Intelligent Signal Processing

Statistical Signal Processing

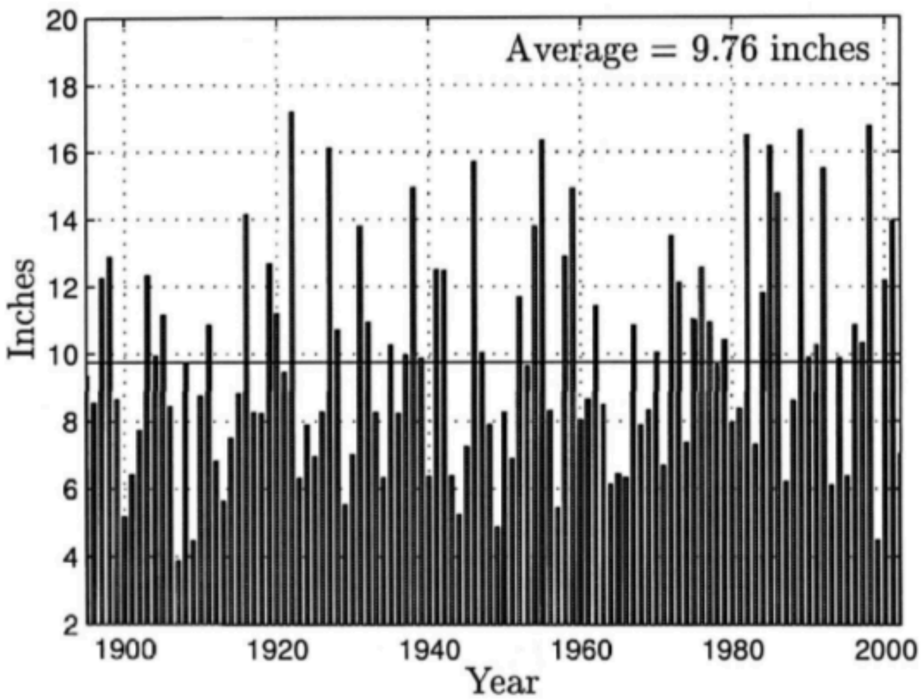
Angelo Ciaramella

Probability

- Ingredients of probability
  - Random experiment
  - A set of outcomes
  - The probabilities associated to these outcomes
  - We cannot predict with certainty the outcome of the experiment
    - We can predict «averages»!

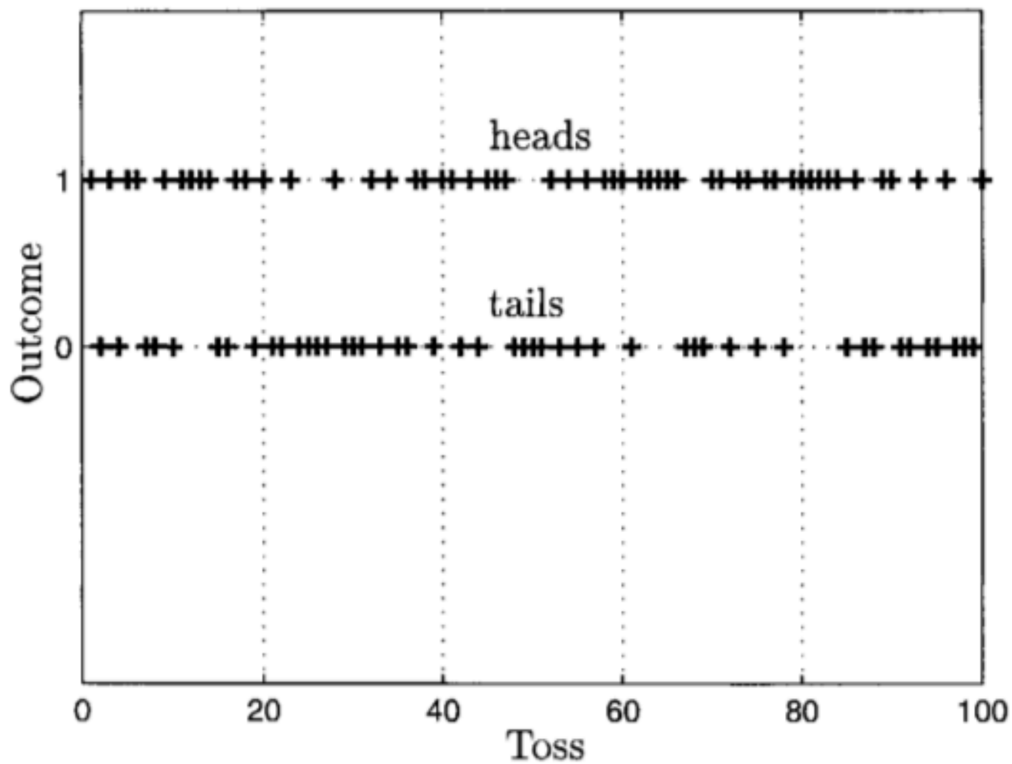
- Philosophical aspects of probability
  - We want a probabilistic description of the physical problem
  - We believe that there’s a statistical regularity that describes the physical phenomenon

Probability



Cannot predict how much rain, but the average suggests not to plant in Arizona

Probability

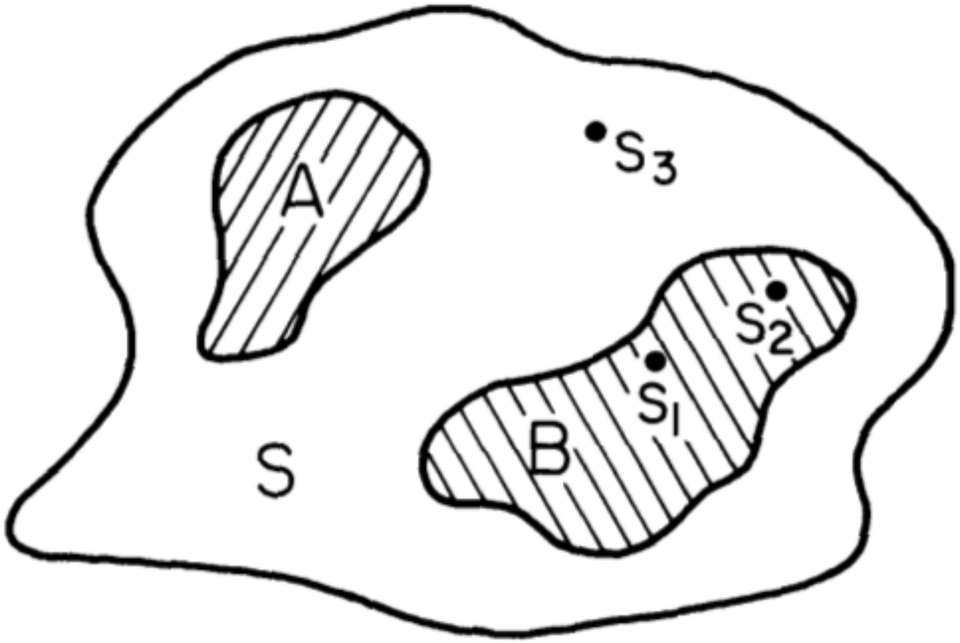


Result of tossing a coin is not predictable, but the average 53% tells me it is fair coin

Sample space

- Sample space
  - Experiment
    - process of observing the state at $t = t_0$
  - Sample point
    - outcome of the experiment
  - Sample space
    - set S of all possible sample points
  - Event
    - event A in S that occurs (happens)

Sample space



s_1, s_2 and s_3: sample points

A, B: events

S: sample space

Probability space

- Probability space
  - The sample space S is a probability space i to every event A there is a number P(A) that fulls
    - $0 \leq P(A) \leq 1$
    - $P(A \cup B) = P(A) + P(B), \quad iff\ A \cap B = \emptyset$
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - $P(S) = 1$

Conditional probability

- Conditional probability

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

- Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{likelihood \times prior}{marginal\ likelihood}$$

Bayes' theorem

- Forward problem
  - Given a specied number of white and black balls in a box, what is the probability of drawing a black ball?
- Reverse problem
  - Given that one or more balls have been drawn, what can be said about the number of white and black balls in the box?

Bayes' theorem

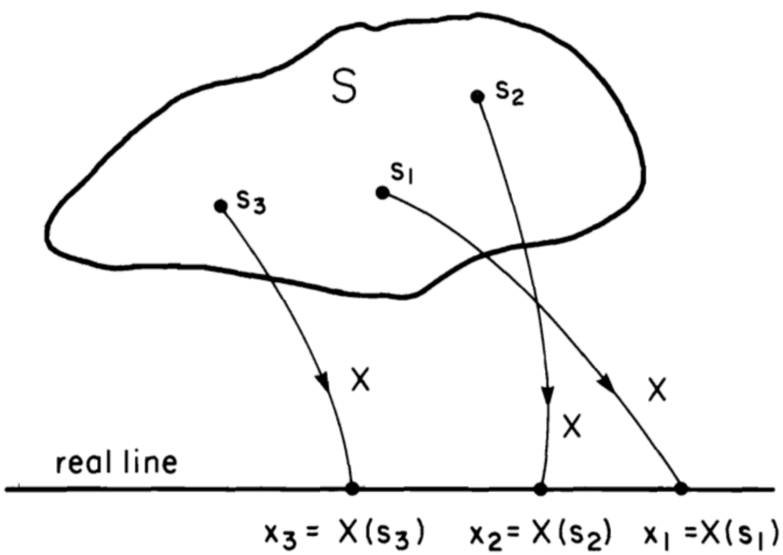
- Example
  - *The department is formed by 60% men and 40% women. Men always wear trousers, women wear trousers or skirts in equal numbers. Which is the probability of meeting a girl with trousers is?*

  - A - I see a girl
  - B - A person is wearing trousers
  - The *probability is*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.5 \times 0.4}{0.5\times0.4 + 1\times0.6} = 0.25$$

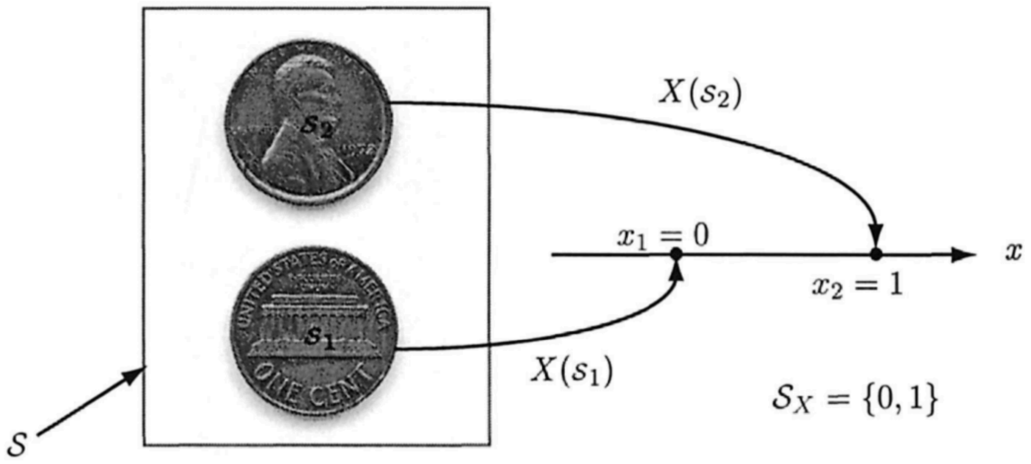
Random variable

- Random variable
  - A *random variable* is a real-valued *function* $X(.)$ of sample points in a sample space:
    - a function that *assigns a real number* $x = X(s)$ *to each sample point* s. The real number x is called *realization*, or *statistical sample* of $X()$



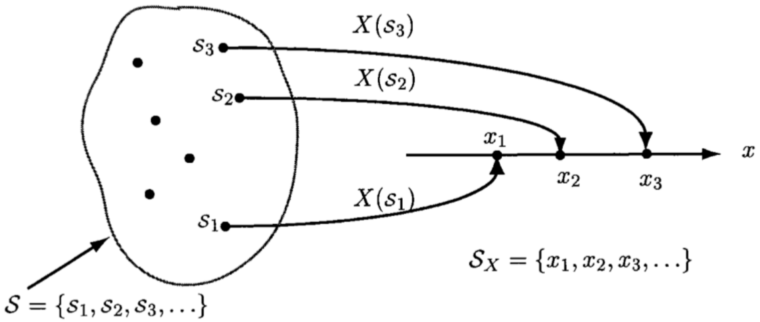
Representation of a random variable

Random variable

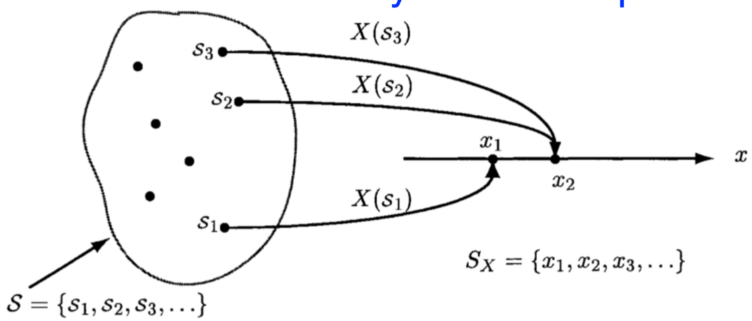


Discrete random variables

One-to-one map



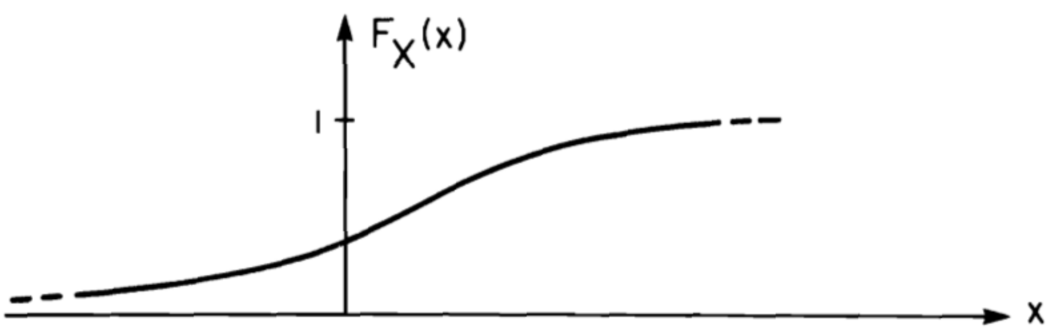
Many-to-one map



Distribution function

- Distribution function (DF) or Cumulative Density Function (CDF)
  - The probability distribution function for a random variable X is denoted by $F_X(.)$

$$F_X(x) = \text{Prob}\{X < x\}$$



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

Distribution function

- Probability Density Function (PDF)
  - Non-negative function

$$f_X(x) \geq 0$$

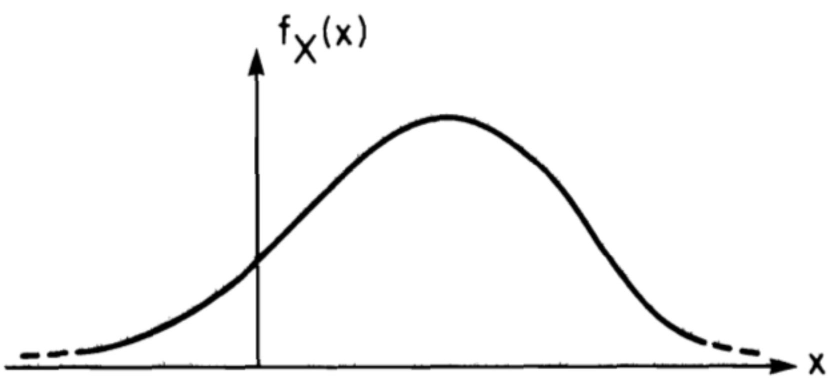
  - Unit area

$$\int_{-\infty}^{+\infty} f_X(x)\,dx \quad = 1$$

  - PDF and CDF

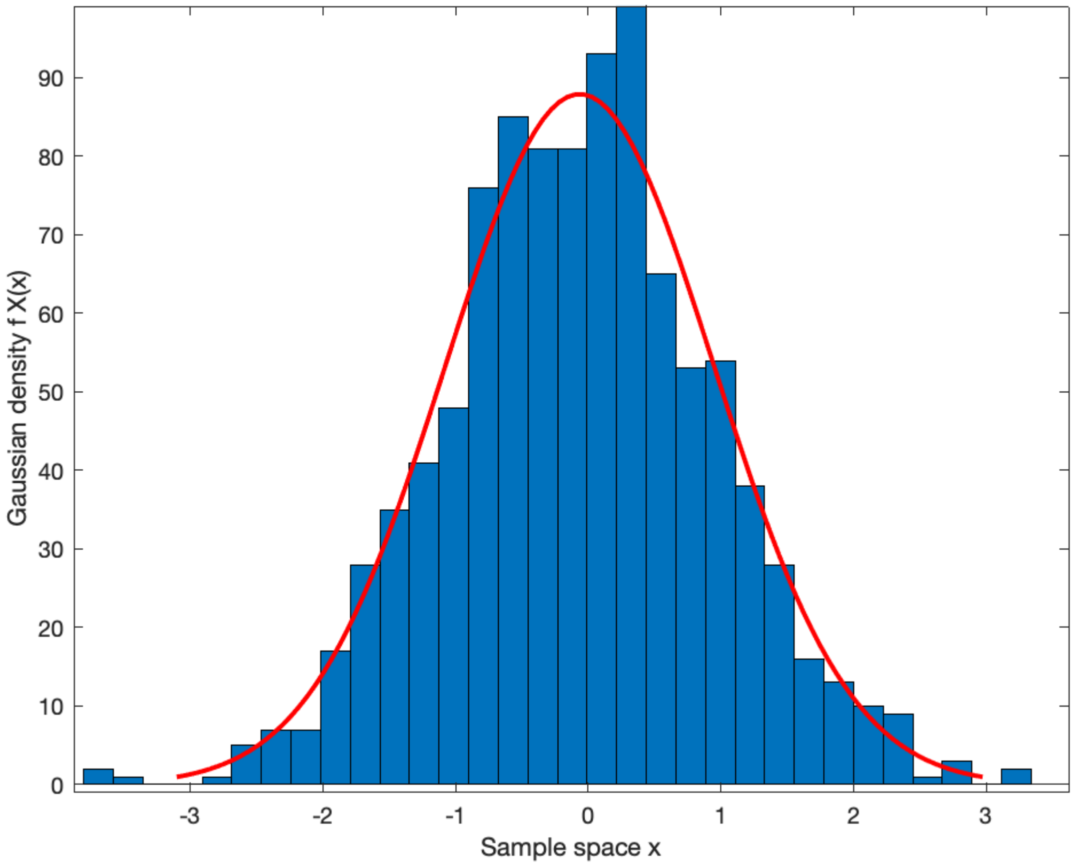
$$f_X(x) = \frac{d}{dx} F_X(x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$



Gaussian density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$



Normal density function

Several random variables

- Joint distribution function

$$F_{XY}(x,y) = \text{Prob}\{X < x \; and \; Y < y\}$$

- Joint density function

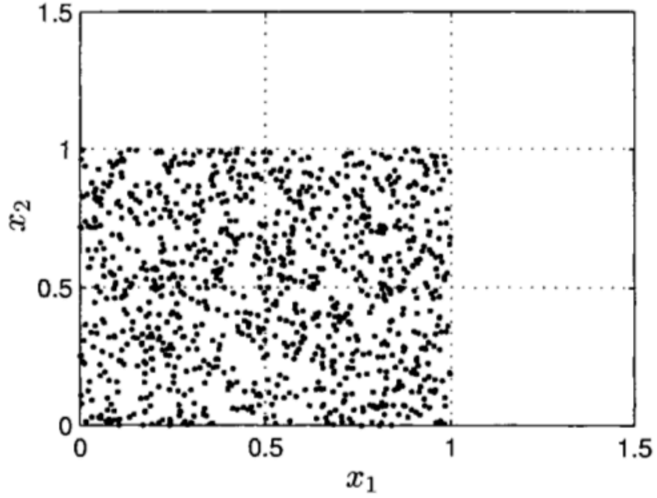
$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$

- Independent random variables

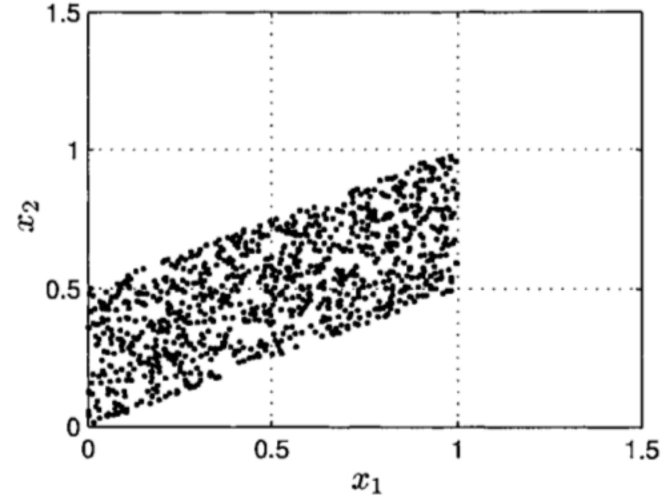
$$F_{XY}(x,y) = F_X(x) F_Y(y)$$

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

Independndt variables



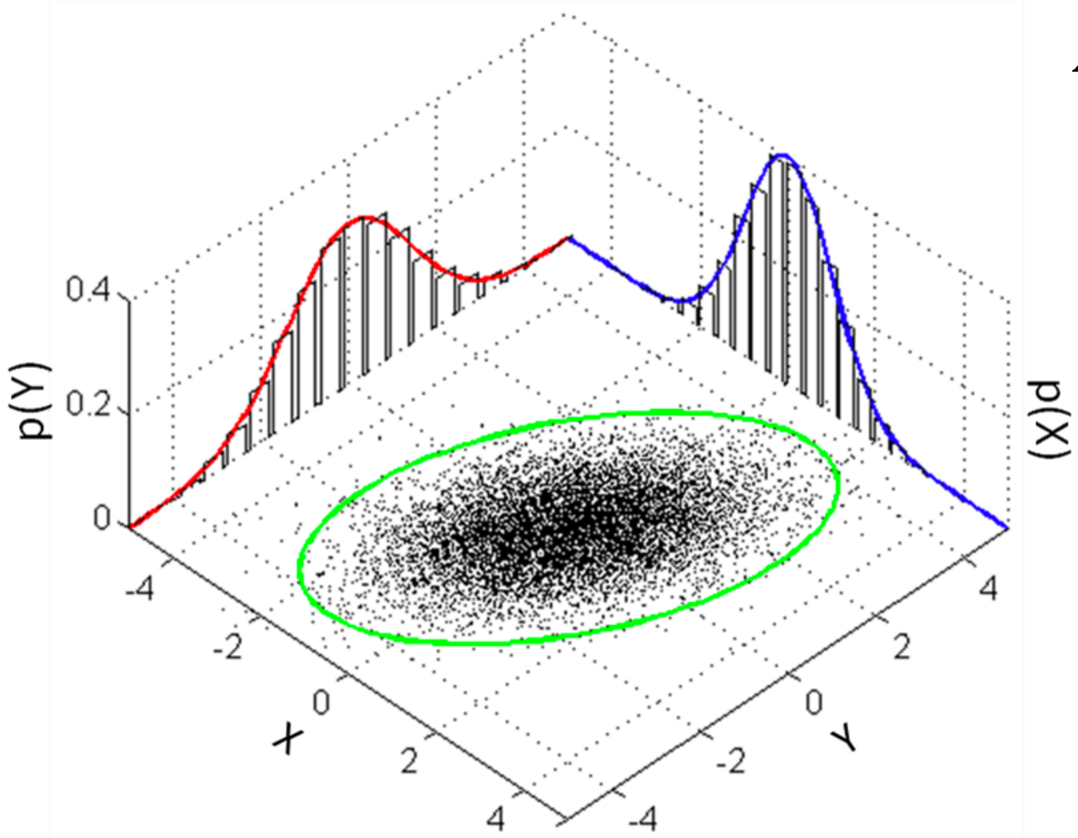
(a) No dependency



(b) Dependency

Independent variables only when you can describe X without the need of observing Y

Independnt variables

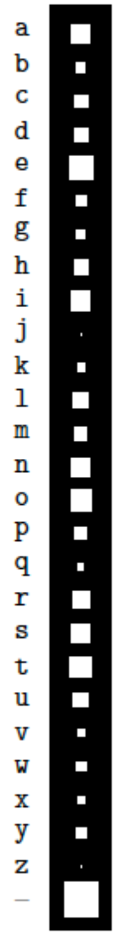


$$X \sim N(\mu, \Sigma)$$

Multivariate normal distribution

Probabilities and ensembles

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928



From «The Frequently Asked Questions Manual for Linux»

Outcome x is the value of a random variable

$(x, \mathcal{A}_X, \mathcal{P}_X)$

$\mathcal{A}_X = \{a_1, a_2, \ldots, a_i, \ldots, a_I\}$ set of values

$\mathcal{P}_X = \{p_1, p_2, \ldots, p_I\}$ probabilities

$P(x = a_i) = p_i$

$\sum_{a_i \in \mathcal{A}_X} P(x = a_i) = 1$

Expected value

- Expected value of a rv X

$$E\{X\} = \sum_{s \in S} X(s) P(s)$$

- For continuous rvs

$$E\{X\} = \int_{s \in S} X(s) dP(s)$$

Expected value

- Linearity Z = aX+bY

$$E\{Z\} = aE\{X\} + bE\{Y\}$$

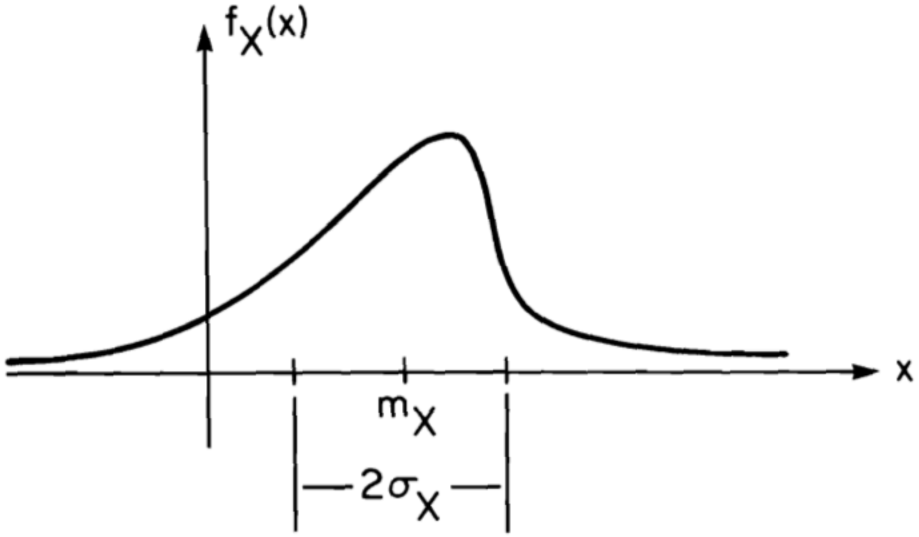
- Function of a rv

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

Moments

- First moment a function $f_X(x)$

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



Higher moments

- First moment - average value

$$m_X = E\{X\}$$

- Second moment - standard deviation and variance

$$\sigma_X = \sqrt{E\{(X - m_X)^2\}}$$

$$\sigma_X^2 = E\{(X - m_X)^2\} = E\{(X)^2\} - m_X^2$$

- 3rd central moment – skewness

$$\sigma_X^3 = E\{(X - m_X)^3\}$$

- 4th central moment – kurtosis

$$\sigma_X^4 = E\{(X - m_X)^4\}$$

Correlation

- Correlation – second joint moment

$$R_{XY} = E\{XY\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_Y(y)f_X(x)dxdy$$

- Covariance

$$K_{XY} = E\{(X - m_X)(Y - m_Y)\} = R_{XY} - m_x m_y$$

- Correlation coefficient

$$\rho = \frac{K_{XY}}{\sigma_X \sigma_y} \qquad -1 \leq \rho \leq +1$$

Correlation

- Correlation matrix – *n*-tuple of rvs $\boldsymbol{X} = [X_1 \ldots X_k]^T$

$$\boldsymbol{R_X} = E\{\boldsymbol{X^T X}\}$$

- Covariance matrix

$$\boldsymbol{K_X} = \boldsymbol{R_X} - \boldsymbol{\mu_X \mu_X^T}$$

Discrete time random processes

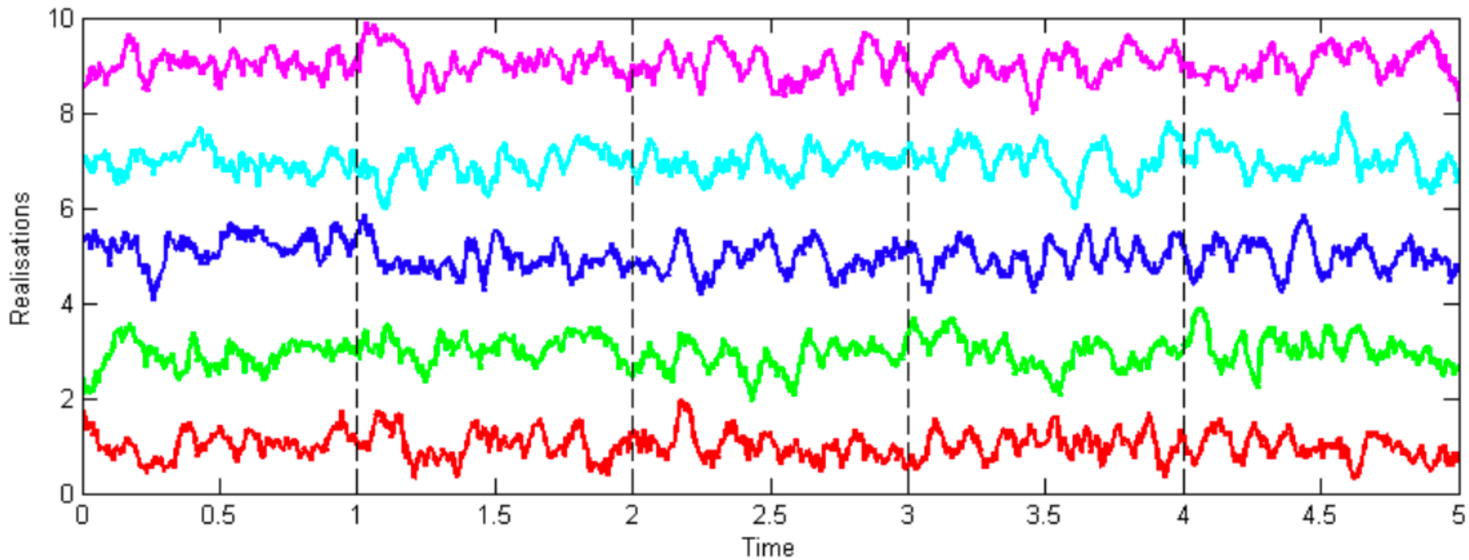
- Process
  - Result of an experiment
- Random processes
  - probabilistic models of ensembles of waveforms and sequences
- Digital signal processing
  - speech,
  - visual signals (images, videos),
  - sonar and radar,
  - geophysical,
  - astrophysical,
  - biological signals, ...

Discrete time random processes

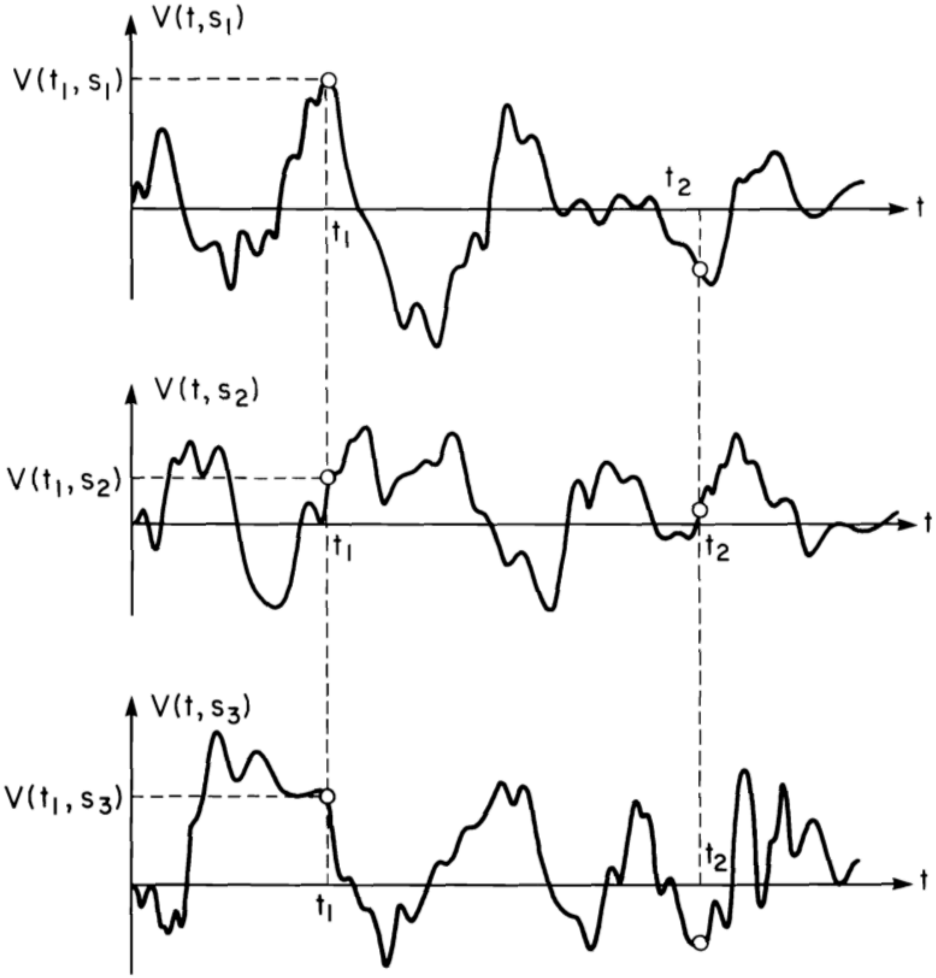
- Random process – definition
  - A random process X(t, s) is a random function of time t and a sample-point variable s
  - X(t, .) is a function of sample points, i.e. a random variable
  - X(. , s) is a function of time, i.e. a sample function
- Intuition
  - A random variable x becomes a function of the possible outcomes (values) s of an experiment and time t: x(s , t)
  - The family of all such functions is called a random process, X(s , t)
  - A random process becomes a random variable for fixed time

Ensemble and realization

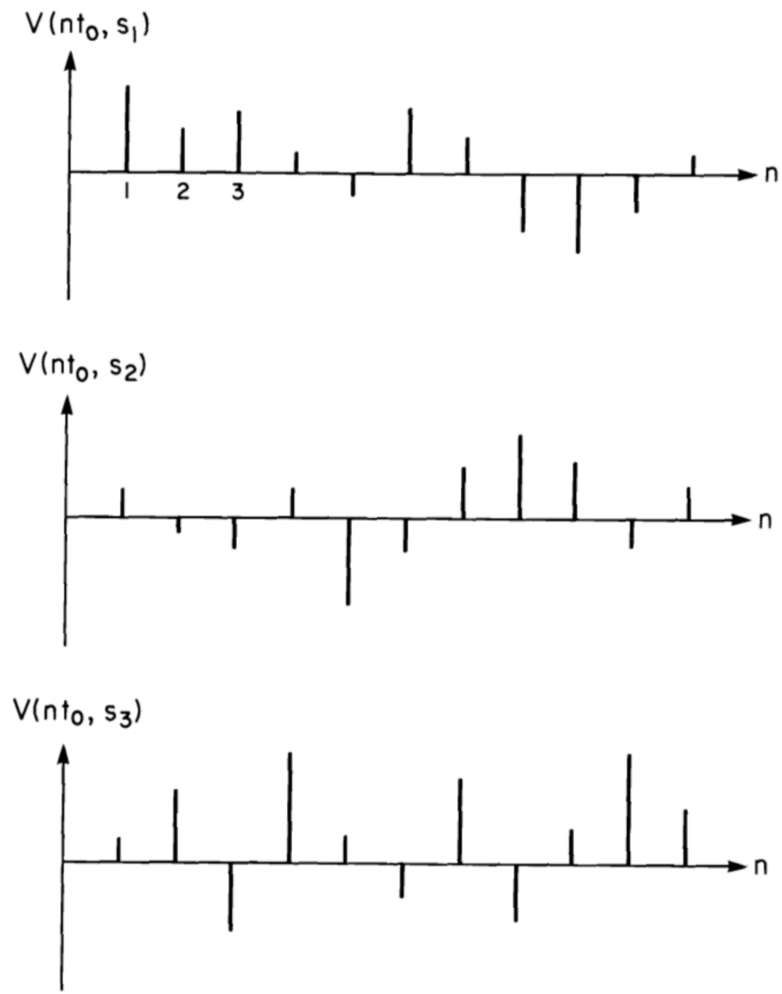
- X(s , t) represents a family or ensemble of time functions
  - Convenient short form x(t) for specific waveform of the random process X(t)
  - Each member time function is called a realization
  - The complete collection of sample functions of a random process is called the ensemble



Ensemble and realization



Continuous-time random process

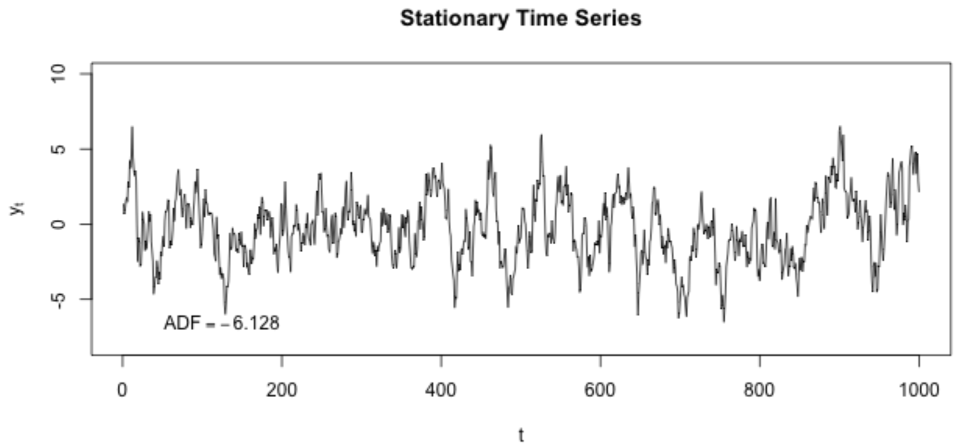


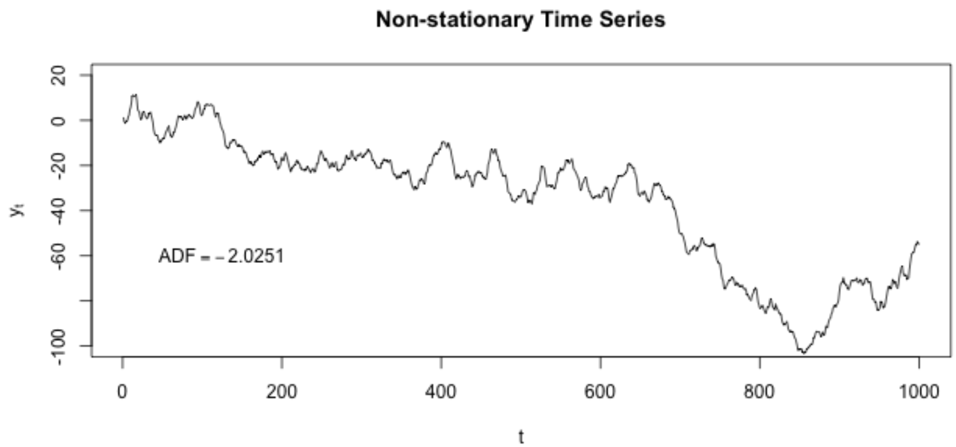
Discrete-time random process

Temporal processes

- Stationary process
  - "A stationary process (or strict(ly) stationary process or strong(ly) stationary process) is a stochastic process whose joint probability distribution does not change when shifted in time"
  - Parameters such as the mean and variance, if they are present, also do not change over time and do not follow any trends
- Cyclostationary process
  - "A cyclostationary process is a signal having statistical properties that vary cyclically with time"
  - A cyclostationary process can be viewed as multiple interleaved stationary processes
  - Examples: temperature, solar radiation, etc.

Stationary process





$\{X_t\}$ stochastic process

$$F_X(x_{t_1+\tau}, \ldots, x_{t_k+\tau}) = F_X(x_{t_1}, \ldots, x_{t_k})$$ Joint distribution

Ergodicity

- Ergodicity
  - An ergodic dynamical system has the same behavior averaged over time as averaged over the space of all the system's states (phase space)
  - Ergodicity is where the ensemble average equals the time average
  - Examples
    - In physics, a system satisfies the ergodic hypothesis of thermodynamics
    - In statistics, a rp for which the time average of one sequence of events is the same as the ensemble average

Discrete ergodicity

- A process is ergodic if the
  - mean is

$$\langle X(n) \rangle = \frac{1}{2N+1} \sum_{n=-N}^{N} X(n) = \mathbb{E}\{X(n)\}$$

  - the autocorrelation is

$$\langle X(n)X(n-l) \rangle = \mathbb{E}\{X(n)X(n-l)\}$$

- Two processes are joint ergodic

$$\langle X(n)Y(n-l) \rangle = \mathbb{E}\{X(n)Y(n-l)\}$$

Definitions

- Mean

$$\mathbb{E}\{X(t)\} = m_X(t)$$

- Autocorrelation

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

- Autocovariance

$$\mathbb{E}\{[X(t_1) - m_X(t_1)][X(t_2) - m_X(t_2)]\} = K_X(t_1, t_2)$$

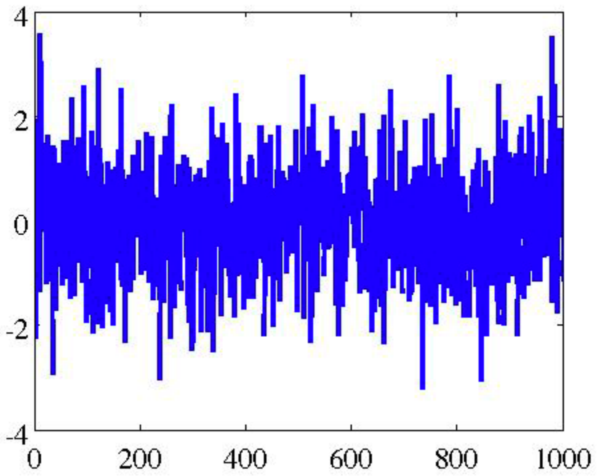
- Cross-correlation

$$\mathbb{E}\{X(t_1)Y(t_2)\} = R_{XY}(t_1, t_2)$$

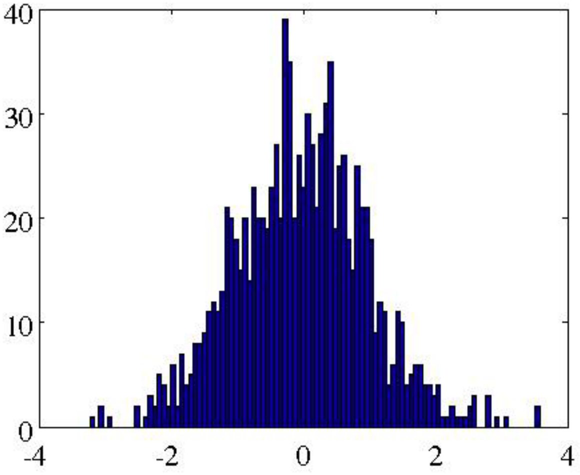
- Cross-covariance

$$\mathbb{E}\{[X(t_1) - m_X(t_1)][Y(t_2) - m_Y(t_2)]\} = K_{XY}(t_1, t_2)$$

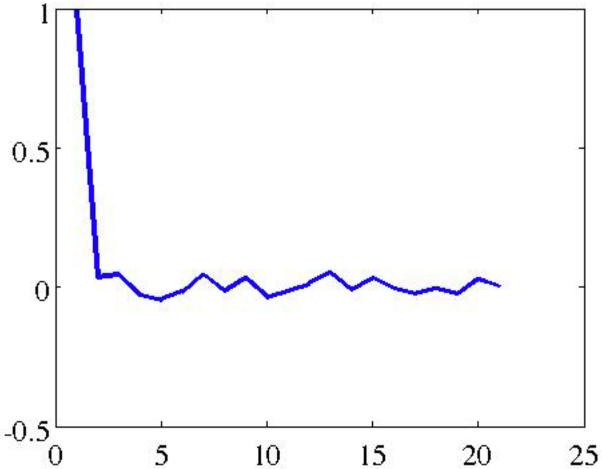
Autocorrelation



signal

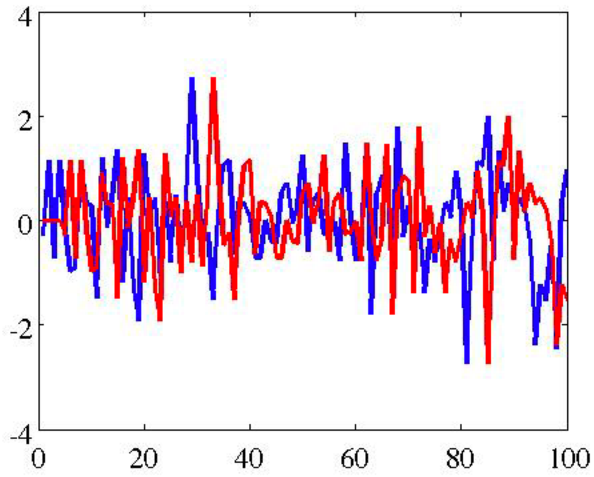


distribution (histogram)



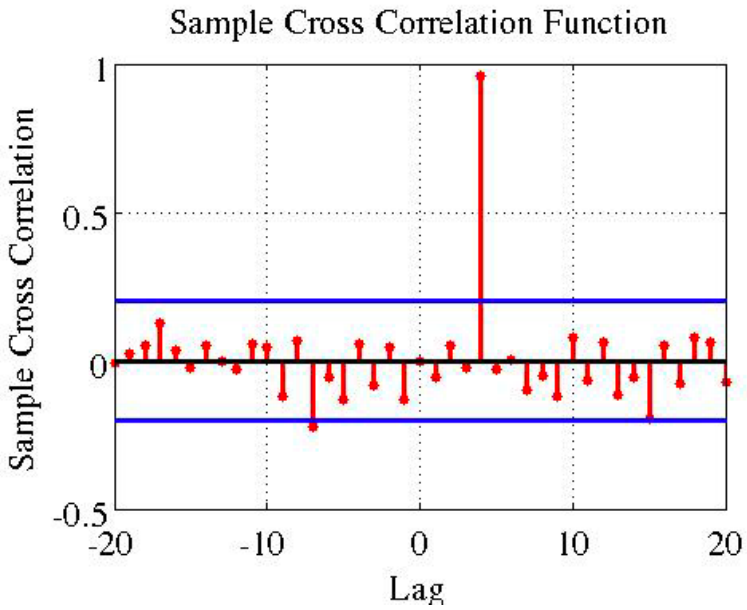
autocorrelation

Cross-correlation function (XCF)



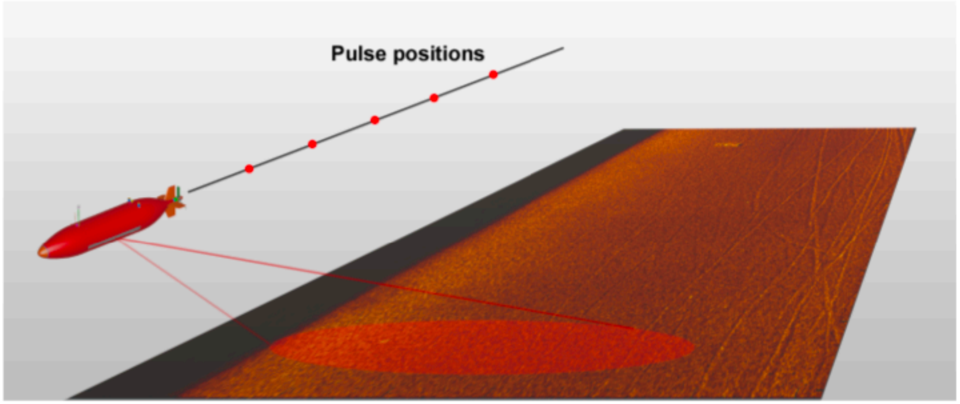
source signals

4 samples delay



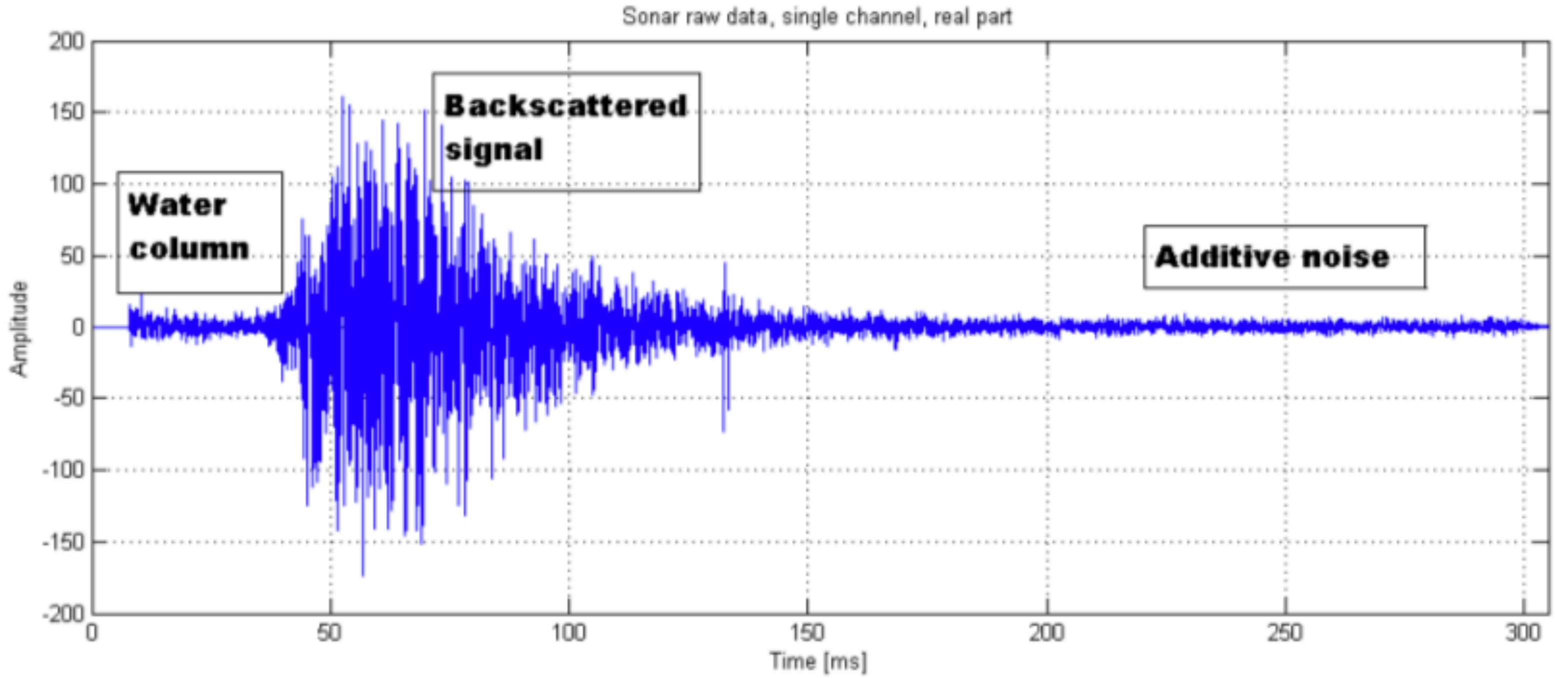
cross-correlation function

Sonar



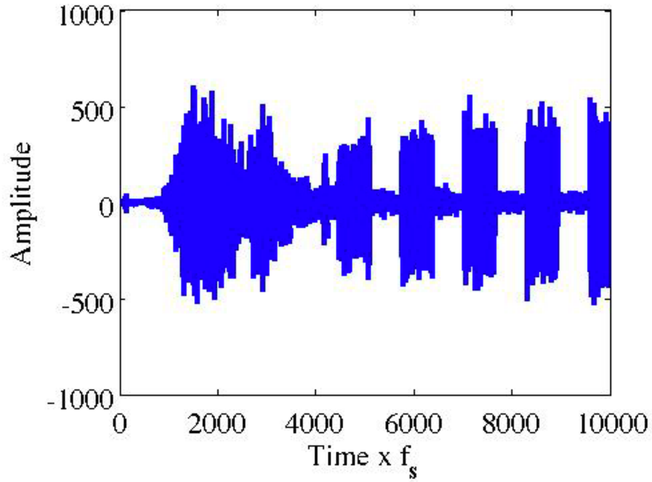
- The HUGIN autonomous underwater vehicle
- Wideband interferometric synthetic aperture sonar
- Transmitter that insonies the seafloor with a LFM pulse
- Array of receivers that collects the echoes from the seafloor
- The signal scattered from the seafloor is considered to be random
- The signal consists of a signal part and additive noise

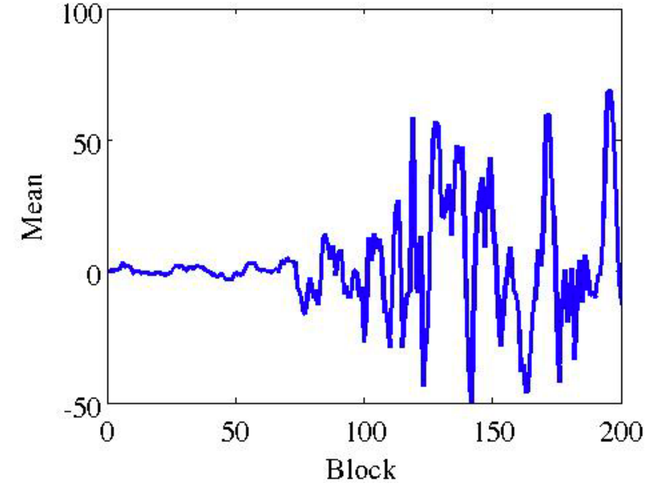
Sonar

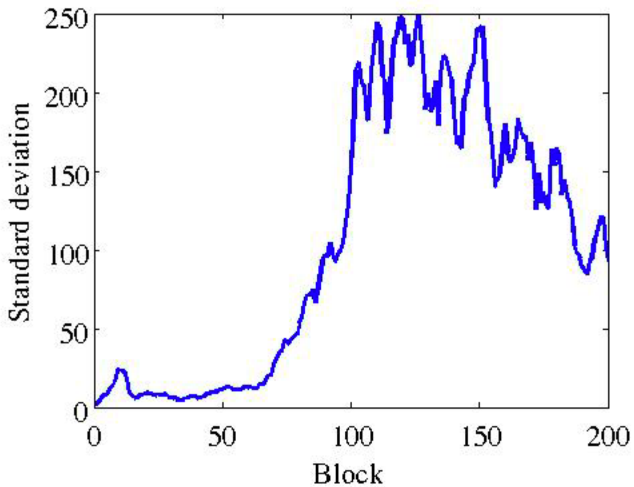


Sonar raw data, single channel, real part
Backscattered signal
Water column
Additive noise
Amplitude
Time [ms]
200
150
100
50
0
-50
-100
-150
-200
0
50
100
150
200
250
300

Sonar

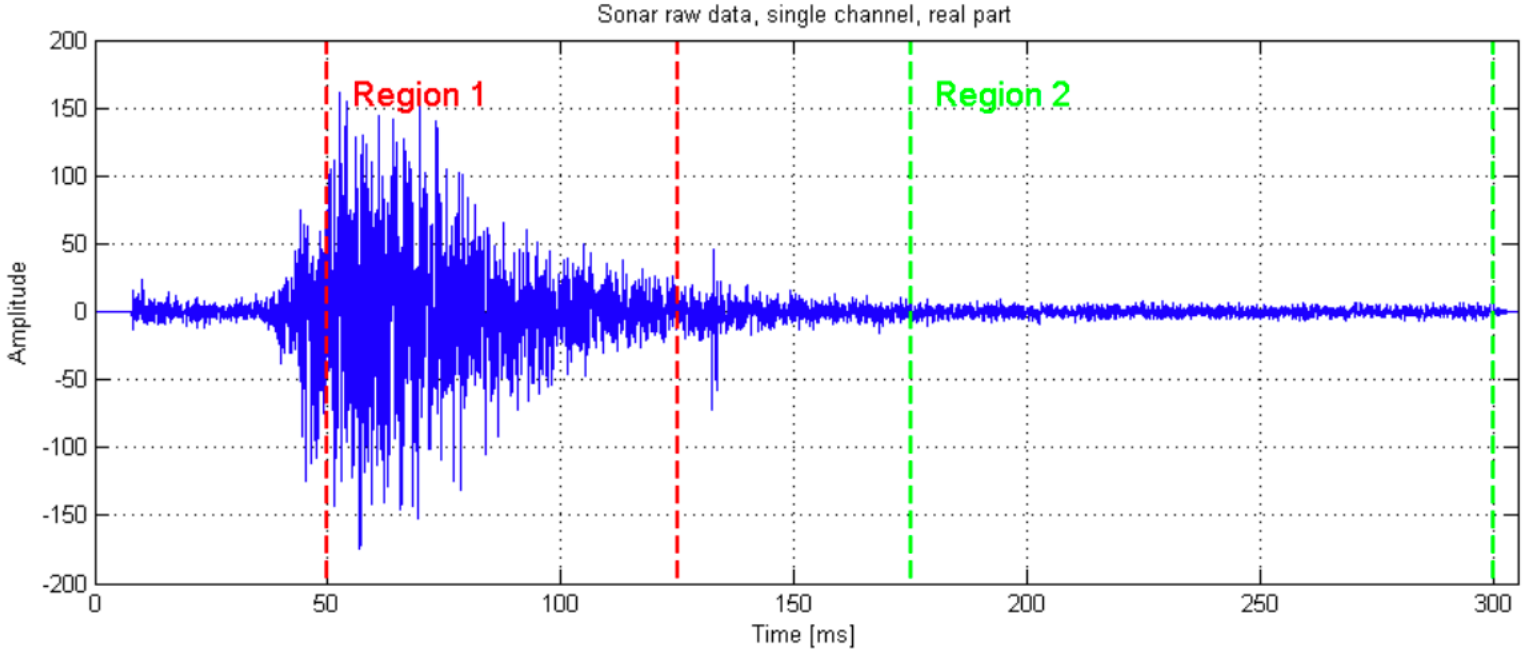






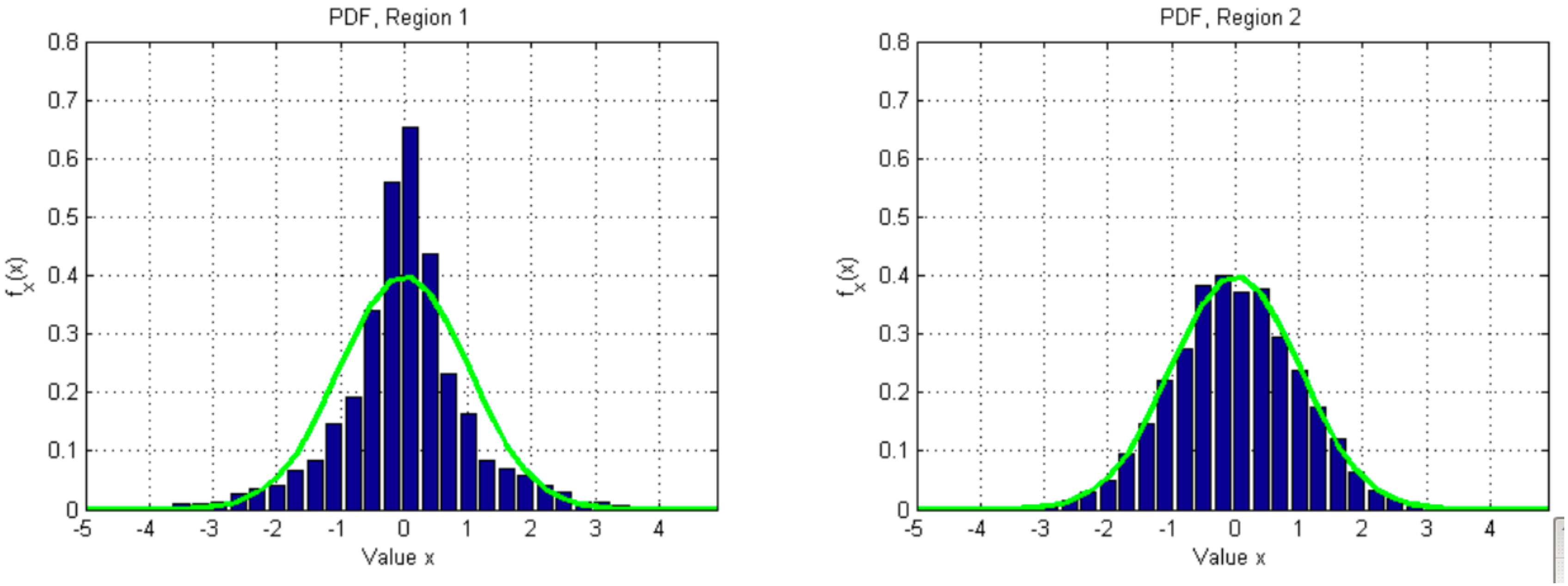
is the process stationary?

Sonar



Two «similar» regions

Sonar



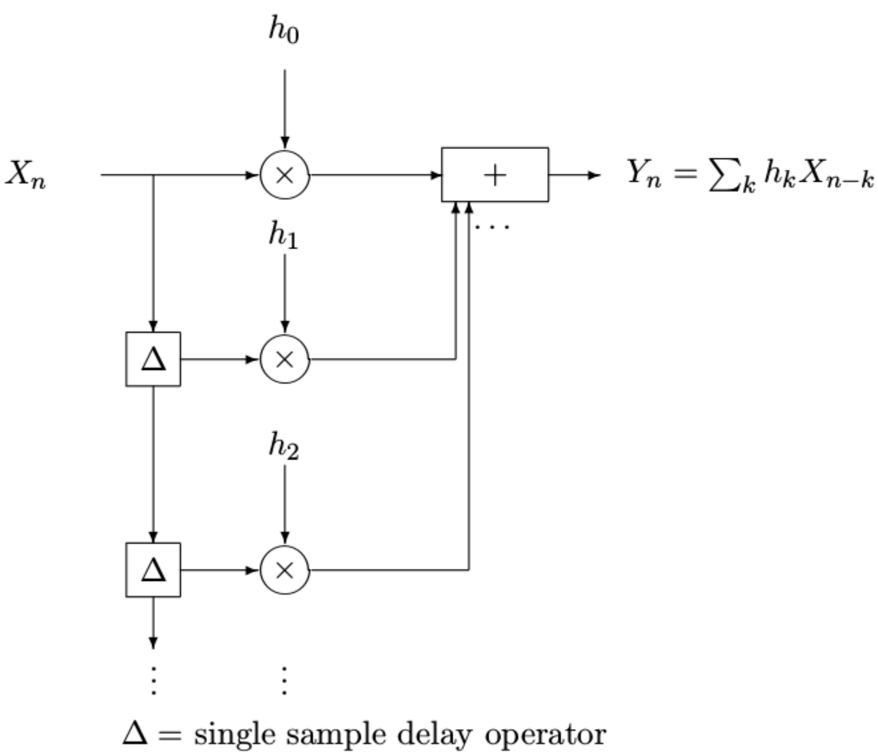
Estimated PDF

is the probability density function Gaussian?

Moving Average Process (MA)

- Many complicated random processes are well modeled as a linear operation on a simple process

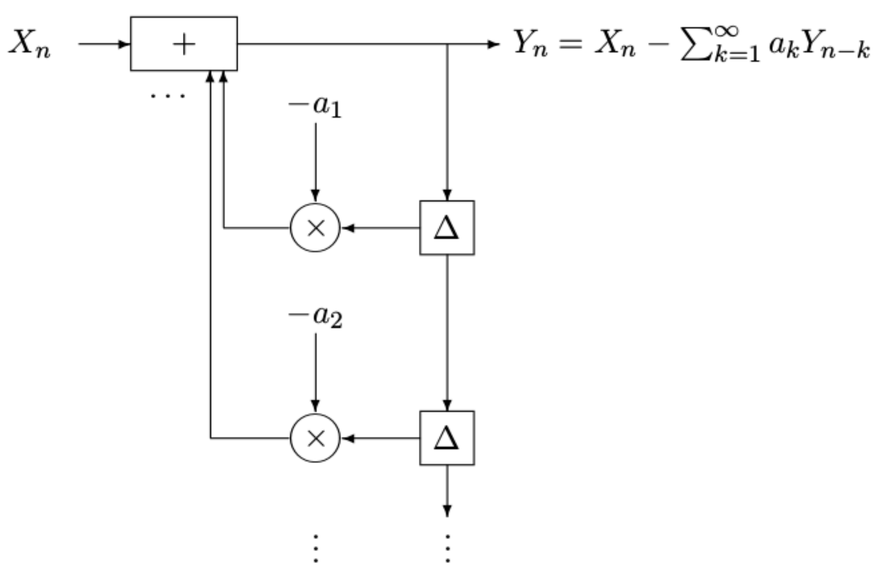
$$Y_n = \sum_k X_{n-k} h_k$$



Autoregressive process (AR)

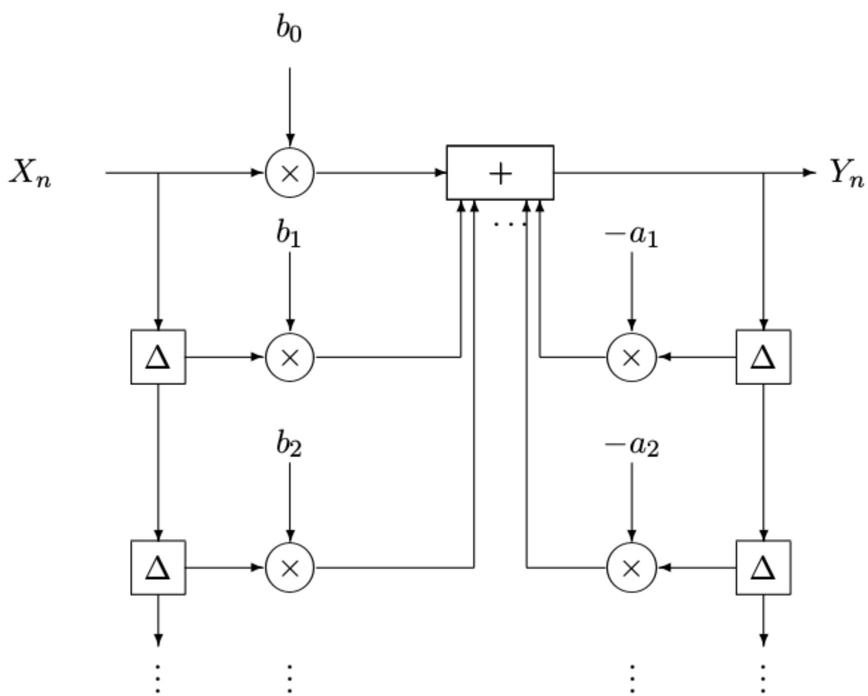
- Convolving the outputs to get the inputs instead of vice versa

$$Y_n = X_n - \sum_{k=1}^{\infty} a_k Y_{n-k}$$



ARMA

$$Y_n = \sum_{k=1}^{P} a_k Y_{n-k} + \sum_{k=0}^{Q} b_k X_{n-k}$$



Estimation principles

- A sequence x(n) is the realization of a random process

- Mean

$$m_x = E[x_n] = \int_{-\infty}^{+\infty} x p_{x_n}(x,n)\,dx$$

- Estimated Mean

$$m_x = \frac{1}{N}\sum_{n=0}^{N-1} x(n)$$

Estimation principles

- Variance

$$\sigma_x^2 = E\left[(x_n - m_x)^2\right]$$

- Autocovariance

$$\gamma_{xx}(m) = E\left[(x_n - m_x)(x_{n+m}^* - m_x^*)\right]$$

- Power Spectral Density

$$P_{xx}(\omega) = \sum_{m=-\infty}^{+\infty} \gamma_{xx}(m) e^{-j\omega m}$$

Estimation principles

- Autocorrelation is defined as the autocovariance with means equals to zero

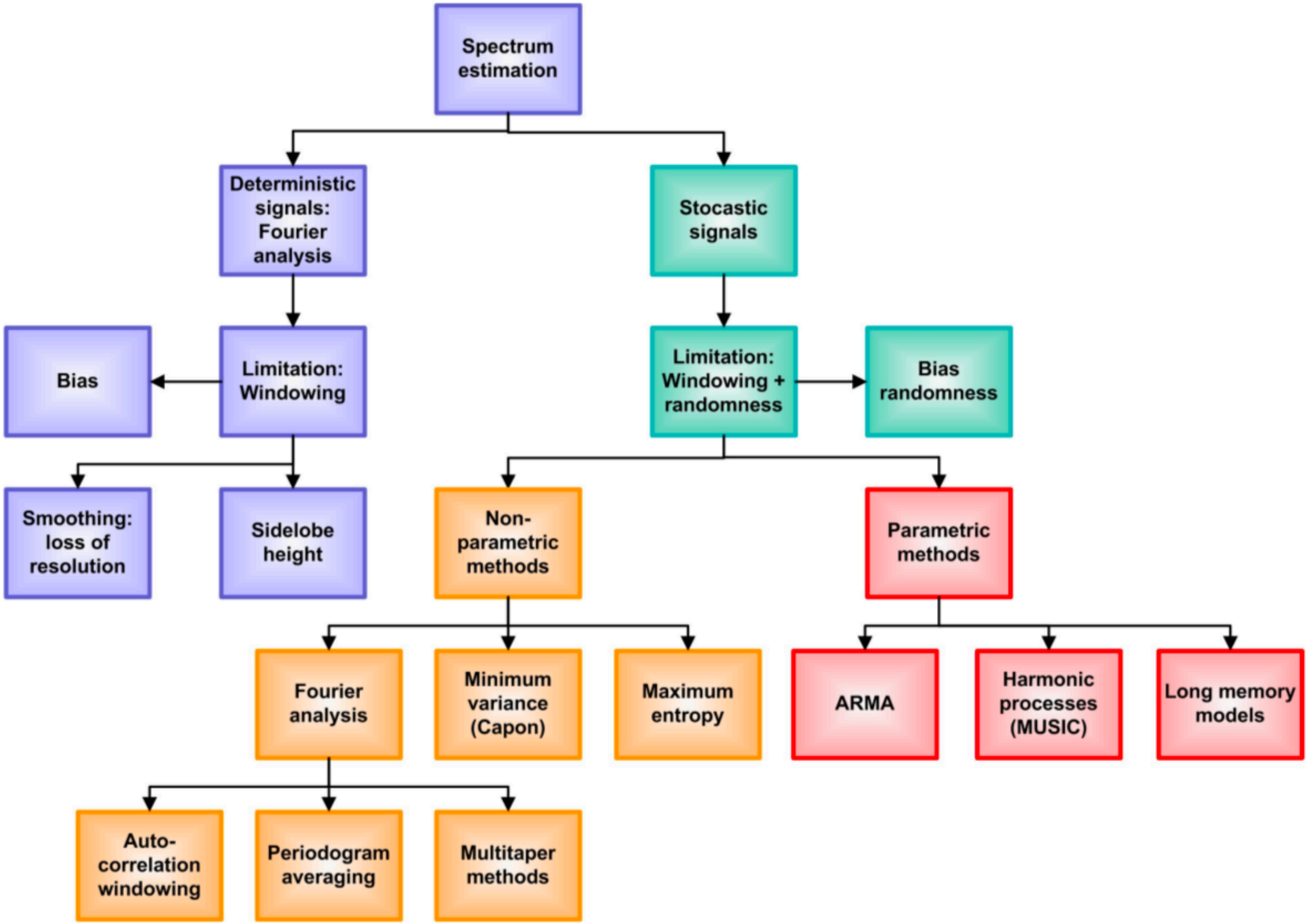
- Estimated autocorrelation

$$c_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n)x(n+m)$$

- Estimated PSD

$$I_N(\omega) = \sum_{m=-(N-1)}^{N-1} c_{xx}(m) e^{-j\omega m}$$

Spectrum estimation techniques



Periodogram

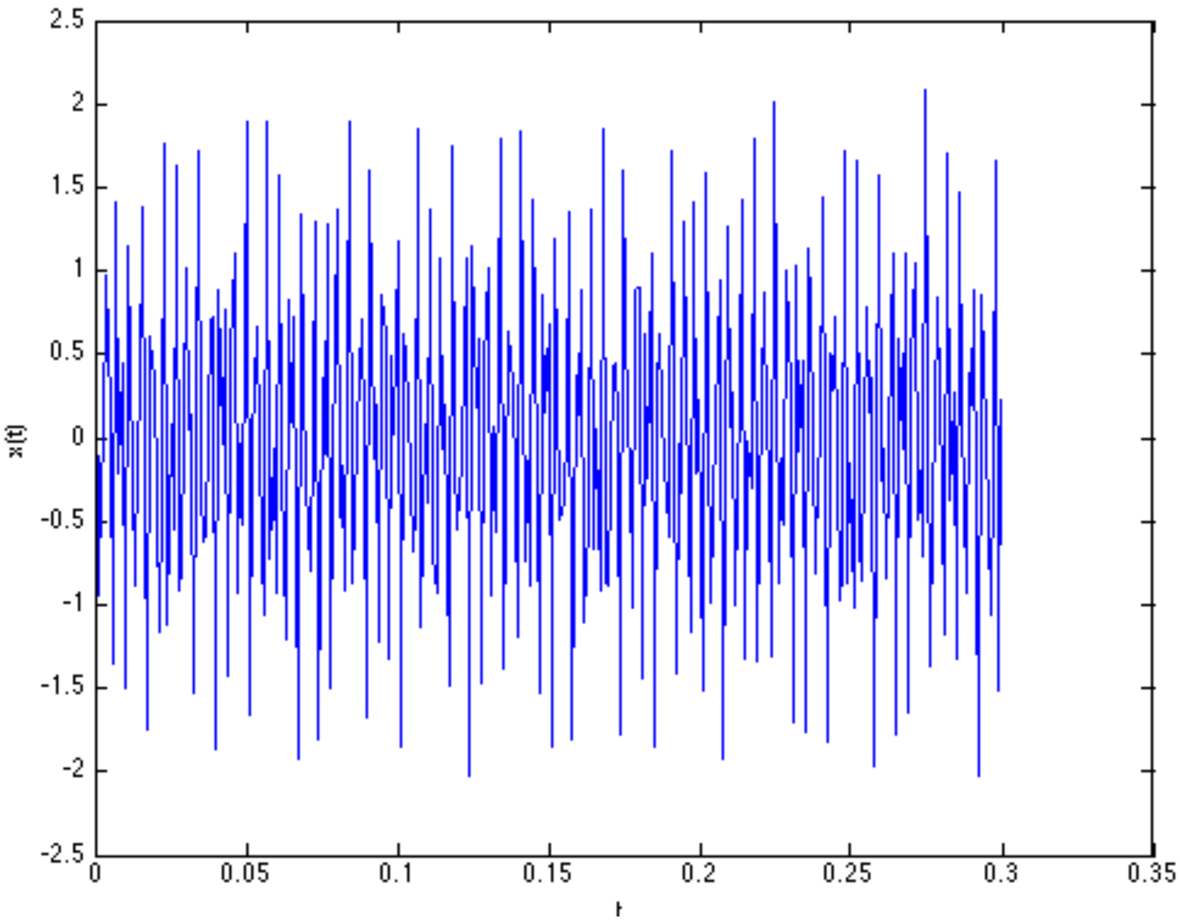
- DTFT of a sequence x(n)

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n) e^{-j\omega n}$$

- Periodogram

$$I_N(\omega) = \frac{1}{N}\left|X(e^{j\omega})\right|^2 = \frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1} x(l)x(m)e^{j\omega m}e^{-j\omega l}$$

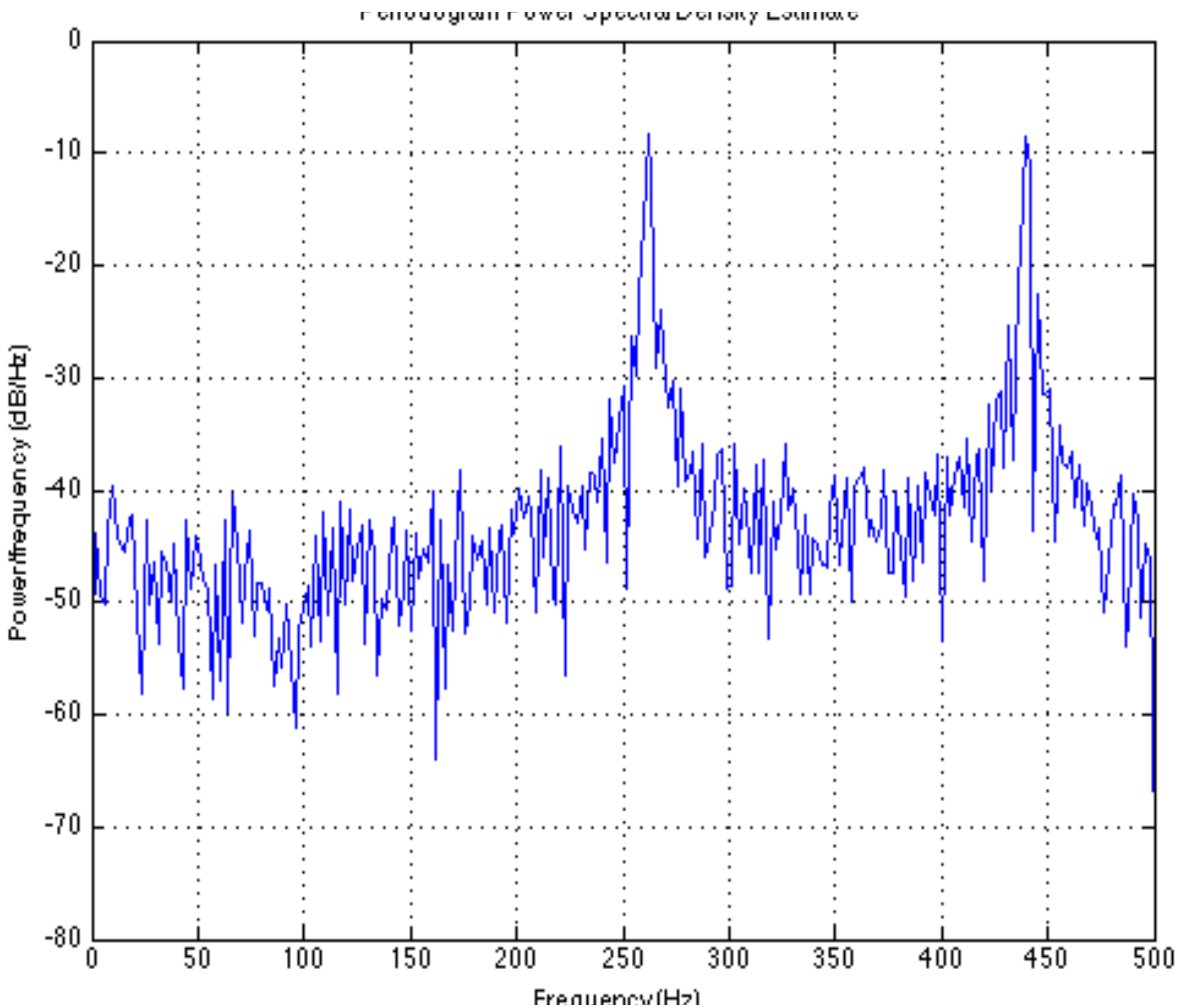
Periodogram



Source signal

$$x(n) = \sin(2\pi \cdot 262 \cdot n) + \sin(2\pi \cdot 440 \cdot n) + 0.1 \cdot randn(n)$$

Periodogram



Estimated Periodogram

Considerations

- The Periodogram
  - We have problems with increasing the N
  - The variance does not approach zero as the data length N increases
  - The periodogram is not a consistent estimator (i.e. converges in some sense to the true value)
  - Why does the variance not decrease with increasing N?
    - Increasing N means increasing the number of individual frequencies (instead of increasing the accuracy of each frequency)
- "smoothed" Periodograms are defined

Bartlett's method

- A sequence x(n) is divided in *K* segments of *M* samples ($N = KM$)

$$x^{(i)}(n) = x(n + iM - M) \qquad 0 \le n \le M-1; 1 \le i \le K$$

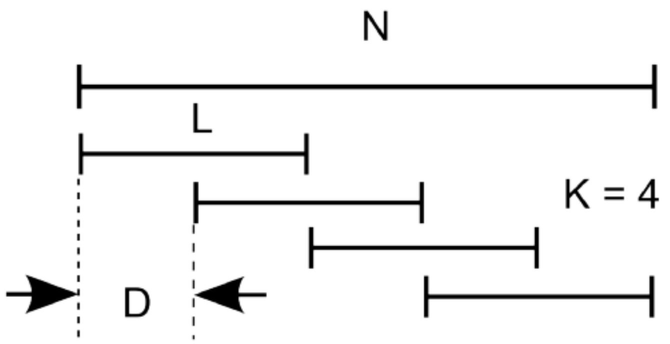
- *K* Periodograms are calculated

$$I_M^{(i)}(\omega) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x^{(i)}(n) e^{-j\omega n}\right|^2$$

Bartlett's method

- Estimation of the spectrum

$$B_{xx}(\omega) = \frac{1}{K} \sum_{i=1}^{K} I_M^{(i)}(\omega)$$



Welch's method

- A window w(n) is applied

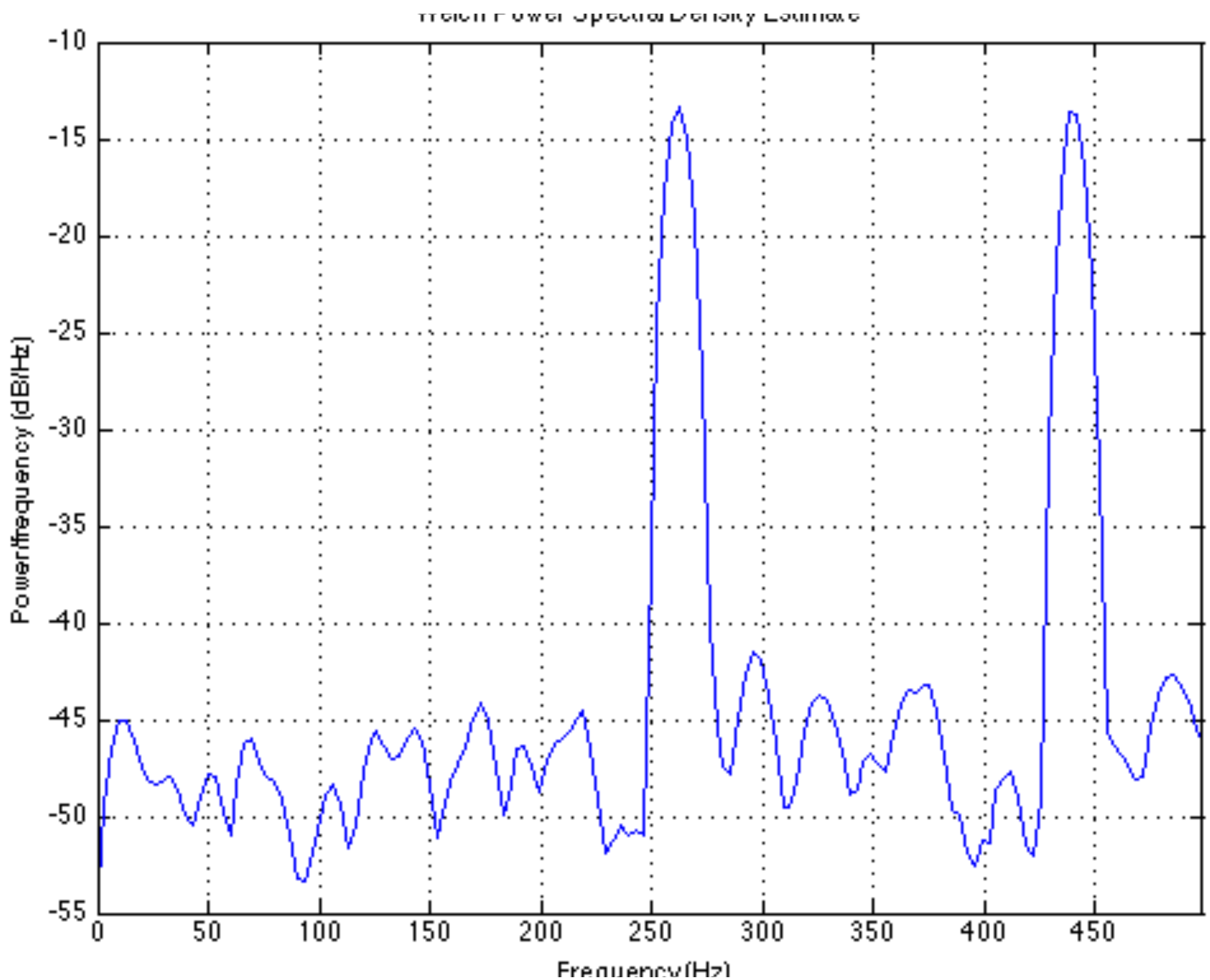
$$J_M^{(i)}(\omega) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x^{(i)}(n)w(n)e^{-j\omega n}\right|^2$$

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n)$$

$$\Rightarrow$$

$$B_{xx}^{\omega}(\omega) = \frac{1}{K}\sum_{i=1}^{K} J_M^{(i)}(\omega)$$

Welch



Estimated Periodogram

MUSIC

- Very important are the methods based on the decomposition by eigenvectors and eigenvalues

- Methods
  - Pisarenko
  - MUltiple Signal Classification (MUSIC)
  - Estimation of Signal Parameters via Rational Invariance Technique (ESPRIT)

- Applications
  - Spectral estimation
  - Direction Of Arrival (DOA)

MUSIC

- The sequence x(n) is

$$x(n) = \sum_{i=1}^{p} A_i e^{jn\omega_i} + w(n)$$

- Autocorrelation matrix

$$r_x(k) = \sum_{i=1}^{p} P_i e^{jk\omega_i} + \sigma_\omega^2 \delta(k)$$

$$P_i = |A_i|^2$$

MUSIC

- We write the autocorrelation matrix as

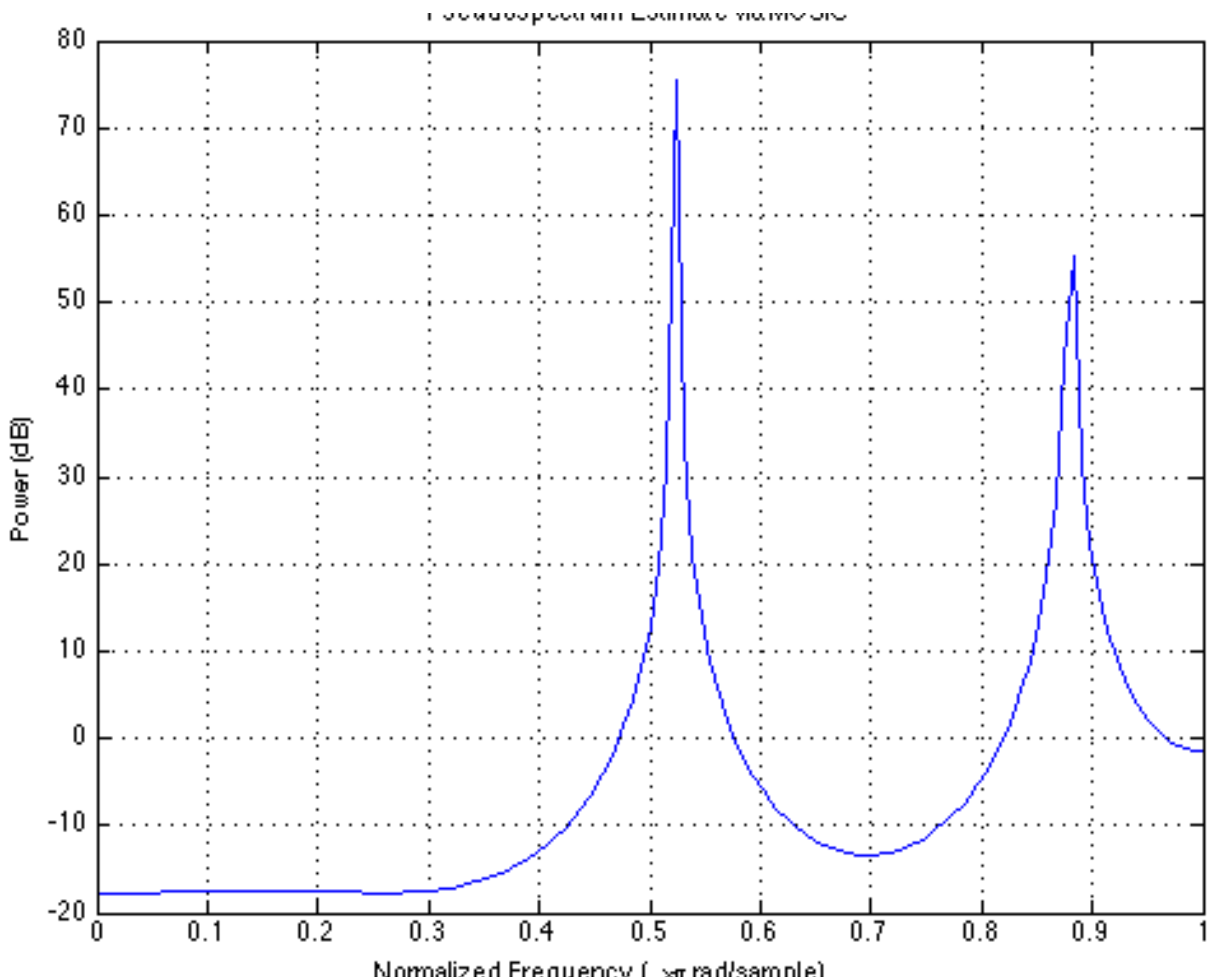
$$\mathbf{R}_x = \mathbf{R}_s + \mathbf{R}_n = \sum_{i=1}^{p} P_i \mathbf{e}_i \mathbf{e}_i^H + \sigma_w^2 \mathbf{I}$$

$$\mathbf{e}_i = \left[1, e^{j\omega_i}, e^{j2\omega_i}, \dots, e^{j(M-1)\omega_i}\right]$$

- The p principal components span the signal space

$$P_{MUSIC}(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^{M} \left|\mathbf{e}^H \mathbf{v}_i\right|^2}$$

Example



Estimated Periodogram