ORI 390R.3 / ME384Q.3

Time-Series Analysis

Midterm Examination

Spring, 2018

Instructor: Dr. Dragan Djurdjanovic

University of Texas at Austin

Thursday, March 8th, 2018.

12:30pm-1:45pm at ETC 5.132

SOLUTIONS

Problem 1. (Total 25 points)

Part (a)
Complete the following 5 sentences using the choices offered in the table accompanying this problem (see the next page). Use the choice that best fits the missing field in the sentence. Also, note that some entries in Table 1.1 are not needed in this problem and that some entries could be used several times. Each field completed correctly brings one point.
(10 points)

Example:

Sentence: Expected value of a stationary time series _________ over time.

Available choices: 1. varies in a polynomial manner; 2 is constant; 3. is unstable; etc

The correct answer is **2** since the expected value of any stationary time series is constant over time. So, the completed sentence in your exam should look like this:

Sentence: Expected value of a stationary time series is ____2____ over time.

YOUR QUESTIONS ARE ON THE NEXT PAGE. PLEASE, TURN THE PAGE!

Sentence 1: The Green's Function $G_k = g_1\lambda_1^k + g_1\lambda_2^k + ... + g_n\lambda_n^k$ corresponds to an ___10___, where $\lambda_1, \lambda_2 ... \lambda_n$ are roots of the ___11___.

Sentence 2: Random walk is a special case of an ___8___ model with the root of the autoregressive characteristic polynomial being equal to ___3___.

Sentence 3: The variance of the eventual forecast error of an unstable system is ___6___. Ultimate forecast of a stable system tends towards ___2___

Sentence 4: Character of the inverse function of a time series following an ARMA model is determined by the ___9___.

Sentence 5: The formula $E[X_t X_{t-l}] = \gamma_l = \sigma_a^2 \sum_{k=0}^{\infty} G_k G_{k+l}$ describing the autocovariance of a discrete time-series X_t governed by an ___4___ model

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - ... - \phi_n X_{t-n} = a_t - \theta_1 a_{t-1} - ... - \theta_m a_{t-m} \quad a_t \sim NID(0, \sigma_a^2)$$

that has the impulse response $G_k, k = 0,1,2,...$ is derived using ___5___ of the discrete time series X_t.

Sentence 6: Time-series described by an ARMA model can be seen as a response of a dynamic system driven by ___1___.

Possible choices to complete sentences 1-6 of Problem 1.

1. White Noise	4. ARMA (n,m)	7. ARMA(1,1)	10. ARMA(n,n-1)
2. Zero	5. Orthogonal decomposition	8. AR(1)	11. Autoregressive characteristic polynomial
3. One	6. Infinite	9. Moving average characteristic polynomial	12. Finite

Part (b)
Circle the most appropriate answer offered to complete the following 2 statements and explain your answer. Each correct choice will bring 1 points and the appropriate explanation for that choice will bring another 4 points. Hence, there are 10 points to be won in this part (5 for each statement).
(total 15 points)

Example: A time series X_t described the ARMA(2,1) model

$$(1-0.4B)(1-0.9)X_t = (1-0.3B)$$

is stable/marginally stable/unstable.

Why:

The most appropriate answer is that this time series is stable and therefore your answer should look like this:

Example: A time series X_t described the ARMA(2,1) model

$$(1-0.4B)(1-0.9)X_t = (1-0.3B)$$

is stable/asymptotically stable/unstable.

The explanation for this choice is that all autoregressive roots of this ARMA model are inside the unit circle.

YOUR QUESTIONS ARE ON THE NEXT PAGE. PLEASE, TURN THE PAGE!

Start of the questions:

Statement 1. The system

$$X_t - 2X_{t-1} + X_{t-2} = a_t - 0.5a_{t-1} \quad a_t \sim NID(0, \sigma_a^2)$$

is stable | marginally stable | unstable

(circle the most appropriate answer and get 1 point)

Explain your answer (3 points)

Autoregressive char poly: $s^2 - 2s + 1 = (s-1)^2$

$\Rightarrow \lambda_1 = \lambda_2 = 1 \Rightarrow$ we have a repeated root at (i.e. ON the unit circle) $\Rightarrow$ unstable!

Statement 2. The system

$$X_t - 3X_{t-1} + 3X_{t-2} - X_{t-3} = a_t - 0.9a_{t-1} + 0.14a_{t-2} \quad a_t \sim NID(0, \sigma_a^2)$$

is invertible | marginally invertible | non-invertible

(circle the most appropriate answer and get 1 point)

Explain your answer (3 points)

Moving average characteristic poly:

$s^2 - 0.9s + 0.14 = (s - 0.2)(s - 0.7)$

$\Rightarrow \nu_1 = 0.2, \nu_2 = 0.7$ $\quad |\nu_1| < 1, |\nu_2| < 1 \Rightarrow$

$\Rightarrow$ this model is invertible!

Statement 3. The system

$$(1-B)(1+B+B^2)X_t = a_t - 2a_{t-1} + 3a_{t-2} \quad a_t \sim NID(0, \sigma_a^2)$$

is stable | marginally stable | unstable

(circle the most appropriate answer and get 1 point)

Explain your answer (3 points)

Autoregressive char. poly: $(s-1)(s^2+s+1)$

$\lambda_1 = 1, \quad \lambda_{2/3} = \frac{1}{2}(-1 \pm \sqrt{1-4}) = -\frac{1}{2} \pm j\frac{\sqrt{3}}{2}$

$|\lambda_1| = |\lambda_2| = |\lambda_3| = 1$ However, all these AR characteristic roots are distinct.

=> we have distinct roots on the unit circle =>

marginally stable

Statement 4. We have a system described by the ARMA model

$$X_t - 0.7X_{t-1} + 0.12X_{t-2} = a_t - \theta_1 a_{t-1}$$

It is known that this model is equivalent to the model

$$X_t - 0.4X_{t-1} = a_t$$

Then, it must be that $\theta_1 = 0.7$ / $\theta_1 = 0.3$ / $\theta_1 = 0$. (1 points)

Why? (2 points)

$X_t - 0.7X_{t-1} + 0.12X_{t-2} = (1-0.3B)(1-0.4B)X_t$

=> to make the original model look like $X_t - 0.4X_{t-1} = a_t$

we need to cancel the AR poly root at 0.3. Hence,

we need $\theta_1 = 0.3$

Problem 2. (Total 40 points)

Gate-opening in the papermaking process was found to be governed by the ARMA(2,1) model

$$X_t - 0.6X_{t-1} + 0.08X_{t-2} = a_t - 0.3a_{t-1},\ a_t \sim NID(0, \sigma_a^2)$$

where $\sigma_a^2 = 4$.

Part a (15 points)

Explicitly express the corresponding Green's function.

$X_t - 0.6X_{t-1} + 0.08X_{t-2} = (1-0.4B)(1-0.2B)X_t \Rightarrow$ AR char. poly.

roots are $\lambda_1 = 0.2$ and $\lambda_2 = 0.4$

$G_l = g_1\lambda_1^l + g_2\lambda_2^l$ where

$g_1 = \dfrac{\lambda_1 - \theta_1}{\lambda_1 - \lambda_2} = \dfrac{0.2-0.3}{0.2-0.4} = \dfrac{1}{2}$; $g_2 = \dfrac{\lambda_2 - \theta_1}{\lambda_2 - \lambda_1} = \dfrac{0.4-0.3}{0.4-0.2} = \dfrac{1}{2}$

$\Rightarrow G_l = \frac{1}{2}\cdot 0.2^l + \frac{1}{2}\cdot 0.4^l$

More on the next page

Part b (15 points)

Explicitly express the covariance function $E[X_l X_{l-l}] = \gamma_l$ for the time-series described by the ARMA(2,1) model given above.

$$\gamma_l = \sigma_a^2 \lambda_1^l \left(\frac{g_1^2}{1-\lambda_1^2} + \frac{g_1 g_2}{1-\lambda_1 \lambda_2} \right) +$$

$$\sigma_a^2 \lambda_2^l \left(\frac{g_2^2}{1-\lambda_2^2} + \frac{g_1 g_2}{1-\lambda_1 \lambda_2} \right)$$

$$= 2.1 \cdot 0.2^l + 2.3 \cdot 0.4^l \quad \text{for } l \geq 0$$

$$\boxed{\gamma_l = \gamma_{-l}}$$

Part c (10 points)

Find the inverse function decomposition

$$X_t = a_t + I_1 X_{t-1} + I_2 X_{t-2} + I_3 X_{t-3} + I_4 X_{t-4} + \dots$$

of the gate-opening time-series.

$$(1 - 0.6B + 0.08B^2) X_t = (1 - 0.3B) a_t$$

$$\Rightarrow a_t = \frac{1 - 0.6B + 0.08B^2}{1 - 0.3B} X_t = \left(1 - 0.3B - 0.01B^2 \frac{1}{1-0.3B}\right) X_t$$

$$= \left(1 - 0.3B - \sum_{l=2}^{\infty} 0.01 \cdot 0.3^{l-2} B^l\right) X_t$$

$$I_0 = -1 \qquad I_1 = 0.3 \qquad I_n = 0.01 \cdot 0.3^{n-2} \quad \text{for } n \geq 2$$

Problem 3. (Total 35 points)

A researcher decided to use ARMA modeling techniques to forecast behavior of a welding process. He performed 4000 welding operations and for every tenth of them recorded the frequency at which the sensor signal displayed the most energy. Thus, he collected a time series of N=400 samples of the frequency locations of energy peaks. In order to forecast their behavior, he decided to fit an ARMA model to this time-series and Table 1 describes the results of fitting several ARMA models.

Parameters	AR(2)	ARMA(2,1)	ARMA(3,2)	ARMA(4,3)
Expected value for X_t μ_X (95% C.I.)	304 ± 20	300 ± 10	301 ± 11	302 ± 12
ϕ_1 (95% C.I.)	1.1 ± 0.7	0.9 ± 0.5	0.85 ± 0.3	0.5 ± 0.2
ϕ_2 (95% C.I.)	-0.3 ± 0.2	-0.2 ± 0.1	0.2 ± 0.2	-0.7 ± 0.1
ϕ_3 (95% C.I.)			0.1 ± 0.3	0.8 ± 0.1
ϕ_4 (95% C.I.)				-0.5 ± 0.2
θ_1 (95% C.I.)		0.3 ± 0.4	0.3 ± 0.3	-0.3 ± 0.2
θ_2 (95% C.I.)			0.1 ± 0.3	-0.7 ± 0.3
θ_3 (95% C.I.)				0.6 ± 0.2
Residual Sum of Squares (RSS)	1490	1472	1465	1440

Part a (12 points)

The researcher used the ARMA(2n,2n-1) modeling technique. What model did he select if he used 95% confidence intervals to test the significance of the change of the corresponding Residual Sum of Squares (RSS). From the standard F-statistics table it is known that $F_{0.95}(1,\infty) = 3.84$, $F_{0.95}(2,\infty) = 3.00$, $F_{0.95}(3,\infty) = 2.60$, $F_{0.95}(4,\infty) = 2.37$, $F_{0.95}(5,\infty) = 2.21$, $F_{0.95}(6,\infty) = 2.10$ etc. (this is enough for you to solve this problem).

Note: For full credit, please conduct the tests in the appropriate order.

Test I : ARMA (4,3) vs ARMA(2,1)

$A_0 = 1440$; $A_1 = 1473$; $s = 4$; $r = 8$; $N = 400$

$$F = \frac{(A_1 - A_0)/s}{A_0/(N-r)} = 2.18 < F_{0.95}(4,\infty) = 2.37$$

=> Stay with ARMA(2,1)

Test 2: ARMA(2,1) VS AR(2)

(doing it since the confidence interval of θ_1 in the ARMA(2,1) model encompasses the zero)

$A_0 = 1472; \; A_1 = 1490; \; s = 1; \; r = 4; N = 400$

$$F = \frac{(A_1 - A_0)/s}{A_0/(N-r)} =$$

$$= 4.842 > F_{0.95}(1, \infty) = 3.84$$

$\Rightarrow$ must keep ARMA(2,1)

Part b (8 points)

For the model found in part a, find the corresponding discrete-time impulse response.

We keep the ARMA(2,1) model

$$X_t - 0.9 X_{t-1} + 0.2 X_{t-2} = a_t - 0.3 a_{t-1}$$

$$\lambda_{1/2} = \frac{1}{2}\left(0.9 \pm \sqrt{0.9^2 - 4 \cdot 0.2}\right) \Rightarrow \lambda_1 = 0.4 ; \lambda_2 = 0.5$$

$$G_l = g_1 \lambda_1^l + g_2 \lambda_2^l , \quad g_1 = \frac{\lambda_1 - \theta_1}{\lambda_1 - \lambda_2} = \frac{0.4 - 0.3}{0.4 - 0.5} = -1$$

$$g_2 = \frac{\lambda_2 - \theta_1}{\lambda_2 - \lambda_1} = \frac{0.5 - 0.3}{0.5 - 0.4} = 2$$

$$\Rightarrow G_l = -(0.4)^l + 2 \cdot (0.5)^l , \quad l = 0, 1, 2, \ldots$$

Part c (6 points)

A young researcher wants to forecast where the frequency locations of the signal energy peaks will be after 20 more welding operations are performed and therefore decides to find a two step ahead prediction at time t=400 based on the model found in part a. Knowing that $\dot{X}_{400} = 305$ Hz, $\dot{X}_{399} = 297$, $a_{400} = 2$, calculate the predicted energy peak location in the frequency domain for him.

Hint: Do not forget to remove the mean value μ_X from the data before employing the ARMA model from part a to find $\hat{X}_{400}(2)$, and then put μ_X back when calculating the final prediction.

$\hat{X}_{400}(2) = ?$ $\quad \underbrace{X_{t+2} - 0.9X_{t+1} + 0.2X_t = a_{t+2} - 0.3a_{t+1}}_{E[\cdot | t]}$

$\hat{X}_t(2) - 0.9\hat{X}_t(1) + 0.2X_t = 0 \Rightarrow$

$\hat{X}_{400}(2) - 0.9\hat{X}_{400}(1) + 0.2X_{400} = 0$ where $X_{400} = \dot{X}_{400} - \mu_X$

$\uparrow ?$

$= 305 - 300 = 5$

$X_{t+1} - 0.9X_t + 0.2X_{t-1} = a_{t+1} - 0.3a_t \Leftarrow E[\cdot|t]$

$\Rightarrow \hat{X}_t(1) - 0.9X_t + 0.2X_{t-1} = -0.3a_t \Rightarrow$

$\hat{X}_{400}(1) - 0.9X_{400} + 0.2X_{399} = -0.3a_{400}$ where

$X_{399} = \dot{X}_{399} - \mu_X = -3$

$\Rightarrow \hat{X}_{400}(1) = -0.3a_{400} - 0.2X_{399} + 0.9X_{400} = 4.5$

$\Rightarrow \hat{X}_{400}(2) = 0.9\hat{X}_{400}(1) - 0.2X_{400} = 3.05$

$\Rightarrow \hat{\dot{X}}_{400}(2) = \mu_X + \hat{X}_{400}(2) = 300 + 3.05 = 303.05$

Part d (9 points)

What is the 95% confidence interval for the 2-step ahead predicted energy peak location in the frequency domain that you calculated for him in the part c.

Hint: The 95% confidence interval will be the predicted value you found in part c plus/minus $1.96\ \sqrt{Var\left[(X_{t+2}-\hat{X}_t(2))\right]} = 1.96\ \sqrt{E\left[(X_{t+2}-\hat{X}_t(2))^2\right]}$ - so, you need to find the variance of the 2-step ahead prediction error for the model in you selected in part a.

$$Var[\hat{e}_{400}(2)] = Var[\hat{e}_t(2)] = \sigma_a^2 [G_0^2 + G_1^2]$$

$$X_t = (G_0 + G_1 B + G_2 B^2 + \dots) a_t$$

$$(1 - 0.9B + 0.2B^2) X_t = (1 - 0.3B) a_t \quad \Rightarrow$$

$$(1 - 0.9B + 0.2B^2)(G_0 + G_1 B + G_2 B^2 + \dots) a_t = (1 - 0.3B) a_t$$

$$B^0: \quad 1 = G_0$$

$$B^1: \quad -0.9\, G_0 + G_1 = -0.3 \Rightarrow G_1 = -0.3 + 0.9 = 0.6$$

$$\sigma_a^2 \approx \frac{RSS}{N-r} = \frac{1472}{400-4} = 3.72$$

r ← number of estimated parameters

$\Rightarrow$ 95% confidence interval will be

$$303.05 \pm 1.96 \cdot \sqrt{Var[\hat{e}_t(2)]} =$$

$$303.05 \pm 1.96 \sqrt{3.72(1^2 + 0.6^2)} =$$

$$303.05 \pm 4.41$$