X. Non-linear clustering, halos, and galaxies

We will touch lightly on some voluminous topics.

Some of this material is covered in §§9.5-9.7 of Huterer.

I have a much more extensive set of lecture notes available from my home page under the "ICTP Advanced Cosmology School" bullet. Lecture 2 is particularly relevant to this section. There are many suggestions for further reading in there.

Some other references are suggested below.

Tools for non-linear clustering

The Zeldovich approximation

We have expressed linear theory in terms of the density contrast $\delta(\mathbf{x}, t)$.

You can also use linear theory to compute the displacements and peculiar velocities of particles given the linear density field.

Comoving displacements are proportional to the gravitational acceleration computed from $\delta(\mathbf{x}, t)$.

Peculiar velocities are proportional to displacements and to $f(\Omega_m) \approx \Omega_m^{0.55}$.

This linear perturbation theory for positions and velocities remains usefully accurate well after linear theory for $\delta(\mathbf{x}, t)$ has broken down, in part because the densities predicted from the particle positions never goes negative.

Because it tracks moving particles rather than fixed comoving positions, this is referred to as linear *Lagrangian* perturbation theory as opposed to *Eulerian* perturbation theory.

This approach was introduced by Yakov Zeldovich (1970). I commend the description and application of this approximation and a powerful extension called the adhesion approximation from one of my thesis papers, Weinberg & Gunn (1990).

A very short version:

The linear continuity equation implies that the displacements $\mathbf{\Delta x}(\mathbf{q}, t)$ obey

$$\vec{\nabla} \cdot \mathbf{\Delta x}(\mathbf{q}, t) = -\delta(\mathbf{q}, t) \ .$$

The associated peculiar velocities are

$$\mathbf{v} = a(\dot{\mathbf{\Delta x}}) = \frac{\dot{D}}{D}(a\mathbf{\Delta x}) \approx \Omega_m^{0.55} \cdot H(a\mathbf{\Delta x}) \ .$$

Higher order perturbation theory

You can go to higher order in perturbation theory, keeping terms of order δ^2, δ^3, etc.

This can be done in either Eulerian or Lagrangian form.

Nth-order perturbation theory is most useful for calculating quantities that are zero in (N-1)th-order pertubation theory, e.g., computing skewness induced by gravity from Gaussian (zero-skewness) initial conditions.

One can also use second-order perturbation theory to predict small departures from the linear theory $P(k)$ or $\xi(r)$.

Predicting galaxy clustering with perturbation theory is more complicated because you need to model the relation between galaxy and matter density fields (a.k.a. galaxy bias, to be discussed more later) to the same order.

Techniques known generically as "effective field theory of large scale structure" try to combine perturbation theory with some free parameters to represent terms that can't be calculated from first principles.

Spherical collapse and other exact solutions

The spherical collapse solution (PS 5 and further discussion below) is extremely useful.

Some other exact non-linear solutions (e.g., a plane-parallel perturbation) exist and are of some use, at least as a source of intuition.

N-body simulations

Fully numerical calculations that model the evolution of a distribution of N particles in a large cosmological volume.

N-body simulations

Initial conditions:

Create a linear density field $\delta(\mathbf{x}, t_i)$ by randomly drawing the amplitudes and phases of Fourier modes from a Gaussian distribution with the desired $P(k)$, then FFT.

Place particles on a uniform grid (or a smooth "glass.")

Compute initial particle displacements and velocities with the Zeldovich approximation (or second-order Lagrangian perturbation theory, a.k.a. 2LPT).

Evolution

At each timestep

1. Compute gravitational accelerations from the particle distribution.
2. Update particle velocities.
3. Update particle positions.
4. Repeat.

Output positions and velocities at desired redshifts.

Numerical considerations

Soften gravitational forces to avoid collisionality and suppress time integration errors.

Tradeoff between resolution and volume.

Many, many details of how to do N-body simulations efficiently and accurately.

Spherical collapse

As you know from PS 5, the evolution of a spherically symmetric perturbation in an $\Omega_m = 1$ universe can be solved analytically up to the point of collapse.

The post-collapse state is more guesswork, confirmed/guided by simulations.

Key features:

$\rho/\bar{\rho} \approx 5.5$ at turnaround

Collapse at epoch when $\delta_{\rm lin} \approx 1.69$

Formation of a bound structure with $\rho/\bar{\rho} \approx 200$

Halos and subhalos

A rough model for a dark matter halo is a singular isothermal sphere, $\rho(r) \propto r^{-2}$, with a faster decline beyond R_{200}.

GMore general isothermal spheres can have constant density cores.

An empirical model that describes the average halo profiles found in N-body simulations remarkably well is the NFW (Navarro, Frenk, & White 1996, 1997) profile:

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2} \ . \tag{1}$$

The profile transitions smoothly from r^{-1} at small radii to r^{-3} at large radii, being roughly r^{-2} at the scale radius r_s.

The behavior at small radii is an r^{-1} "cusp" instead of a constant density core.

The scale radius is often specified in terms of the virial radius (which often but does not always mean R_{200}) and the concentration parameter

$$c = R_{\rm vir}/r_s \ .$$

For a $10^{12}M_\odot$ halo, a typical c is 10-15, but concentrations are lower at higher masses and vice versa.

For more info, see the NFW wikipedia page, read the excellent NFW97 paper, or some of the 8000+ papers that cite it. (NFW96 is also good, but if you are choosing one, choose NFW97.)

A long-standing challenge for CDM is that observed galaxy rotation curves are usually well fit by halo profiles with cores, and the cuspy NFW profile predicts excessively high densities and rotation velocities in the inner regions of galaxies.

This is sometimes called the cusp-core problem, though an NFW profile is usually OK if the concentration is low enough.

If dark matter is warm rather than cold, concentrations are lower and the central profile may be flatter than ρ^{-1}.

Self-interacting dark matter (SIDM) could also change halo profiles.

Baryonic effects might also redistribute dark matter away from the profile predicted by purely gravitational simulations.

The most obvious impact of baryons is to concentrate the DM by pulling it inward, exacerbating the CDM problems. However, models with fairly violent, episodic feedback in the baryon component can reduce DM densities.

For a concise review of this topic, see Weinberg, Bullock, Governato, Kuzio de Naray, and Peter (2015).

Halos contain bound sub-halos, which themselves contain $\sim 10\%$ of the mass within the halo virial radius.

The halo mass function

A characteristic mass scale for non-linear structure is the mass M_* at which linear theory gives

$$\sigma(M_*) = \delta_c \tag{2}$$

where $\delta_c \approx 1.69$ is the critical threshold for spherical collapse.

If the initial fluctuation power spectrum is $P(k) \propto k^n$ The rms fluctuation amplitude scale with mass as

$$\sigma(M) = \delta_c (M/M_*)^{-\alpha}, \qquad \alpha = (3+n)/6$$

(see notes at end of §8). Scales $M \gg M_*$ should still be well described by linear theory.

What matters for purposes of the mass function is the effective slope n on scales close to M_*, not the large scale $n \approx 1$ corresponding to scale-invariance.

For a CDM power spectrum, the relevant slope is typically $n \approx -1$ to -1.5.

If the initial conditions are Gaussian with a power-law $P(k)$, the mass function of halos is approximately

$$\frac{dN}{dM} \propto M^{-1+\alpha} \exp\left[-(M/M_*)^{2\alpha}\right] \ . \tag{3}$$

Equation (3) was first derived with a combination of brilliance and luck by Press & Schechter (1974) and is referred to as the Press-Schechter mass function.

The rough argument: On each scale M, use Gaussian statistics to compute what fraction of the mass distribution smoothed on this scale is above the collapse threshold δ_c and therefore presumed to be in a bound halo. Differentiating this quantity gives the bound mass fraction in the range $M \to M + dM$, and dividing by the halo mass gives the number of halos.

This argument is put on firmer footing (though still with approximations) by Bond, Cole, Efstathiou, & Kaiser (1991).

More accurate models are calibrated on N-body simulations, famously J. Tinker et al. (2008).

Roughly speaking, the sub-halo mass function within a given halo has a Press-Schechter like form truncated at $\sim 5 - 10\%$ of the halo mass.

Again, there are detailed numerical studies (e.g., F. van den Bosch et al. 2005).

Galaxy formation

Big topic!

Big picture sketch

Baryons fall into halos alongside dark matter.

However, a fraction of these baryons can dissipate energy and sink to the center.

These may typically form a disk, supported by angular momentum, which is not easily dissipated. Typical size $R_{\rm disk} \sim 0.05 R_{200}$.

Gas forms stars.

Star formation injects energy, may eject gas out of the galaxy and back to the circum-galactic medium (CGM) or intergalactic medium (IGM).

Galaxies can merge.

Semi-analytic models

Semi-analytic models describe the above steps (and more) by physically motivated prescriptions, with free parameters that are chosen to reproduce some observations.

They can then be tested against other observations.

Cole, Aragon-Salamanca, Frenk, Navarro, & Zepf (1994) is a great paper in this genre, as is its successor, Cole, Lacey, Baugh, & Frenk (2000).

The paper of Mo, Mao, & White (1998) is also very illuminating, based on a simpler model.

Numerical simulations

One can add gas dynamics to N-body simulations and follow the formation of galaxies numerically, either in cosmological volumes or in "zoom" simulations that zero in on individual forming systems.

Voluminous literature. Hopkins et al. (2014), introducing the FIRE simulation suite, is one good starting point for recent work.

Some robust insights

Galaxy formation is inefficient: only a small fraction of baryons within a halo end up in the stellar component of the galaxy.

Either galaxies eject a large fraction of their baryons, or some process prevents accretion in the first place.

This suppression is a bigger effect at lower halo masses.

Below $M_{\rm halo} \sim 10^9 M_\odot$ (more precisely, $v_c \sim 30$ km/s), reionization suppresses galaxy formation (Bullock, Kravtsov, & Weinberg 2000).

In halos $M_{\rm halo} < 10^{12} M_\odot$, much of the gas is accreted "cold" along filamentary streams, without ever heating to the halo virial temperature (Keres, Katz, Weinberg, & Davé 2005; Birnboim & Dekel 2003, 2006).

Usually the most massive galaxy in a halo forms near the halo center.

When smaller halos merge into a larger one, they bring in satellite galaxies that orbit the central galaxy.

Galaxy bias

If galaxies form at special locations, they may be biased tracers of the underlying mass distribution.

The observed color and luminosity dependence of galaxy clustering implies that they *must* be biased tracers to some degree — at most one population of galaxies can have the same clustering as the matter.

Early models of this phenomenon focused on the bias of peaks in the initial density field: high peaks of a Gaussian field are "born clustered" (Kaiser 1984 and the *much* more voluminous tome of Bardeen, Bond, Kaiser, & Szalay 1985).

More generally, if the galaxy density depends on the "local" mass distribution, then the bias of $\xi(r)$ or $P(k)$ is scale-independent in the linear regime. In equations

$$\delta_g(\mathbf{x}) = F[\delta_R(\mathbf{x})]$$

implies

$$\xi_{\rm gg}(r) = b_g^2 \xi_{mm}(r), \qquad P_{\rm gg}(k) = b_g^2 P_{mm}(k) \tag{4}$$

on scales where $r \sim \pi/k \gg R$ *and* $\sigma(r) \ll 1$.

This statement is not obvious, and it may not be totally general, but it is supported by a variety of numerical and analytic arguments.

At second order (i.e., including terms proportional to δ_m^2 or δ_g^2), galaxy bias can be more complex and requires more parameters to describe it.

Early discussions of galaxy bias centered on whether one could reconcile theoretically simpler, $\Omega_m = 1$ models with observations of galaxy clustering and peculiar velocities implying $\Omega_m \sim 0.2$.

This would be possible if the underlying matter was less clustered than the galaxies, with $b_g \sim 2$.

Voids in the galaxy distribution might be underdense in matter but less empty than they appeared, and because they occupy a large volume they could contain a substantial amount of matter.

Although we now have convincing evidence that $\Omega_m < 1$, galaxy bias remains important for trying to infer cosmological parameters from or test cosmological models against any data involving galaxy clustering.

Halo-based models of galaxy bias

Halo bias

The clustering of dark matter halos depends on their mass.

A good approximation (Mo & White 1996, also Cole & Kaiser 1989) expresses the bias factor halos of mass M in terms of

$$\nu \equiv \frac{\delta_c}{\sigma(M)}$$

where $\delta_c \approx 1.69$ is the linear theory collapse threshold and $\sigma(M)$ is the rms fluctuation on mass scale M.

Extending the (modern version of) arguments that lead to the Press-Schechter mass function leads to the approximation

$$b_h(\nu) = 1 + \frac{\nu^2 - 1}{\delta_c} \ .$$

This approximation is pretty good, and Tinker et al. (2010) provide more accurate, numerically calibrated modifications of this formula.

Note that for $P(k) \propto k^n$, our previous formula for $\sigma(M)$ implies

$$\nu = (M/M_*)^{(3+n)/6}.$$

Thus, at redshift z, the bias of halos of mass M depends on the ratio $M/M_*(z)$.

Halo occupation distribution (HOD) modeling

One of the most widely used approaches for modeling non-linear clustering parameterizes the relation between galaxies and DM halos with the halo occupation distribution

$$P(N|M) = \text{probability a halo of mass } M \text{ contains } N \text{ galaxies} \ . \tag{5}$$

One must also specify the spatial and velocity distribution of galaxies within halos.

$P(N|M)$ depends on the luminosity and color (and maybe other proporties) of the galaxy population being considered, e.g., a $10^{12} M_\odot$ halo might contain one galaxy with $L > L_*$ but several with $L > 0.1L_*$.

HOD models usually adopt separate descriptions for central galaxies and satellites, assuming that the former move near the halo mean velocity and the latter trace the dark matter spatial and velocity distribution within the halo.

In HOD modeling, the galaxy correlation function can be separated into the "one-halo term" of galaxy pairs within the same halo and the "two-halo term" of galaxy pairs within different halos.

The one-halo term dominates on small scales but goes to zero on large scales, with overlap of the two on scales corresponding to the virial radii of large halos, $\sim 1-2$ Mpc.

HOD models perform quite well in reproducing observed galaxy clustering from linear to highly non-linear scales.

They transform number density and clustering information into physical characteristics of galaxy formation.

One can marginalize over parameters of the HOD to infer cosmological parameters.

Foundational papers: Berlind & Weinberg (2002), Zheng et al, (2005), Zehavi et al. (2005, 2011), Zheng & Weinberg (2007).

Subhalo abundance matching

Suppose we ignore satellites and just consider one galaxy per halo.

From theory, compute the cumulative halo mass function $n_H(>M)$, the space density of halos with mass $>M$.

From observations, compute the cumulative galaxy luminosity function $n_g(>L)$.

If the galaxy luminosity is a monotonic function of halo mass, we must have

$$n_g(>L) = n_h(>M) \ .$$

This condition allows one to infer the relation between galaxy luminosity and host halo mass.

Subhalo abundance matching (sometimes abbreviated SHAM) extends this idea to populate subhalos with satellite galaxies.

With "no free parameters," this model does astonishingly well at reproducing observed galaxy clustering as a function of luminosity and redshift (Conroy, Wechsler, & Kravtsov 2006).

In practice, in addition to the cosmological model, there is a free parameter to describe the *scatter* between luminosity and halo mass, and there are choices to make about how to assign satellites.

Nonetheless, this is a powerful method for modeling non-linear galaxy bias and for inferring the "galaxy-halo connection" between galaxy properties and halo properties.

Some key extensions of this idea are in excellent papers by Behroozi, Wechsler, & Conroy (2013) and Behroozi, Wechsler, Hearin, & Conroy (2019).

Moster, Naab, & White (2013) is another influential paper in this vein.