Cosmologia Física

Ismael Tereno (FCUL, IA)

2026

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)

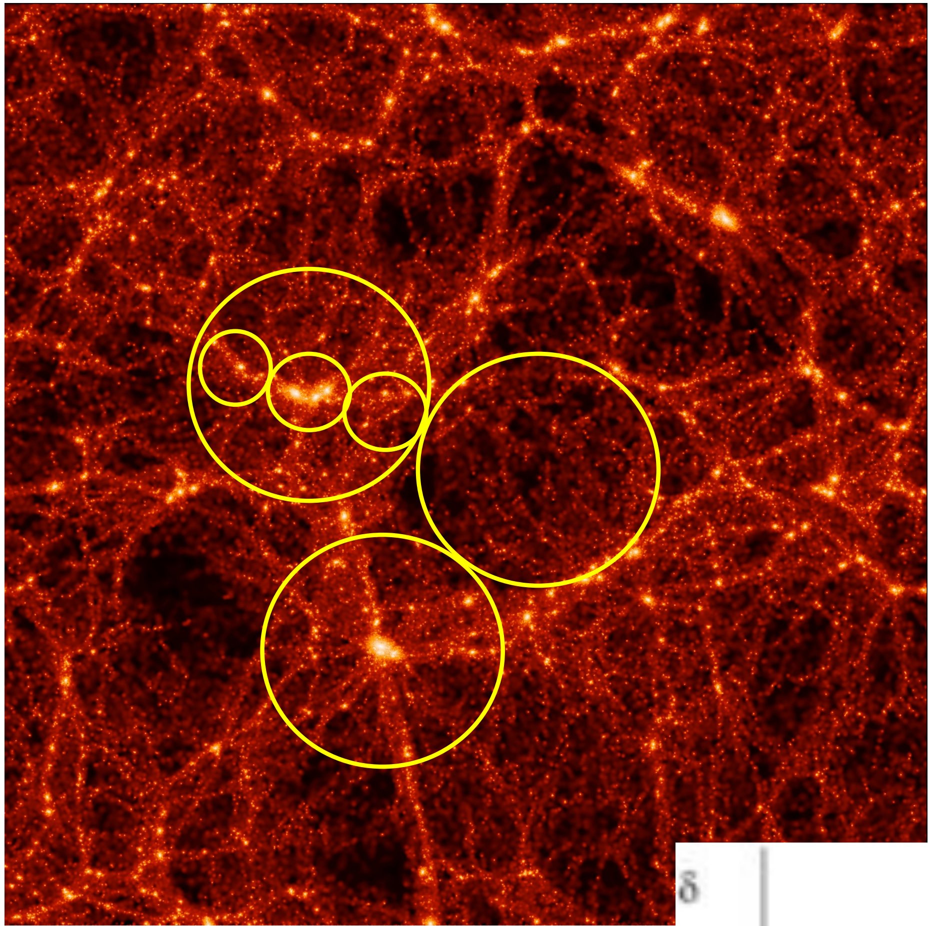
$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}} - 1$$

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.

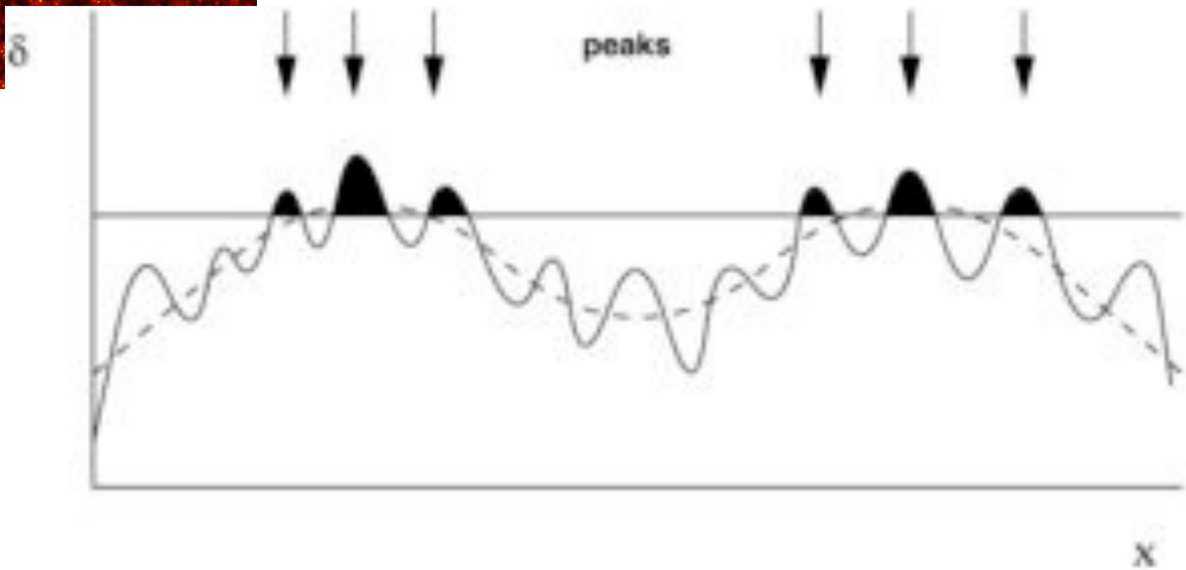
δ can become very large (not a density perturbation anymore) but the associated gravitational potential always remains a perturbation to the metric.



Density map and scales

overdensities
and
undersdensities

on two different scales



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

The reason is: **quantum fluctuations**

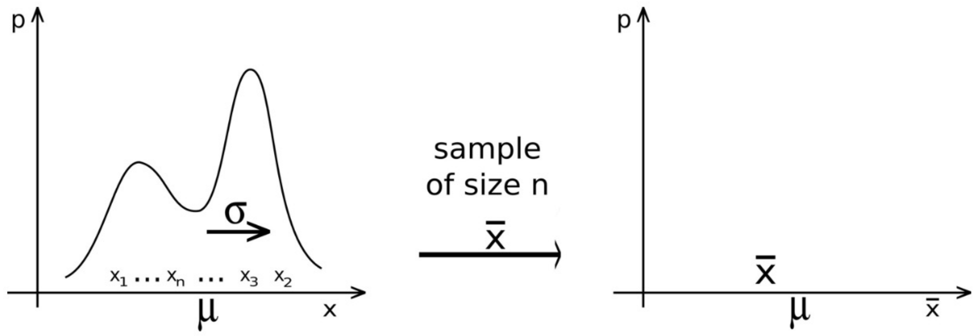
In the quantum universe, there is a large number of random steps, i.e., in the very early Universe the value of density at a given location changes all the time as the result of a stochastic (random) process.

(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.

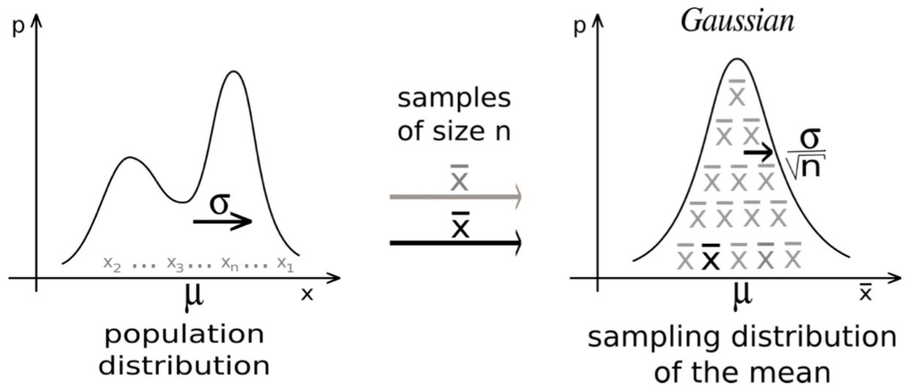
Initial population of *δ* values in a location: histogram of values of *δ* during the stochastic process (which forms a **population**)

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 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 δ(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**, $\boldsymbol{\sigma}$, 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 multivariate 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:

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

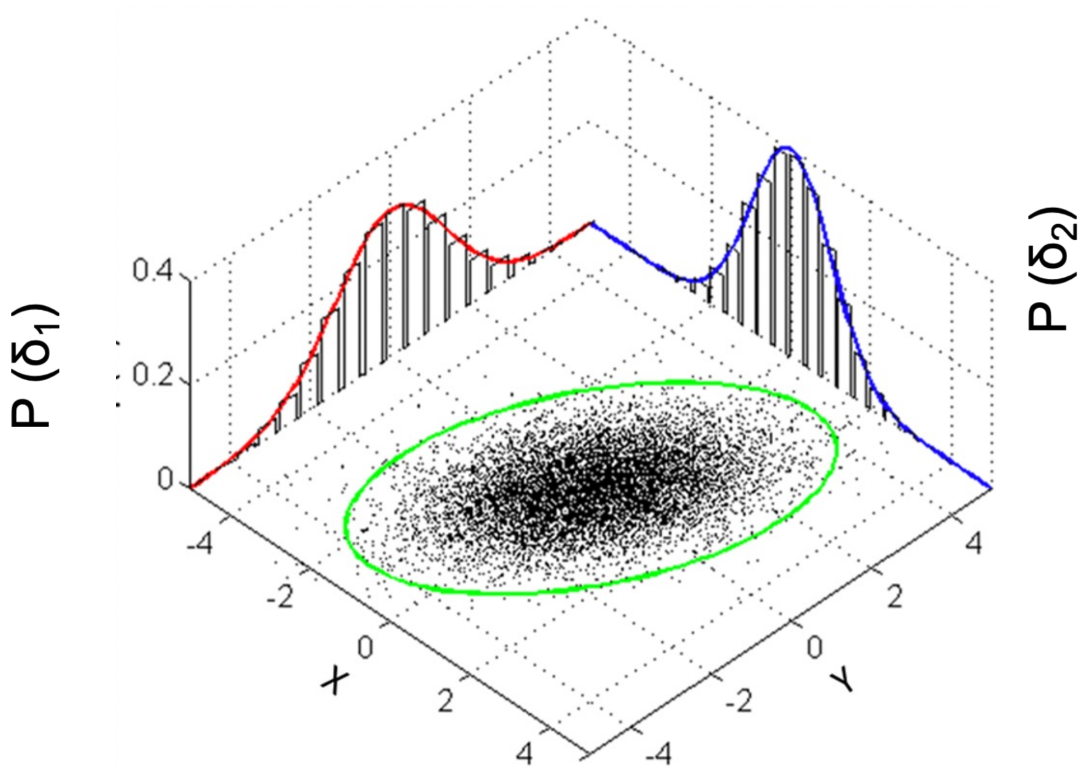
The correlation between the variables can also be written in the form of a **correlation matrix**

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

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)

Apparently, the stochasticity increases the complexity of the treatment of the first-order density field:

If the problem was **deterministic**:

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

Because the problem is **stochastic**:

system not described by the actual values of $\delta(x)$ 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:

$$\langle\delta\rangle=0$$

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:

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

(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.