

Les méthodes statistiques en cosmologie

Thibaut Louis



Plan

- I) Rappel sur les probabilités**
- II) Les fonctions de corrélation à deux points**
- III) Maximum de vraisemblance et estimation de paramètres**

Le paradoxe du faux positif

Gerd Gigerenzer

Test du cancer mammaire:

Une femme de 50 ans (sans symptômes) fait une radiographie mammaire, elle est testée positive et veut connaître sa probabilité d'avoir un cancer ?

Sensibilité: probabilité que le test soit positif pour une personne malade

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

Spécificité: probabilité d'obtenir un test négatif chez les non-malades

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

Prévalence de la maladie dans la population:

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

Quelle est le meilleur estimé de sa probabilité
d'avoir un cancer? :

- A) 90%
- C) 10%

- B) 50%
- D) 1%

Le paradoxe du faux positif

Gerd Gigerenzer

Test du cancer mammaire:

Une femme de 50 ans (sans symptômes) fait une radiographie mammaire, elle est testée positive et veut connaître sa probabilité d'avoir un cancer ?

Sensibilité: probabilité que le test soit positif pour une personne malade

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

Spécificité: probabilité d'obtenir un test négatif chez les non-malades

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

Prévalence de la maladie dans la population:

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

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)} \qquad \text{Théorème de Bayes}$$

Le paradoxe du faux positif

Gerd Gigerenzer

Test du cancer mammaire:

Une femme de 50 ans (sans symptômes) fait une radiographie mammaire, elle est testée positive et veut connaître sa probabilité d'avoir un cancer ?

Sensibilité: probabilité que le test soit positif pour une personne malade

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

Spécificité: probabilité d'obtenir un test négatif chez les non-malades

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

Prévalence de la maladie dans la population:

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

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

Le paradoxe du faux positif

Gerd Gigerenzer

Test du cancer mammaire:

Une femme de 50 ans (sans symptômes) fait une radiographie mammaire, elle est testée positive et veut connaître sa probabilité d'avoir un cancer ?

Sensibilité: probabilité que le test soit positif pour une personne malade

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

Spécificité: probabilité d'obtenir un test négatif chez les non-malades

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

Prévalence de la maladie dans la population:

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

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

Le paradoxe du faux positif

Gerd Gigerenzer

Test du cancer mammaire:

Une femme de 50 ans (sans symptômes) fait une radiographie mammaire, elle est testée positive et veut connaître sa probabilité d'avoir un cancer ?

Sensibilité: probabilité que le test soit positif pour une personne malade

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

Spécificité: probabilité d'obtenir un test négatif chez les non-malades

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

Prévalence de la maladie dans la 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
(probabilité à priori: avant l'acquisition des données)

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

Postérieur
(probabilité A POSTERIORI
étant donné le résultat du test)

Vraisemblance
(Likelihood)

Prior
(probabilité à priori: avant l'acquisition des données)

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

Postérieur
**(probabilité A POSTERIORI
étant donné le résultat du test)**

**Vraisemblance
(Likelihood)**

**Evidence
(marginal likelihood)**

Interlude: comment avoir l'air brillant grâce au théorème de Bayes en utilisant les cotes

Le théorème de Bayes est un peu difficile à calculer de tête

$$P(M|+) = \frac{P(+|M)P(M)}{P(+|M)P(M) + [1 - P(-|\neg M)] [1 - P(M)]}$$

Une version alternative peut être écrite en terme de cotes

$$P(M) = 1\% \rightarrow C(M) = 1 : 99$$

Interlude: comment avoir l'air brillant grâce au théorème de Bayes en utilisant les cotes

Le théorème de Bayes est un peu difficile à calculer de tête

$$P(M|+) = \frac{P(+|M)P(M)}{P(+|M)P(M) + [1 - P(-|\neg M)] [1 - P(M)]}$$

Une version alternative peut être écrite en terme de cotes

$$P(M) = 1\% \rightarrow C(M) = 1 : 99$$

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

Facteur de Bayes

Séparation claire du prior vis à vis de la précision de test

Interlude: comment avoir l'air brillant grâce au théorème de Bayes en utilisant les cotes

Le théorème de Bayes est un peu difficile à calculer de tête

$$P(M|+) = \frac{P(+|M)P(M)}{P(+|M)P(M) + [1 - P(-|\neg M)] [1 - P(M)]}$$

Une version alternative peut être écrite en terme de cotes

$$P(M) = 1\% \rightarrow C(M) = 1 : 99$$

$$\begin{aligned} C(M|+) &= C(M) \frac{P(+|M)}{P(+|\neg M)} \\ &= (1 : 99) \frac{90\%}{9\%} = 10 : 99 \end{aligned}$$

Variables et champs aléatoires

Malade/pas malade, positif/pas positif sont des concepts pas très quantitatifs, en cosmologie on aura tendance à étudier des variables prenant leurs valeurs dans \mathbb{R} ou \mathbb{N} . On utilisera les densités de probabilités, par exemple

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

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

La cosmologie aura aussi pour objet l'étude des champs stochastiques, où on généralisera le concept de probabilité pour un nombre infini de variables aléatoires

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

Propriétés

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

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

Moments de la variable aléatoire

**Espérance
(valeur moyenne)**

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

Propriétés

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

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

Moments de la variable aléatoire

**Espérance
(valeur moyenne)**

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

Variance

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

Propriétés

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

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

Moments de la variable aléatoire

**Espérance
(valeur moyenne)**

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

Variance

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

**Coefficient
d'asymétrie
(Skewness)**

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

**Coefficient
d'aplatissement
(Kurtosis)**

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

Propriétés

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

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

Moments de la variable aléatoire

**Espérance
(valeur moyenne)**

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

Variance

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

**Coefficient
d'asymétrie
(Skewness)**

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

**Coefficient
d'aplatissement
(Kurtosis)**

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

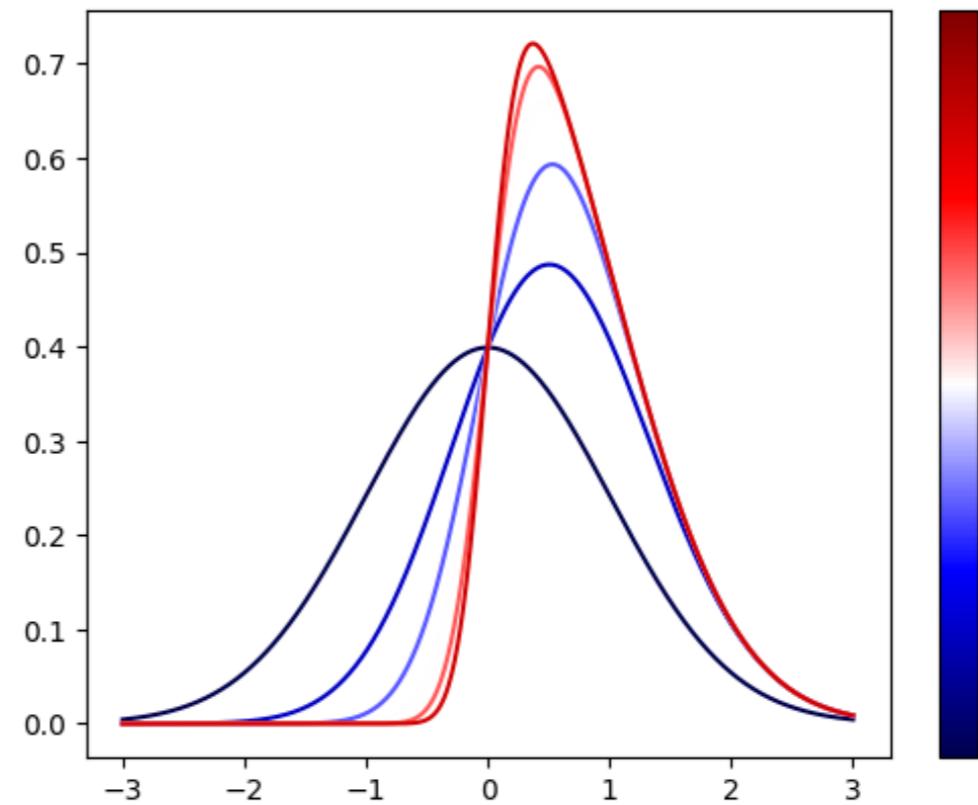
0 pour une variable
aléatoire Gaussienne

Non gaussianité

Loi normale asymétrique

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

$\alpha = 0 \longrightarrow$ **Gaussien**

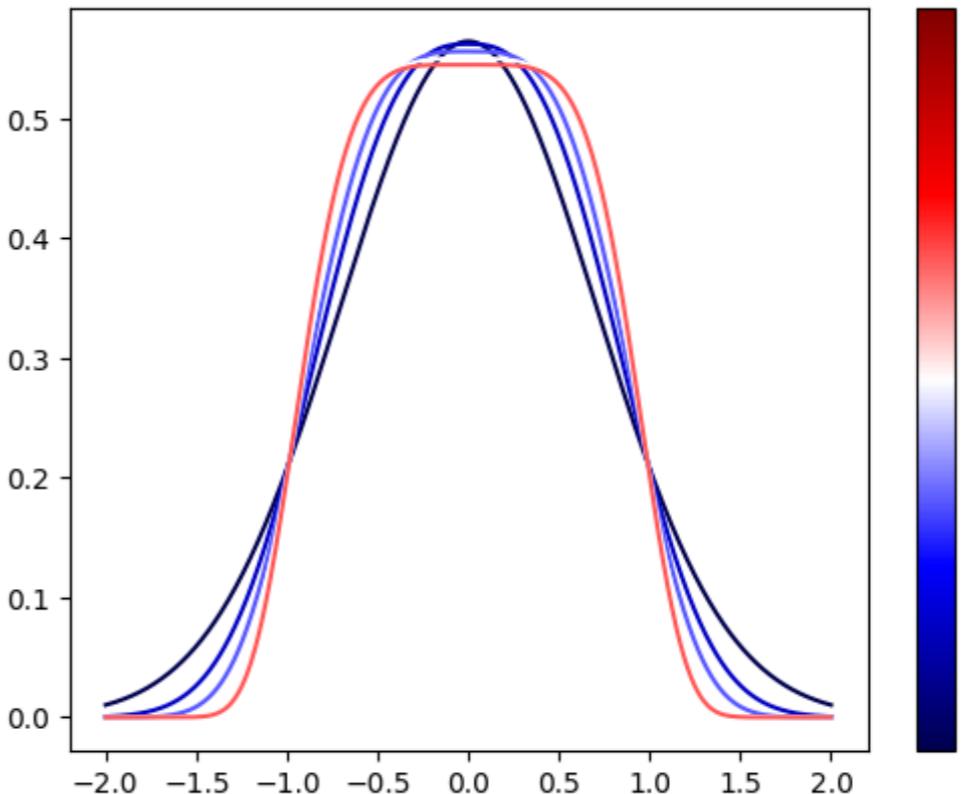


Skewness

Loi normale généralisée

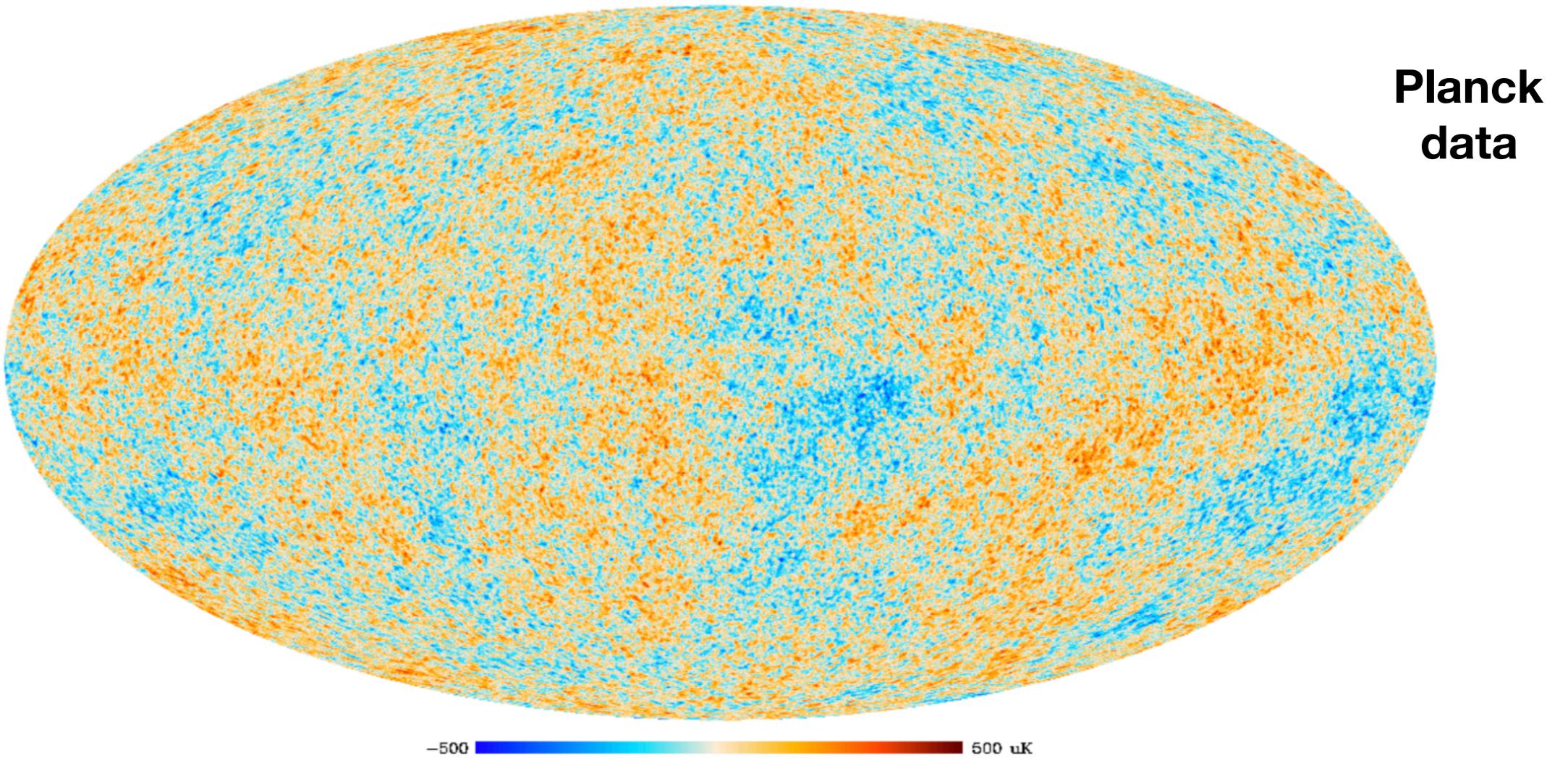
$$\mathcal{P}(x; \alpha\beta, \mu) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^\beta}$$

$\beta = 2 \longrightarrow$ **Gaussien**



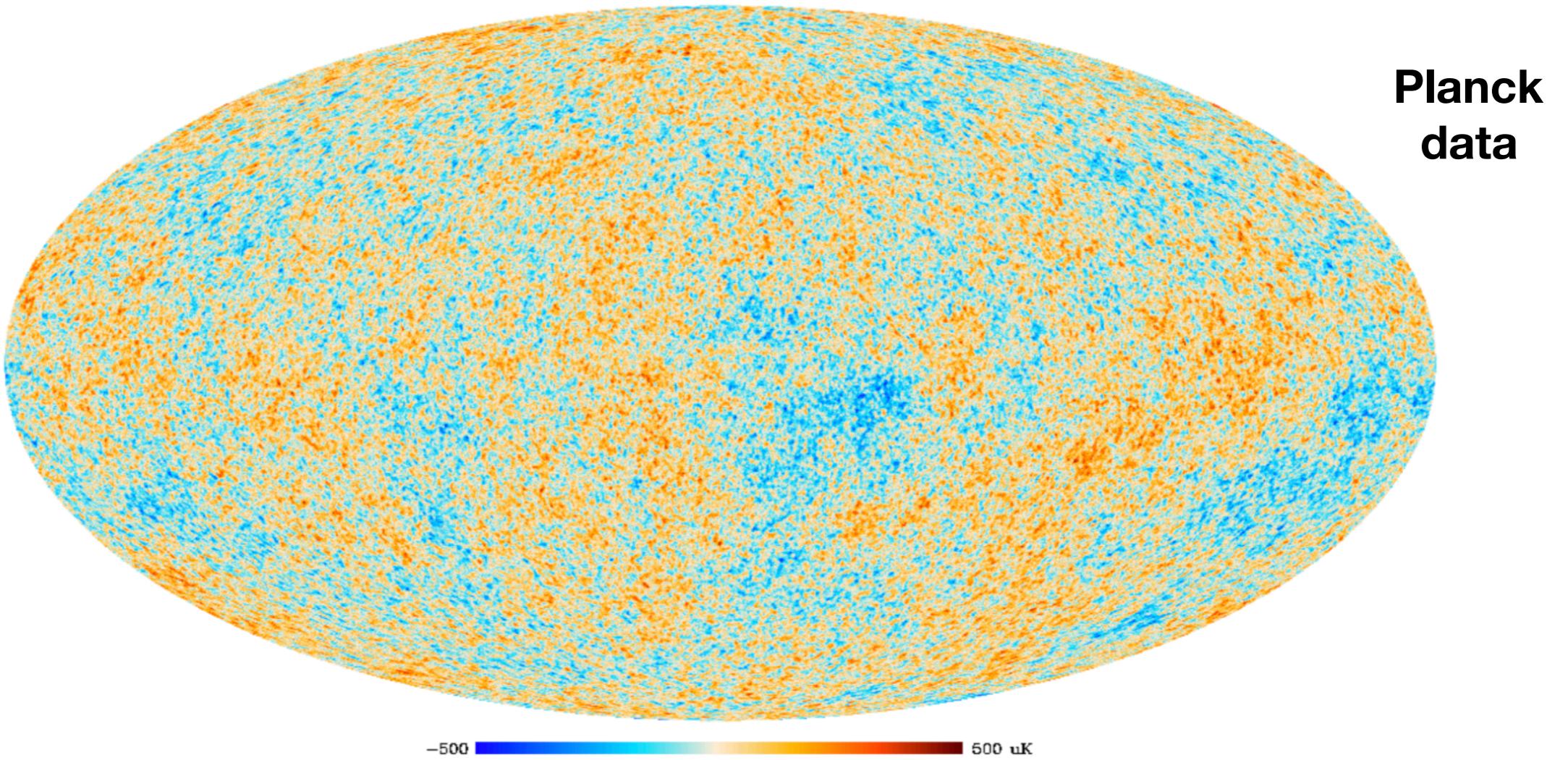
Kurtosis

Interlude: en pratique on ne connaît pas souvent a priori la densité de probabilité d'un processus physique, mais on peut l'inférer à partir de la mesure de ses moments



Le calcul des moments du CMB indique une distribution gaussienne, en particulier $\langle \delta T(\hat{n}_1)\delta T(\hat{n}_2)\delta T(\hat{n}_3) \rangle \sim 0$

Interlude: en pratique on ne connaît pas souvent a priori la densité de probabilité d'un processus physique, mais on peut l'inférer à partir de la mesure de ses moments



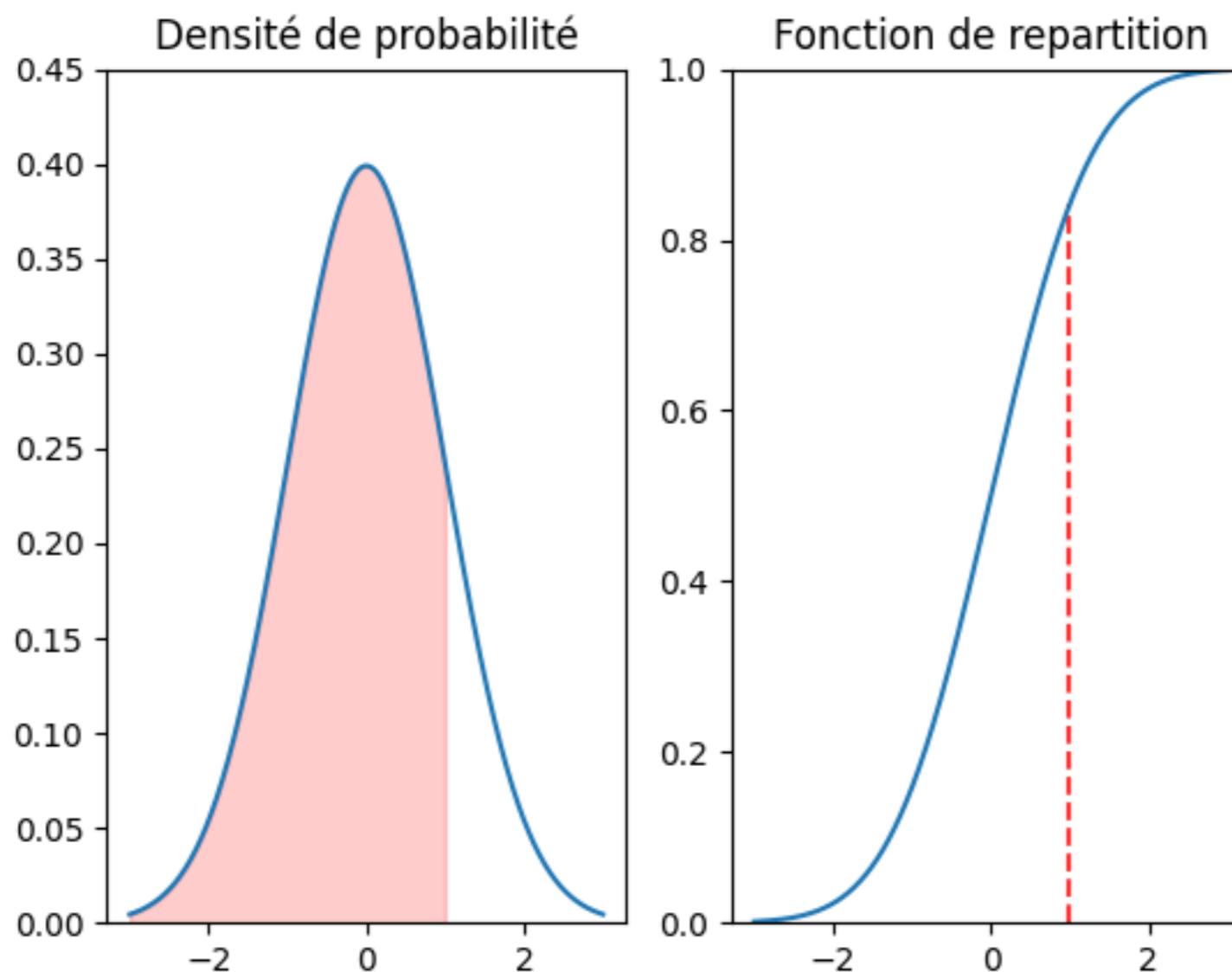
Le calcul des moments du CMB indique une distribution gaussienne, en particulier $\langle \delta T(\hat{n}_1)\delta T(\hat{n}_2)\delta T(\hat{n}_3) \rangle \sim 0$

→ Indication en faveur de l'inflation et contraintes sur les modèles

Cumulative distribution function

Fonction de répartition (cumulative distribution function)

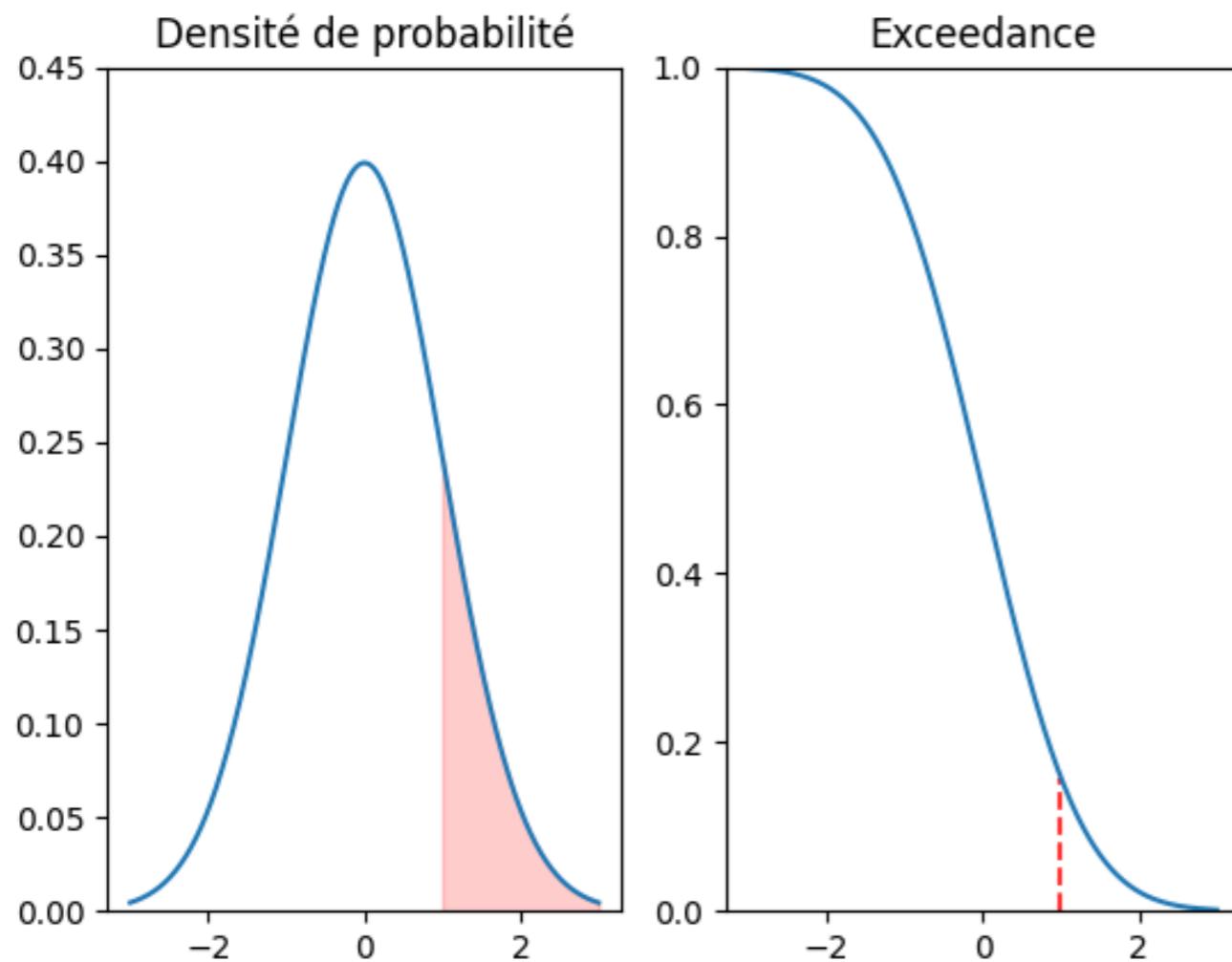
$$F_X(x) = \int_{-\infty}^x \mathcal{P}_X(x')dx' = P(-\infty < X \leq x) = P(X \leq x)$$



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**

Exemple I: le problème de la constante de Hubble

$$H_0^{\text{Planck}} = 67.36 \pm 0.54 \text{ km/s/Mpc}$$

$$H_0^{\text{Riess}} = 73.2 \pm 1.3 \text{ km/s/Mpc}$$

Quel est le degré de désaccord entre ces deux mesures ?

Pour quantifier ça, partons de l'hypothèse que les deux mesures sont non biaisées, indépendantes et que la distribution de probabilité associée à chacune de ces mesures est gaussienne.

Exemple I: le problème de la constante de Hubble

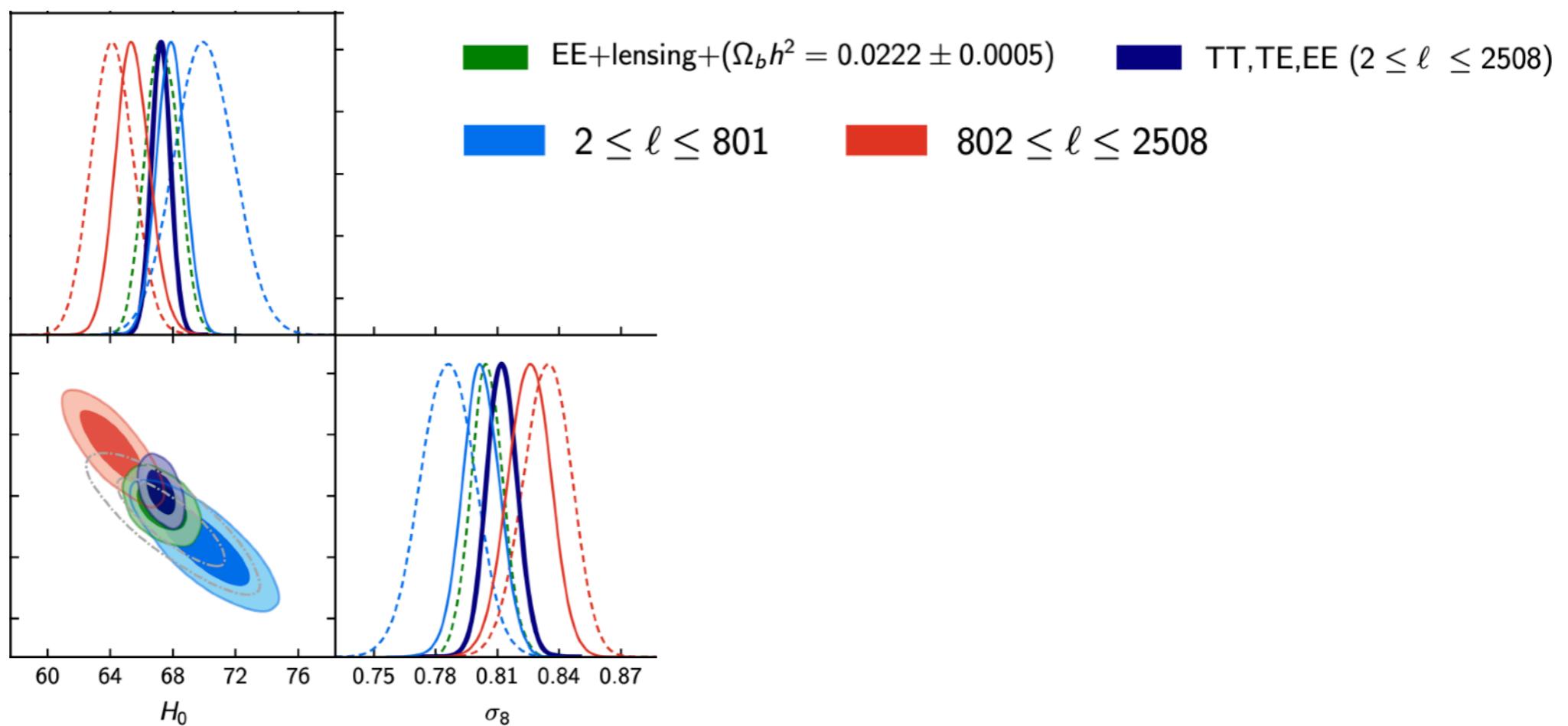
$$H_0^{\text{Planck}} = 67.36 \pm 0.54 \text{ km/s/Mpc}$$

$$H_0^{\text{Riess}} = 73.2 \pm 1.3 \text{ km/s/Mpc}$$

Quel est le degré de désaccord entre ces deux mesures ?

Pour quantifier ça, partons de l'hypothèse que les deux mesures sont non biaisées, indépendantes et que la distribution de probabilité associée à chacune de ces mesures est gaussienne.

Planck postérieur
sur H_0



Exemple I: le problème de la constante de Hubble

Définissons la nouvelle variable aléatoire

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

Elle suit une distribution:

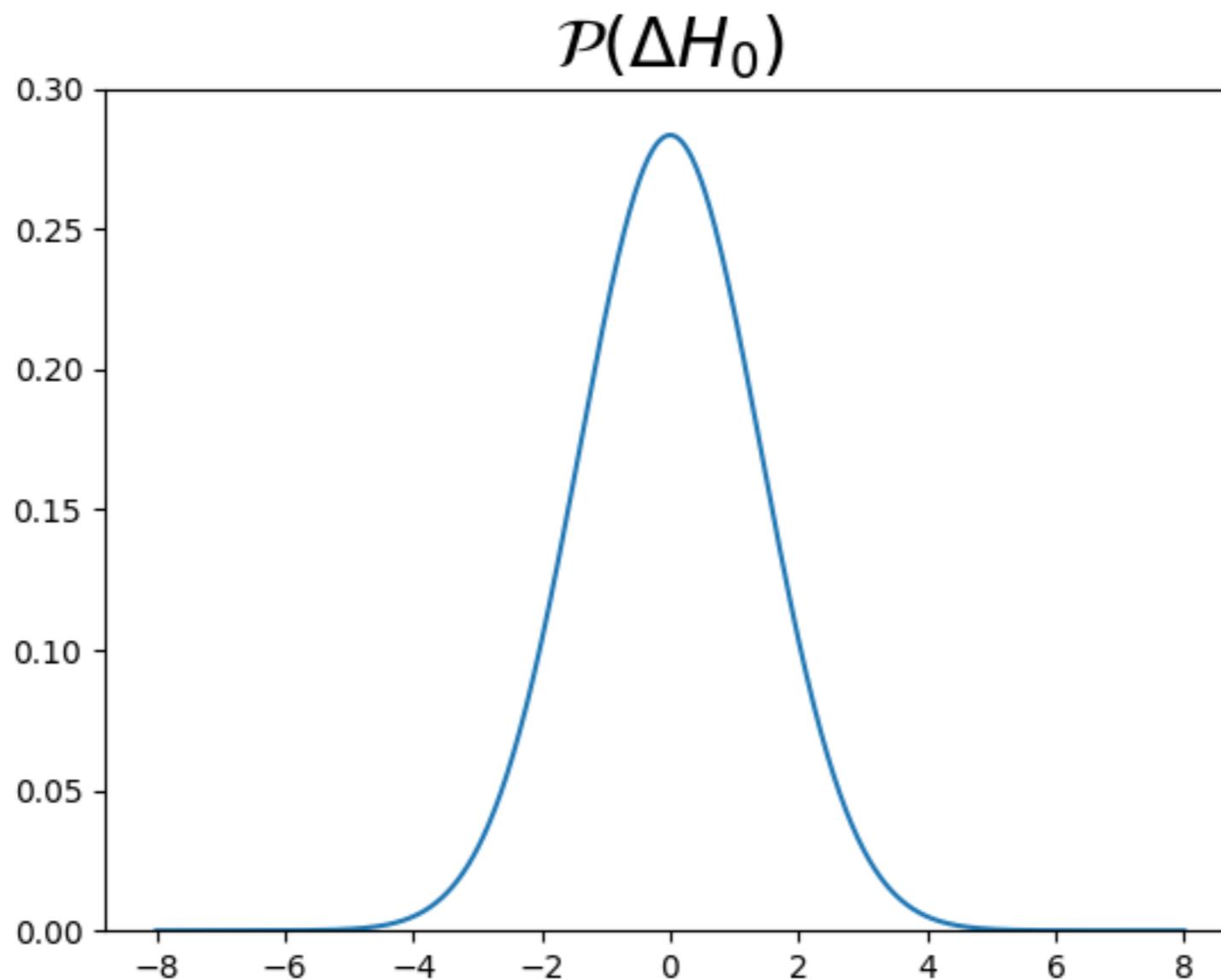
Gaussienne : Toute combinaison linéaire de nombres Gaussiens indépendants suit une distribution Gaussienne

De moyenne nulle : dans notre hypothèse où les deux mesures sont non biaisées

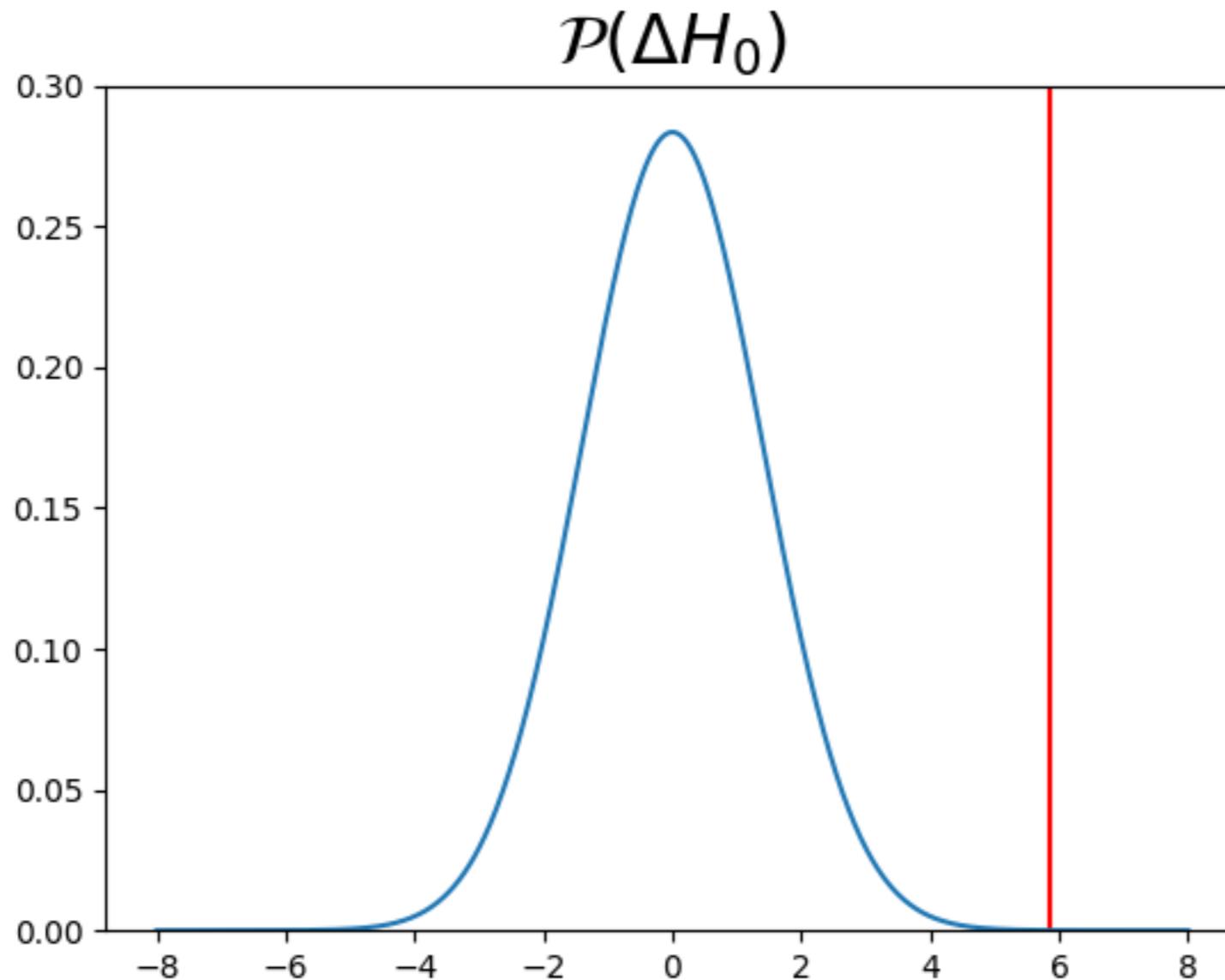
De variance : $\sigma^2(\Delta H_0) = \sigma^2(H_0^{\text{Riess}}) + \sigma^2(H_0^{\text{Planck}})$

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

Exemple I: le problème de la constante de Hubble



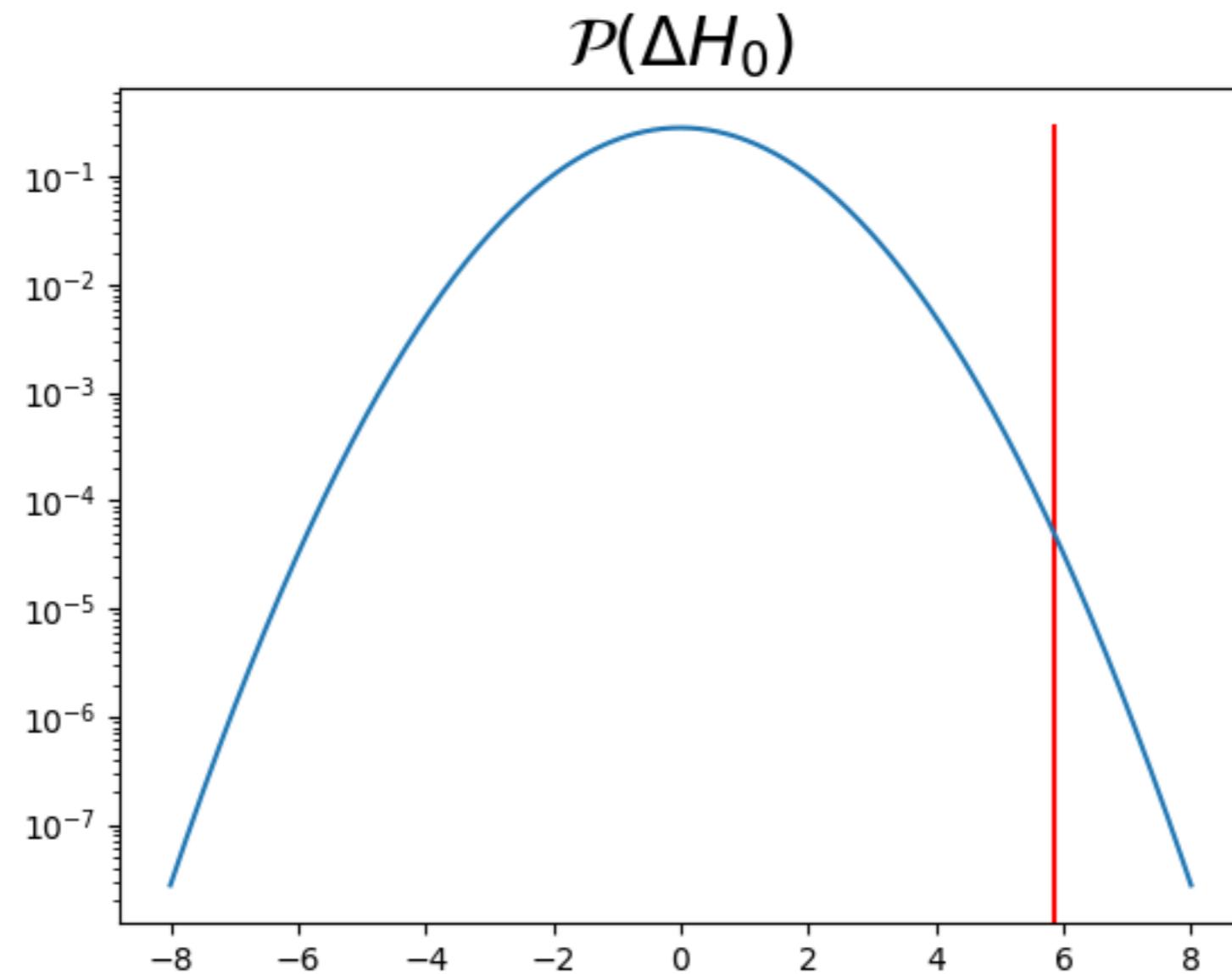
Exemple I: le problème de la constante de Hubble



$$H_0^{\text{Planck}} = 67.36 \pm 0.54 \text{ km/s/Mpc}$$

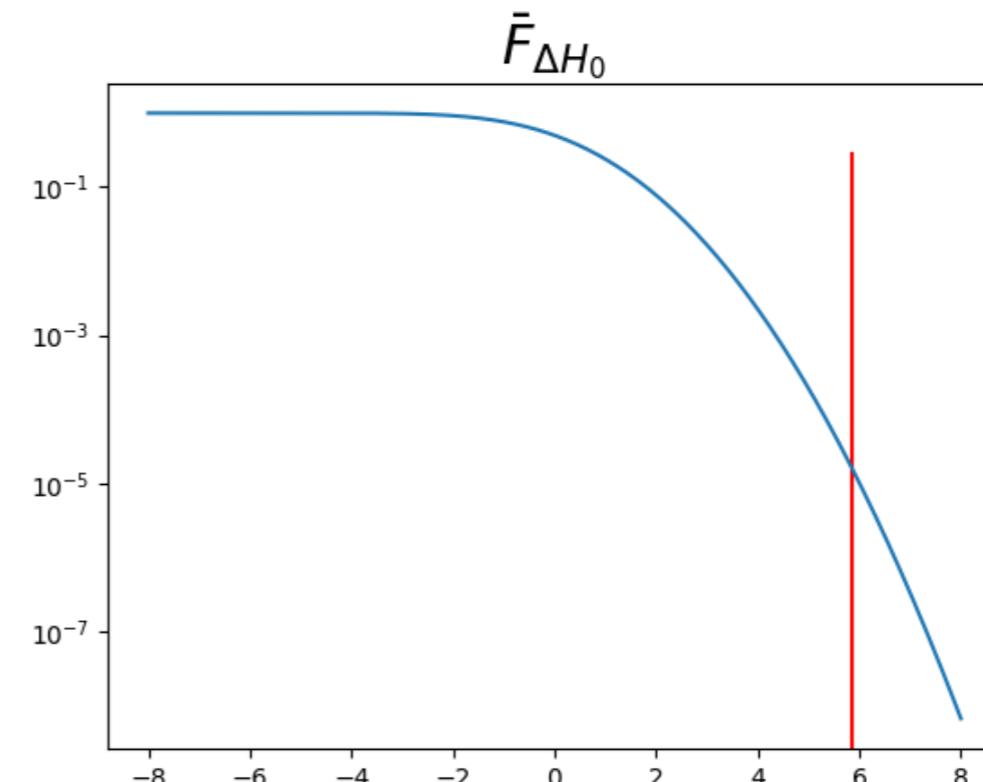
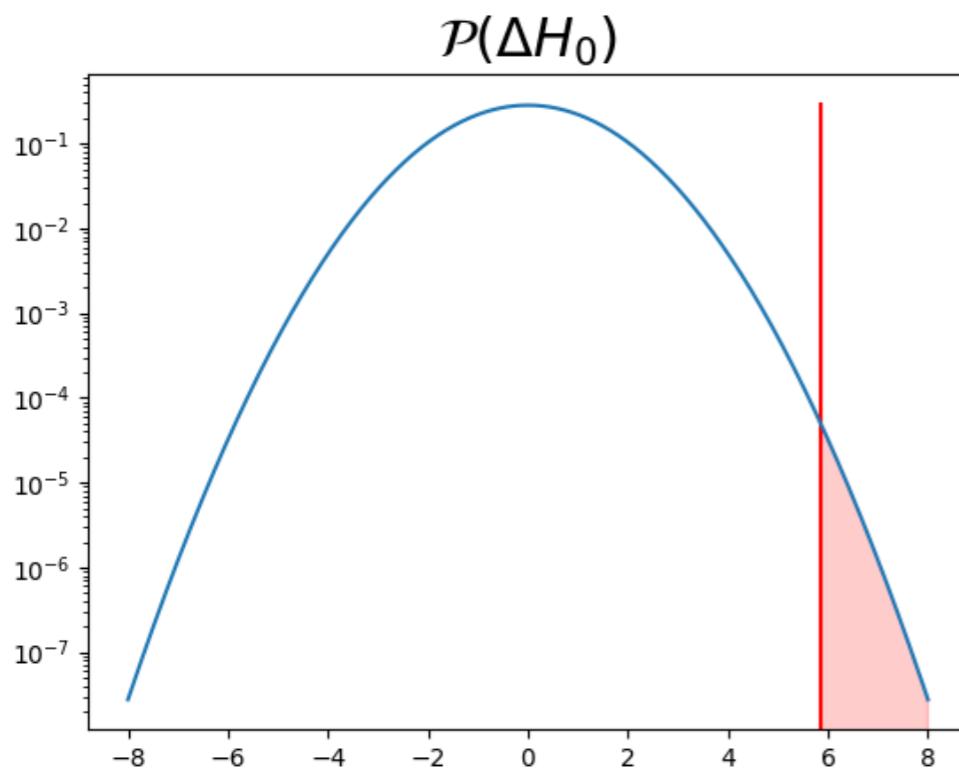
$$H_0^{\text{Riess}} = 73.2 \pm 1.3 \text{ km/s/Mpc}$$

Exemple 1: le problème de la constante de Hubble



Quelle est la probabilité, sous l'hypothèse nulle, d'obtenir une valeur plus extrême que celle observée ?

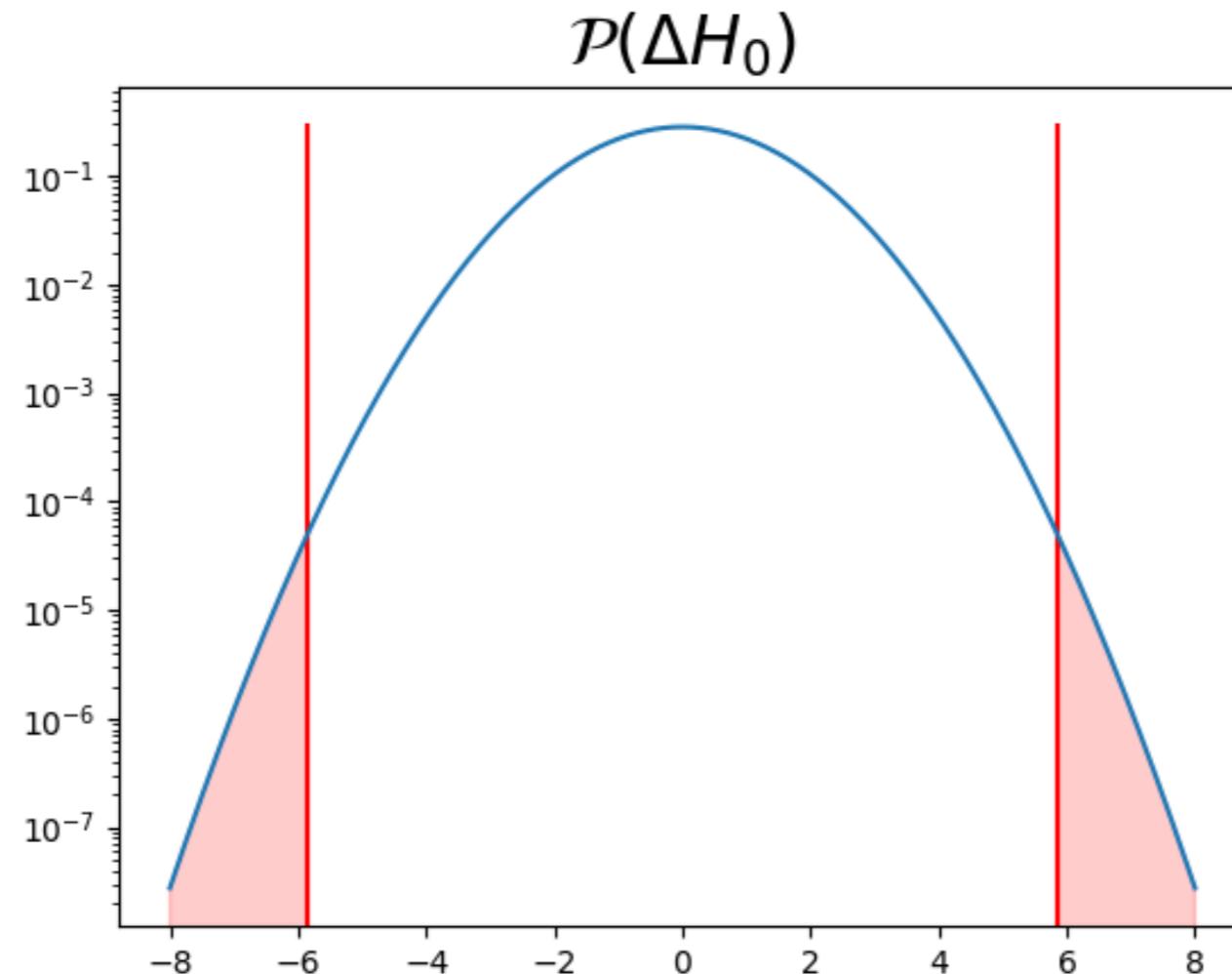
$$\begin{aligned}\bar{F}_{\Delta H_0}(x) = P(\Delta H_0 > x) &= 1 - \int_{-\infty}^x \frac{1}{\sigma_{\Delta H_0} \sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma_{\Delta H_0}^2}\right) dx' \\ &= 1 - \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma_{\Delta H_0}}\right)\right)\end{aligned}$$



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

Quelle est la probabilité, sous l'hypothèse nulle, d'obtenir la même valeur ou une valeur plus extrême que celle observée ?

Soyons honnêtes, le test aurait pu fluctuer dans les deux sens



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

Quelle est la probabilité, sous l'hypothèse nulle, d'obtenir la même valeur ou une valeur plus extrême que celle observée ?

On a détaillé pour des raisons pédagogiques mais un algorithme très simple peut être utilisé

1) Calculer le nombre de sigma entre les 2 valeurs mesurées

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

2) La probabilité est donnée par

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

Exemple II: interprétation des tests du χ^2

La distribution du χ^2 à k degrés de liberté est la distribution de la somme du carré de k nombres gaussiens.

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

La statistique du χ^2 est très fréquemment utilisée en cosmologie,

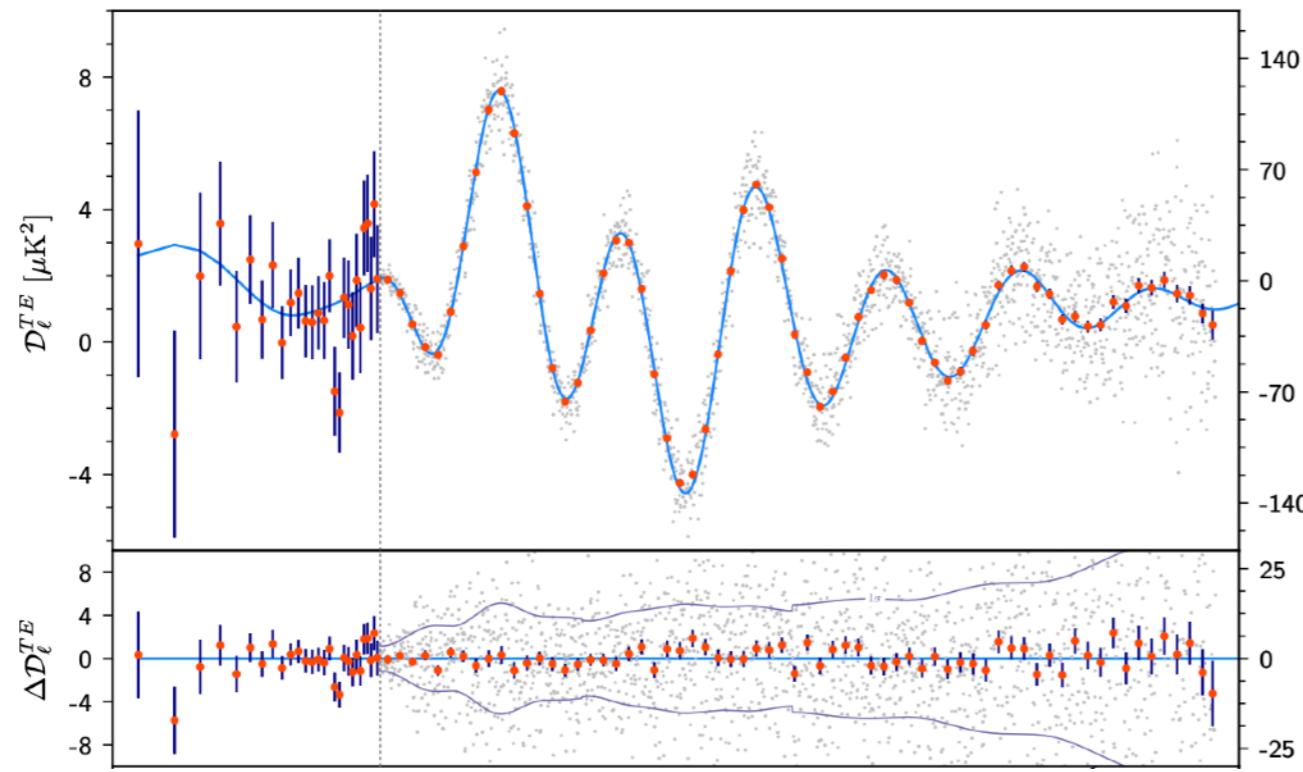
**elle sert notamment à juger de la qualité de l'ajustement (goodness of fit)
d'un modèle aux données observées, la consistance statistique d'un jeu de données,
et est aussi utilisée dans le cadre de la sélection de modèles.**

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

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

Exemple II: interprétation des tests du χ^2

‘Est-ce que le modèle LCDM est un bon fit du spectre TE des données Planck ?’



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

TE, full, binned

30–1996

$\log(\mathcal{L})$

−428.68

χ^2_{eff}

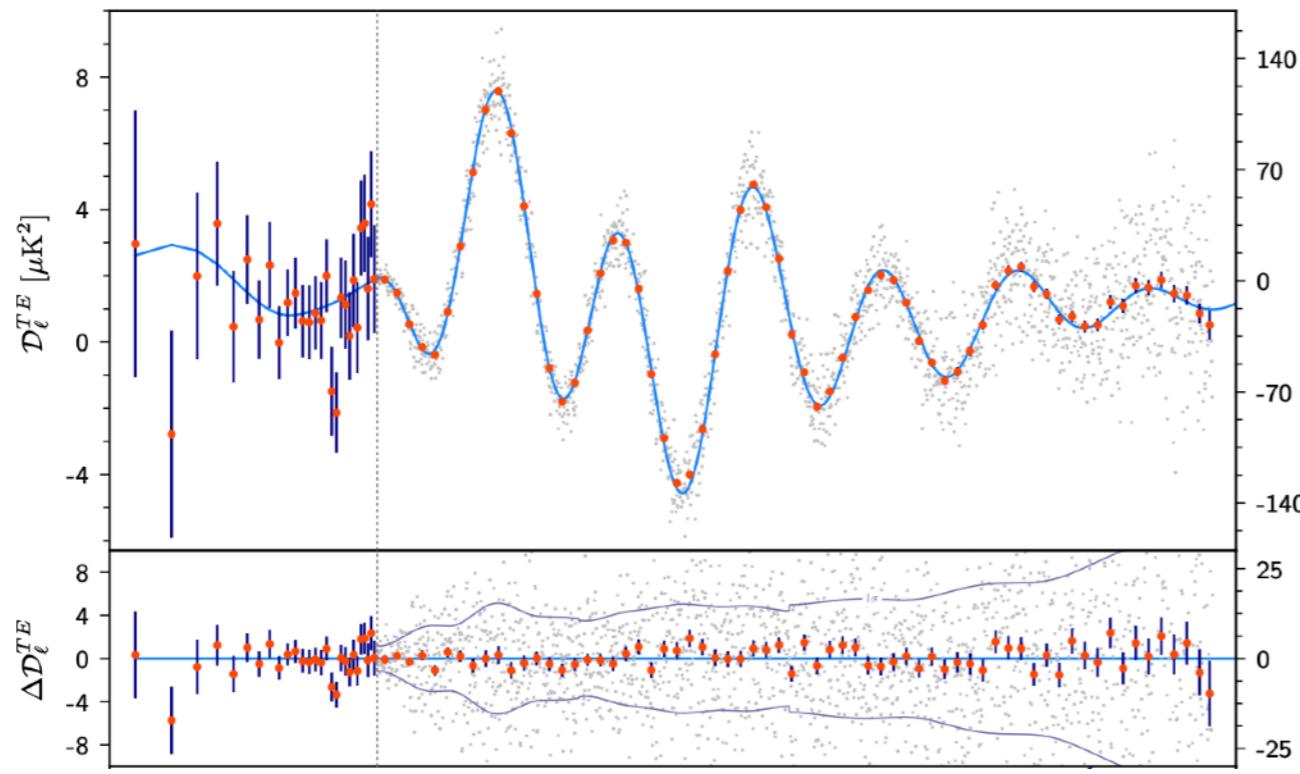
857.36

N_{dof}

762

Exemple II: interprétation des tests du χ^2

‘Est-ce que le modèle LCDM est un bon fit du spectre TE des données Planck ?’



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

```
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
```



$p \sim 1\%$

$$\chi^2/\text{DoF}$$

On requiert comme critère que les χ^2/DoF soient ‘proches’ de 1.

‘Proche’ est extrêmement dépendant du nombre de degrés de liberté considérés.

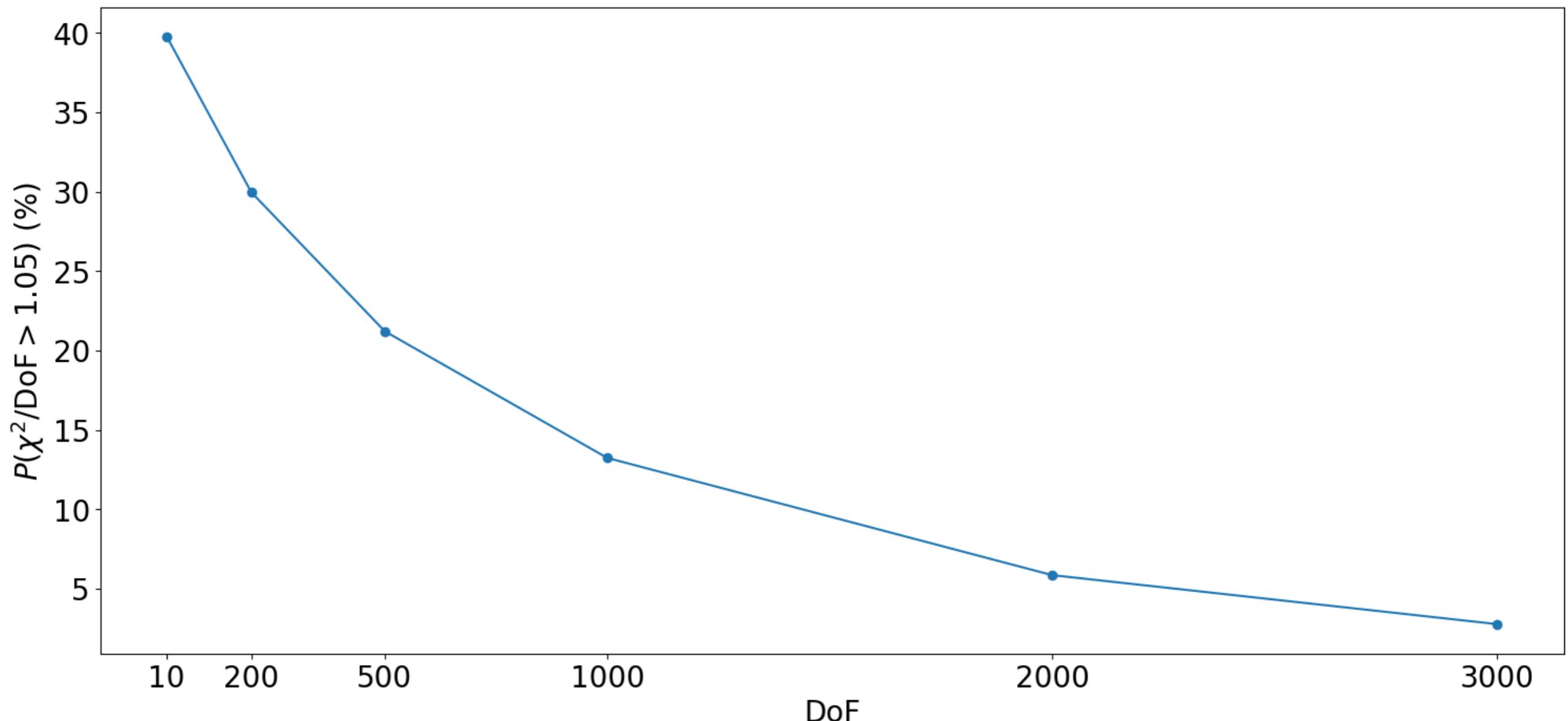


TABLE 3
NULL TESTS USING CUSTOM MAPS (PA1, PA2)

Jackknifes

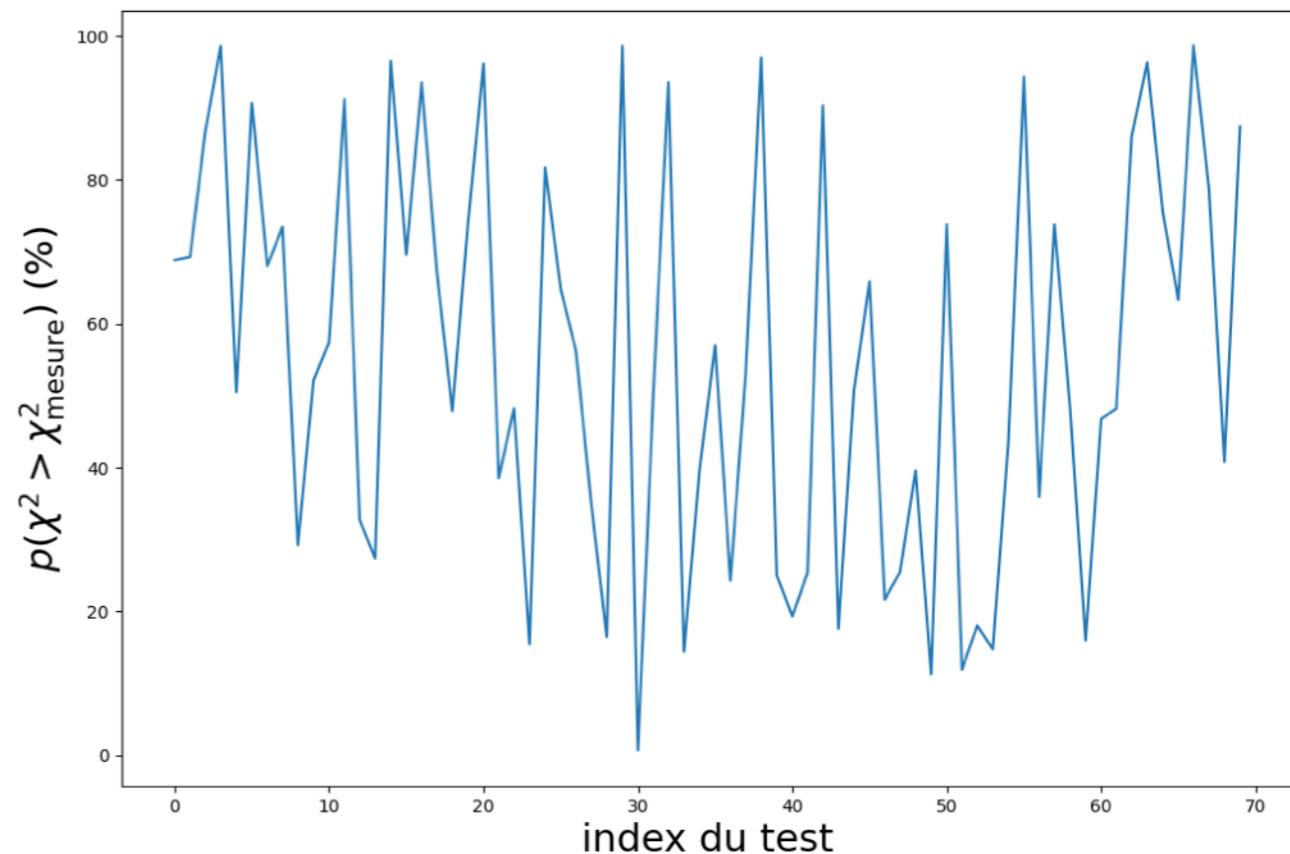
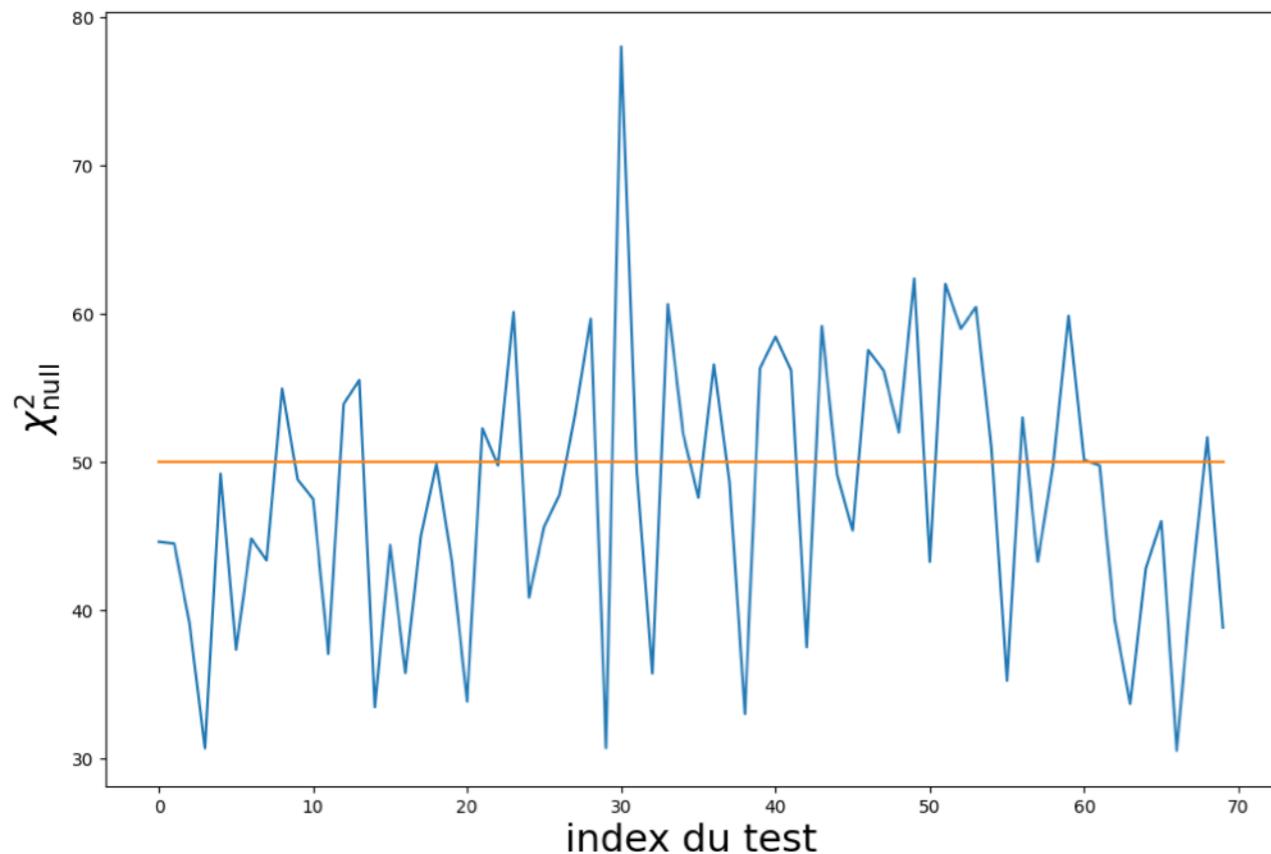
En cosmologie les χ^2 sont aussi utilisés dans le cadre de ‘Jackknifes’ (aussi appelé null tests), un ensemble de tests dans lequel on divise les données en fonction de différents critères pour certifier la stabilité du résultat :

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

Test	Spectrum	PA1 χ^2/dof	P.T.E	PA2 χ^2/dof	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

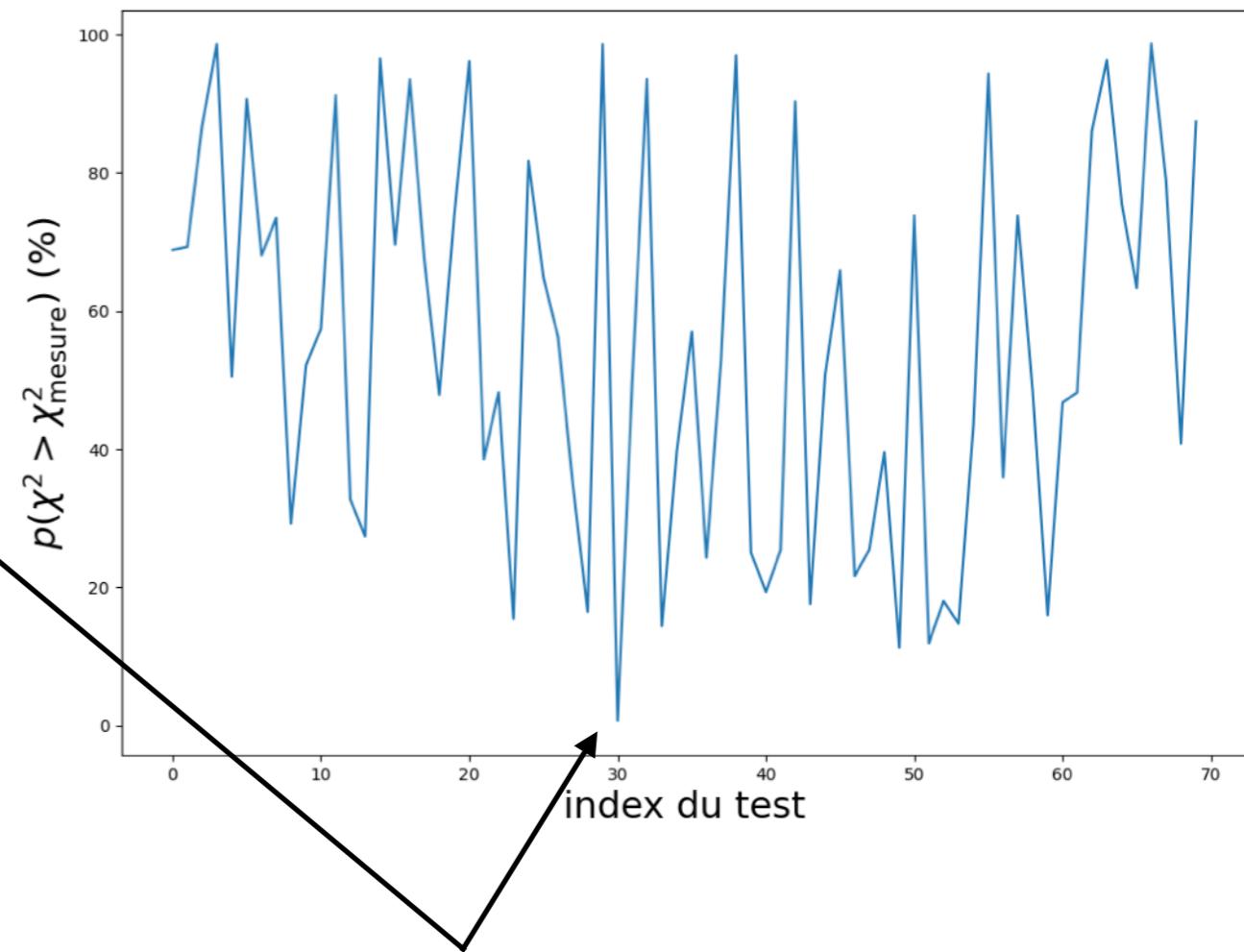
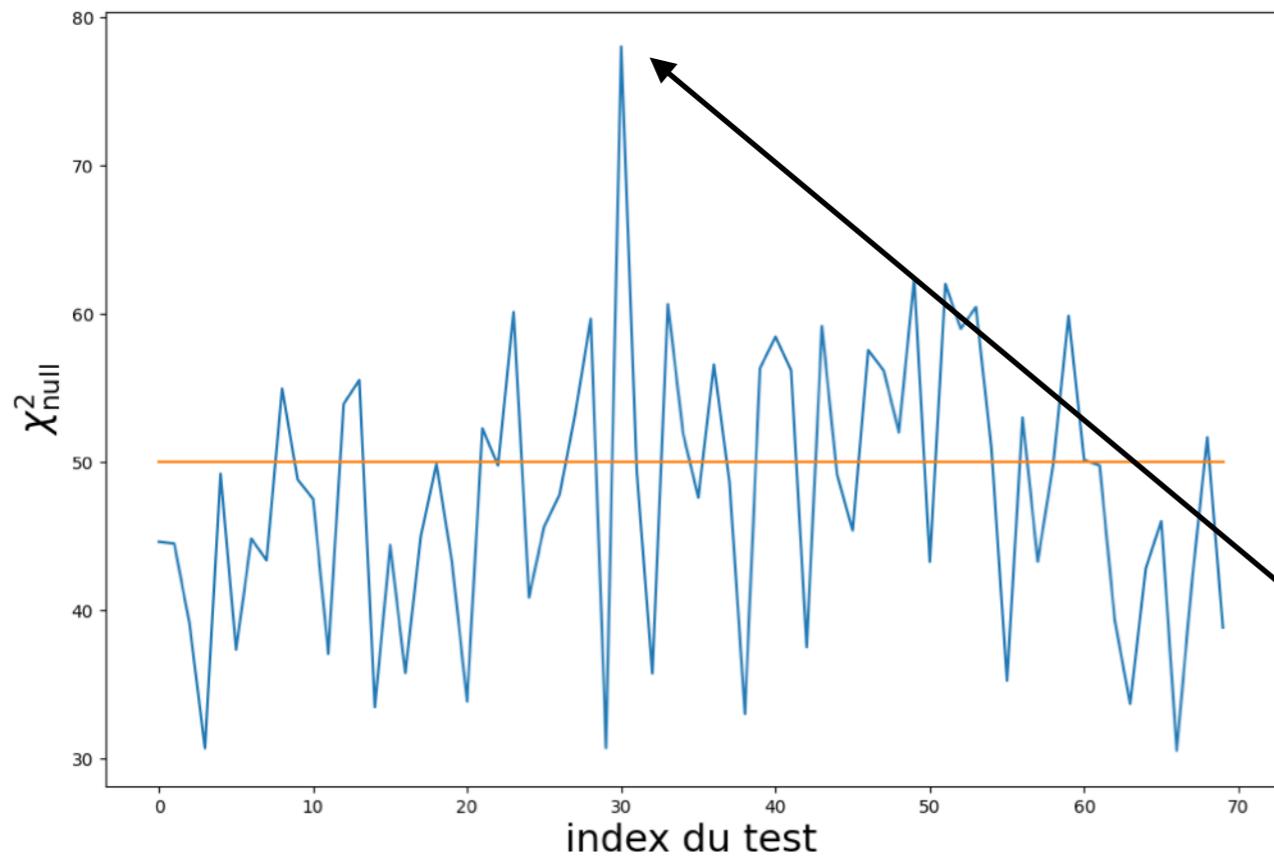
Look elsewhere effect

**Imaginons que l'on conduise un grand nombre de tests sur un jeu de données,
On plot les χ^2 et leurs probabilités associées**



Look elsewhere effect

Imaginons que l'on conduise un grand nombre de tests sur un jeu de données,
On plot les χ^2 et leurs probabilités associées



P = 0.5%, avons nous détecté un problème ?

Look elsewhere effect

(multiple comparisons, multiplicity or multiple testing problem)

On a fait 70 tests, la probabilité qu'un des tests ait une p-value < 0.5% est donnée par

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

Cet effet a été la cause de plusieurs fausses détections en physique

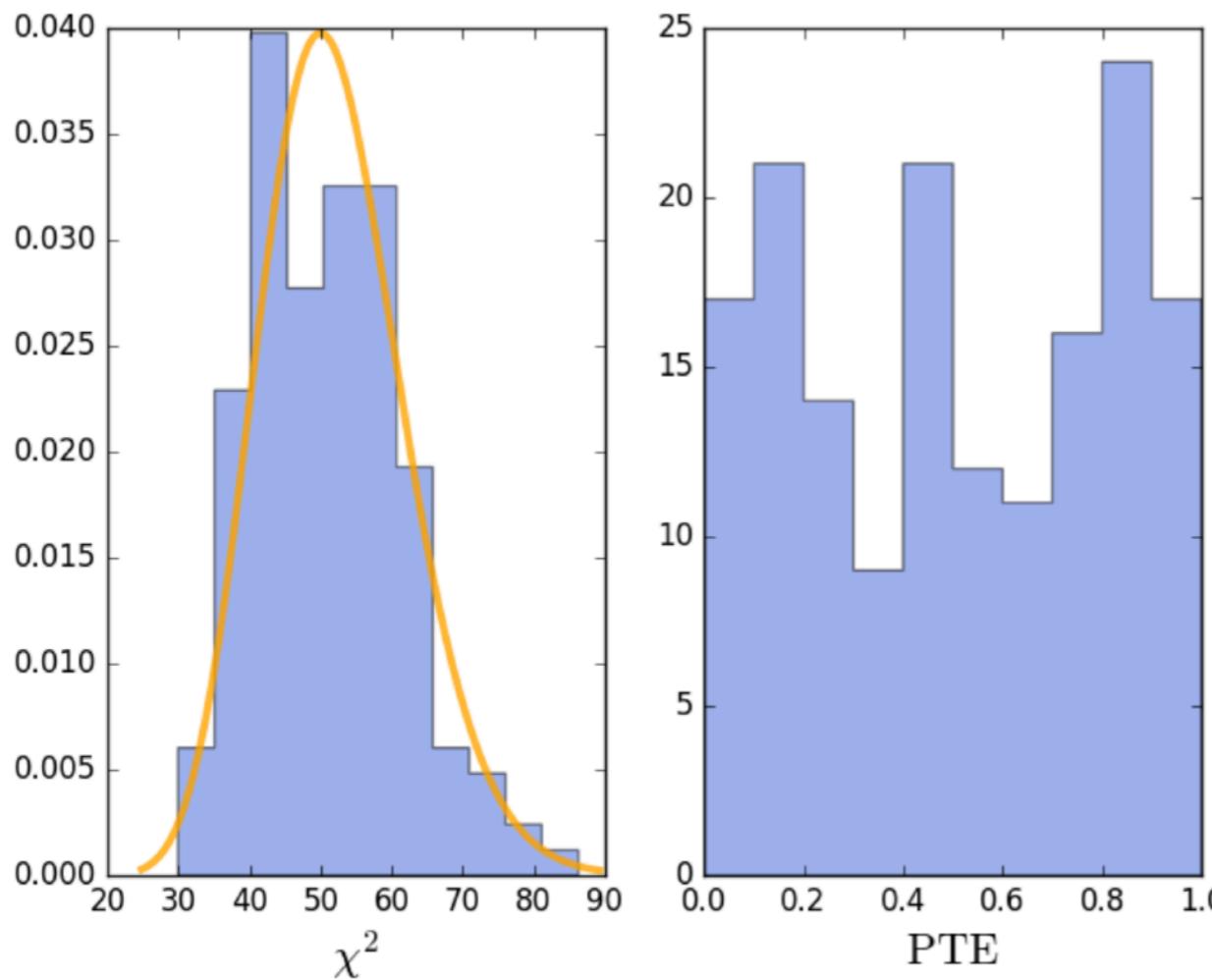
Look elsewhere effect

(multiple comparisons, multiplicity or multiple testing problem)

On a fait 70 tests, la probabilité qu'un des tests ait une p-value < 0.5% est donnée par

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

Cet effet a été la cause de plusieurs fausses détections en physique



En règle générale, on veut toujours analyser le résultat d'un test dans le contexte du nombre de tests effectués, on peut par exemple comparer l'histogramme des tests effectués avec la distribution attendue.

Comparaison de modèles

Un exemple: The H0 Olympics, a fair ranking of proposed models

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

- 1) Pour chaque modèle M, on demande pour un jeu de donnée D (Planck 2018 + BAO + Pantheon) qui n'inclue pas la mesure SHOES, quelle est la tension résiduelle entre le postérieur de H0 inféré à partir de D et celui inféré à partir de SHOES?

$$\frac{\bar{x}_{\mathcal{D}} - \bar{x}_{\text{SH0ES}}}{(\sigma_{\mathcal{D}}^2 + \sigma_{\text{SH0ES}}^2)^{1/2}}$$

- 2) Comment l'addition de SHOES au jeu de données impact le fit d'un modèle particulier ?

$$\Delta\chi^2 = \chi^2_{\min, \mathcal{D}+\text{SH0ES}} - \chi^2_{\min, \mathcal{D}}$$

- 3) Lorsque le jeu de données D inclut SHOES, est-ce que le fit d'un modèle particulier M est significativement meilleur que LCDM ?

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

Critère d'information d'Akaike

Comparaison de modèles

Un exemple: The H0 Olympics, a fair ranking of proposed models

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

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

Interlude

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

```
"""
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')

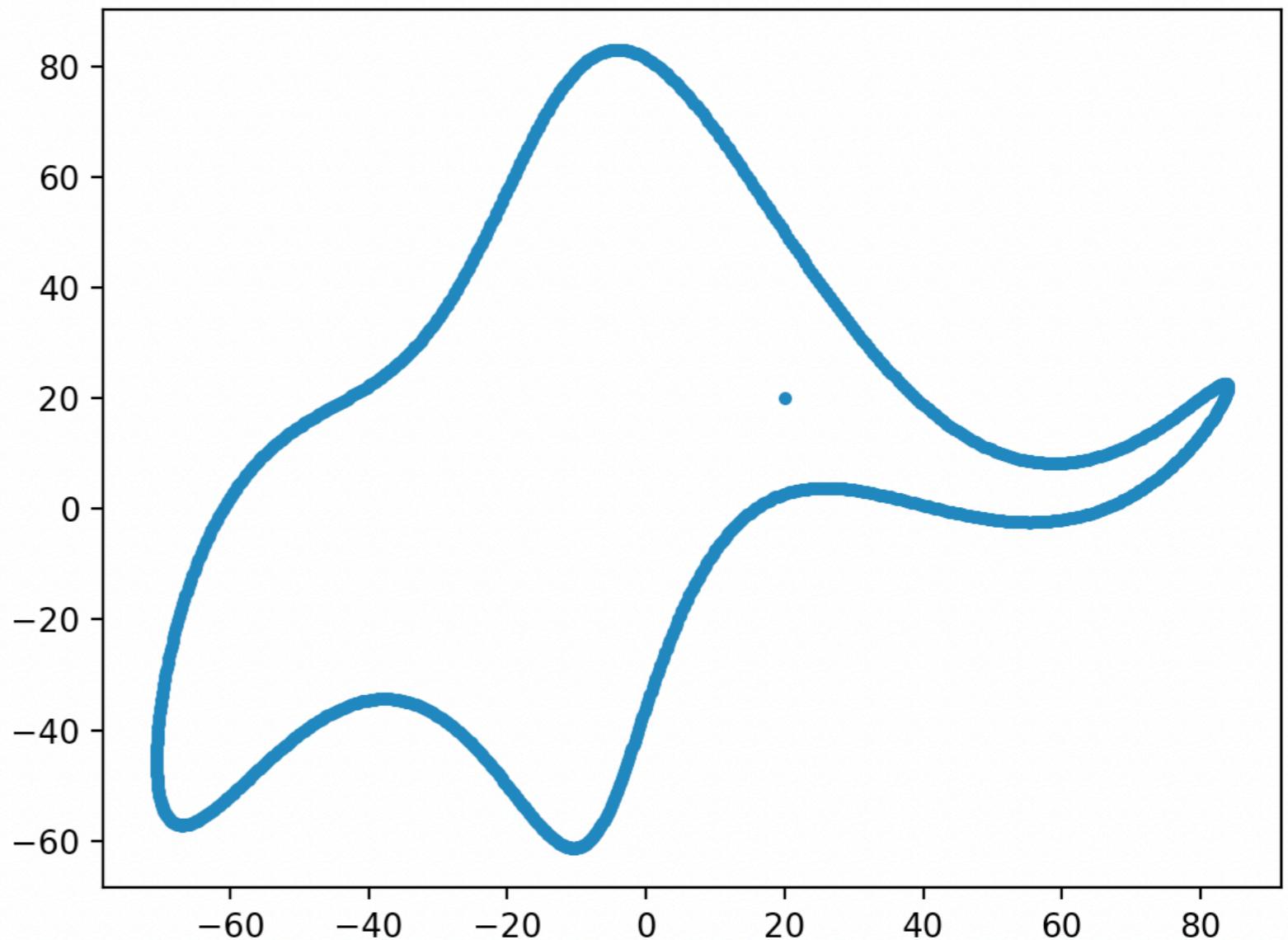
    Cx[1] = p1.real*1j
    Cx[2] = p2.real*1j
    Cx[3] = p3.real
    Cx[5] = p4.real

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

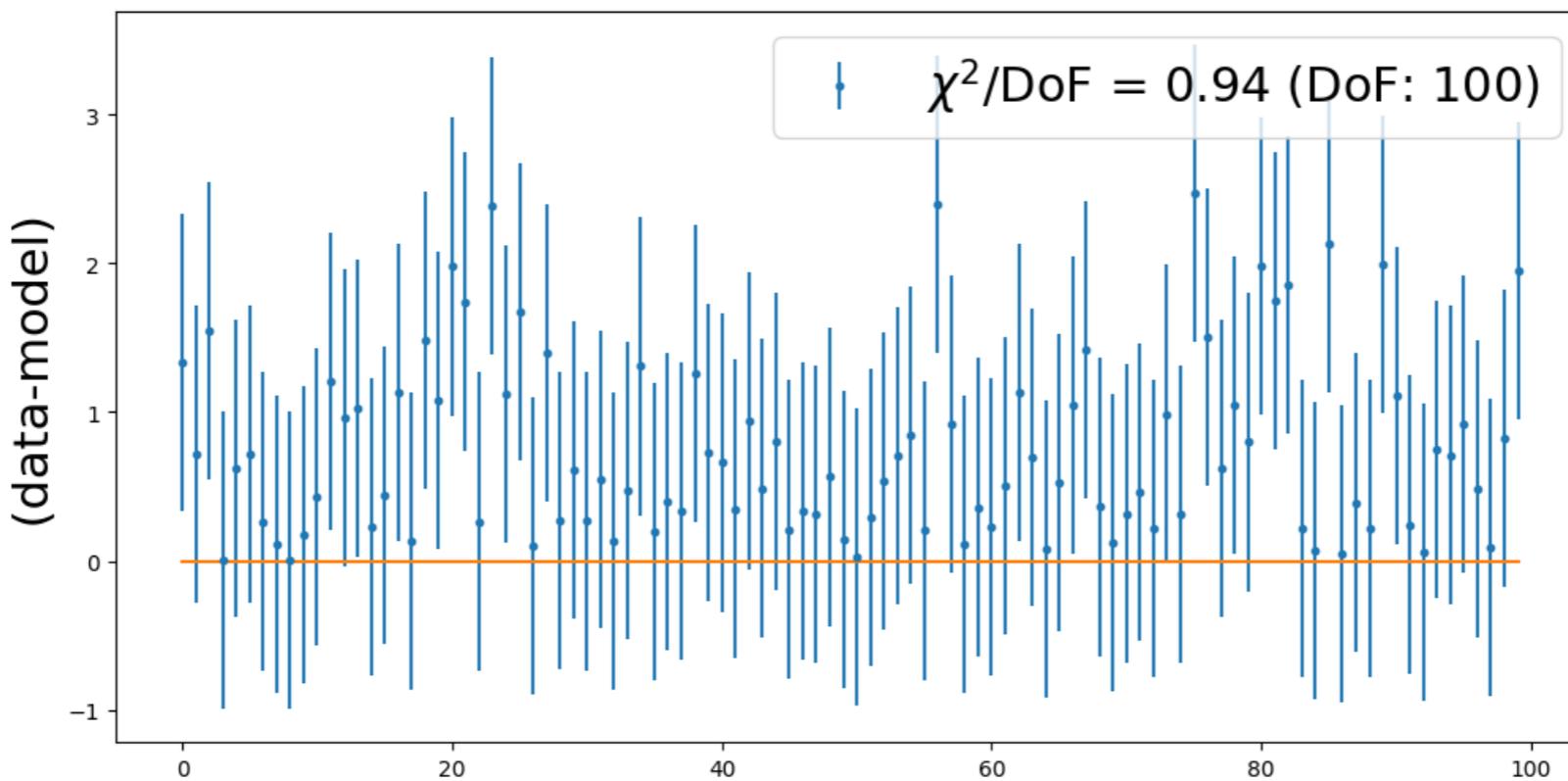
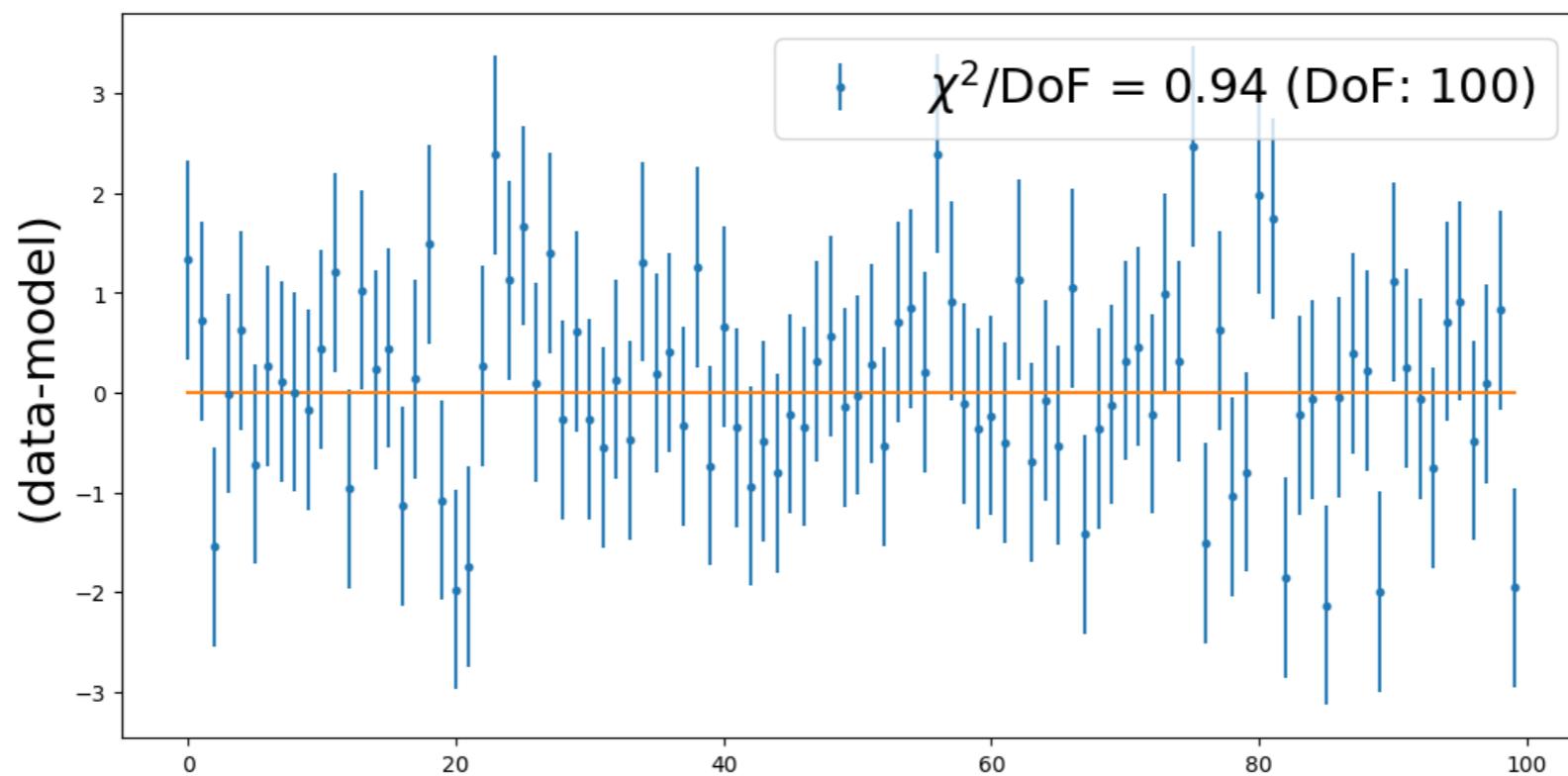
    x = np.append(fourier(t,Cx), [-p5.imag])
    y = np.append(fourier(t,Cy), [p5.imag])

    return x,y

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

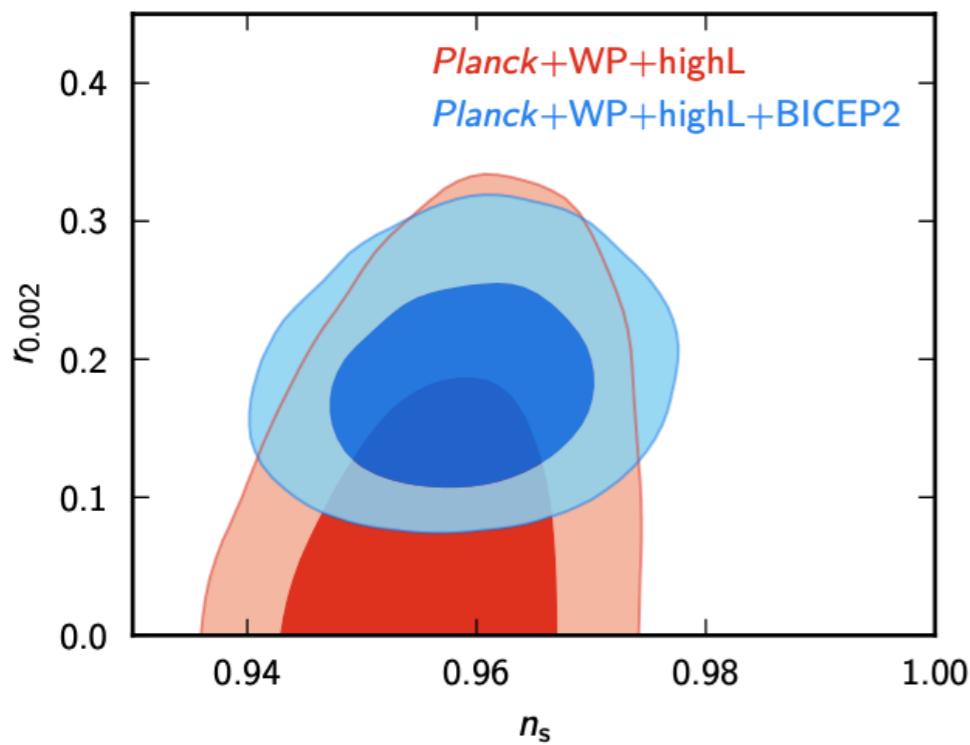
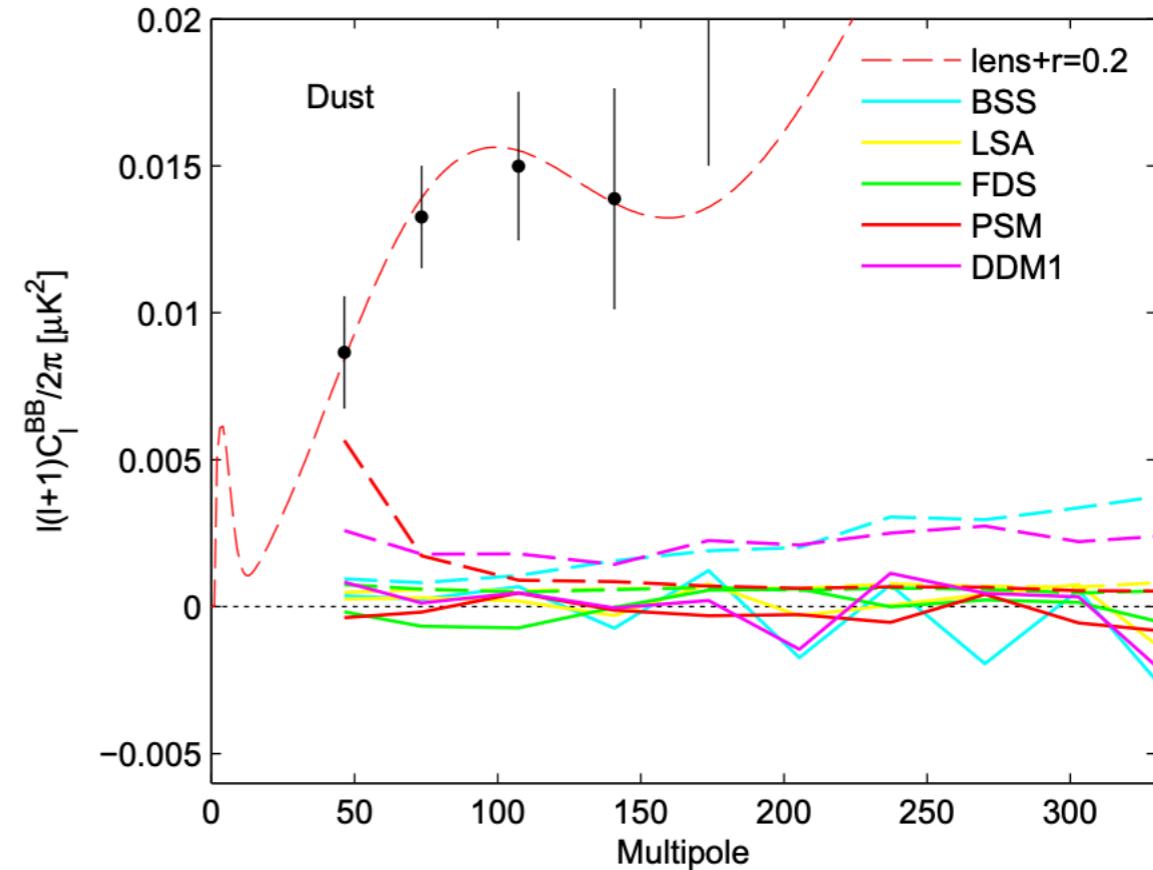
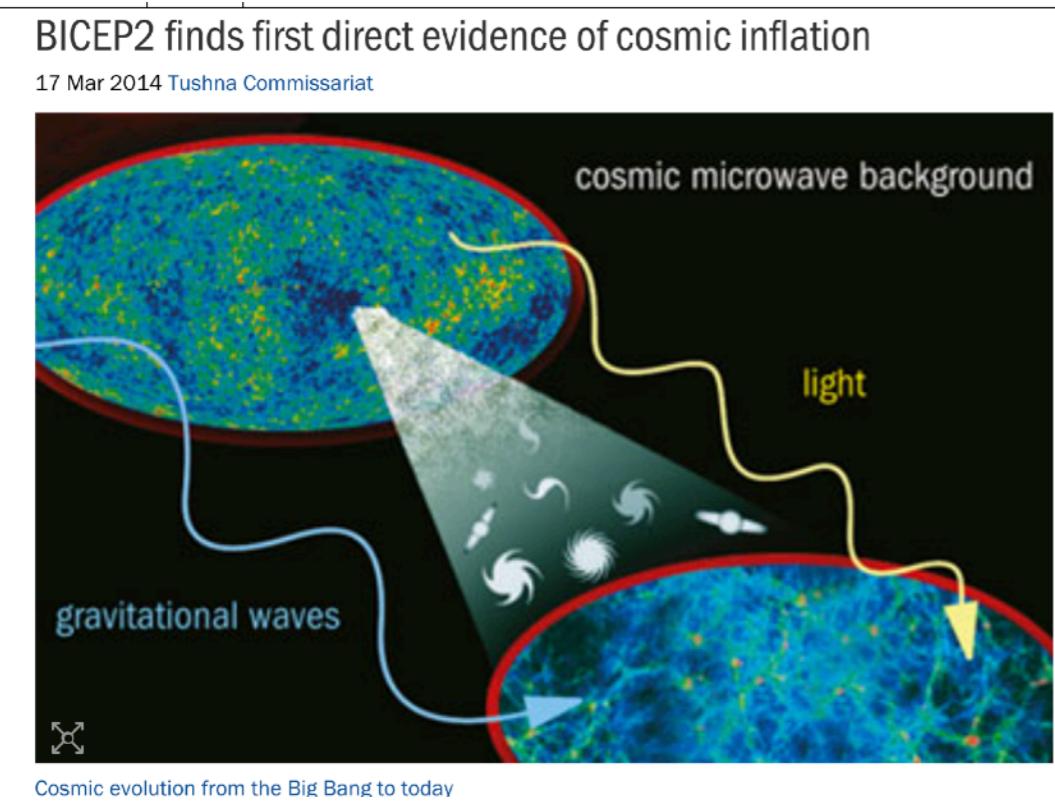


χ^2 is not enough



Une inspection visuelle des résidus est toujours très utile

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



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.

Fonction de probabilité jointe

**On aura souvent à étudier des problèmes décrits par un grand nombre de paramètres
(un grand nombre de variables aléatoires X₁, .., X_N)**

On introduira donc la densité de probabilité jointe:

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

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

On aura la fonction de répartition jointe

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

Définitions utiles I: Marginale

On définit la distribution de probabilité marginale

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

On peut aussi marginaliser sur un sous ensemble de variables aléatoires

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

Pour des variables aléatoires indépendantes, la distribution de probabilité jointe est le produit des marginales

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

Définitions utiles II : Conditionnelle

On définit la distribution de probabilité conditionnelle

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

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

On peut aussi conditionner un sous ensemble de variables aléatoires sur un autre

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

Définitions utiles III : la matrice de covariance

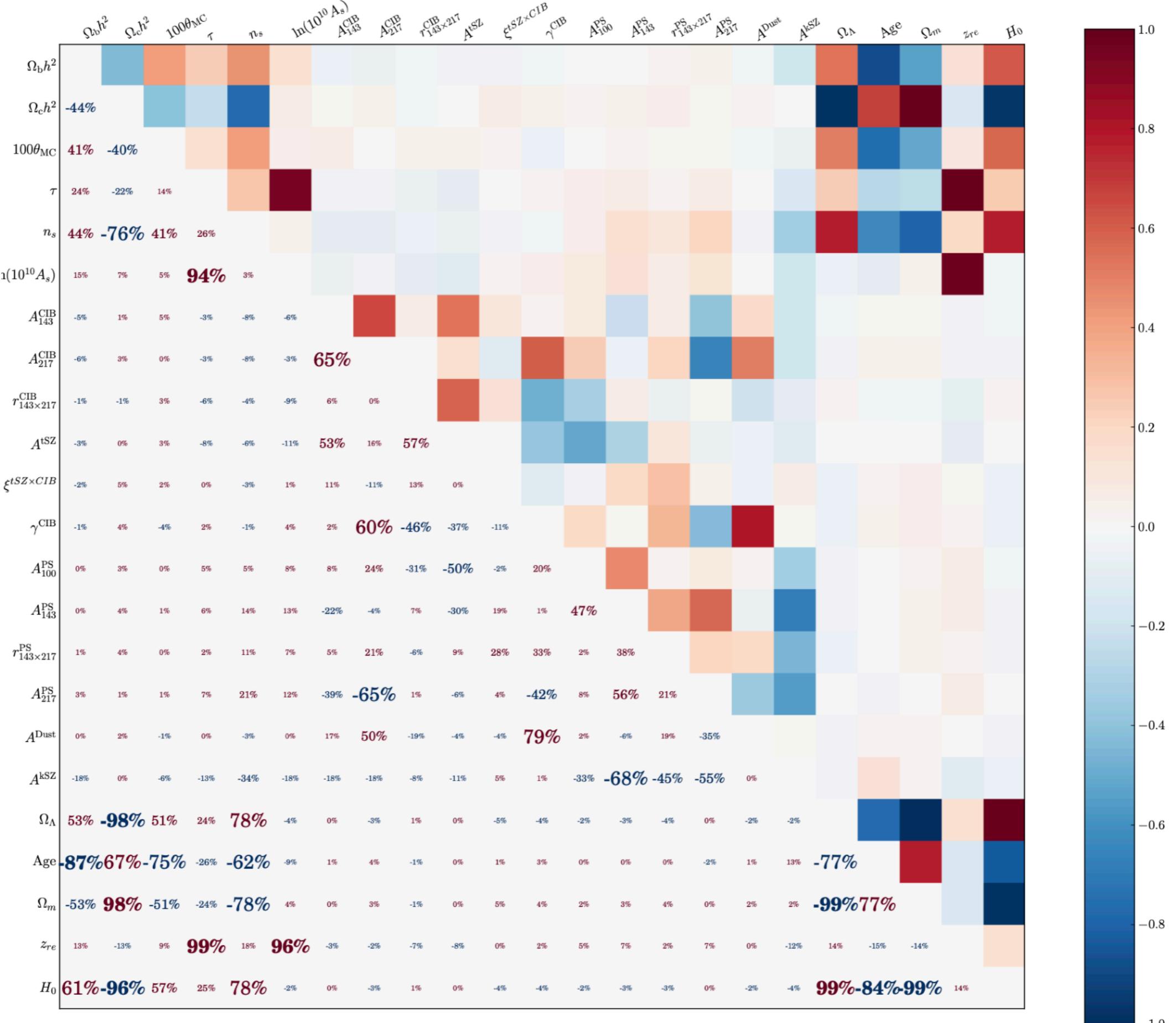
La matrice de covariance d'un vecteur aléatoire est définie par

$$\begin{aligned}\Sigma_{ij} &= \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle \\ &= \int dx_1 \dots dx_N (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \mathcal{P}_{X_1, \dots, X_N}(x_1, \dots, x_N) \\ &= \int dx_i dx_j (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \mathcal{P}_{X_i, X_j}(x_i, x_j)\end{aligned}$$

Le coefficient de corrélation de Pearson (matrice de corrélation) est donné par

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

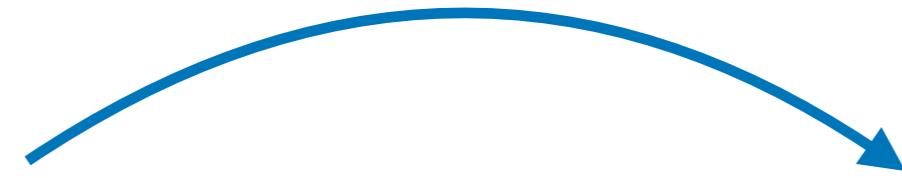
Matrice de corrélation des paramètres de Planck <https://arxiv.org/pdf/1303.5075.pdf>



Deux remarques :

Indépendance

non-corrélé



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

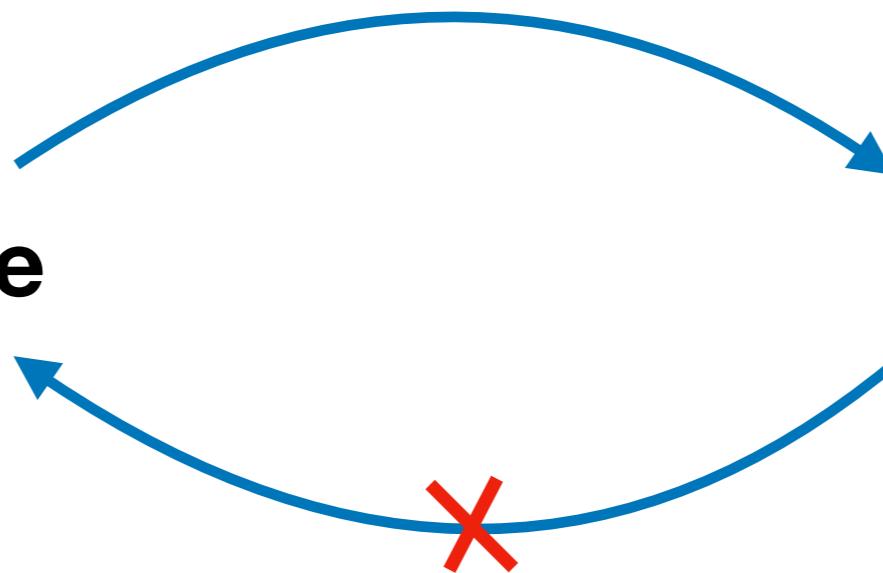
Deux remarques :

Indépendance

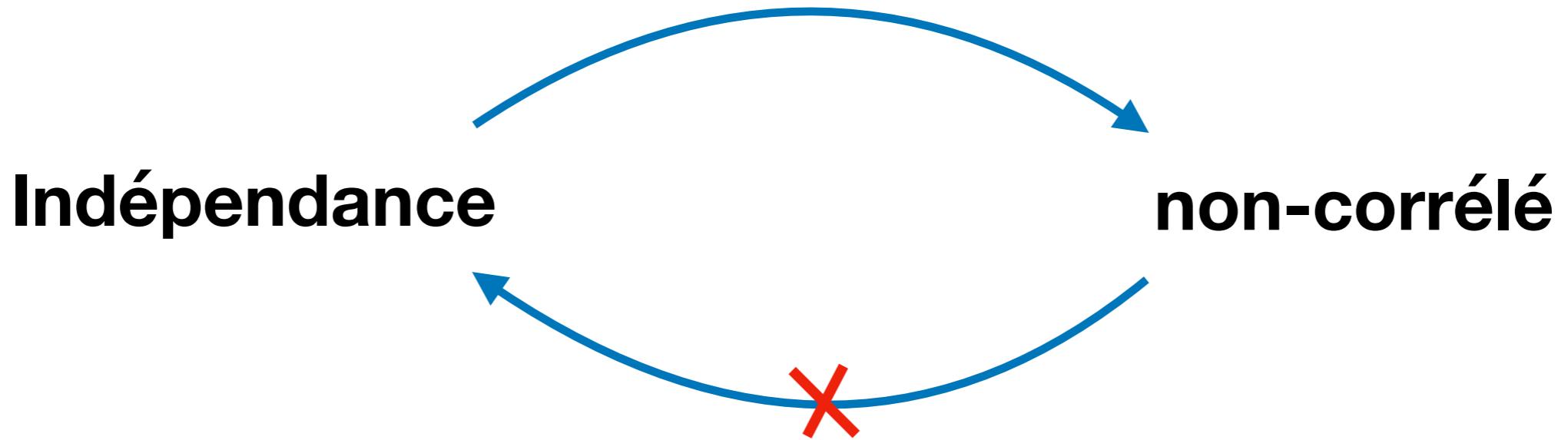
non-corrélé

$$\langle x_i \rangle = 0$$

$$x_j = g(x_i)$$



Deux remarques :



$$\langle x_i \rangle = 0$$

$$x_j = g(x_i)$$

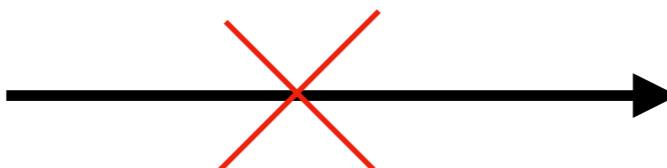
$$\begin{aligned}\Sigma_{ij} &= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i, X_j}(x_i, x_j) \\ &= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i}(x_i) \mathcal{P}_{X_j|X_i}(x_j|x_i) \\ &= \int dx_i dx_j x_i x_j \mathcal{P}_{X_i}(x_i) \delta(x_j - g(x_i)) \\ &= \int dx_i x_i g(x_i) \mathcal{P}_{X_i}(x_i)\end{aligned}$$

Si g et P sont des fonctions paires et si le domaine d'intégration est symétrique autour de zéro

$$\Sigma_{ij} = 0$$

Deux remarques :

Corrélation

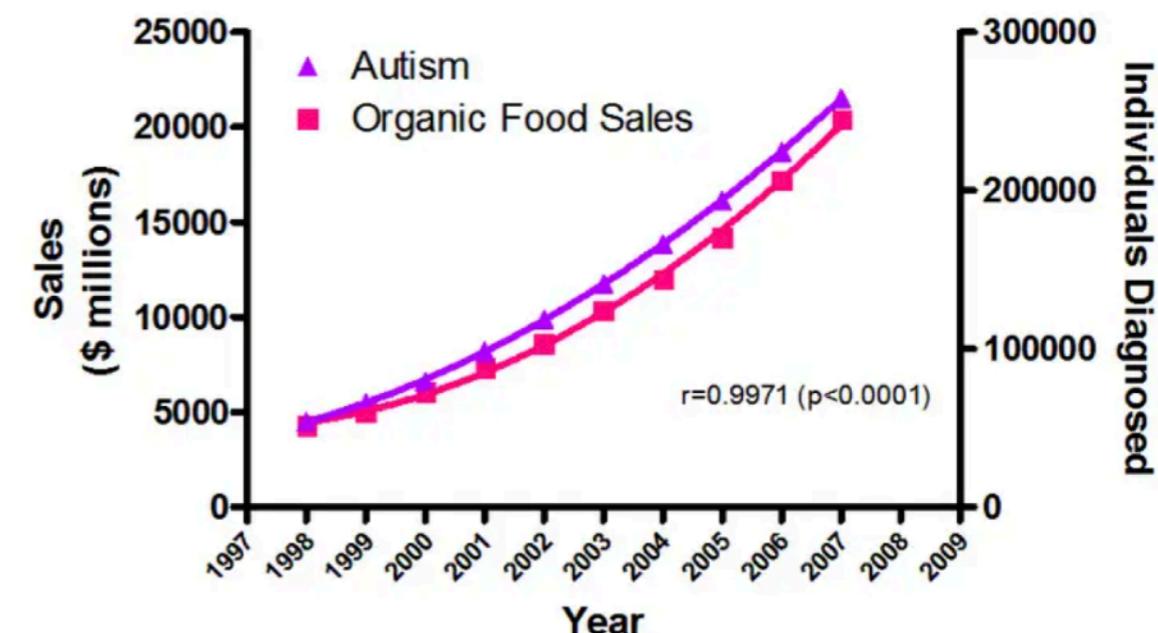
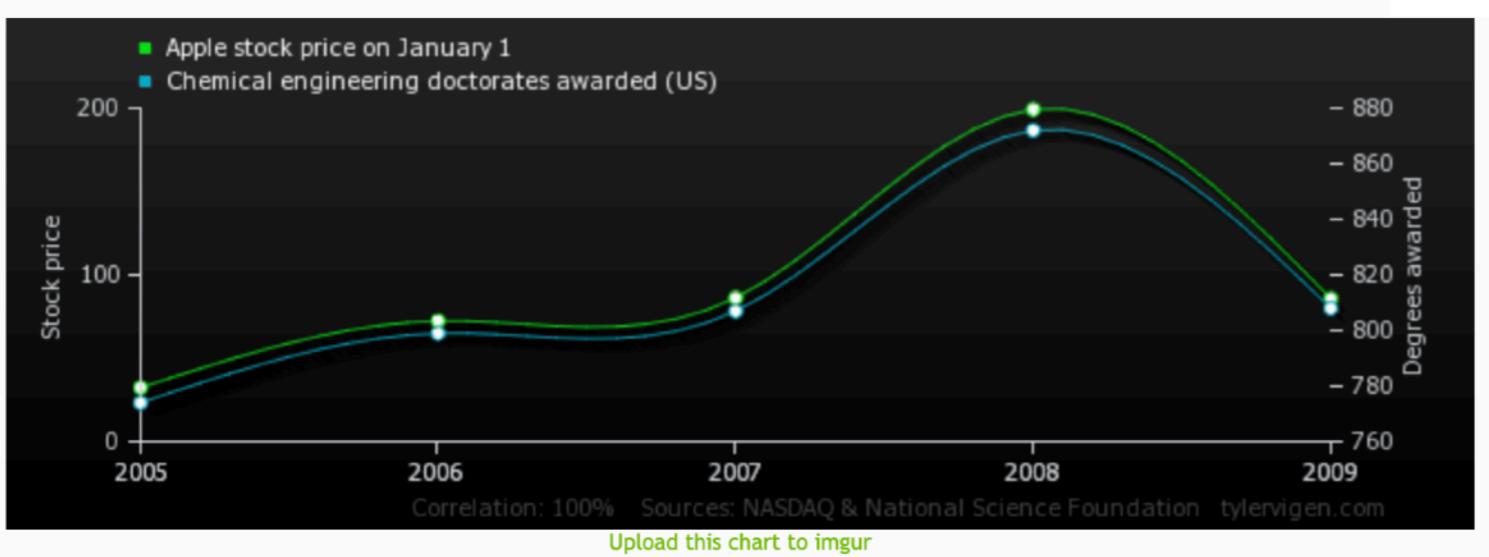


Causalité

Apple stock price on January 1

correlates with

Chemical engineering doctorates awarded (US)



La loi normale multidimensionnelle

Une loi de probabilité multivariée particulièrement utile est la loi normale multidimensionnelle

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

C'est notamment la forme de beaucoup de fonctions de vraisemblance en cosmologie

La loi normale multidimensionnelle

Une loi de probabilité multivariée particulièrement utile est la loi normale multidimensionnelle

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

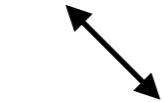
C'est notamment la forme de beaucoup de fonctions de vraisemblance en cosmologie

Si les données prennent la forme

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



Modèle



Bruit Gaussien de moyenne 0 et covariance $\mathcal{N}_{ij} = \langle n_i n_j \rangle$

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

Pour obtenir la distribution marginale d'un sous ensemble de variables aléatoires normales, on doit seulement « enlever » les variables sur lesquelles on marginalise du vecteur qui représente la moyenne et de la covariance

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

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

↓

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

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

Une application de la marginale: Visualisation des ellipses

$$\mathcal{P}_{X_1, X_3}(x_1, x_3) = \frac{1}{2\pi\sqrt{\det \tilde{\Sigma}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1}(\mathbf{x} - \tilde{\boldsymbol{\mu}})\right]$$

$(\mathbf{x} - \tilde{\boldsymbol{\mu}})^T \tilde{\Sigma}^{-1}(\mathbf{x} - \tilde{\boldsymbol{\mu}})$ Suit une distribution du χ^2 à deux degrés de liberté

On veut visualiser l'ensemble des \mathbf{x} qui ont une probabilité supérieure à 5%

On doit trouver la valeur de y qui satisfait $P(\chi^2 < y, 2) = 95\%$

On utilise la « Percent point function » aussi appelée fonction quantile

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

```
[In [2]: chi2.ppf(0.95, 2)           ←———— 95%
Out[2]: 5.991464547107979
```

```
[In [3]: chi2.ppf(0.68, 2)           ←———— 68%
Out[3]: 2.27886856637673
```

Une application de la marginale: Visualisation des ellipses

L'ensemble des \boldsymbol{x} qui ont une probabilité supérieure à 5% satisferont

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

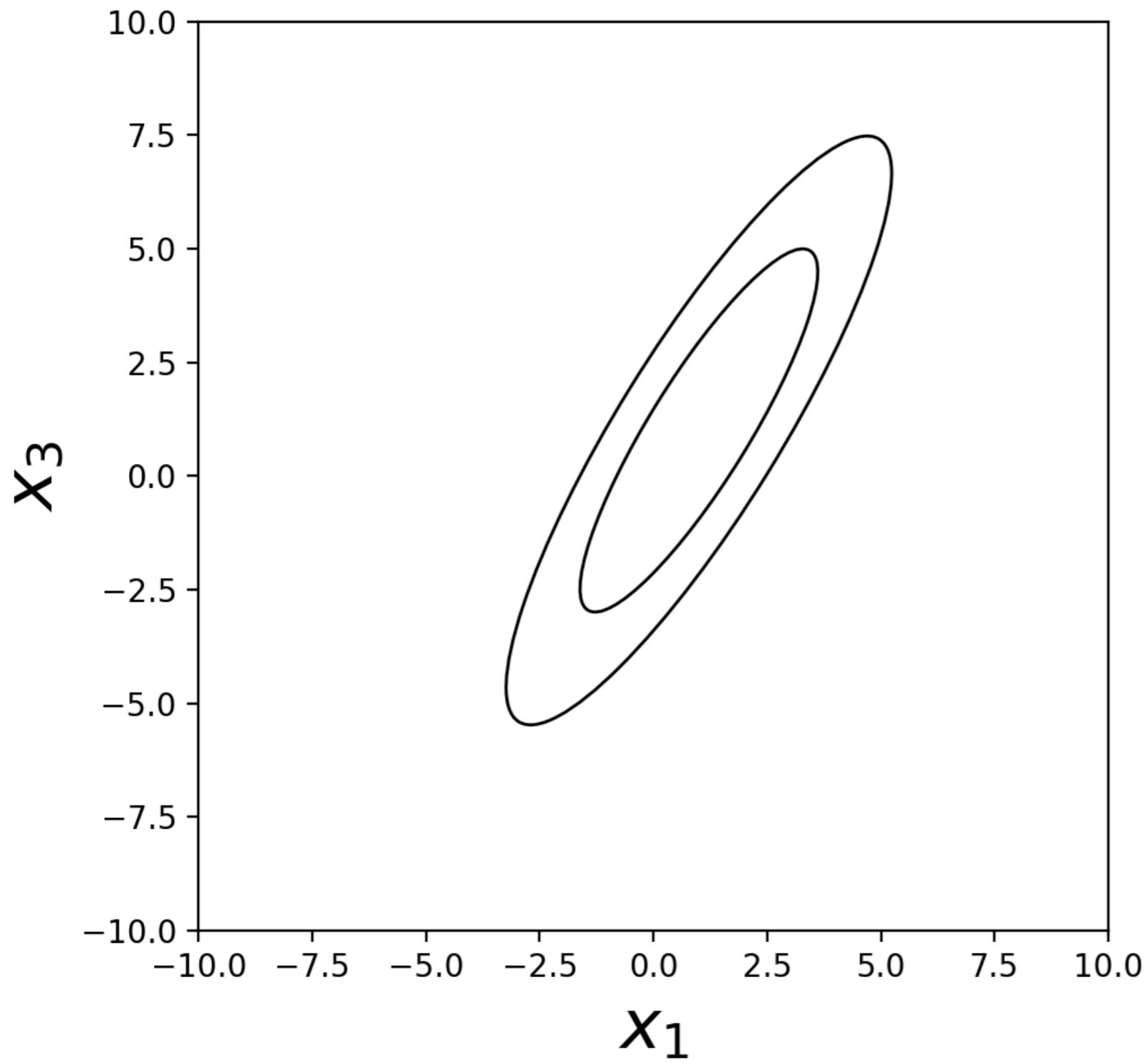
On peut montrer que le contour qui englobe cet ensemble des \boldsymbol{x}

est une ellipse d'axe $\boldsymbol{\mu} \pm \sqrt{5.9915 \lambda_i} \boldsymbol{e}_i$

Avec λ_i les valeurs propres de la matrice de covariance et

\boldsymbol{e}_i ses vecteurs propres

$$\tilde{\Sigma} = \begin{pmatrix} 3 & 4 \\ 4 & 7 \end{pmatrix} \quad \tilde{\mu} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$



$$\tilde{\Sigma} = \begin{pmatrix} 3 & 4 \\ 4 & 7 \end{pmatrix} \quad \tilde{\mu} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

On peut vérifier nos contours en simulant des vecteurs aléatoires à partir de la matrice de covariance

Cholesky

$$\tilde{\Sigma} = LL^T$$

$$\mathbf{x}_{(k)} = \tilde{\mu} + L\mathbf{u}_{(k)}$$

$$\mathbf{u} \sim \mathcal{N}(0, I)$$



$$\langle \mathbf{x} \rangle = \langle \tilde{\mu} \rangle + L \langle \mathbf{u} \rangle = \tilde{\mu}$$

$$\langle \mathbf{x} \mathbf{x}^T \rangle - \tilde{\mu} \tilde{\mu}^T = L \langle \mathbf{u} \mathbf{u}^T \rangle L^T = LL^T = \tilde{\Sigma}$$

$$\tilde{\Sigma} = \begin{pmatrix} 3 & 4 \\ 4 & 7 \end{pmatrix} \quad \tilde{\mu} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

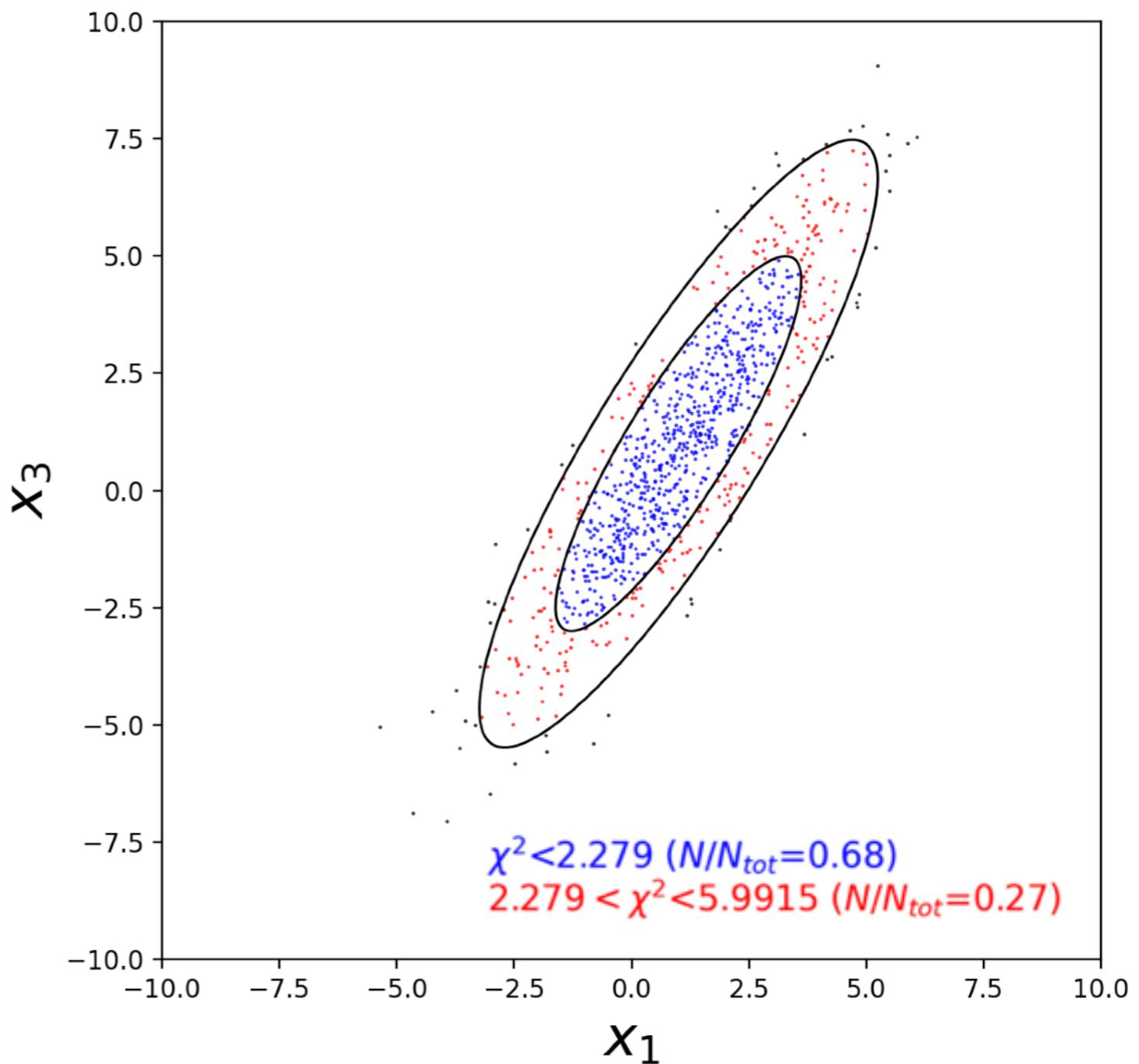
On peut vérifier nos contours en simulant des vecteurs aléatoires à partir de la matrice de covariance

Cholesky

$$\tilde{\Sigma} = LL^T$$

$$x_{(k)} = \tilde{\mu} + Lu_{(k)}$$

$$u \sim \mathcal{N}(0, I)$$



La distribution conditionnelle d'une loi normale multidimensionnelle est aussi une loi normale multidimensionnelle, si on conditionne une partie du vecteur aléatoire

x_1 sur une autre partie x_2 et si on note :

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

La conditionnelle de x_1 étant donné $x_2 = x_2^{\text{obs}}$

est une loi normale multidimensionnelle de moyenne

$$\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2^{\text{obs}} - \mu_2)$$

et covariance

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

La distribution conditionnelle d'une loi normale multidimensionnelle est aussi une loi normale multidimensionnelle, si on conditionne une partie du vecteur aléatoire

x_1 sur une autre partie x_2 et si on note :

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

La conditionnelle de x_1 étant donné $x_2 = x_2^{\text{obs}}$

est une loi normale multidimensionnelle de moyenne

$$\bar{\mu} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2^{\text{obs}} - \mu_2)$$

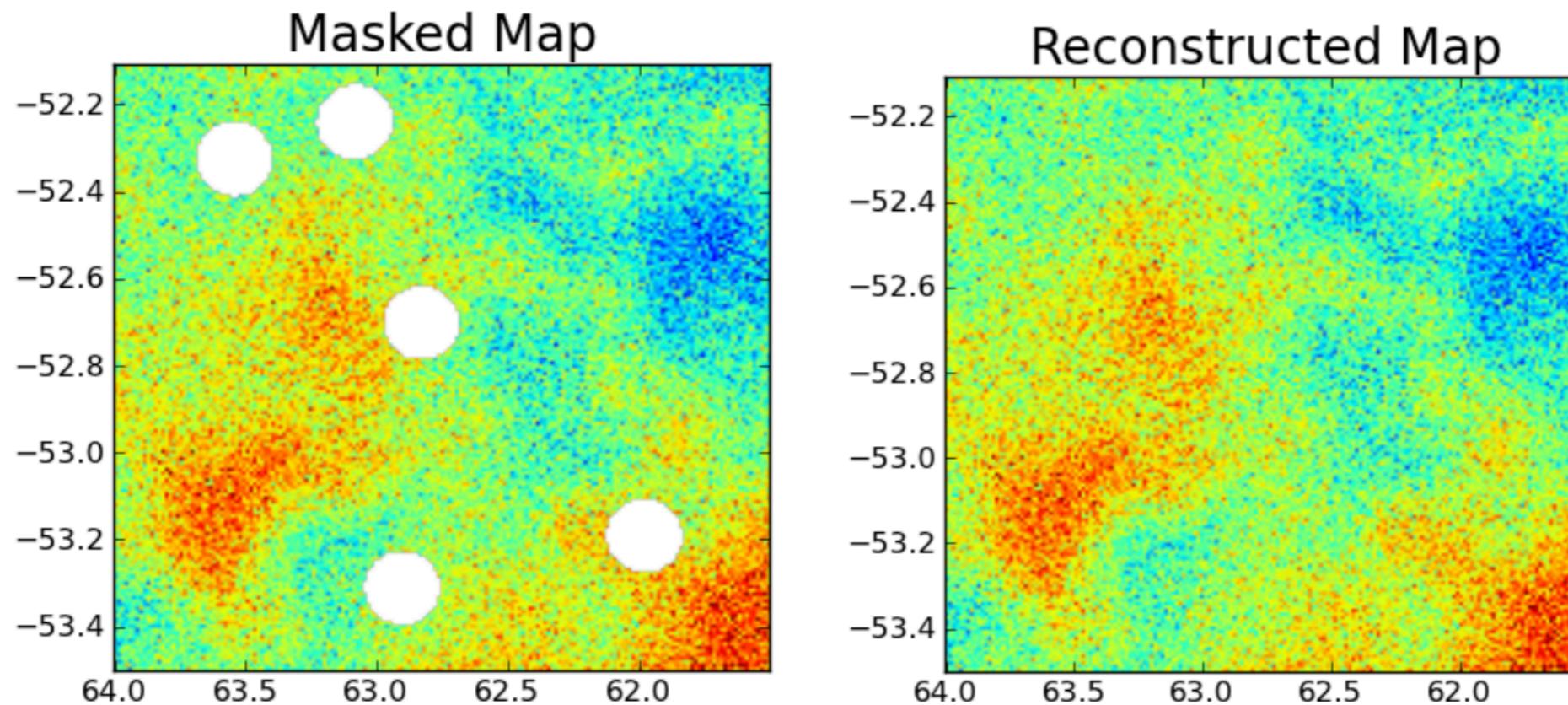
et covariance

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

La conditionnelle d'un vecteur aléatoire de moyenne nulle peut avoir une moyenne non nulle

Une application de la conditionnelle: inpainting

<https://arxiv.org/abs/1109.0286>



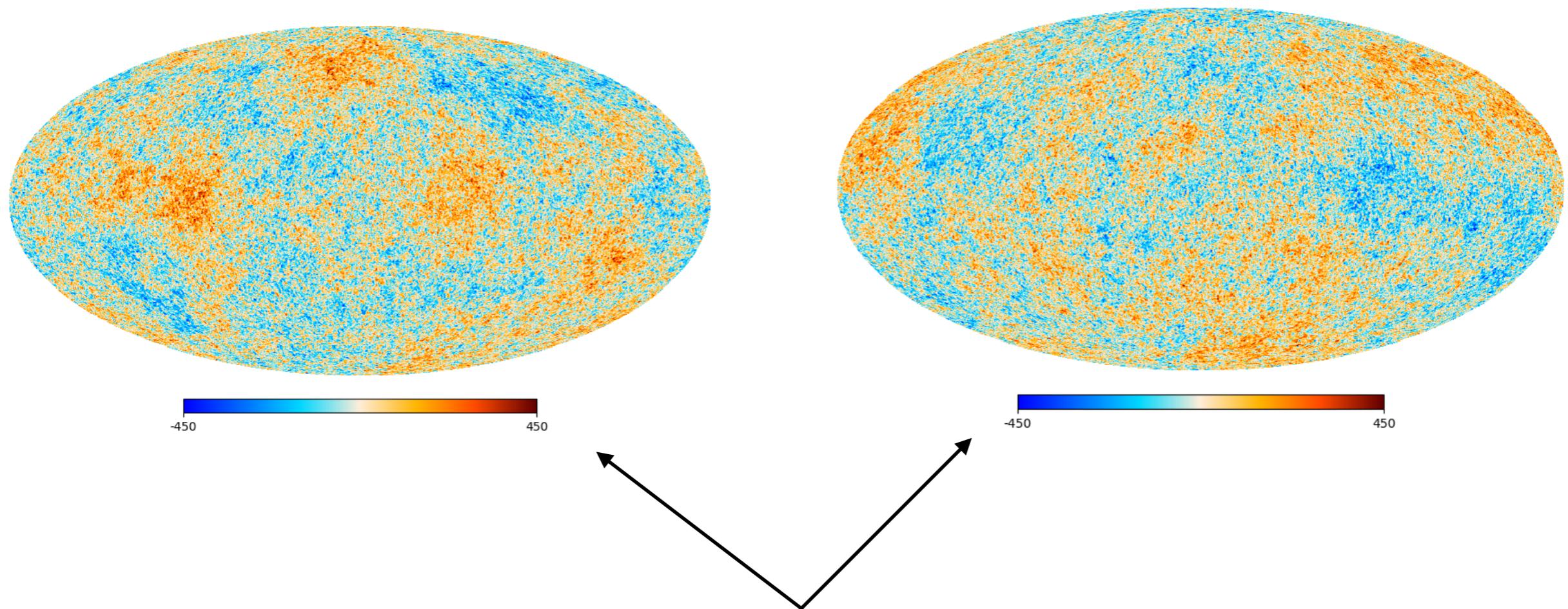
Le CMB est un champ Gaussien, ici on suppose que le bruit instrumental est aussi Gaussien, on conditionne les pixels non observés sur les pixels observés ce qui permet de prédire leur valeur la plus probable.

Champ aléatoire

Du point de vue de nos modèles cosmologiques, la distribution observée des galaxies ainsi que les fluctuations de température et de polarisation observées du fond diffus cosmologique sont contingentes. Nos modèles ne prédisent pas les valeurs que prennent ces observables, ils prédisent simplement leurs propriétés statistiques

Champ aléatoire

Du point de vue de nos modèles cosmologiques, la distribution observée des galaxies ainsi que les fluctuations de température et de polarisation observées du fond diffus cosmologique sont contingentes. Nos modèles ne prédisent pas les valeurs que prennent ces observables, ils prédisent simplement leurs propriétés statistiques

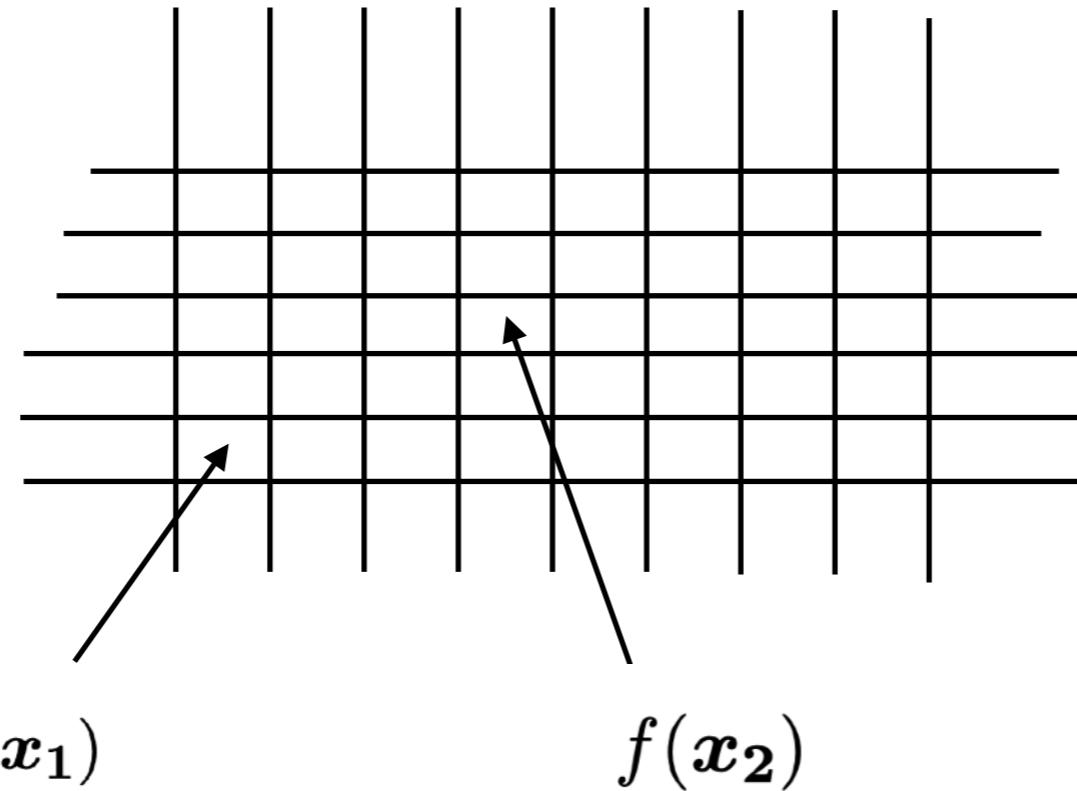


« Même univers » du point de vue de LCDM

Ces observables sont donc promus au titre de champs aléatoires

Champ aléatoire

On peut généraliser le concept de variables aléatoires à celui de champ aléatoire
Imaginons un ensemble de variables aléatoires définies pour chaque élément d'une grille d'un pas donné



En notant $f_n = f(x_n)$ on peut définir un vecteur f de n variables aléatoires

et sa distribution de probabilité associée $\{f, \mathcal{P}(f)\}$

Le champ aléatoire peut être défini comme la limite continue où le pas de la grille tend vers zéro

Fonctions de corrélations à n points

Par analogie aux moments d'une variable aléatoire on peut définir les fonctions de corrélation à n points du champ $f(\mathbf{x})$

$$\xi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \left\langle \prod_{i=1}^n f(\mathbf{x}_i) \right\rangle = \int \mathcal{D}f \mathcal{P}[f] \prod_{i=1}^n f(\mathbf{x}_i)$$

Le principe cosmologique fait l'hypothèse que les propriétés statistiques de l'Univers aux grandes échelles sont homogènes et isotropes

Homogénéité

$$\langle f(x_1) \dots f(x_n) \rangle = \langle f(x_1 + b) \dots f(x_n + b) \rangle$$

Isotropie

Un champ est isotrope autour d'un point z si

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

Une prédiction des modèles les plus simples d'inflation (et confirmée largement par Planck) est que les fluctuations initiales suivent une distribution gaussienne

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

Toute l'information sur un champ Gaussien est contenue dans sa matrice de covariance (sa fonction de corrélation à deux points)

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

Une prédiction des modèles les plus simples d'inflation (et confirmée largement par Planck) est que les fluctuations initiales suivent une distribution gaussienne

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

Toute l'information sur un champ Gaussien est contenue dans sa matrice de covariance (sa fonction de corrélation à deux points)

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

Une combinaison linéaire de champs Gaussiens suit une distribution Gaussienne.

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

\mathcal{F}_1 est une fonctionnelle qui représente l'évolution des contrastes de densité générés par l'inflation jusqu'à l'émission du CMB

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

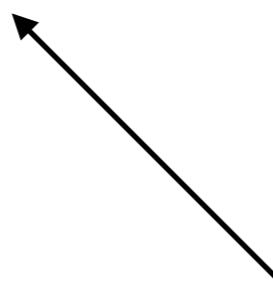
Les phénomènes non linéaires (comme la gravité) vont générer des non-gaussiannités, néanmoins, même dans ce cas les fonctions de corrélation à 2 points contiennent beaucoup d'information

Matrice de covariance

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

Matrice de covariance

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

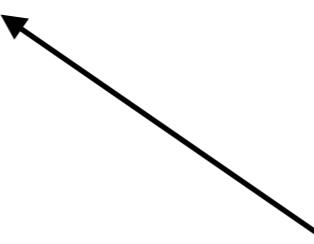


Homogénéité

La matrice de covariance ne dépend que des positions relatives

Matrice de covariance

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



Isotropie

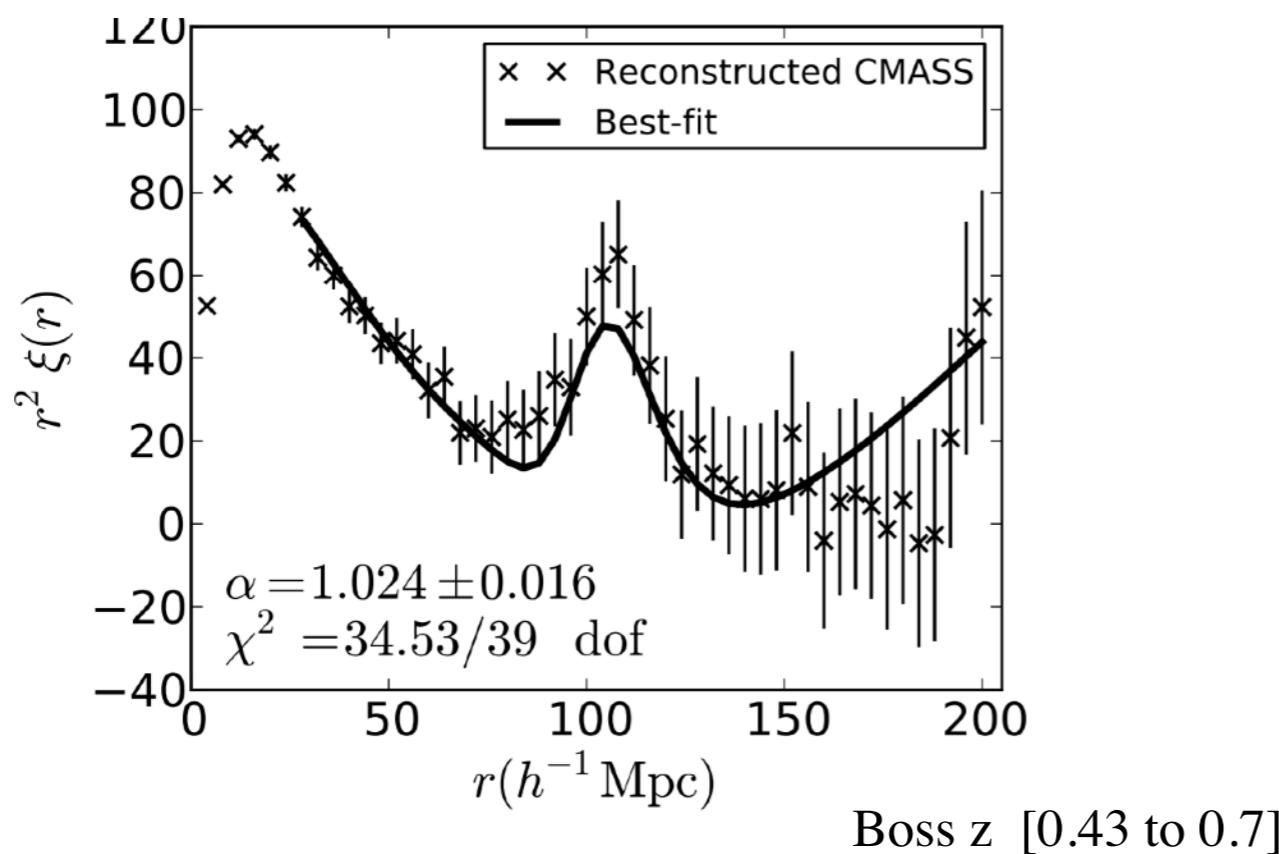
Pas de directionnalité dans la matrice de corrélation

Matrice de covariance

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

Matrice de covariance

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

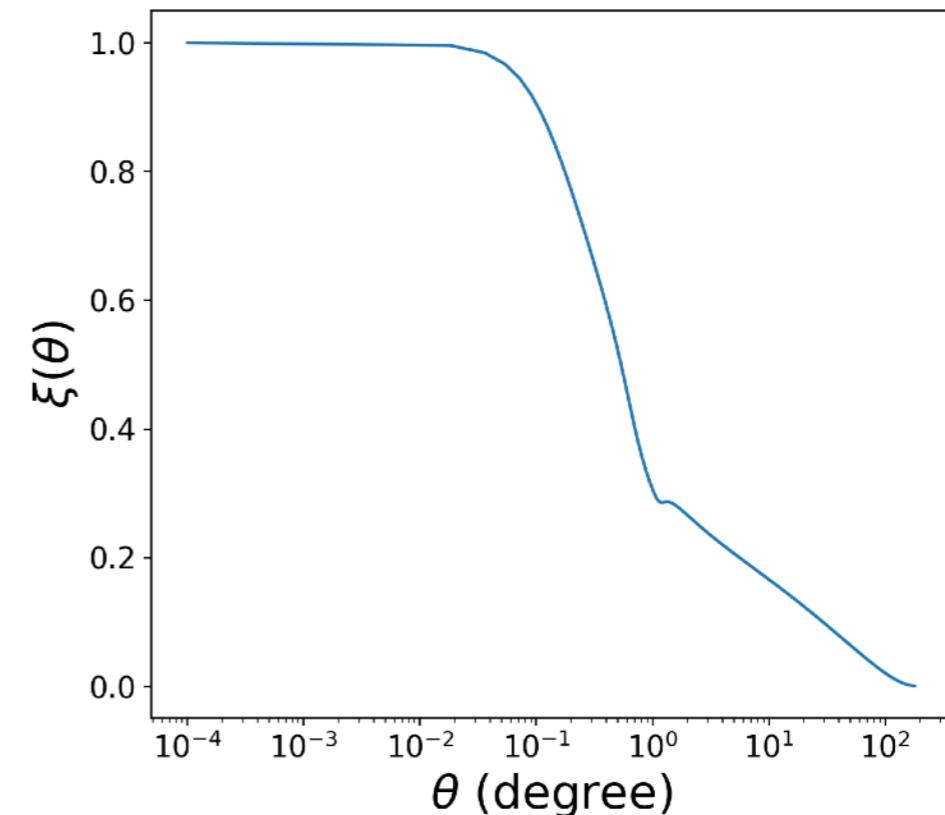
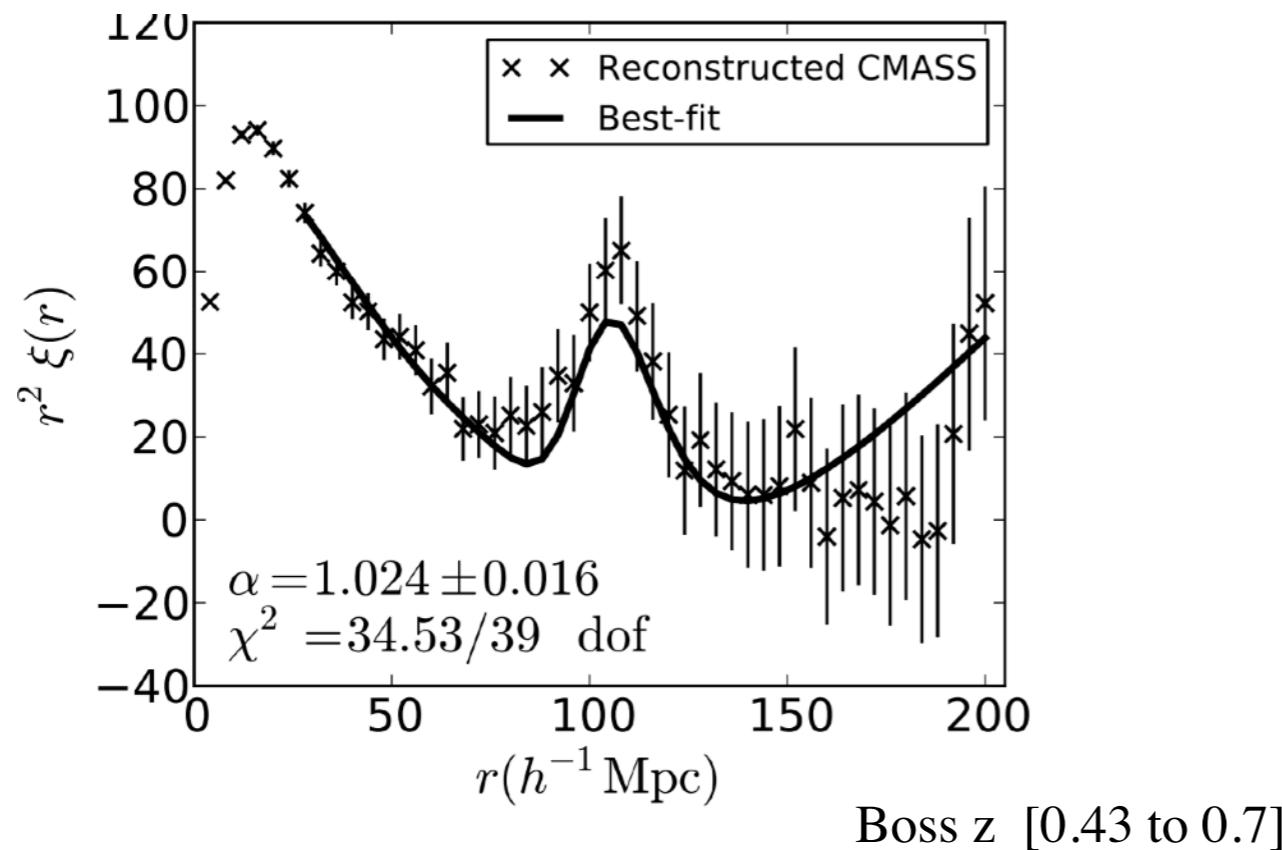


Matrice de covariance

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

$$\begin{aligned} C_{lm} &= \langle \delta T(\hat{\mathbf{n}}_l) \delta T(\hat{\mathbf{n}}_m) \rangle \\ &= \xi(\hat{\mathbf{n}}_l \cdot \hat{\mathbf{n}}_m) \\ &= \xi(\theta) \end{aligned}$$

CMB temperature angular correlation function



Pour représenter des champs à 3 dimensions on utilisera souvent l'espace de Fourier

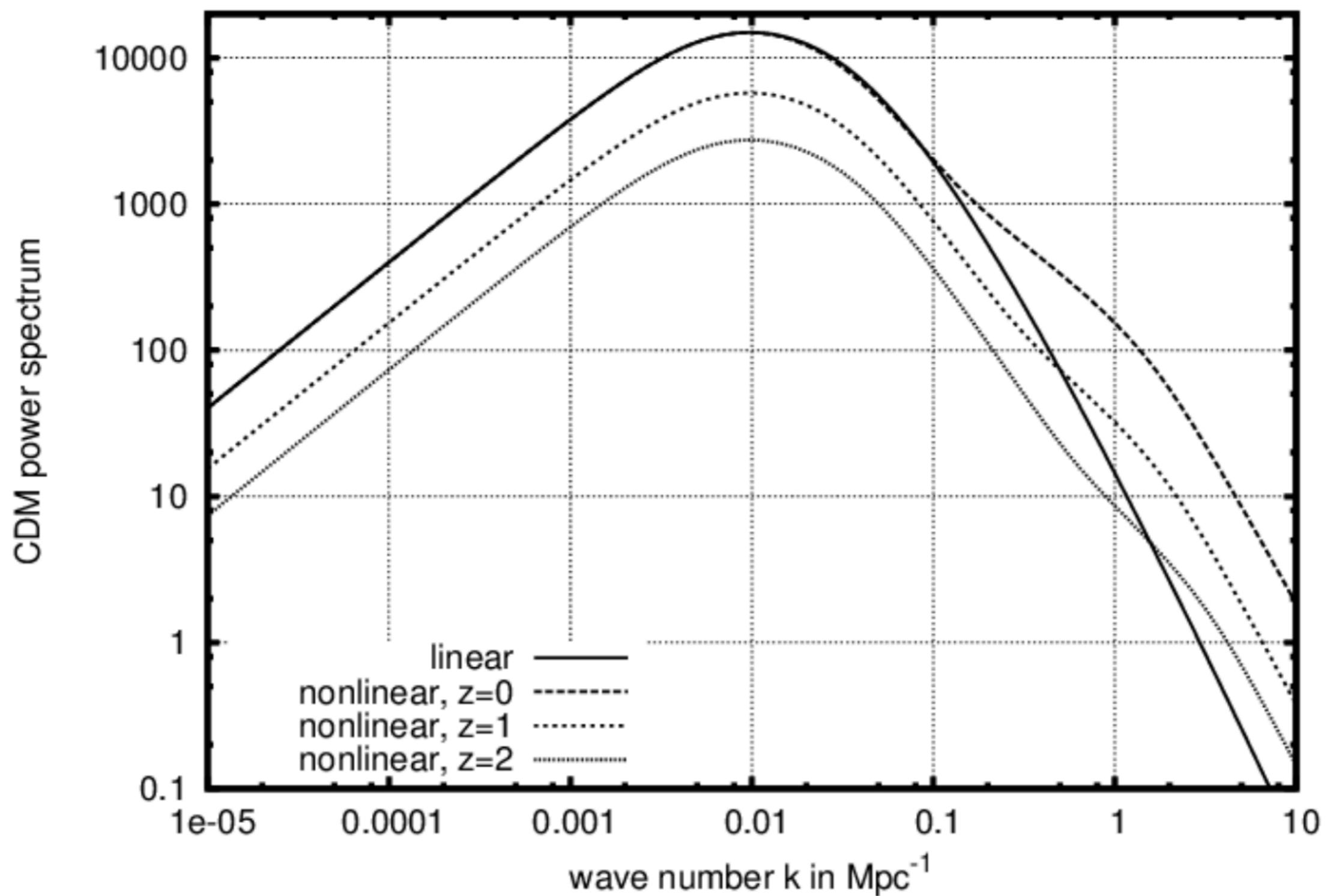
$$\delta(\mathbf{k}) = \int d\mathbf{x} \delta(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}}$$
$$\delta(\mathbf{x}) = \int \frac{d\mathbf{k}}{(2\pi)^3} \delta(\mathbf{k}) e^{-i\mathbf{k}\mathbf{x}}$$

L'analogie de la fonction de corrélation à deux points est le spectre en puissance $P(k)$

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

Cet espace présente différents avantages :

- 1) **Les calculs théoriques d'évolution des perturbations se font traditionnellement dans l'espace de Fourier**
- 2) **La base que représente les modes de Fourier diagonalise la matrice de covariance**
- 3) **Il est naturel d'implémenter des « coupures d'échelles »**



Fonction de corrélation dans l'espace de Fourier

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

Fonction de corrélation dans l'espace de Fourier

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

Coordonnées de Jacobi

$$\begin{aligned}\mathbf{r} &= \mathbf{x} - \mathbf{y} \\ \mathbf{x}_c &= (\mathbf{x} + \mathbf{y})/2\end{aligned}$$

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

Fonction de corrélation dans l'espace de Fourier

$$\begin{aligned}
 \langle \delta(\mathbf{k})\delta(\mathbf{q})^* \rangle &= \left\langle \int d\mathbf{x} \delta(\mathbf{x}) e^{i\mathbf{k}\mathbf{x}} \int d\mathbf{y} \delta(\mathbf{y}) e^{-i\mathbf{q}\mathbf{y}} \right\rangle \\
 &= \int d\mathbf{x} d\mathbf{y} \langle \delta(\mathbf{x})\delta(\mathbf{y}) \rangle e^{i(\mathbf{k}\mathbf{x}-\mathbf{q}\mathbf{y})} \\
 &= \int d\mathbf{x} d\mathbf{y} \xi(\mathbf{x} - \mathbf{y}) e^{i(\mathbf{k}\mathbf{x}-\mathbf{q}\mathbf{y})} \\
 &= \int d\mathbf{x}_c e^{i(\mathbf{k}-\mathbf{q})\mathbf{x}_c} \int d\mathbf{r} \xi(\mathbf{r}) e^{i\frac{(\mathbf{k}+\mathbf{q})}{2}\mathbf{r}} \\
 &= (2\pi)^3 \delta(\mathbf{k} - \mathbf{q}) \int d\mathbf{r} \xi(\mathbf{r}) e^{i\mathbf{k}\mathbf{r}} \\
 &= (2\pi)^3 \delta(\mathbf{k} - \mathbf{q}) P(\mathbf{k})
 \end{aligned}$$

$\mathbf{r} = \mathbf{x} - \mathbf{y}$
 $\mathbf{x}_c = (\mathbf{x} + \mathbf{y})/2$
 $(\mathbf{k} - \mathbf{q})\mathbf{x}_c = \frac{1}{2}[\mathbf{k}\mathbf{x} + \mathbf{k}\mathbf{y} - \mathbf{q}\mathbf{x} - \mathbf{q}\mathbf{y}]$
 $\frac{(\mathbf{k} + \mathbf{q})}{2}\mathbf{r} = \frac{1}{2}[\mathbf{k}\mathbf{x} - \mathbf{k}\mathbf{y} + \mathbf{q}\mathbf{x} - \mathbf{q}\mathbf{y}]$
 $(\mathbf{k}\mathbf{x} - \mathbf{q}\mathbf{y}) = \frac{(\mathbf{k} + \mathbf{q})}{2}\mathbf{r} + (\mathbf{k} - \mathbf{q})\mathbf{x}_c$

Les différents modes de Fourier sont indépendants les uns des autres

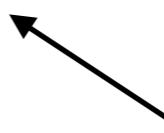
Spectre en puissance angulaire

Sur la sphère l'équivalent de la base de Fourier est la base des harmoniques sphériques

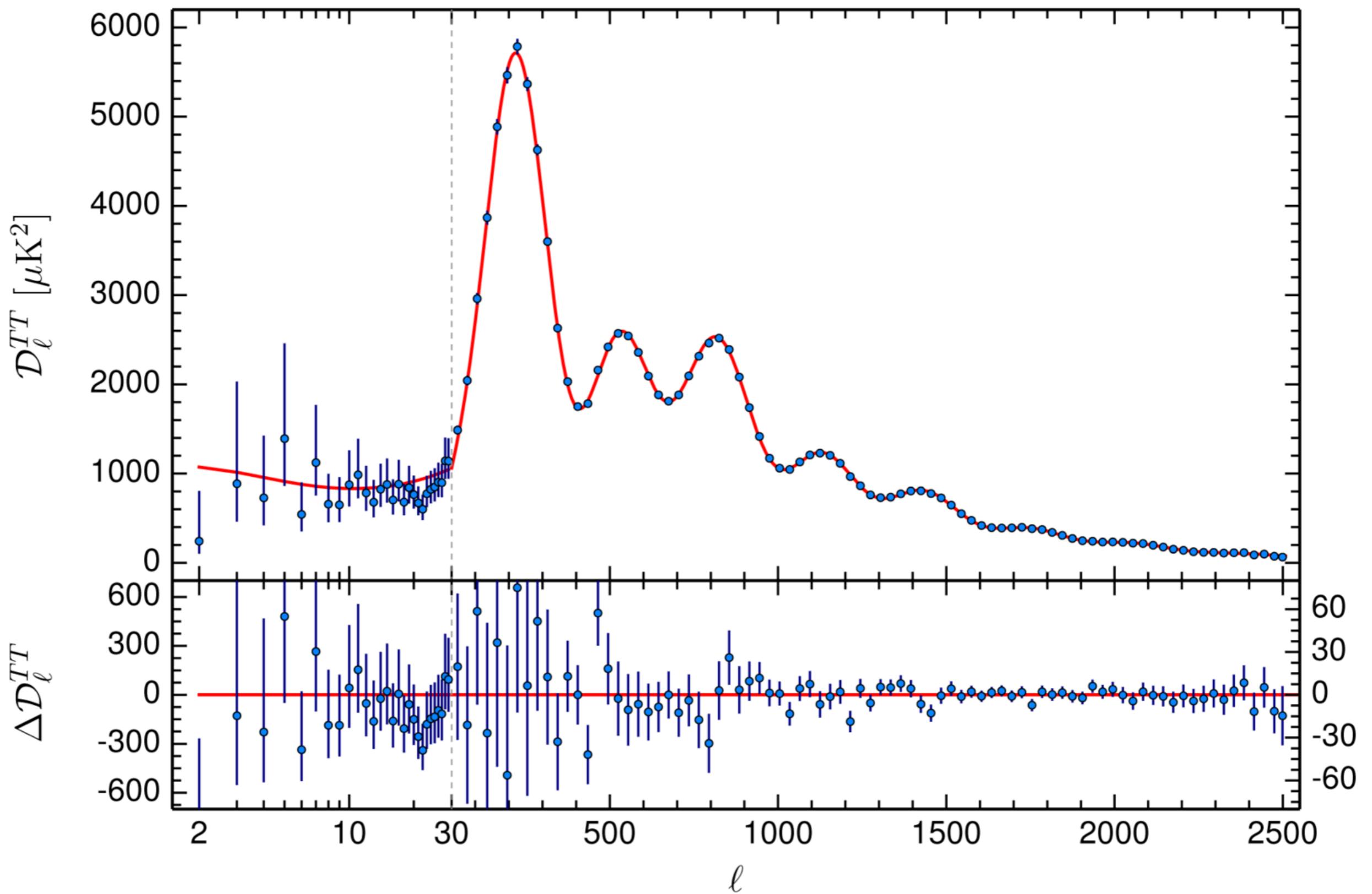
$$\delta T(\hat{n}) = \sum_{\ell m} a_{\ell m} Y_{\ell m}(\hat{n})$$
$$a_{\ell m} = \int \delta T(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n}$$

On peut définir un spectre en puissance angulaire ainsi qu'une fonction de corrélation angulaire

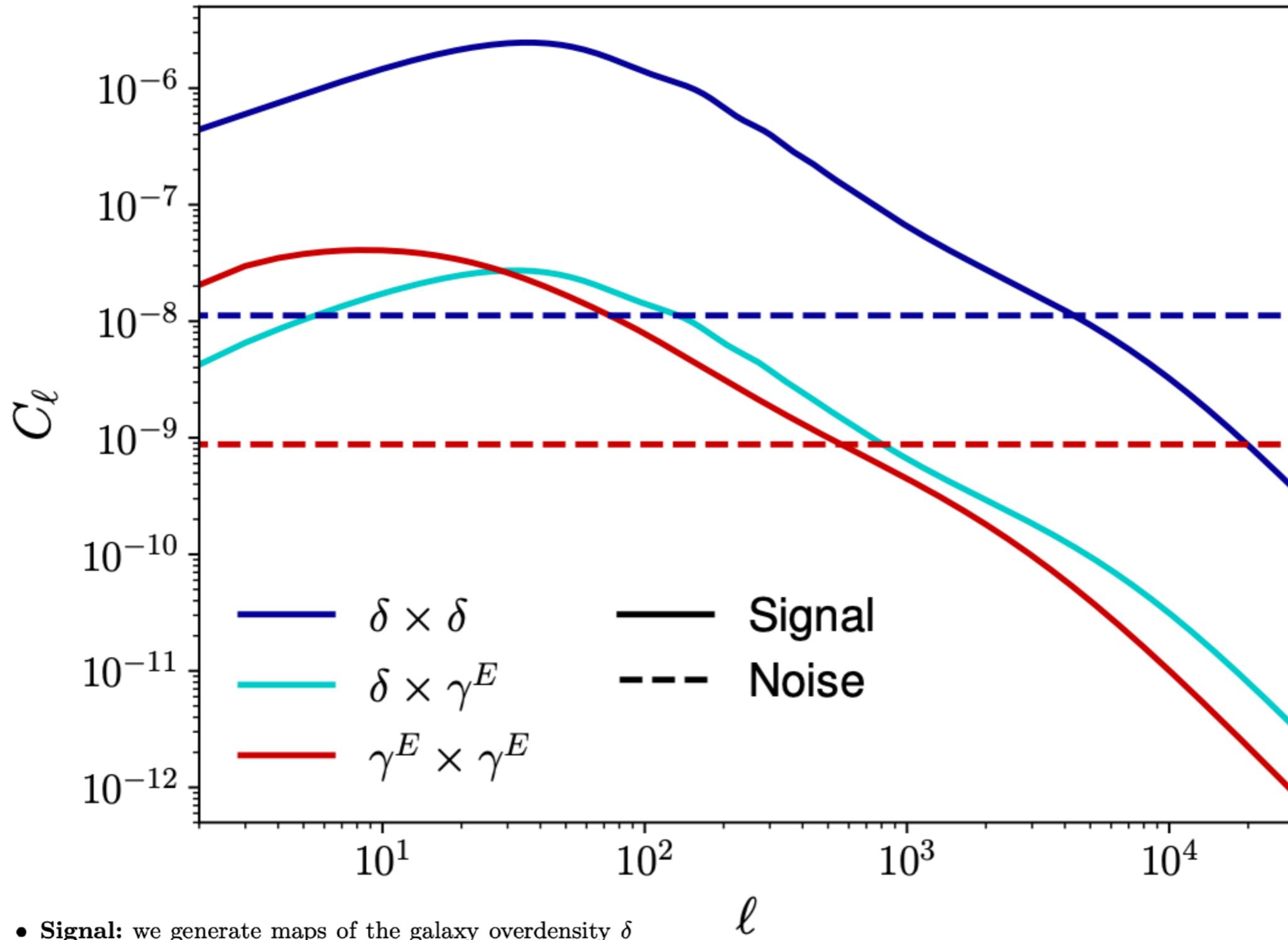
$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell$$
$$\xi(\theta) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} P_\ell(\cos \theta) C_\ell$$



Polynôme de Legendre



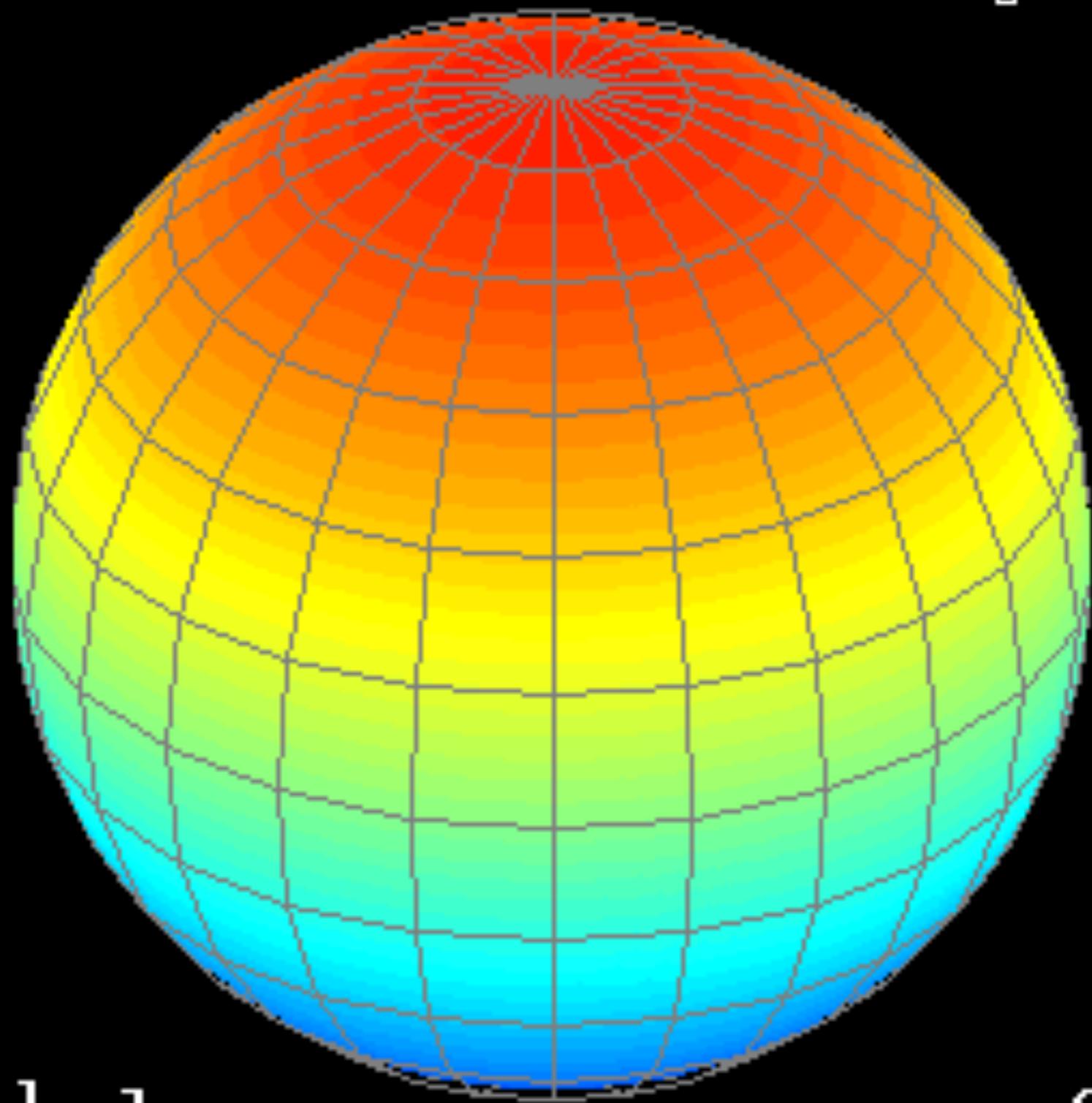
Planck collaboration



- **Signal:** we generate maps of the galaxy overdensity δ and of the shear field (γ_1, γ_2) for a redshift bin centred at $z \simeq 1$ with a width $\Delta z \sim 0.1$ and Gaussian photo- z tails with width $\sigma_z = 0.06$. For δ_g we assume unit bias $b(z) = 1$ and no contributions from redshift-space distortions or magnification. For γ , we assume no contribution from intrinsic alignments. The maps are generated as Gaussian realizations of the corresponding angular power spectra computed with the Core Cosmology Library¹² for cosmological parameters

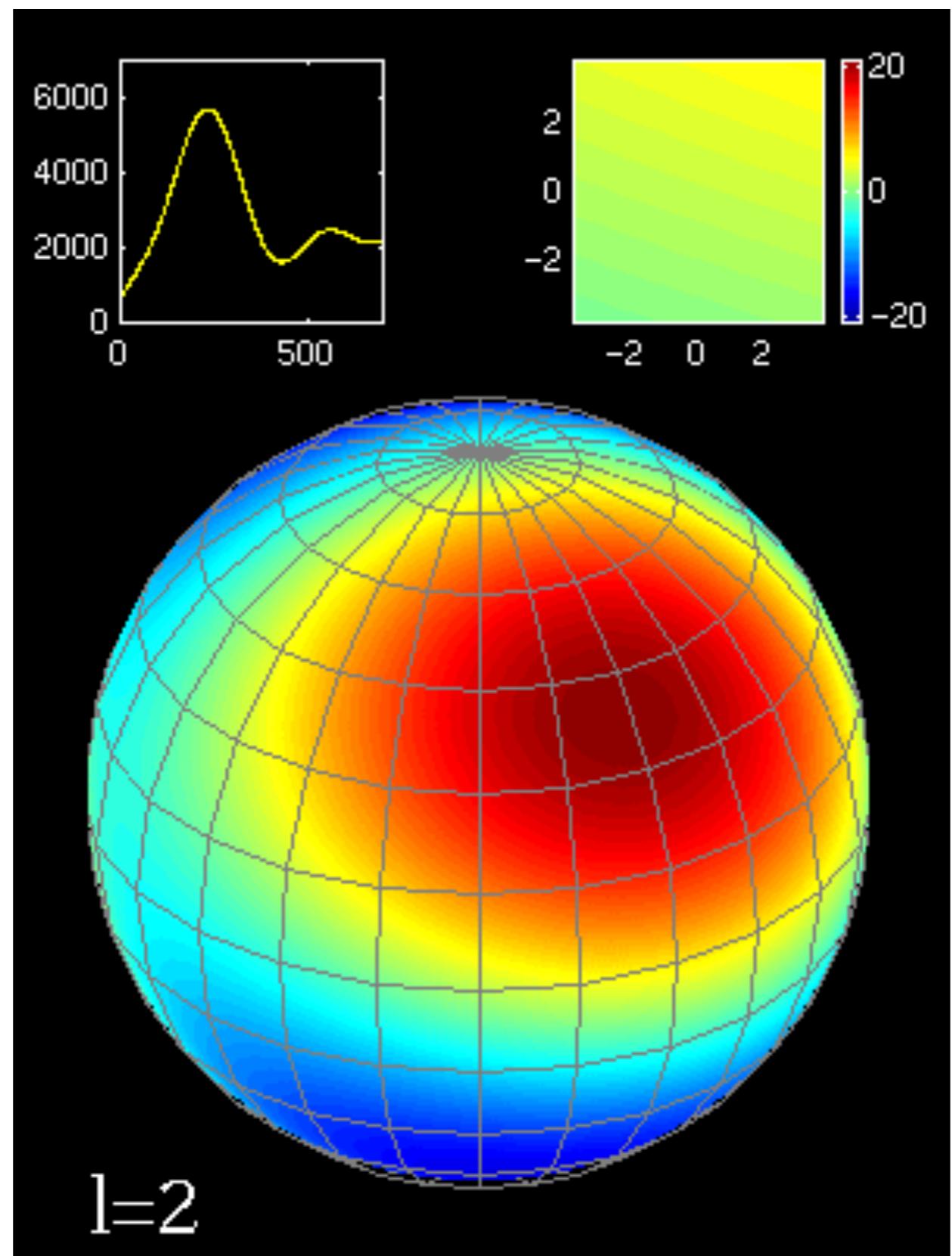
<https://arxiv.org/pdf/1809.09603.pdf>

color range=±0.69



$l=1$

$m=0$



Estimation des spectres en puissance angulaire du CMB

La définition du spectre en puissance étant

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

On peut définir un estimateur

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

Cet estimateur est non biaisé

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

On peut calculer la variance de cet estimateur

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

Pour conclure on a besoin de calculer $\langle \hat{C}_\ell^2 \rangle$

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

Pour calculer ce terme on va utiliser plusieurs propriétés

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

- 1) **Les transformations linéaires sur des champs gaussiens suivent une statistique Gaussienne**

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

↑
Gaussien

Pour calculer ce terme on va utiliser plusieurs propriétés

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

- 1) **Les transformations linéaires sur des champs gaussiens suivent une statistique Gaussienne**

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

Gaussien Gaussien

Pour calculer ce terme on va utiliser plusieurs propriétés

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

- 1) **Les transformations linéaires sur des champs gaussiens suivent une statistique Gaussienne**

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

Gaussien Gaussien

- 2) **Le théorème de Wick**

Le théorème de Wick

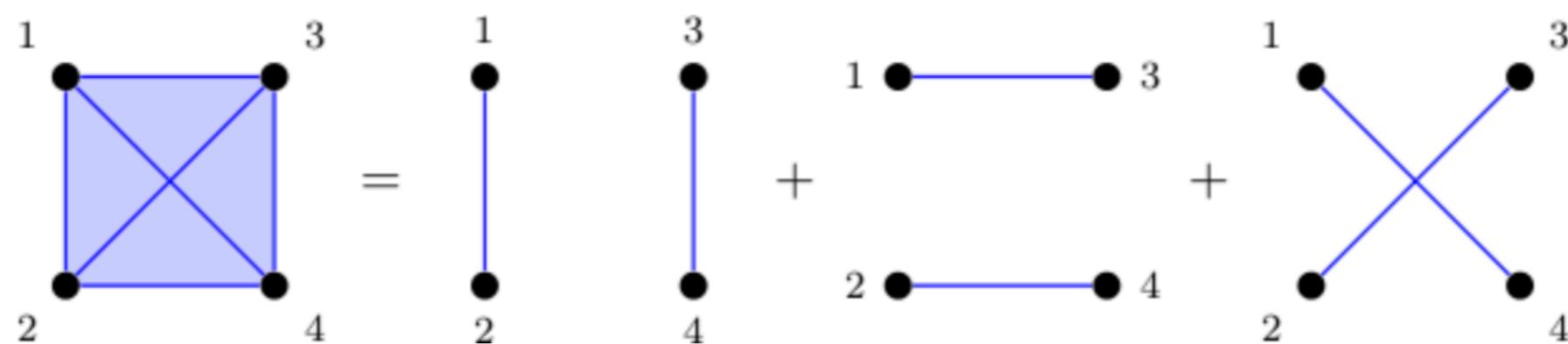
On peut décomposer la fonction à N points d'un champ Gaussien en somme de produits de fonctions de corrélation à deux points

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \mathbb{E}[X_i X_j] = \sum_{p \in P_n^2} \prod_{\{i,j\} \in p} \text{Cov}(X_i, X_j)$$

La somme est faite sur toutes les paires de (1...n)

Dans le cas d'une fonction de corrélation à 4 points

$$\langle X_1 X_2 X_3 X_4 \rangle = \langle X_1 X_2 \rangle \langle X_3 X_4 \rangle + \langle X_1 X_3 \rangle \langle X_2 X_4 \rangle + \langle X_1 X_4 \rangle \langle X_2 X_3 \rangle$$



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

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

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

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

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

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

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

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

Ce terme représente l'incertitude inhérente à toute mesure de spectre en puissance, il est connu sous le nom de cosmic variance et est irréductible. Notre mesure est intrinsèquement limitée par le fait que nous n'avons accès qu'à une « réalisation » de l'univers

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

Le dénominateur est le nombre de modes: le nombre de « m » par l

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

Il existe un analogue pour la variance cosmique du spectre en puissance de la matière à trois dimensions

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

Biais de bruit

En réalité toute observation est impactée par des sources de bruit

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

Le bruit va biaiser l'estimateur et augmenter sa variance

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

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

Biais de bruit

En réalité toute observation est impactée par des sources de bruit

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

Le bruit va biaiser l'estimateur et augmenter sa variance

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

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

Notez que quand $N_\ell=0$ on retrouve bien la variance cosmique

Biais de bruit

En réalité toute observation est impactée par des sources de bruit

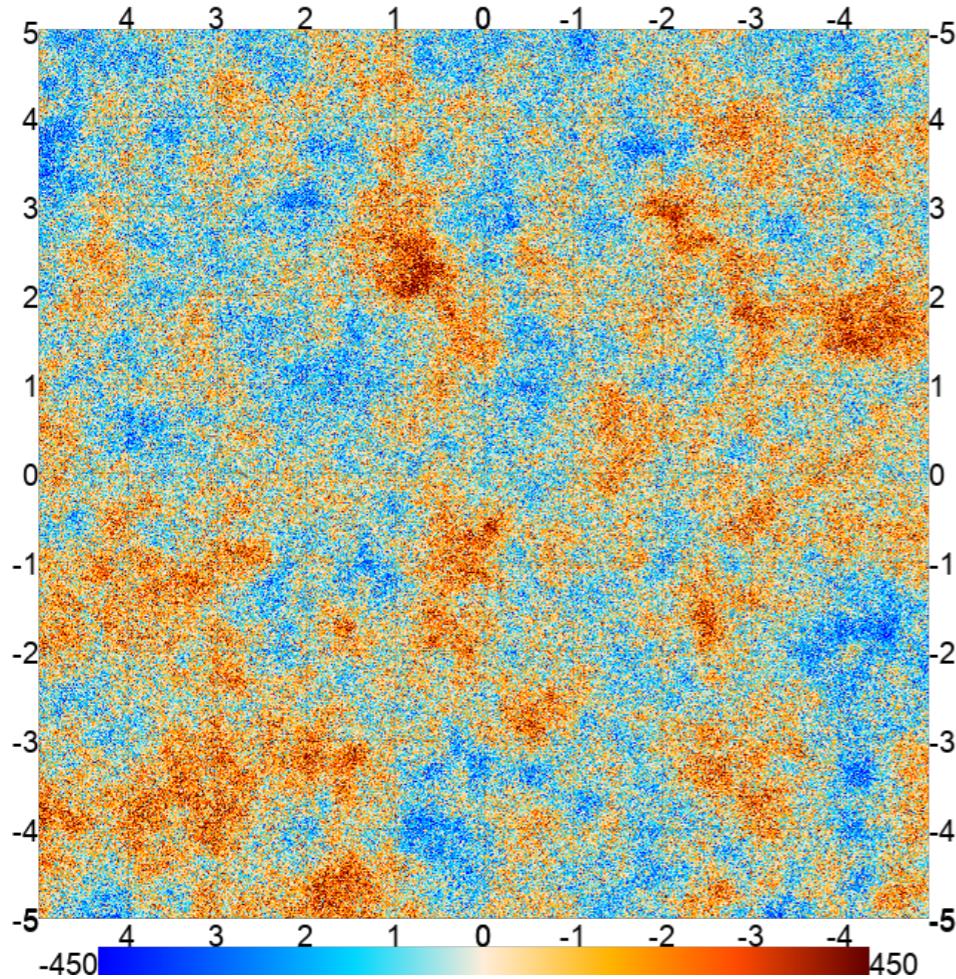
$$\begin{aligned}\delta T^{\text{obs}}(\hat{n}) &= \delta T^{\text{CMB}}(\hat{n}) + n(\hat{n}) \\ a_{\ell m}^{\text{obs}} &= \int \delta T^{\text{obs}}(\hat{n}) Y_{\ell m}^*(\hat{n}) d\hat{n} \\ a_{\ell m}^{\text{obs}} &= a_{\ell m}^{\text{CMB}} + n_{\ell m}\end{aligned}$$

On peut débiaiser l'estimateur de spectre de différentes façons

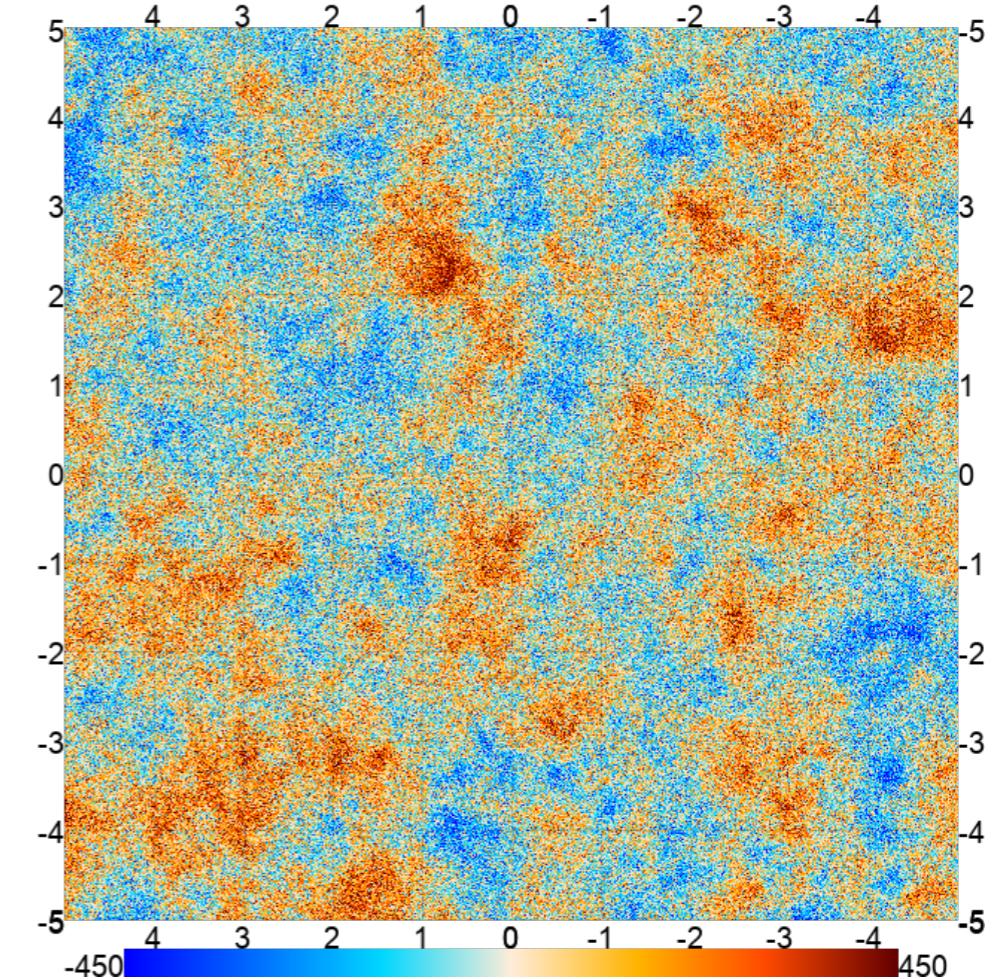
- 1) Soustraire un modèle de bruit formé à partir de nos connaissances de l'instrument
- 2) Faire des cross corrélations

Cross corrélations

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



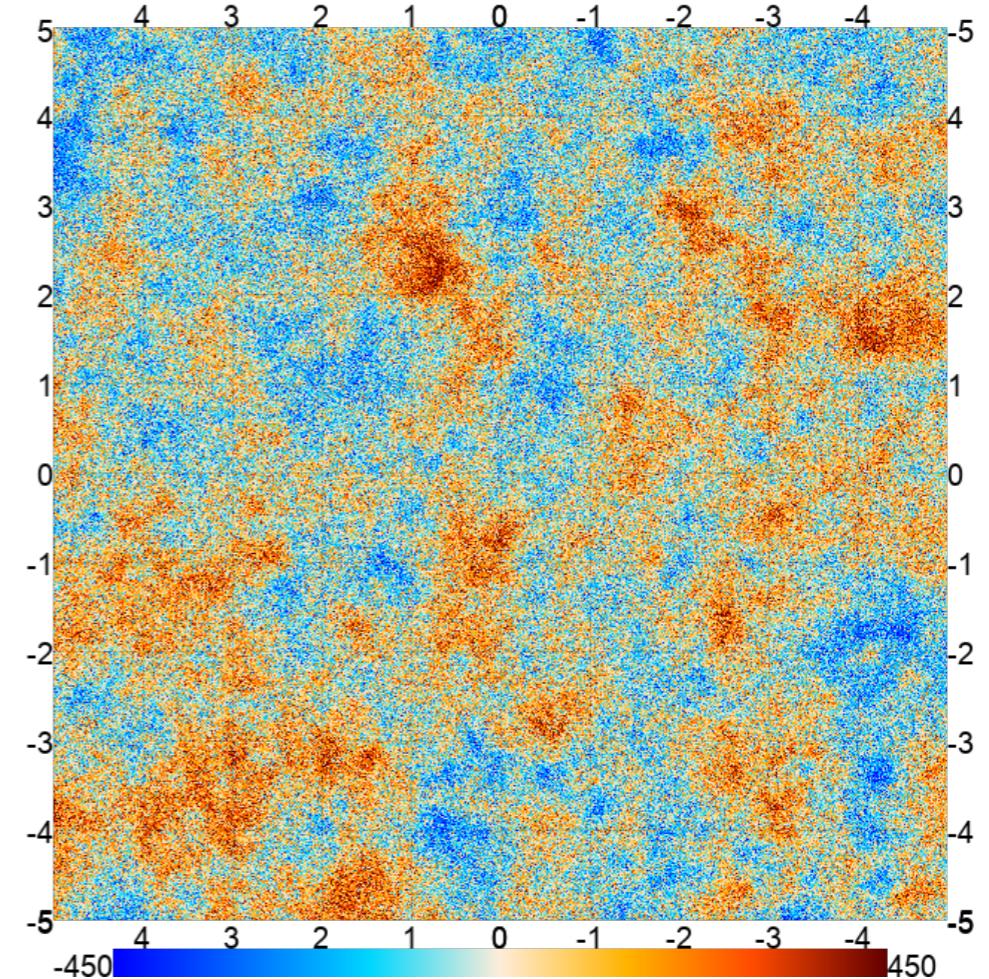
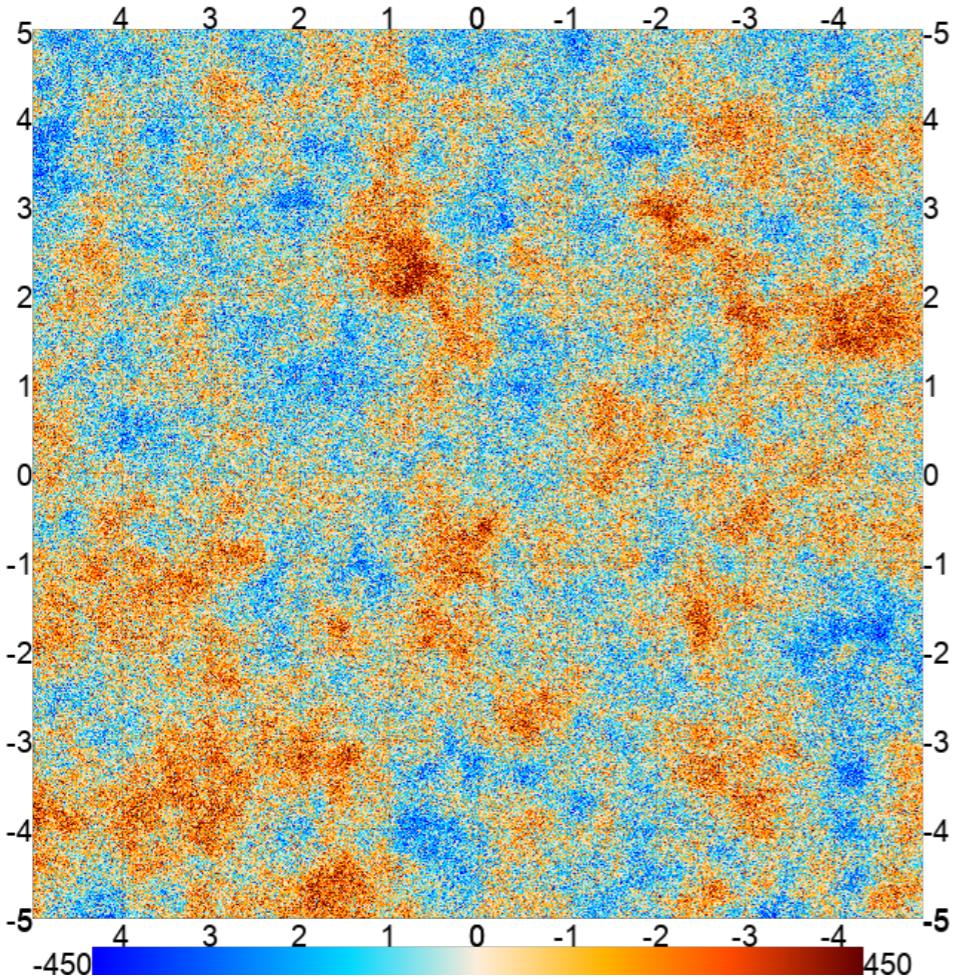
$$\delta T_2^{\text{obs}}(\hat{n}) = \delta T^{\text{CMB}}(\hat{n}) + n_2(\hat{n})$$



Cross corrélations

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

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



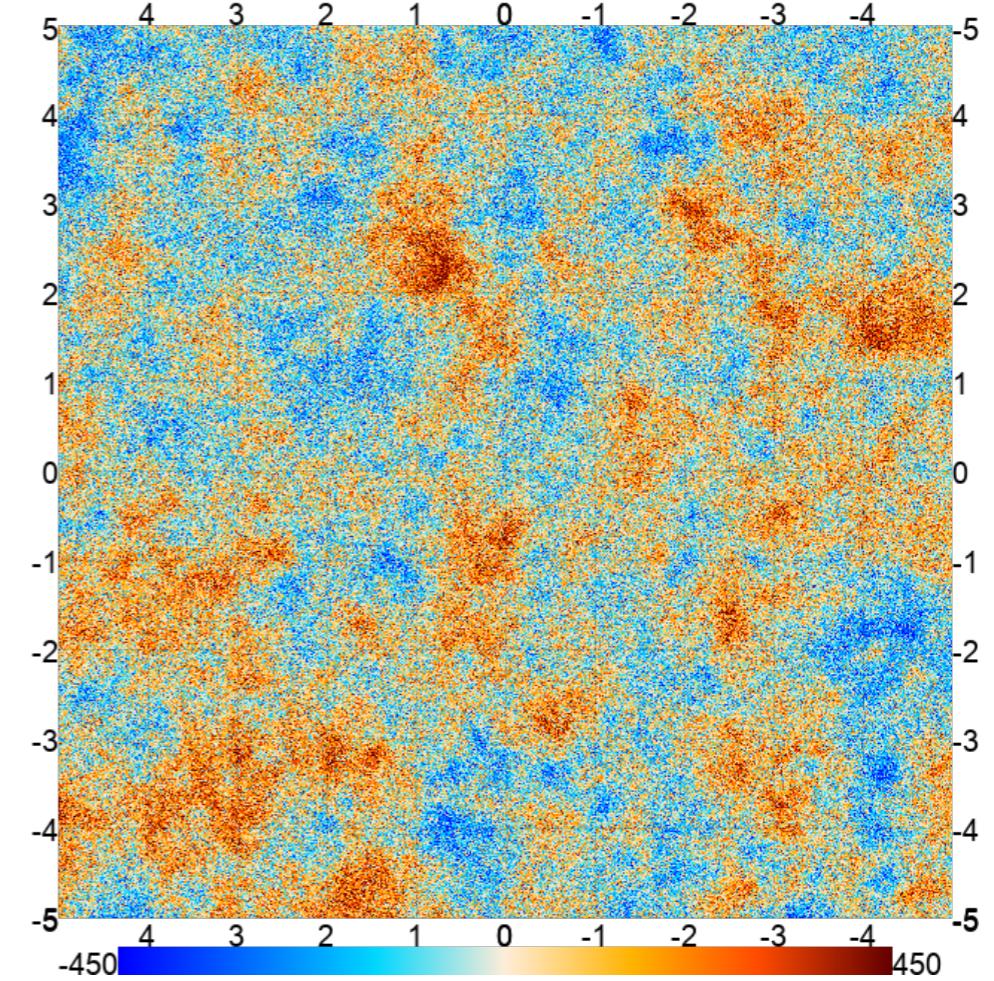
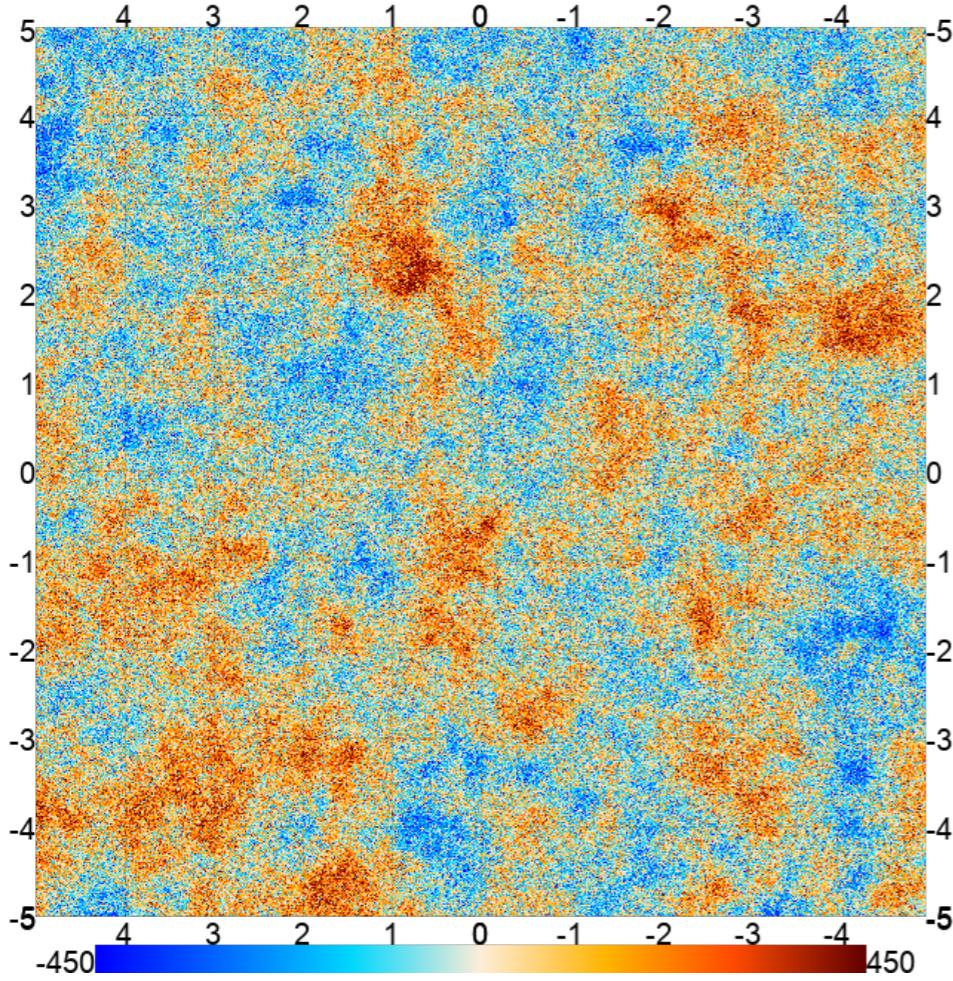
$$a_{1,\ell m}^{\text{obs}} = a_{\ell m}^{\text{CMB}} + n_{1,\ell m}$$

$$a_{2,\ell m}^{\text{obs}} = a_{\ell m}^{\text{CMB}} + n_{2,\ell m}$$

Cross corrélations

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

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



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

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

Cross corrélations

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

vs

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

Cross corrélations

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

vs

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

En général pour K splits de données

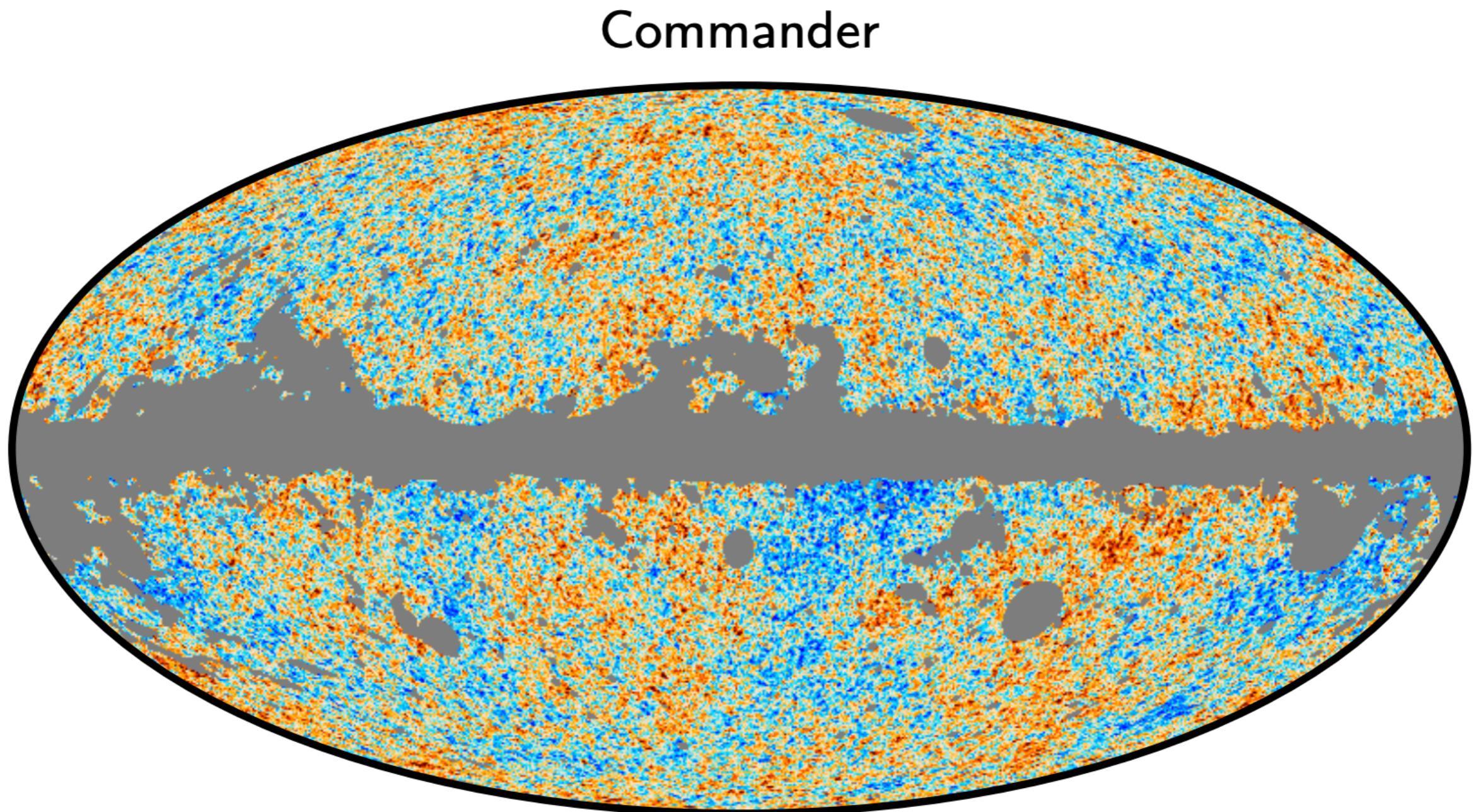
$$\langle \hat{C}_\ell^{\text{obs}} \rangle = \frac{1}{K(K-1)/2} \sum_{i \neq j} \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \langle a_{i,\ell m}^{\text{obs}} a_{j,\ell m}^{\text{obs},*} \rangle = C_\ell^{\text{CMB}}$$

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

Effet du masque

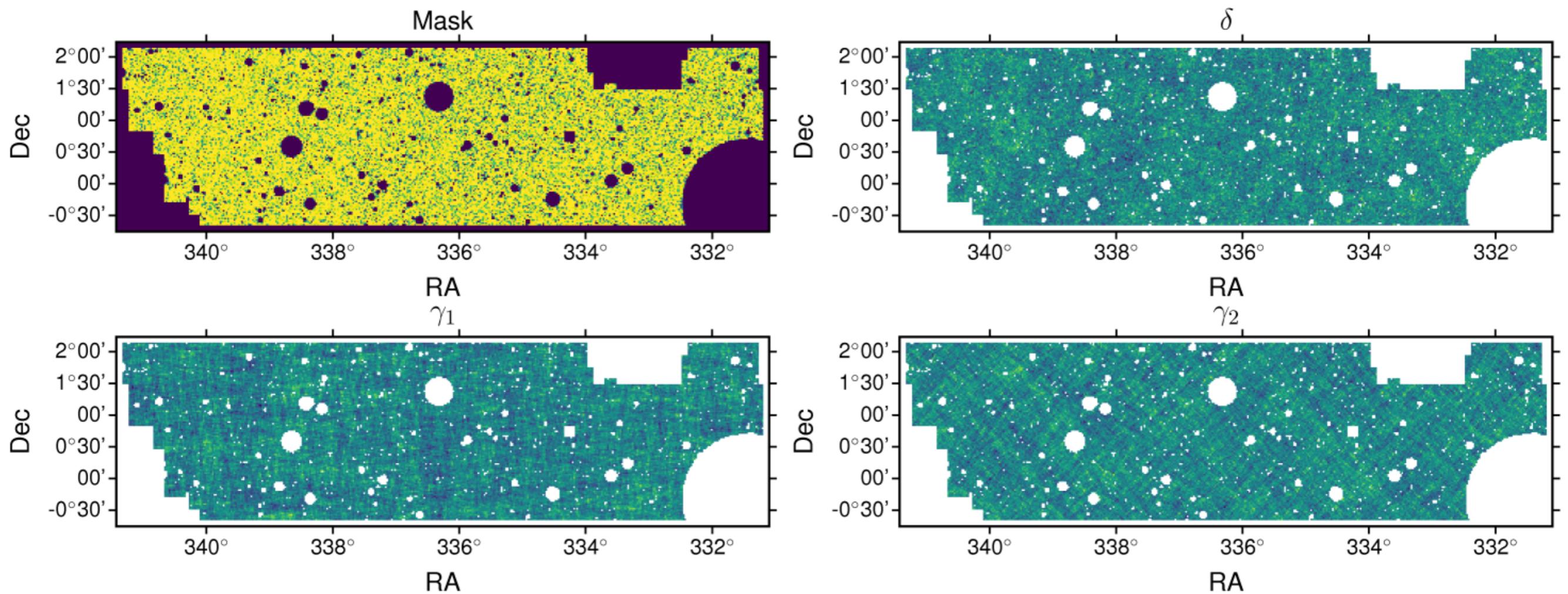


Effet du masque



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

Effet du masque



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

Effet du masque

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

Effet du masque

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

$$\begin{aligned}a_{\ell m}^{\text{obs}} &= \int W(\hat{n})\delta T(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} a_{\ell' m'} \int W(\hat{n})Y_{\ell' m'}(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}\end{aligned}$$

Si $W(n)=1$, cette intégrale vaut $\delta_{\ell\ell'}\delta_{mm'}$

Le masque brise l'homogénéité du champ et introduit un couplage entre les différents modes

Effet du masque

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

$$\begin{aligned}a_{\ell m}^{\text{obs}} &= \int W(\hat{n})\delta T(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} a_{\ell' m'} \int W(\hat{n})Y_{\ell' m'}(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}\end{aligned}$$

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

Effet du masque

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

$$\begin{aligned}a_{\ell m}^{\text{obs}} &= \int W(\hat{n})\delta T(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} a_{\ell' m'} \int W(\hat{n})Y_{\ell' m'}(\hat{n})Y_{\ell m}^*(\hat{n})d\hat{n} \\ &= \sum_{\ell' m'} K_{\ell m, \ell' m'} a_{\ell' m'}\end{aligned}$$

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

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

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-CI (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

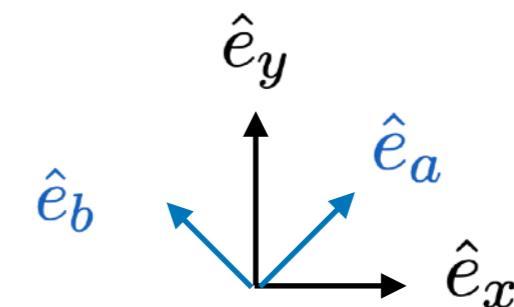
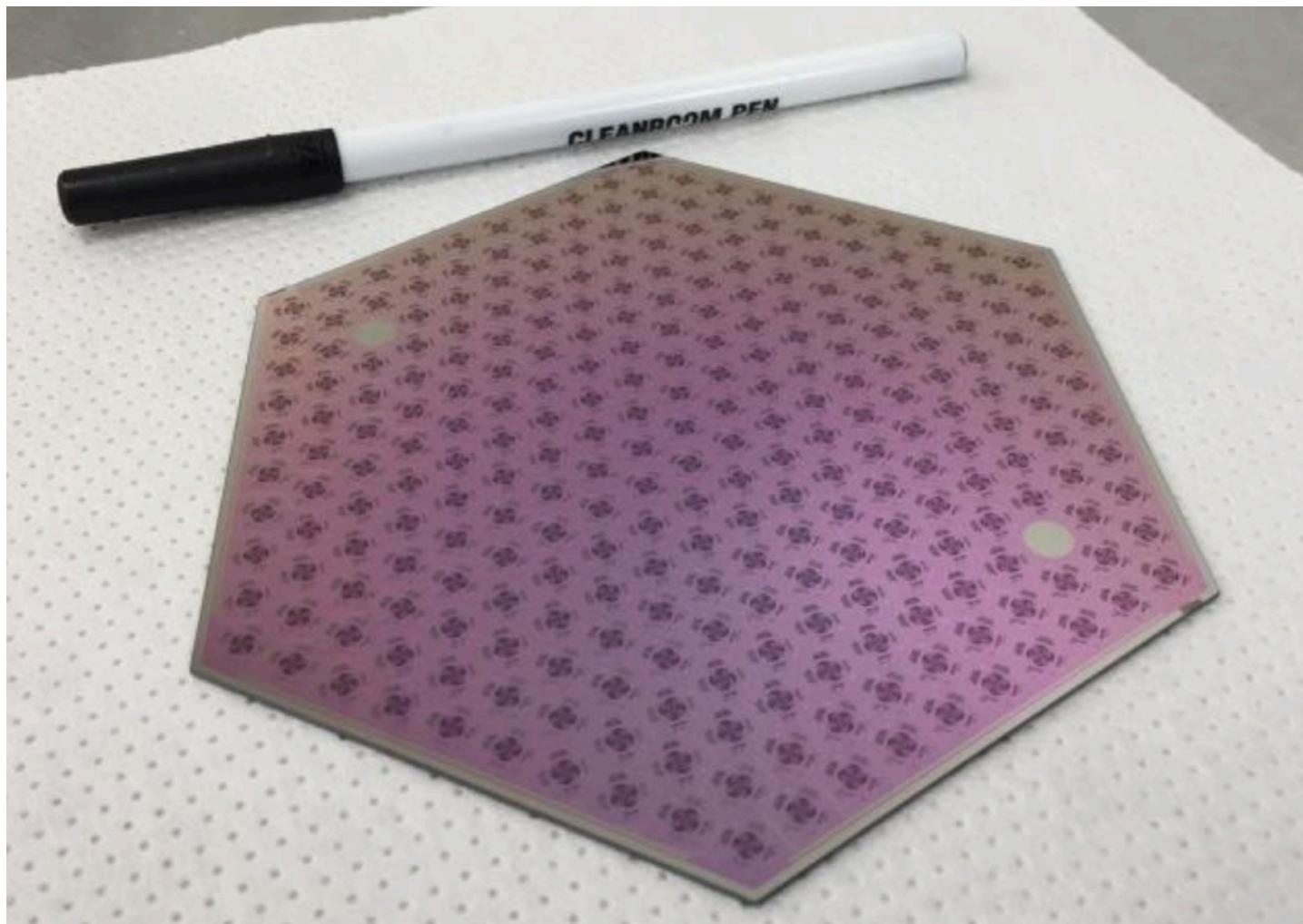
Un mot sur les champs de spin supérieur

Le champ de densité et le champ de température du CMB sont des champs scalaires (i.e de spin 0) leur valeur ne dépend pas du système de coordonnées utilisé.

A l'inverse le champ de polarisation du fond diffus cosmologique ainsi que le champ de cisaillement gravitationnel des galaxies sont des champs de spin 2, leurs valeurs en un point donné dépend du système de coordonnées.

Un mot sur les champs de spin supérieur

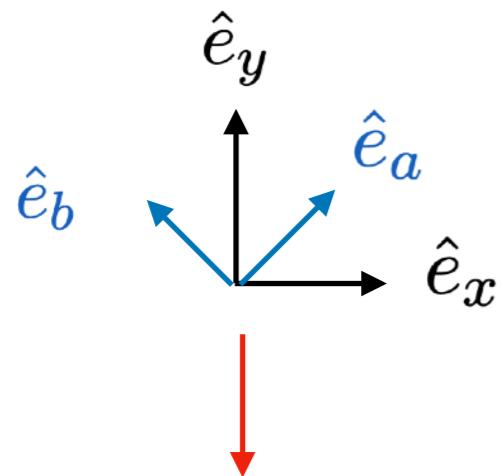
Par exemple le champ de polarisation est défini en termes des paramètres de Stokes qui quantifient la direction de polarisation du champ électrique E des photons du CMB



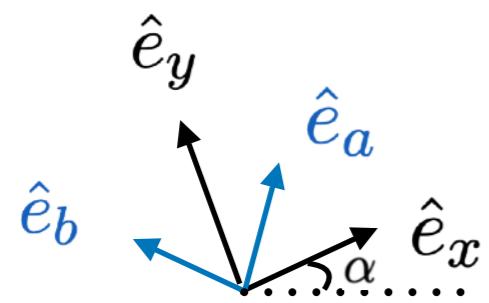
$$\begin{aligned} Q(\hat{n}) &= |E_x(\hat{n})|^2 - |E_y(\hat{n})|^2 \\ U(\hat{n}) &= |E_a(\hat{n})^2| - |E_b(\hat{n})^2| \end{aligned}$$

Un mot sur les champs de spin supérieur

Les paramètres de Stokes ne sont pas invariants sous un changement de coordonnées



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



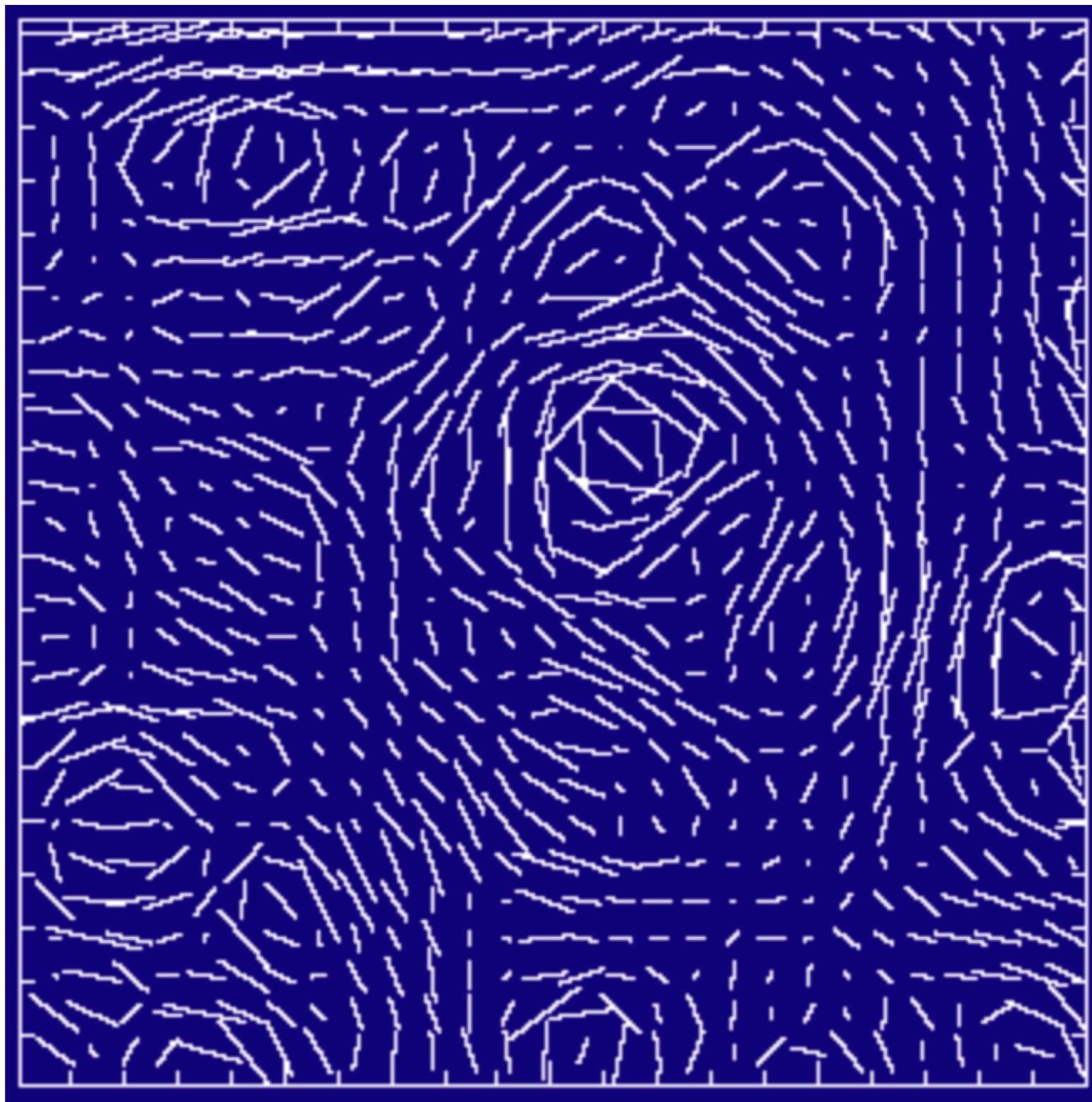
$$\begin{aligned} Q'(\hat{n}) &= Q(\hat{n}) \cos(2\alpha) + U(\hat{n}) \sin(2\alpha) \\ U'(\hat{n}) &= -Q(\hat{n}) \sin(2\alpha) + U(\hat{n}) \cos(2\alpha) \end{aligned}$$

On remarque que si $\alpha = \pi, Q' = Q, U' = U$

Cette propriété fait du champ de polarisation un champ de spin 2

C'est aussi la raison pour laquelle on peut représenter le champ de polarisation comme un champ de vecteurs « sans tête »

Champ de polarisation



Décomposition en modes E et B

Pour prendre en compte la transformation des paramètres de Stokes suivant une transformation de coordonnées on peut les écrire sous la forme d'un tenseur

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

Décomposition en mode E et B

Pour prendre en compte la transformation des paramètres de Stokes suivant une transformation de coordonnées on peut les écrire sous la forme d'un tenseur

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

De la même façon qu'on peut toujours décomposer un champ vectoriel en une partie scalaire et une partie rotationnelle

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

On peut toujours décomposer un champ tensoriel de spin 2 en la somme de deux champs, un champ scalaire et un champ pseudo-scalaire

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


Modes E de polarisation

Modes B de polarisation

Décomposition en mode E et B

Pour prendre en compte la transformation des paramètres de Stokes suivant une transformation de coordonnées, on peut écrire :

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

De la même façon :

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

On peut toutefois décomposer le champ électrique en trois parties : un champ scalaire, un champ pseudo-scalaire et un champ pseudo-scalaire

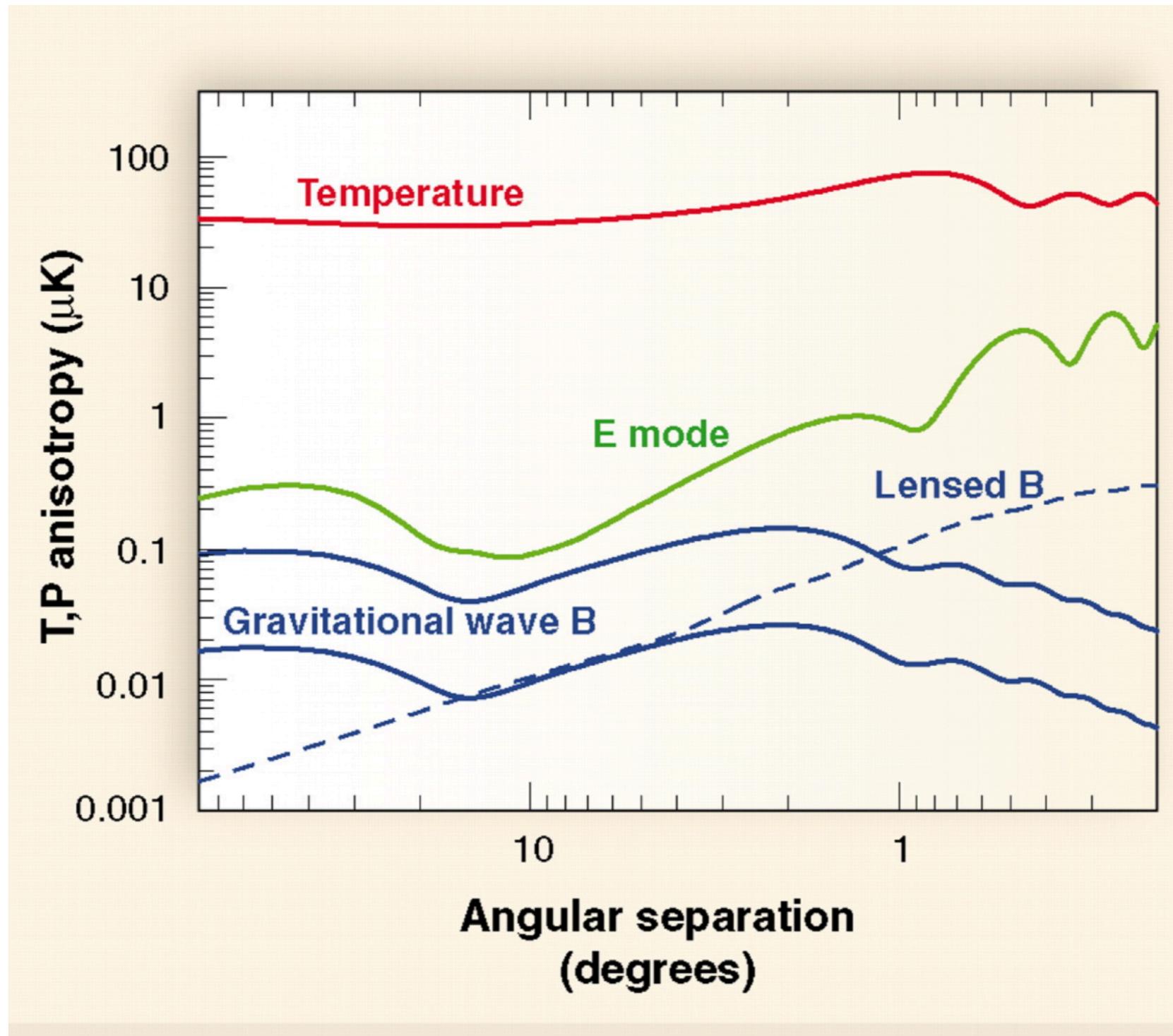
$$P_{ab}(\hat{n}) = \mathcal{E}_{ab}E(\hat{n}) + \mathcal{B}_{ab}B(\hat{n})$$



Modes E de polarisation

Modes B de polarisation

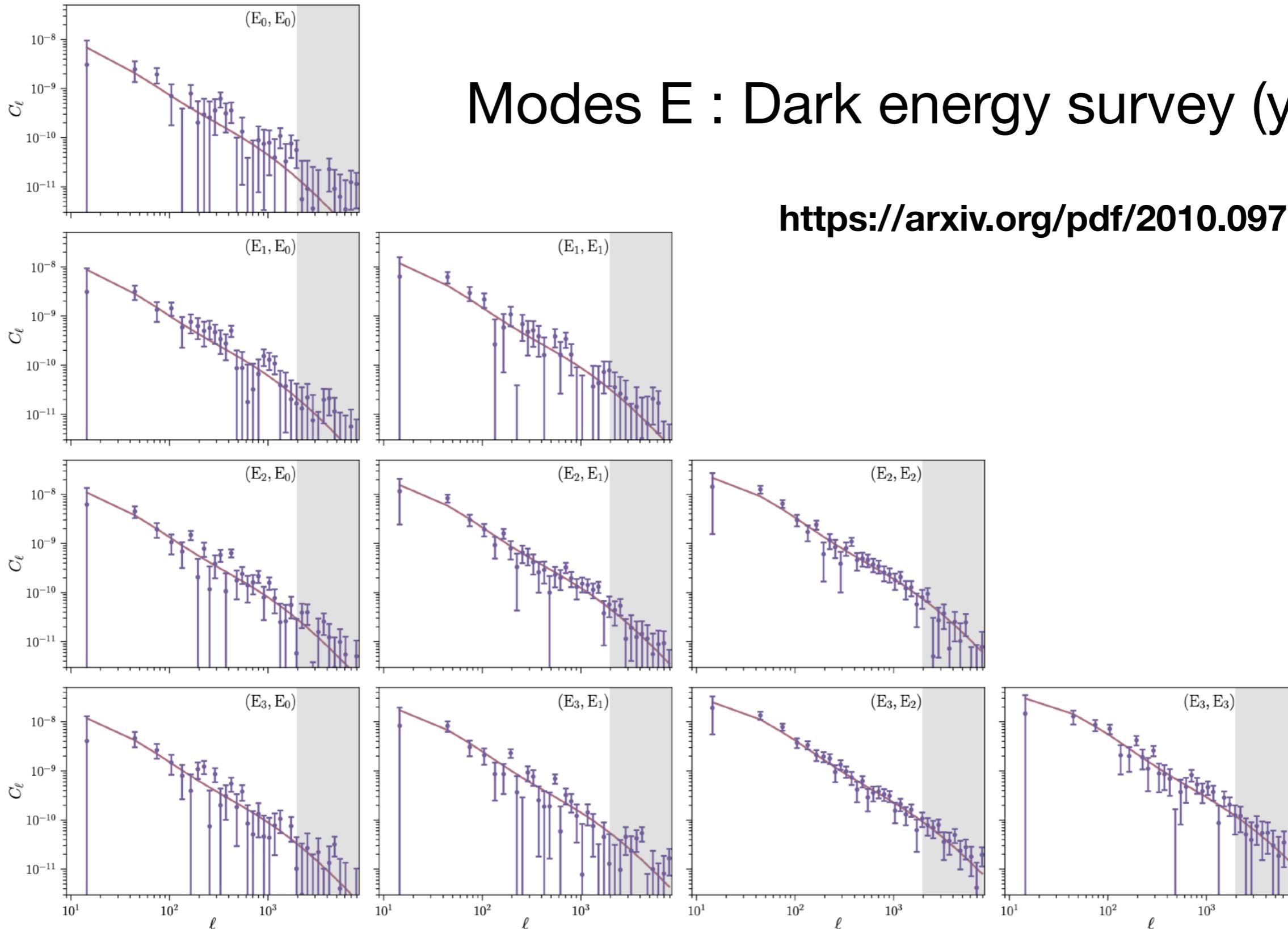
Cette décomposition a un réel intérêt, la plupart des phénomènes physiques dans l'Univers primordial ne génèrent que des modes E. Une possible détection des modes B de polarisation pourrait être la trace de la propagation d'ondes gravitationnelles primordiales générées pendant la phase d'inflation de l'Univers.



En ce qui concerne le cisaillement gravitationnel des galaxies, seul des modes E sont générés au 1er ordre, les modes B peuvent donc servir à tester la présence de possibles contaminations du signal (astrophysique ou instrumental)

Modes E : Dark energy survey (year 1)

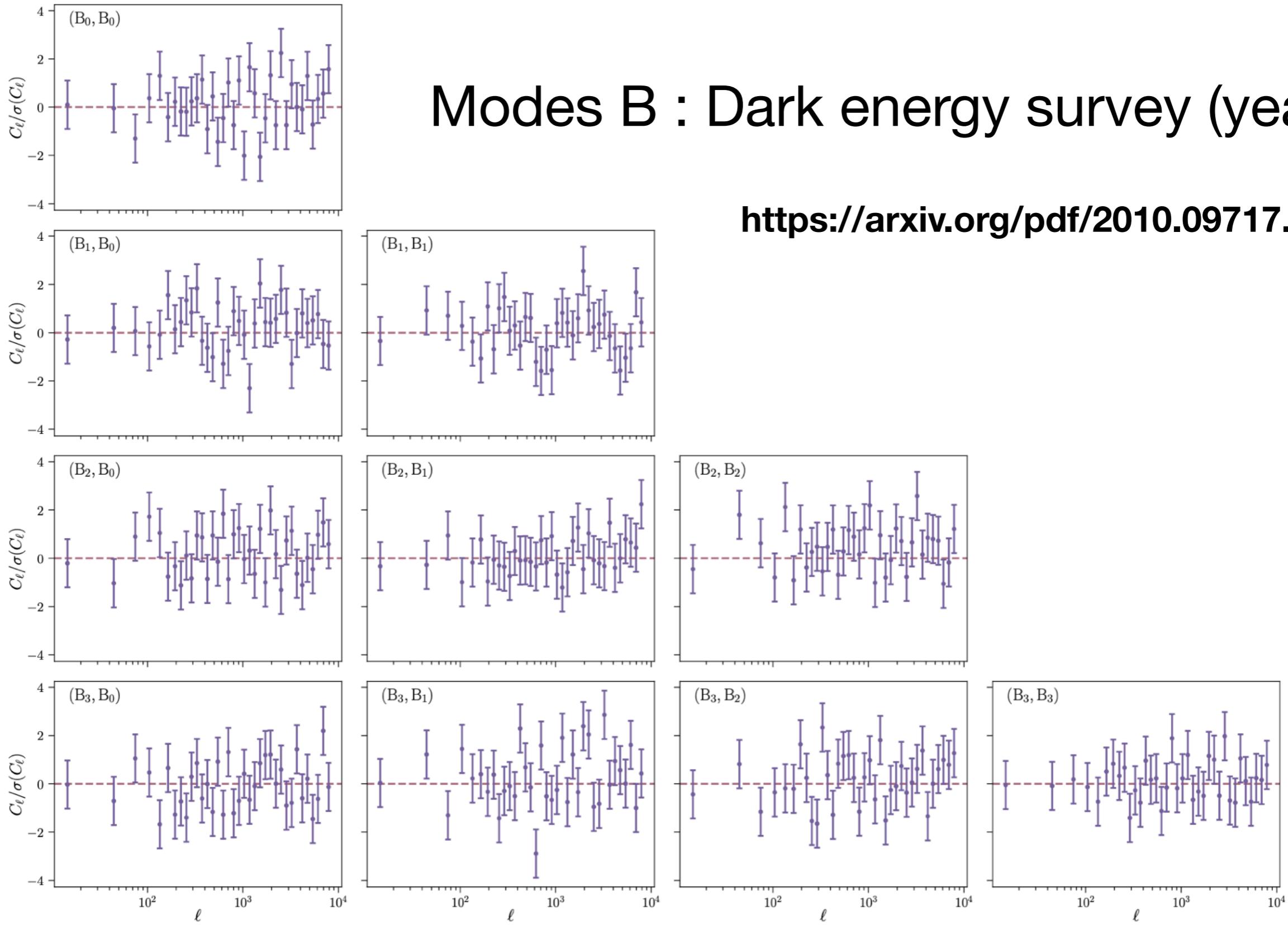
<https://arxiv.org/pdf/2010.09717.pdf>



En ce qui concerne le cisaillement gravitationnel des galaxies, seul des modes E sont générés au 1er ordre, les modes B peuvent donc servir à tester la présence de possibles contaminations du signal (astrophysique ou instrumental)

Modes B : Dark energy survey (year 1)

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



Principe ergodique (if time)

Un champ aléatoire est non-corrélé aux grandes échelles, si

$$\lim_{u \rightarrow +\infty} \left\langle \prod_{i=1}^p f(\mathbf{x}_i + \mathbf{u}) \prod_{j=1}^q f(\mathbf{x}_j - \mathbf{u}) \right\rangle = \left\langle \prod_{i=1}^p f(\mathbf{x}_i) \right\rangle \left\langle \prod_{j=1}^q f(\mathbf{x}_j) \right\rangle$$

Si un champ est homogène et non corrélé aux grandes échelles,
la différence entre la moyenne spatiale d'une région S de volume V

$$\langle f(\mathbf{x}_1) \dots f(\mathbf{x}_n) \rangle_V = \frac{1}{V} \int_S dz \hat{f}(\mathbf{x}_1 + \mathbf{z}) \dots \hat{f}(\mathbf{x}_n + \mathbf{z})$$

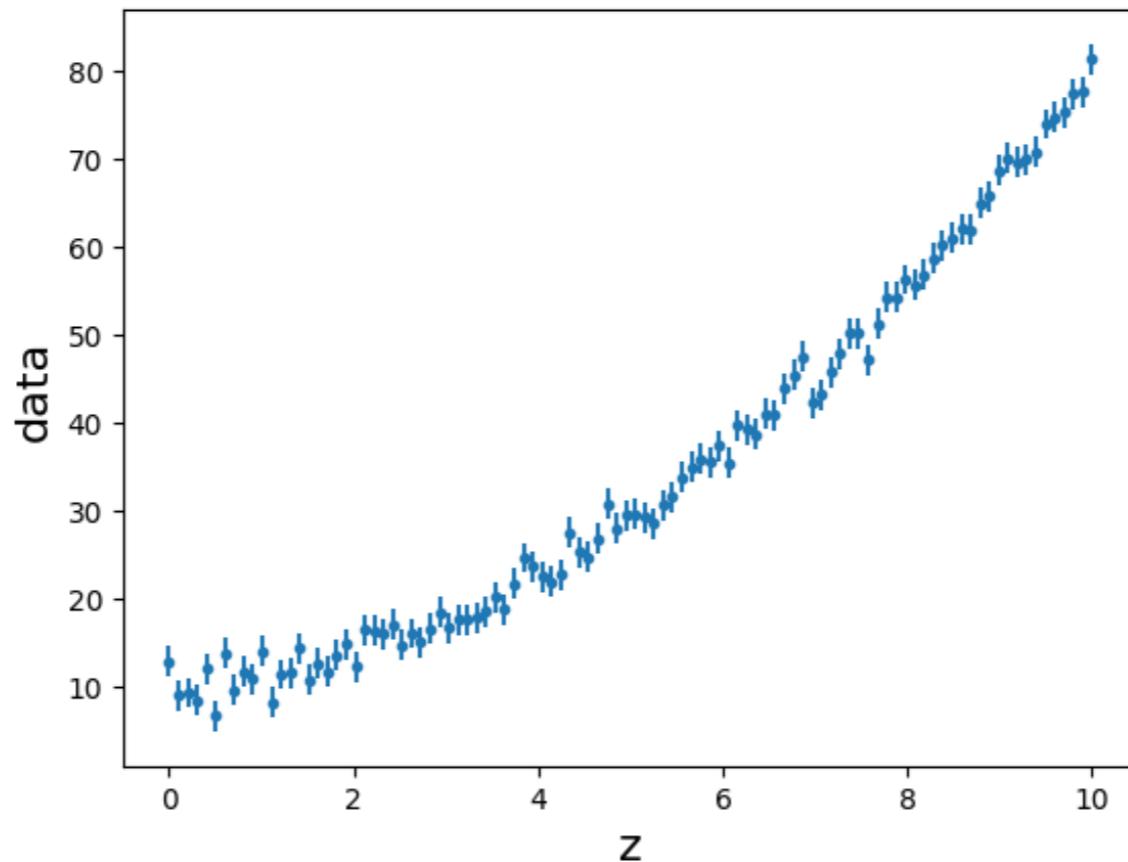
où \hat{f} est une réalisation du champ, et la moyenne d'ensemble

$$\left\langle \prod_{i=1}^n f(\mathbf{x}_i) \right\rangle = \int \mathcal{D}f \mathcal{P}[f] \prod_{i=1}^n f(\mathbf{x}_i) \quad \text{est proportionnelle à } 1/\sqrt{V}$$

→ Ce fait, appelé **principe Ergodique** assure que l'on peut estimer des fonctions de corrélations à partir d'une seule réalisation !

Estimation de paramètres

Exemple: un modèle physique polynomial



On suppose qu'on a N mesures d'un phénomène physique qui évolue en fonction du redshift et un bruit instrumental qui suit une loi normale $\mathcal{N}(0, \Sigma)$

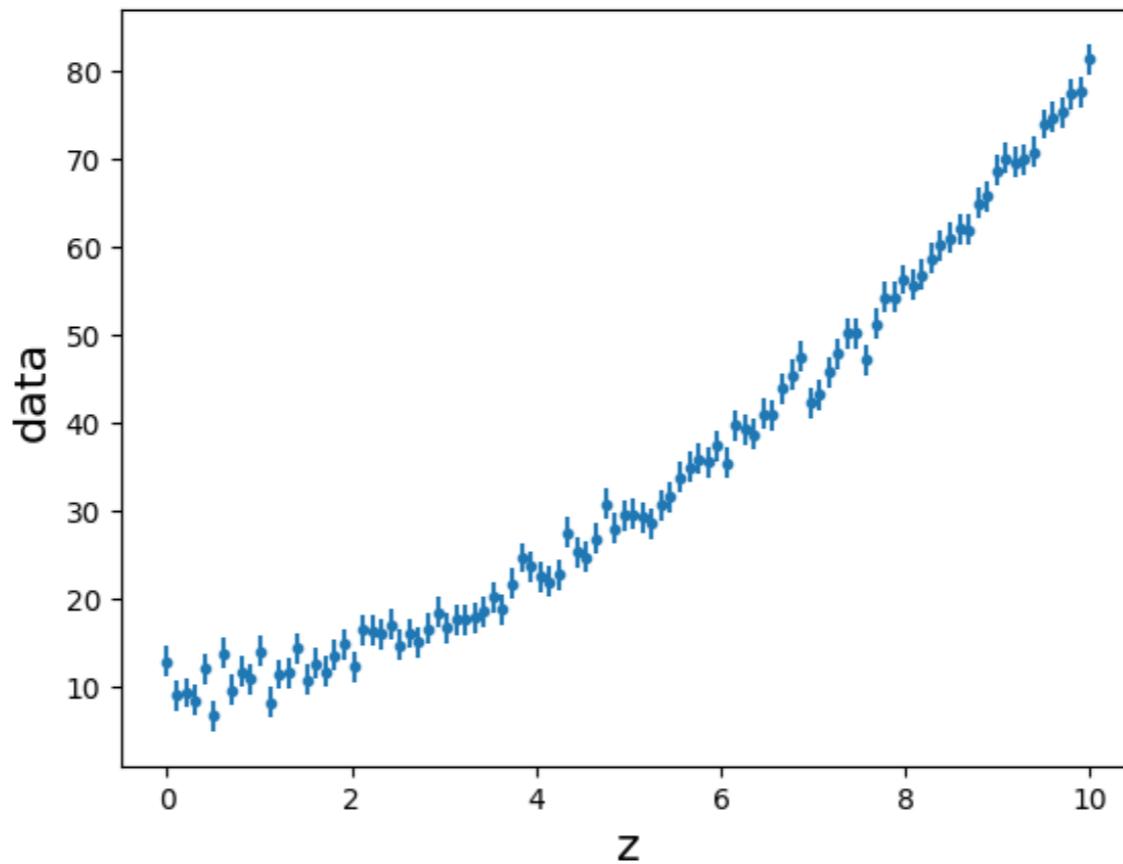
On veut décrire les données par une loi physique qui prend la forme d'un polynôme d'ordre m

$$d(z) = \sum_{k=0}^m a_k z^k + n(z)$$

et estimer les paramètres de cette loi

$$\{a_0, \dots, a_m\}$$

Exemple: un modèle physique polynomial



On a accès à un ensemble de mesures pour différent redshifts $d(z_i) = d_i$

Notre modèle pour ces données peut s'écrire

$$d_0 = a_0 + a_1 z_0 + a_2 z_0^2 + \dots + a_m z_0^m + n_0$$

.....

$$d_{N-1} = a_0 + a_1 z_{N-1} + a_2 z_{N-1}^2 + \dots + a_m z_{N-1}^m + n_{N-1}$$

Exemple: un modèle physique polynomial

$$d_0 = a_0 + a_1 z_0 + a_2 z_0^2 + \dots + a_m z_0^m + n_0$$

.....

$$d_{N-1} = a_0 + a_1 z_{N-1} + a_2 z_{N-1}^2 + \dots + a_m z_{N-1}^m + n_{N-1}$$

On peut l'écrire sous une forme matricielle

$$\mathcal{P} = \begin{pmatrix} 1 & z_0 & z_0^2 & \dots & z_0^m \\ 1 & z_1 & z_1^2 & \dots & z_1^m \\ \dots & \dots & \dots & \dots & \dots \\ 1 & z_{N-1} & z_{N-1}^2 & \dots & z_{N-1}^m \end{pmatrix} \quad \boldsymbol{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_d \end{pmatrix}$$

Paramètres
du modèle



$$\boldsymbol{d}_i = \mathcal{P}_{i\alpha} \boldsymbol{a}_\alpha + \boldsymbol{n}_i$$

Exemple: un modèle physique polynomial

$$\mathbf{d}_i = \mathcal{P}_{i\alpha} \mathbf{a}_\alpha + \mathbf{n}_i$$

Le bruit sur les données suit une loi normale de covariance Σ

On peut écrire la vraisemblance des données étant donné le modèle

$$\mathcal{L}(\mathbf{d}|\mathbf{a}) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) \right]$$

Exemple: un modèle physique polynomial

$$\mathbf{d}_i = \mathcal{P}_{i\alpha} \mathbf{a}_\alpha + \mathbf{n}_i$$

Le bruit sur les données suit une loi normale de covariance Σ

On peut écrire la vraisemblance des données étant donné le modèle

$$\begin{aligned}\mathcal{L}(\mathbf{d}|\mathbf{a}) &= \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) \right] \\ &= \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} \sum_{ij} (\mathbf{d}_i - \sum_\alpha \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\beta \mathcal{P}_{j\beta} \mathbf{a}_\beta) \right]\end{aligned}$$

Exemple: un modèle physique polynomial

$$\mathbf{d}_i = \mathcal{P}_{i\alpha} \mathbf{a}_\alpha + \mathbf{n}_i$$

Le bruit sur les données suit une loi normale de covariance Σ

On peut écrire la vraisemblance des données étant donné le modèle

$$\begin{aligned}\mathcal{L}(\mathbf{d}|\mathbf{a}) &= \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} (\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) \right] \\ &= \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp \left[-\frac{1}{2} \sum_{ij} (\mathbf{d}_i - \sum_\alpha \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\beta \mathcal{P}_{j\beta} \mathbf{a}_\beta) \right]\end{aligned}$$

Estimer les paramètres $\mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix}$ du modèle revient à maximiser la vraisemblance

$$\mathcal{L}(\mathbf{d}|\mathbf{a})$$

Exemple: un modèle physique polynomial

Maximiser la vraisemblance $\mathcal{L}(\mathbf{d}|\mathbf{a})$ **est équivalent à maximiser le log de la vraisemblance** $\ln\mathcal{L}(\mathbf{d}|\mathbf{a})$

$$\ln\mathcal{L}(\mathbf{d}|\mathbf{a}) = -\frac{1}{2}\ln(\det(2\pi\Sigma)) - \frac{1}{2} [(\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a})]$$

Exemple: un modèle physique polynomial

Maximiser la vraisemblance $\mathcal{L}(\mathbf{d}|\mathbf{a})$ **est équivalent à maximiser le log de la vraisemblance** $\ln\mathcal{L}(\mathbf{d}|\mathbf{a})$

$$\ln\mathcal{L}(\mathbf{d}|\mathbf{a}) = -\frac{1}{2}\ln(\det(2\pi\Sigma)) - \frac{1}{2} [(\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a})]$$

$$\frac{\partial \ln\mathcal{L}(\mathbf{d}|\mathbf{a})}{\partial a_\gamma} = -\frac{1}{2} \sum_{ij} \left[\left(\sum_\alpha \mathcal{P}_{i\alpha} \delta_{\alpha\gamma} \right) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \sum_\beta \mathcal{P}_{j\beta} \delta_{\beta\gamma} \right]$$

Exemple: un modèle physique polynomial

Maximiser la vraisemblance $\mathcal{L}(\mathbf{d}|\mathbf{a})$ **est équivalent à maximiser le log de la vraisemblance** $\ln\mathcal{L}(\mathbf{d}|\mathbf{a})$

$$\begin{aligned}\ln\mathcal{L}(\mathbf{d}|\mathbf{a}) &= -\frac{1}{2}\ln(\det(2\pi\Sigma)) - \frac{1}{2} [(\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a})] \\ \frac{\partial \ln\mathcal{L}(\mathbf{d}|\mathbf{a})}{\partial a_\gamma} &= -\frac{1}{2} \sum_{ij} \left[\left(\sum_\alpha \mathcal{P}_{i\alpha} \delta_{\alpha\gamma} \right) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \sum_\beta \mathcal{P}_{j\beta} \delta_{\beta\gamma} \right] \\ &= -\frac{1}{2} \sum_{ij} \left[(\mathcal{P}_{i\gamma}) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \mathcal{P}_{j\gamma} \right]\end{aligned}$$

Exemple: un modèle physique polynomial

Maximiser la vraisemblance $\mathcal{L}(\mathbf{d}|\mathbf{a})$ **est équivalent à maximiser le log de la vraisemblance** $\ln\mathcal{L}(\mathbf{d}|\mathbf{a})$

$$\ln\mathcal{L}(\mathbf{d}|\mathbf{a}) = -\frac{1}{2}\ln(\det(2\pi\Sigma)) - \frac{1}{2} [(\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a})]$$

$$\begin{aligned}\frac{\partial \ln\mathcal{L}(\mathbf{d}|\mathbf{a})}{\partial a_\gamma} &= -\frac{1}{2} \sum_{ij} \left[\left(\sum_\alpha \mathcal{P}_{i\alpha} \delta_{\alpha\gamma} \right) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \sum_\beta \mathcal{P}_{j\beta} \delta_{\beta\gamma} \right] \\ &= -\frac{1}{2} \sum_{ij} \left[(\mathcal{P}_{i\gamma}) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \mathcal{P}_{j\gamma} \right] \\ &= -\sum_{ij} (\mathcal{P}_{i\gamma}) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta)\end{aligned}$$

Exemple: un modèle physique polynomial

Maximiser la vraisemblance $\mathcal{L}(\mathbf{d}|\mathbf{a})$ **est équivalent à maximiser le log de la vraisemblance** $\ln\mathcal{L}(\mathbf{d}|\mathbf{a})$

$$\begin{aligned}\ln\mathcal{L}(\mathbf{d}|\mathbf{a}) &= -\frac{1}{2}\ln(\det(2\pi\Sigma)) - \frac{1}{2} [(\mathbf{d} - \mathcal{P}\mathbf{a})^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a})] \\ \frac{\partial \ln\mathcal{L}(\mathbf{d}|\mathbf{a})}{\partial a_\gamma} &= -\frac{1}{2} \sum_{ij} \left[\left(\sum_\alpha \mathcal{P}_{i\alpha} \delta_{\alpha\gamma} \right) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \sum_\beta \mathcal{P}_{j\beta} \delta_{\beta\gamma} \right] \\ &= -\frac{1}{2} \sum_{ij} \left[(\mathcal{P}_{i\gamma}) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta) + (\mathbf{d}_i - \sum_\beta \mathcal{P}_{i\alpha} \mathbf{a}_\alpha) (\Sigma^{-1})_{ij} \mathcal{P}_{j\gamma} \right] \\ &= -\sum_{ij} (\mathcal{P}_{i\gamma}) (\Sigma^{-1})_{ij} (\mathbf{d}_j - \sum_\alpha \mathcal{P}_{j\beta} \mathbf{a}_\beta)\end{aligned}$$

La vraisemblance sera maximisée si

$$\nabla_{\mathbf{a}} \ln\mathcal{L}(\mathbf{d}|\mathbf{a}) = -\mathcal{P}^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) = 0$$

Exemple: un modèle physique polynomial

$$\nabla_a \ln \mathcal{L}(\mathbf{d}|\mathbf{a}) = -\mathcal{P}^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) = 0$$

La solution de ce système est donnée par l'estimateur

$$\hat{\mathbf{a}} = (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \mathbf{d}$$

L'estimateur de \mathbf{a} est non biaisé

$$\langle \hat{\mathbf{a}} \rangle = (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \langle \mathbf{d} \rangle = (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \langle \mathcal{P}\mathbf{a} + \mathbf{n} \rangle = \mathbf{a}$$

$$(\mathbf{d} = \mathcal{P}\mathbf{a} + \mathbf{n})$$

Cela reste vrai même si notre modèle pour les erreurs sur les données Σ est faux !

Exemple: un modèle physique polynomial

$$\nabla_a \ln \mathcal{L}(\mathbf{d}|\mathbf{a}) = -\mathcal{P}^T \Sigma^{-1} (\mathbf{d} - \mathcal{P}\mathbf{a}) = 0$$

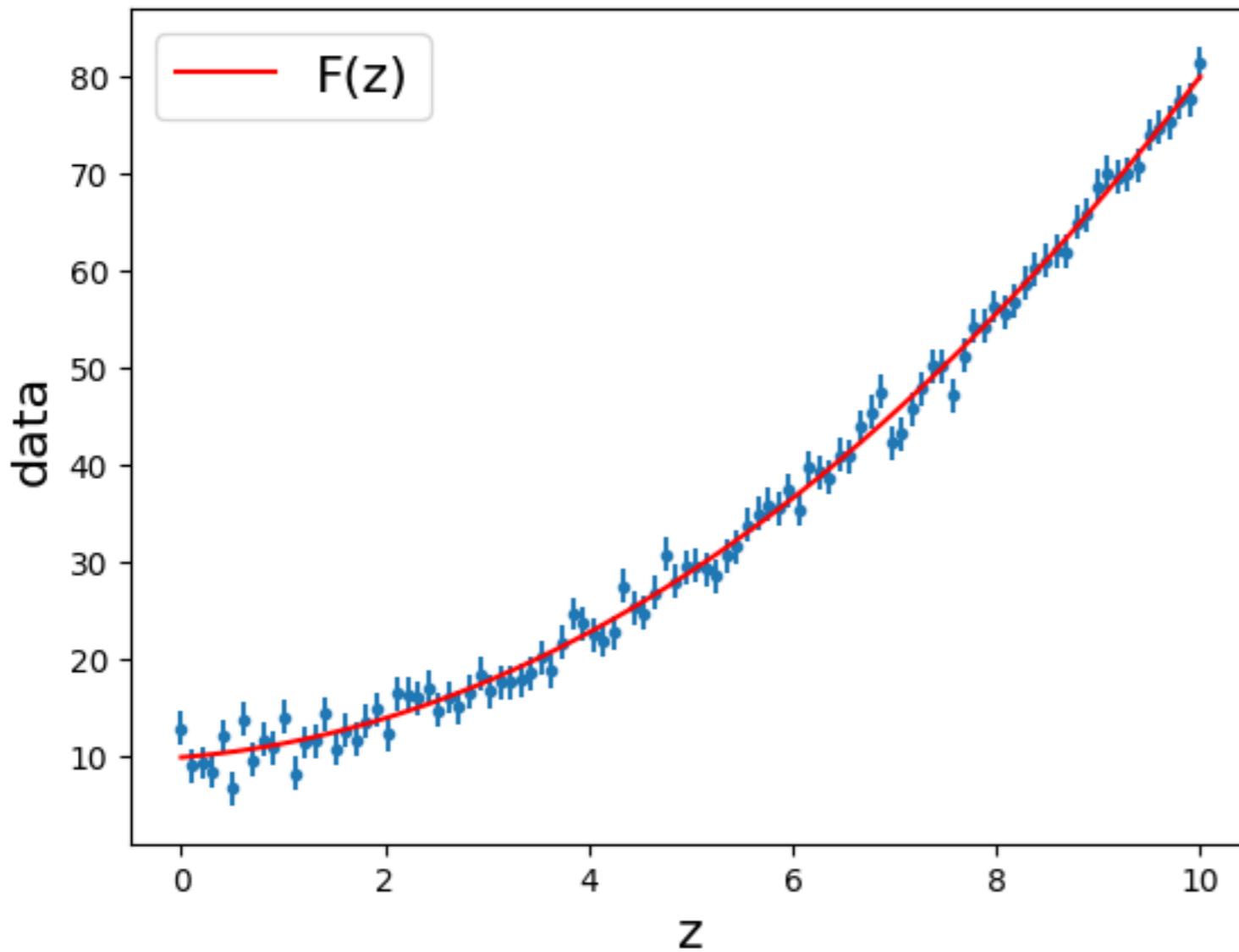
La solution de ce système est donnée par l'estimateur

$$\hat{\mathbf{a}} = (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \mathbf{d}$$

De la même façon on peut calculer la covariance de \mathbf{a}

$$\begin{aligned}\langle \hat{\mathbf{a}} \hat{\mathbf{a}}^T \rangle - \mathbf{a} \mathbf{a}^T &= (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \langle \mathbf{d} \mathbf{d}^T \rangle \Sigma^{-1} \mathcal{P} (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} - \mathbf{a} \mathbf{a}^T \\ &= (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \mathcal{P}^T \Sigma^{-1} \langle (\mathcal{P}\mathbf{a} + \mathbf{n})(\mathcal{P}\mathbf{a} + \mathbf{n})^T \rangle \Sigma^{-1} \mathcal{P} (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} - \mathbf{a} \mathbf{a}^T \\ &= (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1} \\ \text{Cov}(\hat{\mathbf{a}}) &= (\mathcal{P}^T \Sigma^{-1} \mathcal{P})^{-1}\end{aligned}$$

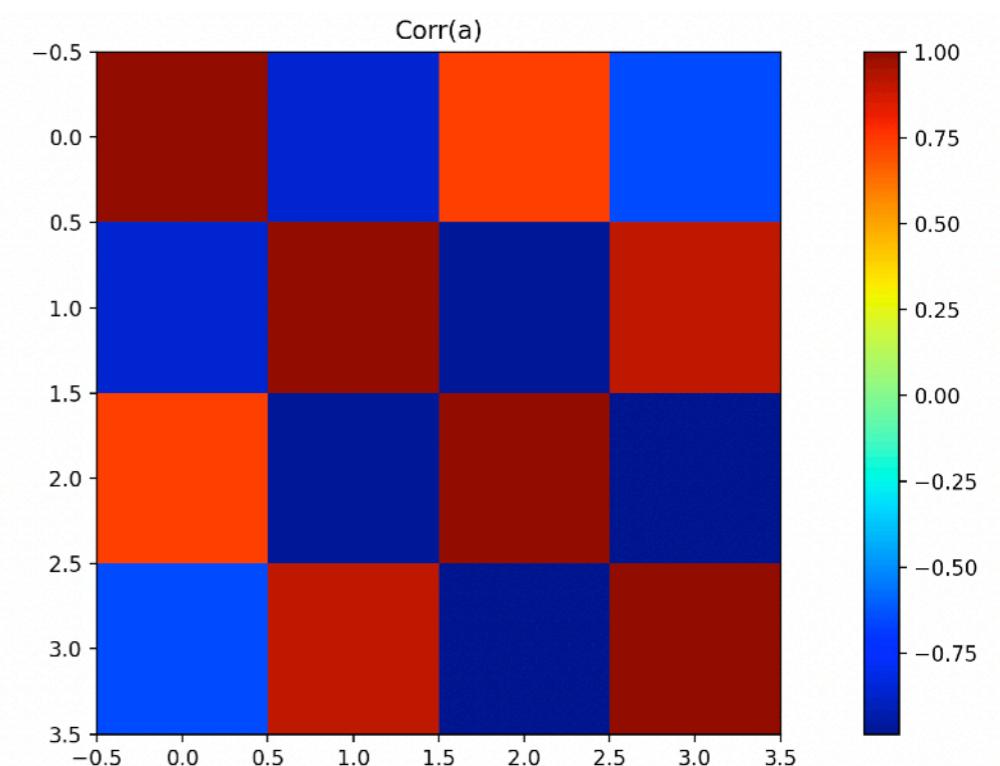
Exemple: un modèle physique polynomial



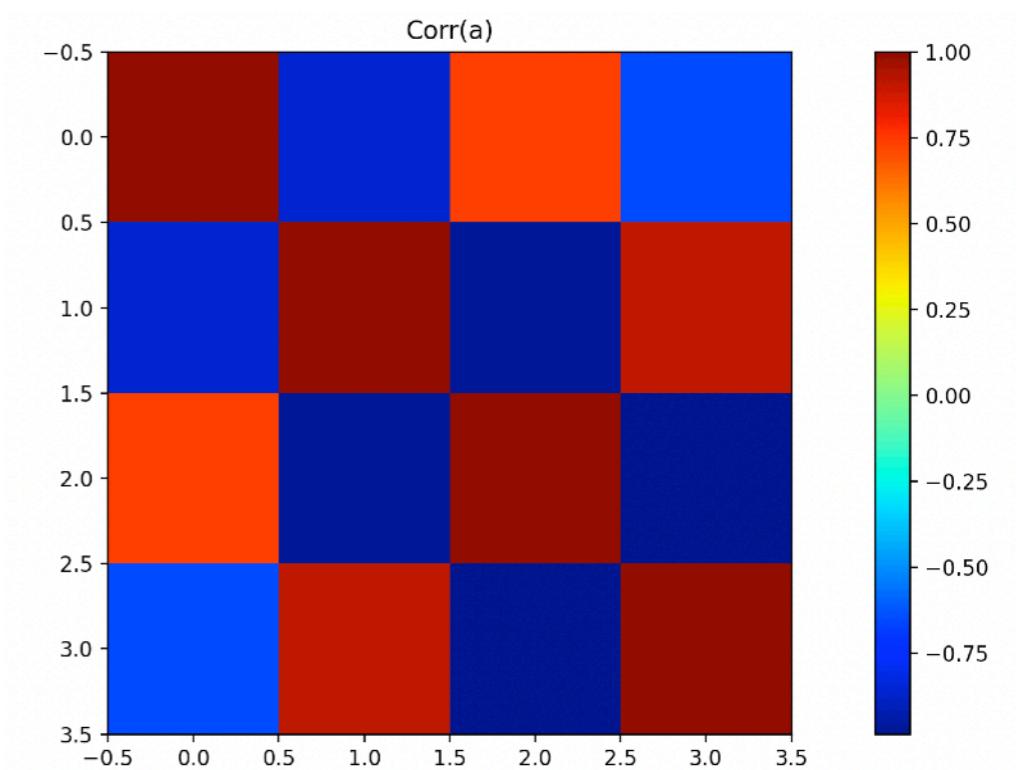
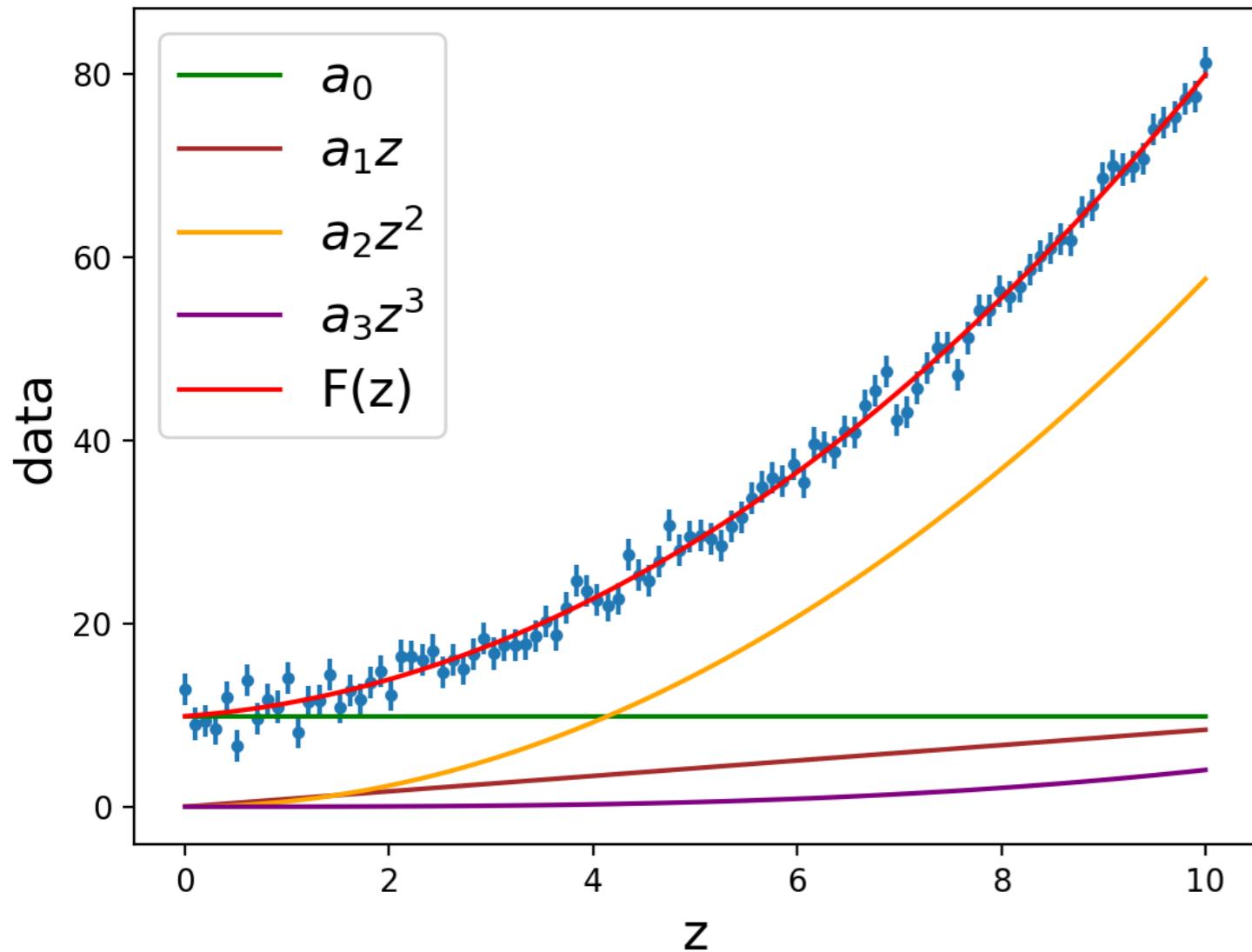
$$F(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3$$

```
param = [10, 1, 0.5, 0.01]
```

$$\begin{aligned}a_0 &= 9.87 \pm 0.67 \\a_1 &= 0.84 \pm 0.58 \\a_2 &= 0.58 \pm 0.13 \\a_3 &= 0.004 \pm 0.0088\end{aligned}$$

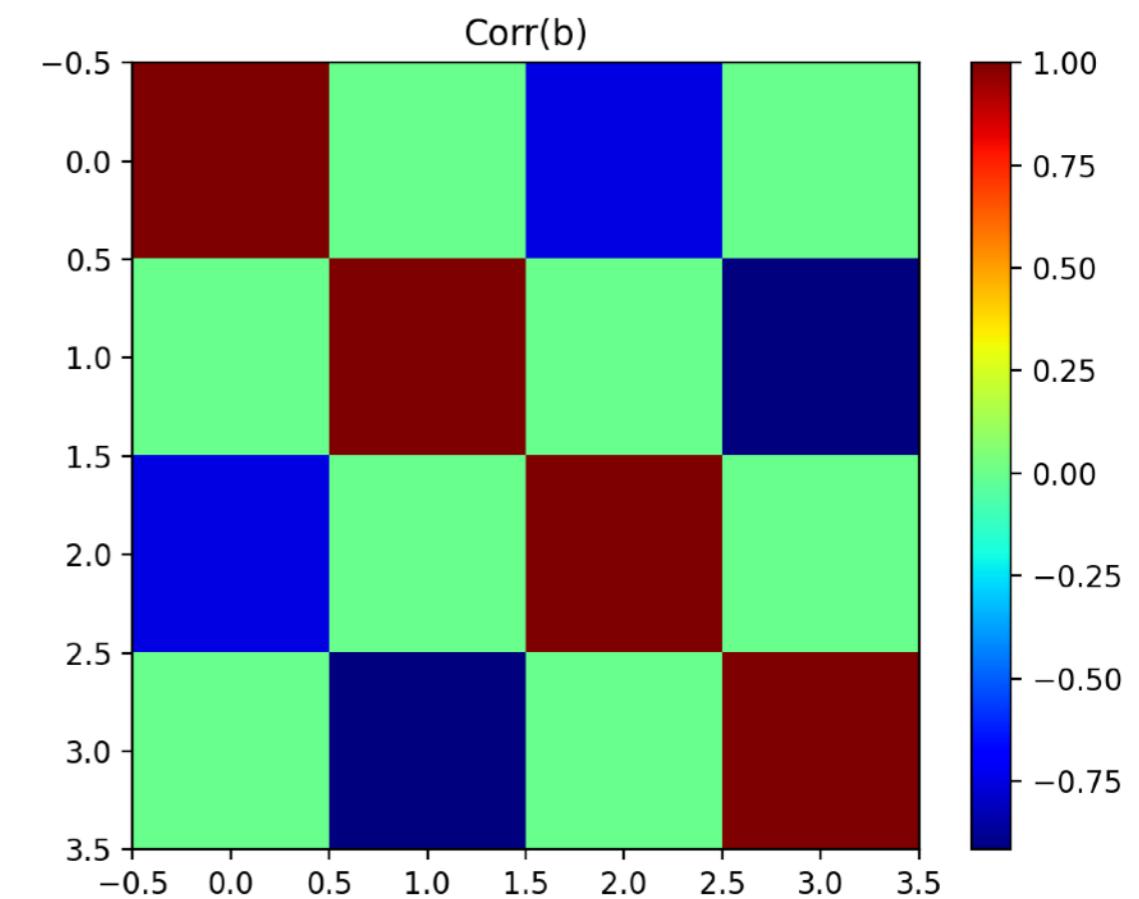
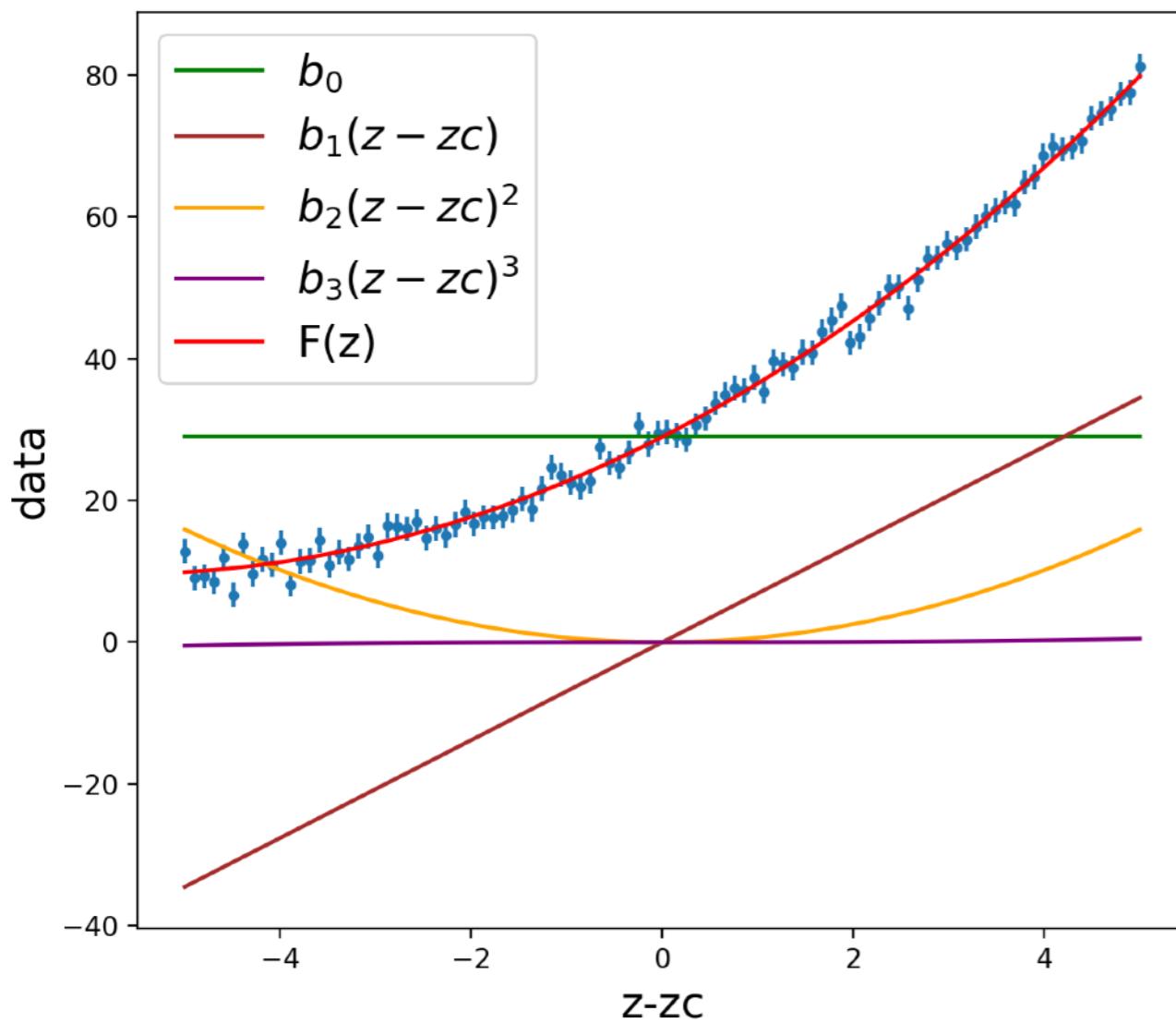


Exemple: un modèle physique polynomial

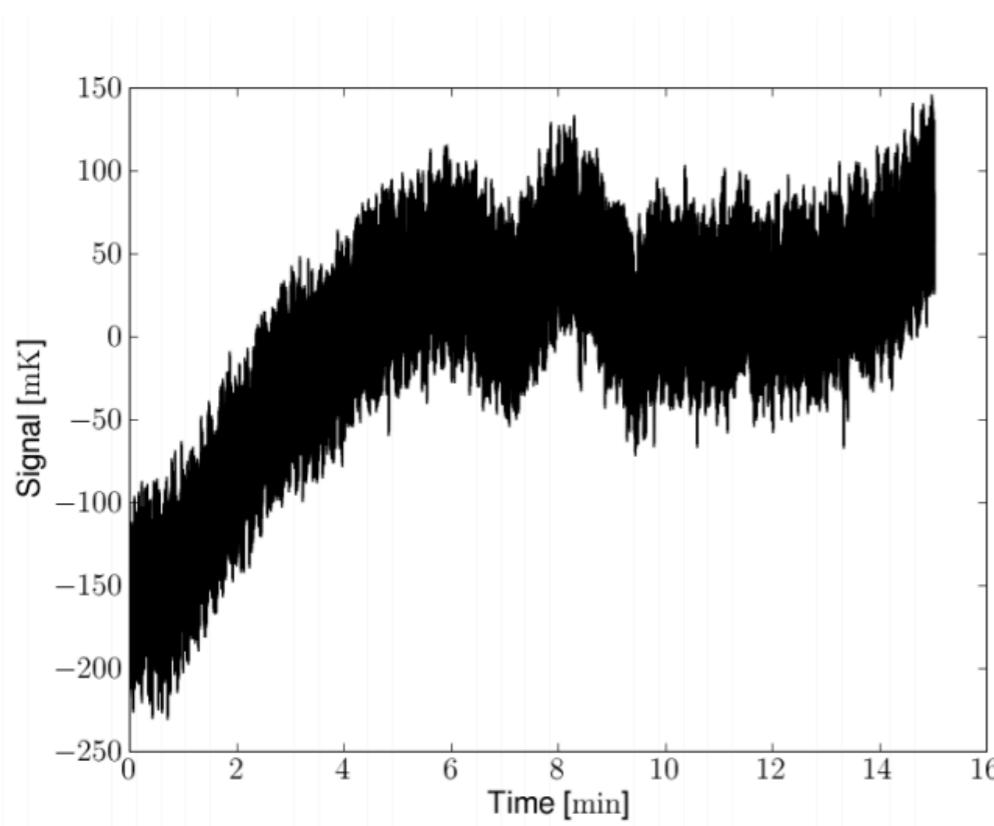


On peut réduire les corrélations entre paramètres en changeant la paramétrisation

$$F(z) = b_0 + b_1(z - z_c) + b_2(z - z_c)^2 + b_3(z - z_c)^3$$



Interlude: l'équation du map making

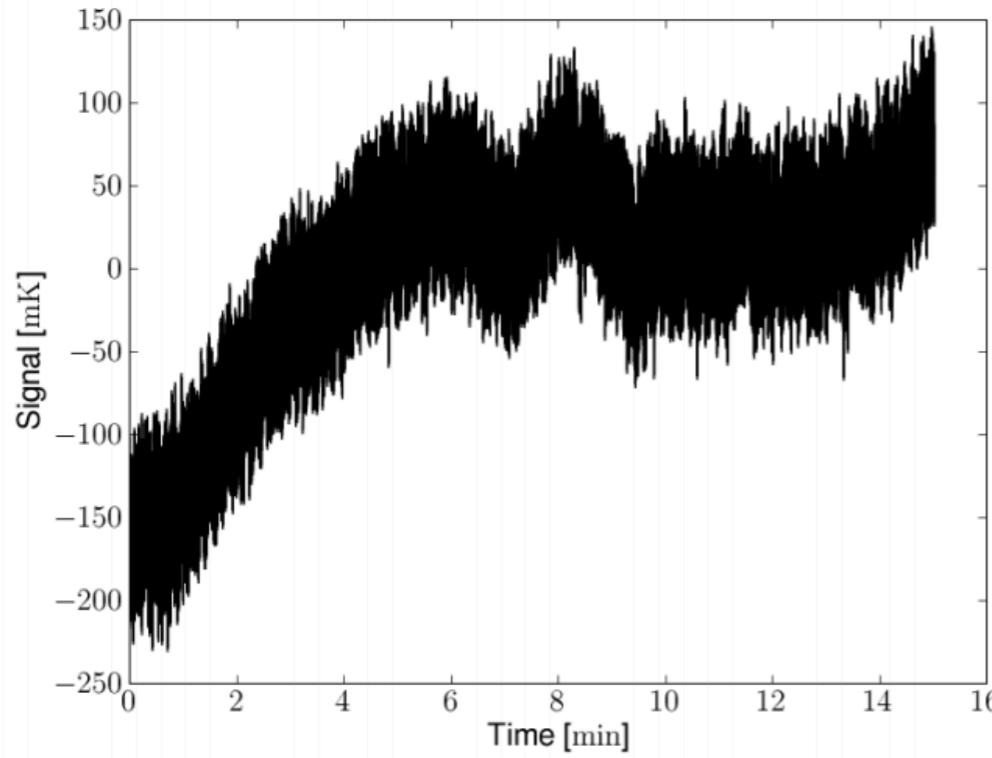


Time sample

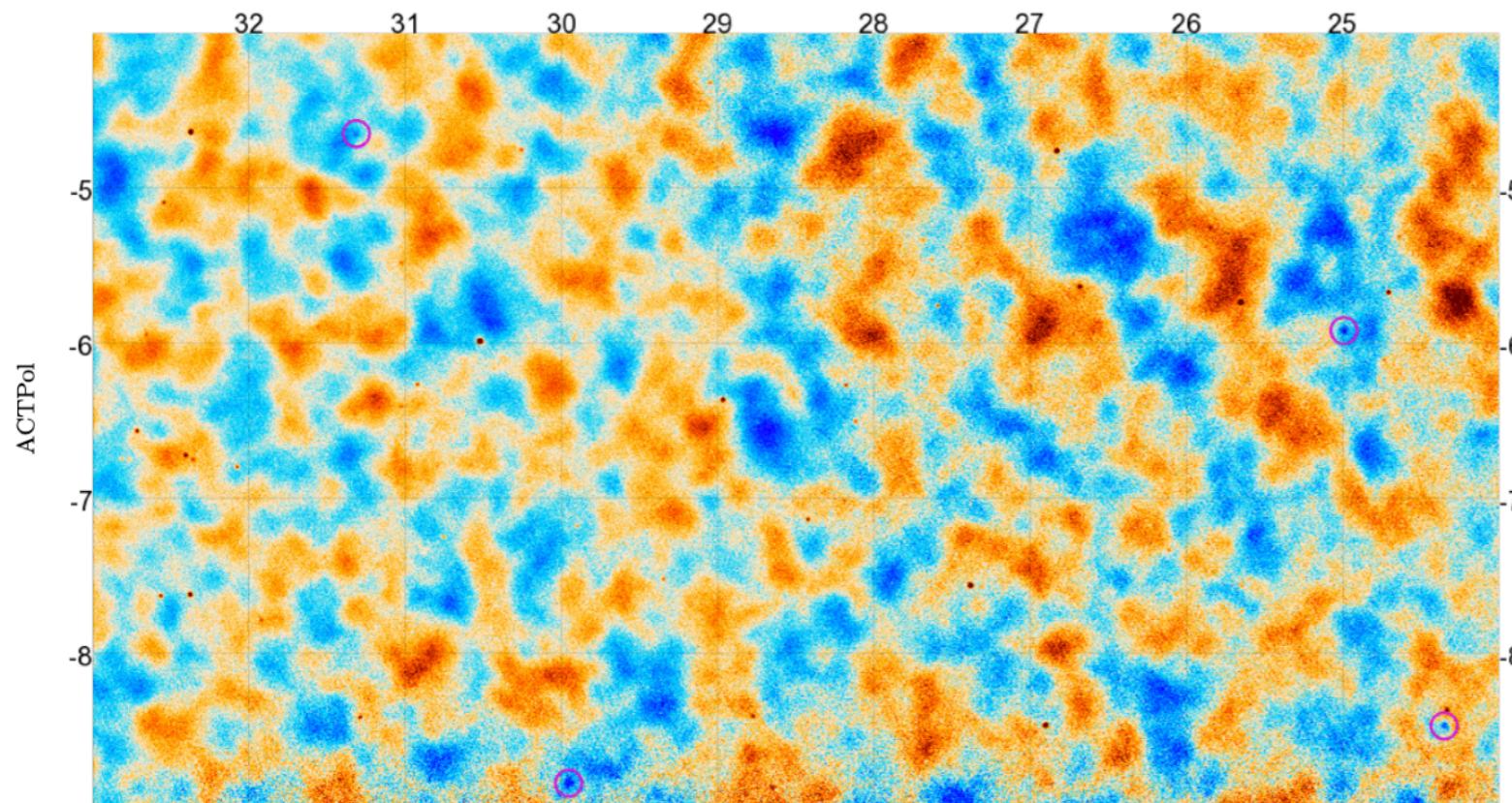
Valeur de la carte dans le pixel α

$$d_t^{\text{TOD}} = \mathcal{A}_{t,\alpha} m_\alpha + n_t$$
$$\mathbf{m} = (\mathcal{A}^T \Sigma^{-1} \mathcal{A})^{-1} \mathcal{A}^T \Sigma^{-1} \mathbf{d}^{\text{TOD}}$$

Interlude: l'équation du map making



$$\begin{aligned} d_t^{\text{TOD}} &= \mathcal{A}_{t,\alpha} m_\alpha + n_t \\ m &= (\mathcal{A}^T \Sigma^{-1} \mathcal{A})^{-1} \mathcal{A}^T \Sigma^{-1} d^{\text{TOD}} \end{aligned}$$



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

Maximisation du postérieur par la méthode de Monte Carlo par chaînes de Markov

Une méthode de Monte-Carlo désigne toute méthode visant à calculer une valeur numérique en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes.

Un processus de Markov est un processus stochastique possédant la propriété de Markov : l'information utile pour la prédiction du futur est entièrement contenue dans l'état présent du processus et n'est pas dépendante des états antérieurs (le système n'a pas de « mémoire »)

Maximisation du postérieur par la méthode de Monte Carlo par chaînes de Markov

Le but est d'obtenir un échantillonnage aléatoire d'une distribution de probabilité quant l'échantillonnage direct est difficile.

$$P(\boldsymbol{\theta}|\text{data}) = \frac{P(\text{data}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\text{data})} = \boxed{\frac{P(\text{data}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}'} P(\text{data}|\boldsymbol{\theta}')P(\boldsymbol{\theta}')d\boldsymbol{\theta}'}}$$



Intégrale à N dimension

Algorithme de Metropolis Hasting

On définit $f(\theta) = P(\text{data}|\theta)P(\theta)$

- 1) Initialisation: choisir un point arbitraire θ_0 pour être le premier échantillon
- 2) Choisir une probabilité de transition (proposal) $g(\theta'|\theta_i)$ un choix usuel est une distribution Gaussienne centrée sur θ_i et de covariance $C_p = \frac{2.4^2}{D}C_\theta$
- 3) A chaque itération i , tirer aléatoirement un candidat θ' selon le proposal.
- 4) Calculer le taux d'acceptation $\alpha = f(\theta')/f(\theta_i)$
- 5) Tirer un nombre aléatoire uniforme u entre $[0,1]$, si

$u \leq \alpha$ Accepter le candidat $\theta_{i+1} = \theta'$

$u > \alpha$ Rejeter le candidat $\theta_{i+1} = \theta_i$

Algorithme de Metropolis Hasting

On se déplace dans l'espace des paramètres, si on cherche à se déplacer vers un état plus probable que l'état actuel le déplacement est toujours accepté.

Si on se déplace vers un état moins probable, le déplacement peut être accepté ou rejeté, le taux de rejet est d'autant plus probable que la chute de densité de probabilité est élevée.

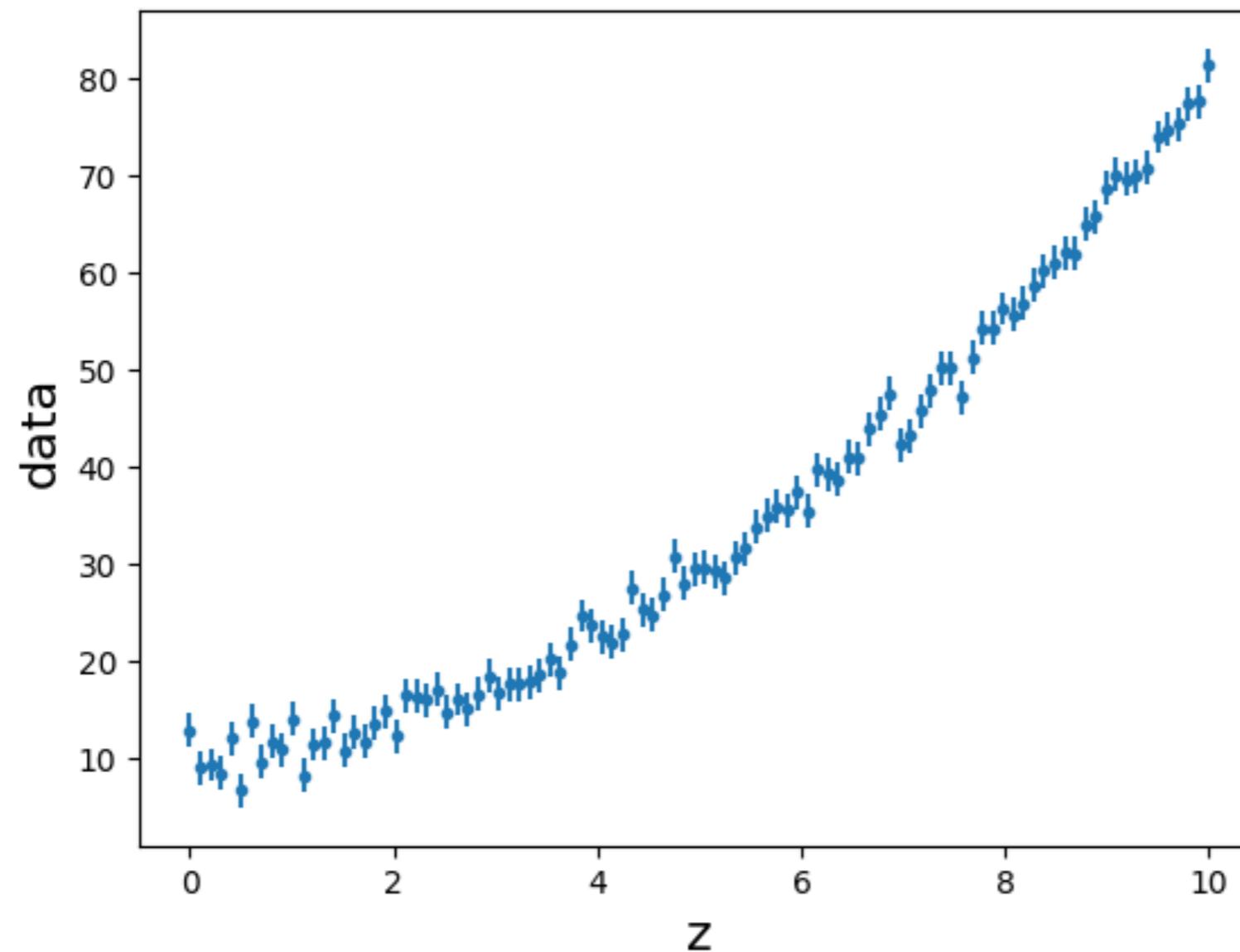
La marche aléatoire tend à visiter préférentiellement les régions de l'espace des paramètres où le postérieur est élevé, mais visite occasionnellement des régions de moindre densité ce qui lui évite de rester piégée dans des maxima locaux.

La méthode permet d'échantillonner efficacement le postérieur sans jamais avoir à calculer des intégrales du type

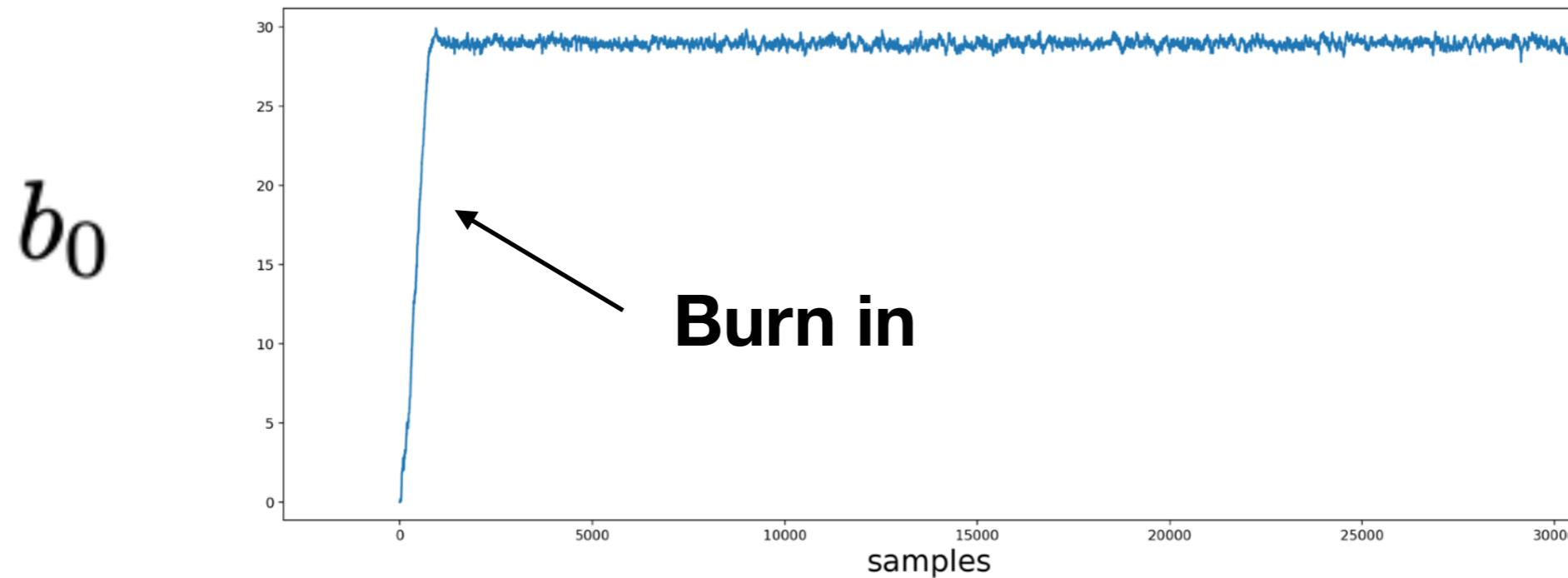
$$\int_{\theta'} P(\text{data}|\theta') P(\theta') d\theta'$$

Retournons à notre exemple

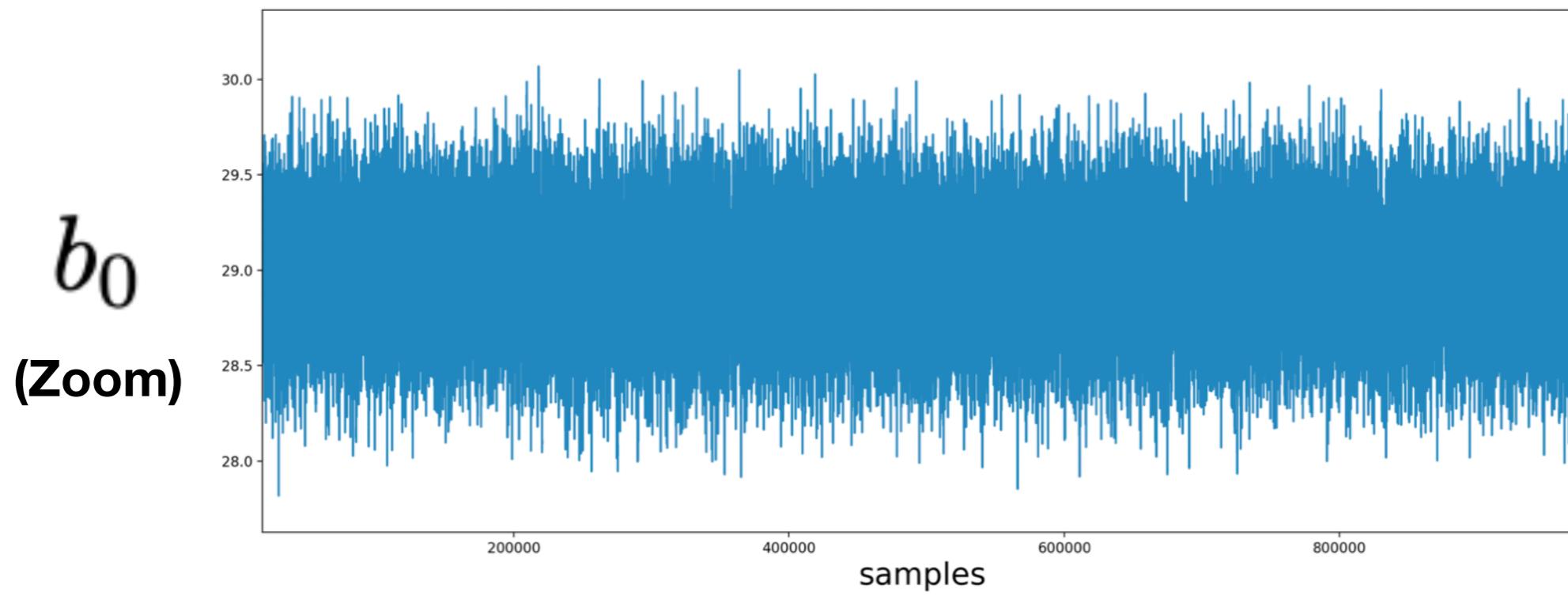
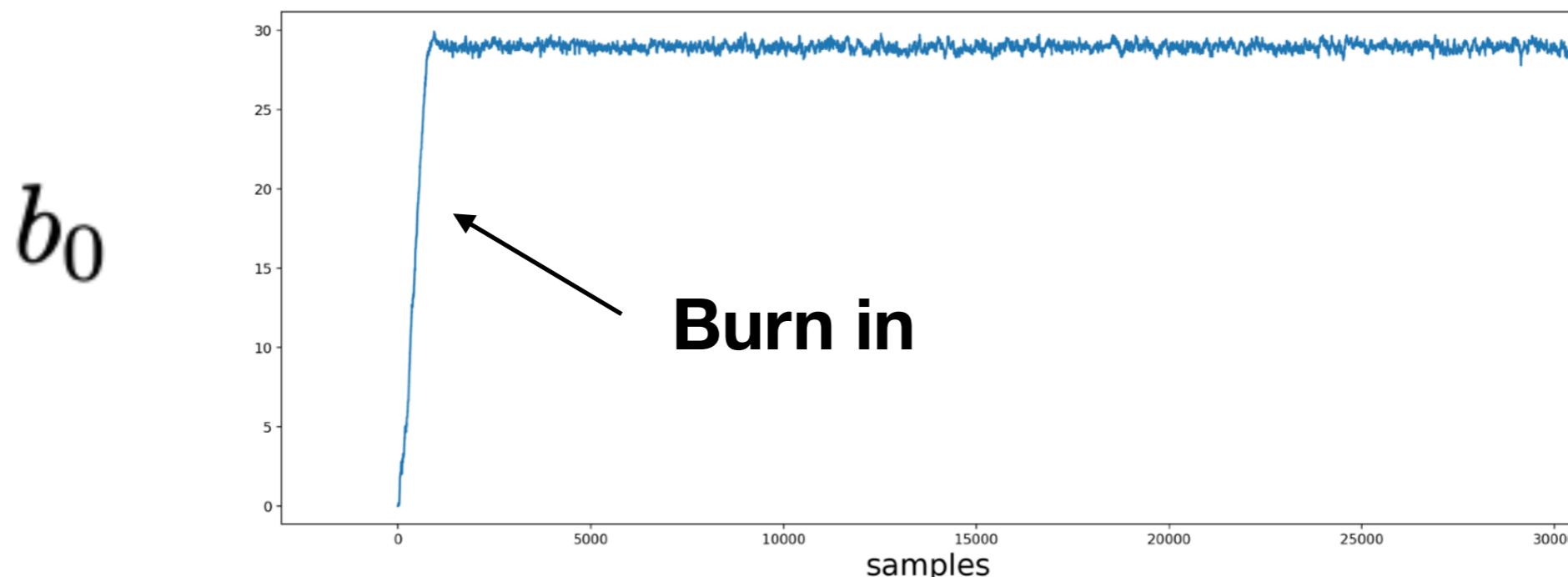
$$F(z) = b_0 + b_1(z - z_c) + b_2(z - z_c)^2 + b_3(z - z_c)^3$$

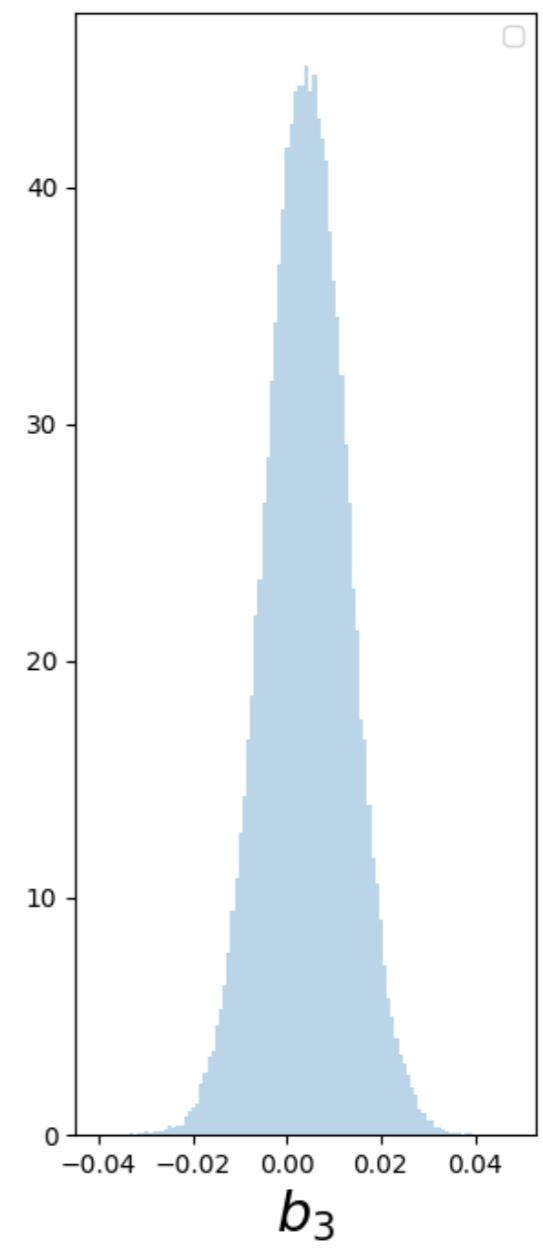
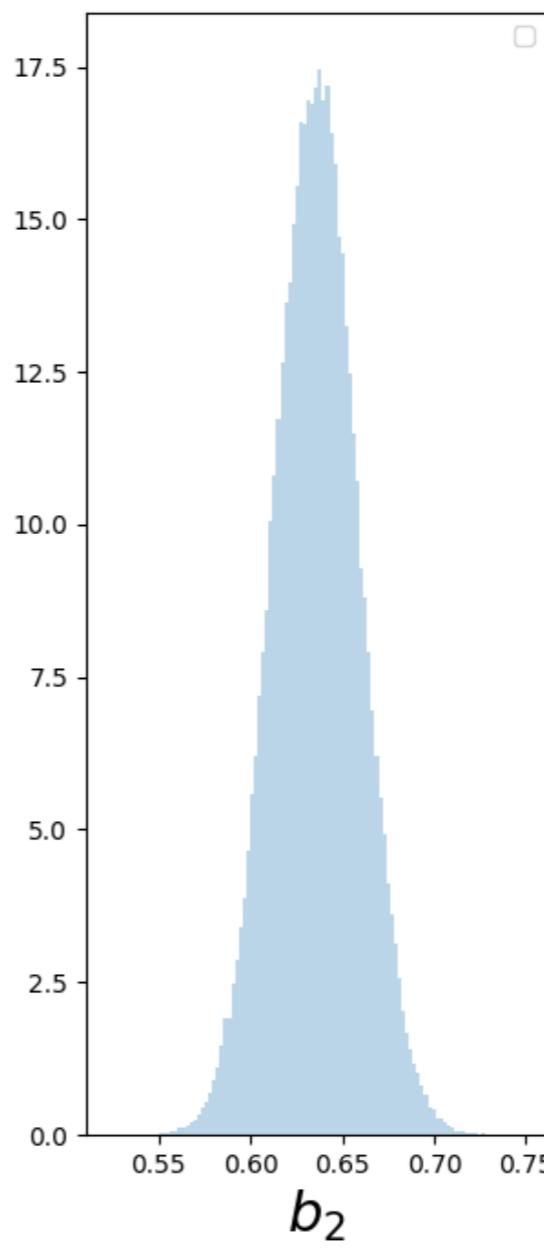
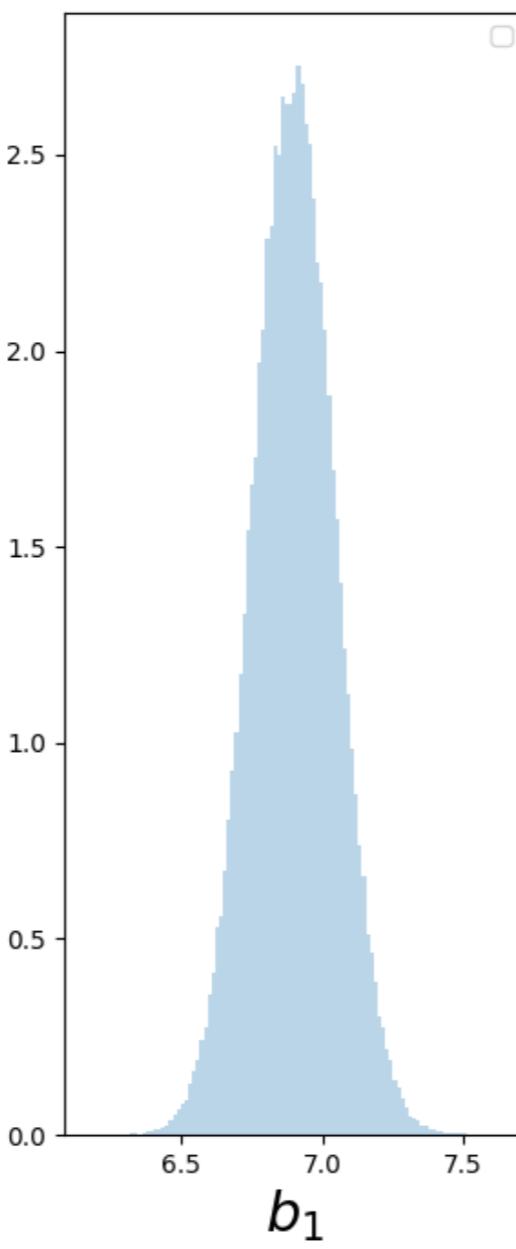
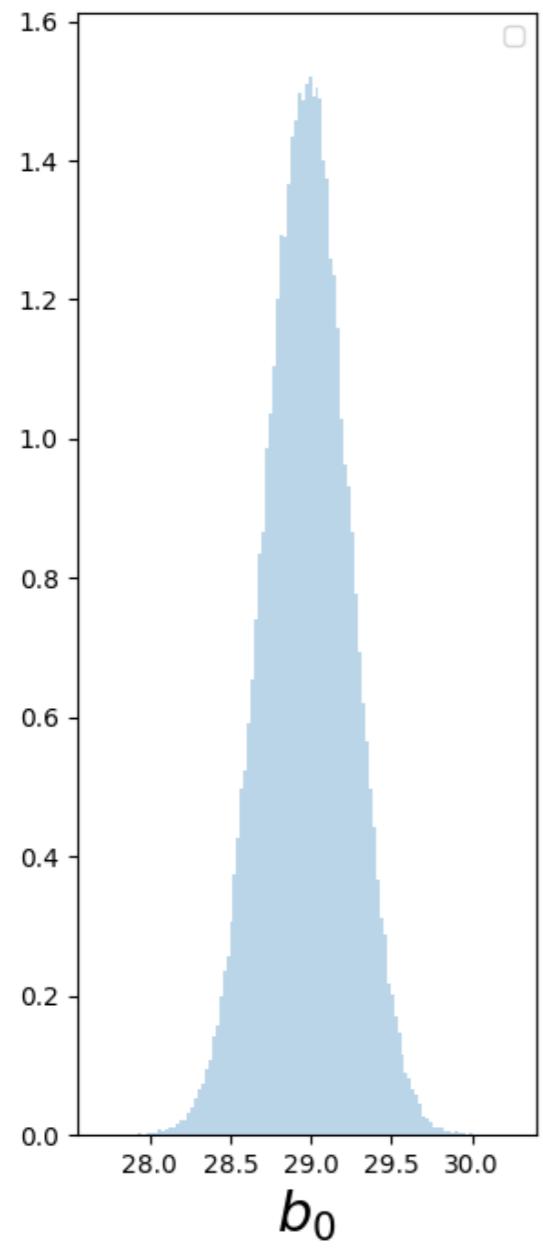


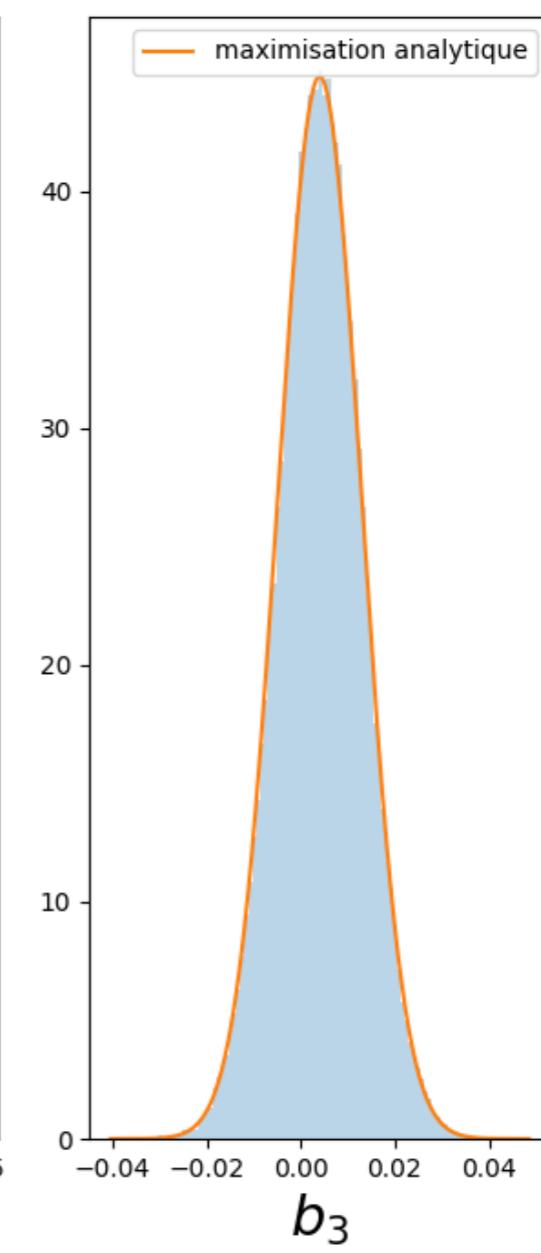
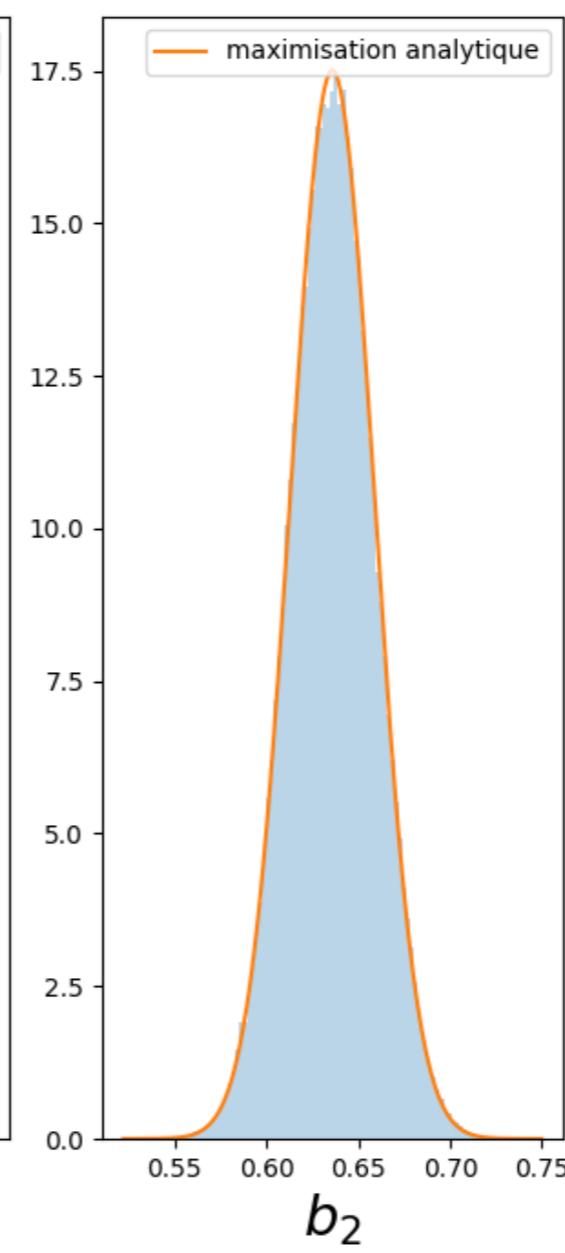
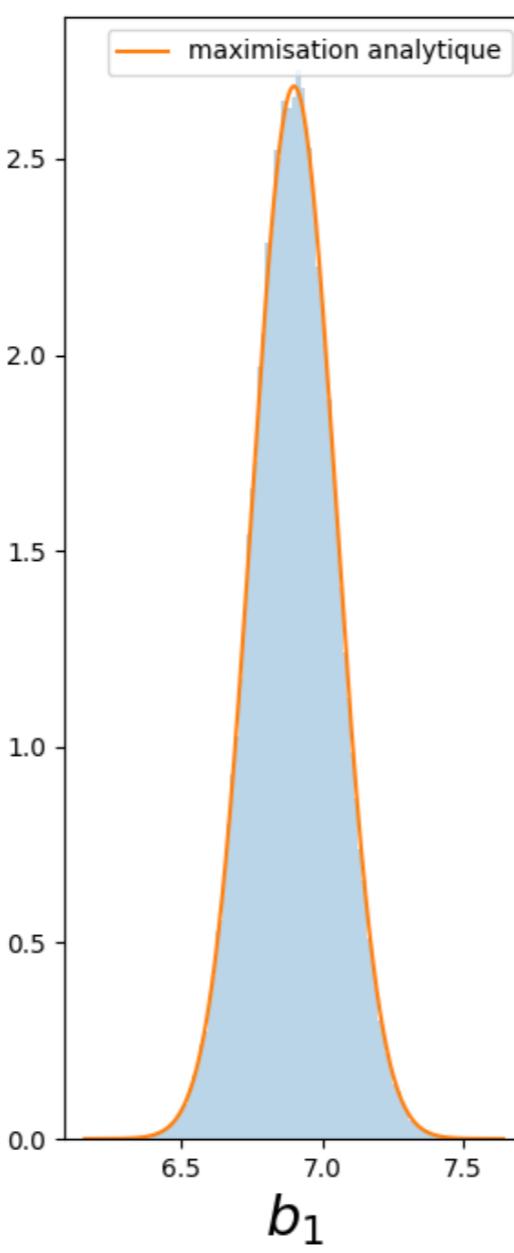
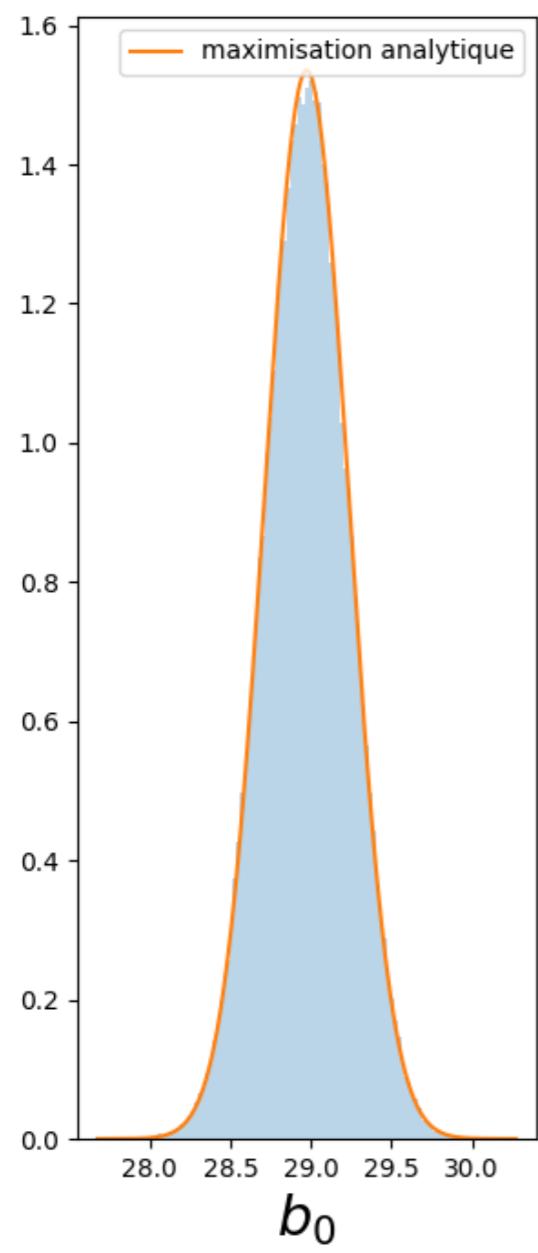
On plot la chaîne qui correspond à b_0

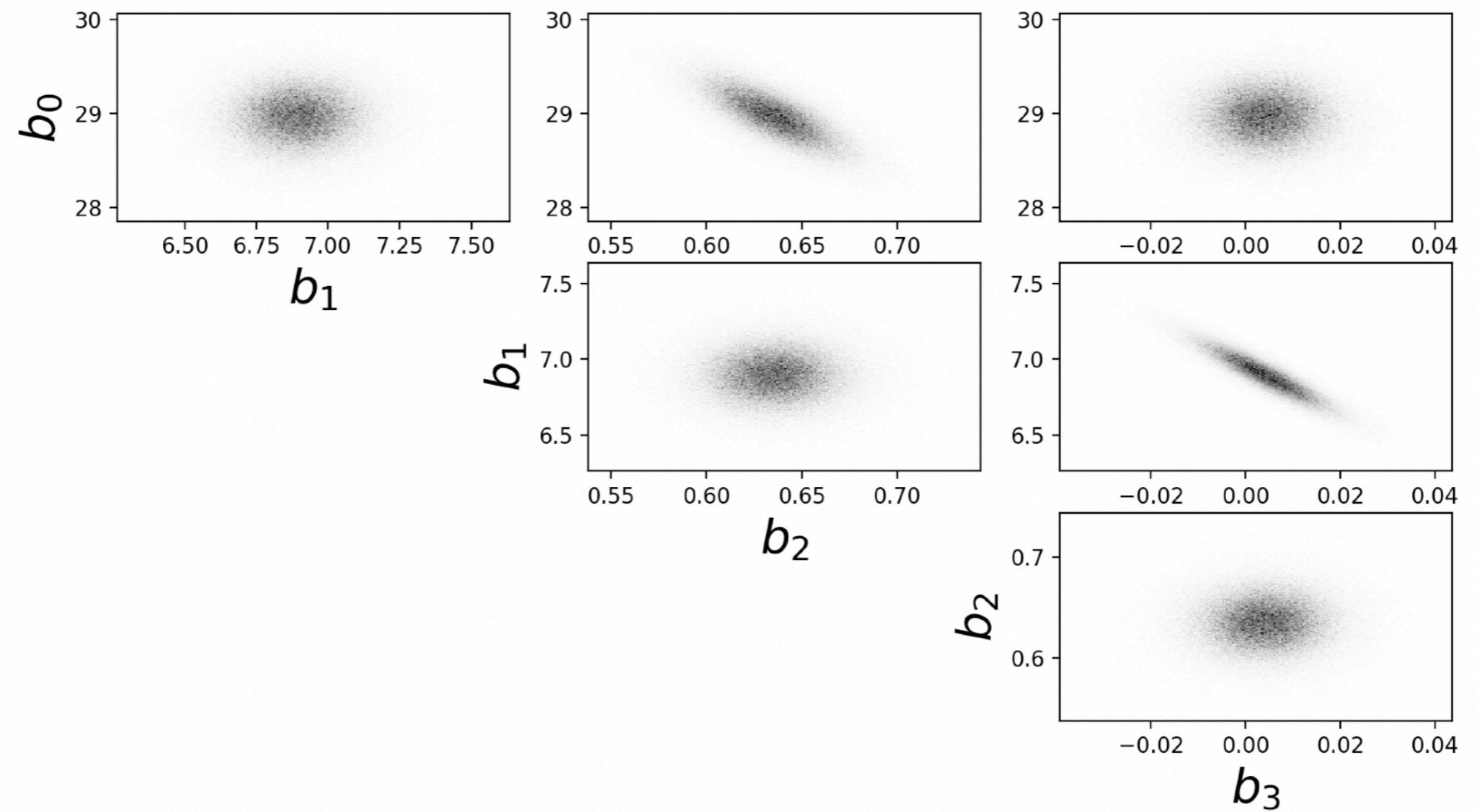


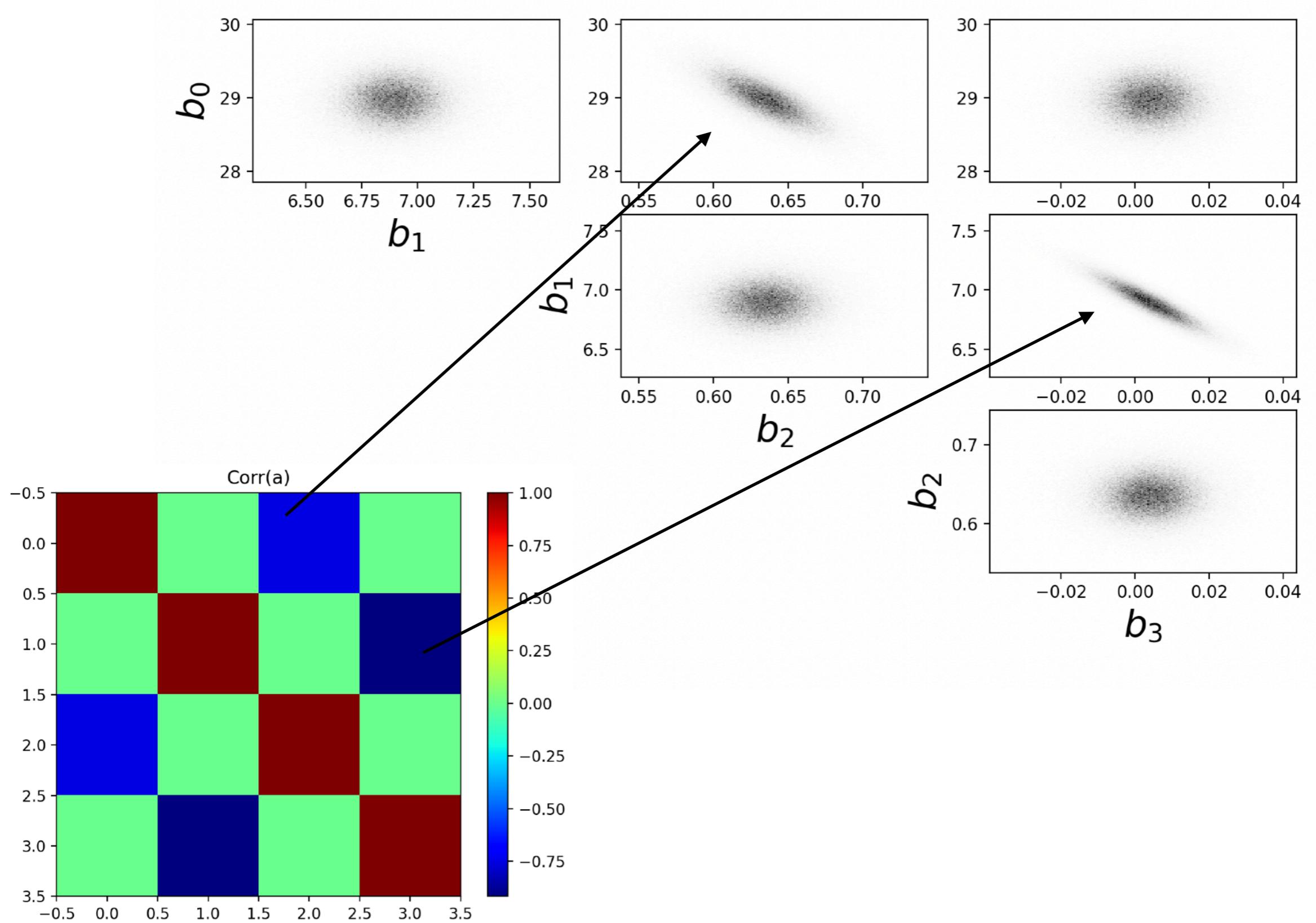
On plot la chaîne qui correspond à b_0











Un diagnostic de la convergence: Gelman-Rubin statistic

Faisons tourner J chaines en parallèle, et dénotons les différents échantillons de la chaîne j (après burn in) par

$$x_1^j, x_2^j, \dots, x_N^j$$

Pour chacune des chaines on peut calculer la valeur moyenne

$$\langle x^j \rangle = \frac{1}{N} \sum_{k=1}^N x_k^j$$

On peut aussi calculer la valeur moyenne de l'ensemble des échantillons de toutes les chaines

$$\langle x \rangle = \frac{1}{J} \sum_{j=1}^J \langle x^j \rangle$$

Un diagnostic de la convergence: Gelman-Rubin statistic

Faisons tourner J chaines en parallèle, et dénotons les différents échantillons de la chaîne j (après burn in) par

$$x_1^j, x_2^j, \dots, x_N^j$$

Pour chacune des chaines on peut calculer la valeur moyenne

$$\langle x^j \rangle = \frac{1}{N} \sum_{k=1}^N x_k^j$$

On peut aussi calculer la valeur moyenne de l'ensemble des échantillons de toutes les chaines

$$\langle x \rangle = \frac{1}{J} \sum_{j=1}^J \langle x^j \rangle$$

On estime 2 variances :

1) la variance de la moyenne de chacune des chaines

$$\frac{B}{N} = \frac{1}{J-1} \sum_{j=1}^J (\langle x^j \rangle - \langle x \rangle)^2$$

2) la moyenne de la variance dans chaque chaîne

$$W = \frac{1}{J} \sum_{j=1}^J \frac{1}{N-1} \sum_{k=1}^N (x_k^j - \langle x^j \rangle)^2$$

Un diagnostic de la convergence: Gelman-Rubin statistic

Faisons tourner J chaines en parallèle, et dénotons les différents échantillons de la chaîne j (après burn in) par

$$x_1^j, x_2^j, \dots, x_N^j$$

Pour chacune des chaines on peut calculer la valeur moyenne

$$\langle x^j \rangle = \frac{1}{N} \sum_{k=1}^N x_k^j$$

On peut aussi calculer la valeur moyenne de l'ensemble des échantillons de toutes les chaines

$$\langle x \rangle = \frac{1}{J} \sum_{j=1}^J \langle x^j \rangle$$

On estime 2 variances :

1) la variance de la moyenne de chacune des chaines

$$\frac{B}{N} = \frac{1}{J-1} \sum_{j=1}^J (\langle x^j \rangle - \langle x \rangle)^2$$

2) la moyenne de la variance dans chaque chaîne

$$W = \frac{1}{J} \sum_{j=1}^J \frac{1}{N-1} \sum_{k=1}^N (x_k^j - \langle x^j \rangle)^2$$

UNE statistique de Gelman-Rubin est donnée par

$$R = \frac{\frac{N-1}{N} W + \frac{B}{N}}{W}$$

Un diagnostic de la convergence: Gelman-Rubin statistic

Faisons tourner J chaines en parallèle, et dénotons les différents échantillons de la chaîne j (après burn in) par

$$x_1^j, x_2^j, \dots, x_N^j$$

WARNING

En préparant ce cours je suis tombé sur une dizaine de définitions différentes pour R, par exemple

$$R = \sqrt{\frac{\frac{N-1}{N}W + \frac{B}{N}}{W}} \quad \sqrt{R} = \sqrt{\frac{\frac{N-1}{N}W + \frac{B}{N} + \frac{B}{JN}}{W} \frac{df + 3}{df + 1}}$$

- Cholesky decompose the mean covariance $\mathbf{C}_x = \mathbf{L}\mathbf{L}^T$ to orthonormalize the parameters by forming $\mathbf{L}^{-1}\mathbf{x}$.
- Calculate the eigenvalues D_i of the between-chain covariance of the orthonormalized parameter means using $\mathbf{L}^{-1}\mathbf{C}_x[\mathbf{L}^{-1}]^T = \mathbf{U}\mathbf{D}\mathbf{U}^T$.
- Define $R-1 \equiv \max(D_i)$ to measure the largest variance between chains of any of the orthonormalized parameter means.

**mais elles sont toute une indication de la convergence de la chaîne,
Quant N est grand et quant B → 0, R → 1**

1) la variance de la moyenne de chaque des chaînes

$$\bar{N} = \frac{1}{J-1} \sum_{j=1}^{J-1} (\langle x^j \rangle - \bar{x})^2$$

2) la moyenne de la variance dans chaque chaîne

$$W = \frac{1}{J} \sum_{j=1}^J \frac{1}{N-1} \sum_{k=1}^N (x_k^j - \langle x^j \rangle)^2$$

UNE statistique de Gelman-Rubin est donnée par

$$R = \frac{\frac{N-1}{N}W + \frac{B}{N}}{W}$$

Il existe beaucoup de méthodes de mcmc pour accélérer la convergence

Adaptative MCMC:

Adapte la taille du proposal pendant le calcul de la chaîne, par exemple :

<https://arxiv.org/pdf/1804.07261.pdf>

Hamiltonian Monte Carlo:

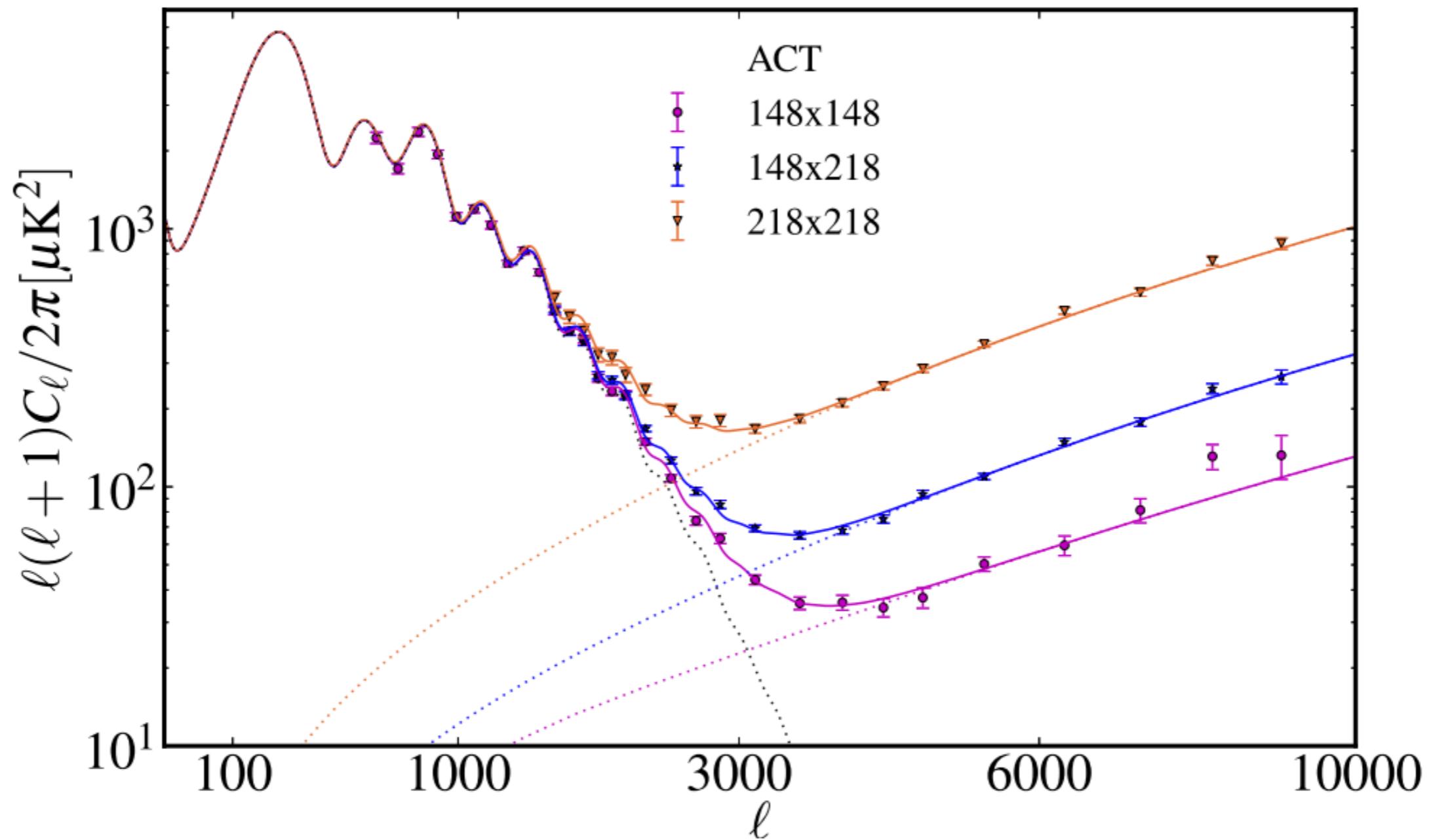
Utilise les dérivées de la likelihood pour réduire la corrélation entre différents états successifs en proposant des sauts vers des états distants qui maintiennent de fortes probabilités d'acceptances.

Echantillonnage de Gibbs (Gibbs sampling):

Une autre méthode mcmc pour les cas où échantillonner les conditionnelles est plus simple qu'échantillonner la distribution jointe

Exemple d'estimation de paramètres pour les observations CMB

<https://arxiv.org/pdf/1301.0776.pdf>



Exemple d'estimation de paramètres pour les observations CMB

<https://arxiv.org/pdf/1301.0776.pdf>

$$C_b^{\text{model}, \nu_1 \times \nu_2} = C_b^{\text{CMB}}[\theta_{\text{CMB}}] + C_b^{\text{fg}, \nu_1 \times \nu_2}[\theta_{\text{fg}}]$$

Paramètres du CMB Paramètres du modèle d'avant plan



Exemple d'estimation de paramètres pour les observations CMB

Paramètres du CMB

Paramètres du modèle d'avant plan

$$C_b^{\text{model}, \nu_1 \times \nu_2} = C_b^{\text{CMB}}[\theta_{\text{CMB}}] + C_b^{\text{fg}, \nu_1 \times \nu_2}[\theta_{\text{fg}}]$$

$$\begin{pmatrix} C_b^{\text{model}, 148 \times 148} \\ C_b^{\text{model}, 148 \times 218} \\ C_b^{\text{model}, 218 \times 218} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} C_b^{\text{CMB}}[\theta_{\text{CMB}}] + \begin{pmatrix} C_b^{\text{fg}, 148 \times 148} \\ C_b^{\text{fg}, 148 \times 218} \\ C_b^{\text{fg}, 218 \times 218} \end{pmatrix} [\theta_{\text{fg}}]$$

CMB ne dépend pas de la fréquence d'observation

Exemple d'estimation de paramètres pour les observations CMB

Paramètres du CMB **Paramètres du modèle d'avant plan**

$$C_b^{\text{model}, \nu_1 \times \nu_2} = C_b^{\text{CMB}}[\theta_{\text{CMB}}] + C_b^{\text{fg}, \nu_1 \times \nu_2}[\theta_{\text{fg}}]$$

$$\begin{pmatrix} C_b^{\text{model}, 148 \times 148} \\ C_b^{\text{model}, 148 \times 218} \\ C_b^{\text{model}, 218 \times 218} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} C_b^{\text{CMB}}[\theta_{\text{CMB}}] + \begin{pmatrix} C_b^{\text{fg}, 148 \times 148} \\ C_b^{\text{fg}, 148 \times 218} \\ C_b^{\text{fg}, 218 \times 218} \end{pmatrix} [\theta_{\text{fg}}]$$

CMB ne dépend pas de la fréquence d'observation

$$\begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} = \begin{pmatrix} C_b^{\text{model}, 148 \times 148} \\ C_b^{\text{model}, 148 \times 218} \\ C_b^{\text{model}, 218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}] + \begin{pmatrix} \Delta C_b^{148 \times 148} \\ \Delta C_b^{148 \times 218} \\ \Delta C_b^{218 \times 218} \end{pmatrix}$$

Exemple d'estimation de paramètres pour les observations CMB

Typiquement on veut sampler le postérieur :

$$\mathcal{P} \left(\theta_{\text{CMB}}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

Et on le fait à l'aide de l'algorithme de Metropolis Hasting, avec la likelihood

$$\ln \mathcal{L} \left(\begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \mid \theta_{\text{CMB}}, \theta_{\text{fg}} \right) \propto \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}] (\Sigma)^{-1} \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}]$$

$$\begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}] = \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} - \begin{pmatrix} C_b^{\text{model}, 148 \times 148} \\ C_b^{\text{model}, 148 \times 218} \\ C_b^{\text{model}, 218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}]$$

Exemple d'estimation de paramètres pour les observations CMB

A chaque étape de la chaîne on doit donc recalculer le résidu

$$\begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}] = \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} - \begin{pmatrix} C_b^{\text{model}, 148 \times 148} \\ C_b^{\text{model}, 148 \times 218} \\ C_b^{\text{model}, 218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}]$$

comme les paramètres du modèle changent.

C'est typiquement ce qui prend du temps dans les MCMC en cosmologie, on doit refaire tourner un code de Boltzmann (type CLASS/CAMB) à chaque pas pour recalculer le modèle (et l'exécution typique de ce genre de code est de quelques secondes)

Généralement ces chaînes mettent quelques jours à converger sur des machines modernes

Cobaya, a code for Bayesian analysis in Cosmology

- Author:** Jesus Torrado [✉](#) and Antony Lewis [✉](#)
- Source:** [Source code at GitHub](#)
- Documentation:** [Documentation at Readthedocs](#)
- Licence:** [LGPL](#) [✉](#) + mandatory bug reporting asap + mandatory arXiv'ing [✉](#) of publications using it (see [LICENCE.txt](#) [✉](#) for exceptions). The documentation is licensed under the [GFDL](#) [✉](#).
- E-mail list:** <https://cosmocoffee.info/cobaya/> [✉](#) – sign up for important bugs and release announcements!
- Support:** For general support, [CosmoCoffee](#) [✉](#); for bugs and issues, use the [issue tracker](#) [✉](#).
- Installation:** `pip install cobaya --upgrade` (see the [installation instructions](#) [✉](#); in general do not clone)

[build](#) [passing](#) [pypi](#) [v3.1.1](#) [docs](#) [passing](#) [codecov](#) [87%](#) [arXiv](#) [2005.05290](#)

[Home](#) [Monte Python](#) [Note Organiser](#) [Markup to Beamer](#) [CV](#) [Github](#)

Monte Python

The Monte Carlo code for [CLASS](#) in Python

[Download the latest version](#)

News

Latest release is **3.0**, and lives in a different repository. Head over to [there](#).

The full changelog is available on the [Github page](#).

To be informed by mail of the latest release, write to [brinckmann at physik.rwth-aachen.de](#), and you will be added to the mailing list.

Description

MONTE PYTHON is a Monte Carlo code for Cosmological Parameter extraction. It contains likelihood codes of most recent experiments, and interfaces with the Boltzmann code [CLASS](#) for computing the cosmological observables.

Several sampling methods are available: Metropolis-Hastings, Nested Sampling (through [MultiNest](#)), EMCEE (through [CosmoHammer](#)) and Importance Sampling.

R & D: accélérer l'estimation des paramètres cosmologiques

On doit échantillonner la distribution de probabilité jointe de $[\theta_{\text{CMB}}, \theta_{\text{fg}}]$

LCDM a 6 paramètres libres (+ pour les extensions) et le modèle considéré ici a 9 paramètres pour décrire les émissions d'avant plan.

Le temps de convergence des chaines dépend du nombre de paramètres

R & D: accélérer l'estimation des paramètres cosmologiques

On doit échantillonner la distribution de probabilité jointe de $[\theta_{\text{CMB}}, \theta_{\text{fg}}]$

LCDM a 6 paramètres libres (+ pour les extensions) et le modèle considéré ici a 9 paramètres pour décrire les émissions d'avant plan.

Le temps de convergence des chaines dépend du nombre de paramètres

L'idée est la suivante, plutôt que d'essayer d'estimer directement les paramètres du CMB, on va plutôt estimer la valeur du spectre en puissance du CMB conjointement avec les paramètres d'avant plan

$$\mathcal{P} \left(\theta_{\text{CMB}}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$



$$\mathcal{P} \left(\{C_b^{\text{CMB}}\}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

R & D: accélérer l'estimation des paramètres cosmologiques

On doit échantillonner la distribution de probabilité jointe de $[\theta_{\text{CMB}}, \theta_{\text{fg}}]$

LCDM a 6 paramètres libres (+ pour les extensions) et le modèle considéré ici a 9 paramètres pour décrire les émissions d'avant plan.

Le temps de convergence des chaines dépend du nombre de paramètres

L'idée est la suivante, plutôt que d'essayer d'estimer directement les paramètres du CMB, on va plutôt estimer la valeur du spectre en puissance du CMB conjointement avec les paramètres d'avant plan

$$\mathcal{P} \left(\theta_{\text{CMB}}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$



A priori c'est une mauvaise idée il y a plus de bins mesurés dans le spectre du CMB que de paramètres dans LCDM

$$\mathcal{P} \left(\{C_b^{\text{CMB}}\}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

R & D: accélérer l'estimation des paramètres cosmologiques

Mais: il est facile d'échantillonner les conditionnelles et on peut donc utiliser l'échantillonnage de Gibbs

L'idée est la suivante, pour échantillonner une distribution de probabilité jointe

$$\mathcal{P} \left(\{C_b^{\text{CMB}}\}, \theta_{\text{fg}} \mid \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

on peut alternativement échantillonner les conditionnelles

$$\begin{aligned} \{C_b^{\text{CMB}}\}^{i+1} &\leftarrow \mathcal{P} \left(\{C_b^{\text{CMB}}\} \mid \theta_{\text{fg}}^i, \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right) \\ \theta_{\text{fg}}^{i+1} &\leftarrow \mathcal{P} \left(\theta_{\text{fg}} \mid \{C_b^{\text{CMB}}\}^{i+1}, \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right) \end{aligned}$$

Après une période de burn in, les échantillons seront bien représentatifs de la distribution de probabilité jointe

Repartons de notre vraisemblance

$$\ln \mathcal{L} \left(\begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \middle| \theta_{\text{CMB}}, \theta_{\text{fg}} \right) \propto \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}] (\Sigma)^{-1} \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\theta_{\text{CMB}}, \theta_{\text{fg}}]$$



qui devient

$$\ln \mathcal{L} \left(\begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \middle| \{C_b^{\text{CMB}}\}, \theta_{\text{fg}} \right) \propto \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}] (\Sigma)^{-1} \begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}]$$

La clef est que maintenant C_b^{CMB} entre de façon linéaire dans le modèle

$$\begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}] = \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} C_b^{\text{CMB}} - \begin{pmatrix} C_b^{\text{fg}, 148 \times 148} \\ C_b^{\text{fg}, 148 \times 218} \\ C_b^{\text{fg}, 218 \times 218} \end{pmatrix} [\theta_{\text{fg}}]$$

$$\boldsymbol{\delta C_b}[\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}] = \boldsymbol{C_b^{\text{data}}} - \boldsymbol{A} C_b^{\text{CMB}} - \boldsymbol{C_b^{\text{fg}}}[\theta_{\text{fg}}]$$

La clef est que maintenant C_b^{CMB} entre de façon linéaire dans le modèle

$$\begin{pmatrix} \delta C_b^{148 \times 148} \\ \delta C_b^{148 \times 218} \\ \delta C_b^{218 \times 218} \end{pmatrix} [\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}] = \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} C_b^{\text{CMB}} - \begin{pmatrix} C_b^{\text{fg}, 148 \times 148} \\ C_b^{\text{fg}, 148 \times 218} \\ C_b^{\text{fg}, 218 \times 218} \end{pmatrix} [\theta_{\text{fg}}]$$

$$\boldsymbol{\delta} C_b [\{C_b^{\text{CMB}}\}, \theta_{\text{fg}}] = \boldsymbol{C}_b^{\text{data}} - \boldsymbol{A} C_b^{\text{CMB}} - \boldsymbol{C}_b^{\text{fg}} [\theta_{\text{fg}}]$$

Pour des paramètres d'avant plan fixés, on peut donc directement écrire le postérieur pour $\{C_b^{\text{CMB}}\}$

$$\mathcal{P}(\{C_b^{\text{CMB}}\} | \boldsymbol{C}_b^{\text{data}}, \theta_{\text{fg}})$$

Qui sera simplement une distribution gaussienne de moyenne

$$\langle C_b^{\text{CMB}} \rangle = [\boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A}]^{-1} [\boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{C}_b^{\text{data}} - \boldsymbol{C}_b^{\text{fg}} [\theta_{\text{fg}}])]$$

et covariance $\boldsymbol{Q} = [\boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A}]^{-1}$

Tirer un nombre aléatoire à partir d'une distribution gaussienne de moyenne et covariance connues est trivial, cette étape est donc triviale:

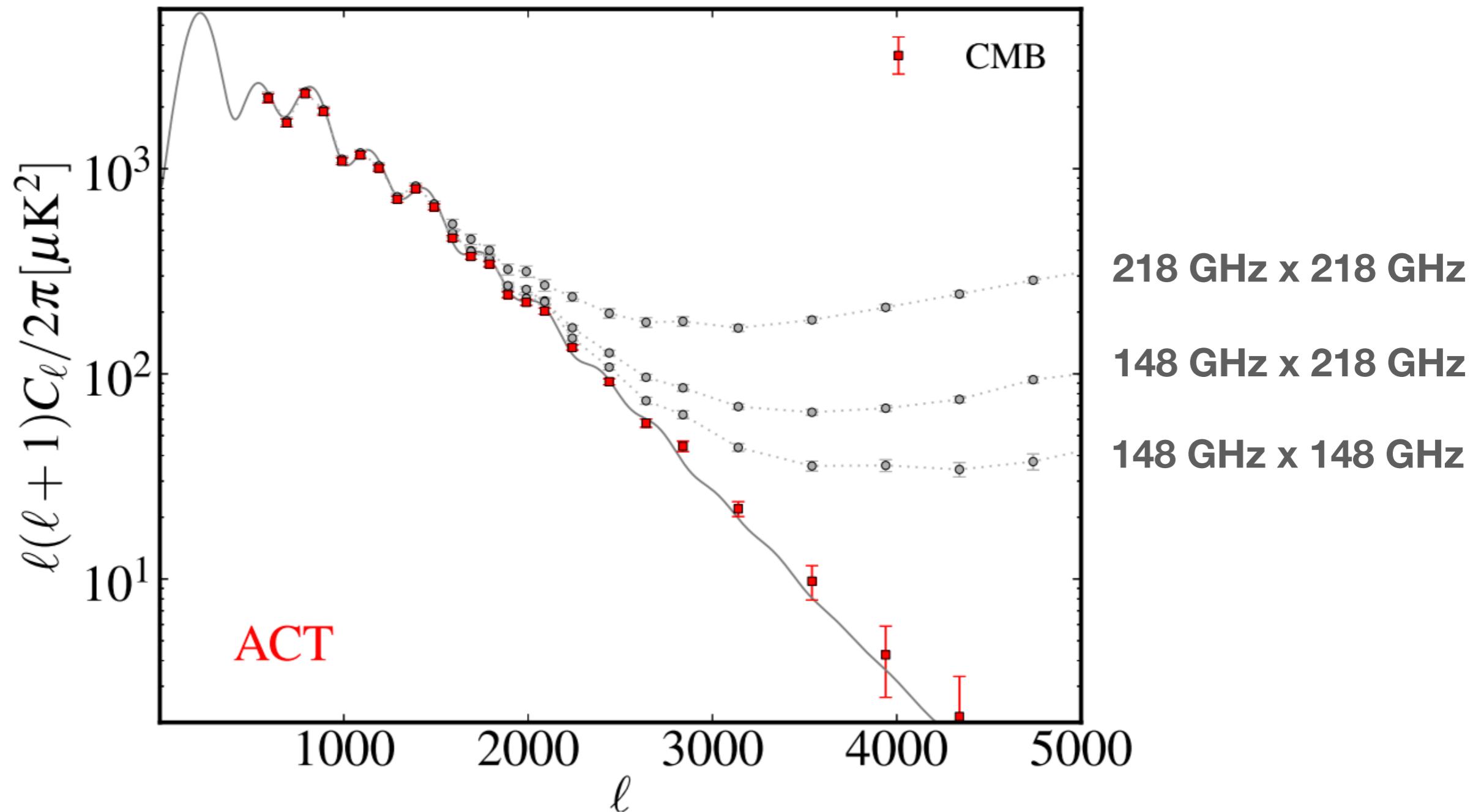
$$\{C_b^{\text{CMB}}\}^{i+1} \leftarrow \mathcal{P} \left(\{C_b^{\text{CMB}}\} \middle| \theta_{\text{fg}}^i, \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

On utilise par contre Metropolis pour échantillonner le postérieur des paramètres de fg (car non Gaussien), mais pas besoin d'utiliser un code de Boltzmann donc rapide

$$\theta_{\text{fg}}^{i+1} \leftarrow \mathcal{P} \left(\theta_{\text{fg}} \middle| \{C_b^{\text{CMB}}\}^{i+1}, \begin{pmatrix} C_b^{\text{data}, 148 \times 148} \\ C_b^{\text{data}, 148 \times 218} \\ C_b^{\text{data}, 218 \times 218} \end{pmatrix} \right)$$

En utilisant les résultats de la chaîne de Markov on a accès à l'ensemble des $\{C_b^{\text{CMB}}\}$ et la matrice de covariance, après marginalisation sur les paramètres d'avant plan

<https://arxiv.org/pdf/1301.0776.pdf>

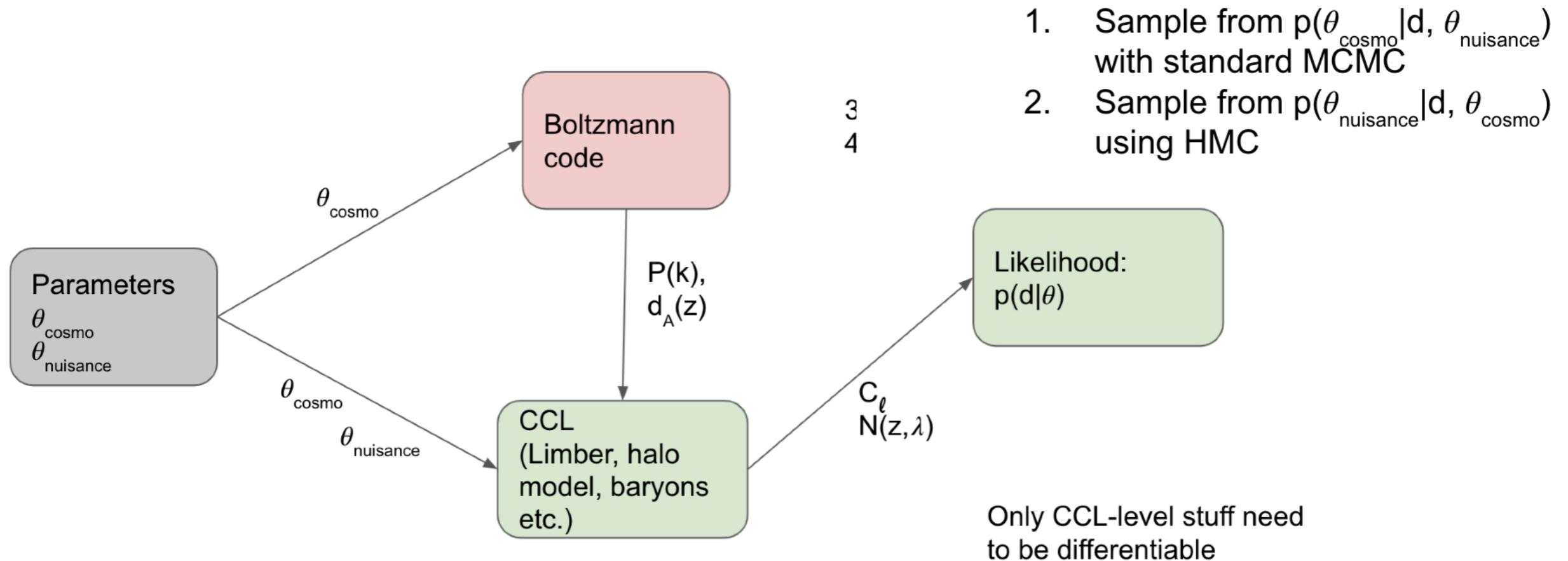


On peut alors directement mesurer les paramètres cosmologiques à partir de ce nouveau jeu de données « CMB only »

Un autre exemple de possibilité de Gibbs sampling : pour les surveys de galaxies type LSST (slide de Anze Slosar et David Alonso)

- E.g. 1, Photo-z uncertainties: $O(3-4)$ parameters for $O(10)$ bins = $O(40)$ params.
- E.g. 2, Baryonic effects: model as PCA basis or as a Gaussian process.
- E.g. 3, Galaxy bias/IAs: $O(3-4)$ parameters per bin in the simples PT schemes.
- E.g. 4, Beyond wCDM stuff: e.g. GP for $w(z)$, or scale-dependent mu-Sigma, primordial $P(k)$?

Bottom line $O(100)$ parameters inference problems with $O(^{\text{second}})$ likelihood evaluations are the pastasciutta that we need to eat.



Un mot sur les priors

(e.g <https://arxiv.org/pdf/2003.07355.pdf>)

Un prior plat sur un paramètre x n'est en général pas plat sur un paramètre dérivé y

Exemple EDE: on suppose qu'il y a un nouveau composant dans l'Univers, un champ scalaire qui évolue dans un potentiel, décrit par trois paramètres

$$\{m, f, \Theta_i\}$$

$$V(\Theta) = m^2 f^2 [1 - \cos(\Theta)]^3$$

On change de paramétrisation pour quelque chose qui a plus de sens d'un point de vue cosmologique:

le redshift pour lequel l'EDE atteint son maximum en tant que fraction de densité de l'univers, et la fraction de densité d'énergie correspondante

$$\{z_c, f_{\text{EDE}}, \Theta_i\}$$

Des priors plats dans la paramétrisation cosmologique ne sont pas plats dans la paramétrisation microphysique

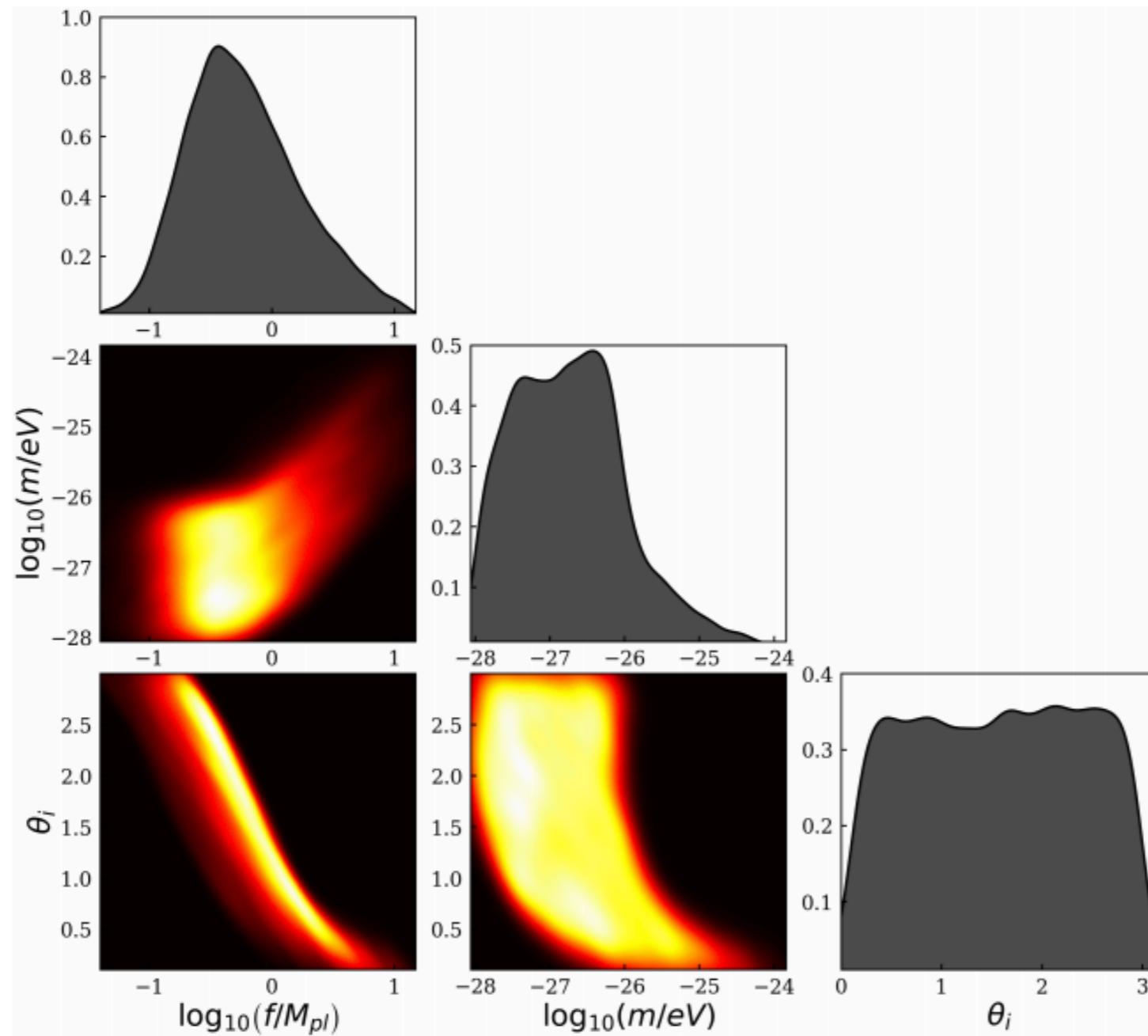
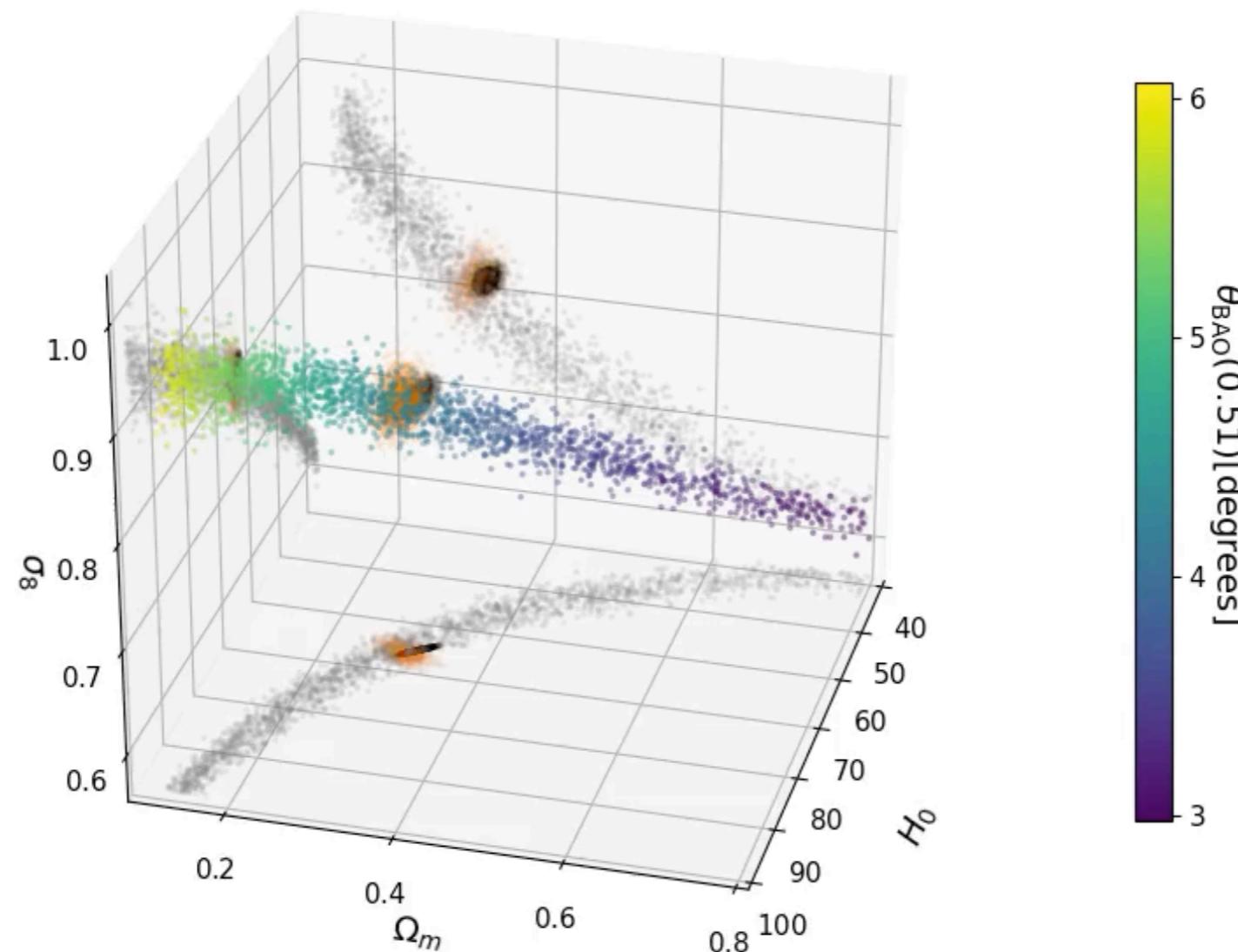


FIG. 13: Effective priors on f and m in EDE with flat priors on f_{EDE} , $\log_{10}(z_c)$, and θ_i . The distribution of f is peaked at $f = 0.59M_{pl}$, where $M_{pl} = 2.435 \times 10^{27}\text{eV}$ is the reduced Planck mass.

Un mot sur la visualisation

(From Antony Lewis)



Planck lensing 2018

+ BOSS BAO ($+\Omega_b h^2$ prior)

Planck 2018 TTTEEE

$\theta_{\text{BAO}}(0.51) \equiv r_s/D_M(z = 0.51)$