Spatial Statistics of the CMB

István Szapudi

Institute for Astronomy
University of Hawaii

Statistical Challenges in Modern Astronomy IV June 12-15, 2006

Outline

Random Fields

Definition

A random field is a spatial field with an associated probability measure: $\mathcal{P}(A)\mathcal{D}A$.

- Random fields are abundant in Cosmology.
- The cosmic microwave background fluctuations constitute a random field on a sphere.
- Other examples: Dark Matter Distribution, Galaxy Distribution, etc.
- Astronomers measure particular realization of a random field (ergodicity helps but we cannot avoid “cosmic errors”)

Isotropy – Anisotropy

Definition

The CMB statistics is assumed to be invariant under (SO(3)) rotations of the sky.

- This follows from the cosmological principles of homogeneity and isotropy.
- CMB temperature fluctuations are called “anisotropy”
- The anisotropy is statistically isotropic.

Gaussian – Non-Gaussian

Higher Order Spatial Statistics

Studies and characterizes the spatial distribution random fields beyond Gaussianity.

- Gaussian: statistics up to second order.
- Non-Gaussian: connected moments are non-zero beyond second order.
- Gaussian/Non-Gaussian: correlated/random phases.
- The CMB is Gaussian to a high degree of accuracy
- Isotropic non-Gaussian or Anisotropic Gaussian?

Definitions

- The ensemble average $\langle A \rangle$ corresponds to a functional integral over the probability measure.
- Physical meaning: average over independent realizations.
- Ergodicity: (we hope) ensemble average can be replaced with spatial averaging.

Joint Moments

$$F^{(N)}(x_1, \ldots, x_N) = \langle T(x_1), \ldots, T(x_N) \rangle$$

Connected Moments

These are the most frequently used spatial statistics

- Typically we use fluctuation fields $\delta = T/\langle T \rangle - 1$

Connected moments are defined recursively

$$\langle \delta_1, \ldots, \delta_N \rangle_c = \langle \delta_1, \ldots, \delta_N \rangle - \sum_P \langle \delta_1 \ldots \ldots \delta_i \rangle_c \ldots \langle \delta_j \ldots \delta_k \rangle_c \ldots$$

- With these the N-point correlation functions are

$$\xi^{(N)}(1, \ldots, N) = \langle \delta_1, \ldots, \delta_N \rangle_c$$

Basic Objects

These are N-point correlation functions.

Special Cases

Two-point functions	$\langle \delta_1 \delta_2 \rangle$
Three-point functions	$\langle \delta_1 \delta_2 \delta_3 \rangle$
Cumulants	$\langle \delta_R^N \rangle_c = S_N \langle \delta_R^2 \rangle^{N-1}$
Cumulant Correlators	$\langle \delta_1^N \delta_2^M \rangle_c$
Conditional Cumulants	$\langle \delta(0) \delta_R^N \rangle_c$

- In the above δ_R stands for the fluctuation field smoothed on scale R (different R's could be used for each δ's).
- Host of alternative statistics exist: e.g. Minkowski functions, void probability, minimal spanning trees, phase correlations, etc.

Complexities

Combinatorial explosion of terms

- N-point quantities have a large configuration space: measurement, visualization, and interpretation become complex.
- CPU intensive measurement: M^N scaling for N-point statistics of M objects.
- Theoretical estimation
- Estimating reliable covariance matrices
- As we will see, a close look at symmetries helps with these issues.

Symmetries in General

Invariance under a (Lie) group G

Expansion according to irreducible representations

$$\begin{aligned} \xi(gx, gy) &= \xi(x, y) \\ \delta(x) &= \sum a_i^\alpha \Psi_i^\alpha \end{aligned}$$

Irreducible reps diagonalize corr. matrix

$$\begin{aligned} \langle a_i^\alpha a_j^\beta \rangle &= \delta_{ij}\delta_{\alpha\beta} \int dx \xi(0, x) \sum_s \Psi_s^\alpha(0)\Psi_s^{\alpha *}(x) \frac{1}{f_\alpha} \int dg \\ \langle a_{l_1 m_1} a_{l_2 m_2} \rangle &= \delta_{l_1 l_2}\delta_{m_1 m_2} C_{l_1} \end{aligned}$$

Expansions for the sphere

Statistics under SO(3) invariance

Angular Power Spectrum

$$\xi(\theta) = \sum \frac{2l+1}{4\pi} C_l P_l(\cos\theta)$$

- Legendre polynomials $P_l \propto$ bipolar spherical harmonics $L = 0$

Bispectrum: Tripolar Expansion

$$S_{l_1 l_2 l_3}(\hat{x}_1, \hat{x}_2, \hat{x}_3) \equiv \sum_{m_1, m_2, m} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} Y_{l_1 m_1}(\hat{x}_1) Y_{l_2 m_2}(\hat{x}_2) Y_{l_3 m_3}(\hat{x})$$

$$\xi^3(\hat{x}_1, \hat{x}_2, \hat{x}_3) = \sum_{l_1 l_2 l_3} B_{l_1, l_2, l_3} S_{l_1 l_2 l_3}(\hat{x}_1, \hat{x}_2, \hat{x}_3)$$

Algorithms: SpICE
Szapudi Etal 2001a,b

- www.ifa.hawaii.edu/cosmowave
- about 258 downloads to date
- Heuristic weighting/pseudo-C_l method
- Use fast harmonic transform (HEALPix Gorski etal 2005) to calculate raw two-point correlators: 2 minutes for $N_{sdie} = 512$.
- The coupling kernel is diagonal in pixel space: invert there
- Use Gausss-Legendre quadrature to calculate C_l's
- Monte Carlo simulations for covariance matrices

Algorithms: SpICE cont'd

- Equivalent technique (MASTER Hivon Etal 2002) directly inverts the (non-diagonal) coupling matrix
- reanalysed WMAP-I with SpICE (Fosalba & Szapudi 2004) coupling matrix in harmonic space
- Part of Planck level 2 pipeline
- Works for $N - M$ cumulant correlators (Szapudi & Szalay 1992) as well
- Direct generalization to bispectrum is not fast enough ($N^{5/2}$) and would calculate $l^3 \simeq N^{3/2}$ bins

Flat Sky Approximation: Translation Invariance

The Dirac δ's are expressing the symmetry

Fourier Transform

$$\langle\delta(k_1)\delta(k_2)\rangle = (2\pi)^D\delta_D(k_1+k_2)P(k_1)$$
$$\langle\delta(k_1)\delta(k_2)\delta(k_3)\rangle = (2\pi)^D\delta_D(k_1+k_2+k_3)B(k_1,k_2,k_3)$$

Generalized Wiener-Khinchin Theorem

$$\xi(x_1,x_2) = \int \frac{d^Dk}{(2\pi)^D}P(k)e^{ik(x_1-x_2)}$$
$$\xi(x_1,x_2,x_3) = \int \Pi_{i=1}^{3}\frac{d^Dk_i}{(2\pi)^D}B(k_1,k_2,k_3)e^{i(\sum k_jx_j)}\delta_D(\sum k_j).$$

Symmetries: Rotation Invariance

These further simplify things...

- The two-point function depends on the distance of the two variables.

The integral is 1D only

$$
\begin{aligned}
\xi(r) &= \int \frac{kdk}{2\pi} P(k) J_0(kr) \text{ 2D}, \\
\xi(r) &= \int \frac{k^2dk}{2\pi^2} P(k) j_0(kr) \text{ 3D}
\end{aligned}
$$

- Can we use the same idea for the 3 point statistics?
- They depend on a triangle size and shape.

Rotation Invariance: 3D

SO(3)

- If we parametrize the triangle by two unit vectors, and two sizes, bipolar spherical expansion can capture the rotation invariance.
- In the bipolar expansion only $L = 0$ (scalar) can happen because of SO(3) invariance: Legendre polynomials.

Multipole Expansion For 3-point Statistics

$$B(k_1, k_2, \theta) = \sum_l B_l(k_1, k_2) P_l(\cos\theta) \frac{2l+1}{4\pi},$$

- Note that tripolar expansion is also possible and more convenient in some cases.

Consequence of Rotation Invariance in 3D

- The above expansion simplifies the connection between Fourier and real space 3-point statistics.
- In particular the 6 dimensional (intractable integral) becomes 2 dimensional.

Double Hankel Transform

$$\xi_l^3(r_1, r_2) = \int \frac{k_1^2}{2\pi^2} dk_1 \frac{k_2^2}{2\pi^2} dk_2 (-1)^l B_l(k_1, k_2) j_l(k_1 r_1) j_l(k_2 r_2)$$

- Multiples do not mix in the transformation!

Rotational Invariance in 2D

U(1) invariance

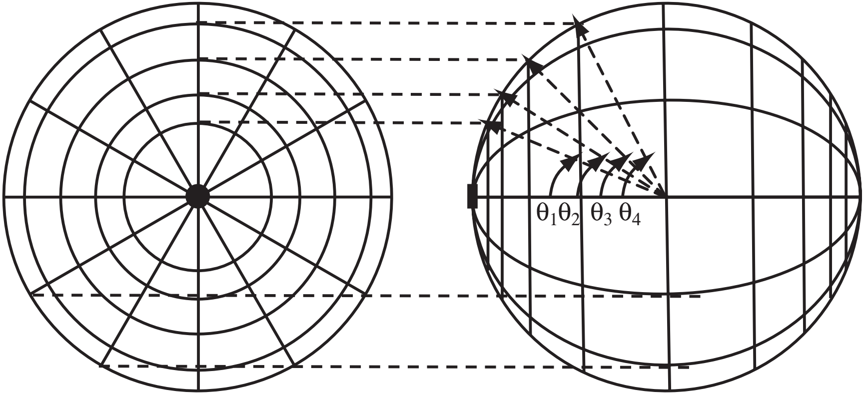
- In 2 dimensions SO(3) $\rightarrow$ U(1), we need to use cosine expansion.

2D Expansion

$$
\begin{aligned}
B(k_1, k_2, \theta) &= \frac{B_0(k_1,k_2)}{2} + \sum_{n<0} B_n(k_1, k_2)\cos(n\theta) \\
\xi_n^3(r_1, r_2) &= \int \frac{k_1}{2\pi} dk_1 \frac{k_2}{2\pi} dk_2 (-1)^n B_n(k_1, k_2) J_n(k_1 r_1) J_n(k_2 r_2)
\end{aligned}
$$

- The same expansion can be applied to spherical geometry in the flat sky limit, and for the (redshift space) projected three-point function.

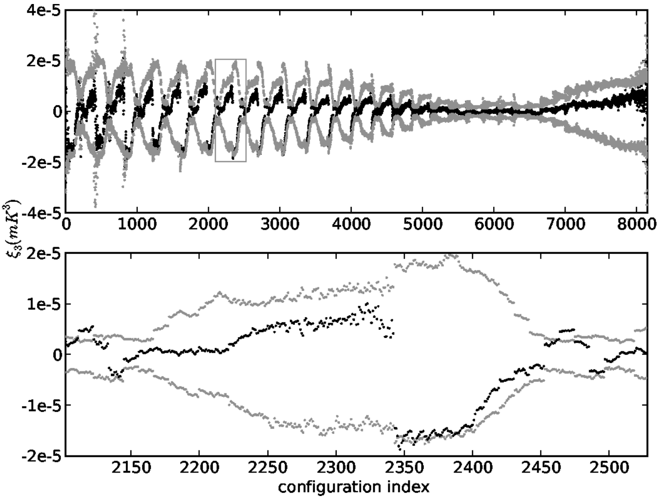
New Algorithm for 3pt



New Algorithm for 3pt Cont'd

- Naively N^3 calculations to find all triplets in the map: overwhelming (millions of CPU years for WMAP)
- Regrid CMB sky around each point according to the resolution
- Use hierarchical algorithm for regridding: $N \log N$
- Correlate rings using FFT's (total speed: 2 minutes/cross-corr)
- The final scaling depends on resolution $N(\log N + N_\theta N_\alpha \log N_\alpha + N_\alpha N_\theta (N_\theta + 1)/4)/2$
- With another cos transform one and a double Hankel transform one can get the bispectrum
- In WMAP-I: 168 possible cross correlations, about 1.6 million bins altogether.
- How to interpret such massive measurements?

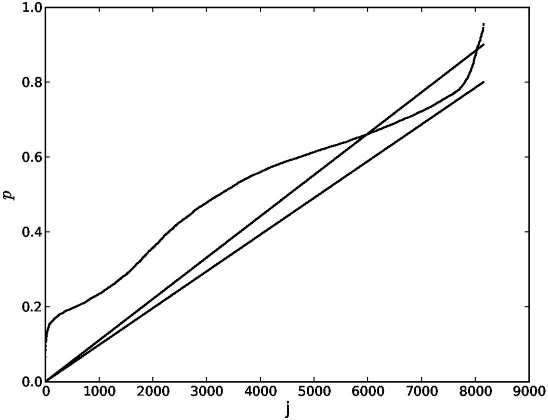
3pt in WMAP


4e-5
2e-5
0
-2e-5
-4e-5
0
1000
2000
3000
4000
5000
6000
7000
8000
$\xi_3(mK^3)$
2e-5
1e-5
0
-1e-5
-2e-5
2150
2200
2250
2300
2350
2400
2450
2500
configuration index

Question1: How consistent is the CMB with Gaussian?
Hypothesis testing with FDR

- Gaussianity: null hypothesis
- Combining 1.6 million (correlated) measurements
- Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)
- Miller etal (2001) Chen and Szapudi (2005)
- False Discovery Rate: fix α s.t. $0 \leq \alpha \leq 1$
- $P_1, ..., P_N$ *sorted* p-values
- let $d = max \left\{ j : P_j < \frac{j\alpha}{c_N N} \right\}$
- Configurations with $i < d$ are rejected, and $\langle FDR \rangle \leq \alpha$ is guaranteed
- $c_N \lesssim \sum_{i=1}^{N} i^{-1}$ for correlated data, otherwise 1 (Hopkins etal 2002)

FDR in WMAP



- $\alpha = 0.8$ produces no rejection ($\alpha = 0.9$ produces many rejections)
- the data are fully consistent with Gaussianity

The Question: Are there Primordial Gaussianities?
Parameter Fitting

- It is difficult to translate FDR results into traditional confidence levels
- Our goal is to constrain primordial non-Gaussianities (others: e.g., point sources, foregrounds, SZ, lensing, etc)
- Since three-point correlations are expected to dominate (unless some symmetry erases three-point correlations)
- $\langle\delta^3\rangle \ll \langle\delta^2\rangle$

Simplest Phenomenological Parametrization

$$\delta = \delta_G + f_{NLT}(\delta_G^2 - \langle\delta_G^2\rangle)$$

A Popular Parametrization

Phenomenology motivated by inflationary models

- $\Phi(k)$ is the primordial curvature perturbation

Perturbative Model

$$\Phi(x) = \Phi_L(x) + f_{NL}(\Phi(x)_L^2 - \langle\Phi_L^2(x)\rangle)$$

- The leading order effect will be bispectrum, (three-point correlations)

Bispectrum (Komatsu and Spergel 2001)

$$b_{l_1 l_2 l_3} = 2\int r^2 dr (b_{l_1}^L(r) b_{l_2}^L(r) b_{l_3}^{NL}(r) + perm.)$$

- where b^L and $b^{NL} = \frac{2}{\pi}\int k^2 dk f(k) g_{Tl}(k) j_l(k)$ with $f = P_\Phi(k)$ or f_{NL} respectively.
- Simple calculation through modification of CAMB (or CMBFAST).

Measurements of $C_l^{2,1}$ in WMAP-I and WMAP-II

- SpICE
- Flat weights (we have tried other weights and it does not make any difference)
- Monte Carlo simulations for covariance matrices
- Gaussian realizations with "Best Fit" C_l's
- Conservative Kp2 mask
- 110 Noise maps for WMAP-I, 200 self generated noise maps for WMAP-II
- Results yield about 1000x168 bins for each simulation or 16800 $C^{2,1}$ measurements.

Theoretical Predictions

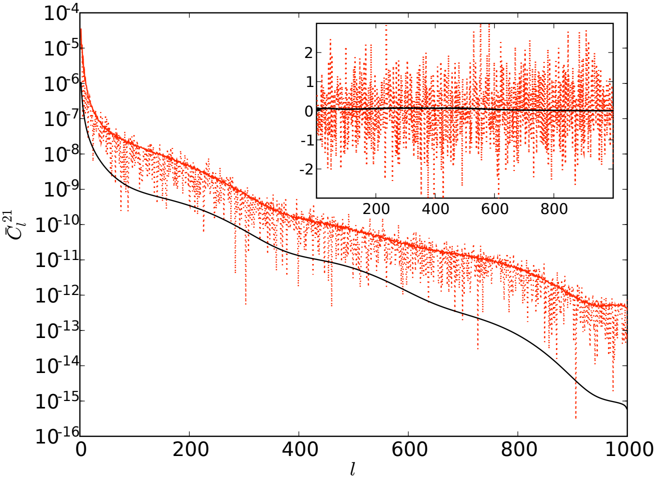
- Theory scales linearly with f_{NL}.
- Modified CAMB to obtain the bispectrum

Cooray (2001)

$$C_l^{2,1} = \sum_{l_1 l_2} B_{l_1 l_2 l} w_{l_1, l_2} \begin{pmatrix} l_1 & l_2 & l \\ 0 & 0 & 0 \end{pmatrix} \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2l+1)}}$$

- Predictions include beam and pixel window functions

$C_l^{2,1}$ Measurements from WMAP-II



Maximum Likelihood Fit
Schematic Plan

- Estimate covariance matrix C of the measurements using a set of Monte Carlo simulations of the data
- Since non-Gaussianity is small, most noise is expected to come from the Gaussian variance
- Gaussian simulations are adequate if f_{NL} is small
- Estimate $\chi^2 \simeq y^T C^{-1} y$, where $y = C_l^{2,1}(WMAP) - C^{2,1}(f_{NL})$
- Minimize χ^2 to find the best fit value and the variance of the parameter

Problem 1: too many bins in three-point functions

- About 1.6 million bins for limited resolution, full three-point analysis.
- 168k bins for full resolution $C_l^{2,1}$, although it can be decreased by using l-bands and averages.
- Averaging cross-correlations and band powers can help that.
- This can be a problem for any statistics with many bins: e.g. full resolution cross-C_l's between many maps.

Covariance Matrix Too Large

- Direct inversion of a $10^6 \times 10^6$ matrix is not feasible, 40-50k matrix is doable, although inconvenient.
- It turns out not even needed!
- Let $M = (p \times q)$ simulation matrix with p is the number of realizations, q is the number of bins (length of the result vector).
- $C = MM^T/p$
- Consider the SVD $M = U\Lambda V$, then $MM^T = U\Lambda^2 U^T$
- Solve the dual problem first $M^T M = V^T \Lambda^2 V$, then find the eigen-vectors U from MV^T.
- In our case the problem is reduced to inverting a 110×110 matrix.
- Corollary : even theoretically, only up to p modes can be extracted.

Problem 2: Too Few Simulations

- At most 110 noise realizations for WMAP-I are available
- Hundreds (maybe thousands) of simulations are feasible, for WMAP-II but not millions or billions, even with fast codes.
- For convergence we need $p \gg q$: unfeasible.
- The problem is slightly less severe for C_l, but it is there: we are never in the $p \gg q$ limit
- A model dependent remedy is to fit a simple analytical form

Random Matrix Theory
Initiated by Jenö Wigner

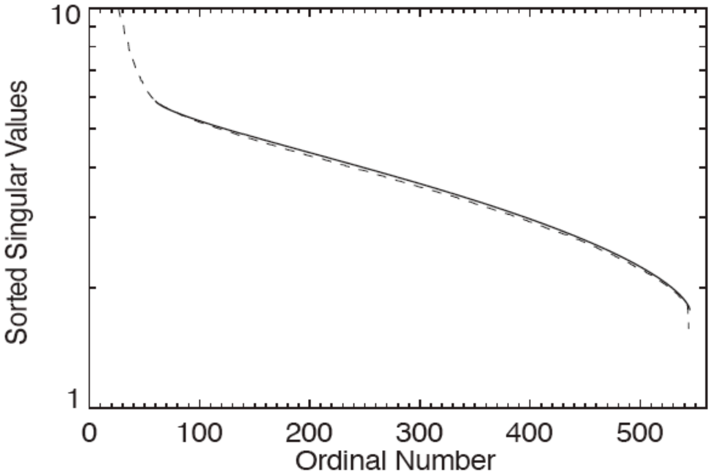
- $M_{ia}M_{jb} = \sigma^2\delta_{ij}\delta_{ab}$

Eigenvalue Spectrum

$$\rho(\lambda) = \frac{q^{1/2}}{\pi\lambda\hat{\sigma}}\sqrt{(\lambda_{max}^2 - \lambda^2)(\lambda^2 - \lambda_{min}^2)}$$
$$\lambda_{max,min} = \sqrt{2}\sigma\sqrt{(p+q)/2 \pm \sqrt{pq}}$$

- Sengupta and Mitra (1997)

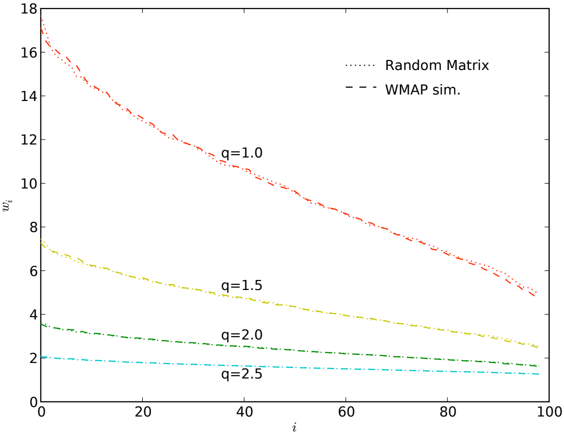
Typical Situation



Power Mapping
Guhr & Kälber (2003)

- A way to decrease the noise and identify significant eigenmodes
- $C_{kl} \rightarrow \mathrm{sgn}(C_{kl}) \mid C_{kl} \mid^{r}$
- variance scales as $1/p^{r/2}$
- e.g., 100 simulations with $r = 2$ scatters like 10,000 simulations
- A series of power maps will reveal hidden structure in otherwise random looking covariance matrices.
- This is more an art than science...

WMAP-II Covariance Matrix vs. Random



The Zen of the Covariance Matrix

- The WMAP-II (and also WMAP-I) covariance matrices are consistent with random
- It is correct to take into account covariance in the χ^2 when those covariances are well known.
- It can be misleading to use the noisy off diagonal terms in a random matrix
- Investment banks paid for this experience dearly
- in this case it appears to be best to use a diagonal χ^2, which is fully consistent with the randomness of the covariance
- One possible refinement is to fit N_{eff} d.o.f. from the shape of the curve.

Constraints for f_{NL} from $C_l^{2,1}$ in WMAP-II

These results are preliminary

- We use the diagonal elements of covariance matrix: the rest is consistent with noise.
- Using an average over the 168 cross-correlations we obtain $f_{NL} = 22 \pm 52 \quad (68\%)$
- These results are comparable with the results in Spergel etal (2006)
- No optimal weighting, but cross correlations and more configuration dependence
- While we used accurate masks which take care of most of the point source contribution, we did not subtract possible contribution from unresolved point sources.

Summary

Three-point statistics of the CMB

- Most of the information in the CMB is in the power spectrum
- Did not talk about parameter estimation from the 2pt and polarization (both are hot topics though)
- Higher order statistics is especially difficult for CMB
- Careful look at the symmetries and the perturbative nature of possible non-Gaussianities help with estimators and algorithms
- The CMB is fully consistent with Gaussian, both in terms of hypothesis testing and parameter fitting