Understanding the CMB

Damian Sowinski

Your abstract.

Contents

PROBABILITY THEORY

One of the most profound results in probability theory is the central limit theorem. Put simply, the statistics of sums of random variables (rv) are approximated by a Gaussian distribution. The approximation becomes better and better the more rv's are used in the sum. This is true irrespective of the distribution that the rv's are drawn from.

Our goal will be to look at how good this approximation is. To that end we will develop the *Edgeworth expansion* to see how to expand a limiting distribution around a Gaussian. We will find that the moments of the underlying distribution play a role in the error terms. In this way we will discover a natural way to quantify *non-gaussianity* in a distribution.

Distributions

The Gaussian Distribution

No discussion of physics is ever devoid of Gaussians, so let's start there. A Gaussian distribution is the maximum entropy distribution with constraints on its mean and variance. More precisely, if we have that

$$\langle X \rangle = \mu$$
$$\langle X^2 \rangle = \sigma^2 \qquad (1)$$

then the MaxEnt distribution density reads:

$$\rho_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \qquad (2)$$

We then say X *is drawn from a normal distribution* and write $X \sim \mathcal{N}(\mu, \sigma)$. If the normal distribution is standardized, that means that the standard deviation has been rescaled to unity, $\sigma = 1$.

The χ^2-Distribution

Instead of a single random variable, what if we draw N standard normal rv's independently, $X_i \sim \mathcal{N}(0,1)$, and form a new rv by taking the sum of their squares:

$$Y = \sum_{i=1}^{N} X_i^2 \tag{3}$$

what is $\rho_Y(y)$? To find it we need to resort to the joint probability distribution of both Y and all the X_i, and use Bayes' rule:

$$\begin{aligned} \rho_Y(y) &= \left[\prod_{i=1}^{N} \int_{-\infty}^{\infty} dx_i\right] \rho(y, x_1, \ldots, x_N) \\ &= \left[\prod_{i=1}^{N} \int_{-\infty}^{\infty} dx_i\right] \rho(y|x_1, \ldots, x_N)\rho_X(x_1) \cdots \rho_X(x_N) \\ &= \frac{1}{(2\pi)^{N/2}} \left[\prod_{i=1}^{N} \int_{-\infty}^{\infty} dx_i\right] \delta(y - \sum_i x_i^2) e^{-\frac{1}{2}\sum_i x_i^2} \end{aligned} \tag{4}$$

In the second line we used the independence of the X_i to factor their joint probability distribution. By examining the integral, we see that we are integrating over $\mathbb{R}^N$, and the delta function is projecting out all but the surface of a hypersphere of radius y. Since the integrand is solely a function of the radius, the integral will evaluate to the surface area of this sphere times the value of the integrand at $r = \sqrt{y}$. Don't forget the extra $\frac{1}{2\sqrt{y}}$ factor coming from the delta function integration! The result is:

$$\rho_Y(y) = \frac{1}{2^{\frac{N}{2}} \Gamma\left(\frac{N}{2}\right)} y^{\frac{N}{2}-1} e^{-\frac{1}{2}y} \tag{5}$$

This is known as a χ^2 distribution with N degrees of freedom.

If instead of standardized random variables we had constructed the weighted sum of squares of nonstandard Gaussians:

$$Y = \frac{1}{N} \sum_{i=1}^{N} X_i^2 \tag{6}$$

$$\langle X_i \rangle = 0 \tag{7}$$

$$\langle X_i X_j \rangle = \sigma^2 \delta_{ij} \tag{8}$$

then the pdf for y will be:

$$\rho_Y(y|\sigma^2) = \frac{1}{\Gamma(\frac{N}{2})} \left(\frac{N}{2}\right)^{\frac{N}{2}} \frac{y^{\frac{N}{2}-1}}{\sigma^N} e^{\frac{N}{2}\frac{y}{\sigma^2}} \tag{9}$$

This form of the χ^2-distribution will be helpful later.

The Central Limit Theorem

A standardized rv X is drawn from a probability density function (pdf), $\rho_X(x)$. Standardized means that $\langle X \rangle = 0$ and $\langle X^2 \rangle = 1$. This can always be accomplished by rescaling the rv:

$$X \to X' = \frac{X - \langle X \rangle}{\langle (X - \langle X \rangle)^2 \rangle^{\frac{1}{2}}} \tag{10}$$

Consider drawing independently N of these rv's,$\{X_i\}$ and defining a new rv

$$Y = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \tag{11}$$

The question we wish to answer is what the pdf of this new variable, $\rho_Y(y)$, is in the limit $N \to \infty$. We know that it is related to the pdf of X by using the joint probability density and Bayes' rule:

$$\begin{aligned}
\rho_Y(y) &= \left(\prod_{i=1}^{N} \int \mathrm{d}x_i\right) \rho(y, x_1, \ldots, x_N) \\
&= \left(\prod_{i=1}^{N} \int \mathrm{d}x_i\right) \rho(x_1, \ldots, x_N)\rho(y|x_1, \ldots, x_N) \\
&= \left(\prod_{i=1}^{N} \int \mathrm{d}x_i \ \rho_X(x_i)\right) \delta(y - \frac{1}{\sqrt{N}} \sum_j x_j)
\end{aligned} \tag{12}$$

where we have used the independence of the X_i's to factor their joint pdf. The probability of a particular value of Y is the sum of the probabilities of all the different ways the X_i sum up to make that value. Rather than knowing this relationship, it would be nice to know the analytical form of ρ_Y. Spoiler Alert: it's a standardized Gaussian.

To prove this remarkable conclusion we will have to examine the characteristic functions of ρ_X and ρ_Y. The characteristic function can be thought of in two ways: as the expectation value of a plane wave, or as the Fourier transform of the pdf:

$$\begin{aligned}
\psi_X(k) &= \int \mathrm{d}x \ \rho_X(x) e^{ikx} \\
&= \langle e^{ikX} \rangle \\
&= \sqrt{2\pi}(\mathcal{F}\rho_X)(k)
\end{aligned} \tag{13}$$

Both points of view will be useful; in particular that $\rho_X = \frac{1}{\sqrt{2\pi}} \mathcal{F}^{-1} \psi_X$. What is ψ_Y?

$$\begin{aligned}
\psi_Y(k) &= \int \mathrm{d}y \ \rho_Y(y) e^{iky} \\
&= \int \mathrm{d}y \int \mathrm{d}x_1 \cdots \int \mathrm{d}x_N \rho_X(x_1) \cdots \rho_X(x_N) \delta(y - \frac{1}{\sqrt{N}}(x_1 + \cdots + x_N)) e^{iky} \\
&= \left(\int \mathrm{d}x \ \rho_X(x) e^{i\frac{k}{\sqrt{N}}x}\right)^N \\
&= \psi_X(\frac{k}{\sqrt{N}})^N
\end{aligned} \tag{14}$$

Ok, so that's really cool. Now if we put on our expectation value glasses for the characteristic function, and consider the large N behavior of the function:

$$\begin{aligned}
\psi_Y(k) &= \lim_{N\to\infty} \langle e^{ikX/\sqrt{N}} \rangle^N \\
&= \lim_{N\to\infty} \langle 1 + \frac{ikX}{\sqrt{N}} - \frac{k^2X^2}{2N} + \cdots \rangle^N \\
&= \lim_{N\to\infty} (1 - \frac{k^2}{2N} + \cdots)^N \\
&= e^{-\frac{1}{2}k^2}
\end{aligned} \tag{15}$$

Putting on our Fourier transform goggles, the result is a simple matter of recalling that the (inverse)Fourier transform of a Gaussian is too a Gaussian:

$$\rho_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \tag{16}$$

Voila! Y is drawn from a standardized Gaussian. All hints of the underlying distribution, ρ_X, have vanished.

Our goal now will be to understand what happens when N is very large, but not infinite. We will see that the moments of ρ_X will play a crucial role in understanding the deviations of ρ_Y from a standardized Gaussian.

Edgeworth Expansion

The Edgeworth expansion is conceptually simple, but algebraically almost a Sisyphean task. We simply have to collect all the terms in the characteristic function in powers of $N^{-1/2}$, without throwing out higher order terms like we did in proving the CLT. Shall we begin, Clarrise?

Let's denote the m^{th} moment of X by $\langle X^m \rangle = \kappa_m$. Everything as before, except for the slight change in notation $\epsilon = N^{-1/2}$, the characteristic function for Y reads

$$\begin{aligned}\psi_Y(k) &= (1 - \frac{k^2}{2N} + \sum_{m=3}^{\infty} \frac{(ik)^m}{m!}\kappa_m \epsilon^m)^N \\ &= (1 - \frac{k^2}{2N})^N (1 + \sum_{\ell=0}^{\infty}\sum_{m=3}^{\infty} \frac{i^m k^{2\ell+m}}{2^\ell m!}\kappa_m \epsilon^{2\ell+m})^N \end{aligned} \tag{17}$$

where the second sum came from a geometric series expansion. Yes, I know I'm mixing ϵ's and N's, and for good reason. In the $N \to \infty$ limit, the term on the outside of this expression is exactly $e^{-\frac{1}{2}k^2}$. With the awesome power of algebra, the binomial expansion, and a Taylor series, one can find the error of this equation:

$$(1 - \frac{k^2}{2N})^N = e^{-\frac{1}{2}k^2}\left(1 - \frac{k^4}{8}\epsilon^2 + (\frac{k^8}{2^8} - \frac{k^6}{3 \cdot 2^3})\epsilon^4 + \cdots\right) \tag{18}$$

We're going to use this, as soon as we expand out the other piece in powers of ϵ. First off let's define

$$\alpha_j = \sum_{\ell=0}^{\lfloor \frac{j-3}{2} \rfloor} \frac{(ik)^j \kappa_{j-2\ell}}{(-2)^\ell (j-2\ell)!} \tag{19}$$

The first few are listed below

$$\begin{aligned}\alpha_3 &= \frac{(ik)^3}{3!}\kappa_3 \\ \alpha_4 &= \frac{(ik)^4}{4!}\kappa_4 \\ \alpha_5 &= \frac{(ik)^5}{5!}\kappa_5 - \frac{(ik)^5}{2 \cdot 3!}\kappa_3\end{aligned}$$

Applying the binomial expansion once more, the characteristic function can now be written

$$\begin{aligned}\psi_Y(k) &= e^{-\frac{1}{2}k^2}\left(1 - \frac{k^4}{8}\epsilon^2 + (\frac{k^8}{2^8} - \frac{k^6}{3 \cdot 2^3})\epsilon^4 + \cdots\right)\left(1 + \sum_{j=3}^{\infty}\alpha_j \epsilon^j\right)^N \\ &= e^{-\frac{1}{2}k^2}\left(1 + \alpha_3\epsilon + (\alpha_4 + \frac{1}{2}\alpha_3^2 - \frac{1}{8}k^4)\epsilon^2 + (\frac{1}{6}\alpha_3^3 - 2\alpha_3 + \alpha_3\alpha_4 + \alpha_5 - \frac{1}{8}\alpha_3 k^4)\epsilon^3 + \cdots\right) \\ &= \left(1 + \frac{\kappa_3}{6}(ik)^3\epsilon + (\frac{\kappa_4 - 3}{24}(ik)^4 + \frac{\kappa_3^2}{72}(ik)^6)\epsilon^2 + \cdots\right)e^{-\frac{1}{2}k^2}\end{aligned} \tag{20}$$

Maybe one day I'll write out the next two terms. Not today. We're almost there; the last step is to inverse Fourier transform, but we have all those pesky ik factors sticking around. They're not so bad if you remember this nice little identity

$$\mathcal{F}^{-1}[(ik)^m \rho] = (-1)^m \frac{\partial^m}{\partial x^m}\mathcal{F}^{-1}\rho \tag{21}$$

Using it to transform back to the pdf of Y we find

$$\rho_Y(y) = (1 - \frac{\kappa_3}{6\sqrt{N}}\frac{\partial^3}{\partial y^3} + \frac{\kappa_4 - 3}{24N}\frac{\partial^4}{\partial y^4} + \frac{\kappa_3^2}{72N}\frac{\partial^6}{\partial y^6} + \cdots)\frac{1}{2\pi}e^{-\frac{1}{2}y^2} \tag{22}$$

One final simplification can be made if one recalls the definition of the Hermite polynomials, which given that physicists are all obsessed with the simple harmonic oscillator you do, the Edgeworth expansion reads

$$\rho_Y(y) = (1 + \frac{\kappa_3}{6\sqrt{N}} H_3(y) + \frac{\kappa_4 - 3}{24N} H_4(y) + \frac{\kappa_3^2}{72N} H_6(y) + \cdots) \frac{1}{2\pi} e^{-\frac{1}{2}y^2} \tag{23}$$

As promised, the skewness and kurtosis of the underlying distribution contribute to the first and second order terms of the Edgeworth expansion. We will primarily be concerned with the first order correction. The appearance of the third order moment here will have an analogy when we start examining the CMB. There we will find the three point correlation function, or in Fourier space, the bispectrum.

THE COSMIC MICROWAVE BACKGROUND

Introduction

The early universe was a hot place; all of the atoms we are familiar with were ionized. Due to the presence of free charges, photons were constantly being scattered. As the universe expanded, it cooled; a phase transition occurred when the temperature was low enough for electrons to become bound to protons. During *recombination*, the neutralization of the cosmic plasma resulted in the very fast growth of the mean free path of the photons. When it became larger than the Hubble radius, the universe became transparent. Photons scattered one last time, carrying the information coming from that scattering in their physical attributes, and then began their long journey through the ever emptier and cooler cosmos.

Along their myriad journeys, some of them were absorbed. Some of them encountered the gravitational wells of giant clusters of matter, and had to climb in and out of them. A small fraction of them encountered a stray free charge, scattered and then kept on flying. An even smaller fraction have even done that twice in the history of all there is! For the most part, however, their journeys have been rather uneventful. In fact, there are on average about 400 of these photons in every cubic centimeter of space, each carrying information about that last scattering that occurred so very long ago.

By measuring these photons coming in from every direction of the sky, we can get a picture of the universe from that time. The main physical quantity we can measure about these photons is their frequency. For a fixed direction in the sky, we can count how many of these photons we find as a function of frequency. Once we have gathered enough of them to be satisfied with the statistics, we can examine the resulting distribution. As it turns out, that distribution happens to be a blackbody spectrum:

$$\frac{dn}{d\nu} = \frac{8\pi\nu^2}{c^3} \frac{1}{e^{\frac{h\nu}{kT}} - 1} \tag{24}$$

In fact, it is the most exquisitely accurate blackbody spectrum every found. We can use the measured distribution to find the temperature of that direction in the sky, $T = T(\hat{n})$. This temperature map is called the Cosmic Microwave Background (CMB).

The average value over the sky is

$$\bar{T} = \frac{1}{4\pi} \int \mathrm{d}\Omega \; T(\hat{n}) \tag{25}$$

To free ourselves of the arbitrariness of a particular temperature scale, we define the fluctuations as the relative differences from the mean in any absolute temperature scale

$$\delta_T(\hat{n}) = \frac{T(\hat{n}) - \bar{T}}{\bar{T}} \tag{26}$$

From this quantity we can calculate the spatial covariance by looking at the average product of two fluctuations separated by a fixed angle:

$$\Sigma(\hat{n}, \hat{n}') = \frac{1}{4\pi} \int \mathrm{d}\Omega \int \mathrm{d}\Omega' \; \delta_T(\hat{n}) \delta_T(\hat{n}') \delta(\hat{n} \cdot \hat{n}' - \cos\theta) = C(\theta) \tag{27}$$

By construction, this function is homogenous under arbitrary rotations, $\Sigma(R(\hat{n}), R(\hat{n}')) = \Sigma(\hat{n}, \hat{n}')$. The variance in the temperature fluctuations makes up the constant "diagonal" of the covariance,

$$\sigma_T^2 = \Sigma(\hat{n}, \hat{n}) = C(0) = \frac{1}{4\pi} \int \mathrm{d}\Omega \; \delta_T(\hat{n})^2 \tag{28}$$

Spherical Expansion

Since the temperature fluctuations form a function on the sphere, δ_T can be expanded in terms of the spherical harmonics. This set of basis functions satisfies certain completeness and closure relations:

$$\int d\Omega\ Y^*_{\ell m}(\hat{n})Y_{\ell' m'}(\hat{n}) = \delta_{\ell\ell'}\delta_{mm'}$$

$$\sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\hat{n})Y_{\ell m}(\hat{n}') = \frac{2\ell+1}{4\pi}P_\ell(\hat{n}\cdot\hat{n}')$$

These relations will prove quite useful. The expansion of the fluctuations reads:

$$\delta_T(\hat{n}) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}Y_{\ell m}(\hat{n}) \quad \text{where} \quad a_{\ell m} = \int \mathrm{d}\Omega\ \delta_T(\hat{n})Y^*_{\ell m}(\hat{n}) \tag{29}$$

We note that the coefficients depend on the axis chosen to define the spherical harmonics.

What happens when we rotate that axis? Consider a rotation of the spherical coordinate system by $\hat{n}' = R(\hat{n})$. Then the spherical harmonics transform as

$$Y_{\ell m}(\hat{n}') = \sum_{m'=-\ell}^{\ell} d^{(\ell)}_{mm'}(R)Y_{\ell m'}(\hat{n}) \tag{30}$$

$$\sum_{m'=-\ell}^{\ell} |d^{(\ell)}_{mm'}(R)|^2 = 1 \tag{31}$$

where the d's form a representation of the rotation operator. Since the temperature transforms as a scalar under rotations, $\delta'_T(R(\hat{n})) = \delta_T(\hat{n})$, this implies that $a_{\ell m}$ transforms as a linear combination of the other $a_{\ell m'}$'s.

The Monopole

A quick calculation will show that $a_{00} = 0$ in the expansion. This is physically obvious since we are examining the deviation from the mean, and the spherical harmonic $Y_{00} = \frac{1}{\sqrt{4\pi}}$ is a constant. It does not need to be included in the expansion.

The Dipole

Consider an observer moving relative to the CMB rest frame with some velocity $\vec{v}$. Due to their motion they will measure Doppler shifted CMB photons, changing the reconstruction of their temperature map as compared to an observer in the CMB rest frame via:

$$T_{obs}(\hat{n}) = \frac{1+\frac{\hat{n}\cdot\vec{v}}{c}}{\sqrt{1-\frac{v^2}{c^2}}}T(\hat{n}) \tag{32}$$

This Doppler shift will introduce mixing between the actual expansion coefficients due to the $\cos\theta$ term coming from the dot product with the observer's velocity wrt the CMB. One can show that this mixing is strongest for the dipole, the observed one being contaminated by the monopole of the actual one. This means that the dipole measurement is mostly contributed to by the relative motion of the galaxy to the comsic microwave background. Disentangling this contamination from the actual dipole is difficult, so we will ignore the dipole is the spherical expansion.

Gaussian Approximation

The Gaussian approximation posits that the expansion coefficients for each ℓ are independently drawn from the same Gaussian distribution, $a_{\ell m} \sim \mathcal{N}(0, C_\ell^{\frac{1}{2}})$:

$$\langle a_{\ell m} \rangle = 0 \tag{33}$$
$$\langle a^*_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_\ell \tag{34}$$

Here we have introduced the C_ℓ's as the variances of these distributions. Under rotations the $a_{\ell m}$'s only mix with themselves for fixed ℓ, this means that the expansion can have a different axis for each ℓ. The choice of axis will have no effect on the C_ℓ's, making them a rotationally invariant description of the expansion. A quantitative model of the cosmos should make predictions about their values.

Note: the bracket averages above are with respect to the distribution the coefficients are drawn from:

$$\langle f(a_{\ell m}) \rangle = \int \mathcal{D}\{a_{\ell m}\} \ f(a_{\ell m}) \rho(\{a_{\ell m}\}) \tag{35}$$

where the measure is over all coefficients.

Estimating the Variance

We introduce the estimator of the variance as

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_m |a_{\ell m}|^2 \tag{36}$$

Imagine observers sprinkled throughout the universe. Each of them will see a different CMB, and therefore measure different values of $\hat{C}_\ell$. If, through some feat of unforeseen physics, these observers could collect all their values, what sort of statistics would they discover. First they would find that the estimator is unbiased, $\langle \hat{C}_\ell \rangle = C_\ell$. The mean of all their measurements would be the actual C_ℓ realized by the universe.

To find the higher moments(cummulants) of the estimator distribution, note that $\hat{C}_\ell$ is just the average of the squares of a bunch of Gaussian variables. We encountered the resulting probability density earlier, and now state the result:

$$\rho(\hat{C}_\ell | C_\ell) = \frac{(\ell + \frac{1}{2})^{\ell + \frac{1}{2}}}{\Gamma(\ell + \frac{1}{2})} \frac{\hat{C}_\ell^{\ell - \frac{1}{2}}}{C_\ell^{\ell + \frac{1}{2}}} e^{-(\ell + \frac{1}{2}) \frac{\hat{C}_\ell}{C_\ell}} \tag{37}$$

One can quickly verify the earlier statement that the estimator is unbiased (exercise). Furthermore, the variance in the estimator is:

$$\langle \hat{C}_\ell^2 \rangle = \frac{C_\ell^2}{\ell + \frac{1}{2}} \tag{38}$$

We see that as $\ell \to \infty$, the variance vanishes. Observers can be very confident that high ℓ estimators are nearly identical to the actual values. For small ℓ, however, they cannot be so confident. The reason for this lies in the number of $a_{\ell m}$'s needed to make each estimator. For large ℓ, there are many, and by the central limit theorem, their average becomes peaked at the true average. On the other end this is not the case. The uncertainty associated with the low ℓ estimators has been dubbed *cosmic variance.*

The Probability of a Cosmos

The Cosmos will have some value for the C_ℓ's. An observer in the Cosmos will have estimators, $\hat{C}_\ell$'s, once their technology is sufficient to measure the CMB. Given an observer's estimators, what is the probability that the Cosmos has realized a particular set of the C_ℓ's? This is a very important question for the observer, since they wish to know what Universe they actually reside in.

This question can be framed in Bayesian analysis. The model realized by the world is the set of actualized values, $M = \{C_\ell\}$. Meanwhile, the data is the set of observed estimators, $D = \{\hat{C}_\ell\}$. The strength of belief in a particular model that the observers have is then the posterior after the measurements are made:

$$p(M|D) = \mathcal{L}(D, M)p(M) \tag{39}$$

$$\mathcal{L}(D, M) = \frac{p(D|M)}{p(D)} \tag{40}$$

Here $p(M|D)$ is the *posterior*, $\mathcal{L}(D, M)$ is the *likelihood*, and $p(M)$ is the *prior*. The relationship between these three is stated as the prior belief that the observer has in the model is updated by the likelihood of the data, leading to an informed posterior belief. The model that maximizes the posterior is the model that the observer should believe in the strongest. As an added feature, Bayesian analysis also tells us exactly how much less nearby models should be believed.

One should also immediately ask about the prior. What distribution should observers choose for the C_ℓ's prior to making any measurements? Here we are once again saved by the Maximum Entropy principle: the distribution chosen should be maximally ignorant under the known constraints. Since the prior is not constrained by any measurements, MaxENT demands that it be uniform. No model should be believed in any more strongly than any other model when there is no data to constrain belief.

With regards to the likelihood, we already have the numerator from the previous section. The denominator does not depend on the model parameters, so it is treated as a constant. The posterior density reads

$$\rho(M|D) \propto \prod_\ell \left(\frac{\hat{C}_\ell}{C_\ell}\right)^{\ell+\frac{1}{2}} e^{-(\ell+\frac{1}{2})\frac{\hat{C}_\ell}{C_\ell}} \tag{41}$$

$$\propto e^{-\frac{1}{2}\chi^2_{eff}} \tag{42}$$

$$\chi^2_{eff} = \sum_\ell (2\ell+1)\left(\frac{\hat{C}_\ell}{C_\ell} + \ln\frac{C_\ell}{\hat{C}_\ell} - 1\right) \tag{43}$$

where we have introduced an effective chi-squared function, χ^2_{eff}. Furthermore, we have normalized it so that the model which maximizes the posterior gives a $\chi^2_{eff} = 0$. The sum is over all the ℓ values that the observations are over. Minimizing this function (maximizing the posterior) with respect to the model parameters gives us:

$$\nabla p(M|D) = 0 \Rightarrow C_\ell = \hat{C}_\ell \tag{44}$$

The model that maximizes the posterior is the one in which the parameters realized by the universe are equal to the estimators.

PHYSICAL COSMOLOGY

Up until now we have been discussing the observational side of cosmology: the CMB. We know why the CMB should be there, and under the Gaussian assumption can judge the observed angular power spectrum with any particular model. The question is, what is the model that describes the statistics of the CMB? To answer this question we need to delve into the realm of physical cosmology to see how the physics of the early universe imprinted itself on the temperature anisotropies of the photons making up the CMB.

The FRW metric

On large scales it is hypothesized that the Universe if isotropic and homogenous. As it turns out, observational data of galaxy cluster distributions supports this assumption. Isotropy means that the universe looks the same in every direction; there is no special direction that breaks this symmetry. Homogeneity means that no point in the universe is any different than any other. Now, of course, on small scales, both these assumptions no longer hold. This is because small scales decouple from cosmic evolution in a process known as hierarchical structure formation, which we shall examine later.

On large scales, however, isotropy and homogeneity hold, so we would like to have an expression for the possible models of spacetime that satisfy these two principles. There is an elegant geometric way to derive this class which we shall now do. The surface of a sphere satisfies both homogeneity and isotropy, so consider a $D+1$ dimensional Euclidean space, and let us embed a D-sphere in it,

$$S^D(R) = \{(x_1, \ldots, x_D, w) \in \mathbb{R}^{D+1} | x_1^2 + \cdots + x_D^2 + w^2 = R^2, R \in \mathbb{R}^+\} \tag{45}$$

where R is the radius of the sphere. Next we will look at what the induced metric on the sphere looks like that it inherits from the embedding space. We do this by treating one of the coordinates as a fucntion of the others, $w = w(x_1, \cdots, x_D)$. The induced line element reads

$$\begin{aligned} dl^2 &= dx_1^2 + \cdots + dx_D^2 + dw^2 \\ &= dx_1^2 + \cdots + dx_D^2 + \frac{(x_1 dx_1 + \cdots + x_D dx_D)^2}{R^2 - x_1^2 - \cdots - x_D^2} \\ &= \frac{dr^2}{1 - \frac{r^2}{R^2}} + r^2 d\Omega^2 \end{aligned} \tag{46}$$

where in the last line we have used spherical coordinates from the D remaining coordinates. The radius of the