Basic Mathematical Notation

Holger Sennhenn-Reulen (`uncertaintree.github.io`)
NW-FVA

September 27, 2022

Contents

1 Function, (Co-)Domain, Inverse Function, and Parameters

2 Function

We need this all the time: Any method transporting observed measures such that the essential information becomes quantified is / can / should be expressed in the form of a mathematical function.

Definition (Function) A mathematical **function** is a special relationship where each input has a single output. A function named f is usually denoted $f(x) = y$, where x is the **input value** and y is the **function value**. $f(x)$ maps all the values of the **domain** X (the space or set of possible input values x) to the **co-domain** Y (the space or set of possible function values y), which is denoted by:

$$f : X \to Y,\ x \mapsto y = f(x).$$

Read: ' f is a function from X to Y. f maps x to y. '

Example (Quadratic function) The relationship $f(x) = x^2$, (Read ' f of x is x squared ', or' ... x to the power of 2 '), $x \in \mathbb{R}^+$ (Read' for x from the positive real numbers '), is a function:

$$f : \mathbb{R}^+ \to \mathbb{R}^+,\ x \mapsto x^2 = f(x),$$

because each input x has a single output x^2:

$$f(1) = 1, \quad f(3) = 9, \quad f\left(\sqrt{z^2}\right) = z^2.$$

f is called the *quadratic function*.

- The *square-root function* $g(x) = \sqrt{x} = x^{\frac{1}{2}}$ is the **inverse function** of $f(x) = x^2$.
- The common notation for this is $g(x) = f^{-1}(x)$.
- The relationship between $f(x)$ and $g(x)$ is visualized by interchanging the axes.

The following figure illustrates the quadratic ('square') and the square-root functions.

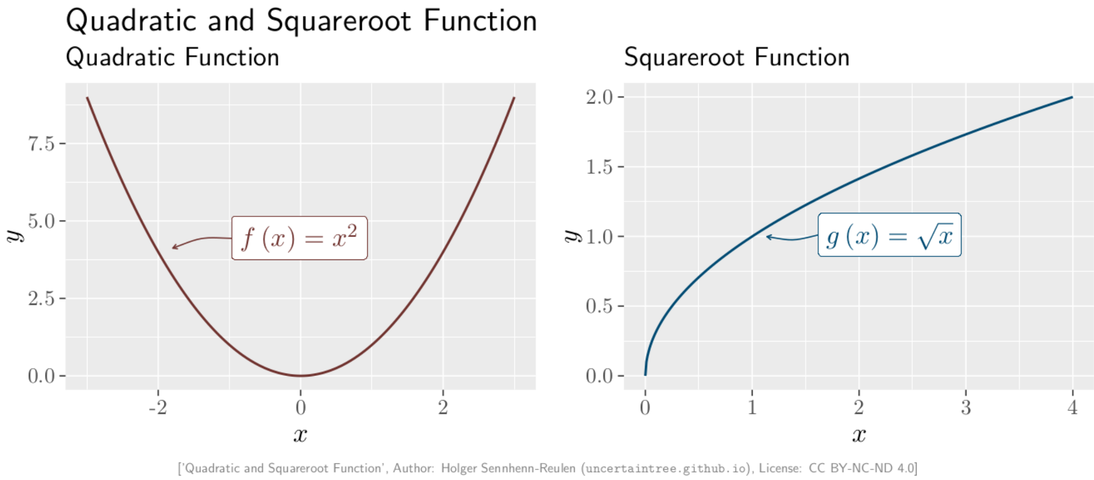

Figure 1: Quadratic and squareroot functions.

3 Function with parameters

The relationship $f(x) = \frac{x}{2}$ *(Read 'f of x is x divided by 2'}) is a function, because each input x has a single output* $\frac{x}{2}$*:*

$$f(1) = 0.5, \quad f(12) = 6.$$

$f(x) = \frac{x}{2}$ *is an example for a simple* linear function}.

Definition (Linear function) A **linear function** is a function of the form:

$$f(x \,|\, a, b) = a + b \cdot x,$$

(' f of x given a and b is a plus b times x ', or' f of x depending on a and b is a plus b times x ') where a and b are *parameters* (constant and given/fixed), often real numbers. a is frequently referred to as the intercept, and b as the slope. In the above example $f(x) = \frac{x}{2}$, we have $a = 0$ and $b = 0.5$.
The following figure illustrates a linear function.
The linear function is a function of a set of variables and parameters that does not contain any powers or cross-products of the quantities. The graph of such a function of one variable x and two parameters a, and b, is a non-vertical line.
Linear functions are frequently applied in statistical models that aim to specify changes in the expected value of a dependent random variable y by changes in an influential variable x (Definitions for *expected value* and *random variable* are given in the Basic Statistics Documents): The value of the linear function then gives the expected value of y conditional on a specific value of x, ie. the goal is to get an expected value for a random variable that varies according to influential characteristics. This type of quantifying the expected value of a dependent variable conditional on an influential variable is called a *linear regression model.*

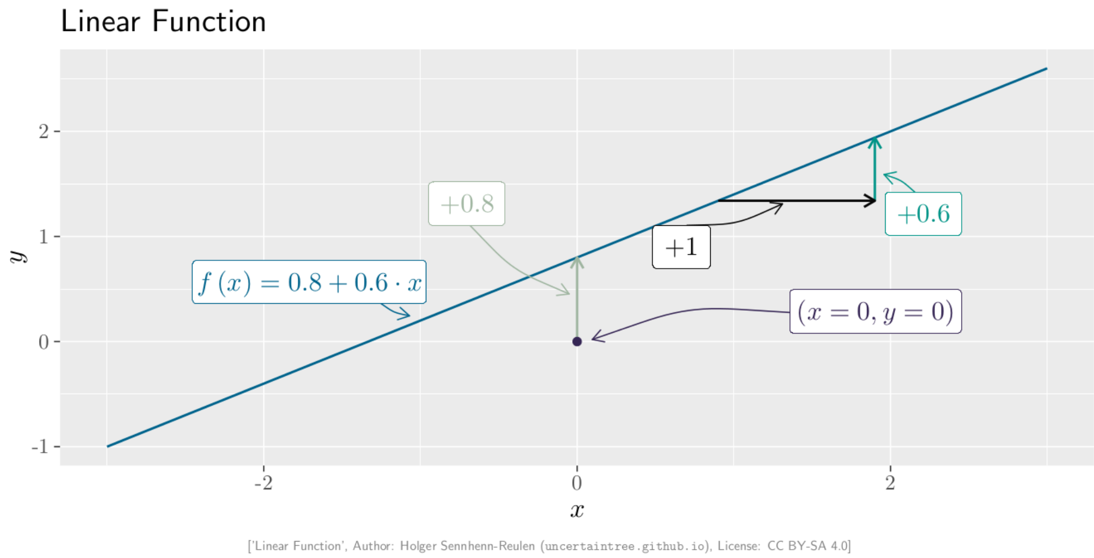

Figure 2: Linear function.

Somme attributes and comments on notation of functions with parameters:

- There is no consensus on the notation of the dependence of a function on parameters: $\rightarrow$ Alternatives $f_{a,b}(x)$, $f(x; a, b)$, and $f(a \mid a, b)$ all mean the same.
- If x is defined on the real values $\mathbb{R}$, than f also maps on the real values for any values of a and b.
- If x is defined on the set of integers $\mathbb{Z}$ (values $0, -1, 1, -2, 2, \ldots$), than f only maps on $\mathbb{Z}$ if a and b are also defined on $\mathbb{Z}$.
- Considering the domain of functions will be important to suitably handle discrete and continuously scaled random variables.
- Often, a and b are notationally simply omitted, e.g. $f(x) = a + b \cdot x$.

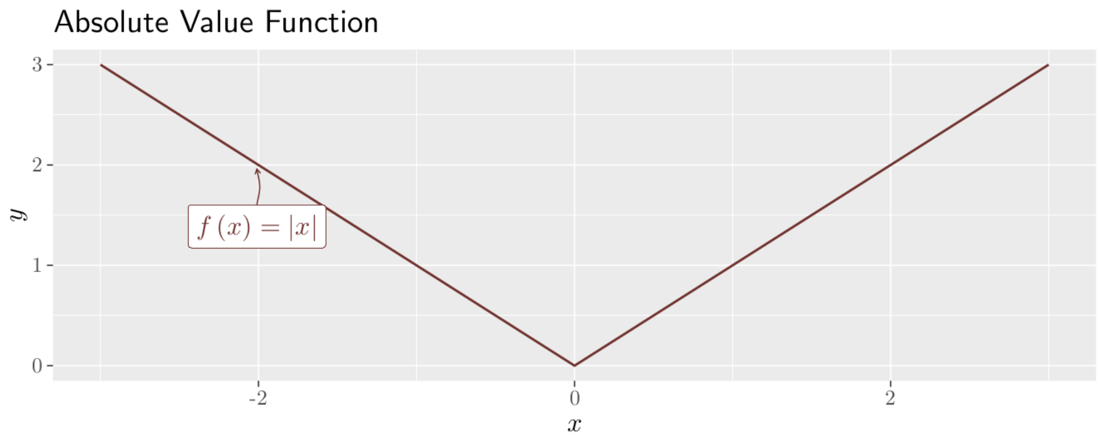

Figure 3: Absolute value function.

3.1 Absolute value function

Definition (Absolute value function) One function that plays a central in non-parametric statistics - but not only there - is the *absolute value function*:

$$f(x) = |x| = \begin{cases} -x, & \text{if } x < 0, \\ x, & \text{else.} \end{cases}$$

- The absolute value function multiplies negative values with -1, and keeps positive values unchanged.
- It therefore reveals the magnitude of an estimated regression effect, for example.

The following figure illustrates the absolute value function.

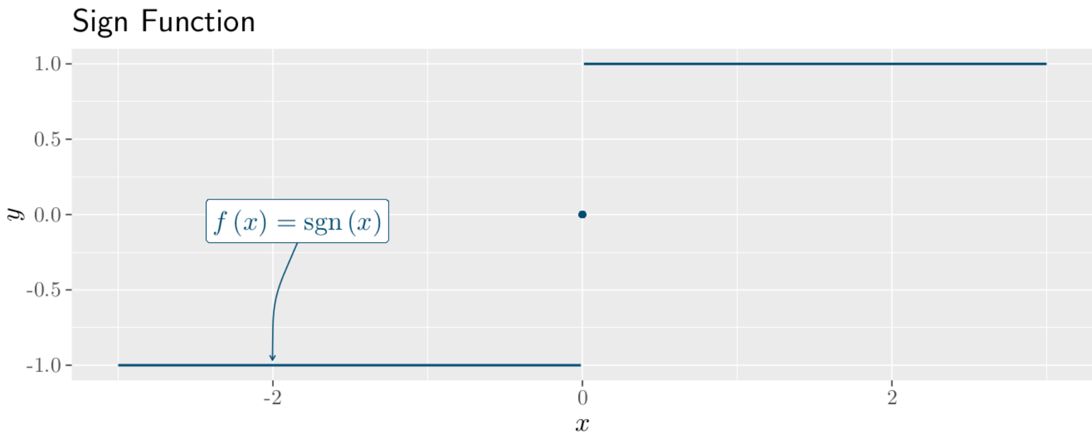

Figure 4: Sign function.

4 Sign function

Definition (Sign function) The *sign function* is closely related to the absolute value function and defined as:

$$\operatorname{sgn}(x) := f(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{else.} \end{cases}$$

Any real number can be expressed as the product of its absolute value and its sign function, ie. elaborate the function $f(x) = \operatorname{sgn}(x) \cdot |x|$.
The following figure illustrates the sign function.

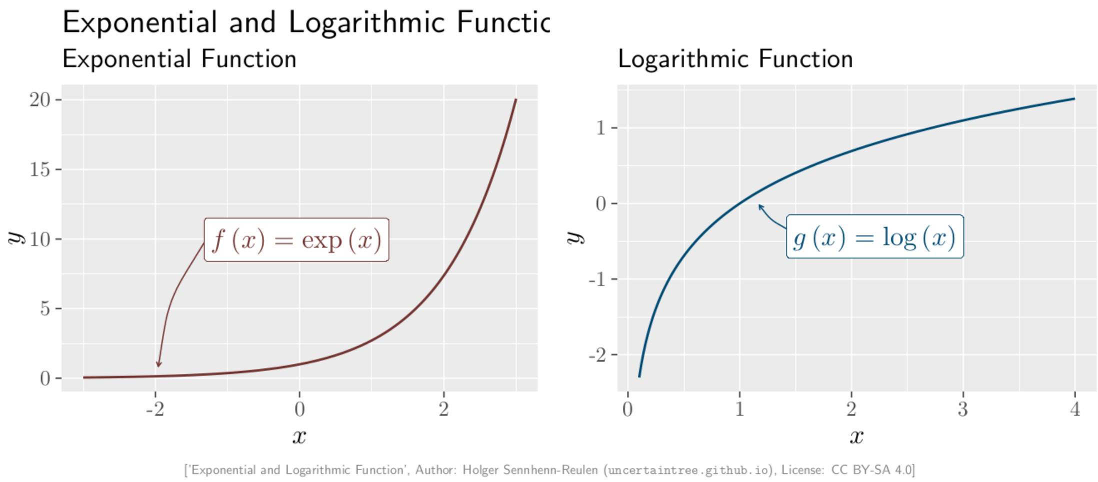

Figure 5: Exponential and logarithmic function.

5 Exponential function and logarithmic function

Definition (Exponential function and logarithmic function) The *exponential function* is a function from $\mathbb{R}$ to the positive real numbers $\mathbb{R}_+$:

$$\exp(x) = e^x.$$

Calculation rules for the exponential function

$$\exp(x) \cdot \exp(y) = \exp(x + y), \exp(1) = e^1 = e \approx 2.7183$$

These rules also hold for any other basis $a \in \mathbb{R}_+$:

$$a^x = \exp_a(x).$$

The *logarithmic function* is then defined as the inverse function of the exponential function:

$$g(x) = \log(x) = \exp^{-1}(x).$$

The inverse function of the exponential function with basis e is also called *natural logarithm* and denoted with $\ln(x)$.
The following figure illustrates the exponential and logarithmic functions.

Example (Application of the logarithmic function in order to achieve symmetry of ratios) Logarithmic transformations are often applied in order to achieve symmetrical probability density function for samples that empirically show a non-symmetrical, left steep and right skewed density. Logarithms are also important when one wants to obtain symmetry of factors around the value 1, the neutral element of multiplication and division (we can multiply or divide anything by 1, it will always stay the same!). Basal area of a certain stand is acquired for two consecutive 5 year periods. We get value of 10 m^2 at the beginning of the first period, a value of 20 m^2 at the time between both periods, and a value of again 10 m^2 at the end of the second period. Absolute increments are equal in both periods:

$$|10 - 20| = 10 = |20 - 10|.$$

If we now compare increment ratios, we get an asymmetry with respect to the value 1, which is the neutral element of multiplication:

$$|\frac{20}{10}| = 2 \quad \neq \quad 0.5 = |\frac{10}{20}|.$$

Taking the log of ratios results in a symmetrical measure around the value~0, which is the neutral element for sums and differences:

$$|\log\left(\frac{20}{10}\right)| = |\log\left(\frac{10}{20}\right)| \approx 0.693.$$

Calculation rules for the logarithmic function The results of the previous application are a direct consequence of the *calculation rules for the logarithmic function*:

$$\log\left(a^b\right) = b \cdot \log(a)$$

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$\log(1) = 0$$

We get:

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

which is equal to $\log(b)$ except of the opposite sign.

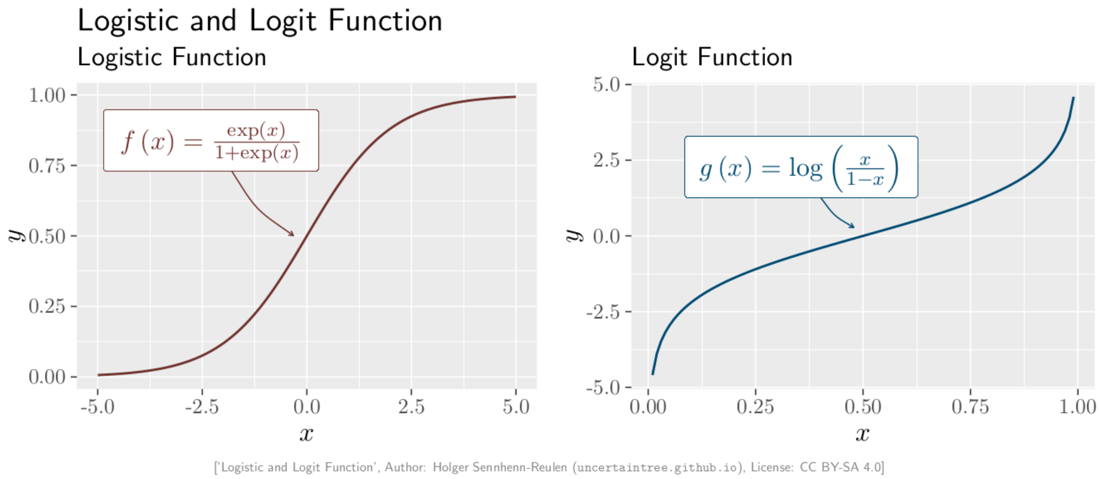

Figure 6: Logistic and logit functions.

6 Logistic and logit function

Definition (Logistic function) The *logistic function* - this function gets it's from the logistic distribution which probability distribution function is equal to the logistic function - is a function from $\mathbb{R}$ to the real numbers between 0 and 1, $\mathbb{R}_{[0,1]}$.
It is defined as:

$$f(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

Definition (Logit function) The inverse function of the logistic function is the *logit function*:

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

The following figure illustrates the logistic and logit functions.

7 Sum

Definition (Sum) A *sum* is denoted using the sum-sign $\sum$ with index i, two values – often 1 and n – that designate the lower and upper limit of the integer values that the index is taking on, and a quantity x_i that is indexed:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_{n-1} + x_n.$$

Example (Arithmetic mean) The *arithmetic mean* – often simply denoted as *'mean'* – is a function that maps observed values x_i, $i = 1, \ldots, n$ (or equivalently denoted by $x_1, x_2, \ldots, x_n$) to their central point:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

- The arithmetic mean is often denoted by $\bar{x}$.
- It is an example for the concept called *statistic* – which is defined as a single measure of some attribute of a sample –, here for the central point in the observed data
  $x_1, x_2, \ldots, x_n$.

Example (BHD) We want to calculate the mean of BHD (in meters m) in `dA`[1]. The first 30 observed values are given as:

$$x_1 = 0.277, x_2 = 0.283, \ldots, x_{30} = 0.133.$$

```
> head(dA$d/100, n = 30)
 [1] 0.277 0.283 0.261 0.266 0.209 0.191 0.187 0.182 0.176 0.161 0.127
[12] 0.116 0.113 0.095 0.079 0.076 0.061 0.048 0.045 0.044 0.328 0.273
[23] 0.281 0.270 0.266 0.252 0.248 0.199 0.175 0.133
```

Their sum and arithmetic mean is:

$$\sum_{i=1}^{1678} x_i = 225.663, \qquad \bar{x} = \frac{1}{1678} \cdot \sum_{i=1}^{1678} x_i = \frac{225.663}{1678} = 0.117.$$

```
> sum(dA$d/100, n = 30); mean(dA$d/100, n = 30)
[1] 225.663
[1] 0.1166049
```

[1]This dataset was not introduced before, I think it is data set `spati` provided by the `lmfor` R add-on package]

Calculation rules for sums An important **calculation rule for sums** is that constant factors for each addend, ie. the terms that are independent of the summation index, may also be written as a factor of the sum:

$$\sum_{i=1}^{n} a \cdot x_i = a \cdot \sum_{i=1}^{n} x_i.$$

Constant addends have to be multiplied by the length of the index in order to be pulled out of the sum:

$$\sum_{i=1}^{n} (x_i + a) = n \cdot a + \sum_{i=1}^{n} x_i.$$

7.1 Indicator function

paragraphDefinition (Indicator function)
The *indicator function* is an auxiliary construction that helps in many data situations:

$$\mathrm{I}_{\text{condition}}\left(x\right)=\begin{cases}1, & \text{if } x \text{ meets the condition,}\\ 0, & \text{else.}\end{cases}$$

paragraphExample (Histogram)
A *histogram* is a graphical representation of the distribution of observed values from a continuous variable.
The underlying calculations are summarized as the table of an artificial partitioning of the originally continuous scale into ordinally scaled, mutually exclusive sub-intervals, as defined by lower and upper boundaries $(b_{\text{low}}, b_{\text{up}}]$.
The value of the histogram *for one sub-interval* is defined[2] as the sum across all indicator functions for the observed data
$x_1, \ldots, x_n$:

$$\operatorname{hist}\left(x_1,\ldots,x_n|\left(b_{\text{low}},b_{\text{up}}\right]\right)=\sum_{i=1}^{n}\mathrm{I}_{\text{condition histogram}}\left(x_i\right)$$

with

$$\text{condition histogram} = x > b_{\text{low}} \text{ and } x \leq b_{\text{up}}$$

And so:

$$\operatorname{hist}\left(x_1,\ldots,x_n|\left(b_{\text{low}},b_{\text{up}}\right]\right)=\sum_{i=1}^{n}\begin{cases}1, & \text{if } x_i > b_{\text{low}} \text{ and } x_i \leq b_{\text{up}},\\ 0, & \text{else.}\end{cases}$$

Example (Histogram as an example for indicator functions) For the BHD values of dataset `dA`, this leads to following results:

```
> x <- dA$d/100
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.0150  0.0620  0.1020  0.1166  0.1470  0.5100 
> table(cut(x, breaks = seq(0, 0.55, by = 0.05)))

  (0,0.05] (0.05,0.1] (0.1,0.15] (0.15,0.2] (0.2,0.25] (0.25,0.3] 
       296        532        459        203         80         49 
(0.3,0.35] (0.35,0.4] (0.4,0.45] (0.45,0.5] (0.5,0.55] 
        27         21          8          2          1 
> hist(x, breaks = seq(0, 0.55, by = 0.05))
```

[2] The notation is here is not what would prefer, but MathJax doesn't seem flexible enough to have more freedom for index notation.

The height of the bars in the resulting histogram gives the absolute frequency of observations within the respective sub-intervals.
For computational convenience, it is often helpful to split up an indicator function into the product of as many indicator functions as needed to have simple conditions:

$$\text{condition low} = x > b_{\text{low}}$$

and

$$\text{condition up} = x \leq b_{\text{up}}$$

With this:

$$\text{hist}\left(x_1, \ldots, x_n | \left(b_{\text{low}}, b_{\text{up}}\right]\right) = \sum_{i=1}^{n} \mathrm{I}_{\text{condition low and condition up}}\left(x_i\right) =$$

$$\sum_{i=1}^{n} \mathrm{I}_{\text{condition low}}\left(x_i\right) \cdot$$

$$\cdot \mathrm{I}_{\text{condition up}}\left(x_i\right),$$

with

$$\mathrm{I}_{\text{condition low}}\left(x_i\right) =$$

$$\begin{cases} 1, & \text{if } x_i > b_{\text{low}}, \\ 0, & \text{else,} \end{cases}$$

and

$$\mathrm{I}_{\text{condition up}}\left(x_i\right) =$$

$$\begin{cases} 1, & \text{if } x_i \leq b_{\text{up}}, \\ 0, & \text{else.} \end{cases}$$

Application (Histogram as probability density visualization) In making a histogram to approximate a probability density function, one divides the number of observations in each sub-interval by the total number of observations times the sub-interval length (for a probability density interpretation of histograms, the area under the histogram needs to add to one).
In *R*, this is achieved with `hist(..., freq = FALSE)`.

8 Product

Definition (Product) A *product* is denoted using the product-sign $\prod$, where indexing is performed in analogy to sums:

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_{n-1} \cdot x_n.$$

Calculation rules for products

$$\prod_{i=1}^{n} (a \cdot x_i) = a \cdot x_1 \cdot a \cdot x_2 \cdot a \cdot x_3 \cdot \ldots \cdot a \cdot x_{n-1} \cdot a \cdot x_n = a^n \cdot \prod_{i=1}^{n} x_i.$$

Applying the calculation rules for the logarithmic function, we get:

$$\log\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \log\left(x_i\right).$$

Factorial

Definition (Factorial) The *factorial* of a positive integer n, denoted by $n!$, is the product of all positive integers less than or equal to n:

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n - 1 \cdot n.$$

Special definition: $0! = 1$.

9 Derivative

Definition (Derivative) The *derivative* of a function maps x to the slope of the function at x, i.e. it gives the current increase, or decrease, in the co-domain at any value in the domain f the function.

- It will be important to link the density of a random variable to the distribution of the same as its derivative.
- Since functions may contain parameters, or may be functions of several random variables, it is important to note which is the measure that the derivative refers to, i.e. the slope in which direction.

The common notation for the derivative of $f(x)$ with respect to x is:

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x),$$

referring for the difference in $y = f(x)$ for an (infinitesimally small) change in x at x.

Calculation rules for derivatives See also: `https://en.wikipedia.org/wiki/Differentiation_rules`
Derivatives of powers:

$$\frac{\mathrm{d}}{\mathrm{d}x} x^a = a x^{a-1},$$

where a is any real number.
Derivatives of constants:

$$\frac{\mathrm{d}}{\mathrm{d}x} a = 0,$$

where a is any real number.
Derivative of the exponential function and the logarithmic function:

$$\frac{\mathrm{d}}{\mathrm{d}x} \exp(x) = \exp(x)$$

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln(x) = \frac{1}{x}$$

Sum rule:

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(f(x) + g(x)\right) = \frac{\mathrm{d}}{\mathrm{d}x} f(x) + \frac{\mathrm{d}}{\mathrm{d}x} g(x).$$

Product rule:

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(f(x) \cdot g(x)\right) = f(x) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}x} g(x)\right) + g(x) \cdot \left(\frac{\mathrm{d}}{\mathrm{d}x} f(x)\right)$$

Quotient rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{f\left(x\right)}{g\left(x\right)} = \frac{g\left(x\right)\cdot\left(\frac{\mathrm{d}}{\mathrm{d}x}f\left(x\right)\right) - f\left(x\right)\cdot\left(\frac{\mathrm{d}}{\mathrm{d}x}g\left(x\right)\right)}{g^{2}\left(x\right)}$$

Chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}f\left(g\left(x\right)\right) = \left(\frac{\mathrm{d}}{\mathrm{d}g\left(x\right)}f\left(g\left(x\right)\right)\right)\cdot\left(\frac{\mathrm{d}}{\mathrm{d}x}g\left(x\right)\right)$$

10 Integral

Definition (Integral) The symbol for an **integral** is a stylish S with lower and upper integration limits, that is a and b. The S is for *sum*, as integration is the technique of summing up infinitesimally thin slices of the area under a function.
Directly next to the integral symbol, we write down the function we want to find the integral of - called the *integrand*, e.g. $f(x)$ -, and then finish our integration notation with $\mathrm{d}x$, by which we note down that the slices refer to the x direction (and approach zero in width):

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

The definition of the integral will be needed, for example, for the definition of the expected value of a continuous random variable.

Example (Integral of a linear function) The integral for a linear function $f(x) = c+d\cdot x$ is:

$$\int_a^b (c + d \cdot x)\ \mathrm{d}x = \int_a^b c\,\mathrm{d}x + \int_a^b d\cdot x\,\mathrm{d}x = c\cdot\int_a^b 1\,\mathrm{d}x + d\cdot\int_a^b x\,\mathrm{d}x = c\cdot(b-a) + 0.5\cdot d\cdot\left(b^2 - a^2\right).$$

We suppressed here adding the integration constant C[3] and used integration rules that allowed us to write the integral of a sum as the sum of integrals, and to extract constant factors from the integrand.

[3] If we would add C to the resulting integral and calculate the derivative, C provides no slope and therefore is dropped.