Angular power spectra and correlation functions

Notes: Martin White

Introduction

An action item on the last telecon was for me to prepare some background on angular clustering. I slightly enlarged upon this task to include some half-baked calculations I've done as background to the "challenge"s. The next few sections cover mostly well-known material using the $z = 0$, dark matter, MICE octant as an explicit example. The "Background" section at the end gives some semi-random references for further reading. Readers interested specifically in BAO can find a reference list at `http://cdm.berkeley.edu/doku.php?id=baopages`.

Notation

Define the dimensionless (3D) power spectrum as

$$\Delta^2(k) \equiv \frac{k^3\,P(k)}{2\pi^2} = \frac{d\sigma^2}{d\ln k} \quad . \tag{1}$$

Sometimes this is written $\mathcal{P}(k)$, and if a different Fourier transform convention is used it can be written $4\pi k^3\,P(k)$. Ignoring redshift space distortions the correlation function is the 3D Fourier transform of this:

$$\xi(r) \equiv \int \frac{dk}{k}\ \Delta^2(k) j_0(kr) \tag{2}$$

where $j_0(x) = \sin x/x$ is the spherical Bessel function of order 0. For an observed number density of tracers, $\bar{n}$, in the Gauss-Poisson limit if

$$\rho_i \equiv \int \frac{dk}{k}\ \Delta^2(k) W_i(k) \tag{3}$$

for some hopefully useful $W_i(k)$ then a survey of volume V would return ρ with errors

$$\mathrm{cov}\left[\rho_i, \rho_j\right] = \frac{2}{V} \int \frac{k^2\,dk}{2\pi^2}\ W_i W_j \left[P(k) + \bar{n}^{-1}\right]^2 + \cdots \tag{4}$$

where the $\cdots$ represent higher order shot-noise terms (see Meiksin & White; MNRAS, 308, 1179 for the full list if you care). I shall assume spatially flat cosmologies below, for simplicity.

Two dimensional surveys

In DES we will not have enough redshift resolution to do a full 3D analysis. The most efficient way to proceed is then to project the survey into slices and analyze the slices. If we include the inter-slice correlation between angular power spectra (or 2-point functions) we have lost no information by first binning into slices. For the photo-z errors anticipated for DES the slice-slice covariance

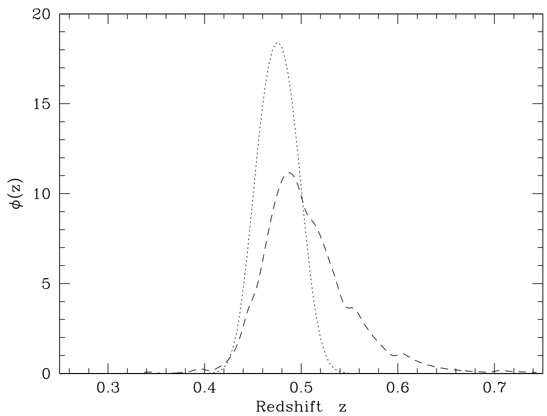

Figure 1: A comparison of two z-distributions each from a slice with $0.45 < z_{\rm photo} < 0.50$. The dotted line assumes a Gaussian photo-z distribution with $\sigma = 1\% \, (1+z)$. The dashed line is actual z-distribution of SDSS LRGs with the same cuts (from Padmanabhan et al. 2007). Both functions are normalized to unit integral.

matrix will be tri-band diagonal to a very good approximation, meaning we only need to keep correlations between slices and their neighbors (see Figure 4). The cross-correlation of more distant slices will be an excellent check for photo-z failures, since any signal indicates the presence of outliers. For this check we can make use of the small-scale clustering, for which we will have excellent statistics.

For 2D surveys we project the 3D density field into some slice, with the probability that an object in the slice is at distance χ or redshift z given by

$$\phi(\chi) = \phi(z) \, \frac{dz}{d\chi} = \phi(z) H(z) \tag{5}$$

conventionally normalized to unit integral. For a distribution of photo-z, z_p, given true redshift, z_t, we have

$$\phi(z_t) \propto \int_{z_1}^{z_2} dz_p \; p\,(z_p|z_t)\, n(z_t) \tag{6}$$

where z_1 and z_2 are the photo-z limits of the slice. In Figure 1 I show the redshift distribution for a photo-z slice selected according to $0.45 < z_{\rm photo} < 0.50$. There are two lines. One is the $\phi(z)$ obtained assuming a Gaussian photo-z distribution with $\sigma = 1\% \, (1+z)$, the other is the actual z-distribution of SDSS LRGs with the same cuts.

If the width of the selection function, $\phi(z)$, is much larger than the scale we are probing and we are interested primarily in small angular scale fluctuations

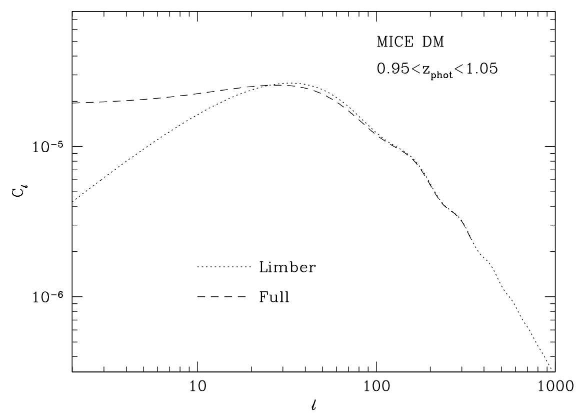

Figure 2: The angular power, C_ℓ or $P_2(K)$, vs. wavenumber ℓ or K, for a thin slice $0.95 < z_{\rm photo} < 1.05$ assuming 1% photo-z errors and the $z = 0$ linear theory DM power spectrum for the parameters of the MICE simulation. The baryon acoustic oscillations are visible near wavenumbers of a few hundred. The dotted line makes the Limber approximation while the dashed line does the full calculation (Eq. 12).

(so $\sin\theta \approx \theta$ and the sky is "flat") then we can make the "Limber approximation":

$$P_2(K) = \int d\chi \left(\frac{\phi}{\chi}\right)^2 P_3\left(\frac{K}{\chi}\right) \quad . \tag{7}$$

An example is plotted in Figure 2. Also in the flat sky limit

$$w(\theta) = \int \frac{K\,dK}{2\pi} P_2(K) J_0(K\theta) \quad ; \quad P_2(K) = 2\pi \int \theta\,d\theta\; w(\theta) J_0(K\theta) \quad . \tag{8}$$

Note that $\phi(\chi)$ has units of inverse length. If we approximate ϕ as a top-hat of width L then its height must be $1/L$ from the normalization condition. Since ϕ comes in squared in Eq. (7) we see immediately that the amplitude of P_2 drops with the width of the survey and angular mode K receives contributions from K/χ_1 to K/χ_2 where χ_1 and χ_2 bracket the peak of ϕ. [In the current best-fit cosmology $\chi \approx 2.5\,h^{-1}$Gpc for objects at $z = 1$.] A wider selection function thus washes out the baryon oscillations and decreases the amplitude of the entire signal. However as we narrow ϕ we become increasingly sensitive to errors in our photo-z distribution.

Writing the angular number density of the sources as $\mathcal{N}$ the covariance is then

$$\mathrm{cov}\left[w_i, w_j\right] = \frac{2}{\Omega} \int \frac{K\,dK}{2\pi} J_0(K\theta_i) J_0(K\theta_j) \left[P_2 + \mathcal{N}^{-1}\right]^2 + \cdots \tag{9}$$

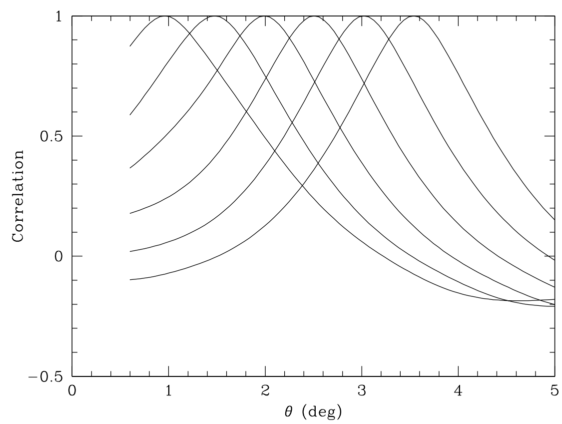

Figure 3: The Gauss-Poisson prediction for the correlation between $w(\theta)$ and $w(\theta')$ for a thin slice $0.95 < z_{\rm photo} < 1.05$ from the DM of the MICE simulation. Different lines represent different rows of the correlation matrix, each centered at the θ where the curve hits unity.

in the limit of infinitesimal bins. For finite size bins you need to integrate $\int \theta d\theta J_0(K\theta)$ which is simple using the identity

$$\int_0^x ydy\ J_0(y) = xJ_1(x) \qquad . \tag{10}$$

If one uses angles larger than a few degrees you can preserve the measure of the full sphere by defining $\widetilde{\omega} \equiv 2\sin(\theta/2)$, with $\widetilde{\omega} \approx \theta$ for $\theta \ll 1$ but $\widetilde{\omega} d\widetilde{\omega} d\phi = d(\cos\theta) d\phi$. Binning $w(\theta)$ in $\widetilde{\omega}$ rather than θ is a trivial change. [Note that if each galaxy has assigned to it a unit vector, $\hat{n}_i$, the 'distance' between two galaxies is $\widetilde{\omega}_{ij}^2 = 2\hat{n}_i \cdot (\hat{n}_i - \hat{n}_j)$.]

We are anticipating $\sim 10^{5-6}$ galaxies in our photo-z shells over the full $\pi/2\,$sr of the survey, so $\mathcal{N}^{-1} \sim 10^{-5}$. Depending on the bias of the tracers and the redshift we become shot-noise limited in the range $\ell = 100 - 1000$.

A plot of some rows of the correlation matrix of $w(\theta)$ for a slice $0.95 < z_{\rm photo} < 1.05$ from the DM of the MICE simulation is shown in Figure 3. Since $w(\theta)$ points with close θ are very highly correlated, there is little to be gained by a very fine binning in θ and indeed too fine a binning places strong demands on the fidelity of the covariance matrix estimator. [An alternate line of thought is that you should dump $w(\theta)$ in very fine bins and then later rebin or use a regularization prescription – then my comment refers to the final step not the intermediate one.] The diagonal elements of the covariance matrix are shown in Figure 4.

Sometimes we write $P_2(K)$ as C_ℓ with $\ell = |K|$ and the dimensionless 2D power spectrum $K^2\,P(K)/2\pi$ as $\ell(\ell+1)C_\ell/(2\pi)$. These Hankel transforms can

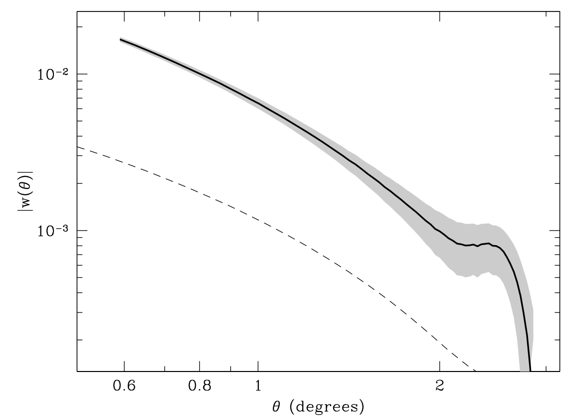

Figure 4: The angular correlation function, $w(\theta)$, for a thin slice $0.95 < z_{\rm photo} < 1.05$ assuming the linear theory matter power spectrum appropriate for the MICE simulation at $z = 0$. The shaded band shows the 1σ error on infinitesimal bins in θ assuming Gaussian statistics. The dashed line shows the cross spectrum between this slice and one with $0.85 < z_{\rm photo} < 0.95$. The next-to-nearest cross-spectrum is below the y-range of the plot.

be related to the (possibly) more familiar Legendre transforms (below) by noting that $P_\ell(\cos\theta) \simeq J_0(\ell\theta)$ for $\theta \ll 1$ and $\ell \gg 1$.

If our photo-z errors are sufficiently small (as we hope they are for DES) then the Limber approximation is no longer accurate. Dropping the small angle approximation at the same time we have

$$C(\theta) = \frac{1}{4\pi} \sum_\ell (2\ell + 1) C_\ell P_\ell(\cos\theta) \tag{11}$$

with

$$C_\ell = 4\pi \int \frac{dk}{k}\ \Delta^2(k) W_\ell^2(k) \tag{12}$$

and

$$W_\ell(k) = \int d\chi\ \phi(\chi) j_\ell(k\chi) \tag{13}$$

where j_ℓ is the spherical Bessel function of order ℓ. For the cross-spectra between slices replace W_ℓ^2 with the product of the W_ℓ for each slice. For very large scales ($\ell < 30$) and very narrow slices there is an additional correction due to redshift space distortions. Figure 2 compares $P_2(K)$ computed within the Limber approximation with C_ℓ computed from Eq. (12). The large discrepancy at low ℓ is well known in the CMB and arises due to "projection bleeding" of high k power into low ℓ modes (e.g. Hu & White 1997; Phys. Rev. D, 56, 596;

Fig. 2). Note however that even at $\ell \sim 10^2$ there is a signficant discrepancy between the Limber approximation and the full calculation. This discrepancy would be larger for thinner slices.

How good is the Gaussian approximation to the covariance matrix? On large scales we expect it to be quite reasonable, though this needs to be tested on simulations. In principle the covariance matrix of the data can be estimated from the survey itself using either jacknife or bootstrap methods. [Currently we can say that the diagonal errors computed above agree roughly with the errors obtained by breaking the octant into 4 pieces.] However, for these methods to converge well requires that the patches into which the survey is divided be approximately independent. The integrals for $w(\theta)$ however encompass modes of quite long wavelength, which correlate the patches and violate this assumption. The observed field thus needs to be filtered before the comparison is made, preferably by using an estimator different from $w(\theta)$ [see the discussion of this point in arxiv:0802.2105 section 3.1.1.]

Background

For those unfamiliar with 2-point statistics and estimators I list here some background reading. Most modern text books contain a discussion of 2-point statistics. Peeble's classic "The large-scale structure of the Universe" still contains one of the best descriptions of the theory of 2- and N-point statistics in cosmology. A standard reference for statistical estimators for 2-point functions is Landy & Szalay (1993; ApJ, 412, 64). [Note that the relevance of the Landy-Szalay estimator to next-generation mega-surveys is questionable. It may be that simple estimators like $DD/RR - 1$ work as well in practice.] For a summary of some of the key ideas, you can find some simple notes from a Berkeley reading group at
`http://astro.berkeley.edu/~mwhite/teachdir/minicourse_03.html`.
A new idea is presented in Padmanabhan et al. (2007; MNRAS, 376, 1702) and applied to the SDSS LRG sample in Padmanabhan et al. (2008; arxiv:0802.2105).

There are numerous algorithmic issues regarding how one computes two point functions, and several publicly available packages. For computing C_ℓ there are iterative, parallel optimal estimators usually run on supercomputers (e.g. the MADspec package at `http://crd.lbl.gov/~borrill` described in astro-ph/9712121). There are equivalent estimators which use an iterative matrix inversion scheme such as (pre-conditioned) conjugate gradient rather than the brute-force approach used above (e.g. astro-ph/0605302, with the technical details given in the appendices of astro-ph/0406004 and earlier work by Tegmark which has to some extent been superceded by the references above). Depending on the science goal you can have fast non-optimal estimators which perform almost as well as the maximum likelihood codes (e.g. Szapudi et al. 2001; ApJ, 548, L11; which inverts the correlation function – this code is publicly available). One of the issues with these fast algorithms is an estimate of the covariance matrix of the results – usually this is handled by Monte-Carlo which can be computationally expensive. There is a fairly large literature on accelerating these algorithms (e.g. Szapudi et al.; 2005, ApJ, 631, L1 or Gray et al.; 2004, ASPC, 314, 249 as only two examples). Many of these codes can be trivially parallelized. Writing these things isn't so hard, so *lots* of people have their own home-grown versions with various advantages or inefficiencies.