MATH 545, TIME SERIES
SELECTED HOMEWORK SOLUTIONS

STEVEN HEILMAN

Contents

1. Homework 1

2. Homework 2

3. Homework 3

Exercise 3.5 (**Sunspot Data, Version 2**). This exercise deals with sunspot data from the following files (the same data appears in different formats)

txt file csv (excel) file

These files are taken from http://www.sidc.be/silso/datafiles#total

To work with this data, e.g. in Matlab you can use the command

```
x=importdata('SN_d_tot_V2.0.txt')
```

to import the .txt file.

The format of the data is as follows.

- Columns 1-3: Gregorian calendar date (Year, Month, then Day)
- Column 4: Date in fraction of year
- Column 5: Daily total number of sunspots observed on the sun. A value of -1 indicates that no number is available for that day (missing value).
- Column 6: Daily standard deviation of the input sunspot numbers from individual stations.
- Column 7: Number of observations used to compute the daily value.
- Column 8: Definitive/provisional indicator. A blank indicates that the value is definitive. A '*' symbol indicates that the value is still provisional and is subject to a possible revision (Usually the last 3 to 6 months)

Date: May 1, 2020

In a previous Exercise, we modelled the number of sunspots U_t at time t, where t is measured in years, by

$$U_t = m_t + a\cos(2\pi t/11) + b\sin(2\pi t/11) + Y_t, \qquad \forall\, t \in \mathbb{R},$$

where $a, b, \theta, \omega \in \mathbb{R}$ are unknown (deterministic) parameters, m_t is an unknown deterministic function of t that is assumed to be a "slowly varying" function of t, and $\{Y_t\}_{t\in\mathbb{R}}$ are i.i.d. mean zero random variables. The quantity m_t is called the **trend** and the quantity $s_t := a\cos(2\pi t/11) + b\sin(2\pi t/11)$ is called the **seasonal component** of the time series $\{U_t\}_{t\in\mathbb{R}}$.

This model was perhaps too simplistic, since it did not seem to fit the data well in some respects. This time, let's not make an a priori assumptions about known periodicities in the data. Consequently, we will just examine the Fourier coefficients of the data U_t directly. If we want to make plots of Fourier coefficients, it is easier to take absolute values. If the units of t were in integers, then we would define $\widehat{U}(s) = \sum_{t\in\mathbb{Z}} U_t e^{2\pi i s n}$ for any $s \in \mathbb{R}/\mathbb{Z}$, as the n^{th} frequency component of the time series. Since the units of t are in integers divided by 365 (or by 365.25), we instead define

$$\widehat{U}(r) := \sum_{t\in\mathbb{Z}/365} U_t e^{2\pi i t r}, \qquad \forall\, r \in \mathbb{R}/365\mathbb{Z}.$$

• Plot $|\widehat{U}(r)|$ versus r, where $r \in [0,1]$ and also when $r \in [0,365]$. Do you observe any large absolute values of $\widehat{U}(r)$ for any values of r near $1/11$?

You should observe some large values of $\widehat{U}(r)$ when r takes the values: .0842, .0921, and .0995, corresponding to frequencies of $11.87, 10.858$, and 10.05, respectively. This large signal should correspond to $r \in [.08, .105]$ (and to $r \in [-.105, -.08]$).

• Plot the inverse Fourier transform of this part of the frequency spectrum $\widehat{U}(r)$. That is, plot the following function of $t \in \mathbb{Z}/365$

$$S_t := \frac{1}{365}\int_{.08}^{.105} \widehat{U}(r)e^{-2\pi i r t}dr + \frac{1}{365}\int_{-.105}^{-.08} \widehat{U}(r)e^{-2\pi i r t}dr.$$

How does S_t compare to U_t when you put them in the same plot? (Instead of plotting U_t itself for this comparison, consider plotting a moving average of U_t.)

When you plotted S_t versus U_t, S_t should follow the oscillations of U_t fairly closely.

When you plotted $\widehat{U}(r)$ versus r, you should have also noticed large values for r near 0. These low frequencies correspond to the long term "trend" in the data. The low frequency signal should correspond roughly to $r \in [0, .016]$

• Plot the inverse Fourier transform of this part of the frequency spectrum $\widehat{U}(r)$. That is, plot the following function of $t \in \mathbb{Z}/365$

$$M_t := \frac{1}{365}\int_{-.016}^{.016} \widehat{U}(r)e^{-2\pi i r t}dr.$$

Plot U_t, M_t and S_t in the same plot. Then plot $U_t - S_t - M_t$. Does $U_t - S_t - M_t$ "resemble" a stationary process? Is this procedure better or worse than what we did on the previous homework?

• If $\{U_t - S_t - M_t\}_{t\in\mathbb{Z}/365}$ were a sequence of i.i.d. random variables with mean zero and variance one, what would its Fourier transform look like? That is, if $\{Z_t\}_{t\in\mathbb{Z}/365}$ were a sequence of i.i.d. random variables with mean zero and variance one, and if

$$\widehat{Z}(r) := \sum_{t \in \mathbb{Z}/365} Z_t e^{2\pi i t r}, \qquad \forall\, r \in \mathbb{R}/365\mathbb{Z},$$

then what would this function look like? Does it have mean zero (when r is fixed)? Can you compute the variance of $\widehat{Z}(r)$ (when r is fixed)? Are the quantities $\widehat{Z}(r)$ and $\widehat{Z}(s)$ independent when $s \neq r$, $s, r \in \mathbb{R}/365\mathbb{Z}$? If they are not independent, could you compute their covariance?

Here is the output and code for a Matlab program answering the above question.

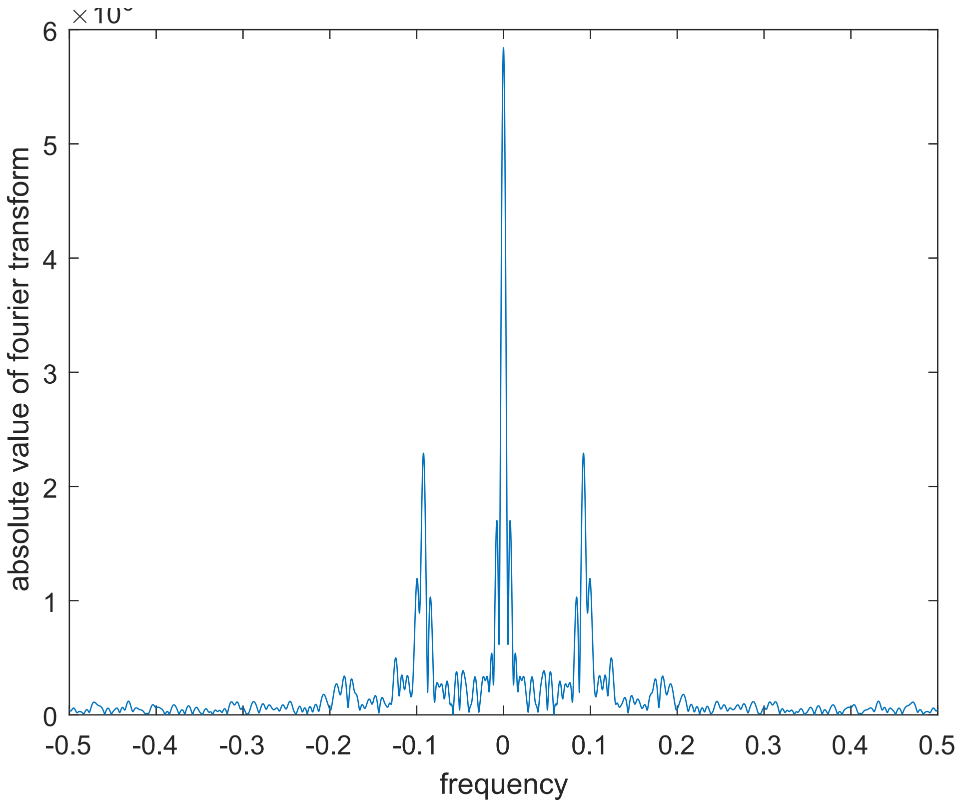

Figure 1. Plot of Absolute Value of Fourier Transform.

```
%%%%%%%

clear all; close all;
x=importdata('SN_d_tot_V2.0.txt');  %import the sunspot data
y=x(:,[4,5]);                       % only 4th and 5th columns are relevant

numb=10000;                           % number of frequencies
r=linspace(-365/2,365/2,numb); r=r/365;  %% frequencies
x=zeros(1,numb);                      % fourier transform
for i=1:numb
    x(i)=sum((y(:,2)).*exp(2*pi*(sqrt(-1))*(y(:,1)).*r(i))); %% i^th FC
```

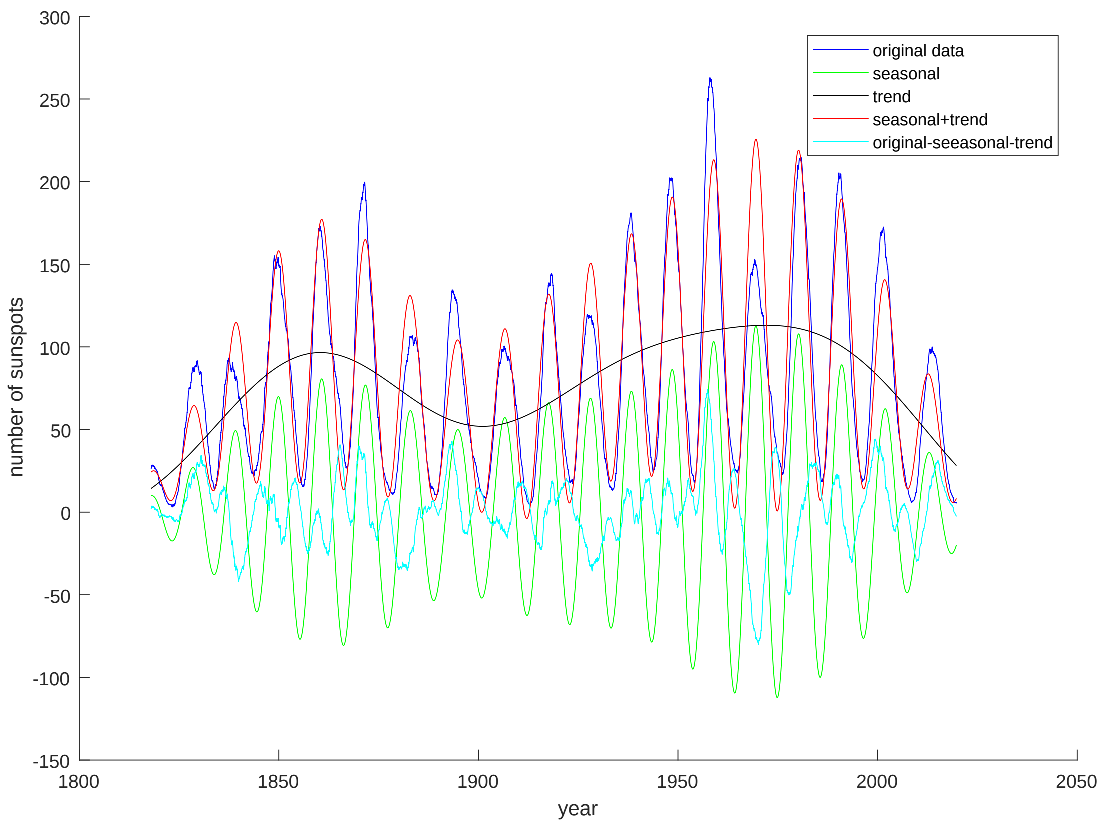

FIGURE 2. Sunspot Data Summary

```
end
 plot(r,abs(x));    %plot the absolute values of the fourier coefficients
 xlabel('frequency');  ylabel('absolute value of fourier transform');
```

```
X=x;

%%now invert the Fourier transform
Y=zeros(length(y),1);  % Y is the seasonal component
startind=find(abs(r-.08)==min(abs(r-.08))); % find indices i
lastind=find(abs(r-.105)==min(abs(r-.105)));
for i=startind:lastind
    Y=Y+(1/numb)*(1/365)*(exp(-2*pi*sqrt(-1)*(y(:,1))*r(i)))*X(i);
end
startind=find(abs(r+.105)==min(abs(r+.105)));
lastind=find(abs(r+.08)==min(abs(r+.08)));
for i=startind:lastind
    Y=Y+(1/numb)*(1/365)*(exp(-2*pi*sqrt(-1)*(y(:,1))*r(i)))*X(i);
end


Z=zeros(length(y),1); %Z is the trend
newfirst=find(abs(r+.016)==min(abs(r+.016)));  % find indices
newlast=find(abs(r-.016)==min(abs(r-.016)));
for i=newfirst:newlast
    Z=Z+(1/numb)*(1/365)*(exp(-2*pi*sqrt(-1)*(y(:,1))*r(i)))*X(i);
end

Y=real(Y);
Z=real(Z);

figure; hold on;
plot(y(:,1),movmean(y(:,2),1000),'b');
plot(y(:,1),Y,'g'); plot(y(:,1),Z,'k'); plot(y(:,1),Y+Z,'r');
plot(y(:,1),movmean(y(:,2),1000)-Y-Z,'c')
legend('original data','seasonal','trend','seasonal+trend','original-seeasonal-trend');
xlabel('year');  ylabel('number of sunspots');
```

4. Homework 4

Exercise 4.4. Let $\{X_n\}_{n\in\mathbb{Z}}$ be a weakly stationary, real-valued time series. For any $n \in \mathbb{Z}$, define the **difference operator** by

$$DX_n := X_n - X_{n-1}, \qquad \forall\, n \in \mathbb{Z}.$$

Define $D^1 := D$. And for any $k \geq 2$, define inductively

$$D^k X_n := D^{k-1}(DX_n), \qquad \forall\, n \in \mathbb{Z}.$$

- For any $k \geq 1$, is $\{D^k X_n\}_{n\in\mathbb{Z}}$ weakly stationary? If so, prove it.
- Let $\{a_n\}_{n\in\mathbb{Z}}$ be a sequence of real numbers with $\sum_{n\in\mathbb{Z}} |a_n|^2 < \infty$. Let $f(x) := \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}\ \forall\, x \in \mathbb{R}/\mathbb{Z}$ be its Fourier transform. Recall that $a_n = \int_0^1 e^{-2\pi i n x} f(x)dx$

$\forall\ n \in \mathbb{Z}$. Write a similar integral expression for $D^k a_n$ for all $k \geq 1$, $n \in \mathbb{Z}$. That is, find $g_k \colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ such that

$$D^k a_n = \int_0^1 g_k(x) e^{-2\pi i n x} f(x) dx, \qquad \forall\, k \geq 1,\ n \in \mathbb{Z}.$$

- What happens to g_k when k is large? For the sunspot data, examine $D^k X_n$ for all $1 \leq k \leq 5$, and examine the Fourier transforms. Are your observations consistent with the behavior of g_k when k is large?.

Solution. For any $k \geq 1$, $\{D^k X_n\}_{n\in\mathbb{Z}}$ is weakly stationary. It suffices to prove this in the case $k = 1$, when $\{D^k X_n\}_{n\in\mathbb{Z}} = \{X_n - X_{n-1}\}_{n\in\mathbb{Z}}$. Since $\{X_n\}_{n\in\mathbb{Z}}$ is weakly stationary, $\mathbf{E}X_n^2 < \infty$, $\mathbf{E}X_n = \mathbf{E}X_0$ and $\mathbf{E}X_{m+n}X_m = \mathbf{E}X_m X_0$ for all $n, m \in \mathbb{Z}$. So $\mathbf{E}(X_n - X_{n-1})^2 < \infty$, $\mathbf{E}(X_n - X_{n-1}) = \mathbf{E}(X_0) - \mathbf{E}X_0 = 0$ for all $n \in \mathbb{Z}$. And for any $m, n \in \mathbb{Z}$,

$$\begin{aligned}\mathbf{E}(X_{m+n} - X_{m+n-1})(X_n - X_{n-1}) &= \mathbf{E}X_m X_0 + \mathbf{E}X_m X_0 - \mathbf{E}X_{m-1}X_0 - \mathbf{E}X_{m+1}X_0 \\ &= \mathbf{E}(X_m - X_{m-1})(X_0 - X_{-1}).\end{aligned}$$

That is, $\{X_n - X_{n-1}\}_{n\in\mathbb{Z}}$ is weakly stationary.

Since $a_n = \int_0^1 e^{-2\pi i n x} f(x) dx\ \forall\ n \in \mathbb{Z}$, we have, for any $n \in \mathbb{Z}$

$$D a_n = a_n - a_{n-1} = \int_0^1 [e^{-2\pi i n x} - e^{-2\pi i (n-1) x}] f(x) dx = \int_0^1 e^{-2\pi i n x}[1 - e^{2\pi i x}] f(x) dx.$$

By inducting on k, we then have, for any $k \geq 1$ and $n \in \mathbb{Z}$,

$$D^k a_n = \int_0^1 e^{-2\pi i n x}[1 - e^{2\pi i x}]^k f(x) dx.$$

Note that $1 - e^{2\pi i x} = -2\pi i x + O(x^2)$, so $[1 - e^{2\pi i x}]^k = (-2\pi i x)^k + O(x^{k+1})$. That is, when k is large

$$D^k a_n = \int_0^1 e^{-2\pi i n x}[(-2\pi i x)^k + O(x^{k+1})] f(x) dx.$$

That is, $D^k a_n$ "dampens" frequencies near zero while accentuating higher frequencies.

□

5. Homework 5

6. Homework 6

Exercise 6.3 (**Sunspot Data, Version 3**)**.** This exercise deals with sunspot data from the following files (the same data appears in different formats)

txt file csv (excel) file

These files are taken from http://www.sidc.be/silso/datafiles#total

To work with this data, e.g. in Matlab you can use the command

```
x=importdata('SN_d_tot_V2.0.txt')
```

to import the .txt file.

The format of the data is as follows.

- Columns 1-3: Gregorian calendar date (Year, Month, then Day)
- Column 4: Date in fraction of year

- Column 5: Daily total number of sunspots observed on the sun. A value of -1 indicates that no number is available for that day (missing value).
- Column 6: Daily standard deviation of the input sunspot numbers from individual stations.
- Column 7: Number of observations used to compute the daily value.
- Column 8: Definitive/provisional indicator. A blank indicates that the value is definitive. A '*' symbol indicates that the value is still provisional and is subject to a possible revision (Usually the last 3 to 6 months)

In two previous exercises, we examined the number of sunspots U_t at time t, where t is measured in years. In the previous exercise, we took the Fourier transform of U, and defined

$$\widehat{U}(r) := \sum_{t\in\mathbb{Z}/365} U_t e^{2\pi i t r}, \qquad \forall\, r \in \mathbb{R}/365\mathbb{Z}.$$

We found that the seasonal component of U closely matched the following function of $t \in \mathbb{Z}/365$

$$S_t := \frac{1}{365}\int_{.08}^{.105} \widehat{U}(r)e^{-2\pi i r t}dr + \frac{1}{365}\int_{-.105}^{-.08} \widehat{U}(r)e^{-2\pi i r t}dr.$$

We also found that the trend component of U closely matched the following function of $t \in \mathbb{Z}/365$

$$M_t := \frac{1}{365}\int_{-.016}^{.016} \widehat{U}(r)e^{-2\pi i r t}dr.$$

In this exercise, we will try to model U_t, and $U_t - S_t - M_t$, as ARMA processes.

• First, try to fit an autoregressive model AR(p) to U_t. That is, use the Yule-Walker estimators for the autoregressive parameters. Examine the values of the partial autocorrelation function, to try to determine a good value of p for the model.

• Now, in order to see whether or not this process really is autoregressive, model it as a general ARMA process, and use both (i) the innovation estimators (from Definition 9.12 in the notes, which then imply estimates for the a and b parameters in the ARMA process) and (ii) the Gaussian maximum likelihood estimators. Do these estimators suggest that the process is autoregressive?

• Give another estimate for p, q using AICC, then create a MLARMA spectral density estimate.

• Do all steps above for $U_t - S_t - M_t$, instead of U_t.

Solution.

Here is a solution to the first part of the problem. It seems like $p = 2$ or $p = 3$ is a reasonable choice.

```
%%%%%%%
clear all; close all;
x=importdata('SN_d_tot_V2.0.txt');  %import the sunspot data
y=x(:,[4,5]);                       % only 4th and 5th columns are relevant
clear x; x=y(:,2);   %  x= data to process

%first, compute gamma
n=length(x);
```

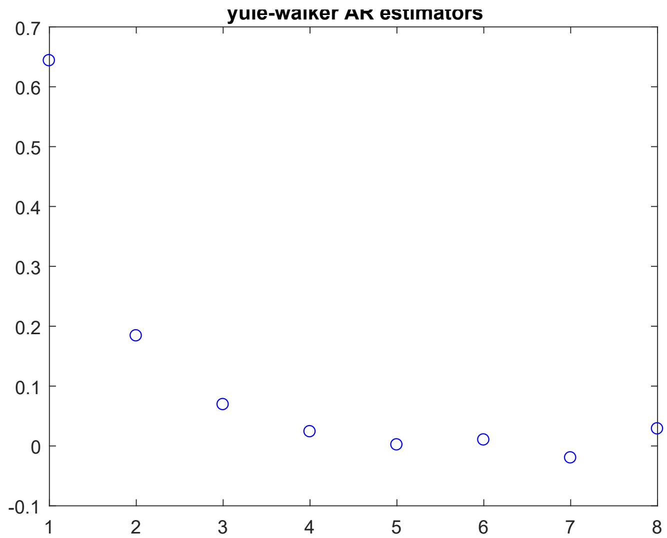

FIGURE 3. Sunspot Yule-Walker Estimators

```
sn=sum(x)/n;
gamma=zeros(n,1);
for m=1:(n-1)
    for j=1:(n-abs(m))           %instead, shift all indices by +n
        gamma(m)=  gamma(m)+(x(j)-sn)*(x(j+m)-sn);
    end
end
gamma=[sum((x-sn).^2);gamma]; % add zero entry, shift all indices by 1
gamma=gamma/n;  % divide by n


%now do the yule-walker stuff
p=8;
covmat=zeros(p,p);
covvec=zeros(p,1);
for i=1:p
    for j=1:p
        covmat(i,j)=gamma(abs(i-j)+1);  % gamma indices are shifted by 1
    end
    covvec(i)=gamma(i+1);
end
A=(inv(covmat))*covvec;
figure;
plot(A,'bo')
title('yule-walker AR estimators');
```

$\square$

USC Department of Mathematics, Los Angeles, CA
E-mail address: stevenmheilman@gmail.com