Higher Criticism: Theory and Applications in Cosmology

Jiashun Jin

Statistics Department
Purdue Uinversity

Collaborators

Alphabetically:

Nabila Aghanim	Université Paris Sud
Laura Cayon	Purdue University
David Donoho	Stanford University
Olivier Forni	Université Paris Sud
Jeau-Luc Starck	Service d'Astrophysique
Anna Treaster	Purdue University

Agenda

- Higher Criticism statistic
- Optimal Adaptivity for detecting sparse Gaussian mixtures
- Application to nonGaussian detection
  - detection of cosmic string
  - WMAP first year data

Tukey's story

- Example: A young scientist administers 250 uncorrelated tests, out of which 11 were significant at the 5% level.
- Question is: Is this surprising?
- Answer: No, we expect

$$250 \times 5\% = 12.5$$

significance at 5% level.

Higher Criticism, Formalization proposed by Tukey

- Higher Criticism statistics:

$$HC_{.05,n} = \sqrt{n}[(\text{Fraction Significant at } .05) - .05]/\sqrt{.05 \times .95}$$

  and typically,

$$\text{Reject } H_0 \text{ if and only if } \quad HC_{.05,n} \geq 2$$

- Solution to previous example:

$$HC_{.05,n} = [11 - 12.5]/\sqrt{250 \times .05 \times .95} = -.43, \quad \Longrightarrow \quad \text{Accept } H_0.$$

- Higher Criticism, or "Second-Level Significance Testing," indicating

Significance of Overall Body of Tests.

Our Proposal

We propose

$$HC^* = \max_{0<\alpha<1} \sqrt{n}[(\text{Fraction Significant at } \alpha) - \alpha]/\sqrt{\alpha(1-\alpha)}$$

- Generalization of Tukey's $HC_{\alpha,n}$ to allow selection of level α
- Looking for unusually large number of "moderate significances"

Only need p-values to implement HC

Obtain individual p-values by:

$$p_i = P\{|N(0,1)| \geq |X_i|\}$$

- sort p-values:

$$p_{(1)} < p_{(2)} \ldots < p_{(n)}$$

- calculate i^{th} z-score:

$$HC_{n,i} = \sqrt{n}\left[\frac{\frac{i}{n} - p_{(i)}}{\sqrt{p_{(i)}(1 - p_{(i)})}}\right]$$

- take maximum:

$$HC_n^* = \max_{\{1 \leq i \leq n\}} \ HC_{n,i}$$

Detection of Sparse Gaussian Mixture

Hypothesis Testing:

$$H_0: \; X_i \stackrel{i.i.d}{\sim} N(0,1), \qquad 1 \le i \le n,$$

$$H_1^{(n)}: \; X_i \stackrel{i.i.d}{\sim} (1-\epsilon_n)N(0,1) + \epsilon_n N(\mu_n, 1), \qquad 1 \le i \le n.$$

- Goal: testing $\epsilon_n = 0$ vs. $\epsilon_n > 0$
- Approach: study for what (ϵ_n, μ_n) H_0 and $H_1^{(n)}$ are separable

Subtlety of the Problem

Calibrate with:

$$\epsilon_n = n^{-\beta}, \qquad \underline{0.5 < \beta < 1},$$

$$\mu_n = \sqrt{2r \log n}, \qquad \underline{0 < r < 1}.$$

Challenges:

- Very sparse: $\epsilon_n \ll \frac{1}{\sqrt{n}}$
- Moderate significance: $\mu_n < \sqrt{2 \log n}$
- Different from contiguity: μ_n increases with n

Detection Boundary

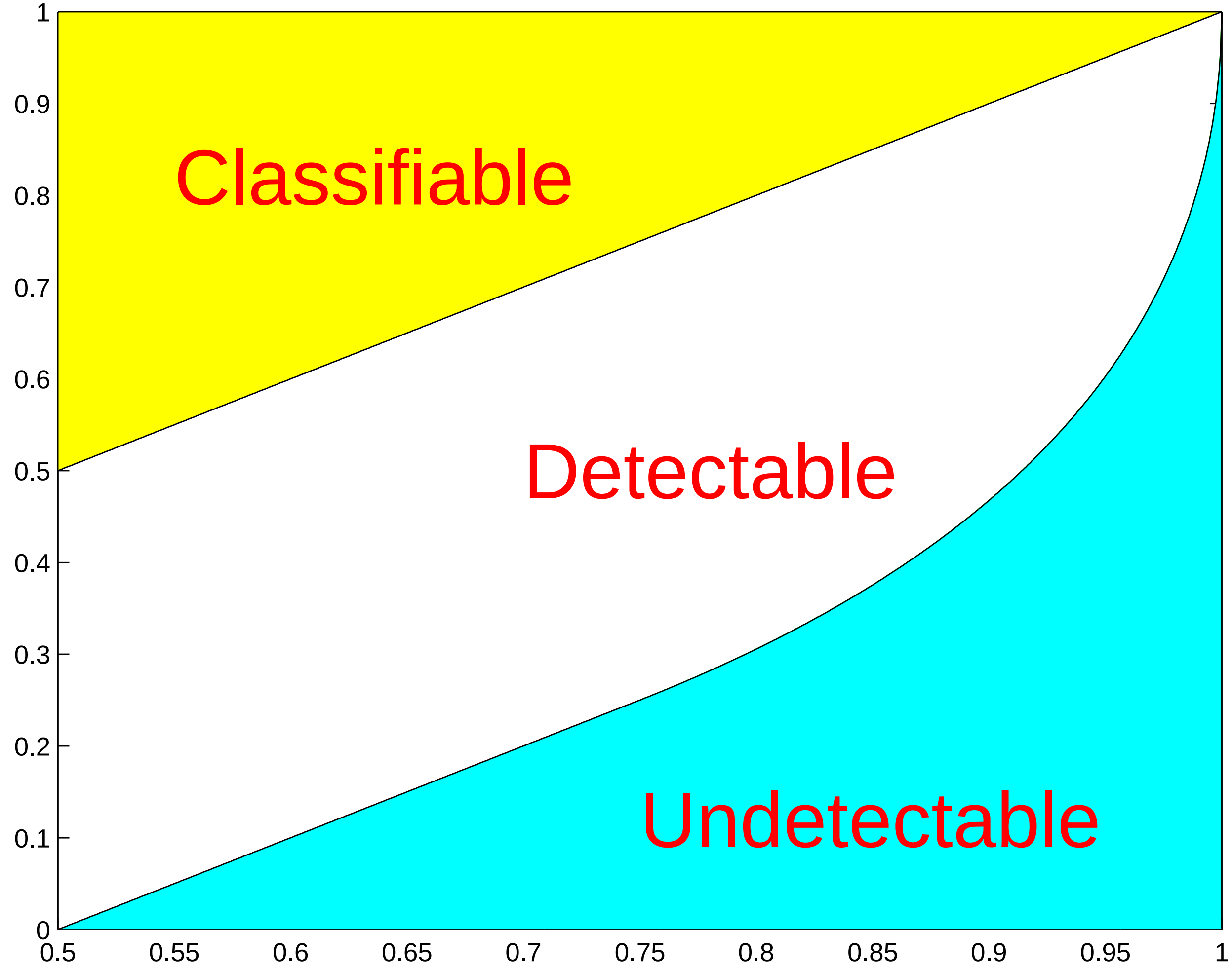
Theorem 1. (*Ingster 1999, Jin 2004*). If $\epsilon_n = n^{-\beta}$, $\mu_n = \sqrt{2r \log n}$, $\frac{1}{2} < \beta < 1$, and $0 < r < 1$, then:

$$\text{If } r > \rho(\beta), \quad H_0 \text{ and } H_1^{(n)} \text{ separate asymptotically,}$$

$$\text{If } r < \rho(\beta), \quad H_0 \text{ and } H_1^{(n)} \text{ merge asymptotically.}$$

We call $r = \rho(\beta)$ the "detection boundary":

$$\rho(\beta) = \begin{cases} \beta - \frac{1}{2}, & \frac{1}{2} < \beta < \frac{3}{4}, \\ (1 - \sqrt{1-\beta})^2, & \frac{3}{4} < \beta < 1. \end{cases}$$


Classifiable
Detectable
Undetectable
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1

Critical Value of Higher Criticism

- let $h(n, \alpha)$ be the critical value that

$$P\{HC^* > h(n, \alpha)\} \leq \alpha$$

- $h(n, \alpha) \approx \sqrt{2 \log \log n}, \quad \forall\, 0 < \alpha < 1$

Call $\alpha_n \to 0$ *slowly enough* if :

$$\frac{h(n, \alpha_n)}{\sqrt{2 \log \log n}} \quad \to \quad 1, \qquad n \to \infty.$$

Optimal Adaptivity of Higher Criticism

Theorem 2. (*Donoho and Jin 2004*). Consider the Higher Criticism test that rejects H_0 when

$$HC^* > h(n, \alpha_n)$$

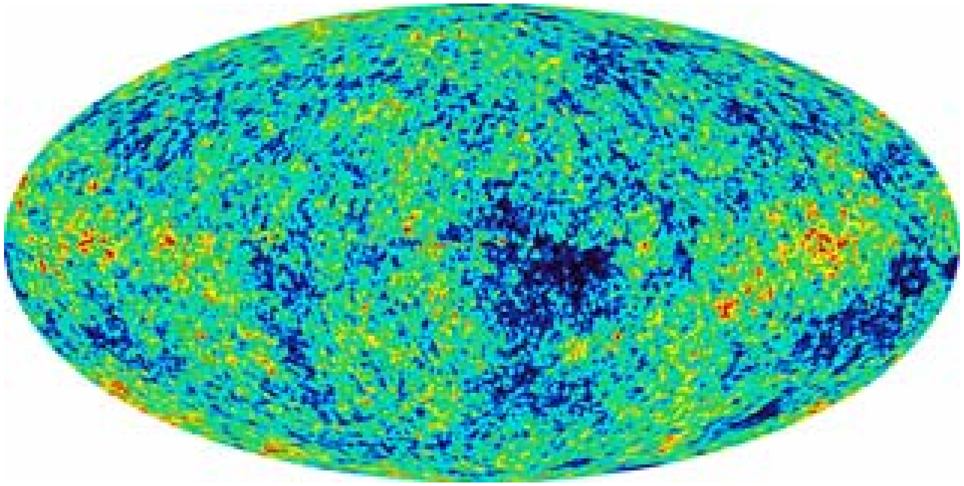
where the level $\alpha_n \to 0$ *slowly enough*. For every alternative $H_1^{(n)}(r, \beta)$ where r exceeds the detection boundary $\rho(\beta)$ — so that the Likelihood ratio test would have full power — Higher Criticism test also has full power:

$$P_{H_1^{(n)}}\{\text{Reject} \quad H_0\} \quad \to \quad 1.$$

Cosmic Microwave Background (CMB)

CMB:

- Oldest light in the universe, a direct link to early universe
- A relic of radiation when the universe $\approx 380,000$ years old
- An almost perfect black body at a temperature ≈ 2.725 Kelvin



Why study CMB

CMB provides a direct link to very early universe:

- Discriminate different models for early universe
- how does it evolve into the large scale galaxies today

From 1965 to 2003

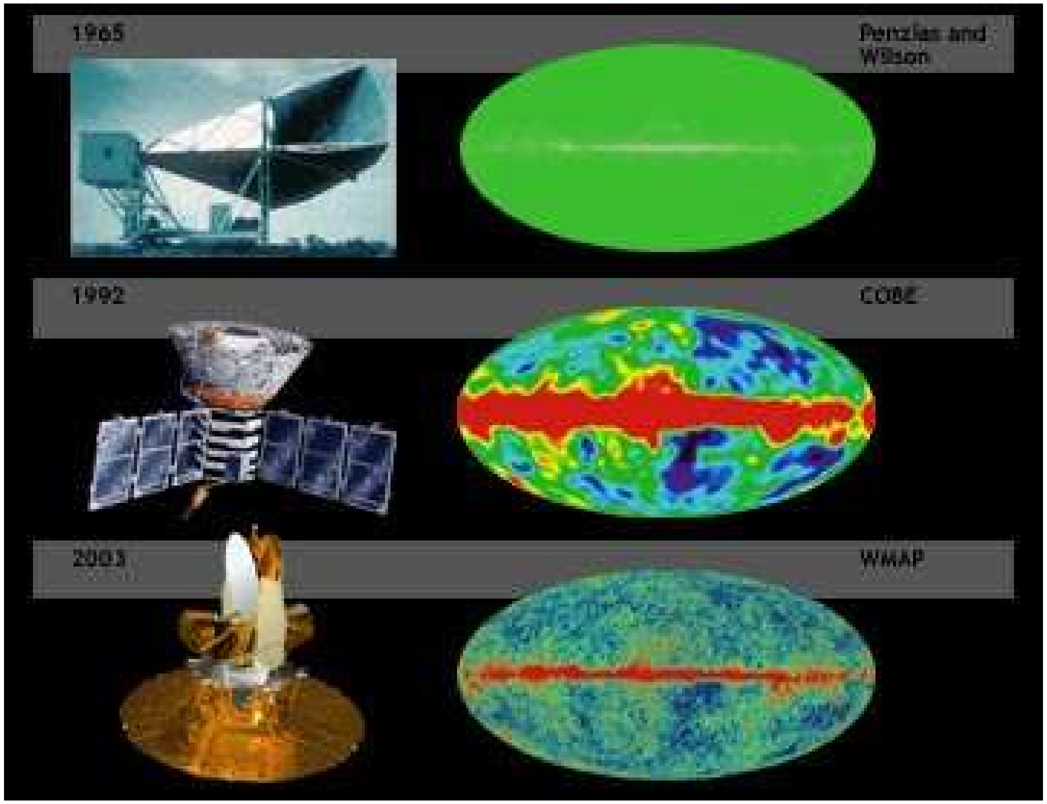


Figure 1: Small angular fluctuations in CMB are predicted as the imprints of initial densities perturbation which gave rise to large scale structures today. Red color: strong emission from the Milky way.

Wavelet Approach for nonGaussian Detection

- Standard inflation model predicts that the CMB is Gaussian
- Other models or secondary effects have nonGaussian signatures
- nonGaussian detection: disentangle different source of nonGaussianity from one to another
- Wavelet transform is a powerful tool for detect nonGaussian signature
  - isotropic à trous algorithm (Starck et al. 1998)
  - bi-orthogonal wavelet transform

For Today

- Consider n transform coefficients of CMB: X_i
- Test the hypothesis:

$$H_0: \qquad X_i \qquad \stackrel{iid}{\sim} \qquad N(0,1), \qquad 1 \leq i \leq n$$

Goal. By comparing different statistics:

- learn the strength and weakness of different tests
- look for the optimal tests in idealized cases

Wavelet Based nonGaussian Tests

1. Excess kurtosis (κ):

$$\kappa(X_1, \ldots, X_n) = \sum_i [X_i^4 - 3]$$

2. Maximum (Max):

$$\text{Max}(X_1, \ldots, X_n) = \max\{|X_1|, |X_2|, \ldots, |X_n|\}$$

3. Higher Criticism (HC)

Heuristic Comparison

A test **only sensitive** to certain type of nonGaussianity:

- κ: deviation of 4-th moment from Gaussian
- Max: unusual behavior of very large observations
- HC:
  - unusual behavior of extreme values
  - unusual behavior of moderately large values

Application I: Detecting Cosmic String

Cosmic string:

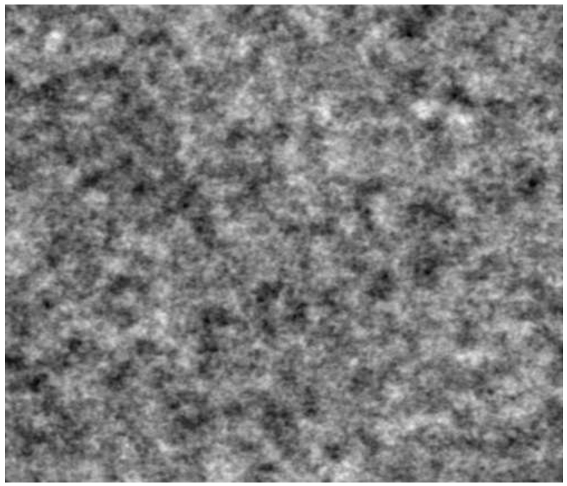
- an important source of nonGaussianity in CMB
- line-like object
- very old: formed within $\frac{1}{100}$ second after Big Bang
- very thin: 10^{-22} m
- very heavy: 10 km weights the same as earth

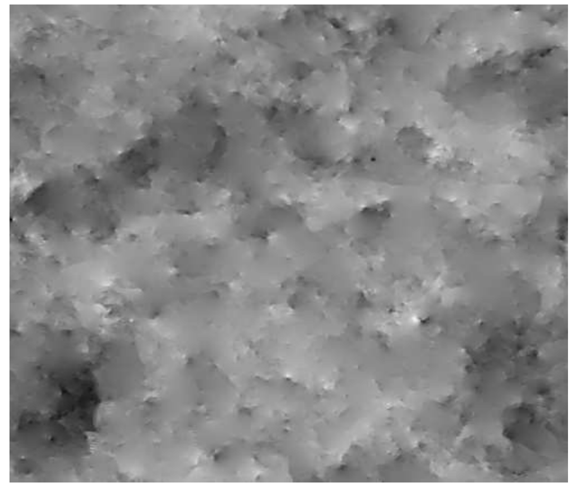
Why Look For Cosmic String

- most potential candidate for forming modern galaxies
- a direct link to very early universe
- not yet detected
- can not be produced in Lab (extremely high energy)

Goal: develop most sensitive detection tools

Detecting nonGassian Convolution Component

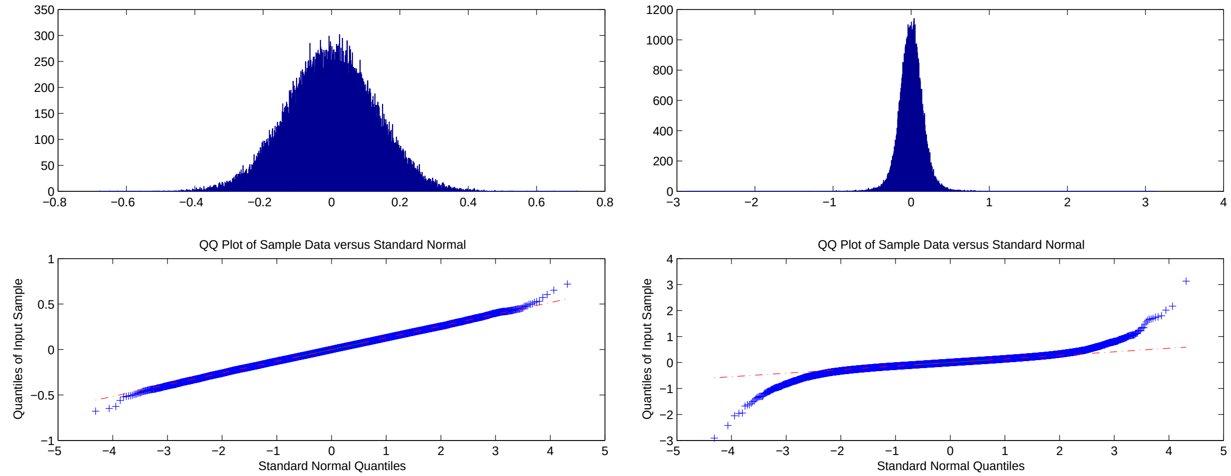




- given superposed image:

$$\sqrt{1-\lambda}\cdot \mathrm{CMB} + \sqrt{\lambda}\cdot \mathrm{CS}, \qquad \lambda \approx 0$$

- test: $\lambda = 0 \qquad \text{vs.} \qquad \lambda > 0$



- equivalent to test:

$$H_0: \; X_i = z_i, \qquad 1 \leq i \leq n,$$

$$H_1^{(n)}: \; X_i = \sqrt{1-\lambda} \cdot z_i + \sqrt{\lambda} \cdot w_i, \qquad 1 \leq i \leq n.$$

  – $z_i \stackrel{i.i.d}{\sim} N(0,1)$: wavelet coefficients of CMB
  – $w_i \stackrel{i.i.d}{\sim} W$: wavelet coefficients of CS
  – W unknown, but symmetric and **heavy tail**

Calibrations

Need careful calibrations for subtle analysis:

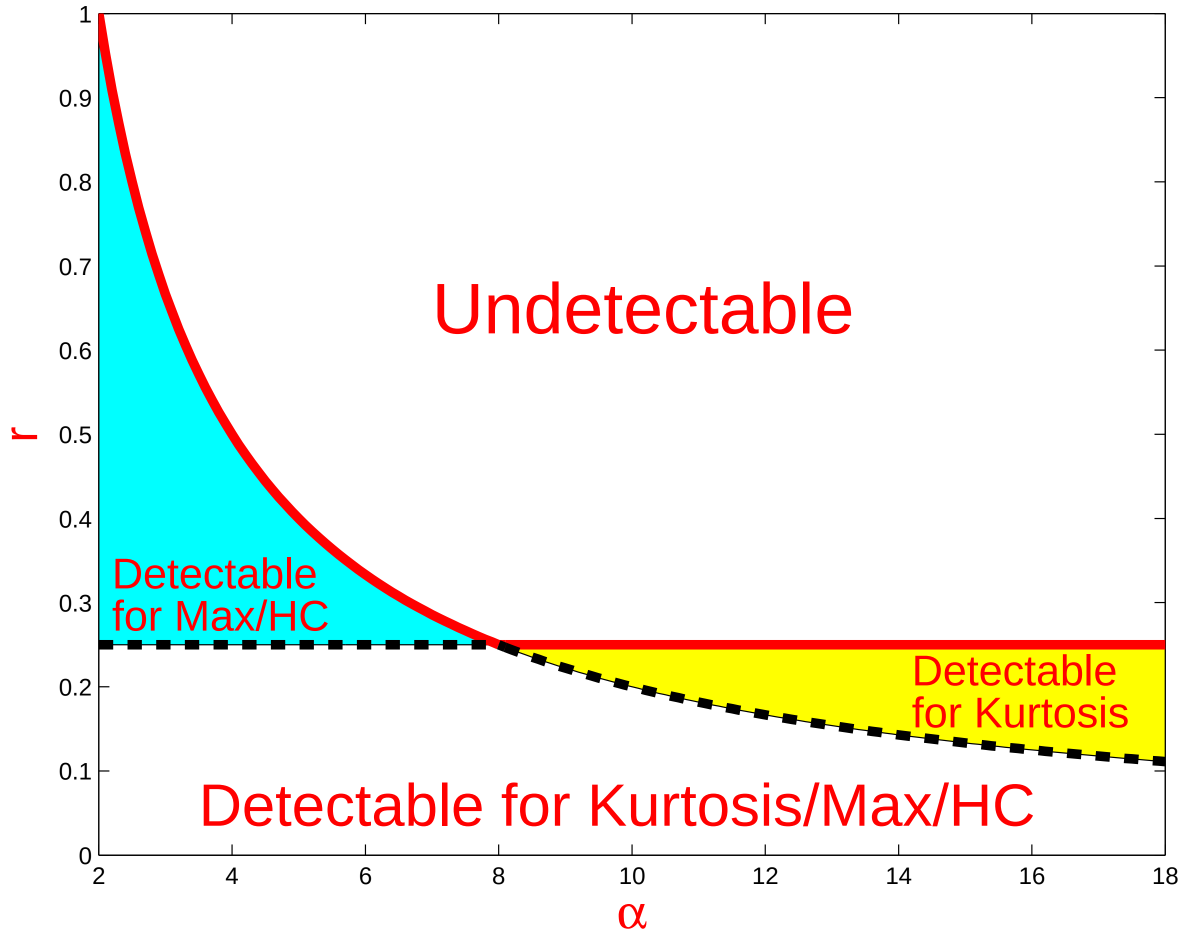
- increasing amount of data are offset by increasingly challenges:

$$\lambda = \lambda_n = n^{-r}, \qquad 0 < r < 1$$

- W: symmetric and has a **power-law** tail with index α:

$$\lim_{x \to \infty} x^{\alpha} p\{|W| > x\} = C_{\alpha}, \qquad C_{\alpha}: \quad \text{constant}$$

Question: Fixed (r, α) and let $n \to \infty$, what is the optimal test?


Undetectable
Detectable
for Max/HC
Detectable
for Kurtosis
Detectable for Kurtosis/Max/HC
r
α
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
4
6
8
10
12
14
16
18

Interpretation

$\alpha = 8$ is the separating line:

- $E[W^8] < \infty$: Kurtosis is better
  - W has a relatively thin tail, nonGaussianity affects the **bulk** of the data
  - best tests: tests based on **moments**
- $E[W^8] = \infty$: HC/Max is better
  - W has relatively heavy tail
  - nonGaussianity has **little** effect on the bulk of data, but large effect on extreme values and moderately large values
  - best tests: tests based on **data tails**

Estimating α

- Analysis supports the power-law tail assumption of W
- A classical estimator for α is the Hill's estimator (*Ann. Statist. 1975*)
- Implementation of Hill's estimator:

$$\hat{\alpha} \approx 6.1, \qquad \text{std}(\hat{\alpha}) \approx 0.9,$$

Large n

- The finer resolution of the image, the larger the n
- Need large n
  - to see the real advantage of HC
  - better answer whether $\alpha < 8$ or $\alpha > 8$
  - better answer which of HC and Kurtosis is better

Application II: WMAP First Year Data

http://map.gsfc.nasa.gov/

- WMAP radiometers observe at 5 frequency bands with one or more receivers: K (1), Ka (1), Q (2), V (2), W (4).
- WMAP team suggested use the weighted average of Q-V-W bands (8 receivers)
- Foreground cleaned
- Mask added: strong emission of Milky way etc..
- Downgraded from $n_{side} = 512$ to $n_{side} = 256$: measurement noise dominant in the smallest scale

Statistical Analysis

1. Generate $5,000$ simulated Gaussian maps of CMB.
2. For WMAP and each simulated map:
   - Use Spherical Mexican Hat Wavelets (SMHW): 2-D-spherical wavelets
   - Normalize the wavelet coefficients
   - Apply kurtosis, Max, and HC to wavelet coefficients

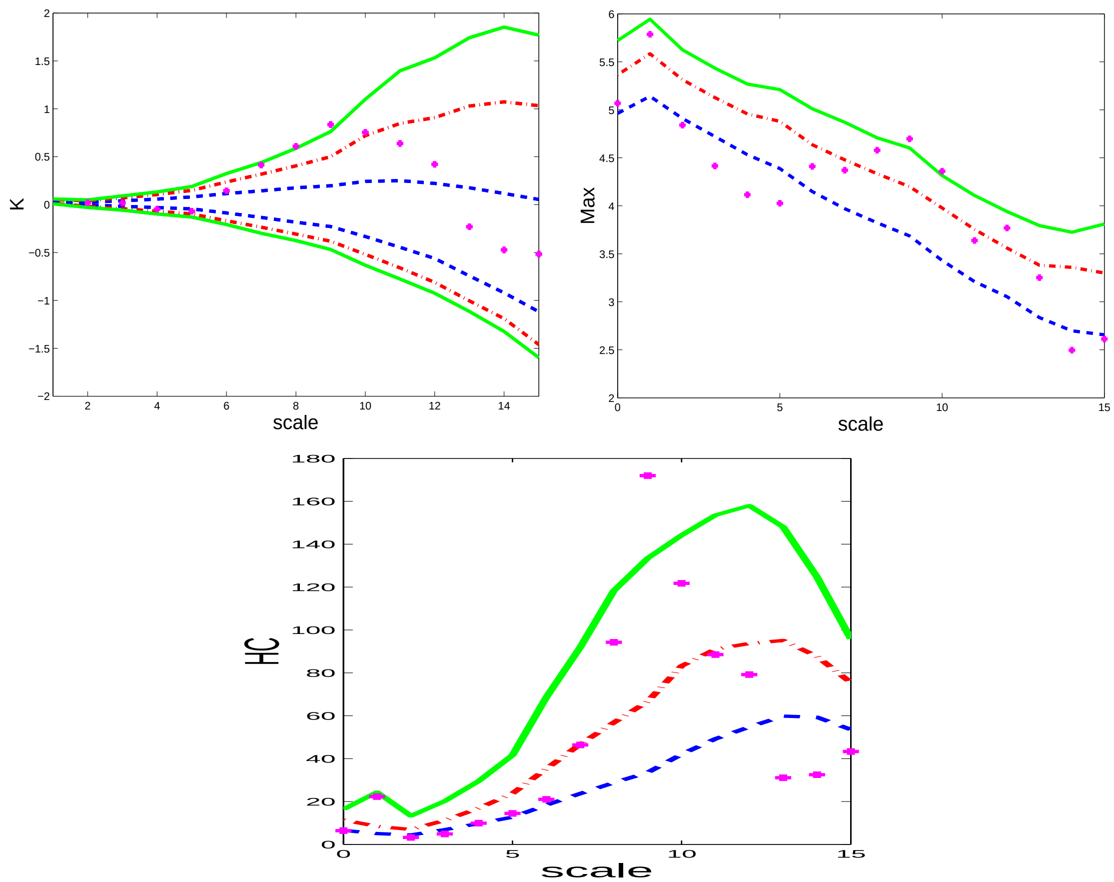


Figure 2: Test scores on WMAP and 67%, 95%, and 99% confidence regions on 5,000 simulated CMB maps.

Comparisons of Different Statistics

1. Almost equally powerful for detection, kurtosis is slightly better
   - Define empirical confidence of detection:

   $$\frac{\#\{\text{test scores based on simulations} \leq \text{score on WMAP}\}}{5000}$$

   - Kurtosis: 99.7%
   - HC: 99.46%
   - Max: 99.44%
2. Higher Criticism: automatically identify a tiny portion data as the source of nonGaussianity

Source of nonGaussianity

- $HC_n^* = \max_{0<\alpha<1}\{HC_{n,\alpha}\}$,

$$HC_{n,\alpha} = \sqrt{n} \cdot [(\text{Fraction Significant at } \alpha) - \alpha]/\sqrt{\alpha(1-\alpha)}$$

- $HC_{n,\alpha} \gg 1$ implies nonGaussiaintiy

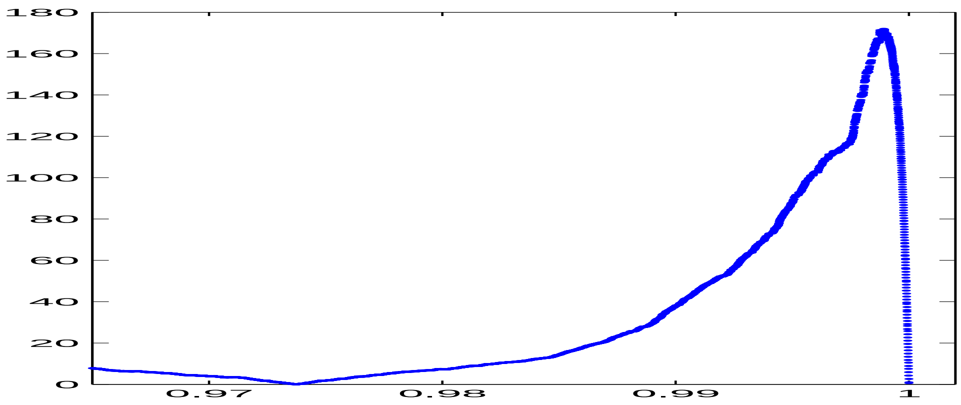


Figure 3: Plot of $HC_{n,\alpha}$ versus $(1-\alpha)$ for wavelet coefficients of WMAP at Scale 9

Figure 4: The selected coefficients maps back to pixels in a ring centered at (209^o, -57^o). We map each coefficient to only one pixel. This doesn't say only pixels over the ring is the source for nonGaussianity.

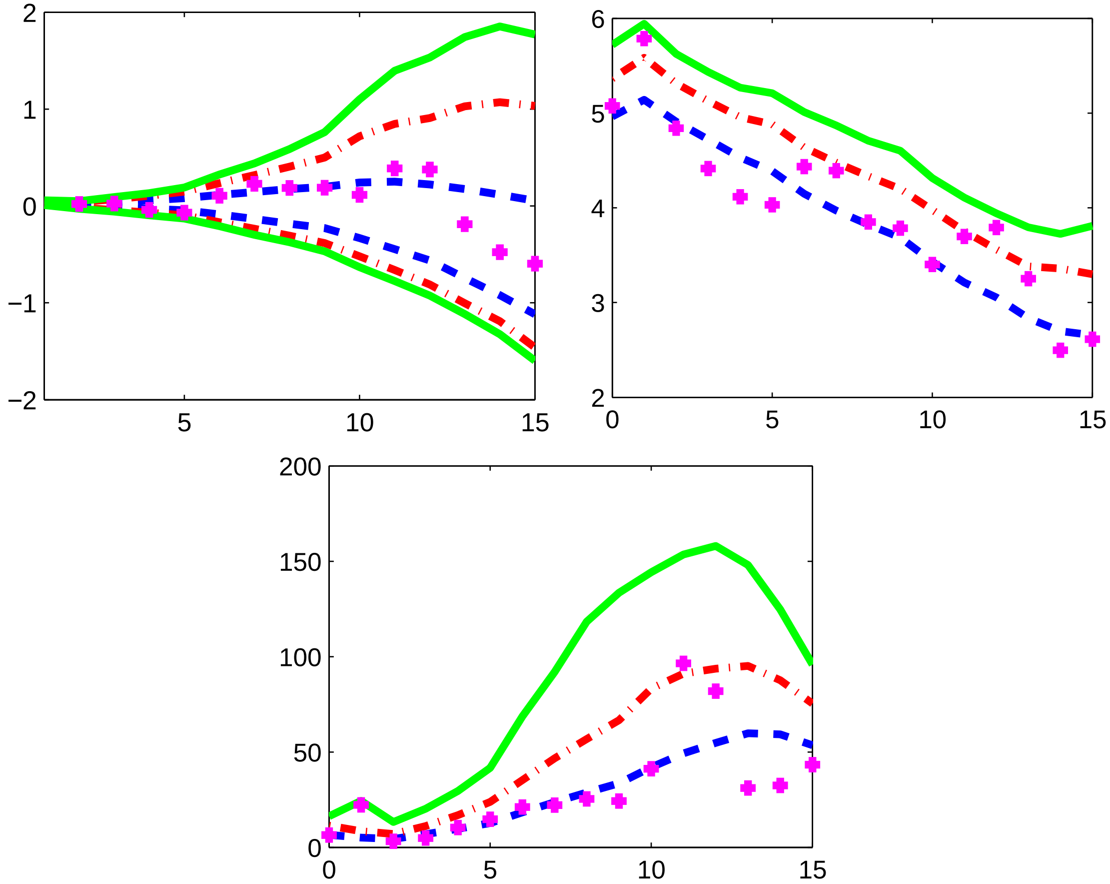


Figure 5: Kurtosis, Max, and HC after the ring removed from the WMAP. No detection of nonGaussianity at the level $\geq 90\%$.

Comparison to Other Works

- Some work has $\geq 99\%$ confidence of nonGaussian detection, and some work identify the cold spot centered at $(209^o, -57^o)$.
- Our contribution:
  - Add new statistics to nonGaussian detection: HC and Max
  - Almost equally powerful as kurtosis
  - HC offers automatical identification of a tiny portion of data as the source of nonGaussianity
  - The location of the ring coincide with the cold spot reported by Viela et al. 2004, Cruz et al. 2005

Take Home Messages

- nonGaussian detection in CMB is an exciting filed
- Higher Criticism is a promising new detection tool, adds more discussion to nonGaussian detection
- better answer is expected in future study with a larger n