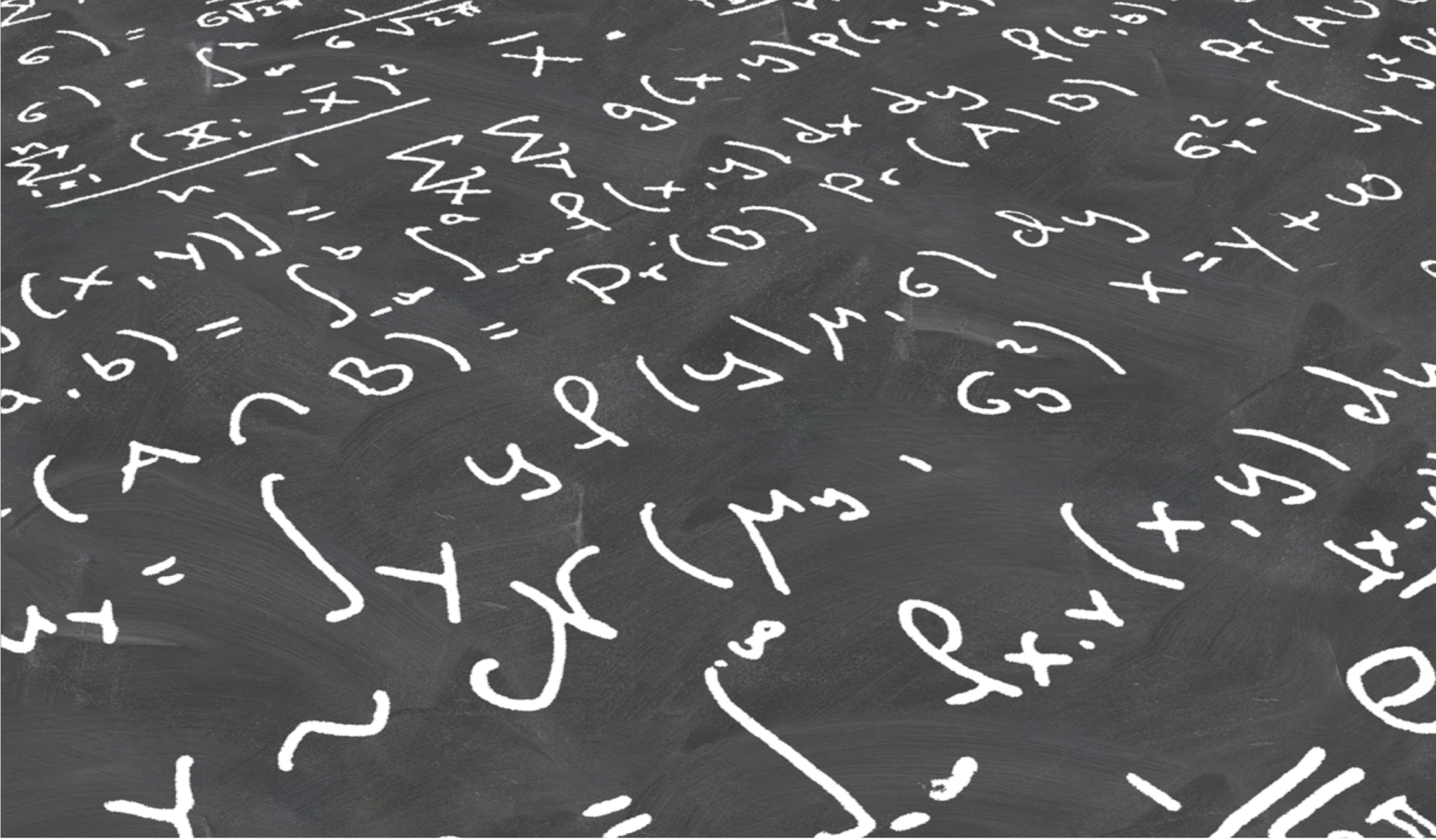
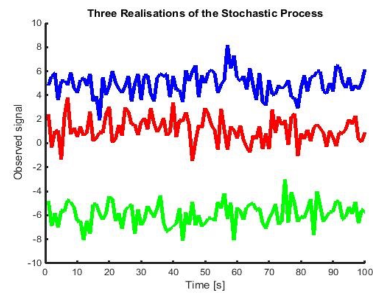


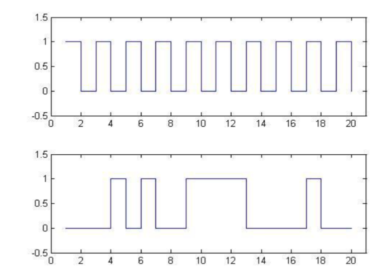
Introduction to Stochastic Modelling and Processes (SMP)

Gunvor Elisabeth Kirkelund
Lars Mandrup

Why Stochastic Modelling and Processes?

- All scientific and engineering work contain some element of randomness
  - How can I get anything out of this noisy signal?
  - How much can I conclude from my measurements?
  - How many tests / size of population do I have to do to validide my system/method/model?





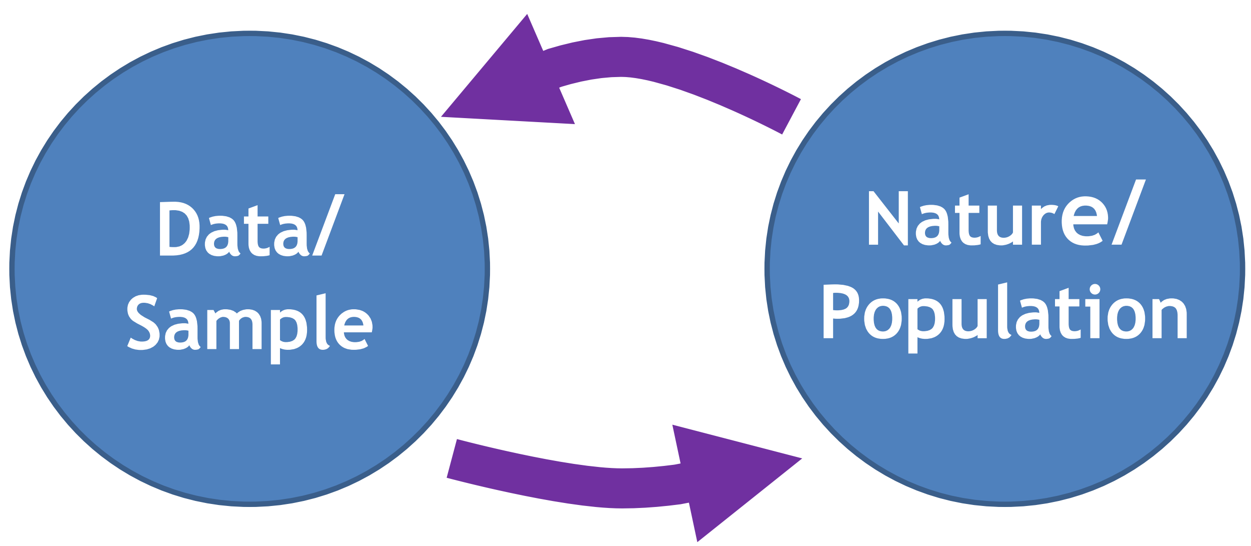
- Stochastic processes is a way to handle and modelling the randomness
- Mandatory if you want to take a master degree

Content of the Course

Probabilistics
- Probability theory ← *Without time variable*
- Stochastic Processes ← *With time variable*
- Statistics

- Probability theory tells us what is in the sample given nature.

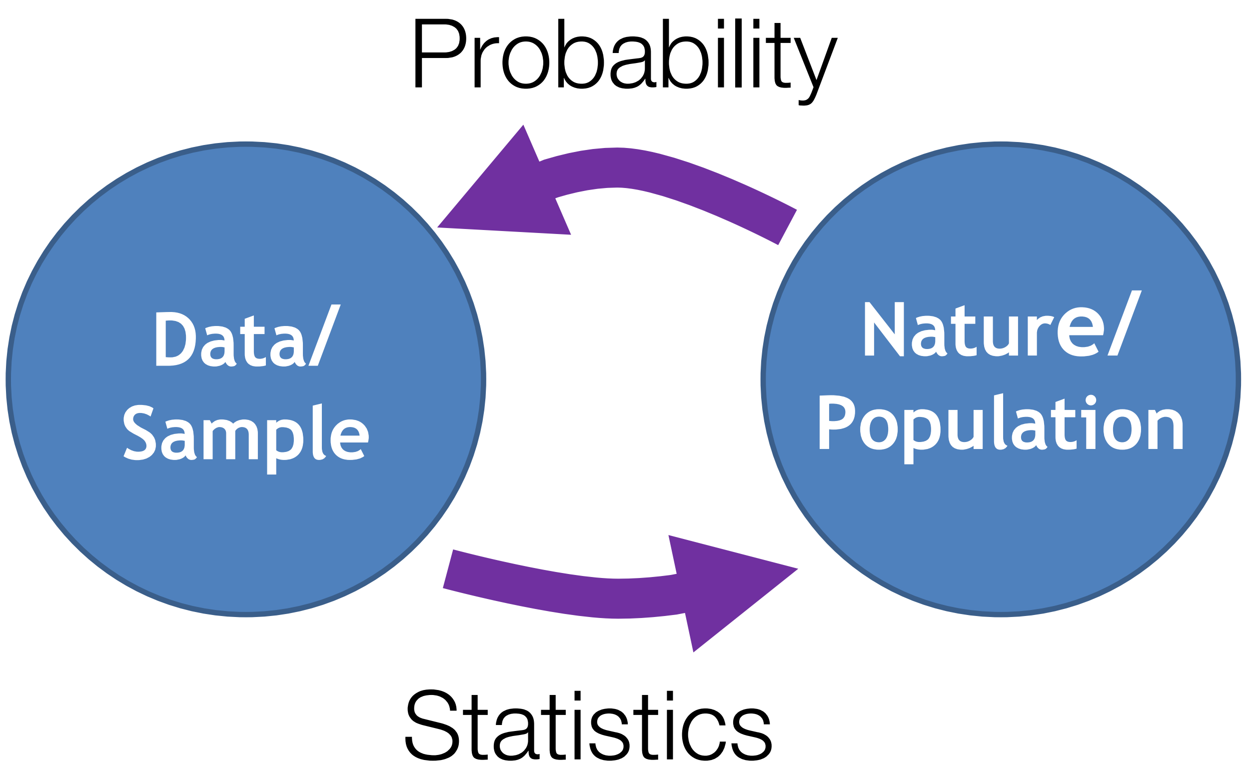
Given a regular dice probability theory can tell me how many times I will get a 6, when I roll the dice 100 times.



- Statistics tells us about nature given the sample.

Rolling an unknown dice 100 times, I 12 times get a 6.
Statistics tell me the nature of the dice (is it regular or not).

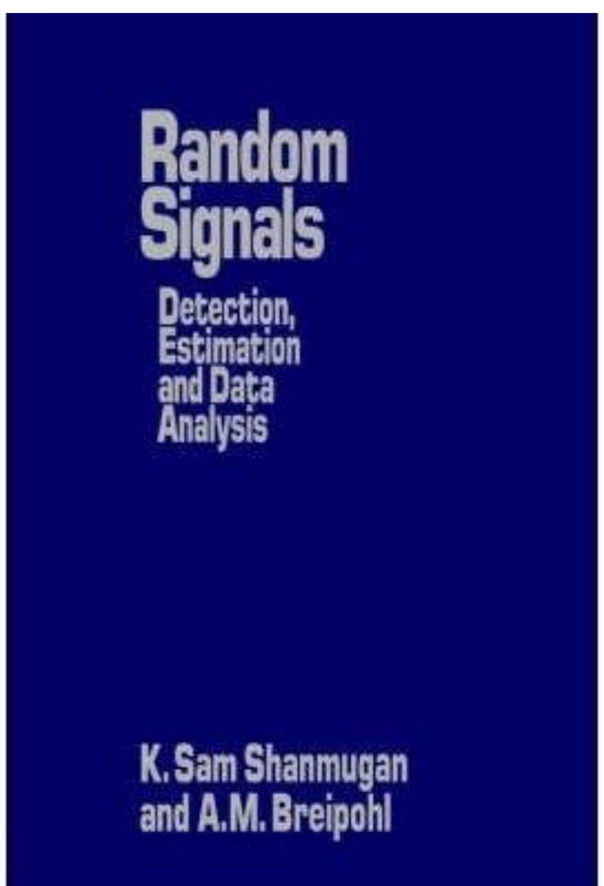
Topics

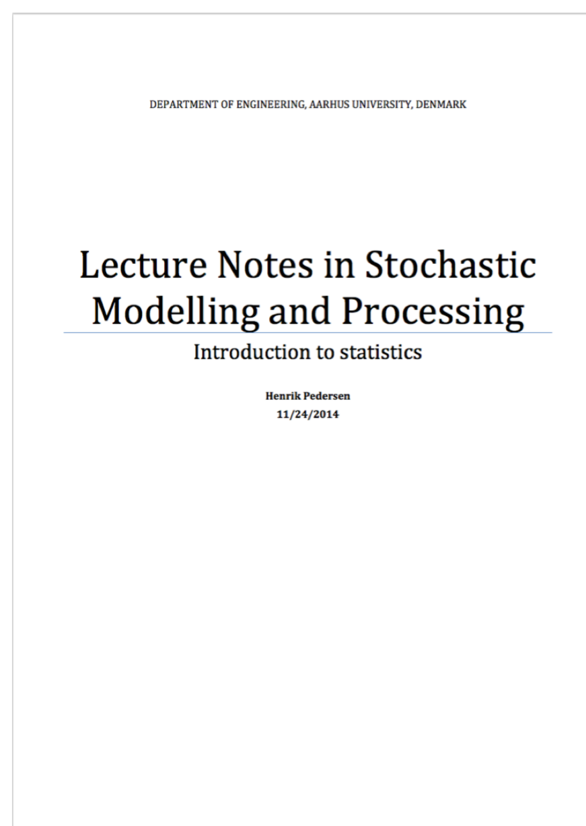


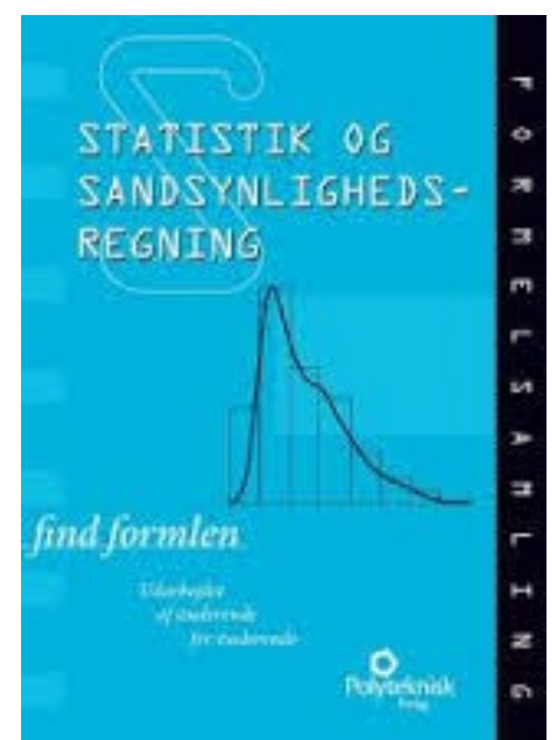
- Probability Theory and Stochastic Processes
  - Basic probability theory
  - Probability distributions
  - Stochastic (random) processes
- Statistics
  - Statistical tests
  - Model estimation
  - Linear regression

Curriculum

- “*Random Signals*”, Shanmugan and Breipohl, chap. 2, 3 and 8
- “*Lecture Notes in Stochastic Modelling and Processing*”, H. Pedersen (supplementary - statistics)
- “*Statistik og Sandsynlighedsregning - find formlen*”







Course Format

- One 4 hours lesson each week in 14 weeks

Exam

- 3 hours written exam

Teachers:

- Lars Mandrup, room 306E, lma@ase.au.dk
- Gunvor Elisabeth Kirkelund, room 300E, gek@ase.au.dk

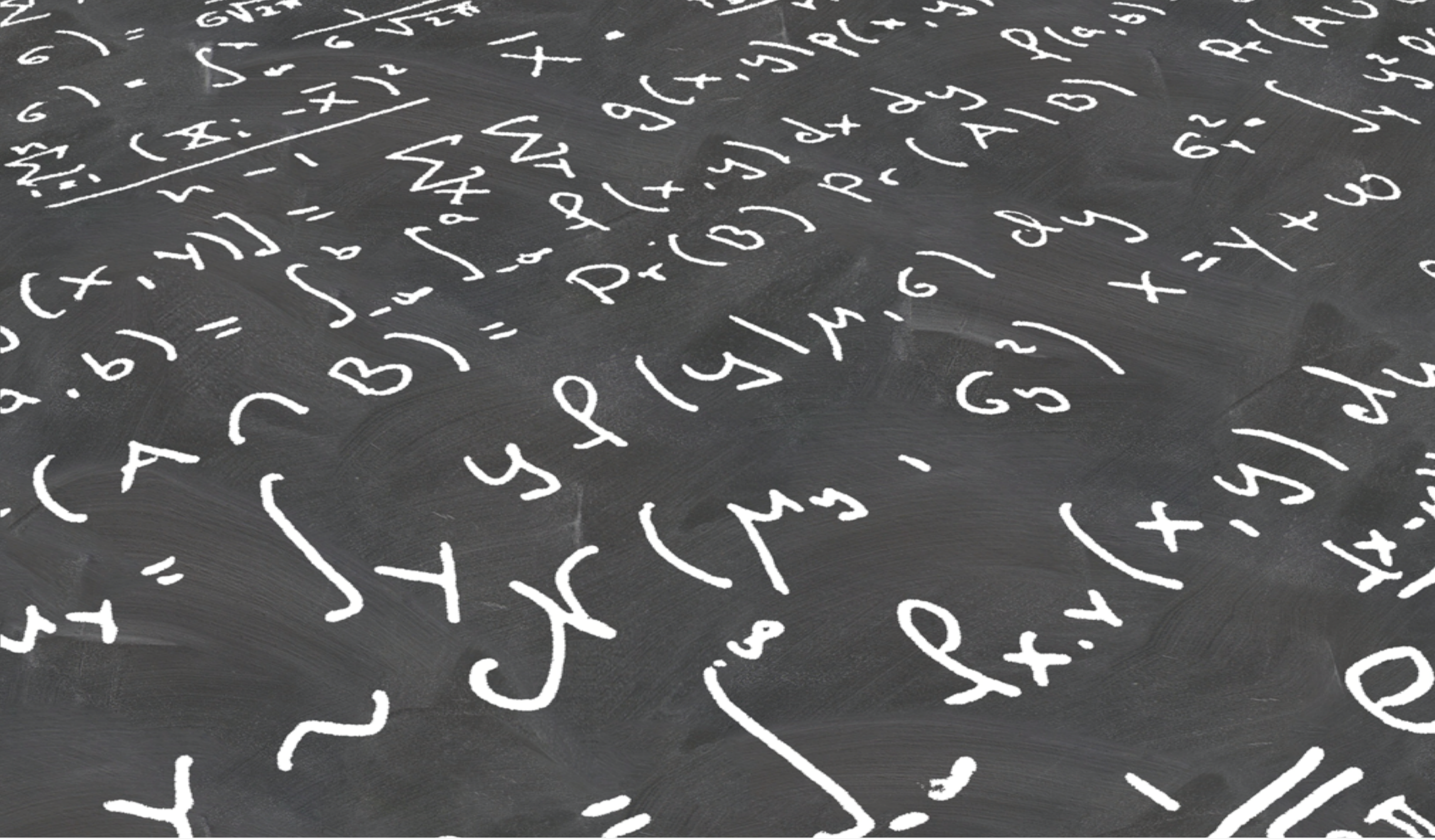
Lecture Format

- The course will consist of 4 hour lessons each week.
- In the first 2 hours we will give introductions to each topic.
- In the second 2 hours you will work in groups to solve different types of Group Assignment problems related to the curriculum.
- As the time during these 4 hour sessions are very limited we expect you to have read the curriculum and solved the introductory assignments before the lecture. Solutions will be provided to all introductory assignments.

What we expect before you begin the course

- You know how to integrate and differentiate
- Know about convolution of two functions
- Have a basic knowledge of Matlab

- And have installed Matlab



1. Introduction to Probability Theory

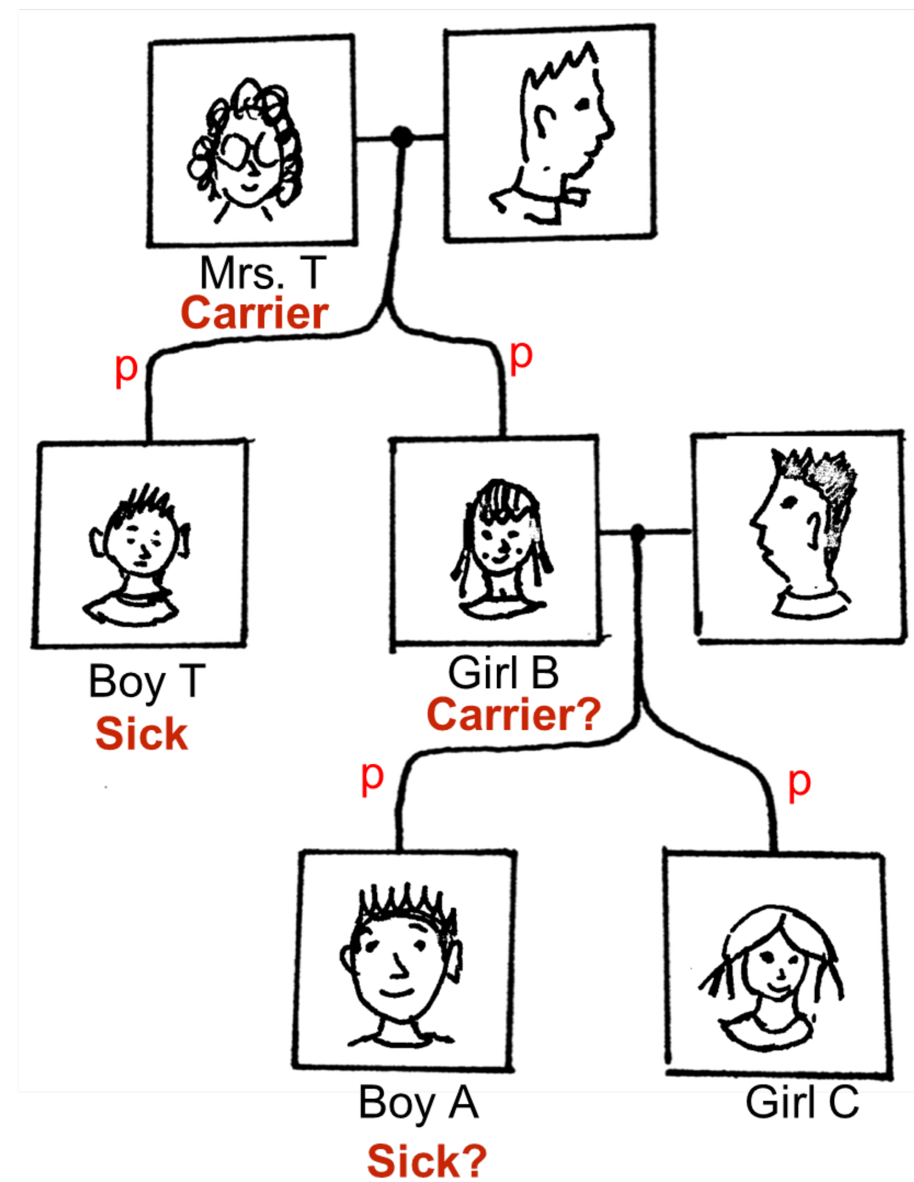
Gunvor Elisabeth Kirkelund
Lars Mandrup

Todays Content

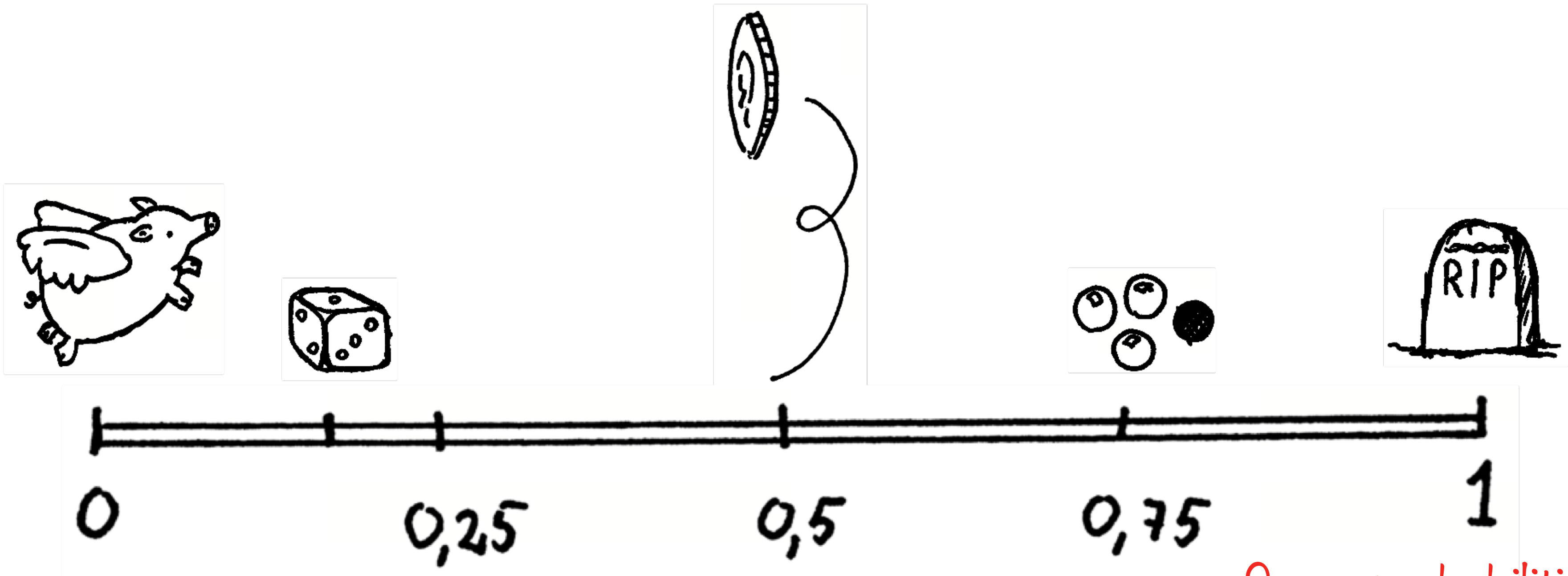
- Introduction to Probability Theory
- Definitions, concepts and notation
- Relative Frequency Approach
- Set theory
- Basic Axions on probabilities

Example - X linked recessive disease

→ Conditional probabilities
→ Don't (only) rely on logic
→ Systematic calculations



Probability Line



Some probabilities you know

- All probabilities are numbers between 0 and 1.
- In percentage, between 0% to 100%.
- We begin with one sample point.

Words to Know

- Experiment/trial (*Forsøg/test*) — Roll a dice
- Sample space (*Udfaldsrum*) — $S=\{1,2,3,4,5,6\}$
- Sample point (*Bestemt udfald*) — $a=\{4\}$
- Event (*Hændelse*) — $A=\{2,4,6\}$ (even number)
  - Elementary event — Event that has one possible outcome
  - Joined event — Event that has many possible outcomes
  - Simultaneous event — Event with two or more sub trials

Relative Frequency Approach

- The number of times event A occurs: N_A

- The number of times that all events occur (sample space):

$$N = N_A + N_B + N_C + \cdots$$

- Then we have the relative frequency:

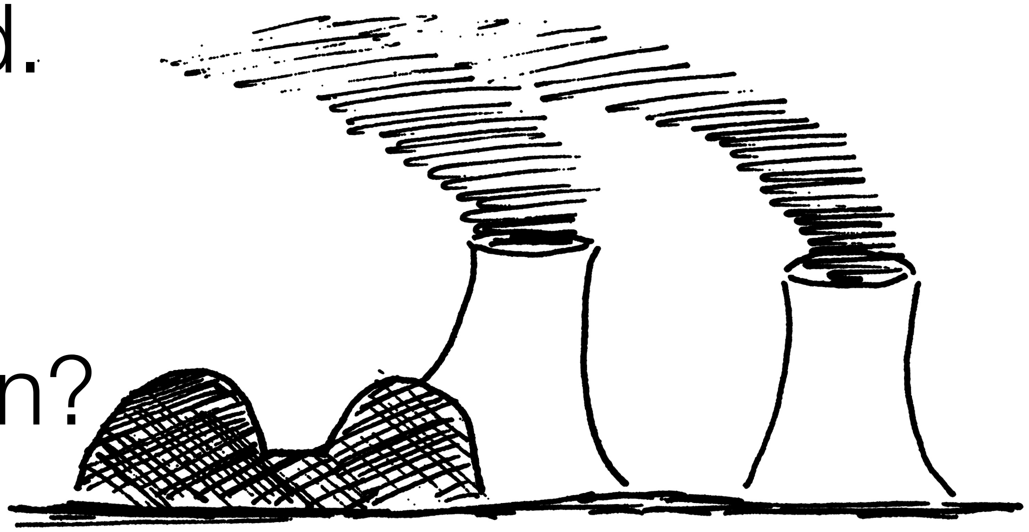
$$\boxed{Pr(A) \sim r(A) = \frac{N_A}{N}}$$

All sample points should have the same a priori probability

- Where: $Pr(A) = \lim_{N\to\infty} r(A)$

Risk of a Meltdown

- There are 437 reactors in the world.
- ~153M operating reactor hours.
- ~Four reactor meltdowns.
- What are the chance of a meltdown?



$$\frac{4}{153M} \text{ pr. reactor pr. hour}$$

$$\sum_{n=1}^{437} \frac{4}{153M} = \frac{1}{87600} \text{ pr. hour}$$

$$\frac{24*365}{87600} = \frac{1}{10} \text{ pr. year}$$

→ Be carefull: Small samples, small probabilities, circumstances, etc.

→ Very uncertain: If number of reactors > 4370 → Pr(Meltdown)>1

Set Theory (*Mængdelære*)

A set:

- A collection of things.
- Elements of sets are not ordered.

Name of set

Some more elements

$$E = \{\alpha_1, \ldots, \alpha_n\}$$

Element

Example:

- The set of all persons in a drug trial group.
- The number of cars i DK.
- All numbers.
- All colours.

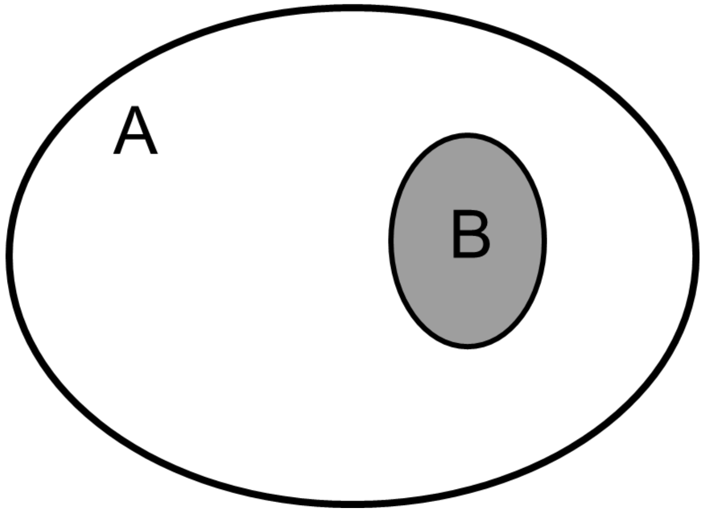
A Subset to a Set *(Delmængder)*

- A subset is any set, where all elements are included in the original set

Notation:

B is a subset to A:

$$B \subset A$$



Example:

For a set $A = \{blue, red, green\}$

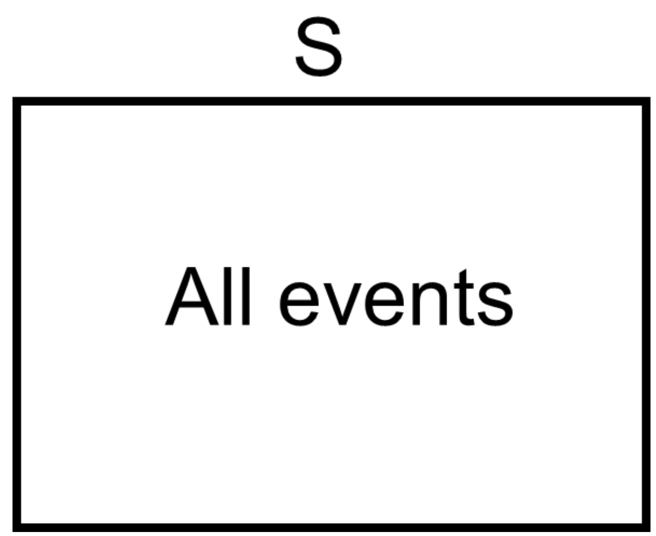
we have a subset $B \subset A$ if B is in A,

e.g. $\{blue, red\}$,$\{blue, red, green\}$,$\{green\}$, $\{\}$

The Sample Space (*Udfaldsrum*)

- The sample space contains all possible events.
- The probability that a sample is from the sample space is 1.

Pr(S)=1

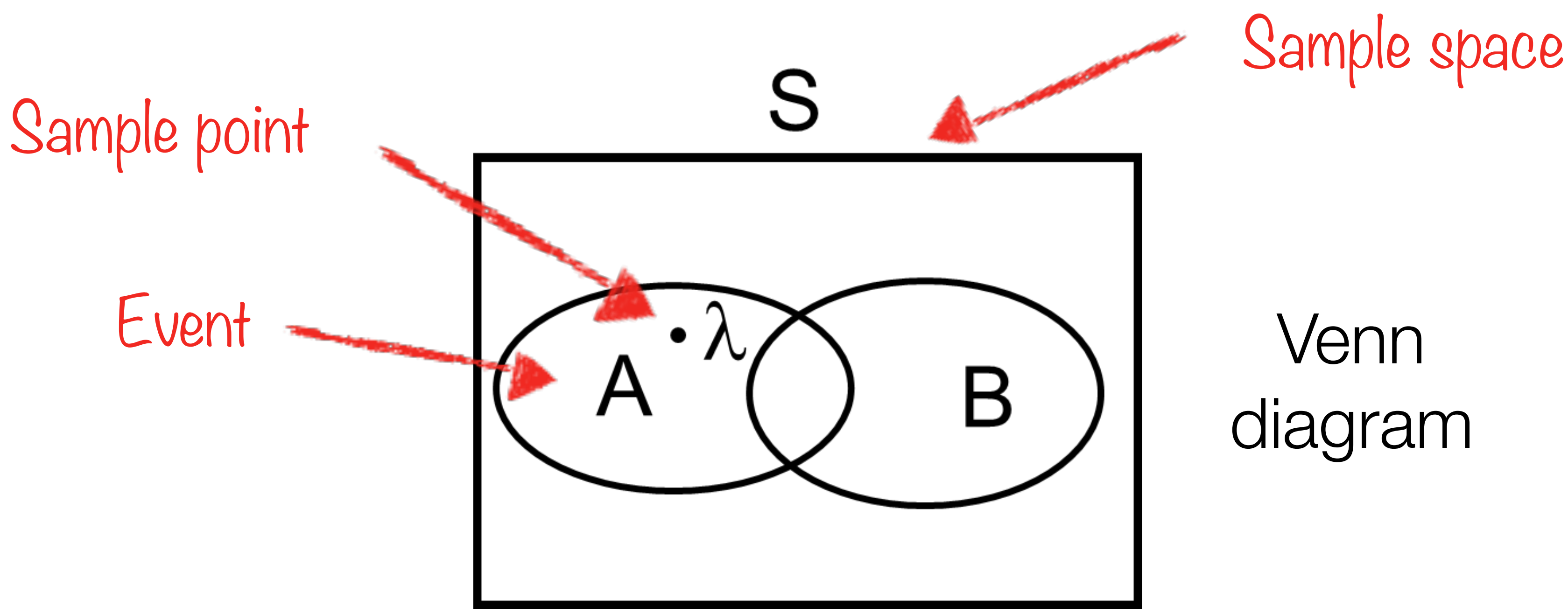


Example:

- A sample space contains 20 people
- 8 has a given disease, 12 is healthy
- Draw a random person.
- What is the chance that a person is a person?
- What is the chance that a person is sick?

A Sample Point (*Udfald*)

- An elementary event.
- Events are collections of sample points.
- Sample space is the collection of all possible sample points.
- Sample points are not ordered.



Venn diagram

Example:

Throw of a dice:

Possible outcomes: 1,2,3,4,5,6 → S={1,2,3,4,5,6}

Events: A={1,2,3} and B={2,4,6}; $A \subset S$; $B \subset S$

Basic Axions of Probability

- The probability of a sample point (element of a sample space).
- The probability of a event (E) (collection of sample points).
- All sample points of a probability space (S) sums up to 1.

➢ Basic Axions of Probability:

Axion 1: $0 \leq Pr(E) \leq 1$

Axion 2: $Pr(S) = 1$

The Empty Set (*Den tomme mængde*)

- The empty set is always a subset of any set.
- This corresponds to the impossible event.

$$\varnothing = \{\}$$

The null set

- The probability of the impossible event is 0.

Example:

- The set of boys in an all girlschool.
- The chance of pigs growing wings and fly.
- To get an 8 when rolling a dice.

Summary

The certain event

S is the certain set.

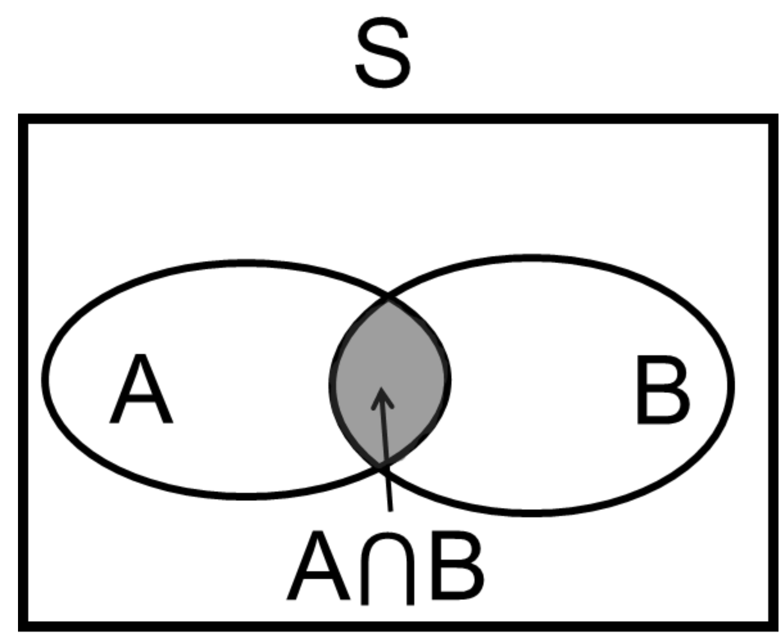
$\varnothing$ is the empty set.

The impossible event

Joint Events (*Fællesmængde*)

- The intersection $A \cap B$ are the common elements of the events A and B
- $A \cap B$ means A <u>and</u> B.

Venn diagram

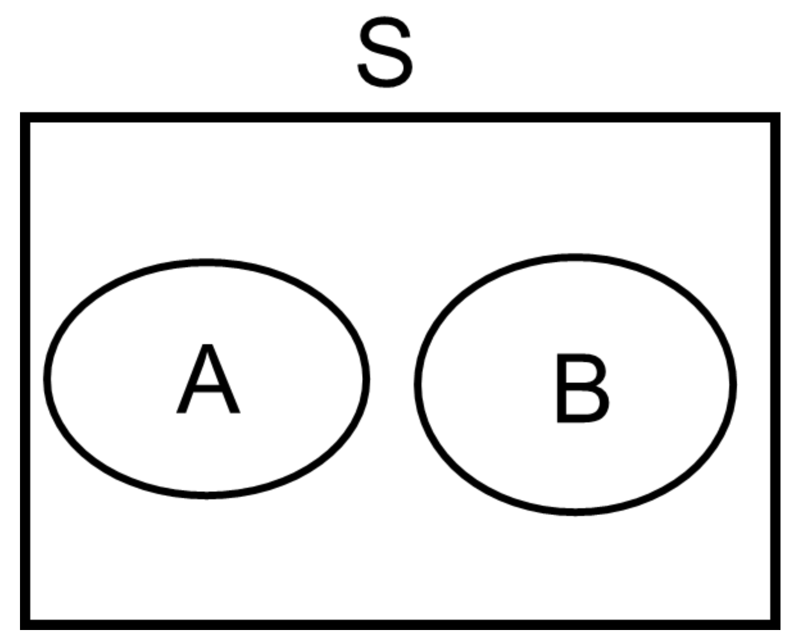


Example:

- A is the event of VW cars i DK
- B is the event of red cars in DK
- The intersection of the events is all red VW in DK.

Mutually Exclusive (Disjoint) Events *(Disjunkte)*

- The sets of A and B are disjoint if: $A \cap B = \emptyset$



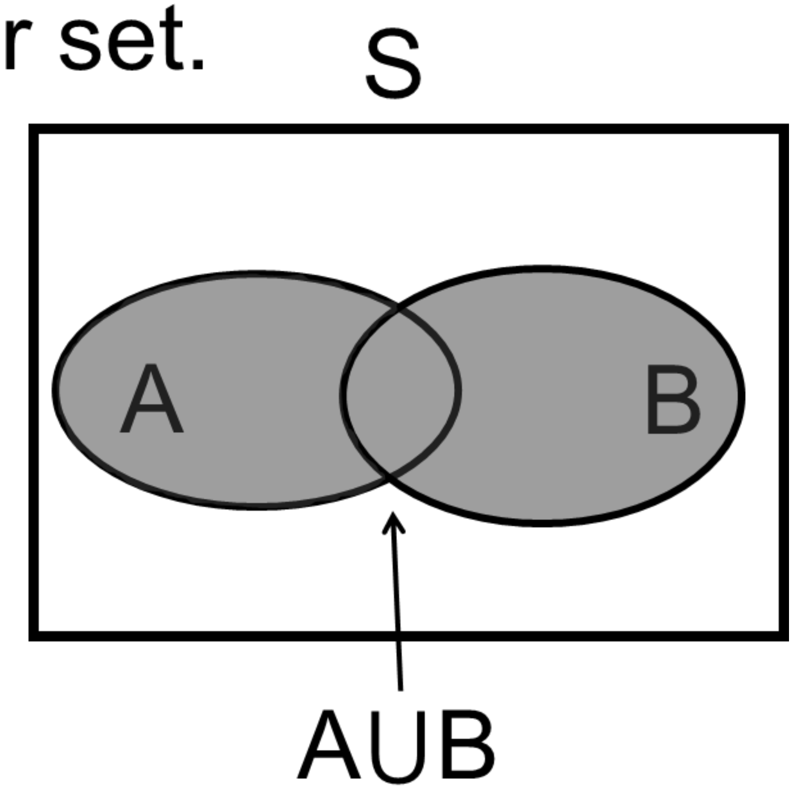
$A \cap B = \varnothing$

Example:

- Event A: The child is a girl.
- Event B: The child is a boy.

Union of Events (*Foreningsmængde*)

- The union of events $A \cup B$ are all the events in one set 'plus' the events in the other set.
- $A \cup B$ means A <u>or</u> B.
- $A \cup B = A + B - A \cap B$

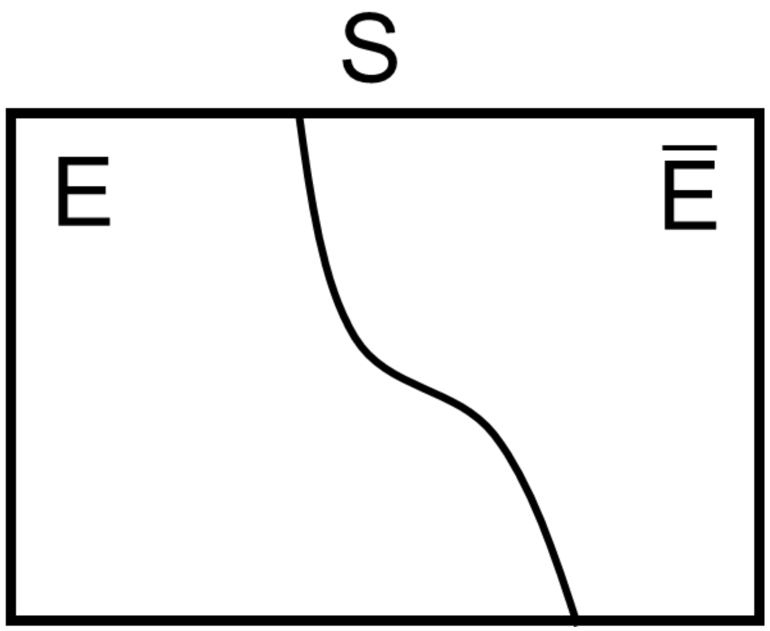


Example:

- I can choose between oatmeal (A) and cornflakes (B) for breakfast.
- The union of the events is that I had breakfast.

The Complement Event (*Komplementær*)

Notation: $S \setminus E = \overline{E} = E^c$ ”not-E”



Notice:

$$E \cup \bar{E} = S$$

The certain event

$$E \cap \bar{E} = \varnothing$$

The impossible event

Example:

- The complement of having a disease is not having a disease

Calculation Rules for Set Theory

Commutative law

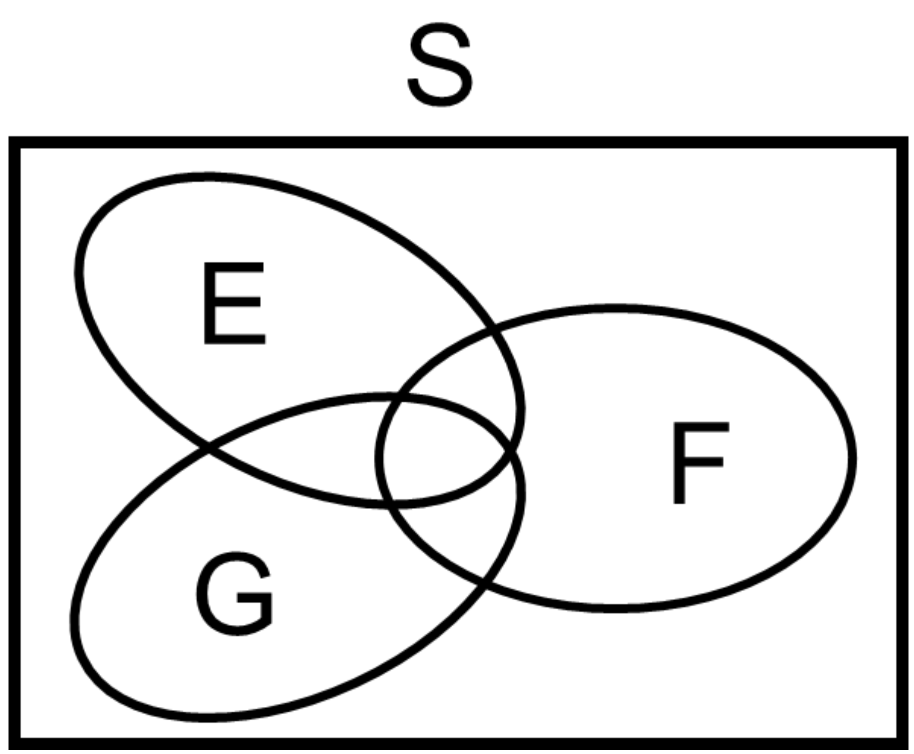
$$E \cup F = F \cup E$$

Associative law

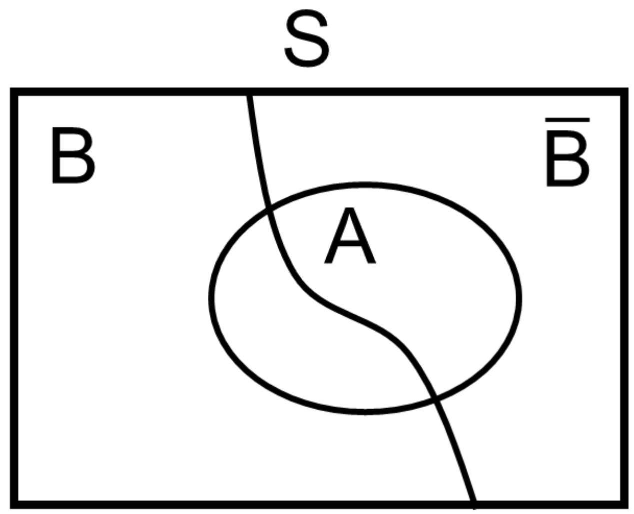
$$E \cup (F \cup G) = (E \cup F) \cup G$$

$$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$$

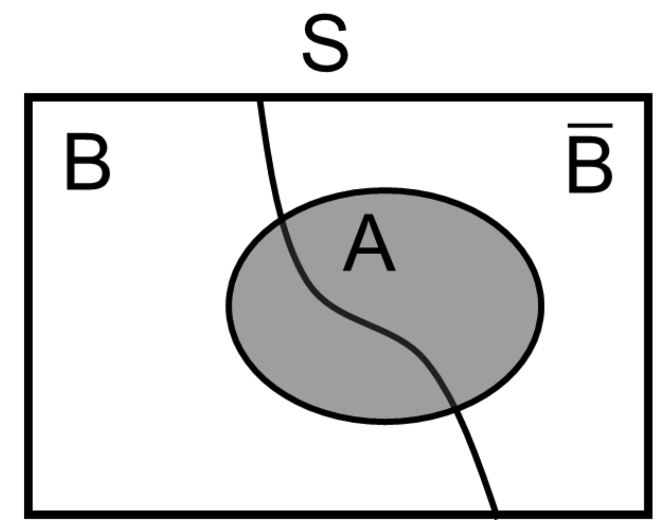
Distributive law



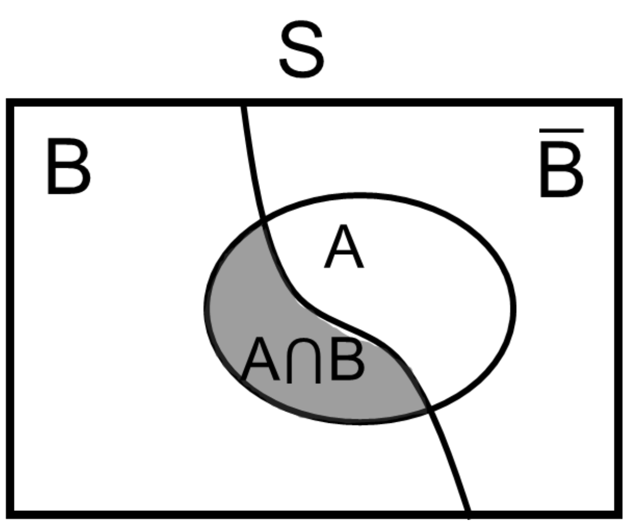
Probability of joint events



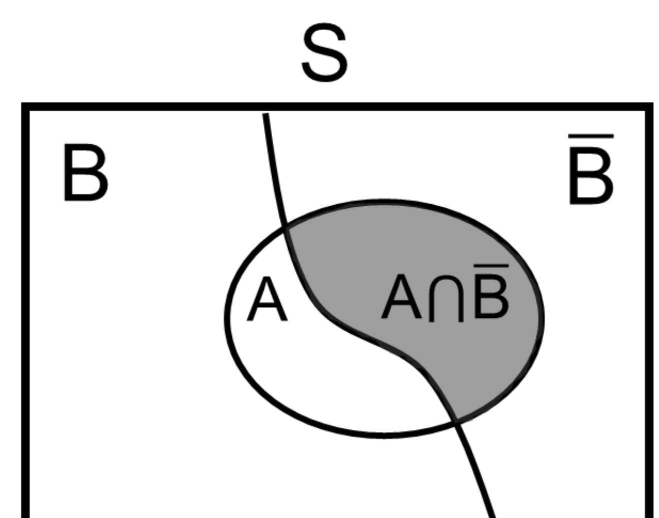
Venn diagram



$$\Pr(A) = \frac{N_A}{N_S}$$



$$\Pr(A \cap B) = \frac{N_{A \cap B}}{N_S}$$



$$\Pr(A \cap \bar{B}) = \frac{N_{A \cap \bar{B}}}{N_S}$$

Independence (*Uafhængighed*)

- We define that two events are independent if and only if:

$$\boxed{Pr(A \cap B) = Pr(A) \cdot Pr(B)}$$

Notice:

- This does not apply if the events A and B are dependent.

Example:

- Two throws with a dice
- The gender of two siblings

Conditional Probability *(Betingede sandsynligheder)*

- We write a conditional probability as:

$$\boxed{Pr(A|B)}$$

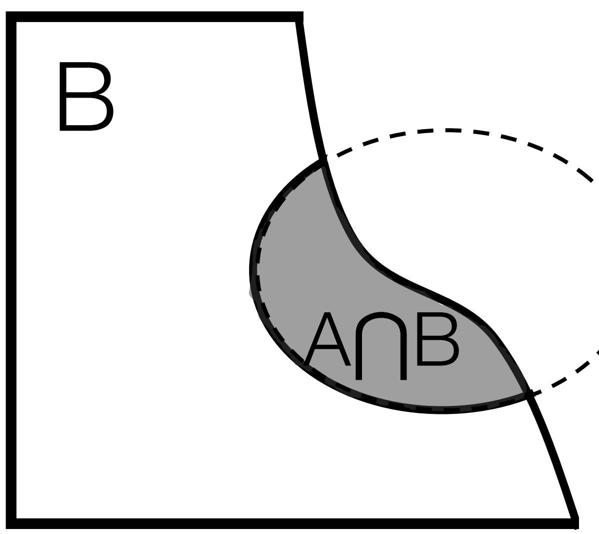
"A given B"

- This means that if the event B has already happened, what is the probability of the event A.
- Reduction of the sample space (possible events) from S to B

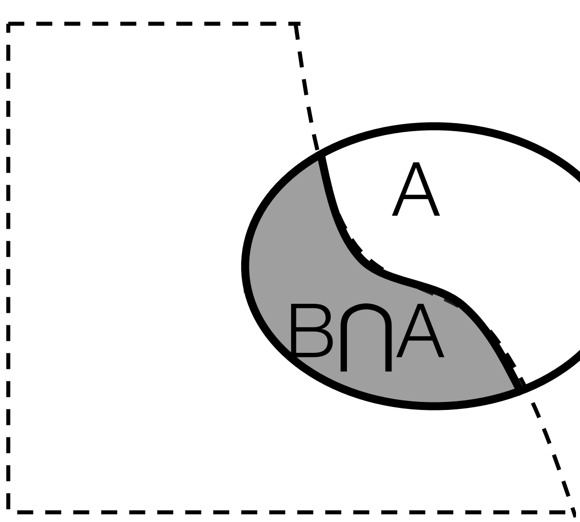
Example:

- From a population, I have selected a female.
- What is the chance that the selected person is below 1.6 m in height?

Conditional Probabilities – Bayes Rule



$$\Pr(A|B) = \frac{N_{A\cap B}}{N_B} = \frac{N_{A\cap B}/N_S}{N_B/N_S} = \frac{\Pr(A \cap B)}{\Pr(B)}$$



$$\Pr(B|A) = \frac{N_{B\cap A}}{N_A} = \frac{N_{B\cap A}/N_S}{N_A/N_S} = \frac{\Pr(B \cap A)}{\Pr(A)}$$

Probabilities of a Joint Event

- We can calculate the probability of a joint event

$$\boxed{Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)}$$

Notice:

- We can extend this rule to multiple events.
- Joint events are not the same as conditional events

- If A and B independent:

$$\boxed{Pr(B|A) = Pr(B) \quad \text{and} \quad Pr(A|B) = Pr(A)}$$

Very important!

Bayes Rule

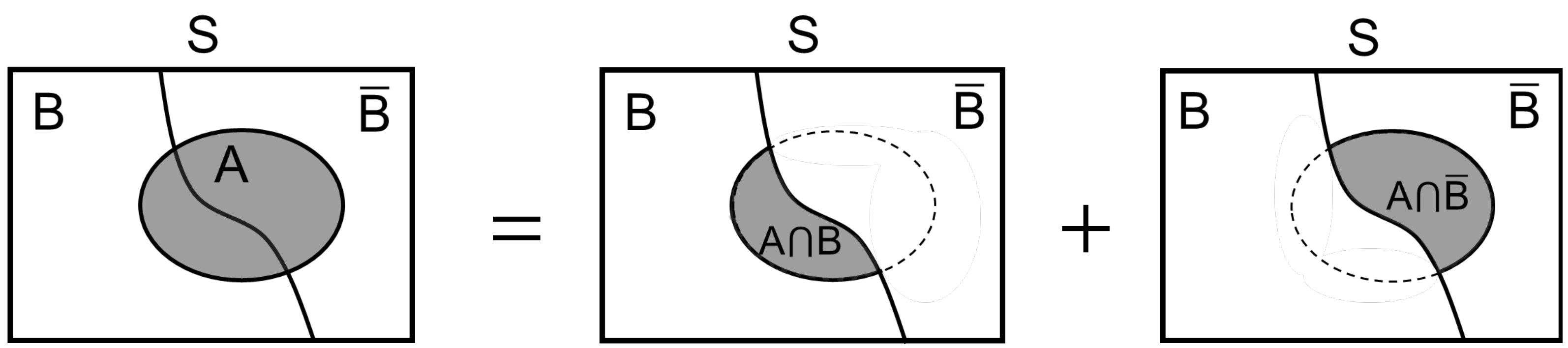
- We can write Bayes rule for two events as:

$$\boxed{Pr(B) \cdot Pr(A|B) = Pr(A) \cdot Pr(B|A)}$$

or

$$\boxed{Pr(A|B) = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)} = \frac{Pr(A \cap B)}{Pr(B)}}$$

Conditional Probabilities – Total Probability



$$\Pr(A) = \frac{N_A}{N_S} = \frac{N_{A\cap B}}{N_S} + \frac{N_{A\cap \bar{B}}}{N_S} = \frac{N_{A\cap B}}{N_B}\cdot\frac{N_B}{N_S} + \frac{N_{A\cap \bar{B}}}{N_{\bar{B}}}\cdot\frac{N_{\bar{B}}}{N_S}$$

$$\boxed{\begin{aligned}\Pr(A) &= \Pr(A\cap B) + \Pr(A\cap \bar{B}) \\ &= \Pr(A|B)\cdot\Pr(B) + \Pr(A|\bar{B})\cdot\Pr(\bar{B})\end{aligned}}$$

Conditional Probabilities - Example

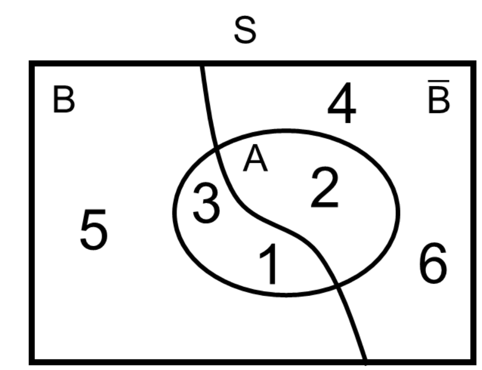
Rolling a dice:

Sample space: $S=\{1,2,3,4,5,6\}$

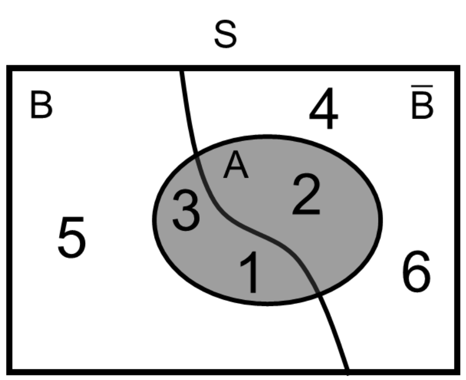
Events: $A=\{1,2,3\}$

$B=\{1,3,5\}$

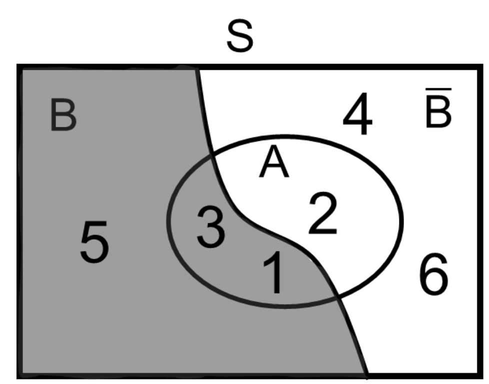
$\bar{B}=\{2,4,6\}$



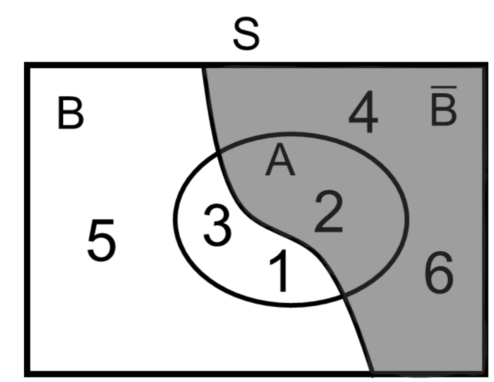
Venn diagram



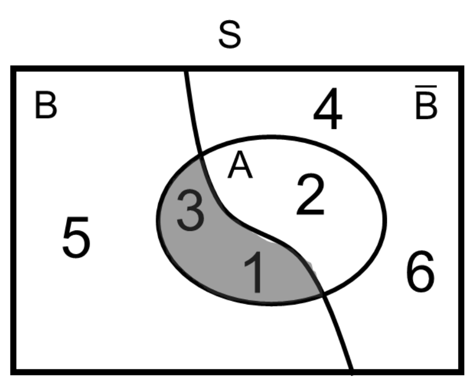
$$\Pr(A)=\frac{N_A}{N_S}=\frac{3}{6}=\frac{1}{2}$$



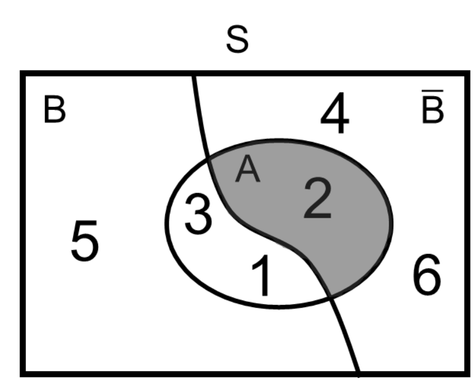
$$\Pr(B)=\frac{N_B}{N_S}=\frac{3}{6}=\frac{1}{2}$$



$$\Pr(\bar{B})=\frac{N_{\bar{B}}}{N_S}=\frac{3}{6}=\frac{1}{2}$$

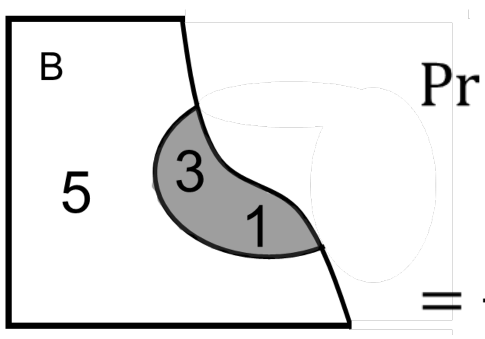


$$\Pr(A\cap B)=\frac{N_{A\cap B}}{N_S}=\frac{2}{6}=\frac{1}{3}$$



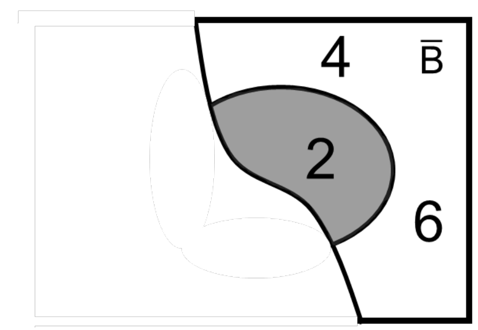
$$\Pr(A\cap\bar{B})=\frac{N_{A\cap\bar{B}}}{N_S}=\frac{1}{6}$$

Conditional Probabilities - Example



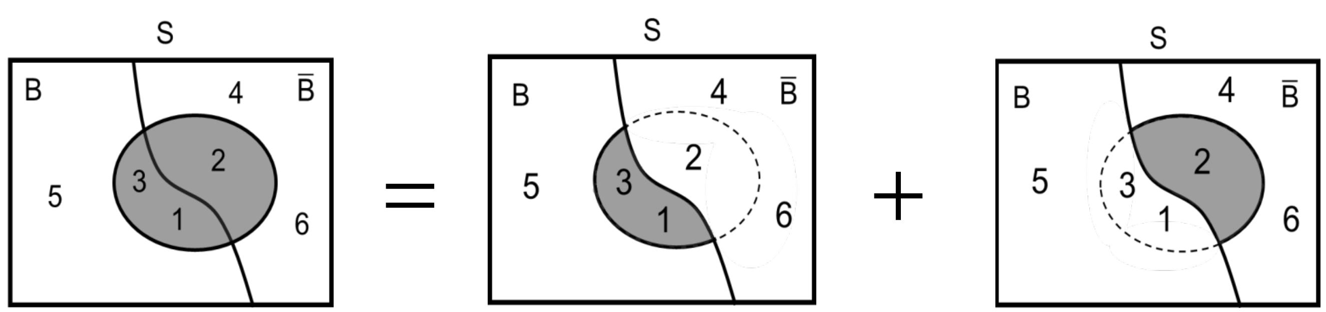
$$\Pr(A|B) = \frac{N_{A\cap B}}{N_B} = \frac{2}{3}$$

$$= \frac{\Pr(A\cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}$$



$$\Pr(A|\bar{B}) = \frac{N_{A\cap\bar{B}}}{N_{\bar{B}}} = \frac{1}{3}$$

$$= \frac{\Pr(A\cap\bar{B})}{\Pr(\bar{B})} = \frac{1/6}{1/2} = \frac{2}{6} = \frac{1}{3}$$



$$\Pr(A) = \frac{N_A}{N_S} = \frac{N_{A\cap B}}{N_S} + \frac{N_{A\cap\bar{B}}}{N_S} = \frac{N_{A\cap B}}{N_B}\cdot\frac{N_B}{N_S} + \frac{N_{A\cap\bar{B}}}{N_{\bar{B}}}\cdot\frac{N_{\bar{B}}}{N_S}$$

$$= \frac{3}{6} = \frac{2}{6} + \frac{1}{6} = \frac{2}{3}\cdot\frac{3}{6} + \frac{2}{3}\cdot\frac{3}{6} = \frac{1}{2}$$

$$\Pr(A) = \Pr(A\cap B) + \Pr(A\cap\bar{B}) = \Pr(A|B)\cdot\Pr(B) + \Pr(A|\bar{B})\cdot\Pr(\bar{B}) = \frac{2}{3}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1}{2} = \frac{3}{6} = \frac{1}{2}$$

Markov Properties

- Bayes rule for tree events:

$$\boxed{Pr(A \cap B \cap C) = Pr(A) \cdot Pr(B|A) \cdot Pr(C|B, A)}$$

- For a Markov chain, it holds that:

$$\boxed{Pr(A \cap B \cap C) = Pr(A) \cdot Pr(B|A) \cdot Pr(C|B)}$$

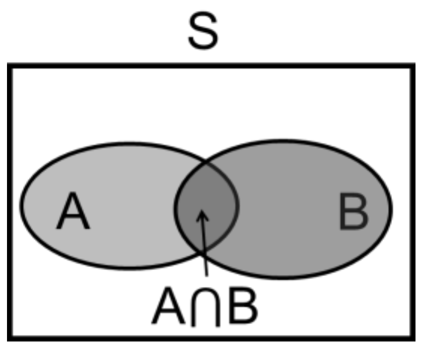
i.e. $$\boxed{Pr(C|B, A) = Pr(C|A \cap B) = Pr(C|B)}$$

(*A* don't give new information)

Probabilities of a Union of Event

- We can calculate the probability of a union of events:

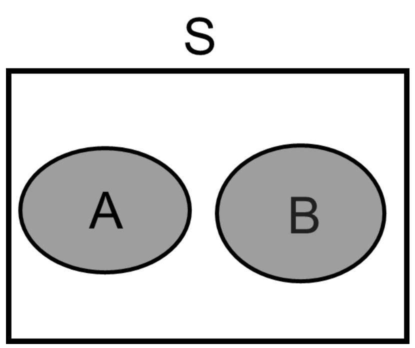
$$\boxed{Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)}$$



Notice:

- If the events are mutually exclusive

$$\boxed{Pr(A \cup B) = Pr(A) + Pr(B)}$$



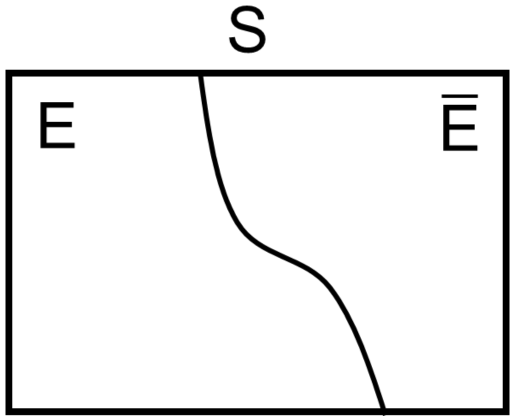
Probabilities of Complement Events

- We can write some rules for the probabilities of a complement event

$$Pr(E \cup \bar{E}) = Pr(S) = 1$$

$$Pr(E) + Pr(\bar{E}) = Pr(S) = 1$$

$$\boxed{Pr(E) = 1 - Pr(\bar{E})}$$



Example:

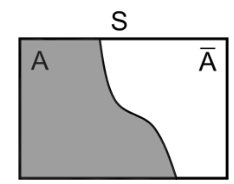
- The probability of not hitting 2 eyes on dice.

$$Pr(\{1,3,4,5,6\}) = 1 - Pr(\{2\}) = 1 - \tfrac{1}{6} = \tfrac{5}{6}$$

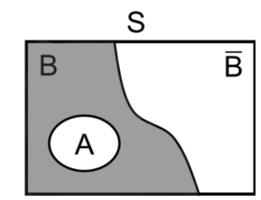
Summary of Probability

Relative frequency: $Pr(A) = \frac{N_A}{N_S}$

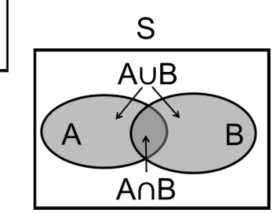
Complement: $Pr(\overline{A}) = 1 - Pr(A)$



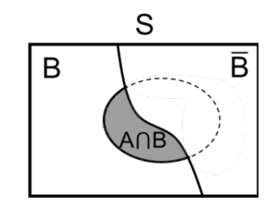
Exclusive: $Pr(\overline{A} \cap B) = Pr(B) - Pr(A)$ if $A \subset B$



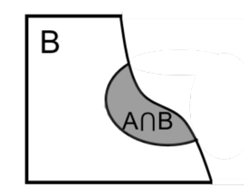
Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



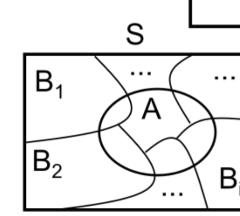
Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



Conditional: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$ if $Pr(B) \neq 0$



Total probability: $Pr(A) = \sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)$



Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

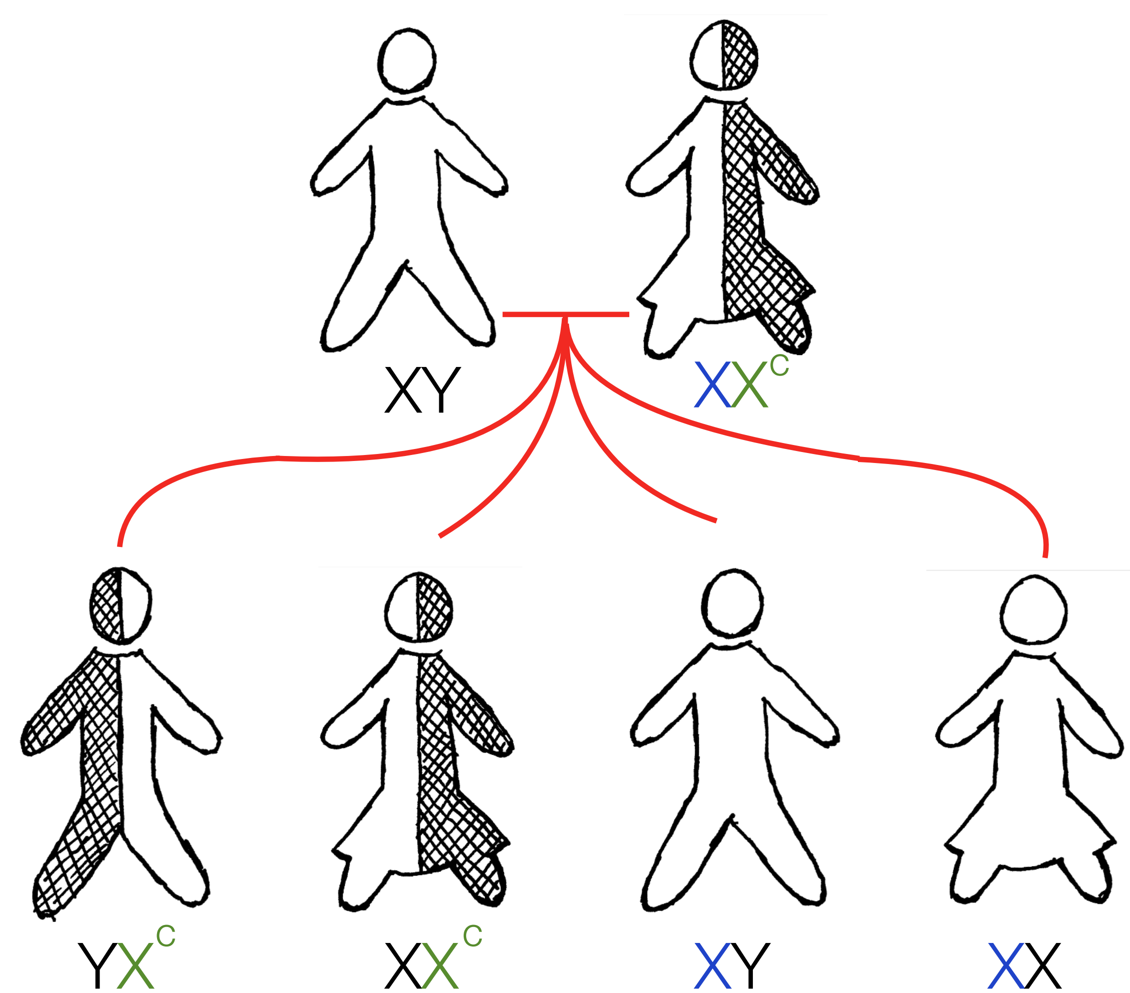
Bayes formula: $Pr(B_i|A) = \frac{Pr(A|B_i) \cdot Pr(B_i)}{\sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)}$

Independence: $Pr(A \cap B) = Pr(A) \cdot Pr(B)$

Neutralized by a healthy X-gene

Example - X linked recessive

- A mother has a sick X gene.
- The chance of giving the sick X gene to a child is 50%.
- A boy with the sick X gene has the disease.
- A girl with the sick X gene is a carrier of the disease.
- Of cause the chance of not giving the sick X gene to a child is also 50%.



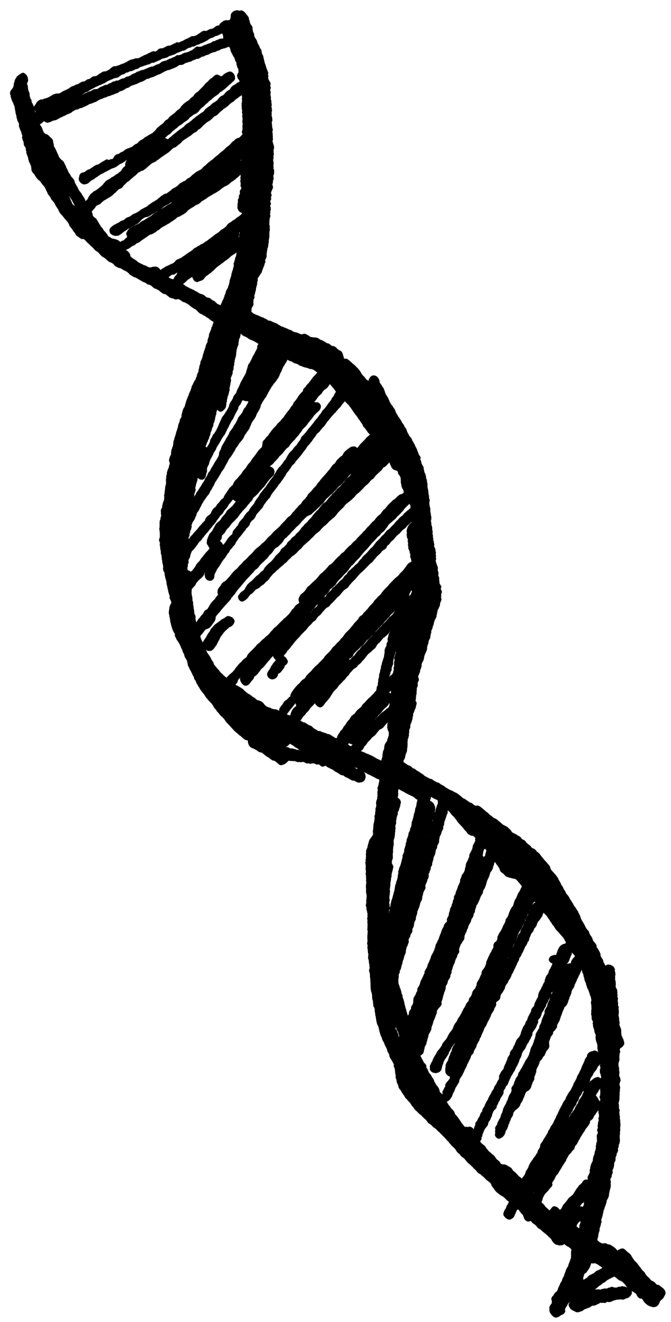
Hunter Syndrome (MPS II)

- X linked recessive
- 1:130.000 male births

- What are the probability that a boy have Hunter?

- Event A: The boy has Hunter.

$$Pr(A) = \frac{1}{130.000} = 7{,}69 \cdot 10^{-6}$$



Genetic Risk Assessment Hunters Syndrome (MPS II)

Information:

- X linked recessive.
- Boy T has Hunter.

Events:

- Event B: Mrs. B is a carrier.
- Event A: Boy A has Hunter.

Find:

- What is Pr(A)?

$Pr(B)=\frac{1}{2}$; $\quad Pr(A|B)=\frac{1}{2}$

$\Downarrow$

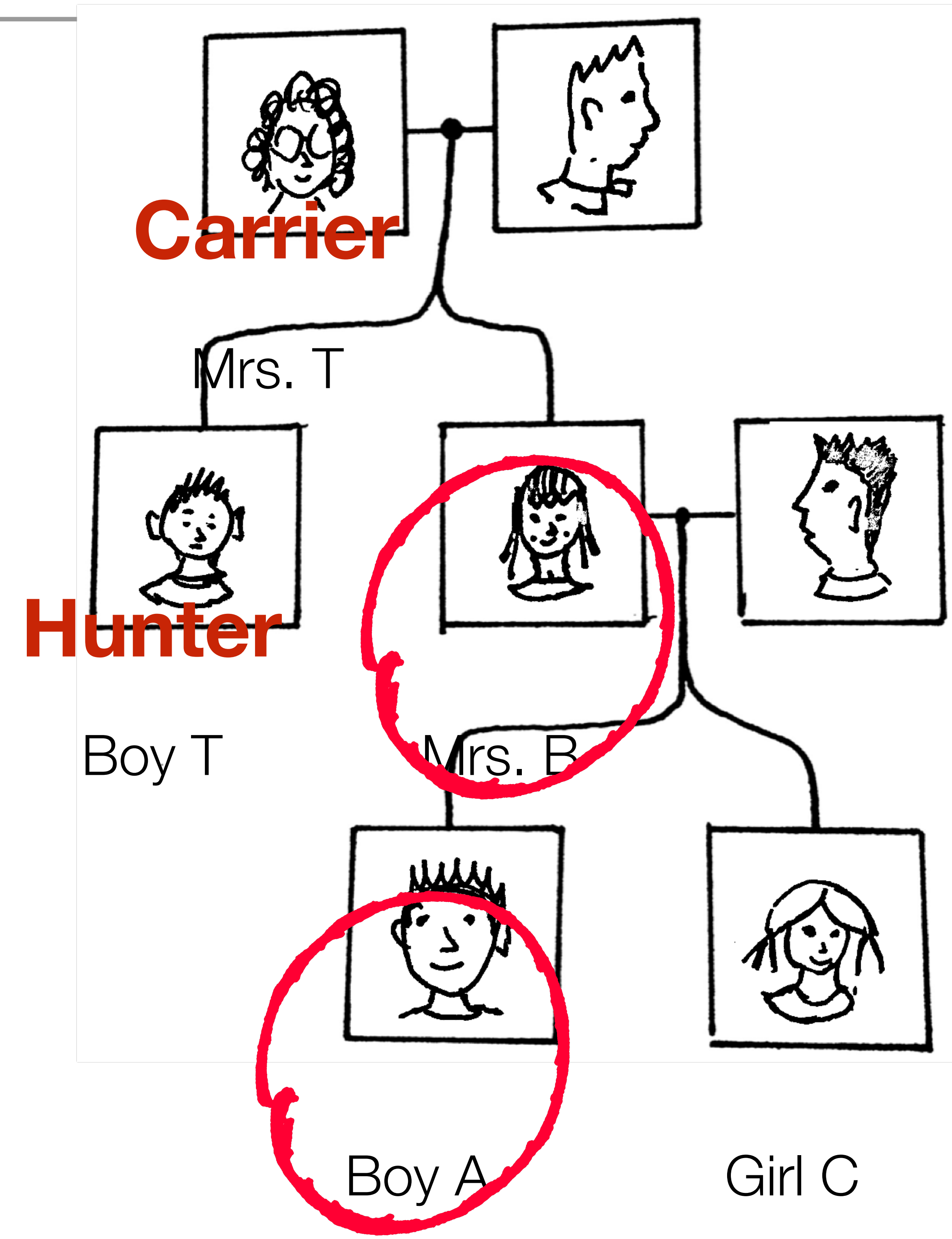
Bayes: $Pr(A\cap B)= Pr(A|B)\cdot Pr(B)=\frac{1}{4}$

$Pr(\bar{B})=1-Pr(B)=\frac{1}{2}$; $\quad Pr(A|\bar{B})=0$

$\Downarrow$

Bayes: $Pr(A\cap \bar{B})= Pr(A|\bar{B})\cdot Pr(\bar{B})=0$

$Pr(A)=Pr(A\cap B)+Pr(A\cap \bar{B})=\frac{1}{4}+0=\frac{1}{4}$



Genetic Risk Assessment Hunters Syndrome (MPS II)

Information:

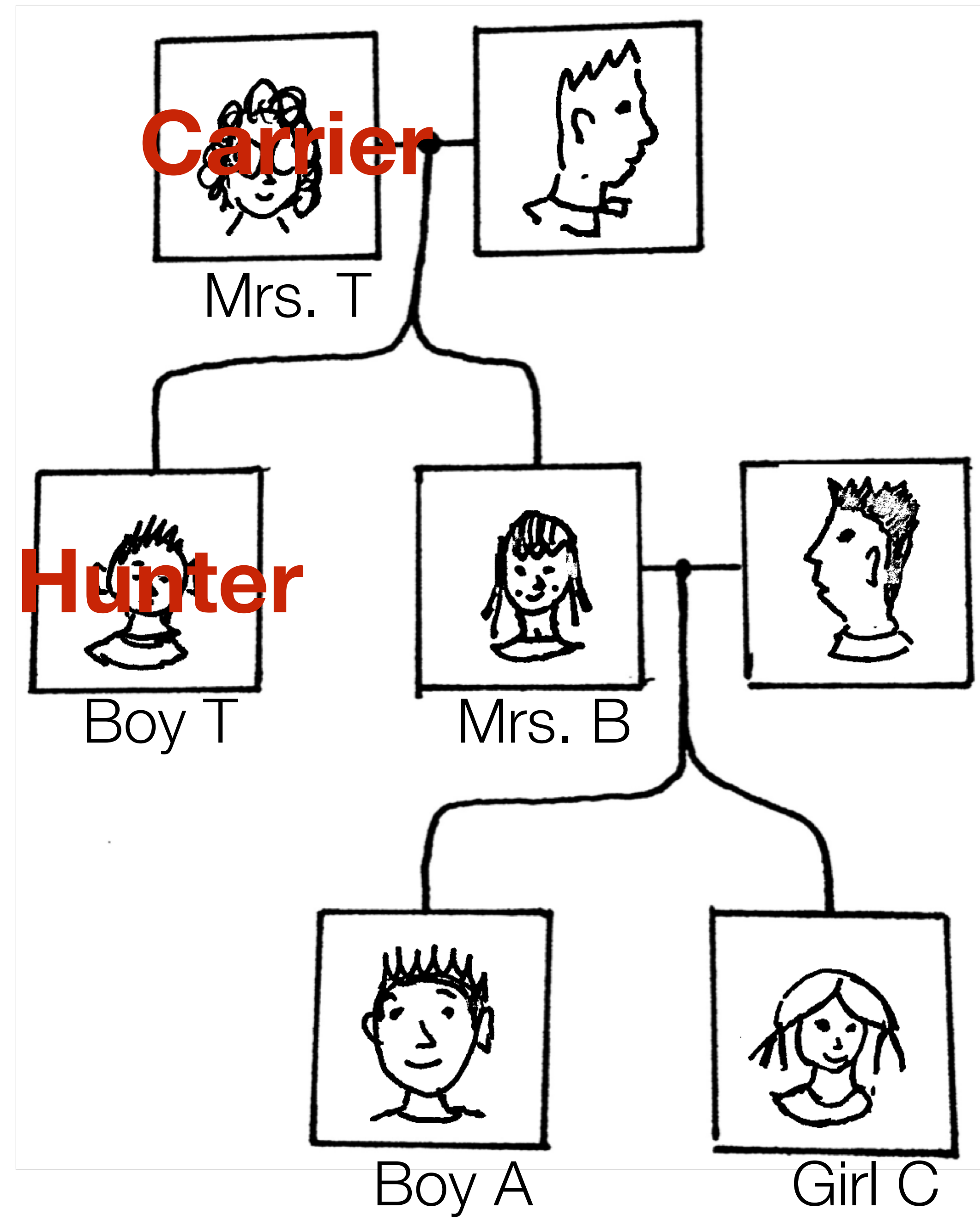
- X linked recessive.
- Boy T has Hunter.

Events:

- Event A: Boy A has Hunter.
- Event B: Mrs. B is a carrier.
- Event C: Girl C is a carrier.

Find:

- What is $Pr(C|\bar{A})$?



Genetic Risk Assessment Hunters Syndrome (MPS II)

Events:

- Event A: Boy A has Hunter.
- Event B: Mrs. B is a carrier.
- Event C: Girl C is a carrier.

$Pr(B)=½;$ $Pr(A)=¼;$ $Pr(\bar{A})=1-Pr(A)=¾;$

$Pr(A|B)=Pr(\bar{A}|B)=½;$ $Pr(A|\bar{B})=0;$ $Pr(\bar{A}|\bar{B})=1;$

$Pr(C|B)=Pr(\bar{C}|B)=½;$ $Pr(C|\bar{B})=0;$ $Pr(\bar{C}|\bar{B})=1;$

$\Downarrow$

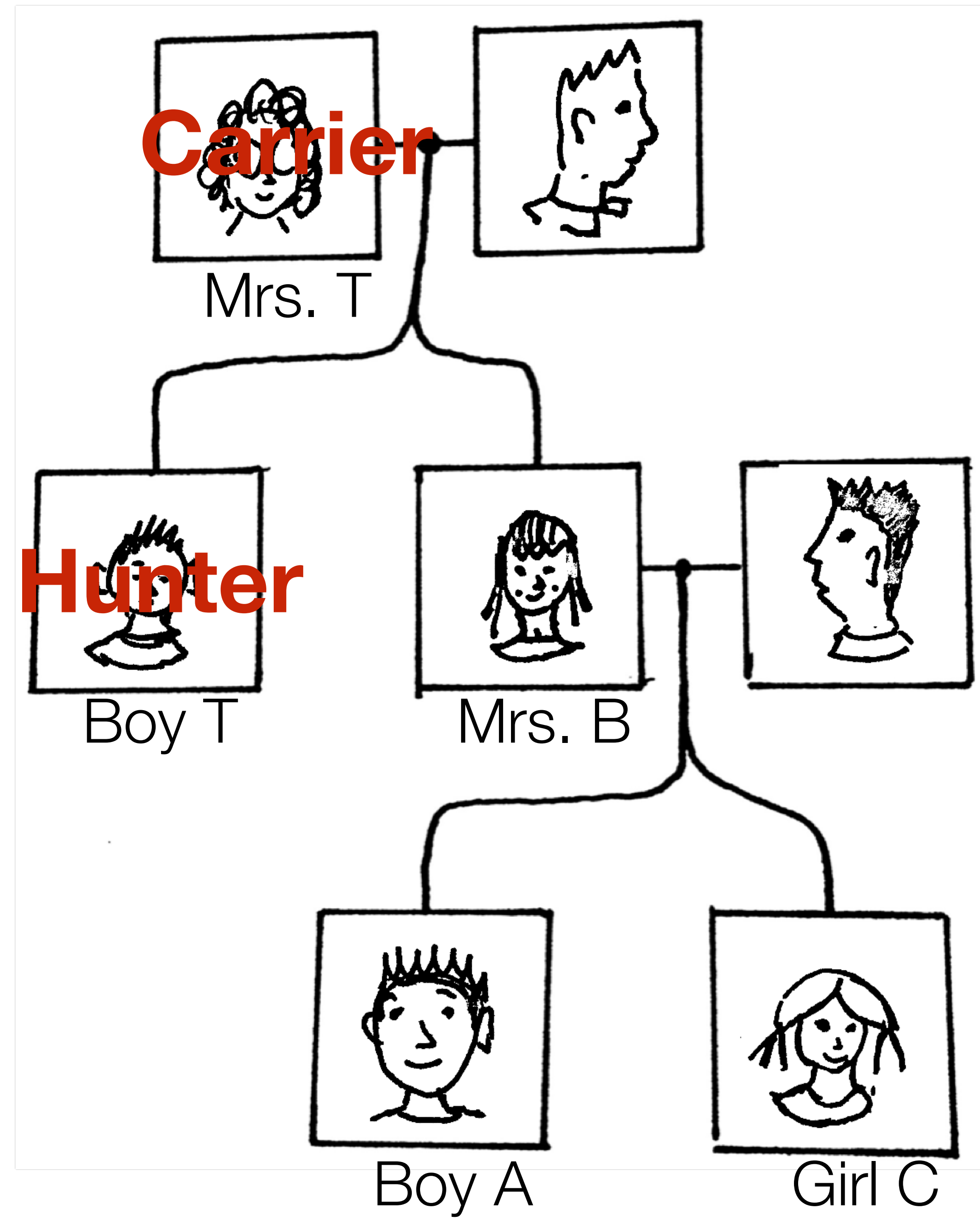
$$\text{Bayes: } Pr(B|\bar{A}) = \frac{Pr(\bar{A}|B)\cdot Pr(B)}{Pr(\bar{A})} = \frac{½\cdot½}{¾} = ⅓$$

$$\text{Markov: } Pr(C\cap B|\bar{A}) = Pr(C|B)\cdot Pr(B|\bar{A}) = ½\cdot⅓ = ⅙$$

$$Pr(C\cap\bar{B}|\bar{A}) = Pr(C|\bar{B})\cdot Pr(\bar{B}|\bar{A}) = 0$$

$\Downarrow$

$$Pr(C|\bar{A}) = Pr(C\cap B|\bar{A}) + Pr(C\cap\bar{B}|\bar{A}) = ⅙+0 = ⅙$$



Words and Concepts to Know

Experiment/Trial

Intersection

Markov chain

Sample space

Mutually Exclusive/Disjoint

Event

Sample point

Union

Complement/not

Relative frequency

Independence

Set

Subset

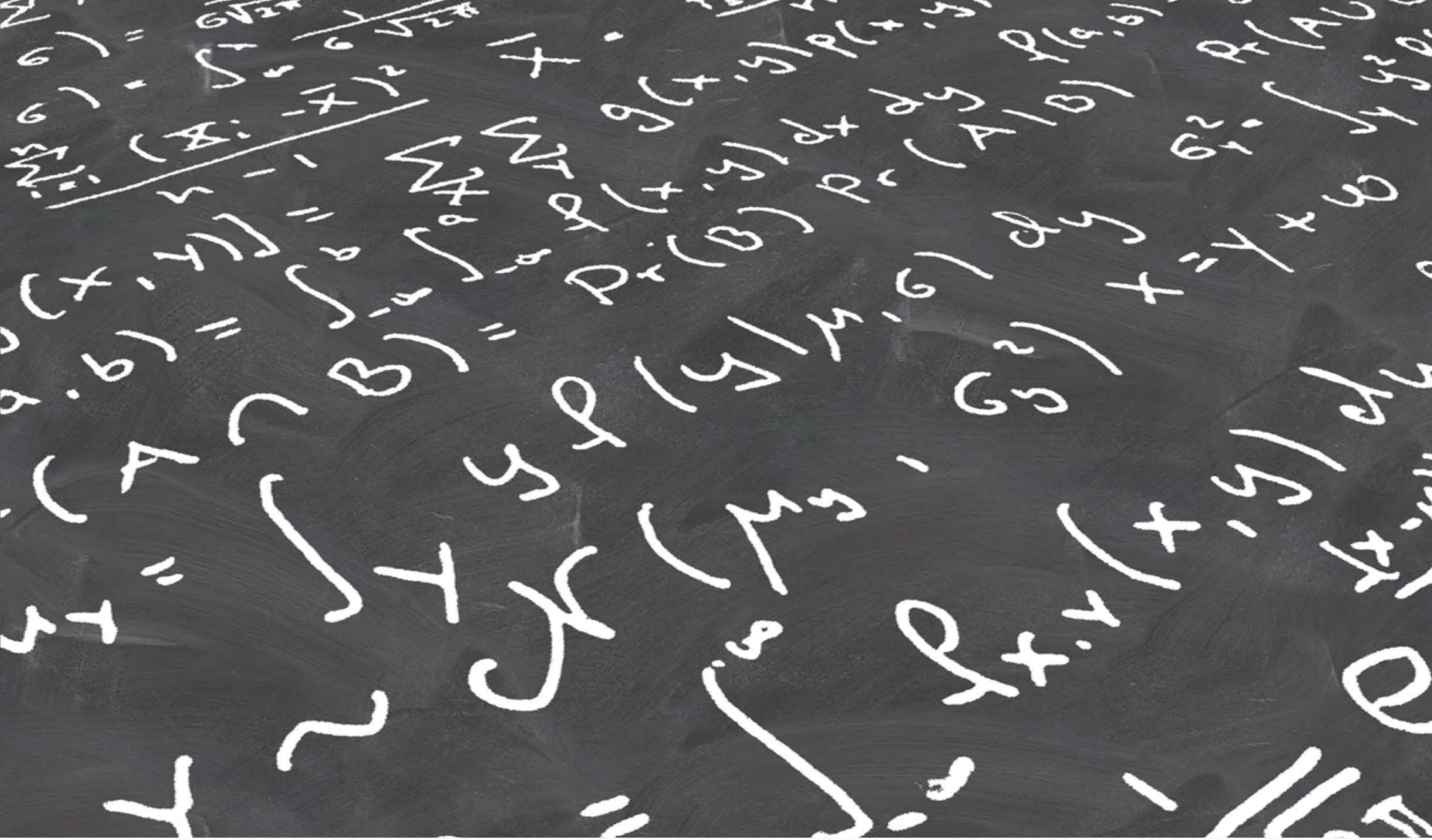
Bayes Rule

Conditional probability

Empty set/Null set

Total probability

Joint events



2. Probability Theory and Combinatorics

Gunvor Elisabeth Kirkelund
Lars Mandrup

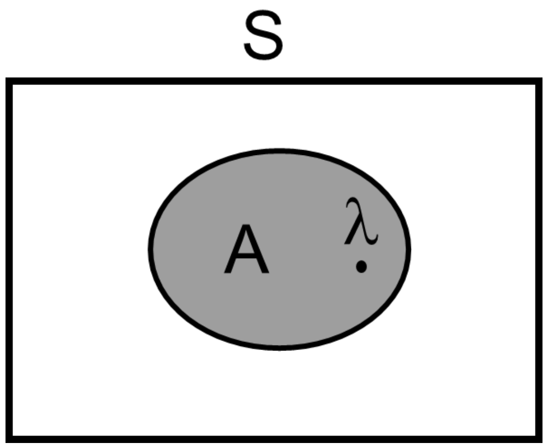
Agenda for Today

- Repetition from last time
- Bayesian probability calculations and total probability
- Bernoulli trials
- Combinatorics
- An experiment

Basic Probability

- Probability theory tells us what is in the sample given nature
- Basic Axions: **Axion 1:** $\boldsymbol{0 \leq Pr(A) \leq 1}$
  Axion 2: $\boldsymbol{Pr(S) = 1}$

S: Sample space
A: Event
λ: Sample point

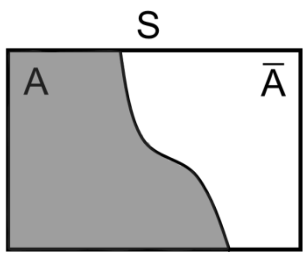


- Often (but not always) we use the relative frequency:

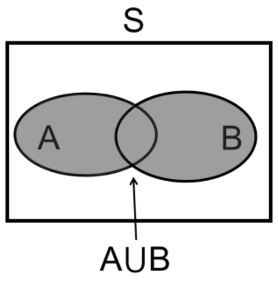
$$\Pr(A) = \frac{N_A}{N}$$

Basic Probability

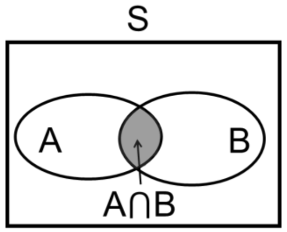
- Complement: $Pr(A) = 1 - Pr(\bar{A})$



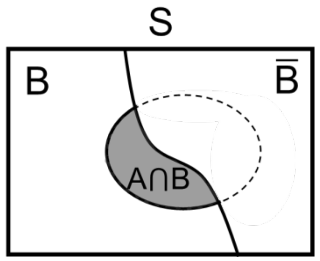
- Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



- Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



- Conditional: $Pr(A|B)$



Bayes Rule and Independence

- Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

- A and B independent:

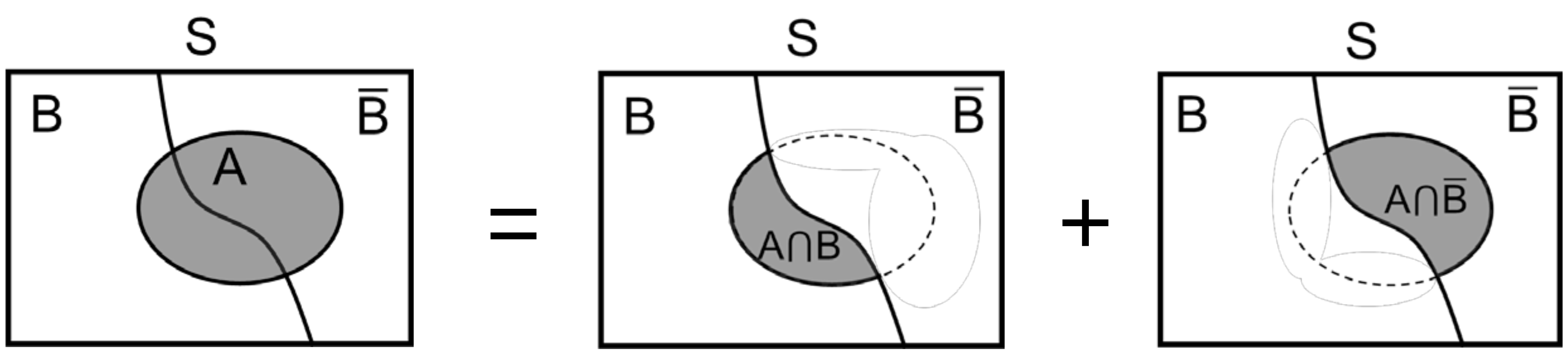
$$\boxed{Pr(A \cap B) = Pr(A) \cdot Pr(B)}$$

$$\boxed{Pr(B|A) = Pr(B) \quad \text{and} \quad Pr(A|B) = Pr(A)}$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.

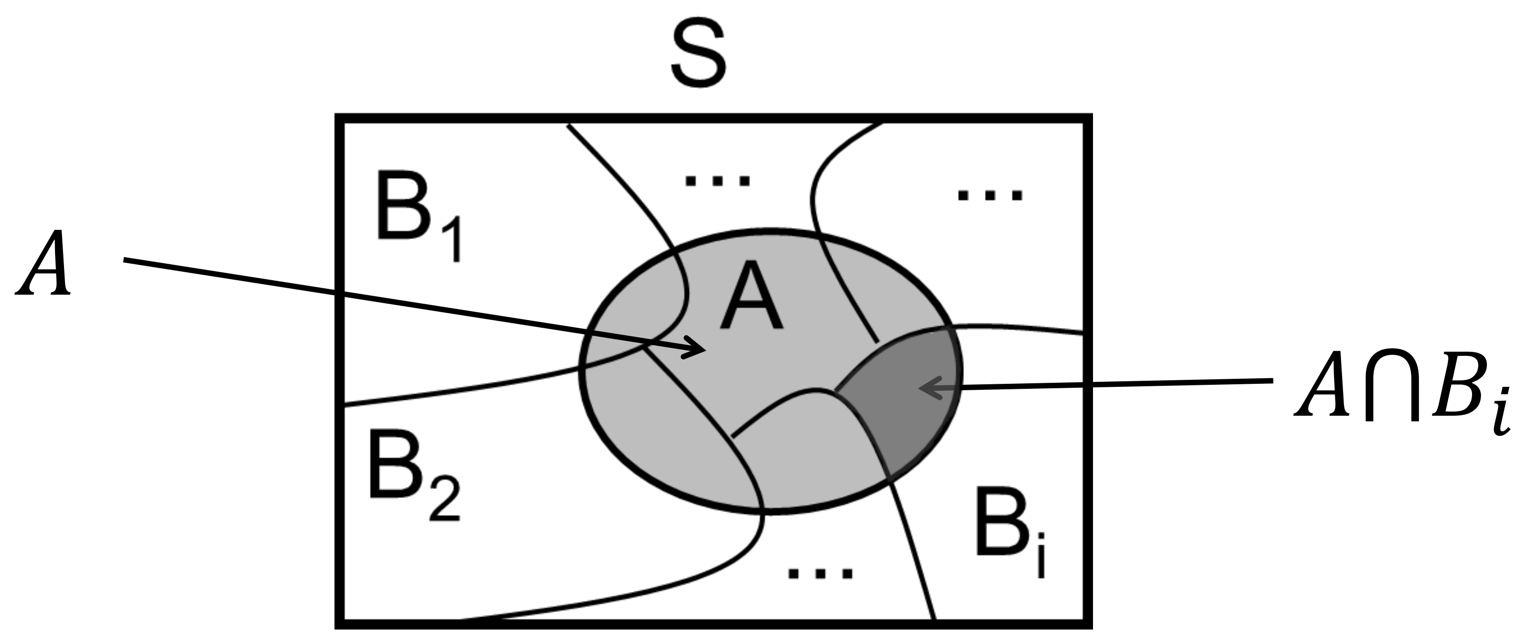


$$\boxed{\begin{aligned}\Pr(A) &= \Pr(A \cap B) + \Pr\left(A \cap \bar{B}\right) \\ &= \Pr(A|B) \cdot \Pr(B) + \Pr\left(A|\bar{B}\right) \cdot \Pr(\bar{B})\end{aligned}}$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.



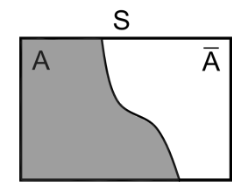
$$Pr(A) = Pr(A \cap B_1) + Pr(A \cap B_2) + \cdots + Pr(A \cap B_i) + \cdots$$
$$= Pr(A|B_1) \cdot Pr(B_1) + Pr(A|B_2) \cdot Pr(B_2) + \cdots$$

where the B_i's are mutually exclusive ($B_i \cap B_j = \varnothing$ for $i \neq j$)
and $S = B_1 \cup B_2 \cup \ldots \cup B_i \cup \ldots$

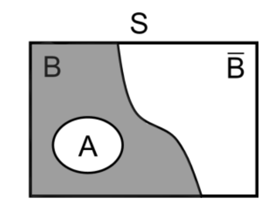
Summary of Probability

Relative frequency: $Pr(A) = \frac{N_A}{N_S}$

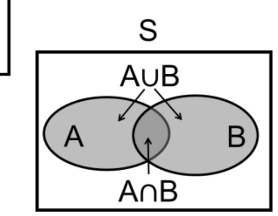
Complement: $Pr(\bar{A}) = 1 - Pr(A)$



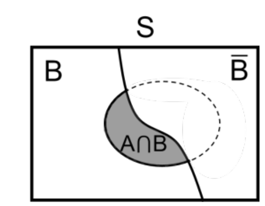
Exclusive: $Pr(\bar{A} \cap B) = Pr(B) - Pr(A)$ if $A \subset B$



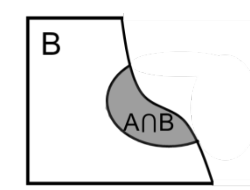
Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



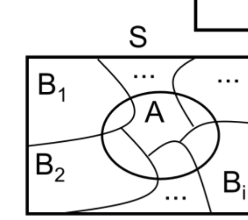
Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



Conditional: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$ if $Pr(B) \neq 0$



Total probability: $Pr(A) = \sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)$



Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

Bayes formula: $Pr(B_i|A) = \frac{Pr(A|B_i) \cdot Pr(B_i)}{\sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)}$

Independence: $Pr(A \cap B) = Pr(A) \cdot Pr(B)$

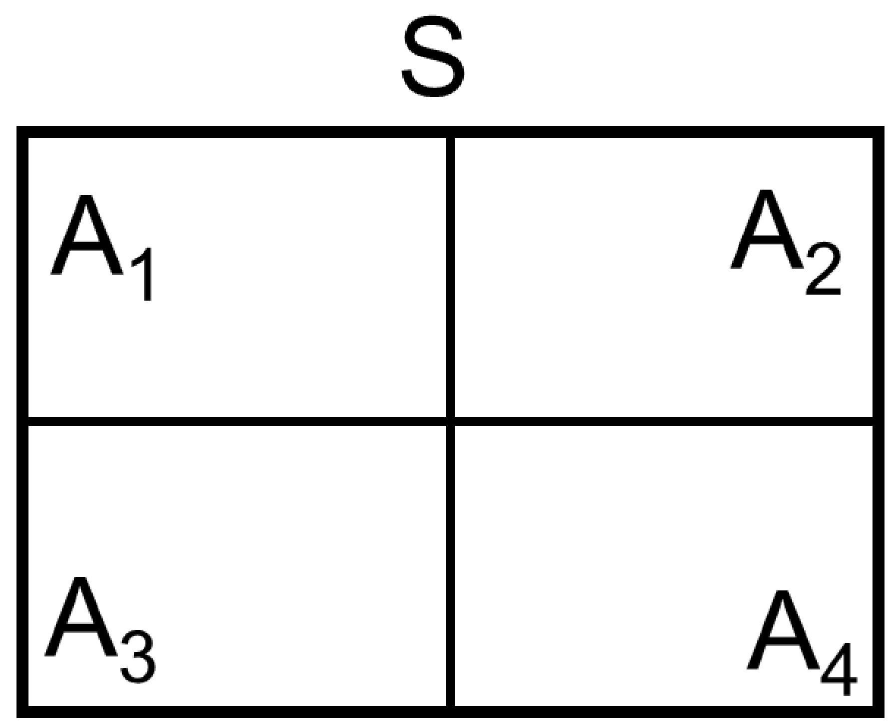
Orca Example

- In a conversation effort, we look for dead orcas when we are visiting an ocean.
- Given (conditioned) that <u>we have selected an ocean to examine</u>, how many males and females orcas will we observe?

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)
Female ($\overline{\mathbf{B}}$)	2	7	11	9
Male ($\mathbf{B}$)	8	3	1	19
Total	10	10	12	28

Orca Example (Cont'd)

- The probability selecting an ocean is identical.



- Event A_1 : Atlantic
- Event A_2 : Antartica
- Event A_3 : Pacific
- Event A_4 : Seaworld

$$Pr(A_1) = Pr(A_2) = Pr(A_3) = Pr(A_4) = \frac{1}{4}$$

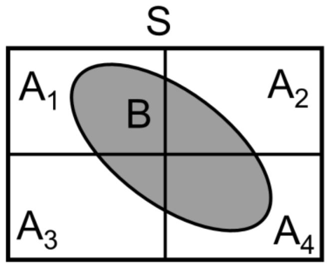
$$Pr(A_1) + Pr(A_2) + Pr(A_3) + Pr(A_4) = 1$$

The events $A_1 - A_4$ are mutually exclusive.

Orca Example Total Probability

- The event B, that the orca is a male, can then be written as:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$$



- The total probability of a found killer whale, being a male, since event $A_1 - A_4$ are mutually exclusive (sum rule):

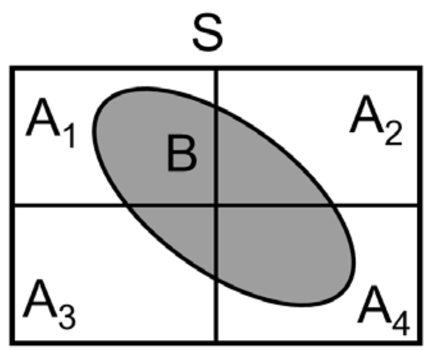
$$\boxed{Pr(B) = Pr(B \cap A_1) + Pr(B \cap A_2) + Pr(B \cap A_3) + Pr(B \cap A_4)}$$

- We rewrite with Bayes rule:

$$Pr(B) = Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2) + Pr(A_3)\,Pr(B|A_3) + Pr(A_4)\,Pr(B|A_4)$$

Orca Example Cont'd

- Total Probability:



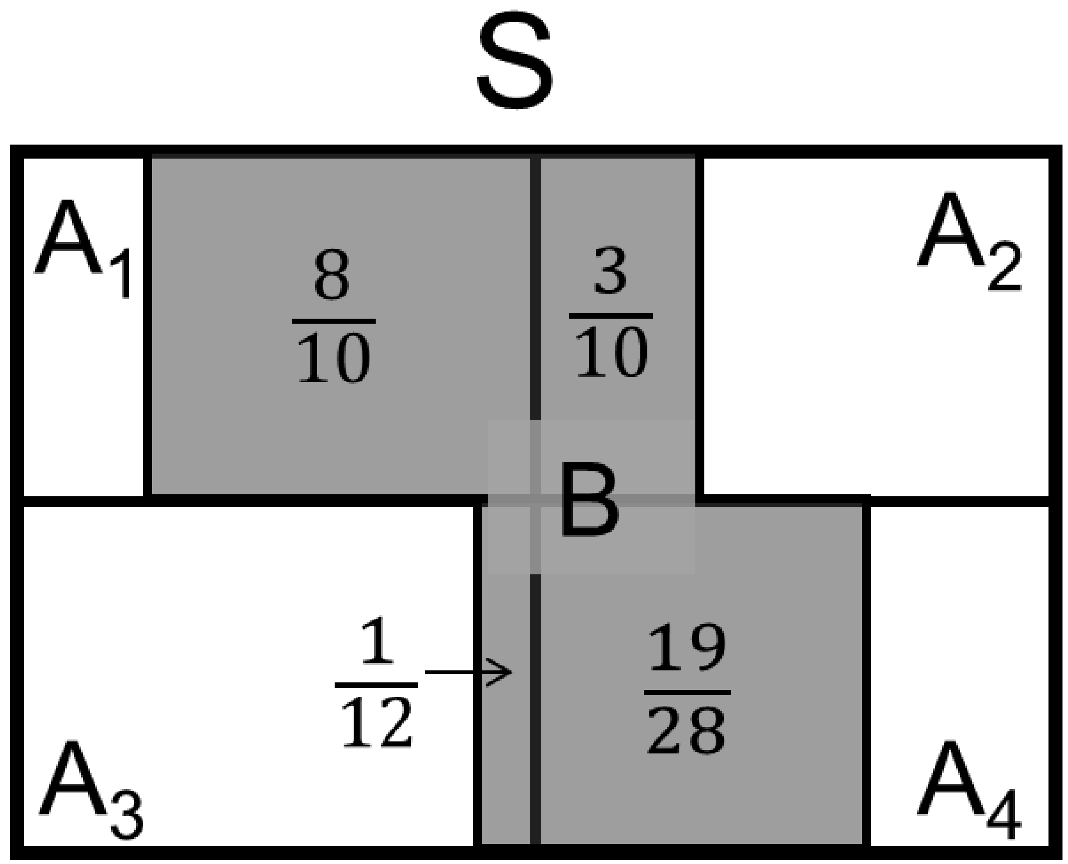
$$Pr(B) = Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2) + Pr(A_3)\,Pr(B|A_3) + Pr(A_4)\,Pr(B|A_4)$$

Gender\ location	**Atlantic (A_1)**	**Antartica (A_2)**	**Pacific (A_3)**	**Seaworld (A_4)**
Female ($\overline{\mathbf{B}}$)	2	7	11	9
Male ($\mathbf{B}$)	8	3	1	19
Total	10	10	12	28

$$Pr(B) = \frac{8}{10}\cdot\frac{1}{4} + \frac{3}{10}\cdot\frac{1}{4} + \frac{1}{12}\cdot\frac{1}{4} + \frac{19}{28}\cdot\frac{1}{4} = 0,465$$

Orca Example Graphical

- We can also use a Graphical approach with Venn diagrams.



- The total probability of B is given by the marked area divided by the area of S.

Orca Example

- If an orca found is a male, what is the probability of us being in the Antartica?

$$Pr(A_2|B)$$

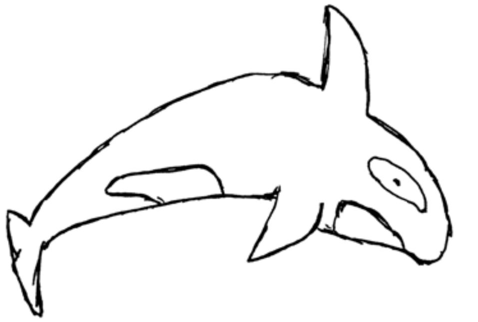
- We use Bayes rule:

$$Pr(A_2|B) = \frac{Pr(A_2 \cap B)}{Pr(B)} = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)}$$

- $Pr(B) = 0{,}47;\;\; Pr(A_2) = 0{,}25;\;\; Pr(B|A_2) = 0{,}3$

$$Pr(A_2|B) = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)} = \frac{0{,}3 \cdot 0{,}25}{0{,}47} = 0{,}16$$

Orca Example



- Is locations of the found orca independent of gender?
- How would you test it?

Gender\ location	**Atlantic (A_1)**	**Antartica (A_2)**	**Pacific (A_3)**	**Seaworld (A_4)**
Female ($\bar{\mathbf{B}}$)	2	7	11	9
Male ($\mathbf{B}$)	8	3	1	19
Total	10	10	12	28

$$Pr(A_2 \mid \bar{B}) = \frac{Pr(\bar{B} | A_2) Pr(A_2)}{Pr(\bar{B})} = \frac{0{,}7 \cdot 0{,}25}{1 - 0{,}47} = 0{,}33 \neq 0{,}16 = Pr(A_2 | B)$$

Orca Example Conclusion

- **Prior**: What is the probability of us being in the Antartica?

$$Pr(A_2) = 0{,}25$$

- **Likelihood**: A tacked orca is found dead in Antartica, what is the probability of it being male?

$$Pr(B|A_2) = 0{,}3$$

- **Posterior**: A tacked orca whale is found dead and is a male, what is the probability of us being in Antartica?

$$Pr(A_2|B) = 0{,}16$$

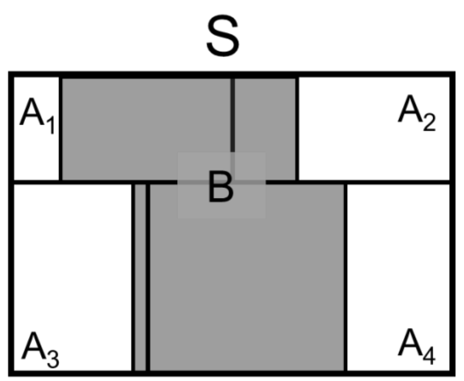
Orca Example – Another test method

- In a conversation effort, we pick up dead orcas from different oceans.
- The dead orcas are marked with the ocean and collected in the same container.
- A dead orca is randomly picked from the container: What is the probability that the orca is a male?

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)
Female ($\bar{B}$)	2	7	11	9
Male (B)	8	3	1	19
Total	10	10	12	28

Orca Example – Another test method

- Total Probability:



$$Pr(B) = Pr(A_1)\,Pr(B|A_1) + Pr(A_2)\,Pr(B|A_2) + Pr(A_3)\,Pr(B|A_3) + Pr(A_4)\,Pr(B|A_4)$$

Gender\ location	Atlantic (A_1)	Antartica (A_2)	Pacific (A_3)	Seaworld (A_4)	Total
Female ($\bar{\mathbf{B}}$)	2	7	11	9	29
Male ($\mathbf{B}$)	8	3	1	19	31
Total	10	10	12	28	60

$$Pr(B) = \frac{10}{60}\cdot\frac{8}{10}+\frac{10}{60}\cdot\frac{3}{10}+\frac{12}{60}\cdot\frac{1}{12}+\frac{28}{60}\cdot\frac{19}{28} = \frac{8+3+1+19}{60} = \frac{31}{60} = 0{,}517$$

Orca Example – Another test method

- If an orca found is a male, what is the probability that it is from the Antartica?

$$Pr(A_2|B)$$

- We use Bayes rule:

$$Pr(A_2|B) = \frac{Pr(A_2 \cap B)}{Pr(B)} = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)}$$

- $Pr(B) = 0{,}517; \; Pr(A_2) = 0{,}167; \; Pr(B|A_2) = 0{,}3$

$$Pr(A_2|B) = \frac{Pr(B|A_2)Pr(A_2)}{Pr(B)} = \frac{0{,}3 \cdot 0{,}167}{0{,}517} = \frac{3}{31} = 0{,}097$$

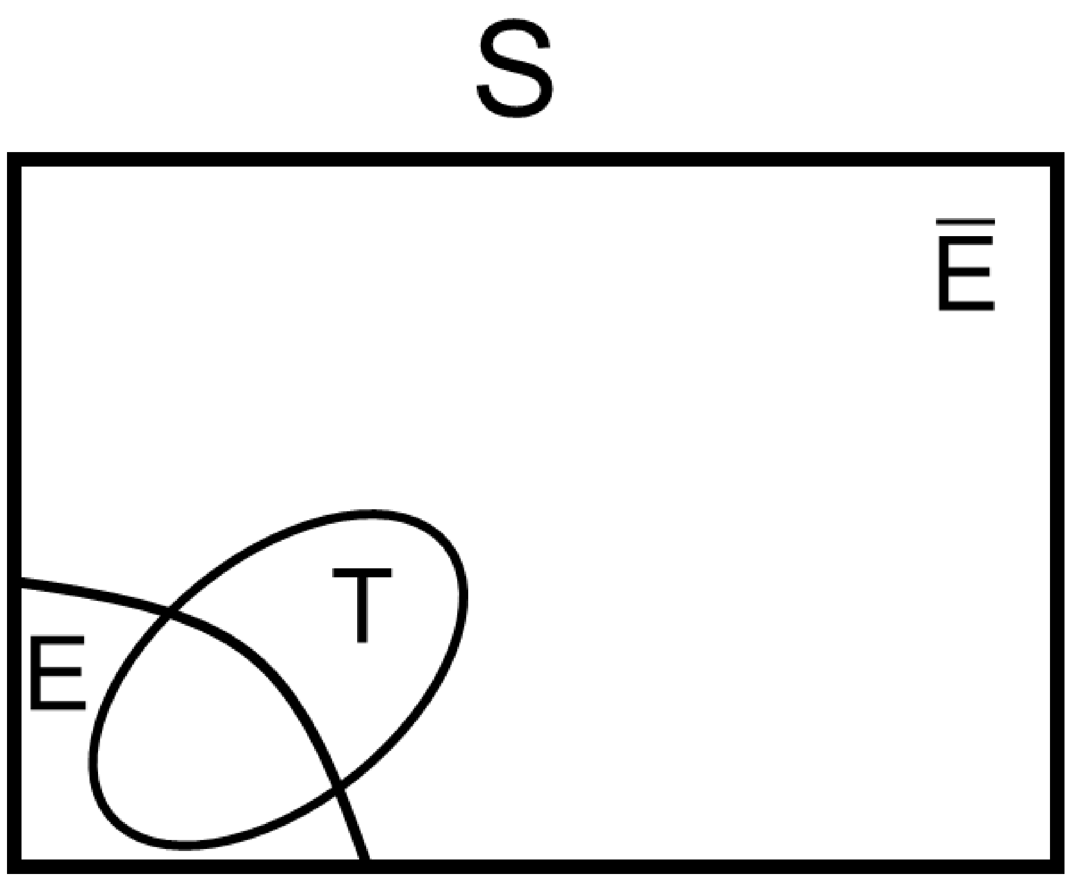
Tests and Types of Errors

- We can classify testing with two outcomes as:

Given / Result	**Disease (True)**	**No disease (False)**
Positive test	Sensitivity	Type I Error
Negative test	Type II Error	Specificity

Example: Ebola Test

- Event E: Patient are infectious with Ebola.
- Event T: The Ebola test is positive.



Example: Ebola Test

- **Prior**: What are the probability of a patient having Ebola?

$$Pr(E)$$

- **Likelihood**: What are the probability of a positive test given infectious with Ebola? Or of a negative test given not infectious with Ebola?

$$Pr(T|E) \quad \text{Sensitivity} \qquad\qquad Pr(\bar{T}|\bar{E}) \quad \text{Specificity}$$

- **Posterior**: What are the probability of being infectious given that a test is positive?

$$Pr(E|T)$$

Example: Ebola Test — Total Probability

- **Prior**: What are the probability of a patient having ebola?

Complement of E

$$Pr(E) = 0,01 \qquad Pr(\bar{E}) = 1 - 0,01 = 0,99$$

- **Likelihood**: What are the probabilities of the tests?

$$Pr(T|E) = 0,9$$ Sensitivity

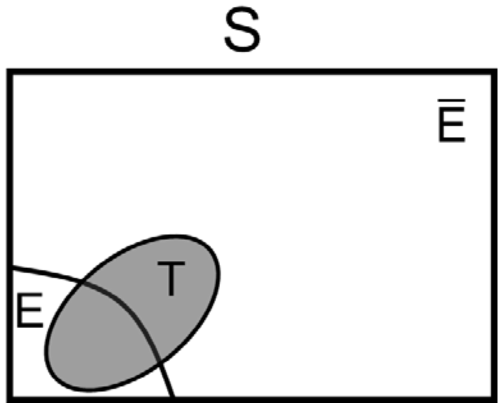
$$Pr(\bar{T}|\bar{E}) = 0,8$$ Specificity

- **Complement**: What are the probability of a patient having a positive test without being infectious?

$$Pr(T|\bar{E}) = 1 - Pr(\bar{T}|\bar{E}) = 0,2$$

Example: Ebola Test — Total Probability

- **Total Probability with the Sum Rule**: What are the probability of a patient having a positive test?



$$Pr(T) = Pr(T \cap E) + Pr(T \cap \bar{E})$$

- **The Product Rule**: We can with Bayes rule find

$$Pr(T) = Pr(T|E)\,Pr(E) + Pr(T|\bar{E})\,Pr(\bar{E})$$

$$= 0{,}9 \cdot 0{,}01 + 0{,}2 \cdot 0.99$$

$$= 0{,}207$$

Ebola Example — Posterior

- **We have**: We now know the probabilities:

$$P(E) = 0,01$$ Prior

$$P(T) = 0,207$$ Total probability

$$P(T|E) = 0,9$$ Likelihood

- **Product Rule**: What are the probability of being infectious given that a test is positive?

Bayes rule

$$Pr(E|T) = \frac{Pr(T|E)Pr(E)}{Pr(T)} = \frac{0,9 \cdot 0,01}{0,207} = 0,043$$

Ebola Example — Posterior

- What are the probability of being infectious given that a test is positive?

$$Pr(E|T) = \frac{Pr(T|E)Pr(E)}{Pr(T)} = \frac{0{,}9 \cdot 0{,}01}{0{,}207} = 0{,}043$$

- What are the probability of <u>not</u> being infectious given that a test is positive?

$$Pr(\bar{E} \mid T) = 1 - Pr(E|T) = 0{,}957$$

- What are the probability of <u>not</u> being infectious given a negative test?

$$Pr(\bar{E}|\bar{T}) = \frac{Pr(\bar{T} \mid \bar{E})Pr(\bar{E})}{Pr(\bar{T})} = \frac{0{,}8 \cdot 0{,}99}{0{,}793} = 0{,}999$$

- What are the probability of being infectious given that a test is negative?

$$Pr(\mathrm{E} \mid \bar{T}) = 1 - Pr(\bar{E}|\bar{T}) = 0{,}001$$

Ebola Example — Conclusion

- If the test is negative, it is allmost certain (99,9%) that you're not being infectious:

$$Pr\left(\bar{E}\middle|\bar{T}\right) = 0{,}999$$

- If the test is positive, there is still only a small risk (4,3%) that you actually are being infectious:

$$Pr(E|T) = 0{,}043$$

Monty Hall Dilemma

- We have three doors
- Behind two of the doors is a goat
- Behind one door is a million dollars ($)

- What is the chance of guessing behind which door the money is?

$$\Pr(\$|1) = \Pr(\$|2) = \Pr(\$|3) = \frac{1}{3}$$

Monty Hall Dilemma cont'd

- We make a selection of a door, say door 2, without open it.
- The quizmaster eliminates one of the doors ($\bar{\$}$) ,which we did not select, based on his knowledge on the goat situation, say door 1.
- We can now reselect between door 2 and 3.
- What are the probabilities of the money being behind the two doors? Should we switch door?

Monty Hall Dilemma cont’d

- What are the probabilities of the money being behind the two doors? Should we switch door?

The Binomial Distribution

- We have n repeated trials.
- Each trial has two possible outcomes
  - **Success** — probability p
  - **Failure** — probability q=1-p

Bernoulli trial

- What is the probability of having k successes out of n trials?
- We write this question as:

$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

- Faculty: $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$
  $0! = 1$

Bernoulli Trial

Definition: The binomial coefficient is defined as:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

Number of ways to select k objects out of a collection of n objects

Example: Out of 10 children, what is the probability that exactly 2 are girls?

$$\begin{aligned} Pr_n(k) &= \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= \frac{10!}{2!(10-2)!} (0,5)^2 (1-0,5)^{10-2} = 0,044 \end{aligned}$$

Combinatorics

- Take an object from a collection of n objects.
- Repeat the test k times.

Types of Experiments:

- With or without replacement
- Ordered or unordered

Example:

What is the probability that if I have two children that the oldest is a girl and the youngest is a boy?

- Ordered.
- With replacement.

Ordered with Replacement

- Take an object from a collection of n objects.
- **Put it back** each time.
- Repeat the test k times.
- **The sequence** of the objects **matters**.

- The number of combinations is: n^k
  - Each trial has n possible outcomes
  - All the trials are independent

Ordered without Replacement

- Take an object from a collection of n objects.
- **Do not** put it back each time.
- Repeat the test k times.
- **The sequence** of the objects **matters**.

- The number of combinations is:

$$ {}_nP_k = P_k^n = \frac{n!}{(n-k)!} = \mathrm{n} \cdot (n-1) \dots (n-k+1) $$

➢ The 1st trial has n possible outcomes, the 2nd trial has n-1 possible outcomes, … , the k'th trial has n-k+1 possible outcomes

Unordered without Replacement

- Take an object from a collection of n objects.
- **Do not** put it back each time.
- Repeat the test k times.
- **The sequence** of the objects **do not matter**.

- The number of combinations is:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

  - The k ordered draws can be shuffled in k! different ways (sequences)

Unordered with Replacement

- Take an object from a collection of n objects.
- **Put it back** each time.
- Repeat the test k times.
- **The sequence** of the objects **do not matter**.

- The number of combinations is:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

➢ Each time we draw an object, we should replace an object (except for the last draw). This correspond to we start with n+k-1 object and draw k objects unordered without replacement.

Summary of Combinatorics

- We can summarise the number of possible outcomes of k trials, sampled from a set of n objects.

		Replacement	
		With	Without
Sam-pling	Ordered	n^k	$P_k^n = \frac{n!}{(n-k)!}$
	Unordered	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$	$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

Experiment: Birthday Example

- k=35 students
- n=365 (number of days in the year)
- What are the probability that at least two have birthday on the same day (E)?

All have different birthdays

Complement rule

Ordered sampling without replacement
(k unique birthsdays in n days)

$$\Pr(E) = 1 - \Pr(\bar{E}) = 1 - \frac{\frac{n!}{(n-k)!}}{n^k} = 1 - \frac{\frac{365!}{(365-35)!}}{365^{35}} > 80\%$$

Ordered sampling with replacement
(all possible combinations of k students birthdays in n days)

- k=50 students: Pr(E)>97%
- k=75 students: Pr(E)>99,97%

Words and Concepts to Know

Prior

Type I Error

Binomial coefficient

Sampling

Unordered

Replacement

Specificity

Likelihood

Combinatorics

Bernoulli Trial

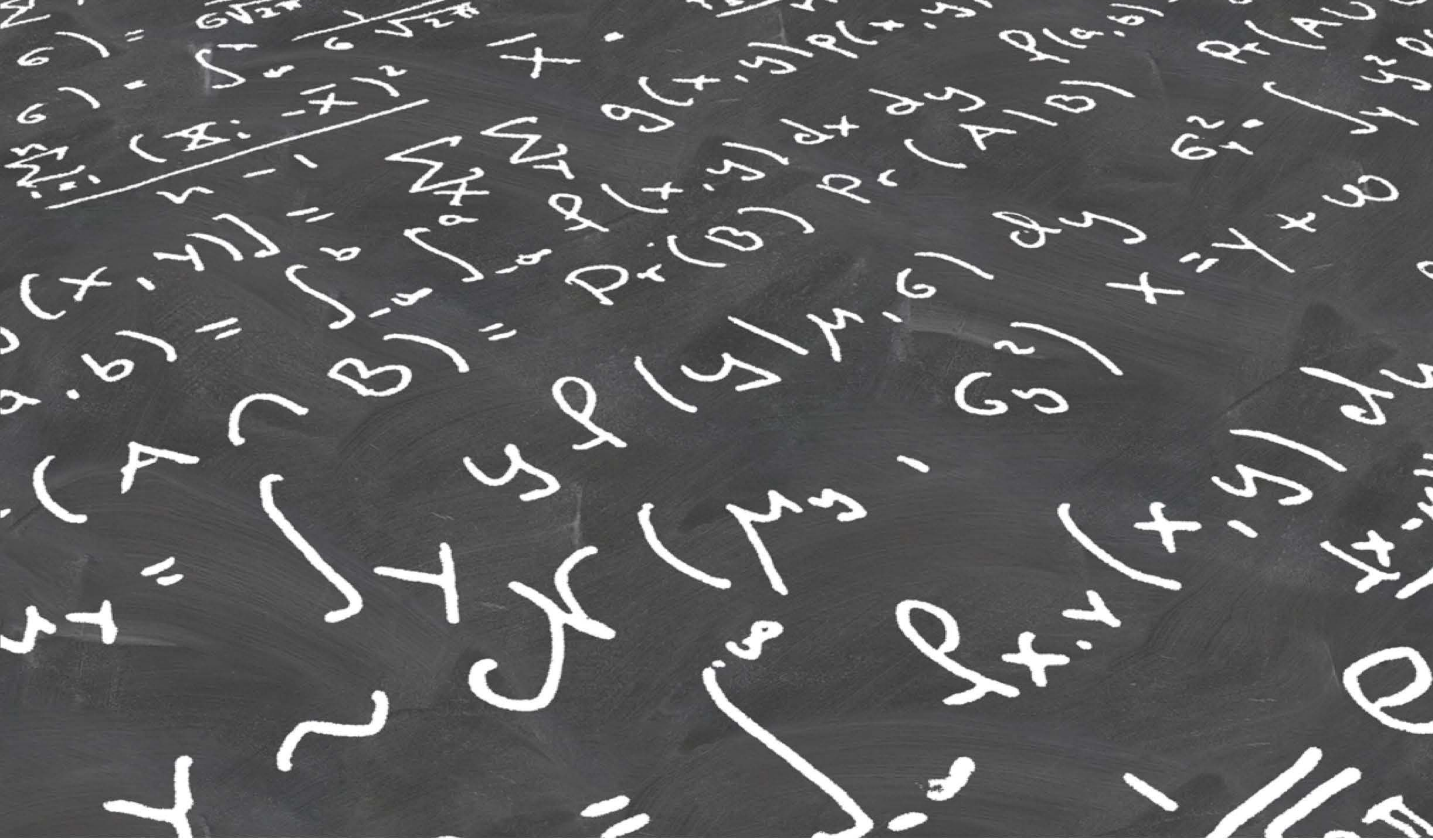
Sensitivity

Posterior

Ordered

Type II Error

Binomial distribution



3. Discrete Random Variables

Gunvor Elisabeth Kirkelund
Lars Mandrup

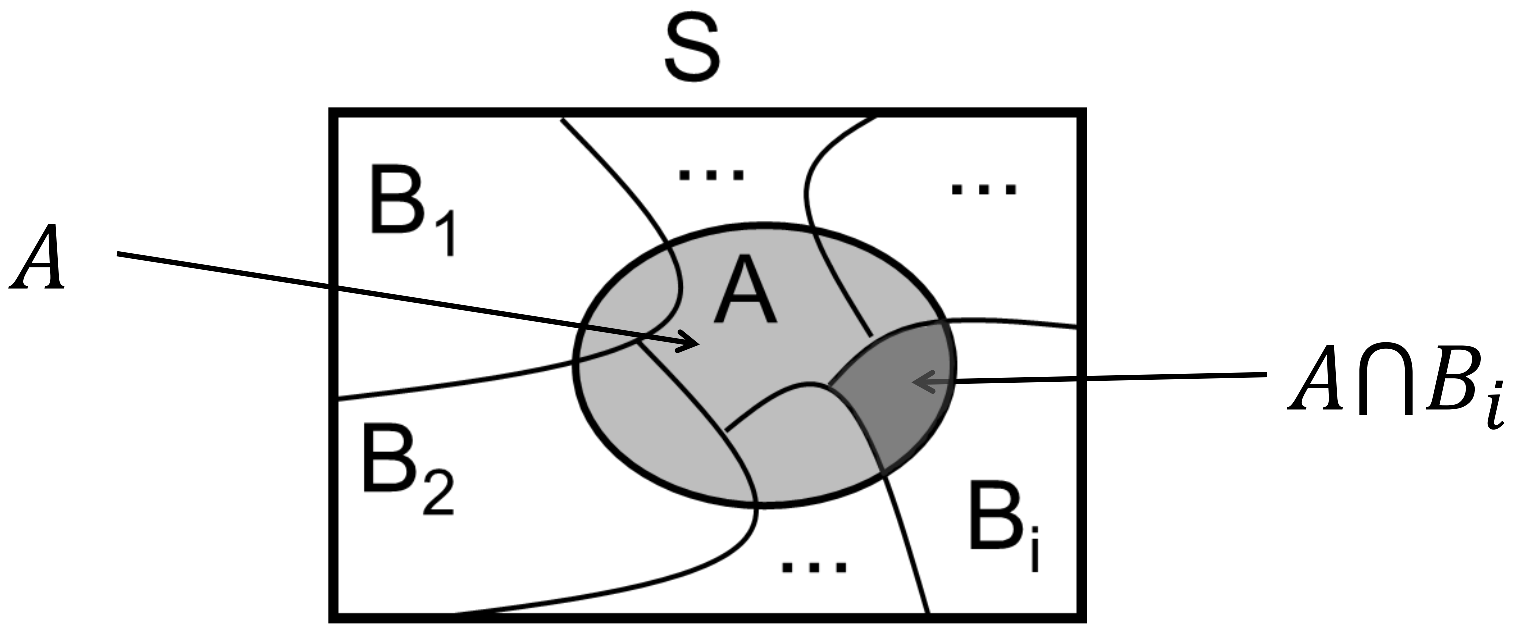
Agenda for Today

- Repetition from last time
- Definition of a Stochastic Random Variable
- Discrete Stochastic Variables

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.



$$Pr(A) = Pr(A \cap B_1) + Pr(A \cap B_2) + \cdots + Pr(A \cap B_i) + \cdots$$
$$= Pr(A|B_1) \cdot Pr(B_1) + Pr(A|B_2) \cdot Pr(B_2) + \cdots$$

where the B_i's are mutually exclusive ($B_i \cap B_j = \emptyset$ for $i \neq j$)
and $S = B_1 \cup B_2 \cup \ldots \cup B_i \cup \ldots$

Bayesian Terms

- **Prior**: What are the overall probability of an event E?

$$Pr(E)$$

- **Likelihood**: What are the probability of a test T given event E?

$$Pr(T|E) = \frac{Pr(T \cap E)}{Pr(E)} = \frac{Pr(E|T) \cdot Pr(T)}{Pr(E)}$$

- **Total Probability**: What is the total probability of the test?

$$Pr(T) = Pr(T|E) \cdot Pr(E) + Pr\left(T|\bar{E}\right) \cdot Pr(\bar{E})$$

- **Posterior**: What are the probability the event given the test T?

$$Pr(E|T) = \frac{Pr(T \cap E)}{Pr(T)} = \frac{Pr(T|E) \cdot Pr(E)}{Pr(T)}$$

Combinatorics

- The number of possible outcomes of k trials, sampled from a set of n objects.

Types of Experiments:

- With or without replacement
- Ordered or unordered

		Replacement	
		With	Without
Sam-pling	Ordered	n^k	$P_k^n = \frac{n!}{(n-k)!}$
	Unordered	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$	$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

The Binomial Distribution

- We have n repeated trials.
- Each trial has two possible outcomes
  - **Success** — probability p
  - **Failure** — probability q=1-p

Bernoulli trial

- What is the probability of having k successes out of n trials?
- We write this question as:

$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

- Faculty: $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$
  $0! = 1$

Also called a random experiment

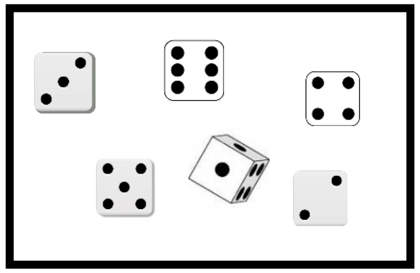
Stochastic Experiment

- An experiment in which you can not predict the outcome

Examples:

- Rolling a dice
- Sample space for the experiment is: {1, 2, 3, 4, 5, 6}

S



- Flip a coin

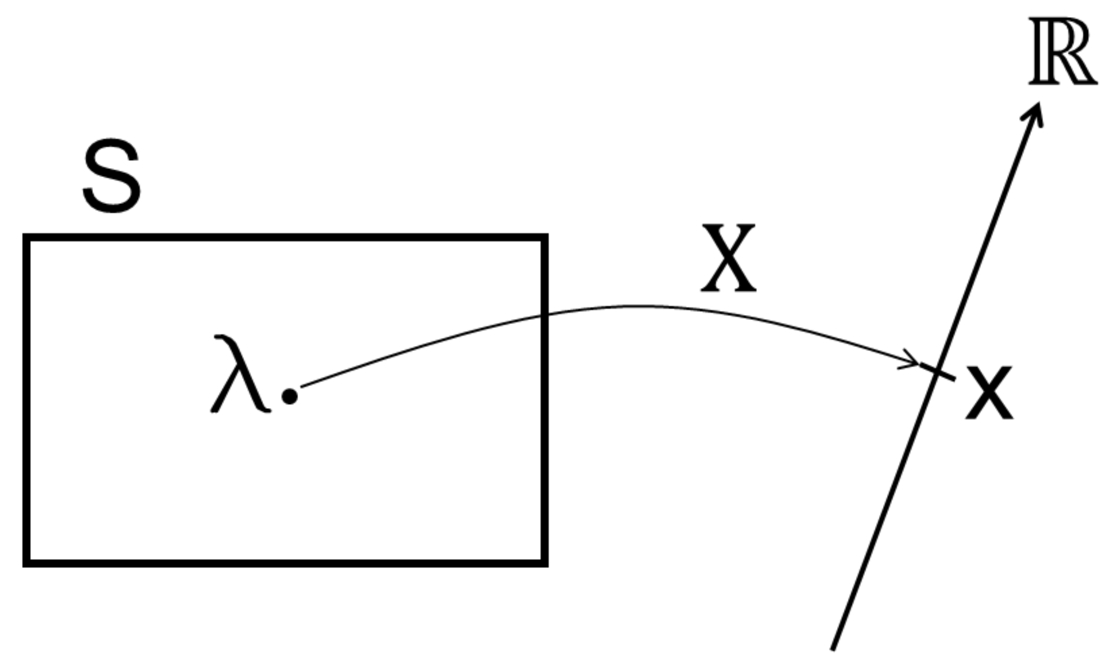
- Sample space for the experiment is: {head, tail}

S

Also just called a random variables

Stochastic Random Variables

- A random variable tells something important about a stochastic experiment.
- Can be discrete or continous



Examples:

- The numbers on a dice (discrete):
  - Sample space for variable X is : $\{1,2,3,4,5,6\}$
  - Sample space for variable Y "Even (1)/Uneven (-1)": $\{1,-1\}$
- The hight of students at IHA (continous):
  - Sample space for variable H is all real numbers: [100;250] cm.

Probability Mass Function (PMF)

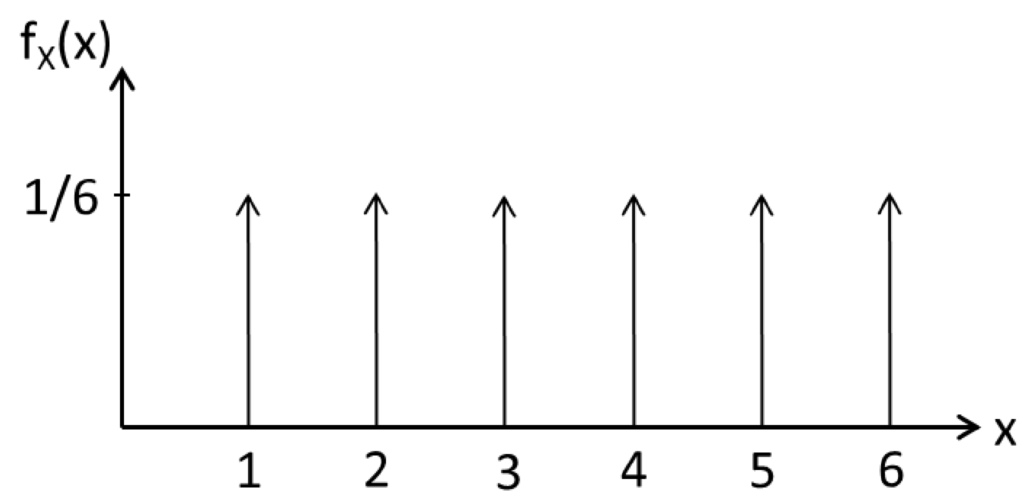
- Sample space for X.
- X is a <u>discreet</u> stochastic variable.

$$\boxed{f_X(x) = \begin{cases} Pr(X = x_i) & for\ X = x_i \\ 0 & otherwise \end{cases}}$$

$$0 \leq f_X(x) \leq 1$$

- We have that: $\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} Pr(X = x_i) = 1$

Example: Laplace Dice
(perfect dice)



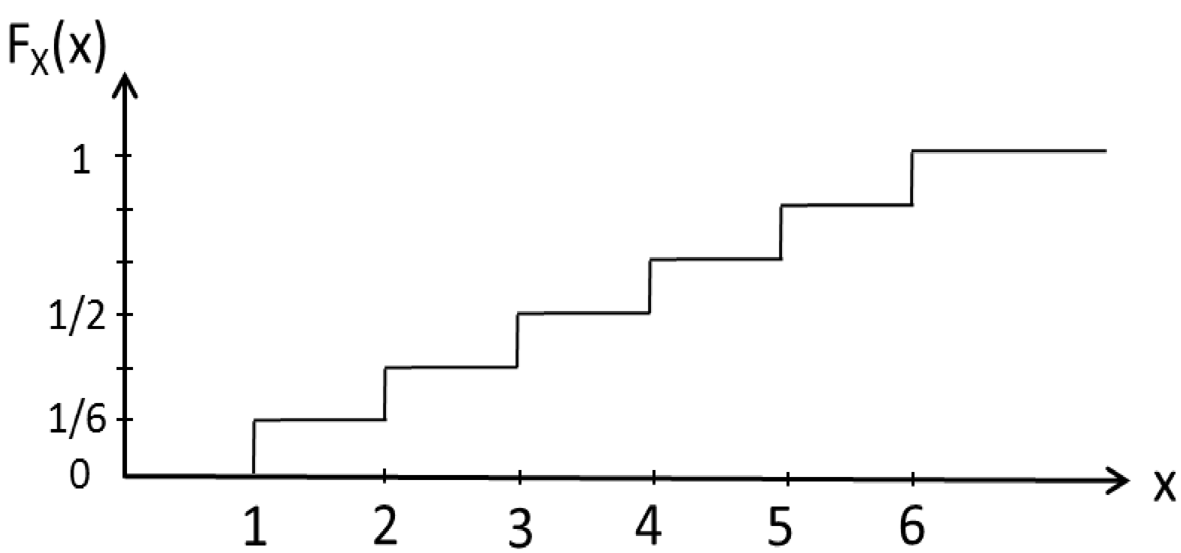
Cumulative Distribution Function (CDF)

- Sample space for X.
- X is a discreet stochastic variable.
- $F_X(x)$ is a non-decreasing step-function.

$$\boxed{F_X(x) = Pr(X \leq x)} \qquad 0 \leq F_X(x) \leq 1$$

- We have that: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$

Example: Laplace Dice
(perfect dice)



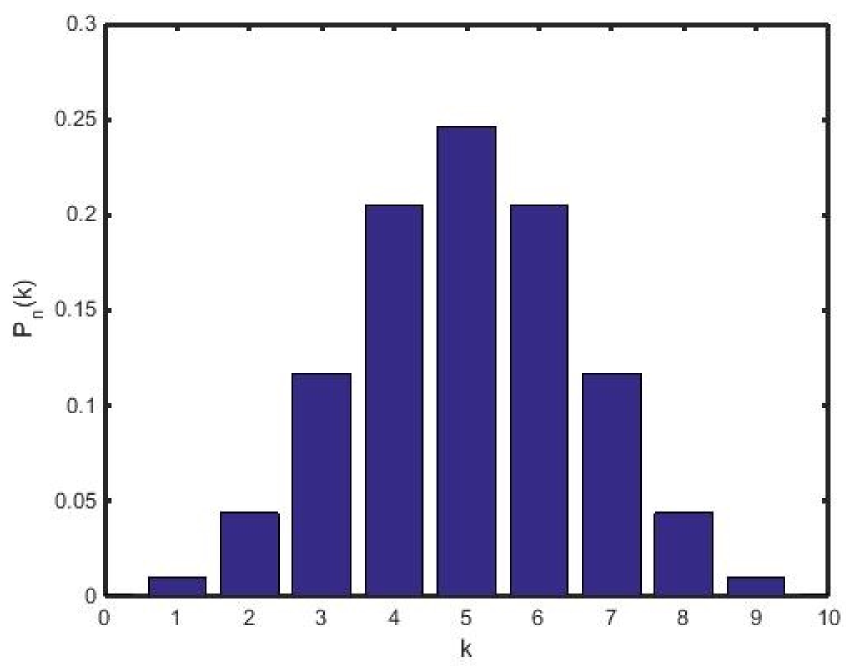
The Binomial Mass Function

- We have n repeated trials.
- Each trial has two possible outcomes
  - **Success** — probability p
  - **Failure** — probability 1-p

Also called a Bernoulli trial

- We write the mass function as:

$$\boxed{f(k|n,p) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}}$$



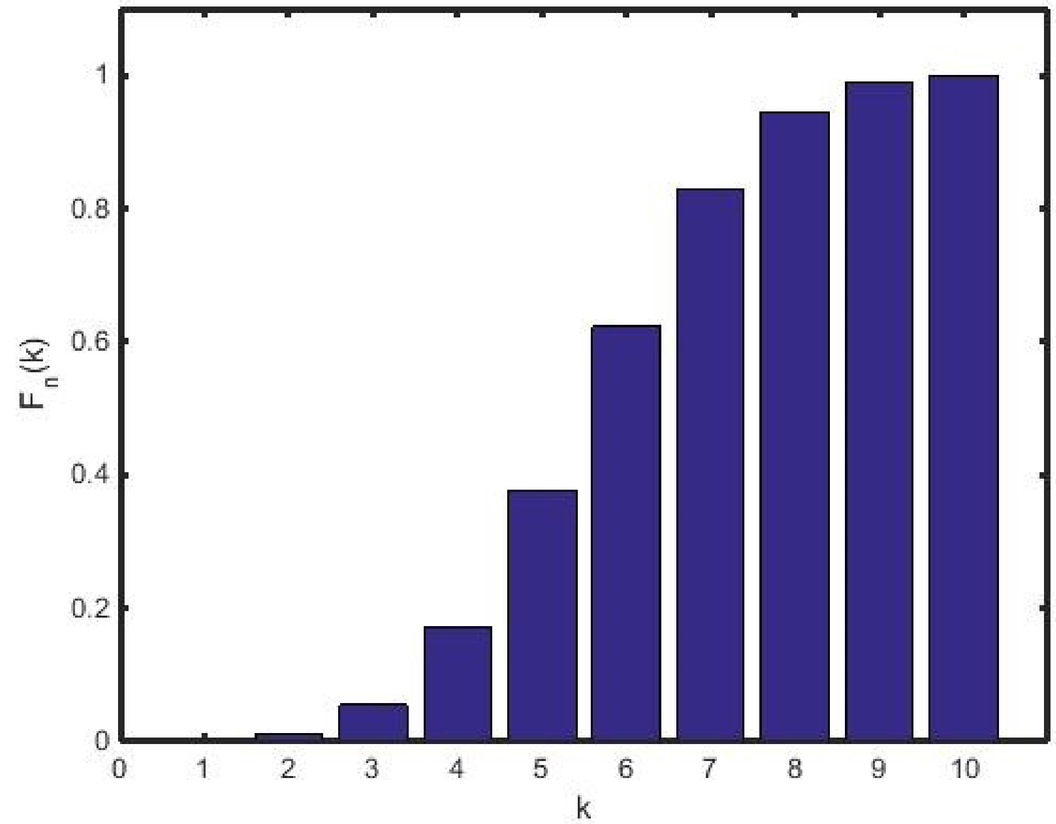
The Binomial Distribution

- The probability mass function is given as:

$$f(k|n,p) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

- We write the distribution as the sum:

$$F(k|n,p) = \sum_{i=0}^{k} f(i|n,p)$$



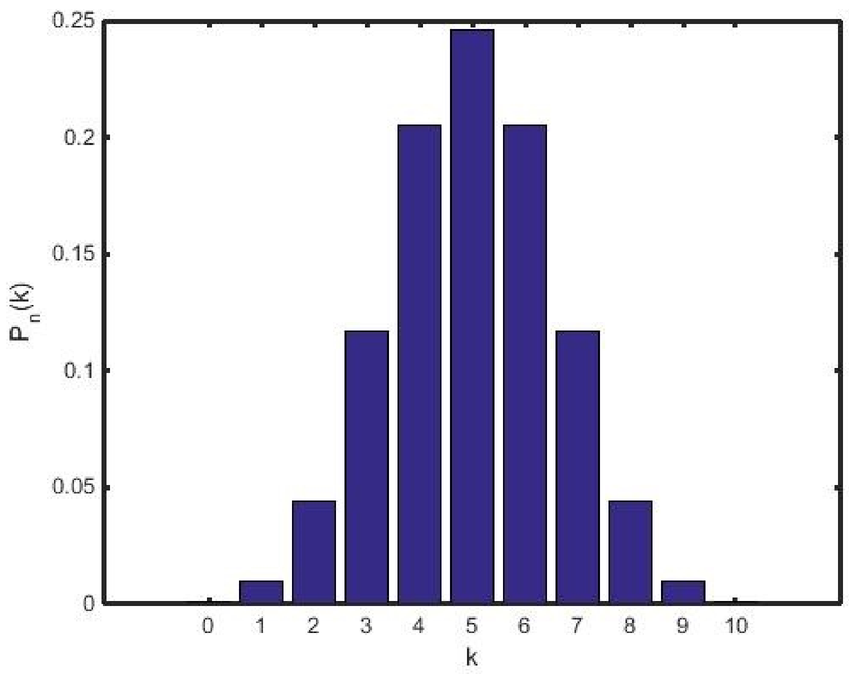
Expectation of a Discrete Random Variable

Example: If I want ten children, how many girls can I expect to get?

Answer: I assume a Binomial distribution with p=0.5:

$$f(k|10{,}0.5) = \binom{10}{k} \cdot 0.5^k \cdot 0.5^{10-k} = \binom{10}{k} \cdot 0.5^{10}$$

$$where \ \binom{10}{k} = \frac{10!}{k!\,(10-k)!}$$



$$E[k] = 0 \cdot f(0|10{,}0.5) + 1 \cdot f(1|10{,}0.5) + \cdots + 10 \cdot f(10|10{,}0.5)$$
$$= \left(0 + 1 \cdot \binom{10}{1} + 2 \cdot \binom{10}{2} \ldots + 10 \cdot \binom{10}{10}\right) \cdot 0.5^{10}$$
$$= (0 + 1 \cdot 10 + 2 \cdot 45 + \cdots + 10 \cdot 1) \cdot 0.5^{10} = 10 \cdot 0.5 = 5$$

Expectation of a Discrete Random Variable

- We define the mean or the expectation of a discreet random variable as:

n is the number of outcomes

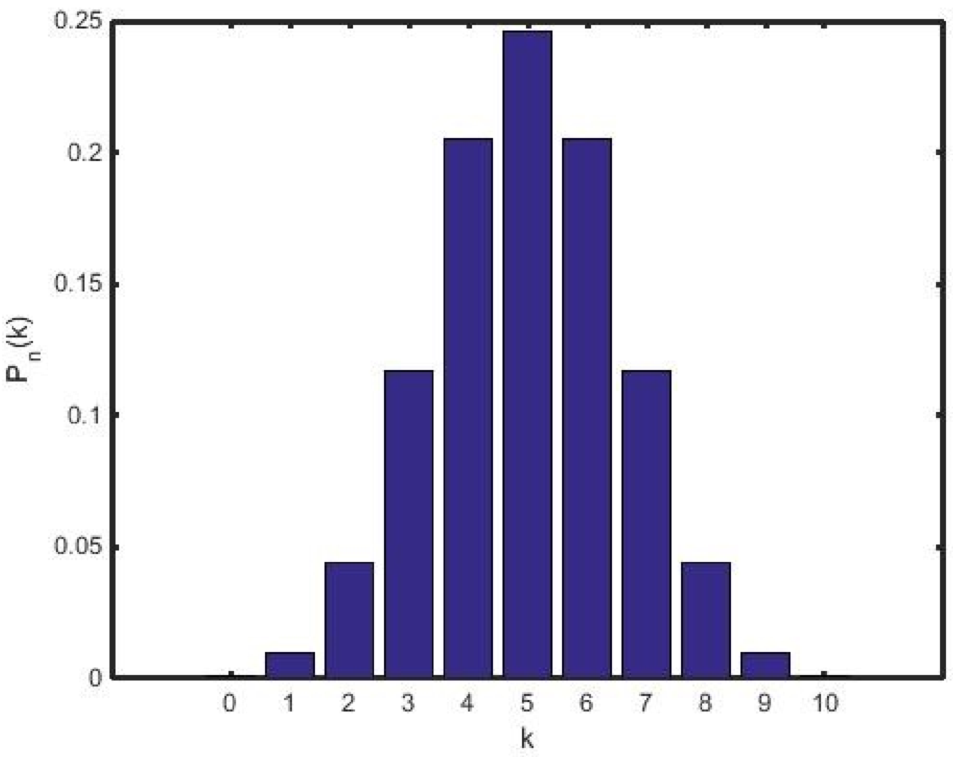
$$\boxed{\bar{X} = E[X] = \sum_{i=1}^{n} x_i \, f_X(x_i)}$$

x_i is its outcome

The Binomial Distribution (cont'd)

- For the Binomial distribution, we have:

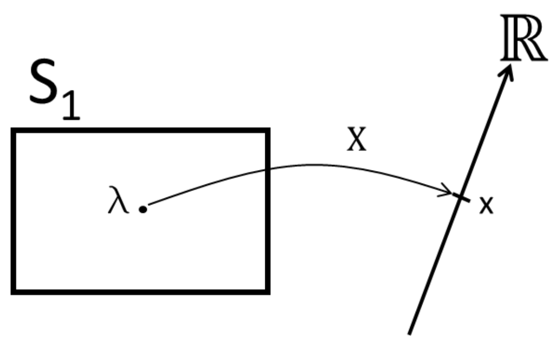
$$\boxed{\begin{aligned} E[k] &= n \cdot p \\ Var(X) &= n \cdot p \cdot (1-p) \end{aligned}}$$

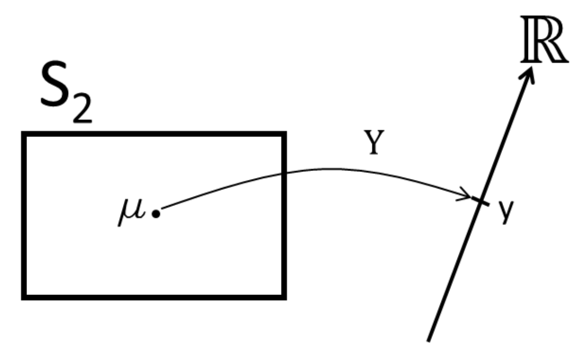


- Where the variance is defined as:

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

Two Simultaneous Discreet Random Variables

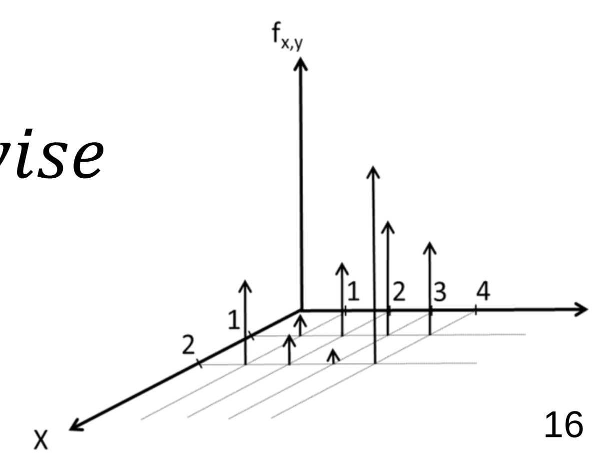




- Two (or more) discreet random variables X and Y
- We can discribe the two probabilities as a simultaneous pmf:

Joint (Simultaneous) pmfs:

$$f_{X,Y}(x,y) = \begin{cases} Pr\Big((X = x_i) \cap (Y = y_j)\Big) & for\ \ X = x_i \wedge Y = y_j \\ 0 & otherwise \end{cases}$$

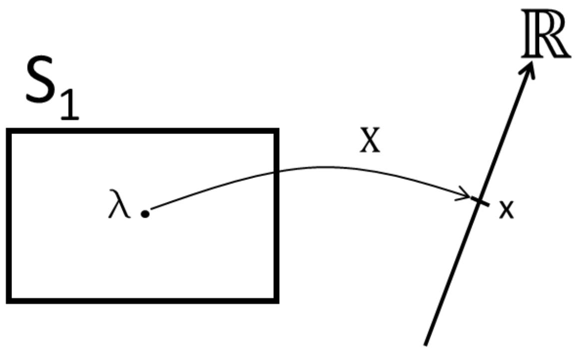


Fx.: X = The number of bicycles in front of IHA
Y = The number of people inside IHA

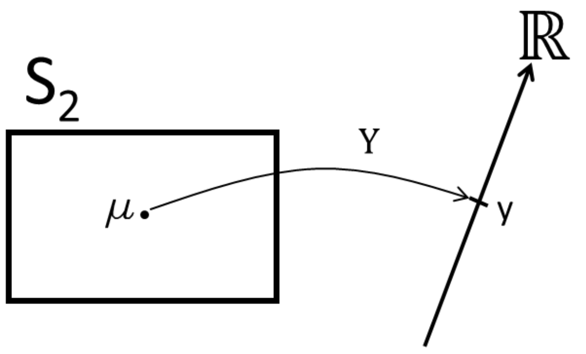
Two Simultaneous Discrete Random Variables

Marginal pmfs:

$$f_X(x) = \sum_y f_{X,Y}(x,y) \qquad f_Y(y) = \sum_x f_{X,Y}(x,y)$$



Conditional pmfs / Bayes Rule:



$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \Pr(X = x|\mathrm{Y} = \mathrm{y})$$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \Pr(Y = y|\mathrm{X} = \mathrm{x})$$

Orca Example

- Let us assume that the discreet simultaneous mass function (pmf) for observing a orca at a specific ocean and its gender is

$f_{X,Y}(x,y)$ $f_X(x)$

Gender (X)\ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total
female (1)	2/60	7/60	11/60	9/60	29/60
male (2)	8/60	3/60	1/60	19/60	31/60
Total	10/60	10(60	12/60	28/60	1

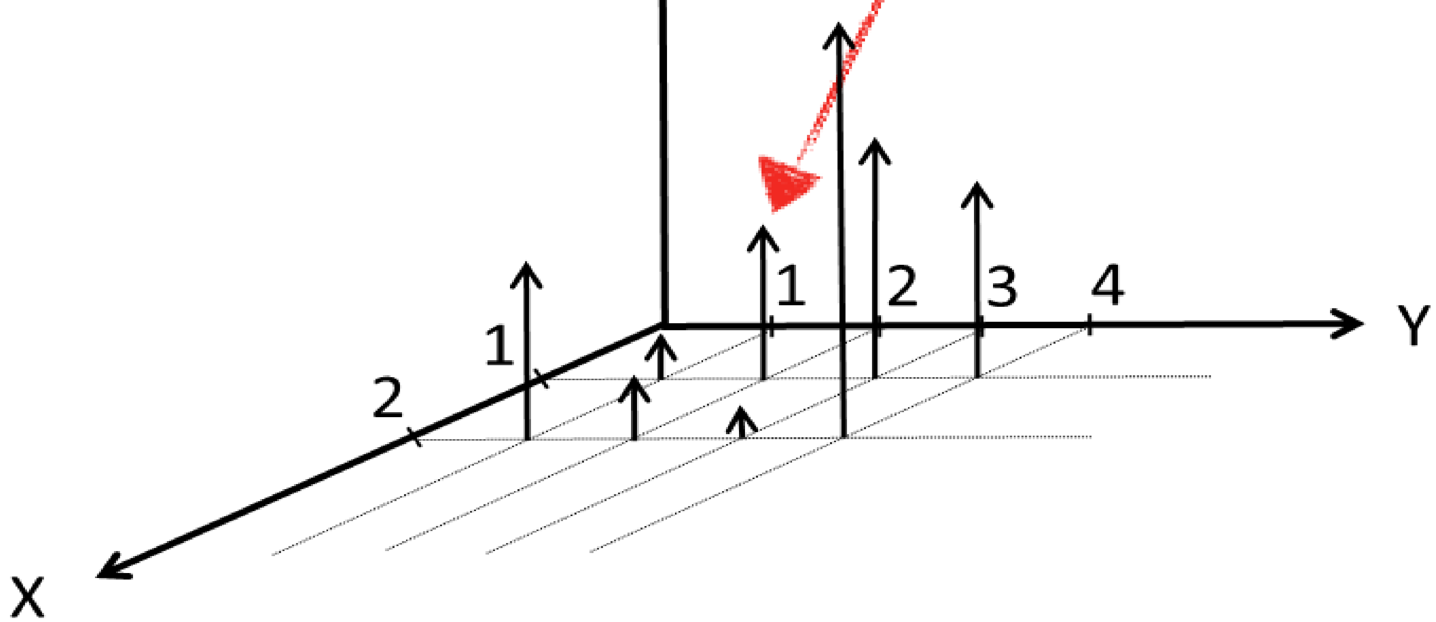
$f_Y(y)$

Fx.: $Pr(Male|Atlantica) = f_{X|Y}(2|1) = \frac{f_{X,Y}(2,1)}{f_Y(1)} = \frac{8/_{60}}{10/_{60}} = \frac{8}{10} = 0{,}8$

Orca Example - Joint pmf

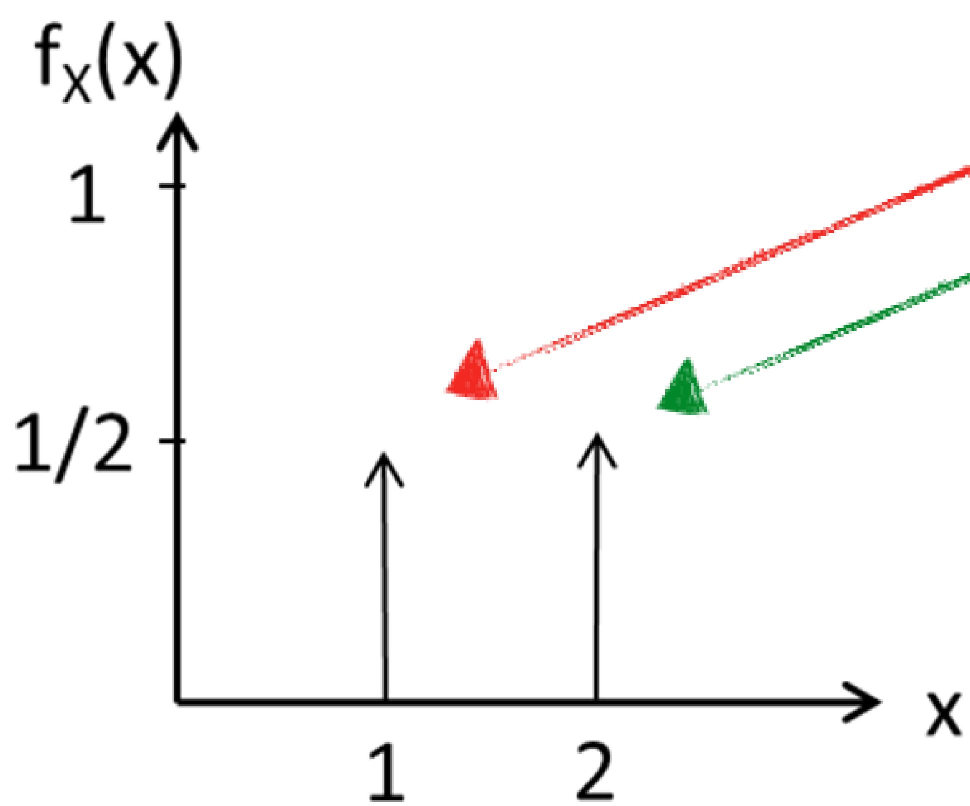
Gender (X)\ Location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total
female (1)	2/60	7/60	11/60	9/60	29/60
male (2)	8/60	3/60	1/60	19/60	31/60
Total	10/60	10(60	12/60	28/60	1

$$f_{X,Y}(1,2) = \frac{7}{60}$$



Orca Example – Marginal pmf

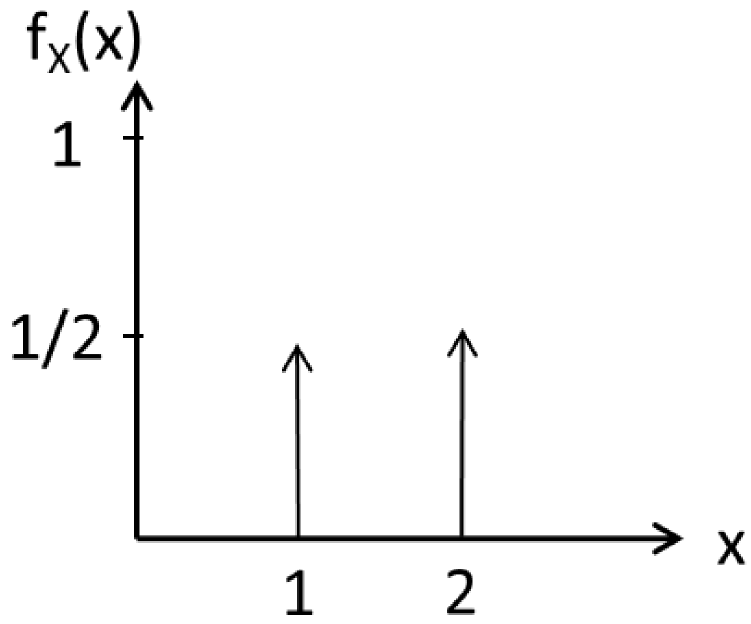
Gender (X)\ location (Y)	Atlantic (1)	Antartica (2)	Pacific (3)	Seaworld (4)	Total
female (1)	2/60	7/60	11/60	9/60	29/60
male (2)	8/60	3/60	1/60	19/60	31/60
Total	10/60	10(60	12/60	28/60	1



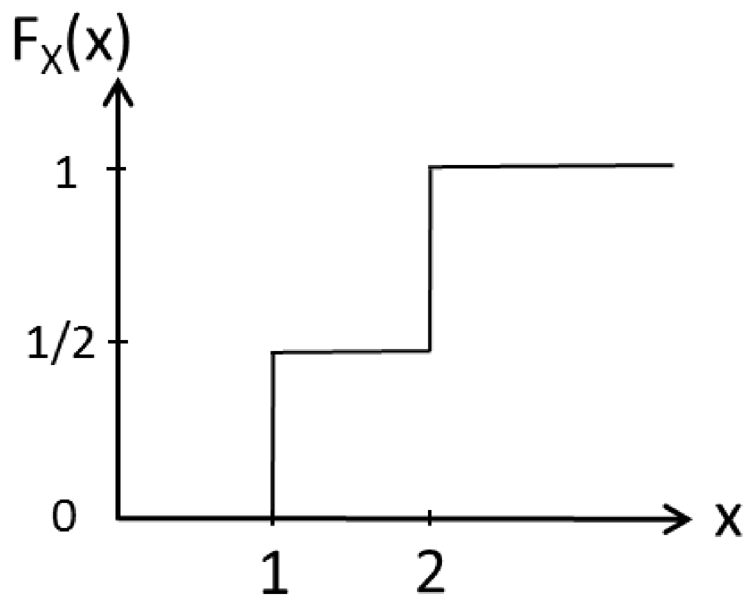
$$
\begin{aligned}
f_X(1) &= f_{X,Y}(1,1) + f_{X,Y}(1,2) + f_{X,Y}(1,3) + f_{X,Y}(1,4) \\
&= \frac{2}{60} + \frac{7}{60} + \frac{11}{60} + \frac{9}{60} = \frac{29}{60} \\
f_X(2) &= f_{X,Y}(2,1) + f_{X,Y}(2,2) + f_{X,Y}(2,3) + f_{X,Y}(2,4) \\
&= \frac{8}{60} + \frac{3}{60} + \frac{1}{60} + \frac{19}{60} = \frac{31}{60}
\end{aligned}
$$

Orca Example – Quick Rewrite to cdf

- We can rewrite the pmf to the cdf



Marginal pmf



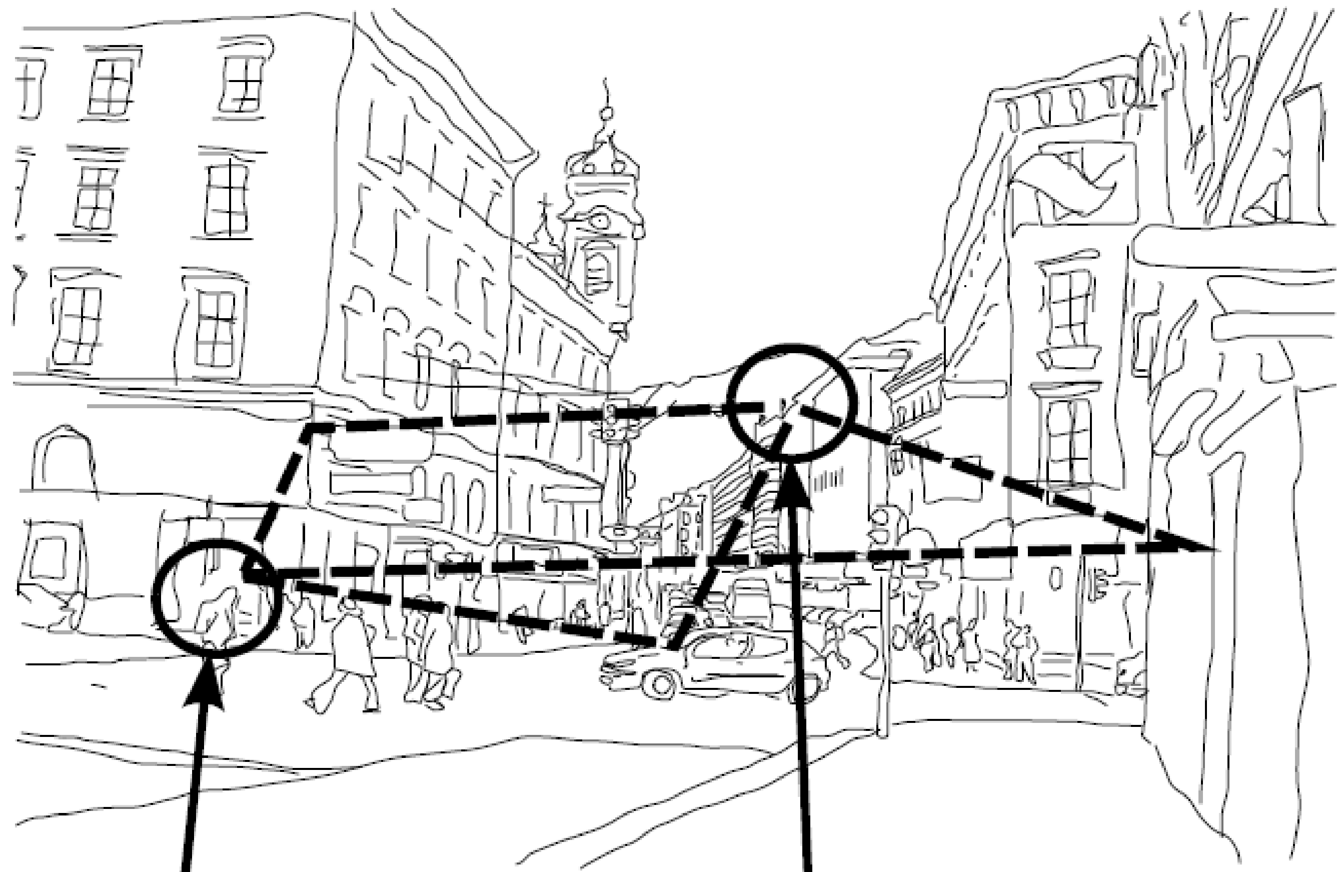
Marginal cdf

$$f_X(1) = \frac{29}{60}$$
$$f_X(2) = \frac{31}{60}$$

$$F_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{29}{60} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } 2 \leq x \end{cases}$$

Example - Wireless Channel

- A signal in a wireless channel travels with equal probability of three different path from transmitter to receiver



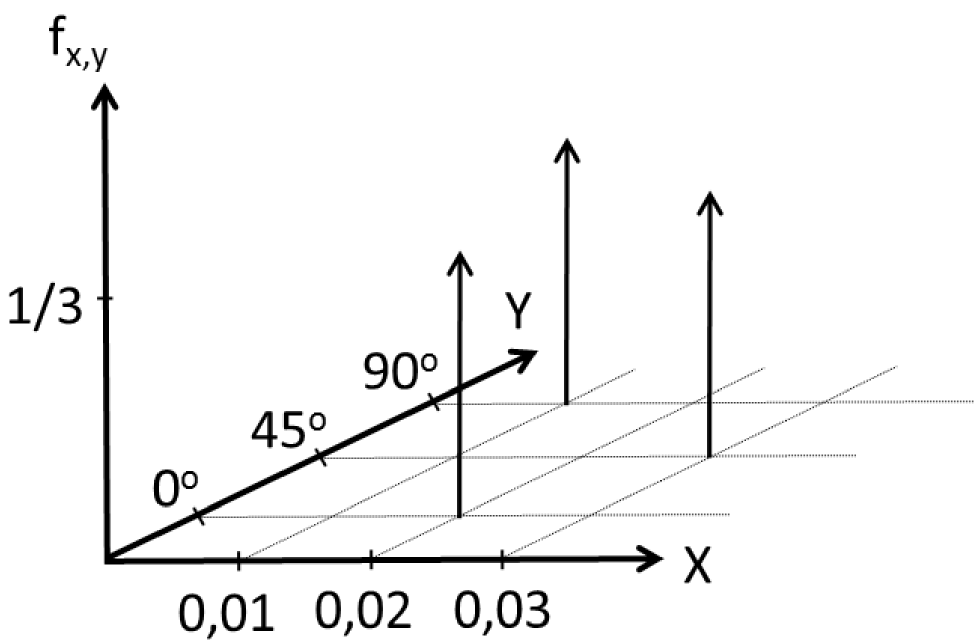
Amplitude\ Phase	0^o	45^o	90^o	**Total**
0.01	0	0	$\frac{1}{3}$	$\frac{1}{3}$
0.02	$\frac{1}{3}$	0	0	$\frac{1}{3}$
0.03	0	$\frac{1}{3}$	0	$\frac{1}{3}$
Total	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	**1**

Example - Wireless Channel: Assignment

- Plot the pmf for the wireless channel.
- What is the Expected Amplitude and Phase?

X Y

Amplitude\ Phase	0^o	45^o	90^o	Total
0.01	0	0	$\frac{1}{3}$	$\frac{1}{3}$
0.02	$\frac{1}{3}$	0	0	$\frac{1}{3}$
0.03	0	$\frac{1}{3}$	0	$\frac{1}{3}$
Total	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	**1**



$$E[X] = (0{,}01 + 0{,}02 + 0{,}03) \cdot \frac{1}{3} = 0{,}02$$

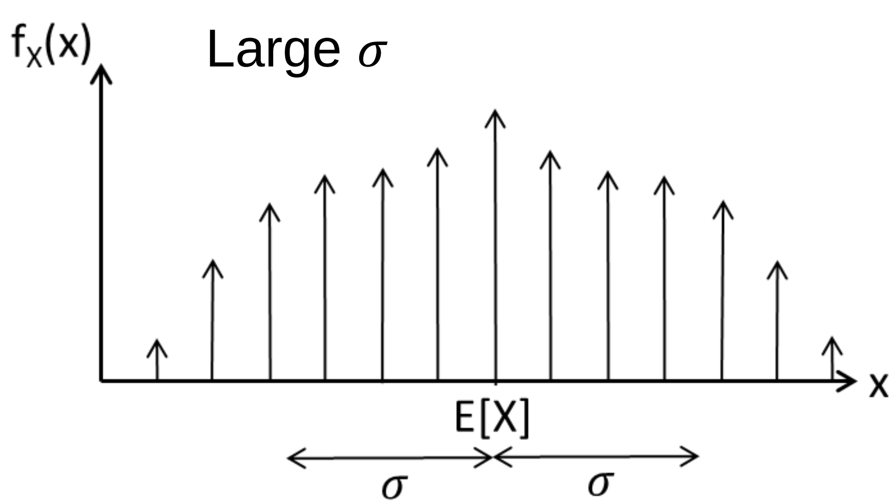
$$E[Y] = (0^o + 45^o + 90^o) \cdot \frac{1}{3} = 45^o$$

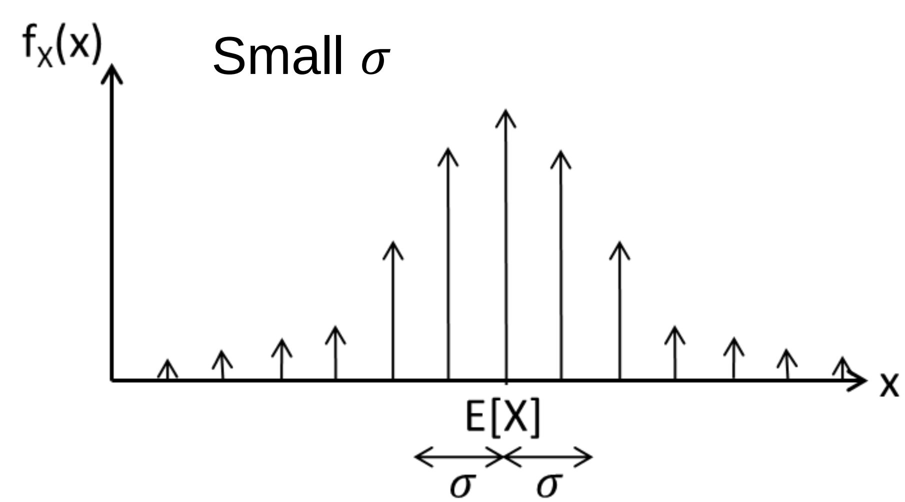
Variance and standard deviation

Variance and standard deviation tells of the spreading of the data

- The variance is an indicator on how much the values of a random variable X are spread around (deviates from) the expectation value.
- The standard deviation σ is the square root of the variance.

$$\boxed{Var(X) = \sigma_X{}^2 = E[X^2] - E[X]^2}$$





Correlation Coefficient

Correlation tells of the coupling between variables

- The correlation coefficient, is an indicator on how much two random variables X and Y are correlated.

$$\boxed{\rho = E\left[\frac{X-\bar{X}}{\sigma_X}\cdot\frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}}$$

- We have that: $-1 \leq \rho \leq 1$

Independence

- We have independence between X and Y if and only if:

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$$

Example of independent random variables:

- A persons height and the current exact distance from the earth to the moon.

Example of dependent random variables:

- The time of day and the amount of bicycles parked the at the engineering college.
- The energy of a mobile signal and the length in meters to a basestation.

Independence

Independence: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

- Bayes Rule: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

  gives that if X and Y are independent, then:

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

- Also:

  $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \Rightarrow E[XY] = E[X]E[Y] \Rightarrow \rho = 0$

  but the opposite is not allways true!

Dependant Variables – Simple Example

- Given a random variable X
- We define a new random variable Y=X

$$f_{X,Y}(1,1) = \frac{1}{2}$$

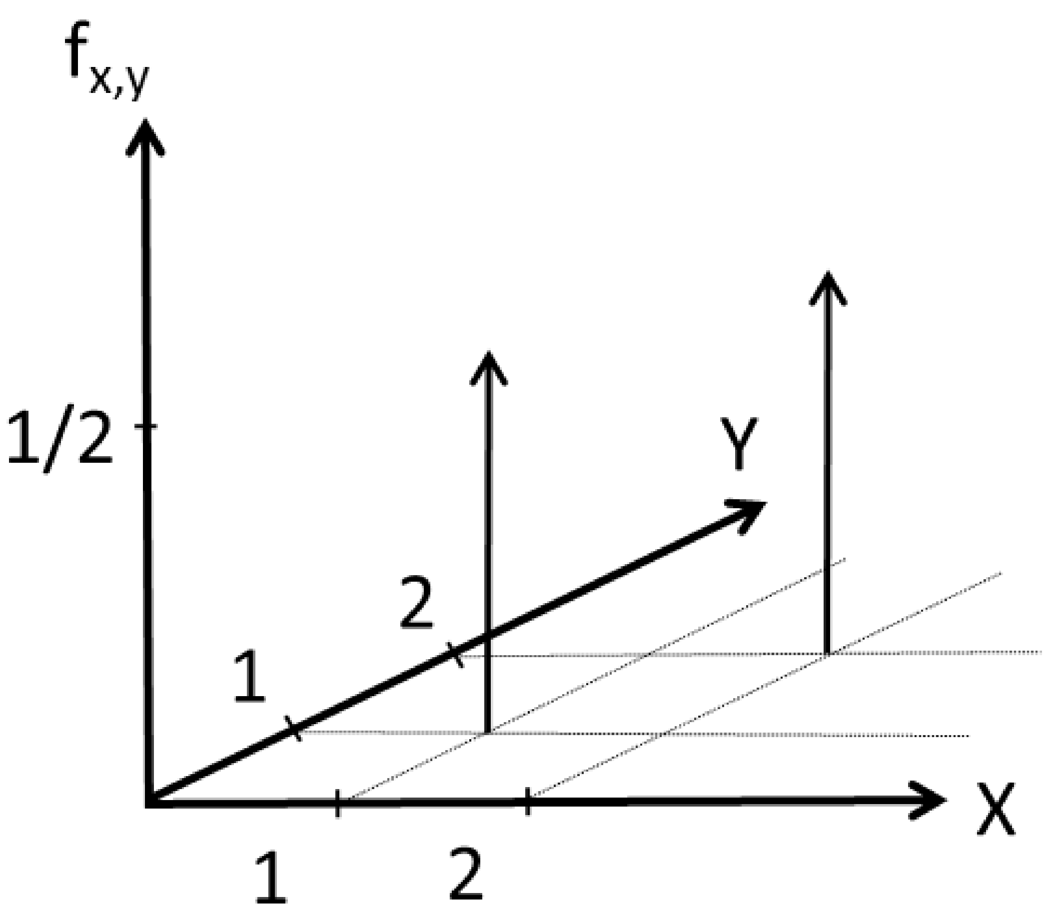
$$f_{X,Y}(2,2) = \frac{1}{2}$$

$$f_{X,Y}(1,2) = 0$$

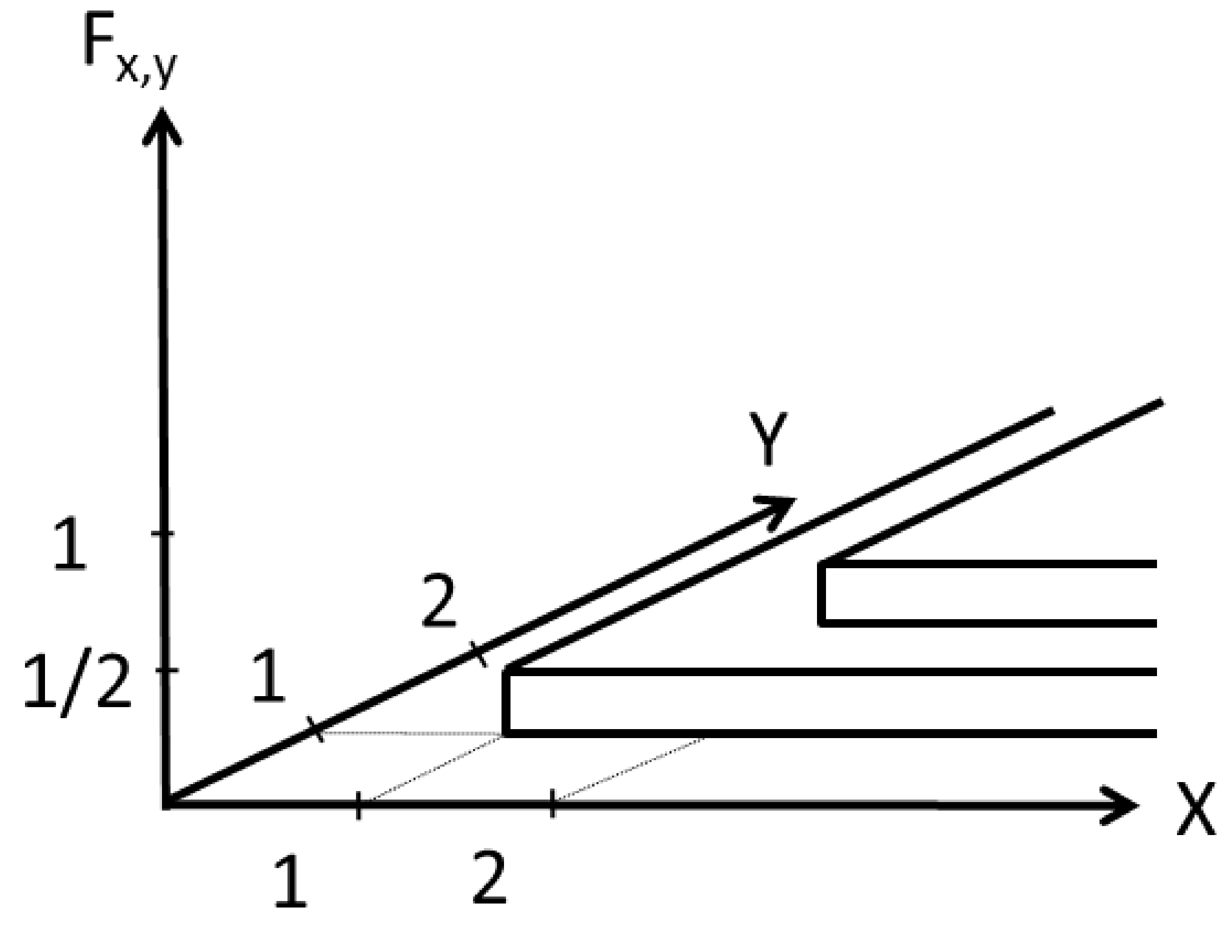
$$f_{X,Y}(2,1) = 0$$

Simple Example - Simultaneous pmf

Plots of the pmf and the cdf:



pmf



cdf

Simple Example – Marginal pmf

$$f_Y(y) = \sum_x f_{X,Y}(x,y)$$

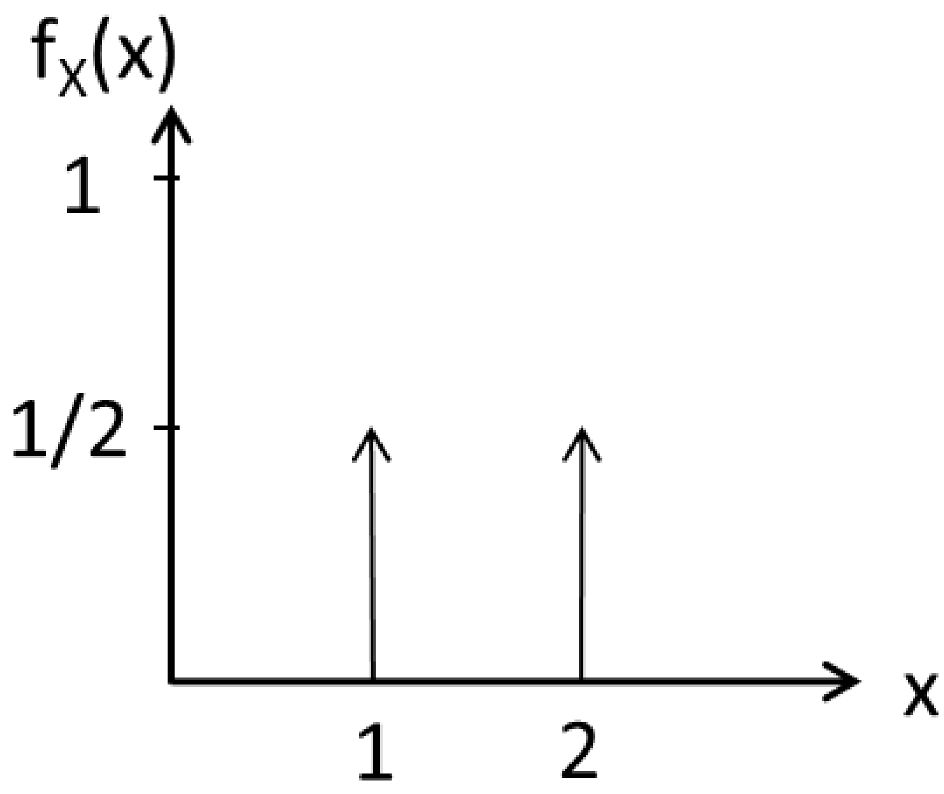
$$f_Y(1) = f_{X,Y}(1,1) + f_{X,Y}(2,1) = \frac{1}{2}$$

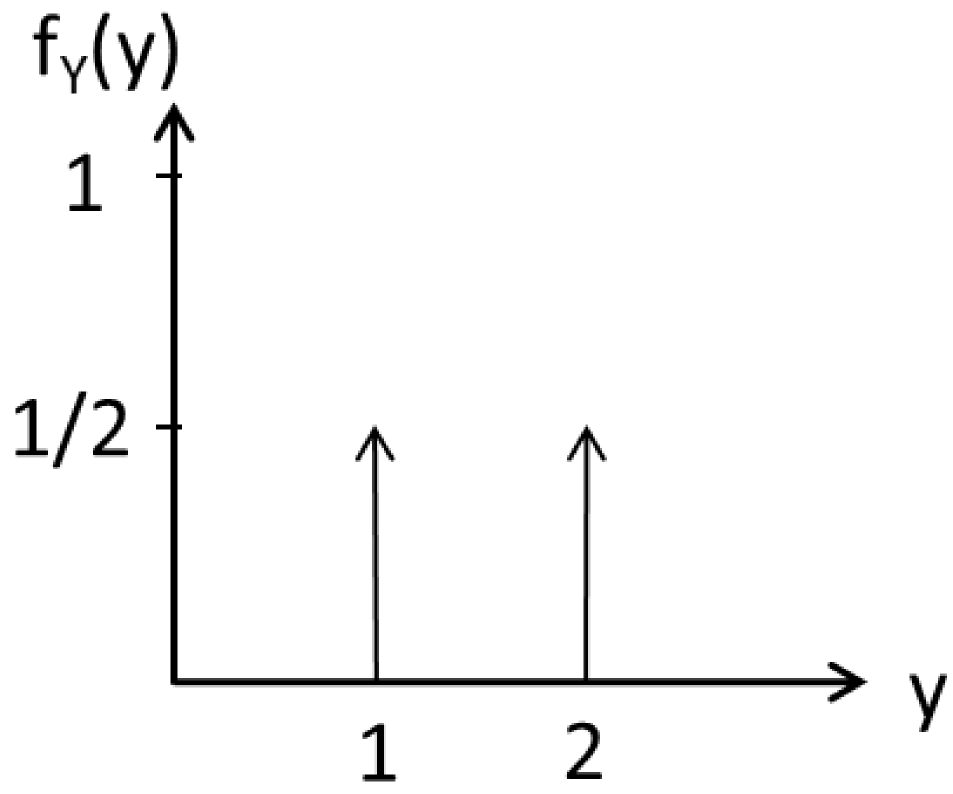
$$f_Y(2) = f_{X,Y}(1,2) + f_{X,Y}(2,2) = \frac{1}{2}$$

$$f_X(x) = \sum_y f_{X,Y}(x,y)$$

$$f_X(1) = f_{X,Y}(1,1) + f_{X,Y}(1,2) = \frac{1}{2}$$

$$f_X(2) = f_{X,Y}(2,1) + f_{X,Y}(2,2) = \frac{1}{2}$$





Dependant Variables – Simple Example

- Are X and Y independent?

$$f_{X,Y}(1,1) = \frac{1}{2} \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = f_X(1) \cdot f_Y(1)$$

$$f_{X,Y}(1,2) = 0 \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = f_X(1) \cdot f_Y(2)$$

...

- No, X and Y are not independent!

Words and Concepts to Know

Stochastic

Cumulative Distribution Function

Marginal

Probability Mass Function

Correlation coefficient

Simultanious pmf

cdf

Joint pmf

pmf

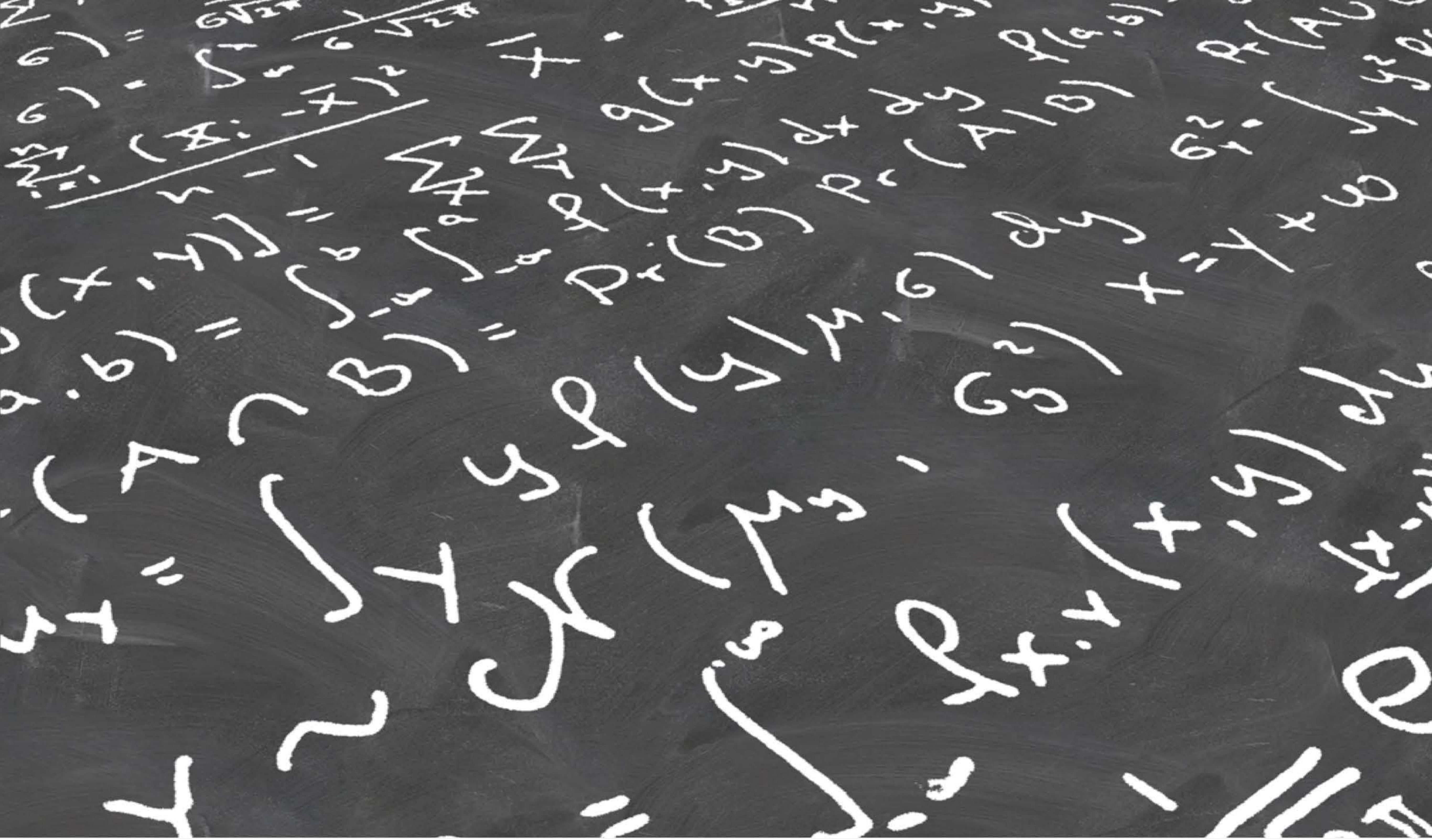
Standard deviation

Binomial Mass Function

Mean

Variance

Expectation



4. Continuous Random Variables

Gunvor Elisabeth Kirkelund
Lars Mandrup

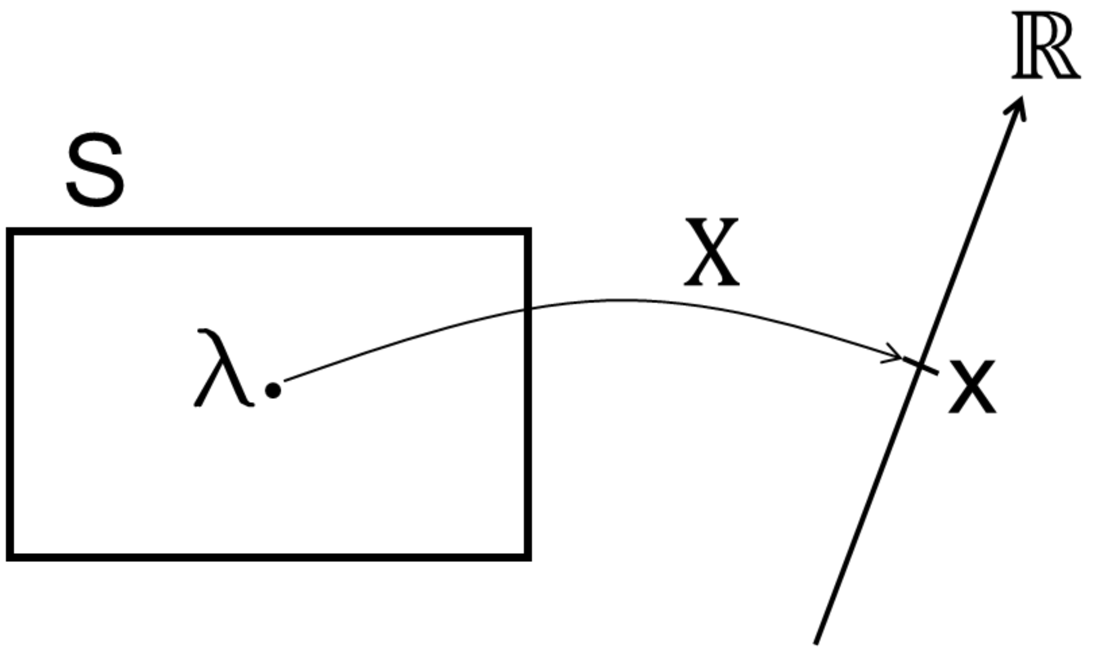
Agenda for Today

- Repetition from last time
    - Discrete Random Variables
- Continuous Random Variables

Also just called a random variables

Stochastic Random Variables

- A random variable tells something important about a stochastic experiment.
- Can be discrete or continous



Examples:

- The numbers on a dice (discrete):
  - Sample space for variable X is : $\{1, 2, 3, 4, 5, 6\}$
  - Sample space for variable Y “Even (1)/Uneven (-1)”: $\{1, -1\}$
- The hight of students at IHA (continous):
  - Sample space for variable H is all real numbers: [100;250] cm.

Probability Mass Function (PMF)

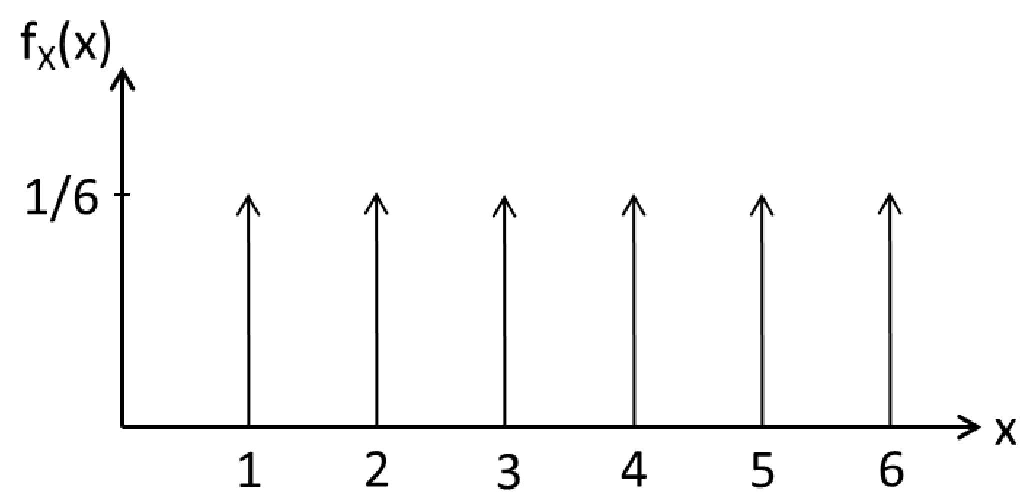
- Sample space for X.
- X is a <u>discreet</u> stochastic variable.

$$\boxed{f_X(x) = \begin{cases} Pr(X = x_i) & for\ X = x_i \\ 0 & otherwise \end{cases}}$$

$$0 \le f_X(x) \le 1$$

- We have that: $\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} Pr(X = x_i) = 1$

Example: Laplace Dice
(perfect dice)



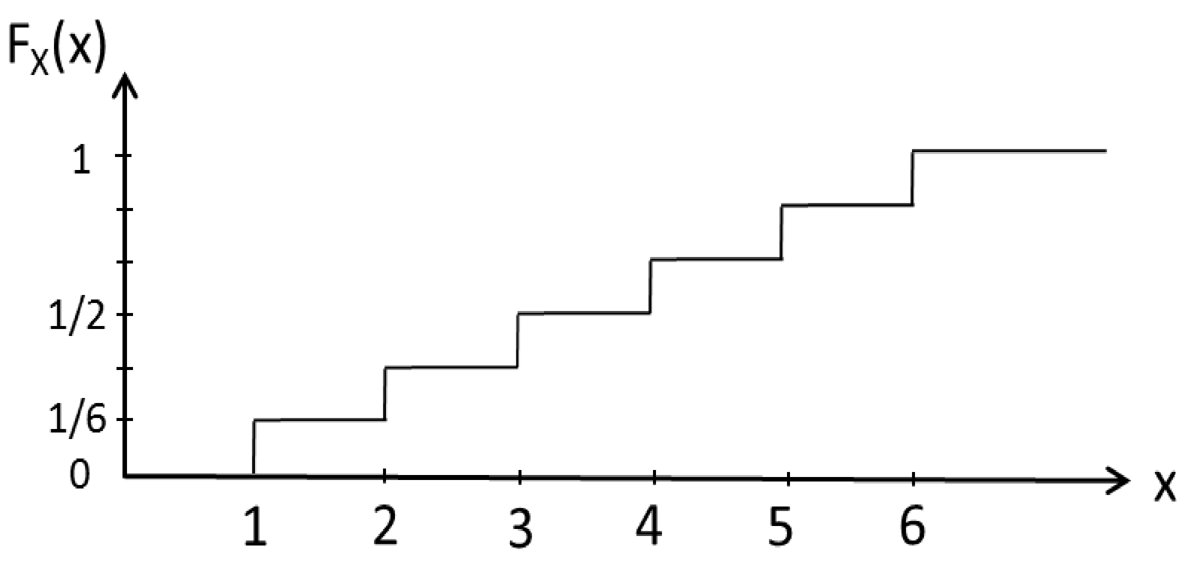
Cumulative Distribution Function (CDF)

- Sample space for X.
- X is a discreet stochastic variable.
- $F_X(x)$ is a non-decreasing step-function.

$$\boxed{F_X(x) = Pr(X \leq x)} \qquad 0 \leq F_X(x) \leq 1$$

- We have that: $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$

Example: Laplace Dice
(perfect dice)



Mean, Variance and Standard deviation

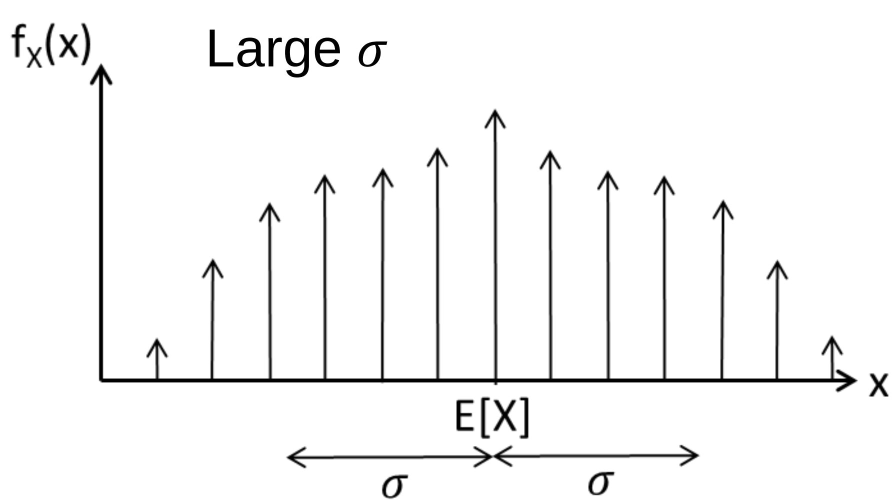
- The <u>mean</u> or the <u>expectation</u> of a discreet random variable X

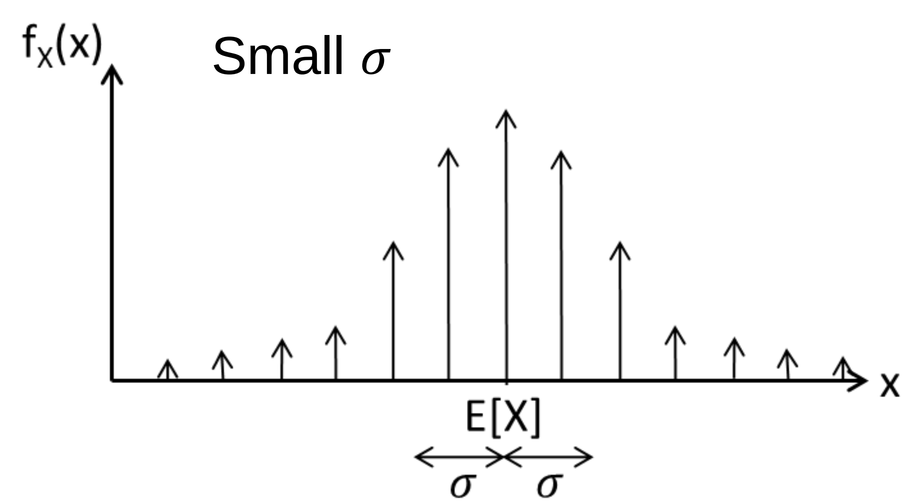
$$\boxed{\bar{X} = E[X] = \sum_{i=1}^{n} x_i\, f_X(x_i)}$$

Variance and standard deviation tells of the spreading of the data

- The <u>variance</u> σ^2 or the <u>standard deviation</u> σ of a random variable X

$$\boxed{Var(X) = \sigma_X^2 = E\left[X^2\right] - E[X]^2}$$

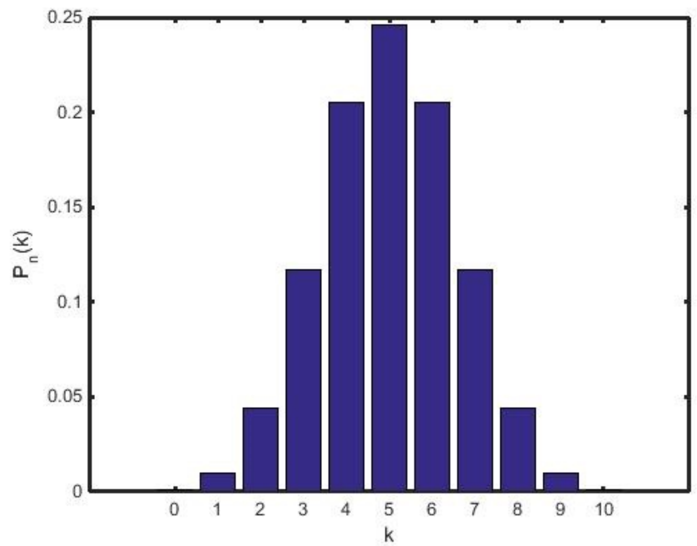




The Binomial Distribution

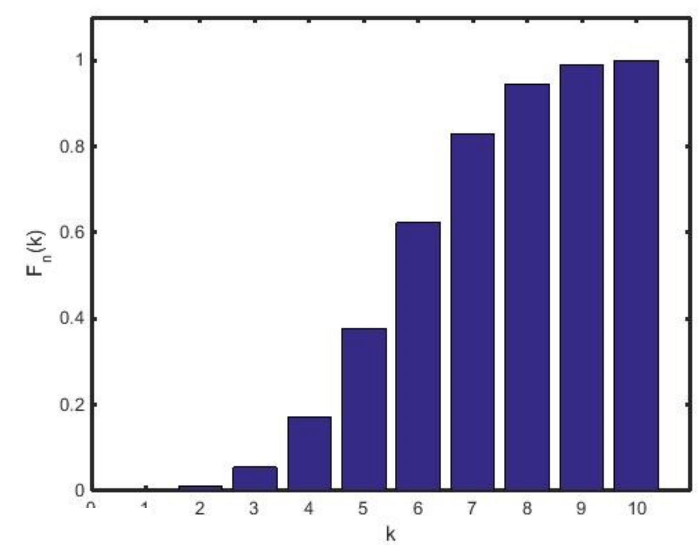
- n repeated trials – each with two possible outcomes
  - **Success** — probability p
  - **Failure** — probability 1-p

Also called a Bernoulli trial



- Probability mass function (pmf):

$$f(k|n,p) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$



- Cumulative distribution function (cdf):

$$F(k|n,p) = \sum_{i=0}^{k} f(i|n,p)$$

- Mean and variance:

$$E[k] = n \cdot p$$
$$Var(X) = n \cdot p \cdot (1-p)$$

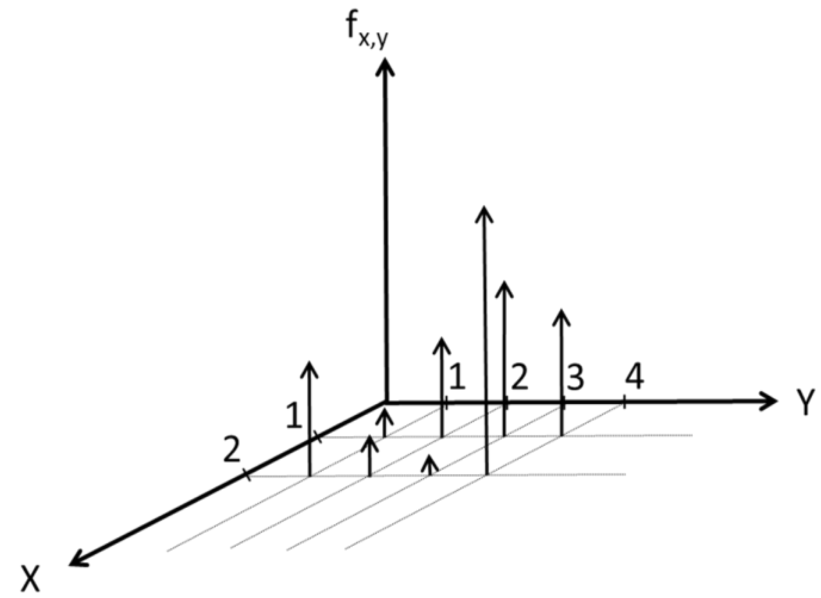
Two Simultaneous Discreet Random Variables

Joint (Simultaneous) pmfs:

$$f_{X,Y}(x,y) = \begin{cases} Pr\Big((X = x_i)\cap(Y = y_j)\Big) & for\ X = x_i \wedge Y = y_j \\ 0 & otherwise \end{cases}$$

Marginal pmfs:

$$f_X(x) = \sum_y f_{X,Y}(x,y) \qquad f_Y(y) = \sum_x f_{X,Y}(x,y)$$



Conditional pmfs / Bayes Rule:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = Pr(X = x|Y = y)$$

Correlation Coefficient

Correlation tells of the coupling between variables

- The correlation coefficient, is an indicator on how much two random variables X and Y are correlated.

$$\boxed{\rho = E\left[\frac{X-\bar{X}}{\sigma_X}\cdot\frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}}$$

- We have that: $-1 \leq \rho \leq 1$

Independence

Independence: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

- Bayes Rule: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

  gives that if X and Y are independent, then:

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

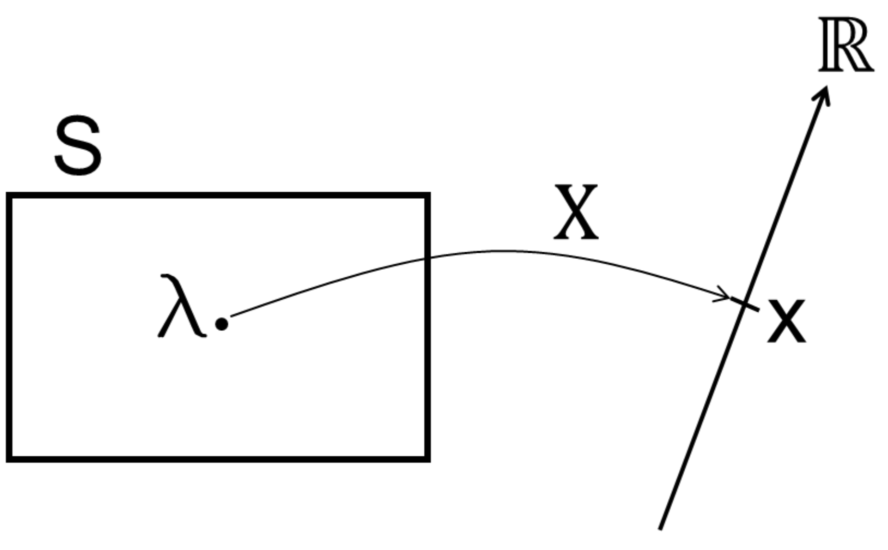
- Also:

  $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \Rightarrow E[XY] = E[X]E[Y] \Rightarrow \ \rho = 0$

  but the opposite is not allways true!

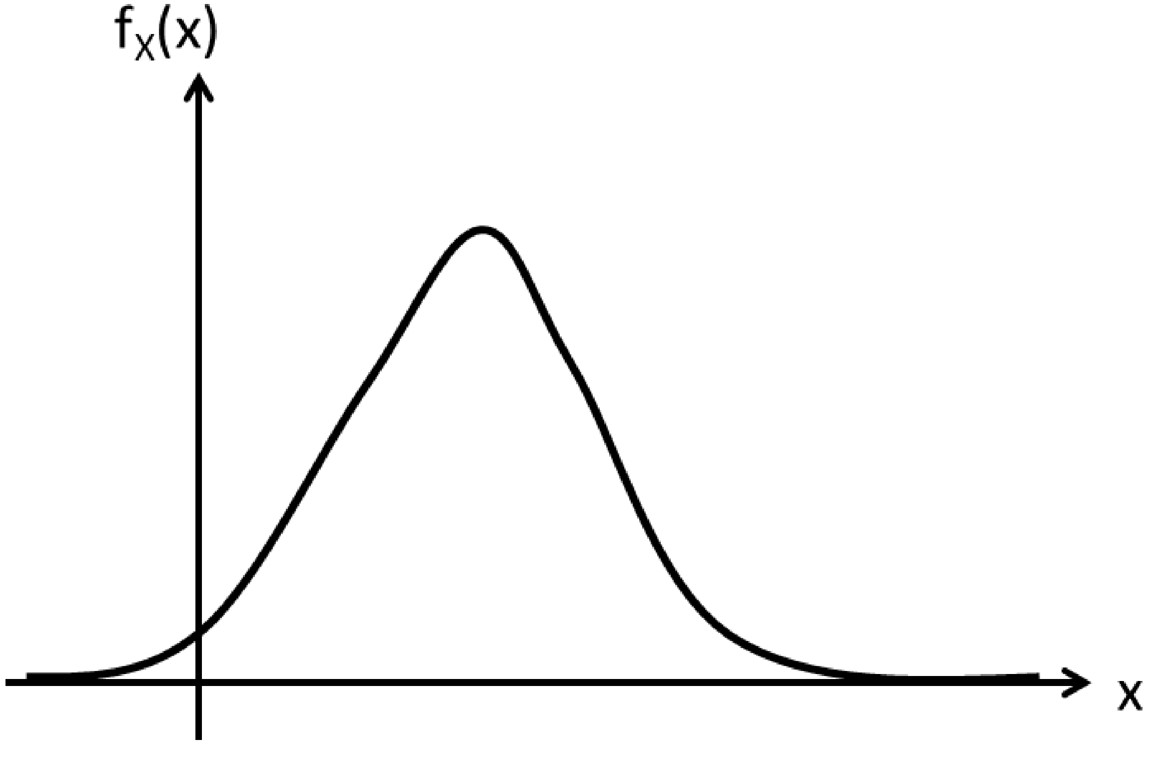
Continuous Random Variables

- We define a stochastic variable X
- X is continuous on $\mathbb{R}$



- Fx. The exact value R of a resistor

- X is defined by a density function $f_X(x)$



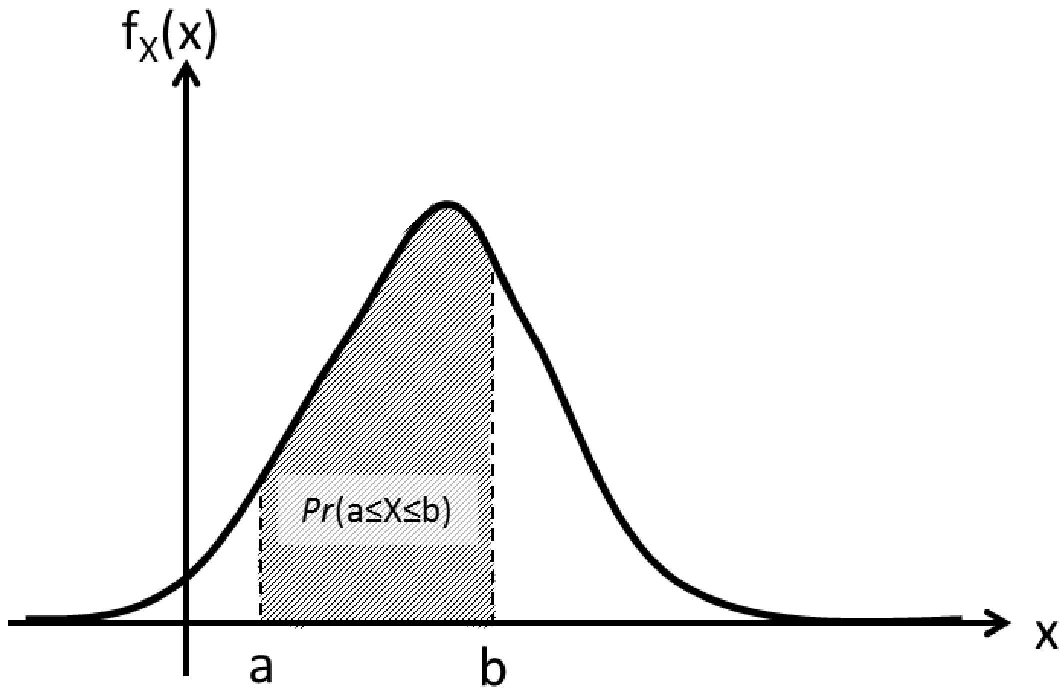
- The probability of one exact value of the variable is always zero:

$$Pr(X = x) = 0$$

Continous Random Variables — PDF

- We define a <u>probability density function</u> (pdf): $f_X(x)$

$$\boxed{Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx}$$



Properties: $f_X(x) \geq 0$

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

Total probability is 1.

Notice: $f_X(x) > 1$ is possible

$$Pr(X = x) = 0$$

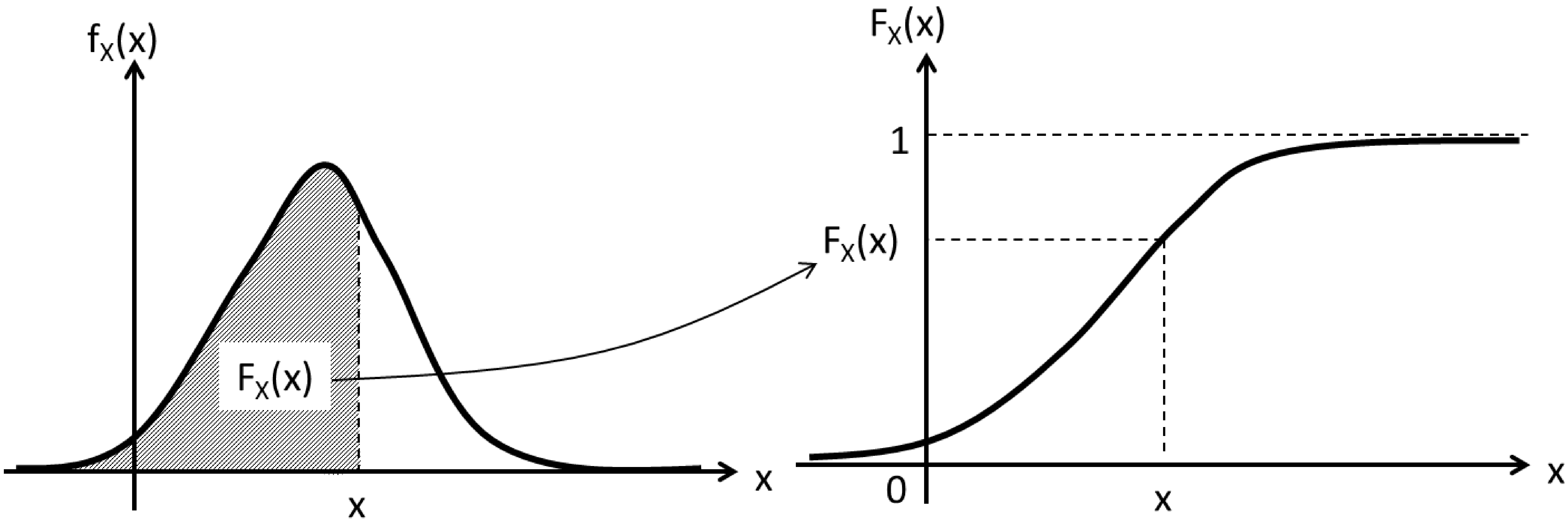
$$Pr(a < X < b) = Pr(a \leq X < b) = Pr(a < X \leq b) = Pr(a \leq X \leq b)$$

Cumulative Distribution Function (CDF)

- We define a <u>cumulative distribution function</u> (cdf): $F_X(x)$

Accumulates the probabilities from minus infinite to x.

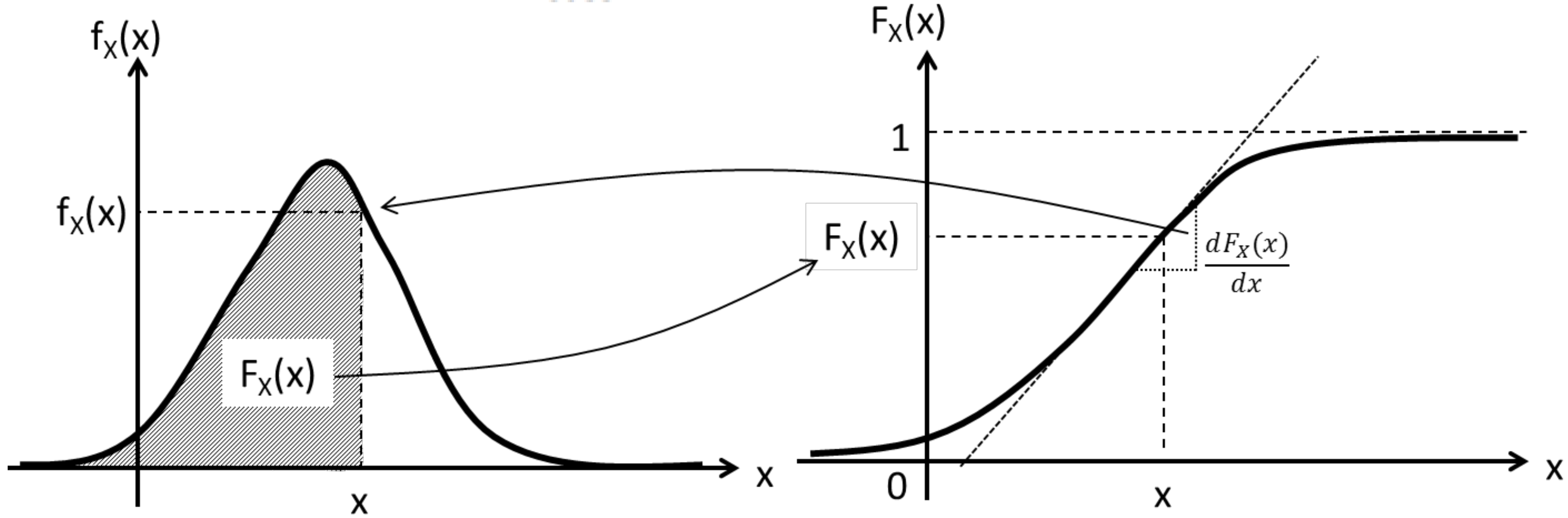
$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = Pr(X \leq x)}$$



The cdf and pdf contains the same information.

Cumulative Distribution Function (CDF)

- From pdf to cdf: $$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = Pr(X \leq x)$$
- From cdf to pdf: $$f_X(x) = \frac{dF_X(x)}{dx}$$



Properties:

- $0 \leq F_X(x) \leq 1$
- $F_X(x)$ is always non−decreasing and continuous
- $Pr(a \leq X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$
- $Pr(X > x) = 1 - Pr(X \leq x) = 1 - F_X(x)$

Definition of Expectation

- We define the expectation of $g(X)$ with respect to a pdf $f_X(x)$ as the integral:

$$\boxed{E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx}$$

Example:

- DC voltage with a noise-signal.

Mean Value

- The mean value is the expectation of X:

$$\boxed{E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)dx}$$

Example:

- The value of 5% 1kΩ resistors.

Expectation

- Linear function: $g(X) = aX + b$

$$E[aX + b\ ] = \int_{-\infty}^{\infty} (ax + b) \cdot f_X\,(x)dx = a \cdot E[X] + b$$

- Square function: $g(X) = X^2$

$$E[g(X)] = E\left[X^2\right] = \int_{-\infty}^{\infty} x^2 \cdot f_X\,(x)dx$$

$$\neq \left(\int_{-\infty}^{\infty} x \cdot f_X\,(x)dx\right)^2 = E[X]^2$$

Definition of Variance

- We define the variance of $g(X)$ with respect to a pdf $f_X(x)$ as the integral:

$$\boxed{\begin{aligned} Var(g(\mathrm{X})) &= \int_{-\infty}^{\infty} \left(g(x) - \overline{g(x)}\right)^2 \cdot f_X(x)dx \\ &= E\left[g(X)^2\right] - E[g(X)]^2 \end{aligned}}$$

- The variance of a continuous random variable X:

$$Var(\mathrm{X}) = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f_X(x)dx = E\left[X^2\right] - E[X]^2$$

Variance

- Linear function: $g(X) = aX + b$

$$Var[aX + b\ ] = E\left[(aX + b)^2\right] - E[aX + b]^2$$

$$= \int_{-\infty}^{\infty} (ax + b)^2 \cdot f_X\,(x)dx - (a \cdot E[X] + b\ )^2$$

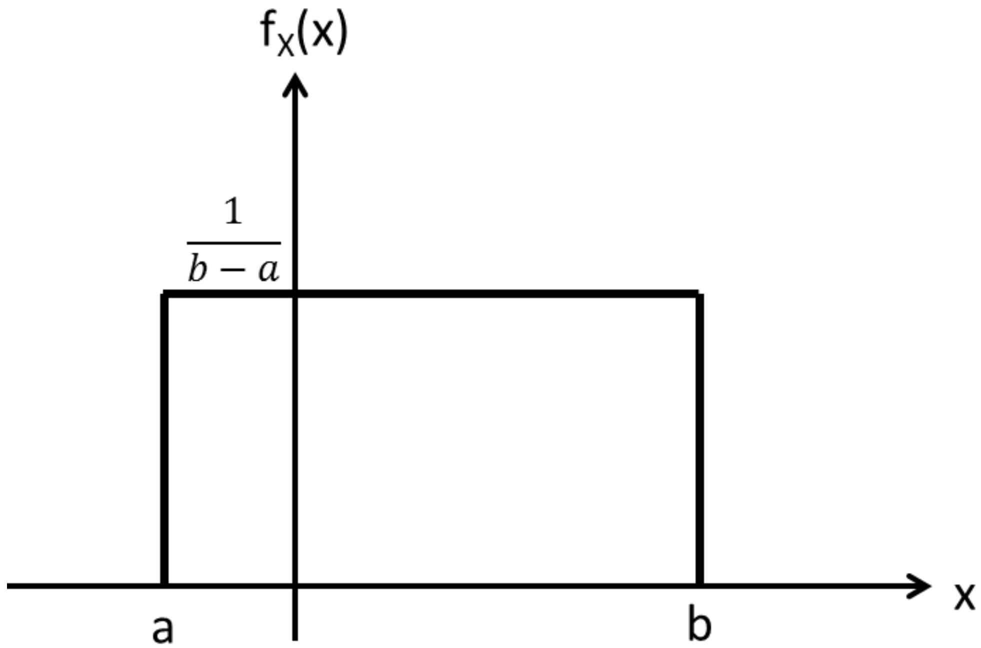
$$= \left(a^2E[X^2] + b^2 + 2abE[X]\right) - \left(a^2E[X]^2 + b^2 + 2abE[X]\right)$$

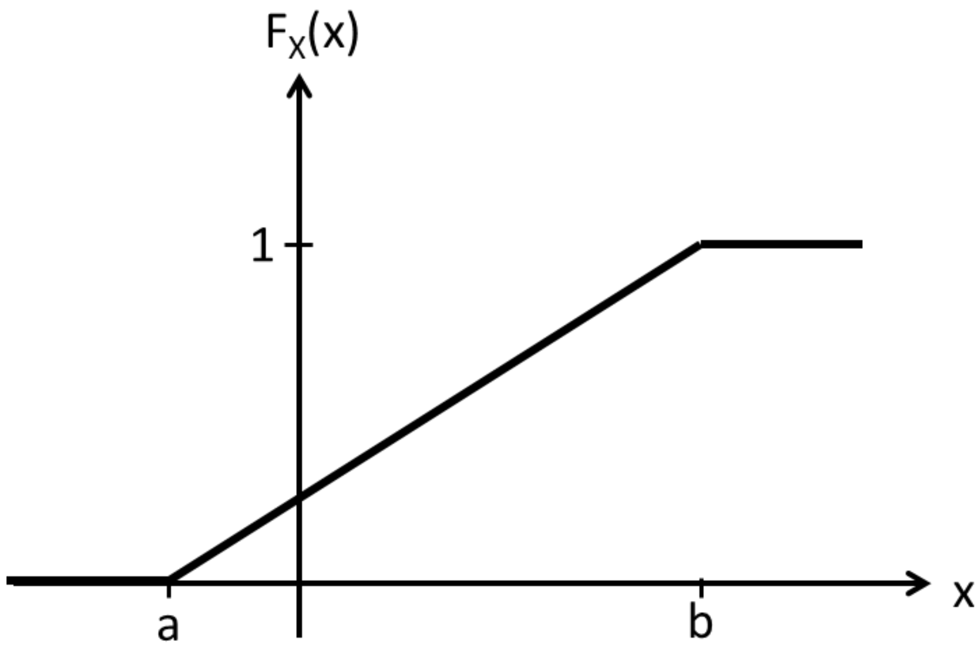
$$= a^2\left(E[X^2] - E[X]^2\right)$$

$$= a^2 \cdot Var(X)$$

Uniform Distribution (continuous)

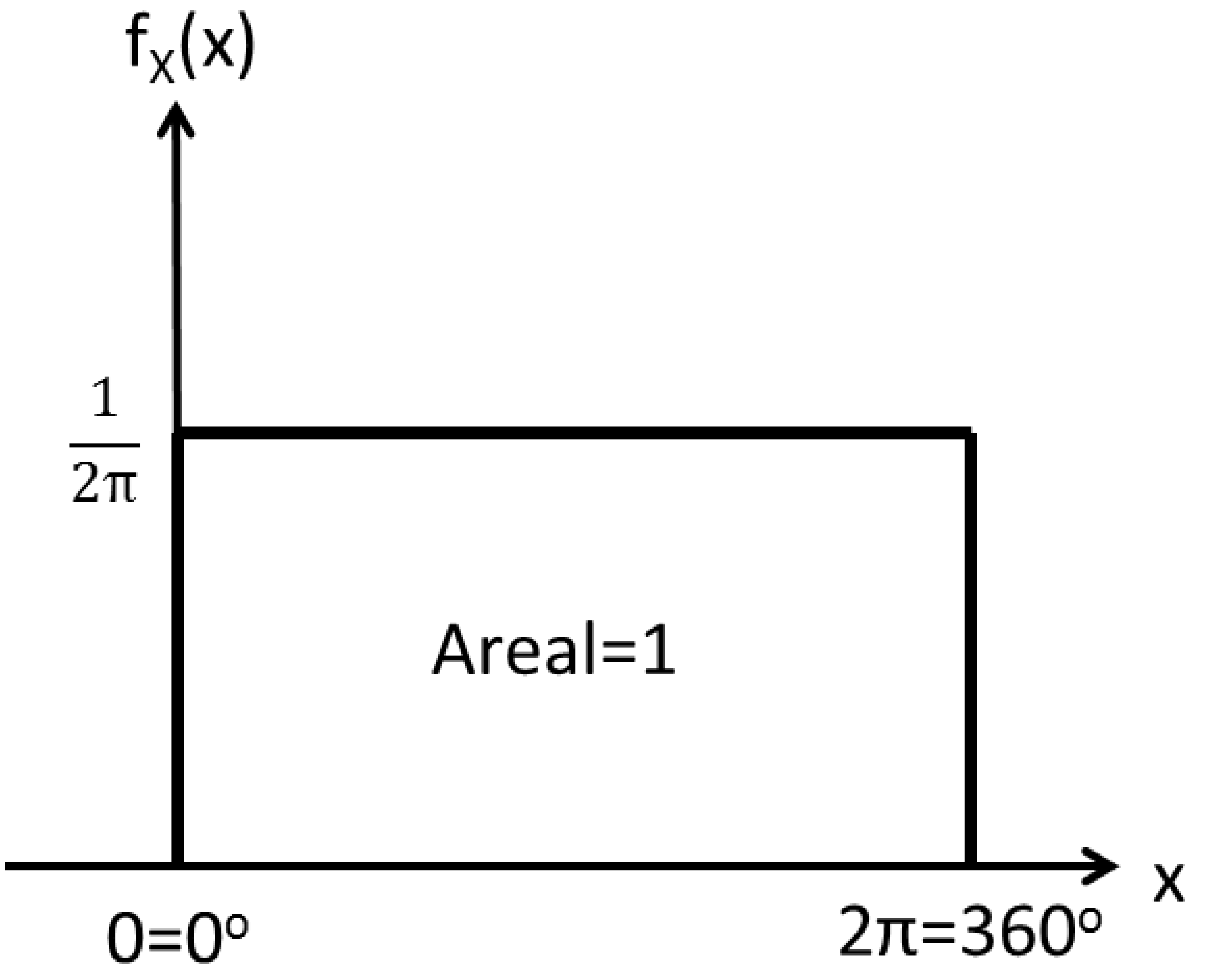
- $\mathcal{U}$(a,b)
- Mean value: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{1}{12}(b-a)^2$
- pdf: $f_X(x) = \begin{cases} \frac{1}{b-a} & for\ a \le x \le b \\ 0 & otherwise \end{cases}$
- cdf: $F_X(x) = \begin{cases} 0 & for\ x \le a \\ \frac{x-a}{b-a} & for\ a \le x \le b \\ 1 & for\ x \ge b \end{cases}$



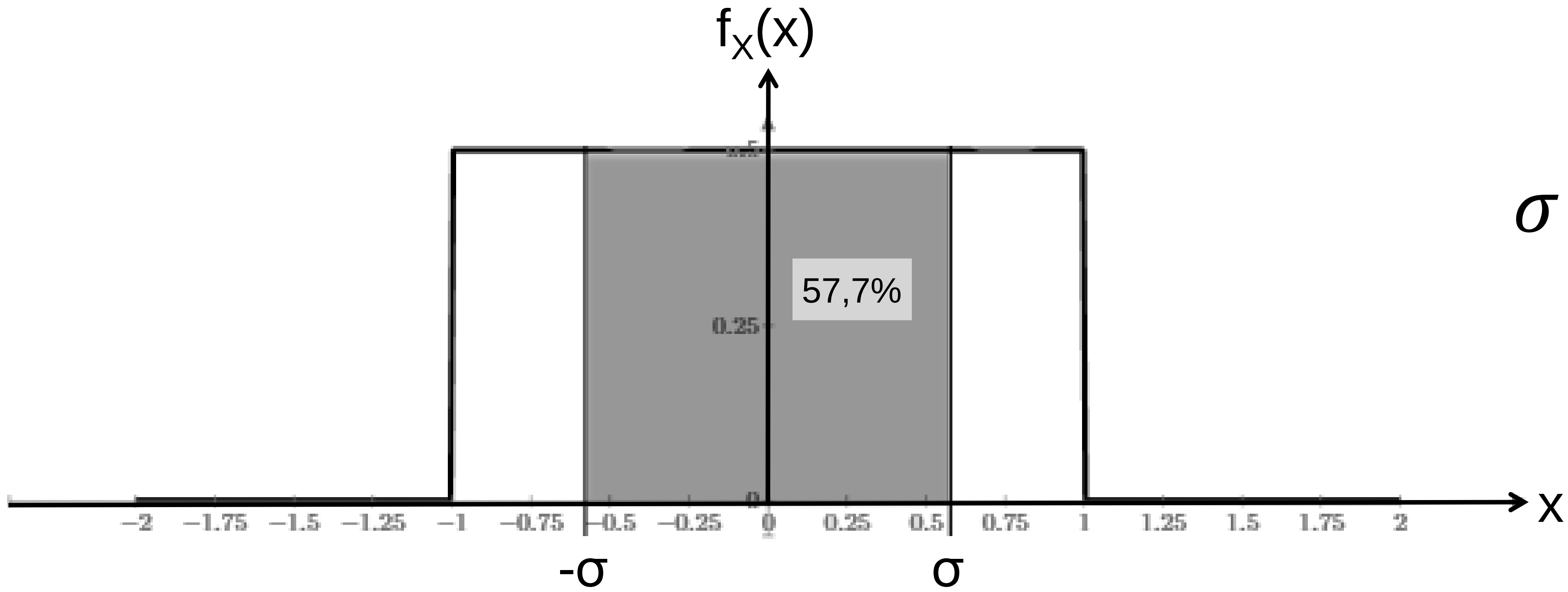


Uniform Distribution — Example

- A phase noise is uniformly distributed.



Uniform Distribution: Standard deviation



$$\sigma = \frac{b-a}{\sqrt{12}}$$

$$\Pr(|X-\mu| \leq \sigma) = 57{,}7\%$$

$$\Pr(|X-\mu| \leq 2\sigma) = 100\%$$

Gaussian Distribution = Normal Distribution

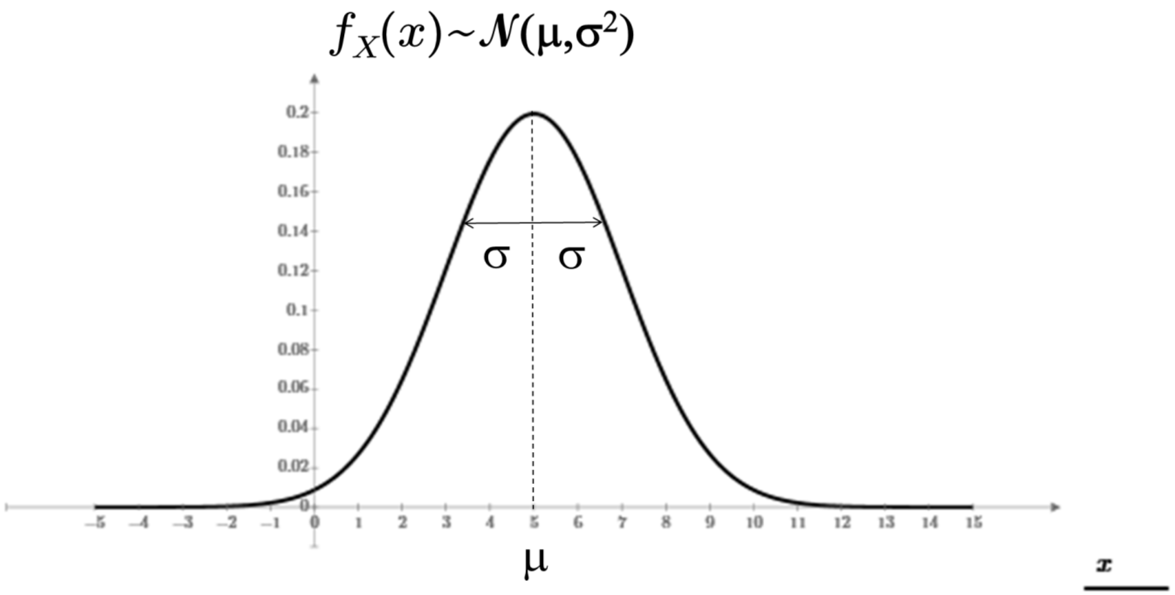
- $\mathcal{N}(\mu,\sigma^2)$
- Mean value: μ
- Variance: σ^2

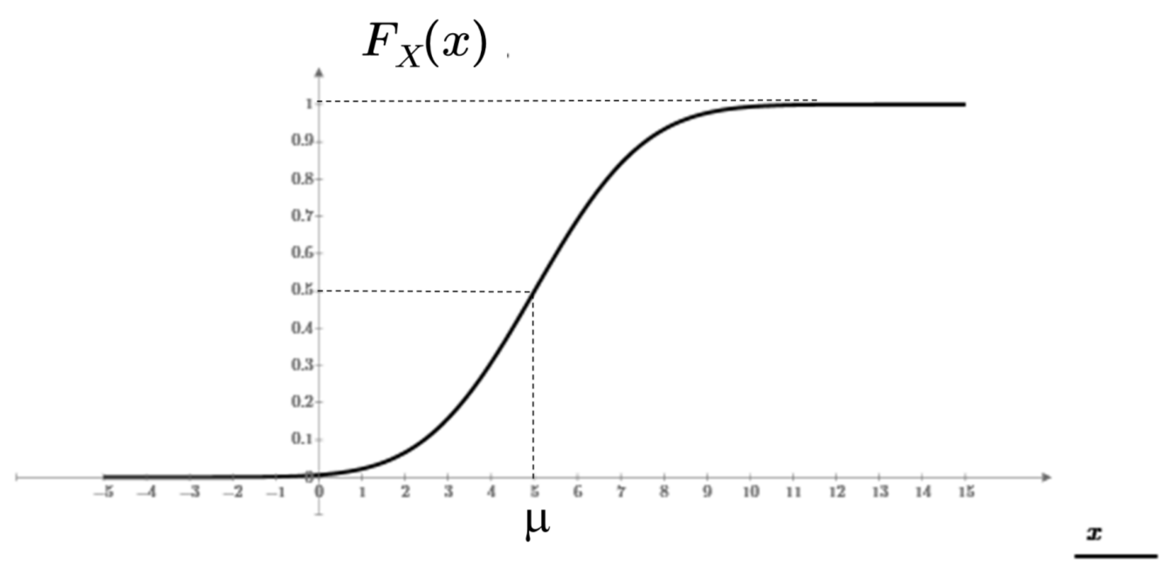
- pdf: $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

- cdf: $F_X(x) = \frac{1}{2}\left[1 + \mathrm{erf}(\frac{x-\mu}{\sigma\sqrt{2}})\right]$

No closed expression for the cdf

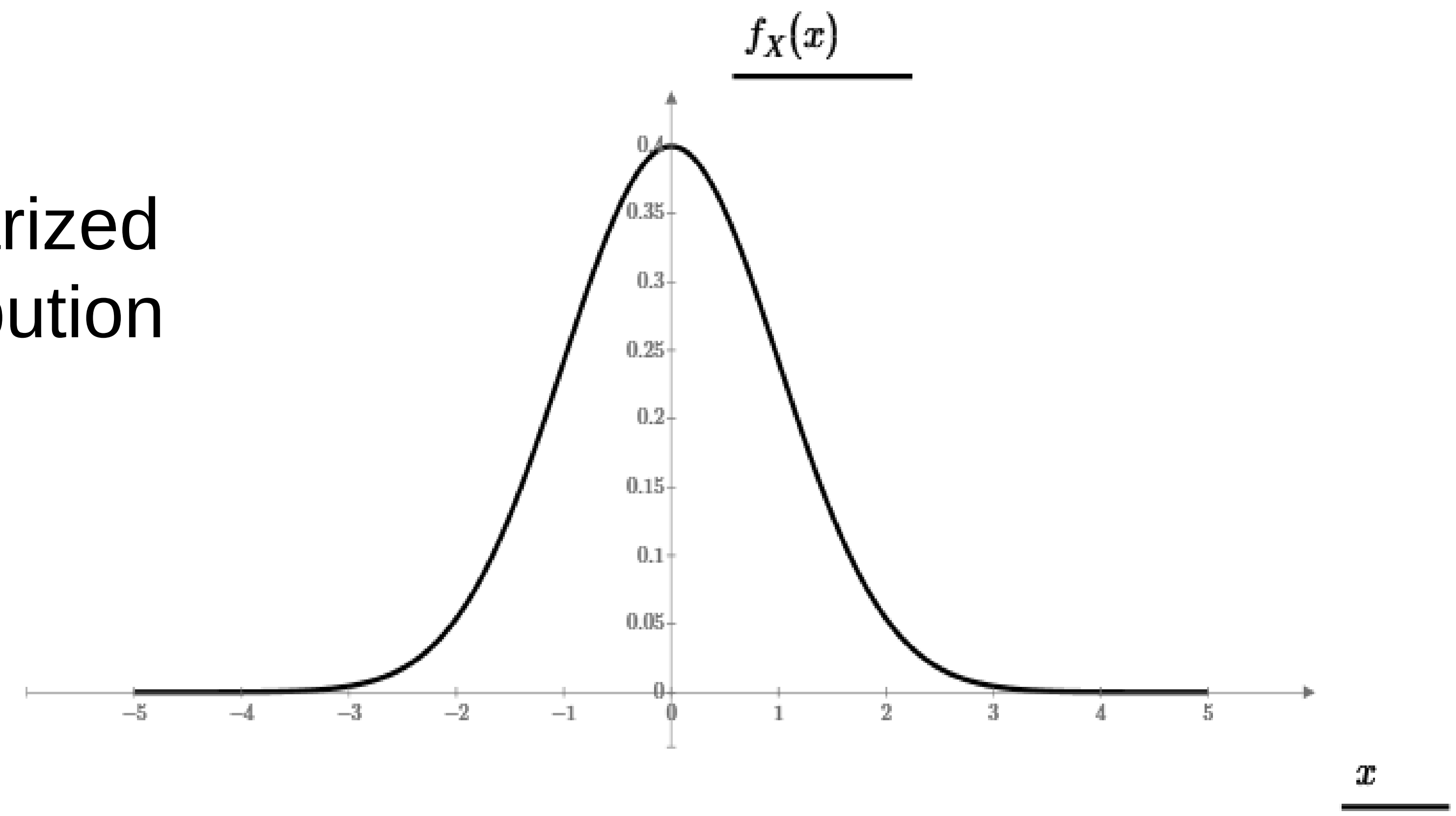
erf = error-function: $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$





Gaussian Distribution = Normal Distribution

$\mathcal{N}(0,1)$
→ the standarized normal distribution

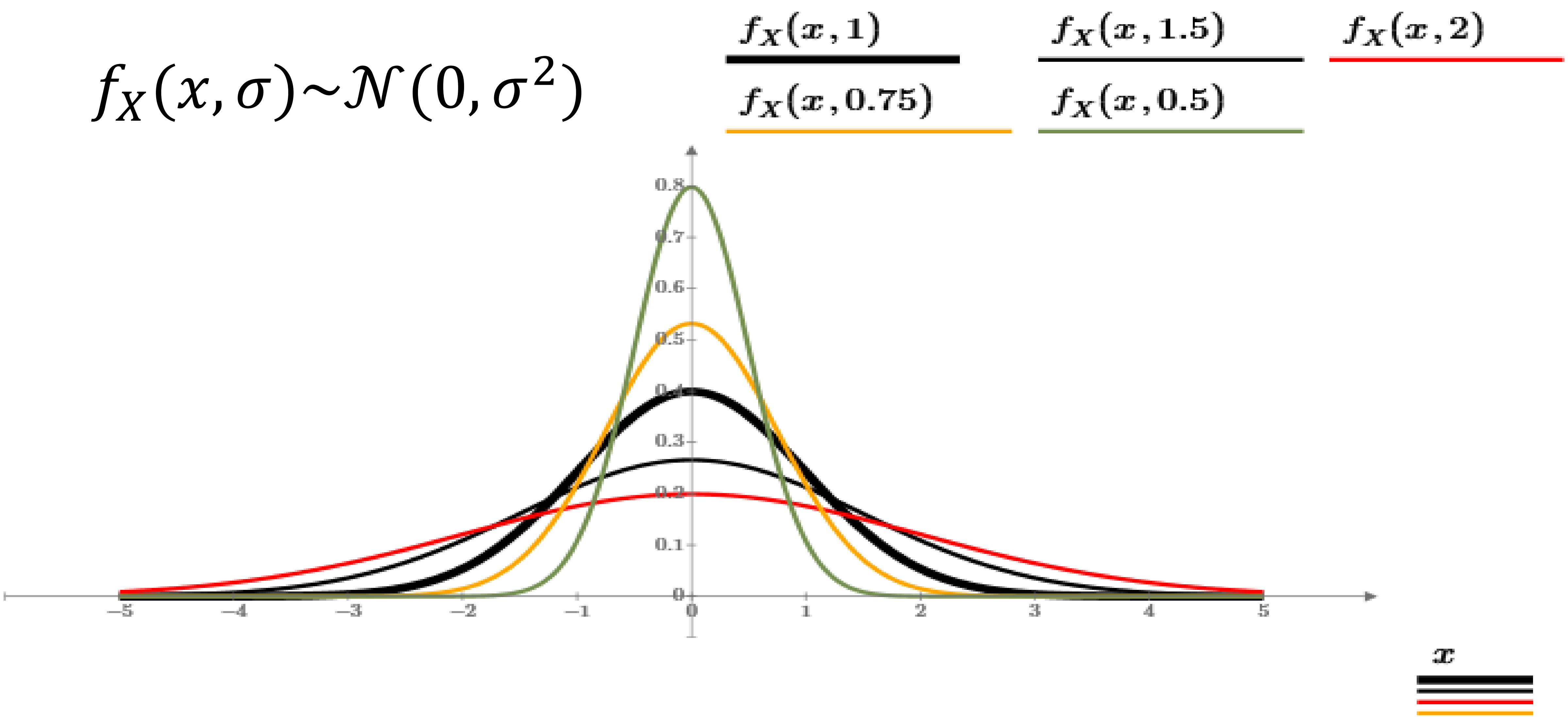


- A lot of things in nature are Gaussian distributed
  - Fx. Examination marks
- Central Limit Theorem → Gaussian distrubution

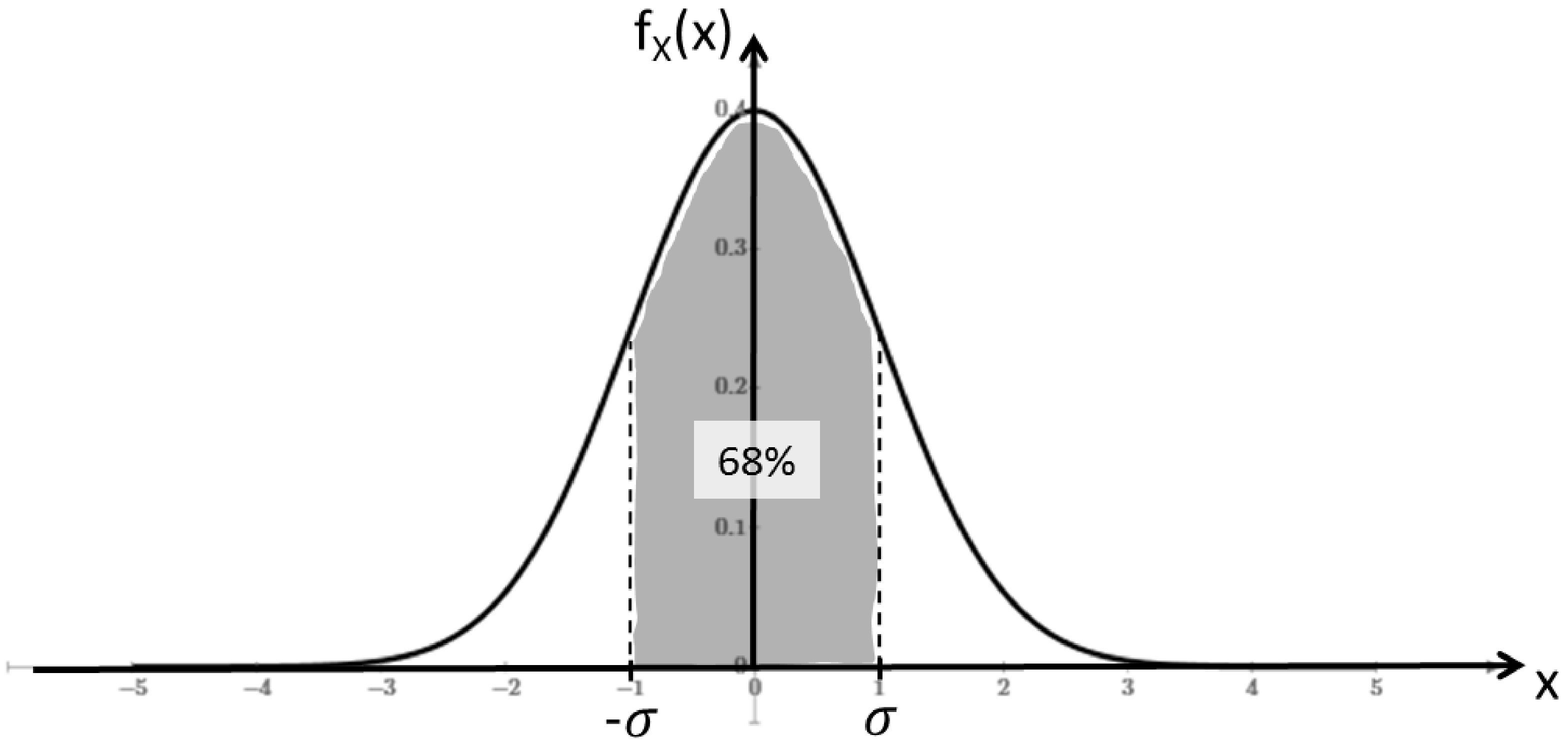
Gaussian Distribution = Normal Distribution

- Maximum probability density at the mean value μ
- The standard deviation (variance) σ determines the form (width and hight)

$$f_X(x, \sigma) \sim \mathcal{N}(0, \sigma^2)$$



Normal Distribution: Standard Deviation



$$\Pr(|X-\mu| \leq \sigma) = 68{,}3\%$$

$$\Pr(|X-\mu| \leq 2\sigma) = 95{,}4\%$$

$$\Pr(|X-\mu| \leq 3\sigma) = 99{,}7\%$$

Gaussian Distribution = Normal Distribution

- Beregninger med normalfordelinger: Tabelopslag og Matlab:
- $X \sim \mathcal{N}(\mu,\sigma^2) \rightarrow Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ (Standard Normal Distribution)
- $F_X(x) = Pr(X \leq x) = Pr\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z(z)$ hvor $z = \frac{x-\mu}{\sigma}$

$$= \begin{cases} \Phi(z) & Tabel\ 1\ (\text{"}Statistik\ og\ Sandsynlighedsregning\text{"}) \\ 1 - \mathrm{Q}(z) & App.D\ (\text{"}Random\ Signals\text{"}) \end{cases}$$

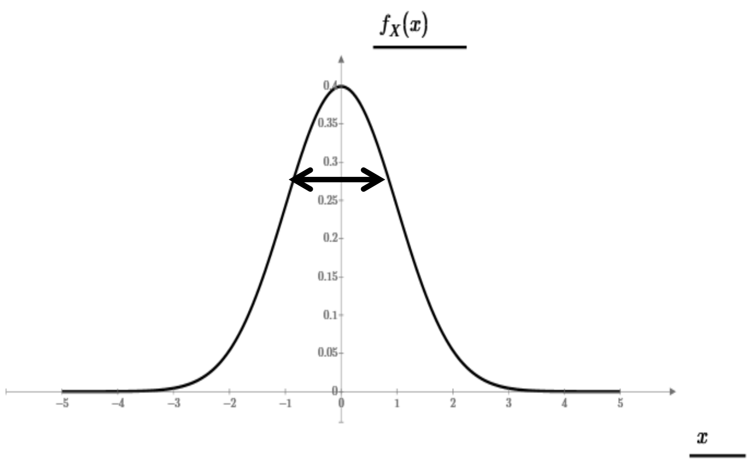
- $\Phi(z) = Pr(Z \leq z)$
- $\Phi(-z) = 1 - \Phi(z)$
- $\mathrm{Q}(z) = Pr(Z \geq z) = 1 - Pr(Z \leq z) = 1 - \Phi(z)$
- $\mathrm{Q}(-z) = 1 - \mathrm{Q}(z)$

Obs: Standard variation

- Matlab:
  - $\Pr(X \leq x) = F_X(x) = normcdf(x,\mu,\sigma)$
  - $\Pr(Z \leq z) = F_Z(z) = normcdf(z,0,1) = normcdf(z)$

Summary of Expectations

- Mean value: $E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)dx \quad (\sum_{i=1}^{n} x_i\, f_X(x_i))$
- Mean square: $E\left[X^2\right] = \overline{X^2} = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx \quad (\sum_{i=1}^{n} x_i^2\, f_X(x_i))$
- Variance: $Var(\mathrm{X}) = {\sigma_X}^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 \cdot f_X(x)dx = E\left[X^2\right] - E[X]^2$
- Standard deviation: $\sigma_X = \sqrt{Var(\mathrm{X})}$



- A function: $E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx \quad (\sum_{i=1}^{n} g(x_i)\, f_X(x_i))$

$$Var\big(g(\mathrm{X})\big) = \int_{-\infty}^{\infty}\left(g(x) - \overline{g(x)}\right)^2 \cdot f_X(x)dx = E\left[g(X)^2\right] - E[g(X)]^2$$

- Linear function: $E[aX + b\ ] = a \cdot E[X] + b$

$$Var[aX + b\ ] = a^2\left(E[X^2] - E[X]^2\right) = a^2 \cdot Var(X)$$

Two Stochastic Variables X,Y

- The simultaneous (joint) density function
- The marginal probability density function
- Bayes rule

- Discreet $\rightarrow$ Continous stochastic random variable

$$\sum \rightarrow \int$$

Continuous Random Variables

- We have a simultaneous (joint) pdf: $f_{X,Y}(x,y)$
- We have the probability:

$$Pr\big((a \le X \le b) \cap (c \le Y \le d)\big) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

- We have for the pdf: $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

$$0 \le f_{X,Y}(x,y)$$

The Marginal PDF

- For a two dimensional pdf $f_{X,Y}(x,y)$, we can find the marginals

Marginals:

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

Relationship between pdf and cdf

- For a two dimensional pdf $f_{X,Y}(x,y)$, the cdf and the pdf correspond to each other

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

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

The Conditional PDF

- For a two dimensional pdf $f_{X,Y}(x,y)$, we can find the conditional pdf with Bayes rule:

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

Independence:

- X and Y are independent if:

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

Correlation

Correlation tells of the (biased) coupling between variables

- Correlation:

$$corr(X,Y) = E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x,y)dxdy$$

➢ If X and Y are independent: $E[XY] = E[X] \cdot E[Y]$

➢ If $X = Y$: $corr(X,X) = E[X^2]$

Covariance

Covariance is without bias from the mean

- Covariance:

$$cov(X,Y) = E\big[(X-\bar{X})(Y-\bar{Y})\big]$$

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

$$= E[XY] - E[X]\cdot E[Y] = corr(X,Y) - E[X]\cdot E[Y]$$

- If X and Y are independent: $corr(X,Y) = 0$

OBS: The opposite not always true

- If $X = Y$: $cov(X,X) = E[X^2] - E[X]^2 = Var(X)$

Correlation Coefficient

Correlation Coefficient is the normalized Covariance

- The correlation coefficient, is an indicator on how much two random variables X and Y are correlated.

$$\rho = E\left[\frac{X-\bar{X}}{\sigma_X} \cdot \frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X \cdot \sigma_Y} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

  - We have that: $-1 \leq \rho \leq 1$
  - If X and Y are independent: $\rho = 0$

Dependence

- We have independence between X and Y if and only if:

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

Example of independent random variables:

- A persons height and the current exact distance from the earth to the moon.

Example of dependent random variables:

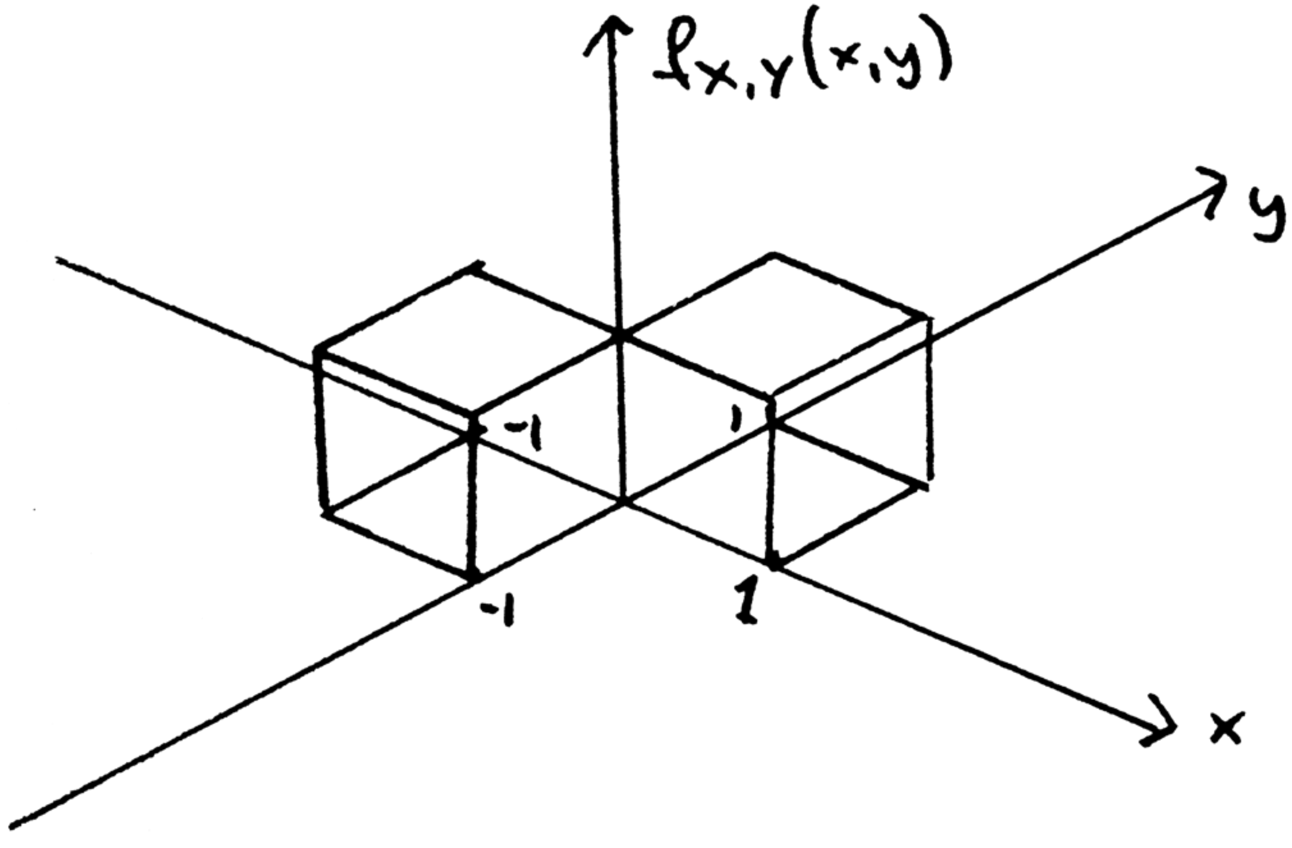
- The time of day and the amount of bicycles parked the at the engineering college.
- The energy of a mobile signal and the length in meters to a basestation.

Dependance - Example

- We want to find out whether two random variables are independent:

Simultaneous pdf for X and Y:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 0 \text{ and } -1 \leq y < 0 \\ \frac{1}{2} & \text{for } 0 \leq x < 1 \text{ and } 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$



Dependance - Example

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 0 \text{ and } -1 \leq y < 0 \\ \frac{1}{2} & \text{for } 0 \leq x < 1 \text{ and } 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find marginals:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \\ &= \begin{cases} \int_{-1}^{0} \frac{1}{2}\, dy & \text{for } -1 \leq x < 0 \\ \int_{0}^{1} \frac{1}{2}\, dy & \text{for } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 0 \\ \frac{1}{2} & \text{for } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \\ &= \begin{cases} \int_{-1}^{0} \frac{1}{2}\, dx & \text{for } -1 \leq y < 0 \\ \int_{0}^{1} \frac{1}{2}\, dx & \text{for } 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{1}{2} & \text{for } -1 \leq y < 0 \\ \frac{1}{2} & \text{for } 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

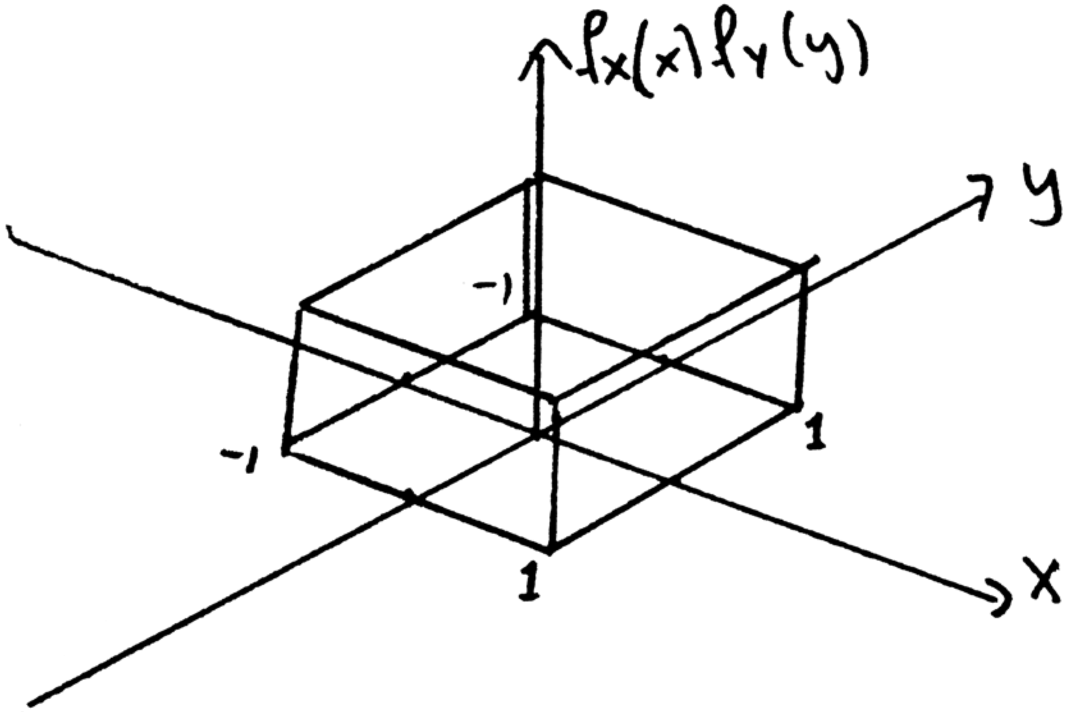
Dependance - Example

- Independence if and only if: $f_{X,Y} = f_X(x) f_Y(y)$

Multiply marginals:

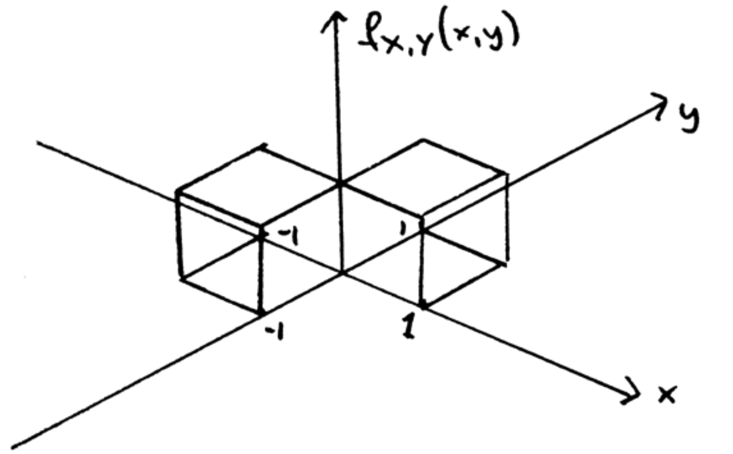
$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) f_Y(y) = \begin{cases} \frac{1}{4} & \text{for } -1 \leq x < 1 \text{ and } -1 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$

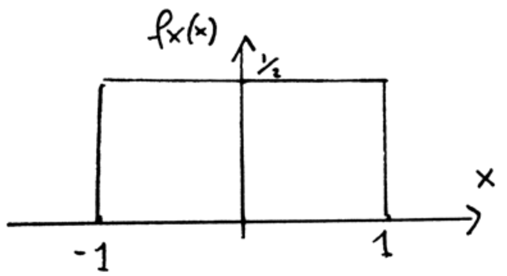


Dependance - Example

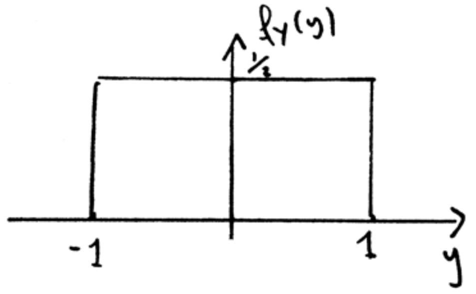
$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 0 \text{ and } -1 \leq y < 0 \\ \frac{1}{2} & \text{for } 0 \leq x < 1 \text{ and } 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$



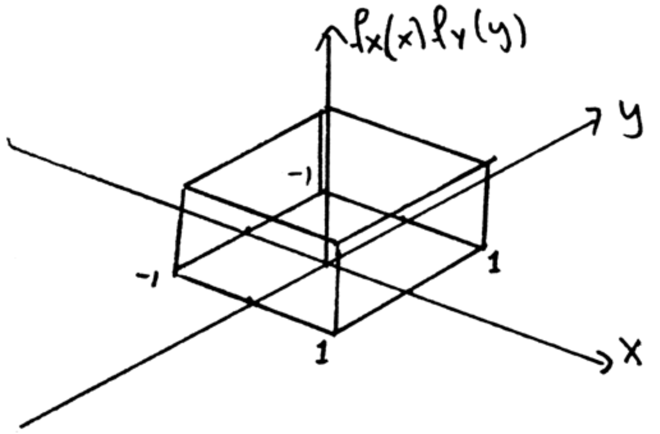
$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$



$$f_Y(y) = \begin{cases} \frac{1}{2} & \text{for } -1 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$



$$f_X(x)f_Y(y) = \begin{cases} \frac{1}{4} & \text{for } -1 \leq x < 1 \text{ and } -1 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$



$f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y) \Rightarrow X$ and Y er <u>ikke</u> uafhængige

Assignment:
Verify the results by doing the detailed calculations

Correlation calculation

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x, y) dx dy = \frac{1}{4}$$

$$E[X] = E[Y] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_{-\infty}^{\infty} y \cdot f_Y(y) dx = 0$$

$$\sigma_X^2 = \sigma_Y^2 = E\left[X^2\right] - E[X]^2 = E\left[Y^2\right] - E[Y]^2 = \frac{1}{3}$$

$$corr(X, Y) = E[XY] = \frac{1}{4}$$

$$cov(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{4} - 0 \cdot 0 = \frac{1}{4}$$

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{1/3} = \frac{3}{4} = 0{,}75$$

Very important!

i.i.d.: Independent and Identically distributed

- We define that for series of random variables that is taken from the same distribution (identically distributed), and are sampled independent of each other, that they are i.i.d.

$$\boxed{\text{i.i.d. = Independent and Identically distributed}}$$

- i.i.d. is a very important characteristic in stochastic variable processing and statistics

Example:

- Quantisation noise.

Words and Concepts to Know

Probability density function

i.i.d.

Correlation

Marginal probability density function

Continuous random variable

Gaussian distribution

Uniform distribution

pdf

Independent and Identically Distributed

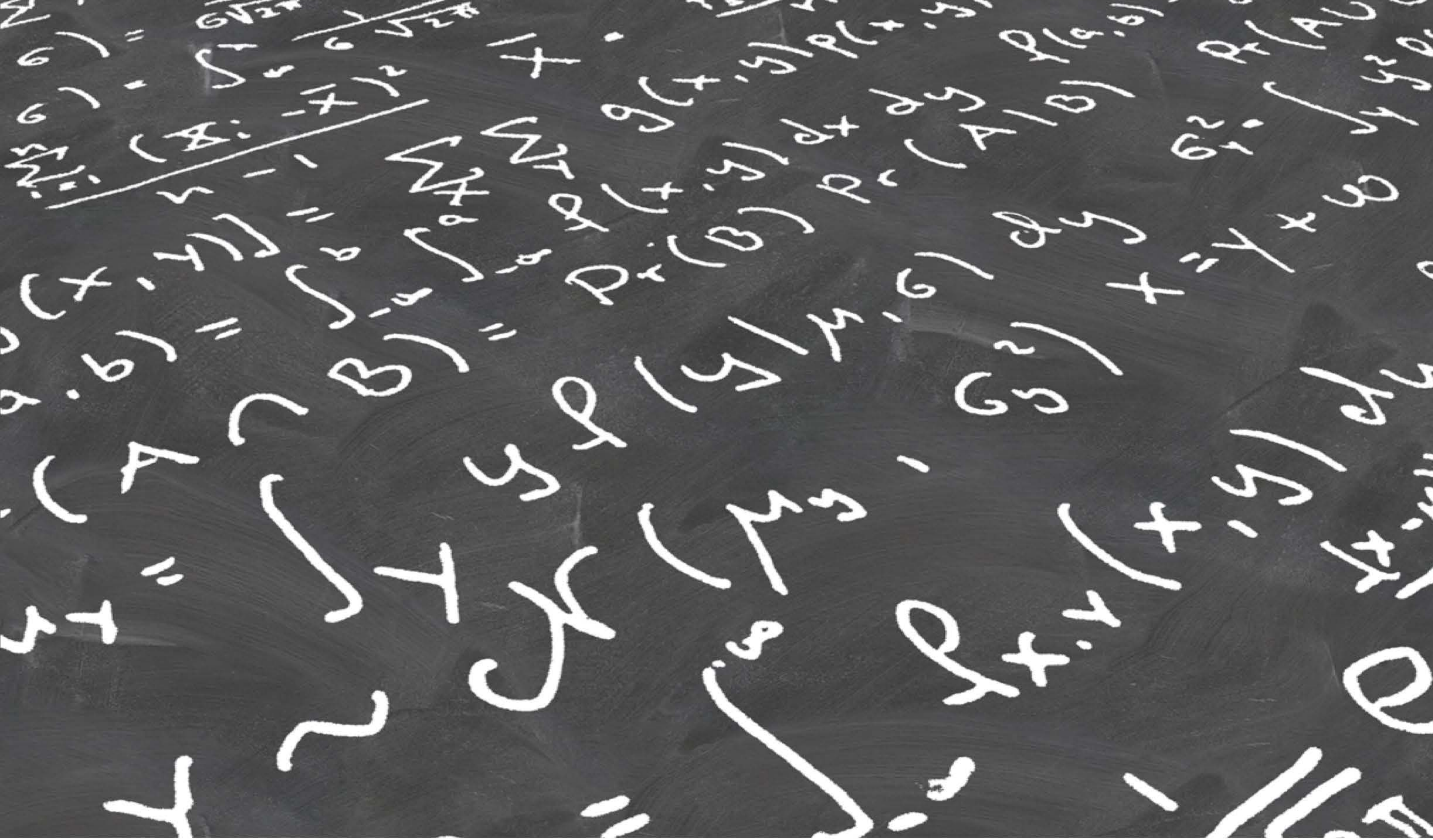
Normal distribution

Correlation coefficient

Simultaneous density function

Covariance

Joint density function



5. Transformations and Multivariate Random Variables

Gunvor Elisabeth Kirkelund
Lars Mandrup

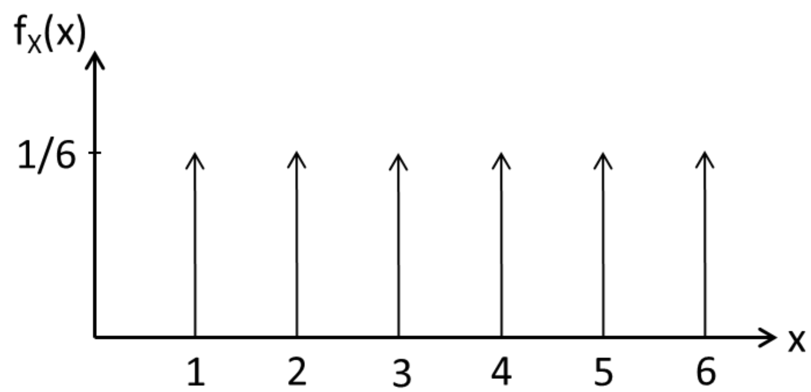
Agenda for Today

- Repetition:
  - One Random Variable
  - Two Random Variables
- Data sampling for test and simulation
- Transformation of random variables
- Sum of two random variables
- Central limit theorem (CLT)

One Stochastic Variable – Discrete

- Probability mass function (pmf):

$$f_X(x) = \begin{cases} Pr(X = x_i) & for\ X = x_i \\ 0 & otherwise \end{cases}$$

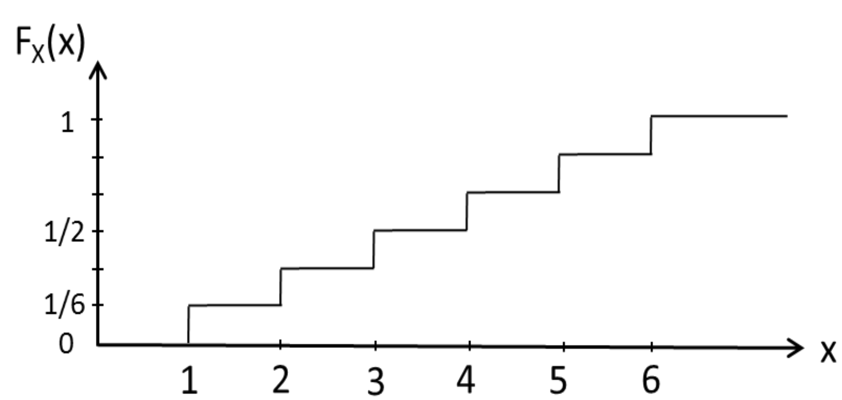


$$0 \leq f_X(x) \leq 1$$

$$\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} Pr(X = x_i) = 1$$

- Cumulative distribution function (cdf):

$$F_X(x) = P\,r(X \leq x) = \sum_{i=1}^{n_x} f_X(x_i)$$



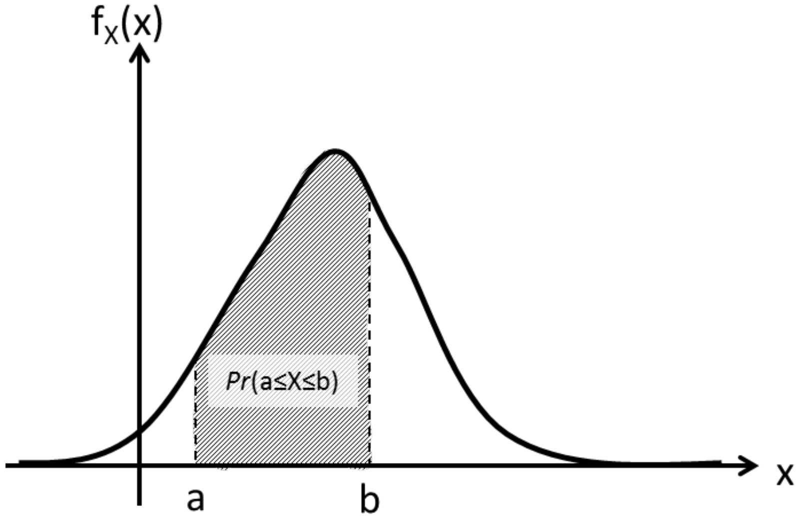
$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

One Stochastic Variable – Continuous

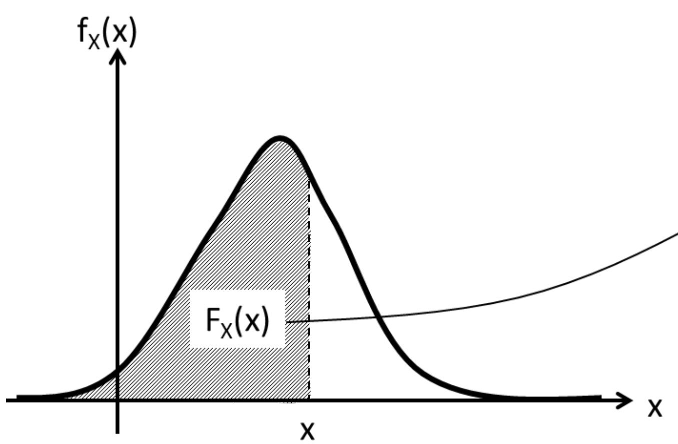
- Probability density function (pdf): $$Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

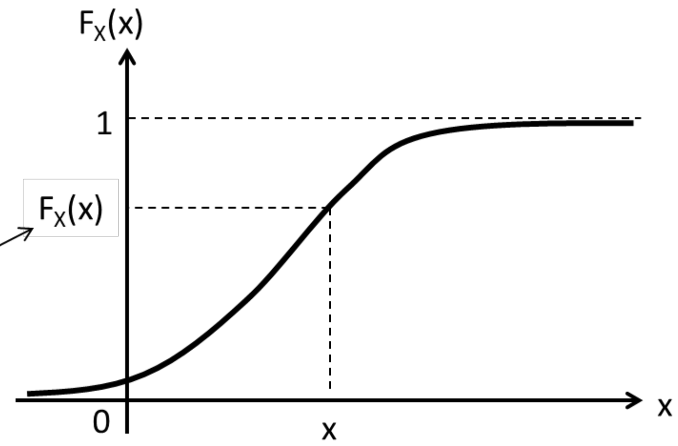


$$f_X(x) \geq 0$$

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

- Cumulative distribution function (cdf): $$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = Pr(X \leq x)$$





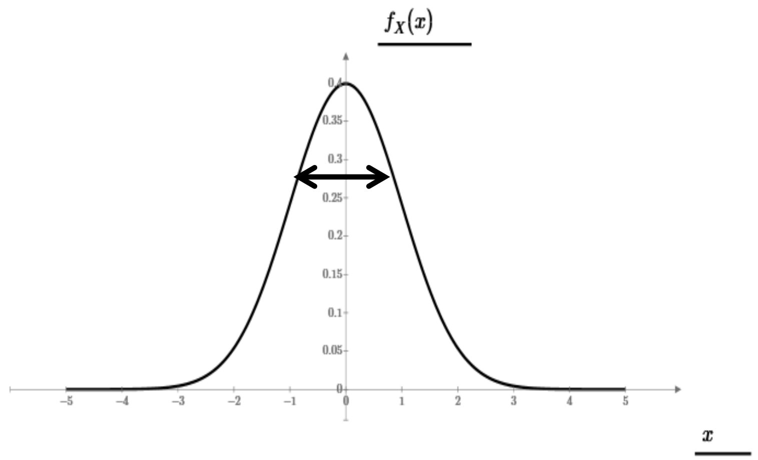
$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

Expectations

- Mean value: $E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X\,(x)dx \quad (\sum_{i=1}^{n} x_i\, f_X(x_i))$
- Variance: $Var(\mathrm{X}) = {\sigma_X}^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f_X\,(x)dx = E\left[X^2\right] - E[X]^2$



- Standard deviation: $\sigma_X = \sqrt{Var(\mathrm{X})}$

- A function: $E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) \cdot f_X\,(x)dx \quad (\sum_{i=1}^{n} g(x_i)\, f_X(x_i))$

$$Var\big(g(\mathrm{X})\big) = \int_{-\infty}^{\infty} \left(g(x) - \overline{g(x)}\right)^2 \cdot f_X\,(x)dx = E\left[g(X)^2\right] - E[g(X)]^2$$

- Linear function: $E[aX + b\,] = a \cdot E[X] + b$

$$Var[aX + b\,] = a^2\left(E[X^2] - E[X]^2\right) = a^2 \cdot Var(X)$$

Two Random Variables X, Y

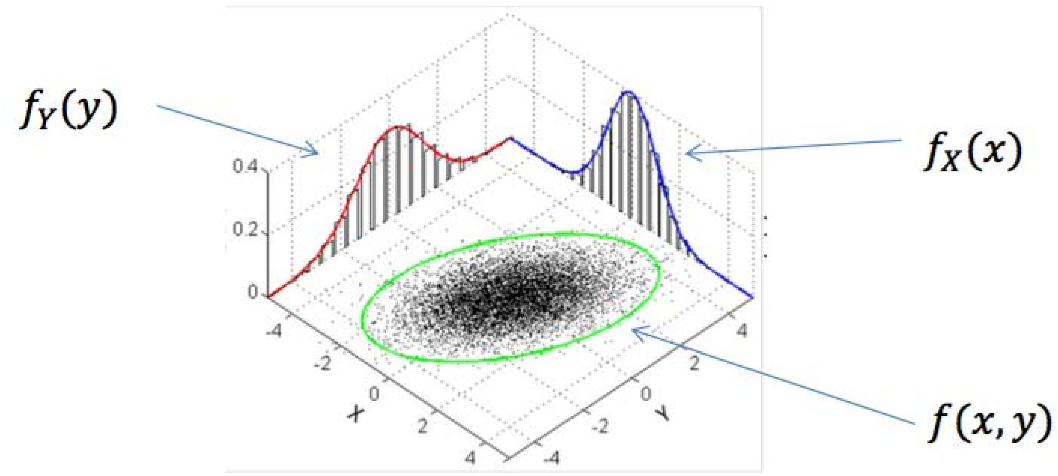
Joint (Simultaneous) pdf: $f_{X,Y}(x,y) \geq 0$ $\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

$$Pr\big((a \leq X \leq b) \cap (c \leq Y \leq d)\big) = \int_{c}^{d}\int_{a}^{b} f_{X,Y}(x,y)dxdy$$

Marginals:

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

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



Cumulative Distribution Function cdf:

cdf $\quad F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy = Pr(X \leq x \wedge Y \leq y)$

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

Bayes Rule, Conditional PDF and Independence

Bayes rule:

- The joint/simultaneous pmf/pdf for two stochastic variables:

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

Conditional pdf:

- For a two dimensional pmf/pdf $f_{X,Y}(x,y)$, we can find the conditional pdf with Bayes rule:

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

Independence:

- X and Y are independent if and only if:

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ or $f_{X|Y}(x|y) = f_X(x)$ for all x and y

Correlation and Covariance

Correlation tells of the (biased) coupling between variables

- Correlation: $corr(X,Y) = E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x,y)dxdy$

Covariance is without bias from the mean

- Covariance: $cov(X,Y) = E\left[(X-\bar{X})(Y-\bar{Y})\right] = E[XY] - E[X] \cdot E[Y]$

Correlation Coefficient is the normalized Covariance

- Correlation coefficient: $\rho = E\left[\frac{X-\bar{X}}{\sigma_X} \cdot \frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY] - E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$

$$-1 \leq \rho \leq 1$$

- If X and Y are independent:

$E[XY] = E[X] \cdot E[Y]$ and $cov(X,Y) = \rho = 0$

Very important!

i.i.d.: Independent and Identically distributed

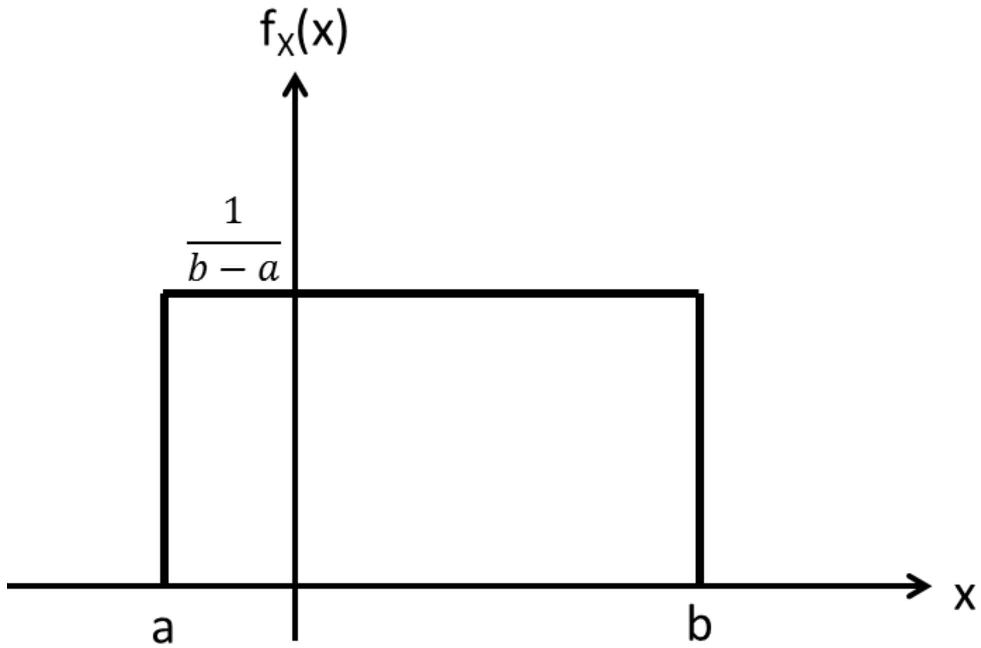
- We define that for series of random variables that is taken from the same distribution (identically distributed), and are sampled independent of each other, that they are i.i.d.

$$\boxed{\text{i.i.d.} = \text{Independent and Identically distributed}}$$

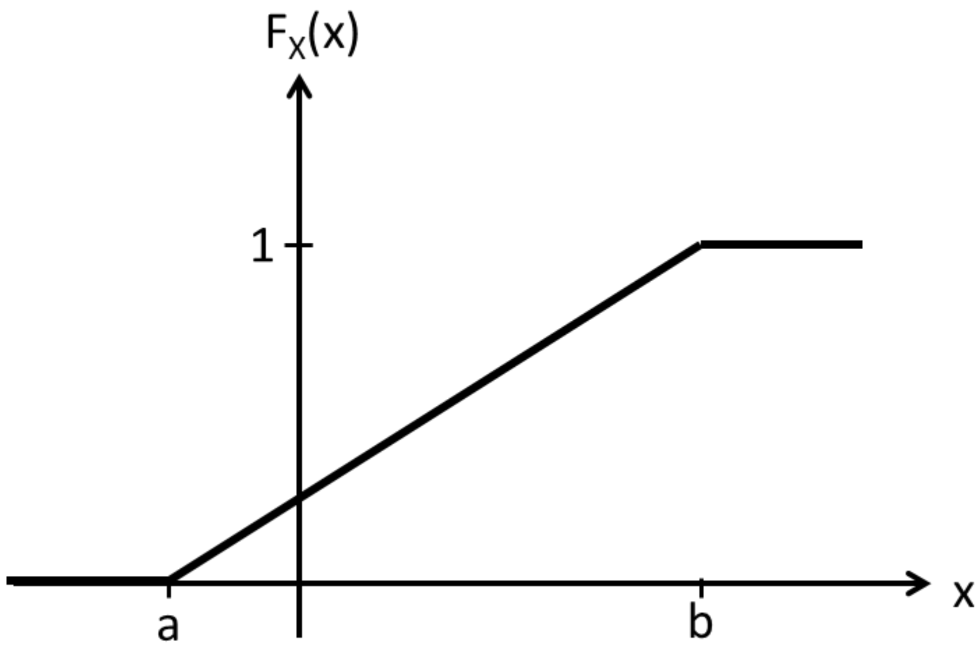
- i.i.d. is a very important characteristic in stochastic variable processing and statistics

Uniform Distribution

- $\mathcal{U}$(a,b)   (Matlab: *rand*)
- Mean value: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{1}{12}(b-a)^2$



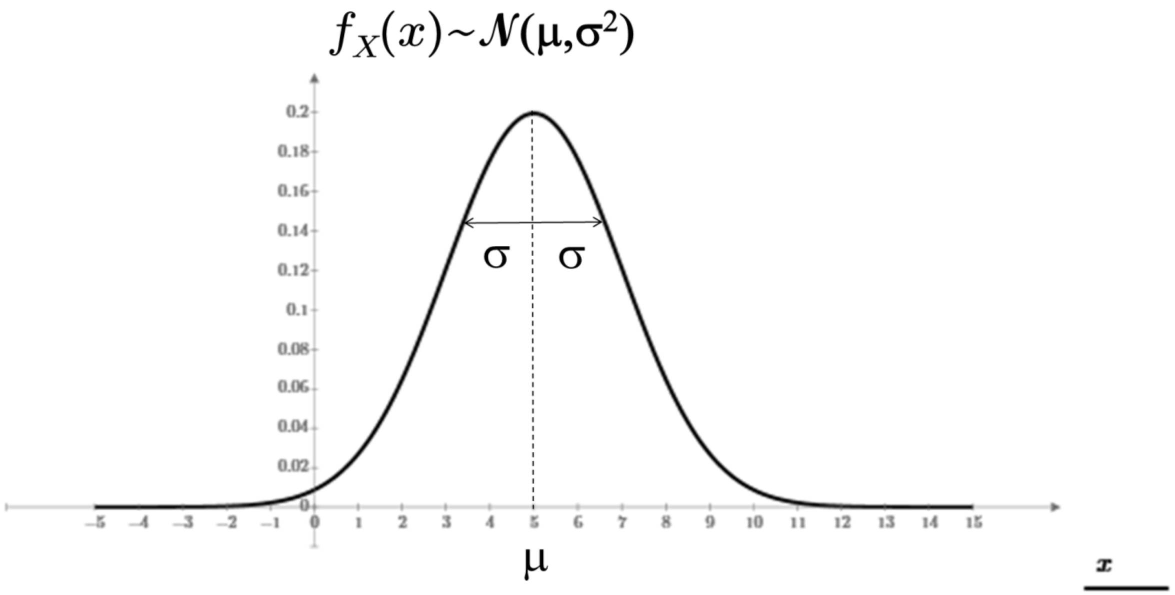
- pdf: $f_X(x) = \begin{cases} \frac{1}{b-a} & for\ a \le x \le b \\ 0 & otherwise \end{cases}$



- cdf: $F_X(x) = \begin{cases} 0 & for\ x \le a \\ \frac{x-a}{b-a} & for\ a \le x \le b \\ 1 & for\ x \ge b \end{cases}$

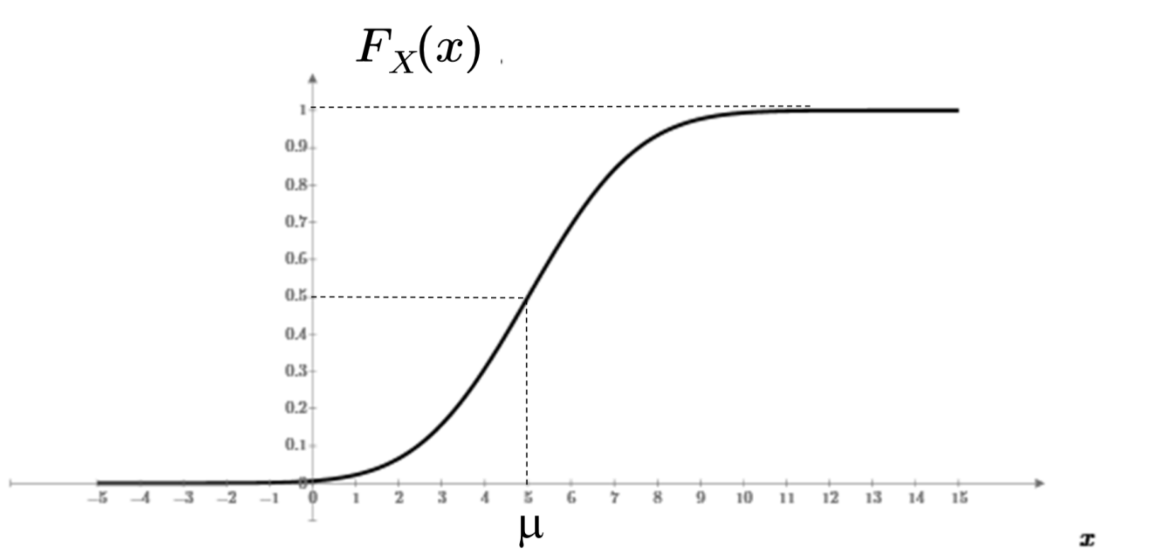
Gaussian Distribution = Normal Distribution

- $\mathcal{N}(\mu, \sigma^2)$
- Mean value: μ
- Variance: σ^2



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

- cdf: $F_X(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$



No closed expression for the cdf

erf = error-function: $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$

Gaussian Distribution = Normal Distribution

- Beregninger med normalfordelinger: Tabelopslag og Matlab:
- $X \sim \mathcal{N}(\mu, \sigma^2) \;\rightarrow\; Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ (Standard Normal Distribution)
- $F_X(x) = Pr(X \leq x) = Pr\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z(z) = \Phi(z)$ hvor $\mathrm{z} = \frac{x-\mu}{\sigma}$

$Tabel\ 1\ ("Statistik\ og\ Sandsynlighedsregning")$

- $\Phi(z) = Pr(Z \leq z)$
- $\Phi(-z) = 1 - \Phi(z) = Pr(Z \geq z) = Pr(Z \leq -z)$

Symmetry of Gaussian distribution

- Matlab:

Obs: Standard deviation

  - $\Pr(X \leq x) = F_X(x) = normcdf(x, \mu, \sigma)$
  - $\Pr(Z \leq z) = F_Z(z) = normcdf(z, 0, 1) = normcdf(z)$

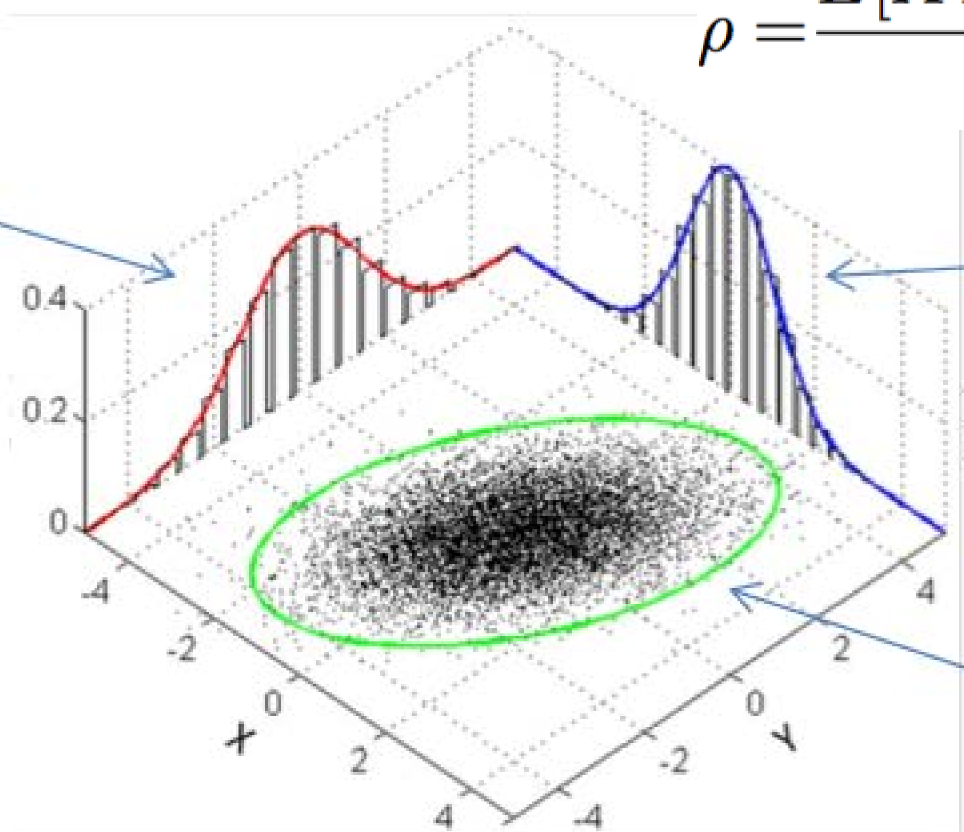
Bivariate (2D) Normal Distribution

Two dimensional Gaussian $$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{z}{2(1-\rho^2)}\right)$$

$$z = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}$$

$$\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X\sigma_Y}$$ Correlation coefficient

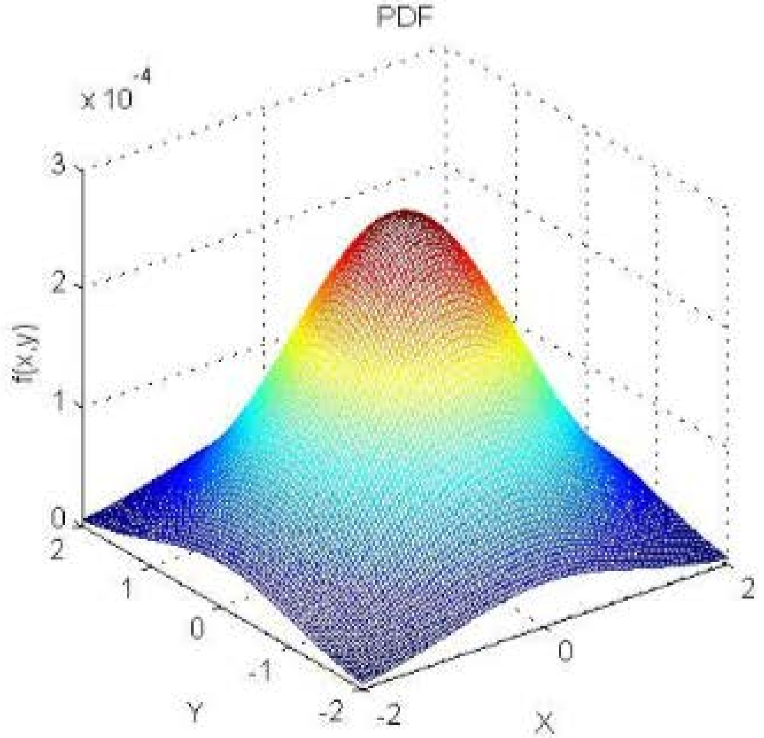
$f_Y(y)$



$f_X(x)$

$f(x,y)$

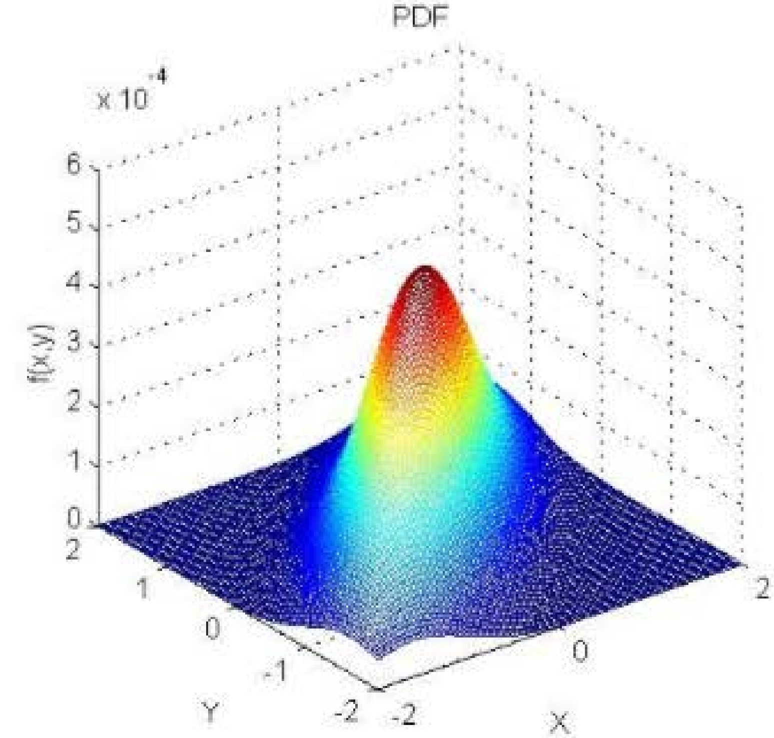
Bivariate Normal Distribution



Symmetric PDF:

$$\rho = 0$$

X and Y independent



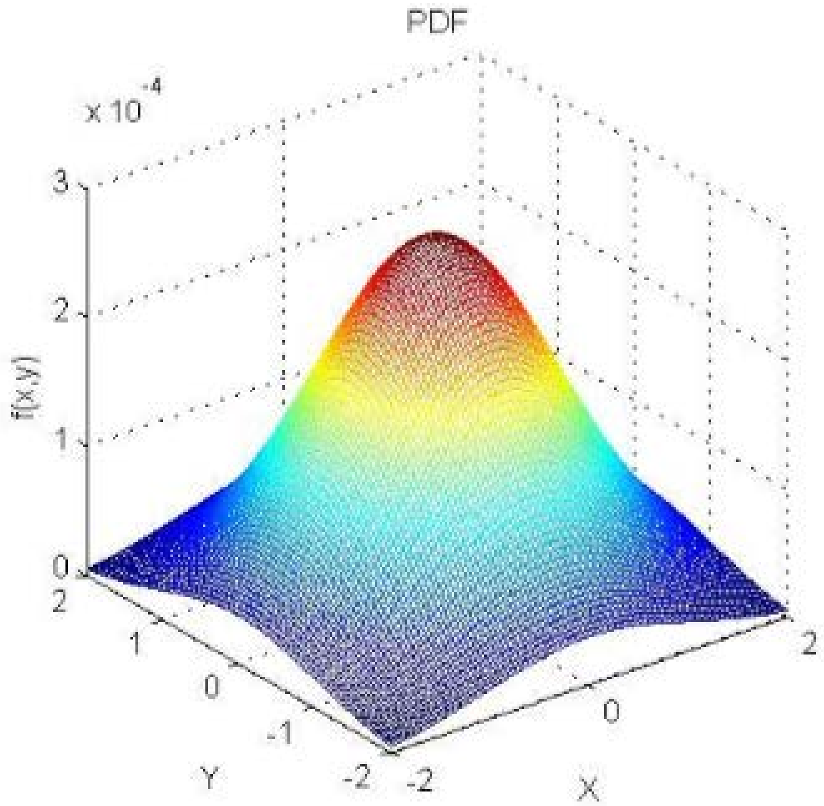
Asymmetric PDF:

$$\rho = 0.8$$

X and Y **dependent**

Symmetric Case

Bivariate Normal Distribution



Symmetric PDF:
$\rho = 0$

X and Y independent

Because of the independence, we should have

$$f(x|y) = f_X(x)$$

$$f(y|x) = f_Y(y)$$

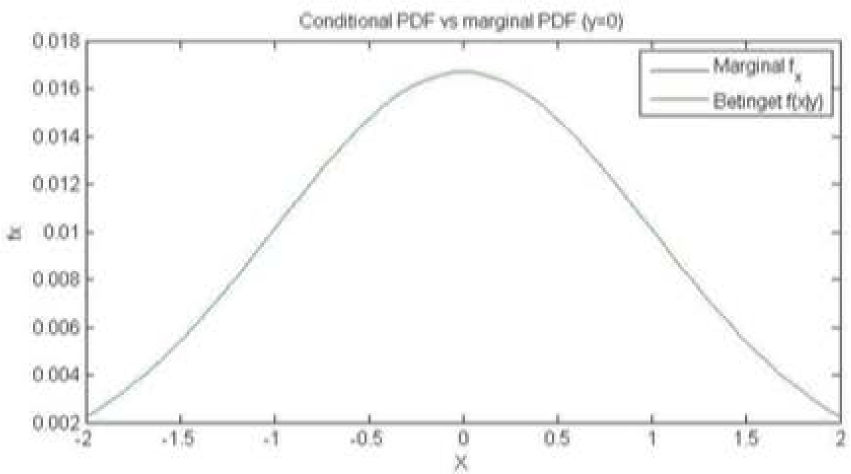
$$f(x,y) = f_X(x) \cdot f_Y(y)$$

Symmetric Case

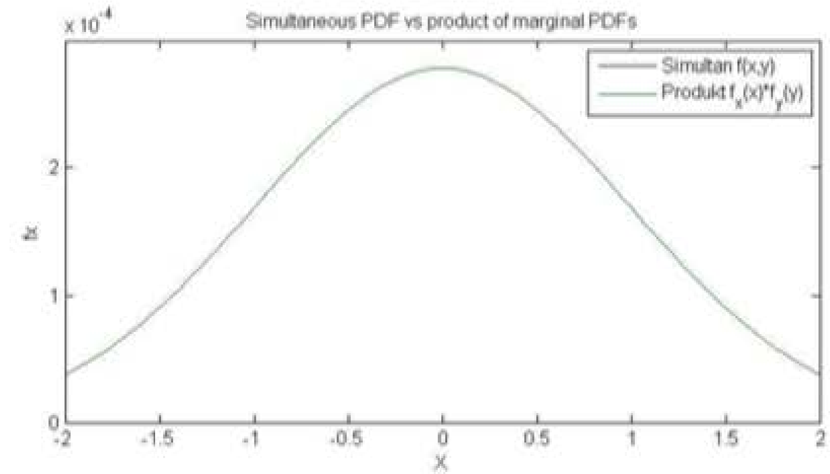
Bivariate Normal Distribution

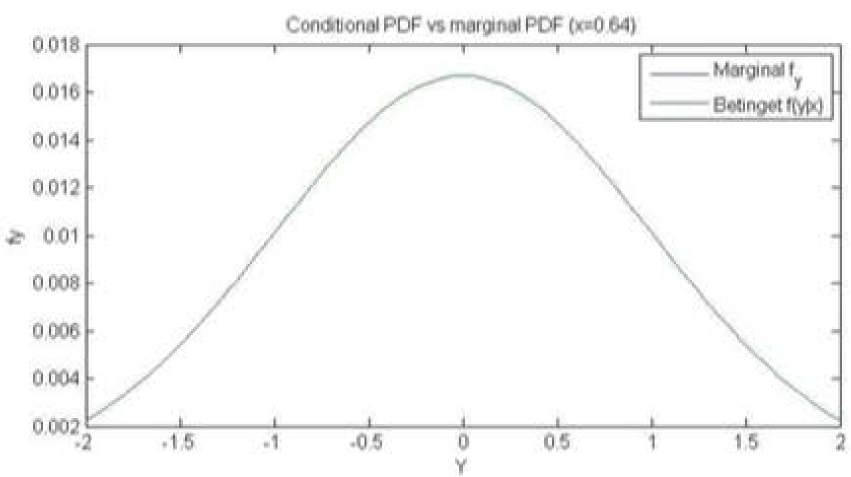
The graphs ($f_{X|Y}(x|y=0)$, $f_{X,Y}(x,y=0)$) and $f_X(x)$ has the same shape (proportional)

$$f(x|y=0) = f_X(x)$$

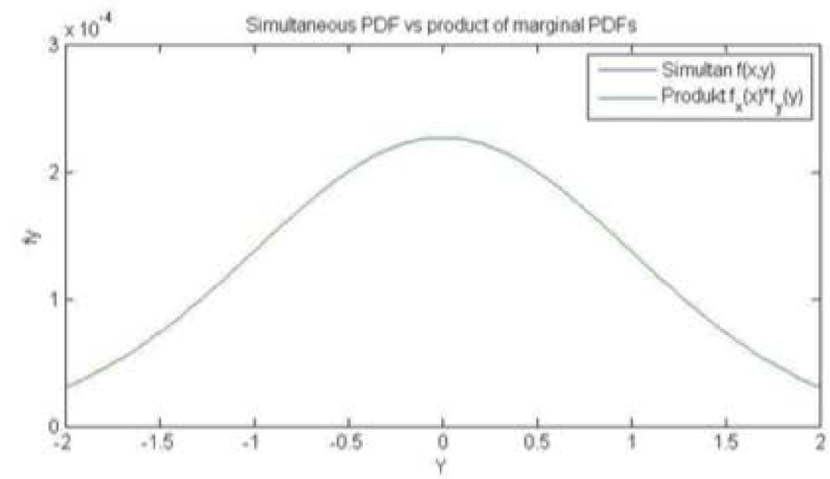


$$f(x,y=0) = f_X(x) \cdot f_Y(y=0)$$





$$f(y|x=0.64) = f_Y(y)$$

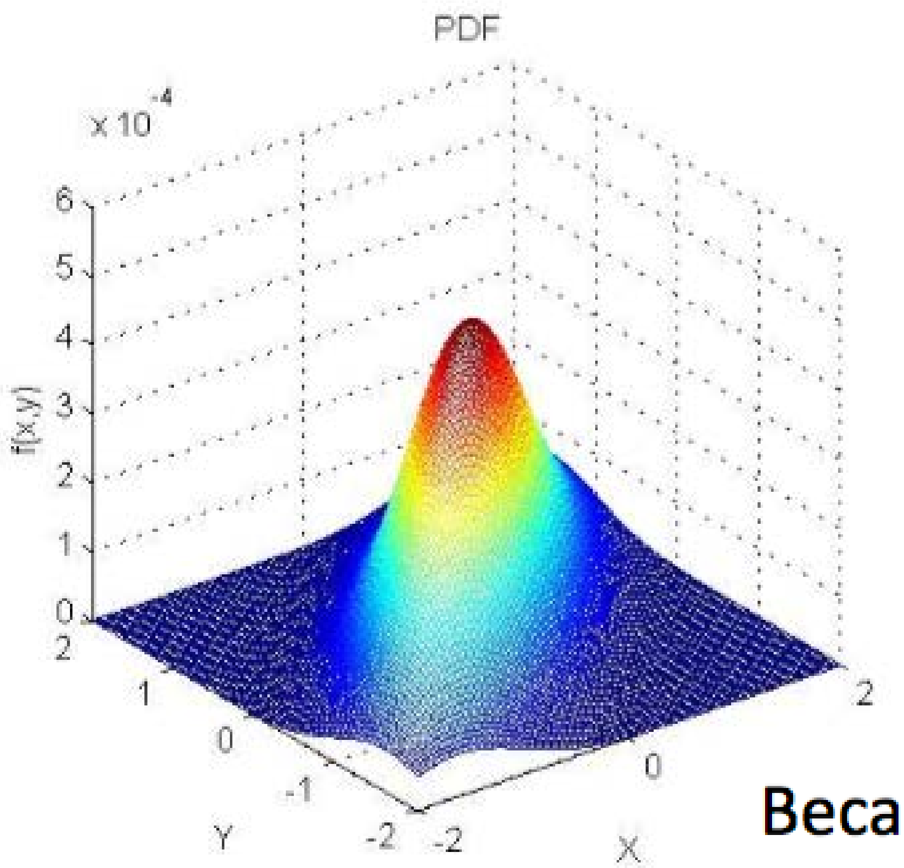


$$f(x=0.64,y) = f_X(x=0.64) \cdot f_Y(y)$$

The graphs $f_{Y|X}(y|x=0.64)$, $f_{X,Y}(x=0.64,y)$ and $f_Y(y)$ has the same shape (proportional)

Asymmetric Case

Bivariate Normal Distribution



Asymmetric PDF:

$\rho = 0.8$

X and Y **dependent**

Because of the dependence, we should have

$$f(x|y) \neq f_X(x)$$

$$f(y|x) \neq f_Y(y)$$

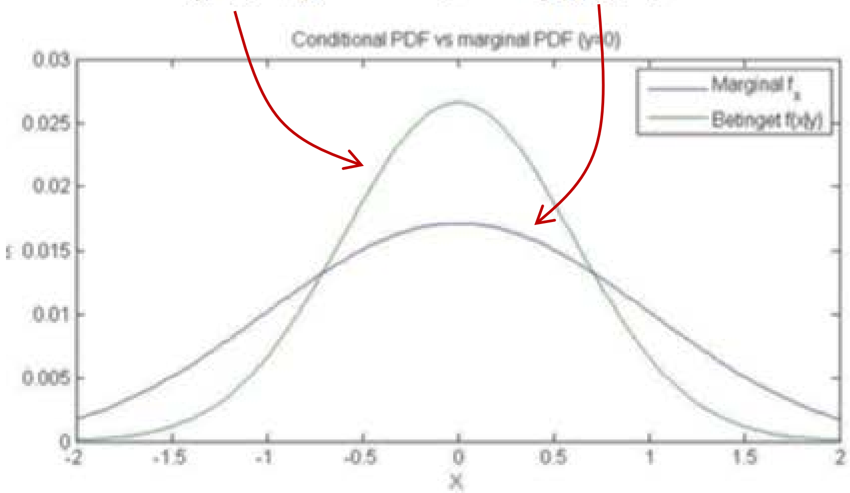
$$f(x,y) \neq f_X(x) \cdot f_Y(y)$$

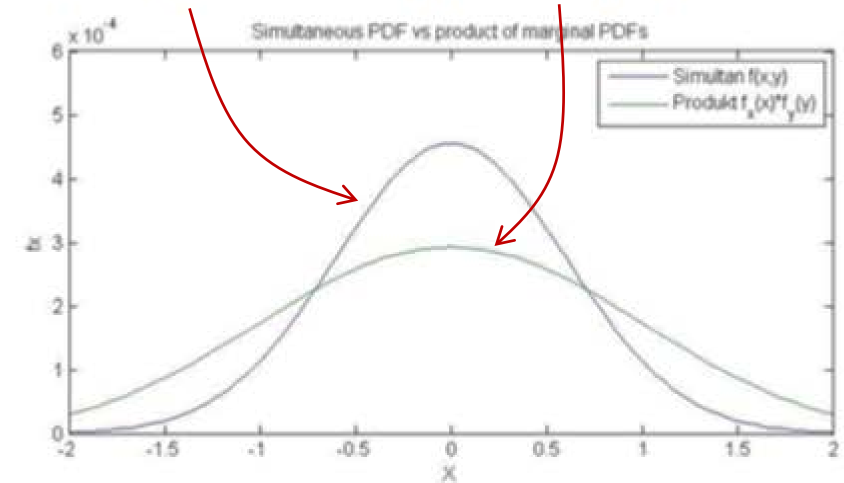
Asymmetric Case

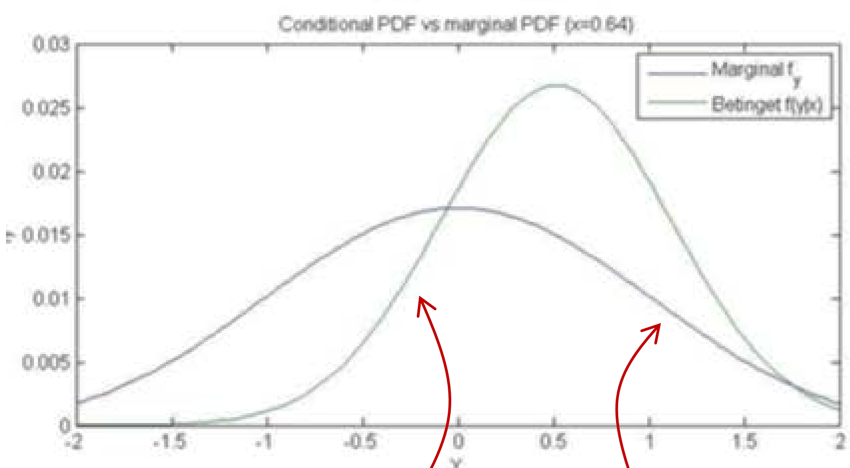
Bivariate Normal Distribution

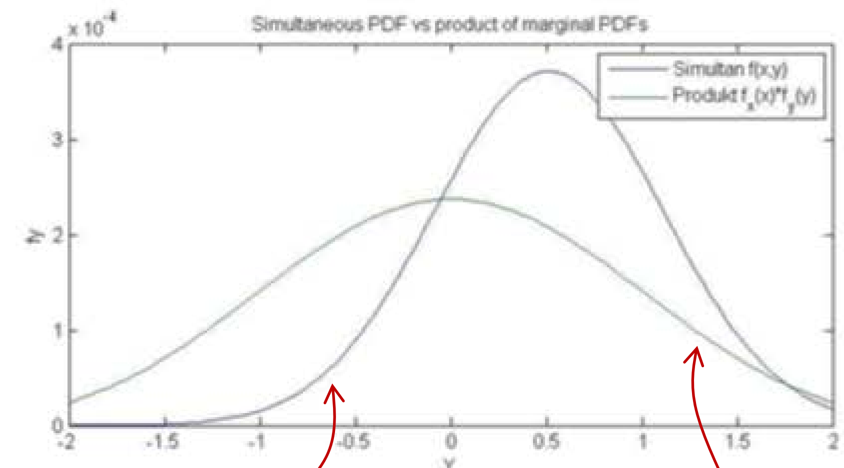
The graphs ($f_{X|Y}(x|y=0)$, $f_{X,Y}(x,y=0)$) and $f_X(x)$ do not have the same shapes.

$$f(x|y=0) \neq f_X(x) \qquad f(x,y=0) \neq f_X(x) \cdot f_Y(y=0)$$







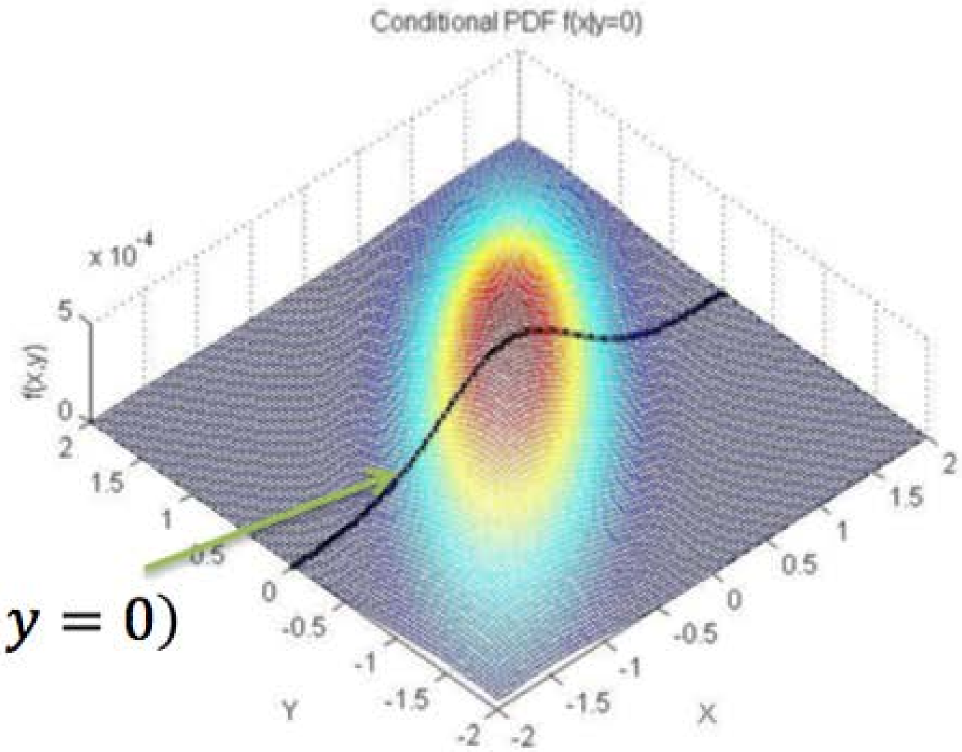


$$f(y|x=0.64) \neq f_Y(y) \qquad f(x=0.64,y) \neq f_X(x=0.64) \cdot f_Y(y)$$

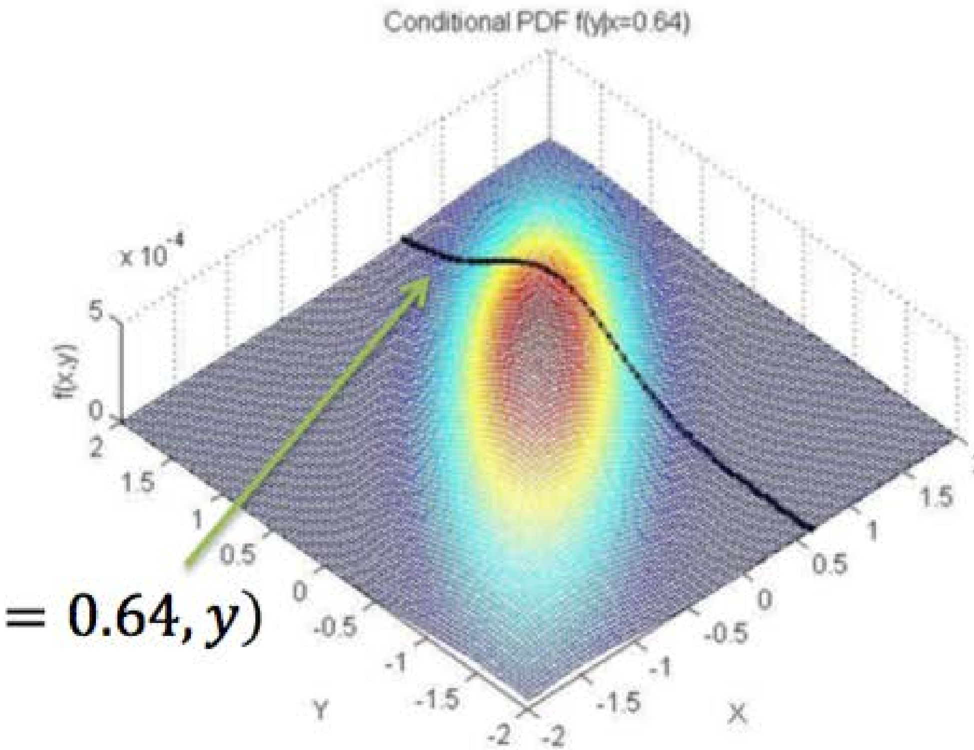
The graphs ($f_{Y|X}(y|x=0.64)$, $f_{X,Y}(x=0.64,y)$) and $f_Y(y)$ do not have the same shapes.

The Conditional pdf's

Bivariate Normal Distribution



$f(x, y = 0)$



$f(x = 0.64, y)$

$$\textit{Area under the curve} = \int_{-\infty}^{\infty} f(x, y = 0)dx = f_Y(y = 0)$$

$$\textit{Area under the curve} = \int_{-\infty}^{\infty} f(x = 0.64, y)dx = f_X(x = 0.64)$$

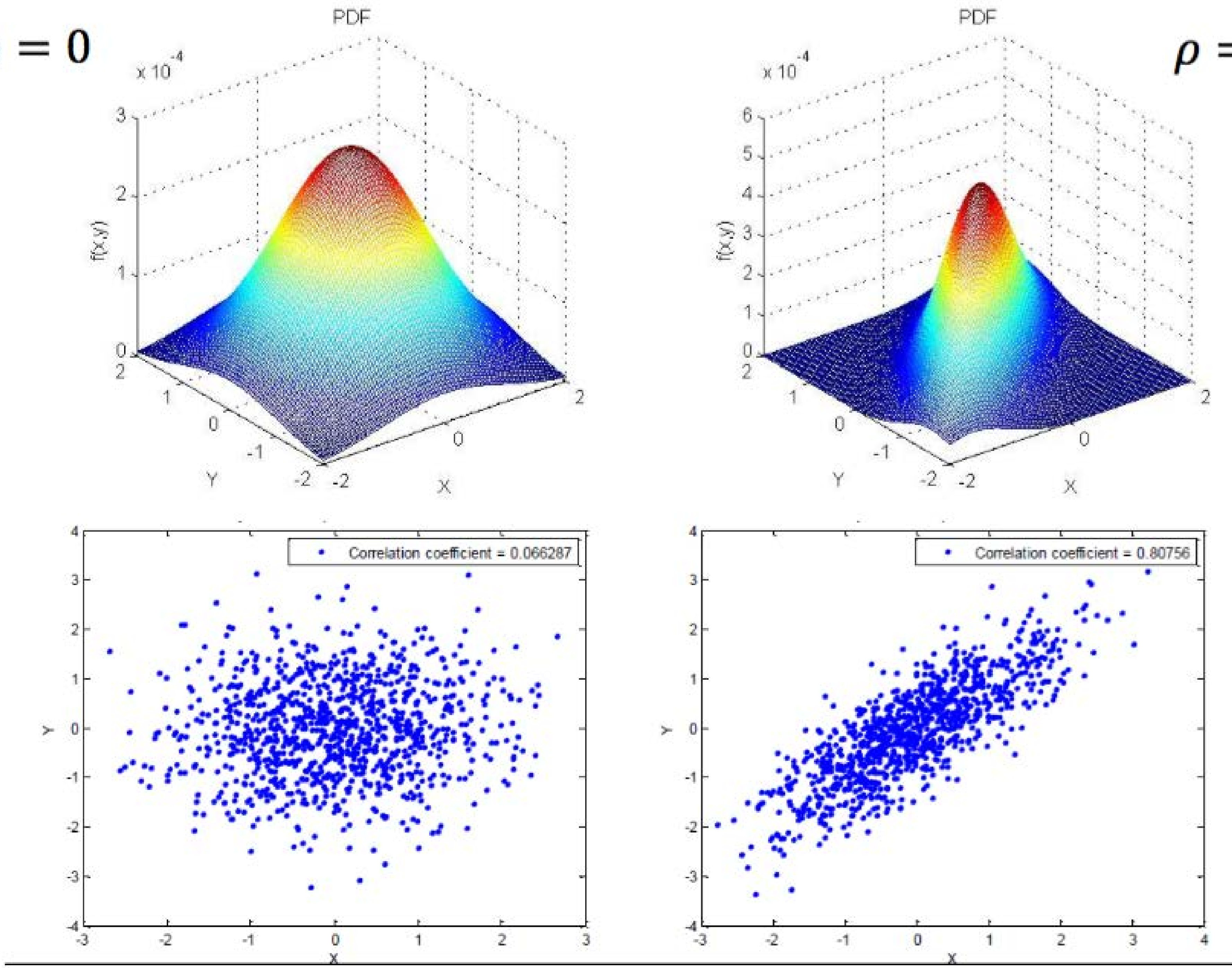
$$f(x|y = 0) = \frac{f(x, y = 0)}{f_Y(y = 0)}$$

$$f(y|x = 0.64) = \frac{f(x = 0.64, y)}{f_X(x = 0.64)}$$

Bivariate Normal Distribution

$\rho = 0$

$\rho = 0.8$



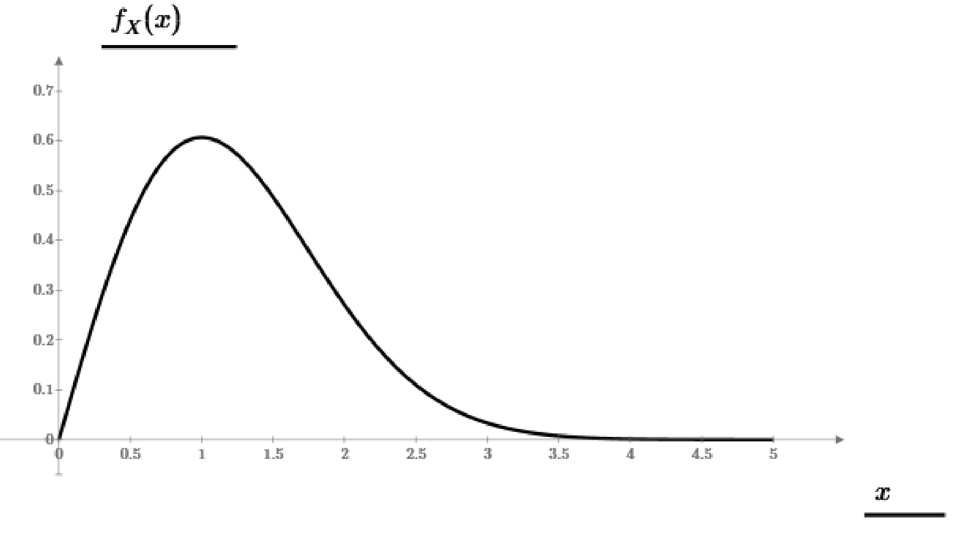
Sampling From Any Distribution

For test or simulation you need testdata (”measurements”) randomly sampled from a given distribution:

- Find the cdf of the distribution: $F_X(x)$
- Find the inverse of the cdf: $y = F_X(x) \;\Rightarrow\; x = F_X^{-1}(y)$
- Draw a ramdom sample: $y \sim \mathcal{U}[0; 1]$
- Insert into the inverse cdf: $x = F_X^{-1}(y)$
- The samples $X = x$ is distributed according to: $F_X(x)$

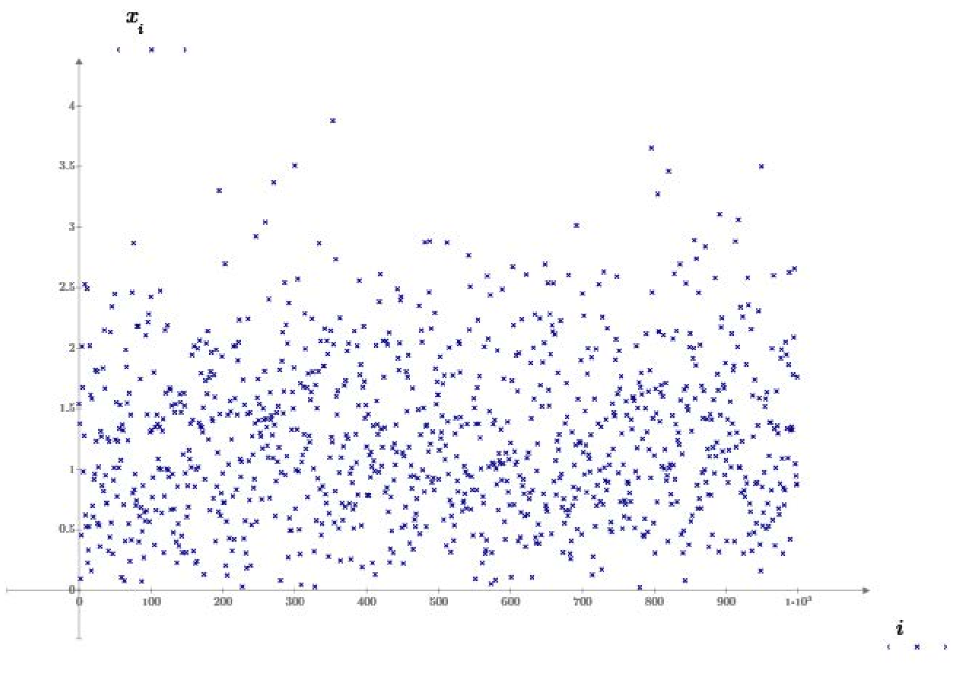
Example – Flight Simulator

- In a flight simulator, the altitude of the plane is simulated to be Rayleigh distributed.
- For a given initial height, draw a Rayleigh distributed sample.



Flight Simulator Example

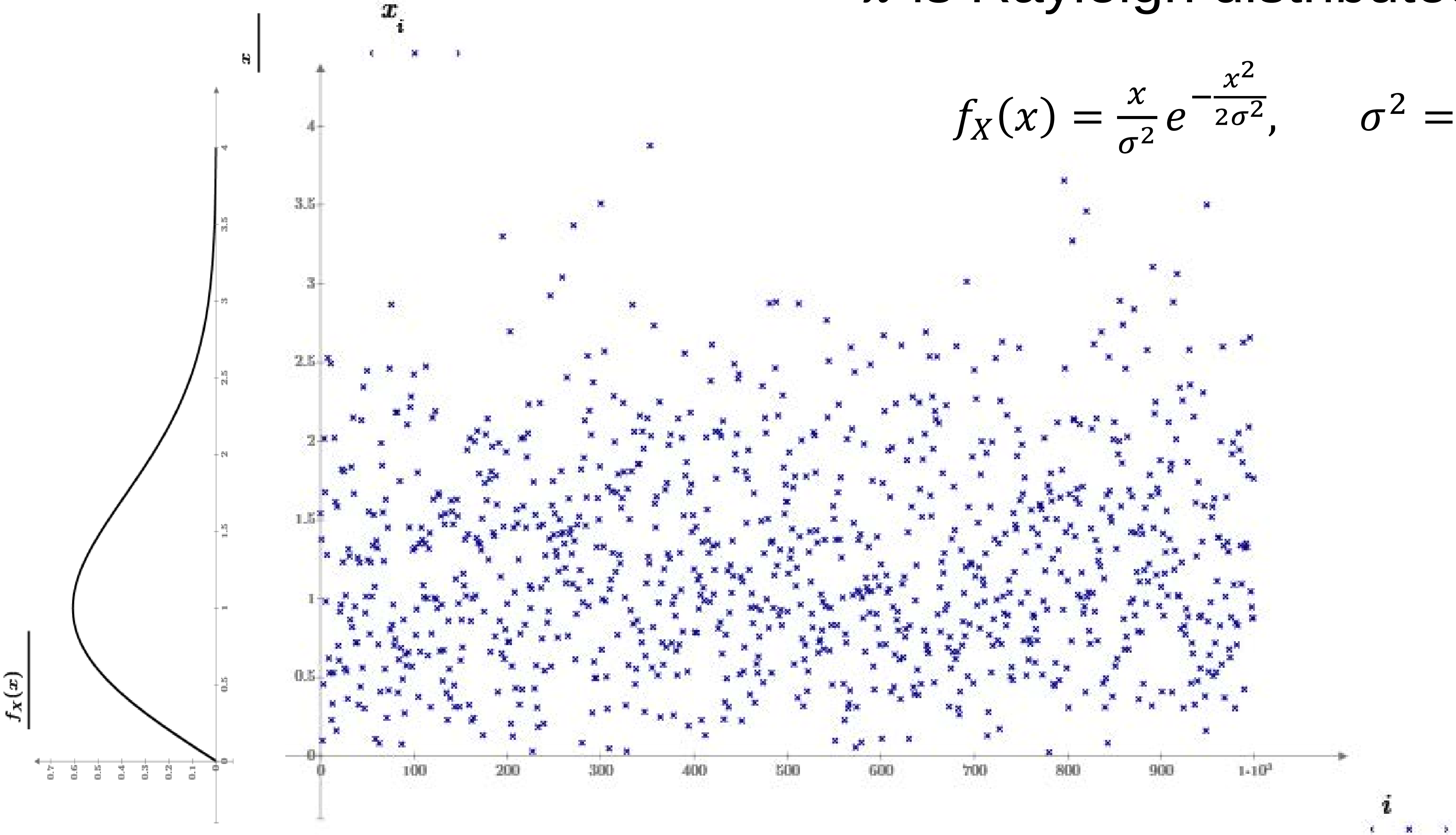
- Rayleigh pdf: $f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ for $x \geq 0$
- Rayleigh cdf: $F_X(x) = \int_0^x \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx = 1 - e^{-\frac{x^2}{2\sigma^2}}$
- Invers of cdf: $\mathrm{y} = 1 - e^{-\frac{x^2}{2\sigma^2}} \Rightarrow x = \sqrt{-2\sigma^2 \ln(1-y)}$
- Draw $\mathrm{y} \sim \mathcal{U}[0;1]$ and insert into $x = \sqrt{-2\sigma^2 \ln(1-y)}$
- x is Rayleigh distributed



Flight Simulator Example

x is Rayleigh distributed:

$$f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \qquad \sigma^2 = 1$$



Assignment

- Choose an exponential pdf: $f_X(x) = \lambda e^{-\lambda \cdot x}$
- Make a Matlab program that samples from that distribution

Transformation of Variable X to Y

- Given:
  - Pdf: $f_X(x)$
  - Function/Transformation: $\mathrm{Y} = g(X)$
  - Limits: $\mathrm{a} \leq X \leq b$

- Find new pdf: $f_Y(y)$:
  1. Inverse: $x = g^{-1}(y)$
  2. Differentiate: $\frac{dg^{-1}(y)}{dy} = \frac{dx(y)}{dy} = \frac{1}{\frac{dg(x)}{dx}}$
  3. Limits: Find $g(a) = a_Y \leq Y \leq b_Y = g(b)$ based on $\mathrm{a} \leq X \leq b$
  4. New pdf: $f_Y(y) = \sum \left|\frac{dx(y)}{dy}\right| f_X\left(g^{-1}(y)\ \right) = \sum \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}$

Example with Transformation of Random Variable

- We have a random sample x.
- The Noise is known to be Gaussian distributed.
- The signal of the noise is amplified.
- What is the pdf of the amplified noise?

- Given:
  - function: $Y = 2X$
  - pdf: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\sim\mathcal{N}(\mu, \sigma^2)$
  - Support: $x \in \mathbf{R}$
- Steps:
  1. Inverse: $x = \frac{1}{2}y$
  2. Differentiate: $\frac{d}{dy}\frac{1}{2}y = \frac{1}{2}$
  3. Support: $y \in \mathbf{R}$
  4. New pdf: $f_Y(y) = \frac{1}{2} f_X(\frac{1}{2}y)$.
- Then: $f_Y(y) = \frac{1}{2}\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\frac{y}{2}-\mu)^2}{2\sigma^2}}$

  $\sim\mathcal{N}(2\mu, 4\sigma^2)$

Distribution of the Sum of Two Random Variables

- Two random variables X and Y have density functions $f_X(x)$ and $f_Y(y)$.

- If we define a new random variable $Z = X + Y$, and Z have density function $f_Z(z)$.

Convolution of Two functions

- Then $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$

Expectation of the Sum of Two Random Variables

- For a random variables $Z = X + Y$.
- X, Y can be both dependent and independent.
- The expectation of Z is:

$$\boxed{E[Z] = E[X] + E[Y]}$$

Proof:

$$\begin{aligned}
E[X+Y] &= \int_x \int_y (x+y) f_{X,Y}(x,y)dxdy \\
&= \int_x \int_y x f_{X,Y}(x,y)dxdy + \int_x \int_y y f_{X,Y}(x,y)dxdy \\
&= \int_x x \int_y f_{X,Y}(x,y)dydx + \int_y y \int_x f_{X,Y}(x,y)dxdy \\
&= \int_x x f_X(x)dx + \int_y y f_Y(y)dy \\
&= E[X] + E[Y]
\end{aligned}$$

Variance of the Sum of Two Random Variables

- We have $Z = X + Y$.

- For independent random variables X, Y, the variance of Z is:

$$\boxed{var(Z) = var(X) + var(Y).}$$

- For correlated random variables X, Y, the variance of Z is:

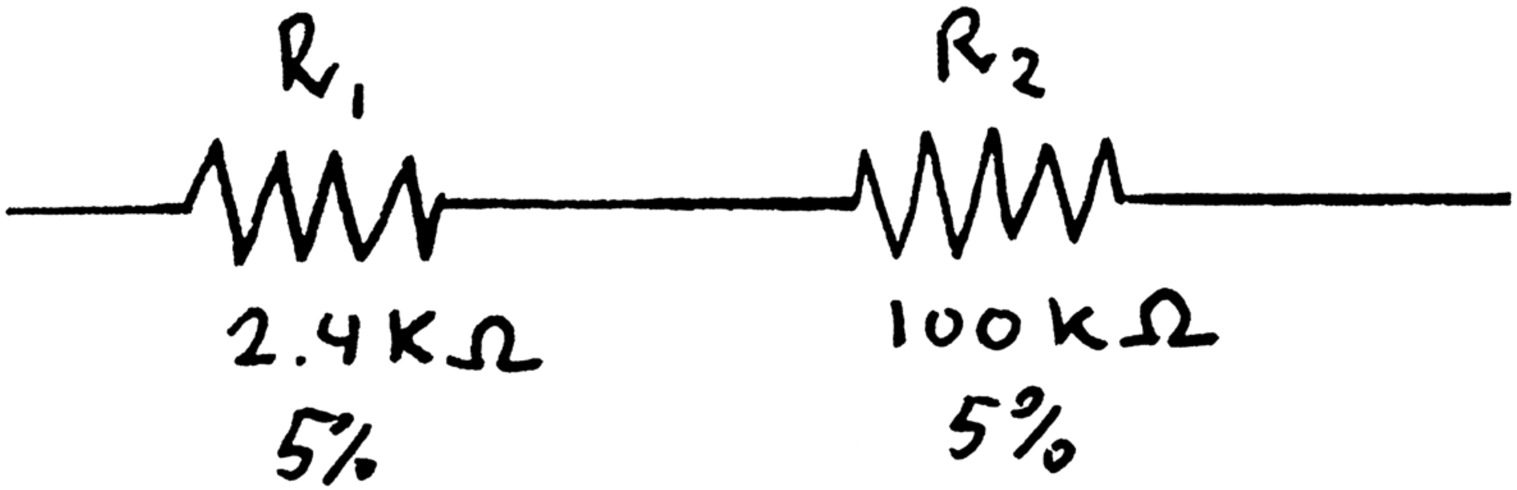
$$\boxed{var(Z) = var(X) + var(Y) + 2cov(X, Y).}$$

where: $cov(X, Y) = E[XY] - E[X]E[Y]$

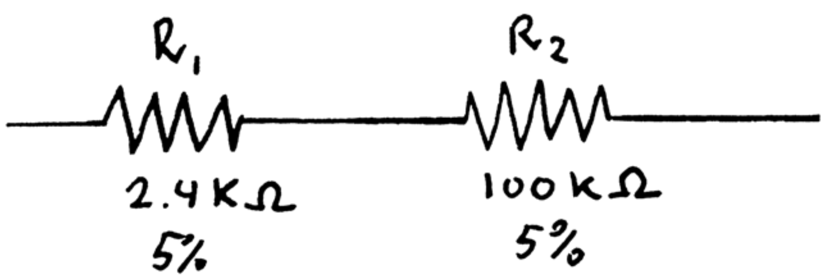
Proof: Similar to the proof of the expectation value

Precision of Resistors in Series

- In a analog filter a resister of size $2.5K\Omega$ is needed.
- We use two 5% resisters of $2.4K\Omega$ and 100Ω respectively.
- What is the resulting uncertainty of the resister?
- X and Y are independent random variables with pdfs: $f_X(x)$ and $f_Y(y)$
- What is the pdf of a random variable Z, where $Z = X + Y$

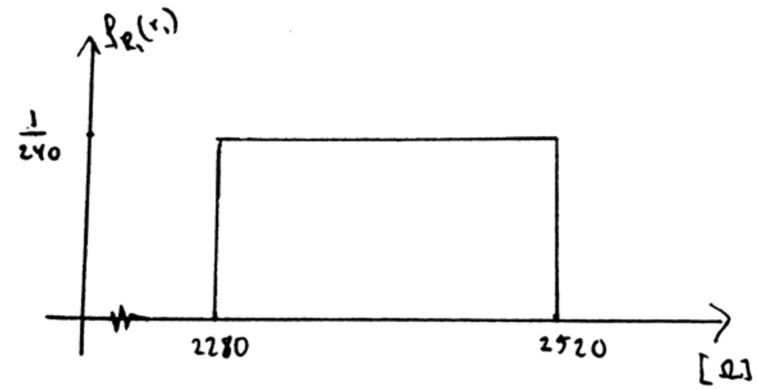


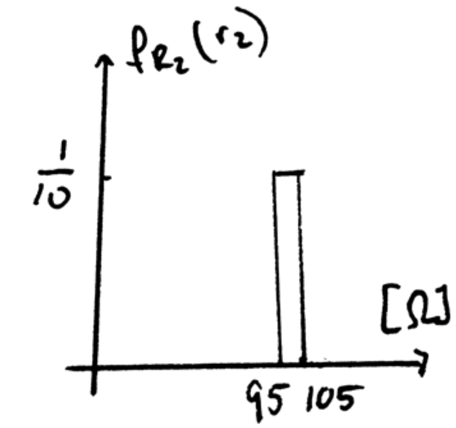
Precision of Resistors in Series

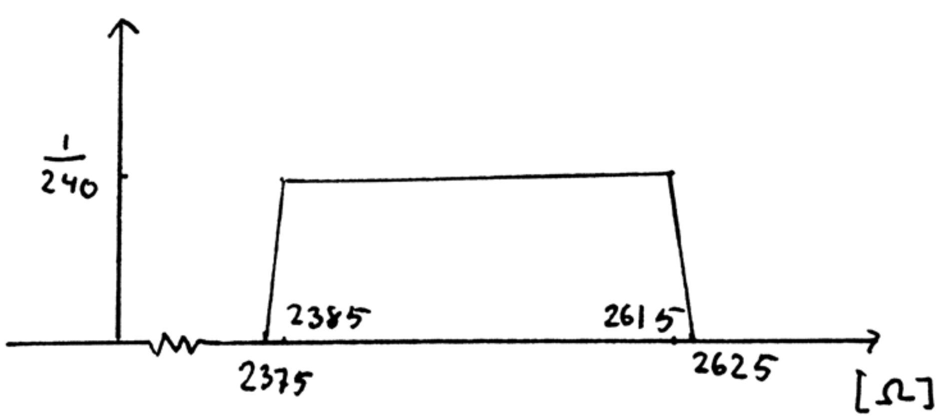


- We assume that the resistance of the resisters are uniformly distributed.
- $R_1 \sim \mathcal{U}[2280; 2520]$
- $R_2 \sim \mathcal{U}[95; 105]$
- The resisters are in series: $R_3 = R_1 + R_2$.
- We have: $f_{R_3}(r_3) = \int_{-\infty}^{\infty} f_X(\rho) f_Y(r_3 - \rho) d\rho$
- We can find that:

$$f_{R_3}(r_3) = \begin{cases} \frac{1}{2400} r_3 - \frac{95}{96} & \text{for } 2375 \leq x < 2385 \\ \frac{1}{240} & \text{for } 2385 \leq x < 2615 \\ -\frac{1}{2400} r_3 + \frac{35}{32} & \text{for } 2615 \leq x < 2625 \\ 0 & \text{otherwise} \end{cases}$$

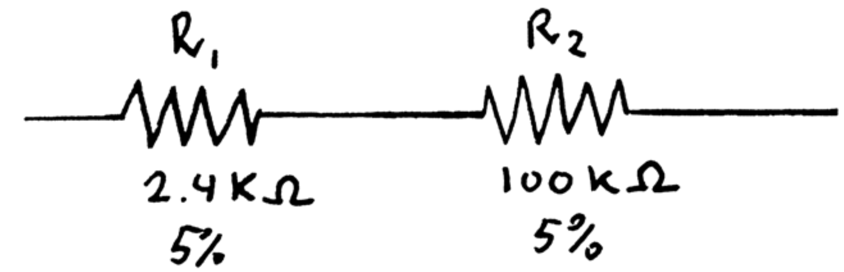






R3 is still a 5% resistor – but no longer uniform distributed!

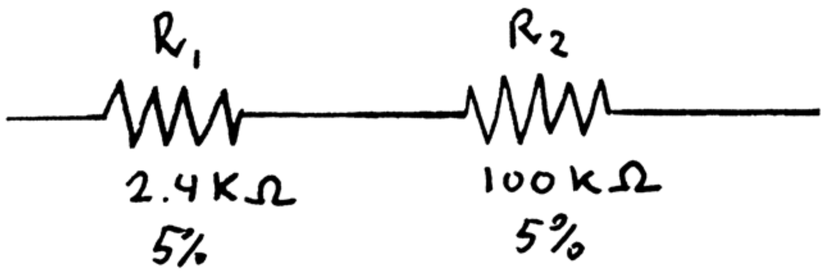
Expected Value of the Resistor



- We assume that R_1 and R_2 are independent
- For a uniform distibution: $\mathrm{E}[R_1] = \frac{1}{2}(2520 + 2280) = 2400\Omega$
- For a uniform distribution: $\mathrm{E}[R_2] = \frac{1}{2}(105 + 95) = 100\Omega$
- For the sum $R_3 = R_1 + R_2$ we have:

$$E[R_3] = E[R_1] + E[R_2] = 2400\Omega + 100\Omega = \underline{2500\Omega}$$

Variance of the Resistor



- We assume that R_1 and R_2 are independent
- For a uniform distibution: $var(R_1) = \frac{1}{12}(2520 - 2280)^2 = 4800$
- For a uniform distribution: $var(R_2) = \frac{1}{12}(105 - 95)^2 = 8{,}333$
- For the sum $R_3 = R_1 + R_2$ we have:

$$var(R_3) = var(R_1) + var(R_2) = 4808 \ \rightarrow \ \sigma_3 = 69\Omega$$

- For one uniform distributed 5%-resistor $R_0 = 2500 \sim \mathcal{U}[2375; 2625]$:

$$var(R_0) = \frac{1}{12}(2625 - 2375)^2 = 5208 \ \rightarrow \ \sigma_0 = 72\Omega$$

- So: $var(R_3) = var(R_1) + var(R_2) < var(R_0)$ $(\sigma_3 < \sigma_0)$

Two Random Variables

Two random variables: X and Y

- Simultaneous pdf: $f_{X,Y}(x,y)$
- Marginal pdf: $f_X(x)$ and $f_Y(y)$
- Conditional pdf: $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$
- Simultaneous cdf: $F_{X,Y}(x,y)$
- Correlation: $corr(X,Y) = E[XY]$
- Covariance: $cov(X,Y) = E[XY] - E[X]E[Y]$
- Correlation coefficient: $\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$
- Sum: $\mathrm{Z} = X + Y$
- Expectation: $\mathrm{E[Z]} = E[X] + E[Y]$
- Variance: $\mathrm{Var[Z]} = Var[X] + Var[Y]$ if <u>independent</u>
  $\mathrm{Var[Z]} = Var[X] + Var[Y] + 2cov(X,Y)$ if <u>dependent</u>

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2
- Let $\bar{X}$ be the random variable (average):

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

- Then in the limit: $n \to \infty$ we have that: $\underline{\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})}$

i.e. in the limit $\bar{X}$ will be normally distributed with mean = μ and variance = $\frac{\sigma^2}{n}$.

The variance is reduced with a factor 1/n

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2
- Let X be the random variable:

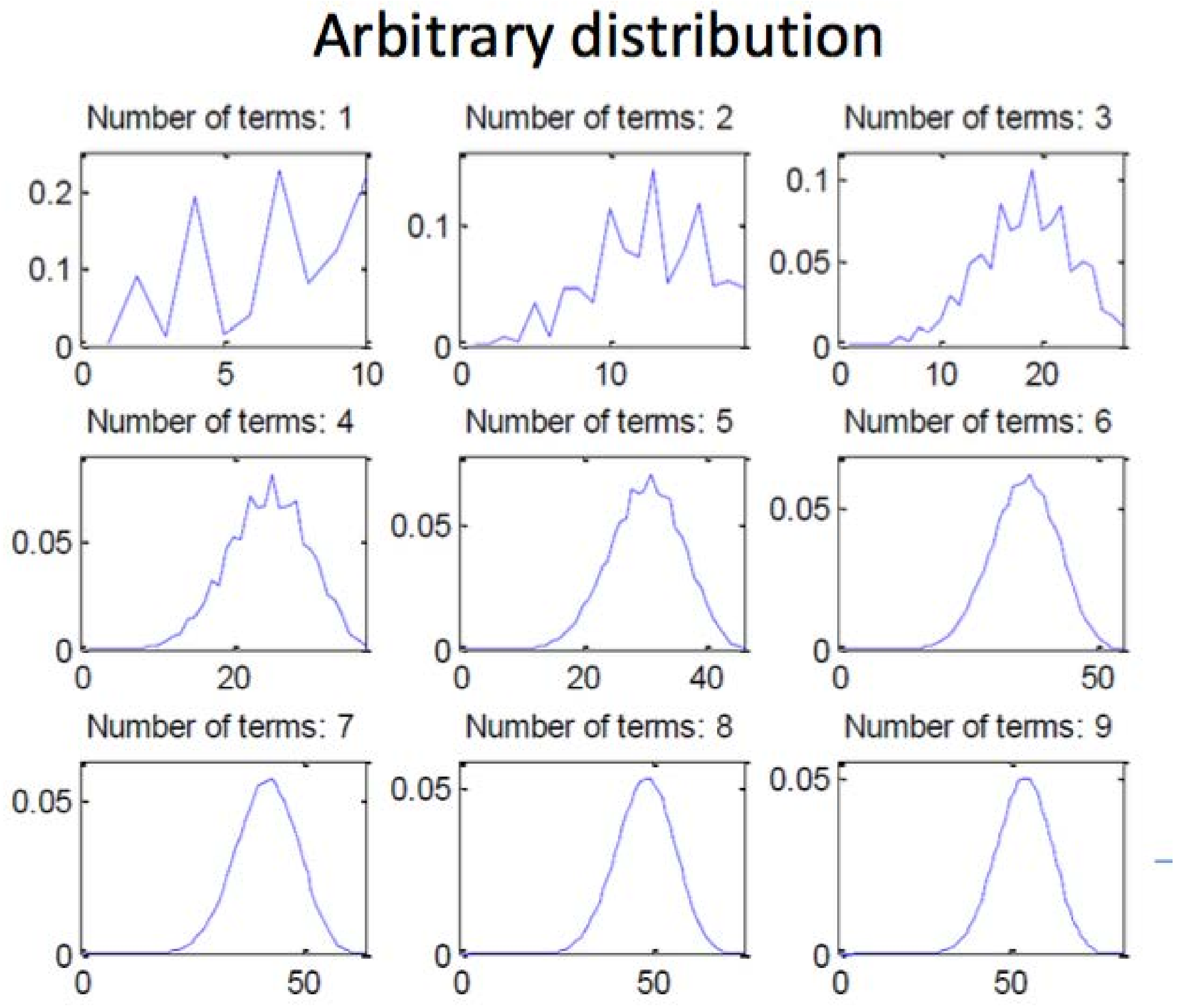
$$X = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^{n} \frac{1}{n} X_i - \mu}{\sqrt{\sigma^2/_n}} = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/_n}}$$

- Then in the limit: $n \to \infty$ we have that: $X \sim \mathcal{N}(0,1)$

  i.e. in the limit X will be normally distributed with mean = 0 and variance = 1 (standard normal distributed).

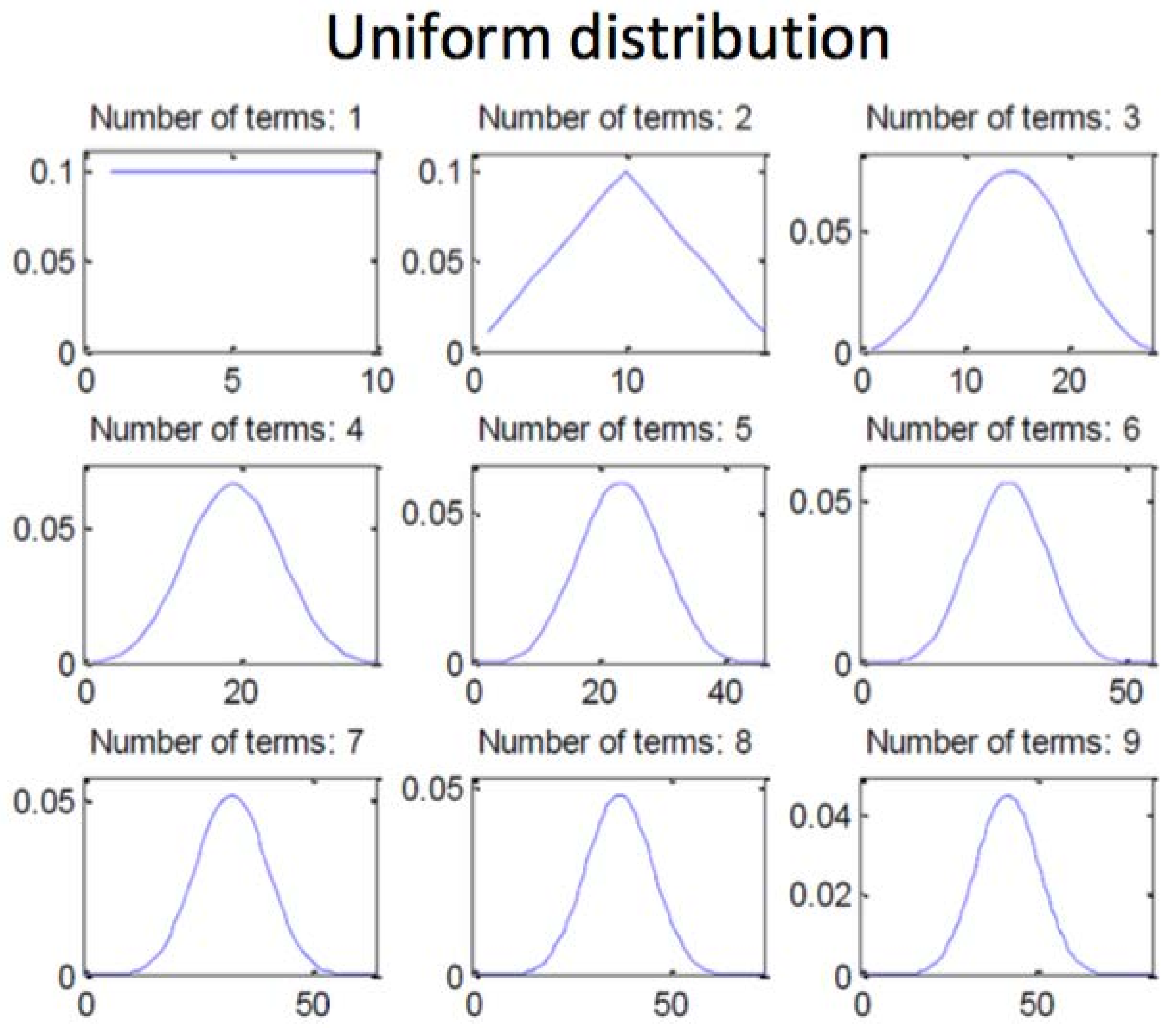
Sum of Random Variables

- The random variables are i.i.d and taken from the same distribution.



Sum of Random Variables

- The random variables are i.i.d and taken from the same uniform distribution.



Words and Concepts to Know

Central Limit Theorem

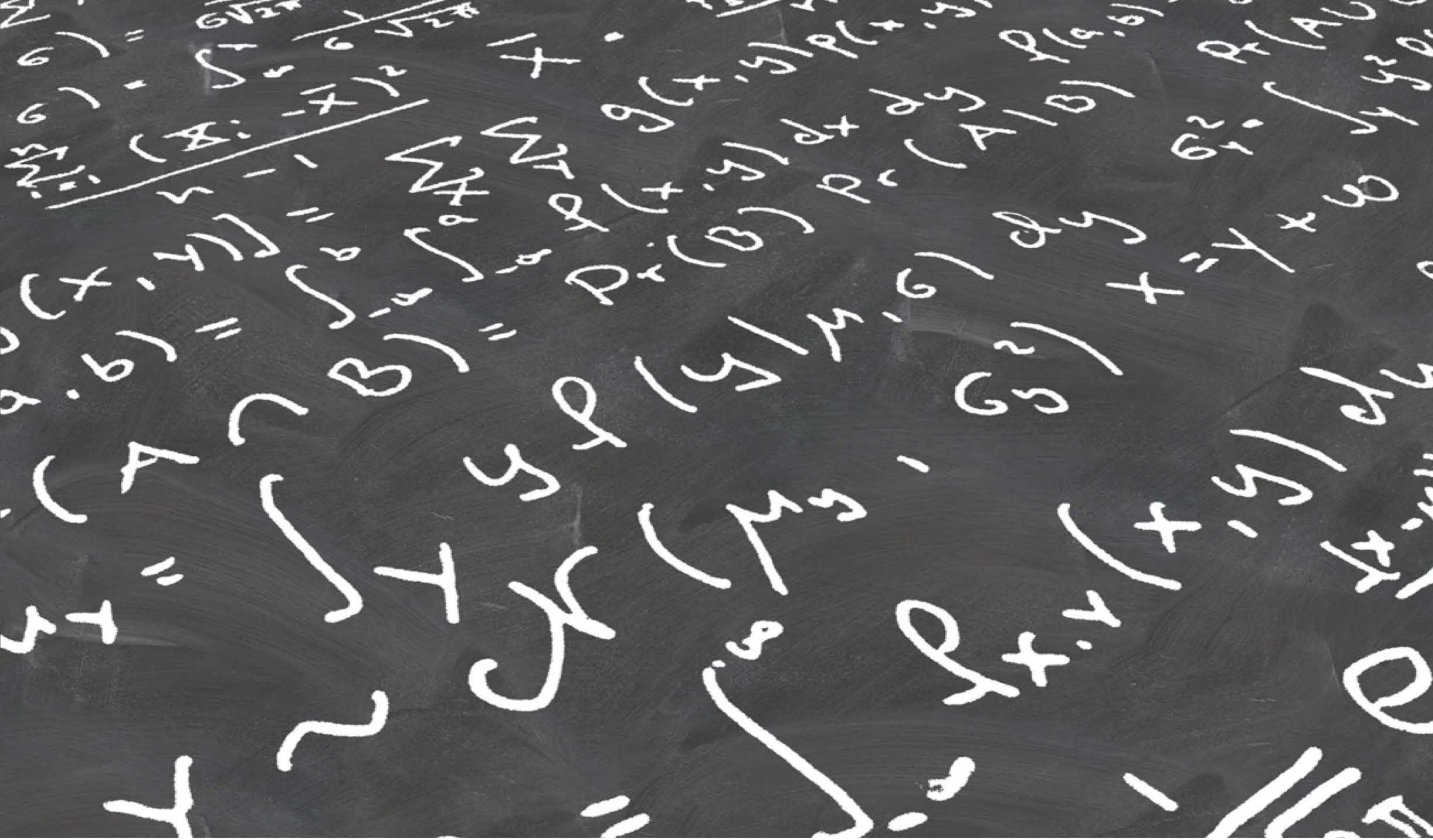
Convolution

Transformation of stochastic variables

Rayleigh Distribution

Randomly Sampled Data

Bivariate Normal Distribution



6. Introduction to Stochastic Processes

Gunvor Elisabeth Kirkelund
Lars Mandrup

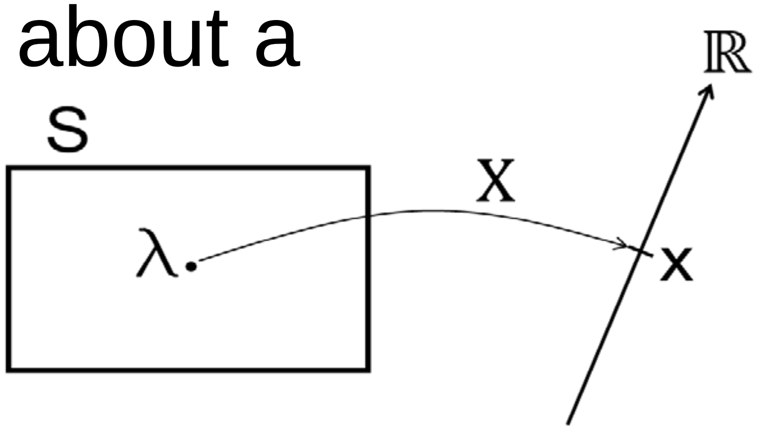
Agenda for Today

- Repetition from last time
  - Random Variables
  - The Central Limit Theorem
- Stochastic Processes
  - Stationarity (WSS, SSS)
  - Ergodic Processes

Also just called a random variables

Stochastic Random Variables

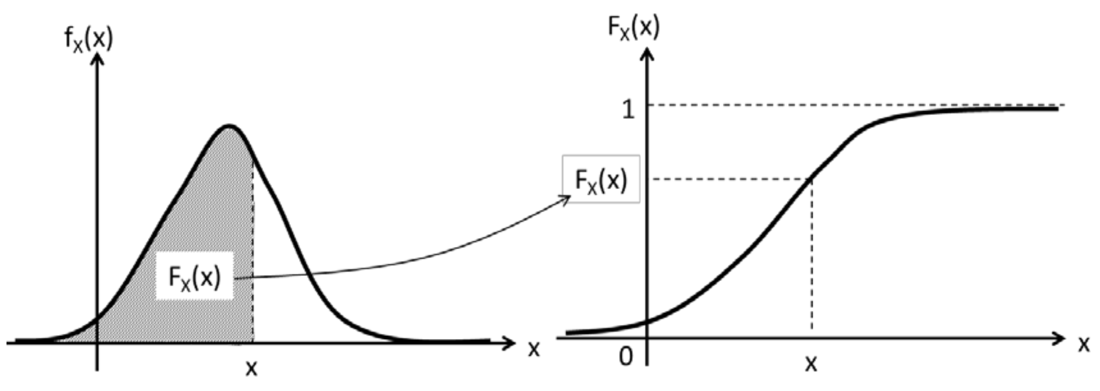
- A random variable tells something important about a stochastic experiment.
- Can be discrete or continous



- Probability density function (pdf):

$$Pr(a \leq X \leq b) = \int_a^b f_X(x)\,dx \qquad f_X(x) \geq 0 \qquad \int_{-\infty}^{\infty} f_X(x)dx = 1$$

- Cumulative distribution function (cdf):



$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = Pr(X \leq x)$$

$$0 \leq F_X(x) \leq 1$$

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad \lim_{x\to\infty} F_X(x) = 1$$

Two Random Variables X, Y

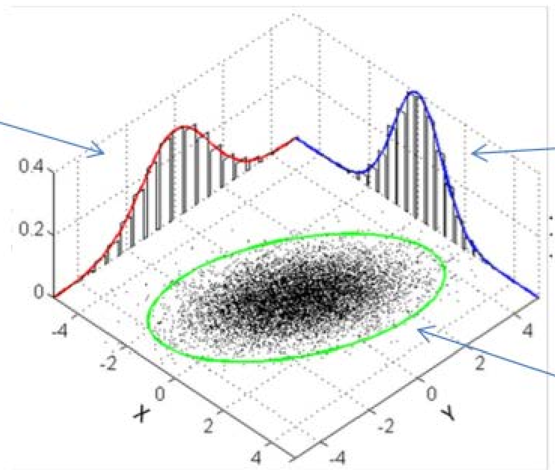
Joint (Simultaneous) pdf: $f_{X,Y}(x,y) \geq 0 \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

$$Pr\big((a \leq X \leq b) \cap (c \leq Y \leq d)\big) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

Marginals:

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

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



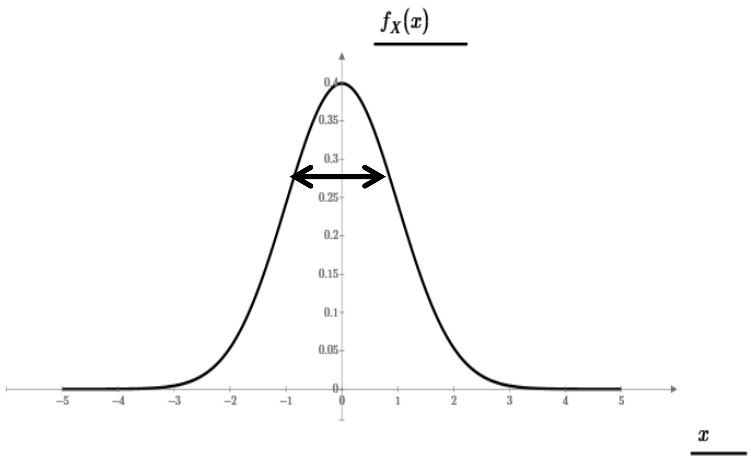
Cumulative Distribution Function cdf:

cdf $\quad F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy = Pr(X \leq x \wedge Y \leq y)$

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

Expectations

- Mean value: $E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X\,(x)dx \quad (\sum_{i=1}^{n} x_i\, f_X(x_i))$
- Variance: $Var(\mathrm{X}) = {\sigma_X}^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f_X\,(x)dx = E\left[X^2\right] - E[X]^2$
- Standard deviation: $\sigma_X = \sqrt{Var(\mathrm{X})}$



- A function: $E[g(X)] = \overline{g(X)} = \int_{-\infty}^{\infty} g(x) \cdot f_X\,(x)dx \quad (\sum_{i=1}^{n} g(x_i)\, f_X(x_i))$

$$Var\big(g(\mathrm{X})\big) = \int_{-\infty}^{\infty} \left(g(x) - \overline{g(x)}\right)^2 \cdot f_X\,(x)dx = E\left[g(X)^2\right] - E[g(X)]^2$$

- Linear function: $E[aX + b\,] = a \cdot E[X] + b$

$$Var[aX + b\,] = a^2\left(E[X^2] - E[X]^2\right) = a^2 \cdot Var(X)$$

Correlation, Covariance and summation

Two random variables: X and Y

- Correlation: $corr(X,Y) = E[XY]$
- Covariance: $cov(X,Y) = E[XY] - E[X]E[Y]$
- Correlation coefficient: $\rho = \frac{E[XY]-E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$ $\quad -1 \leq \rho \leq 1$

Sampling From Any Distribution

For test or simulation you need testdata (”measurements”) randomly sampled from a given distribution:

- Find the cdf of the distribution: $F_X(x)$
- Find the inverse of the cdf: $y = F_X(x) \;\Rightarrow\; x = F_X^{-1}(y)$
- Draw a ramdom sample: $y \sim \mathcal{U}[0;1]$
- Insert into the inverse cdf: $x = F_X^{-1}(y)$
- The samples $X = x$ is distributed according to: $F_X(x)$

Transformation of Variable X to Y

- Given:
  - Pdf: $f_X(x)$
  - Function/Transformation: $\mathrm{Y} = g(X)$
  - Limits: $\mathrm{a} \leq X \leq b$

- Find new pdf: $f_Y(y)$:
  1. Inverse: $x = g^{-1}(y)$
  2. Differentiate: $\frac{dg^{-1}(y)}{dy} = \frac{dx(y)}{dy} = \frac{1}{\frac{dg(x)}{dx}}$
  3. Limits: Find $g(a) = a_Y \leq Y \leq b_Y = g(b)$ based on $\mathrm{a} \leq X \leq b$
  4. New pdf: $f_Y(y) = \sum \left|\frac{dx(y)}{dy}\right| f_X\left(g^{-1}(y) \right) = \sum \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}$

Distribution of the Sum of Two Random Variables

- Two random variables X and Y have density functions $f_X(x)$ and $f_Y(y)$.
- If we define a new random variable $Z = X + Y$, and Z have density function $f_Z(z)$.

Convolution of Two functions

- Then $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$
- Expectation: $E[Z] = E[X] + E[Y]$
- Variance: $var(Z) = var(X) + var(Y) + 2cov(X, Y)$

Very important!

i.i.d.: Independent and Identically distributed

- We define that for series of random variables that is taken from the same distribution (identically distributed), and are sampled independent of each other, that they are i.i.d.

$$\boxed{\text{i.i.d. = Independent and Identically distributed}}$$

- i.i.d. is a very important characteristic in stochastic variable processing and statistics

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2
- Let $\bar{X}$ be the random variable (average):

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

- Then in the limit: $n \to \infty$ we have that: $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

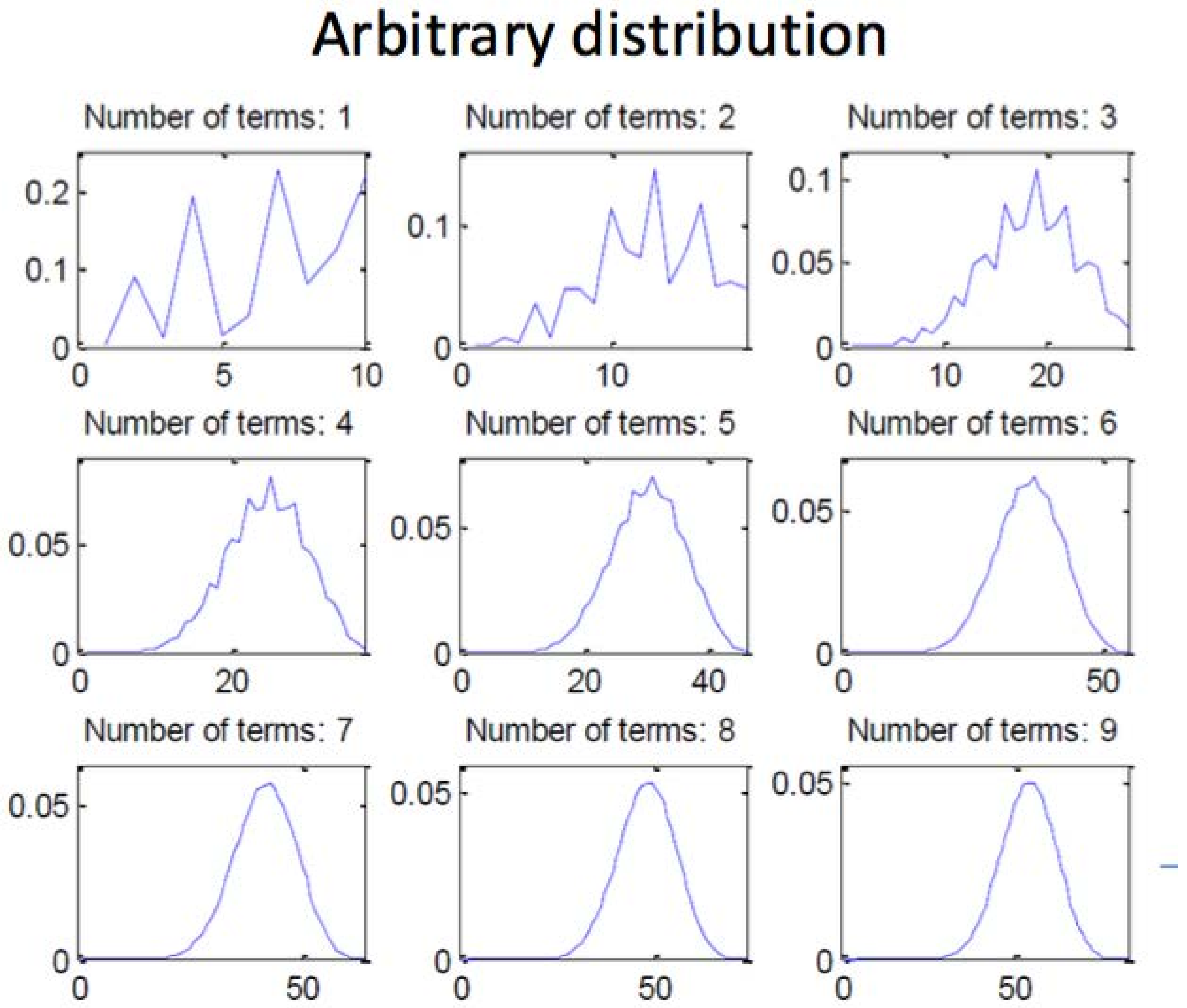
  i.e. in the limit $\bar{X}$ will be normally distributed with

  mean = μ and variance = $\frac{\sigma^2}{n}$.

The variance is reduced with a factor 1/n

Sum of Random Variables

- The random variables are i.i.d and taken from the same distribution.

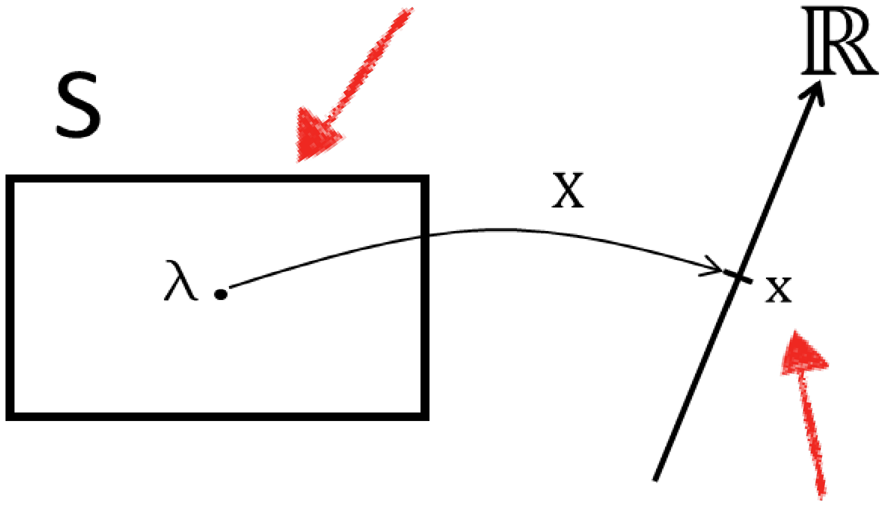


Stochastic Processes

Stochastic Variables

- Sample space for stochastic experiment

Sample space for
stochastic experiment



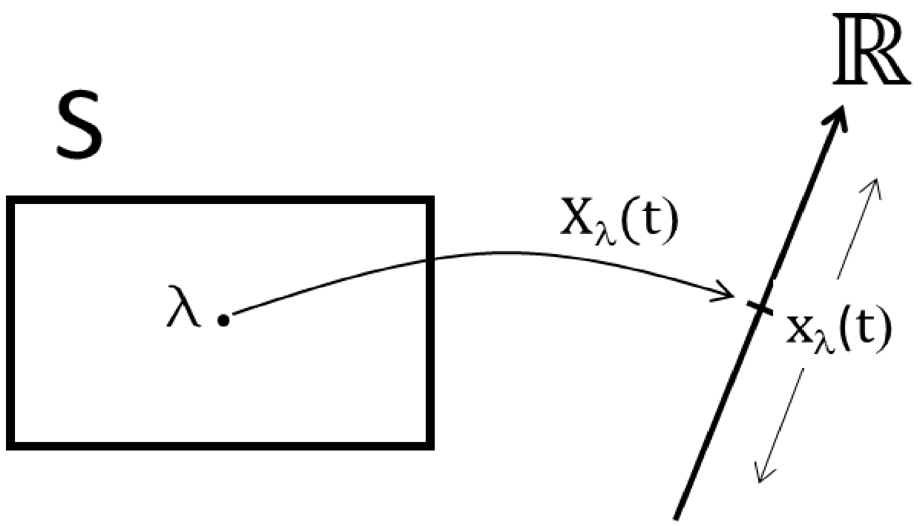
Sample space for
stochastic variable

Time dependent

Stochastic Processes (signals)

- Sample space for stochastic experiment
- Random events that develops in time

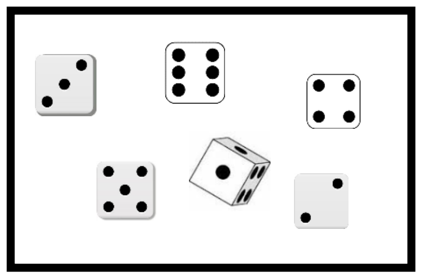
Sample space for
stochastic experiment



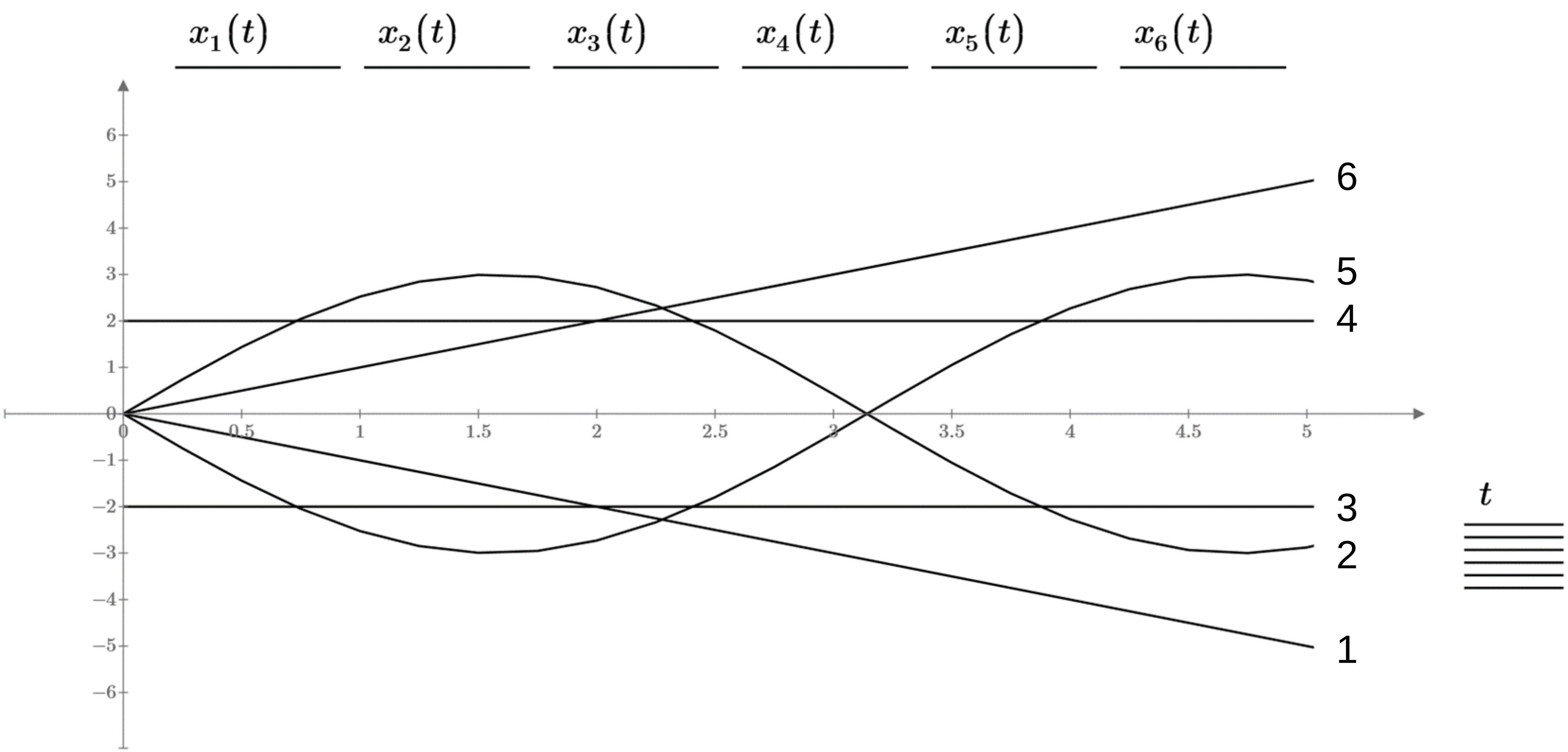
Sample space for
stochastic proces

Stochastic Processes – Example

S



$$X_n(t): \quad x_1(t) = -t \qquad x_2(t) = 3\sin(t)$$
$$x_3(t) = -2 \qquad x_4(t) = 2$$
$$x_5(t) = -3\sin(t) \qquad x_6(t) = t$$

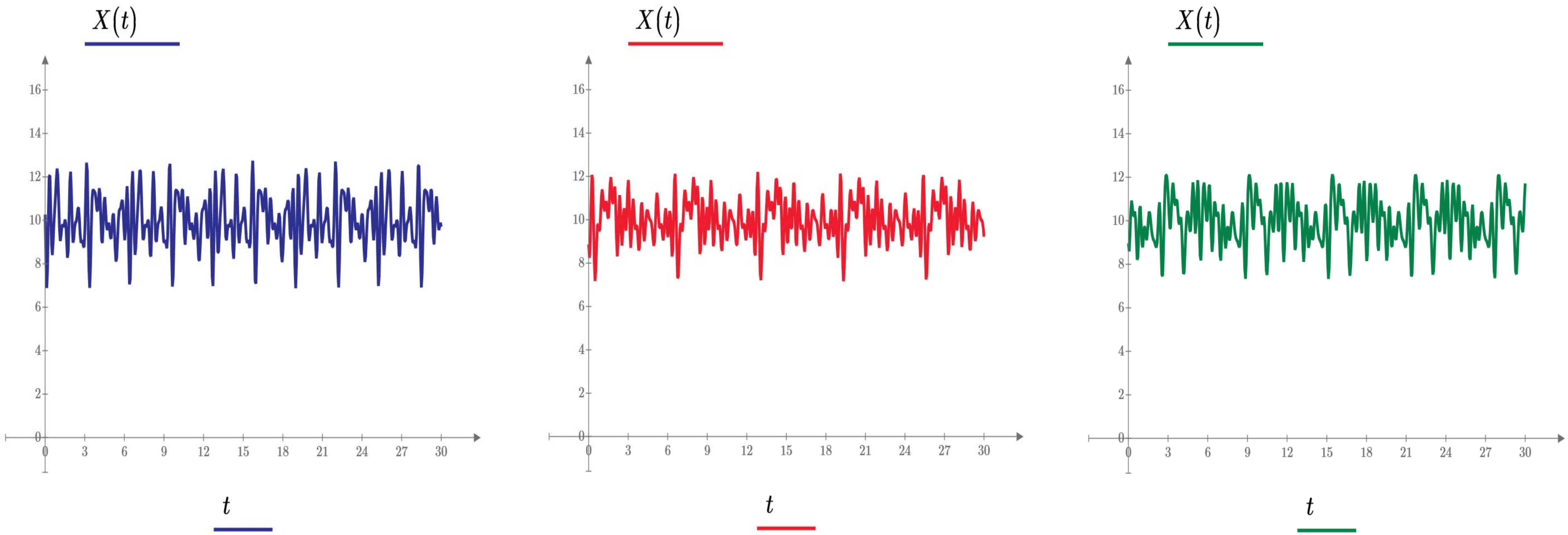


Stochastic Processes – Signals

Additive Noisemodel

$$observed\ signal = signal\ +\ noise$$

Three Realizations of the Stochastic Process



Stochastic Processes

Definitions:

- A stochastic process is a time dependent stochastic variable:

$$X(t)$$

- A discrete stochastic process is given by:

time

$$X[n] = X(nT)$$

where n is an integer.

Notice:

- When we sample a signal from a stochastic process, we observe only one realization of the process

Sample Functions

Definition:

- A sample function $x(t)$ is a realization of a stochastic process X

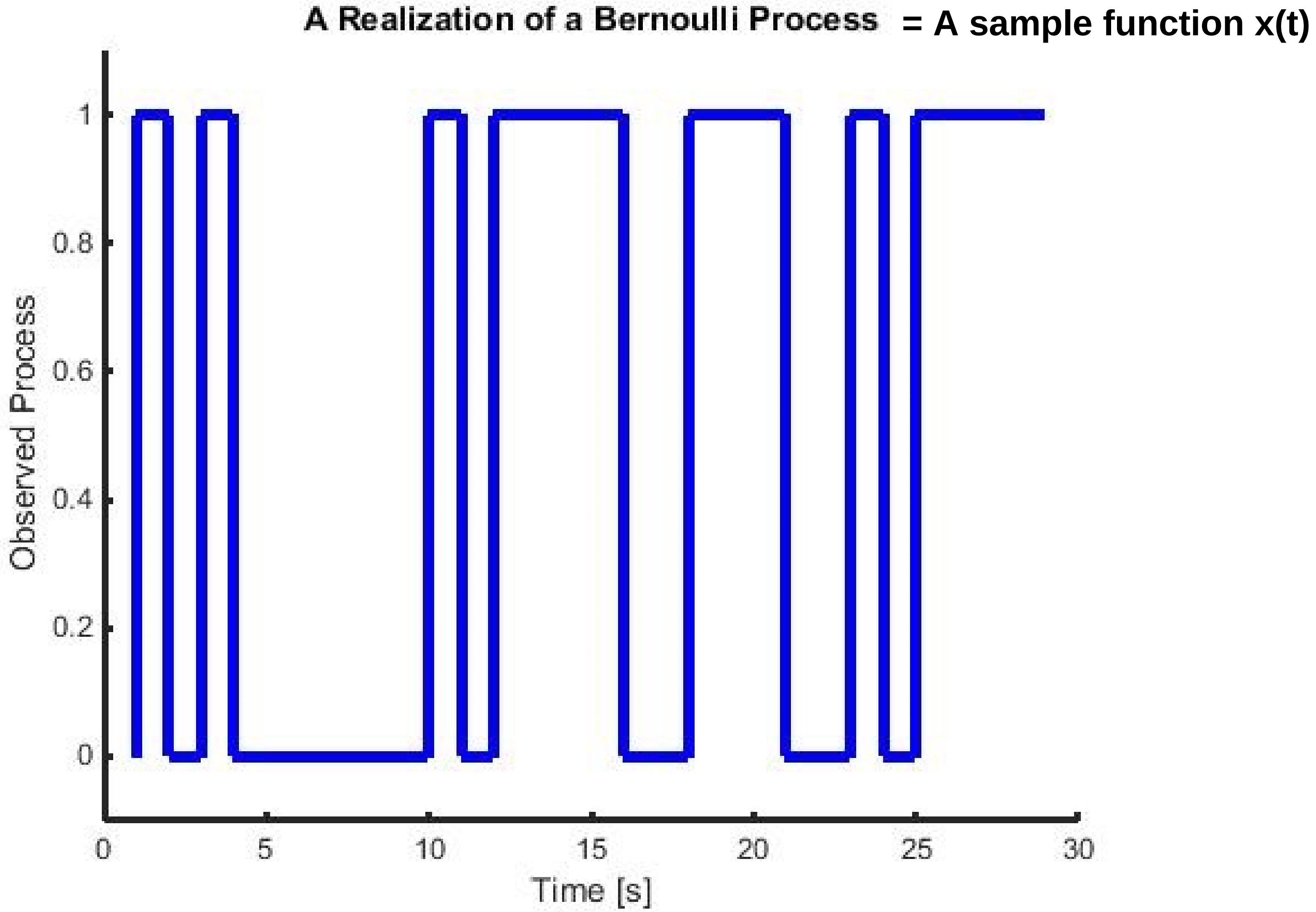
Example:

- A coin is thrown every minute: H = head, T = tail
- One realization of the stochastic signal is:

$$HTHT$$

Example – Random Binary (digital) Signal

- Bernoulli process.
- A sequence of 1 and 0s.
- Is a sequence of i.i.d of Bernoulli trials.



Ensemble

Definition:

- The Ensemble of the Stochastic Process is the collection of all possible realizations $x(t)$ of the Stochastic Process X

Example:

- A coin is thrown every minute: H = head, T = tail
- The Ensemble of the stochastic signals is:

$$HTHT, HHTT, TTHH, THTH, THHT, TTHT, HHHH \ldots$$

Time Dependent Probability Functions

- Probability density function (pdf):

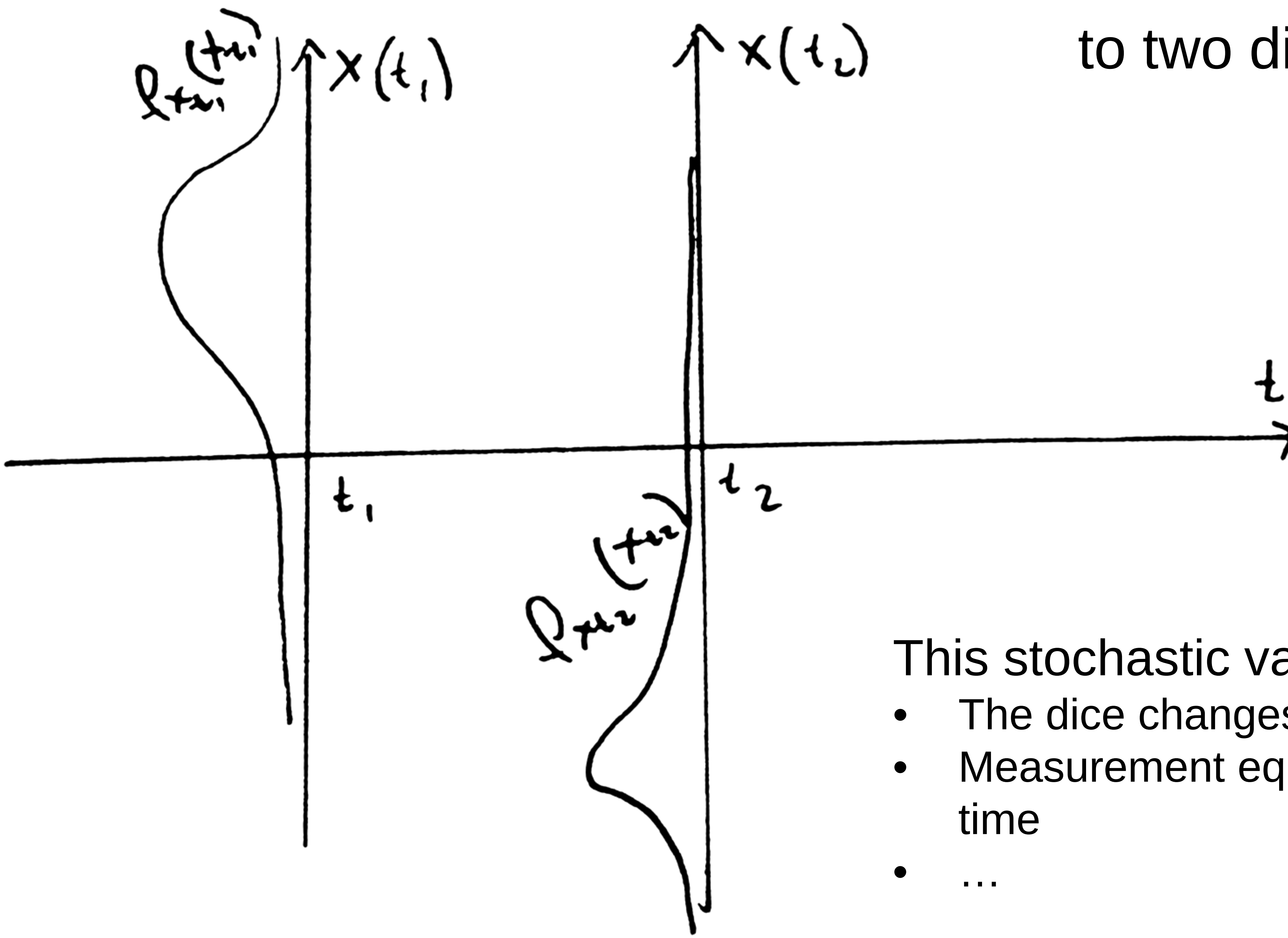
$$f_{X(t)}(x(t))$$

- Cumulative distribution function (cdf):

$$F_{X(t)}\big(x(t)\big) = \int_{-\infty}^{x(t)} f_{X(t)}(x(t))\ dx(t)$$

Time Dependent Stochastic Process

The same stochastic variable
to two different times

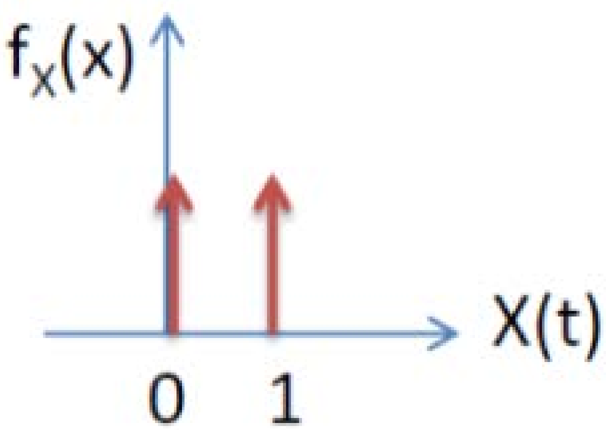


This stochastic variable is not i.i.d.:

- The dice changes it’s properties (wears)
- Measurement equipment changes with time
- …

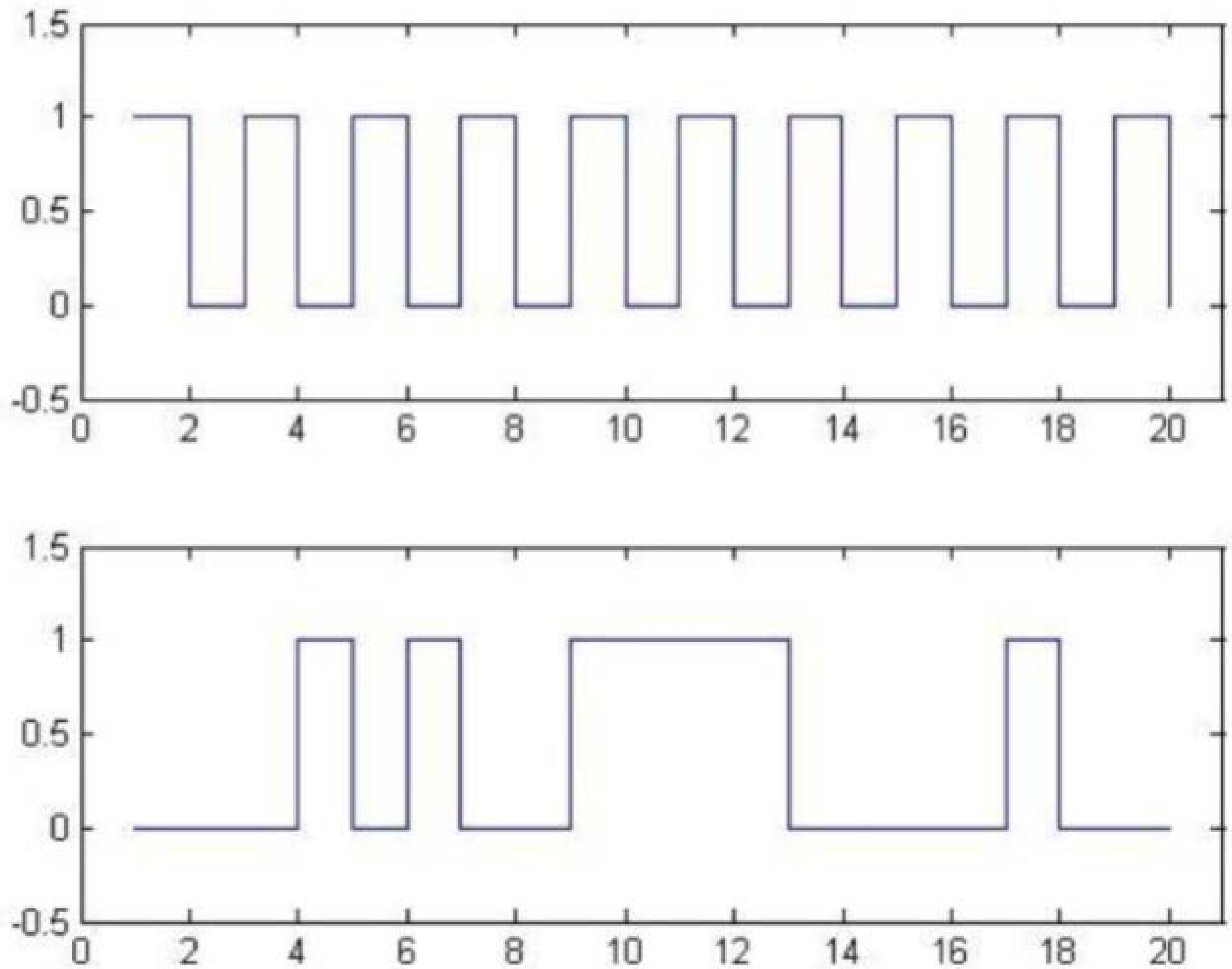
Deterministic Functions

- We find a sample function from a stochastic process.
- The two samples have the same pmf.



pmf

Deterministic



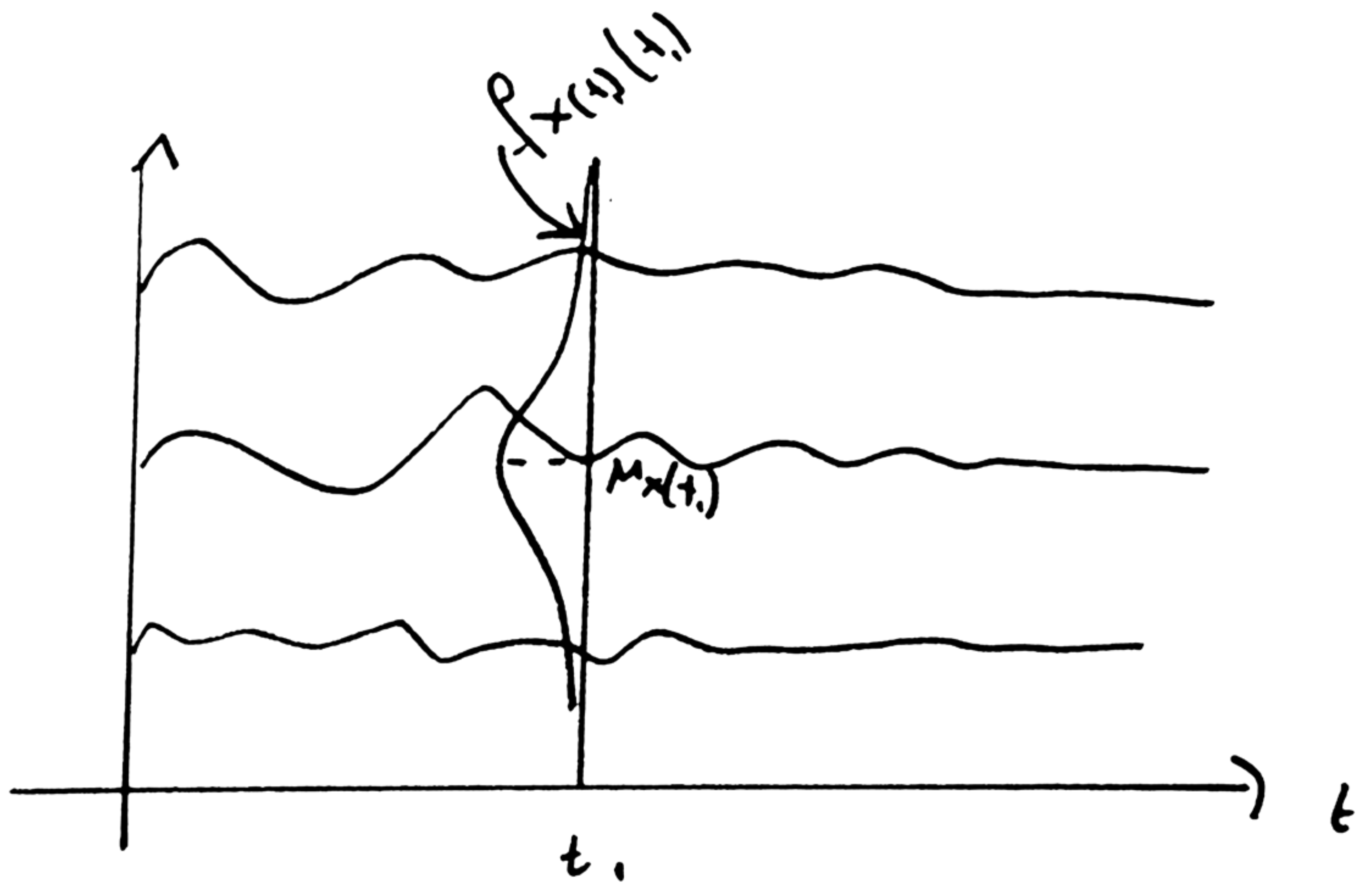
Random (non-deterministic)

Ensemble mean

- The mean value function:

$$\mu_{X(t)}(t) = E[X(t)] = \int_{-\infty}^{\infty} x(t) f_{X(t)}(x(t))\, dx(t)$$

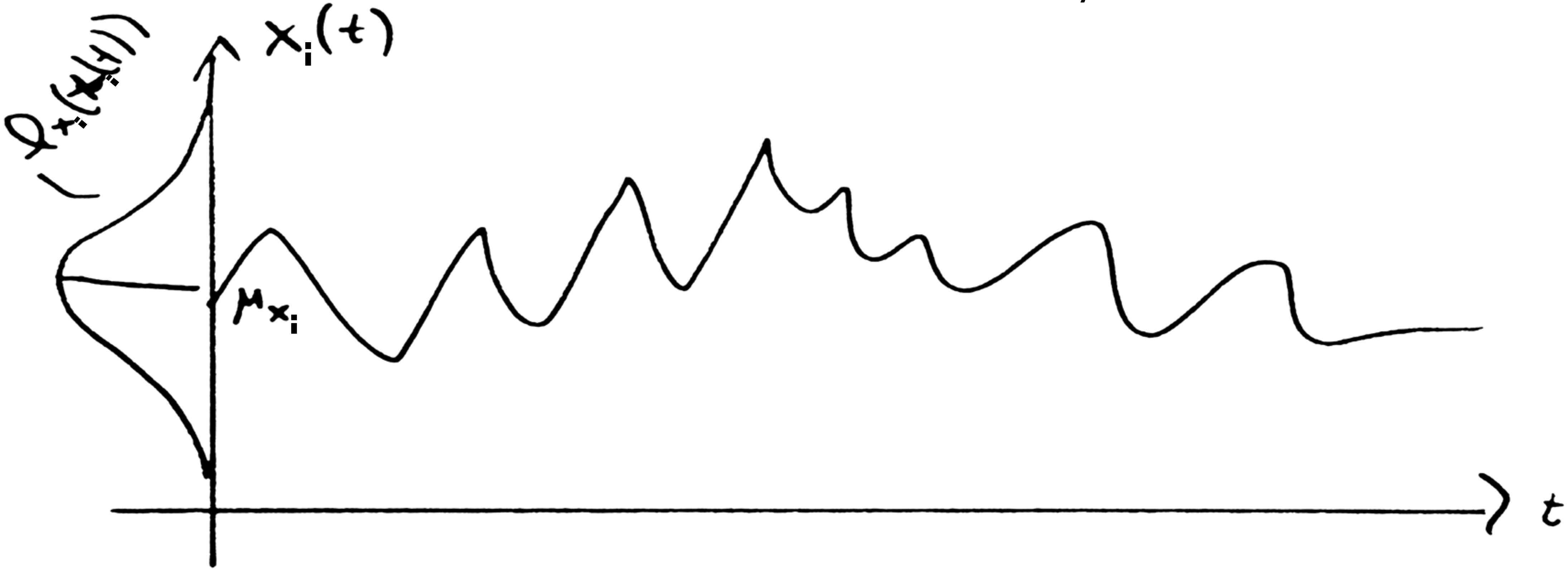
- The mean of <u>all possible realizations</u> to time t



Temporal Mean

- The time average for <u>one realization</u> of the stochastic process
- The temporal mean can differ from the ensemble mean

$$\hat{\mu}_{X_i} = \langle X_i \rangle_T = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt$$

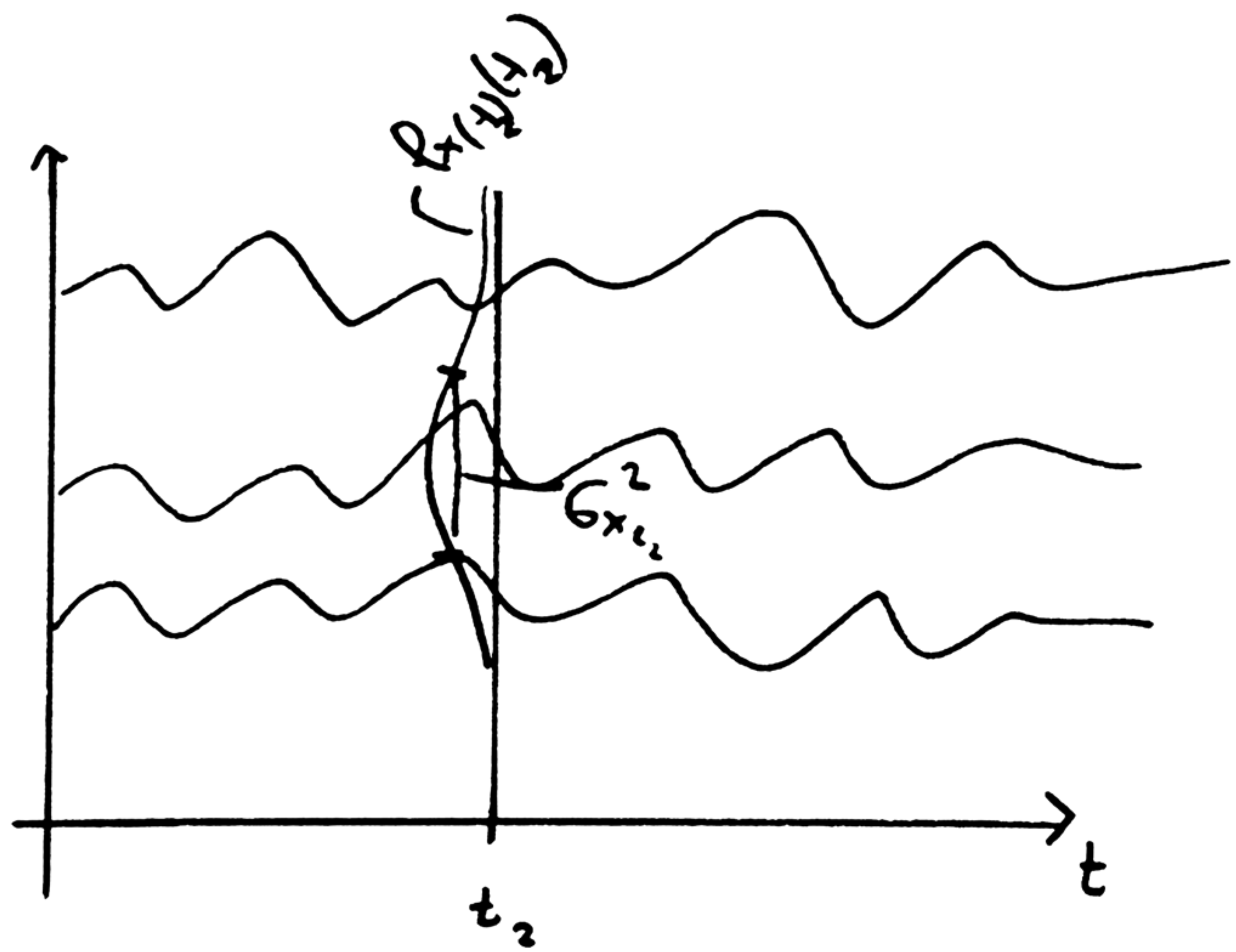


Ensemble Variance

- The variance function:

$$var\big(X(t)\big) = \sigma^2_{X(t)}(t) = E[\big(X(t) - \mu_{X(t)}(t)\big)^2]$$

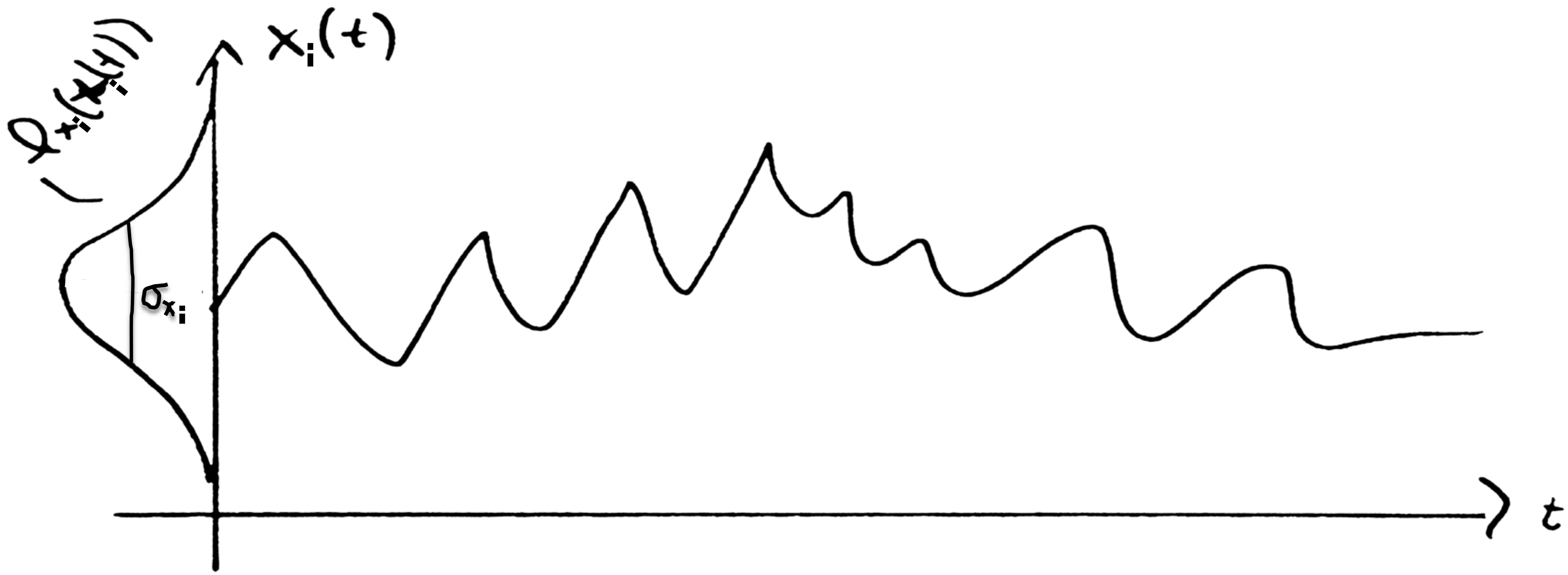
- The variance of <u>all possible realizations</u> to time t



Temporal Variance

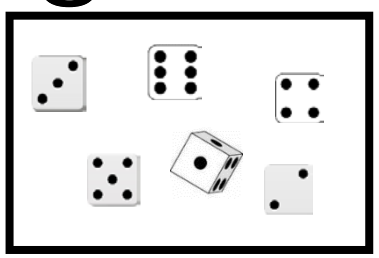
- The variance over time for <u>one realization</u> of the stochastic process
- The temporal variance can differ from the ensemble variance

$$\hat{\sigma}^2_{X_i} = \left\langle X_i^2 \right\rangle_T - \langle X_i \rangle_T^{\;2} = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \left(x_i(t)^2 - \hat{\mu}_{X_i}^{\;2}\right) dt = Var(X_i)$$



Stochastic Processes – Example

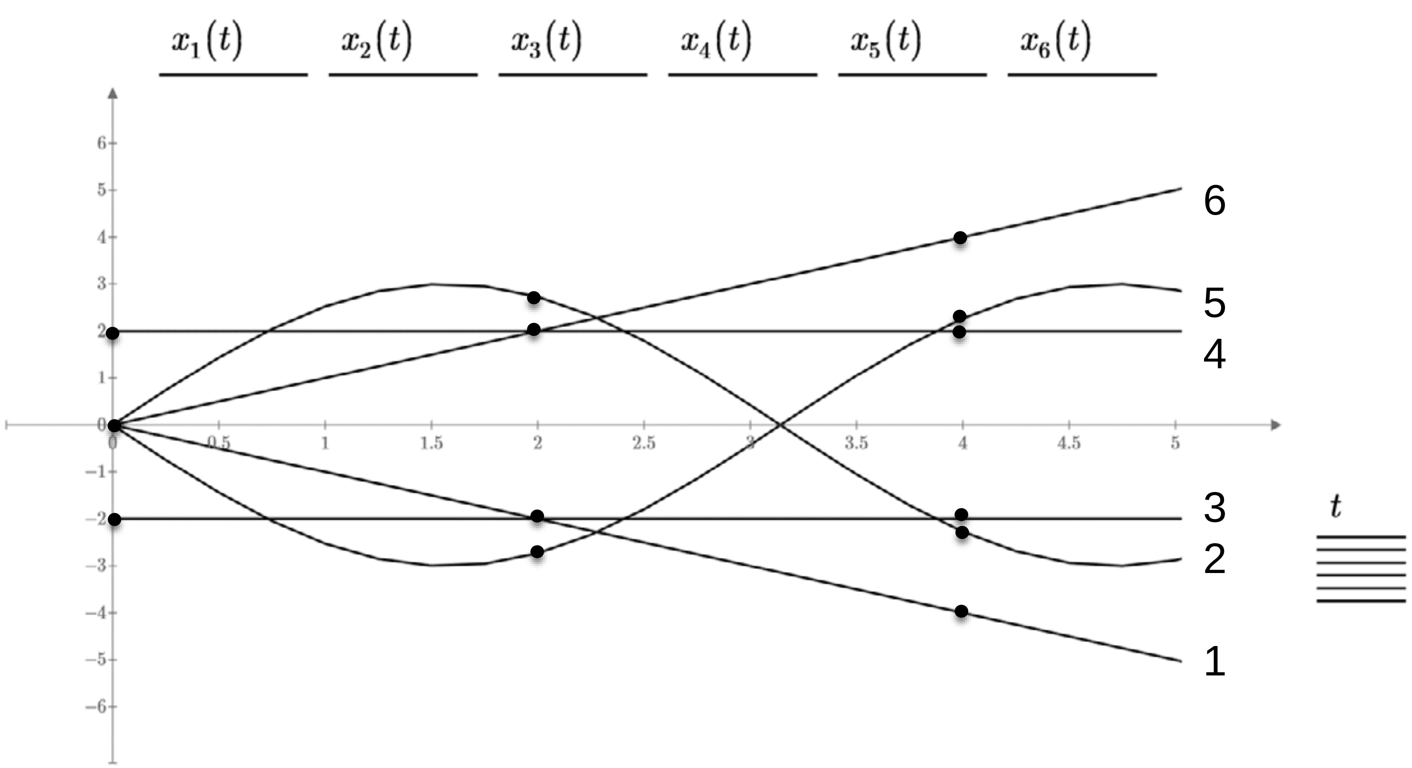
S



$X_n(t)$: $x_1(t) = -t \qquad x_2(t) = 3\sin(t)$

$x_3(t) = -2 \qquad x_4(t) = 2$

$x_5(t) = -3\sin(t) \qquad x_6(t) = t$



$X(0) = \{-2, 0, 2\}$

$X(2) = \{-2.7, -2, 2, 2.7\}$

$X(4) = \{-4, -2.3, -2, 2, 2.3, 4\}$

$\Pr(X(0) = 0) = 2/3$

$\Pr(X(2) = 2) = 1/3$

$\Pr(X(4) = -4) = 1/6$

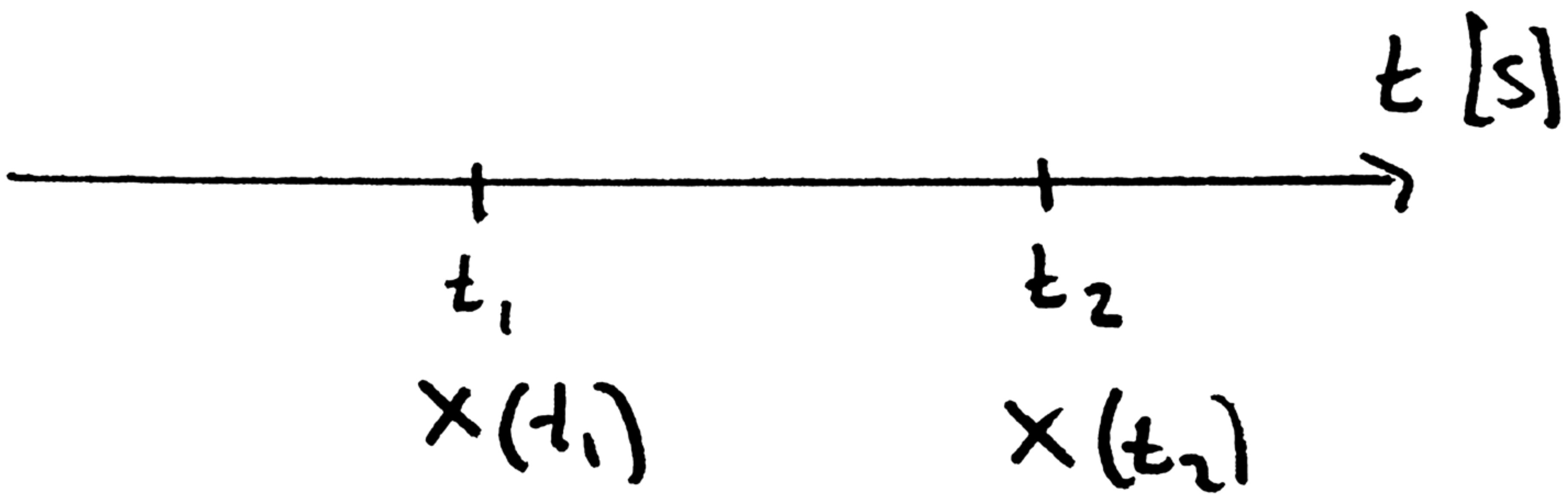
Ensemble: $\mu_{X(t)}(t) = E[X(t)] = 0 \qquad var\big(X(t)\big) = \sigma^2_{X(t)}(t) = \frac{1}{3}(t^2 + 9\sin(t) + 4)$

Temporale: $\hat{\mu}_{X_2} = 0 \qquad \hat{\mu}_{X_3} = -2 \qquad \hat{\sigma}^2_{X_2} = 4.5 \qquad \hat{\sigma}^2_{X_3} = 0$

Correlations

Comparing realizations

- Autocorrelation Correlation of a realization with itself
- Cross-correlations Correlation of two different realizations
- We compare the processes at two different times



Autocorrelations

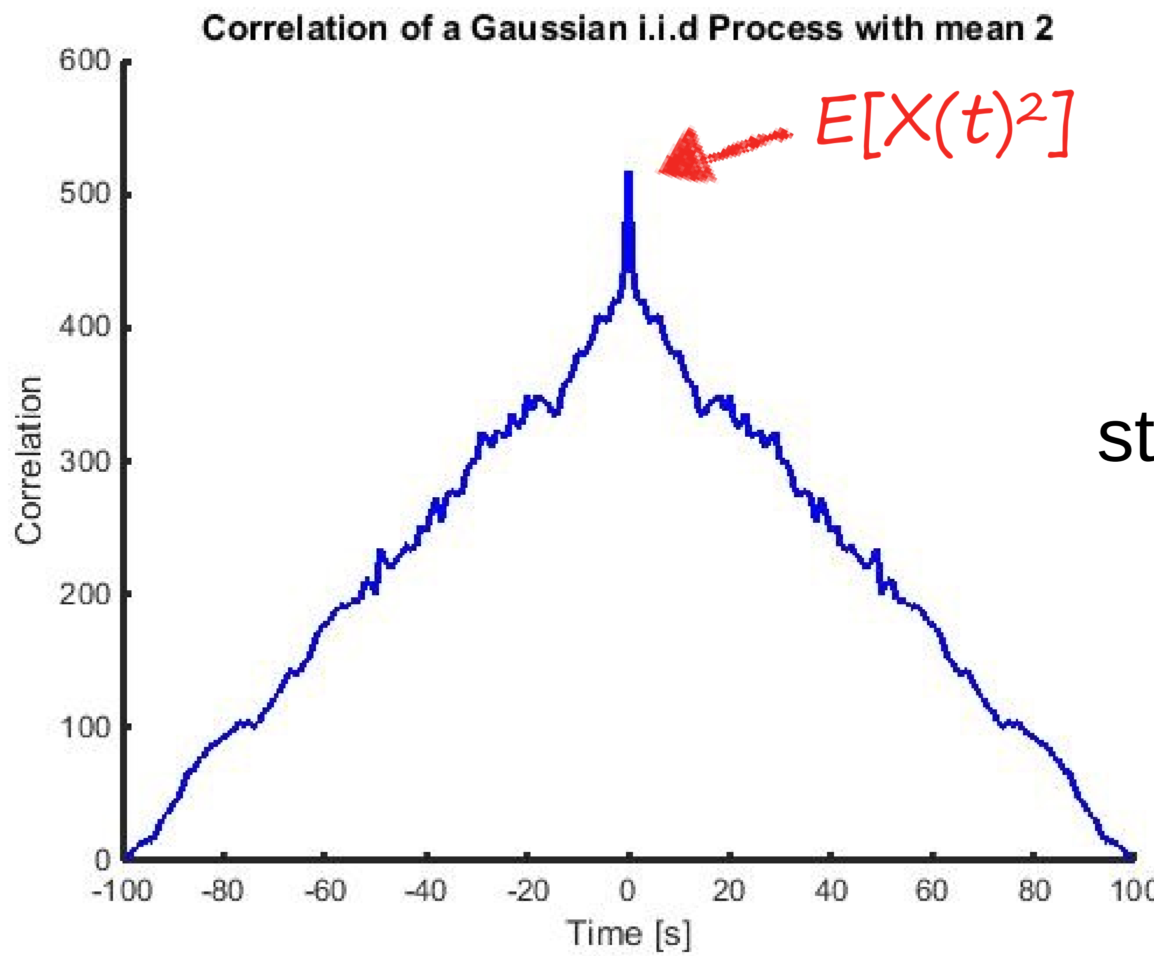
Tells about the connection at two different times

- Autocorrelation function:

Complex conjugated

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$$

$$= \iint_{-\infty}^{\infty} x(t_1)\, x(t_2)^* f_{X(t_1),X(t_2)}\big(x(t_1), x(t_2)\big) dx(t_1) dx(t_2)$$



Autocorrelation of a stationary process at time t_1 as a function of $\tau = t_1$-t_2

Autocovariances

Tells about how much we can predict the future

- Autocovariance function:

$$C_{XX}(t_1,t_2) = E\left[\left(X(t_1) - \mu_X(t_1)\right)\left(X(t_2) - \mu_X(t_2)\right)^*\right]$$
$$= R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X(t_2)$$

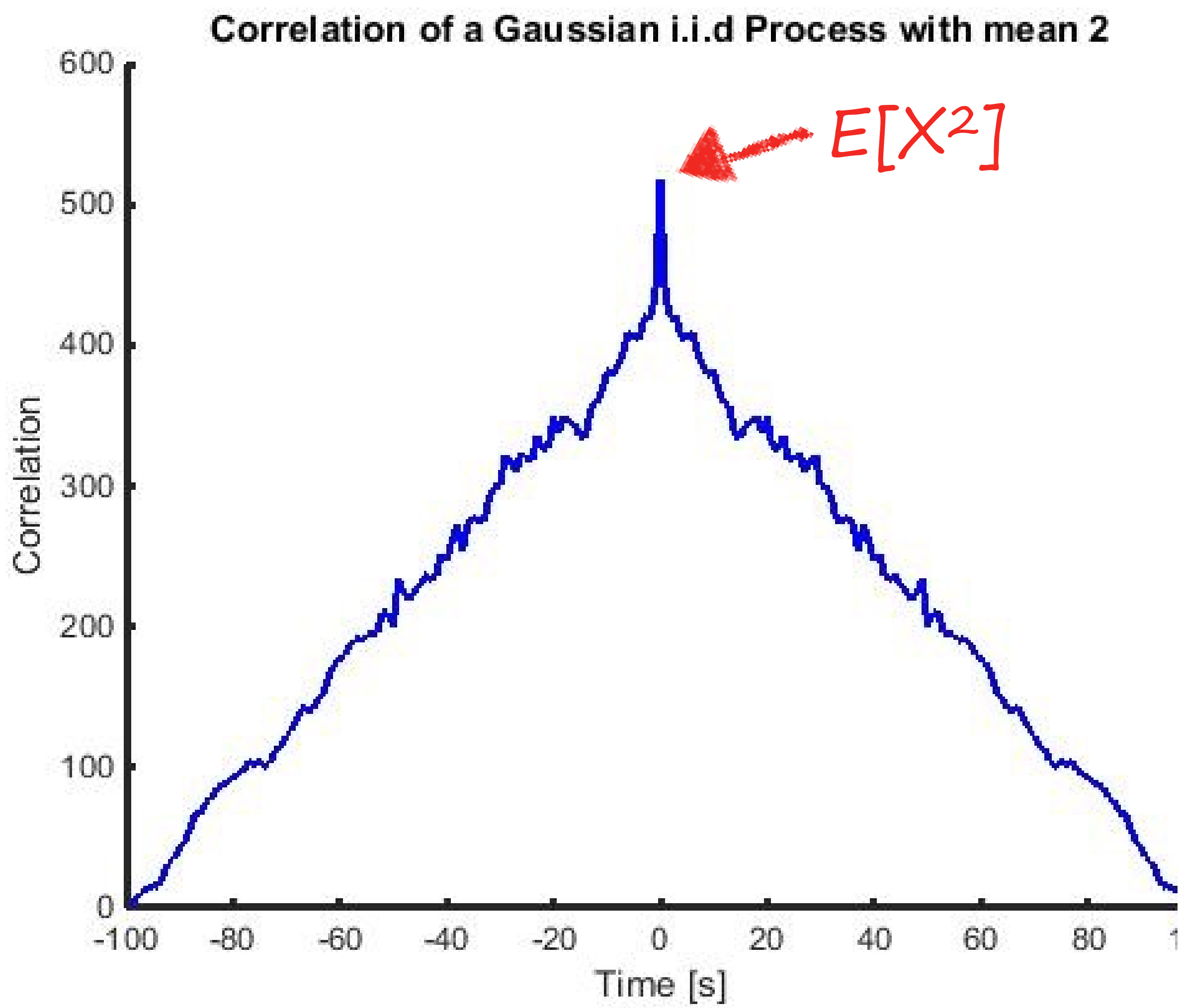
- Autocorrelation coefficient:

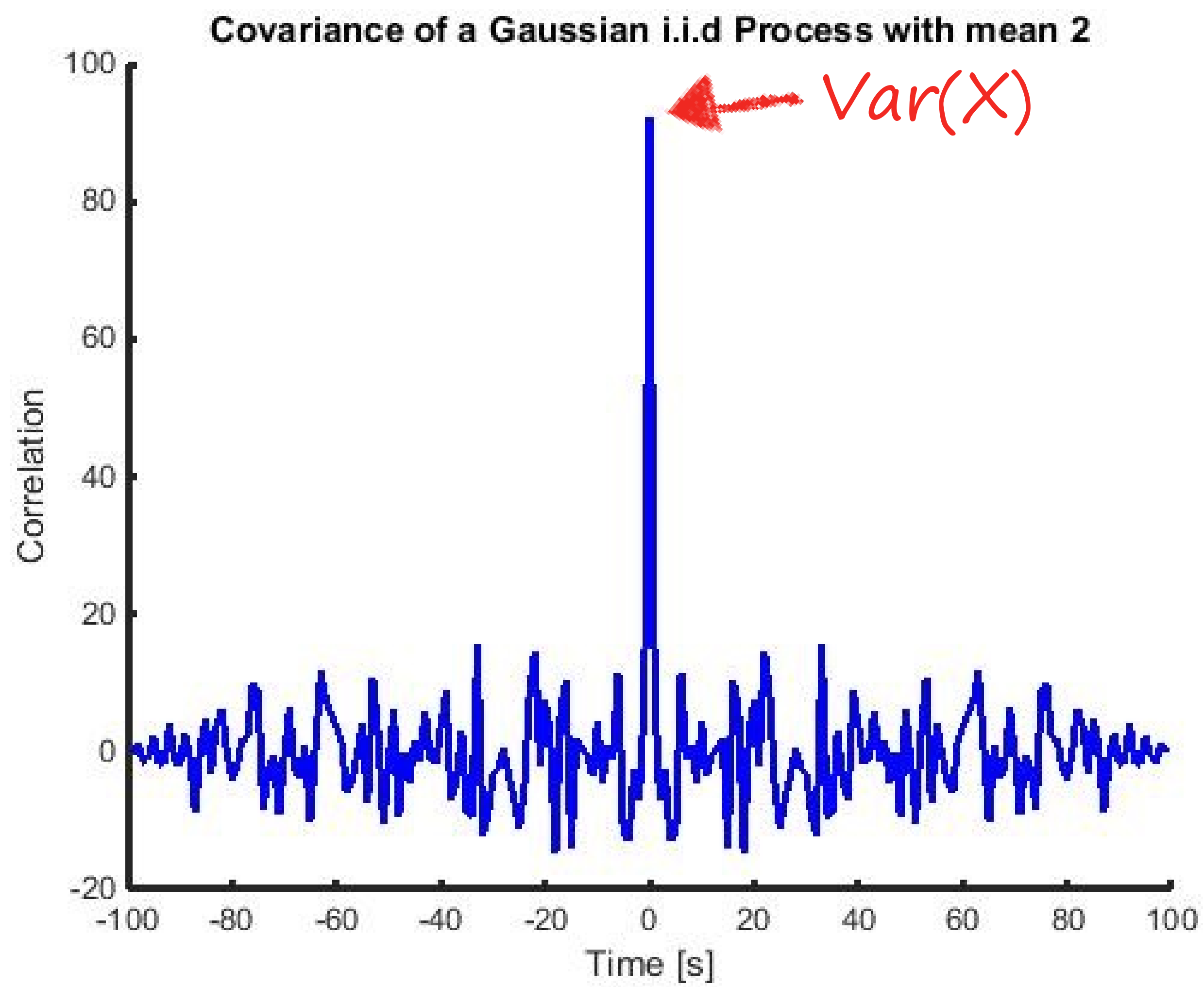
$$r_{XX}(t_1,t_2) = \frac{C_{XX}(t_1,t_2)}{\sqrt{C_{XX}(t_1,t_1)C_{XX}(t_2,t_2)}}; \qquad 0 \le r_{XX}(t_1,t_2) \le 1$$

Autocovariances

For i.i.d. Gaussian (stationary) noise

- Autocorrelation and autocovariance

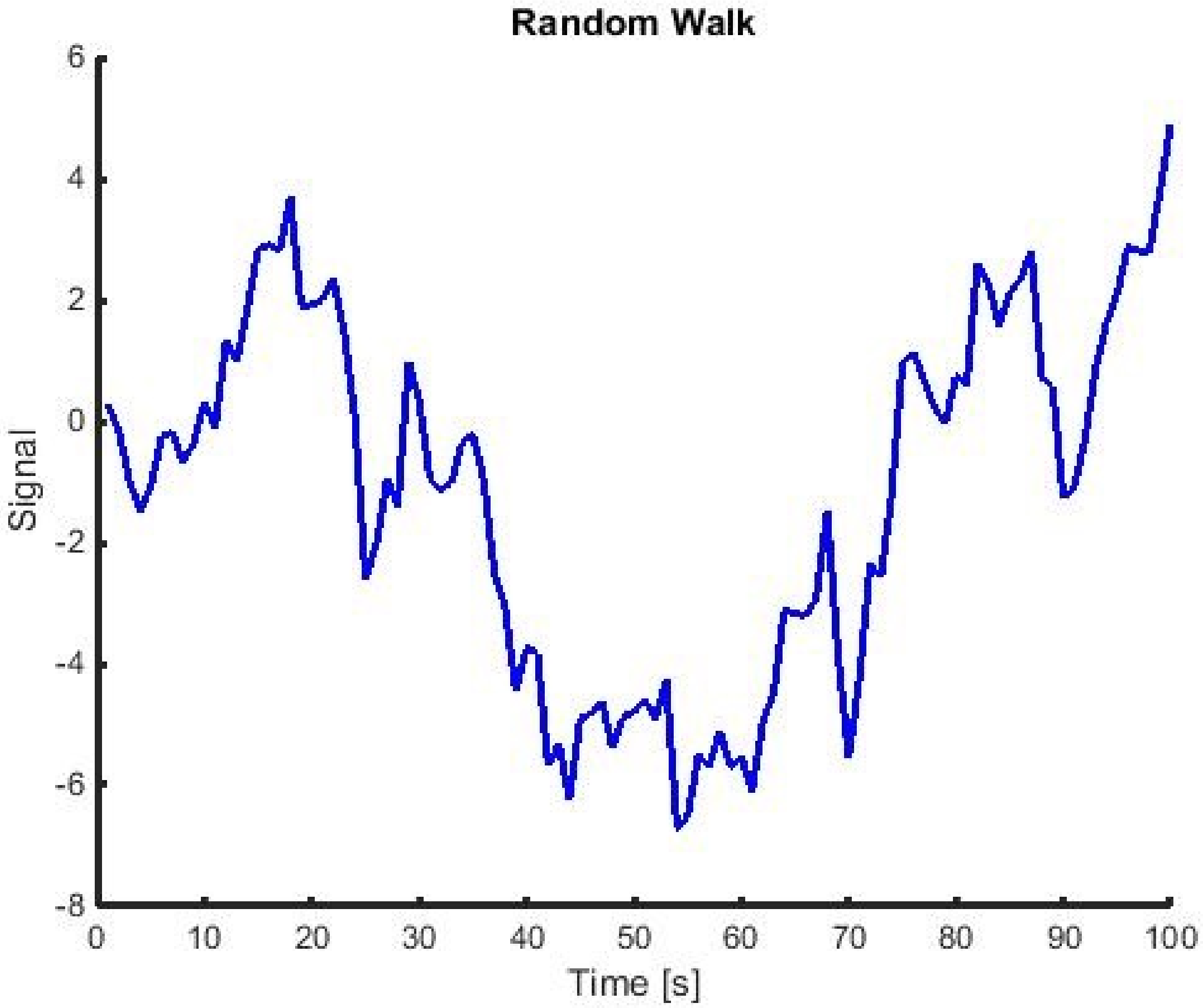




Random Walk – Example

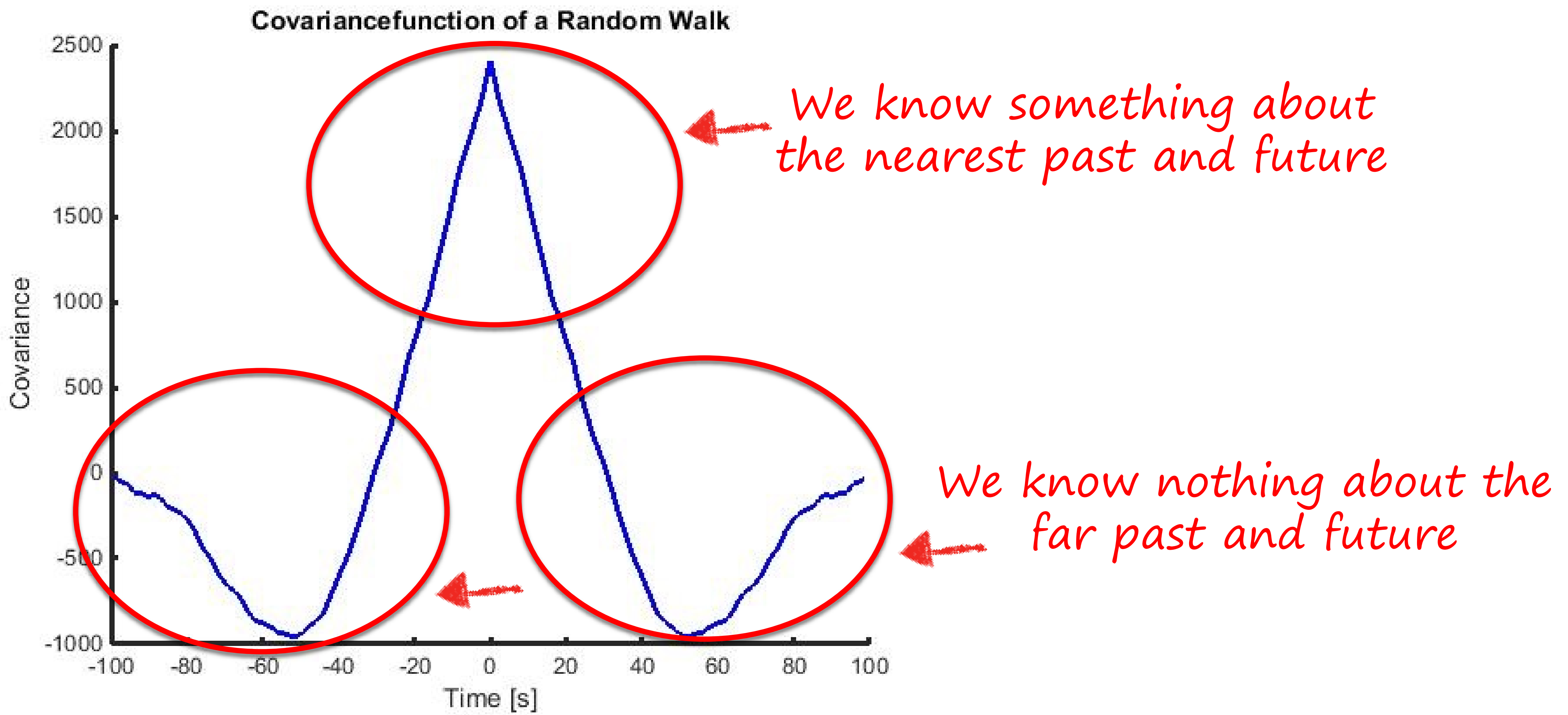
Brownish motions

- We consider a random walk.



Random Walk – Example

- Sample of the autocovariance function:



Stationarity in the Strict Sense (SSS)

Difficult to test in reality

- The density function $f_{X(t)}\big(x(t)\big)$ do not change with time
  - For all choices of t_1 and Δt_1, the marginal pdf:

$$f_{X(t_1)}\big(x(t_1)\big) = f_{X(t_1+\Delta t_1)}\big(x(t_1 + \Delta t_1)\big)$$

  - For all choices of t_1, t_2 and Δt, the simultaneous pdf:

$$f_{X(t_1),X(t_2)}\big(x(t_1), x(t_2)\big) = f_{X(t_1+\Delta t),X(t_2+\Delta t)}\big(x(t_1 + \Delta t), x(t_2 + \Delta t)\big)$$

Stationarity in the Wide Sense (WSS)

Can be tested

- Ensemble mean is a constant

$$\mu_X(t) = E[X(t)] = \mu_X \quad \text{- independent of time}$$

- Autocorrelation depends only on the time difference $\tau = t_2 - t_1$

$$R_{XX}(t, t+\tau) = E[X(t)X(t+\tau)] = R_{XX}(\tau) \quad \text{- independent of time}$$

→ Ensemble variance is a constant

$$\sigma_X^2(t) = E\left[X(t)^2\right] - E[X(t)]^2 = \sigma_X^2 \quad \text{- independent of time}$$

Ergodicity

- We can say something about the properties of the stochastic process in general based on one sample function, as long as we have observed it for long enough.

Example:

- An i.i.d Gaussian noise stream

Ergodicity

- If ensemble averaging is equivalent to temporal averaging:

$$\mu_X(t) = \bar{X}(t) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt = \langle X_i \rangle_T = \hat{\mu}_{X_i}$$

- For any moment:

In practice: n=2 (Variance)

$$\overline{X^n} = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i^n\,(t)\, dt$$

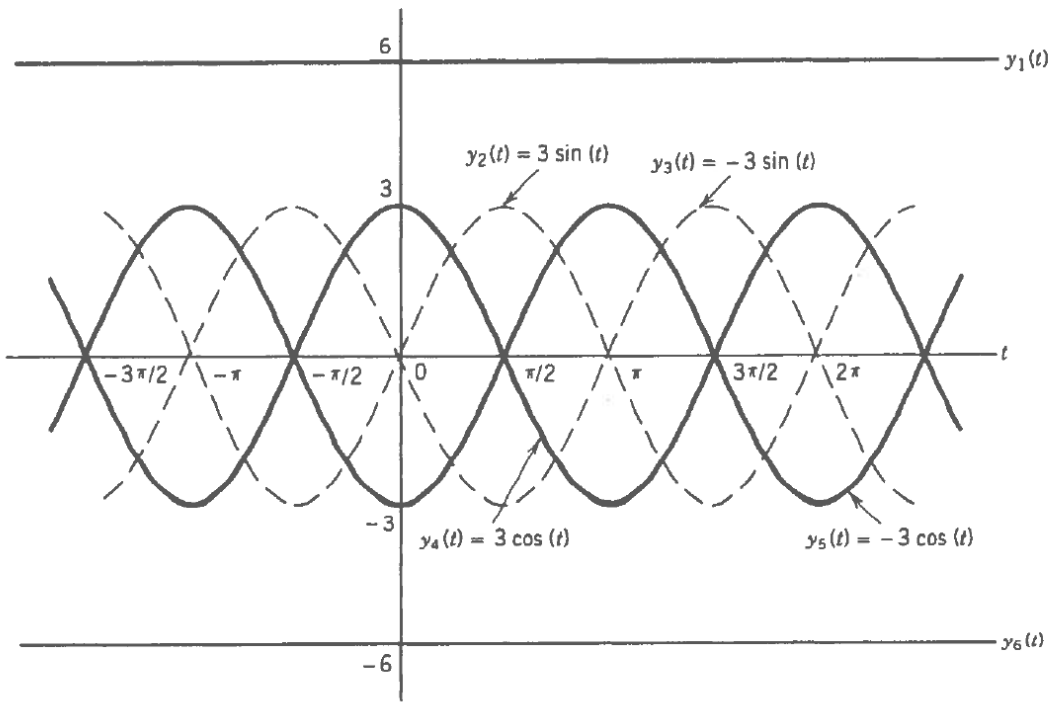
One (any) realization

Ensemple (WSS)

$$\left.\begin{array}{r} \langle X_i \rangle_T = \mu_X \\ \hat{\sigma}_{X_i}^2 = \sigma_X^2 \end{array}\right\} \rightarrow Ergodic$$

All information is achieved with one measurement (realization)

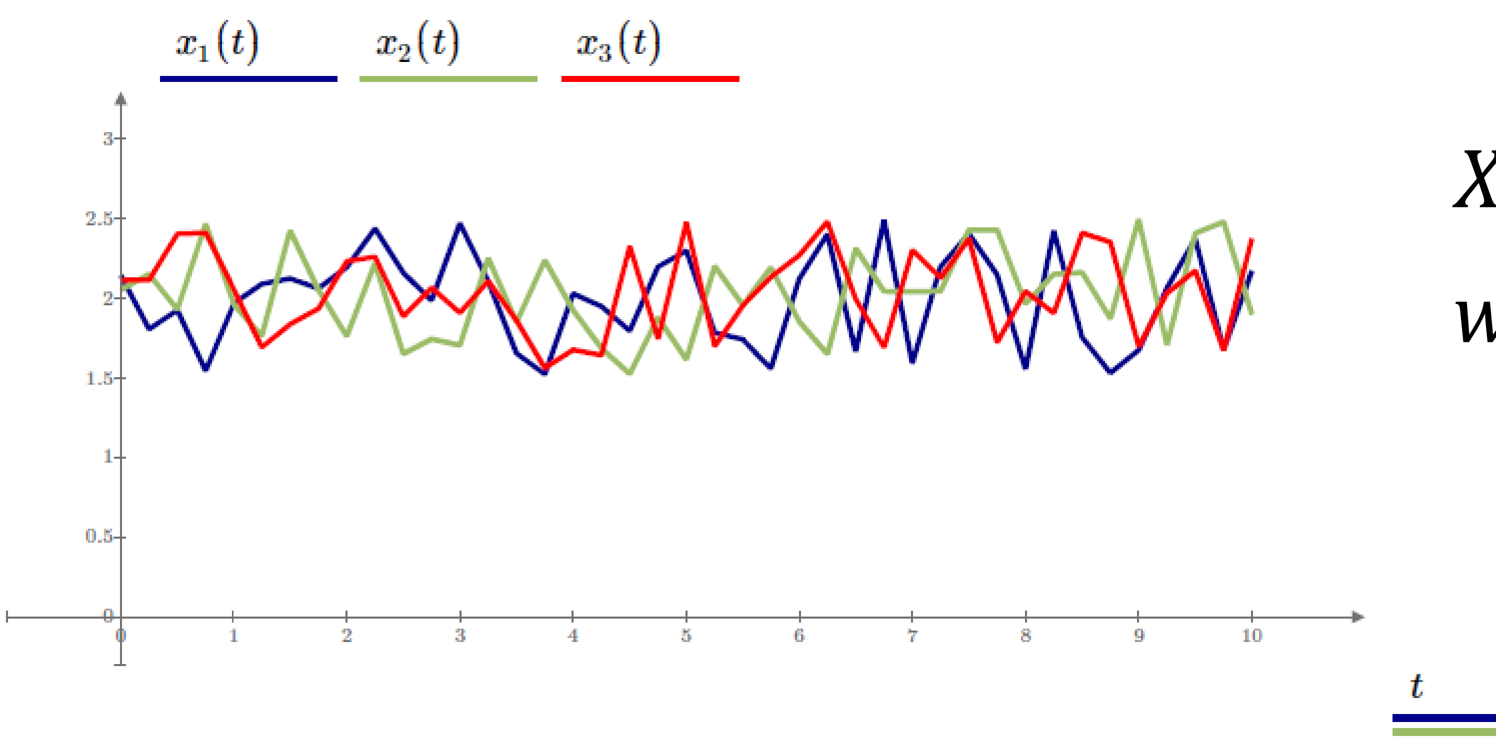
WSS and Ergodicity – Examples



% SSS

✓ WSS

% Ergodic



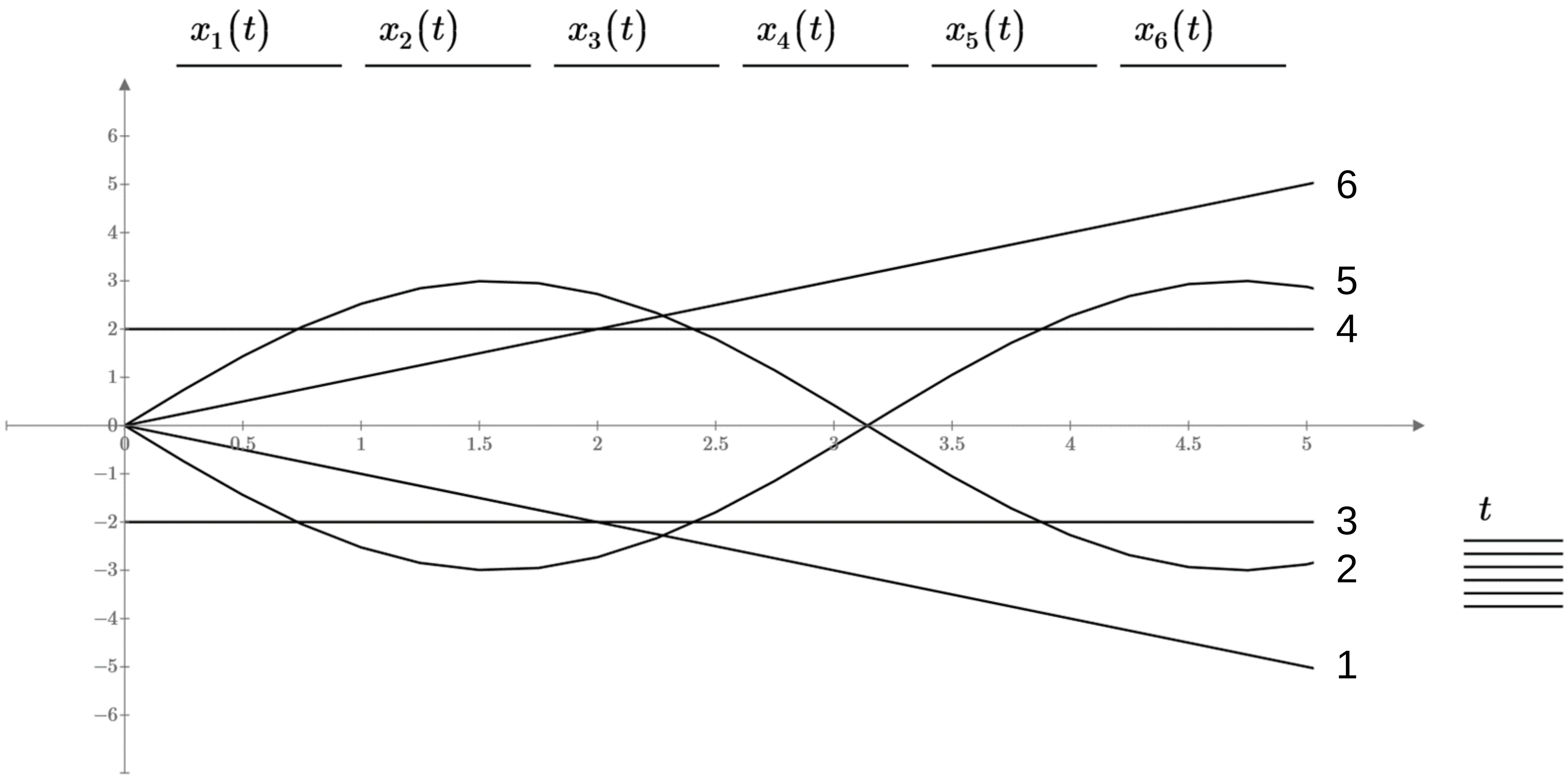
$$X_n(t) = 2 + w_n(t)$$

$$w_n(t) \sim \mathcal{U}[-0{,}5; 0{,}5]$$

✓ WSS

✓ Ergodic

WSS and Ergodicity– Example



- SSS ???
- WSS ???
- Ergodic ???

Words and Concepts to Know

Non-deterministic

Stochastic Processes

Ensemple variance

SSS

Stationarity

Temporal variance

Deterministic

Ergodicity

Autocovariance

WSS

Ensemple mean

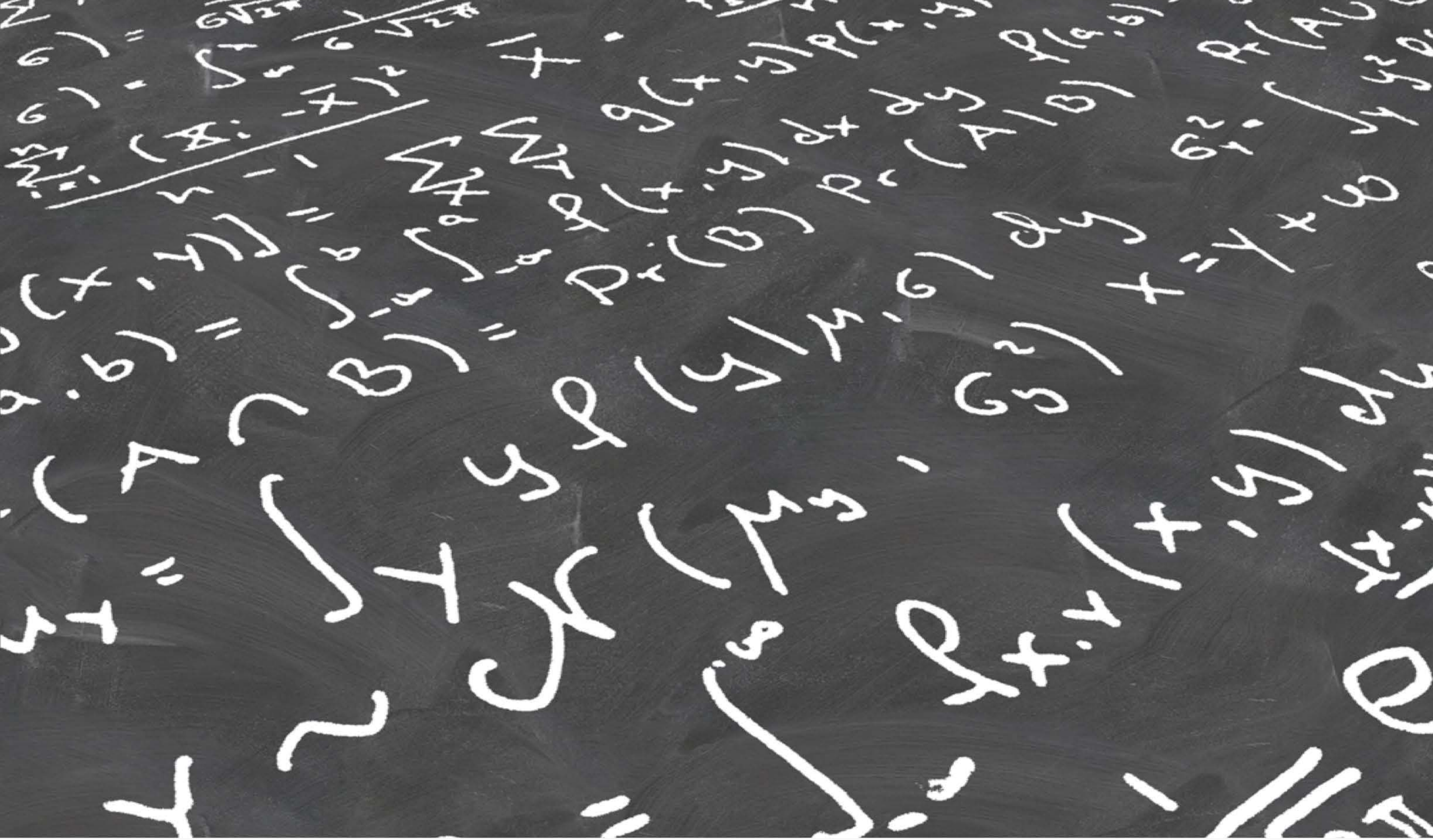
Autocorrelation

Strict Sense Stationary

Realization

Temporal mean

Wide Sense Stationary



7. Stochastic Processes and Correlation Functions

Gunvor Elisabeth Kirkelund
Lars Mandrup

Agenda for Today

- Stochastic Processes (repetition)
  - Mean and variance
  - Stationarity (WSS, SSS)
  - Ergodic Processes
- Correlation functions
  - Autocorrelation functions
  - Cross-correlation functions
- Power spectrum density

Stochastic Processes

Definitions:

- A stochastic process is a <u>time dependent</u> stochastic variable:

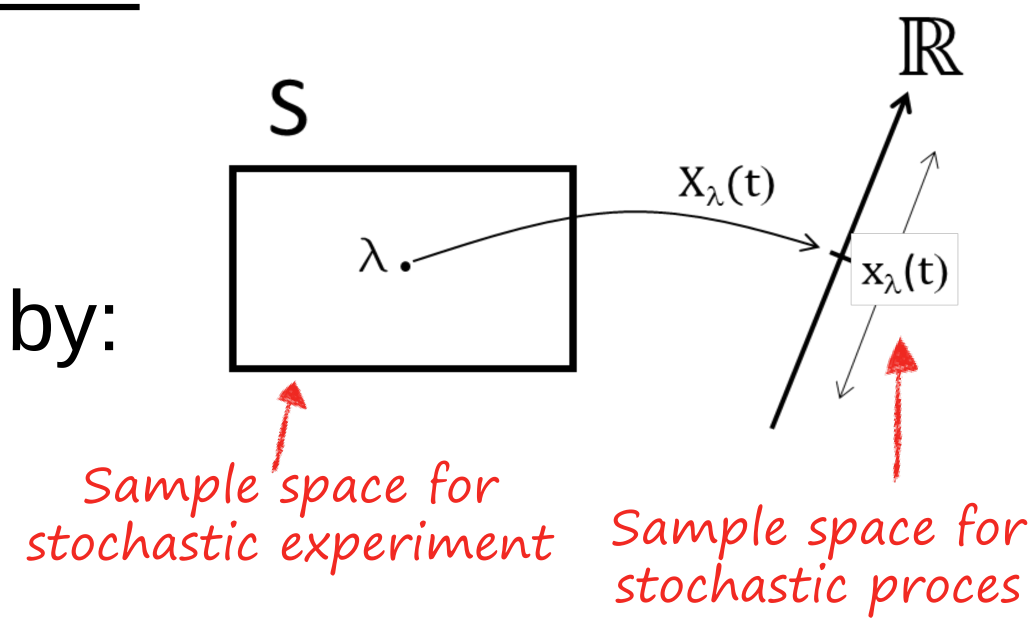
$$X(t)$$

- A discrete stochastic process is given by:

$$X[n] = X(nT)$$

where n is an integer.

- Random events that develops in time



Notice:

- When we sample a signal from a stochastic process, we observe only one <u>realization</u> of the process

Sample Function / Ensemble

Definitions:

- A sample function is a realization of a stochastic process $x(t)$
- The ensemble of a stochastic process is the collection of all possible realizations $x(t)$ of the stochastic process X

Example:

- A coin is thrown every minute: H = head, T = tail
- One realization of the stochastic signal is: $HTHT$ ...
- The ensemble of the stochastic signals is:

$HTHT, HHTT, TTHH, THTH, THHT, TTHT, HHHH\ldots$

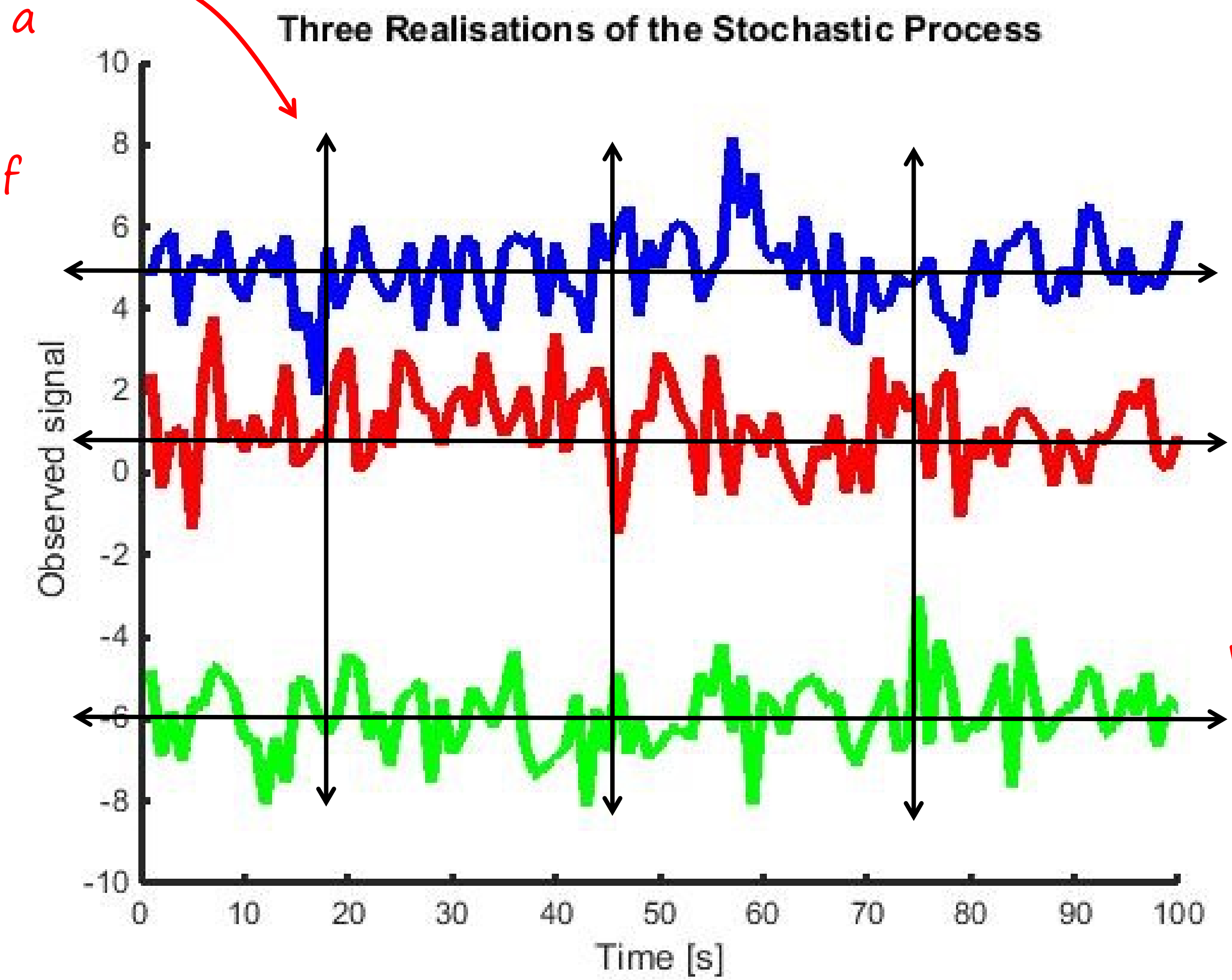
Stochastic Processes (signals)

Additive Noisemodel

$$observed\ signal = signal\ +\ noise$$

Ensemble mean and variance (to a specific time).

If indepentent of time: WSS



Time average and variance of each realization.

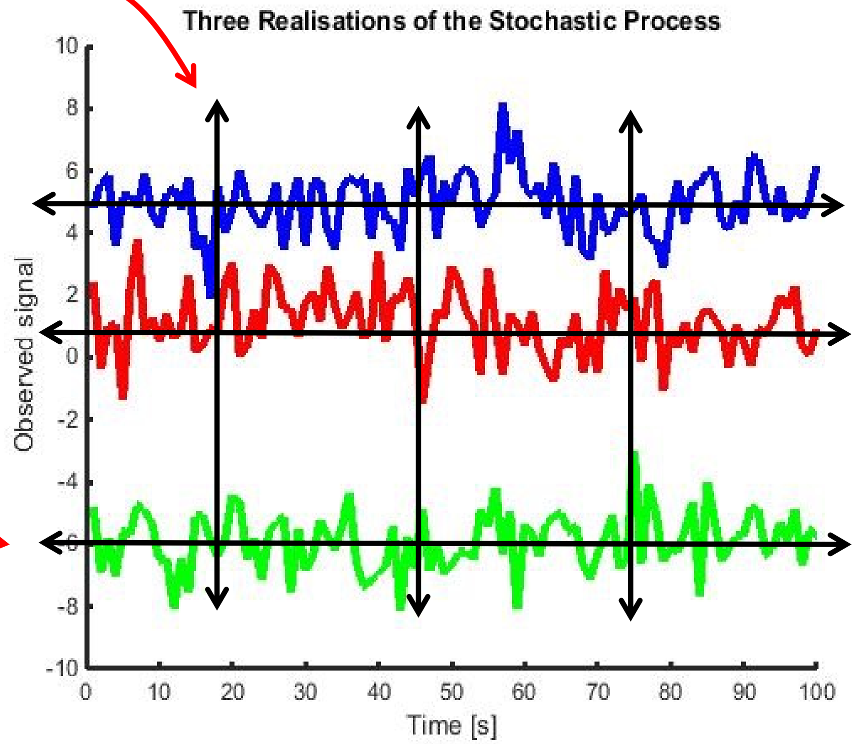
If equal (for all realizations): Ergodic

The Mean Functions

- Ensemple mean:

The mean of all possible realizations to time t

$$\mu_{X(t)}(t) = E[X(t)] = \int_{-\infty}^{\infty} x(t) f_{X(t)}(x(t))\, dx(t)$$



The time average for one realization of the stochastic process

- Temporal mean:

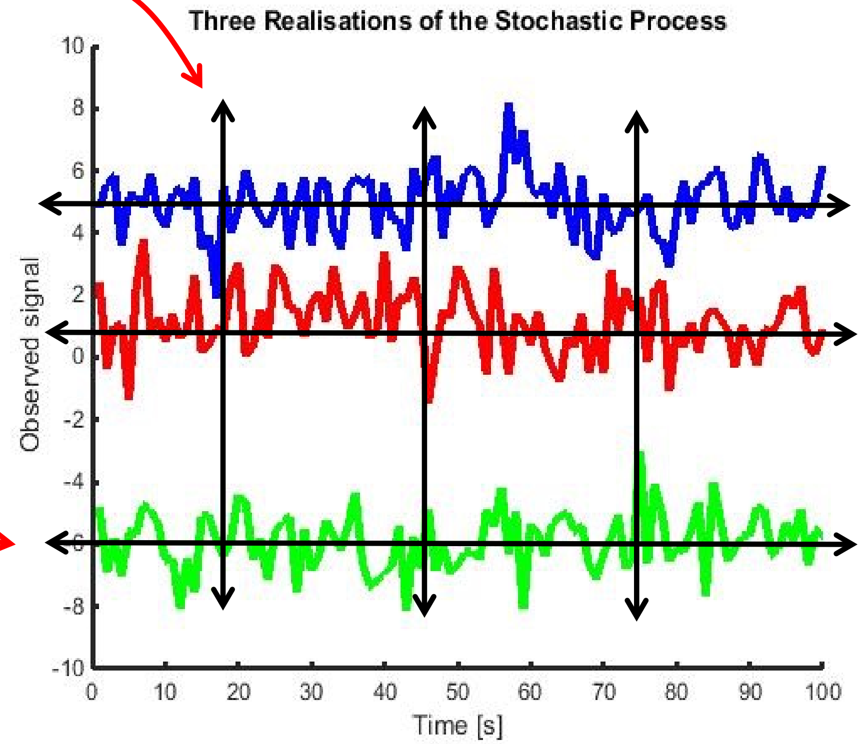
$$\hat{\mu}_{X_i} = \langle X_i \rangle_T = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt$$

The Variance Functions

The variance of all possible realizations to time t

- Ensemple variance:

$$Var\big(X(t)\big) = \sigma_{X(t)}^2(t) = E\Big[\big(X(t) - \mu_{X(t)}(t)\big)^2\Big]$$



The variance over time for one realization of the stochastic process

- Temporal variance:

$$\hat{\sigma}_{X_i}^2 = \langle X_i^2 \rangle_T - {\langle X_i \rangle_T}^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left(x_i(t)^2 - {\hat{\mu}_{X_i}}^2\right) dt = Var(X_i)$$

Stationarity in the Wide Sense (WSS)

- Ensemble mean is a constant

Can be tested.

$$\mu_X(t) = E[X(t)] = \mu_X$$ - independent of time

- Ensemble variance is a constant

$$\sigma_X^2(t) = E\left[X(t)^2\right] - E[X(t)]^2 = \sigma_X^2$$ - independent of time

Stationarity in the Strict Sense (SSS):

- The density function $f_{X(t)}\left(x(t)\right)$ do not change with time

Difficult to test in reality.

Ergodicity

- We can say something about the properties of the stochastic process in general <u>based on one sample function</u>, as long as we have observed it for long enough.

- If ensemble averaging is equivalent to temporal averaging:

Any realization

Ensemble (WSS)

$$\left.\begin{aligned} \langle X_i \rangle_T &= \mu_X \\ \hat{\sigma}^2_{X_i} &= \sigma^2_X \end{aligned}\right\} \rightarrow Ergodic$$

All information is achieved with one measurement (realization)

Realizations / Samples - Example

Discrete stochastic process:

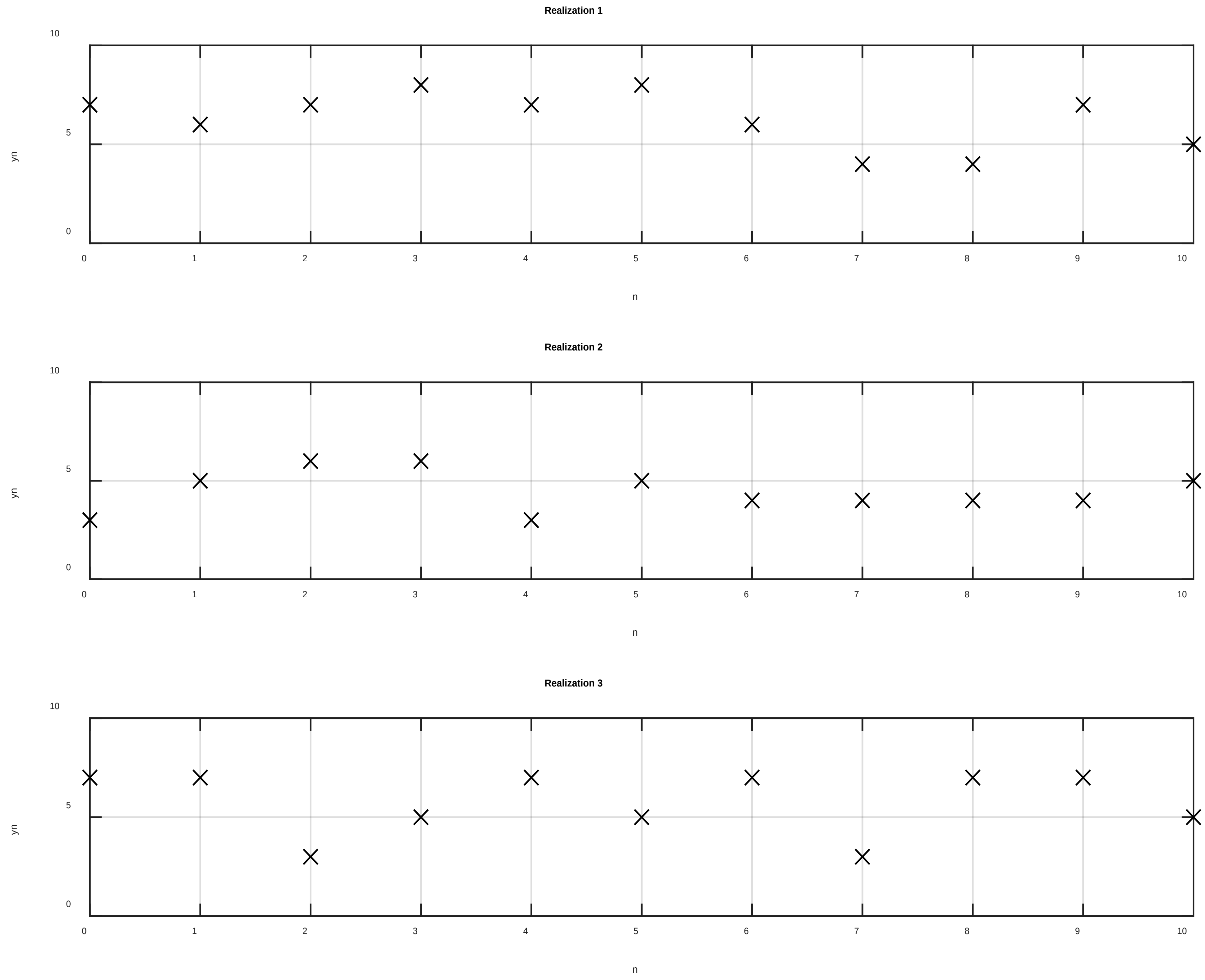
$$Y(n) = X + W(n);$$

$$X \sim \mathcal{B}(10, 0.5)$$
$$W(n) \sim \mathcal{U}_i[-2, 2]$$

3 realizations
11 samples
(n=0,..,10)

WWS

Ergodic



Realizations / Samples - Example

Discrete stochastic process:

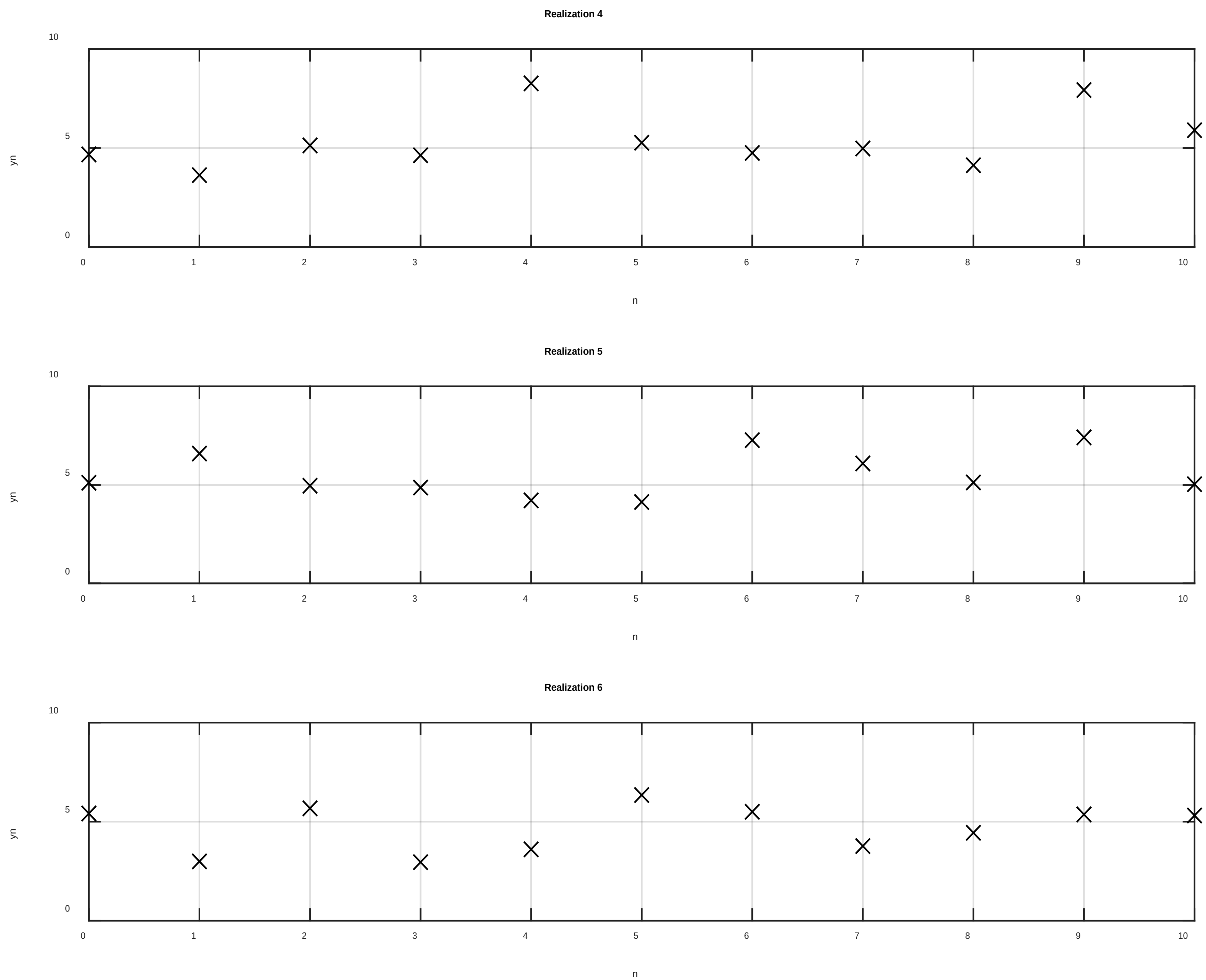
$Y(n) = W(n)$;

$W(n) \sim \mathcal{N}(5,2)$

3 realizations
11 samples
(n=0,..,10)

WWS ✓

Ergodic ✓



Realizations / Samples - Example

Continous
stochastic process:

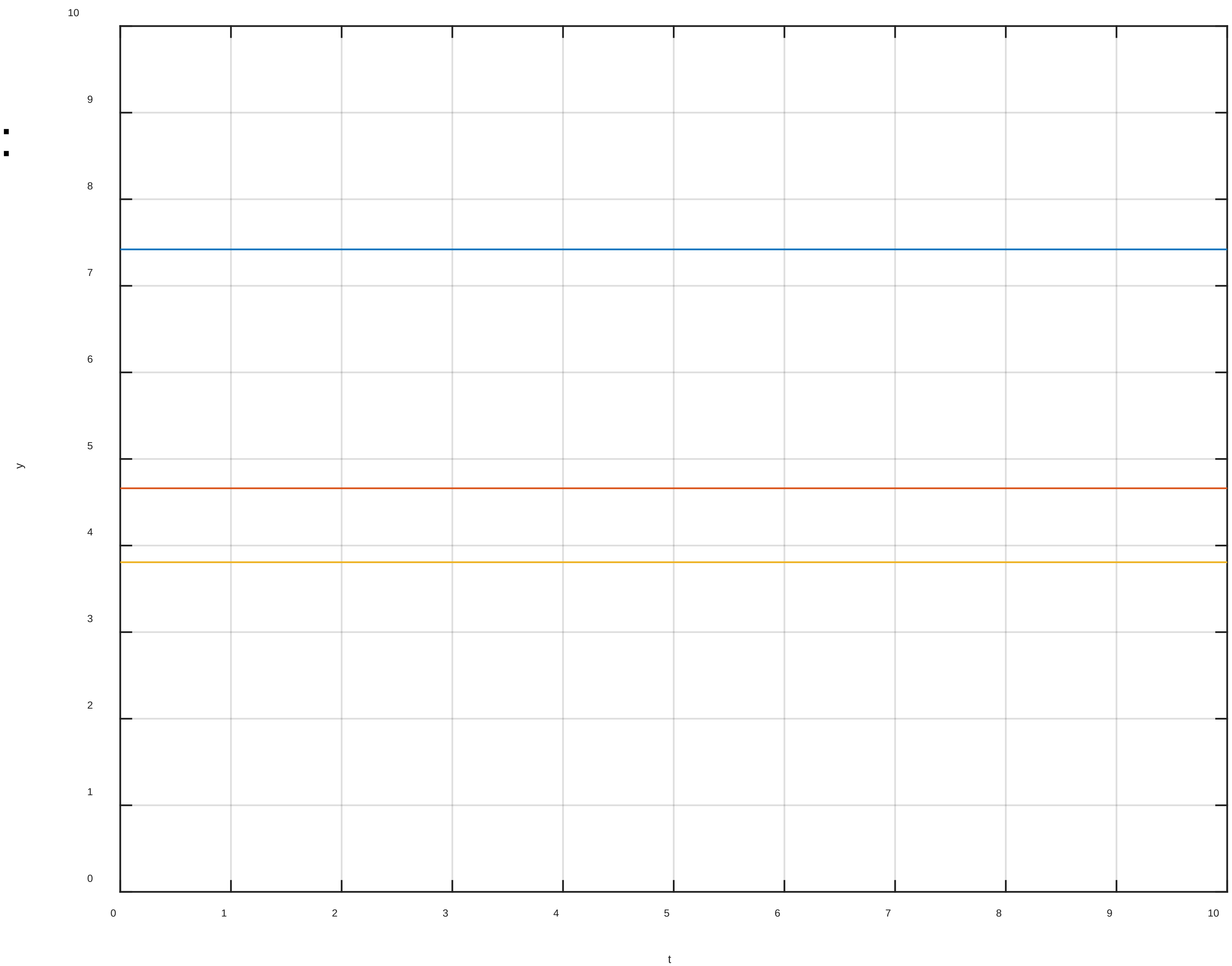
$Y(t) = W;$

$W \sim \mathcal{N}(5,2)$

3 realizations
$0 \leq \mathrm{t} \leq 10$

WWS ✓

Ergodic ÷



Realizations / Samples - Example

Continous stochastic process:

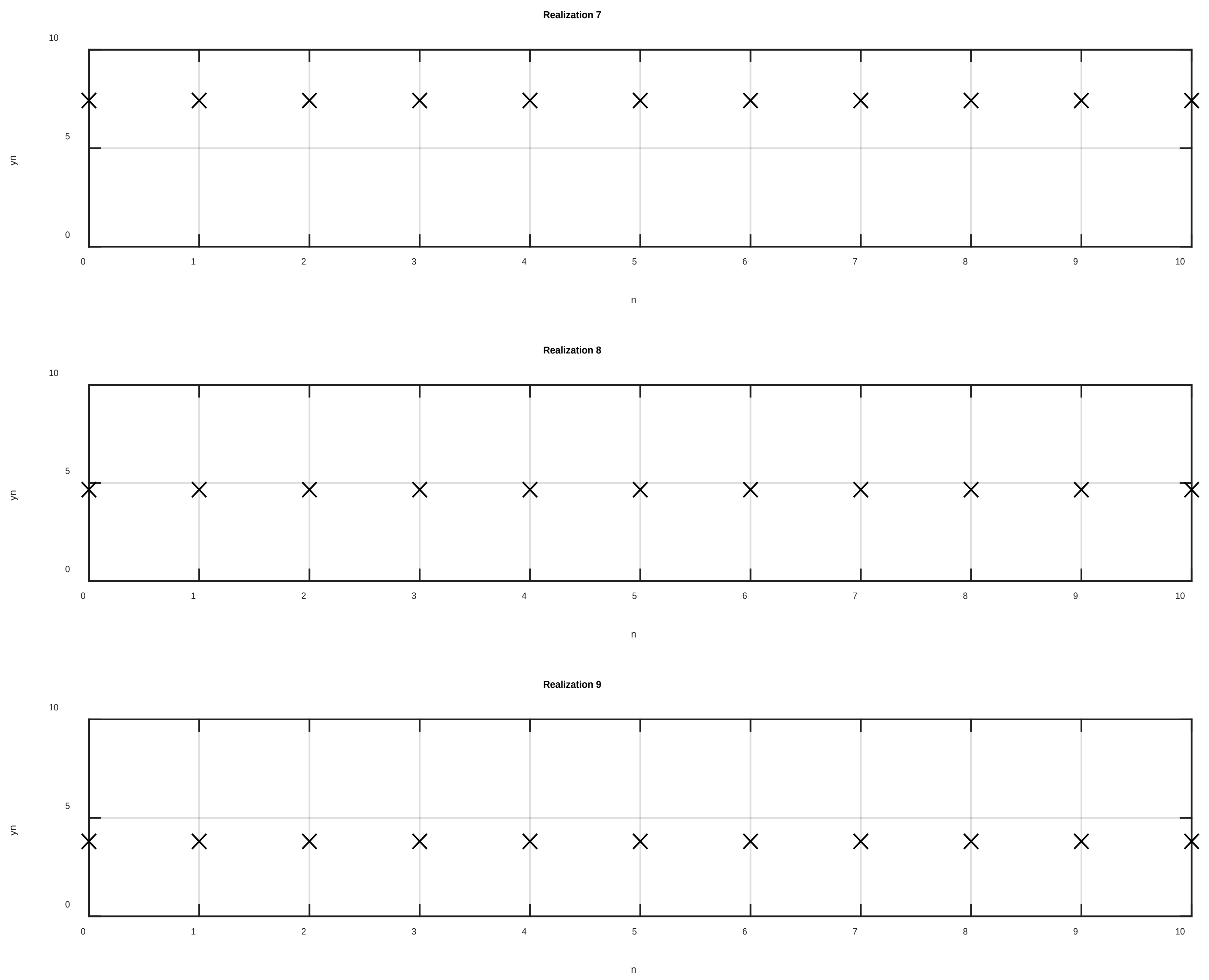
$Y(t) = W;$

$W \sim \mathcal{N}(5,2)$

3 realizations
11 samples
(n=0,..,10)

WWS ✓

Ergodic ÷



Realizations / Samples - Example

Continous stochastic process:

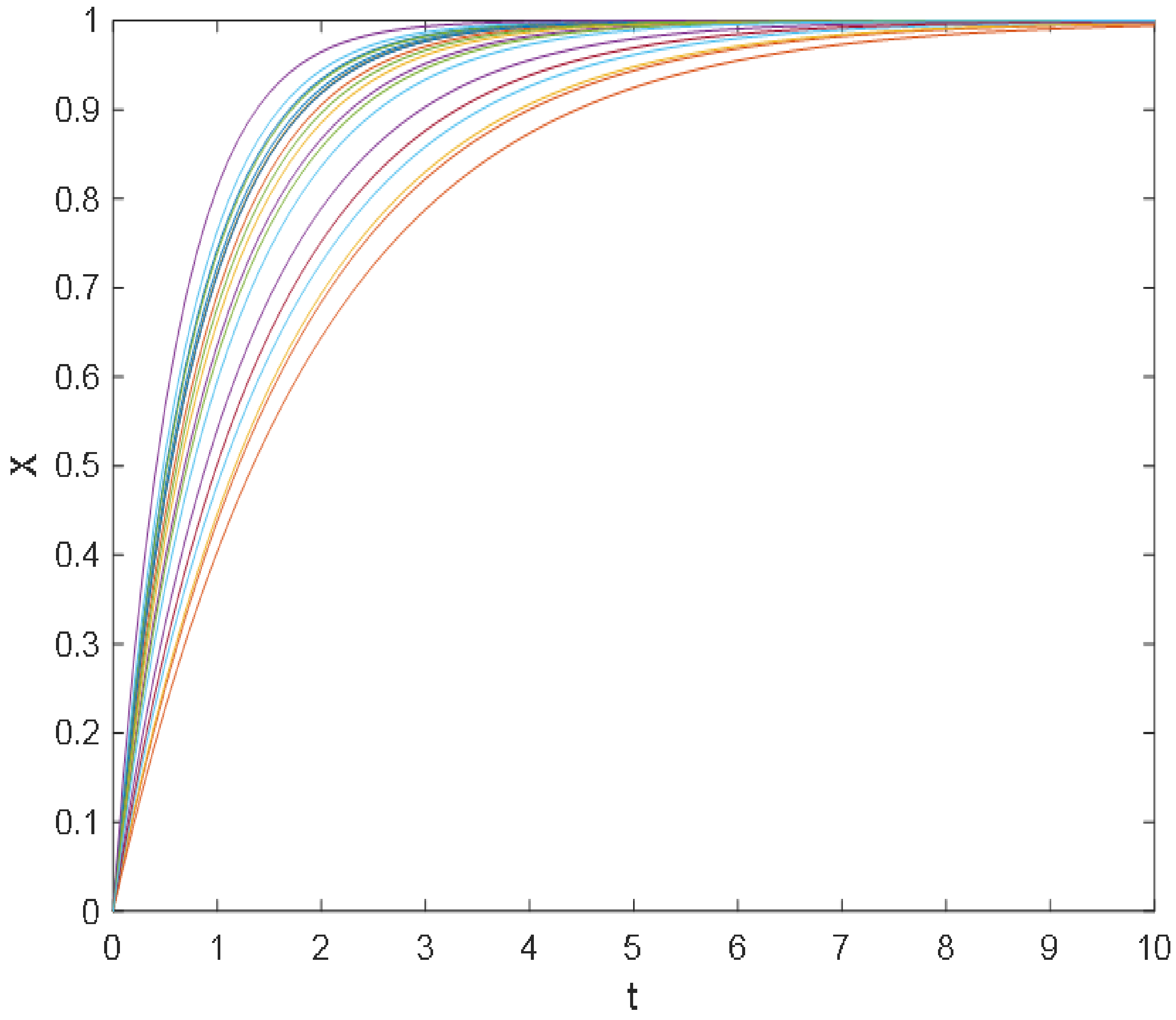
$\mathrm{X}(t) = \mathrm{A}(1 - e^{-k \cdot t});$

$A = 1;\ k \sim \mathcal{N}(1,0.4)$

20 realizations
$0 \leq \mathrm{t} \leq 10$

WWS $\div$

Ergodic $\div$



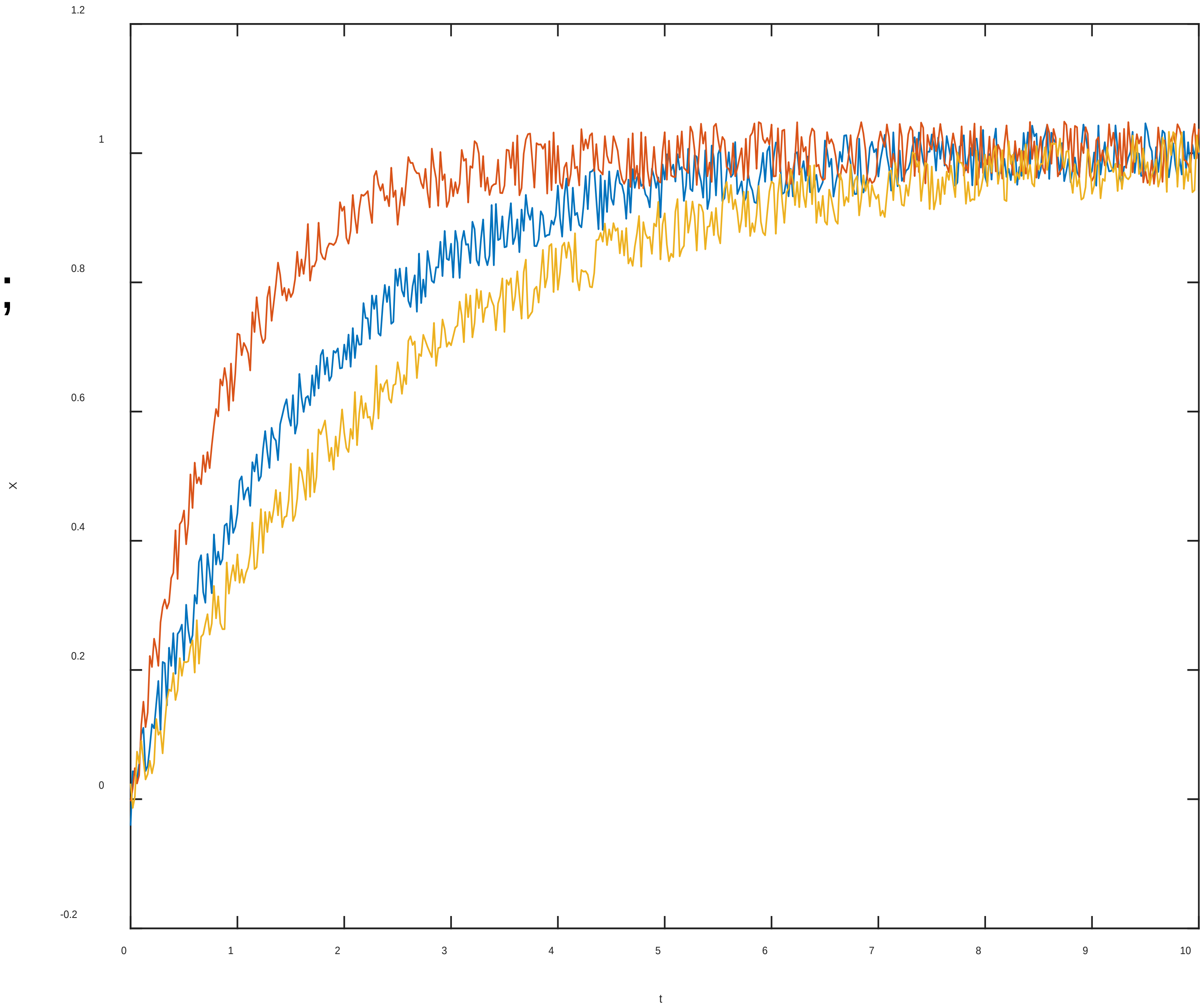
Realizations / Samples - Example

Continous stochastic process:

$$X(t) = A(1 - e^{-k \cdot t}) + w(t);$$

$A = 1;\ k \sim \mathcal{N}(1, 0.4);$
$w(t) \sim \mathcal{U}[-0.1, 0.1]$

3 realizations
$0 \leq t \leq 10$



WWS ÷

Ergodic ÷

Realizations / Samples - Example

Continous stochastic process:

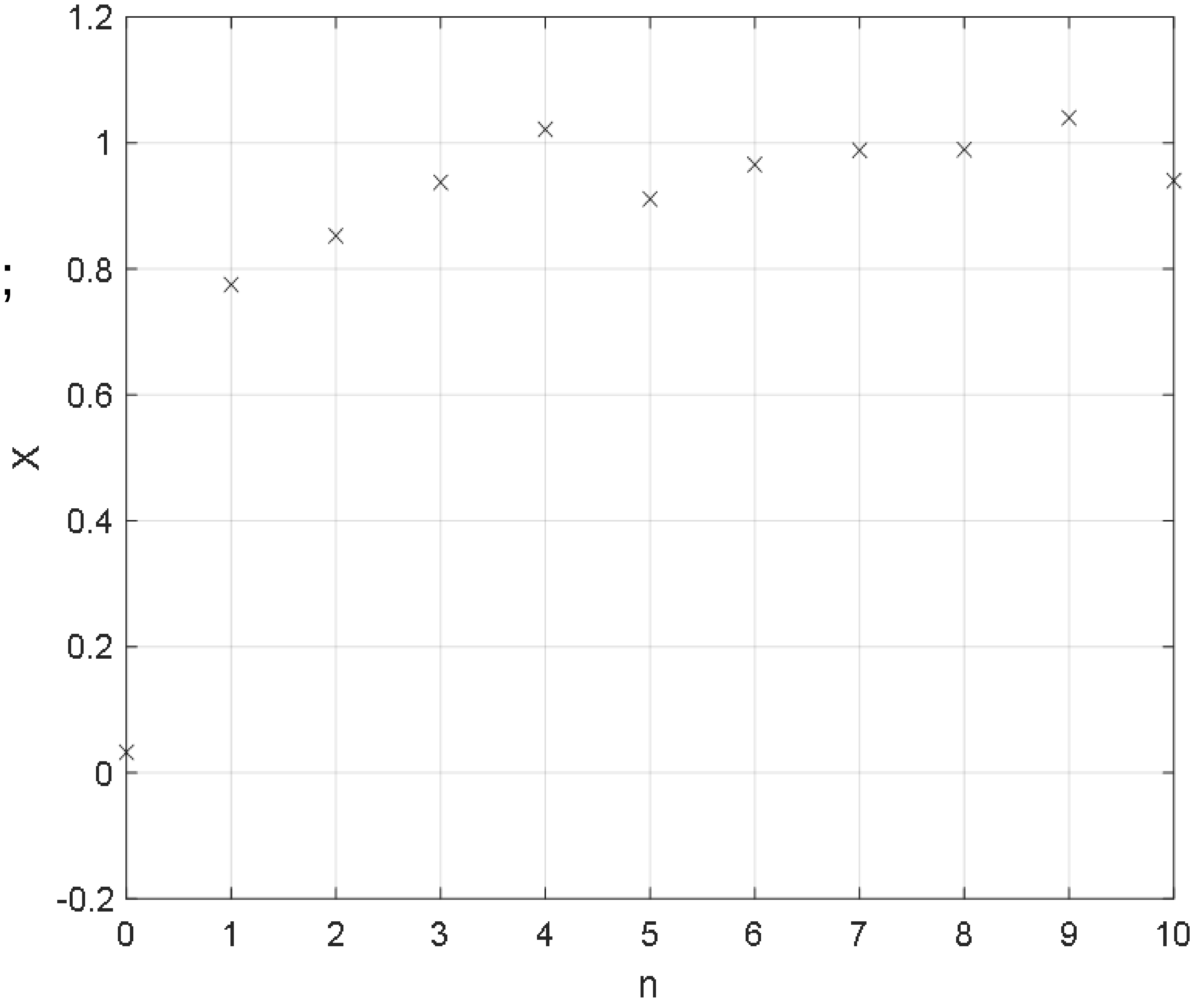
$$X(t) = A(1 - e^{-k \cdot t}) + w(t);$$

$A = 1;\ k \sim \mathcal{N}(1{,}0.4);$
$w(t) \sim \mathcal{U}[-0.1{,}0.1]$

1 realization
11 samples
(n=0,..,10)

WWS ÷

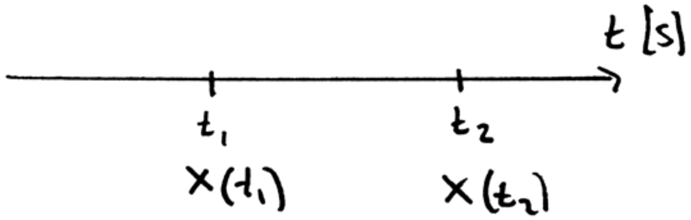
Ergodic ÷



Comparing realizations

Correlations

- We compare the process at two different times.



Correlation of a realization with itself

- <u>Autocorrelation</u>: $R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$
  - Says something about how much the signal $X(t_1)$ resembles itself at time t_2
  - Must depent on how rapidly the signal changes over time
  - Larger if $|t_1 - t_2|$ is small

Correlation of two realizations

- <u>Crosscorrelation</u>: $R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*]$
  - Can be used to look for places where the signal $X(t)$ is similar to the signal $Y(t)$

Autocorrelation

Tells about the connection at two different times

- In general:

Complex conjugated

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$$

$$= \iint_{-\infty}^{\infty} x(t_1)\, x(t_2)^* f_{X(t_1),X(t_2)}\big(x(t_1), x(t_2)\big) dx(t_1) dx(t_2)$$

- For a stationary process (WSS):

$$R_{XX}(t_1, t_2) = R_{XX}(t_1 + T, t_2 + T) = E[X(t_1 + T)X(t_2 + T)^*]$$

Independent of time (t_1)
Depends only on $\tau = t_2 - t_1$

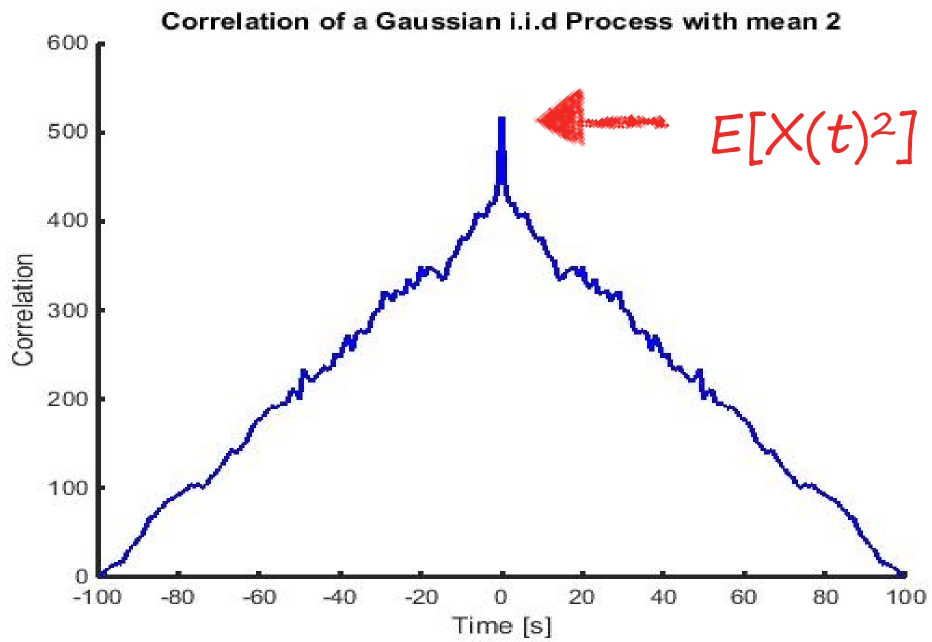
- We rewrite to: $R_{XX}(\tau) = E[X(t)X(t + \tau)^*]$

$\tau = t_2 - t_1$ is the lag!

Autocorrelation

- For Real WSS: $R_{XX}(\tau) = E[X(t)X(t+\tau)]$

- Properties of the autocorrelation function $R_{XX}(\tau)$:
  - An even function of τ ($R_{XX}(\tau) = R_{XX}(-\tau)$)
  - Bounded by: $|R_{XX}(\tau)| \leq R_{XX}(0) = E\left[X^2\right]$ (max. in $\tau = 0$)
  - If X(t) changes fast, then $R_{XX}(\tau)$ decreases fast from $\tau = 0$
  - If X(t) changes slowly, then $R_{XX}(\tau)$ decreases slowly from $\tau = 0$
  - If X(t) is periodic, then $R_{XX}(\tau)$ is also periodic



Temporal Autocorrelation

Temporal only looks at one realization of the stochastic process.

Convolution

- Temporal autocorrelation:

$$\mathcal{R}_{XX}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\,(t) \cdot x\,(t+\tau) dt$$

- If the process is <u>ergodic</u> the temporal autocorrelation is equal to the ensemble autocorrelation:

$$R_{XX}(\tau) = \mathcal{R}_{XX}(\tau)$$

Ensemble

Temporal

Estimate Autocorrelation

We only have measurements of one realization of X(t)

Autocorrelation function:

- In practise, with respect to the lag:

temporal
$$\mathcal{R}_{XX}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x\,(t)\cdot x\,(t+\tau)dt$$

N+1 measurements
$$x(0), x(\Delta t), x(2\Delta t), \dots, x(N\Delta t)$$

- The estimated autocorrelation function:

hat = estimation

t

t+τ

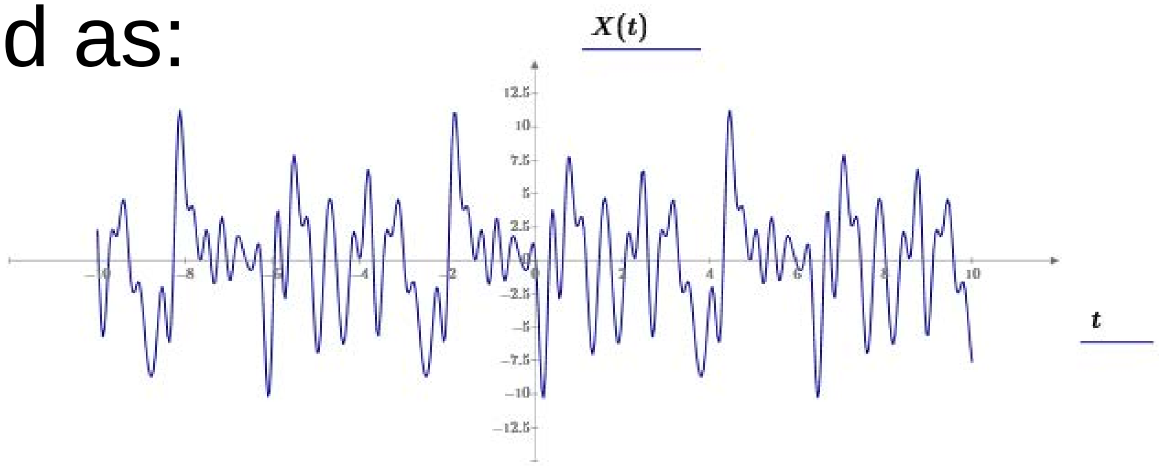
$$\hat{R}_{XX}(n\Delta t) = \frac{1}{N-n+1}\sum_{k=0}^{N-n} x(k\Delta t)\cdot x((k+n)\Delta t)$$

Number of terms (T/ Δt)

Autocorrelation Functions – Example

- Let a stochastic process be defined as:

$$X(t) = \sum_{i=1}^{n} (A_i cos\omega_i t + B_i sin\omega_i t)$$



- where A_i , $B_i \sim \mathcal{N}(0, \sigma^2)$ and i.i.d., and $\omega_i = i \cdot \omega_0$

- Find the autocorrelation:

$$E[X(t)X(t+\tau)] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(A_i cos\omega_i t + B_i sin\omega_i t) \cdot \right.$$
$$\left. \cdot \left(A_j cos\omega_j (t+\tau) + B_j sin\omega_j (t+\tau)\right)\right]$$

Autocorrelation Functions – Example (cont'd)

$$E[X(t)X(t+\tau)] = E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(A_i cos\omega_i t + B_i sin\omega_i t) \cdot \right.$$
$$\left. \cdot (A_j cos\omega_j (t+\tau) + B_j sin\omega_j (t+\tau)) \right]$$

- Since A and B are i.i.d. (and $E[A_i] = E[B_i] = 0$):

$$i \neq j : E[A_i A_j] = 0, E[A_i B_j] = 0, E[B_i A_j] = 0, E[B_i B_j] = 0$$

- We get: $$E[X(t)X(t+\tau)] = \sum_{i=1}^{n}(E\left[A_i^2\right] \cdot cos\,\omega_i t \cdot cos\,\omega_i (t+\tau)$$
$$+ E\left[B_i^2\right] \cdot sin\,\omega_i t \cdot sin\,\omega_i (t+\tau))$$

Autocorrelation Functions – Example (cont'd)

- We can rewrite to:

$$R_{XX}(\tau) = E[X(t)X(t+\tau)]$$

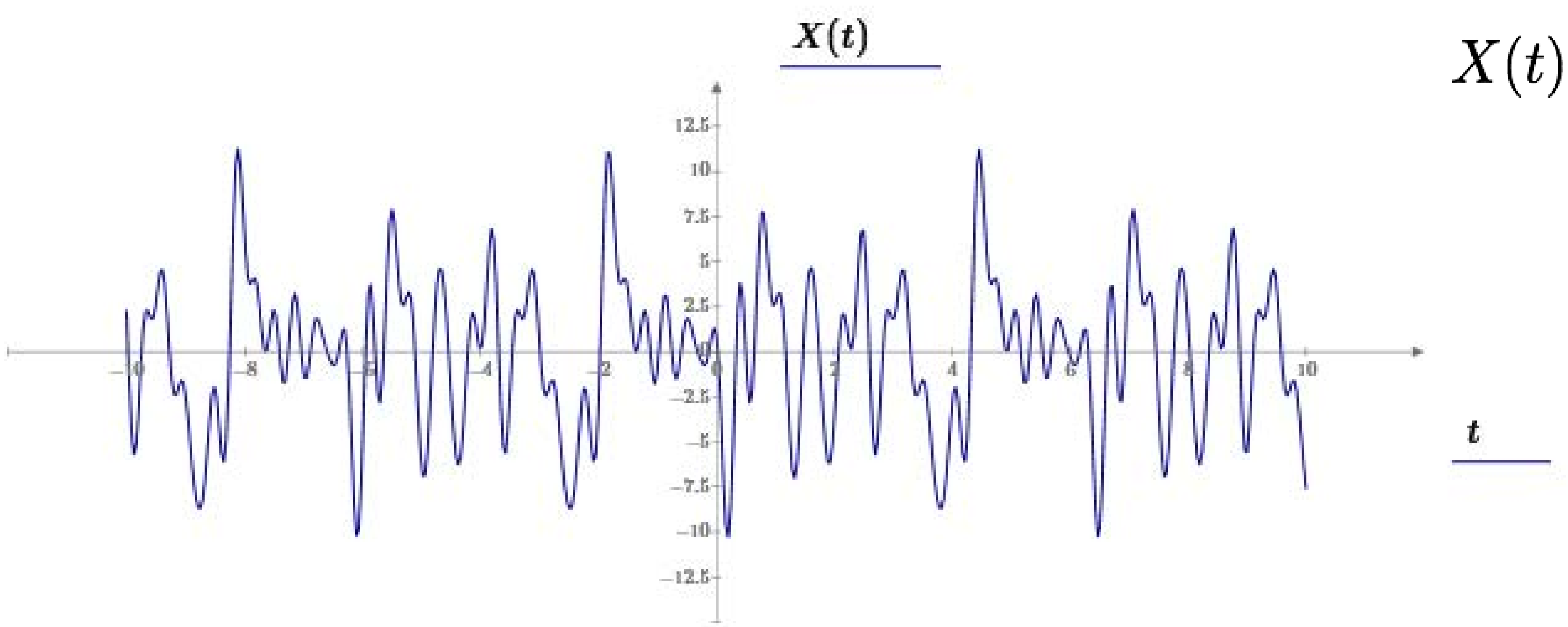
$$= \sum_{i=1}^{n} (E[A_i^2] \cdot cos\ \omega_i t \cdot cos\ \omega_i (t+\tau) + E[B_i^2] \cdot sin\ \omega_i t \cdot sin\ \omega_i (t+\tau))$$

$$= \sigma^2 \sum_{i=1}^{n} cos\ \omega_i \tau$$

(since $E[A_i^2] = E[B_i^2] = \sigma^2$ and $\cos(\theta_1 - \theta_2) = cos\theta_1 \cdot cos\theta_2 + sin\theta_1 \cdot sin\theta_2$)

- We have: $R_{XX}(0) = n\sigma^2$

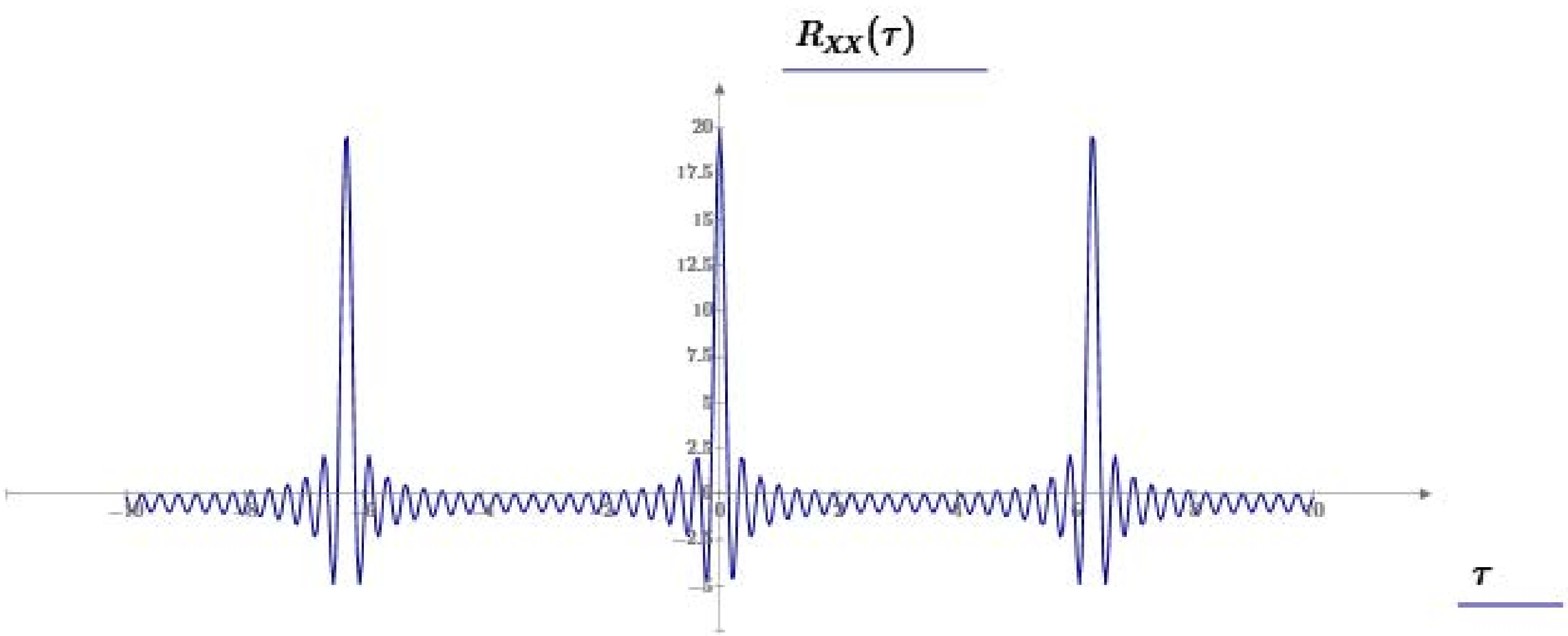
Autocorrelation Functions – Example (cont'd)



$$X(t) = \sum_{i=1}^{n} A_i \cos \omega_i t + B_i \sin \omega_i t$$

$$A_i, B_i \sim \mathcal{N}(0, \sigma^2)$$

$$\omega_i = i \cdot \omega_0$$
$$\omega_0 = 1$$



$$\sigma = 1, n = 20$$

$$R_{XX}(0) = n\sigma^2 = 20$$

Tells about how much we can predict the future

Autocovariances

- Autocovariance function:

$$C_{XX}(t_1, t_2) = E\left[\left(X(t_1) - \mu_X(t_1)\right)(X(t_2) - \mu_X(t_2))^*\right]$$
$$= R_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

Especially: $C_{XX}(t, t\ ) = E\left[(X(t) - \mu_X(t))^2\right] = E\left[X(t)^2\right] - E[X(t)]^2 = \sigma_X^2(t)$

- Autocorrelation coefficient:

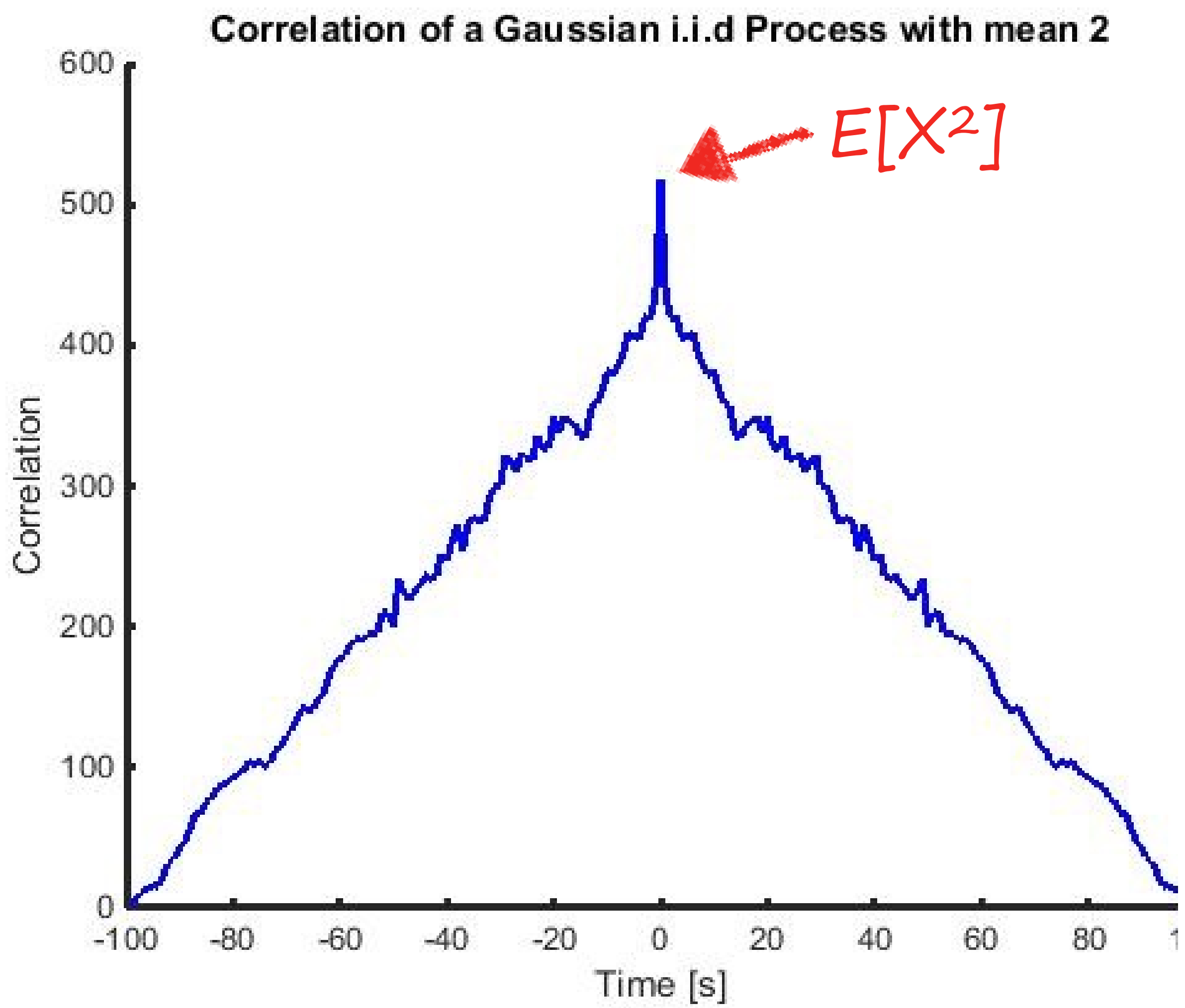
$$r_{XX}(t_1, t_2) = \frac{C_{XX}(t_1,t_2)}{\sqrt{C_{XX}(t_1,t_1)C_{XX}(t_2,t_2)}}; \qquad 0 \le r_{XX}(t_1, t_2) \le 1$$

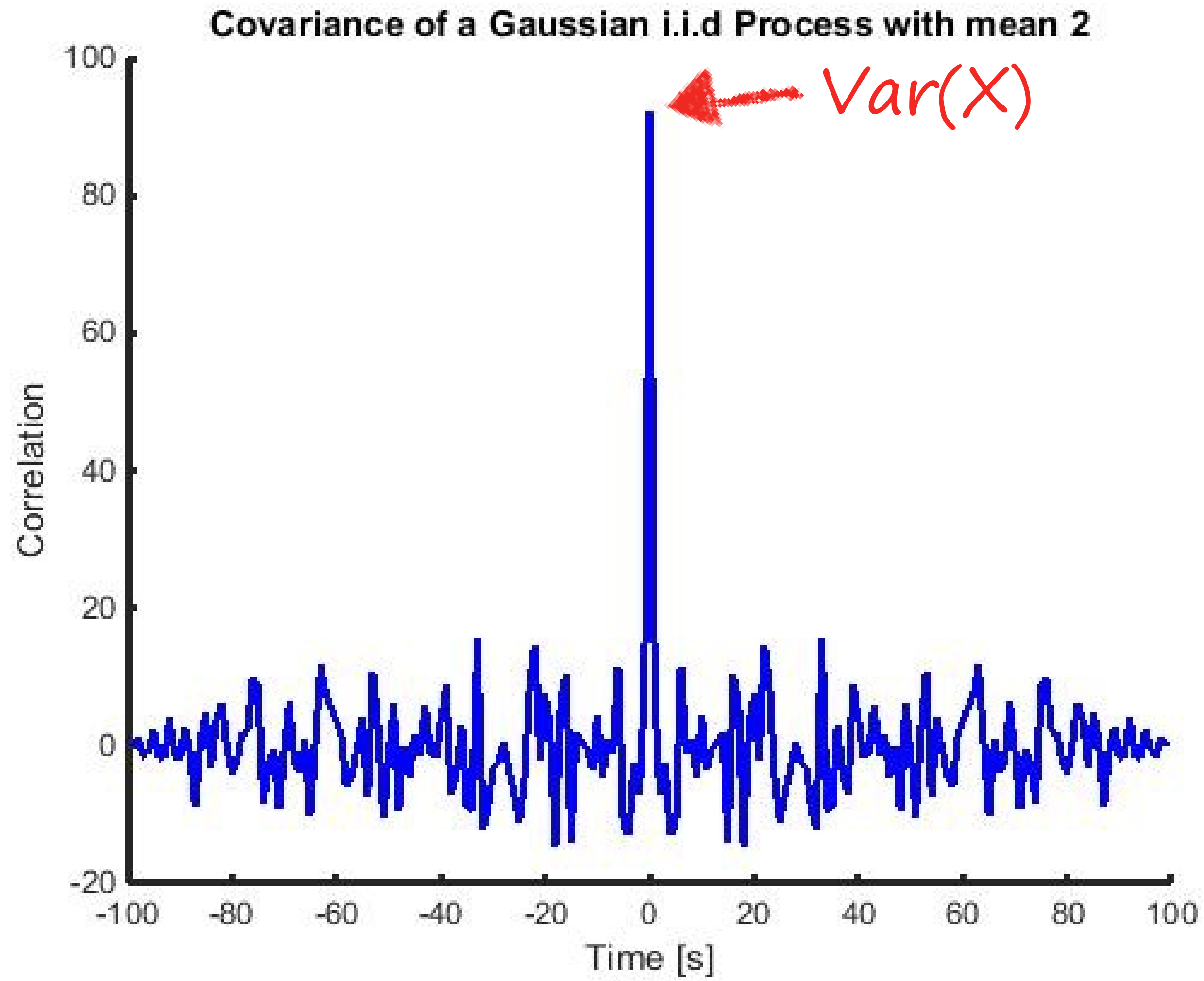
Especially: $r_{XX}(t, t) = 1$ ($X(\mathrm{t})$ is totally correlated to itself!)

Autocovariances

For i.i.d. Gaussian (stationary) noise

- Autocorrelation and autocovariance





Two Stochastic Processes

- If we have two stochastic processes $X(t)$ and $Y(t)$
  - We can compare them by looking at the cross-correlation and cross-covariance:

Cross-correlation $$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*]$$

Cross-covariance $$C_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*] - E[X(t_1)]E[Y(t_2)]$$

Ensemble Cross-correlation

Ensemble means that it applied for the ensemble of the two processes

- In general:

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*]$$
$$= \iint_{-\infty}^{\infty} x(t_1)\, y(t_2)^* f_{X(t_1),Y(t_2)}\big(x(t_1), y(t_2)\big) dx(t_1) dy(t_2)$$

- For two WSS stationary processes:

$$R_{XY}(t_1, t_2) = R_{XY}(t_1 + T, t_2 + T) = E[X(t_1 + T)Y(t_2 + T)^*]$$

- We write: $R_{XY}(\tau) = E[X(t) \cdot Y(t + \tau)^*]$

Cross-Correlation Functions

- For Real WSS processes $X(t)$ and $Y(t)$:

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

- Properties of the cross-correlation function $R_{XY}(\tau)$:
    - $R_{XY}(\tau) = R_{YX}(-\tau))$
    - $|R_{XY}(\tau)| \leq \sqrt{R_{XX}(0)R_{YY}(0)} = \sqrt{E[X^2]E[Y^2]}$
    - $|R_{XY}(\tau)| \leq \frac{1}{2}(R_{XX}(0) + R_{YY}\ (0))$
    - If $X(t)$ and $\mathrm{Y}(t)$ are orthogonal, then $R_{XY}(\tau) = 0$
    - If $X(t)$ and $\mathrm{Y}(t)$ are independant, then $R_{XY}(\tau) = \mu_X \cdot \mu_Y$

Temporal Cross-correlation

Temporal only looks at one realization
of the two stochastic processes.

- The temporal cross-correlation between X and Y:

Convolution

$$\mathcal{R}_{XY}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x\,(t) \cdot y\,(t+\tau) dt$$

- If the two processes are <u>ergodic</u> the temporal cross-correlation is equal to the ensemble cross-correlation:

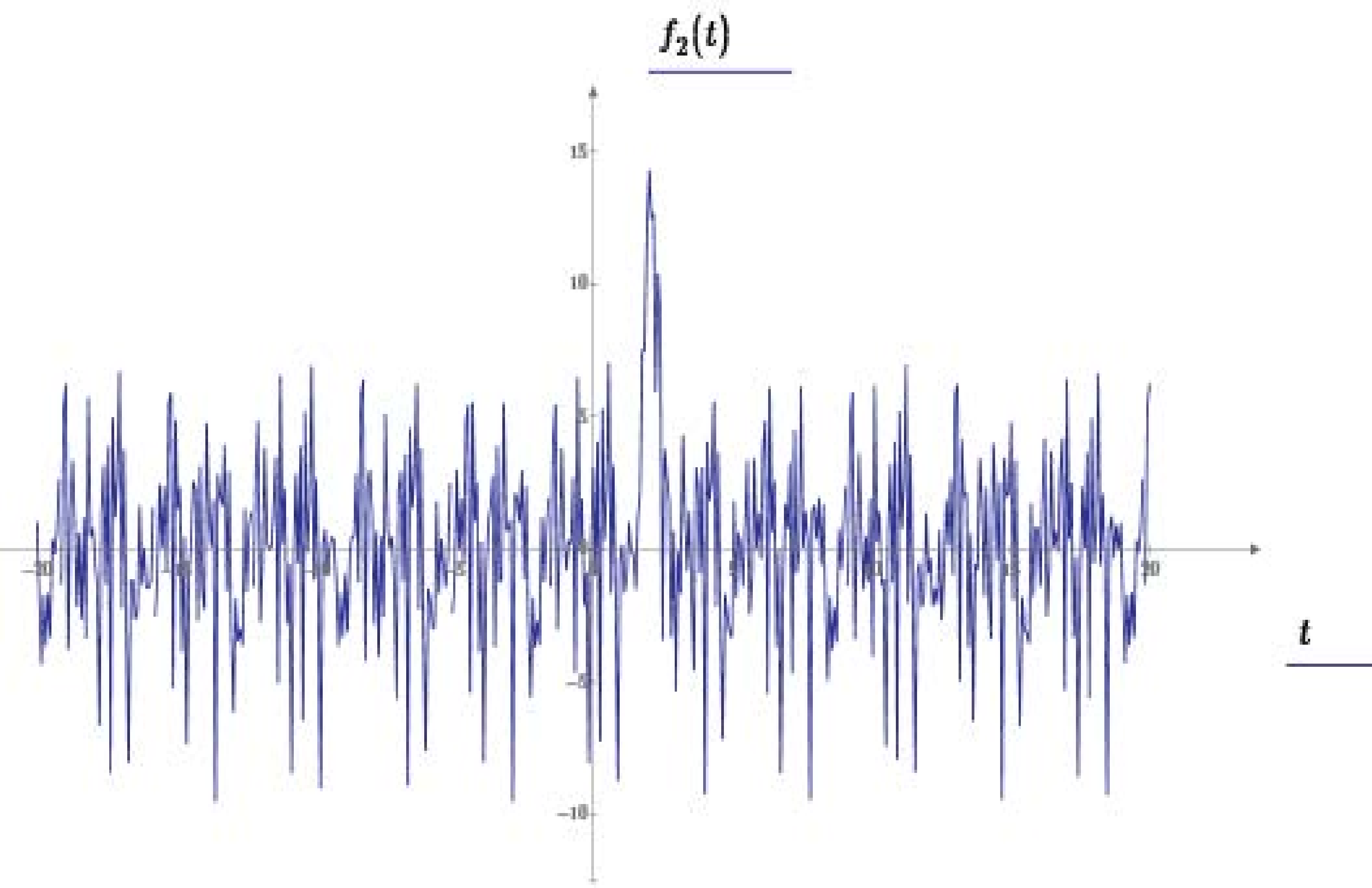
$$R_{XY}(\tau) = \mathcal{R}_{XY}(\tau)$$
$$R_{YX}(\tau) = \mathcal{R}_{YX}(\tau)$$

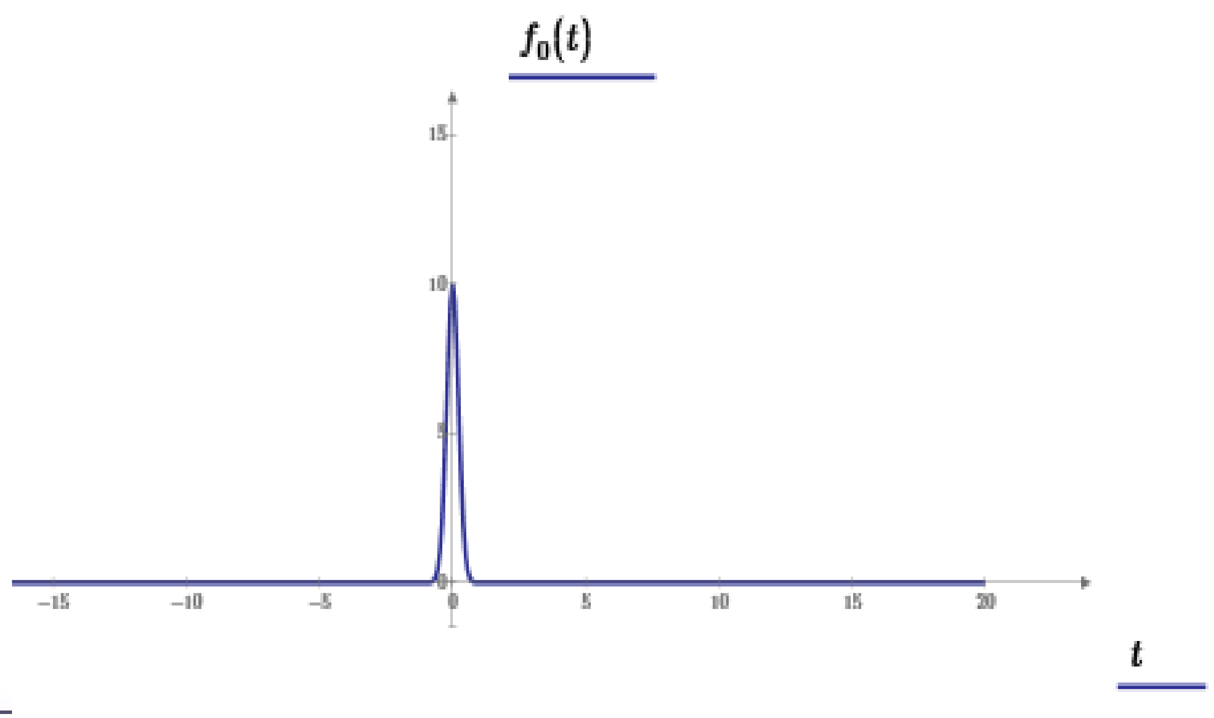
Ensemble

Temporal

Cross-correlation – Uncalibrated noisy signal

- Comparing two signals:
  - An uncalibrated and noisy signal: $f_2(t)$
  - Reference signal: $f_0(t) = 10 \cdot e^{-10t^2}$

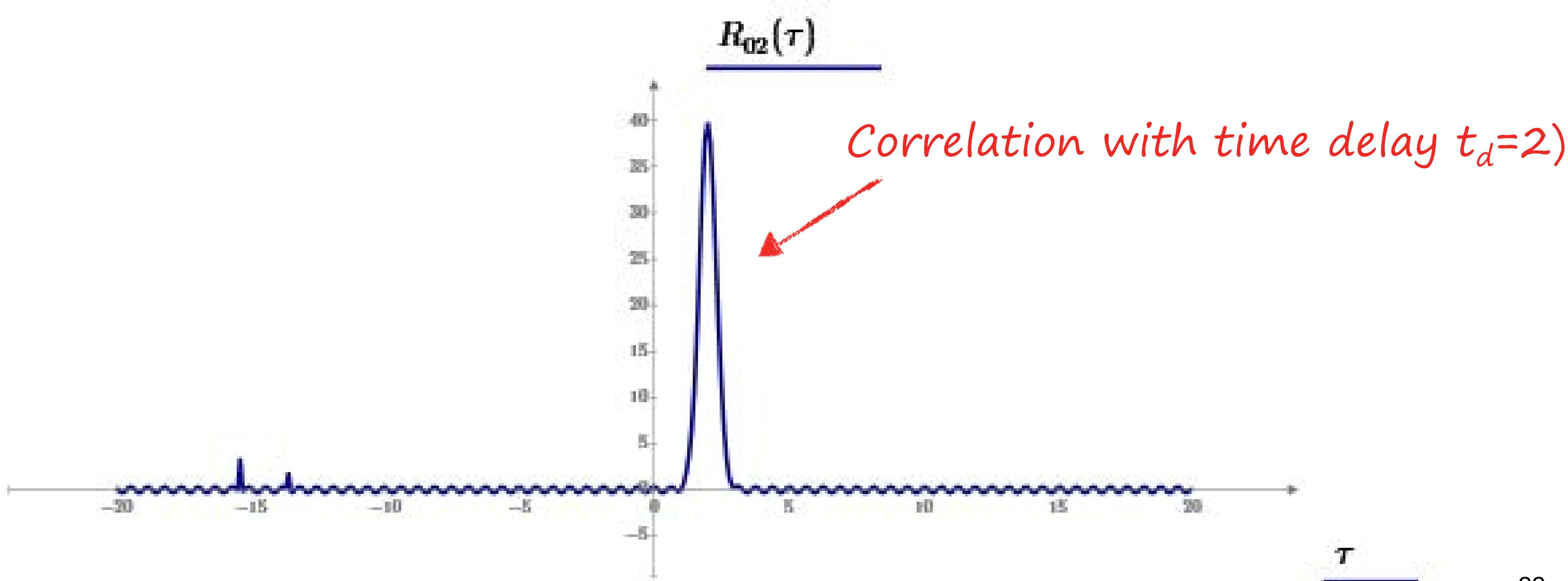




Cross-correlation – Uncalibrated noisy signal

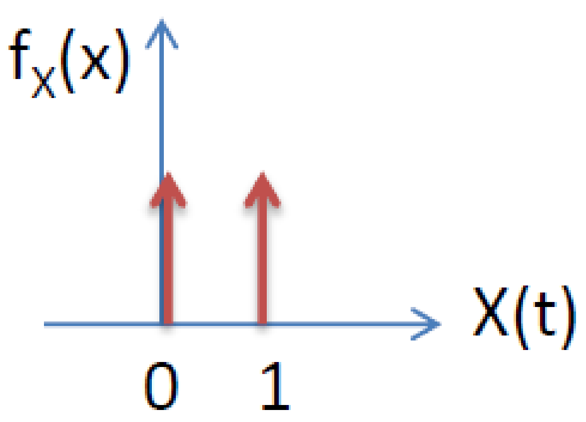
- Comparing two signals:
  - An uncalibrated and noisy signal $f_2(t)$
  - Reference signal $f_0(t) = 10 \cdot e^{-10t^2}$

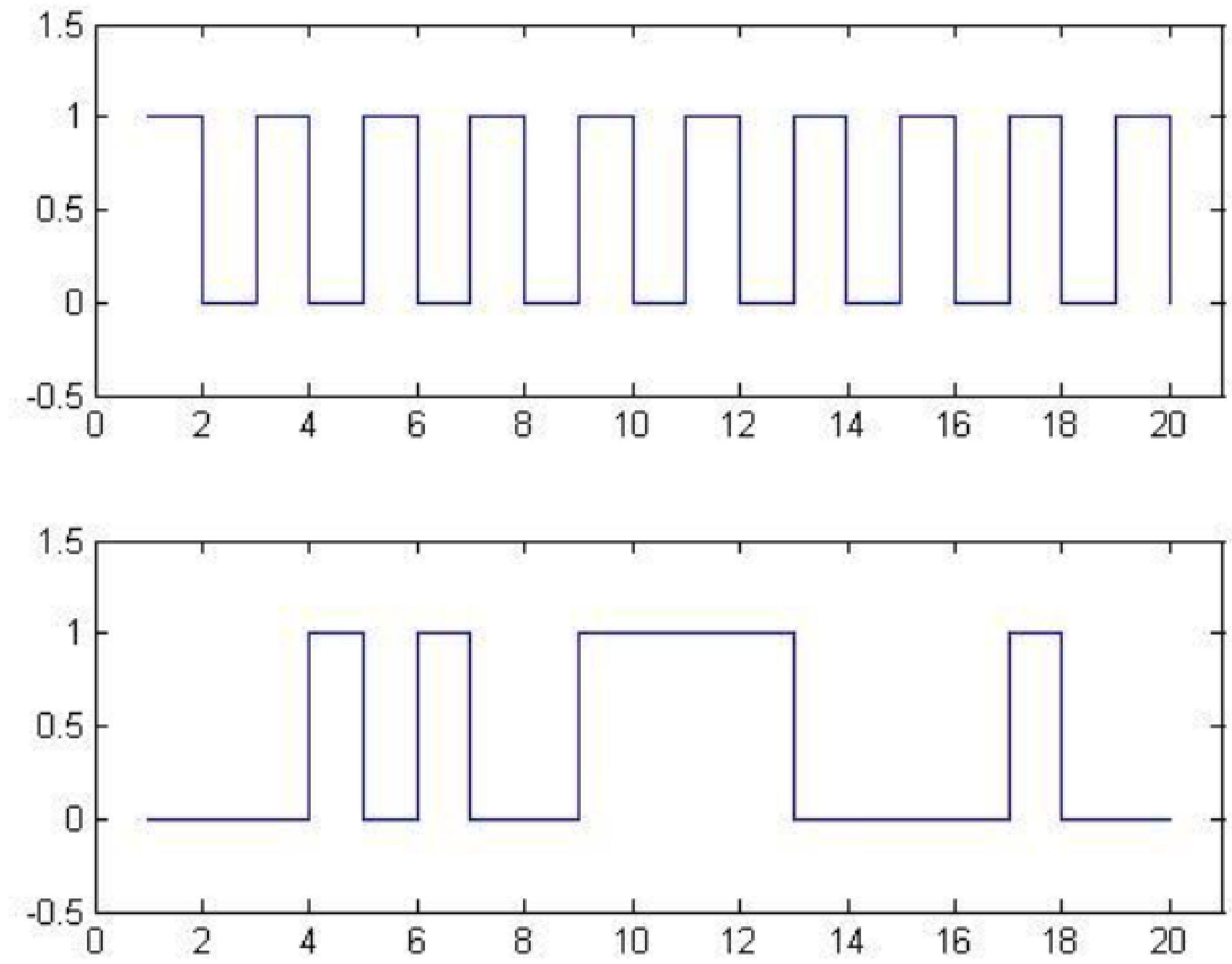
- Cross-correlation: $R_{02}(\tau) = \int_{-\infty}^{\infty} f_0(t) \cdot f_2(t+\tau)dt$



Deterministic vs. Stochastic

The probability mass function:





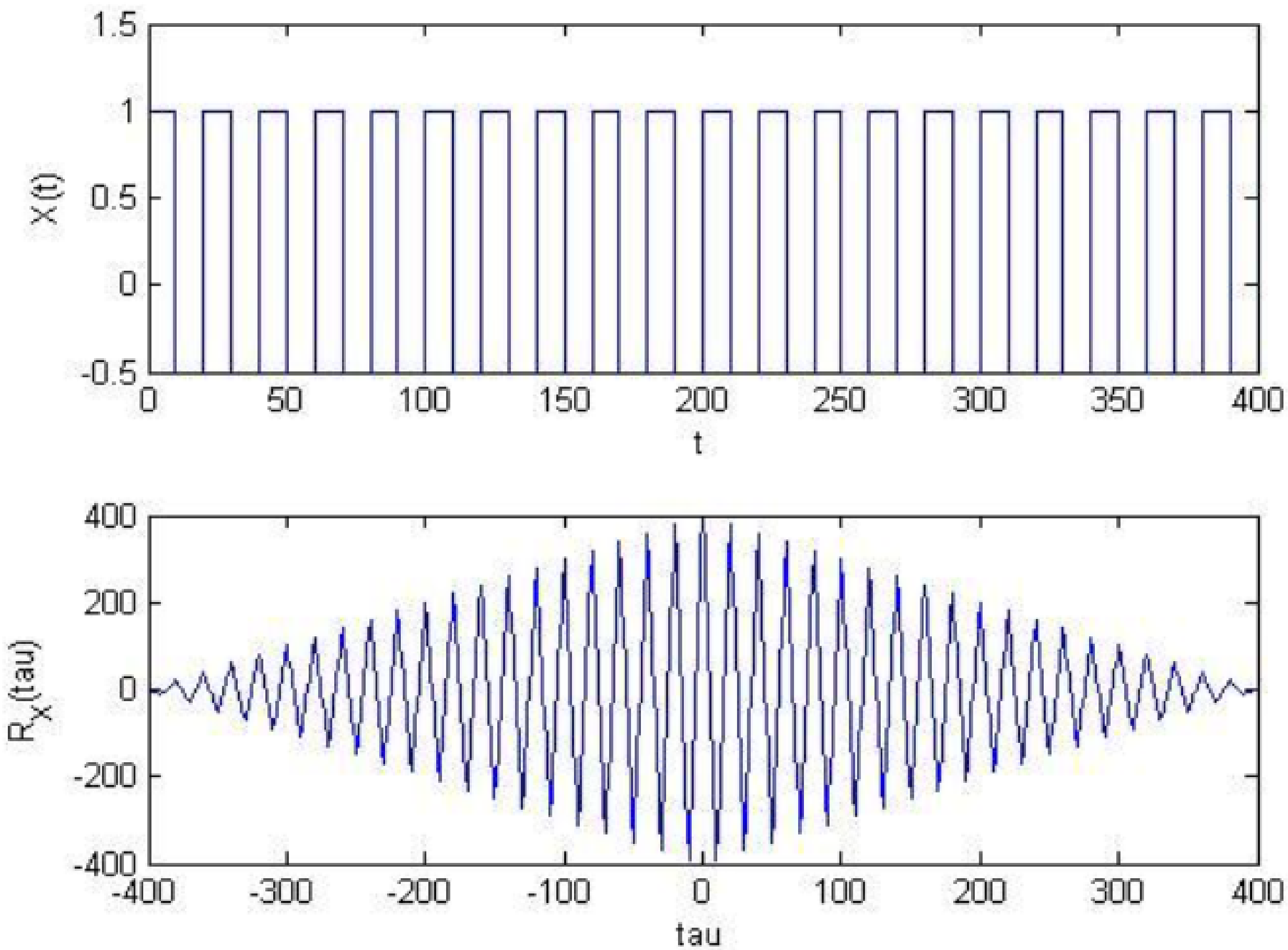
The two random processes have the same pmf.

Deterministic

Periodic signal

↓

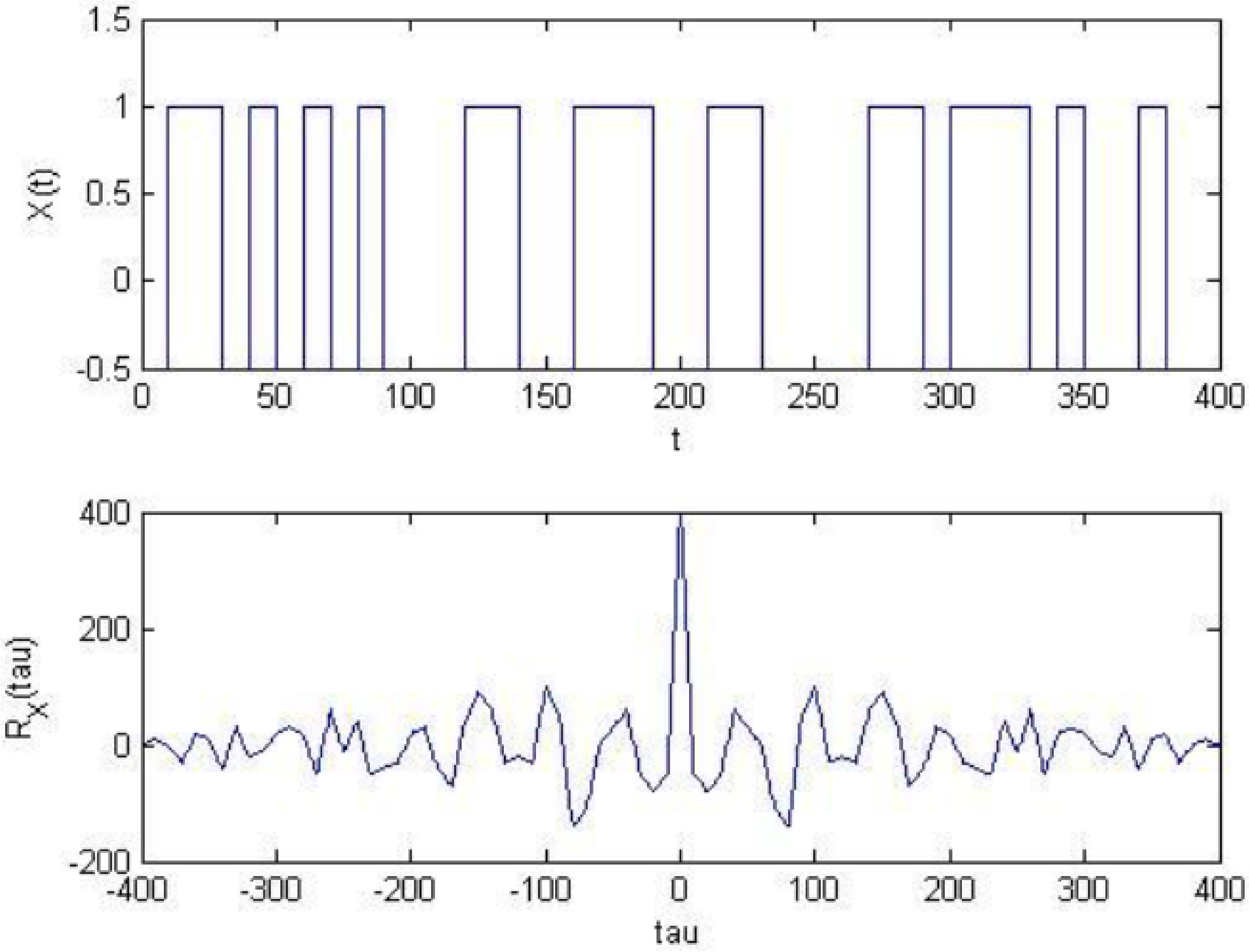
R_{XX} periodic




```
Rx = conv(x,fliplr(x));
```

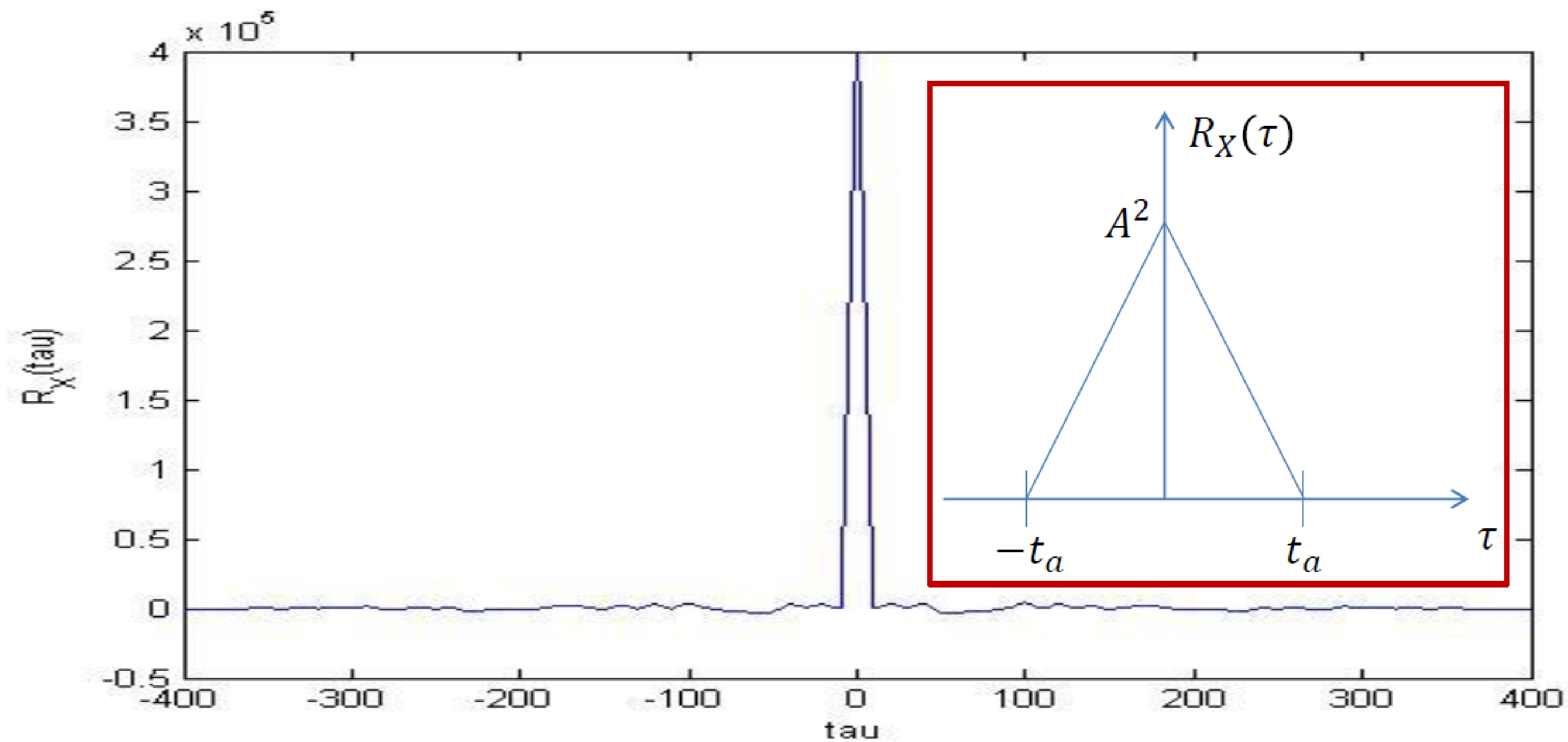
Stochastic

Also called Non-deterministic

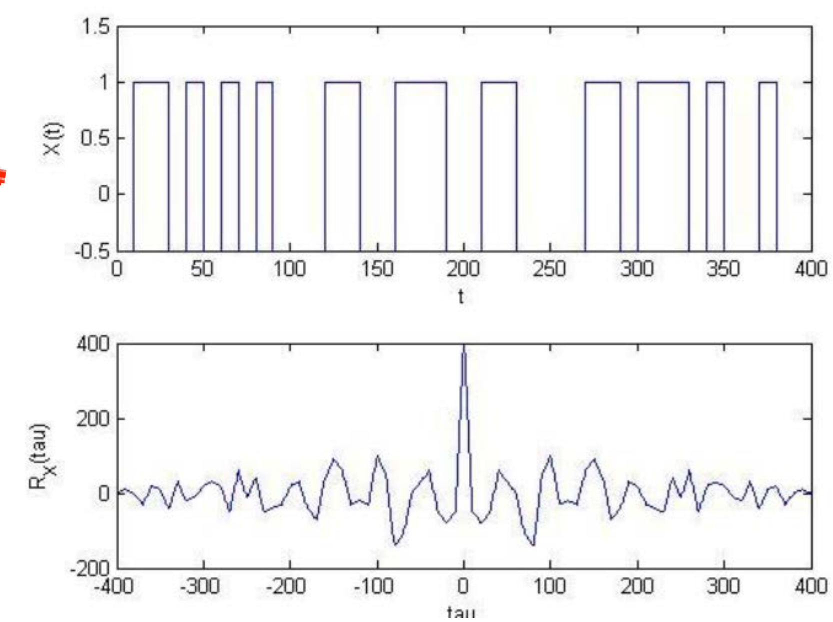



```
Rx = conv(x,fliplr(x));
```

Autocorrelation for Stochastic



Autocorrelation function averaged over 1000 simulations.



Power Spectral Density (psd)

- Frequency domain:
  - Deterministic signals f(t) → Fourier-transformation $\mathcal{F}(\mathrm{f(t)})$
  - Random signals X(t) → ÷Fourier-transformation

- For Real WSS:
- Properties of the autocorrelation function $R_{XX}(\tau)$:
  - If X(t) changes fast, then $R_{XX}(\tau)$ decreases fast from $\tau = 0$
  - If X(t) changes slowly, then $R_{XX}(\tau)$ decreases slowly from $\tau = 0$
  - If X(t) is periodic, then $R_{XX}(\tau)$ is also periodic

→ $R_{XX}(\tau)$ contain information about the frequency content in X(t)

Power Spectral Density (psd)

- Deterministic signals $x(t)$:

Time-average

- Average power: $P_X = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} x(t)^2 dt$
- $x(t)$ periodic T_0: $\langle R_{XX}(\tau)\rangle_{T_0} = \frac{1}{T_0}\int_0^{T_0} x(t)x(t+\tau)dt$
- Power Spectral Density Function (psd):

$$S_{XX}(f) = \mathcal{F}(\langle R_{XX}(\tau)\rangle_{T_0}) \;\Rightarrow\; P_X = \int_{-\infty}^{\infty} S_{XX}(f)df$$

Fourier-transform

Average power in $x(t)$

Power Spectral Density (psd)

- WSS random signals X(t):

- Power Spectral Density Function (psd):

Fourier-transform

$$S_{XX}(f) = \mathcal{F}(R_{XX}(\tau)) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-j\cdot 2\pi f\cdot \tau} d\tau$$

Invers Fourier-transform

$$\Rightarrow \quad R_{XX}(\tau) = \mathcal{F}^{-1}(S_{XX}(f)) = \int_{-\infty}^{\infty} S_{XX}(f) e^{j\cdot 2\pi f\cdot \tau} df$$

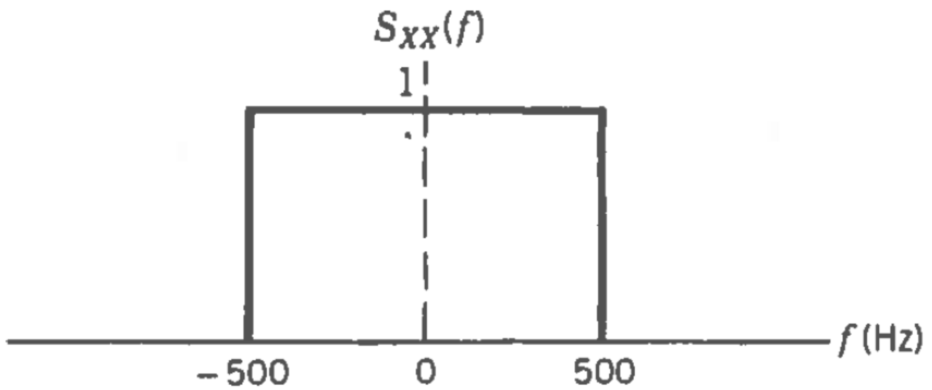


Figure 3.19a Psd of a lowpass random process $X(t)$.

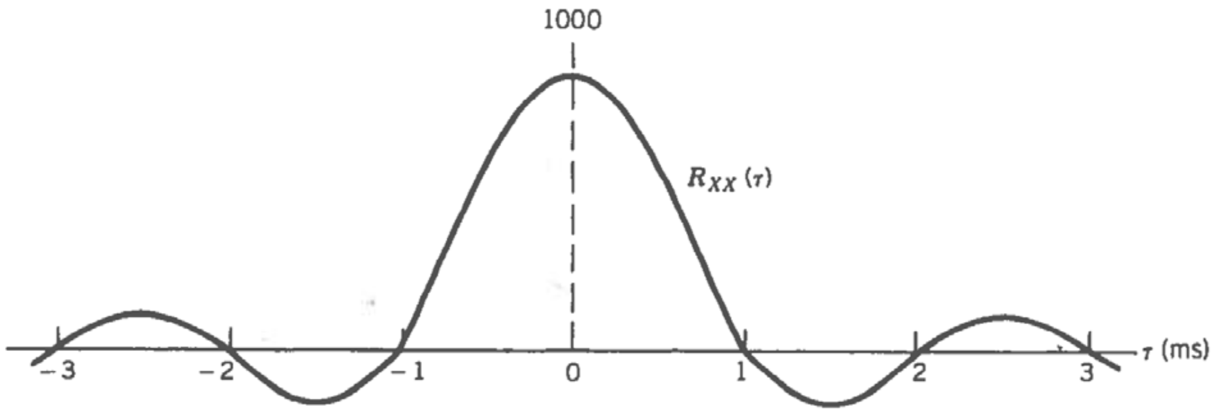


Figure 3.19b Autocorrelation function of $X(t)$.

Figures from ”Random Signals”

Power Spectral Density (psd)

- Properties of psd $S_{XX}(f)$ (spectrum of $X(t)$):
  - $S_{XX}(f) \in \mathbb{R}$
  - $S_{XX}(f) \geq 0$
  - If $X(t) \in \mathbb{R}$: $R_{XX}(-\tau) = R_{XX}(\tau)$ and $S_{XX}(-f) = S_{XX}(f)$ → even functions
  - If $X(t)$ periodic components: $S_{XX}(f)$ will have impulses (δ−functions)
  - $[S_{XX}(f)] = \frac{W}{Hz}$ → Distribution of power with frequency (power spectral density of the stationary random process X(t))
  - $P_X = E\left[X(t)^2\right] = R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(f)df$

    i.e. if $X(t) = V(t)$ (voltage signal)

    → $P_X =$ power in 1Ω-resistor

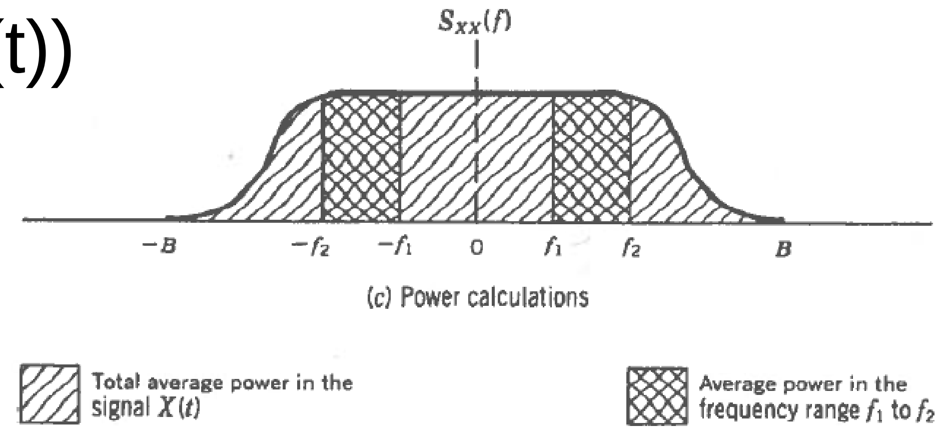


Figure from ”Random Signals”

  - $P_X[f_1, f_2] = 2\int_{f_1}^{f_2} S_{XX}(f)df$ → Power in the frequency-interval $[f_1, f_2]$

Power Spectral Density – Random Binary Signal

Figures from "Random Signals"

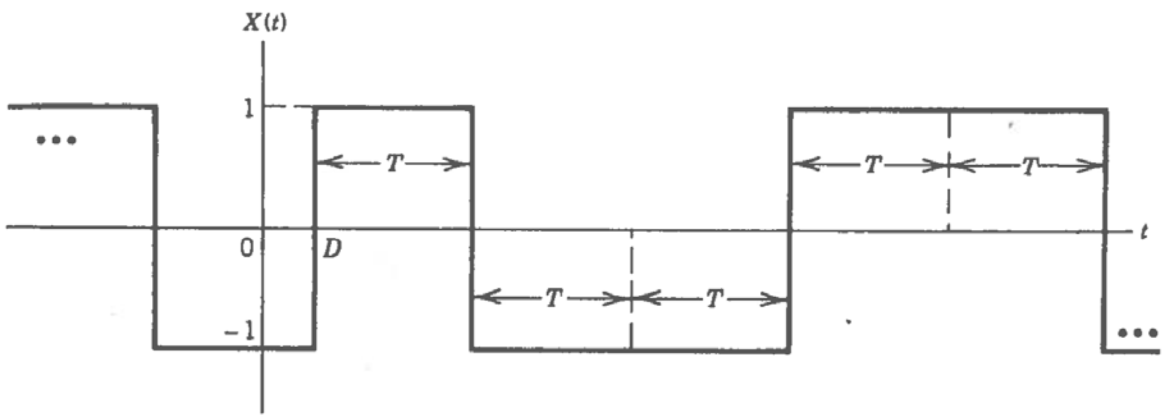


Figure 3.7 Random binary waveform.

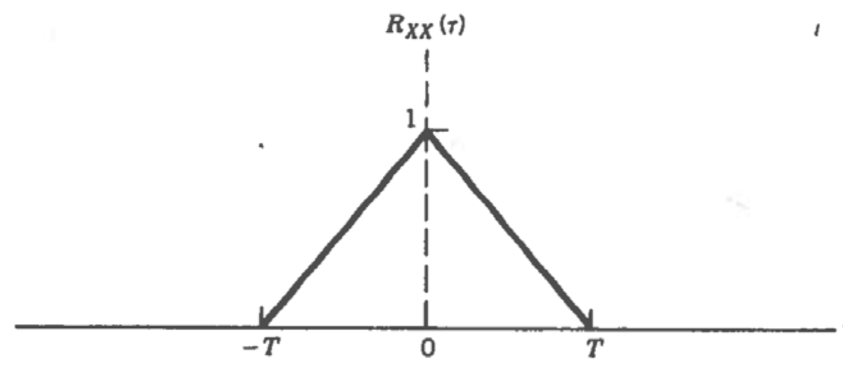


Figure 3.18a Autocorrelation function of the random binary waveform.

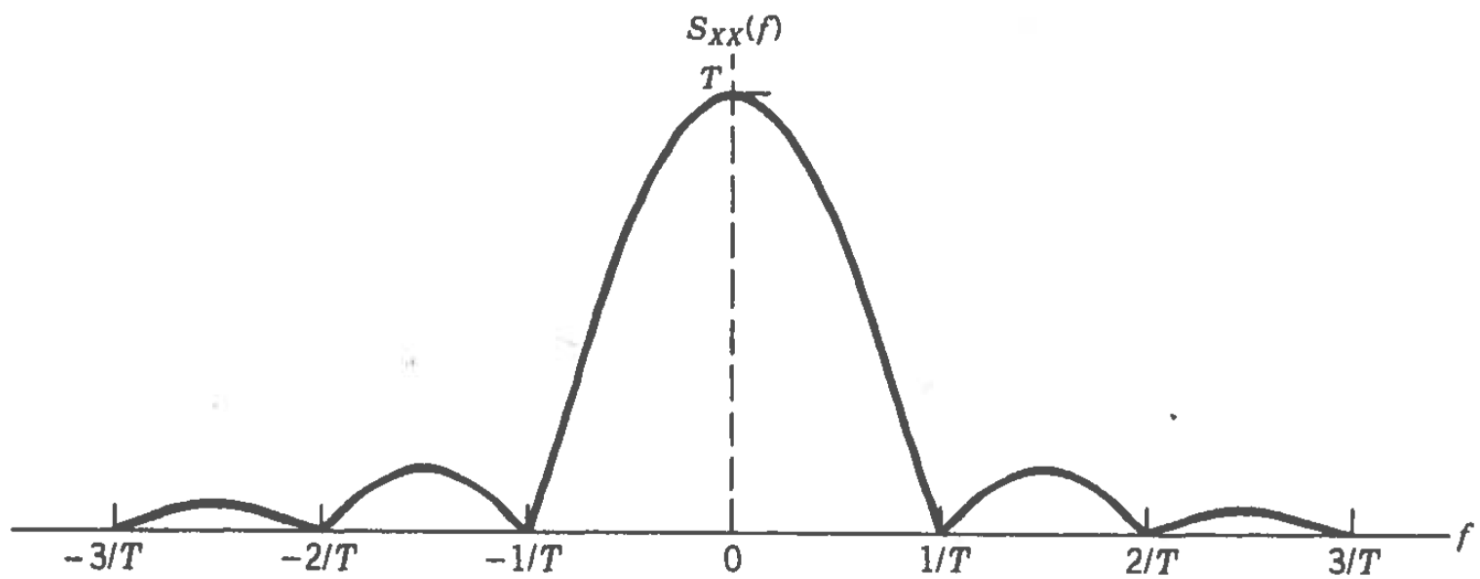


Figure 3.18b Power spectral density function of the random binary waveform.

Words and Concepts to Know

Cross-correlation

Power Spectral Density

Deterministic

Cross-covariance

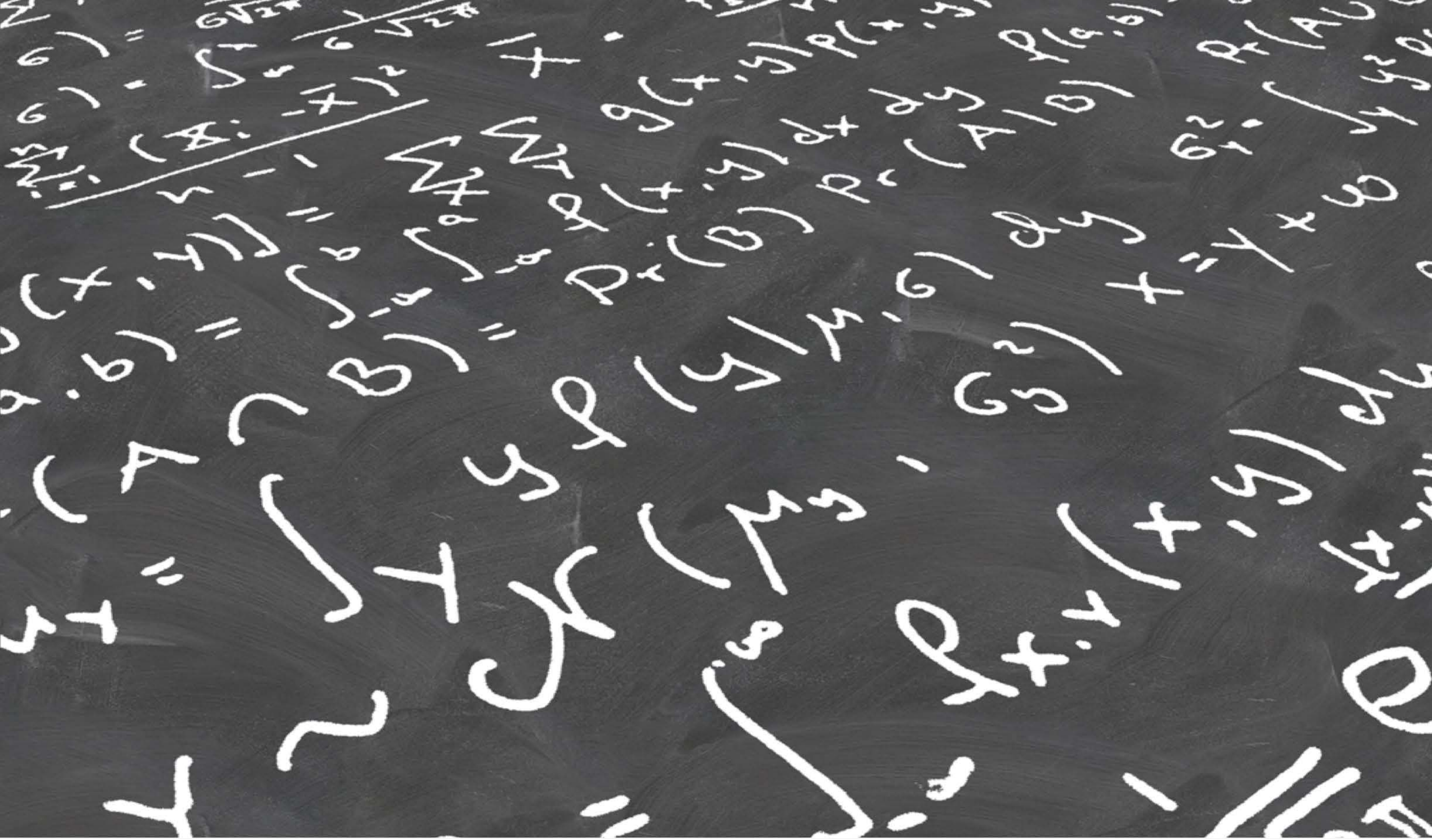
psd

Temporal Autocovariance

Autocorrelation Coefficient

Temporal cross-correlation

Non-deterministic



8. Resume of Probability and Stochastic Processes

Gunvor Elisabeth Kirkelund
Lars Mandrup

Agenda for Today

Resume of stochastic processes:

- Probability
  - Bayes rule
  - Conditional
  - Total
- Stochastic variables
  - pmf/pdf/cdf
  - Joint/marginal/conditional
  - Mean/Variance/Correlation
- Stochastic Processes
  - Ensemble/Sample functions
  - Stationarity and Ergodic Processes
  - Auto- and Cross-correlation functions
  - Power Spectrum Density

Basic Probability

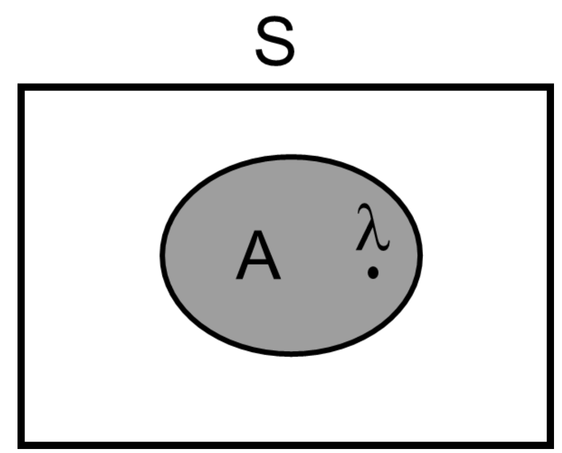
- Probability theory tells us what is in the sample given nature.

- Basic Axions:

Axion 1: $\mathbf{0 \le \boldsymbol{Pr}(\boldsymbol{A}) \le 1}$

Axion 2: $\mathbf{\boldsymbol{Pr}(\boldsymbol{S}) = 1}$

S: Sample space
A: Event
λ: Sample point

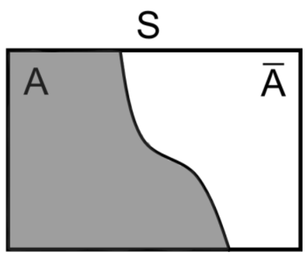


- Often (but not always) we use the relative frequency:

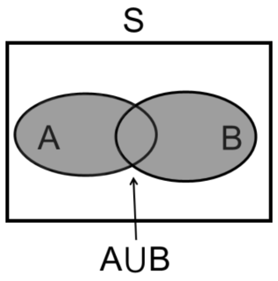
$$\Pr(A) = \frac{N_A}{N}$$

Basic Probability

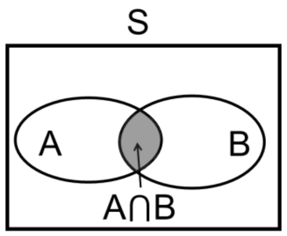
- Complement: $Pr(A) = 1 - Pr(\bar{A})$



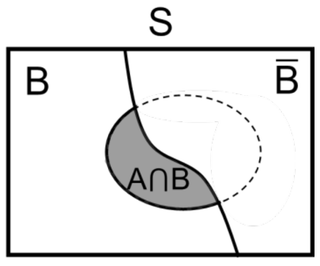
- Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



- Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



- Conditional: $Pr(A|B)$



Bayes Rule and Independence

- Bayes Rule:

$$\boxed{Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B)}}$$

- A and B independent:

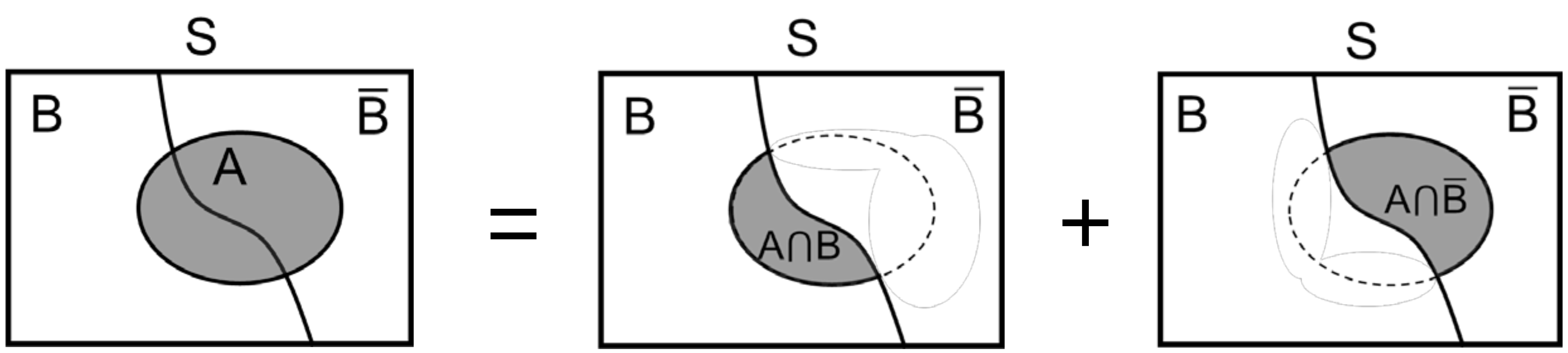
$$\boxed{Pr(A \cap B) = Pr(A) \cdot Pr(B)}$$

$$\boxed{Pr(B|A) = Pr(B) \quad \text{and} \quad Pr(A|B) = Pr(A)}$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.

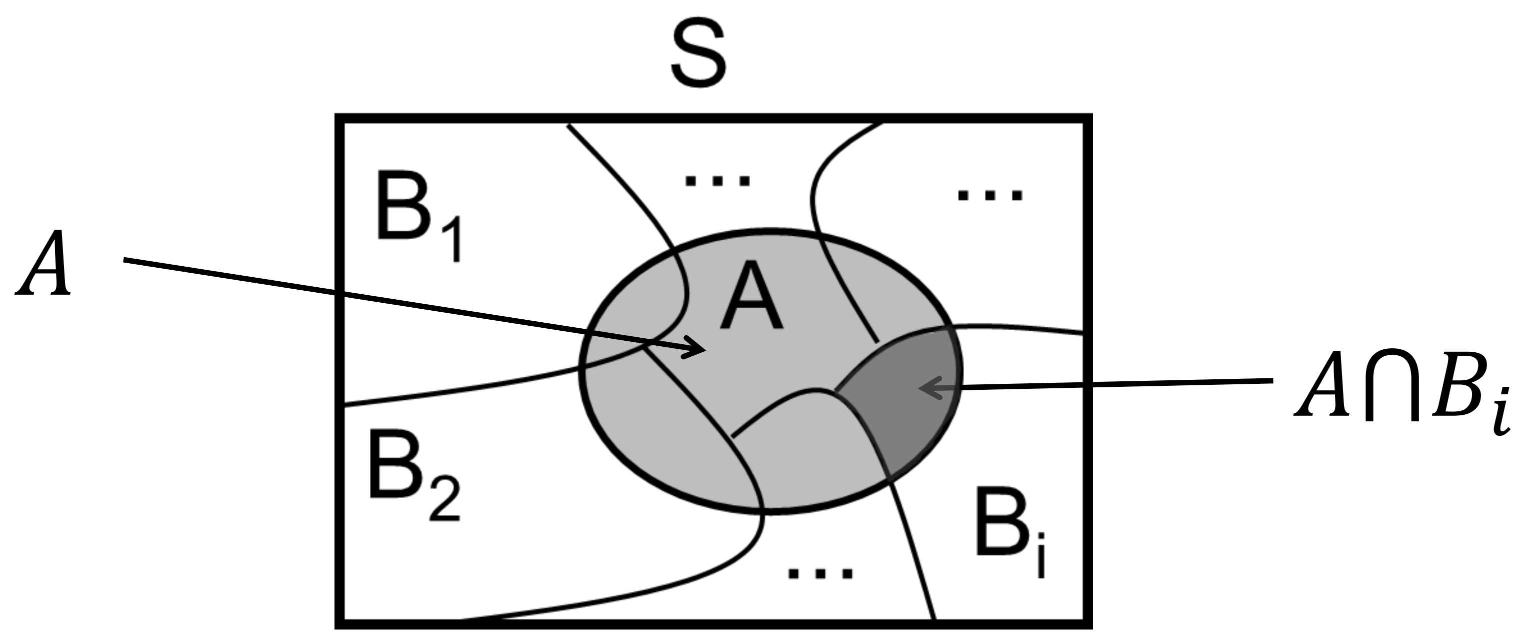


$$\boxed{\begin{aligned} \Pr(A) &= \Pr(A \cap B) + \Pr\left(A \cap \bar{B}\right) \\ &= \Pr(A|B) \cdot \Pr(B) + \Pr\left(A|\bar{B}\right) \cdot \Pr(\bar{B}) \end{aligned}}$$

Total Probability

We sometime call it the marginal

- Pr(A) of an event is the total probability of that event.



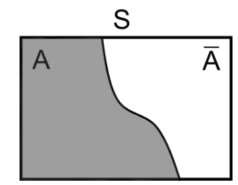
$$Pr(A) = Pr(A \cap B_1) + Pr(A \cap B_2) + \cdots + Pr(A \cap B_i) + \cdots$$
$$= Pr(A|B_1) \cdot Pr(B_1) + Pr(A|B_2) \cdot Pr(B_2) + \cdots$$

where the B_i's are mutually exclusive ($B_i \cap B_j = \varnothing$ for $i \neq j$)
and $S = B_1 \cup B_2 \cup \ldots \cup B_i \cup \ldots$

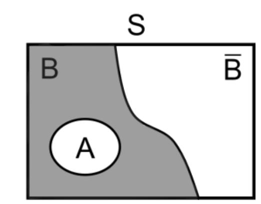
Summary of Probability

Relative frequency: $Pr(A) = \frac{N_A}{N_S}$

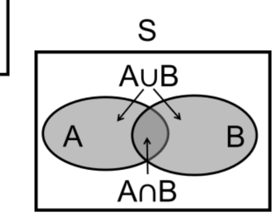
Complement: $Pr(\bar{A}) = 1 - Pr(A)$



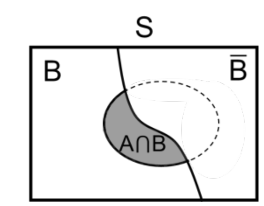
Exclusive: $Pr(\bar{A} \cap B) = Pr(B) - Pr(A)$ if $A \subset B$



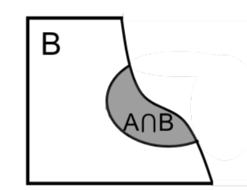
Union: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$



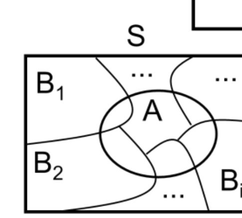
Joint: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$



Conditional: $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$ if $Pr(B) \neq 0$



Total probability: $Pr(A) = \sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)$



Bayes rule: $Pr(B|A) = \frac{Pr(A|B) \cdot Pr(B)}{Pr(A)}$

Bayes formula: $Pr(B_i|A) = \frac{Pr(A|B_i) \cdot Pr(B_i)}{\sum_{i=1}^{n} Pr(A|B_i) \cdot Pr(B_i)}$

Independence: $Pr(A \cap B) = Pr(A) \cdot Pr(B)$

Combinatorics

- The number of possible outcomes of k trials, sampled from a set of n objects.

Types of Experiments:

- With or without replacement
- Ordered or unordered

		Replacement	
		With	Without
Sam-pling	Ordered	n^k	$P_k^n = \dfrac{n!}{(n-k)!}$
	Unordered	$\binom{n+k-1}{k} = \dfrac{(n+k-1)!}{k!\,(n-1)!}$	$\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$

The Binomial Distribution

- We have n repeated trials.
- Each trial has two possible outcomes
  - **Success** — probability p
  - **Failure** — probability $q = 1 - p$

Bernoulli trial

- What is the probability of having k successes out of n trials?
- We write this question as:

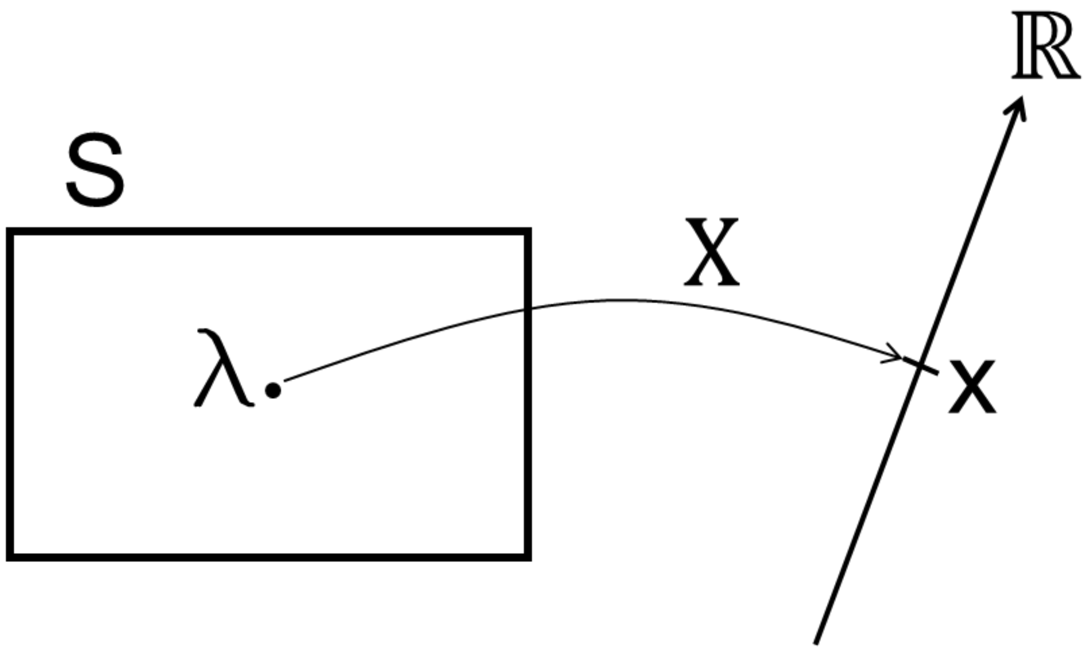
$$\boxed{Pr_n(k) = \frac{n!}{k!\,(n-k)!} p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}}$$

- Faculty: $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$
  $0! = 1$

Also just called a random variables

Stochastic Random Variables

- A random variable tells something important about a stochastic experiment.
- Can be discrete or continous



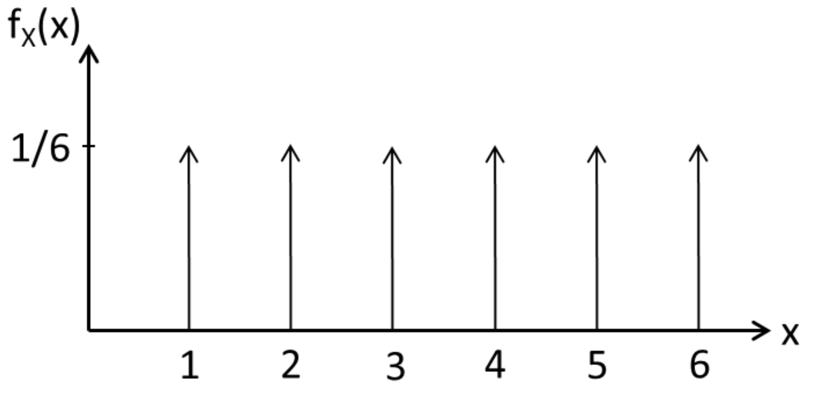
Examples:

- The numbers on a dice (discrete):
  - Sample space for variable X is : $\{1,2,3,4,5,6\}$
  - Sample space for variable Y “Even (1)/Uneven (-1)”: $\{1,-1\}$
- The hight of students at IHA (continous):
  - Sample space for variable H is all real numbers: [100;250] cm.

One Stochastic Variable – Discrete

- Probability mass function (pmf):

$$f_X(x) = \begin{cases} Pr(X = x_i) & for\ X = x_i \\ 0 & otherwise \end{cases}$$

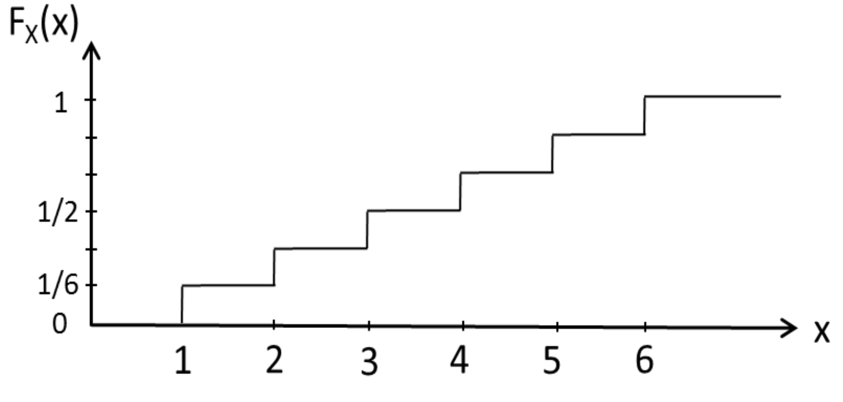


$$0 \le f_X(x) \le 1$$

$$\sum_{i=1}^{n} f_X(x_i) = \sum_{i=1}^{n} Pr(X = x_i) = 1$$

- Cumulative distribution function (cdf):

$$F_X(x) = P\,r(X \le x) = \sum_{i=1}^{n_x} f_X(x_i)$$



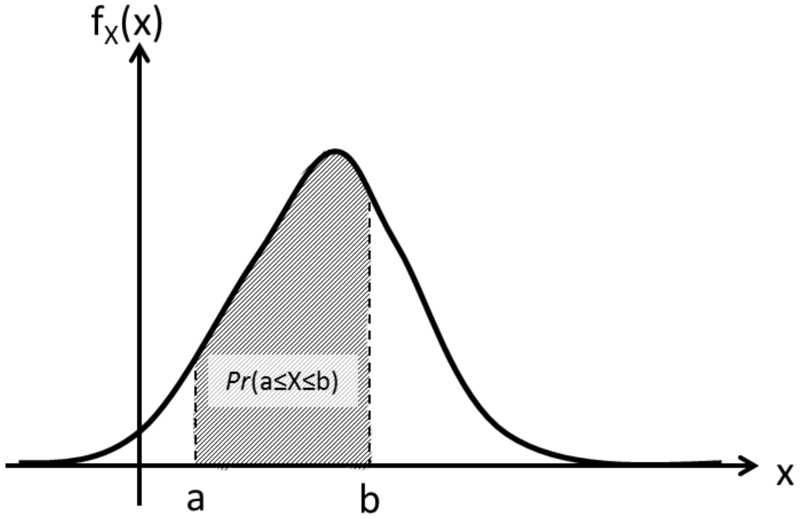
$$0 \le F_X(x) \le 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

One Stochastic Variable – Continuous

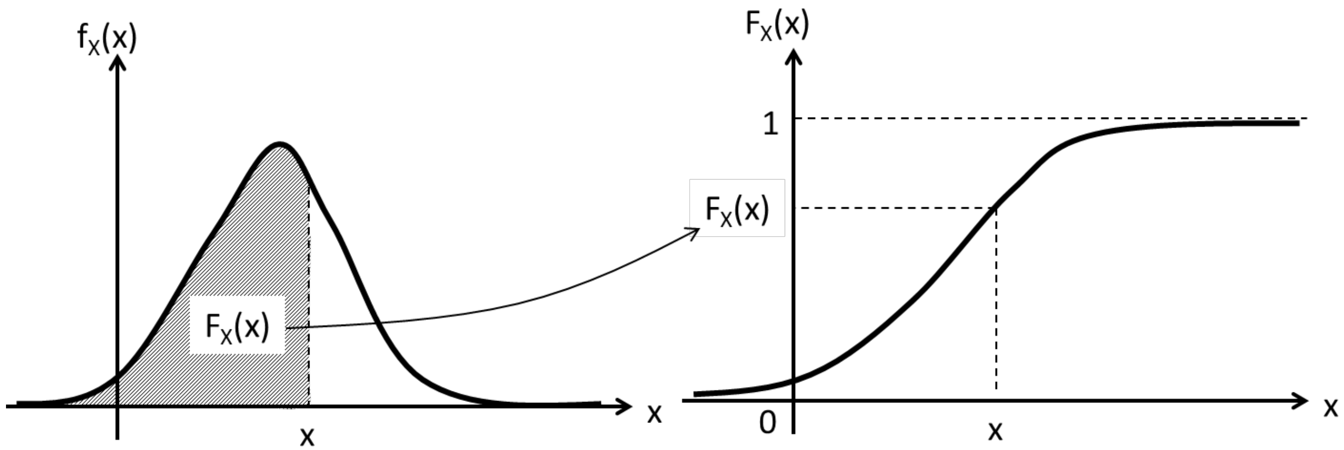
- Probability density function (pdf): $$Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$



$$f_X(x) \geq 0$$

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

- Cumulative distribution function (cdf): $$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = Pr(X \leq x)$$



$$0 \leq F_X(x) \leq 1$$

$$\lim_{x \to -\infty} F_X(x) = 0$$

$$\lim_{x \to \infty} F_X(x) = 1$$

Transformation of Variable X to Y

- Given:
  - Pdf: $f_X(x)$
  - Function/Transformation: $\mathrm{Y} = g(X)$
  - Limits: $\mathrm{a} \leq X \leq b$

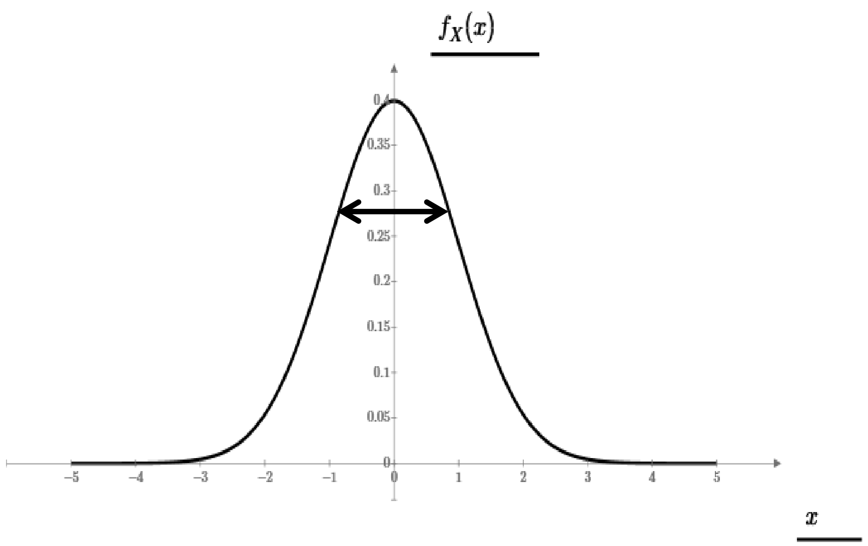
- Find new pdf: $f_Y(y)$:
  1. Inverse: $x = g^{-1}(y)$
  2. Differentiate: $\frac{dg^{-1}(y)}{dy} = \frac{dx(y)}{dy} = \frac{1}{\frac{dg(x)}{dx}}$
  3. Limits: Find $g(a) = a_Y \leq Y \leq b_Y = g(b)$ based on $\mathrm{a} \leq X \leq b$
  4. New pdf: $f_Y(y) = \sum \left|\frac{dx(y)}{dy}\right| f_X\left(g^{-1}(y)\ \right) = \sum \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}$

Expectations

- Mean value: $E[X] = \bar{X} = \mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)dx \quad \left(\sum_{i=1}^{n} x_i\, f_X(x_i)\right)$

- Variance: $Var(\mathrm{X}) = {\sigma_X}^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 \cdot f_X(x)dx = E\left[X^2\right] - E[X]^2$

- Standard deviation: $\sigma_X = \sqrt{Var(\mathrm{X})}$



- Linear function: $E[aX + b] = a \cdot E[X] + b$

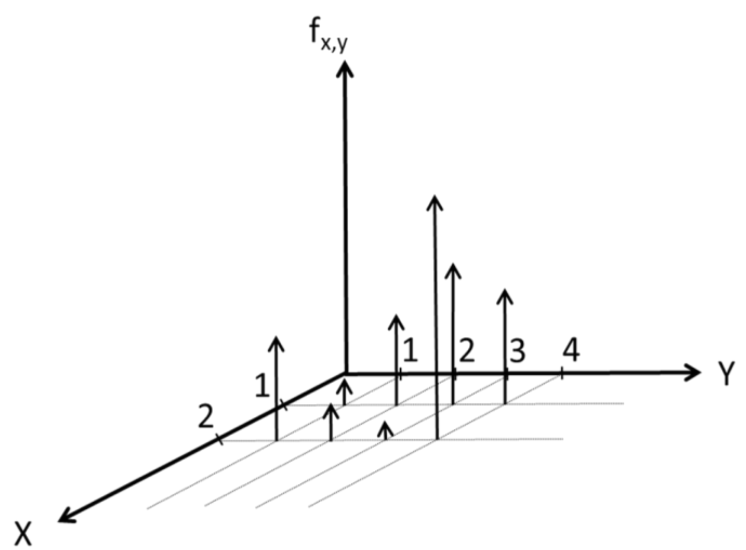
$$Var[aX + b] = a^2\left(E[X^2] - E[X]^2\right) = a^2 \cdot Var(X)$$

Two Stochastic Variables X, Y – Discrete

Joint (Simultaneous) pmf:

$$f_{X,Y}(x,y) = \begin{cases} Pr\left((X = x_i)\cap(Y = y_j)\right) & for\ X = x_i \wedge Y = y_j \\ 0 & otherwise \end{cases}$$

$$0 \le f_{X,Y}(x,y) \le 1 \qquad \sum_{i=1}^{m}\sum_{j=1}^{n} f_{X,Y}(x_i, x_j) = 1$$



Marginal pmfs:

$$f_X(x) = \sum_{y} f_{X,Y}(x,y) \qquad f_Y(y) = \sum_{x} f_{X,Y}(x,y)$$

Cumulative Distribution Function cdf:

$$F_X(x_i, y_j) = Pr\left((X \le x_i)\cap(Y \le y_j)\right) = \sum_{m=1}^{i}\sum_{n=1}^{j} f_{X,Y}(x_m, y_n)$$

Two Stochastic Variables X, Y – Continuous

Joint (Simultaneous) pdf:

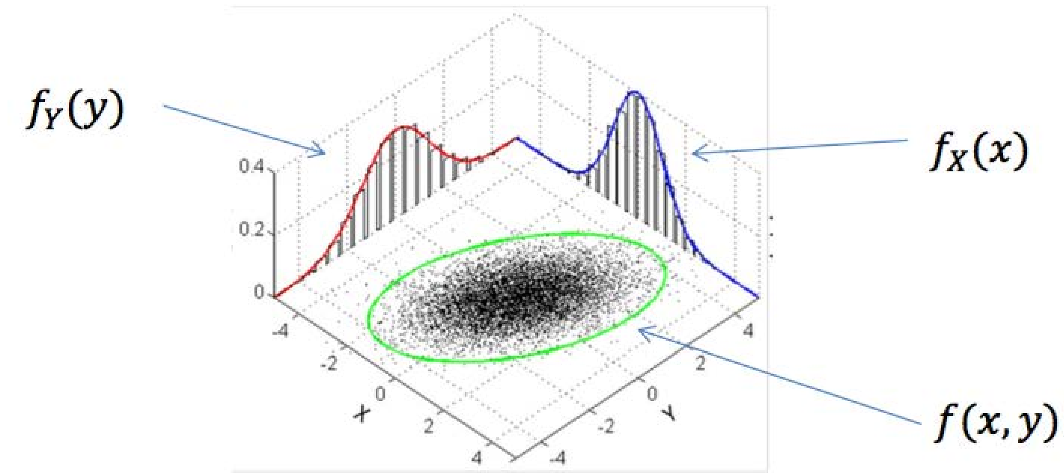
$$f_{X,Y}(x,y) \geq 0$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

Marginals:

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

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



Cumulative Distribution Function cdf:

cdf $$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)dxdy = Pr(X \leq x \wedge Y \leq y)$$

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

Bayes Rule, Conditional PDF and Independence

Bayes rule:

- The joint/simultaneous pmf/pdf for two stochastic variables:

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

Conditional pdf:

- For a two dimensional pmf/pdf $f_{X,Y}(x,y)$, we can find the conditional pdf with Bayes rule:

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

Independence:

- X and Y are independent if and only if:

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ or $f_{X|Y}(x|y) = f_X(x)$ for all x and y

Correlation and Covariance

Correlation tells of the (biased) coupling between variables

- Correlation: $corr(X,Y) = E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x,y)dxdy$

Covariance is without bias from the mean

- Covariance: $cov(X,Y) = E\big[(X-\bar{X})(Y-\bar{Y})\big] = E[XY] - E[X] \cdot E[Y]$

Correlation Coefficient is the normalized Covariance

- Correlation coefficient: $\rho = E\left[\frac{X-\bar{X}}{\sigma_X} \cdot \frac{Y-\bar{Y}}{\sigma_Y}\right] = \frac{E[XY]-E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$

$$-1 \leq \rho \leq 1$$

- If X and Y are independent:

$E[XY] = E[X] \cdot E[Y]$ and $cov(X,Y) = \rho = 0$

Important Rules

- $E[aX + b\,] = a \cdot E[X] + b$
- $Var[aX + b\,] = a^2 \cdot Var(X)$
- $E[aX + bY] = a \cdot E[X] + b \cdot E[Y]$ $\rightarrow$ Linearity of the mean
- $Var[aX + bY] = a^2 \cdot Var[X] + b^2 \cdot Var[Y] + 2ab \cdot Cov(X, Y)$

Correlation

- $Corr(X, Y) = E[XY]$ $(= E[X] \cdot \mathrm{E}[\mathrm{Y}]$ if X and Y are independent)
- $Cov(X, Y) = E\left[(X - \bar{X})(Y - \bar{Y})\right] = E[XY] - E[X] \cdot E[Y]$
- $\rho = E\left[\frac{X - \bar{X}}{\sigma_X} \cdot \frac{Y - \bar{Y}}{\sigma_Y}\right] = \frac{E[XY] - E[X]E[Y]}{\sigma_X \cdot \sigma_Y}$

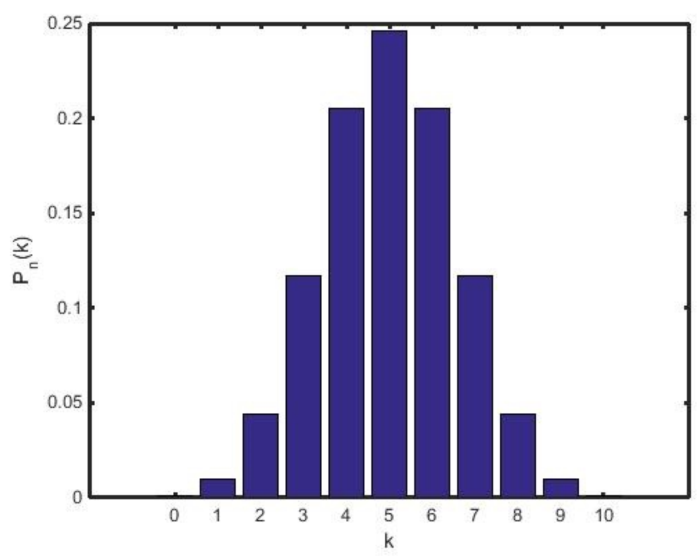
Correlation coefficient

Notice that correlation and correlation coefficient are different, but can have same name and same notation!!

The Binomial Distribution

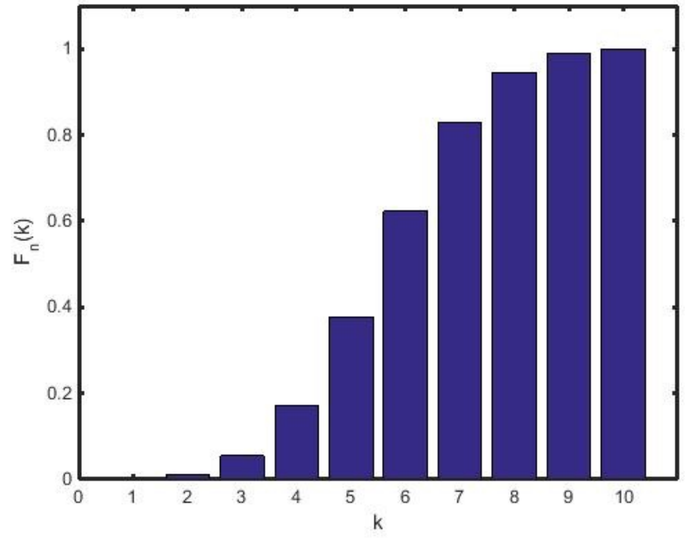
- n repeated trials – each with two possible outcomes
  - **Success** — probability p
  - **Failure** — probability $q = 1 - p$

Also called a Bernoulli trial



- Probability mass function (pmf):

$$f(k|n,p) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$



- Cumulative distribution function (cdf):

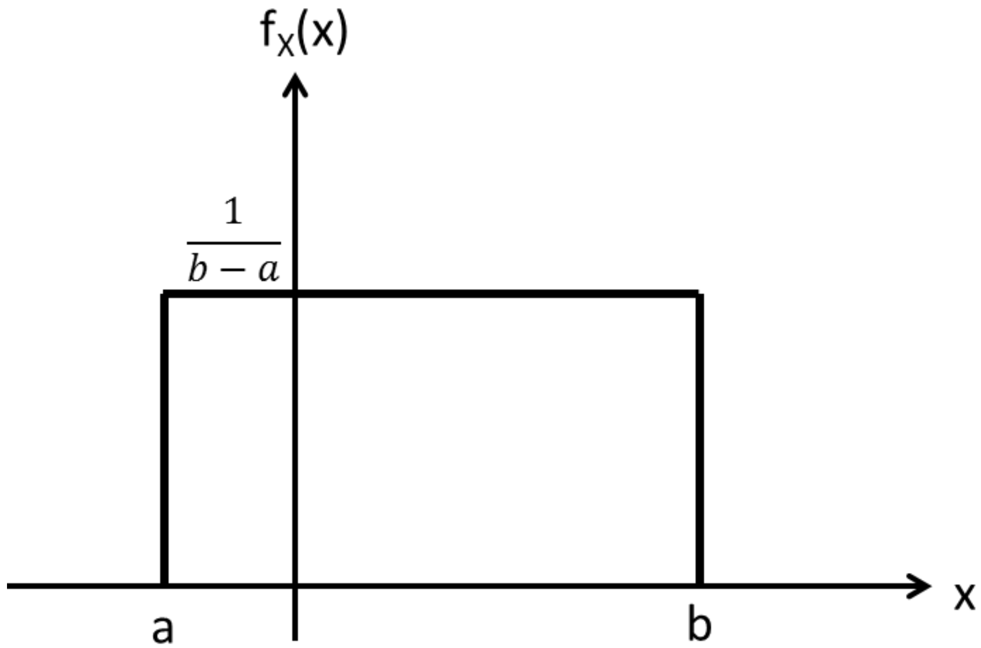
$$F(k|n,p) = \sum_{i=0}^{k} f(i|n,p)$$

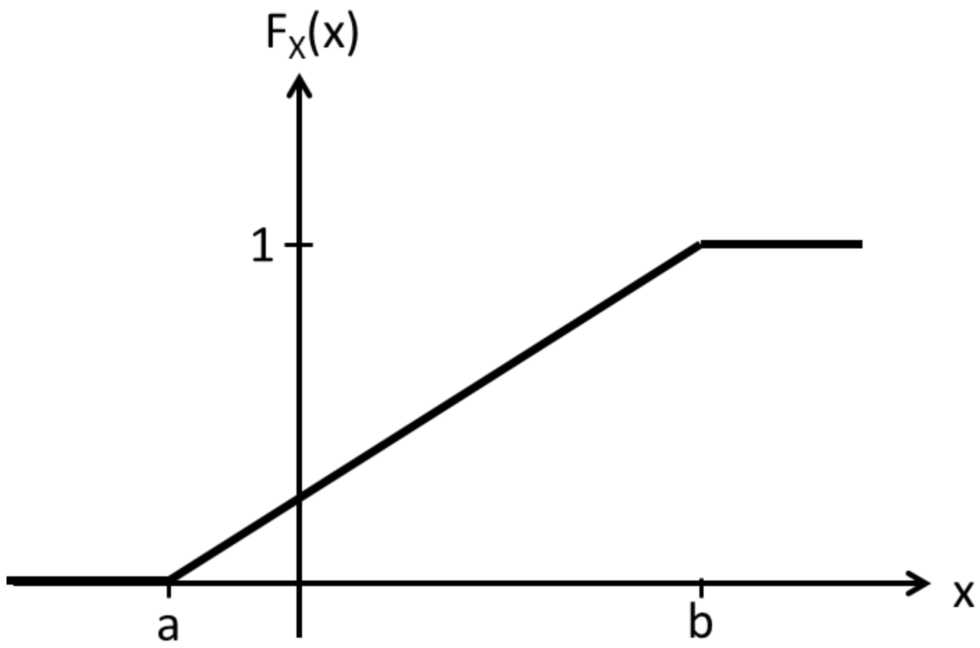
- Mean and variance:

$$E[X] = n \cdot p$$
$$Var(X) = n \cdot p \cdot (1-p)$$

Uniform Distribution (continuous)

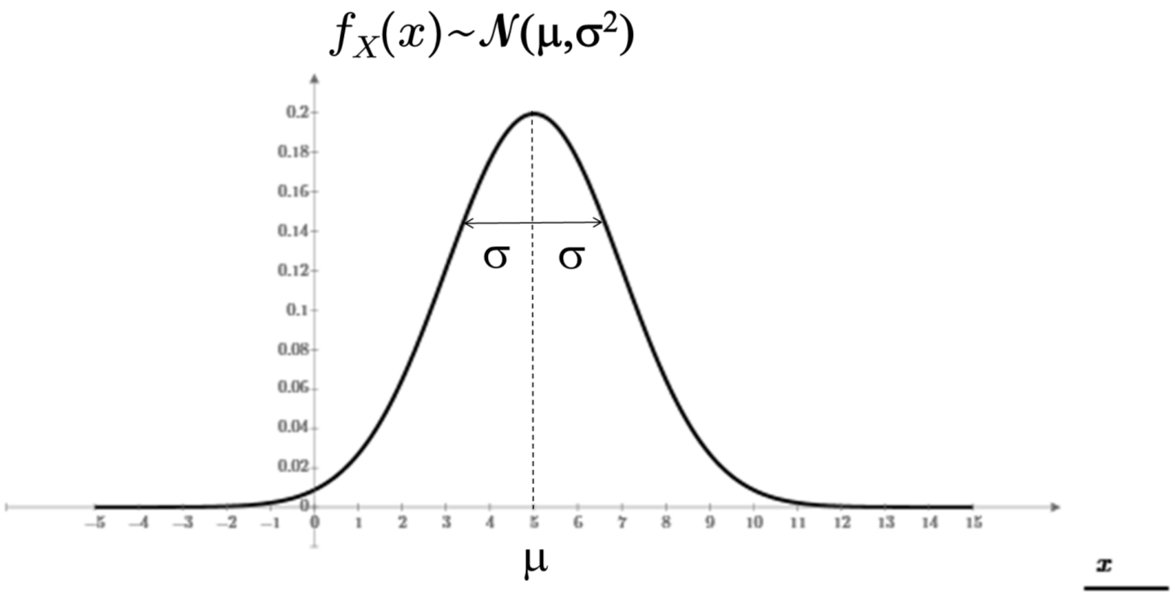
- $\mathcal{U}$(a,b)
- Mean value: $\mu = \frac{a+b}{2}$
- Variance: $\sigma^2 = \frac{1}{12}(b-a)^2$
- pdf: $f_X(x) = \begin{cases} \frac{1}{b-a} & for\ a \le x \le b \\ 0 & otherwise \end{cases}$
- cdf: $F_X(x) = \begin{cases} 0 & for\ x \le a \\ \frac{x-a}{b-a} & for\ a \le x \le b \\ 1 & for\ x \ge b \end{cases}$



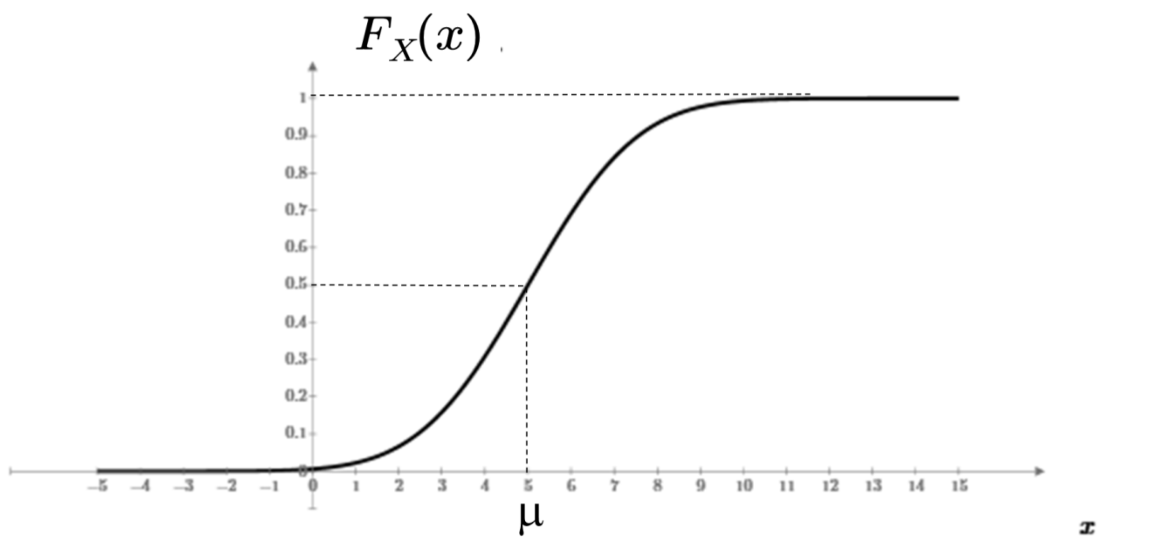


Gaussian Distribution = Normal Distribution

- $\mathcal{N}(\mu,\sigma^2)$
- Mean value: μ
- Variance: σ^2



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



- cdf: $F_X(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$

No closed expression for the cdf

erf= error-function: $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$

Gaussian Distribution = Normal Distribution

- Beregninger med normalfordelinger: Tabelopslag og Matlab:
- $X \sim \mathcal{N}(\mu,\sigma^2) \;\rightarrow\; Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ (Standard Normal Distribution)
- $F_X(x) = Pr(X \leq x) = Pr\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z(z) \qquad \text{hvor } \; \mathrm{z} = \frac{x-\mu}{\sigma}$

$$= \begin{cases} \Phi(z) & Tabel\ 1\ (\text{"}Statistik\ og\ Sandsynlighedsregning\text{"}) \\ 1 - \mathrm{Q}(z) & App.D\ (\text{"}Random\ Signals\text{"}) \end{cases}$$

- $\Phi(z) = Pr(Z \leq z)$
- $\Phi(-z) = 1 - \Phi(z)$
- $\mathrm{Q}(z) = Pr(Z \geq z) = 1 - Pr(Z \leq z) = 1 - \Phi(z)$
- $\mathrm{Q}(-z) = 1 - \mathrm{Q}(z)$

- Matlab:
  - $\Pr(X \leq x) = F_X(x) = normcdf(x,\mu,\sigma)$
  - $\Pr(Z \leq z) = F_Z(z) = normcdf(z,0,1) = normcdf(z)$

Very important!

i.i.d.: Independent and Identically distributed

- We define that for series of random variables that is taken from the same distribution (identically distributed), and are sampled independent of each other, that they are i.i.d.

$$\boxed{\text{i.i.d. = Independent and Identically distributed}}$$

- i.i.d. is a very important characteristic in stochastic variable processing and statistics

Example:

- Quantisation noise.

Very important!

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2
- Let $\bar{X}$ be the random variable (average):

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

- Then in the limit: $n \to \infty$ we have that: $\underline{\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})}$

i.e. in the limit $\bar{X}$ will be normally distributed with mean = μ and variance = $\frac{\sigma^2}{n}$.

The variance is reduced with a factor 1/n

Very important!

Central Limit Theorem

- Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with mean μ and variance σ^2
- Let X be the random variable:

$$X = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^{n} \frac{1}{n} X_i - \mu}{\sqrt{\sigma^2/n}} = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}$$

- Then in the limit: $n \to \infty$ we have that: $X \sim \mathcal{N}(0,1)$

  i.e. in the limit X will be normally distributed with mean = 0 and variance = 1 (standard normal distributed).

Sampling From Any Distribution

For test or simulation you need testdata (”measurements”) randomly sampled from a given distribution:

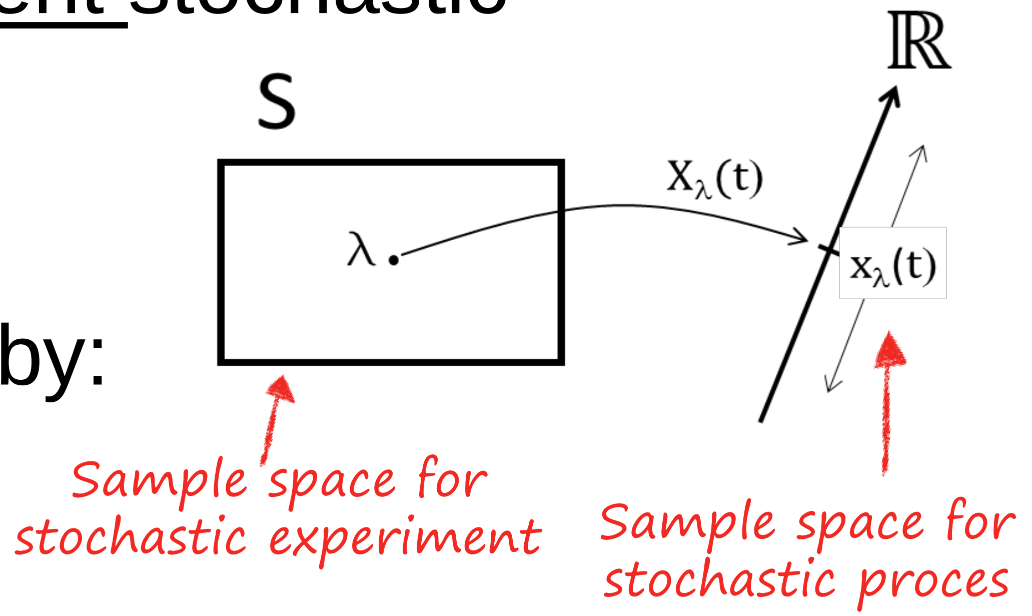
- Find the cdf of the distribution: $F_X(x)$
- Find the inverse of the cdf: $y = F_X(x) \;\Rightarrow\; x = F_X^{-1}(y)$
- Draw a ramdom sample: $y \sim \mathcal{U}[0;1]$
- Insert into the inverse cdf: $x = F_X^{-1}(y)$
- The samples $X = x$ is distributed according to: $F_X(x)$

Stochastic Processes

Definitions:

- A stochastic process is a time dependent stochastic variable:

$$X(t)$$



- A discrete stochastic process is given by:

$$X[n] = X(nT)$$

where n is an integer.

- Random events that develops in time

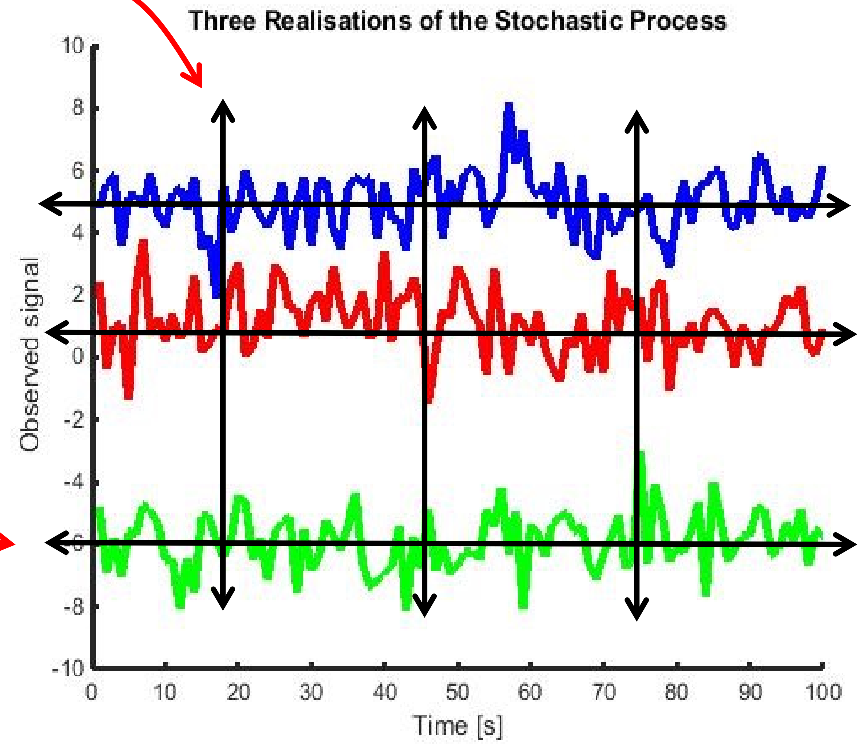
- A sample function (observed signal) is a realization of a stochastic process $x(t)$

The Mean Functions

- Ensemple mean:

The mean of all possible realizations to time t

$$\mu_{X(t)}(t) = E[X(t)] = \int_{-\infty}^{\infty} x(t) f_{X(t)}(x(t))\, dx(t)$$



The time average for one realization of the stochastic process

- Temporal mean:

$$\hat{\mu}_{X_i} = \langle X_i \rangle_T = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt$$

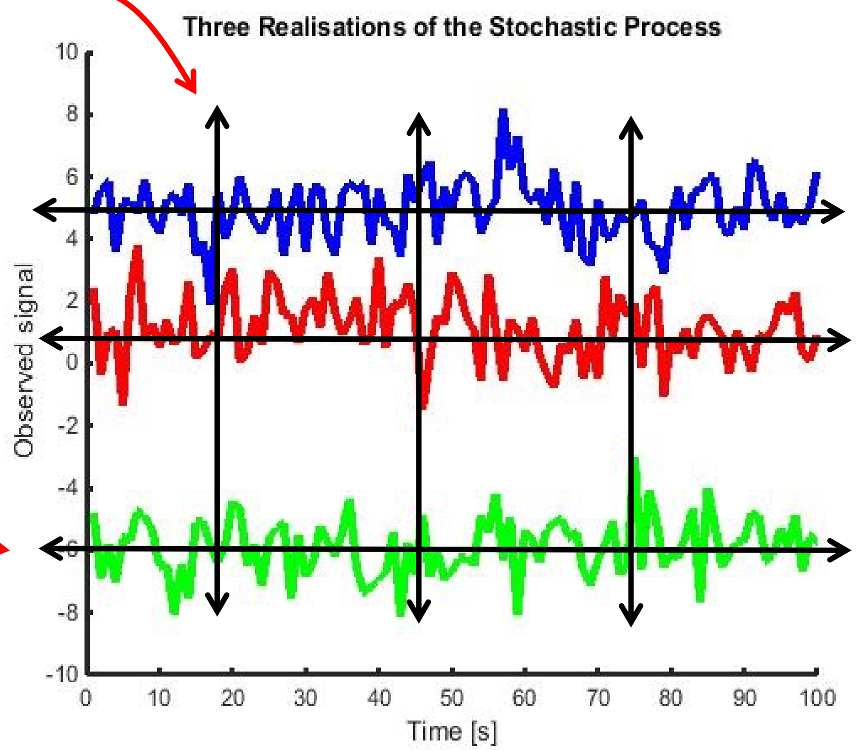
$$\left(\lim_{T\to\infty} \frac{1}{T} \int_{0}^{T} x_i(t)\, dt \right)$$

The Variance Functions

The variance of all possible realizations to time t

- Ensemple variance:

$$Var\left(X(t)\right) = \sigma^2_{X(t)}(t) = E[\left(X(t) - \mu_{X(t)}(t)\right)^2]$$



The variance over time for one realization of the stochastic process

- Temporal variance:

$$\hat{\sigma}^2_{X_i} = \langle X_i^2 \rangle_T - \langle X_i \rangle_T{}^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left(x_i(t)^2 - \hat{\mu}_{X_i}{}^2\right) dt = Var(X_i)$$

$$\left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \left(x_i(t)^2 - \hat{\mu}_{X_i}{}^2\right) dt\right)$$

Stationarity in the Wide Sense (WSS)

- Ensemble mean is a constant

Can be tested.

$$\mu_X(t) = E[X(t)] = \mu_X$$ - independent of time

- Ensemble variance is a constant

$$\sigma_X^2(t) = E\left[X(t)^2\right] - E[X(t)]^2 = \sigma_X^2$$ - independent of time

Stationarity in the Strict Sense (SSS):

Difficult to test in reality.

- The density function $f_{X(t)}\left(x(t)\right)$ do not change with time

Ergodicity

- We can say something about the properties of the stochastic process in general based on one sample function, as long as we have observed it for long enough.

- If ensemble averaging is equivalent to temporal averaging:

$$\mu_X(t) = \bar{X}(t) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i(t)\, dt = \langle X_i \rangle_T = \hat{\mu}_{X_i}$$

- For any moment: In practice: n=2 (Variance)

$$\overline{X^n} = \int_{-\infty}^{\infty} x^n f_X(x)\, dx = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x_i^n\,(t)\, dt$$

One realization Ensemple (WSS)

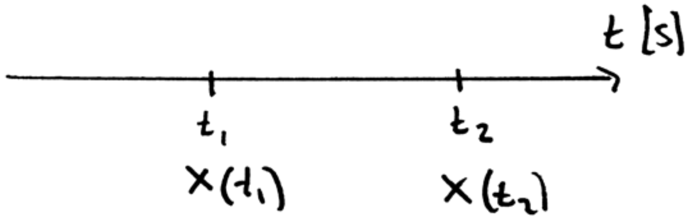
$$\left.\begin{aligned} \langle X_i \rangle_T &= \mu_X \\ \hat{\sigma}_{X_i}^2 &= \sigma_X^2 \end{aligned}\right\} \to Ergodic$$

All information is achieved with one measurement (realization)

Comparing realizations

Correlations

- We compare the process at two different times



Correlation of a realization with itself

- <u>Autocorrelation</u>: $R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)^*]$
  - ➢ Says something about how much the signal $X(t_1)$ resembles itself at time t_2

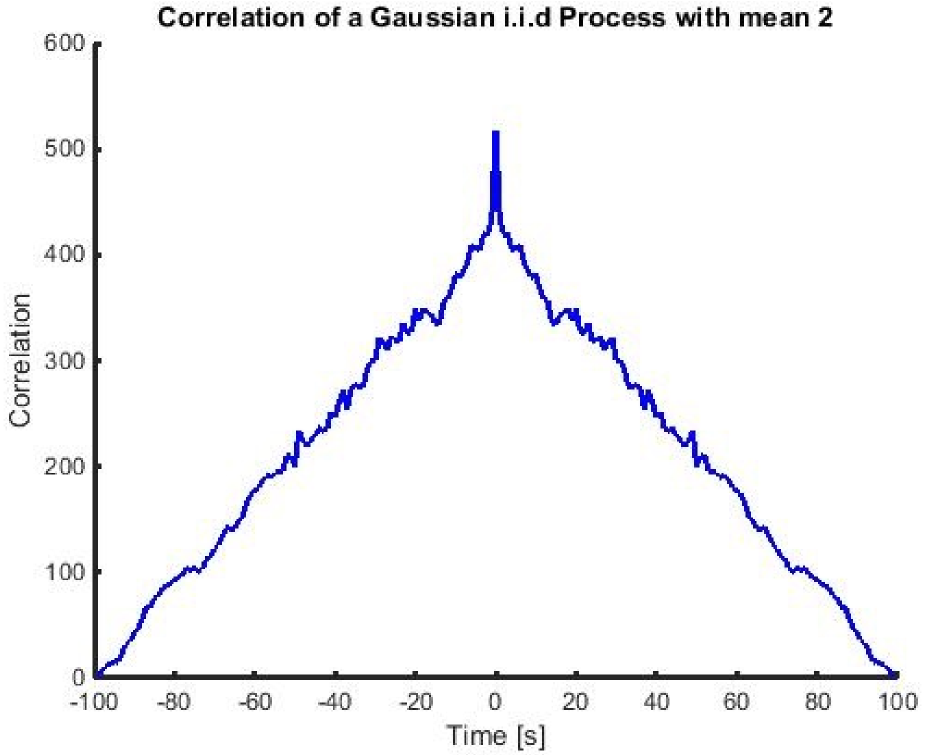
Correlation of two realizations

- <u>Crosscorrelation</u>: $R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*]$
  - ➢ Can be used to look for places where the signal $X(t)$ is similar to the signal $Y(t)$

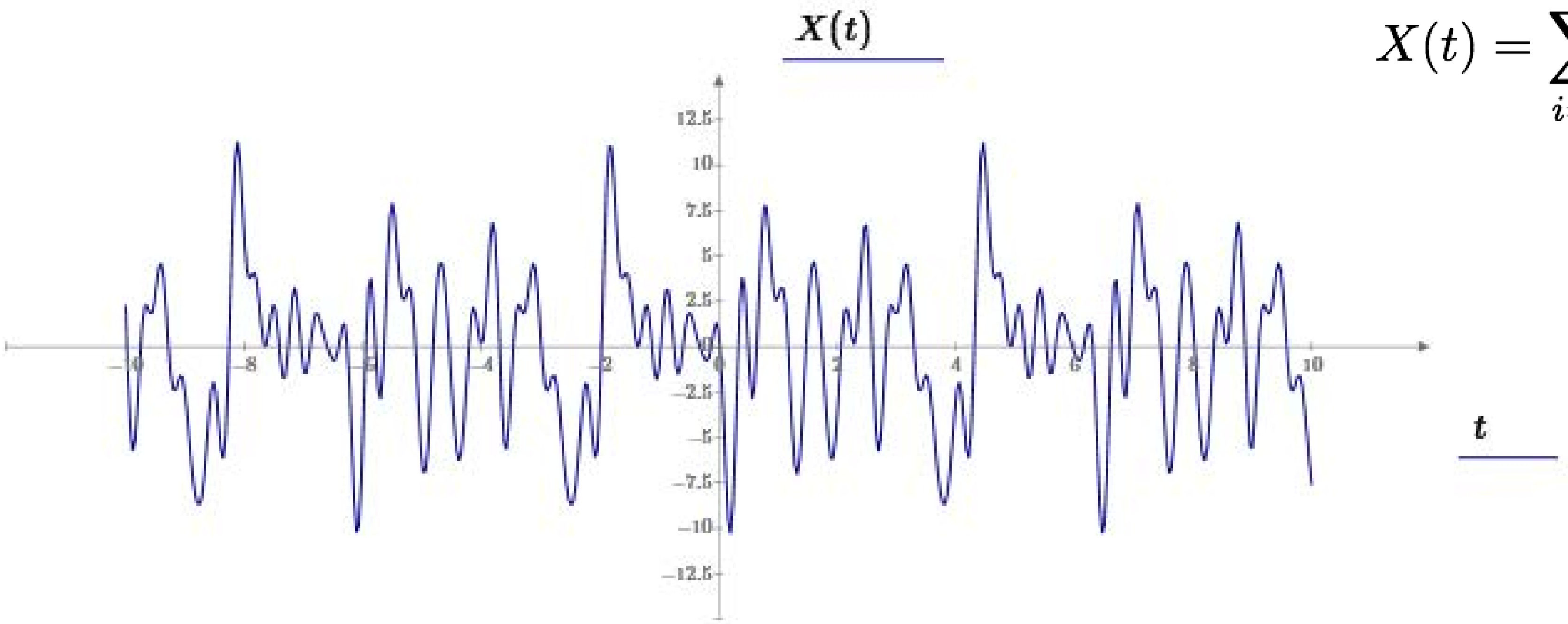
Autocorrelation

- For Real WSS: $R_{XX}(\tau) = E[X(t)X(t+\tau)]$

- Properties of the autocorrelation function $R_{XX}(\tau)$:
  - An even function of τ $(R_{XX}(\tau) = R_{XX}(-\tau))$
  - Bounded by: $|R_{XX}(\tau)| \leq R_{XX}(0) = E\left[X^2\right]$ (max. in $\tau = 0$)
  - If X(t) changes fast, then $R_{XX}(\tau)$ decreases fast from $\tau = 0$
  - If X(t) changes slowly, then $R_{XX}(\tau)$ decreases slowly from $\tau = 0$
  - if X(t) is periodic, then $R_{XX}(\tau)$ is also periodic



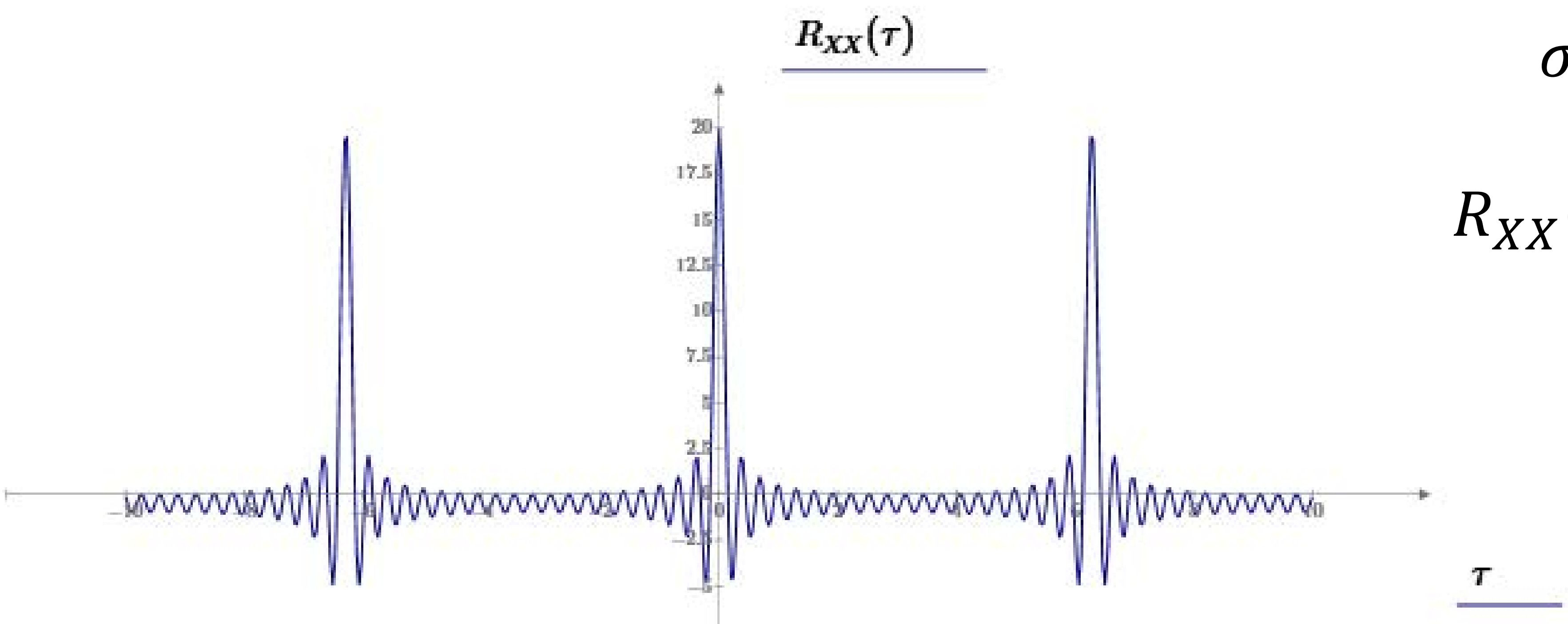
Uncalibrated Noisy Signal



$$X(t) = \sum_{i=1}^{n} A_i \cos \omega_i t + B_i \sin \omega_i t$$

$$A_i, B_i \sim \mathcal{N}(0, \sigma^2)$$

$$\omega_i = i \cdot \omega_0$$
$$\omega_0 = 1$$



$$\sigma = 1, n = 20$$

$$R_{XX}(0) = n\sigma^2 = 20$$

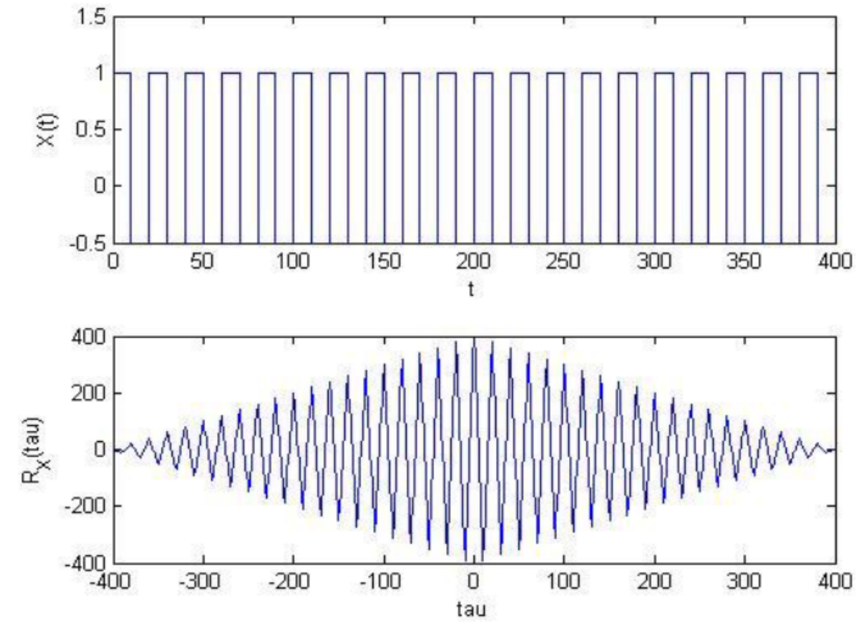
Random Binary (Digital) Signal

Deterministic:

Periodic signal

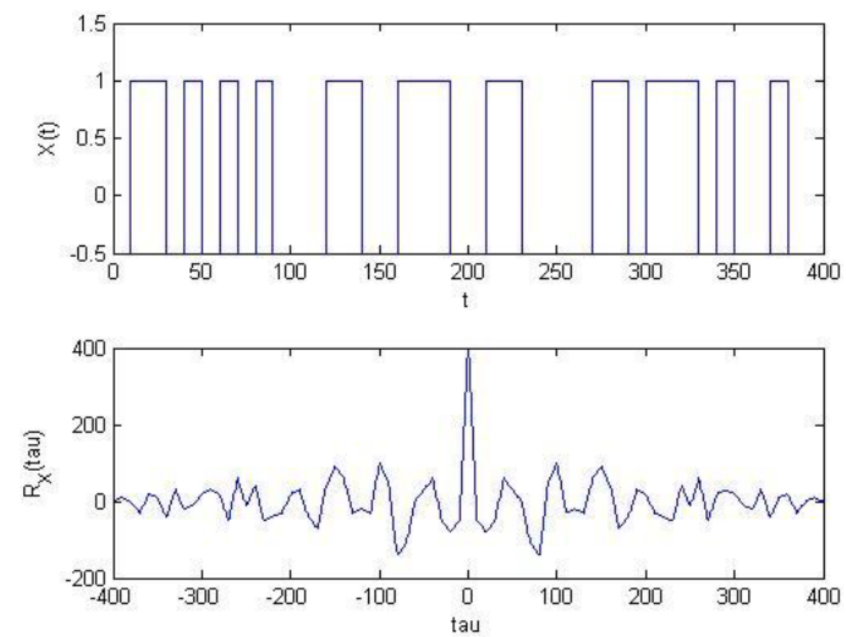
↓

R_{XX} periodic




```
Rx = conv(x,fliplr(x));
```

Non-deterministic
(Stochastic)




```
Rx = conv(x,fliplr(x));
```

Tells about how much we can predict the future

Autocovariances

- Autocovariance function:

$$C_{XX}(t_1,t_2) = E\left[\left(X(t_1) - \mu_X(t_1)\right)(X(t_2) - \mu_X(t_2))^*\right]$$
$$= R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X(t_2)$$

Especially: $C_{XX}(t,t\,) = E\left[(X(t) - \mu_X(t))^2\right] = E\left[X(t)^2\right] - E[X(t)]^2 = \sigma_X^2(t)$

- Autocorrelation coefficient:

$$r_{XX}(t_1,t_2) = \frac{C_{XX}(t_1,t_2)}{\sqrt{C_{XX}(t_1,t_1)C_{XX}(t_2,t_2)}}; \qquad 0 \le r_{XX}(t_1,t_2) \le 1$$

Especially: $r_{XX}(t,t) = 1$ ($X(\mathrm{t})$ is totally dependent of itself!)

Two Stochastic Processes

- If we have two stochastic processes $X(t)$ and $Y(t)$
  - We can compare them by looking at the cross-correlation and cross-covariance:

Cross-correlation $\quad R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*]$

Cross-covariance $\quad C_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)^*] - E[X(t_1)]E[Y(t_2)]$

Cross-Correlation Functions

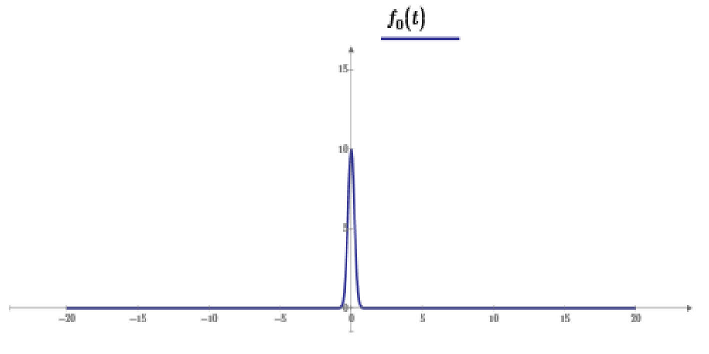
- For Real WSS processes $X(t)$ and $Y(t)$:

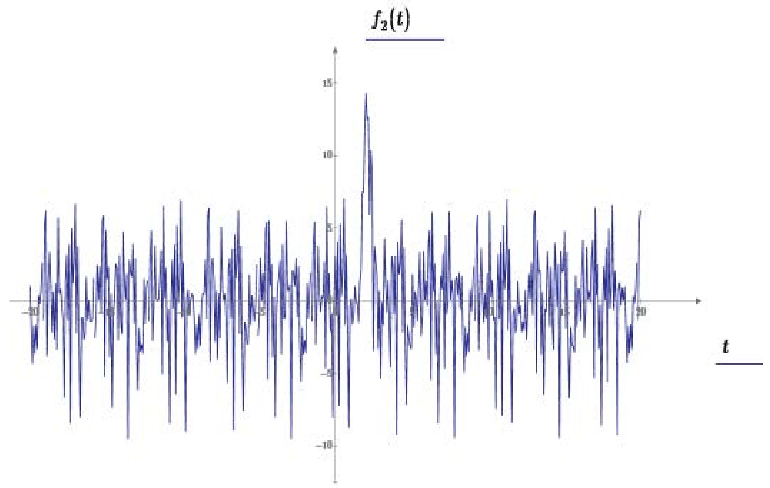
$$R_{XY}(\tau) = E[X(t)Y(t+\tau)]$$

- Properties of the cross-correlation function $R_{XY}(\tau)$:
  - $R_{XY}(\tau) = R_{YX}(-\tau))$
  - $|R_{XY}(\tau)| \leq \sqrt{R_{XX}(0)R_{YY}(0)} = \sqrt{E[X^2]E[Y^2]}$ (max. in $\tau = 0$)
  - $|R_{XY}(\tau)| \leq \frac{1}{2}(R_{XX}(0) + R_{YY}(0))$
  - If $X(t)$ and $\mathrm{Y}(t)$ are orthogonal, then $R_{XY}(\tau) = 0$
  - If $X(t)$ and $\mathrm{Y}(t)$ are independant, then $R_{XY}(\tau) = \mu_X \cdot \mu_Y$

Cross-correlation – Uncalibrated noisy signal

- Comparing two signals:
  - An uncalibrated and noisy signal $f_2(t)$
  - Reference signal $f_0(t) = 10 \cdot e^{-10t^2}$

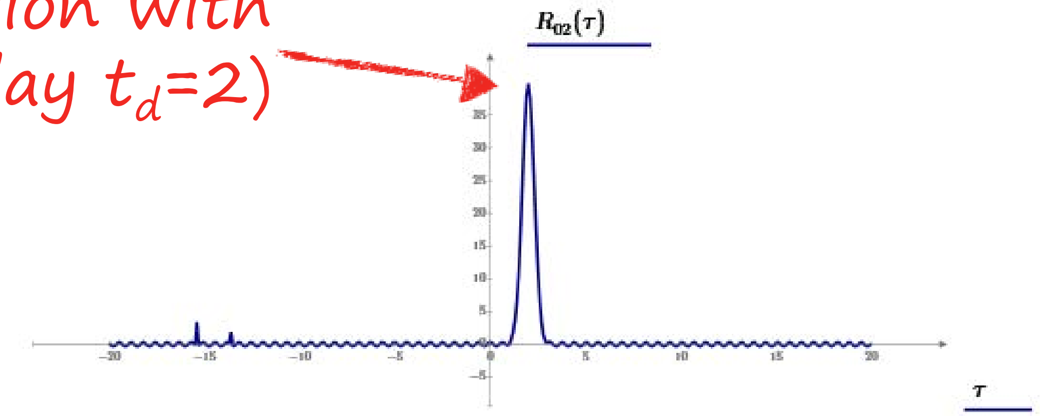




- Cross-correlation:

$$R_{02}(\tau) = \int_{-\infty}^{\infty} f_0(t) \cdot f_2(t+\tau)dt$$

Correlation with time delay t_d=2)



Power Spectral Density (psd)

- WSS random signals X(t):

- Power Spectral Density Function (psd):

Fourier-transform

$$S_{XX}(f) = \mathcal{F}(\langle R_{XX}(\tau)\rangle_{T_0}) = \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j\cdot 2\pi f\cdot\tau}d\tau$$

Invers Fourier-transform

$$\Rightarrow R_{XX}(\tau) = \mathcal{F}^{-1}(\langle R_{XX}(\tau)\rangle = \int_{-\infty}^{\infty} S_{XX}(f)e^{j\cdot 2\pi f\cdot\tau}df$$

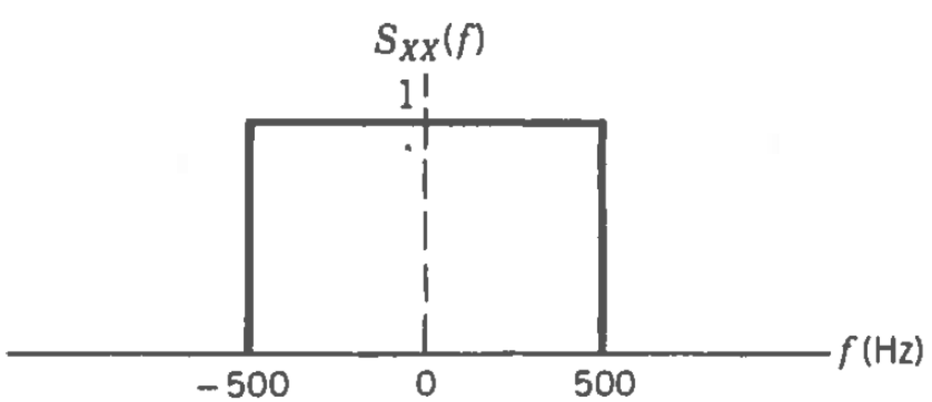


Figure 3.19*a* Psd of a lowpass random process $X(t)$.

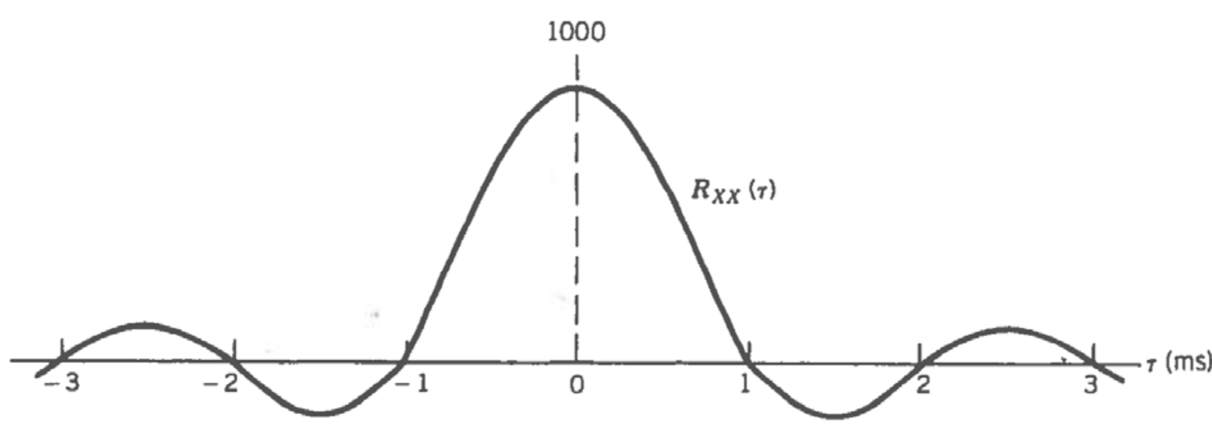


Figure 3.19*b* Autocorrelation function of $X(t)$.

Figures from ”Random Signals”

Power Spectral Density (psd)

- Properties of psd $S_{XX}(f)$ (spectrum of $X(t)$):
  - $S_{XX}(f) \in \mathbb{R}$
  - $S_{XX}(f) \geq 0$
  - If $X(t) \in \mathbb{R}$: $R_{XX}(-\tau) = R_{XX}(f)$ and $S_{XX}(-f) = S_{XX}(f) \rightarrow$ even functions
  - If $X(t)$ periodic components: $S_{XX}(f)$ will have impulses (δ−functions)
  - $[S_{XX}(f)] = \frac{W}{Hz} \rightarrow$ Distribution of power with frequency (power spectral density of the stationary random process X(t)
  - $P_X = E\left[X(t)^2\right] = R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(f)df$

    i.e. if $X(t) = V(t)$ (voltage signal)

    $\rightarrow P_X =$ power in 1Ω-resistor

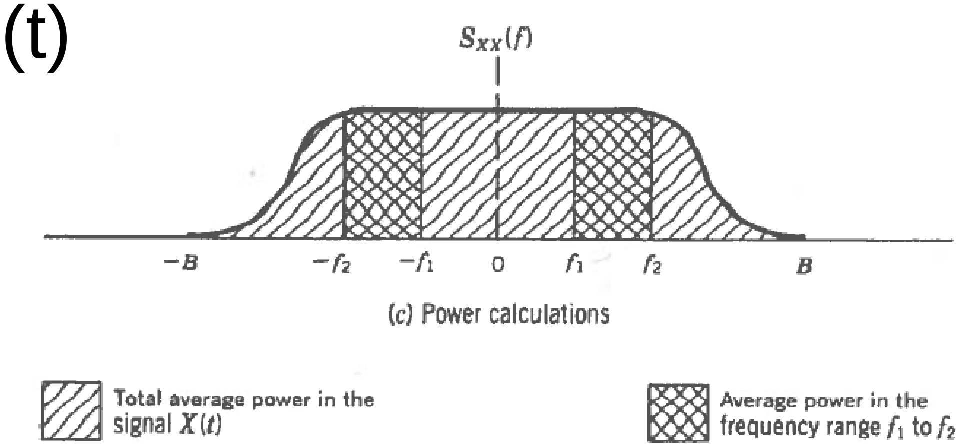


Figure from ”Random Signals”

  - $P_X[f_1, f_2] = 2\int_{f_1}^{f_2} S_{XX}(f)df \rightarrow$ Power in the frequency-interval $[f_1, f_2]$

Words and Concepts to Know

Probability density function Binomial coefficient Cross-covariance Convolution
Deterministic Rayleigh Distribution Deterministic Intersection Type I Error SSS
pdf Temporal cross-correlation Cross-correlation Correlation Markov chain
Probability Mass Function i.i.d. Temporal mean Continuous random variable
Randomly Sampled Data Temporal variance Marginal Correlation coefficient
Stochastic Processes Unordered Mutually Exclusive/Disjoint Ensemple variance
Uniform distribution Replacement Sampling Non-deterministic Ergodicity
Sample point Specificity Stationarity Gaussian distribution Sample space
Central Limit Theorem Experiment/Trial cdf Complement/not Joint pmf WSS
Likelihood Simultanious pmf Covariance
Independent and Identically Distributed Event
Relative frequency Realization Independence Union Correlation coefficient
Normal distribution Sensitivity Combinatorics
Bivariate Normal Distribution
Transformation of stochastic variables Binomial distribution
Joint events
Empty set/Null set Binomial Mass Function Standard deviation
Strict Sense Stationary Ordered Set Conditional probability Total probability
Mean Simultaneous density function Variance Bayes Rule pmf Ensemple mean
Autocovariance Type II Error Autocorrelation Coefficient Joint density function
Power Spectral Density Non-deterministic Stochastic Posterior Autocorrelation
Wide Sense Stationary Bernoulli Trial Prior Expectation Subset
Cumulative Distribution Function psd Marginal probability density function

Assignment 8

- Find a stochastic process in your area

  (discharge of a capacitor, bitrate, failure, hight, weight, …)
- Make a signal model: $X(t) = \cdots$
- Make three realizations
- Determine the ensemble mean and variance
- Determine the temporal mean and variance
- Determine stationarity and ergodicity