UNIVERSITÀ DEGLI STUDI DI TRIESTE

Statistics

Continuos Random Variables

Luca Pennella

September 16th, 2024

Random variable

A **random variable** (r.v.) X is a variable whose value is a numerical outcome of a random phenomenon, that is, it is a well defined but unknown number

There are two main types of random variables: discrete, if it has a finite list of possible outcomes, and continuous, if it can take any value in an interval.

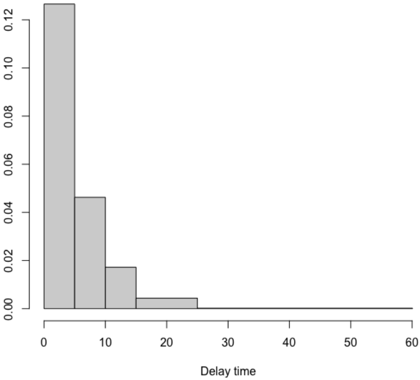
- D the number of tails on three coin tosses
- D the number of defective items in a sample of 20 items from a large shipment
- D the number of students attending the statistics class on a Friday
- C the delay time of the airplane
- C the weight of a newborn
- C the length of a phone call to your mother

For continuous random variables we can assign probabilities only to a range of values, using a mathematical function, we express the probability distribution in the continuous space

Delay time of a flight

About the delay time of a flight, we might observe the following data and plot them using a histogram

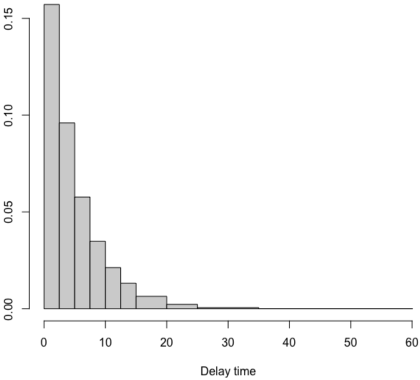
Delay	P
(0,5]	0.633
(5,10]	0.231
(10,15]	0.086
(15,25]	0.043
(25,Inf]	0.006



Delay time of a flight

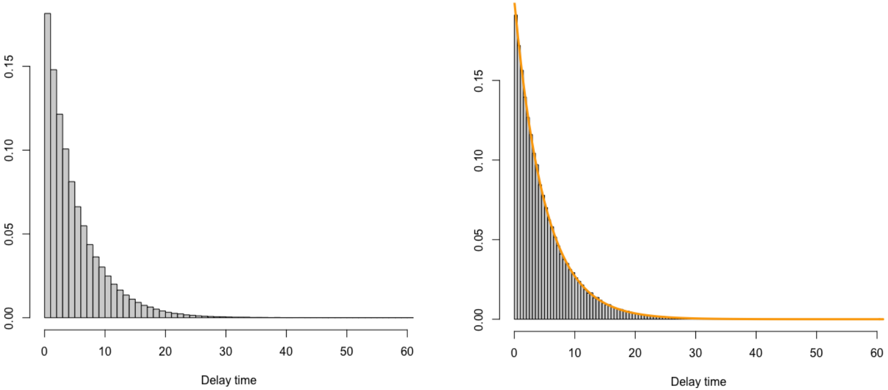
About the delay time of a flight, we might observe the following data and plot them using a histogram

Delay	P
(0,2.5]	0.393
(2.5,5]	0.240
(5,7.5]	0.144
(7.5,10]	0.087
(10,12.5]	0.053
(12.5,15]	0.033
(15,20]	0.032
(20,25]	0.011
(25,35]	0.005
(35,Inf]	0.001

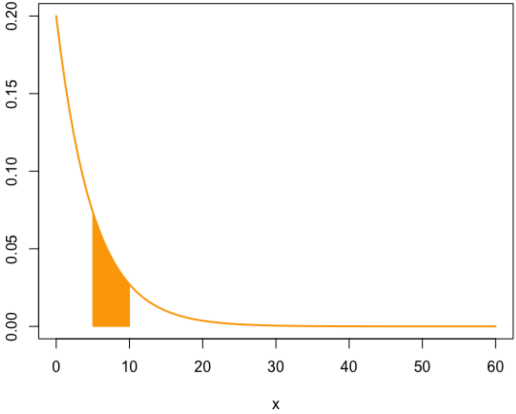


Delay time of a flight

Assuming that we have infinite observations we can increase the number of classes and approximate the histogram with a curve: the probability density function



The interpretation of the probability density function is analogous to the histogram: the areas represent probabilities rather than frequencies



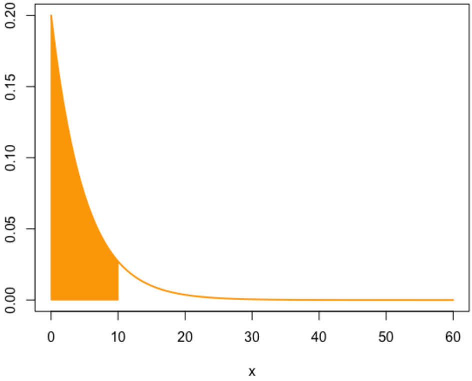
E.g. the probability of observing a value between 5 and 10 is:

$$P(5 \leq X \leq 10) = 0.234$$

Cumulative distribution function

A relevant quantity is

$$P(X \leq x)$$

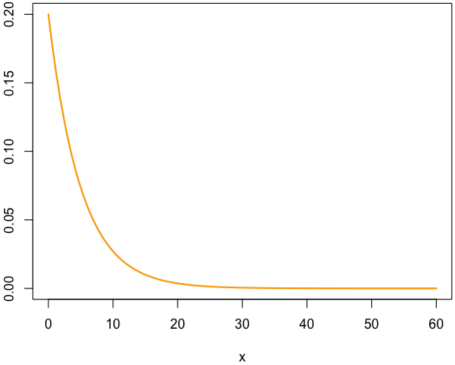


E.g. the probability of observing a value smaller than 10 is:

$$P(X \leq 10) = 0.865$$

Probability density function

The probability density function is a real-valued function assuming non-negative values



The function in the plot is

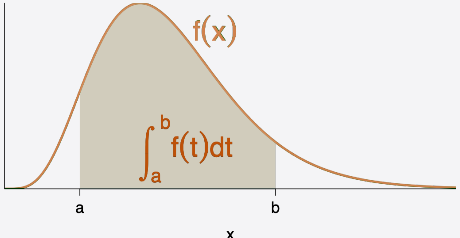
$$f(x) = \frac{1}{5}e^{-x/5}, \quad x \geq 0$$

Probability density function

Probability density function

The probability density function of a continuous random variable X is a non-negative function $f(x)$ whose area under the curve in a range of values is the probability of X in that range:

$$\int_a^b f(t)dt = P(a \leq X \leq b)$$



Probability density function

Probability density function

The probability density function of a continuous random variable X is a non-negative function $f(x)$ whose area under the curve in a range of values is the probability of X in that range:

$$\int_a^b f(t)dt = P(a \leq X \leq b)$$

The probability density function satisfies the following properties:

(i) $f(x) \geq 0$

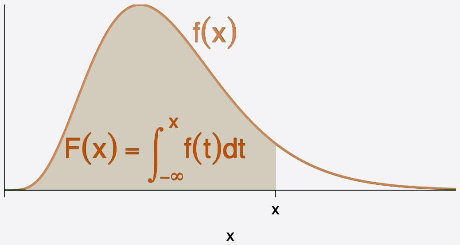
(ii) $\int_{-\infty}^{\infty} f(t)dt = 1$

Cumulative distribution function

Cumulative distribution function

The cumulative distribution function of a random variable X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$



Cumulative distribution function

Cumulative distribution function

The cumulative distribution function of a random variable X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

and satisfies the following properties:

i. $F(x) \geq 0, \qquad \forall x \in \mathbb{R};$

ii. $F(x)$ is not decreasing;

iii. $\lim_{x \to -\infty} F(x) = 0;$

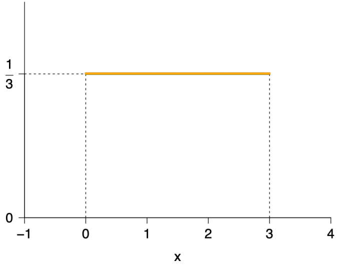
iv. $\lim_{x \to +\infty} F(x) = 1.$

and note that

$$P(a \leq X \leq b) = F(b) - F(a)$$

Example

There are several density functions, any non-negative function that integrates to 1 is a density function



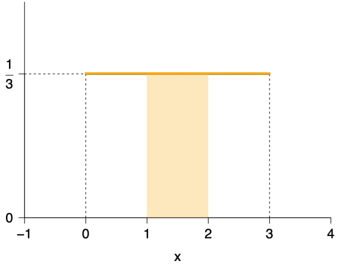
Uniform between 0 and 3

$$f(x) = \begin{cases} 1/3 & \text{for } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Find $P(1 \leq X \leq 2)$

Example

There are several density functions, any non-negative function that integrates to 1 is a density function



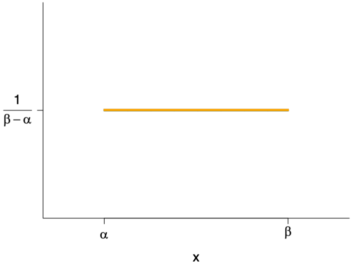
Uniform between 0 and 3

$$f(x) = \begin{cases} 1/3 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

$$P(1 \leq X \leq 2) = (2-1) \times \frac{1}{3} = \frac{1}{3}$$

Probability density functions and parameters

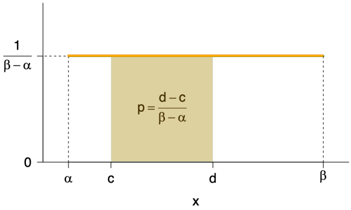
Often it is useful to define a probability density function up to one or more parameters, which is the equivalent of defining a set of density functions



$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

$f(x)$ is a probability density function for every α and β, with $\beta > \alpha$

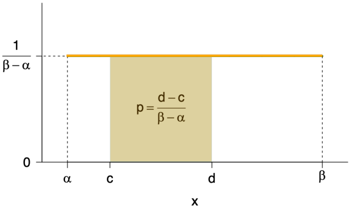
Uniform distribution on $[\alpha, \beta]$



$$P(c \leq X \leq d) = \frac{d-c}{\beta-\alpha}$$

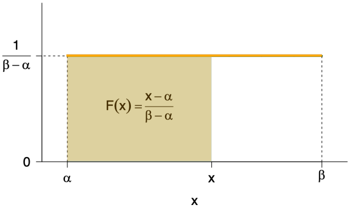
with $\alpha < c < d < \beta$

Uniform distribution on $[\alpha, \beta]$



$$P(c \leq X \leq d) = \frac{d-c}{\beta-\alpha}$$

with $\alpha < c < d < \beta$



$$F(x) = \begin{cases} 0 & \text{if} x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \text{if} \alpha \leq x \leq \beta \\ 1 & \text{if} x > \beta \end{cases}$$

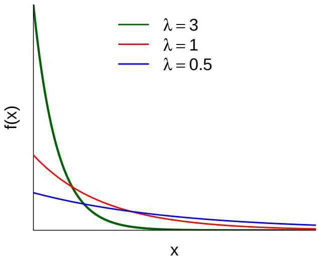
Exponential distribution

The Exponential distribution is a continuous probability distribution that describes the time between events in a Poisson process, where events occur continuously and independently at a constant average rate.

It has rate parameter λ and is defined as

$$f(x) = \lambda e^{-\lambda x}$$

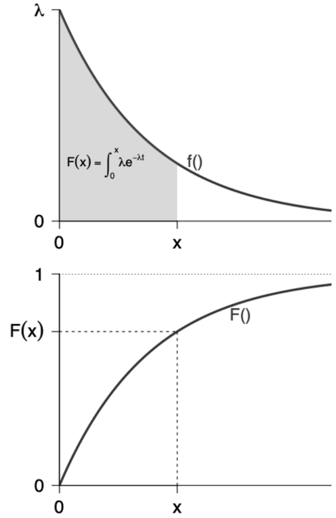
for $x \geq 0$ and $\lambda > 0$



Exponential distribution

The cumulative distribution function is

$$
\begin{aligned}
F(x) &= \int_0^x \lambda e^{-\lambda t} dt \\
&= \left[-e^{-\lambda t}\right]_0^x \\
&= 1 - e^{-\lambda x}
\end{aligned}
$$



Example

Knowing that X follows an exponential distribution with rate parameter λ

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

- If $\lambda = 2$, what is $P(X < 1) =$
- If $\lambda = 2$, $P(X < 0.5) =$
  - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions?

- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Example

Knowing that X follows an exponential distribution with rate parameter λ

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

- If $\lambda = 2$, what is $P(X < 1) = 0.8646647$
- If $\lambda = 2$, $P(X < 0.5) = 0.6321206$
    - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions?

- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Example

Knowing that X follows an exponential distribution with rate parameter λ

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

- If $\lambda = 2$, what is $P(X < 1) = 0.8646647$
- If $\lambda = 2$, $P(X < 0.5) = 0.6321206$
  - So, what is the probability that X is between 0.5 and 1?
- If $\lambda = 4$ what is the answer to the previous questions? (0.9816844, 0.8646647, 0.1170196)
- The delay of a train (in minutes), is distributed according to an exponential of the parameter $\lambda = 0.1$, how likely is it to wait more than ten minutes?
- If the delay were distributed according to an exponential of parameter 0.2, would we wait longer or shorter (probably)?

Expected value and variance of a continuous rv

Let $f(x)$ be the probability density function of a continuous random variable X

$$E(X) = \int xf(x)dx$$

$$E(h(X)) = \int h(x)f(x)dx$$

$$V(X) = \int (x - E(X))^2 f(x)dx$$

Moreover, the following properties hold (using the properties of the integrals)

$$E(aX + b) = aE(X) + b$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$V(aX + b) = a^2 V(X)$$

Uniform distribution: expected value and variance

Let X be a random variable following the Uniform distribution between α and β, thus

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

then

$$E(X) = \int_\alpha^\beta \frac{x}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_\alpha^\beta = \frac{1}{\beta-\alpha}\frac{\beta^2-\alpha^2}{2} = \frac{\alpha+\beta}{2}$$

$$E(X^2) = \int_\alpha^\beta \frac{x^2}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_\alpha^\beta = \frac{1}{\beta-\alpha}\frac{\beta^3-\alpha^3}{3} = \frac{\beta^2+\alpha\beta+\alpha^2}{3}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} = \frac{(\beta-\alpha)^2}{12}$$

Exponential distr: expected value and variance

Let X be a random variable following the Exponential distribution with rate parameter $\lambda > 0$, thus

$$f(x) = \lambda e^{-\lambda x}$$

for $x \geq 0$, then, by integration by parts,

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \left[xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x} dx = -\left[\frac{e^{-\lambda x}}{\lambda}\right]_0^\infty = \frac{1}{\lambda}$$

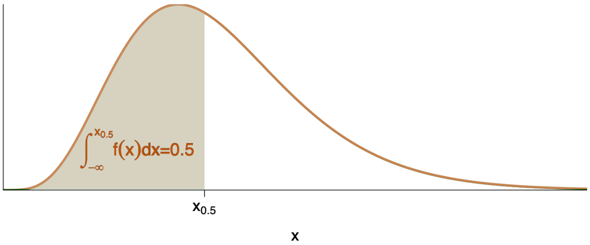
$$E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x} dx = \left[x^2 e^{-\lambda x}\right]_0^\infty + \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Median

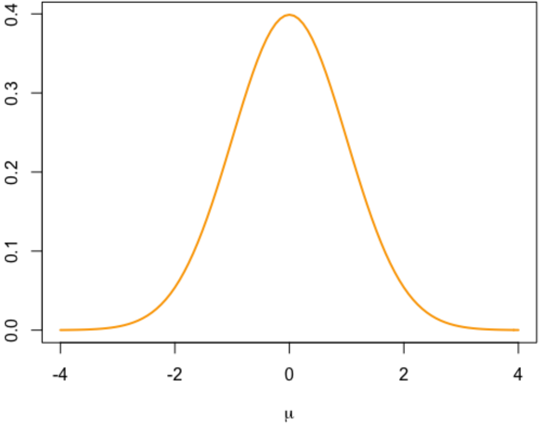
We define the median of a continuous random variable X with pdf $f(x)$ the value $Me(X)$ or $x_{0.5}$ such that

$$F(Me(X)) = \int_{-\infty}^{Me(X)} f(x)dx = 0.5$$



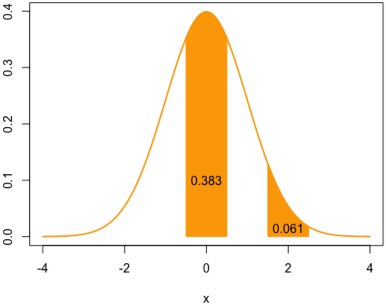
Normal distribution

The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



Normal distribution

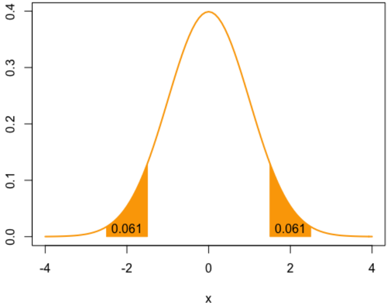
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)

Normal distribution

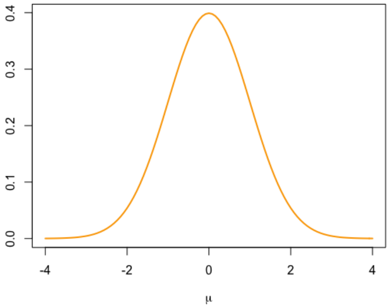
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)
- It is symmetric, then symmetric deviations on the right and left have the same probability

Normal distribution

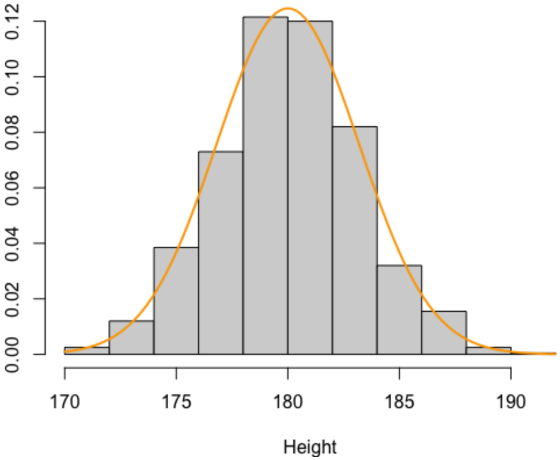
The most popular continuous distribution is the Normal (or Gaussian) distribution and its density has the following shape



- The most likely values to occur are the ones around the center (the mean)
- It is symmetric, then symmetric deviations on the right and left have the same probability
- The pdf is: $f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Example

The distribution of the male students' height

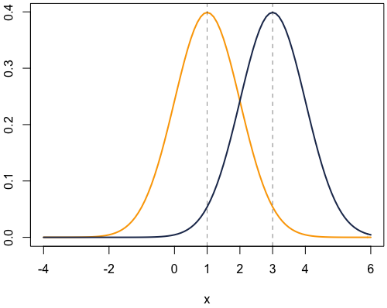


Normal distribution: parameters

The probability density function of the normal distribution is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



μ is the mean of the distribution

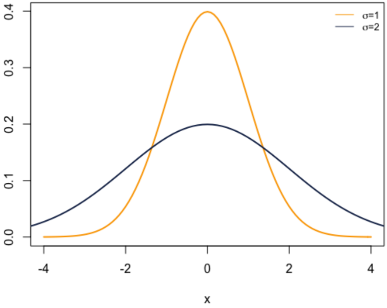
For different μ values, the distribution shifts on the x axis

Normal distribution: parameters

The probability density function of the normal distribution is

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



σ^2 is the variance of the distribution

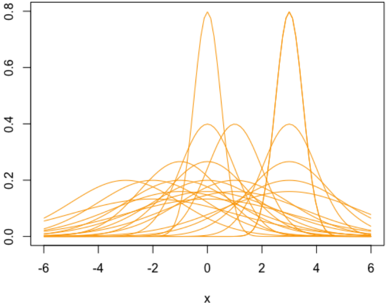
For different σ values, the distribution remains centered on the same value but becomes wider: larger values (in absolute values) of the r.v. are more likely

Normal distribution: parameters

The probability density function of the normal distribution is

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$



Varying μ e σ, we can obtain an infinite number of distributions

Normal distribution: parameters

The probability density function of the normal distribution is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Thus if a r.v. X follows a normal distribution we write $X \sim N(\mu, \sigma^2)$

Normal distribution: probabilities

Assume that $X \sim N(\mu, \sigma^2)$, then its pdf is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

we can use it to compute probabilities of a Normal rv

$$p = P(a \leq X \leq b)$$

where

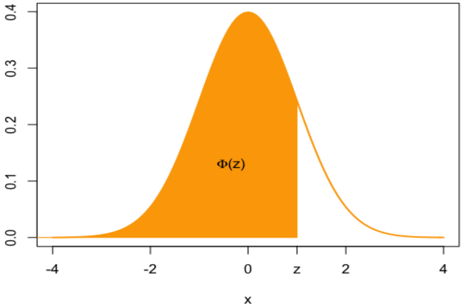
$$p = \int_a^b f(t)dt = F(b) - F(a)$$

Unfortunately, we do not have a closed form to compute this probability, but we can refer to a particular normal distribution...

Standard Normal distribution

The standard normal distribution, that is $\mu = 0$ and $\sigma = 1$, play an important role

$$Z \sim N(0,1) \quad \rightarrow \quad f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$



We define

$$\Phi(z) = P(Z \leq z)$$

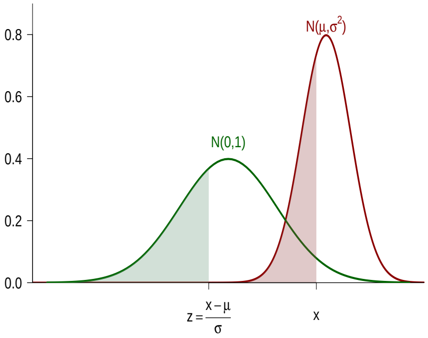
the area under the curve between $-\infty$ and z

Note that with Φ we call the cumulative distribution function of the standard normal distribution

Standard Normal vs Normal(μ, σ)

If $X \sim N(\mu, \sigma^2)$, then

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$



That is, the red area (equal to $P(X \leq x)$) is equal to the green one

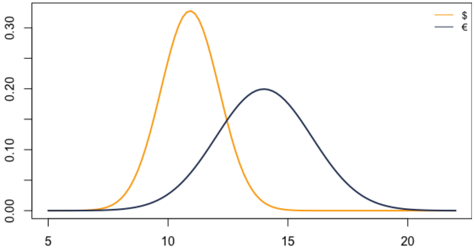
Knowing $\Phi(z)$, we can compute any probability associated with a generic normal distribution $N(\mu, \sigma^2)$

Normal distribution: transformations

Let X be distributed as a normal distribution $N(\mu, \sigma^2)$, and let a, and b be two real numbers, then

$$Y = aX + b$$

follows a normal distribution $Y \sim N(a\mu + b, a^2\sigma^2)$



As an example, if the stock price on a given day expressed in euros follows a Normal distribution $N(14, \sigma^2 = 2)$, then the value in dollars (1 \$ $=$ 0.78 €) is still distributed as a Normal with mean 0.78×14 and variance $2 \times (0.78)^2$