1 Probability Theory

Probability theory is used as a tool in statistics. It helps to evaluate the reliability of our conclusions about the population when we have only information about a sample.
Probability theory is also the main tool in developing a model for describing the populations in inferential statistics. (In inferential statistics the information from a random sample is used to make statements about an entire population.)

Many things in our world depend on randomness, if you observe Head while flipping a coin or you roll a 6 with a "fair" die, if the bus is on time, if the baseball player will hit the ball, if the stock gains at least 10% in a year, etc.. Even if events occur randomly there is an underlying pattern in the occurrence of these events. This is the basis of Probability Theory.

1.1 Introduction

We now provide the vocabulary for probability theory.

Definition:

1. A phenomenon is random if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions.
2. The probability of any outcome (or event) of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

Probability Models
Definition:

1. The **sample space** of an experiment is the set of all possible outcomes.
2. An **event** is a subset of the sample space.

Example:
Examples for random phenomenons are:

- A test grade
- Measure daily snowfall
- Count of leucocytes in a blood sample
- Roll of a die
- Toss of a coin

The sample spaces of above random phenomenon are:

- $\{A, B, C, D, E, F\}$
- All numbers between 0 and 200 cm.

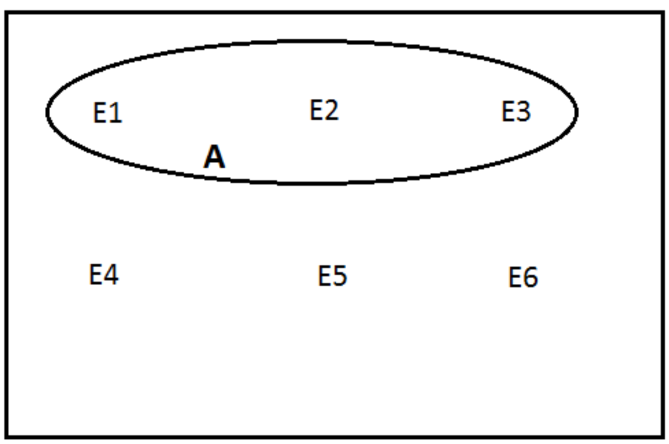
- 1 - 10000/ml ??
- $\{1, 2, 3, 4, 5, 6\}$
- $\{H, T\}$

Examples for events after tossing a die and observing the number on the upper face are:

- Observe a 1 $=\{1\}$
- Observe an odd number $=\{1, 3, 5\}$
- Observe 1 or 6 $=\{1, 6\}$

Example: Suppose you roll an unbiased die with 6 faces and observe the outcome.

Venn Diagrams
The outer box represents the sample space, which contains all of the simple events. Is $A = \{E_1, E_2, E_3\}$ the collection of simple events the appropriate events are circled and labelled with the letter A.



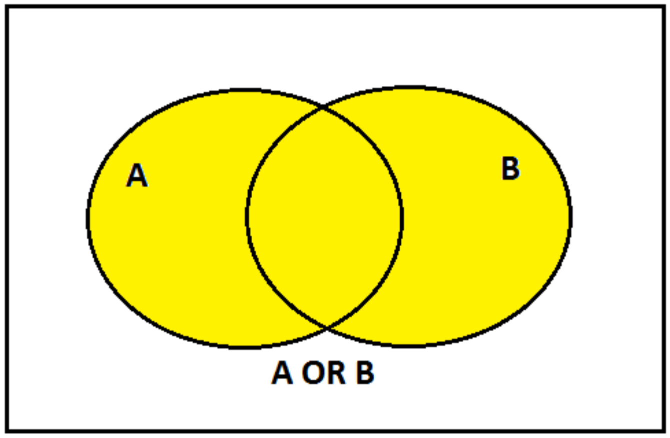
Basic Properties:

- The probability of an event is a number between 0 and 1, since it is a limit of a relative frequency.
- $P(E) = 0$, if the event E never occurs. P(role a 7 with a regular die)=0
- $P(E) = 1$, if the event always occurs. P(role with a regular die a number smaller than 7)=1
- The sum of the probabilities for all simple events in S equals 1.

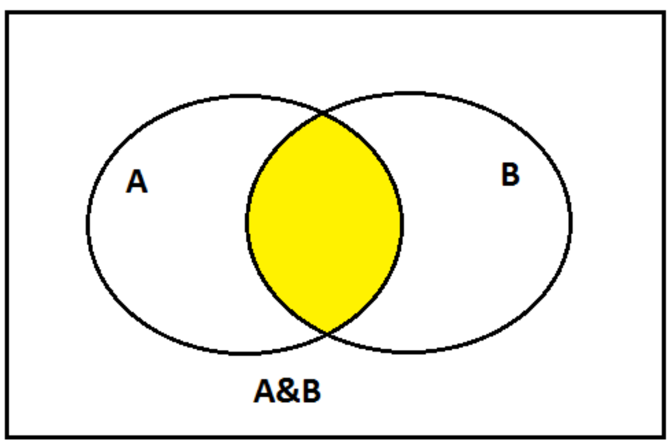
1.2 Basic Rules

Definition:

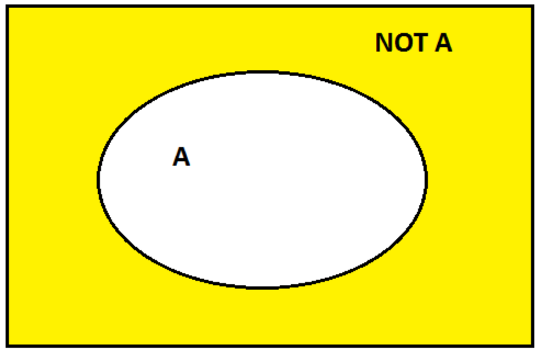
- The probability of an event A is equal to the sum of the probabilities of the simple events contained in A.
- The union of events A and B, denoted by $A\text{or}B$, is the event that either A or B or both occur.



- The intersection of events A and B, denoted by $A\&B$, is the event that both A and B occur.



- The complement of an event A, denoted by $\text{not}A$, is the event that A does not occur.



- Two events, A, B, are called mutually exclusive, if $A\&B = \emptyset$.

Example: Continue the die rolling example.

C is the event to roll an odd number, $C = \{1,3,5\}$. Let D be the event to roll a number smaller than 4, $D = \{1,2,3\}$.

The probability of every simple event is 1/6.

With the definition follows that $P(C) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 1/6 + 1/6 + 1/6 = 3/6 = 1/2$.

$P(C \text{or} D) = P(\{1,3,5,2\}) = 4/6 = 2/3$.

$P(C\&D) = P(\{1,3\}) = 2/6 = 1/3$.

$P(\text{not} D) = P(\{4,5,6\}) = 3/6 = 1/2$.

If E is the event to roll a 2, then events C and E are mutually exclusive.

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!11 One more example !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!1

Remark:
Until this point: In order to calculate the probability of an event,

1. find all the simple events, that belong to the event,
2. find the probabilities for those simple events and
3. add those probabilities to derive the probability for the event of interest.

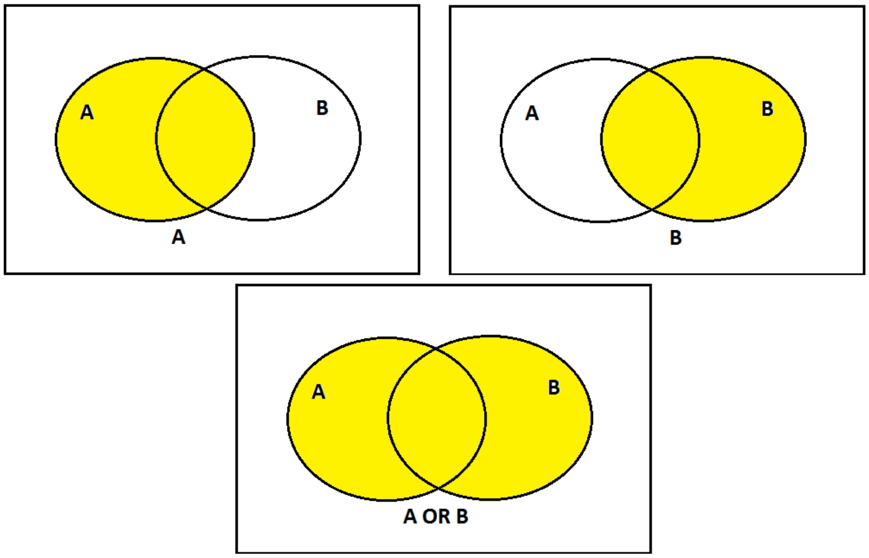
The following rules provide tools for the determination of the probabilities of events like $A \cup B$ and A^c.

Addition Rule

Let A and B be two events. The probability of the union can then be calculated by:

$$P(A \text{ oder B}) = \text{P(AorB)} = \text{P(A)} + \text{P(B)} - \text{P(A\&B)}$$

Remark:
The subtraction of $P(A\&B)$ is necessary because this area is counted twice by the addition of $P(A)$ and $P(B)$, once in $P(A)$ and once in $P(B)$. Check the diagram below.



Complementation Rule (Rule for Complements)

The probability of the complement of an event is 1 minus the probability of this event.

Mathematically:
If $P(\text{not}A) = 1 - P(A)$

Example:

Let $A = \{1, 2\}$ and $B = \{1, 3, 5\}$, with $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, and $P(A\&B) = P(\{1\}) = \frac{1}{6}$.
With the Addition Rule we get $P(A\text{or}B) = P(A) + P(B) = 1/3 + 1/2 - 1/6 = 2/3$.

With the Rule for Complements is $P(\text{not}A) = 1 - 1/3 = 2/3$.

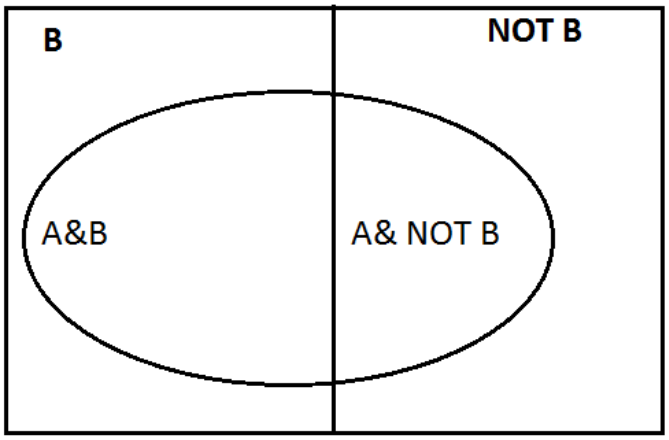
1.3 Independence of Two Events

Two events are considered independent, when the occurrence of one of the events has no impact on the probability for the second event to occur. They don't have an impact on each other.

In this section we will define and do examples to illustrate this concept. For defining the independence of events properly we first introduce the concept of a conditional probability.

Definition
If A and B are events with $P(B) > 0$, the conditional probability of A given B is defined by

$$P(A|B) := \frac{P(A\&B)}{P(B)}.$$



The interpretation of the conditional probability $P(A|B)$ is as follows:
Given that you know already event B occurred, what is the probability that A occurs.

Example:
Consider the event B to roll an even number with a fair 6 sided die.
The Probability for the event A to roll a 1, given B equals 0: $P(A|B) = 0$
Because: If you roll a die, someone peeks and tells you that the number is even, than this number can not be a 1. So the conditional probability of A given B must be 0.

Another way to argue that this probability is as follows:
There are three favourable events in B, out of these zero are a one, therefore $P(A|B) = 0/3 = 0$

Another way to show the dependency of events on each other is in a two-way or contingency table.

Example:
Research has shown, that a mutation in the BRCA gene increases the probability of a woman to develop breast cancer. The following table shows data which is consistent with numbers reported in the Journal of the National Cancer Institute.

		BRCA Gene		
		yes	no	Total
Breast Cancer	yes	16	609	625
	no	4	4371	4375
	Total	20	4980	5000

Then

- P(Breast Cancer = yes) =625/5000=0.125 (marginal probability (ignore the mutation of the gene))
- P(BRCA Gene = yes) =20/5000=0.004 (marginal probability (ignore breast cancer))
- P(BRCA Gene = yes & Breast Cancer = yes) = 16/5000 = 0.0032 (joint probability)

- P(Breast Cancer = yes | BRCA Gene = yes) = 16/20 = 0.8 (conditional probability)
- P(BRCA Gene = yes | Breast Cancer = yes) =16/625=0.0256 (conditional probability)

Given these numbers indicate that a mutation of this gene has a great impact on the probability of developing breast cancer. The two events "mutation of the BRCA gene" and "Develop breast cancer" are not independent.

Joint and Marginal Probabilities

- joint: consider event $P(A\&B)$
- marginal: only consider one event at a time. $P(A), P(B)$ are both marginal probabilities.

In probability theory two events are considered independent, if the knowledge that one of them has occurred does not change the probability for the second to occur.

Definition:
Events A and B, with $P(B) > 0$, are independent if

$$P(A|B) = P(A),$$

(which is equivalent to $P(A\&B) = P(A) \cdot P(B)$).

Remark:

- From this definition we also learn that: If we know that two events A and B are independent then
$$P(A\&B) = P(A)P(B)$$
This is only true for INDEPENDENT events.
- In the case, we do not know, if two events are independent, the general multiplication rule applies:
$$P(A\&B) = P(A|B)P(B) = P(B|A)P(A)$$

The following example shows how to apply these two two concepts to the experiment of rolling a die.

Example:
Consider the experiment to roll a fair die.
Let $A = \{1, 2, 3\}$
$B = \{4\}$
$C = \{3, 4\}$

Then we can calculate, by finding the simple events contained by the described events and adding their probabilities:

- $P(A) = \frac{1}{2}$

- $P(B) = \frac{1}{6}$
- $P(C) = \frac{1}{3}$
- $P(A\&B) = 0$
- $P(A\&C) = \frac{1}{6}$
- $P(B|A) = \dfrac{P(B\&A)}{P(A)} = \dfrac{0}{\frac{1}{2}} = 0$ (by definition of the conditional probability)

  Since $P(B) \neq P(B|A)$, we find that B and A are not independent (apply the definition for independence).

- $P(C|A) = \dfrac{P(C\&A)}{P(A)} = \dfrac{\frac{1}{6}}{\frac{1}{2}} = \dfrac{1}{3}$

  Since $P(C) = P(C|A)$, we find that C and A are independent.

Example:
Consider the experiment of rolling two unbiased dice.
The simple events of this experiment can be described as a pair of numbers between 1 and 6. $(3,4)$ would be interpreted as rolling a 3 with the first die and a 4 with the second die.
The sample space contains of all pairs with numbers between 1 and 6 in the first and the second component. So that the sample space consists of 6·6=36 elements.
Since both dice will fall independently we can calculate the probability of rolling two 6 by

$$P((6,6)) = P(6 \text{ with the first die}) \cdot \text{P}(6 \text{ with the second die}) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$

Example:
A satellite has two power systems, a main and an independent backup system.

Suppose the probability of failure in the first ten years for the main system is 0.05 and for the backup system 0.08.

What is the probability that both systems fail in the first 10 years and the satellite will be lost?

Let M be the event that the main system fails and B the event that the backup systems fails. Since the systems are independent we get:

$$P(M\&B) = P(M) \cdot P(B) = 0.05 \cdot 0.08 = 0.004$$

The event O that at least one of the systems is still operational after 10 years is the complement event of both systems failing. We calculate:

$$P(O) = 1 - P(\text{not}O) = 1 - 0.004 = 0.996$$

Example :
An experiment can result in one of five equally likely simple events E_1, E_2, E_3, E_4, E_5. Events A, B and C are defined as follows:

$A := \{E_1, E_3\}$
$B := \{E_1, E_2, E_4, E_5\}$
$C := \{E_3, E_4\}$
Find the probabilities for the following events:
a. not$A$ b. $A\&B$ c. $B\&C$
d. Aor$B$ e. $B|C$ f. $A|B$ g. Are events A and B independent?

Solution (first try on your own):
a. 3/5 b. 1/5 c. 1/5
d. 1 e. 1/2 f. 1/4 g. No they are not, since $P(A) \neq P(A|B)$

The following example shows the practical relevance of these concepts. In this example they are applied to an AIDS test, but please be aware that the same arguments apply to other clinical tests.
Example:

- ELISA is a test for detecting HIV antibodies in blood samples.
- The test is positive in 99.7% of blood samples containing antibodies.
- The test is positive in 1.5% of blood samples containing no antibodies.

Assuming that 1% of the population do have HIV antibodies in their blood, how big is the probability that a person has HIV if his/her blood tests positive?

For finding an answer to this question we translate the number given above into probability theory and apply the rules introduced.
First choose the denotation:

- Let H_+ be the event of a blood sample containing antibodies.
- Then is $H_- = H_+^C$ the event of a blood sample not containing antibodies.
- Let T_+ be the event of a positive ELISA test result.
- Then is $T_- = T_+^C$ the event of a negative ELISA test result.

Now "translate" the facts stated above into conditional probabilities.
We get from the introductory statements and the definition:
$P(T_+|H_+) = 0.997, P(T_+|H_-) = 0.015$, and $P(H_+) = 0.01$ and want to know $P(H_+|T_+)$.

Calculate using the definition:

$$P(H_+|T_+) = \frac{P(H_+\&T_+)}{P(T_+)}$$

Now calculate numerator and denominator separately.
First the numerator (applying the definition):

$$P(H_+\&T_+) = P(H_+)P(T_+|H_+) = 0.01 \cdot 0.997 = 0.00997$$

Second the denominator:

$$\begin{aligned}
P(T_+) &= P((T_+\&H_+) \text{ } or \text{ } (T_+\&H_-)) && \text{cutting T}_+\text{in two pieces} \\
&= P(T_+\&H_+) + P(T_+\&H_-) && \text{Addition Rule} \\
&= P(H_+)P(T_+|H_+) + P(H_-)P(T_+|H_-) && \text{definition of conditional probability} \\
&= 0.01 \cdot 0.997 + (1 - 0.01) \cdot 0.015 = 0.02482 && \text{Rule for Compliments}
\end{aligned}$$

Put the numerator and the denominator together and we get $P(H_+|T_+) = 0.4017$. This probability is very low and a test result would have to be confirmed with additional tests. This shows that ELISA is rather a test to exclude HIV then to test for HIV. Try on your own to calculate $P(H_-|T_-)$.

Since the probability for a positive ELISA test is different for blood samples with or without antibodies, we conclude that the probability for a positive ELISA test is different in the sample space than it is for positive blood samples.
We can say that the event "positive ELISA test" is NOT independent from the event "antibodies in the blood sample".

2 Probability Distributions

In the chapter about descriptive statistics *samples* were discussed, and tools introduced for describing the *samples* with numbers as well as with graphs.

In this chapter models for the *population* will be introduced. One will see how the properties of a population can be described in mathematical terms. Later we will see how samples can be used to draw conclusions about those properties. That step is called statistical inference.

Definition:
A variable X (we use capital letters for random variables) is a *random variable*(rv) if the value that it assumes, corresponding to the outcome of an experiment, is a chance or random event.

Example:

- X=number of observed "Tail" while tossing a coin 10 times
- X=survival time after specific treatment of a randomly selected patient
- X=GPA of a randomly selected university student

Similar as for variables in sample data, random variables can be categorical or quantitative, and if they are quantitative they can be either discrete or continuous.

$$\text{random variable} \begin{cases} \nearrow \text{categorical} \\ \searrow \text{quantitative} \begin{cases} \nearrow \text{discrete} \\ \searrow \text{continuous} \end{cases} \end{cases}$$

Similar to data description, the models for random variables depend entirely on the type the random variable. The models for continuous random variables will be different than those for categorical or discrete ones.

2.1 Categorical Random Variables

Categorical random variables are described by their *distribution*. Distributions provide the following information about a random variable

- What are the possible values?
- How likely are those values to occur?

Definition
The *probability distribution of a categorical random varibale* is a table giving all possible categories the random variable can assume and their associated probabilities.

Example:

The population investigated are the students of a selected college. The random variable of interest X is the residence status, it can be either resident or nonresident.

If a student is chosen randomly from this college, the probability for being a resident is 0.73. Is X the random variable resident status then write $P(X = \text{resident}) = 0.73$.

The probability distribution is:

resident status	probability
resident	0.73
nonresident	0.27

2.2 Numerical Random Variables

2.2.1 Discrete Random Variables

Remember: A discrete random variable is one whose possible values are isolated points along the number line.

Definition:
The *probability distribution for a discrete random variable* X is a formula or table that gives the possible values of X, and the probability $p(X)$ associated with each value of X.

Value of X	x_1	x_2	x_3	$\cdots$	x_n
Probability	p_1	p_2	p_3	$\cdots$	p_n

The probabilities must satisfy two requirements:

- Every probability p_i is a number between 0 and 1.
- $\sum p_i = 1$.

Example:
Toss two unbiased coins and let x equal the number of heads observed.

The simple events of this experiment are:

coin1	coin 2	X	p(simple event)
H	H	2	1/4
H	T	1	1/4
T	H	1	1/4
T	T	0	1/4

So that we get the following distribution for x=number of heads observed:

X	$p(X)$
0	1/4
1	1/2
2	1/4

With the help of this distribution we can calculate that $P(x \leq 1) = P(x = 0) + P(x = 1) = 1/4 + 1/2 = 3/4$.

Properties of discrete probability distributions:

- $0 \leq p(X) \leq 1$
- $\sum_x p(X) = 1$

The expected value or population mean μ (mu) of a random variable X is the value that you would expect to observe on average if the experiment is repeated over and over again. It is the center of the distribution.

Definition:
Let X be a discrete random variable with probability distribution $p(X)$. The *population mean* μ or *expected value* of X is given as

$$\mu = E(X) = \sum_x X p(X).$$

Example:
The expected value of the distribution of X=the number of heads observed tossing two coins is calculated by

$$\mu = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

Definition:
Let X be a discrete random variable with probability distribution $p(X)$. The *population variance* σ^2 of X is

$$\sigma^2 = E((X - \mu)^2) = \sum_x (X - \mu)^2 p(X).$$

The *population standard deviation* σ (sigma) of a random variable X is equal to the square root of its variance.

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

Example (continued):
The population variance of X=number of heads observed tossing two coins is calculated by

$$\sigma^2 = (0-1)^2 \cdot \frac{1}{4} + (1-1)^2 \cdot \frac{1}{2} + (2-1)^2 \cdot \frac{1}{4} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

and the population standard deviation is:

$$\sigma = \sqrt{\sigma^2} = \frac{1}{\sqrt{2}}.$$

2.2.2 Continuous Random Variables

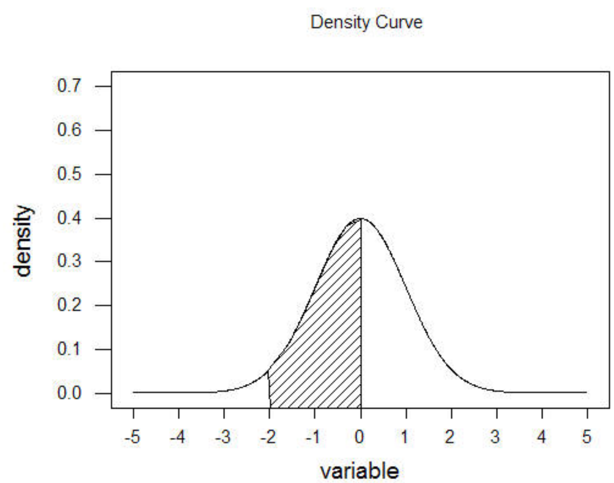
Continuous data variables are described by histograms. For histograms the measurement scale is divided in class intervals and the area of the rectangles put above those intervals is proportional to the relative frequency of the data falling into this interval.

The relative frequency can be interpreted as an estimate of the probability for falling in the associated interval.

With this interpretation the histogram becomes an "estimates" of the probability distribution of the continuous random variable.

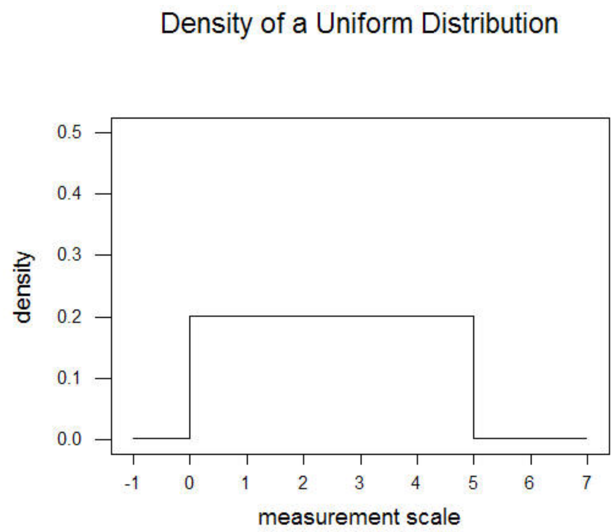
Definition:
The probability distribution of a continuous random variable X is described by a density curve. The probability to fall within a certain interval is then given by the area under the curve above that interval.

1. The total area under a density curve is always equal to 1.
2. The area under the curve and above any particular interval equals the probability of observing a value in the corresponding interval when an individual or object is selected at random from the population.



We can calculate that the probability for falling in the interval $[-2; 0]$ equals 0.4772.

Example:
The density of a uniform distribution in an interval $[0; 5]$ looks like this:



Use the density function to calculate probabilities for a random variable X with a uniform distribution on $[0; 5]$:

- $P(X \leq 3) =$ area under the curve from $-\infty$ to $3 = 3 \cdot 0.2 = 0.6$
- $P(1 \leq X \leq 2) =$ area under the curve from 1 to $2 = 1 \cdot 0.2 = 0.2$
- $P(X > 3.5) =$ area under the curve from 3.5 to $\infty = 1.5 \cdot 0.2 = 0.3$

Remark: Since there is zero area under the curve above a single value, the definition implies for **continuous** random variables and numbers a and b:

- $P(X = a) = 0$
- $P(X \leq a) = P(X < a)$
- $P(X \geq b) = P(x > b)$
- $P(a < X < b) = P(a \leq X \leq b)$

This is generally not true for discrete random variables.

How to choose a model for a given variable of a sample?
The model (density function) should resemble the histogram for the given variable.
Fortunately, many continuous data variables have bell shaped histograms. The normal probability distribution provides a good model for modeling this type of data.

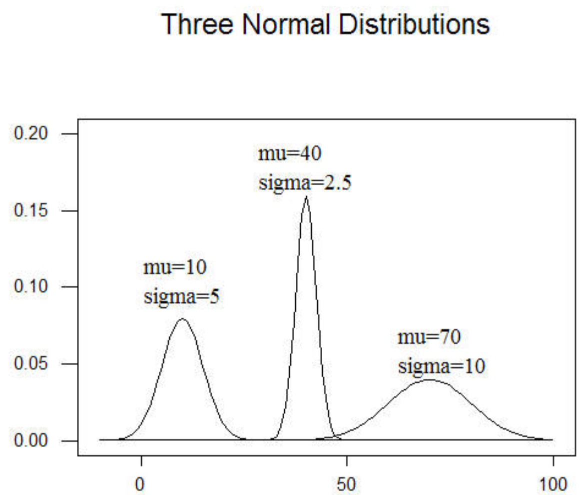
2.2.3 Normal Probability Distribution

The density function of a normal distribution is unimodal, mound shaped, and symmetric.
There are many different normal distributions, they are distinguished from one another by their population mean μ and their population standard deviation σ.

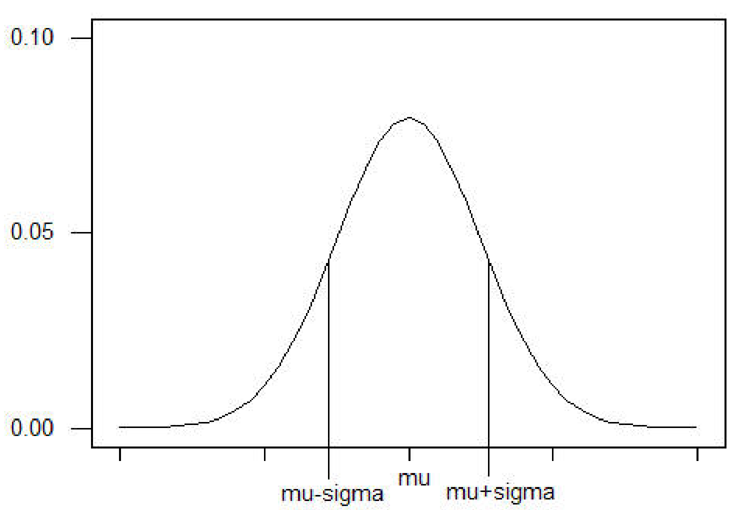
The **density function of a normal distribution** is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty \leq x \leq \infty$$

with $e \approx 2.7183$ and $\pi \approx 3.1416$. μ and σ are called the parameters of a normal distribution. And a normal distribution with mean μ and standard deviation σ is denoted as $N(\mu, \sigma)$.



μ is the center of the distribution, right at the highest point of the density distribution function. At the values $\mu - \sigma$ and $\mu + \sigma$ the density curve has turning points. Coming from $-\infty$ the curve turns from a left to a right curve at $\mu - \sigma$ and again into in a left curve at $\mu + \sigma$.



If the normal distribution is used as a model for a specific situation, the mean and the standard deviation have to be chosen for that situation. E.g. the height of students at a college follow a normal distribution with $\mu = 178$ cm and $\sigma = 10$ cm.
The normal distribution is one example for a quantitative continuous distribution!

Definition:
The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *Standard Normal Distribution*, $N(0, 1)$.

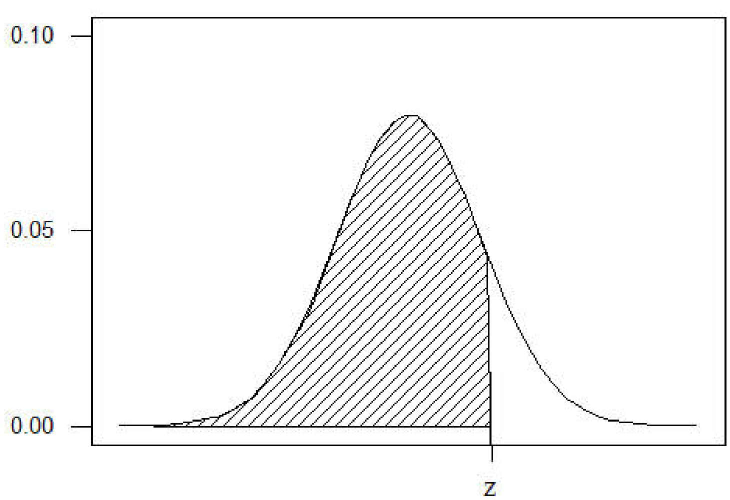
In order to work with the normal distribution, we need to be able to calculate the following:

1. We must be able to use the normal distribution to compute probabilities, which are areas under the normal curve.

2. We must be able to describe extreme values in the distribution, such as the largest 5%, the smallest 1%, the most extreme 10% (which would include the largest 5% and the smallest 5%.

We first look how to compute these for a Standard Normal Distribution.

Since the normal distribution is a **continuous** distribution the following holds for every normal distributed random variable X:
$P(X < z) = P(X \leq z)$ area under the curve from $-\infty$ to z.

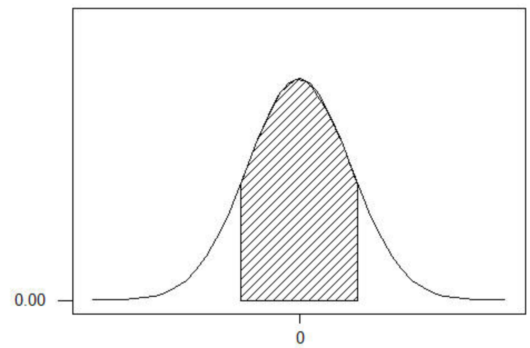


The area under the curve of a normal distribute random variable is very hard to calculate. There is no simple formula that can be used to calculate the area.

Table II in Appendix A (in the text book) tabulates for many different values of z the area under the curve from $-\infty$ to z for standard normal distributed random variables. These values are called cumulative density function.

From now on use Z to indicate a standard normal distributed random variable($\mu = 0$ and $\sigma = 1$). Using the table you find that,

- $P(Z < 1.75) = P(0 < z \leq 1.75) = 0.9599$ therefore
- $P(Z > 1.75) = 1 - P(Z \leq 1.75) = 1 - 0.9599 = 0.0401$.
- $P(-1 < Z < 1) = P(Z < 1) - P(Z \leq -1) = .8413 - 0.1587 = .6826$. (Compare with the Empirical Rule.)



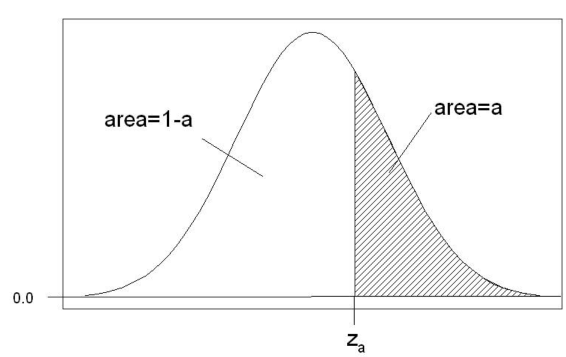
The shaded area equals 0.6826.

The first probability can be interpreted as meaning that, in a long sequence of observations from a Standard Normal distribution, about 95.99% of the observed values will fall below 1.75.

Try this for different values!

Now we will look how to identify extreme values.

Definition:
For any particular number α between 0 and 1, z_α of a distribution is a value such that the cumulative area to the right of z_α is α.



If X is a random variable then z_α is given by:

$$P(X \geq z_\alpha) = \alpha$$

To determine z_α for a standard normal distribution, we can use Table II in Appendix A again.

- Suppose we want to describe the values that make up the largest 2%. So we are looking for $z_{0.02}$, then the area, which falls below $z_{0.02}$ equals 1-0.02=0.98, therefore

  $$P(Z \leq z_{0.02}) = 0.98.$$

  So look in the body of the Table II for the cumulative area 0.98. The closest you will find 0.9798 for $z_{0.02} = 2.05$ This is the best approximation you can find from the table.

  The result is that the largest 2% of the values of a standard normal distribution fall into the interval $[2.05, \infty)$.

- Suppose now we are interested in the smallest 5%. So we are looking for z^*, with

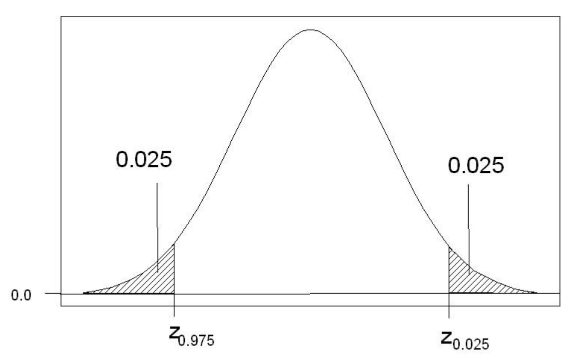
  $$P(z < z^*) = 0.05 \Leftrightarrow P(z > z^*) = 1 - 0.05 = 0.95$$

  Therefore $z^* = z_{0.95}$.

  In the table we have to look for the area to the left, 0.05. Checking the Table we find values 0.0495 and 0.0505, with 0.05 exactly in the middle, so we take the average of the corresponding numbers and get

  $$z_{0.95} = \frac{(-1.64) + (-1.65)}{2} = -1.645$$

- And now we are interested in the most extreme 5%. That means we are interested in the middle 95%. Since the normal distribution is the symmetric the most extreme 5% can be split up in the lower 2.5% and the upper 2.5%. Symmetry about 0 implies that $-z_{0.025} = z_{0.975}$.

  In Table II we find $z_{0.975} = -1.96$, so that $z_{0.025} = 1.96$

  We found the result, that the 5% most extreme values can be found outside the interval $[-1.96, 1.96]$.

Until here we covered how to find probabilities and values for z_α for the standard normal distribution. It remains to show how to find these for any normal distribution, hopefully using the results for the standard normal distribution.

Lemma: Is X normal distributed with population mean μ and population standard deviation σ then the standardized random variable

$$Z = \frac{X-\mu}{\sigma} \quad \text{is normal distributed with } \mu = 0 \text{ and } \sigma = 1$$

or $Z \sim N(0,1)$.

The following example illustrates how the probability and the percentiles can be calculated by using the standardization process from the Lemma.

Example: Let X be normal distributed with $\mu = 100$ and $\sigma = 5$, $x \sim N(100,5)$.

1. Calculate the area under the curve between 98 and 107 for the distribution chosen above.

$$\begin{aligned} P(98 < X < 107) &= P(\tfrac{98-100}{5} < \tfrac{X-100}{5} < \tfrac{107-100}{5}) \\ &= P(-\tfrac{2}{5} < Z < \tfrac{7}{5}) \\ &= P(-0.4 < z < 1.4) \end{aligned}$$

   This can be calculated using Table II. $P(-0.4 < Z < 1.4) = P(Z < 1.4) - P(Z < -0.4) = (0.5 + 0.4192) - (0.5 - 0.1554) = 0.5746$.

2. To find $x_{0.97}$, the value so that the probability to fall above is 0.97, use

$$\begin{aligned} 0.97 &= P(X \geq x_{0.97}) \\ 0.03 &= P(X \leq x_{0.97}) \\ &= P(\tfrac{X-100}{5} \leq \tfrac{x_{0.97}-100}{5}) \\ &= P(Z \leq \tfrac{x_{0.97}-100}{5}) \end{aligned}$$

   But then $\frac{x_{0.97}-100}{5}$ equals the $z_{0.97}$ from a standard normal distribution, which we can find in Table II (look up 0.03).

$$\frac{x_{0.97}-100}{5} = -1.88$$

   This is equivalent to $x_{0.97} = -1.88 \cdot 5 + 100 = 100 - 9.40 = 90.6$.

   So that the lower 3% of a normal distributed random variable with mean $\mu = 100$ and $\sigma = 5$ fall into the interval $(-\infty, 90.6]$, or the top 97% from this distribution fall above 90.6.

Example:
Assume that the length of a human pregnancy follows a normal distribution with mean 266 days and standard deviation 16 days.
What is the probability that a human pregnancy lasts longer than 280 days?

$$\begin{aligned} P(X > 280) &= P(\tfrac{X-266}{16} > \tfrac{280-266}{16}) \quad \text{standardize} \\ &= P(Z > \tfrac{14}{16}) \\ &= 1 - P(Z \le 0.875) \\ &= 1 - 0.8078 \\ &= 0.1922 \end{aligned}$$

How long do the 10% shortest pregnancies last? Find $x_{0.9}$.
(Draw a picture.)

$$\begin{aligned} P(X \le x_{0.9}) &= 0.1 \\ P(\tfrac{X-266}{16} \le \tfrac{x_{0.9}-266}{16}) &= 0.1 \quad \text{standardize} \\ P(Z \le \tfrac{x_{0.9}-266}{16}) &= 0.1 \end{aligned}$$

So $\frac{x_{0.9}-266}{16} = z_{0.9}$.

$$\tfrac{x_{0.9}-266}{16} = z_{0.9} = -1.28 \quad \text{from Table II}$$

This is equivalent to

$$x_{0.9} = 16(-1.28) + 266 = 245.5\text{days}$$

The 10% shortest pregnancies last shorter than 245.5 days.

2.3 Descriptive Methods for Assessing Normality

Process of assessing Normality

1. Construct a histogram or stem-and-leaf plot and note the shape of the graph. If the graph resembles a normal (bell- shaped) curve, the data could be normally distributed. Big deviations from a normal curve will make us decide that the data is not normal.

2. Compute the intervals $\bar{x} \pm s$, $\bar{x} \pm 2 \cdot s$, and $\bar{x} \pm 3 \cdot s$. If the empirical rule holds approximately, the data might be normally distributed. If big deviations occur, the assumption might be unreasonable.

3. Construct a normal probability plot (normal Q-Q Plot) for the data. If the data are approximately normal, the points will fall (approximately) on a straight line.

Construction of normal probability plot:

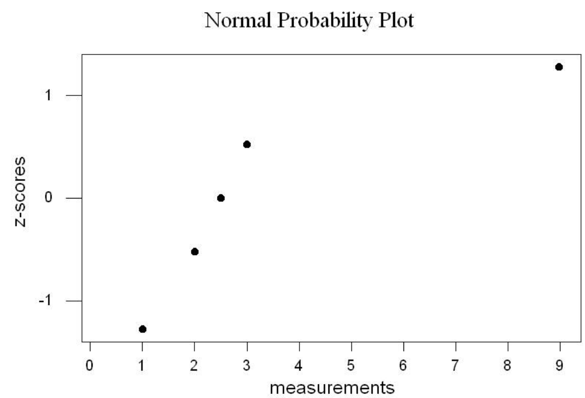
1. Sort the data from smallest to largest and assign the rank, I (smallest gets 1, second 2, etc.)

2. Find the normal scores for the relative rank (Table III).

3. Construct a scatter plot of the measurements and and the expected normal scores.

Example 1
Assume we have the data set 1, 2, 2.5, 3, 9
This is not enough data to do a histogram, so that method does not work. Let's do a normal probability plot to investigate.

x_i	z-score
1	-1.18
2	-0.50
2.5	0
3	0.50
9	1.18



The first 4 points seem to fall on a line, but the fifth does not fit and seems to be an outlier, or uncommon value. The data do not seem to come from a normal distribution. Outliers are highly unlikely in normal distributions.