**Interpretation of Statistics in Astronomy**

**Introduction**

In our article, we're going to discuss some applications of statistics in astronomy and their interpretations. Many problems in astronomy had led to the development of many statistical methods, from traditional and classical methods such as least squares estimation to modern methods like nested sampling.

However, with the development of technology, especially in the field of computer science, it’s now possible for astronomers to collect a large amount of data from the universe. Is there any kind of extraterrestrial life out there? When will the universe come to an end and how? How many stars in each galaxy are similar to the Sun? These are the most common questions, which can only be answered after collecting, analyzing and interpreting the astronomical data.

This is when the need arises for a new field of study that is known as Astro statistics. It is a prominent field of statistics which is used as a tool to analyze and interpret the big databases relating to the universe.

So, the challenge of Astro statistics is to build statistical tools to refine and interpret massive amount of data from space and with the help of that to generate relevant and useful information that answers the most intriguing and big questions of astronomy.

**History of statistics in astronomy**

Astronomy is the oldest observational field of science. Now, let us see, how the application of Statistics, in Astronomy gradually increased day by day.

* The Greek natural philosopher, Hipparchus, made one of the first applications in finding scatter of the length of a year in Babylonian measurements, defined as the time between solstices, he took the middle of the range, rather than the mean or median, to estimate the value. Today, this method is known as the ‘Midrange estimator of location’.
* In the sixteenth century, Tycho Brahe and Galileo Galilei promoted the utility of the mean of discrepant observations to increase precision. Galileo also gave his insights about the properties of errors in nonmathematical language, in his ‘Dialogue on the Two Great World Views, Ptolemaic and Copernican’. Later these properties were incorporated by Gauss into his quantitative theory of errors.

In 18th century, some Astronomers tried to tackle numerous inaccurate astronomical data, combine those observations and estimate the physical quantities through celestial mechanics more earnestly.

* In 1767, British astronomer John Michell applied a significance test, based on the uniform distribution, to show that the ‘Pleiades (The Seven Sister Star Cluster)’ is a physical grouping of stars. Though there were some technical errors in his techniques.
* Bernoulli and Lambert had laid the foundations of the concept of maximum likelihood which was later developed more thoroughly by Fisher in the early 20thcentury.
* Through a complex and difficult course of reasoning, Laplace proved that method of least squares was the most convenient method for finding parameters in orbital models from astronomical observations.

* In 1733, the mathematician Abraham De Moivre used the normal distribution to approximate the distribution of the number of heads resulting from many tosses of a fair coin. This approximation is known as ‘Central Limit Theorem’. Later improvements were developed by Sim’eon Denis Poisson Friedrich Bessel.

In the 20th century, we witnessed a gap between statistics and astronomy. None of the major works regarding Astronomy, like, discovery of the interstellar medium, analysis of the geometry of the Galaxy, discovery of extragalactic nebulae, did not involve statistical theory or application. The modern field of Astro-statistics grew suddenly and rapidly starting in the late 1990s.

* In the 1920s, Fisher formulated the method of maximum likelihood estimation. They were instrumental in discovering galaxy streaming towards the “Great Attractor” (a gravitational anomaly in intergalactic space and the apparent central gravitational point of the “Laniakea Supercluster”) and in computing the ‘galaxy luminosity function’ from flux-limited surveys.
* In the year 1970, the nonparametric Kolmogorov–Smirnov statistic was used for the first time for ‘two-sample’ and ‘goodness-of-fit’ tests.
* Since 1990’s, Bayesian classifiers for discriminating stars and galaxies are used to construct large, automated sky survey catalogs.

**Use of Regression Analysis in Astronomy**

In Astronomy, regression analysis is one of the most frequently used statistical techniques. Most astronomical data analyses feature intrinsic scatter regarding the regression line.

In the year 2007, scientist Kelly described a **Bayesian Method (MLINMIX)** based on the likelihood function of the measured data (Later, in 2015, Scientist Mantz extended the MLINMIX algorithm to the case of multiple response variables). The method will account for measurement errors, intrinsic scatter, multiple independent variables, non-detections, and choice effects within the variable quantity.

Later, in the year 2014, scientist Maughan suggested a model to constrain simultaneously the form and evolution of the scaling relations. The method distinguishes between measured values, intrinsic scattered values, and model values and can constrain the intrinsic scatter and its covariance.

**Use of Linear Scaling of Bayesian Method in Astronomy:**

Case 1: Case of **Scattered Quantities.**

Most of the scaling relations in astronomy are time evolving different power-laws. This straight-forward schematics is supported by different types of observations, theoretical considerations, and numerical simulations.

The general form of the relation between two properties, e.g., the observable O and the mass M, is

**O**  **Mβ Fγz**----------------------(1)

Where, β is the slope and also the red shift evolution within the median scaling relation, which is accounted by the factor **Fz**. According to the context, the redshift factor Fz may be the factor (1 + z). In logarithmic variables, the scaling relation is linear, and the scatter is Gaussian,

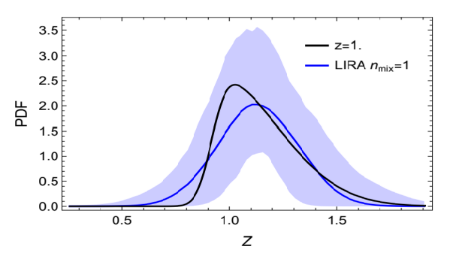
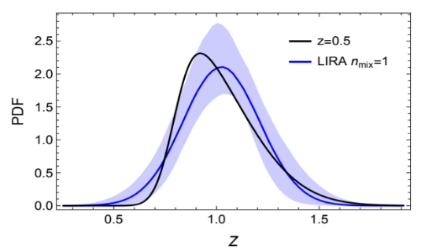
**log O = α + β log M + γ log Fz ---------------------**(2)

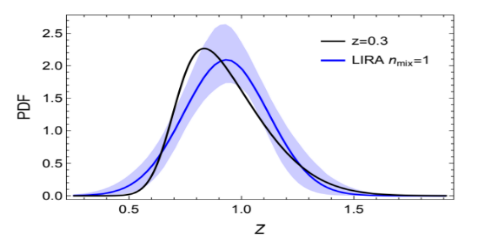
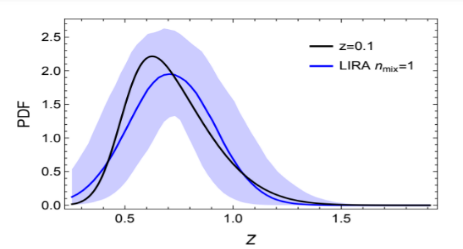
In the usual framework, the time (T=logFz) evolution does not depend on the mass scale and only affects the normalization. However, the interplay between different physical processes that can be more or less effective at different times and can make the slope time dependent, β.

Assuming that the evolution of the slope with redshift is linear in T, Equation (2) may be generalized as,

**Y = α + β X + γ T + δ X T** ----------------------(3)

Where, X = log M, and Y = log O. The time variable T is deterministic, not affected by measurement of errors and the variable X is random.

**Figure 1:**

The reconstructed intrinsic distribution of the independent variable Z at different redshifts, for, z = 1.0, 0.5, 0.3, 0.1. The black line is the input distribution, the blue line is the median reconstructed relation, the shadowed blue region encloses the 1-σ confidence region for each value of Z. For a total of nsample = 100 data.

Case 2: Case of **Unscattered Quantities.**

Also, the linear relation between two **Unscattered quantities** can be expressed, as,

**YZ = αY| Z + βY| Z Z + γY| Z T + δY| Z Z T**---------------------- (4)

Where, α denotes the standardization, the slope β denotes the dependence with Z, the slope γ denotes the time-evolution of the standardization and δ quantifies the tilt of the slope with time.

**DEPARTURE FROM LINEARITY:**



Different physical processes are effective at different scales, which may cause deviation from linearity. Gravity is the actuation behind formation and evolution of galaxy clusters however at small scales baryonic physics will play a distinguished role. As a result, linearity can break. This can be shaped with a knee in this relation, such that before the breaking scale **Zknee**, the scaling as follows

**YZ = αY| Z, knee + βY| Z, knee Z + γY| Z, knee T+ δY| Z, knee Z T**----------------- (5)

The standardization **αY| Z, knee** and the time evolution **γY| Z, knee** is determined by requiring equality at the transition **Zknee**,

**αY| Z, knee= αY| Z + (βY| Z − βY| Z, knee)** --------------- (6)

**γY| Z, knee = γY| Z** ------------ (7)

The transition between the two regimes is often modeled through a transition perform,

**fknee = 1 / 1 + exp [(Z − Zknee) / lknee]** -------------- (8)

where, the size lknee sets the transition length. The relation over the full range reads

**YZ = αY| Z +βY| Z Z +γY| Z T +δY| Z Z T + (Zknee −Z) fknee(Z) × (βY|Z − βY| Z, knee) + (δY| Z − δY| Z, knee) T ---------- (9)**

Similar physical processes will have an effect on the scatter too, which I model as

**σY| Z (Z, zref) = σY| Z, 0 + (σY| Z, 0, knee − σY| Z, 0) fknee(Z)** ------------(10)

Assuming that the redshift evolution of the scatter is not affected.

**Remark:**

Bayesian linear regression models have involved a large number of parameters. Since all relations in the model are expressed as conditional probabilities, thus, the posterior can be efficiently explored. All of these procedures have their own specifications and strengths that can make them preferable under some given circumstances. It also allows the consistent treatment of time-evolution, intrinsic scatter, and selection effects. Deviations from linearity relations with knees can be accounted for. Thus, the feature of a linear regression model to stay simple and to add complexity if needed is then important.

**Inference in Astronomy**

Statistical inference is an art of forming an idea about an unknown population on the basis of a completely known sample. It is not possible to observe all the units in a population when the population size is very large. Astronomers often work with stars and planets. Now the problem is that there are countably infinite numbers of stars in a galaxy. In this situation we take the help of statistical inference to estimate the desired characteristics of the population. Astronomers usually take a sample from the population and measure the properties of the sample to know the properties of the vast underlying population of similar objects in the Universe.

It comes into the picture when the astronomer:

* Smooths over separate observations to know the underlying continuous development
* Seeks to quantify relationship between determined properties
* Tests to know an observation matches with an assumed astrophysical theory
* Subdivides a sample to compensate for flux limits and no detections
* Investigates the temporal behavior of variable sources
* Infers the evolution of cosmic bodies from studies of objects at totally different stages
* Characterizes and models patterns in wavelength, pictures or space and lots of different things.

Now, in order to know some properties of the population i.e., to find the value of the parameter of interest first we need to find a good estimator of this parameter. The method of moments, least squares (LS) and maximum likelihood estimation (MLE) are vital and normally used procedures for constructing estimates of the parameters.

Suppose θ be an unknown parameter of the distribution of a variable X and T is an estimator for estimating θ on the basis of a random sample (X1, X2, ..., Xn). Now, T is said to be a good estimator if, for all >0, >0, however small, it is possible to find a n0, depending on , , such that,

**P [|T- θ| ] > 1-, whenever nn0.**

An alternative method is to provide an interval within which the parameter may be supposed to lie. This is called interval estimation. The confidence interval of a parameter θ, a statistic derived from a dataset X, is outlined by the range of lower and higher values [l (X), u (X)] that depend upon the variable(s) X outlined such that

**P [l (X) < θ < u (X)] = 1 – α [α=level of significance]**

Where 0 <α< 1 is usually a small value like α = 0.05 or 0.01. That is, if θ is the true parameter, then the coverage probability that the interval [l(X), u(X)] contains θ is at least 1 − α. The quality of confidence intervals is judged using criteria including validity of the coverage probability, optimality (the smallest interval possible for the sample size), and invariance with respect to variable transformations.

**Bayesian Parameter Estimation:**

In the Bayesian approach, we can test our model, in the light of our data and see how our degree of belief in its ‘fairness’ evolves, for any sample size, considering only the data that we did actually observe.

**P (model| data, I) =k \* P (data| model, I) \* P (model| I) [k=proportional constant]**

Bayesian inference relates the probability of model parameters θ to experimental data d, and an hypothesis for the data H, via Bayes theorem P(θ| d, H) = π(θ| H) L(d| θ , H) Z(d| H) . Here, P(θ| d, H) is the posterior probability density of the parameters θ given d and H; L(d| θ, H) is the likelihood of d given θ and H; π(θ| H) is the prior probability of θ; and Z(d| H) is the evidence of d given H.

Bayesian parameter estimation is the workhouse of gravitational-wave astronomy, for instance determining the mass and spins of merging black holes, revealing the neutron star equation to state, and unveiling the population properties of compact binaries. It is the tactic by which gravitational-wave data is employed to infer the sources’ astrophysical properties.

There are mainly two scales that verify the overall computational cost of Bayesian inference. They are (i) The price of evaluating parameterized models of the data, and (ii) the rate of convergence of the sampling algorithms. The typical wall time may be calculated firstly by considering the total CPU time. To the leading order, the CPU time,, of Bayesian inference scales such as the average call-time of the data model <>, multiplied by the total number of calls to the likelihood function N of the stochastic sampling algorithm = N<>. We have a tendency to treat N as an overall standardization which is typically N ∼O (107). Once serial sampling algorithms are used, the CPU time is equal to the wall time. The average call-time <> is powerfully passionate about the complexness of the GW signal models, and perhaps depends on the models for the noise.

**Application:**

For example, suppose a spacecraft is sent to a moon of Saturn and, employing a penetrating probe, detects a liquid ocean deep beneath the surface. Now if the thermometer encompasses a fault, then it can’t felicitate to detect the actual temperature of the liquid. By Bayesian hypothesis testing we are able to determine the temperature of the liquid assuming it water and then assuming it fermentation alcohol.

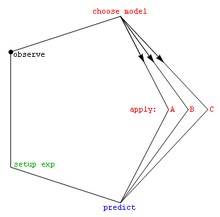
**Prior Distribution:**

Prior distribution is a probability distribution of attainable values for an unknown population characteristic that is developed before one obtains any current data observations about the phenomenon of interest. A posterior probability is the probability that assigns observations to groups for the given data. Samples are accepted subject to the limitation that those drawn on subsequent iterations have a better likelihood than those on previous iterations. The algorithm rule is seeded by drawing a variety of K live points from the prior. These points are arranged from highest to lowest likelihood. The algorithm then proceeds by drawing samples from the prior on each iteration i. The aim is to exchange the live point with the lowest likelihood Lmin, on every iteration.

**Model Selection:**

Model selection is the task of choosing a statistical model from a collection of candidate models, given data. Within the simplest cases, a pre-existing set of data is taken into account. It may be a polynomial also.

A good model selection technique will help to balance goodness of fit with simplicity.



**Figure 2:**

The Scientific Observational Cycle.

**Likelihood Based Model:**

Let us assume that U denote the observed data and let M1, ..., Mk denote the models for U, under consideration.

For every models, Mj, let L(U| θj; Mj) and (θj) = ln f (U| θj; Mj) denote the likelihood and loglikelihood respectively, where θj is a pj-dimensional parameter vector. Here L(U| θj; Mj) denotes the pdf or the p.m.f evaluated at the data U. For most of the cases we have a comparison between two models, M1 and M2. The model M1 is known to be nested in M2 if some elements of the parameter vector θ1 are fixed (and possibly set to zero), i.e., θ2 = (α, γ) and θ1 = (α, γ0), where γ0 is some known fixed constant vector. Comparison between M1 and M2 can then be considered as a classical hypothesis testing problem where the null hypothesis, 𝐻0: γ = γ0. The nested models of this type appear frequently in different astronomical modeling. In the astrophysical modeling, stellar photometry might be modeled as a blackbody (M1) with absorption (M2), the structure of a dwarf elliptical galaxy might be modeled as an isothermal sphere (M1) with a tidal cutoff (M2), or hot plasma might be modeled as an isothermal gas (M1) with nonpolar elemental abundances (M2).

Let, X follows N (μ, σ2) (𝑀1) and Y follows N (0, σ2) (𝑀2), where μ and σ2 both unknown

To test 𝐻0: μ = 0 against 𝐻1: μ ≠ 0

Suppose T is the test statistic and c are the critical point

We reject 𝐻0, if (|T|>c|𝐻0)

Model selection and goodness-of-fit is very much dependent on the sample size, if sample is large, very small discrepancies lead to rejection of the null hypothesis, while for small samples, even large discrepancies might not result in rejection.

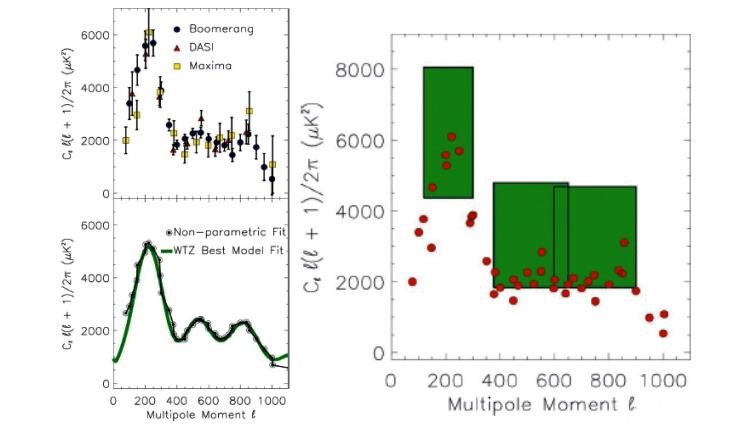
**Nonparametric Statistics**

We have limited knowledge about the planets, stars, galaxies or accretion phenomena which provides very little information about the underlying conditions. In some situations, the mathematical assumptions related to the statistical procedures are often not well explained. As a result, some interesting characteristics or facts about the data may be obfuscated by using a simplistic model. So, astronomers need nonparametric methods because in those cases no assumptions regarding the underlying probability distribution are needed. An example of such a nonparametric method is the Kolmogorov–Smirnov (KS) two-sample test. Modern nonparametric statistical tools may best uncover the underlying mass distribution of galaxies which have complicated tri-axial structures dominated by dark matter without the simplification and physically unrealistic assumptions of the analytical formulae.

Thus, nonparametric approaches to data analysis are very much needed in Astro-statistics. Nonparametric methods can provide us considerable capability to analysis and interpret the astronomical data.

**An application in Astro-statistics:**

Using nonparametric methods, two astrophysicist Robert Nichol, Chris Miller and two statisticians Larry Wasserman, Christopher Genovese have established the three-peak structure of the cosmic microwave background fluctuation spectrum, without using the high precision of parametric modeling. Here, Figure 3, shows the three-peak structure of the cosmic microwave background fluctuation spectrum using nonparametric methods.



**Figure 3:**

The three-peak structure of the cosmic microwave background fluctuation spectrum using nonparametric methods.

**Examples of nonparametric methods:**

Some examples of nonparametric methods are:

1. Nonparametric Density Estimation
2. Tests of hypotheses without parameters
3. Nonparametric Goodness-of-Fit tests

* One-sample Kolmogorov-Smirnov Test
* Two-sample Kolmogorov-Smirnov Test

1. K-sample tests, Correlation coefficients
2. Nonparametric Regression

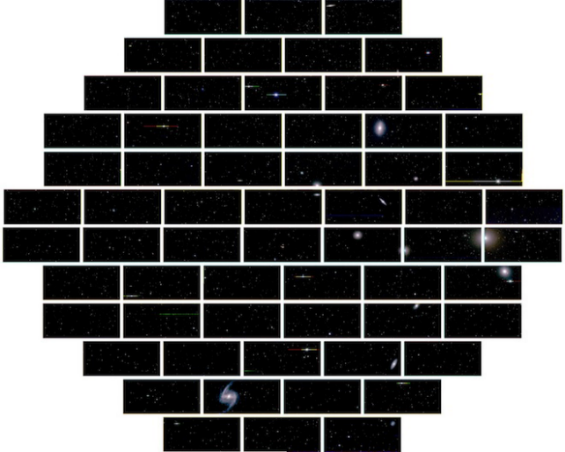
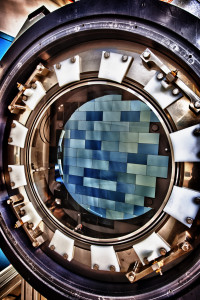
* Kernel Estimation
* k-Nearest Neighbor Estimation
* LOESS Estimation

**TYPES OF DATA USED IN ASTRONOMY**

To make an Astronomical study, with the help of Statistical Methods, we need some data, on the basis of particular field, in Astronomy. For this Astronomical study, we have some special Datum. Mainly, three types of Data are used for this study. These are: (A) Image Data; (B) Spectral Data and (C) Time series and Functional Data.

**Image Data, in Astronomy:**

For astronomical Study, Image data is one of the most essential data types. The images taken by different telescopes is used as the Astronomical Data. One of the most useful telescopes is the Dark Energy Camera (DECam). It takes the images, as the part of the Dark Energy Survey (DES), with a photometric filter which blocks certain light wavelengths. Here, the Figure 4, shows one night sky image, taken by DECam.

**Figure 4:**

The night sky image, taken by the DECam. The white lines are gaps between the CCDs.

**Figure 5:**

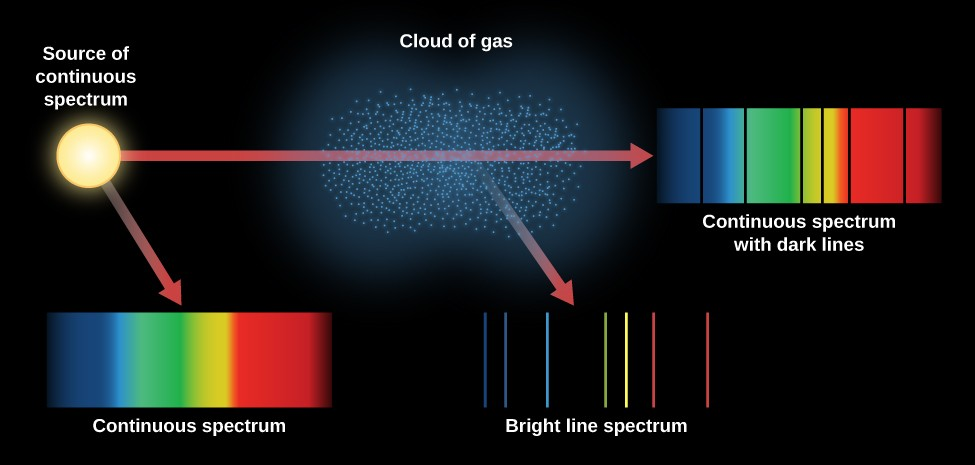
Dark Energy Camera (DECam).

**Spectral data, in astronomy:**

Spectral Data is another important tool for analyzing Astronomy. Basically, spectrum represents the intensity of light in different wavelengths, providing considerably more information than can be directly inferred from image data. Figure 6, shows the structure of the Spectrums. In, Figure 7, it shows the spectrum of the galaxy Messier 77, a barred spiral in the Cetus constellation.

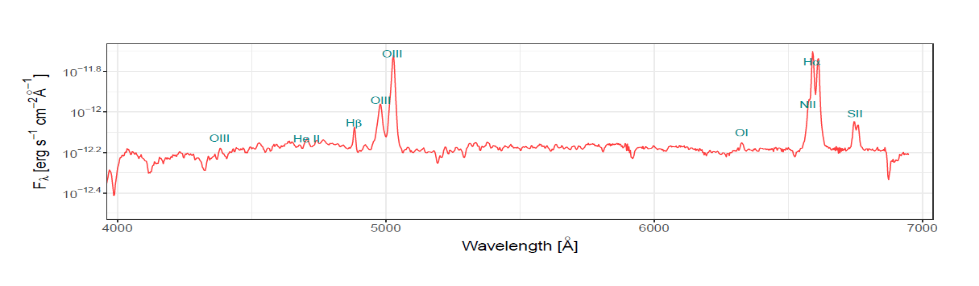
Spectrums basically, carry information about some of the most important properties like, temperature, chemical composition, etc.

With the development of large spectrographic surveys, a large amount of spectral data can be easily obtained through various data mining methods, to assist astronomers’ spectral analysis.

****

**Figure 6:**

Structure of the Spectrums.



**Figure 7:**

Example of a galaxy spectra from Messier 77.

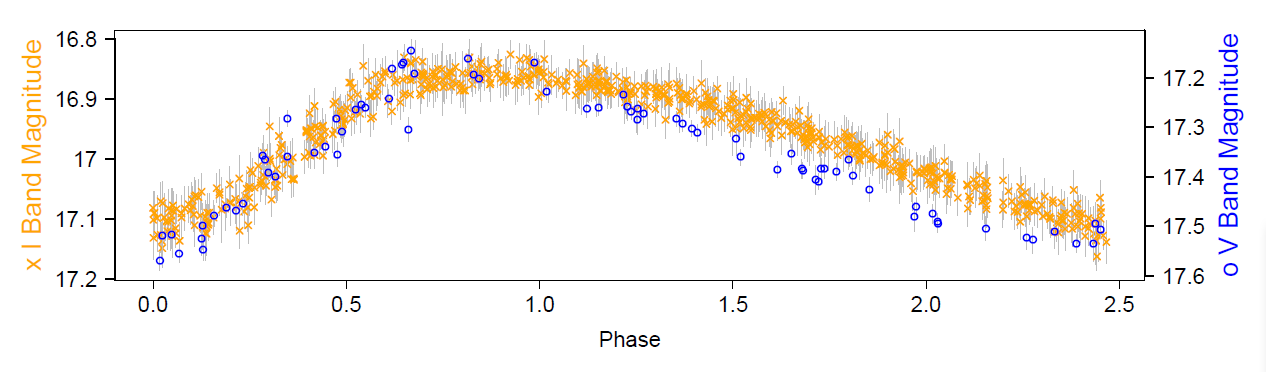
**Time Series and Functional Data:**

Time series data, serves an important role in Astronomical analysis. There are various kind of variable objects exist in this universe, including lots of stars, with their predictable behavior, various kind of objects with their behavior, which are inherently unpredictable, and objects with both predictable and irregular variability, in their patterns. To understand the nature of variability, of different objects in the universe, the Astronomers often use the time series data to predict their behaviors.

In, Figure 8, it shows a time series data, for a certain star, observed by the Optical Gravitational Lensing Experiment (OGLE). The data is represented in two types of structures, one is represented by orange crosses and another one with blue circles, over the course of approximately 10 years. The time spacing between each of the observations, is irregular, which is a typical feature in astronomical data. In this experiment, approximately 400,000 of these light curves are collected. Also, in, Figure 9, it shows the time series data plot on 3 types of random stars (Produced from Rebbapragada et al. [Rebb09]).

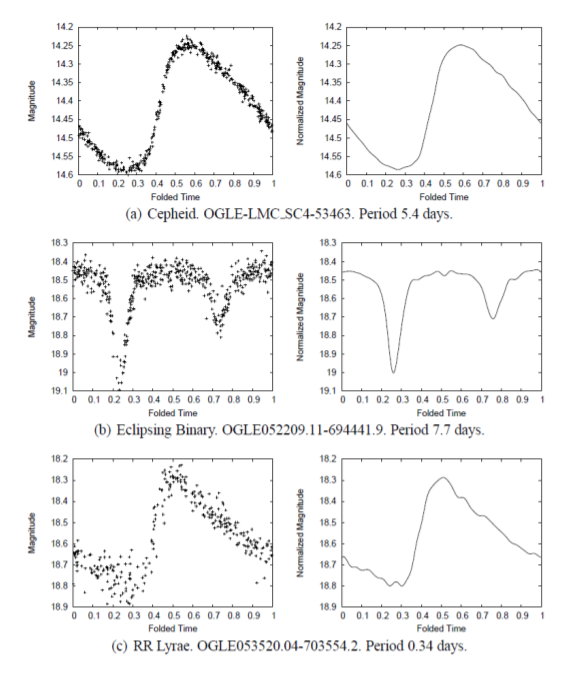
In astronomy, time series analysis of flux measurements is very common. As a consequence of many decades of observations in stars, a large variety of flux variations have been detected by many astronomical objects, including periodic variations, such as, pulsating stars, pulsars, rotators, eclipsing binaries, planetary transits; quasi-periodic variations, such as, star spots, active galactic nuclei, neutron star oscillations, etc. Such astrophysical phenomena are wavelength-specific cases, or were discovered as a result of wavelength-specific flux variations, like soft gamma ray repeaters, X-ray binaries and gravitational waves.

In addition to flux-based time series analysis, astronomical data also include motion-based time series data. These include the trajectories of different planets, comets, and asteroids in the Solar System, the motions of stars around the massive black hole from the center of the Milky Way galaxy.



**Figure 8:**

Light curve of a variable star observed by OGLE. Models from the time series and functional data analysis literature are often used for studying these objects.



**Figure 9:**

Examples of time series data for 3 different types of variable stars (reproduced from Rebbapragada et al. [Rebb09]).

**Conclusion**

In the past astronomers did everything one by one, from the conception of a project to assortment of information and their analysis. As the instruments become complicated, teams had to be come upon and that they more and more enclosed folks specialized in technology. Now it’s not possible to run a project at the forefront of astronomical analysis while not facilitate of those technologies. Thus, we ought to work on larger samples if we wish to require advantage and fully use the information contained all told these data, globally and one by one. A technique to figure with efficiency of massive samples is to use and if necessary to develop associate degree adequate statistical methodology.

Now, we will nearly perceive why statistics is so much important in astronomy. Astronomers are absolutely obsessed with it, without any application of statistics it’s impossible to move on and answer such questions: Can all the eight planets ever line up to an equivalent aspect of sun? Is it possible to see the ring of Saturn from the equator or the poles of it? Is there a center of the universe? What came before the Big Bang? How to measure a galaxy’s distance? What percentages of stars are there within the Milky Way?

In the application of distance scale, more specifically, analysis of errors and calculation of regression coefficients are two important things in measuring the speed of enlargement of the universe (one of the foremost necessary parameters in cosmology), estimating the age of the universe and uncovering massive scale phenomena like super clustering and galaxy streaming.

Multivariate methods such as Principal Components Analysis (PCA) and Cluster Analysis are most common methods that are used in astronomy. Often clustering method is applied to choose anomalous or peculiar objects. These techniques will also work in multidimensional parametric space. In stellar astronomy, one must to study the interface between photometry and spectrographic analysis, particularly within the framework of stellar classification. In star-galaxy separation process also we seek to look for the assistance of regression and interference.

So, it’s clear that in order to figure with astronomical data we will be benefitted if we have a tendency to take help of statistics. In order to manage all kinds of statistical data, its method and mindsets are very crucial during this developing society toward a better and sustainable future.

**References:**

1. **Sources of the different Figures:**

Figure 1:

[https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mn](https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mn%20) [ras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoE CBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs\_\_O](https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mnras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoECBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs__O) (Page 8)

Figure 2:

<https://images.app.goo.gl/8hGwjWeBAvt5vsSQ6>

Figure 3:

Nonparametric and robust statistic by Eric Feigelson (3rd INPE Advanced School in Astrophysics: Astro statistics 2009) (Page 5)

Figure 4:

<https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05>

[834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw 1fNRPIXR\_19HHnxtHBPaQ6](https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw1fNRPIXR_19HHnxtHBPaQ6) (Page 2)

Figure 5:

<https://www.darkenergysurvey.org/the-des-project/instrument/>

Figure 6:

<https://courses.lumenlearning.com/towson-astronomy-2/chapter/formation-of-spectral-lines/>

Figure 7:

[834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw 1fNRPIXR\_19HHnxtHBPaQ6](https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw1fNRPIXR_19HHnxtHBPaQ6) (Page 2)

Figure 8:

[834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw 1fNRPIXR\_19HHnxtHBPaQ6](https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw1fNRPIXR_19HHnxtHBPaQ6) (Page 3)

Figure 9:

<https://r.search.yahoo.com/_ylt=Awrxzw8jfdRhcTwAbAC7HAx.;_ylu=Y29sbwNzZzMEcG9zAzEEdnRpZAMEc2VjA3Ny/RV=2/RE=1641344419/RO=10/RU=https%3a%2f%2fcs.gmu.edu%2f~jessica%2fpublications%2fastronomy11.pdf/RK=2/RS=L3uJLuHMdGQxwYG1EFZlPF2FQF4-> (Page 3)

1. **Sources of the Equations in Regression Analysis Section:**

[https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mn ras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoE CBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs\_\_O](https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mnras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoECBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs__O) (Page 3 and Page 5)

1. **Sources of the Equations in Likelihood Based Model Selection:**

<https://www.iiap.res.in/astrostat/School10/LecFiles/Karandikar_Babu_ModelSelGOF_notes.pdf> (Page 5)

**For Further Readings:**

1. Modern Statistical Methods for Astronomy-by Eric D. Feigelson and G. Jogesh Babu
2. Statistical methods in astronomy by James P. Long and Rafael S. De Souza
3. [https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mn ras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoE CBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs\_\_O](https://www.google.com/url?sa=t&source=web&rct=j&url=https://academic.oup.com/mnras/article/455/2/2149/1111686&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoECBwQAQ&usg=AOvVaw1sJyNI60vBpjPVlfeAs__O)
4. [https://www.google.com/url?sa=t&source=web&rct=j&url=https://core.ac.uk/download/ pdf/25195154.pdf&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoECBsQAQ&u sg=AOvVaw3jxIG7Jj1Dyn8k4MVl7Ub0](https://www.google.com/url?sa=t&source=web&rct=j&url=https://core.ac.uk/download/pdf/25195154.pdf&ved=2ahUKEwiX0YGjxfnzAhUEzzgGHfn9CtQQFnoECBsQAQ&usg=AOvVaw3jxIG7Jj1Dyn8k4MVl7Ub0)
5. [https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05](https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw1fNRPIXR_19HHnxtHBPaQ6)

[834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw 1fNRPIXR\_19HHnxtHBPaQ6](https://www.google.com/url?sa=t&source=web&rct=j&url=https://arxiv.org/pdf/1707.05834&ved=2ahUKEwjo1PrsxPnzAhUkzDgGHUYaBA4QFnoECAQQBg&usg=AOvVaw1fNRPIXR_19HHnxtHBPaQ6)

1. Modern Statistical Methods for Astronomy – NASA/IPAC Extragalactic Database
2. Nonparametric and robust statistic by Eric Feigelson (3rd INPE Advanced School in Astrophysics: Astro statistics 2009)
3. [https://science.psu.edu/science-journal/winter-2020/astronomy-is-better-with-betterstatistics](https://science.psu.edu/science-journal/winter-2020/astronomy-is-better-with-better-statistics)

**Adrija Bhar (Roll: 416); Manopriya Pal (Roll: 430); Tamasha Dutta (Roll: 445)**

**[2nd Year, Department of Statistics]**

**St. Xavier’s College (Autonomous), Kolkata**