Computation of Autocorrelation Function using Data Set

Manjot Kaur

Department of Computer Engineering &Technology
Guru Nanak Dev University
Amritsar, Punjab, India
manjotman123@gmail.com

Jaspreet Singh

Department of Computer Engineering &Technology
Guru Nanak Dev University
Amritsar, Punjab, India
erjaspreetsingh123@gmail.com

Abstract– **This paper focuses on how to calculate the autocorrelation function by using a particular dataset. Autocorrelation is function used to find that data is correlated with itself or not. The computation of autocorrelation function is followed by various steps which is discussed in this paper. This paper mainly focuses on how data is correlated with itself using a company's dataset. With the help of the scatter plot the relationship between the attributes of a company's dataset have been plotted. This paper discusses how to make a plot and analyze the final result of ACF, by using a linear equation whose value lies between the range of (+1 to -1).**

Keywords: correlation, autocorrelation function, scatter plot, correlogram etc.

I.   INTRODUCTION

 The correlation is a statistical relationship involve the dependence or in other words, how two variables are having a linear relationship with each other[1]. For example, the correlation between the demand and the price of the product. The value of correlation can be vary from +1 to -1.The correlation can be classified as positive and negative correlation. The positive correlation exists where the high values of one variable are associated with the high value of the other variable. The high degree of positive correlation is achieved if value is close to +1. The negative correlation exists when the high values of one variable associated with the low values of other values. If the value of correlation is close to -1,so it means high degree of negative correlation. And if the value close to 0, so it means poor correlation or no correlation of either kind. Correlation doesn't prove or disprove the cause and effect relationship between the variables.
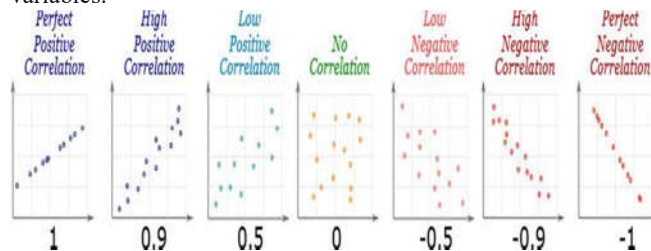

Fig 1: Different Correlation Value charts

Correlation can have a value: **a**. +1 is a perfect positive correlation. **b**. 0 is no correlation (the values don't seem linked at all). **c**. -1 is a perfect negative correlation. **Autocorrelation Function**: It is also known as serial correlation, because it is the correlation of signal with a delayed copy of itself as a function of delay. In another words, ACF refers to a function by which is used to find the

correlation in data collected repeatedly over time or used to find that data is correlated with itself or not. For example, suppose a person analyzing stock market data. The price of a stock today is closely related to the price yesterday and the day before and so on. So here autocorrelation function is used. The autocorrelation function is one of the tool which is used to find the patterns in the data.

ACF is one of the advanced used tool in time series analysis[2]. It is also used to determine the stationary and seasonality in time series analysis. In time series analysis the term **"stationary"** refers to whether the series is going anywhere overtime. The stationary series have persistent value over time. The term **"seasonality"** refers to predictable and recurring trends and patterns over a period of time, normally a year. As an example, following are some of the possible ACF values for a series with discrete time period: The notation is **ACF(n=number of time periods between points)=correlation between points separated by n time periods**.

- ACF(0)=1 (all data are perfectly correlated with themselves)
- ACF(1)=.9 (the correlation between a point and the next point is 0.9)
- ACF(2)=.4 (the correlation between a point and a point two time steps ahead is 0.4)...etc.
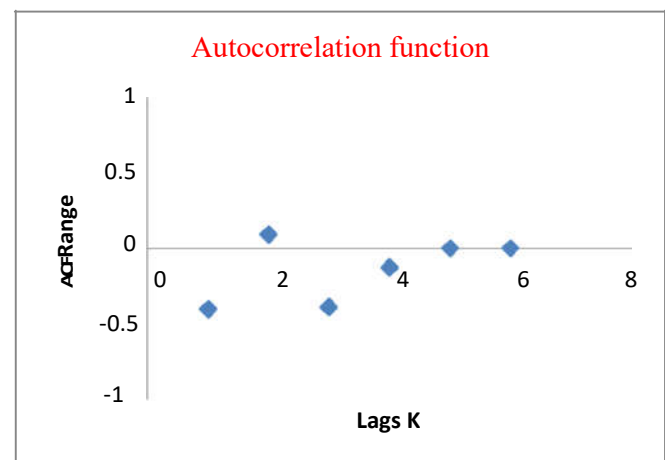

Fig 2: Scatter plot for showing ACF

This plot shows that the attributes count(t) and count(t-1) are perfectly negatively autocorrelated with each other because the linear relationship exists between them, which is clearly understandable by the linear equation whose value is in the range of +1 to -1.

II. TYPES OF AUTOCORRELATION

A. Spatial Autocorrelation

Spatial autocorrelation is the correlation among values of a single variable across a two-dimensional surface that are locationally referenced or tied together by an underlying spatial structure, introducing a violation of the independent observations assumption of classical statistics[3]. According to Tobler's first law of geography,"All things are related, but nearby things are more related than distant things".
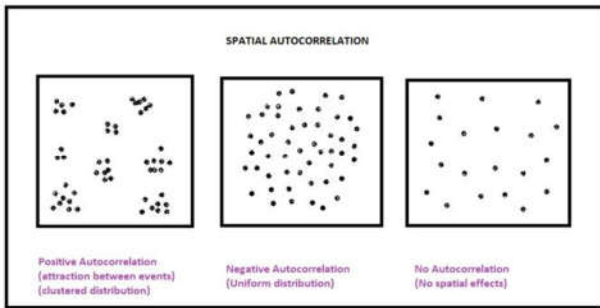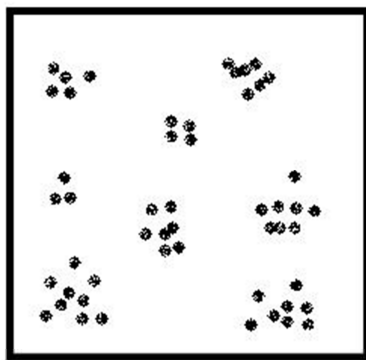

Fig 3: Spatial Autocorrelation

B. Positive Autocorrelation

It refers to the autocorrelation between the events which have similarities and are very near to each other and form a cluster or whose value is very close to positive 1 and not in negative range[4].

Positive Autocorrelation (attraction between events) (clustered distribution)

Fig 4: Positive Autocorrelation

C. Negative Autocorrelation

It refers to autocorrelation whose events are near but dissimilar or whose value is in the range of -0.1 to -1. The Fig 5 shows the negative autocorrelation[5] using dissimilar pattern in data.
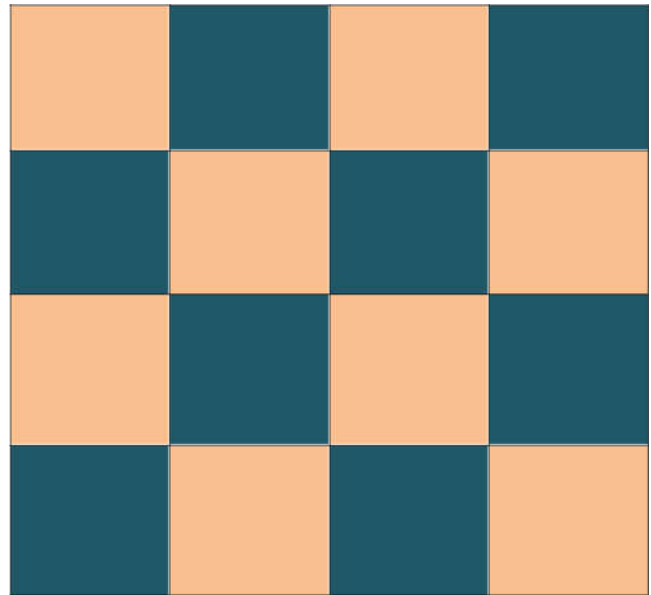

Fig 5: Negative Autocorrelation

III. CALCULATING ACF

To calculate the ACF, the following steps are as:

- Select dataset.
- Choose attribute list.
- Calculate variance of the attributes.
- Calculate Covariance values.
- Calculate the final ACF by dividing variance by covariance for every kth lag.
- At the end to show results pictorially , the scatter plot will be used.

A. Steps to calculate ACF

The autocorrelation function[6] can be calculated by follow the following figure :
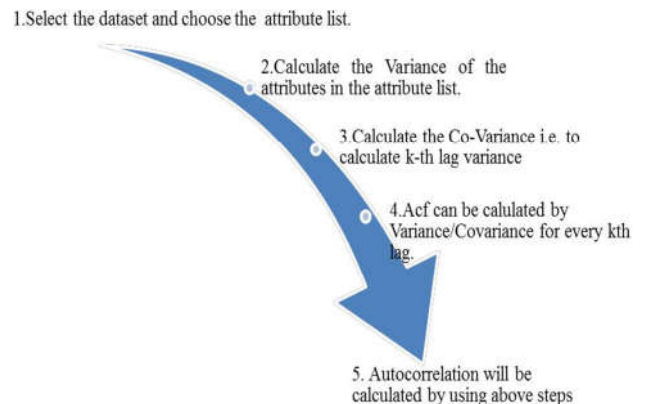

1.Select the dataset and choose the attribute list.

2.Calculate the Variance of the attributes in the attribute list.

3.Calculate the Co-Variance i.e. to calculate k-th lag variance

4.Acf can be caluated by Variance/Covariance for every kth lag.

5. Autocorrelation will be calculated by using above steps

Fig 6: Steps to calculate the ACF

B. Calculating ACF using dataset

Take a dataset of **TESLA company** including main attribute **"Return"**, whose ACF will be calculated using MS-Excel in this section.

Fig 8: Dataset for TESLA Company

Implementation of steps to calculate the ACF for TESLA Company are as:

1. Calculate variance of the column "RETURN".
   by using VAR function in excel ,like
   =VAR(H6:H27) => 0.000108043

2. Calculate the COVARIANCE i.e. to calculate kth lag variance for the column "RETURN" using COVAR function in excel, like
   =COVAR(H7:H27,H6:H26) for lag k=1 …….k=22 in above dataset.

3. Calculate ACF for an individual k th lag by using formula , VARIANCE/COVARIANCE for a particular lag, like
   ACF(k=1) = VAR/COVAR=> -5.98188633---------------ACF(k=22)….. .as shown in figure 8.

4. After calculating the ACF of all k lags then make a scatter plot for the same and check whether the attributes of the "Retrun" column is autocorrelated or not.

C. Scatter plot

A scatterplot is a useful summary of a set of bivariate data (two variables), usually drawn before working out a linear correlation coefficient or fitting a regression line[7]. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model.
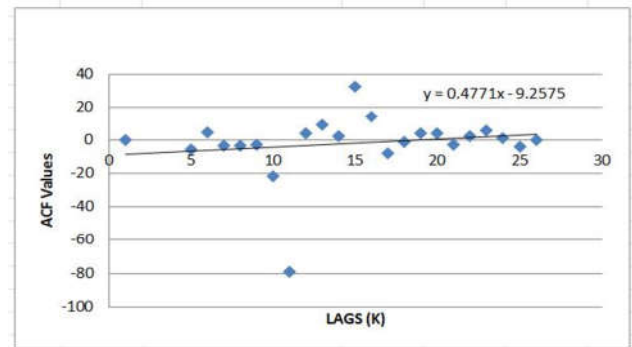

Fig 9: Scatter plot for TESLA company dataset

This plot has a linear equation which indicates that the attributes in this plot are not perfectly autocorrelated with each other. This plot is come under the **No autocorrelation** category.

IV. CORRELOGRAM OR AUTOCORRELATION PLOT

Correlogram is used to analyse the data , checking the randomness if present in the given dataset. In time series examination , the correlogram is also known as autocorrelation plot[8]. The cross correlation is called the cross-correlogram if it is used[9],[10]. "The correlogram is an excellent way of checking for such randomness"[11].
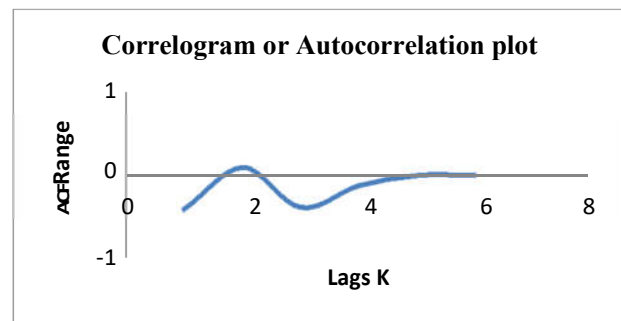
Fig 10: Correlogram or Autocorrelation plot.

V. PROPERTIES OF ACF

The following are properties of autocorrelation function:

- Autocorrelation is symmetrical, i.e. $R(i)= -R(i)$.

- The autocorrelation of a periodic signal, is itself, periodic with the same period.

- The autocorrelation of the sum of two completely uncorrelated functions is the sum of the autocorrelations of each function separately.

- The continuous autocorrelation function reaches its peak at the origin, where it takes a real value.

- It is bounded, i.e. $-1 <= i <= 1$.

- The symmetric autocorrelation function has a real symmetric transform for a real valued function[12].

VI. AUTOCORRELATION MATRIX

A. Introduction

"The **autocorrelation matrix** is used in various digital signal processing algorithms. It consists of elements of the discrete autocorrelation function"[13].

B. Representation

The autocorrelation matrix [14] consists the elements of discrete autocorrelation which are arranged in the following manner:

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^H] = \begin{bmatrix} R_{xx}(0) & R_{xx}^*(1) & R_{xx}^*(2) & \cdots & R_{xx}^*(N-1) \\ R_{xx}(1) & R_{xx}(0) & R_{xx}^*(1) & \cdots & R_{xx}^*(N-2) \\ R_{xx}(2) & R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}^*(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{xx}(N-1) & R_{xx}(N-2) & R_{xx}(N-3) & \cdots & R_{xx}(0) \end{bmatrix}$$

Fig 11: Autocorrelation matrix

VII. APPLICATIONS OF ACF

The autocorrelation function has wide variety of applications which are as[15],[16],[17],[18],[19],[20]:

- Autocorrelation function is used to extract radar signals to improve sensitivity.

- Measuring fast signal decay

- Laser microscopy

- Measuring pulse broadening and distortion

- Fluorescence spectroscopy

- Measuring musical beats (estimating pitch)

- Scanning pulsar frequencies

- Analyzing spatial patterns

- It is used in X-ray diffraction data analysis

- Security system design (pattern recognition utilized)

- Water traffic monitoring

- In signal processing, autocorrelation can give information about repeating events like musical beats.

- In optics, normalized autocorrelations and cross-correlations give the degree of coherence of an electromagnetic field.

- In astrophysics, autocorrelation is used to study and characterize the spatial distribution of galaxies.

- It is used in time series analysis.

- It is used for checking randomness in dataset.

- It is used in stock returns .

VIII. CONCLUSION

This paper focuses on how to calculate the autocorrelation function by using a dataset attributes in MS-Excel by using excel inbuilt functions var and covar. Autocorrelation is a mathematical representation of the degree of similarity between a given time series and a lagged version of itself over successive time intervals. It is the same as calculating the correlation between two different time series, except that the same time series is used twice: once in its original form and once lagged one or more time periods. For analyzing the results, scatter plot and correlogram were discussed. Autocorrelation function plays a major role in stock return for equity analysis and the relationship between attributes of a particular dataset.

REFERENCES

[1]   I. Management, "Study on the Correlations of Chinese Railway Passenger Network Based on Complex Network Theory" Dewei Weight Degree-degree Correlation, pp. 376–379, 2012.

[2]   B. P. Salmon, W. Kleynhans, J. C. Olivier, and C. P. Schwegmann, "Transforming the autocorrelation function of a time series to detect land cover change" in School of Engineering and ICT , University of Tasmania , Australia Remote Sensing Research Unit , Meraka Institute , CSIR , South Africa Department of Electrical , Electronic and Compute," pp. 5181–5184, 2016.

[3]   W. Di and C. Zhongxin, "Spatial autocorrelation and its influencing factors of the sampling units in a spatial sampling scheme for crop acreage estimation." in 2017 6th International Conference on Agro-Geoinformatics,2017.

[4]   D. A. Griffith, "Positive Spatial Autocorrelation , Mixture Distributions, and Geospatial Data Histograms." in IEEE, 2011.

[5]   A. Tsuneda and Y. Miyazaki, "Performance Evaluation of Spreading Sequences with Negative Auto-correlation Based on Chaos Theory and Gold Sequences,"IN IEEE, vol. 0, pp. 169–172, 2009.

[6]   I. Radujkov and V. Senk, " New Direction to Function Estimation," in IEEE no. September, pp. 1–4, 2001.

[7]   R. Nizamee, "Visualizing the web search results with web search visualization using scatter plot," in

IEEE, pp. 5–10, 2009.

[8] W. Rasheed, Y. An, S. Pan, I. Jeong, and J. Park, "Image Retrieval using Maximum Frequency of Local Histogram based Color Correlogram," in IEEE, pp. 322–326, 2008.

[9] Y. Zhao and C. He, "Improving change vector analysis in multi-temporal space to detect land cover changes by using cross-correlogram spectral matching algorithm," in IEEE, pp. 174–177, 2011.

[10] J. Park, S. Han, and Y. An, "Heuristic Features for Color Correlogram for Image Retrieval" in IEEE, pp. 9–13, 2008.

[11] "Correlogram for undersampled data : bias and variance analysis" in IEEE, pp. 3513–3516, 2012.

[12] J. E. Rice, "Properties of Autocorrelation Coefficients," in IEEE, pp. 577–580, 2003.

[13] W. Wang, S. Member, T. Adali, and S. Member, "Eigenanalysis of Autocorrelation Matrices in the Presence of Noncentral and Signal-Dependent Noise," in IEEE, vol. 12, no. 2, pp. 85–88, 2005.

[14] A. Naha, A. K. Samanta, A. Routray, and A. K. Deb, "Determining Autocorrelation Matrix Size and Sampling Frequency for MUSIC Algorithm," IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1016–1020, 2015.

[15] B. Wei, D. Liu, and X. Wang, "The General Autocorrelation and Its Application," in IEEE, 2002.

[16] T. Engineering, "Estimation and Tracking of Pitch for Noisy Speech Signals using EMD based Autocorrelation Function Algorithm," in IEEE, pp. 0–4, 2017.

[17] A. Contin and S. Pastore, "An Algorithm-Based on Auto-Correlation Function Evaluation for the Separation of Partial Discharge Signals," in IEEE no. September, pp. 19–22, 2004.

[18] N. D. Diamantidest, "Multipurpose Electronic Switch for Analog Computer Simulation and Autocorrelation Applications," in IEEE Transactions on Electronic Computers,1956.

[19] M. S. Hromalik, K. Green, H. Philipp, M. W. Tate, and S. M. Gruner, "Asynchronous and synchronous implementations of the autocorrelation function for the FPGA X-ray pixel array detector," 2012 18th IEEE-NPSS Real Time Conference, RT 2012, 2012.

[20] S. Jianquan and M. Xiaoxian, "Warrant pricing model with autocorrelated underlying stock returns," 2009 International Conference on Business Intelligence and Financial Engineering, BIFE 2009, pp. 317–320, 2009.