1. Basic probability concepts

1.1 Definitions & properties

Def: An experiment or trial is a process that terminates with an unpredictable result that will be called its outcome.

E.g.: 1) throwing a coin once
2) throwing a dice once
3) measuring the height of a human

To study an experiment we need to define its possible outcomes.

Def: The set of all possible outcomes of an experiment is called the sample set and is denoted by Ω.

E.g. 1) Coin: $\Omega = \{\text{Head}, \text{Tail}\}$
2) Dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$
3) Height of human: $\Omega = \{x \mid 0 < x \leq 3 \text{ meters}\}$

→ In examples 1) & 2) the sample set is finite while in example 3) it is an intervall of infinitely many outcomes.

Def: An event A is a subset of the sample set $A \subseteq \Omega$.
i) The event $A = \Omega$ is called the certain event.
ii) An event that only contains one element is called elementary event
iii) The empty set is an event and is called the impossible event.

Ex.: Dice: • A = "the outcome is an even number"

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

• B = "the outcome is 2" $= \{2\}$

B is an elementary event

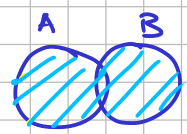
• Ω = "the outcome is a number from 1 to 6"

$= \{1, 2, \dots, 6\}$

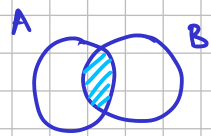
Ω is the certain event

Operations on sets:

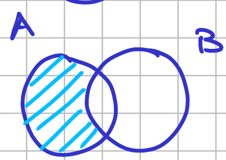
Union: $A \cup B = \{w \mid w \in A \text{ or } w \in B\}$



Intersection: $A \cap B = \{w \mid w \in A \text{ and } w \in B\}$



Difference: $A \setminus B = \{w \mid w \in A \text{ and } w \notin B\}$



Cardinality: $|A| = \#$ of elements in A

<u>Def</u>: A class of events (or σ-field) $\mathcal{F}$ is a set of events that is closed under the following operations:

i) if $A, B \in \mathcal{F}$ are two events, also their union is an event with $A \cup B \in \mathcal{F}$.

ii) given a countable sequence of events $A_i \in \mathcal{F}$ with $i = 1, 2, \dots$, also their union is an event $A_1 \cup A_2 \cup \dots \in \mathcal{F}$

iii) If $A \in \mathcal{F}$, also the complement of A is in $\mathcal{F}$, i.e.

$A^c = \Omega \setminus A \in \mathcal{F}$.

It follows that

1) Given two events $A, B \in \mathcal{F}$ also $A \cap B \in \mathcal{F}$

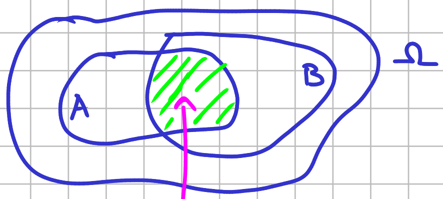
2) The empty set $\emptyset$ and the sample set Ω are events in $\mathcal{F}$

<u>Proof:</u> 1) For $A, B \in \mathcal{F} \underset{iii)}{\Rightarrow} A^c, B^c \in \mathcal{F} \underset{i)}{\Rightarrow} A^c \cup B^c \in \mathcal{F}$

$$\underset{iii)}{\Rightarrow} (A^c \cup B^c)^c \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$$

$$(A^c \cup B^c)^c = A \cap B$$

because



$(A^c \cup B^c)^c = A \cap B$

2) $\emptyset = A \cap A^c \in \mathcal{F}$ per $A \in \mathcal{F}$

$\Omega = A \cup A^c \in \mathcal{F}$ per $A \in \mathcal{F}$

Examples: 1) The smallest class of events is $\mathcal{F} = \{\emptyset, \Omega\}$.

2) For $A \subseteq \Omega$, $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ is a class of events.

Def: Two events $A, B \in \mathcal{F}$ are <u>incompatible</u> if they are disjoint, i.e. $A \cap B = \emptyset$.

→ They can not occur simultaneously.

E.g.: Dice $\Omega = \{1, 2, 3, 4, 5, 6\}$

$A =$ "outcome is even" $= \{2, 4, 6\}$

$B =$ "outcome is odd" $= \{1, 3, 5\}$

$\rightarrow$ A and B are incompatible $A \cap B = \emptyset$

Further, $\mathcal{F} = \{\emptyset, A, B, \Omega\}$ is a class of events because $A^c = \Omega \setminus A = B$.

Exercise: Convince yourself that the following identities are true

a) $A \cup (B \cup C) = (A \cup B) \cup C$

b) $A \cap (B \cap C) = (A \cap B) \cap C$

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

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

$$\left.\begin{aligned} &\text{e) } (A \cup B)^c = A^c \cap B^c \\ &\text{f) } (A \cap B)^c = A^c \cup B^c \end{aligned}\right\} \text{ de Morgan's law}$$

Every event in a class of events we assign a probability to occur.

Def: (probability measure)

Let Ω be the sample space and $\mathcal{F}$ a class of events with $\Omega \in \mathcal{F}$. Further let P be a function definend on $\mathcal{F}$ with values in the intervall $[0,1]$ such that:

$$P : \Omega \rightarrow [0,1].$$

The triple $(\Omega, \mathcal{F}, P)$ is called probability space and $P(A)$ is the probability of the event $A \in \mathcal{F}$ if the following axioms hold:

1) $P(\emptyset) = 0$ & $P(\Omega) = 1$

2) If A and B are two incompatible events, i.e. $A \cap B = 0$, it follows that $P(A \cup B) = P(A) + P(B)$

3) For all $A_i \in \mathcal{F}$ with $i = 1, 2, ..$ with $A_i \cap A_j = 0$ where $i \neq j$ then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Points 2) and 3) say that the probability P is additive for incompatible events.

E.g.: Dice $\quad P(\text{"even number"}) = P(\{2,4,6\}) = \frac{1}{2}$

$P(\text{"odd number"}) = P(\{1,3,5\}) = \frac{1}{2}$

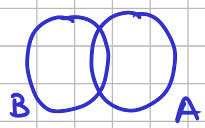
$$\rightarrow \quad P(\underbrace{\{1,2,3,4,5,6\}}_{\Omega}) = P(\{2,4,6\}) + P(\{1,3,5\}) = 1$$

Theorem: (addition law for arbitrary events)

If A and B are events in a probability space $(\Omega, \mathcal{F}, P)$ we have

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

Proof: We write B as a disjoint union $B = (B \setminus A) \cup (A \cap B)$ where $(B \setminus A) \cap (A \cap B) = \emptyset$.



It follows that

$P(A) + P(B) = P(A) + P(B \setminus A) + P(A \cap B)$

but since $\quad A \cap (B \setminus A) = 0$

$\Rightarrow P(A) + P(B \setminus A) = P(A \cup (B \setminus A)) = P(A \cup B)$

and thus $\quad P(A) + P(B) = P(A \cup B) + P(A \cap B)$

Theorem: (law of complements)

If $A \in \mathcal{F}$ is an event and $A^c = \Omega \setminus A$ is the complementary event we have

$$\boxed{P(A) = 1 - P(A^c)}$$

Proof: $\quad 1 = P(\Omega) = P(A \cup A^c) \underset{A \cap A^c = 0}{=} P(A) + P(A^c)$

Theorem: If A and B are events with $A \subseteq B$ it follows that $P(A) \leq P(B)$.

Proof: Decompose B into incompatible events A and $B \setminus A$:

$$B = A \cup (B \setminus A)$$

$$\Rightarrow P(B) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \geq P(A)$$

Exercise (Addition law)

Suppose we know that $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$ and $P(A \cup B) = \frac{3}{4}$. We can compute the probability that A and B do not occur simultaneously $P((A \cap B)^c)$:

$$\begin{aligned} P((A \cap B)^c) = 1 - P(A \cap B) &= 1 - \left(P(A) + P(B) - P(A \cup B)\right) \\ &= 1 - \frac{1}{3} - \frac{1}{2} + \frac{3}{4} \\ &= \frac{11}{12} \end{aligned}$$

1.2 Finite probability spaces

Def: A finite probability space is a probability space $(\Omega, \mathcal{F}, P)$ where the sample set Ω contains finitely many elements $\Omega = \{a_1, a_2, a_3, \dots, a_n\}$ with $n \in \mathbb{N}$.

1.3 Finite equiprobable probability spaces

Def: A probability space is equiprobable if all elementary events $\{a_i\}$ with $a_i \in \Omega$ are assigned the same probability $p_i = P(\{a_i\})$.

Corollary: Let $A \subseteq \Omega = \{a_1, a_2, \dots, a_n\}$ be an event with cardinality $|A| = k$. It follows that

$$\boxed{\begin{aligned} &i) \quad P(\{a_i\}) = \frac{1}{|\Omega|} = \frac{1}{n} \\ &ii) \quad P(A) = \sum_{i: a_i \in A} P(\{a_i\}) = \frac{|A|}{|\Omega|} = \frac{k}{n} \end{aligned}}$$

Example 1) Dice:

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

$$P(\{1\}) = P(\{2\}) = \dots = P(\{6\}) = \frac{1}{|\Omega|} = \frac{1}{6}$$

It follows that the probability for

i) an even number is $P(\{2,4,6\}) = \frac{3}{6} = \frac{1}{2}$

ii) a number lower than 3 is $P(\{1,2\}) = \frac{2}{6} = \frac{1}{3}$

Example 2) Throw a coin three times

Sample space $\Omega = \{(H,H,H), (H,H,T), (H,T,H), \dots\}$

with $|\Omega| = 2^3 = 8$

Probability of event

i) $A =$ "obtain 'tail' once"

$= \{(H,H,T), (H,T,H), (T,H,H)\}$

$\rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{3}{8}$

ii) $B =$ "obtain 'heads' at laest once"

$= \{(T,T,T)\}^c$

$P(B) = 1 - P(B^c) = 1 - \frac{1}{8} = \frac{7}{8}$

Example 3) A box contains 10 screws of which 3 are broken. You extract two screws. What is the probability that none of them is broken?

a) with replacement

b) without replacement

a) $\Omega = \{(v_i, v_j) \mid i,j = 1, \dots, 10\}$ with v_8, v_9, v_{10} broken

$|\Omega| = 10^2 = 100$

$A =$ "first and second screw are not broken"

$= \{(v_i, v_j) \mid i,j = 1, \dots, 7\}$

$\rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{7^2}{10^2} = \frac{49}{100} \hat{=} 49\,\%$

b) $\Omega' = \{(v_i, v_j) \mid i \neq j \text{ with } i,j = 1, \dots, 10\}$

$|\Omega'| = 10 \cdot 9 = 90$

$A' = \{(v_i, v_j) \mid i \neq j \text{ with } i,j = 1, \dots, 7\}$

$|A'| = 7 \cdot 6 = 42$

$\Rightarrow P(A') = \frac{|A'|}{|\Omega'|} = \frac{42}{90} \mathrel{\hat{\approx}} 47\%$

Example 4: An urn contains 20 balls numbered from 1 to 20. We take 4 balls, what is the probability that we extract ball no. "1"?

order does not matter

$\Omega = \{\{i, j, k, l\} \mid 1 \leq i < j < k < l \leq 20\}$

$|\Omega| = \binom{20}{4} = \frac{20!}{4!\,16!} = 4845$

$A = \{\{1, j, k, l\} \mid 2 \leq j < k < l \leq 20\}$

$|A| = \binom{19}{3} = 969$

$\Rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{19!}{3!\,\cancel{16!}} \; \frac{4!\,\cancel{16!}}{20!} = \frac{4}{20} = \frac{1}{5} \mathrel{\hat{=}} 20\%$

1.4 Basic combinatorics

Proposition: The number of sequences without repetition of length k chosen from n objects is

$$D(n,k) = \frac{n!}{(n-k)!}$$

E.g.: $n=3$ $k=2$ chosen from set $\{a,b,c\}$

$\to$ $\{(a,b),(a,c),(b,a),(b,c),(c,a),(c,b)\}$

$$D(3,2) = \frac{3!}{1!} = 6$$

- order matters
- no repetition

Proposition: The number of sequences with repetition of length k chosen from n objects is

$$D_R(n,k) = n^k$$

E.g.: $n=3$ $k=2$ as before

$\to$ $\{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}$

$$D_R(3,2) = 3^2 = 9$$

- order matters
- repetition

Proposition: The number of subsets of cardinality k of a set of cardinality n is

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

E.g: $n=3$ $k=2$ as before

- order does not matter
- no repetition

$\{\{a,b\},\{a,c\},\{b,c\}\}$

$$C(3,2) = \frac{3!}{2!\,1!} = 3$$

$\to$ Compare previous examples!

1.5 Conditional probability and independence

We use the conditional probability to express the probability of an event under the condition that another event happens.

Def: The conditional probability that A occurs given that B occurs is defined as

$$\boxed{P(A|B) = \frac{P(A\cap B)}{P(B)}} \qquad \text{with} \qquad P(B) > 0$$

For finite equiprobable probability spaces we thus have

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

Example 1) Two fair distinguishable dice are thrown. Given the first shows "3" what is the probability that the total exceeds 6?

Intuition: $\Omega = \{(3,1), \dots, (3,6)\}$ with $|\Omega| = 6$

$$P(\{(3,4),(3,5),(3,6)\}) = \frac{3}{6} = \frac{1}{2}$$

Rigorously: $\Omega = \{1,2,3,4,5,6\}^2 = \{(1,1),(1,2),\dots\}$

$|\Omega| = 36$

$B = \{(3,i) \mid i = 1,2,\dots,6\}$ with $|B| = 6$

$A = \{(i,j) \mid i+j > 6\}$ with $|A| = 21$

$A\cap B = \{(3,4),(3,5),(3,6)\}$ with $|A\cap B| = 3$

$$\rightarrow \quad P(A|B) = \frac{|A\cap B|}{|B|} = \frac{3}{6} = \frac{1}{2}$$

Example 2) A family has two children. What is the probability that both are boys, given that at least one is a boy?

$$\Omega = \{(B,B), (G,B), (B,G), (G,G)\}$$

$$P(\{(B,B)\} \mid \{(B,B),(B,G),(G,B)\}) = \frac{1}{3}$$

Corollary: From the definition of the conditional probability follows that

$$P(A \cap B) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B)$$

for $P(A), P(B) \neq 0$.

More generally, for n events $A_1, \dots, A_n$ we get

$$\boxed{P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 \cap A_2) \cdots P\left(A_n \,\Big|\, \bigcap_{i=1}^{n-1} A_i\right)}$$

Example: An urn contains 5 balls numbered from 1 to 5.

Balls 1, 2 & 3 are red and 4 & 5 are white. We extract 3 balls without replacement. What is the probability that all are red?

$\Omega = \{(i,j,k) \mid i \neq j \neq k \text{ with } i,j,k = 1,\dots,5\}$

$|\Omega| = D(5,3) = 5 \cdot 4 \cdot 3 = 60$

$R_1 = \{(i,j,k) \mid i = 1,2,3 \text{ and } i \neq j \neq k \text{ with } j,k = 1,\dots,5\}$

$R_2 = \{(j,i,k) \mid \quad -\text{''}- \quad \}$

$R_3 = \{(j,k,i) \mid \quad -\text{''}- \quad \}$

$|R_i| = 3 \cdot D(4,2) = 3 \cdot 4 \cdot 3 = 36$

$$P(R_1 \cap R_2 \cap R_3) = P(R_1) \cdot P(R_2 \mid R_1)\, P(R_3 \mid R_1 \cap R_2)$$

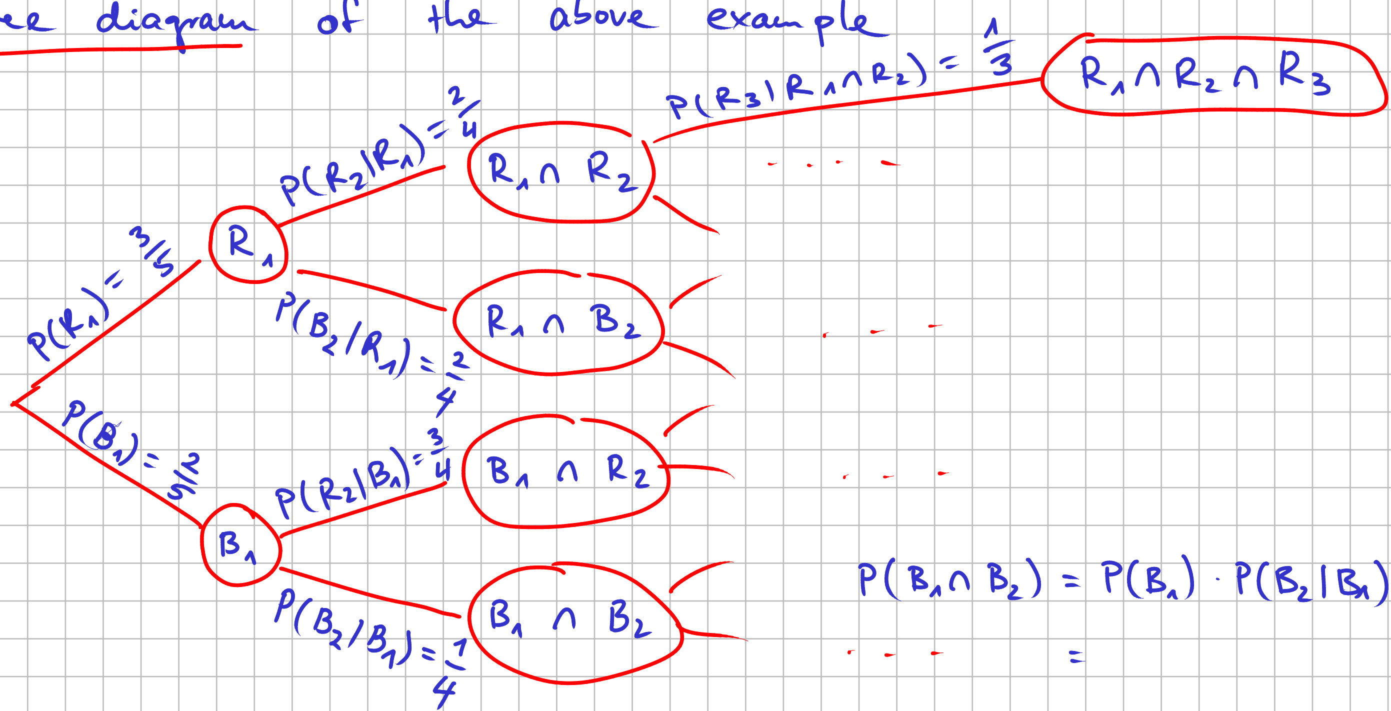
- $P(R_1) = \dfrac{|R_1|}{|\Omega|} = \dfrac{36}{60} = \dfrac{3}{5}$

- $P(R_2 | R_1) = \dfrac{|R_2 \cap R_1|}{|R_1|} = \dfrac{18}{36} = \dfrac{2}{4}$

- $P(R_3 | R_1 \cap R_2) = \dfrac{|R_3 \cap R_1 \cap R_2|}{|R_1 \cap R_2|} = \dfrac{6}{18} = \dfrac{1}{3}$

We obtain the (intuitive) result $P(R_1 \cap R_2 \cap R_3) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = \frac{1}{10}$

Tree diagram of the above example



$P(B_1 \cap B_2) = P(B_1) \cdot P(B_2 | B_1)$
$=$

Can be extended to the case of n colours.

Def: Two events A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

It follows that $P(B|A) = P(B)$ if A and B are independent.

Example: Throw a dice twice.

$\Rightarrow$ Sample space $\Omega = \{1,2,\dots,6\}^2 = \{(1,1),(1,2),\dots\}$

with $|\Omega| = 36$

Let $A = \{(6,a) \mid a = 1,\dots,6\}$ "6 at first throw"

$B = \{(a,6) \mid a = 1,\dots,6\}$ "6 at second throw"

$A \cap B = \{(6,6)\}$

$\rightarrow$ $P(A \cap B) = \frac{1}{36}$

$P(A) = P(B) = \frac{6}{36} = \frac{1}{6}$

$\leadsto$ $P(A \cap B) = P(A)\,P(B)$

$\Rightarrow$ A & B are independent

Theorem: If A and B are independent also A and B^c, A^c and B and A^c and B^c are independent.

Proof: $P(A) = P(A \cap \Omega) = P(A \cap (B \cup B^c)) = P((A \cap B) \cup (A \cap B^c))$

$$= P(A \cap B) + P(A \cap B^c) - \underbrace{P(\overbrace{(A \cap B) \cap (A \cap B^c)}^{\emptyset})}_{=0}$$

$\Leftrightarrow$ $P(A \cap B^c) = P(A) - P(A \cap B) \underset{A \,\&\, B \text{ indep.}}{=} P(A) - P(A)\,P(B)$

$= P(A)(1 - P(B)) = P(A)\,P(B^c)$

* works similarly for the other cases.

Theorem: (Total probability)

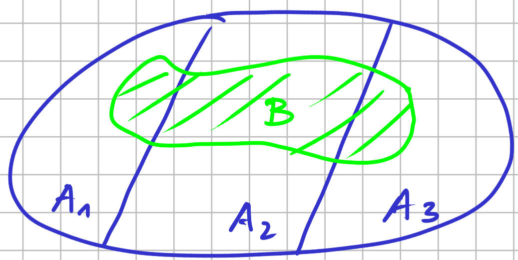
Given a probability space $(\Omega, \mathcal{F}, P)$ and events $A_1, A_2, \dots, A_n \in \mathcal{F}$ with $A_1 \cup \dots \cup A_n = \Omega$ where $A_i \cap A_j = \emptyset$ are pairwise disjoint $(i \neq j)$ and $P(A_i) > 0$ for all i then

$$\boxed{P(B) = \sum_{i=1}^{n} P(A_i)\; P(B \mid A_i)}$$

where $B \in \mathcal{F}$.

Diagram:

$n = 3$



Proof: By definition we have $P(B \mid A_i) = \dfrac{P(A_i \cap B)}{P(A_i)}$ for all i.

$$\Rightarrow \sum_{i=1}^{n} P(A_i)\, P(B \mid A_i) = \sum_{i=1}^{n} P(A_i \cap B)$$

$$\overset{\text{disjoint}}{=} P\left(\bigcup_{i=1}^{n} (A_i \cap B)\right)$$

$$= P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap B\right)$$

$$= P(\Omega \cap B)$$

$$= P(B)$$

Example: Dice $\quad A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$ & $A_3 = \{5, 6\}$

$B = \{1, 5, 6\}$

$$P(B) = \sum_{i=1}^{3} P(A_i)\, P(B \mid A_i) = \sum_{i=1}^{3} \frac{\cancel{|A_i|}}{|\Omega|} \frac{|B \cap A_i|}{\cancel{|A_i|}}$$

$$= \frac{1}{6}(1 + 0 + 2) = \frac{1}{2} \left(= \frac{|B|}{|\Omega|}\right)$$

Bayes' formula

Simple version : given an event B with $P(B) > 0$ we have

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

Proof follows from definition $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

Recall
Example 2) A family has two children. What is the probability that both are boys, given that at least one is a boy?

$$\Omega = \{(B,B), (G,B), (B,G), (G,G)\}$$

$$P(\underbrace{\{(B,B)\}}_{A} | \underbrace{\{(B,B),(B,G),(G,B)\}}_{C}) = \frac{1}{3}$$

Now compute $\quad P(C|A) = \dfrac{P(C)}{P(A)} P(A|C)$

$$= \frac{3}{4} \cdot \frac{4}{1} \cdot \frac{1}{3} = 1$$

the probability that at least one is a boy given that two are boys.

General version : Let $A_1, \dots, A_n \in \mathcal{F}$ with $A_1 \cup \dots \cup A_n = \Omega$ and $A_i \cap A_j = \emptyset$ for $i \neq j$ as before in the total probability theorem. It follows that

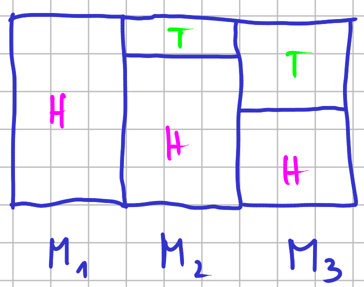
$$\boxed{P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum_{j=1}^{n} P(B|A_j) P(A_j)}}$$

Proof follows from the simple version $P(A_i|B) = \frac{P(A_i)}{P(B)} P(B|A_i)$ when inserting the total probability theorem for $P(B)$.

Example: There are three coins in a box. The first has "head" on both sides, the second yields "head" with 80% probability and "tail" with 20% and the third is a fair coin.

We take a random coin and toss it. The result is "head". What is the probability that we picked the first coin?

Graphic:



M_i = " extract coin i "

H = " result is head "

We have $P(M_i) = 1/3$ and

$$P(H|M_1) = 1$$
$$P(H|M_2) = 4/5$$
$$P(H|M_3) = 1/2$$

<u>Bayes</u>:

$$P(M_1|H) = \frac{P(H|M_1)\, \cancel{P(M_1)}}{\sum_{i=1}^{3} P(H|M_i)\, \cancel{P(M_i)}}$$

$$= \frac{1}{1 + 4/5 + 1/2} = \frac{10}{23} \simeq 0.43$$

Example: We have two coins, one fair and one with "heads" on both sides. We pick one and toss it twice. It shows twice "heads". What is the probability that we picked the fair coin?

M_1	M_2
(H,H) (H,T) (T,H) (T,T)	(H,H)

$$P(M_1 \mid \{(H,H)\})$$

$$= \frac{P(\{(H,H)\} \mid M_1)\, P(M_1)}{P(\{(H,H)\} \mid M_1)\, P(M_1) + P(\{(H,H)\} \mid M_2) P(M_2)}$$

$$= \frac{1/4 \cdot 1/2}{1/4 \cdot 1/2 + 1 + 1/2} = \frac{1/4}{5/4} = 1/5$$

Example: Box 1 contains 1000 bulbs of which 10 % are broken. Box 2 contains 2000 bulbs of which 5 % are broken. Two bulbs are taken from a randomly selected box

a) What is the probability that both are broken

b) If both are broken, what is the probability that they come from box 1?

a) X = "two bulbs are broken"

$$P(X \mid B_1) = \frac{100}{1000} \cdot \frac{99}{999}$$

$$P(X \mid B_2) = \frac{100}{2000} \cdot \frac{99}{1999}$$

Total probability theorem

$$P(X) = P(X \mid B_1) \cdot P(B_1) + P(X \mid B_2) P(B_2)$$

$$= \frac{1}{10} \frac{99}{999} \cdot \frac{1}{2} + \frac{1}{20} \frac{99}{1999} \cdot \frac{1}{2}$$

$$= 0{,}0062$$

b) $P(B_1|X) = \dfrac{P(X|B_1)\,P(B_1)}{P(X)}$

$= 0{,}80$

2 Random Variables

2.1 Definitions and basic properties

Def: Given a probability space $(\Omega, \mathcal{F}, P)$. A random variable X is a function that assigns a real number to every elementary event $\{a\}$ with $a \in \Omega$ such that for any $s, t \in \mathbb{R}$ with $s < t$ we have $\{a \mid s < X(a) \le t\} \subseteq \Omega$.

(introduces order)

Example: Toss a coin three times

$$\Omega = \{(H,H,H), (H,H,T), \dots\} \quad \text{with } |\Omega| = 2^3 = 8$$

Let X denote the number of heads such that

$$X((T,T,T)) = 0$$

$$X((T,T,H)) = X((T,H,T)) = X((H,T,T)) = 1$$

$$X((T,H,H)) = X((H,T,H)) = X((H,H,T)) = 2$$

$$X((H,H,H)) = 3$$

Can compute probabilities:

E.g.: a) $P(X=0) = P(\{T,T,T\}) = \frac{1}{8}$

b) $P(X=1) = P(X=2) = \frac{3}{8}$

c) $P(X = 1{,}5) = 0$

d) $P(0 \le X \le 1) = P(X=0) + P(X=1) = \frac{1+3}{8} = \frac{1}{2}$

* Several random variables can be defined on one probability space. Eg: X = "number of heads"
Y = "number of tails"
$Z = X \cdot Y$

Def: Given a random variable X on $(\Omega, \mathcal{F}, P)$ we define the distribution function $F: \mathbb{R} \to \mathbb{R}$ as

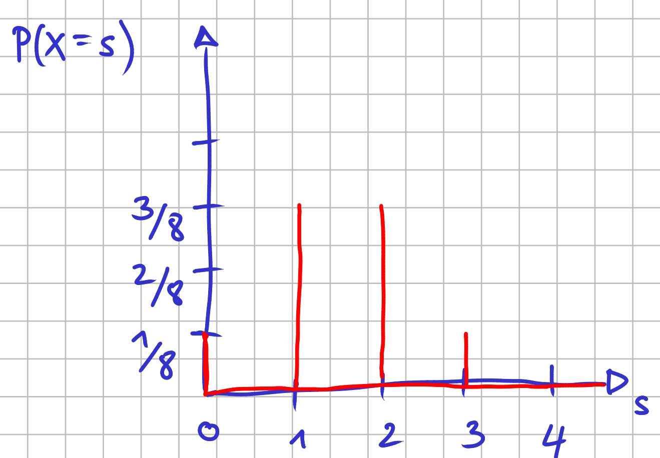
$$\boxed{F_X(s) = P(X \leq s) \quad \text{for } s \in \mathbb{R}}$$

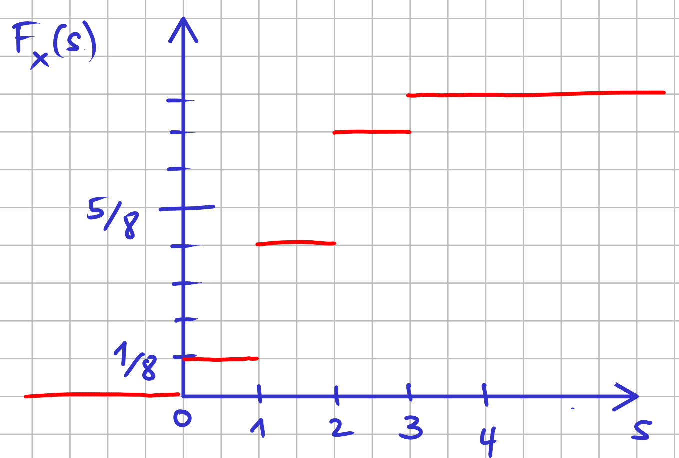
→ $F_X(s)$ measures the probability that X does not exceed "s".

Example: Toss coin three times
X = "# of heads"

$$F_X(s) = \begin{cases} P(X<0) = 0 & \text{for } s<0 \\ P(X=0) = 1/8 & \text{for } 0 \leq s < 1 \\ P(X=0) + P(X=1) = 4/8 & \text{for } 1 \leq s < 2 \\ P(X=0) + \ldots + P(X=2) = 7/8 & \text{for } 2 \leq s < 3 \\ P(X=0) + \ldots + P(X=3) = 1 & \text{for } s \geq 3 \end{cases}$$

Graphs:





General properties

Proposition 1: Let $s, t \in \mathbb{R}$ with $s < t$. We have

$$\boxed{P(s < X \leq t) = F_X(t) - F_X(s)}$$

Follows from definition:

$$F_X(t) = P(X \leq t) = P(X \leq s) + P(s < X \leq t)$$
$$= F_X(s) + P(s < X \leq t) .$$

Prop. 2: The distribution function is monotone non-decreasing

$$\boxed{F_X(s) \leq F_X(t) \quad \text{if} \quad s \leq t}$$

(without proof)

Prop. 3: We have

$$\boxed{\lim_{s \to -\infty} F_X(s) = 0 \quad \text{and} \quad \lim_{s \to +\infty} F_X(s) = 1}$$

We distinguish three types of random variables:

a) discrete

b) continuous

c) mixed (discrete and continuous)

$\left.\begin{array}{l}\text{a)}\\ \text{b)}\end{array}\right\}$ focus on these

2.2 Discrete random variables

Def: A random variable is discrete if there exists a finite set $\{s_1, \dots, s_n\}$ (or coutable $\{s_i \mid i \in \mathbb{N}\}$) such that $P(X = s_i) > 0$ for all i with $\sum_i P(X = s_i) = 1$.

Discrete random variables are often represented as an array

$$X : \begin{pmatrix} s_1 & s_2 & \dots \\ p_1 & p_2 & \dots \end{pmatrix} \quad \text{where} \quad p_i = P(X = s_i)$$

Equivalently a discrete random variable is represented by the (probability) mass function

$$\boxed{f_X(s) = P(X = s)}$$

i.e.
$$f_X(s) = \begin{cases} p_i & \text{if } s = s_i \\ 0 & \text{else} \end{cases}.$$

The distribution function is then written as

$$\boxed{F_X(s) = \sum_{s_i \leq s} f_X(s_i)}$$

since
$$F_X(s) = P(X \leq s) = \sum_{s_i \leq s} P(X = s_i) = \sum_{s_i \leq s} f_X(s_i).$$

2.3 Continuous random variables

Def: A random variable X is continuous if there exists a function f_X with

$$f_X(t) \geq 0 \quad \forall\, t \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{+\infty} f_X(t)\, dt = 1,$$

such that the distribution function F_X is represented as

$$\boxed{F_X(s) = \int_{-\infty}^{s} f_X(t)\, dt \qquad \forall\, s \in \mathbb{R}.}$$

The function f_X is called probability density of the random variable X.

Observation 1: The derivative of the distribution function yields the probability density

$$F_X'(s) = f_X(s)$$

Observation 2: Probabilities are computed by evaluating the integral

$$\boxed{P(s \leq X \leq s') = \int_{s}^{s'} f_X(t)\, dt} \qquad \text{with} \quad s < s'$$

Since $P(s \leq X \leq s') = F_X(s') - F_X(s) = \int_{-\infty}^{s'} f_X(t)\,dt - \int_{-\infty}^{s} f_X(t)\, dt$

$$= \int_{s}^{s'} f_X(t)\, dt .$$

Observation 3: For continuous random variables we have

$$P(X=s) = \int_s^s f_X(t)\,dt = 0.$$

In contrast to the discrete case where

$$P(X=s) = f_X(s).$$

Further in the continuous case we have

$$P(s<X<t) = P(s\le X<t) = P(s<X\le t) = P(s\le X\le t)$$

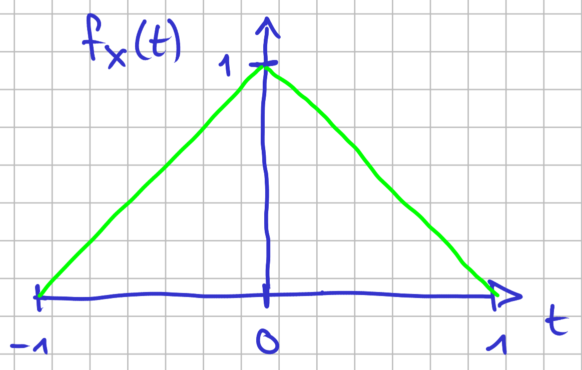
which follows from the continuity of F:

$$\lim_{\varepsilon\to 0} F_X(s\pm\varepsilon) = F_X(s) \quad \text{etc.}$$

Example: Triangle function or customer at bar

A pub has a two meters long counter (bar) $\Omega=[-1,+1]$. The probability density f_X of the random variable X denoting the position of a client arriving at the counter $(X(a)=a\in\Omega)$ shall be approximated by

$$f_X(t) = \begin{cases} 1+t & \text{if} \quad -1\le t\le 0 \\ 1-t & \text{if} \quad 0<t\le 1 \\ 0 & \text{else} \end{cases}$$



We compute the distribution function

$$F_X(s) = \int_{-\infty}^{s} f_X(t)\,dt$$

$$= \begin{cases} 0 & \text{for} \quad s < -1 \\ \int_{-1}^{s} (1+t)\,dt = \frac{1}{2}(s+1)^2 & \text{for} \quad -1 \leq s < 0 \\ \int_{-1}^{0} (1+t)\,dt + \int_{0}^{a} (1-t)\,dt = \frac{1}{2} + \frac{s}{2}(2-s) & \text{for} \quad 0 \leq s < 1 \\ 1 & \text{for} \quad s \geq 1 \end{cases}$$

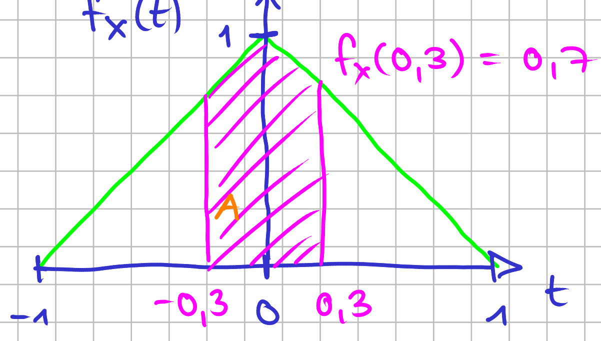
We can compute the probability that a customer arrives less than 30cm away from the center

$$P(-0{,}3 \leq X \leq 0{,}3) = F_X(0{,}3) - F_X(-0{,}3)$$

$$= \frac{1}{2} + \frac{0{,}3}{2}(2-0{,}3) - \frac{1}{2}(-0{,}3+1)^2$$

$$= 0{,}51$$

We are computing the are beyond the probability density



$$A = 2 \cdot \left(0{,}3 \cdot 0{,}7 + \frac{0{,}3 \cdot 0{,}3}{2}\right)$$

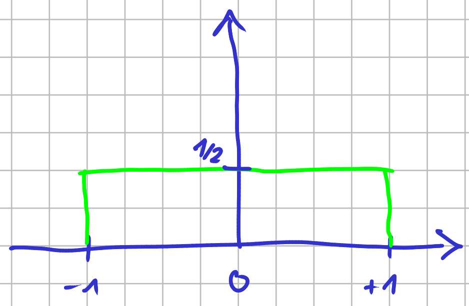
$$= 0{,}51$$

We also observe that $\int_{-\infty}^{+\infty} f_X(t)\,dt = 1$.

Example 2: rectangle function

probability density:

$$f_X(t) = \begin{cases} 0 & \text{for} \quad t < -1 \\ 1/2 & \text{for} \quad -1 \leq t \leq +1 \\ 0 & \text{for} \quad t > 1 \end{cases}$$



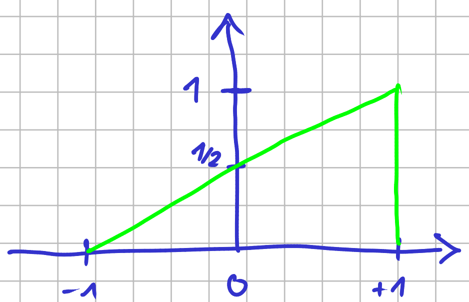
distribution function

$$F_X(s) = \begin{cases} 0 & \text{for} \quad s < -1 \\ \frac{1}{2}\int_{-1}^{s} dt = \frac{1}{2}[t]_{-1}^{s} = \frac{1}{2}(s+1) & \text{for} \quad -1 \leq s < 1 \\ 1 & \text{for} \quad s \geq 1 \end{cases}$$

$$\begin{aligned} P(-0{,}3 \leq X \leq 0{,}3) &= F_X(0{,}3) - F_X(-0{,}3) \\ &= \tfrac{1}{2}(0{,}3+1) - \tfrac{1}{2}(-0{,}3+1) \\ &= 0{,}3 \end{aligned}$$

Example 3: Slope function

$$f_X(t) = \begin{cases} 0 & \text{for} \quad t < -1 \\ \frac{1}{2}(1+t) & \text{for} \quad -1 \leq t \leq 1 \\ 0 & \text{for} \quad t > 1 \end{cases}$$



$$F_X(s) = \begin{cases} 0 & \text{for } s < -1 \\ \int_{-1}^{s} \frac{1}{2}(t+1) = \frac{1}{2}\left[\frac{1}{2}t^2 + t\right]_{-1}^{s} = \frac{1}{4}(s+1)^2 & \text{for } -1 \leq s < 1 \\ 1 & \text{for } s \geq 1 \end{cases}$$

$$P(-0{,}3 \leq X < 0{,}3) = 0{,}3$$

2.4 Expectation and variance

Def: The expectation of a random variable X is denoted by $E(X)$ or μ_X and defined via

a) discrete case: $\boxed{E(X) = \sum_i s_i f_X(s_i)}$

b) continuous case: $\boxed{E(X) = \int_{-\infty}^{+\infty} t \cdot f_X(t)\,dt}$

Here we assume that the sum/integral converges.

Example discrete: Toss a coin three times

$X =$ "# of heads"

$$P(X=0) = P(X=3) = 1/8$$
$$P(X=1) = P(X=2) = 3/8$$

$$E(X) = \sum_{i=1}^{4} s_i f_X(s_i) = \sum_{i=0}^{3} i\, f_X(i) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8}$$
$$= 3/2 = 1{,}5$$

mass function

$s_i = i-1$

Examples: 1) Triangle function

$$E(X) = \int_{-\infty}^{+\infty} t \; f_X(t)\, dt = \int_{-1}^{0} t(1+t) + \int_{0}^{1} t(1-t) = -\frac{1}{6} + \frac{1}{6} = 0$$

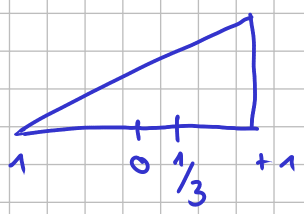
2) rectangle function

$$E(x) = \int_{-\infty}^{+\infty} t \; f_x(t)\, dt = \frac{1}{2}\int_{-1}^{+1} t \; dt = \frac{1}{4}\left[t^2\right]_{-1}^{1} = 0$$

3) slope function

$$E(x) = \int_{-\infty}^{+\infty} t \; f_x(t)\, dt = \frac{1}{2}\int_{-1}^{+1} (1+t)t = \frac{1}{2}\,\frac{1}{3}\left[t^3\right]_{-1}^{1} = \frac{1}{3}$$

The expectation yields the "center of mass" of the density function



→ convenient place for barkeeper

Def: The <u>variance</u> of a random variable X is denoted by $Var(X)$ or σ_X^2 and defined as

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

such that

a) discrete case: $\boxed{Var(X) = \sum_i{}' (s_i - \mu_X)^2 f_X(s_i)}$

b) continuous case: $\boxed{Var(X) = \int_{-\infty}^{+\infty} (t-\mu_X)^2 f_X(t)\, dt}$

The square root of the variance is called the <u>standard deviation</u> and is denoted by σ_X.

Observation:

i) The variance is always non-negative $Var(X) \geq 0$

ii) The variance (and the standard deviation) measure the variance of the random variable from the expectation value: for small $Var(x)$, the values of X concentrate around μ_X and the other way around.

Example: Toss a coin three times

$X =$ "# of heads"

$\mu_X = 3/2$

$$\sigma_X^2 = \sum_{i=1}^{4} (s_i - \mu_X)^2 f_X(s_i) = \sum_{i=1}^{3} (i - 3/2)^2 f_X(i)$$

$$= \frac{9}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{9}{4} \cdot \frac{1}{8}$$

$$= \frac{9+3}{16} = \frac{3}{4}$$

Example 1) Triangular function $\mu_X = 0$

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (t-\mu_X)^2 f_X(t)\, dt = \int_{-1}^{0} t^2 (1+t)\, dt + \int_{0}^{1} t^2 (1-t)\, dt$$

$$= \frac{1}{6} \approx 0{,}17$$

2) rectangular function $\mu_X = 0$

$$\sigma_X^2 = \frac{1}{2} \int_{-1}^{1} t^2\, dt = \frac{1}{3} \approx 0{,}33$$

(bigger than $\frac{1}{6}$)

("barkeeper has to move more")

3) slope function

$$\sigma_X^2 = \int_{-1}^{1} (t - \tfrac{1}{3})^2 f_X(t)\, dt = \frac{2}{9} \approx 0{,}22$$

EX We throw a dice twice and define the random variables X = "sum of eyes" and Y = "# of times the 6 appears".

- compute $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$

$$\Omega = \{(i,j) \mid i,j = 1,\dots,6\}$$

$$\mu_X = \sum_{k=2}^{12} k \, P(X=k)$$

$$= 2\,\frac{1}{36} + 3\,\frac{2}{36} + \dots + 7\,\frac{6}{36} + 8\,\frac{5}{36} + \dots + 12\,\frac{1}{36}$$

$$= 7$$

$$\sigma_X^2 = \sum_{k=2}^{12} (k-\mu_X)^2 \, P(X=k)$$

$$= (-5)^2 \frac{1}{36} + \dots + (-1)^2 \frac{5}{36} + (1)^2 \frac{5}{36} + \dots + 5^2 \frac{1}{36}$$

$$= 2\left(\frac{25}{36} + \frac{16\cdot 2}{36} + \frac{9\cdot 3}{36} + \frac{4\cdot 4}{36} + \frac{5}{36}\right)$$

$$= \frac{1}{18}\left(25 + 32 + 27 + 16 + 5\right)$$

$$= \frac{105}{18}$$

$$= \frac{35}{6}$$

$$\mu_Y = \sum_{k=0}^{2} k \, P(Y=k)$$

$$= 1\cdot\frac{10}{36} + 2\,\frac{1}{36}$$

$$= \frac{12}{36}$$

$$= \frac{1}{3}$$

$$\sigma_Y^2 = \sum_{k=0}^{2} (k-\mu_Y)^2 \, P(Y=k)$$

$$= \frac{1}{9}\cdot\frac{25}{36} + \frac{4}{9}\,\frac{10}{36} + \frac{25}{9}\cdot\frac{1}{36}$$

$$= \frac{5}{18}$$

Def: The expectation of the k-th power

$$E(X^k) = \begin{cases} \sum_i s_i^k f_X(s_i) & \text{discrete} \\ \int_{-\infty}^{+\infty} s^k f_X(s) & \text{continuous} \end{cases}$$

Knowing all "moments" $E(X^k)$ one can obtain probability density.

Proposition: we have
i) $E(X-\mu_X) = 0$
ii) $\sigma_X^2 = E(X^2) - \mu_X^2$

Proof for discrete case:

i) $$E(X-\mu_X) = \sum_i (s_i - \mu_X) f_X(s_i) = \mu_X - \mu_X \overbrace{\sum_i f_X(s_i)}^{1} = 0$$

ii) $$\sigma_X^2 = \sum_i (s_i - \mu_X)^2 f_X(s_i) = \sum_i (s_i^2 - 2\mu_X s_i + \mu_X^2) f_X(s_i)$$
$$= E(X^2) - 2\mu_X^2 + \mu_X^2$$
$$= E(X^2) - \mu_X^2$$

Theorem: (linear transformation of random variables)
If a random variable X has expectiation μ_X and variance σ_X^2 it follows that $X' = c_1 X + c_2$ with $c_1 \neq 0$ has expectation and variance

$$\mu_{X'} = c_1 \mu_X + c_2 \quad \text{and} \quad \sigma_{X'}^2 = c_1^2 \sigma_X^2$$

Proof (discrete case):

For the discrete case we have

$$P(X = s_i) = P(c_1 X + c_2 = c_1 s_i + c_2) = P(X' = s_i')$$

and thus $f_X(s_i) = f_{X'}(s_i')$.

It follows i) $\mu_{X'} = \sum_i' s_i' f_{X'}(s_i') = \sum_i' (c_1 s_i + c_2) f_X(s_i)$

$$= c_1 \mu_X + c_2$$

ii) $\sigma_{X'}^2 = \sum_i' (s_i' - \mu_{X'})^2 f_{X'}(s_i')$

$$= \sum_i' (c_1 s_i + \cancel{c_2} - c_1 \mu_X - \cancel{c_2})^2 f_X(s_i)$$

$$= c_1^2 \sum_i' (s_i - \mu_X)^2 f_X(s_i)$$

$$= c_1^2 \sigma_X^2$$

More tricky for continuous. Start from $P(X' \leq s')$:

$$P(X' \leq s') = P(c_1 X + c_2 \leq s') = P(c_1 X \leq s' - c_2)$$

$$= \begin{cases} P\left(X \leq \frac{s'-c_2}{c_1}\right) & \text{for } c_1 > 0 \\ P\left(X \geq \frac{s'-c_2}{c_1}\right) = 1 - P\left(X \leq \frac{s'-c_2}{c_1}\right) & \text{for } c_1 < 0 \end{cases}$$

and thus

$$f_{X'}(s') = F'_{X'}(s') = \begin{cases} F'_X\left(\frac{s'-c_2}{c_1}\right) \cdot \frac{1}{c_1} & \text{for } c_1 > 0 \\ -F'_X\left(\frac{s'-c_2}{c_1}\right) \frac{1}{c_1} & \text{for } c_1 < 0 \end{cases}$$

$$= \frac{1}{|c_1|} f_X\left(\frac{s'-c_2}{c_1}\right)$$

$$\Rightarrow \mu_{X'} = \int_{-\infty}^{+\infty} s' f_{X'}(s')\, ds' = \frac{1}{|c_1|} \int_{-\infty}^{+\infty} s' f_X\left(\frac{s'-c_2}{c_1}\right) ds'$$

change of variables $r = \frac{s'-c_2}{c_1}$

$$= \begin{cases} \frac{1}{c_1} \int_{-\infty}^{+\infty} (c_1 r + c_2)\, f_X(r) \overbrace{\frac{ds'}{dr}}^{c_1} dr & \text{for } c_1 > 0 \\ -\frac{1}{c_1} \int_{+\infty}^{-\infty} (c_1 r + c_2)\, f_X(r) \overbrace{\frac{ds'}{dr}}^{c_1} dr & \text{for } c_1 < 0 \end{cases}$$

boundaries change!

$$= \int_{-\infty}^{+\infty} (c_1 r + c_2)\, f_X(r)\, dr$$

$$= c_1 \mu_X + c_2$$

Similar for $\sigma^2_{X'}$.

Recall linear transformation $X' = c_1 X + c_2$

i) $\mu_{X'} = c_1 \mu_X + c_2$

ii) $\sigma_{X'}^2 = c_1^2 \sigma_X$

Corollary: (standard variables)

If X has expectation μ_X and variance σ_X^2 the transformed random variable $Z = \frac{X - \mu_X}{\sigma_X}$ has $\mu_Z = 0$ and $\sigma_Z^2 = 1$.

Proof: We fix $c_1 = \frac{1}{\sigma_X}$ and $c_2 = -\frac{\mu_X}{\sigma_X}$

$$\Rightarrow \quad X' = \frac{X - \mu_X}{\sigma_X} = Z$$

$$\Rightarrow \quad \mu_Z = \frac{1}{\sigma_X} \mu_X - \frac{\mu_X}{\sigma_X} = 0$$

and $\sigma_Z^2 = \frac{1}{\sigma_X^2} \sigma_X^2 = 1$

Example: Slope $\mu_X = \frac{1}{3}$ & $\sigma_X^2 = \frac{2}{9} \Rightarrow Z = \frac{3}{\sqrt{2}} \left(X - \frac{1}{3} \right)$

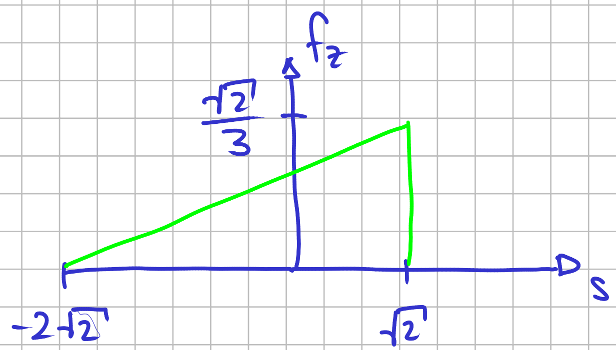
$$f_Z(s) = \frac{1}{|c_1|} f_X\left(\frac{s - c_2}{c_1} \right) = \sigma_X f_X(\sigma_X s + \mu_X)$$

$$= \frac{\sqrt{2}}{3} f_X\left(\frac{\sqrt{2}}{3} s + \frac{1}{3} \right)$$

$$= \frac{\sqrt{2}}{3} \frac{1}{2} \left(1 + \frac{\sqrt{2}}{3} s + \frac{1}{3} \right)$$

$$= \frac{s + 2\sqrt{2}}{9}$$

where $s \in \Omega_Z = \left[\frac{-1-\mu_X}{\sigma_X}, \frac{1-\mu_X}{\sigma_X}\right] = \left[-2\cdot\sqrt{2}, \sqrt{2}\right]$



Quantiles (Percentiles)

Def: Let $F_X(t)$, for $t \in \mathbb{R}$, be a distribution function of the continuous random variable X. We call a solution t_γ of the equation

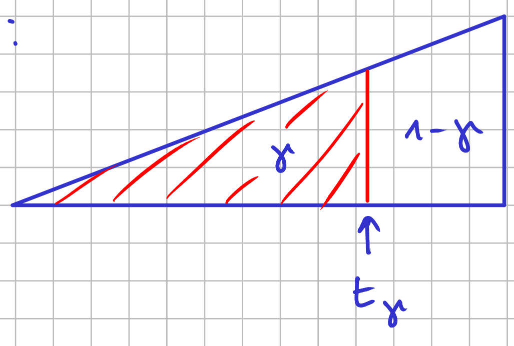
$$\gamma = F_X(t_\gamma)$$

with $\gamma \in (0,1)$ a γ-quantile.

Most common quantiles:

median	$\gamma = \frac{1}{2}$	
quartiles	$\gamma = \frac{k}{4}$	with $k = 1,2,3,4$
percentiles	$\gamma = \frac{k}{100}$	with $k = 1,2,\dots,100$

Graphically:



Remark: The median and mean are two different concepts

Example slope: $\mu = \frac{1}{3}$ but $F_X(\frac{1}{3}) = \frac{4}{9}$

instead $\leadsto t_{1/2} = \sqrt{2} - 1$

Exercise Determine the median and the first and third quartile of r.v. X given by the density

$$f_X(t) = \begin{cases} \frac{2}{49} t & 0 \le t \le 7 \\ 0 & \text{else} \end{cases}$$

$$F_X(t) = \begin{cases} 0 & t \le 0 \\ \frac{2}{49} \int_0^t ds \cdot s = \frac{t^2}{49} & 0 < t \le 7 \\ 1 & t > 7 \end{cases}$$

$\gamma = \frac{1}{4}$: $\frac{1}{4} = F_X(t_{1/4}) \Rightarrow t_{1/4} = \pm\sqrt{\frac{49}{4}} = \pm \frac{7}{2}$

$\rightsquigarrow t = 3.5$

$\gamma = 1/2$: $\frac{1}{2} = F_X(t_{1/2}) \rightsquigarrow t = 4.95$

$\gamma = 3/4$: $3/4 = F_X(t_{3/4}) \longrightarrow t = 6.06$

2.4 Conditional probability (continuous case)

Recall that for two events A & B we have

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

$\rightarrow$ generalizes to discrete random variables

e.g. Throw a coin 3 times

Let $X =$ "# of heads" and

$A = \{(H, H, H)\}$

$$P(X = 0 \mid A) = P(\{(T,T,T)\} \mid \{H,H,H\}) = 0$$

$$P(X = i \mid A) = \begin{cases} 0 & \text{if } 0 \leq i < 3 \\ 1 & \text{if } i = 3 \end{cases}$$

Continuous case

We considered events of the type $A = \{x \mid a \leq x \leq b\}$

E.g. Probability that customer arrives in the segment between -0.3 and 0.3 : $A = \{x \mid -0.3 \leq x \leq 0.3\}$

$$\Rightarrow \quad P(A) = P(-0.3 \leq X \leq 0.3) = \int_A f_X(s)\,ds = \int_{-0.3}^{0.3} f_X(s)\,ds$$

Def: The conditional density function of r.v. X with density function f_X under the condition of an event B is

$$f_X(s|B) = \begin{cases} f_X(s)/P(B) & \text{if } s \in B \\ 0 & \text{if } s \notin B \end{cases}$$

such that for an event A we have

$$P(A|B) = \int_A ds\, f_X(s|B) = \int_{A\cap B} ds\, \frac{f_X(s)}{P(B)} = \frac{P(A\cap B)}{P(B)}.$$

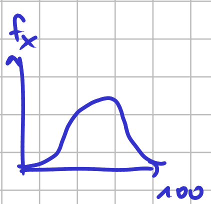
2.4 Conditional probability (example)

Recall that $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Example: Let X denote the time when a person passes away. The probability that $X \le t$ shall be given by

(Insurance company)

$$P(X \le t) = \int_0^t f_X(s)\, ds$$



where $f_X(s) = 3 \cdot 10^{-9}\, s^2 (100-s)^2$ with $0 \le s \le 100$ (years). What is the probability that a person dies between 60 and 70 given (assuming) that he got 50 years old

$$\Omega = \{ s \mid 0 \le s \le 100 \}$$

$$A = \{ s \mid 60 \le s \le 70 \}$$

$$B = \{ s \mid 50 \le s \le 100 \} \qquad \Rightarrow \quad A \cap B = A$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\int_{60}^{70} f_X(t)\,dt}{\int_{50}^{100} f_X(t)\, dt} = 0{,}486$$

using that

$$\int_a^b f_X(s)\, ds = 3 \times 10^{-9} \left[\frac{100}{3} s^3 - \frac{200}{4} s^4 - \frac{1}{5} s^5 \right]_a^b$$

2.4 Generating function

Def: Let X be a random variable. The function

$$G_X(\lambda) = E(e^{\lambda X}) = \begin{cases} \sum_i' e^{\lambda s_i} f_X(s_i) & \text{discrete} \\ \int_{-\infty}^{+\infty} e^{\lambda s} f_X(s)\, ds & \text{continuous} \end{cases}$$

with $\lambda \in \mathbb{R}$ is called moment generating function. (if all $E(X^k)$ exist & $E(e^{\lambda X}) = \sum_{k=0}^{\infty}' \frac{1}{k!} \lambda^k E(X^k)$ is convergent).

It follows that $G_X^{(k)}(\lambda)\big|_{\lambda=0} = E(X^k)$ where $G^{(k)}$ denotes the k-th derivative since $e^x = \sum_{k=0}^{\infty}' \frac{1}{k!} x^k$.

Def: In general we use the notation

$$E(g(X)) = \begin{cases} \sum_i' g(s_i) f_X(s_i) & \text{discrete} \\ \int_{-\infty}^{+\infty} g(s) f_X(s)\, ds & \text{continuous} \end{cases}$$

Example: rectangle

$$E(X^k) = \int_{-\infty}^{+\infty} s^k f_X(s)\, ds = \frac{1}{2} \int_{-1}^{+1} s^k\, ds$$

$$= \frac{1}{2k+2} \left[s^{k+1} \right]_{-1}^{+1} = \begin{cases} 0 & \text{for } k \text{ odd} \\ \frac{1}{k+1} & \text{for } k \text{ even} \end{cases}$$

$$\rightarrow G_X(\lambda) = \sum_{k=0}^{\infty}{}' \frac{1}{k!} \lambda^k E(X^k) = \sum_{\substack{k=0 \\ k \text{ even}}}^{\infty}{}' \frac{1}{(k+1)!} \lambda^k$$

$$= \sum_{k=0}^{\infty}{}' \frac{1}{(2k+1)!} \lambda^{2k} = \frac{\sinh(\lambda)}{\lambda}$$

$$= \frac{e^{\lambda} - e^{-\lambda}}{2\lambda}$$

2.5 Summary

discrete | **continuous**

Mass function

$$f_X(s) = P(X = s)$$

with $\sum_i{}' f_X(s_i) = 1$

Density function

$$f_X(s)$$

with $\int_{-\infty}^{+\infty} f_X(t)\,dt = 1$

Distribution function

$$F_X(s) = P(X \leq s) = \sum_{s_i \leq s}{}' f_X(s_i)$$

$$F_X(s) = P(X \leq s) = \int_{-\infty}^{s} f_X(t)\,dt$$

Expectation

$$\mu_X = \sum_i{}' s_i f_X(s_i)$$

$$\mu_X = \int_{-\infty}^{+\infty} t\, f_X(t)\,dt$$

Variance

$$\sigma_X^2 = \sum_i{}' (s_i - \mu_X)^2 f_X(s_i)$$

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (t - \mu_X)^2 f_X(t)\,dt$$

Generating function

$$G_X(\lambda) = \sum_i{}' e^{\lambda s_i} f_X(s_i)$$

$$G_X(\lambda) = \int_{-\infty}^{+\infty} e^{\lambda s} f_X(s)\,ds$$

3. Common discrete distributions

3.1 Bernoulli distribution

Def: A Bernoulli experiment is an experiment with two outcomes, e.g. toss a coin. One outcome (success) occurs with probability p, the other (failure) with probability $1-p$.

Def: Let $0 < p < 1$. A Bernoulli random variable X of parameter p (denoted by $X \sim Ber(p)$) is characterized by the mass function

$$f_X(s) = P(X=s) = \begin{cases} 1-p & \text{if } s=0 \quad \text{(failure)} \\ p & \text{if } s=1 \quad \text{(success)} \\ 0 & \text{else} \end{cases}$$

The corresponding distribution function reads

$$F_X(s) = \begin{cases} 0 & \text{if } s<0 \\ 1-p & \text{if } 0 \le s < 1 \\ 1 & \text{if } s \ge 1 \end{cases}$$

Theorem: A Bernoulli random variable $X \sim Ber(p)$ has expectation

$$\boxed{E(X) = p}$$

and variance

$$\boxed{Var(X) = p(1-p)}$$

Proof:

i) $E(X) = \sum_{i=1}^{2} s_i \, f_X(s_i) = 0\cdot(1-p) + 1\cdot p = p$

ii) $Var(X) = \sum_{i=1}^{2} (s_i - \mu_X)^2 f_X(s_i) = (0-p)^2\cdot(1-p) + (1-p)^2\cdot p$

$$= p^2 - \cancel{p^3} + p - 2p^2 + \cancel{p^3}$$

$$= p - p^2 = p(1-p)$$

Example: fair coin flip: $\Omega = \{H, T\}$ with $X(H) = 1$ & $X(T) = 0$.

X is Bernoulli random variable with $p = \frac{1}{2}$, $X \sim Ber(\frac{1}{2})$.

Def: Indicator function

Let A be an event and let $I_A : \Omega \to \mathbb{R}$ be the indicator fuction of A, such that

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases}.$$

Then I_A is a Bernoulli random variable with $P(I_A = 1) =$
$= P(A) = p$.

Example: A box contains 10 bulbs of which 1 is broken. We pick one at random

$\Omega = \{b_1, \dots, b_{10}\}$

$A = \{b_1\}$ $\quad \to \quad I_A(b_i) = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i \neq 1 \end{cases}$

$A^c = \{b_2, \dots, b_{10}\}$

The probability that we pick the broken one is

$$P(I_A = 1) = P(A) = \frac{|A|}{|\Omega|} = \frac{1}{10}$$

while $P(I_A = 0) = P(A^c) = \frac{9}{10}$

3.2 Binomial distribution

Thm: Suppose we repeat a Bernoulli experiment n times. The random variable X that counts the number of successes, (here indicated by k) has mass function

$$\boxed{f_X(k) = \binom{n}{k} p^k q^{n-k}}$$

where $q = 1-p$ and $0 \leq k \leq n$. The random variable X is said to have the binomial distribution, $X \sim \mathrm{Bin}(n,p)$.

Proof: $\Omega = \{\underset{(s)}{\text{success}}, \underset{(f)}{\text{failure}}\}^n$

Consider the case where the first k trials are successful

$$P(\{(\underbrace{s,\dots,s}_{k}, \underbrace{f,\dots,f}_{n-k})\}) \overset{\text{independent}}{=} \underbrace{P(\{s\})\cdots P(\{s\})}_{k} \cdot \underbrace{P(\{f\})\cdots P(\{f\})}_{n-k}$$

$$= p^k (1-p)^{n-k}$$

Denote $x_1, \dots, x_k$ the positions of the successes, e.g. $x_i = i$ above. The combinatorical prefactor is obtained by counting the number of subsets with cardinality k of the set of all positions $\{1, \dots, n\}$, i.e. $C(n,k) = \binom{n}{k}$.

Example A fair dice is rolled 5 times. What is the probability that

a) obtain "6" twice

b) obtain "6" at least twice

$p = \frac{1}{6} \qquad n = 5$

a) $k = 2 \qquad P(X=k) = \binom{5}{2}\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3$

$$= 10 \ \frac{1}{36} \ \frac{125}{216} \approx 0{,}16$$

b) $P(X \geq 2) = 1 - P(X=0) - P(X=1)$

$$= 1 - \left(\frac{5}{6}\right)^5 - 5 \ \frac{1}{6} \left(\frac{5}{6}\right)^4 \approx 0{,}20$$

Observation: $\sum_i f_X(s_i) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$

Binomial $\rightarrow$ $= \left(p + (1-p)\right)^n$

$= 1$

Theorem: The random variable with distribution $X \sim Bin(n,p)$ has expectation, variance and generating function

$$E(X) = n \cdot p$$

$$Var(X) = np(1-p)$$

$$G_X(\lambda) = (pe^\lambda + 1 - p)^n$$

proof: $\mu_X = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$

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

shift the sum $\rightarrow$ $= \sum_{k=0}^{n-1} \frac{n!}{k!\,(n-k-1)!} p^{k+1} (1-p)^{n-k-1}$

$$= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!\,(n-1-k)!} p^k (1-p)^{n-1-k}$$

Binomial formula

$$= np \, (p + 1 - p)^{n-1}$$

$$= np$$

$$\sigma_X^2 = \sum_{k=0}^{n} \underbrace{(k - np)^2}_{= k(k-1) - (2np-1)k + n^2p^2} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k}$$

$$- (2np-1)\, np + n^2 p^2$$

$$= \sum_{k=0}^{n-2} \frac{n!}{k!\,(n-k-2)!} p^{k+2} (1-p)^{n-k-2} + np(1-np)$$

$$= n\cdot(n-1)p^2 \underbrace{\sum_{k=0}^{n-2} \binom{n-2}{k} p^k (1-p)^{n-2-k}}_{=1} + np(1-np)$$

$$= \cancel{n^2p^2} - np^2 + np - \cancel{np^2}$$

$$= np\,(1-p)$$

$$G_X(t) = E(e^{tX}) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (p\cdot e^t)^k (1-p)^{n-k}$$

$$= (pe^t + 1 - p)^n$$

Example: Let $p = 1/3$ the probability that a family owns a pet. For a group of 7 families compute the following probabilities

i) two families have a dog

ii) at least two families have a dog

iii) Compute expectation and variance

$$X \sim Bin(7, 1/3) \to f_X(k) = P(X=k) = \binom{7}{k} \left(\tfrac{1}{3}\right)^k \left(\tfrac{2}{3}\right)^{7-k}$$

i)
$$P(X=2) = \binom{7}{2} \left(\tfrac{1}{3}\right)^2 \left(\tfrac{2}{3}\right)^5$$
$$= \frac{7!}{2!\,5!} \; \frac{1}{9} \; \frac{2^5}{3^5}$$
$$= \frac{224}{729} \approx 0{,}31$$

ii)
$$P(X \geq 2) = 1 - P(X \leq 1)$$
$$= 1 - \binom{7}{0}\left(\tfrac{1}{3}\right)^0 \left(\tfrac{2}{3}\right)^7 - \binom{7}{1}\left(\tfrac{1}{3}\right)^1\left(\tfrac{2}{3}\right)^6$$
$$= 1 - \left(\tfrac{2}{3}\right)^7 - \tfrac{7}{3}\left(\tfrac{2}{3}\right)^6$$
$$= \frac{179}{243} \approx 0{,}74$$

iii)
$$\mu_X = n \cdot p = \frac{7}{3} \approx 2{,}33$$
$$\sigma_X = \sqrt{np(1-p)} = \sqrt{\frac{7 \cdot 2}{9}} = \frac{1}{3}\sqrt{14}$$

k	0	1	2	3	4	5	6	7
$P(X=k)$	0,06	0,20	0,31	0,26	0,13	0,04	0,006	0,0005

3.3 Hypergeometric distribution

Def: Hypergeometric distribution

An urn contains $r+b$ balls of which r are red and b are black. We extract n balls without replacement $(r+b \geq n)$. The random variable

$$X = \text{"\# of red among the extracted } n \text{ balls"}$$

has mass function

$$\boxed{f_X(k) = P(X=k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}}$$

with $k = \max(0, n-b), \ldots, \min(r,n)$ and is denoted by $X \sim \text{Hyper}(r,b,n)$. (no derivation)

Example: boundaries

Take $r=3$ $b=2$ $n=4$ Hyper$(3,2,4)$

$$k = \cancel{0}, \cancel{1}, 2, 3, \cancel{4}$$

there are not 4 red balls

can't have less than 2 red balls if we take $n=4$

probability to draw 2 red balls:

$$P(X=2) = \frac{\binom{3}{2}\binom{2}{2}}{\binom{5}{4}} = \frac{3 \cdot 1}{5} = 0{,}6$$

Observation 1: For $n=1$ we have $\text{Hyper}(r,b,1) = \text{Bern}\left(\frac{r}{b+r}\right)$

since

$$P(X=k) = \frac{\binom{r}{k}\binom{b}{1-k}}{\binom{b+r}{1}} = \begin{cases} \frac{b}{b+r} = 1 - \overbrace{\frac{r}{r+b}}^{p} & \text{for } k=0 \\ \underbrace{\frac{r}{r+b}}_{p} & \text{for } k=1 \end{cases}$$

Observation 2: The difference between $\text{Bin}(n, p=\frac{r}{r+b})$ and $\text{Hyper}(r,b,n)$ becomes irrelevant for $N = r+b \to \infty$.

Proof:

$$P(Y=k) = \frac{\binom{r}{k}\cdot\binom{N-r}{n-k}}{\binom{N}{n}} = \frac{r!}{k!(r-k)!} \frac{n!(N-n)!}{N!} \frac{(N-r)!}{(n-k)!(N-r-n+k)!}$$

$$= \binom{n}{k} \frac{r!}{(r-k)!} \frac{(N-n)!}{N!} \frac{(N-r)!}{(N-r+k-n)!}$$

$$= \binom{n}{k} \frac{r\cdot(r-1)\cdots(r-k+1)\,(N-r)\cdot(N-r-1)\cdots(N-r+k-n+1)}{N\cdot(N-1)\cdots(N-n+1)}$$

divide top & bottom $1/N^n$

$$= \binom{n}{k} \frac{p\,(p-1/N)\cdots(p-\frac{k-1}{N})\,(1-p)(1-p-1/N)\cdots(1-p-\frac{n-k-1}{N})}{1\cdot(1-1/N)\cdots(1-\frac{n-1}{N})}$$

$$\underset{N\to\infty}{=} \binom{n}{k} p^k (1-p)^{n-k}$$

Rule of thumb

$$\boxed{\text{For } \frac{n}{r+b} < 0{,}05 \text{ it is sufficient to consider } \text{Bin}(n,\tfrac{r}{r+b}) \text{ instead of } \text{Hyper}(r,b,n).}$$

Theorem: The hypergeometric distribution has

$$\boxed{E(X) = \frac{nr}{b+r} \quad \text{and} \quad var(X) = \frac{nbr(b+r-n)}{(b+r)^2(b+r-1)}}$$

(no proof)

Exercise: Lotto

A bowl contains 90 balls numbered from 1 to 90. We bet on 10 numbers and extract 5 balls without replacement. What is the probability to have guessed at least 4 numbers?

"red balls" $r = 10$

"black balls" $b = 80$

total balls extracted $n = 5$

$$P(X=k) = \frac{\binom{10}{k}\binom{80}{5-k}}{\binom{90}{5}}$$

$$P(X \geq 4) = P(X=4) + P(X=5) \simeq 0{,}0004$$

3.4 Poisson distribution

Def: A random variable X that takes values in the set $\{0, 1, 2, \dots\}$ with mass function

$$\boxed{f_X(k) = \frac{\mu^k}{k!} e^{-\mu}} \qquad k = 0, 1, 2, \dots$$

where $\mu > 0$ is Poisson distributed with parameter μ, $X \sim Pois(\mu)$.

Thm: A Poisson random variable $X \sim Pois(\mu)$ with parameter μ has expectation μ and variance $\sigma_X^2 = \mu$.

<u>Proof:</u> i) $$E(X) = \mu_X = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = e^{-\mu} \underset{\uparrow}{\sum_{k=0}^{\infty}} \frac{\mu^k}{(k-1)!}$$

$k=0$ term vanishes

$$= e^{-\mu} \underbrace{\sum_{k=0}^{\infty} \frac{\mu^{k+1}}{k!}}_{= e^{\mu} \cdot \mu}$$

$$= \mu$$

ii) $$\sigma_X^2 = \sum_{k=0}^{\infty} \underbrace{(k-\mu)^2}_{= k(k-1) - (2\mu-1)k + \mu^2} \frac{\mu^k}{k!} e^{-\mu}$$

$$= e^{-\mu} \left[\sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} - (2\mu - 1) \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} + \mu^2 \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \right]$$

$$= \mu^2 - (2\mu - 1)\mu + \mu^2$$

$$= \mu$$

Proposition: If $X \sim Bin(n,p)$, with parameters $n \to \infty$ and $p \to 0$ while $n \cdot p = \mu$ finite we find that $Bin(n,p) \to Pois(\mu)$, i.e.

$$\binom{n}{k} p^k (1-p)^{n-k} \underset{\substack{n \to \infty \\ n \cdot p = \mu}}{\longrightarrow} \frac{\mu^k}{k!} e^{-\mu} .$$

<u>Proof:</u> $$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \overset{p = \mu/n}{=} \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

$$= \frac{\mu^k}{k!} \left(1 - \mu/n\right)^n \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\mu}{n}\right)^{-k}$$

$$= \frac{\mu^k}{k!}\left(1-\mu/n\right)^n \cdot \underbrace{1 \cdot \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\mu/n\right)^{-k}}_{\xrightarrow[n\to\infty]{} 1}$$

$$\underset{n\to\infty}{=} \frac{\mu}{k!}\, e^{-\mu}$$

where we used $\lim\limits_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n = e^{-\mu}$.

Rule of thumb

For $n \geq 100$ and $p \leq 0.1$ it is sufficient to consider $\mathrm{Pois}(\mu = n\cdot p)$ instead of $\mathrm{Bin}(n, p)$.

Exercise: Suppose that 2% of the world's population have diabetes. What is the probability to find at least 3 diabetics in a group of 100 people? Compute the result using

a) Binomial distr.

b) Poisson distr.

c) Hypergeometric distr.

a) $n = 100$ $\quad p = 0{,}02$ $\qquad P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$\begin{aligned} P(X \geq 3) &= 1 - P(X<3) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &\approx 0{,}323314 \end{aligned}$$

b) $\mu = n\cdot p = 2$ $\qquad P(X=k) = \frac{\mu^k}{k!} e^{-\mu}$

$$\begin{aligned} P(X\geq 3) &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &\approx 0{,}323324 \end{aligned}$$

c) $$P(X=k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$$

$b+r$ = world's pop. $= 7{,}95 \cdot 10^9$

$r = 0{,}02 \cdot 7{,}95 \cdot 10^9$

$= 1{,}59 \cdot 10^8$

$b = 7{,}791 \cdot 10^9$

$n = 100$

$$P(X \geq 3) = 1 - P(X=0) - P(X=1) - P(X=2)$$
$$\approx 0{,}323314$$

Remark: i) $P_{Hyper}(X \geq 3) - P_{Bin}(X \geq 3) = 3 \cdot 10^{-11}$

ii) c) is most precise result, while b) is least precise.

3.4 Geometric & negative binomial distribution

Suppose we perform a sequence of independent Bernoulli experiments Bern(p) with probability of success p, e.g. flipping a coin repeatedly.

Let

X = " # of experiments necessary for a successful outcome "

then X is a geometric random variable.

A sequence of Ber(p): $\underbrace{|x\,x\,x\,x\,x\,x}_{(1-p)^{k-1}}\underbrace{o|}_{p}$

Def: A geometric variable is a random variable X with mass function

$$f_X(k) = (1-p)^{k-1}\,p \quad \text{with } k = 1, 2, \ldots$$

and $0 < p < 1$. It is denoted by $X \sim Geo(p)$.

Example: We throw a dice until the outcome is 6. What is the probability to obtain a 6 at the 9th throw?

$$P(X=9) = (1-p)^8 \cdot p = \left(\frac{5}{6}\right)^8 \cdot \frac{1}{6} \approx 0{,}039$$

Generalisation:

The random variable

$$X = \text{" \# of experiments necessary for } r \text{ successful outcomes"}$$

then X is a geometric random variable.

A sequence of Ber(p):

Def: A negative binomial variable is a random variable X with mass function

$$f_X(k) = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} \cdot p \qquad \text{with } k = 1, 2, \dots$$

$$= \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

and $0 < p < 1$. It is denoted by $NB(r,p)$.

For $r=1$, we have $NB(1,p) = Geo(p)$.

Thm.: The negative binomial distribution has

$$E(X) = \frac{r}{p} \quad \& \quad Var(X) = r\,\frac{1-p}{p^2}.$$

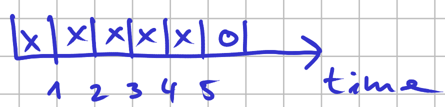
(no proof)

Example (chronological events)

A phone can ring every moment (success) or stay quite. Suppose a phone rings in average 5 times per hour and that the probability is distributed homogeneously.

What is the probability to receive the first call after 10 minutes of switching it on?

i) divide the hour into minutes and think about intervalls as independent Bernoulli experiments with $p = \frac{5}{60}$. Then the waiting time X for the first call is described by $X \sim Geo(p)$.



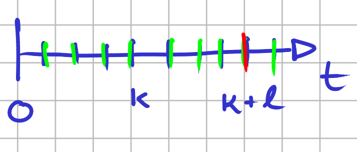
$$P(X=10) = \frac{5}{60}\left(1-\frac{5}{60}\right)^9 \approx 0{,}038$$

ii) The geometric distribution has "no memory", i.e.

$$P(X = k+l \mid X > k) = P(X = l)$$

Probability to wait $k+l$ minutes given that we waited already k minutes coincides with probability of waiting l minutes.

Proof:
$$P(X=k+l \mid X>k) = \frac{P((X=k+l)\cap(X>k))}{P(X>k)}$$



$$= \frac{P(X=k+\ell)}{P(X>k)}$$

$$P(X=k+\ell) = p(1-p)^{k+\ell-1} = P(X=\ell)\,(1-p)^k$$

$$P(X>k) = \sum_{i=k+1}^{\infty} p(1-p)^{i-1} = p \sum_{j=0}^{\infty} (1-p)^{k+j}$$

$$\sum_{k=0}^{\infty} r^k \overset{r<1}{=} \frac{1}{1-r} \qquad = p(1-p)^k \frac{1}{1-(1-p)} = (1-p)^k$$

It then follows:

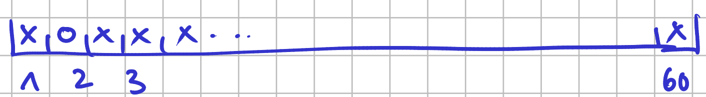
$$P(X=k+\ell \mid X>k) = \frac{P(X=k+\ell)}{P(X>k)} = P(X=\ell)$$

<u>Remark</u>:

Instead of computing the probabilities of the first call, we may ask for the probability of reiceiving the r-th call $\rightsquigarrow$ Negative Binomial $NB(p,r)$.

The probability of reiceiving k calls in a given time interval is obtained from the Binomial distribution.

Example: What is the probability to receive 3 calls in the next hour ?



$p = \frac{5}{60}$ $\qquad n = 60 \qquad k = 3$

$$\Rightarrow \quad P(X=3) = \binom{60}{3} p^3 (1-p)^{57} \approx 0.13893$$

Continuous time limit: We can continue to divide the hour into smaller segments, e.g. seconds. Then we would have $p = \frac{5}{3600}$ & $n = 3600$ such that

$$P(X=3) = \binom{3600}{3} p^3 (1-p)^{3597} \approx 0.14035.$$

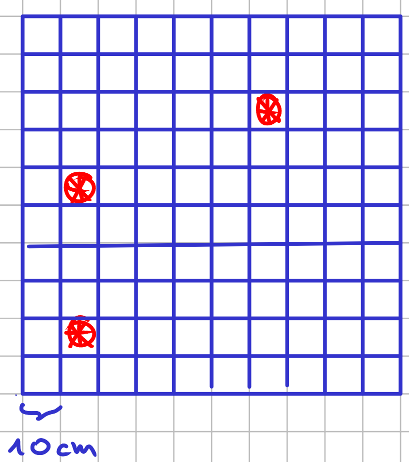
As discussed, for $n \geq 100$ and $p \leq 0{,}1$ the Binomial distribution $\mathrm{Bin}(n,p)$ converges to the Poisson distribution $\mathrm{Pois}(\mu = n \cdot p)$. Thus, for continuous time the number of calls per hour follow a Poisson law $\mathrm{Pois}(\mu = 5)$.

$$\leadsto \quad P(X=3) = \frac{5^3}{3!} e^{-5} \approx 0.14037$$

Remark: for continuous time the Binomial distribution describes a discrete time approximation of the Poisson distribution.

In the same manner the (Hypergeometric), Binomial and Poisson distribution can be used to describe events that appear in space intervalls like distance, area or volume.

Example for discrete case in 2dimensions



Eg.: 3 defect tiles per $1\,m^2$ (or Battleship)

What is the probability to choose 1 defect tiles in 2 picks?

Hypergeometric:

$$P(X=k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}} = 0.0588$$

$r = 3$
$b = 97$
$n = 2$
$k = 1$

Bin(p,n): (approximation)

$p = \frac{3}{100}$ $n = 2$ $P(X=3) = 0.0582$ $\frac{n}{r+b} = 0{,}02 < 0{,}05$

Example for continous case:

A river contains in average 5 bacteria per $1\,cm^3$ of water. What is the probability that a sample of $1\,cm^3$ contains .

a) no bacteria

b) at least 1 bacteria

$$\mu = \underbrace{\frac{5}{1cm^2}}_{=p} \cdot \underbrace{1cm^2}_{n} = 5$$

$\leadsto$ a) $P(X=0) = e^{-5} = 0{,}007$

b) $P(X \geq 1) = 1 - P(X=0) = 0{,}993$

4. Common continuous distributions

4.1. Uniform distribution

Def: Let $a, b \in \mathbb{R}$ with $a < b$. A random variable X is uniformly distributed in the interval $[a,b]$, $X \sim U([a,b])$, if the probability density is given by

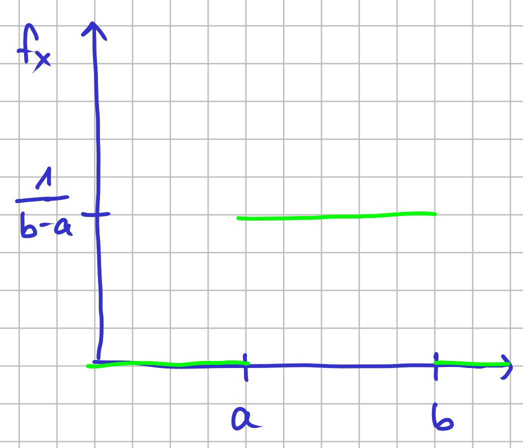
$$f_X(t) = \begin{cases} \frac{1}{b-a} & \text{for} \quad a \leq t \leq b \\ 0 & \text{else} \end{cases}$$

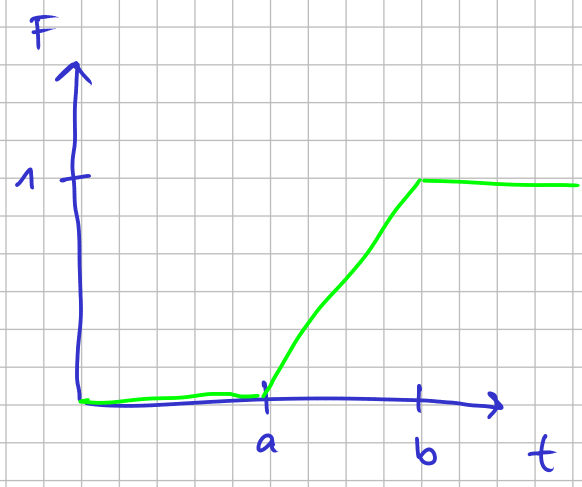
(Rectangle function for $a = -1$ & $b = 1$)

The distribution function reads

$$F_X(t) = \begin{cases} 0 & \text{for} \quad t < a \\ \frac{1}{b-a} \int_a^t ds = \frac{t-a}{b-a} & \text{for} \quad a \leq t \leq b \\ 1 & \text{for} \quad t \geq b \end{cases}$$

Graphs





Theorem The uniformly distributed random variable in the intervall $[a,b]$ has expectation and variance

$$\mu_X = \frac{a+b}{2} \qquad e \qquad \sigma_X^2 = \frac{(b-a)^2}{12}$$

proof :

$$\mu_X = \int_{-\infty}^{+\infty} t\, f_X(t)\, dt = \int_a^b \frac{t}{b-a}\, dt = \frac{1}{b-a}\left[\frac{1}{2}t^2\right]_a^b$$

$$= \frac{1}{b-a} \cdot \frac{1}{2}\left(b^2 - a^2\right)$$

$$= \frac{a+b}{2}$$

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (t - \mu_X)\, f_X(t)\, dt = \frac{1}{b-a}\int_a^b \left(t - \frac{a+b}{2}\right)^2 dt$$

$$= \frac{1}{b-a}\left[\frac{1}{3}\left(t - \frac{a+b}{2}\right)^3\right]_a^b$$

$$= \frac{1}{b-a}\,\frac{1}{3}\left(\left(\frac{b-a}{2}\right)^3 - \underbrace{\left(\frac{a-b}{2}\right)^3}_{-\left(\frac{b-a}{2}\right)^3}\right)$$

$$= \frac{2}{3}\,\frac{1}{b-a}\left(\frac{b-a}{2}\right)^3$$

$$= \frac{1}{12}(b-a)^2$$

4.2 Exponential and gamma distribution

Def: The random variable X is exponentially distributed with parameter $\lambda > 0$ if it has probability density

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{else} \end{cases}$$

The distribution function reads

$$F_X(t) = \begin{cases} \lambda \int_0^t e^{-\lambda s} ds = -\left[e^{-\lambda s} \right]_0^t = 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 . & \text{else} \end{cases}$$

Proposition: The exponential distribution is the continous analog of the geometric distribution.

Recall: waiting time $X \sim Geo(p)$ with discretised time $1h = 60\,min$

|x|x|x|x|x|o|

1 2 3 4 5 . -

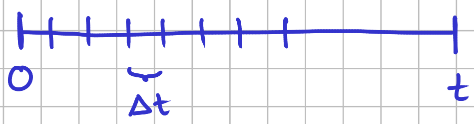
Probability to wait less than or equal k minutes

$$P(X \leq k) = \sum_{j=1}^{k} p(1-p)^{j-1}$$

⊛ This result is an approximation of continuous time

Can we do better ?

discrete time axis :



where $t = k \cdot \Delta t$

Before : i) $\Delta t = 1 \text{min}$

ii) frequency $\lambda = \frac{5}{1h} \Rightarrow p = \lambda \cdot \Delta t = \frac{5}{60}$

Now : send $\Delta t \to 0$

$$P(X \leq k) = P(X \leq t/\Delta t) = \sum_{k=1}^{t/\Delta t} p(1-p)^{k-1}$$

$$= p \sum_{k=0}^{t/\Delta t - 1} (1-p)^k$$

$$\sum_{k=0}^{n-1} a^k = \frac{1-a^n}{1-a}$$

Geometric sum

$$= p \, \frac{1-(1-p)^{t/\Delta t}}{1-(1-p)}$$

$$= 1 - (1-p)^{t/\Delta t}$$

$$= 1 - (1 - \lambda \cdot \Delta t)^{t/\Delta t}$$

$$\overset{\Delta t \to 0}{=} 1 - e^{-\lambda t}$$

where we use $\lim_{n \to \infty} (1 - \lambda/n)^n = e^{-\lambda}$.

We identify the distribution function of the exponential distribution.

Observation: $P(X \leq t) = 1 - e^{-\lambda t} = 1 - P(Y=0)$

where Y is a Poisson r.v. with $\mu = \lambda t$.

This is not a coincidence

Consider Y = "# of calls per time intervall t" discretised as before with $t = k \cdot \Delta t$ such that $p = \lambda \cdot \Delta t = \frac{\lambda \cdot t}{k}$ and $Y \sim \mathrm{Bin}\left(k, \frac{\lambda \cdot t}{k}\right)$. Since $\lim\limits_{n \to \infty} \mathrm{Bin}\left(n, \frac{\lambda \cdot t}{n}\right) = \mathrm{Pois}(\mu)$ with $\mu = \lambda t$ yields the # of calls per time intervall t in the continuum limit. The probability that at least one call arrives in time intervall t is thus

$$P(Y \geq 1) = 1 - P(Y=0) = P(X \leq t)$$

where X is the waiting time for the first call.

The probability that the waiting time X_r for the r-th call is at least t can be expressed as

$$P(X_r \leq t) = 1 - P(Y \leq r-1) = P(Y \geq r)$$

The random variable X_r is Gamma distributed.

Remark: For the negative binomial distribution an analogous limiting procedure as for the geometric yields the Gamma distribution defined via

$$f_X(t) = \begin{cases} \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{else} \end{cases}$$

Exercise: Verify that for $r=2$ we have $P(X_2 \leq t) = P(Y \geq 2)$

$$P(X_2 \leq t) = \lambda^2 \int_0^t s\, e^{-\lambda s}\, ds = \lambda^2 \left(\left[-s \frac{1}{\lambda} e^{-\lambda s} \right]_0^t + \frac{1}{\lambda} \int_0^t e^{-\lambda s}\, ds \right)$$

$$= \lambda^2 \left(-\frac{t}{\lambda} e^{-\lambda t} - \frac{1}{\lambda^2} \left[e^{-\lambda s} \right]_0^t \right)$$

$$= -\lambda t\, e^{-\lambda t} - e^{-\lambda t} + 1$$

$$= 1 - e^{-\lambda t} - \lambda t\, e^{-\lambda t}$$

$$= 1 - P(Y=0) - P(Y=1)$$

$$= P(Y \geq 2)$$

Proposition: We have $\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$

Proof:

$$\int_{-\infty}^{+\infty} f_X(t)\, dt = \frac{\lambda^r}{(r-1)!} \int_0^\infty t^{r-1} e^{-\lambda t}\, dt$$

$$= (-1)^{r-1} \frac{\lambda^r}{(r-1)!} \int_0^\infty \frac{d^{r-1}}{d\lambda^{r-1}} e^{-\lambda t}\, dt$$

$$= -\frac{(-\lambda)^r}{(r-1)!} \frac{d^{r-1}}{d\lambda^{r-1}} \underbrace{\int_0^\infty e^{-\lambda t}\, dt}_{\left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^\infty = \frac{1}{\lambda}}$$

$$= -\frac{(-\lambda)^r}{(r-1)!} \frac{d^{r-1}}{d\lambda^{r-1}} \frac{1}{\lambda}$$

$$= -\frac{(-\lambda)^r}{(r-1)!} \underbrace{(-1)(-2)\cdots(1-r)}_{(-1)^{r-1}\,(r-1)!} \frac{1}{\lambda^r}$$

$$= 1$$

Thm: The Gamma distribution has expectation $E(X) = \frac{r}{\lambda}$ and variance $Var(X) = \frac{r}{\lambda^2}$.

$$\mu_X = \frac{\lambda^r}{(r-1)!} \int_0^\infty t^r e^{-\lambda t} \, dt = (-1)^r \frac{\lambda^r}{(r-1)!} \frac{d^r}{d\lambda^r} \int_0^\infty e^{-\lambda t} \, dt$$

$$= (-1)^r \frac{\lambda^r}{(r-1)!} \frac{d^r}{d\lambda^r} \frac{1}{\lambda}$$

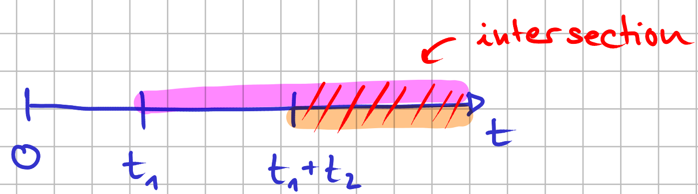
$$= \frac{r}{\lambda}$$

Similarily for $Var(X)$.

Proposition: As for the geometric distribution, the exponential distribution has "no memory", i.e. $P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2)$.

proof: $$P(X > t_1 + t_2 \mid X > t_1) = \frac{P((X > t_1 + t_2) \cap (X > t_1))}{P(X > t_1)}$$

$$= \frac{P(X > t_1 + t_2)}{P(X > t_1)}$$



$$= \frac{1 - F_X(t_1+t_2)}{1 - F_X(t_1)}$$

$$= \frac{1 - (1 - e^{-\lambda(t_1+t_2)})}{1 - (1 - e^{-\lambda t_1})}$$

$$= \frac{e^{-\lambda(t_1+t_2)}}{e^{-\lambda t_1}}$$

$$= e^{-\lambda t_2}$$

$$= 1 - (1 - e^{-\lambda t_2})$$

$$= 1 - P(X \leq t_2)$$

$$= P(X > t_2)$$

Exercise: Suppose the lifetime of a laptop measured in years is an exponentially distributed random variable with parameter $\lambda = \frac{1}{4}$. Find the probability that it lasts at least 5 years given that it has been used already 2 years in the past.

$$P(X > 2+5 \mid X > 2) = P(X > 5) = 1 - P(X \leq 5)$$

no memory

$$= 1 - (1 - e^{-5/4})$$

$$= e^{-5/4}$$

$$\approx 0{,}286$$

Exercise: Suppose a phone rings in average 5 times per hour. After switching it on,

1) what is the probability to have the first call between 5 and 8 minutes?

2) what is the probability to have the second call between 5 and 8 minutes?

i) $X \sim Exp(5)$

$$P\left(\frac{5}{60} < X < \frac{8}{60}\right) = \int_{5/60}^{8/60} 5e^{-5t}\,dt = \left[-e^{-5t}\right]_{5/60}^{8/60}$$

$$= e^{-5/12} - e^{-8/12}$$

$$\approx 0{,}146$$

ii) $X \sim Gamma(5,2)$

$$P\left(\frac{5}{60} < X < \frac{8}{60}\right) = \int_{5/60}^{8/60} 25t\,e^{-5t}\,dt$$

integration by parts

$$= 25\left[-t\,\frac{1}{5}e^{-5t}\right]_{5/60}^{8/60} - 25\int_{5/60}^{8/60}\left(-\frac{1}{5}\right)e^{-5t}\,dt$$

$$= 5\left(\frac{5}{60}e^{-\frac{5}{12}} - \frac{8}{60}e^{-8/12}\right) + 5\left[-\frac{1}{5}e^{-5t}\right]_{5/60}^{8/60}$$

$$= \frac{17}{12}e^{-5/12} - \frac{5}{3}e^{-8/12}$$

$$\approx 0{,}078$$

4.3 Normal (Gaussian) distribution

Def: A random variable X is Gaussian or normal distributed with parameters $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ if its probability density reads

$$\boxed{f_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}}$$

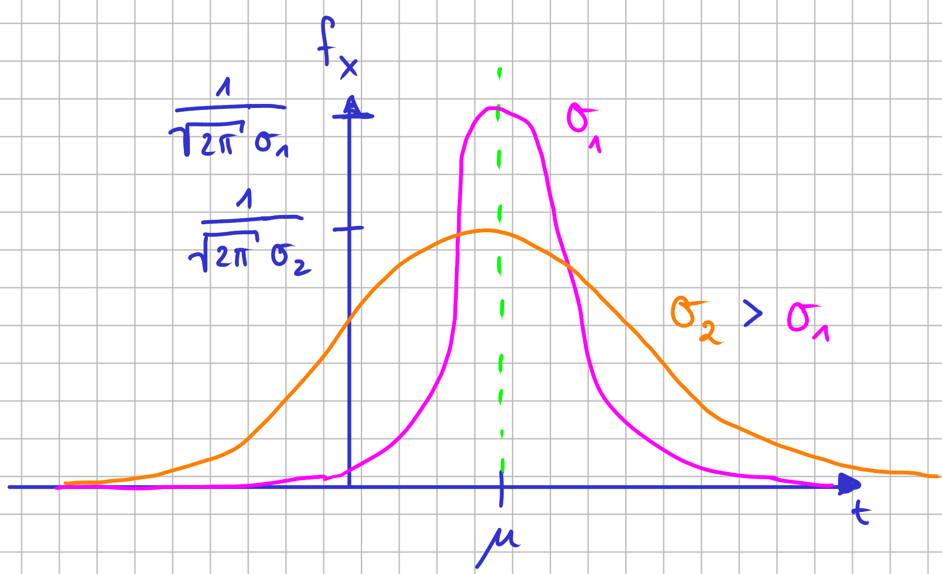
It is denoted as $X \sim N(\mu, \sigma^2)$.

Observations: i) f_X is symmetric around μ, i.e.

$$f_X(\mu + a) = f_X(\mu - a)$$

ii) the maximal value is $f_X(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma}$

iii) f_X is of bell shape



Proposition : We have

$$\int_{-\infty}^{+\infty} f_X(t)\,dt = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{+\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt = 1$$

Proof:

1) With $t = \sqrt{2}\,\sigma s + \mu$ we have

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt = \sqrt{2}\,\sigma \underbrace{\int_{-\infty}^{+\infty} \exp\left(-s^2\right) ds}_{I_1}$$

2) Consider the integral

$$I_2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(-(s^2+t^2)\right) ds\,dt$$

$e^{a+b} = e^a e^b$

$$= \int_{-\infty}^{+\infty} \exp(-s^2)\,ds \cdot \int_{-\infty}^{+\infty} \exp(-t^2)\,dt$$

$$= I_1 \cdot I_1$$

3) Calculate I_2 in polar coordinates

$s = \rho \cos(\theta)$

$t = \rho \sin(\theta)$

with $0 \leq \rho \leq \infty$ and $0 \leq \theta \leq 2\pi$.

It follows that

$$s^2 + t^2 = \rho^2 \overbrace{(\cos^2\theta + \sin^2\theta)}^{=1} = \rho^2$$

Jacobi determinant

$$\begin{vmatrix} \frac{\partial s}{\partial \varrho} & \frac{\partial s}{\partial \theta} \\ \frac{\partial t}{\partial \varrho} & \frac{\partial t}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -\varrho\sin(\theta) \\ \sin\theta & \varrho\cos(\theta) \end{vmatrix} = \varrho \underbrace{\left(\cos^2\theta + \sin^2\theta\right)}_{1} = \varrho$$

So we can write

$$I_2 = \int_0^{2\pi} d\theta \int_0^{\infty} \varrho\, e^{-\varrho^2} = 2\pi \left[-\frac{1}{2} e^{-\varrho^2} \right]_0^{\infty} = \pi$$

It follows that

$$I_1 = \int_{-\infty}^{+\infty} \exp(-s^2)\, ds = \sqrt{\pi}$$

✓

Thm.: The normal distribution has

- Distribution function

$$F_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{t} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2} \right) ds$$

- Expectation value

$$E(X) = \mu$$

- Variance

$$Var(X) = \sigma^2$$

- Moments

$$E((X-\mu)^k) = \begin{cases} 1 \cdot 3 \cdot \dots \cdot (k-1)\,\sigma^k & k \text{ even} \\ 0 & k \text{ odd} \end{cases}$$

Proof of expectation value:

$$E(X) = \int_{-\infty}^{+\infty} t\, f_X(t)\, dt = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} t\, e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

$t = \sqrt{2}\,\sigma s + \mu$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} (\sqrt{2}\,\sigma s + \mu)\, e^{-s^2}\, ds$$

$$= \frac{\sqrt{2}\,\sigma}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{+\infty} s e^{-s^2}\, ds}_{=0} + \frac{\mu}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-s^2}\, ds}_{\sqrt{\pi}}$$

antisymmetric

$$= \mu$$

To derive the variance we use integration by parts.

Remark: Distribution function F_X needs to be evaluated on a computer. The values in standard variables $Z = \frac{X - \mu_X}{\sigma_X}$, i.e. $N(0,1)$ with

$$\boxed{F_Z(t) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\tfrac{1}{2}s^2\right) ds}$$

can be found in the literature. $\Phi(t)$ is used to obtain the distribution function for $N(\mu, \sigma^2)$. We have $F_X(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$ since

$$F_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{t} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds$$

$\frac{s-\mu}{\sigma} = u$

$ds = \sigma\, du$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t-\mu}{\sigma}} \exp\left(-\tfrac{1}{2}u^2\right) du$$

$$= \Phi\left(\frac{t-\mu}{\sigma}\right).$$

Proposition: $\Phi(z) + \Phi(-z) = 1$

Proof:

$$1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\tfrac{1}{2}u^2\right) du$$

$$= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{z} \exp\left(-\tfrac{1}{2}u^2\right) du + \underbrace{\int_{z}^{+\infty} \exp\left(-\tfrac{1}{2}u^2\right) du}_{u=-v} \right)$$

$$-\int_{-z}^{-\infty} \exp\left(-\tfrac{1}{2}v^2\right) dv$$

interchange boundaries

$$= \Phi(z) + \Phi(-z)$$

Observation: For $X \sim N(\mu, \sigma^2)$ it holds that

$$P(\mu - \alpha\sigma < X < \mu + \alpha\sigma) = F_X(\mu + \alpha\sigma) - F(\mu - \alpha\sigma)$$

$$= \Phi\left(\frac{\mu + \alpha\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - \alpha\sigma - \mu}{\sigma}\right)$$

$$= \Phi(\alpha) - \Phi(-\alpha)$$

$\alpha = 1$: $\Phi(1) - \Phi(-1) = \Phi(1) - (1 - \phi(1))$

$$= 2\,\Phi(1) - 1$$

$$\approx 2 \cdot 0{,}84134 - 1$$

$$\approx 0{,}68$$

$\alpha = 2: \quad \Phi(2) - \Phi(-2) \approx 0{,}955$

$\alpha = 3: \quad \Phi(3) - \Phi(-3) \approx 0{,}997$

$\Rightarrow$ 3σ - rule:

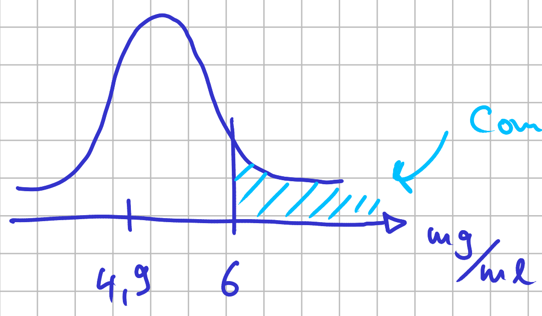
Nearly all values of X are within $\mu - 3\sigma$ and $\mu + 3\sigma$.

Example: find $\Phi(0{,}98) = 0{,}83646$ in the table

Exercise: The concentration of sugar in a solution should be around 4.9 mg/ml. Find the probability that the concentration has above 6,0 mg/ml of sugar given that it is normally distributed with

a) expectation 4,9 and variance 0,36

b) expectation 5,2 and standard deviation 0,4



$$P(X \geq 6) = 1 - P(X \leq 6) = 1 - F_X(6) = 1 - \Phi\left(\frac{6-\mu}{\sigma}\right)$$

a) $P(X \geq 6) = 1 - \Phi\left(\frac{1{,}1}{0{,}6}\right) \approx 1 - \Phi(1{,}83) \approx 1 - 0{,}966 = 0{,}034$

b) $P(X \geq 6) = 1 - \Phi\left(\frac{0{,}8}{0{,}4}\right) = 1 - \Phi(2) \approx 1 - 0{,}977 = 0{,}023$

4.4 Pareto distribution

Def: A random variable X is Pareto distributed, $X \sim Par(\alpha)$, with parameter $\alpha > 0$, if it has density function

$$f_X(t) = \begin{cases} \dfrac{\alpha}{t^{\alpha+1}} & \text{for } t \geq 1 \\ 0 & \text{else} \end{cases}$$

and distribution function

$$F_X(t) = \begin{cases} 0 & \text{for } t < 1 \\ \displaystyle\int_1^t \frac{\alpha}{s^{\alpha+1}}\,ds = \left[-s^{-\alpha}\right]_1^t = 1 - t^{-\alpha} & \text{for } t \geq 1 \end{cases}$$

Observation:

$$\int_{-\infty}^{+\infty} f_X(t)\,dt = \int_1^{\infty} \frac{\alpha}{t^{\alpha+1}}\,dt = \left[-t^{-\alpha}\right]_1^{\infty} = 1$$

Proposition: The mean and variance of a Pareto r.v. are given by

$$\mu_X = \frac{\alpha}{\alpha-1} \qquad \text{for } \alpha > 1$$

and

$$\sigma_X^2 = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} \qquad \text{for } \alpha > 2$$

i) $$\mu_X = \int_{-\infty}^{+\infty} t \cdot f_X(t)\, dt = \int_1^{\infty} \alpha\, t^{-\alpha}\, dt = \alpha \left[\frac{1}{1-\alpha} t^{1-\alpha} \right]_1^{\infty} \overset{\alpha > 1}{=} \frac{\alpha}{\alpha - 1}$$

ii) $$\sigma_X^2 = \int_{-\infty}^{+\infty} (t-\mu_X)^2 f_X(t)\, dt = \int_1^{\infty} (t^2 - 2\mu_X t + \mu_X^2) \frac{\alpha}{t^{\alpha+1}}\, dt$$

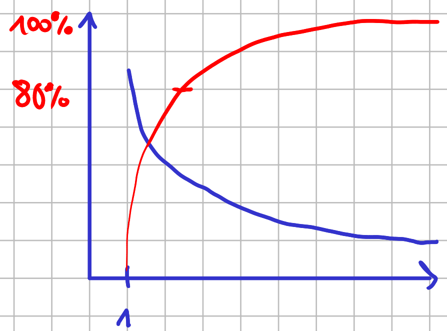
$$= \alpha \int_1^{\infty} t^{1-\alpha}\, dt - 2\mu_X^2 + \mu_X^2$$

$$= \alpha \left[\frac{1}{2-\alpha} t^{2-\alpha} \right]_1^{\infty} - \mu_X^2$$

$$\overset{\alpha > 2}{=} - \frac{\alpha}{2-\alpha} - \left(\frac{\alpha}{\alpha - 1} \right)^2$$

$$= \frac{\alpha}{(\alpha-2)(\alpha-1)^2}$$

- The Pareto distribution has a "heavy tail"



$\leadsto$ mean diverges for $0 < \alpha < 1$

- Pareto principle : "80% of outcome is due to 20% of causes"

$\leadsto$ 80% of Italy is owned by 20% of population

Exercise

Suppose that the turnover of the companies in France (in millons of Euro) can be described by a r.v. $X \sim Par(\alpha = 2)$.

What percentage of companies has a turnover of more than 3 million Euro?

$$P(X \geq 3) = \int_3^\infty \frac{2}{t^3}\, dt = 2\left[-\frac{1}{2}t^{-2}\right]_3^\infty = \frac{1}{9} \approx 0.11$$

$f_X(1) = 2$

$f_X(2) = 1/4$

$f_X(3) = 2/27 \approx 0.074$

Exercise

Compute the 1st 2nd & 3rd quartile for $X \sim Par(\alpha = 2)$

$$F_X(t) = 1 - t^{-\alpha}$$

$$\Rightarrow \quad F_X(t) = \gamma \iff 1 - \gamma = t^{-\alpha} \Rightarrow t = \frac{1}{(1-\gamma)^{1/\alpha}}$$

$\gamma = \frac{1}{4} \quad \Rightarrow \quad t_{1/4} = \frac{2}{\sqrt{3}} \approx 1.15$

$\gamma = 1/2 \quad \Rightarrow \quad t_{1/2} = \sqrt{2} \approx 1.41$

$\gamma = 3/4 \quad \Rightarrow \quad t_{3/4} = 2$

5. Multiple Random Variables

On one sample space we can define several random variables, e.g. height, weight and age of a person or the result of a sequence of coin flips, denoted by a random vector $\vec{X} = (X_1, \dots, X_n)^T$.

sample space　class of events

Def: Given a probability space $(\Omega, \mathcal{F}, P)$ we call the vector

$$\vec{X} = \vec{X}(a) : \Omega \to \mathbb{R}^n, \quad \text{with} \quad a \in \Omega$$

multi-dimensional random variable if for $s_i, t_i \in \mathbb{R}$ with $s_i < t_i$ any set $\{a \mid s_i < X_i(a) \leq t_i\} \subseteq \Omega$.

For simplicity of the presentation we focus on $n=2$. The generalisation to arbitrary n works accordingly.

5.1 Discrete case

Def: A two-dimensional random variable is discrete if there exists a discrete set of points $(s_i, t_j) \in \mathbb{R}^2$ with $i,j \in \mathbb{N}$ such that

i) $P(X_1 = s_i, X_2 = t_j) > 0 \quad \forall \; i,j$

ii) $\sum_{i,j} P(X_1 = s_i, X_2 = t_j) = 1$

Def: The joint probability mass function is defined as

$$f_{\vec{X}}(s,t) = \begin{cases} P(X_1 = s_i, X_2 = t_j) & \text{if} \quad (s,t) = (s_i, t_j) \\ 0 & \text{else} \end{cases}$$

Def: The joint distribution function reads

$$F_{X_1,X_2}(s,t) = \sum_{\{(s',t') \mid s' \leq s \text{ and } t' \leq t\}}' f_{X_1,X_2}(s',t')$$

Def: The marginal mass functions are defined as

$$f_{X_1}(s) = \sum_j' f_{X_1,X_2}(s, t_j)$$

and $$f_{X_2}(t) = \sum_i' f_{X_1,X_2}(s_i, t)$$

Def: The marginal distribution functions are defined via

$$F_{X_1}(s) = \sum_{s' \leq s}' f_{X_1}(s')$$

and $$F_{X_2}(t) = \sum_{t' \leq t}' f_{X_2}(t')$$

Example :

Consider the mass function $f_{X,Y}(s,t)$ with non-vanishing values

$f(2,5) = 0{,}20$ $\quad$ $f(2,9) = 0{,}05$ $\quad$ $f(2,12) = 0{,}35$

$f(4,5) = 0{,}15$ $\quad$ $f(4,9) = 0{,}20$ $\quad$ $f(4,12) = 0{,}05$

Equivalent representation as table

X \ Y	5	9	12
2	0,20	0,05	0,35
4	0,15	0,20	0,05

The marginal mass functions are

$$f_X(s) = f(s,5) + f(s,9) + f(s,12)$$

$$= \begin{cases} 0{,}6 & \text{if } s=2 \\ 0{,}4 & \text{if } s=4 \\ 0 & \text{else} \end{cases}$$

$$X : \begin{pmatrix} 2 & 4 \\ 0{,}6 & 0{,}4 \end{pmatrix} \qquad Y : \begin{pmatrix} 5 & 9 & 12 \\ 0{,}35 & 0{,}25 & 0{,}40 \end{pmatrix}$$

Thus we verify that

$$\sum_{i,j}' f_{X,Y}(s_i, t_j) = \sum_i' f_X(s_i) = \sum_j' f_Y(t_j) = 1$$

Exercise: compute $F_{X,Y}(3,10)$

$$F_{X,Y}(3,10) = F_{X,Y}(2,9) = f_{X,Y}(2,5) + f_{X,Y}(2,9)$$

$$= 0{,}25$$

Exercise : We throw two dice with six sides : two sides show 1, two sides 2 and two sides 3. Consider the random variables

X = "Sum of the resulting numbers"
Y = "Product of the resulting numbers"

i) What is the joint mass function ?
ii) What are the marginal mass functions?
iii) What is the probability for a) $P(X, Y \leq 4)$
b) $P(X, Y \geq 4)$
c) $P(X-Y = 1)$

i) $\Omega = \{1,2,3\}^2$
$|\Omega| = 9$

X \ Y	1	2	3	4	6	9
2	1/9	0	0	0	0	0
3	0	2/9	0	0	0	0
4	0	0	2/9	1/9	0	0
5	0	0	0	0	2/9	0
6	0	0	0	0	0	1/9

(1,1)
(1,2) (2,1)
(1,3) (3,1)
(2,2)
(2,3) (3,2)
(3,3)

ii)

$$\begin{array}{ccccc} (1,1) & (1,2),(2,1) & (1,3),(2,2),(3,1) & & \end{array}$$

$$X : \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 1/9 & 2/9 & 3/9 & 2/9 & 1/9 \end{pmatrix}$$

$$Y : \begin{pmatrix} 1 & 2 & 3 & 4 & 6 & 9 \\ 1/9 & 2/9 & 2/9 & 1/9 & 2/9 & 1/9 \end{pmatrix}$$

This is what we expect.

iii) a) $X, Y \leq 4$

$X \backslash Y$	1	2	3	4	6	9
2	$1/9$	0	0	0	0	0
3	0	$2/9$	0	0	0	0
4	0	0	$2/9$	$1/9$	0	0
5	0	0	0	0	$2/9$	0
6	0	0	0	0	0	$1/9$

$P(X \leq 4, Y \leq 4) = \frac{2}{3}$

b) $X, Y \geq 4$

$X \backslash Y$	1	2	3	4	6	9
2	$1/9$	0	0	0	0	0
3	0	$2/9$	0	0	0	0
4	0	0	$2/9$	$1/9$	0	0
5	0	0	0	0	$2/9$	0
6	0	0	0	0	0	$1/9$

$P(X \geq 4, Y \geq 4) = 4/9$

c) $X - Y = 1$

$X \backslash Y$	1	2	3	4	6	9
2	$1/9$	0	0	0	0	0
3	0	$2/9$	0	0	0	0
4	0	0	$2/9$	$1/9$	0	0
5	0	0	0	0	$2/9$	0
6	0	0	0	0	0	$1/9$

$P(X - Y = 1) = 5/9$

5.2 Continuous case

Def: A two-dimensional random variable $\vec{X} = (X_1, X_2)$ is continuous if there exists a joint probability density f_{X_1,X_2} with $f_{X_1,X_2}(s,t) \geq 0 \quad \forall\, s,t \in \mathbb{R}$ such that

$$\int_{-\infty}^{+\infty} ds \int_{-\infty}^{+\infty} dt\; f_{X_1,X_2}(s,t) = 1$$

The joint probability for a given set $B \subseteq \mathbb{R}^2$ is then computed via

$$P(\vec{X} \in B) = \int_B f_{X_1,X_2}(s,t)\, ds\, dt$$

Def: The joint distribution function reads

$$F_{X_1,X_2}(s,t) = \int_{-\infty}^{s} ds' \int_{-\infty}^{t} dt' \; f_{X_1,X_2}(s',t')$$

where

$$f_{X_1,X_2}(s,t) = \frac{\partial^2}{\partial s\, \partial t} F_{X_1,X_2}(s,t)$$

Def: The marginal probability density and distribution functions are defined via

$$f_{X_1}(s) = \int_{-\infty}^{+\infty} dt\, f_{X_1,X_2}(s,t) \qquad f_{X_2}(t) = \int_{-\infty}^{+\infty} ds\, f_{X_1,X_2}(s,t)$$

and

$$F_{X_1}(s) = \int_{-\infty}^{s} ds' \; f_{X_1}(s') \qquad F_{X_2}(t) = \int_{-\infty}^{t} dt' \; f_{X_2}(t')$$

Remark: All definitions extend to the n-dimensional case $\vec{X} = (X_1, \dots, X_n)$.

Example Consider the density function

$$f_{X,Y}(s,t) = \begin{cases} s \cdot t \cdot \exp\left(-\frac{1}{2}(s^2+t^2)\right) & \text{se } s,t \geq 0 \\ 0 & \text{else} \end{cases}$$

a) compute the marginal denseties

b) show that $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(s,t)\, ds\, dt = 1$

c) compute $P(X+Y \leq 1)$

a) $$f_X(s) = \int_{-\infty}^{+\infty} f_{X,Y}(s,t)\, dt = s e^{-\frac{1}{2}s^2} \int_0^{+\infty} t\, e^{-\frac{1}{2}t^2} dt$$

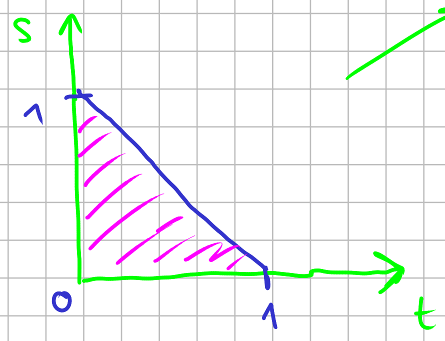
$$= s e^{-\frac{1}{2}s^2} \left[- e^{-\frac{1}{2}t^2}\right]_0^{+\infty}$$

$$= s\, e^{-\frac{1}{2}s^2}$$

$$f_Y(t) = t\, e^{-1/2\, t^2}$$

b) $$\int_{-\infty}^{+\infty} ds\, f_X(s)\, ds = \int_0^{\infty} ds\; s\, e^{-1/2\, s^2} = \left[- e^{-1/2\, s^2}\right]_0^{\infty} = 1$$

c) $$P(X+Y \leq 1) = \int_{\{(s,t) \mid s+t \leq 1\}} f_{X,Y}(s,t)\, ds\, dt$$



$$= \int_0^1 ds\, s e^{-\frac{1}{2}s^2} \int_0^{1-s} dt\, t e^{-\frac{1}{2}t^2}$$

$$= \int_0^1 ds\, s e^{-\frac{1}{2}s^2} \left[-e^{-\frac{1}{2}t^2} \right]_0^{1-s}$$

$$= \int_0^1 ds \left(1 - e^{-\frac{1}{2}(1-s)^2}\right) s e^{-\frac{1}{2}s^2}$$

$$= \int_0^1 ds\, s e^{-\frac{1}{2}s^2} - \int_0^1 ds\, s\, e^{-\frac{1}{2}(1-2s+2s^2)}$$

$$= \left[-e^{-1/2\, s^2} \right]_0^1 - e^{-1/4} \int_0^1 ds\, s\, e^{-(s-\frac{1}{2})^2}$$

$s = \frac{1}{\sqrt{2}} u + \frac{1}{2}$

$ds = \frac{1}{\sqrt{2}} du$

$$= 1 - 1/\sqrt{e} - \frac{e^{-1/4}}{\sqrt{2}} \int_{-1/\sqrt{2}}^{1/\sqrt{2}} \left(\cancel{\frac{1}{\sqrt{2}} u} + \frac{1}{2} \right) e^{-\frac{1}{2}u^2} du$$

antisymmetric
(odd fct.)

$$= 1 - 1/\sqrt{e} - \frac{\sqrt{\pi}}{2 e^{1/4}} \underbrace{\left(\Phi(1/\sqrt{2}) - \Phi(-1/\sqrt{2}) \right)}_{2\Phi(1/\sqrt{2}) - 1}$$

$\approx 0{,}0342$

5.3 Independence and Covariance

Def: Two continuous random variables X_1 and X_2 are independent if

$$f_{X_1,X_2}(s,t) = f_{X_1}(s)\, f_{X_2}(t)$$

Discrete random variables are independent if

$$P(X_1 = s_i, X_2 = t_j) = P(X_1 = s_i)\, P(X_2 = t_j) \quad \forall\ i,j$$

Def: The expectation value is computed via

$$E(g(X_1,X_2)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(s,t)\, f_{X_1,X_2}(s,t)\, ds\, dt$$

Def: The covariance of two random variables X_1, X_2 is defined as

$$\begin{aligned} Cov(X_1,X_2) &= E\big((X_1 - E(X_1))(X_2 - E(X_2))\big) \\ &= E((X_1 - \mu_{X_1})(X_2 - \mu_{X_2})) \\ \text{E is linear} \searrow \quad &= E(X_1 X_2) - \mu_{X_1}\mu_{X_2} \end{aligned}$$

Corollary: For indipendent X_1 & X_2 we have

$$\begin{aligned} E(X_1 X_2) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} s \cdot t\, f_{X_1}(s)\, f_{X_2}(t)\, ds\, dt \\ &= \mu_{X_1} \cdot \mu_{X_2} \end{aligned}$$

and thus $Cov(X_1, X_2) = 0$.

Proposition: For any two random variables X & Y we have

i) $E(X+Y) = E(X) + E(Y)$

ii) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X,Y)$

if expectation and variance exist.

Proof: i)

$$E(X+Y) = \iint ds\,dt\,(s+t)\, f_{X,Y}(s,t)$$

$$= \iint ds\,dt\; s\, f_{X,Y}(s,t) + \iint ds\,dt\; t\, f_{X,Y}(s,t)$$

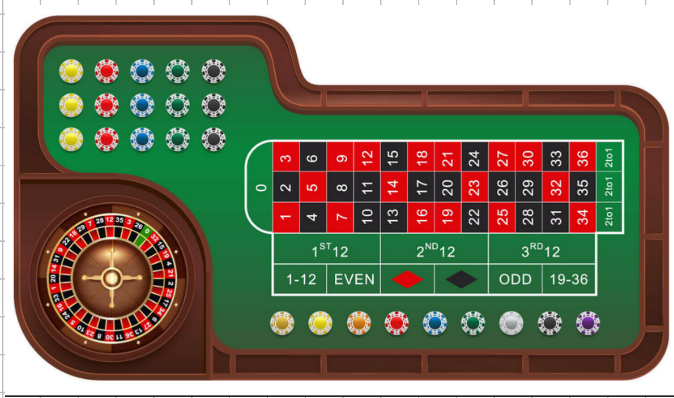
$$= E(X) + E(Y)$$

ii)
$$\mathrm{Var}(X+Y) = \iint ds\,dt\,\big(s+t-\underbrace{E(X+Y)}_{E(X)+E(Y)}\big)^2 f_{X,Y}(s,t)$$

$$= \iint ds\,dt\Big[(s-E(X))^2 + (t-E(Y))^2$$
$$+ 2\,(s-E(X))(t-E(Y))\Big] f_{X,Y}(s,t)$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

Roulette :



The game of roulette yields a random number from 0 – 36. We can bet on the lower dozen $\{1,2,\dots,12\}$ and triple the stake (input) which results in a profit of $X = 2$ in case the outcome is in $\{1,2,\dots,12\}$. Else we lose our stake $X = -1$. The expectation is $E(X) = 2 \cdot \frac{12}{37} - 1 \cdot \frac{25}{37} = -\frac{1}{37}$. (Without "0" we would have $E(X) = 0$.) We also compute the variance

$$Var(X) = \left(2 + \frac{1}{37}\right)^2 \frac{12}{37} + \left(-1 + \frac{1}{37}\right)^2 \frac{25}{37} = 1{,}97.$$

Alternatively we can bet on even or odd with profit $Y = \pm 1$. If the outcome is "0" we loose half stake $Y = -\frac{1}{2}$. So $E(Y) = -\frac{1}{2} \cdot \frac{1}{37} + \frac{18}{37} - \frac{18}{37} = -\frac{1}{74}$ and $Var(Y) =$

$$\left(-\frac{1}{2} + \frac{1}{74}\right)^2 \frac{1}{37} + \left(1 + \frac{1}{74}\right)^2 \frac{18}{37} + \left(-1 + \frac{1}{74}\right)^2 \frac{18}{37} = 0{,}98$$

A player can also combine his bets and bet the same amount on both options $Z = X + Y$. We know $E(Z) = E(X) + E(Y)$ while $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$.

Compute joint probability function:

e.g. $P(X=2, Y=-1) = \frac{6}{37} = \frac{|\text{odd} \cap \{1,\dots,12\}|}{|\Omega|}$

$P(X=2, Y=-\frac{1}{2}) = 0$

$P(X=2, Y=1) = \frac{6}{37}$

$P(X=-1, Y=-1) = \frac{12}{37}$

$P(X=-1, Y=-\frac{1}{2}) = \frac{1}{37}$

$P(X=-1, Y=2) = \frac{12}{37}$

$P(X,Y)$	$Y=1$	$Y=-\frac{1}{2}$	$Y=-1$	
$X=2$	$\frac{6}{37}$	0	$\frac{6}{37}$	$P(X=2)=\frac{12}{37}$
$X=-1$	$\frac{12}{37}$	$\frac{1}{37}$	$\frac{12}{37}$	$P(X=-1)=\frac{25}{37}$
	$P(Y=1)=\frac{18}{37}$	$P(Y=-\frac{1}{2})=\frac{1}{37}$	$P(Y=-1)=\frac{18}{37}$	

$E(X\cdot Y) = \sum_{i,j} s_i t_j \, P(X=s_i, Y=t_j)$

$= 1\cdot 2 \frac{6}{37} - 1\cdot 2 \frac{6}{37} - 1 \frac{12}{37} + \frac{1}{2} \frac{1}{37} + 1 \frac{12}{37} = \frac{1}{74}$

$Cov(X,Y) = E(X\cdot Y) - E(X)E(Y)$

$= \frac{1}{74} - \frac{1}{74}\cdot\frac{1}{37} \approx 0{,}013$

$\Rightarrow Var(X+Y) = 1{,}97 + 0{,}98 + 2\cdot 0{,}013$

$\approx 2{,}98$

5.4 Cauchy-Schwarz inequality

Theorem: The covariance of two random variables X & Y satisfies the inequality

$$\boxed{\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)}$$

Proof: Consider the random variable $Z = (aX - bY)^2$ that is always non-negative for all values of X, Y and $a, b \in \mathbb{R}$. Thus also the mean μ_Z is non-negative

$$0 \leq E(Z) = E((aX-bY)^2) = a^2 E(X^2) - 2ab\,E(X\cdot Y) + b^2 E(Y^2)$$

Now we fix $a = E(Y^2)$ and $b = E(X\cdot Y)$. It follows

$$0 \leq E(Y^2)^2 E(X^2) - 2\,E(Y^2)\,E(X\cdot Y)^2 + E(X\cdot Y)^2 E(Y^2)$$

$$= \underbrace{E(Y^2)}_{\geq 0}\left[E(Y^2)E(X^2) - E(X\cdot Y)^2 \right]$$

$$\Rightarrow \quad 0 \leq E(Y^2)\,E(X^2) - E(X\cdot Y)^2$$

Then for $X \to X - \mu_X$ and $Y \to Y - \mu_X$ we get

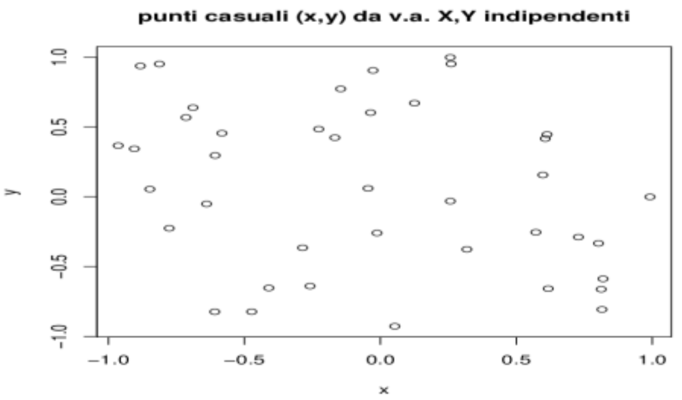
$$\mathrm{Cov}(X,Y)^2 = E((X-\mu_X)(Y-\mu_Y))^2 \leq E((X-\mu_X)^2)\,E((Y-\mu_Y)^2)$$

$$= \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

5.5 Linear correllations

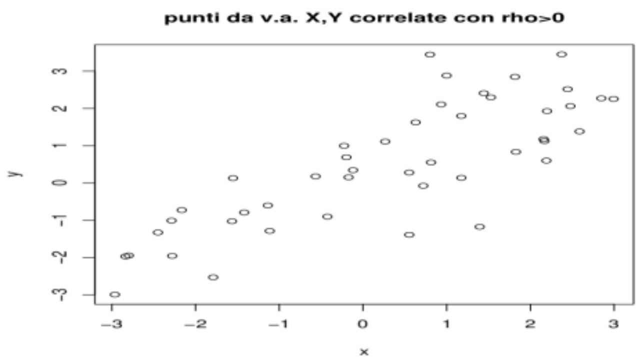
Motivation

Introduce correlation coefficient ρ to measure linear dependendence of two random variables

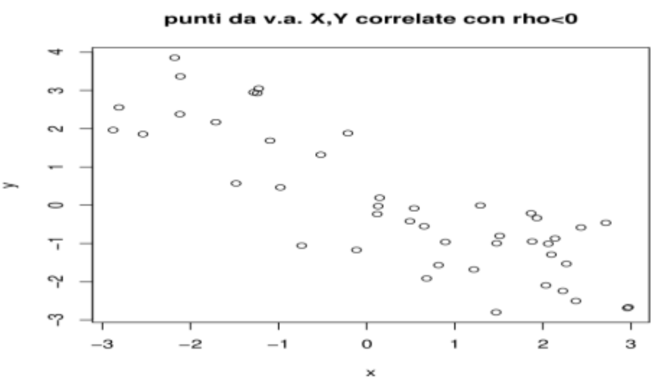


- Random distribution of points (x,y)

  $\rho = 0$



- Distribution of points with a linear trend $\rho > 0$



- Distribution of points with a (negative) linear trend $\rho < 0$

Def: The correlation coefficient is defined as

$$\rho_{X_1 X_2} = \rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)\, Var(X_2)}}$$

Observations: The correlation coefficient takes values in the intervall $\rho_{X_1 X_2} \in [-1, +1]$.

We say

i) X_1 & X_2 are uncorrelated if $\rho_{X_1 X_2} = 0$

ii) X_1 & X_2 are totally correlated if $\rho_{X_1 X_2} = \pm 1$

Relation to independence:

$$X_1, X_2 \text{ independent} \implies Cov(X_1, X_2) = 0$$

$$\Downarrow$$

$$X_1, X_2 \text{ uncorrelated} \iff \rho_{X_1 X_2} = 0$$

Observation: If $\rho_{X_1 X_2} = \pm 1$, the random variables X_1 and X_2 are linearly dependent, i.e

$$\exists\ c_1, c_2 \in \mathbb{R} \quad \text{such that} \quad X_2 = c_1 X_1 + c_2$$

Proof: We know that $\mu_{X_2} = c_1 \mu_{X_1} + c_2$ and $\sigma_{X_2}^2 = c_1^2 \sigma_{X_1}^2$

$$\Rightarrow \sigma_{X_2} = |c_1|\, \sigma_{X_1}$$

So $$\rho_{X_1 X_2} = \frac{Cov(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}$$

$$= \frac{1}{\sigma_{X_1}\sigma_{X_2}} \left(E(X_1 \cdot X_2) - \mu_{X_1}\mu_{X_2} \right)$$

$$= \frac{1}{|c_1| \sigma_{X_1}^2} \left(c_1 E(X_1^2) + \cancel{c_2 \mu_{X_1}} - c_1 \mu_{X_1}^2 - \cancel{c_2 \mu_{X_1}} \right)$$

$$= \frac{c_1}{|c_1|} \frac{1}{\sigma_{X_1}^2} \left(\underbrace{E(X_1^2) - \mu_{X_1}^2}_{\sigma_{X_1}^2} \right)$$

$$= sgn(c_1)$$

$$= \begin{cases} 1 & \text{if} \quad c_1 > 0 \\ -1 & \text{if} \quad c_1 < 0 \end{cases}$$

Exercise : We throw two dice with six sides : two sides show 1, two sides 2 and two sides 3. Consider the random variables

X = "Sum of the resulting numbers"

Y = "Product of the resulting numbers"

i) compute the covariance $Cov(X, Y)$

ii) compute the correlation coefficient

$$\begin{array}{c|cccccc}
X \backslash Y & 1 & 2 & 3 & 4 & 6 & 9 \\
\hline
2 & 1/9 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 2/9 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 2/9 & 1/9 & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 2/9 & 0 \\
6 & 0 & 0 & 0 & 0 & 0 & 1/9
\end{array}$$

$$X : \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 1/9 & 2/9 & 3/9 & 2/9 & 1/9 \end{pmatrix} \qquad Y : \begin{pmatrix} 1 & 2 & 3 & 4 & 6 & 9 \\ 1/9 & 2/9 & 2/9 & 1/9 & 2/9 & 1/9 \end{pmatrix}$$

i) $\mathrm{Cov}(X,Y) = E(X \cdot Y) - \mu_X \mu_Y$

$$\mu_X = \frac{2}{9} + \frac{6}{9} + \frac{12}{9} + \frac{10}{9} + \frac{6}{9} = \frac{36}{9} = 4$$

$$\mu_Y = \frac{1}{9} + \frac{4}{9} + \frac{6}{9} + \frac{4}{9} + \frac{12}{9} + 1 = \frac{27}{9} + 1 = 4$$

$$E(X \cdot Y) = 1 \cdot 2 \, \frac{1}{9} + 2 \cdot 3 \, \frac{2}{9} + 4 \cdot 3 \, \frac{2}{9} + 4 \cdot 4 \cdot \frac{1}{9} + 5 \cdot 6 \cdot \frac{2}{9} + 6 \cdot 9 \, \frac{1}{9}$$

$$= \frac{2}{9} + \frac{12}{9} + \frac{24}{9} + \frac{16}{9} + \frac{60}{9} + 6$$

$$= \frac{114}{9} + 6$$

$$= 18 + \frac{2}{3}$$

$$\approx 18{,}67$$

$$\rightarrow \mathrm{Cov}(X,Y) \approx 18{,}67 - 4 \cdot 4 = 2{,}67$$

ii) $\rho_{X,Y} = \dfrac{\mathrm{cov}(X,Y)}{\sigma_X \, \sigma_Y}$

$$\sigma_X^2 = \sum_i (a_i - \mu_X)^2 f_X(a_i) = (2-4)^2 \tfrac{1}{9} + (3-4)^2 \tfrac{2}{9} + (4-4)^2 \tfrac{3}{9} + (5-4)^2 \tfrac{2}{9} + (6-4)^2 \tfrac{1}{9}$$

$$= \frac{4}{9} + \frac{2}{9} + \frac{2}{9} + \frac{4}{9}$$

$$= \frac{12}{9} = 4/3 \approx 1{,}34$$

$$\sigma_Y^2 = (1-4)^2 \tfrac{1}{9} + (2-4)^2 \tfrac{2}{9} + (3-4)^2 \tfrac{2}{9} + (4-4)^2 \tfrac{1}{9} + (6-4)^2 \tfrac{2}{9} + (9-4)^2 \tfrac{1}{9}$$

$$= 1 + \frac{8}{9} + \frac{2}{9} + \frac{8}{9} + \frac{25}{9}$$

$$= 1 + \frac{43}{9}$$

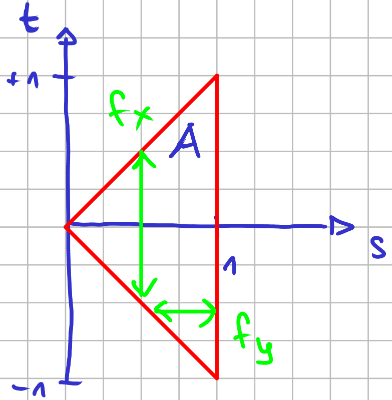
$$\approx 5{,}78$$

$$\Rightarrow \quad \rho_{X,Y} \approx \frac{2{,}67}{\sqrt{5{,}78}\,\sqrt{1{,}34}} \approx 0{,}96$$

Esempio: (uncorrelated but dependent)

Let

$$f_{X,Y}(s,t) = \begin{cases} 1 & \text{for} \quad 0 \le s \le 1 \text{ and } -s \le t \le s \\ 0 & \text{else} \end{cases}$$



$$A = \{(s,t) \mid -s \le t \le +s \text{ and } 0 \le s \le 1\}$$
$$= \{(s,t) \mid |t| \le s \le 1 \text{ and } -1 \le t \le +1\}$$

Marginal densities :
$$f_X(s) = \int_{-s}^{s} dt = 2s \quad \text{for} \quad 0 \le s \le 1$$
$$f_Y(t) = \int_{|t|}^{1} ds = 1 - |t| \quad \text{for} \quad -1 \le t \le +1$$

$$\mu_X = 2\int_0^1 s^2\, ds = \frac{2}{3}$$

$$\mu_Y = \int_{-1}^{+1} t(1-|t|)\, dt = \overbrace{\int_{-1}^{0} t(1+t)\,dt + \int_0^1 t(1-t)\, dt}^{t \to -t} = 0$$

$$E(X \cdot Y) = \int_0^1 ds \int_{-s}^{+s} s \cdot t = \int_0^1 ds\, s\, \left[\tfrac{1}{2}t^2\right]_{-s}^{+s} = 0$$

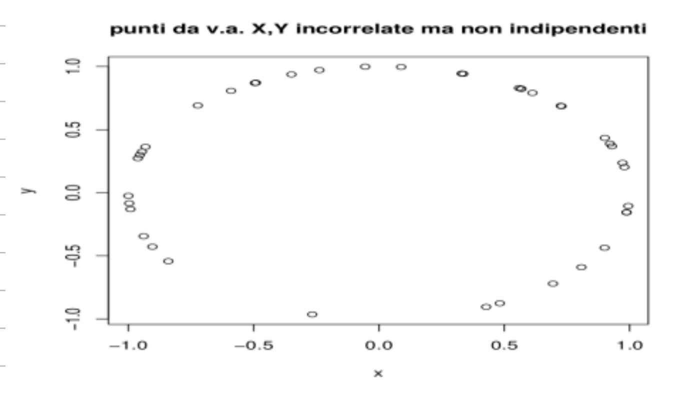
$$\Rightarrow \operatorname{Cov}(X,Y) = 0 \qquad \Rightarrow \rho_{X,Y} = 0$$

(Expected because of uniform distribution)

But X and Y are dependent because, e.g.

$$f_{X,Y}\left(1, \tfrac{1}{3}\right) = 1 \neq f_X(1)\, f_Y\left(\tfrac{1}{3}\right) = 2 \cdot \frac{2}{3} = \frac{4}{3}.$$

Expected since range of Y changes with X or vice versa.
For the square $-1 \le s,t \le +1$ we have indipendence.



· no linear trend but also not independent although $\rho = 0$

Generalizations for n-dimensions

Def: We define the vector of expectations of a random vector $\vec{X} = (X_1, \dots, X_n)^t$ via

$$\mu_{\vec{X}} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_{X_1} \\ \vdots \\ \mu_{X_n} \end{pmatrix}$$

Def: The covariance matrix $\Sigma_{\vec{X}}$ of a random vector $\vec{X} = (X_1, \dots, X_n)$ is an $n \times n$ matrix with entries

$$\boxed{(\Sigma_{\vec{X}})_{ij} = Cov(X_i, X_j)}$$

where $i, j = 1, \dots, n$.

In matrix form for the case $n=2$ we have

$$\Sigma_{\vec{X}} = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{pmatrix}$$

where $Cov(X, X) = Var(X)$.

Observation:

i) The matrix $\Sigma_{\vec{X}}$ is symmetric as $Cov(X_i, X_j) = Cov(X_j, X_i)$

ii) Correlations ρ_{ij} can be defined via $\rho_{ij} = \dfrac{Cov(X_i, X_j)}{\sigma_i \, \sigma_j}$

Proposition: A random vector $\vec{Y} = \vec{a} + B \cdot \vec{X}$ with an $n \times n$ matrix B and $\vec{a} \in \mathbb{R}^n$ has the vector of expectations

$$\boxed{E(\vec{Y}) = \vec{a} + B \cdot E(\vec{X})}$$

and covariance matrix

$$\boxed{\Sigma_{\vec{Y}} = B \cdot \Sigma_{\vec{X}} \cdot B^t}$$

Proof: i) $[E(\vec{Y})]_i = E(Y_i) = E\left(a_i + \sum_j B_{ij} X_j\right)$

$$= a_i + \sum_j B_{ij} E(X_j)$$

$$= [\vec{a} + B \cdot E(\vec{X})]_i$$

for all i.

ii) $[\Sigma_{\vec{Y}}]_{ij} = \mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}\left(\sum_k B_{ik} X_k, \sum_l B_{jl} X_l\right)$

$$= \sum_{k,l} B_{ik} B_{jl} \underbrace{\mathrm{Cov}(X_k, X_l)}_{[\Sigma_{\vec{X}}]_{kl}}$$

$$= \sum_{k,l} B_{ik} [\Sigma_{\vec{X}}]_{kl} B^t_{lj}$$

$$= [B \cdot \Sigma_{\vec{X}} \cdot B^t]_{ij}$$

Example Consider two random variables (X_1, X_2) with vanishing means, variances $\mathrm{Var}(X_1) = 10$ & $\mathrm{Var}(X_2) = 9$ and covariance $\mathrm{Cov}(X_1, X_2) = -3$. Find the covariance matrix of the random vector $\vec{Y} = (Y_1, Y_2)$ given by

$$Y_1 = X_1 + 5X_2$$
$$Y_2 = X_1 - X_2 \, .$$

We have $\vec{Y} = \begin{pmatrix} X_1 + 5X_2 \\ X_1 - X_2 \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = B \cdot \vec{X}$

and $\Sigma_{\vec{X}} = \begin{pmatrix} 10 & -3 \\ -3 & 9 \end{pmatrix}$

Thus

$$\Sigma_{\vec{Y}} = B \cdot \Sigma_{\vec{X}} \cdot B^t = \underbrace{\begin{pmatrix} 1 & 5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 10 & -3 \\ -3 & 9 \end{pmatrix}}_{\begin{pmatrix} 10 & 45 \\ 10 & -9 \end{pmatrix}} \begin{pmatrix} 1 & 1 \\ 5 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 235 & -35 \\ -35 & 19 \end{pmatrix}$$

<u>Remark</u>:

The proposition above also holds for $\vec{Y} = (Y_1, \dots, Y_m)$ and $\vec{X} = (X_1, \dots, X_n)$ with $m \leq n$ and $\vec{a} \in \mathbb{R}^m$ while B is an $m \times n$ matrix.

Proposition: Let $\vec{X}$ be an n-dimensional random vector with independent components $X_i \sim Ber(p)$ where $E(X_i) = p$ and $Var(X_i) = p(1-p)$.
Then, the sum $S_n = X_1 + \dots + X_n$ is a binomial r.v. with $S_n \sim Bin(n,p)$ where $E(S_n) = n \cdot p$ and $Var(S_n) = n \cdot p(1-p)$.

proof:

$$P(X_1 + \dots + X_n = k) = \sum_{\substack{k_i \in \{0,1\} \\ k_1 + \dots + k_n = k}} P(X_1 = k_1, \dots, X_n = k_n)$$

X_i independent

$$= \sum_{\substack{k_i \in \{0,1\} \\ k_1 + \dots + k_n = k}} P(X_1 = k_1) \cdots P(X_n = k_n)$$

$$= p^k (1-p)^{n-k} \underbrace{\sum_{\substack{k_i \in \{0,1\} \\ k_1 + \dots + k_n = k}} 1}_{= C(n,k) = \binom{n}{k}}$$

k times 1 from n

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

We verify that $E(S_n) = np$ & $Var(S_n) = np(1-p)$:

As $S_n = B_n \cdot \vec{X}$ with $B = (\underbrace{1, \dots, 1}_{n})$ we get

$$\mu_{S_n} = B_n \mu_{\vec{X}} = \sum_{i=1}^{n} \mu_{X_i} = n \cdot p$$

& $$Var(S_n) = B \cdot \Sigma_{\vec{X}} \cdot B^t = (1, \dots, 1) \begin{pmatrix} Var(X_1) & & 0 \\ & \ddots & \\ 0 & & Var(X_n) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$
$$= \sum_{i=1}^{n} Var(X_i) = n\,p(1-p)$$

5.4 Multivariant normal distribution

Consider case of two random variables:

Def: The joint probability density of two random variables defined as

$$f_{X,Y}(s,t) = \frac{1}{2\pi\,\sigma_X\sigma_Y\sqrt{1-\rho^2}}\; e^{-Q(s,t)}$$

with

$$Q(s,t) = \frac{1}{2(1-\rho^2)}\left[\left(\frac{s-\mu_X}{\sigma_X}\right)^2 + \left(\frac{t-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\,\frac{s-\mu_X}{\sigma_X}\,\frac{t-\mu_Y}{\sigma_Y}\right].$$

Then X,Y have a bivariant normal distribution.

Observation: For independent X & Y, i.e. $\rho=0$, the probability density factorizes into two densities of the normal distributions $N(\mu_X,\sigma_X^2)$ & $N(\mu_Y,\sigma_Y^2)$.

Def: The n-dimensional normally distributed random vector $\vec{X}=(X_1,\dots,X_n)$ has density

$$\boxed{f(s_1,\dots,s_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma_{\vec{X}}}}\exp\left(-\frac{1}{2}(\vec{s}-\vec{\mu})^T\cdot\Sigma_{\vec{X}}^{-1}\cdot(\vec{s}-\vec{\mu})\right)}$$

where $\vec{s}=(s_1,\dots,s_n)$, $\vec{\mu}=(\mu_{X_1},\dots,\mu_{X_n})$ and the $n\times n$ covariance matrix with entries $(\Sigma_{\vec{X}})_{ij} = \mathrm{cov}(X_i,X_j)$.

Remark: For $n=2$ we have

$$\Sigma_{\vec{X}} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\cdot\rho \\ \sigma_1\sigma_2\cdot\rho & \sigma_2^2 \end{pmatrix}$$

where we used $\rho_{1,2} = \dfrac{\operatorname{cov}(X_1, X_2)}{\sigma_1 \cdot \sigma_2}$.

It follows $\det \Sigma_{\vec{X}} = \sigma_1^2 \sigma_2^2 (1-\rho^2)$ and

$$\Sigma_{\vec{X}}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1 \sigma_2} \\ -\frac{\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

A direct computation shows that

$$Q(s_1, s_2) = \frac{1}{2} (\vec{s} - \vec{\mu})^t \cdot \Sigma_{\vec{X}}^{-1} \cdot (\vec{s} - \vec{\mu}).$$

Iso contours

consider $n=2$ & $\operatorname{Cov}(X_1, X_2) = 0$

$$f_{X_1 X_2}(s,t) = \frac{1}{2\pi \sigma_1 \sigma_2} \exp\left[-\frac{1}{2\sigma_1^2} (s-\mu_1)^2 - \frac{1}{2\sigma_2^2} (t-\mu_2)^2 \right]$$

Isocontours for $c \in \mathbb{R}_+$:

$$I = \{ (s,t) \in \mathbb{R}^2 \mid f_{X_1, X_2}(s,t) = c \}$$

$\Rightarrow$ set $f_{X_1, X_2}(s,t) = c$

$$\Leftrightarrow \quad 2\pi c \sigma_1 \sigma_2 = \exp\left[-\frac{1}{2\sigma_1^2} (s-\mu_1)^2 - \frac{1}{2\sigma_2^2} (t-\mu_2)^2 \right]$$

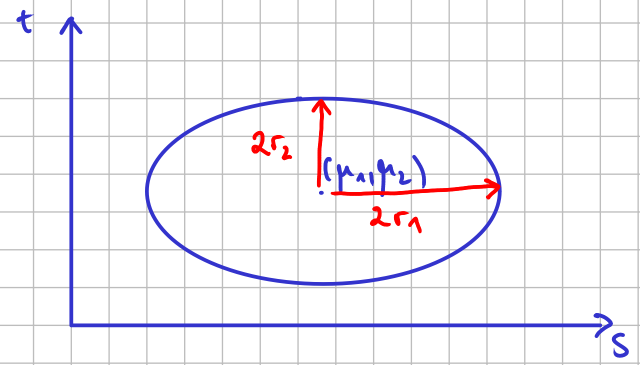
$$\ln \quad \Rightarrow \quad \ln(2\pi c \sigma_1 \sigma_2) = -\frac{1}{2} \left[\frac{1}{\sigma_1^2} (s-\mu_1)^2 + \frac{1}{\sigma_2^2} (t-\mu_2)^2 \right]$$

$$\Leftrightarrow \quad \ln\left(\frac{1}{2\pi \sigma_1 \sigma_2} \right) = \frac{1}{2} \left[\frac{1}{\sigma_1^2} (s-\mu_1)^2 + \frac{1}{\sigma_2^2} (t-\mu_2)^2 \right]$$

$$\Leftrightarrow \quad 1 = \frac{(s-\mu_1)^2}{r_1^2} + \frac{(t-\mu_2)^2}{r_2^2}$$

with $\quad r_i = \sigma_i \sqrt{2 \ln\left(\frac{1}{2\pi\sigma_1\sigma_2}\right)}$

$\rightsquigarrow$ axis-aligned ellipse with center (μ_1, μ_2)



Remark:

i) The smaller σ_i the smaller is r_i

ii) For $Cov(X_1, X_2) \neq 0$ the ellipse is rotated

ii) The n-dimensional counterpart of the ellipse is called ellipsoid.

6. Transformations of random variables

6.1 One single variable

Knowing the density f_X of r.v. X we like to obtain the density f_Y of r.v. $Y = g(X)$ for a given function g.

Indirect method

Proposition: For $Y = g(X)$ the distribution function F_Y is obtained from X as follows

$$F_Y(t) = P(Y \le t) = P(g(X) \in (-\infty, t])$$
$$= P(X \in g^{-1}((-\infty, t]))$$

with the preimage (or inverse image) $g^{-1}(B) = \{a \mid g(a) \in B\}$ of a set B.

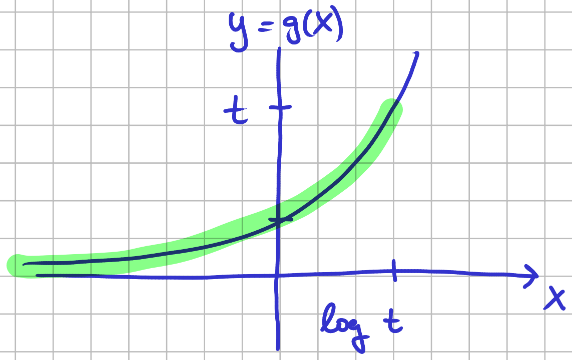
Example of preimage: take $g(x) = x^2$ & $B = \{4\}$

$$\Rightarrow \quad g^{-1}(\{4\}) = \{\pm 2\}$$

because $g(\pm 2) = 4$.

Remark: The density is then obtained from the distribution function via $f_Y(t) = \frac{d}{dt} F_Y(t)$.

Example 1) Let $Y = e^X$



$$g^{-1}(t) = \log(t)$$

preimage $g^{-1}((-\infty, t]) = (0, \log(t)]$

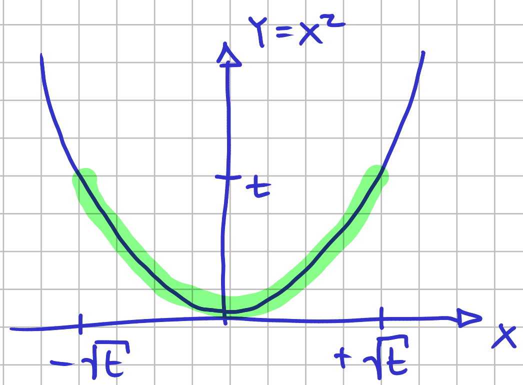
such that $F_Y(t) = P(Y \le t) = P(-\infty \le e^X \le t)$

$$= P(0 \le X \le \log t)$$

$$= F_X(\log t) - F_X(0)$$

$$\Rightarrow f_Y(t) = \frac{1}{t} f_X(\log t)$$

Example 2 Let $Y = X^2$



$$g^{-1}(t) = \pm\sqrt{t}$$

preimage: $g^{-1}([0, t]) = \begin{cases} (-\sqrt{t}, +\sqrt{t}) & \text{for } t \ge 0 \\ \emptyset & \text{for } t > 0 \end{cases}$

$\Rightarrow$ for $t \ge 0$

$$F_Y(t) = P(Y \le t) = P(X^2 \le t) = P(-\sqrt{t} \le X \le \sqrt{t})$$

$$= F_X(\sqrt{t}) - F_X(-\sqrt{t})$$

and $F_Y(t) = 0$ for $t < 0$.

$$f_Y(t) = \begin{cases} \frac{1}{2\sqrt{t}} \left(f_X(\sqrt{t}) + f_X(-\sqrt{t}) \right) & \text{for } t \ge 0 \\ 0 & \text{else} \end{cases}$$

Example 3) Let $X \sim \text{Unif}[-1,1]$. Obtain F_Y and f_Y for $Y = X^2$

$$g([-1,1]) = [0,1]$$

Recall $F_X(t) = \begin{cases} 1 & t > 1 \\ \frac{t+1}{2} & -1 \le t \le 1 \\ 0 & t < -1 \end{cases}$

$$F_Y(t) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) = \begin{cases} 1 & t > 1 \\ \frac{1}{2}(\sqrt{t}+1+\sqrt{t}-1) = \sqrt{t} & 0 \le t \le 1 \\ 0 & t < 0 \end{cases}$$

$$\& \quad f_Y(t) = \begin{cases} \frac{1}{2\sqrt{t}} & \text{for } 0 \le t \le 1 \\ 0 & \text{else} \end{cases}$$

Direct method

The direct method avoids the distribution function

$$\begin{tikzcd} F_X \arrow[r] & F_Y \arrow[d] \\ f_X \arrow[u] \arrow[r] & f_Y \end{tikzcd}$$

indirect

direct

Recall : Indicator function

$$I_A(x) = \begin{cases} 1 & \text{for} \quad x \in A \\ 0 & \text{for} \quad x \notin A \end{cases}$$

Theorem

i) Simple case

Let $U, V \subseteq \mathbb{R}$ and $g : U \to V$ be an invertible, differentiable and single valued function in U. For X with values in U and $Y = g(X)$ with values in V the density f_Y can be expressed via the density f_X as

$$\boxed{f_Y(t) = \left. \frac{f_X(s)}{|g'(s)|} \right|_{s = g^{-1}(t)} \cdot I_V(t)}$$

where $\quad g^{-1}(g(s)) = g(g^{-1}(s)) = s$.

Example 1) Let $Y = aX + b$ with $a \neq 0$ and $U, V = \mathbb{R}$

$$g^{-1}(t) = \frac{t-b}{a} \quad \text{such that}$$

$$g(g^{-1}(t)) = a\, g^{-1}(t) + b = a \frac{t-b}{a} + b = t$$

and $\quad g'(t) = a$

and thus

$$f_Y(t) = \frac{f_X(s)}{|g'(s)|}\Bigg|_{s=g^{-1}(t)} \cdot I_V(t)$$

$$= \frac{f_X\left(\frac{t-b}{a}\right)}{|a|} \quad I_V(t)$$

Example 2) Let $Y = e^X$. $\qquad U = \mathbb{R} \qquad V = \mathbb{R}_+$

$$\Rightarrow \quad g^{-1}(t) = \log(t) \quad \& \quad g'(t) = e^t$$

thus

$$f_Y(t) = \frac{f_X(s)}{|g'(s)|}\Bigg|_{s=g^{-1}(t)} \cdot I_V(t)$$

$$= \frac{f_X(\log(t))}{|t|} \quad I_{\mathbb{R}_+}(t)$$

$$= \frac{f_X(\log(t))}{t} \quad I_{\mathbb{R}_+}(t)$$

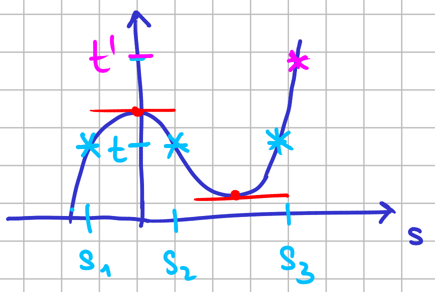
ii) general case

Let $g: U \to V$ not necessarily be invertible and single-valued. We denote by Θ_g the set of singular points of g given by

$$\Theta_g = \{ s \in U \mid \nexists\ g'(s) \text{ or } g'(s) = 0 \}$$

Further for fixed t, we call $m(t)$ the number of points $s_1, \dots, s_{m(t)}$ that are mapped to the same value $g(s_i) = t$.

E.g.



Θ_g

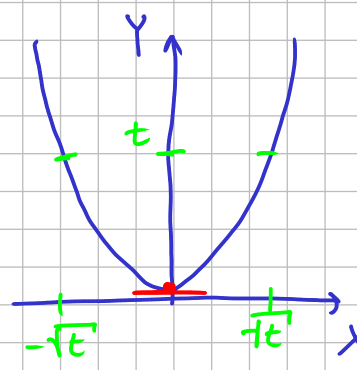
If X is a continuous random variable with density f_X, then r.v. $Y = g(X)$ has density

$$\boxed{f_Y(t) = \sum_{k=1}^{m(t)} \frac{f_X(s_k(t))}{|g'(s_k(t))|} \cdot I_V}$$

Example : Let $Y = X^2$ $\qquad U = \mathbb{R} \qquad V = \mathbb{R}_+$

<u>Singular points</u>

$\Theta_g = \{ s \in U \mid \nexists\, g'(s) \text{ or } g'(s) = 0 \}$



Θ_g

$g(x) = x^2 \quad \Rightarrow \quad g'(x) = 2x$

$\Rightarrow \quad \Theta_g = \{0\}$ because $g'(0) = 0$

multiplicity of $t \in V$

$g(\pm\sqrt{t}) = t \quad \Rightarrow \quad s_1(t) = \sqrt{t}$ & $s_2(t) = -\sqrt{t}$

such that for $t \geq 0$

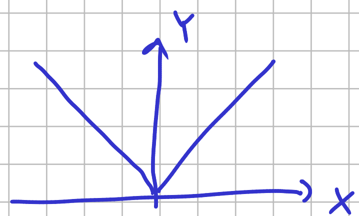
$$f_Y(t) = \frac{f_X(s_1(t))}{|g'(s_1(t))|} + \frac{f_X(s_2(t))}{|g'(s_2(t))|}$$

$$= \frac{f_X(\sqrt{t})}{|2\sqrt{t}\,|} + \frac{f_X(-\sqrt{t})}{|-2\sqrt{t}\,|}$$

$$= \frac{1}{2\sqrt{t}} \left(f_X(\sqrt{t}) + f_X(-\sqrt{t}) \right)$$

Example 2 : $Y = |X|$ $\qquad U = \mathbb{R}$

$V = \mathbb{R}_+$



Derivative of $g(x) = |x|$

$$g'(x) = \begin{cases} 1 & \text{for } x > 0 \\ -1 & \text{for } x < 0 \end{cases}$$

but does not exist for $x = 0$

$\Rightarrow \Theta_g = \{0\}$

Then as before $s_1 = x$ & $s_2 = -x$ and

for $t > 0$ we get

$$f_Y(t) = \frac{f_X(s_1(t))}{|g'(s_1(t))|} + \frac{f_X(s_2(t))}{|g'(s_2(t))|}$$

1 $\qquad\qquad$ 1

$$= f_X(t) + f_X(-t)$$

6.2 Transformation of multiple variables

Proposition: The density of the sum of two independent random variables $Z = X + Y$ is given by

$$f_Z(s) = \int_{-\infty}^{+\infty} f_X(t)\, f_Y(s-t)\, dt$$

Proof:

$$F_Z(t) = P(X+Y \leq t) = \int_{-\infty}^{+\infty} ds \int_{-\infty}^{t-s} ds' \underbrace{f_{X,Y}(s,s')}_{= f_X(s) f_Y(s')} \quad \text{indep.}$$

Transformation: $u_1 = s \in (-\infty, +\infty)$

$u_2 = s + s' \in (-\infty, t]$

$$\Leftrightarrow \quad s = u_1, \quad s' = u_2 - u_1$$

Jacobian $\begin{vmatrix} \frac{\partial s}{\partial u_1} & \frac{\partial s}{\partial u_2} \\ \frac{\partial s'}{\partial u_1} & \frac{\partial s'}{\partial u_2} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$

$$\Rightarrow F_Z(t) = \int_{-\infty}^{t} du_2 \underbrace{\int_{-\infty}^{+\infty} du_1\, f_X(u_1)\, f_Y(u_2 - u_1)}_{f_Z(u_2)}$$

Similarly we find

i) for $X \cdot Y = Z$

$$f_Z(t) = \int_{-\infty}^{+\infty} \frac{1}{|s|} f_X(s) f_Y\left(\frac{t}{s}\right) dt$$

ii) for $X/Y = Z$

$$f_Z(t) = \int_{-\infty}^{+\infty} |s| \, f_X(st) \, f_Y(s) \, ds$$

Example: Let $X, Y \sim N(0,1)$ be independent and $Z = X + Y$. Find the distribution and density function of Z.

$$f_Z(s) = \int_{-\infty}^{+\infty} f_X(t) \, f_Y(s-t) \, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\Big(\underbrace{-\tfrac{1}{2}t^2 - \tfrac{1}{2}(s-t)^2}\Big) \, dt$$

$$= -\frac{1}{2}\left(2t^2 - 2st + s^2\right)$$

$$= -\frac{1}{2}\left[\left(t\sqrt{2} - \frac{1}{\sqrt{2}}s\right)^2 + \frac{1}{2}s^2\right]$$

$$= \frac{1}{2\pi} e^{-\frac{1}{4}s^2} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left(t\sqrt{2} - \frac{1}{\sqrt{2}}s\right)^2\right) dt$$

$u = t - \frac{1}{2}s \quad \Leftrightarrow \quad t = u + \frac{1}{2}s \quad \Rightarrow \quad dt = du$

$$= \frac{1}{2\pi} e^{-\frac{1}{4}s^2} \underbrace{\int_{-\infty}^{+\infty} \exp(-u^2)\, du}_{\sqrt{\pi}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{s^2}{2}}$$

thus we find $Z \sim N(0,2)$.

More generally

<u>Proposition</u>: For $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ independent, we have:

$$Z = X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

(no proof)

Even more general:

<u>Theorem</u>: Let $\vec{X} \sim N(\mu_{\vec{X}}, \Sigma_{\vec{X}})$ an n-dimensional normally distributed random vector. The m-dimensional random vector $\vec{Y} = B \cdot \vec{X}$ with $m \leq n$ and B an $m \times n$ matrix is normally distributed as $\vec{Y} \sim N(\mu_{\vec{Y}}, \Sigma_{\vec{Y}})$ where

$$\mu_{\vec{Y}} = B \cdot \mu_{\vec{X}} \quad \text{and} \quad \Sigma_{\vec{Y}} = B \cdot \Sigma_{\vec{X}} \cdot B^t.$$

Proof (for $m=n$, $\det B \neq 0$):

transformation of variables

$$F_{\vec{X}}(\vec{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\vec{X}}(s_1, \dots, s_n)\, ds_1 \dots ds_n$$

with $$f_{\vec{X}}(s_1, \dots, s_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma'_{\vec{X}}}} \exp\left(-\frac{1}{2}(\vec{s}-\mu_{\vec{X}})^t \Sigma'^{-1}_{\vec{X}} (\vec{s}-\mu_{\vec{X}})\right)$$

trasformation : $\vec{s} = B^{-1}\vec{t}$

i) Jacobian $$|\det B^{-1}| = \left|\frac{1}{\det B}\right| = \sqrt{\frac{1}{(\det B)^2}}$$

$$= \sqrt{\frac{\det \Sigma'_{\vec{X}}}{\det (B \cdot \Sigma'_{\vec{X}} \cdot B^t)}} = \sqrt{\frac{\det \Sigma'_{\vec{X}}}{\det \Sigma'_{\vec{Y}}}}$$

ii) exponential: $$(\vec{s}-\mu_{\vec{X}})^t \Sigma'^{-1}_{\vec{X}} (\vec{s}-\mu_{\vec{X}}) = (B^{-1}\vec{t}-\mu_{\vec{X}})^t \Sigma'^{-1}_{\vec{X}} (B^{-1}\vec{t}-\mu_{\vec{X}})$$

$$= (B^{-1}(\vec{t}-B\mu_{\vec{X}}))^t \Sigma'^{-1}_{\vec{X}} (B^{-1}(\vec{t}-B\mu_{\vec{X}}))$$

$$= (\vec{t}-\underbrace{B\mu_{\vec{X}}}_{\mu_{\vec{Y}}})^t \underbrace{(B^{-1})^t \Sigma'^{-1}_{\vec{X}} B^{-1}}_{(B \cdot \Sigma'_{\vec{X}} \cdot B^t)^{-1} = \Sigma'^{-1}_{\vec{Y}}} (\vec{t}-B\mu_{\vec{X}})$$

$$= (\vec{t}-\mu_{\vec{Y}})^t \Sigma'^{-1}_{\vec{Y}} (t-\mu_{\vec{Y}})$$

$$\Rightarrow \quad F_{\vec{Y}}(\vec{y}) = \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} \frac{1}{(2\pi)^{n/2} \cancel{\sqrt{\det \Sigma_{\vec{X}}}}} \cdot \sqrt{\frac{\cancel{\det \Sigma_{\vec{X}}}}{\det \Sigma_{\vec{Y}}}}$$

$$\cdot \exp\left(-\frac{1}{2}(\vec{t}-\mu_{\vec{Y}})^t \Sigma_{\vec{Y}}^{-1} (\vec{t}-\mu_{\vec{Y}})\right) \, dt_1 \cdots dt_n$$

$\Rightarrow \vec{Y} \sim N(\mu_{\vec{Y}}, \Sigma_{\vec{Y}})$ with $\mu_{\vec{Y}} = B \cdot \mu_{\vec{X}}$ e

$\Sigma_{\vec{Y}} = B \cdot \Sigma_{\vec{X}} \cdot B^t$.

Remark:

The previous proposition is recovered for $B = (1,1)$ where $m = 1$ $n = 2$.

X_1 & X_2 independent $\Rightarrow \Sigma_{\vec{X}} = \begin{pmatrix} Var(X_1) & 0 \\ 0 & Var(X_2) \end{pmatrix}$

$$\Rightarrow \mu_{\vec{Y}} = B \cdot \mu_{\vec{X}} = \mu_{X_1} + \mu_{X_2}$$

$$\Sigma_{\vec{Y}} = B \cdot \Sigma_{\vec{X}} \cdot B^t = Var(X_1) + Var(X_2) = \sigma_1^2 + \sigma_2^2$$

6 Limit theorems

6.1 Empirical probability

A fair coin is thrown 100 times and the outcome is $w_H(100) = 52$ while $w_T(100) = 48$.
Here $w_a(n)$ denotes the absolute frequency of the of outcomes $a \in \{H, T\}$ of n throws.
The relative frequency or empirical probability is defined as

$$W_a(n) = \frac{w_a(n)}{n} \quad \text{with} \quad n = \text{"\# of trials"}$$

	H	T
W	0,52	0,48

For large n we expect that the empirical probability converges to the (theoretical) probability

$$p = \frac{P(\{H\})}{P(\{H,T\})} = \frac{1}{2} = 0{,}5.$$

Defining the independent random variables $X_i(H) = 1$, $X_i(T) = 0$, with $i = 1, \dots, n$, i.e. one for each throw, the empirical probability W_H can be written as

$$W_H(n) = \frac{1}{n} \sum_{i=1}^{n} X_i .$$

More generally, for arbitrary p we consider

$$\overline{W}(n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

where $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$.

Proposition: The expectation and variance of the empirical probability $W = \frac{1}{n} \sum_{i=1}^{n} X_i$ with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$ is

$$E(W) = p \quad \text{and} \quad Var(W) = \frac{1}{n} p(1-p)$$

Proof:

i) $$E(\overline{W}) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{E(X_i)}_{p} = p$$

ii) $$\begin{aligned} Var(W) &= E\left((W - \mu_W)^2\right) \\ &= \frac{1}{n^2} E\left(\left(\sum_{i=1}^{n} X_i - \mu_{\sum_i X_i}\right)^2\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\ &\underset{\text{indep.}}{=} \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{Var(X_i)}_{p(1-p)} \\ &= \frac{1}{n} p(1-p) \end{aligned}$$

6.2 Chebyshev's inequality

Proposition: (Chebyshev's inequality)

Let $c \in \mathbb{R}$ and X be a random variable with $\sigma_X^2 < \infty$. It follows for any $\epsilon > 0$ that

$$P(|X-c| \geq \epsilon) \leq \frac{1}{\epsilon^2} E((X-c)^2).$$

Proof: Note that $|t-c| \geq \epsilon \iff \frac{(t-c)^2}{\epsilon^2} \geq 1$

Then by definition

$$P(|X-c| \geq \epsilon) = \int_{|t-c| \geq \epsilon} f_X(t)\, dt$$

$$\leq \underbrace{\int_{|t-c| \geq \epsilon}}_{} \frac{(t-c)^2}{\epsilon^2} f_X(t)\, dt$$

$$\int_{-\infty}^{c-\epsilon} + \int_{c+\epsilon}^{+\infty} = \int_{-\infty}^{+\infty} - \int_{c-\epsilon}^{c+\epsilon}$$ gives positive contribution

$$\leq \frac{1}{\epsilon^2} \int_{-\infty}^{+\infty} (t-c)^2 f_X(t)\, dt$$

$$= \frac{1}{\epsilon^2} E((X-c)^2)$$

Setting $c = E(X)$ we obtain

$$P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon^2} E((X - E(X))^2)$$
$$= \frac{1}{\epsilon^2} Var(X)$$

$$\Leftrightarrow \quad P(|X - E(X)| < \epsilon) \geq 1 - \frac{1}{\epsilon^2} Var(X)$$

(since $1 - a \leq b$ with $a, b \geq 0$ $\quad -a \leq b - 1 \rightsquigarrow a \geq 1 - b$)

Example: The ages of members of a gym have mean (expectation) of 45 years and standard deviation of 11 years. We can estimate the percentage of gym members aged between $45 \pm c$, e.g. $c = 16{,}5$:

$$P(28{,}5 \leq X \leq 61{,}5) = P(|X - 45| \leq 16{,}5)$$
$$\geq 1 - \frac{1}{(16{,}5)^2} \cdot (11)^2 \approx 1 - 0{,}444 \approx 0{,}556$$

$\rightarrow$ at least $55{,}6\,\%$

Observation:

For the random variable $W = \frac{1}{n}\sum_{i=1}^{n} X_i$ we obtain

$$P(|W-p|<\epsilon) \geq 1 - \frac{1}{\epsilon^2}\frac{1}{n}p(1-p)$$

and thus

as probability can be at most 1.

$$\lim_{n \to \infty} P(|W-p|<\epsilon) = 1$$

We obtain Bernoulli's law of large numbers

6.3 Law of large numbers

Thm: For independent identically distributed random variables X_i with $P(X_i=1)=p$ and $P(X_i=0)=1-p$ it holds for $\sigma_{X_i}^2<\infty$ and for all $\epsilon>0$ that

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| < \epsilon\right) = 1$$

We say: The empirical probability converges to the theoretical probability with probability one.

More generically:

Thm: (Khinchin's law of large numbers)

For independent identically distributed random variables X_i with $E(X_i) = \mu < \infty$ and $\sigma_{X_i}^2 < \infty$ it holds for all $\epsilon > 0$

$$\lim_{n \to \infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1$$

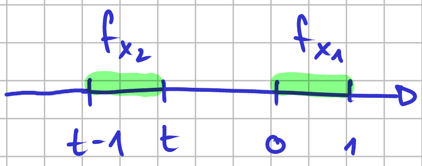
Example: Exponential distr. with Mathematica

6.4 Central limit theorem

Example: Consider the independent uniformly distributed random variables X_1, X_2, X_3, X_4 with $X_i \sim U([0,1])$.

The random variable $X_1 + X_2$ takes values in $[0,2]$ and has density

$$f_{X_1+X_2}(t) = \int_{-\infty}^{+\infty} f_{X_1}(s)\, f_{X_2}(t-s)\, ds$$

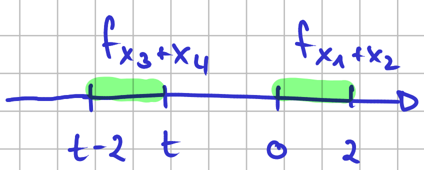


$$= \begin{cases} \int_0^t ds = t & \text{if } 0 < t \leq 1 \\ \int_{t-1}^1 ds = 2-t & \text{if } 1 < t \leq 2 \\ 0 & \text{else} \end{cases}$$

Similarly for $X_3 + X_4$.

We then compute

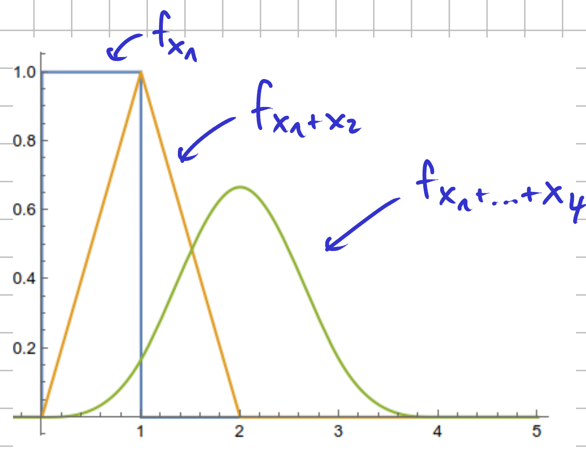
$$f_{X_1+X_2+X_3+X_4}(t) = \int_{-\infty}^{+\infty} f_{X_1+X_2}(s)\, f_{X_3+X_4}(t-s)\, ds$$



$$= \begin{cases} \int_0^t f_{X_1+X_2}(s)\, f_{X_3+X_4}(t-s)\, ds & 0 < t \leq 2 \\ \int_{t-2}^2 f_{X_1+X_2}(s)\, f_{X_3+X_4}(t-s)\, ds & 2 < t \leq 4 \\ 0 & \text{else} \end{cases}$$

$$= \begin{cases} \frac{1}{6} t^3 & 0 < t \leq 1 \\ -\frac{t^3}{2} + 2t^2 - 2t + \frac{2}{3} & 1 < t \leq 2 \\ \frac{t^3}{2} - 4t^2 + 10t - \frac{22}{3} & 2 < t \leq 3 \\ -\frac{t^3}{6} + 2t^2 - 8t + \frac{32}{3} & 3 < t \leq 4 \\ 0 & \text{else} \end{cases}$$

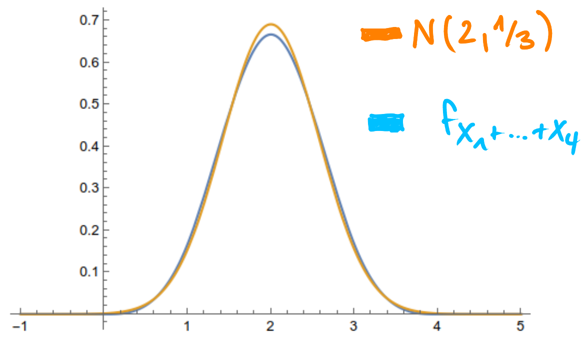
Graphically



We observe that $f_{X_1+\dots+X_4}$ resembles a normal distribution $N(\mu,\sigma^2)$ with

$E(X_1+\dots+X_4) = 2$

$Var(X_1+\dots+X_4) = 1/3$



or equivalently:

The standartised random variable

$$Z_4 = \frac{X_1+\dots+X_4 - E(X_1+\dots+X_4)}{\sqrt{Var(X_1+\dots+X_4)}}$$

resembles a normal distribution $N(0,1)$.

This motivates the central limit theorem.

Thm: (Central limit theorem)

Let $X_1, \dots, X_n$ be a sequence of independent and identically distributed random variables

with $E(X_i) = \mu < \infty$ and $Var(X_i) = \sigma^2 < \infty$ $\forall i$,

then

$$\lim_{n \to \infty} P\left(\frac{S_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}} \leq t \right) = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{s^2}{2}\right) ds}_{\Phi(t)}$$

where $S_n = \sum_{i=1}^{n} X_i$.

We identify the distribution function $\Phi(t)$ of the normal distribution $N(0,1)$.

Observation: The distribution function of the sum of standartized random variables converges to the distribution function of the standard normal distribution.

We also write for $Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n \cdot \sigma^2}}$ that

$$f_{Z_n}(t) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2}$$

or

$$f_{S_n}(t) \xrightarrow[n \text{ large}]{} \frac{1}{\sqrt{2\pi n}\,\sigma} e^{-\frac{1}{2} \frac{(t - n\mu)^2}{\sigma^2 n}}$$

where $S_n = \sum_{i=1}^{n} X_i$, $\mu = E(X_i)$ and $\sigma^2 = Var(X_i)$.

$$\Rightarrow \quad \boxed{P(a \le S_n \le b) \underset{n \text{ large}}{\approx} P\left(\frac{a-\mu_n}{\sigma\sqrt{n}} \le Z_n \le \frac{b-\mu_n}{\sigma\sqrt{n}}\right)}$$

where

$$P\left(\frac{a-\mu_n}{\sigma\sqrt{n}} \le Z_n \le \frac{b-\mu_n}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{b-\mu_n}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-\mu_n}{\sigma\sqrt{n}}\right)$$

This motivates the theorem of de Moivre & Laplace

Theorem: Let $a, b \in \mathbb{N}$ and X be a binomial random variable $X \sim Bin(n, p)$. Let Y be a normal random variable of mean $\mu = n \cdot p$ and variance $\sigma^2 = np(1-p)$ and let Z denote the standardized random variable $Z = \dfrac{Y - \mu_Y}{\sigma_Y}$. Then it holds that

$$P(a \le X \le b) \underset{n \to \infty}{\approx} P(a - 0.5 \le Y \le b + 0.5)$$

with $$P(a-0.5 \le Y \le b+0.5) = P\left(\frac{a-0.5-np}{\sqrt{np(1-p)}} \le Z \le \frac{b+0.5-np}{\sqrt{np(1-p)}}\right)$$

Remark: The terms ± 0.5 arise as corrections when going from discrete to continuous distributions. In particular it is guaranteed that $P(X=k) \approx P(k-0.5 \le Y \le k+0.5)$ does not vanish.

Rule of thumb: The approximation of de Moivre – Laplace can be used when $n \cdot p \ge 5$ and $n(1-p) \ge 5$

Example: A fair coin is tossed 1000 times. What is the probability that heads shows 510 times?

$$p = 1/2 \qquad n = 1000 \qquad \sqrt{n p \cdot (1-p)} = 5\sqrt{10}$$

a) $Bin(p)$

$$P(X=510) = \binom{1000}{510} \left(\tfrac{1}{2}\right)^{510} \left(1-\tfrac{1}{2}\right)^{1000-510} = 0{,}02065$$

b) de Moivre-Laplace $\quad (n \cdot p = 500 \geq 5, \quad n \cdot (1-p) = 500 \geq 5)$

$$P(X=510) \approx P\left(\frac{510-0.5-500}{5\sqrt{10}} \leq Z \leq P\left(\frac{510+0.5-500}{5\sqrt{10}}\right)\right)$$

$$= \Phi\left(\frac{10.5}{5\sqrt{10}}\right) - \Phi\left(\frac{9.5}{5\sqrt{10}}\right)$$

$$= \Phi(0.66) - \Phi(0.60)$$

$$= 0.74537 - 0.72575$$

$$= 0.0196$$

4. A student studies with probability $\frac{1}{2}$. He is aske a question with 4 possible answers. If the stuent studied he/she will certainly give the correct answer, if he/she has not studied he/she will choose one of the 4 at random. Suppose he/she gave the correct answer: how likely is it that he/she actually studied?

Def: $S =$ "did study" $S^c =$ "did not study"

$$P(S) = P(S^c) = \frac{1}{2}$$

Def: $E =$ "gave the right response"

$$P(E|S) = 1 \quad \& \quad P(E|S^c) = \frac{1}{4}$$

We want $P(S|E)$

Bayes' theorem

$$P(S|E) = \frac{P(E|S)\,P(S)}{P(E)} = \frac{P(E|S)\,P(S)}{P(E|S)P(S) + P(E|S^c)\,P(S^c)}$$

$$= \frac{1 \cdot \frac{1}{2}}{1 \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2}} = \frac{1}{1 + 1/4} = \frac{4}{5} = 0{,}8$$

2. An urn contains 6 white and 10 red balls. 5 balls are drawn. Find the probability of 3 being red:

(a) if the extraction is with reinsertion;

(b) if the checkout is without reinsertion.

a) $X \sim \text{Bin}\left(n=5,\ p=\frac{10}{16}\right)$

$$P(X=3) = \binom{5}{3}\left(\frac{10}{16}\right)^3\left(\frac{6}{16}\right)^2 = 0.343$$

b) $X \sim \text{Hyper}(10, 6, 5)$

$$P(X=3) = \frac{\binom{10}{3}\cdot\binom{6}{2}}{\binom{16}{5}} = 0{,}412$$

3. Let $X \simeq N(0.6; 3)$. Find

(a) $P(X \leq -0.53)$

(b) $P(X \geq 1)$

(c) $P(0 < X < 2)$

$\mu = 0{,}6$ & $\sigma^2 = 3$

a) $P(X \leq -0{,}53) = F_X(-0{,}53) = \Phi\left(\frac{-0{,}53 - 0{,}6}{\sqrt{3}}\right)$

$= \Phi(-0{,}65) = 1 - \Phi(0{,}65)$

$= 1 - 0{,}742 = 0{,}258$

b) $P(X \geq 1) = 1 - P(X < 1) = 1 - F_X(1)$

$= 1 - \Phi\left(\frac{1 - 0{,}6}{\sqrt{3}}\right) = 1 - \Phi(0{,}23)$

$= 1 - 0{,}591 = 0{,}409$

c) $P(0 < X < 2) = F_X(2) - F_X(0) = \Phi\left(\frac{2 - 0{,}6}{\sqrt{3}}\right) - \Phi\left(-\frac{0{,}6}{\sqrt{3}}\right)$

$= \Phi(0{,}81) - \Phi(-0{,}35)$

$= \Phi(0{,}81) - (1 - \Phi(0{,}35))$

$= 0{,}791 - (1 - 0{,}636) = 0{,}427$

7. The lifetime of an electric circuit has an exponential distribution with an average two years.

(a) Find the probability that the circuit will last at least 3 years

$$\mu = 2 = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{2}$$

$$T \sim \exp(1/2)$$

a) $P(T > 3) = 1 - P(T \leq 3) = 1 - F_T(3)$

$$= 1 - (1 - e^{-3/2})$$

$$= e^{-3/2} = 0{,}223$$

Recall :

$$F_T(t) = \int_{-\infty}^{t} f_T(s)\,ds = \int_0^t \lambda e^{-\lambda s}\,ds$$

$$= \left[-e^{\lambda s}\right]_0^t = 1 - e^{\lambda t}$$

for $t \geq 0$

and $F_T(t) = 0$ for $t < 0$

3. Let r.v. X, Y be discrete with probability mass function

$$f(1,10) = 6/13 \qquad f(1,20) = 1/13$$
$$f(2,10) = 3/13 \qquad f(2,20) = 3/13$$

Find

(a) the covariance of X and Y;

(b) the variances of X and Y;

		X		
		1	2	f_Y
Y	10	6/13	3/13	9/13
	20	1/13	3/13	4/13
f_X		7/13	6/13	

a) $\mu_X = 1 \cdot \frac{7}{13} + 2\,\frac{6}{13} = \frac{19}{13} = 1{,}46$

$\mu_Y = 10 \cdot \frac{9}{13} + 20\,\frac{4}{13} = \frac{170}{13} = 13{,}07$

$\mathrm{Cov}(X,Y) = E(XY) - \mu_X\mu_Y$

$= 1 \cdot 10\,\frac{6}{13} + 1 \cdot 20\,\frac{1}{13} + 2 \cdot 10\,\frac{3}{13} + 2 \cdot 20\,\frac{3}{13} - \frac{19 \cdot 170}{13^2}$

$= 0{,}887$

b) $\mathrm{Var}(X) = E(X^2) - \mu_X^2 = 1^2 \cdot \frac{7}{13} + 2^2\,\frac{6}{13} - \left(\frac{19}{13}\right)^2 = 0{,}248$

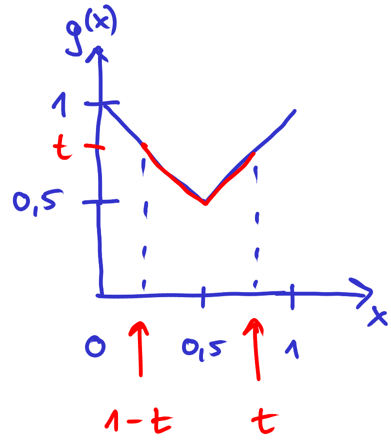
$\mathrm{Var}(Y) = 21{,}483$

4. Let X be uniform in $[0, 1]$. Find:

(a) the distribution function of $Y = \max(X, 1 - X)$;

(b) the probability density Y.

a) Preimage



$g^{-1}((-\infty, t)) = (1-t, t)$ se $0{,}5 \le t \le 1$

$$F_Y(t) = P(\max(X, 1-X) \le t) = \begin{cases} 0 & \text{if } t < 0{,}5 \\ P(1-t \le X \le t) & \text{if } 0{,}5 \le t \le 1 \\ 1 & \text{if } t \ge 1 \end{cases}$$

con $P(1-t \le X \le t) = F_X(t) - F_X(1-t) = t - (1-t) = 2t - 1$

$F_X(t) = \frac{t-a}{b-a}$

$\Rightarrow Y \sim U([\frac{1}{2}, 1])$

b)

$$f_Y(t) = \begin{cases} 2 & 1/2 \le t \le 1 \\ 0 & \text{else} \end{cases}$$

5. The airline VOLAREBASSO estimates that 5% of passengers that booked a ticket do not show up up on boarding. So they decide to overbook and accepts 310 reservations for a flight with 300 seats. How likely is it that some people cannot board the plane?

$p = 0{,}05 \qquad n = 310 \qquad \rightarrow n \cdot p = 15{,}5 \geq 5$

$n(1-p) = 310 - 15{,}5 \geq 5$

$$P(X \leq 9) \approx P\left(Z \leq \frac{9{,}5 - 15{,}5}{\sqrt{14{,}75}}\right)$$

$$= P(Z \leq -1{,}56)$$

$$= \Phi(-1{,}56)$$

$$= 1 - \Phi(1{,}56)$$

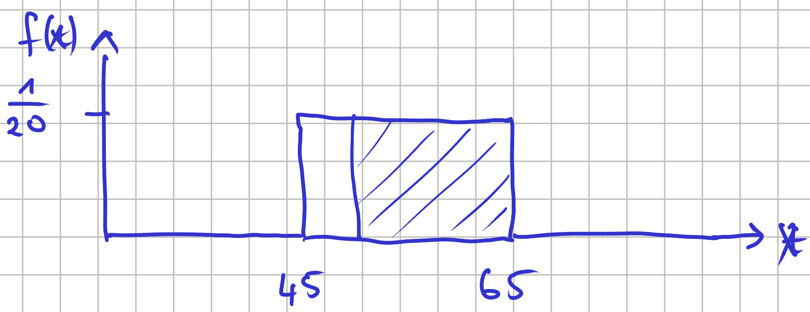
$$= 1 - 0{,}941$$

$$= 0{,}059$$

1. A factory produces bags of cement whose weight is described by a uniform random variable X that takes values between 45 and 65 kg. What is the probability that a bag weighs more than 50 kg?

(A) 0.68
(B) 0.75 ✔
(C) 0.84
(D) 0.77

$$\begin{cases} \frac{1}{\beta-\alpha} & \text{se } x \in (\alpha, \beta) \\ 0 & \text{altrimenti} \end{cases}$$



$\alpha = 45$

$\beta = 65$

$$P(X \geq 50) = \int_{50}^{65} \frac{1}{20}\,dx = \frac{1}{20} \cdot 15 = \frac{3}{4}$$

2. Let X be a continuous random variable with distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-x}(x+1), & x \geq 0. \end{cases}$$

What is the mean $\mathbb{E}(X)$?

(A) 4
(B) 1/2
(C) 2 ✔
(D) 1/4

$$\mu_X = \int_{-\infty}^{+\infty} dt\, t f(t)$$

$$f(x) = F'(x) = \begin{cases} 0 & x < 0 \\ e^{-x}(x+1) + (-1)e^{-x} = xe^{-x} & x \geq 0 \end{cases}$$

$$\mu_X = \int_0^{\infty} x^2 e^{-x}\,dx$$

$$\int_a^b g(x)\,h'(x)\,dx = \left[g(x)\,h(x)\right]_a^b - \int_a^b g'(x)\,h(x)\,dx$$

$g(x) = x^2 \qquad h'(x) = e^{-x}$

$$\underbrace{\left[-x^2 e^{-x}\right]_0^{\infty}}_{=0} + \int_0^{\infty} 2x\, e^{-x} = 2\underbrace{\left[-x e^{-x}\right]_0^{\infty}}_{=0} + 2\int_0^{\infty} e^{-x}$$

$$\Rightarrow \lim_{x \to \infty} x^k e^{-x} = 0$$

$$= 2\left[-e^{-x}\right]_0^{\infty}$$

$$= 2$$

3)

3. Let $X \sim \mathcal{N}(-2, 4)$ be a normal random variable with $\mu = -2$ and $\sigma^2 = 4$. Determine $c \in \mathbb{R}$ such that $P(X \geq c) = 0.2$.

(A) $c = 3.68$
(B) $c = 0.32$
(C) $c = -0.32$ ✔
(D) $c = -0.42$

$$P(X \geq c) = 0.2 = 1 - P(X < c)$$

$$\Leftrightarrow \quad P(X \leq c) = 0.8 = \Phi\left(\frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{c+2}{2}\right)$$

$$= \Phi(0.84)$$

$$\rightsquigarrow \frac{c+2}{2} = 0.84$$

$$\Rightarrow c = -0.32$$

4. In a call center arrive in avarage 3 calls every 15 minutes. What is the probability that more than 2 but less than 5 calls will arrive in the next 10 minutes?

(A) 0.0166
(B) 0.1353
(C) 0.6767
(D) 0.2707 ✔

$$\frac{3 \text{ calls}}{15 \text{ minutes}} = \frac{2 \text{ calls}}{10 \text{ minutes}}$$

$$\text{Pois}(\mu = 2)$$

$$P(2 < X < 5) = P(X=3) + P(X=4) = e^{-2}\left[\frac{2^3}{3!} + \frac{2^4}{4!}\right]$$

$$= 0.2707$$

5. Given two biased (not fair) coins. The first yields heads with probability $\frac{1}{3}$ and the second yields heads with probability $\frac{2}{5}$. They are tossed simultaneously, what is the probability that we obtain heads at least once?

(A) $p = 3/5$

(B) $p = 1/5$

(C) $p = 13/15$

(D) $p = 2/15$

$$S = \{ (H,H), (H,T), (T,H), (T,T) \}$$

$$P(\{(H,H)\}) = \frac{2}{15}$$

$$P(\{(H,T)\}) = \frac{1}{3} \cdot \frac{3}{5} = \frac{1}{5} = \frac{3}{15}$$

$$P(\{(H,T)\}) = \frac{2}{3} \cdot \frac{2}{5} = \frac{4}{15}$$

$$P(A) = \frac{9}{15} = \frac{3}{5}$$

$\rightsquigarrow$ Alternative

$$1 - P(\{(T,T)\}) = 1 - \frac{2}{5}$$

6. Two disjoint events A and B are always independent.

7. Let X_i with $i = 1, \dots, n$ be n independent Bernoulli random variables with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. It follows that the mean of the random variable $W_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ is given by $\mathbb{E}(W_n) = np$.

8. For two independent random variables X_1 and X_2 we have $Cov(X_1, X_2) = 0$.

6. $A \cap B = \emptyset$

Indep. $\quad P(A|B) = P(A) \iff P(A \cap B) = P(A)P(B)$

If $A \cap B = \emptyset \Rightarrow 0 = P(\emptyset) = P(A)P(B)$

$\rightarrow$ False

7) $$E(W_n) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n} E\left(\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} \underbrace{E(X_k)}_{p}$$

$$= \frac{1}{n} \cdot n \cdot p = p$$

False

8) $\mathrm{Cov}(X_1, X_2) = E(X_1 \cdot X_2) - \mu_{X_1}\mu_{X_2} = 0$

True

$$E(X_1 \cdot X_2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} ds\,dt\; s \cdot t\; f_{X_1,X_2}(s,t)$$

$$f_{X_1,X_2}(s,t) \overset{\leftarrow \text{indep.}}{=} f_{X_1}(s)\, f_{X_2}(t)$$

$$\underbrace{\int_{-\infty}^{+\infty} ds\; s \cdot f_{X_1}(s)}_{\mu_{X_1}} \cdot \underbrace{\int_{-\infty}^{+\infty} dt\; t\, f_{X_2}(t)}_{\mu_{X_2}}$$

Definition: A <u>random process</u> is a family $\{X_t \mid t \in T\}$ of random variables indexed by a variable t (usually indicating time) that belongs to a set T.

Def: A random process is said to be of <u>discrete-time</u> if the set T is countably infinite, e.g.
$T = \{0, 1, 2, \dots\} = \mathbb{N}_0$ or $T = \{\dots, -1, 0, +1, \dots\} = \mathbb{Z}$

E.g.: discrete: $T = \{0, 1, 2, 3, \dots\}$, $T = \{1, 2, 3\}$ or continious: $T = \mathbb{R}$, $T = \mathbb{R}_+$

Def: A random process is said to be of continuous-time if the set T is continuous e.g.
$T = (0, \infty) = \mathbb{R}_+$ or $T = (-\infty, +\infty) = \mathbb{R}$

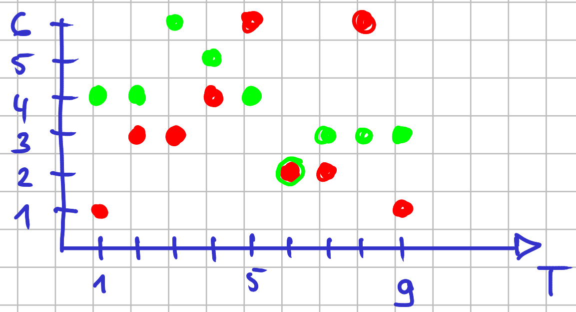
$\rightarrow$ We focus on discrete-time processes

Def: A <u>realization</u> or <u>sample path</u> of a random process is a collection $\{X_t(\omega) \mid t \in T\}$ for $\omega \in \Omega$.

Example 1: $T = \mathbb{N}$, $\Omega = \{(\text{⚀}, \text{⚀}, \dots), (\text{⚁}, \text{⚀}, \dots), \dots\}$

$S = \{1, 2, \dots, 6\}$

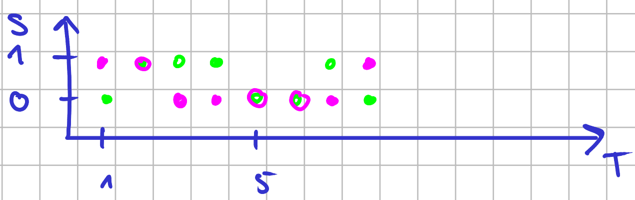
Example of a realisation



Red and green denote two different realizations

Example 2: Bernoulli process

Tossing a coin $X(\text{Head}) = 0$ $X(\text{Tails}) = 1$



Remark: In both cases the variables $X_1, X_2, \dots$ are independent

Example 3: Two gamblers A and B with initial fortunes a & b $(a, b \in \mathbb{N})$. An unfair coin with $P(T) = q$ and $P(H) = p = 1-q$ is flipped.

If the outcome is T, A gives 1€ to B if the outcome is H, B gives 1€ to A.

Let X_n denote the fortune of A after n flips and fix $X_0 = a$. It follows that $X_1, X_2, X_3, \dots$ are not independent as X_{t+1} depends on X_t.

Indeed: Let $m \in \{0, 1, \dots, a+b\}$, we have

$$P(X_{t+1} = m \mid X_t = m-1) = p \quad \text{and}$$
$$P(X_{t+1} = m \mid X_t = m-2) = 0 \quad (\neq p)$$

Thus $P(X_{t+1} = m \mid X_t = n) \neq P(X_{t+1} = m)$ as required for independence.

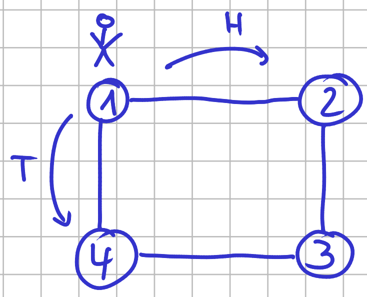
Remark: The process in Example 3 is memory less, i.e. the outcome X_{n+1} depends only on the outcome of X_n but not on $X_0, X_1, \ldots, X_{n-1}$.

"Given the present X_n, the future X_{n+1} does not depend on the past $X_0, \ldots, X_{n-1}$."

$$\Rightarrow P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)$$

$$= P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

Example 4: Random walk on a square (memory less)



Flip a coin $P(H) = p$

$P(T) = 1-p$

and move clockwise for H and anticlockwise for T.

e.g.: $P(X_{n+1} = 3 \mid X_n = 2) = P(H) = p$

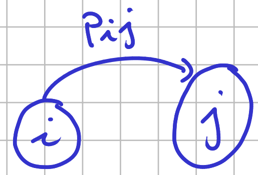
$P(X_{n+1} = 4 \mid X_n = 1) = P(T) = 1-p$

$P(X_{n+1} = 4 \mid X_n = 2) = 0$

Definition: A Markov chain is a memory less process $\{X_t\}_{t\in\mathbb{N}_0}$ with $X_t : \Omega \to S$ where S denotes a finite set.
More precisely, for all $t \in \mathbb{N}_0$ and all $i_0, \dots, i_{n+1} \in S$ we have the Markov condition (memory less)

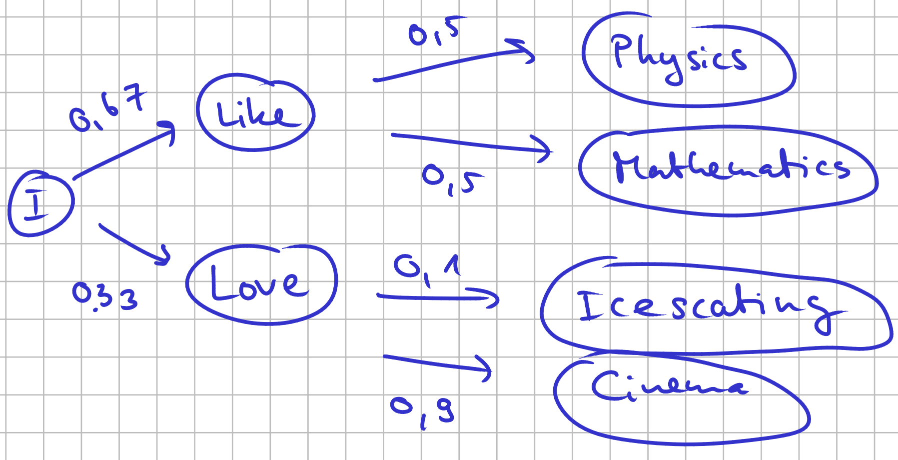
$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$$
$$= P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

Def: We call $p_{ij}(n) = P(X_{n+1} = j \mid X_n = i)$ the transition probability from state i to j.



Applications of Markov chains in the "real world"

- Next word prediction on mobile phone



- Market trends (Amazon suggestions)
- many more

Def: The Markov chain is homogeneous if

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) \qquad \forall\, n \in \mathbb{N}_0,\ i,j \in S$$

such that $p_{ij}(n) = p_{ij}(0) \equiv p_{ij}$.

Def: Let $|S| = N$, the $N \times N$ matrix $\mathbb{P}$ with entries $\mathbb{P}_{ij} = p_{ij} = P(X_1 = j \mid X_0 = i)$ is called <u>transition matrix</u>.

$$\mathbb{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{N1} & \cdots & & p_{NN} \end{pmatrix}$$

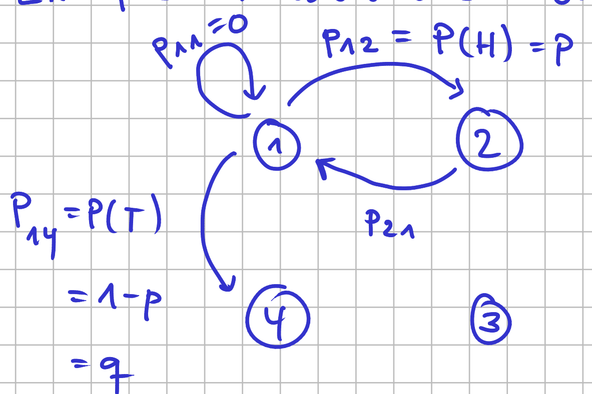
Properties: 1) The transition matrix $\mathbb{P}$ has non-negative entries $0 \leq p_{ij} \leq 1$ (<u>probabilities are positive</u>).

2) The sum of the rows is equal to 1:

$$\sum_{j=1}^{|S|} p_{ij} = 1 \quad \text{for } \forall i \in S \quad (\underline{\text{probability is conserved}})$$

We say that $\mathbb{P}$ is a <u>stochastic matrix</u>.

Example: Random walk on a square



$S = \{1, 2, 3, 4\}$

$P(H) = p$

$P(T) = q = 1 - p$

$$\mathbb{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{pmatrix} \end{array}$$

Properties: 1) $0 \leq p_{ij} \leq 1$

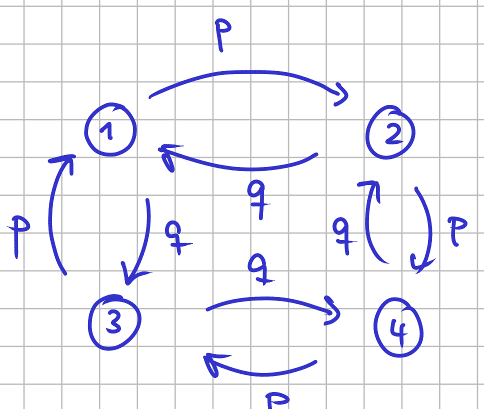
2) $\sum_{j=1}^{4} p_{ij} = 1 \quad \forall i$

Def: The transition probabilities can be denoted in an oriented Markov diagram connecting all states with non-vanishing transition probabilities by an arrow

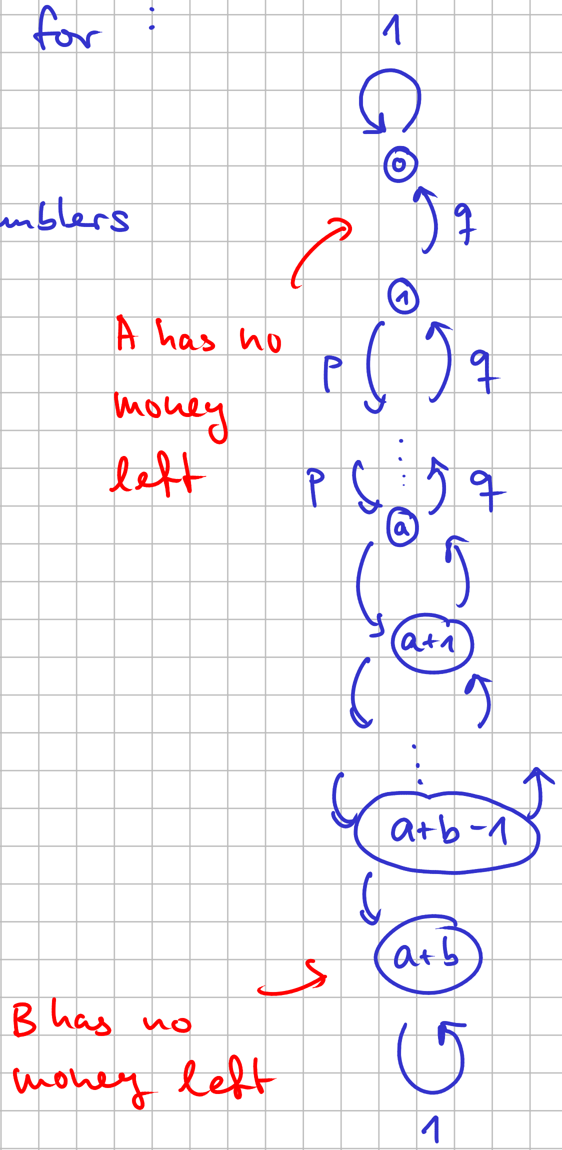
via: $i \xrightarrow{p_{ij}} j$

Exercise: Draw the Markov diagram for:

Random walk on square



Gamblers

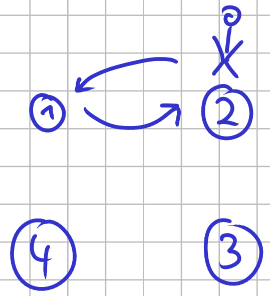


Exercise: (two-step transition)

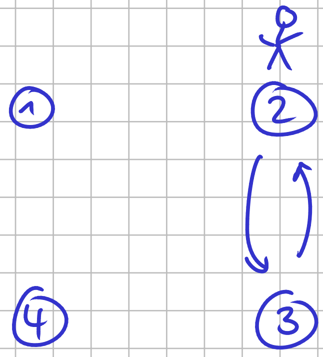
Given that the random walker is at position 2, what is the probability that he ends up at position 2 again after two steps $P(X_2=2 \mid X_0=2)$?

b) What is the probability $P(X_2=4 \mid X_0=2)$?

a)

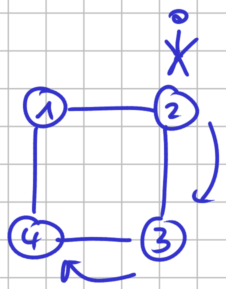


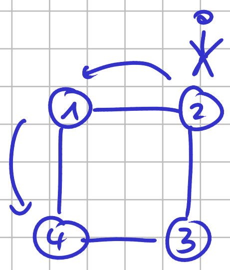
or



$$P(X_2=2|X_0=2) = P(X_2=2|X_1=1)\,P(X_1=1|X_0=2) + P(X_2=2|X_1=3)\,P(X_1=3|X_0=2) = 2p\cdot q$$

b)





$$P(X_2=4|X_0=2) = p^2+q^2$$

More generally for a homogeneous Markov chain with $|\Omega|=N$ we have

$$P(X_2=j|X_0=i) = \frac{P(X_0=i,\,X_2=j)}{P(X_0=i)} = \sum_{r=1}^{N} \frac{P(X_0=i,X_1=r,X_2=j)}{P(X_0=i)}$$

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

$$= \sum_{r=1}^{N} \underbrace{\frac{P(X_0=i,X_1=r,X_2=j)}{P(X_0=i,X_1=r)}}_{P(X_2=j|\cancel{X_0=i},X_1=r)} \underbrace{\frac{P(X_0=i,X_1=r)}{P(X_0=i)}}_{P(X_1=r|X_0=i)}$$

no memory

$$= \sum_{r=1}^{N} \underbrace{P(X_2=j \mid X_1=r)}_{p_{rj}} \; \underbrace{P(X_1=r \mid X_0=i)}_{p_{ir}}$$

$$= \sum_{r=1}^{N} p_{ir}\, p_{rj} \quad = \quad (\mathbb{P}\cdot\mathbb{P})_{ij} = (\mathbb{P}^2)_{ij}$$

For the random walker we have

$$\mathbb{P}^2 = \begin{pmatrix} 2pq & 0 & p^2+q^2 & 0 \\ 0 & 2pq & 0 & p^2+q^2 \\ p^2+q^2 & 0 & 2pq & 0 \\ 0 & p^2+q^2 & 0 & 2pq \end{pmatrix}$$

Remark: $2pq + p^2+q^2 = 2p(1-p) + p^2 + (1-p)^2 = 2p - p^2 + 1 - 2p + p^2$
$= 1$

Proposition: The n-step transition probabilities for a homogeneous Markov chain are

$$P(X_n = i \mid X_0 = j) = P(X_{n+m} = i \mid X_m = j) = (\mathbb{P}^n)_{ij}$$

We call $\mathbb{P}^n$ the n-step transition matrix

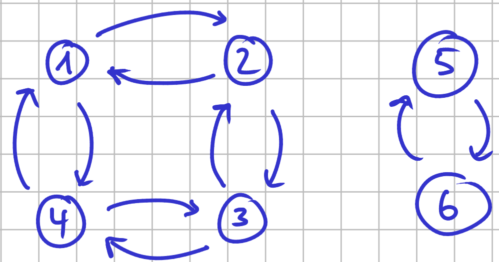
(proof as before)

Apart from the transition probabilities $P(X_n=i \mid X_0=j)$ we are typically interested in the probability to find the Markov chain in state i after n steps $P(X_n=i)$.

Remark: Even if the Markov chain is memory less, the probability $P(X_n=i)$ will usually depend on the initial value X_0.

Example :

$S = \{1, \dots, 6\}$



$\rightarrow$ if X_0 is 1, 2, 3 or 4, the process will never be in state 5 or 4 (and vice versa)

Def: The probability vector $\vec{\mu}^{(n)}$ is a row vector of dimension $|S|$ (the state space) with entries given by

$$\mu_i^{(n)} = P(X_n = i)$$

or

$$\vec{\mu}^{(n)} = (\underset{\mu_1^{(n)}}{\underset{\shortparallel}{P(X_n = 1)}}, \underset{\mu_2^{(n)}}{\underset{\shortparallel}{P(X_n = 2)}}, \dots, \underset{\mu_{|S|}^{(n)}}{\underset{\shortparallel}{P(X_n = |S|)}})$$

Properties: The entries of the probability vector $\vec{\mu}^{(n)}$

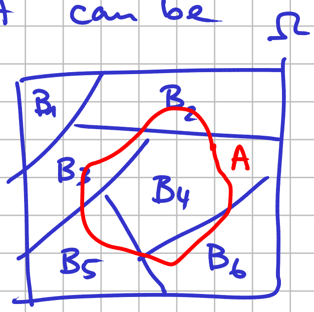
- are all non-negative $\quad \mu_i^{(n)} \geq 0 \quad \forall i \in S$
- the probability is conserved i.e. $\displaystyle\sum_{i=1}^{|S|} \mu_i^{(n)} = 1$

Given the probability distribution $\vec{\mu}^{(0)} = (P(X_0 = 1), \dots, P(X_0 = |S|))$ we can compute $\vec{\mu}^{(1)}$:

- Recall the total probability theorem:

For $B_1, \dots, B_N$ pairwise disjoint $\quad B_i \cap B_k = \emptyset$ and $\bigcup_{i=1}^{N} B_i = \Omega$ the probability of an event A can be written as

$$P(A) = \sum_{i=1}^{N} P(A \mid B_i)\, P(B_i)$$



- Take $A = \{X_1 = j\}$ with j fixed (Note that $|\Omega| = N^2$)

  $B_i = \{X_0 = i\}$ with $i = 1, \dots, N = |S|$

$$\Rightarrow \mu_j^{(1)} = P(X_1 = j) = \sum_{i=1}^{|S|} \underbrace{P(X_1 = j \mid X_0 = i)}_{p_{ij}} \underbrace{P(X_0 = i)}_{\mu_i^{(0)}}$$

$$= \sum_{i=1}^{|S|} \mu_i^{(0)} p_{ij} \qquad \text{for } j = 1, \dots, |S|$$

or in matrix form $\vec{\mu}^{(1)} = \vec{\mu}^{(0)} \cdot \mathbb{P}$

The same argument holds for $\vec{\mu}^{(2)}$:

$$\mu_j^{(2)} = P(X_2 = j) = \sum_{i=1}^{|S|} \underbrace{P(X_2 = j \mid X_1 = i)}_{p_{ij}} \underbrace{P(X_1 = i)}_{\mu_i^{(1)}}$$

$$= \sum_{i=1}^{|S|} \mu_i^{(1)} p_{ij}$$

$$= \sum_{i,k=1}^{|S|} \mu_k^{(0)} p_{ki} p_{ij}$$

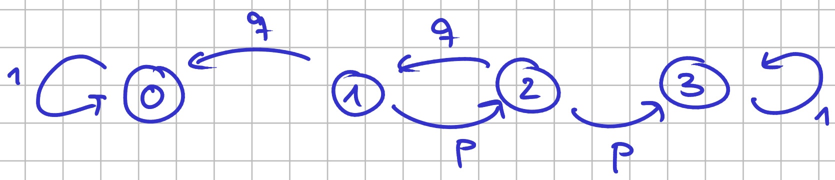
such that $\vec{\mu}^{(2)} = \vec{\mu}^{(0)} \cdot \mathbb{P}^2$.

This generalises to $\vec{\mu}^{(n)} = \vec{\mu}^{(0)} \cdot \mathbb{P}^n$ and motivates the following theorem

Theorem: For a homogeneous Markov chain $X_0, X_1, \dots, X_n, \dots$ with finite state space S, initial distribution $\vec{\mu}^{(0)}$ and transition matrix $\mathbb{P}$, we have that the distribution $\vec{\mu}^{(n)}$ at any time n satisfies: $\boxed{\vec{\mu}^{(n)} = \vec{\mu}^{(0)} \cdot \mathbb{P}^n}$

Example : Consider the problem of the two gamblers with initial condition $a=1$ & $b=2$ and compute the probability distribution after two time steps $\vec{\mu}^{(2)}$

$S = 0,1,2,3$



$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

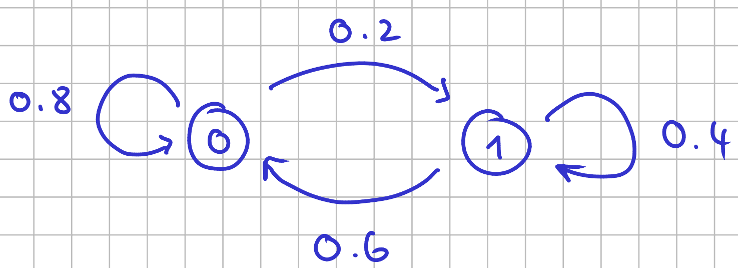
$$\Rightarrow \mathbb{P}^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & pq & 0 & p^2 \\ q^2 & 0 & pq & p \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \vec{\mu}^{(0)} = (0,1,0,0)$$

$$\Rightarrow \vec{\mu}^{(2)} = \vec{\mu}^{(0)} \, \mathbb{P}^2 = (q, pq, 0, p^2)$$

Check: $q + pq + p^2 = 1-p + p(1-p) + p^2 = 1$

<u>More examples:</u> Starting my car engine:

A car engine starts with probability 0.8 and does not with probability 0.2 in case it started the previous time. If the engine did not start it starts with 0.6 and does not start with 0.4.



Probability matrix $\mathbb{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}$

Given that the car was working at the begining $P(X_0 = 0)$, what is the probability that it won't work after 3 days ?

$$\mathcal{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$
$$\mathcal{P}^2 = \begin{bmatrix} 0.76 & 0.24 \\ 0.72 & 0.28 \end{bmatrix}$$
$$\mathcal{P}^3 = \begin{bmatrix} 0.752 & 0.248 \\ 0.744 & 0.256 \end{bmatrix}$$
$$\mathcal{P}^4 = \begin{bmatrix} 0.7504 & 0.2496 \\ 0.7488 & 0.2512 \end{bmatrix}$$
$$\mathcal{P}^5 = \begin{bmatrix} 0.7501 & 0.2499 \\ 0.7498 & 0.2502 \end{bmatrix}$$

1 day : 0.2

2 days : 0.24

3 days : 0,248

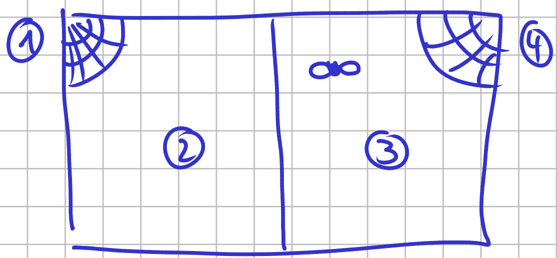
4 days : 0,2496

5 days : 0,2499

We observe that the probability converges to 0,25. This is the same value we get if the car would not have started at day 0. For large n $P(X_n = i)$ is independent of initial state.

We find $\lim_{n \to \infty} \vec{\mu}^{(n)} = (3/4, 1/4)$

<u>Another example</u>: The fly and two spiders



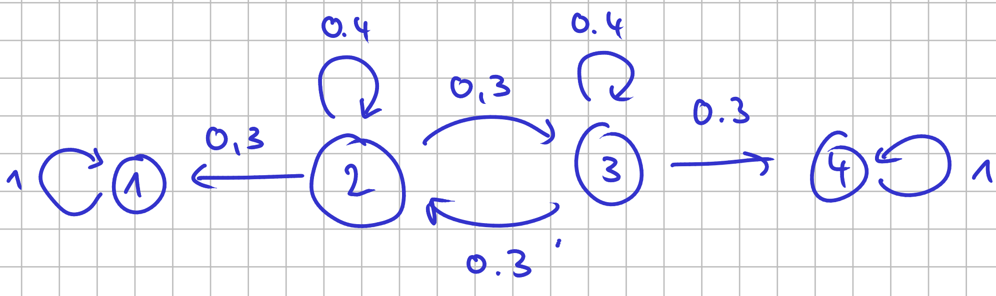
Possible positions of the fly

2: Left part of the room

3: Right part of the room

1: Left spider web

4: Right spider web



$$\mathcal{P} = \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0 & 0.3 & 0.4 & 0.3 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

$$\mathcal{P}^2 = \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0.42 & 0.25 & 0.24 & 0.09 \\ 0.09 & 0.24 & 0.25 & 0.42 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

$$\mathcal{P}^3 = \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0.50 & 0.17 & 0.17 & 0.16 \\ 0.16 & 0.17 & 0.17 & 0.50 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

$$\mathcal{P}^4 = \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0.55 & 0.12 & 0.12 & 0.21 \\ 0.21 & 0.12 & 0.12 & 0.55 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

$$\lim_{n\to\infty} \mathbb{P}^n = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Here $\lim_{n\to\infty} \mu^{(n)}$ depends on the initial state, however we see that the fly is with probability zero in state 2 or 3, thus captured by a spider.

Definition The vector π is called a stationary distribution of the Markov chain if

1) $\pi_i \geq 0$ for all i and $\sum_j \pi_j = 1$

2) $\pi \cdot \mathbb{P} = \pi$

It follows that $\vec{\pi} \cdot \mathbb{P}^n = \vec{\pi} \quad \forall n$ such that $\vec{\pi}$ does not change in time $\Rightarrow$ stationary

Examples: a) two spiders $\quad \vec{\pi} = (p, 0, 0, 1-p)$

fly is stuck in either left or right web

b) Starting a car in winter

$$(a,b)\begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \overset{!}{=} (a,b)$$

$$\Leftrightarrow \quad \begin{aligned} 0.8a + 0.6b &= a \quad \Leftrightarrow \quad 3b = a \\ 0.2a + 0.4b &= b \quad \Leftrightarrow \quad a = 3b \end{aligned}$$

further impose $\quad a+b=1$

$$\Rightarrow 3b = 1-b \quad \Leftrightarrow \quad b = \frac{1}{4}$$

$$\Rightarrow \vec{\pi} = (3/4, 1/4) = \lim_{n\to\infty} \vec{\mu}^{(n)}$$

This is what we observed previously and motivates the ergodic theorem of Markov chains

Theorem: (Ergodic theorem)

Let $\mathbb{P}$ be the transition matrix with entries p_{ij} where $i,j \in \{1,\dots,N\}$ with $N = |S|$. If there is an $n_0 \in \mathbb{N}$ such that $\mathbb{P}^{n_0}$ has strictly positive entries $(\mathbb{P}^{n_0})_{ij} > 0 \;\; \forall\, i,j = 1,\dots,N$, then there exists a probability vector $\vec{\pi} = (\pi_1, \dots, \pi_N)$ such that

$$\pi_i > 0 \quad \forall\, i = 1,\dots,N \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = 1$$

with

$$\lim_{n\to\infty} \vec{\mu}^{(n)} = \lim_{n\to\infty} \mathbb{P}^n \vec{\mu}^{(0)} = \vec{\pi}.$$

$\Rightarrow$ The system "forgets" its initial state if it evolves long enough.

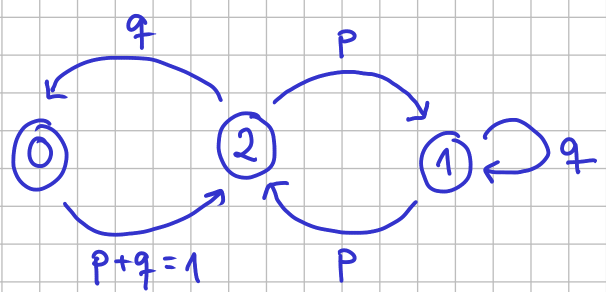
Example : I own two umbrellas. I keep moving between office and home. If it rains, I take an umbrella, if it does not I leave it behind. If it rains and there are no umbrellas where I am, I will get wet.

1) If the probability of rain is p, what is the probability to get wet?

Modena $p = \frac{69}{365} \approx 0{,}2$

2) How many umbrellas do I need to get wet with probability $< 1\%$?

1)



$$\Rightarrow \quad \mathbb{P} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & q & p \\ q & p & 0 \end{pmatrix}$$

$\rightsquigarrow (\mathbb{P}^4)_{ij} > 0 \quad \Rightarrow$ converges to a steady state (ergodic theorem)

$$(a,b,c) \cdot \mathbb{P} \overset{!}{=} (a,b,c) \iff (qc, qb+pc, a+pb) = (a, b, c)$$

$$\iff \begin{aligned} & a = qc \\ & c - b = 0 \end{aligned}$$

$a + bp = qc + bp = c \iff b - c = 0$

$\rightsquigarrow \quad (qc, c, c)$ Normalization $2c + qc = 1$

$$\Rightarrow c = \frac{1}{2+q}$$

$$\Rightarrow \pi = \left(\frac{q}{2+q}, \frac{1}{2+q}, \frac{1}{2+q}\right)$$

Modena: $q = 1 - p = 0.8 \quad \Rightarrow \quad \pi \approx (0.28, 0.36, 0.36)$

$\Rightarrow$ with probability 0.28 there is no umbrella where I am.

$\Rightarrow$ I get wet with probability $0.28 \cdot 0.2 = 0.056$

$\approx 6\%$

2) $k = \#$ umbrellas

$$\mathbb{P} = \begin{pmatrix} & & & 1 \\ & & q & p \\ & q & p & \\ \cdot^{\cdot^{\cdot}} q & p & & \end{pmatrix}$$

Solve $\pi \cdot \mathbb{P} = \pi \iff (\pi_k q, \pi_{k-1} q + \pi_k p, \dots, \pi_1 q + \pi_2 p, \pi_0 + \pi_1 p)$
$= (\pi_0, \pi_1, \dots, \pi_k)$

$\Rightarrow$ i) $\pi_0 = \pi_k q$

ii) $\pi_\ell = \pi_{k-\ell} q + \pi_{k-\ell+1} p \qquad \ell = 1, \dots, k-1$

iii) $\pi_k = \pi_0 + \pi_1 p$

$$\Rightarrow \vec{\pi} = \left(\frac{q}{q+k}, \frac{1}{q+k}, \dots, \frac{1}{q+k}\right)$$

check eqas i) ✓

ii) $\frac{q}{q+k} + \frac{\overset{=1-q}{p}}{q+k} = \frac{1}{q+k}$ ✓

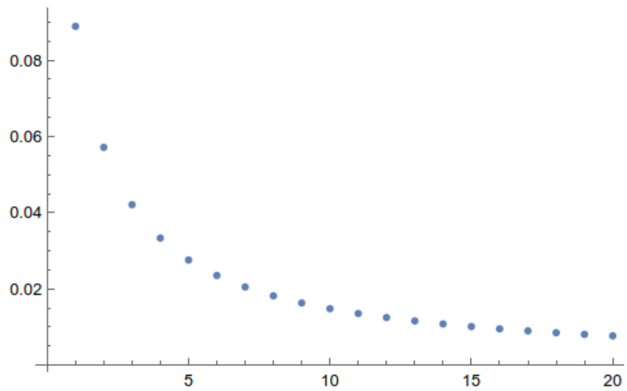
iii) $\frac{q}{q+k} + \frac{\overset{=1-q}{p}}{q+k} = \frac{1}{q+k}$ ✓

$\Rightarrow P(\text{Wet}) = \frac{pq}{q+k} \overset{!}{=} 0{,}01$

$\Rightarrow \quad k = 15.2$

```
f[k_] := 0.8 0.2/(0.8 + k)

f /@ Range[20] // ListPlot
```



Probability model for data

Fundamental assumption:
the measurable quantity associated with each item of a given population is a value assumed by a random variable X with probability distribution $F_X(x) = P(X \le x)$

Goal: Study X and its distribution F_X

We distinguish the two scenarios:

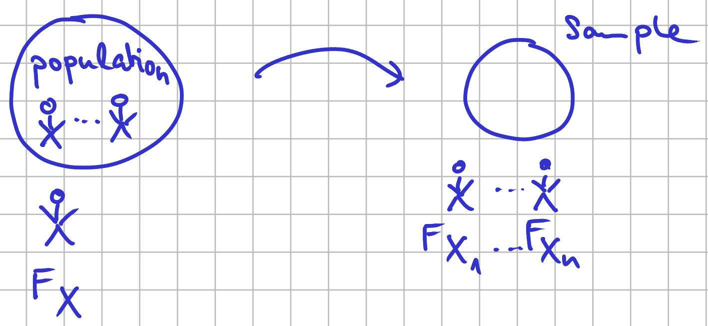
1) **parametric inference problems**
   We know that X has a certain distribution F (e.g. Normal, Binomial, etc) but the parameters are unknown and should be estimated (eg. $N(\mu, \sigma^2)$)
   estimate μ & σ

2) **non parametric inference problems**
   distribution of X unknown

More generally: A population may depend on m parameters $(\Theta_1, \dots, \Theta_m)$ such that $F_X(x) = F_X(x; \Theta_1, \Theta_2, \dots, \Theta_m)$.

→ Introduce **estimators** for $\Theta_1, \dots, \Theta_m$

Def: If $X_1, \dots, X_n$ are independent random variables having a common distribution $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$, then we say that they constitute a **(random) sample** from the distribution F_X



Def: A statistic is a random variable composed out of the elements of a sample $X_1, \dots, X_n$ that does not contain unknown parameters, e.g.

$$g(X_1, \dots, X_n) = X_1 + \dots + X_n$$

Def: An estimator is a statistic used to estimate parameters of a population. The estimate is the value which the estimator yields for a particular sample.

Example: Population : pupils of a school with unknown height distribution F_X

$\downarrow$

Sample : choose at random 30 pupils

$\downarrow$

Estimator : Average height $\frac{1}{30} \sum_{i=1}^{30} X_i = g(X_1, \dots, X_{30})$

$\downarrow$

Estimate : outcome of estimator $\frac{1}{30} \sum_{i=1}^{30} x_i = g(x_1, \dots, x_{30})$

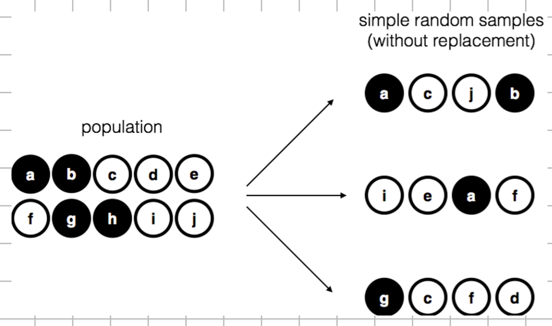
$= 1{,}60\,m$

This gives an estimate for the mean $\mu = E(X)$.

(As discussed in the context of the empirical probability)

Example: 10 coins choose 4

(color not relevant)



Two important statistics are given by the sample mean and the sample variance. They are used to estimate the population mean $\mu = E(X)$ and the population variance $\sigma^2 = Var(X)$.

Def: Given a sample $X_1, \dots, X_n$ i.i.d., the sample mean is the estimator defined as

$$\boxed{\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

The estimate for a given sample then yields $\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$.
(coincides with empirical probability)

Remark: Different samples may give different estimates.
$\rightsquigarrow$ $\overline{X}_n$ has a probability distribution itself

Proposition: The expectation of the sample mean coincides with the population mean, i.e. $\boxed{E(\overline{X}_n) = \mu.}$

Proof: $E(\overline{X}_n) \underset{\text{linear}}{=} \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$.

i.i.d $\Leftrightarrow$ $E(X_1) = E(X_2) = \dots = E(X_n) = E(X) = \mu$

Proposition: The variance of the sample mean $\overline{X}_n$ is the population variance divided by n, i.e

$$\boxed{Var(\overline{X}_n) = \frac{\sigma^2}{n}}$$

Proof: $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \underset{\substack{\uparrow \\ \mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)}}{=} \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$

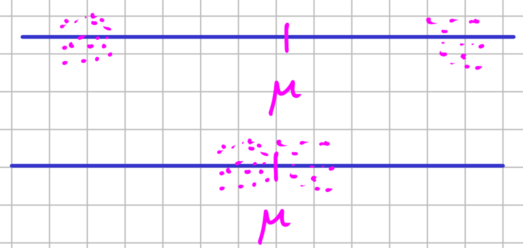
$$\underset{\substack{\uparrow \\ \text{independent}}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \underset{\substack{\uparrow \\ \text{identical}}}{=} \sigma^2/n$$

We conclude that $\bar{X}_n$ is centered around the population mean μ (like the population itself). The variance of $\bar{X}_n$ decreases, as the sample size increases.

<u>Recall variance</u>:

large variance

small variance



It follows that for large sample size the result for the estimate of μ becomes more accurate.

This can be shown using the Chebyshev inequality

$$P(|Y - \mu_Y| \geq \varepsilon) \leq \frac{\sigma_Y^2}{\varepsilon^2} \qquad \forall \varepsilon > 0$$

$$\Rightarrow \quad P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2_{\bar{X}_n}}{\varepsilon^2} = \frac{1}{\varepsilon^2}\frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$

We say that the sample mean X_n converges in probability to the population mean μ as the size of the sample goes to infinity.

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0 \qquad \forall \varepsilon > 0.$$

The following question arises naturally :
What is the probability distribution of the sample mean ?

Remark: If the distribution of the population (and hence the distribution of each element in the sample $X_1, \dots, X_n$) is F, ingeneral the distribution of the sample mean is not F.
(Recall example of uniform distributions that converge to normal distribution but not to another uniform).

Special case: If the population is normally distributed with mean μ and variance σ^2 the sum $X_1 + \dots + X_n$ is normal with mean $n\mu$ and variance $n\sigma^2$.
Thus the sample mean is normal with mean μ and variance σ^2/n, i.e. $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

(Recall)
Proposition: Suppose that $X_i \sim N(\mu_i, \sigma_i^2)$ $i = 1, \dots, n$ and X_i independent, then

$$X_1 + X_2 + \dots + X_n \sim N(\mu_1 + \mu_2 + \dots + \mu_n, \sigma_1^2 + \dots \sigma_n^2).$$

More generally we have

$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

where $a_i \in \mathbb{R}$.

(without proof)

It follows that $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

If the distribution of the population is not normal or even unknown the proposition above holds approximately for large samples.

As a consequence of the central limit theorem we have:

Proposition: Let $X_1, \dots, X_n$ be a sequence of i.i.d. random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Then for n large the distribution of

$$S_n = X_1 + \dots + X_n$$

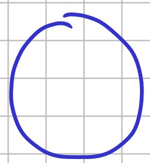
is appoximately normal with $E(S_n) = n\mu$ and $Var(S_n) = n\sigma^2$, i.e. $S_n \underset{n \text{ large}}{\sim} N(n\mu, n\sigma^2)$.

It follows that $\bar{X}_n \underset{n \text{ large}}{\sim} N(\mu, \sigma^2/n)$.

Example: The weights of a population of students have mean 167 pounds and standard deviation 27 pounds.

a) If a sample of 36 students is chosen, approximate the probability that the sample mean of their weight lies between 163 and 170

b) Repeat a) for sample size 144



population

X

$E(X) = 167$

$Var(X) = (27)^2$

a) Take sample of 36 : $X_1, \dots, X_{36}$

Sample mean : $\bar{X}_{36} = \frac{1}{36}(X_1 + \dots + X_{36}) \sim N\left(\mu = 167, \sigma^2 = \frac{(27)^2}{36}\right)$

$$\frac{(27)^2}{36} = \left(\frac{27}{6}\right)^2 = (4.5)^2$$

$$P(163 \leq \bar{X}_{36} \leq 170) = P\left(\frac{163-167}{4.5} \leq Z \leq \frac{170-167}{4.5}\right)$$

$$= \Phi(0{,}75) - \Phi(-0{,}889)$$

$$= 0{,}7734 - (1 - 0{,}81) = 0{,}58$$

b) Take a sample of 144 students : $\bar{X}_{144} \sim N\left(\mu = 167, \sigma^2 = \frac{(27)^2}{144}\right)$

$\Rightarrow \sigma = \sqrt{5.0625} = 2{,}25$

$$\Rightarrow P(163 \leq \bar{X}_{144} \leq 170) = \Phi(1{,}33) - \Phi(-1{,}78)$$

$$= 0{,}91 - (1 - 0{,}96)$$

$$= 0{,}91 - 0{,}04 = 0{,}87$$

Rule of thumb

> For samples with $n > 30$ we can approximately assume that the sample mean is normal with $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

Example : 52 % of the residents of a certain city are in favor of limiting the speed of cars to 30 km/h. Find the approximate probability that at least 50% are in favor when taking a sample of

a) 10

b) 100

c) 1000

$$X_i = \begin{cases} 1 & \text{if in favor} \\ 0 & \text{if not} \end{cases}$$

$$X \sim \text{Bern}(0.52) \quad \Rightarrow \quad \mu = 0{,}52 \quad \& \quad \sigma^2 = 0{,}52(1-0{,}52) = 0{,}2496$$

$$\Rightarrow \bar{X}_n \sim N\left(0.52, \frac{0.2496}{n}\right)$$

$$P(\bar{X}_n \geq 0{,}5) = 1 - P(\bar{X}_n < 0{,}5)$$

$$= 1 - \Phi\Big(\underbrace{\frac{0{,}5 - 0{,}52}{\sqrt{0.2496}}}_{-0{,}04} \sqrt{n} \Big)$$

$$= \Phi(0{,}04\sqrt{n})$$

$n = 10$: $\Phi(0{,}13) \approx 0{,}54$

$n = 100$: $\Phi(0{,}4) \approx 0{,}66$

$n = 1000$: $\Phi(1{,}26) \approx 0{,}898$

$n = 10.000$: $\Phi(4) \approx 1$

Def: Let $T_n = T_n(X_1, \ldots, X_n)$ be an estimator of a parameter θ, we say

a) T_n is unbiased if $E(T_n) = \theta$

b) T_n is consistent if T_n converges to θ in probability as $n \to \infty$:

$$\lim_{n\to\infty} P(|T_n - \theta| < \varepsilon) = 1 \quad \text{for } \forall\, \varepsilon > 0$$

$\Rightarrow$ The sample mean is an unbiased and consistent esimator of the population mean. (unbiased $\not\Leftrightarrow$ consistent)

Sample Variance

Def: Let $X_1, \dots, X_n$ be a sample from a population with mean μ and variance σ^2. The sample variance is the statistics defined by

$$\boxed{S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2}$$

where $\bar{X}_n$ denotes the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Proposition: The sample variance S_n^2 is an unbiased estimator of the variance of a population σ^2, i.e.:

$$\boxed{E(S_n^2) = \sigma^2}$$

Proof:

$$E(S_n^2) = \frac{1}{n-1} \sum_{i=1}^{n} E\Big[\underbrace{(X_i - \bar{X}_n)^2}_{X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2}\Big]$$

linear

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left\{ E(X_i^2) - 2E(X_i \bar{X}_n) \right\} + \frac{n}{n-1} E(\bar{X}_n^2)$$

identical
$\Rightarrow E(X_i^2) = E(X^2)$

$$= \frac{n}{n-1} E(X^2) - \frac{2}{n-1} E\Big(\overbrace{\sum_{i=1}^{n} X_i}^{n\bar{X}_n} \bar{X}_n\Big) + \frac{n}{n-1} E(\bar{X}_n^2)$$

$$= \frac{n}{n-1} E(X^2) - \frac{n}{n-1} E(\bar{X}_n^2)$$

$$= \frac{n}{n-1} \left\{ Var(X) + E(X)^2 - Var(\bar{X}_n) - E(X_n)^2 \right\}$$

$$= \frac{n}{n-1} \left\{ \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \right\}$$

$$= \frac{n}{n-1} \left\{ \frac{n-1}{n} \sigma^2 \right\} = \sigma^2 \qquad \square$$

Recall

$$Var(X) = E(X^2) - \underbrace{E(X)}_{\mu_X}{}^2$$

Remark: The estimator

$$\tilde{S}_n^{\,2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

is a biased estimator of the population variance σ^2

$$E(\tilde{S}_n^2) = \frac{n-1}{n} E(S_n^2) = \left(1 - \frac{1}{n}\right)\sigma^2 \neq \sigma^2 .$$

Only for $n \to \infty$ the expectation of $\tilde{S}_n^2$ converges to the population variance.

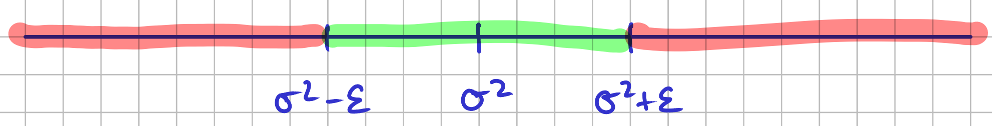
Proposition: The sample variance is a consistent estimator of the population variance

$$\lim_{n\to\infty} P\left(|S_n^2 - \sigma^2| < \varepsilon \right) = 1 \qquad \forall\ \varepsilon > 0$$

sketch of proof: Chebychev inequality (show that Var vanishes)

$$\mathrm{Var}(S_n^2) = \frac{(n-1)\left[(n-1)\, E(X^4) - (n-3)\, E(X^2)^2\right]}{n^3}$$

$$\underset{\text{large } n}{\approx} \frac{1}{n} E(X^4) \xrightarrow[n\to\infty]{} 0$$



Remark: The same holds for the estimator $\tilde{S}_n^2 = \frac{n-1}{n} S_n^2$

as $\mathrm{Var}(\tilde{S}_n^2) = \frac{(n-1)^2}{n^2} S_n^2 = \left(1 - \frac{2}{n} + \frac{1}{n^2}\right) S_n^2 .$

Distribution of the sample variance

Theorem: If $X_1, \dots, X_n$ is a sample from a normally distributed population $N(\mu, \sigma^2)$, then the sample variance S_n^2 is related to the χ^2_{n-1} (chi-square) distribution with $n-1$ degrees of freedom via

$$\boxed{(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Proposition: The sample mean and sample variance are independent. (no proof)

Def: The distribution χ_n^2 is continuous with probability density function

$$f_n(x) = \begin{cases} \dfrac{x^{n/2-1} \, e^{-x/2}}{2^{n/2} \, \Gamma(n/2)} & \text{if} \quad x \geq 0 \\ 0 & \text{else} \end{cases}$$

with $n \in \mathbb{N}$

Definition of the Euler Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt \quad \text{for} \quad x > 0$$

Properties:

1) $\Gamma(x+1) = x\,\Gamma(x)$ for $x \in \mathbb{R}$
2) $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$
3) $\Gamma(1/2) = \sqrt{\pi}$

Observation: The χ^2-distribution is identical to the Gamma distribution $\Gamma(r,\lambda)$ for $n=2r$ and $\lambda=\frac{1}{2}$.

$$\frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} = \frac{x^{n/2-1}}{2^{n/2}\,\Gamma(n/2)} e^{-x/2}$$

Proposition: A random variable X with distribution χ^2_n has mean $E(X)=n$ and variance $Var(X)=2n$.

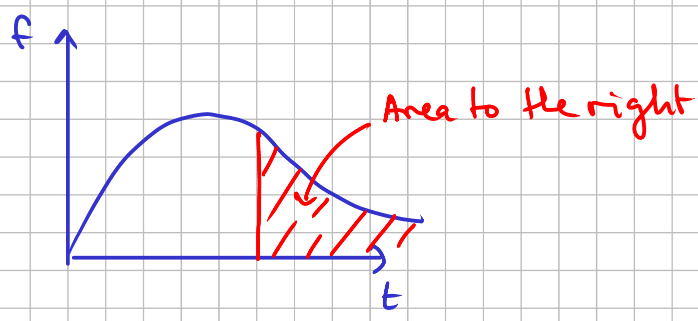
(Proof as for Gamma distribution)

Remark: Similar to the case of the normal distribution the distribution fuction of χ^2_n has to be evaluated numerically.

Chi-Square (χ^2) Distribution

Area to the Right of Critical Value

Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169



Eg. $n = 10$

$P(\chi_n^2 \geq t) = 0{,}9$

$\Rightarrow \quad t = 4{,}865$

Example : The time to serve a customer at a fast food restaurant is approximately distributed normally with $N(20, 8)$. If 15 customers are served what is the probability that the sample variance exceeds 12?

$$P(S_{15}^2 > 12) = P\Big(\underbrace{\frac{15-1}{8} S_{15}^2}_{\chi_{14}^2} > \frac{15-1}{8} 12\Big)$$

$$= P(\chi_{14}^2 > 21) = 0{,}1$$

Consider the sample mean $\bar{X}_n \sim N(\mu, \sigma^2/n)$. We define the standardized version of the sample mean as

$$\bar{X}_n^* = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

such that $E(\bar{X}_n^*) = 0$ & $Var(\bar{X}_n^*) = 1$.

In case σ is not known we can estimate it with square root of the sample variance. This motivates the study of the random variable

$$\sqrt{n} \, \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} .$$

Def: The (Student's) t-distribution is continuous with probability density function

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + x^2/n\right)^{-\frac{n+1}{2}}$$

with $n \in \mathbb{N}$ and $x \in \mathbb{R}$.

Proposition: If $V \sim N(0,1)$ and $U \sim \chi^2_n$ are independent, then the random variable

$$T_n = \frac{V}{\sqrt{U}} \sqrt{n} \qquad n \in \mathbb{N}$$

has a t-distribution with n degrees of freedom.

proof: 1) obtain $f_{U'}$ for $U' = \sqrt{U}/\sqrt{n}$

$$P(U' \le s) = P(\sqrt{U}/\sqrt{n} \le s) = P(U \le n s^2) = F_U(n s^2)$$

$$f_{U'}(s) = f_U(ns^2)\, 2sn = 2sn \frac{s^{n-2}\; e^{-n s^2/2}}{2^{n/2}\; \Gamma(n/2)} \cdot n^{n/2-1}$$

$$= \frac{s^{n-1}\; \exp(-ns^2/2)}{2^{n/2-1}\; \Gamma(n/2)}\, n^{n/2}$$

2) Recall that for $Z = \frac{X}{Y}$ we have

$$f_Z(x) = \int_{-\infty}^{+\infty} |s|\, f_X(st)\, f_Y(s)\, ds$$

Set $X = V$ and $Y = \sqrt{U}/\sqrt{n}$

$$\Rightarrow f_Z(t) = \frac{1}{\sqrt{2\pi}} \frac{n^{n/2}}{2^{n/2-1}\,\Gamma(n/2)} \int_0^\infty \exp\left(-\tfrac{1}{2}s^2(t^2+n)\right) s^n \, ds$$

$\parallel$ (Mathematica)

$$2^{\frac{1}{2}(-1+n)} \left(n+t^2\right)^{-\frac{1}{2}-\frac{n}{2}} \text{Gamma}\left[\frac{1+n}{2}\right]$$

$$= \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)} \qquad \square$$

Now recall that $(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2_{n-1}$ and that $\bar{X}_n^* = \frac{\bar{X}_n-\mu}{\sigma}\sqrt{n} \sim N(0,1)$. It follows

Corollary : Let $X_1,\dots,X_n$ be a sample from a normal population with mean μ. If $\bar{X}_n$ denotes the sample mean and $S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X}_n)^2}$ the sample standard deviation, then

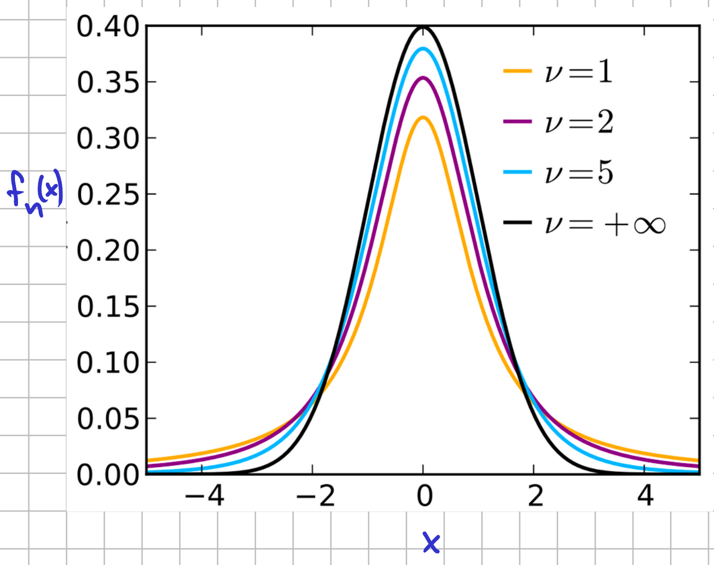
$$\boxed{\frac{\bar{X}_n-\mu}{S_n}\sqrt{n} \sim T_{n-1}}$$

where T_{n-1} denotes the t-distribution.

[We used $\frac{\bar{X}_n-\mu}{\sigma}\sqrt{n}\,\frac{1}{\sqrt{n-1}\,S_n/\sigma}\cdot\sqrt{n-1} = \frac{\bar{X}_n-\mu}{S_n}\sqrt{n}$]

Properties of the density of T_n

- $f(x)$ is continuous
- $f(x)$ is even $\quad f(x) = f(-x)$
- At $x=0$ the density $f(x)$ is maximal
- $\lim\limits_{x \to \pm\infty} f(x) = 0$



$\nu = n$

Theorem: The t-distribution converges to a normal distribution $N(0,1)$ for $n \to \infty$, i.e.

$$\lim_{n\to\infty} P(T_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi(x)$$

Proof:

Uses $\displaystyle \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \underset{n \text{ large}}{\sim} \sqrt{\frac{n}{2}}$ and $\displaystyle \lim_{n\to\infty} \left(1+\frac{\mu}{n}\right)^n = e^{\mu}$

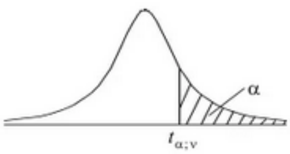
Proposition: A random variable T_n with distributed as the t-distribution of degree n has mean $E(T_n) = 0$ if $n \geq 2$ and variance $Var(T_n) = \frac{n}{n-2}$ if $n \geq 3$.

(without proof)

Remark: Similar to the case of the normal & χ^2 distribution the distribution function of T_n has to be evaluated numerically.

Table of the Student's *t*-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with ν degrees of freedom



ν \ α	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.076	31.821	63.657	318.310	636.620
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

E.g.:

$P(T_{10} \geqslant x) = 0{,}025$

$\Rightarrow \quad x = 2{,}228$

Sampling of finite populations

Def: A sample of size n of a population of size N is a random sample if it is chosen such that each of the $\binom{N}{n}$ population subsets of size n is equally likely.

Example: Population $\{a, b, c\}$

random samples of size 2 $\{a,b\}, \{a,c\}, \{b,c\}$

$$p = \frac{1}{\binom{N}{n}} = \frac{1}{3}$$

A random sample can be chosen sequentially by first choosing randomly among N elements, then $N-1$, etc.

Example:

$\{a\ b\ c\} \longrightarrow a \qquad p = \frac{1}{3}$

$\Downarrow$

$\{b\ c\} \longrightarrow b \qquad p = \frac{1}{2}$

$$P((a,b)) = \frac{1}{6}$$

$$P(\{a,b\}) = P((a,b)) + P((b,a)) = \frac{1}{3}$$

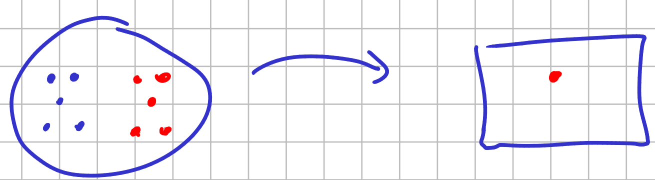
Remark

When chosing sequentially, the elements of the sample are not indipendent!

Example

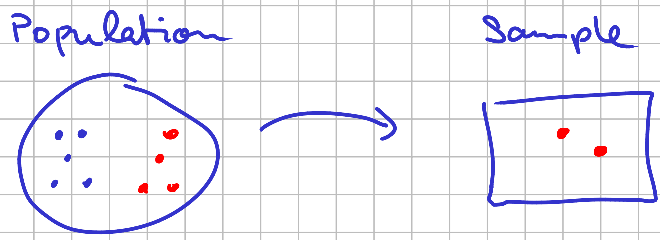
Population

Sample



$P(X_1 = \bullet) = \frac{1}{2}$

$P(X_1 = \bullet) = \frac{1}{2}$



$$P(X_2 = \bullet \mid X_1 = {\color{red}\bullet}) = 5/9$$

$$P(X_2 = {\color{red}\bullet} \mid X_1 = {\color{red}\bullet}) = 4/9$$

total probability

while $\quad P(X_2 = \bullet) = P(X_2 = \bullet \mid X_1 = \bullet)\, P(X_1 = \bullet)$

$$+ P(X_2 = \bullet \mid X_1 = {\color{red}\bullet})\, P(X_1 = {\color{red}\bullet})$$

$$= 4/9 \cdot \frac{1}{2} + 5/9 \cdot \frac{1}{2} = \frac{9}{18} = \frac{1}{2}$$

$\Rightarrow$ Thus $\quad P(X_2 = \bullet \mid X_1 = {\color{red}\bullet}) \neq P(X_2 = \bullet)$

not independent!

However:

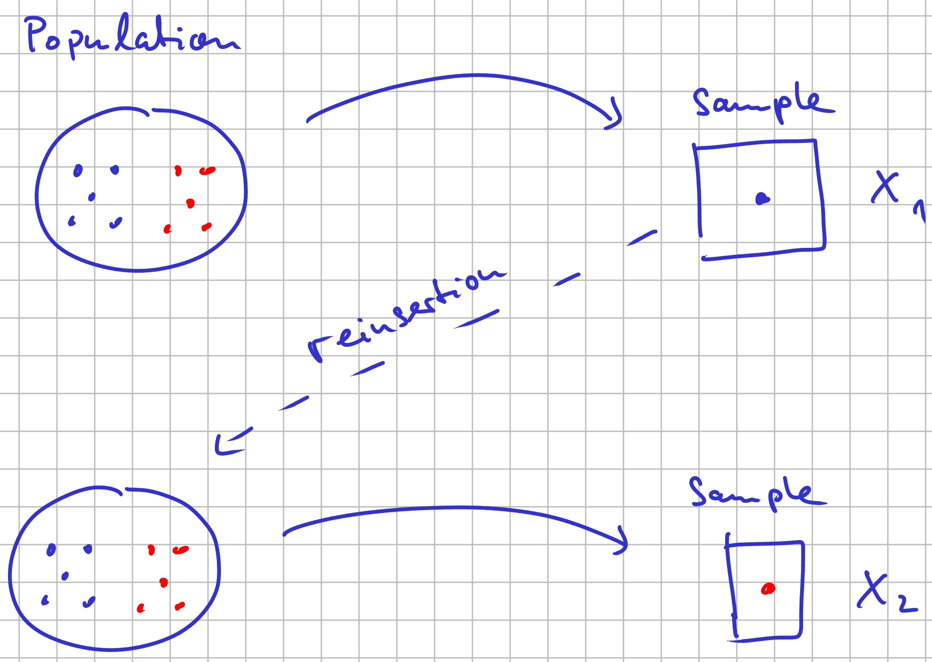
$$\boxed{\text{When the population size } N \text{ is large with respect to the sample size } n, \text{ the variables } X_1, \dots, X_n \text{ are approximately independent}}$$

Recall rule of thumb: Hyp $\to$ Bin for $\frac{n}{N} < 0{,}05$

Remark:

An independent sampling of finite populations is obtained by picking randomly elements from a population and reinserting them into the population before the next extraction (Sampling with replacement)

E.g.



The number of red balls is a Hypergeometric r.v. for sampling without replacement and Binomial r.v. with replacement.

Maximum likelihood (discrete case)

Idea:

Given a sample $X_1, \dots, X_n$ with values $x_1, \dots, x_n$ and an unknown parameter θ, determine θ such that the outcome $x_1, \dots, x_n$ is most likely.

Procedure:

Let $f(x; \theta_1, \dots, \theta_k)$ be a given probability mass fuction of the population X, $P(X = x) = f(x; \theta_1, \dots, \theta_k)$.
The likelihood function is defined as the joint probability mass fuction of a sample $X_1, \dots, X_n$ as

$$L(\theta_1, \dots, \theta_k) := P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} f(x_i; \theta_1, \dots, \theta_k)$$

$\rightarrow$ Maximise $L(\theta_1, \dots, \theta_k)$ for given $x_1, \dots, x_n$ via differentiation

$$\boxed{\frac{\partial L}{\partial \theta_1} \overset{!}{=} 0, \dots, \frac{\partial L}{\partial \theta_k} \overset{!}{=} 0}$$

(necessary but not sufficient)
$\leadsto$ find global maximum

Def: The parameters $\hat{\theta}_1, \dots, \hat{\theta}_k$ for which $L(\theta_1, \dots, \theta_k)$ is maximal is called maximum likelihood estimator set.

Remark: It is common to define the Log-likelihood function $\log L(\theta_1, \dots, \theta_k)$ that has the same maxima (minima) as $L(\theta_1, \dots, \theta_k)$.

__Bernoulli distribution__ : $X \sim Ber(p)$ $\quad p = \theta$ unknown

Sample size n with $P(X_i = 1) = p$ and $P(X_i = 0) = 1-p$ and X_i iid for $i = 1, \dots, n$.

We can write $P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$

$$\Rightarrow \quad P(X_i = 1) = p$$
$$P(X_i = 0) = 1-p$$

Then

$$L(p) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n - \sum_{i=1}^{n} x_i}$$

and thus

$$\log L(p) = \left(\sum_{i=1}^{n} x_i\right)\log(p) + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)$$

Maximise:

$$\frac{\partial}{\partial p} \log L(p) = \frac{1}{p}\left(\sum_{i=1}^{n} x_i\right) - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} x_i\right) \overset{!}{=} 0$$

$$\Leftrightarrow \quad (1 - \cancel{p})\left(\sum_{i=1}^{n} x_i\right) = p\left(n - \cancel{\sum_{i=1}^{n} x_i}\right)$$

$$\Leftrightarrow \quad p = \frac{1}{n}\sum_{i=1}^{n} x_i$$

We obtain the sample mean as estimator for the mean $\mu = E(X) = p$

$$d(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Example: A coin is thrown 100 times, it yields 82 times head and 18 times tail. Thus the estimate of p is

$$p = \frac{1}{100} 82 = 0.82$$

Poisson distribution: $X \sim Pois(\lambda)$ $\quad \lambda = \theta$ unknown

$$P(X_i = x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \qquad \text{with } x_i \in \mathbb{N}_0$$

$$L(\lambda) = \prod_{i=1}^{n} P(X_i = x_i) = e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \frac{1}{x_1! \cdots x_n!}$$

$$\Rightarrow \log L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \log(\lambda) - \log(x_1! \cdots x_n!)$$

$$\Rightarrow \frac{\partial}{\partial \lambda} \log L(\lambda) = -n + \frac{1}{\lambda}\left(\sum_{i=1}^{n} x_i\right) \overset{!}{=} 0$$

$$\Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Again we obtain the sample mean as estimator for the mean $\mu = E(X) = \lambda$

$$d(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

This however is not always the case!

Another example: Consider the distribution

$$P(X=k) = \begin{cases} \frac{1-\theta}{3} & \text{for } k=0 \\ \frac{1}{3} & \text{for } k=1 \\ \frac{1+\theta}{3} & \text{for } k=2 \\ 0 & \text{else} \end{cases}$$

For $X_1, \dots, X_n$ iid estimate θ.

Denote n_k the number of that X takes value k such that $n_0 + n_1 + n_2 = n$. Then

$$L(\theta) = \left(\frac{1-\theta}{3}\right)^{n_0} \left(\frac{1}{3}\right)^{n_1} \left(\frac{1+\theta}{3}\right)^{n_2}$$

$$\Rightarrow \log L(\theta) = n_0 [\log(1-\theta) - \log(3)] + n_1 [\log(1) - \log(3)] + n_2 [\log(1+\theta) - \log(3)]$$

$$\Rightarrow \frac{\partial}{\partial\theta} \log L(\theta) = -\frac{n_0}{1-\theta} + \frac{n_2}{1+\theta} \overset{!}{=} 0$$

$$\Rightarrow (1+\theta) n_0 = (1-\theta) n_2$$

$$\Leftrightarrow \theta(n_0 + n_2) = n_2 - n_0$$

$$\Rightarrow \theta = \frac{n_2 - n_0}{n_0 + n_2}$$

Here $\mu = \frac{1}{3} + 2\frac{1+\theta}{3} = 1 + \frac{2}{3}\theta = 1 + \frac{2}{3} \frac{n_2 - n_0}{n_0 + n_2} = 1 + \frac{2}{3} \frac{2n_2 - n + n_1}{n - n_1}$

$$= \frac{3n - 3n_1 + 2n_2 - 2n + 2n_1}{3n - 3n_1} = \frac{n - n_1 + 2n_2}{3n - 3n_1}$$

$$\neq \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(n_1 + 2n_2)$$

Likelihood for continuous distributions

As before but likelihood function is defined in terms of probability density $f(x;\theta_1,\dots,\theta_k)$

$$L(\theta_1,\dots\theta_k) = \prod_{i=1}^{n} f(x_i;\theta_1,\dots,\theta_k)$$

for sample values $x_1,\dots,x_n$.

Normal distribution

$X \sim N(\mu,\sigma^2)$

$\rightarrow$ estimate $\theta_1=\mu$, $\theta_2=\sigma$

Consider a sample $x_1,\dots,x_n$ iid :

$$L(\mu,\sigma) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}_{f(x_i;\mu,\sigma)}$$

$\log(1)=0$

$$\Rightarrow \log L(\mu,\sigma) = n\left[\log\frac{1}{\sqrt{2\pi}} - \log\sigma\right] - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

(1) $$\frac{\partial}{\partial\mu}\log L(\mu,\sigma) = +\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \overset{!}{=} 0$$

(2) $$\frac{\partial}{\partial\sigma}\log L(\mu,\sigma) = -\frac{n}{\sigma} + \sigma^{-3}\sum_{i=1}^{n}(x_i-\mu)^2 \overset{!}{=} 0$$

from (1) $\Rightarrow$ $$\sum_{i=1}^{n}(x_i-\mu) = -n\mu + \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

from (2) $\Rightarrow \quad \sigma^{-2} \sum_{i=1}^{n} (x_i - \mu)^2 = n$

$$\Leftrightarrow \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Thus we obtain the estimators

$$d_1(X_1, \ldots, X_n) = \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

sample variance

and $d_2(X_1, \ldots, X_n) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2} \neq S_n$

biased unbiased

Exponential distribution : $X \sim Exp(\lambda)$ estimate λ

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\Rightarrow \log L(\lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial \lambda} \log L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \overset{!}{=} 0$$

$$\Rightarrow \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Interval estimation

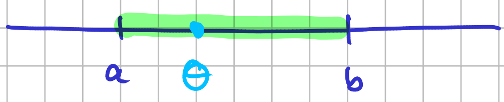
The point estimates, sample mean, sample variance and maximum likelihood estimator, give an estimate for a parameter.
But how accurate is this estimate?

$\rightarrow$ Intervall estimation

Determine $a, b \in \mathbb{R}$ such that $a \leq \theta \leq b$ with a certain probability when repeating the experiment.

The interval (a,b) is called confidence interval.

e.g. 99% probability that θ is contained in the interval (a,b).



Def: Given $0 < \alpha < 1$ and parameter θ, a two-sided (symmetric) $(1-\alpha)$-confidence interval is an interval (Θ_-, Θ_+) where $\Theta_\pm$ are estimators (random variables), such that

capital letters

$$\boxed{P(\theta \in (\Theta_-, \Theta_+)) = 1-\alpha}$$

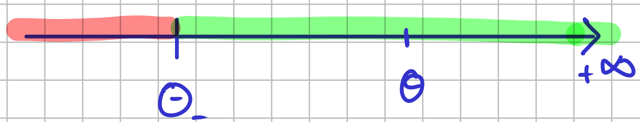
The probability $1-\alpha$ is called confidence level.

One-sided confidence intervals

- Upper confidence interval $(\Theta_-, +\infty)$

$$P(\theta \in (\Theta_-, +\infty)) = P(\theta \geq \Theta_-) = 1-\alpha$$

Probability that θ is at least Θ_-.



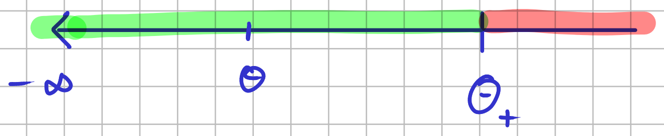
Right interval

- Lower confidence interval $(-\infty, \Theta_+)$

$$P(\theta \in (-\infty, \Theta_+)) = P(\theta \leq \Theta_+) = 1-\alpha$$

Probability that θ is at most Θ_+.

Left interval



Confidence intervall for the mean of a normal population with known variance

Proposition:

Consider $X \sim N(\mu, \sigma^2)$ with σ^2 known. The two-sided $(1-\alpha)$-confidence interval of the population mean μ given a sample $X_1, \dots, X_n$ is determined by the estimators

$$\boxed{\Theta_\pm = \bar{X}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

i.e. $P(\mu \in (\Theta_-, \Theta_+)) = 1-\alpha$, where $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

and sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

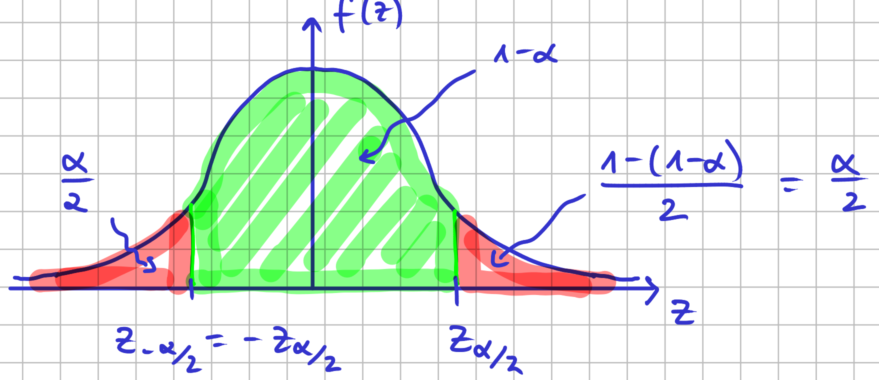
Proof:

Recall that estimator $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$

such that $Z = \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0,1)$

It follows that $P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\alpha/2}) = 1-\alpha$:

$N(0,1)$



$\Phi(-z_{\alpha/2}) = \frac{\alpha}{2}$

$\Phi(z_{\alpha/2}) = \frac{\alpha}{2} + 1 - \alpha$

$= 1 - \frac{\alpha}{2}$

This yields the confidence interval for the population mean:

$$1-\alpha = P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = P\left(-z_{\alpha/2} \le \frac{\bar{X}_n-\mu}{\sigma}\sqrt{n} \le z_{\alpha/2}\right)$$

$$= P\left(-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X}_n \le -\mu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X}_n\right)$$

$$= P\Big(\underbrace{\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\Theta_-} \le \mu \le \underbrace{\bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\Theta_+}\Big)$$

$$= P(\mu \in (\Theta_-, \Theta_+))$$

Remark: For desired α we determine $z_{\alpha/2}$ via

$$\phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

Example: $1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05$

$$\phi(z_{\alpha/2}) = 0{,}975$$

$$\xrightarrow{\text{Table}} \quad \underline{\underline{z_{\alpha/2} = 1.96}}$$

Example: Suppose a signal with value μ is send from Alice to Bob. Bob receives Y where Y is a random variable $Y \sim N(\mu, 4)$ with $\sigma^2 = 4$ denoting some "noise". Alice sends 9 times the same signal. Bob receives 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5. Construct the 95%-confidence interval for μ.

$$1-\alpha \overset{!}{=} 0.95 \quad \rightsquigarrow \quad z_{\alpha/2} = 1.96 \quad \text{(as above)}$$

$$\bar{x}_9 = \frac{1}{9}(5 + 8.5 + \ldots + 10.5) = \frac{81}{9} = 9$$

$$\sigma = \sqrt{4} = 2$$

$\rightsquigarrow$ 95% - confidence interval

$$\mu \in \left(9 - 1.96 \frac{2}{\sqrt{9}} , 9 + 1.96 \frac{2}{\sqrt{9}} \right)$$

$$= (7.6933 , 10.3007)$$

$\rightsquigarrow$ 95% confidence that the sent μ lies between 7.7 and 10.3.

One-sided confidence intervall for the mean of a normal population with known variance

Proposition:

Consider $X \sim N(\mu, \sigma^2)$ with σ^2 known. The upper (lower) $(1-\alpha)$- confidence interval of the population mean μ given a sample $X_1, \dots, X_n$ is determined by the estimator

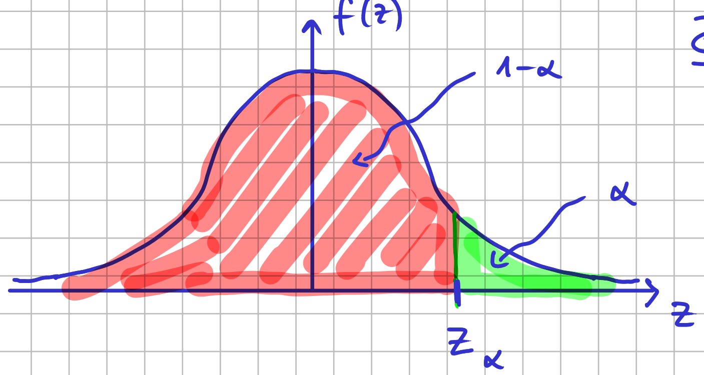
$$\Theta_- = \bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}} \qquad \left(\Theta_+ = \bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}} \right)$$

i.e. $P(\mu \in (\Theta_-, +\infty)) = 1-\alpha$ $\left(P(\mu \in (-\infty, \Theta_+)) = 1-\alpha \right)$,

, where $\Phi(z_\alpha) = 1-\alpha$ and sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Proof (upper interval):

As before $Z \sim N(0,1)$



$$\Phi(z_\alpha) = 1-\alpha$$

$$1-\alpha = P(Z \le z_\alpha) = P\left(\frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \le z_\alpha\right) = P\left(-\mu \le z_\alpha \frac{\sigma}{\sqrt{n}} - \bar{X}_n\right)$$

$$= P(\mu \ge \underbrace{\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}}_{\Theta_-}) = P\left(\mu \in \left(\bar{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right)\right)$$

Confidence interval estimate $\left(\bar{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right)$

Proof: (lower interval)

$$1-\alpha = P(Z \le z_\alpha) = \phi(z_\alpha) = 1 - \phi(-z_\alpha) = 1 - P(Z \le -z_\alpha)$$

$$= P(Z \ge -z_\alpha) = P\left(\frac{\bar{X}_n - \mu}{\sigma}\sqrt{n} \ge -z_\alpha\right) = P\left(-\mu \ge -z_\alpha \frac{\sigma}{\sqrt{n}} - \bar{X}_n\right)$$

$$= P(\mu \le \underbrace{\bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}}_{\Theta_+}) = P\left(\mu \in \left(-\infty, \bar{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)\right)$$

Confidence intervall for the mean of a normal population with unknown variance

Proposition:

Consider $X \sim N(\mu, \sigma^2)$ with σ^2 unknown. The two-sided $(1-\alpha)$-confidence interval of the population mean μ given a sample $X_1, \dots, X_n$ is determined by the estimators

$$\boxed{\Theta_\pm = \bar{X}_n \pm t_{\alpha/2, n-1} \frac{S_n}{\sqrt{n}}}$$

i.e. $P(\mu \in (\Theta_-, \Theta_+)) = 1-\alpha$, where $P(T_{n-1} \le t_{\frac{\alpha}{2}, n-1}) = 1 - \frac{\alpha}{2}$ for Student's t-distribution T_{n-1} of degree $n-1$, sample mean $\bar{X}_n$ and sample variance S_n^2.

Proof: Recall $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

such that $T_{n-1} = \frac{\bar{X}_n - \mu}{S_n}\sqrt{n}$ for $X_i \sim N(\mu, \sigma^2)$.

Proceed as above:

$$1-\alpha = P(-t_{\frac{\alpha}{2},n-1} \leq T_{n-1} \leq t_{\frac{\alpha}{2},n-1}) = \ldots =$$
$$= P(\bar{X}_n - t_{\frac{\alpha}{2},n-1}\frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{\frac{\alpha}{2},n-1}\frac{S_n}{\sqrt{n}})$$

$\Rightarrow$ confidence intervall $\left(\bar{x}_n - t_{\frac{\alpha}{2},n-1}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{\frac{\alpha}{2},n-1}\frac{s_n}{\sqrt{n}}\right)$

Example as above for σ unknown and the sample values $x = \{5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5\}$.

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1-\frac{\alpha}{2} = 0.975$

Determine $t_{\frac{\alpha}{2},n-1}$: $P(T_{n-1} \leq t_{\frac{\alpha}{2},n-1}) = 0.975$

$$\leadsto t_{\frac{\alpha}{2},n-1} = 2.3060$$

$\bar{x}_9 = 9$ $\quad s_9 = \sqrt{\frac{1}{8}\left[(5-9)^2 + (8.5-9)^2 + \ldots + (10.5-9)^2\right]}$

$= 3.082$

$$\leadsto \mu \in \left(9 - 2.306 \cdot \frac{3.082}{3}, 9 + 2.306 \cdot \frac{3.082}{3}\right)$$
$$= (6.63, 11.37)$$

Confidence intervall for the variance of a normal population with unknown mean

Proposition:

Consider $X \sim N(\mu, \sigma^2)$ with μ unknown. The two-sided $(1-\alpha)$-confidence interval of the population variance σ^2 given a sample $X_1, \dots, X_n$ is determined by the estimators

$$\boxed{\Theta_\pm = \frac{(n-1)S_n^2}{\chi^2_{\frac{1}{2} \pm \left(\frac{1-\alpha}{2}\right), n-1}}}$$

i.e. $P\left(\sigma^2 \in (\Theta_-, \Theta_+)\right) = 1-\alpha$, where

$$P\left(\chi^2_{n-1} \leq \chi^2_{\frac{1}{2} \pm \left(\frac{1-\alpha}{2}\right), n-1}\right) = 1 - \left(\frac{1}{2} \pm \frac{1-\alpha}{2}\right) = \frac{1}{2} \mp \frac{1-\alpha}{2}$$

for χ^2-distribution of degree $n-1$ and sample variance S_n^2.

Proof:

Recall $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

such that $\chi^2_{n-1} = (n-1) \frac{S_n^2}{\sigma^2}$ for $X_i \sim N(\mu, \sigma^2)$.

As before

$$1-\alpha = P\left(\chi^2_{1-\frac{\alpha}{2}, n-1} \leq (n-1)\frac{S_n^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}, n-1}\right) = \dots =$$

$$= P\left(\frac{(n-1)S_n^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{1-\alpha/2, n-1}}\right)$$

Summary (S. Ross)

Table 7.1 $100(1-\alpha)$ Percent Confidence Intervals

$X_1,\dots,X_n \sim \mathcal{N}(\mu,\sigma^2)$ $\quad \overline{X}=\sum_{i=1}^{n} X_i/n, \quad S=\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2/(n-1)}$

Assumption	Parameter	Confidence Interval	Lower Interval	Upper Interval
σ^2 known	μ	$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\left(-\infty,\ \overline{X}+z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$	$\left(\overline{X}-z_{\alpha}\frac{\sigma}{\sqrt{n}},\ \infty\right)$
σ^2 unknown	μ	$\overline{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$	$\left(-\infty,\ \overline{X}+t_{\alpha,n-1}\frac{S}{\sqrt{n}}\right)$	$\left(\overline{X}-t_{\alpha,n-1}\frac{S}{\sqrt{n}},\ \infty\right)$
μ unknown	σ^2	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right)$	$\left(0,\ \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}},\ \infty\right)$

where $P(Z \leq z_\alpha) = 1-\alpha$ etc.

Remark: For non-normal populations and large sample size such that as a consequence of the C.L.T. we have $\overline{X}_n \underset{n\ \text{large}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ the confidence interval for μ can be obtained as above. It is determined by the estimators

$$\boxed{\Theta_\pm = \overline{X}_n \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}}}$$

where we note that $t_{\alpha/2,n-1} \underset{n\ \text{large}}{\approx} z_{\alpha/2}$.

Confidence interval for the mean of a Bernoulli random variable

Proposition:

Consider $X \sim Ber(p)$ with p unknown. The two-sided $(1-\alpha)$-confidence interval of the population mean $\mu = p$ given a sample $X_1, \dots, X_n$ with n large is determined by the estimators

$$\boxed{\Theta_{\pm} = \bar{X}_n \pm z_{\alpha/2} \cdot \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}}$$

i.e. $P(\mu \in (\Theta_-, \Theta_+)) = 1-\alpha$, where $\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$ and sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Proof: Noting that $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ for large n, where

$\mu = p$ and $\sigma^2 = p(1-p)$ we have that

$$Z = \frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\sqrt{n} \approx \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \sim N(0,1)$$

Then as before

$$1-\alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \dots =$$

$$= P(\mu \in (\Theta_-, \Theta_+))$$

Example: A sample of 100 transitors is randomly chosen from a large batch and tested. Assuming that 80 of them pass the test, compute the 95 % confidence interval.

$n=100$, $\bar{X}_n = \frac{1}{100} 80 = 0.8$, $1-\alpha = 0.95$

$\Rightarrow z_{\alpha/2} = 1.96$

$$p \in \left(0.8 - 1.96\sqrt{\frac{0.8 \times 0.2}{100}}, 0.8 + 1.96\sqrt{\frac{0.8\times 0.2}{100}}\right)$$

$$= (0.7216, 0.8784)$$

with 95 % probability between 72 % and 88 % of the transistors pass the test.

Example: A newspaper reports that a recent poll indicates, that with a probability of 95 %, a percentage of 52 % of people are in favor of the president with a margin of error of ±4 %. How many people where questioned?

Confidence interval

$$(\underset{\overset{\shortparallel}{\bar{X}_n}}{0.52} - 0.04, 0.52 + 0.04)$$

$$\leadsto \quad 0{,}04 = 1.96\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} = 1.96\sqrt{\frac{0.52 \times 0.48}{n}}$$

$$\Rightarrow \quad n = 599.29$$

Confidence interval for the difference of two normal populations

Setup: Consider two independent normal populations X & Y

- Let $X_1, \dots, X_n$ be a sample from a population $X \sim N(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_m$ be a sample from a population $Y \sim N(\mu_2, \sigma_2^2)$.

We are interested in estimating $\mu_1 - \mu_2$

Sample mean estimator of μ_1 : $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu_1, \frac{\sigma_1^2}{n})$

Sample mean estimator of μ_2 : $\bar{Y}_m = \frac{1}{m} \sum_{i=1}^{m} Y_i \sim N(\mu_2, \frac{\sigma_2^2}{m})$

Proposition :

The maximum likelihood estimator of $\mu_1 - \mu_2$ is $\Delta = \bar{X}_n - \bar{Y}_m$.

Proof:
$$L(\mu_1, \mu_2) = \left(\prod_{i=1}^{n} f_{X_i}(x_i) \right) \left(\prod_{i=1}^{m} f_{Y_i}(y_i) \right)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma_1} \right)^n \left(\frac{1}{\sqrt{2\pi}\,\sigma_2} \right)^m \exp\left[-\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} (x_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{j=1}^{m} (y_j - \mu_2)^2 \right]$$

Set $\delta = \mu_1 - \mu_2$

$$\left.\begin{aligned} \frac{\partial}{\partial \delta} L(\delta, \mu_2) &\stackrel{!}{=} 0 \\ \frac{\partial}{\partial \mu_2} L(\delta, \mu_2) &\stackrel{!}{=} 0 \end{aligned}\right\} \quad \begin{aligned} \mu_2 &= \bar{y}_m \\ \delta &= \bar{x}_n - \bar{y}_m \end{aligned}$$

$$\Rightarrow \quad \Delta = \bar{X}_n - \bar{Y}_m$$

Properties of Δ

1) $E(\Delta) = E(\bar{X}_n) - E(\bar{Y}_m) = \mu_1 - \mu_2$ unbiased

2) $Var(\Delta) = Var(\bar{X}_n) + Var(\bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$

and thus $\Delta \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$

Confidence intervals for Δ

Proposition

For σ_1^2 & σ_2^2 known the confidence interval for $\Delta = \bar{X}_n - \bar{Y}_m$ is (Θ_-, Θ_+) where

$$\boxed{\Theta_{\pm} = (\bar{X}_n - \bar{Y}_m) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

and $\phi(z_{\alpha/2}) = 1 - \alpha/2$.

Proof: since $\bar{X}_n - \bar{Y}_m \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$

we have $Z = \dfrac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1)$

Then as before

$$1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = \dots =$$

$$= P(\Theta_- \leq \mu_1 - \mu_2 \leq \Theta_+)$$

$$= P((\mu_1 - \mu_2) \in (\Theta_-, \Theta_+))$$

Proposition

For $\sigma_1^2 = \sigma_2^2 = \sigma^2$ equal but unknown the confidence interval for $\Delta = \bar{X}_n - \bar{Y}_m$ is (Θ_-, Θ_+) where

$$\boxed{\Theta_\pm = (\bar{X}_n - \bar{Y}_m) \pm t_{\frac{\alpha}{2}, n+m-2} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}\, S_p}$$

with pooled variance $S_p^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$

and $P(T \leq t_{\frac{\alpha}{2}, n+m-2}) = 1 - \frac{\alpha}{2}$ where T denotes Student's t-distribution of degree $n+m-2$.

Proof: Estimator of σ_1^2 : $S_1^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)$

Estimator of σ_2^2 : $S_2^2 = \dfrac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y}_m)$

such that

$$\frac{(n+m-2)}{\sigma^2} S_p^2 \sim \chi^2_{n+m-2}$$

and

$$T = \frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n} + \frac{1}{m}}\, S_p} \sim \text{Student's t-distr. of degree } n+m-2$$

(without proof)

Then we proceed as before

$$1 - \alpha = P(-t_{\alpha/2, n+m-2} \leq T \leq t_{\alpha/2, n+m-2}) = \dots =$$

$$= P(\mu_1 - \mu_2 \in (\Theta_-, \Theta_+))$$

Proposition

For σ_1^2 and σ_2^2 unknown the confidence interval for $\Delta = \bar{X}_n - \bar{Y}_m$ is approximately (Θ_-, Θ_+) where

$$\boxed{\Theta_{\pm} = (\bar{X}_n - \bar{Y}_m) \pm t_{\frac{\alpha}{2}, \ell} \sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

with $P(T \leq t_{\frac{\alpha}{2}, \ell}) = 1 - \frac{\alpha}{2}$ where T denotes Student's t-distribution with ℓ degrees of freedom given by the integer part

$$\ell = \left[\frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m} \right)^2}{\frac{1}{n-1} \left(\frac{S_1^2}{n} \right)^2 + \frac{1}{m-1} \left(\frac{S_2^2}{m} \right)^2} \right].$$

Proof: based on the observation

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim$$

Approximately Student's t-distribution with ℓ d.o.f. as given above