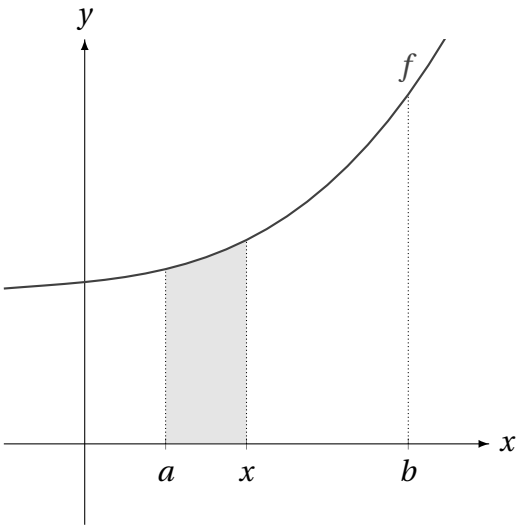
MATHEMATICS B_2

MIKE AUERBACH
WWW.MATHEMATICUS.DK



Mathematics B2
1st English edition, 2017

The document is written using the document preparation system LaTeX, see www.tug.org and www.miktex.org.

Figures and diagrams are written in *pgf/TikZ*, see www.ctan.org/pkg/pgf.

This document and others may be downloaded from www.mathematicus.dk. The content is primarily in Danish.

Preface

This document is a translation of the Danish "Matematik B2", which is a textbook on B level mathematics of the Danish stx. Since English is not my first language, I apologise in advance for errors in translation.

The primary aim is to provide a textbook without too much "clutter". Examples are kept to a minimum, and the text mainly covers the basic mathematics. It would therefore be a good idea to supplement the text with examples and other materials that cover specific uses of the mathematical tools.

Mike Auerbach

ORIGINAL PREFACE (IN DANISH)

Disse matematiknoter dækker kernestoffet (og en smule mere) for det andet år i et studieretningsforløb på B-niveau på stx. Noterne er skrevet med det formål at have en grundbog, som kun indeholder den grundliggende matematiske teori. I forbindelse med samarbejde i studieretningen eller med andre fag er det derfor nødvendigt at supplere med eksempler og andet materiale, der dækker konkrete anvendelser.

Til gengæld dækker noterne den rent matematiske fremstilling af kernestoffet på stx, hvilket ifølge min opfattelse gør dem velegnede til en første behandling af stoffet samt i forbindelse med eksamenslæsningen.

Til slut en stor tak til de mange matematikkolleger, der er kommet med rettelser og gode ændringsforslag. De fejl og mangler, der stadig måtte findes, er naturligvis udelukkende mit ansvar.

Mike Auerbach
Tornbjerg Gymnasium

Contents

1 Differential Calculus

Calculus is a branch of mathematics concerned with functions, and how they change. *Differential* calculus concerns itself with the growth of functions for specific values of x. This growth can be described by the slope of the graph at a specific point $(x, f(x))$. But since only straight lines have slopes, we need to associate the graph with straight lines, for which we can find the slope.

If the graph of a functions is nice and smooth, we can draw at each point on the graph, a straight line which follows the graph at the point in question. Such a line is called a *tangent*. An illustration can be seen in figure 1.1.

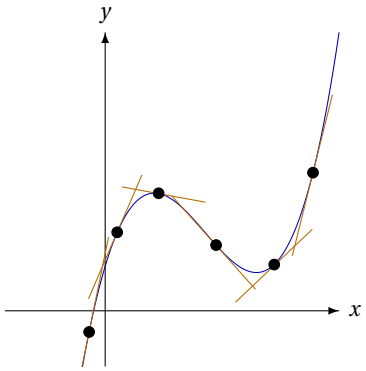


Figure 1.1: At each point on the graph, we may draw a tangent. Here, some of the tangents are illustrated by line segments.

Example 1.1
Here, we look at the function $f(x) = 3x^2 + 7$. The graph of this function passes through the point $P(5, 82)$. At this point, the graph has a tangent, see figure 1.2.

The slope of the tangent at this point is called $f'(5)$ (pronounced "*f prime* of 5"). If we already know that $f'(5) = 30$, then we can find an equation of the tangent.

The tangent is a straight line, so its equation has the form $y = ax + b$. Since we know that $f'(5) = 30$, we also know that the equation is $y = 30x + b$. The *point of tangency* is $(5, 82)$, therefore

$$82 = 30 \cdot 5 + b \qquad \Leftrightarrow \qquad b = 82 - 30 \cdot 5 = -68\,.$$

So, the tangent to the graph of f at the point $P(5, 82)$ has the equation

$$y = 30x - 68\,.$$

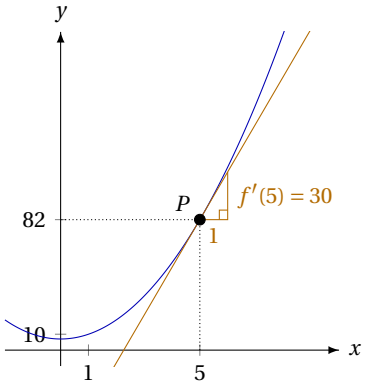


Figure 1.2: P is the point of tangency, so P lies on both the graph and the tangent.

In the example above, we saw that it is possible to find an equation of a tangent, if we already know its slope. The question now is, how do we find this slope?

It is of course possible to draw the tangent and then measure the slope—but this method is not very precise.

1.1 DERIVATIVES

If we can draw a tangent at each point of the graph of a function f, we can define a new function based on f, whose values are the slopes of the tangents at each point on the graph. We call this function the *derivative* of f. The derivative of f is denoted by f'.

It turns out that we can find a formula for $f'(x)$, if we know a formula for $f(x)$. The operation that turns $f(x)$ into $f'(x)$ is called *differentiation*. Not every function can be differentiated, but the ones that can are called differentiable.[1]

[1] We are not going to give a strict definition here, but simply note that for a function to be differentiable, its graph must be continuous (i.e. without holes) and *smooth*.

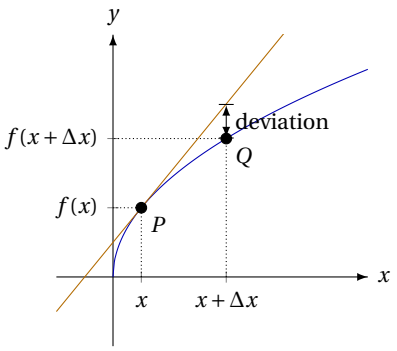


Figure 1.3: The graph of f passes through the points P and Q. Q does not lie on the tangent, but it is close if Δx is small.

To determine the slopes of the tangents at each point on the graph, we need to look at the tangents. Tangents are straight lines, and to determine the slope we need two points on the line. Here we have a problem, since we only know one point, P, the point of tangency.

We do not know the equation of the tangent, so it is not possible to calculate a second point. The best we can do is find another point Q on the graph close to the point of tangency, see figure 1.3.

If we calculate the slope of the tangent using the points P and Q, we will not get the right slope, but we will get a number, which can be used as an approximation. The smaller Δx is, the closer Q is to P, and the better the approximation. This is because the deviation marked in figure 1.3 gets smaller, the smaller Δx is.

A good approximation to the slope $f'(x)$ is therefore

$$f'(x) \approx \frac{\Delta f}{\Delta x},$$

where $\Delta f = f(x+\Delta x) - f(x)$, and Δx is small.

But we actually would like an *exact* value for the slope, and not just an approximation. We can get this by letting Δx have a value which is as small as possible, i.e. $\Delta x = 0$. But we cannot just let $\Delta x = 0$, since this would give us

$$\frac{\Delta f}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x} = \frac{f(x+0) - f(x)}{0} = \frac{0}{0},$$

which makes no sense.

What we then do is to rewrite and simplify the fraction $\frac{\Delta f}{\Delta x}$ to get to an expression, where we can set $\Delta x = 0$ without getting meaningless calculations. What we are trying to find out, is if there exists a number that $\frac{\Delta f}{\Delta x}$ would be equal to, if we were allowed to let $\Delta x = 0$.

Actually, we are investigating the value of $\frac{\Delta f}{\Delta x}$, when Δx approaches 0. We define this number to be the slope, $f'(x)$. [2]

[2] This description of $f'(x)$ is not exact. In reality, we look at something called the *limit* of $\frac{\Delta f}{\Delta x}$ as Δx approaches 0. If f is differentiable, this is a well-defined quantity.

Example 1.2

The graph of $f(x) = 3x^2 + 7$ passes through the point $P(x, f(x))$. The tangent to the graph of f at this point has slope $f'(x)$. To calculate this value, we will first find an approximate value of the slope using the points P and Q, see figure 1.4.

The point Q has coordinates $Q(x+\Delta x, f(x+\Delta x))$, so Δf is:

$$\begin{aligned}\Delta f &= f(x+\Delta x) - f(x)\\ &= (3\cdot(x+\Delta x)^2+7) - (3x^2+7)\\ &= 3x^2+6\cdot x\cdot\Delta x+3\cdot(\Delta x)^2+7-3x^2-7\\ &= 6\cdot x\cdot\Delta x+3\cdot(\Delta x)^2\end{aligned}$$

Now, we divide this by Δx to find an approximate value of $f'(x)$:

$$\frac{\Delta f}{\Delta x} = \frac{6\cdot x\cdot\Delta x+3\cdot(\Delta x)^2}{\Delta x} = 6x+3\cdot\Delta x\,.$$

So the slope of the tangent at $P(x, f(x))$ is approximately $6x+3\cdot\Delta x$.

The smaller Δx is, the closer this expression will be to the true slope. And the smaller Δx is, the closer $6x+3\cdot\Delta x$ will be to $6x$.

We therefore conclude that the derivative of the function $f(x)=3x^2+7$ is

$$f'(x) = 6x\,.$$

This method of finding the derivative gives us the following definition:

Definition 1.3

For a function f, we define the derivative f' to be the function such that

$$\frac{\Delta f}{\Delta x} \to f'(x) \text{ when } \Delta x \to 0,$$

where $\Delta f = f(x+\Delta x) - f(x)$.

The method we use to find derivatives, follows these three steps:

1. Calculate Δf, and simplify as much as possible.
2. Calculate $\frac{\Delta f}{\Delta x}$, and reduce as much as possible.
3. Determine what expression $\frac{\Delta f}{\Delta x}$ approaches when $\Delta x \to 0$. This is $f'(x)$.

For some reason, this is often called the *three-step method*.

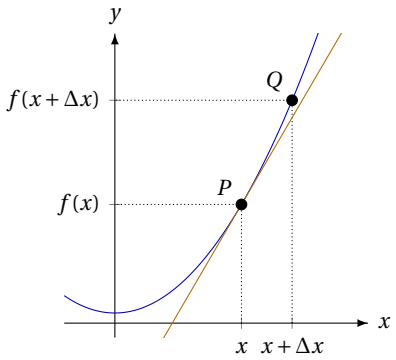


Figure 1.4: P is the point of tangency, so P lies on the graph as well as on the tangent. Q lies only on the graph, but is close to the tangent.

Terms and Notation

Definition 1.3 tells us how to find the derivative $f'(x)$, which is the function whose values are the tangent slopes at each point on the graph of $f(x)$.

To find the derivative, we look at the *difference quotient*[3] $\frac{\Delta f}{\Delta x}$. We investigate what happens to this quantity as Δx approaches 0. Because $f'(x)$ is the "result" of $\frac{\Delta f}{\Delta x}$, we sometimes also use the notation $\frac{\mathrm{d}f}{\mathrm{d}x}$ for the derivative of x.[4]

So, the following are equivalent:

1. The derivative of $f(x)=3x^2+7$ is $f'(x)=6x$.

[3] $\frac{\Delta f}{\Delta x}$ is called the "difference quotient" since Δf and Δx are differences, and the result of a division is called a quotient.

[4] Note that $\frac{\mathrm{d}f}{\mathrm{d}x}$ means exactly the same as $f'(x)$. I.e. $\frac{\mathrm{d}f}{\mathrm{d}x}$ is *not* a fraction or a quotient; we cannot separate $\mathrm{d}f$ and $\mathrm{d}x$.

2. The derivative of $f(x) = 3x^2 + 7$ is $\frac{\mathrm{d}f}{\mathrm{d}x} = 6x$.

The derivative f' is a function. But if we let x have a certain value, e.g. $x = c$, the value of f' at that x is the slope of the tangent to the graph of f at the point $(c, f(c))$.

Example 1.4
The function $f(x) = 3x^2 + 7$ has the derivative

$$f'(x) = 6x\,.$$

The graph of f passes through the point $(1, 10)$ (because $f(1) = 10$). At this point the tangent has the slope

$$f'(1) = 6 \cdot 1 = 6\,.$$

This can also be written as

$$\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=1} = 6\,.$$

1.2 VARIOUS DERIVATIVES

In this section, we find the derivatives of some simple functions.

Theorem 1.5

If $f(x) = c$, where c is a constant, the derivative is $f'(x) = 0$.

This follows from the fact that the graph of $f(x) = c$ is a line parallel to the x-axis, i.e. a line with slope 0. Since the value of $f'(x)$ is the slope of the tangent at each point on the graph, and the graph has slope 0 everywhere, we get $f'(x) = 0$. A more formal proof, using definition 1.3 would be the following:

Proof
If $f(x) = c$, then

$$\Delta f = f(x + \Delta x) - f(x) = c - c = 0\,.$$

Therefore

$$\frac{\Delta f}{\Delta x} = \frac{0}{\Delta x} = 0\,.$$

Since $\frac{\Delta f}{\Delta x} = 0$, regardless of the value of Δx, it follows that

$$\frac{\Delta f}{\Delta x} \to 0\,, \qquad \text{when } \Delta x \to 0\,.$$

I.e.

$$f'(x) = 0\,.$$

■

Theorem 1.6

If $f(x) = x$, then $f'(x) = 1$.

The graph of $f(x) = x$ is a straight line with slope 1. This actually proves the theorem. A more formal proof using definition 1.3 is left as an exercise to the reader.

Theorem 1.7

When $f(x) = x^2$, the derivative is $f'(x) = 2x$.

Proof

First, we calculate

$$\begin{aligned}\Delta f &= f(x+\Delta x) - f(x)\\ &= (x+\Delta x)^2 - x^2\\ &= x^2 + 2x\cdot\Delta x + (\Delta x)^2 - x^2\\ &= 2x\cdot\Delta x + (\Delta x)^2\,.\end{aligned}$$

Next, we calculate the quotient $\frac{\Delta f}{\Delta x}$

$$\frac{\Delta f}{\Delta x} = \frac{2x\cdot\Delta x + (\Delta x)^2}{\Delta x} = 2x + \Delta x\,.$$

If $\Delta x \to 0$, this expression approaches $2x$.

Therefore $f'(x) = 2x$. ■

Theorem 1.8

If $f(x) = \frac{1}{x}$, the derivative is $f'(x) = -\frac{1}{x^2}$.

Proof

If $f(x) = \frac{1}{x}$, we have

$$\begin{aligned}\Delta f &= f(x+\Delta x) - f(x)\\ &= \frac{1}{x+\Delta x} - \frac{1}{x}\\ &= \frac{x}{x\cdot(x+\Delta x)} - \frac{x+\Delta x}{x\cdot(x+\Delta x)}\\ &= \frac{-\Delta x}{x\cdot(x+\Delta x)}\,.\end{aligned}$$

I.e.

$$\frac{\Delta f}{\Delta x} = \frac{\frac{-\Delta x}{x\cdot(x+\Delta x)}}{\Delta x} = \frac{-1}{x\cdot(x+\Delta x)}\,.$$

When $\Delta x \to 0$, this expression approaches $\frac{-1}{x\cdot(x+0)} = \frac{-1}{x^2}$, and therefore

$$f'(x) = -\frac{1}{x^2}\,.$$ ■

Theorem 1.9

If $f(x) = \sqrt{x}$, the derivative is $f'(x) = \frac{1}{2\sqrt{x}}$.

Proof

If $f(x) = \sqrt{x}$, we have

$$\Delta f = f(x+\Delta x) - f(x) = \sqrt{x+\Delta x} - \sqrt{x}\,.$$

We cannot rewrite this expression, so we need to calculate the quotient $\frac{\Delta f}{\Delta x}$ directly. It turns out that we can rewrite the quotient like this:[5]

$$\begin{aligned}
\frac{\Delta f}{\Delta x} &= \frac{\sqrt{x+\Delta x} - \sqrt{x}}{\Delta x} \\
&= \frac{\left(\sqrt{x+\Delta x} - \sqrt{x}\right)\left(\sqrt{x+\Delta x} + \sqrt{x}\right)}{\Delta x \cdot \left(\sqrt{x+\Delta x} + \sqrt{x}\right)} \\
&= \frac{\left(\sqrt{x+\Delta x}\right)^2 - \left(\sqrt{x}\right)^2}{\Delta x \cdot \left(\sqrt{x+\Delta x} + \sqrt{x}\right)} \\
&= \frac{x+\Delta x - x}{\Delta x \cdot \left(\sqrt{x+\Delta x} + \sqrt{x}\right)} \\
&= \frac{\Delta x}{\Delta x \cdot \left(\sqrt{x+\Delta x} + \sqrt{x}\right)} \\
&= \frac{1}{\sqrt{x+\Delta x} + \sqrt{x}}\,.
\end{aligned}$$

This expression approaches $\frac{1}{\sqrt{x+0}+\sqrt{x}} = \frac{1}{2\sqrt{x}}$ when $\Delta x \to 0$, i.e.

$$f'(x) = \frac{1}{2\sqrt{x}}\,.$$

■

In table 1.1, some more derivatives are listed.

Example 1.10

According to theorem 1.9, the derivative of $f(x) = \sqrt{x}$ is $f'(x) = \frac{1}{2\sqrt{x}}$. Since the values of $f'(x)$ are the slopes of the tangents to the graph of f, we find that the tangent at the point $P(4,2)$ has the slope

$$f'(4) = \frac{1}{2\sqrt{4}} = \frac{1}{2\cdot 2} = \frac{1}{4}\,.$$

This is illustrated in figure 1.5.

So, the tangent is a straight line, and its equation is $y = \frac{1}{4}x + b$. If we want to find the value of b, we insert the point of tangency $P(4,2)$ into the equation:

$$2 = \tfrac{1}{4}\cdot 4 + b \qquad \Leftrightarrow \qquad b = 1\,.$$

At the point $P(4,2)$, the graph of $f(x) = \sqrt{x}$ has a tangent, whose equation is

$$y = \tfrac{1}{4}x + 1\,.$$

This can also be seen in figure 1.5.

[5] We multiply by $\sqrt{x+\Delta x} + \sqrt{x}$ in the numerator and the denominator; then we can use the rule

$$(a-b)(a+b) = a^2 - b^2\,.$$

Table 1.1: Various functions and their derivatives.

$f(x)$	$f'(x)$
k	0
x	1
x^2	$2x$
x^3	$3x^2$
x^n	nx^{n-1}
$\frac{1}{x}$	$-\frac{1}{x^2}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
e^x	e^x
e^{kx}	$k\,\mathrm{e}^{kx}$
a^x	$\ln(a)\cdot a^x$
$\ln(x)$	$\frac{1}{x}$

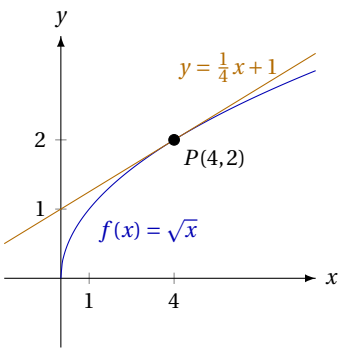


Figure 1.5: The graph of $f(x) = \sqrt{x}$ has a tangent with the equation $y = \frac{1}{4}x + 1$ at the point $P(4,2)$.

1.3 SUM AND DIFFERENCE

It turns out that in order to find the derivative f' of a given function f, it is not necessary to use the method from the previous section every time. We only need to know the derivative of a few simple functions, like the ones in the table above. This is because there are some rules for calculating derivatives of functions, which are "made up of" simpler functions.

Theorem 1.11

If the function p is differentiable, and $f(x) = c \cdot p(x)$, where c is a constant, then $f'(x) = c \cdot p'(x)$.

Proof

If $f(x) = c \cdot p(x)$, then

$$\Delta f = c \cdot p(x + \Delta x) - c \cdot p(x) = c \cdot \big(p(x + \Delta x) - p(x)\big) = c \cdot \Delta p \,.$$

I.e.

$$\frac{\Delta f}{\Delta x} = \frac{c \cdot \Delta p}{\Delta x} = c \cdot \frac{\Delta p}{\Delta x} \,.$$

If we let $\Delta x \to 0$, then $\frac{\Delta p}{\Delta x} \to p'(x)$, and then $c \cdot \frac{\Delta p}{\Delta x} \to c \cdot p'(x)$. Therefore

$$f'(x) = c \cdot p'(x) \,.$$ ■

In the following example, we demonstrate how to use this result.

Example 1.12

According to theorem 1.7, the derivative of $p(x) = x^2$ is $p'(x) = 2x$. But what is the derivative of $f(x) = 7x^2$?

We can now use theorem 1.11. If $f(x) = 7x^2$, then

$$f(x) = c \cdot p(x) \,, \qquad \text{where } c = 7 \text{ and } p(x) = x^2 \,.$$

Since we already know the derivative of $p(x) = x^2$, theorem 1.11 gives us

$$f'(x) = c \cdot p'(x) = 7 \cdot 2x = 14x \,.$$

So, we can find the derivative of $f(x) = 7x^2$, because we already know the derivative of x^2.

Example 1.13

If we want to find the derivative of $f(x) = 4x^3$, we write $f(x)$ as $f(x) = 4 \cdot p(x)$, where $p(x) = x^3$.

Looking up $p(x) = x^3$ gives us $p'(x) = 3x^2$. Then, according to theorem 1.11,

$$f'(x) = 4 \cdot p'(x) = 4 \cdot 3x^2 = 12x^2 \,.$$

Theorem 1.14

Let p and q be differentiable functions. If $f(x) = p(x) + q(x)$, then

$$f'(x) = p'(x) + q'(x) \,.$$

Proof
We use definition 1.3 and calculate

$$\begin{aligned}\Delta f &= f(x+\Delta x) - f(x) = \big(p(x+\Delta x) + q(x+\Delta x)\big) - \big(p(x)+q(x)\big)\\ &= p(x+\Delta x) - p(x) + q(x+\Delta x) - q(x)\\ &= \Delta p + \Delta q\,.\end{aligned}$$

Next, we get

$$\frac{\Delta f}{\Delta x} = \frac{\Delta p + \Delta q}{\Delta x} = \frac{\Delta p}{\Delta x} + \frac{\Delta q}{\Delta x}\,.$$

If we let $\Delta x \to 0$, then $\frac{\Delta p}{\Delta x} \to p'(x)$ and $\frac{\Delta q}{\Delta x} \to q'(x)$, which means that

$$f'(x) = p'(x) + q'(x)\,.$$ ■

Theorem 1.15

Let p and q be differentiable functions. If $f(x) = p(x) - q(x)$, then

$$f'(x) = p'(x) - q'(x)\,.$$

This theorem is very similar to theorem 1.14, and it can be proven in the same manner.

Example 1.16
The theorems 1.11, 1.14 and 1.15 can be combined if we need to differentiate more complicated functions.

The function

$$f(x) = 4x^2 + 5\ln(x) - 3x$$

combines the simpler functions x^2, $\ln(x)$ og x, whose derivatives are all listed in table 1.1.

Using theorems 1.14 and 1.15 we get

$$f'(x) = \big(4x^2\big)' + (5\ln(x))' - (3x)'\,.$$

Next, we use theorem 1.11 to get

$$f'(x) = 4\cdot\big(x^2\big)' + 5\cdot(\ln(x))' - 3\cdot(x)'\,.$$

We now find the derivatives of x^2, $\ln(x)$ and x in the table. Then we have

$$f'(x) = 4\cdot 2x + 5\cdot\frac{1}{x} - 3\cdot 1\,,$$

which can be simplified to

$$f'(x) = 8x + \frac{5}{x} - 3\,.$$

1.4 PRODUCTS, COMPOSITIONS AND QUOTIENTS

If we look at the theorems, we have proven so far, we might get the idea that we can differentiate any function by differentiating each part of the function independently. However, this is not the case, which the next theorem shows.

Theorem 1.17: The product rule

Let p and q be differentiable functions. If $f(x) = p(x) \cdot q(x)$, then

$$f'(x) = p'(x) \cdot q(x) + p(x) \cdot q'(x) .$$

Proof
If $f(x) = p(x) \cdot q(x)$, then

$$\Delta f = f(x + \Delta x) - f(x) = p(x + \Delta x) \cdot q(x + \Delta x) - p(x) \cdot q(x) .$$

In order to rewrite this expression, so it contains Δp as well as Δq, we use a trick: We subtract the term $p(x) \cdot q(x + \Delta x)$ and then add it again. This does not change anything:

$$\begin{aligned} \Delta f &= p(x + \Delta x) \cdot q(x + \Delta x) - p(x) \cdot q(x) \\ &= p(x + \Delta x) \cdot q(x + \Delta x) \underbrace{- p(x) \cdot q(x + \Delta x) + p(x) \cdot q(x + \Delta x)}_{\text{the sum of these two terms is 0}} - p(x) \cdot q(x) . \end{aligned}$$

Now we can factor out like terms to get

$$\begin{aligned} \Delta f &= \big(p(x + \Delta x) - p(x)\big) \cdot q(x + \Delta x) + p(x) \cdot \big(q(x + \Delta x) - q(x)\big) \\ &= \Delta p \cdot q(x + \Delta x) + p(x) \cdot \Delta q . \end{aligned}$$

Then

$$\frac{\Delta f}{\Delta x} = \frac{\Delta p \cdot q(x + \Delta x) + p(x) \cdot \Delta q}{\Delta x} = \frac{\Delta p}{\Delta x} \cdot q(x + \Delta x) + p(x) \cdot \frac{\Delta q}{\Delta x} .$$

If we let $\Delta x \to 0$, we have

$$\begin{aligned} \frac{\Delta p}{\Delta x} &\to p'(x) \\ q(x + \Delta x) &\to q(x) \\ p(x) &\to p(x) \\ \frac{\Delta q}{\Delta x} &\to q'(x) . \end{aligned}$$

Collectively, this gives us

$$f'(x) = p'(x) \cdot q(x) + p(x) \cdot q'(x) . \qquad \blacksquare$$

Example 1.18
We want to find the derivative of $f(x) = \sqrt{x} \cdot \ln(x)$, so we write $f(x)$ as $f(x) = p(x) \cdot q(x)$, where

$$p(x) = \sqrt{x}, \qquad q(x) = \ln(x).$$

In our table, we find that

$$p'(x) = \frac{1}{2\sqrt{x}}, \qquad q'(x) = \frac{1}{x}.$$

Theorem 1.17 then gives us

$$\begin{aligned} f'(x) &= p'(x) \cdot q(x) + p(x) \cdot q'(x) \\ &= \frac{1}{2\sqrt{x}} \cdot \ln(x) + \sqrt{x} \cdot \frac{1}{x}. \end{aligned}$$

This reduces to

$$f'(x) = \frac{\ln(x)}{2\sqrt{x}} + \frac{1}{\sqrt{x}} \quad \Rightarrow \quad f'(x) = \frac{\ln(x)+2}{2\sqrt{x}}.$$

The next theorem is about *composite functions*. These are functions which can best be described as "functions of functions". Some examples are

$$\begin{aligned} f(x) &= (\ln(x))^2, & g(x) &= \sqrt{x^3+4}, \\ h(x) &= \mathrm{e}^{6x+x^2}, & k(x) &= \ln(x^2+\mathrm{e}^x). \end{aligned}$$

To differentiate such a function, we need to separate it into an *outer function* and an *inner function*.[6] How to find the derivative is given by the following theorem:

[6] The function f is composed of an *inner function*, which is $q(x) = \ln(x)$, and an *outer function*, which is $p(q) = q^2$, because $\ln(x)$ is squared.

Theorem 1.19: The chain rule

Let p and q be differentiable functions. If $f(x) = p(q(x))$, then its derivative is

$$f'(x) = p'(q(x)) \cdot q'(x).$$

Proof

If $f(x) = p(q(x))$, then

$$\frac{\Delta f}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x} = \frac{p(q(x+\Delta x)) - p(q(x))}{\Delta x}. \tag{1.1}$$

Δq is by definition $\Delta q = q(x+\Delta x) - q(x)$, which gives us

$$q(x+\Delta x) = q(x) + \Delta q.$$

The quotient in the expression (1.1) can therefore be rewritten as

$$\frac{\Delta f}{\Delta x} = \frac{p(q(x)+\Delta q) - p(q(x))}{\Delta x}.$$

If we do not write explicitly that q depends on x, this can also be written as

$$\frac{\Delta f}{\Delta x} = \frac{p(q+\Delta q) - p(q)}{\Delta x}.$$

As long as Δq is not 0, we can multiply this fraction by Δq in the numerator and the denominator to get

$$\frac{\Delta f}{\Delta x} = \frac{p(q+\Delta q) - p(q)}{\Delta q} \cdot \frac{\Delta q}{\Delta x}. \tag{1.2}$$

The two factors on the right hand side of (1.2) can now be examined in turn:

The fraction $\frac{p(q+\Delta q)-p(q)}{\Delta q}$ can be written as $\frac{\Delta p}{\Delta q}$, where it is implicit that p is a function of q. Since $\Delta q = q(x+\Delta x) - q(x)$, $\Delta q \to 0$ when $\Delta x \to 0$, which means that

$$\frac{\Delta p}{\Delta q} \to p'(q)\,, \qquad \text{when } \Delta x \to 0\,.$$

For $\frac{\Delta q}{\Delta x}$ we have

$$\frac{\Delta q}{\Delta x} \to q'(x)\,, \qquad \text{when } \Delta x \to 0\,.$$

Collectively, we get from the equation (1.2) that

$$\frac{\Delta f}{\Delta x} \to p'(q)\cdot q'(x)\,, \qquad \text{when } \Delta x \to 0\,.$$

Now, if we remember that q is in fact a function of x, we have

$$f'(x) = p'(q(x))\cdot q'(x)\,. \qquad \blacksquare$$

Example 1.20
A function f is given by the formula $f(x) = \sqrt{x^2+3}$. So, f may be written as $f(x) = p(q(x))$, where

$$p(q) = \sqrt{q} \quad \text{and} \quad q(x) = x^2+3\,.$$

The derivatives of both of these functions can be found in a table:

$$p'(q) = \frac{1}{2\sqrt{q}} \quad \text{and} \quad q'(x) = 2x\,.$$

Theorem 1.19 now gives us

$$\begin{aligned} f'(x) &= p'(q(x))\cdot q'(x) \\ &= \frac{1}{2\sqrt{q}}\cdot 2x \\ &\overset{(*)}{=} \frac{1}{2\sqrt{x^2+3}}\cdot 2x \end{aligned}$$

At $(*)$, we replace q by x^2+3, since $q(x) = x^2+3$.

Simplifying this further gives

$$f'(x) = \frac{1}{2\sqrt{x^2+3}}\cdot 2x = \frac{x}{\sqrt{x^2+3}}\,.$$

Example 1.21
A function f has the formula $f(x) = \mathrm{e}^{x^2}$. To differentiate f, we write $f(x) = p(q(x))$, where

$$p(q) = \mathrm{e}^q\,, \qquad q(x) = x^2\,.$$

From our table of derivatives, we get

$$p'(q) = \mathrm{e}^q\,, \qquad q'(x) = 2x\,.$$

Then theorem 1.19 gives us

$$f'(x) = p'(q(x))\cdot q'(x) = \mathrm{e}^q\cdot 2x = \mathrm{e}^{x^2}\cdot 2x\,.$$

The theorems 1.17 and 1.19 can also be used to prove a theorem about the derivative of a quotient of functions. We have the following theorem:

Theorem 1.22: The quotient rule

Let p and q be differentiable, and $q(x) \neq 0$ for all x. Then if $f(x) = \frac{p(x)}{q(x)}$, its derivative is

$$f'(x) = \frac{p'(x) \cdot q(x) - p(x) \cdot q'(x)}{(q(x))^2} .$$

Proof

$f(x) = \frac{p(x)}{q(x)}$ can be rewritten, so we have

$$f(x) = p(x) \cdot \frac{1}{q(x)} .$$

This is a product of two functions. According to theorem 1.17, we must then have

$$f'(x) = p'(x) \cdot \frac{1}{q(x)} + p(x) \cdot \left(\frac{1}{q(x)}\right)' = \frac{p'(x)}{q(x)} + p(x) \cdot \left(\frac{1}{q(x)}\right)' . \tag{1.3}$$

To proceed further, we need to investigate $\left(\frac{1}{q(x)}\right)'$. This is the derivative of a composite function. Using theorem 1.19 gives us[7]

$$\left(\frac{1}{q(x)}\right)' = -\frac{1}{q(x)^2} \cdot q'(x) .$$

If we insert this result into (1.3), we get

$$\begin{aligned} f'(x) &= \frac{p'(x)}{q(x)} + p(x) \cdot \left(-\frac{1}{q(x)^2} \cdot q'(x)\right) \\ &= \frac{p'(x)}{q(x)} - \frac{p(x) \cdot q'(x)}{q(x)^2} \\ &= \frac{p'(x) \cdot q(x)}{q(x)^2} - \frac{p(x) \cdot q'(x)}{q(x)^2} \\ &= \frac{p'(x) \cdot q(x) - p(x) \cdot q'(x)}{q(x)^2} . \end{aligned}$$

■

[7] The expression $\frac{1}{q(x)}$ is composed of $s(q) = \frac{1}{q}$ and $q(x)$. Next, we use that

$$s'(q) = -\frac{1}{q^2} .$$

Example 1.23

Let $f(x) = \frac{x^2}{e^x}$. To find the derivative $f'(x)$, we write $f(x) = \frac{p(x)}{q(x)}$, where

$$p(x) = x^2 , \qquad q(x) = e^x .$$

From our table, we have

$$p'(x) = 2x , \qquad q'(x) = e^x .$$

Using theorem 1.22, we then get

$$f'(x) = \frac{p'(x) \cdot q(x) - p(x) \cdot q'(x)}{(q(x))^2}$$

$$= \frac{2x \cdot e^x - x^2 \cdot e^x}{(e^x)^2} .$$

Simplifying this result gives us

$$f'(x) = \frac{2x - x^2}{e^x} .$$

There are functions, where using one of the rules from the theorems 1.17, 1.19 or 1.22 is not enough to find the derivative. Sometimes we need to combine several rules.

Here, we have an extreme example:

Example 1.24
A function f has the formula

$$f(x) = \frac{1}{\sqrt{x^2 \cdot \ln(x)}} , \qquad x > 1 .$$

How do we differentiate this?

First we write $f(x) = p(q(x))$, where

$$p(q) = \frac{1}{q}, \qquad q(x) = \sqrt{x^2 \cdot \ln(x)} .$$

Here, we can easily differentiate $p(q)$, but what about $q(x)$? We need to write this as $q(x) = s(t(x))$, where

$$s(t) = \sqrt{t} , \qquad t(x) = x^2 \cdot \ln(x) .$$

Now, we need to differentiate t, so we write t as $t(x) = n(x) \cdot m(x)$,

$$n(x) = x^2 , \qquad m(x) = \ln(x) .$$

Here

$$n'(x) = 2x , \qquad m'(x) = \frac{1}{x} .$$

According to theorem 1.17, we now have

$$t'(x) = n'(x) \cdot m(x) + n(x) \cdot m'(x) = 2x \cdot \ln(x) + x^2 \cdot \frac{1}{x} .$$

This can be reduced to $t'(x) = 2x \cdot \ln(x) + x$.

We now have everything we need, and we can begin to work our way backwards through the many parts of the function:

$$q'(x) = s'(t(x)) \cdot t'(x) = \frac{1}{2\sqrt{t}} \cdot (2x \cdot \ln(x) + x) = \frac{1}{2\sqrt{x^2 \cdot \ln(x)}} \cdot (2x \cdot \ln(x) + x) .$$

We reduce this to

$$q'(x) = \frac{2x \cdot \ln(x) + x}{2\sqrt{x^2 \cdot \ln(x)}} .$$

At last, we then have

$$f'(x) = p'(q(x)) \cdot q'(x) = -\frac{1}{q^2} \cdot \frac{2x \cdot \ln(x) + x}{2\sqrt{x^2 \cdot \ln(x)}}$$

$$= -\frac{1}{\left(\sqrt{x^2 \cdot \ln(x)}\right)^2} \cdot \frac{2x \cdot \ln(x) + x}{2\sqrt{x^2 \cdot \ln(x)}} .$$

This may also be reduced, and we get

$$f'(x) = -\frac{2\ln(x) + 1}{2x^2 \cdot \ln(x) \cdot \sqrt{\ln(x)}} .$$

Using the theorems 1.11–1.22 and a table of derivatives allows us to differentiate any function. We therefore conclude this section with an overview of these theorems:

Theorem 1.25

If p and q are differentiable functions, and c is a constant, the following rules hold:

$$\begin{aligned}
f(x) = c \cdot p(x) &\Rightarrow f'(x) = c \cdot p'(x) . \\
f(x) = p(x) + q(x) &\Rightarrow f'(x) = p'(x) + q'(x) . \\
f(x) = p(x) - q(x) &\Rightarrow f'(x) = p'(x) - q'(x) . \\
f(x) = p(x) \cdot q(x) &\Rightarrow f'(x) = p'(x) \cdot q(x) + p(x) \cdot q'(x) . \\
f(x) = p(q(x)) &\Rightarrow f'(x) = p'(q(x)) \cdot q'(x) . \\
f(x) = \frac{p(x)}{q(x)} &\Rightarrow f'(x) = \frac{p'(x) \cdot q(x) - p(x) \cdot q'(x)}{q(x)^2} .
\end{aligned}$$

1.5 TANGENT EQUATIONS

The derivative can be used to find the slope of a tangent at any point on the graph of a function. If we know the slope and a point, it is possible to determine an equation for the tangent. Here, we give some examples.

Example 1.26

The function $f(x) = x^2 + 4x + 6$ has a tangent at the point $P(-1, f(-1))$. What is its equation?

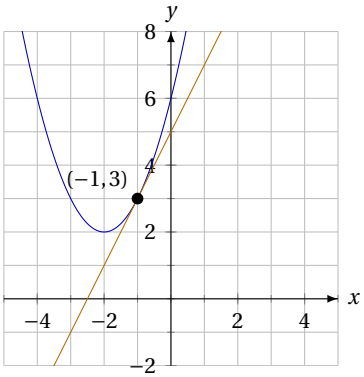


Figure 1.6: The graph of $f(x) = x^2 + 4x + 6$ has a tangent with the equation $y = 2x + 5$ at the point $P(-1,3)$.

The tangent is a straight line, so its equation has the form $y = ax + b$. Thus, we need to determine the two numbers a and b to write down the equation. a is the slope of the tangent, and this slope can be calculated using $f'(x)$. Therefore, we start out by finding the derivative of f:

$$f'(x) = 2x + 4 \cdot 1 - 0 = 2x + 4 .$$

The x-coordinate of the point of tangency is $x_0 = -1$, so the slope is

$$f'(-1) = 2 \cdot (-1) + 4 = 2 ,$$

and the equation of the tangent is $y = 2x + b$.

To determine the last number b, we need to know the point of tangency. The x-coordinate is $x_0 = -1$, the y-coordinate is

$$y_0 = f(-1) = (-1)^2 + 4 \cdot (-1) + 6 = 1 - 4 + 6 = 3 .$$

So, the point of tangency is $(-1,3)$. We insert this point into the equation of the tangent, i.e.

$$3 = 2 \cdot (-1) + b \qquad \Leftrightarrow \qquad b = 5\,.$$

Therefore our tangent has the equation

$$y = 2x + 5\,.$$

The graph and its tangent can be seen in figure 1.6.

Example 1.27
The function $g(x) = 3x + \ln(x)$ has a tangent at $P(1, f(1))$.

To determine the equation of this tangent, we first find the derivative of g,

$$g'(x) = 3 + \frac{1}{x}.$$

The slope of the tangent is then

$$a = f'(1) = 3 + \frac{1}{1} = 4\,,$$

and its equation is $y = 4x + b$.

To calculate b, we find the y-coordinate of the point of tangency

$$y_0 = f(1) = 3 \cdot 1 + \ln(1) = 3\,,$$

and we insert this number along with $x_0 = 1$ into the equation:

$$3 = 4 \cdot 1 + b \qquad \Leftrightarrow \qquad b = -1.$$

So the equation of the tangent is

$$y = 4x - 1\,.$$

As we see from these two examples, we are using the same method each time we determine the equation of the tangent. Therefore, we might turn this method into a formula. This is done in the following theorem.

Theorem 1.28

Let f be a differentiable function. Then the tangent to the graph of f at $P(x_0; f(x_0))$ has the equation

$$y = f'(x_0) \cdot (x - x_0) + f(x_0)\,.$$

Proof
The tangent is a straight line, so its equation has the form $y = ax + b$. Since $f'(x)$ is the slope of the tangent, and the point of tangency is $P(x_0, f(x_0))$, the slope must be

$$a = f'(x_0)\,.$$

Therefore the equation of the tangent is

$$y = f'(x_0) \cdot x + b\,. \tag{1.4}$$

To determine the y-intercept, b, we insert the point of tangency $P(x_0, f(x_0))$ into the tangent equation, which we then solve for b:

$$f(x_0) = f'(x_0) \cdot x_0 + b \qquad \Leftrightarrow \qquad b = -f'(x_0) \cdot x_0 + f(x_0)\,.$$

We insert this expression for b into the tangent equation (1.4), and get

$$y = f'(x_0) \cdot x - f'(x_0) \cdot x_0 + f(x_0)\,,$$

and factoring this equation then gives us

$$y = f'(x_0) \cdot (x - x_0) + f(x_0)\,. \qquad \blacksquare$$

Here, we show a few examples on how to use the formula:

Example 1.29
The function $f(x) = 3x^2 + 10$ has a tangent at $P(5, f(5))$. To find the equation of this tangent we use the formula

$$y = f'(x_0) \cdot (x - x_0) + f(x_0)$$

with $x_0 = 5$, i.e.

$$y = f'(5) \cdot (x - 5) + f(5)\,.$$

Before we can calculate the numbers, we need to find $f'(x)$:

$$f'(x) = 3 \cdot 2x + 0 = 6x\,.$$

Then we calculate

$$f'(5) = 6 \cdot 5 = 30$$
$$f(5) = 3 \cdot 5^2 + 10 = 85\,.$$

Inserting this into our formula, gives us

$$y = 30 \cdot (x - 5) + 85\,,$$

which can be simplified to

$$y = 30x - 65\,.$$

Example 1.30
The function $g(x) = (7x + 1) \cdot \mathrm{e}^x$ has a tangent at $P(0, g(0))$.

The equation of this tangent is

$$y = g'(0) \cdot (x - 0) + g(0) = g'(0) \cdot x + g(0)\,.$$

Now, we find[8]

$$g'(x) = 7 \cdot \mathrm{e}^x + (7x + 1) \cdot \mathrm{e}^x = (7x + 8) \cdot \mathrm{e}^x\,.$$

I.e.

$$g'(0) = (7 \cdot 0 + 8) \cdot \mathrm{e}^0 = 8 \cdot 1 = 8$$
$$g(0) = (7 \cdot 0 + 1) \cdot \mathrm{e}^0 = 1 \cdot 1 = 1\,.$$

Inserting these results into our equation above gives us

$$y = 8x + 1\,.$$

[8] The function is differentiated using the product rule, theorem 1.17.

Determining Points of Tangency

If we know a formula for a function, and a point on its graph, we can find an equation for the tangent at this point. But it is also possible to do the calculations the other way around: If we know the tangent, we can determine the point of tangency.

In this section, we show some examples.

Example 1.31
A function has the formula $f(x) = -x^2 + 3x + 1$.

The graph of this function has a tangent with the equation $y = x+2$. Where is the point of tangency?

The derivative of f is

$$f'(x) = -2x + 3\,,$$

and this gives us the slope of the tangent at each point on the graph.

We know the equation of the tangent, so we know that it has slope 1, i.e. $f'(x) = 1$ at the point of tangency. This gives us the equation

$$-2x + 3 = 1 \qquad \Leftrightarrow \qquad x = 1\,.$$

so, the x-coordinate of the point of tangency is 1. Now we just need to find the y-coordinate, which is

$$f(1) = -1^2 + 3 \cdot 1 + 1 = 3\,.$$

Therefore the point of tangency is $(1,3)$, see figure 1.7.

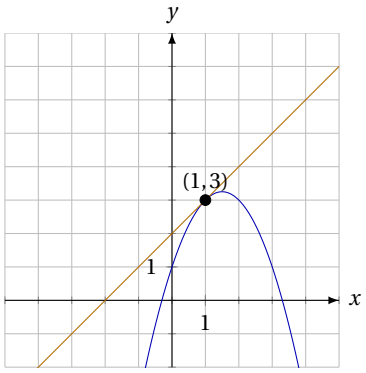


Figure 1.7: The line $y = x+2$ is a tangent to $f(x) = -x^2 + 3x + 1$ at $(1,3)$.

Example 1.32
The function f is given by the formula

$$f(x) = x - \frac{4}{x} + 3\,, \quad x > 0\,.$$

The graph of f has a tangent with slope 2. Where is its point of tangency, and what is its equation?

Since $f'(x)$ is the slope of the tangent, we need to know when $f'(x) = 2$. First, we find $f'(x)$,

$$f'(x) = 1 + \frac{4}{x^2}\,, \quad x > 0\,.$$

Then we solve the equation $f'(x) = 2$,

$$1 + \frac{4}{x^2} = 2 \qquad \Leftrightarrow \qquad \frac{4}{x^2} = 1 \qquad \Leftrightarrow \qquad x = -2 \vee x = 2\,.$$

The equation has two solutions, but since $f(x)$ is only defined for $x > 0$, we discard the negative solution. The x-coordinate of the point of tangency is then $x = 2$.

The y-coordinate is

$$f(2) = 2 - \frac{4}{2} + 3 = 3\,,$$

and the point of tangency is $(2,3)$, see figure 1.8.

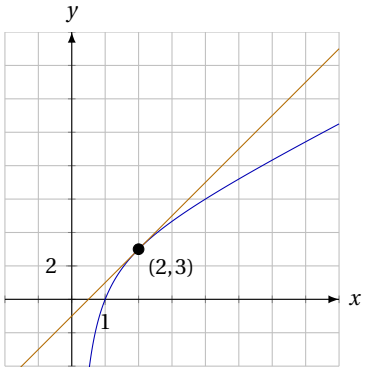


Figure 1.8: The graph of $f(x) = x - \frac{4}{x} + 3$ has a tangent with slope 2 at $(2,3)$.

According to theorem 1.28, the equation of the tangent is

$$y = f'(2) \cdot (x-2) + f(2)\,,$$

but since we already know the slope $f'(2) = 2$, and we know that $f(2) = 3$, this equation becomes

$$y = 2 \cdot (x-2) + 3\,,$$

which may be simplified to

$$y = 2x - 1\,.$$

Example 1.33
The graph of $f(x) = x^3 - 3x^2 - 21x + 5$ has two tangents with slope 3. What are their points of tangency?

The slope is 3, i.e. $f'(x) = 3$. To solve this we need to find $f'(x)$,

$$f'(x) = 3x^2 - 3 \cdot 2x - 21 \cdot 1 = 3x^2 - 6x - 21\,.$$

The equation $f'(x) = 3$ is therefore the quadratic equation

$$3x^2 - 6x - 21 = 3 \qquad \Leftrightarrow \qquad 3x^2 - 6x - 24 = 0\,.$$

If we solve this, we find the solutions

$$x = -2 \ \vee \ x = 4\,.$$

So, the two points of tangency are $(-2, f(-2))$ and $(4, f(4))$. Next, we determine the two y-coordinates

$$\begin{aligned} f(-2) &= (-2)^3 - 3 \cdot (-2)^2 - 21 \cdot (-2) + 5 = 27 \\ f(4) &= 4^3 - 3 \cdot 4^2 - 21 \cdot 4 + 5 = -63\,. \end{aligned}$$

Therefore, the two points of tangency are $(-2, 27)$ and $(4, -63)$. At these two points, the graph of f has a tangent with slope 3.

If we want to know the equations of these two tangents, we can find them in the same way as in example 1.32.

Example 1.34
In example 1.33, we saw how the graph of $f(x) = x^3 - 3x^2 - 21x + 5$ had two tangents with slope 3. Is there a slope a, such that the graph has exactly one tangent with this slope?

This question is a bit more complicated. We find the point of tangency by solving the equation $f'(x) = a$ for a given slope a, so the question may be rephrased as: Is there a number a, so the equation

$$f'(x) = a \tag{1.5}$$

has exactly one solution?

From example 1.33, we have

$$f'(x) = 3x^2 - 6x - 21\,.$$

so the equation (1.5) becomes

$$3x^2 - 6x - 21 = a \qquad \Leftrightarrow \qquad 3x^2 - 6x - 21 - a = 0\,.$$

This is a quadratic equation. If this equation is to have exactly one solution, its discriminant must be 0. The discriminant of this equation is[9]

$$d = (-6)^2 - 4 \cdot 3 \cdot (-21 - a) = 36 - 12 \cdot (-21 - a) = 288 + 12a\,.$$

If this is 0, then

$$288 + 12a = 0 \qquad \Leftrightarrow \qquad 12a = -288 \qquad \Leftrightarrow a = -24\,.$$

Therefore, there is exactly one tangent with slope $a = -24$.

Actually, by taking a closer look at the discriminant, we find that if $a > -24$ there are two tangents with slope a; but there are no tangents with slope a if $a < -24$.

Example 1.35

In this example, we look at the graph of $f(x) = x^2 + 3x + 6$. How many of its tangents pass through the point $P(2,7)$?

This is not a simple question, since the point P is not on the graph. In figure 1.9, we see an illustration of this. Here, we also see that there are two tangents to the graph that pass through P.

According to theorem 1.28 the equation of a tangent is

$$y = f'(x_0) \cdot (x - x_0) + f(x_0)\,.$$

The problem now consists of finding the points of tangency for those tangents that pass through $P(2,7)$. The point of tangency can be calculated from its x-coordinate x_0, but we do not know this coordinate.

What we do know is that the tangents pass through $P(2,7)$, therefore these coordinates fit into the equation of the tangent. This gives us

$$7 = f'(x_0) \cdot (2 - x_0) + f(x_0)\,. \tag{1.6}$$

We want to solve this equation to find x_0. To do this, we must know $f'(x)$, so we differentiate f:

$$f'(x) = 2x + 3\,.$$

We insert this along with the formula for the function f itself into equation (1.6), and get

$$7 = (2x_0 + 3) \cdot (2 - x_0) + (x_0^2 + 3x_0 + 6)\,,$$

which reduces to

$$7 = -2x_0 + x_0 + 6 + x_0^2 + 3x_0 + 6\,.$$

This we can simplify and get the quadratic equation

$$x_0^2 - 4x_0 - 5 = 0\,,$$

which has the solutions

$$x = -1 \;\vee\; x = 5\,.$$

Since there are two points of tangency, there are two tangents.

The y-coordinates of the points of tangency as well as the equations of the tangents can now be found by proceeding in the same way as in example 1.29.

[9] Remember that the discriminant is $d = B^2 - 4AC$, where A, B and C are the coefficients of the equation. (We write them as A, B and C, because we cannot denote the coefficient of the quadratic term by a, since we used this letter to denote the slope of the tangent.)

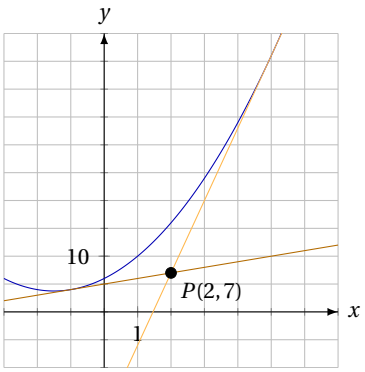


Figure 1.9: The graph of $f(x) = x^2 + 3x + 6$ has two tangents passing through $P(2,7)$.

1.6 MONOTONY INTERVALS AND EXTREMA

If for a certain function, the function value always increases whenever the independent variable increases, we call the function *increasing*.

If, however, the function value always decreases when the independent variable increases, the function is called decreasing.

More formally, we have the following definition:

Definition 1.36

Let a function f be defined on an interval.

1. If for any pair of arbitrary numbers x_1, x_2 in the interval
$$x_1 \leq x_2 \quad \Rightarrow \quad f(x_1) \leq f(x_2)\,,$$
the function is said to be *increasing* in the interval.
2. If for any pair of arbitrary numbers x_1, x_2 in the interval
$$x_1 \leq x_2 \quad \Rightarrow \quad f(x_1) \geq f(x_2)\,,$$
the function is said to be *decreasing* in the interval.

Notice that the definition only mentions the behaviour of functions in an *interval*. If we only look at a single point, it makes no sense to talk about whether the function is increasing or decreasing. The properties *increasing* and *decreasing* only applies to intervals, not points.

Example 1.37
The graph of the function $f(x) = 2x+1$ is a straight line with positive slope. This function is therefore increasing.

Conversely, a straight line with a negative slope is the graph of a decreasing function (e.g. $f(x) = -4x+3$).

A function, which is either increasing or decreasing everywhere, is called a *monotonous* functions. Not every function is monotonous. A lot of functions are decreasing in some intervals and increasing in others.

When we describe where a function is increasing and decreasing, we find what we call the *monotony intervals*, i.e. the intervals in which the function is entirely increasing or decreasing.

Example 1.38
In figure 1.10, we see the graph of the function

$$f(x) = x^2 - 4x + 1.$$

We have also drawn the vertical line $x = 2$. We see that on the left hand side of this line, the functions is decreasing, while it is increasing on the right hand side.

So, we say that $f(x)$ is decreasing when $x \leq 2$, and increasing when $x \geq 2$.

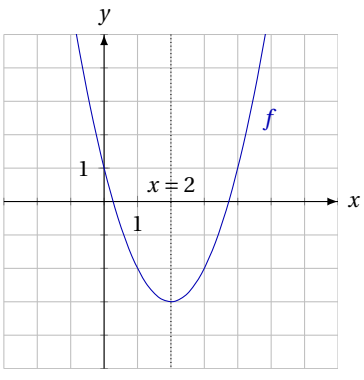


Figure 1.10: The graph of $f(x) = x^2 - 4x + 1$.

These are the monotony intervals.

In example 1.38, we found the monotony intervals by looking at the graph. We can always graph a function to find the monotony intervals, but this method lacks precision.

It would therefore be nice if we had a way of calculating where the graph changes from increasing to decreasing or vice versa; we would like to be able to do this by looking only at a formula of the function.

From example 1.37, we know that if the graph is a straight line, its intervals of monotony are determined by the slope. If the slope is positive, the function is increasing, if it is negative, the function is decreasing for every value of x.

But tangents to a graph of any given function are straight lines, and their slopes are given by $f'(x)$, therefore the following theorem makes sense intuitively,

Theorem 1.39

Let f be a differentiable function.

1. If f is increasing in the interval $[a;b]$, then $f'(x) \geq 0$ for all $x \in]a;b[$.
2. If f is decreasing in the interval $[a;b]$, then $f'(x) \leq 0$ for all $x \in]a;b[$.
3. If f is constant in the interval $[a;b]$, then $f'(x) = 0$ for all $x \in]a;b[$.

Notice that when $f(x)$ is increasing, the tangent slope is not necessarily positive in the entire interval. It may be 0 in some subset of the interval. This actually follows from definition 1.36, where $f(x_1)$ does not need to be greater than $f(x_2)$ when $x_1 \leq x_2$, but only greater than *or equal to*. Increasing and decreasing functions may therefore be constant in an interval.[10]

[10] Actually, it follows from definition 1.36 that a constant function is both increasing and decreasing. This may seem contradictory, but it is the case nonetheless.

Theorem 1.39 can be used to find some of the properties of $f'(x)$, if we already know the monotony intervals. Normally, we would instead try to determine the monotony intervals using $f'(x)$. To do this, we use the following theorem:

Theorem 1.40

Let f be a differentiable function.

1. If $f'(x) > 0$ for all x in the interval $]a;b[$, then f is increasing in $[a;b]$.
2. If $f'(x) < 0$ for all x in the interval $]a;b[$, then f is decreasing in $[a;b]$.
3. If $f'(x) = 0$ for all x in the interval $]a;b[$, then f is constant in $[a;b]$.

If we want to determine the monotony intervals for a function f, we need to investigate f' to find out, when $f'(x)$ changes sign from positive to negative or vice versa. If the value of $f'(x)$ changes from positive to negative, it has to pass through 0. Therefore we need to know, when $f'(x) = 0$.

This is illustrated in the following example.

Example 1.41
Here we look at the same function as in example 1.38,

$$f(x) = x^2 - 4x + 1\,.$$

To find out when the graph changes from increasing to decreasing, we need to find out, where $f'(x) = 0$. Therefore, we first determine $f'(x)$,

$$f'(x) = 2x - 4\,.$$

The equation $f'(x) = 0$ is then

$$2x - 4 = 0 \qquad \Leftrightarrow \qquad x = 2\,.$$

At $x = 2$, the graph has a tangent with slope 0, i.e. a horizontal tangent. We can also see this in figure 1.11.

From the graph, we see that the function is decreasing before $x = 2$ and increasing after. If we do not have the graph, we need to find the sign of $f'(x)$ by calculation.

If we want to find out, whether $f'(x)$ is positive or negative when $x < 2$, we choose some number less than 2, which we insert into the formula for f'. A number less than 2 could e.g. be 0, which would give us

$$f'(0) = 2 \cdot 0 - 4 = -4\,.$$

Since $-4 < 0$ we conclude that $f'(x)$ is negative for all $x < 2$, i.e. f is decreasing for $x \le 2$.[11]

In the same way, we may choose a number greater than 2, e.g. 3, and calculate

$$f'(3) = 2 \cdot 3 - 4 = 2 > 0\,,$$

i.e. $f'(x)$ is positive for all $x > 2$, and $f(x)$ is increasing for $x \ge 2$.

So, we can describe the monotony intervals of f by saying that $f(x)$ is decreasing for all $x \le 2$ and increasing for all $x \ge 2$.[12]

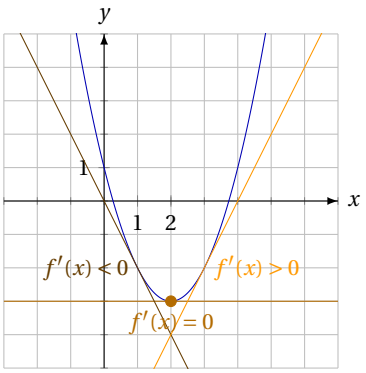


Figure 1.11: The graph of $f(x) = x^2 - 4x + 1$ is decreasing before $x = 2$ and increasing after $x = 2$. At $x = 2$, we have a horizontal tangent.

[11] We know from our earlier calculations that $f'(x)$ will only be 0 for $x = 2$. Therefore the value of $f'(x)$ will have the same sign for alle numbers $x < 2$, and we need only check the sign if $f'(x)$ for one number less than 2—here we used $x = 0$.

[12] The numbers 0 and 3, which we used to calculate the sign of f', have nothing to do with the monotony intervals. We just needed two numbers, one less than 2 and one greater than 2.

Sign Table

The monotony intervals of the function in example 1.41 can also be described in a *sign table*. Such a table could look like this:

x:		2	
$f'(x)$:	−	0	+
$f(x)$:	↘	min.	↗

From the table, we see that before $x = 2$, $f'(x) < 0$, and after $x = 2$, $f'(x) > 0$. This is illustrated by the $-$ and the $+$ in the table. In the last line, we see that this shows us, where $f(x)$ is decreasing, and where it is increasing (illustrated by $\searrow$ and $\nearrow$).

Using this table, we can easily describe the monotony intervals.[13] We can also see something else. At $x = 2$, the function f has a *minimum*,, i.e. a point on the graph, where the function values assumes its lowest possible value.

We see that it is a minimum, because the function decreases and then increases, when we pass $x = 2$. In this case, it is actually a *global* minimum,, because it is the lowest point on the entire graph. If a minimum is not global, we call it a *local* minimum. In the same way, we can talk about global and local maxima. Figure 1.12 illustrates this.

A collective term for these points is *extrema*. An *extremum* is a point on the graph, where it has a minimum or a maximum (local or global).

[13] The sign table is not equal to the monotony intervals, but it *describes* the monotony intervals.

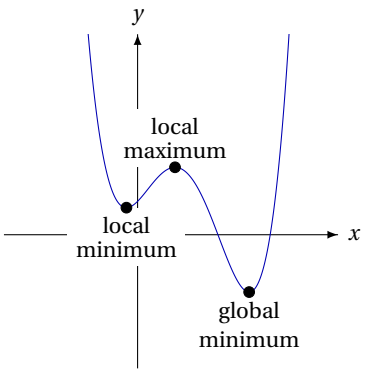


Figure 1.12: Functions may have local and global extrema.

Example 1.42
In this example, we will find the monotony intervals and the extrema of $f(x) = x^3 - 6x^2 + 9x + 1$.

The derivative is

$$f'(x) = 3x^2 - 12x + 9\,,$$

i.e. the equation $f'(x) = 0$ is the quadratic

$$3x^2 - 12x + 9 = 0\,,$$

which has solutions $x = 1$ and $x = 3$.

The two solutions divide the number line into three intervals: The numbers less than 1, the numbers between 1 and 3, and the numbers greater than 3. Now, we choose an arbitrary number from each of these intervals to determine the sign of $f'(x)$ in the intervals:

$$\begin{aligned} x < 1: &\quad f'(0) = 3 \cdot 0^2 - 12 \cdot 0 + 9 = 9 > 0 \\ 1 < x < 3: &\quad f'(2) = 3 \cdot 2^2 - 12 \cdot 2 + 9 = -3 < 0 \\ x > 3: &\quad f'(5) = 3 \cdot 5^2 - 12 \cdot 5 + 9 = 24 > 0 \end{aligned}$$

Now, we may draw a sign table:

x:		1		3	
$f'(x)$:	$+$	0	$-$	0	$+$
$f(x)$:	$\nearrow$	maks.	$\searrow$	min.	$\nearrow$

From this table, we can find the monotony intervals:

> $f(x)$ is increasing for all $x \le 1$ and for all $x \ge 3$, and decreasing for all $1 \le x \le 3$.

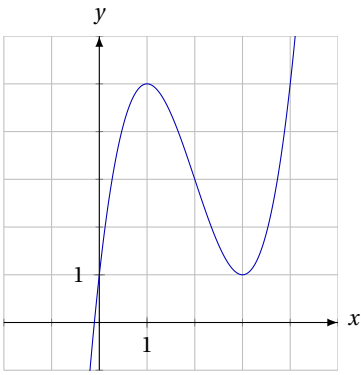


Figure 1.13: The graph of $f(x) = x^3 - 6x^2 + 9x + 1$ has a local maximum and a local minimum.

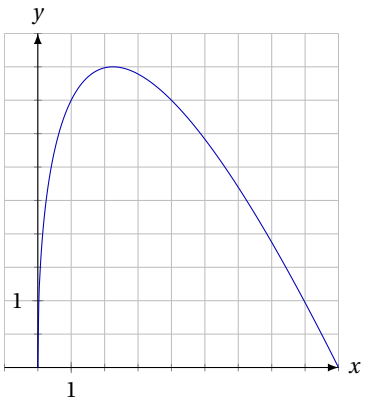


Figure 1.14: The graph $f(x) = 6 \cdot \sqrt{x} - 2x$ appears to have a global maximum.

[14]Here, it is important to remember that f is only defined for $x > 0$, so we cannot let x be 0 or negative.

Since we know that the monotony intervals separate at $x = 1$ and $x = 3$, we can also find the intervals on the graph (see figure 1.13) instead of writing down the table.

From the sign table, we see that there are two local extrema. The first extremum is a local maximum at $x = 1$, the second is a local minimum at $x = 3$.

We find the y-coordinates of these extrema:

$$f(1) = 1^3 - 6 \cdot 1^2 + 9 \cdot 1 + 1 = 5$$
$$f(3) = 3^3 - 6 \cdot 3^2 + 9 \cdot 3 + 1 = 1\,.$$

So, f has a local maximum at $(1,5)$ and a local minimum at $(3,1)$.

Example 1.43

In this example, we will find the extrema of

$$f(x) = 6 \cdot \sqrt{x} - 2x\,, \quad x > 0\,.$$

The graph of this function is seen in figure 1.14. Here, we see that it looks like f has a global maximum near $x = 2$.

To determine, whether the function has a maximum, we first find

$$f'(x) = 6 \cdot \frac{1}{2 \cdot \sqrt{x}} - 2 \cdot 1 = \frac{3}{\sqrt{x}} - 2\,.$$

The equation $f'(x) = 0$ is then

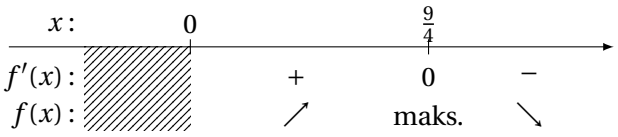
$$\frac{3}{\sqrt{x}} - 2 = 0 \quad \Leftrightarrow \quad 2\sqrt{x} = 3 \quad \Leftrightarrow \quad x = \left(\frac{3}{2}\right)^2 = \frac{9}{4}\,.$$

So there is a possible extremum at $x = \frac{9}{4}$.

To draw a sign table, we look at $f'(x)$ for $x < \frac{9}{4}$ and for $x > \frac{9}{4}$.[14]

$$0 < x < \tfrac{9}{4}: \quad f'(1) = \frac{3}{\sqrt{1}} - 2 = 1 > 0$$
$$x > \tfrac{9}{4}: \quad f'(9) = \frac{3}{\sqrt{9}} - 2 = -1 < 0$$

The sign table looks like this



The hatched area shows that the function is not defined for $x \leq 0$.

Using the sign table, we see that the graph increases until $x = \frac{9}{4}$, and then decreases. So, the function has a global maximum at $x = \frac{9}{4}$. The y-coordinate is

$$f\left(\frac{9}{4}\right) = 6 \cdot \sqrt{\frac{9}{4}} - 2 \cdot \frac{9}{4} = 6 \cdot \frac{3}{2} - \frac{9}{2} = \frac{9}{2}\,.$$

Therefore f has a global maximum at $\left(\frac{9}{4}, \frac{9}{2}\right)$.

Inflection Points

Looking at the examples above, we might think that each time a graph has a horizontal tangent it changes from increasing to decreasing or vice versa. This is, however, not always the case, which we will see in the next example.

Example 1.44
Here, we investigate the function

$$f(x) = x^3 - 12x^2 + 48x - 62$$

in order to find its monotony intervals.

First, we determine $f'(x)$

$$f'(x) = 3x^2 - 12 \cdot 2x + 48 \cdot 1 = 3x^2 - 24x + 48\,,$$

and then we solve $f'(x) = 0$, which is the quadratic

$$3x^2 - 24x + 48 = 0\,.$$

It turns out that this quadratic has only one solution:

$$x = 4\,.$$

Next, we determine the sign of $f'(x)$ for $x < 4$ and $x > 4$,

$$\begin{aligned} x < 4: &\quad f'(0) = 3 \cdot 0^2 - 24 \cdot 0 + 48 = 48 > 0 \\ x > 4: &\quad f'(5) = 3 \cdot 5^2 - 24 \cdot 5 + 48 = 3 > 0\,. \end{aligned}$$

Therefore, our sign table looks like this

x:		4	
$f'(x)$:	+	0	+
$f(x)$:	↗	?	↗

$f'(4) = 0$, which means there is a horizontal tangent at $x = 4$, but we have neither a minimum nor a maximum, since the function is increasing before and after $x = 4$. The situation is illustrated in figure 1.15.

We say that the graph has an *inflection point*; the sign table looks like this

x:		4	
$f'(x)$:	+	0	+
$f(x)$:	↗	infl.	↗

and the function f is increasing for all x.

The term *inflection point* refers to the way the graph curves. It is not the graph itself that is inflected, but its curvature. If we look at the graph, we might notice that it looks like ⌒ before the inflection point, and like ⌣ after the inflection point.

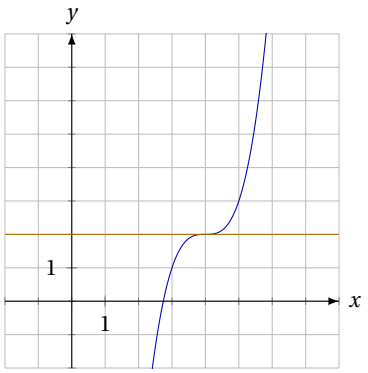


Figure 1.15: The point $(4, f(4))$ is an inflection point on the graph of $f(x) = x^3 - 12x^2 + 48x - 62$.

Summary of the Method

We conclude this section with a general recipe for determining monotony intervals and extrema for a given function $f(x)$:

1. Determine $f'(x)$.
2. Solve the equation $f'(x) = 0$. The solutions are the values of x at which we have extrema or inflection points.
3. The solutions of $f'(x) = 0$ divide the x-axis into intervals. Determine the sign of $f'(x)$ in each of these intervals by inserting a number from each interval into the formula for $f'(x)$.

   It is also possible to graph the function to investigate its behaviour in each of the intervals. In that case, this calculation and the sign table are not needed.
4. Write down a sign table.
5. Use the sign table to draw your conclusions. If we want to determine a maximum or a minimum, we need to calculate the y-coordinate as well.

1.7 OPTIMISATION

In the last section, we described how to find the extrema of a function. We can use this to *optimise* a given quantity. The purpose of optimisation is to find out when a given quantity is as large or as small as possible.

If the quantity we wish to optimise is given as a function of one variable, all we need to do is determine the maximum or the minimum. However, things are not always this simple. E.g. if we want to determine when a given area is as large as possible, the area might depend on both a length and a width. If this is the case, we need to know how the length and the width are connected.

How to actually do this, is most easily illustrated by examples.

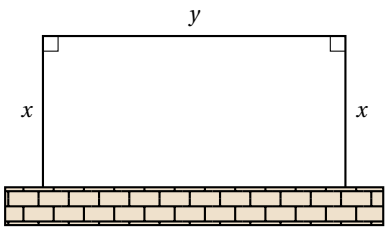


Figure 1.16: A fence around a chicken coop. One side of the area is walled.

Example 1.45
In a garden, we want to build a fence around a chicken coop (see figure 1.16). One side of the garden is walled, so we need only fence 3 sides of a rectangle. If we have 20 m of fence, how should we build the fence, so the enclosure has the largest possible area?

The length and the width of the rectangle, which make up the chicken coop, we call x and y, see figure 1.16. The total length of the fence must then correspond to the length of the three sides, i.e. $2x + y$. Since our fence is 20 m, we have

$$2x + y = 20\,.$$

Isolating y in this equation yields

$$y = 20 - 2x\,.$$

The area of the rectangle is $A = x \cdot y$, and this is the quantity that needs to be as large as possible. The quantity depends on two variables, x, and y,

so we cannot determine the largest value straight away. But we just found out that $y = 20 - x$, therefore this area may also be calculated as

$$A = x \cdot y = x \cdot (20 - 2x) = 20x - 2x^2 \,,$$

and this expression only depends on x.[15]

Where does this area have a maximum? To find the possible extrema of the function, we use the method, we used in the preceeding section, i.e. we solve $A' = 0$.

Since $A = 20x - 2x^2$, we get

$$A' = 20 \cdot 1 - 2 \cdot 2x = 20 - 4x \,,$$

So, the equation $A' = 0$ is

$$20 - 4x = 0 \qquad \Leftrightarrow \qquad x = 5 \,.$$

Now, we know that there is a possible maximum for the area where $x = 5$. We graph $A = 20x - 2x^2$ (see figure 1.17), and here we clearly see that $x = 5$ corresponds to a maximum.

Therefore, the area has a maximum where $x = 5$. This then gives us $y = 10$, and the area $A = 50$, which we also see in the figure.

Example 1.46

We want to build a cylindrical container with a volume of 1 l, such that we use as little material as possible. We can assume that the used material has the same thickness every—which means that we use the least amount of material, when the surface area is as small as possible.

A cylinder can be described by two parameters: Its radius r (at the top and the bottom) and its height h, see figure 1.18. Since the volume is measured in liters, and $1\,\text{l} = 1\,\text{dm}^3$, r and h are measured in decimeters.

The volume of a cylinder is

$$V = \pi r^2 h \,,$$

and since the volume is 1 l, we have

$$\pi r^2 h = 1 \qquad \Leftrightarrow \qquad h = \frac{1}{\pi r^2} \,. \tag{1.7}$$

The surface area of a cylinder is

$$A = 2\pi r^2 + 2\pi r h \,.$$

If we insert the expression for g from (1.7), we get

$$A = 2\pi r^2 + 2\pi r \cdot \frac{1}{\pi r^2} = 2\pi r^2 + \frac{2}{r} \,.$$

Now, the area is a function of r. Where the area is smallest, we have $A' = 0$. Since

$$A' = 4\pi r - \frac{2}{r^2} \,,$$

[15] Notice that $0 < x < 10$. We have $x > 10$ because x is a length, and we have $x < 10$ because we only have 20 m of fence. The two sides with length x must therefore have a total length of less than 20 m. This means that solutions for x which are not in the interval from 0 to 10, must be discarded.

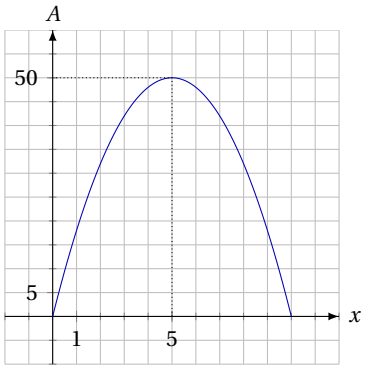


Figure 1.17: At $x = 5$, we have the largest area.

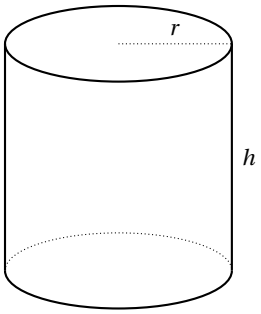


Figure 1.18: A cylinder can be described by its height and radius.

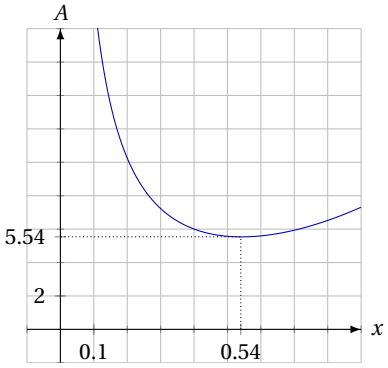


Figure 1.19: We have the smallest surface area, when the radius is 0.54 dm.

we therefore have the equation

$$4\pi r - \frac{2}{r^2} = 0\,,$$

which has the solution

$$r = \sqrt[3]{\frac{1}{2\pi}} \approx 0.54 \text{ dm}\,.$$

That this is indeed a minimum can be seen in figure 1.19.

When we know the radius, $r = 0.54$ dm, we can calculate the height, since equation (1.7) gives us

$$h = \frac{1}{\pi \cdot 0.54^2} = 1.08 \text{ dm}\,.$$

A cylindrical container with a volume of 1 l, therefore, has the least surface area, when the radius is $r = 0.54$ dm and the height is $h = 1.08$ dm.

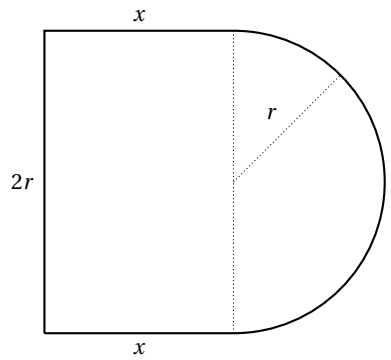


Figure 1.20: A 10 m^2 flower bed is made up of a rectangle and a half circle.

Example 1.47
In a garden, we want to plant a 10 m^2 flower bed. The shape of the flower bed is a figure made of a rectangle and a half circle, see figure 1.20.

We want to place decorative stones around the edge of the flower bed, so we want to minimise the perimeter. In that case, what is then the length of x and r in the figure?

The flower bed is made up of a rectangle with sides x and $2r$, and a half circle, with radius r. Its area is then

$$A = 2r \cdot x + \frac{\pi r^2}{2}\,.$$

Since the area is 10 m^2, this is equal to 10. Next, we isolate x.

$$2rx + \frac{\pi r^2}{2} = 10 \qquad \Leftrightarrow \qquad x = \frac{5}{r} - \frac{\pi r}{4}\,. \tag{1.8}$$

We want to minimise the perimeter. Since the perimeter consists of three straight lines and a half circle, the perimeter is

$$O = 2r + 2x + \pi r\,.$$

We insert the expression for x from (1.8) into this expression for the perimeter, and we get

$$O = 2r + 2 \cdot \left(\frac{5}{r} - \frac{\pi r}{4}\right) + \pi r = \left(2 + \frac{\pi}{2}\right) r + \frac{10}{r}\,.$$

To minimise this expression, we differentiate and find

$$O' = 2 + \frac{\pi}{2} - \frac{10}{r^2}\,.$$

We let this be equal to 0, and get the equation

$$2 + \frac{\pi}{2} - \frac{10}{r^2} = 0\,,$$

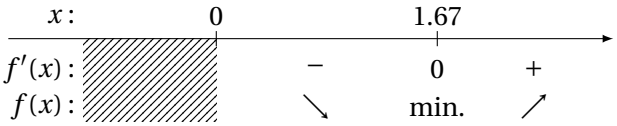
which has the solution

$$r = \frac{10}{\sqrt{20+5\pi}} \approx 1.67\,\text{m}\,.$$

To find out, if this really *is* a minimum, we construct a sign table for O' for values of x greater than or less than 1.67.

$$0 < r < 1.67: \quad O'(1) = 2 + \frac{\pi}{2} - \frac{10}{1^2} = -6.43 < 0$$
$$r > 1.67 \quad O'(2) = 2 + \frac{\pi}{2} - \frac{10}{2^2} = 1.07 > 0$$

The sign table looks like this



and we have a minimum, when the radius of the circle $r = 1.67$ m.

The length x is then (we use the result from (1.8))

$$x = \frac{5}{1.67} - \frac{\pi \cdot 1.67}{4} = 1.67\,\text{m}\,.$$

Summary of the Method

The method we used in the examples above can be described in the following way.

1. Translate a condition (e.g. fixed perimeter, fixed area, fixed volume) into an equation. Then isolate one of the variables in this equation.
2. Write down an expression for the quantity, you wish to optimise, and replace one of the variables with the expression found in step 1. You now have a function of one variable.
3. Determine the extrema of the function found in step 2. Now, you can determine the remaining measurements.

In principle, it is possible to have more than two variables in the expression, we wish to optimise. Then we need more than one condition to write the expression as a function of one variable. This corresponds to repeating steps 1–2.

1.8 RATE OF CHANGE

Using differentiation, we may find out where certain quantities have maxima and minima. This can, for instance, be used for optimisation. But we can also use differentiation to determine how fast certain quantities grow at certain points.

We have the following definition.[16]

[16] Notice that in this definition, the independent variable is called t instead of x. In principle, we could have used x, but we use t to emphasise that we are talking about *tid*.

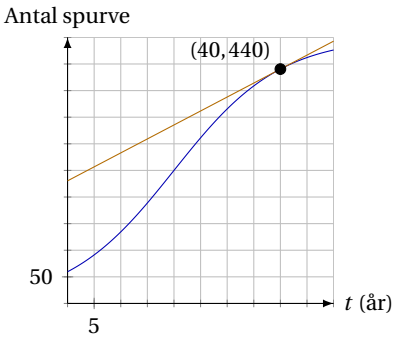


Figure 1.21: The population of sparrows at time t (in years).

Definition 1.48

Let $f(t)$ be a function, where t is the time. Then $f'(t)$ is the *rate of change* at the time t.

Example 1.49

In figure 1.21, we see the graph of $f(t)$, which shows us how the amount of sparrows on a certain island increases over time (measured in years).

In the figure, we see the graph passing through the point $(4,440)$. We have also drawn a tangent through this point—the slope of the tangent is 5.25. In other words

$$f(40) = 440 \qquad \text{and} \qquad f'(40) = 5.25\,.$$

This is a purely mathematical description, which may be translated into

1. After 40 years, there are 440 sparrows on the island.
2. After 40 years, the amount of sparrows increases at a rate of 5.25 sparrows per year.

Example 1.50

A jug of lukewarm water is put into a refrigerator. The temperature of the water can then be described by the function

$$f(t) = 5 + 15 \cdot e^{-0.01 \cdot t}\,,$$

where the time t is measured in minutes.

From this function, we can determine the rate of change $f'(45)$. First we calculate

$$f'(t) = 0 + 15 \cdot (-0.01) \cdot \mathrm{e}^{-0{,}01 \cdot t} = -0.15 \cdot \mathrm{e}^{-0.01 \cdot t}\,,$$

and then

$$f'(45) = -0.15 \cdot \mathrm{e}^{-0.01 \cdot 45} = -0.096\,.$$

What does this number tell us?

First of all, we notice that the number is negative, i.e. the temperature is *decreasing*. The value of the number shows us how much. Since $f'(45) = -0{,}096$, we have the following interpretation:

After 45 minutes in the refrigerator, the temperature of the water decreases at a rate of 0.096°C per minute.

Descriptive Statistics 2

Descriptive statistics is a branch of mathematics concerned with *describing* certain data. The purpose is to arrive at a description or a summary of a (possibly large) data set.

We can describe a data set in several ways. We might

- list the data set in a table, possibly grouping some of the numbers,
- calculate some *descriptors*, i.e. certain numbers that describe the data set, or
- draw diagrams to illustrate the data set.

2.1 STATISTICS WITH UNGROUPED DATA

The term "ungrouped data" refers to—as the name implies—data that have not been sorted into groups, i.e. raw data.

Imagine, we ask a class of 25 students how many hours of TV they watched yesterday. The answers might be like those listed in table 2.1.

Table 2.1: TV habits of students.

Hours of TV (unsorted)				
1	2	1	1	3
3.5	0	0.5	2	1
2.5	2	0.5	1	1.5
2	2.5	0	3	1
1.5	0	1	2	0

The usefulness of this table is limited. The first thing we do is therefore to sort the numbers.This is done in table 2.2.

Table 2.2: Hours of TV, sorted ascendingly.

Hours of TV (sorted)				
0	0	0	0	0.5
0.5	1	1	1	1
1	1	1	1.5	1.5
2	2	2	2	2
2.5	2.5	3	3	3.5

As we see in table 2.2, some of the numbers occur several times. It is, therefore, a good idea to write down a table of the different numbers and their *frequencies* (i.e. how *frequently* they occur). This table still consists of *ungrouped* data, since we do not group different observations, but only list how many times, the different numbers occur. As well as counting the frequencies, we calculate a few other numbers for each observation. The table might look like this:.

Observation, x	Frequency, n	Relative freq., f	Cum. rel. fr., F
0	4	16%	16%
0.5	2	8%	24%
1	7	28%	52%
1.5	2	8%	60%
2	5	20%	80%
2.5	2	8%	88%
3	2	8%	96%
3.5	1	4%	100%
Total	25	100%	

The meaning of the different columns is explained in the following definition:

Definition 2.1

For a data set with N observations, which is made up of M *different* observations, $x_1, x_2, \ldots, x_M$, we define:

1. The *frequency* n_i is the number of times x_i occurs in the data set.

   Note that $n_1 + n_2 + \cdots + n_M = N$.
2. The *relative frequency* f_i is the frequency divided by the total number of observations, i.e. $f_i = \frac{n_i}{N}$.
3. The *cumulative relative frequency* F_i is the sum of relative frequencies up to and including the relative frequency of the observation x_i, i.e.
$$F_i = f_1 + f_2 + \cdots + f_i \,.$$

The relative frequency shows the percentage of students, which have watched TV for 0 hours, 0.5 hours, etc. This is useful if we want to compare two classes with a different number of students. We usually write the relative frequency as a percentage, but this is not necessary—we might just as well write the relative frequencies as numbers between 0 and 1, i.e. 0.16 instead of 16%.

The cumulative relative frequency shows how many students watched TV for e.g. 1 hour *or less*. The cumulative relative frequency F_3 for the observation x_3 (1 hour) is 52%. This means that 52% of the students have watched TV for 1 hour or less. We find the number by adding the relative frequencies for x_1, x_2 and x_3:

$$F_3 = f_1 + f_2 + f_3 = 16\% + 8\% + 28\% = 52\% \,.$$

2.2 MEAN AND STANDARD DEVIATION

Eventhough a table greatly increases our ability to compare different data sets, it is sometimes easier if we can describe the data sets by a few numbers, so-called *descriptors*.

A descriptor could be e.g. the mean, which tells us what the average observation is. We find the mean by adding all the observations and dividing by the total number of observations. For the numbers in table 2.2, the mean is

$$\mu = \frac{0+0+0+0+0.5+\cdots+3+3+3.5}{25} = 1.42\,.$$

Since we already counted the frequencies for the different observations (e.g. in the table above we see that the observation "0" occurs 4 times), we can also use the frequencies, and the calculation becomes

$$\mu = \frac{0\cdot 4+0.5\cdot 2+1\cdot 7+\cdots+3.5\cdot 1}{25} = 1.42\,.$$

The result is, of course, still the same.

Since we obtain the relative frequencies by dividing the frequencies by the total number of observations, we could have just divided all of the frequencies by 25 to begin with, and get[1]

$$\mu = 0\cdot 0.16+0.5\cdot 0.08+1\cdot 0.28+\cdots+3\cdot 0.08+3.5\cdot 0.04 = 1.42\,.$$

[1] Notice that the relative frequencies are written as decimals, instead of percentages. E.g. the relative frequency of the first observation is not 16, but 16%, which is the same as 0.16.

The last calculation is the one we use in our definition:

Definition 2.2

For a data set of M different observations $x_1, x_2, \ldots, x_M$ with corresponding relative frequencies $f_1, f_2, \ldots, f_M$, we define the mean μ and the standard deviation σ as

1. $\mu = x_1\cdot f_1 + x_2\cdot f_2 + \cdots + x_M\cdot f_M$.
2. $\sigma = \sqrt{(x_1-\mu)^2\cdot f_1 + (x_2-\mu)^2\cdot f_2 + \cdots + (x_M-\mu)^2\cdot f_M}$.

The mean tells us what the average observation is. So, when $\mu = 1.42$ for the data set above, it means that each of the 25 students on average watched 1.42 hours of TV.

The standard deviation is a little more complicated, but it describes how far the observations on average are from the mean. If every student had watched TV for the same amount of time, the standard deviation would be $\sigma = 0$. So, in a sense, the standard deviation measures how "spread out" the data are.

For the data set in question, the standard deviation is

$$\begin{aligned}\sigma &= \sqrt{(0-1.42)^2\cdot 0.16+(0.5-1.42)^2\cdot 0.08+\cdots+(3.5-1.42)^2\cdot 0.04}\\ &= 0.987\,.\end{aligned}$$

Most CAS have built-in tools to calculate the mean and the standard deviation from raw/ungrouped data.

2.3 QUARTILES

The mean of any data set is highly sensitive to extreme values. If one student had watched TV for 20 hours, the mean would have been a lot greater. Therefore, it sometimes makes sense to describe instead a data set using the *median*, which is the "middle" value of the data set.

If we list all of the 25 numbers from table 2.2, the median is the number in the middle, i.e. the 13th number:[2]

[2] If we have an even number of observations, the median is the average of the two observations in the middle.

median

0, 0, 0, 0, 0.5, 0.5, 1, 1, 1, 1, 1, 1, **1**, 1.5, 1.5, 2, 2, 2, 2, 2, 2.5, 2.5, 3, 3, 3.5

So, the median is 1. This means that half of the students watched TV for 1 hour or less. The other half watched TV for 1 hour or more. It is important to remember that this has nothing to do with the mean, and, as we can see, the two numbers are indeed different.

Sometimes we want more information than what we get from just the median. We find the median by dividing the data set into two halves. So, we might get more information by dividing the data set into four quarters. Then we find the so-called *quartiles*:

lower quartile median upper quartile

0, 0, 0, 0, 0.5, 0.5, 1, 1, 1, 1, 1, 1, **1**, 1.5, 1.5, 2, 2, 2, 2, 2, 2.5, 2.5, 3, 3, 3.5

The *lower quartile* is the median of the lower half of the data. Since, in this case, the lower half has an even number of observations (12), the lower quartile is the average of the two values in the middle (the 6th and the 7th). Thus the lower quartile is

$$Q_1 = \frac{0.5+1}{2} = 0.75\,.$$

The median is the same as before, i.e. the median is

$$Q_2 = 1\,.$$

The *upper quartile* is the median of the upper half of the data. Here, we must again take the average of two values, i.e.

$$Q_3 = \frac{2+2}{2} = 2\,.$$

So, the quartiles are the three numbers Q_1, Q_2 og Q_3.

The quartiles of the hours of TV watched by the students are $(0.75, 1, 2)$.

Instead of "lower quartile, median, and upper quartile" we sometimes call them "1st, 2nd and 3rd quartile".

Definition 2.3

For an ungrouped data set, we define the following quantities:

- The *median* (or *2nd quartile*) Q_2, which is the middle value of the observations. If there is an even number of observations, the median is the average of the two middle values.
- The *lower* (or *1st*) *quartile* Q_1, which is the median of the lower half of the observations.
- The *upper* (or *3rd*) *quartile* Q_3, which is the median of the upper half of the observations.

The *quartiles* is the ordered set of numbers (Q_1, Q_2, Q_3) of all the quartiles.

When, in our case, the lower quartile is 0.75, it shows us that a quarter (25%) of the students watched TV for 0.75 hours or less, while three quarters (75%) watched TV for 0.75 hours or more.

The upper quartile, $Q_3 = 2$, shows us that three quarters of the students watched TV for 2 hours or less, while a quarter watched TV for 2 hours or more.

So, the quartiles provide a useful, short description of our data set.

2.4 DIAGRAMS

In this section, we show 3 different types of diagrams, which can be used to describe a data set:

- A *bar chart*, which is useful if we want to illustrate a single data set.
- A *cumulative relative frequency graph*.
- A *box plot*, which is useful when we want to compare different data sets.

Bar Chart

Our investigation of the TV habits of students yielded the following table:

Observation, x	Frequency, n	Relative freq., f	Cum. rel. fr., F
0	4	16%	16%
0.5	2	8%	24%
1	7	28%	52%
1.5	2	8%	60%
2	5	20%	80%
2.5	2	8%	88%
3	2	8%	96%
3.5	1	4%	100%
Total	25	100%	

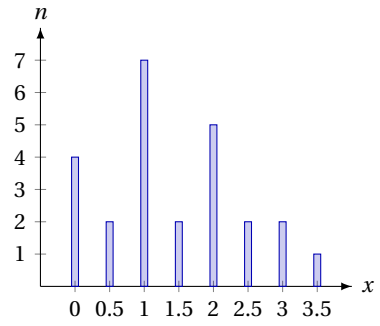


Figure 2.1: The TV hours of the students as a bar chart. The y-axis represents the frequency.

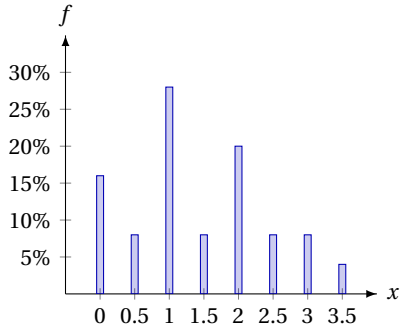


Figure 2.2: The TV hours of the students as a bar chart. The y-axis represents the relative frequency.

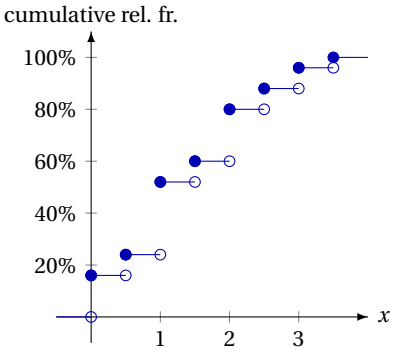


Figure 2.3: Cumulative relative frequency graph of the TV hours of the students.

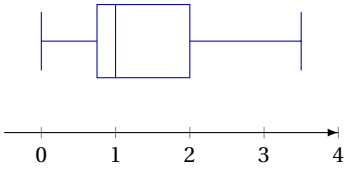


Figure 2.4: Box plot of the TV hours.

This table enables us to draw a bar chart. The x-axis represents the individual observations, and at each observation we draw a bar, whose height equals the frequency of the observation.

In figure 2.1, we see a bar chart where the height of the columns indicate the frequency. Figure 2.2 shows the same bar chart, but here the heights indicate the relative frequencies. The two charts are identical except for the numbers on the y-axis.

If we just want a quick description of the data set, we might just as well use the frequencies. But if we want to compare two data sets, it is easier when we use the relative frequencies—especially if the two data sets contain a different number of observations. This might be the case if we were comparing two school classes with a different number of students.

Cumulative Relative Frequency Graph

A cumulative relative frequency graph is a graph of the cumulative relative frequencies. We plot the cumulative relative frequency at the corresponding observation, and the we move horizontally until we get to the next observation, where we jump to the next cumulative relative frequency. In this way, we get a graph that looks a bit like a set of steps—a function, which has a such a graph is called a *step function*.

It is possible to use cumulative relative frequency graphs to compare different data sets. But the box plot, which we describe below, is a much easier tool to use for comparisons.

Box Plot

A box plot is a diagram drawn using only the quartiles. When we do this, we discard a lot of information. But in return, we get a diagram which shows us how the numbers are distributed in way that is easy to read.

A box plot of our data set can be seen in figure 2.4. We draw vertical lines at the minimum value (0), the lower quartile (0.75), the median (1), the upper quartile, and at the maximum value (3.5). Then we connect the vertical lines as shown in the figure.

The box contains the middle half of the observations, while the horizontal lines at both ends show the hours of TV watched for the lower and the upper quarter of the class.

When we draw box plots of different distributions, they are easy to compare. If we measured the minimum and maximum values, and the quartiles for two different classes, we might get something like this table (class A is the one we have looked at all along):

Class	Minimum value	Q_1	Q_2	Q_3	Maximum value
A	0	0.75	1	2	3.5
B	0	1	1.5	1.75	4

If we just look at the numbers, it is hard to tell what the difference is between the two classes. If, however, we draw a box plot for both them (see figure 2.5), they are quite easy to compare.

Here, we see that even though class B has at least one student who watched more TV than any student in class A, the lower 75% of class B have watched a little less TV than the lower 75% of class A. The middle half of class B is also closer than the middle half of A, which means that the number of TV hours is not as spread out for B as it is for A.

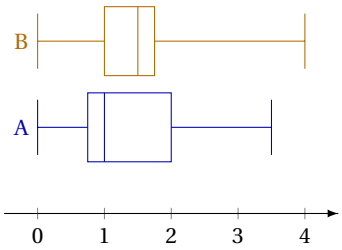


Figure 2.5: Using box plots to compare hours of TV watched.

2.5 STATISTICS WITH GROUPED DATA

We talk about grouped data, when the data is grouped in intervals. It is useful to group the data if we have a large data set with many different observations.

In table 2.3, we see a listing of Danish gross incomes. Here we have so many observations that it makes no sense to list all the observed incomes behind the table. We have therefore grouped the different incomes in intervals.

Table 2.3: Annual gross incomes in Denmark.[1]

Interval	Frequency
0–100 000	628 423
100 000–200 000	1 074 944
200 000–300 000	1 054 700
300 000–400 000	889 807
400 000–500 000	469 896
500 000–750 000	334 062

Looking at the table, it is impossible to see, whether an income of exactly DKK 100 000 should be included in the first or the second interval. It is therefore a good idea to use mathematical notation for the intervals to describe if such incomes are in one interval or another.

The frequencies in the table turn out to be quite large numbers, and so it makes sense to calculate the relative frequencies instead. A table with that information looks like this:

Interval	Frequency, n	Relative fr., f	Cum. rel. fr., F
[0; 100 000[	628 423	14.1%	14.1%
[100 000; 200 000[	1 074 944	24.1%	38.3%
[200 000; 300 000[	1 054 700	23.7%	62.0%
[300 000; 400 000[	889 807	20.0%	81.9%
[400 000; 500 000[	469 896	10.6%	92.5%
[500 000; 750 000[	334 062	7.5%	100.0%
Total	4 451 832	100,0%	

2.6 MEAN AND STANDARD DEVIATION

We cannot calculate the mean and the standard deviation like we did for ungrouped data. This is impossible because we do not know how the incomes are distributed in the different intervals, since we do not have the raw data from which the table is made.

Instead, we assume that the incomes are evenly distributed in the intevals. This allows us to use the midpoints as though *they* were the observations.

Definition 2.4

For a data set, which is grouped into the intervals $[a_1; b_1[$, $[a_2; b_2[$, …, $[a_N; b_N[$, we define the mean μ and the standard deviation σ:

1. $\mu = m_1 \cdot f_1 + m_2 \cdot f_2 + \cdots + m_N \cdot f_N$.
2. $\sigma = \sqrt{(m_1 - \mu)^2 \cdot f_1 + (m_2 - \mu)^2 \cdot f_2 + \cdots + (m_N - \mu)^2 \cdot f_N}$.

f_i is the relative frequency of the interval, and m_i is the midpoint of the interval, $m_i = \frac{a_i + b_i}{2}$.

To calculate the mean and the standard deviation of the data set above, we add a new column of interval midpoints:

Interval	Interval midpoint, m	Relative frequency, f
[0; 100 000[	50 000	14.1%
[100 000; 200 000[	150 000	24.1%
[200 000; 300 000[	250 000	23.7%
[300 000; 400 000[	350 000	20.0%
[400 000; 500 000[	450 000	10.6%
[500 000; 750 000[	625 000	7.5%

The mean is then

$$\mu = 50\,000 \cdot 0.141 + 150\,000 \cdot 0.241 + \cdots + 625\,000 \cdot 0.075 = 266\,859\,.$$

This means the average income in the table is DKK 266 859.

The standard deviation is

$$\begin{aligned}\sigma &= \sqrt{(50\,000 - 266\,859)^2 \cdot 0.141 + \cdots + (625\,000 - 266\,859)^2 \cdot 0.075} \\ &= 156\,684\,.\end{aligned}$$

So, the standard deviation is DKK 156 684.

2.7 DIAGRAMS

In this section, we describe three ways of illustrating grouped data:

- Histograms, which correspond to bar charts of ungrouped data.
- Cumulative relative frequency graphs, or *ogives* which can be used to determine the quartiles.
- Box plots, which is exactly the same type of diagram as a box plot of ungrouped data.

Histograms

In a histogram, the relative frequencies of the intervals are drawn as columns. For ungrouped data, we could draw a bar chart—and the height

of the bars corresponded to the relative frequencies. Here, we cannot do that, since then wider intervals would carry more weight than narrow intervals.

Instead, we let the frequency determine the *area* of the corresponding column, see figure 2.6.

When the relative frequency is given by the area, we need to show which area corresponds to a certain percentage. This is illustrated in the figure, where the rectangle in the upper right hand corner shows, which area corresponds to 10%.

Since the area shows the relative frequency, we have no use for a y-axis, so this is usually omitted.

When we draw histograms, it is important to remember that

> in a histogram, the relative frequency is the *area* of the corresponding column.

If, however, all the intervals are of equal width, we *can* let the height correspond to the relative frequency. Many CAS work this way. But it is important to remember that the intervals then *have to be of equal width*.

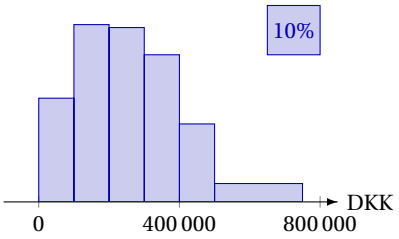


Figure 2.6: Histogram of the income distribution.

Ogives

An ogive is a plot of the cumulative relative frequencies. The graph illustrates how many percent of the data set is below a certain value. Since the graph shows how many percent is *below* the value, the cumulative relative frequencies are plotted as a function of the end point of the intervals.

We therefore add a column of interval end points to the table above:

Interval	Interval end point	Cumulative rel. freq.
[0; 100 000[	100 000	14.1%
[100 000; 200 000[	200 000	38.3%
[200 000; 300 000[	300 000	62.0%
[300 000; 400 000[	400 000	81.9%
[400 000; 500 000[	500 000	92.5%
[500 000; 750 000[	750 000	100.0%

We then draw the ogive by plotting the cumulative relative frequencies as a function of the end points of the intervals.

Since the ogive shows us, how many percent of the data is below a certain value, we can use it to investigate e.g. how many percent have an income below DKK 250 000, or what the maximum income is for the 80% lowest paid. This last number is called the 80th percentile; we have the following definition:

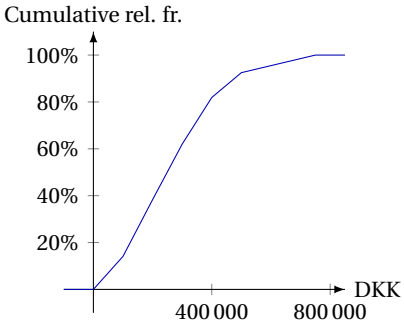
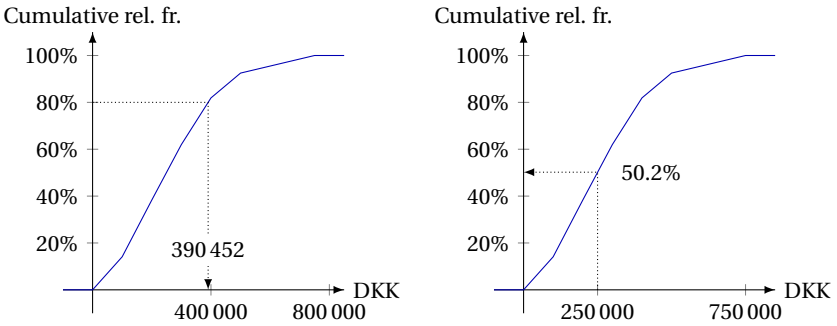


Figure 2.7: Ogive of the income distribution.



(a) The 80th percentile.

(b) Incomes below DKK 250 000?

Figure 2.8: In the figure to the left, we find the 80th percentile. This numbers shows us that 80% have an income below DKK 390 452.
To the right, we find the number corresponding to 250 000 on the x-axis. This number tells us that 50.2% have an income below DKK 250 000.

Definition 2.5

For a statistical data set, the pth percentile is the observation that has a cumulative relative frequency of p%.

In figure 2.8, we see how to find the 80th percentile. We start at 80% on the y-axis and find the corresponding value on the x-axis. The number 390 452 shows us that 80% of the people in the statistic have an income below DKK 390 452. Then we also know that 20% have an income above DKK 390 452.

The figure also shows us, which percentile corresponds to an income of DKK 250 000. Here, we start at 250 000 on the x-axis and find the corresponding number on the y-axis. This number is 50.2%, which means that 50.2% have an income below DKK 250 000. So, 49.8% have an annual income above DKK 250 000.

Using the ogive we can find the quartiles. We have this definition:

Definition 2.6

Using the ogive of a statistical data set, we find the quartiles (Q_1, Q_2, Q_3):

1. The *lower quartile*, Q_1 is the 25th percentile.
2. The *median*, Q_2 is the 50th percentile.
3. The *upper quartile*, Q_3 is the 75th percentile.

In figure 2.9, we see how to find the quartiles. We start at 25%, 50%, and 75% on the y-axis and then find the corresponding values on the x-axis. We find the quartiles

$$(145\,041, 249\,367, 365\,327)\,.$$

These numbers show that

- 25% have an income below DKK 145 041.
- 50% have an income below DKK 249 367.
- 75% have an income below DKK 365 327.

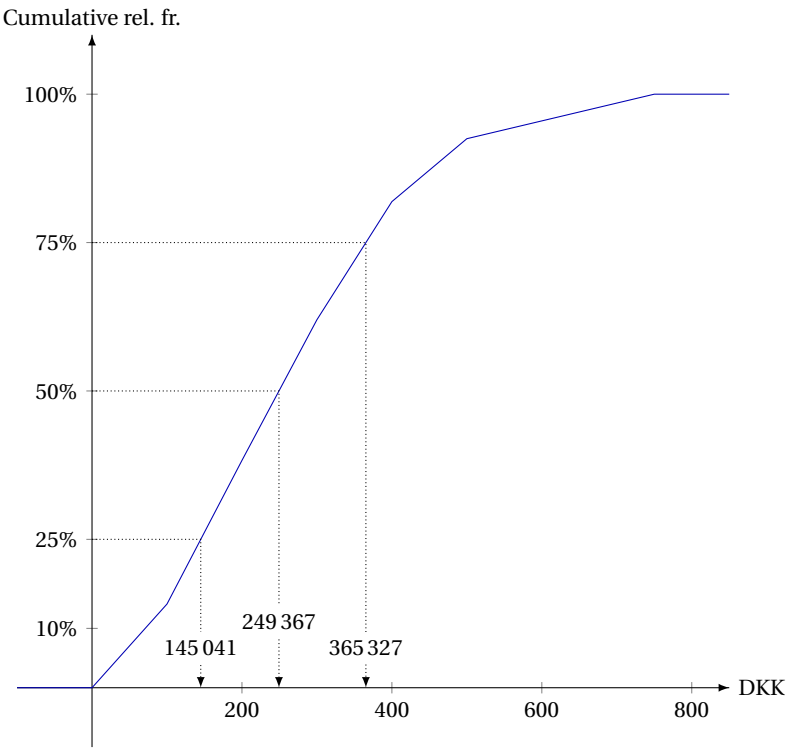


Figure 2.9: How to find the quartiles on the ogive of the income distribution.

Box Plots

Drawing a box plot of a grouped data set is no different from drawing a box plot of an ungrouped data set. The only difference between the two is how we find the quartiles. After they are found, we do exactly the same.

For the income distribution we looked at above, the quartiles were

$$(145\,041, 249\,367, 365\,327)\,.$$

The minimum value was 0, and the maximum value was 750 000.

A box plot of this distribution will, therefore, look like figure 2.10.

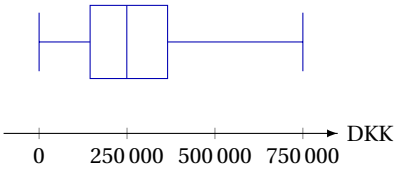


Figure 2.10: Box plot of the income distribution.

Integral Calculus 3

Integral calculus is, in essence, the opposite of differential calculus. In chapter 1, we looked at derivatives, which describe the slope of tangents to the graph. If we do this calculation "the other way around", we find the so-called *primitive function* or *antiderivative*.

3.1 PRIMITIVE FUNCTIONS

We have the following definition:

Definition 3.1

Let f be a function. A differentiable function F, where

$$F'(x) = f(x)\,,$$

is called a *primitive function* of f.

A primitive function[1] F of a function f is a function that has f as its derivative. Examining whether a given function is a primitive function of another can therefore be done by differentiation.

[1] Primitive functions are often denoted by capital letters, such that a primitive function of $f(x)$ is called $F(x)$, and a primitive function of $h(x)$ is called $H(x)$. In principle, we can denote primitive functions by whichever letter we want, but it is easier to spot their origin, if we use this convention.

Example 3.2
Is $F(x) = x^3 + 2x - 5$ a primitive function of $f(x) = 3x^2 + 2$?

We can find out by differentiating F:

$$F'(x) = 3x^2 + 2 \cdot 1 - 0 = 3x^2 + 2\,.$$

When we differentiate F, we get the formula for f, i.e. $F'(x) = f(x)$ and F *is* a primitive function of f.

Example 3.3
Is $H(x) = 4x + \ln(x)$ a primitive function of $g(x) = 2x + \frac{1}{x}$?

When we differentiate $H(x)$, we get

$$H'(x) = 4 \cdot 1 + \frac{1}{x} = 4 + \frac{1}{x}\,.$$

This is *not* the same as $h(x)$, i.e. H is *not* a primitive function of $h(x)$.

Example 3.4
$F_1(x) = x^2 + e^x + 4$ and $F_2(x) = x^2 + e^x - 17$ are both primitive functions of $f(x) = 2x + e^x$.

This is because, when we differentiate F_1 and F_2, we get

$$F_1'(x) = 2x + e^x + 0 = 2x + e^x$$
$$F_2'(x) = 2x + e^x - 0 = 2x + e^x\,.$$

So, $F_1'(x) = F_2'(x) = f(x)$ and both functions are primitive functions of $f(x)$.

From example 3.4, we see that it is possible for a function to have more than one primitive function. The two primitive functions in the example are, however, not *that* different. The only difference is a constant. Actually, the reason a function has several primitive functions is that when we differentiate a constant, we always get 0.

This means we can always find a new primitive function by adding a constant to another primitive function, since added constants disappear when we differentiate.

Theorem 3.5

If $F_1(x)$ and $F_2(x)$ are both primitive functions of f, then

$$F_1(x) - F_2(x) = C\,,$$

where C is a constant.

Proof
Since $F_1(x)$ and $F_2(x)$ are both primitive functions of f, we have[2]

$$\big(F_1(x) - F_2(x)\big)' = F_1'(x) - F_2'(x) = f(x) - f(x) = 0\,.$$

If we differentiate the difference $F_1(x) - F_2(x)$ we get 0.

It is only possible to get 0, when we differentiate a constant, and therefore

$$F_1(x) - F_2(x) = C\,,$$

where C is a constant. ■

[2] In this calculation, we use that since F_1 and F_2 are both primitive functions of f, we must have $F_1'(x) = f(x)$ and $F_2'(x) = f(x)$.

So, theorem 3.5 states that although any given function has an infinite amount of primitive functions, we can find them all by adding constants to another primitive function.

Example 3.6
$F(x) = x^2 + \ln(x)$ is a primitive function of $f(x) = 2x + \frac{1}{x}$, because

$$F'(x) = 2x + \frac{1}{x} = f(x)\,.$$

But then

$$F_1(x) = x^2 + \ln(x) + 3$$
$$F_2(x) = x^2 + \ln(x) - 14$$
$$F_3(x) = x^2 + \ln(x) + 365\,749$$

are also primitive functions of $f(x)$.

3.2 INDEFINITE INTEGRALS

Calculating a primitive function of a function $f(x)$ is called *integrating* $f(x)$. We have the following definition:[3]

[3]The notation $\int \cdot \,\mathrm{d}x$ means that we integrate everything between $\int$ and $\mathrm{d}x$.

So, the symbol $\mathrm{d}x$ is not a mathematical quantity. It merely shows us, where the integration ends, and that our independent variable is called x.

Definition 3.7

Let f be a function. *The indefinite integral of* $f(x)$ is the set of all primitive functions of $f(x)$. It is denoted by

$$\int f(x)\,\mathrm{d}x\,.$$

The function $f(x)$ is called the *integrand*.

When we calculate the definite integral, we show that $\int f(x)\,\mathrm{d}x$ is the set of all primitive functions by writing an added constant.

Example 3.8
Here, we determine $\int (2x+3)\,\mathrm{d}x$.

$$\int (2x+3)\,\mathrm{d}x = x^2+3x+C\,.$$

x^2+3x is a primitive function of $2x+3$, and the constant C shows that we have found *every* primitive function.

The constant C in example 3.8 is called the *constant of integration*.

Table 3.1 shows the indefinite integrals of a few simple functions. If we want to be certain that these are indeed the indefinite integrals, we can differentiate the right hand column to see if this yields the left hand column.

Table 3.1: Indefinite integrals of some simple functions.

$f(x)$	$\int f(x)\,\mathrm{d}x$
a	$ax+C$
x	$\frac{1}{2}x^2+C$
x^2	$\frac{1}{3}x^3+C$
x^n	$\frac{1}{n+1}x^{n+1}+C$
$\frac{1}{x}$	$\ln(x)+C$
e^x	e^x+C
e^{ax}	$\frac{1}{a}\mathrm{e}^{ax}+C$

3.3 CALCULATION RULES

Just as there are rules for differentiation, we have rules for integration.

Theorem 3.9

Let f be a function, and c be an arbitrary constant. Then

$$\int c\cdot f(x)\,\mathrm{d}x = c\cdot\int f(x)\,\mathrm{d}x\,.$$

Proof
If we differentiate the right hand side of the equation in the theorem, we get

$$\left(c\cdot\int f(x)\,\mathrm{d}x\right)' = c\cdot\left(\int f(x)\,\mathrm{d}x\right)' = c\cdot f(x)\,.$$

The first equality follows from a differentiation rule, theorem 1.11. The second follows from the fact that $\int f(x)\,\mathrm{d}x$ is the set of primitive functions of f.

We have now shown that $c \cdot \int f(x)\,\mathrm{d}x$ is the set of primitive functions of $c \cdot f(x)$, but this means

$$\int c \cdot f(x)\,\mathrm{d}x = c \cdot \int f(x)\,\mathrm{d}x\,,$$

and this proves the theorem. ■

Theorem 3.9 can be used to find integrals of functions that are not listed in table 3.1.

Example 3.10
What is $\int 6x^2\,\mathrm{d}x$?

$6x^2$ is not listed in table 3.1, but x^2 is. We can now use theorem 3.9 to get

$$\int 6x^2\,\mathrm{d}x = 6 \cdot \int x^2\,\mathrm{d}x = 6 \cdot \tfrac{1}{3}x^3 + C = 2x^3 + C\,.$$

The next important rules are:

Theorem 3.11

Let f and g be functions. Then

1. $\int \big(f(x) + g(x)\big)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x$, and
2. $\int \big(f(x) - g(x)\big)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x - \int g(x)\,\mathrm{d}x\,.$

Proof
We only prove the first part of the theorem. The second part may be proven analogously.

Since

$$\begin{aligned}\left(\int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x\right)' &= \left(\int f(x)\,\mathrm{d}x\right)' + \left(\int g(x)\,\mathrm{d}x\right)' \\ &= f(x) + g(x)\,,\end{aligned}$$

we know that $\int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x$ is a primitive function of $f(x) + g(x)$, i.e.

$$\int \big(f(x) + g(x)\big)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x\,,$$

which proves the theorem. ■

Example 3.12
What is $\int (\mathrm{e}^x + x)\,\mathrm{d}x$?

$\mathrm{e}^x + x$ is not listed in table 3.1, but e^x and x are. We can therefore use theorem 3.11, and get

$$\int (\mathrm{e}^x + x)\,\mathrm{d}x = \int \mathrm{e}^x\,\mathrm{d}x + \int x\,\mathrm{d}x = \mathrm{e}^x + \tfrac{1}{2}x^2 + C\,.$$

We can also use theorems 3.9 and 3.11 in the same calculation, as in this example:

Example 3.13
We calculate the indefinite integral $\int (9x^2+4x-3)\,\mathrm{d}x$ like this:

$$\begin{aligned}\int (9x^2+4x-3)\,\mathrm{d}x &= 9\cdot\int x^2\,\mathrm{d}x + 4\cdot\int x\,\mathrm{d}x - \int 3\,\mathrm{d}x\\ &= 9\cdot\tfrac{1}{3}x^3 + 4\cdot\frac{1}{2}x^2 - 3\cdot x + C\\ &= 3x^3+2x^2-3x+C\,.\end{aligned}$$

Here, we use both of the theorems 3.9 and 3.11 as well as table 3.1.

3.4 FINDING PRIMITIVE FUNCTIONS

In the preceding section, we saw how to find the indefinite integral of a function. The indefinite integral is the set of all primitive functions. There are infinitely many primitive functions, owing to the fact that the derivative of a constant is 0.

If we are looking for a certain primitive function, we therefore need more information than a formula for the integrand.

This information might be

1. a point, which the graph of the primitive function passes through, or
2. the equation of a tangent to the graph of the primitive function.

A Point on the Graph of the Primitive Function

Since the primitive functions of any given function are equal up to a constant, the graphs of the primitive functions can be found by shifting a graph of one primitive function vertically.

So, if we know a point, which the graph of our primitive function passes through, we can find the value of the constant of integration, C. And then, we have the formula for the primitive function we are looking for.

Example 3.14
Here, we find the primitive function $F(x)$ of $f(x) = x^3+2x-1$, whose graph passes through $P(2,10)$.

First, we set $F(x)$ equal to the indefinite integral of $f(x)$:

$$F(x) = \int (x^3+2x-1)\,\mathrm{d}x = \tfrac{1}{4}x^4+x^2-x+C\,.$$

C now has a fixed value, i.e. the value for which the graph of $F(x)$ passes through $P(2,10)$.

In figure 3.1, we see a few of the graphs of the primitive functions of f. The graph that passes through $P(2,10)$ is the graph of the primitive function, we are looking for.

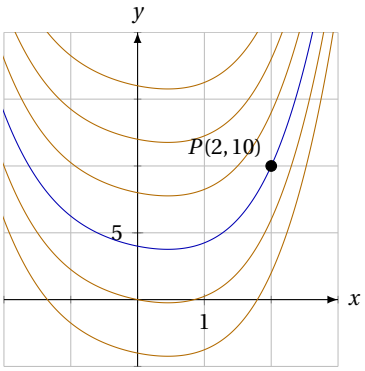


Figure 3.1: Graphs of some of the primitive functions of $f(x) = x^3+2x-1$.

We know that the primitive function has the formula

$$F(x) = \tfrac{1}{4}x^4+x^2-x+C\,.$$

We also know that its graph passes through the point $P(2,10)$. If this is true, then $F(2) = 10$, which gives us the equation

$$F(2) = \tfrac{1}{4} \cdot 2^4 + 2^2 - 2 + C = 10\,.$$

Solving this equation, we get

$$\tfrac{1}{4} \cdot 2^4 + 2^2 - 2 + C = 10 \qquad \Leftrightarrow \qquad C = 4\,.$$

We can now write a formula for F:

$$F(x) = \tfrac{1}{4}x^4 + x^2 - x + 4\,.$$

The primitive function of $f(x)$, whose graph passes through $P(2,10)$, has this formula.

Example 3.15
Which primitive function of $g(x) = \mathrm{e}^x - 3x$ has a graph passing through the point $Q(0,-7)$?

The primitive function has the formula

$$G(x) = \int (\mathrm{e}^x - 3x)\,\mathrm{d}x = \mathrm{e}^x - \tfrac{3}{2}x^2 + C\,.$$

Since the graph of G passes through $Q(0,-7)$, we have

$$G(0) = \mathrm{e}^0 - \tfrac{3}{2} \cdot 0^2 + C = -7\,.$$

We solve this equation

$$\mathrm{e}^0 - \tfrac{3}{2} \cdot 0^2 + C = -7 \qquad \Leftrightarrow \qquad 1 - 0 + C = -7 \qquad \Leftrightarrow \qquad C = -8\,.$$

Our primitive function therefore has the formula

$$G(x) = \mathrm{e}^x - \tfrac{3}{2}x^2 - 8\,.$$

Primitive Function, Whose Graph has a Certain Tangent

We can also determine a formula for a certain primitive function if we know the equation of a tangent to its graph.

Example 3.16
In this example, we determine the primitive function of $f(x) = \frac{4}{x}$, whose graph has tangent with the equation $y = 4x + 1$.

We call the primitive function $F(x)$. Its formula is

$$F(x) = \int \frac{4}{x}\,\mathrm{d}x = 4 \cdot \ln(x) + C\,.$$

In figure 3.2, we see the graphs of some of the primitive functions of $f(x)$ and the line $y = 4x + 1$. One of the graphs has the line as a tangent.

If $y = 4x + 1$ is tangent to the graph of the primitive function, we know that the graph has slope 4 at some point. The slope of the tangents of $F(x)$ are given by $F'(x)$, but since F is a primitive function of $f(x)$, $F'(x) = f(x)$.

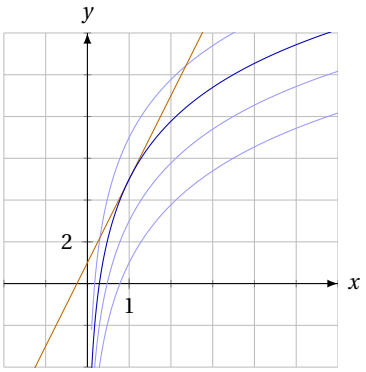


Figure 3.2: Graphs of some of the primitive functions of $f(x) = \frac{4}{x}$ and the line $y = 4x + 1$.

Determining where F has slope 4 is then the same as determining the value of x for which $f(x) = 4$. We therefore solve this equation:

$$f(x) = 4 \qquad \Leftrightarrow \qquad \frac{4}{x} = 4 \qquad \Leftrightarrow \qquad x = 1\,.$$

Now we know that the point of tangency has x-coordinate $x = 1$. To find the y-coordinate of the point, we look again at the equation of the tangent. The graph and the tangent both pass through the point of tangency. We cannot find the point of tangency using the formula for $F(x)$, because we do not yet know C, but we can use the equation of the tangent.

Inserting $x = 1$ into the equation $y = 4x + 1$ yields

$$y = 4 \cdot 1 + 1 = 5\,.$$

So, the graph of $F(x)$ and the line both pass through the point $(1,5)$. We can now use this to determine C, because

$$F(1) = 4 \cdot \ln(1) + C = 5 \qquad \Leftrightarrow \qquad 4 \cdot 0 + C = 5 \qquad \Leftrightarrow \qquad C = 5\,.$$

Now that we know C, we know the entire formula for F:

$$F(x) = 4 \cdot \ln(x) + 5\,.$$

When we know a point that the graph of a primitive function passes through, only one primitive functions fits this piece of information. But if we know the equation of a tangent, the graphs of several primitive functions may have this tangent.

Example 3.17

In this example, we find the primitive function of $f(x) = -x^3 + 3x$, whose graph has the line $y = -2x + 8$ as a tangent.

The formula for the primitive function is

$$F(x) = \int (-x^3 + 3x)\,\mathrm{d}x = -\tfrac{1}{4}x^4 + \tfrac{3}{2}x^2 + C\,.$$

We now look for the value of x where the graph has slope -2 (since this is the slope of the tangent):

$$f(x) = -x^3 + 3x = -2\,.$$

Solving this equation yields *two* solutions:

$$x = -1 \qquad \vee \qquad x = 2\,.$$

So apparently, two different primitive functions have the line $y = -2x + 8$ as a tangent to their graphs.

We therefore need to determine *two* points of tangency. The first point has x-coordinate $x = -1$ and y-coordinate

$$y = -2 \cdot (-1) + 8 = 10\,,$$

and the second point has x-coordinate $x = 2$ and y-coordinate

$$y = -2 \cdot 2 + 8 = 4\,.$$

The graph of the first primitive function passes through the point $(-1, 10)$, and here we find C by solving the equation

$$F(-1) = -\tfrac{1}{4} \cdot (-1)^4 + \tfrac{3}{2} \cdot (-1)^2 + C = 10 \qquad \Leftrightarrow \qquad C = \tfrac{35}{4}\,.$$

The graph of the second primitive function passes through $(2, 4)$, so here we solve the equation

$$F(2) = \tfrac{1}{4} \cdot 2^4 + \tfrac{3}{2} \cdot 2^2 + C = 4 \qquad \Leftrightarrow \qquad C = 2\,.$$

So, the function $f(x) = -x^3 + 3x$ has two primitive functions, whose graphs have the line $y = -2x + 8$ as a tangent:

$$F_1(x) = -\tfrac{1}{4}x^4 + \tfrac{3}{2}x^2 + \tfrac{35}{4}$$
$$F_2(x) = -\tfrac{1}{4}x^4 + \tfrac{3}{2}x^2 + 2\,.$$

In figure 3.3, we see the graphs of the two functions as well as the tangent.

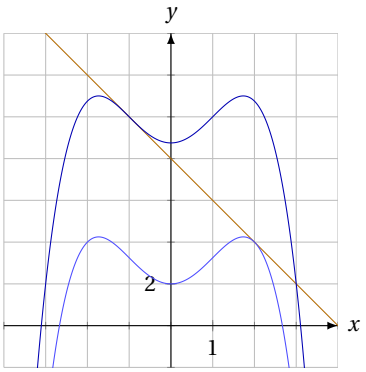


Figure 3.3: The two primitive functions of $f(x) = -x^3 + 3x$, whose graphs have the line $y = -2x + 8$ as a tangent.

3.5 DEFINITE INTEGRALS

Until now, we have only looked at *indefinite* integrals. But if we have *in*definite integrals, there is of course also something called *definite* integrals.

If we have a function f and two numbers a and b, we define a number called the *definite integral of f in the interval $[a; b]$*. This number is calculated using a primitive function.[4]

[4] Which primitive function we use does not matter, therefore we usually choose the simplest, i.e. where the constant of integration is $C = 0$.

Definition 3.18

Let F be a primitive function of f, and let a and b be numbers. *The definite integral of f in the interval $[a; b]$* is the number

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

The two numbers a and b are called the *limits of integration.*

Note that the indefinite integral $\int f(x)\,\mathrm{d}x$ is a function,[5] whereas the definite integral $\int_a^b f(x)\,\mathrm{d}x$ is a number.

[5] Actually infinitely many functions, since the constant of integration C may have any value.

When we calculate the number $\int_a^b f(x)\,\mathrm{d}x$, we first find a primitive function $F(x)$ and then calculate $F(b) - F(a)$ by inserting the numbers a and b.

Example 3.19

If we want to calculate $\int_1^4 x^2\,\mathrm{d}x$, we first need to find a primitive function of x^2. This could be $F(x) = \frac{1}{3}x^3$. Then we calculate $F(4) - F(1)$, i.e. $\frac{1}{3} \cdot 4^3 - \frac{1}{3} \cdot 1^3$.

We can write the calcultion like this:

$$\int_1^4 x^2\,\mathrm{d}x = \left[\tfrac{1}{3}x^3\right]_1^4 = \tfrac{1}{3}\cdot 4^3 - \tfrac{1}{3}\cdot 1^3 = 21.$$

The indefinite integral has the value 21.

Notice that we wrote the primitive function in brackets (with the limits attached), before we inserted the numbers—this is done to make the calculation easier to read.

Corresponding to the theorems 3.9 and 3.11 we have the following:

Theorem 3.20

Let the function f and g, and the interval $[a;b]$ be given. Then

1. $\int_a^b c\cdot f(x)\,\mathrm{d}x = c\cdot\int_a^b f(x)\,\mathrm{d}x$,
2. $\int_a^b (f(x)+g(x))\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x$, and
3. $\int_a^b (f(x)-g(x))\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x - \int_a^b g(x)\,\mathrm{d}x$.

We omit the proof, since it follows immediately from the corresponding theorems for indefinite integrals.

For definite integrals we also have the following theorem:

Theorem 3.21

Let f be a function defined in an interval containing the numbers a, b and c. Then

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$$

Proof

Let F be a primitive function of f. From definition 3.18, we get

$$\begin{aligned}\int_a^b f(x)\,\mathrm{d}x &= F(b)-F(a)\\ &= F(b)-F(c)+F(c)-F(a)\\ &= \int_c^b f(x)\,\mathrm{d}x + \int_a^c f(x)\,\mathrm{d}x,\end{aligned}$$

which proves the theorem. ■

Theorem 3.21 merely states that we can split a definite integral into several integrals by dividing the interval $[a;b]$ into subintervals.

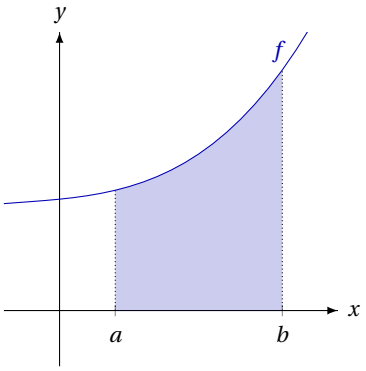


Figure 3.4: The marked area is $\int_a^b f(x)\,\mathrm{d}x$.

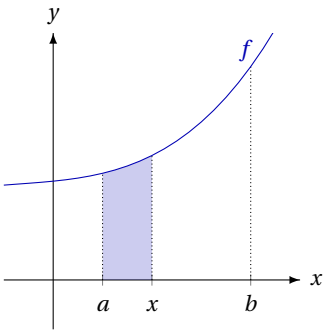


Figure 3.5: The value of $A(x)$ corresponds to the marked area.

3.6 AREAS BELOW GRAPHS

It turns out that there is a connection between the definite integral and the area below the graph of a function. For a function f, the area between the graph and the x-axis between the two values $x = a$ and $x = b$ is $\int_a^b f(x)\,\mathrm{d}x$, see figure 3.4. However, this is only true if the graph lies above the x-axis in the entire interval $[a;b]$.

We can write this statement as a theorem:

Theorem 3.22

Let f be a function defined in the interval $[a;b]$. If $f(x) \geq 0$ for all $x \in [a;b]$, then the area A of the region bounded by the graph of f and the x-axis in the interval $[a;b]$ is

$$A = \int_a^b f(x)\,\mathrm{d}x\,.$$

Proof

First, we assume that f is increasing in the entire interval $[a;b]$.

Then, we define a function $A(x)$ in the interval $[a;b]$. The function value of $A(x)$ is the area between the graph of f and the x-axis in the interval $[a;x]$, see figure 3.5.

We see from this definition that $A(a) = 0$, and $A(b)$ must be the entire area between the graph of f and the x-axis in the interval $[a;b]$. Therefore the entire area below the graph is

$$A(b) - A(a)\,.$$

If A is a primitive function of f, this is the same as the definite integral of f from a to b.[6] We must therefore show that A is a primitive function of f.

[6] This follows from the definition of the definite integral.

Since f is increasing, we have (see figure 3.6(a))

$$A(x+\Delta x) - A(x) \geq f(x)\cdot\Delta x\,,$$

and (see figure 3.6(b))

$$A(x+\Delta x) - A(x) \leq f(x+\Delta x)\cdot\Delta x\,.$$

We can write this collectively as a double inequality:

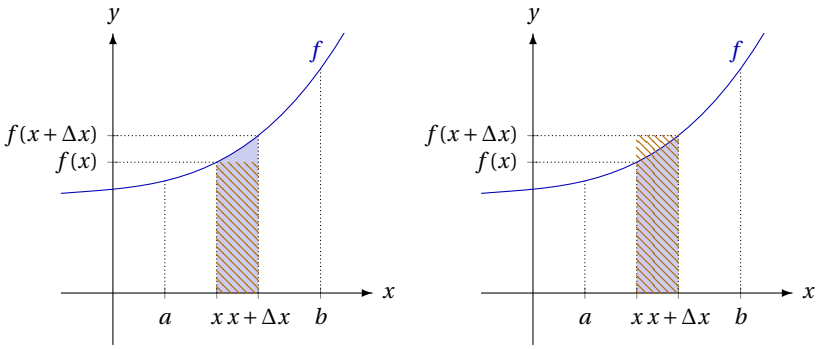
$$f(x)\cdot\Delta x \leq A(x+\Delta x) - A(x) \leq f(x+\Delta x)\cdot\Delta x\,. \tag{3.1}$$

If we divide by Δx everywhere, we get

$$f(x) \leq \frac{A(x+\Delta x) - A(x)}{\Delta x} \leq f(x+\Delta x)\,. \tag{3.2}$$

Now, we let $\Delta x \to 0$. Then

$$f(x) \to f(x),$$
$$f(x+\Delta x) \to f(x),$$



(a) $A(x+\Delta x)-A(x) \geq f(x)\cdot\Delta x$. (b) $A(x+\Delta x)-A(x) \leq f(x+\Delta x)\cdot\Delta x$.

Figure 3.6: The marked area between the graph and the x-axis is $A(x+\Delta x)-A(x)$. The two hatched areas $f(x)\cdot\Delta x$ and $f(x+\Delta x)\cdot\Delta x$ are less than resp. greater than this area.

and

$$\frac{A(x+\Delta x)-A(x)}{\Delta x} \to A'(x).$$

The inequality (3.2) then becomes

$$f(x) \leq A'(x) \leq f(x),$$

and we conclude that $A'(x) = f(x)$. This means A is a primitive function, and we have proven the theorem (for increasing functions).

If the function is decreasing, the proof is the same in essence. However, the inequality signs in (3.1) and (3.2) will point in the opposite direction.

If the function is not monotonous, we can divide the x-axis into the monotony intervals of f. In these intervals, the theorem applies, and it must therefore be true for the entire interval because of theorem 3.21. ■

Here, a few examples of how to calculate the area between a graph and the x-axis are given:

Example 3.23
If we want to find the area between the graph of the function $f(x) = \frac{1}{\sqrt{x}}$ and the x-axis in the interval $[4;9]$, we calculate

$$\int_4^9 \frac{1}{\sqrt{x}}\,\mathrm{d}x = \left[2\sqrt{x}\right]_4^9 = 2\cdot\sqrt{9} - 2\cdot\sqrt{4} = 2\,.$$

Therefore, the area (see figure 3.7) is 2.

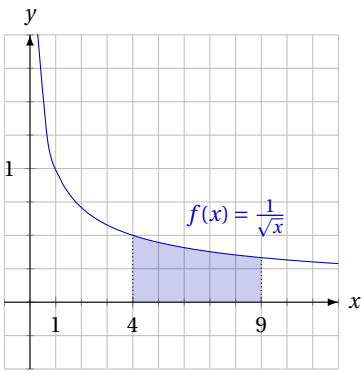


Figure 3.7: The area between the graph of f and the x-axis in the interval $[4;9]$ is 2.

Example 3.24
In figure 3.8, we see that the region M is bounded by the graph of the function

$$f(x) = 3\mathrm{e}^{-x} - \frac{x}{4}$$

together with the x- and the y-axes. We can find the area of M by calculating a definte integral. But before we do that, we need to know the limits of integration.

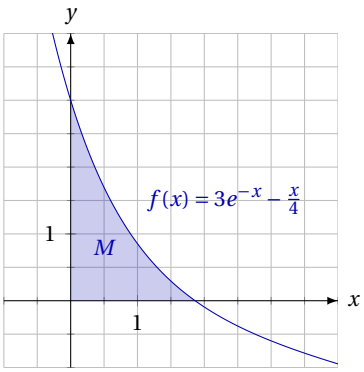


Figure 3.8: A region M is bounded by the graph of $f(x) = 3e^{-x} + \frac{x}{4}$ and the coordinate axes.

The lower limit is 0, since the region begins at the y-axis. The upper limit can be found where the graph intersects the x-axis. To find this value of x, we need to solve the equation $f(x) = 0$, i.e.

$$3e^{-x} - \frac{x}{4} = 0\,.$$

We cannot solve this equation analytically, so we need to use a CAS. We then find the solution

$$x = 1.8628\,.$$

Therefore, we need to integrate $f(x)$ in the interval $[0; 1.8628]$.

This yields

$$\begin{aligned}\int_0^{1.8628} \left(3e^{-x} - \frac{x}{4}\right) dx &= \left[-3e^{-x} - \frac{x^2}{8}\right]_0^{1.8628} \\ &= \left(-3e^{-1.8628} - \frac{1.8628^2}{8}\right) - \left(-3e^{-0} - \frac{0^2}{8}\right) \\ &= 2.1005\,.\end{aligned}$$

So, the area of the region M is 2.1005.

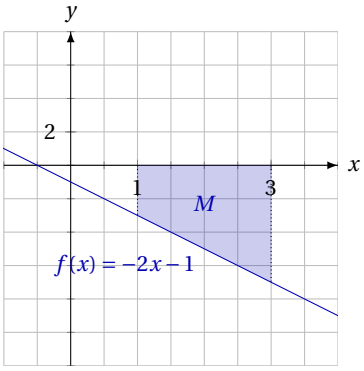


Figure 3.9: The region M lies below the x-axis.

If the Graph is Below the x-Axis

Theorem 3.22 can be used to calculate the area between a graph and the x-axis, but only if the graph lies above the x-axis. What happens if the graph lies below the x-axis?

Example 3.25
In figure 3.9, we see the region M bounded by the graph of $f(x) = -2x - 1$ and the x-axis in the interval $[1; 3]$. Here, we cannot use the definite integral to calculate the area of M, since the graph of f lies below the x-axis.

If we just calculate the definite integral anyway, we find

$$\int_1^3 (-2x - 1)\,dx = \Big[-x^2 - x\Big]_1^3 = (-3^2 - 3) - (-1^2 - 1) = -10\,,$$

which cannot be an area, since an area is never negative.

But the function is linear, hence M is a trapezoid and we can calculate the area geometrically. We find that the area is 10. So, we see that the definite integral still gives us the area, albeit with a negative sign.

The conclusion from this example can be found to hold true in general. The definite integral is the area between the graph and the x-axis *with sign*. If the graph is above the x-axis, we get the area, but if it is below, we find the negative area.

3.7 AREAS BETWEEN GRAPHS

We can also use the definite integral to calculate areas between graphs. If one graph lies completely above another, we can find the area between them by subtracting the area beneath the lower graph from the area beneath the upper graph. We have the following theorem:

Theorem 3.26

Let two functions f and g be given such that $f(x) \geq g(x) \geq 0$ for all x in the interval $[a;b]$. Then the area of the region between the graphs of f and g in the interval $[a;b]$ is

$$\int_a^b (f(x) - g(x))\,\mathrm{d}x\,.$$

Proof

Since the graph of f lies above the graph of g everywhere, the area below the graph of f is greater than the area below the graph of g, and the area between the two graphs is

$$\int_a^b f(x)\,\mathrm{d}x - \int_a^b g(x)\,\mathrm{d}x\,,$$

which, according to theorem 3.20, equals

$$\int_a^b (f(x) - g(x))\,\mathrm{d}x\,. \qquad \blacksquare$$

Example 3.27

In this example, we calculate the area between the graphs of

$$f(x) = \frac{x^2}{3} + 6 \qquad \text{and} \qquad g(x) = -\frac{x^2}{2} + 2x + 3$$

in the interval $[-1;2]$ (see figure 3.10).

In the figure, we see that the graph of f lies above the graph of g. The area is then, according to theorem 3.26,

$$\begin{aligned}
\int_{-1}^{2} (f(x) - g(x))\,\mathrm{d}x &= \int_{-1}^{2} \left(\left(\frac{x^2}{3} + 6\right) - \left(-\frac{x^2}{2} + 2x + 3\right)\right)\mathrm{d}x \\
&= \int_{-1}^{2} \left(\tfrac{5}{6}x^2 - 2x + 3\right)\mathrm{d}x \\
&= \left[\tfrac{5}{18}x^3 - x^2 + 3x\right]_{-1}^{2} \\
&= \left(\tfrac{5}{18}\cdot 2^3 - 2^2 + 3\cdot 2\right) - \left(\tfrac{5}{18}\cdot(-1)^3 - (-1)^2 + 3\cdot(-1)\right) \\
&= \tfrac{17}{2}\,.
\end{aligned}$$

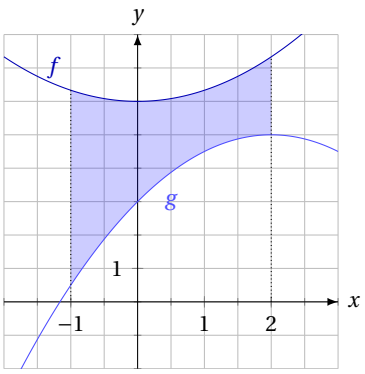


Figure 3.10: The area between the graphs of $f(x) = \frac{x^2}{3} + 6$ and $g(x) = -\frac{x^2}{2} + 2x + 3$ in the interval $[-1;2]$.

Example 3.28

In figure 3.11, we see that a region M is bounded by the graphs of

$$f(x) = x^2 + 1 \qquad \text{and} \qquad g(x) = -x + 3$$

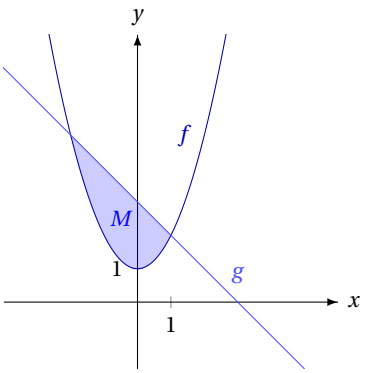


Figure 3.11: M is bounded by the graphs of $f(x) = x^2 + 1$ and $g(x) = -x + 3$.

The area of this region can be calculated as an integral, but in order to do this, we need to know the limits of integration.

We therefore need to find the values of x where the two graphs intersect. We find these values by solving the equation $f(x) = g(x)$:

$$x^2 + 1 = -x + 3 \quad \Leftrightarrow \quad x^2 + x - 2 = 0 \quad \Leftrightarrow \quad x = -2 \vee x = 1\,.$$

So, we need to integrate over the interval $[-2; 1]$. Since the graph of g lies above the graph of f in this interval, we calculate

$$\begin{aligned}\int_{-2}^{1} (g(x) - f(x))\,\mathrm{d}x &= \int_{-2}^{1} \big((-x+3) - (x^2+1)\big)\,\mathrm{d}x \\ &= \int_{-2}^{1} (-x^2 - x + 2)\,\mathrm{d}x \\ &= \Big[-\tfrac{1}{3}x^3 - \tfrac{1}{2}x^2 + 2x\Big]_{-2}^{1} \\ &= \big(-\tfrac{1}{3}\cdot 1^3 - \tfrac{1}{2}\cdot 1^2 + 2\cdot 1\big) \\ &\qquad - \big(-\tfrac{1}{3}\cdot(-2)^3 - \tfrac{1}{2}\cdot(-2)^2 + 2\cdot(-2)\big) \\ &= \tfrac{9}{2}\,,\end{aligned}$$

and this is the area of the region M.

Theorem 3.26 only applies when both graphs lie above the x-axis. But it is, in fact, only necessary that one graph lies above the other. A more general version of theorem 3.26 is therefore

Theorem 3.29

Let f and g be two continuous functions such that $f(x) \geq g(x)$ for all x in the interval $[a; b]$. Then the area between the graphs of f and g in the interval $[a; b]$ is

$$\int_a^b (f(x) - g(x))\,\mathrm{d}x\,.$$

Proof

If the graphs of both functions are above the x-axis, the theorem is the same as theorem 3.26. If this is not the case, the graph of g will have a minimum $-M$ for some positive number M. The two functions

$$f_1(x) = f(x) + M \qquad \text{and} \qquad g_1(x) = g(x) + M\,,$$

will therefore have graphs that are vertical shifts of f and g, and these graphs are above the x-axis. Since the graphs of f_1 and g_1 are above the x-axis, the area between them can be calculated using theorem 3.26, and we get the area

$$\begin{aligned}\int_a^b (f_1(x) - g_1(x))\,\mathrm{d}x &= \int_a^b \big((f(x) + M) - (g(x) + M)\big)\,\mathrm{d}x \\ &= \int_a^b (f(x) - g(x))\,\mathrm{d}x\,.\end{aligned}$$

But the area between the graphs of f_1 and g_1 must be the same as the area between the graphs of f and g, since the two graphs have been shifted vertically by the same quantity. Thus, the area between the graphs of f and g is

$$\int_a^b (f(x) - g(x))\,\mathrm{d}x\,. \qquad \blacksquare$$

Example 3.30
In this example, we calculate the area between the graphs of

$$f(x) = x + 1 \qquad \text{and} \qquad g(x) = 2^{-x} - 3$$

in the interval $[0;3]$ (see figure 3.12).

Since the graph of f lies above the graph of g, we calculate

$$\int_0^3 (f(x) - g(x))\,\mathrm{d}x = \int_0^3 \big((x+1) - (2^{-x} - 3)\big)\,\mathrm{d}x = \int_0^3 (-2^{-x} + x + 4)\,\mathrm{d}x$$

to find the area.

This integral can be calculated by hand, but we can also use a CAS. We then find the area to be

$$\int_0^3 (-2^{-x} + x + 4)\,\mathrm{d}x = 15{,}24\,.$$

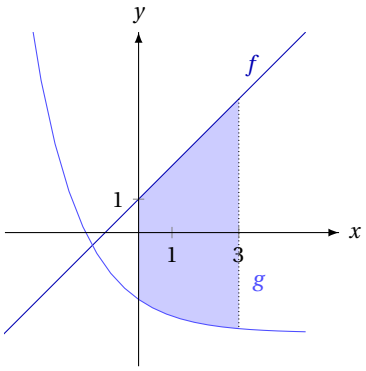


Figure 3.12: The area between the graphs of $f(x) = x + 1$ and $g(x) = 2^{-x} - 3$.

Probability Theory

4

Probability theory is a branch of mathematics, which tries to quantify random phenomena. The first books on probability theory dealt with different games of chance.[3]

Probability theory deals with *outcomes* and their *probabilities*. An *outcome* is the result of an "experiment". This might be

- the result of rolling a die,
- the winnings on lottery ticket, or
- the values of 5 random playing cards.

The *probabilities* are a description of how often a certain outcome of the experiment happens. E.g. how often we get a 5 when we roll a die.

If we roll a die, the probability of getting a 5 is $\frac{1}{6}$, but what does that actually mean? The interpretation is that if we roll the die a huge number of times, about $\frac{1}{6}$ of the rolls will gives us a 5. We find this probability using the formula

$$\text{probability} = \frac{\text{number of wanted outcomes}}{\text{number of possible outcomes}} . \tag{4.1}$$

This formula only works in those cases where each outcome is equally probable. If we want to simplify calculations involving probabilities, we therefore want to describe the outcomes in such a way that they are all equally probable.

To describe which outcomes are "wanted", we describe the outcomes using a so-called *random variable*. This is a quantity which assigns a number to each outcome. The random variable might asign the number 1 to the wanted outcomes and 0 to the rest of the outcomes.[1]

[1] Because a random variable assigns a number to each outcome, it is not actually a variable, but a function. However, this distinction is not important for our purposes.

4.1 OUTCOMES, EVENTS, AND RANDOM VARIABLES

If we roll a die, there are 6 possible outcomes. These outcomes are shown in table 4.1. The 6 outcomes are equally probable, and they make up the *sample space* S, which is the set of all possible outcomes.

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

Table 4.1: Possible outcomes when rolling a die.

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

For each element $s \in S$ in the sample space, there is an associated probability $P(s)$, which is listed in table 4.1. The sum of all the probabilities is 1.

Any subset of the sample space is called an *event*. So, an event is a collection of certain outcomes, which we happen to be looking at (corresponding to the "wanted outcomes" in formula 4.1).

A few events are:

$$E_1 = \{6\}$$
$$E_2 = \{1,2\}$$
$$E_3 = \{1,3,5\}\ .$$

Table 4.2: Outcomes of rolling a die, and the values of the random variables X, Y and Z.

s	$P(s)$	X	Y	Z
1	$\frac{1}{6}$	1	2	1
2	$\frac{1}{6}$	2	2	2
3	$\frac{1}{6}$	3	-1	1
4	$\frac{1}{6}$	4	-1	2
5	$\frac{1}{6}$	5	-1	1
6	$\frac{1}{6}$	6	-1	2

The event E_1 corresponds to rolling a 6, E_2 corresponds to rolling a 1 or a 2, while E_3 corresponds to rolling an odd number. The three events might be described by three random variables, X, Y, and Z, which assign a number to the events, we are looking at ("6", "1 or 2", "odd number"). The values of the three random variables might be chosen as in table 4.2.

We have chosen the values of the random variables, such that every outcome, which is part of the corresponding event, has the same value. Apart from that, we can chose the values completely at random, but we usually choose them so that they describe the event to some extent.

The probability of an event E is denoted by $P(E)$ or $P(X = t)$, where t is the value of X which makes up the event. By looking at the table, we see that with the chosen values of the random variables, we have

$$P(E_1) = P(X = 6)\,, \quad P(E_2) = P(Y = 2) \quad \text{and} \quad P(E_3) = P(Z = 1)\,.$$

We calculate the probability of an event by adding the probabilities of the outcomes, which make up the event:

$$P(X = 6) = P(6) = \tfrac{1}{6}$$
$$P(Y = 2) = P(1) + P(2) = \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{3}$$
$$P(Z = 1) = P(1) + P(3) + P(5) = \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{1}{2}\,.$$

4.2 DISCRETE PROBABILITY DISTRIBUTIONS

The sample space and the random variables, we looked at in the previous sections, are examples of what we call *discrete probability distributions*. We talk about *discrete* distributions when the outcomes are separate and countable. Here, the methods we use are basically the same as when we look at ungrouped data in statistics. The probabilities $P(X = t)$ can be interpreted as relative frequencies.

As mentioned previously, when we want to calculate the probability of an event, it is easier if we choose the sample space, so that every outcome is equally probable.

If we toss a coin 3 times and count the number of "heads", then we have 4 possible outcomes for the number of "heads": 0, 1, 2 or 3 times. But these

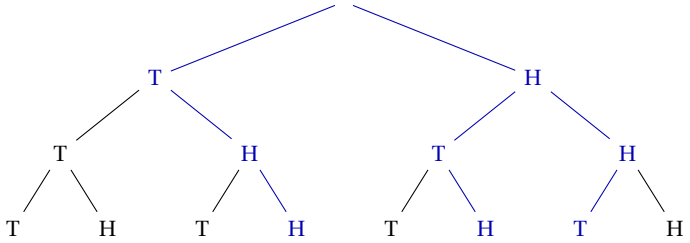


Figure 4.1: If we toss a coin 3 times, there are 8 possible outcomes. The outcomes, which have 2 "heads", are marked.

outcomes are not equally probable, and we might therefore consider them to be events rather than outcomes.

We instead define the outcomes to be the actual results of the tosses, i.e. combinations of "heads" and "tails":

TTT, TTH, THH, etc.

If we look at the number of "heads" in 3 tosses of a coin, then three possible outcomes will have 2 "heads". The event, which is made up of these three outcomes is

$$E = \{\text{HHT}, \text{HTH}, \text{THH}\} .$$

To analyse all the possible outcomes, we can draw a so-called *tree*, see figure 4.1. Here, we can see that there are 8 possible outcomes, which are equally probable, and that the event E contains 3 elements.

To calculate the probabilities, we list all 8 possible outcomes in a table, and let the random variable X count the number of "heads" in the outcomes (see table 4.3).

We see in the table, that $X = 2$ in 3 places, i.e.

$$P(X = 2) = 3 \cdot \tfrac{1}{8} = \tfrac{3}{8} .$$

The total probability distribution for the number of "heads" in 3 coin tosses can be seen in table 4.4.

Another way of presenting the probability distribution is in a bar chart. A bar chart for the distribution in table 4.4 can be seen in figure 4.2.

Here, we provide a few examples of how to find the probability distribution of different discrete random variables.

Example 4.1

If a sample space $S = \{s_1, s_2, s_3, s_4\}$ has the associated probabilities seen in table 4.5, and we also have a random variable X, whose values are as in the table, we can find the probability distribution in the following way.

The random variable has three possible values, -1, 0 and 2. The probabilities of these are calculated by adding the probabilities of the outcomes, which represent these values:

$$P(X = -1) = P(s_1) + P(s_3) = 0.2 + 0.3 = 0.5$$

Table 4.3: The value of the random variable X for all possible outcomes of 3 coin tosses.

s	$X(s)$
TTT	0
TTH	1
THT	1
THH	2
HTT	1
HTH	2
HHT	2
HHH	3

Table 4.4: The probability distribution of X.

t	$P(X = t)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

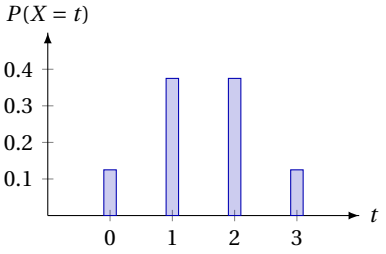


Figure 4.2: The probability distribution of X as a bar chart.

Table 4.5: A series of outcomes with corresponding probabilities, and a random variable X.

s	$P(s)$	X
s_1	0.2	-1
s_2	0.4	0
s_3	0.3	-1
s_4	0.1	2

Table 4.6: X: the result of rolling two dice.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Table 4.7: The probability distribution of the roll of two dice.

t	$P(X=t)$
2	$\frac{1}{36}$
3	$\frac{1}{18}$
4	$\frac{1}{12}$
5	$\frac{1}{9}$
6	$\frac{5}{36}$
7	$\frac{1}{6}$
8	$\frac{5}{36}$
9	$\frac{1}{9}$
10	$\frac{1}{12}$
11	$\frac{1}{18}$
12	$\frac{1}{36}$

$$P(X=0) = P(s_2) = 0.4$$
$$P(X=2) = P(s_4) = 0.1\,.$$

These three values describe the probability distribution of the random variable X.

Example 4.2

If we roll two dice, the sample space S consists of pairs (d_1, d_2), where d_1 is the roll of the first die, and d_2 is the roll of the second. In total, there are 36 such pairs, so the sample space has 36 elements:

$$\begin{aligned} S = \{&(1,1),\,(1,2),\,(1,3),\,(1,4),\,(1,5),\,(1,6),\\ &(2,1),\,(2,2),\,(2,3),\,(2,4),\,(2,5),\,(2,6),\\ &(3,1),\,(3,2),\,(3,3),\,(3,4),\,(3,5),\,(3,6),\\ &(4,1),\,(4,2),\,(4,3),\,(4,4),\,(4,5),\,(4,6),\\ &(5,1),\,(5,2),\,(5,3),\,(5,4),\,(5,5),\,(5,6),\\ &(6,1),\,(6,2),\,(6,3),\,(6,4),\,(6,5),\,(6,6)\}\end{aligned}$$

All of these outcomes are equally probable, i.e. the probability of one of the outcomes is $\frac{1}{36}$.

We now define the random variable X to be the *sum* of the two rolls. The possible values of X are then the numbers from 2 to 12, but these values are not equally probable. As we see from table 4.6, some of the values occur more often than others.

If we want to know the probability of rolling a 9 with two dice, we count the number of 9s in table 4.6 and multiply by $\frac{1}{36}$:

$$P(X=9) = 4 \cdot \frac{1}{36} = \frac{1}{9}\,.$$

The probability distribution of X can be seen in table 4.7.

Mean and Standard Deviation

The probability distributions of random variables show, how probable it is to get certain results of an experiment. The probabilities can be treated in exactly the same way as relative frequencies in statistics. This means we can describe the probability distribution via certain descriptors.

We have the following definition:

Definition 4.3

Let X be a discrete random variable with possible values $x_1, \ldots, x_n$, and let $p_i = P(X = x_i)$.

We then define the mean μ_X and the standard deviation σ_X of X to be the numbers

$$\mu_X = x_1 \cdot p_1 + \cdots + x_n \cdot p_n$$
$$\sigma_X = \sqrt{(x_1 - \mu_X)^2 \cdot p_1 + \cdots + (x_n - \mu_X)^2 \cdot p_n}$$

The mean shows us, which value we should expect to get on average, if we perform the experiment a large number of times. The standard deviation shows, how far the results are on average from the mean.

Example 4.4
If we look at the number of "heads" in 3 coin tosses, then the probability distribution is given in table 4.4.

The mean can then be calculated using the tabulated values:

$$\begin{aligned}\mu_X &= 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) \\ &= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5\,.\end{aligned}$$

This is the number of "heads", we would expect to get in 3 coin tosses. Of course, it is not possible to get 1.5 "heads", the number 1.5 means that *on average* we will get 1.5 "heads" in 3 tosses.

Example 4.5
In the previous example, we calculated the mean of the number of "heads" in 3 coin tosses to be 1.5.

We now calculate the standard deviation σ_X:

$$\begin{aligned}\sigma_X &= \sqrt{(0-1.5)^2 \cdot \frac{1}{8} + (1-1.5)^2 \cdot \frac{3}{8} + (2-1.5)^2 \cdot \frac{3}{8} + (3-1.5)^2 \cdot \frac{1}{8}} \\ &= \sqrt{0.75} \\ &= 0.866\,.\end{aligned}$$

This is a measure of how far the values of X are on average from the mean 1.5.

4.3 CONTINUOUS PROBABILITY DISTRIBUTIONS

The methods we use to describe discrete probability distributions are the same as the ones we use when doing statistics with ungrouped data. Just as we separated statistics into ungrouped and grouped data, we have two possibilities for probability distributions.

When doing statistics with grouped data, the observations are grouped in intervals with a corresponding relative frequency, which we can use to draw a histogram. The histogram provides an illustration of the entire data set, and the area of the histogram is always 1 (or 100%).

In probability theory, we instead talk about *continuous* probability distributions. Here, we know the distribution of probabilities so well that the intervals are infinitely small. Instead of a histogram, we then get a smooth curve. The area beneath this curve is 1 (see figure 4.3).

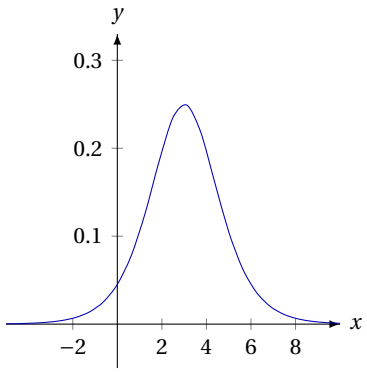


Figure 4.3: The graph of a probability density function. The area beneath the graph is 1.

The function, which is graphed in figure 4.3 is called the *probability density function* of the probability distribution.

When we talk about continuous distributions, it makes no sense to talk about the probability of getting a certain value. We instead talk about the

probability of getting a value in some interval. We calculate this probability as the area beneath the graph of the probability density function between the start and end point of the interval.[2] I.e. we calculate the probability of the random variable X assuming values in the interval $[t_1; t_2]$ as

$$P(t_1 \leq X \leq t_2) = \int_{t_1}^{t_2} f_X(x)\,\mathrm{d}x\,,$$

where f_X is the probability density function of the random variable X.

So, a continuous random variable is a quantity which assumes values in some interval. The probability distribution of a continuous random variable X can then be calculated using the corresponding probability density function $f_X(x)$.

[2] This is exactly how we use histograms when doing statistics with grouped data.

Example 4.6

The probability density function of a random variable X is given by

$$f_X(x) = \frac{\mathrm{e}^{x-3}}{(\mathrm{e}^{x-3}+1)^2}\,.$$

This is the probability density function, which is graphed in figure 4.3.

The probability that the random variable assumes values in the interval $[1;2]$ is then

$$\int_1^2 f_X(x)\,\mathrm{d}x = \int_1^2 \frac{\mathrm{e}^{x-3}}{(\mathrm{e}^{x-3}+1)^2}\,\mathrm{d}x = 0.1497\,.$$

I.e.

$$P(1 \leq X \leq 2) = 0.1497\,.$$

This probability corresponds to the marked area in figure 4.4.

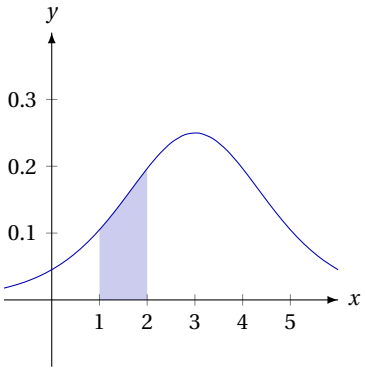


Figure 4.4: The probability $P(1 \leq X \leq 2)$ corresponds to the marked area below the graph of the probability density function.

Example 4.7

A random variable Y has the probability density function

$$f_Y(x) = \begin{cases} \frac{2}{125}x^3 - \frac{21}{125}x^2 + \frac{11}{25}x & \text{for } 0 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}\,.$$

This function has the value 0 outside of the interval $[0;5]$, which means that the probability of getting values outside this interval is 0.

If we want to calculate the probability of the random variable assuming values in the interval $[2;3]$, we calculate

$$\int_2^3 f_Y(x)\,\mathrm{d}x = \int_2^3 \left(\tfrac{2}{125}x^3 - \tfrac{21}{125}x^2 + \tfrac{11}{25}x\right)\mathrm{d}x = \frac{37}{125}\,.$$

I.e.

$$P(2 \leq Y \leq 3) = \frac{37}{125} = 0.296\,.$$

This probability corresponds to the area marked in figure 4.5.

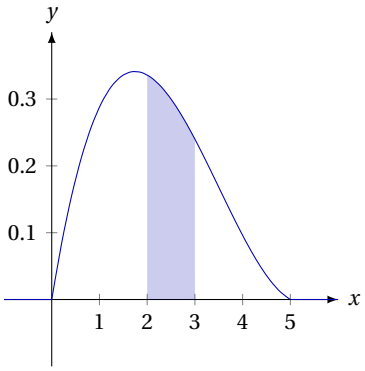


Figure 4.5: The probablity $P(2 \leq Y \leq 3)$ is equal to the marked area below the graph of the probability density function.

Cumulative Distribution Functions

Having to integrate the probability density function, every time we want to calculate a probability can get somewhat tiresome. Therefore, we would like to do the integration once and for all, and so we define the *cumulative distribution function*, F_X:

Definition 4.8

If the random variable X has the probability density function f_X, we define the *cumulative distribution function*

$$F_X(t) = P(X \leq t) = \int_{-\infty}^{t} f_X(x)\,\mathrm{d}x\,.$$

The cumulative distribution function gives us the area below the graph of the probability density function up to the value t. The graph of the cumulative density function is therefore a graph where the values increase from 0 to 1, see figure 4.6.

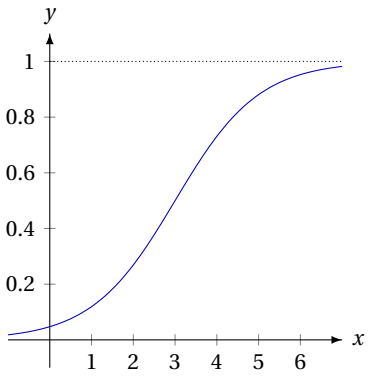


Figure 4.6: The graph of a cumulative distribution function. Notice, how the function values increase from 0 to 1.

The graph of the cumulative distrution function is actually an ogive just like the ones we drew when we looked at statistics with grouped data. So, the function values of F_X correspond to cumulative relative frequencies.

We can use the cumulative distribution function to calculate probabilities. We have the following theorem:

Theorem 4.9

If the random variable X has the cumulative distribution function F_X, then

1. $P(X \leq t) = F_X(t)$,
2. $P(X \geq t) = 1 - F_X(t)$, and
3. $P(t_1 \leq X \leq t_2) = F_X(t_2) - F_X(t_1)$.

Proof

1 follows from the definition of the cumulative distribution function.

2 holds, because

$$P(X \geq t) = 1 - P(X \leq t) = 1 - F_X(t)\,.$$

To prove 3, we calculate

$$\begin{aligned} P(t_1 \leq X \leq t_2) &= \int_{t_1}^{t_2} f_X(x)\,\mathrm{d}x \\ &= \int_{t_1}^{t_2} f_X(x)\,\mathrm{d}x + \int_{-\infty}^{t_1} f_X(x)\,\mathrm{d}x - \int_{-\infty}^{t_1} f_X(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{t_2} f_X(x)\,\mathrm{d}x - \int_{-\infty}^{t_1} f_X(x)\,\mathrm{d}x \\ &= F_X(t_2) - F_X(t_1)\,. \end{aligned}$$

■

Example 4.10
We can calculate the probability $P(1 \le X \le 2)$ from example 4.6 using the cumulative distribution function.

The probability density function is

$$f_X(x) = \frac{e^{x-3}}{(e^{x-3}+1)^2}\,.$$

So, the cumulative distribution function is

$$F_X(t) = \int_{-\infty}^{t} f_X(x)\,dx = \int_{-\infty}^{t} \frac{e^{x-3}}{(e^{x-3}+1)^2}\,dx = \frac{1}{e^{3-t}+1}\,.$$

I.e.

$$P(1 \le X \le 2) = F_X(2) - F_X(1) = \frac{1}{e^{3-2}+1} - \frac{1}{e^{3-1}+1} = 0.1497\,,$$

which is just what we found previously.

We can also find the function values $F_X(2)$ og $F_X(1)$ on the graph, see figure 4.7.

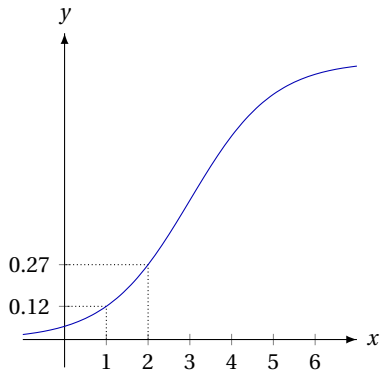


Figure 4.7: The probability $P(1 \le X \le 2) = F_X(2) - F_X(1)$.

Example 4.11
The probability $P(2 \le X \le 3)$ from example 4.7 can also be found by first determining the cumulative distribution function.

The probability density function is

$$f_Y(x) = \begin{cases} \frac{2}{125}x^3 - \frac{21}{125}x^2 + \frac{11}{25}x & \text{for } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}\,.$$

Since this function only has non-zero values in the interval $[0;5]$, the cumulative distribution function will yield 0 for $t < 0$ and 1 for $t > 5$.

In the interval $[0;5]$, the cumulative distribution function is

$$\begin{aligned} F_Y(t) &= \int_{-\infty}^{t} f_Y(x)\,dx = \int_{0}^{t} \left(\tfrac{2}{125}x^3 - \tfrac{21}{125}x^2 + \tfrac{11}{25}x\right) dx \\ &= \tfrac{1}{250}t^4 - \tfrac{7}{125}t^3 + \tfrac{11}{50}t^2\,. \end{aligned}$$

In total, the cumulative distribution function is

$$F_Y(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{250}t^4 - \frac{7}{125}t^3 + \frac{11}{50}t^2 & \text{for } 0 \le t \le 5 \\ 1 & \text{for } t > 5 \end{cases}\,.$$

This function is graphed in figure 4.8.

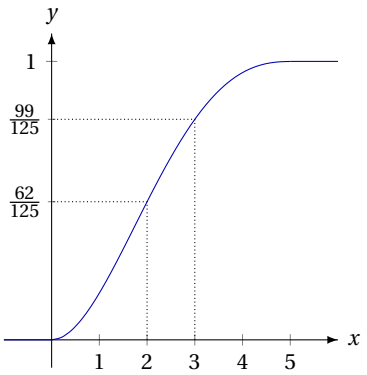


Figure 4.8: The graph of F_Y. $P(2 \le Y \le 3) = F_Y(3) - F_Y(2)$.

We can now calculate the probability $P(2 \le Y \le 3)$:

$$P(2 \le Y \le 3) = F_Y(3) - F_Y(2) = \frac{99}{125} - \frac{62}{125} = \frac{37}{125}\,,$$

which is the same value as in example 4.7.

The two function values $F_Y(2)$ and $F_Y(3)$ can be seen in figure 4.8.

Mean and Standard Deviation

We can also describe continuous probability distributions using certain descriptors. As for discrete distributions, we have the mean and standard deviation. These are calculated using the probability density function.

Definition 4.12

Let X be a continuous random variable, and let f_X be the probability density function of X.

The mean μ_X and standard deviation σ_X of X are then

$$\mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x$$
$$\sigma_X = \sqrt{\int_{-\infty}^{\infty} (x-\mu_X)^2 \cdot f_X(x)\,\mathrm{d}x}$$

Because these integrals are sometimes quite difficult to calculate, we normally use a CAS to calculate them.

Example 4.13

In example 4.6, we looked at a random variable with probability density function

$$f_X(x) = \frac{\mathrm{e}^{x-3}}{(\mathrm{e}^{x-3}+1)^2}\,.$$

The mean is then

$$\mu_X = \int_{-\infty}^{\infty} x \cdot f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x \cdot \frac{\mathrm{e}^{x-3}}{(\mathrm{e}^{x-3}+1)^2}\,\mathrm{d}x = 3\,,$$

and the standard deviation is

$$\begin{aligned}\sigma_X &= \sqrt{\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,\mathrm{d}x} \\ &= \sqrt{\int_{-\infty}^{\infty} (x-3)^2 \cdot \frac{\mathrm{e}^{x-3}}{(\mathrm{e}^{x-3}+1)^2}\,\mathrm{d}x} \\ &= \frac{\pi}{\sqrt{3}} \approx 3.29\,.\end{aligned}$$

The mean and the standard deviation are shown on the graph of the probability density function in figure 4.9.

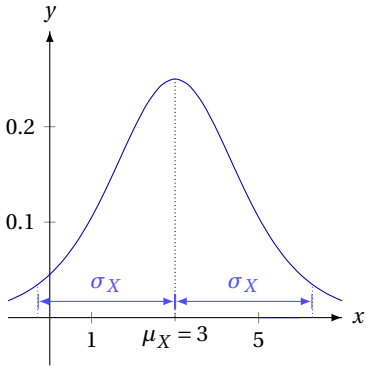


Figure 4.9: The graph of f_X (the probability density function). The mean μ_X and the standard deviation σ_X are shown on the graph.

Example 4.14

The random variable from example 4.7 has the probability density function

$$f_Y(x) = \begin{cases} \frac{2}{125}x^3 - \frac{21}{125}x^2 + \frac{11}{25}x & \text{for } 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}\,.$$

We can now calculate the mean using this function. When we calculate the integral, we need to remember that the probability density function only has non-zero values in the interval $[0;5]$, i.e.

$$\mu_Y = \int_{-\infty}^{\infty} x \cdot f_Y(x)\,\mathrm{d}x = \int_0^5 x \cdot \left(\tfrac{2}{125}x^3 - \tfrac{21}{125}x^2 + \tfrac{11}{25}x\right)\mathrm{d}x = \frac{25}{12} \approx 2.08\,.$$

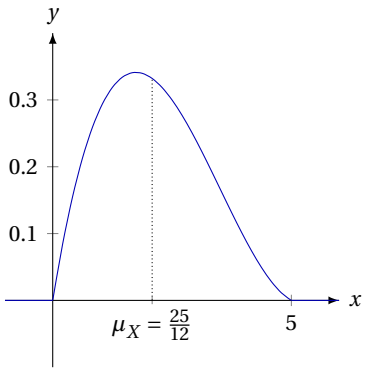


Figure 4.10: The graph of f_Y, and the mean μ_Y.

The graph of the probability density function with the mean μ_Y marked is shown in figure 4.10.

Example 4.15

The standard deviation of the random variable from example4.7 can be found using the mean $\mu_Y = \frac{25}{12}$.

The standard deviation is

$$\begin{aligned}
\sigma_Y &= \sqrt{\int_{-\infty}^{\infty} \left(x - \tfrac{25}{12}\right)^2 \cdot f_Y(x)\,\mathrm{d}x} \\
&= \sqrt{\int_0^5 \left(x - \tfrac{25}{12}\right)^2 \cdot \left(\tfrac{2}{125}x^3 - \tfrac{21}{125}x^2 + \tfrac{11}{25}x\right)\mathrm{d}x} \\
&= \sqrt{\frac{155}{144}} \\
&= \frac{\sqrt{155}}{12} \\
&\approx 1.037\,.
\end{aligned}$$

The Binomial Distribution 5

The *binomial distribution* is a probability distribution used to calculate the probability of a certain number of successes in a number of repeated experiments. For example if we look at the probability of getting three 6s in five rolls of a die.

In this case, we start with an experiment called the *binomial trial*, which we perform a number of times. Every time we do the experiment, there is a probability of success p, and a probability of failure $1-p$.[1]

[1] If the experiment is not a success, it is a failure. Hence the probability of success and the probability of failure must add up to 1.

In the example above, the binomial trial is the roll of a die. We perform this experiment five times. The probability of success (i.e. a 6) is $\frac{1}{6}$. The probability of failure is then $1-\frac{1}{6}=\frac{5}{6}$, which is the probability of getting anything but a 6.

Getting three 6s in the five rolls can happen in several ways. The first three rolls might be 6s—or the last three. So, getting three 6s can happen in quite a lot of different ways.

Therefore, to be able to calculate the probability of three 6s in five rolls, we first need to know, in how many ways it can happen.

5.1 THE BINOMIAL COEFFICIENT

The binomial coefficient is a number, which tells us in how many ways it is possible to choose a certain number out of a larger set, e.g. how many ways to choose 3 out of 5.

We want to find a formula for the binomial coefficient, but first we need some notation, which will make the formulas easier to read.

Definition 5.1

If n is a natural number larger than 0, we define $n!$ as

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 \,.$$

$0!$ is defined to be $0! = 1$.
The number $n!$ is called "n factorial".

Example 5.2
The number 6! is

$$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720\,.$$

As we see from this, $n!$ can be quite a large number, even for small values of n.

Choosing 3 from 5 is a manageable problem, we can actually just count how many different combinations there are. If we want to choose three letters out of ABCDE, we can do it in the ways listed in table 5.1. So, there are 10 ways of choosing 3 out of 5.

Table 5.1: The different ways of choosing 3 letters from ABCDE.

ABC	ACD	BCD	CDE
ABD	ACE	BCE	
ABE	ADE	BDE	

We can also arrive at this number by calculation. We can choose the first letter in 5 different ways, the next letter in 4 ways, and the last in 3 ways. This gives us

$$5 \cdot 4 \cdot 3 = 60$$

different ways of choosing the three letters. This is a lot more than in the table, so clearly we are missing something. What we forgot to consider, is the fact that choosing ABC is no different from choosing CBA, since these are the same three letters. We can arrange three letters in

$$3 \cdot 2 \cdot 1 = 6$$

different ways, so the 60 ways we just calculated actually fall in groups of 6 equal choices. Therefore, we actually only have

$$\frac{60}{6} = 10$$

ways of choosing 3 out of 5. Luckily, this is the same number we found by counting.

So, to find out in how many ways we can choose 3 out of 5, we calculate

$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10\,,$$

which we can rewrite to get

$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = \frac{5!}{3! \cdot 2!} = \frac{5!}{3! \cdot (5-3)!}\,.$$

This calculation tells us in how many ways we can choose 3 out of 5. From this we can extrapolate a general formula.

Definition 5.3: The binomial coefficient

The *binomial coefficient* C_n^r tells us in how many ways we can choose r out of n. The number C_n^r is

$$\mathsf{C}_n^r = \frac{n!}{r! \cdot (n-r)!}\,.$$

Example 5.4
A deck of playing cards contains 52 cards. If we want to choose 5 of these, we can do that in

$$\mathsf{C}_{52}^5 = \frac{52!}{5! \cdot (52-5)!} = \frac{52!}{5! \cdot 47!} = 2\,598\,960$$

different ways.

5.2 THE BINOMIAL DISTRIBUTION

We found in the previous section that we can choose 3 out of 5 in 10 different ways, i.e. $\mathrm{C}_5^3 = 10$. If we want to calculate the probability of three 6s in five rolls of a die, we now know that we can get the three 6s in 10 different ways.

One of these is getting three 6s in the first three rolls. The probability of getting a 6 in one roll of a die is $\frac{1}{6}$. We now want this to happen for the first three rolls, and we want to not get a 6 in the fourth and the fifth roll. The probability of not getting a 6 is $\frac{5}{6}$. The total probability of first getting three 6s and then two of something else is

$$\overbrace{\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}}^{\text{The first three}} \cdot \overbrace{\frac{5}{6}\cdot\frac{5}{6}}^{\text{The last two}} = \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 .$$

All the ways in which we might get three 6s have to be equally probable. So, if we are interested in knowing the probability of three 6s in five rolls (and not just three 6s in the first three rolls), we need to multiply the probability we just found by the number of ways, we can get three 6s. Therefore the probability is

$$10 \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^2 = \frac{125}{3888} \approx 0.0322 . \tag{5.1}$$

The calculation (5.1) can also be written as

$$\mathrm{C}_5^3 \cdot \left(\frac{1}{6}\right)^3 \cdot \left(1 - \frac{1}{6}\right)^{5-3} . \tag{5.2}$$

Here, we have shown more clearly, where the numbers 10 and $\frac{5}{6}$ come from. And this calculation only uses the original numbers, i.e. the number of 6s (3), the number of rolls (5), and the probability of getting a 6 in one roll ($\frac{1}{6}$).

The General Formula

We want to find a general formula for the binomial distribution, and we define a random variable X, which counts the number of successes in n trials. For each repeated trial, the probability of success is p.

So, X is binomially distributed with the number of trials n and probability of success p. The probability of r successes, $P(X = r)$, can then be calculated using the following formula, which is a generalisation of the calculation in (5.2).

Theorem 5.5

If the random variable X is binomially distributed with number of trials n and probability of success p, then the probability of r successes is

$$P(X = r) = \mathrm{C}_n^r \cdot p^r \cdot (1-p)^{n-r} .$$

Example 5.6
What is the probability of getting exactly four 1s in 15 rolls of a die?

The random variable, which counts the number of 1s in the 15 rolls, is binomially distributed with the number of trials $n = 15$, and the probability of success $p = \frac{1}{6}$. The probability of getting four 1s is therefore

$$P(X = 4) = \mathsf{C}_{15}^{4} \cdot \left(\frac{1}{6}\right)^{4} \cdot \left(1 - \frac{1}{6}\right)^{15-4}$$
$$= 1365 \cdot \left(\frac{1}{6}\right)^{4} \cdot \left(\frac{5}{6}\right)^{11} = 0.1418\,.$$

So, there is a probability of 14.18% to get exactly four 1s in 15 rolls of a die.

Example 5.7
A small tropical island in the Pacific is flooded during the summer on average every 4 years. So, the probability of the island flooding during a single summer is $\frac{1}{4}$.

The random variable, which counts the number of floodings in a 5-year period, is binomially distributed with number of trials $n = 5$ and probability of success $\frac{1}{4}$. In a 5-year period, the island may be flooded anywhere between 0 and 5 times. The probability distribution can the be found by calculating $P(X = 0)$, $P(X = 1)$, ..., $P(X = 5)$.

E.g.

$$P(X = 3) = \mathsf{C}_{5}^{3} \cdot \left(\frac{1}{4}\right)^{3} \cdot \left(\frac{3}{4}\right)^{2} = 0.0879\,.$$

This is the probability of the island flooding 3 times during a 5-year period. The total distribution is listed in table 5.2. A bar chart is shown in figure 5.1.

Table 5.2: The probability of r floodings in a 5-year period.

r	$P(X = r)$
0	0.2373
1	0.3955
2	0.2637
3	0.0879
4	0.0146
5	0.0010

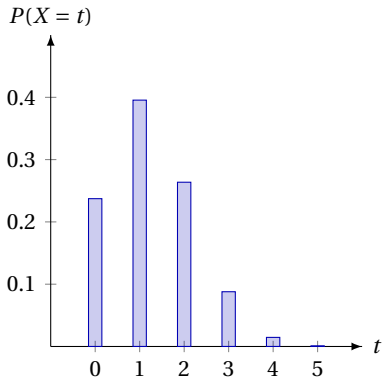


Figure 5.1: The probability distribution of X: The number of floodings in a 5-year period.

The table and the figure tells us that the most probable event is one flooding during the 5 years. We can also see that the probability of no floodings is quite large, but a flooding in every one of the 5 years is very improbable (probability 0.0010 = 0.10%).

If we want to find the probability of no more than one flooding in 5 years, we calculate

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.2373 + 0.3955 = 0.6328\,.$$

So, it is quite probable that we will have no more than one flooding during the 5 years. But, we also have a probability of

$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.6328 = 0.3672$$

for more than 1 flooding during the 5 years.

5.3 MEAN AND STANDARD DEVIATION

The mean and the standard deviation of a binomially distributed random variable can be found using these formulas, which we do not prove:

Theorem 5.8

If the random variable X is binomially distributed with number of trials n and probability of success p, then

$$\mu_X = np$$
$$\sigma_X = \sqrt{np(1-p)}\,.$$

Example 5.9

If we roll a die 10 times and count the number of 5s, then the random variable representing the number of 5s is binomially distributed with number of trials $n = 10$ and probability of success $p = \frac{1}{6}$.

The mean is then

$$\mu_X = n \cdot p = 10 \cdot \frac{1}{6} \approx 1.667\,.$$

So, if we roll a die 10 times, we will get 1.667 5s on average.

The standard deviation is

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{10 \cdot \frac{1}{6} \cdot \left(1 - \frac{1}{6}\right)} = 1.179\,.$$

Example 5.10

In examples 5.7, we looked at a random variable, where the number of trials was 5, and the probability of success was $\frac{1}{4}$.

Here, the mean is

$$\mu_X = n \cdot p = 5 \cdot \frac{1}{4} = 1.25\,,$$

and the standard deviation is

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{5 \cdot \frac{1}{4} \cdot \left(1 - \frac{1}{4}\right)} = 0.9682\,.$$

The Normal Distribution 6

A lot of statistical measurements can be described by the probability distrubtion called the *normal distribution*. An example could be something like the thickness of bread sliced using a machine.

No machine slices perfectly. Table 6.1 lists measurements for a machine, which is supposed to slice bread into 1 cm slices. Some of the slices are too thick, and some are too thin; but it would seem that most of the slices have a thickness around 1 cm.

Using the table, we can find the mean μ and the standard deviation σ of the thickness of the slices. We get

$$\mu = 1.151 \qquad \text{and} \qquad \sigma = 0.248\,.$$

In figure 6.1, we have drawn a histogram of the distribution from table 6.1. In the figure, we have also drawn the graph of the probability density function of the *normal distribution with mean 1.151 and standard deviation 0.248*. The graph of the probability density function is a bell-shaped curve, which seems to fit the distribution of the thickness of the slices quite well.

The probability density function of the normal distribution is given by this definition:

Definition 6.1

The *normal distribution* with mean μ and standard deviation σ has the probability density function

$$f_{\text{norm}}(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,.$$

The corresponding cumulative distribution funcion is

$$F_{\text{norm}}(t;\mu,\sigma) = \int_{-\infty}^{t} f_{\text{norm}}(x;\mu,\sigma)\,\mathrm{d}x\,.$$

A lot of phenomena result in normally distributed data. Some of them are

- Experimental measurement errors.

Table 6.1: The thickness of 100 slices of bread.

Thickness (cm)	Frequency
0.55–0.65	2
0.65–0.75	4
0.75–0.85	6
0.85–0.95	9
0.95–1.05	14
1.05–1.15	16
1.15–1.25	15
1.25–1.35	10
1.35–1.45	10
1.45–1.55	8
1.55–1.65	6

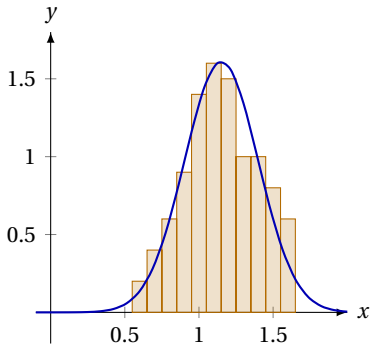
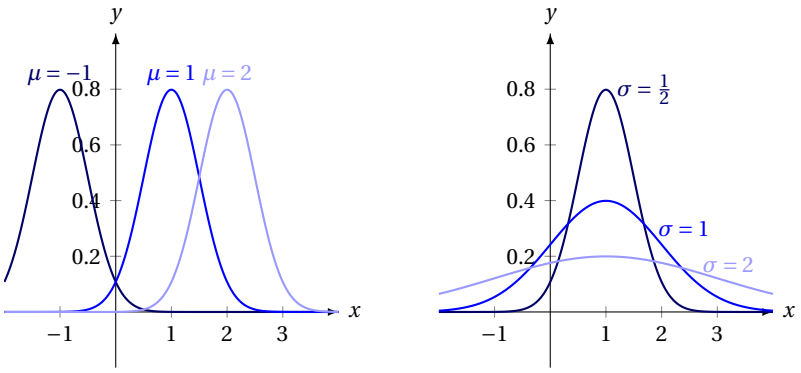


Figure 6.1: Histogram of the thickness of the slices.



(a) Different mean.

(b) Different standard deviation.

Figure 6.2: If the probability density functions have the same standard deviation, but different means, the graphs are shifted horizontally. If they have the same mean, but different standard deviations, then the curves have different widths.

- The size of anything produced using a machine (like the thickness of the slices of bread).
- Biological variables such as height and weight.[1]

[1] A lot of biological variables are actually only approximately normally distributed. In reality they are often log-normally distributed.[4].

The mean and the standard deviation change the shape of the curve. If we change the mean, the curve shifts horizontally. If the standard deviation decreases, the curve gets narrower; and if the standard deviation increases, the curve widens (see figure 6.2).

Since a normally distributed random variable is a continuous random variable, we can find the probability of measurements in a certain interval by calculating the area below the graph of the probability density function.

Example 6.2

In a factory, a machine fills 1 kg bags of sugar. The weight of the bags is normally distributed with mean $\mu = 1000$ g and standard deviation $\sigma = 25$ g. The probability density function is then

$$f_{\text{norm}}(x; 1000, 25) = \frac{1}{\sqrt{2\pi} \cdot 25} \cdot \mathrm{e}^{-\frac{(x-1000)^2}{2 \cdot 25^2}} .$$

If we want to calculate probabilities, it is, however, easier to use the cumulative distribution function. The probability of a bag weighing between 950 and 975 g can then be found:

$$\begin{aligned} P(950 \leq X \leq 975) &= F_{\text{norm}}(975; 1000, 25) - F_{\text{norm}}(950; 1000, 25) \\ &= 0.1359 = 13.59\% . \end{aligned}$$

So, it is actually quite probable to get a bag, which weighs a little below 1 kg.

The graph of the probability density function of the normal distribution is symmetric around the mean. Actually the probability density function is so nicely behaved that we find (we will not prove this theorem):

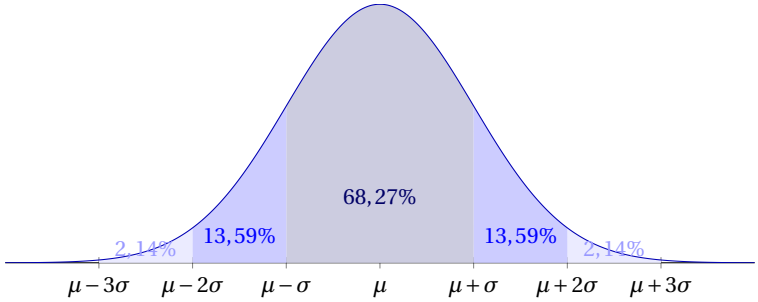


Figure 6.3: For a normally distributed random variable, the probability of X being in a symmetric interval of 1 standard deviation to each side of the mean is a fixed number. The same applies to an interval of 2 standard deviations to each side of the mean, etc.

Theorem 6.3

If X is a normally distributed random variable with mean μ and standard deviation σ, then

$$P(\mu-\sigma \le X \le \mu+\sigma) = 0.6827$$
$$P(\mu-2\sigma \le X \le \mu+2\sigma) = 0.9545$$
$$P(\mu-3\sigma \le X \le \mu+3\sigma) = 0.9973\,.$$

This theorem tells us that 68.27% of the measurements will be in an interval which covers 1 standard deviation to each side of the mean, 95.45% will be in an interval covering 2 standard deviations to each side of the mean, etc. This is illustrated in figure 6.3.

Example 6.4

Here, we take a further look at our previous concerning the slices of bread. The random variable, which measured the thickness of the slices, turned out to be normally distributed with mean $\mu = 1.151$ and standard deviation $\sigma = 0.248$.

Now that we know this, we can answer questions like

1. What is the probability of getting a slice of bread with a thickness between 0.9 cm and 1 cm?
2. What is the probability of getting a slice with a thickness of more than 1.3 cm?

We can find both answers by using the cumulative distribution function

$$F_{\text{norm}}(t; 1.151, 0.248)$$

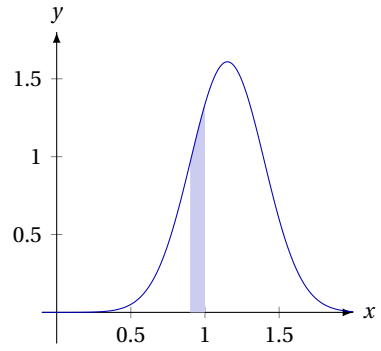
to calculate the area below the graph of the probability density function. The probability of getting a slice with a thickness between 0.9 cm and 1 cm is then

$$\begin{aligned} P(0.9 \le X \le 1) &= F_{\text{norm}}(1; 1.151, 0.248) - F_{\text{norm}}(0.9; 1.151, 0.248) \\ &= 0.1156\,. \end{aligned}$$

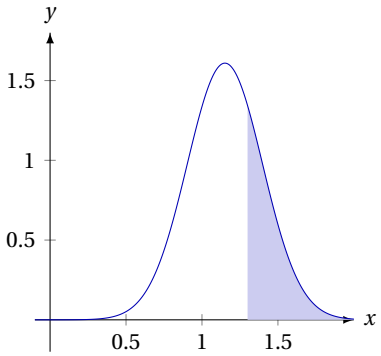
The probability of getting a slice of bread, which is more than 1.3 cm thick is

$$P(X \ge 1.3) = 1 - P(X \le 1.3) = 1 - F_{\text{norm}}(1.3; 1.151, 0.248) = 0.2740\,.$$

The areas representing these two probabilities are shown in figure 6.4.



(a) $P(0.9 \le X \le 1) = 0.1156$.



(b) $P(X \ge 1.3) = 0.2740$.

Figure 6.4: The probability that the thickness of the slices lie in certain intervals is given by the area below the graph of the probability density function.

Hypothesis Tests

7

In this chapter, we will describe two methods for testing a hypothesis based on a random sample. Both tests use the probability distribution called the χ^2-distribution.[1]

[1] The symbol χ is the greek letter *chi* (pronounced "ki").

7.1 STATISTICAL TESTS

A candy company producing a certain kind of jelly beans writes on their website that the colour distribution of their jelly beans is: 20% red, 30% yellow and 50% green.

If we want to test this claim, the obvious thing to do is to take a random sample. So, we buy a large bag of 100 jelly beans and look at the colour distribution. The colour distribution in our sample is listed in table 7.1.

In the table, we have also listed the expected values, i.e. how many jelly beans of each colour, we would expect to find in the bag. The expected values are calculated simply by taking 20%, 30% and 50% of the 100 jelly beans—because we expect to find a distribution which corresponds exactly to the company's claim.

But in reality, we cannot determine based on just one sample, if the company's claim is true. Because the company produces a huge amount of jelly bean bags, we might find several bags where the distribution is not *exactly* 20%-30%-50%.

Table 7.1: The distribution in a bag with 100 jelly beans.

Colour	Number	Expected
Red	16	20
Yellow	40	30
Green	44	50

On the other hand, it is quite unlikely to find a bag where almost every jelly bean is red. If our sample had contained 98 red jelly beans, we might conclude that the company's claim is untrue. But the questions is then, how far from the expected distribution must our sample be, before we reject the company's claim?

To quantify this, we first need to find out what we mean by "far". We therefore calculate the so-called χ^2-statistic:

Definition 7.1

The difference between the observed distibution $o_1, \ldots, o_n$ and the expected distribution $e_1, \ldots, e_n$ is measured by the χ^2-statistic:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_n - e_n)^2}{e_n}.$$

The number n in the definition is the number of categories. In our case $n = 3$, because there are 3 different colours of jelly beans. So, in our example we add 3 fractions to calculate the χ^2-statistic:

$$\chi^2 = \frac{(16-20)^2}{20} + \frac{(40-30)^2}{30} + \frac{(44-50)^2}{50} = 4.85\,.$$

7.2 THE DISTRIBUTION OF THE χ^2-STATISTIC

Before we can determine if this value of χ^2 is large or small, we need to know what to expect from a random sample, if the claim is true. To determine what we can expect, we can do a computer simulation of random samples. We simulate what random samples would be if the company's claim is actually true.

For each simulated sample, we get a new χ^2-statistic. The one χ^2-statistic, we found from our real sample cannot tell us, what to expect—but if we simulate a large amount of samples, we might get an idea.

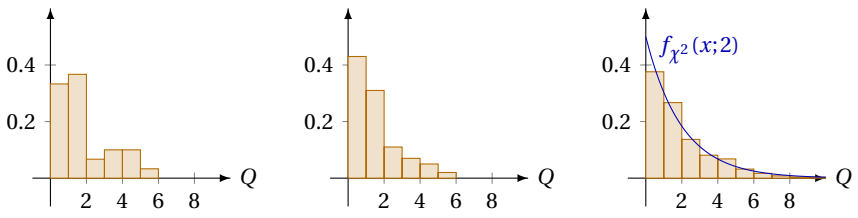
In table 7.2, we have listed how the χ^2-statistics are distributed for 30, 100 and 1000 different simulated samples. As we see from the table, a lot of samples have small values of the χ^2-statistic, but a few have larger values—this shows us that even though most of the samples will be close to the expected values, a few of them will be quite different. This happens because we take random samples.

We might get a better idea of what the distribution looks like if we illustrate the data. In figure 7.1, we see three histograms based on table 7.2. The histograms are scaled, so that the area of a column corresponds to the relative frequency of the interval.

Looking at the figure, we see how the distribution of the χ^2-statistic approaches a certain distribution when the number of samples increases. If we repeated the simulation, we would not get the exact same numbers

Table 7.2: The distribution of χ^2 in 30, 100 and 1000 simulated samples. Values above 10 are omitted from the table.

	Number of samples		
χ^2	30	100	1000
[0;1[	10	43	376
[1;2[	11	31	267
[2;3[	2	11	137
[3;4[	3	7	81
[4;5[	3	5	68
[5;6[	1	2	32
[6;7[	0	0	17
[7;8[	0	0	8
[8;9[	0	0	4
[9;10[	0	0	3



(a) 30 samples. (b) 100 samples. (c) 1000 samples.

Figure 7.1: The simulated distribution of the χ^2-statistic for 30, 100 and 1000 random samples.
In the last figure, the graph of the probability density function of the χ^2-distribution with 2 degrees of freedom is included.

as those in table 7.1; but for a large number of samples, the shape of the histogram would look just like the one in figure 7.1(c).

It is possible to calculate how the χ^2-statistics will be distributed for a large number of samples. In figure 7.1(c), we have added a curve, which follows this theoretical distribution. As we can see, this curve is consistent with the histogram.

So, the curve in figure 7.1(c) shows how the χ^2-statistics are distributed for a very large number of samples (if the original distribution of jelly bean colours is as the company claims).

The curve is the graph of the probability density function of a certain probability distribution: *The χ^2-distribution with 2 degress of freedom.* How the χ^2-statistics are distributed depends on the number of categories. It is this we describe by the *degrees of freedom*, which is

$$df = \text{number of categories} - 1\,.$$

In the jelly bean example, $df = 2$ because there are 3 different colours.[2]

The probability density function of the χ^2-distribution with n degrees of freedom is called $f_{\chi^2}(x;n)$. In figure 7.2, we see the graphs of the probability density functions with $df = 1,\ldots,5$.

[2]There are 2 degrees of freedom because if we know that there are 100 jelly beans, we only need to know 2 of the numbers in table 7.1 to know them all.

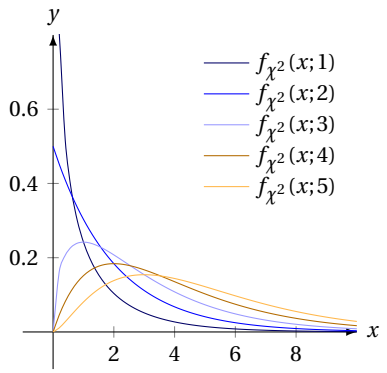


Figure 7.2: χ^2-distributions with 1, 2, 3, 4 and 5 degrees of freedom.

7.3 GOODNESS OF FIT

The χ^2-distribution is the basis of a statistical test called *goodness of fit*, first described in an article by Karl Pearson in 1900.[5]

The idea behind the goodness-of-fit test is that if we know the distribution of N variables, then the χ^2-statistic for an infinite amount of samples will be distributed according to $f_{\chi^2}(x;N-1)$.

In the previous example with the jelly beans, we found that the χ^2-statistic has the value 4.85. If the company's claim is correct, then all of the possible samples have χ^2-statistics distributed according to $f_{\chi^2}(x;2)$. We can use this to calculate the probability of a random sample having a χ^2-statistic with this value or more if the company's claim is true. We calculate this probability using the cumulative distribution function:

$$P(\chi^2 \geq 4.85) = 1 - P(\chi^2 < 4.85) = 1 - F_{\chi^2}(4.85;2).$$

The function $F_{\chi^2}(x;n)$ is usually quite complicated, but fortunately we can just use a CAS to calculate

$$P(\chi^2 \geq 4.85) = 1 - P(\chi^2 < 4.85) = 1 - F_{\chi^2}(4.85;2) = 0.088 = 8.8\%\,.$$

So, if the company's claim is true, then there is a probability of 8.8% to get a random sample, whose χ^2-statistic has a value of 4.85 or more. We call this the P-value, and it is illustrated in figure 7.3.

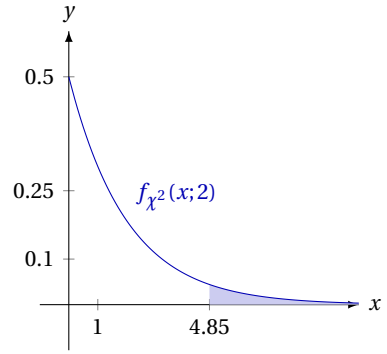


Figure 7.3: The marked area is 0.088, i.e. $P(\chi^2 \geq 4.85) = 8.8\%$.

Null Hypothesis and Significance Level

Even though we now know that there is a probability of 8.8% to get a random sample, whose χ^2-statistic has a value of 4.85 or more, we still do not know if we want to accept the company's claim or not.

Before we calculate the χ^2-statistic, we must therefore have decided how small this probability has to be for us to reject the company's claim. We call this the *significance level*.

We typically choose 1%, 5% or 10%.

When we want to find out, whether to accept the company's claim or not, we perform a *hypothesis test*:

1. First, we describe the *hypothesis* we want to test. This is called the *null hypothesis*, H_0. In this case, the null hypothesis is

   $$H_0: \text{ There are 20\% red, 30\% yellow and 50\% green jelly beans.}$$

2. Then we choose a significance level, e.g. 5%.
3. Next, we use the null hypothesis to calculate the expected values, and then the χ^2-statistic of the sample.

   Here, the χ^2-statistic is $\chi^2 = 4.85$.
4. We then calculate the probability of a random sample having a χ^2-statistic with this value or more. The degrees of freedom in the χ^2-distribution we use for the calculations is the number of categories minus 1.

   Here, we get

   $$P(Q \geq 4.85) = 8.8\% .$$

5. Lastly, we compare the probability to the chosen significance level. If the probability is less than the significance level, we *reject* the null hypothesis, otherwise we *accept* the hypothesis.

   In this case, we accept the hypothesis since $8.8\% > 5\%$, which was the chosen significance level.

If, instead of choosing 5%, we had chosen a significance level of 10%, we would have rejected the hypothesis. This is why it is important to choose the significance level, *before* we do the investigation—otherwise we can choose the significance level, so that we accept or reject the hypothesis according to our wishes.[3]

[3] If we choose a low significance level, we have a greater chance of accepting a hypothesis, even though it might turn out to be wrong (false positive). On the other hand, there is a greater chance of rejecting a true hypothesis if we choose a high significance level (false negative).

Critical Value

In our previous investigation, we calculated the probability of the χ^2-statistic having a value of 4.85 or more. Instead, we could have found out how large the χ^2-statistic has to be before the probability falls below the significance level. This is called the *critical value*. In the example with the jelly beans, we can find the critical value C by solving the equation

$$P(\chi^2 \geq C) = 0.05 \quad \Leftrightarrow \quad 1 - F_{\chi^2}(C; 2) = 0.05 .$$

We solve this equation using a CAS, and get

$$C = 5.99\,.$$

I.e. if the χ^2-statistic is below this value, we accept the hypothesis, see figure 7.4.

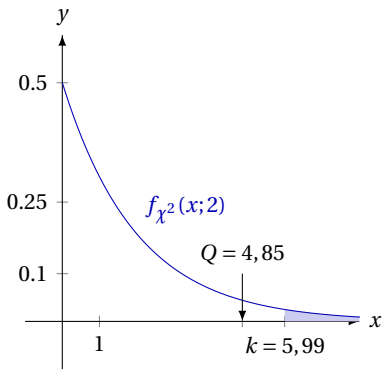


Figure 7.4: The marked area is 0.05 = 5%. Here, $\chi^2 = 4.85$ is below the critical value $C = 5.99$ and we accept the null hypothesis.

7.4 TEST FOR INDEPENDENCE

In table 7.3, we see the results of a poll preceding an election. It is possible to vote for the parties D, M and Q. As we see from the table, men and women do not vote exactly alike.

Since there is a difference, we might ask, whether the choice of party depends on sex. We can investigate this using the χ^2-distribution.

In order to do this, we first need to calculate a χ^2-statistic, but to do this, we need the expected values. We calculate this based on the assumption that votes and sex are *in*dependent. We therefore start by calculating how many percent are going to vote for the different parties, independent of sex.

In table 7.4, we calculate how many people in total are voting for the different parties. We then calculate, how many percent this corresponds to.

We see that 386 men and 297 women are going to vote for party D. This corresponds to a total of 683 votes. Since 1301 people participated in the poll, this corresponds to

$$\frac{683}{1301} = 0.525 = 52.5\%\,.$$

The rest of the relative totals are calculated in the same way.

We now calculate the expected values based on the assumption that men and women vote alike, i.e. $\frac{683}{1301}$ of the men vote for party D, and so do $\frac{683}{1301}$ of the women. Since 695 participated in the poll, the expected number of men who vote for party D is

$$\frac{683}{1301} \cdot 695 = 364.9\,.$$

There are 606 women, so the expected number of women who vote for party D is

$$\frac{683}{1301} \cdot 606 = 318.1\,.$$

Table 7.3: The result of a poll.

	Men	Women
D	386	297
M	127	134
Q	158	145
Undecided	24	30

	Men	Women	Total	Relative total
D	386	297	683	$\frac{683}{1301} = 52.5\%$
M	127	134	261	$\frac{261}{1301} = 20.1\%$
Q	158	145	303	$\frac{303}{1301} = 23.3\%$
Undecided	24	30	54	$\frac{54}{1301} = 4.2\%$
Total	695	606	1301	

Table 7.4: In this table, we have calculated the total number of votes for each party as well as the relative total.

Table 7.5: The expected values are calculated using the relative totals in table 7.4. Notice that the totals are the same for the expected and the observed values.

	Men	Women	Total
D	$\frac{683}{1301} \cdot 695 = 364,9$	$\frac{683}{1301} \cdot 606 = 318.1$	683
M	$\frac{261}{1301} \cdot 695 = 139,4$	$\frac{261}{1301} \cdot 606 = 121.6$	261
Q	$\frac{303}{1301} \cdot 695 = 161,9$	$\frac{303}{1301} \cdot 606 = 141.1$	303
Undecided	$\frac{54}{1301} \cdot 695 = 28,8$	$\frac{54}{1301} \cdot 606 = 25.2$	54
Total	695	606	1301

The rest of the calculations are shown in table 7.5.

Now that we know the expected values, we can calculate the χ^2-statistic. By definition 7.1

$$\begin{aligned}\chi^2 &= \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \cdots + \frac{(o_n - e_n)^2}{e_n} \\ &= \frac{(386 - 364.9)^2}{364.9} + \frac{(297 - 318.1)^2}{318.1} + \cdots + \frac{(30 - 25.2)^2}{25.2} \\ &= 6.95\,.\end{aligned}$$

The sum consists of 8 fractions, since there are 8 values in the original table (table 7.3).

Degrees of Freedom

We know the χ^2-statistic, but we still need the degrees of freedom to perform a hypothesis test based on the χ^2-distribution.

We determine the degrees of freedom by looking at the number of rows and columns in the original poll (table 7.3). The degrees of freedom are[4]

$$df = (\text{number of rows} - 1) \cdot (\text{number of columns} - 1)\,.$$

In the example we have 4 rows and 2 columns, so

$$df = (4 - 1) \cdot (2 - 1) = 3\,.$$

[4]The degrees of freedom is the number of entries in the table, we need to know to calculate the rest (if we know the totals).

Test and Conclusion

Now that we know the χ^2-statistic and the degrees of freedom, we can perform the test for independence:

1. Formulate a null hypothesis. In our example we have

   $$H_0: \text{ Choice of party and sex are independent.}\,.$$

2. Choose a significance level, e.g. 10%.
3. Calculate the expected values and the χ^2-statistic. In the example $\chi^2 = 6.95$.
4. Determine the degrees of freedom, df, and the critical value C. In our example, $df = 3$, so the critical value can be found by solving the equation[5]

   $$1 - F_{\chi^2}(C; 3) = 0.10$$

   Solving this equation, we find $C = 6.25$.

[5]Remember that we chose a significance level of 10% = 0.10.

5. Since the χ^2-statistic is above the critical value (see figure 7.5), we reject the null hypothesis. Therefore, we conclude that choice of party is *not* independent of sex.

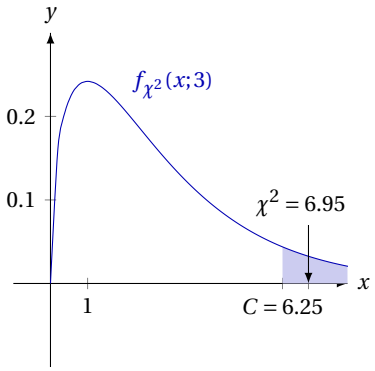


Figure 7.5: The marked area is 0.10 = 10%. Here, $\chi^2 = 6.95$ is above the critical value $C = 6.25$ and we reject the null hypothesis.

7.5 CHOICE OF TEST

We have now looked at two different χ^2-tests: The goodness-of-fit test and the test for independence. When we have a data set and need to choose a test, it is important that we know what both tests are for.

Goodness of Fit

The goodness-of-fit test is used to compare a sample to a distribution, which we already know. This could be election data from previous years, earlier sales numbers of a newspaper, etc.

So, we use the goodness-of-fit test to compare a sample to numbers we already knew *before we took the sample*.

Test of Independence

In a test of indepence we do *not* know previous data. Here we investigate a sample for independence between different categories. It could be whether men and women choose alike, or if young and old people have different political views, etc.

So, we use the test of independence when we are looking for independence of different categories based *solely on the numbers within the sample*.

To sum it up:

- When we compare a sample to data we know *before* we take the sample, then we use a *goodness-of-fit test*.
- When we compare the numbers within the sample itself, we use a *test for independence*.

7.6 CONCLUSIONS OF HYPOTHESIS TESTS

When we analyse a data set in terms of a known distribution or we test for independence, we usually want to find out if there *is* a difference.

But this is not testable. The only thing we can test is if the sample matches an already known distribution (goodness of fit) or an equal distribution (test for independence).

The null hypothesis is always that there is *no* significant difference. Even if this is the opposite of what we are really interested in.

If we are doing a goodness-of-fit test, H_0 is that the measured distribution is the same as the expected distribution—i.e. that there is *no difference*.

If we are testing for independence, H_0 is that our categories are *independent*.

If it then turns out that there is a difference or the variables are not independent, this will be because we reject H_0. Rejecting a hypothesis corresponds to finding a χ^2-statistic above the critical value. It also corresponds to finding a P-value *below* the significance level.

When analysing H_0, we can do it in one of two ways:

1. Use the significance level to determine the critical value. Reject H_0 if χ^2 is larger than the critical value.
2. Determine the P-value. Reject H_0 if P is less than the significance level.

Set Theory

A

Set theory is the basis of a lot of mathematics (e.g. the concept of functions, and probability theory) is built upon. In this section, we therefore present some of the basics of set theory.

A.1 SET

A set in mathematics is a collection of objects. In principle, an "object" can be anything, but we normally restrict ourselves to mathematical objects. In this section, we only look at sets of numbers.

A *set* can be defined in the following way:

Definition A.1

A *set* is a well-defined collection of mutually different objects.

Notice that the objects that make up the set have to be different. We cannot talk about a set made up of four 2s. We can only have one 2 in a set.

Sets are usually denoted by capital letters, e.g. the set S. If we want to show that the set S consists of the numbers 1, 2, 3 and 10, we write

$$S = \{1, 2, 3, 10\} \ . \tag{A.1}$$

So, to show which objects, or *elements*, are in a set, we write the elements in braces, $\{\ldots\}$.[1]

[1] The elements do not have to be ordered, i.e. $\{1,2,3\}$ and $\{3,1,2\}$ are the same set.

The number 3 is an element of the set S. We therefore write

$$3 \in S \ ,$$

which means "3 is an element of the set S".

If we want to say the opposite, we just strike out the symbol. So, because 7 is not an element of S, we write

$$7 \notin S \ .$$

Large Sets

Sometimes sets contain a large amount of numbers. If this is the case, it might be basically unreadable if we list all the numbers as in (A.1). If we want to list all the positive integers between 0 and 100, we write instead

$$H = \{1, 2, 3, \ldots, 100\} .$$

Here, the ellipsis ... shows us that there are numbers missing from the list. But from the pattern we can figure out what numbers are meant to be included in the set.

There are also infinite sets, i.e. sets that do not end at a certain number. The set of positive, odd numbers contains an infinite amount of elements and can be written as

$$O = \{1, 3, 5, 7, \ldots\} .$$

Special Sets

Some sets are used repeatedly in different contexts. These sets are denoted by special symbols:

$\emptyset$ The *empty set*, i.e. the set which contain no elements.

$\mathbb{N}$ The *natural numbers*: The set of all positive integers, $\mathbb{N} = \{1, 2, 3, 4, 5, \ldots\}$. Note that 0 is not included.

$\mathbb{Z}$ The *whole numbers*, or *integers*: The set $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}$.

$\mathbb{Q}$ The *rational numbers*: The set of all numbers that can be written as fractions, e.g. $\frac{1}{7}$ and $-\frac{5}{2}$.

$\mathbb{R}$ The *real numbers*: The set of all the numbers on the number line. A few examples of real numbers are -1, $\frac{1}{6}$, π, e and $\sqrt{2}$.

Here, it is important to note that all of the natural numbers are also integers, i.e. the set $\mathbb{N}$ is a part of the set $\mathbb{Z}$. In the same way, the set of integers $\mathbb{Z}$ is a part of the set of rational numbers $\mathbb{Q}$,[2] and the set of rational numbers is a part of the set of real numbers (see figure A.1).

[2] This is because every integer can be written as a fraction, e.g. $2 = \frac{6}{3}$ and $-4 = -\frac{8}{2}$.

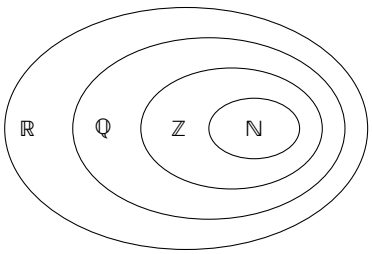


Figure A.1: All of the natural numbers are integers, and all of the integers are rational numbers, etc.

A.2 SET BUILDER

Sometimes it is easier to describe a set by listing properties of the numbers in the set. An example is

$$A : \text{ all of the numbers between 1 and 6 .}$$

We can write this set using *set-builder notation*:

$$A = \{x \in \mathbb{R} \mid 1 < x < 6\} .$$

This expression contains two parts. The part before the vertical line ($x \in \mathbb{R}$) shows what larger set the numbers in our set come from. Here, $x \in \mathbb{R}$ means that we are looking at the set of real numbers, i.e. every possible number. The part after the line is a condition on the numbers in the set—in this case that the numbers must be between 1 and 6.[3]

[3] The notation $1 < x < 6$ is a contraction of $1 < x$ and $x < 6$, i.e. we are looking at numbers both greater than 1 and less than 6.

Other examples of set-builder notation are

$$B = \{x \in \mathbb{Z} \mid 1 < x < 6\}\ , \qquad C = \left\{x \in \mathbb{Q} \,\middle|\, x^2 < 9\right\}\ .$$

Here, B is the set of all integers between 1 and 6, so we might just as well write

$$B = \{2,3,4,5\}\ .$$

C is the set of all fractions, whose squares are less than 9. This set cannot be written as a simple list, since there are infinitely many numbers in the set.

A.3 INTERVALS

The set A from the previous section is an example of an *interval*. An interval is a set containing every real number between two given values, e.g. "the set of real numbers between 1 and 6" or "the set of real numbers from -5 to 80, including 80". Sets containing every number greather than or less than some given value, are also called intervals.

In mathematics, we often need to talk about intervals, and therefore we have invented a notation for intervals. The set A can be written in this way:

$$A =]1;6[\ .$$

This means that the set A is made up of every number between 1 and 6. Because the brackets point away from the numbers, neither 1 nor 6 is included in the interval (see figure A.2).

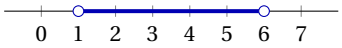


Figure A.2: The interval $]1;6[$. The empty circles at 1 and 6 show that these numbers are not included in the interval.

If, instead, we want the interval from 1 to 6 including 1 and 6, we write

$$D = [1;6]\ .$$

In set-builder notation D can also be written as $D = \{x \in \mathbb{R} \mid 1 \leq x \leq 6\}$.

Some other examples are (see figure A.3).

$$]-3;2] = \{x \in \mathbb{R} \mid -3 < x \leq 2\}$$
$$\left[-4;\tfrac{1}{2}\right[= \left\{x \in \mathbb{R} \,\middle|\, -4 \leq x < \tfrac{1}{2}\right\}\ .$$

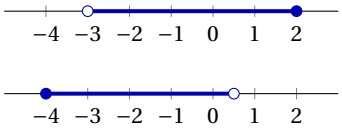


Figure A.3: The intervals $]-3;2]$ and $\left[-4;\frac{1}{2}\right[$.

If we want to look at the interval containing every number greater than 3, we use the symbol ∞ (inifinity):

$$E =]3;\infty[\ .$$

So, the set E contains every number greater than 3. If, instead, we want to talk about all the numbers less than or equal to 5, we write

$$F =]-\infty;5]\ .$$

Notice that when we use the symbol ∞, the bracket has to point away from the symbol (this is because ∞ is not a number, but a symbol we use to show that the interval does not end in that direction).

If we let our interval be infinite in both directions, we get an interval containing every real number, i.e.

$$\mathbb{R} =]-\infty;\infty[\ .$$

A.4 COMBINING SETS

If we have two sets, A and B, we can form new sets in different ways. We might look at all the numbers which are in both A and B, or the numbers, which are in A, but not in B.

The following definitions list some of the ways in which we can build new sets.

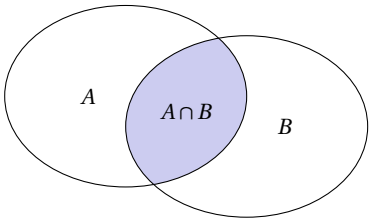


Figure A.4: The intersection $A \cap B$.

Definition A.2

The *intersection* of two sets A and B contains the numbers, which are both in A and in B. The intersection of A and B is denoted by $A \cap B$.

Example A.3
If $A = \{1,2,3,4,5\}$ and $B = \{2,4,6,8\}$, then

$$A \cap B = \{2,4\} .$$

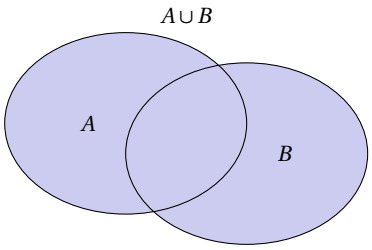


Figure A.5: The union $A \cup B$.

Definition A.4

The *union* of two sets A and B contains all of the numbers that are in either A or B (or both). The union of A and B is denoted by $A \cup B$.

Example A.5
If $A = \{1,2,3,4,5\}$ and $B = \{2,4,6,8\}$, then

$$A \cup B = \{1,2,3,4,5,6,8\} .$$

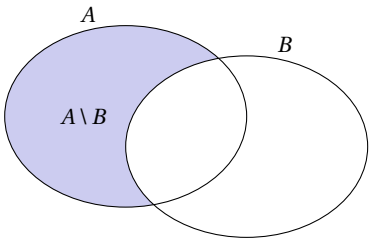


Figure A.6: The difference $A \setminus B$.

Definition A.6

The *difference* between two sets A and B contains the numbers, which are in A but not in B. The difference between A and B is denoted by $A \setminus B$.

Example A.7
If $A = \{1,2,3,4,5\}$ and $B = \{2,4,6,8\}$, then

$$A \setminus B = \{1,3,5\} .$$

and

$$B \setminus A = \{6,8\} .$$

So, when we are looking at differences between sets, the order matters.

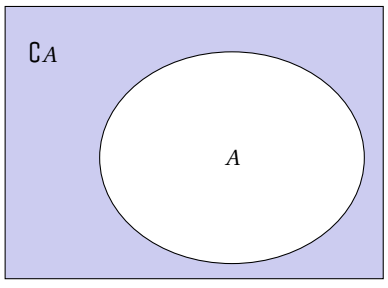


Figure A.7: The complement $\complement A$.

Lastly, we define the *complement*, which is the set of all elements that are not in a given set. This is only well-defined if we first define a *universal set*, which contains all the numbers, we allow a set to contain in the present context.[4]

Definition A.8

Let U be the *universal set*. The *complement* $\complement A$ of the set A contains all of the elements in U that are not in A, i.e. $\complement A = U \setminus A$.

[4]The universal set can be all of the real numbers, i.e. $\mathbb{R}$, but it could also be the natural numbers $\mathbb{N}$, or some limited set, e.g. $\left\{-2, 0, \frac{1}{3}, 7\right\}$.

A.5 RELATIONS BETWEEN SETS

Sometimes we need to compare different sets. Then we need to know what it actually means for two sets to be equal.

Definition A.9

Two sets A and B are equal if they contain the exact same elements. This is denoted by $A = B$.

If every element in A is an element in B, but all of the elements in B are not necessarily in A, we call A a subset of B.

Definition A.10

The set A is a *subset* of the set B if every element in A is also an element in B. This is denoted by $A \subseteq B$.

Example A.11
The set $A = \{-1, 1\}$ is a subset of $B = \{-2, -1, 0, 1, 2, 3, 4\}$, i.e.

$$\{-1, 1\} \subseteq \{-2, -1, 0, 1, 2, 3, 4\} .$$

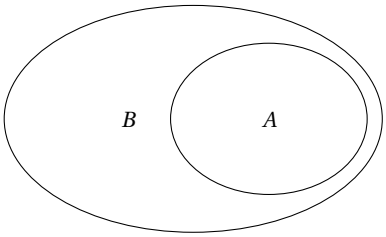


Figure A.8: A is a subset of B, $A \subseteq B$.

Previously we argued that every natural number is an integer, and every integer is a rational number,, etc. We can write this using the idea of subsets as

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} .$$

This is shown in figure A.1.

If two sets have common elements, they are called *disjoint*.

Definition A.12

Two sets A and B are called *disjoint* if no element in A is an element in B (and no element in B is an element in A), i.e. when $A \cap B = \emptyset$.

Example A.13
The sets $A = \{1, 2, 3\}$ and $B = \{-1, 0, 7\}$ are disjoint.

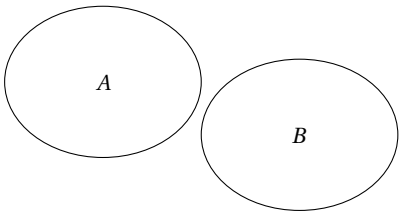


Figure A.9: A and B are disjoint.

More Derivatives

B

In this section, we show how to find the derivatives of $\ln(x)$, e^x, a^x and x^n. To find the first of these, we use the three-step method—the rest are found using the rules in sections 1.3 and 1.4.

This will prove the last of the assertions in table 1.1.

Theorem B.1

If $f(x) = \ln(x)$, the derivative is $f'(x) = \frac{1}{x}$.

Proof

Here, we use the three-step method. First, we find[1]

$$\Delta f = \ln(x+\Delta x) - \ln(x) = \ln\left(\frac{x+\Delta x}{x}\right) = \ln\left(1+\frac{\Delta x}{x}\right).$$

[1] In this calculation, we use the rule

$$\ln(a) - \ln(b) = \ln\left(\frac{a}{b}\right).$$

Next, we look at

$$\frac{\Delta f}{\Delta x} = \frac{\ln\left(1+\frac{\Delta x}{x}\right)}{\Delta x} = \frac{1}{\Delta x}\cdot\ln\left(1+\frac{\Delta x}{x}\right). \tag{B.1}$$

We cannot seem to simplify this further.

Now, we need to let $\Delta x \to 0$, but the expression (B.1) is too complicated to see what that gives us. We therefore use a little trick: We introduce a new variable t, which is equal to $\frac{\Delta x}{x}$. Letting $\Delta x \to 0$ corresponds to letting $t \to 0$.

(B.1) can now be rewritten as

$$\frac{\Delta f}{\Delta x} = \frac{1}{xt}\cdot\ln(1+t),$$

which then corresponds to

$$\frac{\Delta f}{\Delta x} = \frac{1}{x}\cdot\frac{1}{t}\cdot\ln(1+t) = \frac{1}{x}\cdot\ln\left((1+t)^{\frac{1}{t}}\right). \tag{B.2}$$

It is well-know that[2]

$$(1+t)^{\frac{1}{t}} \to e \quad \text{når} \quad t \to 0. \tag{B.3}$$

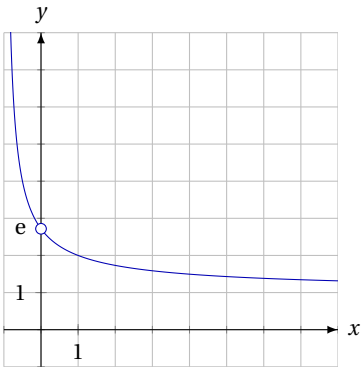


Figure B.1: The graph of $(1+t)^{\frac{1}{t}}$.

Actually, this is sometimes used as the definition of the number e. We are not going to prove the result in (B.3), but that it is correct may be inferred from the graph of $(1+t)^{\frac{1}{t}}$ shown in figure B.1.

Now, letting $\Delta x \to 0$ is the same as letting $t \to 0$ in (B.2), and using the result from (B.3), we get

$$f'(x) = \frac{1}{x} \cdot \ln(e) = \frac{1}{x}\,.$$ ■

Theorem B.2

When $f(x) = e^x$, the derivative is $f'(x) = \mathrm{e}^x$.

Proof

Since e^x is the inverse of $\ln(x)$, we have the following equation:

$$\ln(\mathrm{e}^x) = x\,. \tag{B.4}$$

If we differentiate both sides of this equation, it will still hold.

On the left hand side, we need to differentiate a composite function, and we get[2]

$$\left(\ln(\mathrm{e}^x)\right)' = \frac{1}{\mathrm{e}^x} \cdot (\mathrm{e}^x)' = \frac{1}{\mathrm{e}^x} \cdot f'(x)\,.$$

On the right hand side we get

$$(x)' = 1\,.$$

Since the left hand side is equal to the right hand side, we have

$$\frac{1}{\mathrm{e}^x} \cdot f'(x) = 1 \qquad \Leftrightarrow \qquad f'(x) = \mathrm{e}^x\,.$$ ■

[2] Here, it is important to remember that we do not yet know the derivative of e^x. Therefore, we must write $(\mathrm{e}^x)'$, which is the same as $f'(x)$.

Theorem B.3

If $f(x) = a^x$, where $a > 0$, then $f'(x) = \ln(a) \cdot a^x$.

Proof

We can rewrite the function f as

$$f(x) = a^x = \left(\mathrm{e}^{\ln(a)}\right)^x = \mathrm{e}^{\ln(a)\cdot x}\,.$$

This is a composite function, and its derivative is

$$f'(x) = \mathrm{e}^{\ln(a)\cdot x} \cdot \ln(a) = a^x \cdot \ln(a) = \ln(a) \cdot a^x\,.$$ ■

Theorem B.4

If $f(x) = x^n$, the derivative is $f'(x) = nx^{n-1}$.

Proof
First, we rewrite the formula for $f(x)$:

$$f(x) = x^n = \mathrm{e}^{\ln(x^n)} = \mathrm{e}^{n\cdot\ln(x)} .$$

So, f can be written as a composite function, where the outer function is

$$p(q) = \mathrm{e}^q ,$$

and the inner function is

$$q(x) = n\cdot\ln(x) ,$$

where n is a constant.

If we differentiate the outer function, we get

$$p'(q) = \mathrm{e}^q .$$

The inner function yields

$$q'(x) = n\cdot\frac{1}{x} .$$

So,

$$\begin{aligned} f'(x) &= p'(q)\cdot q'(x) = \mathrm{e}^q\cdot n\cdot\frac{1}{x} \\ &= \mathrm{e}^{n\cdot\ln(x)}\cdot n\cdot\frac{1}{x} = x^n\cdot n\cdot\frac{1}{x} \\ &= n\cdot x^{n-1} . \end{aligned}$$

■

Limits and Differentiability

C

In chapter 1, we defined the derivative $f'(x)$ as the value of the fraction $\frac{\Delta f}{\Delta x}$ as Δx approaches 0. What we actually mean by "approaches 0" was never really investigated. We just rewrote the fraction, so it was possible to let $\Delta x = 0$ and get a useful result.

If we want to do this in a mathematically meaningful way, we need to define, what we mean by such vague terms as "approaches" and "close to". To do that, we need to introduce the concept of a *limit*.

C.1 LIMITS

To describe the concept of a limit, we first look at the function

$$f(x) = \frac{x^2 - 1}{x - 1}\,.$$

This function is undefined for $x = 1$, since

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

which has no mathematical meaning.

If we draw the graph of f, we get figure C.1. Here, we see that even though f is undefined for $x = 1$, using the graph we might say something about, what $f(1)$ should be, if this value were defined.

If we calculate $f(x)$ for values of x that are "close to" 1, we get table C.1. Looking at figure C.1 and table C.1, it seems reasonable to suggest that the closer x gets to 1, the closer $f(x)$ gets to 2.

So, although $f(1)$ is undefined, $f(x)$ approaches a fixed number, when x approaches 1. We therefore say that $f(x)$ has a *limit for x approaching* 1. The value of this limit is 2. We write

$$\lim_{x \to 1} f(x) = 2\,.$$

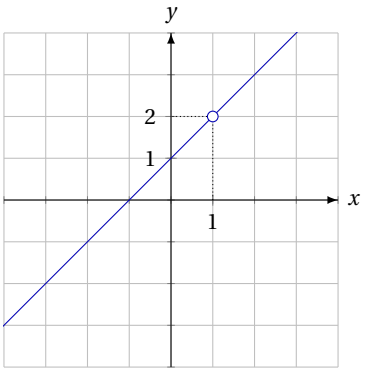


Figure C.1: The graph of $f(x) = \frac{x^2-1}{x-1}$.

Table C.1: Function values of $f(x) = \frac{x^2-1}{x-1}$.

x	$f(x)$
0.5	1.5
0.9	1.9
0.99	1.99
1	undefined
1.01	2.01
1.1	2.1
1.5	2.5

Mathematically, we define a limit by investigating if the function value approaches some fixed number, when the independent variable approaches some (other) fixed number (like $x = 1$ above).

We have the following definition of the limit of a function:

Definition C.1

Let f be a function, and let x_0 and L be numbers. If for any small number ε, we can ensure that $f(x)$ is closer to L than ε as long as x is closer to x_0 than some other number δ, then we call L the *limit of* $f(x)$ *for* x *approaching* x_0, and we write

$$\lim_{x \to x_0} f(x) = L\,.$$

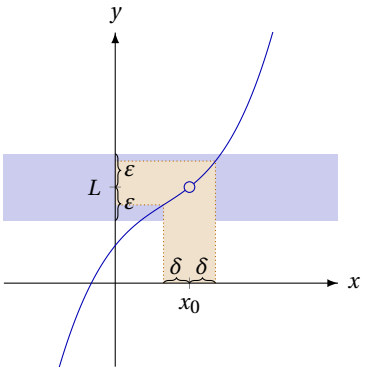


Figure C.2: The function value is closer to L than ε if we let x be closer than δ to x_0.

Figure C.2 shows the meaning of this definition: The function value $f(x)$ is closer to L than a given distance ε, as long as x is closer than δ to x_0.

Example C.2

For the function $f(x) = \frac{x^2-9}{x-3}$, we have

$$\lim_{x \to 3} f(x) = 6\,.$$

This means that $f(x)$ can get as close as we want to 6, as long as x is close enough to 3.

E.g., if we want $f(x)$ to be closer to 6 than $\varepsilon = 0.1$ (i.e. $f(x)$ is between 5.9 and 6.1), x has to be closer to 3 than $\delta = 0.07$—we might choose $x = 3.05$:

$$f(3.05) = \frac{3.05^2 - 9}{3.05 - 3} = \frac{0.3025}{0.05} = 6.05\,,$$

which is closer to 6 than 0.1.

If we want to get even closer, we might demand that $f(x)$ must be between 5.99 and 6.01. Here, we might let $x = 2.995$, and we get

$$f(2.995) = \frac{2.995^2 - 9}{2.995 - 3} = \frac{-0.029975}{-0.005} = 5.995\,.$$

Again we are as close as we want to the number 6.

We therefore conclude[1] that

$$\lim_{x \to 3} \frac{x^2 - 9}{x - 3} = 6\,.$$

The concept of a limit makes perfect sense, when we investigate how functions behave at undefined values of the independent variable. But what if we want to investigate a limit at a value of x, where the function *is* defined?

Example C.3

What is $\lim_{x \to 5} x^2 + 3$?

The expression $x^2 + 3$ is defined for $x = 5$, where the value is

$$5^2 + 3 = 28\,.$$

[1] Two examples are not really enough to conclude that the limit is 6. We would actually need to find out how to choose δ, when ε is given.

If we let x approach 5, the value of x^2+3 will approach 28, and we therefore get

$$\lim_{x\to 5} x^2 + 3 = 28\,.$$

Sometimes we can just calculate the value of the expression for our value of x.

We can find the limit of $f(x)$ for $x \to x_0$ in some cases by calculating $f(x_0)$. But this is not always the case, even if the function might be defined for $x = x_0$.

Example C.4
Here, we look at the function

$$f(x) = \begin{cases} x+1 & \text{for } x < 2 \\ 4-x & \text{for } x \geq 2 \end{cases}.$$

The function f is defined to be equal to $x+1$, as long as $x < 2$, after that it is equal to $4-x$. Such a function is called a "piecewise linear function". The graph of f is shown in figure C.3.

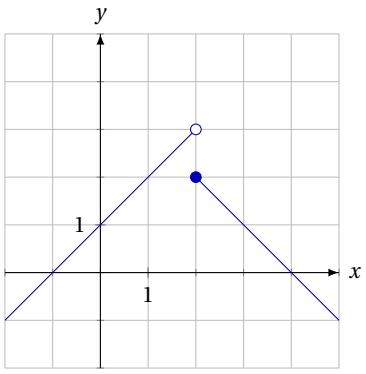


Figure C.3: The graph of the piecewise linear function in example C.4.

The function value at $x = 2$ is

$$f(2) = 4 - 2 = 2\,,$$

but what is the limit as $x \to 2$?

When $x < 2$, then $f(x) = x+1$, i.e. $f(x)$ gets closer and closer to $2+1$, the closer x gets to 2. If we examine the limit by letting x get closer to 2 from below, we find the value 3.

If we investigate the limit of $f(x)$ by letting x approach 2 from above, we follow the graph of $4-x$, and here the value gets closer to 2, when x approaches 2.

So, we get two different answers depending on which way, we approach $x = 2$. Therefore, we must conclude that the limit $\lim_{x\to 2} f(x)$ *does not exist*—even though the function is defined for $x = 2$.

Example C.5
Previously, we looked at the function $f(x) = \frac{x^2-1}{x-1}$ and concluded that

$$\lim_{x\to 1} f(x) = 2\,.$$

In principle, we might argue that we cannot know this from figure C.1 and table C.1, since it is impossible solely from the figure and the table to see if the true limit is e.g. 2.00000326 and not exactly 2.

However, it turns out that $x^2 - 1$ can be rewritten as $(x+1)(x-1)$, and therefore

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

as long as $x \neq 1$.

But, since the the definition of the limit does not depend on how the function behaves at $x = 1$, but only when x is close to 1, we have

$$\lim_{x\to 1} \frac{x^2-1}{x-1} = \lim_{x\to 1} x+1\,.$$

So, here we only need to find out, which number $x+1$ approaches, when x approaches 1. This number is exactly 2.

Therefore

$$\lim_{x\to 1} \frac{x^2-1}{x-1} = 2\,.$$

If it is possible to simplify the formula of the function we are investigating, it is easier to investigate its limits.

C.2 CONTINUITY

Most of the functions, we investigate, have graphs that are connected. A function, whose graph is connected, is called *continuous*. We can define the concept of continuity using limits.

In example C.4, we looked at a function, whose graph was *not* connected (see figure C.3). In the example, we showed that this function has no limit at the point where the graph "jumps".

But in example C.3, we investigated a limit which was equal to the function value. The graph of the function in question is connected, because when we approach a value of x, the function value automatically approaches $f(x)$—from above as well as from below—and $(x, f(x))$ is a point on the graph.

We therefore define continuity the following way:

Definition C.6

A function f is called *continuous* in an interval $]a;b[$ if for all $x_0 \in]a;b[$

$$\lim_{x\to x_0} f(x) = f(x_0)\,.$$

Example C.7
The function

$$f(x) = x^2 + 4\,,$$

is continuous for all $x \in \mathbb{R}$.

If we choose e.g. $x_0 = 3$, we find

$$\lim_{x\to 3} f(x) = \lim_{x\to 3} x^2 + 4 = 3^2 + 4 = f(3)\,,$$

and we can do this calculation for any value of x_0—not just 3.

So, $f(x)$ is continuous.

Example C.8
The function

$$f(x) = \begin{cases} x & \text{for } x \neq 3 \\ 4 & \text{for } x = 3 \end{cases}$$

is not continuous.

We see from the graph (figure C.4) that

$$\lim_{x \to 3} f(x) = 3\,,$$

but

$$f(3) = 4\,.$$

So, $\lim_{x\to 3} f(x) \neq f(3)$, and the function is not continuous.

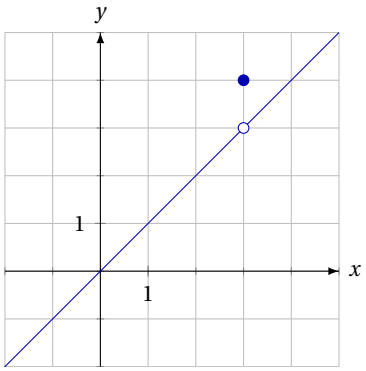


Figure C.4: This function is not continuous.

C.3 DIFFERENTIABILITY

Using limits, we can define the derivative of a function in a more precise way than we did in chapter 1. A function, where the derivative exists for all x in an interval, is called *differentiable* on this interval.

Definition C.9

Let f be a function defined on the interval $]a;b[$. If the limit

$$\lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{C.1}$$

exists for all $x_0 \in]a;b[$, the function is called *differentiable* on the interval $]a;b[$.

The limit (C.1) is called the *derivative at* x_0 and denoted by $f'(x_0)$.

There is nothing new in this definition. But the definition in chapter 1 does not mention limits, so the definition is not as exact as this one. But the content is the same.

Lastly, it is worth mentioning that if a function is differentiable, it is also continuous. So, if a function is differentiable, its graph is connected. The opposite does not apply—functions exist, whose graphs are connected, but that are not differentiable.

Differentiability roughly corresponds to the function having a "smooth" curve. The graph is not allowed to have "kinks". In figure C.5, we see the graph of a function that is continuous, but not differentiable.

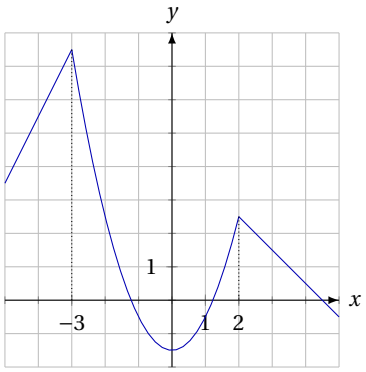


Figure C.5: This function is not differentiable at $x = -3$ or at $x = 2$, but it is continuous everywhere.

Bibliography

[1] Danmarks Statistik. *INDKP102: Indkomst i alt efter region, enhed, køn og indkomstinterval*. URL: http://www.statistikbanken.dk/10331 (visited on 08/04/2015).

[2] Demetrios Kanoussis. *Euler's Number e*. May 9, 2015. URL: http://www.goldenratiopublications.com/eulers-number-e/ (visited on 08/03/2015).

[3] Victor J. Katz. *A History of Mathematics. An Introduction*. 2nd ed. Addison-Wesley Educational Publishers, Inc., 1998.

[4] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. "Log-normal Distributions across the Sciences: Keys and Clues". In: *BioScience* 51.5 (2001), pp. 341–352.

[5] R.L. Plackett. "Karl Pearson and the Chi-Squared Test". In: *International Statistical Review/Revue Internationale de Statistique* 51.1 (Apr. 1983), pp. 59–72.

Index