STAT 720 sp 2019 midterm

assigned: Wednesday, March 20th, 2019

due on Friday, March 22nd, 2019 at 5:00 pm in my mailbox or to me in my office.

This is a take-home exam. You may use all the course notes and any books. Do not work together with any classmates or with any other person.

1. Consider the MA(2) time series $\{X_t, t \in \mathbb{Z}\}$ defined by

$$X_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, \qquad \text{for all } t \in \mathbb{Z}, \tag{1}$$

where $\{Z_t, t \in \mathbb{Z}\} \sim \mathrm{IID}(0, \sigma^2)$.

(a) Find the autocovariance function $\gamma(\cdot)$ of the time series $\{X_t, t \in \mathbb{Z}\}$.

For any $h \in \mathbb{Z}$ We have

$$\begin{aligned}
\mathrm{Cov}(X_t, X_{t+h}) &= \mathrm{Cov}(\mu + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, \mu + Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_2 Z_{t+h-2}) \\
&= \mathbb{E}(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2})(Z_{t+h} + \theta_1 Z_{t+h-1} + \theta_2 Z_{t+h-2}) \\
&= \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2), & h = 0 \\ \sigma^2(\theta_1 + \theta_1\theta_2), & h = \pm 1 \\ \sigma^2(\theta_2 + \theta_1\theta_2), & h = \pm 2 \\ 0, & |h| > 2. \end{cases} \\
&=: \gamma(h).
\end{aligned}$$

(b) Find $2\pi f(0)$, where $f(\cdot)$ is the spectral density of the time series $\{X_t, t \in \mathbb{Z}\}$.

We have

$$f(\lambda) = \frac{\sigma^2}{2\pi}|1 + \theta_1 \exp(-\iota\lambda) + \theta_2 \exp(-\iota 2\lambda)|^2,$$

so that

$$2\pi f(0) = \sigma^2(1 + \theta_1 + \theta_2)^2.$$

(c) Verify that $2\pi f(0) = \sum_{h=-\infty}^{\infty} \gamma(h)$ for the time series in (1).

We have

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2(1 + \theta_1^2 + \theta_2^2) + 2\sigma^2(\theta_1 + \theta_1\theta_2) + 2\sigma^2(\theta_2 + \theta_1\theta_2) = \sigma^2(1 + \theta_1 + \theta_2)^2 = 2\pi f(0).$$

(d) Give the form of an asymptotic 95% confidence interval for μ based on $X_1, \ldots, X_n$, assuming θ_1, θ_2, and σ^2 are known.

Since $\sqrt{n}(\bar{X} - \mu) \to \text{Normal}(0, 2\pi f(0))$ in distribution as $n \to \infty$, the interval

$$\bar{X} \pm 1.96\sigma(1 + \theta_1 + \theta_2)/\sqrt{n}$$

is an asymptotic 95% confidence interval for μ.

(e) Give finite values of the parameters θ_1 and θ_2, if they exist, for which the time series $\{X_t, t \in \mathbb{Z}\}$ is

i. invertible.

The time series is invertible for $\theta_1 = 1/2$ and $\theta_2 = 1/4$, since the polynomial $1 + u/2 + u^2/4$ has roots $-1 \pm \sqrt{3}$, which have complex modulus equal to 2. We find that the time series is invertible if and only if

$$|\theta_2| < 1 \text{ and } |\theta_1| < |1 + \theta_2|.$$

ii. non-invertible.

The time series is non-invertible for $\theta_1 = 1$ and $\theta_2 = 1$, since the polynomial $1 + u + u^2$ has roots $-1/2 \pm \iota\sqrt{3}/2$, which have complex modulus equal to 1.

iii. non-stationary.

There are no values of θ_1 and θ_2 for which the time series is non-stationary.

iv. m-dependent with $m = 1$.

Set $\theta_2 = 0$. Then we have m-dependence with $m = 1$.

For each case explain your answer.

(f) Let $f(\cdot)$ be the spectral density of a stationary time series with autocovariance function $\gamma(\cdot)$. Show that $\int_{-\pi}^{\pi} f(\lambda)d\lambda = \gamma(0)$.

We have

$$\begin{aligned}
\int_{-\pi}^{\pi} f(\lambda)d\lambda &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)\exp(-\iota h\lambda)d\lambda \\
&= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \int_{-\pi}^{\pi} [\cos(h\lambda) - \iota\sin(h\lambda)]d\lambda \\
&= \frac{1}{2\pi}\gamma(0) \int_{-\pi}^{\pi} [\cos(0) - \iota\sin(0)]d\lambda \\
&= \gamma(0).
\end{aligned}$$

(g) For the time series $\{X_t, t \in \mathbb{Z}\}$ in (1), give $\int_{-\pi}^{\pi} f(\lambda)d\lambda$ in terms of θ_1, θ_2, and σ^2.

By (f), we have

$$\int_{-\pi}^{\pi} f(\lambda)d\lambda = \gamma(0) = \sigma^2(1 + \theta_1^2 + \theta_2^2).$$

(h) Assuming that μ, θ_1, θ_2, and σ^2 are known, give the one-step-ahead predictor $\hat{X}_1$ of X_1 based on $X_1, \ldots, X_n$ and give the mean squared error of prediction (MSEP) associated with the predictor.

The predictor of X_1 is μ, and the MSEP is the value of v_0 from either the innovations or the Durbin-Levinson algorithm is $v_0 = \gamma(0) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$.

(i) Assuming that μ, θ_1, θ_2, and σ^2 are known, give the h-step-ahead predictors $\hat{X}_{n+h}$ of X_{n+h} for $h > 2$ based on $X_1, \ldots, X_n$ and give the MSEP associated with the predictors.

For $h > 2$, the predictor of X_{n+h} is simply μ, because the time series is m-dependent with $m = 2$. The MSEP is again equal to $\gamma(0) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$.

(j) Describe in words how you would generate a realization of length 100 of the time series $\{X_t, t \in \mathbb{Z}\}$, where $\{Z_t, t \in \mathbb{Z}\}$ are independent $\text{Normal}(0, \sigma^2)$ random variables.

Generate $Z_{-1}, Z_0, Z_1, \ldots, Z_{100}$ as independent $\text{Normal}(0, \sigma^2)$ random variables and then follow the formula in (1) to generate $X_1, \ldots, X_{100}$.

2. For each of the following time series, indicate which statements among the following four statements, if any, must be true.

(I) it is strictly stationary	(III) it is not strictly stationary
(II) it is stationary	(IV) it is not stationary

(a) Let $\{Y_t, t \in \mathbb{Z}\}$ be a time series such that the covariance matrix containing the covariances among $Y_1, \ldots, Y_5$ is given by

$$(\mathrm{Cov}(Y_i, Y_j))_{1 \le i,j \le 5} = \begin{bmatrix} 1 & 0.9 & 0 & 0 & 0 \\ 0.9 & 1 & 0.9 & 0 & 0 \\ 0 & 0.9 & 1 & 0.8 & 0 \\ 0 & 0 & 0.8 & 1 & 0.8 \\ 0 & 0 & 0 & 0.8 & 1 \end{bmatrix}$$

It is not stationary. It is not strictly stationary.

(b) Let $\{Y_t, t \in \mathbb{Z}\}$ be a time series such that $\mathbb{E}Y_t = \mu$ for all $t \in \mathbb{Z}$ and $\mathrm{Cov}(Y_i, Y_j) = (0.9)^{|i-j|}$ for $i, j \in \mathbb{Z}$.

It is stationary. We do not know if it is strictly stationary.

(c) Let $\{Y_t, t \in \mathbb{Z}\}$ be a time series such that

$$Y_t = (3/4)Y_{t-1} + (1/4)Y_{t-2} + Z_t, \quad t \in \mathbb{Z},$$

where $\{Z_t, t \in \mathbb{Z}\}$ are independent $\mathrm{Normal}(0, \sigma^2)$ random variables.

It is not stationary because the polynomial $1 - (3/4)u - (1/4)u^2$ has a root with unit complex modulus. Since it is not stationary it is also not strictly stationary.

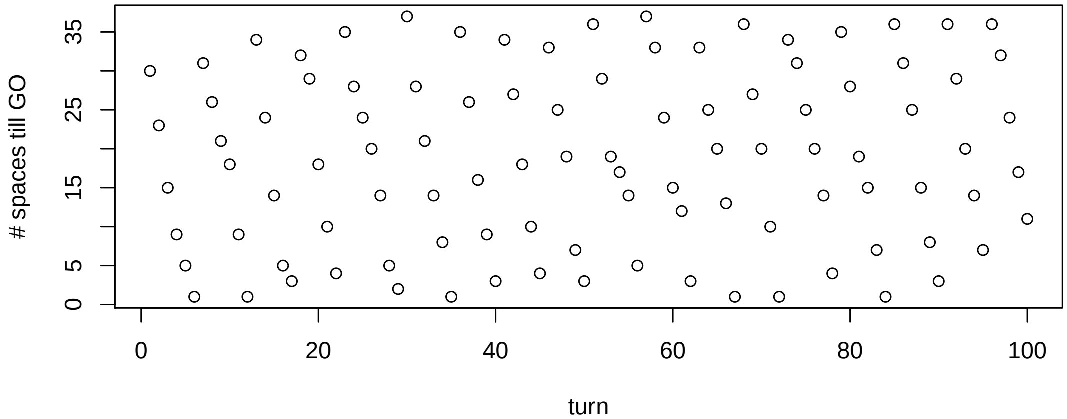
(d) Let $\{Y_t, t \in \mathbb{Z}\}$ be a time series such that $Y_t \sim \mathrm{Poisson}(\lambda)$ for all $t \in \mathbb{Z}$.

None of the statements must be true.

(e) Let $\{Y_t, t \in \mathbb{Z}\}$ be a time series such that $Y_t \sim \mathrm{Normal}(0, \sigma^2)$ for all $t \in \mathbb{Z}$ and $\mathrm{Cov}(Y_i, Y_j) = \mathbf{1}(|i - j| > 0)$ for $i, j \in \mathbb{Z}$, where $\mathbf{1}(\cdot)$ is the indicator function.

It is stationary. It is strictly stationary.

3. Suppose you play a game of Monopoly and you record after each turn the number of spaces you must move in order to reach the GO tile, resulting in a time series $\{X_t, t = 1, 2, \ldots\}$. There are 40 tiles on the Monopoly board; you begin on the GO tile and advance by moving a number of tiles equal to the sum of two dice rolls (the dice are six-sided). Your observed time series might look like this after 100 turns:



(a) Find $\mathbb{E}X_1$, where X_1 is the number of spaces you must move to reach the GO tile after your first turn.

Let Z_1 be the roll on the first turn. Then $X_1 = 40 - Z_1$. Since $\mathbb{E}Z_1 = 7$ we have $\mathbb{E}X_1 = 33$.

(b) Find $\mathbb{E}X_2$, where X_2 is the number of spaces you must move to reach the GO tile after your second turn.

Let Z_2 be the roll on the second turn. Then $X_2 = 40 - Z_1 - Z_2$, so $\mathbb{E}X_2 = 40 - 7 - 7 = 26$.

(c) Suggest a value for $\lim_{n\to\infty} \mathbb{E}X_t$ and argue why it is correct.

For each $t = 1, 2, \ldots,$ we can write

$$X_t = 40 - \left(\sum_{j=1}^{t} Z_t - \left\lfloor \frac{\sum_{j=1}^{t} Z_j}{40} \right\rfloor 40 \right),$$

where $Z_1, Z_2, \ldots$ are the rolls. For large t we find that X_t behaves like a discrete uniform distribution over the integers $1, \ldots, 40$, which has expected value 20.5. We can see this via simulation:

```
n <- 5000 # play for 5000 turns
N <- 40
S <- 1000 # do this 1000 times
XX <- matrix(NA,S,n)
for(s in 1:S)
{

        Z <- sample(1:6,n,replace=TRUE) + sample(1:6,n,replace=TRUE)
```

```
        XX[s,] <- N - cumsum(Z) %% N

}

hist(XX[,n])
mean(XX[,n])
```

(d) Describe a transformation of the time series $\{X_t, t = 1, 2, \ldots\}$ which would result in a strictly stationary time series.

Transform to get back to the rolls. Like this in R:

```
n <- 100
N <- 40
Z <- sample(1:6,n,replace=TRUE) + sample(1:6,n,replace=TRUE)
X <- N - (cumsum(Z) - floor(cumsum(Z)/N)*N)

ZfromX <- numeric(n)
ZfromX[1] <- N - X[1]
for(i in 2:n)
{
        ZfromX[i] <- ifelse(X[i] < X[i-1],X[i-1] - X[i],40 - X[i] + X[i-1])
}
```

4. Let U and V be random variables such that $\operatorname{Var} U = \sigma_U^2$, $\operatorname{Var} V = \sigma_V^2$, and $\operatorname{Cov}(U, V) = \sigma_{UV}$. You wish to predict the value of U using the value of V with a predictor of the form

$$\hat{U} = a_0 + a_1 V$$

for some $a_0, a_1 \in \mathbb{R}$.

(a) Find the values a_0 and a_1 which minimize the MSEP of the predictor $\hat{U}$; that is, find a_0 and a_1 which minimize $\mathbb{E}(U - \hat{U})^2$.

We obtain

$$a_0 = \mathbb{E}U - a_1 \mathbb{E}V \quad \text{and} \quad a_1 = \frac{\operatorname{Cov}(U, V)}{\operatorname{Var}(V)}.$$

(b) Give the MSEP of $\hat{U}$ under the choices of a_0 and a_1 from part (a).

The MSEP is

$$\operatorname{Var} U - \frac{\operatorname{Cov}(U, V)^2}{\operatorname{Var}(V)} = \operatorname{Var} U - \operatorname{Cov}(U, V)[\operatorname{Var}(V)]^{-1} \operatorname{Cov}(V, U).$$

5. Let $V_1, \dots, V_n, V_{n+1}$ be random variables with zero mean such that

$$\mathrm{Cov}(V_i, V_j) = \begin{cases} 1, & i = j \\ 1/4, & |i-j| \leq 3 \\ 0, & |i-j| > 3 \end{cases}$$

Consider a predictor $\hat{V}_{n+1}$ of V_{n+1} based on $V_1, \dots, V_n$ which is of the form

$$\hat{V}_{n+1} = \sum_{i=1}^{n} a_i V_{n+1-i}.$$

For $n = 2$, find the values of $a_1, \dots, a_n$ which minimize the MSEP of the predictor $\hat{V}_{n+1}$.

The values a_1 and a_2 which minimize the MSEP of the predictor are

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1/4 \\ 1/4 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 1/5 \end{bmatrix}.$$

6. Consider the causal $\mathrm{ARMA}(2,1)$ time series given by

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t + \theta_1 Z_{t-1}, \quad t \in \mathbb{Z},$$

where $\{Z_t, t \in \mathbb{Z}\}$ is $\mathrm{WN}(0, \sigma^2)$. Let $\{\psi_j, j = 0, 1, \dots\}$ be the coefficients such that we may write

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}.$$

Give ψ_0, ψ_1, ψ_2, and ψ_3, in terms of ϕ_1, ϕ_2, and θ_1.

We have

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \theta_1 + \phi_1 \\ \psi_2 &= \phi^2 + \phi_1\theta_1 + \phi_2 \\ \psi_3 &= \phi_1^3 + \phi_1^2\theta_1 + 2\phi_1\phi_2 + \phi_1\theta_1 \end{aligned}$$