NORMAL (OR GAUSSIAN) DISTRIBUTION (FR: LOI NORMALE)

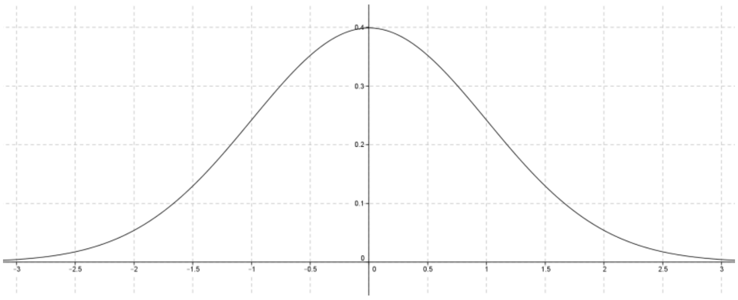
Ecole Internationale PACA
à Manosque
Tle S

1. THE STANDARD NORMAL DISTRIBUTION (FR: LOI NORMALE CENTREE REDUITE)

Definition :
A random variable X follows a standard normal distribution when its density function is defined on $\mathbb{R}$ by :

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

We write $X \sim \mathcal{N}(0;1)$



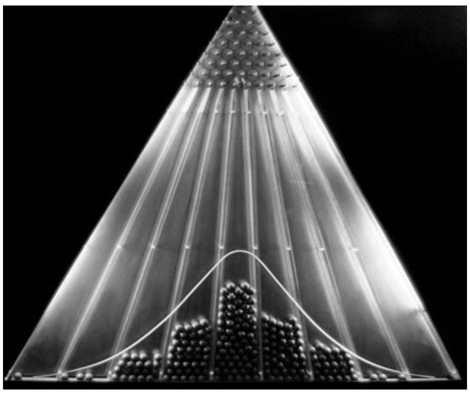
Notes :

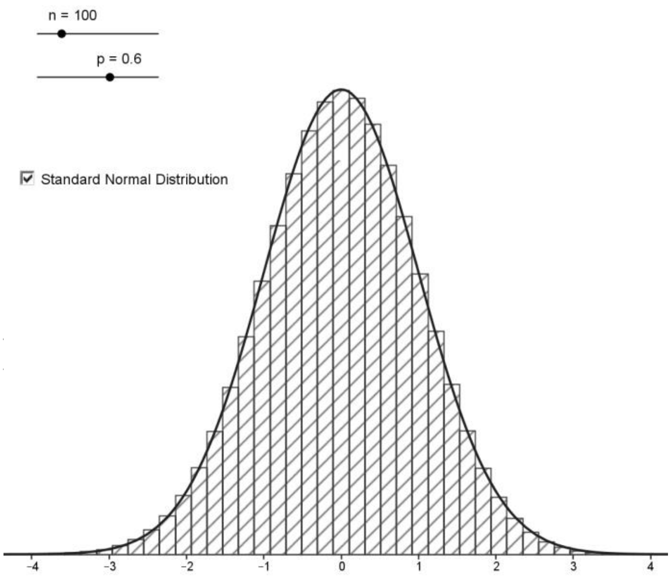
1. This graph is known as the **Gaussian curve** or the bell-shaped graph.
2. If is not possible to express antiderivatives of this function with standard functions, hence the cumulative distribution function (CDF) of the normal distribution is unknown.

Theorem : De Moivre-Laplace[1]
Consider n a non-zero natural integer and p a real number in $[0;1]$ ***,*** X_n ***a random variable following the binomial distribution*** $\mathcal{B}(n;p)$ ***and*** Z_n ***the random variable*** $Z_n=\frac{X_n-np}{\sqrt{np(1-p)}}$.
Then, for any real numbers a and b ($a<b$ ***) :***

$$\lim_{n\to\infty}\left[P\left(a\le Z_n\le b\right)\right]=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

Note : This theorem allows us to approximate (under given conditions) a probability linked to a binomial distribution (usually difficult to calculate) by a probability linked to a normal distribution.

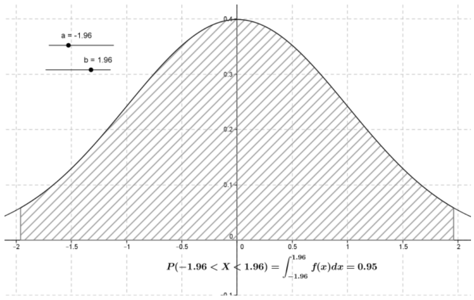




[1] Abraham de Moivre (1667-1654) and Pierre-Simon de Laplace (1749-1827)

Property :

If $X \sim \mathcal{N}(0;1)$ then for any real number α in u_α, there exists a unique positive real number such that $P(-u_\alpha \le X \le u_\alpha) = 1-\alpha$.



Special cases (to be known) :

$u_{0.05} \approx 1.96$ $\quad$ $u_{0.01} \approx 2.58$

Proof :

Quoted in the curricula

b being a positive real number, consider the function F defined on $[0;+\infty[$ by

$$F(b) = \int_0^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

This function is the unique antiderivative of the density function f which vanishes at 0. It is continuous and strictly increasing on $[0;+\infty[$. Since the area below the graph of f , from 0 to $+\infty$ is 0.5, we have $\lim_{b\to+\infty} F(b) = 0.5$. $F(0)$ being 0, F takes its value in $[0;0.5]$.

As $0 < \alpha < 1$, $0 < \frac{1-\alpha}{2} < 0.5$.

Hence according to the IVT, there exits a unique solution (denoted u_α) such that $F(u_\alpha) = \frac{1-\alpha}{2}$.

The graph of f being symmetrical about 0, the finally have $P(-u_\alpha \le X \le u_\alpha) = 1-\alpha$.

Definition :

The expected value of a random variable X defined on $\mathbb{R}$ whose probability density function is f, is defined by : $E(X) = \lim_{x\to+\infty} \int_{-x}^{x} tf(t)dt$

Note : Its variance can be defined as for discrete variables by $V(X) = E\left[(X - E(X))^2\right]$ or $V(X) = E(X^2) - (E(X))^2$

Property :

If $X \sim \mathcal{N}(0;1)$ then its expected value is $\mu = 0$ and its standard deviation is $\sigma = 1$

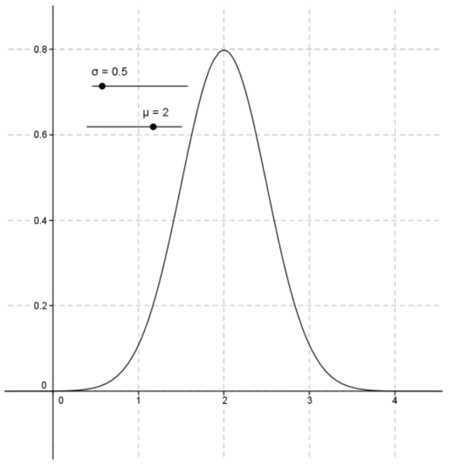
Proof :

Calculate $\int_{-x}^{x} tf(t)dt = \ldots$ and then take the limit …….

2. THE NORMAL DISTRIBUTION (FR: LOI NORMALE)

Definition :
A random variable X follows a normal distribution with expected value μ and standard deviation σ when the random variable $\frac{X-\mu}{\sigma}$ follows a standard normal distribution.

We write $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$

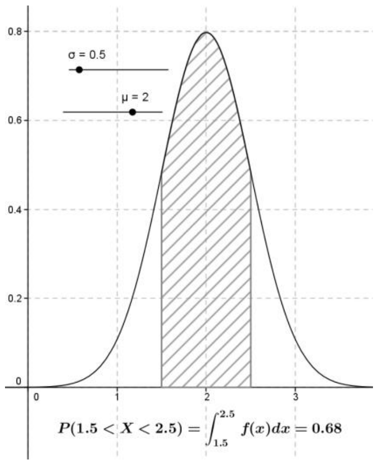


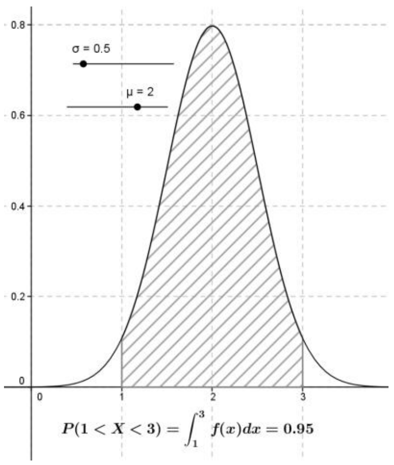
Notes :

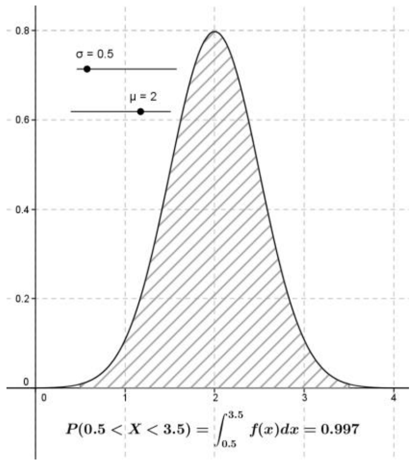
1. The density function is then : $f\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
2. The graph is symmetrical about the line with equation $x=\mu$
3. μ picks out the peak of the graph and σ its spread (small values lead to tall and narrow graphs, larger values give short, fat graphs).

Properties:
If $X \sim \mathcal{N}\left(\mu;\sigma^2\right)$ then :

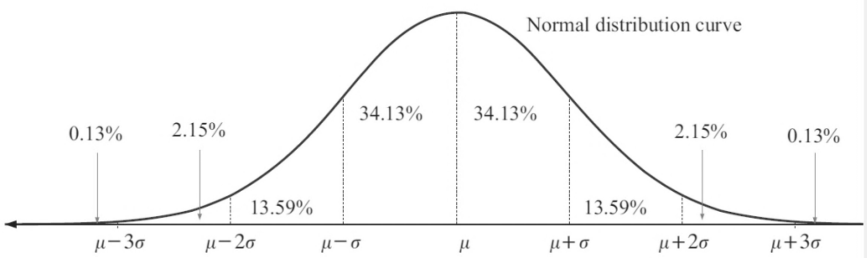
$P\left(\mu-\sigma \leq X \leq \mu+\sigma\right)\approx 0.68$ $\qquad$ $P\left(\mu-2\sigma \leq X \leq \mu+2\sigma\right)\approx 0.95$ $\qquad$ $P\left(\mu-3\sigma \leq X \leq \mu+3\sigma\right)\approx 0.99$







For a normal distribution with mean μ and standard deviation σ, the proportional breakd of where the random variable could lie is given below.



Notice that:
- $\approx 68.26\%$ of values lie between $\mu-\sigma$ and $\mu+\sigma$
- $\approx 95.44\%$ of values lie between $\mu-2\sigma$ and $\mu+2\sigma$
- $\approx 99.74\%$ of values lie between $\mu-3\sigma$ and $\mu+3\sigma$.

Note :

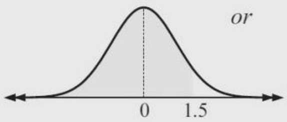
1. **Since we can't calculate directly (we don't know a primitive) , the calculator is very helpful to calculate approximate values of $P(a \le X \le b)$ with the function normalFrép .**

Ex : if $X \sim \mathcal{N}(100;25)$, then $P(70 \le X \le 95) =$ normalFrép(70,95,100,5)

2. **With a spreadsheet : $P(X \le a) =$ LOI.NORMALE.N(a; μ ; σ ;VRAI) and LOI.NORMALE.INVERSE.N(k; μ ; σ) gives the number a such that $P(X \le a) = k$.**

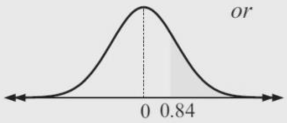
If Z is a standard normal variable, find:
a $P(Z \leqslant 1.5)$ **b** $P(Z > 0.84)$ **c** $P(-0.41 \leqslant Z \leqslant 0.67)$

a $P(Z \leqslant 1.5)$
≈ 0.933



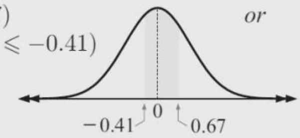
or $P(Z \leqslant 1.5)$
$=$ normalcdf(−E99, 1.5)
≈ 0.933

b $P(Z > 0.84)$
$= 1 - P(Z \leqslant 0.84)$
$\approx 1 - 0.79954$
≈ 0.200



or $P(Z > 0.84)$
$=$ normalcdf(0.84, E99)
≈ 0.200

c $P(-0.41 \leqslant Z \leqslant 0.67)$
$= P(Z \leqslant 0.67) - P(Z \leqslant -0.41)$
$\approx 0.7486 - 0.3409$
≈ 0.408



or $P(-0.41 \leqslant Z \leqslant 0.67)$
$=$ normalcdf(−0.41, 0.67)
≈ 0.408

EXAMPLE

The chest measurements of 18 year old male footballers is normally distributed with a mean of 95 cm and a standard deviation of 8 cm.

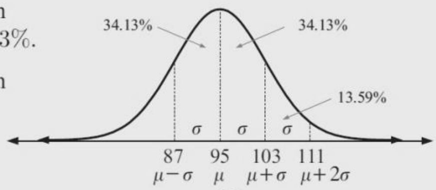
a Find the percentage of footballers with chest measurements between:
  i 87 cm and 103 cm **ii** 103 cm and 111 cm

b Find the probability that the chest measurement of a randomly chosen footballer is between 87 cm and 111 cm.

a **i** We need the percentage between $\mu - \sigma$ and $\mu + \sigma$. This is $\approx 68.3\%$.

ii We need the percentage between $\mu + \sigma$ and $\mu + 2\sigma$.

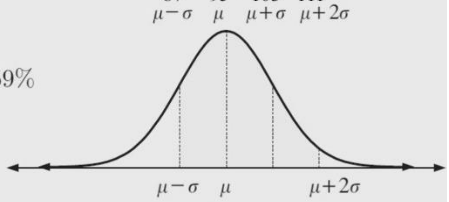
This is $\approx 13.6\%$.



b This is between $\mu - \sigma$ and $\mu + 2\sigma$.

The percentage is $68.26\% + 13.59\%$
$\approx 81.9\%$.

So, the probability is ≈ 0.819.



Bullet points of the chapter

- ✓ Knowing the graph of the density function of normal probability distributions
- ✓ Proof spotted in the lesson
- ✓ Knowing the values $u_{0.05} \approx 1.96$ and $u_{0.01} \approx 2.58$
- ✓ Knowing that If $X \sim \mathcal{N}(\mu;\sigma^2)$ then $P(\mu - \sigma \le X \le \mu + \sigma) \approx 0.68$
- ✓ Knowing that If $X \sim \mathcal{N}(\mu;\sigma^2)$ then $P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 0.95$
- ✓ Knowing that If $X \sim \mathcal{N}(\mu;\sigma^2)$ then $P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.99$
- ✓ Using a calculator or a spreadsheet to get $P(a \le X \le b)$