Data Analysis

CHABERT Benjamin
Université Paul Sabatier - Toulouse III
AstroParticules et Cosmologie - Paris

September 2, 2020

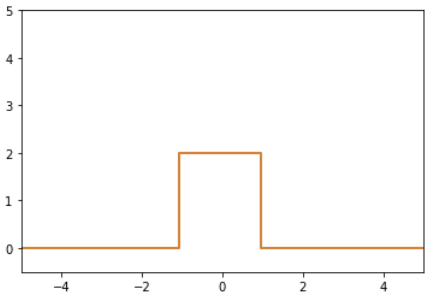
SECTION A - correlation function

Which function to use ?

There are a lots of correlation functions in the python library. First of all, we have compared four of them : the numpy.corrcoef and scipy.stats pearsonr, spearmanr and kendalltau.

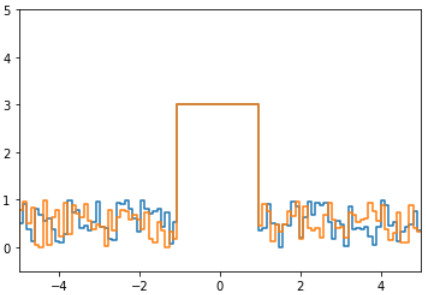
We perform an analysis of these four function thanks to gate functions. The first analysis shows a similarity between them.

```
corr= 1.0
pearson= 1.0
spearman= 0.9999999999999998
kendall= 1.0
```



However, when we put noise on the gate function background, disparities appear.

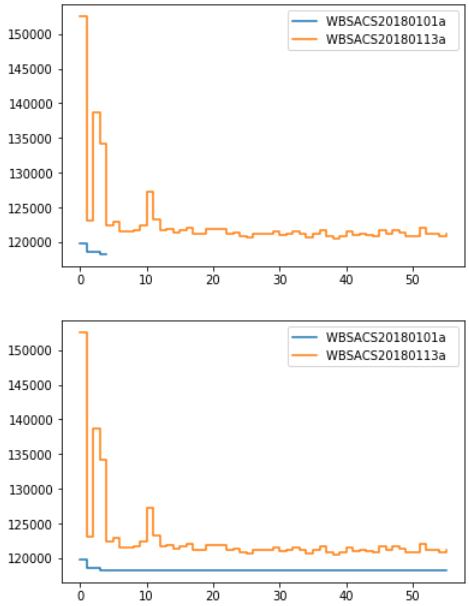
```
corr= 0.9418710757467017
pearson= 0.9418710757467024
spearman= 0.4964069683038954
kendall= 0.342436974789916
```



So, for the following analysis, we will use the pearsonr function (More information)

How to split bursts in families?

With our previous analyzes, we have seen that the most important parameter in the correlation is the peak. So we have to center all the peak before to use the correlation function. However, to use the correlation function the two arrays have to have the same lenght so we add a point at the end with a background value.



We choose a lower limit of 0.7 to define two bursts as similar bursts. With this limit we have statistical pollution. Indeed, in the short family we have 63 "not short" bursts. As the separation is not satisfying, we decide to separate bursts with the number of peak.

SECTION B - number of peak

For that part, we will define what we called a peak. If a point is above the 5-sigma limit and if it is not in an existing peak we define the point as a peak. With this definition, we can classified burst in two categories (short and long) with a limit of 5 peak. Then, inside this two categories, we separe the burst thanks their peak's number.