Generating Gaussian Random Field using Fourier Transforms

Héctor Gil Marín*

November 28, 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

The key feature of a Gaussian Random Field (GRF) is that their structure is fully determined by the second order moment of the Gaussian distribution. The 2-point correlation function (2PCF) or its Fourier Transform (FT), the power spectrum (PS), fully characterises a GRF, containing all their (cosmological) information. The PS or the 2PCF specifies the relative contribution of the different scales to the density fluctuation filed. Furthermore, in a GRF all Fourier modes are mutually independent and Gaussian distributed.

A generic GRF can be characterise by 3 variables: the mean of the field, μ, the variance, σ^2 and their amplitude, A, which can, in general, be functions of the scale or Fourier mode, k. In cosmology we usually set the field to fluctuate around 0 by construction of the field, which leave the GRF two variables: the variance and the amplitude. The variance can be understood as how large are the differences between dense and empty regions, whereas the amplitude is related to the clustering strength over all scales or positions.

For cosmologist GRFs are important as the Cosmic Microwave Background observations by WMAP and more recently by Planck suggest that the primordial matter field at the epoch of recombination is Gaussian, or nearly Gaussian. After that, the gravitational evolution of the dark matter field induces correlation among these Gaussian modes plus non-Gaussian correlations. Also, the relation between the galaxy and matter field may also induce extra correlations in the measured galaxy density field.

In these notes I explain and motivate the basic steps to generate a Gaussian random field which can be used as an initial seed for N-body codes, or for generating fast mocks.

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

2 Fourier Transform

The FT decomposes a function of time or position into a functions of the frequencies that makes up the original signal. Such transformation is reversible and fully preserves the original information content.

Given a continuous field $g(\mathbf{r})$, we define its Fourier Transform as,

$$G(\mathbf{k}) \equiv \int d\mathbf{r}\, e^{+i\mathbf{k}\cdot\mathbf{r}} g(\mathbf{r}). \tag{1}$$

Note that even if g is a real function, now G is not necessarily a real function anymore. Its inverse relation, the inverse Fourier Transform (iFT) can be written as,

$$g(\mathbf{r}) \equiv \frac{1}{(2\pi)^3} \int d\mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{r}} G(\mathbf{k}). \tag{2}$$

The choice of the $\pm$ signs in the exponential function of the above two equations, as well as the normalisation factor, $1/(2\pi)^3$, in the iFT expression (instead of the FT) is arbitrary. In these notes I always work in this specific convention, but different expressions for different conventions are equally valid.

We define the 2PCF, ξ, of the g-field as,

$$\langle g(\mathbf{r}) g(\mathbf{r}') \rangle \equiv \xi(\mathbf{r} - \mathbf{r}'), \tag{3}$$

We define the power spectrum, P as just the FT of ξ,

$$P(\mathbf{k}) \equiv \int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \xi(\mathbf{r}) \tag{4}$$

In this fashion one can see that the relation between the G-field and its power spectrum is given by[1],

$$\langle G(\mathbf{k}) G^*(\mathbf{k}') \rangle = (2\pi)^3 \delta^D(\mathbf{k} + \mathbf{k}') P(\mathbf{k}) \tag{5}$$

where δ^D is the Dirac-delta distribution, which can be written as,

$$\delta^D(\mathbf{k}) = \frac{1}{(2\pi)^3} \int d\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{6}$$

Note that the fields g and G do not have the same dimension units. If we think the field g to be dimensionless, the field G would have $[L]^3$ dimensions, as well as P and δ^D.

[1] As a simple exercise one can try to demonstrate the expression below.

3 Discrete Fourier Transform

In real life problems, we will not have access to a continuous $g(\mathbf{r})$ field and we will need to discretise this field in cells. For example, when we measure the density of objects in the sky as a function of position, we need to sub-divide the area or volume of the sky in small cells and then perform the object counting. Therefore the resolution of such field will be given by the size of the cells we have chosen.

We define a continuous g-field to be contained in a cubic box of volume L^3, which we divide in N^3 quadrangular and equal cells, whose volume will be $\Delta V = \Delta L^3_{\rm cell} = (L/N)^3$. In this case the g-field will be sampled in certain positions multiples of the size of the fundamental cell,

$$r_{abc} = \{o\frac{L}{N}, j\frac{L}{N}\, k\frac{L}{N}\}, \tag{7}$$

where the indices $\{a, b, c\}$ run from 0 to $N-1$.

The discrete Fourier Transform (DFT) of a field $f(r_{abc}) \equiv f_{abc}$, will be given by,

$$F_{\ell mn} \equiv \sum_{abc} f_{abc}\, e^{ik_{\ell mn}\cdot r_{abc}}, \tag{8}$$

where $F_{\ell mn} \equiv F(k_{\ell mn})$ and where,

$$k_{\ell mn} \equiv \{\ell\frac{2\pi}{L}, m\frac{2\pi}{L}, n\frac{2\pi}{L}\}. \tag{9}$$

Note that imposing that the field f is discrete in configuration space and given a value of L, defines a set of $k_{\ell mn}$ vectors in the Fourier space. A priori, looking at Eq. 8 one could evaluate F in whatever set of vector they want, but the above given set $k_{\ell mn}$ guarantees the minimum number of vectors containing the same information that the original r_{abc}. This is $k_{\ell mn}$ is an complete orthogonal base for the information given by f evaluated at r_{abc}.

Also, if f belongs to $\Re$, the f field is characterised by N^3 independent entries, whereas the F field, which is $\Im$, contains nearly two times this amount. Since the information contained in both fields have to be the same, not all real and complex values of $F(k_{\ell mn})$ will be independent. In fact one can demonstrate that if the f field is $\Re$, then F will satisfy the following relation,

$$F^*_{\ell,m,n} = F_{-\ell,-m,-n} = F_{N-\ell,N-m,N-n}, \tag{10}$$

where the second equality comes from the fact that adding or subtracting N to the $\{\ell, m, n\}$ indices will not change the argument of the exponential in Eq. 8. The inverse of Eq. 8 reads (iDFT),

$$f_{abc} = \frac{1}{N^3}\sum_{\ell mn} e^{ik_{\ell mn}\cdot r_{abc}} F_{\ell mn}, \tag{11}$$

where N^3 is the normalisation factor, similarly to the $(2\pi)^3$ factor in Eq. 2. In order to relate $F_{\ell mn}$ to $G(\mathbf{k})$ (and f_{abc} to $g(\mathbf{r})$), we can re-write the integration as a summation under the usual transformation $\int d\mathbf{r} \rightarrow \Delta V \sum_i$,

$$G(\mathbf{k}) = \int d\mathbf{r}\, e^{i\mathbf{k}\cdot r} g(\mathbf{r}) \sim \Delta V^3 \sum_{abc} e^{ik_{\ell mn}\cdot r_{abc}} f_{abc} = \left(\frac{L}{N}\right)^3 F_{\ell mn}. \tag{12}$$

The discrete power spectrum resulting from the discrete field F, P^F will read,

$$\begin{aligned}
\langle F_{\ell mn} F^*_{\ell' m' n'} \rangle &= \sum_{abc} \sum_{a'b'c'} e^{ik_{\ell mn}\cdot r_{abc}} e^{-ik_{\ell' m' n'}\cdot r_{a'b'c'}} \langle f_{abc} f_{a'b'c'} \rangle & (13)\\
&= \sum_{abc} \sum_{a''b''c''} e^{ik_{\ell mn}\cdot r_{abc}} e^{-ik_{\ell' m' n'}\cdot (r_{a''b''c''}+r_{abc})} \xi^f_{r''} & (14)\\
&= \sum_{a''b''c''} e^{-ik_{\ell' m' n'}\cdot r_{a''b''c''}} \xi^f_{r''} \sum_{abc} e^{i(k_{\ell mn}-k_{\ell' m' n'})\cdot r_{abc}} & (15)\\
&= N^3 \delta^K_{\ell\ell', mm', nn'} P^F_{\ell mn} & (16)
\end{aligned}$$

where a change of variables has been performed, $r - r' \equiv r''$; we have followed the notation $r'' \equiv r_{a''b''c''}$, and defined the discrete 2PCF $\xi^f_{r''} \equiv \langle f_{abc} f_{a'b'c'} \rangle$, and P^F the corresponding power spectrum, as its DFT as in Eq. 8.

$$P^F_{\ell mn} \equiv \sum_{abc} e^{ik_{\ell mn} r_{abc}} \xi^f_r = \frac{1}{\Delta V} P(k_{\ell mn}). \tag{17}$$

The δ^K is the Kronecker delta function,

$$\delta^K_{\ell\ell', mm', nn'} \equiv \frac{1}{N^3} \sum_{abc} e^{i(k_{\ell mn}-k_{\ell' m' n'})\cdot r_{abc}}, \tag{18}$$

We can relate the discrete P^F defined above with the power spectrum of the continuous G-field as defined in Eq. 5 with the following expression,

$$P^F_{\ell mn} = \frac{1}{\Delta V} P(k_{\ell mn}), \tag{19}$$

which can be derived from Eq. 12. Note that P has dimensions of $[L]^3$, whereas P^F is dimensionless. Therefore the relation between the power spectrum of the continuous G-field and the dimensionless and discrete F field is,

$$\langle F_{\ell mn} F^*_{\ell mn} \rangle = N^3 P^F_{\ell mn} = \frac{N^6}{L^3} P(k_{\ell mn}) \tag{20}$$

4 Fast Fourier Transform algorithm

When performing the computation of DFT described by Eq. 8 and 11, one should naively expect that they perform as a $\sim N^2_{\rm tot}$ processes (where $N_{\rm tot} = N^3$ in this case). However there are algorithms which take advantage of symmetries of the problem to reduce the computational time such that the DFTs perform as $\sim N_{tot}\log(N_{tot})$ processes. This is the case of the algorithms provided by FFTW[2].

When using the FFTW algorithms to perform DFTs from a real field to a complex field there are several points to be considered.

- FFTW provides the un-normalised computation of Eq. 8 and 11, which means that given a discrete field f when we apply a DFT followed by an iDFT, we won't recover f, but f times N^3. One should therefore add an N^3 factor wherever is needed.
- For multidimensional DFT, the input and output vectors should correspond to an array of $r_{abc\ldots}$ and $k_{abc\ldots}$ given by a single index number. Such index follows (in C) the following notation, $aN^{n-1}+bN^{n-2}+cN^{n-3}+\ldots$, where the $a, b, c, \ldots$ indices run from 0 to $N-1$, and where n is the dimension of the problem. For the usual 3-dimensional situations, $n=3$, the input and output index reads as $N^2a+Nb+c$, which runs from 0 to N^3.
- When using real-to-complex and complex-to-real algorithms by FFTW, such ordering is different and can be misleading (see the FFTW documentation for a full explanation). However, one can always work in the general framework of complex-to-complex transformation (at the expenses of some computation time and resources) by imposing the complex part to be 0 in configuration space, and by imposing the $G^*(-\mathbf{k})=G(\mathbf{k})$ condition in Fourier Space. In such cases, take into account that the negative frequencies correspond to some positive frequencies under the $+N$ shift of indices. For instance, one can have $k_\ell=\ell(2\pi/L)$, where ℓ goes from 0 to $N-1$. However, this case is equivalent to have $k_\ell=\ell(2\pi/L)$ for $0\le\ell\le N/2$ and $k_\ell=-(N-\ell)(2\pi/L)$ for $N/2+1\le\ell\le N-1$. Recall these relation when using complex-to-complex DFT transformation and imposing the $G^*(-\mathbf{k})=G(\mathbf{k})$ conditions.

When connecting the notation described in section 3 to physical cosmological parameters you may consider the following points,

- You can think of your physically motivated dark matter or galaxy over-density fields, $\delta(\mathbf{r})$ and $\delta(\mathbf{k})$ as the $g(\mathbf{r})$ and $G(\mathbf{k})$ variables, respectively, in section 2.
- f_{abc} is your discrete $\delta(r_{abc})$ matter or galaxy field in configuration space. Such quantity is dimensionless.
- $F_{\ell mn}$ is the DFT of f_{abc} and is dimensionless. The connection of that variable with $G(\mathbf{k})$ (i.e., with $\delta(\mathbf{k})$)) is given by Eq. 12. In this fashion you can define $H_{\ell mn}$ a discrete variable with the same dimension than $G(\mathbf{k})$ or $\delta(\mathbf{k})$ using the ΔV volume element: $H_{\ell mn}\equiv(L/N)^3F_{\ell mn}$. In this case,

$$\langle H_{\ell mn}H^*_{\ell' m' n'}\rangle L^{-3}=P(\mathbf{k})\delta^K_{\ell\ell',mm',nn'} \tag{21}$$

  or simply,

$$\langle H_{\ell mn}H^*_{\ell mn}\rangle L^{-3}=P(\mathbf{k}) \tag{22}$$

[2]Fastest Fourier Transform of the West: http://fftw.org

where P is the power spectrum of the $\delta(\mathbf{k})$ field.

Recalling the definition of the power spectrum from the delta field,

$$\langle \delta(\mathbf{k})\delta^*(\mathbf{k}')\rangle = (2\pi)^3 \delta^D(\mathbf{k}+\mathbf{k}')P(\mathbf{k}) \tag{23}$$

one can make an obvious connection between Eq. 21 and 23, by making the changes,

$$H_{\ell mn} \rightarrow \delta(\mathbf{k}_{\ell mn}) \tag{24}$$

$$\delta^K_{\ell\ell',mm',nn'} \rightarrow (2\pi)^3 L^{-3}\delta^D(\mathbf{k}+\mathbf{k}'), \tag{25}$$

where the second relation can be actually seen when changing $\int d\mathbf{r} \rightarrow \Delta V \sum_{abc}$ in Eq. 6 and connecting it with the definition in 18.

5 Gaussian Random Fields

We aim to use the above described formalism to generate a real over-density field which is Gaussian and which contains a specific 2PCF or PS signal.

In order to do so, we will first generate the corresponding Fourier space over-density field, $\delta(\mathbf{k})$ (in the above notation $H_{\ell mn}$). Such field will contain a real and complex components such,

$$H_{\ell mn} \equiv H_n = a_n + ib_n. \tag{26}$$

We impose that the field $H_{\ell mn}$ has to contain a specific power spectrum signature (see Eq. 22),

$$\langle H_{\ell mn}H^*_{\ell mn}\rangle = \langle (a_n+ib_n)(a_n-ib_n)\rangle = \langle a_n^2\rangle + \langle b_n^2\rangle = L^3 P(\mathbf{k}) \tag{27}$$

Since the variance of the real and imaginary part should be the same we can relate the variance of a_n and b_n to be $\langle a_n^2\rangle = \langle b_n^2\rangle = L^3P(\mathbf{k})/2$. In this fashion, in order to generate $H_{\ell mn}$ to be a GRF with a power spectrum $P(k)$ you will need to perform the following points.

- Generate a_n and b_n random numbers according to a Gaussian distribution with 0 mean and variance $\sigma^2 = L^3P(k)/2$. Note that the 0 mean is required as by definition $\langle\delta(\mathbf{k})\rangle = 0$. In order to generate a Gaussian random number, g_i with 0 mean and variance 1 you need to generate two flat random numbers between 0 and 1, x_1, x_2, and make $g_i(\mu = 0, \sigma^2 = 1) = \cos(2\pi x_1)\sqrt{-2\ln(x_2)}$. In order to rescale such number to have a different variance just multiply $g(0,1)$ by the desired variance: $\{a_n,\, b_n\} = g(0,1)\sqrt{L^3P(k)/2}$.
- For the $k = 0$ mode both real and imaginary parts need to be 0, $a_0 = 0$, $b_0 = 0$. This required to impose that $\sum \delta(\mathbf{r}) = 0$.
- Generate another real value for the Nyquist mode (that mode with $\ell, m, n = N/2$) and set its imaginary part to 0 in order to ensure the reality of $\delta(\mathbf{r})$.

- Impose the $H_{\ell mn} = H^*_{-\ell,-m,-n}$ condition on the negative $k_{\ell mn}$ vectors (or on those with $\ell, m, n > N/2$) if you are using a generic complex-to-complex DFT, in order to ensure the reality of $\delta(\mathbf{x})$. This corresponds to set $a_{\ell mn} = a_{\ell' m' n'}$ and $b_{\ell mn} = -b_{\ell' m' n'}$ for those conjugate pairs of frequencies, such that $\ell + N/2 = \ell'$, or $\ell + N = \ell'$ if you think in terms of negative frequencies. Note among the first half of modes: $\ell, m, n = 0, \ldots, N/2$, all of them have a corresponding negative frequency, except for 0 and $N/2$. You won't impose the $H_{\ell mn} = H^*_{-\ell,-m,-n}$, when all indices are 0 or $N/2$. In other words, the Nyquist Frequency should not be treated as the complex conjugate of the 0-mode.
- Apply the DFT to convert this $H_{\ell mn}$ field into a $\delta(\mathbf{r})$ field.

If you want to impose the condition that $\delta(\mathbf{x}) \geq -1$[3] the GRF condition will only hold for small variances of the field. This means that the $\delta(\mathbf{x})$ can only be Gaussian for very small perturbations, and in general, as perturbation grows (even linearly) their distribution deviates from the Gaussian condition. Such condition will be necessary if you want to generate a particle distribution which follows the δ-field, as you cannot put less particles than 0 in a cell, which will be the case for $\delta(\mathbf{x}) = -1$.

[3]Note that this feature is not imposed by any of the above conditions.

Exercise A: Compute the power power spectrum

Given a $\delta(r_{abc})$ field compute its angle averaged power spectrum (monopole). You can download the $\delta(r_{abc})$ field from[4]. The size of the box is $L = 1000\,\mathrm{Mpc}/h$. It corresponds to $N^3 = 256^3$ and it is ordered according to $\ell = N^2 i + Nj + k$, where ℓ is the line number in the file (starting with $\ell = 0$ on the first line of the file).

Tips:

- For simplicity use the complex-to-complex DFT from FFTW and set the imaginary part of the vector to 0 before transforming it.
- Once you have the transformed field ($F_{\ell mn}$ in the above notation) you can compute the power spectrum according to Eq. 20.
- In order to perform the $\langle \ldots \rangle$ bin your $F_{\ell mn}$ field according to $|k| = \sqrt{k_\ell^2 + k_m^2 + k_n^2}$, and average over directions. Averaging over directions should be equivalent to do $\langle \ldots \rangle$
- Plot the resulting $P(|\mathbf{k}|)$ as a function of $|\mathbf{k}|$.

Exercise B: Generate a GRF $\delta(x_{abc})$ with a specific power spectrum signal.

Choose an analytic shape for $P(k)$, for eg. $P(k) = Ak^n$ and generate a $\delta(r_{abc})$ which contains such power spectrum. With the code from exercise A, check that such $\delta(r_{abc})$ field has the original $P(k)$ signal. Repeat the the process, but now imprint the shape of the power spectrum given by the following file [5]. If you manage to do so, you will have generated a δ-field galaxy mock with a cosmologically motivated power spectrum.

Exercise C: Anisotropic power spectrum

Modify the code from Exercise A in order to additionally compute the power spectrum quadrupole. Compute the quadrupole of the fields used in the above exercises. Why the quadrupole has this value?

Following what you did in exercise B, generate a GRF $\delta(x_{abc})$, with an anisotropic signal. Compute again the quadrupole. What have changed?

Which has to be the input power spectrum signal in order to have two different fields, the first one isotropic, and the second one anisotropic, but with identical power spectrum monopole signal. Generate them.

[4] https://www.dropbox.com/s/9y9jbvm5009pkwu/deltaxfield.dat?dl=0
[5] https://www.dropbox.com/s/9xf0zfgbve8rlej/Pkfiducial.txt?dl=0

Exercise D: Non-Gaussian δ field

Think how you should modify the above instructions in order to generate a non-Gaussian $\delta(\mathbf{x})$ signal.

You can do so just by generating a non-Gaussian distribution of δ (be aware that $\langle\delta(\mathbf{k})\rangle = 0$ by construction).

A more physically motivated approach would be to generate a primordial potential field Φ such,

$$\Phi(\mathbf{x}) = \phi(\mathbf{x}) + f_{\mathrm{nl}}(\phi^2(\mathbf{x}) - \langle\phi^2(\mathbf{x})\rangle) + g_{\mathrm{nl}}(\phi^3(\mathbf{x}) - \langle\phi^3(\mathbf{x})\rangle) + \dots \tag{28}$$

where $\phi(\mathbf{x})$ is a Gaussian gravitational potential field. Recall that the Poisson Equation relates the δ and Φ fields,

$$\nabla^2\Phi(\mathbf{x}) = 4\pi G\bar{\rho}\delta(\mathbf{x}) \tag{29}$$

or in Fourier Space,

$$\Phi(\mathbf{k}) \propto \frac{\delta(\mathbf{k})}{k^2} \tag{30}$$

Note that if the power spectrum associated to δ is of the type, $P_\delta = Ak^n$ (motivated from inflation), the power spectrum associated to the Φ-field is of the type, $P_\Phi(k) = Ak^{n-2}$, where the $n = 2$ case would correspond a scale-independent $P_\Phi(k)$. CMB observations favour a nearly scale-independent Φ-power spectrum with $n \simeq 1.96$.

In the case of f_{nl}, the skewness of the field would be $\sim f_{\mathrm{nl}}$ and the kurtosis $\sim f_{\mathrm{nl}}^2$, whereas g_{nl} does not introduce any (extra) skewness, but only kurtosis $\sim g_{\mathrm{nl}}$.

Such non-Gaussian components on the $\delta(\mathbf{x})$ field will generate a 3PCF moment or bispectrum.

Generate a non-Gaussian primordial field, δ_{NG}, of the above f_{NL}-type. How does the distribution of δ_{NG} look like with respect to gaussian one for a given scale? How does the power spectrum of both compare? What would you need to do if you want to generate a non-Gaussian primoerdial density field with a given bispectrum signature?

Further reading about primordial non-Gaussianities with a given bispectrum signal, arXiv:1006.5793

For simplicity one could assume an only-f_{nl}-type of non-Gaussianity, which is just a correction of the Gaussian part (i.e. $\phi(\mathbf{x}) \gg \phi^2(\mathbf{x})$). In this case one can write,

$$\nabla^2\Phi(\mathbf{x}) = \nabla^2\phi(\mathbf{x}) + 2f_{\mathrm{nl}}[\phi(\mathbf{x})\nabla^2\phi(\mathbf{x}) + |\nabla\phi^2(\mathbf{x})|] \tag{31}$$

For high peaks of the density field, Dalal et al. 2007 found $\delta_{\mathrm{NG}}(\mathbf{x}, a) \simeq \delta_{\mathrm{G}}(\mathbf{x}, a)[1+2f_{\mathrm{nl}}\phi(\mathbf{x}, a)/g(a)]$, where $g(a) \propto D(a)/a$, which predicts a particular scale dependent bias $\propto k^2$ for biased objects in the late Universe.

Further reading about primordial non-Gaussianities effect on massive halo abundances: arXiv:0710.4560