Normal pdf

Also called the "Gaussian" pdf

$$p_X(x;\mu,\sigma) = N(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$P(X \in [x, x+dx]) = p_X(x)dx$: probability of getting a value in $[x, x+dx]$

μ : mean of the pdf

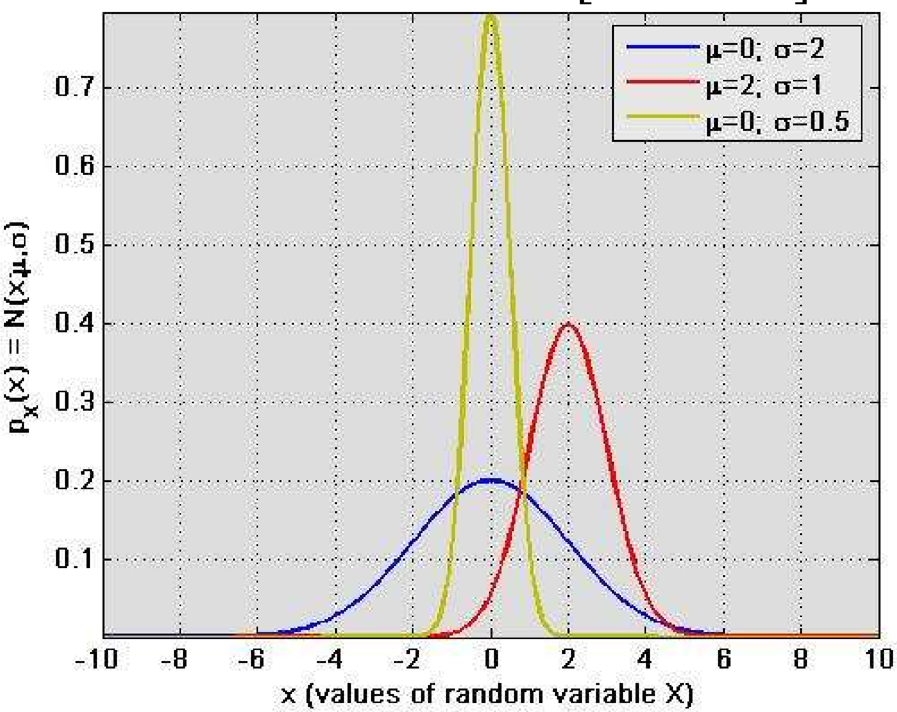
$\mu = E[X]$

σ^2 : variance of the pdf

$\sigma^2 = E[(X - E[X])^2]$

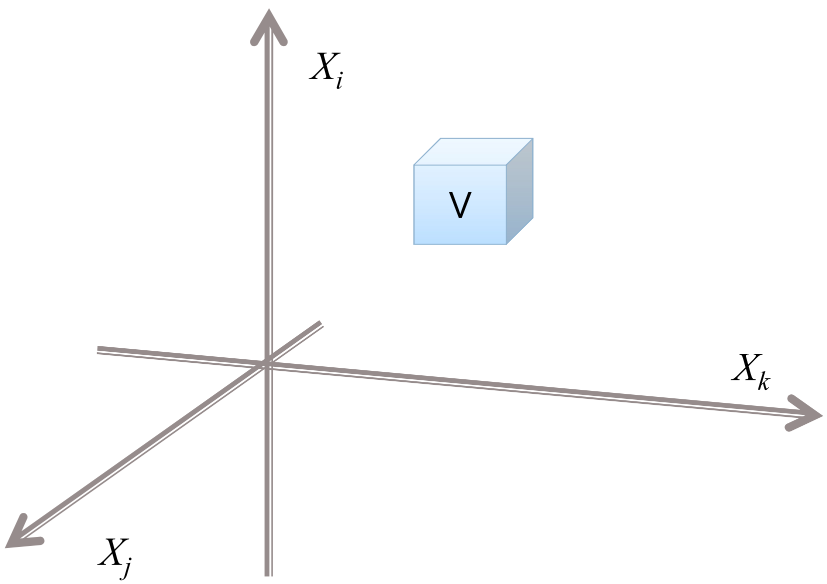
σ : standard deviation

Usually we say μ and σ^2 are the mean and variance of the random variable X



N random variables

- $\{ X_1, X_2, \ldots, X_N \}$
- Outcome of a trial is a vector $\bar{x} = (x_1, x_2, \ldots, x_N)$ in $\mathbb{R}^N$



Probability that a trial outcome will fall in a volume V

$$P(X \in V) = \int_V p_X(\bar{x}) \mathrm{d}^n x$$

$p_X(\bar{x})$: Joint pdf

$p_{Y \subset X}(x)$: marginal pdf of $Y \subset X$ obtained by integrating $p_X(\bar{x})$ over $X \backslash Y$

$X_i \sim p_{X_i}(x)$: X_i has the pdf $p_{X_i}(x)$

Expectation

$$E[f(X_1,\ldots,X_N)] = \int_{-\infty}^{\infty} f(x_1,\ldots,x_N)\, p_{X_1,\ldots,X_N}(x_1,\ldots,x_N)\, dx_1,\ldots,dx_N$$

$E[(X_1 - E[X_1])^{m_1}(X_2 - E[X_2])^{m_2}\ldots(X_N - E[X_N])^{m_N}]$ is called the $m_1 + m_2 + \ldots + m_N$ order (central) moment of the joint pdf

- Of course, there are many possible moments of a given order
- Of special importance is the central moment of second order

$$C_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]$$

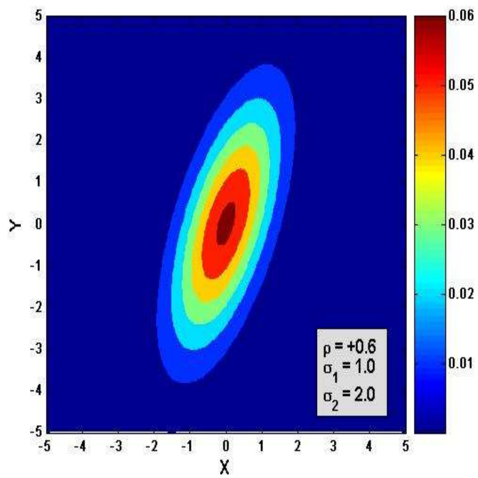
C_{ij} is called the Covariance of X_i and X_j

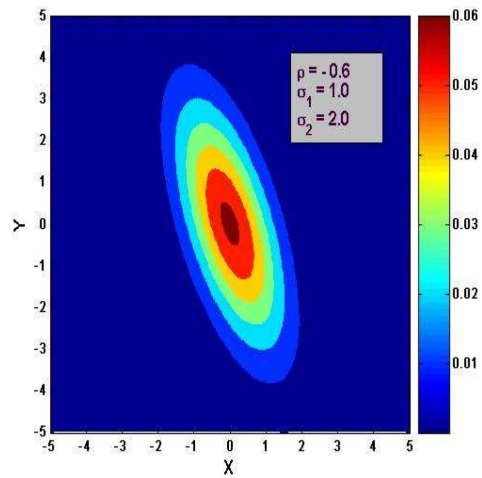
The matrix with elements C_{ij} is called the Covariance Matrix of the random variables $\{X_1,\ldots X_N\}$

Bivariate Normal pdf

$$p_{XY}(\bar{x};\bar{\mu},\mathbf{C}) = \frac{1}{2\pi\,|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\bar{x}-\bar{\mu})^T \mathbf{C}^{-1}(\bar{x}-\bar{\mu}) \right)$$

$$\mathbf{C} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix};\ \bar{x} = \begin{pmatrix} x \\ y \end{pmatrix};\ \bar{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

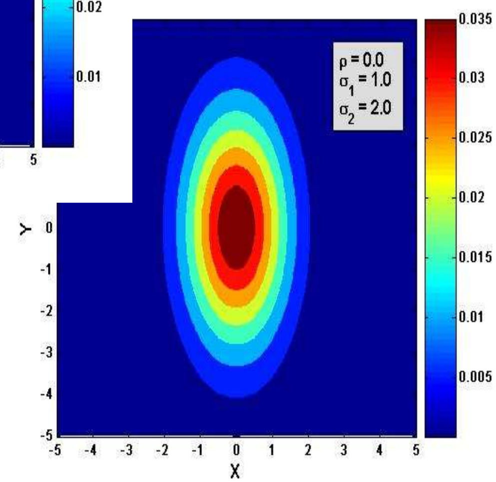




$$\mu_x = \mu_y = 0$$

$$\sigma_x = 1.0, \sigma_y = 2.0$$

$$\rho \in [-0.6, 0, 0.6]$$



One can prove that **C** is also the covariance matrix of the bivariate normal pdf

Stochastic process

A sequence of random variables

$$\{\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots\}$$

- The sequence need not be a vector in a finite dimensional space
- The indices can be taken to indicate "time" instants$\rightarrow$ "Time Series"
  - True continuous time theory for stochastic processes is more complicated

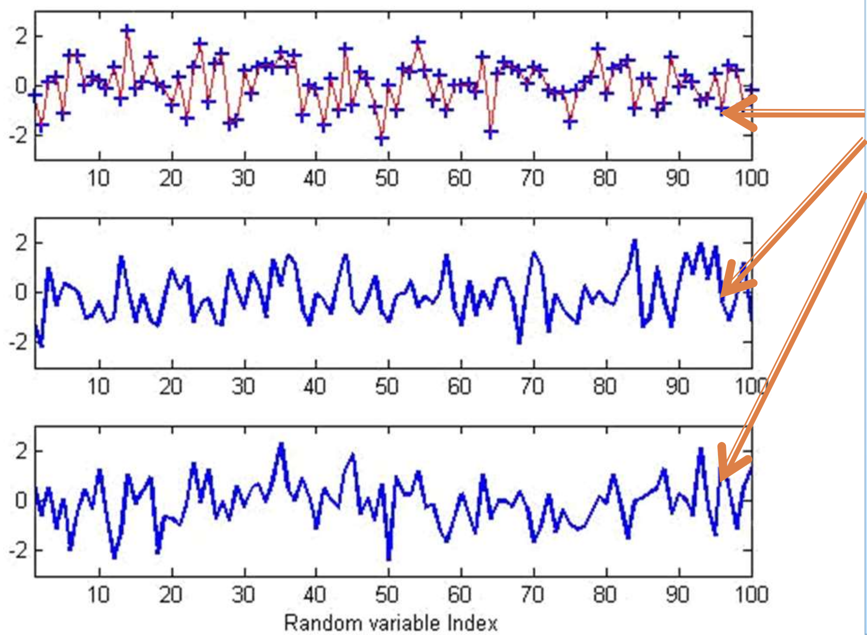
Each trial gives a set of values (a "realization")

$$\{\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots\}$$

- The stochastic processes we will mostly consider are time-series (i.e, only one index)
- The study of stochastic processes extends to spatial stochastic processes
- Notation: $X_{\mathrm{k}} \equiv X[\mathrm{k}]$

Textbook: *Stochastic Processes* by Ross; *Statistical Optics* by Goodman

Mathematical description of a stochastic process



•Different realizations of the same stochastic process
•Each time series is the outcome of **one** trial
• X-axis is the index number of the random variables
• The **ordering** of random variables distinguishes a stochastic process from a simple collection of variables: change the ordering and you change the stochastic process

Ensemble average

Ensemble average

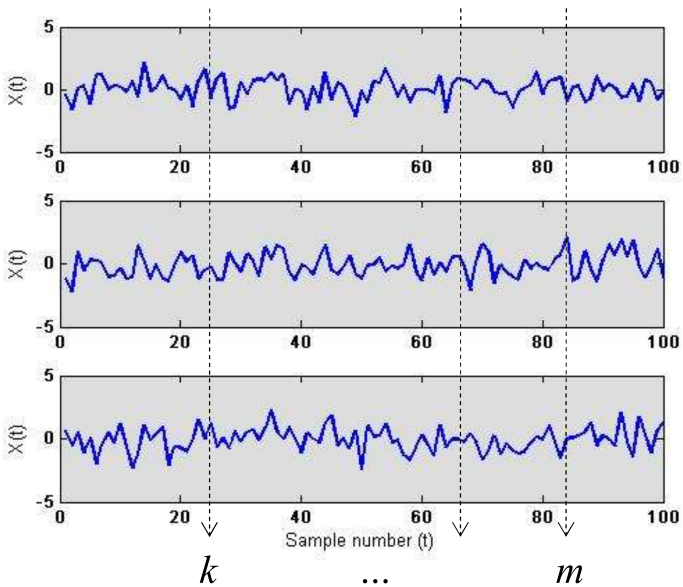
Same as the joint expectation defined earlier

$$E[f(X_k,\ldots,X_m)] = \int_{-\infty}^{\infty} f(x_k,\ldots,x_m)\, p_{X_k,\ldots,X_m}(x_k,\ldots,x_m)\,dx_k,\ldots,dx_m$$

Ensemble averages are constructed by averaging $f(X_k, \ldots X_m)$ over an infinite number of realizations of the stochastic process

- Compute $f_p = f(X_k, \ldots X_m)$ for the p^{th} realization ($1 \le p \le N$)

- Then compute $\frac{1}{N}\sum_{p=1}^{N} f_p$ with $N \to \infty$



Stationary stochastic process

- Wide-sense stationary process
  - $E[\,X_k\,]$ is constant, independent of k
  - $E[\,X_k X_{k+m}\,]$ is independent of k and dependent only on m
    - *First and second order moments are time-translation independent*
  - Makes sense to talk about the "mean and variance of the stochastic process"

- Strictly stationary stochastic process: ***All*** joint moments are time-translation independent

Gaussian noise: multivariate Normal pdf

- Multivariate Normal pdf:
- $p_{\bar{X}}(\bar{x}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}|^{1/2}} \exp(-\frac{1}{2}\|\bar{x} - \bar{\mu}\|^2)$
- $\bar{x} \in R^N$ (row vector)
- $E[X_i] = \mu_i$
- $C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$: Covariance matrix
- $|\boldsymbol{C}|$: Determinant of $\boldsymbol{C}$
- $\|\bar{x}\|^2 = \langle \bar{x}, \bar{x} \rangle$ where $\langle \bar{x}, \bar{y} \rangle = \bar{x}\boldsymbol{C}^{-1}\bar{y}^T$
- Gaussian noise: The joint pdf of any subset of the random variable sequence is a multivariate normal pdf

Textbook: *Introduction to multivariate statistics*, T.W. Anderson

Wide-sense stationary Gaussian noise

- $E[X_i] = \mu$ (Independent of i)
- $E[(X_i - \mu)(X_{i\pm m} - \mu)] = \phi(m)$ (dependent only on the separation): Autocovariance sequence of the noise
- Covariance matrix is a Toeplitz matrix

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

Power spectral density

- $\phi(0) = E[(X_i - \mu)^2] = \sigma^2$
- $\phi(k)$: even function of k (Because $\boldsymbol{C}$ is always symmetric)

We use $\tilde{s}$ to denote the discrete Fourier transform (DFT) of $\bar{s}$,

$$\tilde{s}^T := \mathbf{F}\bar{s}^T, \quad (1)$$

$$F_{\mathrm{km}} = e^{-2\pi i k m/N}, \quad (2)$$

with $\tilde{s}_j$ being its jth element. The inverse DFT is given by,

$$\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^{\dagger}. \quad (3)$$

The symbol "./" denotes element-by-element division

$$\begin{aligned} \|\bar{x}\|^2 &= \frac{1}{N}\tilde{x}^*(\mathbf{FCF}^{-1})^{-1}\tilde{x}^T \\ &\approx \frac{1}{Nf_s}\tilde{x}(\tilde{x}^{\dagger}./\bar{S}^T), \end{aligned} \quad (6)$$

where $\bar{S}$ is the two-sided power spectral density of the noise

$$\bar{S} = \frac{1}{f_s}\boldsymbol{F}\bar{\phi}$$

(Where $\bar{\phi}$ is a circular sequence)

$$S_i = \frac{1}{Nf_s}E[|\tilde{n}_i|^2]$$

Convenient definition

$$\delta_f \sum_{m=0}^{N-1} S_m = \sigma^2,$$

Power Spectral Density

The definitions are more convenient if one switches to continuous time

$$S_n(f) = \int_{-\infty}^{\infty} d\tau\, \phi(\tau) e^{-2\pi i f \tau}$$

$S_n(f)$ is real and symmetric as $\phi(\tau)$ is an even function

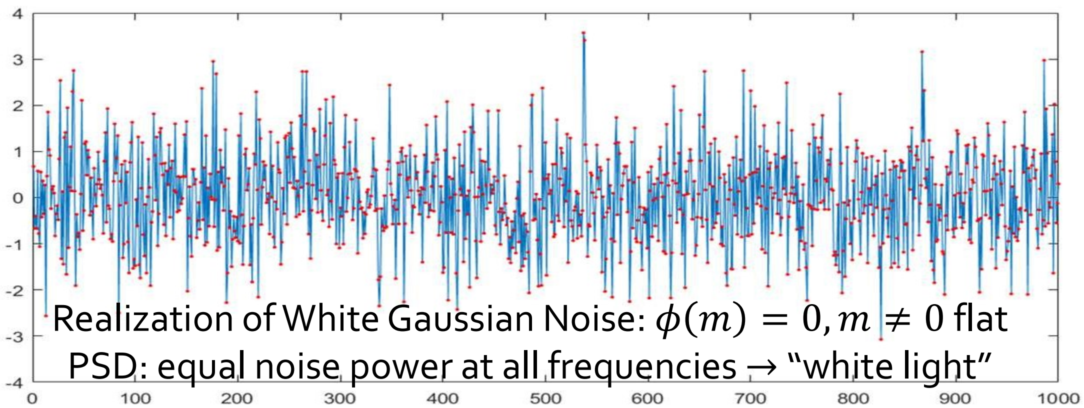
(Sign of phase in the Fourier transform does not matter)

Physical interpretation

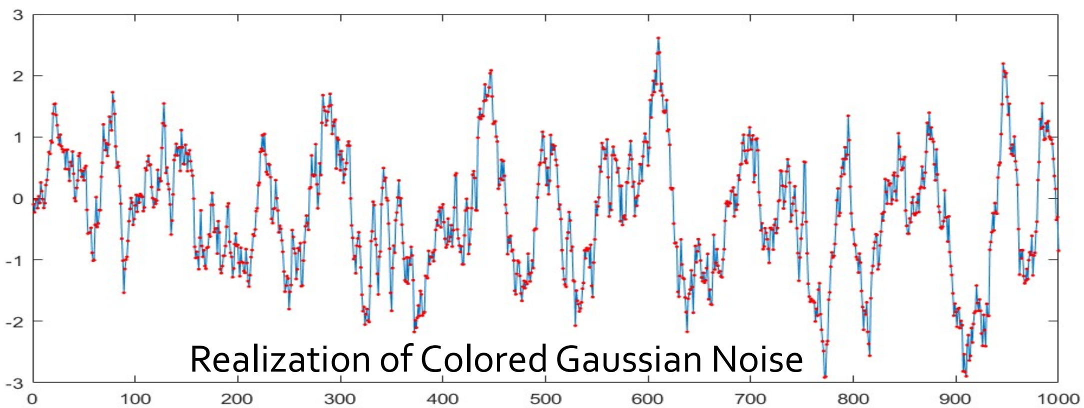
The variance of the noise process $\sigma^2 = \phi(0) = \int_{-\infty}^{\infty} df\, S_n(f)$

Hence, $S_n(f)df$ can be interpreted as the noise variance contributed by the band $[f, f + df]$

Gaussian noise nomenclature



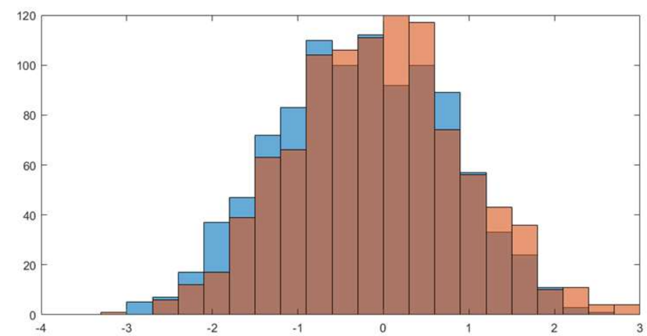
Realization of White Gaussian Noise: $\phi(m) = 0, m \neq 0$ flat PSD: equal noise power at all frequencies → “white light”



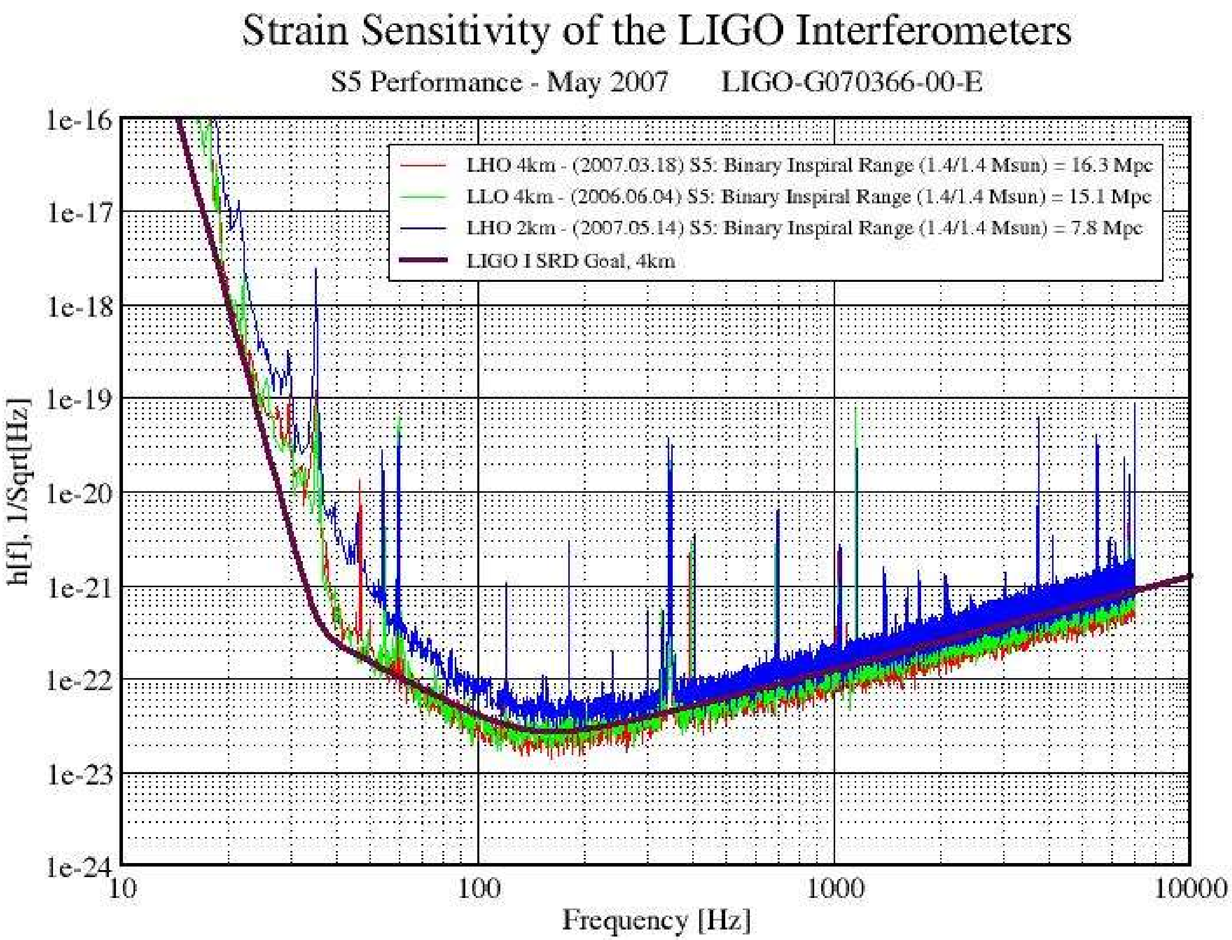
Realization of Colored Gaussian Noise

The terms ‘white’ and ‘colored’ refer to the shape of the PSD

Here, the two noise processes have the same marginal pdf $\mathrm{X_i} \sim N(0,1) \Rightarrow$ marginal pdf is not enough to describe noise



$\sqrt{S(f)}$


Strain Sensitivity of the LIGO Interferometers
S5 Performance - May 2007 LIGO-G070366-00-E
LHO 4km - (2007.03.18) S5: Binary Inspiral Range (1.4/1.4 Msun) = 16.3 Mpc
LLO 4km - (2006.06.04) S5: Binary Inspiral Range (1.4/1.4 Msun) = 15.1 Mpc
LHO 2km - (2007.05.14) S5: Binary Inspiral Range (1.4/1.4 Msun) = 7.8 Mpc
LIGO I SRD Goal, 4km
h[f], 1/Sqrt[Hz]
1e-16
1e-17
1e-18
1e-19
1e-20
1e-21
1e-22
1e-23
1e-24
10
100
1000
10000
Frequency [Hz]

Estimation problem in GW

$\bar{x}$: given data from GW detector

If a signal is present

$\bar{x} = \bar{h}(\Theta) + \bar{n}$

$\bar{h}$: Signal time series

Θ : Parameters defining the signal

$\bar{n}$: noise realization

Let $p_{\bar{X}}(\bar{x})$ be the pdf of noise.

Then, $p_{\bar{X}}(\bar{x};\Theta) = p_{\bar{X}}(\bar{x} - \bar{h}(\Theta))$

Examples

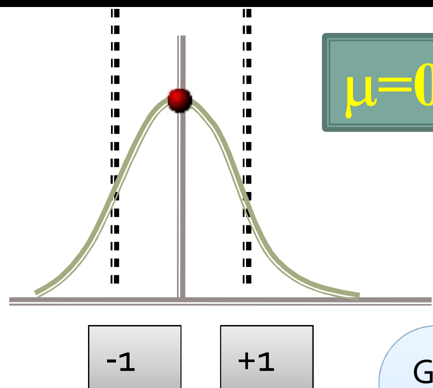
$$\bar{h} = h_0 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \rightarrow \text{"dc signal"}$$

Θ Is h_0

$$h_k = A_0 \sin(\omega_0 k \Delta)$$

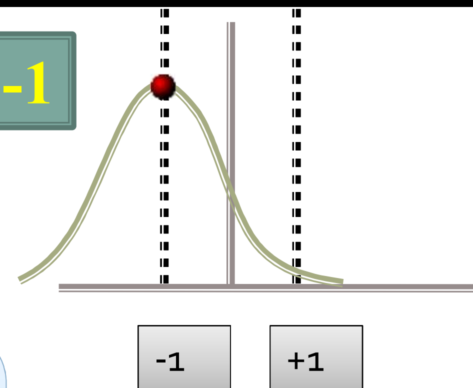
Θ Is (A_0, ω_0)

A simple estimation problem

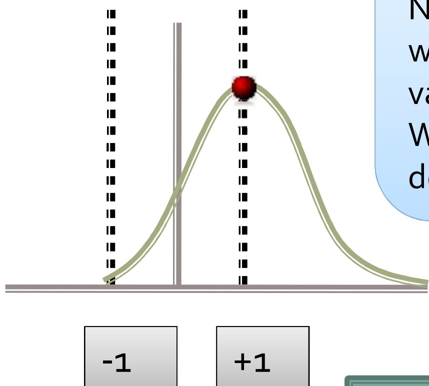


$\mu=0$

$\mu=-1$

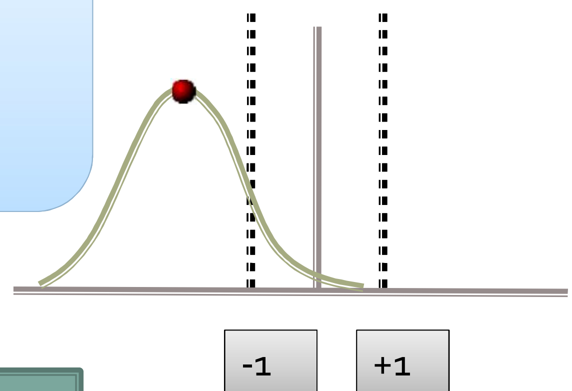


Given data: 0.5, -0.08, 1.2, 1.5, 0.8,1.4 , -0.1,0.6
Numbers drawn from a normal pdf with unknown mean but known variance ($\Theta=(\mu,\sigma)$, μ is unknown)
Which of the mean values best describes the data?



$\mu=+1$

$\mu=-2$



Likelihood

- In the first example, our judgment was based on which pdf appeared to be more "likely" as the correct one
- One way to make this idea mathematically precise is the to use the **likelihood function**
- Joint pdf of the data: $p_{\bar{X}}(\bar{x};\Theta)$
- Likelihood function: consider the joint pdf as a function of Θ for fixed x (data)
  - Alternative notation: $\Lambda(\Theta;\bar{x})$
- A high likelihood value means the corresponding pdf gives higher probability of occurrence for the data

In example 1: estimating the mean of the normal pdf,
The joint pdf of the data is

$$p_{\bar{X}}(\{x_1, x_2, \ldots, x_8\}; \mu) = \prod_{i=1}^{8} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

$$\Lambda(\mu; \{x_1, x_2, \ldots, x_8\}) = \prod_{i=1}^{8} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \mu)^2\right)$$

Which μ gives highest value of Λ?

Example of likelihood function: dc signal in WGN

Data (realization)

$$\overline{x} = \{x_1, x_2, \ldots, x_N\}$$

Parameter μ

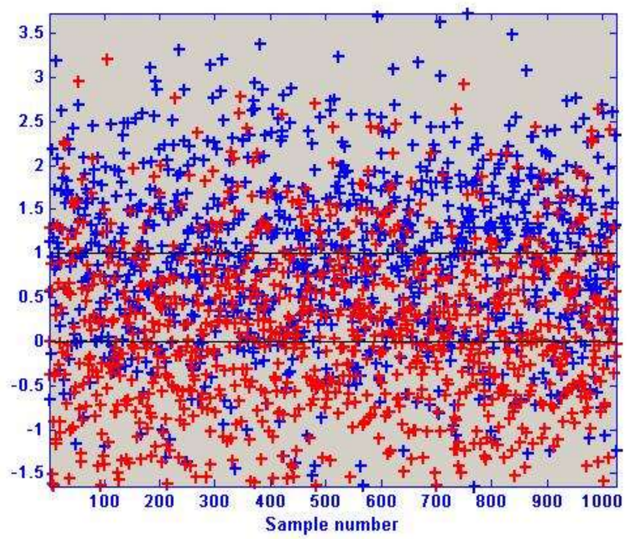
$$\overline{\mu} = \mu\,(1,1,\ldots,1)^T$$

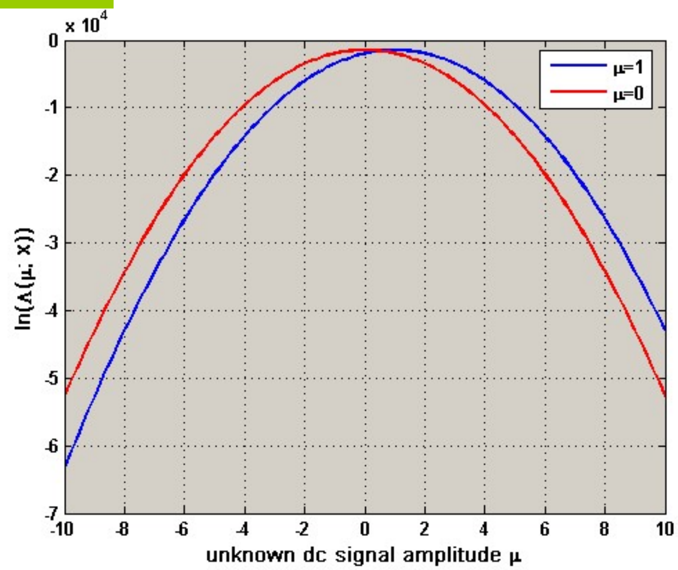
$1 \times N$ vector

"dc signal"

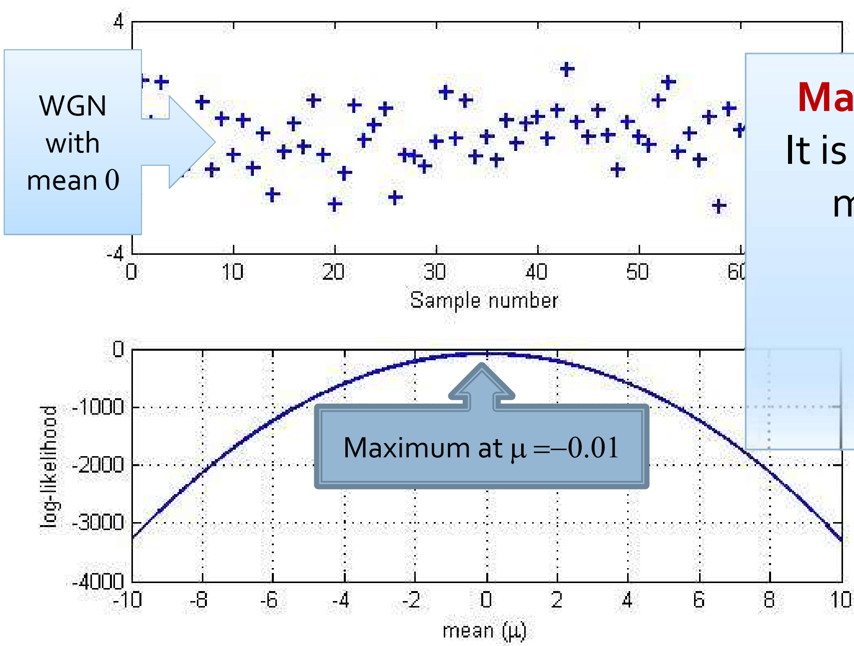
Likelihood

$$\frac{1}{\left(\sqrt{2\pi}\right)^N} \exp\left(-\frac{1}{2}(\overline{x}-\overline{\mu})^T(\overline{x}-\overline{\mu})\right)$$





Maximum Likelihood Estimator



Maximum Likelihood Estimator (MLE)

It is the point at which the Likelihood has maximum value

$\hat{\Theta}$ is such that

$$\max_{\Theta}\left(\Lambda(\Theta;\bar{x})\right) = \Lambda(\hat{\Theta};\bar{x})$$

Instead of maximizing the likelihood, we can use any monotonic function of the likelihood $\rightarrow$ e.g., $\ln(\Lambda(x;\Theta))$ (“log-likelihood”)

MLE for GW signals in Gaussian noise

GW data : $\bar{x} = \bar{h}(\Theta) + \bar{n}$

$\bar{n}$: realization of Gaussian, stationary noise $\quad p_{\bar{X}}(\bar{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\bar{x}^T \mathbf{C}^{-1} \bar{x}\right)$

Let $\langle \bar{z}, \bar{y} \rangle = \bar{z}^T \mathbf{C}^{-1} \bar{y}$, where $\mathbf{C}$ is the covariance matrix of the noise,

and $\|\bar{z}\|^2 = \langle \bar{z}, \bar{z} \rangle$. Then, the likelihood is $\quad \Lambda(\Theta; \bar{x}) = p_{\bar{X}}(\bar{x} - \bar{h}(\Theta)) = \mathrm{N} \exp\left(-\frac{1}{2}\|\bar{x} - \bar{h}(\Theta)\|^2\right)$

$$\ln \Lambda(\Theta; \bar{x}) = \ln \mathrm{N} - \frac{1}{2}\|\bar{x} - \bar{h}(\Theta)\|^2 \qquad \mathrm{N} = \frac{1}{(2\pi)^{N/2} |\mathbf{C}|^{1/2}}$$

$$= \left(\ln \mathrm{N} - \frac{1}{2}\|\bar{x}\|^2\right) + \langle \bar{x}, \bar{h}(\Theta) \rangle - \frac{1}{2}\|\bar{h}(\Theta)\|^2$$

MLE for Θ : find $\hat{\Theta}$ such that $\max_{\Theta}\left(\ln \Lambda(\Theta; \bar{x})\right) = \ln \Lambda(\hat{\Theta}; \bar{x})$

Or, get $\hat{\Theta}$ by maximizing $\lambda(\Theta; \bar{x}) = \langle \bar{x}, \bar{h}(\Theta) \rangle - \frac{1}{2}\|\bar{h}(\Theta)\|^2$

Least Squares

- Least squares estimation is simply MLE for the case of Gaussian noise
- Maximizing Log-likelihood

$$\ln \Lambda(\bar{x} \,|\Theta) = \ln N \; - \frac{1}{2}\left\|\bar{x} - \bar{h}(\Theta)\right\|^2$$

- Is the same as minimizing Least squares

$$\left\|\bar{x} - \bar{h}(\Theta)\right\|^2$$

- For White Gaussian Noise: $\phi(m) = 0, m \neq 0$ (Random variables in the sequence are statistically independent) $\Rightarrow C_{ij} = \delta_{ij} \Rightarrow$ $\left\|\bar{x} - \bar{h}(\Theta)\right\|^2 = \sum_{k=0}^{N-1}\left(x_k - h_k(\Theta)\right)^2$

Matched filtering for Colored noise

- Maximizing log-likelihood (= minimizing least squares) is equivalent to maximizing
$$\left\langle \bar{x}, \bar{h}(\Theta) \right\rangle - \frac{1}{2} \left\| \bar{h}(\Theta) \right\|^2 = \left\langle \bar{x}, \bar{h}(\Theta) \right\rangle - \frac{1}{2} \langle \bar{h}(\Theta), \bar{h}(\Theta) \rangle$$
- (See Lecture 2)
- Everything stays the same except
$$\langle \bar{x}, \bar{y} \rangle = \frac{1}{N f_s} \tilde{x} (\tilde{y}^\dagger \, ./ \, \bar{S}^T)$$

Wiener-Khinchin theorem

Input noise process $x(t)$
with PSD $S_{in}(f)$

Linear time invariant system

$$y(t) = \int_{-\infty}^{\infty} x(t')\, h(t-t')dt'$$

$h(t)$: system impulse response

$T(f) = \int_{-\infty}^{\infty} dt\; h(t)e^{-2\pi i\, ft}$: System transfer function

Output noise process
$y(t)$ with PSD $S_{out}(f)$

$$S_{out}(f) = S_{in}(f)|T(f)|^2$$

Generation of colored noise

- Use Wiener-Khinchin theorem
- Pass white noise through an LTI system
- Discrete time: Pass white noise sequence through a digital filter
- Example: $y_i = b_0 x_i + b_1 x_{i-1} + \cdots + b_m x_{i-m}$ for some coefficients $\bar{b}$ and white noise sequence $\bar{x}$
- Fourier domain: Multiply discrete fourier transform of white noise sequence with $\sqrt{S(f)}$ where $S(f)$ is the desired output PSD (Caution: this will generate a circular auto-covariance noise sequence!)