Abstract—The outcomes of exoplanet search over the last two decades have given us a wide variety of different types of exoplanets, with a lot of different characteristics, changing our understanding of planetary systems. The study of habitability and life[1] has been one of the primary focus in planetary sciences. But, to understand this, first it is needed to know more about the architectures, stabilities and configurations of these systems. With that aim, a lot of surveys have been done, modelling the occurrence rate of exoplanets around stars with specific physical conditions. Here, a comparison between different models is presented using Bayesian statistics to find the best fit for distribution models of eccentricity of planetary orbits, using the day to day increasing exoplanet catalogue. A trend is observed for massive planets to have more elliptic orbits.

Keywords—exoplanets, Methods: data analysis, statistical, planetary systems.

I. INTRODUCTION

The development of better technologies and multidisciplinary efforts in the last ten to twenty years, trying to understand how planetary systems are formed, have had extraordinary results[2]. The database of exoplanets has been increasing for the last five to six years at a very fast pace, almost one new exoplanet is discovered every two to three days. It has become relevant to figure it out, the architecture and orbital parameters of these exoplanet systems. There has been different surveys [3] , [4] characterizing different distributions of the exoplanets occurrence rate. Most of these have been done with a much smaller sample of exoplanets and highly biased for a particular detection method, and sometimes, without any physical or statistical fundamentation. Although, many interesting correlations have been found between orbital properties of exoplanetary systems. For example, a possible correlation between obliquity and rotation of hot exoplanets. [5] . Additionally the implementation of Bayesian Statistics, Machine Learning and Big data techniques are now a trend in almost all the branches of Astronomy and Astrophysics [6]. Here, I try to compare already existing relations for exoplanet surveys that were done with a much smaller sample and single detection techniques, mixing all the exoplanet catalogue available. As a reference, just one third of the exoplanets catalogued were discovered after 2015. The eccentricity of the orbits is one of the best studied orbital parameters of exoplanets, also it is very important to the stability of the system. Surprisingly, most of the surveys done with Doppler technique, show that there is a population of planets that tend to have more eccentric orbits than expected. A characteristic that hasn’t been yet explained by the current planet formation theories. It is important to consider the detection technique used to find the exoplanet, because they are highly biases [7]. For example the Transit technique works mainly for edge-on systems, While Direct Imaging works just for giant planets that orbit farter out their system.

II. DATA

The Data catalogue was obtained from http://exoplanet.eu/catalog/, the catalogue contained 3610 exoplanets at June 16, 2017, with a maximum number of features of 98, but most of those parameters weren’t used, because some exoplanets lack of some info. At the end, just six of them were considered: The Orbital distance, the Orbital period, the Metallicity of the host star, the eccentricity, the mass, and the bound mass. Between the other features, it could be parameters, such as effective temperature of host stars, planet size, orbit inclination, stellar type, stellar age, etc. Nevertheless, there are actual research teams involved in obtaining these parameters, as quickly as possible, in order to obtain more robust statistics about the formation and the architecture of these systems.
III. METHODOLOGY

To fit the data, Bayesian Regression will be used. That means, the minimization of the Negative Log Likelihood to find the best fit and model to the data. The goal is to do a model comparison and find the best description of eccentricity distribution. The problem will be analyzed dividing a sample of the catalogue in two. First, a subsample of 600 of the planets that were found using the Doppler technique (Radial Velocity), chosen randomly were used to do the Bayesian regression, and then the model comparison will be done with the rest of the sample, 243 planets. That’s because they represent the more heterogeneous sample of exoplanets with less bias. The rest of the exoplanets are less distributed in eccentricity, so the correlation is less representative, more data and the improvement of the techniques already developed is needed.

The two models to compare are the following:

1. The first model uses a Rayleigh distribution and was proposed by Juric & Tremaine [8]
   \[ f(e, a) = \frac{dN}{de} = \frac{e}{a^2} \exp\left(-\frac{e^2}{2a^2}\right) \]  

2. The second one by Shen & Turner [9]
   \[ g(e, a) = \frac{dN}{de} \propto \frac{1}{(1+e)^a} - \frac{e}{2a} \]  

The function to minimize is the negative log likelihood:

\[ NLL(a, e, \sigma) = \sum \frac{1}{2\sigma^2}(N(e) - y(a, e)) \]  

where \( y(a, e) \) is an integration of the function \( f \) or \( g \) depending on the case. To minimize the NLL, I used the BFGS algorithm, and to find out the better model. The mean squared error will be used as a parameter of model comparison, to all the sample of exoplanets that were detected using the Radial Velocity method. So, the model with a lower MSE will be chosen as the best fit for the data given.

\[ \frac{1}{n} \sum (\hat{Y}(e_i) - Y(e_i))^2 \]  

Then, using the mass and the eccentricity, I will obtain the frequency for a quantity of bins and then estimate a kde with different kernels. I will use an exponential, a cosine, a tophat and a gaussian kernel, but a priori, I assume that the distribution should behave as a gaussian. Finally, I will do parameter estimation for the mean and standard deviation of the sample.

The kernels that will be used are the following:

- **Gaussian**: \( K(x, h) \propto \exp\left(-\frac{x^2}{2h^2}\right) \)
- **Tophat**: \( K(x, h) \propto 1 \), if \( x < h \)
- **Exponential**: \( K(x, h) \propto \exp\left(-\frac{x}{h}\right) \)
- **Cosine**: \( K(x, h) \propto \cos\left(\frac{\pi x}{2h}\right) \)

IV. RESULTS

For the eccentricity distribution the parameters obtained minimizing the NLL, and the accuracy (MSE) are shown in the Table 1.

The Table 2. shows the result of the parameter estimation of the kde for the eccentricity distribution for the two subsamples.
Fig. 4. Distribution of eccentricity for massive and small planets

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juric &amp; Tremaine($\sigma$)</td>
<td>405</td>
<td>2.57</td>
</tr>
<tr>
<td>Shen &amp; Turner ($\alpha$)</td>
<td>2.05</td>
<td>2.38</td>
</tr>
</tbody>
</table>

**TABLE I. MODEL COMPARISON FOR THE ECCENTRICITY DISTRIBUTION**

The Fig 5. shows the distribution for massive and giant planets with the separation at fifty Earth masses, and the Fig 6. the distribution of smaller planets.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Small Planets $e$</th>
<th>Massive Planets $e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$0.152 \pm 0.164$</td>
<td>$0.232 \pm 0.113$</td>
</tr>
<tr>
<td>TopHat</td>
<td>$0.148 \pm 0.168$</td>
<td>$0.230 \pm 0.117$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$0.152 \pm 0.163$</td>
<td>$0.233 \pm 0.112$</td>
</tr>
<tr>
<td>Cosine</td>
<td>$0.152 \pm 0.170$</td>
<td>$0.234 \pm 0.122$</td>
</tr>
</tbody>
</table>

**TABLE II. ESTIMATED PARAMETERS FOR THE ECCENTRICITY MEAN, AND STANDARD DEVIATION WITH THE DIFFERENT KERNELS.**

V. CONCLUSIONS

First of all, It is noticeable, that the distribution changes considering all the detection or just the Radial Velocity method, so there is a great bias in the models inferred using those surveys. Just comparing the two models, they fit similarly to the Radial Velocity data. But when, all the planets are considered, the Rayleigh fails at fitting the distribution, because it seems to work better in the high eccentricity portion of the distribution. On the other hand, the model of Shen & Turner is more sensitive to the circular orbits, so it fits them better. The bias happens, because planets with low eccentricity, meaning less elliptical orbits are easier to detect and confirm by the other methods. It is clearly seen in the graphics, than were all the catalogue is considered, the histogram is more concentrated in the lower eccentricity part of the distribution.

However planet formation models predicts that planets tend to follow circular orbits because of gas drag during their formation. It is yet unclear if the exoplanetary distribution for the RV technique is a deficiency of the detection methods, or a real tendency in the planet formation process. In the near future, with more sensitive telescopes and dedicated surveys, it should be possible to discriminate the answer.

Studying the distribution of eccentricity according to their mass, it was found that smaller planets tend to have more stable and circular orbits than giant planets. However, a mean of $\sim0.15$ is still high enough to produce some instabilities in their parent systems in the case of multiplanetary systems. For example, except for Mercury, none of the planets of the Solar System have a eccentricity higher than 0.1. In the case of
giant planets, Neptune-size or bigger, there is a clear tendency to higher eccentricity. A effect that is still puzzeling for planet formation theories.

Previous studies had remarked that very eccentric orbits are associated with stars with higher metallicities. Although, Fig 3 doesn’t show a significant difference for the stars with higher abundance of metalsit seems that most of those planets orbits around these types of stars. However it seems to be a difference in the distribution it is less pronounced than previous studies, but it still exists. The reasons that could possibly explain this behaviour are not yet understood, but could be associated at the stability and weight of heavier elements in these second generation star systems. It is worthwhile study these particular planets in more details to see if any other correlation exist among them.

VI. FUTURE WORK AND DISCUSSION

Although the exoplanet catalogue available is increasing. There is a lot of work to do finding parameters and characterizing the incomplete data in the catalogue. Several attempt are not just possible, but mandatory in order to get a confident and robust result of occurrence rate modeling. Not just characterize the architecture of these systems, but to physically model them. Also, there are a lot of other possible combinations of parameters that show some kind of correlation.

The detection methods are getting more sensitive and they are now available to detect Earth size planets in the habitability zone. So, it is not wrong to say that we are getting closer to finding more and more planetary systems suitable for the formation of life. Future instruments and facilities aims at this target. The quantity and quality of data in astronomy is already increasing, so efficient and fast algorithms are needed to keep the pace, the use of Machine Learning, Big Data and Bayesian Statistics are becoming more important.

In particular, the relevance and long term stability of exoplanetary systems with giant planets that present a very elongated orbit, and how this affects the suitability for life birth.

REFERENCES