Bias, the Halo Model & Halo Occupation Statistics

Halo Bias

- the matter distribution is highly clustered.
- halos form at the peaks of this distribution.
Halo Bias

- Halos are biased tracers of the "background" dark matter field with a bias \( b(M) \) that is given by spherical collapse and the form of the mass function
  \[
  \frac{\delta n_M}{n_M} = [1 + b(M)] \delta
  \]

- For Press-Schechter
  \[
  b(M) = 1 + \frac{\mu^2 - 1}{\delta_c}
  \]

- Improved by the Sheth-Tormen mass function
  \[
  b(M) = 1 + \frac{\alpha \mu^2 - 1}{\delta_c} + \frac{2p}{\delta_c[1 + (\alpha \mu^2)^p]}
  \]

  with \( \alpha = 0.75 \) and \( p = 0.3 \) to match simulations.

In general a given model should simultaneously give \( b(M) \) and \( n(M) \)

Halo Bias in Simulations

\[
\xi_h = b^2 \xi_{DM}
\]

- \( b > 1 \) : bias
- \( b < 1 \) : anti-bias

Hu & Kravtsov 2002

see also
- Mo & White 1996, Sheth & Tormen 1999,
- Sheth, Mo & Tormen 2001, etc.

Halo Bias

- if halos are formed without regard to the underlying density, then
  \[
  \frac{\delta n_h}{n_h} = \frac{\delta \rho}{\rho}
  \]

- but spherical collapse model indicates that the probability of forming a halo depends on the initial density field: large scale density acts as a background enhancement

"peak-background split" (more on the theory from Zentner)
evolution of dark matter clustering

- evolves rapidly with redshift
- 2PCF not a power law
- evolution is a strong function of matter density and dark energy

Colin et al 1999

bias at high z

Wechsler et al 1998

evolution and scale dependence of galaxy clustering

- galaxy clustering is approximately a power law
- large-scale galaxy clustering is not a strong function of redshift
- clustering is more strongly scale dependent at high redshift

Adelberger et al
Zehavi et al

stochastic, non-linear bias
Dekel & Lahav 1999 (see also Kravtsov & Klypin 1999, Somerville et al 2001)

biasing may also be a function of scale, epoch, galaxy properties...
Dekel & Lahav 1999 (see also Kravtsov & Klypin 1999, Somerville et al 2001)
galaxy clustering

a function of luminosity, color & type:
galaxies must have a non-trivial relation to the matter distribution

\[ b = \sqrt{\xi_{gg}/\xi_{mm}} \]

4 orders of magnitude

\[ \text{halo mass} \]

galaxy abundance

clusterung

Mo & White 1996

Seljak & Warren 2004

\[ \text{brighter} \]

Masjedi et al. 2006

e.g., Sheth & Tormen 1999

Jenkins et al. 2001

Warren et al. 2005

zehavi et al. 2004

\[ 4 \text{ orders of magnitude} \]
The Halo Model

- Basic idea:
  - assume that stuff (e.g.: mass, galaxies, quasars, gas) lives in dark matter halos.
  - use knowledge of dark matter halo properties + relation of stuff to halos to determine the clustering properties of the stuff (e.g., non-linear power spectrum, galaxy clustering, etc...)

Halo Model Ingredients

what do we need to calculate power spectrum or clustering of mass or a population of things from knowledge of halos?

- \( n(M) \)
- \( b(M) \)
- spatial distribution of stuff within halos
- velocity distribution of stuff within halos

The Halo Model

- NFW halos, of abundance \( n_M \) given by mass function, clustered according to the halo bias \( b(M) \) and the linear theory \( P(k) \)
- Power spectrum example:
Halo Occupation

- relation of stuff to halos
- a prediction of physical models
- can be constrained by clustering
- simple example: if there is one galaxy per halo you can connect galaxies to halos by figuring out what halos match the abundance + clustering of the galaxies.

• the catch: halos may be occupied by more than one galaxy
• the number of galaxies per halo may depend on halo mass and galaxy luminosity

\[ N_g = \left( \frac{M}{M_H} \right)^{\alpha} \quad M_H > M_{\text{min}} \]
The missing piece: how galaxies populate halos

occupation function constraints for LBGs (z=3)

natural assumption: galaxies live in dark matter subhalos
**Halo Occupation of Galactic Halos**

- Average number of galactic (sub)halos

**Equation:**

\[ \alpha^2 \equiv \frac{\langle N(N-1) \rangle}{\langle N \rangle^2} \]

- \( N_{\text{sub}} \sim M \)

- A physically motivated way of characterizing non-linear bias of galaxies/subhalos
- No smoothing scale; naturally incorporates stochasticity, naturally brings about scale dependence

**Figure:**

- Host halo mass
- \( n = 5.85 \times 10^{-4} \) (halo Mpc\(^{-3}\))
- \( \langle N \rangle \), \( \langle N \rangle \), \( \langle N \rangle \)

---

**The Subhalo Counting Approach:**

- Instead, go from subhalos directly to galaxies using some mass connected property
- Assign luminosities to subhalo circular velocities by matching \( n(>v_{\text{max}}) \) to \( n(>L) \)

**Figure:**

- Luminosity function
- Velocity function

**References:**

- Kravtsov, Berlind, Wechsler, et al. 2004
- Tassis, Kravtsov, Wechsler & Primack 2005
- Conroy, Wechsler & Kravtsov 2006
- See also Vale & Ostriker 2006, 2007

---

**Galaxy-Galaxy Correlation Function at z=0**

- Data: Zehavi et al. 2004
- SDSS, z=0

**Figure:**

- Bright galaxies
- Dark matter
- Dimmer galaxies

**References:**

- Wechsler et al. 2002
- Conroy, Wechsler & Kravtsov 2006

---

**Distinct Halo Evolution**

- Constant increase

**Subhalo Evolution**

- Increase
- Decrease

**Figure:**

- Time
- Time

**References:**

- SDSS, z=0
- Data: Zehavi et al. 2004
- Conroy, Wechsler & Kravtsov 2006
Galaxies have a tight correspondence to their dark matter subhalos: a simple model which relates galaxy luminosity to subhalo velocity with no free parameters matches galaxy clustering from $z=0$ to $z=5$ (including the powerlaw nature and departures from it, luminosity and redshift dependence, scale dependence).

The hierarchical clustering predicted by LCDM (~30% matter, ~70% dark energy) has been directly observed. Gravity and the dynamical evolution of dark matter halos and subhalos are the primary drivers of galaxy clustering.

The halo occupation approach

- Assume galaxies live in halos
- Assume $P(N|M)$ is just a function of $M$

The relation between the clustering of Dark Matter and any class of galaxies (luminosity, type, etc.) is fully defined by the Halo Occupation Distribution (HOD):
- The probability distribution $P(N|M)$ that a halo of mass $M$ contains $N$ galaxies of that class.
- The relation between the spatial distributions of galaxies and DM within halos.
- The relation between the velocity distributions of galaxies and DM within halos.

How do we compute clustering statistics?

**Correlation function**

- **Small scales**: All pairs come from same halo.
  - The 1-halo term
  
  $$1 + \xi_{1h}(r) = \left\langle \frac{2 \pi r^2 n_s^2}{M} \int_{0}^{\infty} \frac{dn}{dM} \frac{N(N-1)}{2} \lambda(r|M) \right. \frac{dM}{dM}$$

- **Large scales**: Pairs come from separate halos.
  
  $$\xi_{2h}(r) = b_g^2 \xi_{gg}(r)$$

  where $b_g = n_s^2 \int_{0}^{\infty} \frac{dn}{dM} \frac{N}{M} b_h(M)$$

Berlind & Weinberg (2002)
1. Parameterize the halo occupation \( P(N|M) \) for a given set of galaxies, with 3-5 parameters
2. Assume that you know \( b(M) \) and \( n(M) \) for halos, plus the distribution of dark matter and/or galaxies within halos (either directly measured in simulations, or use analytic approximations. The later is typically referred to as “The Halo Model”)
3. Assume that halo mass is the only thing that impacts clustering
5. Need to do this as a function of luminosity: “conditional luminosity function”: use clustering + global LF to constrain LF as a function of cluster mass (e.g. van den Bosch et al 2003, Yang et al 2005, etc.)

Why is the Halo Occupation Distribution (HOD) the right way to think about galaxy clustering?

- **Powerful**: Halo properties have nearly universal relations.
- **Complete**: It tells us everything a theory of galaxy formation has to say about galaxy clustering (all statistics, all scales).
- **Physically illuminating**: Discrepancies offer guidance about their physical origin.

Nice conceptual division between roles of “cosmological model” and “theory of galaxy formation”.

The HOD contains information about physics!

**Cosmological Model**
- \( \Omega, P(k), \text{etc.} + \text{Gravity} \)

**Galaxy Formation**
- Gas cooling, Star formation, Feedback, Mergers, etc.

**Dark Halo Population**
- \( n(M), \rho(r|M), \bar{v}(r|M), v(r|M) \)

**Halo Occupation Distribution**
- \( P(N|M) \): Spatial distribution within halos, Velocity distribution within halos

Galaxy clustering

Baryon/DM fraction
- Gas cooling
- Star formation efficiency

DM halo merger statistics

Dynamical friction
- Tidal disruption
the halo model

The Halo Occupation Distribution:
- $P(N|M)$ -- typically parameterized as the first two moments of $N(M)$
- Typically assumed to be a function of only $M$, and not any other halo property that correlates with environment
- In principle, can be extended to other properties of the halo: $P(N|M, \text{formation time, density profile, shape, angular momentum, large scale environment, etc...})$

The Halo Model:
- A way to calculate galaxy clustering statistics, which assumes that galaxies live in halos, and that their distribution in halos can be specified by $P(N|M)$ and $n(r)$. Uses this plus $b(M)$ [measured in sims or analytic] to calculate clustering.

The Standard Halo Model
- $N$ and its moments are only a function of $M$, and not of any other property of the halo.
- This is equivalent to the assumption that extended Press-Schechter makes: halo formation is a random walk, and the future of a halo depends only on its mass, and not on its past history or environment
- Is this correct?

40 Mpc slice of 120 Mpc box
one box: $\sim 3 \times 10^{11} \text{ Msun/h}$ halos; earliest forming 15% red, latest forming 15% blue
other box: $\sim 3 \times 10^{12} \text{ Msun/h}$ halos (red) and $\sim 3 \times 10^{11} \text{ Msun/h}$ halos (blue)

Relative halo bias as a function of concentration
- Concentrated/early formers
- Unconcentrated/late formers

Wechsler et al 2006

Relative halo bias correlates with formation time and halo concentration

Zentner et al 2005; Wechsler et al 2006

see also Gao et al 05; etc etc
Halo Occupation and Clustering

First direct indication that the assumption of the Standard Halo Model is incorrect

Wechsler et al 2006

Implications

The Result

- Halo clustering is not only a function of M; depends on other observables
- High mass, high occupation halos are early forming and less clustered; low mass, high occupation halos are late forming and more clustered.
- The trends with environment are negligible near M/M*, and for mass selected samples: at the <5-10% level for luminosity selected samples around L*

Implications for the Halo Model

- Halo model is still extremely useful for understanding features in the correlation function, and how clustering properties of galaxies and depend on halo mass, type, etc,
- It may need improvement before being used for "precision cosmology", and for understanding rare/formation selected samples.

In particular:

- LSB galaxies: less clustered --> lower concentration halos?
- If galaxy formation is inefficient in small mass halos (e.g., reionization, feedback, etc), dwarf galaxies should avoid the voids more than typical halos of that mass
- Self-calibration of the cluster mass function may be affected (number or x-ray selected samples are likely to be early formers, and thus biased high compared to typical halos of their mass)