Vocabulary

Sunday, September 21, 2025 4:40 PM

- Set
  - Collection of objects
  - $A = \{$ Jane, Bill, Jessica, Fred $\}$ (enumeration)

    $A = \{$ Students : each student is enrolled in the Prob. class $\}$ (description)

- ' : '
  - Reas as " such that "

- Element
  - Each object in the set
  - Each element is distinct
  - Sets $\{1, 2, 3\}$ and $\{1, 2, 1, 3\}$ are equivalent
  - $C_1 = \{$ Bill, Fred $\}$, $C_2 = \{$ Male members in the Prob. class $\}$

    $C_1 = C_2$

- '$\in$'
  - Discribes something contained in something.
  - $x \in A$ is read as " x is contained in A "

- Empty set / null set
  - Denoted $\emptyset$
  - $D = \{\ \}$, $D = \emptyset$ } Same thing

- Universal set
  - Denoted S
  - The set of <u>all</u> students enrolled in Prob. class

- Sample space
  - The list of all possible outcomes of a random experiment

- Outcome
  - a possible result of the random experiment

- Event

- Event

  - an event is a possible outcome of the experiment and forms a subset of the sample space

- Trial

  - when a random experiment is repeated many times each one is known as a trial

- Strict Stationary Process

  - Mean and variance doesn't change with time fordelingen skal vare ens for hver tidspunkt

- Wide sense Stationary

  - Mean, μ_x, and variance, σ, is not dependent on time t (they are constant)
  - ACF is only dependent on time lag τ

Lec 0

Thursday, September 4, 2025 12:57 PM

S00_stochasticSyste...

Stochastic Systems

Probability Spaces and Random Elements

Troels Pedersen

Section Wireless Communication Networks,
Department of Electronic Systems, Aalborg University

Fall 2025

11 / 21

Outline for Probability Spaces and Random Elements

Welcome to The Course!

Probability Spaces and Random Elements

12 / 21

Outline for Probability Spaces and Random Elements

Welcome to The Course!

Probability Spaces and Random Elements

13 / 21

Random Experiment

In propbability theory, an *experiment* is a mathematical model of any procedure that has a well-defined set of possible outcomes.

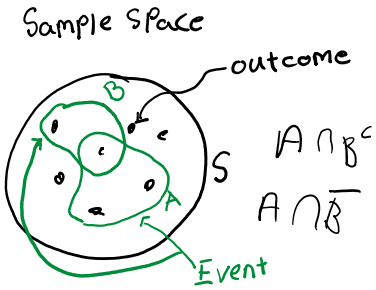
Each time a random experiment is conductioned one (and only one) outcome results.
The experiment is called *random* if it has more that one possible outcome and *deterministic* if it has only one.

The set of possible outcomes is called a *sample space*, $\mathbb{S}$.

Some examples of random experiments:

- Coin toss Head & tailes
- Dice through amount of eyes
- Dart through
- Voltage measurement

Ex1: Discuss the examples above. Identify a sample space in each case.
Ex2: Sample space is not unique — can you define an additional sample space for the coin toss?



Random Experiment

In propbability theory, an *experiment* is a mathematical model of any procedure that has a well-defined set of possible outcomes.

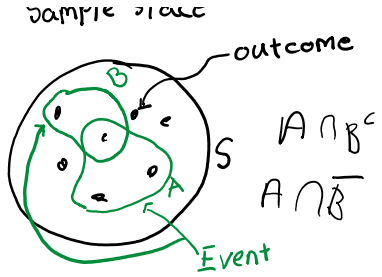
Each time a random experiment is conductioned one (and only one) outcome results.
The experiment is called *random* if it has more that one possible outcome and *deterministic* if it has only one.

The set of possible outcomes is called a *sample space*, $\mathbb{S}$.

Some examples of random experiments:

- Coin toss Head & tailes
- Dice through amount of eyes
- Dart through
- Voltage measurement

Ex1: Discuss the examples above. Identify a sample space in each case.
Ex2: Sample space is not unique — can you define an additional sample space for the coin toss?



Sample Space, $\mathbb{S}$, outcomes and events

The *sample space* $\mathbb{S}$ is a set of all possible *outcomes* of a random experiment. Two conditions apply:

- only one outcome results from the experiment (mutually exclusive)
- all possible outcomes included (collectively exhaustive)

Rather than talking about single outcomes, we can think of sets of them. Such a set of outcomes is called an *event*.

An event A is a sub-set of a sample space $\mathbb{S}$, that is $A \subset \mathbb{S}$.

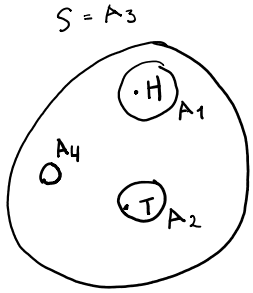
Event lingo:

- "Event A occurred" means that the random experiment resulted in one of the outcomes in A.
- "Event B did not occur" means that the random experiment resulted in one of the outcomes in the complement of A (denoted by A^c.
- The sample space $\mathbb{S}$ is called the "certain event".
- The empty set $\emptyset$ is called the "impossible event".
- An event $A = \{a\}$ with only one outcome is called a "simple event"
- Disjoint events $A \cap B = \emptyset$ are called "mutually exclusive".

Ex3: List all events for the coin toss and give examples where you use the event lingo.

$\varphi = \emptyset$ = Empty set

$S = \{H, T\}$
$A_1 = \{H\}$
$A_2 = \{T\}$
$A_3 = S = \{H, T\}$
$A_4 = \varphi$



A family of events (an algebra)

To make it possible to define probabilities for event, we first need to consider which events to include. It turns out that it is not precise enough to consider "all possible events in $\mathbb{S}$" due to some pathological cases.
We can prevent the pathological cases by only considering a famliy of events $\mathcal{A}$ which is a *σ-algebra*, i.e. that it obeys certain rules:

- $\mathbb{S} \in \mathcal{A}$,
- If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ (closed under complements)
- If $A_1, A_2, \dots \in \mathcal{A}$ then so is $A = A_1 \cup A_2 \cup \dots$ (closed under countable unions)

Ex4: Check for the coin toss example that the events listed in Ex3 fulfills the above conditions for being a σ-algebra.

Note: The σ-algebra of events is important for the mathematical machinery to work, but not very important for understanding the topic of the course — we shall largely ignored it!

The probability function $\mathbb{P}$ (probability measure)

A *probability* is a number between zero an one assigned to each event $A \in \mathcal{A}$.

The assignment is done by a *probability function* also called a *probability measure* denoted by $\mathbb{P}$ which maps $A \in \mathcal{A}$ into the number $0 \leq \mathbb{P}(A) \leq 1$.

The probability function should fulfill three axioms: Prob. of an action

Axiom 1 $\boxed{\mathbb{P}(A)} \geq 0$ for every event $A \in \mathcal{A}$
Axiom 2 $\mathbb{P}(\mathbb{S}) = 1$
Axiom 3 For a countable sequence of disjoint events $A_1, A_2, A_3 \dots$ in $\mathcal{A}$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Together, the tripple $(\mathbb{S}, \mathcal{A}, \mathbb{P})$ of the sample space $\mathbb{S}$, the σ-algebra A and the propobability function is called a *probability space*.

Ex5: Consider again the coin toss. Write up a probability function for a fair coin. Check that the axioms hold.
Ex6: Adapt the probabiity function for the case that the coin is not fair.

Probability function:

$\mathbb{P}: A \to [0,1]$:
$A \to \mathbb{P}(A)$

Fair:

event	$\mathbb{P}(A)$
$\{H\}$	0.5
$\{T\}$	0.5
$\{H,T\}$	1
$\emptyset$	0

Unfair:

event	$\mathbb{P}(A)$
$\{H\}$	p
$\{T\}$	$1-p$
$\{H,T\}$	1
$\emptyset$	0

Some remarks

- The specific assignment of probabilities is up to us — as long as the axioms are fulfilled. We encounter many different probability functions in applications.
- In applications, it is usually unecessary to define the probability measure explicitly, but rather to define more abstract entities, e.g a "random variable following a normal probabiltiy distribution".
- The *interpretation* of the probabilities is up to us — several interpretations are used!
  - Frequency interpretation : $\mathbb{P}(A)$ = relative frequency of event A occurring if the experiment is repeated infinitely many times.
  - Evidential (or Bayesian) interpretation: $\mathbb{P}(A)$ = degree of belief that event A will occur if the experiment is performed once.
- Note that regardless the interpretation, the underlying axioms are the same — the math does not care!

Random Elements[1]

To avoid the hassle of inventing new probability spaces for every new application, we simply map outcomes to new types of mathematical entities (numbers, vectors, functions, ...).

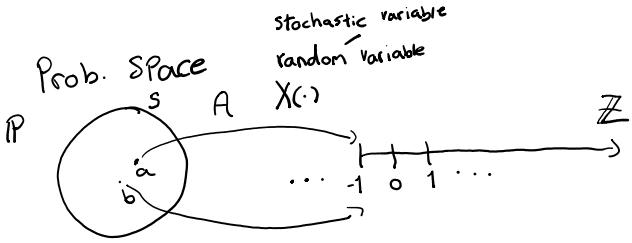
Define a function X to map the outcomes in $\mathbb{S}$ to another range $\mathbb{S}_X$. The function X maps Each outcome s is mapped to a *realization* $X(s) \in \mathbb{S}_x$. Such a construction is called a *random element*.

Examples we encounter in this course:

- Discrete random variables: $\mathbb{S}_X = \mathbb{N}$
- Continuous random variables: $\mathbb{S}_X = \mathbb{R}$
- Random vectors: $\mathbb{S}_X = \mathbb{R}^N$
- Random Processes: $\mathbb{S}_X$ is a space of time-functions
- Point Processes: $\mathbb{S}_X$ set of countable sets of points

Ex5: Consider again the probability space of a fair coin toss. Now define a random variable by inventing a function X that assigns numerical values to each outcome. Now redefine X to create a new type of random element with $\mathbb{S}_x$ being the set of your two favorite animals.

[1] Any concievable may be mapping used, but only particular mappings are feasible to analysis. Further details require some measure theory, which is out of scope of this course. There is and endless list of types of random elements: Wikipedia lists some of them https://en.wikipedia.org/wiki/Random_element



Conditional Probability

The *conditional probability* $\mathbb{P}(A|B)$ is the probability that A occur given that B occurred.

Although the joint probability $\mathbb{P}(A \cap B)$ gives the probabiltiy that both A and B occurs, it is not good enough, since we allready know that B occurred.

Knowing that event B occurred, the sample space is no longer $\mathbb{S}$ but reduces to B. We have to to convert the joint probability $\mathbb{P}(A \cap B)$ into a normalized probability measure defined on this reduced sample space.

Therefore, the conditional probability is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad \text{(Undefined for } \mathbb{P}(B) = 0)$$

It is often more convenient to recast the definition as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B), \qquad \text{(Undefined for } \mathbb{P}(B) = 0)$$

Ex6: Consider the through of a fair six-sided die. What is $\mathbb{P}(A|B)$ with the events A ='six eyes' and B ='Three or more eyes'? Compared to $\mathbb{P}(A \cap B)$ and $\mathbb{P}(A)$.

Bayes' Theorem

With the definition of conditional probability, it is now straight-forward to proved the famous Bayes' Theorem

By the definition,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B), \qquad \text{(Undefined for } \mathbb{P}(B) = 0) \tag{1}$$

Similarly, by swapping the roles of A and B in the definition, we obtain

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A), \qquad \text{(Undefined for } \mathbb{P}(A) = 0) \tag{2}$$

Equating the two expressions and isolating gives obtain *Bayes' Theorem*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \qquad \mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0. \tag{3}$$

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

$$= \frac{\left(\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}\right) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

Lec 0) exercises

Sunday, September 21, 2025 4:26 PM

Intended learning outcome:

- Explain the key concepts:
  - random experiment
  - Sample space
  - events
  - Probability space
- Explain the basics of how to map a Probability space to a random element
- Use and understand conditional probability and apply Baye's Theorem

Random experiment:

- A process or activity which we do not know the outcome before peforming the experiment
- Has more than one possible outcome

Def:

- Experiments that can be peformed many times under the same conditions and their outcome cannot be predicted with complete certainty

Sample space:

- The list of all possible outcomes of a random experiment

Event

- an event is a possible outcome of the experiment and forms a subset of the sample space

Probability space

- A mesure space such that the measure of the whole space is equal to 1
- Mathmatical triplet $(\Omega, \mathcal{F}, P)$
  - Sample space Ω: The set of all possible outcomes
  - σ-algebra $\mathcal{F}$: collection of all the events we would like to consider
  - Probability measure P: a set func. returning an events probability

How to map Prob. Space to random element?

Random element:

- Random element / variable, is a way to map outcomes $\omega \in \Omega$, to another space where analysis can be done

- $X: \Omega \to S$ ← state space $(\mathbb{R}, \mathbb{R}^n$

  Assigns to each outcome, ω, a "realized value", $X(\omega)$.

- Must be measurable to mathmatically valid.
  - $X^{-1}(A) \in \mathcal{F}$ for all measurable $A \subseteq S$
  - ensures that events like "X falls in A" is legitimate events in the prob. space.

Distribution law:

X defines a new prob. measure on S

$$P_X(A) := P(X^{-1}(A))$$

for measurable $A \subseteq S$

- P_X distribution (law) of X
- P_X tells us the prob. X takes a value within A

Mapping

See it as 3 steps:

1. Base	$(\Omega, \mathcal{F}, P)$	Prob. Space
2. Mapping	$X: \Omega \to S$	Maps the sample space to concrete values
3. Inducing	P_X on S	Prob. law of those values (Dist. law)

Overall flow:

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (S, S, P_X)$$

Example:

Rolling a fair die:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{F} = 2^{\Omega}$ (all subsets)
- $P(\omega) = 1/6$

Define random var.:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \text{ even} \\ 0 & \text{if } \omega \text{ odd} \end{cases}$$

Then:

$$P_X(1) = P(\text{even roll}) = \frac{3}{6} = \frac{1}{2}$$

$$P_X(0) = P(\text{odd roll}) = \frac{3}{6} = \frac{1}{2}$$

X maps the die into a "parity outcome" and induces a new prob. dist. on $\{0,1\}$

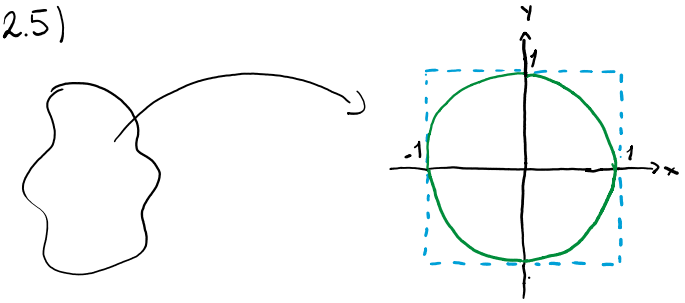
Lec 1 exercises

Thursday, November 13, 2025 11:06 AM

Session 0)

Exercises

12.5)



$S_{X,Y} = \{(x,y) : x^2+y^2 \leq 1\}$

Relate $\mathbb{P}[x^2+y^2 \leq 1]$ to π

area of circle $= \pi r^2 : r=1$

$\mathbb{P}[x^2+y^2 \leq 1] = \frac{\pi}{4} \rightarrow \pi = 4\,\mathbb{P}[x^2+y^2=1]$

Compute:

Python code

12.11)

If (X,Y) has a <u>Standart</u> bivariate Gaussian PDF
find $\mathbb{P}[x^2+y^2 = 10]$

Standart bivariate Gaussian PDF is the PDF for two independent Standart normal random variables (Z_1 and Z_2) each with a mean of 0 and a variance of 1.

Formula:

$$f(z_1, z_2) = \frac{1}{2\pi} e^{\left(-\frac{1}{2}(z_1^2+z_2^2)\right)}$$

"Standart", variables has 0 mean and unit variance, Simply the general bivariate PDF by setting variables:

$\mu_x = \mu_y = 0$

$\sigma_x = \sigma_y = 1$

$\rho = 0$ (For independance)

- - - -

$$\mathbb{P}[x^2+y^2 = 10] = \frac{1}{2\pi} e^{\left(-\frac{1}{2}(x^2+y^2)\right)}$$

Lec 1 - 2)

Tuesday, September 9, 2025 8:44 AM

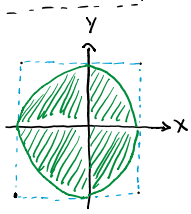
Exercises)

12.5)

$X \sim U(-1,1)$

$Y \sim U(-1,1)$

$P[X^2+Y^2 \leq 1]$



$$P[X^2+Y^2=1] = \frac{\text{Area}}{\text{Area}}$$

$$= \frac{\pi}{4}$$ (Relation of square & circle)

12.11)

Standart bivariate Gaussian:

$$P_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)} \qquad -\infty < x < \infty, \quad -\infty < y < \infty$$

$\rho = \quad -1 < \rho < 1$

$P[X^2+Y^2=10]$

$$P_{X,Y}(x,y)\Big|_{\rho=0} = e^{-\frac{10}{2}(x^2+y^2)}$$

Betinger $= P(X|Y)$

Simultant $= P(x,y)$

Marginal $= P(x)$

12.15)

if $P_{X,Y}(x,y) = \begin{cases} 2 & 0<x<1,\ 0<y<x \\ 0 & \text{otherwise} \end{cases}$

Find marginals PDFs

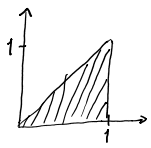
<u>Marginal of X:</u>

$$P(x) = \int P_{X,Y}(x,y)\,dy$$

$$= \int_0^x P(x,y)\,dy \qquad (0<x<1,\ \text{Valid y-values } 0<y<x)(P_{X,Y}(x,y)=2)$$

$$= \int_0^x 2\,dy \qquad (\text{for } 0<x<1)$$

$$= \int_0^x 2[y]_0^x = 2(x-0) = \underline{\underline{2x}}$$



<u>Marginal of Y:</u>

$$P(y) = \int P_{X,Y}(x,y)\,dx \qquad (0<y<x<1 \rightarrow 0<1-y<1)$$

$$P(y) = \int P_{X,Y}(x,y)\,dx \qquad (0<y<x<1 \rightarrow 0<1-y<1)$$

$$= \int_y^1 2\,dx = \int_y^1 2[x]_y^1 = \underline{\underline{2(1-y)}}$$

13.2)

$$P_{Y|X}(y|x) = \begin{cases} c e^{-y/x} & y \geq 0,\ x>0 \\ 0 & \text{otherwise} \end{cases}$$

$$P_{Y|X}(y|x) = c\,e^{-y/x} \rightarrow \int_0^\infty c e^{-y/x}\,dy = -xce^{-y/x}$$

$$P(\infty) - P(0) = 0 - (-xc) = 1$$

$$= xc = 1$$

$$= c = \underline{\underline{\frac{1}{x}}}$$

13.4)

if $P_{X,Y}(x,y) = \begin{cases} 2e^{-(x+y)} & 0 \leq y \leq x,\ x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

$$P_{X,Y}(x,y) = \frac{2e^{-(x+y)}}{P(x)} \qquad P(x) = \int_0^x 2e^{-(x+y)}\,dy =$$

$$= -(x+y)2e^{-(x+y)}\,dy$$

$$= -\cancel{x}\,\overset{1}{\cancel{-y}}\,2e^{-(x+y)}\,dy$$

$$= -1 \cdot 2e^{-(x+y)}$$

$$= -2e^{-(x+y)}$$

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

$$P(x) = \int P(y,x)\,dy = 2e^{-x} - 2e^{-2x}$$

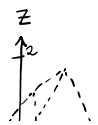
$$P(y|x) = \frac{2e^{-x} \cdot e^{-y}}{2e^{-x} - 2e^{-x}e^{-x}}$$

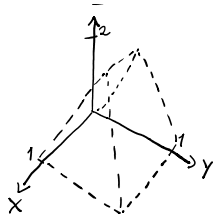
$$= \frac{\cancel{2e^{-x}} \cdot e^{-y}}{\cancel{2e^{-x}}(1-e^{-x})}$$

$$= \frac{e^{-y}}{1-e^{-x}}$$

13.5) Plot the joint PDF

$$P_{X,Y}(x,y) = \begin{cases} 2x & 0<x<1,\ 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

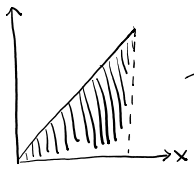




$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

13.8) If $X \sim U(0,1)$ and $Y|(X=x) \sim U(0,x)$
find the joint PDF for X and Y
and also the marginal PDF for Y

PDF



$$\rightarrow P_{X,Y}(x,y) = \begin{cases} \frac{1}{x}, & 0<x<1, \quad 0<y<x \rightarrow 0<1-y<1 \\ 0 & \text{otherwise} \end{cases}$$

$$P(Y) = \int_Y^1 P_{X,Y}(x,y)\, dx$$

$$= \int_Y^1 \frac{1}{x} [x]_Y^1$$

$$= \ln(1-y)$$

$$= \underline{\underline{-\ln(y)}}$$

Lec 1 Learning outcomes

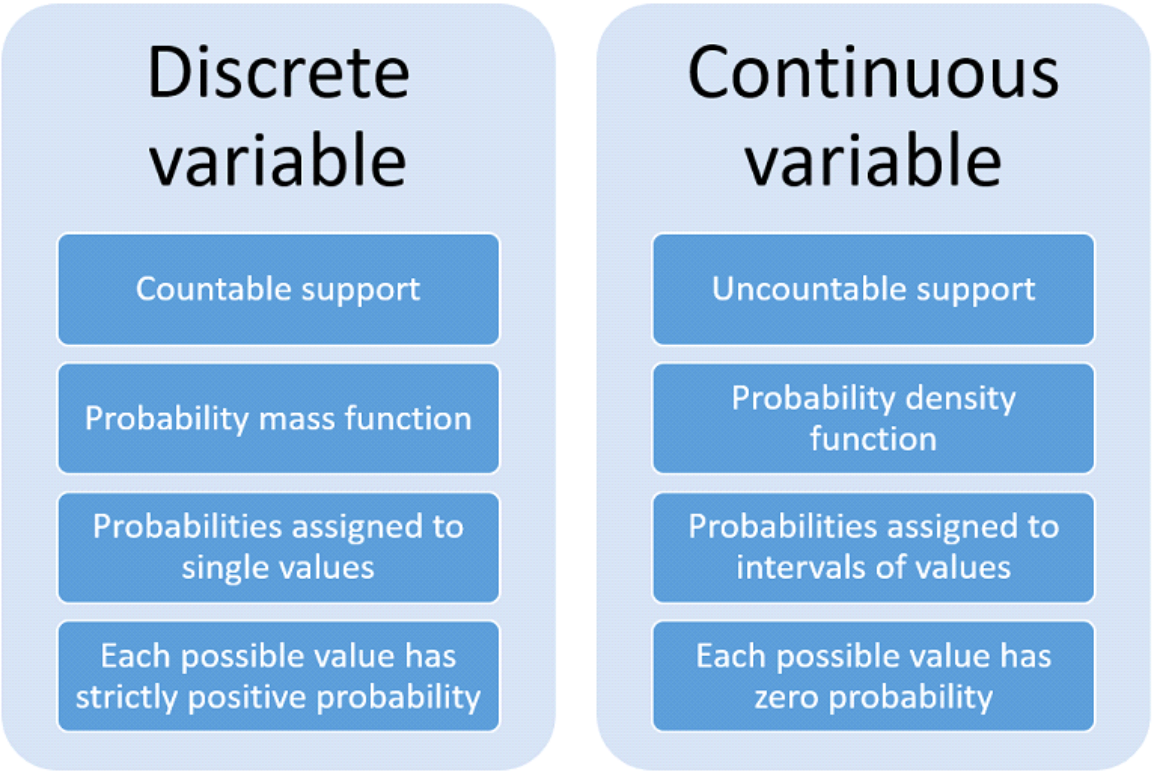
Tuesday, November 25, 2025 10:09 AM

Intended Learning Outcomes

After having attended this session solved the exercises you should be able to:

- Explain the difference between a discrete and a continuous random variable to a fellow student.
- Relate the concepts of joint, conditional and marginal pdfs.
- Write up and interpret the covariance matrix of a random vector.
- Compute the expectation of simple functions of random variables, such as a sum of random variables or the product of independent random variables.

Explain the difference between a discrete and a continuous random variable to a fellow student.



Relate the concepts of joint, conditional and marginal pdfs.

- **Marginal Probability** refers to the probability of a single event occurring, without considering any other events.
- **Joint Probability** is the probability of two or more events happening at the same time. It is the probability of the intersection of these events.
- **Conditional Probability** deals with the probability of an event occurring given that another event has already occurred.

In this article, we will discuss these probabilities in detail, including examples and differences between them as well.

Lec 3)

Tuesday, September 16, 2025 8:21 AM

14.1)

if $Y = X_1 + X_2 + X_3$

where

$$X \sim \mathcal{N}(\mu, C)$$

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{bmatrix}$$

find the mean and variance of Y

Mean: $\bar{X}$ is an unbiased estimator of μ_X

$$\mu_{\bar{X}} = \mu_X$$

Variance:

$$\sigma_X^2 = \frac{1}{N} \sum_{K=-(N-1)}^{N-1} \left[1 - \frac{|K|}{N}\right] C_{XX}(K)$$

X

$$\mu_Y = \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \mathbb{E}[Y_3]$$

$$= 1 + 2 + 3 = \underline{\underline{6}}$$

$$Var(Y) = \mathbb{E}((Y - \mu_Y)^2)$$

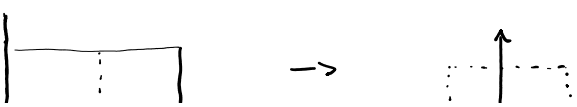
$$= \mathbb{E}((X_1 + X_2 + X_3 - \mu_Y)^2)$$

14.12)

$$\sum_{i=1}^{N} i \cdot \sigma^2 = \sigma^2 \sum_{i=1}^{N} i = \frac{(N+1)N}{2}$$

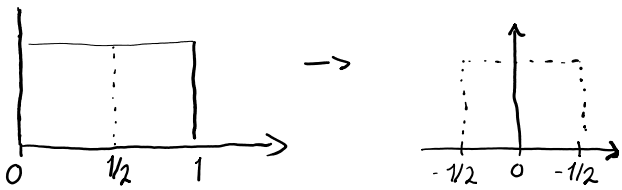
14.19)

$$Y = \sum_{i=1}^{12} \left(U_i - \frac{1}{2}\right)$$



$$Y = \sum_{i=1}^{12} \left(U_i - \frac{1}{2}\right)$$

$U_i \sim \mathcal{U}(0,1)$



$\text{mean}(Y) = \underline{\underline{0}}$

upper- lower-bound

$$\text{Var}(Y) = \frac{(b-a)^2}{12} \cdot 12 = \underline{\underline{1}}$$

Lec 5)

Tuesday, September 23, 2025 8:18 AM

S05_stochasticSyste...

Stochastic Systems

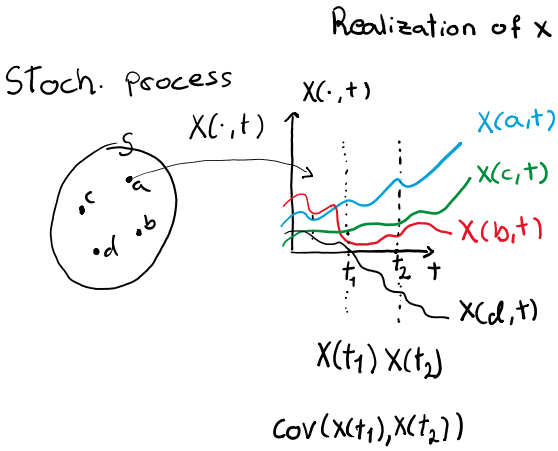
Session 5 — Lecture

Troels Pedersen and Carles Navarro Manchón

Section Wireless Communication Networks,
Department of Electronic Systems, Aalborg University

Fall 2022

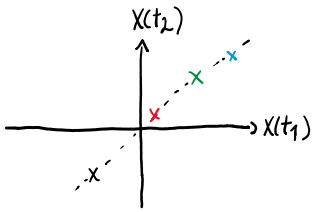
1 / 26



Outline for Session 5 — Lecture

- ILOs for Session 5
- The autocorrelation and its properties
- Weak sense stationary (WSS) processes
- Gaussian WSS processes are SSS
- Autocorrelation quiz

2 / 26



$t_1 = t_2$

$cov(x(t_1), x(t_2)) = Var(x(t_1)) \geq 0$

$Var(X(t_1)) = \mathbb{E}[(x(t_1) - \mu_x(t_1))^2]$

Outline for Session 5 — Lecture

- ILOs for Session 5
- The autocorrelation and its properties
- Weak sense stationary (WSS) processes
- Gaussian WSS processes are SSS
- Autocorrelation quiz

3 / 26

ILOs for Session 5

After attending this lecture and solving the exercises you should be able to:

- Explain the meaning of an ACF to a fellow student.
- Compute the ACF for simple stochastic process using its definition.
- Use the theoretical properties of any ACF as sanity checks of your derivations.
- Determine whether or not a stochastic process is WSS.
- Relate the definitions of SSS and WSS.
- Know (without hesitation and computation) the ACF of an iid. process.

4 / 26

Outline for Session 5 — Lecture

ILOs for Session 5

The autocorrelation and its properties

Weak sense stationary (WSS) processes

Gaussian WSS processes are SSS

Autocorrelation quiz

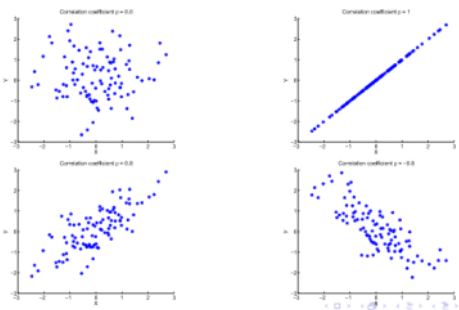
5 / 26

Statistical dependency revealed via second-order characteristics.

Example:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

Correlation coefficient $\rho = \mathrm{Cov}(X, Y)/\sqrt{Var(X) \cdot Var(Y)}$



6 / 26

Covariance function

Linear statistical dependency of two RVs is revealed via second-order characteristics, i.e. in their covariance or correlation.

Similarly, for a random process $\{X(t)\}$ we can measure the dependency between any pair of its random variables, say $X(t_1)$ and $X(t_2)$ using the

Covariance function:

$$\begin{aligned} C_X(t_1, t_2) &:= \mathbb{C}ov\big(X(t_1), X(t_2)\big) \\ &= \mathbb{E}\left[\big(X(t_1) - \mu_X(t_1)\big) \cdot \big(X(t_2) - \mu_X(t_2)\big)\right] \end{aligned}$$

where $\mu_X(t) = \mathbb{E}[X(t)]$.

Relation to the (auto-)correlation coefficient:

$$\rho_X(t_1, t_2) := \frac{C_X(t_1, t_2)}{\sqrt{C_X(t_1, t_1) C_X(t_2, t_2)}} \in [-1, 1]$$

Which we will not use much in the course, but is good to know.

7 / 26

Covariance function:

$$C_X(t_1, t_2) = \mathrm{cov}(X(t_1), X(t_2))$$
$$= \mathbb{E}[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]$$

Mean func.

$$\mu_X(t) = \mathbb{E}[X(t)]$$

Example: Independent Identically Distributed (i.i.d.) Random Process

The random process $X(n)$ is a random process made of independent and identically distributed samples with mean $\mathbb{E}[X(n)] = \mu_X$ and variance $\mathbb{V}\mathrm{ar}[X(n)] = \sigma_X^2$.

The covariance function of such i.i.d. random process is:

$$C_x(n_1, n_2) = \sigma_X^2 \delta(n_2 - n_1) = \begin{cases} \sigma_X^2, & n_1 = n_2 \\ 0, & n_1 \neq n_2 \end{cases}$$

(Check this!)

variancen (tager cov af "de to" der er ens, så er cov = σ^2 (variancen))

covariancen (tager cov af "de to" der ikke er ens, så er covariancen = 0 (derfor er de i.i.d.))

$$C_x(n_1, n_2) = \mathrm{cov}[X(n_2), X(n_1)]$$
$$= \mathbb{E}[(X(n_2) - \mu_x(n_2)) \cdot (X(n_1) - \mu_x(n_1))]$$

Autocorrelation function (ACF)

The same information of statistical dependency is captured by the simpler **Autocorrelation function (ACF):**[1]

$$R_x(t_1, t_2) := \mathbb{E}[X(t_1)X(t_2)]$$

which is related the covariance function as

$$R_x(t_1, t_2) = C_x(t_1, t_2) + \mu_x(t_1)\mu_x(t_2)$$ *(Show this!)*

(Obviously, for the special case $\mu_x(t) = 0$, we have $R_x(t_1, t_2) = C_x(t_1, t_2)$.)

Properties of the ACF for general processes:

Non-negative at $t_1 = t_2$: $R_x(t_1, t_1) \geq 0$ *(Show this!)*

Symmetry: $R_x(t_1, t_2) = R_x(t_2, t_1)$ *(Show this!)*

[1] [Kay] uses the term "autocorrelation sequence" (ACS) when discussing discrete-time processes and reserve the term "autocorrelation function" (ACF) for the continuos time case. We do not distinguish in the name.

Example: Independent Identically Distributed (i.i.d.) Random Process (cont'd)

The random process $X(n)$ is a random process made of independent and identically distributed samples with mean $\mathbb{E}[X(n)] = \mu_X$ and variance $\mathbb{V}\mathrm{ar}[X(n)] = \sigma_X^2$.

The covariance function of such i.i.d. random process is:

$$C_x(n_1, n_2) = \sigma_X^2 \delta(n_2 - n_1) = \begin{cases} \sigma_X^2, & n_1 = n_2 \\ 0, & n_1 \neq n_2 \end{cases}$$

What is the ACF of X(n)?

$$R_x(t_1, t_2) = \mathrm{Cov}(t_1, t_2) + \mu_x(t_1) \cdot \mu_x(t_2)$$
$$= \sigma_X^2 \delta(n_2 - n_1) + \mu_x(n_2) \cdot \mu_x(n_1)$$

Example: Independent Identically Distributed (i.i.d.) Random Process (cont'd)

The random process $X(n)$ is a random process made of independent and identically distributed samples with mean $\mathbb{E}[X(n)] = \mu_X$ and variance $\mathbb{V}\mathrm{ar}[X(n)] = \sigma_X^2$.

The covariance function of such i.i.d. random process is:

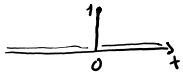
$$C_x(n_1, n_2) = \sigma_X^2 \delta(n_2 - n_1) = \begin{cases} \sigma_X^2, & n_1 = n_2 \\ 0, & n_1 \neq n_2 \end{cases}$$

What is the ACF of X(n)?

$$R_x(n_1, n_2) = \mu_X^2 + \sigma_X^2 \delta(n_2 - n_1)$$

Kronecker delta

$$\delta(t) = \begin{cases} 1 & t = 0 \\ 0 & t \neq 0 \end{cases}$$



Dirac Delta

$$\int_{-\infty}^{\infty} f(t)\,\delta(t)\,dt = f(0)$$ (Sampling property)

Auto correlation func. (ACF)

$$R_x(t_1, t_2) = \mathbb{E}[X(t_1)\,X(t_2)]$$

$$R_x(t_1, t_2) = C_x(t_1, t_2) + \mu_x(t_1)\,\mu_x(t_2)$$

- $R_x(t_1, t_2) = \mathbb{E}[X(t_1) \cdot X(t_2)] = \mathbb{E}[X^2(t_1)] \geq 0$
- $R_x(t_1, t_2) = \mathbb{E}[X(t_1) \cdot X(t_2)] = R_x(t_2, t_1)$

Examples of processes with easily computable ACFs

Example 1: For a discrete-time process $X(n)$ with independent samples according to $X(n) \sim \mathcal{N}(an, n\sigma^2)$, the ACF reads

$$R_x(n_1, n_2) = \begin{cases} \mathbb{E}[X^2(n_1)], & n_1 = n_2 \\ \mathbb{E}[X(n_1)]\mathbb{E}[X(n_2)], & n_1 \neq n_2 \end{cases}$$
$$= \begin{cases} a^2 n_1^2 + n_1\sigma^2, & n_1 = n_2 \\ a^2 n_1 n_2, & n_1 \neq n_2 \end{cases} \qquad \textit{(Check this!)}$$

(Can you find $C_X(t_1, t_2)$?)

Example 2: The extreme case where $X(t) = g(t)$ is deterministic,

$$R_x(t_1, t_2) = g(t_1)g(t_2). \qquad \textit{(Check this!)}$$

(What is $C_X(t_1, t_2)$ in this case?)

Outline for Session 5 — Lecture

ILOs for Session 5

The autocorrelation and its properties

Weak sense stationary (WSS) processes

Gaussian WSS processes are SSS

Autocorrelation quiz

Strict-sense stationarity is too restrictive in practice

Recall the definition of SSS:

Definition of SSS:
A stochastic process $X(t)$ is said to be *strict-sense stationary* (SSS) if for each $n \in \mathbb{N}$ and any $\tau \in \mathbb{T}$:

$$P\Big(X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n\Big) = P\Big(X(t_1 + \tau) \leq x_1, \ldots, X(t_n + \tau) \leq x_n\Big)$$
$$\forall t_1, t_2, \ldots, t_n \in \mathbb{T} \qquad \text{and} \qquad \forall x_1, x_2, \ldots, x_n \in \mathbb{R}$$

In words:
The process $X(t)$ is SSS if *all* of its n'th-order joint cdf's are invariant under *arbitrary* time shifts.

SSS is too strict for practical purposes:

- It is usually hard to verify that all n'th-order joint cdf's are shift invariant.
- It is very rarely fulfilled in practice.

Therefore we seek a "weaker" stationarity definition.

First and second moment properties of SSS processes:

Recall that a SSS process $X(t)$ has the following properties:
The *mean and variance are constants*:

$$\mu_X(t) = \mathbb{E}[X(t)] = \mu_X \quad \text{and} \quad \sigma_X^2(t) = \mathbb{V}ar[X(t)] = \sigma^2$$

The covariance *$C_X(t, t+\tau)$ is a function of the time lag τ* only:

$$C_X(t, t+\tau) = C_X(\tau)$$

The latter property carries over to the autocorrelation *(Check this!)*:

$$R_x(t, t+\tau) = R_x(\tau)$$

Warning: While a SSS process has constant mean, variance and a covariance depending only on the time difference, these properties are not enough to ensure that a process is SSS!
Motivation: Often we care only about first and second moments. Thus we consider the class of processes with the same first and second moment properties as SSS processes.

Weak-sense (or wide-sense) stationary (WSS) processes

We consider processes with the same first and second moment properties as holds for SSS processes. We call such processes weak-sense stationary (WSS):

Definition:
A stochastic process $X(t)$ is said to be *weak-sense stationary* (WSS) if:

1) $\mathbb{E}[X(t)] = \mu_x$, i.e. *no* dependency on time t
2) $\mathbb{E}\big[X(t)X(t+\tau)\big] = R_x(\tau)$, i.e. dependency *only* on the time lag τ

In plain words:
The process $X(t)$ is WSS if it has constant mean and its autocorrelation function depends only on the time lag (not on absolute time instances).

SSS versus WSS

We see from our definition of WSS that:

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

The converse is not true!

Properties of the autocorrelation function for WSS processes

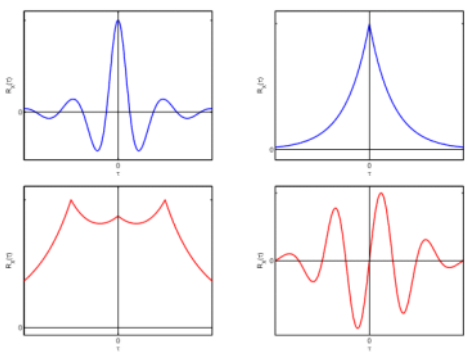
The ACF of a WSS process has a number of properties useful for checking your calculations:

1) $R_x(0) \geq 0$
2) $R_x(-\tau) = R_x(\tau) = R_x(|\tau|)$ for all $\tau \in \mathbb{T}$
3) $|R_x(\tau)| \leq R_x(0)$ for all $\tau \in \mathbb{T}$
4) $R_x(0) = \mathbb{E}[X(t)^2] = \sigma_X^2 + \mu_X^2$

In plain words this means that $R_x(\tau)$:

1) is nonnegative at zero *(When does the case $R_x(0) = 0$ occur?)*
2) is an even function (i.e. symmetric around zero)
3) has maximum magnitude at zero
4) The value at zero is the mean square of $X(t)$

Graphical examples (and counterexamples)



Positive semi-definiteness of the ACF is necessary and sufficient

Not all functions qualify as the autocorrelation function of a WSS process.

Definition:
A function $g : \mathbb{R} \to \mathbb{C}$ is called *positive semi-definite* if for each $n \in \mathbb{N}$ and $\forall t_1, t_2, \ldots, t_n \in \mathbb{R}$ and $\forall a_1, a_2, \ldots, a_n \in \mathbb{C}$ it holds that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i g(t_i - t_j) a_j^* \geq 0$$

which can be stated in matrix form as $\mathbf{a}^{*T}\mathbf{G}\mathbf{a} \geq 0$.

$R_x(\tau)$ is the autocorrelation function of a WSS random process $X(t)$
if and only if
$R_x(\tau)$ is a positive semi-definite function

The positive definiteness property of an autocorrelation function leads to all previously mentioned properties!

Proof of necessity

It is easy to prove that an autocorrelation function is positive semi-definite: Define a random variable

$$Y := \sum_{i=1}^{n} a_i X(t_i) =: \mathbf{a}^T \mathbf{X}$$

for arbitrary n and a_i's. Trivially $\mathbb{E}[|Y|^2] \geq 0$ and therefore

$$0 \leq \mathbb{E}[|Y|^2] = \mathbb{E}[\mathbf{a}^{*T}\mathbf{X}^*\mathbf{X}^T\mathbf{a}] = \mathbf{a}^{*T}\mathbb{E}[\mathbf{X}\mathbf{X}^T]\mathbf{a} = \mathbf{a}^{*T}\mathbf{R}_\mathbf{X}\mathbf{a}$$

Entry (i, j) of $\mathbf{R}_\mathbf{X}$ is $R_x(t_i - t_j)$, hence the $R_x(\tau)$ is positive semidefinite.

Remark: It is a bit harder to prove the converse, i.e. to prove that for any positive semidefinite function there exist a stochastic process with this as autocorrelation. To prove this, we essentially need to construct a process (a Gaussian process will do) with the wanted autocorrelation. (This proof is outside the scope of the course.)

Outline for Session 5 — Lecture

ILOs for Session 5

The autocorrelation and its properties

Weak sense stationary (WSS) processes

Gaussian WSS processes are SSS

Autocorrelation quiz

Gaussian WSS processes are SSS

Let $X(t)$ be a Gaussian WSS processes with covariance function $C_X(\tau)$ and define the two vectors

$$\mathbf{X} := \begin{bmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_n) \end{bmatrix} \sim \mathcal{N}(\mu_\mathbf{X}, \boldsymbol{C}_\mathbf{X}) \quad \text{and} \quad \mathbf{X}' := \begin{bmatrix} X(t_1 + \tau) \\ X(t_2 + \tau) \\ \vdots \\ X(t_n + \tau) \end{bmatrix} \sim \mathcal{N}(\mu_{\mathbf{X}'}, \boldsymbol{C}_{\mathbf{X}'}).$$

The vector $\mathbf{X}$ is Gaussian with mean vector $\mu_\mathbf{X} = \mathbb{E}[\mathbf{X}]$ and covariance matrix

$$\boldsymbol{C}_\mathbf{X} := \mathbb{E}[(\mathbf{X} - \mu_\mathbf{X})(\mathbf{X} - \mu_\mathbf{X})^T].$$

Similarly, $\mathbf{X}'$ has mean $\mu_{\mathbf{X}'}$ and covariance $\boldsymbol{C}_{\mathbf{X}'}$.

To prove that $X(t)$ is SSS we check that $\mu_\mathbf{X} = \mu_{\mathbf{X}'}$ and $\boldsymbol{C}_\mathbf{X} = \boldsymbol{C}_{\mathbf{X}'}$ hold.

Gaussian WSS processes are SSS (contd.)

Since $X(t)$ is WSS it has constant mean. Thus,

$$\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} \mu_X(t_1) \\ \mu_X(t_2) \\ \vdots \\ \mu_X(t_n) \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_X \\ \vdots \\ \mu_X \end{bmatrix} = \begin{bmatrix} \mu_X(t_1+\tau) \\ \mu_X(t_2+\tau) \\ \vdots \\ \mu_X(t_n+\tau) \end{bmatrix} = \boldsymbol{\mu}_{\mathbf{X}'}$$

Since $X(t)$ is WSS, its covariance function is shift invariant. Therefore,

$$\begin{aligned} \boldsymbol{C}_{\mathbf{X}} &= \mathbb{E}[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})^T] \\ &= \begin{bmatrix} C_X(0) & C_X(t_1-t_2) & \dots & C_X(t_1-t_n) \\ C_X(t_2-t_1) & C_X(0) & \dots & C_X(t_2-t_n) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(t_n-t_1) & C_X(t_n-t_2) & \dots & C_X(0) \end{bmatrix} \\ &= \mathbb{E}[(\mathbf{X}'-\boldsymbol{\mu}_{\mathbf{X}'})(\mathbf{X}'-\boldsymbol{\mu}_{\mathbf{X}'})^T] = \boldsymbol{C}_{\mathbf{X}'} \end{aligned}$$

We conclude that

$$\boxed{\text{If } X(t) \text{ is Gaussian and WSS it is also SSS.}}$$

24 / 26

Outline for Session 5 — Lecture

ILOs for Session 5

The autocorrelation and its properties

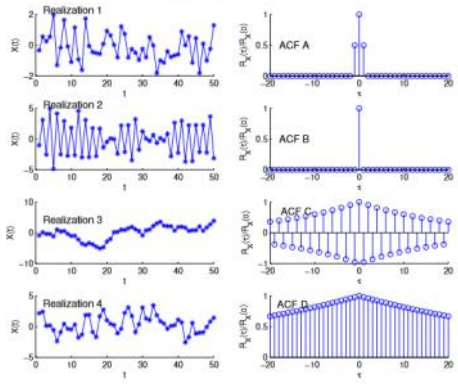
Weak sense stationary (WSS) processes

Gaussian WSS processes are SSS

Autocorrelation quiz

25 / 26

Quiz: Which realizations belong to which ACF?



26 / 26

1 - B

2 - C

3 - D

4 - A

Lec 6)

Thursday, September 25, 2025 12:41 PM

6.1)

$$X(n) = U(n) - \frac{1}{2}U(n-1), \quad n = 1, 2 \ldots$$

with $U(n) \overset{iid}{\sim} \mathcal{N}(3,9)$

a) verify that the process is WSS by computing its mean & ACF.

$$R_X(t_1, t_2) := \mathbb{E}[X(t_1)X(t_2)]$$

$$= \mathbb{E}[X(n_1) \cdot X(n_2)]$$

$$= \mathbb{E}\left[\left(U(n_1) - \frac{1}{2}U(n_1-1)\right) \cdot \left(U(n_2) - \frac{1}{2}U(n_2-1)\right)\right]$$

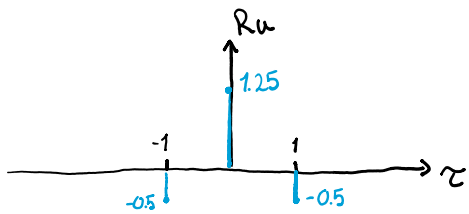
$$= \mathbb{E}\left[\left(U(n_1)U(n_2) - \frac{1}{2}U(n_1)U(n_2-1) - \frac{1}{2}U(n_1-1) \cdot U(n_2) + \frac{1}{4}U(n_1-1) \cdot U(n_2-1)\right)\right]$$

$$= \underbrace{\mathbb{E}[(U(n_1)U(n_2)]}_{R_u(n_1-n_2)} - \frac{1}{2}\underbrace{\mathbb{E}[U(n_1)U(n_2-1)]}_{R_u(n_1-(n_2-1))} - \frac{1}{2}\underbrace{\mathbb{E}[U(n_1-1)U(n_2)]}_{R_u((n_1-1)-n_2)} + \frac{1}{4}\underbrace{\mathbb{E}[U(n_1-1) \cdot U(n_2-1)]}_{R_u((n_1-1)-(n_2-1))}$$

$$= R_u(n_1-n_2) - \frac{1}{2}R_u(n_1-(n_2-1)) - \frac{1}{2}R_u((n_1-1)-n_2) + \frac{1}{4}R_u((n_1-1)-(n_2-1))$$

$$= 1 \cdot R_u(n_1-n_2) - \frac{1}{2}R_u(n_1-n_2+1) - \frac{1}{2}R_u(n_1-1-n_2) + \frac{1}{4}R_u(n_1-n_2)$$

$$= 1 \cdot \frac{1}{4}R_u(n_1-n_2) - \frac{1}{2}R_u(n_1-n_2+1) - \frac{1}{2}R_u(n_1-1-n_2)$$



Lec 7)

Tuesday, September 30, 2025 8:24 AM

S07_stochasticSyste...

Stochastic Systems

Session 7 — Lecture

Troels Pedersen

Section Wireless Communication Networks,
Department of Electronic Systems, Aalborg University

Fall 2024

1 / 24

Outline for Session 7 — Lecture

Power spectral density

Estimators — definition and associated terms

Estimation of autocorrelation and power spectrum

2 / 24

ILOs for Session 7

After having attended this lecture and solved the exercises you should be able to:

- Explain the definition and the meaning of a PSD to a fellow student.
- Compute the PSD given a particular ACF.
- Use the theoretical properties of any PSD as sanity checks of your derivations.
- Know (without hesitation and computation) the PSD of a white process.
- Know the definition of bias and MSE of an estimator and explain their meaning.

3 / 24

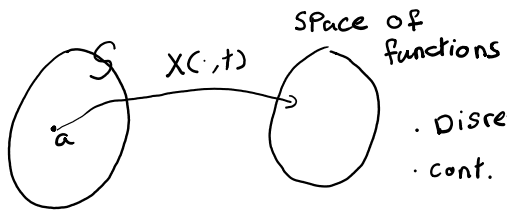
Outline for Session 7 — Lecture

Power spectral density

Estimators — definition and associated terms

Estimation of autocorrelation and power spectrum

Prob. Space



- Discrete time $t \in \mathbb{Z}$
- cont. time $t \in \mathbb{R}$

Mean:

$$\mathbb{E}[X(t)] = \mu_x(t) = \int x\, p(x)\, dx$$

Discrete-time versus continuous-time

In this lecture we need to distinguish whether a random process is defined with respect to *discrete-time* or *continuous-time*, i.e. whether $\mathbb{T} \subseteq \mathbb{Z}$ or $\mathbb{T} \subseteq \mathbb{R}$.

Discrete-time notation:
Let $\{X_n\}$ denote a real-valued discrete-time WSS random process with autocorrelation function $R_x(k) = \mathbb{E}[X_n X_{n+k}]$.

Continuous-time notation:
Let $\{X(t)\}$ denote a real-valued continuous-time WSS random process with autocorrelation function $R_x(\tau) = \mathbb{E}[X(t)X(t+\tau)]$.

Autocorrelation func.

$$R_x(t_1, t_2) = \mathbb{E}[X(t_1) \cdot X(t_2)]$$

If X is WSS:

$$\mu_x(t) = \mu_x$$

$$R_x(t, t+\tau) = R_x(\tau)$$

Power spectral density (PSD) or power spectrum

Definition: (discrete-time)
The *power spectral density* (PSD) of the process $\{X_n\}$ is the discrete-time Fourier transform of the autocorrelation function $R_x(k)$, i.e.

$$S_x(f) := \mathcal{F}\{R_x\}(f) = \sum_{k=-\infty}^{\infty} R_x(k)\exp(-j2\pi kf), \qquad |f| \le \frac{1}{2}$$

Definition: (continuous-time)
The *power spectral density* (PSD) of the process $X(t)$ is the continuous-time Fourier transform of the autocorrelation function $R_x(\tau)$, i.e.

$$S_x(f) := \mathcal{F}\{R_x\}(f) = \int_{-\infty}^{\infty} R_x(\tau)\exp(-j2\pi\tau f)\mathrm{d}\tau, \qquad f \in \mathbb{R}$$

PSD:

$$S_x(f) = \sum_{K=-\infty}^{\infty} R_x(K)\, e^{-j2\pi Kf} \qquad -\frac{1}{j2} \le f < \frac{1}{2}$$

$$= \mathcal{F}\{R_x(K)\}(f)$$

$$R_x(K) = \mathcal{F}^{-1}\{S_x(f)\}(f)$$

$$= \int_{-1/2}^{1/2} S_x(f) \cdot e^{j2\pi Kf}\, df$$

Remarks

The PSD $S_x(f)$ of a discrete-time random process is *periodic* in f with unit period since

$$\exp\big(-j2\pi k(f+1)\big) = \exp(-j2\pi kf), \qquad \forall k \in \mathbb{Z},\ \forall f \in \mathbb{R}$$

Hence, $S_x(f)$ is completely specified on any connected interval of unit length.

By convention, we use the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$

The PSD $S_x(f)$ of a continuous-time random process is *not periodic* in f.

Inverse transforms

Discrete-time:

$$R_x(k) = \mathcal{F}^{-1}\{S_x\}(k) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_x(f)\exp(j2\pi kf)\mathrm{d}f, \qquad k \in \mathbb{Z}$$

Continuous-time:

$$R_x(\tau) = \mathcal{F}^{-1}\{S_x\}(\tau) = \int_{-\infty}^{\infty} S_x(f)\exp(j2\pi\tau f)\mathrm{d}f, \qquad \tau \in \mathbb{R}$$

Interpretation: Distribution of average power (in frequency)

The mean square $\mathbb{E}[X_n^2]$ can be interpreted as the average power of a process.

The PSD can be related to the average power: *(Show this!)*

Discrete-time:

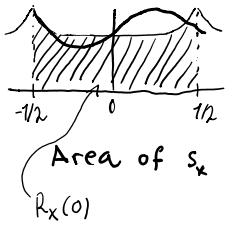
$$R_x(0) = \mathbb{E}\left[X_n^2\right] = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_x(f)\mathrm{d}f$$

Continuous-time:

$$R_x(0) = \mathbb{E}\left[X^2(t)\right] = \int_{-\infty}^{\infty} S_x(f)\mathrm{d}f$$

$$R_x(0) = \mathbb{E}[X(n)X(n+0)] = \mathbb{E}[X^2(n)]$$
$$= \sigma^2 + \mu_x^2$$

$$R_x(0) = \mathcal{F}^{-1}\{S_x(f)\}(0)$$
$$= \int_{-1/2}^{1/2} S_x(f)\underbrace{e^{j2\pi f 0}}_{1}\,df = \int_{-1/2}^{1/2} S_x(f)\,df$$



Properties of $S_x(f)$

Recall that $\{X_n\}$ and $X(t)$ are real-valued WSS random processes.
The properties below are valid no matter if time is discrete or continuous.

Properties:

1) $S_x(f) \in \mathbb{R}$ for all $f \in \mathbb{R}$
2) $S_x(-f) = S_x(f)$ for all $f \in \mathbb{R}$
3) $S_x(f) \geq 0$ for all $f \in \mathbb{R}$

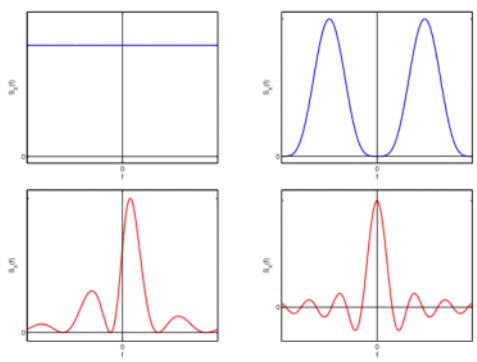
In plain words this means that $S_x(f)$:

1) is real-valued (even though the Fourier transform is a complex operation)
2) is an even function
3) is not just real-valued but in fact non-negative

Properties of PSD

1) $S_x(f)$ is a real function
2) $S_x(-f) = S_x(f)$ (Symmetric)
3) $S_x(f) \geq 0$ (Positive / non-negative)

Graphical examples (and counterexamples)



White processes

Motivated by the observation that "white light" has constant PSD, we define a white process:

Definition: A discrete-time[1] process $\{X(n)\}$ is *white* if it is WSS, and its PSD is constant, i.e.

$$S_X(f) = \begin{cases} \sigma_X^2, & |f| < \frac{1}{2} \\ 0, & |f| > \frac{1}{2} \end{cases}$$

Correspondingly, the ACF reads

$$R_X(k) = \sigma_X^2 \delta(k)$$

where $\delta(k)$ is the Kronecker delta.

Example:
Let $\{X_n\}$ be an (discrete time) i.i.d. process with zero mean and variance σ_X^2. The autocorrelation function is $R_X(k) = \mathbb{E}[X_n X_{n+k}] = \sigma_X^2 \delta(k)$ and the spectrum is $S_X(f) = \sigma_X^2, |f| \leq \frac{1}{2}$.

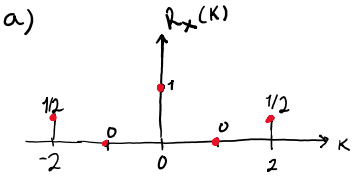
Show that the mean of a white process is zero.

[1] A similar definition applies to continuous time processes. Replace the frequency interval by the real line and the $\delta(k)$ by Dirac's delta.

12 / 24

Exercise 7.1)

a)



$$R_X(K) = \delta(K) + \tfrac{1}{2}\delta(K-2) + \tfrac{1}{2}\delta(K+2) + 0\delta(K-1) + 0\delta(K+1)$$
$$= \delta(K) + \tfrac{1}{2}\delta(K-2) + \tfrac{1}{2}\delta(K+2)$$

$$S_X(f) = \sum_{K=-\infty}^{\infty} R_X(K)\, e^{-j2\pi K f}$$

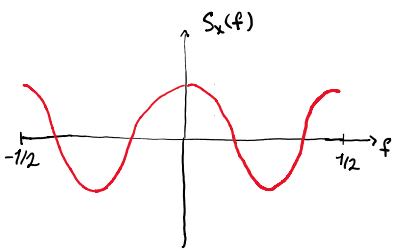
$$= R_X(-2)e^{-j2\pi(-2)f} + R_X(2)\, e^{-j2\pi(2)f} + R_X(0)e^{-j2\pi(0)f}$$

$$= \frac{1}{2}\cdot e^{j4\pi f} + \frac{1}{2}\cdot e^{-j4\pi f} + 1\cdot \underbrace{e^{-j0}}_{1}$$

euler identity:
$e^{j\varphi} = \cos(\varphi) + j\sin(\varphi)$

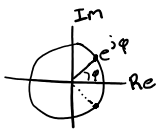
$$= \frac{1}{2}(\cos(4\pi f) + j\sin(4\pi f)) + \frac{1}{2}(\cos(-4\pi f) + j\sin(-4\pi f)) + 1\cdot 1$$

$$= \cos(4\pi f) + 1$$ (2 gange rundt om enheds cirkel



Troels note

$$e^{j\varphi} = \cos(\varphi) + j\sin(\varphi)$$



$$e^{-j\varphi} = (e^{j\varphi})^*$$

$$z + z^* = 2\,\mathrm{Re}\{z\}$$
$$e^{j\varphi} + e^{-j\varphi} = 2\cos(\varphi)$$

b) $S_X(f) = a + b\cos(2\pi f)$

b,1)

$a > b$

b,2)

when $b = 0$

b,3)

$$R_X(K) = \int_{-1/2}^{1/2} S_X(f)\, e^{j2\pi K f}\, df$$

b,3)

$$R_x(k) = \int_{-1/2}^{1/2} S_x(f) \cdot e^{j2\pi kf}\, df$$

$$= \int_{-1/2}^{1/2} (a + b\cos(2\pi f))\, e^{j2\pi kf}\, df$$

$b < |a| < \infty$

$-\infty < |b| < a$

c) $X(n) = U(n) + \mu$

$U(n)$ = white process

$Var = \sigma_u^2$

ACF:

$$R_x(t_1, t_2) = \mathbb{E}[X(t_1) \cdot X(t_2)]$$

$$= \mathbb{E}[(U(n_1) + \mu_1) \cdot (U(n_2) + \mu_2)]$$

$$= \mathbb{E}[(U(n_1)U(n_2) + U(n_1) + \mu_2) + (U(n_2)\mu_1 + \mu_1\mu_2)]$$

$$= \mathbb{E}[U(n_1)U(n_2)] + \mu_2\, \mathbb{E}[U(n_1)] + \mu_1\, \mathbb{E}[U(n_2)] + \mu_{1,2}^2$$

Lec 8

Thursday, October 2, 2025 12:33 PM

ACF for WSS x:

$R_x(k) = \mathbb{E}[x(n)x(n+k)]$

PSD $\mathcal{F}$

$S_x(f) = \mathcal{F}\{R_x(k)\}(f), \quad |f| \leq 1/2$

S07_stochasticSyste...

Outline for Session 7 — Lecture

Power spectral density

Estimators — definition and associated terms

Estimation of autocorrelation and power spectrum

13 / 24

Estimation problem

Definition:
An *estimator* is a function $g(\cdot)$ of data X (i.e. a *statistic*) used to infer on the value of an unknown quantity θ. The value $\hat{\theta} = g(X)$ is called an *estimate*.

We mark an estimate by a "hat" over the quantity we estimate.

Example: Mean estimation
Let $\{X_k\}$ be a WSS process with unknown mean μ_X. We estimate μ_X from the observed data vector $\mathbf{X} = [X_1, X_2, \ldots, X_N]^T$ as

$$\hat{\mu}_X = g(\mathbf{X}) := \frac{1}{N}\sum_{n=1}^{N} X_n = \frac{1}{N}\mathbf{1}^T\mathbf{X},$$

where $\mathbf{1} = [\underbrace{1, 1, \ldots, 1}_{N \text{ entries}}]^T$.

14 / 24

$$\mathbb{E}[x] \overset{\text{cont.}}{=} \int x\, p_x(x)\, dx$$

$$\overset{\text{disc.}}{=} \sum_x x\, \mathbb{P}(X = x)$$

Biased and unbiased estimators, mean square error

The estimate $\hat{\theta}$ is a function $g(X)$ of random data X and is therefore random.

Estimator bias: $\text{bias}_{\hat{\theta}} = \mathbb{E}[\hat{\theta} - \theta]$
An estimator is *unbiased* if $\text{bias}_{\hat{\theta}} = 0$, i.e. if $\mathbb{E}[\hat{\theta}] = \mathbb{E}[\theta]$.

Expected value of mean = bias

Mean square error (MSE): $\text{MSE}_{\hat{\theta}} = \mathbb{E}[(\hat{\theta} - \theta)^2]$

$\text{Bias}_{\hat{\theta}} = \mathbb{E}[\hat{\theta} - \theta]$

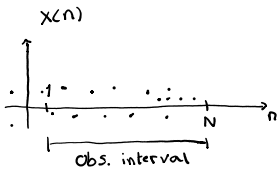
$\text{MSE}_{\hat{\theta}} = \mathbb{E}[(\hat{\theta} - \theta)^2]$

Example: Mean estimation (cont.)

$$\mathbb{E}[\hat{\mu}_X - \mu_X] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} X_n - \mu_X\right] = \underbrace{\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[X_n]}_{N \cdot \mu_X} - \mu_X = 0$$

Thus $\hat{\mu}_X$ is unbiased.

15 / 24



$X(1), \ldots, X(N)$

Mean square error for mean estimation (contd.)

The MSE of μ_X: Observe first that $\hat{\mu}_X - \mu_X = \frac{1}{N}\mathbf{1}^T(\mathbf{X} - \mathbf{1}\mu_X)$, then

$$\begin{aligned}\mathbb{E}[(\hat{\mu}_X - \mu_X)^2] &= \mathbb{E}\left[\frac{1}{N^2}\mathbf{1}^T(\mathbf{X} - \mathbf{1}\mu_X)(\mathbf{X} - \mathbf{1}\mu_X)^T\mathbf{1}\right] \\ &= \frac{1}{N^2}\mathbf{1}^T\mathbb{E}\left[(\mathbf{X} - \mathbf{1}\mu_X)(\mathbf{X} - \mathbf{1}\mu_X)^T\right]\mathbf{1} = \frac{1}{N^2}\mathbf{1}^T\mathbf{C_X}\mathbf{1}.\end{aligned}$$

where $\mathbf{1}^T\mathbf{C_X}\mathbf{1}$ is the sum of all entries of the covariance matrix $\mathbf{C_X}$:

$$\mathbf{C_X} = \begin{bmatrix} C_X(0) & C_X(1) & \ldots & C_X(N-1) \\ C_X(-1) & C_X(0) & \ldots & C_X(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(-(N-1)) & C_X(-(N-2)) & \ldots & C_X(0) \end{bmatrix}$$

$R_x(0) = \sigma_x^2 + \mu_x^2 = \mathbb{E}[X^2(n)]$

where $\mathbf{1}^T\mathbf{C_X}\mathbf{1}$ is the sum of all entries of the covariance matrix $\mathbf{C_X}$:

$$\mathbf{C_X} = \begin{bmatrix} C_X(0) & C_X(1) & \dots & C_X(N-1) \\ C_X(-1) & C_X(0) & \dots & C_X(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_X(-(N-1)) & C_X(-(N-2)) & \dots & C_X(0) \end{bmatrix}$$

Due to the special structure of $\mathbf{C_X}$ we achieve

$$\mathbb{E}[(\hat{\mu}_X - \mu_X)^2] = \frac{1}{N^2} \sum_{k=-(N-1)}^{N-1} (N - |k|) \cdot C_X(k).$$

Special case: For *uncorrelated* $\{X_k\}$ the MSE is $\mathbb{E}[(\hat{\mu}_X - \mu_X)^2] = \frac{\sigma_X^2}{N}$.

16 / 24

Outline for Session 7 — Lecture

Power spectral density

Estimators — definition and associated terms

Estimation of autocorrelation and power spectrum

17 / 24

The spectral estimation problem

Autocorrelation and Spectrum Estimation problem:
Estimate $R_X(k)$ and/or $S_X(f)$ of a discrete-time WSS process $\{X_k\}$ from $\mathbf{X} = [X_0, \dots, X_{N-1}]^T$.

I.e. find functions $\hat{R}_X(k) = r_k(\mathbf{X})$ and $\hat{S}_X(f) = s_f(\mathbf{X})$.

We will approach this problem by first estimating the autocorrelation and then obtain an estimate of the spectrum.

Here we give only a brief introduction. Details in the lecture notes [LN - Section 6.1].

18 / 24

Sample autocorrelation function

Given a N point sample $X(1), \dots, X(N)$ of a discrete-time stochastic process $\{X(n)\}$ we can estimate the ACF via the *sample autocorrelation function* defined as:

$$\hat{R}_X(k) = \begin{cases} \frac{1}{N} \sum_{n=1}^{N-k} X(n)X(n+k), & k = 0, 1, \dots, N-1 \\ \hat{R}_X(-k), & k = -(N-1), \dots, -1 \\ 0, & |k| \geq N. \end{cases}$$

$R_X(K) = R_X(-K)$

Alternatively, we can express the sample ACF in form of a convolution of a signal $X_{obs}(n)$ with its time-reverse:

$$\hat{R}_X(k) = \frac{1}{N} X_{obs}(n) * X_{obs}(-n) \qquad \underbrace{X_{obs}(n)}_{\text{subset}} = \begin{cases} X(n), & n = 1, \dots, N \\ 0, & \text{otherwise} \end{cases}$$

where $*$ denotes the convolution operator.

19 / 24

Estimation of the PSD using the periodogram

We define the *periodogram* as the Fourier transform of the sample ACF:

$$\hat{S}_X(f) = \mathcal{F}\{\hat{R}_X\}(f) = \sum_{k=-(N-1)}^{N-1} \hat{R}_X(k) \exp(-j2\pi kf), \qquad f \in [-\tfrac{1}{2}, \tfrac{1}{2}]$$

The periodogram has a simple form useful for numerical implementation:

$$\begin{aligned} \hat{S}_X(f) &= \mathcal{F}\{\tfrac{1}{N} X_{obs}(k) * X_{obs}(-k)\}(f) \\ &= \tfrac{1}{N} \mathcal{F}\{X_{obs}(k)\}(f) \cdot \mathcal{F}\{X_{obs}(-k)\}(f) \\ &= \tfrac{1}{N} \mathcal{F}\{X_{obs}(k)\}(f) \cdot \mathcal{F}\{X_{obs}(k)\}(f)^* \\ &= \tfrac{1}{N} |\mathcal{F}\{X_{obs}(k)\}(f)|^2 \end{aligned}$$

For discrete frequencies $f = \frac{m}{N}, m = 0, \dots, N-1$, the periodogram may be computed in Matlab by a one-liner:

```
Sx = 1/N * abs(fft(X)).^2;
```

20 / 24

$X(1), \dots, X(N)$

$R_X(0) = \sigma_X^2 + \mu_X^2 = \mathbb{E}[X^2(n)]$

$\hat{R}_X(0) = \frac{1}{N} \sum_{n=1}^{N} X^2(n)$

$R_X(1) = \mathbb{E}[X(n)X(n+1)]$

$= \frac{1}{N-?} \sum_{n=1}^{N-1} X(n)X(n+1)$

?

$R_X(2) = \frac{1}{N-?} \sum_{n=2}^{N-2} X(n)X(n+2)$

Bias of the the sample autocorrelation function

The sample ACF is a *biased* estimator of $R_X(k)$ which can be shown by taking the expectation of $\hat{R}_X(k)$ for $k = 0, \dots, N$:

$$\mathbb{E}[\hat{R}_X(k)] = \frac{1}{N}\sum_{n=1}^{N-k}\mathbb{E}[X(n)X(n+k)] = \frac{N-k}{N}\cdot R_X(k).$$

For general k, we have $\mathbb{E}[\hat{R}_X(k)] = w_B(k)\cdot R_X(k)$ where $w_B(k)$ is called the *Bartlett window*:

$$w_B(k) = \begin{cases} \frac{N-|k|}{N}, & |k| \le N \\ 0, & otherwise \end{cases}$$

Since $\mathbb{E}[\hat{R}_X(k)] \neq R_X(k)$, we conclude that the sample ACF is biased.

Bias of the periodogram

Also the periodogram is biased:

$$\begin{aligned}\mathbb{E}[\hat{S}_X(k)] &= \mathbb{E}[\mathcal{F}\{\hat{R}_X(k)\}(f)] \\ &= \mathcal{F}\{\mathbb{E}[\hat{R}_X(k)]\}(f) \\ &= \mathcal{F}\{w_B(k)\cdot R_X(k)\}(f) \\ &= \mathcal{F}\{w_B(k)\}(f) * S_X(f)\end{aligned}$$

Thus the spectrum is on average smeared by the Fourier transform of the Bartlett window:

$$\mathcal{F}\{w_B(k)\}(f) = \left(\frac{sin(\pi f N)}{sin(\pi f)}\right)^2$$

which is sometimes called the "Fejér kernel".

Unbiased sample autocorrelation function

In an attempt to "repair" the sample ACF and the periodogram, we may look at the *unbiased sample ACF*:

$$\begin{aligned}\check{R}_X(k) &= \frac{\hat{R}_X(k)}{w_B(k)} \\ &= \begin{cases} \frac{1}{N-k}\sum_{n=1}^{N-k}X(n)X(n+k), & k = 0,1,\dots,N-1 \\ \check{R}_X(-k), & k = -(N-1),\dots,-1 \\ 0, & |k| \ge N. \end{cases}\end{aligned}$$

It is straightforward to check that $\check{R}_X(k)$ is unbiased for $|k| \le N-1$:

$$\mathbb{E}[\check{R}_X(k)] = w_r(k)\cdot R_X(k)$$

where $w_r(k)$ is the *rectangular window* defined as

$$w_r(k) = \begin{cases} 1, & |k| \le N-1 \\ 0, & \text{otherwise.} \end{cases}$$

Caveat: The unbiased sample ACF is essentially worthless for large time-lags $|k|$ because it exhibits high variance!

PSD estimator for the unbiased sample ACF

Similar to the periodogram, we can define a PSD estimator based un the unbiased ACF as

$$\check{S}_X(f) = \mathcal{F}\{\check{R}_X(k)\}(f)$$

This does not yield an unbiased spectral estimate of the PSD:

$$\begin{aligned}\mathbb{E}[S_X(f)] &= \mathbb{E}[\mathcal{F}\{\check{R}_X(k)\}(f)] \\ &= \mathcal{F}\{\mathbb{E}[\check{R}_X(k)]\}(f) \\ &= \mathcal{F}\{w_r(k)\cdot R_X(k)\}(f) \\ &= \mathcal{F}\{w_r(k)\}(f) * S_X(f)\end{aligned}$$

The discrete Fourier transform of a rectangular function (called the Dirichlet kernel) reads

$$\mathcal{F}\{w_r(k)\}(f) = \frac{\sin(2\pi f N)}{\sin(\pi f)}$$

The Dirichlet kernel is negative for some f, and gives rise to negative PSD estimates at some frequencies. What a high price to pay for an unbiased ACF estimate!

8.1)

$$X(n) = U(n) + aU(n-1) + b, \quad U(n) \overset{i.i.d.}{\sim} \mathcal{N}(0,1) \quad n \in \mathbb{Z}$$

· 8.1.1)

$$\begin{aligned}R_X(k) &= \mathbb{E}[X(n_1)\cdot X(n_2)] \\ &= \mathbb{E}[(U(n_1) + aU(n_1-1) + b)\,(U(n_2) + aU(n_2-1) + b)] \\ &= \mathbb{E}[(U(n_1)U(n_2) + U(n_1)aU(n_2-1) + U(n_1)b) + (aU(n_1-1)U(n_2) + aU(n_1-1)aU(n_2-1) + aU(n_1-1)b)(bU(n_2) + b\cdot aU(n_2-1) + b^2)] \\ &= \mathbb{E}[U(n_1)U(n_2)] + a\,\mathbb{E}[U(n_1)U(n_2-1)] + b\,\mathbb{E}[U(n_1)] + a\,\mathbb{E}[U(n_1-1)U(n_2)] + a^2\,\mathbb{E}[U(n_1-1)U(n_2-1)] + ab\,\mathbb{E}[U(n_1-1)] + b\,\mathbb{E}[U(n_2)] + ba\,\mathbb{E}[U(n_2-1)] + b^2 \\ &= \mathbb{E}[\underbrace{U(n_1)-U(n_2)}_{n_1-n_2=0}] + a\,\mathbb{E}[\underbrace{U(n_1)-(U(n_2-1))}_{n_1-(n_2-1)=n_1-n_2+1}] + b\,\mathbb{E}[\underbrace{U(n_1)}_{n_1}] + a\,\mathbb{E}[\underbrace{(U(n_1-1))-U(n_2)}_{(n_1-1)-n_2=n_1-1-n_2}] + a^2\,\mathbb{E}[\underbrace{(U(n_1-1))-(U(n_2-1))}_{(n_1-1)-(n_2-1)=n_1-n_2}] + ab\,\mathbb{E}[\underbrace{U(n_1-1)}_{n_1-1}] + b\,\mathbb{E}[\underbrace{U(n_2)}_{n_2}] + ba\,\mathbb{E}[\underbrace{U(n_2-1)}_{n_2-1}] + b^2 \\ &= R_X(0) + aR_X(n_1-n_2+1) + b\cdot 0 + aR_X(n_1-1-n_2) + a^2R_X(n_1-n_2) + abR_X(n_1-1) + b\cdot 0 + baR_X(n_2-1) + b^2 \\ &= R_X(0) + aR_X(+1) + 0 + aR_X(-1) + a^2R_X(0) + abR_X(-1) + b\cdot 0 + ba(-1) + b^2\end{aligned}$$

$R_X(n_1-n_2)$

$\mathcal{N}(0,1)$ variable $n=0$

$\mathcal{N}(0,1)$
varible $n=0$

$$= R_x(0) + a R_x(n_1-n_2+1) + b\cdot 0 \quad + a R_x(n_1-1-n_2) + a^2 R_x(n_1-n_2) + ab R_x(n_1-1) + b\cdot 0 \quad + ba R_x(n_2-1) + b^2$$

$$= R_x(0) + a R_x(+1) + \quad 0 + a R_x(-1) + a^2 R_x(0) + ab R_x(-1) + b\cdot 0 + ba(-1) + b^2$$

$$= (1+a^2) R_x(0) + a R_x(+1) + (a^3+b^2) R_x(-1) + b^2$$

idk noget er galt

Lec 9

Tuesday, October 7, 2025 8:20 AM

S09_stochasticSyste...

Stochastic Systems

Session 9 — Lecture

Troels Pedersen

Section Wireless Communication Networks,
Department of Electronic Systems, Aalborg University

Fall 2025

1 / 16

Outline for Session 9 — Lecture

- Linear time-invariant system with WSS input
- 1st and 2nd order characterization
- Summary of results
- Examples

2 / 16

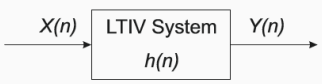
Outline for Session 9 — Lecture

- Linear time-invariant system with WSS input
- 1st and 2nd order characterization
- Summary of results
- Examples

3 / 16

Linear time-invariant system with WSS input

We consider a *Linear time-invariant (LTIV) system* with input $\{X(n)\}$ and output $\{Y(n)\}$:



From linear system theory, we know that the input-output relation of a LTIV system is in the form of a *convolution* with an *impulse response* $h(n)$:

$$Y(n) = \sum_{k=-\infty}^{\infty} h(k)X(n-k) = h(n) \star X(n)$$

We consider for a weak sense stationary (WSS) $\{X(n)\}$ with mean μ_X and autocorrelation $R_X(k)$, the questions:

- How is the mean of $\{Y(n)\}$ related to properties of $\{X(n)\}$?
- What is the autocorrelation of $\{Y(n)\}$— how does it relate to properties of $\{X(n)\}$?
- Is $\{Y(n)\}$ WSS?
- If so, what can we say about the PSD?

4 / 16

Convolution and Fourier Relations

Let $x(n)$ be $\delta(n)$ then $Y(n) = h(n)$ ← imp. response

$$Y(n) = \sum_{k=-\infty}^{\infty} h(n)X(n-k) = h(n) * X(n) = (h * X)(n)$$

$$H(f) = \mathcal{F}\{h(n)\}(f) = \sum^{\infty} h(n) e^{-j2\pi f n}$$

- is $\{Y(n)\}$ WSS?
- If so, what can we say about the PSD?

$$H(f) = \mathcal{F}\{h(n)\}(f) = \sum_{n=-\infty}^{\infty} h(n) e^{-j2\pi f n}$$

Convolution and Fourier Relations

Knowledge of linear systems and Fourier transforms are prerequisites for the course. In this lecture, we make use of the following:

Convolution is a linear operator defined as:

$$g(n) \star h(n) := \sum_{k=-\infty}^{\infty} g(k)h(n-k) = \sum_{i=-\infty}^{\infty} g(n-i)h(i) = h(n) \star g(n)$$

Discrete-time Fourier transform is a linear operator defined as:

$$\mathcal{F}\{g(n)\}(f) := \sum_{n=-\infty}^{+\infty} g(n)e^{-j2\pi nf} = G(f).$$

Fourier transforms of convolutions simplify due to the *Convolution Theorem*:

$$\mathcal{F}\{g(n) \star h(n)\}(f) = G(f) \cdot H(f)$$

Fourier transforms of signals flipped in time *(Time Reversal Identity)*:

$$\mathcal{F}\{g(-n)\}(f) = G^*(f)$$

where * denotes complex conjugation.

In combination, the convolution theorem and the time reversal identity, give

$$\mathcal{F}\{g(n) \star g(-n)\}(f) = G(f) \cdot G^*(f) = |G(f)|^2$$

Outline for Session 9 — Lecture

Linear time-invariant system with WSS input

1st and 2nd order characterization

Summary of results

Examples

1st $\mathbb{E}[X(n)]$

2st $\mathbb{E}[X(n_1)X(n_2)]$

3st $\mathbb{E}[X(n_1)X(n_2)X(n_3)]$

$\vdots$

Mean function of $\{Y(n)\}$

We compute the mean function of the output $\{Y(n)\}$ as[1]

$$\mu_Y(n) := \mathbb{E}[Y(n)] = \mathbb{E}[h(n) \star X(n)]$$

Linearity

$$= \mathbb{E}\left[\sum_{j=-\infty}^{\infty} h(j)X(n-j)\right]$$

$$= \sum_j h(j)\mathbb{E}[X(n-j)] = \mu_X \sum_j h(j).$$

The mean of $\{Y(n)\}$ is constant and can be related to the mean of $\{X(n)\}$ and the impulse response $h(n)$.

It is sometimes more convenient to write this relation in terms of the systems *transfer function* $H(f) = \mathcal{F}\{h(n)\}(f)$:

$$\mu_Y = H(0) \cdot \mu_X$$

Check this relation!

[1] If not otherwise mentioned, the summations are from $-\infty$ to $+\infty$.

Autocorrelation function for $\{Y(n)\}$ — Part 1 of 3

The derivation of ACF the output process is tedious but straightforward: Insert $\{Y(n)\}$ in the definition of autocorrelation and then simplify (carefully).

Since $\{Y(n)\}$ could be nonstationary, we use the general form of the ACF:

$$\begin{aligned} R_Y(n, n+k) &= \mathbb{E}[Y(n)Y(n+k)] \\ &= \mathbb{E}[Y(n)\sum_j h(j)X(n+k-j)] \\ &= \sum_j h(j)\underbrace{\mathbb{E}[Y(n)X(n+k-j)]}_{=:R_{YX}(n,n+k-j)} \\ &= \sum_j h(j)R_{YX}(n, n+k-j) = h(k) \star R_{YX}(n, n+k) \end{aligned}$$

where it still remains to derive the *cross-correlation function* $R_{YX}(n, n+k)$.

Autocorrelation function for $\{Y(n)\}$ — Part 2 of 3

We now derive an expression for the *cross-correlation function* $R_{YX}(n, n+k)$:

$$\begin{aligned} R_{YX}(n, n+k) &= \mathbb{E}[Y(n)X(n+k)] \\ &= \mathbb{E}\left[\sum_j h(j)X(n-j)X(n+k)\right] \\ &= \sum_j h(j)\underbrace{\mathbb{E}\left[X(n-j)X(n+k)\right]}_{=R_X(k+j)} \\ &= \sum_j h(j)R_X(k+j) \end{aligned}$$

This is a function of k only, and thus we write $R_{YX}(n, n+k) = R_{YX}(k)$.
Then the cross-correlation can be written is the form

$$R_{YX}(k) = h(-k) \star R_X(k)$$

Autocorrelation function for $\{Y(n)\}$ — Part 3 of 3

Combining the pieces of the derivation of the ACF we finally achieve:

$$\begin{aligned} R_Y(n, n+k) &= h(k) \star R_{YX}(n, n+k) \\ &= h(k) \star R_{YX}(k) \\ &= h(k) \star h(-k) \star R_X(k). \end{aligned}$$

Since it has constant mean and its ACF is a function of k only, *the output process is WSS* with autocorrelation

$$R_Y(k) = h(k) \star h(-k) \star R_X(k).$$

Remark: The symmetric and positive semidefinite function $h(k) \star h(-k)$ is some-times called the "autocorrelation of the impulse response $h(k)$" and denoted by $R_h(k)$. This is because

$$h(k) \star h(-k) = \sum_j h(j)h(j+k).$$

The term is used for historical reasons, but is strictly speaking a misnomer, since it is no in accordance with the definition of ACF.

$$\underbrace{R_Y(n, n+k)}_{R_Y(k)} = h(n) * \underbrace{R_{Y_X}(n, n+k)}_{R_{YX}(k)}$$

$$\boxed{R_Y(k) = R_h(k) * R_X(k)}$$

$$\boxed{\mu_Y(n) = \mu_Y = \mu_X \sum_j h(j)}$$

$$= H(0)\mu_X$$

DC gain

$$S_Y(f) = \mathcal{F}\{R_h(k)\}(f) \cdot S_X(f)$$

$$= \mathcal{F}\{h(k) * h(-k)\}(f) \cdot S_X(f)$$

$$= H(f) \cdot H^*(f) \cdot S_X(f)$$

$$\boxed{S_Y(f) = |H(f)|^2 S_X(f)}$$

Power spectral density of the output process

Since the output process is WSS, its power spectral density is well defined.
Starting from the autocorrelation of $\{Y(n)\}$

$$R_Y(k) = h(k) \star h(-k) \star R_X(k)$$

we achieve by Fourier transforming and using the convolution theorem and time reversal identity the relation:

$$S_Y(f) = H(f) \cdot H^*(f) \cdot S_X(f) = |H(f)|^2 S_X(f).$$

Outline for Session 9 — Lecture

Linear time-invariant system with WSS input

1st and 2nd order characterization

Summary of results

Examples

Summary of results: 1st and 2nd order input-output relation

	Input	Output
Signal	$X(n)$	$Y(n) = h(n) \star X(n)$
Mean	μ_X	$\mu_Y = \mu_X \sum_n h(n) = H(0)\mu_X$
ACF	$R_X(k)$	$R_Y(k) = h(k) \star h(-k) \star R_X(k)$
WSS	Yes	Yes
PSD	$S_X(f)$	$S_Y(f) = \lvert H(f)\rvert^2 S_X(f)$

Outline for Session 9 — Lecture

Linear time-invariant system with WSS input

1st and 2nd order characterization

Summary of results

Examples

Example: MA(1) process with zero mean

Consider an MA(1) process defined as

$$Y(n) = \overbrace{X(n) + X(n-1)}^{\text{LTI System}}, \qquad \{X(n)\} \overset{iid.}{\sim} \mathcal{N}(0,3)$$

Observe that $Y(n) = h(n) \star X(n)$ with impulse response $h(n) = \delta(n) + \delta(n-1)$.

The input $X(n)$ is a white process with $R_X(k) = 3\delta(k)$.

Mean: $\mathbb{E}[Y(n)] = \mathbb{E}[X(n)] \cdot (h(0) + h(1)) = 0 \cdot 2 = 0$

ACF: Note that $R_h(k) = h(k) \star h(-k) = \delta(k+1) + 2\delta(k) + 1\delta(k-1)$.
Then

$$\begin{aligned} R_Y(k) &= \left[\delta(k+1) + 2\delta(k) + \delta(k-1)\right] \cdot 3 \\ &= 3\delta(k+1) + 6\delta(k) + 3\delta(k-1) \end{aligned}$$

PSD : Note that $X(f)$ has PSD $S_X(f) = 3$.
The transfer function reads
$H(f) = 1 + \exp(-j2\pi f)$.
Then,

$$S_Y(f) = |1+\exp(-j2\pi f)|^2 3 = 6\cdot(1+\cos(2\pi f))$$

Example: MA(1) process with non-zero mean

Now, consider an MA(1) process with non-zero mean defined as

$$Y(n) = X(n) + X(n-1), \qquad \{X(n)\} \overset{iid.}{\sim} \mathcal{N}(3,15)$$

As before, $Y(n) = h(n) \star X(n)$ for $h(n) = \delta(n) + \delta(n-1)$.

The input has mean 3 and ACF $R_X(k) = 9 + 15\delta(k)$.

Mean: $\mathbb{E}[Y(n)] = \mathbb{E}[X(n)] \cdot (h(0) + h(1)) = 3 \cdot 2 = 6$

ACF: Since $h(k) \star h(-k) = \delta(k+1) + 2\delta(k) + 1\delta(k-1)$, we have

$$\begin{aligned} R_Y(k) &= (9 + 15\delta(k+1)) + 2 \cdot (9 + 15\delta(k)) + (9 + 15\delta(k-1)) \\ &= 36 + 15\delta(k+1) + 30\delta(k) + 15\delta(k-1). \end{aligned}$$

PSD: The PSD for $X(f)$ is $S_X(f) = 15 + 9\delta(f)$ and the transfer function reads $H(f) = 1 + \exp(-j2\pi f)$, thus the PSD for $Y(n)$ is [2]

$$\begin{aligned} S_Y(f) &= |1 + \exp(-j2\pi f)|^2 (15 + 9\delta(f)) \\ &= 2 \cdot (1 + \cos(2\pi f)) \cdot (15 + 9\delta(f)) \\ &= 30 \cdot (1 + \cos(2\pi f)) + 36\delta(f) \end{aligned}$$

[2] The Dirac impulse $36\delta(f)$ is due to the non-zero mean of $Y(n)$. The coefficient 36 equals the squared mean.

$$\begin{aligned} R_h(k) &= h(n) * \overbrace{h(-k)}^{g(k)} \\ &= h(n) * g(k) \\ &= \sum_n h(n)\, g(k-n) \\ &= \sum_n h(n)\, h(-(k-n)) \\ &= \sum_n h(n)\, h(-k+n) \\ &= \sum_n (\delta(n) + \delta(n-1))\, h(-k+n) \\ &= h(-k) + h(-k+1) \\ &= \delta(-k) + \delta(-k-1) + \delta(-k+1) + \delta(-k+1-1) \\ &= 2\delta(k) + \delta(k+1) + \delta(k-1) \end{aligned}$$

Lec 10

Thursday, October 9, 2025 12:32 PM

Autoregresive Moving-average, ARMA

10.1 ARMA(1,1) Process:

$$X(n) = aX(n-1) + bZ(n-1) + Z(n) \quad , \quad n = 1,2,3,\dots$$

a & b Known Parameters
$\{Z(n)\} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_Z^2)$

1)
Z er normal fordeit.

folding af two PDF'er er en ny PDF

Derfor er den stadig normal fodeit (Hvis den er skaleret, så er den stadig normalt fordeit)

Linear transformation. (LTI System, garantere linear

2)
$b = 0$
AR = (1)

Calculate mean & variance for $\{X(n)\}$ $(X(0) = 0)$

$$\mathbb{E}[X(0)] = 0$$
$$\mathbb{E}[X(1)] = \mathbb{E}[aX(1-1) + bZ(1-1) + Z(1)] = 0$$
$$\mathbb{E}[X(2)] = \mathbb{E}[aX(2-1) + bZ(2-1) + Z(2)] = 0$$

$$\mu_X = a \cdot \mathbb{E}[X(n-1)] + b\,\mathbb{E}[Z(n-1)] + \mathbb{E}[Z(n)]$$
$$= a \cdot \mathbb{E}[X(1-1)] + b \cdot 0 + 0$$
$$= a \cdot 0 + b \cdot 0 + 0$$
$$= 0$$

$\mathbb{E}[X(n)] = 0$

$$\sigma_X^2 = \mathbb{E}[(X(n) - \mathbb{E}[X(n)])^2]$$
$$= \mathbb{E}[X(n)^2]$$
$$= \mathbb{E}[(a \cdot X(n-1) + bZ(n-1) + Z(n)) \cdot (a \cdot X(n-1) + bZ(n-1) + Z(n))]$$
$$= \mathbb{E}[((a \cdot X(n-1)) \cdot (aX(n-1)) + ((aX(n-1)) \cdot (bZ(n-1)) + ((aX(n-1)) \cdot (Z(n)))]$$
$$= \mathbb{E}[(a^2\, 2X(n-1)) + (ab\, X(n-1)\, Z(n-1)) + aX(n-1) \cdot Z(n)]$$

Lec 12 - self study

Saturday, October 18, 2025 2:34 PM

Hidden Markov Models

- Typiske instancer er i.i.d.
  - Kan finde sandsynligheden af et sample ved at tage productet af hver individuel sample
- Instancer der er dependent, kan man ikke gøre det samme som i.i.d. instancer
  - Exp.: i engelsk er sandsynligheden for at 'h' kommer efter 't' ikke den samme som 'h' efter 'x' i et ord

    (Dette kaldes sekvenser)
  - En sekvens kan karakteriseres som at blive genereret af "Parametric random process"

Discrete Markov Processes

- A system (at any time) is in a set of N distinct states:

  $S_1, S_2, \dots, S_N$
  - The state at time t: q_t, $t = 1, 2, \dots$

    $q_t = S_i$ (at time t, the state is in S_i)
- A system switches states with a given prob. depending on the values of previous states:

  $P(q_{t+1} = S_j | q_t, q_{t-1} = S_k, \dots)$
- Markov model
  - For a 1. order Markov model the state at time t+1 only depends on the state at time t, regardless of previous times of states:

    $P(q_{t+1} = S_j | q_t = S_i, q_{t-1} = S_k, \dots) = P(q_{t+1} = S_j | q_t = S_i)$

    (The future is independent of the past)
- Transition probabilities

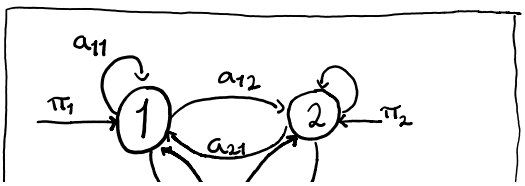
  $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$

  Assuming that the transition probabilities are independent of time

  Satisfying:

  $a_{ij} \geq 0 \quad \& \quad \sum^{N} a_{ij} = 1$

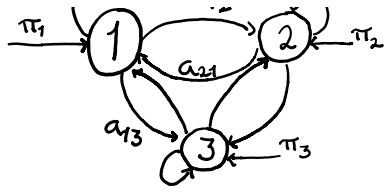
Markov model



satisfying:

$$a_{ij} \geq 0 \quad \& \quad \sum_{j=1}^{N} a_{ij} = 1$$

- **Stochastic automation**
  - Going from S_i to S_j is the same prob. no matter the time
  - From each state S_i, the system moves to state S_j with prob. a_{ij}. This prob. is the same for any t



a Markov model with 3 states is a stochastic automation

- **Initial Probabilities**
  - The only speciel case is the first state
  - Initial probabilities, π_i, which is the prob. that the first state is in the sequence S_i:

    $\pi_i = P(q_1 = S_i)$

    Satisfying

    $$\sum_{i=1}^{N} \pi_i = 1$$

Product symbol

$\bar{\pi} = [\pi_i]$ is a vector of N elements that sum to 1

$\bar{\bar{A}} = [a_{ij}]$ is a N×N Matrix whose rows sum to 1

- **Observable Markov Model**
  - Observation sequence: O
  - State sequence: $O = Q = \{q_1, q_2, \ldots, q_T\}$

    Prob. given by:

    $$P(O = Q \mid \bar{\bar{A}}, \bar{\pi}) = P(q_1) \prod_{t=2}^{T} P(q_t \mid q_{t-1}) = \pi_{q_1} a_{q_1 q_2} \cdots a_{q_{T-1} q_T}$$

    (π_{q_1} = Prob. of first state being q_1)

    ($a_{q_1 q_2}$ = Prob. of going from q_1 to q_2 & so on)

  - Multiply all these prob. to get the prob. of whole sequence.