3. Extraction of parameters from a physiological signal

A statistical parameter in signal processing is any measured quantity of a number of signal samples that summarises or describes an aspect of a signal, such as a mean or standard deviation.

There are several statistical parameters. In this tutorial we will look at the most commonly used parameters, which are: mean, median, variance, standard deviation, kurtosis (kurtosis coefficient, degree of curvature, acuity coefficient), skewness (asymmetry coefficient), entropy and energy.

II.1 Average (mean)

This is calculated as the average of the samples, i.e. the sum of all the amplitudes, divided by the number of samples, and is given by equation (I.1).

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \qquad (2.1)$$

For example, the amplitudes of a signal for 10 samples: 301, 358, 298, 320, 407, 302, 299, 280, 305, 346.

The average is then calculated as follows:

```
(301+358+298+320+407+302+299+280+305+346)/10 = 321.6000.
```

In Matlab we use the mean(x) instruction, such that :

```
x=[301, 358, 298, 320, 407, 302, 299, 280, 305, 346] ;
```

II.2 Median

The median represents the midpoint of the data set. This midpoint is the one that divides the observations into two equal halves, one above the value, the other below. The median is determined by ranking the observations, then taking the observation with rank [N + 1] / 2 in the order obtained. If the number of observations is even, the median is equal to the average of the observations of rank N/2 and [N/2] + 1.

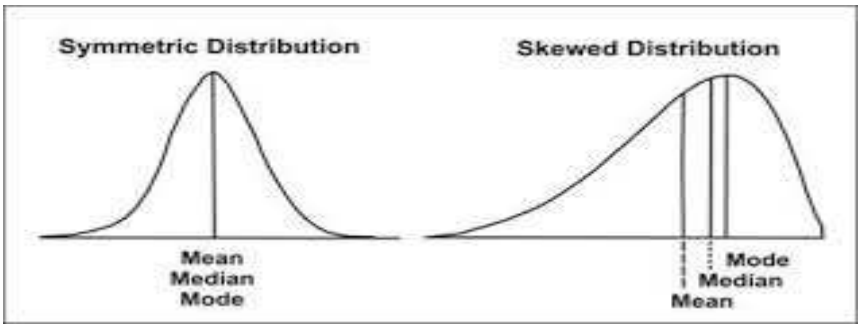


Figure 1: Graphical examples of distributions showing median, mean and mode.

Note:

The mode is the most frequent value in a sample. Median: the median is a number that divides the samples into 2 parts so that each part contains the same number of values.

Matlab uses the instruction: `M = median(x),`

II.3. Variance

This measures the dispersion around the mean.

It is given by equation (I.2).

$$var = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N} \qquad (2.2)$$

II.4 Standard deviation

This is the most commonly used measure of dispersion, along with the standard deviation, which corresponds to the square root of the variance. It is given by equation (I.3).

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N}} \qquad (2.3)$$

Matlab code: `std(x)`

II.5 Kurtosis

Also known as the kurtosis coefficient, degree of curvature or acuity coefficient.

It is a direct measure of acuity and an indirect measure of the flatness of a signal distribution. It also represents the 4th order moment and is given by equation (I.4).

$\beta_2 = E[(\frac{x-\mu}{\sigma})^4]$

(I.4)

Matlab code : `k = kurtosis(X)`

II.6. **Skewness**

The skewness coefficient is a measure of the asymmetry of the sample distribution.

It is the first of the shape parameters, along with kurtosis (moments of order 5).

In general terms, the skewness of a distribution is positive if the right tail (at high values) is longer or fatter, and negative if the left tail (at low values) is longer or fatter.

It is given by equation (I.5).

$\beta_1 = E[(\frac{x-\mu}{\sigma})^3]$ (2.5)

Matlab code : `y = skewness(X)`

II.7. Energy

The energy of a signal x(t), where x is a function of time t, is called Es. Its energy is calculated by the relation:

$E_s = \int_{-\infty}^{+\infty} |x(t)|^2 dt$ (2.6)

Matlab code : `E=x.^2 ;`

II.9. Correlation function:

One of the best-known methods for obtaining information about the relationship and degree of dependence between two signals is the correlation function. If we are interested in the characteristics of a single signal, univariate analysis can perfectly well carry out such a task; but sometimes we want to study the characteristics of two signals, in particular the relationship between two physiological signals.

The correlation function measures the degree of dependence or independence between two signals X and Y as a function of the delay time (τ), since the latter reflects the causal relationship between the two signals.

The correlation function is given by equation (2.7).

$C_{xy}(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k+\tau)y(k)$ (2.7)

Where N is the number of samples

τ: is the delay time between two signals X and Y.

An example of the correlation function between two sinusoidal X and Y signals is shown in Figure 2.

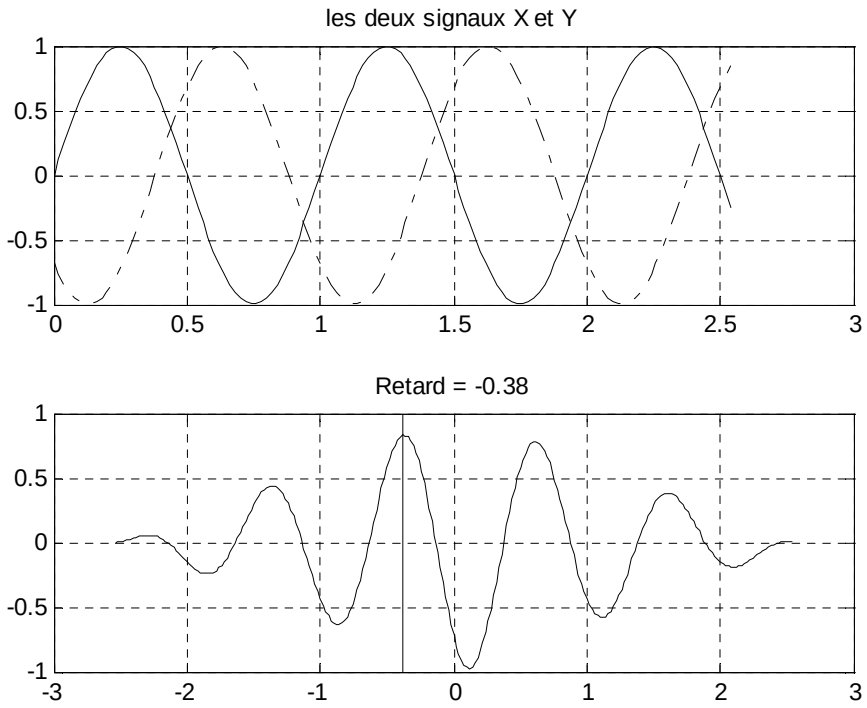


Figure 2: (a) two sinusoidal signals X and Y, (b) the correlation function of the two signals X and Y.

This function varies between -1 (complete inverse correlation) and +1 (complete direct correlation), where C_xy (τ)=0 means that the two signals are independent.

C_xy (τ)<0 implies inverse correlation, i.e. the two signals have similar absolute values but opposite signs. Whereas C_xy (τ)>0 implies a direct correlation .i.e. the two signals have similar values with the same sign.

The value of τ which maximises this function is always taken as an estimate of the delay between the two signals.

MATLAB Code :

```
t = [0:127]*0.02; % le temps
f = 1.0;   % la fréquence
s1 = sin(2*pi*f*t); % le signal sinusoidal s1
s2 = sin(2*pi*f*(t-0.35)); % s1 est en retard par rapport s2 par 0.35s
subplot(2,1,1);
% plot(t,s1,'k',t,s2,'k-.');
```

```
plot(t,s1), hold on, plot(t,s2); % tracer les deux signaux
grid
%title('signaux')
title('…….')

% la correlation des deux signaux
x = xcorr(s1,s2,'coeff');
tx = [-127:127]*0.02;
subplot(2,1,2)
plot(tx,x)
grid
[mx,ix] = max(x);
lag = tx(ix);
hold on
tm = [lag,lag];
mm = [-1,1];
plot(tm,mm,'K')
hold off
S = sprintf('Retard = %5.2f',lag);
title(S)
```

Manipulation :

Determine the following parameters and interpret the results obtained.

1- Determine the mean, median, variance, and standard deviation of the signal RR1, PR1, RR2,

PR2, EEG1, EEG2, and EEG3.

2- Determine the skewness and kurtosis of the signal nb1, nb2, AGb1, AGb2. Compare the results obtained.

3- Calculate the energy of the ‘n.wav’ and ‘AG.wav’ signal.

4- Compare the results obtained.

5- Determine the correlation function between the signal RR1 and PR1, and RR2 and PR2.

6- Determine the delay rate. Conclude.