Cosmo Tools ICE Winter School

Marcos Pellejero Ibáñez

November 2022

1 General context and goal of the exercise

In the last decade, spectroscopic surveys have been a core piece in the study of cosmology. Spectroscopic surveys consist of measurements of as many spectra and sky positions as possible in the biggest possible volumes. The sky position allows us to pin down two of the spatial dimensions, while the redshift can be translated into the third spatial component. In figure 1 we can see an example of such a survey (the BOSS survey). Therefore, we end up with a 3D representation of galaxy positions. This distribution of galaxies traces the large-scale structure of the Universe (LSS). This structure is used to infer cosmological parameters. A very simple example is the amount of matter in the Universe $\Omega_{\rm m}$. It is clear that depending on the amount of $\Omega_{\rm m}$, structures will be more or less collapsed at a given redshift.

In this exercise, we will have some LSS data and will explore ways of extracting cosmological information from them. The goal is to have a realistic scenario of the typical quantities that spectroscopic surveys provide and to make the steps followed by the collaborations to pin down cosmological parameters. The procedure is based on spectroscopic quantities but the ideas can be generalized to any kind of cosmological survey. We will start with an introduction to the theoretical concepts that we will use, but feel free to jump over to the exercise section if you are already familiar with the concepts explained here.

1.1 The Universe has structure

Let's start with a quick recap of the theory behind the models that we will be using in this practice. I am not going to be very formal nor detailed since you will have lectures on many of the concepts discussed here during the School.

Just by observing our surroundings, it becomes evident that the Universe is not homogeneous and isotropic. We see places more populated than others, we observe structure. However, on large enough scales (such as those of the CMB), we know the isotropy and homogeneity are satisfied to a high degree. A natural explanation of this structure formation is that it is the result of the gravitational amplification of small primordial fluctuations. This picture is supported by the measurement of small anisotropies in the temperature of the CMB and by

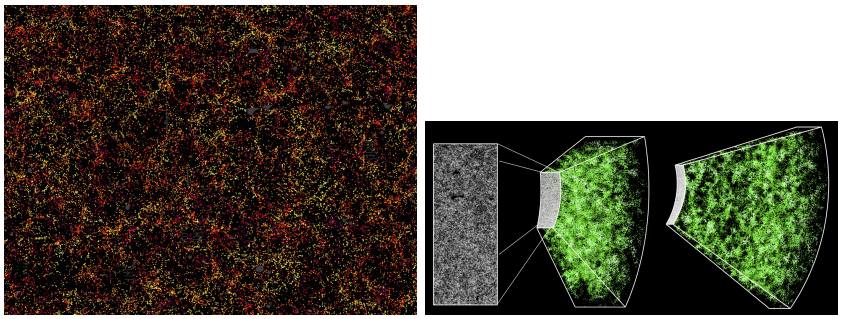

Figure 1: Measured BOSS galaxies. Upper panel: 2D projection where the color represents redshift bins. Lower panel: mapping of the 2D positions to a 3D wedge through the redshift-distance relation.

observations of structure growth at different redshifts (or cosmological time). Let's see two approaches to some general equations for structure formation.

1.1.1 "Hard" (but more complete) way

Let us consider a set of particles with mass m that interact only gravitationally in an expanding universe. The equation of motion for a particle of velocity $\vec{u}$ at position $\vec{r}$ is

$$\frac{d\vec{u}}{dt} = Gm \sum_i \frac{\vec{r}_i - \vec{r}}{|\vec{r}_i - \vec{r}|^3} \,. \tag{1}$$

In the limit of a large number of particles, we can define a smooth gravitational potential satisfying

$$\phi(\vec{r}) = G \int d^r \vec{r}\prime \frac{\rho(\vec{r}\prime)}{|\vec{r}\prime - \vec{r}|} \;\Rightarrow\; \frac{d\vec{u}}{dt} = -\frac{\partial \phi}{\partial \vec{r}} \,. \tag{2}$$

The background evolution is governed by the Friedmann equations and, in order to use Newtonian physics, we consider that positions of particles are described by their comoving coordinates $\vec{x}$. Their physical coordinates are then $\vec{r} = a(\tau)\vec{x}$. We choose to describe the equations of motion in terms of the conformal time, τ, satisfying $dt = a(\tau)d\tau$. This coordinate transformation makes the FRWL metric Minkowskian with a multiplying factor. Defining the density contrast $\delta(\vec{x}, \tau)$

$$\rho(\vec{x}, \tau) \equiv \bar{\rho}(\tau)\left[1 + \delta(\vec{x}, \tau)\right] \,, \tag{3}$$

where $\bar{\rho}$ is the mean (background) density, and the cosmological gravitational potential $\Phi(\vec{x}, \tau)$ is given by

$$\phi \equiv \Phi(\vec{x}, \tau) \;-\; \underbrace{\frac{1}{2}\frac{\partial \mathcal{H}}{\partial \tau} x^2}_{\text{Hubble kinetic term}} \,, \tag{4}$$

where $\mathcal{H}(\tau) = d\ln a/d\tau = Ha$ is the conformal expansion rate. The Poisson equation becomes

$$\nabla^2\Phi(\vec{x},\tau) = \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(\vec{x},\tau)\,, \tag{5}$$

where we used the definition of peculiar velocity $\vec{v}$

$$u(\vec{x},\tau) = \mathcal{H}\vec{x} + \vec{v}(\vec{x},\tau)\,. \tag{6}$$

The Hubble kinetic term is removed from the gravitational potential because we only want to have the potential coming from the overdensities and not from kinetic terms due to the coordinate change. The spatial derivatives are taken over comoving coordinates $\vec{x}$ and we will keep this notation unless stated otherwise. The equation of motion 2 can be written as

$$\frac{d\vec{p}}{d\tau} = -am\nabla\Phi(\vec{x}) \tag{7}$$

with $\vec{p} = am\vec{v}$. We now define the particle number density in phase space by $n(\vec{x},\vec{p},\tau)$. From Liouville's theorem, phase-space number density is conserved, yielding the Vlasov equation,

$$\begin{aligned}\frac{dn(\vec{x},\vec{p},\tau)}{d\tau} &= \frac{\partial n}{\partial\tau}\underbrace{\frac{d\tau}{d\tau}}_{1} + \frac{\partial n}{\partial\vec{p}}\underbrace{\frac{d\vec{p}}{d\tau}}_{-am\nabla\Phi} + \frac{\partial n}{\partial\vec{x}}\underbrace{\frac{d\vec{x}}{d\tau}}_{\vec{v}=\frac{\vec{p}}{am}} \\ &= \frac{\partial n}{\partial\tau} + \frac{\vec{p}}{ma}\cdot\nabla n - am\nabla\Phi\cdot\frac{\partial n}{\partial\vec{p}} = 0\,.\end{aligned} \tag{8}$$

This equation is a non-linear partial differential equation involving seven variables. Therefore, exact solutions are far from our reach except in very specific cases. In practice, however, we are not interested in solving the full phase-space dynamics but rather in the evolution of the spatial distribution. Thus we marginalize the unwanted extra variables. The most common way of doing so is to compute the moments on the phase space distribution. The zeroth order moment is

$$\int n(\vec{x},\vec{p},\tau)d^3p \equiv \rho(\vec{x},\tau)\,, \tag{9}$$

which, after expanding $\int d^3p\frac{df}{d\tau}$ and using integration by parts, gives the continuity equation

$$\frac{\partial\delta(\vec{x},\tau)}{\partial\tau} + \nabla\cdot\{[1+\delta(\vec{x},\tau)]\,\vec{v}(\vec{x},\tau)\} = 0\,. \tag{10}$$

The first and second order moments define the peculiar velocity flow, $\vec{v}(\vec{x},\tau)$, and the stress tensor, $\sigma_{i,j}(\vec{x},\tau)$, fields as

$$\begin{aligned}&\int\frac{\vec{p}}{am}n(\vec{x},\vec{p},\tau)d^3p \equiv \rho(\vec{x},\tau)\vec{v}(\vec{x},\tau)\,,\\ &\int\frac{p_ip_j}{a^2m^2}n(\vec{x},\vec{p},\tau)d^3p \equiv \rho(\vec{x},\tau)v_i(\vec{x},\tau)v_j(\vec{x},\tau) + \sigma_{i,j}(\vec{x},\tau)\,.\end{aligned} \tag{11}$$

The second moment of the Vlasov equation leads to the Euler equation

$$\frac{\partial \vec{v}(\vec{x},\tau)}{\partial \tau} + \mathcal{H}(\tau)\vec{v}(\vec{x},\tau) + \vec{v}(\vec{x},\tau)\cdot\nabla\vec{v}(\vec{x},\tau) = -\nabla\Phi(\vec{x},\tau) - \frac{1}{\rho}\nabla_j(\rho\sigma_{ij})\,, \quad (12)$$

describing the conservation of momentum. Both equations 10 and 12 shows the full evolution of the overdensity and velocity matter fields (assuming baryons do not interact). These are the equations we would require to solve to understand structure formation. Therefore, these are the equations N-body simulations have to solve at each step of their computations. Note that each moment is recursive, the continuity equation couples the zeroth (ρ) to the first moment ($\vec{v}$), the Euler equation couples the first moment to the second (σ_{ij}) and so on and so forth. Equation 11 shows that σ_{ij} characterizes the deviation of particle motions from a single coherent flow. Under the assumption that structure formation is driven by cold dark matter (CDM), which equation of state relies on the assumption that it has negligible velocity dispersion or pressure, it is safe to assume $\sigma_{ij} = 0$ at least in the first stages of gravitational instability, when the structures had no time to collapse. When $\sigma_{ij} \neq 0$, there is a velocity dispersion generation due to multiple streams known as shell crossing.

1.1.2 "Easy" way

Note that one could have arrived at the same set of equations by redefining the "field" derivatives in an expanding space in comoving coordinates as

$$\nabla_r = a^{-1}\nabla_x\,, \quad (13)$$

and

$$\begin{aligned}\left(\frac{\partial}{\partial t}\right)_{\vec{r}} &= \left(\frac{\partial}{\partial t}\right)_{\vec{x}} + \left(\frac{\partial \vec{x}}{\partial t}\right)_{\vec{r}}\nabla_{\vec{x}} \\ &= \left(\frac{\partial}{\partial t}\right)_{\vec{x}} + \left(\frac{\partial a^{-1}(t)\vec{r}}{\partial t}\right)_{\vec{r}}\nabla_{\vec{x}} \\ &= \left(\frac{\partial}{\partial t}\right)_{\vec{x}} - H\vec{x}\cdot\nabla\,.\end{aligned} \quad (14)$$

Then one can just substitute these field derivatives into the usual continuity, Euler and Poisson fluid equations

$$\begin{aligned}\nabla^2_{\vec{r}}\Phi &= 4\pi G\rho \\ \frac{\partial\rho}{\partial t} &= -\nabla_{\vec{r}}\cdot(\rho\vec{u}) \\ \left(\frac{\partial}{\partial t} + \vec{u}\cdot\nabla\right)\vec{u} &= -\frac{\nabla_{\vec{r}}P}{\rho} - \nabla_{\vec{r}}\Phi\end{aligned} \quad (15)$$

with the pressure P vanishing in a matter-dominated universe, and recover 10 and 12 under the assumption of $\sigma_{ij} = 0$. Our original approach is much more general than this one and shows the full difficulty of the problem at hand. Moreover, following the previous ways of reasoning we would have never arrived at the existence of the σ_{ij} field and the dependencies of higher moments.

1.1.3 Solutions to structure formation

Due to their nonlinear nature, equations 10 and 12 are unmanageable even in the approximation of $\sigma_{ij} = 0$. A natural way of approaching the problem is by means of perturbation theory. Within this framework, the fields of interest are expanded as

$$\begin{aligned} \delta(\vec{x}, t) &= \sum_{n=1}^{\infty} \delta^{(n)}(\vec{x}, t)\,, \\ \theta(\vec{x}, t) &= \sum_{n=1}^{\infty} \theta^{(n)}(\vec{x}, t)\,, \end{aligned} \tag{16}$$

where $\theta(\vec{x}, t) \equiv \nabla \cdot \vec{v}(\vec{x}, t)$ is the divergence of the velocity field. The assumption here is that it is possible to expand the density and velocity fields about the linear solutions.

Linear order perturbations

To linear order, expressions 10 and 12 become

$$\begin{aligned} &\frac{\partial \delta^{(1)}(\vec{x}, \tau)}{\partial \tau} + \theta^{(1)}(\vec{x}, \tau) = 0\,, \\ &\frac{\partial \vec{v}(\vec{x}, \tau)}{\partial \tau} + \mathcal{H}\vec{v}(\vec{x}, \tau) = -\nabla\Phi(\vec{x}, \tau)\,. \end{aligned} \tag{17}$$

Taking the time derivative of this second relation and replacing it into the first one, we obtain a single expression for $\delta^{(1)}$ of the form

$$\frac{d^2\delta^{(1)}(\vec{x}, \tau)}{d\tau^2} + \mathcal{H}\frac{d\delta^{(1)}(\vec{x}, \tau)}{d\tau} - \frac{3}{2}\Omega_m \mathcal{H}^2 \delta^{(1)}(\vec{x}, \tau) = 0\,, \tag{18}$$

where we used relation 5. Coming back from comoving time (τ) to physical time (t), the previous equation becomes

$$\ddot{\delta}^{(1)} + 2H\dot{\delta}^{(1)} - 4\pi G\bar{\rho}\delta^{(1)} = 0\,, \tag{19}$$

which is the differential equation of a spring-damper system. The dot denotes derivatives with respect to time t. The damping comes from the term $2H\dot{\delta}^{(1)}$, so it is caused by an expanding (or contracting) universe where $H \neq 0$.

Because equation 19 is a linear equation only depending on the time component, we can split the overdensity field dependencies as $\delta^{(1)}(\vec{x}, t) = D^{(1)}(t)\delta^{(1)}(\vec{x}, 0)$ and remove the spatial dependence coming from the previous equation. $D^{(1)}(t)$ is the linear growth factor. This equation, together with Friedmann equations determines the growth of density perturbations in the linear regime as a function of cosmology. Since e.g in a matter-dominated universe $a \propto t^{2/3}$, we have $H = 2/3t$ and consequently

$$\ddot{D}^{(1)}(t) + \frac{4}{3t}\dot{D}^{(1)}(t) - \frac{2}{3t^2}D^{(1)}(t) = 0 \tag{20}$$

where we used $4\pi G\bar{\rho} = \frac{3}{2}H^2$. Trying a solution of the form $D^{(1)}(t) \propto t^\alpha$ gives the following results,

$$D^{(1)}(t) \propto \begin{cases} t^{-1} \propto a^{-3/2} & \text{Decaying mode} \\ t^{2/3} \propto a & \text{Growing mode} \end{cases} \tag{21}$$

where we will dismiss the decaying mode as it will tend to zero at late times. With this result in hand, we would only need to provide the initial conditions of the overdensity field $\delta(\vec{x}, 0)$ and evolve them with $D(t)$ to obtain the linear structure at any desired cosmological time. For a flat universe with a cosmological constant and matter, the solution to the differential equation 19 can be written in integral form

$$D^{(1)}(a) = \frac{H(a)}{H_0}\left[\int_0^1 \frac{da}{a^3 H(a^3)}\right]^{-1} \int_0^a \frac{da'}{a'^3 H^3(a')} \tag{22}$$

with the term in brackets guaranteeing the normalisation $D^{(1)} = 1$ when $a = 1$. For other equations of state of dark energy, or for more components (curvature, radiation), no general integral expression exists, and the differential equation must be solved directly.

Finally, just note that with the continuity equation defined by $\dot{\delta} = -a^{-1}\nabla\vec{v}$ one can rewrite

$$\frac{d\delta}{dt} = \frac{d\delta}{da}\frac{da}{dt} = \delta_0 \frac{dD}{da}\frac{da}{dt} = \delta_0 \frac{D}{a}\frac{d\ln D}{d\ln a}\frac{da}{dt} = \delta_0 D \frac{\dot{a}}{a} f = \delta f H = -a^{-1}\nabla\vec{v}, \tag{23}$$

with $f \equiv \frac{d\ln D}{d\ln a} = \frac{1}{\mathcal{H}}\frac{d\ln D}{d\tau}$ the growth factor, D is the growing mode, and we have used the notation $\delta_0 = \delta(\vec{x}, t = 0)$. This is our case so we can relate overdensities and velocities in the linear regime as

$$\delta^{(1)} = -\frac{\nabla\cdot\vec{v}}{afH}\,. \tag{24}$$

As a result $\nabla\cdot\vec{v} \propto f$, the growth factor in the linear approximation. It is important to keep this relation in mind for the understanding of parameter dependencies in redshift space distortions that will be studied in a following section.

Full solutions

The only way of finding full non-perturbative solutions to Vlasov's equation is to numerically solve it. The most common way is through large-scale N-body simulations. The importance of numerical solutions for structure formation is that they provide currently the only way to predict the highly nonlinear universe, and potentially extract cosmological information from observations at all scales. In contrast, if one is restricted to scales where perturbative approaches are valid and shell-crossing effects are negligible (i.e., have only a sub-per cent

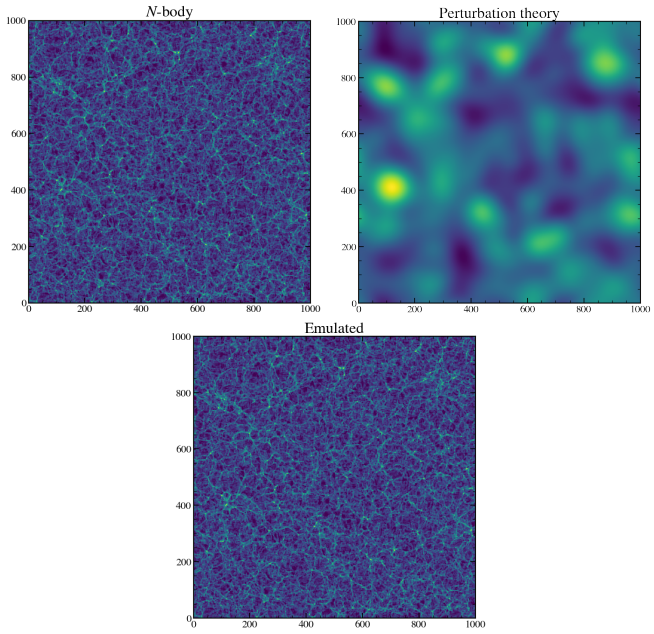

Figure 2: Comparison between the predictions of perturbation theory, an N-body simulation, and an emulator for a volume of 1000Mpc/h.

impact on summary statistics), then a huge amount of cosmological information is potentially lost.

In figure 2 we show the visual differences between perturbation theory predictions and full N-body simulations. Numerical simulations are able to explore smaller scales than perturbative approaches.

The main drawback of these simulations is their computational costs. They usually take hundreds or thousands of CPU hours. Comparing the volumes of observed surveys to such simulations is possible, as long as one does not vary cosmological parameters. Once they come into play, an N-body per value of each cosmological parameter is needed and we are not able to perform the fit, minimization, MCMC, ... any kind of parameter estimation you can think of. For this reason, a very extended practice nowadays is the use of emulators. Emulators are approximated methods to get the same results as full numerical solutions but in a fraction of the time. The most commonly used in cosmology are those based on Neural Networks (NN). The idea is very simple. You run a bunch of N-body simulations with a variety of cosmological parameters. Then, you use them as training data for a feed-forward NN. As a result, you end up with a glorified interpolator that, given some cosmological parameters, returns N-body-like results. Again in figure 2 we find an example of an emulator that interpolated between the displacement field of N-body simulations. The small-scale structure is recovered at a fraction of the computational cost. See figure

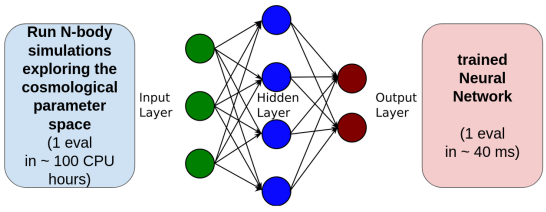

Figure 3: Schematic picture of the procedure to build a typical emulator.

3 for a schematic view.

1.2 We do not have access to the Initial Conditions of the Universe directly so we use statistics

Because we cannot observe the past evolution of our surrounding volume, we do not know the initial conditions, $\delta(\vec{x}, 0)$ and $\vec{v}(\vec{x}, 0)$, that led to the structure we see today. In general, unless we are working with constrained simulations (and even so), we work with statistical quantities that are able to summarise the cosmological dependencies independently of the concrete realization one chooses as the initial field. It seems reasonable that the statistics we get by comparing different parts of a large volume should be similar to the statistics of a given part over different realizations, i.e., they provide a fair sample of the probability distribution. This property is called ergodicity. Fields that satisfy $\bar{f} \to \langle f \rangle$ when $V \to \infty$ are called ergodic. We assume that the density field is ergodic. The equality of $\bar{f}$ with $\langle f \rangle$ does not hold for a finite volume V. The difference is called cosmic variance if we are thinking about the finite volume of the observable Universe. It is called sample variance if the restrictions are related to some finite volume survey geometry. The larger the volume, the smaller the difference. Since cosmological theory predicts $\langle f \rangle$, whereas observations probe $\bar{f}$ for a limited volume, cosmic variance limits how accurately we can compare theory with observations.

1.2.1 Correlation function

From statistical homogeneity, $\langle \rho(\vec{x}) \rangle = \langle \rho \rangle$ and so $\langle \delta \rangle = 0$. Thus, we cannot use $\langle \delta \rangle$ as a measure of homogeneity. Instead, we can use the squared of δ. Its expectation value $\langle \delta^2 \rangle$ is the variance of the density perturbation. It tells us about how strong the inhomogeneity is, but nothing about the shapes or sizes of the inhomogeneities. To get this information we move to the correlation function $\xi(\vec{x}_1, \vec{x}_2) \equiv \langle \delta(\vec{x}_1)\delta(\vec{x}_2) \rangle$. Note that $\xi(\vec{x}_1, \vec{x}_2)$ is positive if the density perturbation has the same sign at both $\vec{x}_1$ and $\vec{x}_2$, and negative when there is overdensity at one and underdensity at the other. Hence it shows how different density perturbations are correlated. Because of homogeneity, ξ can only depend on the distance $\vec{r} = \vec{x}_2 - \vec{x}_1$, and because of isotropy it can only depend on its

modulus r. Thus we can rewrite

$$\xi(r) = \langle \delta(\vec{x})\delta(\vec{x}+r)\rangle \,. \tag{25}$$

This describes the two-point statistics of the sample. For the case of observations, a nice (but not completely accurate) way of thinking about this quantity is in terms of probabilities. Pick a random object in your sample and compute what is the excess of probability of finding another object at a distance r from you. If the sample is random then this probability is a constant and the over-probability is zero (so is $\xi(r)$). In the gravitational case, there is a higher probability of finding objects closer to you since gravity acts as an attractive force. Therefore, one would expect a peak next to your position and a decrease for further distances (until we reach the BAO peak which I won't talk about here).

1.2.2 Power spectrum

We can also study statistics in Fourier space. The advantage of working in Fourier space is that, in the linear regime, different modes are decoupled thus evolving independently. We now expand the density perturbation as a Fourier series

$$\delta(\vec{x}) = \sum_k \delta_k e^{i\vec{k}\cdot\vec{x}} \,, \tag{26}$$

where $\delta_k = \frac{1}{V}\int_V \delta(\vec{x}) e^{-i\vec{k}\cdot\vec{x}} d^3x$. In analogy with the correlation function $\xi(r)$, we may ask what is the corresponding correlation in Fourier space, $\langle \delta_k^* \delta_{k'}\rangle$, where superscript * indicates the complex conjugate. The answer is the power spectrum, $P(k)$,

$$\langle \delta_k^* \delta_{k'}\rangle = \frac{1}{V}\delta_{\rm D}(\vec{k}-\vec{k}') \int d^3r \xi(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} \equiv \frac{1}{V}\delta_{\rm D}(\vec{k}-\vec{k}') P(\vec{k}) \,, \tag{27}$$

with δ_D being the Dirac delta.

We show in figure 4 examples of power spectra. Both the linear approximation and the N-body result are shown.

1.3 Galaxies are tracers of the cosmic web

A key extra ingredient needed when matching our theories with observations is the fact that we only observe luminous objects. Thus we are only able to see tracers emitting light of the underlying density field. Bias models describe, in a statistical sense, the relationship of the distribution of these objects to that of matter. To better understand how this affects the statistical quantities we introduced (power spectrum, correlation function, etc), it is helpful to consider the following example. Because of gravity, there is a higher chance of baryonic gas falling into the deepest potential wells. In this case, let us assume luminous objects such as galaxies will only form in collapsed regions. As a consequence, when observing these galaxies, they will be tracing only the collapsed regions,

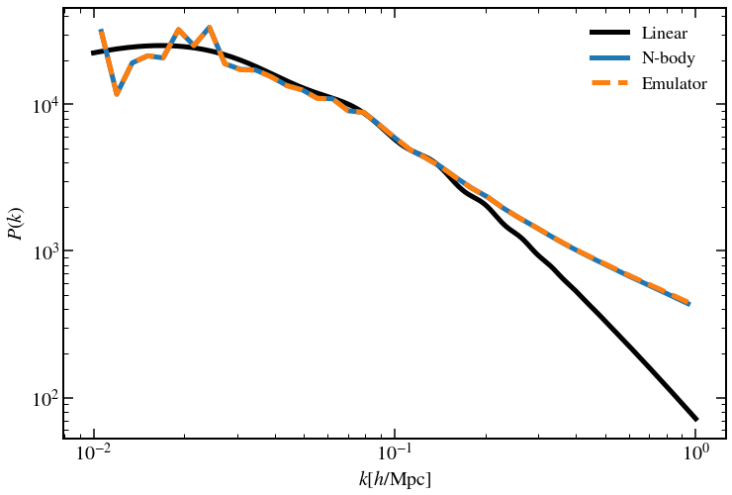

Figure 4: Power spectra of previously shown linear approximation, $N-$body simulation, and emulator. Large scatter on large scales (small k) are due to cosmic variance. Discrepancies on small scales (large k) come from larger clustering of small structures from non-linear evolution captured by the numerical solution and not by the perturbative approach.

i.e, the highly clustered parts of the cosmic web (such as knots). Because the clustering in those regions is bigger, we are systematically choosing higher-density parts of the cosmic web. In a naive way, we can model this with a constant bias b_{L} so that the field that we are actually measuring is $\delta_{\mathrm{tr}} = b_{\mathrm{L}}\delta_m$, where δ_{tr} is the overdensity field of the tracers, δ_m is the dark matter field and b_{L} is the so-called "linear" bias (although a better name would be large scale bias). Moreover, from the definition of the power spectrum

$$P_{\mathrm{tr}}(k) = b_{\mathrm{L}}^2 P_m(k)\,, \tag{28}$$

which tells us that the amplitude of the power spectrum is higher for the tracers than for the underlying dark matter. We already expected this since the clustering was higher for these objects than for the full field.

With this example, we have shown that even with very basic assumptions there is a bias parameter on the measured statistics of the Universe. Of course, the relation between each kind of tracer and the matter field will be different. In reality, this bias is much more involved.

1.4 Galaxy position measurements are affected by velocities

An extra key ingredient for the step between theory and observations is the redshift space distortions in our measurements due to the peculiar velocity of

tracers. Redshift space distortions (RSDs) are important because we measure any cosmological quantity only as a function of redshift space distance. For any cosmological observable involving radial distances (or averages over them), we should take RSDs into account. RSDs are implicitly measuring the cosmological velocity field determined through the gravitational potential. To study this effect, we will take the RSD transformation, compute its jacobian and impose conservation of mass.

1.4.1 Redshift space transformation

When mapping out objects such as galaxies in the 3-dimensional space, the radial distance is determined by its measured redshift z. However, we should be aware of the existence of two contributions, the Hubble flow $x(z_{\cos}) = \int_0^{z_{\cos}} c\,dz/H(z)$ and the peculiar velocity of the object,

$$z = Hx + \frac{\vec{v}}{a}\cdot\hat{x} \underset{1/H}{\Longrightarrow} s = x + \frac{\vec{v}\cdot\hat{x}}{aH}\,. \tag{29}$$

The first term is the Hubble law, which states that the cosmological redshift is proportional to the distance; the second is the velocity along the line of sight. The second term is often ignored in astronomy. However, the second term has a non-negligible effect on the clustering statistics of the matter density field. Usually, we would be able to write the power spectrum $P(\vec{k})$ as a scalar function $P(k)$ thanks to its rotational invariance. However, this does not hold in the case of the observed power spectrum in redshift space, simply because the peculiar velocity term breaks down rotational invariance.

Density field in redshift space

Taking into account that the number of galaxies is invariant and so is the mass (always in the Newtonian approximation),

$$\bar{n}\,(1+\delta^s(\vec{s}))\,d^3s = \bar{n}\,(1+\delta(\vec{x}))\,d^3x\,, \tag{30}$$

where $\delta(\vec{x})$ at $\vec{x}$ is in real space, $\delta^s(\vec{s})$ at $\vec{s}$ is in redshift space are the overdensities, and $\bar{n}$ is the mean density of galaxies. Hence

$$1+\delta^s(\vec{s}) = (1+\delta(\vec{x}))\,J\,, \tag{31}$$

as the angular volume elements are the same. Here the Jacobian J is given by

$$J = \left\|\frac{d^3x}{d^3s}\right\| = \frac{dx}{ds}\frac{x^2}{s^2}\,, \tag{32}$$

describing the redshift space transformation from equation 30. The Jacobian becomes

$$J = \left(1+\frac{\partial}{\partial x}\left[\frac{\vec{v}\cdot\hat{x}}{aH}\right]\right)^{-1}\left(1+\frac{\vec{v}\cdot\hat{x}}{xaH}\right)^{-2}\,. \tag{33}$$

The partial derivative term is much more important than the second term of the product. The first term would be proportional to $\vec{k} \cdot \vec{v}/aH$, thus the difference between these would be roughly of a factor $\vec{k} \cdot \vec{x}$. While $\vec{x}$ is of the order of the size of the survey, $\vec{k}$ can be small. Nevertheless, the number of modes satisfying a small $\vec{k}$ is very low because they are of the size of the survey. On top of that, we are averaging over single modes. This means that the weight of low $\vec{k}$ modes is smaller, leading to $\vec{k} \cdot \vec{x} \gg 1$. Taking this into account, and expanding equation 33 about $v = 0$, we obtain

$$J \simeq \left(1 - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{aH}\right]\right). \tag{34}$$

Substituting in equation 31 we get

$$1 + \delta^s(\vec{s}) = (1 + \delta(\vec{x}))\left(1 - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{aH}\right]\right) \Rightarrow \delta^s(\vec{s}) \simeq \delta(\vec{x}) - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{x}}{aH}\right], \tag{35}$$

The distant observer approximation assumes that $\hat{x}$ points in the same direction as $\hat{z}$. In the context of this approximation, we can Fourier transform

$$\begin{aligned}
\tilde{\delta}^s(\vec{k}) &= \int d^3x e^{-i\vec{k}\vec{x}}\left(\delta(\vec{x}) - \frac{\partial}{\partial x}\left[\frac{\vec{v} \cdot \hat{z}}{aH}\right]\right) \\
&= \tilde{\delta}(\vec{k}) - if\int d^3x e^{-i\vec{k}\vec{x}}\frac{\partial}{\partial x}\left[\int \frac{d^3k'}{(2\pi)^3}e^{i\vec{k}'\vec{x}}\tilde{\delta}(\vec{k}')\frac{\vec{k}'}{k'^2} \cdot \hat{z}\right] \\
&= \tilde{\delta}(\vec{k}) + \int \frac{d^3k'}{(2\pi)^3}\tilde{\delta}(\vec{k}')f(\hat{k}' \cdot \hat{z})^2 \int d^3x e^{i(\vec{k}'-\vec{k})\cdot\vec{x}} \\
&= \left[1 + f\mu_k^2\right]\tilde{\delta}(\vec{k})
\end{aligned} \tag{36}$$

where we introduced in the first step the linear relation $\vec{v}(\vec{k}) = ifaH\delta(\vec{k})\frac{\vec{k}}{k^2}$ from 24 (this relation is then only valid in linear scales) and in the last step the Dirac delta definition. Here μ_k is defined as the cosine of the angle between the line of sight and the wavevector $\hat{k}$. The term multiplying $\tilde{\delta}(\vec{k})$ is the so-called "Kaiser factor".

Power spectrum in redshift space

Suppose we measure the power spectrum in redshift space. How is this distorted spectrum related to the underlying spectrum in real space? As a consequence of equation 36 and by definition of power spectrum

$$P^s(\vec{k}) = \left[1 + f\mu_k^2\right]^2 P(k). \tag{37}$$

The Kaiser factor here stands for a very simple physical effect. At large scales, where we can assume linear behavior, objects tend to coherently infall into high-density regions. Because closer objects in the line of sight have velocities pointing away from us and those further have velocities toward us we measure

them as if they were further and closer respectively. This has the effect of squashing the density field and the clustering amplitude becomes stronger along the line of sight. This is the so-called Kaiser effect.

We can relate now this purely dark matter power spectrum to the galaxy (tracers) one using the linear relation $\delta_g = b\delta$ (where we dropped the subscript L from the previously introduced $b_{\rm L}$). For these linear regimes in which there is a linear bias between our tracers and the underlying dark matter field, the velocity field is sensitive to the mass overdensity $\vec{v} \propto \delta \propto \delta_g/b$. We can easily take the constant bias factor out in the second term of 36 and rewrite

$$P_g^s(\vec{k}) = b^2\left[1+\beta\mu_k^2\right]^2 P(k). \tag{38}$$

where $\beta = f/b$. It is important to realize that the anisotropic term originates from the velocity and, consequently, it does not depend on the bias. This is the reason why the RSD measurement is often parametrized by the amplitude of the peculiar velocity field, $f\sigma_8(z)$.

Let us expand the power spectra in its multipoles using Legendre polynomials,

$$P^s(\vec{k}) = \sum_\ell P_\ell^s(k)\mathcal{P}_\ell(\mu_k)\,, \tag{39}$$

where $\mathcal{P}_\ell$ are the Legendre polynomials and

$$P_\ell^s(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu P^s(k,\mu)\mathcal{P}_\ell(\mu_k)\,. \tag{40}$$

Since the Kaiser formula only contains terms up to μ^4, only $\ell = 0$ (monopole), 2 (quadrupole), and 4 (hexadecapole) are non-vanishing as in odd poles the integrand of equation 40 becomes symmetric and gives zero. Expanding

$$\left[1+\beta\mu_k^2\right]^2 = \left[1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right]\mathcal{P}_0(\mu_k) + \left[\frac{4}{3}\beta+\frac{4}{7}\beta^2\right]\mathcal{P}_2(\mu_k) + \frac{8}{35}\beta^2\mathcal{P}_4(\mu_k), \tag{41}$$

we can see right away how β enters in the different multipoles of the power spectrum

$$P^s_{g,\ell=0}(k) = \left[1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right] b^2 P(k)\,, \tag{42}$$

$$P^s_{g,\ell=2}(k) = \left[\frac{4}{3}\beta+\frac{4}{7}\beta^2\right] b^2 P(k)\,, \tag{43}$$

$$P^s_{g,\ell=4}(k) = \frac{8}{35}\beta^2 b^2 P(k)\,. \tag{44}$$

In this exercise we will only use equation 42, but, if you feel brave enough, feel free to use the three multipoles.

This is everything we will need from the theoretical point of view. We will use equation 42 to compare against the data we will provide. For that comparison, we will use an already implemented Markov Chain Monte Carlo method (MCMC). We will briefly describe what MCMCs are and explain the exercise.

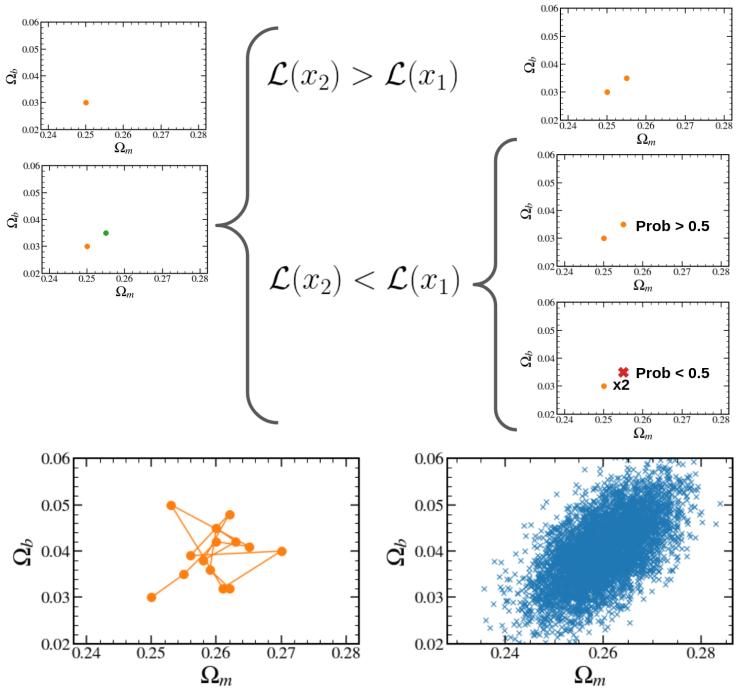

Figure 5: Schematic view on how an MCMC works. The symbol $\mathcal{L}$ stands for the likelihood function. Therefore $\mathcal{L}(x_1)$ is the likelihood at point x_1 (orange point) and $\mathcal{L}(x_2)$ is the likelihood at point x_2 (green point). The green point is the trial point. The orange points are already accepted points. Red cross means that the point will not be accepted and a new random point will be tested.

1.5 Markov Chain Monte Carlo (MCMC)

We will assess here the procedure of inference that is commonly used for parameter constraining. For a schematic view of how the algorithm works, see figure 5.

From the bayesian point of view, there is no distinction between observables and parameters of a statistical model; all of them are considered "unknown" quantities and they can be associated with a probability. Let us define D as the observed data and θ as the model parameters. Formally, one can build a joint probability $P(D, \theta)$ over the unknown parameters. This joint probability is formed by two bits, a prior knowledge $P(\theta)$ and a likelihood function $P(D|\theta)$. This likelihood shows how probable is getting a result after the data are known. Specifying $P(D, \theta)$ and $P(\theta)$ gives a complete probabilistic model, where $P(D, \theta) = P(D, \theta)P(\theta)$. Having observed D, one can use Bayes' theorem

to determine the distribution of θ conditioned to D,

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta}\,, \tag{45}$$

which is called the posterior distribution. Thus, the expected values of any function $f(\theta)$ can be expressed as,

$$E\left[f(\theta), D\right] = \frac{\int f(\theta)P(\theta)P(D|\theta)d\theta}{\int P(\theta)P(D|\theta)d\theta}\,. \tag{46}$$

This very simple expression encodes a very complicated computation. The integral is highly dimensional and very complicated to solve in general.

Markov Chain Monte Carlo methodologies were invented in order to deal with equations of the form of 46. They are composed of two parts, the Monte Carlo integration, and the Markov Chain.

The Monte Carlo integration evaluates the expected values $E\left[f(X)\right]$ taking samples $\{X_t; t = 1, \cdots, n\}$ from the function $P(\theta|D)$ and then approximates,

$$E\left[f(X)\right] \approx \frac{1}{n}\sum_{1}^{n} f(X_t)\,. \tag{47}$$

Therefore, the mean of the population is estimated with the mean of the sample. When the samples $\{X_t\}$ are independent the bigger the n, the more accurate this approximation will be. Note that n is not the number of data points but the size of the sample.

Let us assume that we generate a sequence of random variables $\{X_0, X_1, X_2, \cdots\}$ in such a way that, for each $t \geq 0$, the following variable X_{t+1} is extracted from a distribution $P(X_{t+1}|X_t)$ that only depends on the current variable X_t. This sequence is called a Markov Chain and $P(X_{t+1}|X_t)$ is the transition kernel of the chain. The chain will lose the information coming from the first step X_0, and the distribution from which the variables are sampled will tend to a stationary distribution $\phi(X_t)$.

After m steps far from $\phi(X_t)$, called the "burn in" steps, the points $\{X_t; t = m+1, \cdots, n\}$ will be sampled from the distribution $\phi(X_t)$. After the burn in, the estimator $E\left[f(X)\right]$ becomes,

$$E\left[f(X)\right] = \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t)\,. \tag{48}$$

It is necessary to check that the chain is such that its stationary distribution $\phi(X_t)$ coincides with the posterior $P(\theta|D)$. One possibility is to use the Metropolis-Hastings algorithm.

For this algorithm, at each step t, the following step X_{t+1} is chosen first sampling a candidate point Y from a proposed distribution $q(Y|X_t)$. For the candidate point Y one computes a probability $\alpha(X_t, Y)$ such that

$$\alpha(X, Y) = \min\left(1, \frac{\mathrm{P(Y)q(X|Y)}}{\mathrm{P(X)q(Y|X)}}\right)\,. \tag{49}$$

At the same time, a random number is chosen in a range from 0 to 1. Then this number is compared with $\alpha(X_t, Y)$, if it is lower, one takes $X_{t+1} = Y$, otherwise, $X_{t+1} = X_t$. Finally, one increases t. Note that the proposed distribution q can have any shape while the stationary distribution is $P(\theta|D)$.

The question that arises is whether the simulated Markov chain has fully explored the target posterior distribution or whether we need longer simulations. The so-called Gelman-Rubin diagnostic evaluates MCMC convergence by analyzing the difference between multiple Markov chains. The convergence is assessed by comparing the between-chains and within-chain estimated variances for each model parameter. Large differences between these variances indicate non-convergence. It is based on the notion that, if multiple chains have converged, by definition they should appear very similar to one another; if not, one or more of the chains has failed to converge.

The Gelman-Rubin diagnostic uses an analysis of variance approach to assess convergence. It calculates both the between-chain variance B and within-chain variance W and tests whether they are different enough to worry about convergence. Assuming m chains, each of length n, variances are calculated by

$$\begin{aligned} B &= \frac{n}{m-1} \sum_{j=1}^{m} \left(\hat{\theta}_j - \hat{\theta}\right)^2 , \\ W &= \frac{1}{m} \sum_{j=1}^{m} \left[\frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{\theta}_{ij} - \hat{\theta}_j\right)^2 \right] , \end{aligned} \tag{50}$$

for each model parameter θ. Using these values, two estimates of the posterior variance of θ can be calculated as

$$\widehat{\mathrm{Var}}(\theta) = \frac{n-1}{n} W + \frac{1}{n} B \,. \tag{51}$$

Although this estimator would be unbiased if the initial points were extracted from $P(\theta|D)$, it is most probable that θ was initialized to arbitrary starting points in each chain. Then this quantity will overestimate the true marginal posterior variance. On the other hand, W will tend to underestimate the within-chain variance early in the sampling run. In the limit of $n \to \infty$, both quantities will converge to the true variance of θ. Thus, the Gelman-Rubin statistic monitors convergence using the ratio

$$\hat{R} = \sqrt{\frac{\widehat{\mathrm{Var}}(\theta)}{W}} \,. \tag{52}$$

It is usually accepted that a value of $\hat{R} - 1 < 0.1$ indicates that the chain is converged.

2 Exercise

Let's start with the real stuff. Most of the exercise is clearly stated in the notebook and I will guide you through it during the lecture, so no worries if you

find yourself a bit lost while reading this. We will make use of several codes, namely:

- `Python notebook`: if you have never used a python notebook before, they are super cool. You will get used to them in no time. They are like running python but in an interactive way, plus you can write equations and text there, which makes things more beautiful.
- `baccoemu`: this is a public tool made to compute the key theory ingredients we have presented in the last section. Mainly, the power spectrum of the matter content. It is based on neural emulators and we will use it to compute most of the theory quantities.
- `emcee`: this is a public code that makes affine Monte Carlo integrations. For our purposes, we will think about it as an MCMC solver.

Installing all of these packages might be a bit tedious and difficult. That's why we have created a Docker so that you just need to install it and everything else is there for you to use. The only drawback is that the Docker takes 6Gb of disk space. Apologies about this, but it was the simplest way. As the main plan, we will run everything in `Google colab`, and you won't have to install anything, but we cannot guarantee that we will all have access to a good internet connection. Therefore, installing the docker on your computer is something that we will really appreciate as a backup plan. Anyways, let me write down a few instructions to download and install the Docker:

Instalation

1. Windows

1.1. Install Windows terminal

https://docs.microsoft.com/en-us/windows/terminal/get-started

It is not compulsory but having all terminals centralized makes things simpler.

1.2. Instal WSL2

1. Enable bios virtualization, the steps are a bit dependent on each computer.
2. In Start “Activate or deactivate Windows characteristics” and activate “Windows Subsystem for Linux” and “Virtual machine platform”.
3. Restart the computer.
4. Update Linux Kernel: https://wslstorestorage.blob.core.windows.net/wslblob/wsl_update_x64.msi Execute as administrator if there is any problem.
5. Download Ubuntu 20 from Microsoft Store and configure the user and password. One can add Ubuntu as a preferred terminal in Windows Terminal.
6. Check that the wsl version is 2: in PowerShell `wsl -l -v`
7. If it is not version 2, we update https://learn.microsoft.com/es-es/windows/wsl/install-manual: `wsl --set-version <Ubuntu:verion> 2`
8. If the wsl state is stopped, start it with the command `wsl`.

3. Instal docker en WSL2

1. Download docker desktop for Windows: https://docs.docker.com/desktop/install/windows-install/
2. Enable Linux containers:
    1. Select Docker in the toolbar (down right) > Right click > Switch to Linux containers…
3. Enable “WSL Integration” in Docker desktop > Settings > Resources > WSL Integration
4. To check that it works, run in Windows terminal with Ubuntu: `docker image ls`

2. Ubuntu

Follow the steps:

https://docs.docker.com/engine/install/ubuntu/

Run docker

1. Windows:

1. Copy the docker-compose.yml file to the “work” folder (the folder where you will be working from) inside WSL. As an example, in the path: `\\wsl$\Ubuntu\home\user\exercise`
2. Go to Ubuntu’s terminal previously created.
3. Run: `sudo docker-compose up`
4. Control+C to stop

2. Ubuntu:

1. Copy the docker-compose.yml file to the “work” folder (the folder where you will be working from).
2. Run: `sudo docker-compose up`
3. Control+C to stop

$\Omega_{\rm m}$	σ_8	$\Omega_{\rm b}$	n_s	h	m_ν	w_0	w_a	a
0.307	0.8	0.048252	0.9611	0.6777	0	-1	0	0.6355

Table 1: Table summarising the cosmology assumed for the galaxy mock.

Once you run the Docker, an automatic window should have opened with the exercise notebook. There, you will find the exercise explained step by step. Just follow the notebook and fill in the places where the questions are asked. Here we will go through the exercise showing what the results should look like. Again, do not worry if you are lost, we will provide the solutions at the same time as the exercise is done. You can always have a look at them to get inspired!

2.1 Data

The data we will use is the monopole of a realistic mock survey. This data has been computed following the procedure of the work appearing in [3]. In a very simplified way, we created these data by running a simulation, funding the collapsed regions (halos and subhalos), and populating those with galaxies with a prescription that has been shown to reproduce the clustering of hydrodynamical simulations. The painted galaxies reproduce the clustering of Luminous Red Galaxies (LRG's). These are the kind of galaxies observed in the BOSS collaboration. In principle, we could have used the public data from the mentioned collaboration. However, dealing with observational systematics is beyond the scope of this exercise, and having access to the cosmology used to create the galaxy sample will be of utmost importance. Find in Table 2.1 the cosmological parameters used for the creation of the mock. Find in the left panel of figure 6 the computed monopole of the power spectrum. The volume of the mock is 1440 Mpc/h and its number density $\bar{n} = 3.16 \times 10^{-3}[h/\text{Mpc}]^3$. As an aside, note that the number density is an order of magnitude larger than the BOSS sample. We chose this number density so that there were no issues with the shot-noise contribution. The error bars provided correspond to those of the Gaussian approximation. This approximation is very accurate on large scales but fails to describe small scales. We will use it as a proxy for the correct one since there is no definitive answer yet as to how to treat the errors on small scales.

The first step of the exercise is to learn to read the data and plot it.

2.2 Models

We will use two models to test the cosmological content of the data. Both of them are provided by the public tool `baccoemu` (see [1]). The models under study are:

- **Linear model**: As we saw, this model can be solved "analytically" given some Hubble function $H(t)$. Of course, getting solutions is way more

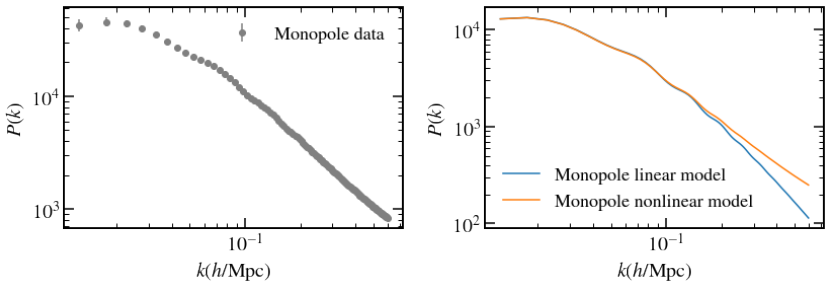

Figure 6: Monopoles of the galaxy mock, the linear and the non-linear predictions.

complicated in practice and one needs to solve the Boltzmann equation for different species. Two very well-known codes that do this are CAMB and CLASS (see [5] and [2] respectively). Even though they are very fast in their computations, getting the results we seek, would take an amount of time we do not have. That's why we will make use of the bacco emulator. The only thing this emulator does is interpolate between the predictions computed with the previous codes. For this, it uses a neural network that makes predictions extremely fast. For any application, this is like using CLASS or CAMB, but faster.

- **Non-linear model**: As already discussed, full solutions to structure formation in the Universe can only be achieved through numerical simulations. This is what we will use here as the nonlinear model, solutions to the power spectrum through N-body simulations. Now, running an N-body simulation at the volumes we need for the data provided, takes an insane amount of CPU time. Therefore, we will use the BACCO emulator once again. This was trained over a set of 4000 N-body simulations to give the non-linear power spectrum in a matter of microseconds.

The second step of the exercise is to learn how to call and represent the theoretical predictions. The results will look like the right panel of 6. Don't be shy and feel free to play around with the cosmological parameters to see their effect on the predicted power spectra.

2.3 Model comparison

Once we have some experience with the data and models, it is time to compare them. Naively, the first thing anyone would do is simply plot one vs the other to see the similarities and differences. If one does this, the plot would look like the left panel of figure 7. It becomes clear that there is a difference in the amplitude between the data and the models. Why is that?

From equation 42 it seems that we missed two effects, the fact that galaxies are tracers of the full matter distribution (predicted by our models), and the

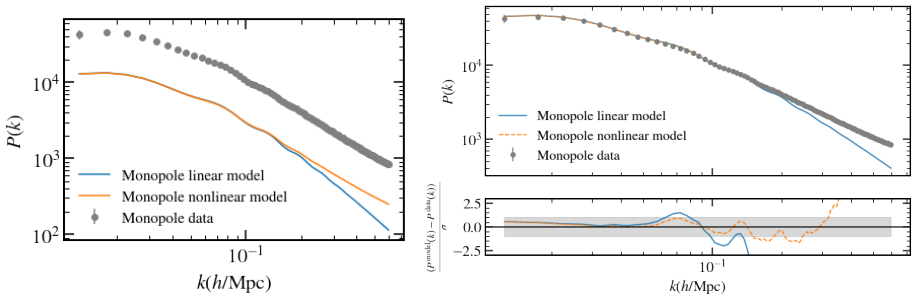

Figure 7: Comparison between data and theory with and without bias and RSDs.

effect of velocities in the line of sight of our measurements i.e. Redshift Space Distortions. Following this equation, we have to include two free parameters, b and β. Note that β is not independent of cosmology and it actually has a functional form completely defined by the growth, we will come back to this later.

2.3.1 Find the parameters b and β such that the models fit reasonably well the data

Starting with the predictions from `baccoemu`, write code up equation 42 and play around with the free parameters b and β until you find, by eye, a fit that suits your sophisticated taste. Note that the models are designed to fit large scales (small k), so our advice would be not to waste time trying to fit the small scales. The solution would look something like figure 7.

2.3.2 Parameter constraints with the linear model

This was just a warm-up. Needless to say, no one fits the theory to data by eye. We will use more powerful techniques to find out the parameter constraints. Concretely, we will make use of the Markov Chain Monte Carlo implemented in `emcee` (see [4]). You will find an example of how to use it in the notebook, together with some functions that we included that make its use simpler.

Write a function to compute the χ^2 of the linear model down to a given $k_{\rm max}$ and with free Ω_m (called omega_cold in params), b and β.

The next step is to write down a function to get the χ^2 of a given data, model, and data variance. Remember that the χ^2 is defined in relation to the gaussian likelihood function $\mathcal{L} \propto \exp(-\chi^2/2)$:

$$\chi^2 = (\mathcal{X}_i - \mathcal{X}_{i,\rm theo})^{\rm T} M_{ij} (\mathcal{X}_j - \mathcal{X}_{j,\rm theo})\,. \tag{53}$$

Here $^{\mathrm{T}}$ denotes the transpose, $\mathcal{X}$ are the data taken, $\mathcal{X}_{\rm theo}$ are the theoretical predictions depending on the cosmology and $M = C^{-1}$ is the inverse of the covariance matrix, defined as the matrix whose element in the i, j position is a measure of the joint probability of two variables between the i-th and j-th elements. Concretely,

$$C_{ij} = \mathrm{cov}(\mathrm{X_i}, \mathrm{X_j}) = \mathrm{E}\left[(\mathrm{X_i} - \mathrm{E}(\mathrm{X_i}))\,(\mathrm{X_j} - \mathrm{E}(\mathrm{X_j}))\right] , \tag{54}$$

where X is a vector of random entries and E[ ] refers to the expected (mean) value. Note that from this definition, the elements of the diagonal correspond to the error in the measurement. This is our case here, where I did not provide any non-diagonal terms in the covariance. Thus we can simplify the previous equation to

$$\chi^2 = \sum_i \frac{(\mathcal{X}_i - \mathcal{X}_{i,\mathrm{theo}})^2}{\sigma_i^2} = \sum_i \frac{(P_0^{\mathrm{data}}(k_i) - P_0^{\mathrm{model}}(k_i))^2}{\sigma^2(k_i)} . \tag{55}$$

And this is the exercise, write a function to compute the χ^2 of the linear model down to a given $k_{\rm max}$ and with free Ω_m (called omega_cold in params), b and β.

To check that we have defined a sensible χ^2 function I usually plot the values of χ^2 computed for an array of values of a free parameter fixing the others. If the plot free parameter vs chi2 looks like a parabola, then there are good chances you did it right.

Use this function to constrain the parameters down to $k_{\rm max} = 0.1$ using `emcee`.

Use the χ^2 function that you defined in the previous section to run a Markov Chain down to $k_{\rm max} = 0.1$. What do you find? Plot the best fit (an example of how to read it from `emcee` is in the notebook). Is the fit a good one? Plot the chains with `corner` (again, there is an example in the notebook). Is $\Omega_{\rm m}$ well recovered?

The solutions would look something like the upper panels in figure 8. Would you say the model made a good job constraining the parameters down to this $k_{\rm max}$ scale?

Do the same as in 2.2 for several values of $k_{\rm max}$ until you see that the model stops working.

After trying several values of $k_{\rm max}$ you will likely find something like the lower panels of figure 8. Once the model tries to fit scales smaller than $k_{\rm max} > 0.1$, we find biases on the recovered values of $\Omega_{\rm m}$ and very bad fits. This means that the model is not able to explore smaller scales than $k_{\rm max} = 0.1$. Is this a big deal? We will try to answer this in the next question.

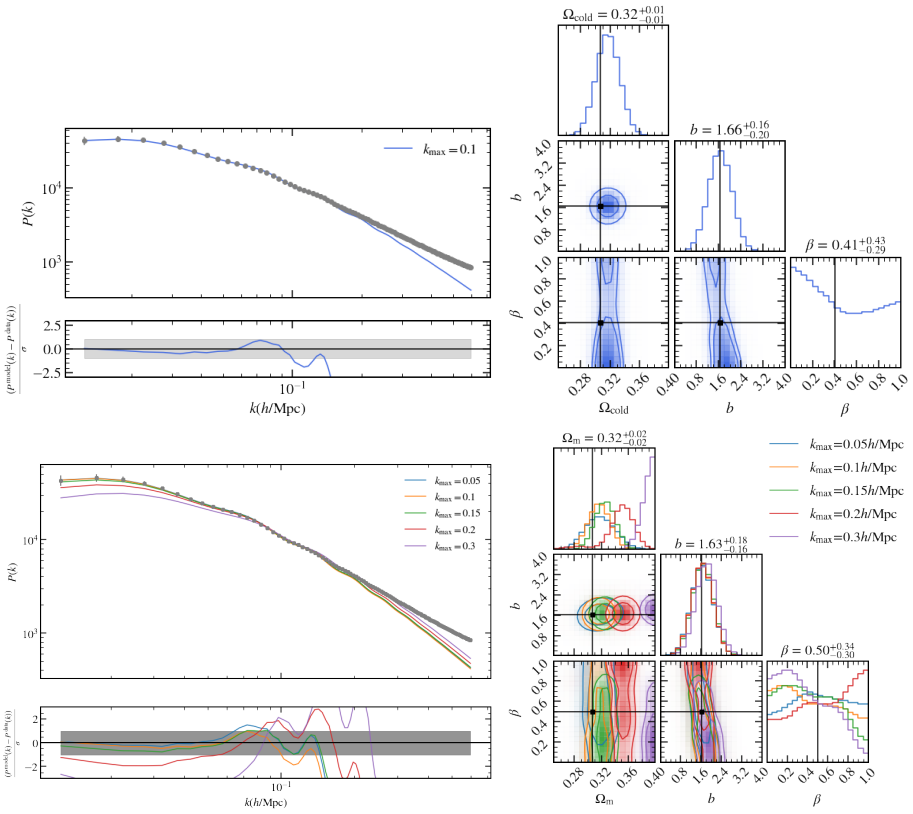

Figure 8: Comparison between data and linear theory with and without bias and RSDs.

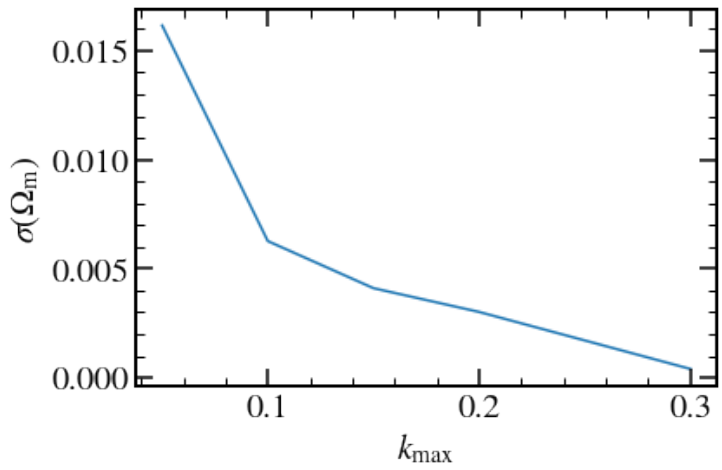

Figure 9: Variance of the cosmological parameter Ω_m as a function of scale for the linear model.

Plot the variance found in Ω_m as a function of scale to see if including extra scales increases the accuracy of the measurements.

Does including extra scales increase the accuracy of our measurements? If the answer is yes, then it would make sense to try to get a model that also explores these scales. Otherwise, we can simply use linear approximations and focus on something more interesting. By studying the variance of Ω_m we can have an idea of the gain we might get in a real survey. So let's plot this quantity!

You can find an example of doing so in figure 9. Even though we cannot trust the results to lower scales than 0.15, there is a huge decrease in variance thanks to the incorporation of more data. This is great news! It means that, by making a more complete model, we can make better constraints on cosmological parameters.

2.3.3 Parameter constraints with the non-linear model

Repeat the process we have gone through but substitute the linear model with the emulator of N-body results and discuss what are the changes you see with respect to the previous results. You will probably find very similar results to those appearing in figure 10.

Discuss how the different models compare. Is one better behaved than the other? Is there a compelling reason to try to make models that explore smaller scales than the linear model? What if the errorbars would shrink, would the nonlinear model presented here still work?

ATTENTION: Since the model we are using for structure formation is non-linear, we should in principle also consider non-linear models for both the RSDs

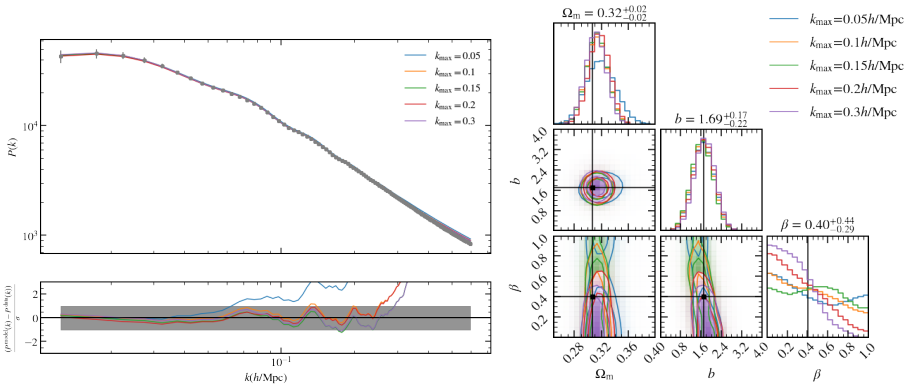

Figure 10: Comparison between data and non-linear theory with and without bias and RSDs.

and the bias model. However, these are complicated models cosmologists are still working on and are beyond the scope of this lecture.

If we plot the same variance versus scale plot (do it), what one finds is something similar to what is shown in figure 11. Note that there is still an increase in constraining power from small scales in this case. Since we are not biased at these scales, it is safe to say that the model benefits from exploring these scales and thus, should be included in our analysis (as long as they do not break other tests we could put the model through).

2.4 Extra credit

We have assumed β to be a free parameter, but it is not. Check that all the β values you have found are consistent with the theory prediction.

In order to compute the expected β prediction by the theory from the outputs of `baccoemu`, we can use the relation,

$$D(a) = \sqrt{\frac{P_{\rm lin}(k,a)}{P_{\rm lin}(k,a=1)}}. \tag{56}$$

Since the growth factor is defined as $f = d\ln D/d\ln a$, and $\beta = f/b$, we can easily recover the expected value of the growth factor for a given cosmology. If we compute it with the true cosmology and the value is consistent with what we found in the chains, no problem. Otherwise, we might have pushed the model too much. The value you will likely find if you do the exercise will be around $\beta \approx 0.48$.

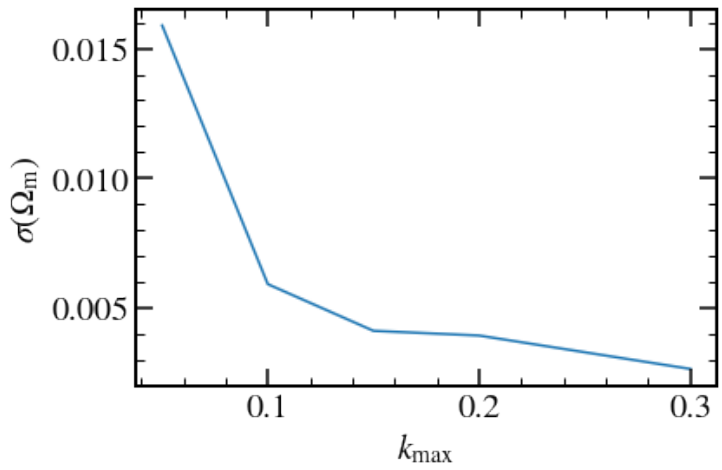

Figure 11: Variance of the cosmological parameter Ω_m as a function of scale for the non-linear model.

References

[1] Raul E. Angulo et al. "The BACCO simulation project: exploiting the full power of large-scale structure for cosmology". In: 507.4 (Nov. 2021), pp. 5869–5881. DOI: 10.1093/mnras/stab2018. arXiv: 2004.06245 [astro-ph.CO].

[2] Diego Blas, Julien Lesgourgues, and Thomas Tram. "The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes". In: 2011.7, 034 (July 2011), p. 034. DOI: 10.1088/1475-7516/2011/07/034. arXiv: 1104.2933 [astro-ph.CO].

[3] S. Contreras, R. E. Angulo, and M. Zennaro. "A flexible subhalo abundance matching model for galaxy clustering in redshift space". In: 508.1 (Nov. 2021), pp. 175–189. DOI: 10.1093/mnras/stab2560. arXiv: 2012.06596 [astro-ph.CO].

[4] Daniel Foreman-Mackey et al. "emcee: The MCMC Hammer". In: 125.925 (Mar. 2013), p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

[5] Antony Lewis and Sarah Bridle. "Cosmological parameters from CMB and other data: A Monte Carlo approach". In: 66 (2002), p. 103511. DOI: 10.1103/PhysRevD.66.103511. arXiv: astro-ph/0205436 [astro-ph].