# Part 1 – Applied Statistics
# Review (and extension) of Probability Theory

## 1.1 Important results of Probability Theory

### Introduction

For this course "Applied Statistics" we will assume that basic probability theory and basic statistics are known to students. Nevertheless we will repeat the highlights of probability theory in section 1. In section 2 we will extend the theory with some specific results and joint continuous distributions.
Statistics can be considered as "applied probability theory": the basics of probability concepts, such as distributions, expectations and variances, play an important role in the introduction, interpretation and construction of the statistical tools. In statistics an observation (measurement) $x$ is a real value that can be interpreted as the realization of the random variable $X$ in our probability model. In this model usually the distribution of the variable is specified, sometimes completely, sometimes parameters of the distribution are unknown.

Remember that a definition of the concept "random sample" of a series of observations can be given as follows:

**Definition 1.1.1** If a $\boldsymbol{X_1, \dots, X_n}$ **is a random sample of** $\boldsymbol{X}$, or: from the distribution of $X$, then:
**1.** $\boldsymbol{X_1, \dots, X_n}$ are **independent** and
**2.** $\boldsymbol{X_1, \dots, X_n}$ all have **the same distribution as** $\boldsymbol{X}$ (the population distribution).

The observed values $x_1, \dots, x_n$ (the realization of the sample) is called a "random sample" as well. But the independence and distributions are based on the "underlying model" with the random variables $X_1, \dots, X_n$. Furthermore, if we state that we have "a random sample from the normal distribution", we indicate that the population is assumed to have a normal distribution and the observations are independent variables $X_1, \dots, X_n$, having this distribution.

In probability theory we distinguished discrete and continuous random variables.

### Discrete random variables

If the probabilities can be given by a probability function $P(X = x)$, the variable is discrete: it attains a finite or a numerable infinite number of values: the range is $S_X = \{x_1, \dots, x_N\}$ or $S_X = \{x_1, x_2, x_3, \dots\}$.

Mostly a discrete $X$ is defined as an integer number, e.g., a number of events that occur.
The expected value can be computed as a "weighted average" $\mu = E(X) = \sum_x x \cdot P(X = x)$
and the variance is $\sigma^2 = var(X) = E(X - \mu)^2$
the standard deviation, having the same unit as $X$, is simply $\sigma = \sqrt{var(X)}$

Remember that the expectation μ is called "mean" in common language, which should be interpreted as the **population mean**: this value is usually unknown, unless the probability distribution of the population is

completely specified. $\overline{x}$ is referred to as the "mean" as well, but should be interpreted as the **sample mean**, resulting from a random sample $x_1, \ldots, x_n$, drawn from the population.

Four basic "families" of discrete distributions are:

- The **binomial distribution** for $n$ independent and identical "trials" with two possible outcomes ("success" or "failure") and success probability $p$ for each trial.
  $X$, the number of successes in $n$ trials, has a $B(n,p)$-distribution (for short: $X \sim B(n,p)$):
  $$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, .., n.$$
  Expectation and variance are $E(X) = np$ and $var(X) = np(1-p)$.
  In the formulas above you may replace $1-p$ by $\boldsymbol{q}$. In this reader we will use $1-p$.
  For example, if $X$ is the number of 6`s in 25 rolls with a dice, then $X \sim B\left(25, \frac{1}{6}\right)$.

- The **Poisson distribution** provides often an adequate model if we count the number of (rare) events in a period/area, e.g., the number of heart attacks on a day in a large town.
  The parameter μ is the mean number of events and increases proportionally as the period or the area in which is counted, is enlarged.
  $P(X=x) = \frac{\mu^x}{x!} e^{-\mu}$, where $x = 0, 1, 2, \ldots$ and $\mu = E(X) = var(X)$.

  The relation between the two discrete distributions above is given by the following approximation: if the number of trials is large enough ($n \geq 25$) and the success probability $p$ is small enough (such that $np < 10$), then the **$\boldsymbol{B(n,p)}$-distribution can be approximated by the Poisson($\boldsymbol{\mu = np}$)-distribution**.

- The **hypergeometric distribution** is applied whenever we count the number of successes if we draw a random sample (without replacement) from a finite dichotomous population (consisting of "Successes" and "Failures"). Using the concept of the "vase model" with $R$ red and $N-R$ white balls and count the number $X$ of red balls in $n$ random draws without replacement, then $X$ has a hypergeometric distribution:
  $P(X=x) = \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}, x = 0,1, \ldots, n$
  $E(X) = np$ and $var(X) = np(1-p)\frac{N-n}{N-1}$, where $p = \frac{R}{N}$ and $\frac{N-n}{N-1}$ is the correction factor for a finite population.
  These last formulas and the factor suggest the link between the hypergeometric and binomial distribution. Indeed, for large populations, rule of thumb $N > 5n^2$, the hypergeometric distribution can be approximated by the $B\left(n, \frac{R}{N}\right)$-distribution.

- The **geometric distribution** is used when we count the number $X$ of Bernoulli trials until we obtain a (first) success: $P(X=x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$
  Useful properties: $E(X) = \frac{1}{p}$, $var(X) = \frac{1-p}{p^2}$, $P(X > x) = (1-p)^x$ and the geometric distribution is the only discrete distribution that has the **"lack of memory" property**:
  $$P(X > x + y | X > x) = P(X > y), \quad \text{for all } x, y = 0,1,2, \ldots..$$

The concept of independence of variables is important, especially in statistics, where we use random samples. The "randomness" of the sample variables (observations) implies **independence**:

- In general two random variables are independent if for any pair of sets $A_1 \subset \mathbb{R}$ and $A_2 \subset \mathbb{R}$:

  $$P(X_1 \in A_1 \text{ and } X_2 \in A_2) \overset{ind.}{=} P(X_1 \in A_1) \cdot P(X_2 \in A_2).$$

  For example we can state that: $P(X_1 \leq x_1 \text{ and } X_2 > x_2) \overset{ind.}{=} P(X_1 \leq x_1) \cdot P(X_2 > x_2)$
- For two discrete variables we have: $P(X_1 = x_1 \text{ and } X_2 = x_2) \overset{ind.}{=} P(X_1 = x_1) \cdot P(X_2 = x_2)$
- For two continuous variables we will discuss the following properties in the next section:

  $$f_{X,Y}(x,y) \overset{ind.}{=} f_X(x) \cdot f_Y(y) \quad \text{and } F_{X,Y}(x,y) \overset{ind.}{=} F_X(x) \cdot F_Y(y).$$
- Variances: $var(X_1 + \cdots + X_n) \overset{ind.}{=} var(X_1) + \cdots + var(X_n)$

For the expectations we have in general (the equality is valid for dependent variables as well!):

- $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

If we have a random sample, taken from a population with expectation μ and variance $\sigma^2$, it follows from the properties above that the summation $X_1 + \cdots + X_n$ has expectation $n\mu$ and variance $n\sigma^2$.

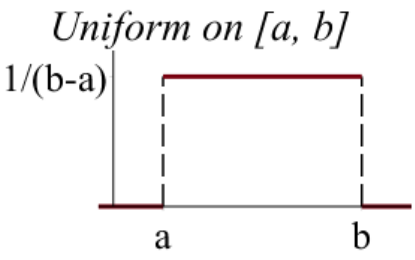
Furthermore, if the transition to other units of measurement or a linear transformation of the variables is considered, then the following properties apply:

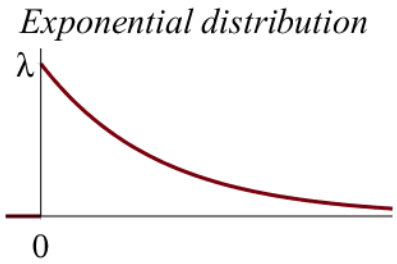
- $E(aX + b) = aE(X) + b$
- $var(aX + b) = a^2 var(X)$
  and the standard deviation of $aX + b$ is $\sigma_{aX+b} = |a|\sigma$ ($|a|$ is the absolute value of $a$)

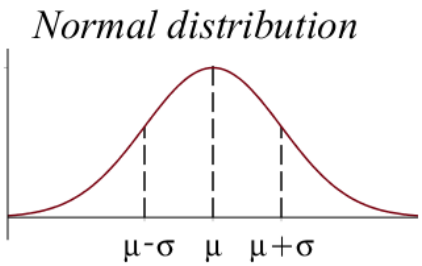
Note that the properties for expectation and variance of the linear transformation $Y = aX + b$ and of the sum $X_1 + \cdots + X_n$ hold for both discrete and continuous variables.

## Continuous Distributions

Giving an appropriate model for the sample observations is not a simple task: however, the shape of a histogram or a bar graph could give you an indication whether one of the common distributions applies.
The shapes of the most common continuous distributions are given in the following graphs:




Usually we will give the density function $f(x)$ to define a continuous distribution, but sometimes it is given by its (cumulative) distribution function $F(x) = P(X \leq x)$.
Since $F(x) = \int_{-\infty}^{x} f(u)du$, we have $f(x) = \frac{d}{dx}F(x)$.
The main characteristics of the common continuous distributions above are as follows:

- **The uniform distribution $\boldsymbol{U(a,b)}$** applies to numbers drawn at random from an interval. Especially random numbers from the interval $(0,1)$ are often used in simulations. The general case is a uniform distribution on $(a,b)$, an interval with length $b-a$. The density function is simply $f(x) = \frac{1}{b-a}$ on the interval en $f(x) = 0$ outside the interval. The expected value of such a number is the middle of the interval: $E(X) = \frac{a+b}{2}$. And $var(X) = \frac{(b-a)^2}{12}$.

- **The exponential distribution $\boldsymbol{Exp(\lambda)}$** is often an appropriate model when waiting times, inter-arrival times of clients and life times are observed. The density function is clearly "skewed to the right" and is non-zero for positive values of the times: $f(x) = \lambda e^{-\lambda x}$.
  The **parameter** λ can be determined if the "mean" $E(X) = \frac{1}{\lambda}$ is known. Furthermore $\sigma = \frac{1}{\lambda}$ equals the expectation and the "survival" probability is given by $P(X > x) = e^{-\lambda x}$ $(x \geq 0)$.
  This is the only continuous distribution that has the "lack of memory" property.

  The exponential distribution is a special case of the **Gamma distribution with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$**:

$$f(x) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}},$$

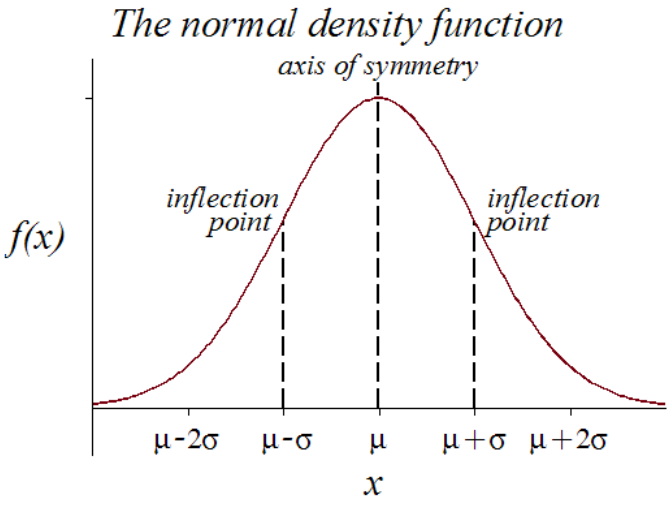
  where $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}e^{-u}du$, having properties $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and $\Gamma(n) = n!$.
  Verify that for $\alpha = 1$ and $\beta = \frac{1}{\lambda}$ we have an exponential distribution with parameter $\lambda$.

- **The normal distribution $\boldsymbol{N(\mu,\sigma^2)}$ and random samples drawn from this distribution.**

$$\boldsymbol{f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}, \text{with expectation } \boldsymbol{\mu} \text{ and variance } \boldsymbol{\sigma^2}$$

  The importance of the normal distribution and its central role in probability theory have been emphasized before. In this course we will see that many statistical techniques are based on the assumption of a normal model of variables in applications: in physics, nature, economy, etc.

"$X$ is $N(\mu, \sigma^2)$" means that the population shows a **bell (mound) shaped distribution**, symmetric about the line $x = \mu$ and having a standard deviation σ. The probabilities of the "Empirical rule" apply and can be determined with the table of standard normal probabilities.




The "**Empirical rule**":

| **Interval** | **Probability** of "value in interval" |
|---|---|
| $(\mu - \sigma, \mu + \sigma)$ | **≈ 68%** |
| $(\mu - 2\sigma, \mu + 2\sigma)$ | **≈ 95%** |
| $(\mu - 3\sigma, \mu + 3\sigma)$ | **≈ 99.7%** |

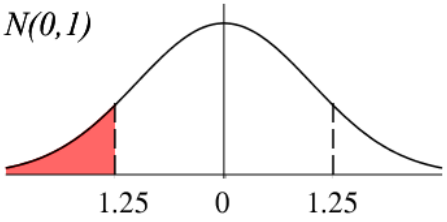
Probabilities can be computed using standardization: if $X \sim N(\mu, \sigma^2)$, then $\boldsymbol{Z = \frac{X-\mu}{\sigma} \sim N(0,1)}$.

Then we can use the $N(0,1)$-table, containing values of the standard normal (cumulative) distribution function $\Phi(z) = P(Z \leq z)$, for positive values $z \geq 0$.
Note that in $N(\mu, \sigma^2)$ the second parameter is the **variance $\boldsymbol{\sigma^2}$**, **not** the standard deviation $\sigma$.

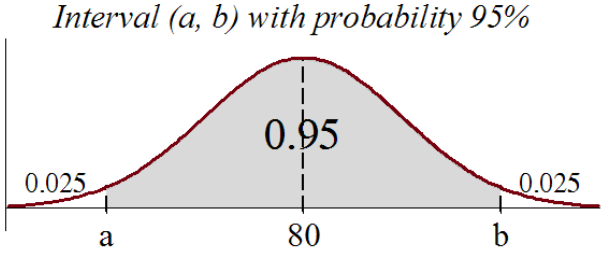
**Example 1.1.2** Consider a population of persons with weights in kg, that are $N(80, 64)$-distributed, so our model is: $X =$ "the weight of an arbitrarily chosen person", $X$ is $N(80, 64)$.

a. Compute the probability $P(X \leq 70)$.
Solution:
$P(X \leq 70) = P\left(Z \leq \frac{70-80}{8}\right) = P(Z \leq -1.25)$
Using the symmetry of the standard normal distribution we know that the probability of $Z \leq -1.25$ equals the probability of $Z \geq 1.25$ so that:

$$P(X \leq 70) = P(Z \geq 1.25) = 1 - P(Z \leq 1.25)$$
$$= 10.56\%$$




b. What is the 95th percentile $c$ of these weights?
Solution: we will determine $c$ such that:
$P(X \leq c) = 0.95$
As in a. we will compute the z-score, in this case for $c$: $P(X \leq c) = P\left(Z \leq \frac{c-80}{8}\right) = 0.95$,
Using the table: $\frac{c-80}{8} = 1.645$, we find the 95th percentile $c = 80 + 1.645 \cdot 8 \approx 93.2$ kg.



c. Determine an interval $(a, b)$, symmetric about μ = 80 $kg$, such that $P(a < X < b) = 0.95$
Solution: $P(X < b) = 0.975$, or: $P\left(Z < \frac{b-80}{8}\right) = 0.975$
From the $N(0,1)$-table we find the z-score $z = \frac{b-80}{8} = 1.96,$ so $b \approx 95.7\ kg$.
Because of the symmetry about 80, $a = 64.3$. So $P(64.3 < X < 95.7) = 0.95$ ■

In probability theory we discussed that both the sum and the mean of independent, normally distributed variables $X_1, \ldots, X_n$ are normally distributed as well.

**Property 1.1.3** For a random sample $X_1, \ldots, X_n$, taken from a $N(\mu, \sigma^2)$-distribution we have:

$$\sum_{i=1}^{n} \boldsymbol{X_i} \sim \boldsymbol{N(n\mu, n\sigma^2)} \quad \text{and} \quad \boldsymbol{\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)}$$

Note that the expected values of the sum and the mean differ a factor $n$, but the variances a factor $n^2$. This a consequence of the property $var(aX + b) = a^2 var(X)$:

$$var\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i\right) = \left(\tfrac{1}{n}\right)^2 var(\textstyle\sum_{i=1}^{n} X_i) = \tfrac{1}{n^2}[var(X_1) + \cdots + var(X_n)] = \tfrac{1}{n^2} \cdot n\sigma^2 = \tfrac{\sigma^2}{n},$$

resulting in a standard deviation of the sample mean, being $\boldsymbol{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$.

The **Central Limit Theorem** (CLT) makes the statements in property 1.1.3 *approximately* applicable for large samples, drawn from not normally distributed populations. As a rule of thumb we consider $\boldsymbol{n \geq 25}$ "large enough".

**Property 1.1.4 The Central Limit Theorem (CLT)**
If $X_1, \ldots, X_n$ are independent and all identically distributed with expectation $\mu$ and variance $\sigma^2$,

$$\boldsymbol{\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = \Phi(z)}$$

According the CLT, $\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ **converges in distribution** to the $N(0,1)$-distributed $Z$.

For the proof see advanced text books on probability theory.

**Example 1.1.5**
Suppose we observe 1000 random numbers between 0 and 1: than the result of this sample can be modelled as independent variables $X_1, X_2, \ldots, X_{1000}$, which all have a $U(0,1)$-distribution with $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$. Then, approximately according to the CLT:

$$\sum_{i=1}^{1000} X_i \sim N\left(1000 \cdot \frac{1}{2}, 1000 \cdot \frac{1}{12}\right) = N\left(500, \frac{1000}{12}\right) \text{ and } \overline{X} = \frac{1}{1000}\sum_{i=1}^{1000} X_i \sim N\left(0.5, \frac{1}{12000}\right)$$ ■

## The binomial distribution and the normal approximation of the binomial probabilities.

When considering properties of a population we might be interested in non-numerical aspects, such as: being or not being married of an adult, whether or not a product is substandard, etc.
In these cases we would like to know which part or proportion in the population has the property. The **variable** with two values "possesses the property" and "does not possess the property" is the simplest **categorical** variable. A categorical variable with two possible values is called **dichotomous**: each element of the population has, or does not have, the property. Remember that in probability theory we indicated these outcomes as "success" and "failure". We are interested in the unknown population proportion $p$ of successes (having the property). Based on a random sample of $n$ elements taken from the population we might try to determine the value of $p$: under conditions (independence implies e.g. sampling **with replacement**) we can assume that the number $X$ of successes is a $B(n,p)$-distributed variable. Based on the actually observed value $x$ of $X$ one could give an estimate of the **population proportion $\boldsymbol{p}$** by computing the **sample proportion** $\hat{p} = \frac{\boldsymbol{X}}{\boldsymbol{n}}$.

A more refined model of this binomial situation can be given by defining a variable $X_i$ for each element: $X_i = 1$ if the element has the property and $X_i = 0$, if not. So $X = \sum_{i=1}^{n} X_i$, since the sum of all 1`s and 0`s equals the observed number of successes. Reasoning from the actually observed values $x_i$, then, instead of sample proportion $\frac{x}{n}$, we can write $\frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$ : the sample proportion is a mean of a series of 1-0 variables ("alternatives")!
The model with the independent alternatives $X_i$`s reminds us that the CLT applies for large $n$: then $X = \sum_{i=1}^{n} X_i$ is approximately normal with parameters met $\mu = E(X) = np$ and

$$\sigma^2 = var(X) = np(1-p).$$

The rule of thumb for applying this approximation: $\boldsymbol{n \geq 25,\ np > 5}$ **and** $\boldsymbol{n(1-p) > 5}$.

**Property 1.1.6** If $X \sim B(n,p)$, then we have approximately (CLT) for sufficiently large $n$

$$\boldsymbol{X \sim N(np, np(1-p))} \quad \text{and} \quad \boldsymbol{\hat{p} = \frac{X}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)}$$

Similarly as in property 1.1.3, the expectations differ a factor $n$, and the variances a factor $n^2$.

**Continuity correction** is mandatory when applying property 0.2.3 with respect to $X$, but, when computing probabilities w.r.t. $\frac{X}{n}$, we will **not** apply continuity correction, as shown below:

**Example 1.1.7**
In a referendum on the separation of Scotland less than 50% of the voters were in favour of separation. Prior to the referendum many opinion polls showed a variety of possible outcomes: some predicted that at most 47% would be in favour of separation, others predicted a majority of 51% or more.
Let us assume that (exactly) 50% was in favour of separation and a researcher wants to predict the result of the referendum, based on a random sample of $n = 1600$ Scots. What is, in that case, the probability that the sample proportion deviates at least 2% from the real proportion (50%)?

**Model**: $X =$ "the number in favour of separation in the sample of 1600 Scots",
then $X \sim B(1600, p)$, where $p$ is assumed to be 0.5.
Consequently the expected number $E(X) = np = 800$ and $var(X) = np(1-p) = 400$.
$X$ has, according to the CLT, a $N(800, 400)$-distribution.
A deviation of 2% is $0.02 \cdot 1600 = 32$ Scots. Using symmetry we find the requested probability, applying **continuity correction** (c.c.), because of the switch from a discrete (binomial) distribution to a continuous (normal) variable:

$$2 \cdot P(X \geq 832) \overset{\text{c.c.}}{=} 2 \cdot P(X \geq 831.5) \overset{\text{CLT}}{\approx} 2 \cdot P\left(Z \geq \tfrac{831.5-800}{\sqrt{400}}\right)$$
$$\approx 2 \cdot [1 - P(Z \leq 1.58)] \approx 12.6\%$$

An alternative computation uses the approximately normal distribution $N\left(0.5, \sqrt{\tfrac{0.5 \cdot 0.5}{1600}}\right)$ of the sample proportion $\frac{X}{n}$ (without continuity correction).
We will use that $X \geq 832$ is equivalent to $\frac{X}{1600} \geq \frac{832}{1600} = 0.52$, so

$$2 \cdot P(X \geq 832) = 2 \cdot P\left(\frac{X}{1600} \geq 0.52\right) \overset{\text{CLT}}{\approx} 2 \cdot P\left(Z \geq \frac{0.52-0.50}{\sqrt{\frac{0.5 \cdot 0.5}{1600}}}\right) = 2\big(1 - \Phi(1.60)\big) \approx 11.0\%$$

The resulting probability 11.0% is less than the 12.6% probability before: the difference is caused by the absence of continuity correction in the last computation. ■

In general we will always apply continuity correction if we are normally approximating probabilities with respect to integer valued variables.

If we have to determine the distribution of functions of variables with a known distribution, such as $Y = X^2$ or $M = \max(X_1, \ldots, X_n)$, we will start to express the distribution function (cdf) of the function in the known distribution(s), as illustrated in the following example.

**Example 1.1.8**
Determine the distribution of $Z_1^2 + Z_2^2$, if $Z_1$ and $Z_2$ are independent and both $N(0,1)$-distributed.
First, we will determine the distribution of $X = Z_1^2$ (and of $Y = Z_2^2$):
If $x > 0$ we have

$$F_X(x) = P(Z_1^2 \leq x) = P\big(-\sqrt{x} \leq Z_1 \leq \sqrt{x}\big) = \Phi\big(\sqrt{x}\big) - \Phi\big(-\sqrt{x}\big) = 2\Phi\big(\sqrt{x}\big) - 1$$

So $f_X(x) = \frac{d}{dx} F_X(x) = 2 \cdot \frac{1}{2\sqrt{x}} \varphi\big(\sqrt{x}\big) = \frac{1}{\sqrt{2\pi x}} e^{-x/2}$, for $x > 0$ (and $f_X(x) = 0$ elsewhere.)
Now we can apply the convolution integral to find the distribution of $X + Y = Z_1^2 + Z_2^2$:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)dx = \int_0^z \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \frac{1}{\sqrt{2\pi(z-x)}} e^{-\frac{z-x}{2}} dx$$
$$= \frac{e^{-z/2}}{2\pi} \int_0^z \frac{1}{\sqrt{x(z-x)}} dx$$

In the calculus book we can find the result of the latter integral: $\int_0^z \frac{1}{\sqrt{x(z-x)}}dx = \pi$.

So $f_{X+Y}(z) = \frac{1}{2}e^{-\frac{1}{2}z}$, for $z > 0$. We found that $X + Y \sim \text{Exp}\left(\lambda = \frac{1}{2}\right)$.
Similarly, by repeatedly applying the convolution integral, we can find the distribution of $Z_1^2 + Z_2^2 + \cdots + Z_n^2$, which for independent and standard normally distributed $Z_1, \ldots, Z_n$ has, by definition, a **Chi-square distribution with $n$ degrees of freedom** ($df = n$).
Above we derived the density of this distribution for $Z_1^2$ $(df = 1)$ and for $Z_1^2 + Z_2^2$ $(df = 2)$. ■

In example 1.1.8 we applied the formula of the convolution integral for 2 independent continuous variables. The **convolution sum** gives a similar expression for 2 independent discrete variables:

$$P(X + Y = n) = \sum_k P(X = k)P(Y = n - k)$$

Applying this property to two independent, Poisson distributed variables $X$ and $Y$, with parameters $\mu_1$ and $\mu_2$, respectively, we can derive that the sum $X + Y$ has a Poisson distributed as well, with parameter $\mu_1 + \mu_2$.

**Definition 1.1.9 The Chi-square distribution**
If $Z_1, \ldots, Z_n$ are independent and all $N(0, 1)$-distributed, then $Y = Z_1^2 + \cdots + Z_n^2$ has a **Chi-square distribution with $n$ degrees of freedom.**

Brief notation: $\boldsymbol{Y \sim \chi_n^2}$

**Property 1.1.10** If $Y$ has a Chi-square distribution with $n$ degrees of freedom , then the expectation $\boldsymbol{E(Y) = n}$ and variance $\boldsymbol{var(Y) = 2n}$.

The Chi-square density function is given by

$$f(x) = \frac{\frac{1}{2}e^{-\frac{1}{2}x}\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)} \ (x > 0), \ \text{ where } \Gamma(t) = \int_0^\infty e^{-x}x^{t-1}dx \ (t > 0).$$

This is a special case of the Gamma-distribution (see page 1-5): parameters $\alpha = \frac{n}{2}$ and $\beta = 2$.
For $n = 2$, $Y$ has the $Exp\left(\lambda = \frac{1}{2}\right)$-distribution (see also example 1.1.5).

# 1.2 Theoretical extensions of Probability Theory

**Definition 1.2.1** The **moment generating function** of a variable $X$ is $\boldsymbol{M_X(t) = E(e^{tX})}$ .

Moment generating functions characterize distributions and can be used to determine the moments of a distributions and can simplify the proof of properties of distributions., as the following property and example illustrate.

**Property 1.2.2**
**a.** The $k$-the derivative of $M(t)$ equals, for $t = 0$, the $k^{\text{th}}$ moment $E(X^k)$: $\boldsymbol{M^{(k)}(0) = E(X^k)}$.
**b.** Two variables $X$ and $Y$ are **independent** $\Leftrightarrow \boldsymbol{M_{X+Y}(t) = M_X(t) \cdot M_X(t)}$

**Example 1.2.3** If $X$ and $Y$ are independent and both $Exp(\lambda)$-distributed, then:

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x} dx = \left[\frac{\lambda}{t-\lambda}\right]_{x=0}^{x\to\infty} = \frac{\lambda}{\lambda - t}$$

So $M_X^{`}(t) = -1 \cdot -1 \cdot \frac{\lambda}{(\lambda-t)^2} = \frac{\lambda}{(\lambda-t)^2}$.

Similarly: $M_X^{``}(t) = \frac{2\lambda}{(\lambda-t)^3}, \ldots\ldots, M_X^{(k)}(t) = \frac{\lambda \cdot k!}{(\lambda-t)^{k+1}}$, so $E(X^k) = M_X^{(k)}(0) = \frac{k!}{\lambda^k}$.

And: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = \left(\frac{\lambda}{\lambda-t}\right)^2$ ■

You can consult an advanced Probability Theory textbook to verify that we can use generating functions to show that the following important results b. and c. hold (a. repeats property 1.1.4):

**Property 1.2.4** If $X_1, \ldots, X_n$ are independent and all $N(\mu, \sigma^2)$-distributed, then for the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ we have:

- **a.** $\overline{X}$ has a $N\left(\mu, \frac{\sigma^2}{n}\right)$-distribution.
- **b.** $\frac{(n-1)S^2}{\sigma^2}$ has a Chi-square distribution with $n-1$ degrees of freedom.
- **c.** $\overline{X}$ and $S^2$ are independent.

Property 1.2.4 is, of course, a key result for application in statistics: the distributions of the sample mean and the sample variance and their independence!

**Property 1.2.5 Markov's inequality**

If the first and second moment of a random variable $Y$ exist, then $\boldsymbol{P(|Y| \geq c) \leq \frac{E(Y^2)}{c^2}}$, for any $c > 0$.

**Proof**: We will give the proof for discrete variables (the continuous case is similar).
Note that the events $|Y| \geq c$ and $Y^2 \geq c^2$ are equivalent.

$$EY^2 = \sum_x y^2 P(Y=y) = \sum_{y^2 \geq c^2} y^2 P(Y=y) + \sum_{y^2 < c^2} y^2 P(Y=y)$$

$$\geq \sum_{y^2 \geq c^2} y^2 P(Y=y) \geq \sum_{y^2 \geq c^2} c^2 \cdot P(Y=y) = c^2 \cdot P(|Y| \geq c)$$ ■

**Property 1.2.6 Chebyshev`s inequality**

If the expectation $\mu$ and variance $\sigma^2$ of a random variable $X$ exist, then

$$\boldsymbol{P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}}, \text{for any } c > 0$$

This follows directly from Markov's inequality: substitute $Y = X - \mu$, so that $E(Y^2) = \sigma^2$.

If we choose $c = k\sigma$, intervals $(\mu - k\sigma, \mu + k\sigma)$ are symmetrical about the mean $\mu$: Chebyshev`s inequality claims that there is a probability of at most $\frac{1}{k^2}$ to observe a value outside the interval.
For $k = 2$ and $k = 3$ these maximum probabilities are $\frac{1}{4} = 25\%$ and $\frac{1}{9} \approx 89\%$ , respectively.
Chebyshev's inequality is often used in probability to proof general properties of all kind of distributions, such as the following:

**Property 1.2.7 (Weak law of large numbers)**
If $X_1, \ldots, X_n$ are independent and all identically distributed with expectation $\mu$ and variance $\sigma^2$, then for the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ we have: $\lim_{n\to\infty} P\left(\left|\overline{X} - \mu\right| \geq c\right) = 0$ for any $c > 0$.

This property follows directly from Chebyshev's inequality and the property $var\left(\overline{X}\right) = \frac{\sigma^2}{n}$.
According to this weak law the sample mean $\overline{X}$ is said to "converge to $\mu$ in probability": $\overline{X} \xrightarrow{P} \mu$.
Since the sample proportion $\hat{p} = \frac{X}{n}$ is a sample mean as well (if we consider $X$ to be a sum of 1-0 variables $X_i$ for the $n$ Bernoulli trials), the weak law also applies here: $\hat{p} \xrightarrow{P} p$.
The strong law of large numbers states that $\overline{X}$ converges to $\mu$ with probability 1.

# 1.3 Exercises

**1.** The constant in the Gamma distribution is: $\Gamma(\alpha) = \int_{u=0}^{\infty} u^{\alpha-1} e^{-u} du \quad (\alpha > 0)$. Show that:

- **a.** $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$
- **b.** $\Gamma(1) = 1$ and $\Gamma(n) = (n-1)!$
- **c.** $\Gamma(½) = \sqrt{\pi}$ (use substitution $u = ½x^2$) and $\Gamma(n + ½) = (n - ½)\,(n - 1½) \ldots (½)\sqrt{\pi}$

**2.** $X$ has a known density function $f_X$ and $Y = -3X + 5$.
- **a.** Express the (probability) density function of $Y$ in $f_X$.
- **b.** Determine the density function of $Y$ if $X$ has a $N(0,1)$-distribution and verify that this is the density of a normal distribution (which parameters?).

**3.** Derive the density function of $Z = min(X, Y)$, if $X$ and $Y$ are independent and exponentially distributed variables with the same parameter λ.

**4.** $X_1, \ldots, X_n$ are independent and all $U(0,1)$-distributed.

- **a.** Determine the distribution function of $X_1$.
- **b.** Find the density function of $Z = max(X_1, \ldots, X_n)$ is: $f_Z(z) = nz^{n-1}$, for $0 < z < 1$.
- **c.** Determine $E(Z)$ and $var(Z)$.

**5.** $X_1, \ldots, X_n$ are independent and all $Exp(\lambda)$-distributed: $S = \sum_{i=1}^{n} X_i$ and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

- **a.** Prove the formulas $E(X_1) = \frac{1}{\lambda}$ and $var(X_1) = \frac{1}{\lambda^2}$.

**b.** Approximate the de probability $P(S > 55)$ if $n$ = 100 and λ = 2.
**c.** Approximate the de probability $P(\overline{X} > 0.55)$ if $n$ = 50 and λ = 2.

**6.** $Z_1, Z_2, \ldots, Z_n$ are independent and all standard normal.

**a.** Compute $E(Z_1^2)$ and $var(Z_1^2)$, and $E(Z_1^2 + \cdots + Z_n^2)$ and $var(Z_1^2 + \cdots + Z_n^2)$
**b.** Find the density function of $Z_1^2$, which is the $\chi_1^2$-density function.
**c.** Use the convolution integral $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$ and $\int_0^z \frac{1}{\sqrt{x(z-x)}}dx = \pi$ (see your calculus book) to show that $Z_1^2 + Z_2^2$ has an exponential distribution.
**d.** Check that the $\chi_2^2$-distribution you found in c. is a gamma distribution with parameters $\alpha = 1$ and $\beta$ = 2

**7.** Determine the moment generating function of $X$ and check the formula of $E(X)$ and $var(X)$:
**a.** if $X \sim N(0,1)$
**b.** if $X \sim Exp(\lambda)$
**c.** if $X \sim Poisson\ (\mu)$

**8.** The random behavior of complicated waiting time systems is sometimes statistically assessed by computer simulations. For that goal waiting times (service times) are generated, using random number generators, that produce random numbers between 0 and 1.
Assume $X$ is such a random number (uniformly distributed on $(0, 1)$ ), then we can generate a random waiting time, having an exponential distribution with parameter $\lambda = 1$, by computing $Y = \ln\left(\frac{1}{X}\right)$.

**a.** Show that $Y$ has an exponential distribution with parameter $\lambda = 1$.
**b.** Verify whether $E(Y) = \ln\left(\frac{1}{E(X)}\right)$,

**9.** The times between two consecutive clients, logging on to a company`s computer system are considered to be independent and exponentially distributed. The mean time between two consecutive log on`s is 12 seconds (during office hours). We consider the log on`s during one minute: $X_1$ is the time (in seconds) from the start to the first log on, $X_2$ is the time between the first and second log on, etc.

**a.** Compute $E(X_i)$, $var(X_i)$ and $E(\sum_{i=1}^{6} X_i)$
**b.** Determine $P(X_1 > 12)$ and $P(X_1 > 15 | X_1 > 3)$.

**10.** $X_1, X_2, \ldots, X_n$ are independent waiting times: they all are exponentially distributed with parameter $\lambda = \frac{1}{4}$.
The sum of waiting times is $S_n = \sum_{i=1}^{n} X_i$ and the mean waiting time is $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

**a.** Give $E(S_n)$ and $var(S_n)$.
**b.** Derive the density function of $S_2$ from the density functions of $X_1$ and $X_2$.
(Apply the convolution integral).
**c.** Compute (for $n = 2$): $P\left(\overline{X}_2 > 5\right)$.
**d.** Approximate (for $n = 100$): $P\left(\overline{X}_{100} > 5\right)$.

**Answers exercises part 1**

**1.** **a**. $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1}e^{-u}du = -u^{\alpha-1}e^{-u}|_{u=0}^{u\to\infty} + \int_0^{\infty}(\alpha-1)u^{\alpha-2}e^{-u}du = (\alpha-1)\Gamma(\alpha-1)$

**2.** **a.** $f_Y(y) = \frac{1}{3} f_X\left(\frac{y-5}{-3}\right)$

**b.** $f_Y(y) = \frac{1}{\sqrt{2\pi \times 9}} e^{-\frac{1}{2} \times \frac{(y-5)^2}{9}}$

**3.** $f_Z(z) = 2\lambda e^{-2\lambda z}, for\ z \geq 0$ (elsewhere $f_Z(z) = 0$)

**4.** **a**. $E(X_i) = \frac{1}{2}$ (symmetry), $E(X_i^2) = \int_0^1 x^2 dx = \left[\frac{1}{3}x^3\right]_{x=0}^{x=1} = \frac{1}{3}$, so $var(X) = E(X^2) - (EX)^2 = \frac{1}{12}$.

**b**. $F_{X_1}(x) = x$ (if $0 \leq x \leq 1$). $F_{X_1}(x) = 0$ for $x < 0$ and $F_{X_1}(x) = 1$ for $x > 1$.

**c.** $f_Y(y) == \frac{1}{2} e^{-\frac{1}{2}y}$ $(y \geq 0)$.

**d.** $f_Z(z) = nz^{n-1}$, for $0 \leq z \leq 1$ (and $f_Z(z) = 0$ elsewhere)

$E(Z) = \frac{n}{n+1}$, $E(Z^2) = \frac{n}{n+2}$, so $var(Z) = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+2)(n+1)^2}$.

**5.** (We are using the notation $X$ instead of $X_1$ in a.)

**a.** $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \left[x \cdot -e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx \ = \ 0 \ + \frac{1}{\lambda}$

**b.** $P(S > 55) = 15.87\%$.

**c.** $P(\overline{X} > 0.55) = 0.2389$.

**d.** $f_M(m) = n\lambda e^{-n\lambda m}$ $(\geq 0)$, $E(M) = \frac{1}{10 \cdot 2} = 0.05$.

**6.** **a**. $E(Z_1^2) = 1,\ E(Z_1^4) = 3,\ var(Z_1^2) = 2, E(Z_1^2 + \cdots + Z_n^2) = n$ and $var(Z_1^2 + \cdots + Z_n^2) = 2n$

**b.** $(Y = Z_1^2)$ $f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y},$ for $y > 0$.

**c.** $f_{X_1+X_2}(z) = \frac{1}{2} e^{-\frac{1}{2}z}$, *for* $z > 0$.

***d.*** $f(z) = \frac{z^{\alpha-1} e^{-\frac{z}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}} = \frac{1}{2} e^{-\frac{1}{2}z}$.

**7.** **a**. $M(t) = e^{\frac{1}{2}t^2}$, $M`(t) = M`(0) = 0 = E(X)$. $M``(0) = 1 = E(X^2) = var(X)$

**b.** $M(t) = \frac{\lambda}{\lambda - t}$ if $t < \lambda$, $M`(0) = \frac{1}{\lambda} = E(X)$, $M``(0) = \frac{2}{\lambda^2}$ and $var(X) = E(X^2) - (EX)^2 = \frac{1}{\lambda^2}$

**c.** $M(t) = e^{\mu(e^t - 1)},\ M`(0) = \mu,\ M``(0) = \mu^2 - \mu = E(X^2),\ var(X) = \ E(X^2) - (EX)^2 = \mu$

**8.** **a.** $f_Y(y) = e^{-y}$ if $y \geq 0$. ($Y = \ln\left(\frac{1}{X}\right)$ is exponentially distributed with parameter $\lambda = 1$.

**b.** $E(Y) = \frac{1}{\lambda} = 1$ and $E(X) = \frac{1}{2}$, so $1 = E(Y) \neq \ln\left(\frac{1}{EX}\right) = \ln(2)$

**9.** **a**. $\lambda = \frac{1}{12}$. $var(X_i) = 144$ and $E(\sum_{i=1}^{6} X_i) = 72$.

**b.** $P(X_1 > 15 | X_1 > 3) = P(X_1 > 12) = e^{-1}$.

**10.** **a**. $4n, 6n$

**b**. $f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z - x) dx = \lambda^2 z e^{-\lambda z},$ *for* $z \geq 0$ *(where* $\lambda = \frac{1}{4}$*). And* $f_{X_1+X_2}(z) = 0$, *if* $z < 0$.

**c**. $P\left(\overline{X}_2 > 5\right) = e^{-\frac{5}{2}}$

**d.** $P\left(\overline{X}_{100} > 5\right) \overset{\text{CLT}}{\approx} 0.62\%$.