

# chapter-6

## Introduction:

Signals are subdivided into two classes, namely,

- ❑ Deterministic signals
- ❑ Random signals

Signals that can be modeled exactly by a mathematical formula are known as deterministic signals. Deterministic signals are not always adequate to model real-world situations. Random signals, on the other hand cannot be described by a mathematical equation; they are modeled in probabilistic terms. In this chapter we shall use the power of MATLAB to describe some fundamental aspects of random signals.

## Random number generation:

Random numbers conforming to particular density functions can be generated via MATLAB. MATLAB has two built-in functions to generate pseudo-random numbers, namely **rand** and **randn**. The **rand** function produces pseudo-random numbers uniformly distributed between zero and one, and the function **randn** generates pseudo-random numbers selected from a normal distribution with zero mean and unity variance (standard normal). Both functions have the same syntax. For example, **rand(n)** returns a n-by-n matrix of random numbers, **rand(n,m)** returns a n-by-m matrix with randomly generated entries distributed uniformly between 0 and 1., and **rand(1)** returns a single random number.

## Practice:

```
>>%Generate one thousand uniform pseudo-random numbers
>>rand(1,1000)           %return a row vector of 1000 entries
>>%Generate one thousand gaussian pseudo-random numbers
>>randn(1,1000);        %return a row vector of 1000 entries
```

In some applications, it is desirable to generate random integers. The following code illustrates one possible way of generating 5 uniform pseudo-random integers over the interval (-20,20).

```
>>x=fix(40*rand(5,1))-20      %generate 5 uniform integers
```

x =

```
    18
   -11
     4
    -1
    15
```

## Randomization:

It is a good practice to provide a **seed** before using the **rand** command; otherwise you will get the same sequence of random numbers every time you restart the computer. A good way to reseed the random number generator is to use the clock, as follows:

```
>>rand('state', sum(100*clock));
```

### **Histogram construction:**

When the probability density function (pdf) is not available, it can be estimated using a **histogram**. A histogram is constructed by subdividing the interval [a,b] containing a collection of data points into sub-intervals known as **bins** and count for each bin the number of the data points that fall within that bin. The function **hist** provides the histogram of sample values of a random variable.

### **Syntax:**

```
>>[n,y]=hist(x,N);
```

The function **hist** divides the interval [min(x), max(x)] into N bins and yields the output [n,y], where n is a vector whose elements are the number of samples in each bin, and y is a vector whose elements are the centers of the bins. When used in this manner, the **hist** function does not produce a graph; instead we use the **bar** function.

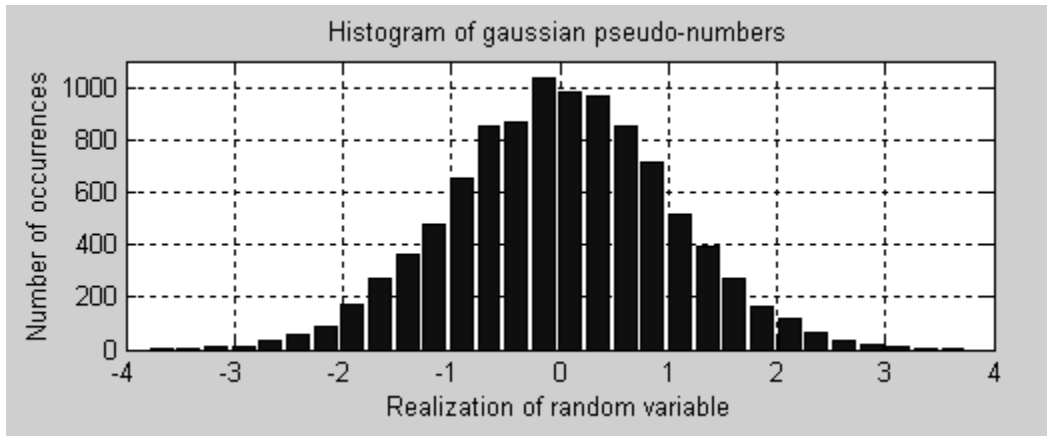
### **Bar graphs:**

Bar graphs are a good way of examining trends (rising or falling) in one or more variables over a period of time. MATLAB bar graphs can be created to plot either vertically or horizontally. The **bar** function produces a bar graph which for each value of y, there is a bar whose height is proportional to n.

### **Practice:**

Generate 10000 gaussian distributed pseudo-random numbers, and then plot the histogram.

```
>>%Generate 10000 gaussian pseudo-random numbers and draws a histogram
>>x=randn(1,10000);           %generate a random vector
>>N=30;                       %specify the number of bins
>>[n,y]=hist(x,N);           %y=vector of centers of bins; n=centers of bins
>>bar(y,n)                   %plot histogram using the bar function
>>xlabel('Realization of random variable')
>>ylabel('Number of occurrences')
```



Practice:

Find the approximate distribution of two resistors in a parallel connection assuming that they each have measured values, which vary uniformly about their nominal values by  $\pm 5\%$ .

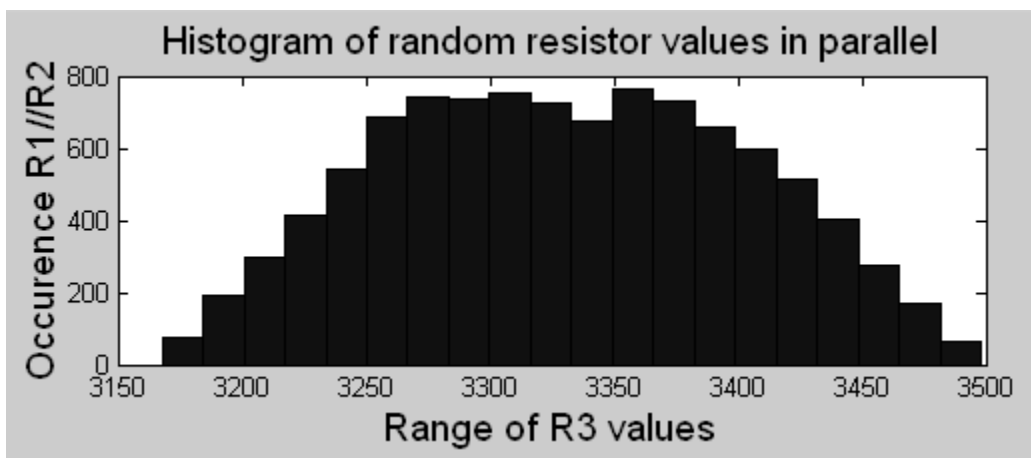
$$R_1 = 10k\Omega$$

$$R_2 = 5k\Omega$$

$$R_{eq} = R_1 // R_2 = \frac{R_1 R_2}{R_1 + R_2}$$

Compute 10000 trials and histogram the results.

```
n=10000;
r1=rand(n,1)*(10500-9500)+9500;
r2=rand(n,1)*(5250-4750)+4750;
r3=r1.*r2./(r1+r2);
subplot(2,1,1);hist(r3,20)
title('Histogram of random resistor values in parallel','FontSize',14);
ylabel('Occurrence R1//R2','FontSize',14)
xlabel('Range of R3 values','FontSize',14)
```



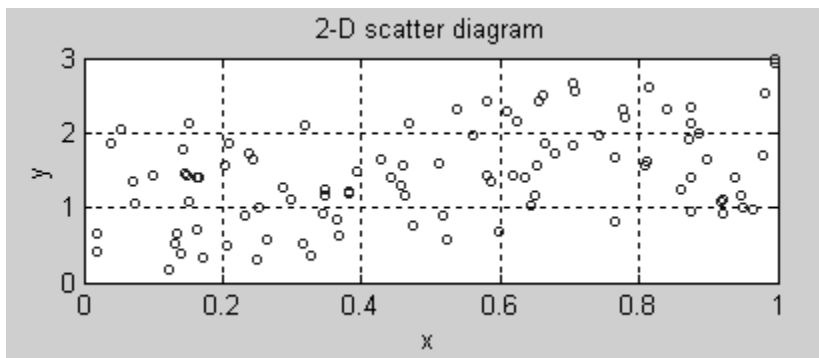
**Scatter diagrams (or scattergram) & correlation coefficient:**

The **scatter diagram** is a useful tool for identifying a potential relationship or correlation between two variables. Correlation implies that as one variable changes, the other also changes. Sometimes if we know that there is good correlation between two variables, we can use one to predict the other. The first step in determining the presence of a relationship between two variables is to plot the scattergram (graph of the observed data).

The **correlation coefficient** is a measure of the degree of linear relationship that exists between two variables.

Practice:

```
>>x=rand(100,1);
>>y=x+rand(100,1);
>>plot(x,y,'o','MarkerSize',2)           %2-D scatter plot
>>corrcoef(x,y);
>>z=randn(100,1);
>>plot3(x,y,z,'.');                     %3-D scatter plot
```



**Mean, standard deviation, and median:**

Many times we wish to characterize the probability density function (pdf) with a few numbers. The **mean** is a measure of the center or most likely value of a distribution. The **variance** (or **standard deviation**) is a measure of dispersion (spread), and the **median** is also a measure of the center. The median is less sensitive to extreme scores (**outliers**) than the mean. The MATLAB commands **mean**, **std**, and **median** determine the sample mean, standard deviation, and median, respectively. The standard deviation is measured in the same units as the mean and the median.

Practice:

```
>>x=randn(1,10000);                    %generate gaussian numbers
>>[mean(x); std(x); median(x)]         %compute the mean, standard deviation, and median
```

```
ans =
    0.0066
    1.0036
    0.0098
```

It is a straightforward matter to simulate from any normal distribution with a specified mean value and a specified standard deviation. In MATLAB one can produce normally distributed numbers with mean zero and a standard deviation of unity directly using the function **randn**. To produce random numbers from a gaussian distribution of mean **m** and a standard deviation of **sd**, proceed as follows:

```
>>r=randn; %gaussian number: mean zero, standard deviation unity
>>z=m+r*sd; %gaussian number: mean m, standard deviation sd.
```

The **rand** function generates random numbers uniformly distributed from zero to one. Numbers uniform on the interval [0,1] can be transformed to numbers uniform on [a,b] using the following transformation:

```
>>r=rand; %uniform number in [0,1]
>>x=(b-a)*r+a; %uniform number in [a,b]
```

### Gaussian probabilities:

For a gaussian random variable X with mean m and standard deviation s, the cumulative distribution function is given by

$$F_X(x) = P\{X \leq x\} = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x-m}{s\sqrt{2}}\right)$$

where erf is the error function. MATLAB has a built-in error function erf defined by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

The following results are important when evaluating gaussian probabilities:

$$P\{x_1 < X \leq x_2\} = \frac{1}{2} \operatorname{erf}\left(\frac{x_2-m}{s\sqrt{2}}\right) - \frac{1}{2} \operatorname{erf}\left(\frac{x_1-m}{s\sqrt{2}}\right)$$

$$P\{X > x\} = 1 - P\{X \leq x\} = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{x-m}{s\sqrt{2}}\right)$$

### Practice:

A gaussian voltage has a mean value of 5 and a standard deviation of 4.

1. Find the probability that an observed value of the voltage is greater than zero.
2. Find the probability that an observed value of the voltage is greater than zero but less than or equal to 5.

```
>>% Answer to part (1)
>>m=5; s=4; x=0; %specify the mean, standard deviation, and x
>>z=(x-m)/(s*sqrt(2)); %define an intermediate variable z
>>P1=1/2-1/2*erf(z) %compute P(X>0)
>>%Answer to part (2)
>>x1=0; x2=5; %define the values of x1 and x2
>>z1=(x1-m)/(s*sqrt(2)); z2=(x2-m)/(s*sqrt(2)); %define two intermediate variables
```

$P2=1/2*\text{erf}(z2)-1/2*\text{erf}(z1);$

`%compute the probability  $P\{0 < X \leq 5\}$`

### White noise:

#### □ Uniform white noise:

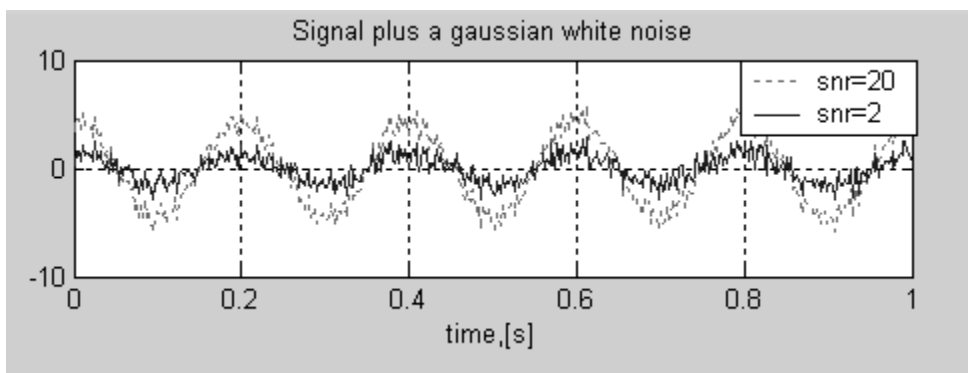
A uniform white noise is a sequence of independent samples with zero mean. The **rand** function generates a sequence of uniform pseudo numbers with mean of 0.5 and variance of 1/12. Therefore, the average power of (rand-0.5) is 1/12. A uniform white noise with a specific average power P can be generated using  $\sqrt{12P}(\text{rand} - 0.5)$ .

#### □ Gaussian white noise:

Similarly, the function **randn** provides a gaussian sequence with zero mean and a variance of unity. Therefore, one can generate a white gaussian noise having an average power P via  $\sqrt{P}\text{randn}$ .

### Practice:

```
>>%Signal-to-noise ratio=2
>>t=[0:512]/512; %define a time vector
>>signal=sqrt(2)*cos(2*pi*5*t); %define a signal sequence (average power=1 W)
>>noise=sqrt(0.5)*randn(1,length(t)); %define a noise sequence (average power=0.5 W)
>>sn=signal+noise; %compute the signal+noise sequence
>>subplot(2,1,1);plot(t,sn);grid %plot the signal+noise sequence
>>%Signal-to-noise ratio=20
>>signal2=sqrt(20)*cos(2*pi*5*t); %define a signal sequence (average power=10 W)
>>sn2=signal2+noise; %computer the signal+noise sequence
>>subplot((2,1,2);plot(t,sn2);grid %plot the signal+noise sequence
>>legend('snr=20','snr=2'); %add legend to plot
```



### Some useful functions:

sort	reorders elements of a vector to ascending order
sum	computes the summation of a vector x
max	finds the largest entry of the vector x
min	finds the smallest entry of the vector x
randint	generates matrix of uniformly distributed random integers
norm(x, arg)	computes the norm of a matrix or a vector

note:

*The commands max and min return the maximum and minimum values of an array, and with a slight change of syntax they will also return the indices of the array at which the maximum and minimum occur.*

```
>>[y,k]=max(y)           %y=max; k=index
>>[y,k]=min(y)          %y=min; k=index
```

**Spinning coins:**

When a fair coin is spun, the likelihood of having heads or tails is 0.5. Since a value returned by the function **rand** is equally likely to be anywhere in the interval [0,1), we can represent heads, say, with a value less than 0.5, and tails otherwise.

Practice:

Simulate spinning a fair coin 30 times.

```
N=input('Enter the number of simulations: ')
for k=1:N
    rand_number=rand;
    if random_number<0.5
        fprintf(' H ')
    else
        fprintf(' T ')
    end
end
fprintf('\n')
```

Practice:

Write a script function that will simulate n throws of a pair of dice

This entails the generation of random integers in the range 1 to 6. This can be accomplished as follows:

```
>>floor(1+6*rand);
```

```
function r=dice(n)
%simulate n throws of a pair of dice
%Input: n, the number of throws
```

```

%Output: an n-by-2 matrix, each row corresponds to one throw
%Usage: r=dice(3)
r=floor(1+6*rand(n,2));
%end of dice

ans =

     1     1
     3     2
     5     2
     1     4

>>sum(dice(100))/100           %compute average value over 100 throws

```

Practice:

The average information measure of a digital source is defined by

$$H = -\sum_{k=1}^n P_k \log_2(P_k) \text{ bits}$$

where n is the number of possible distinct source messages and  $P_k$  is the probability of sending the k-th message. This average information is often known as the source entropy.

A digital source puts out -1.0 V and 0.0 V levels with a probability of 0.2 each and 3.0 V and 4.0 V levels with a probability of 0.3 each. Evaluate the entropy of the source.

```

>>p=[0.2 0.2 0.3 0.3];           %define the vector of probabilities
>>H=-sum(p.*log2(p))           %compute the entropy

H =

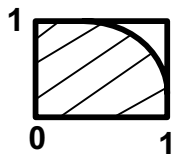
    1.9710

```

Practice:

Monte Carlo approximation of  $\pi$ .

Consider the following diagram of quarter unit circle inside a unit square. The ratio of the area of the quarter circle to the area of the square is  $\pi/4$ , so if we pick a point in the square at random, the probability of it landing within the circle is  $\pi/4$ . This idea provides a simple way of approximating  $\pi$ , by finding a sample proportion of random points falling in the circle and multiplying by 4.





Write a MATLAB script to estimate the value of pi.

```
function estimate=randpi(n)
s=0;
for k=1:n
    if(rand^2+rand^2<=1)
        s=s+1;
    end
end
estimate=4*(s/n);
```

### **The cumsum function:**

The **cumsum** function creates a vector in which each element is the cumulative sum of all the elements up to and including the comparable position in the original vector.

#### Practice:

Let x be a vector of values, sorted in ascending order, and p a vector of the probabilities associated with each of the corresponding values. As depicted in table below

x	10	20	30
p	0.20	0.30	0.50

```
>>x=[10 20 30];
>>p=[0.20 0.30 0.50];
>>y=cumsum(p)
```

y =

0.2000 0.5000 1.0000

### **Evaluation of binomial coefficients:**

$$C_k^n = \frac{n!}{k!(n-k)!}$$

If this expression is used, the factorials can get very big, causing an overflow. This can be avoided by using the following procedure:

```
nck=1;
n=input('Enter the value of n: ');
k=input('Enter the value of k: ');
for m=1:k
    nck=nck*(n-m+1)/m;
end
disp(['nck= ', num2str(nck)])
```

#### Practice:

The probability that n trials will result exactly in k successes in a Bernoulli trial is given by

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

Write a script file that computes recursively (using odds ratio relation) the binomial probabilities. Try plotting the probability mass function (PMF) and the cumulative distribution function.

```
close all
clc
%Evaluate binomial probabilities recursively
n=input('Enter the number of trials: ');
p=input('Enter the probability of success: ');
% q=probability of failure (p=1-q)
q=1-p;
y(1)=q^n;
for i=1:n
    oddsratio=((n-i+1)/i)*(p/q);
    y(i+1)=y(i) .* oddsratio;
end
k=0:n;
subplot(2,1,1);h=stem(k,y,'filled');grid
set(h,'LineWidth',2);
set(gca,'FontSize',10);
title('Binomial probabilities')
ylabel('P(X=k)')
xlabel('Number of successes [k]')
mean_value=sum(k.*y)
z=cumsum(y);
disp([' k ' ' P(X=k)' ' CDF'])
disp('-----')
disp(['k' y' z'])
subplot(2,1,2);
h=stairs(k,z);grid
axis([0 n 0 1.2])
set(h,'LineWidth',3)
set(gca,'YTick',0:0.2:1.2)
set(gca,'XTick',1:n)
ylabel('Cumulative probability')
```

Enter the number of trials: 12  
Enter the probability of success: 0.3

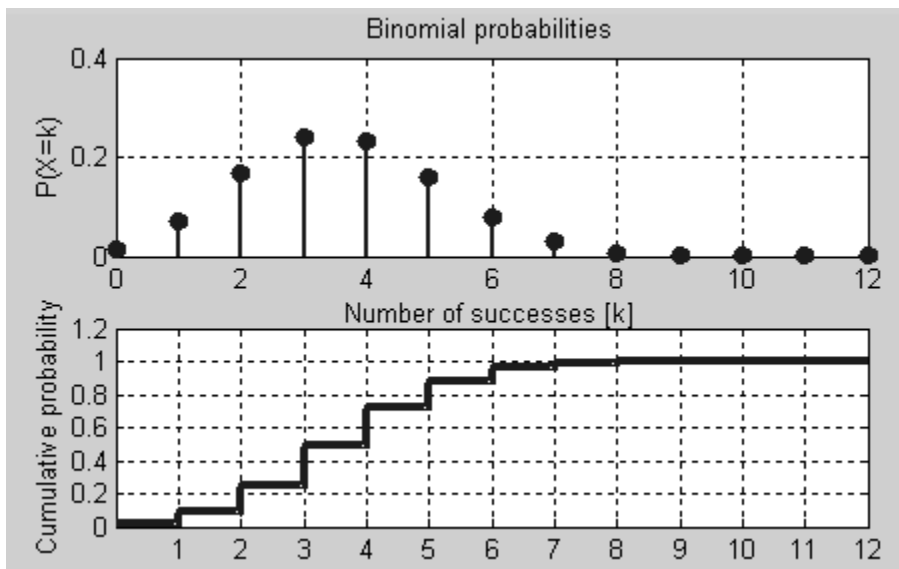
mean\_value =

3.6000

k	P(X=k)	CDF
---	--------	-----

0	0.0138	0.0138
1.0000	0.0712	0.0850
2.0000	0.1678	0.2528
3.0000	0.2397	0.4925
4.0000	0.2311	0.7237
5.0000	0.1585	0.8822
6.0000	0.0792	0.9614
7.0000	0.0291	0.9905
8.0000	0.0078	0.9983
9.0000	0.0015	0.9998
10.0000	0.0002	1.0000
11.0000	0.0000	1.0000
12.0000	0.0000	1.0000

Figure below shows the associated cumulative probability distribution. Note that it is a staircase function, reflecting the discrete nature of the outcomes.



### **Correlation:**

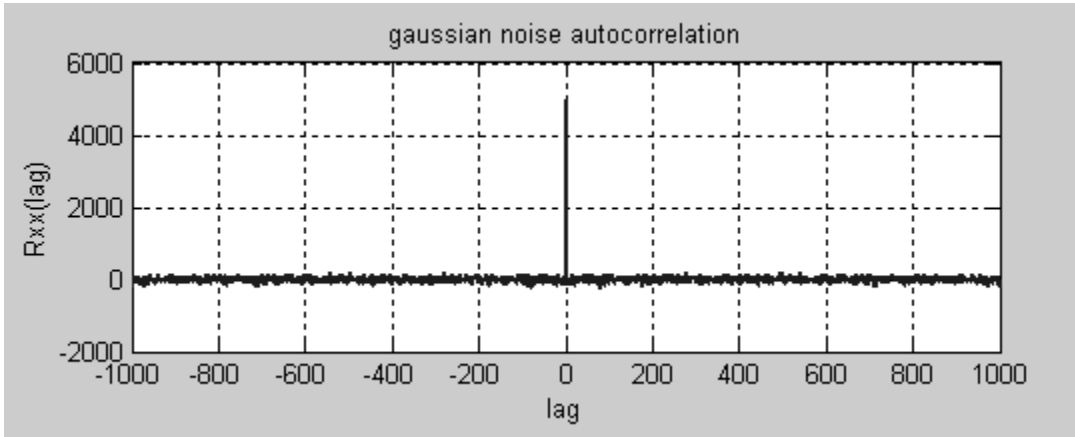
The **cross-correlation** between two signals tells how “identical” the signals are. If there is correlation between the signals, then the signals are more or less dependent on each other. **Auto-correlation** means the cross-correlation of a signal with itself.

### **Practice:**

Determine the auto-correlation of a gaussian noise.

```
noise=randn(5000,1);
[acf,lag]=xcorr(noise,noise,1000);
subplot(2,1,1); plot(lag,acf);grid
title('gaussian noise autocorrelation')
```

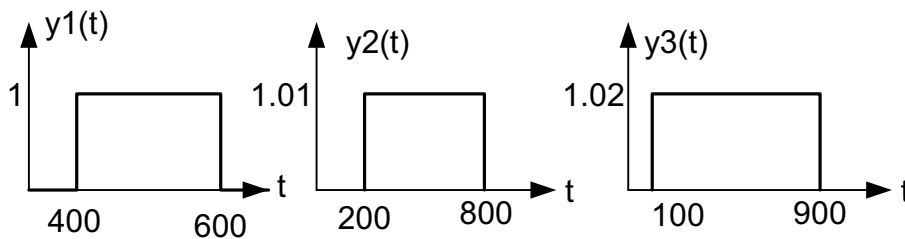
```
ylabel('Rxx(lag)')
xlabel('lag')
```



*Note: `xcorr(x,y, 'coeff')` is the same as `xcorr(x,y)` but with the maximum set to 1.0.*

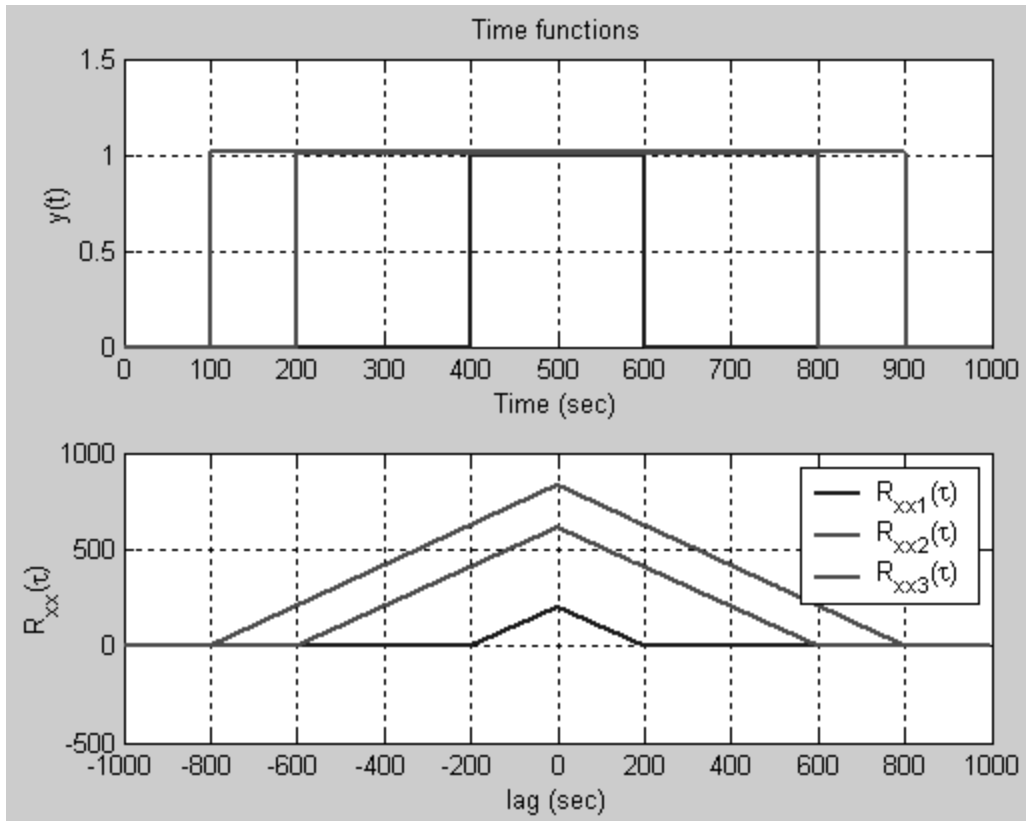
Practice:

Determine and plot the autocorrelation functions of the following rectangular functions.



```
%specify time functions;
N=1000;
x=zeros(N,1);
y1=x; y1(400:600)=1;
y2=x; y2(200:800)=1.01;
y3=x; y3(100:900)=1.02;
t=1:length(x);
subplot(2,1,1);plot(t,y1,t,y2,t,y3,'LineWidth',2);grid
title('Time functions')
xlabel('time (sec)')
ylabel('y_i(t)')
%Determination of autocorrelation function
acf1=xcorr(y1);
acf2=xcorr(y2);
acf3=xcorr(y3);
tau=(-(N-1):(N-1));
subplot(2,1,2);plot(tau,acf1,tau,acf2,tau,acf3,'LineWidth',2);grid
ylabel('R_{xx}(\tau)')
legend('R_{xx1}(\tau)', 'R_{xx2}(\tau)', 'R_{xx3}(\tau)')
```

```
xlabel('lag (sec)')
axis([0 2000 -100 900])
```

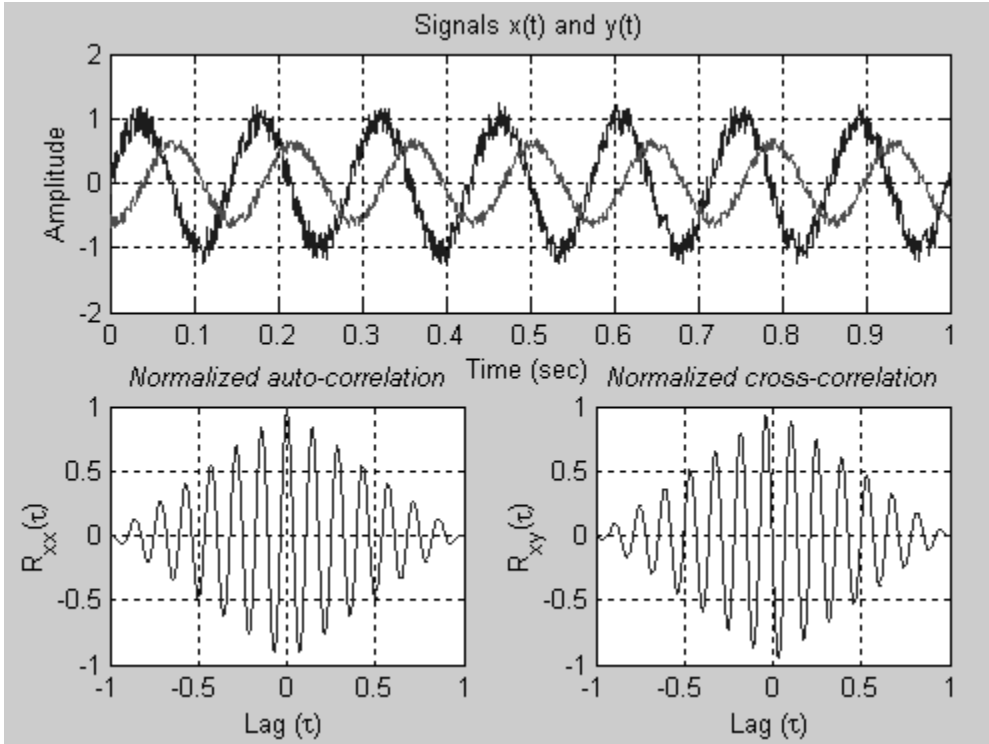


Practice:

%Auto-correlation and cross-correlation functions

```
Fs=1000;
f=7;
t=0:1/Fs:1;
noise=rand(size(t));
noise=noise-mean(noise);
x=sin(2*pi*f*t)+0.5*noise;
y=0.6*sin(2*pi*f*(t-0.04))+0.2*noise;
subplot(2,1,1); plot(t,x,t,y);grid
xlabel('Time (sec)')
ylabel('Amplitude')
title('Signals x(t) and y(t)')
tau=-1:1/Fs:1;
acf=xcorr(x,'coeff');
subplot(2,2,3);plot(tau,acf);grid
xlabel('Lag (\tau)')
ylabel('R_{xx}(\tau)')
title('\itNormalized auto-correlation')
ccf=xcorr(x,y,'coeff');
```

```
subplot(2,2,4);plot(tau,ccf);grid
xlabel('Lag (\tau)')
ylabel('R_{xy}(\tau)')
title('Normalized cross-correlation')
```



**Power spectral density (PSD)**

The power spectral density describes how the power of a time series is distributed across frequency. Different algorithms are used for the estimation of PSD, some of which are:

- Periodogram method
- Welch’s method
- Maximum entropy method

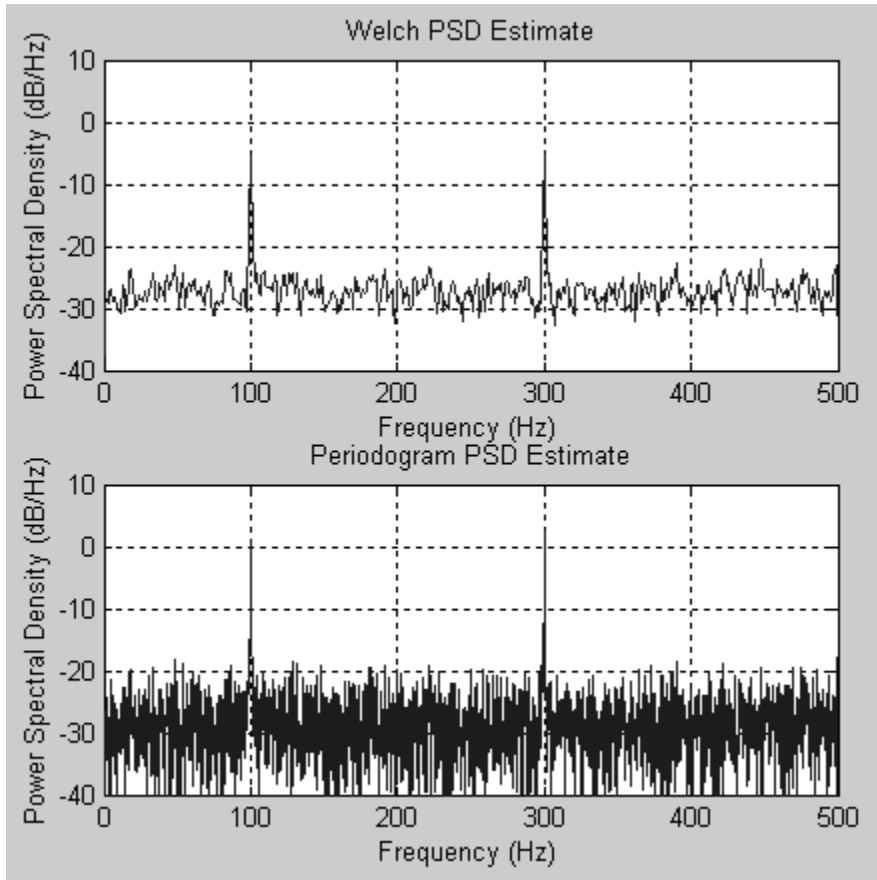
The periodogram calculates a straight fft-based PSD, while the Welch method averages several sub-spectra to give a smoother estimate of the PSD.

Practice:

Find and sketch the PSD of a signal consisting of two tones added to white noise.

```
Fs=1000; %sample rate
time=0:1/Fs:4; %time base
noise=randn(size(time)); %white gaussian noise
signal=sin(2*pi*100*time)+sin(2*pi*300*time)+noise; %signal+noise
```

```
subplot(2,1,1);pwelch(signal,[],[],Fs);           %psd estimate
subplot(2,1,2);periodogram(signal,[],[],Fs);     %psd estimate
```



### **Signal-to-Noise Ratio:**

In the process of information transmission, one of the nasty things that happens to a signal is that it is corrupted by additive noise. A measure of the extent of corruption is the signal-to-noise ratio or SNR. This is the ratio of the mean-square value of the signal to the mean-square value of the noise, expressed typically in dB.

### **Practice:**

```
Fs=1000;
t=0:1/Fs:4;
wave=5*sin(2*pi*200*t);
noise=randn(size(wave));
pwave=mean(wave.^2);
disp(['mean-square value of signal is: ',num2str(pwave)])
pnoise=mean(noise.^2);
disp(['mean-square value of noise is: ',num2str(pnoise)])
disp(['signal-to-noise ratio: ', num2str(10*log10(pwave/pnoise)), ' dB'])
```

```
filt=butter(10,0.3);  
output=filter(filt,1,noise);  
subplot(2,1,1);plot(t,noise);grid  
subplot(2,1,2);plot(t,output);grid
```

Sample output:

mean-square value of signal is: 12.4969  
mean-square value of noise is: 1.0019  
signal-to-noise ratio: 10.9596 dB