# EXERCISE CLASS 1 (Part 3/3)

Final remarks and conclusions

The normal (gaussian) distribution

Probability Distributions

The normal (gaussian) distribution

*Properties of the normal distribution:*

If $X$ is normally distributed with mean $\mu$ and variance $\sigma^2$, then the variable $Y = aX + b$, for any real numbers $a$ and $b$, is also normally distributed, with mean $a\mu + b$ and variance $a^2\sigma^2$.

If $X1$ and $X2$ are two independent normal random variables, with means $\mu 1$, $\mu 2$ and variances $\sigma 1^2$, $\sigma 2^2$, then their sum $X1 + X2$ will also be normally distributed with mean $\mu 1 + \mu 2$ and variance $\sigma 1^2 + \sigma 2^2$

More generally:
**any linear combination of independent normal variables is a normal variable**

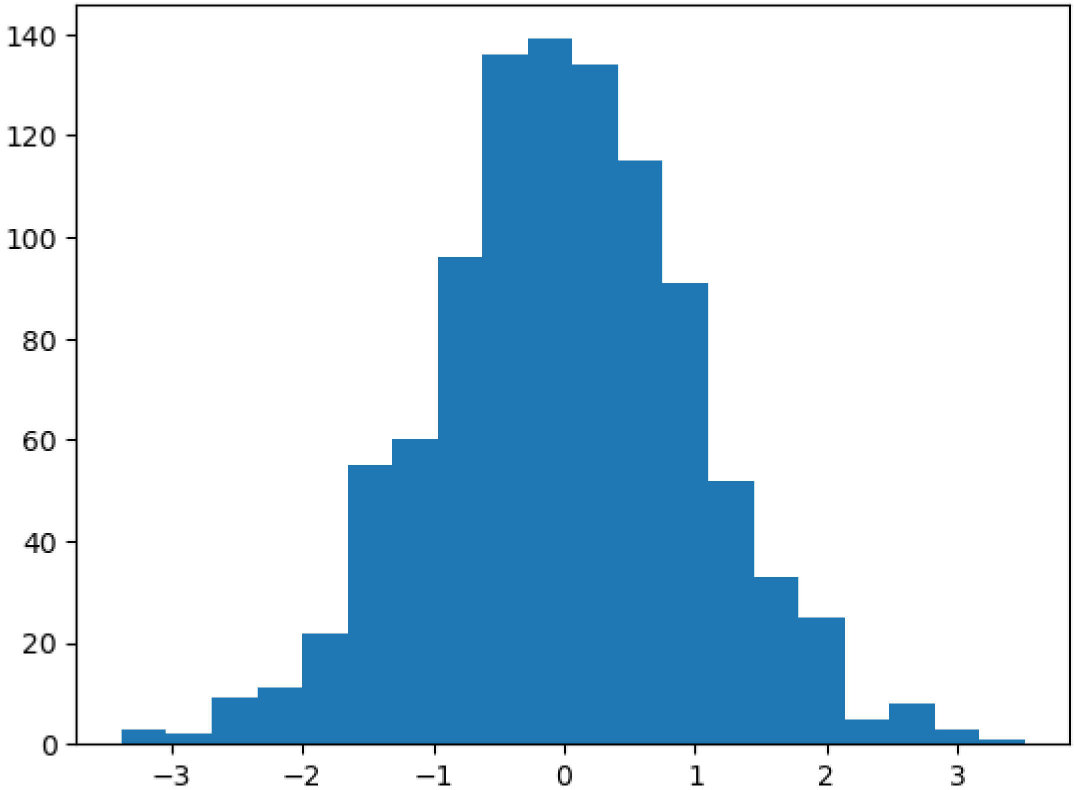
How to generate random numbers from a given distribution?

In [ ]:

```
# Import the necessary libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

#generate 1000 random data from normal (0,1)
mu=0
sigma=1
n=1000

data = np.random.normal(mu,sigma,n)

#plot the data
plt.hist(data, bins=20)
```

Out[ ]:

```
(array([  3.,   2.,   9.,  11.,  22.,  55.,  60.,  96., 136., 139., 134.,
        115.,  91.,  52.,  33.,  25.,   5.,   8.,   3.,   1.]),
 array([-3.37996247, -3.03529502, -2.69062757, -2.34596013, -2.00129268,
        -1.65662523, -1.31195779, -0.96729034, -0.62262289, -0.27795544,
         0.066712  ,  0.41137945,  0.7560469 ,  1.10071434,  1.44538179,
         1.79004924,  2.13471669,  2.47938413,  2.82405158,  3.16871903,
         3.51338647]),
 <BarContainer object of 20 artists>)
```




Normality and independence

REMIND:

The random variables $x_1, x_2, \dots, x_i, \dots, x_n$ are **independent** if:

$$P(x_1 \in E_1, x_2 \in E_2, \dots, x_n \in E_n) = P(x_1 \in E_1)\, P(x_2 \in E_2) \dots P(x_n \in E_n)$$

For any sets $E_1, E_2, \dots, E_n$

Two random variables are independent if the realization of one does not affect the probability distribution of the other

If random variables are independent, their covariance is zero
If the distribution is **normal**, than this is sufficient and necessary.

The Central Limit Theorem (1/2)

It's one of the reason why the normal distribution plays a so important role in statistics

If $x_1, x_2, \dots, x_i, \dots, x_n$ are ***independent random variables*** with mean $\mu_i$ and variance $\sigma_i^2$, and
If $y = x_1 + x_2 + \dots + x_i + \cdots + x_n$,

THEN the distribution of:

$$\frac{y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

**approaches a standard normal N(0,1) distribution** as *n* approaches infinity

The Central Limit Theorem (2/2)

*What does this theorem imply*?

The sum (or **average**) of a large number *n* of independently distributed random variables is approaximately normal, regardless of the distribution of the individual variables.

THUS:

The sampling distribution of $\bar{X}$ is approximately normal (for a large enough *n*), regardless of the distribution of *X* !

In the end…

# CONCLUDING SUMMARY

*(what you have learned – or just reviewed)*

- What is a random variable and its main properties
- How to graphically depict a set of data (histogram, time plot, boxplot)
- Probability distributions and descriptive statistics
- The normal distribution, its properties and the central limit theorem