# Bispectrum estimator

Héctor Gil Marín*

November 30, 2018

These notes are the complementary material to the seminar 'High order statistics in large scale structure' given at the Instituto de Física at the Universidad Nacional Autónoma de Mexico.

## 1 Motivation

In these notes I present the isotropic signal of the 3-point correlation function in Fourier space, the bispectrum. We know that the late-time distribution of galaxies in the Universe is non-Gaussian with $\sim 40\%$ skewness in $50\mathrm{Mpc}/h$ spheres, encoding relevant cosmological information, which is independent (although correlated) from the one in the power spectrum. Using a series of iFT, the bispectrum signal can be efficiently computed. Such estimator was used in the analysis of the galaxy spectroscopic galaxy BOSS surveys.

## 2 Estimator formalism

We start from the field of density fluctuations $\delta(\mathbf{x})$ described in the previous chapter of these notes.

The bispectrum estimator is defined as the angle-average of closed triangles defined by the $\mathbf{k}$-modes, $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$,

$$B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \frac{1}{V_{123}} \int_{\mathcal{S}_1} d\mathbf{q}_1\, \delta(\mathbf{q}_1) \int_{\mathcal{S}_2} d\mathbf{q}_2\, \delta(\mathbf{q}_2) \int_{\mathcal{S}_3} d\mathbf{q}_3\, \delta(\mathbf{q}_3) \delta^D(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3) \tag{1}$$

where, $\delta(\mathbf{q})$ is the Fourier transform $\delta(\mathbf{x})$, $k_f$ is the fundamental frequency and $V_{123}$ is the number of fundamental triangles,

$$V_{123} \equiv k_f^{-3} \int_{\mathcal{S}_1} d\mathbf{q}_1 \int_{\mathcal{S}_2} d\mathbf{q}_2 \int_{\mathcal{S}_3} d\mathbf{q}_3\, \delta^D(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3), \tag{2}$$

*Institut de Ciències del Cosmos. Universitat de Barcelona. hectorgil@icc.ub.edu

inside the shell defined by $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$, where $\mathcal{S}_i \equiv \mathcal{S}(k_i|\Delta k)$ is the $k$-region contained by $k_i - \Delta k/2 \leq k \leq k_i + \Delta k/2$, given a $k$-bin, $\Delta k$,

$$\int_{\mathcal{S}_i} d\mathbf{q} \equiv \int_{k_i-\Delta k/2}^{k_i+\Delta k/2} dq\, q^2 \int d\Omega_q, \tag{3}$$

and $\delta^D$ is the Dirac delta distribution that ensures the condition of only including closed triangles.

Writing the Dirac-delta as the exponential expression, $\delta^D(\mathbf{x}) \equiv \frac{1}{(2\pi)^3}\int d\mathbf{y}\, e^{i\mathbf{x}\cdot\mathbf{y}}$, we can re-write Eq. 1 as a separate product of iFT,

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{(2\pi)^6}{V_{123}} \int d^3\mathbf{r}\, \mathcal{D}_{\mathcal{S}_1}(\mathbf{r}) \mathcal{D}_{\mathcal{S}_2}(\mathbf{r}) \mathcal{D}_{\mathcal{S}_3}(\mathbf{r}) \tag{4}$$

where,

$$\mathcal{D}_{\mathcal{S}_j}(\mathbf{r}) \equiv \frac{1}{(2\pi)^3} \int_{\mathcal{S}_j} d\mathbf{q}_j\, \delta(\mathbf{q}_j) e^{i\mathbf{q}_j\cdot\mathbf{r}} = \frac{1}{(2\pi)^3} \int d\mathbf{q}_j\, \tilde{\delta}(\mathbf{q}_j|\mathcal{S}_i) e^{i\mathbf{q}_j\cdot\mathbf{r}} \tag{5}$$

Here, $\tilde{\delta}(\mathbf{k}|\mathcal{S}_i)$ is a field which is equal to $\delta(\mathbf{k})$ within the $\mathcal{S}_i$ region, and 0 otherwise.

Note that $V_{123}$ can be estimated in the same way as $B$, just by changing $\delta(\mathbf{k})$ by 1. In this fashion,

$$\mathcal{U}_{\mathcal{S}_j}(\mathbf{r}) \equiv \frac{1}{(2\pi)^3} \int_{\mathcal{S}_j} d\mathbf{q}_j\, e^{i\mathbf{q}_j\cdot\mathbf{r}} = \frac{1}{(2\pi)^3} \int d\mathbf{q}_j\, u_{\mathcal{S}_i} e^{i\mathbf{q}_j\cdot\mathbf{r}} \tag{6}$$

where $u_{\mathcal{S}_i}$ is unity within the $\mathcal{S}_i$ region, and 0 otherwise. Thus,

$$V_{123} \equiv V(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = (2\pi)^6 k_f^{-3} \int d^3\mathbf{r}\, \mathcal{U}_{\mathcal{S}_1}(\mathbf{r}) \mathcal{U}_{\mathcal{S}_2}(\mathbf{r}) \mathcal{U}_{\mathcal{S}_3}(\mathbf{r}) \tag{7}$$

Note that the $(2\pi)^6$ factors in Eq. 4 and 7 cancel out when computing the bispectrum.

The total number of non-equivalent triangular configurations is $\sim N_b^3/2$, where $N_b$ is the number of bins considered between the fundamental frequency, $k_f$, and the Nyquist frequency, $k_{\rm Ny} \equiv N^{1/3} k_f$. Here, $N$ is the total number of cartesian grid-cells chosen for performing the DFT. In the computation of Eq. 4, the bispectrum of a triplet $\{k_1, k_2, k_3\}$, requires only 3 FT and most triplets share one or two $k_i$-vectors, which means that they share as well the output of the FT of the shared vectors. Consequently, the estimator of Eq. 4 is much faster than the naive implementation of Eq. 1, which would require to produce random orientations of the triangles within the binned region.

Finally, when we discretise the FT and take into account the normalisation factors, the bispectrum estimator looks like,

$$B_{123} = k_f^3 \frac{\sum_{\mathbf{r}'} \left[\sum_j \tilde{\delta}(\mathbf{q}_j|\mathcal{S}_1) e^{i\mathbf{q}_j\cdot\mathbf{r}'}\right] \left[\sum_j \tilde{\delta}(\mathbf{q}_j|\mathcal{S}_2) e^{i\mathbf{q}_j\cdot\mathbf{r}'}\right] \left[\sum_j \tilde{\delta}(\mathbf{q}_j|\mathcal{S}_3) e^{i\mathbf{q}_j\cdot\mathbf{r}'}\right]}{\sum_{\mathbf{r}''} \left[\sum_j u_{\mathcal{S}_1} e^{i\mathbf{q}_j\cdot\mathbf{r}''}\right] \left[\sum_j u_{\mathcal{S}_2} e^{i\mathbf{q}_j\cdot\mathbf{r}''}\right] \left[\sum_j u_{\mathcal{S}_3} e^{i\mathbf{q}_j\cdot\mathbf{r}''}\right]} \tag{8}$$

## Exercise A: Power spectrum estimator

We can write a similar type of estimator for the power spectrum,

$$P(k|\Delta k) = \frac{1}{V_P(\Delta k)} \int_{\mathcal{S}_{\Delta k}} d\mathbf{q}_1 \, \delta(\mathbf{q}_1) \int_{\mathcal{S}_{\Delta k}} d\mathbf{q}_2 \delta(\mathbf{q}_2) \delta^D(\mathbf{q}_1 + \mathbf{q}_2) \tag{9}$$

where,

$$V_P(\Delta k) \equiv k_f^{-3} \int_{\mathcal{S}_{\Delta k}} d\mathbf{q}_1 \int_{\mathcal{S}_{\Delta k}} d\mathbf{q}_2 \, \delta^D(\mathbf{q}_1 + \mathbf{q}_2) = 4\pi (k/k_f)^2 (\Delta k/k_f) \left[1 + \Delta k^2/(12k^2)\right] \tag{10}$$

Implement this methodology to write a code to compute the power spectrum using the above estimator. Compare the output of this estimator with the estimator written in the previous chapter.

In order to do so, you would need to find a way to rapidly write a $\tilde{\delta}(\mathbf{k})$ given $\Delta k$ and $k$. Think that $\mathbf{k} = (k_x, k_y, k_z)$ and each of these $k_i$ runs from $0, k_f, 2k_f, \ldots, (N-1)k_f$. However, nearly the 2nd half of frequencies should be converted into negative frequencies such as, $k_i$ runs from

$$k_i = \left\{0, k_f, 2k_f, \ldots, (N/2-1)k_f, (N/2)k_f, -(N/2-1)k_f, -(N/2-2)\ldots, -k_f\right\}$$

for,

$$i = \{0, 1, 2, \ldots, N/2-1, N/2, N/2+1, N/2+2\ldots, N-1\}$$

so, given a $|k| \pm \Delta k/2 = \sqrt{k_x^2 + k_y^2 + k_z^2} \pm \Delta k/2$, you would need to find the set of values of $k_x, k_y, k_z$ and their corresponding $i$-indices which contribute to that bin.

This a purely geometrical problem, and may have different solution with different efficiency. Of course you could loop through all $k_x, k_y, k_z$ combination and select those which satisfy this relation, but you want to find a more efficient solution to this problem, as when you have to repeat this many times would considerably slow your computation.

## Exercise B: Bispectrum normalisation factor

Demonstrate that the bispectrum normalization factor is approximately (for $k \gg \Delta k$),

$$V_{123} \equiv k_f^{-3} \int_{\mathcal{S}_1} d\mathbf{q}_1 \int_{\mathcal{S}_2} d\mathbf{q}_2 \int_{\mathcal{S}_3} d\mathbf{q}_3 \, \delta^D(\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3) \simeq 8\pi^2 k_1 k_2 k_3 (\Delta k/k_f)^3 \tag{11}$$

where, the regions $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$ , $k_i - \Delta k/2 \leq k_i \leq k_i + \Delta k/2$. Also, use the FT formalism developed in the previous chapter of the notes to implement this calculation in your computer, and check the result of this integral as a function of $k_1$, $k_2$ and $k_3$ for a given value of $\Delta k$ (for example, $\Delta k = k_f$).