Gaussian Auto Regressive Model for Scat-Sat Data Set

Polaiha B¹, Sunitha M², Rajesh Anand B³, Khadar Babu SK⁴

¹Assistant Professor, Mallareddy University, Hyderabad
²Statisticum cum Assistant Professor, Department of community Medicine, SVIMS, Tirupathi
³Department of Mathematics, Sri Venkateswara University, Tirupathi
⁴Associate Professor, VIT University, Vellore, TN, India
Corresponding Author Mail ID: khadar.babu36[at]gmail.com

Abstract: The Gaussian distribution plays a very important role in the statistical theory as well as methods. In this paper introduces the identification of the phenological stages of the rice crop located at Godavari region, Andhra Pradesh, India. The data taken from satellite images using ARCGIS platform and got the backscatter values. At present we make the following objectives for research work carried out. The major objective is to identify the phenological stages using graphical approach and to apply Gaussian probability approach for generation of the values using random Gaussian generator. Next step of the process is to fit an Auto-regressive model to predict and forecasting of the future observations and also obtain missing observations in the bulky data set. While fitting a Gaussian probability approach applied the method of moments and the maximum likelihood method of estimation for parameter estimation.

Keywords: Probability distribution, frequency analysis, Gaussian distribution, Auto Regressive model etc

1. Introduction

The Gaussian distribution has a significant theoretical and practical impact on probability theory and mathematical statistics. The distribution can be found in a variety of sectors, such as high technology, natural phenomena, and industrial production. In general, a Gaussian distribution is an index that is influenced by numerous explanatory factors, each of which has a negligible impact. For example, the outcomes of quality indexes include tool size, fiber strength, and single feature groups like class vital capacity, plant strength of rice stem diameter in a particular place, recorded data on maximum air temperature, and average rainfall. The gaussian probability distribution's laws and regulations are being followed by all of the aforementioned indices. The phenological stages of the rice crop in the Godavari region of Andhra Pradesh, India, are identified in this research. The backscatter values were obtained from the data obtained from satellite photos using the ARCGIS platform. The main goal is to use a graphical approach to determine the phenological stages. The next step is to use the Gaussian probability technique and a random Gaussian generator to produce the values. The procedure then moves on to fitting an auto-regressive model to anticipate and predict future observations, as well as extract missing observations from a large amount of data. For time series data sets, the autoregressive model is a generator that fits the data the best.

2. Review of Literature

T. Mayooran and A laheetharan (2014) is to identify the best fit probability distribution of annual maximum rainfall flow time series in Colombo district for each period of study. In this paper for estimation of parameters they applied maximum likelihood method of estimation and goodness of fit tests were carried out In order to find the best fitting probability distribution among 45 probability distributions for annual rainfall flow time series for four seasons separately but the author cannot compare the parameters calculated by different methods.

Dr. Rafa H ALL-Suhili and Dr.Reza Khan Bilvardi (2014) studied about different frequency distribution models were fitted to the monthly rainfall data in saloimonia region north Iraq.

Haoge Liu and Jianhe (2016) studied about design of practical models of statistical events to optimize the Gaussian probability density function and to provide a significant method in statistics.

M Sunita et al (2019) studied about novel research methodology to analyze the parameters of the big data sets such as rainfall flow time series, wind speed time series, hydrological datasets under fuzzy randomized approach.

Mahesh Palakuru et.al 2019, focused on the phenological parameters estimating using S map soil moisture active passive (S map), MODIS NDVI and Scat Sat – 1 scatrometer data and they adopted Gaussian distribution and two parameter Logistic distribution model for analysis.

3. Methods and Discussions

Let us consider the data, Zi, i=1,2, 3,…, n observations in a data set, for these values, by using the method of moments, to estimate the parameters of the probability distributions.

The Gaussian distribution plays a very important role in the statistical theory as well as methods. The great mathematician like Gauss, Laplace, Legendre and others are associated with the discovery and use of the distribution of errors of measurement.
The method of moments was discovered by Karl Pearson in 1894. It is a method of estimation of population parameters such as mean, variance etc. (which need not be moments), by equating sample moments with unobservable population moments and then solving those equations for the quantities to be estimated. The method of moments is special case when we need to estimate some known function of finite number of unknown moments. Suppose \( f(x; \lambda_1, \lambda_2, \ldots, \lambda_p) \) be the density function of the parent population with \( p \) parameters \((\lambda_1, \lambda_2, \ldots, \lambda_p)\). Let \( \mu_i \) be the \( s \)-th moment of a random variable about origin and is given by 

\[
\mu_i = \int_{-\infty}^{\infty} x^i f(x; \lambda_1, \lambda_2, \ldots, \lambda_p) \, dx
\]

In general \((\mu_1, \mu_2, \ldots, \mu_p)\) will be the functions of parameters \((\lambda_1, \lambda_2, \ldots, \lambda_p)\).

Let \((x_i; i=1, 2, \ldots, n)\) be a random sample of size \(n\) from the given population. The method of moments consists in solving the \(p\)-equations (i) for \((\lambda_1, \lambda_2, \ldots, \lambda_p)\) in terms of \((\mu_1, \mu_2, \ldots, \mu_p)\). Then replacing these moments \((\mu_i; i=1, 2, 3, \ldots, p)\) by the sample moments \(\hat{\lambda}_i = \lambda_i(m_1, m_2, \ldots, m_p) ; i = 1, 2, \ldots, p\) where \(m_i\) is the \(i\)-th moment about origin in the sample.

Then by the method of moments \((\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_p)\) are the estimators respectively.

Autoregressive models can be effectively coupled with moving average models to form a general and useful class of time series models called autoregressive moving average models. However, they can only be used when the data are stationary. This differencing of the data series. For all three stages, they are identified using Autoregressive model building using Excel sheet. The data analysis clearly given in the following tables and made different graphical methodology is given.

### 3.1 Rice Crop Phenology Stage II Statistical Analysis

It is the major stage in rice crop phenology called heading. In this stage seed yield buildup is constructed in this stage. Here needs continuous observation. For the data, calculated all normally distributed simulated values for the corresponding observations. Build a model for data at heading stage and analysis is given below.

<table>
<thead>
<tr>
<th>Table 3.1: Auto regressive model building for stage – II analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image_url" alt="Image" /></td>
</tr>
</tbody>
</table>

**Volume 12 Issue 11, November 2023**
International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Mean (X_t) = -9.9136 \quad \sigma_1 = 0.13467526
\sigma_e^2 = 0.23179562 \quad \sigma_e = 0.481451576

Therefore AR (1) equation for the problem can be written as

\begin{equation}
X_t - (-9.9136) = 0.13467526(X_{t-1} + 9.9136) + 0.481451576
\end{equation}

\begin{equation}
X_t = 0.13467526(X_{t-1} + 9.9136) + 0.481451576 - 9.9136
\end{equation}

If X_{t-1} = -9.8480 then the predicted observation is

X_{16} = 0.13467526(-9.8480 + 9.9136) + 0.481451576 - 9.9136

X_{16} = -9.42331

4. Conclusions

Gaussian probability distribution using Auto Regressive model is perfectly suitable for ScatSat datasets. We suggest that Gaussian probability distribution is exactly fitted for ScatSat data and also data simulation uses to build a auto regressive model for prediction and forecasting of the backscatter values of the ScatSat data. In this chapter we especially focus on regression model approach is also perfectly a novel application to generate, predict and forecasting of the vegetations at identified crop phenological stages in the present study area.

ScatSat datasets are the data given by the scatterometer fixed in satellites and the data called scatterometer data and it gives the backscatter values of the rice crop vegetations in identified study area. Finally, we conclude that the Gaussian probability random generated simulated datasets are perfectly applicable to the Scatterometer data.
References


