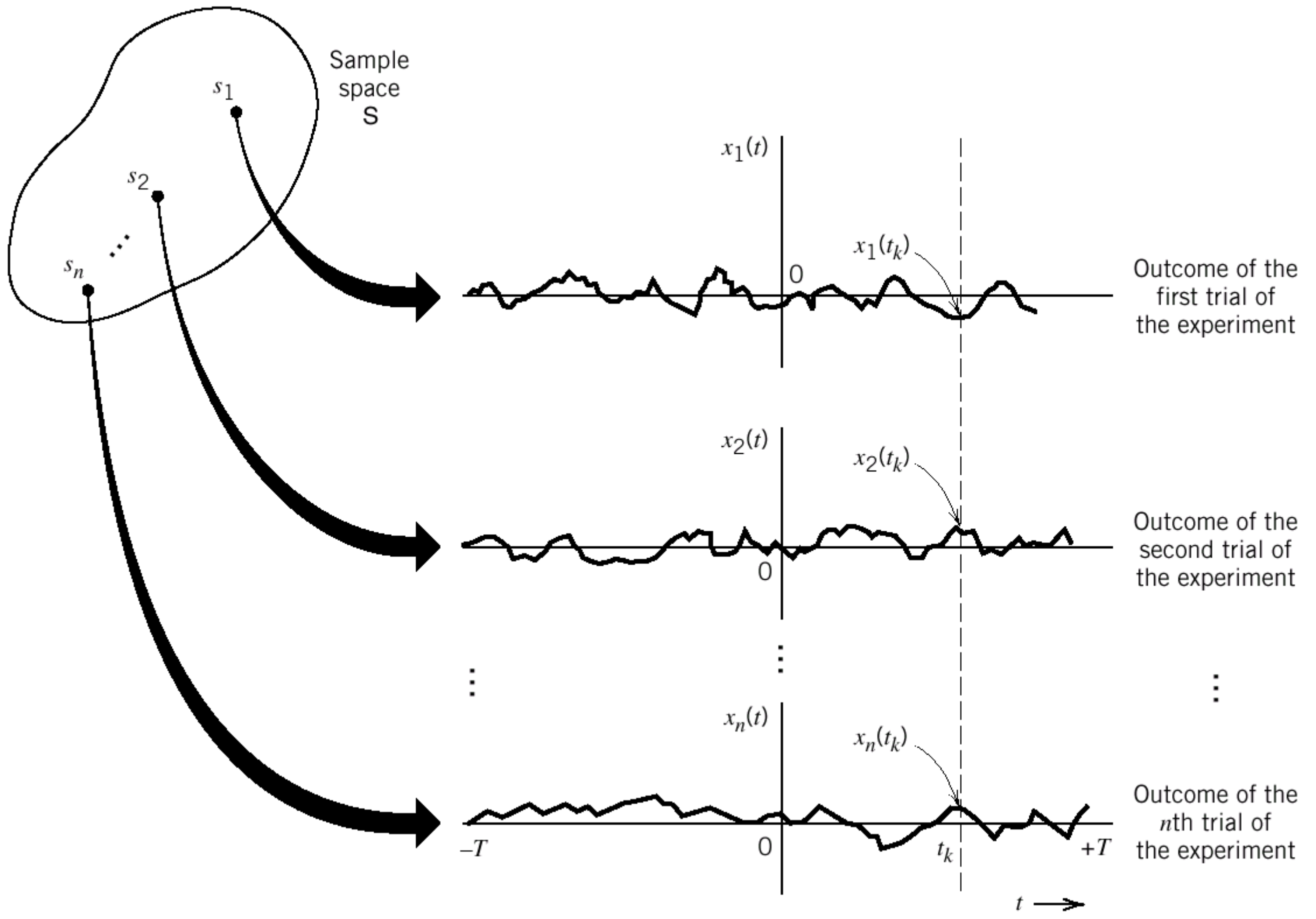
Stochastic Processes map the outcome of a random experiment to a ***signal*** (function of time).

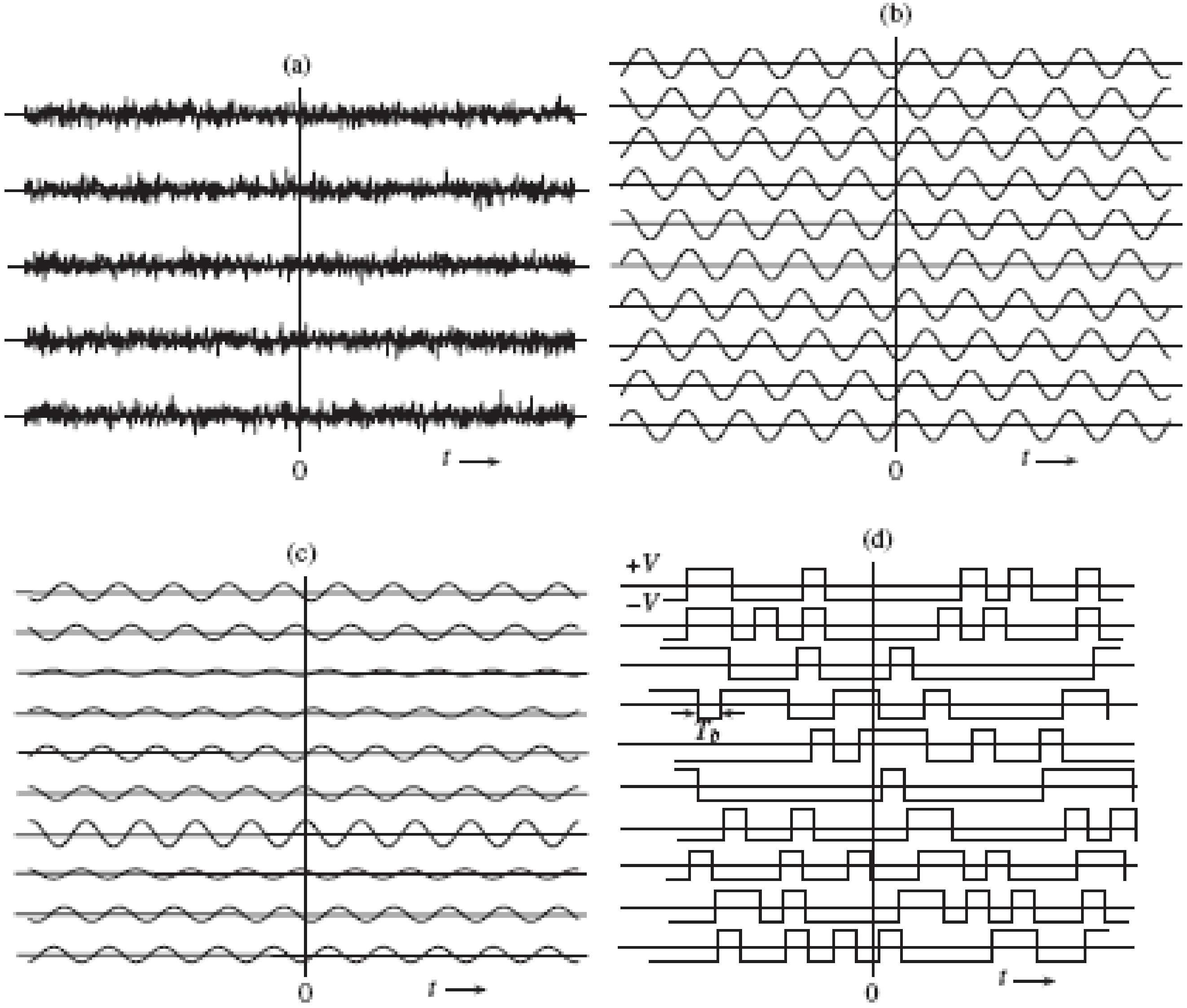
signal associated with the outcome:



A stochastic process evaluated at a particular time is a random variable

stochastic processes

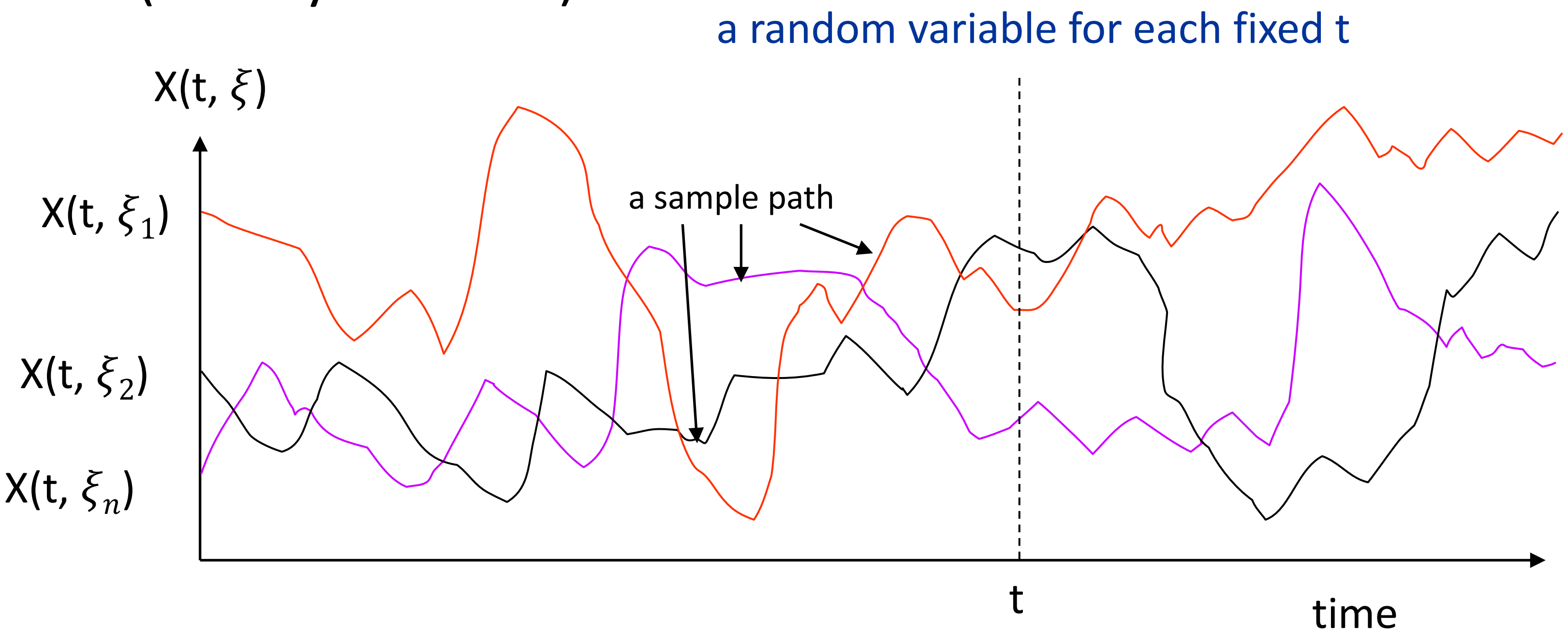




Typical ensemble members for four random processes commonly encountered in communications (a) thermal noise (b) uniform phase. (c) Rayleigh fading process and (d) binary random data process.

STOCHASTIC PROCESSES

A stochastic process $X_t(\xi)$(or $X(t, \xi)$) is a family of random variables indexed by a parameter t (usually the time).



STOCHASTIC PROCESSES

Basic notions

Formally, a stochastic process is a mapping from the sample space S to functions of t. With each element ξ of S is associated a function $X_t(\xi)$.

("a lottery ticket e with a plot of a function is drawn from an urn")

- For a given value of ξ, $X_t(\xi)$ is a function of time
  (this is a sample path)
  (ensemble : $X_t(\xi)$ for all outcomes of sample space ξ)
- For a given value of t, $X_t(\xi)$ is a random variable
  ($\{X_t(\xi)\}$ for different t): this view is a collection of RVs and thus RP is defined by all joint CDFs for all possible set of times.
- For a given value of ξ and t, $X_t(\xi)$ is a (fixed) number

The function $X_t(\xi)$ associated with a given value ξ is called the realization of the stochastic process (also trajectory or sample path).

Classification of Stochastic Processes

State space: the set of possible values of X_t

Parameter (e.g. time) space: the set of values of t

Stochastic processes can be classified according to whether these spaces are discrete or continuous:

		time, Index set T	
		Discrete	*Continuous*
State/Sample Space I	*Discrete*	Discrete-time stochastic chain	Continuous-time stochastic chain
	Continuous	Discrete-time continuous-state process	Continuous-time continuous-state process

- discrete-state process → chain
- discrete-parameter (e.g. discrete-time) process → stochastic sequence $\{X_n \mid n \in T\}$ (e.g., probing a system every 10 ms.)

STOCHASTIC PROCESSES

Example 1. A random process is of the form of sinusoid $x(n) = A\cos(n\omega_0)$ where $A \in \Omega = \{ 1,2,.., 6 \}$, the amplitude is a random variable that assumes any integer number between one and six, each with equal probability $\Pr(A=k)=1/6$ (k=1, 2, …, 6).

This random process consists of an ensemble of six different discrete-time signals $x_k(n)$,

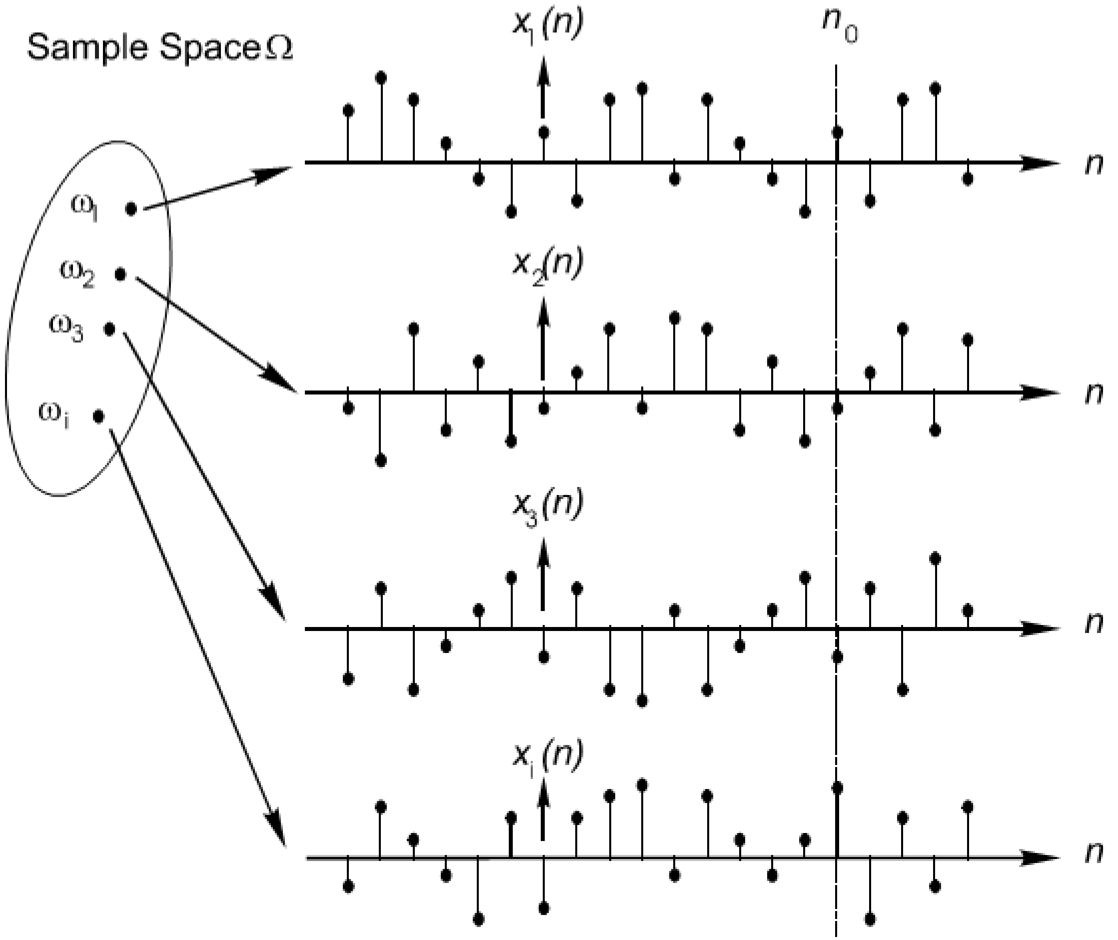
$x_1(n) = \cos(n\omega_0)$, $x_2(n) = 2\cos(n\omega_0)$, … $x_6(n) = 6\cos(n\omega_0)$,
each of which shows up with equal probability.

STOCHASTIC PROCESSES

Question: Given a random process $x(n) = A(n)\cos(n\omega_0)$, where the amplitude $A(n)$ is a random variable (at instant n) that assumes any integer number between one and six, each with equal probability, how many equally probable discrete-time signals are there in the ensemble?

SP-view 1 –collection of waveforms- Analytic description

A random process as in the figure has an ensemble of different discrete-time signals, each occurring according to a certain probability. From a sample space point of view, to each experimental outcome ω_i in the sample space, there is a corresponding discrete-time signal $x_i(n)$ (i.e. sample path).



SP-view 2 –Sequence of RVs- Statistical description

at a certain 'fixed' time instant *n*, e.g., $n = n_0$, the signal value $x(n_0)$ is a RV that is defined on the sample space and has an underlying CDF and pdf

$$F_{x(n_0)}(\alpha) = \Pr\{x(n_0) \leq \alpha\},$$

And

$$f_{x(n_0)}(\alpha) = df_{x(n_0)}(\alpha)/d\alpha$$

for a different n_0 , $x(n_0)$ is a random variable at a different time instant. Therefore, a *discrete-time random process* is an indexed sequence of random variables $x(n)$ that is an ensemble of elementary events $x_i(n)$ at *n*.

at a certain 'fixed' time instant *n*, e.g., $n = n_0$, the signal value $x(n_0)$ is a RV that is defined on the sample space and has an underlying CDF and pdf

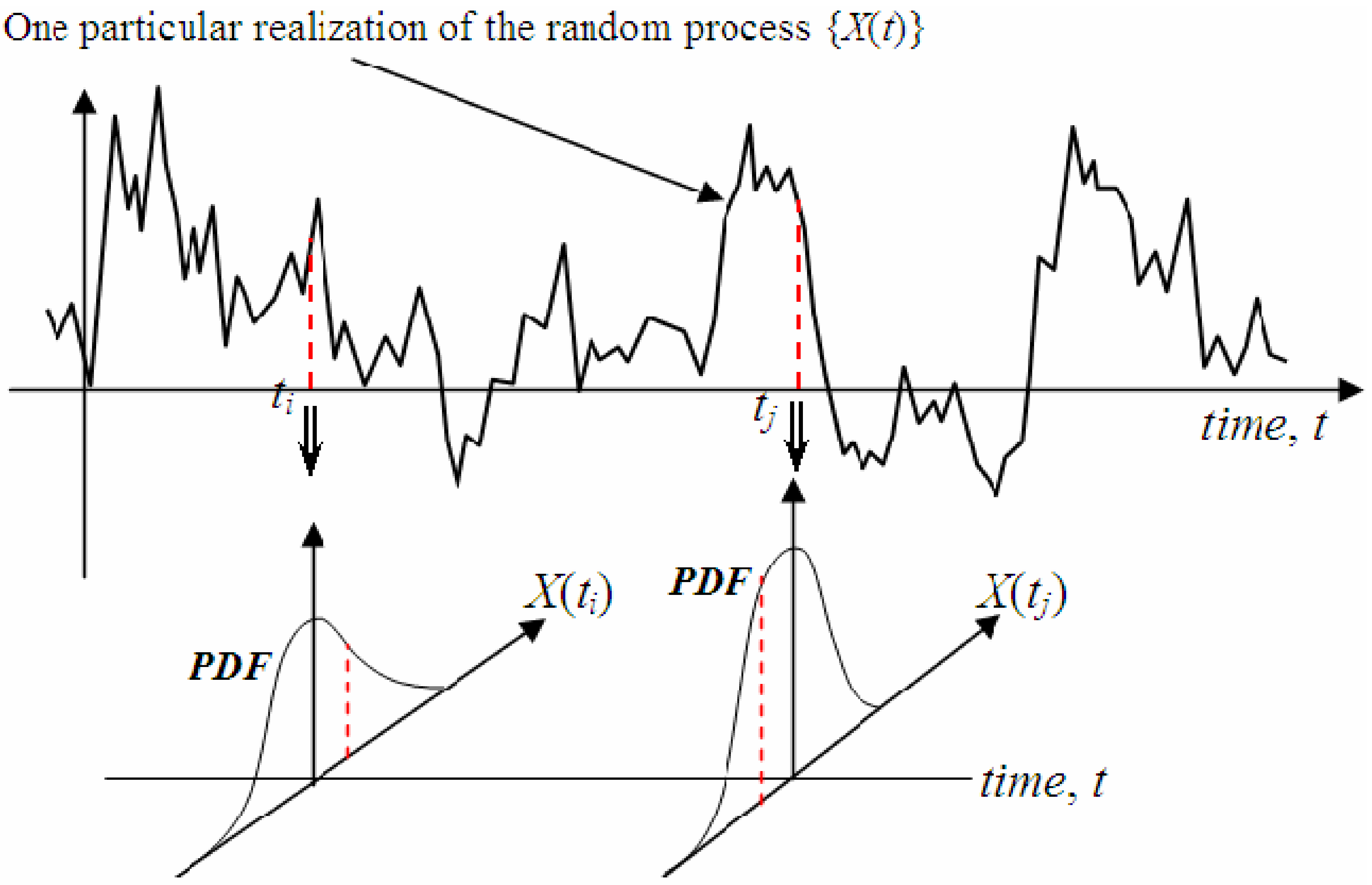
$$F_{x(n_0)}(\alpha) = \Pr\{x(n_0) \leq \alpha\},$$

And

$$f_{x(n_0)}(\alpha) = df_{x(n_0)}(\alpha)/d\alpha$$

for a different n_0 , $x(n_0)$ is a random variable at a different time instant. Therefore, a *discrete-time random process* is an indexed sequence of random variables $x(n)$ that is an ensemble of elementary events $x_i(n)$ at *n*.

STOCHASTIC PROC ESS



In considering stochastic processes we are often interested in quantities like:

- Time-dependent distribution: defines the probability that X_t takes a value in a particular subset of S at a given instant t
- Stationary distribution: Stationarity refers to time invariance of some, or all, of the statistics of a random process, such as mean, autocorrelation, n-th-order distribution (We define two types of stationarity: strict sense (SSS) and wide sense (WSS))
- The relationships between X_s and X_t for different times s and t (e.g. covariance or correlation of X_s and X_t)
- Hitting probability: the probability that a given state is S will ever be entered
- First passage time: the instant at which the stochastic process first time enters a given state or set of states starting from a given initial state

Description of Random Processes

(Stochastic Process Characterization)

Stochastic Process Characterization

- ***two*** types of descriptions:
- ***analytic expressions***
- If the random process can be viewed as a collection of signals,

In this description, analytic expressions are given for each sample in terms of one or more random variables; i.e., the random process is given as $X(t) = f(t; \boldsymbol{\omega})$ where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$ is, in general, a random vector with a given joint PDF. This is a very informative description of a random process because it completely describes the analytic form of various realizations of the process.

- ***statistical description***
  For real-life processes, it is hardly possible to give such a complete description.

If an analytic description is not possible, a *statistical description* may be appropriate. Such a description is based on the second viewpoint of random processes, regarding them as a collection of random variables indexed by some index set.

Stochastic Process Characterization- statistical description

- **Definition 6.1.** *A complete statistical description* of a random process $X(t)$ is known if for any integer n and any choice of $(t_1, t_2, . . . , t_n) \in R^n$ the joint PDF of $(X(t_1), X(t_2), . . . , X(t_n))$ is given.
  - If the complete statistical description of the process is given, for any n the joint density function of $(X(t_1), X(t_2), . . . , X(t_n))$ is given by

$$F(\boldsymbol{x};\boldsymbol{t}) = F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) = P\{X(t_1) \le x_1, \ldots, X(t_n) \le x_n\}$$

- **Definition 6.2.** A process $X(t)$ is described by its Mth *order statistics* if for all $n \le M$ and all $(t_1, t_2, . . . , t_n) \in R^n$ the joint PDF of $(X(t_1), X(t_2), . . . , X(t_n))$ is given.
  - important special case:
    $M = 2$, in which second-order statistics are known. This simply means that, at each time instant t, we have the density function of $X(t)$, and for all choices of (t_1, t_2) the joint density function of $(X(t_1), X(t_2))$ is given.

Stochastic Process Characterization- full description-n-th order statistics

- The n^{th} order statistics of a stochastic process X_t is defined by the joint distribution

$$F(\boldsymbol{x};\boldsymbol{t}) = F_{X(t_1),\dots,X(t_n)}(x_1,\dots,x_n) = P\{X(t_1) \le x_1,\dots,X(t_n) \le x_n\}$$

where,

$$\boldsymbol{x} = (x_1,\dots,x_n) \epsilon \mathcal{R}^n$$

and

$$\boldsymbol{t} = (t_1,\dots,t_n) \epsilon \mathcal{T}^n$$

A complete characterization of a stochastic process X_t requires knowing the stochastics of the process of all orders n.

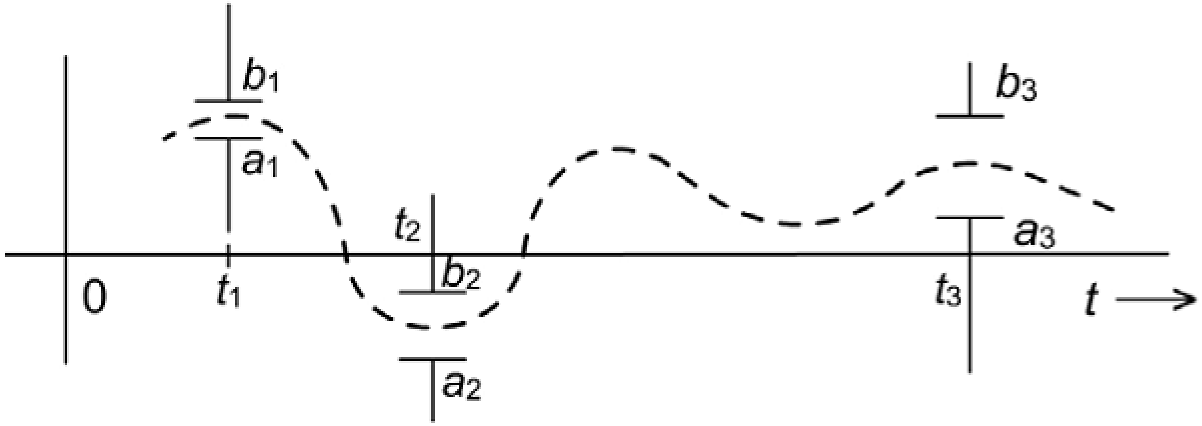
Stochastic Process Characterization- full description-n-th order statistics

Once $f_{X(t)}(x)$ is known, we might compute the probability of various events. E.g., we might calculate the probability of the random process $X(t)$ passing through a set of windows as shown in the Fig.

Let A be the event:
$A = \{s : a_1 < X(t_1) \leq b_1, a_2 < X(t_2) \leq b_2, a_3 < X(t_3) \leq b_3, \}$

That is, the event A consists of all those sample points $\{s_j\}$ such that the corresponding sample functions $\{x_j(t)\}$ satisfy the requirement, $a_i < X_j(t_i) \leq b_i$, for i=1, 2, 3 . We need to calculate $P(A)$.

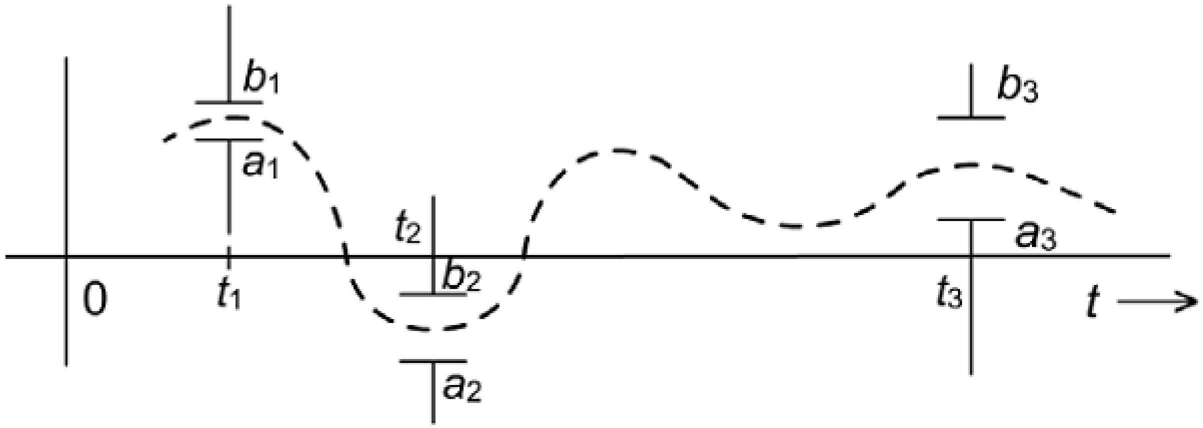
A typical sample function which would contribute to $P(A)$ is shown in the same figure.



Stochastic Process Characterization- full description-n-th order statistics

P (*A*) can be calculated as

$$P(A) = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} f_{\boldsymbol{X}(\boldsymbol{t})}(\boldsymbol{x})\, \boldsymbol{d}\,\boldsymbol{x}$$



where $\boldsymbol{x}\ (\boldsymbol{t}\ ) = X(t_1)\ ,X(t_2), X(t_3)$,
and $\boldsymbol{d}\ \boldsymbol{x} = d\ x_1\ d\ x_2\ d\ x_3$

The above step can be generalized to the case of a random vector with n-components.
This complete description of RP is virtually impossible to use for practical applications!
Usually make do with **1st and 2nd order PDF's:**

Stochastic Process Characterization- full description-1st order statistics

1st order statistics: $F(x_1;t_1) = F_{X(t_1)}(x_1) = P\{X(t_1) \le x_1\}$

1st order statistics can be characterized (but not necessarily completely) by two parameters mean and variance

- The ***expected value***, ***ensemble average*** or ***mean*** of a RP at time t is:

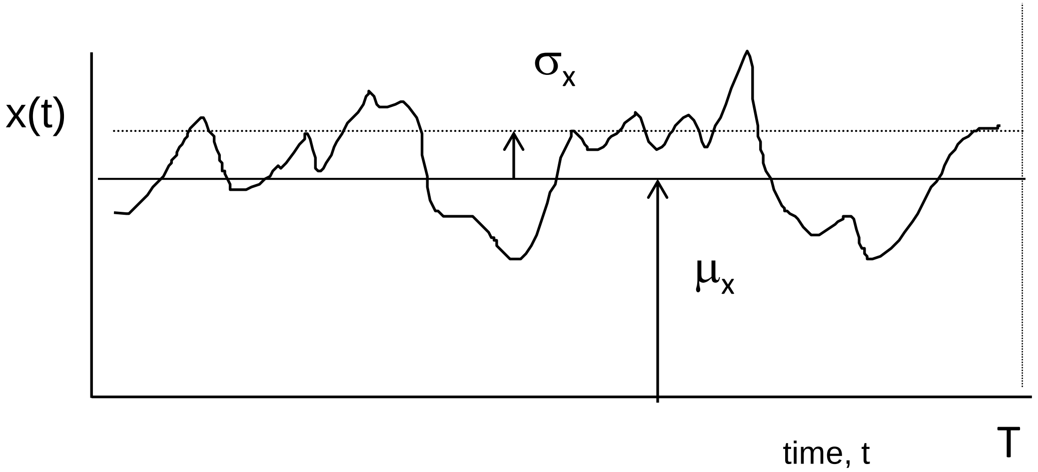
$$\overline{X}_t = E[X_t] = m_x(t) = \int_{-\infty}^{\infty} x_t p(x_t)\mathrm{dx}$$

Mean: shows "center of concentration" of possible values of x(t) as a Function of time in general

- ***Variance*** of X(t)

$$\sigma_{x(t)}^2 = E[(x(t) - m_x(t))^2] = \int_{-\infty}^{\infty} (x(t) - m_x(t))^2 p(x_t)\mathrm{dx}$$

Shows a measure of expected range of variation around the mean, as a function of time in general



Q: What does the 1st order PDF's tell us ?

Ans: 1st order $F(x_1;t_1)$ tells, as a function of time, what values are likely and unlikely to occur (tells the likelihood of values occurring at each given time)

Stochastic Process Characterization- full description-2nd order statistics

2nd order statistics: $F(x_1, x_2; t_1, t_1) = F_{X(t_1),X(t_2)}(x_1, x_2) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2\}$

we want sth to capture most of the essence of the 2nd order PDF

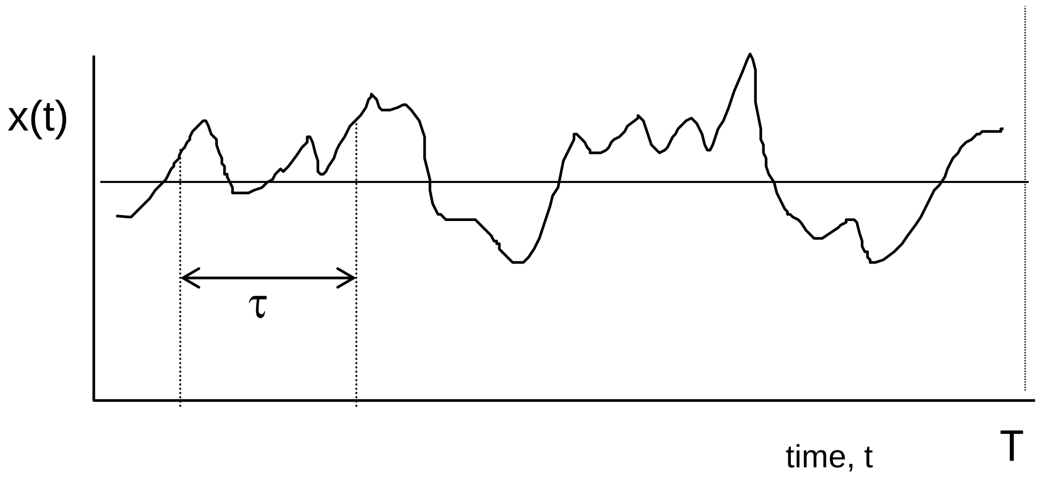
- Covariance (auto covariance) $C_X(t_1, t_2) = E[\{X_{t_1} - \overline{X}_{t_1}\}\{X_{t_2} - \overline{X}_{t_2}\}] = R_X(t_1, t_2) - m_X(t_1)m_X^*(t_2)$

- The ***autocorrelation*** function (ACF) is:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_x(x_1, x_2; t_1, t_2) dx_1 dx_2$$

Autocorrelation is a measure of how alike the random process is from one time instant to another.

Q: What does the 2nd order PDF's tell us ?
Ans 2nd Order PDF Characterizes "Probabilistic Coupling" between RP values at each pair of times t_1 and t_2
Example : What is the probability $x(t_1)$ and $x(t_2)$ are…
…. both Positive?
…. both Negative?
…. of Opposite Signs



Characterizing Stochastic processes: Highlight

Specifying RPs in terms of n-th order statistics : e.g.:

Gaussian Random process

Sinusoid with random phase

IID processes

Specifying RPs in terms of 1st order statistics:

Specifying full desc. Of RPs using 2nd order statistics :

Markov processes

Specifying RPs in terms of moment 1 and moment 2:

e.g. Poisson ((see Papoulis)): **note** : for Poisson we may find full desc. As well

1st order statistics: denote $n(0,t)$ with $n(t)$

mean $E[X(t)] = E[n(0,t)] = E[n(t)] = \lambda t$

and variance $\sigma_x^2 = E[x^2(t)] - E[X(t)]^2 = \lambda t$ since $E[x^2(t)] = E[n^2(t)] = \lambda t + \lambda^2 t^2$

2nd order statistics $R_X(t_1,t_2) = E[n(t_1)n(t_2)] = E[n(t_1)\{n(t_2) - n(t_1) + n(t_1)\}] = E[n^2(t_1)] + E[n(t_1)]E[n(t_2) - n(t_1)]$

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

$= \lambda \min\{t_1, t_2\} + \lambda^2 t_1 t_2$ in general

Specifying whether a RP process is stationary (strict sense or WSS)

Specifying whether a RP is ergodic (mean-ergodic , egodic in correlation(covariance) , distribution ergodic ,..., Papoulis ch. 12)

Mean and Autocorrelation

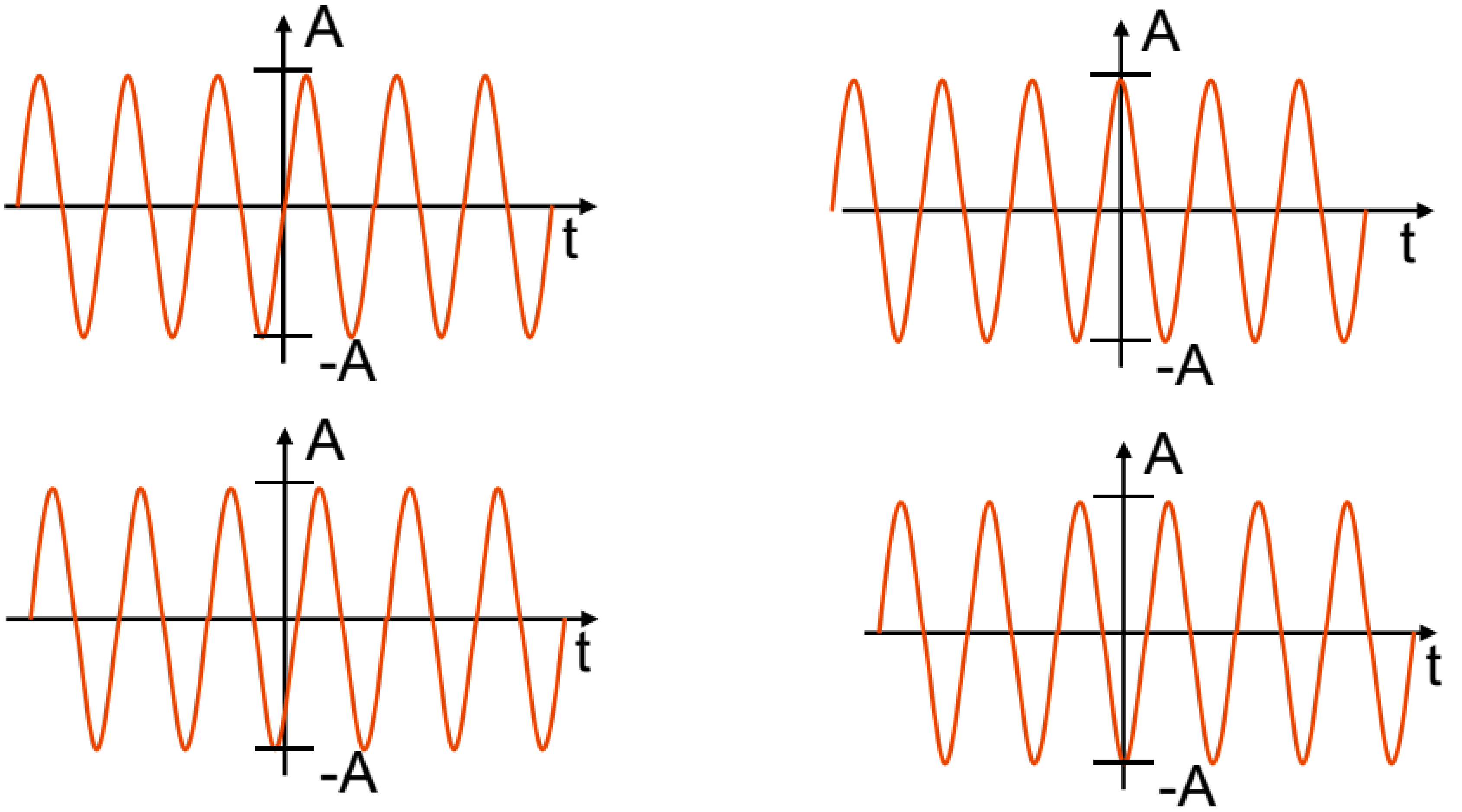
- Finding the mean and autocorrelation is not as hard as it might appear! (although higher moments are not easy to derive)
  - Since many times a stochastic process can be expressed as a *function* of a random variable (you know functions of random variables).

- Example2: consider $x(t)=A\cos(2\pi t+\theta)$, where θ is a RV according to a given distribution and thus x(t) is just a function $g(\theta)$ of θ:

  - **Q: Which of our two "view points" is easier to think of for this example?**
    1. Sequence of RVs???
    2. Collection of Waveforms & Pick One???
  - **A :** Clearly it is easier to view this random process as a collection of waveforms from which you randomly pick one
    Remember!!! – Both Views are Still Correct

Mean and Autocorrelation-example 2

- Here are **4 realizations (sample functions)** of the ensemble of this process



Each one has a different Phase

Which signal you get ? The signal is randomly chosen according to the PDF of Phase

Mean and Autocorrelation-example 2

- Looking at any one sample function doesn't give the appearance of being a random process !
- **BUT IT IS RANDOM! You don't know ahead of time which you were going to get.**
- **In this case, randomness is best viewed as "not knowing which of the infinite possible sample functions you will get"**
- So… now we have a model for a practical signal scenario. Now What???
  Do analysis to characterize the model !!!!

Mean and Autocorrelation-example 2

- Example2: consider $x(t)=A\cos(2\pi t+ \theta)$, where θ is a RV according to a given distribution and thus x(t) is just a function $g(\theta)$ of θ:

- To characterize this process:
- **Task** : Find the **mean** and the **ACF** of this process & Ask: **Is It WSS?**

- We need to find the expected value of a function of a random variable: i.e. $g(\theta)=A\cos(2\pi t+ \theta)$
- To do so you need to know the pdf of θ.

$$E[x(t)]=E[g(\theta)]=E[A\cos(2\pi t+ \theta)]$$

Example2 (Papoulis P.377)

x(t)=Acos(2πt+ θ), θ~U(-π, π):

$m_x(t)=$

$E[A\cos(2\pi t+\theta)]=A\int_{-\infty}^{\infty}\cos(2\pi t+\theta)p_{\theta}(\theta)d\theta$

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

$=A\int_{-\pi}^{\pi}\{\cos(2\pi t)\cos(\theta)-\sin(2\pi t)\sin(\theta)\}\frac{1}{2\pi}d\theta$

$=\frac{A}{2\pi}\cos(2\pi t)\ \int_{-\pi}^{\pi}\cos(\theta)\ d\theta$

$+\frac{A}{2\pi}\sin(2\pi t)\int_{-\pi}^{\pi}\sin(\theta)d\theta=0$

Example2 (Papoulis P.377)

$x(t)=A\cos(2\pi t+\theta)$, $\theta \sim U(-\pi, \pi)$:

$R(t_1,t_2)= E[A\cos(2\pi t_1+\theta)A\cos(2\pi t_2+\theta)]$

$= A^2 E[1/2\{\cos(2\pi(t_1+t_2)+2\theta) + \cos(2\pi(t_1-t_2))\}$

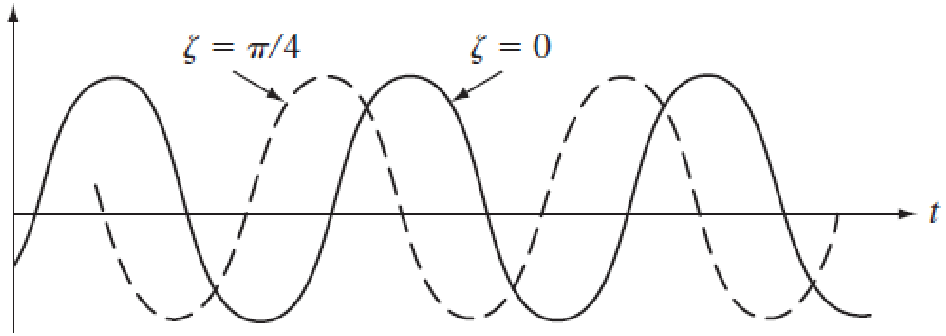
$=A^2 \int_{-\pi}^{\pi} 1/2\{\cos(2\pi(t_1+t_2)+2\theta) + \cos(2\pi(t_1-t_2)\} p_\theta(\theta)d\theta$

$=A^2 \int_{-\pi}^{\pi} 1/2\cos(2\pi(t_1+t_2)+2\theta)\frac{1}{2\pi}d\theta + A^2 \int_{-\pi}^{\pi} 1/2\cos 2\pi(t_1-t_2)\frac{1}{2\pi}d\theta$

$=0 + \frac{A^2}{4\pi}\cos 2\pi(t_1-t_2) \int_{-\pi}^{\pi} d\theta$

$=\frac{A^2}{2}\cos 2\pi(t_1-t_2)$

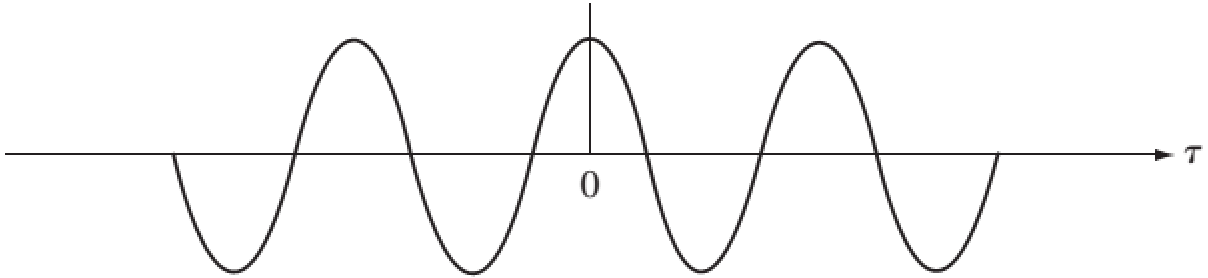
$R(\tau) = \frac{A^2}{2}\cos 2\pi(\tau)$ where $\tau = (t_1-t_2)$



$\cos(A + B) = \cos A \cos B - \sin A \sin B$

$\cos(a)\cos(b)= ½\{\cos(a + b) +\cos(a - b)\}$

$R(\tau)=\frac{A^2}{2}\cos(2\pi\tau)$



Autocorrelation function of a sinusoid with random phase

Note1 we didn't use joint pdf for $R(t_1,t_2)$, since we could separate the cosine multiplication

Note2 we see later that: x(t) is WSS

Note3 the samples at time t_1 and t_2 are uncorrelated, not proved here

Stationary Processes

Stationary Processes

- In a complete statistical description of a random processes, for any *n*, and any$(t_1, t_2, \ldots, t_n)$, the joint PDF
$F(\boldsymbol{x};\boldsymbol{t}) = F_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) = P\{X(t_1) \le x_1,\ldots,X(t_n) \le x_n\}$
is given.
- This joint PDF in general depends on the choice of the time origin.
- In a very important class of random processes, the joint density function is independent of the choice of the time origin. These processes whose statistical properties are time independent are called *stationary processes*.

A Process is (strictly) stationary if, for *all* orders of *n*,

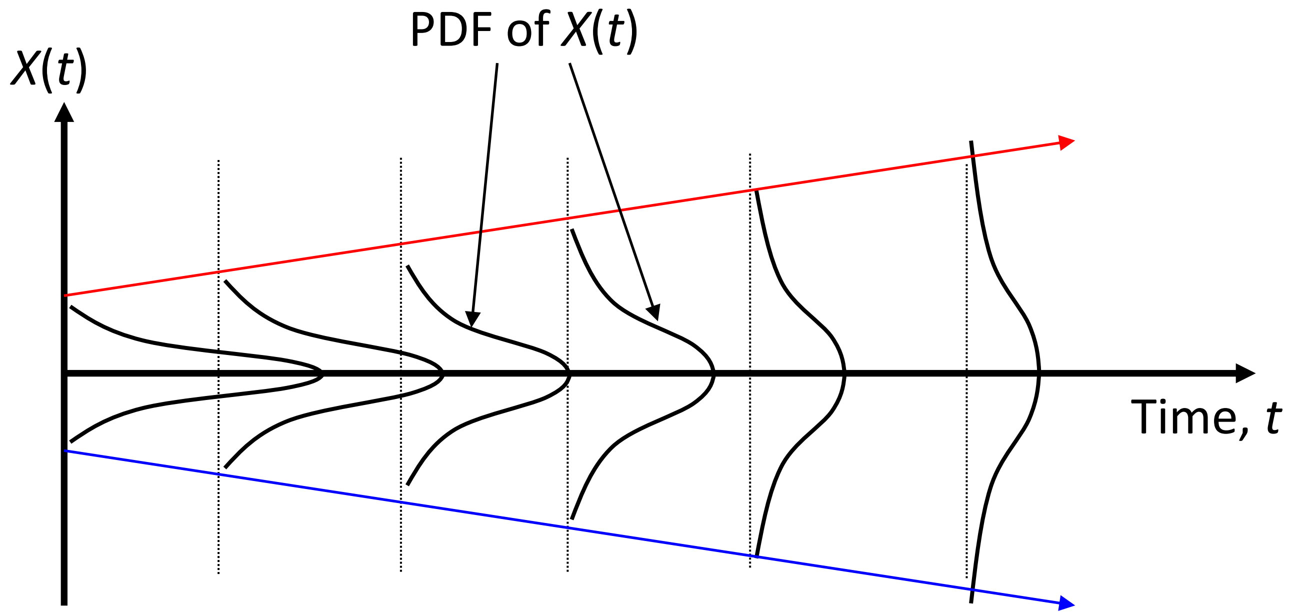
$$p(x_1,x_2,\ldots,x_n;\ t_1,\ t_1+\Delta_2,\ldots,t_1+\Delta_n)$$

does not depend on t_1 but only on $\Delta_2,\ldots\Delta_n$
i.e. it depends *only* on relative time between points

Stationarity: Summary

There are different notions of stationarity.

- **Strictly stationary:** (SSS: Strict Sense Stationary process) If none of the statistics of the random process are affected by a shift in the time origin. Strict-sense stationarity seldom holds for random processes, except for some Gaussian processes. Therefore, weaker forms of stationarity are needed.



Wide sense stationary (WSS): If the mean (e.g. above) and autocorrelation function do not change with a shift in the origin time.

Cyclostationary: If the mean and autocorrelation function are periodic in time.

SSS: Strict Sense Stationary process

The statistics of all the orders are unchanged by a shift in the time axis

$F_{X(t_1+\tau),..,X(t_n+\tau)}(x_1, \ldots, x_n) = F_{X(t_1),..,X(t_n)}(x_1, \ldots, x_n)$ $\forall$n, $\forall t_1, \ldots, t_n$

first-order strict sense stationary process, when $F(x) = \lim_{t\to\infty} P\{X_t \leq x\}$

- Or from n-th order we have $f_X(\boldsymbol{x}; \boldsymbol{t}) = f_X(\boldsymbol{x}; \boldsymbol{t} + c)$ for any *c*.

- In particular *c* = – *t* gives $f_X(\boldsymbol{x}; \boldsymbol{t}) = f_X(\boldsymbol{x})$
- i.e., the first-order density of *X*(*t*) is independent of *t*. In that case

$E[X(t)] = \int_{-\infty}^{\infty} x f(x) \mathrm{dx} = \mu$ which is a constant

- Similarly, for a

second-order strict-sense stationary process we have

$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1 + c, t_2 + c)$ for any *c*.

- For $c = -t_2$ we get $f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1 - t_2)$

SSS: Strict Sense Stationary process

One implication of stationarity is that the probability of the set of sample functions of this process which passes through the windows of the Fig.(a) is equal to the probability of the set of sample functions which passes through the corresponding time shifted windows of Fig. (b). Note, however, that it is not necessary that these two sets consist of the same sample functions. The distributions (which are probabilities are equal) . Equality in distribution

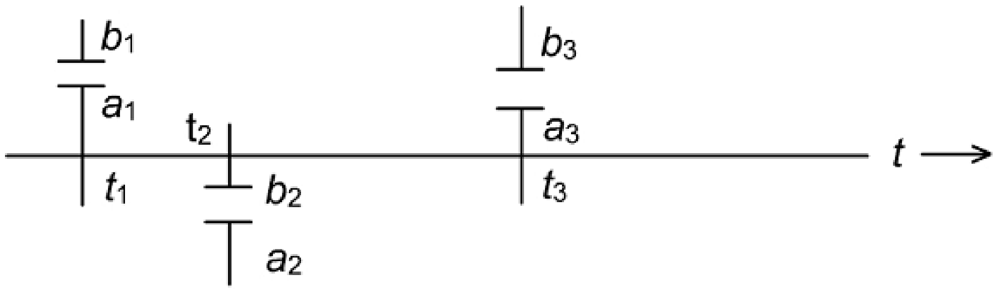


Fig.(a)

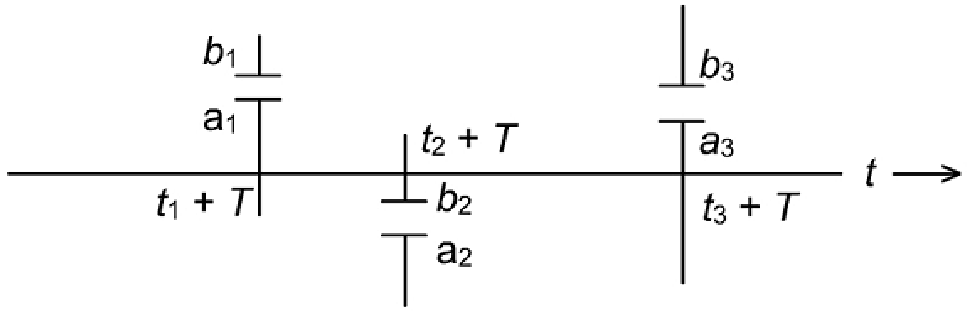


Fig.(b)

WSS: Stationarity in wide sense (or weak sense)

Specifying a stochastic process X(t) with its moments is a weaker condition than its distributions

A RP X(t) is WSS if its 1st and 2nd order statistics (mean and autocorrelation functions) are time invariant, i.e.

$\overline{X}_t$ = constant (e.g. $E[X(t)] = \mu$) and

$R_{t+\tau,s+\tau} = R_{t,s}$ ∀τ , (i.e. R(t1,t2)=R(τ))

in another word, correlation depends on the difference of t1,t2

Note: A WSS process has a constant variance

By definition:

$$\sigma_x^2 = E[(x(t) - \bar{x})^2] = E[x^2(t) - 2\bar{x}x(t) + \bar{x}^2] = E[x^2(t)] - 2\bar{x}E[x(t)\} + \bar{x}^2 = R_X(0) - \bar{x}^2$$

Example of WSS: example2: $x(t)=A\cos(2\pi t+\theta)$, $\theta \sim U(-\pi, \pi)$: Have shown that this process is WSS, i.e.

Mean = constant

ACF = function of τ only

What is variance for this example? Variance of Sinusoid w/ Random Phase is:

$$\sigma_x^2 = R_X(0) - \bar{x}^2 = \frac{A^2}{2} - 0 = \frac{A^2}{2}$$

Strict-sense stationarity → *wide-sense* stationarity

Reverse is not true

However, for Gaussian RP WSS implies SSS

Autocorrelation and stationarity

Note: For WSS :

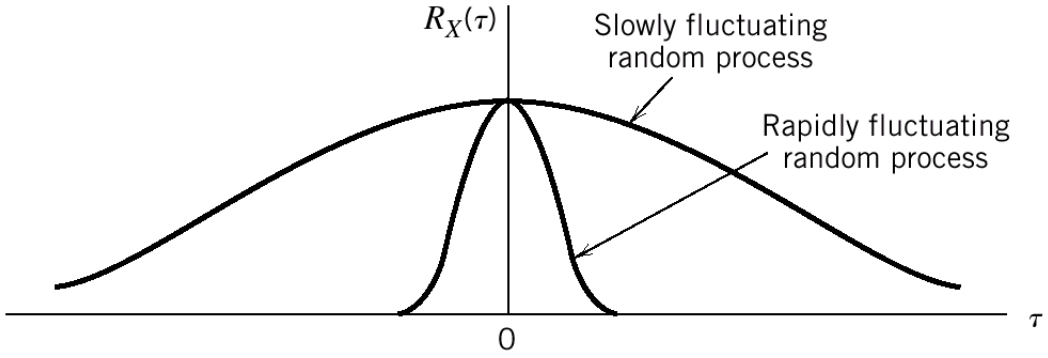
$$R_X(\tau) = R_X(t_1 - t_2, 0) = E[X(t_1 - t_2)X^*(0)]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X(0),X(t_2-t_1)}(x_1, x_2)dx_1 dx_2$$

$$R_X(t_1, t_2) = R_X(t_1, t_1 + \tau) = \underbrace{E[X(t_1)X^*(t_1 + \tau)]}_{} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_x(x_1, x_2; \tau)dx_1 dx_2 = R_X(\tau)$$

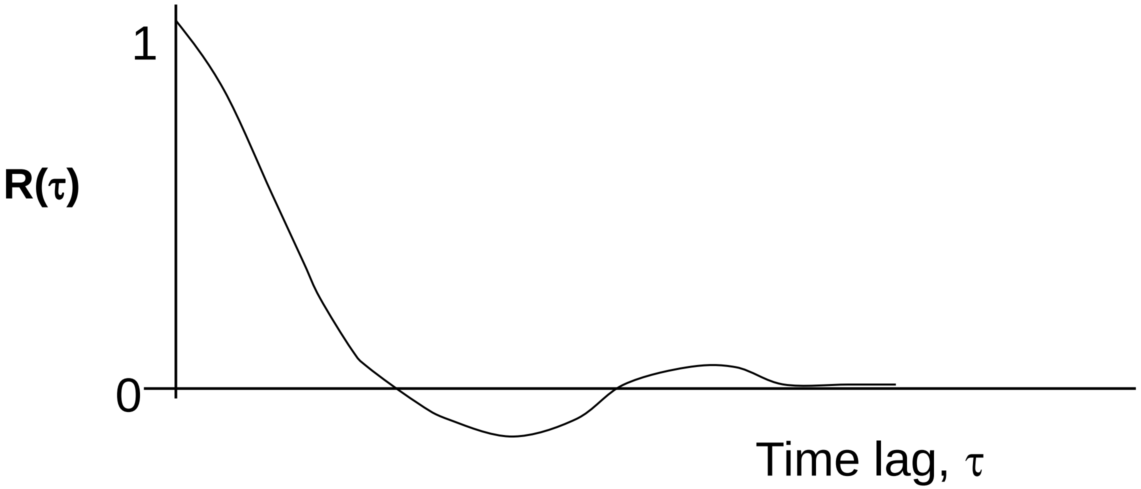
Does not depend on t_1 if x(t) is stationary

Autocorrelation

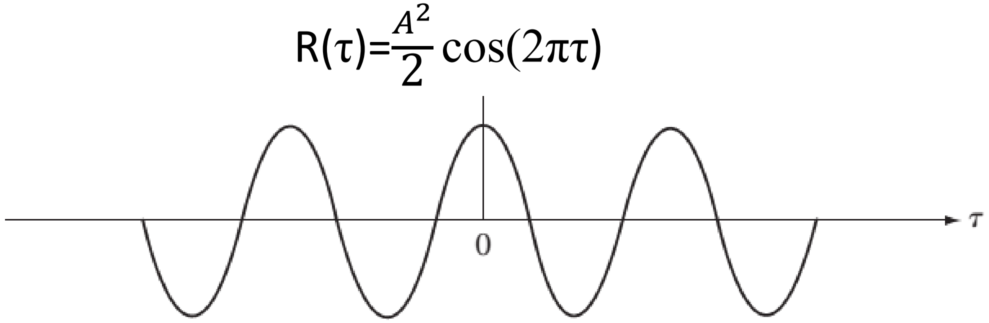
- Illustrating the autocorrelation functions of slowly and rapidly fluctuating random processes



- The autocorrelation for a random process eventually decays to zero at large τ



- The autocorrelation for a sinusoidal process is a cosine function which does not decay to zero



Properties of the Autocorrelation Function

- If x(t) is Wide Sense Stationary, then autocorrelation is a real function and has the following properties:

$$R(0) = E\left\{ |x(t)|^2 \right\}$$ this is the second moment

$$R(\tau) = R(-\tau)$$ even symmetry

$$R(0) \geq |R(\tau)|$$

Process of independent increments

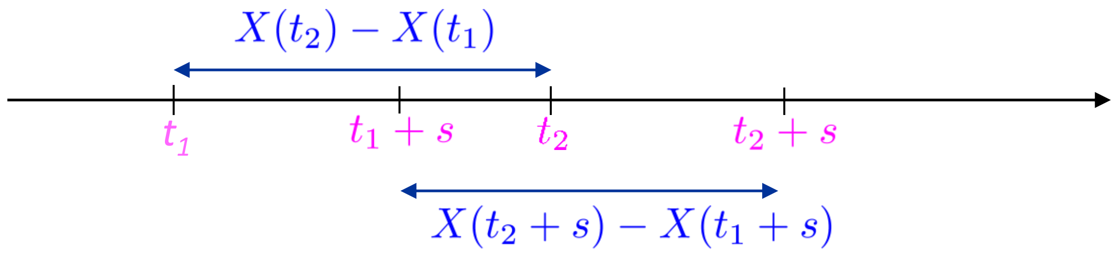
$X_{t_2} - X_{t_1}, \ \ldots, \ X_{t_n} - X_{t_{n-}}$ are independent $\forall t_1 < t_2 < \cdots < t_n$ (c.f. disjoint intervals)

- $P(N_1=n_1, N_2=n_2) = P(N_1=n_1)P(N_2=n_2)$ if N_1 and N_2 are disjoint intervals

$P(N_1 = n_1, N_2 = n_2) = P(N_1 = n_1)P(N_2 = n_2)$

Stationary Increments :

- A stochastic process X(t) is said to have stationary increments if $X(t_2+s)-X(t_1+s)$ and $X(t_2)-X(t_1)$ have the same distribution for all $t_1 < t_2$, $s > 0$.
- i.e. $X_{t+\tau} - X_t$ is a stationarity process $\forall\tau$ (i.e. does not depend on t)



- Examples: see Garcia's book

Ergodicity

Ergodic Processes - averages

For a strictly stationary process $X(t)$ and for any function $g(x)$, we can define two types of averages: statistical average , time average

Time Average versus Ensemble Average

- Example.
  - $X(t)$ is stationary.
  - For any t, $X(t)$ is uniformly distributed over {1, 2, 3, 4, 5, 6}.
  - Then *ensemble average* is equal to:

$$1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=3.5$$

Time Average versus Ensemble Average

- We make a series of observations at time 0, T, $2T$, …, $10T$ to obtain 1, 2, 3, 4, 3, 2, 5, 6, 4, 1. (They are deterministic!)
- Then, the *time average* is equal to:

$$\frac{1+2+3+4+3+2+5+6+4+1}{10}=3.1$$

Ergodic Processes - averages

For a strictly stationary process *X(t)* and for any function *g(x)*, we can define two types of averages: statistical average , time average

1. By looking at a given time t_0 and different realizations of the process, we have a RV $X(t_0)$ with density function $f_{X(t_0)}(x)$, which is independent of t_0 since the process is strictly stationary.

For this RV, we can find the statistical average (or *ensemble average*) of any function *g(X)* as

$$E[g(X(t_0))] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) f_{X(t_0)}(x) dx$$

This value is, of course, independent of t_0.

Ergodic Processes - averages

2. By looking at an individual realization (e.g realization *i*), we have a deterministic function of time $x(t;\omega_i)$. Based on this function, we can find the time average for a function *g(x)*, defined as

$$< g(x) >_i = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} g(X(t;\omega_i)dt$$

- $< g(x) >_i$ is, of course, a real number independent of *t* but, in general, it is dependent on the particular realization chosen (ω_i or *i*). Thus, for each ω_i we have a corresponding real number $< g(x) >_i$. Hence, $< g(x) >_i$ is the value assumed by a **random variable**. We denote this random variable by $< g(x) >_i$.

If it happens that for all functions *g(x)*, $< g(x) >_i$ is independent of *i* and equals $E[g(X(t_0))]$, then the process is called *ergodic*

Ergodic Processes

Definition 6.3. A stationary process(strict sense or WSS) *X(t)* is *ergodic* if for all functions *g(x)* and all $\omega_i \epsilon \Omega$

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(X(t;\omega_i))dt = E[g(X(t_0))]$$

i.e., if all time averages are equal to the corresponding statistical averages, then the stationary process is ergodic.

A natural consequence of ergodicity is that in measuring various statistical averages (mean and autocorrelation, for example), it is sufficient to look at *one realization* of the process and find the corresponding time average, rather than considering a large number of realizations and averaging over them.

Ergodic Processes

Note1: many types of ergodic process e.g.

1-iid- see garcia

2- WSS

- Mean Ergodic, if $g(x) = X(t)$
- Correlation(covariance) Ergodic, if $g(x) = R_X(T, \tau)$
- Distribution Ergodic (Papoulis ch.12)

Note2: Ergodicity is a very strong property. Unfortunately, in general, there exists no simple test for ergodicity. For the important class of Gaussian processes, however, there exists a simple test.

Note3: for Stationary processes ensemble average is constant, while statistical average is a RV

Exercise

- For any RP we are interested to know whether it is Stationary (in what sense) and ergodic
- Q: Characterize **Sinusoidal Wave with Random Phase** in terms of stationarity and ergodicity?
- Q: what is mean-ergodicity?
- Q: what is necessary and sufficient condition for mean ergodicity?
- Q: what is sufficient condition for ergodicity?
- A: see Papoulis Ch. 12
- Note: $C_X(t_1, t_2) = E[\{X_{t_1} - \overline{X}_{t_1}\}\{X_{t_2} - \overline{X}_{t_2}\}] = R_X(t_1, t_2) - m_X(t_1)m_X^*(t_2)$

Ergodic Process

- Experiments or observations on the same process can only be performed at different time.
- “Stationarity” only guarantees that the observations made at different time come from the same distribution.
- It can be shown, for instance, that a WSS process with finite variance at each instant and with a covariance function that approaches 0 for large lags is ergodic in the mean.
- Note that a (nonstationary) process with time-varying mean cannot be ergodic in the mean.

Ergodic Process

- Frequently, however, such ergodicity is simply assumed for convenience, in the absence of evidence that the assumption is not reasonable. Under this assumption, "the mean and autocorrelation can be obtained from time-averaging on a single ensemble member, through the following equalities: (here we don't require the pdf and joint pdf)

$$E\{x(t)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)dt \tag{9.40}$$

$$E\{x(t)x(t+\tau)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)dt \tag{9.41}$$

- A random process for which (9.40) and (9.41) are true is referred as second-order ergodic.

Ergodicity

If all of the sample functions of a random process have the same statistical properties the random process is said to be **ergodic**. The most important consequence of ergodicity is that ensemble moments can be replaced by time moments.

$$\mathrm{E}\left(X^n\right)=\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}X^n\left(t\right)dt$$

Every ergodic random process is also stationary.

Stochastic processes

- Reference
  - Kobayashi, Mark, Turin, Probability, Random Processes, and Statistical Analysis, Cambridge University Press 2012.
  - Papoulis, Probability,Random Variables, and Stochastic Processes 4th Edition, McGraw-Hill,2004
  - S.M. Ross, Stochastic Process, 2nd ed., John Wiley & Sons, Inc., 1996, chapter 2.
  - Leon Garcia, Probability and Random Processes for Electrical Engineering, **Addison-Wesley, Third Edition, 2008.**
  - Bertsekas D.P., Tsitsiklis J.N.- Introduction to Probability 2e. 2008
  - Yates, Goodman, Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers**, 3rd ed. 2014, wiley**

Appendix

Power of
Random Processes

Power of a Random Process

Recall : For deterministic signals…

instantaneous power is $x^2(t)$

For a random signal, $x^2(t)$ is a random variable for each time t. Thus there is no single # to associate with "instantaneous power". To get the Expected Instantaneous Power (i.e., on average) we compute the statistical (ensemble) average of $x^2(t)$:

Power of a Random Process

Often drop the "Average" terminology

This is also called the "mean square value" of the process

Avg. Power of x(t) : $P_X(t) = E\{ x^2(t)\}$

In general, can depend on time. But we'll see it doesn't for WSS (& stationary) processes

Relationship of Power to ACF

Recall: $R_X(t, t+\tau) = E\{x(t)\,x(t+\tau)\}$

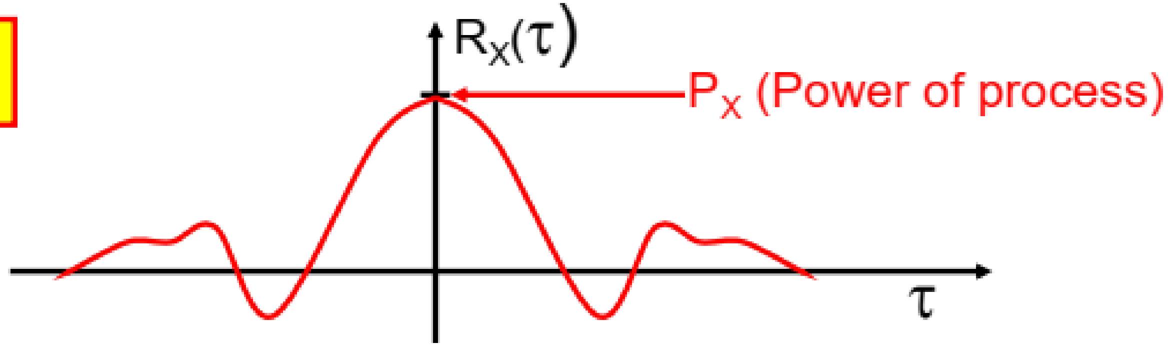
Clearly, setting $\tau = 0$ makes this equal to $P_X(t)$

$$\Rightarrow \mathbf{P_X(t) = R_X(t, t)}$$

If the Process is WSS (or SS) $R_X(t,t) = R_X(0)$

For WSS or SS

$$P_X = R_X(0)$$



Power vs. Variance

Recall Variance: $\sigma_x^2 = E\left\{[x(t) - \overline{x}(t)]^2\right\}$

$= \underbrace{R_x(0)}_{P_x} - \overline{x}^2$ for WSS

$$P_x = \sigma_x^2 + \overline{x}^2$$

"DC" Power ($\overline{x}^2$)

"AC" Power (σ_x^2)

Note: If the WSS process is **Zero Mean**, then:

$$P_X = \sigma_x^2$$

Power & Variance are Equal for Zero mean Process

Power Spectral Density of a Random Process

Recall: PSD for Deterministic Signal x(t) :

$$S_x(\omega) = \lim_{T \to \infty} \frac{|X_T(\omega)|^2}{T}$$

where $$X_T(\omega) = \int_{-T/2}^{T/2} x(t) e^{-j\omega t} dt$$

Power Spectral Density of a Random Process

For a random Process: each realization (sample function) of process x(t) has different FT and therefore a different PSD.

We again rely on averaging to give the **"Expected" PSD** or **"Average" PSD**…

But… **Usually just call it "PSD"**.

Define PSD for WSS RP

We define PSD of WSS process x(t) to be :

$$S_x(\omega) = \lim_{T \to \infty} E\left\{ \frac{|X_T(\omega)|^2}{T} \right\} \quad \bigstar$$

This definition isn't very useful for analysis so we seek an alternative form

The **Wiener-Khinchine Theorem** provides this alternative!!!

Weiner- Khinchine Theorem

Let x(t) be a WSS process w/ ACF $R_X(\tau)$ and w/ PSD $S_X(\omega)$ as defined in ( )… Then $R_X(\tau)$ and $S_X(\omega)$ form a FT pair :

$$S_X(\omega) = \mathcal{F}\{ R_X(\tau) \}$$

or Equivalently

$$R_X(\tau) \leftrightarrow S_X(\omega)$$

Some Properties of PSD & ACF

(1) The PSD is an <u>even function</u> of ω for a <u>real</u> process x(t)

<u>proof</u> : since each sample function is real valued then we know that $|x(\omega)|$ is even.

$\Rightarrow$ $|x(\omega)|^2$ is even: (even x even = even)

$\Rightarrow$ So is $S_X(w)$ (this is clear from ★)

Some Prop. of PSD & ACF

(2) $S_X(\omega)$ is real-valued and ≥ 0.

proof : again from (★) – since $|x(w)|^2$ is real-valued & ≥ 0, so is $S_X(w)$

(3) $R_X(\tau)$ is an even function of τ

$$
\begin{aligned}
R_x(-\tau) &= E\{x(t)x(t-\tau)\} \\
&= E\{x(\sigma)x(\sigma+\tau)\} \\
&= R_x(\tau)
\end{aligned}
$$

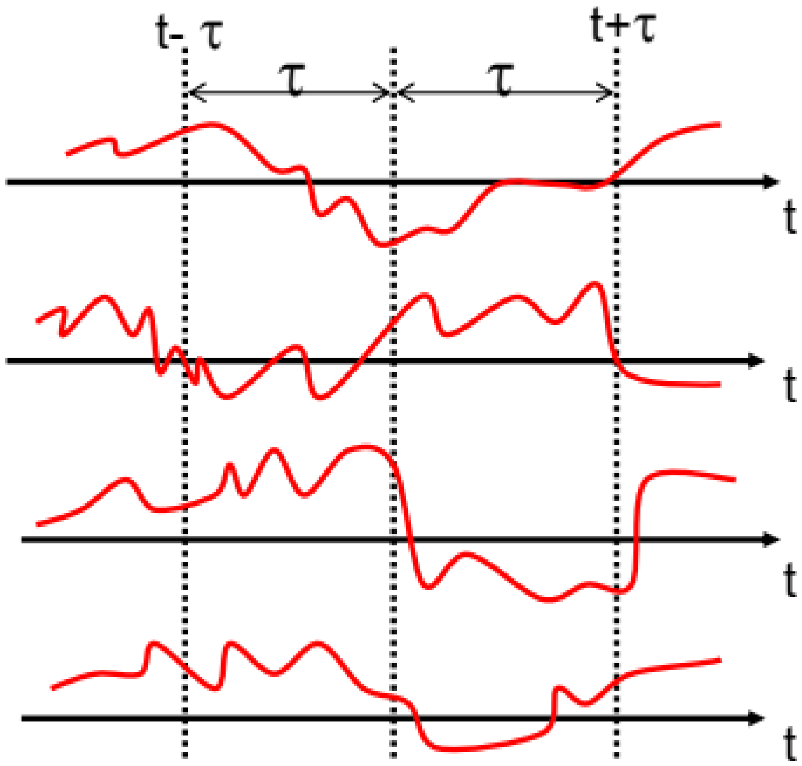
Also follows from IFT{ Real & Even } = Real & Even

<Property of the FT>

Graphical View of Prop #3

For WSS, ACF does not depend on absolute time only relative time

Doesn't matter if you look forward by τ or look backward by τ



PSD of RP- method 1
Wiener-Khintchine Theorem

- It can be shown that the PSD is also given by the FT of the autocorrelation function (ACF), $R_X(\tau)$,

$G_X(\omega) = F\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}d\tau$ or

$$S_X(f) = F\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)e^{-2j\pi f\tau}d\tau$$

where, $R_X\ (\tau) = E[x(t)x(t+\tau)]$

so it applies to a WSS random process X(t) with autocorrelation function $R_X(\tau)$

- We may find $R_X(\tau)$, by the inverse Fourier transform of $S_X(f)$

$$R_X(\tau) = F^{-1}\{S_X(f)\} = \int_{-\infty}^{\infty} S_X(f)e^{2j\pi f\tau}df$$

The Wiener-Khintchine result holds for random signals, choosing for *x(t)* any randomly selected realization of the signal

PSD of RP- method 1

- For real-valued $X(t)$,

$R_X(\tau)$ is an even, real-valued function of τ.

- From the properties of the Fourier transform, we find that

1. $S_X(f)$ is also real-valued and an even function of f
   i.e. $S_X(-f) = S_X(f)$ for all f

2. $S_X(f) \geq 0$

PSD of RP- method 1

- We use PSD to find power of the random signal, to do so, choose τ=0

  e.g. total power is:

  P_X= E[$(X^2(t))$]=$\frac{1}{2\pi}\int_{-\infty}^{\infty} G_X(\omega)d\omega = R_X(0) = \int_{-\infty}^{\infty} S_X(f)df$

  Proof: E[$(X^2(t))$]= E[$X(t)x(t-\tau)]|_{\tau=0} = R_X(\tau)|_{\tau=0} = R_X(0)R_X(\tau)$

  $= F^{-1}\{S_X(f)\} = \int_{-\infty}^{\infty} S_X(f)e^{2j\pi f.0}df$= $\int_{-\infty}^{\infty} S_X(f)df$

- we may find the expected power of $X(t)$ in a specific frequency range by integrating the PSD over that specific range

PSD, random signals-method 2

we use statistical average of the random signal (compare this with deterministic signal that we used x(t) to compute PSD i.e. $G_X(\omega) = \lim_{T\to\infty} \frac{|X_T(\omega)|^2}{T}$)

$$G_X(\omega) = \lim_{T\to\infty} \frac{E[|X_T(\omega)|^2]}{T}$$

Where $X_T(\omega) = \int_{-T}^{T} X(t)e^{-j\omega t}dt$ is the Fourier transform of X(t) limited to [−T, T]

Why we call $G_X(\omega)$ PSD: For a random signal power is obtained as:

$$P_X \stackrel{\text{def}}{=} E[\lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt]$$

Then it holds:

$$P_X = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_X(\omega) d\omega$$

PSD, random signals-method 2

Proof. First of all, we recall that P_X is the expected average power of $X(t)$. Let

$$X_T(t) = \begin{cases} X(t) & -\frac{T}{2} \le t \le \frac{T}{2} \\ 0 & otherwise \end{cases}$$

Then we can show that integrating over $-\infty$ to ∞ is equivalent to

$$\int_{-\infty}^{\infty} |X_T(t)|^2 = \int_{-T}^{T} |X(t)|^2$$

By Parseval's theorem, energy preserves in both time and frequency domain:

$$\int_{-\infty}^{\infty} |X_T(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_T(\omega)|^2 d\omega$$

Therefore, we can show that P_X satisfies

$$P_X \stackrel{\text{def}}{=} E[\lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt]$$

$$= E\left[\lim_{T\to\infty} \frac{1}{2\pi} \frac{1}{T} \int_{-\infty}^{\infty} |X_T(\omega)|^2 d\omega\right]$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{1}{T} E[|X_T(\omega)|^2] d\omega$$

PSD, random signals

The Wiener-Khintchine result holds for random signals, choosing for $x(t)$ any randomly selected realization of the signal

Note: Only applies for ergodic signals where the time averages are the same as the corresponding ensemble averages

Examples

Example1. Let $R_X(\tau) = e^{-2\alpha|\tau|}$, then

$S_X(\omega) = F\{R_X(\tau)\} = \frac{4\alpha}{4\alpha^2+\omega^2}$

Example2. Let $X(t) = A\cos(\omega_0 t + \Theta)$, $\Theta \sim$ Uniform$[0, 2\pi]$. Then we can show that

$$R_X(\tau) = \frac{A^2}{2}\cos(\omega_0\tau) = \frac{A^2}{2}\left(\frac{e^{j\omega_0\tau} + e^{-j\omega_0\tau}}{2}\right)$$

Then, by taking Fourier transform of both sides, we have

$$S_X(\omega) = \frac{A^2}{2}\left(\frac{2\pi\delta(\omega-\omega_0)+2\pi\delta(\omega+\omega_0)}{2}\right)$$

$$= \frac{\pi A^2}{2}[\delta(\omega - \omega_0) + 2\pi\delta(\omega + \omega_0)]$$

Example3. Given that $S_X(\omega) = \frac{N_0}{2} rect\left(\frac{\omega}{2W}\right)$, then

$R_X(\tau) = \frac{N_0}{2}\frac{W}{\pi} sinc(W\tau)$

<u>ERGODICITY QULATITAIVE DESC</u>.

- Qualitative (but not precise) description: a random process is said to be ergodic if its statistical properties (such as its mean and autocorrelation function) can be deduced from a single, sufficiently long sample (realization) of the process. One can discuss the ergodicity of various properties of a random process.

Ergodicity

☐**Definition**. A stationary process *X*(*t*) is *ergodic in the mean* if

1. $P\{\lim_{T\to\infty} < X(t) >_T = m_X\} = 1$ and
2. $\lim_{T\to\infty} Var[< X(t) >_T] = 0$

where

$$< X(t) >_i = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t;\omega_i)dt$$

Ergodicity

☐**Theorem:** Let $X(t)$ be a WSS process with $m_X(t) = m$ then

$$\lim_{T\to\infty} < X(t) >_T = m$$

in the mean square sense, if and only if

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} 1 - \frac{|u|}{2T} C_X(u) du = 0$$

where, $C_X(\tau) = E[X_t X_{t+\tau}]$

Proof: Garcia's book page 541

In keeping with engineering usage, we say that a WSS process is **mean ergodic** if it satisfies the conditions of the above theorem.

Ergodicity

☐**Definition**. A stationary process $X(t)$ is *ergodic in the autocorrelation function* if

1. $P\{\lim_{T\to\infty} R_X(\tau;T) = R_X(\tau)\} = 1$
2. $\lim_{T\to\infty} Var[R_X(\tau;T)] = 0$

where

$$R_X(\tau;T) =< X(t)X(t+\tau) >_T = \frac{1}{2T}\int_{-T}^{T} X(t)X^*(t+\tau)dt$$

Ergodicity

☐**Theorem:** Let $X(t)$ be a WSS process with $R_X(t_1, t_2) = R_X(t_2 - t_1) = R_X(\tau)$ then

$$\lim_{T \to \infty} < X(t)X(t+\tau) >_T = R_X(\tau)$$

in the mean square sense, if and only if

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} 1 - \frac{|u|}{2T} C_X(u) du = 0$$

Ergodicity-sufficient condition

☐**Theorem:** a WSS random process is mean ergodic if

$$\int_{-\infty}^{\infty} |C(u)|du < \infty$$

a discrete-time WSS random process is mean ergodic if

$$\sum_{k=-\infty}^{\infty} |C(k)| < \infty$$

Proof:See papapoulis

Ergodicity

☐**Example 9.48- Garcia 3rd ed. Page 542**

Q: Is the random telegraph process mean ergodic?

A:The covariance function for the random telegraph process is $C_X(\tau) = e^{-2\alpha|\tau|}$ so the variance of $< X(t) >_T$of is

$$Var[< X(t) >_T] = \frac{2}{2T}\int_0^{2T}\left(1 - \frac{|u|}{2T}\right)C_X(u)du = \frac{2}{2T}\int_0^{2T}\left(1 - \frac{|u|}{2T}\right)e^{-2\alpha u}du$$

$$< \frac{1}{T}\int_0^{2T} e^{-2\alpha u}du = \frac{1 - e^{-4\alpha T}}{2\alpha T}$$

The bound approaches zero as $\mathrm{T} \to \infty$ so $Var[< X(t) >_T] \to 0.$

Therefore the process is mean ergodic.

Ergodicity- Discrete time

the time-average estimate for the mean and the autocorrelation functions of discrete time X_n are given by

$$< X_n >_T = \frac{1}{2T+1} \sum_{n=-T}^{T} X_n$$

$$< X_{n+k} X_n >_T = \frac{1}{2T+1} \sum_{n=-T}^{T} X_{n+k} X_n$$

If $X_n$is a WSS random process, then $E[< X_n >_T]$ and so $< X_n >_T$ is an unbiased estimate for *m*. It is easy to show that the variance of $< X_n >_T$is

$$Var[< X_n >_T] = \frac{1}{2T+1} \sum_{k=-2T}^{2T} \left(1 - \frac{|k|}{2T+1}\right) C_X(k)$$

Therefore, $< X_n >_T$approaches *m* in the mean square sense and is mean ergodic if the expression in the above Eq. approaches zero with increasing *T*.

Ergodicity- Discrete time

Example 9.49 Ergodicity and Exponential Correlation p. 543 Garcia 3rd.ed.
Let X_n be a wide-sense stationary discrete-time process with mean *m* and covariance function $C_X(k) = \sigma^2\rho^{-|k|}$ for $|\rho| < k$ and $k + 0, \pm 1, \pm 2, \ldots$Show that X_n is mean ergodic. The variance of the sample mean i.e. $< X_n >_T$ is

$$Var[< X_n >_T] = \frac{1}{2T+1}\sum_{k=-2T}^{2T}\left(1 - \frac{|k|}{2T+1}\right)\sigma^2\rho^{-|k|}$$

$<\frac{1}{2T+1}\sum_{k=0}^{\infty}\sigma^2\rho^{-k} = \frac{2\sigma^2}{2T+1}\frac{1}{1-\rho}$

Therefore, $< X_n >_T$ approaches *m* in the mean square sense and is mean ergodic if the expression in the above Eq. approaches zero with increasing *T*.

Ergodicity- Discrete time

Example 9.50 Ergodicity of Self-Similar Process and Long-Range Dependence
Let $X_n$be a WSS discrete-time process with mean *m* and covariance function

$C_X(k) = \frac{\sigma^2}{2}\{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\}$for $1/2 < H < 1$ and $k + 0, \pm 1, \pm 2, \ldots$

X_n is said to be **second-order self-similar**. investigate the ergodicity of X_n. We rewrite the variance of the sample mean (i.e.$< X_n >_T = \frac{1}{2T+1}\sum_{n=-T}^{T} X_n$) as follows:

$$Var[< X_n >_T] = \frac{1}{(2T+1)^2}\sum_{k=-2T}^{2T}(2T+1-|k|)\,C_X(k)$$
$$= \frac{1}{(2T+1)^2}\{(2T+1)C_X(0) + 2\big(2TC_X(1)\big) + \cdots + 2C_X(2T)\}$$

It is easy to show (See Problem 9.105) that the sum inside the braces is$\sigma^2(2T+1)^{2H}$Therefore the variance becomes:

$$Var[< X_n >_T] = \frac{1}{(2T+1)^2}\sigma^2(2T+1)^{2H} = \sigma^2(2T+1)^{2H-2}$$

9.105. In Example 9.50 show that $Var[< X_n >_T] = \sigma^2(2T+1)^{2H-2}$

Problem:9.105. In Example 9.50 show that $Var[< X_n >_T] = \sigma^2(2T+1)^{2H-2}$

$$(2T+1)\times \quad C_X(0) = (2T+1)\frac{\sigma^2}{2}\{1^{2H} - 2|0|^{2H} + 1^{2H}\}$$

$$2\times 2T\times \quad C_X(1) = 2(2T)\ \frac{\sigma^2}{2}\{2^{2H} - 2\times 1^{2H} + 0^{2H}\}$$

$$2\times(2T-1)\times \quad C_X(2) = 2(2T-1)\frac{\sigma^2}{2}\{3^{2H} - 2\times 2^{2H} + 1^{2H}\}$$

…

$$2\times 2\times \quad C_X(2T-1) = 2\times 2\times\frac{\sigma^2}{2}\{(2T)^{2H} - 2\times(2T-1)^{2H} + (2T-2)^{2H}\}$$

$$2\times 1\times \quad C_X(2T) = 2\times 1\times\frac{\sigma^2}{2}\{(2T+1)^{2H} - 2\times(2T)^{2H} + (2T-1)^{2H}\}$$

$$Var[< X_n >_T] = \frac{1}{(2T+1)^2}(2T+1)^{2H}\times 2\times 1\times\frac{\sigma^2}{2} = \sigma^2(2T+1)^{2H-2}$$

Ergodicity- Discrete time

Example 9.50 Ergodicity of Self-Similar Process and Long-Range Dependence

The value of *H*, which is called the **Hurst parameter**, affects the convergence behavior of the sample mean. Note that if H=1/2, the covariance function $C_X(k) = \frac{1}{2}\sigma^2\delta_k$ becomes which corresponds to an iid sequence. In this case, the variance becomes $\frac{\sigma^2}{(2T+1)^2}$ which is the convergence rate of the sample mean for iid samples. However, for H>1/2 the variance becomes

$$Var[< X_n >_T] = \frac{\sigma^2}{(2T+1)^2}(2T+1)^{2H-1}$$

so the convergence of the sample mean is slower by a factor of $(2T+1)^{2H-1}$ than for iid samples.

The slower convergence of the sample mean when H>1/2 results from the long-range dependence of X_n. It can be shown that for large *k*, the covariance function is approximately given by:

$$C_X(k) = \sigma^2 H(2H-1)k^{2H-2}$$

For ½<H<1 , $C_X(k)$decays as $\frac{1}{k^n}$ where $0 < \alpha < 1$,which is a very slow decay rate. Thus the dependence between values of X_n decreases slowly and the process is said to have a long memory or long-range dependence.