4. Correlation Concepts

4.1 Correlation in the time domain

4.1.1 Mean and Variance

The mean, variance and autocorrelation functions have their origins in stationary random processes. These are functions which are described by the probability structure of the data which is invariant (to some degree) with time.

To obtain an understanding of data of this type, consider the average value of a random variable, $x(t)$:

$$\mu_x = E[x(t)] = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} x(t)\,dt \qquad (4.1)$$

This definition implies that $x(t)$ exists everywhere over the domain, since if it did not, the average would go to zero in the limit. An impulse function is an example where $\mu_x = 0$ by the above definition. To deal with this, a *transient mean value* is defined as:

$$\mu_x = \int_{-\infty}^{\infty} x(t)\,dt \qquad (4.2)$$

Other functions are also subject to dual definitions depending on the nature of the data; transient or periodic.

The calculation of the average using *sampled* data would be:

$$\overline{x} = \overline{\mu}_x = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (4.3)$$

Since a finite number of data points are used to determine the mean, we have obtained a *sample mean* which in the long run average is equal to the true mean:

$$\mu_x = E[\overline{x}] \qquad (4.4)$$

Another quantity of interest is the variance:

$$\sigma_x^{\ 2} = E\left[\left(x(t) - \mu_x\right)^2\right] \qquad (4.5)$$

This is a measure of the spread of the data around the mean value.

Again, using sampled data results in a sample variance:

$$s_x^{\ 2} = E\left[\left(x(t) - \overline{x}\right)^2\right] \qquad (4.6)$$

or:

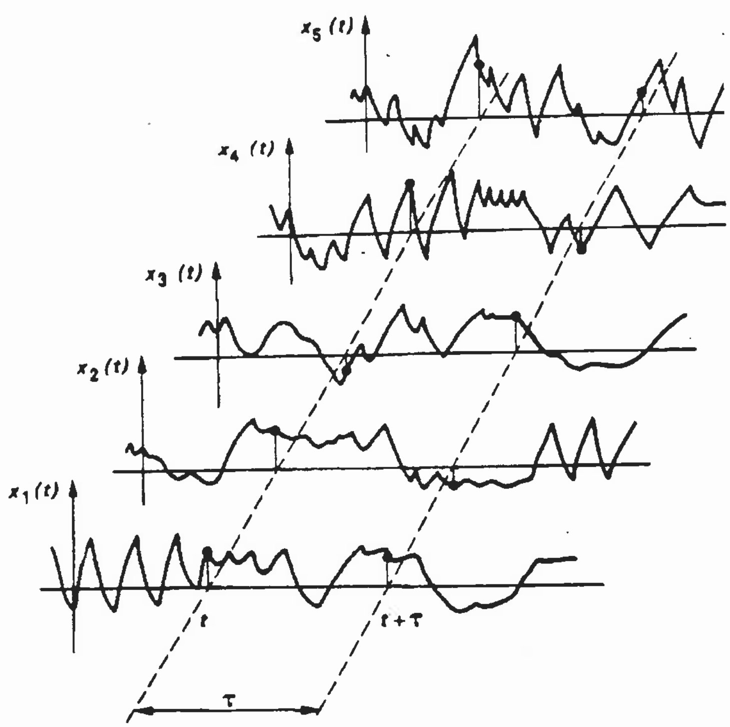
$$s_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 \qquad (4.7)$$

4.1.2 Autocorrelation

The autocorrelation function is defined as:

$$R_{xx}(\tau) \quad = \quad E[x(t)\,x(t+\tau)] \tag{4.8}$$

Where τ is a fixed time interval and t is a fixed time among ensembles (sets) of data. This is shown below:



If $\tau = 0$, the autocorrelation becomes:

$$R_{xx}(0) \quad = \quad E[x(t)\,x(t)] = E\left[x(t)^2\right] \quad = \quad \psi_x^{\;2} \tag{4.9}$$

Which is called the *mean square value of x*. This will be used when we look at the signal in the frequency domain. The mean square is related to the mean and the *variance* by:

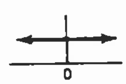
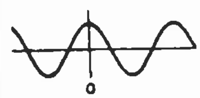
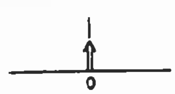
$$\psi_x^{\;2} = \mu_x^{\;2} + \sigma_x^{\;2} \tag{4.10}$$

- The autocorrelation function is used to identify periodic trends in sampled data.
- As with many types of data that we will be dealing with, the autocorrelation function is defined for *negative time* as well as positive time. The function is an even function, so:

$$R_{xx}(\tau) = R_{xx}(-\tau) \tag{4.11}$$

Some common autocorrelation functions are shown below:

Special Autocorrelation Functions

Type	Autocorrelation Function
Constant	 $R_{xx}(\tau) = c^2$
Sine wave	 $R_{xx}(\tau) = \frac{X^2}{2}\cos 2\pi f_0\tau$
White noise	 $R_{xx}(\tau) = a\delta(\tau)$
Low-pass, white noise	 $R_{xx}(\tau) = aB\left(\frac{\sin 2\pi B\tau}{2\pi B\tau}\right)$

For a *periodic* signal, the autocorrelation function is computed by:

$$R_{xx}(\tau) \quad = \quad \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,x(t+\tau)\,dt \tag{4.12}$$

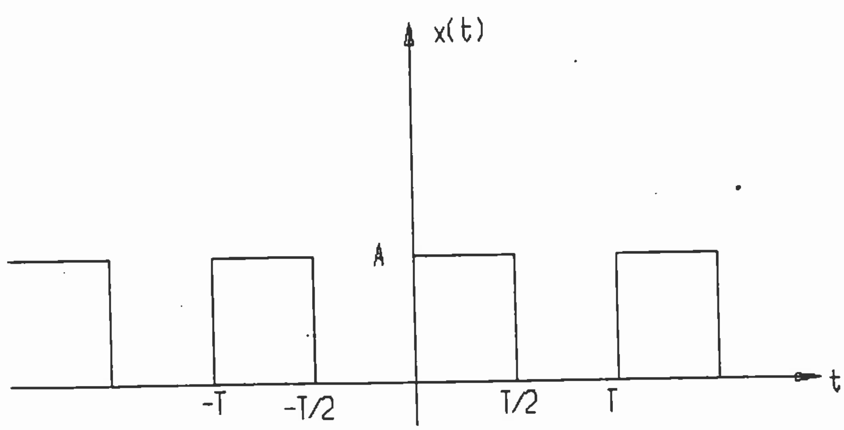
A *random* signal, like the time domain plot shown on the previous page, has an autocorrelation defined by:

$$R_{xx}(\tau) \quad = \quad \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x(t)\,x(t+\tau)\,dt \tag{4.13}$$

Since a random signal is uncorrelated to past and future events, $R_{xx}(\tau)$ will be zero for large τ.

An example for a periodic function is given:

For the square wave shown, determine the autocorrelation function.



$$x(t) = A \qquad 0 < x \le \frac{T}{2}$$

$$x(t) = 0 \qquad \frac{T}{2} < x \le T$$

Computing the autocorrelation from Eqn. 4.12 for the interval from $0 < \tau \le \frac{T}{2}$:

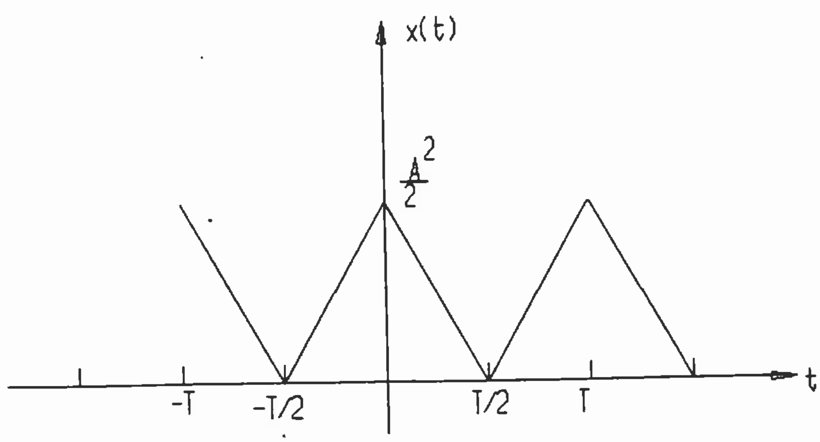
$$R_{xx}(\tau) = \frac{1}{T}\left[\int_0^{T/2-\tau} A^2\,dt + \int_{T/2-\tau}^{T/2} 0\,dt + \int_{T/2}^{T-\tau} 0\,dt + \int_{T-\tau}^{T} 0\,dt\right]$$

$$R_{xx}(\tau) = \frac{A^2}{T}\left(\frac{T}{2} - \tau\right) \qquad \text{for} \quad 0 < \tau \le \frac{T}{2}$$

Likewise:

$$R_{xx}(\tau) = \frac{A^2}{T}\left[\tau - \frac{T}{2}\right] \qquad \text{for} \quad \frac{T}{2} < \tau \le T$$

These functions plot out a saw-tooth wave:



Note that the maximum value is $\frac{A^2}{2}$, which corresponds to a $\tau = 0$.

This should make sense if we recall that $R_{xx}(0) = E\left[x^2\right]$; the long run average is comprised of half A^2 and half 0.

Also note that the function is periodic and even.

4.1.3 Cross-correlation

The cross-correlation function compares two signals in a similar manner as the autocorrelation function:

$$R_{xy}(\tau) \quad = \quad E\left[x(t)\,y(t+\tau)\right] \quad = \quad \frac{1}{T}\int_{-T/2}^{T/2} x(t)\,y(t+\tau)\,dt \qquad (4.14)$$

- Unlike the autocorrelation function, the cross-correlation function is not even
- In general, $R_{xy} \neq R_{yx}$.

4.2 Summary

The four functions defined above are used to quantify data of any type. Although "random" data (noise) is an obvious random variable, periodic functions can also have a random variable in the form of a zero-crossing.