White noises, Moving Averages, Tests for Randomness

**Goals:**
Learn how to use \texttt{R} to manipulate with white noises and moving averages, and how to test time series for randomness.

**Assignments:** There are three assignments formulated in the handouts.

**Report:** All assignments listed below require printed report. All assignments involve “research” questions, which by their nature do not assume a single correct answer. Your goal is to support your conclusions using theoretical and/or practical arguments using \texttt{R} (do not forget about nice graphs.)

**Due date:** Wednesday, September 27, 2006 in the class.

\texttt{R} commands used in these handouts are collected in R-file \texttt{Lab-2.R}, which is available on the course Web site [http://unr.edu/homepage/zal/STAT758_Fall06.htm](http://unr.edu/homepage/zal/STAT758_Fall06.htm)
1 White noises

A sequence of uncorrelated rvs with mean zero and variance $\sigma^2$ is called white noise. Any sequence of iid rvs is white noise, but not conversely (recall HW 1.5). Below are examples of white noise simulations (use help to learn about the details of these commands and find similar ones):

\>
\> e1<-rnorm(n=500,mean=0,sd=1)
\>
\> e1<-rnorm(500) (a short version of the same command)
\>
\> e2<-rexp(n=500,rate=2)
\>
\> e2<-(e2-0.5)

Now $e_1$ and $e_2$ are white noise (in fact, iid) samples of length 500. Next, we will create dependent white noises.

\>
\> tmp<-rnorm(501)
\>
\> tmp1<-tmp[1:500]
\>
\> tmp2<-tmp[2:501]
\>
\> e3<-tmp1*tmp2
\>
\> rm(tmp1,tmp2)
\>
\> e4<-rnorm(501)
\>
\> even<-seq(2,500,by=2)
\>
\> odd<-seq(1,500,by=2)
\>
\> e4[even]<-(e4[odd]^2-1)/sqrt(2)

To check that our samples $e_i, i = 1, \ldots, 4$ are indeed white noises use the command

\>
\> acf(e1,lag=20)

The confidence lines are drawn by default at $\pm 1.96 \sqrt{1/N}$. Recall that for large $N$, the distribution of sample autocorrelation coefficients can be well approximated by $N(0, 1/N)$. How many observations would you expect to lie outside the confidence lines? How many actually do? You can plot the autocorrelation function at different lags using the option `lag`. Play with it, count the numbers of observations outside the confidence intervals, discuss.

Let us reveal now the dependence among members of $e_3$ and $e_4$.

\>
\> acf(e3^2,lag=20)
\>
\> acf(e4^2,lag=20)

You can do the same with $e_1$ and $e_2$. What do you observe? Play with `lag` and discuss. Sometimes, a better way to reveal dependencies is to plot the data:

\>
\> plot(e4[1:499],e4[2:500])

Do the same with $e_3$. What do you see? Try the following approach and discuss:

\>
\> plot(log(abs(e3[1:499])),log(abs(e3[2:500])))
2 MA($p$) models

Here we work with moving averages (linear filters) of white noises. Filters can be causal (output does not depend on the future values of filtered sequence) or non-causal (output depends on the future values of filtered sequence). We start by causal filtering of $e_1$.

$$f_1 <- \text{filter}(e_1, c(.5,.5), \text{side}=1)$$

This is nothing but a moving average of the sequence $e_1$, this is why this model is also called Moving Average, or MA(1), which reflects the fact that there is one-step dependence in the resulting series. Similarly, we can create MA($p$) series with arbitrary $p \geq 1$:

$$> f2 <- \text{filter}(e1, c(.2,.3,.5), \text{side}=1)$$
$$> f3 <- \text{filter}(e1, c(.1,.2,.3,.4), \text{side}=1)$$

Here, the vector $c(...)$ specifies the weights $\{w_0, ..., w_n\}$ of the filter and the resulting series is given by

$$f_t = \sum_{i=0}^{n} w_i e_{t-i}$$

Assignment 1: Simulate MA($p$) models with different values of $p$ and analyze their acf. What general conclusion can you make about the autocorrelation function of a MA($p$) process? Show analytically that your conclusion is true.

3 Tests for randomness

We start with plotting both the original series $e_1$ and the filtered one $f_1$ in the same window:

$$> \text{split.screen}(c(2,1))$$
$$> \text{plot}(e1, \text{type}='o', \text{pch}=20)$$
$$> \text{screen}(2)$$
$$> \text{plot}(f1, \text{type}='o', \text{pch}=20)$$
$$> \text{close.screen}(\text{all=TRUE})$$

Can you see the difference in the dynamics of the two series? Would you be able to tell which is white noise and which is not just by looking at the plot?

There exist formal methods for detecting internal structure within time series. One of them is indeed the autocorrelation function. (Plot acf for both time series and discuss what you see.) Another approach is to use statistical tests that are designed to decide whether the series is white noise or not. One of such tests looks for lengths of consecutive positive and negative observations (called runs):

$$> \text{runs.test(factor(sign(e1)))}$$
$$> \text{runs.test(factor(sign(f1)))}$$

Another test with the same purpose is Box-Pierce or Ljung-Box test, which consider the sum of first squared autocorrelation coefficients (the number of coefficients to consider is given by lag):

$$> \text{Box.test}(e1, \text{lag}=5, \text{type}='\text{Box-Pierce}')$$
> Box.test(e1,lag=5,type='Ljung-Box')
> Box.test(f1,lag=5,type='Box-Pierce')
> Box.test(f1,lag=5,type='Ljung-Box')

**Assignment 2:** Perform Ljung-Box test for different sequences (white noises and MA models) changing the value of `lag`. What general recommendation can you make about the value of `lag` to be used in practice, why?

**Assignment 3:** What test (we consider three: acf, runs, and Box-Pierce-Ljung) is the best for practical purposes? Why? What is your personal (and may be subjective) feeling about the usefulness of the formal tests to check for dependence among observations vs. visual analysis of time series plots?