The Inhomogeneous Universe

The density contrast random field

First Principles

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

(At first) 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)**:

the deviation with respect to the mean density (averaged over space)

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

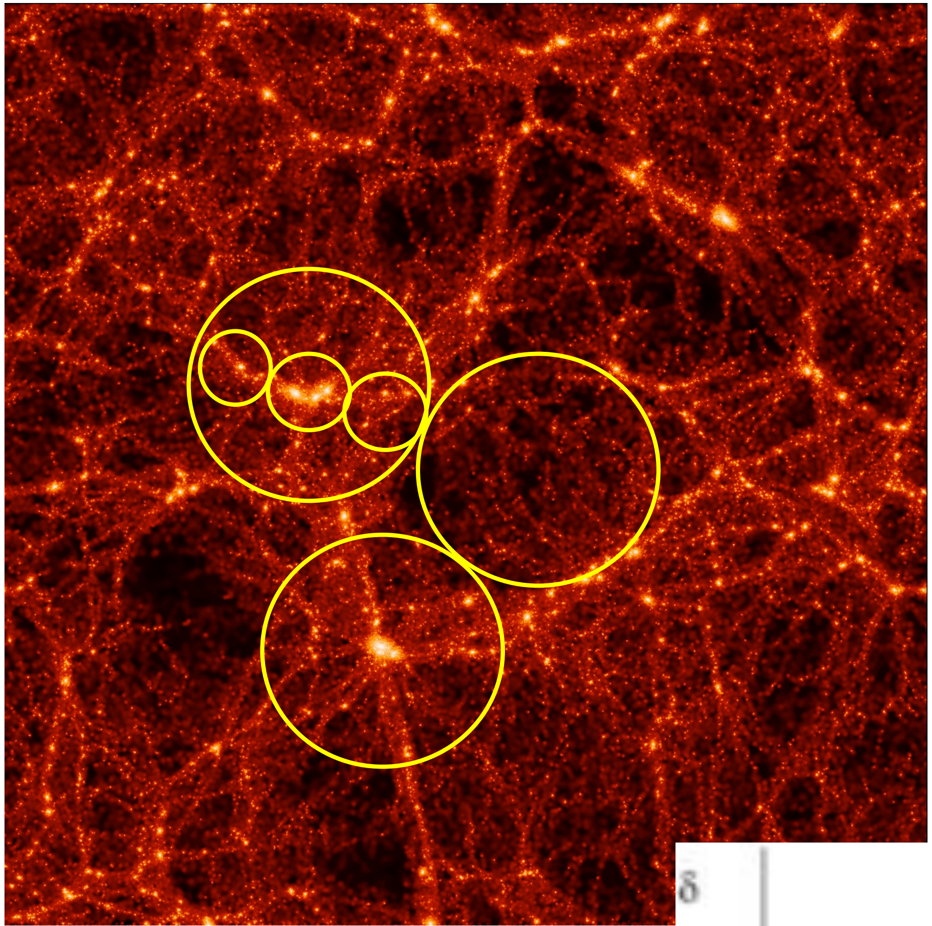
During the **evolution of the Universe** (evolution of the mean density), the density contrast at each point also evolves, either increasing or decreasing, driven by gravity.

An increase of δ means clustering of matter → in practice a local region of the Universe expands slower than the global expansion.

The process of evolution of the density contrast is called **structure formation**, turning density fluctuations in cosmological and astrophysical structures.

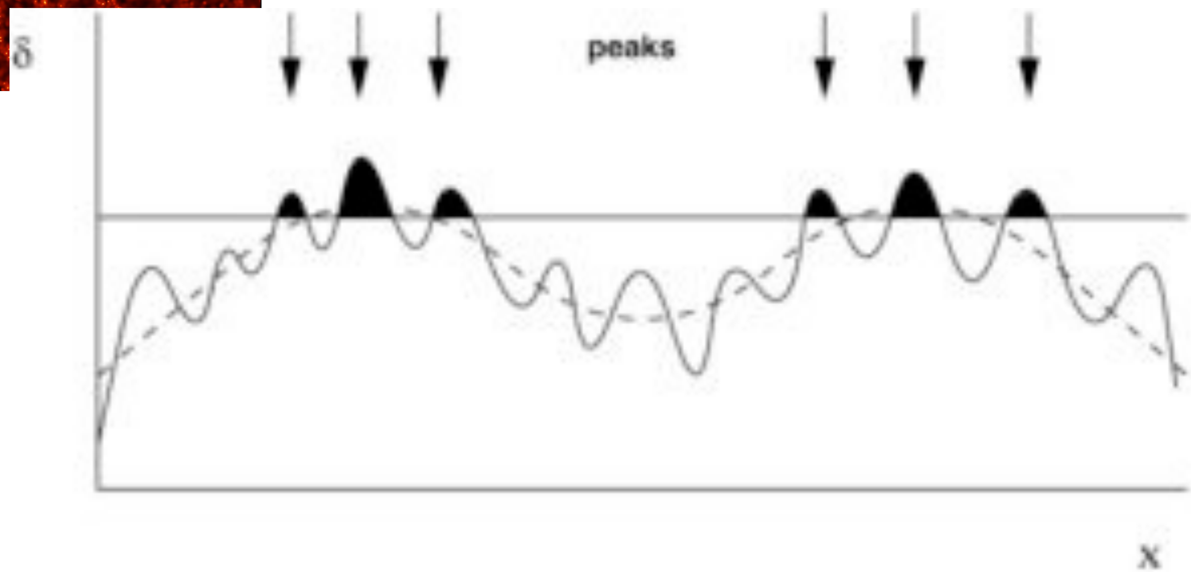
δ can become very large (not a density perturbation anymore) but the associated gravitational potential always remains a perturbation to the metric.

Density map and scales



overdensities
and
undersdensities

on two different scales



Why the very early Universe is not exactly homogeneous?
(how do initial fluctuations around the mean arise?)

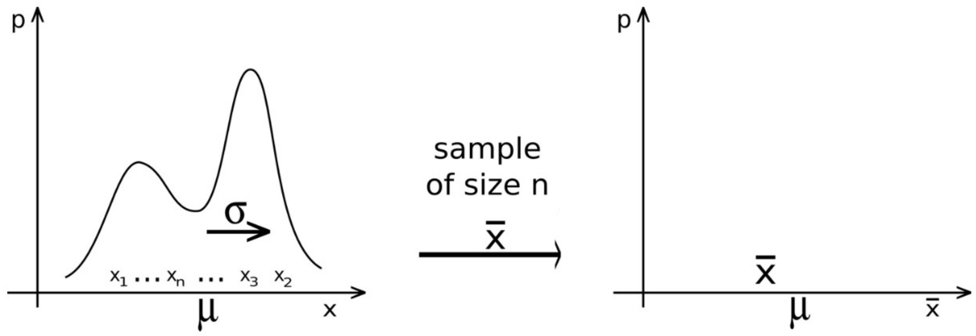
The reason is: **quantum fluctuations**

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 changes all the time as the result of a stochastic (random) process.

(In very short timescales as compared with the expansion rate of the Universe)

It is not possible for the cosmological model to determine the value of density at a given location at a given time, in a deterministic way.

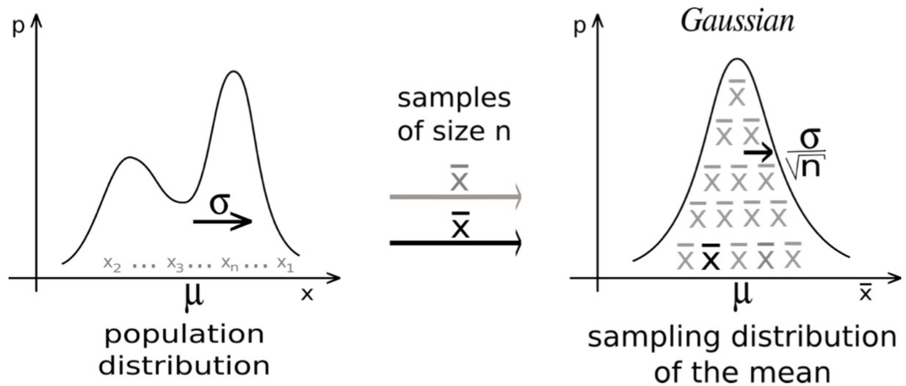
Initial population of *δ* values in a location: histogram of values of *δ* in during the stochastic process (which forms a **population**)



After a certain time there is an average value of *δ* there

In other time intervals, the average value is different → the average values will also form a distribution (the **sampling distribution**).

Since there is a large number of independent random processes involved, **the sampling distribution of the averages is a Gaussian distribution**, centered on the true mean, whatever the form of the population distribution.
(Central limit theorem)



→ the quantum density field is a Gaussian random field.

Later, the **inflationary mechanism** makes the transition 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 $\delta(x)$, but again its actual value is not known, it is a particular realization among all possible realizations of the average δ value.

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)

In standard inflation, the Gaussianity is preserved.

The value of density at a given location is then a value taken from a Gaussian distribution → the actual values of $\delta(x)$ at each point are not known.

We only 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,

$$P(\delta_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\, e^{-\frac{1}{2}\frac{(\delta_1-\mu_1)^2}{\sigma_1^2}}$$

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

The **mean**, $\boldsymbol{\mu}$, can be estimated from a sample of M elements of the population of δ_1 as

$$\bar{\delta}_1 = \frac{1}{M}\sum_{i=1}^{M}\delta_{1i}$$

The **variance**, ***σ***, can be estimated from a sample of M elements of the population of δ_1 as

$$\sigma_1^2 = \frac{1}{M}\sum_{i=1}^{M}(\delta_{1i} - \mu_1)^2$$

If both the mean and the variance are estimated from the sample, then the variance can be estimated in an alternative way:

$$\sigma_1^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\delta_{1i} - \bar{\delta}_1)^2$$

(the square root of the variance is known as the **standard deviation**)

The value of density at a given location is a realization of this distribution. The Universe has only one value $\delta_1(t)$, i.e., one specific realization.

So, what are the other elements of the population? They could be realizations in alternative Universes.

Note 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** → the stochastic processes may be different in each location, leading to different values of mean and variance.

Let us consider the full density contrast field (assuming a discretization)

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

An important point is that with time the N variables δ_1 ... δ_N cease to be 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 multi-variate Gaussian:

$$P(\delta_1, ..., \delta_N) = \frac{1}{\sqrt{(2\pi)^N \det C}} \exp\left(-\frac{1}{2}(\vec{\delta} - \vec{\mu})^T C^{-1} (\vec{\delta} - \vec{\mu})\right)$$

(The random variable δ has dimension N, and the N-dimension Gaussian distribution has a N-dimension vector of means μ and a N x N covariance matrix C.)

For example, **in the case of only 2 random variables** (we could bin the density field such that it would have only two locations), we would need a 2-dimensional Gaussian, with covariance:

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\,\sigma_1\sigma_2 \\ \rho_{12}\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

The diagonal of the matrix contains the **variances** of each variable and the off-diagonal contains the **covariance** between the variables:

$$\sigma_{12} = \frac{1}{M-1}\sum_{i=1}^{M}(\delta_{1i} - \bar{\delta}_1)(\delta_{2i} - \bar{\delta}_2)$$

The alternative form of the covariance is written introducing the **correlation coefficient** between the variables:

$\boldsymbol{\rho}_{12} = \sigma_{12} / (\sigma_1 \sigma_2)$

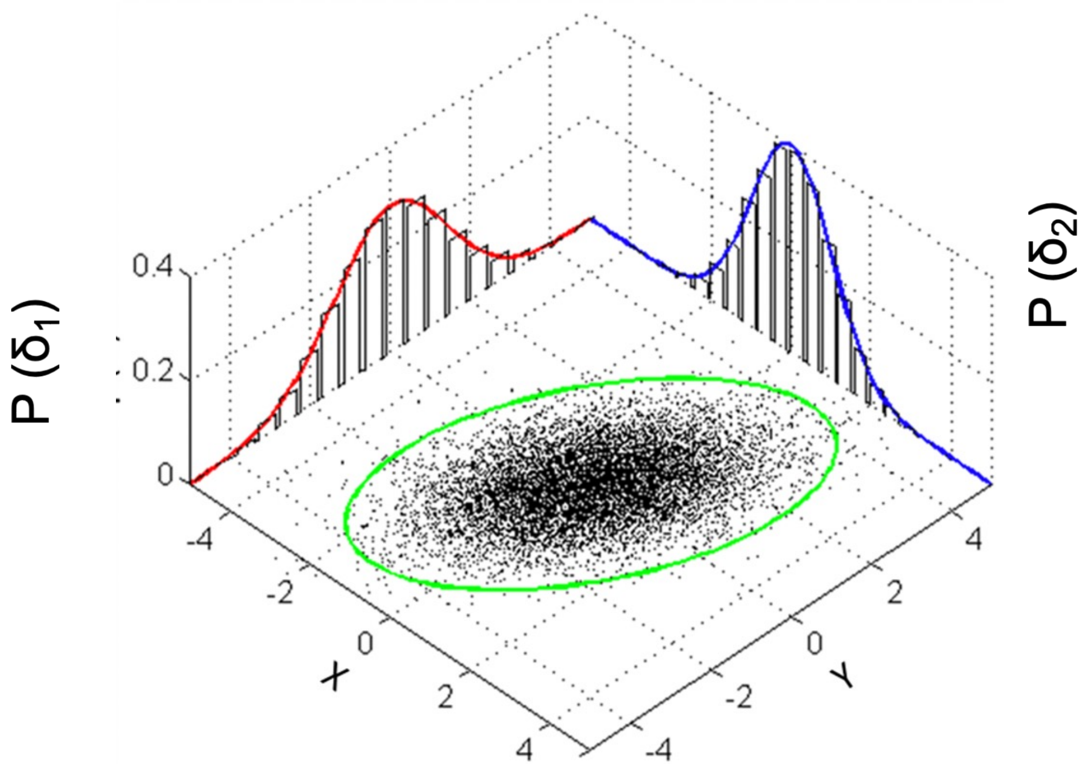
The correlation between the variables can also be written in the form of a **correlation matrix**

$$\begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}$$

So, the 2-dimensional Gaussian distribution is

$$P(\delta_1, \delta_2) = \frac{1}{(2\pi)\sqrt{\det C}} \, e^{-\frac{1}{2}(\delta_1-\mu_1)^2 \, C_{11}^{-1} + 2(\delta_1-\mu_1)(\delta_2-\mu_2) \, C_{12}^{-1} + (\delta_2-\mu_2)^2 \, C_{22}^{-1}}$$

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



The two distributions have different variances → **different widths**

and non-zero correlation → the iso-probability contour (an ellipse) is **not aligned with the axes**

The joint probability of having a value δ_1 at the location 1 and having at the same time a value δ_2 at the location 2 can be written as:

$P(\delta_1, \delta_2) = P(\delta_1)\, P(\delta_2 | \delta_1)$ (which introduces $P(\delta_2 | \delta_1)$ the conditional probability)

It seems that the stochasticity increases the complexity of the treatment of the first-order density field:

If the problem was **deterministic**:

system described by the field δ(x) → N values

Because the problem is **stochastic**:

system not described by the actual values of δ(x) but by the moments of the N-dimensional 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)

Since 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).

However, the complexity is reduced thanks to the

Generalized cosmological principle:

"The universe is statistically homogenous and isotropic"

This means that perturbations to the homogeneity are not completely free. 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**.

(unlike 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**.

$$\bar{\delta} = \langle \delta \rangle$$

(sample average equals spatial average)

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

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

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:

$$\boldsymbol{\langle\delta\rangle=0}$$

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 in the usual way:

$$\langle \delta^2 \rangle = \frac{1}{M} \sum_{i=1}^{M} \delta_i^2$$

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)

and hence it is not more complex than the deterministic problem.

Correlation function

Definition and standard computation

The N-1 covariances define 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.

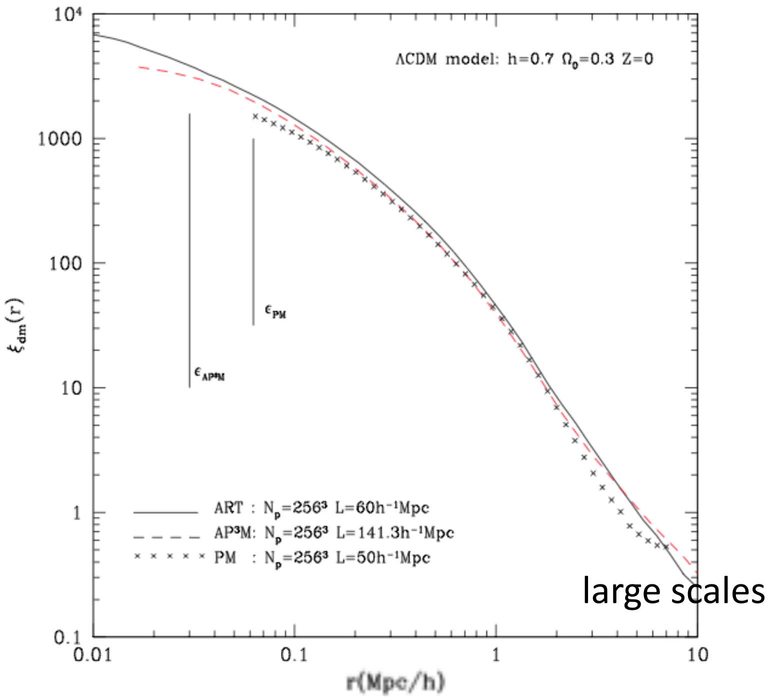
The randomness aspect and the generalized cosmological principle, imply that ***the most natural spatial variables to use in the treatment of the inhomogeneous Universe are not locations but separations between locations.***

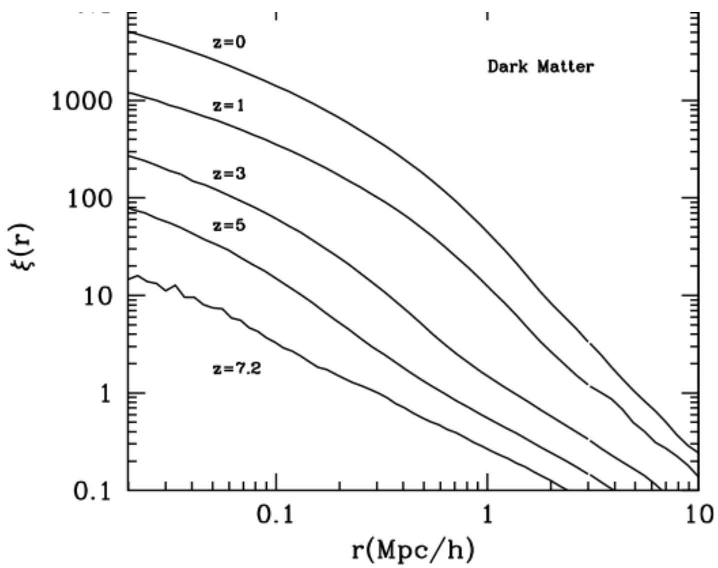
For a given δ(x), we can compute the correlation from its definition

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

The dark matter density correlation function of the overdensity field predicted by the ΛCDM model is **positive and decreases with separation.**
(Theoretical predictions are computed from the linear structure formation mechanism, and the non-linear gravitational collapse).

Its amplitude increases with structure formation (as the clustering of matter increases) → **it decreases with redshift**.





The correlation function of the density contrast field contains all the statistical information on the Gaussian density contrast field → and so **it describes how matter is distributed in the Universe,** because it is all information we need to compute the joint probability of having a value δ_1 at a location “1” and having a value δ_2 at a location “2”.

The joint probability is written as:

$$P(\delta_1 ,\delta_2) = P(\delta_1)\, P(\delta_2 |\delta_1)$$

and depends on 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

In this form it becomes explicit that the correlation describes the **clustering** properties of the field.

Alternative computation

The correlation function can be estimated in an alternative way. Instead of making a direct application of its formula, **we may use use its role in the probability distribution**.

Let us consider N galaxies on a volume V, with a number density of n=N/V (and assume that the position of a galaxy indicates a matter overdensity)

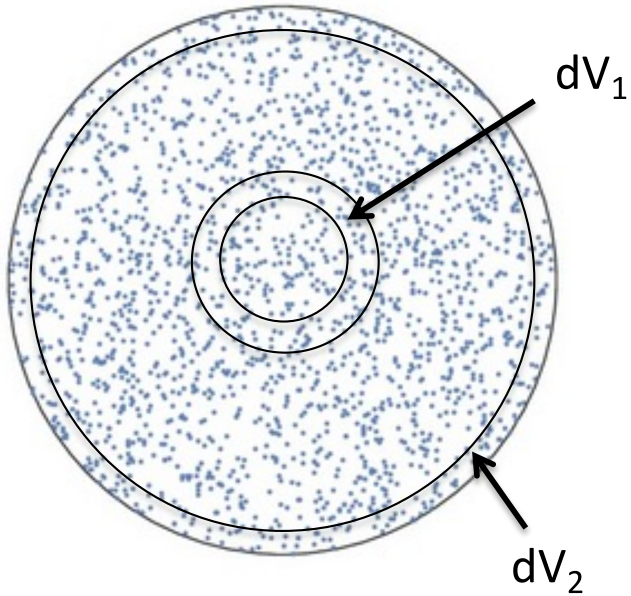
(i) Case of an uncorrelated distribution

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 uncorrelated distribution

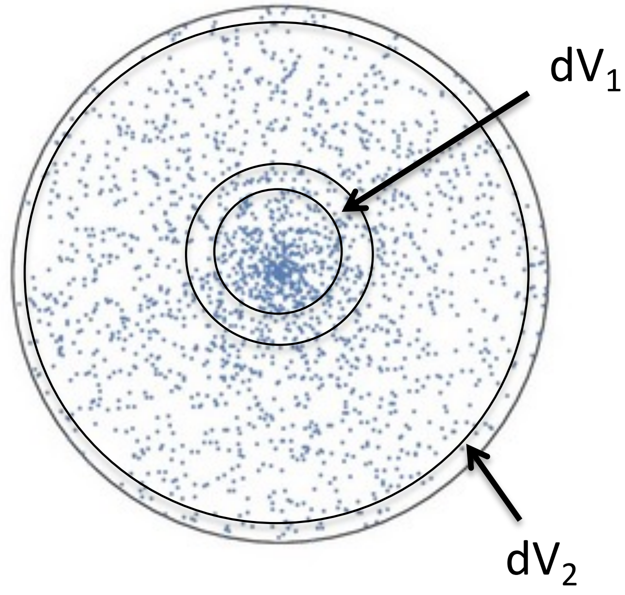
(ii) Case of a correlated distribution

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} :

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



Case of correlated distribution

So, the number of galaxies found is no longer just a function of the size of dV_2 but it also depends on the way the galaxies are distributed in the volume (which depends on correlation with the neighbors, i.e., on the correlation function)

Note that the correlation can be positive or negative:

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

We can compute the **total number of galaxies in a volume up to a radius r.**

It is given the integral of the quantity N multiplied by its **weight function**. The weight function is the "histogram" of the distribution of galaxies per bins of r, i.e. it is a **"distance function", the number of objects per distance bin dN (r).**

So, $N(r) = \int N \, dP(r)$

In the uncorrelated case (the conditional probability is 1), N(r) is simply

$N(r) = \int N/V \, dV = n \int dV/dr \, dr \sim r^3$ → the number increases with the volume

In the correlated case, $N(r) = n \int (1+ \xi(r)) \, dV/dr \, dr$ → the slope will be different

from r^3, depending on the correlation function $\xi(r)$ → the number is higher on a highly correlated area (usually on small separations).

From this result, we see that the **correlation function can be equivalently defined as the excess N(r) between the clustered and the random cases:**

If we compare the probabilities dP(r) 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):

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

Note on discrete distributions

We can define a $\boldsymbol{\delta_g}$ **(x)**, which is basically N_gal (x).

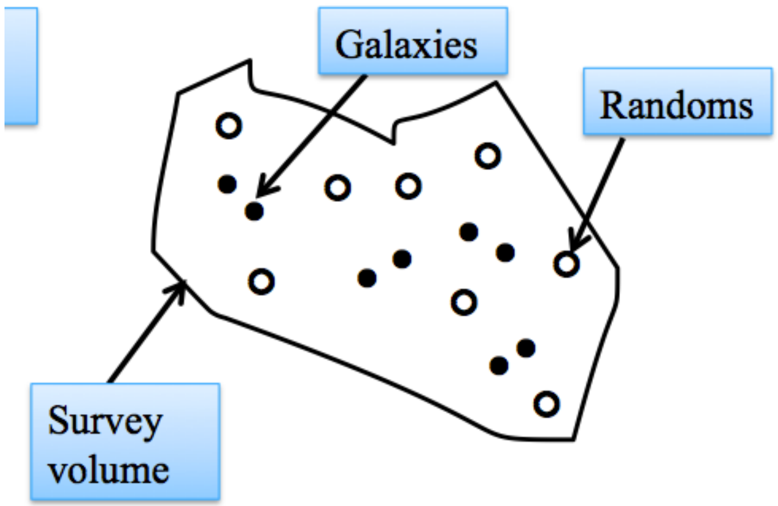
the number of galaxy pairs as function of separation can be written schematically as 1x1 + 1x0 + 1x0 + 1x1 + …. → it is "a kind of" $<\delta_g(x)\ \delta_g(x)>$

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:

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

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

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

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

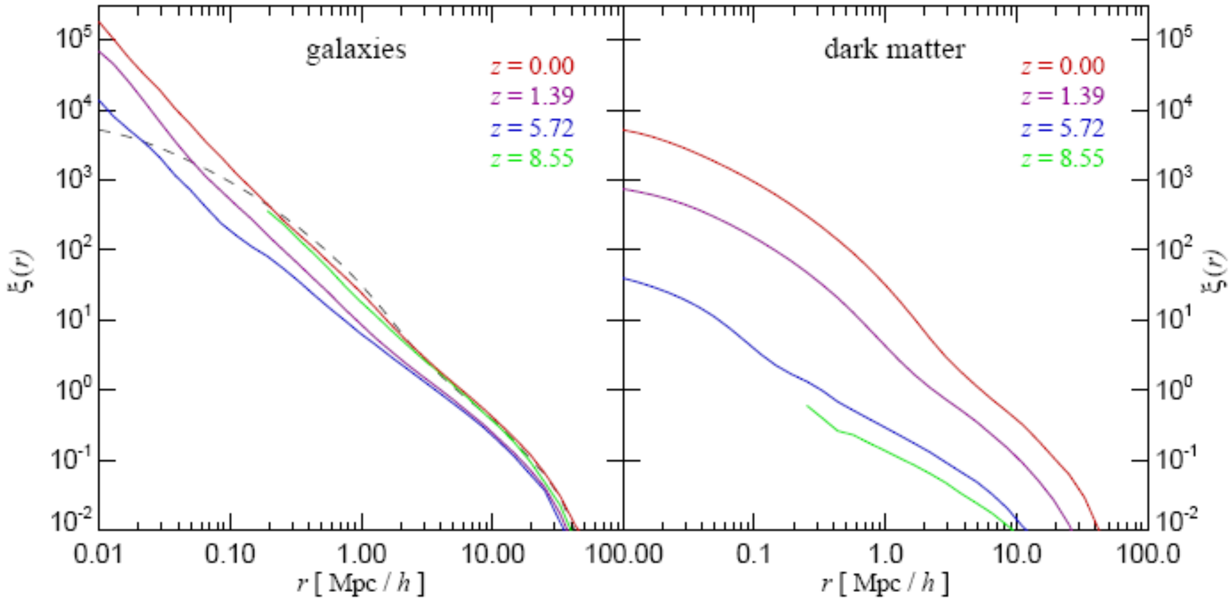
The 4 estimators have different noise properties.

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

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

(r_0 is a critical separation that depends on the type of galaxies, a typical value is r_0 ~ 5 Mpc/h)



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.

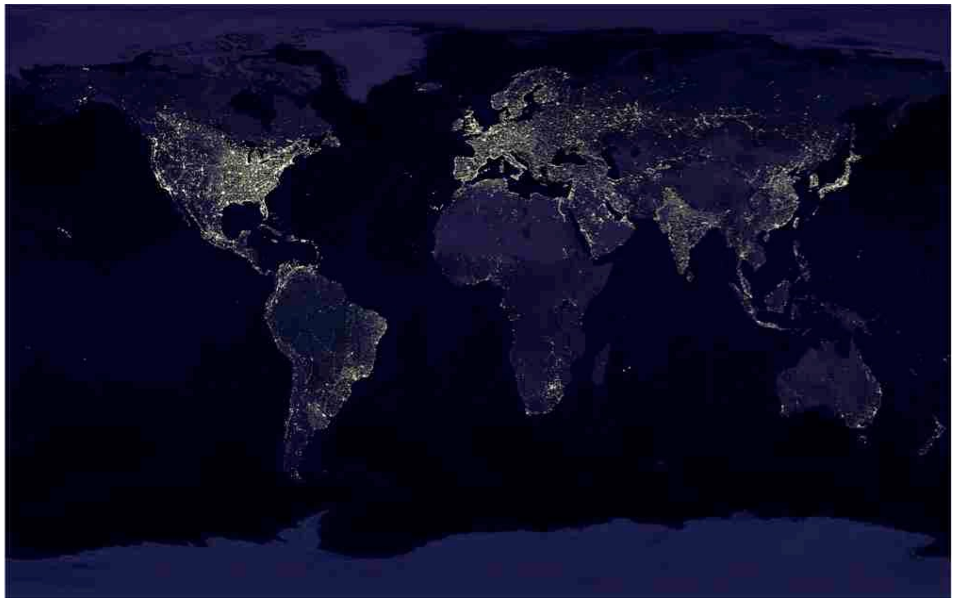
This shows that there is a **bias** between the spatial distributions of galaxies and dark matter, i.e.,

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

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



Correlation Function in Fourier space

Power Spectrum

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) → there is a halo 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 r.

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)

Convention:

- we are writing the plane waves as ikx and not i2πkx → this makes a factor $(2\pi/k)^3$ to appear

- **the integrals are normalised by the volume V**, which ensures that δ_k is dimensionless if δ(x) is also dimensionless

Let us compute the 2-point correlation function in k-space :

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

The ergodic hypothesis allows us to put the brackets inside the integrals

Inserting the definition of the correlation function, we can write:

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

where y is the separation vector between x and x’,

Note that for fixed x the integration over x’ is the same as an integration over y.

So, we are left with an integral in x with no function dependent on x (except the plane waves),

and an integral in y that is a (normalised) Fourier transform of the correlation function:

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

The first integral is the (dimensionless) Dirac delta.

Recall the Dirac delta is the (standard) Fourier transform of f(x)=1:

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

The second integral is the (normalised) Fourier transform of the correlation function, which is called the **dimensionless power spectrum:**

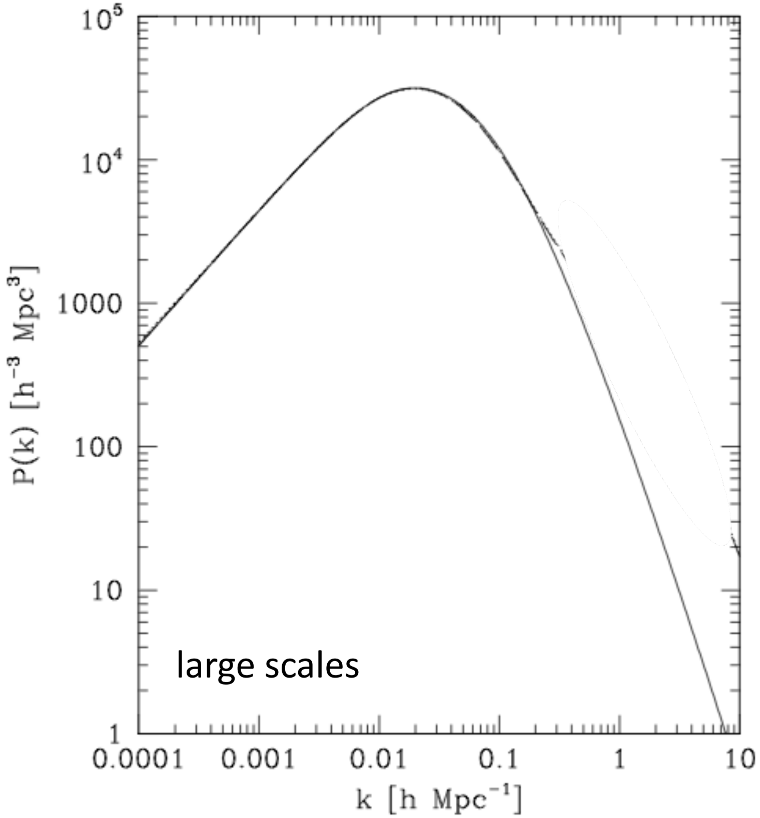
$$P_\delta(|k|) / V$$

Note that due to isotropy it only depends on the modulus of the k-mode vector.

The **power spectrum** of a random field is defined as the (standard) Fourier transform of the correlation function of the same field,

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

(and reciprocally, the correlation function is the Fourier transform of the power spectrum)



(the ΛCDM power spectrum of the density contrast field looks like this)

So the result is $$\frac{(2\pi)^3}{V}\delta_D(\vec{k}-\vec{k}')P_\delta(|\vec{k}|)$$

where δ_D here is the dimensionless Dirac delta

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

where we used the fact that the length associated to a Fourier mode k is $2\pi/k$, and so the corresponding volume is $V = (2\pi/k)^3$

Notice that the power spectrum P(k) has dimensions of volume [$(\mathrm{Mpc}/h)^3$]

and $\Delta^2(k) = k^3 P(k)$ is 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 density contrast field in Fourier space
is the (standard) Fourier transform of the correlation function multiplied by
the Fourier volume k^3 and by a dimensionless 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,
and

the elements of the correlation function in Fourier space are independent,
as are the elements of the power spectrum

Variance

It is also useful to compute the **auto-correlation function** of the density contrast field, i.e. the **variance**:

$$\sigma^2 = \langle \delta(x)\delta^*(x') \rangle = \langle \delta^2(x) \rangle \quad \text{where x=x'}$$

$$\sigma^2 = \frac{V}{(2\pi)^3}\frac{V}{(2\pi)^3}\int d^3k \delta_k e^{ikx} \int d^3k' \delta_{k'} e^{-ik'x}$$

$$= \frac{V}{(2\pi)^3}\frac{V}{(2\pi)^3}\int d^3k \int d^3k' \langle \delta_k \delta^*_{k'} \rangle e^{i(k-k')x}$$

Inserting the result for $\langle \delta_k \delta^*_{k'} \rangle$

$$\sigma^2 = \frac{V}{(2\pi)^3}\frac{V}{(2\pi)^3}\int d^3k \int d^3k' e^{i(k-k')x} k^3 P_\delta(k) \frac{\delta_D(k-k')(2\pi)^3}{V}$$

one of the integrals is just the Fourier transform of the Dirac delta, which is 1 (and also cancels with one of the volumes);

k^3 cancels with the other volume

and we are left with:

$$\sigma^2 = \int \frac{d^3k}{(2\pi)^3} P(k)$$

So, the variance of the delta field (in real space) is a 3d integral of the power spectrum. Since the power spectrum is isotropic, we can integrate the angular part of

$$d^3k = k^2 \sin\theta \, dk \, d\theta \, d\phi$$

which is $4\pi k^2$

resulting in:

$$\sigma^2 = \int_0^\infty \frac{dk}{k} \frac{k^3 P(k)}{2\pi^2}$$

Writing k^2 as k^3/k shows explicitly that:

to integrate k^2 P(k) on the linear domain dk is equivalent to integrate the dimensionless power spectrum in the logarithmic domain dk/k

This is the reason why the dimensionless power spectrum is known as the power spectrum per interval of ln(k).

This result tells us that the variance of the density contrast field has contributions from all scales of the power spectrum.
Each logarithmic bin contributes with a certain value (the value of the dimensionless power spectrum of that scale)

and so, **the amplitude of the dimensionless power spectrum is a direct indication of the amplitude of clustering**

$\Delta < 1$ - weak clustering, linear structure

$\Delta > 1$ - strong clustering, non-linear structure : large over-densities, or large under-densities (voids)

Covariance

Let us now consider 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).

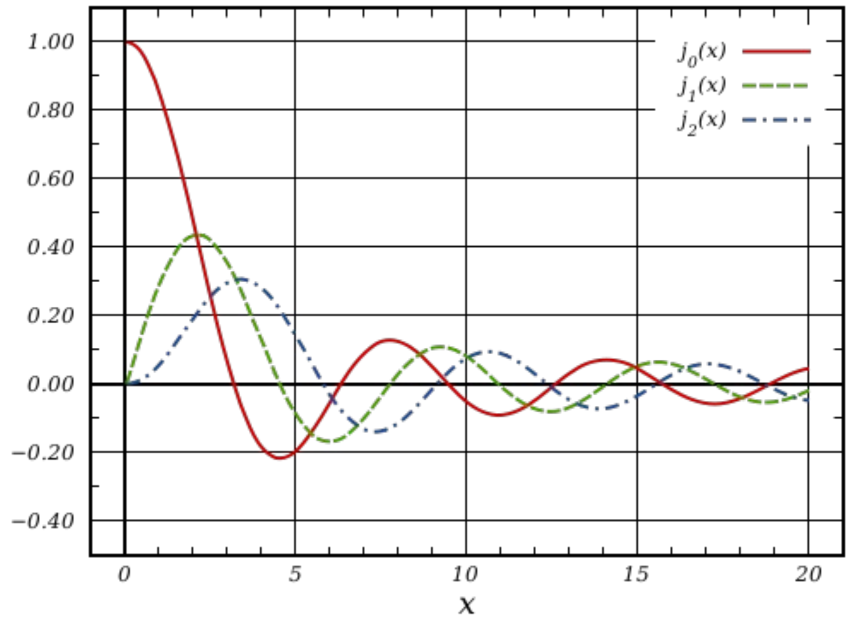
The filter (the function that multiplies k^2 P(k) in the integral) is the **spherical Bessel function** of the first kind for n=0 : j_0 (kr)

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

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

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

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



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.

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 the variance is a moment of a distribution → **the fact that a certain scale has a certain amplitude does not mean that all regions of the Universe of that size will have that same value of density contrast,**

- The value of the density contrast of a region of a given scale will be a realization of a Gaussian with the variance at that scale (which is given by the amplitude of the dimensionless power spectrum).

- **Each scale has a different variance**
(unlike the real-space description, where all locations have the same variance and the information is on the correlation function between locations)

- Recall that the for a random variable of zero mean, its amplitude is indicated by its variance - and not by its the mean! -

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 $\langle \delta_k^2 \rangle$

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).

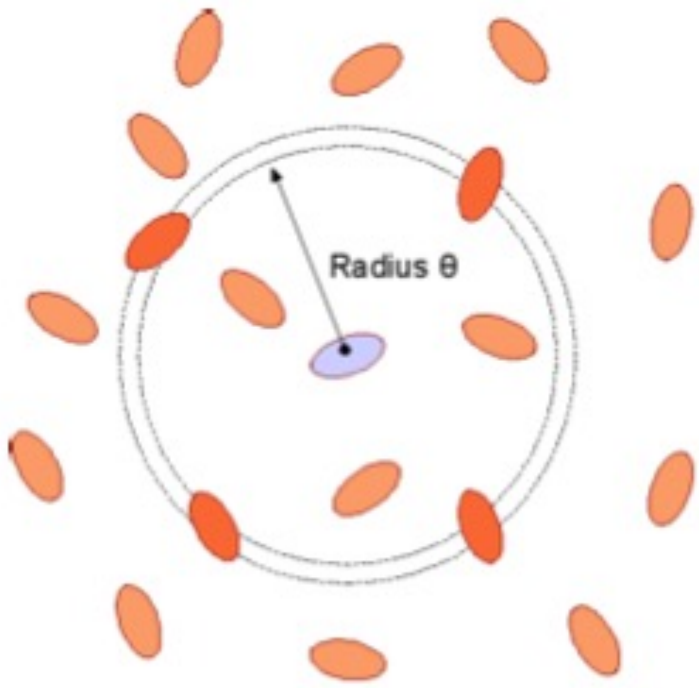
Power spectrum estimator: shot noise

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).

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

Note this is just a sophisticated way of writing 0 or 1.

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 integral over the Dirac delta sets $x=x_i$ in the plane wave and compute the correlation function in Fourier space $< \delta(k)\delta^*(k') >$

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

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

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

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 $\langle \exp(-ikx_i)\ \exp(ik'x_j) \rangle$
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:

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

(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.

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

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**.

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

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, $(2\pi)^6 \delta_D(\vec{\kappa})\delta_D(\vec{\kappa}')$

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

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

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

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 :

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

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).