Generating and studying time series based on AR(1) model

Pranav Gopalkrishna, 1940223

2022-02-08

# INTRODUCTION ON AUTOREGRESSIVE MODELS

An autoregressive (AR) model is a statistical time series model wherein each observation is linearly dependent on past observations. The general equation of an AR model of order *p* is given by:

where

* *p* is the maximum lag at which past observations are included, i.e.  depends on all past observations until
* is the error term for time t i.e. difference between the actual observation at time t and the estimated observation at time t based on the other deterministic components
* is the *i*th constant coefficient i.e. the constant coefficient for the observation at lag *i*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Our focus is on AR(1) models i.e. AR models wherein current observations only depend on the immediately preceding observation. The general equation is given by:

# GENERATING RANDOM AR(1) PROCESSES

Here, we will be generating a random time series based on AR(1) model, with a given ϕ (phi) value. In other words, we will be generating a time series model that can be accurately modelled by AR(1) model for the given phi value.

## Note on ARIMA

**ARIMA** => **A**utoregressive **I**ntegrated **M**oving **A**verage

ARIMA is a generalization of AR models, MA (moving average) models and I (integrated) models. These components are explained below:

1. ***AR (Autoregression)***  
Current values depend on some number of past observations up to a certain lag.

2. ***I (Integrated)***  
A time series is said to be integrated of order d if taking repeated differences between zt values results in a stationary process. For example, if zt is integrated of order 2, then  
zt - zt-1 - zt-2  
is a stationary process. Here, ‘integrated’ refers the property of the time series being integrated at some order, i.e. it refers to time series that are non-stationary with respect to mean, but can become stationary after some number of repeated differencing.

3. ***MA (Moving Average)***  
The value at each time point is smoothened by averaging some number of past and future values around it.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ARIMA model has three integer components, namely:

1. ***Order of autoregression*** of the time series (p). This indicates the maximum lag at which observations are included i.e. extent to which past values are used to model current values.

2. ***Order of integration*** of the time series (d). This specifies the non-seasonal (i.e. trend) part of the model.

3. ***Order of moving average*** of the time series (q). This indicates the number of values around and including the current time point’s value that are averaged, in order to smoothen the current time point’s value. For example, MA(3) implies that each time point’s value is replaced by the average of the just preceding, current and just succeeding value.

## Generating AR(1) based time series using ARIMA model

*This is important, since R provides a function for generating time series based on ARIMA models, but not AR models in particular.*

ARIMA models are denoted as ARIMA(p, d, q). Note that here, only trend is considered here, not seasonality. To include a seasonal component, we must use seasonal ARIMA, which is beyond the scope of this assignment.

ARIMA generalizes AR models, differencing models and MA models. Hence, we can use the function for generating time series based on ARIMA model to generate time series based on AR model as well. In particular, ARIMA(1, 0, 0) = AR(1) (i.e. p = 1, d = 0, q = 0).

## Generating AR(1) based time series using R code

To generate a time series based on AR(1), we can use the function **arima.sim**, which generates time series based on ARIMA model. The arguments required are as follows:

1. ***A list with component ‘ar’, ‘ma’ (or both)*** giving the AR and MA coefficients respectively. For our purposes, we only need to give coefficients for AR, and only one coefficient at that, since we want to generate a time series based on AR(1) model. An empty list gives an ARIMA(0, 0, 0) model, that is white noise.
2. ***The length of the output series (n)***. This is a strictly positive integer.

### General function

(For generating time series based on AR(1) model, along with its time plot & ACF plot)

new\_AR1\_ts = function(phi, n)  
{  
 AR1 = arima.sim(model = list(ar = phi), n = n)  
 print(summary(AR1))  
   
 # Time plot  
 ts.plot(  
 AR1,  
 main = paste("AR(1) based time series (ϕ = ", phi, ")", sep = ''),  
 xlab = "Time",  
 ylab = "Value")  
   
 # ACF plot  
 acf(  
 AR1,  
 main = paste("ACF plot for AR(1) based time series (ϕ = ", phi, ")", sep = ''))  
}

**Given phi values:**

phiValues = c(-0.8, -0.5, -0.3, 0.3, 0.7, 0.8)

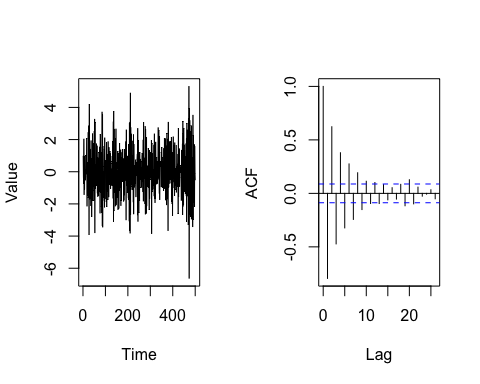
**Generating time series, time plots & ACF plots:**

par(mfrow = c(1, 2))  
for(phi in phiValues){new\_AR1\_ts(phi, 500)}

#### TIME SERIES BASED ON AR(1) WITH ϕ = -0.8

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -6.63186 | -1.09345 | -0.06418 | 0.00962 | 1.16269 | 5.30616 |



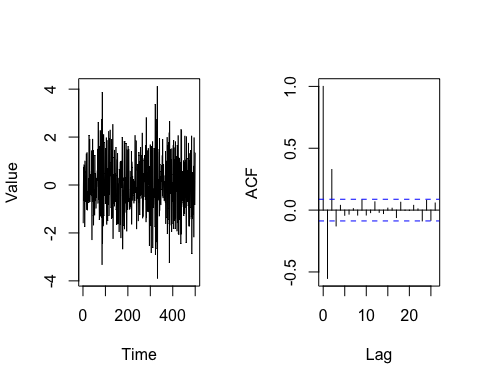
The time plot shows no clear periodic motion, neither seasonal nor cyclic. There seems to be no trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance. Due to these features, the time series resembles a white noise sequence.

The ACF plot indicates significant autocorrelation for lags below 10, suggesting that the time series is autoregressive with an order between 5 and 10 i.e. an observation at a given time point may be significantly affected by up to between 5-10 past observations. Since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations. The oscillating autocorrelation coefficients suggest that there is no clear increase or decrease of in the level of the values, indicating the lack of trend, as we noticed in the time plot.

#### TIME SERIES BASED ON AR(1) WITH ϕ = -0.5

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -3.89724 | -0.73979 | -0.01935 | 0.01604 | 0.84169 | 4.11420 |



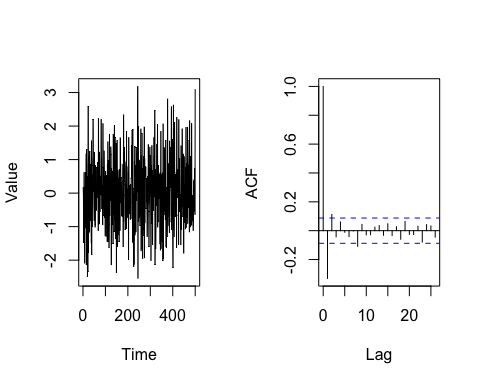
The time plot shows no clear periodic motion, neither seasonal nor cyclic. There seems to be no trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance. Due to these features, the time series resembles a white noise sequence.

The ACF plot indicates significant autocorrelation for lags up to 2, suggesting that the time series is autoregressive with an order between 1 and 2 i.e. an observation at a given time point may be significantly affected by up to 2 past observations. Since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations. The opposing autocorrelation coefficient signs for lags 1 and 2 indicate the lack of trend, since the values do not seem to move in a clear direction upwards or downwards.

#### TIME SERIES BASED ON AR(1) WITH ϕ = -0.3

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -2.54373 | -0.71046 | 0.02612 | 0.03276 | 0.73020 | 3.18307 |



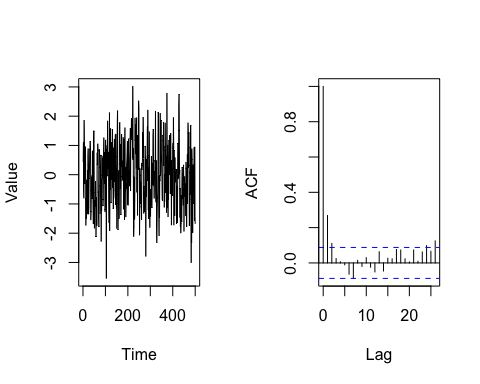
The time plot shows no clear periodic motion, neither seasonal nor cyclic. There seems to be no trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance. Due to these features, the time series resembles a white noise sequence.

The ACF plot indicates significant autocorrelation only for lags up to 2, with autocorrelation at lag 2 being barely above the threshold value, suggesting that the time series is autoregressive with order 1 i.e. an observation at a given time point may be significantly affected by the immediately past observations. Since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations. The autocorrelation at lag 2 is also in the opposite direction as the autocorrelation at lag 1, suggesting lack of trend.

#### TIME SERIES BASED ON AR(1) WITH ϕ = 0.3

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -3.545375 | -0.776052 | 0.002996 | -0.011191 | 0.760251 | 3.017933 |



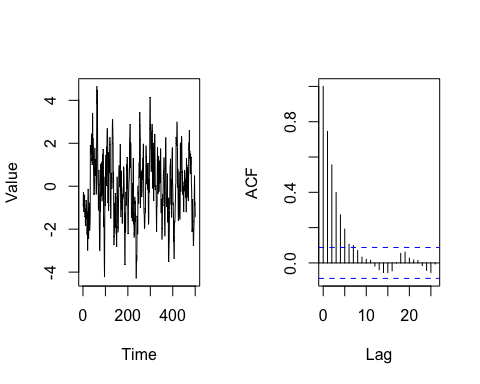
The time plot shows no clear periodic motion, neither seasonal nor cyclic. There seems to be no trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance. However, there seems to be a long-term convex movement, peaking at around t = 200. This may be a random feature, not indicative of any feature of the time series. Alternatively, it may indicate potential non-stationary movement, maybe even cyclic movement.

The ACF plot indicates significant positive autocorrelation only for lags up to 2, with autocorrelation at lag 2 being barely above the threshold value, suggesting that the time series is autoregressive with order 1 i.e. an observation at a given time point may be significantly affected by the immediately past observations. Since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations. The autocorrelation at lag 2 is also in the same direction as the autocorrelation at lag 1, unlike all the above time series, indicating slightly more long-term movements in one direction in the time series.

#### TIME SERIES BASED ON AR(1) WITH ϕ = 0.7

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -4.29261 | -1.06220 | -0.09355 | -0.06457 | 0.90465 | 4.64858 |



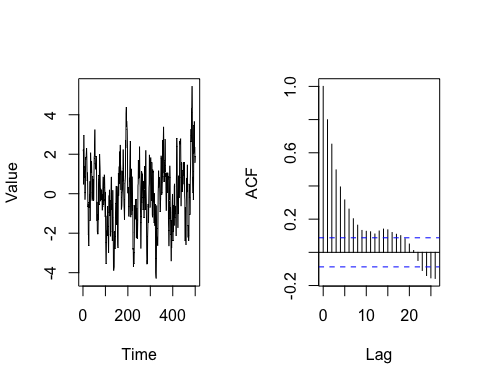
The time plot shows no clear periodic motion, neither seasonal nor cyclic. However, we see more long-term movements, that are not necessarily periodic in nature. As a result, unlike most of the above time plots, its observations oscillate less frequently around the mean, creating a less visually dense time plot. There seems to be no trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance.

The ACF plot indicates significant positive autocorrelation for lags below 7, suggesting that the time series is autoregressive with an order between 5 and 7 i.e. an observation at a given time point may be significantly affected by up to between 5-7 past observations. Since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations. One noticeable difference between this time series and the ones above it is that of its significant autocorrelation coefficients are all positive, indicating more long-term movements in one direction in the time series (other than trend or seasonality). This is apparent in the time plot, which is less visually dense compared to the other time plots, even though it is plotted in the same scale. This is due to observations oscillating less frequently around the mean, due to the more long-term movements in the singular directions.

#### TIME SERIES BASED ON AR(1) WITH ϕ = 0.8

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -4.289528 | -1.121804 | 0.056698 | 0.004155 | 1.054646 | 5.435909 |



The time plot shows no clear periodic motion, neither seasonal nor cyclic. However, we see more long-term movements, that are not necessarily periodic in nature. As a result, unlike most of the above time plots, its observations oscillate less frequently around the mean, creating a less visually dense time plot. There seems to be no clear trend, with the mean of the time plot being constant at around zero. There are many and varied irregular fluctuations, which seem to lie within a fixed range and do not skew in any direction, suggesting constant variance.

The ACF plot indicates significant positive autocorrelation for lags below 20, suggesting that the time series is autoregressive with an order between 15 and 20 i.e. an observation at a given time point may be significantly affected by up to between 15-20 past observations. We also see significant negative correlation lag 23 onwards. This may suggest a much longer long-term movement, or may simply be a random effect. Regardless, since every 12 observations are clearly not autocorrelated, we can confirm the lack of seasonality in the time series. Similarly, we do not see similar autocorrelation between observations at certain fixed intervals, ruling out cyclic fluctuations, at least for now. One noticeable difference between this time series and the ones above it is that most of its significant autocorrelation coefficients are positive, indicating more long-term movements in one direction in the time series (other than trend or seasonality). This is apparent in the time plot, which is less visually dense compared to the other time plots (even the previous time plot, which had significant positive autocorrelation up to 7 lags), even though it is plotted in the same scale. This is due to observations oscillating less frequently around the mean, due to the more long-term movements in the singular directions.

# GENERATING AND COMPARING TWO ACF(1) BASED TIME SERIES

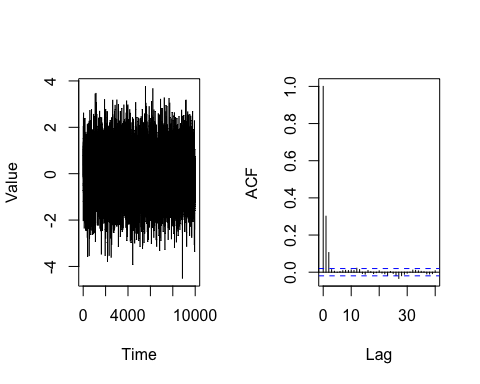
Here, we put n = 10000. The phi values for the two AR(1) models (based on which the two time series will be generated) are 0.3 and 0.9 respectively.

par(mfrow = c(1, 2))  
for(phi in c(0.3, 0.9)){new\_AR1\_ts(phi, 10000)}

## TS1: TIME SERIES BASED ON AR(1) WITH ϕ = 0.3

**Summary**

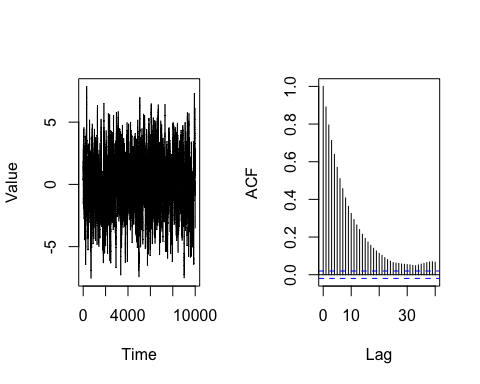
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -4.515424 | -0.716101 | -0.001655 | -0.003016 | 0.711233 | 3.764128 |



## TS2: TIME SERIES BASED ON AR(1) WITH ϕ = 0.9

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Minimum | 1st Quarter | Median | Mean | 3rd Quarter | Maximum |
| -7.55264 | -1.42917 | 0.09445 | 0.03950 | 1.49600 | 7.87188 |



## Observations & comparisons

The time plots for both time series are centered around a constant mean of zero, and have many and varied irregular fluctuations.

One clear difference we can notice in the ACF plots is that for TS1, the autocorrelation coefficients are significant and positive only for up to lag 2, after which they are insignificant, while for TS2, autocorrelation coefficients are significant and positive for up to lag 40.

This suggests that TS2 exhibits much more long-term movement in singular directions (potentially cyclical movement, which we cannot confirm, since the autocorrelation pattern does not repeat for the given number of lags). This is visible in the time plots, as the time plot for TS2 is visibly less dense than the time plot for TS1, due to fewer oscillations around the mean.