

Decision and Estimation in Data Processing

Chapter I. Random Signals

I.1 Random variables

- ▶ A **random variable** is a variable that holds a value produced by a (partially) random phenomenon
 - ▶ basically it is a *name* attached to an arbitrary value
 - ▶ short notation: r.v.
- ▶ Typically denoted as X , Y etc..
- ▶ Examples:
 - ▶ X = The value of a dice
 - ▶ V_{in} = The value of the voltage in one point of a circuit

Off-topic: Glossary

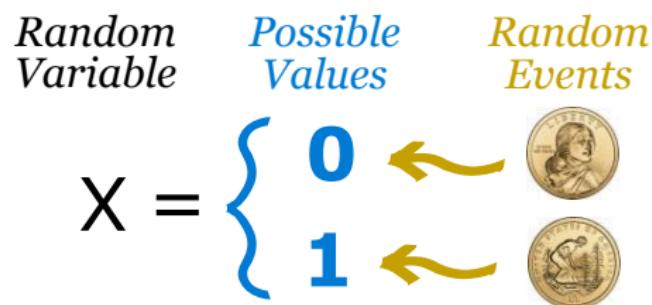
- ▶ "*i.e.*" = *id est* = "that is" = "adică"
- ▶ "*e.g.*" = *ex ampli gratia* = "for example" = "de exemplu"

- ▶ **A realization** of a random variable = one possible value it can take
 - ▶ e.g. the value 3 of a dice
 - ▶ at different times, one may get different realizations
- ▶ **Sample space** Ω = the set of all values that can be taken by a random variable X
 - ▶ i.e. the set of all possible realizations
- ▶ Example: rolling a dice
 - ▶ The r.v. is denoted as X
 - ▶ We might get a realization $X = 6$
 - ▶ But we could have got any value from the sample space

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Rolling a die

- Random variable X = "the face obtained by throwing a coin"



(image from <https://www.mathsisfun.com/data/random-variables.html>)

- ▶ **Discrete** random variable: if Ω is a discrete set
 - ▶ Example: value of a dice
- ▶ **Continuous** random variable: if Ω is a continuous set
 - ▶ Example: a voltage value

Why random variables?

- ▶ Random variables are a great model for **noise**
- ▶ Examples:
 - ▶ Measure a voltage in a circuit
 - ▶ Measure several times, the value is never precisely the same. The values always *varies* a little.
 - ▶ i.e. it is affected by noise

Probability Mass Function

- ▶ Consider a **discrete** r.v. A
- ▶ The **probability mass function (PMF)** = the probability that A has value x

$$w_A(x) = P\{A = x\}$$

- ▶ Also known as the **distribution** of A
- ▶ Example: what is the PMF of a dice? Plot on board.

Computing probability based on PMF

- ▶ Probability that A is equal to some value v

$$P\{A = v\} = w_A(v)$$

- ▶ Probability that A is between a and b (including):

$$P\{a \leq A \leq b\} = \sum_{x=a}^b w_A(x)$$

Cumulative Distribution Function

- ▶ The **cumulative distribution function (CDF)** = the probability that the value of A is smaller or equal than x

$$F_A(x) = P\{A \leq x\}$$

- ▶ In Romanian: "*funcție de repartitie*"
- ▶ Example: what is the CDF of a dice? Plot on board.
- ▶ For discrete r.v., the CDF is "stairwise"

Computing probability based on CDF

- ▶ Probability that A is equal to some value v

$$P\{A = v\} = F_A(v) - F_A(v - 1)$$

- ▶ Probability that A is between a and b (including):

$$P\{a \leq A \leq b\} = F_A(b) - F_A(a - 1)$$

Relation between PMF and CDF

- ▶ CDF is the *cumulative sum* (i.e. the integral) of PMF

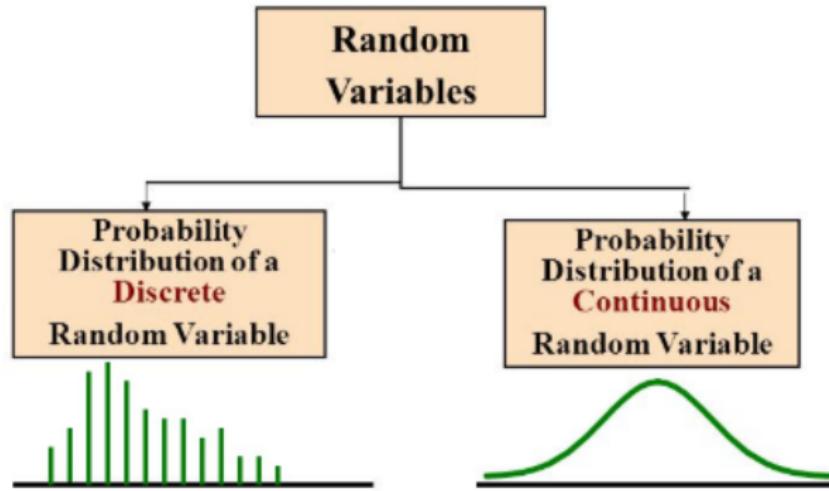
$$F_A(x) = \sum_{t=-\infty}^{t=x} w_A(t)$$

- ▶ Example for dice: easy to notice graphically

Probability Density Function

- ▶ Consider a **continuous** r.v. A
 - ▶ assume it takes values in some interval $[a, b]$
- ▶ The **Probability Density Function (PDF)** of A = probability that the value of A is in a small vicinity ϵ around x , divided by ϵ
- ▶ Denoted as $w_A(x)$, also known as **the distribution** of A
- ▶ Informally, the PDF gives the probability that the value of A is **close to x**

Continuous and discrete random variables



(image from "Probability Distributions: Discrete and Continuous", Seema Singh, <https://towardsdatascience.com/probability-distributions-discrete-and-continuous-7a94ede66dc0>)

Probability of an exact value

- ▶ The probability that a continuous r.v. A is **exactly** equal to a value x is **zero**
 - ▶ because there are an infinity of possibilities (continuous)
 - ▶ That's why we can't define a probability mass function like for discrete r.v.
- ▶ That's why the PDF says **in a small vicinity** around some value x , and not precisely equal to x

Computing probability based on PDF

- ▶ Probability that A is equal to some value v is always 0

$$P\{A = v\} = 0$$

- ▶ Probability that A is between a and b = integral of PDF from a to b :

$$P\{a \leq A \leq b\} = \int_a^b w_A(x)dx$$

Cumulative Distribution Function

- ▶ The **cumulative distribution function (CDF)** = the probability that the value of A is smaller or equal than x

$$F_A(x) = P\{A \leq x\}$$

- ▶ In Romanian: "*funcție de repartiție*"
- ▶ Same definition as for discrete r.v.

Computing probability based on CDF

- ▶ Probability that A is between a and b :

$$P\{a \leq A \leq b\} = F_A(b) - F_A(a)$$

- ▶ Doesn't matter if we consider closed or open interval
 - ▶ $[a, b]$ or (a, b)
 - ▶ why?

Relation between PDF and CDF

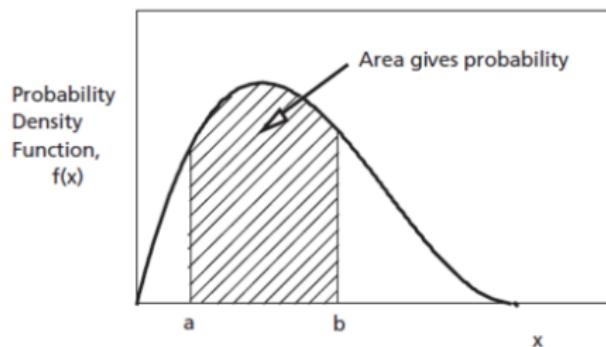
- CDF is **the integral** of PMF
- PDF is **the derivative** of CDF

$$F_A(x) = \int_{-\infty}^x w_A(x)dx$$

$$\begin{aligned}w_A(x) &= \frac{dF_A(x)}{dx} \\&= \lim_{\epsilon \rightarrow 0} \frac{F_A(x + \epsilon) - F_A(x - \epsilon)}{2\epsilon} \\&= \lim_{\epsilon \rightarrow 0} \frac{P(A \in [x - \epsilon, x + \epsilon])}{2\epsilon}\end{aligned}$$

Graphical interpretation

- ▶ Probability that a continuous r.v. A is between a and b is **the area below the PDF**
 - ▶ i.e. the integral from a to b
- ▶ Probability that A is exactly equal to a certain value is zero
 - ▶ the area below a single point is zero



(image from "<https://intellipaat.com/blog/tutorial/statistics-and-probability-tutorial/probability-distributions-of-continuous-variables/>")

Discrete vs continuous r.v.

Comparison of discrete vs continuous random variables:

- ▶ The CDF $F_A(x)$ is defined identically, means same thing
- ▶ The PDF/PMF $w_A(x)$ is the derivative of CDF
 - ▶ for continuous r.v.:
 - ▶ it is a proper derivative
 - ▶ it means probability to be "around" x
 - ▶ for discrete r.v.:
 - ▶ sort of "discrete derivative"
 - ▶ it means probability to be exactly equal to x

Properties of random variables

CDF:

- ▶ The CDF is always ≥ 0
- ▶ The CDF is always monotonously increasing (non-decreasing)
- ▶ The CDF starts from 0 and goes up to 1

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

PDF/PMF:

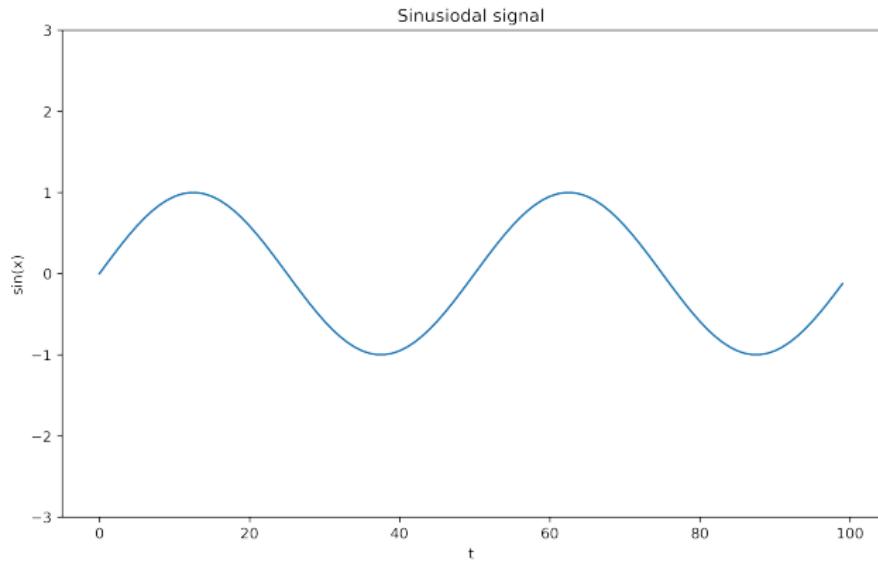
- ▶ The PDF/PMF are always ≥ 0
- ▶ Integral/sum over all of the PDF/PMF = 1

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

$$\sum_{x=-\infty}^{\infty} w_A(x) = 1$$

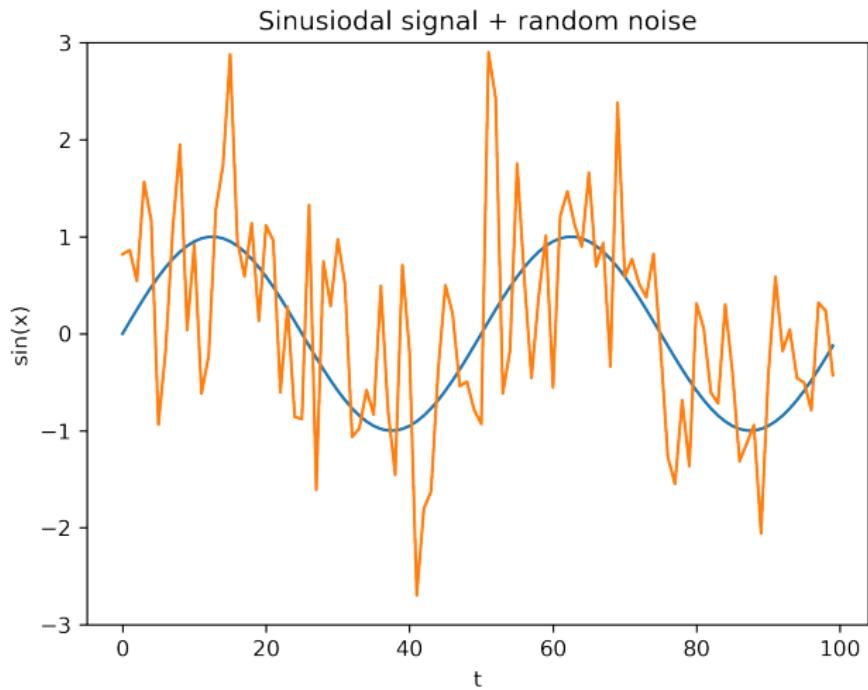
Different distributions

► Normal sine signal



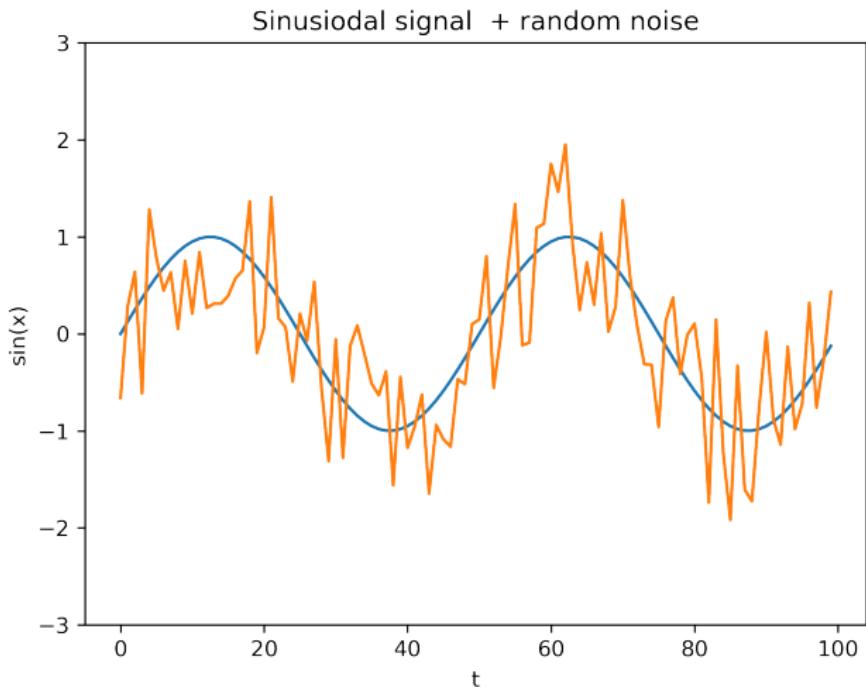
Different distributions

- ▶ Sine + noise 1 (normal, $\mu = 0, \sigma^2 = 1$)



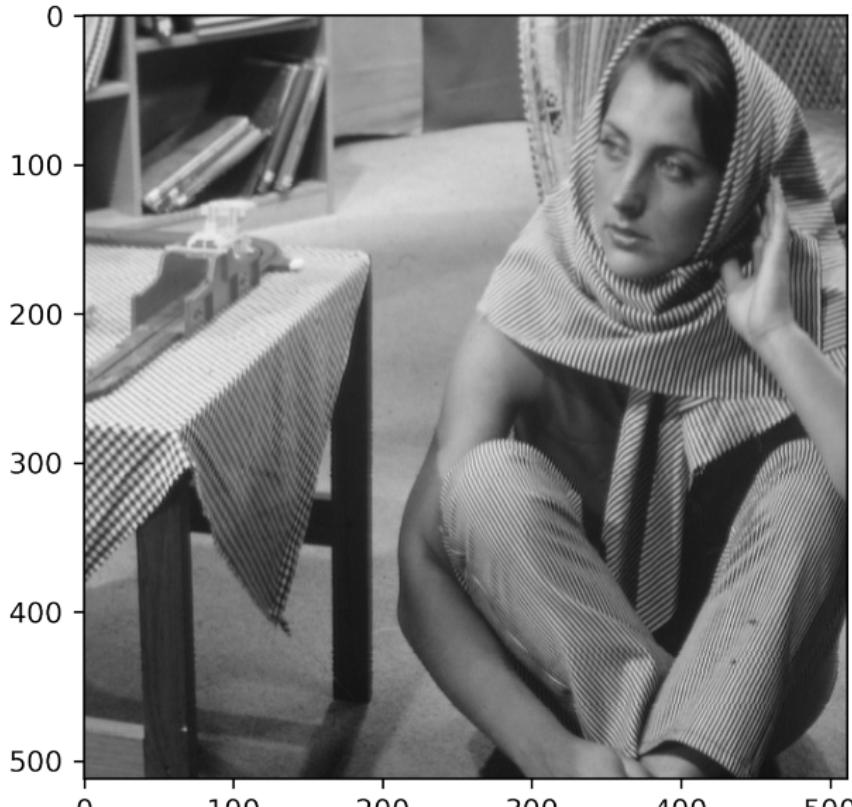
Different distributions

- ▶ Sine + noise 2 (uniform $\mathcal{U}[-1, 1]$)
- ▶ What's different? The distribution type



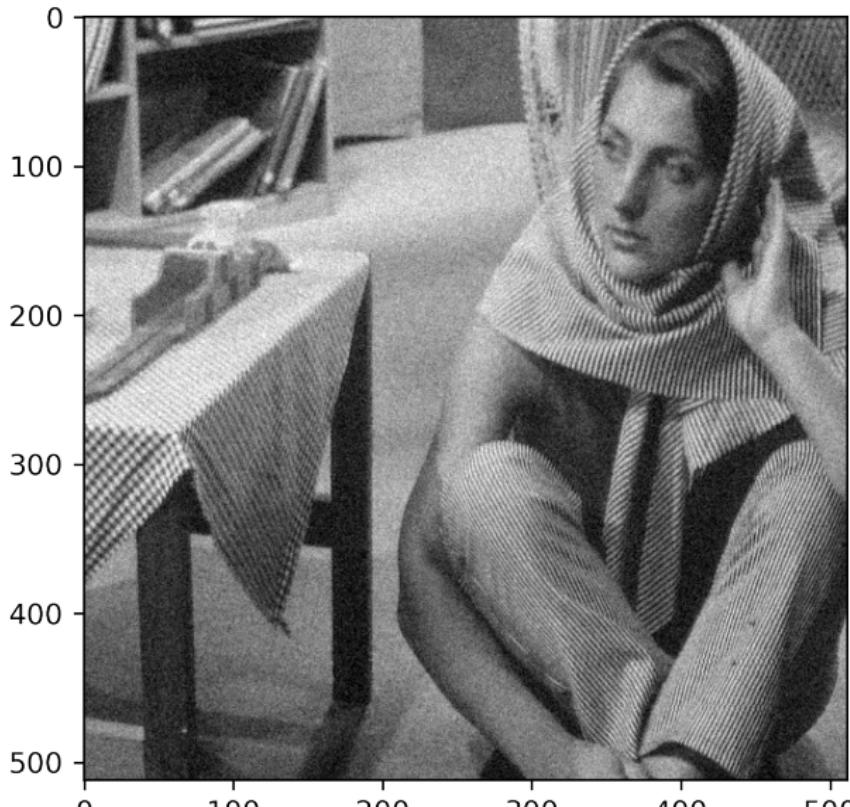
Different distributions

► Clean Image



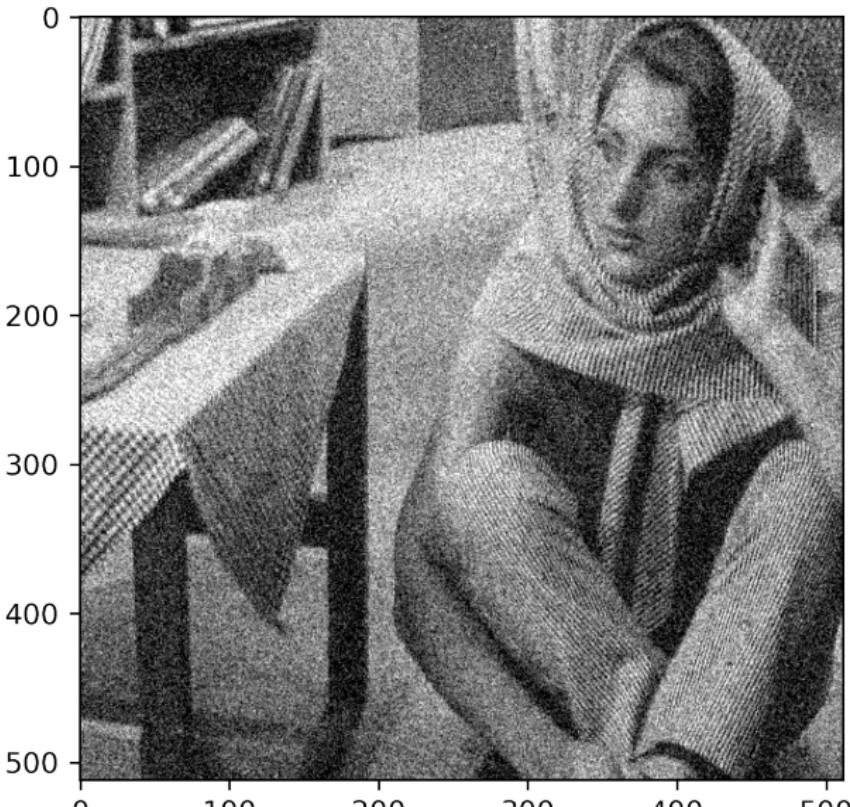
Different distributions

- ▶ Image + noise (normal, $\mu = 0, \sigma^2 = 1$)



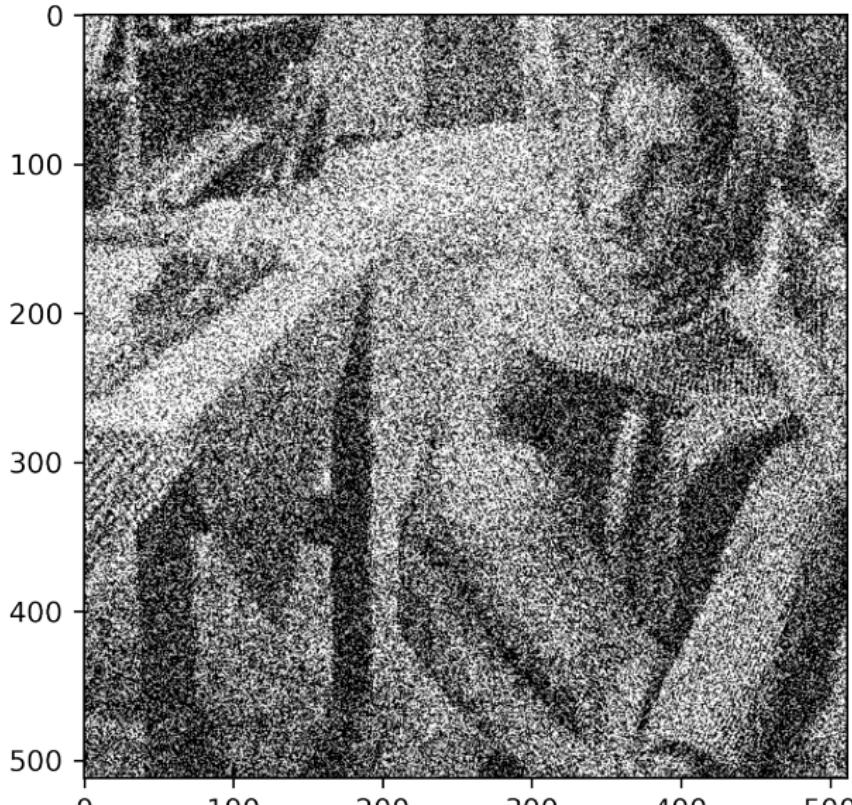
Different distributions

- ▶ Image + larger noise (normal, $\mu = 0, \sigma^2 = 10$)



Different distributions

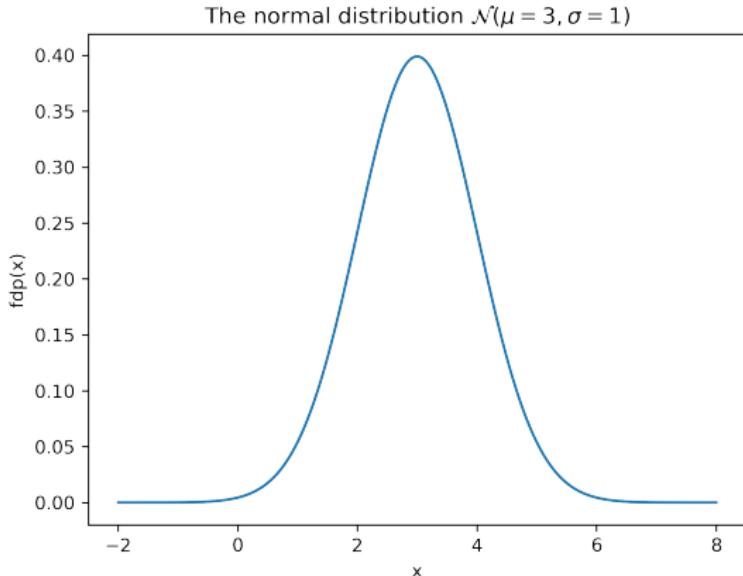
- ▶ Image + noise (uniform, $\mathcal{U}[-5, 5]$)



The normal distribution

► Probability density function

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



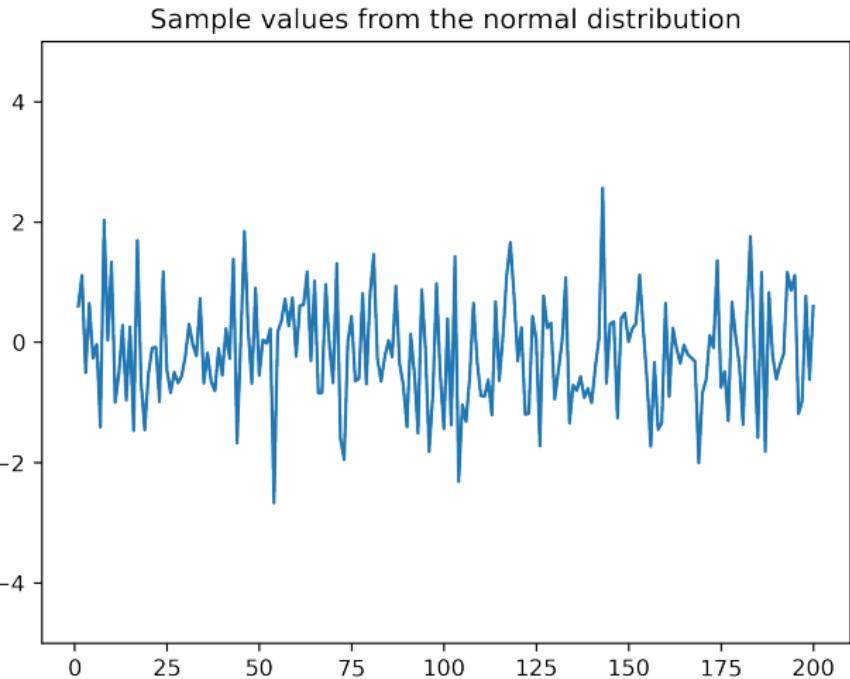
The normal distribution

- ▶ Has two parameters:
 - ▶ **Average value** μ = "center" of the function
 - ▶ **Standard deviation** σ = "width" of the function
 - ▶ Small σ = narrow and tall
 - ▶ Big σ = wide and low
- ▶ The front constant is just for normalization (ensures that integral = 1)
- ▶ Extremely often encountered in real life
- ▶ Any real value is possible ($w_A(x) > 0, \forall x \in \mathbb{R}$)
- ▶ Usually denoted as $\mathcal{N}(\mu, \sigma^2)$

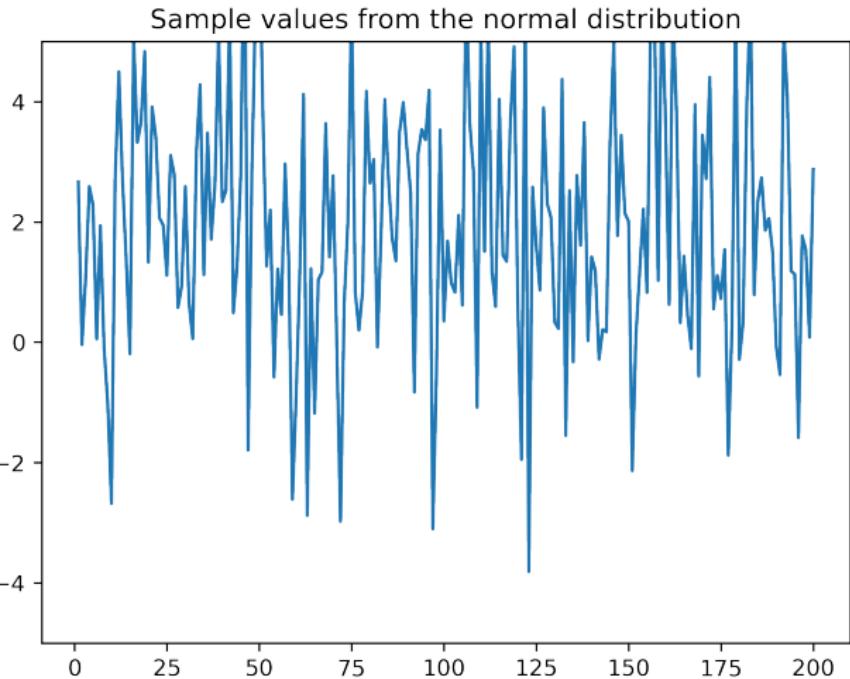
The normal distribution

- ▶ The distribution decreases as x gets farther from μ
 - ▶ Because of the term $-(x - \mu)^2$ at the exponent
 - ▶ Most likely values: around μ ($x - \mu = 0$)
 - ▶ Values closer to μ are more likely, values farther from μ are less likely
- ▶ The function describes a preference for values around μ , with decreasing preference when getting farther from μ

Example of values from the normal distribution ($\mu=0$,
 $\sigma=1$)



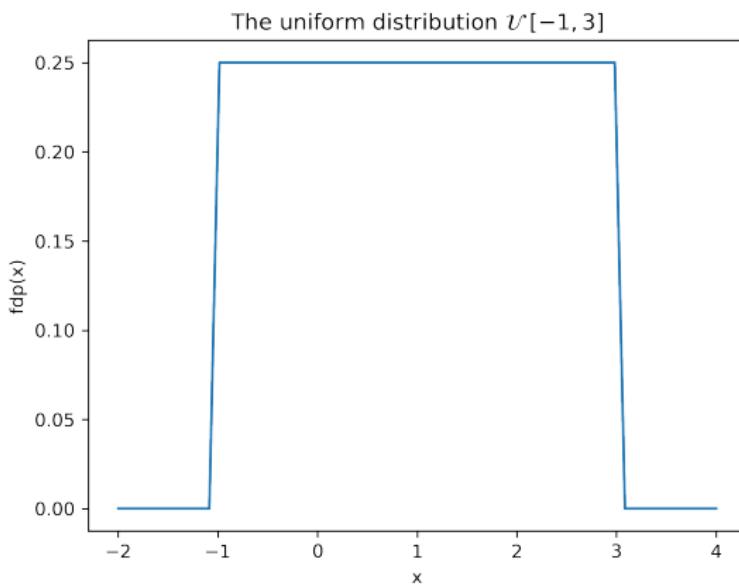
Example of values from the normal distribution ($\mu=2$,
 $\sigma=4$)



The uniform distribution

- The probability density function = a constant, between two endpoints

$$w_A(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{elsewhere} \end{cases}$$



The uniform distribution

- ▶ Has two parameters: the limits a and b of the interval
- ▶ The "height" of the function is $\frac{1}{b-a}$
 - ▶ in order for the integral to be 1
- ▶ Only values from the interval $[a, b]$ are possible
 - ▶ value cannot be outside interval (probability is 0)
- ▶ Denoted as $\mathcal{U}[a, b]$

Other distributions

- ▶ Many other distributions exist, relevant for particular applications

Computing probabilities for the normal distribution

- ▶ How to compute \int_a^b for a normal distribution?
 - ▶ Can't be done with algebraic formula, non-elementary function
- ▶ Use *the error function*:

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

- ▶ The CDF of a normal distribution $\mathcal{N}(\mu, \sigma^2)$
$$F_A(x) = \frac{1}{2} \left(1 + \text{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right) \right)$$
- ▶ The values of $\text{erf}()$ are available / are computed numerically
 - ▶ e.g. on Google, search for $\text{erf}(0.5)$
 - ▶ Other useful values:
 - ▶ $\text{erf}(-\infty) = -1$
 - ▶ $\text{erf}(\infty) = 1$

Exercise

Exercise:

- ▶ Let X be a r.v. with distribution $\mathcal{N}(3, 2)$. Compute the probability that $X \in [2, 4]$

Sum of constant + random variable

- ▶ Consider a random variable A
- ▶ What is $B = 5 + A$?

Answer:

- ▶ B is also a random variable
- ▶ B has same type of distribution, but the function is "shifted" by 5 to the right

Example:

- ▶ A is normal variable with $w_A(x) = \mathcal{N}(\mu = 3, \sigma^2 = 2)$
- ▶ What is the distribution of $B = 5 + A$?
- ▶ Answer: $w_B(x) = \mathcal{N}(\mu = 8, \sigma^2 = 2)$

- ▶ A function applied to a r.v. produces another r.v.
- ▶ Examples: if B is a r.v. with distribution $\mathcal{U}[0, 10]$, then
 - ▶ $C = 5 + B$ is another r.v., with distribution $\mathcal{U}[5, 15]$
 - ▶ $D = B^2$ is also another r.v.
 - ▶ $E = \cos(B)$ is also another r.v.
- ▶ Reason: since B is random, the values B, C, D are also random
- ▶ A, B, C, D are *not independent*
 - ▶ A certain value of one of them automatically implies the value of the others

Multiple random variables

- ▶ Consider a system with two continuous r.v. A and B
- ▶ What is the probability that the pair (A, B) has values around (x, y) ?
- ▶ Distribution of the values of (A, B) is described by:
 - ▶ joint probability density function $w_{AB}(x, y)$
 - ▶ joint cumulative density function $F_{AB}(x, y)$

Multiple random variables

- ▶ Joint cumulative distribution function:

$$F_{AB}(x, y) = P \{A \leq x \cap B \leq y\}$$

- ▶ Joint probability density function:

$$w_{AB}(x, y) = \frac{\partial^2 P_{AB}(x, y)}{\partial x \partial y}$$

- ▶ The joint PDF gives the probability that the value of the pair (A, B) is in a vicinity of (x, y)
- ▶ Similar for discrete random variables

$$w_{AB}(x, y) = P \{A = x \cap B = y\}$$

- ▶ Two v.a. A and B are **independent** if the value of one of them does not influence in any way the value of the other
- ▶ For independent r.v., the probability that A is around x and B is around y is **the product** of the two probabilities

$$w_{AB}(x, y) = w_A(x) \cdot w_B(y)$$

- ▶ Relation holds for CDF / PDF / PMF, continuous or discrete r.v.
- ▶ Same for more than two r.v.

Exercise:

- ▶ Compute the probability that three r.v. X , Y and Z i.i.d. $\mathcal{N}(-1, 1)$ are all positive simultaneously
 - ▶ **i.i.d** = "independent and identically distributed"

Multiple normal variables

- ▶ Consider a set of N normal r.v. (A_1, \dots, A_N) , with different μ_i , but same σ
- ▶ Then probability that (A_1, \dots, A_N) is around (x_1, \dots, x_N) is

$$w_{A_1, \dots, A_N}(x_1, \dots, x_N) = \frac{1}{(\sigma\sqrt{2\pi})^N} e^{\frac{(x_1 - \mu_1)^2 + \dots + (x_N - \mu_N)^2}{2\sigma^2}}$$

- ▶ The probability depends on the **Euclidean distance** between $\mathbf{x} = (x_1, \dots, x_N)$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)$

Euclidean distance

- ▶ **Euclidean (geometric) distance** between two N-dimensional vectors:

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + \dots + (u_N - v_N)^2}$$

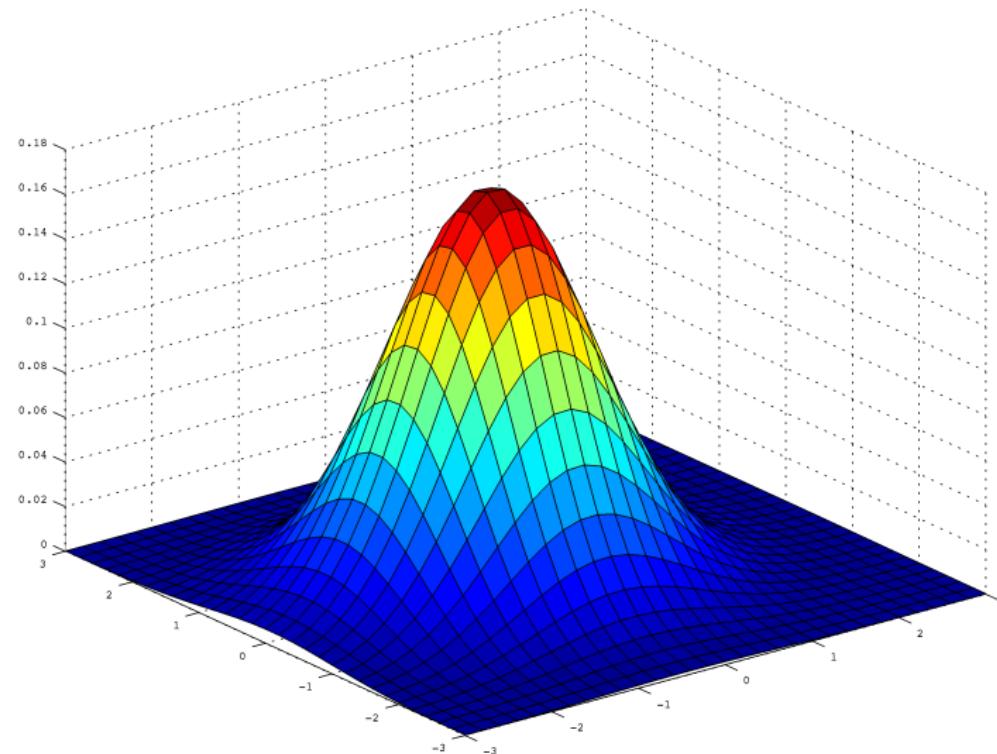
- ▶ One-dimensional: $\|\mathbf{u} - \mathbf{v}\| = |u - v|$
- ▶ 2D: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$
- ▶ 3D: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$
- ▶ ...
- ▶ N-dimensional: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\sum_{i=1}^N (u_i - v_i)^2}$
- ▶ ...
- ▶ Continuous signals: $\|\mathbf{u} - \mathbf{v}\| = \sqrt{\int_{-\infty}^{\infty} (u(t) - v(t))^2 dt}$

Multiple normal variables

- ▶ Probability of N normal random variables, independent, with same σ but possibly different μ ; depends on the **squared Euclidean distance to the mean vector** $\mu = (\mu_1, \dots, \mu_N)$
 - ▶ Close to μ : higher probability
 - ▶ Far from μ : lower probability
 - ▶ Two points at same distance from μ have same probability

2D normal distribution

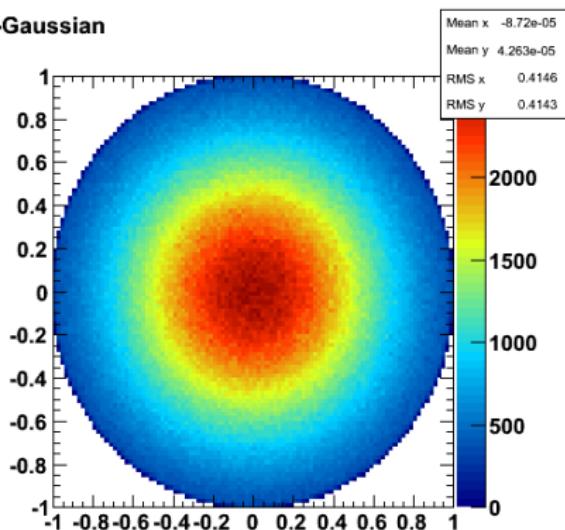
- Distribution of 2 normal random variables (2D normal distribution)



2D normal distribution - top view

- ▶ View from top
- ▶ Here, $\mu = (0, 0)$
- ▶ Probability density decreases as distance from center increases, in circles (symmetrically)

2D-Gaussian



Statistical averages

- ▶ R.v. are described by statistical averages ("moments")
- ▶ **The average value** (or "expected value", or "moment of order 1")
- ▶ Continuous r.v.:

$$\bar{A} = E\{A\} = \int_{-\infty}^{\infty} x \cdot w_A(x) dx$$

- ▶ Discrete r.v.:

$$\bar{A} = E\{A\} = \sum_{x=-\infty}^{\infty} x \cdot w_A(x)$$

- ▶ (Example: the entropy of $H(X)$ = the average value of the information)
- ▶ Usual notation: μ

Meaning of average value

- ▶ What does the average value of a random variable *mean*?
 - ▶ If we have $N \rightarrow \infty$ values drawn from this distribution, it is their average;
 - ▶ If you'd have to guess the value of a random variable value X , and you pay a price proportional to square of the error of your guess u , $(u - X)^2$, the average value μ would be the best choice, as it minimizes the overall cost:

$$\mu = \arg \min_u \int_{-\infty}^{\infty} (u - x)^2 \cdot w(x) dx$$

- ▶ Proof: on blackboard: derivate and set derivative to 0

Meaning of average value

- ▶ Values with high probability will "pull" the average value closer to them
- ▶ For distributions with symmetrical shape (like the normal distribution), the average value is always the middle value
 - ▶ Proof: both sides of the function "pull" the average value equally, so it remains in the middle
- ▶ For the normal distribution, $\bar{X} = \mu$ (obvious from notation)
- ▶ For the uniform distribution $\mathcal{U}[a, b]$, $\bar{X} = \frac{a+b}{2}$ (middle)

Properties of the average value

- ▶ Computing the average value is a **linear** operation
 - ▶ because the underlying integral / sum is a linear operation
- ▶ If A and B are two (independent) random variables:
- ▶ Linearity

$$E\{c_1A + c_2B\} = c_1E\{A\} + c_2E\{B\}$$

- ▶ Or:

$$E\{cA\} = cE\{A\}, \forall c \in \mathbb{R}$$

$$E\{A + B\} = E\{A\} + E\{B\}$$

- ▶ No proof given here

Average squared value

- ▶ **Average squared value** = average value of the squared values
- ▶ "Moment of order 2"
- ▶ Continuous r.v.:

$$\overline{A^2} = E\{A^2\} = \int_{-\infty}^{\infty} x^2 \cdot w_A(x) dx$$

- ▶ Discrete r.v.:

$$\overline{A^2} = E\{A^2\} = \sum_{-\infty}^{\infty} x^2 \cdot w_A(x)$$

- ▶ Interpretation: average of squared values = average power of a signal

Variance

- ▶ **Variance**= average squared value of the difference to the average value
- ▶ Continuous r.v.:

$$\sigma^2 = \overline{\{A - \mu\}^2} = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot w_A(x) dx$$

- ▶ Discrete r.v.:

$$\sigma^2 = \overline{\{A - \mu\}^2} = \sum_{-\infty}^{\infty} (x - \mu)^2 \cdot w_A(x)$$

- ▶ Interpretation: how much do the values vary around the average value
 - ▶ σ^2 = large: large spread around the average value
 - ▶ σ^2 = small: values are concentrated around the average value

Relation between the three values

- ▶ Relation between the average value, the average squared value, and the variance:

$$\begin{aligned}\sigma^2 &= \overline{\{A - \mu\}^2} \\ &= \overline{A^2 - 2 \cdot A \cdot \mu + \mu^2} \\ &= \overline{A^2} - 2\mu\overline{A} + \mu^2 \\ &= \overline{A^2} - \mu^2\end{aligned}$$

Sum of random variables

- ▶ Sum of two or more **independent** r.v. is also a r.v.
- ▶ Its distribution = the **convolution** of the distributions of the two r.v.
- ▶ If $C = A + B$

$$w_C(x) = w_A(x) * w_B(x)$$

- ▶ Particular case: if A and B are normal r.v., with $\mathcal{N}(\mu_A, \sigma_A^2)$ and $\mathcal{N}(\mu_B, \sigma_B^2)$, then:

- ▶ C is also a normal r.v., with $\mathcal{N}(\mu_C, \sigma_C^2)$, having:
- ▶ average = sum of the two averages: $\mu_C = \mu_A + \mu_B$
- ▶ variance = sum of the two variances: $\sigma_C^2 = \sigma_A^2 + \sigma_B^2$

I.2 Random processes

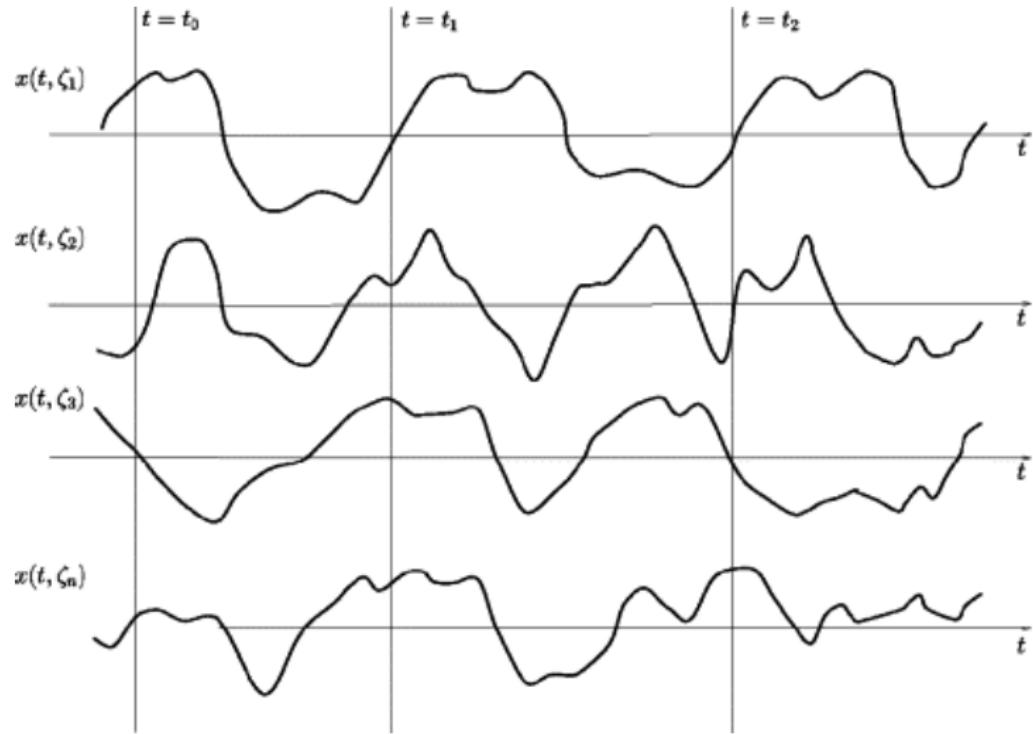
- ▶ A **random process** = a sequence of random variables indexed in time
- ▶ **Discrete-time** random process $f[n]$ = a sequence of random variables at discrete moments of time
 - ▶ e.g.: a sequence of 50 throws of a dice, the daily price on the stock market
- ▶ **Continuous-time** random process $f(t)$ = a continuous sequence of random variables at every moment
 - ▶ e.g.: a noise voltage signal, a speech signal
- ▶ Every sample from a random process is a (different) random variable!
 - ▶ e.g. $f(t_0)$ = value at time t_0 is a r.v.

- ▶ A **realization** of the random process = a particular sequence of realizations of the underlying r.v.
 - ▶ e.g. we see a given noise signal on the oscilloscope, but *we could have seen any other realization just as well*
- ▶ Typically denoted as $f^{(k)}[n]$ or $f^{(k)}(t)$
 - ▶ k indicates the particular realization that we consider
- ▶ When we consider a random process = we consider the set of all possible realizations

Random process is 2-D

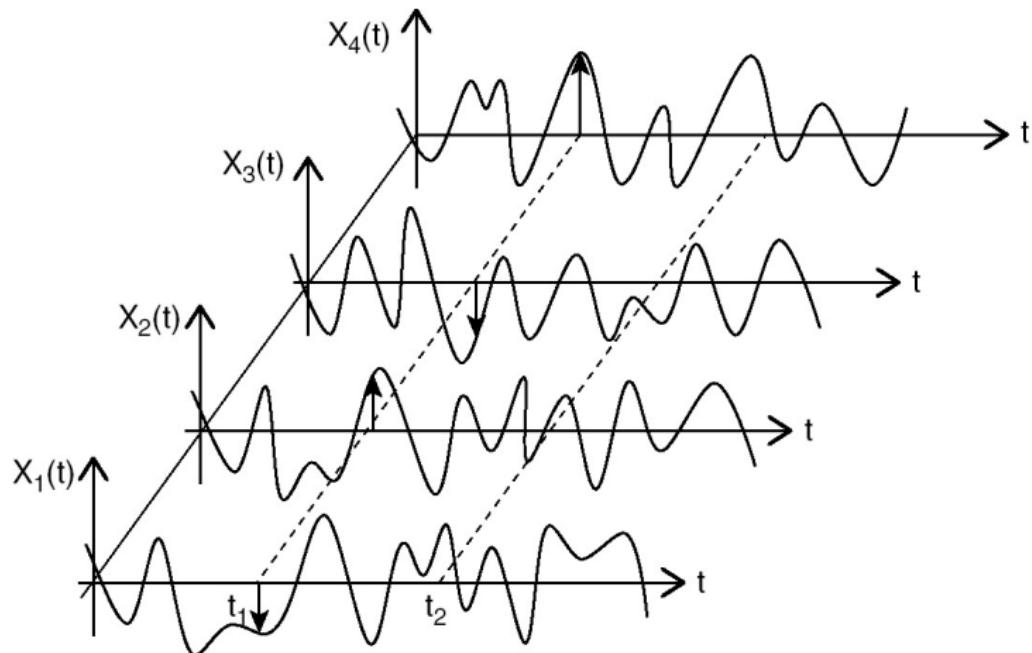
- ▶ A random process is a 2-Dimensional thing
 - ▶ $f^{(k)}[n]$ or $f^{(k)}(t)$ depends on two variables:
 - ▶ k = the particular realization
 - ▶ t or n = time

Random process is 2-D



- ▶ from "Information-Based Inversion and Processing with Applications"
Edited by Tadeusz J. Ulrych, Mauricio D. Sacchi, Volume 36,

Random process is 2-D



- ▶ from: Razdolsky, L. (2014). Random Processes. In Probability-Based Structural Fire Load (pp. 89-136). Cambridge: Cambridge University Press

Random process is 2-D

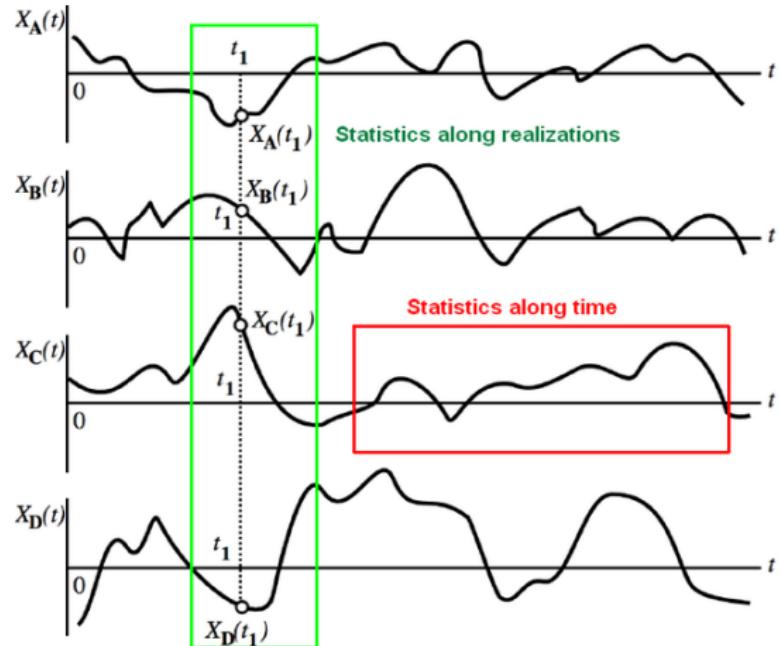


- ▶ from: <https://www.quora.com/What-is-the-difference-between-a-stationary-ergodic-and-a-stationary-non-ergodic-process>

Two types of averages

- ▶ Random processes have two types of averages:
 - ▶ **Statistical** averages = for a given time t or n , across all possible realizations
 - ▶ **Temporal** averages = for a given realization k , across all time

Two types of averages



- ▶ from: <https://www.quora.com/What-is-the-difference-between-a-stationary-ergodic-and-a-stationary-non-ergodic-process>

Distributions of order 1 of random processes

- ▶ Every sample $f(t_1)$ from a random process is a random variable
 - ▶ it is described by a **distribution of order 1**
 - ▶ has a CDF $F_1(x; t_1)$
 - ▶ has a PDF $w_1(x; t_1) = \frac{dF_1(x; t_1)}{dx}$
 - ▶ everything depends on the time moment t_1
- ▶ The sample at time t_2 is a different random variable with **possibly different** functions
 - ▶ has a different CDF $F_1(x; t_2)$
 - ▶ has a different PDF $w_1(x; t_2) = \frac{dF_1(x; t_2)}{dx}$
- ▶ These functions specify how the value of one sample is distributed
- ▶ The index w_1 indicates we consider a single random variable (distribution of order 1)
- ▶ Same for discrete-time random processes

Distributions of order 2

- ▶ A pair of random variables $f(t_1)$ and $f(t_2)$ form a system of 2 r.v.
 - ▶ they are described by a **distribution of order 2**
 - ▶ have a joint CDF $F_2(x_i, x_j; t_1, t_2)$
 - ▶ have a joint PDF $w_2(x_i, x_j; t_1, t_2) = \frac{\partial^2 F_2(x_i, x_j; t_1, t_2)}{\partial x_i \partial x_j}$
 - ▶ depend on time moments t_1 and t_2
- ▶ These functions specify how the pair of values is distributed
- ▶ Same for discrete-time random processes

Distributions of order n

- ▶ Generalize to n samples of the random process
- ▶ A set of n random variables $f(t_1), \dots, f(t_n)$ from the random process $f(t)$
 - ▶ are described by **distribution of order n**
 - ▶ have joint CDF $F_n(x_1, \dots, x_n; t_1, \dots, t_n)$
 - ▶ have joint PDF $w_n(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{\partial^2 F_n(x_1, \dots, x_n; t_1, \dots, t_n)}{\partial x_1 \dots \partial x_n}$
 - ▶ depend on time moments t_1, t_2, \dots, t_n
- ▶ These functions specify how the whole set of n values is distributed
- ▶ Same for discrete-time random processes

Random processes are characterized using **statistical** and **temporal** averages ("moments")

For continuous random processes:

1. Average value

$$\overline{f(t_1)} = \mu(t_1) = \int_{-\infty}^{\infty} x \cdot w_1(x; t_1) dx$$

2. Average squared value ("valoarea pătratică medie")

$$\overline{f^2(t_1)} = \int_{-\infty}^{\infty} x^2 \cdot w_1(x; t_1) dx$$

3. Variance ("variantă")

$$\sigma^2(t_1) = \overline{\{f(t_1) - \mu(t_1)\}^2} = \int_{-\infty}^{\infty} (x - \mu(t_1))^2 \cdot w_1(x; t_1) dx$$

- ▶ Relation between these three:

$$\begin{aligned}\sigma^2(t_1) &= \overline{\{f(t_1) - \mu(t_1)\}^2} \\ &= \overline{f(t_1)^2 - 2f(t_1)\mu(t_1) + \mu(t_1)^2} \\ &= \overline{f^2(t_1)} - \mu(t_1)^2\end{aligned}$$

- ▶ Note:

- ▶ these three values are calculated across all realizations, at time t_1
- ▶ they characterize only the sample at time t_1
- ▶ at a different time t_2 , the r.v. $f(t_2)$ is different so all average values might be different

Statistical averages for a pair of **two samples**:

4. The autocorrelation function

$$R_{ff}(t_1, t_2) = \overline{f(t_1)f(t_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 w_2(x_1, x_2; t_1, t_2) dx_1 dx_2$$

5. The correlation function (for different random processes $f(t)$ and $g(t)$)

$$R_{fg}(t_1, t_2) = \overline{f(t_1)g(t_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 y_2 w_2(x_1, y_2; t_1, t_2) dx_1 dy_2$$

► Note:

- these functions may have different values for a different pair of samples at times (t_1, t_2)

Discrete random processes

For **discrete random processes**: integral becomes sum, notation $f(t)$ becomes $f[t]$):

$$1. \overline{f[t_1]} = \mu(t_1) = \sum_{x=-\infty}^{\infty} x \cdot w_1(x; t_1)$$

$$2. \overline{f^2[t_1]} = \sum_{x=-\infty}^{\infty} x^2 \cdot w_1(x; t_1)$$

$$3. \sigma^2(t_1) = \overline{(f[t_1] - \mu(t_1))^2} = \sum_{x=-\infty}^{\infty} (x - \mu(t_1))^2 \cdot w_1(x; t_1)$$

$$4. R_{ff}(t_1, t_2) = \overline{f[t_1]f[t_2]} = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 x_2 w_2(x_1, x_2; t_1, t_2)$$

$$5. R_{fg}(t_1, t_2) = \overline{f[t_1]g[t_2]} = \sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} x_1 y_2 w_2(x_1, y_2; t_1, t_2)$$

Temporal averages

- ▶ What to do when we only have access to a single realization $f^{(k)}(t)$?
- ▶ Compute values **for a single realization $f^{(k)}(t)$, across all time moments**

Temporal averages

Temporal averages for continuous random processes:

1. Temporal average value

$$\overline{f^{(k)}(t)} = \mu^{(k)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^{(k)}(t) dt$$

2. Temporal average squared value

$$\overline{[f^{(k)}(t)]^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [f^{(k)}(t)]^2 dt$$

3. Temporal variance

$$\sigma^2 = \overline{\{f^{(k)}(t) - \mu^{(k)}\}^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T (f^{(k)}(t) - \mu^{(k)})^2 dt$$

- ▶ Relation between the three:

$$\sigma^2 = \overline{[f^{(k)}(t)]^2} - [\mu^{(k)}]^2$$

- ▶ Note:

- ▶ these values do not depend anymore on time t (integrated)

4. The temporal autocorrelation function

$$\begin{aligned} R_{ff}(t_1, t_2) &= \overline{f^{(k)}(t_1 + t)f^{(k)}(t_2 + t)} \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^{(k)}(t_1 + t)f^{(k)}(t_2 + t)dt \end{aligned}$$

5. The temporal correlation function (for different random processes $f(t)$ and $g(t)$)

$$\begin{aligned} R_{fg}(t_1, t_2) &= \overline{f^{(k)}(t_1 + t)g^{(k)}(t_2 + t)} \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^{(k)}(t_1 + t)g^{(k)}(t_2 + t)dt \end{aligned}$$

Discrete random processes

For **discrete random processes**: replace \int with \sum , T with N , and divide to $2N + 1$ instead of $2T$

$$1. \overline{f^{(k)}[t]} = \mu^{(k)} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N f^{(k)}[t]$$

$$2. \overline{[f^{(k)}[t]]^2} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N (f^{(k)}[t])^2$$

$$3. \sigma^2 = \overline{\{f^{(k)}[t] - \mu^{[k]}\}^2} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N (f^{(k)}[t] - \mu^{(k)})^2$$

4. Temporal autocorrelation:

$$\begin{aligned} R_{ff}(t_1, t_2) &= \overline{f^{(k)}[t_1 + t]f^{(k)}[t_2 + t]} \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N f^{(k)}[t_1 + t]f^{(k)}[t_2 + t] \end{aligned}$$

5. Temporal correlation:

$$\begin{aligned} R_{fg}(t_1, t_2) &= \overline{f^{(k)}[t_1 + t]g^{(k)}[t_2 + t]} \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N f^{(k)}[t_1 + t]g^{(k)}[t_2 + t] \end{aligned}$$

Finite length realizations

- ▶ We may have a realization of **finite length** (e.g. a vector with 1000 elements)
- ▶ How do we compute the temporal averages?
- ▶ Just use $\int_{t_{min}}^{t_{max}}$ or $\sum_{t_{min}}^{t_{max}}$ instead of $-\infty$ to ∞
- ▶ Example: Compute the temporal averages (all five) for the finite-length realization:

$$\{1, -1, 2, -2, 3, -3, 4, -4, 5, -5\}$$

Statistical and temporal averages

- ▶ Statistical averages are usually the ones we want
 - ▶ but they require to know the distributions $w(x)$, which is most often unknown
- ▶ In real life, typically we can only measure one realization
 - ▶ so we can only compute the temporal values for one realization
- ▶ Fortunately, in many cases statistical and temporal averages **are the same** (ergodicity)

Stationary random processes

- ▶ Until now, all the statistical averages are dependent on time
 - ▶ i.e. they might be different for a sample at t_1 and another at t_2
- ▶ **Stationary** random process = when all statistical averages are identical if we shift the time origin (e.g. delay the signal)
- ▶ Equivalent definition: if all the PDF are identical when shifting the time origin

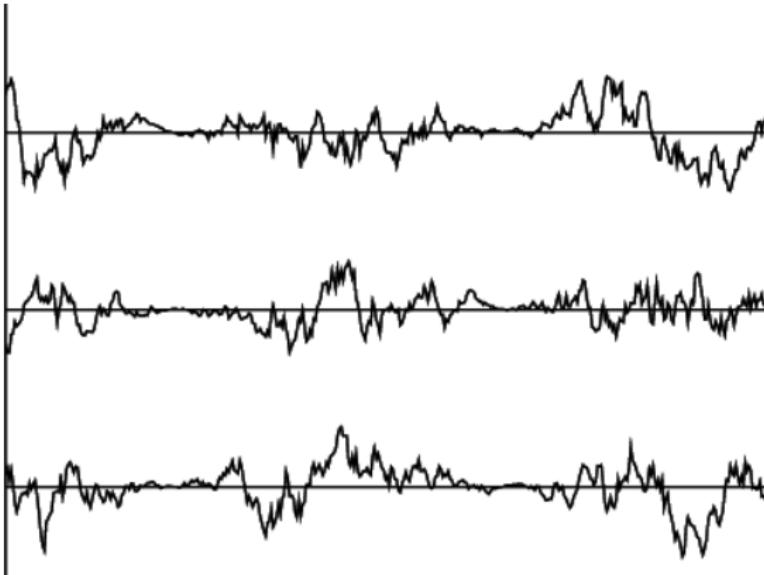
$$w_n(x_1, \dots x_n; t_1, \dots t_n) = w_n(x_1, \dots x_n; t_1 + \tau, \dots t_n + \tau)$$

- ▶ Basically, stationary = **all statistical averages do not depend on time t**

- ▶ **Strictly** stationary / **strongly** stationary / **strict-sense** stationary:
 - ▶ relation holds for every n (i.e. for distributions of any order)
 - ▶ average value, average squared value, variance, autocorrelation, **and all other** higher-order statistics do not depend on origin of time t
- ▶ **Weakly** stationary / **wide-sense** stationary:
 - ▶ relation holds only for $n = 1$ and $n = 2$ (i.e. for distributions of a single sample, and of a pair of samples)
 - ▶ **only** the average value, average squared value, variance, autocorrelation do not depend on time t , but high-order statistics may depend

Stationary random processes

- ▶ Is the Random Process below stationary or non-stationary?



- ▶ from: SEX, LIES & STATISTICS, Ned Wright,
<http://www.astro.ucla.edu/~wright/statistics/>

- ▶ Answer: non-stationary
- ▶ You can see that the variance is not the same at all moments of time

Consequences of stationarity

- ▶ For distributions of a single sample (order $n = 1$):

$$w_1(x_i; t_1) = w_1(x_i; t_2) = w_1(x_i)$$

- ▶ The average value, average squared value, variance of a sample are all **identical at any time t**

$$\overline{f(t)} = \text{constant} = \bar{f}, \forall t$$

$$\overline{f^2(t)} = \text{constant} = \bar{f^2}, \forall t$$

$$\sigma^2(t) = \text{constant} = \sigma^2, \forall t$$

Consequences of stationarity

- ▶ For distributions of a pair of samples (order $n = 2$):

$$w_2(x_i, x_j; t_1, t_2) = w_2(x_i, x_j; 0, t_2 - t_1) = w_2(x_i, x_j; t_2 - t_1)$$

- ▶ The autocorrelation & correlation function depend only on the **time difference** $\tau = t_2 - t_1$ between the samples

$$R_{ff}(t_1, t_2) = R_{ff}(0, t_2 - t_1) = R_{ff}(\tau) = \overline{f(t)f(t + \tau)}$$

- ▶ Depend on a single value $\tau = t_2 - t_1 =$ time difference of the two samples

Consequences of stationarity

Definition of autocorrelation function for a **stationary** r.p:

- ▶ Statistical autocorrelation: no change
- ▶ Temporal autocorrelation:
 - ▶ for continuous r.p.

$$R_{ff}(\tau) = \overline{f(t)f(t+\tau)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^{(k)}(t)f^{(k)}(t+\tau)dt$$

- ▶ for discrete r.p.

$$R_{ff}(\tau) = \overline{f(t)f(t+\tau)} = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N f^{(k)}[t]f^{(k)}[t+\tau]$$

- ▶ finite length realizations: use $\int_{t_{min}}^{t_{max}}$ or $\sum_{t_{min}}^{t_{max}}$

Consequences of stationarity

- ▶ For **correlation**, similar to the autocorrelation definition on the previous slide
- ▶ Correlation depends only on the **time difference** $\tau = t_2 - t_1$ between the samples

$$R_{fg}(t_1, t_2) = R_{fg}(0, t_2 - t_1) = R_{fg}(\tau) = \overline{f(t)g(t + \tau)}$$

Interpretation of autocorrelation

- ▶ $R_{ff}(\tau) = \overline{f(t)f(t + \tau)}$ = the **average value of the product of two samples which are time τ apart**
 - ▶ e.g. tells us if the two samples vary in same direction or not
- ▶ Same for correlation, but the samples are taken from different r.p f and g

Interpretation of autocorrelation

- ▶ Example:

- ▶ $R_{ff}(0.5) > 0$ means two samples separated by 0.5 seconds tend to vary in same direction (both positive, both negative => their product is mostly positive)
 - ▶ knowing one of them, you can "guess" the other one
- ▶ $R_{ff}(1) < 0$ means two samples separated by 1 second tend to vary in opposite directions (when one is positive, the other is negative => their product is mostly negative)
 - ▶ knowing one of them, you can "guess" the other one
- ▶ $R_{ff}(2) = 0$ means two samples separated by 2 seconds are **uncorrelated** (their product is 0 on average, so the two samples have equal chances of being in the same or opposite directions)
 - ▶ knowing one of them, you **can't** "guess" the other one

- ▶ In practice, we have access to a single realization
- ▶ **Ergodic** random process = if the temporal averages on any realization are **equal to the statistical** averages
- ▶ Ergodicity means:
 - ▶ We can compute / estimate all averages from a single realization (any)
 - ▶ but the realization must be very long ($\text{length} \rightarrow \infty$) for precise results
 - ▶ Realizations are all similar to the others, in terms of averages
 - ▶ so a single realization is characteristic of the whole process

- ▶ Most random processes we consider in this course are both ergodic and stationary
 - ▶ e.g. voltage noises
- ▶ Example of **non-ergodic** process:
 - ▶ throw a dice, then the next 50 values are identical to the first
 - ▶ a single realization is not characteristic

```
int getRandomNumber()
{
    return 4; // chosen by fair dice roll.
              // guaranteed to be random.
}
```

- ▶ XKCD 221 (link here: <https://xkcd.com/221/>)
- ▶ Consider all numbers which could have been obtained by the "fair dice roll"
- ▶ What's the problem here?
 - ▶ stationary or non-stationary?
 - ▶ ergodic or non-ergodic?

I.3 More on autocorrelation

The Power Spectral Density of a signal

- ▶ The **Power Spectral Density (PSD)** $S_{ff}(\omega) =$ a function indicating the power of the signal depending on the frequency f ($\omega = 2\pi f$)
- ▶ For a **deterministic** signal (i.e. non-random), it is given by the modulus of the Fourier transform (squared):

$$S_{ff}(\omega) = |F(\omega)|^2$$

- ▶ The power in the frequency band $[f_1, f_2]$ is equal to $\int_{f_1}^{f_2} S_{ff}(\omega) d\omega$
- ▶ The whole power of the signal is $P = \int_{-\infty}^{\infty} S_{ff}(\omega) d\omega$
- ▶ The PSD is a measurable quantity
 - ▶ it can be determined experimentally
 - ▶ it is important in practical (engineering) applications

The Power Spectral Density of a random process

- ▶ How about for a random process?
 - ▶ we don't have one signal, we have an infinite number of possible realizations
 - ▶ we have a (different) Fourier transform for each possible realization
 - ▶ the PSD for every realization is, itself, a random process
- ▶ The **PSD of a random process** = the average value of the PSD for all the realizations
- ▶ Has the same meaning and use as for a deterministic signal, but **in average** across all possible realizations
 - ▶ for a particular realization, the values vary around the average

The Wiener-Khinchin theorem

The Wiener-Khinchin theorem ("teorema Wiener-Hincin")

- ▶ The Power Spectral Density of a random process = **the Fourier transform of the autocorrelation function**

$$S_{ff}(\omega) = \int_{-\infty}^{\infty} R_{ff}(\tau) e^{-j\omega\tau} d\tau$$

$$R_{ff}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ff}(\omega) e^{j\omega\tau} d\omega$$

- ▶ No proof
- ▶ Notes:
 - ▶ Makes a relation between two rather different domains:
 1. The autocorrelation function: a *statistical* property
 2. The PSD function: a *physical* property (relevant for engineering purposes)

- ▶ **White noise** = a random process with **autocorrelation function equal to a Dirac** function

$$R_{ff}(\tau) = \delta(\tau)$$

- ▶ is a random process: every sample of white noise is a random variable
- ▶ autocorrelation is a Dirac: autocorrelation is 0 for any $\tau \neq 0$
- ▶ any two different samples ($\tau \neq 0$) have zero correlation (are uncorrelated)
 - ▶ values of any two different samples are not related

White noise

- ▶ Power spectral density of white noise = Fourier transform of a Dirac
= **a constant** = does not depend on ω

$$S_{ff}(\omega) = \text{constant}, \forall \omega$$

- ▶ equal distribution of power at all frequencies up to ∞
- ▶ White noise can have **any distribution** (normal, uniform etc.)
 - ▶ the term "white noise" doesn't refer to the distribution of sample values, but to the fact that all samples are unrelated to each other

Band-limited white noise

- ▶ In real life, power spectral density goes to 0 at very high frequencies
 - ▶ because total power $P = \int_{-\infty}^{\infty} S_{ff}\omega$ cannot be infinite
 - ▶ known as "**band-limited white noise**"
- ▶ In this case, autocorrelation = **approximately** a Dirac, but not infinitely thin
 - ▶ samples which are very close are necessarily a bit correlated
 - ▶ e.g. due to small parasitic capacities

- ▶ **AWGN** = Additive White Gaussian Noise
 - ▶ is the usual type of noise considered in applications
- ▶ It means:
 - ▶ additive: the noise is added to the original signal (e.g. not multiplied with it)
 - ▶ white: the samples are uncorrelated (unrelated) with each other
 - ▶ gaussian: the samples have normal distribution
 - ▶ noise: it is a random process, every sample is random, every realization is different

- ▶ Chapter 1 ends here for 2020-2021 exam. Following slides not needed.

Properties of the autocorrelation function

1. Is even

$$R_{ff}(\tau) = R_{ff}(-\tau)$$

- ▶ Proof: change variable in definition

2. At infinite it goes to a constant

$$R_{ff}(\infty) = \overline{f(t)^2} = \text{const}$$

- ▶ Proof: two samples separated by ∞ are independent

3. Is maximum in 0

$$R_{ff}(0) \geq R_{ff}(\tau)$$

- ▶ Proof: start from $\overline{(f(t) - f(t + \tau))^2} \geq 0$
- ▶ Interpretation: different samples might vary differently, but a sample always varies identically with itself

Properties of the autocorrelation function

4. Value in 0 = the power of the random process

$$R_{ff}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ff}(\omega) d\omega$$

- ▶ Proof: Put $\tau = 0$ in inverse Fourier transform of Wiener-Khinchin theorem

5. Variance = difference between values at 0 and ∞

$$\sigma^2 = R_{ff}(0) - R_{ff}(\infty)$$

- ▶ Proof: $R_{ff}(0) = \overline{f(t)^2}$, $R_{ff}(\infty) = \overline{f(t)}^2$

- ▶ Consider a stationary random process applied as input to a LTI system
 - ▶ either continuous-time: input $x(t)$, system $H(s)$, output $y(t)$
 - ▶ or discrete-time: input $x[n]$, system $H(z)$, output $y[n]$
- ▶ How does the autocorrelation of y depend on that of the input x ?
 - ▶ y is the convolution between x and the impulse response h

Computations

- For discrete-time processes

$$\begin{aligned} R_{yy}(\tau) &= \overline{y[n]y[n + \tau]} \\ &= \overline{\sum_{k_1=-\infty}^{\infty} h[k_1]x[n - k_1] \sum_{k_2=-\infty}^{\infty} h[k_2]x[n + \tau - k_2]} \\ &= \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h[k_1]h[k_2] \overline{x[n - k_1]x[n + \tau - k_2]} \\ &= \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h[k_1]h[k_2]R_{xx}[\tau - k_1 + k_2] \end{aligned}$$

- From Wiener-Hincin theorem:

$$S_{ff}(\omega) = \sum_{\tau=-\infty}^{\infty} R_{ff}(\tau) e^{-j\omega\tau}$$

Computations

- ▶ Therefore

$$S_{yy}(\omega) = \sum_{\tau=-\infty}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h[k_1]h[k_2]R_{xx}[\tau - k_1 + k_2]e^{-j\omega\tau}$$

- ▶ Change of variable: $\tau - k_1 + k_2 = u$

- ▶ then $\tau = u + k_1 - k_2$

$$\begin{aligned} S_{yy}(\omega) &= \sum_{u=-\infty}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h[k_1]h[k_2]R_{xx}[u]e^{-j\omega(u+k_1+k_2)} \\ &= \sum_{u=-\infty}^{\infty} R_{xx}[u]e^{-j\omega u} \sum_{k_1=-\infty}^{\infty} h[k_1]e^{-j\omega k_1} \sum_{k_2=-\infty}^{\infty} h[k_2]e^{j\omega k_2} \\ &= S_{xx}(\omega) \cdot H(\omega) \cdot H * (\omega) \\ &= S_{xx}(\omega) \cdot |H(\omega)|^2 \end{aligned}$$

Result

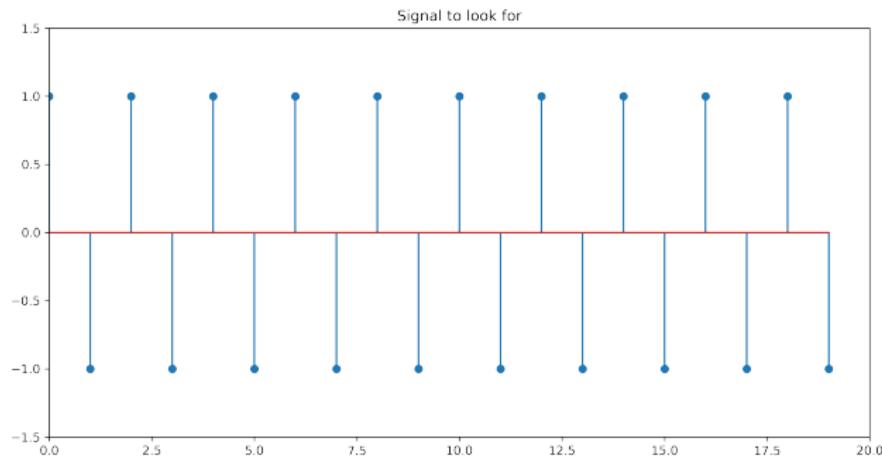
$$S_{yy}(\omega) = S_{xx}(\omega) \cdot |H(\omega)|^2$$

- ▶ The PSD of y = the PSD of x multiplied with the squared amplitude response of the filter
- ▶ Same relation is valid for continuous processes as well

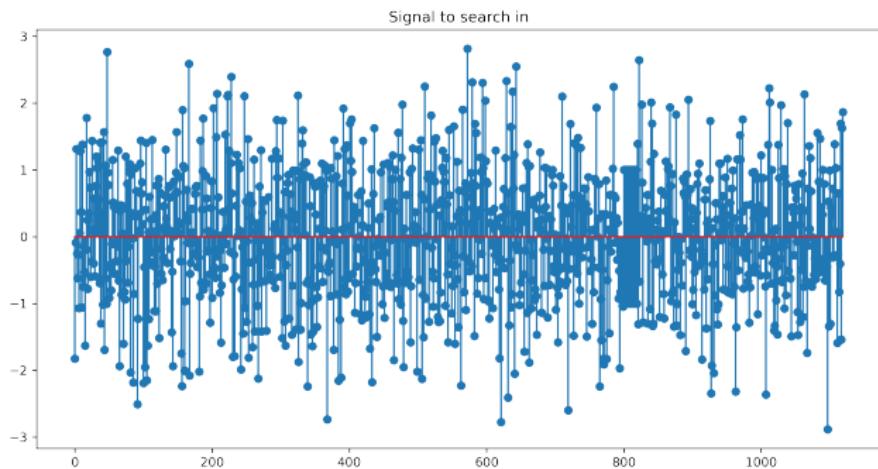
Applications of (auto)correlation

- ▶ Searching for a certain part in a large signal
- ▶ Correlation of two signals = measure of **similarity** of the two signals
 - ▶ The correlation function measures the similarity of a signal with all the shifted versions of the other
 - ▶ Example at blackboard
- ▶ Correlation can be used to locate data
 - ▶ The (auto)correlation function has large values when the two signals match
 - ▶ Large value when both positive and negative areas match,
 - ▶ Small values when they don't match

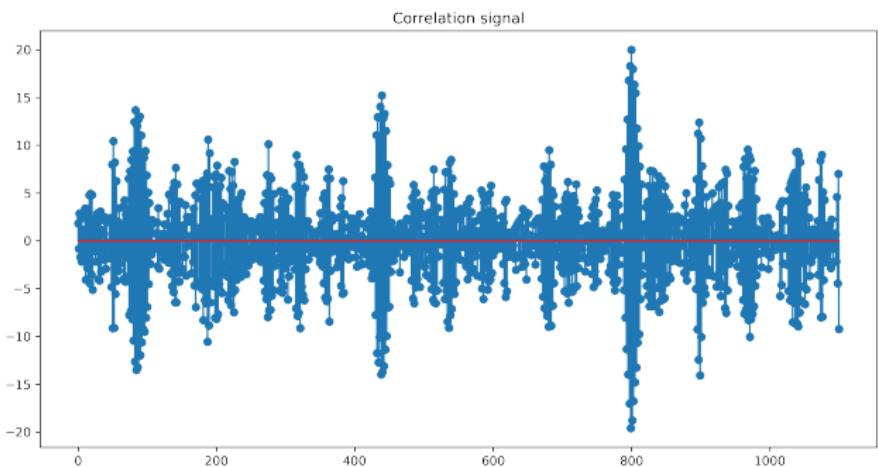
The signal to look for



The complete signal



Correlation result



System identification

- ▶ Determining the impulse response of an unknown LTI system
- ▶ Based on correlation between input and output of the system

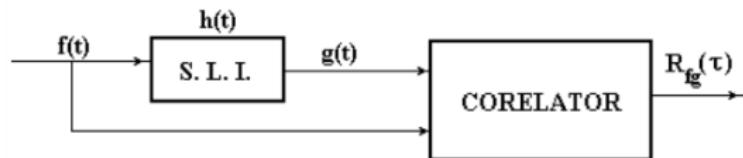


Figure 1: System identification setup

System identification

$$\begin{aligned} R_{fg}(\tau) &= \overline{f[n]g[n + \tau]} \\ &= f[n] \sum_{k=-\infty}^{\infty} h[k]f[n + \tau - k] \\ &= \sum_{k=-\infty}^{\infty} h[k]\overline{f[n]f[n + \tau - k]} \\ &= \sum_{k=-\infty}^{\infty} h[k]R_{ff}[\tau - k] \\ &= h[\tau] \star R_{ff}[\tau] \end{aligned}$$

- If the input f is **white noise** with power A , $R_{ff}[n] = A \cdot \delta[n]$, and

$$R_{fg}(\tau) = h[\tau] \star R_{ff}[\tau] = A \cdot h[\tau] \star \delta[\tau] = A \cdot h[\tau]$$

- Then the correlation is proportional with the impulse response of the unknown system