Exercise

Simulate a different random walk than the one we have worked with in this chapter. You can simply change the seed and get new values:

3.5.1 Simulate and forecast a random walk

1 Generate a random walk of 500 timesteps. Feel free to choose an initial value different from 0. Also, make sure you change the seed by passing a different integer to np.random.seed().

2 Plot your simulated random walk.

3 Test for stationarity.

4 Apply a first-order difference.

5 Test for stationarity.

6 Split your simulated random walk into a train set containing the first 400 timesteps. The remaining 100 timesteps will be your test set.

7 Apply different naive forecasting methods and measure the MSE. Which method yields the lowest MSE?

8 Plot your forecasts.

9 Forecast the next timestep over the test set and measure the MSE. Did it decrease?

10 Plot your forecasts.

3.5.2 Forecast the daily closing price of GOOGL

Using the GOOGL dataset that we worked with in this chapter, apply the forecasting techniques we've discussed and measure their performance:

1 Keep the last 5 days of data as a test set. The rest will be the train set.

2 Forecast the last 5 days of the closing price using naive forecasting methods and measure the MSE. Which method is the best?

3 Plot your forecasts.

4 Forecast the next timestep over the test set and measure the MSE. Did it decrease?

5 Plot your forecasts.

3.5.3 Forecast the daily closing price of a stock of your choice

The historical daily closing price of many stocks is available for free on finance.yahoo.com. Select a stock ticker of your choice, and download its historical daily closing price for 1 year:

Summary

- A random walk is a process where the first difference is stationary and not autocorrelated.
- We cannot use statistical or deep learning techniques on a random walk, since it moves at random in the future. Therefore, we must use naive forecasts.
- A stationary time series is one whose statistical properties (mean, variance, autocorrelation) do not change over time.
- The augmented Dickey-Fuller (ADF) test is used to assess stationarity by testing for unit roots.
- The null hypothesis of the ADF test is that there is a unit root in the series. If the ADF statistic is a large negative value and the p-value is less than 0.05, the null hypothesis is rejected, and the series is stationary.
- Transformations are used to make a series stationary. Differencing can stabilize the trend and seasonality, while logarithms stabilize the variance.
- Autocorrelation measures the correlation between a variable and itself at a previous timestep (lag). The autocorrelation function (ACF) shows how the autocorrelation changes as a function of the lag.
- Ideally, we will forecast a random walk in the short term or the next timestep. That way, we do not allow for random numbers to accumulate, which will degrade the quality of our forecasts in the long term.