Project Proposal

Datawaves - Libin N George, Ajeet Singh, Saharsh Bhave, 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.

Generic AR Model Equation

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

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

where:

- X_t is the value of the time series at time t,
- $\phi_1, \phi_2, \ldots, \phi_p$ 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.

Data Generation

To understand the behavior of AR (AutoRegressive) models under different configurations, we can simulate time series data using the arima.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 will use it to generate data for AR models with varying parameters.

Data Analysis Plan

We will consider the following variations in AutoRegressive models, We will generate data for these models and create visualizations from the data.

1. AR(1) - Random Walk AR(1) - Having moderate positive correlation, decaying toward the mean over time.
2. AR(1) - Data will oscillate, alternating signs and showing a negative correlation
3. AR(2) - Data with oscillatory behavior
4. AR(2) - Data without oscillatory behavior
5. AR(3) - Data with Complex behavior

We will create a line graph and associated Autocorrelation graph for the above models to visualize the statistical information and how the data behaves. We will also create a shiny app that enables the user to change parameters and plot the line graph.

Final Repository Organization

- **data/**: Contains generated datasets and relevant metadata. It includes README, which details dataset parameters and generation processes.
- **plots/**: Includes static ggplot2 visualizations for AR models (AR(1), AR(2), and AR(3)) with various parameter combinations. README explains each plot and parameter choices.

- **shiny_app/**: Holds the Shiny app files for interactive plotting of AR models. README provides instructions for using the app and explains the interactive features.
- **docs/**: Project report, analysis write-ups, and weekly task updates. README outlines project goals, methods, and findings.
- **project_proposal.qmd:** Project proposal file in qmd format.
- **project_proposal.pdf**: Project proposal auto generated from project_proposal.qmd

Each folder will contain a README.md file explaining its contents and purpose in the project workflow.