The Inhomogeneous Universe

Statistical properties of the density contrast field

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

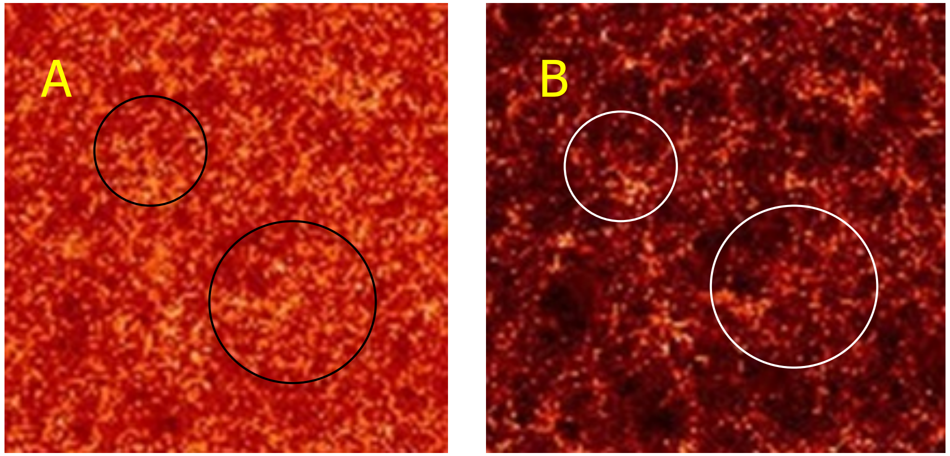
Placing cells of a fixed size R on a δ map (discrete or continuous) allows us to smooth the map on a scale R, defining a δ_R as a convolution of δ(x) with a window function (a filter) of size R → **δ_R is a weighted average of δ in a cell of size R.**

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).

i)

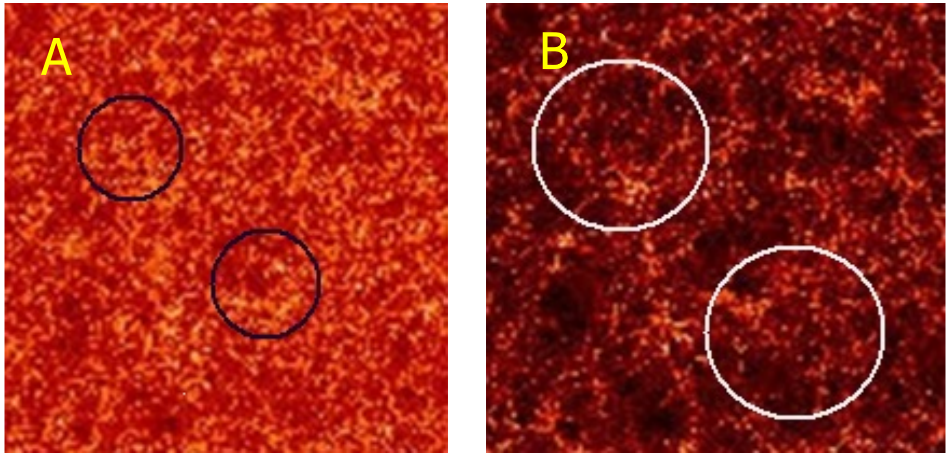


Compute δ_R in each map for the two different values of R, obtaining 4 quantities.

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.

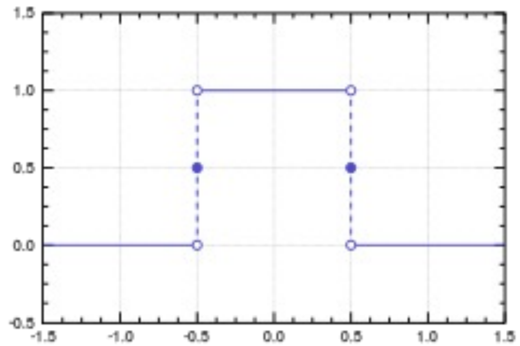
Let us derive that relation.

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:

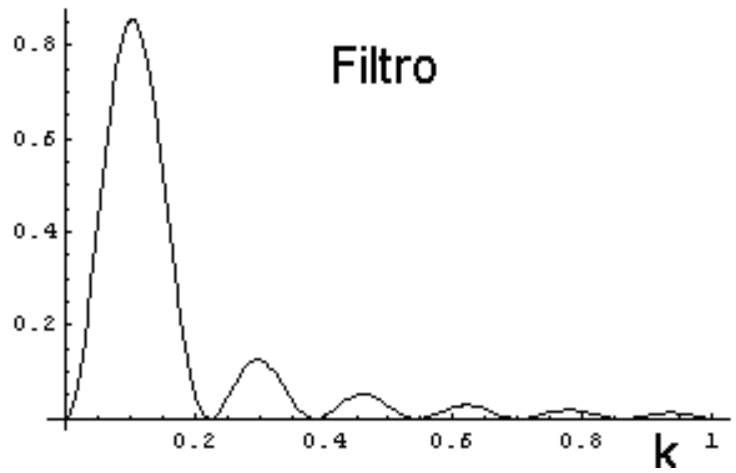
$$\delta_R(k) = \delta(k)\, W_R(k)$$

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 σ_8 = 1 reference universe.

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)

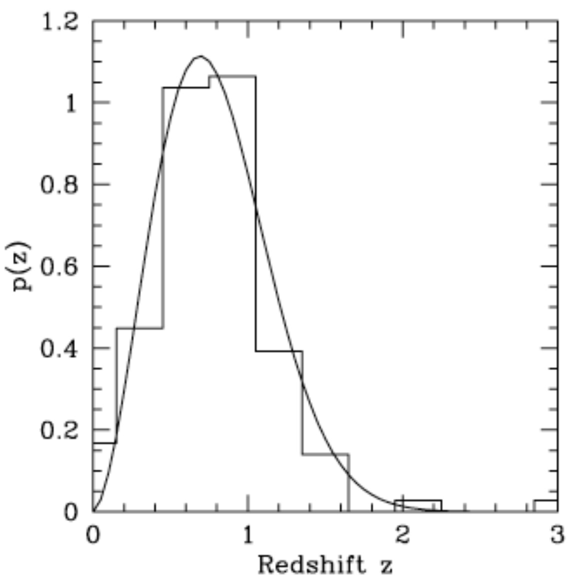
  θ_x is the angular separation (in the x direction) to a reference axis (the line-of-sight)

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

- $g(\chi)$ 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

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

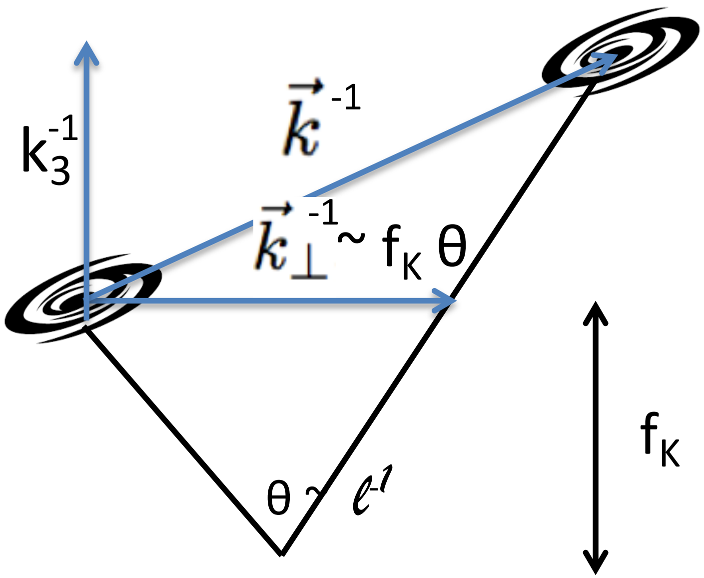
Note the power spectrum evolves in time, and so it also depends explicitly on the redshift z (which is related to χ).

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 $2\pi\,\delta_D(k_3)$ → $k_3 = 0$, i.e:

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

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:

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

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:

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

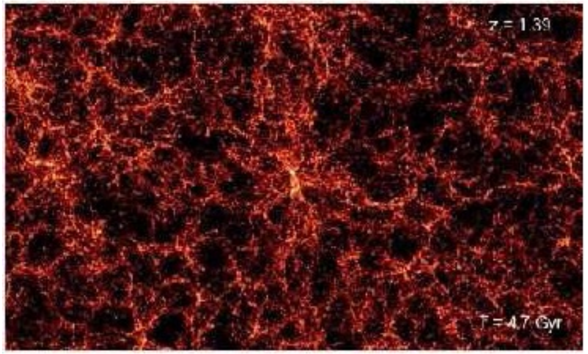
This shows that the amplitude of C for a given angular scale l, is a weighted sum of the amplitudes of P at scales $l/f_K(\chi)$

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

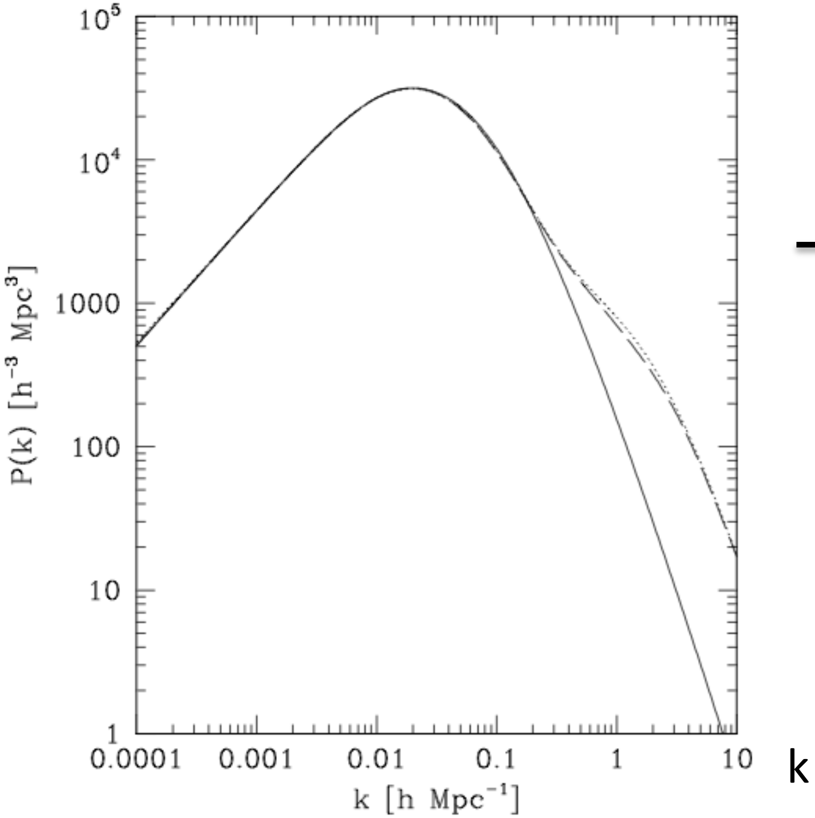
Due to statistical isotropy, the correlation functions only depend on the separation modulus → C(l) is only function of the modulus of ‘l’, as P(k) was function of the modulus of ‘k’.

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)

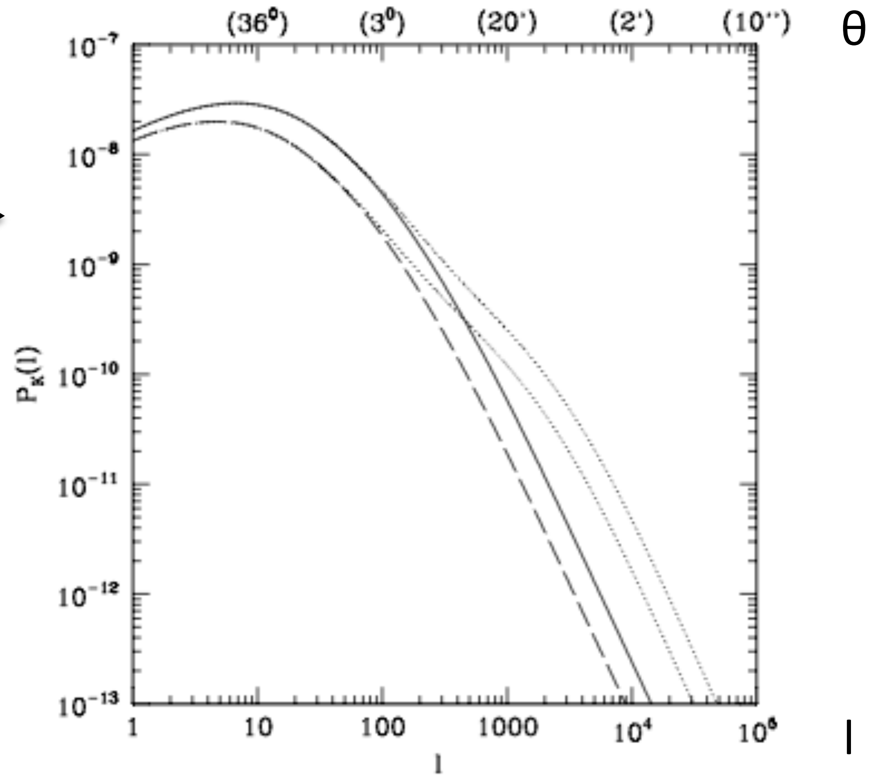


3D matter power spectrum $P_{\delta\delta}(k)$



$k = l / f_K$

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,

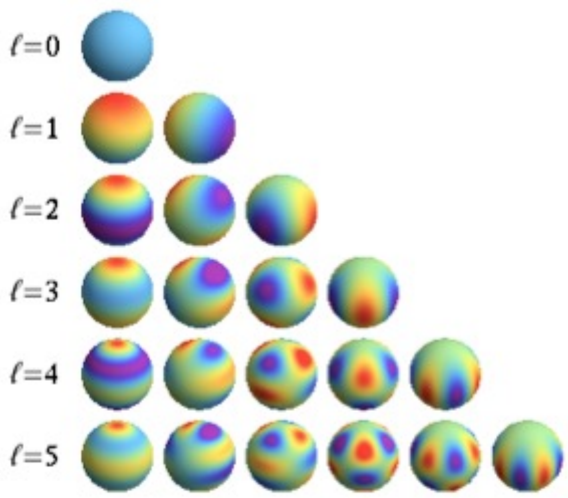
$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left[(x^2-1)^n\right]$$

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:

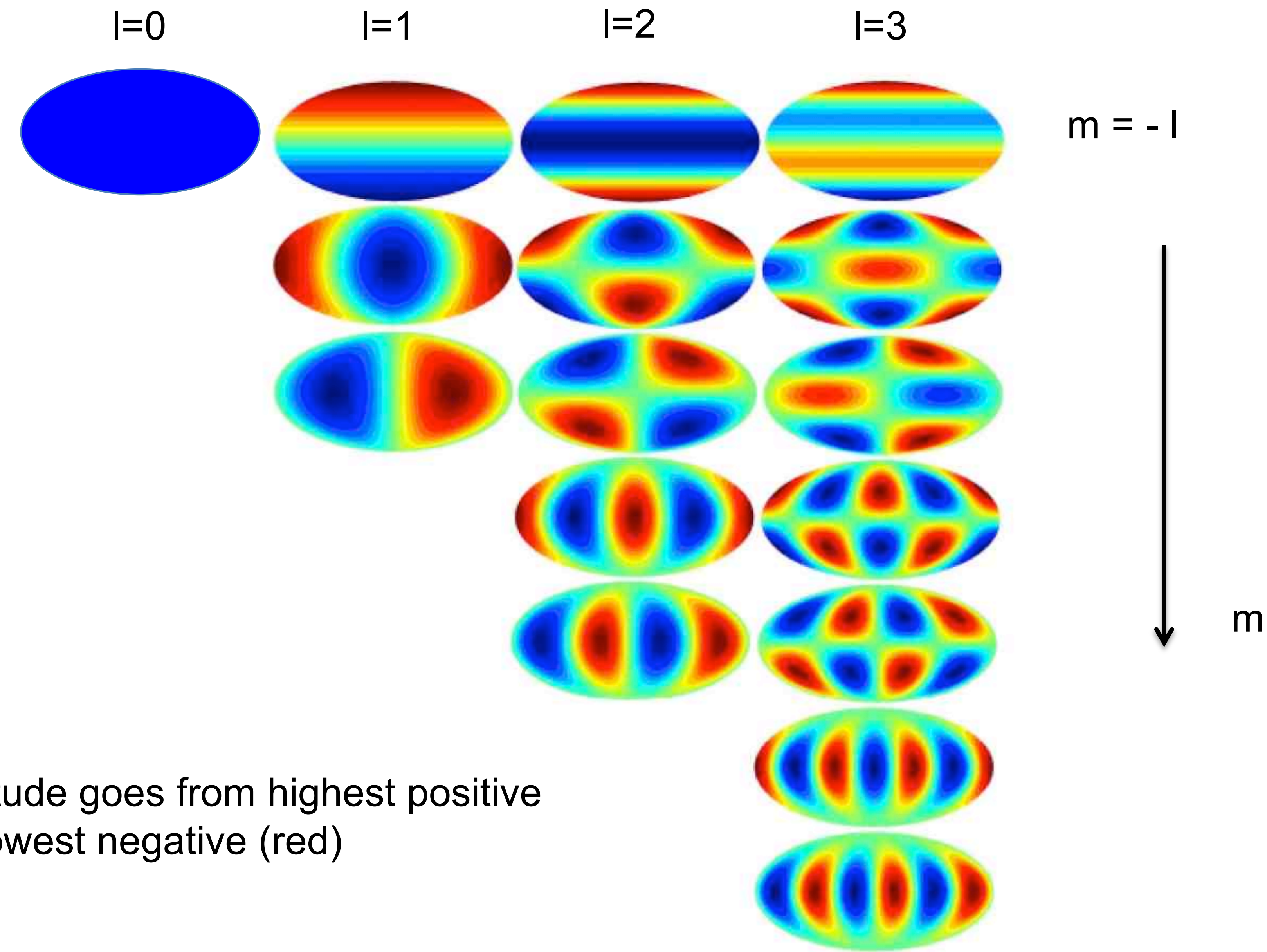
$$\sum_m |Y_{\ell m}(\theta,\phi)|^2 = \frac{2\ell+1}{4\pi}$$

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 $<a_{lm}\, a_{l'm'}>$.
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**).

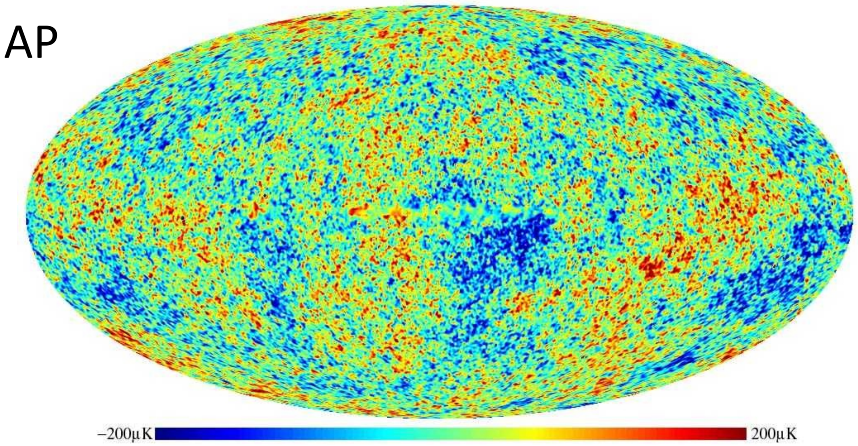
Finally, we can also write the **correlation function in real space**, as function of the isotropic angular power spectrum,

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

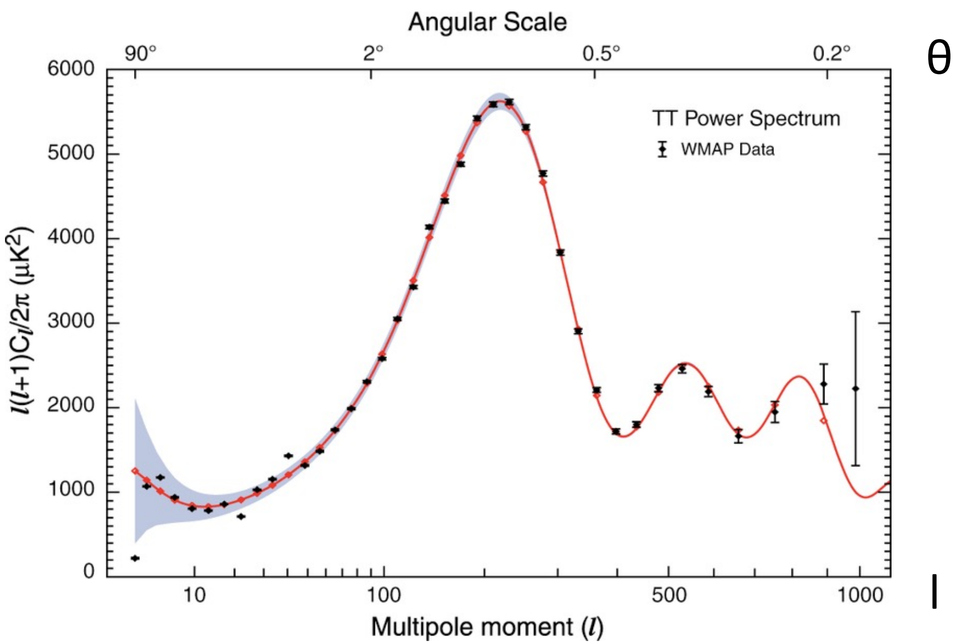
Decomposing a map in spherical harmonics: the CMB temperature contrast

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
$<a_{lm}\ a_{lm}>$ (any m, all are equivalent)

WMAP



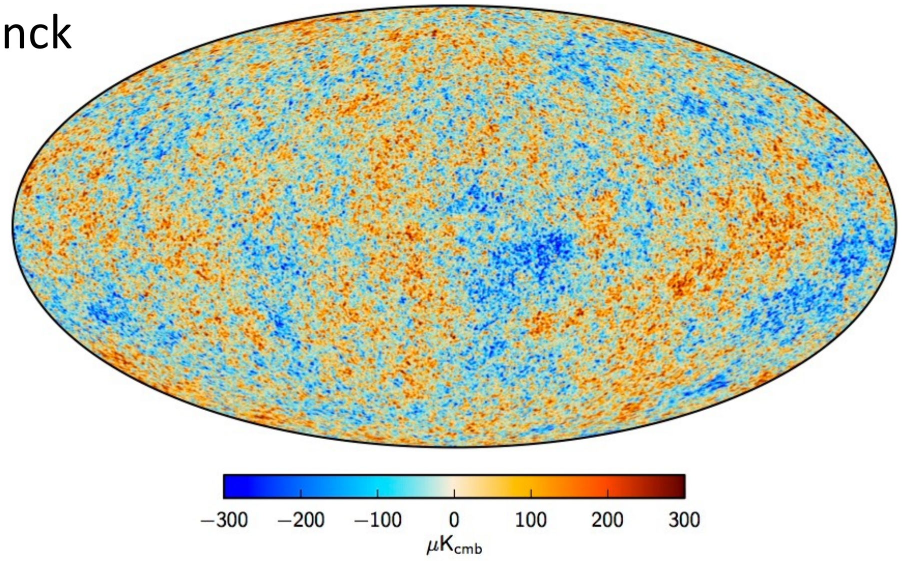
2D angular temperature dimensionless power spectrum $l(l+1)\ C_{TT}(l)$



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

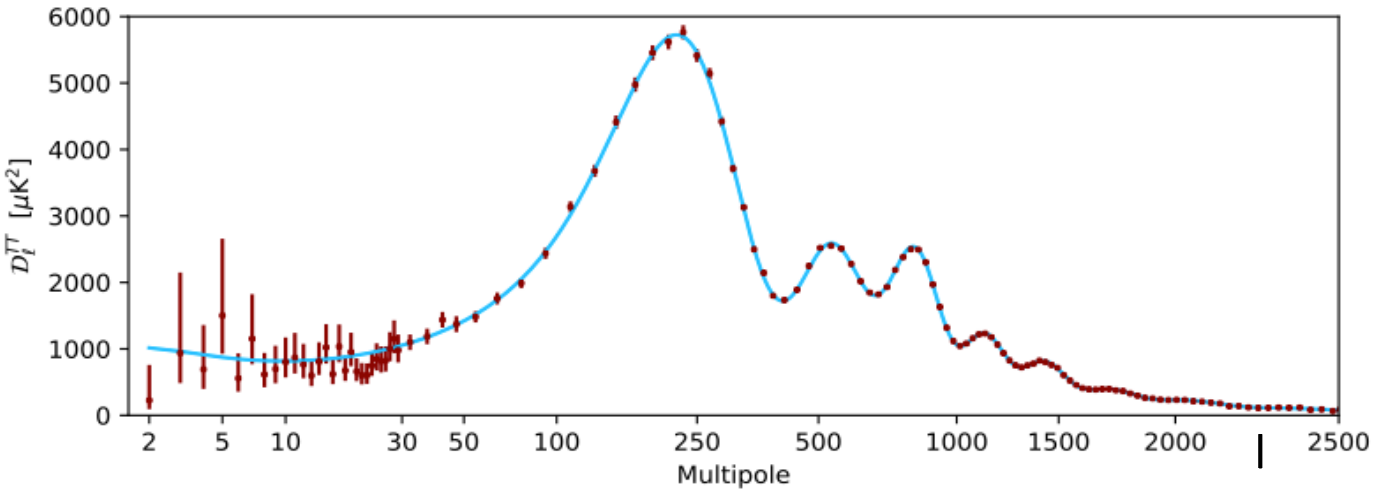


The Planck map was obtained to higher order of the spherical harmonics than the WMAP one.

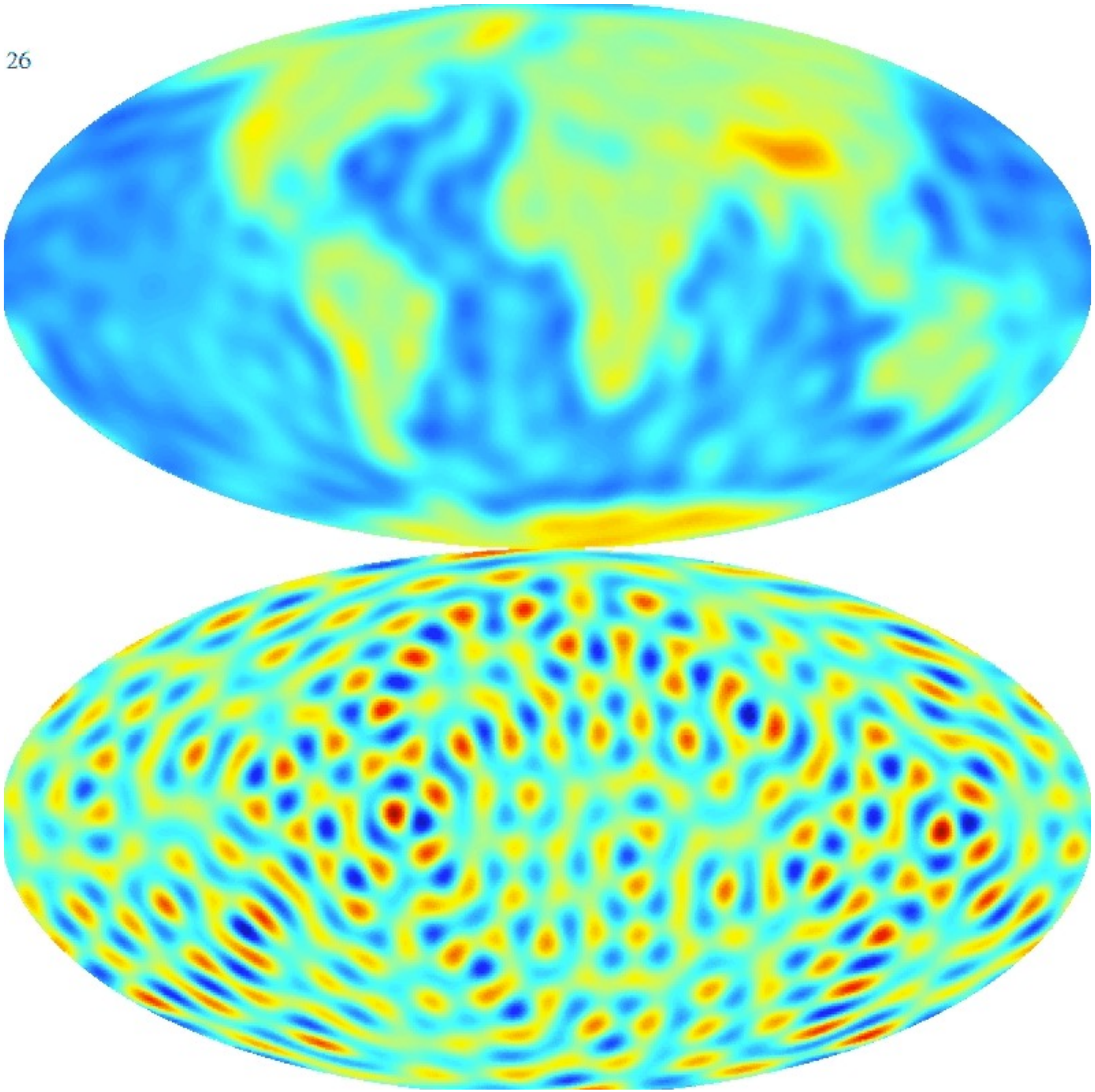
This is noticeable in the map (better resolution and better defined small-scale features)

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 realization, it is not averaged over all the m functions.

Higher-order statistics

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

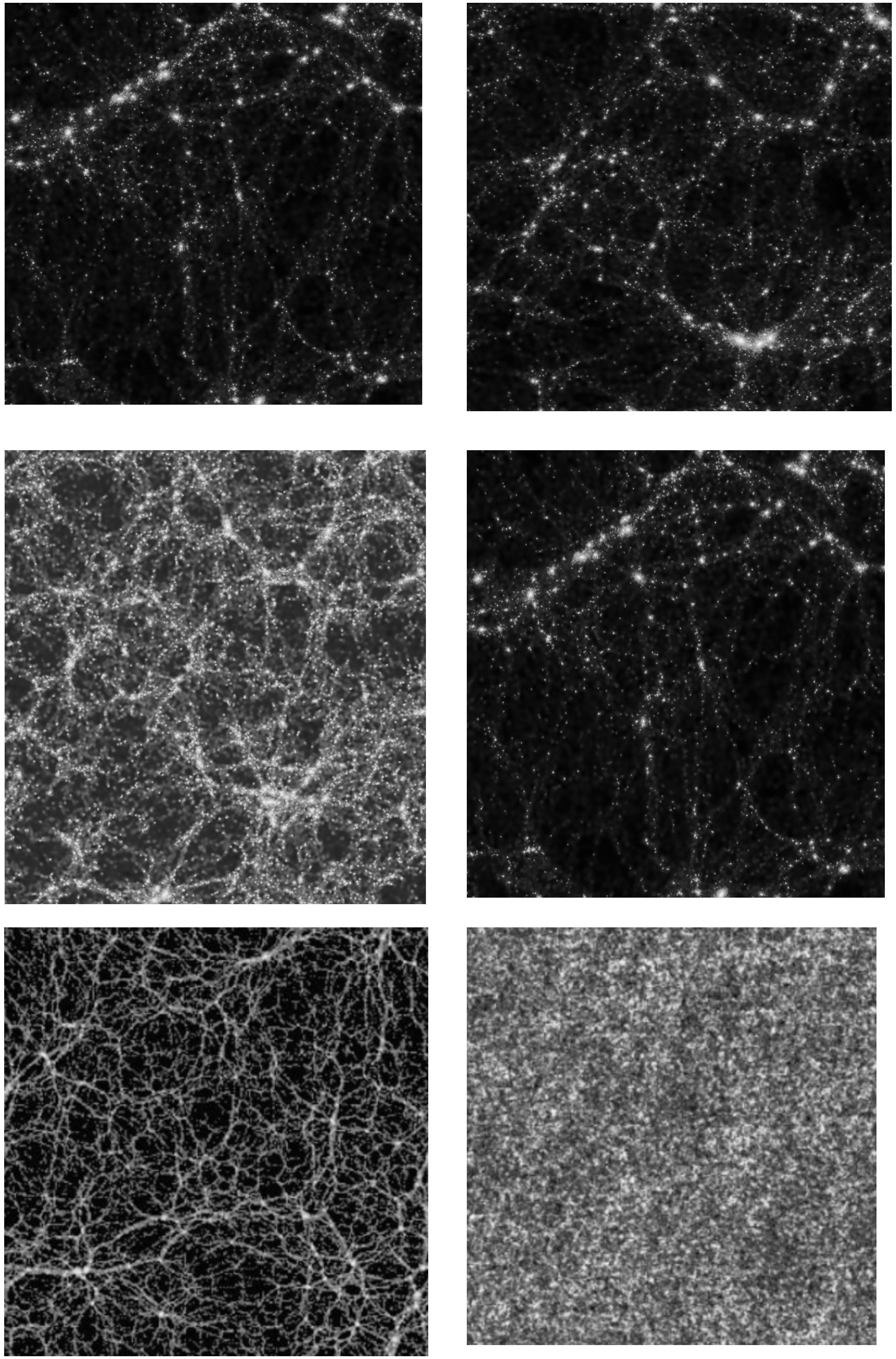
- **Secondary non-Gaussianities**

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

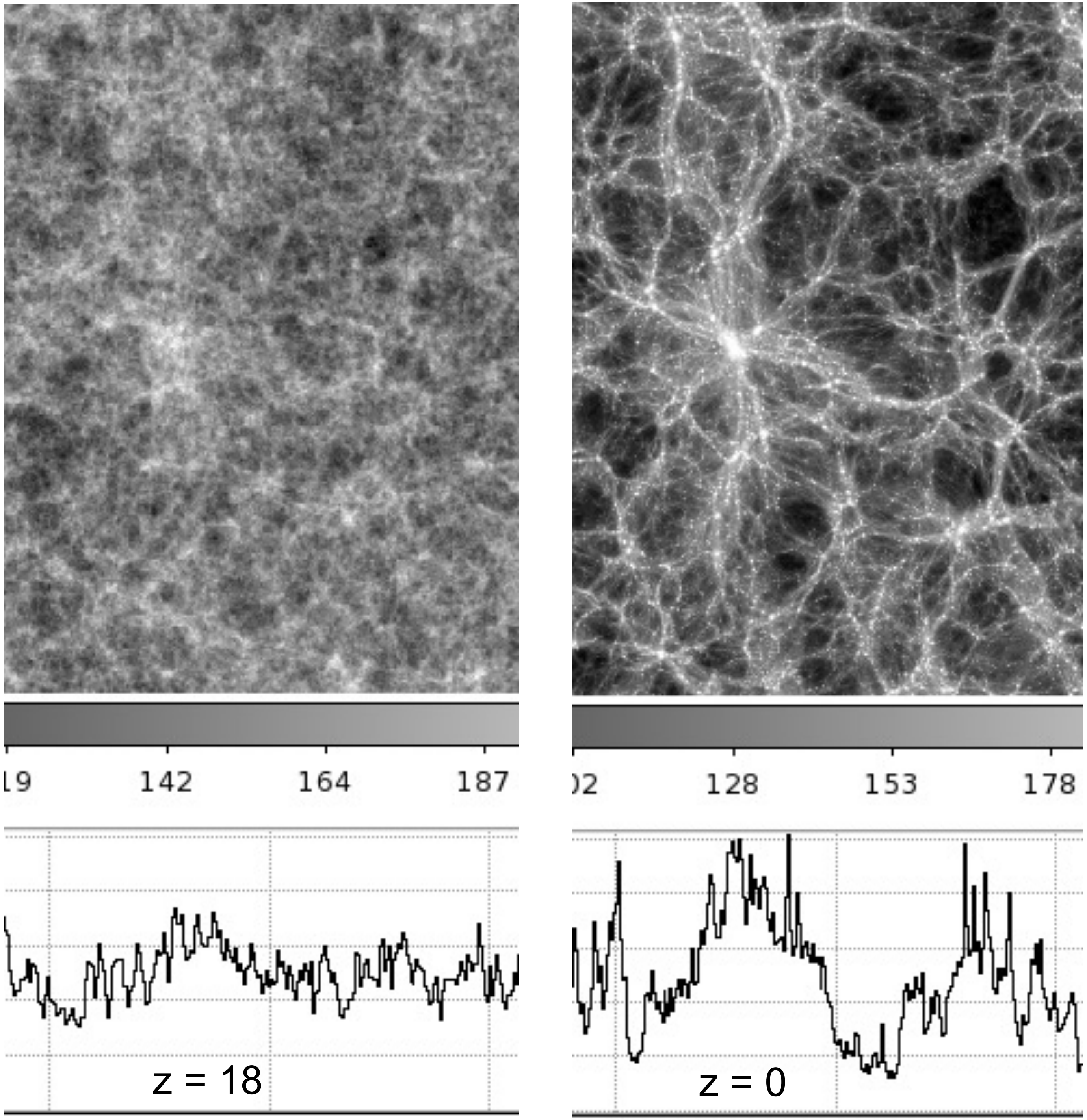


Comparing left and right panels

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



Structure formation: redshift evolution of

- amplitude of the density contrast

- statistical distribution of the density contrast field

Scales: a useful description of clustering → work in modes (Fourier or harmonic space)

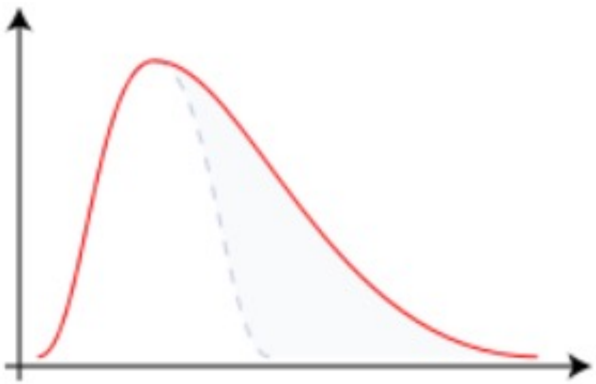
If the distribution is not Gaussian, the covariance matrix (and consequently the 2-pt correlation function and power spectra) do not contain the whole cosmological information

→ we need to consider **higher-order moments.**

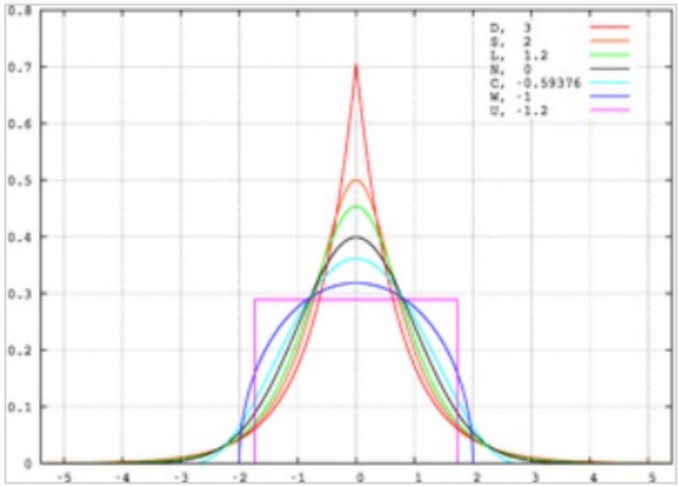
For example:

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



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

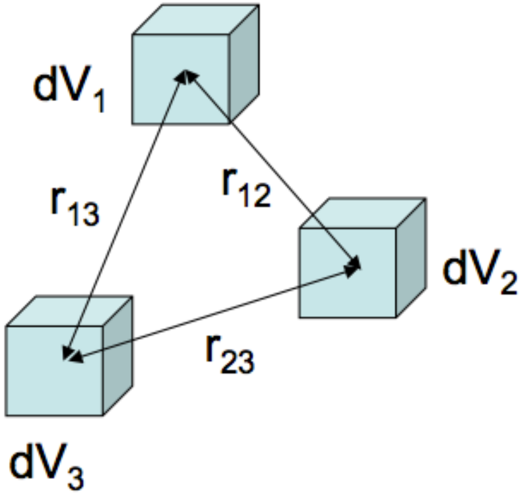


$$K = \frac{\langle \delta^4 \rangle}{\langle \delta^2 \rangle^2} - 3$$

Higher-order moments are computed from 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 galaxies in locations 1,2,3 depends on the full conditional probability between the triplet, and also on all combinations of conditional probabilities between pairs:



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

An **n-point correlation function** can be written as a sum of terms involving lower-order correlations,
plus an irreducible (also called connected) term

→ this is the **Isserlis theorem** of probability theory (1918).

$$\mathrm{E}[\,X_1 X_2 \cdots X_n\,] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \mathrm{E}[\,X_i X_j\,] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \mathrm{Cov}(\,X_i, X_j\,)$$

This also implies that

for **variables of zero mean → the reducible part** of an **odd n-point correlation function** is zero

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

Wick's theorem (1950) - Note there is a version of Isserlis' theorem used in particle physics that allows to reduce the operators in creation/annihilation processes into sums of products of pairs, which is the basis of the description of the process in terms of Feynman diagrams.

3-pt function

Using "Wick's" theorem, the 3-pt correlation function ζ_{123} - zeta - 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**.

In the case of 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

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

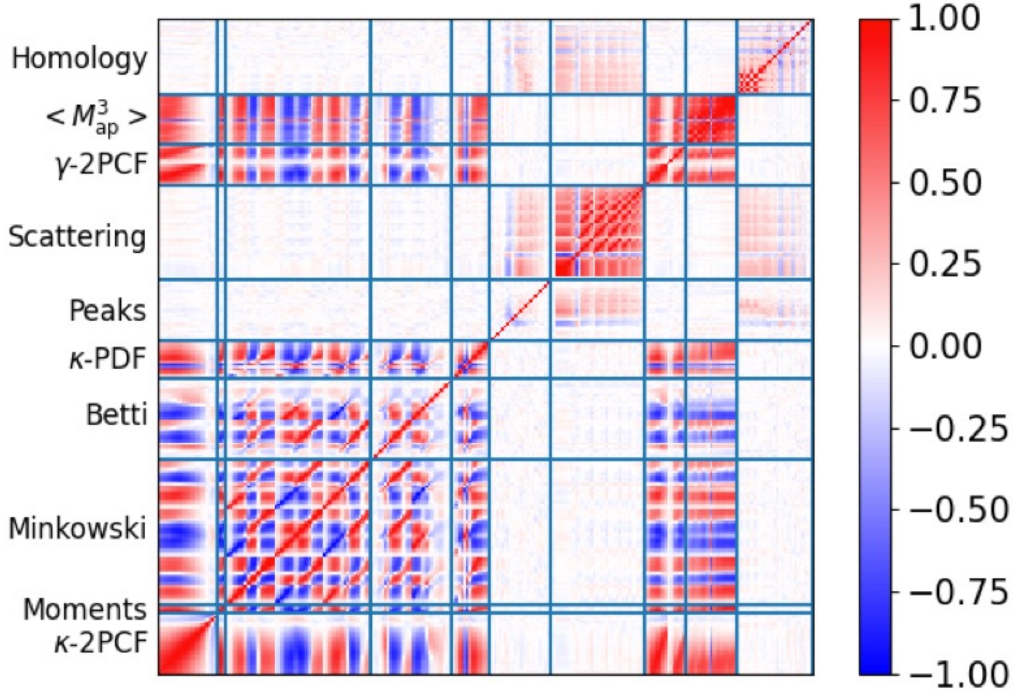
note the **number of terms in the sum** is $\frac{n!}{2^{n/2}\,(n/2)!}$ (in this case, n=4 → n_terms=3)

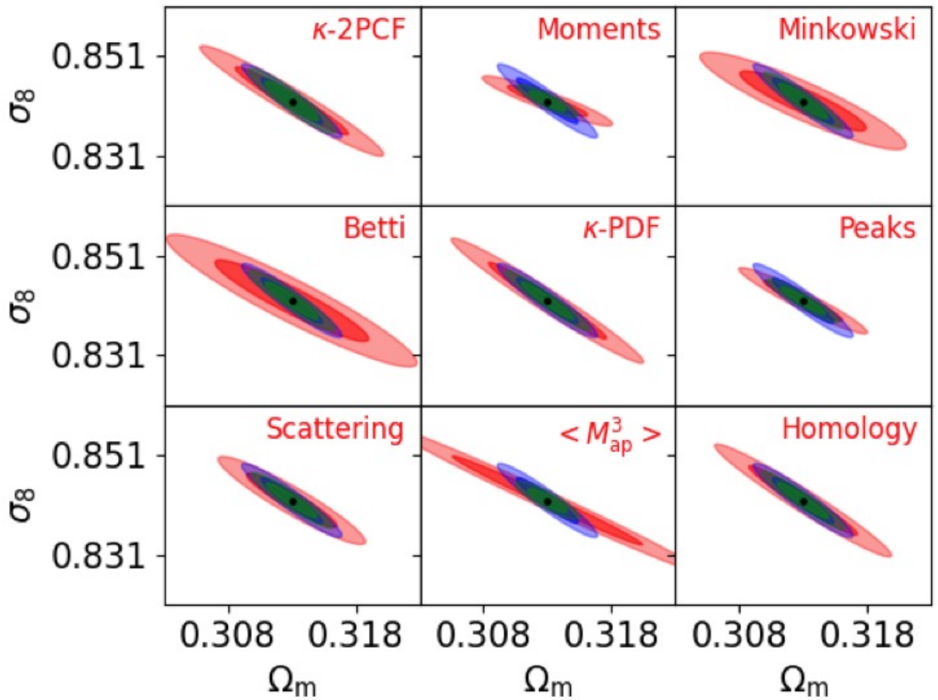
In the case of a Gaussian distribution the connected term is zero**, but the 4-pt function is not zero** → however the "4-point feature" is zero (notice the definition of kurtosis).

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

There are **different ways of defining higher-order statistics**.

Some are based on n-point correlation functions (like the skewness and kurtosis), others are based on different properties of the field (e.g. number of peaks, area of connected regions - topological features - etc).





Each one combines the underlying cosmological information in a different way → they depend on the cosmological parameters in different, sometimes complementary, ways.

Cosmic variance

Besides being important cosmological functions with valuable information needed to characterize cosmological maps and models**, higher-order statistics are also needed to compute the uncertainty of 2-pt functions.**

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,** that we want to estimate from measurements in a map, is C_l with any value of m.

The best way to **estimate** the theoretical power spectrum from a map 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, meaning that if many measurements were made (N → ∞) its average would give exactly the theoretical power spectrum:

$$\langle \widehat{C}_\ell \rangle = C_\ell \pm \sigma / N$$

This is the same as when estimating the mean of a distribution by computing the average of N measurements (the larger is N, the more precise is the estimate).

However, the maximum independent measurements that can be made of each multipole is limited: it is given by 2l+1,

so 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 defined as the covariance (dispersion) of the estimator. This can be computed theoretically, which is more rigorous than just measuring the dispersion between various measurements (which depends on the specific sample measured).

The expression is:

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

Note that in general this expression 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'**

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

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

(where $<C_l> = C_l$ is the theoretical value)

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

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

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

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

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

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:

$$\sum_{m,m'} \langle a_{lm} a_{lm} a_{lm'} a_{lm'} \rangle = \sum_{m,m'} \left[\langle a_{lm} a_{lm} \rangle \langle a_{lm'} a_{lm'} \rangle + 2 \langle a_{lm} a_{lm'} \rangle \langle a_{lm} a_{lm'} \rangle + \langle a_{lm} a_{lm} a_{lm'} a_{lm'} \rangle_c \right]$$

Now, using the result

$$\sum_{m,m'} \langle a_{lm} a_{lm} \rangle = (2l+1) C_l$$

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

$$\langle \hat{C}_l^2 \rangle = \frac{C_l^2}{(2l+1)^2} \left[(2l+1)^2 + 2(2l+1) \right]$$

And so the cosmic variance is:

$$\langle \hat{C}_l^2 \rangle - C_l^2 = C_l^2 \left(1 + \frac{2}{2l+1} - 1 \right)$$

i.e.,

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

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

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.**

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.

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:

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

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 $\sigma_l / C_l = (2/3)^{0.5} = 81\%$

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