Basics of Probability & Random Processes

Ref Material : Stark & Woods book.

Let us briefly understand what a probability space is. For this, we need to understand the defns. behind fields & σ- fields.

Consider a universal set Ω and a collection of subsets of Ω.
Let $E, F, \ldots$ denote the subsets in this collection.
This collection of subsets forms a field $\mathcal{M}$ if

1) $\phi \in \mathcal{M}$, $\Omega \in \mathcal{M}$ (null set & universal set are part of the field)
2) If $E \in \mathcal{M}$ and $F \in \mathcal{M}$, then $E \cup F \in \mathcal{M}$, $E \cap F \in \mathcal{M}$
3) If $E \in \mathcal{M}$, then $E^c \in \mathcal{M}$.

A σ-field $\mathcal{F}$ is a field that is closed under any countable set of unions, intersections & combinations.
i.e., if $E_1, E_2, \ldots, E_n \ldots$ belong to $\mathcal{F}$ so do
$\bigcup_{i=1}^{\infty} E_i$ (Set of all elements in every E_i) and $\bigcap_{i=1}^{\infty} E_i$ (Set of all elements in at least one E_i)

NOTE :

A set S is called countable if there is an injective function f from $S \longrightarrow$ natural nos $\{0, 1, 2, \cdots\}$

Ponder

1) Rationals are countably infinite
2) Real nos are uncountably infinite

Consider an experiment $\mathcal{H}$ with a sample description space Ω. If Ω has a countable number of elements, then every subset of Ω may be assigned a 'probability' consistent with the axioms such that for every event $E \in \mathcal{F}$

a) $P(E) \geq 0$

b) $P(\Omega) = 1$

c) $P(E \cup F) = P(E) + P(F)$ if $P(E \cap F) = 0$.

Now, the class of all subsets make up a σ field and each subset is an event.

When Ω is not countable i.e., $\Omega = \mathbb{R}$ (real line)
not every subset of Ω can be assigned a probability
consistent with the above axioms.

Only those subsets for which prob. can be assigned are
called 'events'.

The collection of those subsets is smaller than the
collection of all possible subsets that one can define on Ω.

This smaller collection is called a σ. field a.k.a Borel field.

So, the 3 objects $(\Omega, \mathcal{F}, P)$ form a "probability space"

Sample space description

Field

Prob. measure

Example :

Suppose we do a fair coin toss 'once'

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

σ field of events consists of the following sets:

$\{H\}$, $\{T\}$, ϕ , Ω

Prob. measure

$P(H) = \frac{1}{2}$

$P(T) = \frac{1}{2}$

$P(\phi) = 0$

$P(\Omega) = 1$

Ω

$\mathcal{F}$

$\mathcal{P}$

Recap of basics

$P(AB)$: joint prob. of events A and B.

Example :

A : Event when it rains

B : Event when it is sunny

C : Event when it rains & is sunny.

$$P(C) = P(AB)$$

Independence

Two events A & B are statistically independent iff

$$P(AB) = P(A)\,P(B)$$

(occurrence of one $\Rightarrow$ non-occurrence of the other)

Mutually Exclusive

$$P(AB) = 0$$

i.e., $P(A \cup B) = P(A) + P(B)$

Baye's Theorem

If A_i's $i = 1, \dots, n$ be a set of disjoint & exhaustive events over a prob. space $\mathcal{P}$

$$\bigcup_{i=1}^{n} A_i = \Omega \qquad A_i \cap A_j = \phi \text{ for } i \neq j$$

For any event B over $\mathcal{P}$

$$P(A_j \mid B) = \frac{P(B \mid A_j)\, P(A_j)}{\underbrace{\sum_{i=1}^{n} P(B \mid A_i)\, P(A_i)}_{P(B)}}$$

Probability Distribution Function $F_X(x)$

The PDF contains the information necessary to compute $P(E)$ for any event E in the Borel field of events.

$$F_X(x) = P\left[\{\xi : X(\xi) \leq x\}\right] = P_X([-\infty, x])$$

ξ in the underlying sample space

Set of all outcomes ξ such that $X(\xi)$ assumes values less than x.

i.e., $\{\xi : X(\xi) \leq x\} \subset \Omega$ is a subset of outcomes that under the mapping $X(\cdot)$ generates the set $[-\infty, x]$

Properties of PDF

1) $F_X(\infty) = 1$ $\quad F_X(-\infty) = 0$

2) $x_1 \leq x_2 \implies F_X(x_1) \leq F_X(x_2)$ (non decreasing property)

3) $F_X(x)$ is continuous from the right

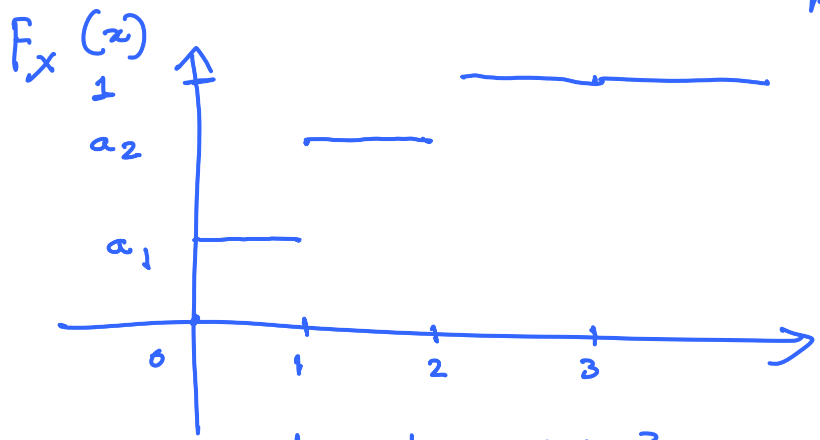
i.e., $F_X(x) = \lim_{\varepsilon \to 0} F_X(x+\varepsilon)$ $\quad \varepsilon > 0$

If $F_X(x)$ is discontinuous @ a point, say x_0, then $F_X(x_0)$ will be taken to mean the value of the PDF immediately to the right of x_0.

Example : Suppose x is a binomial random variable

$$P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$



Evaluated at $n = 3$

(a) Evaluate $P(1.5 < x < 3)$

$$= F_X(3) - P(X=3) - F_X(1.5)$$

$$= 1 - P(x=3) - a_2$$

Probability density function

If $F_X(x)$ is continuous and differentiable,

$$f_X(x) = \frac{d\, F_X(x)}{dx}$$

Properties

1) $f_X(x) \geq 0$

2) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ $\quad \left(F_X(\infty) - F_X(-\infty) = 1 - 0 \right)$

3) $F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx = Pr\left(X \leq x\right)$

4) $F_X(x_2) - F_X(x_1) = Pr\left(x_1 < X \leq x_2\right) = \int_{x_1}^{x_2} f_X(x)\, dx$

Continuous, discrete, mixed random variables

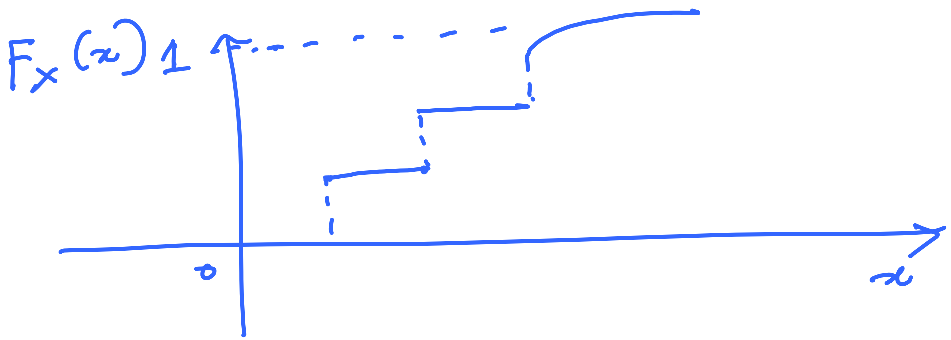
Examples:

(a) Gaussian random variable (Continuous)

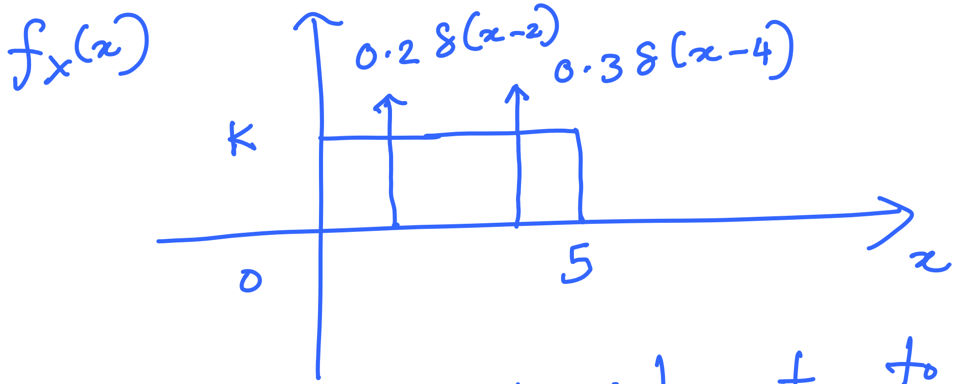
(b) Binomial distribution (Discrete r.v)

(c) Mixture of continuous & discrete random variables

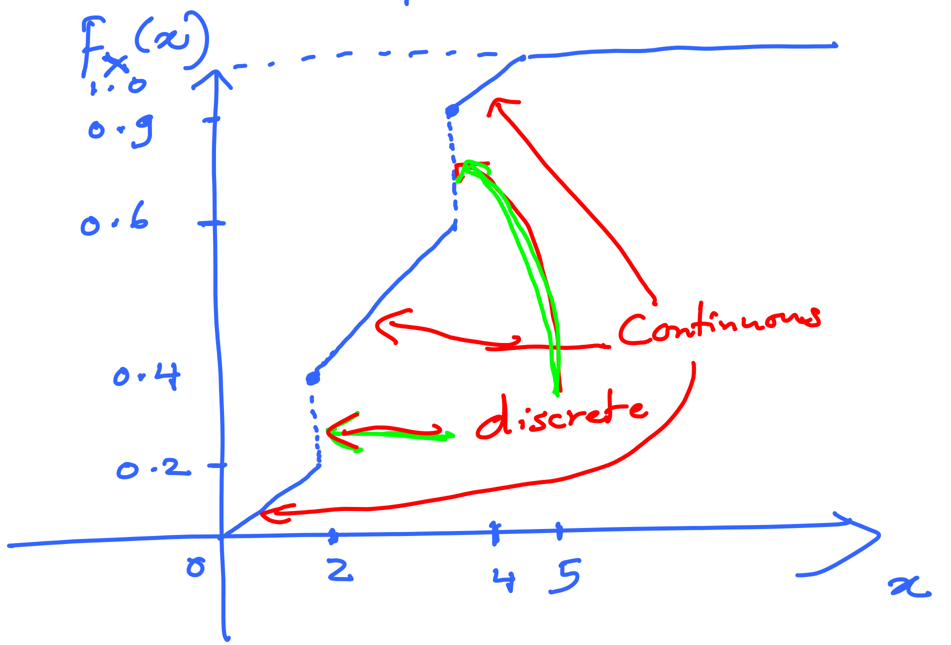
For a mixed random variable



Example : Suppose the pdf of a mixed random variable is



Sketch of the prob. distribution function



Since density must integrate to '1'

$$5k + 0.2 + 0.3 = 1$$

$$\Rightarrow \quad k = \frac{0.5}{5} = 0.1$$

Mean and Variance of a random variable (Discrete)

The mean of a random variable X with a prob. mass function (pmf) P_X is $E(X)$ given by

$$E(X) = \sum_i x_i \, P_X(x_i)$$

x_i's are possible values of X

Variance of a random variable X indicates the "spread" of the pmf of X.

$$\text{var} = E\left(\left(X - \mu_X\right)^2\right) = \sigma_X^2$$

σ_X is the std. deviation $\quad \sigma_X = \sqrt{\text{Var}(X)}$

NOTE: For a cont. r.v.

$$E(X) = \int_{\mathbb{R}} x\, f_X(x)\, dx$$ if the integral exists

Expectation is a linear operation

$$E\left(a\,g(X) + b\,h(X) + c\right) = a\,E(g(X)) + b\,E(h(X)) + c$$

One can compute higher moments $E(X^k)$ for different values of 'k'

Additional material

a) characteristic functions

b) moment generating functions

If X and Y are random variables with finite second moments,

Correlation : $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cancel{x}\,y\, f_{XY}(x,y)\,dx\,dy$

Covariance : $\mathrm{cov}(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right)$

$= E(XY) - E(X)\,E(Y)$

Correlation Coefft : $\rho_{XY} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$

Few things to note

1) If either X or Y has zero mean
$$E(XY) = \text{cov}(X,Y)$$

2) Random variables X and Y are uncorrelated if
$$\text{cov}(X,Y) = 0 \implies \rho_{X,Y} = 0$$

3) If X and Y are independent,
$$E(XY) = E(X)\,E(Y) \implies \text{cov}(X,Y) = 0$$
$\implies$ X and Y are uncorrelated statistical

The converse is not true i.e., uncorrelated $\not\Longrightarrow$ independence

Example : Consider 2 rvs X and Y with joint pmf populated below in the table

	$x_1 = -1$	$x_2 = 0$	$x_3 = 1$
$y_1 = 0$	0	1/3	0
$y_2 = 1$	1/3	0	1/3

$P_{XY}(0\ 1)$ = Joint Prob $(X=0, Y=1) = 0$

Let us Consider compute $P_{r_X}(0)$ $P_Y(1)$

$$P_Y(1) = \frac{2}{3} = \sum_{x \in \{-1,0,1\}} P_{xy}(X=x, Y=1)$$

$$P_X(0) = \frac{1}{3}$$

$$P(X=0\ Y=1) = 0 \neq P_X(0)\, P_Y(1) = \frac{2}{9}$$

(They are not independent)

$$E(X) = -1 \times \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = 0$$

$$E(XY) = \text{cov}(X, Y) = -1 \times 1 \cdot \frac{1}{3} + 1 \cdot 1 \cdot \frac{1}{3} + 0 \cdot 0 \cdot \frac{1}{3}$$

$$= 0$$

$\text{cov}(X, Y) = 0 \implies$ RVs X and Y are uncorrelated

So, uncorrelated $\not\Longrightarrow$ statistical independence

Orthogonal random variables

$E(XY) = 0 \implies$ rvs are orthogonal

Now,

$Cov(X, Y) = 0 \implies$ rvs are uncorrelated

When either $E(X)$ or $E(Y)$ are zero and (X and Y are orthogonal)

$$Cov(X\,Y) = E(XY) - E(X)\,E(Y)$$
$$= 0$$

For zero mean rvs orthogonality $\implies$ Uncorrelatedness

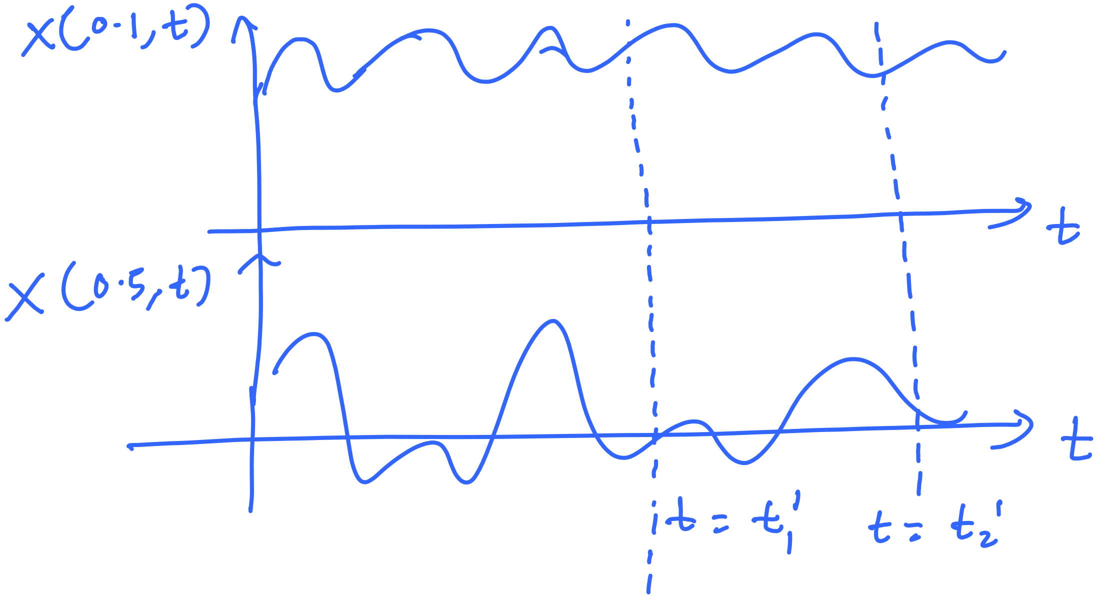
Overview of basics of random processes

A random process $x(t)$ (continuous/discrete) is a family of functions real/complex, scalar/vector defined on a probability space. At specified times $t_1, t_2, \ldots$ the samples $x(t_1)\ x(t_2) \ldots$ are random variables/random vectors.

Given $(\Omega, \mathcal{F}, P)$ $X(t, \xi) \in \mathcal{F}$ for a fixed 't' on the real line $-\infty < t < \infty$

When ξ is fixed, $X(t, \xi)$ is an ordinary function (time function)

when t is fixed, $X(t, \xi)$ becomes a random variable



$\xi = 0.1$

$\xi = 0.5$

$t = \pi/\omega$; $A \sin(\pi + \phi)$

$0, -A, 0$

$2/3$

$1/3$

$$x(t, \phi) = A \sin(\omega t + \phi)$$

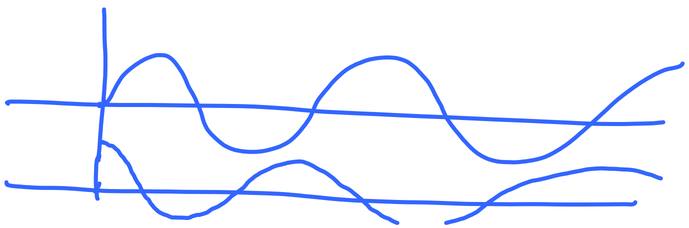
$\phi \in U[0, 2\pi]$

$t = 0$

$\phi = 0, \pi/2, \pi$

$A \sin(\phi) = \{0, A, 0\}$

$\{0, A\}$



$P^{(0)}(0) = 2/3$; $Pr^{(0)}(A) = \frac{1}{3}$

Mean & Correlations

The mean of a R.P. is $E(x(t))$ and can be a function of the time index.

$$\mu_x(t) \triangleq E[x(t)] \qquad -\infty < t < \infty$$

Auto correlation function

$$R_{xx}(t_1, t_2) = E[x(t_1)x^*(t_2)] \qquad -\infty < t_1, t_2 < \infty$$

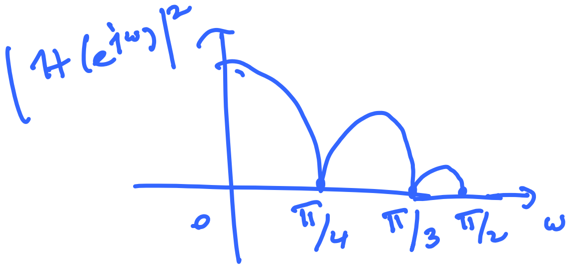
Eg.

$$x(t) = A e^{j2\pi ft}$$

Covariance function

$$\text{cov}_{xx}(t_1, t_2) = E\left[(x(t_1) - \mu_x(t_1))(x(t_2) - \mu_x(t_2))^*\right]$$

$$= R_{xx}(t_1, t_2) - \mu_x(t_1)\mu_x^*(t_2)$$



1 1 1 1 . . . .

1 −1 1 −1

1 1 −1 −1 1 1 −1 −1 . . .

1 −1 1 −1 1 1 1 −1 . . . .

<u>Example</u> : (Sinusoidal random process)

Suppose $X(t) = A \sin(\omega_0 t + \Theta)$ where $\Theta \sim U[-\pi, \pi]$

Suppose A is also a r.v. & let A & Θ be statistically independent.

$$\mu_X(t) = E\left[A \sin(\omega_0 t + \theta)\right]$$

$$= E[A]\, E\left[\sin(\omega_0 t + \theta)\right]$$

$$= E[A] \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(\omega_0 t + \theta)\, d\theta$$

$$= E[A] \cdot 0 \qquad = 0$$

Auto correlation

$$R_{xx}(t_1, t_2) = E\left(x(t_1)\, x^*(t_1)\right)$$

$$= E\left(A^2 \sin(\omega_0 t_1 + \theta)\, \sin(\omega_0 t_2 + \theta)\right)$$

$$= E(A^2)\, E\left(\sin(\omega_0 t_1 + \theta)\, \sin(\omega_0 t_2 + \theta)\right)$$

$$\frac{1}{2}\left[\underbrace{\cos\left(\omega_0 (t_1 - t_2)\right)}_{\text{constant}} - \underbrace{\cos\left(2\theta + (t_1 + t_2)\omega_0\right)}_{\text{depends on } \theta}\right]$$

that gets averaged to 0 over the cycle.

$$= \frac{1}{2} E(A^2) \cdot \cos\left[\omega_0 (t_1 - t_2)\right]$$

Auto correlation depends on the time lag!

Statistical Specification of a random sequence

A random sequence $X(n)$ is said to be statistically 'specified' by knowing the N^{th} order prob. dist. fns. for all integers $N \geq 1$ & times $n, n+1, \ldots, n+N-1$

$$F_X(x_n, x_{n+1}, \ldots, x_{n+N-1}; n, n+1, \ldots, n+N-1)$$

$$= Pr\left(X[n] \leq x_n, X[n+1] \leq x_{n+1}, \ldots, X[n+N-1] \leq x_{n+N-1}\right)$$

The representation we had is an infinite set of PDFs for each N because for all times $-\infty < n < \infty$, we need to know the joint PDF / CDF

$$\mu_X[n] = E\{X[n]\} = \int_{-\infty}^{\infty} x f_X(x; n)\, dx \quad \text{Cont. case}$$

$$= \sum_{k=-\infty}^{\infty} x_k \, P[X[n] = x_k]$$

Some extensions/ classifications of R.P.

Let X and Y be R.P. They are

a) uncorrelated : $R_{XY}(t_1, t_2) = \mu_X(t_1)\, \mu_Y^*(t_2)$

where $R_{XY}(t_1, t_2) \triangleq E[X(t_1)\, Y^*(t_2)]$

b) Orthogonal : $R_{XY}(t_1, t_2) = 0 \quad \forall\, t_1, t_2$

c) Independent : If for +ve integers n, the n^{th} order PDF of X and Y factors i.e.,

$$F_{XY}(x_1\, y_1\, x_2\, y_2 \ldots x_n\, y_n ;\, t_1\, t_2 \ldots t_n)$$
$$= F_X(x_1\, x_2 \ldots x_n ;\, t_1\, t_2 \ldots t_n)\, F_Y(y_1\, y_2 \ldots y_n ;\, t_1\, t_2 \ldots t_n)$$

Stationarity

A R. P. $X(t)$ is stationary if it has the same n^{th} order prob. dist. fns. as $X(t+T)$

$$F_X(x_1\, x_2 \dots x_n;\, t_1\, t_2 \dots t_n) = F_X(x_1\, x_2 \dots x_n;\, t_1+T, \dots, t_n+T)$$

If the PDF is differentiable,

$$f_X(x_1\, x_2 \dots x_n;\, t_1\, t_2 \dots t_n) = f_X(x_1\, x_2 \dots x_n;\, t_1+T, \dots, t_n+T)$$

Properties

1) Mean of a stationary process is a 'constant'

$$f(x; t) = f(x; t+T)$$

$$\Rightarrow f(x; t) = f(x; t+T)\Big|_{t=-T} = f(x; 0)$$

$$E[X(t)] = \mu_X(0)$$

2) 2^{nd} order density is shift invariant

$$f(x_1\ x_2; t_1\ t_2) = f(x_1\ x_2; t_1+T, t_2+T)$$

Upon choosing $T = -t_2$

$$f(x_1\ x_2; t_1\ t_2) = f(x_1\ x_2; t_1-t_2, 0)$$

$$E(X(t_1) X^*(t_2)) = R_{XX}(t_1-t_2, 0)$$

Weak form of stationarity (Wide Sense Stationary)

Defn: A R.P. is WSS if

$$E[X(t)] = \mu_X = \text{a constant}$$

$$E[X(t+\tau)\, X^*(t)] = R_{XX}(\tau) \qquad -\infty < \tau < \infty$$

Mean is a Const.
Auto Corr. depends only on the time lag

Example : Suppose $x(t) = A e^{j2\pi f t}$; f is known (real const.)

A is a real valued r.v. with $E(A) = 0$ $E(A^2) < \infty$

$$E[X(t)] = E(A e^{j2\pi f t}) = 0$$ ✓ (mean is a const.)

$$E[x(t+\tau)\, x^*(t)] = E\left[A e^{j2\pi f (t+\tau)}\, A e^{-j2\pi f t}\right]$$

$$= E(A^2)\, e^{j2\pi f \tau}$$

$$= R_{xx}(\tau)$$ ✓ (depends only on lag τ)

$\boxed{X(t) \text{ is WSS}}$

Exercises

1) If $X(t) = \sum_{k=1}^{M} A_k e^{j2\pi f_k t}$ & A_k's are uncorrelated
$E(A_k) = 0$
Examine if $X(t)$ is a W.S.S.

2) **Prove the following:**

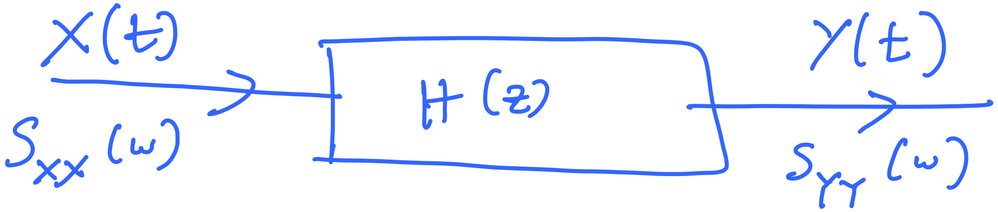
1) $|R_{xx}(\tau)| \leq R_{xx}(0)$

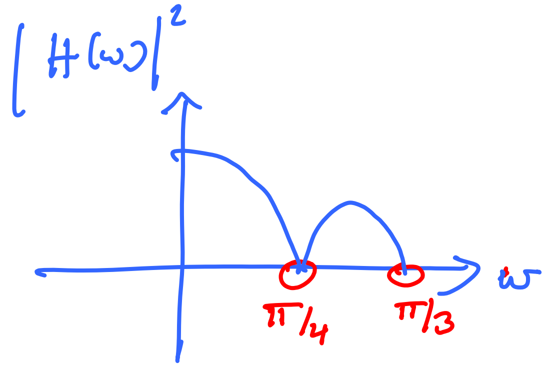
2) $|R_{xy}(\tau)| \leq \sqrt{R_{xx}(0)\, R_{yy}(0)}$

3) $R_{xx}(\tau) = R_{xx}^{*}(-\tau)$ complex
$R_{xx}(\tau) = R_{xx}(-\tau)$ real

4) $\forall N > 0$ $\forall t_1 < t_2 \ldots < t_N$ & complex $a_1 \; a_2 \ldots a_N$
$$\sum_{k=1}^{N} \sum_{l=1}^{N} a_k a_l^{*} R_{xx}(t_k - t_l) \geq 0$$

Filtering a WSS process through an LTI system





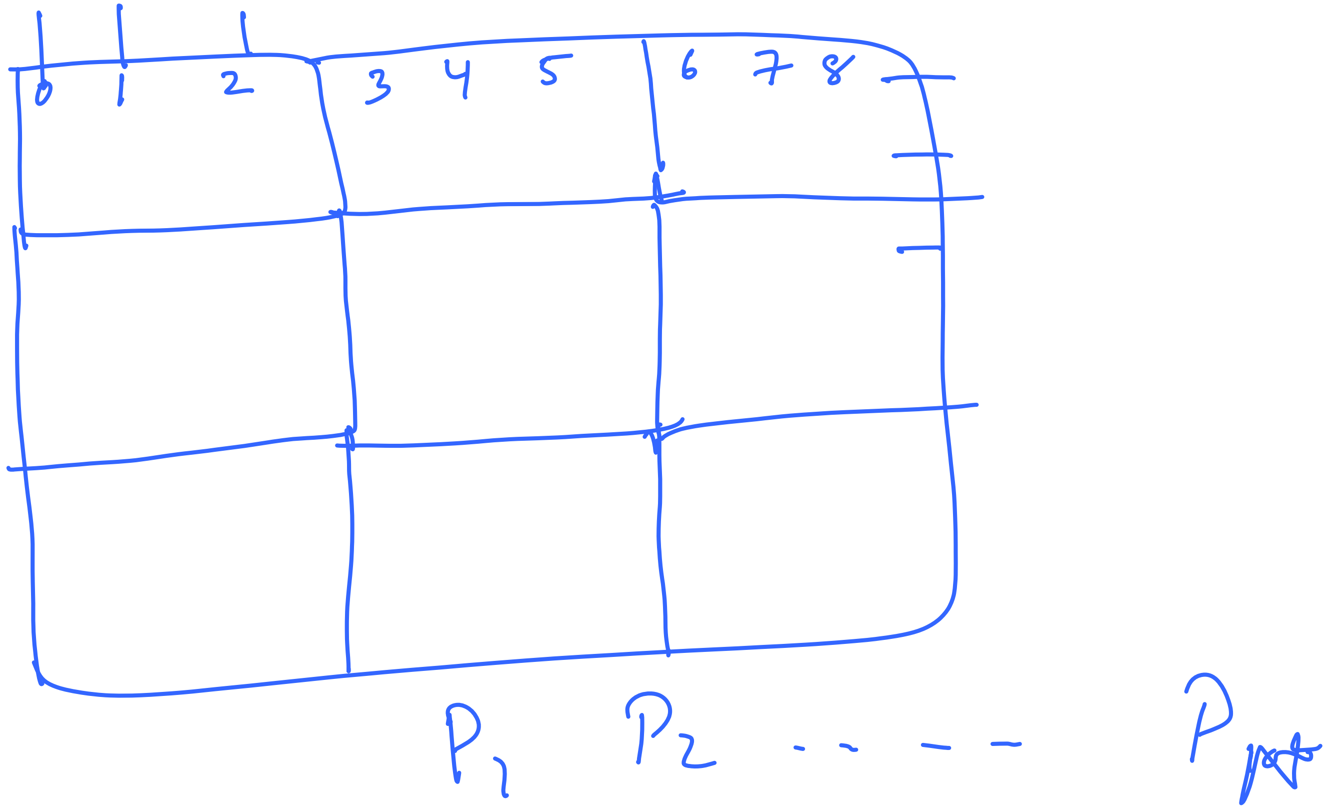
$$S_{YY}(\omega) = |H(\omega)|^2 \, S_{XX}(\omega)$$

where

$$S_{YY}(\omega) \triangleq \int_{-\infty}^{\infty} R_{YY}(\tau)\, e^{-j\omega\tau}\, d\tau$$

$$S_{XX}(\omega) \triangleq \int_{-\infty}^{\infty} R_{XX}(\tau)\, e^{-j\omega\tau}\, d\tau$$

The role of modulation codes is crucial to examine $S_{XX}(\omega)$ so that it does not coincide with a spectral null of $|H(\omega)|^2$



$\{P_i\}$