Signals and systems class, HSE, Spring 2015. A. Ossadtchi, Ph.D.

Lecture 2

Basic random sequences (discrete time random processes)

In real world we are often dealing with signals that could be best characterized as random. Random sequence is NOT a set of numbers listed in some particular order. **Random sequence is a concept**, a law that prescribes how to generate a realization of this random sequence. A realization is a set of numbers, no longer random as we already know this set of values! This should be crystal clear! Read it several times and try to understand what I mean here!

Sometimes, the law for generating a realization is genuinely discrete. Sometimes, however, the sequences we are dealing with are, in fact, sampled (discretized) realizations of continuous time random process. To formalize the concept of a random process, consider a set of outcomes Ω .When dealing with random number we used to associate each element of this space with a particular realization of a random variable. For example, for a dice, Ω has six outcomes each being (or corresponding to) a face of the dice. Each face has some symbol, or number written on it that is getting assigned to the a value that the random variable takes at a particular realization.

Random variable is a **rule** $X(\omega),\ \omega\in\Omega$ that maps the space of experimental outcomes Ω to the set of numbers. In case of the example with the dice, realization of a random variable is this number written on the face corresponding to the randomly picked element ω from the set of experimental outcomes Ω .

Random process is a **rule** that maps a set of experimental outcomes to a set of functions. Experimental outcome ω_k is mapped to function $X(t,\omega_k)$ of time (or any other carrier (independent) variable) that is called a realization of random process. When using discrete time signal processing, we will be dealing with random sequences $x[n]=X(t_n,\omega)$ where t_n is a set of time instances at which we register the values of this random process. Again, a random process is usually thought of as a function of time, but it can equally well be a function of other independent (carrier) variable(s), such as, for instance, spatial coordinates.

A random process is basically a giant family of functions indexed by experimental outcome variable ω This family of functions is called an **ensemble**. Note that we used word "rule" in the definition of the random process. We did this in order to spare the word "function" as we needed it further in the definition.

Consider an example. A random process that is called sinusoid with random phase. In the discrete time case it is given by the following function $x[n]=\sin(2\pi f n+\phi(\omega))$. Here, we can see that phase shift $\phi(\omega)$ is a random variable, its distribution is given, usually, for this signal model $\phi(\omega)$ is supposed to be uniformly distributed in the $[0,2\pi]$ interval. To generate a realization of this process, we sample random variable $\phi(\omega)$ and obtain, say, $\phi(\omega_i)=\pi/3$, then the realization of this discrete time random process is simply $x[n]=\sin(2\pi f n+\pi/3)$ that is no longer random and looks like a shifted sinusoid, shifted by the value that is a realization of the random variable $\phi(\omega)$. If we sample $\phi(\omega)$ *N* times, we will obtain the ensemble of *N* realizations that can be used to compute the estimates of statistical characteristics of the random process.

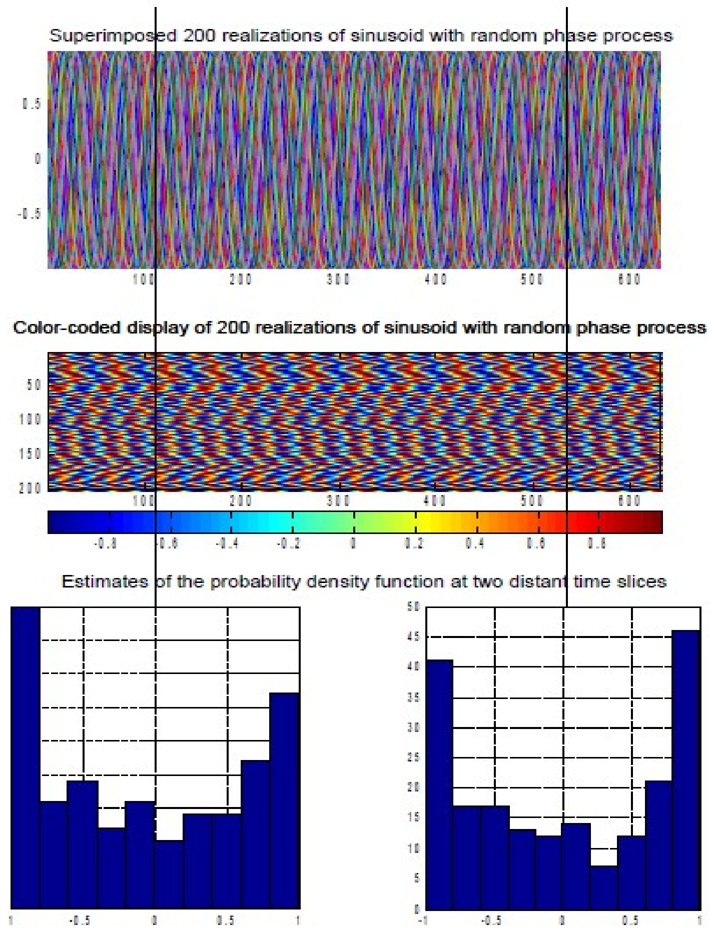

Figure 1: The best we can do to characterize a random process is to specify the joint pdf of the random variables for every value of t_n *. The bottom panel of figure shows two marginal distributions corresponding to the two different moments in time. To plot the histograms, we fixed the time slice index and took values of the random process corresponding to the 200 realizations of the random process at this particular time index. Importantly, specifying these marginal distributions apparently does not uniquely specify the sinusoid with random phase random process, as specifying marginal distributions and not specifying "connection" between the time slices does not allow to specify the dynamical properties of the process, i.e. describe its smoothness and etc. To do so using the pdf approach we need to specify the joint pdf of values of the random process for all time slices In other words we need to specify a full blown multivariate pdf* $P([x_{1,} x_{2,} .. x_N])$ *if we are talking about the realizations with size of N s*amples.

Fixing time moment, we get $x[n]=X(t_n,\omega)$ is simply a random variable. The best we can do to characterize a random *variable* is to specify its probability density function (pdf). The best we can do to characterize a random *process* is to specify the joint pdf of the random variables for every value of t_n . The bottom panel of figure 1 shows two marginal distributions corresponding to the two different moments in time. To plot the histogram, we fixed the time slice index and took values of the random process corresponding to the 200 realizations of the random process at this particular time index. Importantly, specifying these marginal distributions apparently does not uniquely specify the sinusoid with random phase random process, as specifying marginal distributions and not specifying "connection" between the time slices does not allow to constrain the dynamical properties of the process, i.e. describe its smoothness and etc. To do so using the pdf approach we need to specify the joint pdf of values of the random process for all time slices. In other words we need to specify a full blown multivariate pdf (that in general may be a function of time) $p([x_{1,} x_{2,} .. x_N]) = p([x_{1,} x_{2,} .. x_N], t_{1,} t_{2,} .., t_N)$ if we are talking about the realizations with size of N time samples. It is only for the white noise process whose time samples are independent that the joint distribution can be represented as the product of distributions i.e. $p([x_{1,} x_{2,} .. x_N]) = p(x_{1,} t_1) p(x_{2,} t_2) ... p(x_N, t_N)$. In other cases the joint distribution is the product of conditional distributions reflecting the fact that the process values at different time slices are not independent. This means that most processes have some sort of temporal smoothness that contrasts them from the white noise process (sequence of independent identically distributed (iid) random variables) that does not have any temporal structure and whose samples are independent.

Consider this almost trivial example. The triviality of this example lies in considering a realization of the random process containing only $N = 2$ samples and is dictated by our limitations to visualize higher than a 3D surface. Let this process be sinusoid with random phase process. $x[n] = \sin(2\pi f n + \phi(\omega))$. Take n = [0,1] and calculate 2000 realizations of this random process. We will thus obtain a [2000 x 2] matrix with each row being a realization of the random variable.

The top panel of figure 2 shows the estimated histogram (approximation of the joint pdf $p(x_{1,} x_{2,} t_{1,} t_2)$) of the random process for $N = 2$ samples. Each realization (combination of x[1], x[2]) have different probabilities to occur as a result of sampling the experimental space. The left plot on the bottom panel shows colorcoded histogram (same as on the top panel) and shows two realizations of the random process. We can see that different realization have different probability as can be judged by the brightness of the corresponding element of the colorcoded histogram.

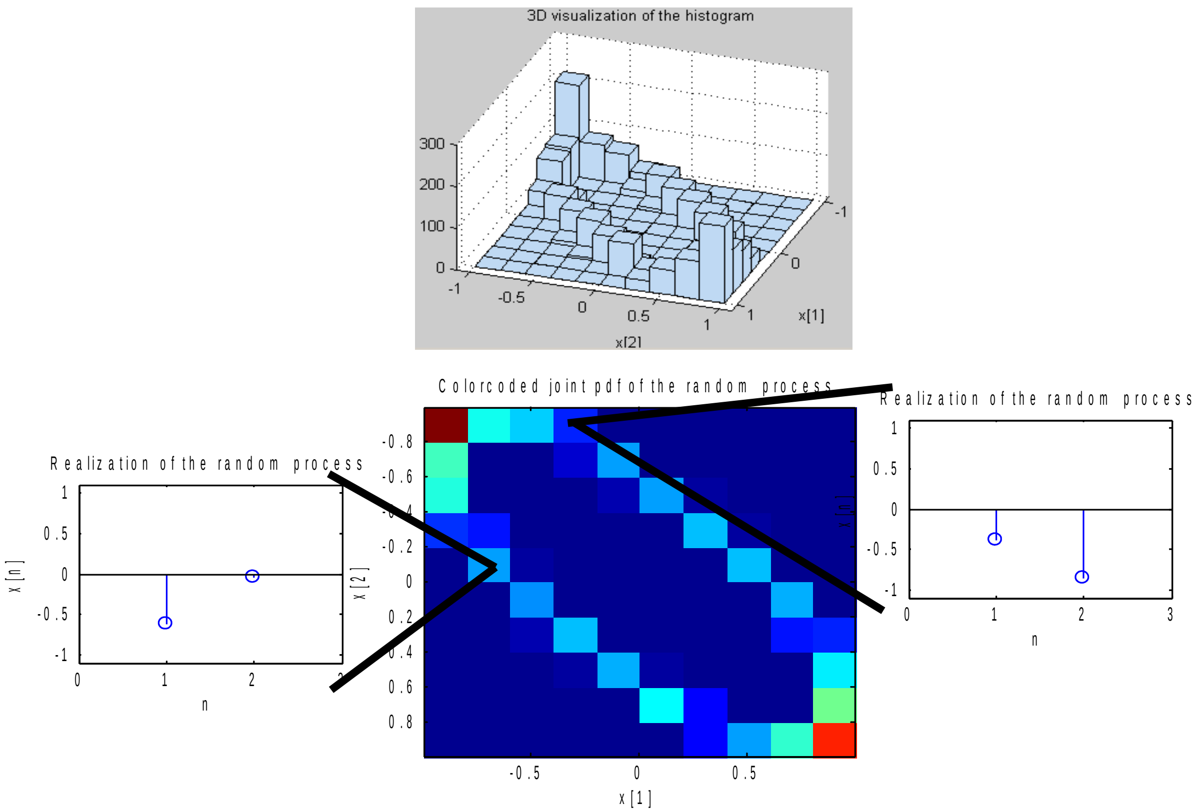

Figure 2: The top panel shows estimated histogram (approximation of the joint pdf) of the random process for N = 2 samples. Each combination of x[1], x[2] have different probabilities to be realized during the sampling of the experimental space. The left plot on the bottom panel shows colorcoded histogram (same as on the top panel) and shows two realizations of the random process. We can see that different realization have different probability as can be judged by the brightness of the corresponding element of the colorcoded histogram.

<u>Describing random process with moments</u>

Clearly, in general case specifying the distribution for longer realizations appears to be not really feasible. To alleviate this problem we can use time-dependent moments to describe statistical properties of the random process. Also, it is useful to know that the ubiquitous Gaussian distribution is completely defined by its first and second moments. Indeed,

The *k*-th moment of the random process is a function of time (or other carrier) is defined as the following integral

$$m_k(t)=E\{X^k(t)\}=\int_{-\infty}^{\infty} x^k p(x,t)\,dx \quad .$$

The first moment $\mu(t)=m_1(t)$ of a random process is called mean. There also exist so called central moments defined similarly but using the random process with subtracted mean

$$\mathring{m}_k(t)=E\{(X(t)-\mu(t))^k\}=\int_{-\infty}^{\infty}(x-\mu(t))^k p(x,t)\,dx$$

The most prominent central moment is the second order moment called variance.

Note, that for each moment of time, the probability distribution $p(x,t)$ may, in general, be different. This is not the case, however, for the sinusoid with random phase process as we can see from the histograms in figure 1 that are the approximations of the distributions for the two different time moments. The same can be seen from the actual expression of the process $x[n]=\sin(2\pi f n+\phi(\omega))$ in which random variable $\phi(\omega)$ is independent of time index *n*.

Also, note that for discrete time processes we replace *t* with *n*, corresponding to the *n*-th moment of time $\mathring{m}_k[n]=\mathring{m}_k(t_n)=E\{(X(t)-\mu(t_n))^k\}=\int_{-\infty}^{\infty}(x-\mu(t_n))^k p(x,t_n)dx$

The equations presented here are formal definitions of the moment and require knowledge of the pdf to be computed. The obtained values are not random. In real life we need to estimate the moments of the random process based on the ensemble of realizations of the random process. Unlike the actual moments, the obtained estimates are the functions of the random numbers and therefore are random themselves.

Imaging that we have collected *M* realizations $x_m[n]$ of a random process x. Each realization has duration of *N* time points (samples). To compute the estimate of the first moment, or in other words average, we just sum all of these M realizations and divide the sum by M. In other words $\hat{\mu}_x[n]=\frac{1}{M}\sum_{m=1}^{M}x_m[n]$, so we get the sequence $\hat{\mu}_x[n]$ that is an estimate of the first moment of the random process *x*. The hat on the top of the letter μ denotes that this is an estimate an not the exact value of the first moment.

Consider a set of evoked responses recorded during an EEG experiment. At each trial (or epoch) each EEG channel records a realization of the random process, called evoked response. When we perform the operation of averaging these evoked responses and obtain time varying average exhibiting some peaks and deeps we, in fact, estimate the first moment of the evoked response random process. In other words, the averaged evoked response is exactly the first moment of evoked response random process. Note that in this case, the original random process is a continuous process that we sampled in accordance with the sampling rate settings of our amplifier (usually 500 samples per second) and thus obtained the set of sequences – discretized in time realizations of this random process. The example of EEG and at each trial (epoch).

To estimate the variance of the random process we can use a similar averaging procedure i.e. $\hat{\sigma}_x^2[n]=\frac{1}{M-1}\sum_m^M(x_m[n]-\hat{\mu}[n])^2$. Note that this estimate is using the estimate of the mean and then is simply computing the average square of the deviation from the mean.

None of these two statistics tell us about the relation between the random variables being samples of the random process with different time indices, i.e. n_1 and n_2. To characterize this relation we will use so called autocorrelation function defined as

$$R[n_1,n_2]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mu[n_1])(x_2-\mu[n_2])\,p(x_1,x_2,n_1,n_2)\,dx_1dx_2$$

Each element of this autocorrelation function is simply a correlation between the values of the random process at the two different time moments.

We can estimate it from the data using the following expression

$$\hat{R}[n_1, n_2] = \frac{1}{M-1}\sum_{m=1}^{M}(x_m[n_1]-\hat{\mu}[n_1])(x_m[n_2]-\hat{\mu}[n_2])$$

As you can the estimation expression is simply estimating the average value of the product of centered values of the process. Also note that the autocorrelation function is in general a function of two variables, two indices n_1 and n_2 .

<u>Examples of analytic calculation of the moments</u>

Consider the sinusoid with random phase process $x[n]=\sin(\omega_0 n+\phi(\omega))$. Now, as prescribed by the definition lets write

$$\mu_x[n]=\int_0^{2\pi}\sin(\omega_0 n+\phi)\frac{1}{2\pi}d\phi=\frac{1}{2\pi}\int_0^{2\pi}\sin(\omega_0 n+\phi)d\phi$$
$$=\frac{-1}{2\pi}\cos(\omega_0 n+\phi)\Big|_0^{2\pi}=\frac{-1}{2\pi}(\cos(\omega_0 n+2\pi)-\cos(\omega_0 n+0))=0$$

The difference of cosines appears to be equal to zero because the arguments are 2π apart. Importantly, note that the mean of this process is not a function of time. This results should be expected by you based on examining the two distributions shown in figure 1. You can see that the distributions, that are the approximations of the pdf, are symmetric around zero.

Let's compute the variance remembering that $\mu_x[n]=0\ \ \forall n$.

$$\sigma_x^2[n]=\int_0^{2\pi}\sin^2(\omega_0 n+\phi)\frac{1}{2\pi}d\phi=\frac{1}{2\pi}\int_0^{2\pi}\frac{1-\cos(2\omega_0 n+2\phi)}{2}d\phi$$
$$=\frac{1}{4\pi}\phi\Big|_0^{2\pi}-\frac{1}{8\pi}\sin(2\omega_0 n+2\phi)\Big|_0^{2\pi}=\frac{1}{2}$$

Note that the second term in the final summation vanishes. Also, note that similarly to the mean of this process, the variance does not depend on the time index value and is identical for all time moments.

Now, let's compute the autocorrelation function for $x[n]=\sin(\omega_0 n+\phi(\omega))$

$$R_x[n_1, n_2]=\int_0^{2\pi}\sin(\omega_0 n_1+\phi)\sin(\omega_0 n_2+\phi)\frac{1}{2\pi}d\phi=\frac{1}{4\pi}\int_0^{2\pi}\cos(\omega_0(n_1-n_2))d\phi+$$
$$+\frac{1}{4\pi}\int_0^{2\pi}\cos(\omega_0(n_1+n_2)+2\phi)d\phi=\frac{1}{4\pi}\cos(\omega_0(n_1-n_2))\phi\Big|_0^{2\pi}=\frac{1}{2}\cos(\omega_0(n_1-n_2))$$

As we can see the autocorrelation for this process does not depend on the absolute values of the time indices but only on the distance between them $d=n_1-n_2$. This value *d* is usually called lag. Figure 3 shows the autocorrelation for a range of lags. Note that the value of the autocorrelation function corresponding to lag *d = 0* is exactly the variance of the random process, i.e. $\sigma_x=C[0]=1/2;$ The same can be concluded from analysis of the definition for autocorrelation function by substituting $n_1=n_2=n$. What else does the autocorrelation sequence tell us? For example, the periodicity of maxima with T = 10 samples tells us that if we know the value of some particular *m*-th sample for a realization at hands then the realization will exhibit exactly the same values for all samples with indices m+kT. This is obvious since the process we have at hands is a sinusoid with random phases and is peridoci with period T = 10 samples. Very often, analysis of the autocorrelation sequence of some measured realization of a random process allows to reveal hidden regularities present in the data and not visible wih "bare eye".

Exercise 1.1

Put together a simple Matlab program that plots autocorrelation sequence of $x_k[n]=\cos(\frac{2\pi}{20}n+\phi(\omega))$ with uniformly distributed $\phi(\omega)$ in -pi:pi range for the range of lags d =*-40:40.*

Exercise 1.2

Imaging that the autocorrelation sequence of a random process *x*[*n*] is given by $R_x[d]=ae^{-bd}$. Calculate the second moment for random variable $y=x[4]-x[2]$ Put together a simple Matlab program that plots autocorrelation sequence of $x_k[n]=\cos(\frac{2\pi}{20}n)$ for the range of lags d =*-40:40.*

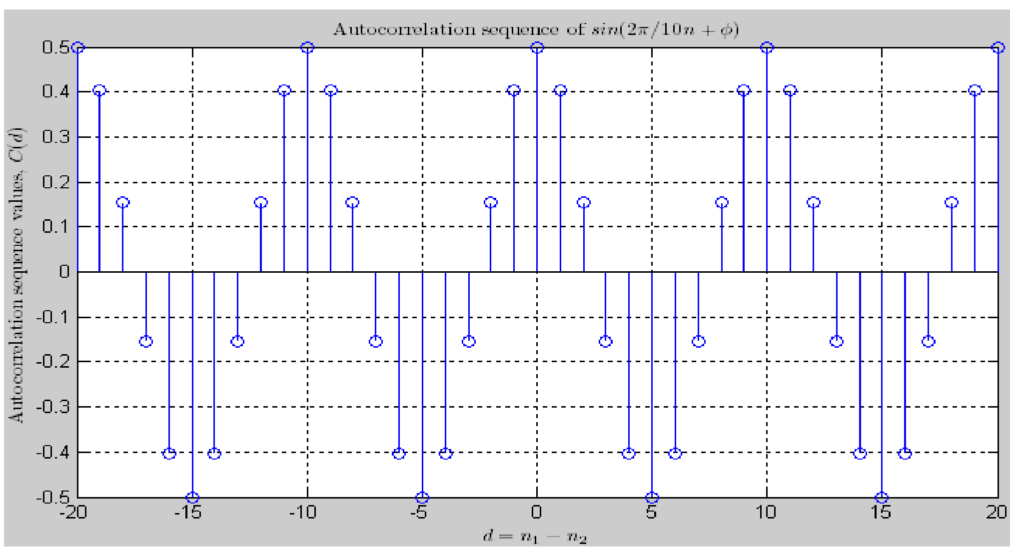

Figure 3: Autocorrelation sequence of $x[n]=\sin(2\pi/10n+\phi)$ *discrete time random process.*

<u>Stationary processes</u>

Stationary process is an important class of processes that significanly simplify our life as researchers and data analysts. A stationary random process is the random process whose joint probability density function does not change with time. This means, that the moments of any order, since they are dependent on the joint pdf, do not change with time. This is a very strict characteristics. Softer characteristics involves requirement for the processes up to order 2 to be not dependent on time. In other words if the mean is constant over time and the autocorrelation sequence depends only on the difference of time indices then the process is called wide sense stationary (WSS). Most of the time we will be dealing with WSS processes.

Ergodic processes

The strict definition of ergodicity is very very technical. However, since the ergodicity is a very important property that oftentimes implicitly assumed by many signal processing algorithms, we will attempt to define it here. Consider the following situation. You have obtained a record of data. It could be, for example a record of resting state EEG. Having explored the data visually, you are see that the record is nice and clean. Now you want to calculate the mean, the variance and the autocorrelation sequence of your data. Theoretically, what you have in hands is just a single realization of the random process and in order to employ the general estimator you need to collect many more such realizations. One strategy that may come to you mind is to take you data record split in some significant number of segments and then use these segments as you realizations. Essentially, you are replacing the averaging over ensemble by averaging over time. To be able to legitimately do it the process at hands have to be **ergodic**.

Estimation of autocorrelation sequence from a realization of an ergodic process

A practical estimator of the autocorrelation sequence is shown in the figure 4. The corresponding expression is $\hat{R}[k]=\frac{1}{N-k}\sum_{i=k}^{N} x[i]x[i-k]$.

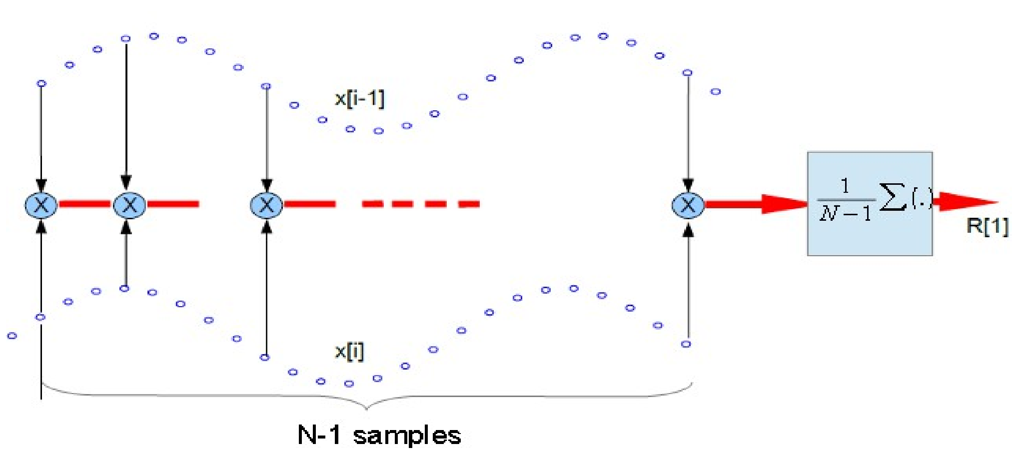

Figure 4: Calculating the first lag of the autocorrelation function for an ergodic random process. The top sequence is shifted by 1 sample with respect to the bottom sequence. This automatically aligns the samples that are 1 time slice apart. Sample by sample multiplication (for all but two samples on both extremities that do not have the pair) and calculating the average of this product results into the estimate of the first lag of the autocorrelation function. For simplicity, this example assumes that te process is zero mean and therefore we used the original and not centered (mean subtracted) timeseries.

Exercise 1.3
Write an expression and draw a diagram for calculating the *k*-th lag of the auto**covariance** sequence.