Incomplete

General procedure

Our input data is a collection of redshift-space galaxy positions (s_g, θ_g, ϕ_g) for $1 \leq g \leq N_{\text{galaxies}}$. We collect this data into a (redshift-space) galaxy number density

$$n(\boldsymbol{s}) = \sum_g \delta_D(\boldsymbol{s} - \boldsymbol{s}_g), \tag{1}$$

where δ_D is the Dirac delta function. According to the standard assumption, these galaxy positions are determined by a Poisson point process with intensity function $\bar{n}(1+\delta_g)$, where the galaxy density contrast δ_g is linearly related to the underlying matter density contrast δ on large scales. Since we are only interested in the large-scale behavior of $n(\boldsymbol{s})$, we project it onto a subspace spanned by a suitable set of smooth functions over the survey volume V.

We introduce the standard inner product for functions over $\mathbb{R}^3$,

$$(f, g) \equiv \int d^3x \; f(\boldsymbol{x}) g(\boldsymbol{x}) \tag{2}$$

We choose a set of basis functions $\{\phi_a(\boldsymbol{s})\}$ that span a reasonable space of smooth functions over V, with $a \leq 1 \leq N_{\text{basis}}$, and define basis pixels

$$x_a \equiv (n, \phi_a) = \int d^3s \; n(\boldsymbol{s}) \phi_a(\boldsymbol{s}) = \sum_g \phi_a(\boldsymbol{s}_g). \tag{3}$$

Following the standard assumptions, we have

$$\langle n(\boldsymbol{s}) \rangle = \bar{n}(\boldsymbol{s}), \tag{4}$$

$$\langle n(\boldsymbol{s}_1) n(\boldsymbol{s}_2) \rangle = \bar{n}(\boldsymbol{s}_1) \bar{n}(\boldsymbol{s}_2)[1 + \xi_g(\boldsymbol{s}_1, \boldsymbol{s}_2)] + \bar{n}(\boldsymbol{s}_1) \delta_D(\boldsymbol{s}_2 - \boldsymbol{s}_1). \tag{5}$$

where $\xi_g(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is the (wide-angle, redshift-space) galaxy correlation function. The covariance of the pixel values x_a is therefore

$$\text{Cov}[x_a, x_b] = \int d^3s_1 \; d^3s_2 \; \phi_a(\boldsymbol{s}_1) \phi_b(\boldsymbol{s}_2) \bar{n}(\boldsymbol{s}_1) \bar{n}(\boldsymbol{s}_2) \xi_g(\boldsymbol{s}_1, \boldsymbol{s}_2) + \int d^3s \; \phi_a(\boldsymbol{s}) \phi_b(\boldsymbol{s}) \bar{n}(\boldsymbol{s}) \tag{6}$$

$$\equiv S_{ab} + N_{ab}. \tag{7}$$

The first contribution S_{ab} carries the cosmologically interesting signal, while the second contribution N_{ab} is simply Poisson noise.

While our original set of basis functions $\{\phi_a(\boldsymbol{s})\}$ may be chosen arbitrarily, we are primarily interested in integral combinations of the galaxy density that carry the most cosmological signal. To that end, we first carry out a Karhunen-Loève transform to obtain a more suitable basis. Let $\{\lambda_i\}$ and $\{\boldsymbol{b}_i\}$ be the complete set of real eigenvalues and orthonormal eigenvectors of the generalized eigenvalue problem $\mathsf{S}\boldsymbol{b} = \lambda\mathsf{N}\boldsymbol{b}$, arranged in order of decreasing eigenvalue. (A complete set of such eigenvalues/eigenvectors is guaranteed to exist since S is symmetric and N is a positive definite matrix. By "orthonormal" we mean with respect to the noise matrix, i.e. $\boldsymbol{b}_i^T N \boldsymbol{b}_j = \delta_{ij}$.) Define a new basis of functions $\{\psi_i(\boldsymbol{s})\}$ by

$$\psi_i(\boldsymbol{s}) = \sum_{a=1}^{N_{\text{basis}}} b_{i,a}\phi_a(\boldsymbol{s}), \tag{8}$$

Then the new pixels defined by

$$y_i \equiv (n, \psi_i) = \int d^3s\; n(\boldsymbol{s})\psi_i(\boldsymbol{s}) = \sum_g \psi_i(\boldsymbol{s}_g) \tag{9}$$

have covariance

$$\text{Cov}[y_i, y_j] = \delta_{ij}(1 + \lambda_i), \tag{10}$$

so that the eigenvalue λ_i measures the signal-to-noise ratio.

We refer to the functions ψ_i as "mode functions" or "KL mode functions," and call the set of these mode functions the "KL basis." Since all we have done so far is perform a change of basis, we have lost no information about the cosmological signal. However, since the mode functions are arranged in order of decreasing signal-to-noise, most of the interesting signal is confined to the first several modes (where "several" may of course be a large number, but hopefully is small compared to N_{basis}). We achieve a substantial data compression by simply discarding those modes with lowest signal-to-noise ratio. That is, we restrict the index i to run only from 1 to some N_{modes}. We let B be the $N_{\text{modes}} \times N_{\text{basis}}$ matrix with the eigenvectors $\boldsymbol{b}_i$ as rows, i.e. $B_{ia} = b_{i,a}$. This matrix serves as a projection operator from the original space of smooth functions to the KL subspace: $y_i = \sum_a B_{ia}x_a$ and $\text{Cov}[y_i, y_j] = \sum_{a,b} B_{ia}B_{jb}\,\text{Cov}[x_a, x_b]$.

Count-in-cell basis

From the definitions of the pixels x_a and the matrices S_{ab} and N_{ab}, we see that the basis functions $\phi_a(\boldsymbol{s})$ are always multiplied by a factor of the selection function $\bar{n}(\boldsymbol{s})$, and therefore the values of the basis functions outside the survey volume are irrelevant. In order to get the most out of our basis, without including members, we should choose our functions ϕ_a to vanish outside the survey volume. Since modern surveys tend to have rather complicated angular boundaries, the simplest way to enforce this constraint is to partition our survey into a collection of discrete cells V_a, and choose our basis functions to be indicator functions over these cells, i.e.

$$\phi_a(\boldsymbol{s}) = \begin{cases} \bar{\phi}_a, & \boldsymbol{s} \in V_a, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The constant value $\bar{\phi}_a$ is of course arbitrary, but we can achieve a significant simplification if we choose $\bar{\phi}_a = \bar{N}_a^{-1/2}$ where

$$\bar{N}_a = \int_{V_a} d^3s \; \bar{n}(\boldsymbol{s}), \tag{12}$$

which fixes the noise matrix to be the identity:

$$N_{ab} = \int d^3s \; \phi_a(\boldsymbol{s})\phi_b(\boldsymbol{s})\bar{n}(\boldsymbol{s}) = \delta_{ab}. \tag{13}$$

Now the generalized eigenvalue problem $\mathsf{S}\boldsymbol{b} = \lambda\mathsf{N}\boldsymbol{b}$ reduces to the regular one, $\mathsf{S}\boldsymbol{b} = \lambda\boldsymbol{b}$.