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

Due April 10, at the beginning of class.

Homework 5

Exercise 1. Let $\{X_n\}_{n\in\mathbf{Z}}$ be a real-valued weakly stationary time series. Let $\{Z_n\}_{n\in\mathbf{Z}}$ be $WN(0,1)$. Assume that

$$X_n = (1/2)X_{n-3} + Z_n, \qquad \forall\, n \in \mathbf{Z}.$$

Find explicit expressions for the autocovariance function and partial autocorrelation function of $\{X_n\}_{n\in\mathbf{Z}}$.

Also, find an explicit expression for the spectral density (i.e. the Fourier transform of $\gamma\colon \mathbf{Z} \to \mathbf{R}$).

Exercise 2. Let p, q be positive integers. Let $a_1, \ldots, a_p, b_1, \ldots, b_q$ be real numbers. For any $z \in \mathbf{C}$, define $\phi(z) := 1 - \sum_{j=1}^{p} a_j z^j, \qquad \theta(z) := 1 + \sum_{j=1}^{q} b_j z^j$. Suppose ϕ, θ have no common zeros. Suppose also that $\phi(z) = 0$ for some $z \in \mathbf{C}$ with $|z| = 1$. Show that there does not exist a unique weakly stationary stochastic process $\{X_n\}_{n\in\mathbf{Z}}$ satisfying

$$\phi(S)X_n = \theta(S)Z_n, \qquad \forall\, n \in \mathbf{Z}.$$

(Argue by contradiction. Assume a stationary solution exists. Then, equate the spectral densities of $\{\phi(S)X_n\}_{n\in\mathbf{Z}}$ and $\{\theta(S)Z_n\}_{n\in\mathbf{Z}}$ to derive a contradiction.)

Exercise 3. Let $X_1, \ldots, X_n$ be n samples from a weakly stationary time series. For any $n \geq 1$, define the n^{th} sample mean

$$\overline{X}_n := \frac{1}{n}(X_1 + \cdots + X_n).$$

For any $n \geq 1$ define the n^{th} sample autocovariance function $\Gamma_n\colon \mathbf{Z} \to \mathbf{R}$ by

$$\Gamma_n(m) := 1_{\{|m|<n\}} \frac{1}{n} \sum_{j=1}^{n-|m|} (X_j - \overline{X}_n)(X_{j+|m|} - \overline{X}_n), \qquad \forall\, m \in \mathbf{Z}.$$

Note: we have added an absolute value on the right term that does not appear in the book, or in our previous definition of $\Gamma_n(m)$.

This exercise demonstrates that the sample (auto)covariance matrix $\{\Gamma_n(j-k)\}_{1\leq j,k\leq n}$ is invertible for any $n \geq 1$, if $\Gamma_n(0) > 0$ for all $n \geq 1$.

- Show that for any $n \geq 1$, the function $m \mapsto \Gamma_n(m)$ is an even, positive semidefinite function. So, from Proposition 4.32 in the notes, there exists a (strongly) stationary time series $\{Y_m\}_{m\in\mathbf{Z}}$ whose autocovariance function $\gamma_Y\colon \mathbf{Z} \to \mathbf{R}$ satisfies $\gamma_Y(m) = \Gamma_n(m)$ for all $m \in \mathbf{Z}$.

- Using Lemma 7.23 in the notes, conclude that, if $\Gamma_n(0) > 0$, then the matrix $\{\Gamma_n(j-k)\}_{1\leq j,k\leq n}$ is invertible.

Exercise 4 (Exponential and Logistic Growth)**.** Recall the **logistic growth** model. Suppose $y(t)$ is the amount of bacteria in a petri dish at time t and $k > 0$ is a constant. Let c be the maximum possible population of the bacteria. Let $y_0 > 0$ be the initial population of bacteria. We model the growth of the bacteria by the formula

$$y'(t) = ky(t)(c - y(t)), \qquad y(0) = y_0, \qquad \forall\, t \in \mathbf{R}.$$

So, when y is small, $y'(t)$ is proportional to y. However, when y becomes close to C, y' becomes very small. That is, the rate of growth of bacteria is constrained by the environment. Recall that the following logistic function satisfies the above differential equation.

$$y(t) = \frac{C}{1 + (C^{-1}y_0^{-1} - 1)e^{-kt}}, \qquad \forall\, t \in \mathbf{R}.$$

A discrete, randomized version of the logistic growth model would be

$$X_n - X_{n-1} = kX_{n-1}(c - X_{n-1}), \qquad \forall\, n \in \mathbf{Z}$$

This model has a nonlinearity due to the X_{n-1}^2 term on the right side. However, when $y(t)$ is much smaller than c, this model resembles that of exponential growth. So, a reasonable approximation to the logistic growth model would be an AR(1) model of the form

$$X_n - X_{n-1} = kX_{n-1} + Z_n, \qquad \forall\, n \in \mathbf{Z}.$$

That is,

$$X_n = aX_{n-1} + Z_n, \qquad \forall\, n \in \mathbf{Z}.$$

for some $a > 1$. Since $a > 1$, $\phi(z) = 1 - az$ has a zero in the unit disc ($z = 1/a$), so $\{X_n\}_{n\in\mathbf{Z}}$ would not be causal. To get an equivalent causal process, we use Theorem 7.19 in the notes to get $\widetilde{\phi}(z) := (1 - z/a)$, i.e. we may as well start the original model with $a < 1$. In any case, we will try to estimate the parameters for a few data sets for the model

$$X_n = aX_{n-1} + Z_n, \qquad \forall\, n \in \mathbf{Z},$$

where we presume that $a < 1$.

- Let's use our data from the cumulative number of confirmed coronavirus cases in the US, starting February 26, through March 24. The daily cumulative number of cases is e.g.

  15 15 19 24 42 57 85 111 175 252 353 497 645 936 1205 1598

  2163 2825 3501 4373 5662 8074 12018 17438 23710 32341 42752 51825

  The most updated data appears at a wikipedia link here. Note that the number of confirmed cases is much different than the actual number of cases, which is anticipated to be an order of magnitude higher.
- Use the Durbin-Levinson algorithm to compute the first few values of the partial autocorrelation function $\alpha\colon \{1, 2, 3, \ldots\} \to \mathbf{R}$ (where the sample autocovariance function is substituted for the actual autocovariance function). You should observe that $\alpha(2)$ is small, indicating that the AR(1) model could be accurate. In this exercise, it would

be more convenient to force the sample autocovariance to be symmetric, i.e. for any $n \geq 1$ define the n^{th} sample autocovariance function $\Gamma_n \colon \mathbf{Z} \to \mathbf{R}$ by

$$\Gamma_n(m) := 1_{\{|m|<n\}} \frac{1}{n} \sum_{j=1}^{n-|m|} (X_j - S_n)(X_{j+|m|} - S_n), \qquad \forall\, m \in \mathbf{Z}.$$

- Use the Yule-Walker estimators for the autoregressive parameters. (Assume a priori that the process you are sampling from is AR(p) for $p = 1, 2$ or 3.) You should find that a_{11} is around .767 and your estimate for the first parameter should be about .83, and the remaining parameters are small. Going back through our derivation, this indicates that $X_n \approx (1/.83)X_{n-1}$ for all $n \geq 2$. Does this look to be true? (Of course we could have just plotted $\log X_n$ versus n at the outset.)
- An ARMA model pre-supposes that the data is weakly stationary, e.g. its mean is time invariant. This does not at all seem to be the case with this data, despite the fact that the estimators worked somewhat correctly. So, re-do what you did above, but now replace $\{X_n\}_{n\in\mathbf{Z}}$ with the logarithmically differenced sequence

  $$\{\log X_n - \log X_{n-1})\}_{n\in\mathbf{Z}} = \{\log(X_n/X_{n-1})\}_{n\in\mathbf{Z}}.$$

  The partial autocorrelation and Yule-Walker estimators should look like random numbers, so perhaps it is not unreasonable to just assume that there are random variables $\{Z_n\}_{n\in\mathbf{Z}}$ that are $WN(\mu, \sigma^2)$ such that

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

  Using e.g. mean and covariance estimators, estimate μ and σ^2. Using this model, estimate the number of cases for tomorrow. Estimate the mean squared error of your prediction.
- Do the same thing for data from Italy (link here) and Malaysia (link here). Using your estimates for the mean of Z_n what are your estimates for the doubling time (i.e. the number of days it takes for the number of cases to double) in each country?