Statistical signal processing
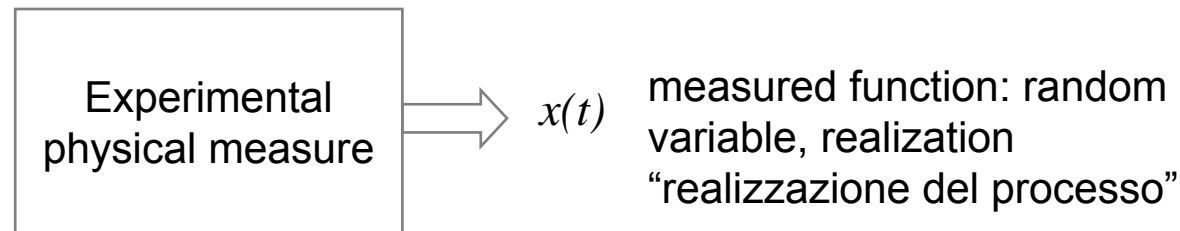
Short overview of the fundamentals

Gloria Menegaz

Outline

- Random variables

- Random processes
  - Stationarity
  - Ergodicity

- Spectral analysis

Random variable and processes

- Intuition: A random variable can be considered as the outcome of an experiment. This can be
  - A number (es lottery)
  - A function of time (es. EGC signal)
  - A function of space (es. noise in photographic images)

- Random (or stochastic) process: set of random variables $X_1(t), X_2(t), \ldots, X_n(t)$
  - The repetition of the experiment results in a set of random variables

Experimental physical measure	$\Rightarrow$ $x(t)$	measured function: random variable, realization "realizzazione del processo"

- NOTATIONS: X↔RV, x ↔realization

Gloria Menegaz

Deterministic and Random variables

Deterministic phenomenon

- Precise relation between causes and effects

- Repeatability

- Predictability

Stochastic phenomenon

- The relation between causes and effects is not given in mathematical sense

- There is a "stochastic regularity" among the different observations

- The "regularity" can be observed if a large number of observations is carried out such that the "expectations" of the involved variables can be inferred

Gloria Menegaz

Random variables

- A random variable (RV) is a real valued variable which depends on the outcomes of an experiment A and it is such that

  $$\{X \le x, X \in \mathbb{R}\} \text{ is an event}$$

  - (it gathers the outcomes of an experiment for which a probability is defined)
  - The probability of the *events* x=-∞ and x=+ ∞ is zero

  $$p\{x = -\infty\} = p\{x = +\infty\} = 0$$

- Otherwise stated: x(A) is a function which is defined over the domain A (that corresponds to the set of possible results) and whose value is a real number

Gloria Menegaz
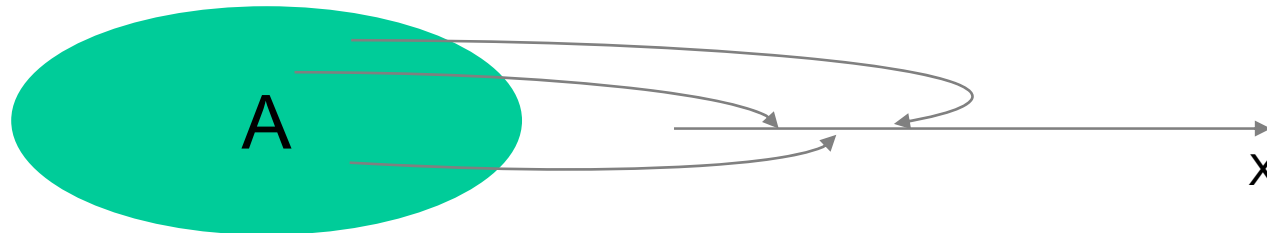
Random variables: résumé

- A random variable is a mapping between the sample space and the real line (*real-valued* RV) or the complex plan (*complex valued* RV)

Suppose we have a probability space $\{S, \Im, P\}$.

Let $X : S \rightarrow \Re$ be a function mapping the sample space s into the real line such that

For each $s \in S$, there exists a unique $X(s) \in \Re$. Then X is called a random variable.

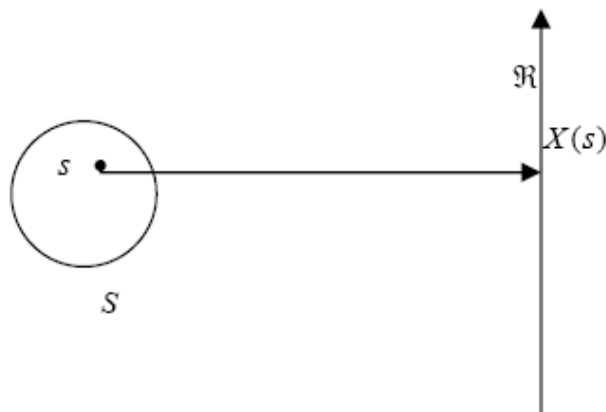Thus a random variable associates the points in the sample space with real numbers.

Notations:
- Random variables are represented by upper-case letters.
- Values of a random variable are denoted by lower case letters
- $Y = y$ means that y is the value of a random variable X.

Figure Random Variable

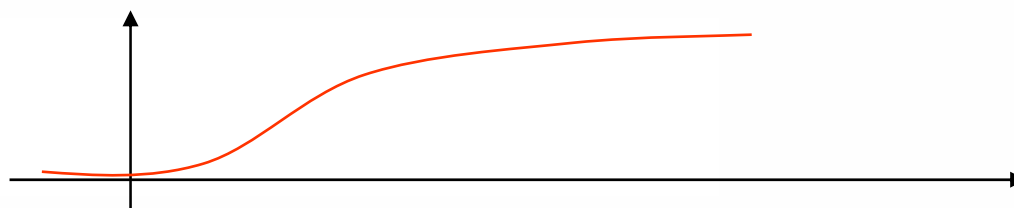Gloria Menegaz

Probability distribution function

We can define an event $\{X \le x\} = \{s \, / \, X(s) \le x, s \in S\}$

The probability

$F_X(x) = P\{X \le x\}$ is called the probability distribution function.

Given $F_X(x)$, we can determine the probability of any event involving values of the random variable X.

- $F_X(x)$ is a non-decreasing function of X.

- $F_X(-\infty) = 0$

- $F_X(\infty) = 1$

- $P\{x_1 < X \le x\} = F_X(x) - F_X(x_1)$

Discrete and Continuous RV

- The distinction concerns the *values* that can be taken by the RV
  - Continuous: there is no restriction of the set of values that x can take
  - Discrete: there exists a countable sequence of distinct real numbers x_i that the RV can take

- Discrete RV
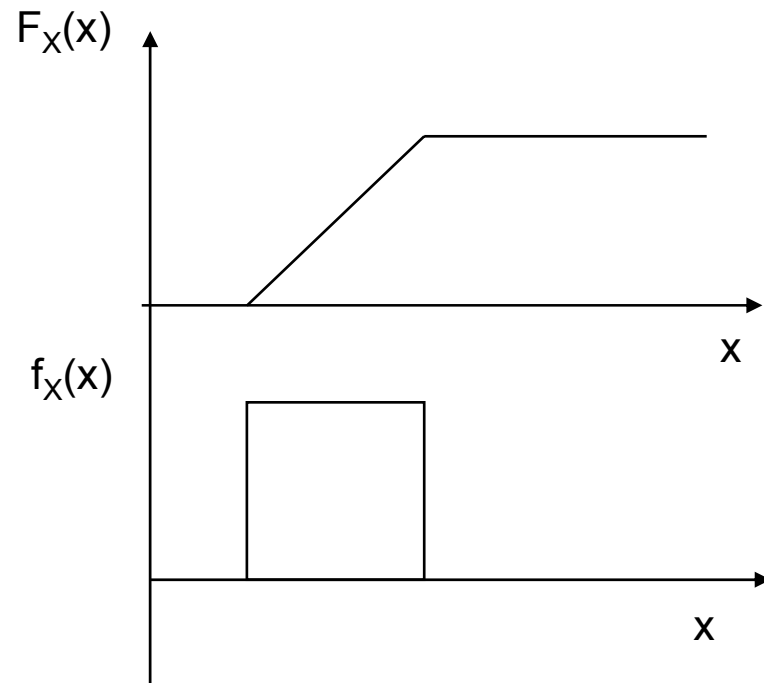 - Probability mass function replaces the distribution function
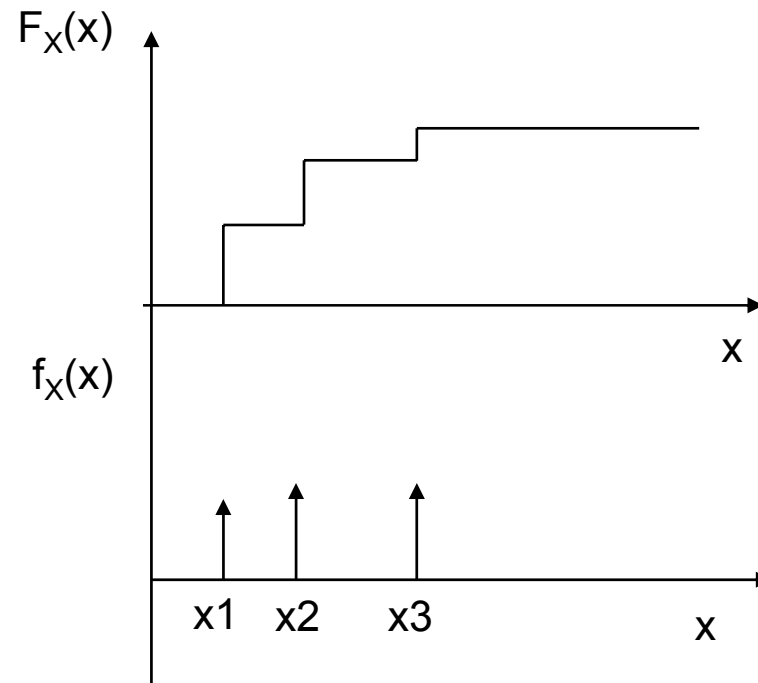
$$P_m = P\{x = x_i\}$$

$$\sum_i P_m\{x_i\} = 1$$

- Mixed RV: combination of the two types

Examples

Continuous RV
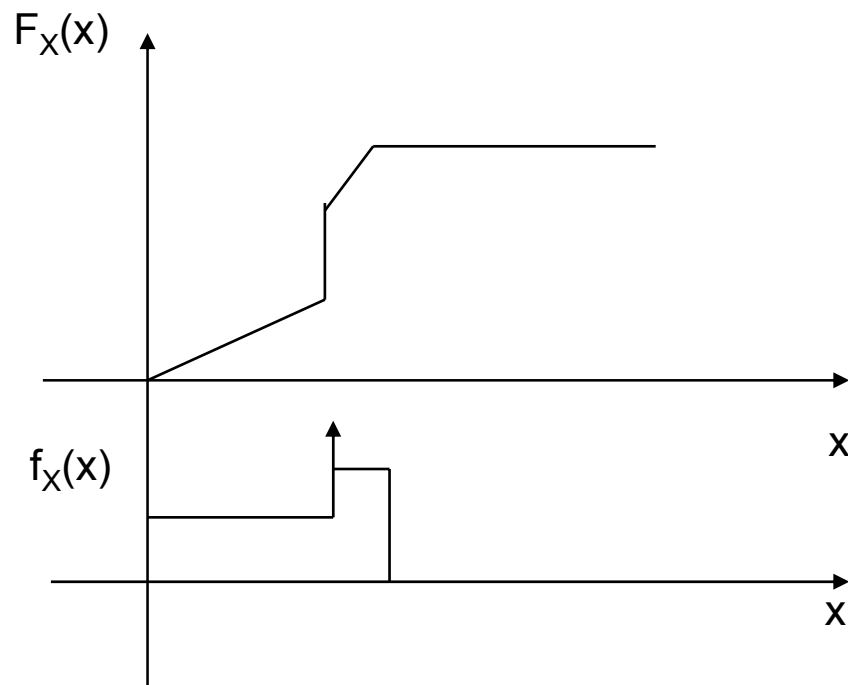
$F_X(x)$

$f_X(x)$

Discrete RV

$F_X(x)$

$f_X(x)$

x1   x2   x3   x

Deltas allow to derive the pdf of a discrete variable.The height of the deltas represents the probability of the event.

Gloria Menegaz

Examples

Mixed variable

$F_X(x)$

$f_X(x)$

x

x

Probability density function

If $F_X(x)$ is differentiable $f_X(x) = \dfrac{d}{dx} F_X(x)$ is called the probability density function and

has the following properties.

- $f_X(x)$ is a non- negative function

- $\displaystyle\int_{-\infty}^{\infty} f_X(x)dx = 1$

- $P(x_1 < X \le x_2) = \displaystyle\int_{-x_1}^{x_2} f_X(x)dx$

Gloria Menegaz

Joint random variables

- Given two random variables defined over the same space S, then

$$F_{X,Y}(x,y) = P\{x \le X, y \le Y\} \text{ joint distribution function}$$

$$\begin{cases} F_X(x) = F_X(x, \infty) \\ F_Y(y) = F_Y(\infty, y) \end{cases}$$ the events x=+∞ and y=+∞ are certain

$$\begin{cases} F_{XY}(x, -\infty) = 0 \\ F_{XY}(-\infty, y) = 0 \end{cases}$$ the events x=-∞ and y=-∞ have probability zero

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y) dx dy$$

Marginal density functions

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

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

$$= \frac{d}{dx} \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \right) dx$$

$$= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

$$\text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$$

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

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

Gloria Menegaz

Conditional density function

- Conditional density of Y given X

$$f_{Y|X}(y \mid X = x) = f_{Y|X}(y \mid x)$$

- Conditional distribution function

$$F_{Y|X}(y \mid x) = \frac{\displaystyle\int_{-\infty}^{y} f_{XY}(x,u)\,du}{f_X(x)}$$

- Thus

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} \qquad f_{Y|X}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Independent RV

Let X and Y be two random variables characterised by the joint density function

$$F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\}$$

and $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$

Then X and Y are independent if $\boxed{f_{X/Y}(x/y) = f_X(x) \quad \forall x \in \Re}$

and equivalently

$\boxed{f_{X,Y}(x,y) = f_X(x) f_Y(y)}$, where $f_X(x)$ and $f_Y(y)$ are called the marginal density functions.

Independent RV

- In case of more than two RV are involved

$$X_1, X_2 \ldots, X_n \text{ are independent } \rightarrow$$

$$F_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \ldots F_{X_n}(x_n)$$

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n)$$

Moments of a RV

- Expectations: provide a description of the RV in terms of a few parameters instead of specifying the entire distribution function or the density function

$$m_k = E\{X^k\} = \int_{-\infty}^{+\infty} x^k f_X(x)dx$$

- Mean (expectation value)

$$m_X = \mu_X = E\{X\} = \sum_{i=1}^{n} x_i P(x_i) \qquad \text{discrete random variable}$$

$$\mu_X = \int_{-\infty}^{+\infty} x f_X(x)dx \qquad \text{continuous random variable}$$

- For any piecewise constant function g(x), the expectation value is

$$Y = g(X)$$

$$E\{Y\} = E\{g(X)\} = \int_{-\infty}^{+\infty} g(x) f_X(x)dx$$

Moments of a RV

- Second order moment

$$E\{X^2\} = \sum_{i=1}^{n} x^2_i P(x_i)$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x) dx$$

- Variance
    - Measures the dispersion of the distribution around the mean
    - σ_x: standard deviation

$$\sigma^2_X = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx$$

Joint moments

- Joint moment of order $(k+r)$

$$m_{k,r} = E\left\{X^k Y^r\right\} = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} x^k y^r f_{XY}(x,y)\,dxdy$$

- Central joint moment

$$\mu_{k,r} = E\left\{X^k Y^r\right\} = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} (x-\mu_X)^k (y-\mu_Y)^r f_{XY}(x,y)\,dxdy$$

$$\mu_{2,0} = \sigma^2{}_X$$

$$\mu_{0,2} = \sigma^2{}_Y$$

Joint moments

- Covariance

$$\mu_{1,1} = \sigma_{XY} = E\{X^1 Y^1\} = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x,y) dx dy$$

$$\boxed{\sigma_{XY} = E\{XY\} - E\{X\}\{Y\} = E\{XY\} - \mu_X \mu_Y}$$

- Correlation coefficient
  - Measures the "similarity" among two random variables

$$\rho(X,Y) = \frac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sqrt{E\{(X - \mu_X)\}^2 E\{(Y - \mu_Y)\}^2}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\left\| \rho(X,Y) \right\| \leq 1$$

$$\left\| x \right\|^2 = \langle x, x \rangle = \int\limits_{-\infty}^{+\infty} x^2 f_X(x) dx$$

Uncorrelated RV

- Two RV are uncorrelated iif

$$\sigma_{XY} = 0$$

$$E\{XY\} = E\{X\}E\{Y\}$$

- – **Warning**: uncorrelation does not imply statistical independence BUT statistical independence implies uncorrelation
- – Two uncorrelated RV are also independent iif they are jointly Gaussian

  Uncorrelated RV can de dependent
  Independent RV are uncorrelated

Orthogonal RV

- *Orthogonality condition*

$$E\{XY\} = 0$$

- *For zero-mean RV orthogonality ↔ independence*

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

$$\begin{cases} E\{XY\} = 0 \\ E\{X\} = E\{Y\} = 0 \end{cases} \to Cov(X,Y) = 0$$

Sum of uncorrelated variables

- If Z is the sum of X and Y that are uncorrelated, its variance is given by the sum of the respective variances

$$Z = X + Y$$

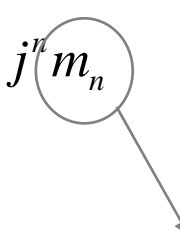$$\sigma^2{}_Z = \sigma^2{}_X + \sigma^2{}_Y$$

Characteristic function

- Expectation of the exponential function

$$M_X(\omega) = E\left\{e^{j\omega X}\right\} = \int_{-\infty}^{+\infty} e^{j\omega X} f_X(x)dx$$

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-j\omega X} M_X(\omega)dx$$

$$\left.\frac{\partial^n M}{\partial \omega^n}\right|_{\omega=0} = j^n m_n$$

n-order moment of X

The characteristic function only depends on the moments of X and can be used to estimate the pdf

Law of large numbers

- Let $X_1, X_2, ..., X_n$ be a sequence RV and let their expected value be μ_{X_i}

- Then, let $\hat{\mu}_N$ be the RV corresponding to their sum

$$\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} X_i \text{ then, if the RV are uncorrelated and have finite variance}$$

$$\hat{\mu}_N \to \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mu_{X_i} \text{ (mean square convergence)}$$

Then, if all the RV have the same mean $\mu_{X_i} = \mu_X$

$$\begin{cases} \hat{\mu}_N \to \mu_X \\ \sigma^2_{\hat{\mu}_x} = E\left\{\left[\hat{\mu}_x - E\{\hat{\mu}_x\}\right]^2\right\} = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2_{X_i} \end{cases}$$

The variance in the estimation of the expected value can be made arbitrarily small by adding a sufficiently large number of RV.

Central limit theorem

- The distribution of the sum of n *independent* RV having means μ_{Xi} and variance σ^2_{Xi} is a RV having mean and variance given by the following relations

$$X = \sum_{i=1}^{n} X_i$$

$$\mu_X = \sum_{i=1}^{n} \mu_{X_i}$$
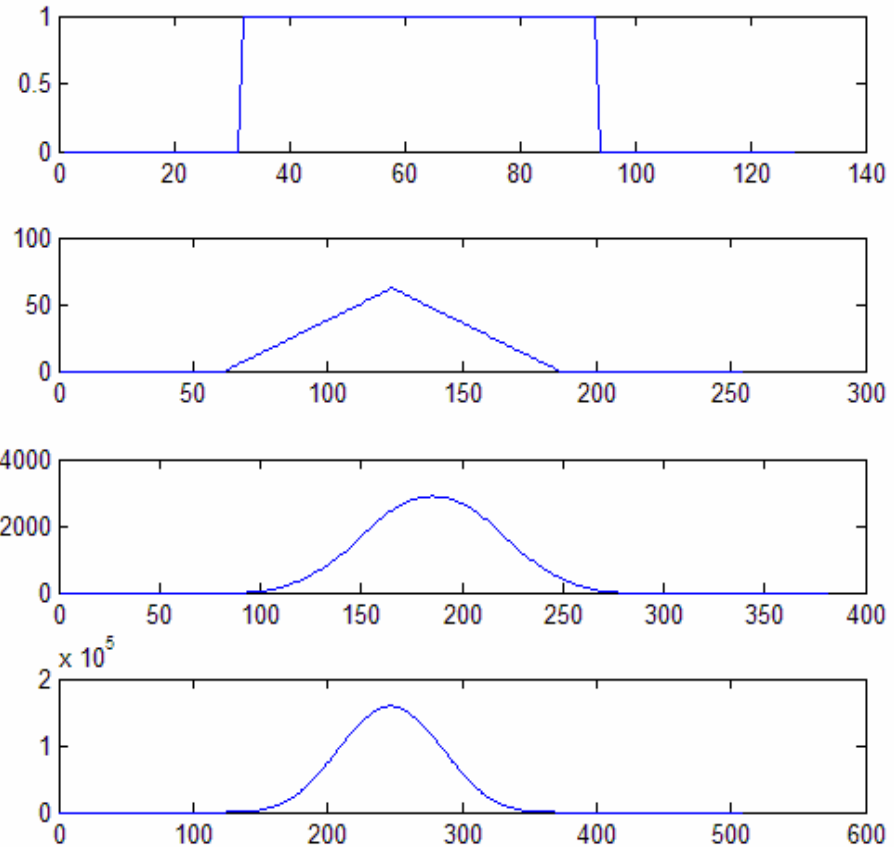
$$\sigma^2{}_X = \sum_{i=1}^{n} \sigma^2{}_{X_i}$$

- and its pdf is given by the convolution of the pdfs of the RV

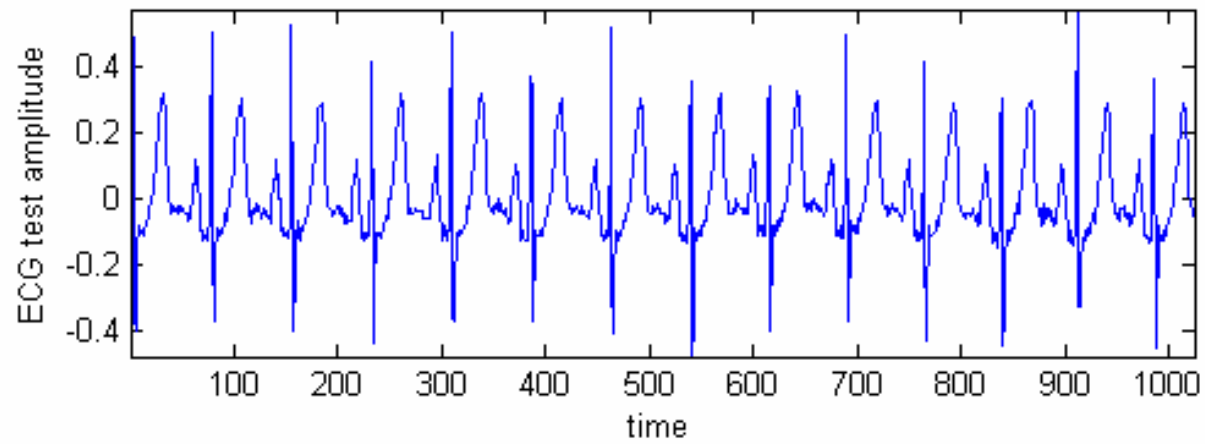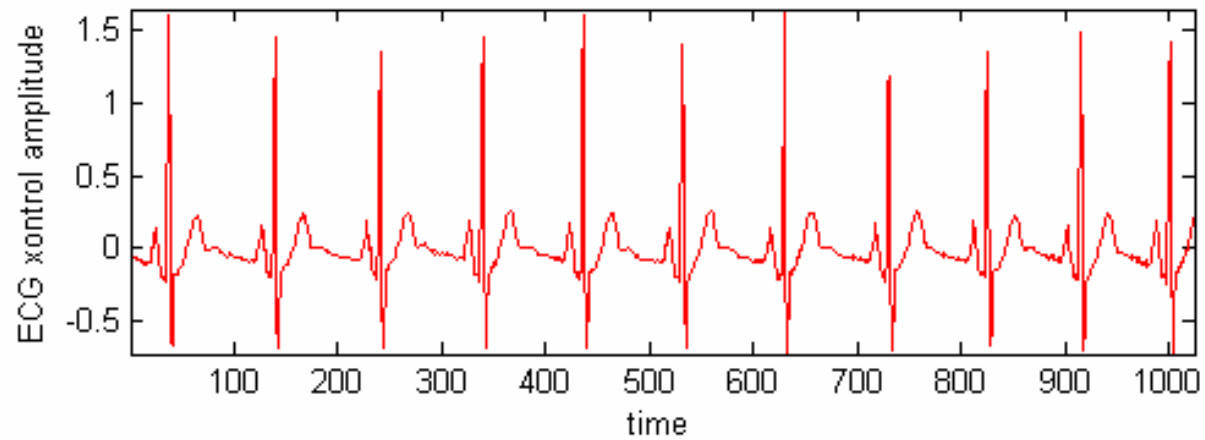$$f_X(x) = f_{X_1}(x_1) * f_{X_1}(x_1) * \ldots * f_{X_n}(x_n)$$

> It converges to a Gaussian irrespectively of the type of the single pdfs
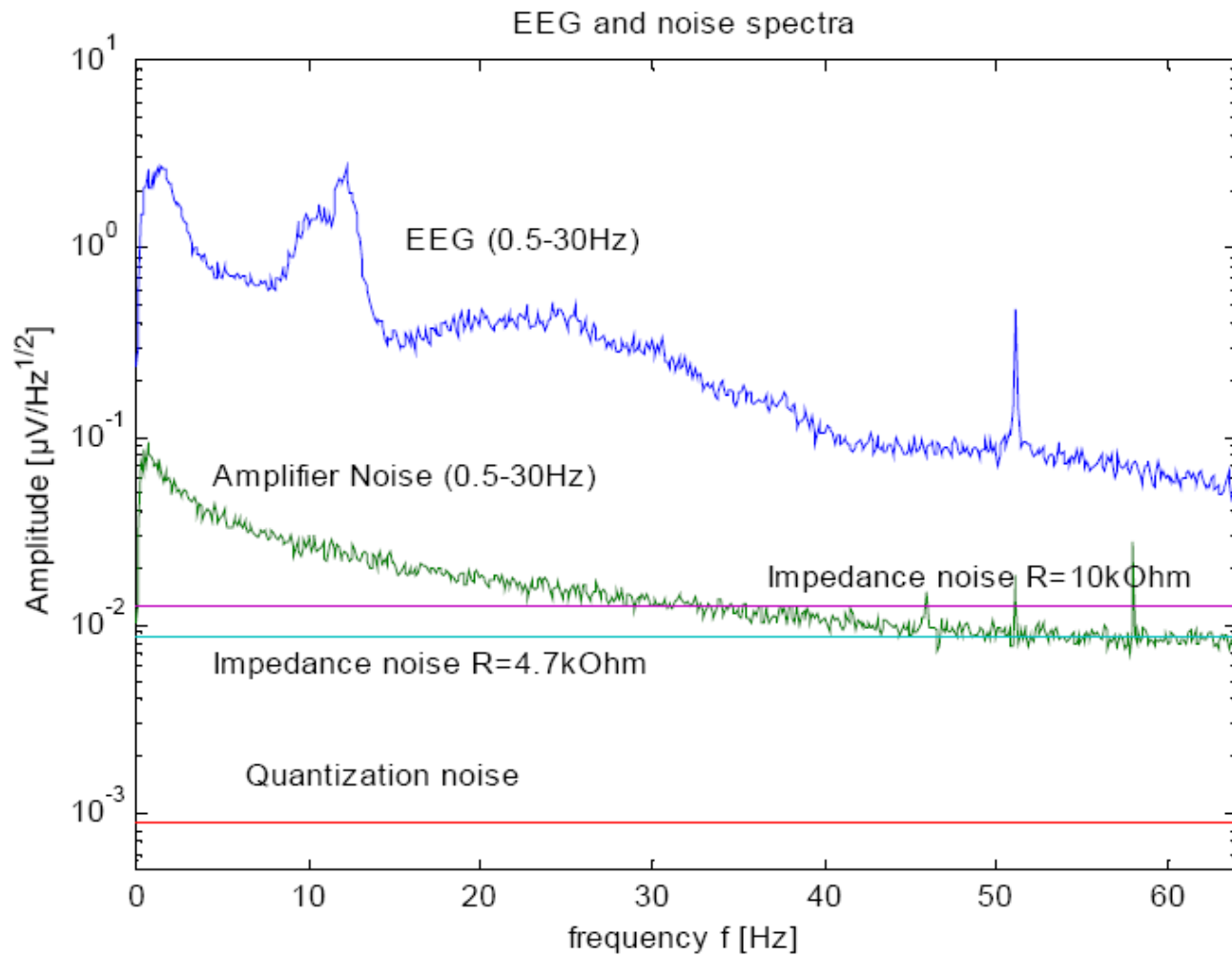
Central limit theorem

$$f_X(x) \underset{n \to \infty}{\to} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu_X)^2}{2\sigma^2 x}}$$

Random processes

Random processes

EEG and noise spectra

Random processes

- A random process can be defined as an indexed family of random variables $\{X(t), \; t \in T\}$ where T is an index set which may be discrete or continuous usually denoting time.

- The random process is defined on a common probability space $\{S, \Im, P\}$.

- A random process is a function of the sample point ξ and index variable t and may be written as $X(t, \xi)$.

- For a fixed $t(= t_0)$, $X(t_0, \xi)$ is a random variable.

- For a fixed $\xi(= \xi_0)$, $X(t, \xi_0)$ is a single realization of the random process and is a deterministic function.

- When both t and ξ are varying we have the random process $X(t, \xi)$.

The random process $X(t, \xi)$ is normally denoted by $X(t)$.

Gloria Menegaz

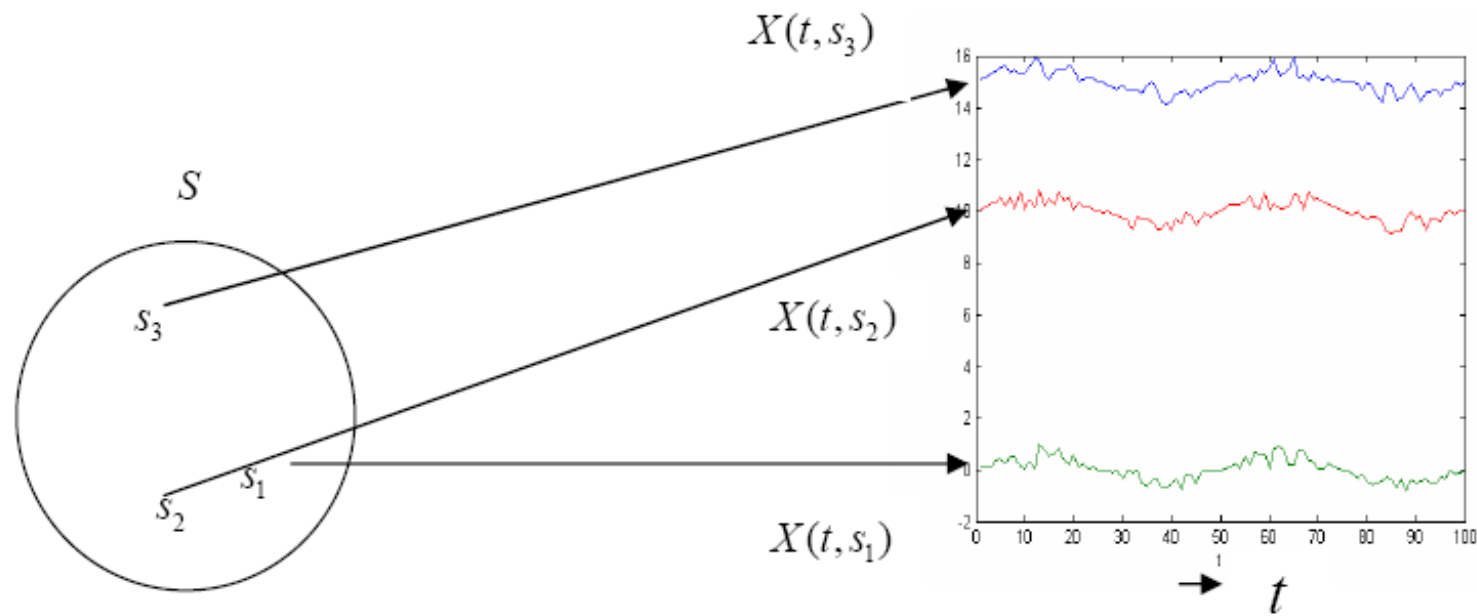# Random processes

realizations (functions of time)

Figure   Random Process

Space of realizations
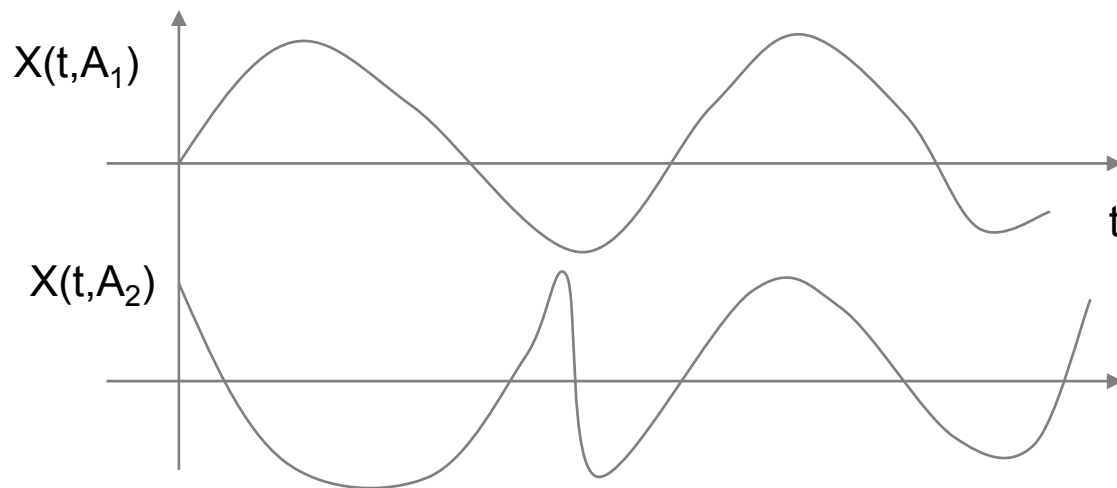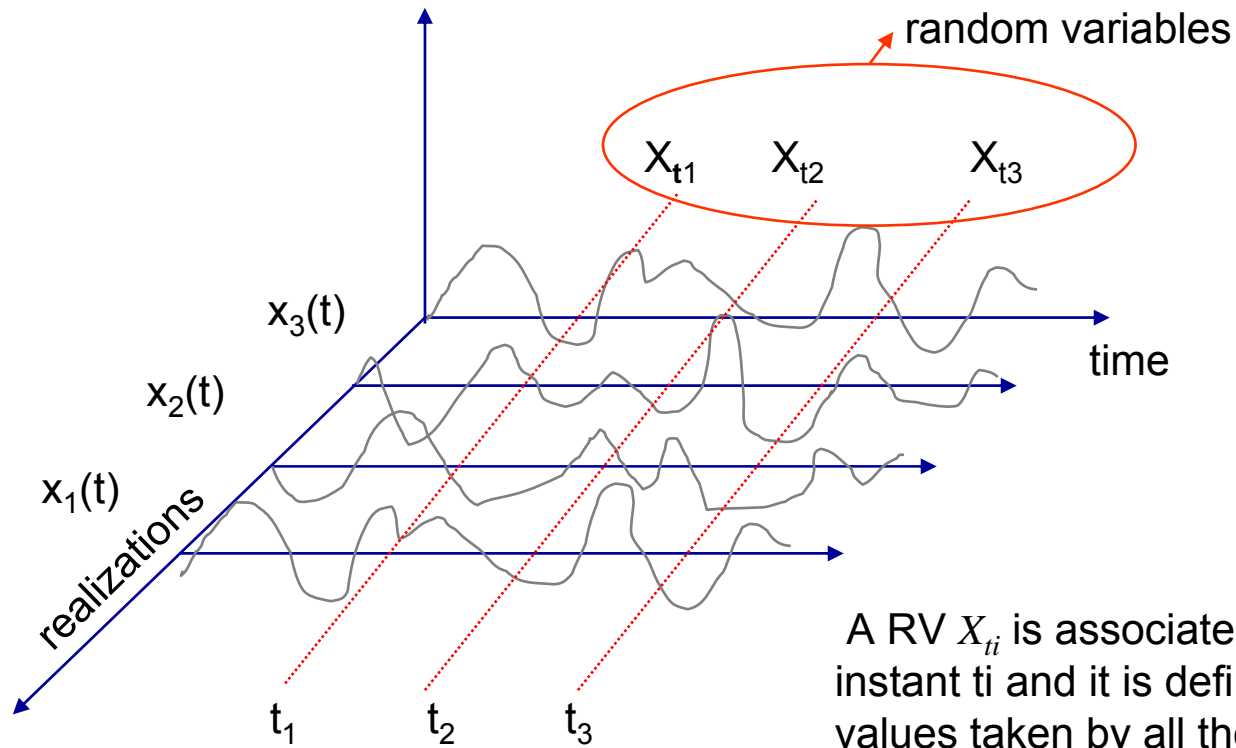
Random processes

- In a stochastic or random process there is some indeterminacy in its future evolution described by probability distributions. Even if the initial condition (or starting point) is known, there are many possibilities the process might go to, but some paths are more probable and others less

Single realizations are **deterministic** functions

$X(t,A_1)$

$X(t,A_2)$

Random processes

random variables

X_{t1} X_{t2} X_{t3}

$x_3(t)$

$x_2(t)$

$x_1(t)$

realizations

time

t_1 t_2 t_3

A RV X_{ti} is associated to any time instant ti and it is defined by the set of values taken by all the realizations at $t=t_i$

The set of functions concerning all the values of t completely define the Random process

Random variables

- The set of values taken by the realizations $x_k(t)$ in a given instant t_i generates a sequence describing a random variable $X(t_m)$

- This random variable can be described by a probability density function
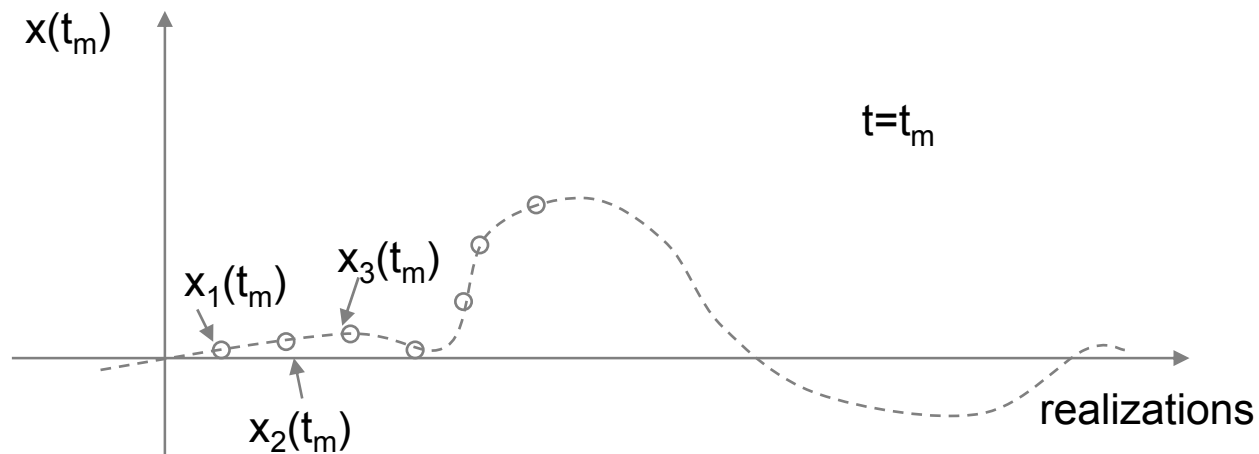
$$ f_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}} \left(x_{t_1}, x_{t_2}, \ldots, x_{t_n} \right) = f_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}} \left(x_1, x_2, \ldots, x_n, t_1, t_2, \ldots, t_n \right) $$

Random processes

To describe $X(t)$ we have to use joint density function of the random variables at different t.

For any positive integer n, $X(t_1), X(t_2), \ldots X(t_n)$ represents n jointly distributed random variables. Thus a random process can be described by the joint distribution function $F_{X(t_1), X(t_2), \ldots X(t_n)}(x_1, x_2, \ldots x_n) = F(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n), \forall n \in N$ and $\forall t_n \in T$

Otherwise we can determine all the possible moments of the process.

$E(X(t)) = \mu_x(t)$ = mean of the random process at t.

$R_X(t_1, t_2) = E(X(t_1)X(t_2))$ = autocorrelation function at t_1, t_2

$R_X(t_1, t_2, t_3) = E(X(t_1), X(t_2), X(t_3))$ = Triple correlation function at t_1, t_2, t_3, etc.

We can also define the auto-covariance function $C_X(t_1, t_2)$ of $X(t)$ given by

$$C_X(t_1, t_2) = E(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))$$
$$= R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

Parameters

- Expectation value of a random process

$$\eta_X(t) = E\{X(t)\} = \int_{-\infty}^{+\infty} x f_{X(t)}(x)dx$$

- Autocorrelation: expectation of the product of the RV at instants t_1 and t_2

$$R_X(t_1, t_2) = E\{X(t_1)X(t_2)\} = \iint_{R^2} x_1 x_2 f_{X(t_1)X(t_2)}(x_1, x_2) dx_1 dx_2$$

- Autocovariance: corresponding central moment

$$C_X(t_1, t_2) = E\{(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))\} = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

$$C_X(t_1, t_2) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

Gloria Menegaz

Parameters

- Mutual correlation

$$R_{XY}(t_1, t_2) = E\{X(t_1)Y(t_2)\} = \iint_{R^2} x_1 y_2 f_{X(t_1)Y(t_2)}(x_1, x_2)\, dx_1 dx_2$$

- Covariance

$$C_{XY}(t_1, t_2) = E\{(X(t_1) - \mu_X(t_1))(Y(t_2) - \mu_Y(t_2))\} = R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2)$$

$$\boxed{C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2)}$$

  – Warning

$$R_{XY} \neq R_{YX}$$
$$C_{XY} \neq C_{YX}$$

Uncorrelated processes

- If the covariance of two random processes X(t) and Y(t), evaluated at time instants t_1 and t_2, is zero for any t_1 and t_2, then the two processes are uncorrelated

$$C_{XY}\left(t_1,t_2\right)=0, \forall t_1,t_2 \rightarrow$$
$$R_{XY}\left(t_1,t_2\right)=\mu_X\left(t_1\right)\mu_Y\left(t_2\right), \forall t_1,t_2$$

Stationarity

A random process $X(t)$ is called <u>strict-sense stationary</u> if its probability structure is invariant with time. In terms of the joint distribution function

$$F_{X(t_1),X(t_2)\ldots X(t_n)}(x_1,x_2\ldots x_n) = F_{X(t_1+t_0),X(t_2+t_0)\ldots X(t_n+t_0)}(x_1,x_2\ldots x_n) \ \forall n \in N \text{ and } \forall t_0, t_n \in T$$

The pdfs which define the random process are *time-invariant*, i.e. they don't change when moving the origin of the time axis.
$X(t_1)$ **and** $X(t_1+\tau)$ **have the same probability distribution function**

For a stationary random process the following relations hold

$$f_{X(t)}(x) \rightarrow f_X(x)$$

$$f_{X(t_1)X(t_2)}(x_1,x_2) \rightarrow f_{X(t_1+\Delta t)X(t_2+\Delta t)}(x_1,x_2) \qquad \forall t_n, \Delta t$$

Stationary random processes

- The first order probability density function don't depend on time

- The second order pdfs only depend on the time delay $\tau = t_1 - t_2$

- Thus

  - The expectation value is independent on time
  - The covariance and autocorrelation depend on the *time lag*

A random process $X(t)$ is called **wide sense stationary process (WSS)** if

$$\mu_X(t) = \text{constant}$$

$$R_X(t_1, t_2) = R_X(t_1 - t_2) \text{ is a function of time lag.}$$

For a Gaussian random process, WSS implies strict sense stationarity, because this process is completely described by the mean and the autocorrelation functions.

The autocorrelation function $R_X(\tau) = EX(t + \tau)X(t)$ is a crucial quantity for a WSS process.

Autocorrelation funct of WSS processes

- The value of R in the origin ($\tau=0$) is the second order moment and corresponds to the maximum value

$$R_X(\tau) = E\{X(t)X(t+\tau)\}$$

$$R_X(0) = E\{X^2(t)\} = \mu_2$$

$$\boxed{\left|R_X(\tau)\right| \leq R_X(0) \qquad \forall\tau}$$

- "Symmetry"

$$R_{XY}(\tau) = E\{X(t)Y(t+\tau)\} \rightarrow$$

$$R_{XY}(-\tau) = E\{X(t)Y(t-\tau)\} \rightarrow$$

letting $t' = t - \tau \rightarrow t = t' + \tau \rightarrow$

$$R_{XY}(-\tau) = E\{X(t'+\tau)Y(t')\} = R_{YX}(\tau)$$

$$\boxed{\begin{aligned} R_{XY}(-\tau) &= R_{YX}(\tau) \\ R_X(-\tau) &= R_X(\tau) \end{aligned}}$$

$$f_{X(t+\tau)}(x) = f_{X(t)}(x) = f_X(x)$$

Autocorrelation funct of WSS processes

- The autocorrelation function of a wide sense stationary process is *symmetric about the origin*

- The width of the autocorrelation function is related to the correlation among the signal samples
  - If $R(\tau)$ drops quickly the samples are weakly correlated which means that they go through fast changes with time
  - Viceversa, $R(\tau)$ drops slowly the samples take "similar" values at close time instants, thus slow signal changes are expected
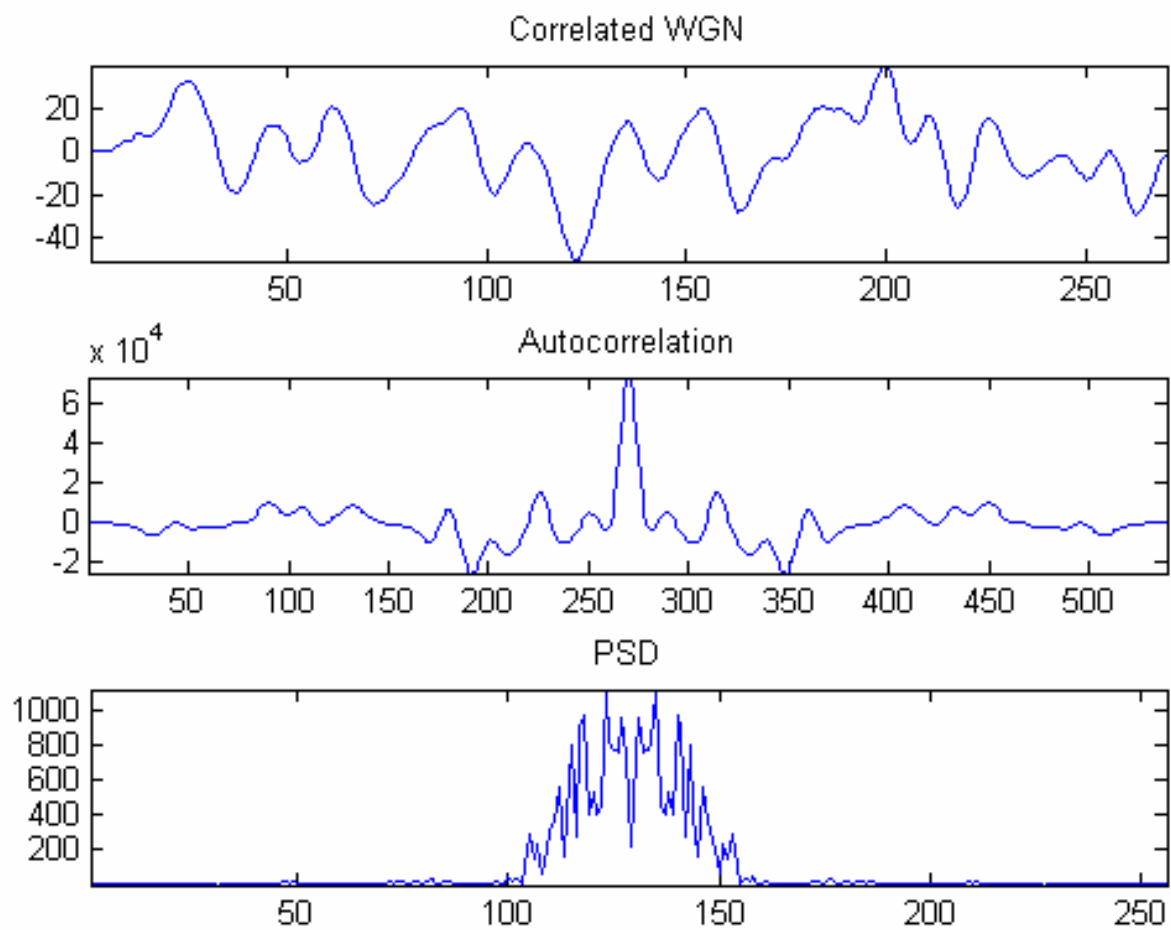 - $R(\tau)$ is related to the frequency content of the signal

Example: White Gaussian Noise
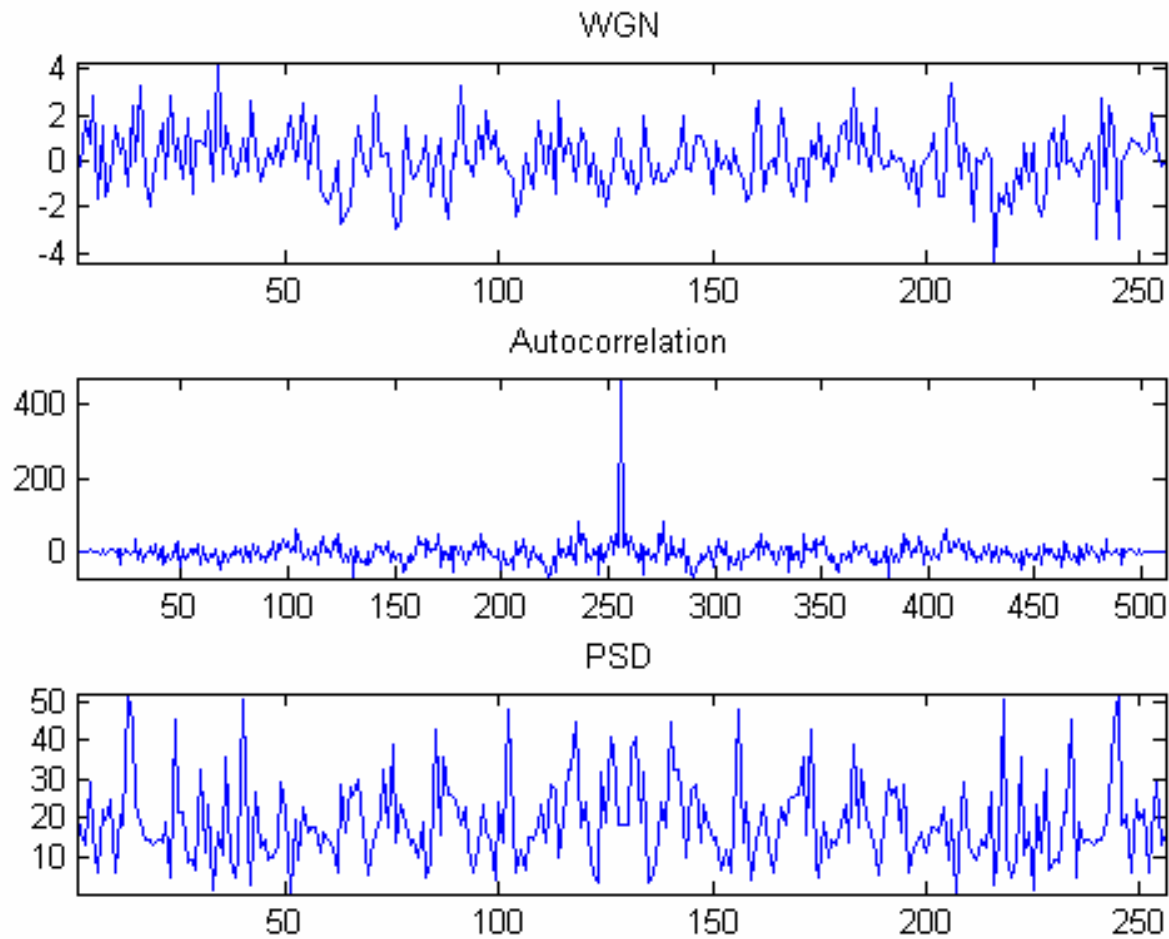
Example: Filtered White Gaussian Noise
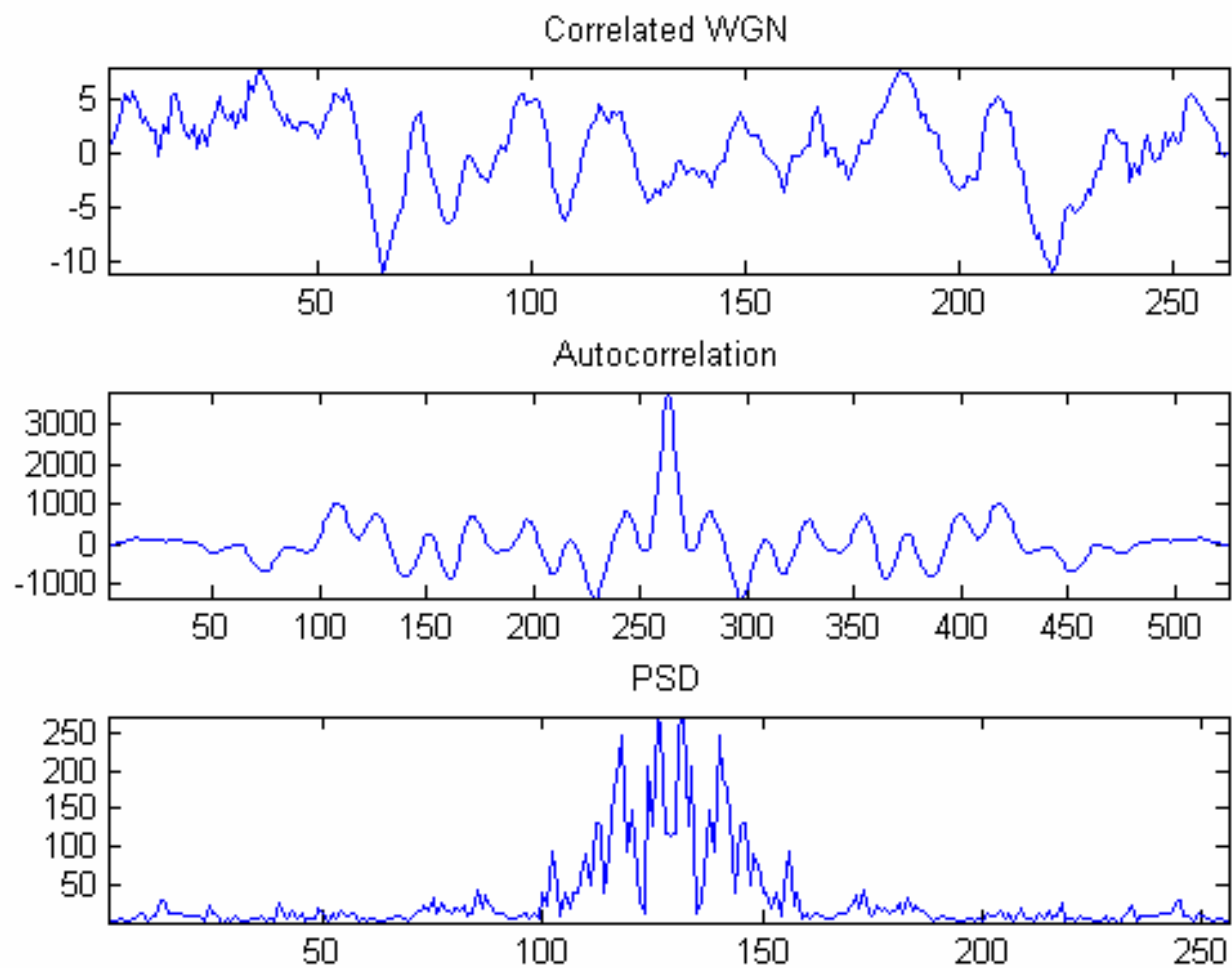
Example: Filtered White Gaussian Noise

Time averages

- From a practical point of view, it is preferable to deal with a single sequence rather than an infinite ensemble of sequences.

- When the pdfs are independent of time (e.g. for *stationary processes*), it is reasonable to expect that the amplitude distribution of a long sequence corresponding to a single realization should be approximately equal to the probability density

  – Similarly, the arithmetic average of a large number of samples of a single realization should be very close to the mean of the process.

- Time averages

  – NB: Such time averages are functions of an infinite set of RV, and thus are properly viewed as RV themselves!

Time averages of single realizations

$$m_x = \langle x(t) \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) \, dt$$

temporal mean

$$MSE_x = \langle x(t)^2 \rangle = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)^2 \, dt$$

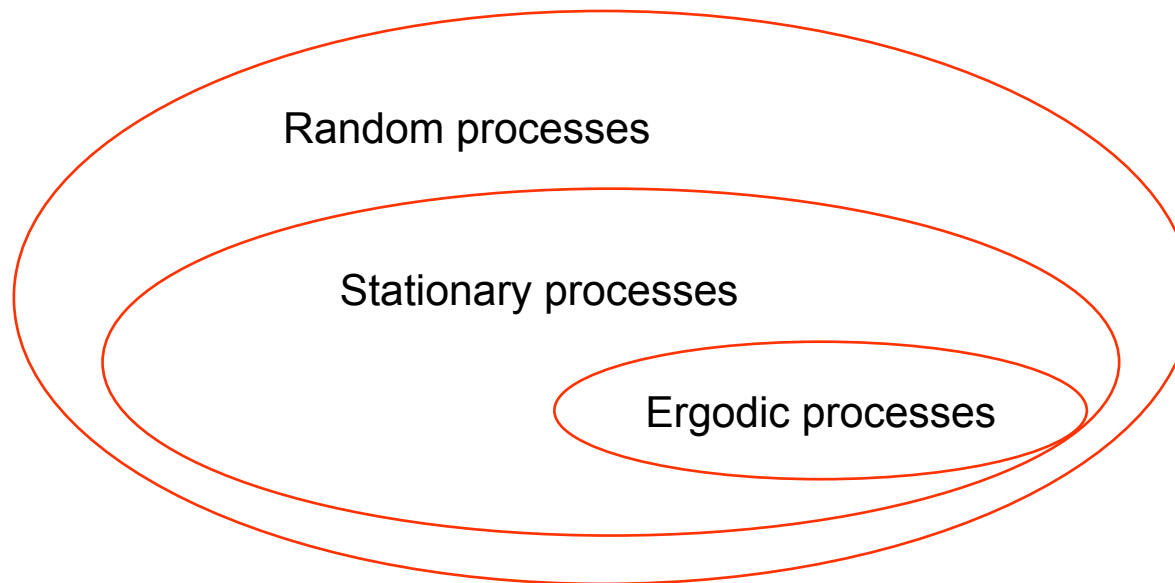mean square value

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

temporal autocorrelation

Gloria Menegaz

Ergodicity

- When _temporal averages_ _are equal with probability 1_ (namely for almost all the realizations) _to the corresponding ensemble averages_ the process is said to be ergodic

Random processes

Stationary processes

Ergodic processes

Ergodicity

- Ergodicity implies stationarity
  - Otherwise the ensemble averages would depend on time, which contradicts the hypothesis

- Temporal averages are the same for almost all the realizations
  - So that we can talk about "temporal average"

- <u>Temporal and ensemble averages are the same</u>

- For ergodic processes, a single realization is sufficient to completely characterize the process!

$$\langle X(t) \rangle = \lim_{T \to \infty} \frac{1}{2L+1} \int_{-T}^{T} x(t)dt = E\{X(t)\} = \mu_X$$

$$\langle X(t)X^*(t+\tau) \rangle = \lim_{T \to \infty} \frac{1}{2T} = \int_{-T}^{T} X(t)X^*(t+\tau)dt = E\{X(t)X^*(t+\tau)\}$$

Gloria Menegaz

Discrete time formulation

Gloria Menegaz

Discrete *time* formulation for RP

- The temporal axis is sampled and the integer valued index n is used. All the rest remains the same
  - It's only a matter of using different notations and replacing integrals in time domain with discrete summations

- A sequence *{x[n]}* is considered one of an ensemble of sample sequences

- A *random process* is an indexed set of *random variables* X_n
  - The family of random variables is characterized by a set of probability distribution functions that in general may be functions of the index n (unless it is stationary)
  - In the case of <u>*discrete time signals*</u>, the index n is associated to the discrete time variable
  - An individual RV X_n is described by the probability distribution function

$$P_{X_n}\left(x_n, n\right) = \mathrm{Pr}\left(X_n \le x_n, n\right)$$

(1)

where X_n denotes the RV and x_n is a particular value.
  - The probability density function is obtained from (1) by differentiation and represents the probability of the RV to be in the infinitesimal interval dX_n around x_n

Discrete time random processes

- Each variable X_n is a random variable. The values it takes over the different realizations of the corresponding process are its observations

- Ensemble averages
  - Since a random process is an indexed set of RV, it can be characterized by statistical averages of the RV comprising the process (over the different realizations). Such averages are called *ensemble averages*.

- Definitions
  - Average, or mean

$$m_{X_n} = \mathrm{E}\{X_n\} = \int_{-\infty}^{\infty} x p_{X_n}(x_n, n) dx \qquad p_{X_n}(x_n, n) = \frac{\partial P_{X_n}(X_n, n)}{\partial X_n}$$

  - where E denotes the expectation.
  - In general, the expected value (mean) depends on n

Discrete time random processes

– Mean square value (average power)

$$rms[X_n] = \mathrm{E}\left\{|X_n|^2\right\} = \int_{-\infty}^{\infty} |x|^2 \; p_{X_n}(x_n, n)dx$$

– Variance

$$\sigma^2_{X_n} = \mathrm{var}[X_n] = \mathrm{E}\left\{|(X_n - m_{X_n})|^2\right\} = \int_{-\infty}^{\infty} |x_n|^2 \; p_{X_n}(x_n, n)dx$$

$$= E\left\{|X_n|^2\right\} - |m_{X_n}|^2 = \sigma^2_{X_n}$$

– In general, the mean and the variance are functions of time (index n), while they are constant for stationary processes
– The absolute value has been introduced to allow dealing with complex random processes (X_n and Y_n are real random processes)

$$W_n = X_n + jY_n$$

Discrete time random processes

- Autocorrelation sequence

$$\varphi_{XX}[n,m] = \mathrm{E}\left\{X_n X_m^{*}\right\} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} X_n X_m^{*} \, p_{X_n,X_m}(x_n,n,x_m,m)\,dx_n dx_m$$

  – Samples of X_n and X_m are taken on different realizations

- Autocovariance sequence

$$\gamma_{XX}[n,m] = \mathrm{E}\left\{(X_n - m_{X_n})(X_m - m_{X_m})^{*}\right\} = \varphi_{XX}[n,m] - m_{X_n} m_{X_m}^{*}$$

- Cross-correlations and cross-covariance are obtained in the case the same quantities are evaluated between two different processes (ex. X_n and Y_m)

Uncorrelation and Independence

- In general, the average of the product of 2 RV is not equal to the product of the averages. If this is the case, the RV are said to be *uncorrelated*

(2)

$$\gamma_{XX}(m,n) = 0 \rightarrow \quad E\{X_n X_m\} = E\{X_n\}E\{X_m\} = m^2_{X_n}$$

$$\gamma_{XY}(m,n) = 0 \rightarrow \quad E\{X_n Y_m\} = E\{X_n\}E\{Y_m\} = m_{X_n} m_{Y_m}$$

- Statistically independent processes

$$(3) \quad p_{X_n Y_m}(X_n, n, Y_m, m) = p_{X_n}(X_n, n) p_{Y_m}(Y_m, m)$$

  – Condition (3) is stronger than condition (2): statistically independent processes are also uncorrelated, but NOT viceversa.

Stationary random processes

- A stationary random process is characterized by an equilibrium condition in which the statistical properties are invariant to a shift in the time origin. Accordingly
  - The first-order probability distribution is independent of time
    - The pdf is the same for all n
  - The joint probability distributions are also invariant to a shift in the time origin
  - The first order averages, like the mean and variance, are independent of time
  - The second order averages, like the autocorrelation, depend on the *time difference* (m-n)
  - Slightly different notations: $X_n \rightarrow X[n]$

(1) $$\mu_{X_n} = E\{X[n]\} = \mu$$ Independent of n

$$\sigma_{X_n}^2 = E\left\{|(X[n] - \mu_{X_n})|^2\right\} = E\left\{|(X[n] - \mu)|^2\right\} = \sigma^2$$

$$\varphi_{XX}[n, n+m] = E\left\{X[n]X^*[m]\right\} = \varphi_{XX}[m]$$ Dependent on the time shift m

Stationary random processes

- *Strict* stationarity: the full probabilistic description is time invariant

- *Wide-sense* stationarity: the probability distributions are not time-invariant but the relations (1) still hold
  - In particular, relations (1) show 2-nd order stationarity

- Linear operations preserve wide-sense stationarity
  - Filtering by a linear time invariant system (LTIS) conserves wide-sense stationariety

Time averages

- From a practical point of view, it is preferable to deal with a single sequence rather than an infinite ensemble of sequences. When the pdfs are independent of time (e.g. for stationary processes), it is reasonable to expect that the amplitude distribution of a long sequence corresponding to a single realization should be approximately equal to the probability density

  – Similarly, the arithmetic average of a large number of samples of a single realization should be very close to the mean of the process.

- Time averages

$$\langle x_n \rangle = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x_n \qquad \text{Time average of a random process}$$

$$\langle x_{n+m} x_m^* \rangle = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x_{n+m} x_m^* \qquad \text{Autocorrelation sequence}$$

  – NB: Such time averages are functions of an infinite set of RV, and thus are properly viewed as RV themselves!

Ergodicity

- For an ergodic process, time averages coincide with ensemble averages
  - That is, for a single realization (sequence {x[n]})

$$\left\langle x[n] \right\rangle = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x[n] = E\left\{ X_n \right\} = \mu_X$$

$$\left\langle x[n+m]x[n]^* \right\rangle = \lim_{L \to \infty} \frac{1}{2L+1} \sum_{n=-L}^{L} x_{n+m} x_m^* = E\left\{ X_{n+m} X_m^* \right\} = \varphi_{XX}[m] = R[m]$$

  - Sample means and variances are estimates of the corresponding RV, and as such are corrupted by estimation errors.

$$\hat{m}_x = \frac{1}{L} \sum_{n=0}^{L-1} x[n]$$

$$\hat{\sigma}_x^{\ 2} = \frac{1}{L} \sum_{n=0}^{L-1} |x[n] - \hat{m}_x|^2$$

Ergodic random processes

- We don't need to keep the index n for X_n and we can abbreviate it with X

- Let's consider a zero-mean wide-sense stationary random process
  - The autocorrelation and the autocovariance coincide

Covariance matrix

- Given a sequence of RV, X_1, X_2,…,X_n, we can calculate the covariance between any couple of them, and organize the results in a matrix
  - The sequence of RV represent the observations at given time instants

- Referring to the continuous time case
  - The formalization generalizes to the discrete time by replacing $t_n \rightarrow n$

$$C_{XX}(t_1, t_2) = E\{(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))\} = R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

$$C_X(t_1, t_1) = E\{(X(t_1) - \mu_X(t_1))^2\} = C_{1,1} = \sigma_{X_1}^2 = \sigma_1^2$$

$$C_X(t_1, t_2) = E\{(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))\} = C_{1,2} = \sigma_{X_1 X_2} = \sigma_{1,2}$$

$$C_X(t_1, t_3) = E\{(X(t_1) - \mu_X(t_3))(X(t_3) - \mu_X(t_3))\} = C_{1,3} = \sigma_{X_1 X_3} = \sigma_{1,3}$$

…

$$C_X(t_1, t_n) = E\{(X(t_1) - \mu_X(t_1))(X(t_n) - \mu_X(t_n))\} = C_{1,n} = \sigma_{X_1 X_n} = \sigma_{1,n}$$
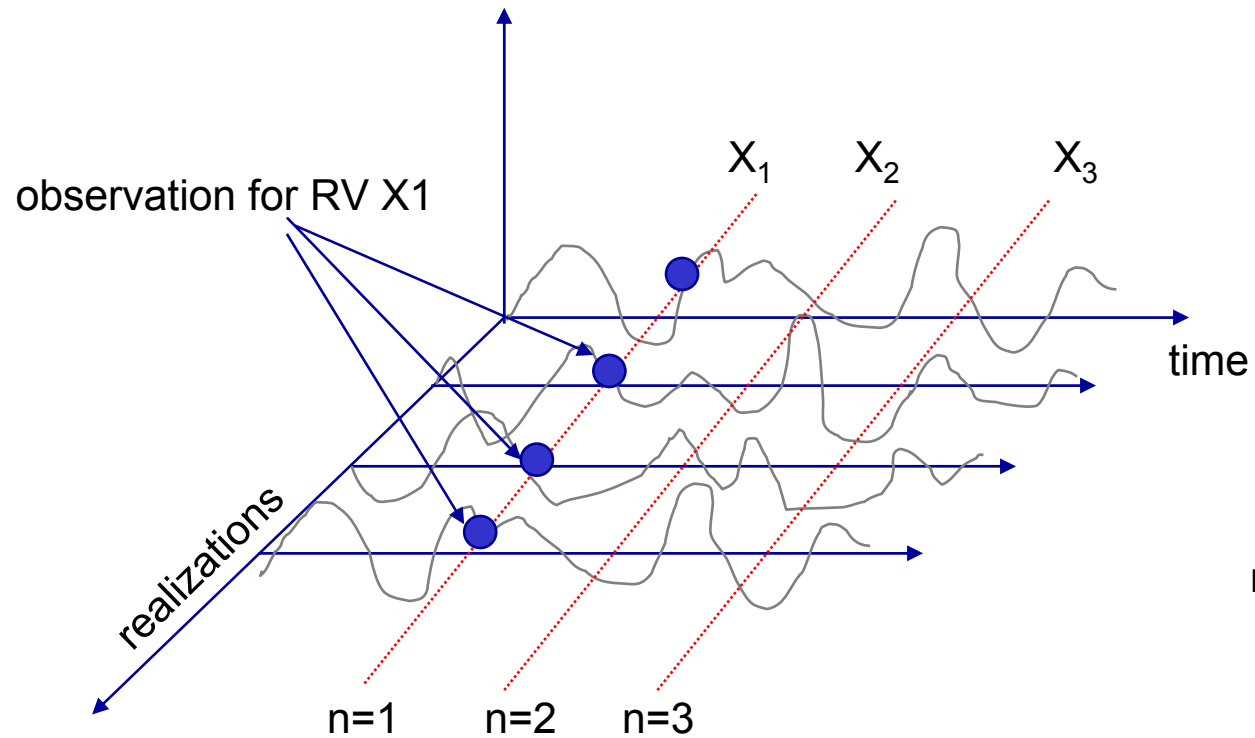
Covariance matrix

- These data can be put in matrix form

$$C = \begin{bmatrix} C_1 & C_{12} & \cdots & C_{1n} \\ C_{21} & C_2 & \cdots & C_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ C_{n1} & C_{n2} & \cdots & C_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

- The matrix element at position (n,m) represents the covariance between the RV X_n and X_m. If the two RV are uncorrelated, the element is null. THUS

- *The covariance matrix of uncorrelated RV is diagonal*

Covariance matrix

observation for RV X1

X_1 X_2 X_3

time

realizations

n=1   n=2   n=3

Vectors of RV can be built by gathering the RV in a column vector

random variables

$$\vec{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix}$$

Covariance matrix

- Each RV X_n corresponds to k observations that can be put in vector form as well

$$X_1 = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \end{bmatrix}$$

- Then, the covariance matrix can be written as

$$C_X = E\left\{ \left(\vec{X} - \vec{\mu}_X \right) \cdot \left(\vec{X} - \vec{\mu}_X \right)^T \right\}$$

$$\vec{\mu}_X = \begin{bmatrix} \mu_1 \\ \cdots \\ \mu_n \end{bmatrix}$$

Covariance matrix

- Proof

$$C_X = E\left\{\left(\vec{X} - \vec{\mu}_X\right) \cdot \left(\vec{X} - \vec{\mu}_X\right)^T\right\} =$$

$$= E\left\{\begin{bmatrix} \left(X_1 - \mu_1\right) \\ \left(X_2 - \mu_2\right) \\ \cdots \\ \left(X_n - \mu_n\right) \end{bmatrix} \cdot \begin{bmatrix} \left(X_1 - \mu_1\right) & \left(X_2 - \mu_2\right) & \cdots & \left(X_n - \mu_n\right) \end{bmatrix}\right\} =$$

$$= E\left\{\begin{bmatrix} \left(X_1 - \mu_1\right)^2 & \cdots & \left(X_1 - \mu_1\right)\left(X_n - \mu_n\right) \\ \cdots & \cdots & \cdots \\ \left(X_n - \mu_n\right)\left(X_1 - \mu_1\right) & \cdots & \left(X_n - \mu_n\right)^2 \end{bmatrix}\right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

Covariance matrix for WSS processes

- For wide sense stationary processes

$$R_X\left(-\tau\right) = R_X\left(\tau\right)$$
$$C_X\left(-\tau\right) = C_X\left(\tau\right)$$

- Thus the covariance matrix is symmetric about the diagonal

$$C = \begin{bmatrix} C_1 & C_{12} & \cdots & C_{1n} \\ C_{12} & C_2 & \cdots & C_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ C_{1n} & C_{2n} & \cdots & C_n \end{bmatrix} = \begin{bmatrix} \sigma_1^{\,2} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^{\,2} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^{\,2} \end{bmatrix}$$

Gaussian Random Process

For any positive integer n, $X(t_1), X(t_2),.....X(t_n)$ represent n jointly random variables. These n random variables define a random vector $\mathbf{X} = [X(t_1), X(t_2),.....X(t_n)]'$. The process $X(t)$ is called Gaussian if the random vector $[X(t_1), X(t_2),.....X(t_n)]'$ is jointly Gaussian with the joint density function given by
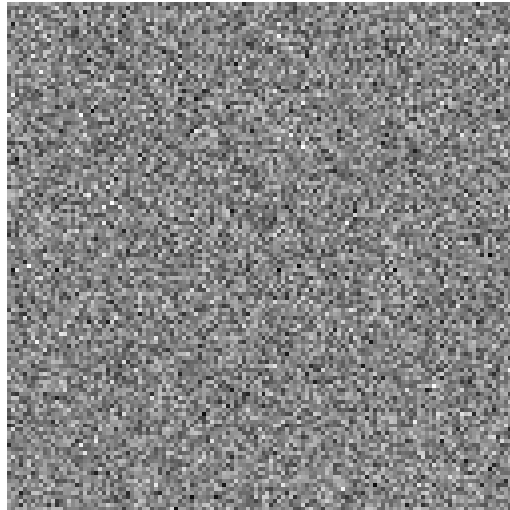
$$f_{X(t_1),X(t_2)...X(t_n)}(x_1, x_2,...x_n) = \frac{e^{-\frac{1}{2}\mathbf{x}'\mathbf{c}_X^{-1}\mathbf{x}}}{\left(\sqrt{2\pi}\right)^n \sqrt{\det(\mathbf{C_X})}} \text{ where } \boxed{\mathbf{C_X} = E(\mathbf{X}-\mathbf{\mu_X})(\mathbf{X}-\mathbf{\mu_X})'}$$

and $\mathbf{\mu_X} = E(\mathbf{X}) = [E(X_1), E(X_2)......E(X_n)]'$.

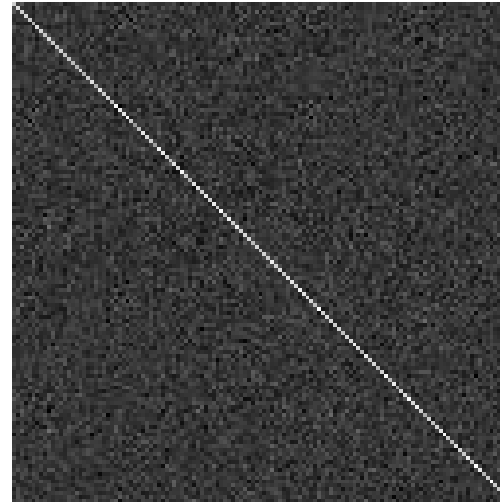C_X: *covariance matrix*

Gloria Menegaz

Gaussian random processes

Gaussian process

Covariance matrix

The columns of the matrix
are iid RV

The matrix is symmetric
The elements out of the diagonal
are close to zero

Covariance matrix: properties

- The covariance matrix is symmetric and nonnegative definite

- The elements along the principal diagonal are the variances of the elements of the random vector

- The elements out of the principal diagonal are the correlation coefficients between couples of elements

- Uncorrelated vector elements correspond to a diagonal covariance matrix

➢ *Is it possible to define a linear transformation mapping the RP X to the RP such that the RP Y has a covariance matrix in diagonal form?*

Gloria Menegaz

Karunen-Loeve transform

- The KLT is a linear transform that maps the random process X to a random process Y whose covariance matrix is diagonal ↔ whose components are uncorrelated
  - If X is a generalized Gaussian, then the components of Y are independent
    - For Gaussian processes, uncorrelation is necessary and sufficient for independence

- Given a wide sense stationary process X, with covariance matrix C_X, we look for a linear transform T such that $Y=T^TX$ such that C_Y is diagonal

- It can be proved that T consists of the eigenvectors of C_X

$$C_X \phi_k = \lambda_k \phi_k \qquad k = 0,...,N-1$$

$$\Phi = \begin{bmatrix} \phi_0 & \phi_1 & \cdots & \phi_{N-1} \end{bmatrix} \qquad \text{eigenvectors matrix (eigenvectors are the columns)}$$

$$\Lambda = diag\left(\lambda_0, \lambda_1,...,\lambda_{N-1}\right) \qquad \text{(diagonal) eigenvalues matrix}$$

Properties

- The eigenvector matrix is square (NxN), unitary and orthogonal
  - The eigenvectors form an orthonormal basis

$$\Phi^T \Phi = I$$

$$\Phi^{-1} = \Phi^T$$

$$\langle \phi_i, \phi_j \rangle = \phi_i^T \phi_j = \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Projection on a basis

$$\vec{y} = \begin{bmatrix} y_0 \\ y_1 \\ ... \\ y_{N-1} \end{bmatrix} = A\vec{x} = \begin{pmatrix} a_{0,0} & \cdots & a_{0,N-1} \\ \vdots & \ddots & \vdots \\ a_{N-1,0} & \cdots & a_{N-1,N-1} \end{pmatrix} \begin{bmatrix} x_0 \\ x_1 \\ ... \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} \vec{a}_0 \\ \vec{a}_1 \\ ... \\ \vec{a}_{N-1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ ... \\ x_{N-1} \end{bmatrix}$$

$$y_i = \langle \vec{a}_i, \vec{x} \rangle = \sum_{k=0}^{N-1} a_{i,k} x_k$$

KLT

- Projection on the eigenvector basis

$$\vec{y} = \begin{bmatrix} y_0 \\ y_1 \\ ... \\ y_{N-1} \end{bmatrix} = \Phi^T \vec{x} == \begin{bmatrix} \phi_0^T \\ \phi_1^T \\ ... \\ \phi_{N-1}^T \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ ... \\ x_{N-1} \end{bmatrix}$$

$$y_i = \left\langle \phi^T{}_i, \vec{x} \right\rangle$$

Analysis

$$\boxed{\vec{y} = \Phi^T \vec{x}}$$

$$\Phi^{-1} = \Phi^T \rightarrow \vec{x} = \left(\Phi^T \right)^{-1} \vec{y} = \Phi \vec{y}$$

Synthesis

$$\boxed{\vec{x} = \Phi \vec{y}}$$

KLT

- The KLT diagonalizes the covariance matrix $\quad C_Y = \Phi^T C_X \Phi = \Lambda$

$(X-M_x)^T \Sigma_x^{-1}(X-M_x)=C_1$

$Y = \emptyset^T X$

KLT

$(Y-M_y)^T \Sigma_y^{-1}(Y-M_y)=C_1$
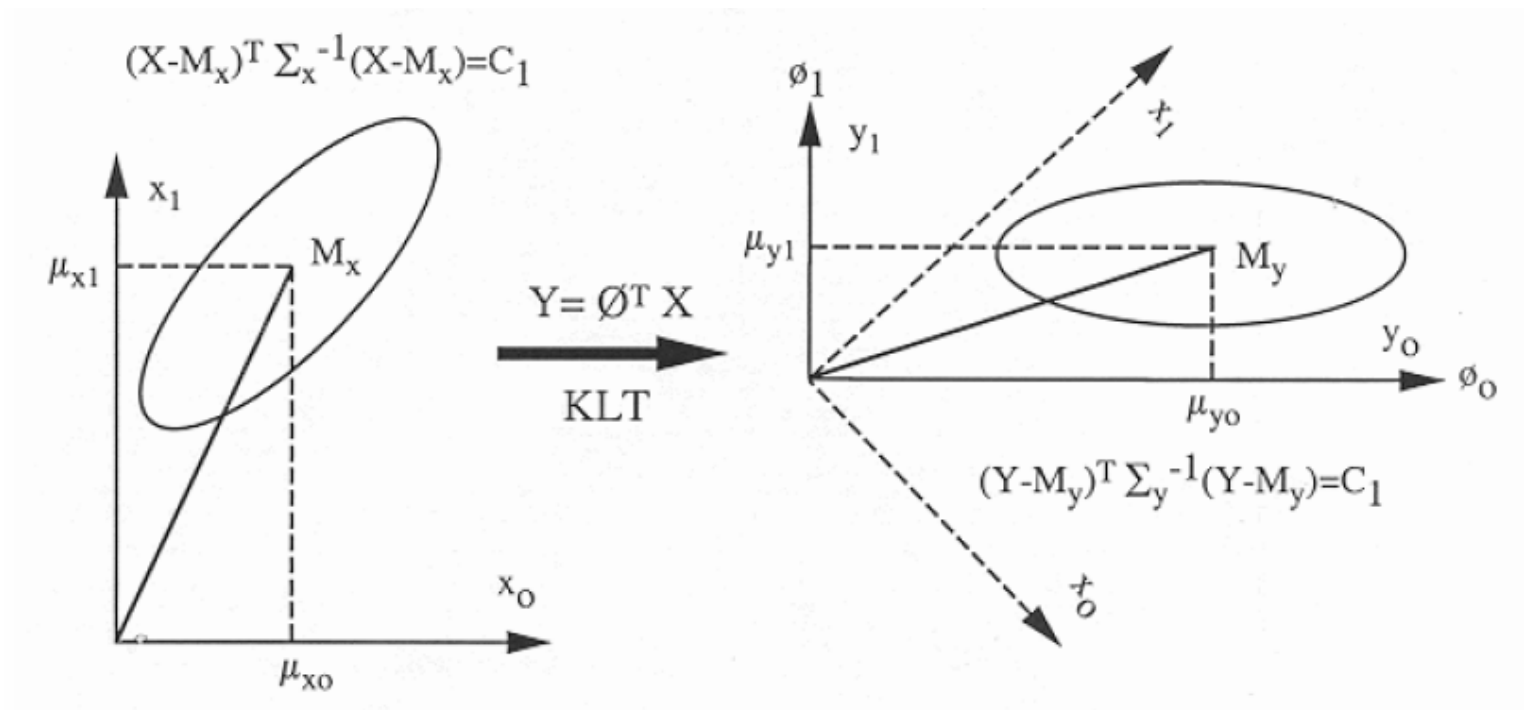
Principal Component Analysis (PCA)

- **Principal component analysis** (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

- This is derived from the KLT
  - Due to its properties of signal decorrelation the KLT can be used for compression by reducing the size of the dataset

Gloria Menegaz

PCA

- Algorithm
    1. Find the mean vector and the covariance Cx
    2. Find the eigenvalues (λi, i=0,..,N-1) and sort them in descending order, and sort the eigenvectors φi, i=0,..,N-1 accordingly
    3. Choose a lower dimensionality m<N (following an energy-based criterion)
    4. Construct an Nxm transform matrix composed by the m eigenvectors corresponding to the largest eigenvalues

$$\Phi_m = \begin{bmatrix} \phi_0 & ... & \phi_{m-1} \end{bmatrix} \quad \text{basis vectors}$$

$$\vec{y} = \Phi^T_{\ m} \vec{x} \qquad \text{Analysis}$$

$$\vec{x} = \Phi_m \vec{y} \qquad \text{Synthesis}$$

$$\downarrow$$

Nxm

Example
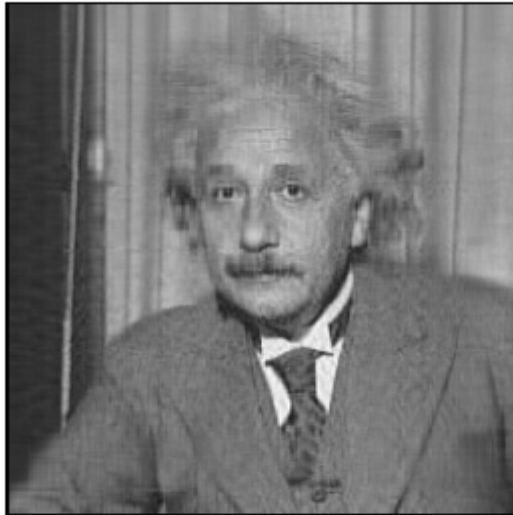
m=32   m=16
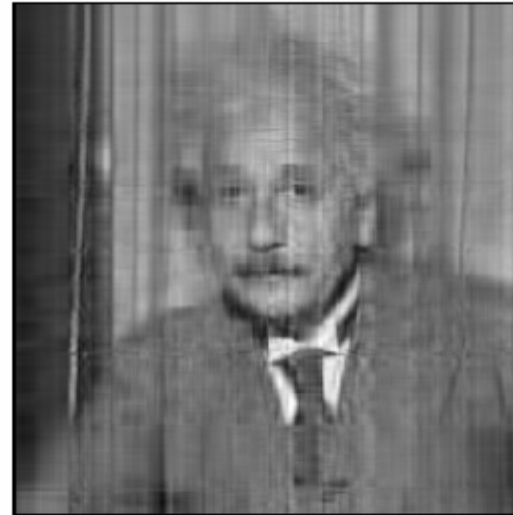
Example

m=8

Karunen-Loeve Transform

$$Y = T^T X$$

$$C_Y = T^T C_X T = \Lambda \text{ diagonal matrix of the eigenvalues}$$

$$C_X v_i = \lambda_i v_i \quad \text{eigenvector equation for } C_X$$

v_i eigenvector associated to the eigenvalue λ_i

$$T = \begin{bmatrix} v_0 & v_1 & \cdots & v_{N-1} \end{bmatrix} \quad \text{The columns are the eigenvectors}$$

The matrix T transforms X into Y whose covariance matrix is diagonal with elements $\lambda_{ii} = \text{var}[y_i]$

Properties of the KLT

- It is optimal for Gaussian sources
  - namely it minimizes the MSE between the vector and its approximation when only k out of K transform coefficients are retained

- It basically removes the redundancy in the input vector allowing better compression performance

- The KLT transforms a Gaussian random vector to a Gaussian random vector with statistically independent components.
  - If the vector is not Gaussian, the components of **Y** will be uncorrelated but not independent

- Under some conditions, it is well approximated by the DCT, which in addition allows fast algorithms for its implementation
  - JPEG

What about ergodicity?

- The hypothesis of ergodicity (which encloses stationarity) is often assumed in applications. This is because it allows to *focus on the single realization* to estimate the probability density function (and its parameters) of a random process

- What does this mean?
  - For 1D signals (ECG, EEG): the measurements correspond to the realizations, and thus are used to study the signals through the estimation of the stochastic parameters
  - However, for 1D signals many realizations of a given process are often available

Images

X_1

X_k
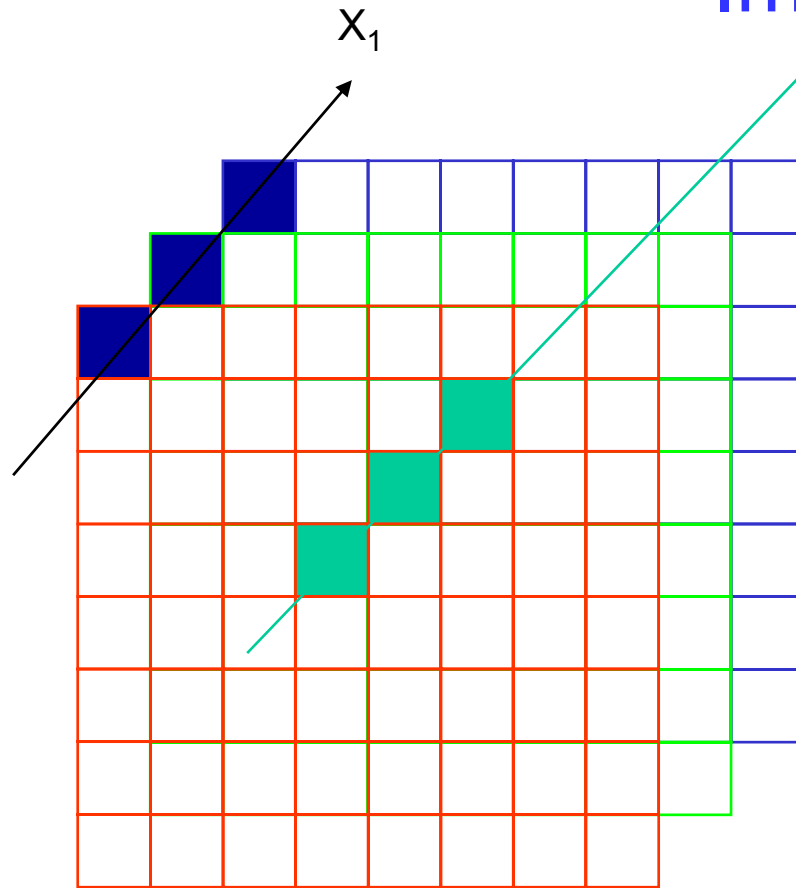
- Each image is the realization of a 2D random process X

- The process consists of NxM RV

- The observations of each RV "run orthogonally to the image plan", that is, gather the pixel values at position (n,m) in the set of images

- Ensemble averages should be evaluated on such RV

- Assuming stationarity and ergodicity facilitates the task by allowing to perform all the computations on the single image

Images

- **Stationarity**
  - Averages are assumed to be equal
  - What matters is the *distance* among the pixels: $C_X(d)$ is the same irrespectively of the direction

- **Ergodicity**
  - All the calculations are performed locally: instead of looking at different realizations, the different moments are calculated on the image

- **Simplification**
  - Columns represent the RV
  - Rows represents the realizations

$$R_X(-\tau) = R_X(\tau) \rightarrow \quad R_X(-d) = R_X(d)$$
$$C_X(-\tau) = C_X(\tau) \rightarrow \quad C_X(-d) = C_X(d)$$

$X_1 \qquad X_k$

d
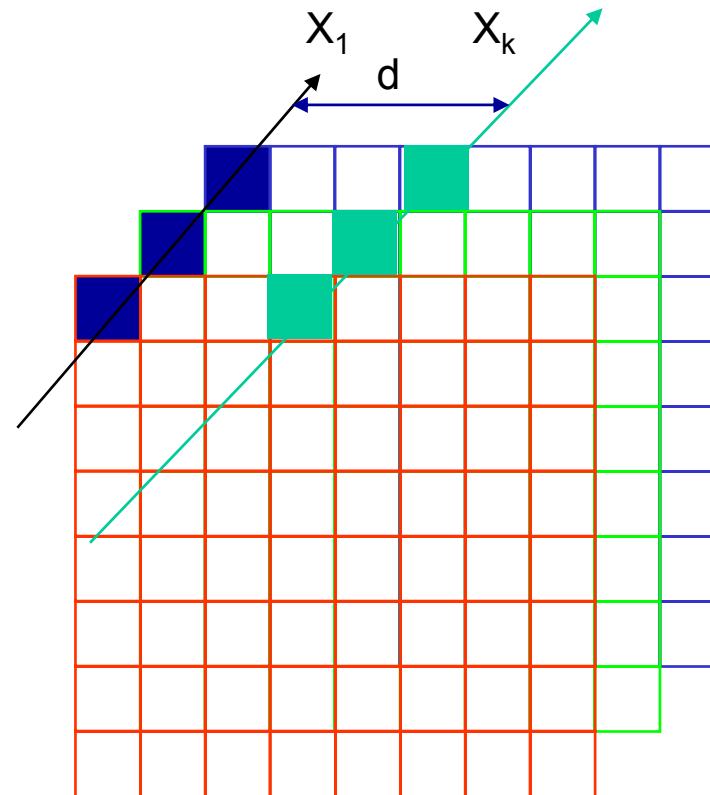
Images
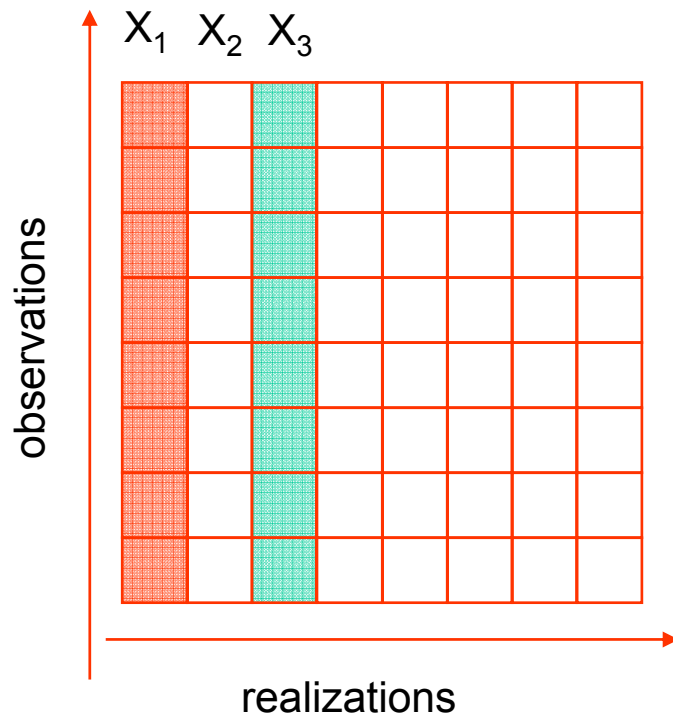
- The covariance matrix is symmetric about the major diagonal

- Covariances and correlations are evaluated between columns

- **Limit**: the stationarity and ergodicity assumptions are asymptotic: they assume that the number of realizations (k) and the size of each realization (n in 1D, NxM in 2D) tend to infinity. When dealing with signals of finite size the hypothesis are not satisfied and the estimations are "poor"

Gloria Menegaz

Example
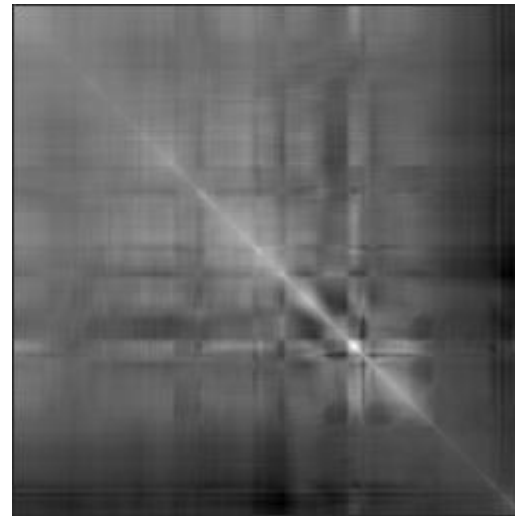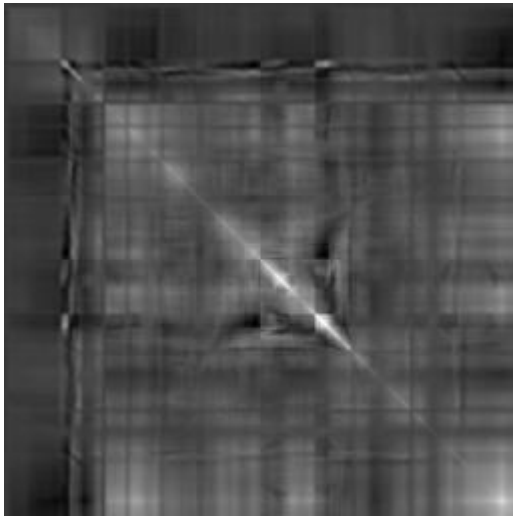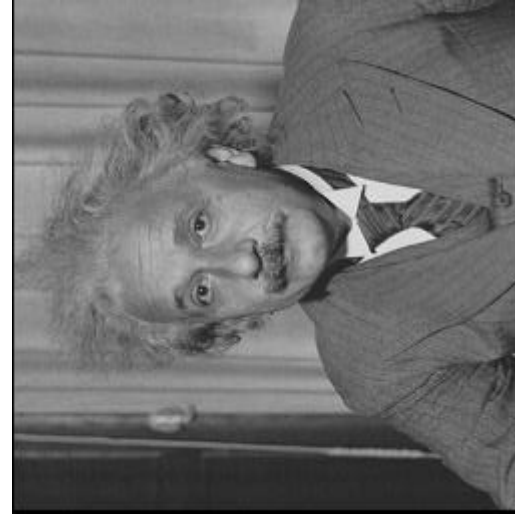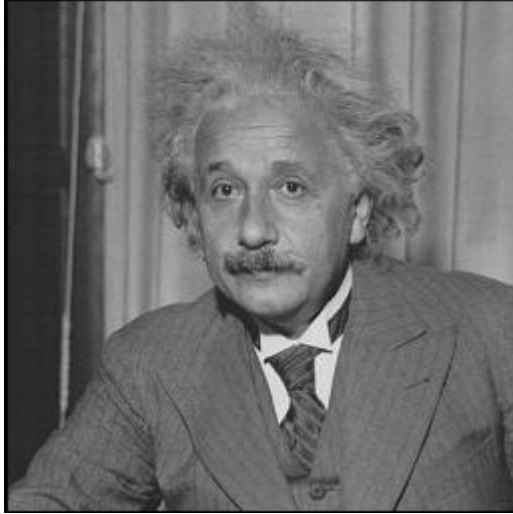
Other solution

Consider the image as a set of subimages

X_1

X_3

X_2
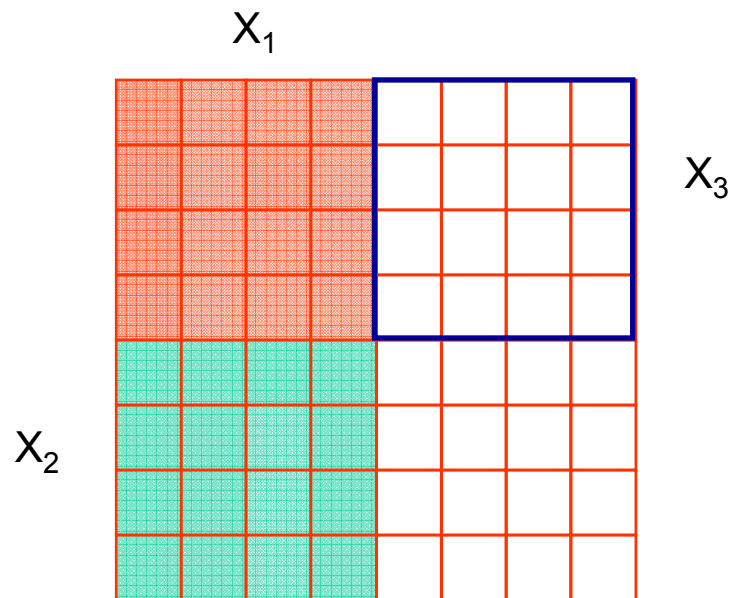
Assuming that the stationarity holds also locally, each subimage is considered as a realization. The covariance matrix is calculated on subimages.

X_1 X_2