Cosmology data analysis from BOSS catalogue

Rafael Ribeiro

University of Sao Paulo
rafaelmgr@usp.br

December 13, 2019

Abstract

We analyse the Baryon Acoustic Oscillation Survey (BOSS) data release (DR12). Through ,we will analyze the North and South BOSS galaxy maps, where galaxies in this work start with $z = 0.5$ and go to $z = 0.75$. Using the Λ_{cdm} model with no curvature, and calculate the power spectrum from this data. From the power spectrum estimate the uncertainties of the measurements, and for this we use a covariance matrix not just to estimate the uncertainties but to calculate the correlation between different data points.Soon after we will retrieve the Baryon Acoustic Oscillation (BAO) from the power spectrum. Finally we will do a MCMC to retrieve cosmological information.

I. Introduction

This work presents cosmological results from the final galaxy clustering data set of the Baryon Oscillation Spectroscopic Survey (BOSS,[4]), conducted as part of the Sloan Digital Sky Survey III(SDSS-III,[5]). The BOSS main goal is the measure the cosmic expansion history by means of baryon acoustic oscillation (BAO), which imprint a characteristic scale detectable in the clustering of galaxies and of intergalactic.

So to continue our work we need to define the BAO, and with this definition we can measure the the cosmological distance with a effective redshift in a distribution of galaxy and with this technique we can map the expansion of the universe.

We choose to make our measurements of the BAO using the data set DR12 CMASS and LOWZ combined. This data have two maps of the sky, the north and south. So with this, we will estimate and compute the cosmological parameters.

After we chosen the data, we make numerical calculation to extract and estimate our cosmological parameters. First we extract the power spectrum using the south and the north BOSS data. Then we using a covariance matrix to calculate uncertainties of the power spectrum and also compute the correlation between different data points.

To extract the BAO signal we will use a model for the power spectrum where we remove all but the oscillation signal. To obtain such thing, we can do fit with a power spectrum without a BAO signal to a power spectrum with BAO signal ,that we will discuss this model in the next sections, therefore we will obatain the the signal.

Finally we can make a Monte Carlo Markov Chain to constraint the cosmological parameters and together with the MCMC we use the Gelman Rubin[6] method to be sure of the convergence.

II. Baryon Acoustic Oscillation

The baryon acoustic oscillation (BAO) signal in the distribution of galaxies is an imprint of primordial sound waves that have propagated in the very early Universe through the plasma of tightly couple with photons and baryons(

e.g. [9] ; [10]). The corresponding BAO signal in photons has been observed in the Cosmic Microwave Background (CMB) and has revolutionised cosmology in the last two decades (e.g. [1]).

The BAO signal has a characteristic physical scale that represent the distance that the sound waves have traveled before the epoch of decoupling. In the distribution of galaxies, the BAO scale is measured in angular and redshift coordinates, and this observational metric is related to the physical coordinates through the angular diameter distances and Hubble parameters. Therefore, Comparing the BAO scale measured in the distribution of galaxies with the true physical BAO scale, i.e., the sound horizon scale that is independently measured in the CMB, allows us to make cosmological distance measurements to the effective redshift of the distribution of galaxies. With this "standard rule" technique one can map the expansion history of the universe(e.g., [7]; [3]; [8]).

In brief, pressure waves in the pre-recombination universe imprint a characteristic scale on late-time matter clustering at the radius of the sound horizon,

$$r_d = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)} dz \tag{1}$$

evaluated at the drag epoch z_d, shortly after recombination, when photons and baryons decouple. With this information it can be calculated the comoving distance angular

$$D_A(z) = \frac{c}{1+z} \int_0^z \frac{dz'}{H(z')} \tag{2}$$

And we compute the distance

$$D_V = \left[D_A^2 \frac{cz}{H(z)} \right]^{1/3} \tag{3}$$

III. The BOSS dataset

The map that we will use is the data set of the Baryon Oscillation Spectroscopic Survey (BOSS). The BOSS survey is part of SDSS-III ([5];[4]) and used the SDSS multi-fibre spectrographs to measurements spectroscopic redshifts of 1.1987006 million galaxies.

In this analysis we used the data set DR12 CMASS and LOWZ combined. The catalogs have two maps of the sky with redshift of $0. < z < 0.9$. This maps are divided in north and south, there is no overlapping, where we have galaxies obtains by this maps. We also use random map of the galaxies given by BOSS catalogue. For the covariance matrices we use the simulated and random Mock catalogs also from the north and south skies.

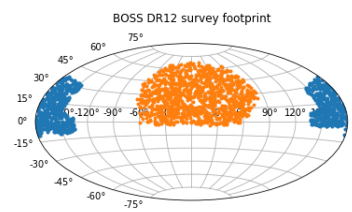

Figure 1: *The sky distribution of galaxies in the BOSS galaxy survey. Only 1000 points for each region are plotted.*

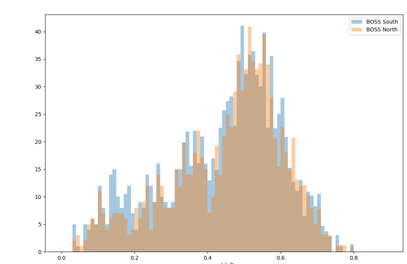

Figure 2: *Redshift distribution of the BOSS galaxy sample. Redshift* 0.7 *corresponds to about 6 billion light-years.*

IV. Methodology

In this section will be show the aspects of our analysis. As mentioned above, we analyze galaxies from the catalogue of BOSS, here we consider the redshift $z = 0.5$ and $z = 0.75$, to calculate the power spectrum, covariance matrices and cosmological parameters from both north and south from BOSS catalogues.

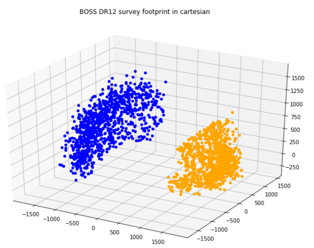

Figure 3: *3D distribution of BOSS galaxies. Only 1000 randomly selected galaxies out of more than 1 million in this dataset are included in this plot*

To calculate the power spectrum we combine the data catalogues and random catalogue, assign the point distributions to a 3D grid and calculate the power spectrum, all using the python package "nbodykit".Here we use a grid with 512 grid points in each of the 3 dimensions. There are also two weightings included in this calculation, the "Weight" and "WEIGHT FKP" columns. The first weight refers to a correction of incompleteness in the data catalogue, while the second weight tries to optimise the signal-to-noise.

To continue our analysis, we use the covariance matrix to calculate the uncertainties of the power spectrum. Also we calculate the correlation between different data points. All the information is contained in the covariance matrix, which is defined as

$$\mathrm{C(k}_i, k_j) = \begin{pmatrix} \sigma_{k_1}^2 & cor(k_1,k_2)\sigma_{k_1}\sigma_{k_2} & \dots & cor(k_1,k_n)\sigma_{k_1}\sigma_{k_n} \\ cor(k_2,k_1)\sigma_{k_2}\sigma_{k_1} & \sigma_{k_2}^2 & \dots & cor(k_2,k_n)\sigma_{k_2}\sigma_{k_n} \\ \vdots & \vdots & \ddots & \vdots \\ cor(k_n,k_1)\sigma_{k_n}\sigma_{k_1} & cor(k_n,k_2)\sigma_{k_n}\sigma_{k_2} & \dots & \sigma_{k_n}^2 \end{pmatrix} \quad (4)$$

Here σ_{ki} is the standard deviation of the power spectrum $P(k_i)$ at wavelength k_i. The off-diagonal elements of this matrix contain the uncertainties on a data point contributed by neighbouring data points, which are proportional to the correlation $cor(k_i, k_j)$ between the two data points.

To make the calculation we user MOCK catalogues of the BOSS together with the correspond random catalogue. Given that the simulated catalogues we again use the package,nbodykit, to read and calculate.

The Next step is extract the BAO signal. To do this we use a model for the power spectrum where we remove all but the oscillation signal. To obtain such thing, we can do fit with a power spectrum without a BAO signal to a power spectrum with BAO signal.

This model uses the noBAO power spectrum together with 5 additional polynomial terms. These parameters have no physical motivation, but their whole purpose is to make this fit independent of the broadband and focus on the oscillation signature.

$$\mathrm{P}^{model}(k) = p_0^2 P^{noBAO}(k) + \tfrac{p_1}{k^3} + \tfrac{p_2}{k^2} + \tfrac{p_3}{k} + p_4 + p_5 k \quad (5)$$

This procedure removed the entire broadband power spectrum and isolated the oscillation signature but before we continue the model need to be extended. Here we introduce the smooth model, but modify it by introducing a smoothing scale and a scaling parameter.

Finally we will user a Monte Carlo Markov Chain(MCMC) to constraint the cosmological parameters. Here I use the χ^2 and write a function to transform into a log-likelihood to apply in MCMC. The MCMC has a problem with the convergence. To solve this problem, we will use the Gelman Rubin criterium [[6]]. In the end of the MCMC we will obtain our cosmological parameters mainly the scaling parameters α

V. Discussion and results

Now we discuss our result obtain by the dataset. First we will see the power spectrum obtained with the galaxy maps(figure 4).

Here we see that the power spectra of the northern and southern parts are slightly different. Whether these differences are significant is not clear though, since we don't have uncertainties on these measurements. Now we need to obtain the uncertainties in the power spectrum, to see this we will make a plot of ther correlation matrix that can be obtained with

$$R = \frac{C_{ij}}{\sqrt{C_{ij}C_{ji}}} \quad (6)$$

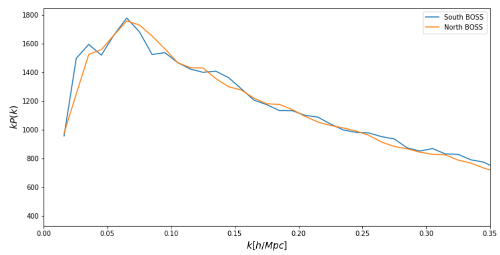

Figure 4: *Power spectrum measurements for the northern (orange) and southern (blue) parts of the BOSS dataset.*

And now we can make the plot of the this matrix

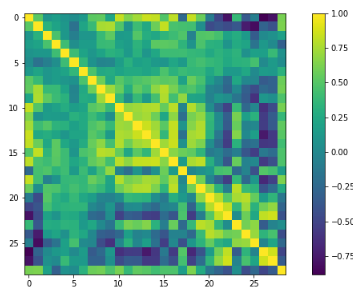

Figure 5: *Correlation matrix, derived from the covariance matrix of the southern part of BOSS using 1000 simulated catalogues.*

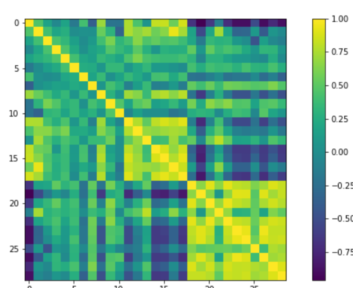

Figure 6: *Correlation matrix, derived from the covariance matrix of the northern part of BOSS using 1000 simulated catalogues.*

Figure 5 and figure 6 shows that there is some correlation between the power spectrum bins close to each other, while some bins which are far apart are anti-correlated. Now with the daigonal of covariance matrix we can obtain the variance in the power spectrum if the different data points would not be correlated.

$$\sigma_{P(k_i)} = \sqrt{C(k_i, k_i)}. \tag{7}$$

Using these uncertainties we can plot the BOSS measurements with error bars as show in the figure 7.

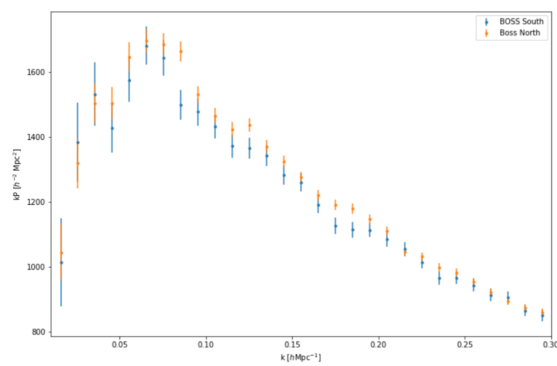

Figure 7: *The BOSS power spectrum measurements including the diagonal elements of the covariance matrix as error bars.*

With this data, we can now constrain the cosmological parameters. As we discuss in the methodology we will extract baryon acoustic oscillation, to do so we gonna first minimizing the χ^2 to obtain the parameters with the best fitting to our model.

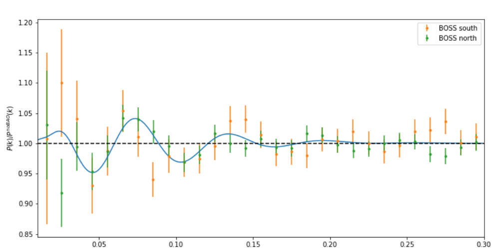

Figure 8: *BOSS measurements (data points) together with the best fitting model.*

And the best fitiing parameters are in the table 1.

2.38774786	$-1.40192579 \times 10^{-1}$
4.43605352×10^{1}	$-3.79576413 \times 10^{3}$
1.67608592×10^{4}	2.38719897×10^{4}
$9.97996589 \times 10^{-1}$	1.10862632×10^{1}

Table 1: *Best fitting values for* χ^2

We obtain the BAO, as show in the figure 8, and the respective error with the south and north. However the fitting parameters do not have uncertainties attach to them. And without this we cannot constraint the cosmological parameters.

To constraint the uncertainties we use the MCMC with the criteria of convergence the Gelman and Rubin. The meaning of the convergence can see in the figure 9

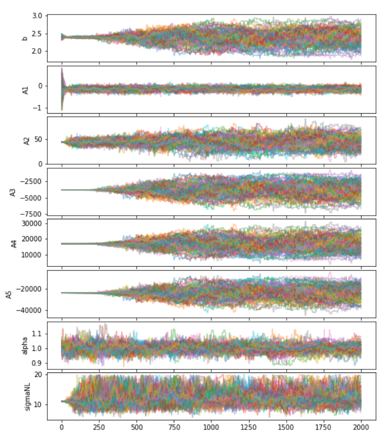

Figure 9: *Variance in the different parameters.*

The convergence has been reached if the variance is constant with time. In the early stages of the chain the variance changes, indicating the the convergence has not been reached.But as we can see with time the parameters converge if so, we can calculate the likelyhood.

Now we obtain our α scaling parameter, that is defined as

$$\alpha = \frac{D_V}{D_V^{\text{fid}}} \tag{8}$$

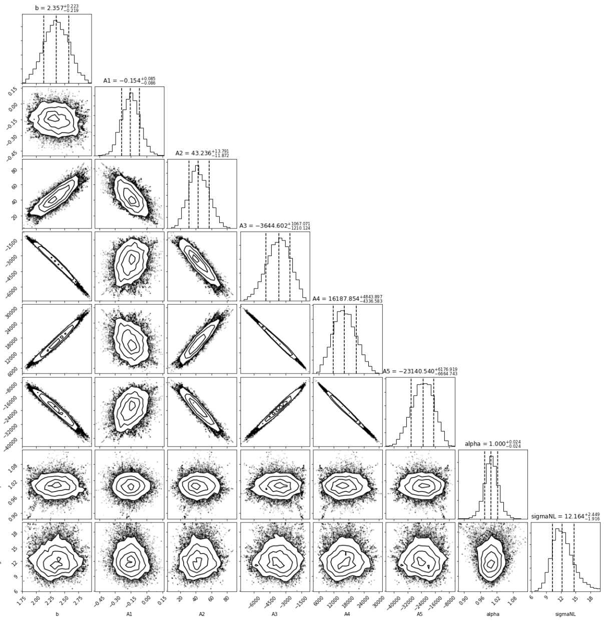

Figure 10: *2D likelihood distributions for the 7 parameters in our fit.The 1D marginalised likelihood is given on the right together with the 68% confidence levels.*

where the the distance D_v is defined as

$$D_V = \left[(1+z)^2 D_A^2 \frac{cz}{H(z)} \right]^{1/3} \tag{9}$$

The D_V^{fid} is the fiducial model, that we defined in the beginning in the class of nbodykit.

The distance $D_V(z)$ is an average of the angular diameter distance $D_A(z)$ and the Hubble parameter $H(z)$

$$D_A(z) = \frac{c}{1+z} \int_0^z \frac{dz'}{H(z')} \tag{10}$$

$$H(z) = H_0 \left[(\Omega_{cdm} + \Omega_b)(1+z)^3 + \Omega_\Lambda \right]^{1/2} \tag{11}$$

Which is numerically calculated

$$D_v = 2163.810118 \tag{12}$$

The other parameters can be constraint comparing with the Cosmic Microwave Background (CMB), and then obtaining the actual cosmological parameters. But the main purpose in this work is the calculated α and the distance $D_v(z)$.

VI. Conclusion

We have presented measurements of the cosmological distance redshift relation using a numerical calculation from the SDDS-III Baryon Acoustic Oscillation Survey where we use a redshift $0.5 < z < 0.75$. So we extract the BAO with errors related to the south and north maps of BOSS, given a the best values for the fit and constraint the α scaling parameter to obtain the cosmological distance eq.12.

We can compare our result to the [2] and we will see that our cosmological distance diverge $1,074766355\%$ that Alam measure in his work.

References

[1] P. A.R. Ade et al. "Planck 2015 results: XIII. Cosmological parameters". In: *Astronomy and Astrophysics* 594 (2016). ISSN: 14320746. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.

[2] Shadab Alam et al. "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample". In: *Monthly Notices of the Royal Astronomical Society* 470.3 (2017), pp. 2617–2652. ISSN: 13652966. DOI: 10.1093/mnras/stx721. arXiv: 1607.03155.

[3] Chris Blake and Karl Glazebrook. "Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler". In: *The Astrophysical Journal* 594.2 (2003), pp. 665–673. ISSN: 0004-637X. DOI: 10.1086/376983. arXiv: 0301632 [astro-ph].

[4] Kyle S. Dawson et al. "The baryon oscillation spectroscopic survey of SDSS-III". In: *Astronomical Journal* 145.1 (2013). ISSN: 00046256. DOI: 10.1088/0004-6256/145/1/10. arXiv: 1208.0022.

[5] Daniel J. Eisenstein et al. "SDSS-III: Massive spectroscopic surveys of the distant Universe, the MILKY WAY, and extra-solar planetary systems". In: *Astronomical Journal* 142.3 (2011). ISSN: 00046256. DOI: 10.1088/0004-6256/142/3/72. arXiv: 1101.1529.

[6] Andrew Gelman and Donald B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". In: *Statistical Science* 7.4 (Nov. 1992), pp. 457–472. ISSN: 0883-4237. DOI: 10.1214/ss/1177011136. arXiv: 1011.1669. URL: http://stacks.iop.org/1751-8121/44/i=8/a=085201?key=crossref.abc74c979a75846b3de48a5587bf708f%20http://projecteuclid.org/euclid.ss/1177011136.

[7] Wayne Hu and Martin White. "Acoustic Signatures in the Cosmic Microwave Background". In: *The Astrophysical Journal* 471.1 (1996), pp. 30–51. ISSN: 0004-637X. DOI: 10.1086/177951. arXiv: 9602019 [astro-ph].

[8] Eric V. Linder. "Cosmic growth history and expansion history". In: *Physical Review D - Particles, Fields, Gravitation and Cosmology* 72.4 (2005), pp. 1–8. ISSN: 15507998. DOI: 10.1103/PhysRevD.72.043529. arXiv: 0507263 [astro-ph].

[9] J. T. Peebles, P. J. E.; Yu. "Primeval Adiabatic Perturbation in an Expanding Universe". In: *The Astrophysical Journal.* (1970).

[10] Ya. B Sunyaev, R.; Zeldovich. "Small-Scale Fluctuations of Relic Radiation". In: (1970).