Cosmologia Física

Ismael Tereno (FCUL, IA)

Ciências
ULisboa

instituto de astrofísica
e ciências do espaço

2026

The Inhomogeneous Universe

Statistical properties of the density contrast field I

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.

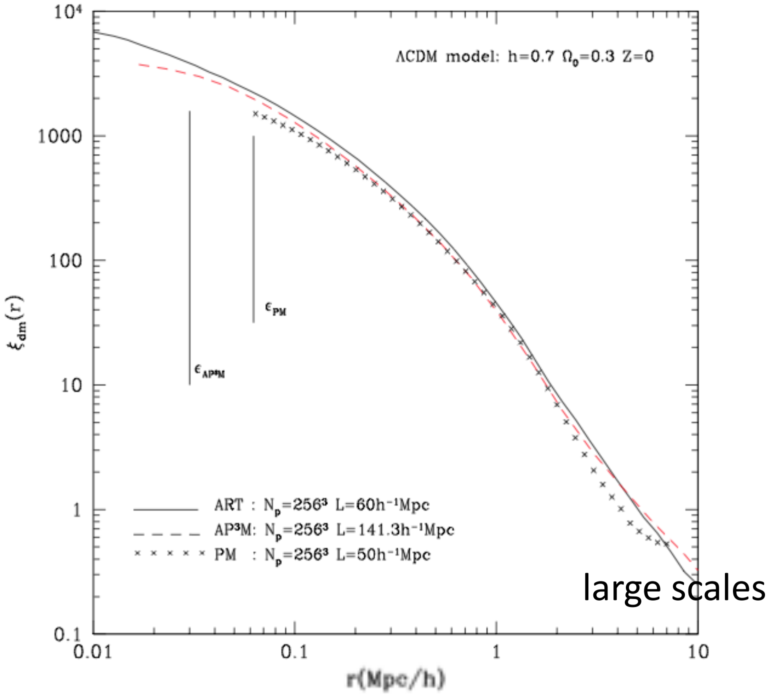
The randomness aspect and the generalized cosmological principle, imply that ***the most natural spatial variables to use in the treatment of the inhomogeneous Universe are not locations but separations between locations.***

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

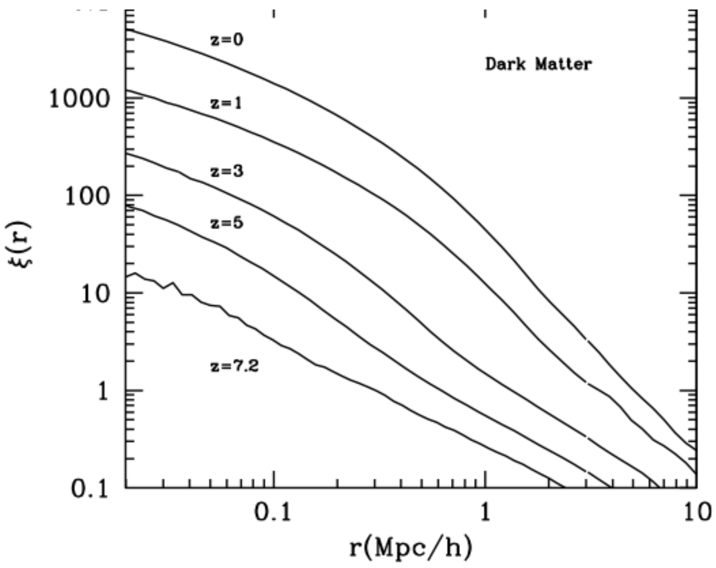
$$\xi_{\delta\delta}(r) = \langle \delta(x)\delta^*(x') \rangle$$

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**.



large scales



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 the 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

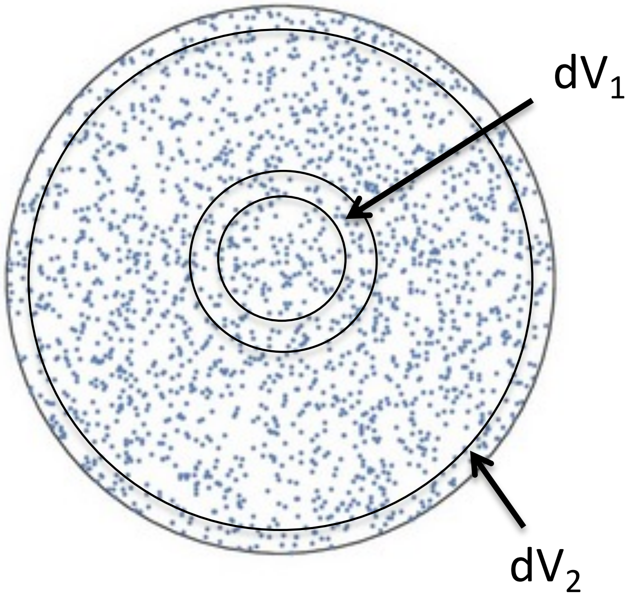
The correlation function can be estimated in an alternative way. Instead of making a direct application of its formula, **we may use use its role in the probability distribution**.

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

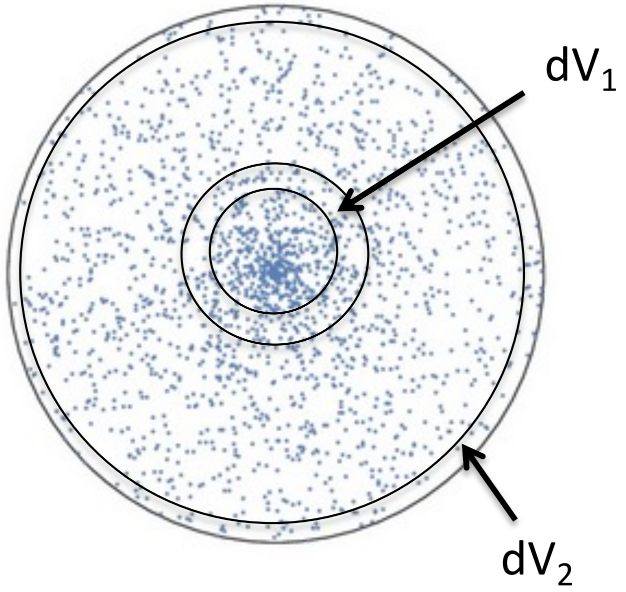
The probability of having a galaxy in the shell volume dV_1 is given by the number of galaxies within that volume divided by the total number of galaxies N:
$dP_1 = n\, dV_1 / N = dV_1 / V$

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



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 but it also depends on the way the galaxies are distributed in the volume (which depends on the 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 by 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 **δ_g (x)**, which is basically N_{gal} (x).

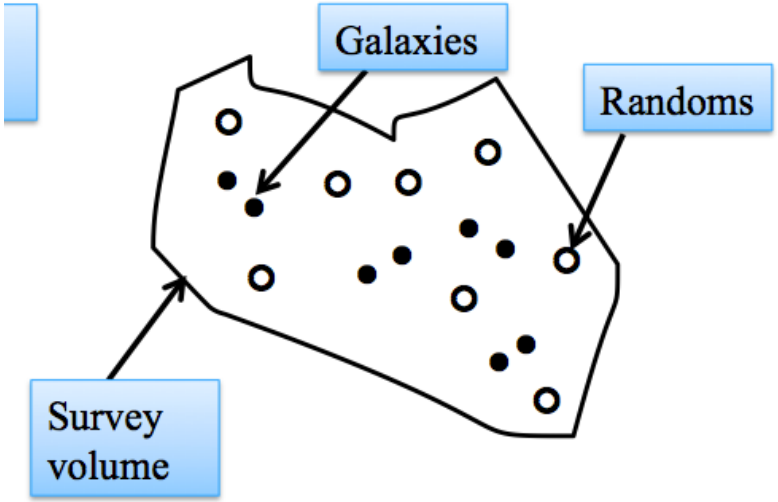
The number of galaxy pairs as function of separation can be written schematically as 1x1 + 1x0 + 1x0 + 1x1 + …. → it is “a kind of” $<\delta_g(x)\ \delta_g(x)>$

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:

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

$$\begin{aligned} 1 + \xi_2 &= \frac{\langle DD \rangle}{\langle DR \rangle}, \\ 1 + \xi_3 &= \frac{\langle DD \rangle \langle RR \rangle}{\langle DR \rangle^2}, \\ 1 + \xi_4 &= 1 + \frac{\langle (D - R)^2 \rangle}{\langle RR \rangle^2} \end{aligned}$$

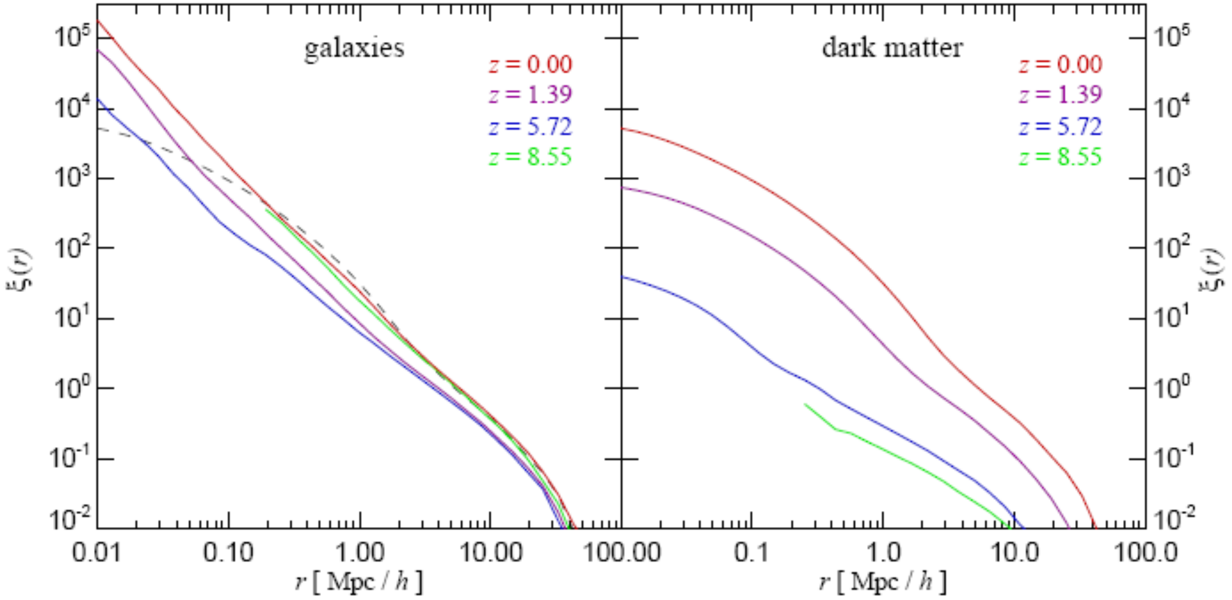
The 4 estimators have different noise properties.

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.

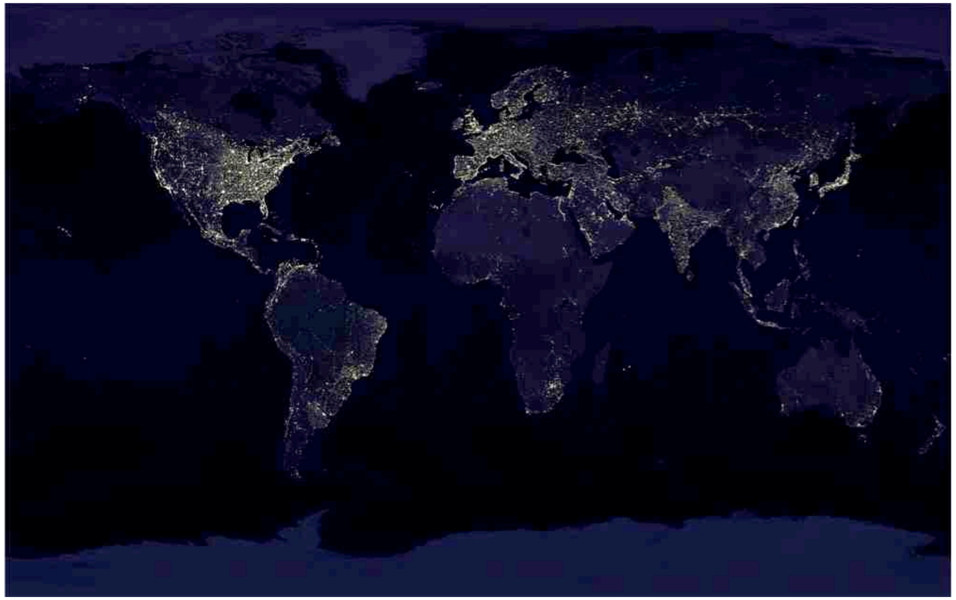
This shows that there is a **bias** between the spatial distributions of galaxies and dark matter, i.e.,

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

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

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

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:**

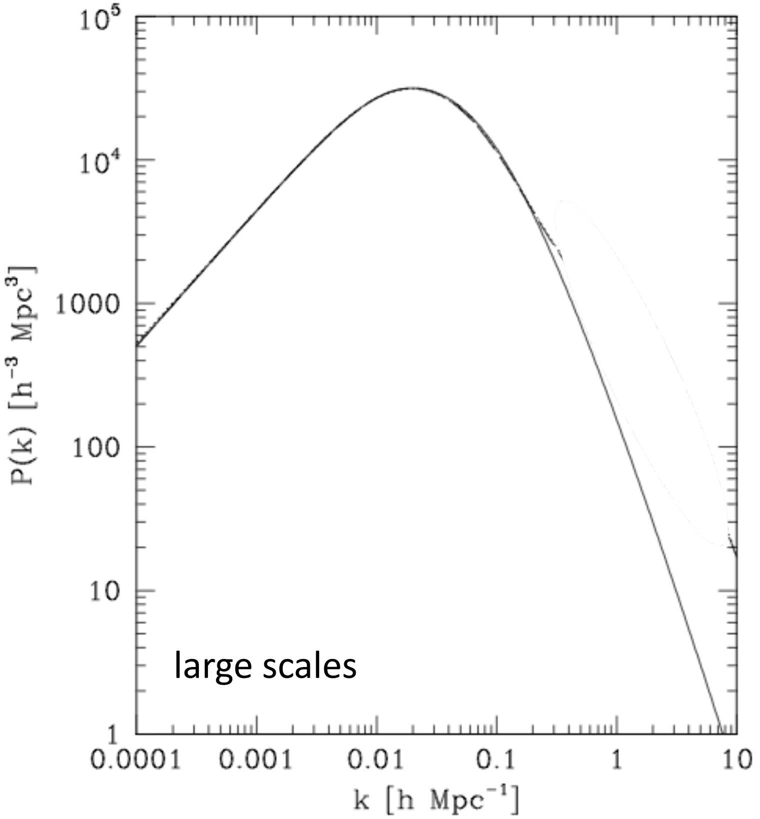
$$P_\delta(|k|) / V$$

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

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

where we used the fact that the length associated to a Fourier mode k is 2π/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**:

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

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

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

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

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

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).

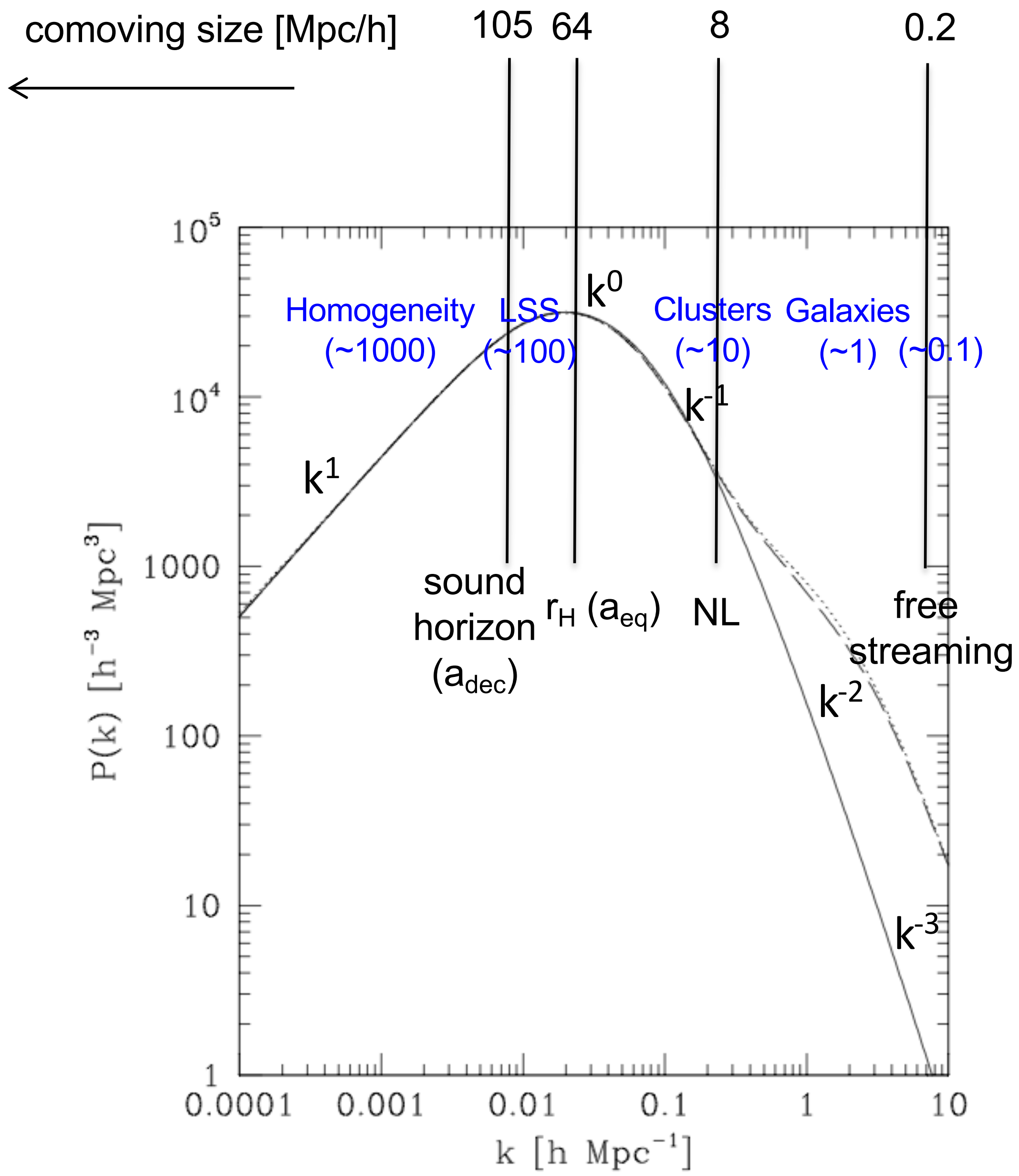
This result tells us that the variance of the density contrast field has contributions from all scales of the power spectrum.
Each logarithmic bin contributes with a certain value (the value of the dimensionless power spectrum of that scale)

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

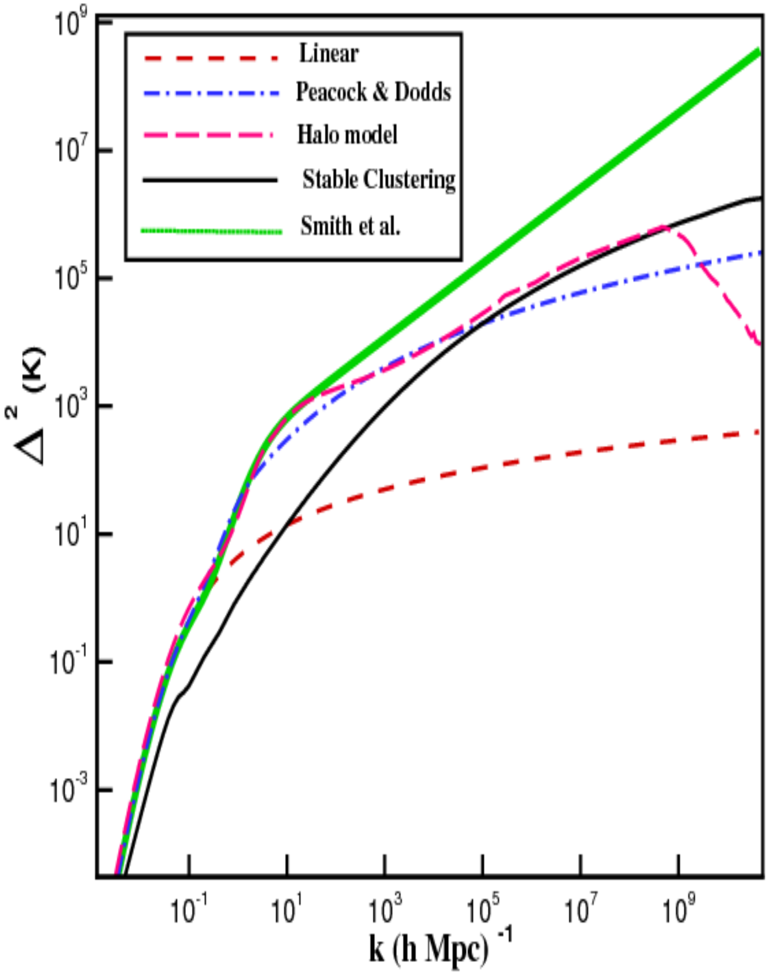
$\Delta < 1$ - weak clustering, linear structure

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

Power spectrum (z=0, concordance model)



Dimensionless power spectrum
(z=0, concordance model)



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).

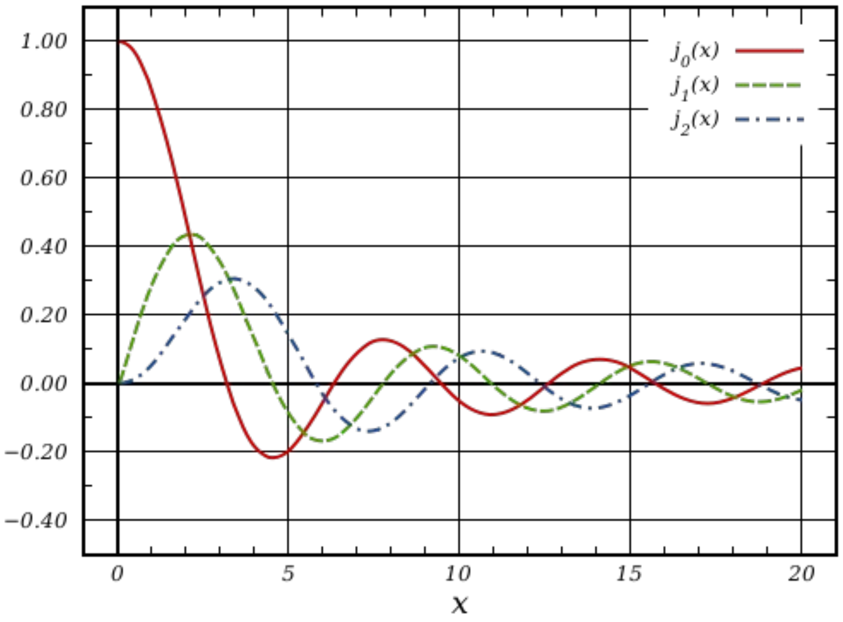
The filter (the function that multiplies k^2 P(k) in the integral) is the **spherical Bessel function** of the first kind for n=0 : j_0 (kr)

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

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

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

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



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 $\langle \delta_k^2 \rangle$

0	1	2	3	4
1	0	1	2	3
2	1	0	1	2
3	2	1	0	1
4	3	2	1	0

Correlation function (real space: locations)

1	-	-	-	-
-	2	-	-	-
-	-	3	-	-
-	-	-	4	-
-	-	-	-	5

Power spectrum (Fourier space: scales)

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).

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

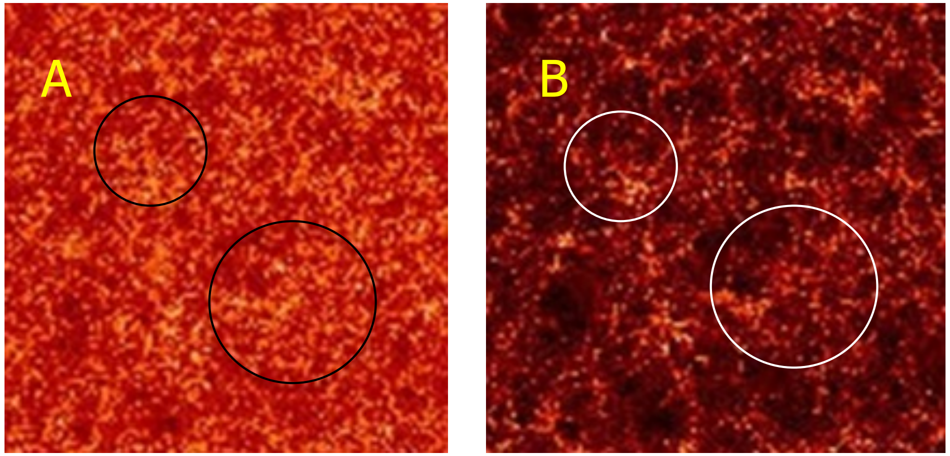
Placing cells of a fixed size R on a δ map (discrete or continuous) allows us to smooth the map on a scale R, defining a δ_R as a convolution of $\delta(x)$ with a window function (a filter) of size R → **δ_R is a weighted average of δ in a cell of size R.**

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).

i)

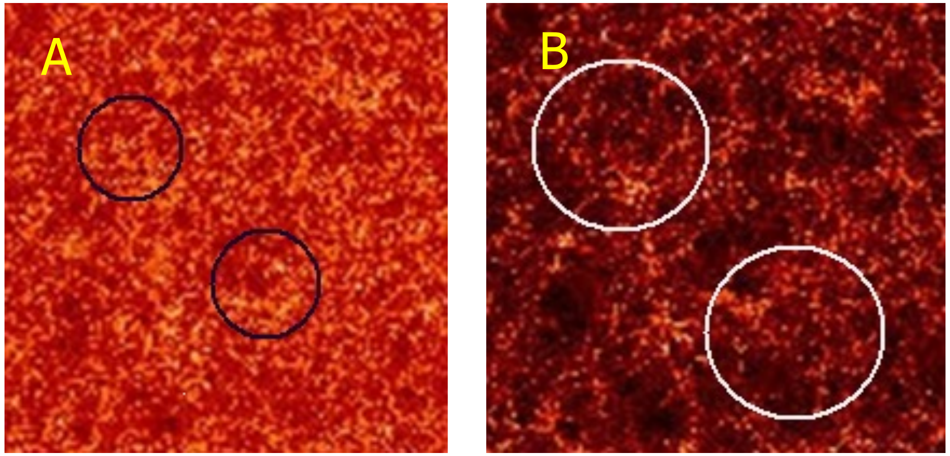


Compute δ_R in each map for the two different values of R, obtaining 4 quantities.

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)



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.

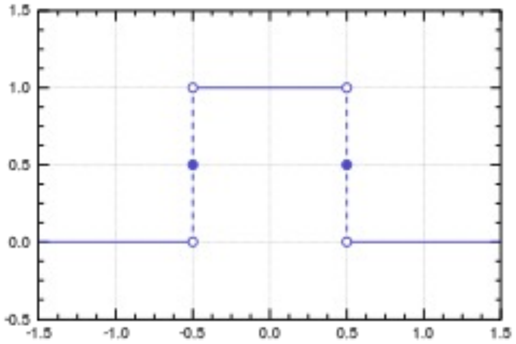
Let us derive that relation.

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:

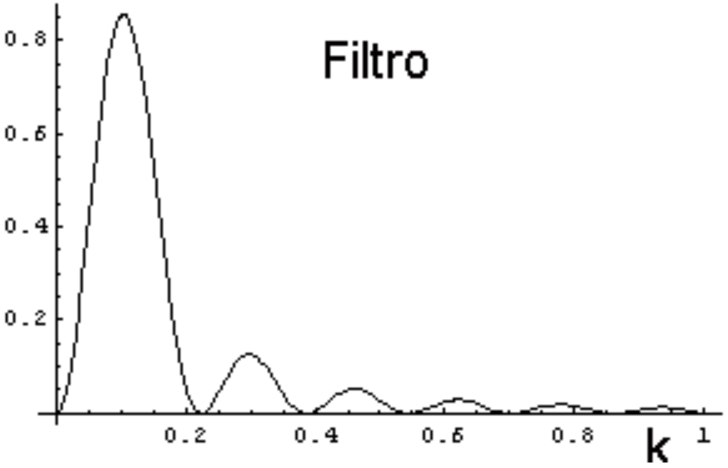
$$\delta_R(k) = \delta(k)\, W_R(k)$$

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 $\delta(x,t)$.

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.