Statistical methods for cosmology

Thibaut Louis

IDPASC

IJCLab Irène Joliot-Curie
Laboratoire de Physique des 2 Infinis

Outline:

I) Reminder on Probabilities

II) χ^2 statistics

III) Multivariate Probability Functions

IV) The Multidimensional Normal Distribution

Reminder on Probabilities

The False Positive Paradox

Gerd Gigerenzer

Breast Cancer test:

A 50-year-old woman (with no symptoms) undergoes a mammogram. She tests positive and wants to know: what is the probability that she actually has breast cancer?

Sensitivity: the probability that the test is positive given that the person has the disease

$$P(+|M) = 90\%$$

Specificity: The probability of getting a negative test result given that the person does not have the disease

$$P(-|\neg M) = 91\%$$

Disease Prevalence in the Population:

$$P(M) = 1\%$$

What is the best estimate of her probability of having cancer? :

A) 90%
B) 50%
C) 10%
D) 1%

The False Positive Paradox

Gerd Gigerenzer

Breast Cancer test:

A 50-year-old woman (with no symptoms) undergoes a mammogram. She tests positive and wants to know: what is the probability that she actually has breast cancer?

Sensitivity: the probability that the test is positive given that the person has the disease

$$P(+|M) = 90\%$$

Specificity: The probability of getting a negative test result given that the person does not have the disease

$$P(-|\neg M) = 91\%$$

Disease Prevalence in the Population:

$$P(M) = 1\%$$

$$P(M|+) \quad = \quad \frac{P(+|M)P(M)}{P(+)}$$

Bayes theorem

The False Positive Paradox

Gerd Gigerenzer

Breast Cancer test:

**A 50-year-old woman (with no symptoms) undergoes a mammogram.
She tests positive and wants to know: what is the probability that she actually has breast cancer?**

Sensitivity: the probability that the test is positive given that the person has the disease

$$P(+|M) = 90\%$$

Specificity: The probability of getting a negative test result given that the person does not have the disease

$$P(-|\neg M) = 91\%$$

Disease Prevalence in the Population:

$$P(M) = 1\%$$

$$P(M|+) \quad = \quad \frac{P(+|M)P(M)}{P(+)} = \frac{P(+|M)P(M)}{P(+|M)P(M) + P(+|\neg M)P(\neg M)}$$

The False Positive Paradox

Gerd Gigerenzer

Breast Cancer test:

**A 50-year-old woman (with no symptoms) undergoes a mammogram.
She tests positive and wants to know: what is the probability that she actually has breast cancer?**

Sensitivity: the probability that the test is positive given that the person has the disease

$$P(+|M) = 90\%$$

Specificity: The probability of getting a negative test result given that the person does not have the disease

$$P(-|\neg M) = 91\%$$

Disease Prevalence in the Population:

$$P(M) = 1\%$$

$$\begin{aligned} P(M|+) &= \frac{P(+|M)P(M)}{P(+)} = \frac{P(+|M)P(M)}{P(+|M)P(M) + P(+|\neg M)P(\neg M)} \\ &= \frac{P(+|M)P(M)}{P(+|M)P(M) + [1 - P(-|\neg M)]\,[1 - P(M)]} \end{aligned}$$

The False Positive Paradox

Gerd Gigerenzer

Breast Cancer test:

A 50-year-old woman (with no symptoms) undergoes a mammogram.
She tests positive and wants to know: what is the probability that she actually has breast cancer?

Sensitivity: the probability that the test is positive given that the person has the disease

$$P(+|M) = 90\%$$

Specificity: The probability of getting a negative test result given that the person does not have the disease

$$P(-|\neg M) = 91\%$$

Disease Prevalence in the Population:

$$P(M) = 1\%$$

$$\begin{aligned} P(M|+) &= \frac{P(+|M)P(M)}{P(+)} = \frac{P(+|M)P(M)}{P(+|M)P(M) + P(+|\neg M)P(\neg M)} \\ &= \frac{P(+|M)P(M)}{P(+|M)P(M) + [1 - P(-|\neg M)]\,[1 - P(M)]} = \frac{90\% \times 1\%}{90\% \times 1\% + 9\% \times 99\%} \approx 9\% \end{aligned}$$

Prior (prior probability): the probability assigned before any data is observed

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)}$$

Posterior: probability given the test result

Likelihood

Prior (prior probability): the probability assigned before any data is observed

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)}$$

Likelihood

Posterior: probability given the test result

Evidence (also called Marginal Likelihood) The total probability of observing the data under all possible hypotheses.

Random Variables and Random Fields

Sick/not sick, positive/negative are not very quantitative concepts. In cosmology, we tend to study variables taking values in $\mathbb{R}$ or $\mathbb{N}$ we use probability densities, for example:

$$\mathcal{P}(T_{\rm CMB}|d_{\rm cobe})$$

$$\mathcal{P}(\Omega_m h^2, \Omega_b h^2, n_s, A_s, \theta_{\rm MC}, \tau, \omega_0, \omega_a, \sum m_\nu | d_{\rm cosmology})$$

Cosmology also studies stochastic fields, where the concept of probability is generalized to an infinite number of random variables.

$$\mathcal{P}[\delta T_{\rm CMB}(\hat{n})], \mathcal{P}[\rho(\vec{x})]$$

Properties

$$\int_a^b \mathcal{P}_X(x)dx = P(a < X \leq b)$$

$$\int_{-\infty}^{+\infty} \mathcal{P}_X(x)dx = 1$$

Moments of a random variable

Mean $$\mathbb{E}(X) = \langle X \rangle = \bar{X} = \mu = \int_{-\infty}^{+\infty} x\mathcal{P}_X(x)dx$$

Properties

$$\int_a^b \mathcal{P}_X(x)dx = P(a < X \leq b)$$

$$\int_{-\infty}^{+\infty} \mathcal{P}_X(x)dx = 1$$

Moments of a random variable

Mean $$\mathbb{E}(X) = \langle X \rangle = \bar{X} = \mu = \int_{-\infty}^{+\infty} x\mathcal{P}_X(x)dx$$

Variance $$\mathbb{V}(X) = \sigma^2 = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{+\infty} (x-\mu)^2\mathcal{P}_X(x)dx$$

Properties

$$\int_a^b \mathcal{P}_X(x)dx = P(a < X \leq b)$$

$$\int_{-\infty}^{+\infty} \mathcal{P}_X(x)dx = 1$$

Moments of a random variable

Mean $$\mathbb{E}(X) = \langle X \rangle = \bar{X} = \mu = \int_{-\infty}^{+\infty} x\mathcal{P}_X(x)dx$$

Variance $$\mathbb{V}(X) = \sigma^2 = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{+\infty} (x-\mu)^2\mathcal{P}_X(x)dx$$

Skewness $$\gamma_1 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Kurtosis $$\gamma_2 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

Properties

$$\int_a^b \mathcal{P}_X(x)dx = P(a < X \leq b)$$

$$\int_{-\infty}^{+\infty} \mathcal{P}_X(x)dx = 1$$

Moments of a random variable

Mean $$\mathbb{E}(X) = \langle X \rangle = \bar{X} = \mu = \int_{-\infty}^{+\infty} x\mathcal{P}_X(x)dx$$

Variance $$\mathbb{V}(X) = \sigma^2 = \mathbb{E}[(X - \mathbb{E}(X))^2] = \int_{-\infty}^{+\infty} (x-\mu)^2\mathcal{P}_X(x)dx$$

Skewness $$\gamma_1 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Kurtosis $$\gamma_2 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$$

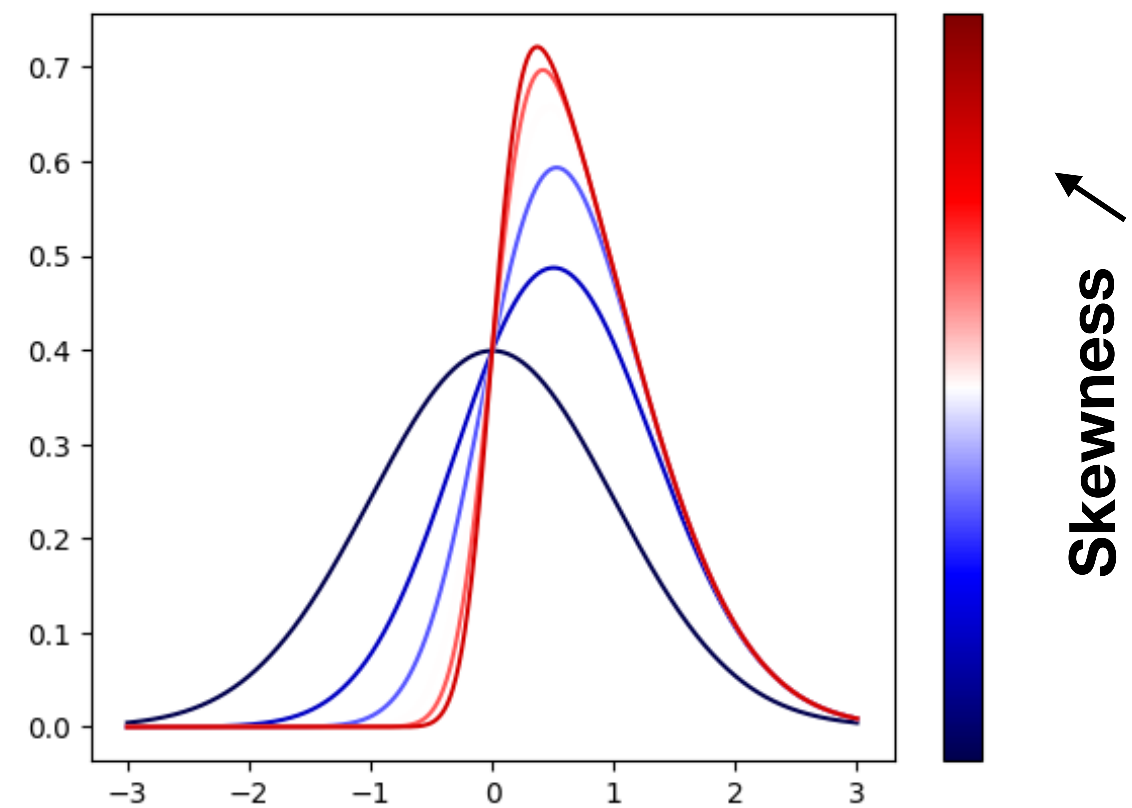
0 For a Gaussian random variable

Non-Gaussianity

Asymmetric Normal Distribution

$$\mathcal{P}(x;\epsilon,\omega,\alpha) = \frac{1}{\omega\pi} e^{-\frac{(x-\epsilon)^2}{2\omega^2}} \int_{-\infty}^{\alpha\frac{x-\epsilon}{\omega}} e^{-t^2/2} dt$$

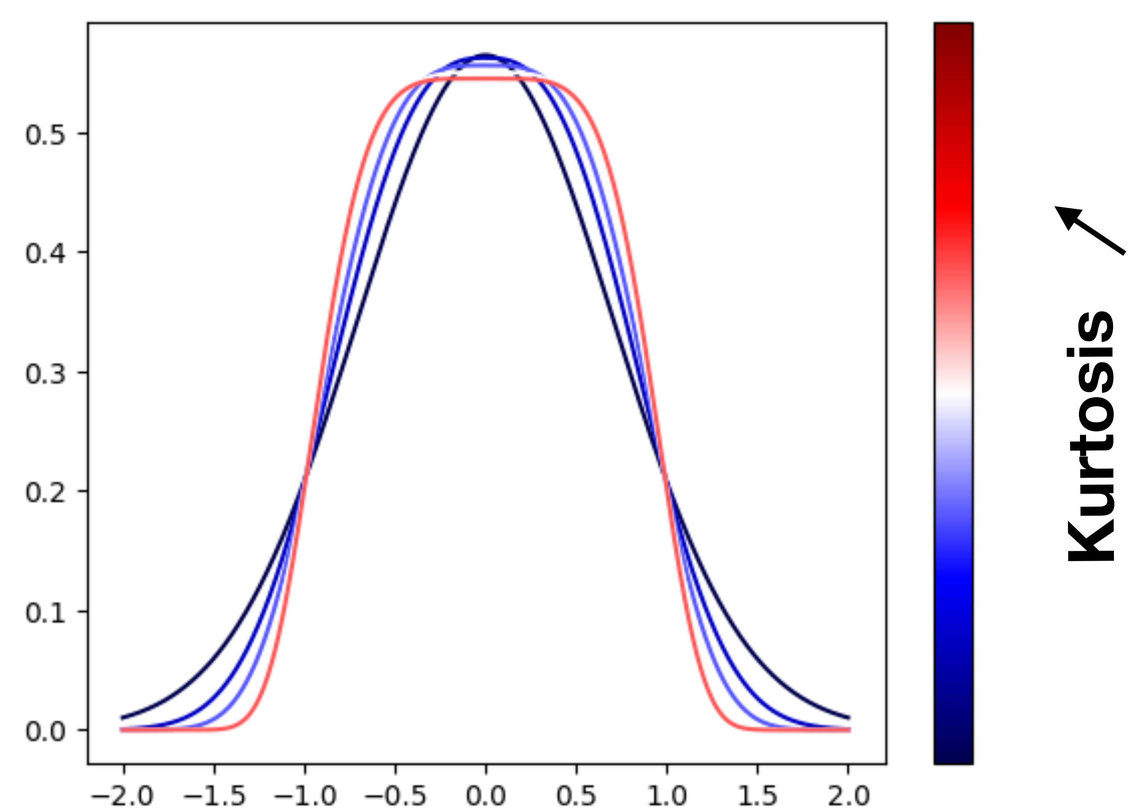
$\alpha = 0$ ⟶ **Gaussian**



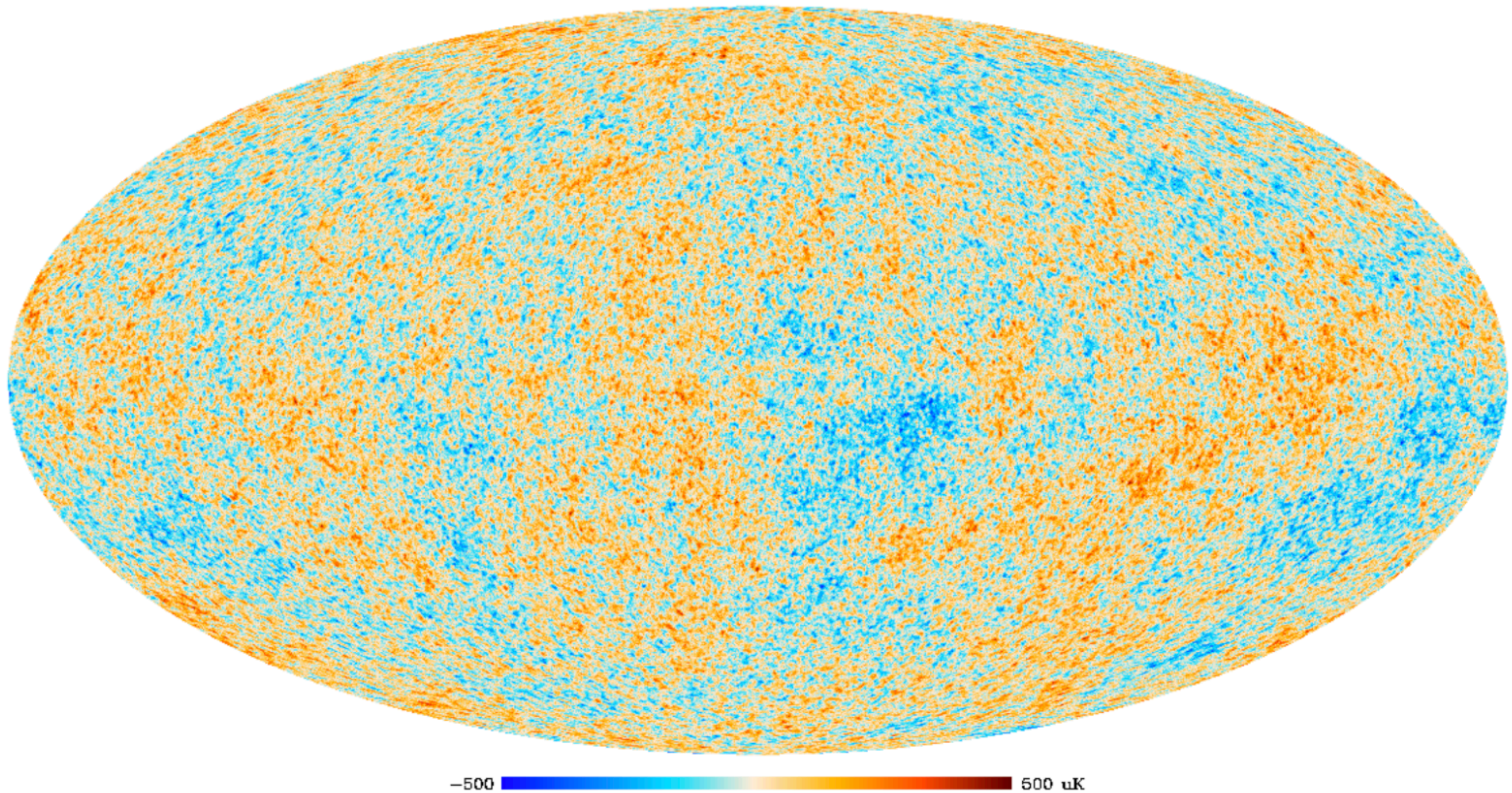
Generalized Normal Distribution

$$\mathcal{P}(x;\alpha\beta,\mu) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}}$$

$\beta = 2$ ⟶ **Gaussian**



Interlude : In practice, we often do not know a priori the probability density of a physical process, but we can infer it from the measurement of its moments

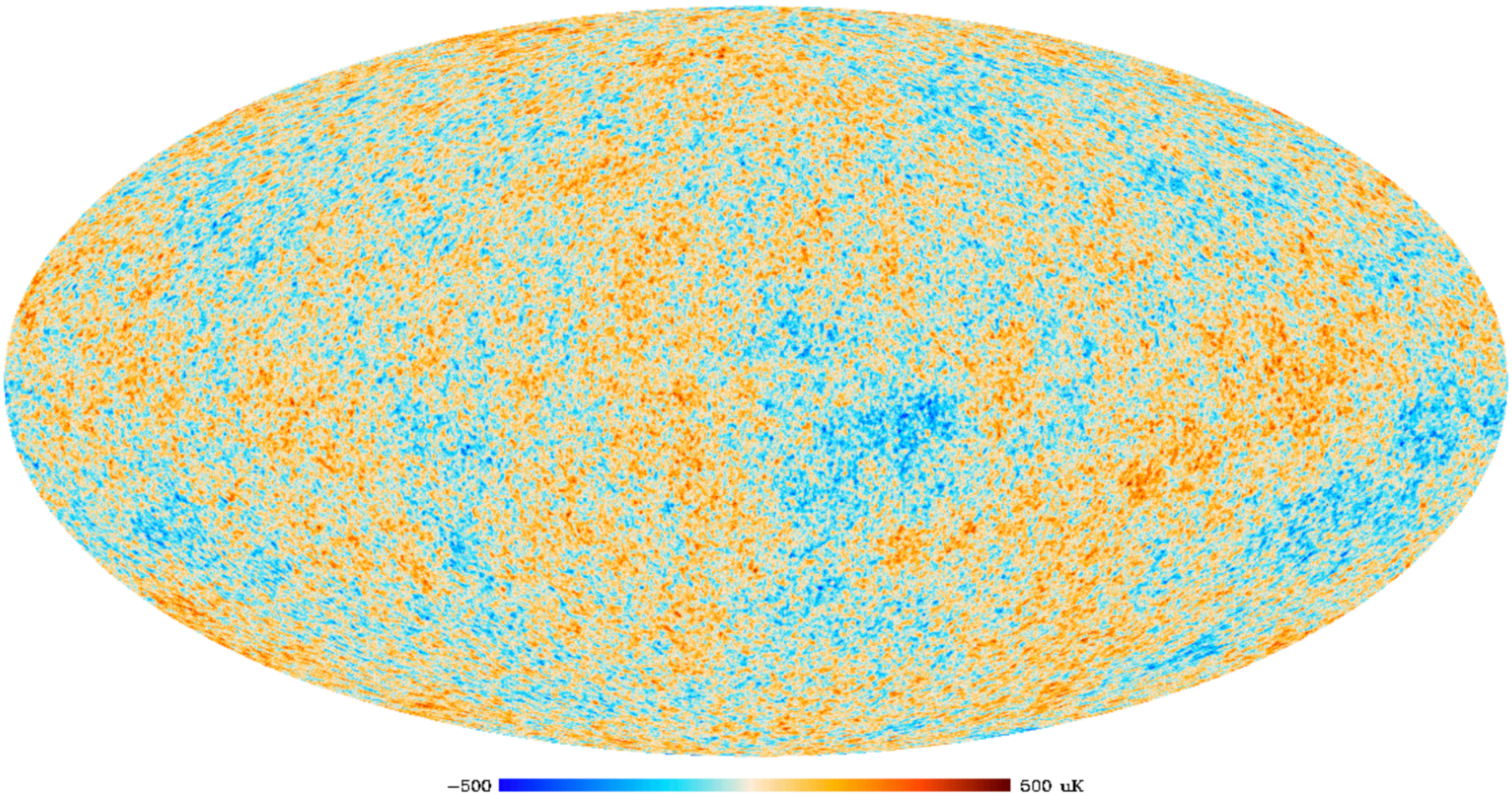


Planck data

The calculation of the moments of CMB temperature anisotropies indicates a Gaussian distribution, in particular

$$\langle \delta T(\hat{n}_1)\delta T(\hat{n}_2)\delta T(\hat{n}_3)\rangle \sim 0$$

Interlude : In practice, we often do not know a priori the probability density of a physical process, but we can infer it from the measurement of its moments



Planck data

The calculation of the moments of CMB temperature anisotropies indicates a Gaussian distribution, in particular

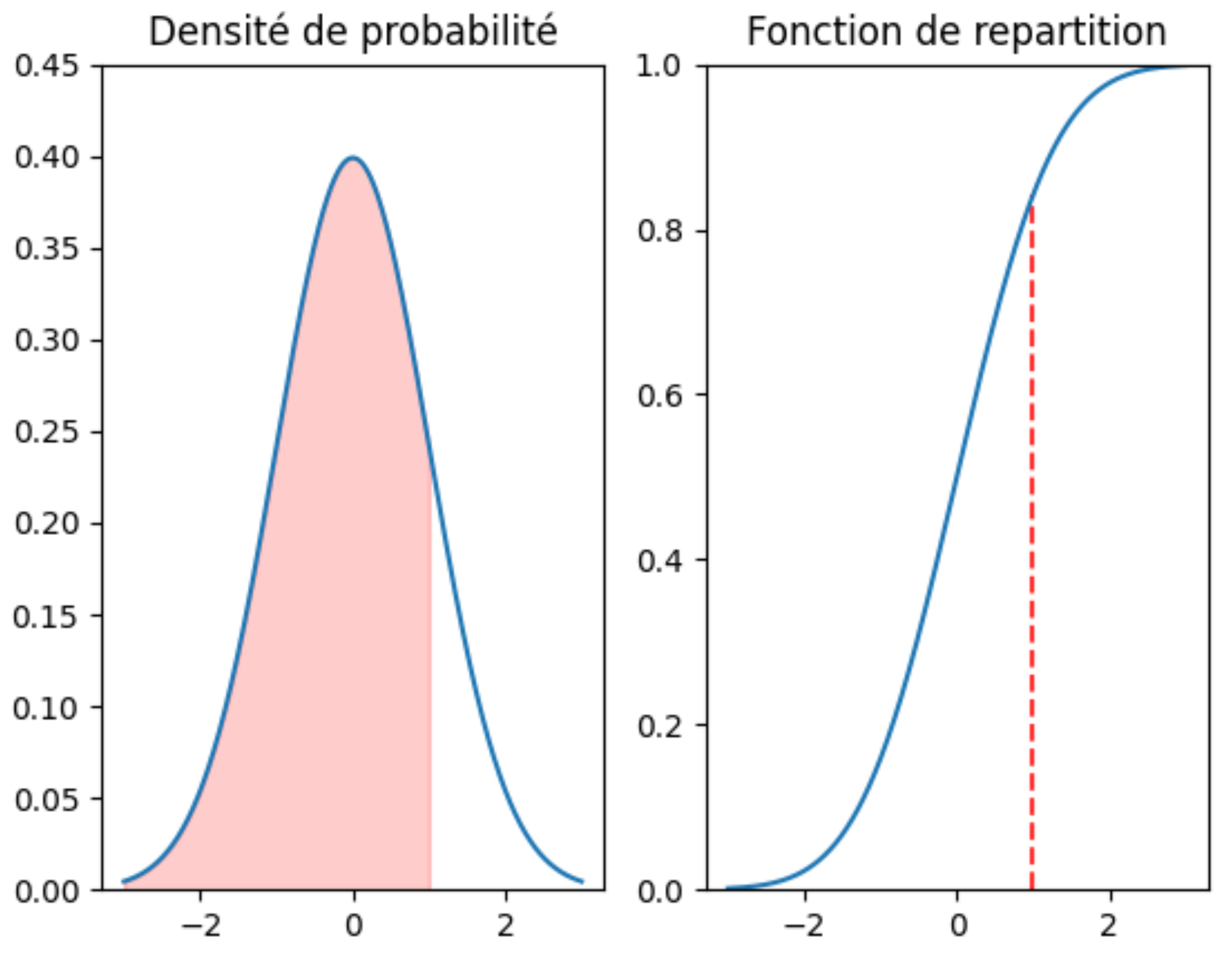
$$\langle \delta T(\hat{n}_1)\delta T(\hat{n}_2)\delta T(\hat{n}_3)\rangle \sim 0$$

$\longrightarrow$ **Indication in favor of inflation and constraints on models**

Cumulative distribution function

Cumulative distribution function

$$F_X(x) = \int_{-\infty}^{x} \mathcal{P}_X(x')dx' = P(-\infty < X \leq x) = P(X \leq x)$$

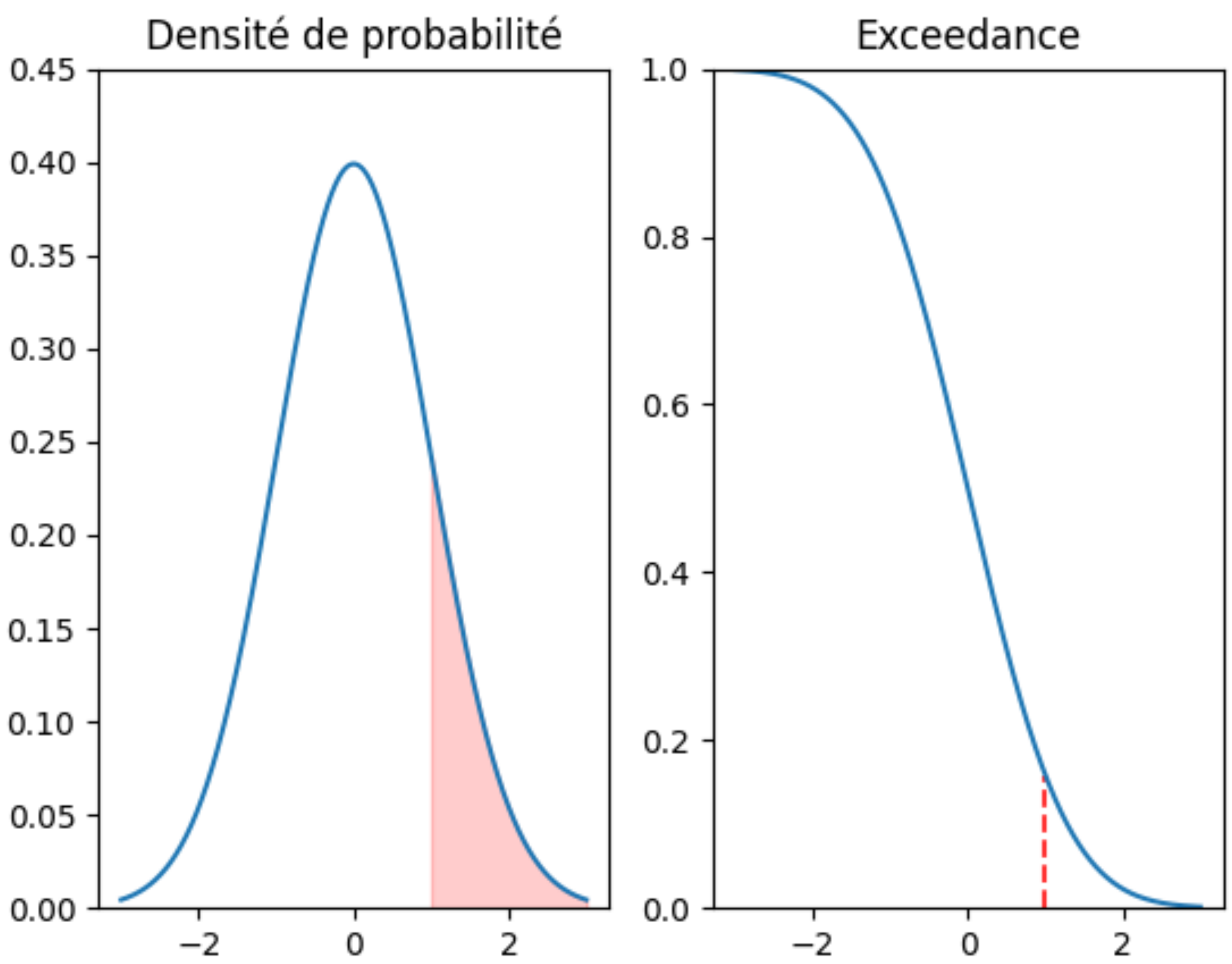


Cumulative distribution function

Complementary cumulative distribution function

$$\bar{F}_X(x) = P(X > x) = 1 - F_X(x)$$

Tail distribution
Exceedance
One sided p-value
Survival function
Reliability function



This thing has so many names
it must be useful

Example I: The Hubble constant Problem

$$
\begin{aligned}
H_0^{\mathrm{Planck}} &= 67.36 \pm 0.54 \ \mathrm{km/s/Mpc} \\
H_0^{\mathrm{Riess}} &= 73.2 \pm 1.3 \ \mathrm{km/s/Mpc}
\end{aligned}
$$

Are these two measurements in agreement or in disagreement?

Example I: The Hubble constant Problem

$$
\begin{aligned}
H_0^{\mathrm{Planck}} &= 67.36 \pm 0.54 \ \mathrm{km/s/Mpc} \\
H_0^{\mathrm{Riess}} &= 73.2 \pm 1.3 \ \mathrm{km/s/Mpc}
\end{aligned}
$$

Are these two measurements in agreement or in disagreement?

To quantify this, let's start from the hypothesis (also called the null hypothesis) that the two measurements are unbiased, independent, and that the probability distribution associated with each measurement is Gaussian.

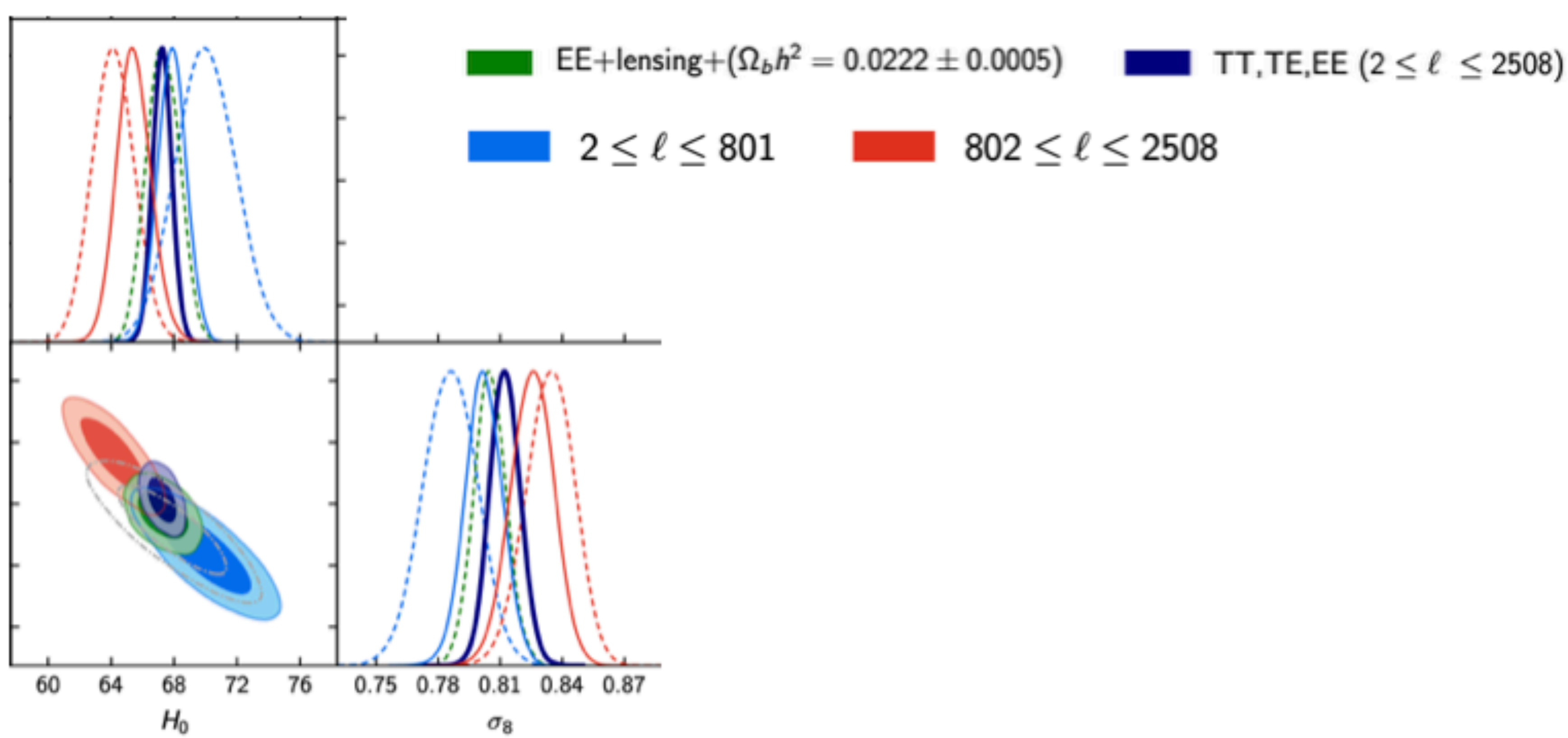
Example I: The Hubble constant Problem

$$\begin{aligned} H_0^{\rm Planck} &= 67.36 \pm 0.54 \ \mathrm{km/s/Mpc} \\ H_0^{\rm Riess} &= 73.2 \pm 1.3 \ \mathrm{km/s/Mpc} \end{aligned}$$

Are these two measurements in agreement or in disagreement?

To quantify this, let's start from the hypothesis (also called the null hypothesis) that the two measurements are unbiased, independent, and that the probability distribution associated with each measurement is Gaussian.

Planck H0 Posterior



Example I: The Hubble constant Problem

Let us define the new random variable

$$\Delta H_0 = H_0^{\text{Riess}} - H_0^{\text{Planck}}$$

It follows a distribution

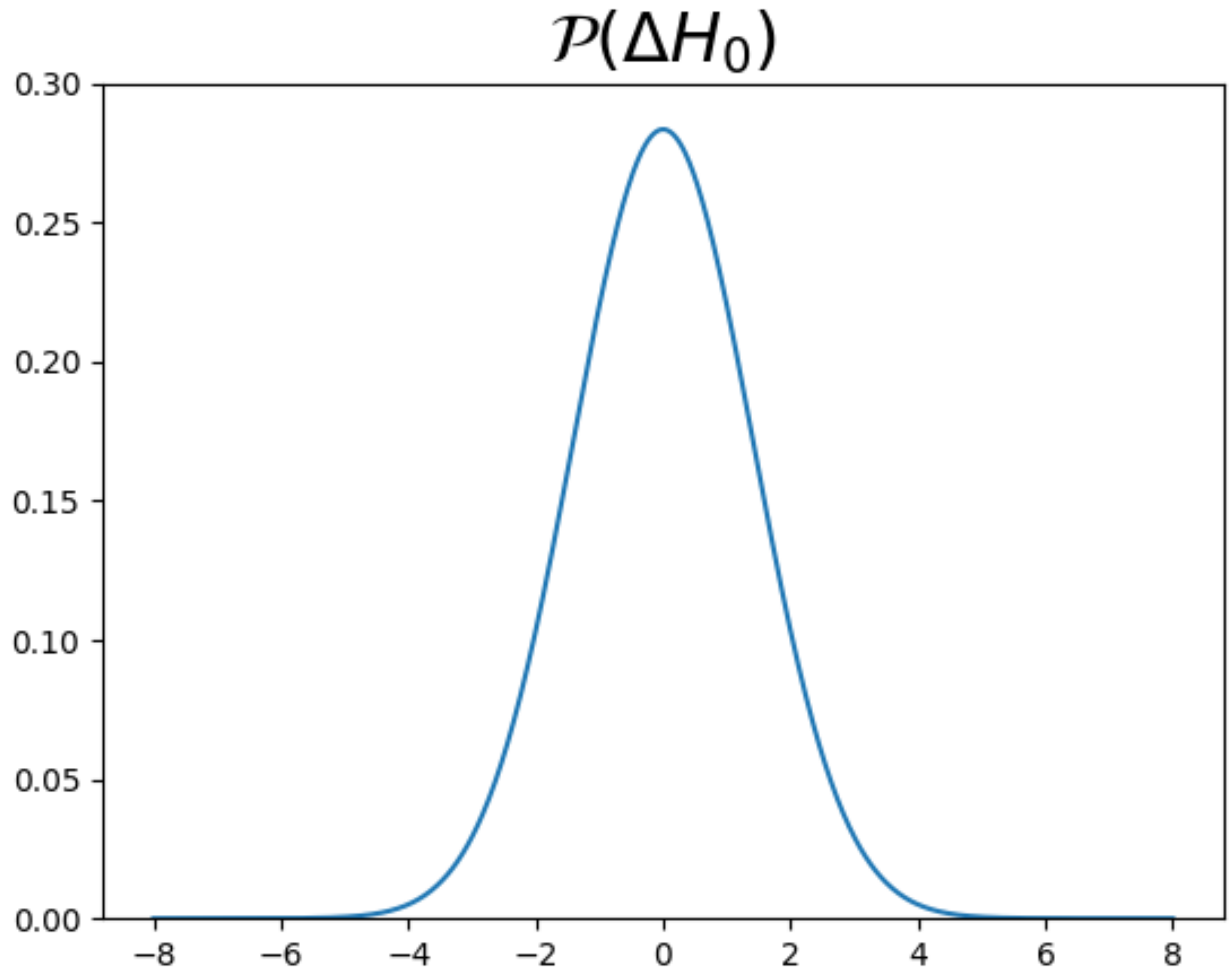
Gaussian : Any linear combination of independent Gaussian random variables follows a Gaussian distribution

With zero mean : Under our hypothesis that the two measurements are unbiased

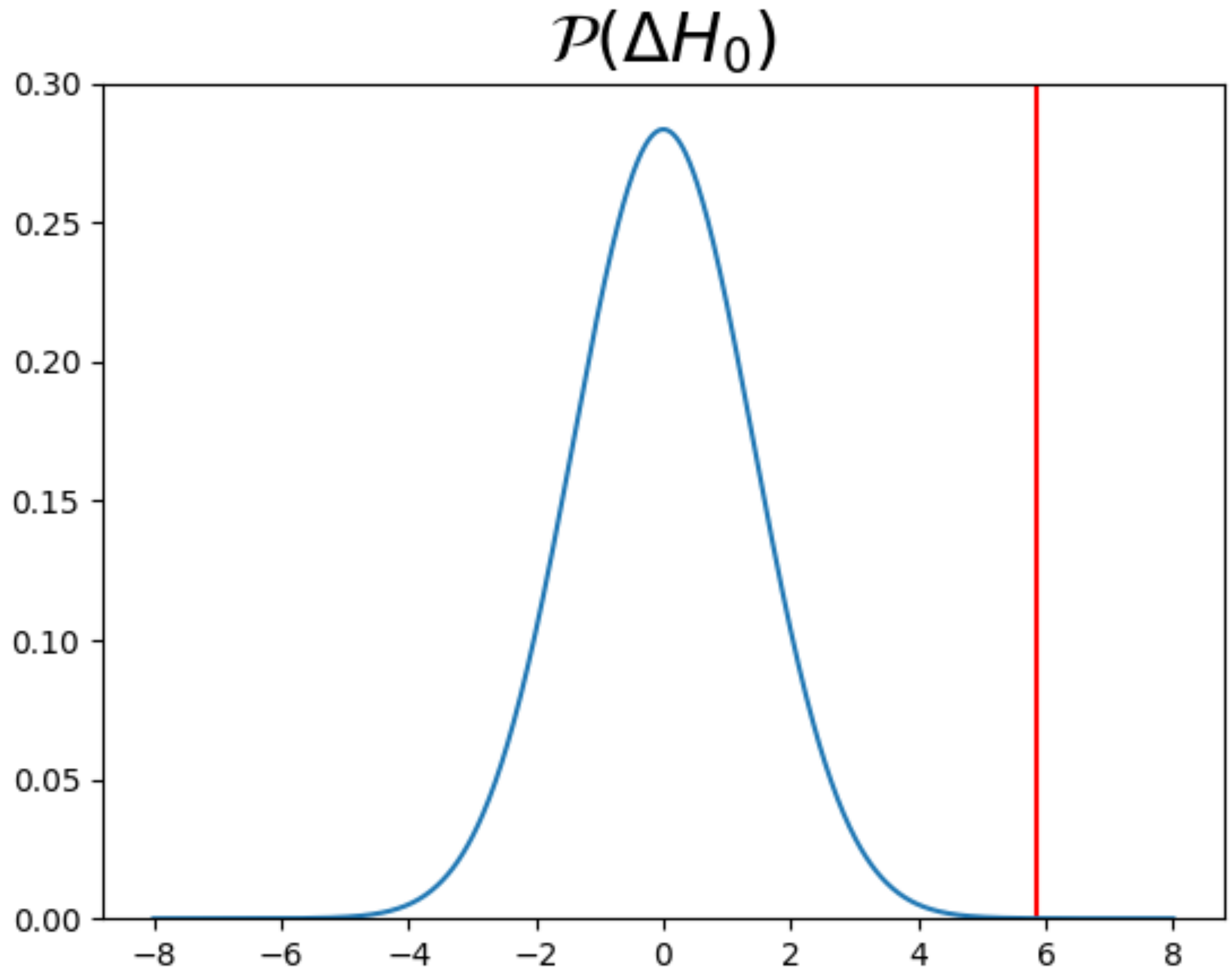
And variance : $\sigma^2(\Delta H_0) = \sigma^2(H_0^{\text{Riess}}) + \sigma^2(H_0^{\text{Planck}})$

$$\boxed{\mathcal{P}(\Delta H_0) = \mathcal{G}(0, \sigma^2(\Delta H_0))}$$

Example I: The Hubble constant Problem

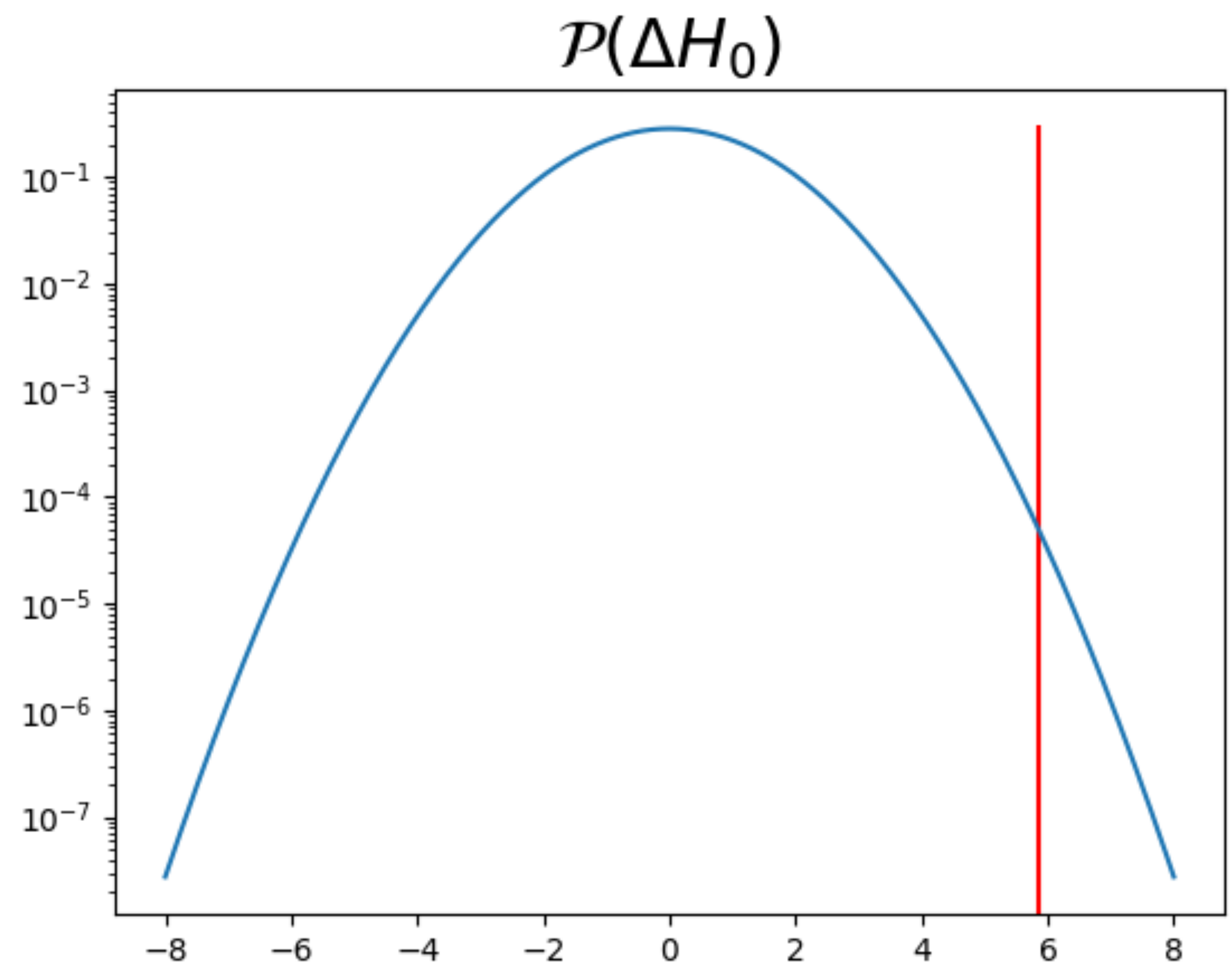


Example I: The Hubble constant Problem



$$
\begin{aligned}
H_0^{\mathrm{Planck}} &= 67.36 \pm 0.54 \ \mathrm{km/s/Mpc} \\
H_0^{\mathrm{Riess}} &= 73.2 \pm 1.3 \ \mathrm{km/s/Mpc}
\end{aligned}
$$

Example I: The Hubble constant Problem



Example I: The Hubble constant Problem

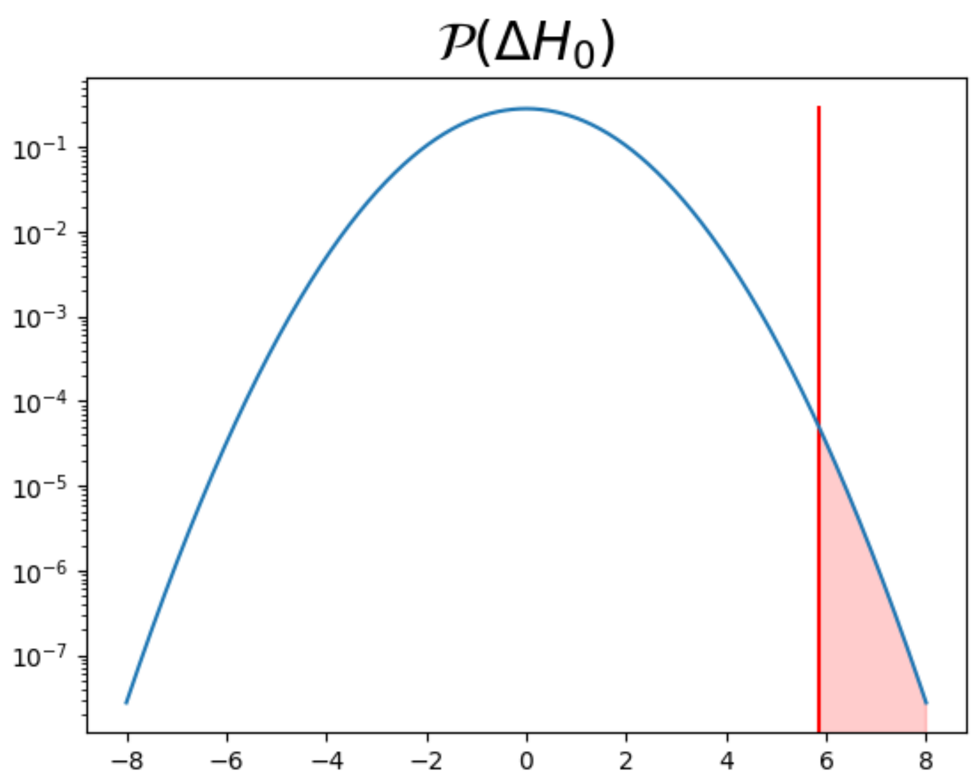
$$\begin{aligned} H_0^{\text{Planck}} &= 67.36 \pm 0.54 \text{ km/s/Mpc} \\ H_0^{\text{Riess}} &= 73.2 \pm 1.3 \text{ km/s/Mpc} \end{aligned}$$

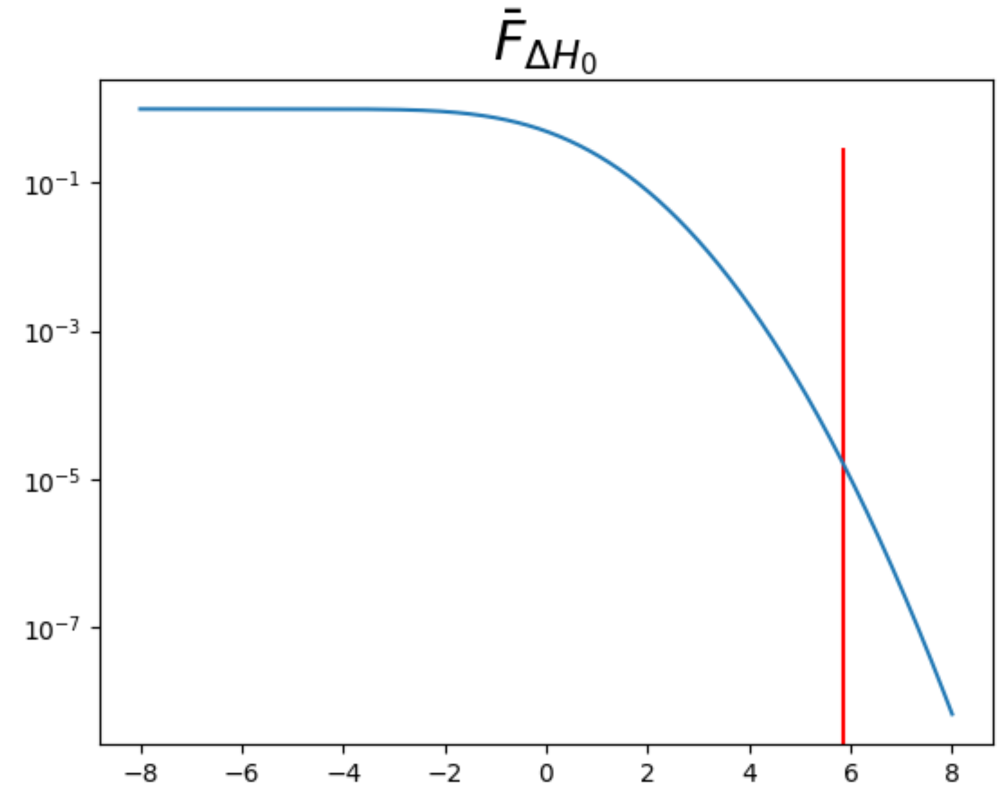
Are these two measurements in agreement or in disagreement?

What is the probability, under the null hypothesis (that the two measurements agree), of obtaining a value more extreme than the one observed?

What is the probability, under the null hypothesis, of obtaining a value ***more extreme*** *than the observed one?*

$$\bar{F}_{\Delta H_0}(x) = P(\Delta H_0 > x) \quad = \quad 1 - \int_{-\infty}^{x} \frac{1}{\sigma_{\Delta H_0}\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma^2_{\Delta H_0}}\right) dx'$$
$$= \quad 1 - \frac{1}{2}\left(1 + \text{erf}\ \left(x/\sqrt{2}\sigma_{\Delta_{H_0}}\right)\right)$$



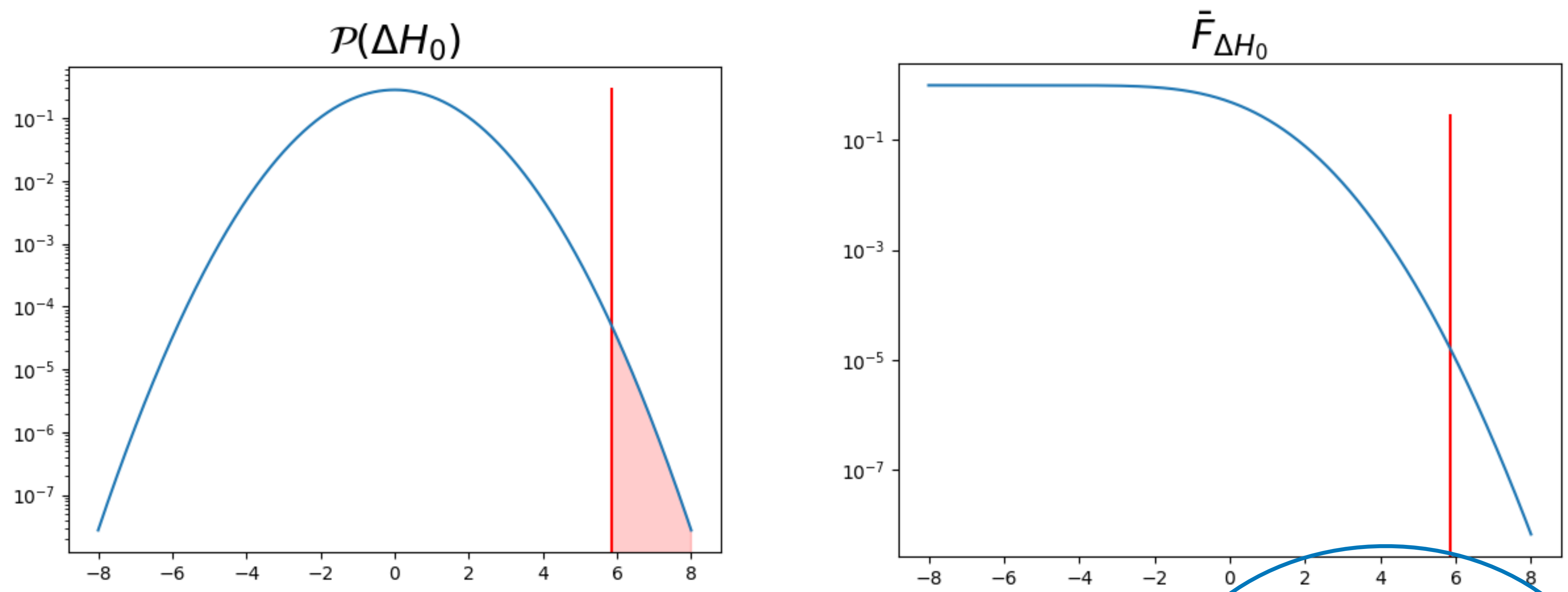


$$P(\Delta H_0 > 5.84) \quad = \quad 1.67 \times 10^{-5}$$

What is the probability, under the null hypothesis, of obtaining a value ***more extreme*** *than the observed one?*

$$\bar{F}_{\Delta H_0}(x) = P(\Delta H_0 > x) \;=\; 1 - \int_{-\infty}^{x} \frac{1}{\sigma_{\Delta H_0}\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma^2_{\Delta H_0}}\right) dx'$$

$$= 1 - \frac{1}{2}\left(1 + \mathrm{erf}\;\left(x/\sqrt{2}\sigma_{\Delta_{H_0}}\right)\right)$$

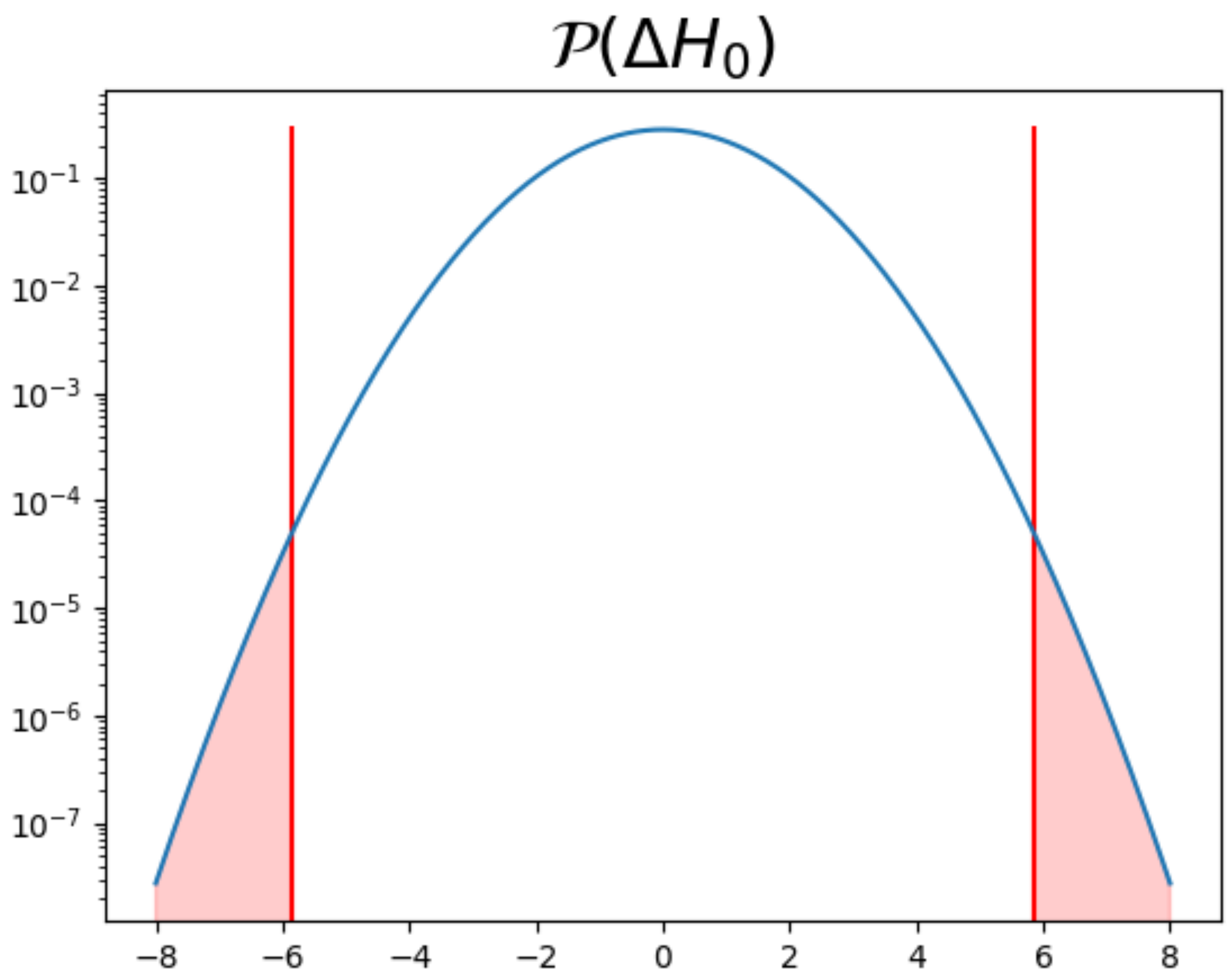


$$P(\Delta H_0 > 5.84) \;=\; 1.67 \times 10^{-5}$$

Tail distribution
Exceedance
One sided p-value
Survival function
Reliability function

What is the probability, under the null hypothesis, of obtaining a value ***more extreme*** *than the observed one?*

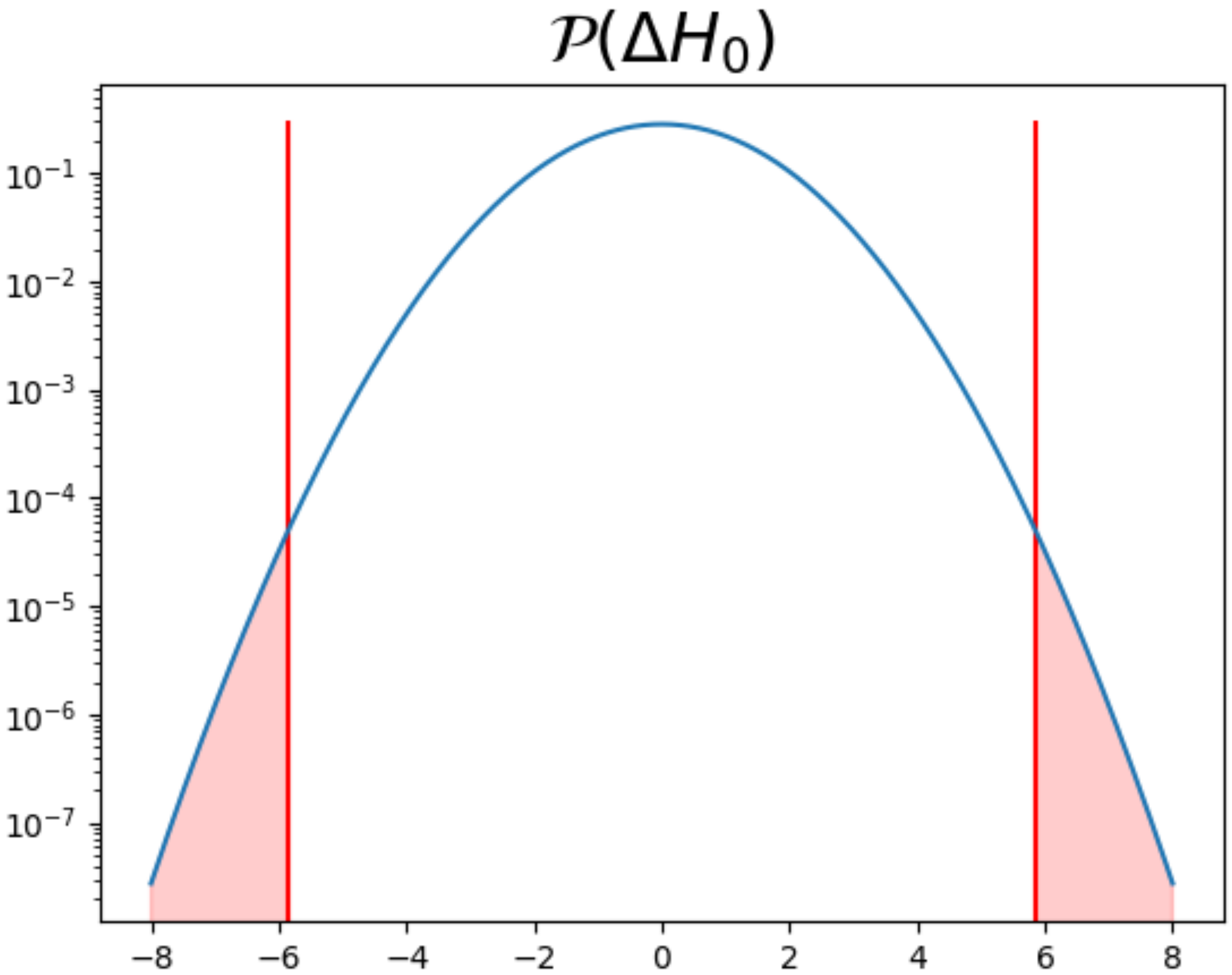
The test could have fluctuated in either direction, so "more extreme » should be understood in terms of the absolute value.



$$P(|\Delta H_0| > 5.84) \quad = \quad 3.34 \times 10^{-5}$$

*What is the probability, under the null hypothesis, of obtaining a value **more extreme** than the observed one?*

The test could have fluctuated in either direction, so "more extreme » should be understood in terms of the absolute value.



$$P(|\Delta H_0| > 5.84) \quad = \quad 3.34 \times 10^{-5}$$

The null hypothesis is therefore extremely unlikely

- **New physics ??**
- **Problem with one of the two measurements ?**

*What is the probability, under the null hypothesis, of obtaining a value **more extreme** than the observed one?*

We have detailed this for pedagogical reasons, but a very simple algorithm can be used.

$$
\begin{aligned}
H_0^{\text{Planck}} &= 67.36 \pm 0.54 \ \text{km/s/Mpc} \\
H_0^{\text{Riess}} &= 73.2 \pm 1.3 \ \text{km/s/Mpc}
\end{aligned}
$$

1) Calculate the number of sigmas between the two measured values

$$
N_\sigma = \frac{|H_0^{\text{Riess}} - H_0^{\text{Planck}}|}{\sqrt{\sigma^2(H_0^{\text{Riess}}) + \sigma^2(H_0^{\text{Planck}})}}
$$

2) The probability for two measurements to be N_σ apart is given by

$$
\begin{aligned}
P(|\Delta H_0 \,|/\sigma(\Delta H_0) > N_\sigma) &= 1 - \text{erf}\left(\frac{N_\sigma}{\sqrt{2}}\right) \\
&= 3.34 \times 10^{-5}
\end{aligned}
$$

χ^2 statistics

Chi-squared statistics is very commonly used in cosmology
It is notably used to assess the goodness of fit of a model to observed data, the statistical consistency of a dataset, and is also employed in the context of model selection.

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{d_i - m_i}{\sigma_i} \right)^2$$

Model

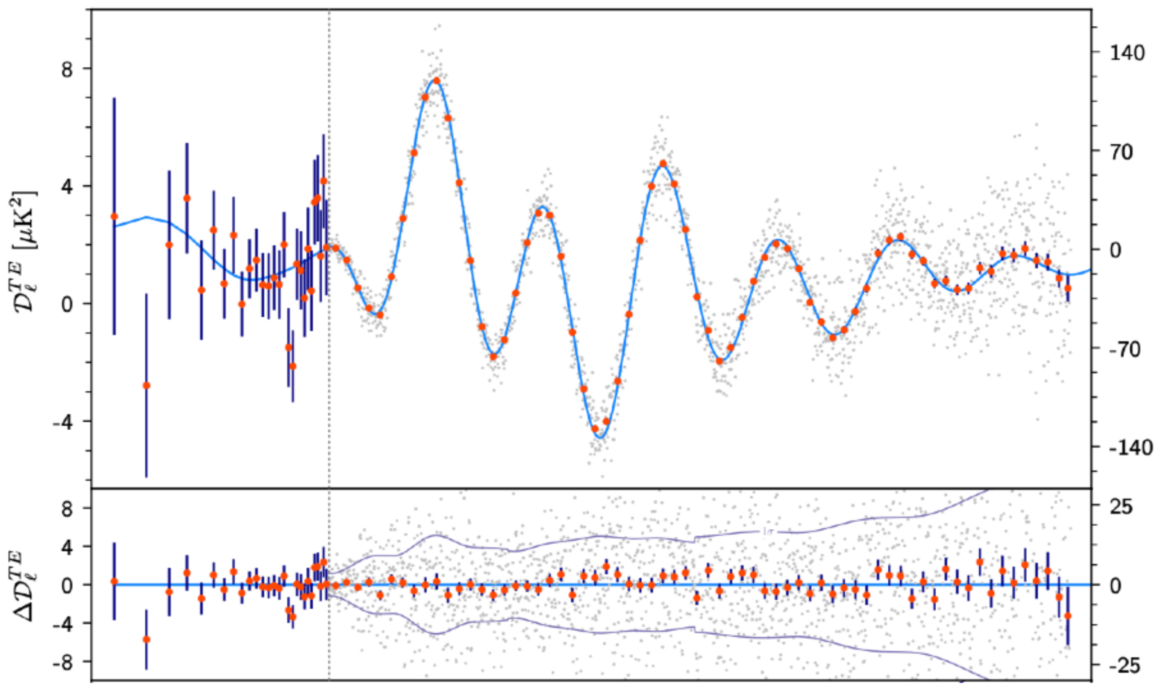
Data

Error

In the case where the measurements are not independent, a covariance matrix will be used:

$$\chi^2 = (\boldsymbol{d} - \boldsymbol{m})^T \boldsymbol{C}^{-1} (\boldsymbol{d} - \boldsymbol{m})$$

Example II: Interpretation of Chi-squared Tests

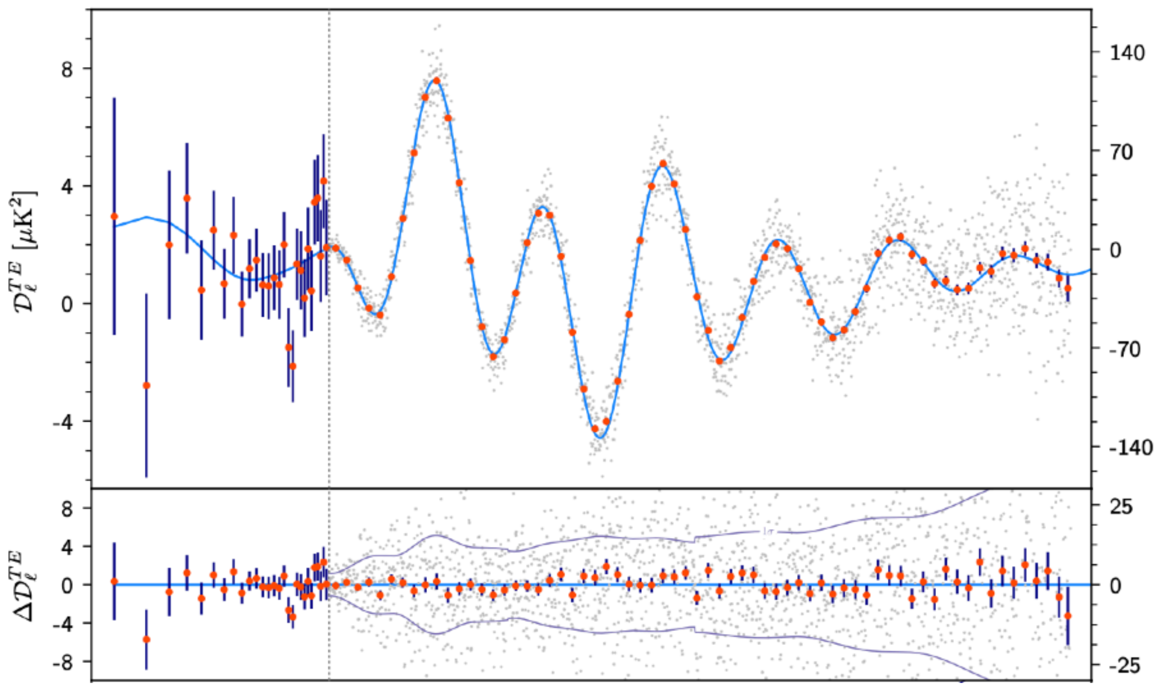


A&A 641, A5 (2020)
https://doi.org/10.1051/0004-6361/201936386
© Planck Collaboration 2020

		$\log(\mathcal{L})$	χ^2_{eff}	N_{dof}
TE, full, binned	30–1996	−428.68	857.36	762

‘Is the ΛCDM model a good fit to the Planck TE power spectrum data?’

Example II: Interpretation of Chi-squared Tests



A&A 641, A5 (2020)
https://doi.org/10.1051/0004-6361/201936386
© Planck Collaboration 2020

		$\log(\mathcal{L})$	$\chi^2_{\rm eff}$	$N_{\rm dof}$
TE, full, binned	30–1996	−428.68	857.36	762

‘Is the ΛCDM model a good fit to the Planck TE power spectrum data?’

*What is the probability, under the null hypothesis (that the ΛCDM model is correct), of obtaining a value **more extreme** than the observed chi-squared value for the Planck TE power spectrum?*

Example II: Interpretation of Chi-squared Tests

The probability distribution of the value of a χ^2 statistic wit k degrees of freedom is the distribution of the sum of the squares of k Gaussian random variables

$$\mathcal{P}_{\chi^2}(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} e^{-x/2} x^{k/2-1}$$

We will use our famous

Tail distribution
Exceedance
One sided p-value
Survival function
Reliability function

and compute

$$P(\chi^2 > 857.36; k = 762)$$

Example II: Interpretation of Chi-squared Tests

The probability distribution of the value of a χ^2 statistic wit k degrees of freedom is the distribution of the sum of the squares of k Gaussian random variables

$$\mathcal{P}_{\chi^2}(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} e^{-x/2} x^{k/2-1}$$

We will use our famous

Tail distribution
Exceedance
One sided p-value
Survival function
Reliability function

and compute

$$P(\chi^2 > 857.36; k = 762)$$

```
thibautlouis@mbp-de-thibaut ~ % ipython
Python 3.7.6 (default, Dec 22 2019, 01:09:06)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.12.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: from scipy.stats import chi2

In [2]: print(1 - chi2.cdf(857.36, 762))
0.009025337969493563
```

→ ***p ~ 1%***

Example II: Interpretation of Chi-squared Tests

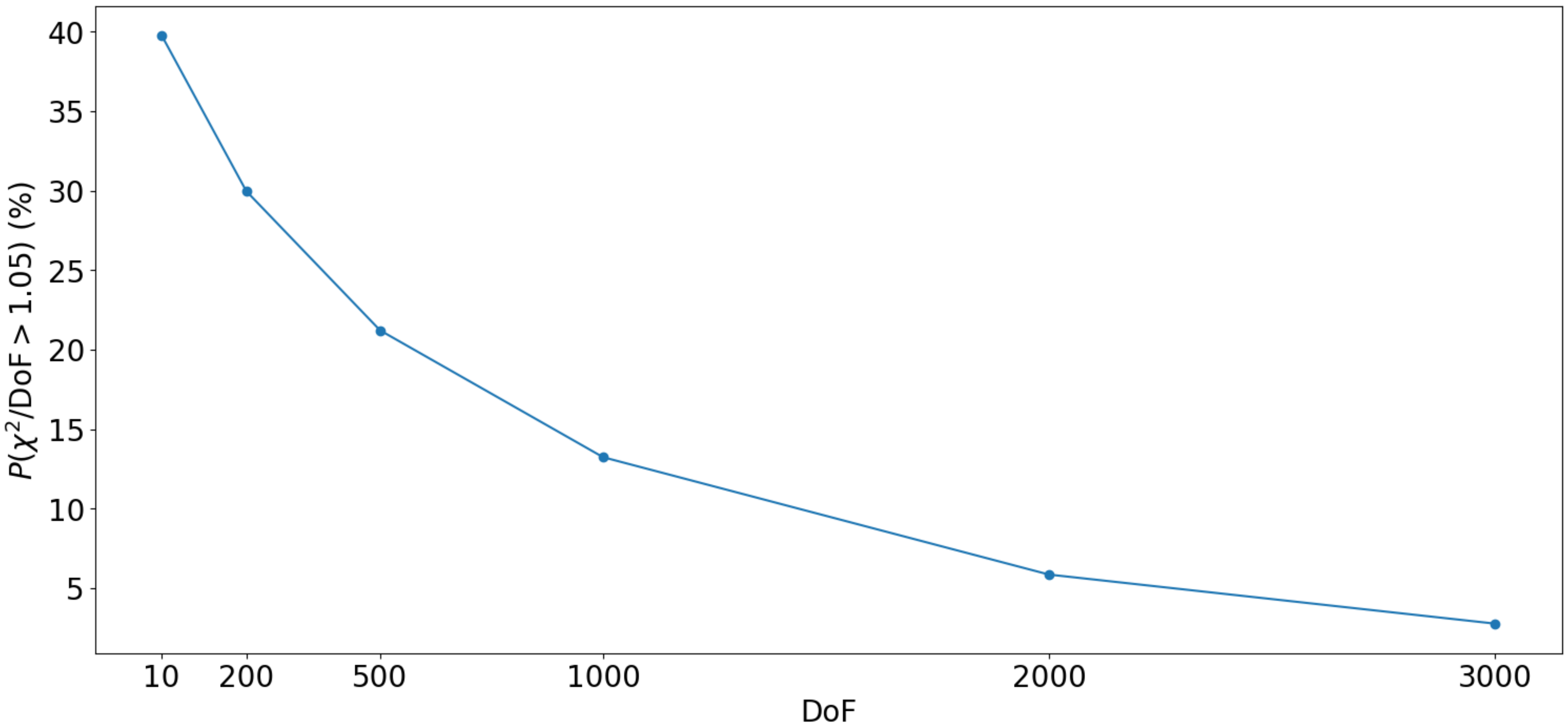
We have therefore answered our question:

*What is the probability, under the null hypothesis (that the ΛCDM model is correct), of obtaining a value **more extreme** than the observed chi-squared value for the Planck TE power spectrum?*

p ~ 1%

χ^2/DoF

A commonly used criterion is that the reduced chi-squared values should be *close* to 1. « Close" is highly dependent on the number of degrees of freedom considered.



For 10 degrees of freedom, it is quite likely to have a reduced chi-squared greater than 1.05. This is much less likely for 3000 degrees of freedom

Jackknifes

In cosmology χ^2 tests are also used in the context of *jackknifes* (also called *null tests*), a set of tests in which the data are split according to various criteria to verify the stability of the result:

$$\chi^2_{\rm null} = \left(\frac{d_{\rm first-half} - \ d_{\rm second-half}}{\sigma(\Delta d)} \right)^2$$

TABLE 3
Null tests using custom maps (PA1, PA2)

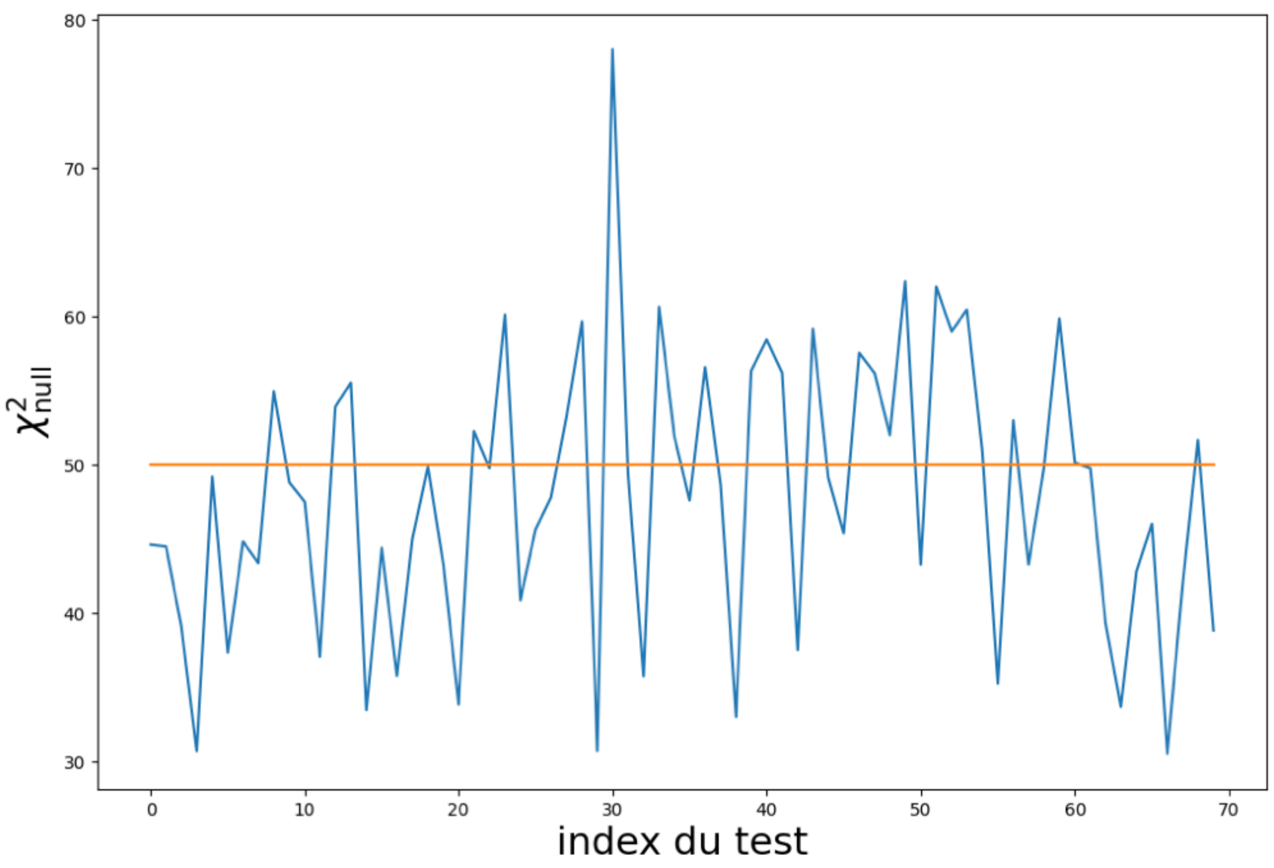
Test	Spectrum	PA1 χ^2/dof	PA1 P.T.E	PA2 χ^2/dof	PA2 P.T.E
Scan pattern 1 v Scan pattern 2: (0-1)x(2-3)	TT	0.82	0.83	1.00	0.47
	EE	0.91	0.66	0.72	0.94
	TE	0.99	0.49	0.80	0.85
	TB	1.13	0.25	0.86	0.76
	EB	1.15	0.21	0.93	0.61
	BB	0.66	0.97	0.83	0.81
Scan pattern 1 v Scan pattern 2: (0-3)x(1-2)	TT	1.13	0.24	1.19	0.17
	EE	0.67	0.97	1.12	0.25
	TE	0.99	0.50	0.83	0.80
	TB	0.85	0.77	0.81	0.84
	EB	0.95	0.58	0.98	0.53
	BB	0.96	0.55	0.75	0.91
Detectors: Fast v slow	TT	0.98	0.51	0.89	0.69
	EE	0.78	0.88	0.72	0.94
	TE	0.94	0.59	0.87	0.74
	TB	1.07	0.34	0.78	0.88
	EB	0.81	0.84	0.68	0.96
	BB	1.02	0.42	1.00	0.48
PWV: High v low	TT	0.99	0.49	1.18	0.18
	EE	0.84	0.78	0.90	0.68
	TE	0.72	0.94	0.71	0.94
	TB	0.75	0.91	0.77	0.89
	EB	0.98	0.52	0.96	0.56
	BB	0.65	0.98	0.94	0.60
Pick up:	TT	1.14	0.22	0.94	0.61
	EE	0.83	0.81	0.64	0.98
	TE	0.87	0.74	0.88	0.72
	TB	0.83	0.80	0.95	0.58
	EB	0.64	0.98	0.95	0.58
	BB	1.00	0.47	0.83	0.81
Moon: more aggressive cut	TT	0.82	0.82	1.08	0.32
	EE	1.40	0.03	1.18	0.17
	TE	1.30	0.07	0.68	0.97
	TB	0.92	0.64	0.91	0.66
	EB	1.01	0.45	0.96	0.55
	BB	0.90	0.67	1.22	0.13
Wafers: Hex1+hex3 v hex 2+semis	TT			1.02	0.44
	EE			1.08	0.33
	TE			1.29	0.07
	TB			0.59	0.99
	EB			1.03	0.42
	BB			0.54	0.99

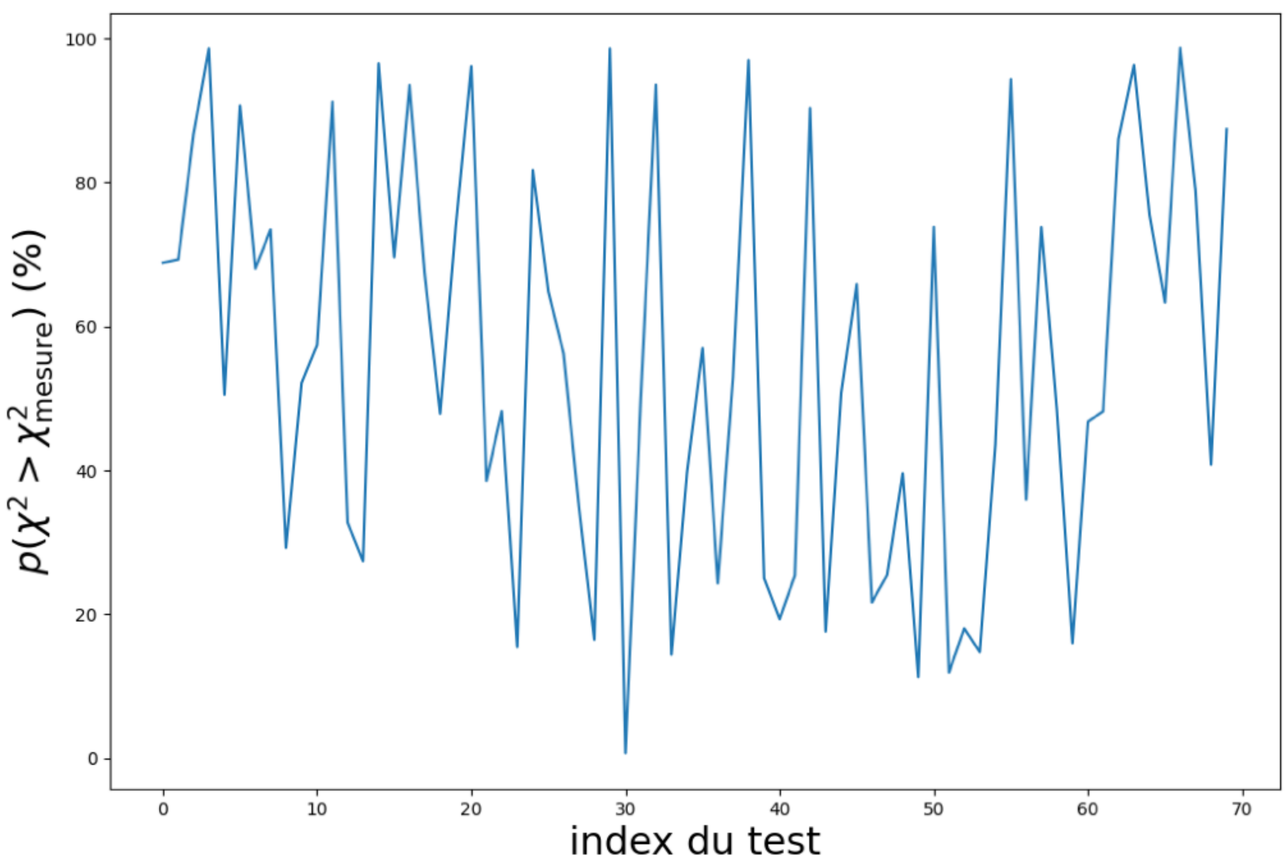
https://arxiv.org/pdf/1610.02360.pdf

Look elsewhere effect

Let us imagine conducting a large number of tests on a dataset.

We plot the χ^2 values and their associated probabilities

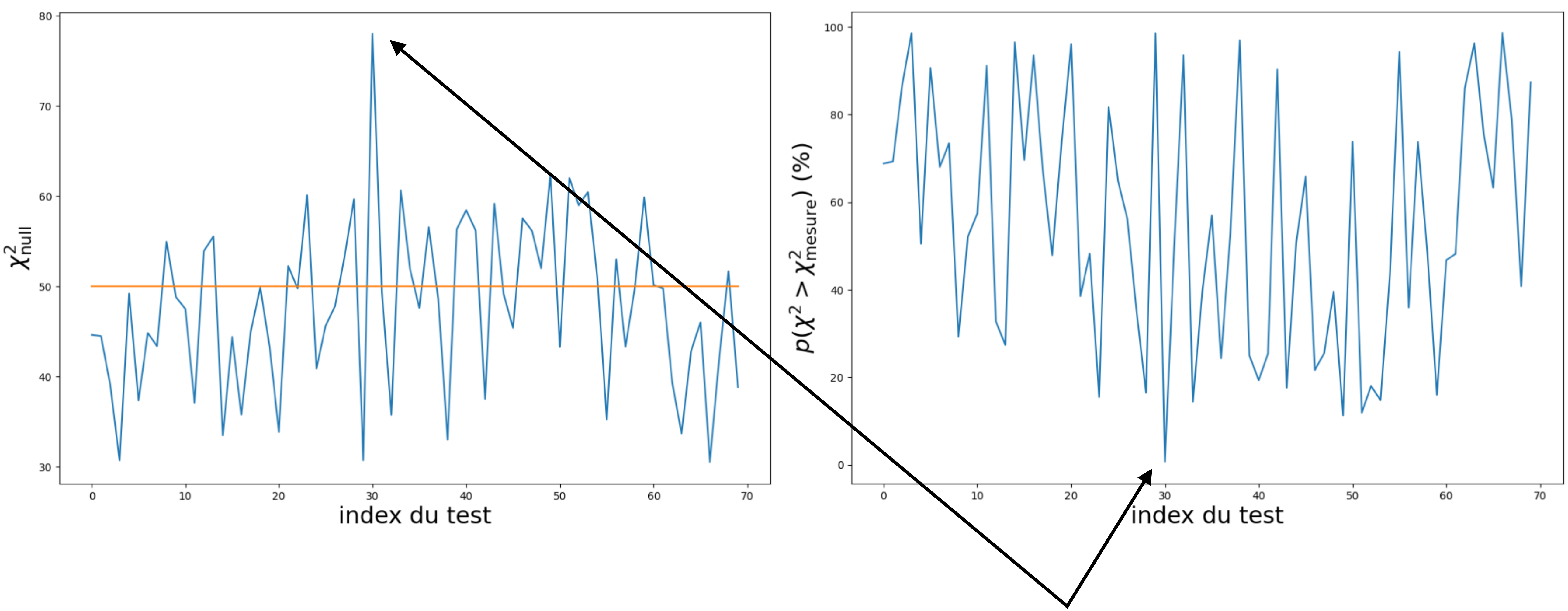




Look elsewhere effect

Let us imagine conducting a large number of tests on a dataset.

We plot the χ^2 values and their associated probabilities



P = 0.5%, have we detected a problem ?

Look elsewhere effect

(multiple comparisons, **multiplicity** or **multiple testing problem)**

We performed 70 tests. What is the probability that at least one of the tests has a p-value less than 0.5%?

Look elsewhere effect

(multiple comparisons, **multiplicity** or **multiple testing problem)**

We performed 70 tests. What is the probability that at least one of the tests has a p-value less than 0.5%?

To calculate it, let's break down the different possibilities:

1) **a null hypothesis test has a probability greater than 0.5%**
2) **a null hypothesis test has a probability smaller than 0.5%**

Look elsewhere effect

(multiple comparisons, **multiplicity** or **multiple testing problem)**

We performed 70 tests. What is the probability that at least one of the tests has a p-value less than 0.5%?

To calculate it, let's break down the different possibilities:

1) **a null hypothesis test has a probability greater than 0.5%**
2) **a null hypothesis test has a probability smaller than 0.5%**

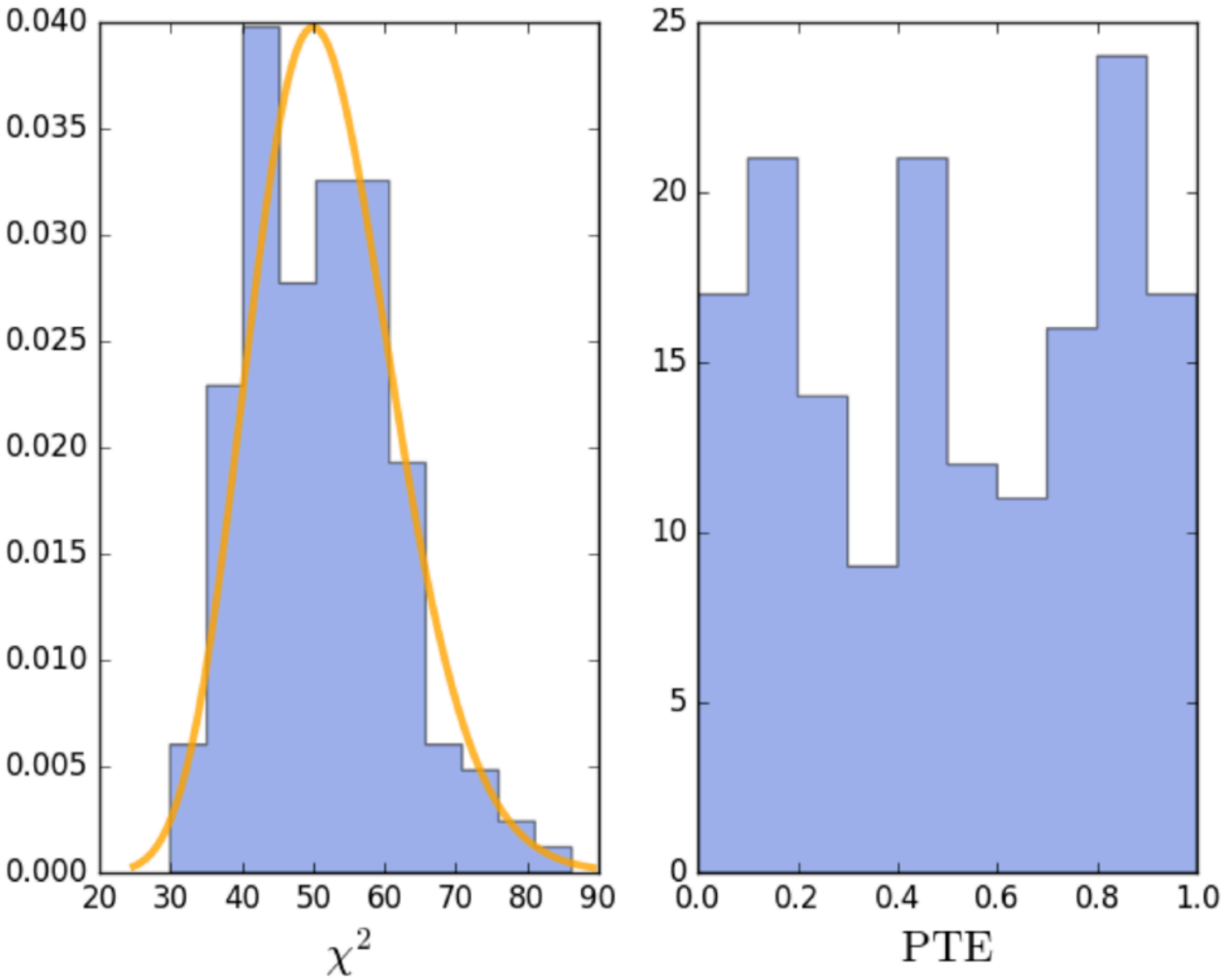
The probability of obtaining a p-value < 0.5% in 70 tests (hold on tight) is therefore calculated using the binomial distribution.

$$p = 1 - (0.995)^{N_{\rm test}} = 30\%$$

This effect has been the cause of several false detections in physics.

Look elsewhere effect

(multiple comparisons, multiplicity or **multiple testing problem)**



arXiv:2007.07289

As a general rule, one always wants to analyze the result of a test in the context of the number of tests performed. For example, one can compare the histogram of the tests conducted with the expected distribution.

Model Comparison

An example: The H0 Olympics, a fair ranking of proposed models

https://arxiv.org/pdf/2107.10291.pdf

Is the fit of a particular physical model M significantly better than ΛCDM?

Akaike Information Criterion (AIC)

$$\Delta \mathrm{AIC} = \chi^2_{\min,\mathcal{M}} - \chi^2_{\min,\Lambda\mathrm{CDM}} + 2(N_{\mathcal{M}} - N_{\Lambda\mathrm{CDM}})$$

Model Comparison

An example: The H0 Olympics, a fair ranking of proposed models

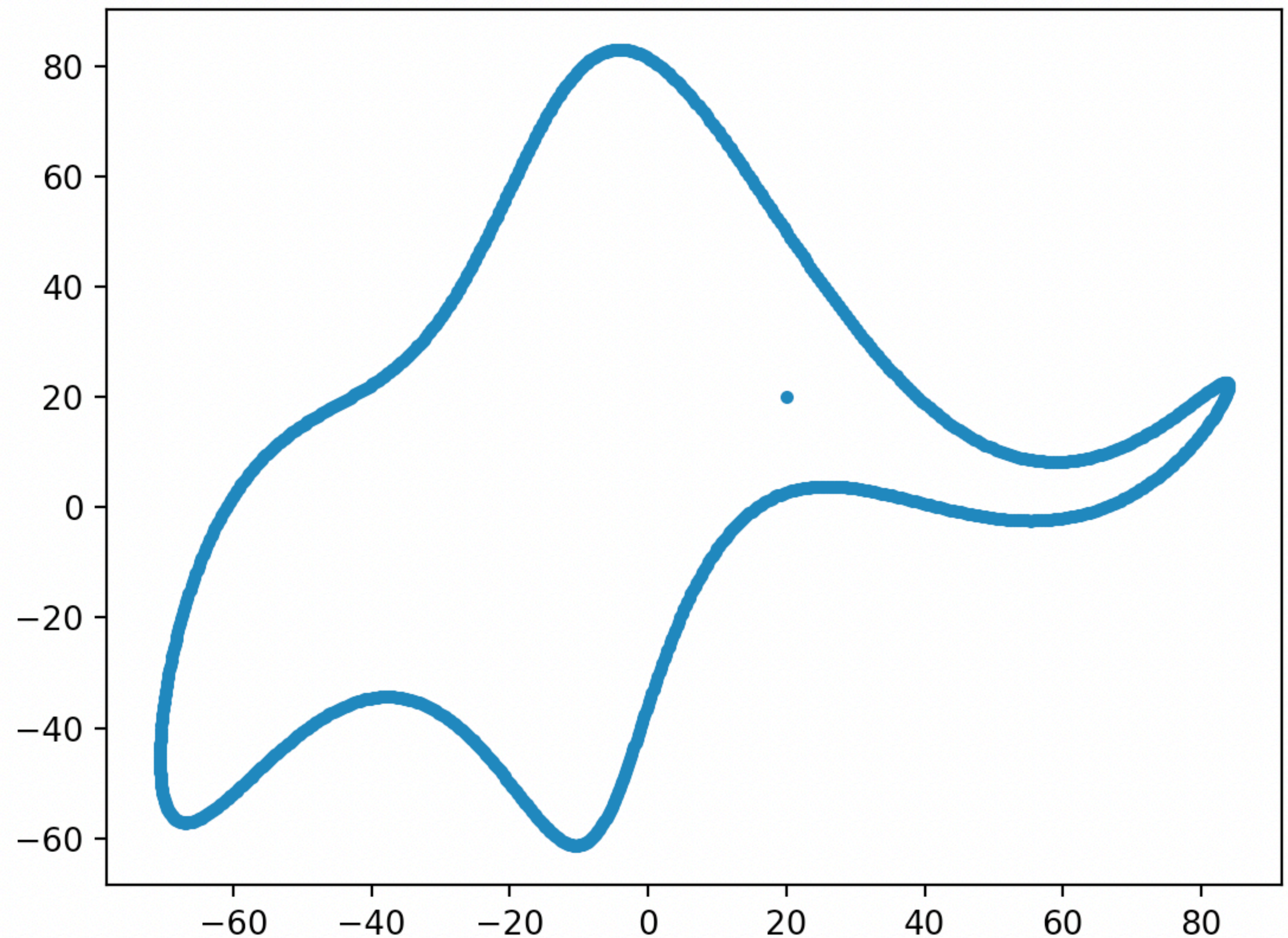
https://arxiv.org/pdf/2107.10291.pdf

Is the fit of a particular physical model M significantly better than ΛCDM?

Model	$\Delta N_{\rm param}$	M_B	Gaussian Tension	$Q_{\rm DMAP}$ Tension		$\Delta\chi^2$	ΔAIC		Finalist
ΛCDM	0	-19.416 ± 0.012	4.4σ	4.5σ	✗	0.00	0.00	✗	✗
$\Delta N_{\rm ur}$	1	-19.395 ± 0.019	3.6σ	3.9σ	✗	-4.60	-2.60	✗	✗
SIDR	1	-19.385 ± 0.024	3.2σ	3.6σ	✗	-3.77	-1.77	✗	✗
DR-DM	2	-19.413 ± 0.036	3.3σ	3.4σ	✗	-7.82	-3.82	✗	✗
mixed DR	2	-19.388 ± 0.026	3.2σ	3.7σ	✗	-6.40	-2.40	✗	✗
SIν+DR	3	-19.440 ± 0.038	3.7σ	3.9σ	✗	-3.56	2.44	✗	✗
Majoron	3	-19.380 ± 0.027	3.0σ	2.9σ	✓	-13.74	-7.74	✓	✓ 2
primordial B	1	-19.390 ± 0.018	3.5σ	3.5σ	✗	-10.83	-8.83	✓	✓ 3
varying m_e	1	-19.391 ± 0.034	2.9σ	3.2σ	✗	-9.87	-7.87	✓	✓ 3
varying m_e+Ω_k	2	-19.368 ± 0.048	2.0σ	1.7σ	✓	-16.11	-12.11	✓	✓ 1
EDE	3	-19.390 ± 0.016	3.6σ	1.6σ	✓	-20.80	-14.80	✓	✓ 2
NEDE	3	-19.380 ± 0.021	3.2σ	2.0σ	✓	-17.70	-11.70	✓	✓ 2
CPL	2	-19.400 ± 0.016	3.9σ	4.1σ	✗	-4.23	-0.23	✗	✗
PEDE	0	-19.349 ± 0.013	2.7σ	2.0σ	✓	4.76	4.76	✗	✗
MPEDE	1	-19.400 ± 0.022	3.6σ	4.0σ	✗	-2.21	-0.21	✗	✗
DM $\to$ DR+WDM	2	-19.410 ± 0.013	4.2σ	4.4σ	✗	-4.18	-0.18	✗	✗
DM $\to$ DR	2	-19.410 ± 0.011	4.3σ	4.2σ	✗	0.11	4.11	✗	✗

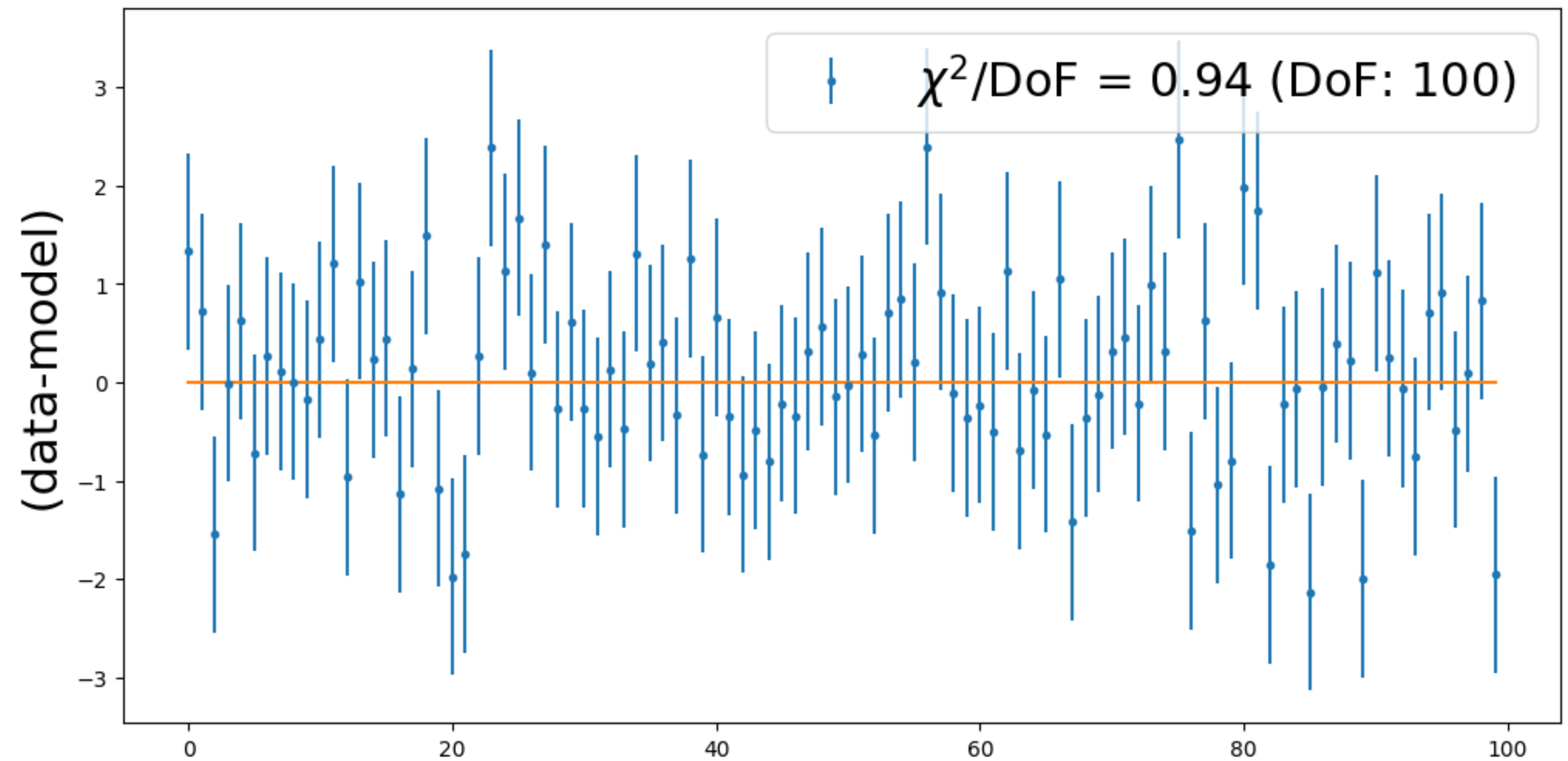
Interlude
Johny Von Neumann: With four parameters I can fit an elephant, and with five I can make him wiggle his trunk

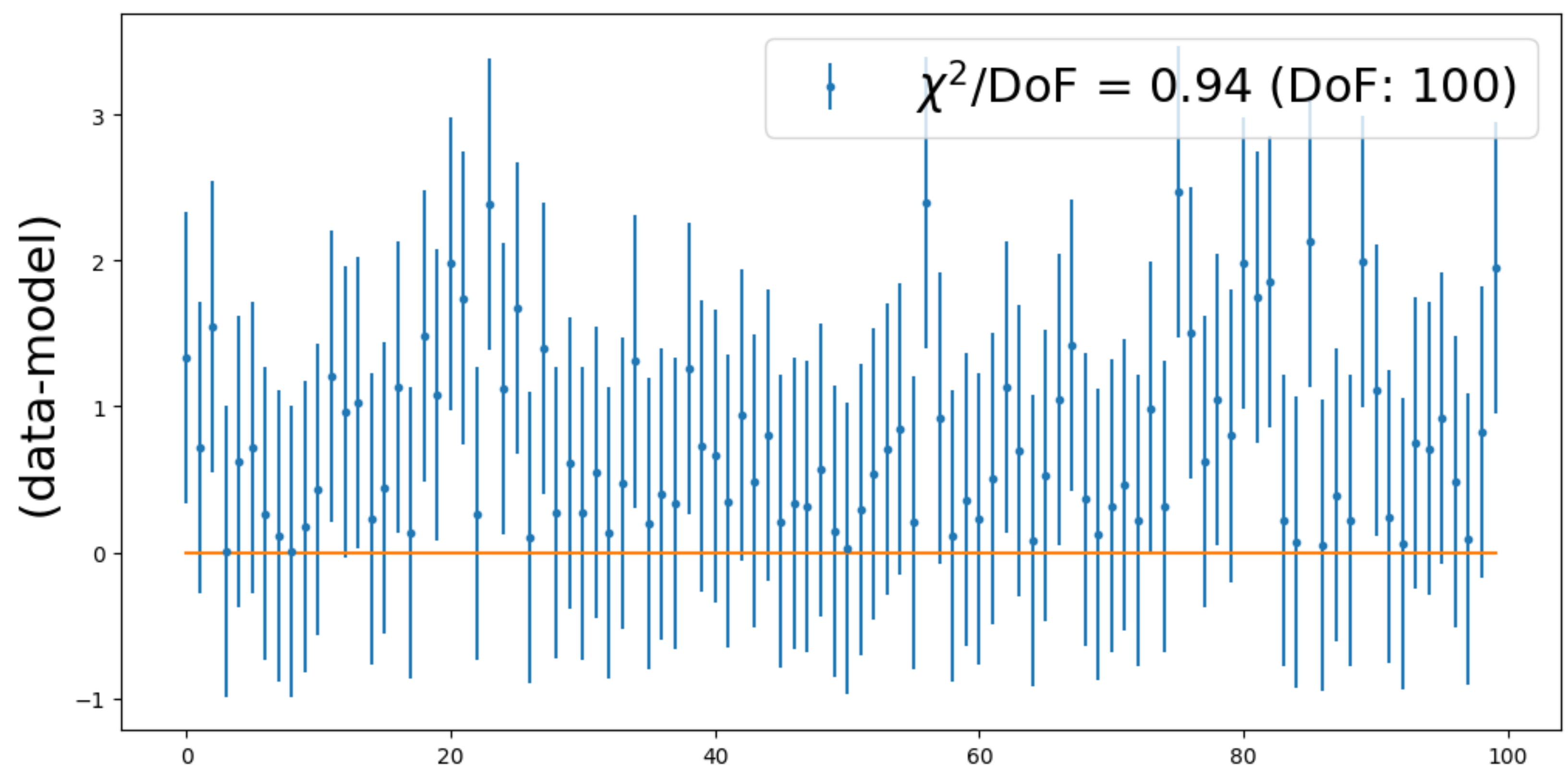
```python
"""
Author: Piotr A. Zolnierczuk (zolnierczukp at ornl dot gov)

Based on a paper by:
Drawing an elephant with four complex parameters
Jurgen Mayer, Khaled Khairy, and Jonathon Howard,
Am. J. Phys. 78, 648 (2010), DOI:10.1119/1.3254017
"""
import numpy as np
import pylab

# elephant parameters
p1, p2, p3, p4 = (50 - 30j, 18 +  8j, 12 - 10j, -14 - 60j )
p5 = 40 + 20j # eyepiece

def fourier(t, C):
    f = np.zeros(t.shape)
    A, B = C.real, C.imag
    for k in range(len(C)):
        f = f + A[k]*np.cos(k*t) + B[k]*np.sin(k*t)
    return f

def elephant(t, p1, p2, p3, p4, p5):
    npar = 6
    Cx = np.zeros((npar,), dtype='complex')
    Cy = np.zeros((npar,), dtype='complex')

    Cx[1] = p1.real*1j
    Cx[2] = p2.real*1j
    Cx[3] = p3.real
    Cx[5] = p4.real

    Cy[1] = p4.imag + p1.imag*1j
    Cy[2] = p2.imag*1j
    Cy[3] = p3.imag*1j

    x = np.append(fourier(t,Cx), [-p5.imag])
    y = np.append(fourier(t,Cy), [p5.imag])

    return x,y

x, y = elephant(np.linspace(0,2*np.pi,1000), p1, p2, p3, p4, p5)
pylab.plot(y,-x,'.')
pylab.show()
```



https://gwang.umn.edu/story/2020/03/20/How-to-fit-an-elephant.html

χ^2 *is not enough*



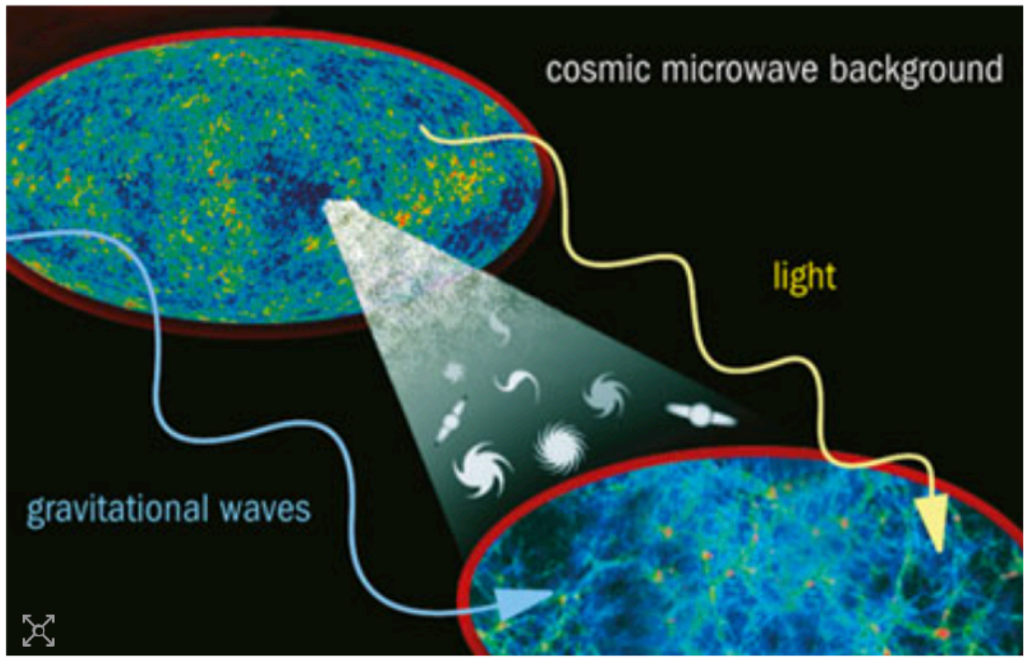


A visual inspection of the residuals between data and model is always very useful.

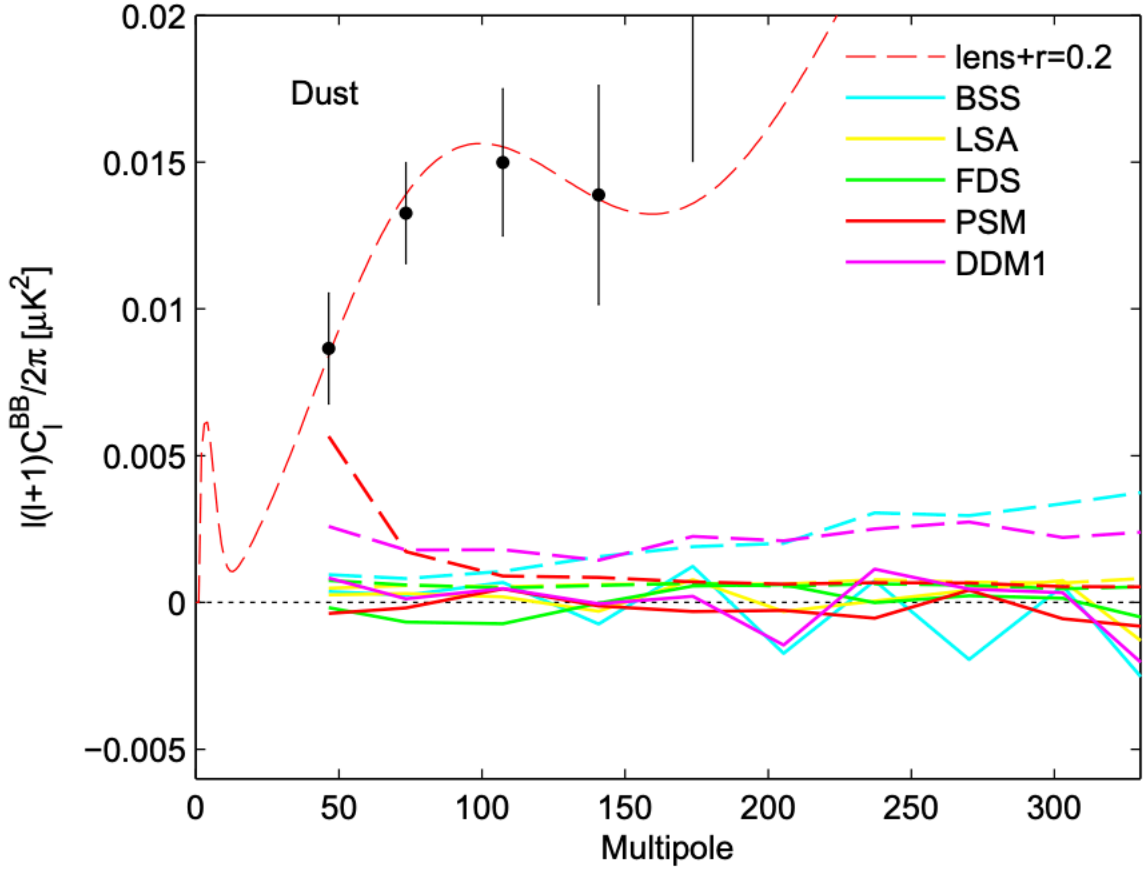
The devil you know and the one you don't know

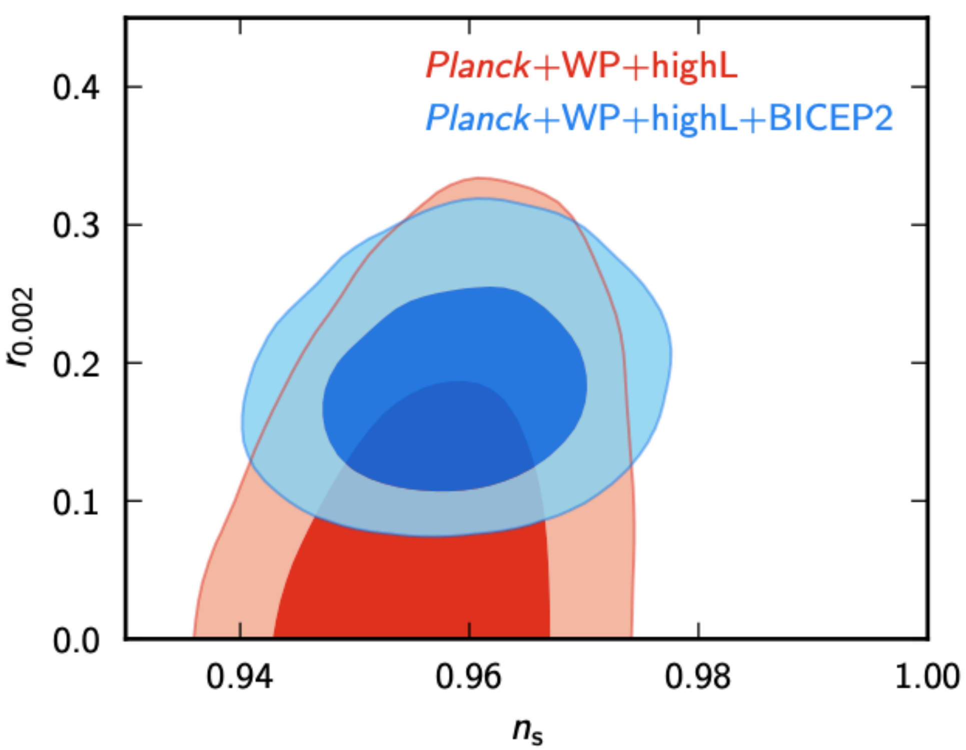
BICEP2 finds first direct evidence of cosmic inflation

17 Mar 2014 Tushna Commissariat



Cosmic evolution from the Big Bang to today





The BICEP2 team assumed a polarization fraction of 5% for dust in their field, based on a preliminary map presented at a conference [9]. A visual comparison of this map with the new version in [7] suggests that there is imperfect agreement between the two in many regions and that the polarization fractions are significantly higher in the new maps relative to the old ones. One reason for the discrepancy is the CIB, which was not corrected for in the old maps; since CIB is not polarized correcting for it reduces intensity but not polarization, increasing the polarization fraction.

Multivariate Probability Functions

Joint Probability Function

We will often have to study problems described by a large number of parameters (a large number of random variables (X1 ,…, XN).

We will therefore introduce the joint probability density:

$$\mathcal{P}_{X_1,..,X_N}(x_1,...,x_N)$$

$$\int dx_1 \; ... \; dx_N \mathcal{P}_{X_1,..,X_N}(x_1,...,x_N) = 1$$

We will have the joint cumulative distribution function (CDF).

$$F_{X_1,..,X_N}(x_1,...,x_N) = P(X_1 \leq x_1,...,X_N \leq x_N)$$

Useful Definitions I: Marginal (Distribution)

The marginal probability distribution is defined as

$$\mathcal{P}_{X_i}(x_i) = \int dx_1 \;...\; dx_{i-1} dx_{i+1} \;...\; dx_N \mathcal{P}_{X_1,\;...\;,X_N}(x_1,\;...\;,x_N)$$

We can also marginalize over a subset of random variables.

$$\mathcal{P}_{X_1,\;...\;,X_i}(x_1, ..., x_i) = \int dx_{i+1} \;...\; dx_N \mathcal{P}_{X_1,\;...\;,X_N}(x_1,\;...\;,x_N)$$

For independent random variables, the joint probability distribution is the product of the marginals.

$$\mathcal{P}_{X_1,..,X_N}(x_1, ..., x_N) = \prod_{i=1}^{N} \mathcal{P}_{X_i}(x_i)$$

Useful Definitions II: Conditional (Distribution)

The conditional probability distribution is defined as

$$\mathcal{P}_{X_i|X_1, \ldots X_{i-1}, X_{i+1}, \ldots, X_N}(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$$
$$= \frac{\mathcal{P}_{X_1, \ldots, X_N}(x_1, \ldots, x_N)}{\mathcal{P}_{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N}(x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_N)}$$

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

We can also condition a subset of random variables on another subset.

$$\mathcal{P}_{X_1 \ldots X_i \ |X_{i+1}, \ldots, X_N}(x_1, \ldots, x_i|x_{i+1}, .., x_N) = \frac{\mathcal{P}_{X_1, \ldots, X_N}(x_1, \ldots, x_N)}{\mathcal{P}_{X_{i+1}, \ldots, X_N}(x_{i+1}, \ldots, x_N)}$$

Useful Definitions III: Covariance Matrix

The covariance matrix of a random vector is defined by

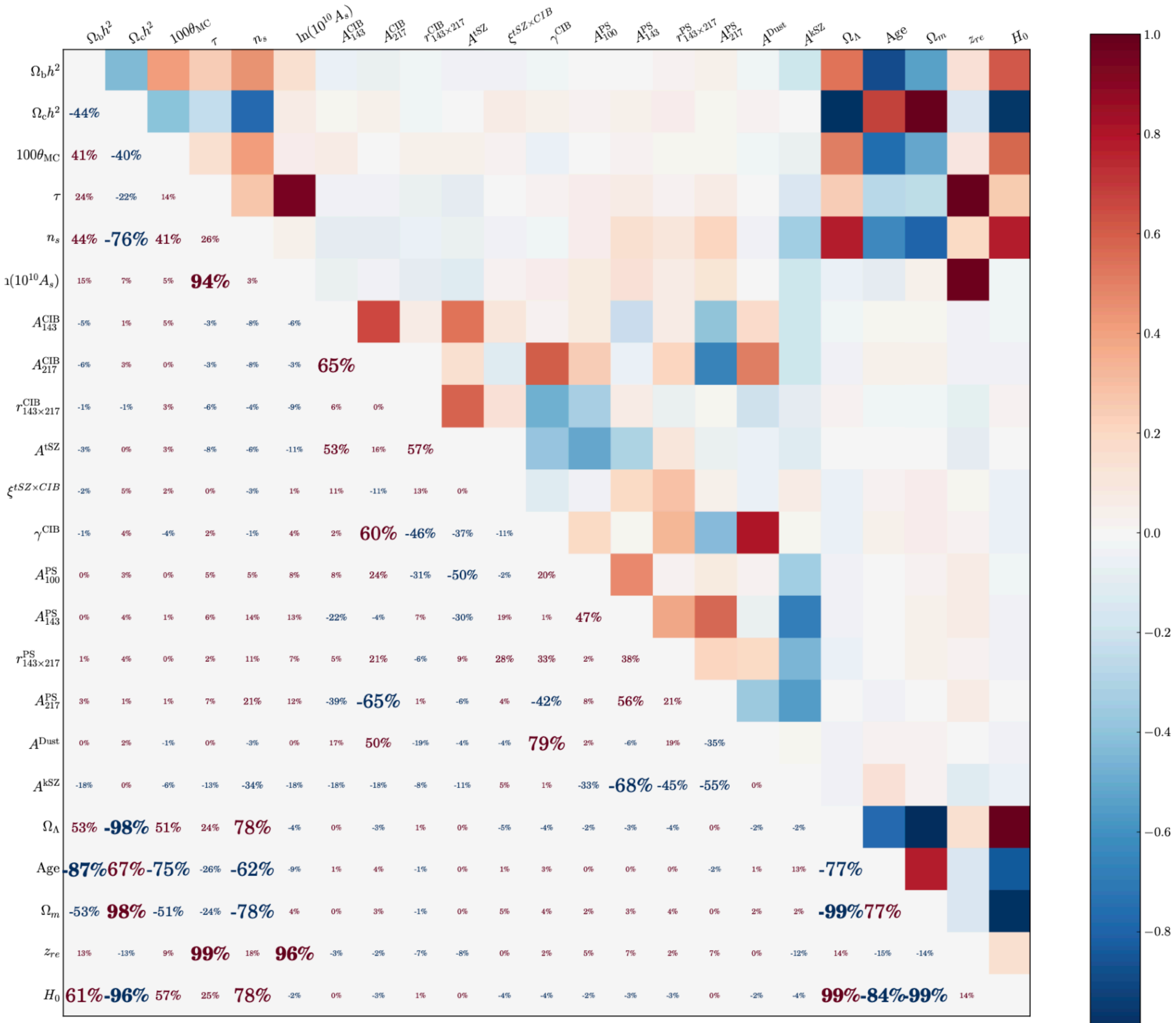
$$\Sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

The Pearson correlation coefficient (correlation matrix) is given by

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} \qquad -1 \leq \rho_{ij} \leq 1$$

Planck correlation matrix

https://arxiv.org/pdf/1303.5075.pdf



Two remarks :

Independence $\rightarrow$ **Uncorrelated**

$$
\begin{aligned}
\Sigma_{ij} &= \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \\
&= \int dx_i dx_j (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \mathcal{P}_{X_i, X_j}(x_i, x_j) \\
&= \int dx_i (x_i - \langle x_i \rangle) \mathcal{P}_{X_i}(x_i) \int dx_j (x_j - \langle x_j \rangle) \mathcal{P}_{X_j}(x_j) \\
&= 0
\end{aligned}
$$

Two remarks :

$$\begin{tikzcd} \textbf{Independence} \arrow[r, bend left] & \textbf{Uncorrelated} \arrow[l, bend left, "\times"] \end{tikzcd}$$

$$\langle x_i \rangle = 0$$

$$x_j = g(x_i)$$

Two remarks :

$$\textbf{Independence} \;\underset{\not\Longleftarrow}{\Longrightarrow}\; \textbf{Uncorrelated}$$

$$\langle x_i \rangle = 0$$

$$x_j = g(x_i)$$

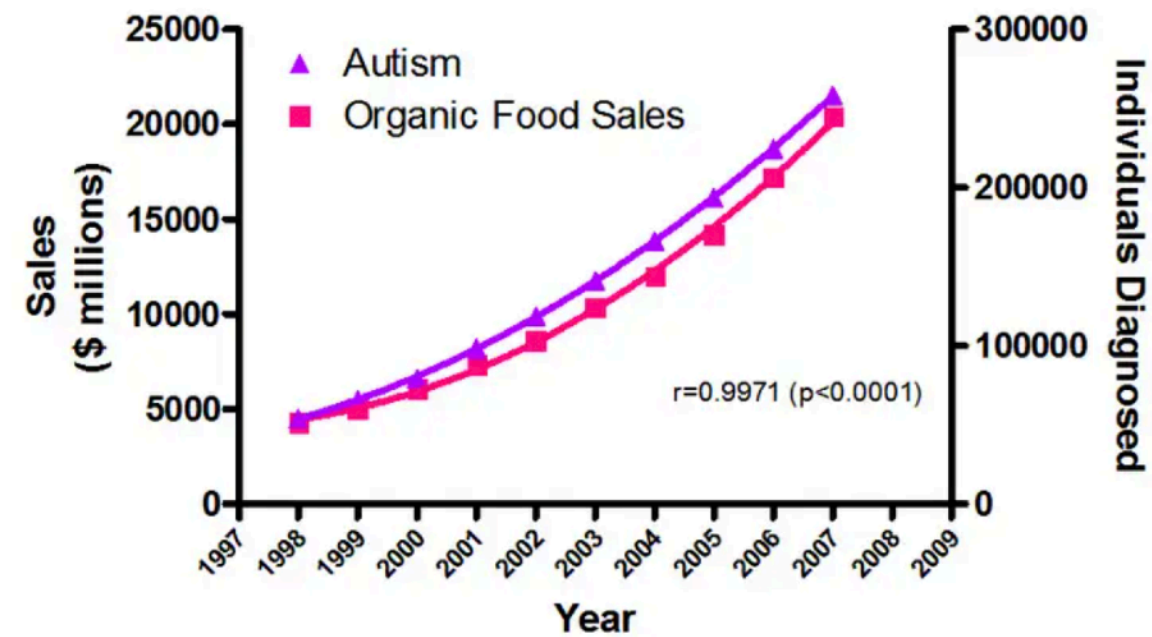
$$\begin{aligned}
\Sigma_{ij} &= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i,X_j}(x_i, x_j) \\
&= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i}(x_i) \mathcal{P}_{X_j|X_i}(x_j|x_i) \\
&= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i}(x_i) \delta(x_j - g(x_i)) \\
&= \int dx_i x_i g(x_i) \mathcal{P}_{X_i}(x_i)
\end{aligned}$$

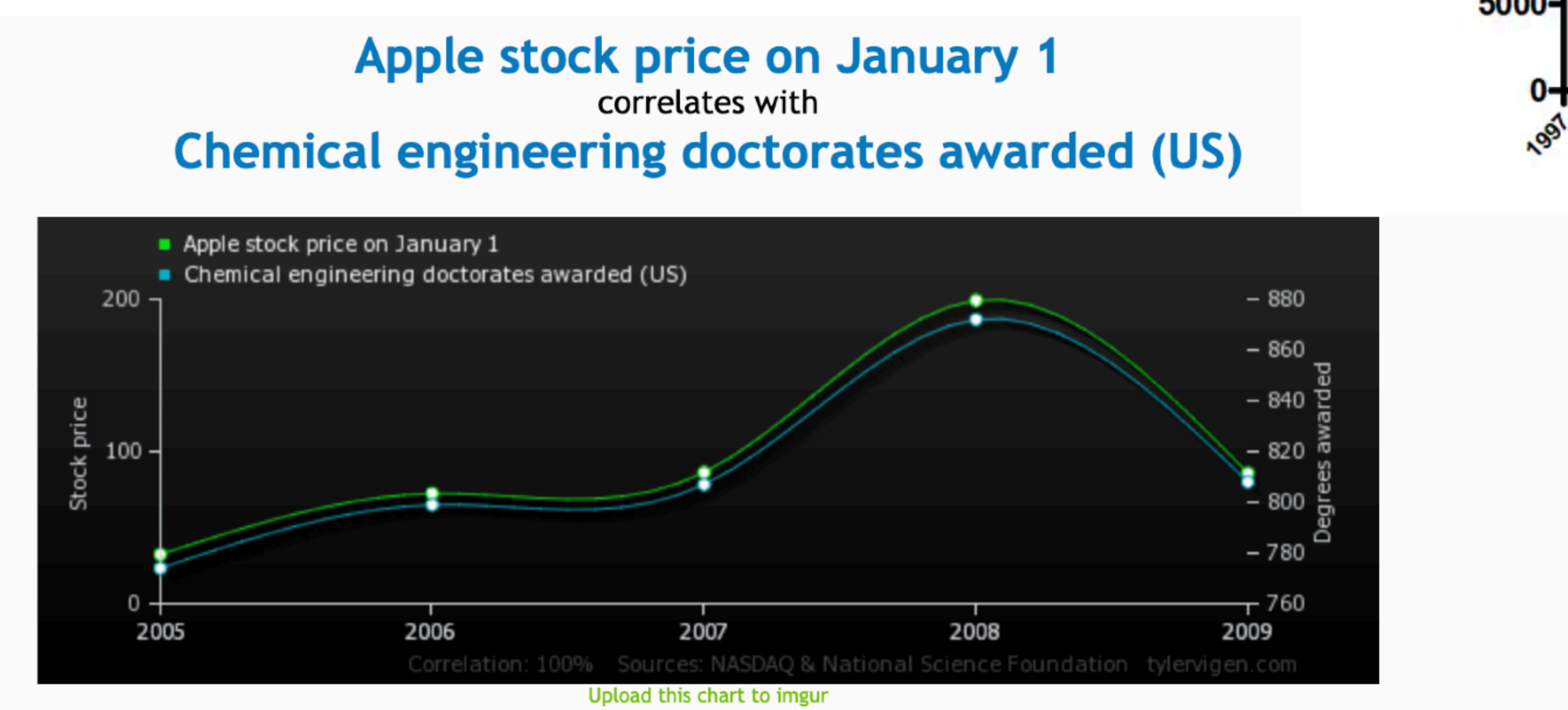
If g and P are even functions and the integration domain is symmetric around zero

$$\Sigma_{ij} = 0$$

Two remarks :

Correlation $\nrightarrow$ **Causality**





Multivariate Normal Distribution

A particularly useful multivariate probability distribution is the multivariate normal distribution.

$$\mathcal{P}_{X_1,..,X_N}(x_1, ..., x_N) = \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det\,\Sigma}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

It is notably the form of many likelihood functions in cosmology.

A particularly useful multivariate probability distribution is the multivariate normal distribution.

$$\mathcal{P}_{X_1,..,X_N}(x_1,...,x_N) = \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det\,\Sigma}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

It is notably the form of many likelihood functions in cosmology.

If the data take the form

$$d_i = M_i(\{\theta\}) + n_i$$

Model **Gaussian noise with mean 0 and covariance** $\mathcal{N}_{ij} = \langle n_i n_j \rangle$

$$\mathcal{L}(\boldsymbol{d}|\{\theta\}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det\,\mathcal{N}}}\exp\left[-\frac{1}{2}[\boldsymbol{d}-\boldsymbol{M}(\{\boldsymbol{\theta}\})]^T\mathcal{N}^{-1}[\boldsymbol{d}-\boldsymbol{M}(\{\boldsymbol{\theta}\})]\right]$$

To obtain the marginal distribution of a subset of normal random variables, one only needs to "remove" the variables over which we marginalize from the mean vector and the covariance matrix.

$$\mathcal{P}_{X_1,..,X_N}(x_1,...,x_N) = \mathcal{P}_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det\,\Sigma}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

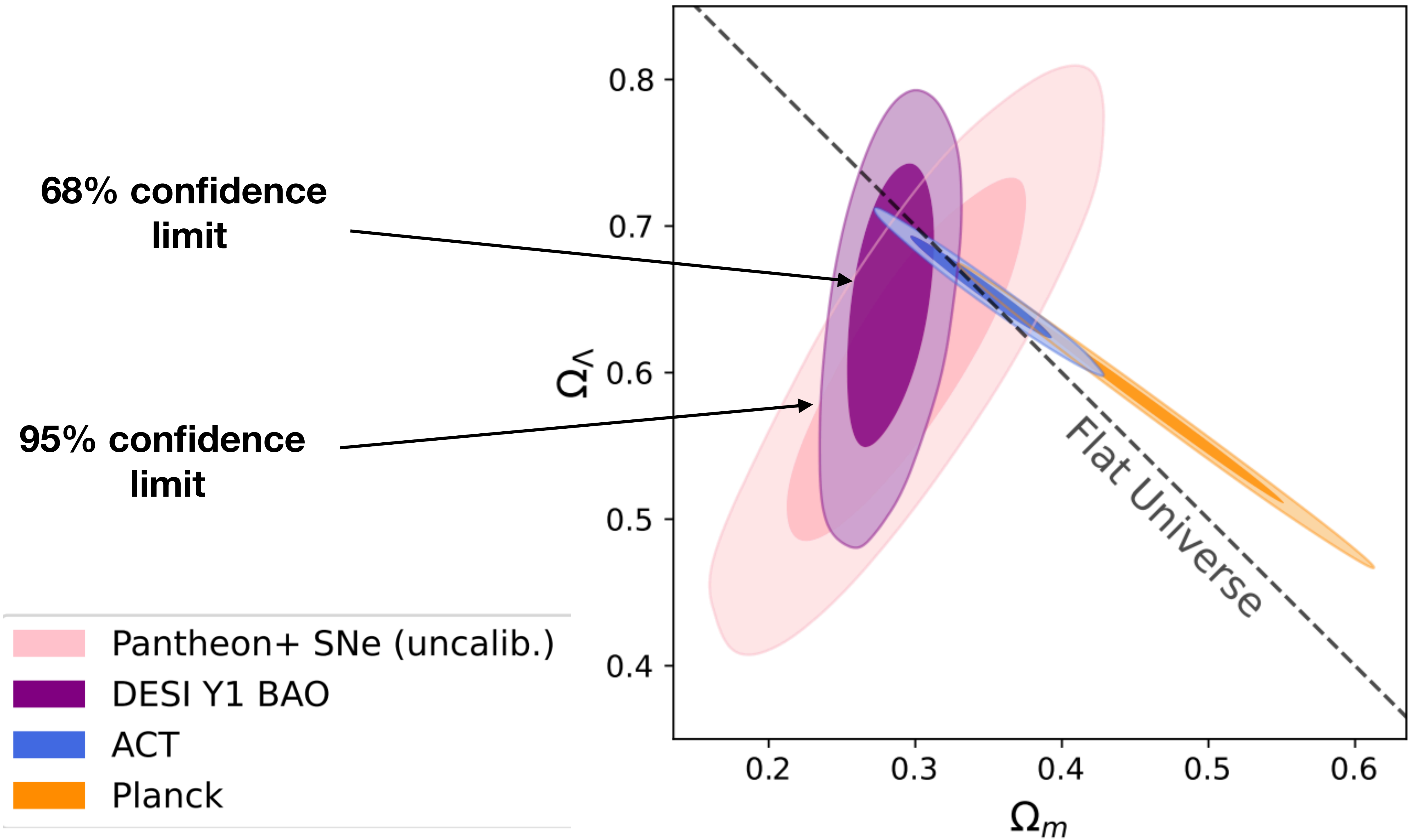
$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ .... \\ \mu_N \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & .... & \Sigma_{1N} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & .... & \Sigma_{2N} \\ .... & & & & \\ \Sigma_{N1} & \Sigma_{N2} & \Sigma_{N3} & .... & \Sigma_{NN} \end{pmatrix}$$

$$\downarrow$$

$$\begin{aligned} \mathcal{P}_{X_1,X_3}(x_1,x_3) &= \int dx_2 dx_4 \ldots dx_N \mathcal{P}_{X_1,\,\ldots\,,X_N}(x_1,\ \ldots\ ,x_N) \\ &= \frac{1}{2\pi\sqrt{\det\,\tilde{\Sigma}}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\tilde{\boldsymbol{\mu}})^T\tilde{\Sigma}^{-1}(\boldsymbol{x}-\tilde{\boldsymbol{\mu}})\right] \end{aligned}$$

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix} \qquad \tilde{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{pmatrix}$$

An application of marginalization: Visualization of ellipses



An application of marginalization: Visualization of ellipses

Let's write down the posterior distribution of cosmological parameters

$$\mathcal{P}(\Omega_m, \Omega_\Lambda, n_s, \tau, A_s, \Omega_b)$$

The marginal distribution for Omega matter and Omega Lambda is

$$\begin{aligned}\mathcal{P}(\Omega_m, \Omega_\Lambda) &= \int dn_s dA_s d\tau d\Omega_b \mathcal{P}(\Omega_m, \Omega_\Lambda, n_s, \tau, A_s, \Omega_b) \\ &= \frac{1}{2\pi\sqrt{\det \tilde{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x} - \tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1} (\boldsymbol{x} - \tilde{\boldsymbol{\mu}})\right]\end{aligned}$$

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \bar{\Omega}_m \\ \bar{\Omega}_\Lambda \end{pmatrix} \qquad \tilde{\Sigma} = \begin{pmatrix} \Sigma_{\Omega_m\Omega_m} & \Sigma_{\Omega_m\Omega_\Lambda} \\ \Sigma_{\Omega_\Lambda\Omega_m} & \Sigma_{\Omega_\Lambda\Omega_\Lambda} \end{pmatrix}$$

An application of marginalization: Visualization of ellipses

$$
\begin{aligned}
\mathcal{P}(\Omega_m, \Omega_\Lambda) &= \int dn_s dA_s d\tau d\Omega_b \mathcal{P}(\Omega_m, \Omega_\Lambda, n_s, \tau, A_s, \Omega_b) \\
&= \frac{1}{2\pi\sqrt{\det\, \tilde{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1}(\boldsymbol{x}-\tilde{\boldsymbol{\mu}})\right]
\end{aligned}
$$

The quantity $\chi^2 = (\boldsymbol{x}-\tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1}(\boldsymbol{x}-\tilde{\boldsymbol{\mu}})$

Follows a chi-squared distribution with two degree of freedom

We want to visualize the set of x values that have a probability greater than 5%, we need to find the value of y such that

$$P({\chi^2}_{\mathrm{2DoF}}(\boldsymbol{x}) < y) = 95\%$$

We use the Percent Point Function (also called the quantile function).

```
In [1]: from scipy.stats import *

In [2]: chi2.ppf(0.95,2)
Out[2]: 5.991464547107979

In [3]: chi2.ppf(0.68,2)
Out[3]: 2.27886856637673
```

← **95%** (for `chi2.ppf(0.95,2)`)

← **68%** (for `chi2.ppf(0.68,2)`)

x values with a probability greater than 5% will satisfy

$$(\boldsymbol{x} - \tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1} (\boldsymbol{x} - \tilde{\boldsymbol{\mu}}) < 5.9915$$

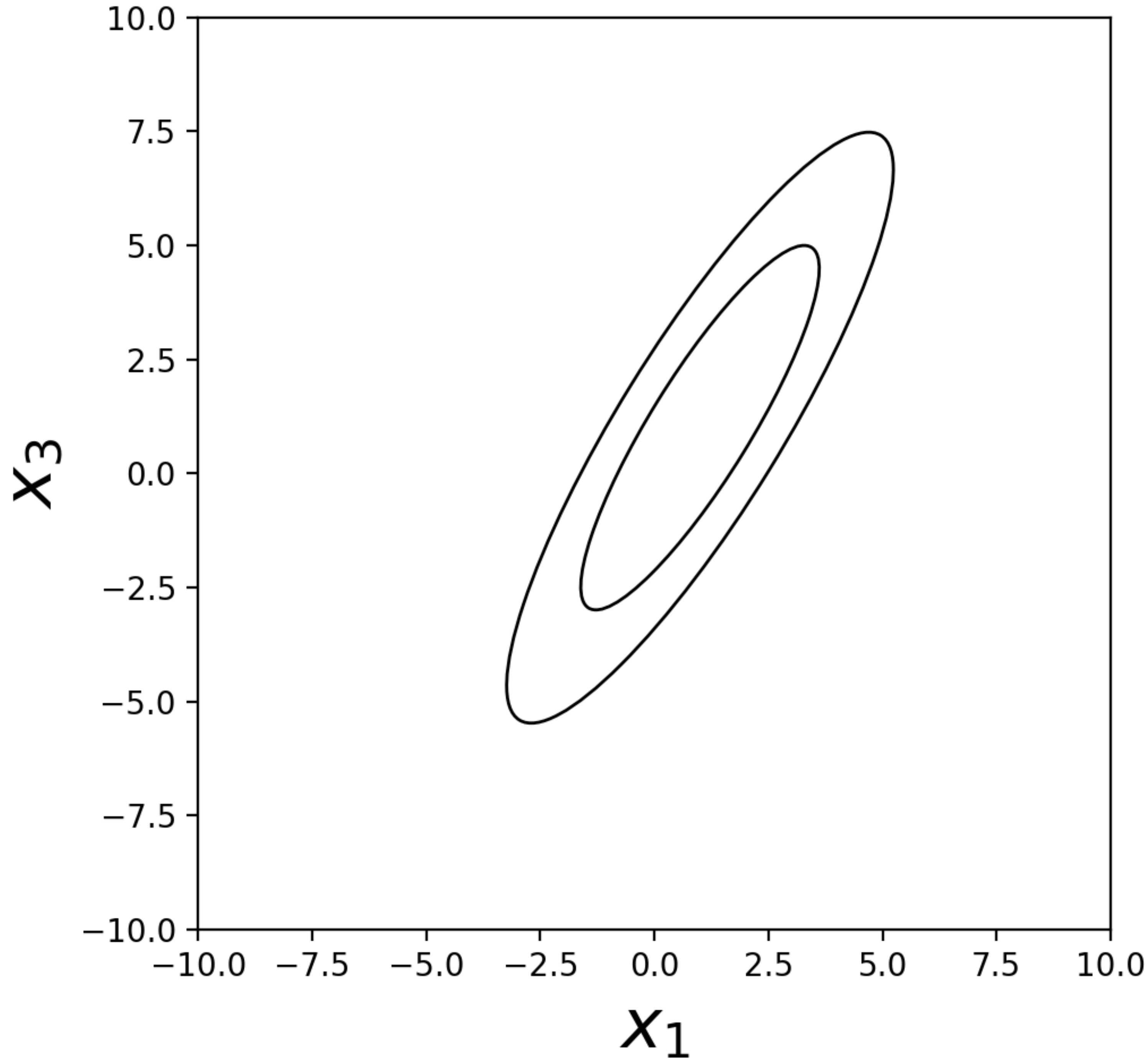
It can be shown that the contour enclosing this set of x values is an ellipse with axes

$$\boldsymbol{\mu} \pm \sqrt{5.9915 \lambda_i} \boldsymbol{e}_i$$

With λ_i **the eigenvalues of the covariance matrix and**

$\boldsymbol{e}_i$ **its eigenvectors**

An application of marginalization: Visualization of ellipses



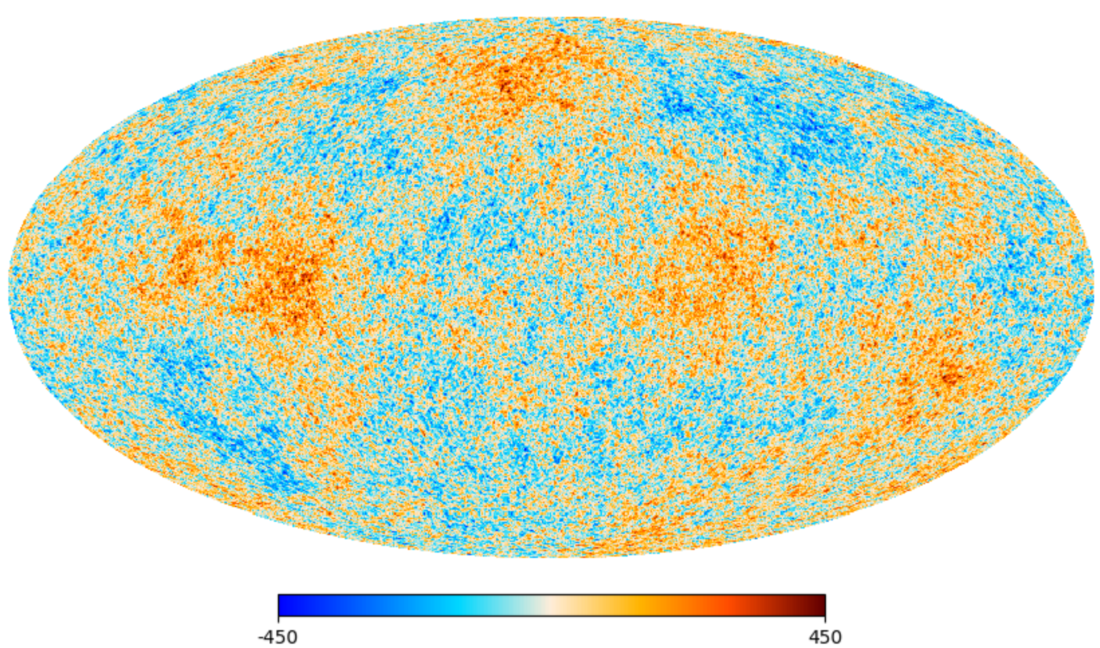
$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

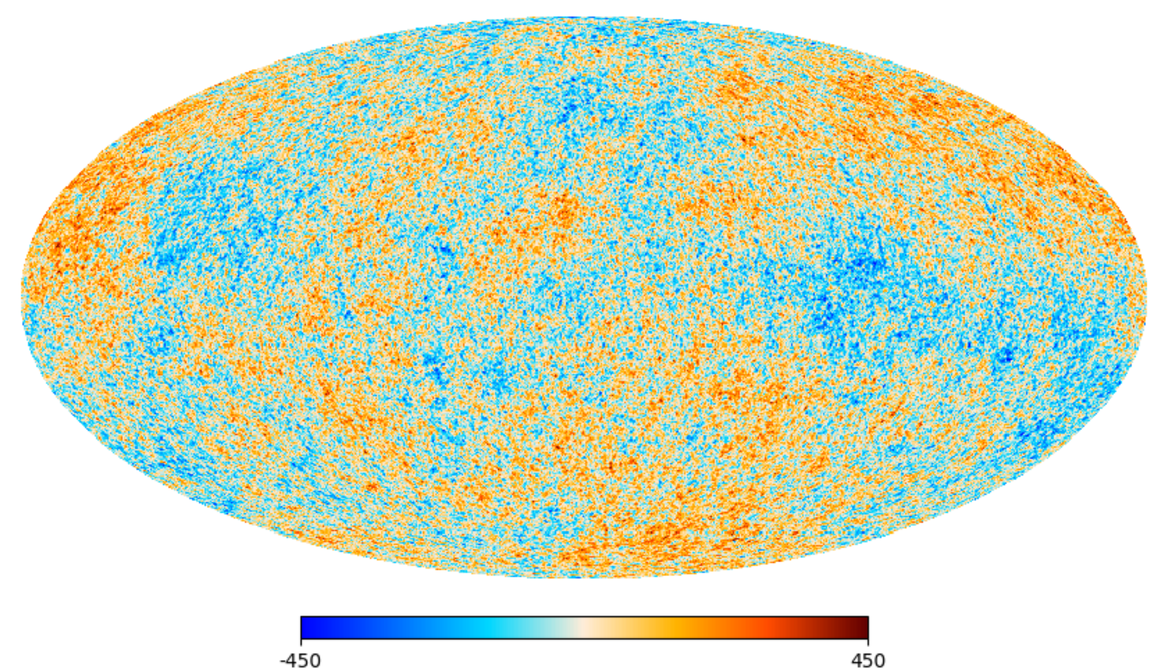
$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 3 & 4 \\ 4 & 7 \end{pmatrix}$$

$$\tilde{\boldsymbol{\rho}} = \begin{pmatrix} 1 & 0.87 \\ 0.87 & 1 \end{pmatrix}$$

Cosmological random fields

From the point of view of our cosmological models, the observed distribution of galaxies, as well as the observed temperature and polarization fluctuations of the cosmic microwave background, are contingent. Our models do not predict the specific values taken by these observables — they only predict their statistical properties.

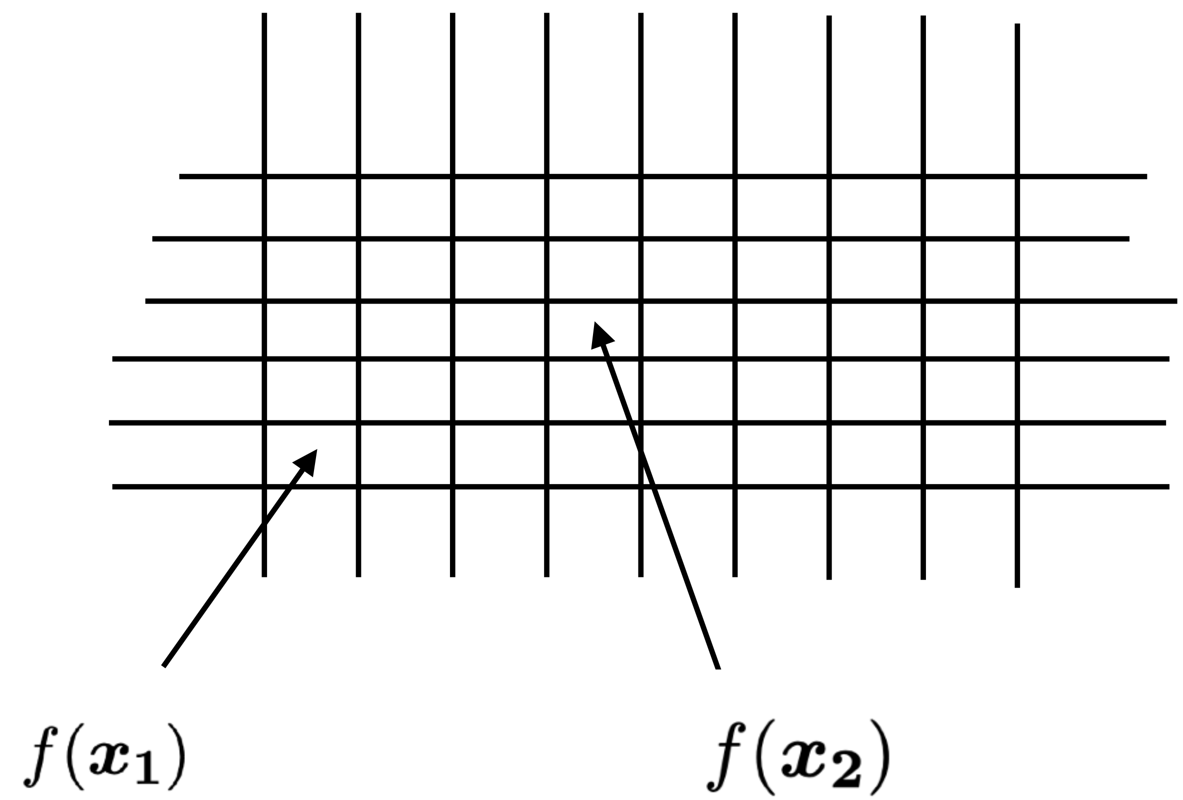




“Same universe” from the ΛCDM point of view

These observables are therefore promoted to the status of random fields.

We can generalize the concept of random variables to that of a random field. Imagine a set of random variables defined at each point of a grid with a given spacing.



By denoting $f_n = f(\boldsymbol{x_n})$ **one can define a vector** $\boldsymbol{f}$ **of n random variables.**
Its associated probability distribution is $\{\ \boldsymbol{f}, \mathcal{P}(\boldsymbol{f})\}$

The random field can be defined as the continuous limit in which the grid spacing tends to zero.

n-point correlation functions

By analogy with the moments of a random variable, one can define the *n*-point correlation functions of the field. $f(\boldsymbol{x})$

$$\xi(\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}) = \left\langle \prod_{i=1}^{n} f(\boldsymbol{x_i}) \right\rangle$$

The cosmological principle assumes that the statistical properties of the Universe on large scales are homogeneous and isotropic.

Homogeneity

$$\langle f(\boldsymbol{x_1})...f(\boldsymbol{x_n})\rangle = \langle f(\boldsymbol{x_1}+\boldsymbol{b})...f(\boldsymbol{x_n}+\boldsymbol{b})\rangle$$

Isotropy

A field is isotropic around a point z if

$$\begin{aligned}\xi(\boldsymbol{x_1},\boldsymbol{x_2},...,\boldsymbol{x_n}) &= \xi(\boldsymbol{x'_1},\boldsymbol{x'_2},...,\boldsymbol{x'_n})\\ \boldsymbol{x}' &= \boldsymbol{z}+R(\boldsymbol{x}-\boldsymbol{z})\end{aligned}$$

A prediction of the simplest inflation models (and widely confirmed by Planck) is that the initial fluctuations follow a Gaussian distribution.

$$\mathcal{P}(\boldsymbol{\delta_i}) \quad = \quad \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\delta^i})^{\boldsymbol{T}} C^{-1} \boldsymbol{\delta^i}\right]}{\sqrt{\det(2\pi C)}}$$

All the information about a Gaussian field is contained in its covariance matrix (its two-point correlation function).

$$C_{lm} \quad = \quad \langle \delta^i(\boldsymbol{x}_l)\delta^i(\boldsymbol{x}_m)\rangle$$

A prediction of the simplest inflation models (and widely confirmed by Planck) is that the initial fluctuations follow a Gaussian distribution.

$$\mathcal{P}(\boldsymbol{\delta_i}) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\delta^i})^{\boldsymbol{T}} C^{-1}\boldsymbol{\delta^i}\right]}{\sqrt{\det(2\pi C)}}$$

All the information about a Gaussian field is contained in its covariance matrix (its two-point correlation function).

$$C_{lm} = \langle \delta^i(\boldsymbol{x}_l)\delta^i(\boldsymbol{x}_m)\rangle$$

A linear combination of Gaussian fields follows a Gaussian distribution.

$$\delta T_{\rm CMB}(z=1100) = \mathcal{F}_1(\delta^i, \boldsymbol{v}^i)$$

$\mathcal{F}_1$ **is a functional that represents the evolution of the density contrasts generated by inflation up to the emission of the CMB**

$$\delta_m(z=0) = \mathcal{F}_2(\delta^i, \boldsymbol{v}^i) + {\rm NL}$$

Non-linear phenomena (such as gravity) generate non-Gaussianities; nevertheless, even in this case, the two-point correlation functions contain a lot of information.

Covariance Matrix

$$C_{lm} = \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m) \rangle$$

Covariance Matrix

$$
\begin{aligned}
C_{lm} &= \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m)\rangle \\
&= \xi(\boldsymbol{x}_l - \boldsymbol{x}_m)
\end{aligned}
$$

Homogeneity

The covariance matrix depends only on the relative positions.

Covariance Matrix

$$
\begin{aligned}
C_{lm} &= \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m) \rangle \\
&= \xi(\boldsymbol{x}_l - \boldsymbol{x}_m) \\
&= \xi(|\boldsymbol{x}_l - \boldsymbol{x}_m|)
\end{aligned}
$$

Isotropy

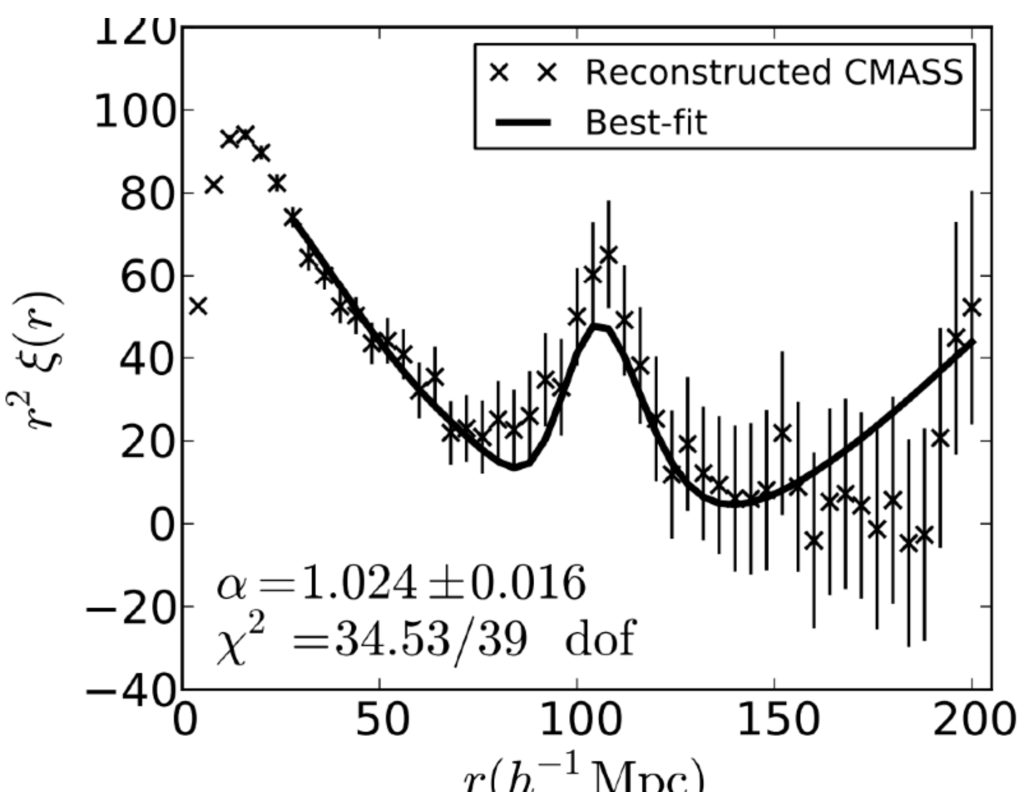
There is no directionality in the correlation matrix.

Covariance Matrix

$$
\begin{aligned}
C_{lm} &= \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m) \rangle \\
&= \xi(\boldsymbol{x}_l - \boldsymbol{x}_m) \\
&= \xi(|\boldsymbol{x}_l - \boldsymbol{x}_m|) \\
&= \xi(r)
\end{aligned}
$$

Covariance Matrix

$$
\begin{aligned}
C_{lm} &= \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m) \rangle \\
&= \xi(\boldsymbol{x}_l - \boldsymbol{x}_m) \\
&= \xi(|\boldsymbol{x}_l - \boldsymbol{x}_m|) \\
&= \xi(r)
\end{aligned}
$$



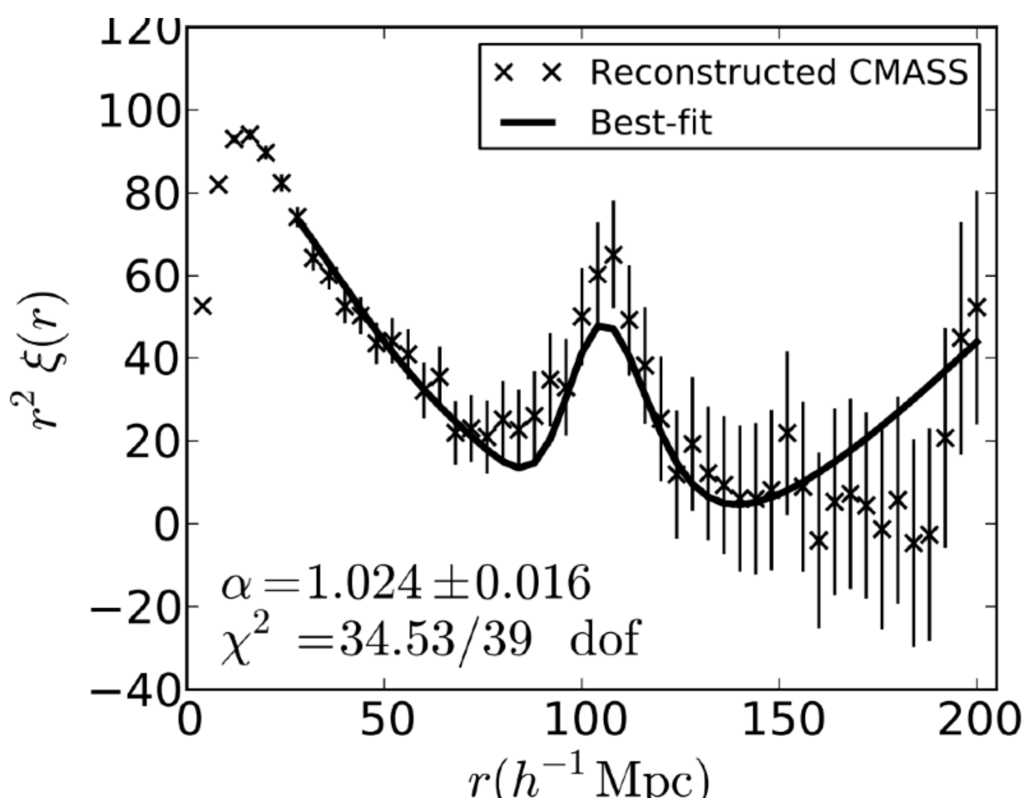
Boss z [0.43 to 0.7]

https://arxiv.org/pdf/1203.6594.pdf

Covariance Matrix

$$\begin{aligned} C_{lm} &= \langle \delta(\boldsymbol{x}_l)\delta(\boldsymbol{x}_m)\rangle \\ &= \xi(\boldsymbol{x}_l - \boldsymbol{x}_m) \\ &= \xi(|\boldsymbol{x}_l - \boldsymbol{x}_m|) \\ &= \xi(r) \end{aligned}$$

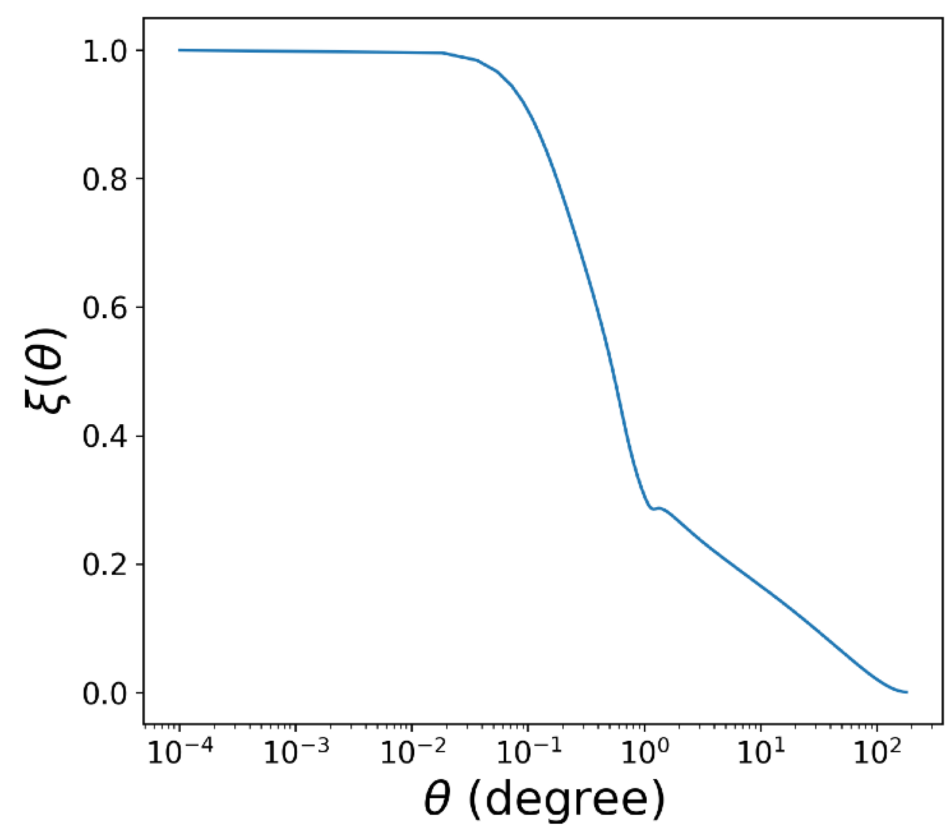
$$\begin{aligned} C_{lm} &= \langle \delta T(\hat{\boldsymbol{n}}_{\boldsymbol{l}})\delta T(\hat{\boldsymbol{n}}_{\boldsymbol{m}})\rangle \\ &= \xi(\hat{\boldsymbol{n}}_{\boldsymbol{l}}.\hat{\boldsymbol{n}}_{\boldsymbol{m}}) \\ &= \xi(\theta) \end{aligned}$$



Boss z [0.43 to 0.7]

https://arxiv.org/pdf/1203.6594.pdf

CMB temperature angular correlation function



To represent fields in three dimensions, the Fourier space is often used.

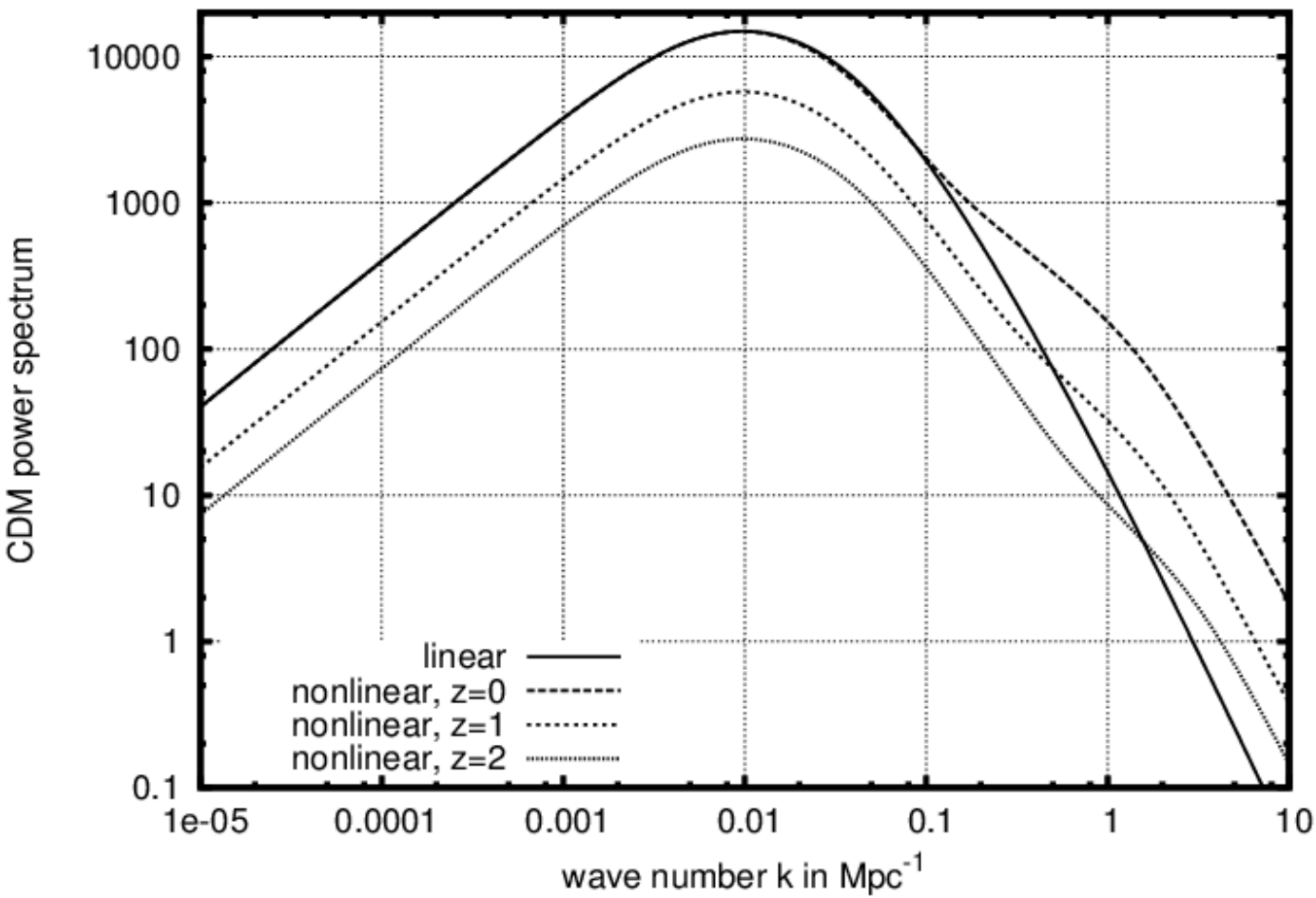
$$\begin{aligned}\delta(\boldsymbol{k}) &= \int d\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{k}\boldsymbol{x}} \\ \delta(\boldsymbol{x}) &= \int \frac{d\boldsymbol{k}}{(2\pi)^3}\delta(\boldsymbol{k})e^{-i\boldsymbol{k}\boldsymbol{x}}\end{aligned}$$

The analogue of the correlation function is the power spectrum. $P(k)$

$$P(\boldsymbol{k}) = \int d\boldsymbol{r}e^{i\boldsymbol{k}\boldsymbol{r}}\xi(\boldsymbol{r})$$

Fourier space presents advantages

1) Theoretical calculations of the evolution of perturbations are traditionally performed in Fourier space.
2) The basis represented by the Fourier modes diagonalizes the covariance matrix.
3) II It is natural to implement "scale cuts."


CDM power spectrum
10000
1000
100
10
1
0.1
1e-05
0.0001
0.001
0.01
0.1
1
10
wave number k in Mpc^{-1}
linear
nonlinear, z=0
nonlinear, z=1
nonlinear, z=2

Correlation function in Fourier space

$$\begin{aligned}
\langle \delta(\boldsymbol{k})\delta(\boldsymbol{q})^* \rangle &= \langle \int d\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{k}\boldsymbol{x}} \int d\boldsymbol{y}\delta(\boldsymbol{y})e^{-i\boldsymbol{q}\boldsymbol{y}} \rangle \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\langle \delta(\boldsymbol{x})\delta(\boldsymbol{y})\rangle e^{i(\boldsymbol{k}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y})} \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\xi(\boldsymbol{x}-\boldsymbol{y})e^{i(\boldsymbol{k}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y})}
\end{aligned}$$

Correlation function in Fourier space

$$
\begin{aligned}
\langle \delta(\boldsymbol{k})\delta(\boldsymbol{q})^* \rangle &= \langle \int d\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{k}\boldsymbol{x}} \int d\boldsymbol{y}\delta(\boldsymbol{y})e^{-i\boldsymbol{q}\boldsymbol{y}} \rangle \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\langle \delta(\boldsymbol{x})\delta(\boldsymbol{y})\rangle e^{i(\boldsymbol{k}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y})} \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\xi(\boldsymbol{x}-\boldsymbol{y})e^{i(\boldsymbol{k}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y})}
\end{aligned}
$$

Jacobi coordinates

$$
\boxed{
\begin{aligned}
\boldsymbol{r} &= \boldsymbol{x}-\boldsymbol{y} \\
\boldsymbol{x_c} &= (\boldsymbol{x}+\boldsymbol{y})/2
\end{aligned}
}
$$

$$
\begin{aligned}
(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x_c} &= \frac{1}{2}[\boldsymbol{k}\boldsymbol{x}+\boldsymbol{k}\boldsymbol{y}-\boldsymbol{q}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y}] \\
\frac{(\boldsymbol{k}+\boldsymbol{q})}{2}\boldsymbol{r} &= \frac{1}{2}[\boldsymbol{k}\boldsymbol{x}-\boldsymbol{k}\boldsymbol{y}+\boldsymbol{q}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y}] \\
(\boldsymbol{k}\boldsymbol{x}-\boldsymbol{q}\boldsymbol{y}) &= \frac{(\boldsymbol{k}+\boldsymbol{q})}{2}\boldsymbol{r}+(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x_c}
\end{aligned}
$$

Correlation function in Fourier space

$$\begin{aligned}
\langle \delta(\boldsymbol{k})\delta(\boldsymbol{q})^* \rangle &= \langle \int d\boldsymbol{x}\delta(\boldsymbol{x})e^{i\boldsymbol{kx}} \int d\boldsymbol{y}\delta(\boldsymbol{y})e^{-i\boldsymbol{qy}} \rangle \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\langle \delta(\boldsymbol{x})\delta(\boldsymbol{y})\rangle e^{i(\boldsymbol{kx}-\boldsymbol{qy})} \\
&= \int d\boldsymbol{x}d\boldsymbol{y}\xi(\boldsymbol{x}-\boldsymbol{y})e^{i(\boldsymbol{kx}-\boldsymbol{qy})} \\
&= \int d\boldsymbol{x_c}e^{i(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x_c}} \int d\boldsymbol{r}\xi(\boldsymbol{r})e^{i\frac{(\boldsymbol{k}+\boldsymbol{q})}{2}\boldsymbol{r}} \\
&= (2\pi)^3\delta(\boldsymbol{k}-\boldsymbol{q}) \int d\boldsymbol{r}\xi(\boldsymbol{r})e^{i\boldsymbol{kr}} \\
&= (2\pi)^3\delta(\boldsymbol{k}-\boldsymbol{q})P(\boldsymbol{k})
\end{aligned}$$

Jacobi coordinates

$$\boxed{\begin{aligned}
\boldsymbol{r} &= \boldsymbol{x}-\boldsymbol{y} \\
\boldsymbol{x_c} &= (\boldsymbol{x}+\boldsymbol{y})/2
\end{aligned}}$$

$$\begin{aligned}
(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x_c} &= \frac{1}{2}[\boldsymbol{kx}+\boldsymbol{ky}-\boldsymbol{qx}-\boldsymbol{qy}] \\
\frac{(\boldsymbol{k}+\boldsymbol{q})}{2}\boldsymbol{r} &= \frac{1}{2}[\boldsymbol{kx}-\boldsymbol{ky}+\boldsymbol{qx}-\boldsymbol{qy}] \\
(\boldsymbol{kx}-\boldsymbol{qy}) &= \frac{(\boldsymbol{k}+\boldsymbol{q})}{2}\boldsymbol{r}+(\boldsymbol{k}-\boldsymbol{q})\boldsymbol{x_c}
\end{aligned}$$

Different Fourier modes are independent

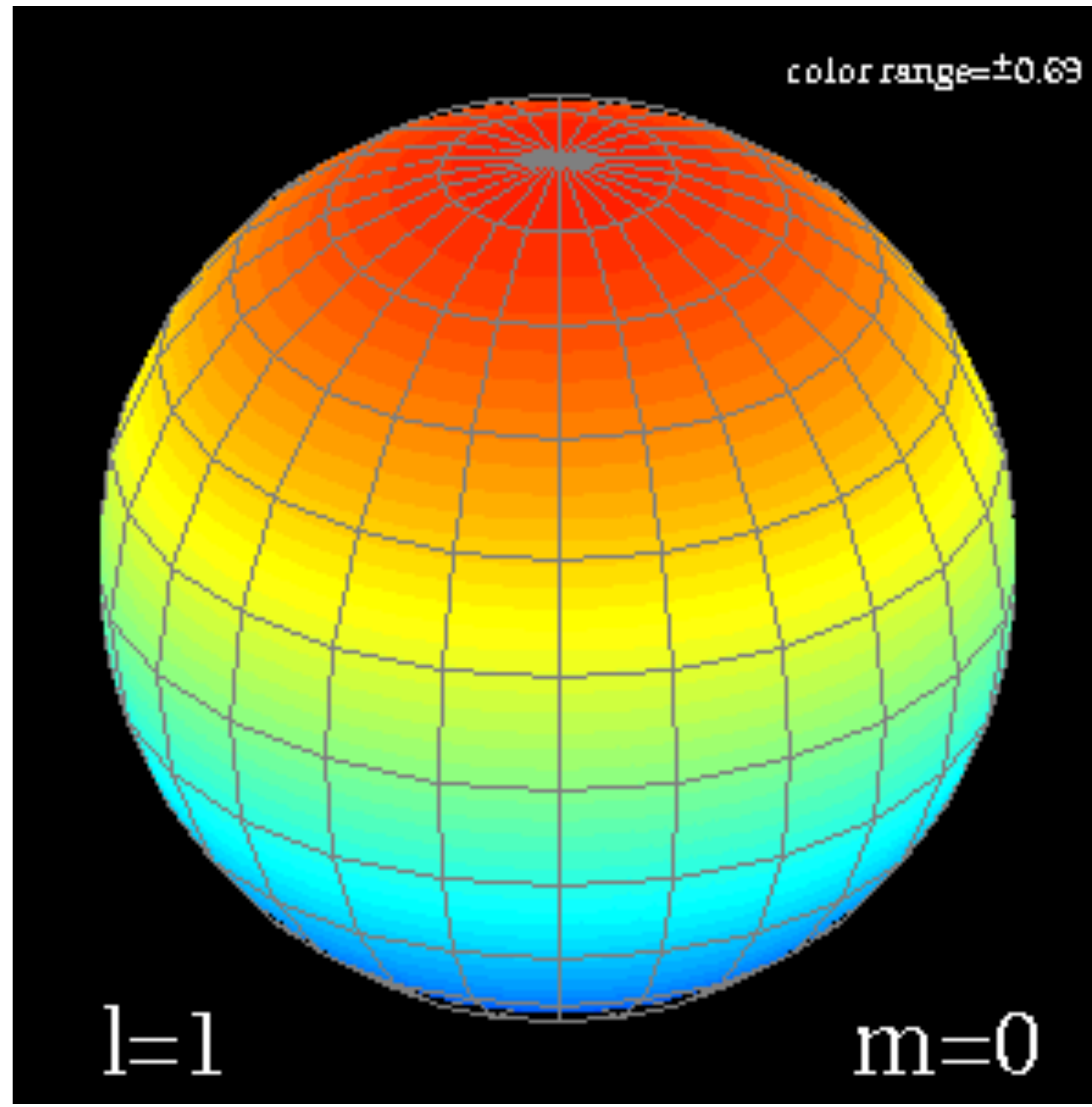
Angular power spectrum

On the sphere, the harmonic basis is the basis of spherical harmonics.

$Y_{\ell m}(\hat{n})$

$$\delta T(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n})$$

$$a_{\ell m} = \int \delta T(\hat{n}) Y^*_{\ell m}(\hat{n}) d\hat{n}$$



Angular power spectrum

One can define an angular power spectrum as well as an angular correlation function.

Legendre polynomial

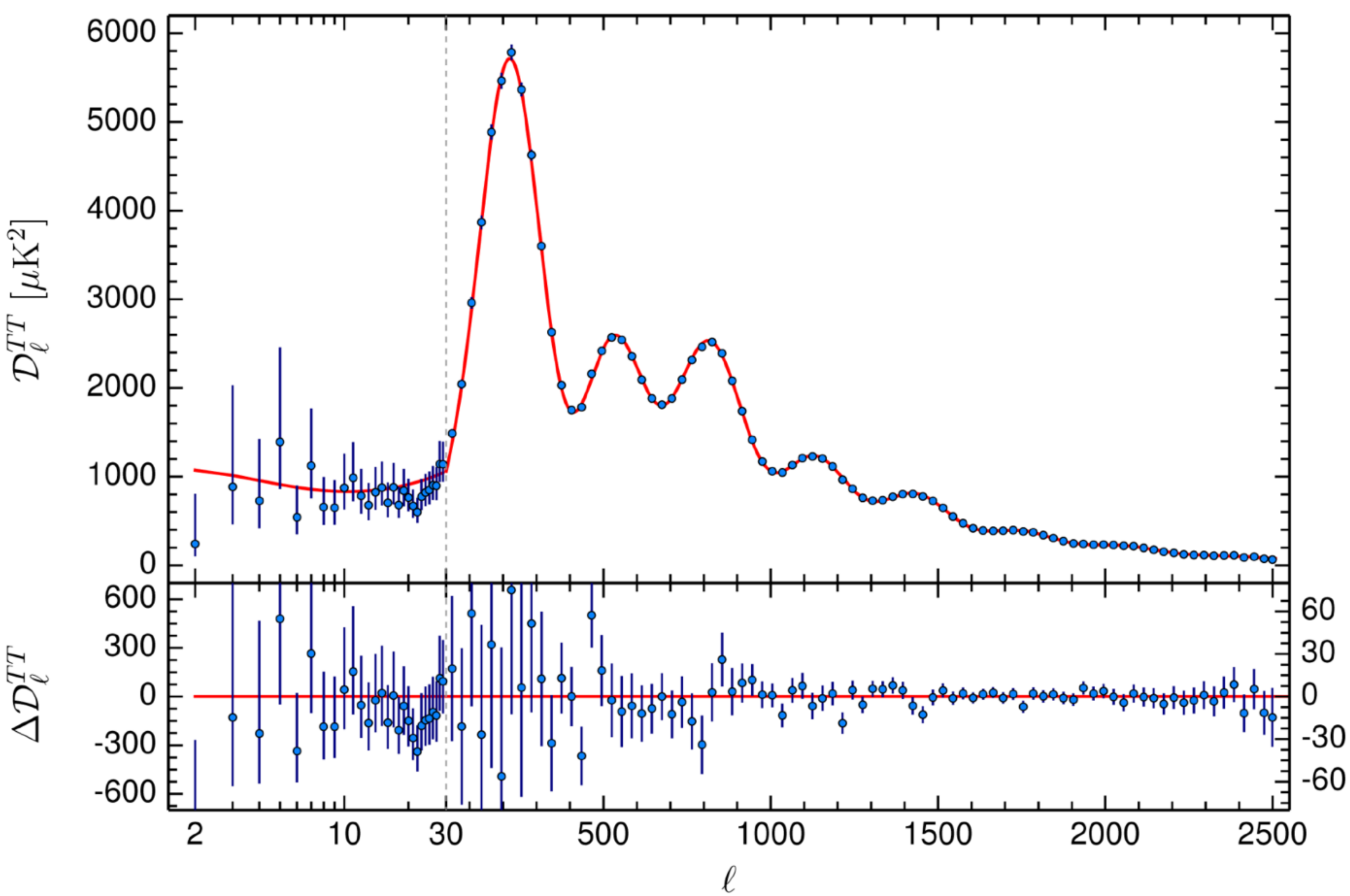
$$
\begin{aligned}
\langle a_{\ell m} a^*_{\ell' m'} \rangle &= \delta_{\ell\ell'} \delta_{mm'} C_\ell \\
\xi(\theta) &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} P_\ell(\cos\theta) C_\ell
\end{aligned}
$$

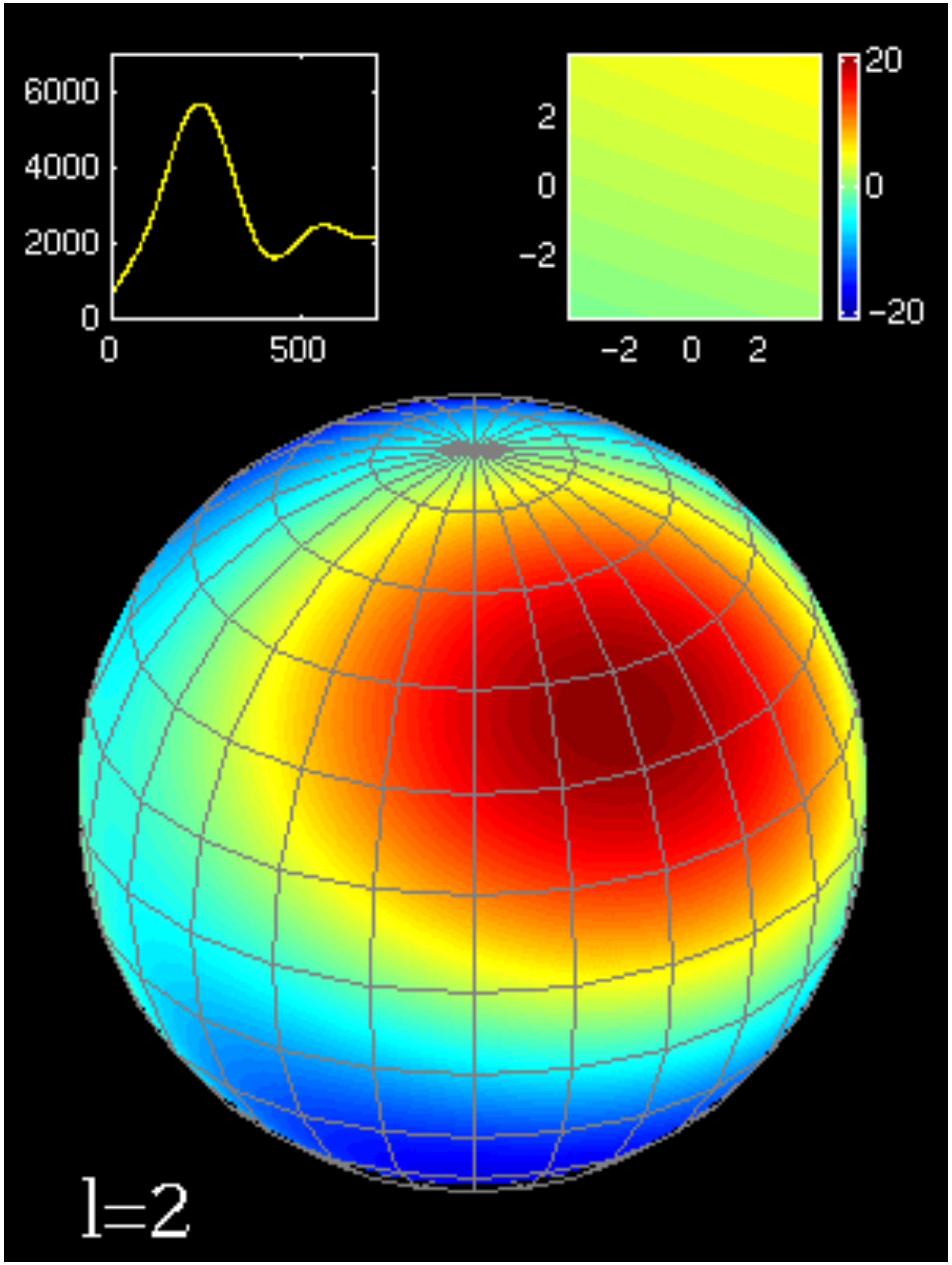
Angular power spectrum

One can define an angular power spectrum as well as an angular correlation function.

Legendre polynomial

$$
\begin{aligned}
\langle a_{\ell m} a^*_{\ell' m'} \rangle &= \delta_{\ell\ell'}\delta_{mm'}C_\ell \\
\xi(\theta) &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} P_\ell(\cos\theta) C_\ell
\end{aligned}
$$




6000
4000
2000
0
0
500
2
0
−2
−2
0
2
20
0
−20
l=2

Estimation of the angular power spectra of the CMB

The angular power spectrum definition is

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

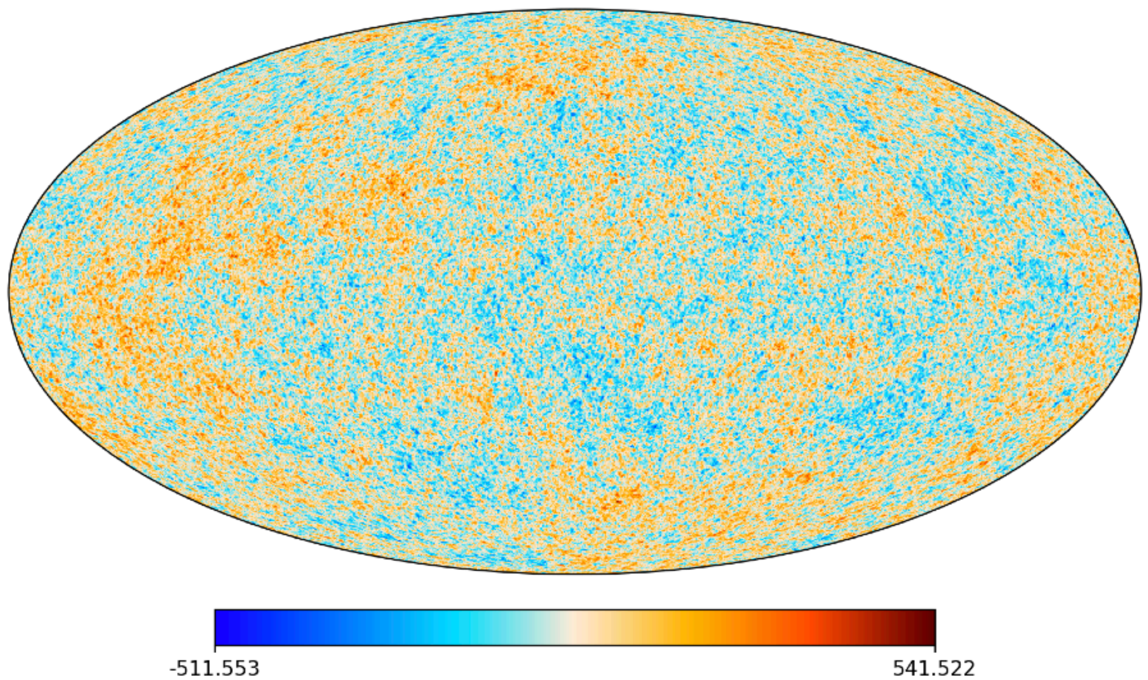
We can define and estimator

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m} a^*_{\ell m}$$

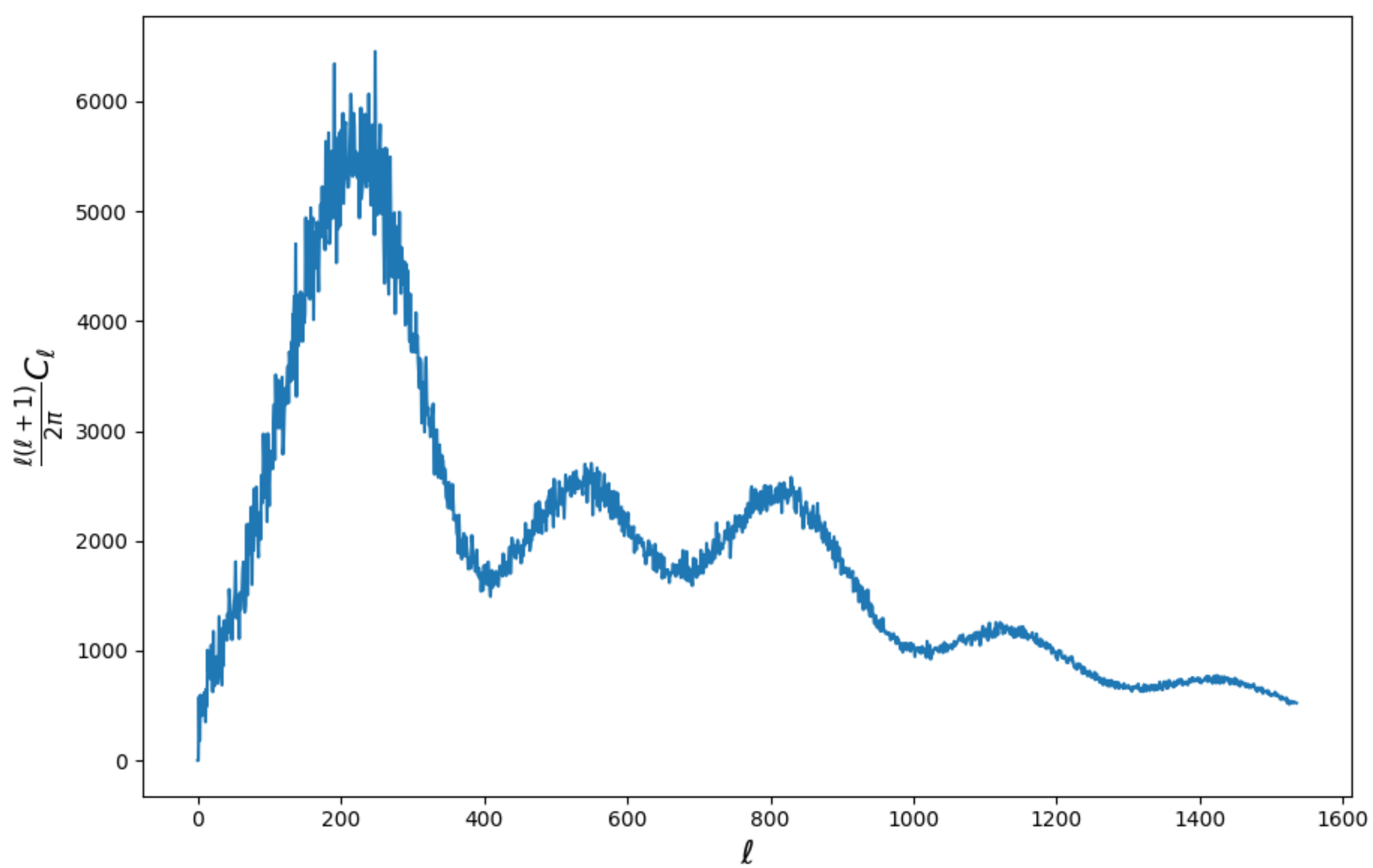
Where the $a_{\ell m}$ **are computed from**

$$a_{\ell m} = \int \delta T(\hat{n}) Y^*_{\ell m}(\hat{n}) d\hat{n}$$

Numerics :




```python
import healpy as hp
import pylab as plt
import numpy as np

cmb_map = hp.fitsfunc.read_map("cmb.fits")
alms = hp.sphtfunc.map2alm(cmb_map)
cls = hp.sphtfunc.alm2cl(alms)

```



When defining an estimator of a statistical property, one must always answer (at least) two questions:

-> is the estimator biased ?
-> what is its variance ?

Bias

An estimator is unbiased if the ensemble average of the estimator equals the quantity one wishes to estimate

$$\langle \hat{C}_\ell \rangle = C_\ell$$

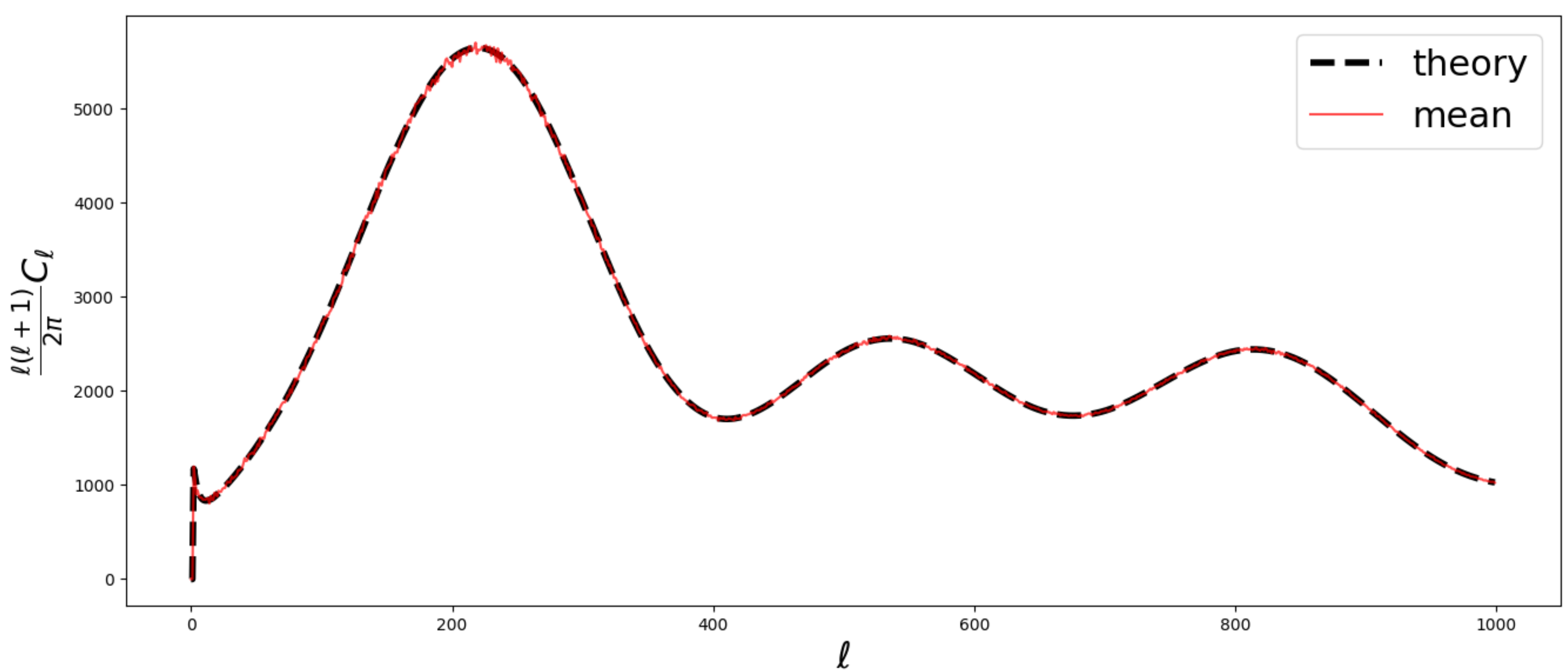
We can check whether this is the case for our estimator

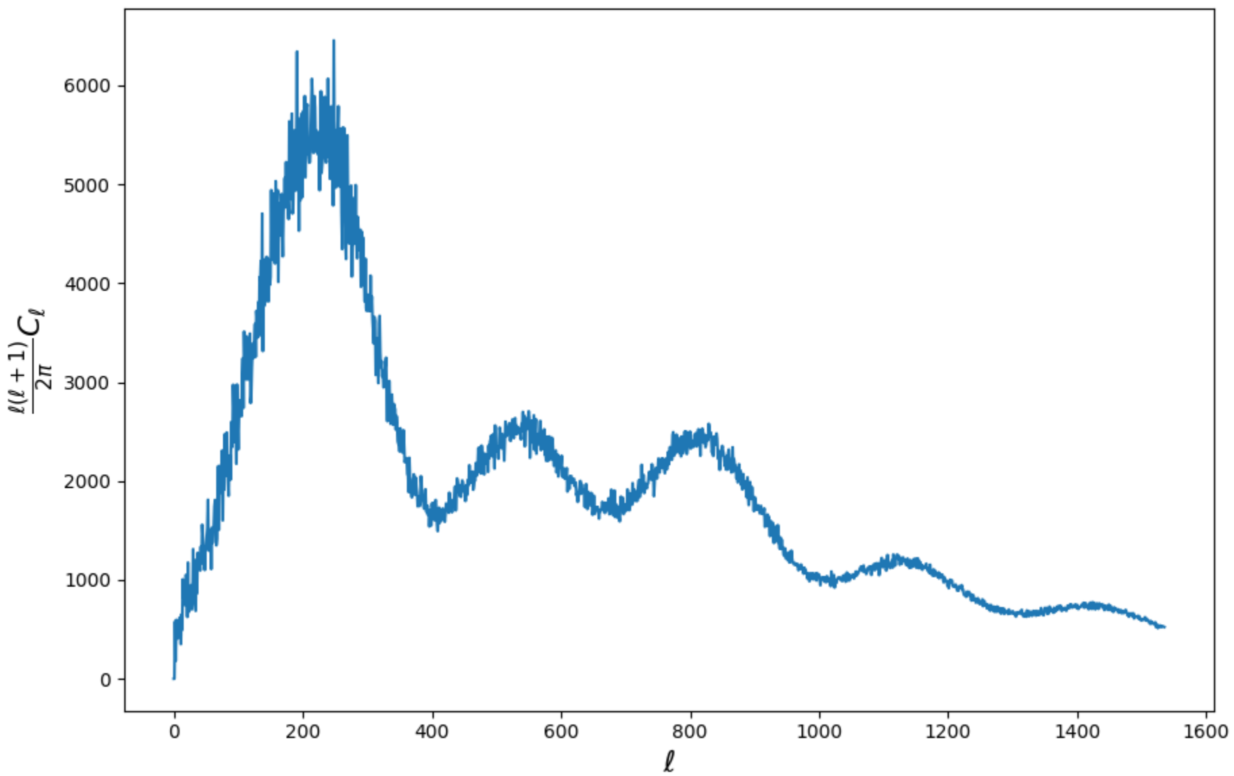
$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m} a^*_{\ell m}$$

$$\begin{aligned} \langle \hat{C}_\ell \rangle &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} \rangle \\ &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} C_\ell \\ &= C_\ell \end{aligned}$$

Numerics :

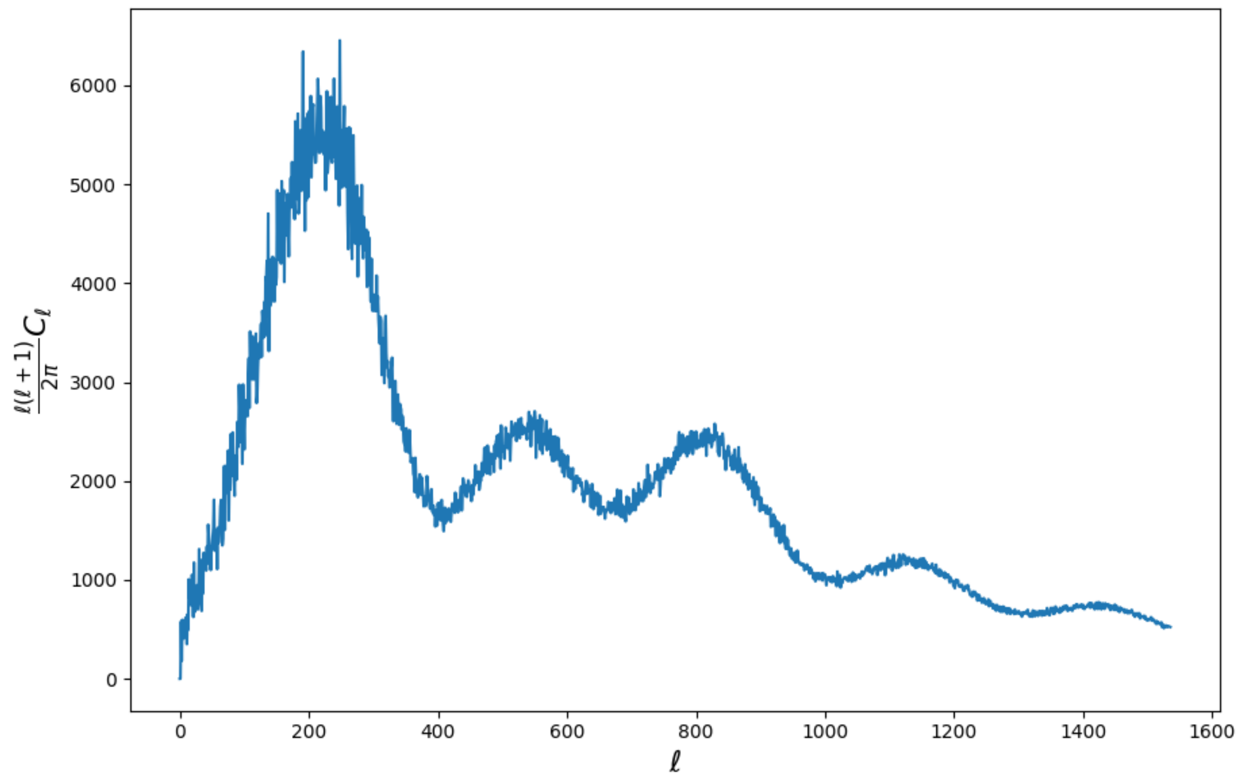
```
nside = 512
nsims = 100
cl_list = []
for i in range(nsims):
    print(i)
    cmb_sim = hp.sphtfunc.synfast(cl_th, nside=nside) #generate simulation
    alms_sim = hp.sphtfunc.map2alm(cmb_sim)
    cls_sim = hp.sphtfunc.alm2cl(alms_sim)
    cl_list += [cls_sim]


mean = np.mean(cl_list, axis=0)
```





We can also compute the variance of this estimator

$$\sigma^2(\hat{C}_\ell) = \langle(\hat{C}_\ell - C_\ell)^2\rangle \quad = \quad \langle\hat{C}_\ell^2\rangle - 2C_\ell\langle\hat{C}_\ell\rangle + C_\ell^2 = \langle\hat{C}_\ell^2\rangle - C_\ell^2$$



We can also compute the variance of this estimator

$$\sigma^2(\hat{C}_\ell) = \langle (\hat{C}_\ell - C_\ell)^2 \rangle \quad = \quad \langle \hat{C}_\ell^2 \rangle - 2C_\ell \langle \hat{C}_\ell \rangle + C_\ell^2 = \langle \hat{C}_\ell^2 \rangle - C_\ell^2$$

We now need to compute $\langle \hat{C}_\ell^2 \rangle$

$$\langle \hat{C}_\ell^2 \rangle = \left(\frac{1}{2\ell + 1} \right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle$$

To compute this term, we will use several properties.

$$\langle \hat{C}_\ell^2 \rangle = \left(\frac{1}{2\ell+1} \right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \rangle$$

1) Linear transformations of Gaussian fields follow Gaussian statistics.

$$a_{\ell m} = \int \underbrace{\delta T(\hat{n})}_{\text{Gaussian}} Y^*_{\ell m}(\hat{n}) d\hat{n}$$

To compute this term, we will use several properties.

$$\langle \hat{C}_\ell^2 \rangle = \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle$$

1) Linear transformations of Gaussian fields follow Gaussian statistics.

$$a_{\ell m} = \int \delta T(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n}$$

Gaussian **Gaussian**

To compute this term, we will use several properties.

$$\langle \hat{C}_\ell^2 \rangle = \left(\frac{1}{2\ell+1} \right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle$$

1) Linear transformations of Gaussian fields follow Gaussian statistics.

$$a_{\ell m} = \int \delta T(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n}$$

Gaussian

Gaussian

2) The Wick theorem

The Wick theorem

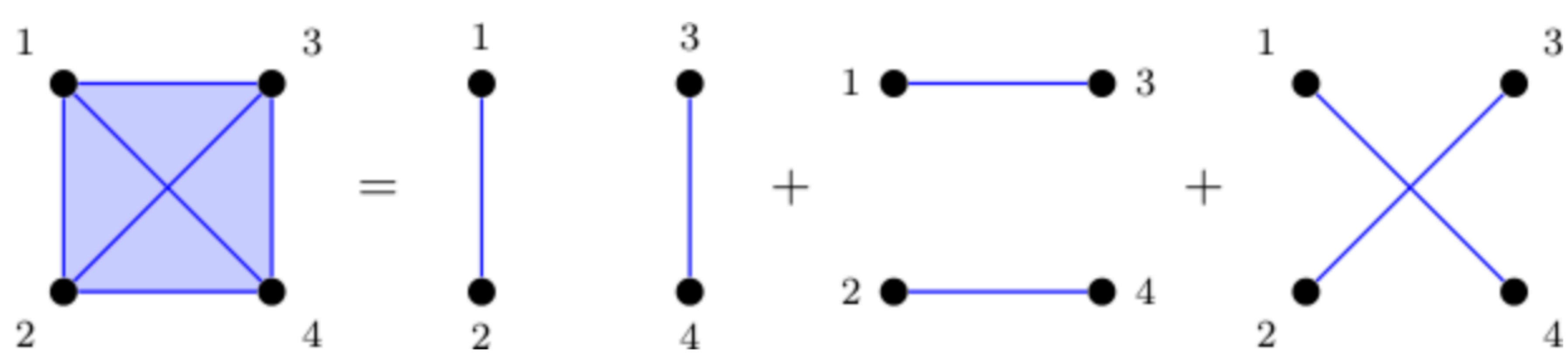
The N-point function of a Gaussian field can be decomposed into a sum of products of two-point correlation functions.

$$\mathrm{E}[\,X_1X_2\cdots X_n\,] = \sum_{p\in P_n^2}\prod_{\{i,j\}\in p}\mathrm{E}[\,X_iX_j\,] = \sum_{p\in P_n^2}\prod_{\{i,j\}\in p}\mathrm{Cov}(\,X_i, X_j\,)$$

The sum is taken over all pairings of (1…n)

In the case of a four-point correlation function.

$$\langle X_1X_2X_3X_4\rangle = \langle X_1X_2\rangle\langle X_3X_4\rangle + \langle X_1X_3\rangle\langle X_2X_4\rangle + \langle X_1X_4\rangle\langle X_2X_3\rangle$$



$$\langle \hat{C}_\ell^2 \rangle = \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \rangle$$

$$
\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} \rangle \langle a_{\ell m'} a^*_{\ell m'} \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a^*_{\ell m} a^*_{\ell m'} \rangle + \langle a_{\ell m} a^*_{\ell m'} \rangle \langle a^*_{\ell m} a_{\ell m'} \rangle
\end{aligned}
$$

$$
\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} \rangle \langle a_{\ell m'} a^*_{\ell m'} \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a^*_{\ell m} a^*_{\ell m'} \rangle + \langle a_{\ell m} a^*_{\ell m'} \rangle \langle a^*_{\ell m} a_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} C_\ell^2 + C_\ell^2 \delta_{m,-m'} + C_\ell^2 \delta_{m,m'}
\end{aligned}
$$

$$
\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* \rangle \langle a_{\ell m'} a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a_{\ell m}^* a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'}^* \rangle \langle a_{\ell m}^* a_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} C_\ell^2 + C_\ell^2 \delta_{m,-m'} + C_\ell^2 \delta_{m,m'} \\
&= C_\ell^2 + \frac{2}{2\ell+1} C_\ell^2
\end{aligned}
$$

$$
\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* a_{\ell m'} a_{\ell m'}^* \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a_{\ell m}^* \rangle \langle a_{\ell m'} a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a_{\ell m}^* a_{\ell m'}^* \rangle + \langle a_{\ell m} a_{\ell m'}^* \rangle \langle a_{\ell m}^* a_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} C_\ell^2 + C_\ell^2 \delta_{m,-m'} + C_\ell^2 \delta_{m,m'} \\
&= C_\ell^2 + \frac{2}{2\ell+1} C_\ell^2
\end{aligned}
$$

$$
\sigma^2(\hat{C}_\ell) = \langle \hat{C}_\ell^2 \rangle - C_\ell^2 = \frac{2}{2\ell+1} C_\ell^2
$$

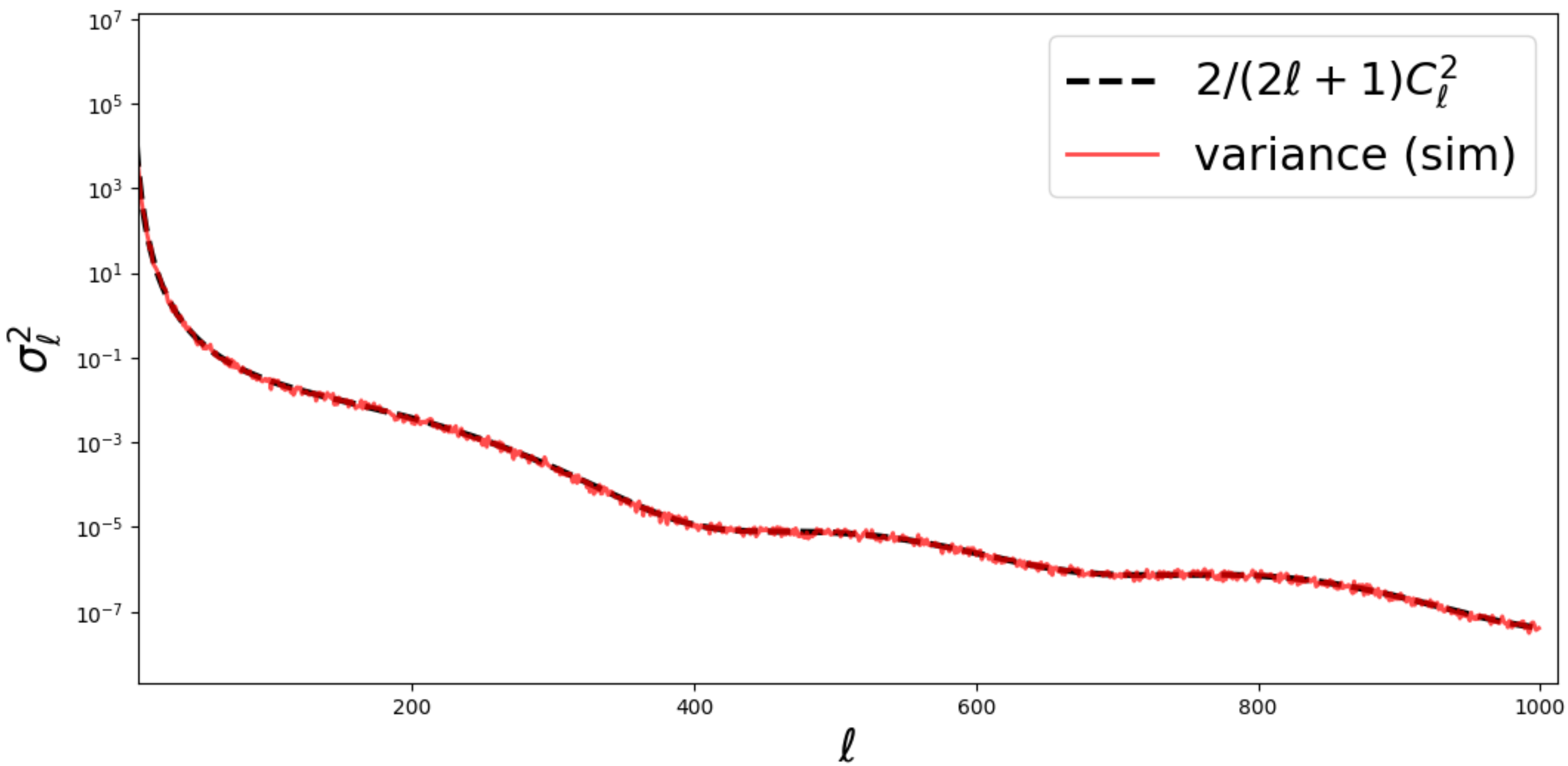
$$
\begin{aligned}
\langle \hat{C}_\ell^2 \rangle &= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} a_{\ell m'} a^*_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} \langle a_{\ell m} a^*_{\ell m} \rangle \langle a_{\ell m'} a^*_{\ell m'} \rangle + \langle a_{\ell m} a_{\ell m'} \rangle \langle a^*_{\ell m} a^*_{\ell m'} \rangle + \langle a_{\ell m} a^*_{\ell m'} \rangle \langle a^*_{\ell m} a_{\ell m'} \rangle \\
&= \left(\frac{1}{2\ell+1}\right)^2 \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell}^{\ell} C_\ell^2 + C_\ell^2 \delta_{m,-m'} + C_\ell^2 \delta_{m,m'} \\
&= C_\ell^2 + \frac{2}{2\ell+1} C_\ell^2
\end{aligned}
$$

$$
\sigma^2(\hat{C}_\ell) = \langle \hat{C}_\ell^2 \rangle - C_\ell^2 = \boxed{\frac{2}{2\ell+1} C_\ell^2}
$$

This term represents the uncertainty inherent to any measurement of the power spectrum; it is known as cosmic variance and is irreducible.
Our measurement is intrinsically limited by the fact that
we have access to only a single “realization” of the Universe.

Numerics :

```
nside = 512
nsims = 100
cl_list = []
for i in range(nsims):
    print(i)
    cmb_sim = hp.sphtfunc.synfast(cl_th, nside=nside) #generate simulation
    alms_sim = hp.sphtfunc.map2alm(cmb_sim)
    cls_sim = hp.sphtfunc.alm2cl(alms_sim)
    cl_list += [cls_sim]


mean = np.mean(cl_list, axis=0)
var = np.var(cl_list, axis=0)


var_analytic = 2 / (2 * l_th + 1) * cl_th ** 2
```



$$\sigma^2(\hat{C}_\ell) = \langle \hat{C}_\ell^2 \rangle - C_\ell^2 = \frac{2}{2\ell+1} C_\ell^2$$

The denominator is the number of modes: the number of "m" for each "l."

$$\sigma(\hat{C}_\ell) = \sqrt{\frac{2}{N_{\rm modes}}} C_\ell$$

There is an analogue of cosmic variance for the three-dimensional matter power spectrum.

$$\sigma^2(P(k)) = \frac{2}{\frac{4\pi k^2 dk}{(2\pi)^3} V_{\rm eff}} P(k)^2$$

Noise bias

In reality, every observation is affected by sources of noise.

$$\begin{aligned}
\delta T^{\mathrm{obs}}(\hat{n}) &= \delta T^{\mathrm{CMB}}(\hat{n}) + n(\hat{n}) \\
a_{\ell m}^{\mathrm{obs}} &= \int \delta T^{\mathrm{obs}}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\
a_{\ell m}^{\mathrm{obs}} &= a_{\ell m}^{\mathrm{CMB}} + n_{\ell m}
\end{aligned}$$

Noise will bias the estimator and increase its variance.

$$\begin{aligned}
\langle \hat{C}_{\ell}^{\mathrm{obs}} \rangle &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{\ell m}^{\mathrm{obs}} a_{\ell m}^{\mathrm{obs},*} \rangle = C_{\ell}^{\mathrm{CMB}} + N_{\ell} \\
N_{\ell} &= \langle n_{\ell m} n_{\ell m}^{*} \rangle
\end{aligned}$$

$$\sigma^2(\hat{C}_{\ell}^{\mathrm{obs}}) = \frac{2}{2\ell+1} (C_{\ell} + N_{\ell})^2$$

Noise bias

In reality, every observation is affected by sources of noise.

$$\begin{aligned} \delta T^{\rm obs}(\hat{n}) &= \delta T^{\rm CMB}(\hat{n}) + n(\hat{n}) \\ a_{\ell m}^{\rm obs} &= \int \delta T^{\rm obs}(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n} \\ a_{\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{\ell m} \end{aligned}$$

Noise will bias the estimator and increase its variance.

$$\begin{aligned} \langle \hat{C}_\ell^{\rm obs} \rangle &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{\ell m}^{\rm obs} a_{\ell m}^{\rm obs,*} \rangle = C_\ell^{\rm CMB} + N_\ell \\ N_\ell &= \langle n_{\ell m} n_{\ell m}^* \rangle \end{aligned}$$

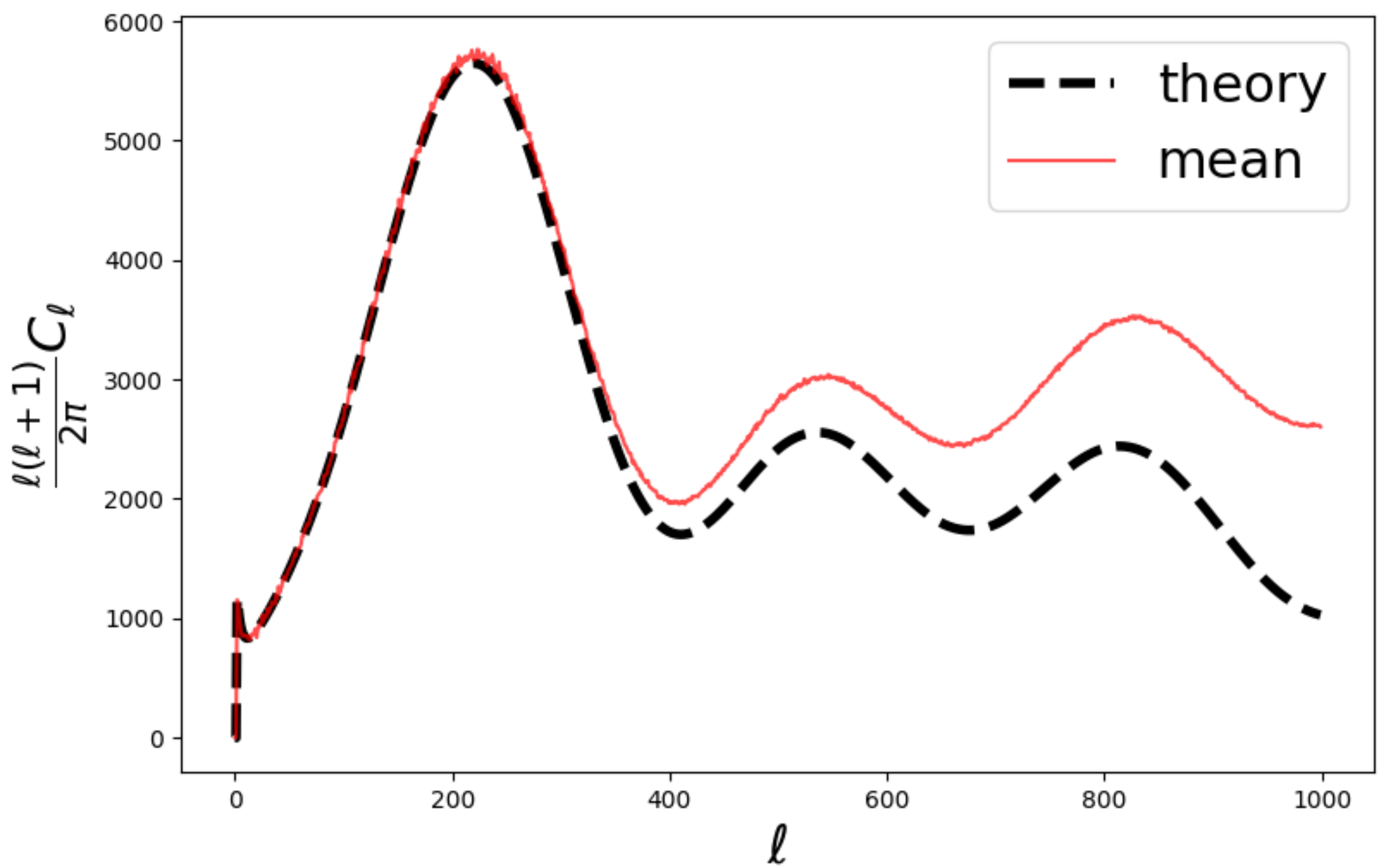
$$\sigma^2(\hat{C}_\ell^{\rm obs}) = \frac{2}{2\ell+1}(C_\ell + N_\ell)^2$$

Note that when Nℓ =0, we recover

The cosmic variance.

Numerics

```
nside = 512
nsims = 100
cl_list = []
for i in range(nsims):
    print(i)
    cmb_sim = hp.sphtfunc.synfast(cl_th, nside=nside) #generate simulation

    cmb_sim += np.random.randn(cmb_sim.shape[0])*50 #add noise
    alms_sim = hp.sphtfunc.map2alm(cmb_sim)
    cls_sim = hp.sphtfunc.alm2cl(alms_sim)
    cl_list += [cls_sim]


mean = np.mean(cl_list, axis=0)
```



Noise bias

In reality, every observation is affected by sources of noise.

$$\begin{aligned}\delta T^{\rm obs}(\hat{n}) &= \delta T^{\rm CMB}(\hat{n}) + n(\hat{n}) \\ a_{\ell m}^{\rm obs} &= \int \delta T^{\rm obs}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\ a_{\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{\ell m}\end{aligned}$$

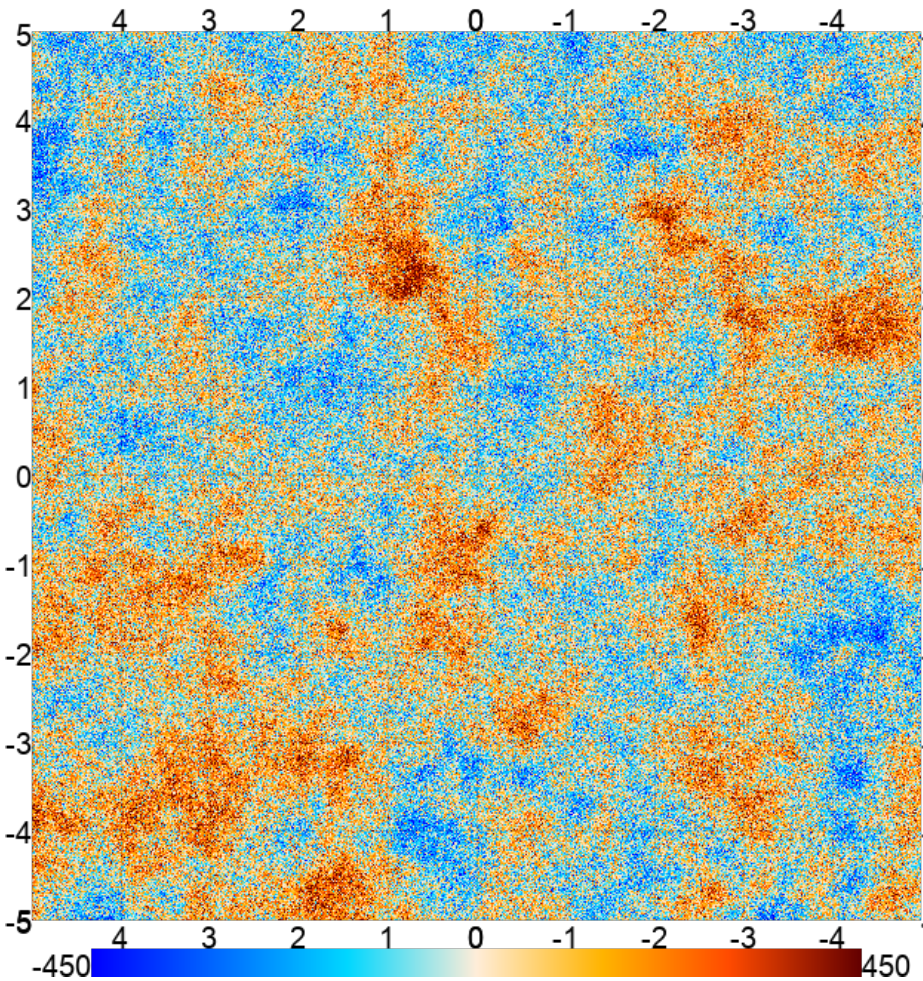
We can debias the power spectrum estimator in different ways:

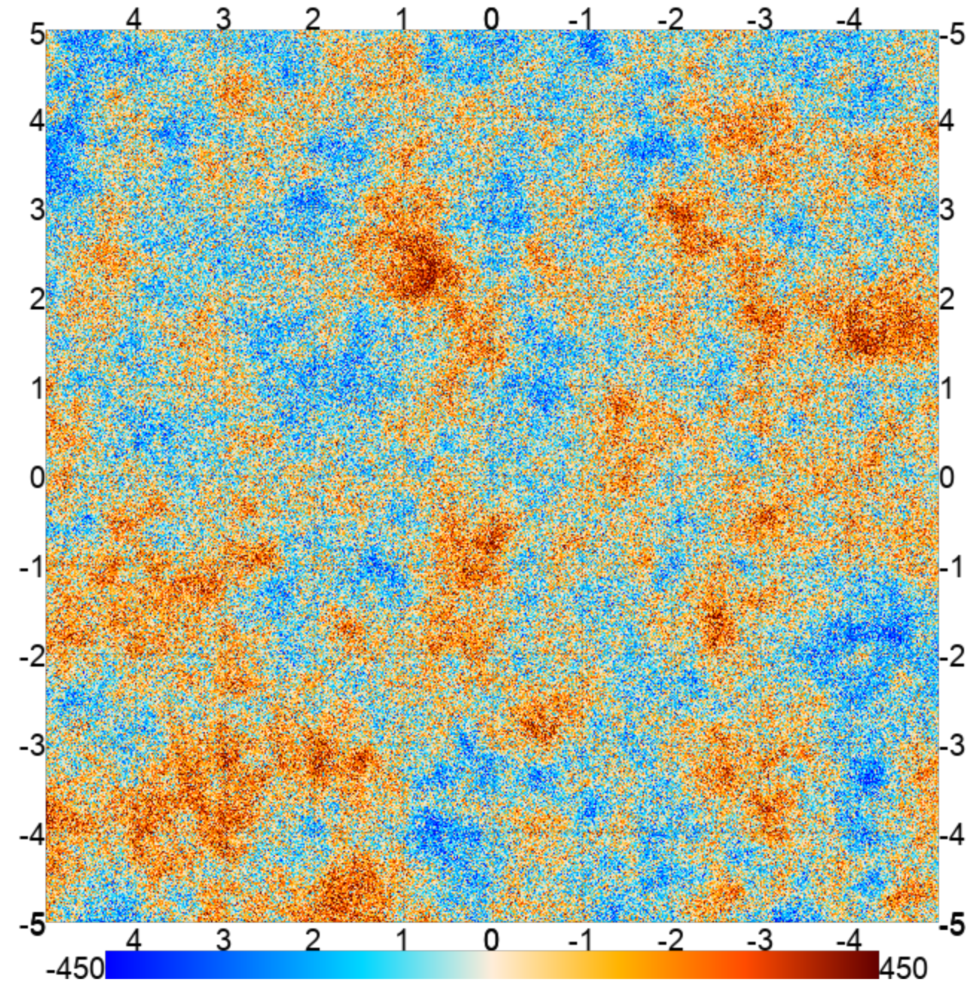
- **Subtracting a noise model built from our knowledge of the instrument**
- **Performing cross-correlations**

Cross-correlations

$$\delta T_1^{\mathrm{obs}}(\hat{n}) \quad = \quad \delta T^{\mathrm{CMB}}(\hat{n}) + n_1(\hat{n})$$

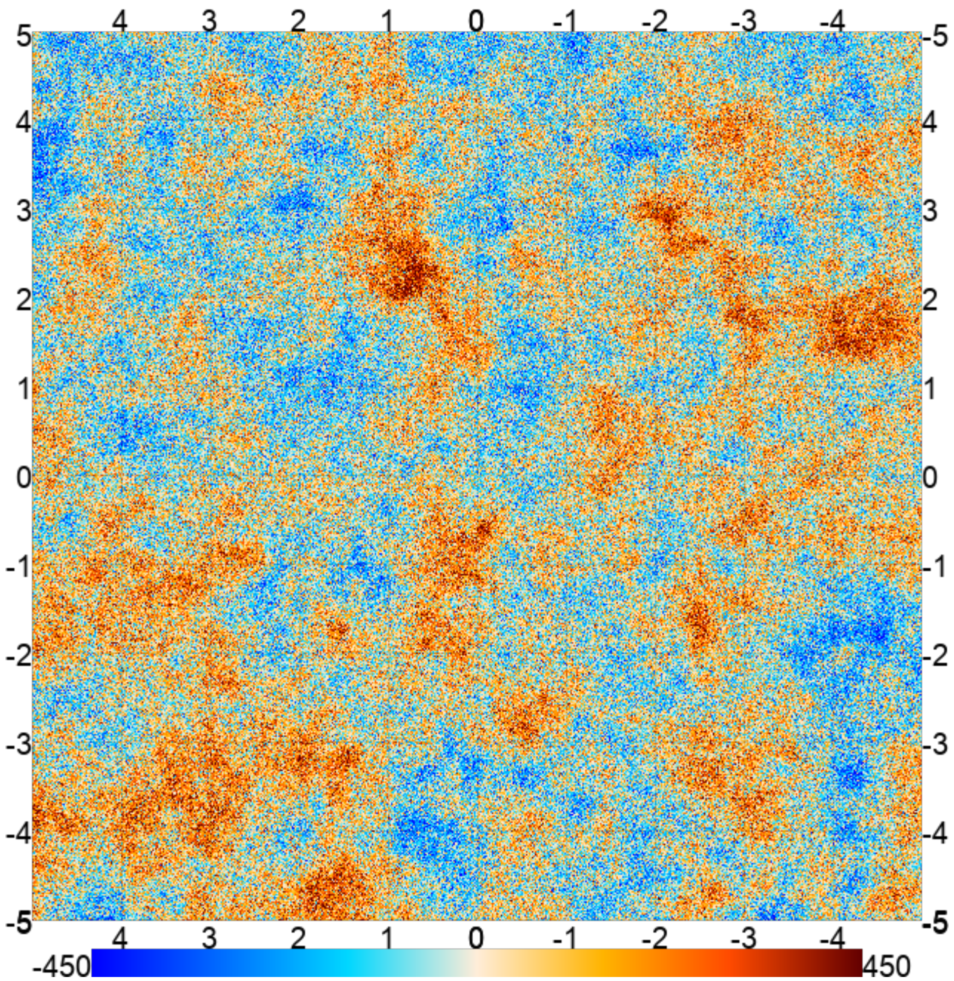
$$\delta T_2^{\mathrm{obs}}(\hat{n}) \quad = \quad \delta T^{\mathrm{CMB}}(\hat{n}) + n_2(\hat{n})$$



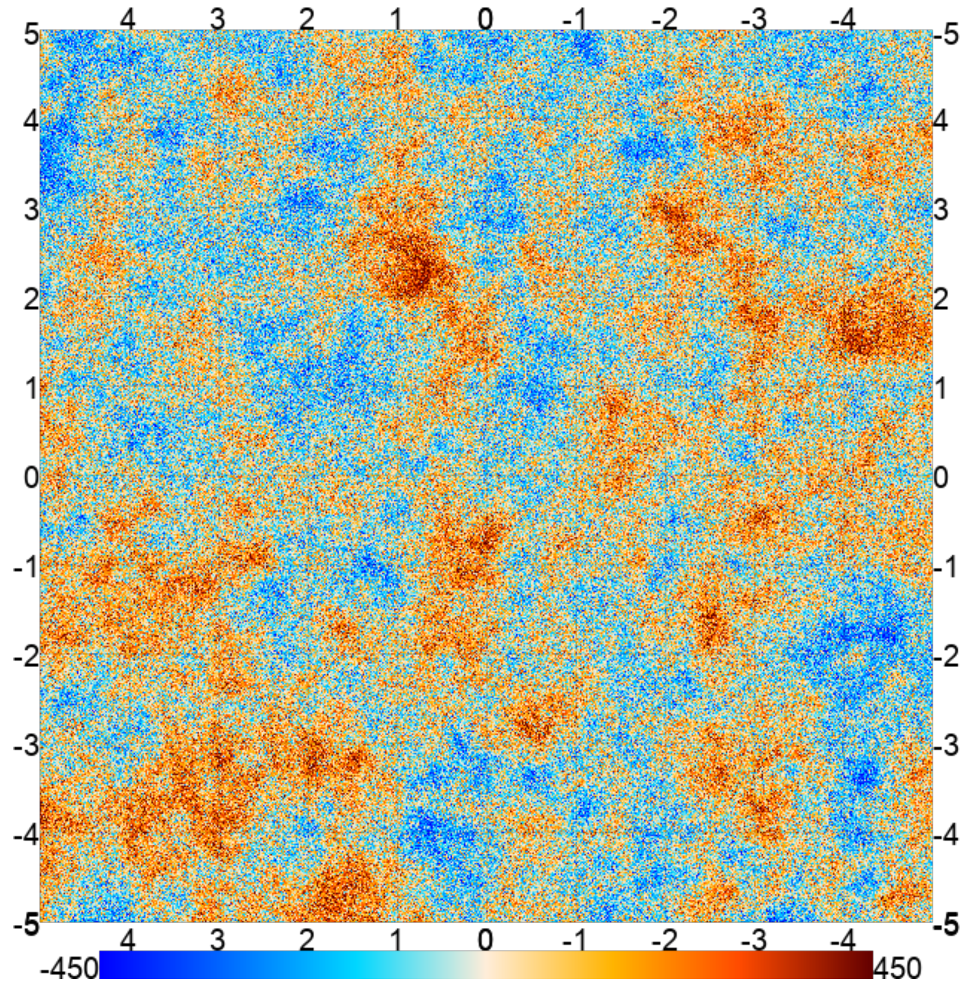


Cross-correlations

$$\delta T_1^{\rm obs}(\hat{n}) \quad = \quad \delta T^{\rm CMB}(\hat{n}) + n_1(\hat{n})$$



$$\delta T_2^{\rm obs}(\hat{n}) \quad = \quad \delta T^{\rm CMB}(\hat{n}) + n_2(\hat{n})$$

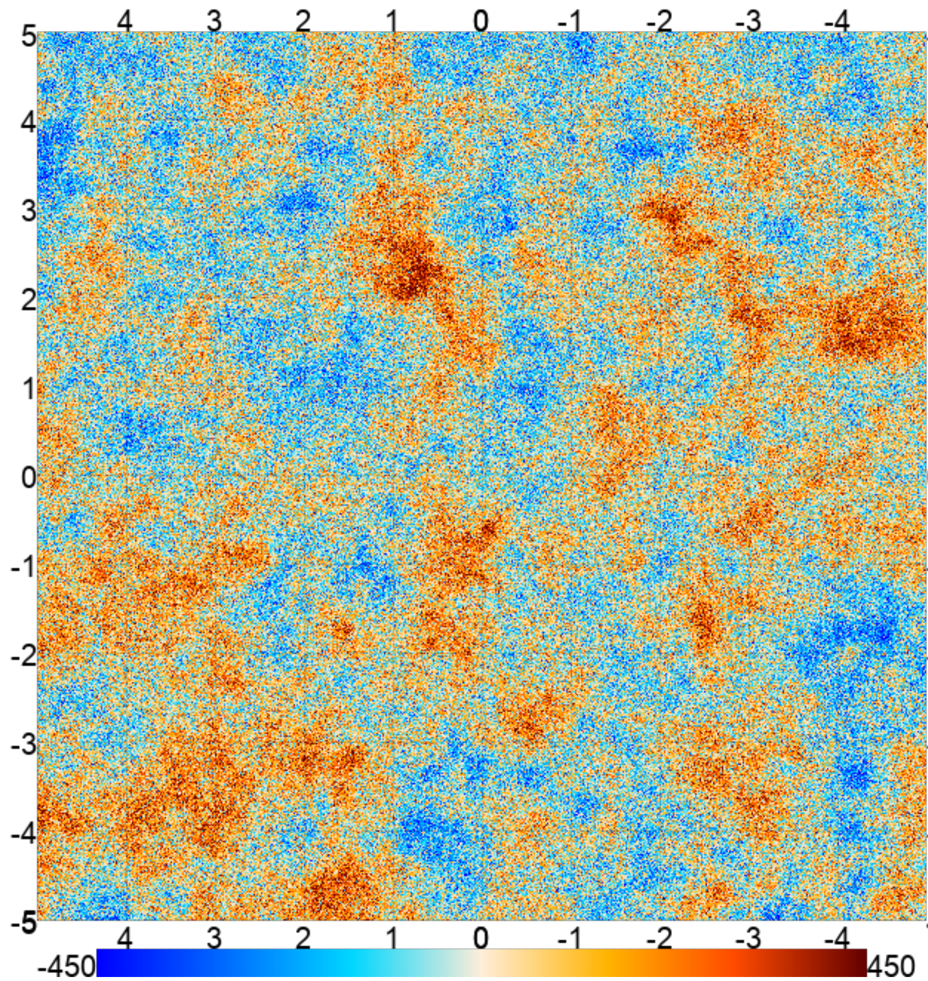


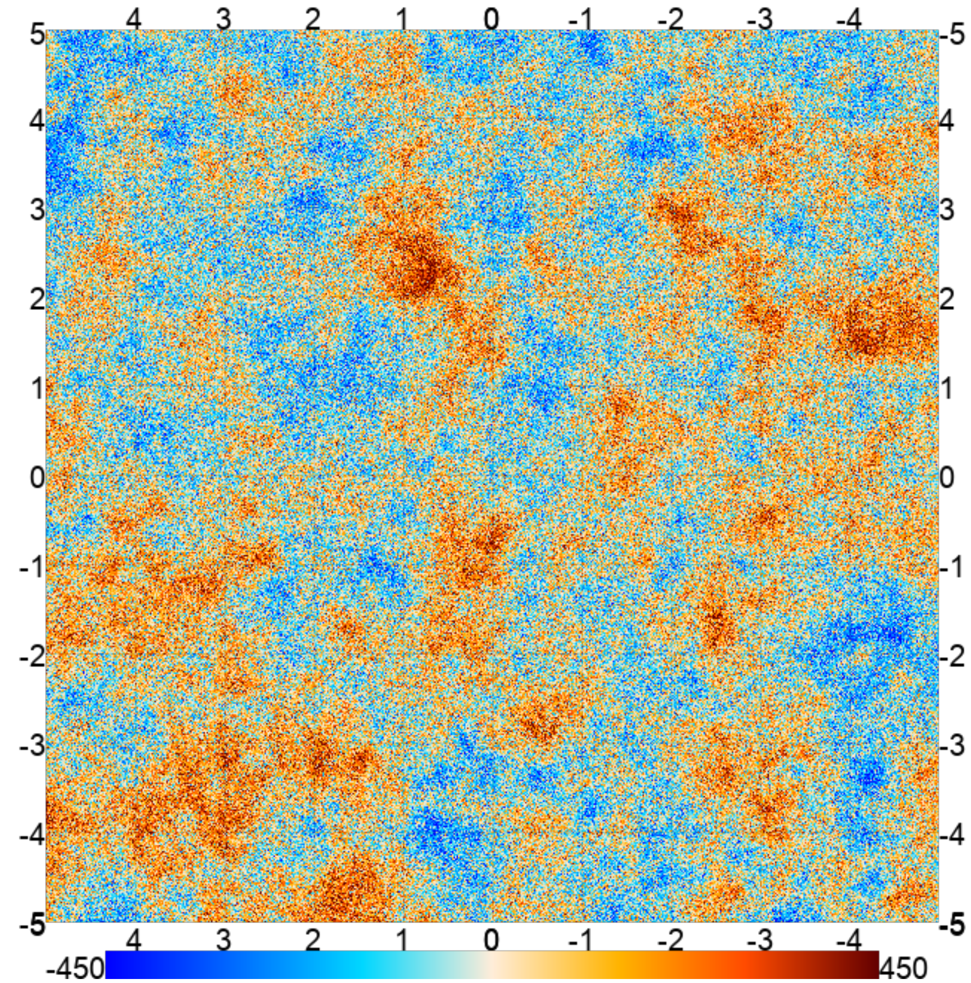
$$\begin{aligned} a_{1,\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{1,\ell m} \\ a_{2,\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{2,\ell m} \end{aligned}$$

Cross-correlations

$$\delta T_1^{\rm obs}(\hat{n}) \;=\; \delta T^{\rm CMB}(\hat{n}) + n_1(\hat{n})$$

$$\delta T_2^{\rm obs}(\hat{n}) \;=\; \delta T^{\rm CMB}(\hat{n}) + n_2(\hat{n})$$





$$\begin{aligned} a_{1,\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{1,\ell m} \\ a_{2,\ell m}^{\rm obs} &= a_{\ell m}^{\rm CMB} + n_{2,\ell m} \end{aligned}$$

$$\begin{aligned} \langle \hat{C}_\ell^{\rm obs} \rangle &= \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{1,\ell m}^{\rm obs} a_{2,\ell m}^{{\rm obs},*} \rangle = C_\ell^{\rm CMB} + \frac{1}{2\ell+1} \langle n_{1,\ell m} n_{2,\ell m}^{*} \rangle \\ &= C_\ell^{\rm CMB} \end{aligned}$$

Cross-correlations

Variance $$\sigma^2(\hat{C}_\ell^{\rm obs}) \;\;=\;\; \frac{2}{2\ell+1}\left(C_\ell^2 + 2C_\ell N_\ell + 2N_\ell^2\right)$$

Vs

$$\sigma^2(\hat{C}_\ell^{\rm obs}) = \frac{2}{2\ell+1}(C_\ell + N_\ell)^2$$

Cross-correlations

Variance $$\sigma^2(\hat{C}_\ell^{\rm obs}) \;=\; \frac{2}{2\ell+1}\left(C_\ell^2 + 2C_\ell N_\ell + 2N_\ell^2\right)$$

Vs

$$\sigma^2(\hat{C}_\ell^{\rm obs}) = \frac{2}{2\ell+1}(C_\ell + N_\ell)^2$$

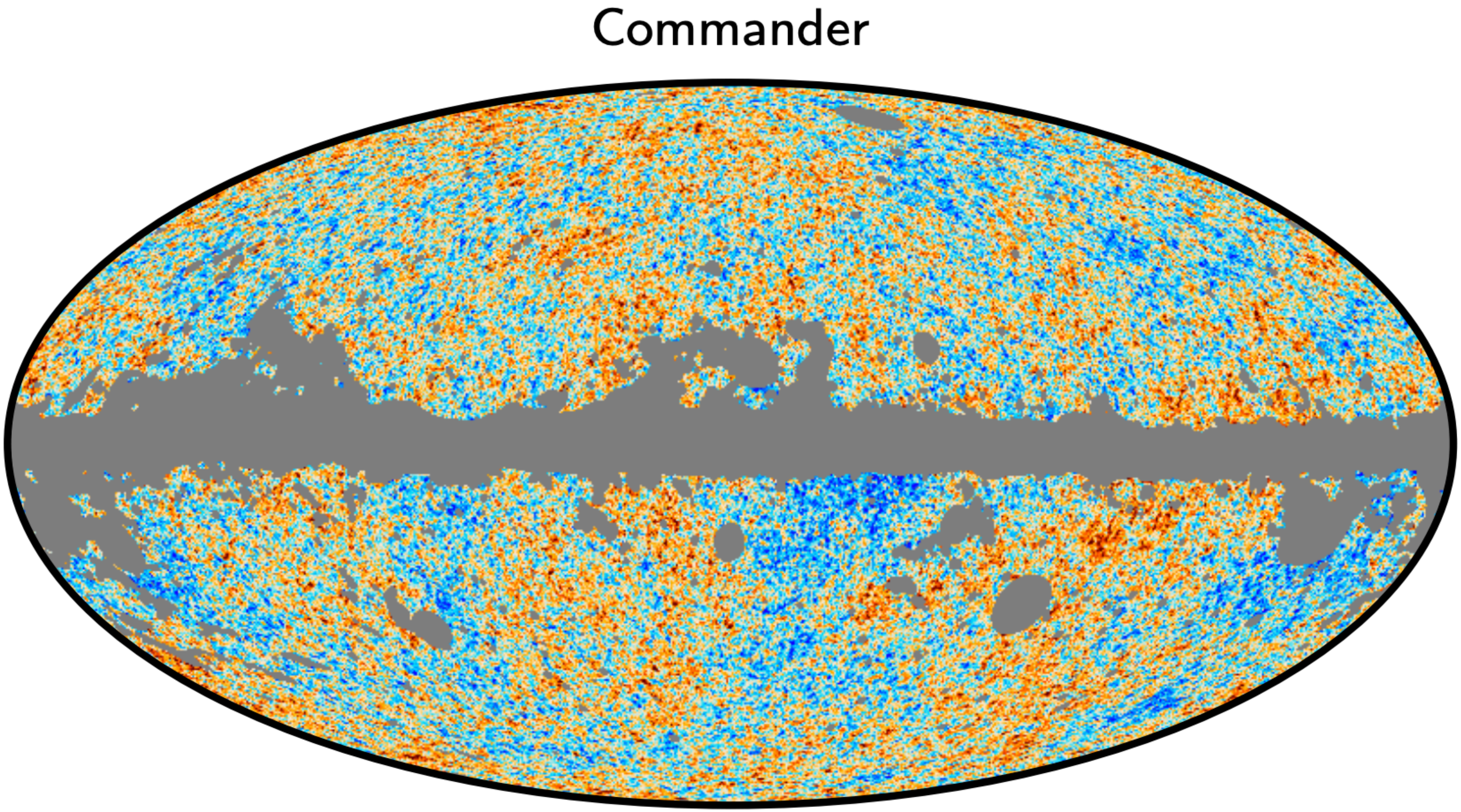
In general, for K data splits

$$\langle \hat{C}_\ell^{\rm obs} \rangle \;=\; \frac{1}{K(K-1)/2}\sum_{i\neq j}\frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell}\langle a_{i,\ell m}^{\rm obs} a_{j,\ell m}^{\rm obs,*}\rangle = C_\ell^{\rm CMB}$$

$$\sigma^2(\hat{C}_\ell^{\rm obs}) \;=\; \frac{2}{2\ell+1}\left(C_\ell^2 + 2C_\ell N_\ell + \frac{K}{K-1}N_\ell^2\right)$$

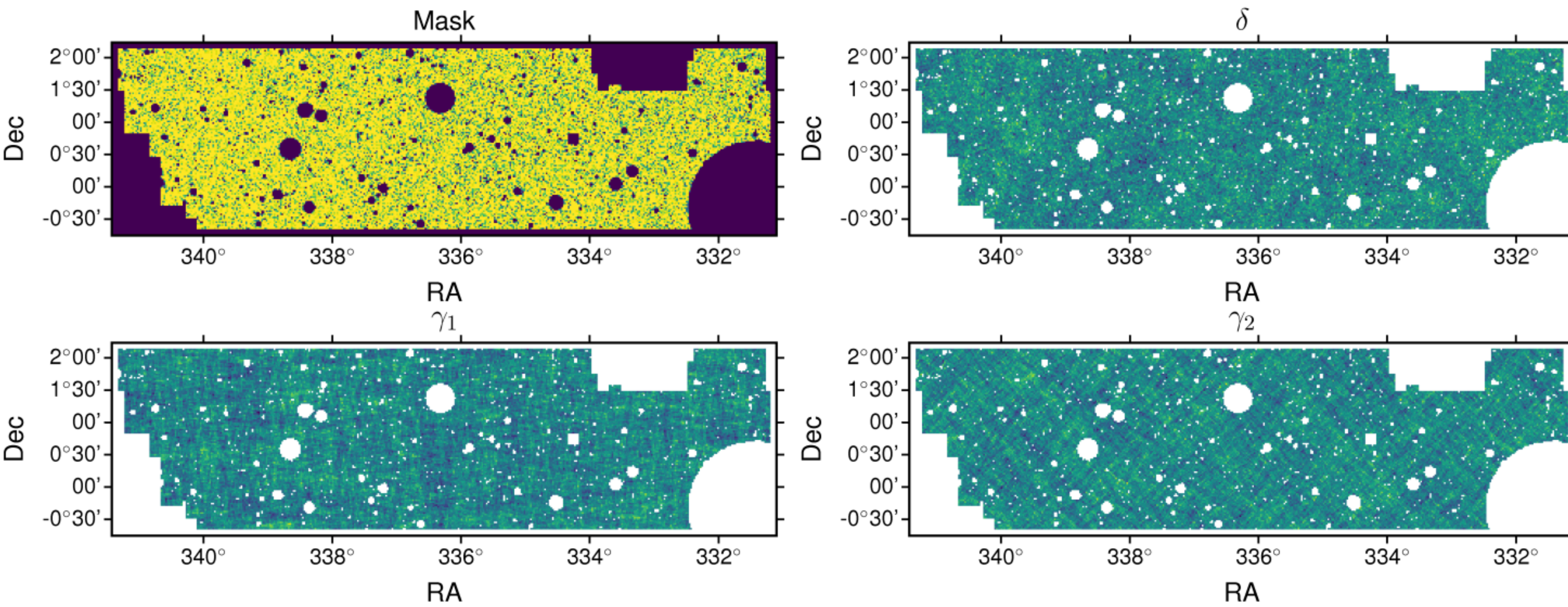
Impact of the Mask

Impact of the Mask



https://www.aanda.org/articles/aa/pdf/2016/10/aa25936-15.pdf

Impact of the Mask



https://arxiv.org/pdf/1809.09603.pdf

Impact of the Mask

$$
\begin{aligned}
\delta T^{\mathrm{masked}}(\hat{n}) &= W(\hat{n})[\delta T^{\mathrm{CMB}}(\hat{n}) + n(\hat{n})] \\
&= W(\hat{n})\delta T^{\mathrm{fullsky}}(\hat{n})
\end{aligned}
$$

Impact of the Mask

$$\begin{aligned}\delta T^{\rm masked}(\hat{n}) &= W(\hat{n})[\delta T^{\rm CMB}(\hat{n}) + n(\hat{n})] \\ &= W(\hat{n})\delta T^{\rm fullsky}(\hat{n})\end{aligned}$$

$$a_{\ell m}^{\rm masked} = \int W(\hat{n})\delta T^{\rm fullsky}(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n}$$

Impact of the Mask

$$\begin{aligned}\delta T^{\rm masked}(\hat{n}) &= W(\hat{n})[\delta T^{\rm CMB}(\hat{n}) + n(\hat{n})] \\ &= W(\hat{n})\delta T^{\rm fullsky}(\hat{n})\end{aligned}$$

$$a_{\ell m}^{\rm masked} = \int W(\hat{n})\delta T^{\rm fullsky}(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n}$$

$$\boxed{\delta T^{\rm fullsky}(\hat{n}) = \sum_{\ell' m'} a_{\ell' m'}^{\rm fullsky} Y_{\ell' m'}(\hat{n})}$$

Impact of the Mask

$$
\begin{aligned}
\delta T^{\mathrm{masked}}(\hat{n}) &= W(\hat{n})[\delta T^{\mathrm{CMB}}(\hat{n}) + n(\hat{n})] \\
&= W(\hat{n})\delta T^{\mathrm{fullsky}}(\hat{n})
\end{aligned}
$$

$$
a_{\ell m}^{\mathrm{masked}} = \int W(\hat{n})\delta T^{\mathrm{fullsky}}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n}
$$

$$
\boxed{\delta T^{\mathrm{fullsky}}(\hat{n}) = \sum_{\ell' m'} a_{\ell' m'}^{\mathrm{fullsky}} Y_{\ell' m'}(\hat{n})}
$$

$$
a_{\ell m}^{\mathrm{masked}} = \sum_{\ell' m'} a_{\ell' m'}^{\mathrm{fullsky}} \int W(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n}
$$

Impact of the Mask

$$
\begin{aligned}
a_{\ell m}^{\text{masked}} &= \sum_{\ell' m'} a_{\ell' m'}^{\text{fullsky}} \int W(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\
&= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}^{\text{fullsky}}
\end{aligned}
$$

Impact of the Mask

$$
\begin{aligned}
a_{\ell m}^{\text{masked}} &= \sum_{\ell' m'} a_{\ell' m'}^{\text{fullsky}} \int W(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\
&= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}^{\text{fullsky}}
\end{aligned}
$$

If W(n)=1, this integral equals $\delta_{\ell \ell'} \delta_{m m'}$

The mask breaks the homogeneity of the field and introduces coupling between different modes

Impact of the Mask

$$
\begin{aligned}
a_{\ell m}^{\text{masked}} &= \sum_{\ell' m'} a_{\ell' m'}^{\text{fullsky}} \int W(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\
&= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}^{\text{fullsky}}
\end{aligned}
$$

If we know the shape of the mask, we can calculate the matrix K analytically

$$
\begin{aligned}
K_{\ell m, \ell' m'} &= \int W(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n} \\
&= \sum_{\ell", m"} w_{\ell", m"} \int Y_{\ell" m"}(\hat{n}) Y_{\ell' m'}(\hat{n}) Y_{\ell m}^{*}(\hat{n}) d\hat{n}
\end{aligned}
$$

$$
\begin{aligned}
\int d\hat{n} Y_{\ell_1 m_1}^{*}(\hat{n}) Y_{\ell_3 m_3}(\hat{n}) Y_{\ell_2 m_2}(\hat{n}) &= (-1)^{m_1} \left[\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi} \right]^{1/2} \\
&\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ -m_1 & m_2 & m_3 \end{pmatrix}.
\end{aligned}
$$

pspy

`pspy` is a cosmology code for calculating CMB power spectra and covariance matrices. See the python example notebooks for an introductory set of examples on how to use the package.

pypi v1.4.4 | license BSD | build passing | docs passing | launch binder

- Free software: BSD license
- `pspy` documentation: https://pspy.readthedocs.io.
- Scientific documentation: https://pspy.readthedocs.io/en/latest/scientific_doc.pdf

Installing

```
$ pip install pspy [--user]
```

You can test your installation by running

NaMaster

build passing | docs passing | coverage 97%

NaMaster is a C library, Python module and standalone program to compute full-sky angular cross-power spectra of masked fields with arbitrary spin and an arbitrary number of known contaminants using a pseudo-Cl (aka MASTER) approach. The code also implements E/B-mode purification and is available in both full-sky and flat-sky modes.

Installation

There are different ways to install NaMaster. In rough order of complexity, they are:

Conda forge

Unless you care about optimizing the code, it's worth giving this one a go. The conda recipe for NaMaster is currently hosted on conda-forge (infinite kudos to Mat Becker for this). In this case, installing NaMaster means simply running:

```
conda install -c conda-forge namaster
```

+ Xpol, Xpure, Polspice

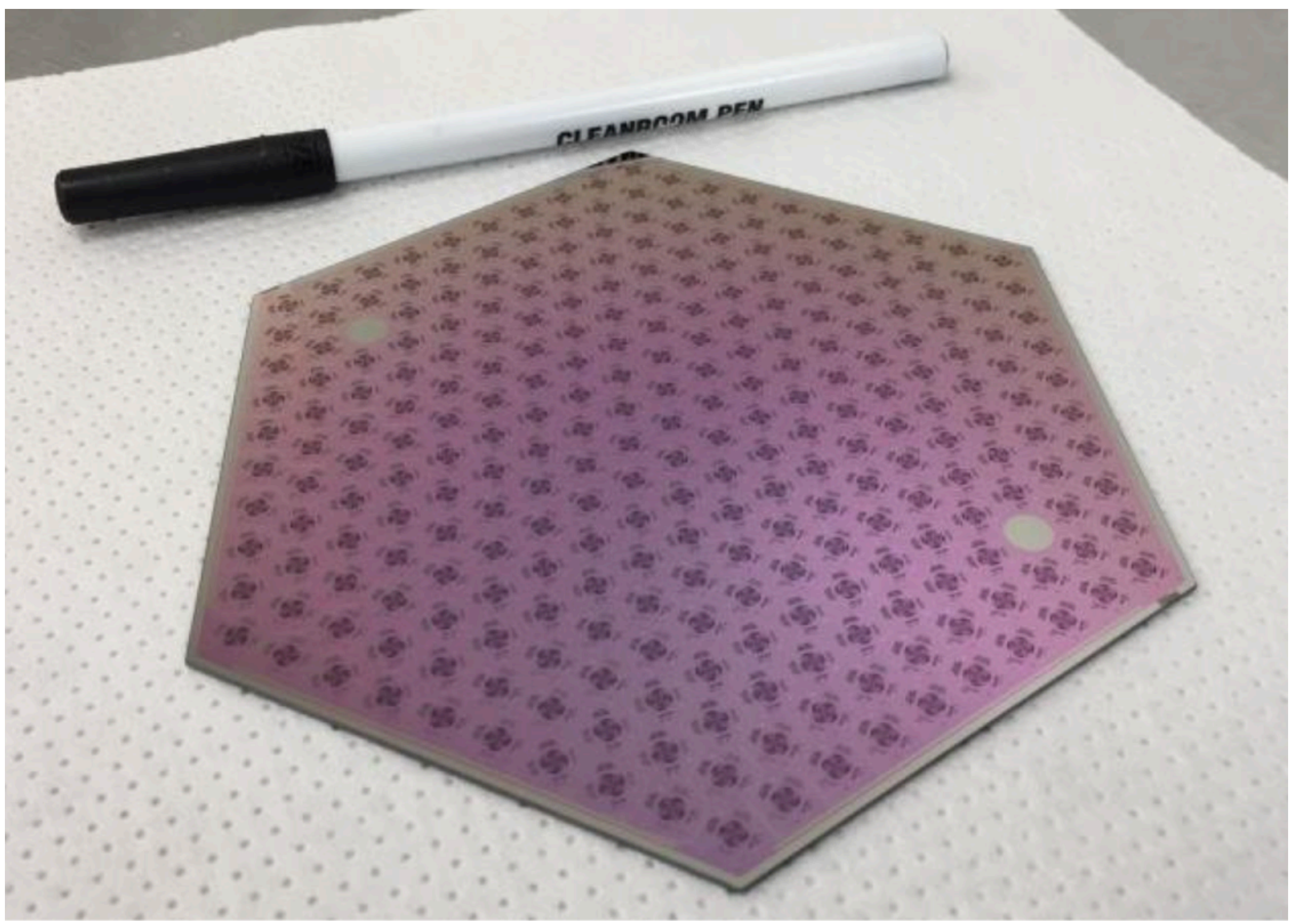
A word about higher spin fields.

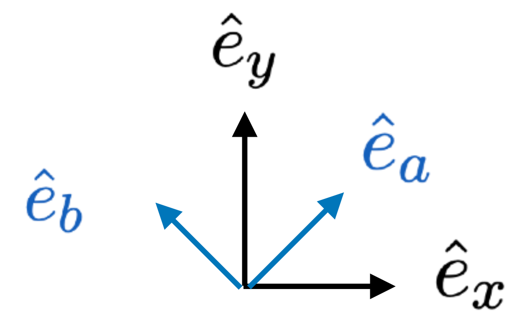
The density field and the CMB temperature field are scalar fields (i.e., spin 0); their values do not depend on the coordinate system used.

Conversely, the polarization field of the cosmic microwave background, as well as the gravitational shear field of galaxies, are spin-2 fields; their values at a given point depend on the coordinate system.

A word about higher spin fields.

For example, the polarization field is defined in terms of the Stokes parameters, which quantify the direction of polarization of the electric field E of the CMB photons.

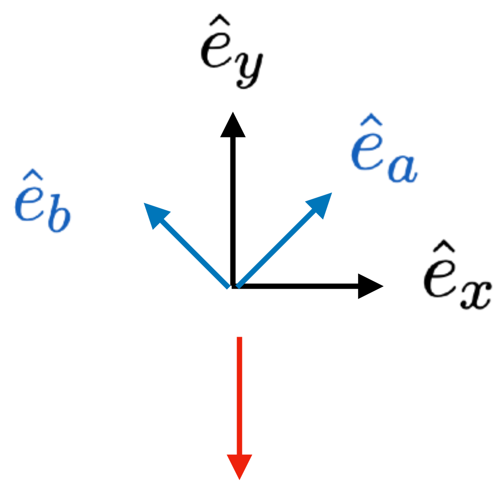




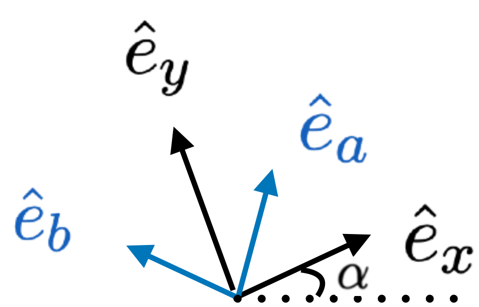
$$
\begin{aligned}
Q(\hat{n}) &= |E_x(\hat{n})|^2 - |E_y(\hat{n})|^2 \\
U(\hat{n}) &= |E_a(\hat{n})^2| - |E_b(\hat{n})^2|
\end{aligned}
$$

A word about higher spin fields.

The Stokes parameters are not invariant under a change of coordinates.



$$
\begin{aligned}
Q(\hat{n}) &= |E_x(\hat{n})]^2 - |E_y(\hat{n})]^2 \\
U(\hat{n}) &= |E_a(\hat{n})^2| - |E_b(\hat{n})^2|
\end{aligned}
$$



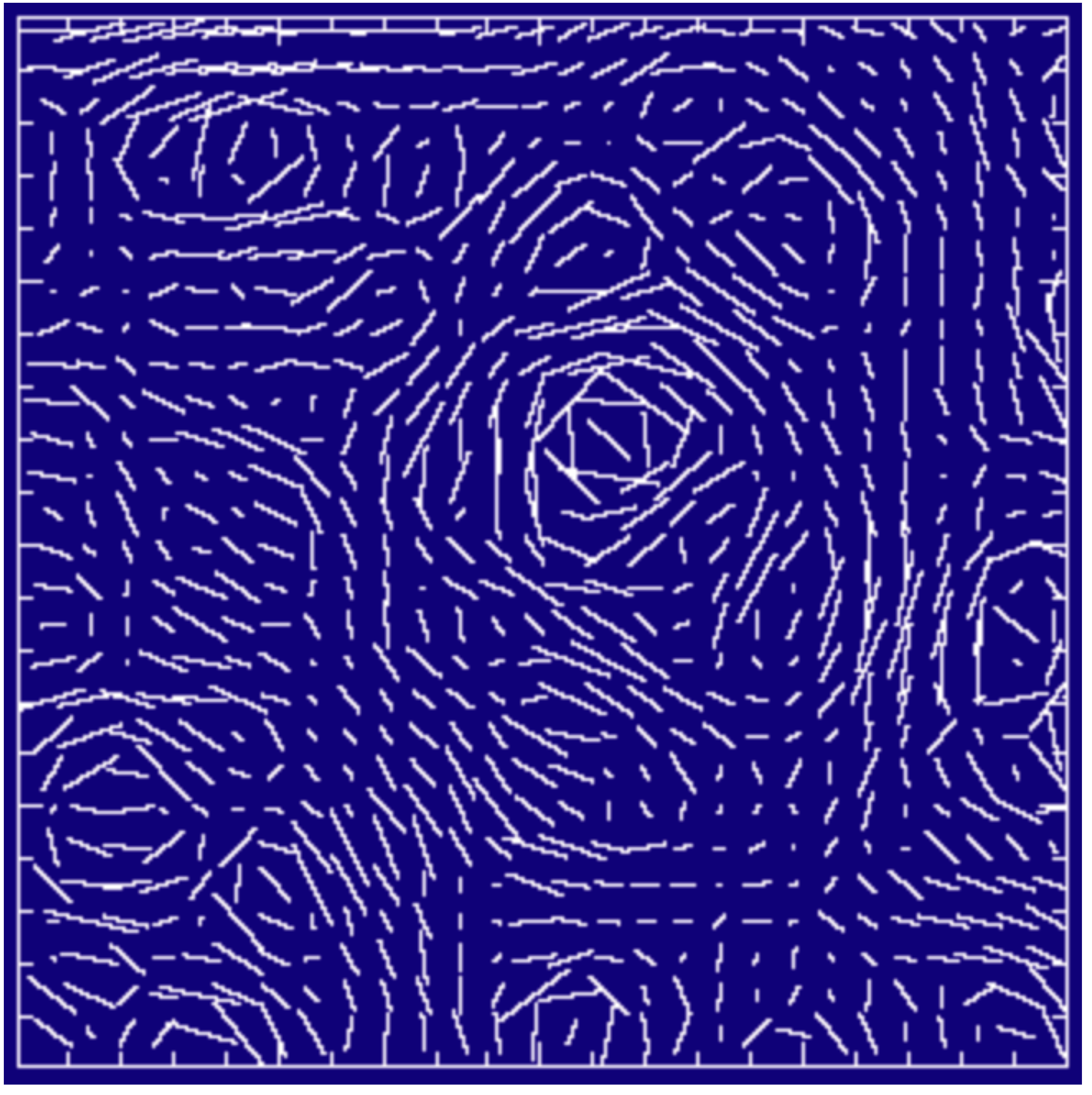
$$
\begin{aligned}
Q'(\hat{n}) &= Q(\hat{n})\cos(2\alpha) + U(\hat{n})\sin(2\alpha) \\
U'(\hat{n}) &= -Q(\hat{n})\sin(2\alpha) + U(\hat{n})\cos(2\alpha)
\end{aligned}
$$

If $\alpha = \pi, Q' = Q, U' = U$

This property makes the polarization field a spin-2 field.

This is also why the polarization field can be represented as a "headless" vector field.

Polarisation field



E and B modes decomposition

To account for the transformation of the Stokes parameters under a change of coordinates, they can be written in the form of a tensor.

$$P_{ab}(\hat{n}) = \frac{1}{2}\begin{pmatrix} Q(\hat{n}) & U(\hat{n}) \\ U(\hat{n}) & -Q(\hat{n}) \end{pmatrix}$$

E and B modes decomposition

To account for the transformation of the Stokes parameters under a change of coordinates, they can be written in the form of a tensor.

$$P_{ab}(\hat{n}) = \frac{1}{2}\begin{pmatrix} Q(\hat{n}) & U(\hat{n}) \\ U(\hat{n}) & -Q(\hat{n}) \end{pmatrix}$$

Just as a vector field can always be decomposed into a scalar part and a rotational part.

$$\vec{V} = \vec{\nabla}\phi + \vec{\nabla} \times A$$

E and B modes decomposition

To account for the transformation of the Stokes parameters under a change of coordinates, they can be written in the form of a tensor.

$$P_{ab}(\hat{n}) = \frac{1}{2}\begin{pmatrix} Q(\hat{n}) & U(\hat{n}) \\ U(\hat{n}) & -Q(\hat{n}) \end{pmatrix}$$

Just as a vector field can always be decomposed into a scalar part and a rotational part.

$$\vec{V} = \vec{\nabla}\phi + \vec{\nabla} \times A$$

A spin-2 tensor field can always be decomposed into the sum of two fields: a scalar field and a pseudoscalar field.

$$P_{ab}(\hat{n}) = \mathcal{E}_{ab}E(\hat{n}) + \mathcal{B}_{ab}B(\hat{n})$$

E modes **B modes**

E and B modes decomposition

To account for the transformation of the Stokes parameters under
a change c

$$\mathcal{E}_{ab}E = \left(-\partial_a\partial_b + \frac{1}{2}\delta_{ab}\nabla^2\right)E$$

Just as a ve
into a scala

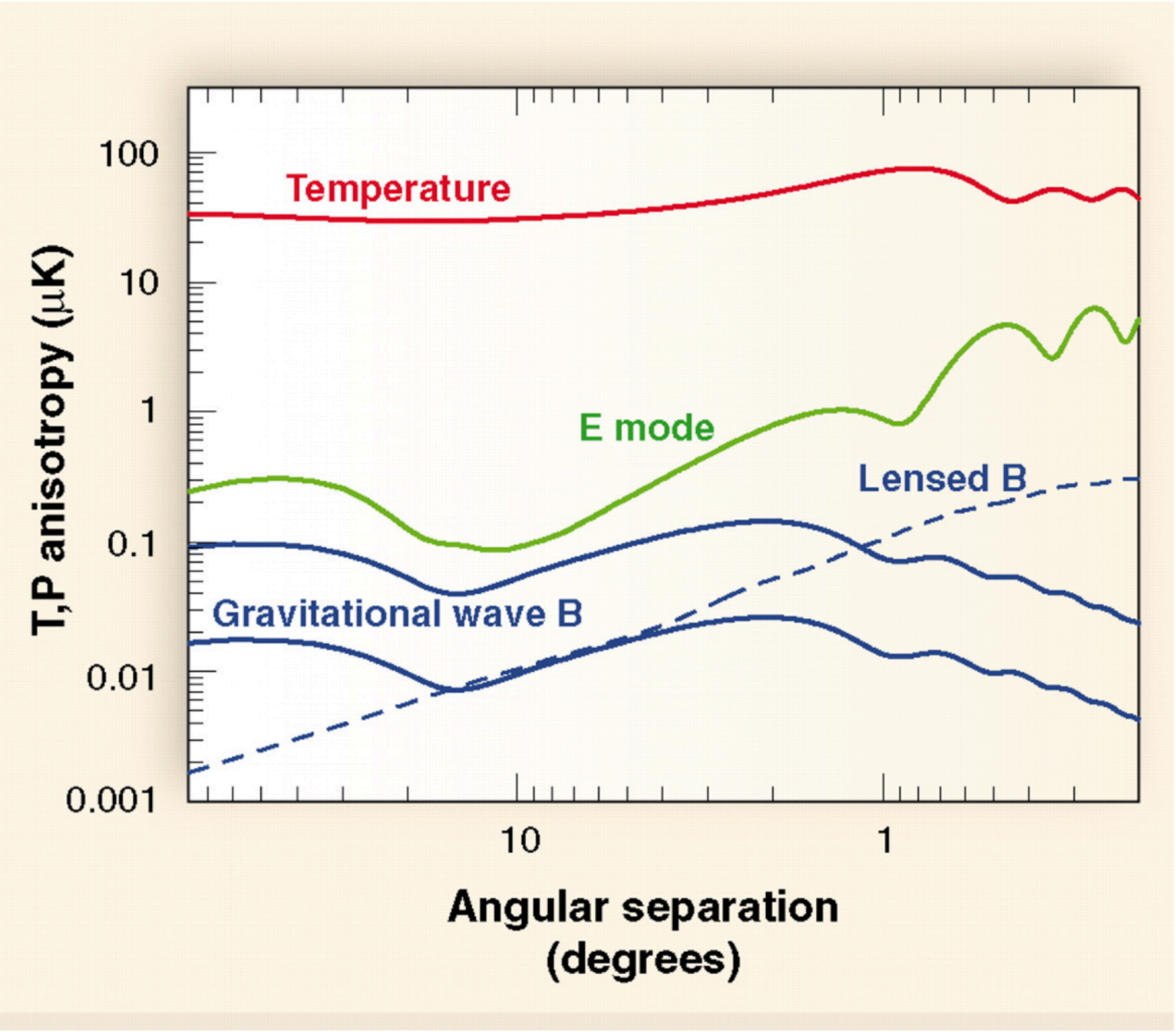
$$\mathcal{B}_{ab}B = \frac{1}{2}\left(\epsilon_{ac}\partial^c\partial_b + \epsilon_{bc}\partial^c\partial_a\right)B$$

A spin-2 ten
sum of two fields: a scalar field and a pseudoscalar field.

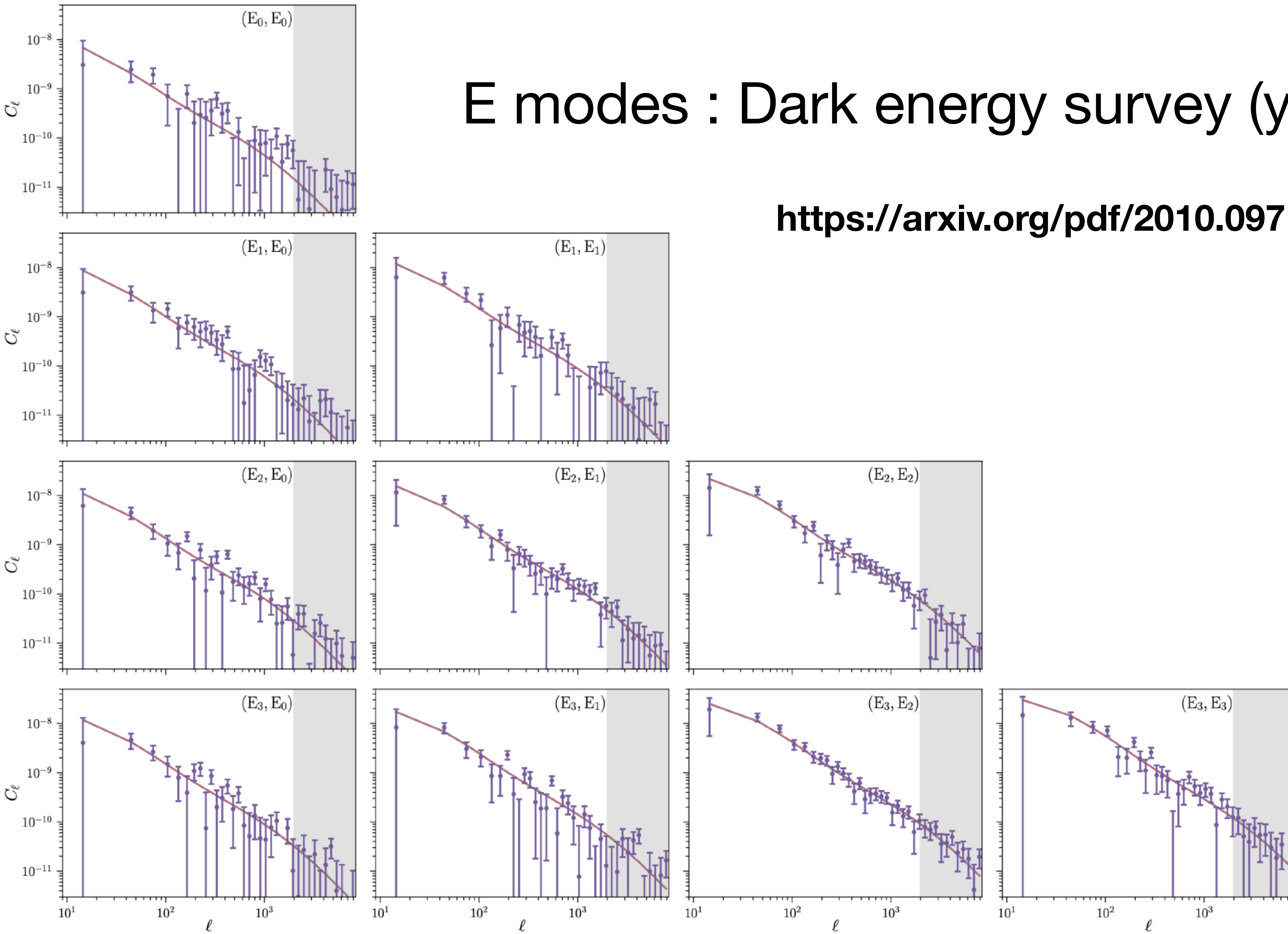
$$P_{ab}(\hat{n}) = \mathcal{E}_{ab}E(\hat{n}) + \mathcal{B}_{ab}B(\hat{n})$$

E modes **B modes**

This decomposition is truly useful: most physical phenomena in the primordial Universe generate only E-modes. A possible detection of B-mode polarization could be the signature of primordial gravitational waves produced during the inflationary phase of the Universe.



Regarding the gravitational shear of galaxies, only E-modes are generated at first order, so B-modes can be used to test for the presence of possible contaminations of the signal (astrophysical or instrumental).



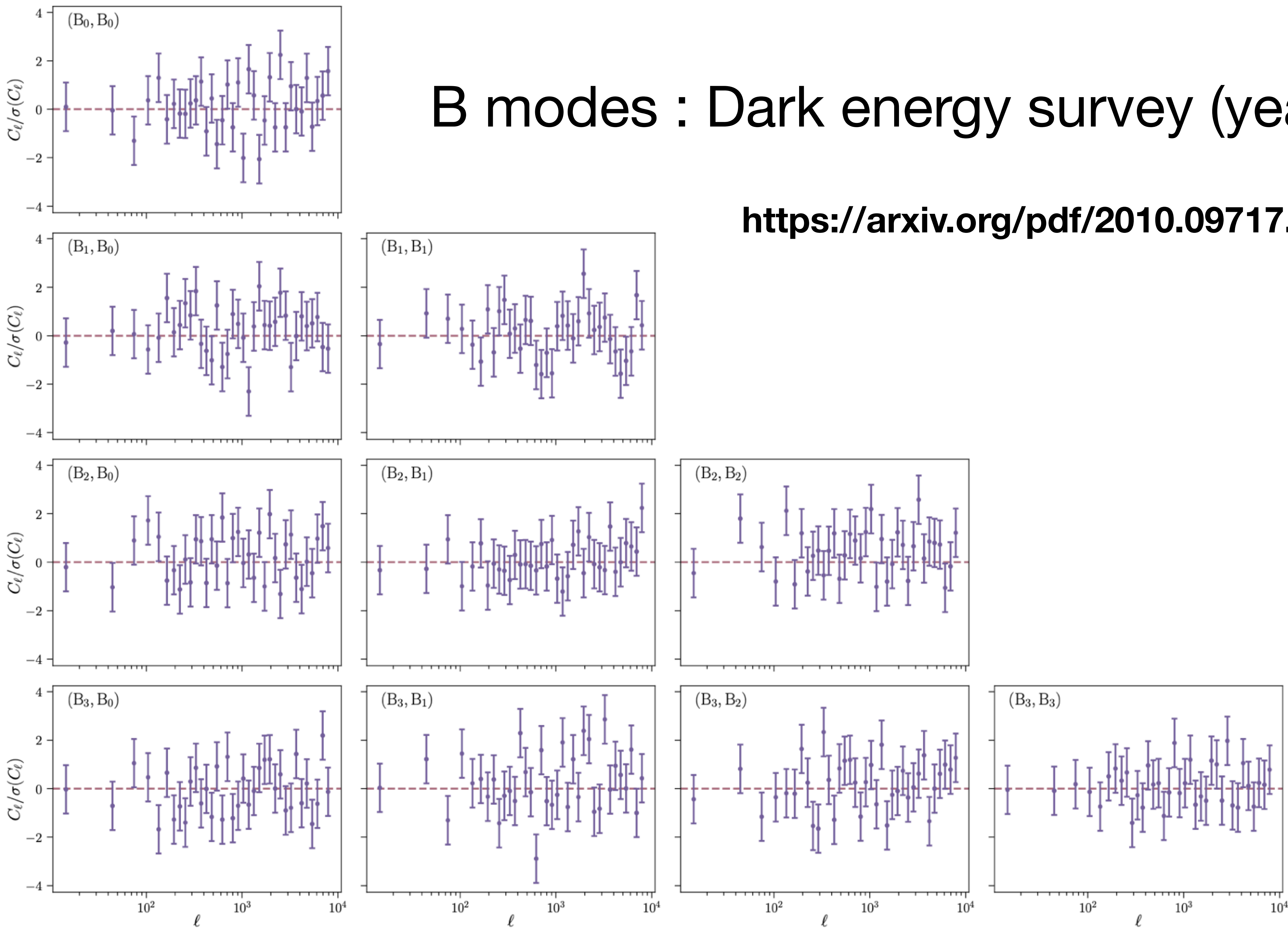
E modes : Dark energy survey (year 1)

https://arxiv.org/pdf/2010.09717.pdf

Regarding the gravitational shear of galaxies, only E-modes are generated at first order, so B-modes can be used to test for the presence of possible contaminations of the signal (astrophysical or instrumental).

B modes : Dark energy survey (year 1)

https://arxiv.org/pdf/2010.09717.pdf



For the numerical tutorial:

https://github.com/thibautlouis/TD