ASTM21: Project 1

Robin Emanuelsson

November 2019

Introduction

The two point correlation function, $\omega(\theta)$, is used for description of data clusters, such as galaxy distribution. Clustering increase the likelihood that two points are close to each other and $\omega(\theta)$ determines a probability dP to find two objects on an angle θ from each other within two solid angle elements $d\Omega_1$ and $d\Omega_2$:

$$\mathrm{d}P = \rho(1 + \omega(\theta))\mathrm{d}\Omega_1\mathrm{d}\Omega_2 \tag{1}$$

where ρ is the surface density.
The correlation function is not without its flaws. Things like phase information is lost, its sensitive to shot noise and the error on the measured $\omega(\theta)$ is difficult to compute for small survey areas. But despite all this the correlation is a quick and easy measure on the clustering.
To estimate $\omega(\theta)$ we must take the distances between the data points, called DD. Take distances in a random distribution of data points, to compare with DD, called RR. The first type of estimator of $\omega(\theta)$, called the natural estimator, is as follows

$$\omega_1 = \frac{r(r-1)}{n(n-1)}\frac{DD}{RR} - 1 \tag{2}$$

where r and n is the size of R and D.
In the litterateur there are 3 more improved estimators. They are:

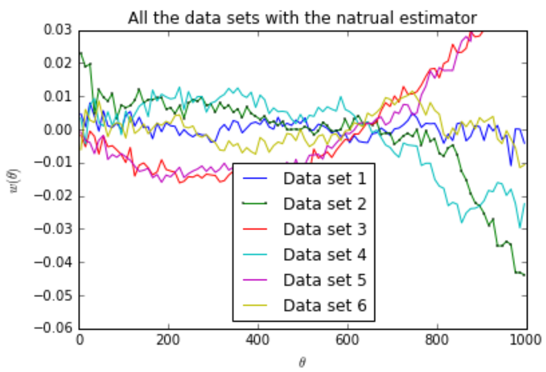
$$\omega_2 = \frac{2r}{n-1}\frac{DD}{DR} - 1 \tag{3}$$

$$\omega_3 = \frac{r(r-1)}{n(n-1)}\frac{DD}{RR} - \frac{r-1}{n}\frac{DR}{RR} + 1 \tag{4}$$

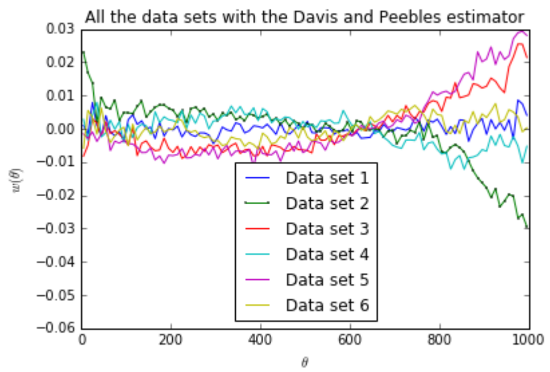
$$\omega_4 = \frac{4nr}{(n-1)(r-1)}\frac{DD \times RR}{(DR)^2} - 1 \tag{5}$$

These are called the Davis & Peebles, Landy & Szalay and Hamilton estimators respectively. Here DR is a distance from the data set to the random data points.
In this report I will use these estimators to determine which one of a given set of 6 different data sets is data from the *Hubble Ultra Deep Field* and which are just random data.

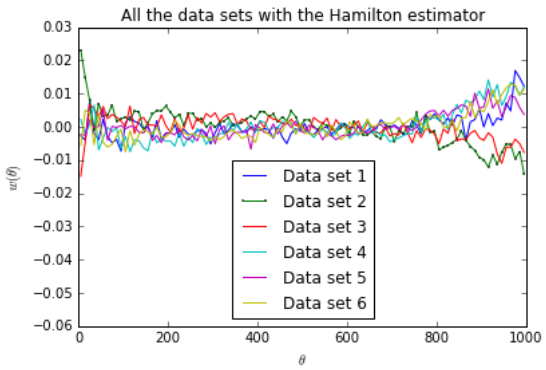
Results



Figur 1: The natural estimator for the different data sets. The dotted line is the second data set



Figur 2: The Davis & Peebles estimator for the different data sets. The dotted line is the second data set



Figur 3: The Hamilton estimator for the different data sets. The dotted line is the second data set

Figur 4: The Landy & Szalay estimator for the different data sets. The dotted line is the second data set

The figures show the different estimators as a function of angular separation, for all of the data sets. The dotted line is the second data set. It's dotted because of its high value, relative to the other data sets, in the first bin.

Conclusion

The estimators over all behave similarly, with the noticeable diffrence that the Hamilton and Landy & Szalay estimator does not spread out so much for large θ. This is due to the problem with handling the variance. For the natural and Davis & Peebles estimator this is more noticeable. In all cases we can see that data set two stands out for the first bin, also indicated by having a dotted line. This indicates that this is the data from the *Hubble Ultra Deep Field* since clustering should increase $\omega(\theta)$ for smaller θ's. For larger scales we assume that the distribution of the galaxies are uniform. We this we also see in that data set two, which we assume is the *Hubble Ultra Deep Field* data sets, starts

to group up in with the other, assumed, random data for the middle part of θ.
The different estimators preformed similarly in detecting what seems to be the *Hubble Ultra Deep Field* data sets. So for this problem the natural estimator is preferable since there is no need to calculate DR, which saves some computational time.