Please provide complete and well-written solutions to the following exercises.

Due May 1, at the beginning of class.

Homework 6

Exercise 1. Suppose we have a real-valued AR(p) process of the form

$$X_n = \sum_{j=1}^{p} a_j X_{n-j} + Z_n, \qquad \forall\, n \in \mathbf{Z},$$

where $\{Z_n\}_{n\in\mathbf{Z}}$ are $WN(0,\sigma^2)$ and $0 < \sigma < \infty$. Show that the Gaussian likelihood $\ell(a_1,\ldots,a_p,\sigma)$ takes the form

$$\ell(a_1,\ldots,a_p,\sigma) = (2\pi\sigma^2)^{-n/2}(\det(G_p))^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}\Big[X^T G_p^{-1}X + \sum_{t=p+1}^{n}(X_t - \sum_{j=1}^{p} a_j X_{t-j})^2\Big]\Big),$$

where

$$X := (X_1,\ldots,X_p)^T, \qquad G_p = \frac{1}{\sigma^2}\mathbf{E}XX^T.$$

Exercise 2 (Financial Data Analysis)**.** In this exercise, we will attempt to model financial data by ARMA processes. You can create a spreadsheet of financial data if you open up a spreadsheet in google, and input the following command into one of the cells in the spreadsheet (and then press enter):

```
=GOOGLEFINANCE("GOOG", "price", "1/1/2010", "12/31/2019", "DAILY")
```

The first argument in the command is the stock price ticker symbol (which corresponds to google stock). The closing price of the stock is then listed every day, starting from January 1st, 2010, and ending on December 31st, 2019. Once you have the spreadsheet data, you should be able to import it into Matlab or your favorite mathematical software.

Since stock prices generally increase in the long term, they are probably not weakly stationary (as their expected value is not constant). Moreover, stock prices change in a way that is proportional to their value. So, if X_t is a stock price on day t, we anticipate that the logarithmically differenced value

$$Y_n := \log(X_{n+1}/X_n) = \log X_{n+1} - \log X_n \qquad (*)$$

could be a weakly stationary ARMA process. Let's try to fit data like this to such a model.

(Logarithmic differencing to get a weakly stationary process is also reasonable since a typical model of a stock price in discrete time is $W_n = e^{Z_n}$, where $Z_1, Z_2, \ldots$ is a random walk, so that $\log(W_{n+1}/W_n)$ is a sequence of independent random variables.)

• First, take the Google stock closing price from all of 2009 through all of 2019, and examine Y_n versus n (where n is the day of the year, neglecting weekend days; so e.g. $n = 6$ is the sixth non-weekend day of 2009). 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. (Recall from the notes that the order m of the innovation estimators should satisfy $m = o(n^{1/3})$ in order to be consistent, and we have $n \approx 2 \cdot 10^3$, so you should not take m larger than about 14 in order to have a consistent estimator. Moreover, with $m \leq 14$, you should also restrict p, q to be at most 14.)

• A classic barometer of US economy is the ten year US treasury note yield. You can access this data with the command

```
=GOOGLEFINANCE("TNX", "price", DATE(2010,1,1), DATE(2019,12,31), "DAILY")
```

A discrete version of Vasicek interest rate model is

$$X_{n+1} - X_n = a(b - X_n) + Z_n, \qquad \forall\, n \in \mathbf{Z}.$$

where a, b are unknown parameters. So, let's try to see if the interest rate is autoregressive or not. Try to fit an autoregressive model AR(p) to the interest rate data. 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. 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. Does the order selection (AICC) estimator of p, q agree with your findings? (Examine the data itself, do not do any logarithmic differencing.)

Exercise 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

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

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

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

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

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

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.