Statistics and Sensor Data Fusion

2. Probability Calculus

Probability Calculus

Goal: Mathematical treatment of random phenomena

2.1 Fundamental Concepts of Probability Calculus

Fundamental Concepts of Probability Calculus

Random Experiments:

A random experiment...

- ...is a procedure with **two or more possible outcomes**, where the outcome of one trial is not deterministic but **depends on chance**
- ...is carried out by following **verifiable instructions**
- ...can in principle be repeated an **arbitrary number of times**

The most elementary random experiment can be realized by **tossing a coin** and is called heads or tails.

Fundamental Concepts of Probability Calculus

Possible Outcomes of a Random Experiment:

The possible outcomes of random experiment are called simple or elementary events, which are **mutually exclusive**.

Examples:

- Tossing a coin $\rightarrow$ Possible outcomes: "Head" (H) or "Tail" (T)
- Rolling a die $\rightarrow$ Possible outcomes: "1", "2", "3", "4", "5", "6"
- Tossing a coin twice $\rightarrow$ Possible outcomes:
  (H, H), (H, T), (T, H), (T, T)
- Rolling two distinguishable dice $\rightarrow$ Possible outcomes:
  $(1,1),\ldots,(6,6)$

Fundamental Concepts of Probability Calculus

Sample Space and Events:

The set of all elementary events of a specific random experiment is called sample space

$$\Omega = \{\omega_1, \dots, \omega_n\}$$

where $\omega_1, \dots, \omega_n$ denote the outcomes or elementary events. Any **subset** $A \subseteq \Omega$ of the sample space Ω is called an event.

Examples:

- Sample space when rolling a die:

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

- Event $A \subseteq \Omega$ indicating that the outcome is odd:

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

Fundamental Concepts of Probability Calculus

Exercise:

Consider the sample space Ω for the random experiment of rolling two dice, where the two dice can be distinguished by their color (e.g. blue and red).

Indicate the event $A \subseteq \Omega$

(a) that the sum of the two dice can be divided by 4

(b) that the product of the two dice is larger than 8 but smaller than 20

Fundamental Concepts of Probability Calculus

Occurrence and Complement of an Event:

It is said that an event $A \subseteq \Omega$ occurs, if the actual outcome ω_j of the random experiment is an element of A, i.e. if $\omega_j \in A$.

With respect to a given event $A \subseteq \Omega$, the set

$$\bar{A} = \Omega \setminus A$$

defines the complementary event or complement of A.

The complement of A occurs **if and only if** A does not occur.

Example:

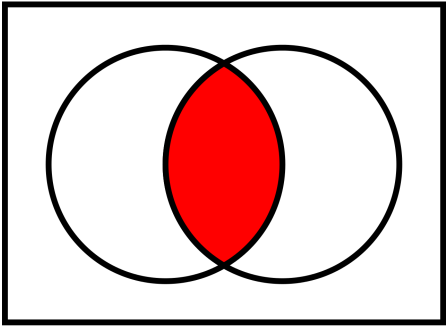
Considering the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the event $A = \{1, 3, 5\}$, the complementary event is $\bar{A} = \Omega \setminus A = \{2, 4, 6\}$, i.e. the outcome of rolling a die once is either “odd” or “even”.

Fundamental Concepts of Probability Calculus

Venn Diagrams as Illustrations of Events:

Sets and set operations can often be illustrated by Venn diagrams.

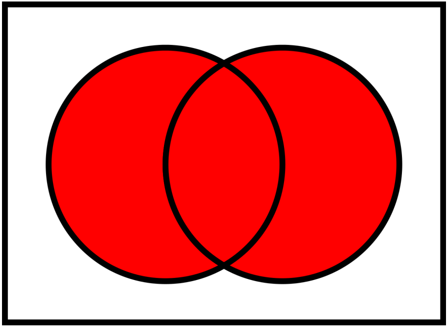
Intersection of Two Sets:



$A \cap B$:   "A and B"

Fundamental Concepts of Probability Calculus

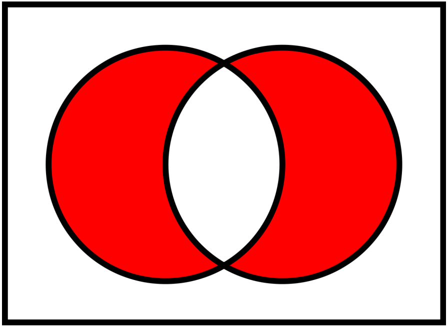
Union of Two Sets:



$A \cup B$: "A or B"

Fundamental Concepts of Probability Calculus

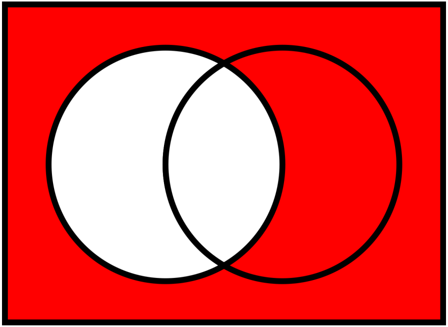
Symmetric Difference of Two Sets:



$$A \Delta B = (A \cup B) \setminus (A \cap B): \quad \text{“either A or B”}$$

Fundamental Concepts of Probability Calculus

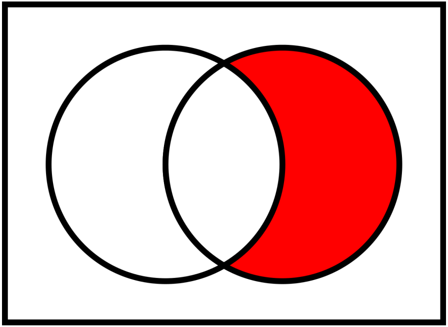
Complement of A:



$$\bar{A} = \Omega \setminus A : \quad \text{“not A”}$$

Fundamental Concepts of Probability Calculus

Relative Complement of A in B:



$B \setminus A = \bar{A} \cap B$: "B but not A"

Fundamental Concepts of Probability Calculus

For **arbitrary sets** $A, B, C \subseteq \Omega$, the following calculation rules for the set operations $\cap$ and $\cup$ hold:

Associative Laws:

$$(A \cap B) \cap C = A \cap (B \cap C) \qquad (A \cup B) \cup C = A \cup (B \cup C)$$

Commutative Laws:

$$A \cap B = B \cap A \qquad A \cup B = B \cup A$$

Distributive Laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Further Rules:

$$A \cap A = A \qquad A \cup A = A$$

$$A \cap \varnothing = \varnothing \qquad A \cup \varnothing = A$$

$$A \cap B = A \quad \text{if } A \subseteq B \qquad A \cup B = B \quad \text{if } A \subseteq B$$

Fundamental Concepts of Probability Calculus

Classical Approach to Probability:

A Laplace experiment is a random experiment with a finite number of outcomes, where all the possible outcomes or elementary events are **equally probable**.

For an event $A \subseteq \Omega$, the probability $P(A)$ of the event A is in this case defined by

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of all possible outcomes}} = \frac{|A|}{|\Omega|}$$

Example: For the Laplace experiment "rolling a die", the probability of the event $A = \{1, 3, 5\}$ would be

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = 0.5$$

Fundamental Concepts of Probability Calculus

Exercise:

Consider again the random experiment of rolling two dice, where the two dice can be distinguished by their color (e.g. blue and red).

What is the probability of the event $A \subseteq \Omega$

(a) that the result of the red die is smaller than the result of the blue die?

(b) that the result of the red die can be divided by the result of the blue die?

Fundamental Concepts of Probability Calculus

Empirical Approach to Probability:

In the **empirical** or **frequentist approach**, probabilities are interpreted as limits of relative frequencies.

In order to determine the unknown probability $P(A)$ of an event A, the underlying random experiment is repeated n times.

During these n trials, the event A is observed to occur in total h_n times, i.e. h_n is the **absolute frequency** of A. In turn, the **relative frequency** of A is given by

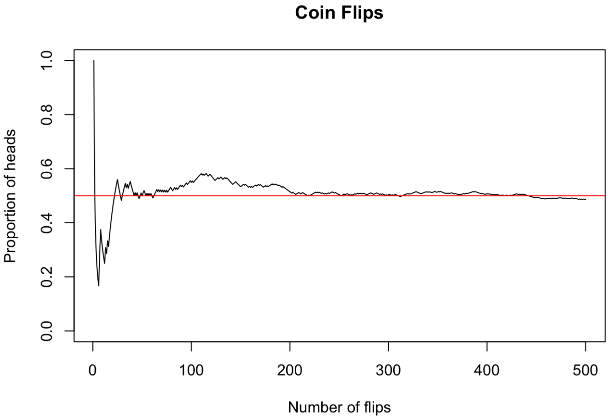
$$f_n = \frac{h_n}{n}$$

With increasing n, the relative frequency f_n converges to the probability $P(A)$ of the event A ("law of large numbers"):

$$\lim_{n \to \infty} f_n = P(A)$$

Fundamental Concepts of Probability Calculus

Law of Large Numbers – Illustration:



Relative frequency of observing "head" when tossing a coin repeatedly.

https://uw-statistics.github.io/Stat311Tutorial/limit-theorems.html

Fundamental Concepts of Probability Calculus

General Approach to Probability – Probability Measure:

- **Goal:** Every event $A \subseteq \Omega$ shall be assigned a number $P(A)$, which indicates the probability for the occurrence of A.
- The set of **all subsets** of Ω (i.e., the set of all events) is represented by the power set $\mathcal{P}(\Omega)$:

$$\mathcal{P}(\Omega) = \{A \mid A \subseteq \Omega\}$$

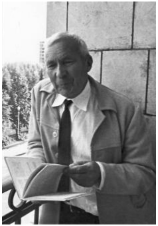
- Formally, a probability measure P is a mapping

$$P : \mathcal{P}(\Omega) \to \mathbb{R}, \quad A \mapsto P(A)$$

- **Question:** What properties are necessary for the mapping P, such that thereupon a consistent calculus of probabilities can be established?

Fundamental Concepts of Probability Calculus

General Approach to Probability –
The Axioms of Kolmogorov:



The Kolmogorov axioms given by A.N. Kolmogorov in 1933 are the foundation of probability theory.

With Ω the sample space of a random experiment, for arbitrary events $A, B \subseteq \Omega$ it holds:

(A1) $\quad P(A) \geq 0$

(A2) $\quad P(\Omega) = 1$

(A3) $\quad P(A \cup B) = P(A) + P(B) \quad \text{if} \quad A \cap B = \varnothing$

Based on these axioms, the further rules of probability calculus can be derived by **logical deduction**.

Fundamental Concepts of Probability Calculus

Exercise:

We derive the following **calculation rules for probabilities** as conclusions from the Kolmogorov axioms:

(a) $P(A) \leq 1$

(b) $P(\varnothing) = 0$

(c) $P(A \setminus B) = P(A) - P(A \cap B)$

(d) $P(\bar{A}) = 1 - P(A)$

(e) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(f) $A \subseteq B \implies P(A) \leq P(B)$

Exercise:

There are two local newspapers published in a country town, the **Morning Post** and the **Evening Standard**.

The probability that a resident reads

- the Morning Post (event A) is 0.6
- the Evening Standard (event B) is 0.5
- at least one newspaper is 0.9

Determine the probability that a resident reads

(a) both newspapers

(b) no newspaper at all

(c) exactly one of the two newspapers

Fundamental Concepts of Probability Calculus

Sometimes one is interested in the probability of an event A **under the condition** that another event B has occurred:

Conditional Probability:

The conditional probability of an event A under the condition that an event B has occurred, is defined by

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

where it should hold that $P(B) > 0$.

By multiplying this equation by $P(B)$, one gets

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

Fundamental Concepts of Probability Calculus

Conditional Probability – Example: Rolling a die

What is the conditional probability of "tossing a 6" **under the condition** that the result is even?

- The corresponding events are $A = \{6\}$ and $B = \{2, 4, 6\}$
- The conditional probability of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{6\})}{P(\{2,4,6\})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3} \neq P(A)$$

- In this case, the knowledge that the event B has occurred **affects** the probability of event A

Fundamental Concepts of Probability Calculus

In some situations, the knowledge that a certain event B has occurred **does not affect** the probability of another event A:

Statistical Independence:

If for two events A and B it holds that

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

the events A and B are called statistically independent.

Events that are **not** independent are called dependent.

For two **independent events** A and B, it holds that

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

Fundamental Concepts of Probability Calculus

Statistical Independence – Example: Rolling a die

Are the two events "tossing a 6" ($A = \{6\}$) and "the result is even" ($B = \{2, 4, 6\}$) independent?

- The conditional probability of A given B is

$$P(A|B) = \tfrac{1}{3}$$

- It holds that

$$P(A|B) = \tfrac{1}{3} \neq \tfrac{1}{6} = P(A)$$

$$P(A \cap B) = P(A) = \tfrac{1}{6} \neq \tfrac{1}{6} \cdot \tfrac{1}{2} = P(A) \cdot P(B)$$

i.e. the events A and B are **dependent**

Question: What are the values for $P(B)$ and $P(B|A)$?

Fundamental Concepts of Probability Calculus

Exercise:

Consider the random experiment of rolling two dice, where the dice can be distinguished by their color (e.g. blue and red).

(a) What is the **conditional probability** that the red die shows the number 4 under the assumption that the sum of the two dice is 9?

(b) Are the two events "the sum of the two dice is 7" and "the red die shows the number 1" **statistically independent**?

Fundamental Concepts of Probability Calculus

Law of Total Probability:

Assume that the events $B_1, \ldots, B_n$ are a complete decomposition of the sample space Ω into **mutually exclusive** or **disjoint events**:

(i) $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$ (complete decomposition)

(ii) $B_i \cap B_j = \varnothing$ for $i \neq j$ (mutually exclusive)

(iii) $P(B_i) > 0$ (no impossible events)

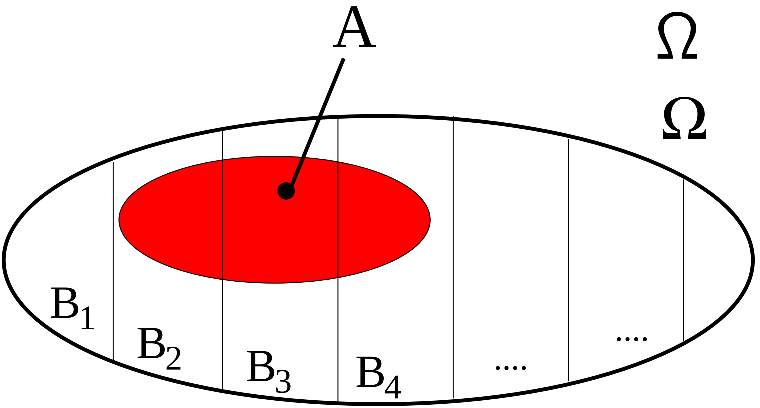
Under the above assumptions, for any event $A \subset \Omega$ it holds that

$$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)}$$

Law of total probability

Fundamental Concepts of Probability Calculus

Law of Total Probability – Illustration:



Fundamental Concepts of Probability Calculus

Exercise:

A production facility orders manufacturing components from three different suppliers S_1, S_2 and S_3:

- S_1 delivers 60 % of the components at a scrap rate of 9 %
- S_2 delivers 25 % of the components at a scrap rate of 12 %
- S_3 delivers 15 % of the components at a scrap rate of 4 %

Out of the whole delivery (100 %), one manufacturing component is chosen randomly.

(a) What is the probability that the chosen component is scrap and what is the probability that it is okay?

(b) After choosing a component randomly it turns out to be scrap. What is the probability that it was delivered by supplier S_3?

Fundamental Concepts of Probability Calculus

Bayes' Rule:

Bayes' rule enables conclusions about the probability of events that are **not directly observable**.

For a complete decomposition of the sample space Ω into mutually disjoint events $B_1, \ldots, B_n$ and an event $A \subseteq \Omega$ it holds:

$$\boxed{P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^{n} P(B_i) \cdot P(A|B_i)}}$$

$P(B_k)$ is called prior probability of B_k (before A is observed).

$P(B_k|A)$ is called posterior probability of B_k (after A was observed).

Bayes' rule plays a major role in **statistical pattern recognition**.

Fundamental Concepts of Probability Calculus

Exercise:

During the operation hours of a power plant, there are three different **operation modes** called “high”, “low” or ”medium”:

- operation mode “high” has a probability of 0.1
- operation mode “low” has a probability of 0.3
- operation mode “medium” has a probability of 0.6

The different operation modes have different **outage probabilities**:

- during “high” the outage probability is 0.5
- during “low” the outage probability is 0.1
- during “medium” the outage probability is 0.2

There was an outage of the power plant. What is the probability that the plant was in the operation mode “high”?

2.2 Discrete Random Variables

Discrete Random Variables

Concept of a Random Variable:

Consider a random experiment with sample space Ω.
A random variable X is a **real-valued function**

$$X : \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega)$$

which assigns a meaningful **numerical value** $X(\omega)$ to every possible outcome $\omega \in \Omega$ of the random experiment.

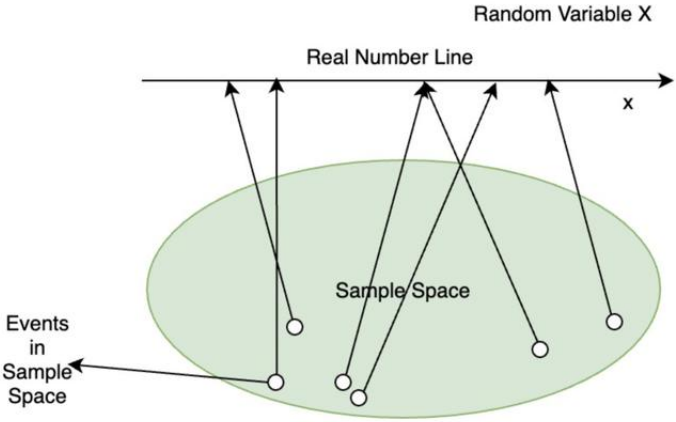
The values $X(\omega) \in \mathbb{R}$ are called the realizations of X.

The **probability of a value** $x \in \mathbb{R}$ is equal to the probability of the corresponding event $A \subseteq \Omega$ consisting of all outcomes $\omega \in \Omega$ which lead to the realization $X(\omega) = x$:

$$P(X = x) = P(\underbrace{\{\omega \in \Omega \mid X(\omega) = x\}}_{A}) = P(A)$$

Discrete Random Variables

Concept of a Random Variable – Illustration:



https://www.geeksforgeeks.org/probability-distribution/

Discrete Random Variables

Random Variable – Example:

- Consider the scenario of **picking randomly one student** out of all currently enlisted students of THWS
- The **sample space** Ω would then consist of all students, i.e.

  $$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$$

  with n the total number of enlisted students
- After the student ω_j was selected randomly, you could e.g. consider to
  - ask for his/hers matriculation number
  - measure his/hers height in cm
  - perform a COVID-19 test

Discrete Random Variables

Random Variable – Example:

Let us consider for example the **height measurement**:

- The obtained value for the height of student ω_j in cm would then correspond to the realization of a random variable X defined on the set of all students of THWS:

$$X : \underbrace{\Omega}_{\text{all students}} \to \mathbb{R}, \qquad \omega_j \mapsto \underbrace{X(\omega_j)}_{\text{height of student } \omega_j}$$

- The probability of measuring the height $x \in \mathbb{R}$ is equal to the probability of selecting **any** student with the height $X(\omega_j) = x$:

$$\boxed{P(X = x) = P(\{\omega_j \in \Omega \mid X(\omega_j) = x\})}$$

Discrete Random Variables

More Examples for Random Variables:

- Rolling two dice and considering the sum of the two results:

$$\Omega = \{\omega = (n_1, n_2) \mid n_1, n_2 \in \{1, 2, \ldots, 6\}\}, \quad |\Omega| = 36$$

$$X : \Omega \to \mathbb{R}, \quad \omega \mapsto n_1 + n_2$$

- A company receives a delivery of 100 items, the random variable X should indicate the number of damaged items:

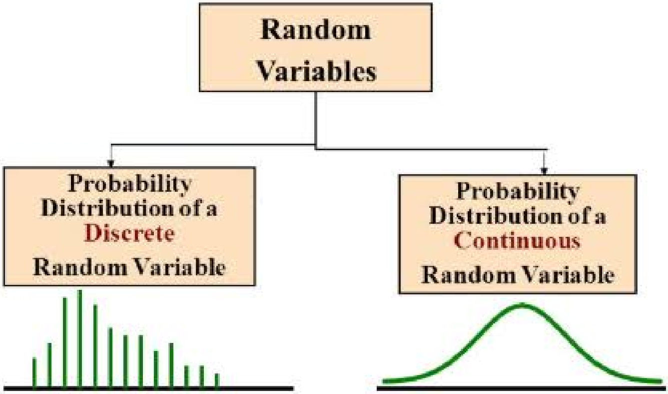
The status of item i is $x_i = \begin{cases} 0 & \text{if item } i \text{ is in order} \\ 1 & \text{if item } i \text{ is damaged} \end{cases}$

$$\Omega = \{\omega = (x_1, \ldots, x_{100}) \mid x_i \in \{0, 1\}\}, \quad |\Omega| = 2^{100}$$

$$X : \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega) = \sum_{i=1}^{100} x_i$$

Discrete Random Variables

Discrete vs. Continuous Random Variables:

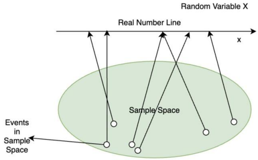


https://prwatech.in/blog/tag/discrete-random-variable/

Discrete Random Variables

Discrete Random Variables:

A random variable X is called discrete, if the set of all possible realizations $X(\omega)$ are **isolated points along the real line**



Example: When tossing two dice and considering the sum of the two results, we have a discrete random variable with the possible realizations $2, 3, \ldots, 12$.

For discrete random variables, all possible realizations can be **enumerated**: $x_1, x_2, x_3, \ldots$

Nevertheless, it is possible that a discrete random variable X can take on **infinitely** many values, e.g. if X indicates the number of coin flips needed to get the result “Head”.

Discrete Random Variables

Probability Mass Function and Cumulative Distribution Function of a Discrete Random Variable:

A discrete random variable X can be described in **two ways**:

The probability mass function (pmf) of a discrete random variable X assigns to every possible realization x_k of X the probability that the value x_k is realized:

$$\boxed{p(x_k) = P(X = x_k)}$$

The cumulative distribution function (cdf) of a discrete random variable X assigns to any value $x \in \mathbb{R}$ the probability that X realizes a value $x_k \leq x$:

$$\boxed{F(x) = P(X \leq x) = \sum_{x_k \leq x} p(x_k)}$$

Discrete Random Variables

Example: Tossing two dice

- Tossing two dice and considering the sum:

$$\Omega = \{\omega = (n_1, n_2) \mid n_1, n_2 \in \{1, 2, \ldots, 6\}\}$$

$$X : \Omega \to \mathbb{R}, \quad \omega \mapsto n_1 + n_2$$

- Exemplary values of the pmf:

$$p(4) = P(X = 4) = P\big(\{(1,3),(2,2),(3,1)\}\big) = \frac{3}{36} = \frac{1}{12}$$

$$p(12) = P(X = 12) = P\big(\{(6,6)\}\big) = \frac{1}{36}$$

$$p(1) = P(X = 1) = P(\varnothing) = 0$$

Discrete Random Variables

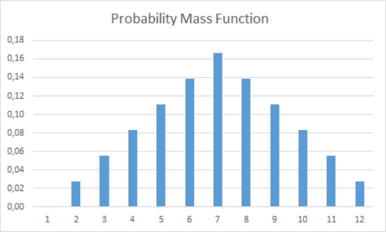
Example: Tossing two dice

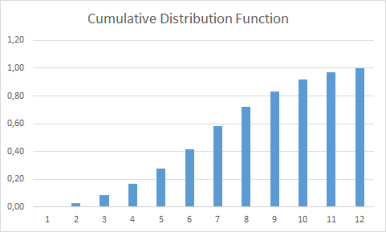
Realization x_k	Corresponding Event	Probability	Cumulated
2	$\{(1,1)\}$	1/36	1/36
3	$\{(1,2),(2,1)\}$	2/36	3/36
4	$\{(1,3),(2,2),(3,1)\}$	3/36	6/36
5	$\{(1,4),(2,3),(3,2),(4,1)\}$	4/36	10/36
6	$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$	5/36	15/36
7	$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$	6/36	21/36
8	$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$	5/36	26/36
9	$\{(3,6),(4,5),(5,4),(6,3)\}$	4/36	30/36
10	$\{(4,6),(5,5),(6,4)\}$	3/36	33/36
11	$\{(5,6),(6,5)\}$	2/36	35/36
12	$\{(6,6)\}$	1/36	36/36

Discrete Random Variables

Example: Tossing two dice

x_k	2	3	4	5	6	7	8	9	10	11	12
$p(x_k)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$
$F(x_k)$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{5}{12}$	$\frac{7}{12}$	$\frac{13}{18}$	$\frac{5}{6}$	$\frac{11}{12}$	$\frac{35}{36}$	1





Discrete Random Variables

Exercise:

Consider the random experiment of tossing two dice and interpret the sum of the numbers as a discrete random variable X.

(a) What is the probability that X lies between 5 and 9?

(b) What is the probability that X is larger than 6?

Discrete Random Variables

A discrete random variable X can be described in terms of its distribution parameters:

Expectation Value, Variance and Standard Deviation:

For a discrete random variable X, we define

- the expectation value or mean

$$\boxed{\mu = E(X) = \sum_k x_k \cdot p(x_k)}$$

- the variance

$$\boxed{\sigma^2 = V(X) = \sum_k (x_k - \mu)^2 \cdot p(x_k) = \left(\sum_k x_k^2 \cdot p(x_k)\right) - \mu^2}$$

- the standard deviation $\sigma = \sqrt{V(X)}$

Discrete Random Variables

Exercise:

Consider again the random experiment of tossing two dice and interpret the sum of the numbers as a discrete random variable X.

Determine the expectation value μ and the variance σ^2 of X.

Discrete Random Variables

The common approach to determine a discrete random variable X is to specify a so-called probability distribution:

Probability Distributions:

Probability distributions for discrete random variables can be characterized by specifying

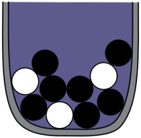
1. the probability mass function (pmf)
2. the cumulative distribution function (cdf)

The most important **discrete probability distributions** are

- the hypergeometric distribution
- the binomial distribution
- the Poisson distribution

Discrete Random Variables

Hypergeometric Distribution:



- The hypergeometric distribution is based on the urn model "sampling without replacement"
- There are in total n balls in the urn from which n_1 balls are black and $n_2 = n - n_1$ balls are white
- By random selection, $m \leq n$ balls are taken out of the urn **without** replacement
- The discrete random variable X should indicate the **total number of black balls** taken out of the urn
- In this case, we say that X is distributed according to the hypgergeometric distribution $\text{Hyp}(n, n_1, m)$ or in short

$$\boxed{X \sim \text{Hyp}(n, n_1, m)}$$

Discrete Random Variables

Hypergeometric Distribution – Properties:

A discrete random variable $X \sim \text{Hyp}(n, n_1, m)$ is characterized by the probability mass function

$$p(k) = P(X = k) = \frac{\binom{n_1}{k} \cdot \binom{n_2}{m-k}}{\binom{n}{m}}, \quad k = 0, \ldots, m$$

In this case, the random variable X has the expectation value

$$\mu = E(X) = m \cdot \frac{n_1}{n}$$

and the variance

$$\sigma^2 = V(X) = m \cdot \frac{n_1}{n} \cdot \left(1 - \frac{n_1}{n}\right) \cdot \frac{n-m}{n-1}$$

Discrete Random Variables

Hypergeometric Distribution – Example:

$$X \sim \text{Hyp}(30, 20, 8)$$

$$p(k) = P(X = k) = \frac{\binom{20}{k} \cdot \binom{10}{8-k}}{\binom{30}{8}}, \quad k = 0, \ldots, 8$$

k	0	1	2	3	4	5	6	7	8
$p(k)$	0.00	0.00	0.01	0.05	0.17	0.32	0.30	0.13	0.02
$F(k)$	0.00	0.00	0.01	0.06	0.23	0.55	0.85	0.98	1.00

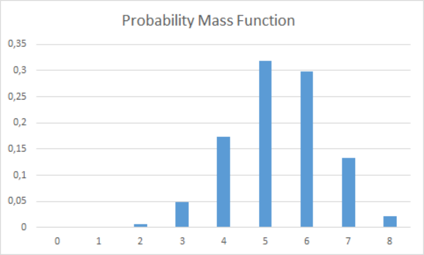
Remark: Table values are rounded at the second decimal digit.

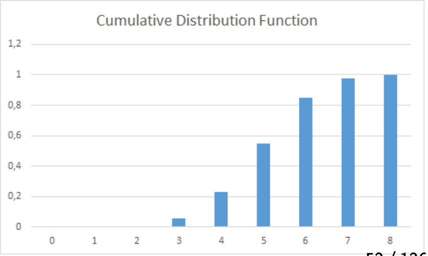
Discrete Random Variables

Hypergeometric Distribution – Example:

$$X \sim \text{Hyp}(30, 20, 8)$$

$$p(k) = P(X = k) = \frac{\binom{20}{k} \cdot \binom{10}{8-k}}{\binom{30}{8}}, \quad k = 0, \ldots, 8$$





Discrete Random Variables

Hypergeometric Distribution – Example:

The expectation value, the variance and the standard deviation for the discrete random variable

$$X \sim \mathrm{Hyp}(30, 20, 8)$$

are given by

$$\mu = m \cdot \frac{n_1}{n} = 8 \cdot \frac{20}{30} \approx 5.33$$

$$\sigma^2 = m \cdot \frac{n_1}{n} \cdot \left(1 - \frac{n_1}{n}\right) \cdot \frac{n-m}{n-1} = 8 \cdot \frac{20}{30} \cdot \left(1 - \frac{20}{30}\right) \cdot \frac{30-8}{30-1} \approx 1.34$$

$$\sigma = \sqrt{\sigma^2} \approx 1.16$$

Discrete Random Variables

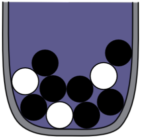
Exercise:

After the legal warranty period is over, 20 out of 30 airbag triggers are still functional.

What is the probability that from 8 randomly selected airbags at least 5 trigger?

Discrete Random Variables

Binomial Distribution:



- The binomial distribution is based on the urn model "sampling with replacement"
- In total, m repetitions **with** replacement are performed (in the special case of $m = 1$, it is also called Bernoulli distribution)
- The probability for selecting a black ball in one trial is $p \in [0, 1]$, the probability for a white ball is thus $1 - p$
- By the random variable X we denote the **total number of black balls** taken out of the urn in m trials
- In this case, X is distributed according to

$$\boxed{X \sim \text{Bin}(m, p)}$$

https://michael-franke.github.io/intro-data-analysis/Chap-03-01-probability-basics.html

Discrete Random Variables

Binomial Distribution – Properties:

A discrete random variable $X \sim \mathrm{Bin}(m, p)$ is characterized by the probability mass function

$$p(k) = P(X = k) = \binom{m}{k} \cdot p^k \cdot (1-p)^{m-k}, \quad k = 0, \ldots, m$$

In this case, the random variable X has the expectation value

$$\mu = E(X) = m \cdot p$$

and the variance

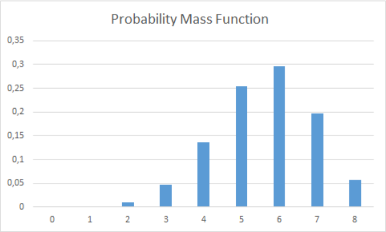
$$\sigma^2 = V(X) = m \cdot p \cdot (1-p)$$

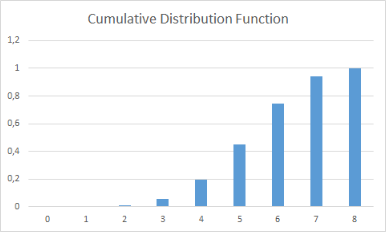
Discrete Random Variables

Binomial Distribution – Example:

$$X \sim \text{Bin}(8, \tfrac{2}{3})$$

$$p(k) = P(X = k) = \binom{8}{k} \cdot \left(\tfrac{2}{3}\right)^k \cdot \left(\tfrac{1}{3}\right)^{8-k}, \quad k = 0, \ldots, 8$$





Discrete Random Variables

Binomial Distribution – Example:

The expectation value, the variance and the standard deviation for the discrete random variable

$$X \sim \mathrm{Bin}(8, \tfrac{2}{3})$$

are given by

$$\mu = m \cdot p = 8 \cdot \tfrac{2}{3} \approx 5.33$$

$$\sigma^2 = m \cdot p \cdot (1 - p) = 8 \cdot \tfrac{2}{3} \cdot \tfrac{1}{3} \approx 1.77$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\tfrac{16}{9}} = \tfrac{4}{3} \approx 1.33$$

Discrete Random Variables

Exercise:

Donald misses the 8 p.m. news each day with a probability of 0.3. What is the probability that Donald misses the 8 p.m. news during one week **at most twice**?

https://www.freepik.com/free-vector/news-tv-illustration_5894635.htm

Discrete Random Variables

Under certain conditions, the binomial distribution can be used as an approximation of the hypergeometric distribution:

Approximation of the Hypergeometric Distribution by the Binomial Distribution:

Consider the hypergeometric distribution $\text{Hyp}(n, n_1, m)$.

For $m \ll n$, the hypergeometric distribution can be **approximated** by the binomial distribution according to

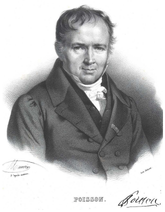
$$\text{Hyp}(n, n_1, m) \approx \text{Bin}(m, \frac{n_1}{n})$$

A possible **rule of thumb** for the condition $m \ll n$ could be

$$\frac{m}{n} \leq 0.1$$

Discrete Random Variables

Poisson Distribution:

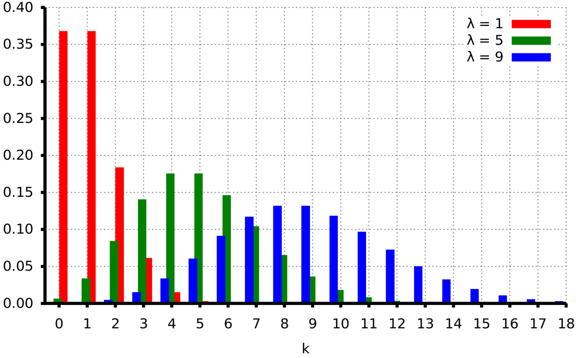


- The Poisson distribution is the limit of the binomial distribution when p gets very small and m gets very large
- In practical scenarios, the binomial distribution can be replaced by the Poisson distribution, when $m \geq 100$ and $\lambda = mp \leq 10$
- Let X be a discrete random variable which can take values in the natural numbers including zero $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$
- Then X is distributed according to a Poisson distribution with the parameter $\lambda > 0$ or in short $X \sim \text{Ps}(\lambda)$, if it holds that

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad k = 0, 1, 2, 3, \ldots$$

Discrete Random Variables

Poisson Distribution – Examples:



For a discrete random variable $X \sim \mathrm{Ps}(\lambda)$, it holds that

$$\mu = E(X) = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda$$

Discrete Random Variables

Exercise:

On average 5 % of the output of a production facility do not possess the required quality for further processing.

Determine the probability that out of a sample of 100 output items **exactly ten** do not meet the quality requirements

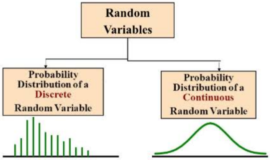
(a) with the help of the binomial distribution

(b) with the help of the Poisson distribution

2.3 Continuous Random Variables

Continuous Random Variables

Discrete vs. Continuous Random Variables:



So far, we have dealt with discrete random variables, where the possible realizations are a **collection of isolated points** along the real line $\mathbb{R}$.

For discrete random variables, all possible values could be **enumerated**: $x_1, x_2, x_3, \ldots$

In this section, we consider the situation that the realization of a random variable is a real number which can assume any value **in a whole interval** of the real line $\mathbb{R}$.

Mathematically, this corresponds to the concept of a continuous random variable.

https://prwatech.in/blog/tag/discrete-random-variable/

Continuous Random Variables

Continuous Random Variables – Introductory Example:

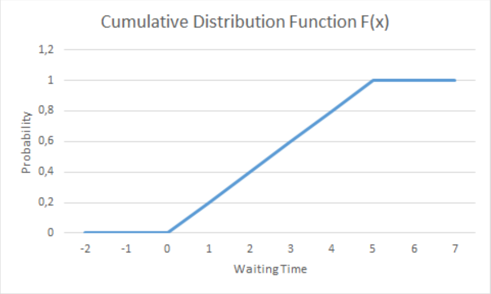
- You arrive at a streetcar stop **randomly** at some point in time
- Without knowing the timetable exactly, you know that a new streetcar arrives **every 5 minutes**
- Your **waiting time** can now be described as a continuous random variable X:
  - Probability of waiting **at most** 5 minutes: $P(X \leq 5) = 1$
  - Probability of waiting **at most** 2.5 minutes: $P(X \leq 2.5) = \frac{1}{2}$
  - Probability of waiting **at most** 4 minutes: $P(X \leq 4) = \frac{4}{5}$
- How to describe this in general?

Continuous Random Variables

Cumulative Distribution Function:

The probability $P(X \leq x)$ to wait **at most** x minutes is given by the cumulative distribution function (cdf) $F(x)$ as follows:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{5} & \text{if } 0 \leq x \leq 5 \\ 1 & \text{if } x > 5 \end{cases}$$



Continuous Random Variables

For continuous random variables, the probability density function takes the role of the probability mass function defined for discrete random variables:

Probability Density Function:

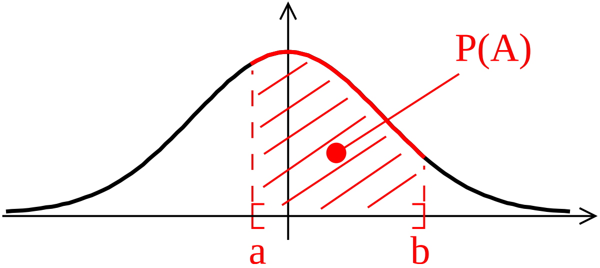
- The probability density function (pdf) $f(x)$ of a continuous random variables X takes on positive values on whole **intervals** of the real line, not only on isolated points
- Mathematically, **integration** now takes the role of summation (compute an area below a function instead of summing up probabilities for isolated points)
- A specific **probability** is now represented by the corresponding area below the probability density function $f(x)$

Continuous Random Variables

Probabilities for Continuous Random Variables:

For a **continuous random variable** X with the pdf $f(x)$ and the cdf $F(x)$, it holds for the interval $A = [a, b] \subset \mathbb{R}$:

$$P(A) = P(X \in A) = P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)$$



Continuous Random Variables

Relationship Between Cumulative Distribution Function and Probability Density Function:

1. The cdf $F(x)$ is now the integral of the pdf $f(x)$ up to the point $x \in \mathbb{R}$:

$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\, ds}$$

2. Vice versa, the pdf $f(x)$ is the derivative of the cdf $F(x)$ in the case that $F(x)$ is differentiable:

$$\boxed{f(x) = \frac{d}{dx} F(x)}$$

Continuous Random Variables

Probabilities for Continuous Random Variables:

Due to the properties of the integral, for a continuous random variable X with pdf $f(x)$ and cdf $F(x)$ it holds:

1. $P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)\,dx = F(a) - F(a) = 0$
2. $P(X < a) = P(X \leq a) - \underbrace{P(X = a)}_{=0} = P(X \leq a) = F(a)$
3. $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) =$

$$= P(a \leq X \leq b) = F(b) - F(a)$$

Continuous Random Variables

A probability density function (pdf) must fulfill certain properties:

Probability Density Function - Properties:

For the probability density function $f(x)$ of a continuous random variable X, the following two conditions must hold:

(i) $f(x) \geq 0$ for all $x \in \mathbb{R}$ (*non-negativity*)

(ii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (*normalization*)

Property (ii) corresponds to the fact that

$$P(X \in \mathbb{R}) = P(-\infty < X < \infty) = 1$$

Continuous Random Variables

Exercise:

Consider the following function:

$$f(x) = \begin{cases} \frac{x}{8} & \text{if } 3 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Check that $f(x)$ is the pdf of a continuous random variable X.

(b) Determine the cdf $F(x)$ and compute the probabilities $P(X \leq 4)$ and $P(4 < X < 5)$.

Continuous Random Variables

Continuous Random Variables – Remarks:

A continuous random variable X with the pdf $f(x)$ does not imply that the pdf $f(x)$ itself is a continuous function, so $f(x)$ is allowed to have **jumps**.

Since probabilities are now expressed as the area below the pdf $f(x)$ and not by $f(x)$ directly, the pdf $f(x)$ can take on values which are **larger than 1**.

For a given pdf $f(x)$, the corresponding cdf $F(x)$ is always a **continuous function** which is differentiable at every point where the pdf $f(x)$ is continuous. There it holds that

$$\frac{d}{dx}F(x) = f(x)$$

Continuous Random Variables

The **distribution parameters** of a continuous random variable X are obtained by integration:

Expectation Value, Variance and Standard Deviation:

For a continuous random variable X with the pdf $f(x)$, we define

- the expectation value or mean

$$\boxed{\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx}$$

- the variance

$$\boxed{\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx = \left(\int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx\right) - \mu^2}$$

Again, the standard deviation is given by $\sigma = \sqrt{V(X)}$.

Continuous Random Variables

Exercise:

Consider the following function:

$$f(x) = \begin{cases} 6(x - x^2) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Check that $f(x)$ is the pdf of a continuous random variable X and compute the expectation value and the variance of X.

Continuous Random Variables

As for discrete random variables, a continuous random variable is frequently specified by a certain probability distribution:

Continuous Probability Distributions:

Probability distributions for continuous random variables can be characterized by specifying

1. the probability density function (pdf)
2. the cumulative distribution function (cdf)

The most important **continuous probability distributions** are

- the uniform distribution
- the exponential distribution
- the normal distribution

Continuous Random Variables

Uniform Distribution:

- The uniform distribution is the **most simple type** of a continuous probability distribution
- For a uniform distribution, one specifies
  - the **smallest possible value** $a \in \mathbb{R}$
  - the **largest possible value** $b \in \mathbb{R}$
- On the obtained interval $[a, b] \subset \mathbb{R}$, the pdf $f(x)$ is assumed to be **constant**
- For a uniformly distributed random variable X on the interval $[a, b] \subset \mathbb{R}$, we write shortly

$$\boxed{X \sim \mathrm{U}(a, b)}$$

Continuous Random Variables

The uniform distribution can be described by the pdf or the cdf:

Uniform Distribution:

The pdf $f(x)$ and the cdf $F(x)$ of the uniform distribution $\mathrm{U}(a, b)$ are given by

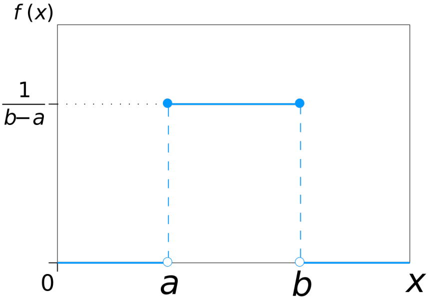
$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{if } x > b \end{cases} \qquad F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

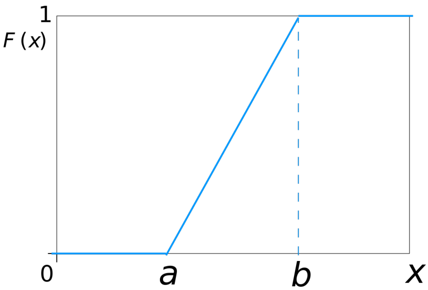
The expectation value and the variance of $\mathrm{U}(a, b)$ are obtained as

$$\mu = E(X) = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

Continuous Random Variables

Uniform Distribution – Graphical Illustration:





Continuous Random Variables

Exponential Distribution:

- The exponential distribution is often applied for random variables measuring the length of a **time span**:
  - time between the random arrivals of trucks at a gate (interarrival time)
  - duration of telephone conversations
  - life span of wearless equipment (e.g. light bulbs)
- For an exponentially distributed random variable X, the **conditional distribution** of the remaining time span is independent of the already reached or consumed time:

$$P(X \leq t + s \,|\, X \geq t) = P(X \leq s)$$

- Therefore, the exponential distribution is also called **"distribution without memory"**

Continuous Random Variables

Exponential Distribution:

The pdf $f(x)$ and the cdf $F(x)$ of the exponential distribution with the parameter $\lambda > 0$ are given by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases} \qquad F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

The expectation value and the variance are obtained as

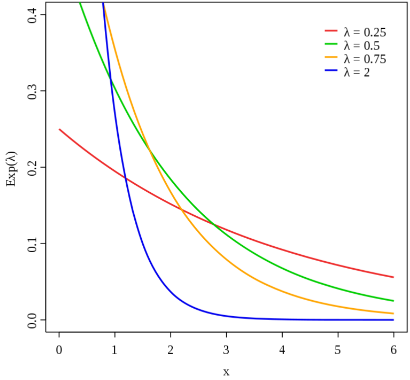
$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

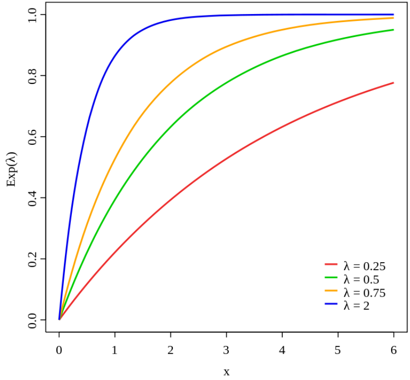
For an exponentially distributed random variable X with parameter $\lambda > 0$, we write shortly

$$\boxed{X \sim \mathrm{Exp}(\lambda)}$$

Continuous Random Variables

Exponential Distribution – Graphical Illustration:

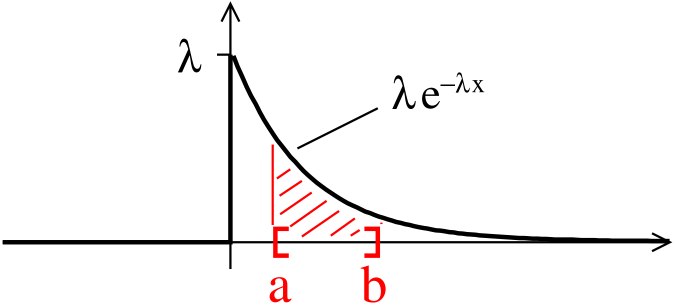




Continuous Random Variables

Exponential Distribution – Computing Probabilities:

The pdf $f(x)$ of the exponential distribution takes the form



For the probability $P(a \leq X \leq b)$, it holds that

$$P(a \leq X \leq b) = \int_a^b \lambda e^{-\lambda x}\, dx = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}$$

Continuous Random Variables

Exercise:

The time span between the arrivals of trucks at a gate (so-called interarrival time) in minutes is assumed to be an exponentially distributed random variable $X \sim \text{Exp}(\lambda)$.

After t minutes have passed since the last arrival, the probability that the next truck will arrive within the next minute should be $\frac{1}{2}$.

(a) What is the value of the parameter λ?

(b) What is the probability that the interarrival time of the next truck is larger than 1 minute but smaller than 3 minutes?

Continuous Random Variables

Normal Distribution:

- The normal distribution or Gaussian distribution is the **most important** continuous probability distribution
- A normally distributed random variable X can take values on the whole real line $\mathbb{R}$ and is characterized by the **mean** $\mu \in \mathbb{R}$ and the **variance** $\sigma^2 > 0$
- The corresponding pdf $f(x)$ is also called Gaussian bell curve or normal curve
- The normal distribution plays also a central role in **inductive** or **inferential statistics**

Von Deutsche Bundesbank, Frankfurt am Main, Germany - http://www.bundesbank.de/Redaktion/DE/Standardartikel/Kerngeschaeftsfelder/Bargeld/dm_banknoten.html #doc18118bodyText2, Gemeinfrei, https://commons.wikimedia.org/w/index.php?curid=3813487

Continuous Random Variables

Normal Distribution:

The pdf $f(x)$ of the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ is given by

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

Accordingly, the cdf $F(x)$ is obtained as

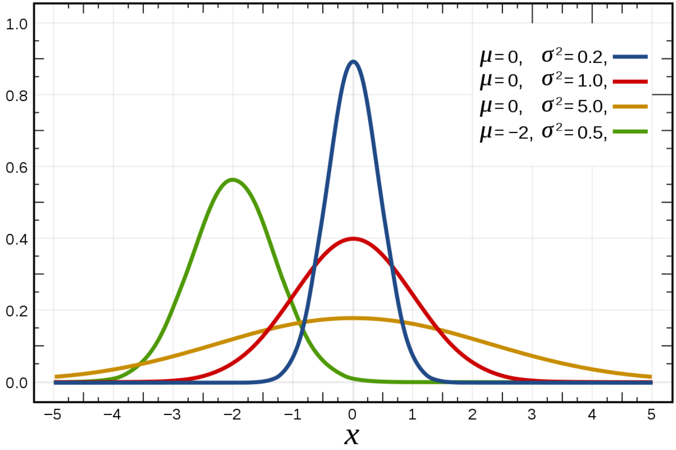
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\, ds = \underbrace{\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s-\mu)^2}{2\sigma^2}}\, ds}_{\text{not directly computable}}$$

For a normally distributed random variable X with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, we write shortly

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

Continuous Random Variables

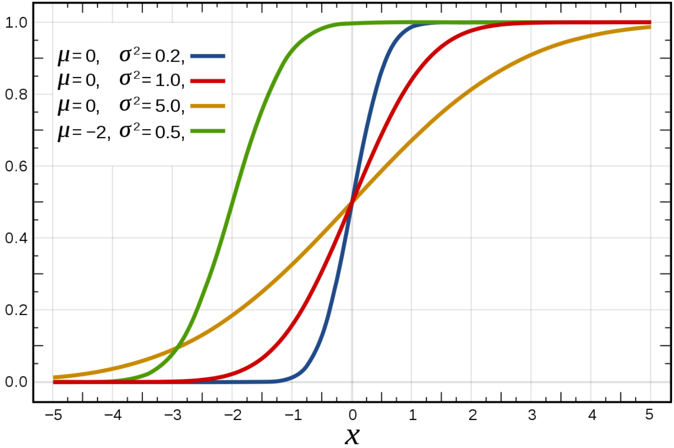
Normal Distribution – Different Pdfs:



By Inductiveload - self-made, Mathematica, Inkscape, Public Domain,
https://commons.wikimedia.org/w/index.php?curid=3817954

Continuous Random Variables

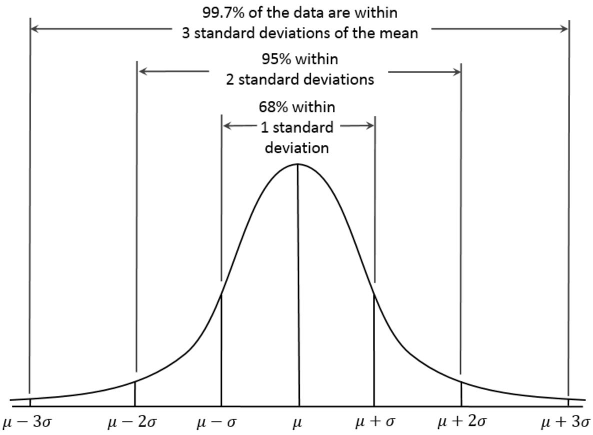
Normal Distribution – Different Cdfs:



By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817960

Continuous Random Variables

Normal Distribution – Mean and Standard Deviation:



Continuous Random Variables

Normal Distribution – Properties:

- It holds that $f(x) > 0$ for all $x \in \mathbb{R}$
- The maximum of $f(x)$ is attained at $x = \mu$
- $f(x)$ has inflection points at $x = \mu - \sigma$ and $x = \mu + \sigma$
- $f(x)$ is symmetric around the mean μ, i.e.

$$f(\mu - x) = f(\mu + x) \quad \text{and} \quad F(\mu - x) = 1 - F(\mu + x)$$

Attention: The pdf $f(x)$ of the normal distribution <u>does not</u> have an elementary antiderivative, so the computation of specific values has to be done by means of **calculators** or **tables** which make use of the standard normal distribution.

Continuous Random Variables

Standard Normal Distribution:

Since the cdf $F(x)$ of $X \sim \mathcal{N}(\mu, \sigma^2)$ cannot be computed directly, values are transformed to the standard normal distribution $\mathcal{N}(0,1)$ which is supported by **table values**.

For the standard normal distribution, it holds that

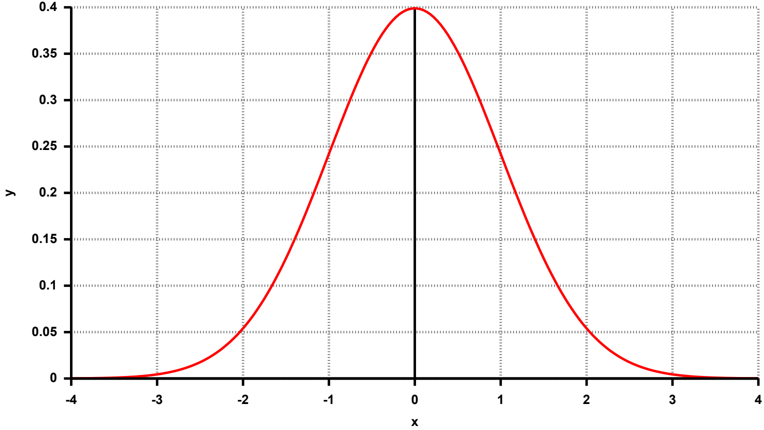
$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{and} \quad F_0(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \, ds$$

The **transformation rule** for a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ to the standard normal distribution $\mathcal{N}(0,1)$ is according to

$$F(x) = P(X \leq x) = F_0\left(\frac{x-\mu}{\sigma}\right) = F_0(z) \quad \text{where} \quad z = \frac{x-\mu}{\sigma}$$

Continuous Random Variables

Standard Normal Distribution – Graphical Illustration:



$$\boxed{f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}$$

Continuous Random Variables

Standard Normal Distribution – Table Values:

Table Standard Normal Distribution

z \ *	0	1	2	3	4	5	6	7	8	9
0.0*	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1*	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2*	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3*	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4*	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5*	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6*	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7*	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8*	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9*	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0*	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1*	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2*	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3*	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4*	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5*	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408

Continuous Random Variables

Exercise:

Past experience has shown that the time needed for the successful deployment of an SAP ERP system is normally distributed with the mean $\mu = 39$ (weeks) and the standard deviation $\sigma = 2$ (weeks).

What is the probability that the time for the deployment of a system lies between 37 and 41 weeks?

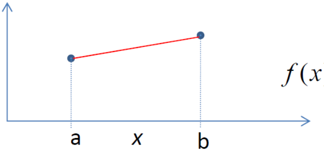
Continuous Random Variables

Standard Normal Distribution – Remarks on Using the Table:

Because of the **symmetry** of the pdf $f_0(x)$, only half of the values are required in the table ($z \geq 0$).

The table can also answer questions like "How big do we have to choose the window around $\mu = 39$ when we want to have a probability of at least 0.9?"

In the case that a requested value is located **between** two table values, a linear interpolation may be applied (connect the two table values with a straight line):

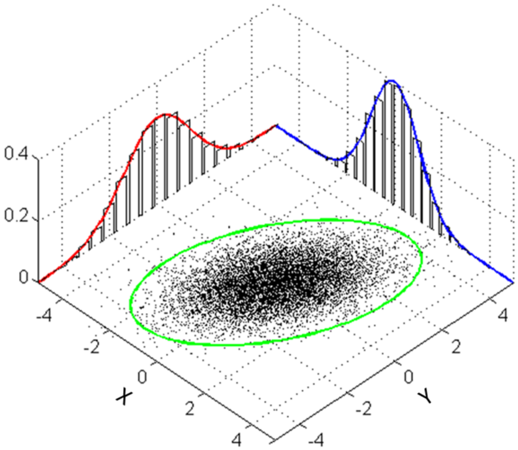


$$f(x) = \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

2.4 Multivariate Distributions

Multivariate Distributions

The joint realization of **two or more random variables** is described by multivariate distributions:



Multivariate Distributions

Two or more simultaneously observed random variables are combined to a random vector:

Random Vectors:

For real-valued random variables $X_1, \dots, X_D$ defined on the same sample space Ω, the mapping

$$\boldsymbol{X} : \Omega \to \mathbb{R}^D, \quad \omega \mapsto \boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_D(\omega))^T$$

defines a *D*-dimensional random variable or random vector, interpreted as a **column vector**.

Multidimensional random variables or random vectors are also called multivariate random variables.

The **joint probability distribution** of the random vector $\boldsymbol{X}$ will be denoted by $P_{\boldsymbol{X}}$, and we write $\boldsymbol{X} \sim P_{\boldsymbol{X}}$.

Multivariate Distributions

Random Vectors – Example:

Let us consider the joint measurement of the **height** (in cm) and the **weight** (in kg) of a randomly selected student of FHWS, in this way obtaining a bivariate data sample.

The obtained values for height and weight of student ω_j would then correspond to the realization of a random vector $\boldsymbol{X} = (X_1, X_2)^T$ defined on the set Ω of all students of FHWS:

$$\boldsymbol{X} : \underbrace{\Omega}_{\text{all students}} \to \mathbb{R}^2, \quad \omega_j \mapsto \boldsymbol{X}(\omega_j) = (\ \underbrace{X_1(\omega_j)}_{\text{height of student } \omega_j}\ ,\ \overbrace{X_2(\omega_j)}^{\text{weight of student } \omega_j}\ )$$

How to determine the **joint distribution** of the random vector $\boldsymbol{X}$?

Multivariate Distributions

Joint Distribution of Random Vectors:

Let $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ be a D-dimensional random vector.

The mapping $F_{\boldsymbol{X}} : \mathbb{R}^D \to \mathbb{R}$ with the property

$$F_{\boldsymbol{X}}(x_1, \ldots, x_D) = P(\{\omega \in \Omega \,|\, X_1(\omega) \le x_1, \ldots, X_D(\omega) \le x_D\})$$
$$= P(X_1 \le x_1, \ldots, X_D \le x_D)$$

is called joint cumulative distribution function (joint cdf) of the random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$.

The joint probability distribution $P_{\boldsymbol{X}}$ of the random vector $\boldsymbol{X}$ is **determined uniquely** by the joint cdf $F_{\boldsymbol{X}}$, we therefore also write

$$\boldsymbol{X} \sim F_{\boldsymbol{X}}$$

Multivariate Distributions

Joint Probability Density Function:

An integrable function $f_{\boldsymbol{X}} : \mathbb{R}^D \to \mathbb{R}_+$ is called joint probability density function (joint pdf) of the random vector $\boldsymbol{X}$, if it holds that

$$F_{\boldsymbol{X}}(x_1, \ldots, x_D) = \int_{-\infty}^{x_D} \cdots \int_{-\infty}^{x_1} f_{\boldsymbol{X}}(s_1, \ldots, s_D)\, ds_1 \ldots ds_D$$

for all $x_1, \ldots, x_D \in \mathbb{R}$.

In this case, the random vector $\boldsymbol{X}$ is called continuous with the joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, we then also write

$$\boldsymbol{X} \sim f_{\boldsymbol{X}}$$

Frequently, we will determine the probability distribution of a random vector $\boldsymbol{X}$ by specifying its joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$.

Multivariate Distributions

The joint pdf can be used to determine the probability of events:

Joint Probability Density Function and Probability of Events:

For a continuous random vector $\boldsymbol{X}$ with the joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, the probability $P_{\boldsymbol{X}}(B)$ of a measurable event $B \subseteq \mathbb{R}^D$ can be computed via **integration**:

$$\boxed{P_{\boldsymbol{X}}(B) = P(\boldsymbol{X} \in B) = \int_B f_{\boldsymbol{X}}(x_1, \ldots, x_D)\, dx_1 \ldots dx_D}$$

Here, a multiple integral has to be evaluated, since the probability of the event $B \subseteq \mathbb{R}^D$ is given by a definite integral of the joint pdf

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{X}}(x_1, \ldots, x_D)$$

which is a real-valued function of D real variables.

Multivariate Distributions

Properties of the Joint Cdf and Pdf:

1. The joint cdf $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{X}}(x_1, \ldots, x_D)$ is monotonically increasing in all variables x_i, $i = 1, \ldots, D$.

2. It holds that
$$\lim_{x_i \to -\infty} F_{\boldsymbol{X}}(x_1, \ldots, x_D) = 0$$
i.e. $F_{\boldsymbol{X}}(\boldsymbol{x})$ approaches 0, whenever one x_i goes to $-\infty$.

3. It holds that
$$\lim_{x_1, \ldots, x_D \to +\infty} F_{\boldsymbol{X}}(x_1, \ldots, x_D) = 1$$
i.e. $F_{\boldsymbol{X}}(\boldsymbol{x})$ approaches 1, if all x_i go to $+\infty$.

4. It holds that
$$f_{\boldsymbol{X}}(x_1, \ldots, x_D) = \frac{\partial^D F_{\boldsymbol{X}}(x_1, \ldots, x_D)}{\partial x_1 \ldots \partial x_D}$$

Multivariate Distributions

Marginal Cdf and Pdf:

For a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ with the joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, the marginal pdf of the component X_i is obtained via **integration** of $f_{\boldsymbol{X}}(\boldsymbol{x})$ with respect to all other variables:

$$f_{X_i}(x_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\boldsymbol{X}}(x_1, \ldots, x_D)\, dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_D$$

The corresponding marginal cdf $F_{X_i}(x_i)$ of the component X_i is obtained via integration of the marginal pdf $f_{X_i}(x_i)$:

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} f_{X_i}(s_i)\, ds_i$$

Based on the relation between the marginal cdfs/pdfs and the joint cdf/pdf of the random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$, the question of statistical independence of the components X_i can be addressed.

Multivariate Distributions

Statistical independence of random variables is one of the key concepts in the modeling of stochastic phenomena:

Statistical Independence of Random Variables:

Let $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ be a D-dimensional random vector with the joint cdf $F_{\boldsymbol{X}}(\boldsymbol{x})$. The components $X_1, \ldots, X_D$ of $\boldsymbol{X}$ are called statistically independent, if it holds that

$$F_{\boldsymbol{X}}(x_1, \ldots, x_D) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_D}(x_D)$$

for all $x_1, \ldots, x_D \in \mathbb{R}$.

In the case that the random variables $X_1, \ldots, X_D$ are statistically independent and all have the same marginal distribution, they are called independent and identically distributed (i.i.d.).

Multivariate Distributions

For continuous random vectors, the characterization of statistically independent random variables in terms of the **joint** and **marginal probability density functions** is of primary importance:

Statistical Independence of Random Variables:

For a continuous random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ with the joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, it holds:

1. $f_{\boldsymbol{X}}(x_1, \ldots, x_D) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_D}(x_D)$ for all $x_1, \ldots, x_D \in \mathbb{R}$

   $\implies X_1, \ldots, X_D$ are statistically independent

2. $X_1, \ldots, X_D$ are statistically independent

   $$\implies \prod_{i=1}^{D} f_{X_i}(x_i) \text{ is a joint pdf of } \boldsymbol{X}$$

Multivariate Distributions

Exercise:

The joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the bivariate random vector $\boldsymbol{X} = (X_1, X_2)^T$ is given by

$$f_{\boldsymbol{X}}(x_1, x_2) = \begin{cases} 3x_1^2 + 3x_2^2 & \text{for } 0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq x_1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Check that $f_{\boldsymbol{X}}(\boldsymbol{x})$ is actually a pdf.

(b) Compute the probabilities $P\big(\boldsymbol{X} \in [0, \frac{1}{2}] \times [\frac{3}{4}, \frac{4}{5}]\big)$ and $P\big(\boldsymbol{X} \in [\frac{1}{2}, 1] \times [0, \frac{1}{2}]\big)$.

(c) Determine the marginal pdfs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$.

(d) Are the random variables X_1 and X_2 statistically independent?

Multivariate Distributions

Mean of a Random Vector:

For a random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$, the expression

$$\boldsymbol{\mu} = E(\boldsymbol{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_D) \end{pmatrix} \in \mathbb{R}^D$$

is called mean vector or mean of $\boldsymbol{X}$. If $\boldsymbol{X}$ is a continuous random vector with joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, the components of the vector $\boldsymbol{\mu} \in \mathbb{R}^D$ can be computed as

$$\boxed{\mu_i = E(X_i) = \int_{\mathbb{R}^D} x_i \cdot f_{\boldsymbol{X}}(x_1, \ldots, x_D)\, dx_1 \ldots dx_D}$$

The mean vector $\boldsymbol{\mu} \in \mathbb{R}^D$ indicates the **location** of the distribution of $\boldsymbol{X}$ and is therefore a measure of central tendency .

Multivariate Distributions

Covariance Matrix of a Random Vector:

For a random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$, the **symmetric matrix** of the pairwise covariances of X_i and X_j according to

$$\boldsymbol{\Sigma} = Cov(\boldsymbol{X}) = (\sigma_{ij})_{i,j} = (Cov(X_i, X_j))_{i,j} \in \mathbb{R}^{D \times D}$$

is called the covariance matrix of $\boldsymbol{X}$. If $\boldsymbol{X}$ is a continuous random vector with joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$, the components σ_{ij} of the symmetric matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$ can be computed as

$$\boxed{\sigma_{ij} = Cov(X_i, X_j) = \int_{\mathbb{R}^D} (x_i - \mu_i)(x_j - \mu_j) f_{\boldsymbol{X}}(x_1, \ldots, x_D)\, dx_1 \ldots dx_D}$$

The covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$ captures both the dispersion of the individual components of $\boldsymbol{X}$ (on the main diagonal) as the statistical relationship between the components (on the secondary diagonals) and is thus a measure of dispersion and correlation.

Multivariate Distributions

Properties of the Covariance Matrix:

The covariance matrix $\boldsymbol{\Sigma}$ of a random vector $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ is always positive semidefinite, i.e.

$$\forall \boldsymbol{a} \in \mathbb{R}^D : \quad \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a} \geq 0$$

It holds further that

$$\begin{aligned}
\boldsymbol{\Sigma} &= E((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T) \\
&= E(\boldsymbol{X}\boldsymbol{X}^T) - E(\boldsymbol{X})\boldsymbol{\mu}^T - \boldsymbol{\mu}E(\boldsymbol{X}^T) + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= E(\boldsymbol{X}\boldsymbol{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}$$

The expression $E(\boldsymbol{X}\boldsymbol{X}^T)$ is also called auto-correlation matrix.

Multivariate Distributions

Frequently, it is important to know the mean and covariance matrix of a random vector obtained by **linear transformation**:

Linear Transformation of Random Vectors:

Let $\boldsymbol{X} = (X_1, \ldots, X_D)^T$ be a random vector with mean $\boldsymbol{\mu} \in \mathbb{R}^D$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$. Furthermore, let $\boldsymbol{A} \in \mathbb{R}^{C \times D}$ be an arbitrary real matrix and $\boldsymbol{b} \in \mathbb{R}^C$ be an arbitrary real vector.

Then it holds for the linearly transformed random vector

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$$

that

1. $E(\boldsymbol{Y}) = E(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b} \in \mathbb{R}^C$
2. $Cov(\boldsymbol{Y}) = Cov(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T \in \mathbb{R}^{C \times C}$

Multivariate Distributions

Exercise:

The joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the bivariate random vector $\boldsymbol{X} = (X_1, X_2)^T$ is given by

$$f_{\boldsymbol{X}}(x_1, x_2) = \begin{cases} 3x_1^2 + 3x_2^2 & \text{for } 0 \leq x_1 \leq 1,\ 0 \leq x_2 \leq x_1 \\ 0 & \text{otherwise} \end{cases}$$

Determine the mean vector $\boldsymbol{\mu} = E(\boldsymbol{X}) \in \mathbb{R}^2$ and the covariance matrix $\boldsymbol{\Sigma} = Cov(\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$ of the bivariate random vector $\boldsymbol{X}$.

Multivariate Distributions

The Multivariate Normal Distribution:

- The multivariate normal distribution or multivariate Gaussian distribution is a generalization of the one-dimensional or univariate normal distribution to higher dimensions
- A possible definition is that a D-dimensional random vector $\boldsymbol{X}$ is said to be normally distributed, if **every linear combination** of its D components has a univariate normal distribution
- The multivariate normal distribution is often used as an **approximation** to describe any set of correlated real-valued random variables, each of which clusters around a mean value
- In particular, the multivariate normal distribution plays a key role in the derivation of the classical Kalman filter (Chapter 6)

Multivariate Distributions

The Multivariate Normal Distribution:

Let $\boldsymbol{\mu} \in \mathbb{R}^D$ be a real vector and let $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$ be a symmetric and positive definite real matrix.

The **joint pdf** of the *D*-dimensional normal distribution is given by

$$\boxed{f_{\boldsymbol{X}}(x_1, \ldots, x_D) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right)}$$

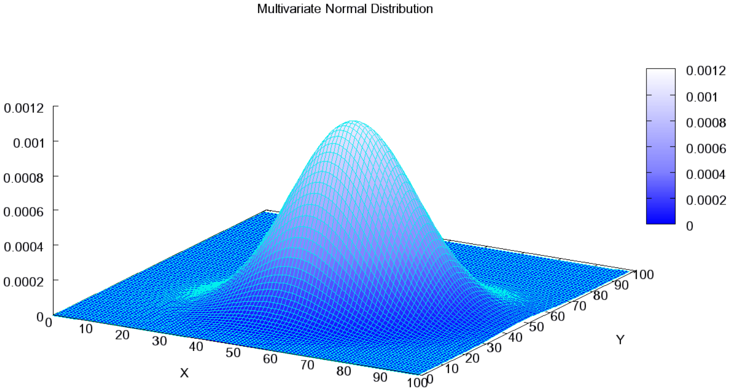
where $|\boldsymbol{\Sigma}| = \det(\boldsymbol{\Sigma})$. For the joint pdf $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the multivariate normal distribution, it holds that

$$f_{\boldsymbol{X}}(\boldsymbol{x}) > 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^D$$

We write shortly $\boldsymbol{X} \sim \mathcal{N}_D(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For $D = 1$, one gets the known univariate normal distribution $\mathcal{N}(\mu, \sigma^2)$.

Multivariate Distributions

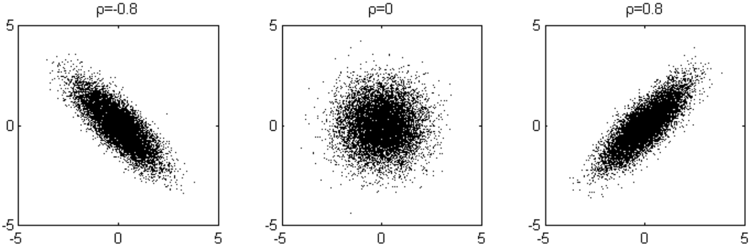
Example: Joint pdf of a bivariate normal distribution



Multivariate Distributions

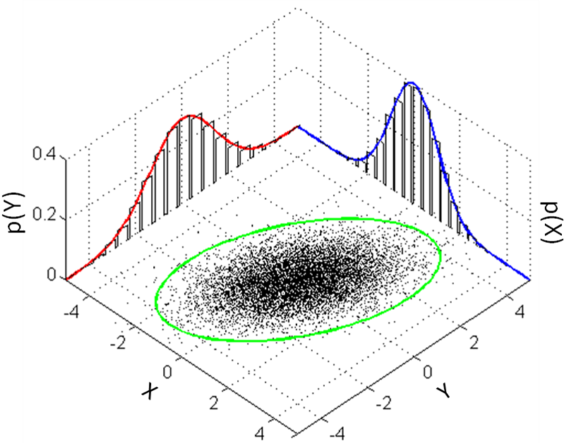
Example: Joint pdf of a bivariate normal distribution with $\mu_1 = \mu_2 = 0$, $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{12} = \sigma_{21} = \varrho$

$$f_{\boldsymbol{X}}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}(x_1^2 - 2\varrho x_1 x_2 + x_2^2) \right)$$



Multivariate Distributions

Example: Bivariate normal distribution and its marginal pdfs with $\mu_1 = \mu_2 = 0$, $\sigma_{11} = 1$, $\sigma_{22} = 2$ and $\sigma_{12} = \sigma_{21} = \frac{3}{5}$



By Bscan - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=25235145

Multivariate Distributions

Properties of the Multivariate Normal Distribution:

1. $\boldsymbol{X} \sim \mathcal{N}_D(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \implies X_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$ for $i = 1, \ldots, D$

2. $\boldsymbol{X} \sim \mathcal{N}_D(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{X} = \begin{pmatrix} \tilde{\boldsymbol{X}}_1 \\ \tilde{\boldsymbol{X}}_2 \end{pmatrix}$ (partitioning of $\boldsymbol{X}$)

   $\implies \tilde{\boldsymbol{X}}_1 \sim \mathcal{N}_{D_1}(\tilde{\boldsymbol{\mu}}_1, \boldsymbol{\Sigma}_{11}), \quad \tilde{\boldsymbol{X}}_2 \sim \mathcal{N}_{D_2}(\tilde{\boldsymbol{\mu}}_2, \boldsymbol{\Sigma}_{22})$

3. $\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2$ statistically independent $\iff \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$

   (i.e. $\tilde{\boldsymbol{X}}_1, \tilde{\boldsymbol{X}}_2$ uncorrelated)

4. $\boldsymbol{X} \sim \mathcal{N}_D(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{A} \in \mathbb{R}^{C \times D}, \quad \boldsymbol{b} \in \mathbb{R}^C$

   $\implies \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \sim \mathcal{N}_C(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$

Multivariate Distributions

Exercise:

Consider the normally distributed random vector

$$\boldsymbol{X} = (X_1, X_2)^T \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & \frac{1}{2} \\ \frac{1}{2} & 3 \end{pmatrix}$$

Determine the pdf of $Y = X_1 + X_2$.

Multivariate Distributions

In some situations, the observed realizations of a random vector are discrete, e.g. if the measured values are **quantized**:

Binary Random Vectors:

An important special case are binary random vectors, i.e. one considers a random vector

$$\boldsymbol{X} : \Omega \to \{0,1\}^D, \quad \omega \mapsto \boldsymbol{X}(\omega) = (X_1(\omega), \dots, X_D(\omega))^T$$

where the components $X_1, \dots, X_D$ are binary random variables, i.e.

$$X_i : \Omega \to \{0,1\}, \quad \omega \mapsto X_i(\omega), \quad i = 1, \dots, D$$

In this case, there are in total 2^D possible realizations of $\boldsymbol{X}$.

For the treatment of binary random vectors, it is important to know whether the components $X_1, \dots, X_D$ are **independent** or not.

Multivariate Distributions

Independent Binary Components:

In the special case of independent binary components $X_1, \ldots, X_D$, the **joint probability mass function** (joint pmf) $f_{\boldsymbol{X}}(\boldsymbol{x})$ of the binary random vector $\boldsymbol{X}$ takes the form

$$f_{\boldsymbol{X}}(x_1, \ldots, x_D) = \prod_{i=1}^{D} f_{X_i}(x_i)$$

where each $f_{X_i}(x_i)$ follows a **Bernoulli distribution** according to

$$f_{X_i}(x_i) = p_i^{x_i} \cdot (1 - p_i)^{1-x_i}, \quad x_i \in \{0, 1\}$$

Therefore, for the joint pmf of $\boldsymbol{X}$ it follows that

$$f_{\boldsymbol{X}}(x_1, \ldots, x_D) = \prod_{i=1}^{D} p_i^{x_i} \cdot (1 - p_i)^{1-x_i} = \prod_{i=1}^{D} \left(\frac{p_i}{1 - p_i}\right)^{x_i} \cdot (1 - p_i)$$

2.5 Conditional Distributions

Conditional Distributions

Sometimes it is important to know the conditional distribution of a random vector $\boldsymbol{X}$ given the realization of another random vector $\boldsymbol{Y}$:

Conditional Density and Conditional Distribution:

Let $(\boldsymbol{X}, \boldsymbol{Y}) \in \mathbb{R}^{D+C}$ be a continuous random vector with the joint pdf $f_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{x}, \boldsymbol{y})$ and the marignal pdfs $f_{\boldsymbol{X}}(\boldsymbol{x})$ and $f_{\boldsymbol{Y}}(\boldsymbol{y})$.

Then the conditional density of $\boldsymbol{X}$ under $\boldsymbol{Y} = \boldsymbol{y} \in \mathbb{R}^C$ is given by

$$f_{\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}}(\boldsymbol{x}) = f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) = \begin{cases} \frac{f_{(\boldsymbol{X},\boldsymbol{Y})}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})} & \text{if } f_{\boldsymbol{Y}}(\boldsymbol{y}) > 0 \\ f_{\boldsymbol{X}}(\boldsymbol{x}) & \text{if } f_{\boldsymbol{Y}}(\boldsymbol{y}) = 0 \end{cases}$$

The corresponding distribution is the conditional distribution of $\boldsymbol{X}$ under $\boldsymbol{Y} = \boldsymbol{y}$, in short $P_{\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}}$. For events $B \subseteq \mathbb{R}^D$, it holds that

$$P_{\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}}(B) = P(\boldsymbol{X} \in B\,|\,\boldsymbol{Y} = \boldsymbol{y}) = \int_B f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y})\, dx_1 \ldots dx_D$$

Conditional Distributions

Conditional Density and Conditional Distribution – Remarks:

With the conditional density $f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$, the marginal pdf of the random vector $\boldsymbol{X}$ can be obtained as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\mathbb{R}^C} f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \cdot f_{\boldsymbol{Y}}(\boldsymbol{y})\, dy_1 \dots dy_C$$

Furthermore, for a mapping

$$g : \mathbb{R}^D \to \mathbb{R}, \quad \boldsymbol{x} \mapsto g(\boldsymbol{x})$$

the conditional expectation of $g(\boldsymbol{X})$ under $\boldsymbol{Y} = \boldsymbol{y} \in \mathbb{R}^C$ can be expressed as

$$E\big(g(\boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}\big) = \int_{\mathbb{R}^D} g(\boldsymbol{x}) \cdot f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\, dx_1 \dots dx_D$$

Conditional Distributions

Exercise:

Consider a uniformly distributed random variable $Y \sim \mathrm{U}(0,1)$.

The conditional distribution of the random variable X under $Y = y$ is assumed to be characterized by the conditional density

$$f_{X|Y}(x\,|\,y) = \frac{1}{2y} \cdot \mathbb{I}_{[-y,y]}(x)$$

i.e. it should hold that $X\,|\,Y = y \sim \mathrm{U}(-y, y)$.

Determine the marginal density $f_X(x)$ of the random variable X.