

Project Write-up of ‘Visualization of Autoregressive(AR) Models’ Project

DataWaves - Libin N George, Ajeet Singh, Saharsh Bhawe, Navyasree Madhu

Objective

Presenting and visualizing technical topic in statistics, focusing on Autoregressive (AR) models, to demonstrate how parameters affect patterns in time series data.

Introduction

This project aims to explore and visually demonstrate how autoregressive (AR) models operate, specifically how different parameters and AR orders impact the structure and behavior of a time series. AR models are essential tools in time series analysis, with broad applications across economics, finance, weather forecasting, and more. Understanding the visual and statistical characteristics of AR(1), AR(2), and AR(3) models, as well as how different parameter values influence the model's output, provides insight into data pattern recognition and prediction capabilities.

We will use a generated dataset to simulate various AR model scenarios. This approach allows us to precisely control parameters and observe how each setting impacts the time series and how to enable precise, tailored demonstrations of AR model behavior. Our visualizations will feature multiple plots for AR(1), AR(2), and AR(3) models, showcasing stationary and non-stationary processes, cyclical patterns, and convergence to the mean. We will also develop an interactive Shiny app that allows users to work on different AR model parameters and visualize the resulting time series. This approach helps to understand through direct engagement.

Real-World Applications of AR Models

Autoregressive(AR) models are foundational tools in time series analysis and have significant real-world applications across various domains. In finance, they are used to forecast stock

prices, assess risks, and optimize portfolios. In meteorology, AR models aid in weather prediction and climate analysis, while in healthcare, they help track disease outbreaks and forecast hospital resource needs. Industries like energy and telecommunications rely on AR models for demand forecasting and network traffic prediction. These applications highlight the power of AR models in analyzing historical patterns to make informed predictions, enabling better decision-making and resource optimization.

Generic AR Model Equation

The general formula for an auto regressive model of order p , or $AR(p)$, is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

where:

- X_t is the value of the time series at time t ,
- $\phi_1, \phi_2, \dots, \phi_p$ are the auto regressive coefficients,
- p is the order of the AR model, and
- ϵ_t is a white noise error term at time t , with mean zero and constant variance.

Justification of approach

To understand the behavior of AR (AutoRegressive) models under different configurations, we can simulate time series data using the `arma.sim()` function in R. This function is well-suited for generating synthetic data that follows ARIMA processes, allowing us to study how different AR and MA configurations affect the patterns in a time series. Specifically, we use it to generate data for AR models with varying parameters. In the data analysis plan, we did the following variations in AutoRegressive models and generate data for these models and created visualizations from the data.

1. $AR(1)$ - Random Walk
2. $AR(1)$ - Having moderate positive correlation, decaying toward the mean over time.
3. $AR(1)$ - Data will oscillate, alternating signs and showing a negative correlation
4. $AR(2)$ - Data with oscillatory behavior
5. $AR(2)$ - Data without oscillatory behavior
6. $AR(3)$ - Data with Complex behavior We created a line graph and associated Autocorrelation graph for the above models to visualize the statistical information and how the data behaves. We have included a shiny app that enables the user to change parameters and plot the line graph.

Code

```
library(ggplot2)
```

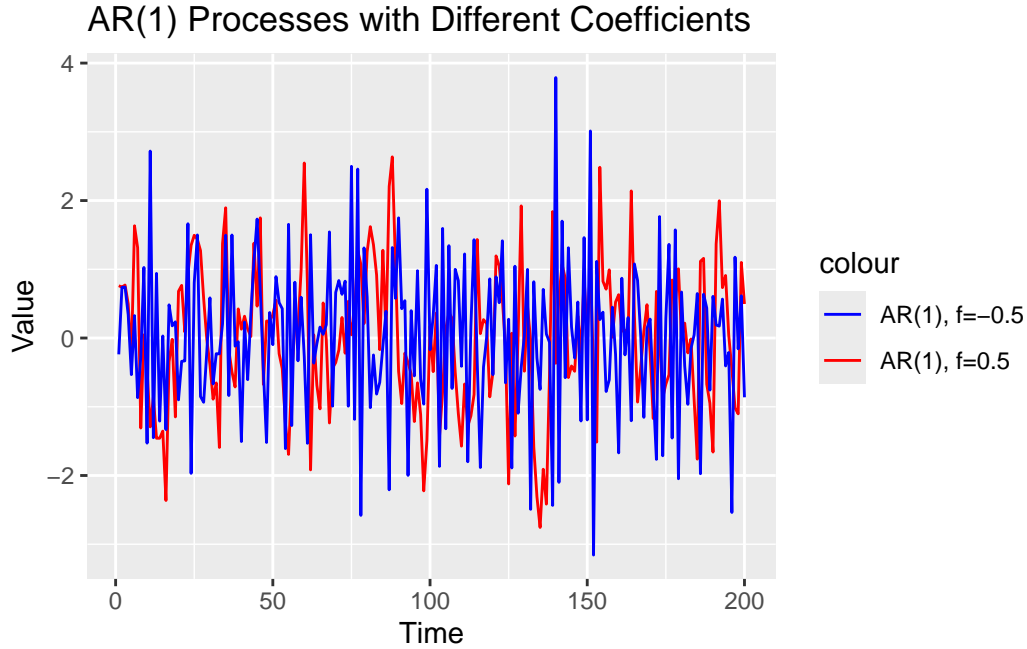
AR(1) model

- Formula: $X_t = \phi_1 X_{t-1} + \epsilon_t$
- The series depends only on the immediately preceding value.
- Effect of ϕ_1 :
 - o $|\phi_1| < 1$: The series will be stationary and will show a gradual decay toward the mean, with fluctuations around the mean.
 - o $|\phi_1| = 0.5$: The series will show moderate positive correlation, decaying toward the mean over time.
 - o $|\phi_1| = -0.5$: The series will oscillate, alternating signs and showing negative correlation.
 - o $\phi_1 = 1$: The series will behave like a random walk, non-stationary.
 - o $\phi_1 > 1$: The series will diverge, growing without bounds.

1. AR(1) with positive

```
set.seed(123)
ar1_pos <- arima.sim(model = list(ar = 0.5), n = 200)
ar1_neg <- arima.sim(model = list(ar = -0.5), n = 200)
library(ggplot2)
plot_data <- data.frame(Time = 1:200, AR_Positive = ar1_pos, AR_Negative = ar1_neg)
ggplot(plot_data, aes(x = Time)) +
  geom_line(aes(y = AR_Positive, color = "AR(1), =0.5")) +
  geom_line(aes(y = AR_Negative, color = "AR(1), =-0.5")) +
  labs(title = "AR(1) Processes with Different Coefficients", y = "Value") +
  scale_color_manual(values = c("blue", "red"))
```

Don't know how to automatically pick scale for object of type <ts>. Defaulting to continuous.



The Behavior of the Model Low Coefficient ($\phi=0.3$) The value of $\phi=0.3$ is relatively low, which means the influence of X_{t-1} is moderate. The series will exhibit limited persistence; therefore, the deviations from the mean will fade quickly as the lagged influence diminishes over time. Stationarity The process is stationary because $\phi < 1$. This implies that the time series will fluctuate around a constant mean (0 for this example). The variance of the series remains finite and does not grow over time. White Noise Contribution Random white noise (ϵ_t) plays a more significant role in shaping the series than lagged values. This results in a more stochastic (less deterministic) series. Smoothness and Short-Term Correlation The small ϕ value results in limited smoothness; is only weakly correlated with X_{t-1} . Short-term trends are minimal, and the series appears more random. Mean-Reverting Property The series will revert to the mean quickly after a deviation because the influence of the lagged term diminishes rapidly.

2. AR(1) with negative coefficients

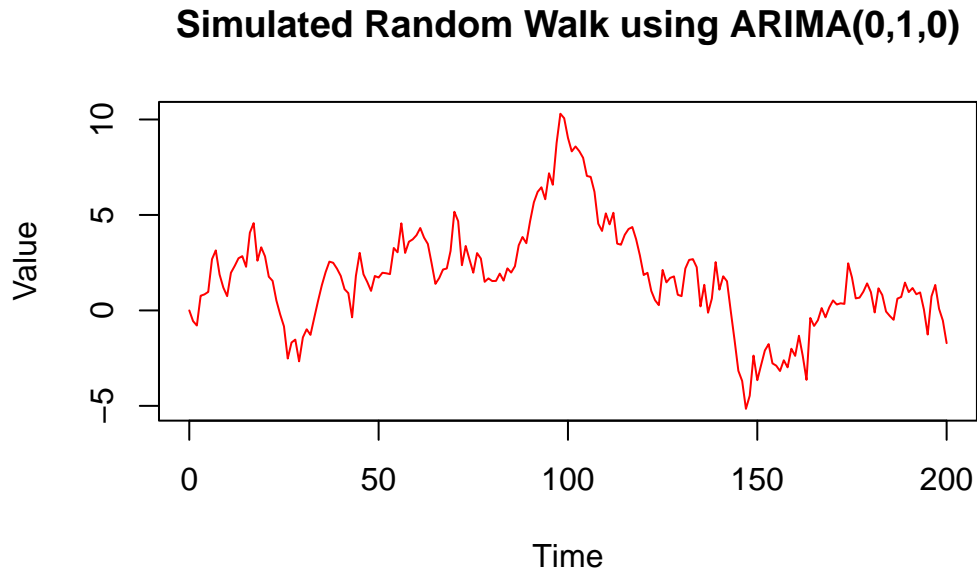
The Behavior of the Model Low Negative Coefficient ($\phi=-0.3$) A small negative coefficient indicates: A weak negative correlation between X_t and X_{t-1} . X_t tends to move in the opposite direction to X_{t-1} but with low persistence. Oscillatory Behavior The negative coefficient introduces an alternating pattern: If X_{t-1} is positive, X_t is more likely to be negative, and vice versa. However, the oscillations are not pronounced due to the small magnitude of ϕ . Stationarity The process is stationary because $\phi < 1$. This ensures the series fluctuates around a constant mean (0 for this example). The variance remains stable over time, preventing explosive growth or decay. Randomness and White Noise Random white noise (ϵ_t) dominates the series: The weak ϕ value means that the current value is heavily influenced by random

noise rather than the lagged term. Limited Persistence The impact of deviations from the mean dissipates quickly because ϕ is small. The series lacks strong trends or long-term dependencies.

3. Random Walk

```
set.seed(123)
random_walk_arima <- arima.sim(model = list(order = c(0, 1, 0)), n = 200)

# Plotting the random walk (ARIMA(0,1,0))
plot(random_walk_arima, type = "l", col = "red",
      main = "Simulated Random Walk using ARIMA(0,1,0)",
      xlab = "Time", ylab = "Value")
```



4. White noise

AR(2) Model

- Formula: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$
- More complex dynamics emerge as the series now depends on the previous two points.
- Effect of ϕ_1 and ϕ_2 :
 - Oscillations: If $\phi_2 < 0$, the series

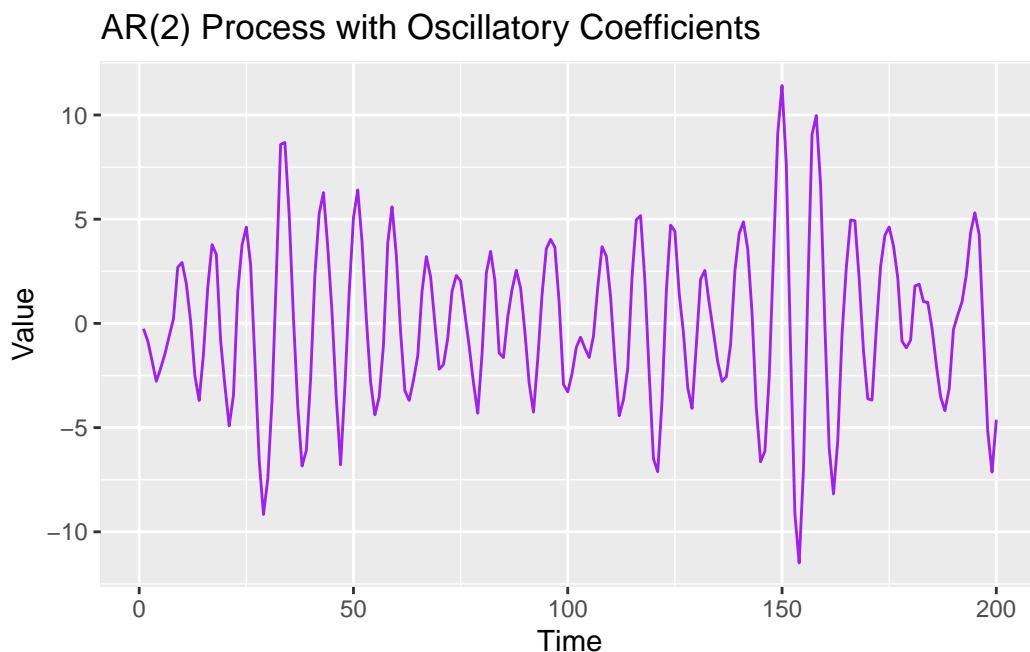
may display oscillatory or cyclical behavior.

- o Damping: If both coefficients are positive but less than 1, the series will dampen oscillations toward the mean.
- o Explosive behavior: Larger values of ϕ_1 and ϕ_2 lead to explosive growth in the series.

AR(2) with oscillatory behavior

```
set.seed(123)
ar2_oscillatory <- arima.sim(model = list(ar = c(1.4, -0.9)), n = 200)
plot_data <- data.frame(Time = 1:200, AR_Oscillatory = ar2_oscillatory)
ggplot(plot_data, aes(x = Time, y = AR_Oscillatory)) +
  geom_line(color = "purple") +
  labs(title = "AR(2) Process with Oscillatory Coefficients", y = "Value")
```

Don't know how to automatically pick scale for object of type <ts>. Defaulting to continuous.



AR-2 negative coefficient

The Behavior of the Model Negative Coefficients ($\phi_1 = -0.1$, $\phi_2 = -0.3$) The negative coefficients imply that both X_{t-1} and X_{t-2} exert negative influences on X_t . X_t moves in the opposite

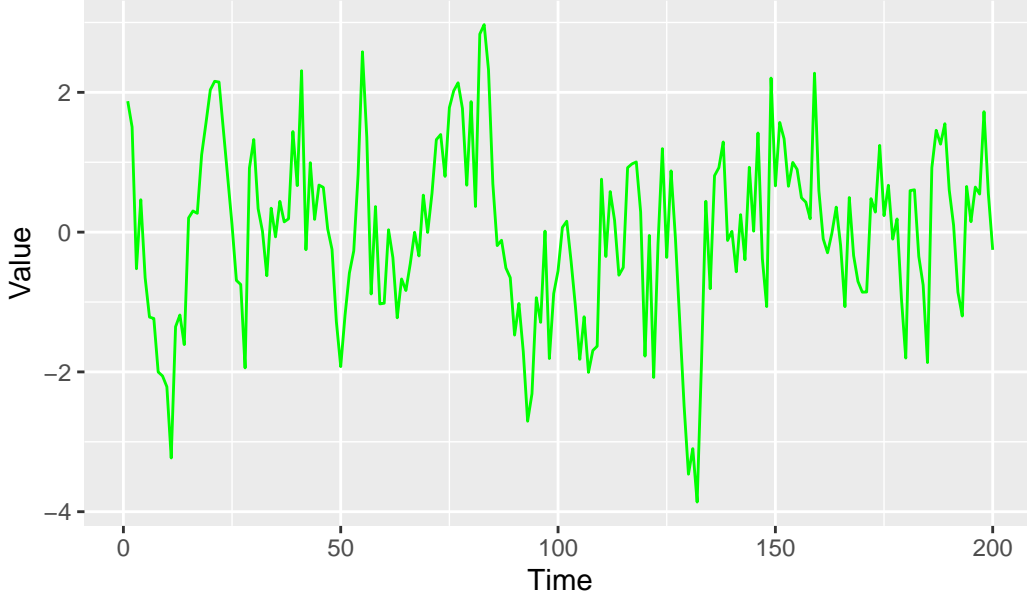
direction of both X_{t-1} and X_{t-2} . The degree of negative correlation is higher with X_{t-2} ($\phi_2=-0.3$) than X_{t-1} ($\phi_1=-0.1$). Oscillatory Behavior The presence of negative coefficients, especially for both lags, results in oscillations. If X_{t-1} and X_{t-2} have positive values, the current value X_t will be negative and vice versa. Due to the small size of the coefficients, the oscillations are relatively weak and less pronounced, but still observable. Stationarity The process remains stationary as long as the roots of the equation (based on the AR coefficients) lie outside the unit circle. In this case, the AR(2) model with $\phi_1=-0.1$ and $\phi_2=-0.3$ will be stationary since the magnitude of the coefficients is less than 1. White Noise Contribution The noise term ϵ_t is the reason for random variation in the series. This contribute to the stochastic nature of the process. Even with the negative autoregressive terms, random noise still largely influences the process. Decay of Influence Given that the coefficients are relatively small, the impact of past values on the current value decays over time, leading to weak persistence. This means that shocks or deviations from the mean will not last long, and the series will quickly revert back to the mean.

4. AR(3) model with complex cycle

```
set.seed(123)
ar3_complex <- arima.sim(model = list(ar = c(0.6, 0.3, -0.2)), n = 200)
plot_data <- data.frame(Time = 1:200, AR_Complex = ar3_complex)
ggplot(plot_data, aes(x = Time, y = AR_Complex)) +
  geom_line(color = "green") +
  labs(title = "AR(3) Process with Complex Cycles", y = "Value")
```

Don't know how to automatically pick scale for object of type <ts>. Defaulting to continuous.

AR(3) Process with Complex Cycles



The Behavior of the Model Positive and Negative Coefficients The AR(3) model has a combination of both positive and negative coefficients. $\phi_1=0.6$: The current value X_t has a positive dependence on the previous value X_{t-1} , meaning that if X_{t-1} is large, X_t will likely also be large. $\phi_2=0.3$: The current value X_t has a moderate positive dependence on X_{t-2} but is weaker than X_{t-1} . $\phi_3=-0.2$: The current value X_t has a negative dependence on X_{t-3} , meaning that if X_{t-3} is large, X_t will likely be smaller. Oscillatory Behavior The process exhibits oscillations when it has positive and negative coefficients. Positive coefficients contribute to values moving in the same direction, while negative coefficients introduce reversals. This can lead to a complex, oscillatory pattern, where the influence of earlier terms (both positive and negative) generates alternating cycles. The size of the coefficients determines the amplitude of these oscillations. Large positive coefficients (like $\phi_1=0.6$) cause the series to follow the previous term closely, while the negative coefficient $\phi_3=-0.2$ creates some counteracting force. Stationarity As long as the roots of the characteristic equation (derived from the AR coefficients) are outside the unit circle, the process is stationary. Given that the coefficients here are moderate in size, this process is stationary, meaning that the mean and variance will remain constant over time. Impact of White Noise The white noise term ϵ_t adds randomness to the model, making the series unpredictable and variable. While the autoregressive terms provide structure, the random noise plays a major role in shaping the series, especially when the influence of past values is not very strong. Persistence of Shocks Because of the combination of positive and negative coefficients, shocks (deviations from the mean) may persist longer than in models with only one or two coefficients, as multiple lags influence the series. However, the persistence of shocks is still limited by the relatively small size of the negative coefficient ($\phi_3=-0.2$).

Discussion

This project is a demonstration of the characteristics and behavior of autoregressive (AR) models through simulations and visualizations. The insights are relayed to AR(1), AR(2), and AR(3) models under varying parameters. For AR(1) models, positive coefficients, such as $\phi = 0.5$, result in moderate positive correlations with values decaying toward the mean, while negative coefficients, such as $\phi = -0.5$, introduce oscillatory behavior with alternating signs. Both cases maintain stationarity as long as $|\phi| < 1$. AR(2) models exhibit more complex dynamics due to including a second lag term. Oscillatory patterns emerge when one coefficient is negative, as seen with $\phi_1 = 1.4$ and $\phi_2 = -0.9$, while positive coefficients produce dampened oscillations. In AR(3) models, the interplay of three lag terms creates intricate cyclical behavior, with positive and negative coefficients driving alternating cycles and persistence influenced by the magnitude of these coefficients. The model white noise plays a crucial role in introducing randomness and variability, while the AR terms define structure. We have introduced an interactive Shiny app, which allows users to manipulate AR parameters and observe their effects in real-time. In future work, it incorporatesis possible to real-world datasets and expands the analysis to mixed ARIMA models or higher-order processes. This study provides a comprehensive and engaging exploration of AR models, highlighting their versatility in time series analysis and real-world applications.