Unit 7: Smoothing

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2016

Readings for Unit 7

Textbook chapter 2.4.

Last Unit

1. Periodic functions
2. Exploratory data tools to access frequency

This Unit

Smoothing techniques:

1. Averaging
2. Kernel Smoothing
3. Nearest Neighbors Regression

Motivation

Sometimes, the time series data we have can be too noisy to be able to detect long term trends. Smoothing is used to smooth out short term random fluctuations so that longer term trends can be emphasized.

1. Averaging
2. Kernel Smoothing
3. Nearest Neighbors Regression

So Far...

So far, we have looked to fit models of the form

$$x_t = \mu_t + y_t$$

where μ_t is a function and y_t is stationary. We've already discussed when μ_t is of the form

$$\beta_1 + \beta_2 t + ... + \beta_{p+1} t^p$$

or

$$\beta_1 cos(2\pi\omega t) + \beta_2 sin(2\pi\omega t).$$

In the context of smoothing, we will denote μ_t by f_t.

Averaging

One way to approximate f_t is to take a moving average of the time series. Averaging, in general, ________________. It can also reduce "seasonal" fluctuations. Averaging can help in viewing ________________, because the seasonal variations will be damped. In general we may write a moving average as

$$m_t = \sum_{j=-k}^{k} a_j x_{t-j}, \tag{1}$$

where $a_j \geq 0$ and $\sum_{j=-k}^{k} a_j = 1$.

Averaging

Equation (1) can be used as an estimate of f_t. The smoothed value for a particular time is calculated as a ________________ of observations for surrounding times.

Equation (1) is sometimes also called a centered moving average, since we average with k times before and k times after. For example, a centered moving average of length 5 with equal weights would be the average of values at times $t-2, t-1, t, t+1, t+2$.

Averaging

Smoothing using (1) has the advantage of being adaptable to slow changes in f_t across time. The disadvantage is that there may still be a substantial amount of variability in our estimate f_t, and we may not know ________ what the window size, k, should be. With some data, there will be a natural window size; with others there may not be.

Window Size

Question: What is an appropriate window size, k, to smooth away seasonality in quarterly data, in order to identify yearly trends? What about monthly data? Weekly data?

Variance Reduction with Averaging

It was mentioned earlier that averaging reduces variation, in general. Let's look at an example. Assume that the original series x_t is stationary, so that $var(x_t)$ is constant, denoted by σ^2. Let's create another time series

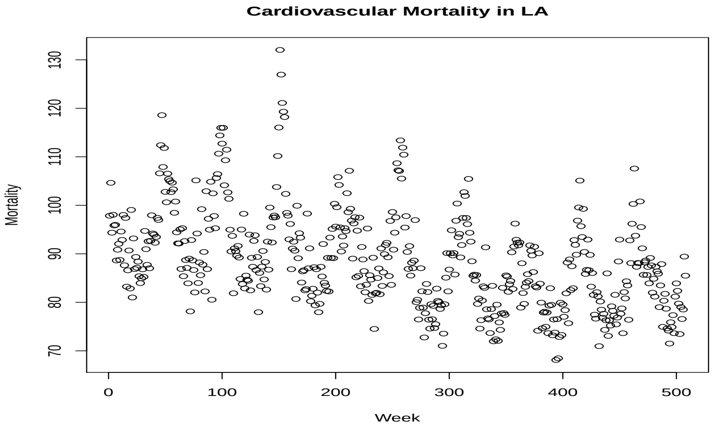
$$y_t = \frac{1}{3}x_{t+1} + \frac{1}{3}x_t + \frac{1}{3}x_{t-1}.$$

Question: Derive the variance of y_t.

Variance Reduction with Averaging

Example: Cardiovascular Mortality in Los Angeles

In this example, we look at the weekly cardiovascular mortality in Los Angeles county. We have a plot of the data below.

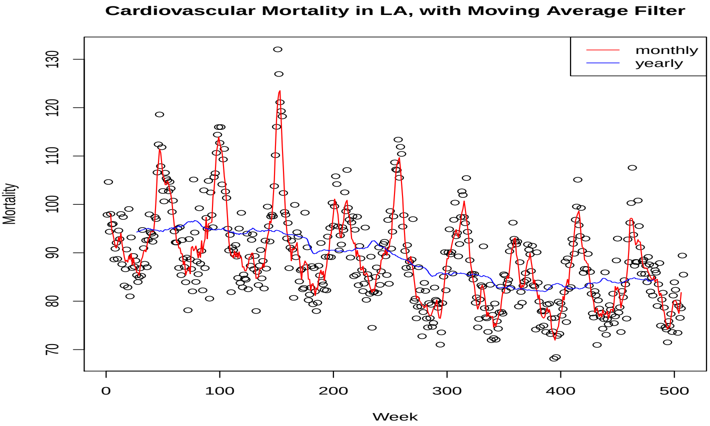


Comments on monthly trend? Yearly trend?

Example: Cardiovascular Mortality in Los Angeles

To make things a little easier, we use a centered moving average of length 5 ($k = 2$) to obtain the monthly average, to bring out the monthly trend. We also use a centered moving average of length 53 ($k = 26$) to obtain the yearly average, to bring out the yearly trend.

Example: Cardiovascular Mortality in Los Angeles



Comments on seasonal and/or yearly trend?

Example: Cardiovascular Mortality in Los Angeles

```
mort<-scan("cmort.dat")
plot(mort))
ma5<-filter(mort,sides=2,rep(1,5)/5)
ma53<-filter(mort,sides=2,rep(1,53)/53)
lines(ma5, col="red")
lines(ma53, col="blue")
```

1. Averaging
2. Kernel Smoothing
3. Nearest Neighbors Regression

Kernel Smoothing

The idea with kernel smoothing is similar to the moving average; however, the contribution to the estimate of the smooth function at a point t from local points declines as a function of distance from the current point. The smooth function is estimated by

$$\hat{f}_t = \sum_{i=1}^{n} w_t(i) x_j \tag{2}$$

where

$$w_t(i) = \frac{K\left(\frac{t-i}{b}\right)}{\sum_{i=1}^{n} K\left(\frac{t-i}{b}\right)}. \tag{3}$$

In (3), $K(.)$ is the ________________, and b is the ____________.

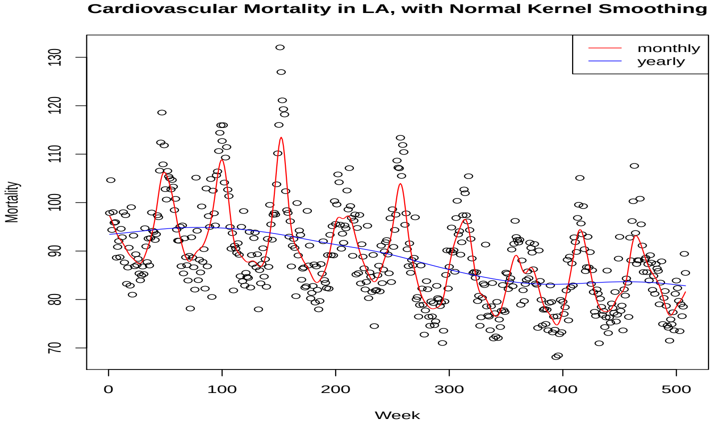
Kernel Smoothing

A common choice for $K(.)$ is a standard normal density; in this case, $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$. Other choices include the uniform density.

An issue with kernel smoothing is the choice of b. This will help determine how much of f_t is determined by neighboring points. The _____ the bandwidth, the smoother the result.

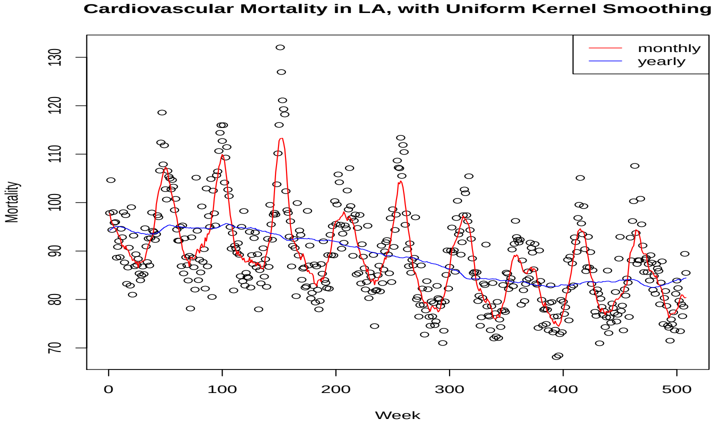
Example: Cardiovascular Mortality in Los Angeles

We apply normal kernel smoothing to the cardiovascular data.



Example: Cardiovascular Mortality in Los Angeles

We apply uniform kernel smoothing to the cardiovascular data.



Example: Cardiovascular Mortality in Los Angeles

```
t<-1:length(mort)
plot(t,mort)
lines(ksmooth(t,mort, "normal", bandwidth=10))
lines(ksmooth(t,mort, "normal", bandwidth=104))
```

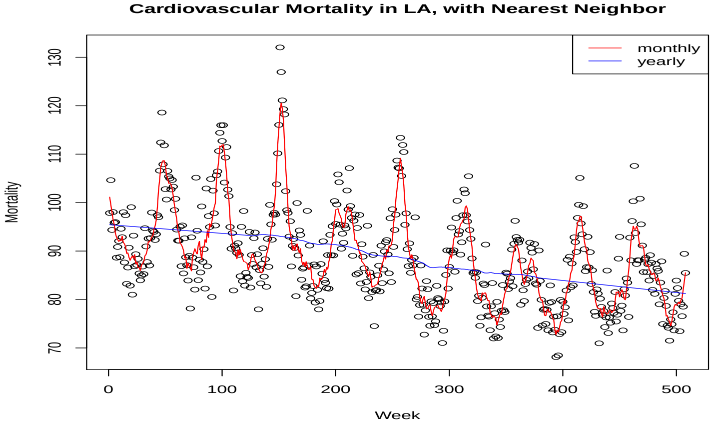
1. Averaging
2. Kernel Smoothing
3. Nearest Neighbors Regression

Nearest Neighbors Regression

Another technique is nearest neighbors regression. This is based on k-nearest linear regression, where $\{x_{t-k/2}, \cdots, x_t, \cdots, x_{t+k/2}\}$ to predict x_t via linear regression.

Example: Cardiovascular Mortality in Los Angeles

We apply nearest neighbors regression to the cardiovascular data.



Example: Cardiovascular Mortality in Los Angeles

```
plot(t,mort)
lines(supsmu(t,mort,span=0.01))
lines(supsmu(t,mort,span=0.5))
```

Under-smoothing Vs Over-smoothing

Question: What are the consequences of under-smoothing and over-smoothing?

Summary

It can be difficult to detect trends and patterns in time series plots. Smoothing can help us dampen irregularities so we get a clearer idea on the time series. Note that all of the techniques discussed can be used, however, the most important aspect in using them is knowing the right parameters to use (e.g. bandwith, window etc.) These techniques aid us in develop an appropriate model for our data and should be used as a guide.