I STATISTICS IN COSMOLOGY

I.1 *Statistical description of structure and random fields*

Structures in the Universe require a statistical description: On large scales, they look statistically identical and similar in every direction, and predictions from cosmological theories such as structure formation concern statistical properties rather than for instance individual formation scenarios for single galaxies, the principal exception being our own Milky Way. Suitable tools for a statistical description of e.g. the density field are random fields: There, one specifies a distribution of field amplitudes and possible correlations between them taken at different points. Conceptually, a Gaussian distribution such as

$$p(\delta(\boldsymbol{x})) = \frac{1}{\sqrt{2\pi\langle\delta(\boldsymbol{x})^2\rangle}}\exp\left(-\frac{1}{2}\frac{\delta(\boldsymbol{x})^2}{\langle\delta(\boldsymbol{x})^2\rangle}\right) \tag{I.469}$$

predicts values of the field δ taken at a specified point $\boldsymbol{x}$ for an ensemble of statistically equivalent universes. Over this ensemble, the variance $\langle\delta(\boldsymbol{x})^2\rangle$ is defined. Now, descriptive statistics always concerns moments or cumulants of δ over this ensemble of universe, and the same is true of symmetries like statistical isotropy or homogeneity, that is invariance of statistical quantities if $\boldsymbol{x}$ is rotated or shifted. Naturally, there is only one Hubble-volume in which we can carry out observations, such that accessing the ensemble for computing statistical quantities is impossible. But there is the concept of ergodicity, implying that one can construct estimates for ensemble averages from volume averages, provided that the random field is Gaussian and has a continuous spectrum, both concepts will be explained below. Gaussian random fields, i.e. a Gaussian distribution of field amplitudes are of particular relevance in cosmology, as at least at early times there are very good indications that all fields have Gaussian statistical properties.

The fluctuations of the cosmic density field $\delta(\boldsymbol{x})$, which are defined as the relative deviation of the density field $\rho(\boldsymbol{x})$ from the mean background density $\langle\rho\rangle = \Omega_m\rho_\mathrm{crit}$,

$$\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x}) - \langle\rho\rangle}{\langle\rho\rangle}, \tag{I.470}$$

are assumed to be Gaussian with a certain correlation length, meaning that the probability of finding the amplitudes $\delta_1 \equiv \delta(\boldsymbol{x}_1)$ and $\delta_2 \equiv \delta(\boldsymbol{x}_2)$ and positions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in a hypothetical ensemble of universes is given by a multivariate Gaussian probability density,

$$p(\delta_1), \delta_2) = \frac{1}{\sqrt{(2\pi)^2\det(\mathrm{Q})}}\exp\left[-\frac{1}{2}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}^t \mathrm{Q}^{-1}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}\right] \tag{I.471}$$

with the covariance matrix Q:

$$\mathrm{Q} = \begin{pmatrix}\langle\delta_1\delta_1\rangle & \langle\delta_1\delta_2\rangle\\ \langle\delta_2\delta_1\rangle & \langle\delta_2\delta_2\rangle\end{pmatrix} \tag{I.472}$$

The off-diagonal variance in Q is the correlation function $\xi(\boldsymbol{x}_1,\boldsymbol{x}_2) \equiv \langle\delta_1\delta_2\rangle$ of the random field, which describes how fast with increasing distance $|\boldsymbol{x}_2 - \boldsymbol{x}_1|$ the field loses its memory on the amplitude at $\boldsymbol{x}_1$. A length scale in $\xi(\boldsymbol{x}_1,\boldsymbol{x}_2)$ can be interpreted as a correlation length. Due to the Cauchy-Schwarz inequality,

$$\langle\delta_1\delta_2\rangle^2 \leq \langle\delta_1^2\rangle\langle\delta_2^2\rangle \quad \rightarrow \quad r = \frac{\langle\delta_1\delta_2\rangle}{\sqrt{\langle\delta_1^2\rangle\langle\delta_2^2\rangle}} \tag{I.473}$$

the correlation function is always smaller than the geometrical mean of the variances at a single point, i.e. the covariance Q is positive definite, and the Pearson correlation coefficient $|r|$ is smaller than unity. Therefore, the Cauchy-Schwarz inequality makes sure that the distribution I.471 is normalisable, as the determinant det(Q) is ensured to be positive.

Clearly, if $\xi(\boldsymbol{x}_1, \boldsymbol{x}_2)$ vanishes the Gaussian probability density separates,

$$p(\delta_1, \delta_2) = p(\delta_1)p(\delta_2) \tag{I.474}$$

and the amplitudes are uncorrelated: The covariance Q becomes diagonal if c is zero, which has two important consequences: (i) The determinant factorises,

$$\det(Q) = \langle\delta_1^2\rangle\,\langle\delta_2^2\rangle \tag{I.475}$$

as well as (ii) the quadratic form in the exponent of the distribution I.471,

$$\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix}^t Q^{-1}\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} = \frac{\delta_1^2}{\langle\delta_1^2\rangle} + \frac{\delta_2^2}{\langle\delta_2^2\rangle}. \tag{I.476}$$

The Pearson correlation coefficient r vanishes simultaneously with the correlation function $\xi(\boldsymbol{x}_1, \boldsymbol{x}_2)$. It is sensible that the correlations in a random field decrease with increasing distance between the points where the amplitudes are measured and correlated, therefore, in the limit $r \rightarrow \infty$ we get $\xi \rightarrow 0$ as well as $r \rightarrow 0$, such that $p(\delta_1, \delta_2) = p(\delta_1)p(\delta_2)$ for sufficiently separated points.

The knowledge of the variance is sufficient because all moments of a Gaussian distributed random variable with zero mean are proportional to the variance, $\langle\delta^{2n}\rangle \propto \langle\delta^2\rangle^n$. Hence the characteristic function $\varphi(t) = \int \mathrm{d}\delta p(\delta)\exp(\mathrm{i}t\delta) = \sum_n \langle\delta^n\rangle(\mathrm{i}t)/n!$ only requires the estimation of the variance $\langle\delta^2\rangle$ for reconstructing $p(\delta)\mathrm{d}\delta$ from the moments $\langle\delta^{2n}\rangle$ by inverse Fourier transform.

If the correlation function $\xi(\boldsymbol{r}) = \langle\delta_1\delta_2\rangle$ only depends on the separation vector $\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, the density field has homogeneous fluctuation properties: Pictorially, this is a case where the fluctuations are similar (and in fact, statistically equivalent) at every point in space. In this case it is convenient to transform to Fourier space,

$$\delta(\boldsymbol{k}) = \int \mathrm{d}^3x\, \delta(\boldsymbol{x})\exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{x}) \quad \leftrightarrow \quad \delta(\boldsymbol{x}) = \int \frac{\mathrm{d}^3k}{(2\pi)^3}\,\delta(\boldsymbol{k})\exp(+\mathrm{i}\boldsymbol{k}\boldsymbol{x}), \tag{I.477}$$

and to consider the variance between two Fourier modes $\delta(\boldsymbol{k}_1)$ and $\delta(\boldsymbol{k}_2)$

$$\langle\delta(\boldsymbol{k}_1)\delta^*(\boldsymbol{k}_2)\rangle = (2\pi)^3\delta_D(\boldsymbol{k}_1 - \boldsymbol{k}_2)P(\boldsymbol{k}_1) \quad \text{with} \quad P(\boldsymbol{k}) = \int \mathrm{d}^3r\,\xi(\boldsymbol{r})\exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{r}). \tag{I.478}$$

Therefore, Fourier modes of **homogeneous** random fields are mutually independent and their variance in Fourier-space defines the power spectrum $P(\boldsymbol{k})$ as the Fourier transform of the correlation function $\langle\delta_1\delta_2\rangle$. If, in addition, the random field is **isotropic**, $P(k)$ only depends only the wave number k instead of the wave vector $\boldsymbol{k}$. Pictorially, this would be a random field whose fluctuation properties are identical in every direction. An intuitive counter-example wound be waves of the surface of the ocean close to a beach, where the wave fronts are roughly parallel to the seafront and isotropy, which one would expect from the open ocean, is broken.

In this case, the angular integrations in eqn. I.478 can be carried out by introducing spherical coordinates in Fourier-space, yielding:

$$P(k) = 2\pi \int_0^\infty r^2 \mathrm{d}r\, \xi(r) j_0(kr), \tag{I.479}$$

with the spherical Bessel function of the first kind $j_0(kr)$ of order $\ell = 0$, being equal to

$$j_0(kr) = \mathrm{sinc}(kr) = \frac{\sin(kr)}{kr} \tag{I.480}$$

Cosmological inflation provides a mechanism for generating Gaussian fluctuation fields with the spectrum $P(k)$,

$$P(k) \propto k^{n_s} T^2(k) \tag{I.481}$$

with the CDM transfer function $T(k)$. $T(k)$ describes the scale-dependent suppression of the growth of small-scale modes between horizon-entry and matter-radiation equality by the Meszaros-mechanism. It is well approximated with the polynomial fit of the type

$$T(q) = \frac{\ln(1 + 2.34q)}{2.34q} \left(1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4\right)^{-\frac{1}{4}}, \tag{I.482}$$

The asymptotic behaviour of the transfer function is such that $T(k) \propto$ const for $k \ll 1$ and $T(k) \propto k^{-2}$ at $k \gg 1$, such that $P(k) \propto k^{n_s}$ on large scales and $P(k) \propto k^{n_s-4}$ on small scales. The wave vector is rescaled with the shape parameter $\Gamma \simeq \Omega_m h$, which corresponds to the horizon size at the time of matter-radiation equality $a_{\gamma m}$, and describes the peak shape of the CDM power spectrum $P(k)$. There are weak corrections due to a nonzero baryon density Ω_b

$$\Gamma = \Omega_m h\, \exp\left[-\Omega_b \left(1 + \frac{\sqrt{2h}}{\Omega_m}\right)\right], \tag{I.483}$$

where Γ is measured in units of $(\mathrm{Mpc}/h)^{-1}$, such that $q = k/\Gamma$ is a dimensionless wave vector. The spectrum is usually normalised to the variance of the linearly evolved density field at zero redshift on a scale of $R = 8$ Mpc/h,

$$\sigma_R^2 = \frac{1}{2\pi^2} \int_0^\infty \mathrm{d}k\, k^2 P(k) W^2(kR), \tag{I.484}$$

with a Fourier transformed spherical top hat filter function,

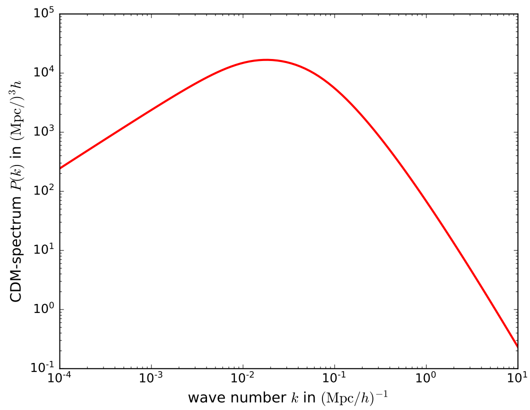

Figure 13: CDM-spectrum $P(k)$ *today for linear growth*

$$W(x) = \frac{3j_1(x)}{x} \tag{I.485}$$

$j_1(x)$ is the spherical Bessel function of the first kind of order $\ell = 1$.

This particular definition of σ_R^2, along with the fact that the power spectrum has the dimension of a volume, motivates the definition of the dimensionless power spectrum $\Delta^2(k) \propto k^3 P(k)$,

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P(k) \quad \rightarrow \quad \sigma_R^2 = \int_0^\infty \mathrm{d}\ln k\, \Delta^2(k) W^2(kR), \tag{I.486}$$

such that $\Delta^2(k)$ reflects the fluctuation variance per logarithmic band in k, $\mathrm{d}\sigma_R^2/\mathrm{d}\ln k \propto \Delta^2$. It is common to normalise $P(k)$ by the variance σ_8^2 on scales of comoving $R = 8\,\mathrm{Mpc}/h$, and typical values are $\sigma_8 = 0.8 \ldots 0.9$. The spectrum $P(k)$ is shown in Fig. 13 for linear evolution at the current cosmic epoch.

I.2 *Fluctuations on the sky and Limber-projections*

A Gaussian random field $\gamma(\theta)$ on the celestial sphere can be characterised by the correlation function

$$C_{\gamma\gamma}(\alpha) = \langle \gamma(\theta)\gamma^*(\theta') \rangle \tag{I.487}$$

with the separation $\alpha = \sphericalangle(\theta, \theta')$, because a Gaussian distribution is determined by the variance, following the argument using the characteristic function of a distribution outlined in Sect. **??**. The averaging brackets $\langle \ldots \rangle$ denote a hypothetical ensemble average over realisations of the random field, but can be replaced by spherical averages for estimating the correlation function because of the ergodicity of the ensemble provided that the random process has a continuous correlation function up to cosmic variance. The correlation function and the angular power spectrum can be converted into each other under transformations with Legendre polynomials $P_\ell(\cos\alpha)$,

$$C_{\gamma\gamma}(\ell) = 2\pi \int \mathrm{d}\cos\alpha\, C_{\gamma\gamma}(\alpha) P_\ell(\cos\alpha) \quad \leftrightarrow \quad C_{\gamma\gamma}(\alpha) = \frac{1}{4\pi}\sum_{\ell=0}^{\infty}(2\ell+1)C_{\gamma\gamma}(\ell)P_\ell(\cos\alpha) \tag{I.488}$$

using the orthonormality of the Legendre-polynomials $P_\ell(\cos\alpha)$:

$$\int_{-1}^{+1} \mathrm{d}x\, P_\ell(x)P_{\ell'}(x) = \frac{2}{2\ell+1}\delta_{\ell\ell'}. \tag{I.489}$$

Fluctuations of a quantity like the sky temperature $\tau(\theta)$ or the galaxy density $\gamma(\theta)$ on the celestial sphere with homogeneous fluctuation properties can be decomposed in using the spherical harmonics $Y_{\ell m}(\theta)$, because they are a complete orthonormal set of basis functions:

$$\gamma(\theta) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{+\ell} \gamma_{\ell m} Y_{\ell m}(\theta) \quad \leftrightarrow \quad \gamma_{\ell m} = \int_{4\pi} \mathrm{d}\Omega\, \gamma(\theta) Y^*_{\ell m}(\theta). \tag{I.490}$$

The orthonormality relation of the spherical harmonics $Y_{\ell m}(\theta)$ reads

$$\int_{4\pi} \mathrm{d}\Omega\, Y_{\ell m}(\theta) Y^*_{\ell' m'}(\theta) = \delta_{\ell\ell'}\delta_{mm'}, \tag{I.491}$$

and is not identical to the completeness relation:

$$\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{+\ell} Y_{\ell m}(\theta) Y^*_{\ell m}(\theta') = \delta(\theta - \theta'), \tag{I.492}$$

because the spherical harmonics $Y_{\ell m}(\theta)$ are a discrete basis system. Orthonormality and completeness are identical in the case of a continuous basis system like the plane waves $\exp(\pm ikx)$ of the Fourier transform. The variance of the spherical harmonics expansion coefficients $\gamma_{\ell m}$ can be related to the angular power spectrum,

$$\langle \gamma_{\ell m}\gamma^*_{\ell' m'}\rangle = \int_{4\pi} \mathrm{d}\Omega \int_{4\pi} \mathrm{d}\Omega'\, C_{\gamma\gamma}(\alpha) Y_{\ell m}(\theta) Y^*_{\ell' m'}(\theta'), \tag{I.493}$$

by substituting the decomposition eqn. (I.490) and using the definition of the correlation function eqn. (I.487). The correlation function $C_{\gamma\gamma}(\alpha)$ can be replaced with the angular spectrum $C_{\gamma\gamma}(\ell)$, and the Legendre polynomial can be substituted with the spherical harmonic's addition theorem, $\alpha = \sphericalangle(\hat{\theta}, \hat{\theta}')$:

$$\sum_{m=-\ell}^{+\ell} Y_{\ell m}(\theta) Y^*_{\ell m}(\theta') = \frac{2\ell+1}{4\pi} P_\ell(\cos\alpha). \tag{I.494}$$

Using the orthonormality relation twice and contracting the Kronecker δ-symbols yields the final result

$$\langle \gamma_{\ell m}\gamma^*_{\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}\, C_{\gamma\gamma}(\ell), \tag{I.495}$$

i.e. that the variance of the expansion coefficients $\gamma_{\ell m}$ is equal to the angular spectrum $C_{\gamma\gamma}(\ell)$ and that there is no cross-correlation between coefficients on different angular scale ℓ or different propagation direction m in the case of homogeneous and isotropic fields.

The Limber equation is used for relating the fluctuation statistics of the 3d source field to the fluctuation statistics of the projected observable. Both observables, the iSW-temperature perturbation $\tau(\theta)$ and the tracer density $\gamma(\theta)$ are derived as line of sight projections from the source fields $\varphi(\chi\theta, \chi)$ and $\delta(\chi\theta, \chi)$ with weighting functions $W_\tau(\chi)$ and $W_\gamma(\chi)$:

$$\gamma(\theta) = \int_0^{\chi_H} d\chi\, W_\gamma(\chi)\delta(\chi\theta, \chi) \quad \text{and} \quad \tau(\theta) = \int_0^{\chi_H} d\chi\, W_\tau(\chi)\varphi(\chi\theta, \chi) \tag{I.496}$$

The angular correlation function $C_{\gamma\gamma}(\alpha)$ can be then related to the correlation of the source field $\delta(\theta\chi, \chi)$:

$$C_{\gamma\gamma}(\alpha) = \int_0^{\chi_H} d\chi\, W_\gamma(\chi) \int_0^{\chi_H} d\chi'\, W_\gamma(\chi') \int dk\, k^2 P(k, \chi, \chi') \int_{4\pi} d\Omega_k \exp(i\boldsymbol{k}(\boldsymbol{x} - \boldsymbol{x}')), \tag{I.497}$$

with the spatial comoving coordinates $\boldsymbol{x} = (\theta\chi, \chi)$ and the solid angle element $d\Omega_k$ in Fourier space. The power spectrum $P(k, \chi, \chi')$ follows from the Fourier transform of the correlation function of the source field,

$$\langle\gamma(\theta\chi, \chi)\gamma^*(\theta'\chi', \chi')\rangle = \int \frac{d^3k}{(2\pi)^3} P(k) \exp(i\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{x}')) = \int dk k^2 P(k) \int_{4\pi} d\Omega_k \exp(i\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{x}')) \tag{I.498}$$

In order to solve the angular integration, one can take advantage of the Rayleigh expansion of a plane wave in terms of spherical waves:

$$\exp(i\boldsymbol{k}\boldsymbol{x}) = 4\pi \sum_{\ell=0}^{\infty} i^\ell j_\ell(kx) \sum_{m=-\ell}^{+\ell} Y_{\ell m}(\hat{k}) Y^*_{\ell m}(\theta). \tag{I.499}$$

The angular integration can be carried out while substituting the orthonormality relation of the spherical harmonics,

$$\int_{4\pi} d\Omega_k \exp(i\boldsymbol{k}(\boldsymbol{x} - \boldsymbol{x}')) = (4\pi)^2 \sum_{\ell=0}^{\infty} j_\ell(k\chi) j_\ell(k\chi') \sum_{m=-\ell}^{+\ell} Y_{\ell m}(\theta) Y^*_{\ell m}(\theta') = 4\pi \sum_{\ell=0}^{\infty} j_\ell(k\chi) j_\ell(k\chi')\, (2\ell + 1)\, P_\ell(\cos\alpha) \tag{I.500}$$

where in the last step the addition theorem has been used, yielding

$$C_{\gamma\gamma}(\alpha) = 4\pi \int_0^{\chi_H} d\chi W_\gamma(\chi) \int_0^{\chi_H} d\chi' W_\gamma(\chi') \int dk k^2 P(k,\chi,\chi') \sum_{\ell=0}^{\infty} j_\ell(k\chi) j_\ell(k\chi')(2\ell+1) P_\ell(\cos\alpha) \tag{I.501}$$

Inverting the expression for the angular correlation function $C_{\gamma\gamma}(\ell)$ by multiplying both sides with $P_{\ell'}(\cos\alpha)$ and integrating over $d(\cos\alpha)$ results in

$$C_{\gamma\gamma}(\ell) = (4\pi)^2 \int_0^{\chi_H} d\chi\, W_\gamma(\chi) \int_0^{\chi_H} d\chi'\, W_\gamma(\chi') \int dk\, k^2 P(k,\chi,\chi') j_\ell(k\chi) j_\ell(k\chi') \tag{I.502}$$

by using the orthonormality relation of the Legendre-polynomials. The expression can be further simplified if $P(k,\chi,\chi')$ is slowly varying in comparison to the spherical Bessel functions, i.e. if the angles involved are small, which corresponds to approximating the sky as being flat. In this case $P(k,\chi,\chi')$ can be moved in front of the dk-integration, which can then be carried out using the orthogonality relation of the spherical Bessel functions,

$$\int_0^{\infty} k^2 dk\, j_\ell(k\chi) j_\ell(k\chi') = \frac{\pi}{2\chi^2} \delta_D(\chi - \chi'). \tag{I.503}$$

We can approximately set $P(k) \simeq P(\ell/\chi)$, giving the final result

$$C_{\gamma\gamma}(\ell) \simeq \int_0^{\chi_H} \frac{d\chi}{\chi^2} W_\gamma^2(\chi) P(k = \ell/\chi, \chi). \tag{I.504}$$

A slightly better approximation to the correct result in spherical coordinates can be obtained by replacing $k = \ell/\chi$ with $k = (\ell + 1/2)/\chi$. The small-angle approximation of the Limber eqn. (I.504) generally overestimates the angular power spectrum in comparison to the correct solution in eqn. (I.502).

I.3 *Cosmic microwave background anisotropies*

The spectrum $C_{TT}(\ell)$ of the CMB-fluctuations is given in Fig. 14.

Fig. 15 shows the size of the CMB photosphere from the moment of decoupling until today, for three different ΛCDM cosmologies.

I.4 *Secondary anisotropies in the cosmic microwave background*

I.4.1 Gravitational lensing of the CMB

There is a very interesting gravitational lensing effect in the cosmic microwave background: A typical lensing deflection that photons from the CMB experience amounts to a few arcminutes, which is small compared to the typical scales on which the temperatures in the cosmic microwave backgrounds vary, which is roughly on the degree-scale. Therefore, one expects a small distortion of the CMB-fluctuation pattern, as hot and cold patches are deformed by roughy a percent. There is no energy input into the CMB by the gravitational lensing effect, if one assumes the gravitational potentials to be static (which is a good approximation but which is ultimately flawed

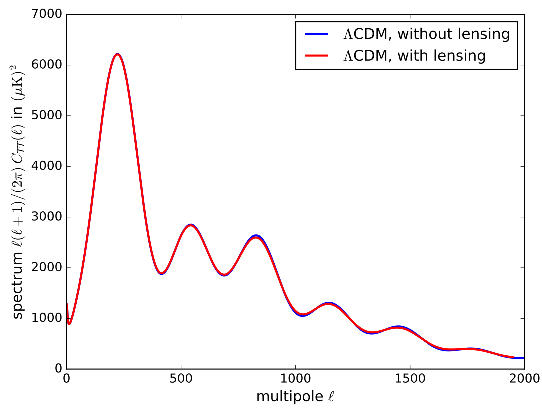

Figure 14: Angular spectrum $C_{TT}(\ell)$ of the temperature anisotropies of the CMB

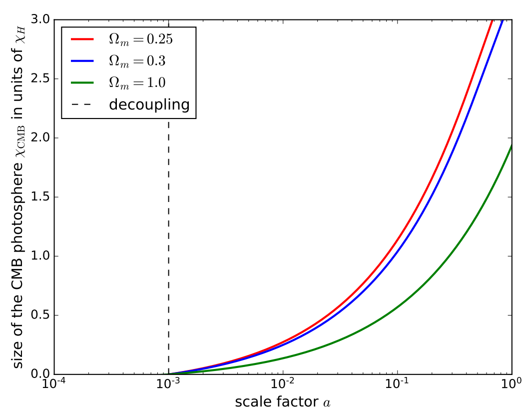

Figure 15: Comoving radius of the CMB photosphere as a function of scale factor, for three ΛCDM cosmologies

because of the integrated Sachs-Wolfe effect). Lensing, being a gravitational effect, can not differentiate between photons of different energy and is completely achromatic, as well as perfectly conserving photon density, energy flux and spectral distribution. Therefore, we expect that lensing conserves the Planck-distribution of the photons of the CMB as a thermal source. Because the lensing effect only redistributes photons, it could not generate structures in a completely isotropic CMB, contrarily, it needs structures to work on.

In gravitational lensing in astronomy it is rarely the case that one can access the unlensed situation, where the Solar eclipse of 1919 is a very notable exception. In almost all other cases of gravitational lensing, one has to make an assumption about the unlensed case in order to detect the gravitational lensing effect. Obviously, one would like to make assumptions that are as weak as possible and generic from a physical point of view. Gravitational lensing changes the statistical properties of the cosmic microwave background and breaks statistical homogeneity as a symmetry. With the assumption of a statistically homogeneous unlensed CMB one can quantify the magnitude of the broken statistical symmetry and therefore measure the weak lensing effect. It is practical to define a dimensionless amplitude $T(\boldsymbol{\theta})$ of the temperature fluctuations $T(\boldsymbol{\theta})$ in the cosmic microwave background relative to the mean temperature $T_{CMB} = 2.725$ Kelvin,

$$T(\boldsymbol{\theta}) = \frac{T(\boldsymbol{\theta}) - T_{CMB}}{T_{CMB}}, \tag{I.505}$$

Statistical homogeneity has a very interesting implication for the Fourier-modes $T(\boldsymbol{\ell})$ of the temperature field $T(\boldsymbol{\theta})$. Defining

$$T(\boldsymbol{\ell}) = \int d^2\theta\, T(\boldsymbol{\theta}) \exp(-i\boldsymbol{\theta}\boldsymbol{\ell}) \quad \leftrightarrow \quad T(\boldsymbol{\theta}) = \int \frac{d^2\ell}{(2\pi)^2}\, T(\boldsymbol{\ell}) \exp(+i\boldsymbol{\theta}\boldsymbol{\ell}), \tag{I.506}$$

one can ask the question how the Fourier-modes are correlated, if there is a nonzero correlation in configuration space. In fact,

$$\langle T(\boldsymbol{\ell}) T(\boldsymbol{\ell}')^* \rangle = (2\pi)^2 T_D(\boldsymbol{\ell} - \boldsymbol{\ell}') C_{TT}(\ell), \tag{I.507}$$

with the Dirac-function T_D,

$$T_D(\ell) = \int d^2\theta \, \exp(+i\boldsymbol{\theta}\boldsymbol{\ell}). \tag{I.508}$$

$C_{TT}(\ell)$ is the spectrum of the temperature fluctuations and is given by

$$C_{TT}(\ell) = \int d^2\theta\, \xi(\theta) \exp(-i\boldsymbol{\theta}\boldsymbol{\ell}), \tag{I.509}$$

where in the case of statistically isotropic fields the integration can be simplified according to $\boldsymbol{\theta}\boldsymbol{\ell} = \theta\ell\cos\varphi_\ell$ and $d^2\theta = \theta d\theta d\varphi_\ell$ by introducing polar coordinates. Therefore, one observes uncorrelated Fourier-modes in the case of statistically homogeneous fields. As we will see, the signature of CMB-lensing is to change the fluctuation statistics of the cosmic microwave background. In particular, it will make a statistically homogeneous CMB statistically inhomogeneous and will generate correlations $\langle T(\boldsymbol{\ell}) T(\boldsymbol{\ell}')^* \rangle \neq 0$ even if $\boldsymbol{\ell} \neq \boldsymbol{\ell}'$. Under the assumption that the unlensed CMB has been statistically homogeneous, which is a weak assumption that is supported by models of cosmic inflation, any measurement of these correlations would be an indication for the gravitational lensing effect.

As the photons of the CMB propagate to us, they have to transverse the cosmic large-scale structure and experience gravitational lensing. As they are deflected by gravitational potentials, they seem to change their propagation direction: Instead of measuring the temperature field $T(\theta)$ in the direction θ, the arrival direction is changed to $\theta + \boldsymbol{\alpha}$, where the deflection angle $\boldsymbol{\alpha}$ is the gradient of the lensing potential $\psi(\theta)$:

$$T(\theta) \rightarrow \hat{T}(\theta) = T(\theta + \boldsymbol{\alpha}) \tag{I.510}$$

These deflections distort the pattern of hot and cold patches in the CMB, which can be quantified in a statistical way. To this purpose, assuming that the deflections are small compared to the angular size of structures in the microwave background, one can expand the temperature field in a Taylor-series,

$$\hat{T}(\theta) = T(\theta + \boldsymbol{\alpha}) = T(\theta) + \sum_i \alpha_i \partial^i T + \frac{1}{2}\sum_{ij} \alpha_i \alpha_j \partial^i \partial^j T + \dots. \tag{I.511}$$

Computing a correlation function of the lensed temperature field yields

$$\langle \hat{T}(\theta)\hat{T}(\theta')\rangle = \langle T(\theta + \boldsymbol{\alpha})T(\theta' + \boldsymbol{\alpha}')\rangle \tag{I.512}$$

and consequently

$$= \langle T(\theta)T(\theta')\rangle + \sum_i \sum_k \langle \alpha_i(\theta)\alpha_k(\theta')\rangle \times \langle \partial_i T(\theta)\partial'_k T(\theta')\rangle + 2\sum_{ij} \langle \alpha_i(\theta)\alpha_j(\theta)\rangle \times \langle \partial^2_{ij} T(\theta)T(\theta')\rangle + \dots \tag{I.513}$$

if one assumes that the deflection field is uncorrelated with the temperature field, and that the distribution of the lensing deflection angle components are symmetric with zero mean. Both assumptions are physically sensible, because the deflecting large-scale structure responsible for the gravitational lensing effect is separated by a very large distance from the CMB, and because the structures responsible for lensing do not define a preferred direction. The two terms appearing as a correction to the unlensed temperature fluctuations can be interpreted as a correlated deflection $\langle \alpha_i(\theta)\alpha_k(\theta')\rangle$ where the lensing deflection $\alpha_i(\theta)$ at θ is not independent from the deflection $\alpha_k(\theta')$ at θ', and as an effect caused by a second-order deflection $\langle \alpha_i(\theta)\alpha_j(\theta)\rangle$ at a single point. Especially the last effect can be visualised by imagining that CMB photons reaching us are deflected by some amount into a random direction, leading to a blurring of the CMB. In fact, the lensed CMB has less structure compared to the unlensed one, as the blurring causes the contrast of structures to decrease.

Transforming the Taylor-series to Fourier-space then yields

$$\hat{T}(\boldsymbol{\ell}) = T(\boldsymbol{\ell}) + \mathrm{i}\int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2} \sum_i \alpha_i(\boldsymbol{\ell}_1)(\boldsymbol{\ell}-\boldsymbol{\ell}_1)_i T(\boldsymbol{\ell}-\boldsymbol{\ell}_1) - \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int \frac{\mathrm{d}^2\ell_2}{(2\pi)^2} \sum_{ij} \alpha_i(\boldsymbol{\ell}_1)\alpha_j(\boldsymbol{\ell}_2)(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)_i(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)_j T(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2) + \ldots, \tag{I.514}$$

applying the two properties of Fourier transforms, i.e. that products become convolutions and that every derivative ∂_i generates a prefactor $\mathrm{i}\boldsymbol{\ell}_i$. Furthermore, one can replace the deflection angle $\boldsymbol{\alpha}(\boldsymbol{\ell})$ by the derivative $-\mathrm{i}\boldsymbol{\ell}\psi(\boldsymbol{\ell})$ of the lensing potential ψ. Inspection of the last relationship shows that the lensed CMB temperature field is given by the unlensed field, with a series of corrections that involve n-fold derivatives of T, contracted with n factors of the lensing deflection field $\boldsymbol{\alpha}$.

Assembling a correlation function $\langle\hat{T}(\boldsymbol{\ell})\hat{T}(\boldsymbol{\ell}')^*\rangle$ of the lensed CMB then yields a series of correction terms to the correlation function $\langle T(\boldsymbol{\ell})T(\boldsymbol{\ell}')^*\rangle$ of the unlensed CMB. If one assumes that the structures that are responsible for gravitational lensing are separated by a large distance from the structures that cause the temperature fluctuations of the CMB, one can again factorise the mixed correlation functions

$$\langle\alpha_i(\boldsymbol{\ell}_1)\alpha_j(\boldsymbol{\ell}_2)T(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)T(\boldsymbol{\ell}')\rangle = \langle\alpha_i(\boldsymbol{\ell}_1)\alpha_j(\boldsymbol{\ell}_2)\rangle \times \langle T(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)T(\boldsymbol{\ell}')\rangle, \tag{I.515}$$

and using the assumption, that the lensing deflection field is isotropic, implying that the distributions of each of the components of α is symmetric with mean zero, sets $\langle\alpha_i\rangle = 0$. Then, one obtains for the correlations of $\hat{T}$ in Fourier space:

$$\langle\hat{T}(\boldsymbol{\ell})\hat{T}(\boldsymbol{\ell}')^*\rangle = \langle T(\boldsymbol{\ell})T(\boldsymbol{\ell}')^*\rangle + \int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int \frac{\mathrm{d}^2\ell_1'}{(2\pi)^2} \sum_i\sum_k (\boldsymbol{\ell}-\boldsymbol{\ell}_1)_i(\boldsymbol{\ell}'-\boldsymbol{\ell}_1')_k\langle\alpha_i(\boldsymbol{\ell}_1)\alpha_k(\boldsymbol{\ell}_1')\rangle \times \langle T(\boldsymbol{\ell}-\boldsymbol{\ell}_1)T(\boldsymbol{\ell}'-\boldsymbol{\ell}_1')\rangle + 2\int \frac{\mathrm{d}^2\ell_1}{(2\pi)^2}\int \frac{\mathrm{d}^2\ell_2}{(2\pi)^2} \sum_{ij} (\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)_i(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)_j\langle\alpha_i(\boldsymbol{\ell}_1)\alpha_j(\boldsymbol{\ell}_2)\rangle \times \langle T(\boldsymbol{\ell}-\boldsymbol{\ell}_1-\boldsymbol{\ell}_2)T(\boldsymbol{\ell}')\rangle. \tag{I.516}$$

The two correction terms that appear at second order have a clear physical interpretation. They involve the correlations $\langle\alpha_i(\boldsymbol{\ell}_1)\alpha_k(\boldsymbol{\ell}_1')\rangle$ and $\langle\alpha_i(\boldsymbol{\ell}_1)\alpha_j(\boldsymbol{\ell}_2)\rangle$, showing that the correlations of the temperature field in fact get changed due to correlations in the deflection field, which can both be traced back to the spectrum $C_{\alpha\alpha}(\ell)$. Both correction terms introduce correlations between $\boldsymbol{\ell}$ and $\boldsymbol{\ell}'$ as an expression of breaking of statistical homogeneity. The effect is proportional to $C_{\alpha\alpha}(\ell)$, such that the lensing effect can be measured in a quantitative way.

I.4.2 Thermal and kinetic Sunyaev-Zel'dovich effect

Of all secondary CMB-anisotropies, the thermal Sunyaev-Zel'dovich effect is the most subtle: There is a redistribution of the CMB-photons in energy in scattering processes

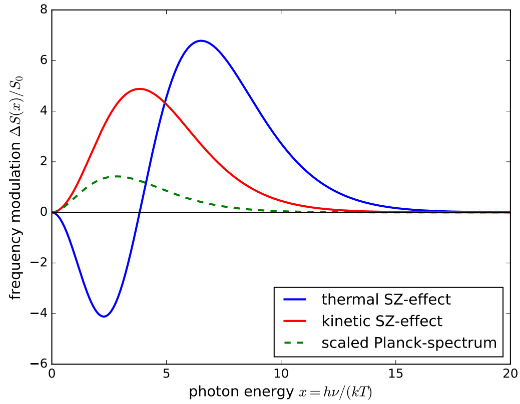

Figure 16: Spectral modulation of the CMB due to the thermal and kinetic Sunyaev-Zel'dovich effects

with electrons in galaxy clusters, as illustrated by Fig. 16. Essentially, Compton-collisions between the CMB-photons and electrons of the intra-cluster medium put the CMB as a very cold reservoir into thermal contact with the intra-cluster medium of a galaxy cluster as a very hot reservoir. Consequently, there will be a flow of thermal energy from the hot electron gas to the cold photon gas, causing a spectral distortion of the CMB: This is illustrated in Fig. 16, where the peculiar modulation of the CMB-spectrum if it is observed through a galaxy cluster is depicted. There is a secondary Sunyaev-Zel'dovich effect caused by the bulk motion of the cluster itself: In the cluster's rest frame, the CMB appears anisotropic, and likewise the radiation pressure exerted on it through Compton collisions, causing effectively the cluster to be slowed down until it comes to rest in a frame where the CMB appears isotropic. The kinetic energy of the cluster is transfered to the CMB, and therefore one perceives photons of higher energy from the direction of a cluster that is approaching the observer.

I.4.3 Integrated Sachs-Wolfe effect

The integrated Sachs-Wolfe effect is essentially a gravitational lensing effect: In the same way as spatial gradients $\partial^i\Phi$ of the gravitational potential Φ have an influence on the direction of propagation k^i of the photons, the time derivative $\partial^t\Phi$ changes the frequency (or colour) of photon. Again, working with a Newtonian perturbation on a flat, Minkowksian background

$$\mathrm{d}s^2 = (\eta_{\mu\nu} + h_{\mu\nu})\,\mathrm{d}x^\mu \mathrm{d}x^\nu = -\left(1 + \frac{2\Phi}{c^2}\right)\mathrm{d}\eta^2 + \left(1 - \frac{2\Phi}{c^2}\right)\mathrm{d}\boldsymbol{x}^2 \tag{I.517}$$

one can write for the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with $\left|h_{\mu\nu}\right| \ll 1$ in this preferred frame. Photons follow null-geodesics defined by $k_\mu k^\mu = 0$ where $k^\mu = (k^0, \boldsymbol{k})^t$ as the wave vector is tangent to $x^\mu(\lambda)$; it is parameterised by an affine parameter λ and for convenience normalised to $k^0 = 1$ and $\boldsymbol{k}^2 = 1$. Again, it is sufficient to consider a static background because of the conformal invariance of null-geodesics, which do not change under conformal transformations of the type $g_{\mu\nu} \rightarrow a^2 g_{\mu}$ of the metric $g_{\mu\nu}$. Effectively, this amounts to ignoring cosmological redshifts while focusing on the gravitational interaction.

The geodesic equation, which describes the change δk^α in k^α due to gravitational interaction now reads:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\delta k^\alpha = -\delta\Gamma^\alpha_{\ \mu\nu}\ k^\mu k^\nu \tag{I.518}$$

where the Christoffel symbol of the weakly perturbed metric transforming the time-component of k^α is given by

$$\delta\Gamma^t_{\ \mu\nu} = -\frac{1}{2}\left[\partial_\nu h_{\mu t} + \partial_\mu h_{\nu t} - \partial_t h_{\mu\nu}\right]. \tag{I.519}$$

In this approximation of $\delta\Gamma^t_{\ \mu\nu}$, the multiplication with the metric $g_{\mu\nu}$ was dropped because it would give rise to terms quadratic in the perturbation $h_{\mu\nu}$.

The first two terms give rise to the conventional Sachs-Wolfe effect, and the last term with the time derivative $\partial_t h_{\mu\nu}$ of the metric perturbation $h_{\mu\nu}$ causes the iSW-effect. Substitution into the geodesic equation yields:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\delta k^t = -\frac{1}{2}\,\partial_t h_{\mu\nu}\, k^\mu k^\nu \tag{I.520}$$

The energy shift δk^0 of a photon is given by subsequent integration,

$$\delta k^0 = \frac{1}{c^2}\int \mathrm{d}\lambda\,\left[(k^t)^2 + \boldsymbol{k}^2\right]\frac{\partial\Phi}{\partial\eta} = \frac{2}{c^2}\int \mathrm{d}\lambda\,\frac{\partial\Phi}{\partial\eta} \tag{I.521}$$

such that the energy perturbation is a measure of the integrated growth rate along the line of sight. Curiously, the iSW-effect is a direct probe of dark energy, as $\partial\Phi/\partial\eta$ vanishes in flat cosmologies with only matter, $\Omega_m = 1$.

The integral should be evaluated along the photon geodesic, but one assumes Born's approximation, such that the energy shift is obtained perturbatively while the geodesic remains characterised by the conditions $(k^0)^2 = 1$ and $\boldsymbol{k}^2 = 1$ mentioned above. In a cosmological context, the photon geodesic $\mathrm{d}s^2 = 0$ is given by $\mathrm{d}\chi = c\mathrm{d}t/a = c\mathrm{d}\eta$ with the conformal time η such that η is would be the natural choice for the affine parameter λ in the comoving frame. η and λ are linearly related such that their ratio can be absorbed in the normalisation of k. As a purely gravitational interaction, the iSW-effect conserves the spectral distribution of photons: Due to the equivalence principle, gravity treats all photons in the same way, which is true for the iSW-effect and lensing alike.

I.5 *Weak gravitational lensing by the large-scale structure*

In Sect. H.7.3 we have already derived the deflection angle $\hat{\alpha}$ in gravitational light deflection

$$\hat{\alpha}^i = -\frac{2}{c^2}\int \mathrm{d}\lambda\,\partial^i\Phi \tag{I.522}$$

into the direction of the gradients of Φ perpendicular to the line of sight. To be exact, $\hat{\alpha}$ is the change in direction between the spatial wave vector entering and leaving the gravitational potential, but not yet the change in position as observed on the sky. If a source is situated at a comoving distance χ_s and the gravitational potential acting as the light deflector is at a comoving distance χ, the change in position α^i of the source being at position θ^i without lensing is given by

$$\alpha = \theta' - \theta \tag{I.523}$$

so that the source appears at θ'. Writing $\theta = x/\chi_s$, $\theta' = (x+d)/\chi_2$ and $\hat{\alpha} = d/(\chi_s - \chi)$ suggests that the angular displacement on the sky generated by lensing is

$$\alpha = \left(1 - \frac{\chi}{\chi_s}\right)\hat{\alpha} \tag{I.524}$$

where $\hat{\alpha}$ is computable with eqn. I.522. With comoving distance as the integration variable and by rewriting the spatial as an angular derivative $\chi\partial^i = \partial_\theta$ we get

$$\alpha^i = \partial^i_\theta \psi \quad \text{with} \quad \psi = 2\int \mathrm{d}\chi\, \frac{\chi_s - \chi}{\chi_s \chi} \frac{\Phi}{c^2} \tag{I.525}$$

defining the lensing potential ψ.

In this entire discussion it would be important to realise that the change in propagation direction $\hat{\alpha}^i$ is only defined because the lens is embedded in a flat spacetime (or at least a conformally flat spacetime). Then, there is a parallel transport around the lens through essentially flat space as a reference wave vector, to which one can compare the actual wave vector that has been properly parallel transported through the gravitational potential along the physical trajectory, defining a deflection angle.

Clearly, one needs to make an assumption about the unlensed situation: Generally one does not know the positions of objects without lensing (Eddington's Solar eclipse from 1919 being a very notable exception). Instead, one could try to observe a differential deflection across the image of a distant object like galaxy: If the deflection field depends on position, there is such a differential deflection and one observes a change in shape of the image. Similarly, one could invoke Raychaudhuri's equation, as the light bundle of the galaxy forms a geodesic congruence. Then, changes in shape and size of the light bundle are related to the tidal gravitational fields, or relativistically speaking, to the curvature experienced by the light bundle. With this idea, if the observable are galaxy shapes, a weak assumption about the unlensed situation would be uncorrelated shapes, which would get coherently distorted, as light bundles from neighbouring galaxies would experience similar tidal distortions.

For changes in shape and size to emerge one needs variations of the deflection field α^i across the image of a galaxy, and as α^i is defined as the gradient of the lensing potential ψ, the changes are induces by second derivatives $\psi_{ij} = \partial_i \partial_j \psi$ of ψ. The index pair ij runs over x and y and as partial derivatives interchange, ψ_{ij} is a real symmetric 2×2 matrix. A suitable basis system are the real-valued Pauli matrices,

$$\psi_{ij} = \sum_n a_n \sigma^{(n)}_{ij} \quad \text{with} \quad a_n = \frac{1}{2}\sum_{ij} \sigma^{(n)}_{ji} \Psi_{ij} \tag{I.526}$$

The role of the three different components of $\psi_{ij} = \partial_i\partial_j\Psi$ are:

- convergence $\kappa = a_0 = \frac{1}{2}(\partial_i\partial_j\psi\delta_{ij}) = \Delta\psi/2$, which changes the angular size of a galaxy isotropically, i.e. by the same amount in the x and y-direction
- shear $\gamma_+ = a_1 = \frac{1}{2}\sigma^{(1)}_{ij}\psi_{ij} = \frac{1}{2}(\partial_x^2\psi - \partial_y^2\psi)$, which elongates the image in x-direction while compressing it in the y-direction
- shear $\gamma_\times = a_3 = \frac{1}{2}\sigma^{(3)}_{ij}\psi_{ij} = \partial_x\partial_y\psi$, which stretches an image into the $(x+y)$-direction while compressing in the $(x-y)$-direction

The two components of shear are often combined into a single complex number $\gamma = \gamma_+ + \mathrm{i}\gamma_\times$.

The convergence κ provides a mapping of the matter density:

$$\kappa = \frac{1}{2}\Delta_\theta\Psi = \Delta_\theta\int_0^{\chi_s}\mathrm{d}\chi\,\frac{\chi_s-\chi}{\chi_s\chi}\frac{\Phi}{c^2} = \int_0^{\chi_s}\mathrm{d}\chi(\chi_s-\chi)\frac{\chi}{\chi_s}\Delta_x\frac{\Phi}{c^2} \tag{I.527}$$

using $\Delta_\theta = \chi^2\Delta_x$ and the small angle approximation $x = \theta\chi$. Substituting the Poisson-equation

$$\Delta\frac{\Phi}{c^2} = \frac{3\Omega_m}{2\chi_H^2}\frac{\delta}{a} \tag{I.528}$$

yields

$$\kappa = \frac{3\Omega_m}{2\chi_H^2}\int_0^{\chi_s}\mathrm{d}\chi\,\frac{\chi_s-\chi}{\chi_s}\chi\frac{D_+}{a}\delta_0 \tag{I.529}$$

Statistically, line of sight expressions like $\kappa = \int_0^{\chi_s}\mathrm{d}\chi\,W(\chi)\delta$ can be used in Limber's equation to give the angular spectrum of the shear or convergence fields,

$$C^{\kappa\kappa}(\ell) = \int_0^{\chi_s}\frac{\mathrm{d}\chi}{\chi^2}W(\chi)^2\,P(k=\ell/\chi) \tag{I.530}$$

as a function of the spectrum $P(k)$ of the source field, in our case δ. The spectrum $C^{\gamma\gamma}(\ell)$ is identical to that of κ.

A quantification of shape could be the ellipticity ϵ measured in terms of the second moments of the brightness distribution $I(\theta)$

$$Q_{ij} = \int\mathrm{d}^2\theta\,I(\theta)\,\theta_i\theta_j \tag{I.531}$$

from which one defines the ellipticity

$$\epsilon = \frac{Q_{xx}-Q_{yy}}{Q_{xx}+Q_{yy}} + 2\mathrm{i}\frac{Q_{xy}}{Q_{xx}+Q_{yy}} \tag{I.532}$$

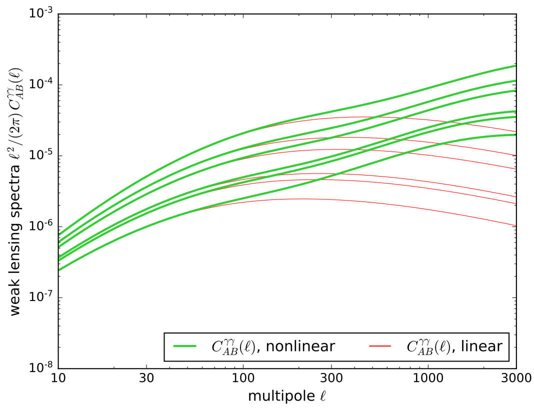

Figure 17: Tomographic spectra $C^{\gamma\gamma}(\ell)$ of the weak lensing shear

Similarly to lensing shear γ, ellipticity is a tensor with two components, and has the property to be invariant under rotations of π, as one can easily imagine by rotating an actual ellipse, and there is a practical notational advantage to combine both components into a complex ellipticity. In the weak lensing limit, the shear γ operates on ellipticity according to

$$\epsilon \rightarrow \epsilon + \gamma \tag{I.533}$$

such that an estimate of correlation functions with the observable ϵ provides an estimate of γ, if there is no intrinsic correlation between the ellipticities without lensing. The angular spectrum $C^{\gamma\gamma}(\ell)$ is shown in Fig. 17, for a so-called tomographic measurement, where the galaxies are divided up in redshift intervals.

I.6 *Bayes-inference in cosmology*

Science knows two types of truths: empirical truths correspond to reproducible, objective observations, and logical truths to statements that are derived from axioms in a mathematically consistent way. The way in which science operates is by making predictions for theories, and comparing them to observations, possibly discarding the theories in the process: As such, science is a self-correcting process guided by deduction and inference. Here, inference refers to deriving statistical statements about model parameters from data, or about the validity of entire model classes. The issue in this is that Nature provides data only with an added experimental error or by providing only finite amounts of data with a restricted statistical power: After all, the Hubble volume is finite. Therefore, an observation can not tell in an absolute sense if a theory is true, rather, it provides confidence regions and statistical error estimates, and only allows to differentiate theories that differ by significantly more than the inherent error of the measurement.

One might wonder how randomness in a measurement comes about: but after all, it is simply the result of all variables in the measurement process that can not be controlled in the experiment, because the experimental setup itself is perfectly predictable by the laws of Nature, as it is clearly part of Nature and does not exist in a transcendent way, and a better experiment will essentially allow a better control, resulting in reduced errors. Recording a set of values y_i in a measurement that are Gaussian distributed with error σ_i allows to compute the likelihood $\mathcal{L}(\{y_i\}|\theta_\mu)$, here with the simplifying assumption that all data points are statistically independent,

$$\mathcal{L}(\{y_i\}|\theta_\mu) \propto \prod_i \exp\left(-\frac{1}{2}\left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2\right) = \exp\left(-\frac{1}{2}\sum_i\left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2\right) \tag{I.534}$$

that one would make the measurement if the values result from a theoretial model $y(x)$ with model parameters θ_μ. A likelihood is, for all intents and purposes, a probability as it is a number obeying Kolmogorov's axioms. But there is an important difference in perspective: Usually, one imagines in a probability that there is a fixed random experiment that is able to produce outcomes at different probability, but in a likelihood there is a fixed outcome (the data values y_i) for which one considers now variable models $y(x)$ that differ by the value of their model parameters θ_μ. For the Gaussian error process as in eqn. I.534 it is possible to work with the χ^2-functional instead, which is linked to the likelihood by

$$\mathcal{L}(\{y_i\}|\theta_\mu) \propto \exp\left(-\frac{\chi^2(\theta_\mu)}{2}\right) \quad \text{with} \quad \chi^2 = -\sum_i\left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2 \tag{I.535}$$

Therefore, the likelihood is a function of the model parameters and depends of course on the actual data set. Now, one suspects the true model parameters in the value that maximises $\mathcal{L}$ (or minimises χ^2), as the data that one has is most easily generated by the true model: This is exactly the principle of maximum likelihood. At the same time, eqn. I.535 shows that the origin of least squares-rule originates from the Gaussian error in the data.

But there is a very important catch: The likelihood $\mathcal{L}(\{y_i\}|\theta_\mu)$ is able to quantify how probable it would have been to observe the data for a given choice of θ_μ, but what one actually would like to know is the distribution $p(\theta_\mu|\{y_i\})$ of the model parameters given the observation of the data points. For interchanging the random variable and the condition one needs to use the Bayes-theorem:

$$p(\theta_\mu|\{y_i\}) = \frac{\mathcal{L}(\{y_i\}|\theta_\mu)}{p(\{y_i\})}\,p(\theta_\mu) \tag{I.536}$$

In Bayes' reasoning, the posterior distribution $p(\theta_\mu|\{y_i\})$, i.e. the distribution of the model parameters taking the data into account is given by the likelihood $\mathcal{L}(\{y_i\}|\theta_\mu)$, for which one needs the model to predict the data and the knowledge on the error process, and the prior distribution $p(\theta_\mu)$, which reflects the uncertainty in the model parameters before one has carried out the experiment, normalised by the evidence $p(\{y_i\})$,

$$p(\{y_i\}) = \int \mathrm{d}^n\theta\,\mathcal{L}(\{y_i\}|\theta_\mu)\,p(\theta_\mu) \tag{I.537}$$

which is the probability to obtain the data in the first place given the prior information. If a new experiment is carried out, the posterior distribution from the last experiment would serve as a prior for the next experiment.

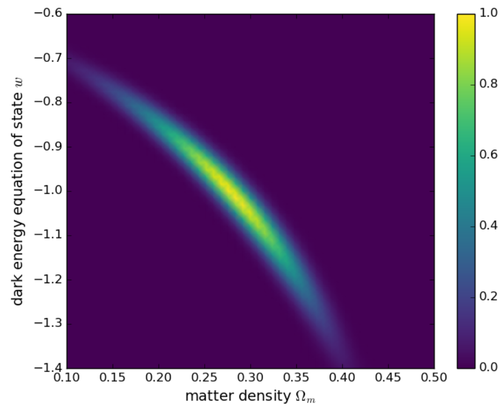

Figure 18: Direct gridded evaluation of the supernova likelihood in the parameters Ω_m and the dark energy equation of state w.

As one essentially multiplies peaked distributions in this process, the resulting distribution will be more peaked as the original ones, indicating that the uncertainty on a model parameter has been reduced by including more data.

Even though the posterior or the likelihood are computable for a given model $y(x)$ parameterised by θ_μ and for a given data set $\{y_i\}$, this is in practise numerically very challenging for highly-dimensional parameter spaces. Instead, one uses the Metropolis-Hastings algorithm (or more efficient variants of it) to generate samples θ_μ that are distributed according to the posterior distribution $p(\theta_\mu|\{y_i\})$.

In the Metropolis-Hastings algorithm one performs a random walk in parameter space on a potential given by the logarithmic likelihood (or the sum of logarithmic likelihood and logarithmic prior, those quantities exist for distributions from the exponential family). For evaluating the random walk, one takes a step from θ_μ to $\theta_\mu + \delta_\mu$, where δ_μ is a random vector from the so-called proposal distribution. Comparing $\mathcal{L}(\theta_\mu)$ with $\mathcal{L}(\theta_\mu + \delta_\mu)$ with the logarithmic likelihood ratio

$$r = \ln \frac{\mathcal{L}(\theta_\mu + \delta_\mu)}{\mathcal{L}(\theta_\mu)} \tag{I.538}$$

gives two possible options: Either $r > 0$, in which case one allows the process to jump $\theta_\mu \to \theta_\mu + \delta_\mu$, as the new point is a better explanation for the data. Or, $r < 0$, in which one accepts the step to a point with lower likelihood only in $\exp(r)$ of all cases. In this way, the samples θ_μ will follow the distribution $\mathcal{L}(\theta_\mu)$. The comparison between the gridded evaluation of a supernova likelihood in Fig. 18 with the Metropolis-Hastings evaluated result in Fig. 19 at a fraction of the computational cost is striking.

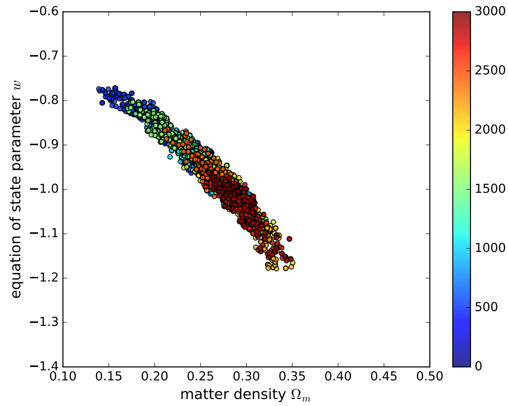

Figure 19: Samples ($n = 3 \times 10^3$) from the supernova likelihood in the parameters Ω_m and w. The colour indicates the number of the sample resulting from the Metropolis-Hastings random walk.

ACKNOWLEDGEMENTS

Ana Marta Machado de Pinho and Marco Hübner deserve my gratitude for helping with the typesetting, my colleagues Burkhard Fuchs, Matteo Maturi, Tilman Plehn and Markus Pössel for lecturing with me, and Robert Reischke, Elena Sellentin, Sarah Konrad, Hans-Martin Rieser, Maximilian Düll and Carsten Littek for accompanying the lecture as tutors. And it is a pleasure to thank Jens Jasche, Philipp M. Merkel, Francesco Pace, ln(a) Sellentin and Britta Zieser, who always understood things really, really well and always found a new perspective. If you're interested, please visit cosmology question of the week for weekly appearing questions on cosmology!

There is a large number of excellent textbooks on cosmology, and my script is not supposed to be a replacement for them. In no particular order I would like to mention:

- M.P. Hobson, G.P. Efstathiou, A.N. Lasenby: General Relativity: An Introduction for Physicists, Cambridge University Press, 2006
- A.R. Liddle, D.H. Lyth: Cosmological Inflation and the Large-Scale Structure, Cambridge University Press, 2000
- V. Mukhanov: Physical Foundations of Cosmology, Cambridge University Press, 2005
- L. Amendola, S. Tsujikawa: Dark Energy - Theory and Observations, Cambridge University Press, 2010
- E. Guyon, J.-P. Hulin, L. Petit, C.D. Mitescu: Physical Hydrodynamics, Oxford University Press, 2001
- R. Durrer: The Cosmic Microwave Background, Cambridge University Press 2008

I would like to acknowledge the lecture LaTeX-class vhbelvadi.com/latex-lecture-notes-class/ by V.H. Belvadi.