Lecture 2: Description of random fields
Licia Verde

1 catching up

1.1 Likelihoods

One can be more sophysticated than χ^2, if $P(D)$ (D is data) is known. Remember from the Bayes theorem (eq.??) the probability of the data given the model (Hypothesis) is the likelihood. If we set $P(D) = 1$ (after all we got the data) and ignore the prior by maximizing the likelihood we find the most likely Hypothesis, or , often, the most likely parameters of a given model.

Note that we have ignored $P(D)$ and the prior so in general this technique does not give you a goodness of fit and not an absolute probability of the model, only relative probabilities. Frequentists rely on χ^2 analyses where a goodness of fit can be established.

In many cases (thanks to the central limit theorem) the likelihood can be well approximated by a multi-variate Gaussian:

$$\mathcal{L} = \frac{1}{(2\pi)^{n/2}|detC|^{1/2}} \exp\left[-\frac{1}{2}\sum_{ij}(D-y)_i C^{-1}_{ij}(D-y)_j\right] \tag{1}$$

where $C_{ij} = \langle (D_i - y_i)(D_j - y_j)\rangle$ is the covariance matrix.

Exercise: when are likelihood analyses and χ^2 analyses the same?

1.2 Confidence levels for likelihood

For Bayesian statistics, confidence regions are found as regions R in *model space* such that $\int_R P(\vec{\alpha}|D)d\vec{\alpha}$ is ,say, 0.68 for 68% confidence level and 0.95 for 95% confidence. Note that this encloses the prior information. To report results independently of the prior the likelihood ratio is used. In this case compare the likelihood at a particular point in model space $\mathcal{L}(\vec{\alpha})$ with the value of the maximum likelihood $\mathcal{L}_{max}$. Then a model is said acceptable if

$$-2\ln\left[\frac{\mathcal{L}(\vec{\alpha})}{\mathcal{L}_{max}}\right] \leq \text{threshold} \tag{2}$$

Then the threshold should be calibrated by calculating the distribution of the likelihood ratio in the case where a particular model is the true model. There are some cases however when the value of the threshold is the corresponding confidence limit for a χ^2 with m degrees of freedom, for m the number of parameters. (The data must have Gaussian errors, the model must depend linearly on the parameters, the gradients of the model wrt the parameters are not degenerate, the parameters do not affect the covariance).

1.3 Marginalization, combining different experiments

Of all the model parameters α_i some of them may be uninteresting. Typical examples of nuisance parameters are calibration factors, galaxy bias parameter etc, but also it may be that we are interested on constraints on only one cosmological parameter at the time rather than on the *joint* constraints on 2 or more parameters simultaneously. One then marginalizes over the uninteresting parameters by integrating the posterior distribution:

$$P(\alpha_1..\alpha_j|D) = \int d\alpha_{j+1}, ...\alpha_m P(\vec{\alpha}|D) \qquad (3)$$

If there are in total m parameters and we are interested in j of them ($j < m$).

Note that if you have two independent experiments the combined likelihood of the two experiments is just the product of the two likelihoods. (of course if the two experiments are non independent then one would have to include their covariance). In many cases one of the two experiments can be used as a prior. A word of caution is on order here. We can always combine independent experiments by multiplying their likelihoods, and if the experiments are good and sound and the model used is a good and complete description of the data all is well. However it is always important to a) think about the priors one is using and to quantify their effects. b) to make sure that results from independent experiments are consistent: by multiplying likelihood from inconsistent experiments you can always get some sort of results but it does not mean that the result actually makes sense....

Sometimes you may be interested in placing an prior on the uninteresting parameters before marginalization. The prior may come from a previous measurement or from your "belief".

Typical examples of this are: marginalization over calibration uncertainty, over point sources amplitude or over beam errors for CMB studies. It is useful to know of the following trick for Gaussian likelihoods:

$$\begin{aligned} P(\alpha_1..\alpha_{m-1}|D) &= \int \frac{dA}{(2\pi)^{m/2}||C||^{1/2}} e^{\left[-\frac{1}{2}(C_i-(\hat{C}_i+AP_i))\Sigma^{-1}_{ij}(C_j-(\hat{C}_j+AP_j))\right]} \\ &\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(A-\hat{A})^2}{\sigma^2}\right] \end{aligned} \qquad (4)$$

repeated indices are summed over. A the amplitude of , say, a point source constribution ot the C_ℓ angular power spectrum is the $m-th$ parameter which we want to marginalize over with a Gaussian prior with variance σ^2. The trick is to recognize that this integral can be written as:

$$P(\alpha_1..\alpha_{m-1}|D) = C_0 \exp\left[-\frac{1}{2}C_1 - 2C_2 A + C_3 A^2\right] dA \qquad (5)$$

(where $C_{0...3}$ denote constants) and that this kind of integral is evaluated by using the substitution $A \longrightarrow A - C_2/C_3$ giving something $\propto \exp[-1/2(C_1 - C_2^2/C_3)]$.

It is left as an exercise to write the constants explicitly.

1.4 An example

Let's say you want to constrain cosmology by studying clusters number counts as a function of redshift. The observation of a discrete number N of clusters is a Poisson process, the probability of which is given by the product

$$P = \Pi_{i=1}^{N}[e_i^{n_i}\exp(-e_i)/n_i!] \tag{6}$$

where n_i is the number of clusters observed in the $i-th$ experimental bin and e_i is the expected number in that bin in a given model:$e_i = I(x)\delta x_i$ with i being the proportional to the probability distribution. Here δx_i can represent an interval in clusters mass and/or redshift. Note: this is a product of Poisson distributions, thus one is assuming that these are independent processes. Clusters may be clustered, so when can this be used?

For unbinned data (or for small bins so that bins have only 0 and 1 counts) we define the quantity:

$$C \equiv -2\ln P = 2(E - \sum_{i=1}^{N}\ln I_i) \tag{7}$$

where E is the total expected number of clusters in a given model. The quantity ΔC between two models with different parameters has a χ^2 distribution! (so all that was said in the χ^2 section applies, even though we started from a highly non-Gaussian distribution.

(This is from the paper of Cash 1979)

2 Introduction

Let's take a break from probabilities for a while...

In comparing the results of theoretical calculations with the observed Universe, it would be meaningless to hope to be able to describe with the theory the properties of a particular patch, i.e. to predict the density contrast of the matter $\delta(\vec{x}) = \delta\rho(x)/\rho$ at any specific point $\vec{x}$. Instead, it is possible to predict the average statistical properties of the mass distribution [1]. Following the *cosmological principle* (e.g. Peebles 1980), models of the Universe have to be homogeneous on the average, therefore, in widely separated regions of the Universe (i.e. independent), the density field must have the same statistical properties.

A crucial assumption of standard cosmology is that the part of the Universe that we can observe is a *fair sample* of the whole. This is closely related to the *cosmological principle* since it implies that the statistics like the correlation functions have to be considered as averages over the ensemble. But the peculiarity in cosmology is that we have just one Universe, which is just one realization from the ensemble (quite fictitious one: it is the ensemble of all possible Universes). The fair sample hypothesis states that samples from well separated part of the Universe are independent realizations of the same physical process, and that, in the observable part of the Universe, there are enough independent samples to be representative of the statistical ensemble. The hypothesis of ergodicity follows: averaging over many realizations is equivalent to averaging over a large (enough) volume. The cosmological field we are interested in, in a given volume, is taken as a realization of the statistical process and, for the hypothesis of ergodicity, averaging over many realizations is equivalent to averaging over a large volume.

[1] A very similar approach is taken in statistical mechanics.

Theories can just predict the statistical properties of $\delta(\vec{x})$ which, for the cosmological principle, must be a homogeneous and isotropic random field, and our observable Universe is a random realization from the ensemble.

In cosmology the scalar field $\delta(\vec{x})$ is enough to specify the initial fluctuations field, and we hope also the present day distribution of galaxies and matter. Here lies one of the big challenges of modern cosmology.

A fundamental problem in the analysis of the cosmic structures, is to find the appropriate tools to provide information on the distribution of the density fluctuations, on their initial conditions and subsequent evolution. here we concentrate on power spectra and correlation functions.

3 Gaussian random fields

Gaussian random fields are crucially important in cosmology, for different reasons: first of all it is possible to describe their statistical properties analytically, but also there are strong theoretical motivations, namely inflation, to assume that the primordial fluctuations that gave rise to the present-day cosmological structures, follow a Gaussian distribution. Without resorting to inflation, for the central limit theorem, Gaussianity results from a superposition of a large number of random processes.

The distribution of density fluctuations δ defined as [2] $\delta = \delta\rho/\rho$ cannot be exactly Gaussian because the field has to satisfy the constraint $\delta > -1$, however if the amplitude of the fluctuations is small enough, this can be a good approximation. This seems indeed to be the case: by looking at the CMB anisotropies we can probe fluctuations when their statistical distribution should have been close to its primordial one; at present there is no uncontroversial evidence that the primordial density field is not Gaussian[3]. If δ is a Gaussian random field with average 0, its probability distribution is given by:

$$P(\delta)d\delta = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\delta^2}{2\sigma^2}\right] d\delta \tag{8}$$

where σ is the r.m.s. of the δ field. Also the N-variate joint distribution of a set of $\delta_i \equiv \delta(\mathbf{x}_i)$ can easily be written as a multivariate Gaussian:

$$P_n(\delta_1, \cdots, \delta_n) = \frac{\sqrt{\mathrm{Det}\mathbf{C}^{-1}}}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\delta^T \mathbf{C}^{-1} \delta\right] \tag{9}$$

where δ is a vector made by the δ_i, $\mathbf{C}^{-1}$ denotes the inverse of the correlation matrix which elements are $\mathbf{C}_{ij} = \langle \delta_i \delta_j \rangle$.

An important property of Gaussian random fields is that the Fourier transform of a Gaussian field is still Gaussian. The phases of the Fourier modes are random and the real and imaginary part of the coefficients have Gaussian distribution and are mutually independent.

[2] Note that $\langle \delta \rangle = 0$

[3] Note the wordings

Let us denote the real and imaginary part of $\delta_{\mathbf{k}}$ by $Re\delta_{\mathbf{k}}$ and $Im\delta_{\mathbf{k}}$ respectively. Their joint probability distribution is the bivariate Gaussian:

$$P(Re\delta_{\mathbf{k}}, Im\delta_{\mathbf{k}})dRe\delta_{\mathbf{k}}dIm\delta_{\mathbf{k}} = \frac{1}{2\pi\sigma_k^2}\exp\left[-\frac{Re\delta_{\mathbf{k}}^2 + Im\delta_{\mathbf{k}}^2}{2\sigma_k^2}\right] dRe\delta_{\mathbf{k}}dIm\delta_{\mathbf{k}} \qquad (10)$$

where σ_k^2 is the variance in $Re\delta_{\mathbf{k}}$ and $Im\delta_{\mathbf{k}}$ and for isotropy it depends only on the magnitude of $\mathbf{k}$. Equation (10) can be re-written in terms of the amplitude $|\delta_{\mathbf{k}}|$ and the phase $\phi_{\mathbf{k}}$:

$$P(|\delta_{\mathbf{k}}|, \phi_{\mathbf{k}})d|\delta_{\mathbf{k}}|d\phi_{\mathbf{k}} = \frac{1}{2\pi\sigma_k^2}\exp\left[-\frac{|\delta_{\mathbf{k}}|^2}{2\sigma_k^2}\right] |\delta_{\mathbf{k}}|d|\delta_{\mathbf{k}}|d\phi_{\mathbf{k}} \qquad (11)$$

that is $|\delta_k|$ follows a Rayleigh distribution.

From this follows that the probability that the amplitude is above a certain threshold X is:

$$P(|\delta_{\mathbf{k}}|^2 > X) = \int_{\sqrt{X}}^{\infty} \frac{1}{\sigma_k^2}\exp\left[-\frac{|\delta_{\mathbf{k}}|^2}{2\sigma_k^2}\right] |\delta_{\mathbf{k}}|d|\delta_{\mathbf{k}}| = \exp\left[-\frac{X}{\langle|\delta_{\mathbf{k}}|^2\rangle}\right] . \qquad (12)$$

Which is an exponential distribution.

The fact that the phases of a Gaussian field are random, implies that the two point correlation function (or the power spectrum) completely specifies the field.

P.S. If your advisor now asks you to generate a Gaussian random field you know how to do it. (It you are not familiar with Fourier transforms see next section)

The observed fluctuation field however is not Gaussian. The observed galaxy distribution is highly non-Gaussian principally due to gravitational instability; the observed CMB fluctuations are non-Gaussian[4]. To completely specify a non-Gaussian distribution higher order correlation functions are needed[5]; conversely deviations from Gaussian behavior can be characterized by the higher-order statistics of the distribution.

4 Basic tools

The structure of the Universe on large scales is largely dominated by the force of gravity (which we think we know well) and no too much by complex mechanisms (baryonic physics, galaxy formation etc.)- or at least that's the hope...

The Fourier transform of the (fractional) overdensity filed δ is defined as:

$$\delta_{\vec{k}} = A\int d^3r\delta(\vec{r})\exp[-i\vec{k}\cdot\vec{r}] \qquad (13)$$

with inverse

$$\delta(\vec{r}) = B\int d^3k\delta_{\vec{k}}\exp[i\vec{k}\cdot\vec{r}] \qquad (14)$$

[4]This does not mean that the CMB temperature fluctuation is intrinsically non-Gaussian

[5]For "non pathological" distributions. For a discussion see e.g. Kendall and Stuardt 1977.

and the Dirac delta is then given by

$$\delta^D(\vec{k}) = BA\int d^3r \exp[\pm i\vec{k}\cdot\vec{r}] \tag{15}$$

Here I chose the convention $A = 1$, $B = 1/(2\pi)^3$, but always beware of the FT conventions.

The two point **correlation function** (or correlation function) is defined as:

$$\xi(x) = \langle\delta(\vec{r})\delta(\vec{r}+\vec{x})\rangle = \int <\delta_{\vec{k}}\delta_{\vec{k}'}> \exp[i\vec{k}\cdot\vec{r}]\exp[i\vec{k}\cdot(\vec{r}+\vec{x})]d^3kd^3k' \tag{16}$$

because of isotropy $\xi(|x|)$ (only a function of the distance not orientation). Note that in some cases when isotropy is broken one may want to keep the orientation information (see e.g. redshift space distortions).

The definition of the power spectrum $P(k)$ follows :

$$<\delta_{\vec{k}}\delta_{\vec{k}'}>= (2\pi)^3P(k)\delta^D(\vec{k}+\vec{k}') \tag{17}$$

again for isotropy $P(k)$ depends only on the modulus of the k-vector, although in special cases where isotropy is broken one may want to keep the direction information.

Since $\delta(\vec{r})$ is real. we have that $\delta^*_{\vec{k}} = \delta_{-\vec{k}}$, so

$$<\delta_{\vec{k}}\delta^*_{\vec{k}'}>= (2\pi)^3\int d^3x\xi(x)\exp[-i\vec{k}\cdot\vec{x}]\delta^d(\vec{k}-\vec{k}') \tag{18}$$

this the power spectrum and the correlation function are fourier transform pairs:

$$\xi(x) = \frac{1}{(2\pi)^3}P(k)\exp[i\vec{k}\cdot\vec{r}]d^3k \tag{19}$$

$$P(k) = \int\xi(x)\exp[-i\vec{k}\cdot\vec{x}]d^3x \tag{20}$$

At this stage the same amount of information is enclosed in $P(k)$ as in $\xi(x)$.

From here the variance is

$$\sigma^2 =<\delta^2(x)>= \xi(0) = \frac{1}{(2\pi)^3}\int P(k)d^3k \tag{21}$$

or better

$$\sigma^2\int\Delta^2(k)d\ln k \text{ where } \Delta^2(k) = \frac{1}{(2\pi)^3}k^3P(k) \tag{22}$$

and the quantity $\Delta^2(k)$ is independent form the FT convention used.

Now the question is: on what scale is this variance defined?

Answer: in practice one needs to use filters: the density filed is convolved with a filter (smoothing) function. There are two typical choices:

$$f = \frac{1}{(2\pi)^{3/2}R_G^3}\exp[-1/2x^2/R_G^2] \text{ Gaussian} \rightarrow f_k = \exp[-k^2R_G^2/2] \tag{23}$$

$$f = \frac{1}{(4\pi)R_T^3}\Theta(x/R_T) \;\; \text{TopHat} \rightarrow f_k = \frac{3}{y^3}[\sin(kR_T) - kR_T\cos(kR_T)] \tag{24}$$

roughly $R_t \simeq \sqrt{5}R_G$.

Remember: **Convolution in real space is a multiplication in Fourier space**; **Multiplication in real space is a convolution in Fourier space**.

Exercise: consider a multi-variate gaussian distribution:

$$P(\delta_1..\delta_n) = \frac{1}{(2\pi)^{n/2}} det\mathbf{C}^{1/2} \exp[-\frac{1}{2}\delta^T\mathbf{C}^{-1}\delta] \tag{25}$$

where $C_{ij} =< \delta_i\delta_j >$ is the covariance. Show that is δ_i are Fourier modes then C_{ij} is diagonal. This is an ideal case, of course but this is telling us that for Gaussian fields the different k modes are independent! which is always a nice feature.

Another question for you: f you start off with a gaussian distribution (say from Inflation) and then leave this Gaussian field δ to evolve under gravity, will it remain gaussian forever? Hint: think about present-time Universe, and think about the dark matter density at,say, the center of a big galaxy and in a large void.

4.1 The importance of the Power spectrum

Theory (see lectures on inflation) give us a prediction for the primordial power spectrum:

$$P(k) = A\left(\frac{k}{k_0}\right)^n \tag{26}$$

n - the **spectral index** is often taken to be a constant and the power spectrum is a power law power spectrum. However there are theoretical motivations to generalize this to

$$P(k) = A\left(\frac{k}{k_0}\right)^{n(k_0)+\frac{1}{2}\frac{dn}{d\ln k}\ln(k/k_0)} \tag{27}$$

as a sort of Taylor expansion of $n(k)$ around the pivot point k_0. $dn/d\ln k$ is called the **running of the spectral index**.

[figure]

Note that different authors often use different choices of k_0 (sometimes the same author in the same paper uses different choices...) so things may get confused.... so let's report explicitly the conversions:

$$A(k_1) = A(k_0)\left(\frac{k}{k_0}\right)^{n(k_0)+1/2(dn/d\ln k)\ln(k_1/k_0)} \tag{28}$$

Prove the equations above.

Show that given the above definition of the running of the spectral index, $n(k) = n(k_0) + dn/d\ln k \ln(k/k_0)$.

It can be shown that as long as linear theory applies $\delta << 1$ different Fourier modes evolve independently and the Gaussian field remains Gaussian. In addition, $P(k)$ changes only in amplitude and not in shape except in the radiation to matter dominated era and when there are baryon-photon interactions and baryons-dark matter interactions (see CMB). In detail, this is described by linear perturbation growth and by the "transfer function".

[see handwritten notes if there is time]

5 Examples of real world issues

Say that now you go and try to measure a $P(k)$ from a realistic galaxy catalog. What are the real world effects you may find? We have mentioned before redshuft space distortions. Here we concentrate on other effects that are more general (and not so specific to large-scale structure analysis).

5.1 Discrete Fourier transform

In the real world when you go and take the FT of your survey or even of your simulation box you will be using something like a fast Fourier transform code (FFT) which is a discrete Fourier transform.

If your box has side of size L, even if $\delta(r)$ in the box is continuous, δ_k will be discrete. The k-modes sampled will be given by

$$\vec{k} = \left(\frac{2\pi}{L}\right)(i, j, k) \quad \text{where} \quad \Delta_k = \frac{2\pi}{L} \tag{29}$$

The discrete Fourier transform is obtained by placing the $\delta(x)$ on a lattice of N^3 grid points with spacing L/N. Then:

$$\delta_k^{DFT} = \frac{1}{N^3}\sum_r \exp[-i\vec{k}\cdot\vec{r}]\delta(\vec{r}) \tag{30}$$

$$\delta^{DFT}(\vec{r}) = \sum_k \exp[i\vec{k}\cdot\vec{r}]\delta_k^{DFT} \tag{31}$$

Beware of the mapping between r and k, some routines use a weird wrapping!

There are different ways of placing galaxies(or particle in your simulation) on a grid: Nearest grid point, Clou din cell, trangular shaped cloud etc... For each of these *remember*(!!) then deconvolve the resulting $P(k)$ for their effect. Note that

$$\delta_k \sim \left(\frac{\Delta x}{2\pi}\right)^3 N^3 \delta_k^{DFT} \simeq \frac{1}{\Delta k^3}\delta_k^{DFT} \tag{32}$$

and thus

$$P(k) \simeq \frac{<|\delta^{DFT}|^2>}{(\Delta k)^3} \quad \text{since} \quad \delta^D(k) \simeq \frac{\delta^K}{(\Delta k)^3} \tag{33}$$

The discretization introduces several effects:
The **Nyquist frequency**: $k_{Ny} = \frac{2\pi}{L}\frac{N}{2}$ is that of a mode which is sampled by 2 grid points. Higher frequencies cannot be properly sampled and give you aliasing (spurious transfer of power) effects. You should always work at $k < k_{Ny}$. There is also a minimum k (largest possible scale) that you finite box can test :$k_{min} > 2\pi/L$. This is one of the reason why one needs ever larger N-body simulations...

In addition DFT assume periodic boundary conditions [figure], if you do not have periodic boundary conditions then his also introduces aliasing.

5.2 Window, selection function, masks etc

Selection function: As you look further away you start missing some galaxies. The selection function tells you the probability for a galaxy at a given distance (or z) to enter the survey. It is a multiplicative effect along the line of sight in real space.

Window or mask You can never observe a perfect (or even better infinite) box (unless you throw away a lot of observations..) and in CMB studies you can never have a perfect full sky map (we live in a galaxy...). The mask (sky cut in CMB jargon) is a function that usually takes values of 0 or 1and is defined on the plane of the sky (i.e. it is constant along the same line of sight). The mask is also a real space multiplication effect.

Let's recall that a multiplicaton in real space (where $W(\vec{x})$ denotes the effects of window and selection functions)

$$\delta^{true}(\vec{x}) \longrightarrow \delta^{obs}(\vec{x}) = \delta^{true}(\vec{x})W(\vec{x}) \tag{34}$$

is a convolution in Fourier space:

$$\delta^{true}(\vec{k}) \longrightarrow \delta^{obs}(\vec{k}) = \delta^{true}(\vec{k}) * W(\vec{k}) \tag{35}$$

the sharper $W(\vec{r})$ is the messier and delocalized $W(\vec{k})$ is. As a result it will couple different k-modes even if the underlying ones were not correlated!

Discreteness While the dark matter distribution is almost a continuous one the galaxy distribution is discrete. We usually assume that the galaxy distribution is a sampling of the dark matter distribution. The discreteness effect give the galaxy distribution a Poisson contribution (also called shot noise contribution). Note that the Poisson contribution is non Gaussian: it is only in the limit of large number of objects (or of modes) that it approximates a Gaussian. Here it will suffice to say that as long as a galaxy number density is high enough (which will need to be quantified and checked for any practical application) and we have enough modes, we say that we will have a superposition of our random field (say the dark matter one characterized by its $P(k)$) plus a white noise contribution coming from the discreteness which amplitude depends on the average number density of galaxies (and should go to zero as this go to infinity), and we treat this additional contribution as if it has the same statistical properties as the underlying density field (which is an approximation). What is the shot noise effect on the correlation properties?

Our random field is now given by

$$f(\vec{x}) = n(\vec{x}) = \bar{n}[1 + \delta(\vec{x})] = \sum_i \delta^D(\vec{x} - \vec{x_i}) \tag{36}$$

where$\bar{n}$ denotes average number of galaxies: $\bar{n} =< \sum_i \delta^D(\vec{x} - \vec{x_i})$. Then, as done when introducing the Poisson distribution, we divide the volume in infinitesimal volume elements δV so that their occupation can only be 0 or 1. For each of these volumes the probability of getting a galaxy is $\delta P = \rho(\vec{x})\delta V$, the probability of getting no galaxy is $\delta P = 1 - \rho(\vec{x})\delta V$ and $< n_i >=< n_i^2 >= \bar{n}\delta V$. We then obtain a double stochastic process with one level of randomness coming from the underlying random field and one level coming from the Poisson sampling. The correlation function is obtained as:

$$\langle \sum_{ij} \delta^D(\vec{r_1} - \vec{r_i})\delta^D(\vec{r_2} - \vec{r_j})\rangle = \bar{n}^2(1 + \xi_{12}) + n\delta^D(\vec{r_1} - \vec{r_2}) \tag{37}$$

thus

$$< n_1 n_2 >= \bar{n}^2[1+ < \delta_1\delta_2 >^d] \quad \text{where} \quad < \delta_1\delta_2 >^d = \xi(x_{12}) + \frac{1}{\bar{n}}\delta^D(\vec{r_1} - \vec{r_2}) \tag{38}$$

and in Fourier space

$$< \delta_{k_1}\delta_{k_2} >^d = (2\pi)^3 \left(P(k) + \frac{1}{\bar{n}}\right) \delta^d(\vec{k_1} + \vec{k_2}) \tag{39}$$

This is not a complete surprise: the power spectrum of a superposition of two independent processes is the sum of the two power spectra....

5.3 pros and cons

Recap of pros and cons of corr functions vs power spectra.

5.4 And for CMB?

If we can observe the full sky the the CMB temperature fluctuation filed can be nicely expanded in spherical harmonics:

$$\Delta T(\hat{n}) = \sum_{\ell > 0} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{n}). \tag{40}$$

where

$$a_{\ell m} = \int d\Omega_n \Delta T(\hat{n}) Y^*_{\ell m}(\hat{n}). \tag{41}$$

and thus

$$< |a_{\ell m}|^2 >= \langle a_{\ell m} a^*_{\ell' m'} \rangle = \delta_{\ell\ell'}\delta_{mm'} C_\ell \tag{42}$$

C_ℓ is the angular power spectrum and

$$C_\ell = \frac{1}{(2\ell+1)} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2 \tag{43}$$

Now what happens in the presence of real world effects such as a sky cut? Analogously to the real space case:

$$\widetilde{a}_{\ell m} = \int d\Omega_n \Delta T(\hat{n}) W(\hat{n}) Y^*_{\ell m}(\hat{n}) \tag{44}$$

where $W(\hat{n})$ is a position dependent weight that in particular is set to 0 on the sky cut.

As any CMB observation gets pixelized this is

$$\widetilde{a}_{\ell m} = \Omega_p \sum_p \Delta T(p) W(p) Y^*_{\ell m}(p) \tag{45}$$

where p runs over the pixels and Ω_p denotes the solid angle subtended by the pixel.

The pseudo-C_ℓ' (Hivon et al 2002)s are defined as:

$$\widetilde{C}_\ell = \frac{1}{(2\ell+1)} \sum_{m=-\ell}^{\ell} |\widetilde{a}_{\ell m}|^2 \tag{46}$$

Clearly $\widetilde{C}_\ell \neq C_\ell$ but

$$\langle \widetilde{C}_\ell \rangle = \sum_\ell G_{\ell \ell'} \langle C_{\ell'} \rangle \tag{47}$$

where<> denotes the ensamble average.

We notice already two things: as expected the effect of the mask is to couple otherwhise uncorrelated modes, in large scale structure studies usually peopl sto here: convolve the theory with the various real world effects including the mask and compare that to the observed quantities. In CMB usually we go beyond this step and try to deconvolve the real world effects.

First of all note that

$$G_{\ell_1 \ell_2} = \frac{2\ell_2+1}{4\pi} \sum_{\ell_3} (2\ell_3+1) W_{\ell_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \tag{48}$$

where

$$W_\ell = \frac{1}{2\ell+1} \sum_m |W_{\ell m}|^2 \;\text{ and }\; W_{\ell m} = \int d\Omega_n W(\hat{n}) Y^*_{\ell m}(\hat{n}) \tag{49}$$

So if you are good enough to be able to invert G and you can say that $< C_\ell >$ is what you want then

$$C_\ell = \sum_{\ell'} G^{-1}_{\ell \ell'} \widetilde{C}_{\ell'} \tag{50}$$

The instrument has other effects such as **noise** and a finite **beam**.

5.5 Noise and beams

The effect of the noise is easily found: the instrumental noise is an independent random process with a Gaussian distribution superposed to the temperature field. In a_{lm} space $a_{lm} \longrightarrow a_{\ell m}^{signal} + a_{\ell m}^{noise}$. It the power spectrum this gives rise to the so-called noise bias:

$$C_\ell^{measured} = C_\ell^{signal} + C_\ell^{noise} \tag{51}$$

where $C_\ell^{noise} = 1)2\ell + 1) \sum_m |a_{\ell m}^{noise}|^2$. As the expectation value of C_ℓ^{noise} is non zero this is a **biased estimator**. Note that the noise bias disappears if one computes the so-called cross C_ℓ obtained as a cross-correlation between different, uncorrelated, detectors as $\langle a_{\ell m}^{noise,a} a_{\ell m}^{noise,b} \rangle = 0$. One is however not getting something for nothing, when one computes the covariance (or the associated error) for auto and for cross correlation C_ℓ (exercise!) the covariance is the same and includes the extra contribution of the noise. It is only that the cross-C_ℓ are *unbiased* estimators.

Every experiment sees the CMB with a finite resolution given by the experimental beam (similar concept to the Point Spread Function for optical astronomy. The observed temperature filed is smoothed on the beam scales. Smoothing is a convolution in real space:

$$T_i = \int d\Omega'_n T(\hat{n}) b(|\hat{n} - \hat{n'}) \tag{52}$$

where the beam is often well approximated by a Gaussian of a given Full Width at Half Maximum. remember that $\sigma_b = 0.425 FWHM$.

Thus in harmonic space the beam effect is a multiplication:

$$C_\ell^{measured} = C_\ell^{sky} e^{-\ell^2 \sigma_b^2} \tag{53}$$

and in the presence of instrumental noise

$$C_\ell^{measured} = C_\ell^{sky} e^{-\ell^2 \sigma_b^2} + C_\ell^{noise} \tag{54}$$

Of course one can always deconvolve for the effects of the beam to obtain $C_\ell^{measured}$ as close as possible to C_ℓ^{sky} obtaining:

$$C_\ell^{measured'} = C_\ell^{sky} + C_\ell^{noise} e^{\ell^2 \sigma_b^2} \tag{55}$$

The effective noise "blows up" at high ℓ (small scales). This is why it is important to know you beams well.

Exercise: what happens if you use cross-C_ℓ's?

Note that the signal to noise of a CMB map depends on the pixel size (by smoothing the map and making larger pixels the noise per pixel will decrease as $\sqrt{\Omega_{pix}}$, Ω_{pix} being the new pixel solid angle), on the integration time $\sigma_{pix} = s/\sqrt{t}$ where s is the detector sensitivity and t the time spent on a given pixel and on the number of detectors $\sigma_{pix} = s/\sqrt{M}$ where M is the number f detectors.

To compare maps of different beam sizes it is useful to have a noise measure that in independent of Ω_{pix}: $w = (\sigma^2_{pix}\Omega_{pix})^{-1}$.

To be more precise: $\langle a^{noise}_{\ell m} a^{noise*}_{\ell' m'}\rangle = \frac{s^2}{tM}\delta_{\ell\ell'}\delta_{mm'}$ where the time spent on a pixel is approximated by total observing time / number of pixels and number of pixels is given by are of the sky covered / pixel solid angle.

Remember that while FWHM may be given to you in all sort of units σ_b should be in sterradiants.

6 Higher orders correlations

From what we have learned so far we can conclude that the power spectrum (or the correlation function) completely characterizes the statistical properties of the density filed if it is Gaussian. But what if it is not? Indeed as the field evolves it cannot remain Gaussian.

[figure]

Higher order correlations start then to develop.
Higher order correlations are defined as: $\langle\delta_1...\delta_m\rangle$ where the deltas can be in real space giving the correlation function or in Fourier space giving power spectra.

At this stage is useful to present here the Wick's theorem (or cumulant expansion theorem). The correlation of order m can in general be written as sum of products of unreducible (*connected*) correlations of order ℓ for $\ell = 1...m$. For example for order 3 we obtain:

$$\langle\delta_1\delta_2\delta_3\rangle = \tag{56}$$
$$\langle\delta_1\rangle\langle\delta_2\rangle\langle\delta_3\rangle + \tag{57}$$
$$\langle\delta_1\rangle\langle\delta_2\delta_3\rangle + (3cyc.terms) \tag{58}$$
$$\langle\delta_1\delta_2\delta_3\rangle \tag{59}$$

and for order 6 (but for a distribution of zero mean):

$$< \delta_1...\delta_6 >_f = \tag{60}$$
$$< \delta_1\delta_2 >< \delta_3\delta_4 >< \delta_5\delta_6 > +..(15terms) \tag{61}$$
$$< \delta_1\delta_2 >< \delta_3\delta_4\delta_5\delta_6 > +..(15terms) \tag{62}$$
$$< \delta_1\delta_2\delta_3 >< \delta_4\delta_5\delta_6 > +...(10terms) \tag{63}$$
$$< \delta_1..\delta_6 > \tag{64}$$