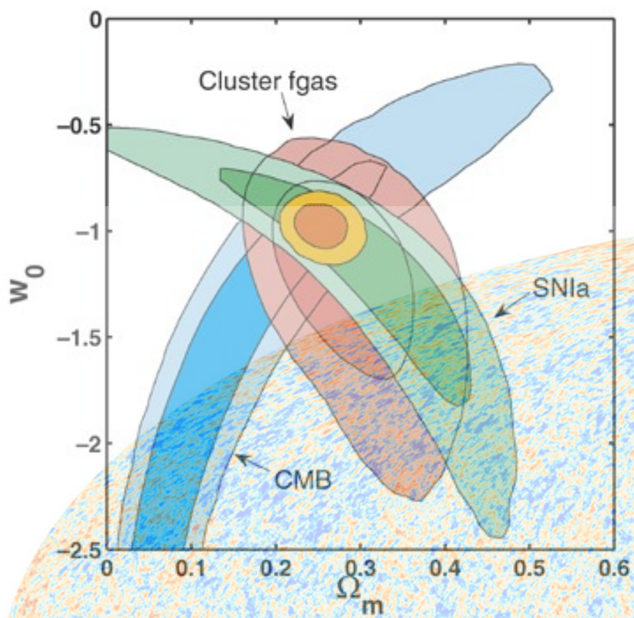
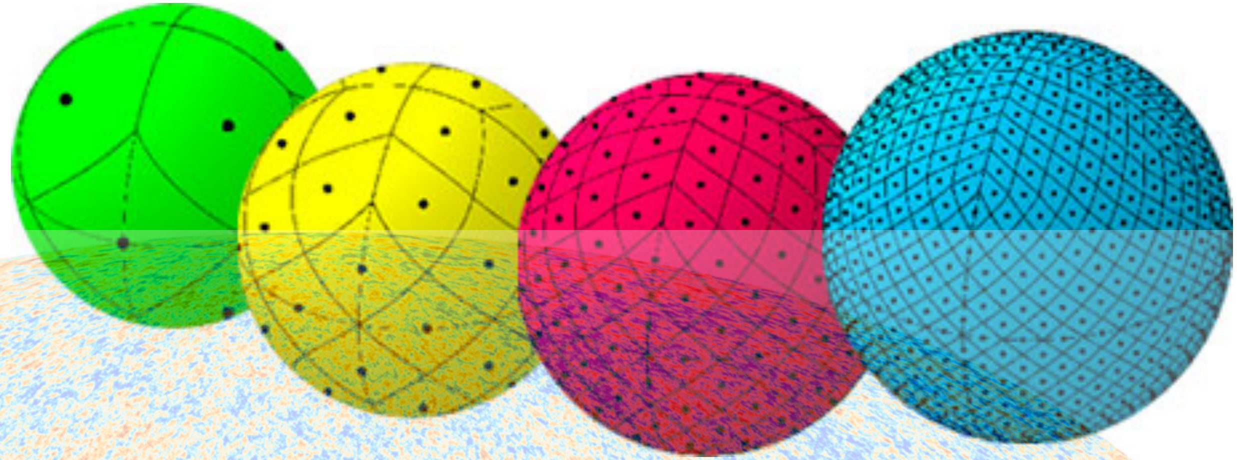
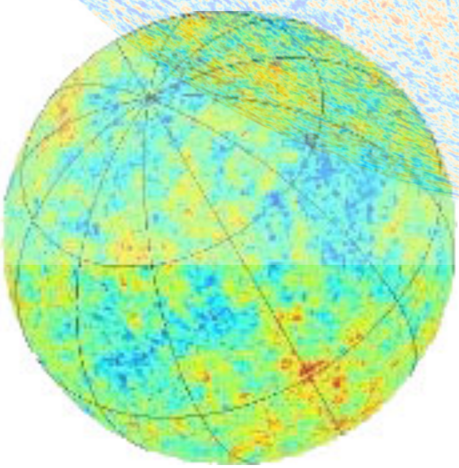




Bayesian Statistics and the CMB

Theory and challenges



Dimitrios Tanoglidis

AST 448 Presentation
31 May 2017

Part I:

Basic Theory
of
Bayesian Statistics

Conditional Probabilities

$$P(A|B)$$

(Probability of A given B)

Notice:

$$\boxed{P(A|B) \neq P(B|A)}$$

$$P(\text{pregnant}|\text{woman}) \sim< 1\%$$

But:

$$P(\text{woman}|\text{pregnant}) = 1$$

Bayes Theorem

$$P(A,B) = P(A|B)P(B) \qquad \text{joint probability}$$

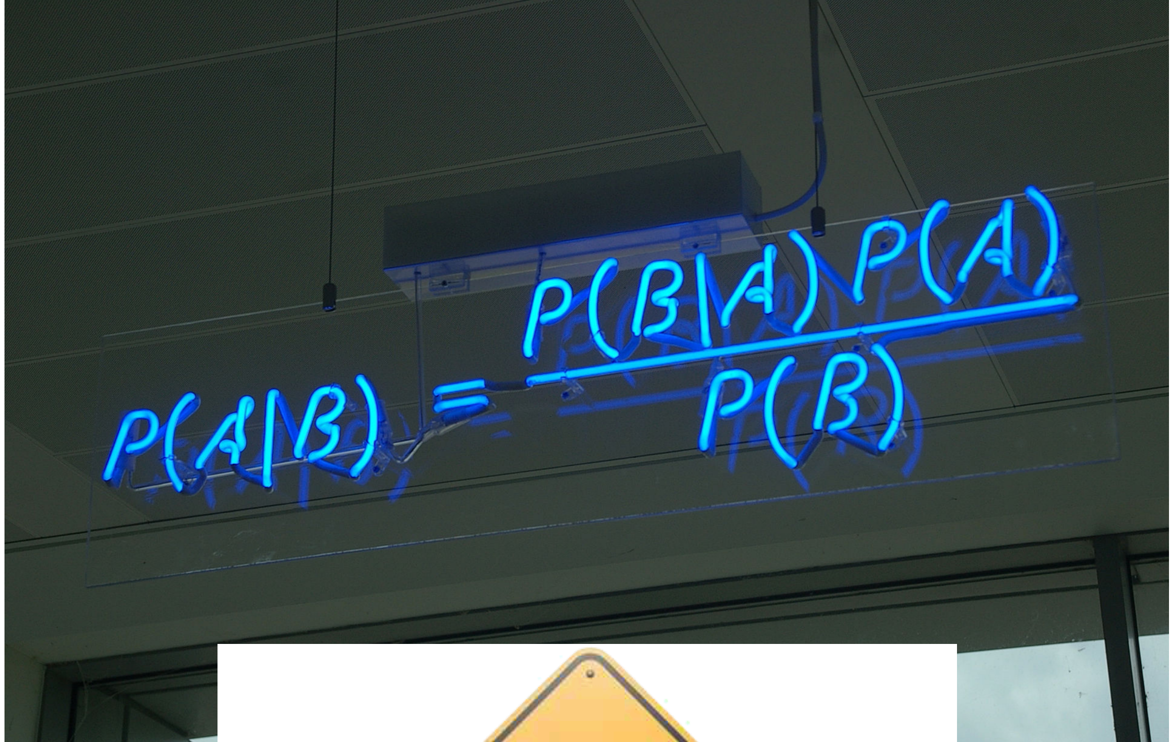
Also:

$$P(B,A) = P(B|A)P(A)$$

But: $$P(A,B) = P(B,A)$$

Combining: $$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

This is the famous Bayes theorem....


P(A|B) = P(B|A) P(A) / P(B)


SO
WHAT?

Consider again Bayes theorem…

Now: A → T = theory, B → D = data

Likelihood

Prior

$$P(T|D) = \frac{P(T)P(D|T)}{P(D)}$$

Posterior

Evidence

Bayesians

Vs

Frequintists

Usually we assume a flat prior, so posterior $\propto$ Likelihood

$$P(T|D) \propto \mathcal{L}(\mathcal{T})$$

Model (theory) with parameters ϑ

Likelihood function
(Gaussian, Poisson ...)

What are the best values ϑ given the available data?

Maximum likelihood estimator for ϑ: $\theta_{ML} \equiv \max_{\theta} \mathcal{L}(\theta)$

Recipe:

$$\left. \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta} \right|_{\theta_{ML}} = 0$$

Example of (Gaussian) likelihood:

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

Where:

$$C_{ij} = \langle (d_i - \theta_i)(d_j - \theta_j) \rangle$$

Is the covariance matrix

Marginalization

What is ϑ? For example, it can be:

$$\theta = \{\Omega_m, \Omega_\Lambda, H_0, \dots\}$$

You want, for example, to know the probability distribution of Ω_m

Regardless of the values of the other parameters (sometimes referred as **nuisance** parameters). We simply integrate out these parameters:

$$P(\Omega_m) = \int d\Omega_\Lambda dH_0 d \dots P(\Omega_m, \Omega_\Lambda, H_0, \dots)$$

This process is known as Marginalization

Confidence Intervals and Combination of Experiments

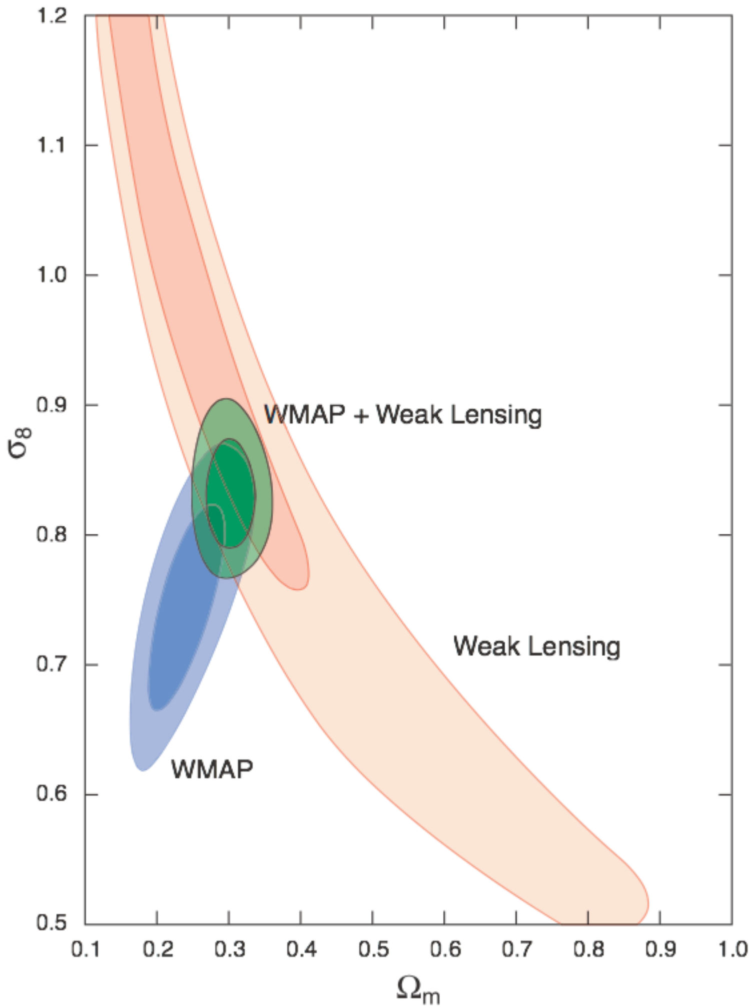
Regions of confidence (or *belief*). These are defined as regions R(α) of constant likelihood, for which:

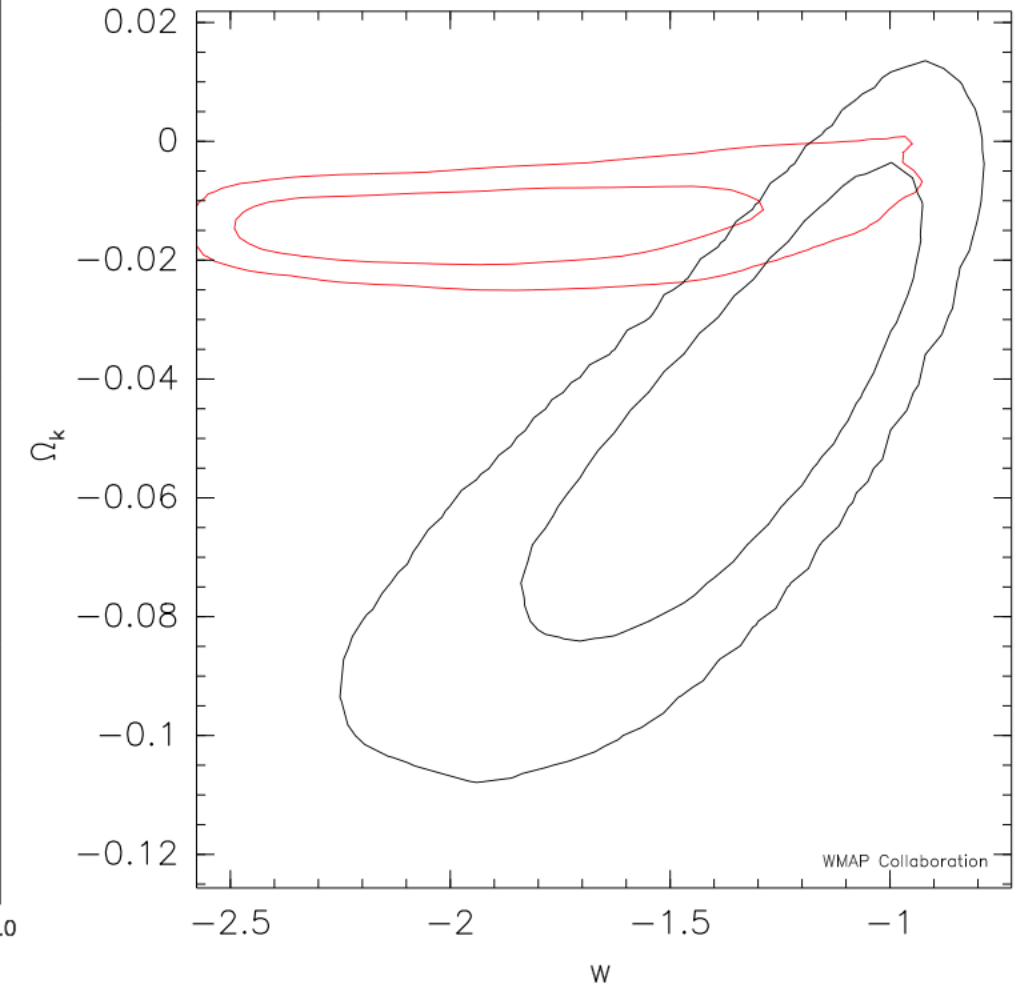
$$\int_{R(\alpha)} \mathcal{L}(\theta_i) d^n\theta = \alpha$$

Where: $0 < \alpha < 1$

Typical choices: $\alpha = 0.683, 0.954, 0.997$

We'll say more things about errors… See Fisher matrix





Q: Are all data sets compatible?

Fisher Matrix Formalism
and
Error Forecasts

Let's find an elegant way to handle models with many parameters....

Consider a general log-likelihood (one parameter ϑ) and expand around the maximum likelihood estimator:

$$\ln \mathcal{L}(\theta) = \ln \mathcal{L}(\theta_{ML}) + \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta}|_{\theta_{ML}} (\theta - \theta_{ML}) + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}\theta}{\partial^2 \theta} (\theta - \theta_{ML})^2 + \ldots$$

Second term vanishes, likelihood can be written:

$$\mathcal{L}(\theta) \cong \mathcal{L}(\theta_{ML}) \exp\left(-\frac{1}{2} \frac{(\theta - \theta_{ML})^2}{\Sigma_\theta} \right) + \ldots$$

Where:

Error → $$\frac{1}{\Sigma_\theta^2} = -\frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial^2 \theta}|_{\theta_{ML}}$$

Easy to generalize this:

$$\ln\mathcal{L}(\theta) = \ln\mathcal{L}(\theta_{ML}) + \frac{1}{2}\sum_{ij}(\theta_i - \theta_{ML,i})\frac{\partial^2 \ln\mathcal{L}(\theta)}{\partial^2\theta}(\theta_j - \theta_{ML,j}) + \ldots$$

If I define the **Fisher Matrix:**

$$F_{ij} = -\left\langle \frac{\partial^2 \ln\mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle$$

You'll not be surprised if I tell you that:

$$\sigma_{\theta_i} \geq (F^{-1})_{ii}^{1/2}$$

Equal sign for gaussian likelihood

Cramer – Rao inequality

This is known as the marginalized error

We can estimate the parameters errors before doing the experiment

We'll see later the Fisher Matrix for the CMB

Actually one has to use stochastic methods

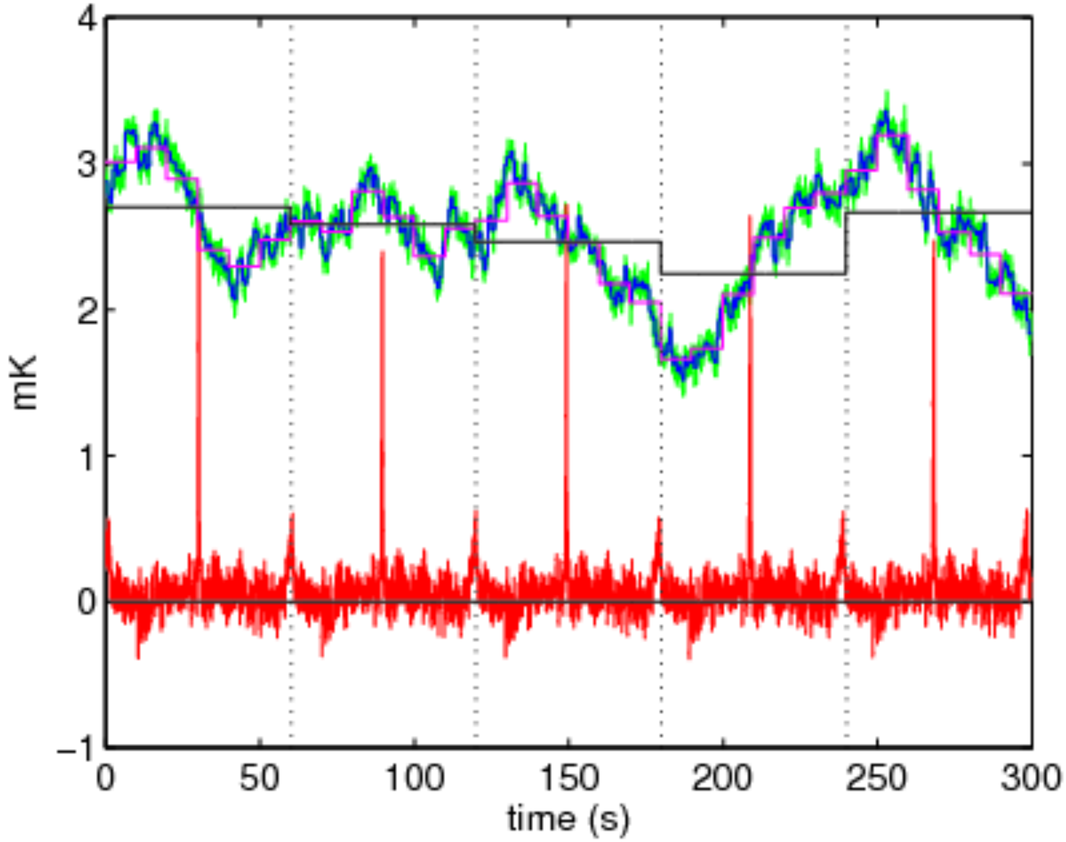
(MCMC)

Details not covered here…

Part II:

Statistical Challenges of Cosmic Microwave Background Analysis

From Time Ordered Data (TOD)



To cosmological Parameters

Where is the info?

No information in any feature of the CMB map

Information encoded in the statistical properties of the CMB map – or the invariant under SO(3) quantities of the temperature / polarization anisotropies.

Simplest models → Only 2p function counts....

A model can have 10 – 20 parameters we want to constrain.

Recipe. Given a model, calculate:

$$\left\langle \frac{\Delta T}{T_0}(\hat{x}) \frac{\Delta T}{T_0}(\hat{y}) \right\rangle = C(\arccos(\hat{x} \cdot \hat{y}))$$

Since we are working on the surface of a sphere, it is easier to work with spherical harmonics:

$$T(\hat{n}) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{+\ell} a_{\ell m} Y_{\ell m}(\hat{n})$$

Note starting point

Diagonal correlations: $\langle a_{\ell m} a^*_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}$

Correlation function

Power Spectrum

$$C(\theta) = \sum_{\ell} \frac{2\ell + 1}{4\pi} C_\ell P_\ell(\cos\theta)$$

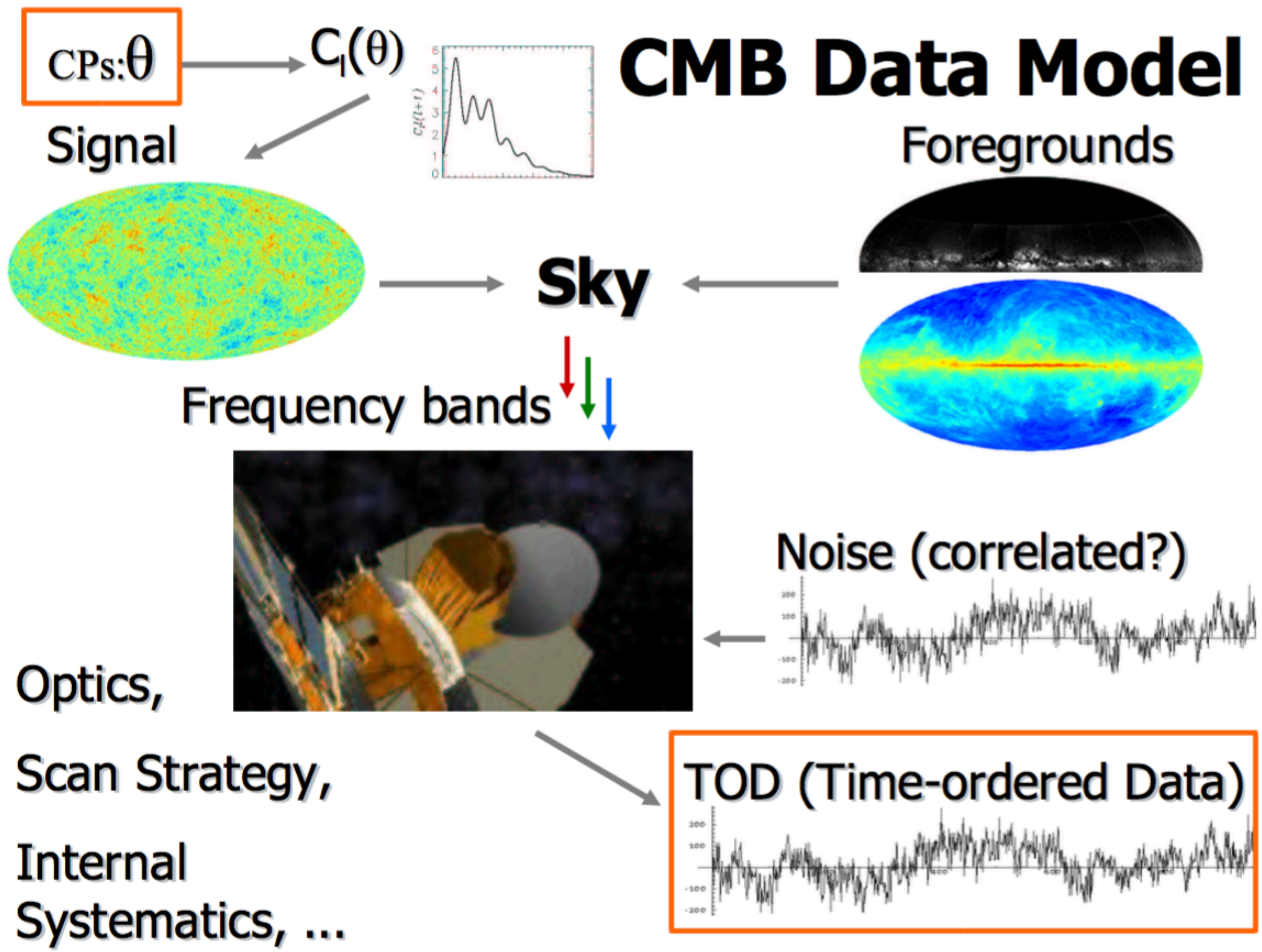
Power spectrum MLE estimator:

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

Compare theory and observations

That easy???

Data from CMB observations



(Wandelt 2003)

Data from CMB observations

- Restricted from our position on the sky → other sources of microwaves (foregrounds)
- Instrument systematics. Microwaves have macroscopic wavelengths → diffract around the edges of the instrument.
- Detector adds noise
- Maps from different bands

The data program:

For PLANCK: Complete TOD ~ 1 Terabyte of storage

↓

100 detectors → Each one results in a map of order 10 – 100 Mb

↓

Grouping into 10 frequency bands

↓

CMB maps

↓

Calculation of the power spectrum (~ a few thousand Cl s)

↓

Cosmological parameters

Time Ordered Data - Inference

Action of the detector

data

$$d_i = \sum_p A_{ip} T_p + n_i$$

Noise

$$i = (f, t)$$

Temperature at pixel p

frequency

pixel

In matrix form:

$$\mathbf{d} = \mathbf{A}\,\mathbf{T} + \mathbf{n}$$

$$\langle n_i \rangle = 0, \qquad \langle n_i n_j \rangle = N_{ij}$$

$$A, N \Rightarrow \text{known}$$

Apply Bayes Theorem to get posterior:

$$P(C_\ell, \Theta, A|d) = \frac{P(d|C_\ell, \Theta, A)P(C_\ell, \Theta, A)}{P(d)}$$

We have to explore this (Likelihood), or
equivalently
The **CMB Fisher Information matrix**

Let' s just give the CMB Fisher matrix

Gaussian data:

$$F_{ij} = \frac{1}{2} Tr \left[C^{-1} C,_i C^{-1} C,_j + C^{-1} M_{ij} \right]$$

$$M_{ij} = \theta,_i \theta^T,_j + \theta,_j \theta^T,_i$$

Signal for CMB: $\mathrm{s} = (a_\ell^T, a_\ell^E, a_\ell^B)$

$$F_{ij}^{CMB} = \sum_{XY} \sum_\ell \frac{\partial C_\ell^X}{\partial \theta_i} \left(\mathcal{C}_\ell^{XY}\right)^{-1} \frac{\partial C_\ell^Y}{\partial \theta_j}$$

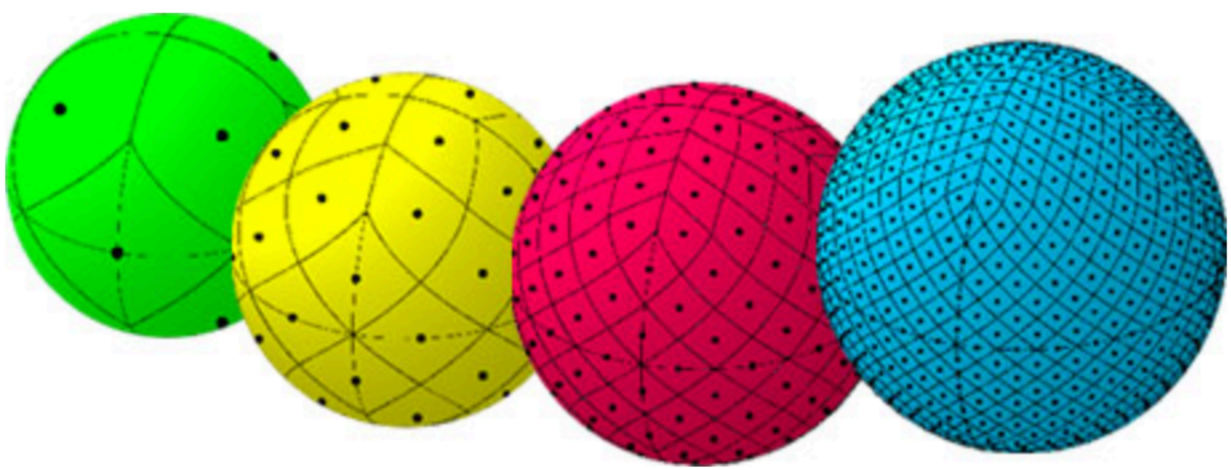
$$X, Y = TT, TE, EE, BB, etc....$$

More challenges….

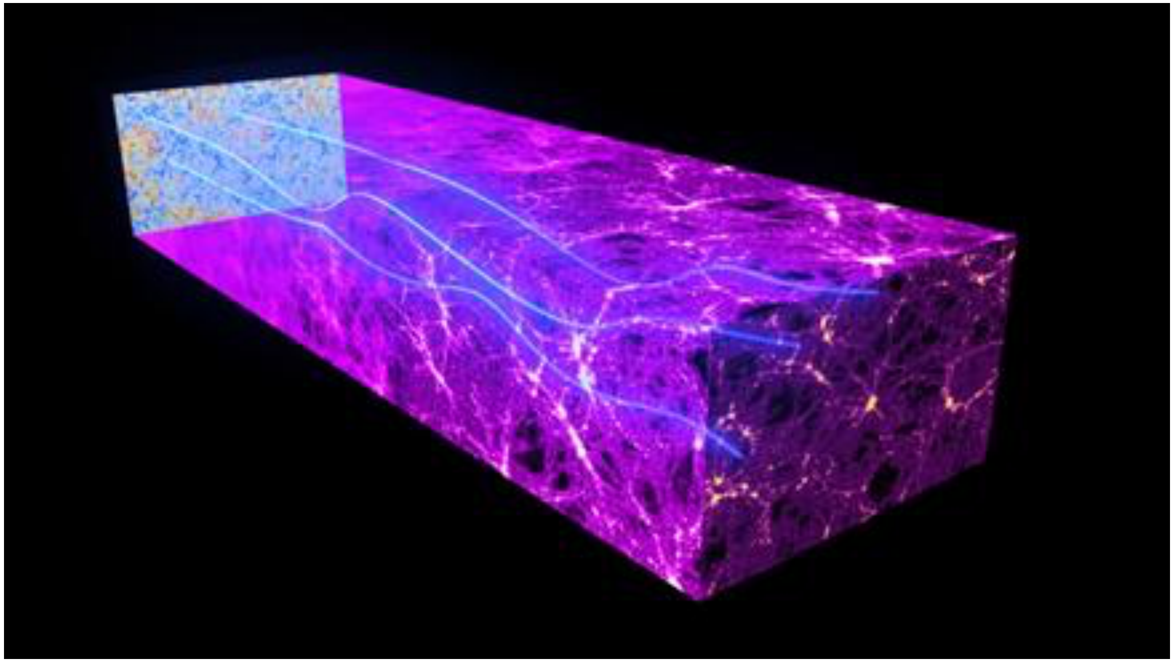
How to pixelize on the sphere?

Algorithms:

HEALPix is an acronym for **H**ierarchical **E**qual **A**rea iso**L**atitude **Pix**elization of a sphere. As suggested in the name, this pixelization produces a subdivision of a spherical surface in which each pixel covers the same surface area as every other pixel. The figure below shows the partitioning of a sphere at progressively higher resolutions, from left to right. The green sphere represents the lowest resolution possible with the HEALPix base partitioning of the sphere surface into 12 equal sized pixels. The yellow sphere has a HEALPix grid of 48 pixels, the red sphere has 192 pixels, and the blue sphere has a grid of 768 pixels (~7.3 degree resolution).



Weak Lensing → Needs delensing



Beam decomvolution

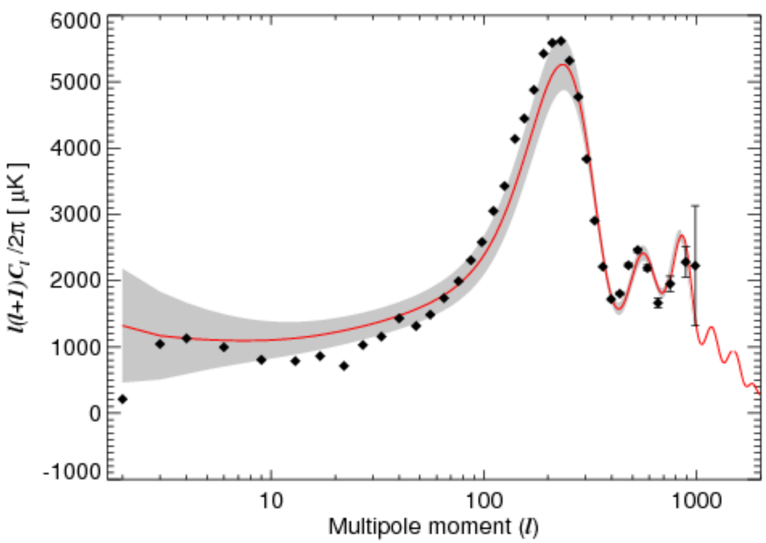
Component separation

See again the formula for the estimation of power spectrum:

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

For low *l* just a few values of m
→ Poor statistics

Cosmic Variance

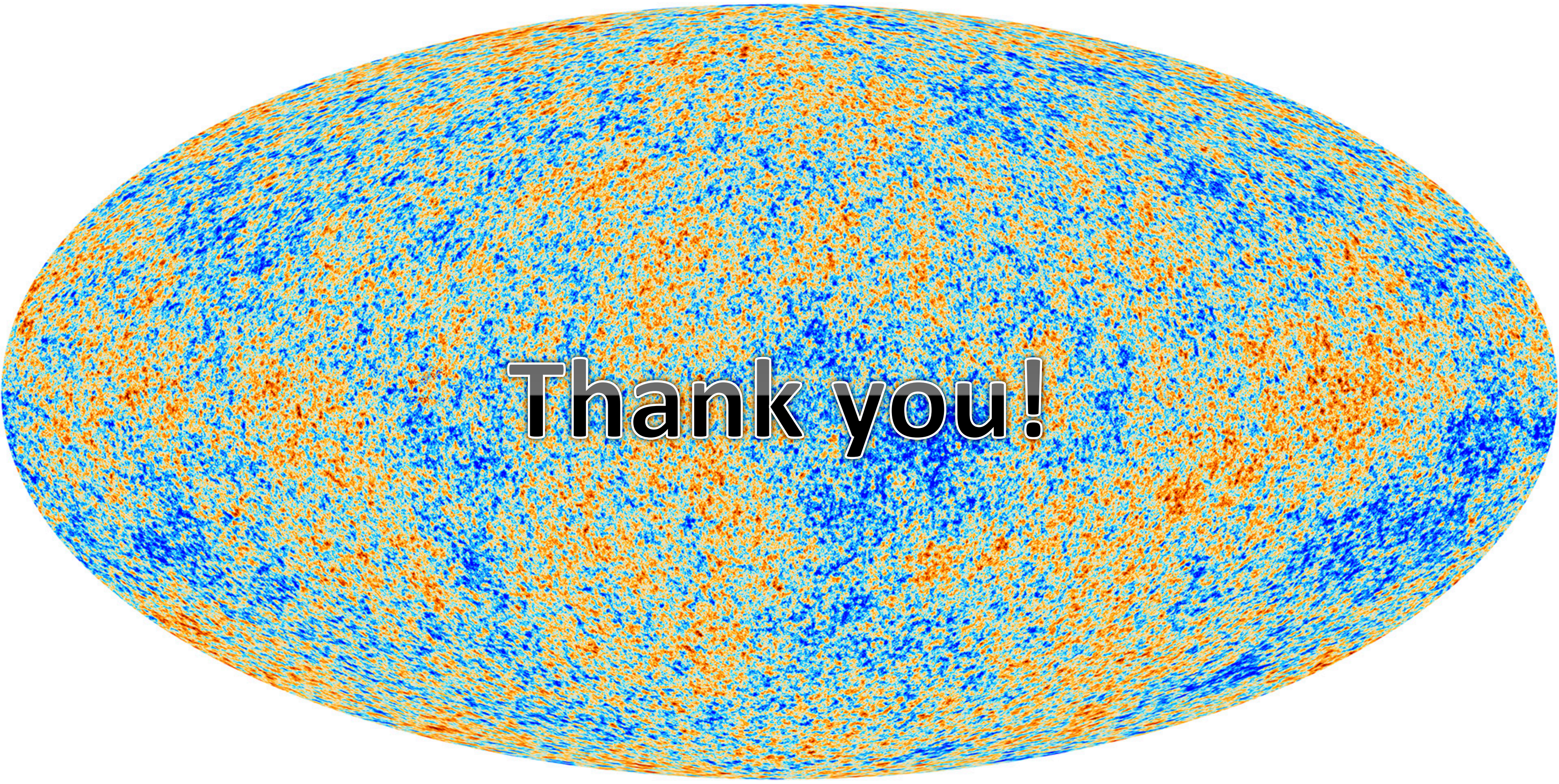


Computational limitations due to the time needed to perform operations as calculations of C_ls and then estimate parameters …

For one likelihood evaluation with Planck data: 10^{21}

Operations → thousands of CPU years

Still things to be done


Thank you!