Non-Gaussianity analysis of CMB interferometric data

A.M. Aliaga[a,b], J.A. Rubiño-Martín[c], E. Martínez-González[a], R.B. Barreiro[a], J.L. Sanz[a]

[a]Instituto de Física de Cantabria, CSIC-Univ. de Cantabria, Santander, Spain

[b]Dpto. de Física Moderna, Univ. de Cantabria, Santander, Spain

[c]Instituto de Astrofísica de Canarias, Tenerife, Spain

TALK OUTLINE

- What an interferometer measures
- Signal-to-noise eigenmodes
- Smooth goodness-of-fit tests (Rayner and Best)
- Very Small Array (VSA) simulations
- Non-Gaussian simulations: Edgeworth expansion
- Non-Gaussian simulations: strings signal
- Conclusions

WHAT AN INTERFEROMETER MEASURES

If we assume that the plane approximation of the observed region is valid, an interferometer measures visibilities at a frequency ν:

$$V(u,\nu) = \int P(x,\nu)B(x,\nu)\exp\left(i2\pi u\cdot x\right)d^2x + V_{\mathrm{noise}}$$

x angular position of the observed point on the sky
u baseline vector in units of the wavelength of the observed radiation
P primary beam of the antennas (with $P(0) = 1$)
B brightness distribution ($\propto \Delta T/T$ for the CMB)

$\Rightarrow$ We want to test if V is Gaussian distributed.

Correlation matrix: $\langle V(u_i)V^*(u_j)\rangle =$

$$\int \widetilde{P}(u_i - u)\widetilde{P}^*(u_j - u)C(u)d^2u + \langle V_{\rm noise}(u_i)V^*_{\rm noise}(u_j)\rangle$$

$\widetilde{P}$ Fourier transform of P and $C(u)$ CMB power spectrum

Then

$$\Rightarrow \text{DATA} = \text{SIGNAL} + \text{NOISE}$$

It is desirable to analyse data such that $\frac{\text{SIGNAL}}{\text{NOISE}} \gg 1 \;\Rightarrow \ldots$

SIGNAL-TO-NOISE EIGENMODES

- Data: $\Delta = s + n$
  Correlation matrices $C_s = \langle ss^t \rangle$ and $C_n = \langle nn^t \rangle$
- Signal-to-noise eigenmodes:
  $\xi = RL_n^{-1}\Delta = RL_n^{-1}s + RL_n^{-1}n = \tilde{s} + \tilde{n}$
  (Bond J. R., 1995, Phys. Rev. Lett., 74, 4369)
  – $C_n = L_n L_n^t$
  – $R(L_n^{-1} C_s L_n^{-t})R^t = E = \mathrm{diag}(E_1, \ldots, E_n)$
- $\xi_k \to E_k$: signal-to-noise eigenvalues: they are large when the signal $\tilde{s}$ dominates over the noise $\tilde{n}$.
  $\langle \xi \xi^t \rangle = E + I$.
- We define $y_k = \xi_k / \sqrt{E_k + 1} \Rightarrow$ uncorrelated and normalized by construction (very interesting for our method!).

SMOOTH GOODNESS-OF-FIT TESTS

These tests were applied to the MAXIMA data by Cayón L. et al., 2003, MNRAS, 344, 917.

Given n independent realizations $\{x_i\}$ of a statistical variable x $(i = 1, \ldots, n)$, we want to test if x has a distribution function compatible with $f(x)$ (null hypothesis).

Rayner J.C.W. and Best D.J. (1989, *Smooth Tests of Goodness of Fit*, Oxford University Press, New York; 1990, International Statistical Review, 58, 9) propose some statistics to discriminate between f and another distribution (alternative hypothesis) which deviates smoothly from f:

- f is a Gaussian with zero mean and unit variance

$$
\begin{aligned}
U_1^2 &= n(\hat{\mu}_1)^2 \\
U_2^2 &= n(\hat{\mu}_2 - 1)^2/2 \\
U_3^2 &= n(\hat{\mu}_3 - 3\hat{\mu}_1)^2/6 \\
U_4^2 &= n[(\hat{\mu}_4 - 3) - 6(\hat{\mu}_2 - 1)]^2/24
\end{aligned}
$$

where $\hat{\mu}_\alpha = (\sum_{j=1}^{n} x_j^\alpha)/n$ (estimated moment of order α).

If the data follow a distribution given by f and n is large enough:

$$U_i^2 \sim \chi_1^2$$

If the data do not follow an f distribution we expect values of the statistics which are not compatible with a χ_1^2 distribution.

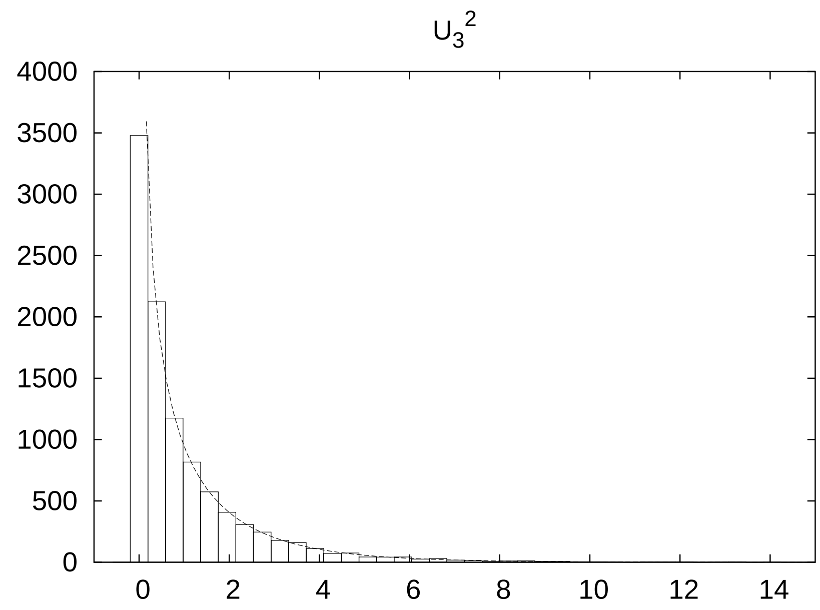
VERY SMALL ARRAY (VSA) SIMULATIONS

- VSA is an interferometer array sited at the Teide Observatory (Watson, R. A. et al., 2003, MNRAS, 341, 1057).
- We use an observation template of the so-called extended configuration ($\ell = 300 - 1500$, $\nu = 33$ GHz, FWHM $= 2.1^\circ$).
- Noise correlation matrix: diagonal.
- We take an integration time of 2500 hours (comparable to the whole signal-to-noise level of the dataset presented in Dickinson et al., 2004, astro-ph/0405341).
- $V(u) = V^R(u) + iV^I(u)$. The primary beam can be approximated by a Gaussian function (Hobson M. P. and Maisinger K., 2002, MNRAS, 334, 569) $\Rightarrow \langle V^R(u)V^I(u)\rangle = 0$.
- To perform our analysis we bin the data into cells of size of 9λ (Hobson and Maisinger, 2002).

- We transform $V^R(u)$ to give the y_k^R quantities ($k = 1, \ldots, N_{\text{vis}}$).
- $V^I(u) \rightarrow y_k^I$.
- Finally, we have the y_k quantities ($k = 1, \ldots, 2 \times N_{\text{vis}}$) with the associated signal-to-noise eigenvalues E_k such that:
  (1) $\langle y_i y_j \rangle = \delta_{ij}$
  (2) The larger is E_k the more signal contains y_k.
- By (1): if the visibilities are Gaussian distributed then $y_k \sim N(0,1)$ and they are independent.
- By (2): we can select only data with high signal-to-noise.
- Application of the smooth tests of goodness-of-fit: we compute the statistics U_i^2 for y_k.

- Gaussian simulations: Given $y_i \sim N(0,1) \Rightarrow V_i^R = \sum_j L_{s,ij}^R y_j$ are multinormal variables with the desired correlation matrix ($L_{s,ij}^R = \sigma_i (E_j^R)^{1/2} R_{ji}^R$).
  - Gaussian noise is added to each visibility.
  - In an analogous way we construct V_i^I.

Distribution of the U_3^2 statistic using 10000 simulated observations of our reference template (Gaussian CMB signal plus Gaussian noise)



The dashed line is a χ_1^2 function normalized to the total number of simulations.

NON-GAUSSIAN SIMULATIONS: EDGEWORTH EXPANSION

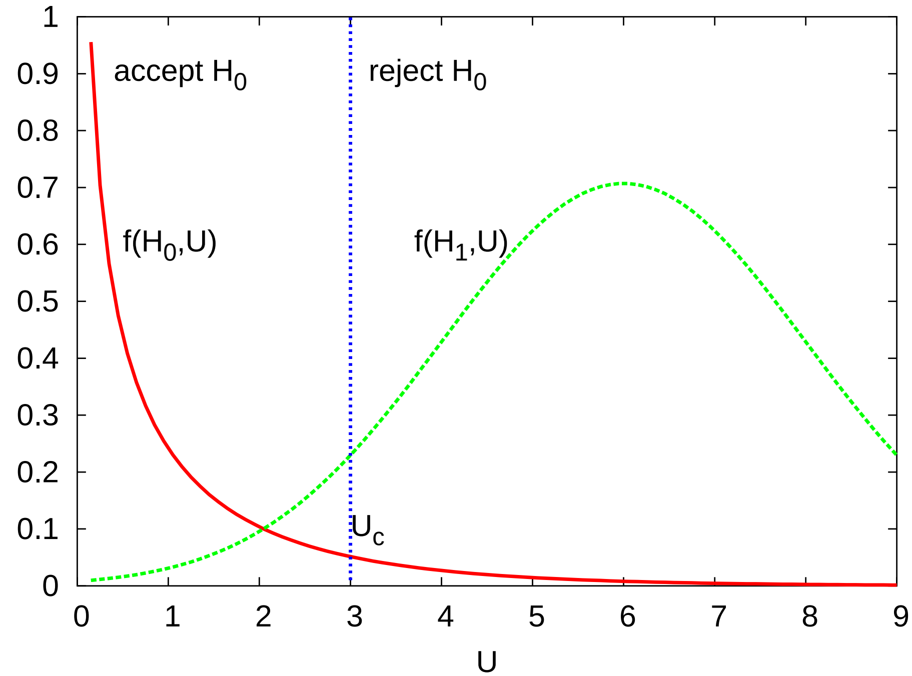
We construct non-Gaussian simulations using the Edgeworth expansion (Martínez-González E. et al., 2002, MNRAS, 336, 22).

- 5000 realizations of independent values y_k with zero mean, unit variance and skewness and kurtosis $S = K = 1$. The output simulations have a mean skewness equal to 0.96 ± 0.07 and a mean kurtosis equal to 0.85 ± 0.32 (i.e., there are simulations with kurtosis ~ 0.5).
- We generate $V_i^R = \sum_j L_{s,ij}^R y_j$ (and the same for V_i^I).
- We add Gaussian noise.
- The statistics U_i^2 of these simulations are calculated and the power to discriminate between them and Gaussian simulations is calculated. We take as significance level $\alpha = 5\%/1\%$.

- Significance level (α) and power of the test (p):

$$\alpha = \int_{U_c}^{\infty} f(H_0|U)dU$$

$$p = \int_{U_c}^{\infty} f(H_1|U)dU$$



Given $\alpha \downarrow \Rightarrow \; p \uparrow$

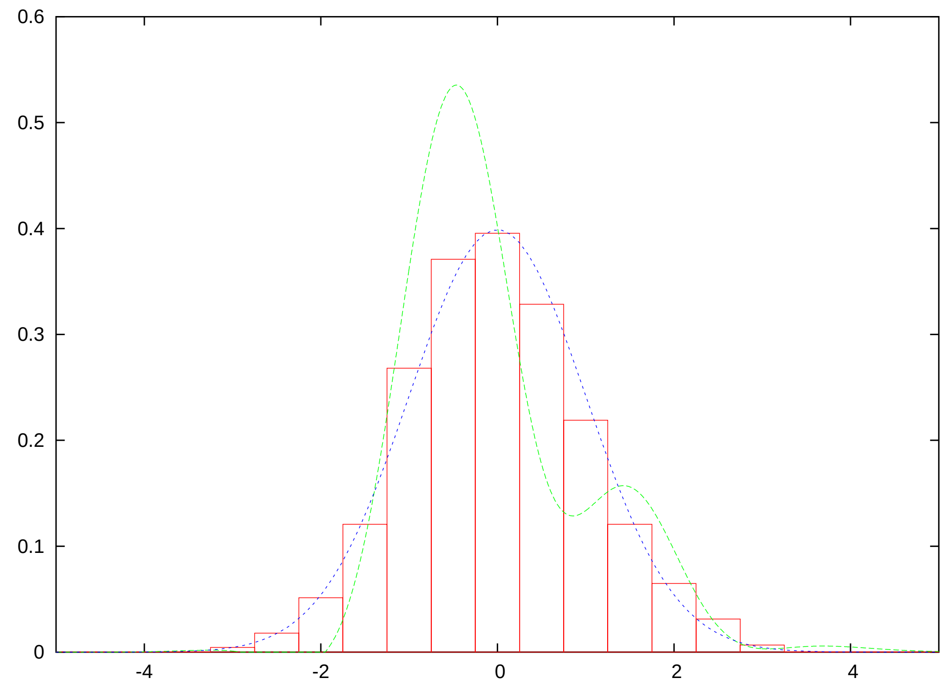
We take $\alpha = 0.05/0.01$

- The power of U_2^2, U_3^2 and U_4^2 is (U_1^2 does not give power):

$E^{\rm cut}(\sqrt{E^{\rm cut}})$	num.	U_2^2	U_3^2	U_4^2
0.0 (0.0)	1790	5.52/1.22	57.96/36.08	18.20/10.04
0.1 (0.32)	513	6.48/1.62	95.92/88.62	35.58/25.64
0.2 (0.45)	441	7.36/1.82	97.66/91.94	36.24/26.82
0.3 (0.55)	402	7.60/2.22	98.26/93.28	37.42/28.52
0.4 (0.63)	371	7.52/2.02	98.72/94.28	38.56/29.68
0.5 (0.71)	245	7.80/2.38	98.82/94.44	39.24/30.80

Note the gap between (i) to use all the data and (ii) to use only data with $E_k > 0.1$.

Edgeworth distribution
Gaussian distribution
Histogram of one simulation of the normalized eigenmodes y_k

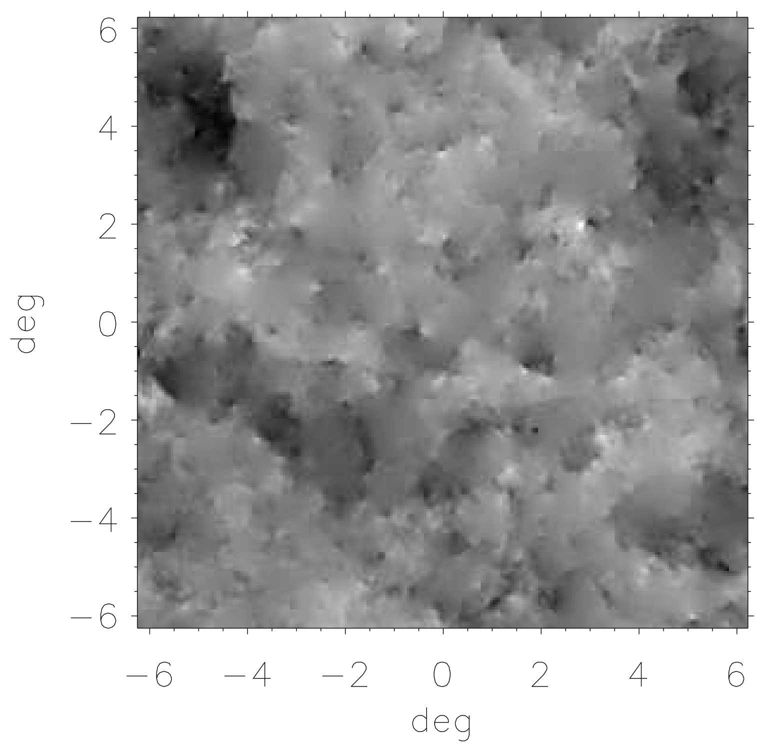


The histogram is very close to a Gaussian function but we can detect the non-Gaussian signal when we cut low signal-to-noise eigenvalues.

NON-GAUSSIAN SIMULATIONS: STRINGS SIGNAL

We apply our method to a strings map (Bouchet F. R., Bennett D. P., Stebbins A., 1988, Nature, 335, 410).

- Visibilities are obtained from the real map.
- The visibilities correlation matrix is calculated from the power spectrum obtained from the real map: $C(u) = \langle |a(\vec{u})|^2 \rangle \big|_{|\vec{u}|=u}$; a: Fourier transform of the real map.



- Analysis without noise: statistics values:

$U_1^2 \simeq 1.4 \cdot 10^3$ $(\hat{\mu}_1 \simeq -0.89)$

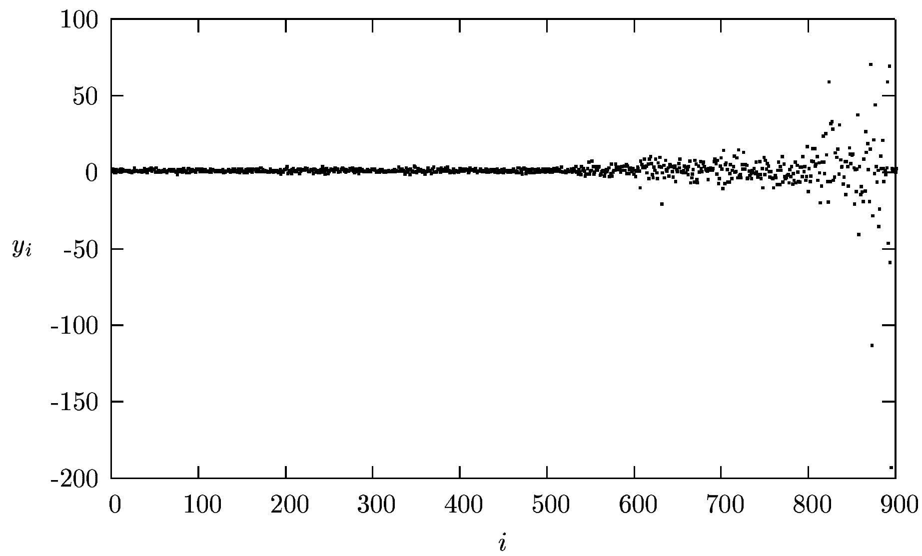
$U_2^2 \simeq 3.4 \cdot 10^8$ $(\hat{\mu}_2 \simeq 6.2 \cdot 10^2)$

$U_3^2 \simeq 2.9 \cdot 10^{13}$ $(\hat{\mu}_3 \simeq -3.1 \cdot 10^5)$

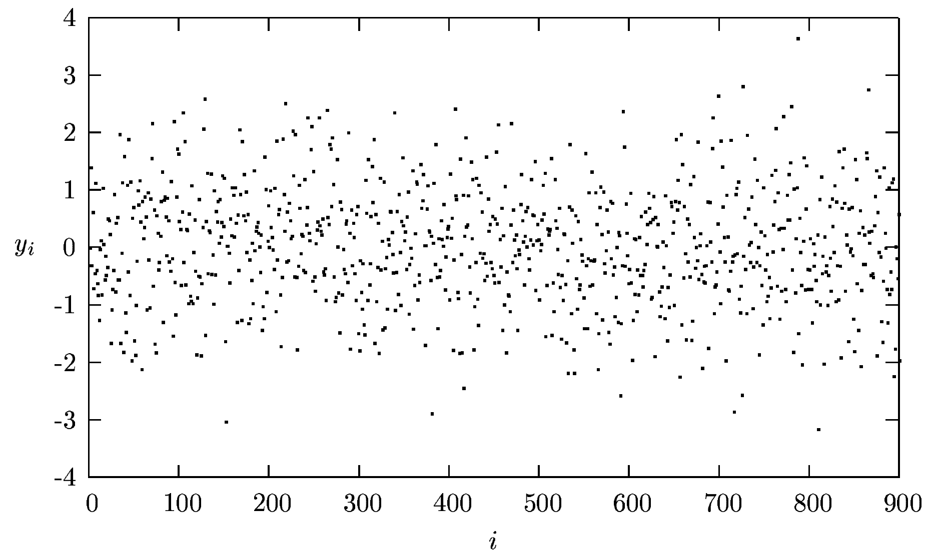
$U_4^2 \simeq 4.4 \cdot 10^{18}$ $(\hat{\mu}_4 \simeq 2.5 \cdot 10^8)$

Values incompatible with a Gaussian realization.

y_i values associated to the real part of the visibilities for the strings map. Remember that $\langle y_i y_j \rangle = \delta_{ij} \Rightarrow$ the data are not well decorrelated or normalized.



y_i values associated to the real part of the visibilities for a simulated Gaussian CMB map.



- Then, the behaviour of the y_i in the case of the strings map is a feature of the non-Gaussianity of the signal
- We are estimating the power spectrum with the expression which maximizes the likelihood under the hypothesis of Gaussianity for the temperature field (Bond et al., 2000, ApJ, 533, 19). Moreover the cosmic variance in the case of the

strings is higher that in the Gaussian case.

⇒ This is reflected in the fact that we are not able to estimate the correlation matrix from one strings realization and then the y_i quantities are not well decorrelated.

- Analysis with noise:

E^{cut}	U_1^2	U_2^2	U_3^2	U_4^2
0.	5.30/0.98	24.26/8.44	4.20/0.86	4.28/1.04
0.1	6.56/0.84	79.56/49.32	1.58/0.12	1.30/0.22
0.2	7.18/0.92	86.72/59.60	1.08/0.08	1.24/0.34
0.3	2.94/0.26	87.62/60.44	0.80/0.04	0.70/0.10
0.4	3.74/0.10	90.74/64.16	0.48/0.02	0.66/0.08
0.5	3.20/0.12	92.90/69.58	0.48/0.08	0.38/0.08

CONCLUSIONS

- We have developed a method to study the non-Gaussianity of CMB interferometric data.
- We have integrated the signal-to-noise formalism in the smooth goodness-of-fit tests developed by Rayner and Best.
- The signal-to-noise formalism allows to work with data with high signal-to-noise.
- Finally, we have studied the power of the test to discriminate between Gaussian and non-Gaussian simulations and we have seen that the method is able to detect non-Gaussian signals. The method detects Edgeworth simulations with a power equal to 99% ($\alpha = 5\%$) when the noise is the equivalent to the one of the dataset presented in Dickinson et al., 2004.