Fundamental concepts

Statistical Cosmology

First Principles

The density field of the inhomogeneous Universe is not constant everywhere, but it varies with spatial location.

The density values at different locations do not differ much from the mean density

→ they are **perturbations**.

It is usual to define the **density contrast δ (x)**, i.e., the deviation with respect to the mean density (averaged over space):

$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}} - 1$$

During the evolution of the Universe, the density contrast at each point also evolves, either increasing or decreasing, driven by gravity. The process of evolution of the density contrast is called **structure formation**, turning density fluctuations in cosmological and astrophysical structures.

How do initial fluctuations around the mean arise?

from quantum fluctuations of density.

In the quantum universe, there is a large number of random steps, i.e., in the very early Universe the value of density at a given location is changing all the time as the result of a stochastic (random) process. It is not possible to know the value of density at a given location at a given time, in a deterministic way.

We just know that the value is a realization of a probability distribution. Due to the large number of random processes involved, the central limit theorem tell us that the resulting probability distribution is a Gaussian

→ **the quantum density field is a Gaussian random field**.

Later, the **inflationary mechanism** makes the passage from quantum to macroscopic world

→ it produces a density field of macroscopic perturbations - called the primordial perturbations - this field is the **initial condition** for the subsequent time evolution of δ(x), but again its actual value is not known, it is a particular realization among all possible realizations.

Note that the depending on the inflationary model, the Gaussianity of the density random fields may or may not be preserved during inflation → search for possible primordial non-Gaussianity is a test of inflation.

(This is the goal of the measurements of the f_NL parameter in CMB observations)

Now, the value of density at a given location is then (most likely) a value taken from a Gaussian distribution.

So the actual values of δ(x) at each point are not known.

We just know that the density contrast at each point is a random variable, and its value is one among the various possible realizations of a Gaussian distribution,

$$P(\delta_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(\delta_1 - <\delta_1>)^2}{2\sigma_1^2}}$$

The density contrast random field is then described by the parameters of its Gaussian distribution. As we know, a Gaussian distribution has only parameters (its moments): mean and variance.

Note however that there is one Gaussian distribution for each spatial location (hence the subscript in δ above). In principle each location may have its own mean and variance.

Consider a discretization of the density contrast field.

We need N distributions $P(\delta_i)$ (one for each position x; of course the problem is continuous N→infinity).

However, the N variables δ_1 ... δ_N are not independent

→ The value at a point depends on the values of neighboring points (due to the gravitational interactions between them).

So we cannot describe the system by considering N independent Gausian distributions, but we need a N-dimensional Gaussian:

$$f_{\mathbf{x}}(x_1, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

(In our case the vector of k δ random fields on k locations is the random variable x with dimension k, and the k-dimension Gaussian distribution has a k-dim vector of means μ and a k x k covariance matrix Σ)

For example, if there were only 2 random variables (i.e., binning the density field such that it would have only two locations), we would need a 2-dimensional Gaussian;

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

with $$\boldsymbol{\mu} = \begin{pmatrix}\mu_x \\ \mu_y\end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}$$

where ρ is the correlation coefficient $\rho = \sigma_{xy} / (\sigma_x \sigma_y)$

Since the two random variables are not independent, the correlation coefficient is different from zero, and the covariance matrix is not diagonal.

The joint probability of having a value δ_1 at the location 1(called x in the notation above) and having at the same time a value δ_2 at the location 2 (called y in the notation above) is

$P(\delta_1, \delta_2) = P(\delta_1)\, P(\delta_2 | \delta_1)$ (where $P(\delta_2 | \delta_1)$ is the conditional probability)

It seems the complexity of the problem increases with the stochasticity!

If the problem was **deterministic**:

system described by the field $\delta(x)$ → N values

But the problem is **stochastic**:

system not described by the actual values of $\delta(x)$ but by the moments of the N-dim distribution (of which the values of δ are realizations).

The number of moments of an N-dimensional Gaussian is

→ N(N+1) (N values of mean, NxN values in the covariance matrix)

In case the correlations are symmetric, there are only N(N-1)/2 off-diagonal correlation coefficients → a total of N(N+1)/2 elements in the covariance matrix

→ a total of N(N+3)/2 moments.

So the N Gaussian random variables are described by N(N+3)/2 variables (the moments of the distribution).

Fortunately, the complexity is reduced by introducing the

Generalized cosmological principle:

"The universe is statistically homogenous and isotropic"

This means that there are perturbations to the homogeneity but they are described by a probability distribution with a homogeneous and isotropic set of moments.

→ The moments of the distribution do not depend on location or orientation.

(instead of the values of the density field themselves)

Statistical Homogeneity

implies that:

i) *The means do not depend on location* → **all N means are identical** (one for each random variable δ_i).

Can we measure the means of the distributions?

If we had a sample from the distribution, we could just measure its average in the usual way (summing the values and dividing by their number) - this is called the ensemble average. This statistic (the ensemble average) is known to give an unbiased estimate of the mean of a distribution (if the sample is large enough).

Problem: However we only have one realization - which is the Universe itself - instead of a full sample (unless there are parallel universes), i.e., we can only measure one value of δ in a given location, and we cannot repeat the experiment to get more values.

Solution: We assume that the whole Universe provides a representative set of all possibilities, i.e., the Universe includes in itself all possible realizations of the distribution.

In other words, distant parts of the field in separate parts of the Universe are independent of each other. The values of δ there are not correlated with the values of δ here. Those values are independent realizations of the same distribution that provides the values here (the distributions are the same due to statistical homogeneity).

In this way we can have access to different realizations of the same distribution, and get a sample

→ we can then make spatial averages instead of ensemble averages in order to find the moments. This is called the **ergodic hypothesis**.

Using the ergodic hypothesis, we can easily compute the mean of the distribution of δ.

From its definition, the values of δ are:

$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}} - 1$$

the mean value of the distribution can then be computed by the ensemble (now equivalent to spatial) average of the values of δ across the spatial field.

The result follows immediately:

$$\langle \delta \rangle = 0$$ (*Note: <> denotes ensemble or spatial averages*)

This means that the value of δ on any point of the Universe is a random value around the mean δ = 0.

This also implies that the amplitude of cosmological perturbations will not be given by the mean value of their distribution but by the variance of the distribution (a larger variance allows for the possibility of producing realizations with larger values of δ).

The N-dimensional distribution is then essentially described by the NxN covariance matrix. Its elements are:

Variance: i.e. the N terms of the diagonal (also called auto-correlation)

Covariances: i.e., the N(N-1) off-diagonal terms (also called the cross-correlations)

Statistical homogeneity further implies that:

ii) *The variances do not depend on location* → **all N terms of the diagonal are identical.**

Can we measure the variances of the distributions?

Yes, by measuring a sample of values of δ at different locations and computing the variance with the usual statistic:

$$\frac{1}{k-1}\sum_{i=1}^{k}\delta_i^2 = \left\langle \delta^2 \right\rangle$$

iii) *The correlation coefficients do not depend on location* → this does not mean that all N(N-1) terms of the off-diagonal are identical. It means that **the correlation coefficient between a pair of points separated by a given vector is the same for all pairs separated by identical vectors.**

Statistical Isotropy

implies that:

iv) *The correlation coefficients do not depend on orientation* → **the correlation coefficient between a pair of points separated by a given vector modulus (i.e. a given distance, irrespective of the orientation) is the same for all pairs separated by the same distance.**

Eg: $\sigma_{14} = \sigma_{37}$ (covariance between locations 1 and 4 and between locations 3 and 7)

Can we measure the variances of the distributions?

Yes, by measuring a sample of values of δ at different locations and computing the covariance using only pairs of points at the same separations:

$$\frac{1}{n_{\text{pairs}}} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \delta_i \delta_j \delta_D(|i-j| - d) = \langle \delta_i \delta_j \rangle (d)$$

(the Dirac delta indicates the sum only includes points at a separation d from each other)

In summary, the density contrast random field (discretized in N positions of a regular grid) is described by N values:

- 1 variance (auto-correlation)
- N-1 covariances (since the condition iv reduces the original N(N-1) correlation coefficients to N-1)

Two-point functions

Correlation Function

The N-1 covariances form a function known as the **2-point correlation function** :

$$\xi_{\delta\delta}(r) = \langle \delta(x)\delta^*(x') \rangle$$ (r=|x -x' |)

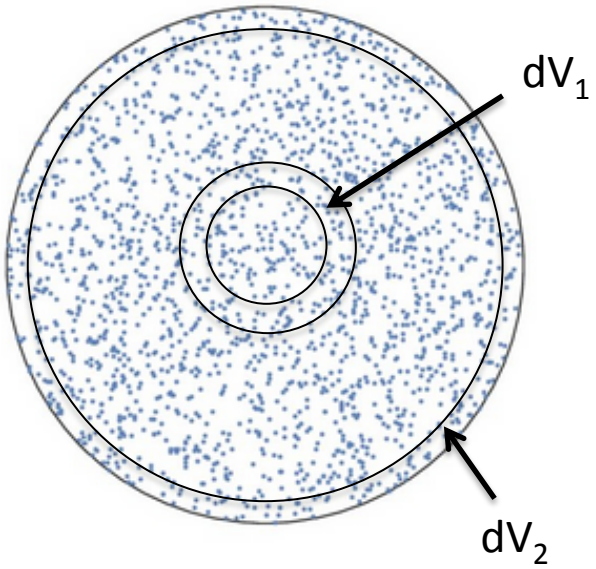
(δ* accounts fot the possibility of having complex fields)

These N quantities contain the full cosmological information of a Gaussian δ(x) map.

The randomness aspect and the generalized cosmological principle, make that ***the most natural spatial quantities to use in the treatment of the inhomogeneous Universe are not locations but separations between locations.***

The correlation function of the density contrast field contains all the statistical information on the Gaussian density contrast field. In particular it tell us the conditional probability of having a value δ_2 at a location "2" separated by "r" from a location "1" where there is a value δ_1 → it describes the **clustering** properties of the field.

For example, consider that we have sampled the field δ on N locations (for example by measuring the positions x,y,z of N galaxies, assuming the galaxies trace the locations of the overdensities)



Case of uncorrelated distribution

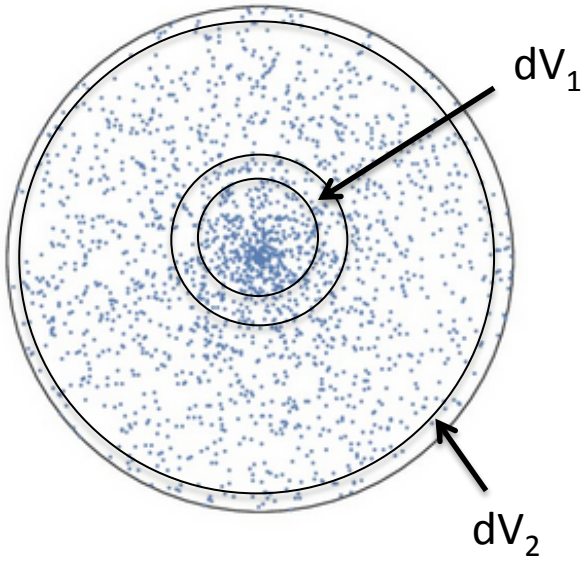
There are N galaxies on the full volume V, and a number density of n=N/V

The probability of having a galaxy in the shell volume dV_1 is given by the number of galaxies within that volume divided by the total number of galaxies N:

$dP_1 = n\, dV_1 / N = dV_1 / V$

The probability of having a galaxy in the shell volume dV_2 is independent of dP_1 :

$dP_{2u} = n\, dV_2 / N = dV_2 / V$



Case of correlated distribution

In the case of correlated distribution, the probability of having a galaxy in the shell volume dV_2 depends on dP_1 .

In other words, the value of dP_2 depends on the correlation between the locations 1 and 2,

i.e., it depends on the correlation at the separation r_{12} :

$$dP_{2c} = n\, dV_2 \,(\, 1+\xi(r_{12})\,) \,/\, N = dV_2\,(1+\xi(r_{12})\,) \,/\, V$$

So, the number of galaxies found is no longer just a function of the size of dV_2

If there is a:

correlation, $\xi > 0 \rightarrow dP_{2c} > dP_{2u}$
(anti-)correlation, $\xi < 0 \rightarrow dP_{2c} < dP_{2u}$

The total number of galaxies as function of r, on the full volume, is given by N times the integral of the probability dP(r):

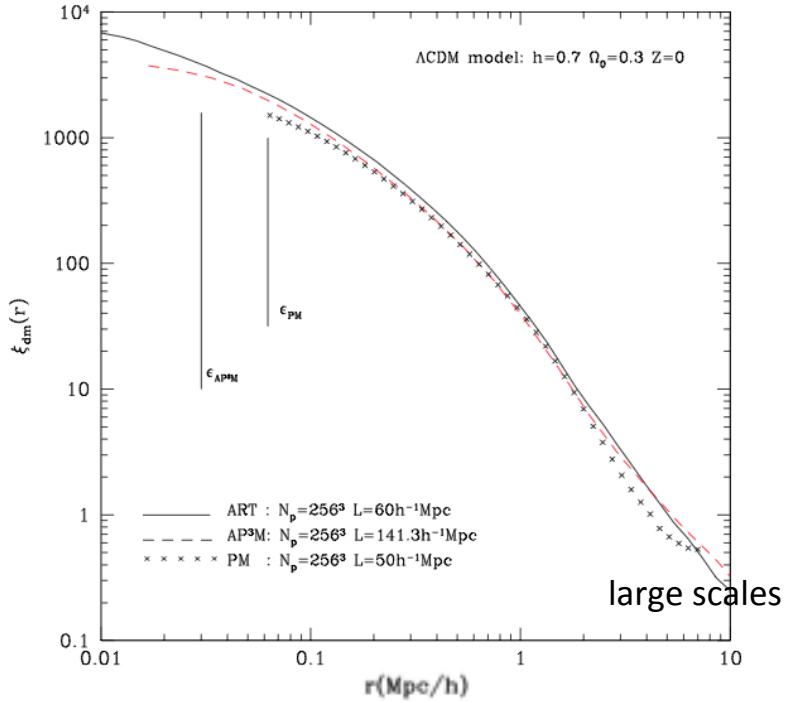
$$N(r) = \int n\,(1+\xi(r))\,dV/dr\,dr$$

Note that n dV(r) is a "distance function", the number of objects per distance bin dN (r)

(the use of shell volumes dV is very practical to obtain a function of r)

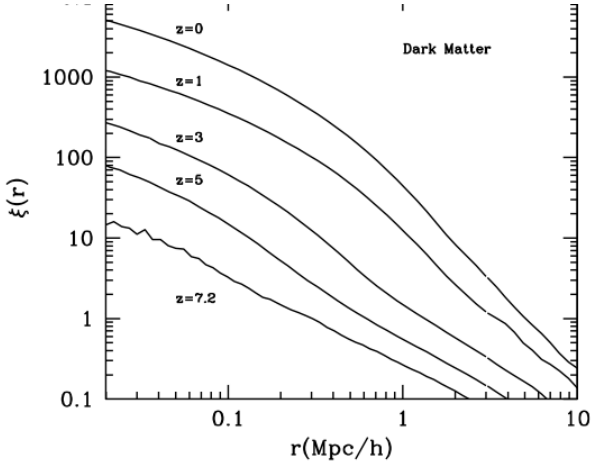
In the uncorrelated case $N(r) \sim r^3$, but in general the slope will be different, depending on the correlation function slope ξ(r) → the number is higher on a highly correlated area (small separations).

The dark matter correlation function predicted in the ΛCDM model is positive and decreases with separation:



large scales

The amplitude of the correlation function naturally increases with structure formation (as the clustering of matter increases) → **it decreases with redshift**.



Correlation as an excess ratio between the clustered and the random cases:

Now, if we compare the probabilities dP(r) we just found for the correlated and the uncorrelated cases,

$dP_{2u} = n\, dV_2 / N$

$dP_{2c} = n\, dV_2\, (\, 1+\xi(r)\,) / N$

we see that **1+ξ(r) is given by the ratio of the probabilities**, i.e., by the ratio of the two "distance functions" (the number of galaxies as function of r):

$$\mathbf{1+\xi(r) = N_c\,(r) / N_u\,(r)}$$

(see homework)

Correlation Function in Fourier space

The correlation coefficient of 2 points separated by r tells us about **structure** - the central property of the inhomogeneous universe that we want to describe. It quantifies the clustering of the density field (**the "degree of collapse"**) - the **formation of structure**.

For example, if there is correlation on all separations up to a separation r and then the correlation drops, it shows that (on average) there are overdensity regions from x to (x+r), i.e. halos of size r

However the relation between correlation as function of separation, and size of the overdensity is not a one-to-one relation → from this example, we see that we need to know the correlation at various separations to find out if there is an overdensity of a given size.

We would like to have a function that directly shows the clustering amplitude on a given size. Is this possible?

Let us consider the Fourier transform of the density contrast field

$$\delta_k = \frac{1}{V}\int \delta(x)\, e^{-ik.x}\, d^3x \qquad\qquad \delta(x) = \frac{V}{(2\pi)^3}\int \delta_k\, e^{+ik.x}\, d^3k$$

This defines a set of Fourier modes k (3d vectors), with associated sizes 2π/k (or wave numbers)

(Note that the factor (2π/k)^3 is needed in this convention of Fourier transforms because we are writing the plane waves as ikx and not i2πkx)

(Note that V accounts for finite volume, and δ(x) and δ_k are dimensionless)

Let us compute the correlation function in this approach (i.e., compute the 2-point correlation function in k-space) :

$$\langle \delta_k \delta^*_{k'} \rangle = \frac{1}{V}\left\langle \int d^3x\ \delta(x)\, e^{i\vec{k}.\vec{x}} \frac{1}{V}\int d^3x'\ \delta^*(x')\, e^{-i\vec{k}'.\vec{x}'} \right\rangle = \frac{1}{V}\int d^3x\ e^{i\vec{k}.\vec{x}} \frac{1}{V}\int d^3x'\ e^{-i\vec{k}'.(\vec{x}+\vec{y})}\, \xi(|\vec{y}|) =$$

where y is the separation vector between x and x', and for fixed x the integration over x' is the same as an integration over y.

Continuing,

$$= \frac{1}{V}\int d^3x\; e^{i\vec{x}.(\vec{k}-\vec{k}')} \frac{1}{V}\int d^3y\, \xi(|\vec{y}|)\, e^{-i\vec{k}'.\vec{y}} = \frac{(2\pi)^3}{V}\delta_D(\vec{k}-\vec{k}')P_\delta(|\vec{k}|)$$

where the first term is a Dirac delta (the Fourier transform of f(x)=1),

$$\int e^{i(k-k').x}\, d^3x = (2\pi)^3\, \delta_D(k-k')$$

and the second term is the Fourier transform of the correlation function, which is called the **power spectrum** P(k) (it is its definition). Note that due to isotropy it only depends on the modulus of the k-mode vector.

So the result is,

$$\langle \delta_k \, \delta'_k{}^* \rangle = \langle \delta_k^2 \rangle = \frac{(2\pi)^3}{(2\pi/k)^3} \, \delta_D(\vec{k} - \vec{k}') P_\delta(|\vec{k}|) = k^3 P_\delta(k) \; = \Delta^2(k)$$

Note that the power spectrum P(k) has dimensions of volume, and $\Delta^2(k) = k^3 P(k)$ is called the **dimensionless power spectrum**, also known as the power spectrum per interval of ln(k).

The important result we obtained here is that

the correlation function of the Fourier transform of the density contrast field is the Fourier transform of the correlation function in real space multiplied by the Fourier volume k^3 and by a Dirac delta function, i.e.

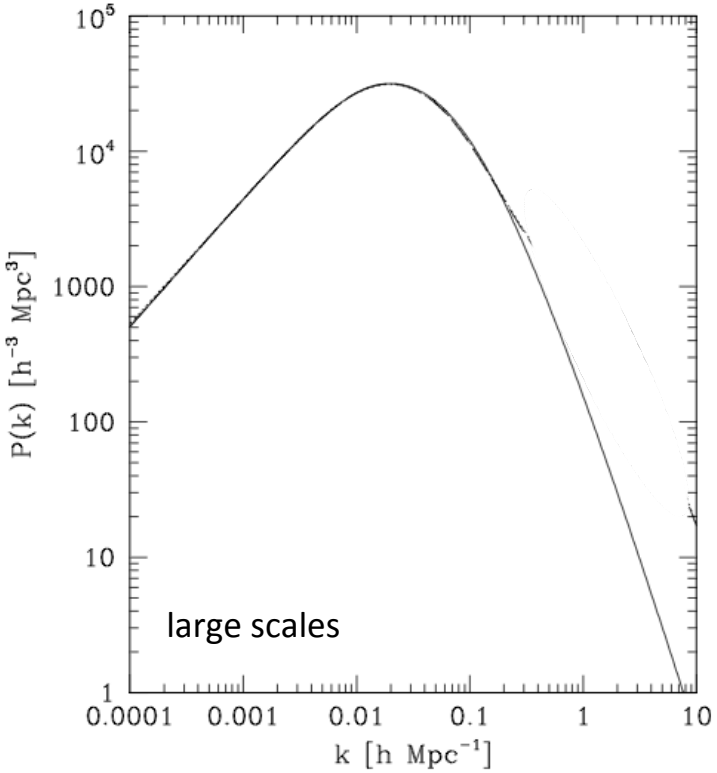
it is the dimensionless power spectrum multiplied by a Dirac delta function

The presence of the Dirac delta makes the coefficients δ_k to be independent, i.e.,

the correlation between two different modes k, k' is zero.

So, while the original correlation function describes the density contrast field using a set of N-1 non-independent covariance (cross-correlations) variables (plus one variance) that depend on separation on the real space,

the power spectrum describes the same field using a set of N independent variance (auto-correlations) variables in the harmonic space: the set of $\langle \delta_k^2 \rangle$



Power spectrum vs. Correlation function

Both descriptions - in real and Fourier space - have the same information.
Both are valid to describe the cosmological field.

The fact that the dimensionless power spectrum contains variances instead of covariances, means that it gives directly the information of a mode - or ***scale*** *- (instead of relying on separation between points).*

Note that A small value of k is called a large scale
A large value of k is called a small scale

because the inverse of the scale - $2\pi/k$ - corresponds to a physical size

So the value of the dimensionless power spectrum on a given Fourier mode, is the variance on that scale, i.e., the degree of clustering (the clustering amplitude) that exists on that scale of the Universe on average.

(Remember it is a moment of a distribution, so it does not mean that all regions of the Universe of that size will have that density contrast, it only means that their values will be realizations of that distribution with that variance).

(Note again that the dispersion of a random variable of mean zero is a direct indication of its amplitude - and not the mean! -)

Let us consider now the power spectrum as the basic quantity and compute the correlation function from it:

We need to compute the inverse Fourier transform of the power spectrum:

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) e^{-ik.r} d^3k$$

The correlation function is real so we just need to consider:

$$\mathrm{Re}(e^{-ikr\cos\theta}) = \cos(kr\cos\theta)$$

and the power spectrum is isotropic (it depends only on the radius |k| → we can integrate over the angular part:

$$\int_0^\pi \cos(kr\cos\theta)\sin\theta\, d\theta = -\frac{\sin(kr\cos\theta)}{kr}\Big|_0^\pi = 2\frac{\sin kr}{kr}$$

(in spherical coordinates the integral element is $d^3k = k^2 \sin\theta\, dk\, d\theta\, d\phi$

The result is:

$$\xi(r) = \frac{1}{2\pi^2} \int_0^\infty P(k) \, \frac{\sin kr}{kr} \, k^2 \, dk$$

This means that the correlation function is a ***filtered linear combination of the power spectrum*** → one separation r is a combination of various scales k → **k are the independent and fundamental cosmological scales, the separations r are not independent.**

There is not a one-to-one correspondence between separation and scale (unless the filter in the integral, also called window function, is very narrow).

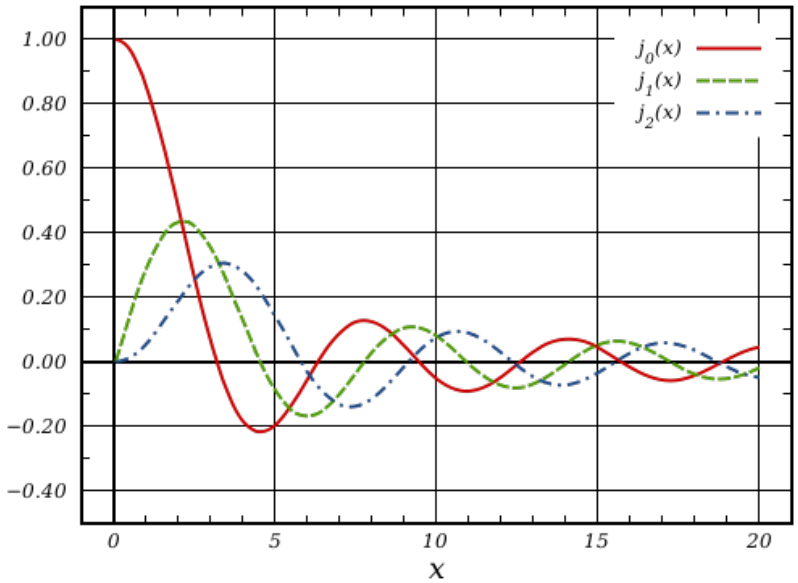
The filter (the function that multiplies k^2 P(k) in the integral) is the **spherical Bessel function** of the first kind for n=0 : j_0 (kr)

$$j_n(x) = (-x)^n \left(\frac{1}{x} \frac{d}{dx} \right)^n \frac{\sin(x)}{x}$$

$$j_0(kr) = \frac{\sin kr}{kr}$$

$$j_1(kr) = \frac{\sin kr}{(kr)^2} - \frac{\cos kr}{kr}$$

$$j_2(kr) = \frac{3\sin kr}{(kr)^3} - \frac{3\cos kr}{(kr)^2} - \frac{\sin kr}{kr}$$



The shape of j_0 (the solid line) shows that most of the contribution for the correlation at a separation r -ξ (r)- comes from larger scales: k < 2.6/r (the range where the contribution is large, with filter amplitude > ~0.2)

In summary: power spectrum and correlation function have the same information, but the N components of the power spectrum are independent and give directly the amplitude of clustering as function of scale, while the N components of the correlation function do not.

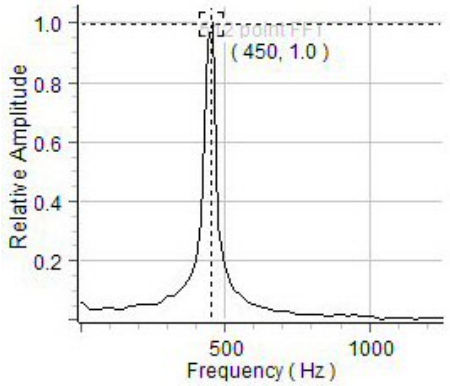
Even though the 2-pt correlation function is highly correlated and does not give direct information on an individual scale, it is a useful quantity to consider because

it is defined in real space → it can be **measured directly** from data measured in the sky.

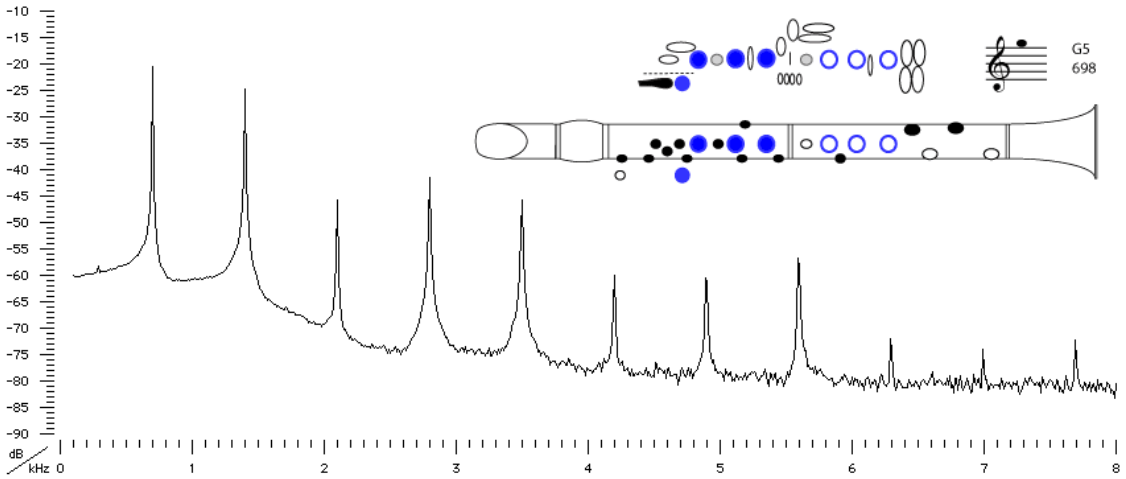
(The power spectrum needs to be **estimated** from data in an indirect procedure).

The fundamental modes in the harmonic space (i.e., the wavenumber or scale k) are thus the natural choice to define the cosmological scales.

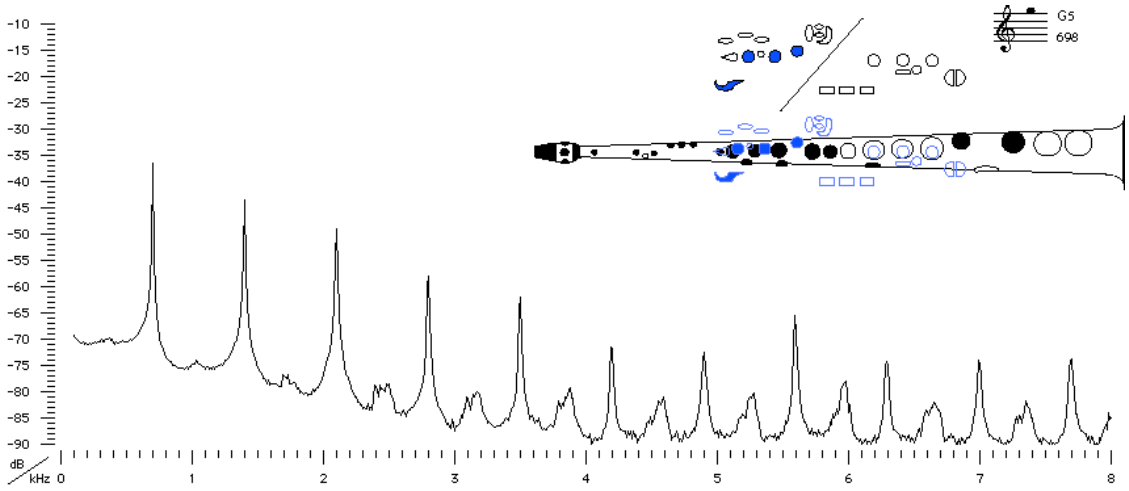
Note: an analogy, is that the notes (A,B,C,D,...) produced by a musical instrument are not independent (they are like the separations), each one contain various fundamental notes defined by a tuning fork (which are the fundamental ones, like the cosmological scales). **Each instrument has a sound spectrum, which in fact is a Power spectrum**:



G tuning fork is independent from the other fundamental notes (the “scales”)



G clarinet is a linear combination of the fundamental notes



G saxophone is a different linear combination of the fundamental notes

(The sound power spectrum defines the timbre of the instrument)

Smooth spatial distribution: counts in cells and sigma_8

Alternatively to using discrete quantities (i.e. separations r between discrete locations x, x'), the clustering properties in the real space can be determined using a **smoother measure of density**:

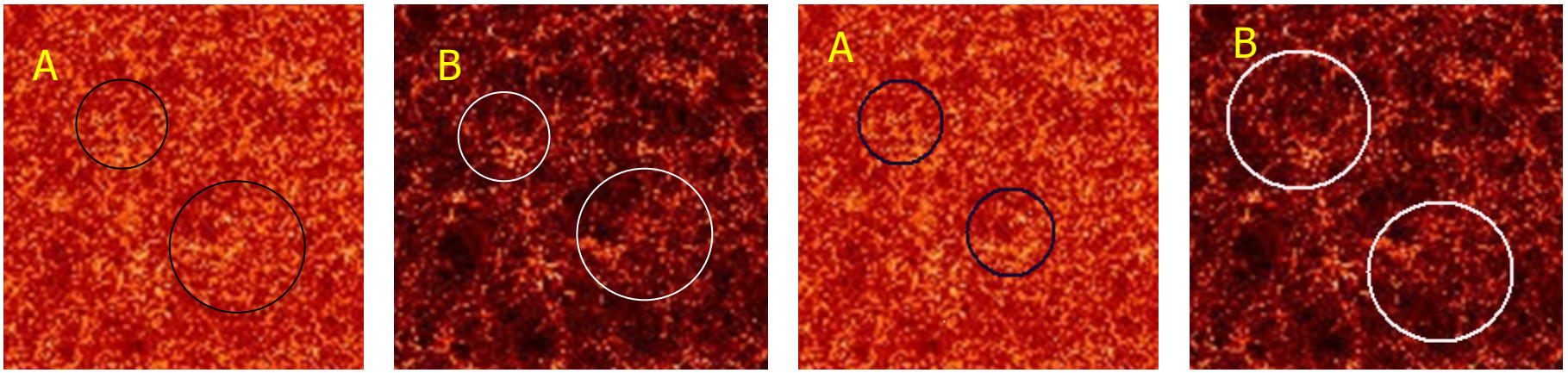
the variance of number counts in cells

Placing cells of a fixed size R on a δ map (discrete or continuous) allows to smooth the map on a scale R, defining a δ_R as a convolution of δ(x) with a window function (a filter) of size R → **δ_R is a weighted average of δ in a cell of size R.**

We can then compute the variance of this δ_R on cells R across the whole map.

Doing this for N values of R, we can define a vector of variances of δ_R.

Example: Consider two density maps A and B and two different scales R (shown by the circles).



Compute δ_R in each map for the two different values of R, obtaining 4 quantities.

Then compute the variance of each of those quantities, by moving the circles on the maps. The result is:

i) The variances in B are larger than in A (for both scales R), because B has more density contrast than A. In B the circles can fall in high-density regions or in low-density regions → large variance. While in A all regions are more similar → **B has more structure than A.**

ii) Placing the larger circle (for both A and B) it is more likely to find similar regions along the maps than with the smaller circle → the variance decreases with R → the smallest cell R to approach zero variance defines the homogeneity scale → there is no structure above that scale.

Now, since the variance of δ_R is a second-order moment, it is certainly related to the power spectrum.

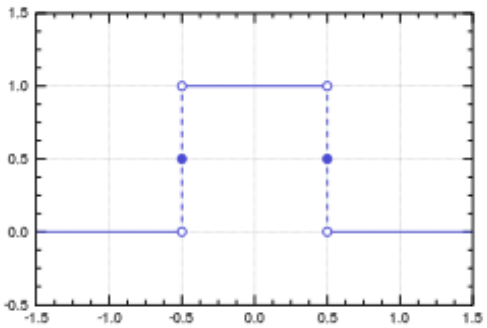
Let us derive that relation.

First, how can we write a theoretical expression for the smooth density δ_R ?

Let us consider a top-hat window function W_R, i.e., a filter of constant amplitude.

δ_R can be written as the convolution of δ with the top-hat:

$$\delta_R(x) = \int d^3y\, \delta(y)\, W_R(|x-y|)$$



The Fourier transform of the smooth field is simply the product of the Fourier transforms of δ and the top-hat:

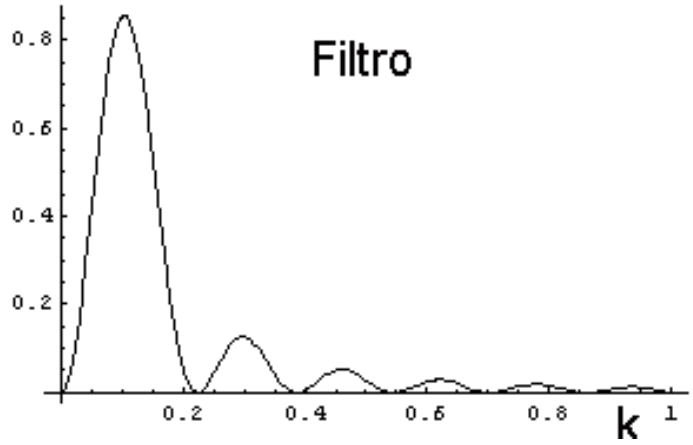
$$\delta_R(k) = \delta(k)\, W_R(k)$$

The variance of the smooth density is then,

$$\sigma_R^2 = \left\langle \delta^2(k)\, W_R^2(k) \right\rangle = \frac{1}{(2\pi)^3} \int d^3k\, W_R^2(k) P(k)$$

i.e., it is a filtered integral of the power spectrum, where the filter is the square of the Fourier transform of the top-hat $W_R(k)$:

$$W_R(k) = 3\frac{\sin kR - kR\cos kR}{(kR)^3}$$



This filter is very diferent from the j_0 Bessel function.
It is relatively narrow and peaked at k ~ 2π/R.

We conclude that a vector of σ^2_R (for various cell sizes R) is a linear combination of the power spectrum amplitudes, just like the correlation function was.

However, its components are less correlated than the correlation function ones → since the filter is very peaked, there is roughly a one-to-one correspondence between R and scale k.

For this reason, the value of σ^2_R gives a good indication of the clustering amplitude at the scale R (like the power spectrum also does).

As we will see later, to compute structure formation (i.e., the time evolution of the density contrast field), we need an initial condition for the density contrast field δ(x,t).

As we know, the field is fully represented by a 2-pt quantity. So the initial condition must be the value of a 2-pt function at a fixed time (redshift). In particular, the amplitude of an initial 2-pt function at a given scale is a **comological parameter of the inhomogeneous Universe.**

There are two alternative parameters that set the primordial amplitude of the density contrast field:

- The amplitude of the primordial power spectrum at a large scale k = 0.02 h/Mpc→ parameter A_s

- The amplitude of today's power spectrum (z=0) at a smaller scale R = 8 Mpc/h → parameter σ_8 ("sigma eight")

From early times to late times, the power spectrum evolves in amplitude and shape → the two amplitude parameters are related; the relation between the values of A_s and σ_8 depends on all cosmological parameters.

- *Why is a large scale [k=0.02 h/Mpc → R ~ 300 Mpc/h] used for early-times normalization?*

The scale factor is small → there is no resolution to access small scales

- *Why is R=8 Mpc/h used for late-time normalization?*

It is the scale where the observed dark matter power spectrum P(k,z=0) has amplitude ~1 → It is the threshold that separates linear scales (the larger ones) from non-linear scales (the smaller ones) today → so the value of σ_8 in a given model shows immediately the level of clustering in the universe today, compared with a $\sigma_8 = 1$ reference universe.

Correlation function estimator

The 2-pt correlation function of a continuous density contrast field δ(x) - with positive and negative values around a zero mean - may be measured in a δ(x) map by:

- computing the covariances of δ(x) $\xi_{\delta\delta}(r) = \langle \delta(x)\delta^*(x') \rangle$

- computing the covariances of the smoothed $\delta_R(x)$

However, we usually do not have a continuous field:

We may have measurements of δ(x) on some points of the field (or on some points of some other related field, not directly δ(x))

We may observe the positions of galaxies, and assume their concentration traces the δ(x) field → i.e., there are more galaxies where δ(x) is larger.
With this assumption we can define a $\boldsymbol{\delta_g(x)}$, which is basically N_gal (x).

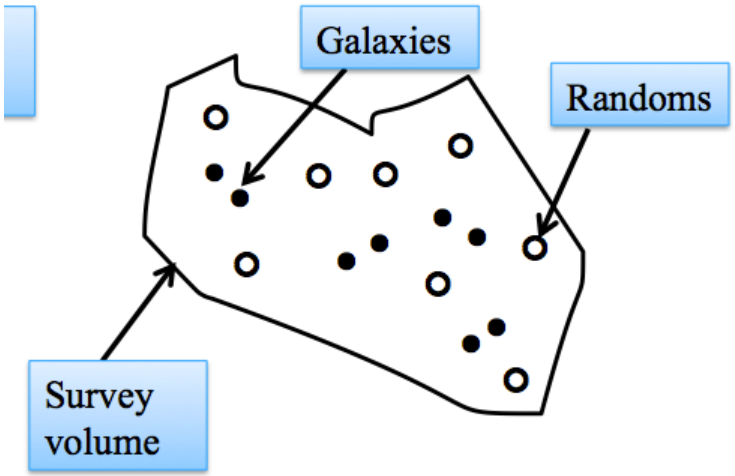
the number of galaxy pairs as function of separation can be written as schematically as 1x1 + 1x0 + 1x0 + 1x1 + …. → it is "a kind of" $\langle \delta_g(x)\, \delta_g(x) \rangle$

Note however that the number of galaxies at a location is 0 or 1; it cannot be negative → the **N_gal (x) is not entirely equivalent to a δ(x) field**

In other words, the correlation found from this method is not normalized, its absolute value is not correct. **What we can do, to be able to use this information, is to compare the N_pairs (x) with the N_pairs (x) from a uncorrelated field.**

The ratio of the two has the correct information.

This method requires that we build a sample of mock galaxies (the "randoms"), in the same survey volume and geometry, with the same spatial sampling as the data sample, but with uncorrelated positions, (i.e. with P(1) independent of P(2)).



Using this we can measure:

DD (r) - number of galaxy-galaxy pairs as function of separation
RR (r) - number of mock-mock pairs as function of separation
DR (r) - number of galaxy-mock pairs as function of separation

Several **estimators** of the correlation function can be defined, based on different ways of making the data-random comparison:

$$1 + \xi_1 = \frac{\langle DD \rangle}{\langle RR \rangle}$$

$$1 + \xi_2 = \frac{\langle DD \rangle}{\langle DR \rangle},$$

$$1 + \xi_3 = \frac{\langle DD \rangle \langle RR \rangle}{\langle DR \rangle^2},$$

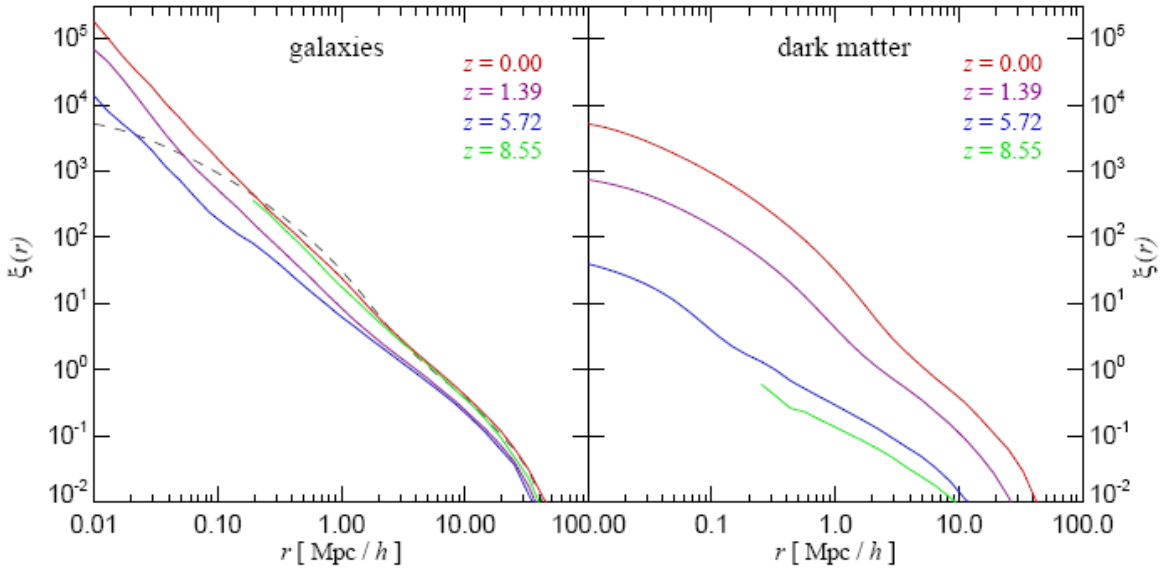
$$1 + \xi_4 = 1 + \frac{\langle (D - R)^2 \rangle}{\langle RR \rangle^2}$$

The 4 estimators have different noise properties.

Number 4 has the best signal-to-noise ratio.

The typical result obtained for the correlation function (of galaxies positions) is a power-law, with slope γ= 1.7 (where r_0 is a critical separation that depends on the type of galaxies, a typical value is r_0 ~ 5 Mpc/h)

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}$$



Note that the correlation function obtained from galaxy surveys is different from the one measured directly on the δ(x) field (from simulated dark matter fields using N-body simulations), which is not a power-law slope.

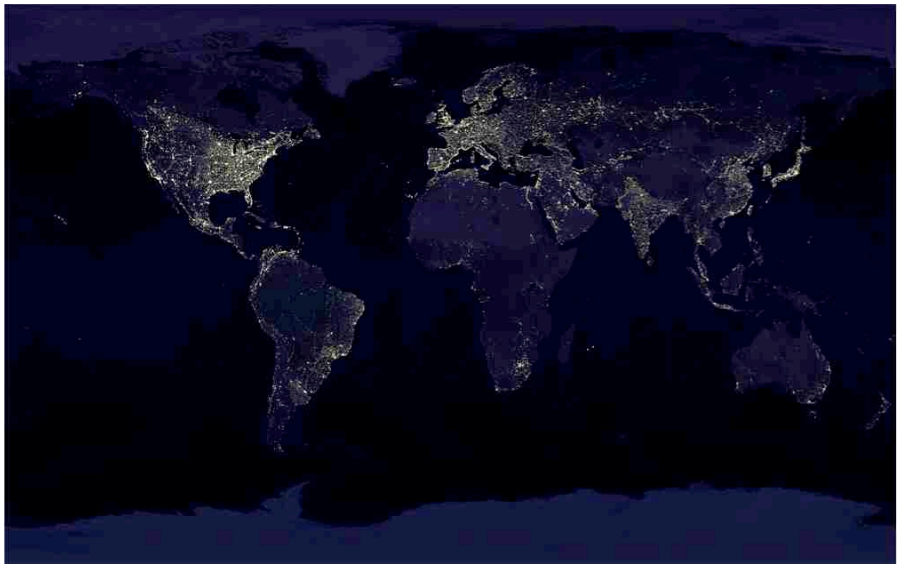
This shows that there is an important **bias** between the spatial distributions of galaxies and dark matter, i.e.,

$$\delta_g(x) = b(r,z)\,\delta(x) \quad \text{(in a linear approximation)}$$

The bias “b” is not a constant. It can be modeled as function of redshift and scale, introducing additional nuisance parameters.

(It is known to be larger for brighter galaxies - like the galaxies in clusters - → there is also an environment dependence)

So, light only follows matter in an approximate way



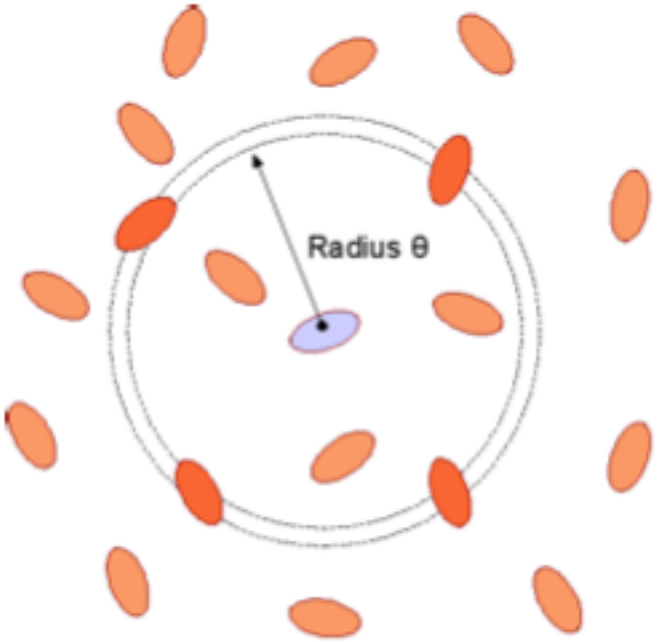
Power spectrum estimator: shot noise

Measurements of discrete galaxies positions can also be used to estimate the power spectrum of the underlying continuous δ field.

Consider N galaxies (particles) of mass m=1 in a volume V, corresponding to a mean density

$$\bar{\rho} = \frac{M}{V} = \frac{N}{V}$$

Assume there is no galaxy bias, i.e.,
galaxy positions trace perfectly
the mass distribution



The density ρ at a location takes values
0 (at a point x with no particle)
or 1 (at a point x with a particle).

(This also implies that the volume occupied by 1 galaxy is V_g=1 → ρ= 1/1)

With this set up, the density contrast may be written using the Dirac delta function (which will be convenient later on):

Note that the integral of the Dirac delta is 1 (over the full infinity range), or zero (if the sum range does not contain the peak).

$$\delta(x) = \frac{\rho(x) - \bar{\rho}}{\bar{\rho}} = \frac{\Sigma_{i=1}^{N} \delta_D(x - x_i)}{N/V} - 1$$

Now, in order to compute the power spectrum, we need first to Fourier transform δ(x):

$$\begin{aligned}
\tilde{\delta}(\vec{\kappa}) &= \int \mathrm{d}^3 x \, \exp(-\mathrm{i}\vec{\kappa}\vec{x})\delta(\vec{x}) \\
&= \int \mathrm{d}^3 x \, \exp(-\mathrm{i}\vec{\kappa}\vec{x}) \left[\frac{V}{N}\sum_{i=1}^{N} \delta_{\mathrm{D}}(\vec{x} - \vec{x}_i) - 1\right] \\
&= \frac{V}{N}\sum_{i} \exp(-\mathrm{i}\vec{k}\vec{x}_i) - (2\pi)^3\delta_{\mathrm{D}}(\kappa)
\end{aligned}$$

where the Dirac delta sets $x=x_i$ in the plane wave

and compute the correlation function in Fourier space $< \delta(k)\delta^*(k') >$

$$\begin{aligned}\langle\tilde{\delta}(\vec{\kappa})\tilde{\delta}^*(\vec{\kappa}')\rangle &= \frac{V^2}{N^2}\sum_{i,j}\langle\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_j)\rangle + (2\pi)^6\delta_{\mathrm{D}}(\vec{\kappa})\delta_{\mathrm{D}}(\vec{\kappa}') \\ &- (2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa})\frac{V}{N}\sum_i\langle\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\rangle - (2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa})\frac{V}{N}\sum_j\langle\exp(\mathrm{i}\vec{\kappa}'\vec{x}_j)\rangle\end{aligned}$$

To evaluate the 1st term - we may separate the terms i=j from i≠j :

$$\begin{aligned}\frac{V^2}{N^2}\sum_{i,j}\langle\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_j)\rangle &= \frac{V^2}{N^2}\sum_{i=j}\langle\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_i)\rangle \\ &+ \frac{V^2}{N^2}\sum_{i\neq j}\langle\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_j)\rangle\end{aligned}$$

Note: What is the sum of a 'bracketed' quantity?

The ensemble average of a random variable 'x' is the sum over all its realizations (all elements in a sample).

If we do not have a sample but know the probability function of 'x' we could generate a sample and average.
Or, more precisely (and without recurring to numerical methods), we need to sum over 'x' multiplied by its probability → it is a weighted sum.

In general an ensemble average of a function f is then

<f> = integral (dx f(x) p(x))

or, in 2 dimensions:

$$< f >= \int \int \, dx_1 \, dx_2 \, p(x_1, x_2) \, f(x)$$

So in order to proceed with the derivation and compute the ensemble averages in this first term, we need first to write the probabilities.

In the case i=j, we need to compute <exp(-ikx$_i$) exp(ik'x$_j$)>
It is a 1-dimensional problem, the ensemble average is an integral over x_i

What is the probability of having a particle in x_i?

It is just $P(x_i)=1/V$

So now we can proceed and get:

$$\frac{V^2}{N^2}\sum_{i=j}\langle \exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_i)\rangle = \frac{V^2}{N^2}\sum_{i=j}\int \mathrm{d}^3 x_i \,\frac{1}{V}\exp(-\mathrm{i}(\vec{\kappa}-\vec{\kappa}')\vec{x}_i)$$
$$= \frac{V^2}{N^2}\frac{N}{V}(2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa}-\vec{\kappa}')$$

(where the integral gives a Dirac delta and the sum is over the N cases i=j)

In the case i≠j, we need to consider the joint probability of having two particles, one in x_i and another in x_j.

$$\frac{V^2}{N^2} \sum_{i \neq j} \langle \exp(-i\vec{\kappa}\vec{x}_i) \exp(i\vec{\kappa}'\vec{x}_j) \rangle$$

This is the probability of x_i times the conditional probability of x_j given x_i.

If they are independent this is just $P(xi,xj) = P(xi)\, P(xj) = (1/V)^2$

But if there is a correlation, the probability of finding a particle in x_j depends on having or not a particle in x_i.

If they are (positively) correlated the joint probability is larger than $(1/V)^2$:

$$P(xi,xj) = P(xi)\, P(xj|xi) = (\, 1+\xi\,(|xi-xj|)\,) \,/\, (V^2)$$

This is, of course, the definition of correlation function.

So **the ensemble average introduces in a natural way the correlation function of the continuous field in the derivation**.

$$
\begin{aligned}
&= \frac{V^2}{N^2}\sum_{i\neq j}\int \mathrm{d}^3 x_i \mathrm{d}^3 x_j \; \frac{1}{V^2}[1+\xi(|\vec{x}_i-\vec{x}_j|)]\exp(-\mathrm{i}\vec{\kappa}\vec{x}_i)\exp(\mathrm{i}\vec{\kappa}'\vec{x}_j)\\
&= \frac{N-1}{N}\int \mathrm{d}^3 x_i \mathrm{d}^3 z \exp(-\mathrm{i}\vec{\kappa}'\vec{z})\exp(-\mathrm{i}(\vec{\kappa}-\vec{\kappa}')\vec{x}_i)[1+\xi(|\vec{z}|)]\\
&= +(2\pi)^6\delta_{\mathrm{D}}(\vec{\kappa})\delta_{\mathrm{D}}(\vec{\kappa}') + \frac{}{}(2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa}-\vec{\kappa}')\int \mathrm{d}^3 z \exp(-\mathrm{i}\vec{\kappa}'\vec{z})\xi(z)
\end{aligned}
$$

The sum has N(N-1) cases and (1+ξ) separates in 2 terms:

- an integral over the plane waves → giving 2 delta functions

- and the Fourier Transform of the correlation function (where z=|xi-xj|).

Going back to the expression for $< \delta(k)\delta^*(k') >$

The 2nd term has nothing to compute, $(2\pi)^6 \delta_{\rm D}(\vec{\kappa})\delta_{\rm D}(\vec{\kappa}')$

and **the 3rd and 4th terms**

are similar to the i=j part of the 1st term:

$$
\begin{aligned}
(2\pi)^3\delta_{\rm D}(\vec{\kappa})\frac{V}{N}\sum_i \langle \exp(-{\rm i}\vec{\kappa}\vec{x}_i)\rangle &= (2\pi)^3\delta_{\rm D}(\vec{\kappa})\frac{V}{N}\sum_j \langle \exp({\rm i}\vec{\kappa}'\vec{x}_j)\rangle \\
&= (2\pi)^3\delta_{\rm D}(\vec{\kappa})\frac{V}{N}N\int {\rm d}^3 x_i \, \frac{1}{V}\exp(-{\rm i}\vec{\kappa}\vec{x}_i) \\
&= (2\pi)^6\delta_{\rm D}(\vec{\kappa})\delta_{\rm D}(\vec{\kappa}')
\end{aligned}
$$

Putting **all terms together**:

The first term of the i≠j term and the 2nd, 3rd and 4th terms are all double Dirac deltas, and all cancel each other.

The result is then the i=j term, plus the second term of the i≠j term :

$$\langle \tilde{\delta}(\vec{\kappa})\tilde{\delta}^*(\vec{\kappa}')\rangle = (2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa}-\vec{\kappa}')\left[\frac{V}{N}+\int \mathrm{d}^3 z \exp(-\mathrm{i}\vec{\kappa}'\vec{z})\xi(z)\right]$$

$$= (2\pi)^3\delta_{\mathrm{D}}(\vec{\kappa}-\vec{\kappa}')\left[\frac{V}{N}+P(|\vec{\kappa}|)\right]$$

We derived that the correlation function in the Fourier space is the power spectrum plus a constant term (V/N).

(Instead of being just the power spectrum, as we had seen before)

This is a general property of any power spectrum estimated from a discrete spatial distribution.

Why is now the result P(k)+V/N instead of P(k) ?

The extra contribution comes from the i=j term of the derivation → it is a term of auto-correlation and not a term of covariance → it has no cosmological information related to a scale, because a scale needs a separation → it is a monopole term.

In our derivation, starting from measurements in the real space, it would be very easy to avoid ending up with this term → we just needed to discard auto-correlations in the estimator → consider only pairs of galaxies where the 2 galaxies are different.

But when we estimate directly the power spectrum from a discrete map,in a more indirect way, the result will always implicitly include this monopole → this term cannot be avoided:

$$\hat{P}(k) = P(k) + \frac{V}{N}$$

Notice that, since a scale k is a linear combination of all separations r within the window function, the i=j monopole affects the estimated amplitudes of P(k) for all scales → **it is an overall constant shift in amplitude.**

However, the fact that the monopole amplitude is given by V/N tells us that its amplitude will decrease in future surveys → larger V and larger N
(with V being limited while N can tend to ∞)

So, **the galaxy power spectrum estimator is not biased**:

$$\left\langle \hat{P}(k) \right\rangle = P(k) + \left\langle \frac{V}{N} \right\rangle = P(k)$$

The monopole adds uncertainty to the estimated power spectrum, but does not bias the measurement. It does not to be subtracted, it is part of the noise and contributes to the error bars. The monopole term is known as the shot noise (also called discreteness noise).

If we want to limit the shot noise in a future survey, we should build a deeper survey rather than a wider one (i.e., increase the density of galaxies n = N/V).

Projected two-point functions

The 2-pt functions that we saw until now are defined in the cosmological volume, i.e., in a 3D density contrast field.

We can also define angular two-point functions, which are function of two-dimensional (angular) separations and are obtained by ***projecting the 3D 2-pt functions on the sky*.**

A projected 2-pt function is more directly measured in the sky than the original 3-dimensional one → we can always measure an angular separation, but not a radial separation (which needs redshift information) → in general what we really observe is a map of the projected density.

Angular correlation function

An angular correlation function is a 2D correlation function, i.e., obtained by ***projecting the 3D correlation function on the sky***.

A projected quantity may be written in general as a weighted (filtered) integral over the third dimension:

$$F(\theta) = \int d\chi\, g(\chi)\, F(f_K(\chi)\theta, \chi)$$

where,

- the 3D coordinates are $\vec{x}$ = ($f_K(\chi)\,\theta_x$, $f_K(\chi)\theta_y$, χ), with

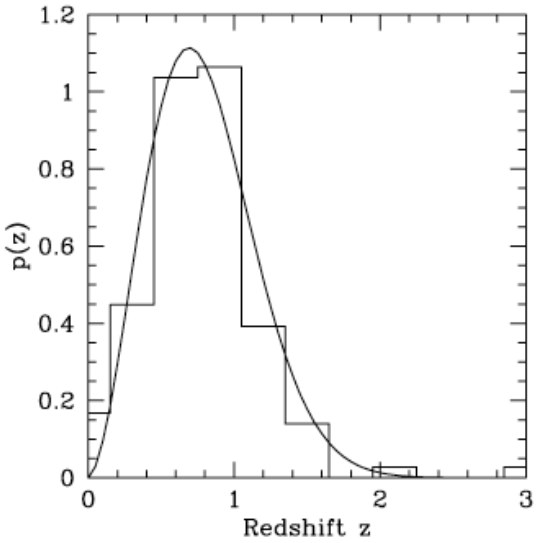
  χ is the radial coordinate (comoving)

  θ_x is the angular separation (in the x direction) to a reference axis (the line-of-sight)

  $f_K(\chi)\,\theta_x$ is the comoving physical separation corresponding to that angular separation, i.e, the angular separation times the comoving angular diameter distance.

- g(χ) is the weight function used in the projection: for example the redshift distribution of the density tracers (galaxies). In this case coordinates χ (redshift z) with more galaxies contribute more to the integral.

$$g(\chi)\, d\chi = p(z)\, dz$$



(A filter (or window or weight function) is needed to account for the various contributions to a given position θ on the sky).

Let us then use this general form to write the projected correlation function:

$$w(\vartheta) = \int d\chi\, g_1(\chi) \int d\chi'\, g_2(\chi')\xi(|\vec{x} - \vec{x}'|)$$

which is function of separation $\vartheta = |\vec{\theta} - \vec{\theta}'|$.

In the projection, each angular separation has contributions from pairs with elements at any radial distance.

We may aproximate it by considering that

- since the 3D correlation function is a decreasing function of separation, **only physically close pairs contribute** (i.e., close in the 3D space and not only in the projected sky) → we consider only pairs with $\chi \sim \chi'$

- the window function has a slow variation in redshift: $g(\chi) \sim g(\chi')$

This is called the Limber approximation.

In the Limber approximation, the two window functions are function of χ and can be written inside the first integral.

Notice that the product of the two window functions is g^2 only in the case that they are not correlated. In general, they are correlated by the correlation function itself → the joint probability $P(g_1, g_2)$ is a conditional probability → there is source clustering and so we should write:

$$w(\vartheta) = \int d\chi\, g^2(\chi)\, [1 + \xi(|\vec{x} - \vec{x}'|)] \int d\chi'\, \xi(|\vec{x} - \vec{x}'|)$$

But this is a second-order effect (order ξ^2). **To first order, the angular correlation function is linear in the 3D correlation function:**

$$w(\vartheta) = \int \int d\chi\, d\chi'\, g^2(\chi)\, \xi(|\vec{x} - \vec{x}'|)$$

The 3d correlation function $\xi = \langle \delta(x)\delta(x') \rangle$ is the Fourier transform of the power spectrum, and so we can write

$$w(\vartheta) = \int \int d\chi\, d\chi'\, g^2(\chi) \int \frac{d^3k}{(2\pi)^3} P(\vec{k}, z)\, e^{-i\vec{k}.(\vec{x}-\vec{x}')}$$

Note the power spectrum evolves in time, and so it also depends explicitly on the redshift z (which is related to χ).

The 3D vector scale can be decomposed in a 2D transversal and a 1D longitudinal component,

$$\vec{k} = (\vec{k}_\perp, k_3)$$

and we can write,

$$w(\vartheta) = \int d\chi\, g^2(\chi) \int d\chi'\, e^{-ik_3(\chi-\chi')} \int \frac{d^3k}{(2\pi)^3} P(\vec{k}, z)\, e^{-i\vec{k}_\perp.(\vec{\theta}-\vec{\theta}')} f_K(\chi)$$

In this expression, there remains no dependence on χ' → the integral over dχ' (or over d(χ-χ') which is the same) is a Dirac delta function $2\pi\,\delta_D\,(k_3) \rightarrow k_3 = 0$, i.e:

$$w(\vartheta) = \int d\chi\, g^2(\chi) \int \frac{d^2 k_\perp}{(2\pi)^2} P(k_\perp, z)\, e^{-i\vec{k}_\perp . \vec{\vartheta} f_K(\chi)}$$

This is the result, also called the Limber equation - the relation between the angular 2-pt correlation function and the power spectrum.

It shows that only scales in the plane contribute to the angular 2-pt function.

Note on notation - the standard notation is:

ξ(r) - 2-pt correlation function
w(ϑ) - 2-pt angular correlation function
P(k) - power spectrum
C(l) - angular power spectrum

Angular power spectrum: flat sky

The angular power spectrum is the transform of the angular correlation function in the harmonic space.

For **flat-sky (valid for small fields)**, plane-waves $e^{i\vec{\ell}.\vec{\theta}}$ are an orthonormal basis of functions that can be used to make the Fourier transform.

This introduces the 2D angular scale 'l', the reciprocal of the real-space angular separation θ.

The relation between the Fourier angular scale and the real-space angular separation is:

$$\theta = 2\pi/l$$

→ the scale l=100 corresponds to a separation of 3.6 deg
→ the scale l=1000 corresponds to a separation of 21.6 arcmin

Now, the **Fourier transform of the 2-pt angular correlation function** is:

$$C(\ell) = \int d^2\vartheta \, e^{i\vec{\ell}.\vec{\vartheta}} \, w(\vartheta)$$

Inserting in the Limber equation, we find the relation between the angular power spectrum and the power spectrum:

$$C(\ell) = \int d\chi \, g^2(\chi) \int \frac{d^2k_\perp}{(2\pi)^2} P(k_\perp, z) \int d^2(\vartheta) \, e^{-i\vec{k}_\perp.\vec{\vartheta} f_K(\chi)} \, e^{i\vec{\ell}.\vec{\vartheta}}$$

The last integral is a Dirac delta: $(2\pi)^2 \, \delta_D(\vec{\ell} - k_\perp f_K(\chi))$

This means that 'l' only depends on the transversal components of k, and not on the full 3D k vector,

and allows us to make the dk integration setting

$$k\_transverse = l / f_K(\chi).$$

The result is:

$$C(\ell) = \int d\chi\, g^2(\chi)\, P\left(\frac{\ell}{f_K(\chi)}, z\right)$$

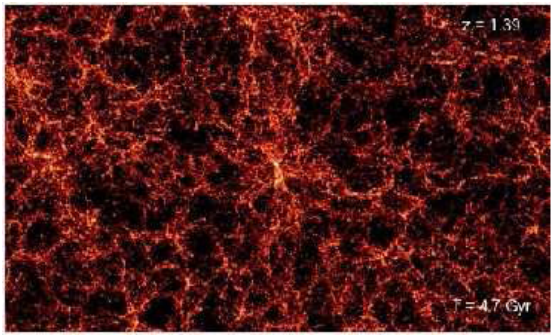
This shows that the amplitude of C for a given angular scale l, is a weighted sum of the amplitudes of P at scales $l/f_K(\chi)$

i.e., at different redshifts, the scales k that contribute to the same angular scale l are different.

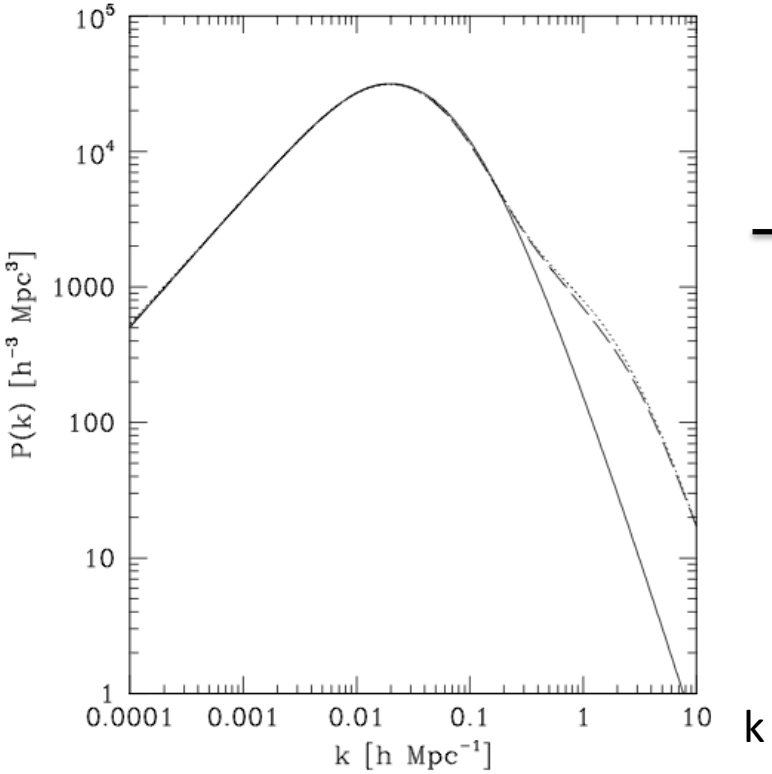
Due to statistical isotropy, the correlation functions only depend on the separation modulus → C(l) is only function of the modulus of 'l', as P(k) was function of the modulus of 'k'.

Decomposing a map in plane waves: the dark matter density contrast

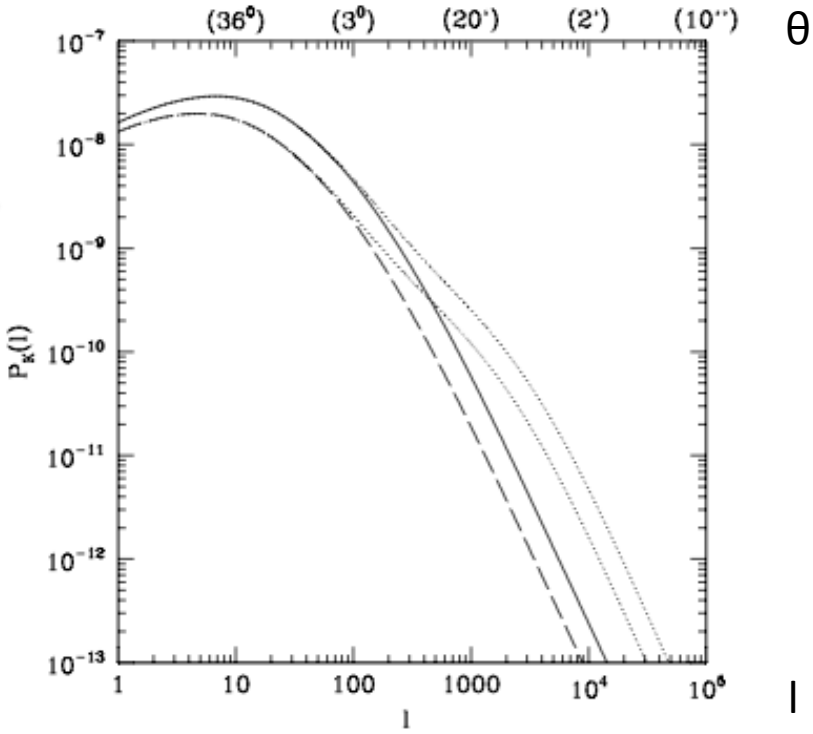
(Note that obviously we still need to study structure formation to find the power spectra, we are just looking at the relations between the various power spectra and correlation functions)



3D matter power spectrum $P_{\delta\delta}(k)$



2D angular matter power spectrum $C_{\delta\delta}(l)$



Angular power spectrum: spherical sky

In the spherical full-sky, the flat-sky approximation is not valid for large scales → plane waves are no longer an orthonormal basis.

A better basis are the spherical harmonics Y_{lm}

Since we are in 2D there are 2 indexes to these functions, just like for Fourier modes $l=(l_x,l_y)$. For spherical harmonics the indexes are called (l,m) and are associated with spherical coordinates θ and φ.

The spherical harmonics form an orthonormal set of functions on the spherical surface:

$$\int d\hat{\mathbf{n}} Y_\ell^{m*}(\hat{\mathbf{n}}) Y_{\ell'}^{m'}(\hat{\mathbf{n}}) = \delta_{\ell\ell'}\delta_{mm'}$$

$$\sum_{\ell m} Y_\ell^{m*}(\hat{\mathbf{n}}) Y_\ell^{m}(\hat{\mathbf{n}}') = \delta(\phi-\phi')\delta(\cos\theta - \cos\theta')$$

$$Y_\ell^{m*} = (-1)^m Y_\ell^{-m}$$

The spherical harmonics are defined from the associated Legendre polynomials P_{lm}

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi}$$

which in turn are defined from the ordinary Legendre polynomials P_l

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x)$$

which are the solutions of Legendre's differential equation

$$\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}P_n(x)\right] + n(n+1)P_n(x) = 0$$

and can be written as,

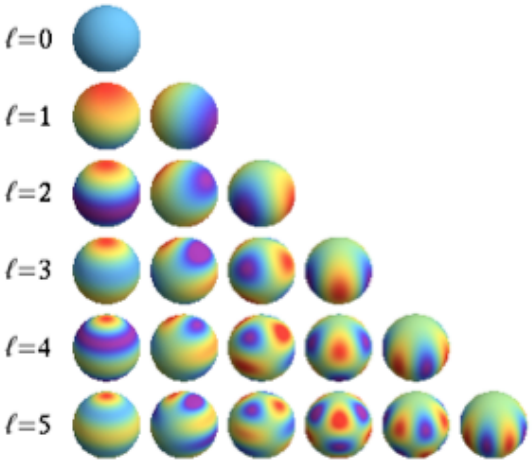
$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left[(x^2-1)^n\right]$$

Contrary to cartesian coordinates (where the range of l_x and l_y are independent), in spherical coordinates the range of l and m are not independent: **for each 'l', 'm' runs from -l to l.** → there are 2l+1 values of 'm' for each 'l' → summing over 'm', for a fixed 'l' gives the closure relation:

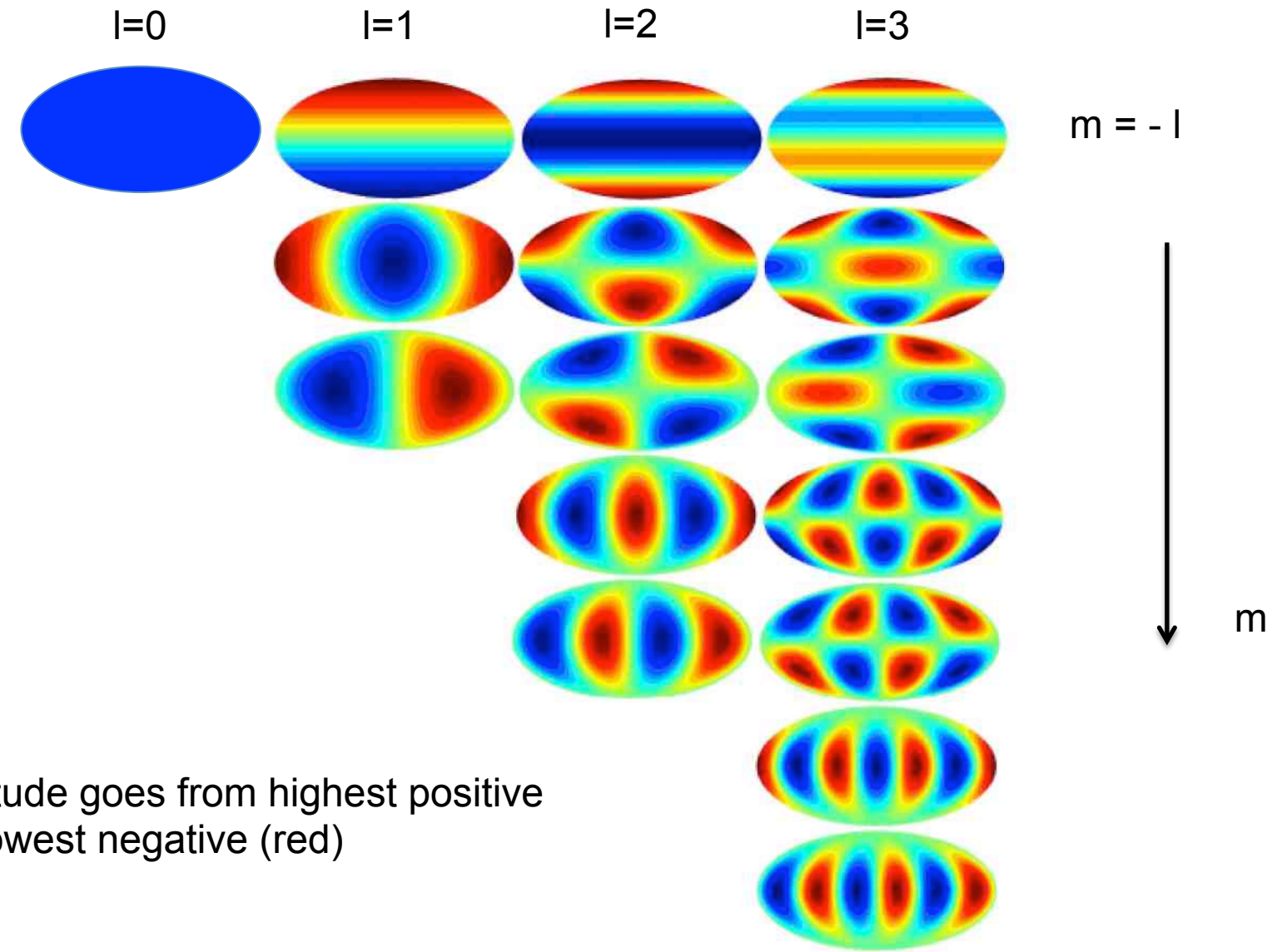
$$\sum_m |Y_{\ell m}(\theta,\phi)|^2 = \frac{2\ell+1}{4\pi}$$

The first spherical harmonics are:

$P_0^0(x) = 1$	$Y_{00} = \sqrt{\frac{1}{4\pi}}$
$P_1^1(x) = -(1-x^2)^{1/2}$	$Y_{11} = -\sqrt{\frac{3}{8\pi}}\sin\theta e^{i\phi}$
$P_1^0(x) = x$	$Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta$
$P_2^2(x) = 3(1-x^2)$	$Y_{22} = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta e^{2i\phi}$
$P_2^1(x) = -3(1-x^2)^{1/2}x$	$Y_{21} = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta e^{i\phi}$
$P_2^0(x) = \frac{1}{2}(3x^2-1)$	$Y_{20} = \sqrt{\frac{5}{4\pi}}(\frac{3}{2}\cos^2\theta - \frac{1}{2})$



The first spherical harmonics look like this:



The amplitude goes from highest positive (blue) to lowest negative (red)

We see that the (2l+1) 'm' configurations of spherical harmonics for a given 'l' have a similar pattern → they divide the surface of a sphere in (2l) regions of equal area.

l = 0 is constant → monopole
l = 1 is a gradient between 2 poles (the maximum and a minimum) → dipole
(the different basis configurations show the gradient along latitude or along longitude)
l = 2 → quadrupole
l = 3 → octopole

Notice that *the relation between the spherical harmonics angular scale and the real-space angular separation is not unique.*

As an approximation, we may consider that the 2l regions of equal area that divide the surface of the sphere are placed along the meridians. In that case, the width of each region at the equator is

$$\theta = 2\pi/(2l)$$

and so a good indicator is $\theta \sim \pi/l$ (different from the flat sky case)

→ scale l=2 corresponds to a separation of 90 deg (the quadrupole)
→ scale l=100 corresponds to a separation of 1.8 deg
→ scale l=220 corresponds to a separation of 49 arcmin (CMB first peak)
→ scale l=2500 corresponds to a separation of 4.2 arcmin (Planck last data point)

Now, the **spherical harmonic transform of the delta field** is:

$$\delta(\theta,\varphi) = \sum a_{\ell m} Y_{\ell m}(\theta,\phi)$$

The **multipole coefficients** a_{lm} are the equivalent to δ_k in Fourier space (to be precise, this notation a_{lm} is usually reserved for the transform of the CMB temperature contrast δT)

The **correlation function of the transform of the delta field** is $\langle a_{lm}\, a_{l'm'} \rangle$.
As we saw for the 3D case, the derivation can be made by inserting the inverse transform, which makes appear the correlation function in real space, and various spatial integrals that will result in Dirac deltas and the power spectrum.

The result is:

$$< a_{\ell m} a_{\ell' m'} >= \delta_D(\ell - \ell')\, \delta_D(m - m')\, C_{\ell m}$$

where, once again, the Dirac deltas show the independence of the power spectrum scales.

The correlation function is isotropic → it depends only on the angular separation (l ←→ θ), and not on the direction (m ←→φ). We can thus integrate over m, and get:

$$\sum_m < a_{\ell m} a_{\ell m} > = \frac{2\ell+1}{4\pi} C_\ell$$

This defines the isotropic angular power spectrum, as an average over all directions

$$C_\ell = \frac{4\pi}{2\ell+1} \sum_m < a_{\ell m} a_{\ell m} >$$

This has an impact on observations → the power spectrum on large scales (low multipoles l) corresponds to an average over a small number of independent functions → the large scales are measured with much less precision than small scales →there is a fundamental limit of statistical uncertainty on large scales (called the **cosmic variance**).

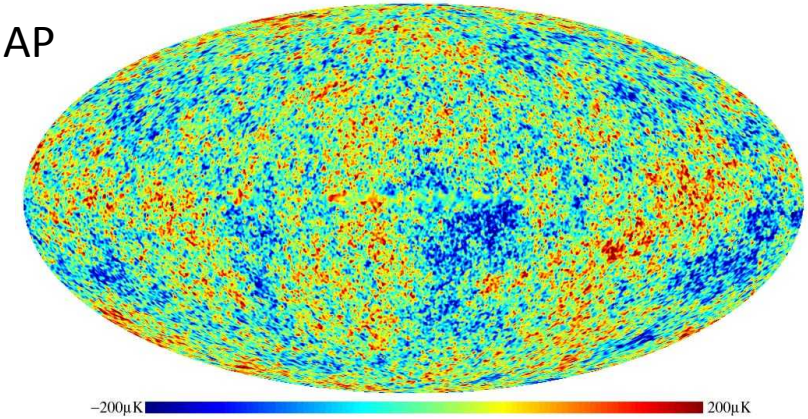
Finally, we can also write the **correlation function in real space**, as function of the isotropic angular power spectrum,

$$<\delta_i \, \delta_j> = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell P_\ell(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j)$$

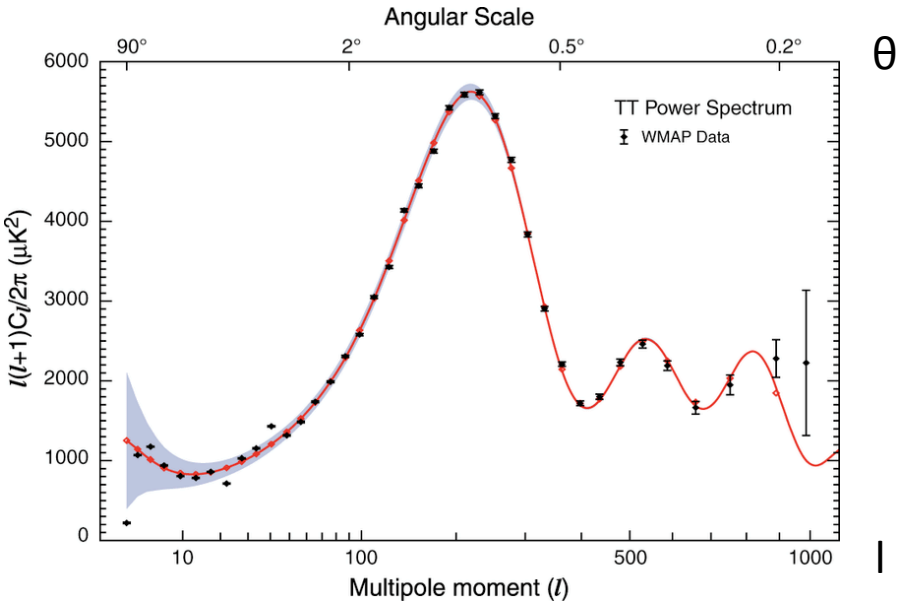
Decomposing a map in spherical harmonics: the CMB temperature contrast

The observed map is one realization (i.e., one specific m for each l) of the theoretical C_l computed from the cosmological model, which is $\langle a_{lm}\, a_{lm} \rangle$ (any m, all are equivalent)

WMAP



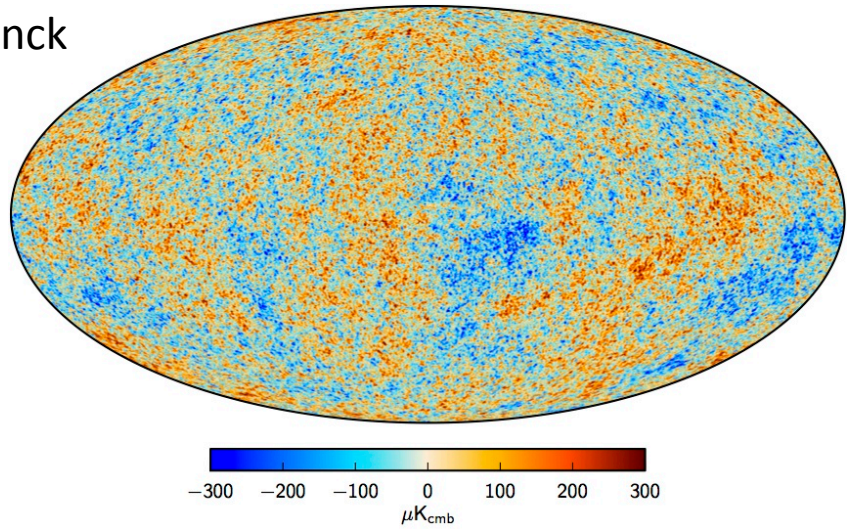
2D angular temperature dimensionless power spectrum $l(l+1)\, C_{TT}(l)$



Cosmological perturbations are functions defined as perturbations around a mean value → its own mean value is zero → the monopole is zero for density contrast fields.

So in cosmology, the monopole is not used and the 'l' range is 1,2,…∞

Planck

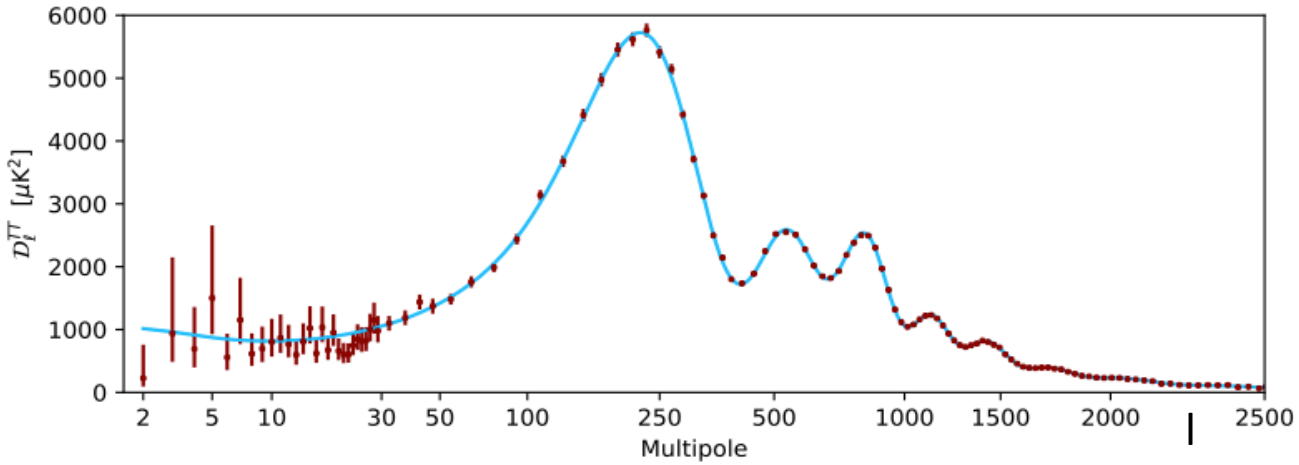


The Planck map was obtained to higher order of the spherical harmonics than the WMAP one.

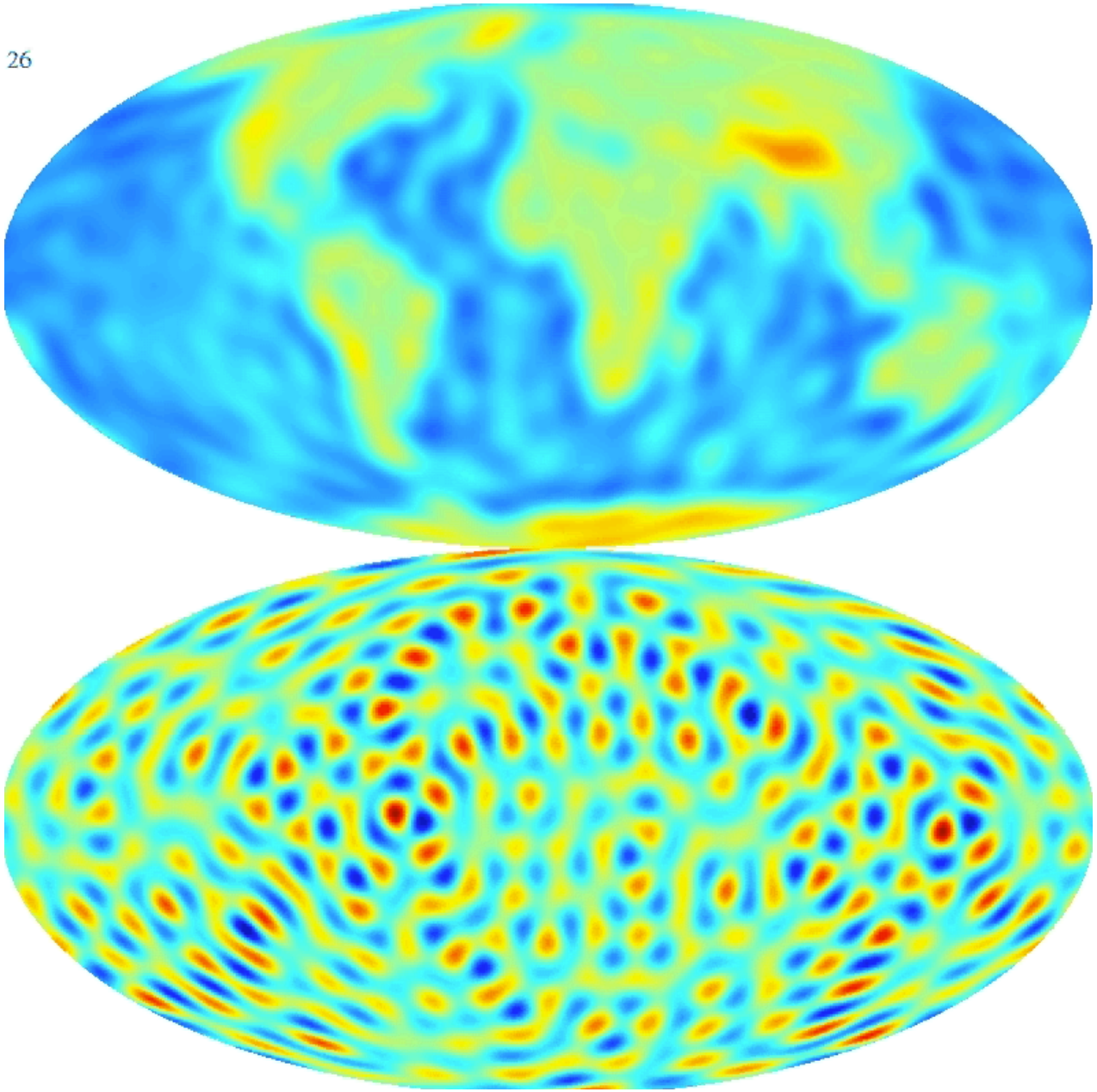
This is noticeable in the map (better resolution and better defined small-scale features)

and in the power spectrum (function measured to higher 'l')

2D angular temperature dimensionless power spectrum $l(l+1)\,C_{TT}(l)$



Decomposing a map in spherical harmonics: the Earth



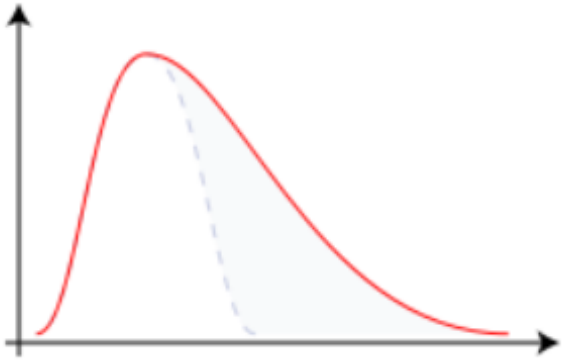
The land distribution is not an isotropic field → the "theoretical" power spectrum is a specific realizations, it is not averaged over all the m functions.

Higher-order statistics

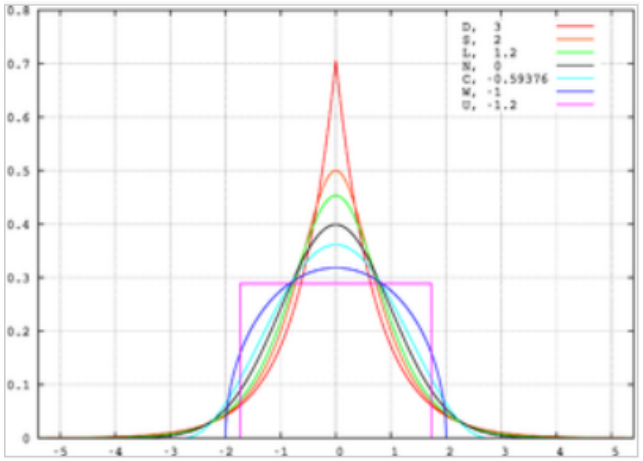
If the distribution is not Gaussian, the covariance matrix (and consequently the 2-pt correlation functions and power spectra) do not contain the whole cosmological information.

If the δ distribution is not symmetric → there is a non-zero skewness

If the δ distribution is cuspy → there is a non-zero kurtosis



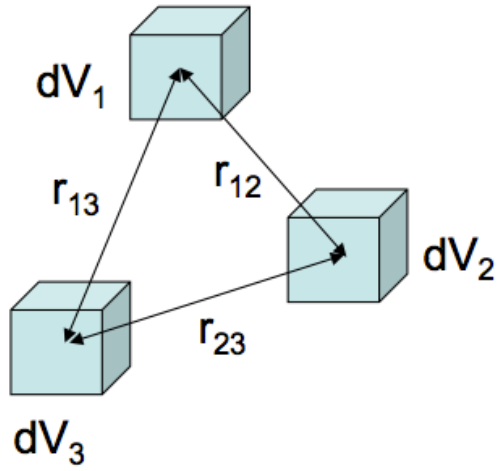
$$S = \frac{\langle \delta^3 \rangle}{\langle \delta^2 \rangle^{3/2}}$$



$$K = \frac{\langle \delta^4 \rangle}{\langle \delta^2 \rangle^2} - 3$$

In this case, we can define n-point correlation functions:

$\langle \delta_1 \delta_2 \delta_2 \rangle$ $\langle \delta_1 \delta_2 \delta_3 \delta_4 \rangle$



The joint probability (and so the clustering properties) of having a galaxy in locations 1,2,3 depends on the full conditional probability between the triplet, and also on all combinations of conditional probabilities between pairs:

$$dP_{123} = n^3 (1+\xi(r_{12})+ \xi(r_{13})+ \xi(r_{23})+\zeta(r_{12}, r_{13}, r_{23})) \, dV_1 \, dV_2 \, dV_3$$

Note that an **n-point correlation function** can be written as a sum of terms involving low-order correlations plus an irreducible (or connected) term → this is the Isserlis theorem of probability theory.

$$\mathrm{E}[X_1 X_2 \cdots X_{2n}] = \sum \prod \mathrm{E}[X_i X_j],$$
$$\mathrm{E}[X_1 X_2 \cdots X_{2n-1}] = 0,$$

(there is a version of this theorem used in quantum mechanics → Wick's theorem).

3-pt function

Using Wick's theorem, the 3-pt correlation function ζ_123 may be decomposed as

$$< \delta_1 \, \delta_2 \, \delta_3 >=< \delta_1 >< \delta_2 \, \delta_3 > + < \delta_2 >< \delta_1 \, \delta_3 > + < \delta_3 >< \delta_1 \, \delta_2 > + < \delta_1 \, \delta_2 \, \delta_3 >_c$$

This shows that for variables with zero mean → the 3-pt function is just the connected term.

For a Gaussian distribution the connected term is zero, and the 3-pt function is zero → note that this does not imply that the joint probability becomes just the product of the 3 individual probabilities (with zero correlation) since the conditional probability also depends on the 2-pt correlations.

We can also define the harmonic transformation of the 3-pt function, which is called the bispectrum:

$$\langle \tilde{\delta}(\vec{k}) \tilde{\delta}(\vec{q}) \tilde{\delta}(\vec{p}) \rangle = (2\pi)^3 B(k,q,p) \delta_D(\vec{k} + \vec{q} + \vec{p})$$

4-pt function

In this case the joint probability is

$$dP_{1234} = n^4(1 + \xi_{12} + \xi_{13} + \xi_{14} + \xi_{23} + \xi_{24} + \xi_{34} + \zeta_{123} + \zeta_{124} + \zeta_{134} + \zeta_{234} + \mu_{1234})\, dV_1 dV_2 dV_3 dV_4$$

Using Wick's theorem, the 4-pt correlation function, μ_1234, may be written as

$$< \delta_1\, \delta_2\, \delta_3\, \delta_4 >=< \delta_1\, \delta_2 >< \delta_3\, \delta_4 > + < \delta_1\, \delta_3 >< \delta_2\, \delta_4 > + < \delta_1\, \delta_4 >< \delta_2\, \delta_3 > + < \delta_1\, \delta_2\, \delta_3\, \delta_4 >_c$$

note the number of terms in the sum is $\dfrac{n!}{2^{n/2}\,(n/2)!}$ (n=4, in this case)

For a Gaussian distribution the connected term is zero, but the 4-pt function is not zero → however the kurtosis is zero (notice its definition).

The harmonic transformation of the 4-pt function is called the trispectrum: T(k,p,q,s)

Is the density contrast really a Gaussian random field?

- **Primordial non-Gaussianities**

Perhaps not: certain models of inflation can produce non-Gaussian features from the original Gaussian quantum fluctuations

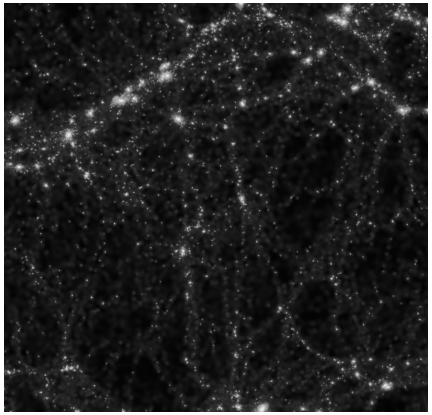
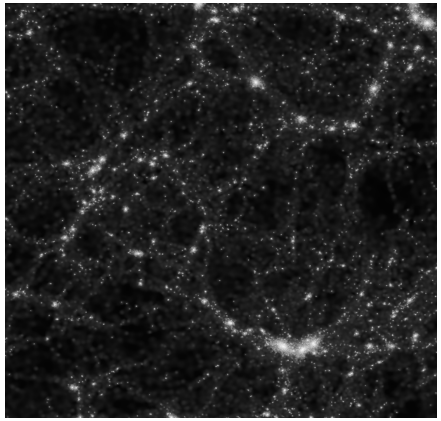
- **Secondary non-Gaussianities**

Definitely not: late-time evolution and other late-time effects produce mode coupling and the cosmological random fields are no longer Gaussian today

The dark matter density field becomes non-Gaussian in the recent universe due to non-linear evolution → δ may only be Gaussian in the linear regime, i.e., while its value is small.

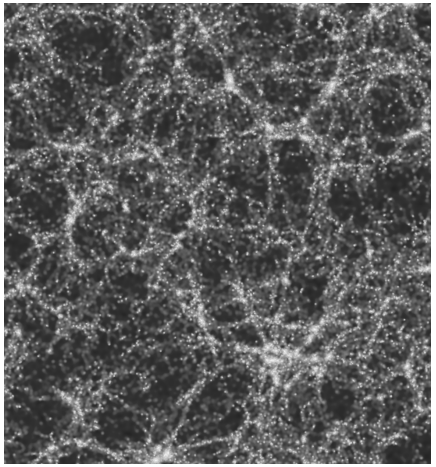
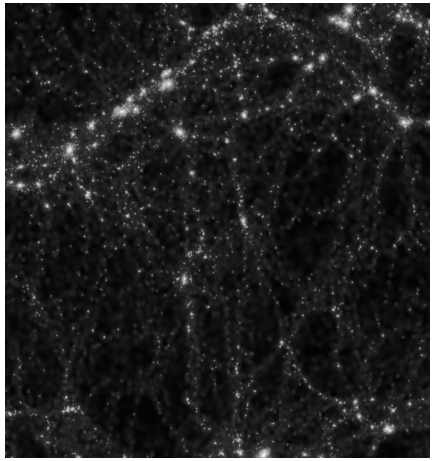
Higher-order moments (eg: order 3 and 4) are in reality non-zero and contain additional cosmological information.

Examples of density contrast fields

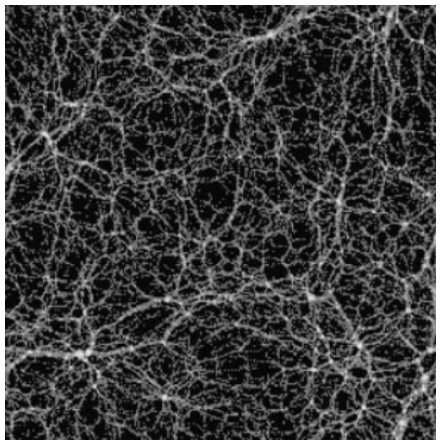
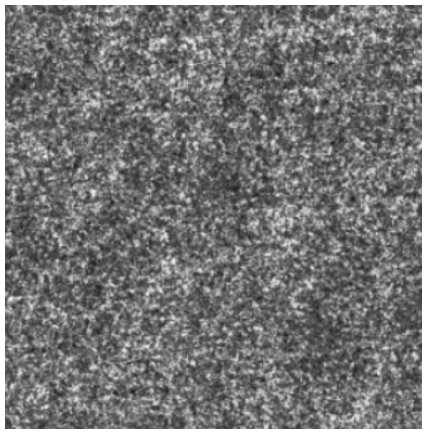



Comparing left and right panels

Same cosmological model (identical statistical moments, P(k), etc)
Same distribution (Gaussian)
Different realizations
→
The maps are statistically equivalent, although not identical

Different cosmological models (different statistical moments, P(k), etc)
Same distribution (Gaussian)
Different realizations
→
Fundamentally non-equivalent

Same cosmological model (identical statistical moments, P(k), etc)
Gaussian distribution (left) and non-Gaussian with identical Gaussian part (right)
Different realizations
→
Non-equivalent from NG effects

Uncertainty of the angular power spectra estimator: cosmic variance

The power spectrum measured from a map is one realization of the theoretical power spectrum predicted from the cosmological model.

For example, for a given multipole l, the **measured power spectrum** amplitude may be:

$C_l = \langle a_{l4}\, a_{l4} \rangle$ (or any other value of m) (and other values of m for other multipoles).

Other parts of the map may correspond to other realizations (each sub-map is independent). **The maximum number of independent measurements of C_l from a map is 2l+1**

On the other hand, the **theoretical power spectrum** we want to estimate is C_l with any value of m.

However, the measurements have uncertainty and are not exactly identical. So, the best way to **estimate** the theoretical power spectrum is to take the average of all possible measurements:

$$\widehat{C}_\ell \equiv \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2$$

This estimator is unbiased and if there were many measurements it would give exactly the theoretical power spectrum:

$$\langle \hat{C}_\ell \rangle = C_\ell$$

However, since there is a limiting number of independent measurements of each multipole (2l+1), the measured value will estimate the theoretical value with some **minimum uncertainty**. This is called the **cosmic variance**.

The total uncertainty is in general larger than this, since other measurement errors need to be added to this minimal one.

The uncertainty of the estimator (i.e. the cosmic variance) is computed from the definition of a covariance:

$$\sigma^2_{C\ell} = \left\langle (\hat{C}_\ell - C_\ell)(\hat{C}_{\ell'} - C_{\ell'}) \right\rangle$$

Note this is written as a covariance, i.e., considering l and l'. However, since the multipoles are independent the covariance matrix is diagonal → **only the variances are non-zero → l = l'**

$$\sigma^2_{C\ell} = \left\langle (\hat{C}_\ell - C_\ell)^2 \right\rangle$$

$$\left\langle (\hat{C}_\ell - \langle C_\ell \rangle)^2 \right\rangle = \left\langle \hat{C}_\ell^2 - 2\hat{C}_\ell \langle C_\ell \rangle + \langle C_\ell \rangle^2 \right\rangle$$

$$= \langle \hat{C}_\ell^2 \rangle - 2C_\ell \langle \hat{C}_\ell \rangle + C_\ell^2$$

$$= \langle \hat{C}_\ell^2 \rangle - 2C_\ell^2 + C_\ell^2 = \langle \hat{C}_\ell^2 \rangle - C_\ell^2$$

(where $\langle C_l \rangle = C_l$ is the theoretical value)

To evaluate the cosmic variance we need then to compute $\langle \hat{C}_\ell^2 \rangle$ as function of C_l . Naturally, this is:

$$\langle \hat{C}_\ell^2 \rangle = \frac{1}{(2\ell+1)^2} \left\langle \sum_m \sum_{m'} (a_{\ell m} a_{\ell m}) \sum_{m'} (a_{\ell m'} a_{\ell m'}) \right\rangle$$

$$= \frac{1}{(2\ell+1)^2} \sum_{m,m'} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'} \rangle$$

Notice the variances are power spectra squared, i.e., **4-pt functions**

Wick's theorem allows us to write a four-point function in terms of lower order functions. In particular for **Gaussian fields of zero mean**, the 1-pt and 3-pt functions are zero, and we can write:

$$\sum_{m,m'} \langle a_{\ell m} a_{\ell m} a_{\ell m'} a_{\ell m'} \rangle = \sum_{m,m'} \Big[\langle a_{\ell m} a_{\ell m} \rangle \langle a_{\ell m'} a_{\ell m'} \rangle + 2 \langle a_{\ell m} a_{\ell m'} \rangle \langle a_{\ell m} a_{\ell m'} \rangle + \langle a_{\ell m} a_{\ell m} a_{\ell m'} a_{\ell m'} \rangle_c \Big]$$

Using the result $\sum_{m,m'} \langle a_{\ell m} a_{\ell m} \rangle_c = (2\ell+1) C_\ell$

it is just a question of counting all the terms contributing to the various sums, to find the result:

(*see homework)* $\langle \hat{C}_\ell^2 \rangle = \frac{C_\ell^2}{(2\ell+1)^2} \left[(2\ell+1)^2 + 2(2\ell+1) \right]$

And so the cosmic variance is:

$$\langle \hat{C}_\ell^2 \rangle - C_\ell^2 = C_\ell^2 \left(1 + \frac{2}{2\ell+1} - 1 \right)$$

i.e.,

$$\sigma^2_{C\ell} = \frac{2}{2\ell+1} C^2_\ell$$

This result shows that this ultimate limit of cosmological observations depends on the amplitude of the angular power spectrum and the scale.

Thinking of the **ergodic hypothesis**, independent regions of the sky are different realizations → could correspond to different universes (with different parameter values) → creating an intrinsic variance on the measurements → (this is the reason for this limit to be called the cosmic variance).

Also note that since cosmic variance depends on the cosmological parameters, it is not taken into account in Fisher matrix analyses.

Since each scale has (2l+1) independent 'measures' contributing to it → large scales have less independent measures in the full sky than smaller ones → **cosmic variance dominates on large scales**, **we only have 1 universe to observe.**

The calculation is valid for a **full sky survey**. If the survey covers a smaller area, by a factor **f_sky = Area_survey / Area_fullsky**, there are less independent measures contributing to each scale, and the cosmic variance scales accordingly:

$$\sigma^2_{C\ell} = \frac{1}{f_{sky}} \frac{2}{2\ell+1} C_\ell^2$$

If we want to limit the cosmic variance in a future survey, we should build a wider survey rather than a deeper one (i.e., increase the survey area).

Note that for the largest possible angular scale (l=1), the minimum uncertainty achievable (in the ideal case of a full sky survey and no experimental noise) is a fractional uncertainty of
σ_l / C_l = (2/3)^0.5 = 81%

This is the large uncertainty seen in CMB plots, and is a fundamental limitation of cosmological data.