Nonparametric Inference for The Cosmic Microwave Background

Christopher R. Genovese
Department of Statistics
Carnegie Mellon University
http://www.stat.cmu.edu/~genovese/

This work partially supported by NSF Grants ACI-0121671 and SES 9866147.

AstroStatistics: A Long History

- Hipparchus of Nicaea (190 BCE – 120 BCE)
  - Estimated an Earth-Moon distance in the range 59–67 Earth radii to be consistent with observations, accounting for uncertainty in Solar parallax measurement. (The actual value is about 60.3.) First "confidence interval"?
  - Tested his model with empirical observations.
  - Ptolemy (*Almagest* V.11) criticized Hipparchus for obtaining a range

- New technologies have turned what was once a trickle of astronomical data into a flood.

  These data can reveal more complex and subtle effects than ever, but they also demand new statistical methods.
- Many of the big questions in astronomy, astrophysics, and cosmology hinge on *statistical* issues.

WMAP: The Cosmic Microwave Background

We see the Cosmic Microwave Background (CMB) as a pattern of hot and cold spots on the sky.

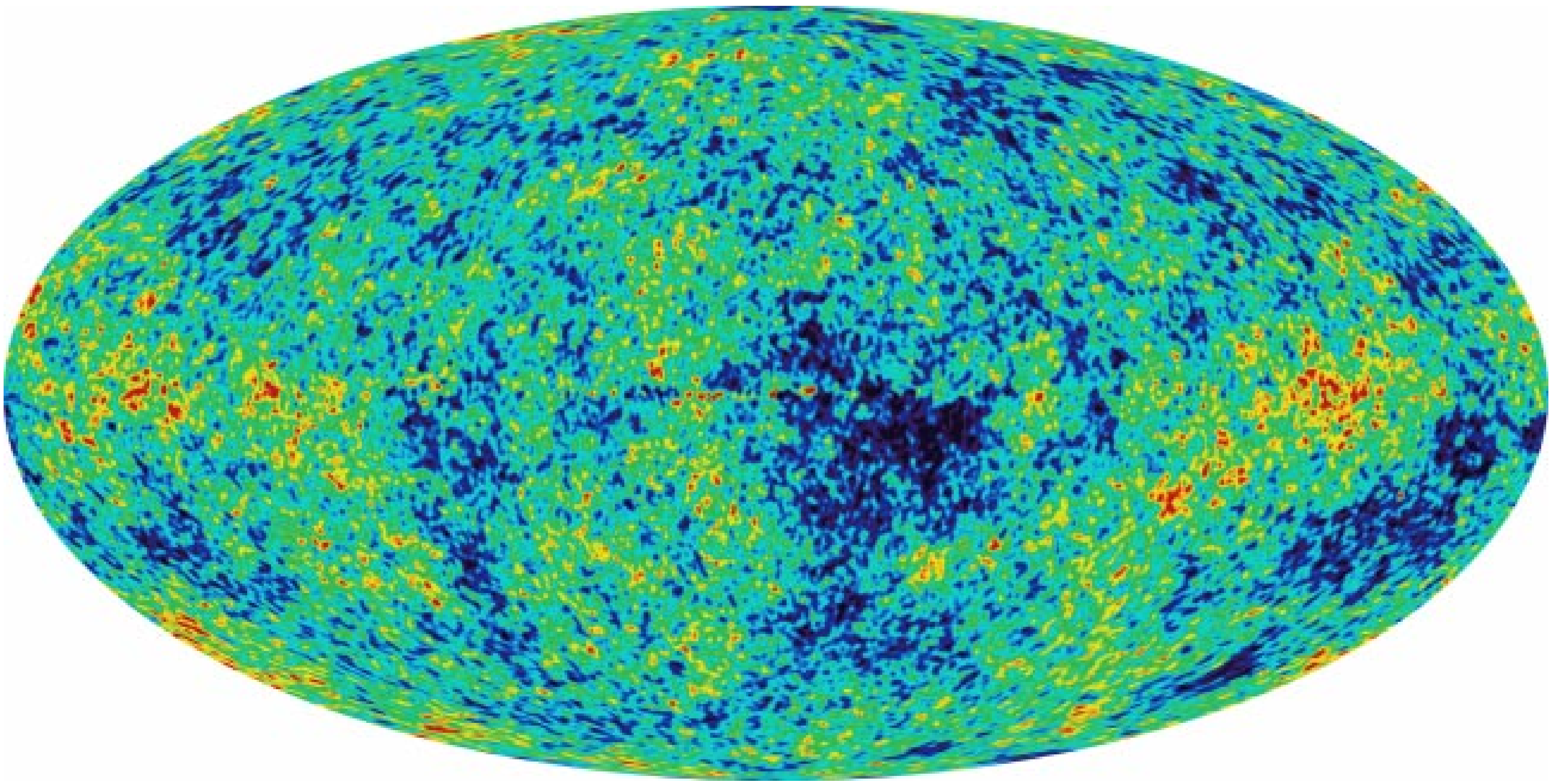


Image: NASA/WMAP Science Team

WMAP in the News

‘‘Breakthrough of the Year, 2003’’ — *Science*

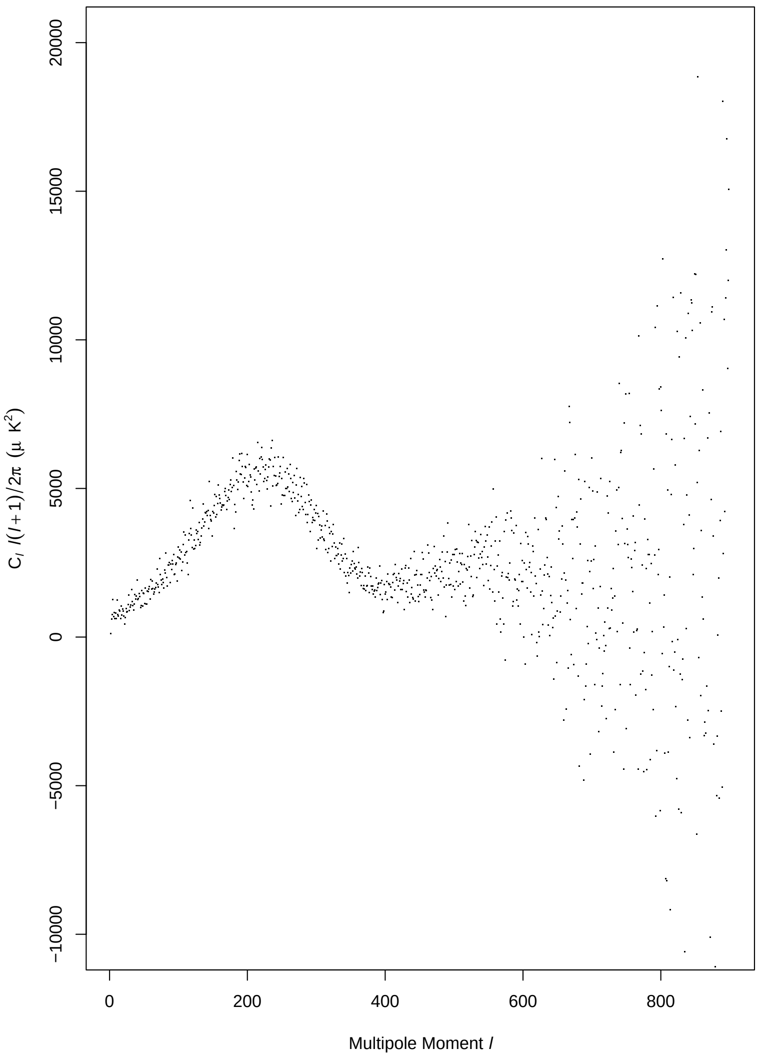
‘‘Most precise, detailed map yet produced of universe just after its birth ... confirms Big Bang theory’’ — *New York Times*

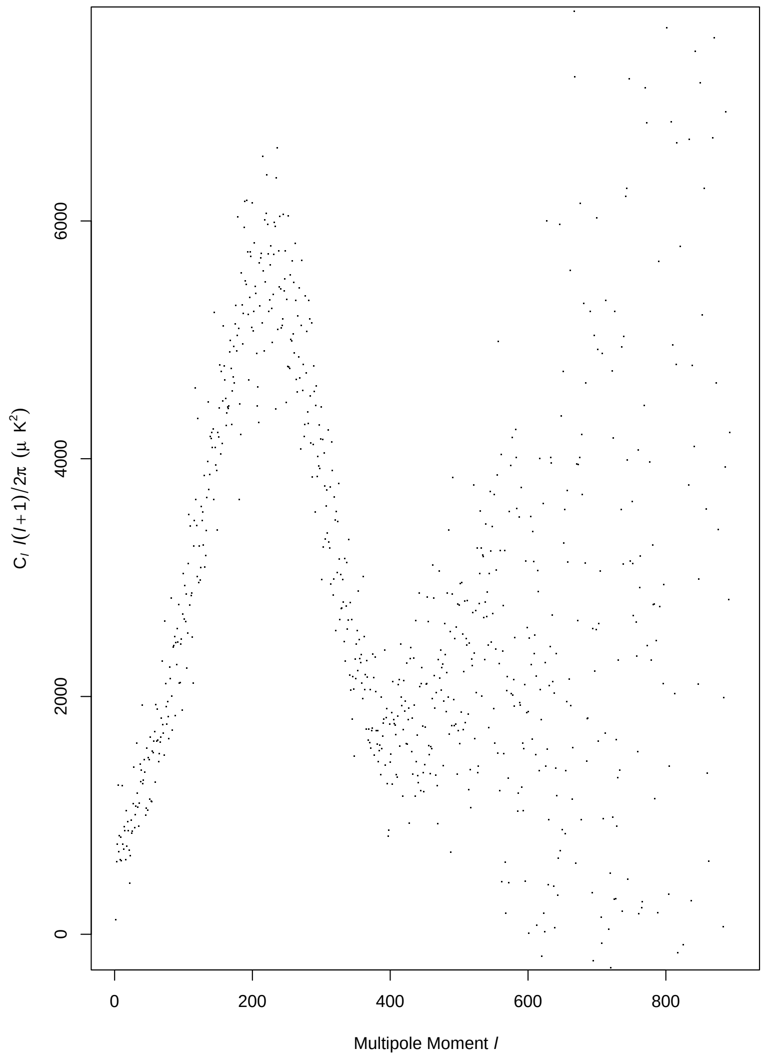
‘‘As of today we know better than ever when the universe began, how it behaved in its earliest instants, how it has evolved since then, and everything it contains.’’ — *Sky & Telescope*

‘‘The WMAP data pinpointed -- with unprecedented accuracy -- the universe’s age at 13.7 billion years; its flat shape; and its makeup of just 4 per cent "ordinary" matter, 23 per cent dark matter, and 73 per cent dark energy.’’ — *New Scientist*

‘‘I think every astronomer will remember where they were when they heard these results. ... I certainly will. This announcement represents a rite of passage for cosmology from speculation to precision science.’’ — John Bahcall, Princeton astrophycist in *Washington Post*

It's Just Regression After All





Pittsburgh AstroStatisics Team (PiCA)

Astrophysics

Robert Nichol

Chris Miller

Andrew Connolly

Peter Freeman

. . .

Statistics

Christopher Genovese

Larry Wasserman

Julie McIntyre

Alex Rojas

Computer Science

Andrew Moore

Jeff Schneider

Alex Gray

Road Map

1. Cosmology and the Cosmic Microwave Background

   – A Brief History

   – Can you hear the shape of the Universe?

2. Constructing Confidence Sets for Unknown Functions

   – Simultaneity, Bias, and Relevance

   – The Pivot-Ball Method (Beran and Dümbgen 1998)

   – Extensions and Alternatives

3. Case Study: The Cosmic Microwave Background

   – The WMAP Spectrum

   – Keeping Our Eyes on the Ball: Parametric Probes and Model Checking

Road Map

1. Cosmology and the Cosmic Microwave Background
   - A Brief History
   - Can you hear the shape of the Universe?
2. Constructing Confidence Sets for Unknown Functions
   - Simultaneity, Bias, and Relevance
   - The Pivot-Ball Method (Beran and Dümbgen 1998)
   - Extensions and Alternatives
3. Case Study: The Cosmic Microwave Background
   - The WMAP Spectrum
   - Keeping Our Eyes on the Ball: Parametric Probes and Model Checking

Physics of the Early Universe

The Big Bang model posits an expanding universe that began hot and dense.
A concise history starting 13.7 billion years ago:

- Temperature $\approx$ 1 trillion K (about 1 second)

  Density high enough to stop neutrinos
- Temperature $>$ 1 billion K (about 3 minutes)

  Atoms cannot form. Space filled with a stew of photons, baryons (e.g., protons and neutrons), electrons, neutrinos, and other matter.
- Temperature 12000 K

  Photons and baryons became coupled in a mathematically perfect fluid.
  Dark matter begins to clump under gravity. Acoustic waves propagate.
- Temperature 3000 K (about 380,000 years). "Recombination"

  Atoms form, photons are released.
- Temperature 2.7K (today). The Cosmic Microwave Background (CMB).

  Photons released at recombination observed in microwave band.
  Nearly uniform across the sky.

The Cosmic Microwave Background Today

Photons released at recombination seen as a pattern of hot and cold spots on the sky.

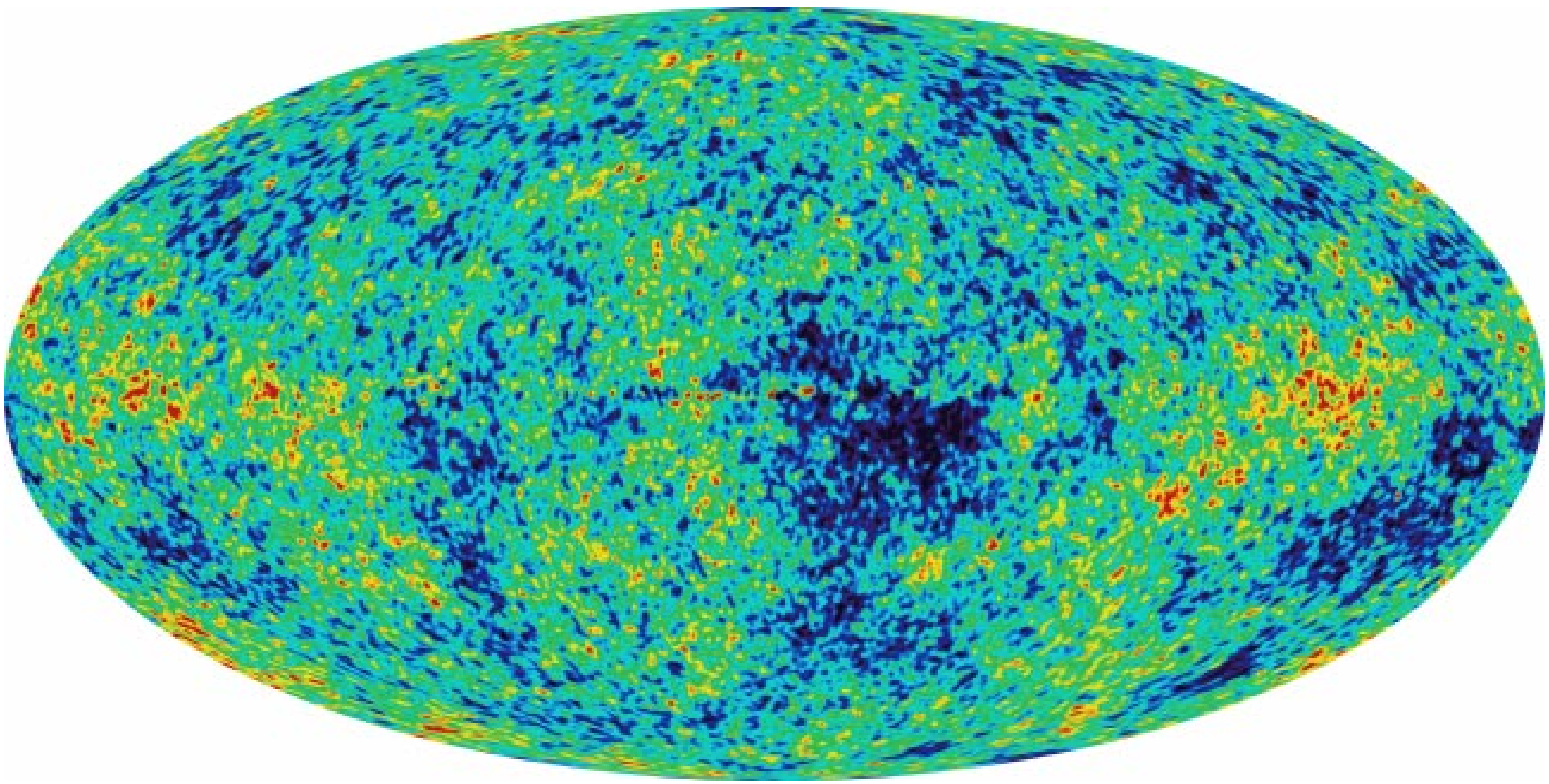


Image: NASA/WMAP Science Team

Illustration: The One Mode Universe

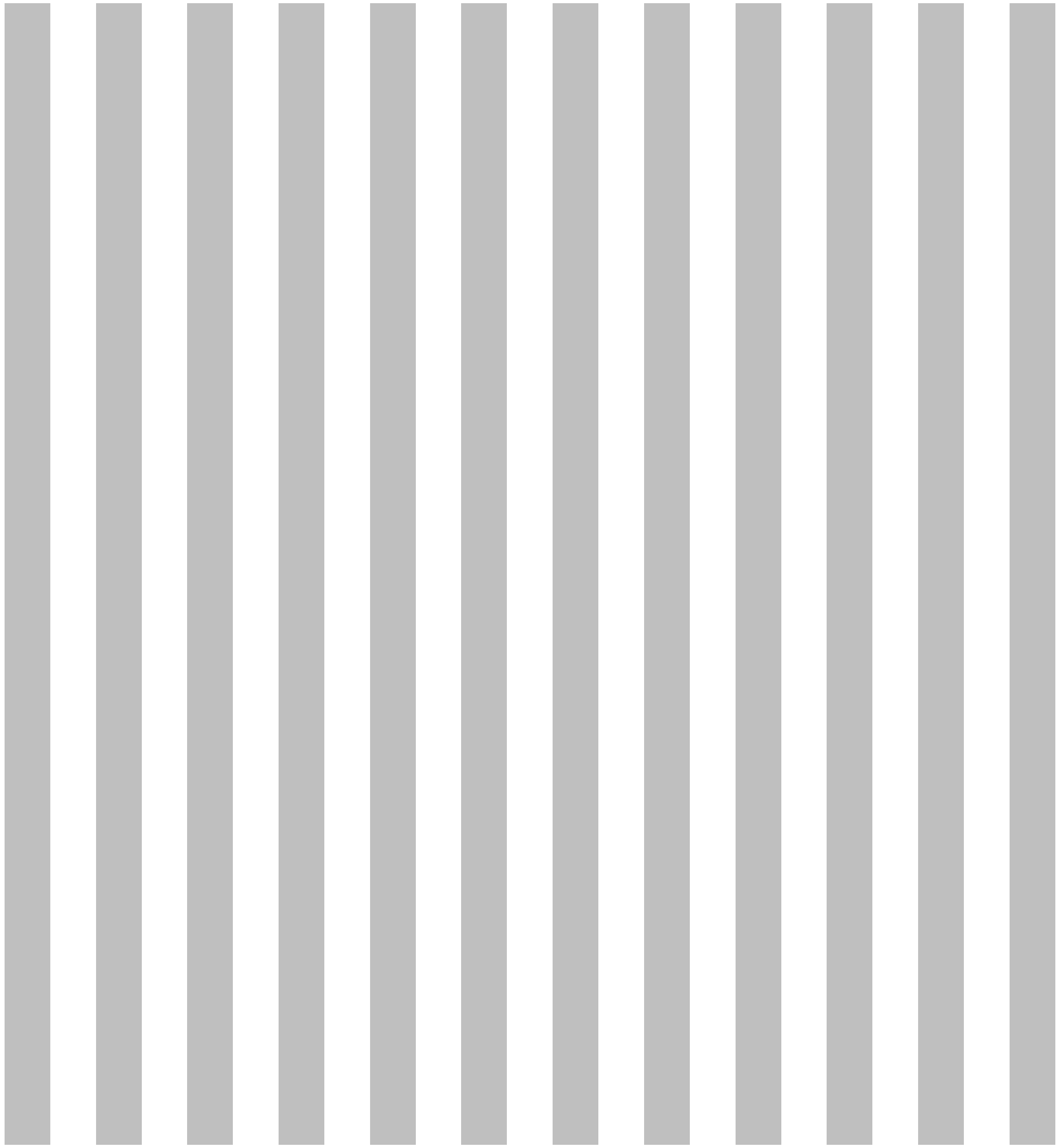


Illustration: The One Mode Universe (cont'd)

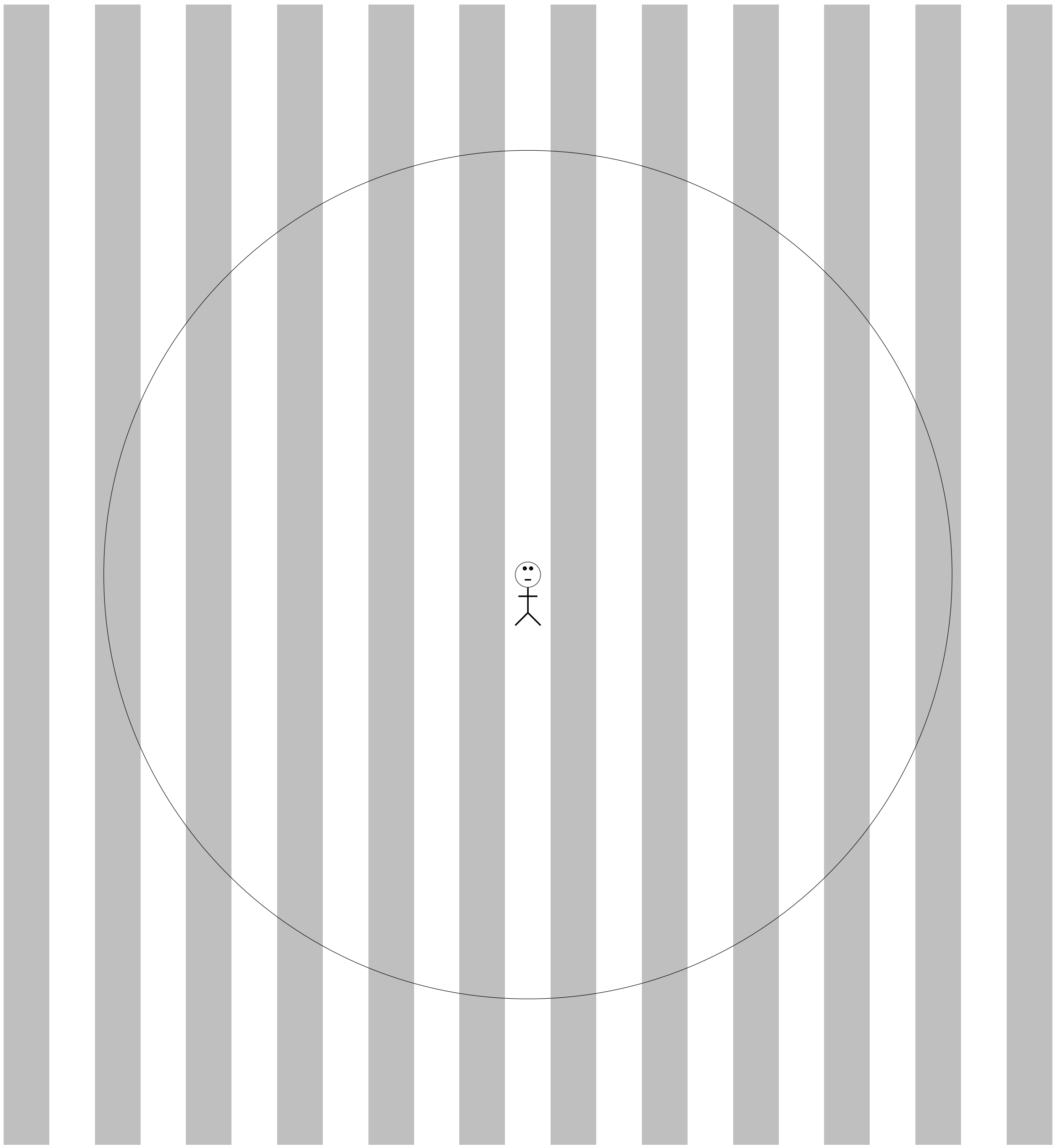
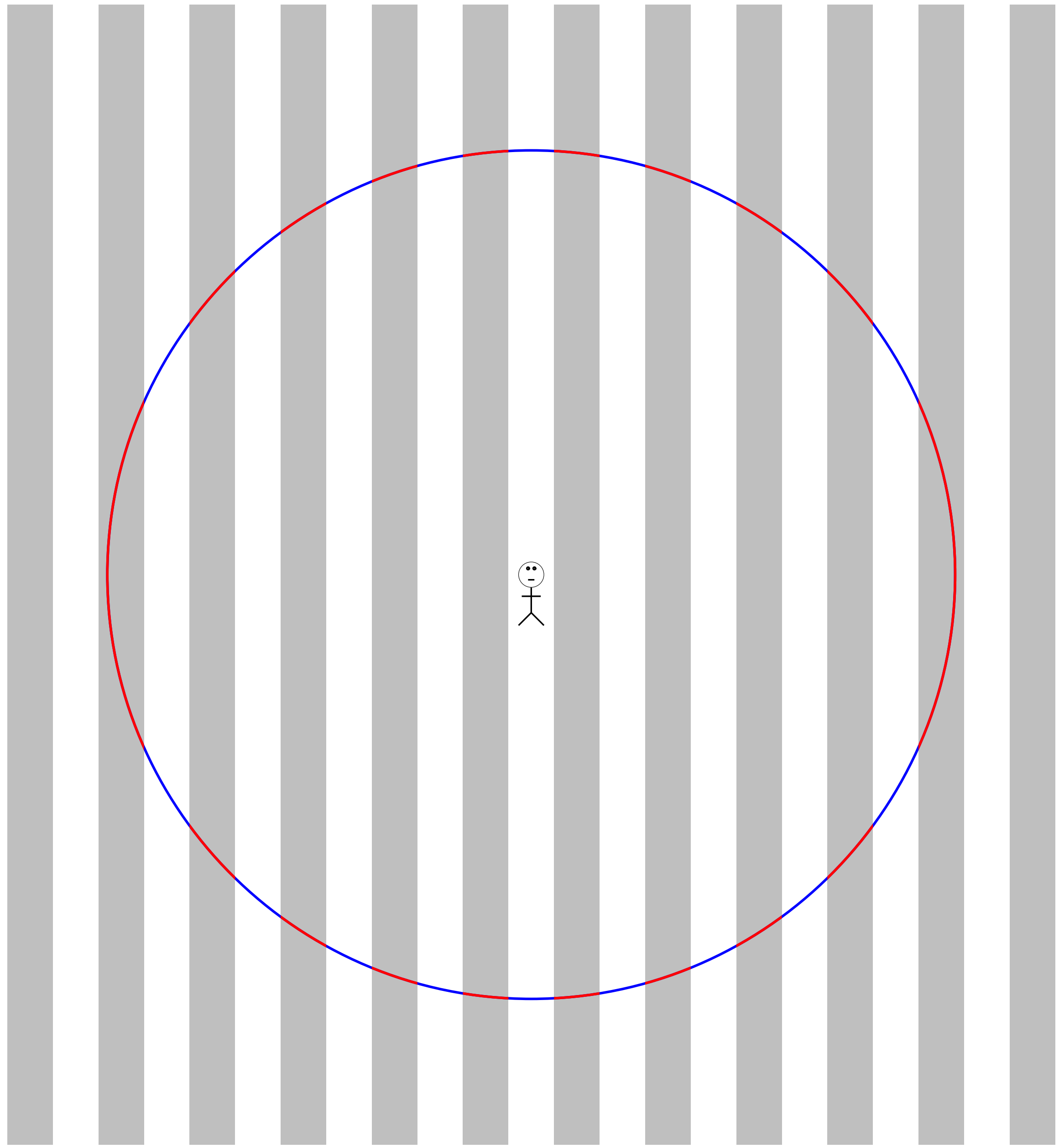


Illustration: The One Mode Universe (cont'd)

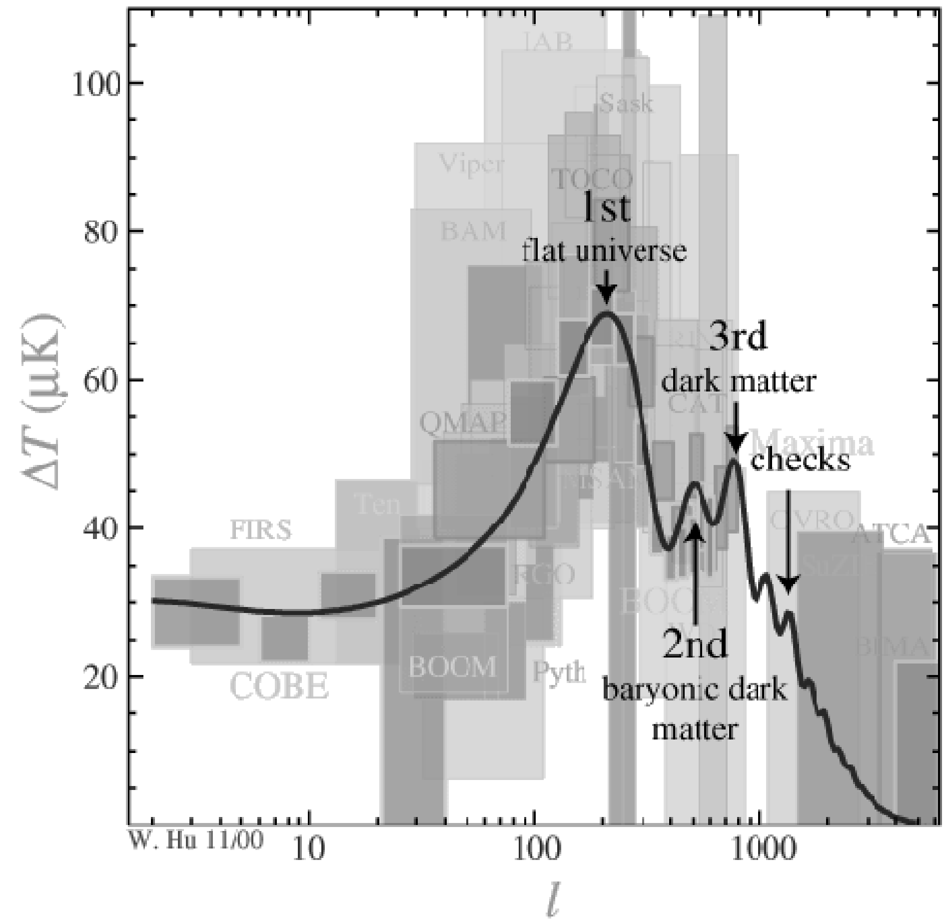


Can you hear the shape of the universe?

The acoustic oscillations before recombination carry information about the geometry and composition of the early universe.

Cosmologists decompose the sky map into an orthonormal basis (spherical harmonics).

The variance of the basis coefficients at each angular scale "ℓ" gives CMB temperature power spectrum $f(\ell)$.



The location, height, and width of the peaks in this spectrum relates to fundamental cosmological parameters, such as the Universe's total energy density (Ω_{total}) or the fraction of baryons (Ω_{baryon}).

Cosmological Models

Cosmologists use a finite-dimensional model (typically 6–13 parameters) for the power spectrum. It's common to incorporate other data to obtain the "Concordance model".

But the likelihood is complicated, multi-modal, and degenerate.

Other serious concerns with the model fits have been raised and are being actively investigated.

Bottom line:

– The assessment of uncertainty in the inferences has likely been optimistic.

– The underlying models still need checking.

What can Statistics offer?

Road Map

1. Cosmology and the Cosmic Microwave Background
   - A Brief History
   - Can you hear the shape of the Universe?

2. Constructing Confidence Sets for Unknown Functions
   - Simultaneity, Bias, and Relevance
   - The Pivot-Ball Method (Beran and Dümbgen 1998)
   - Extensions and Alternatives

3. Case Study: The Cosmic Microwave Background
   - The WMAP Spectrum
   - Keeping Our Eyes on the Ball: Parametric Probes and Model Checking

The Nonparametric Regression Problem

Observe data (x_i, Y_i) for $i = 1, \ldots, n$ where

$$Y_i = f(x_i) + \epsilon_i$$

Assume E$\boldsymbol{\epsilon} = 0$ and Var $\boldsymbol{\epsilon} = \Sigma$.

Leading case: $x_i = i/n$ and $\Sigma = \mathrm{diag}(\sigma^2, \ldots, \sigma^2)$ with σ^2 known.

Model f as belonging to some infinite dimensional space $\mathcal{F}$.
Example: Sobolev ball $\mathcal{F}_p(C) = \left\{ f \in \mathcal{L}^2\colon \int |f^{(p)}|^2 \leq C^2 \right\}$.

Usually specify a loss function over $\mathcal{F}$, such as $L(\widehat{f}, f) = \int (\widehat{f} - f)^2$, and define *risk* $R(\widehat{f}, f) = \mathsf{E}_f L(\widehat{f}, f)$.

Rate-Optimal Estimators

Finding good estimators in this problem hinges on deciding how much (and how to) smooth.

Optimize rate of convergence r_n for minimax risk over $\mathcal{F}$:

$$\inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}} R(\widehat{f}_n, f) \asymp r_n,$$

Want an estimator that achieves this rate.

For infinite-dimensional $\mathcal{F}$, necessarily have $r_n\sqrt{n} \to \infty$.

For example, $r_n = n^{-\frac{2p}{2p+1}}$ on $\mathcal{F}_p$.

Rate-optimal estimators exist. In fact, *adaptive* estimators exist.

For example, can find $\widehat{f}_n \to f$ in $\mathcal{F}_p$ at rate $n^{-2p/2p+1}$ *without knowing* p.

Inference about the Unknown Function

But we usually need more than $\widehat{f}$.

We want to make inferences about features of f: shape, magnitude, peaks, inclusion, derivatives.

Would like to contruct a $1-\alpha$ confidence set for f, a set $\mathcal{C}$ such that $\mathsf{P}\Big\{\mathcal{C} \ni f\Big\} = 1-\alpha$.

Typically, $\mathcal{C}$ is the set of functions within a confidence band over all (or a finite set of) points in the domain.

Three challenges:

1. Bias
2. Simultaneity
3. Relevance

Bias

In nonparametric problems, rate-optimal tuning parameter gives $\text{bias}^2 \approx \text{var}$.

Loosely, if $\widetilde{f} = \mathsf{E}\widehat{f}$ and $s = \sqrt{\mathsf{Var}\,\widehat{f}}$, then

$$\frac{\widehat{f} - f}{s} = \frac{\widehat{f} - \widetilde{f}}{s} + \frac{\widetilde{f} - f}{s} \approx \mathsf{N}(0,1) + \frac{\text{bias}}{\sqrt{\text{var}}}.$$

So, "$\widehat{f} \pm 2s$" undercovers.

Two common solutions in the literature:

– Bias Correction: Shift confidence set by estimated bias.

Often increases variance more than it reduces bias.

– Undersmoothing: Smooth so that var dominates bias^2.

Requires additional calibration; typically non-uniform coverage.

Simultaneity

We observe f on a finite set of points $x_1, \ldots, x_n$ but often want to extend inferences to the whole object.

Require additional assumptions to constrain f between design points.

For confidence bands, one solution is the "volume of tubes" formula (Sun and Loader 1994).

If $\widehat{f}(x) = \sum_{i=1}^{n} \ell_i(x) Y_i$, then for a suitable class $\mathcal{F}$,

$$\inf_{f \in \mathcal{F}} \mathsf{P}\Big\{ \widehat{f}(x) - c\widehat{\sigma}\, \|\ell(x)\| \leq f(x) \leq \widehat{f}(x) + c\widehat{\sigma}\, \|\ell(x)\|, \forall x \Big\} = 1 - \alpha,$$

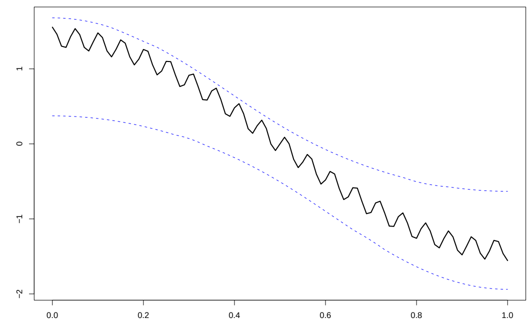
where c solves $\alpha = K\phi(c) + 2(1 - \Phi(c))$ and the constant K depends on $T(x) = \ell(x)/\|\ell(x)\|$.

Still must account for bias.

Relevance

In small samples, confidence balls and bands need not constrain all features of interest.

For example, number of peaks:



Alternative: confidence intervals for specific functionals of f

Two practical problems:

1. Many relevant functionals (e.g., peak locations) hard to work with.
2. One often ends up choosing functionals post-hoc.

We prefer to obtain construct a confidence set for the whole object with post-hoc protection for inferences about many functionals.

An Inferential Strategy

- Build confidence sets for the whole function f that obtain either finite-sample coverage or uniform asymptotic coverage ($\sup_{f \in \mathcal{F}} \left| \mathsf{P}\left\{ \mathcal{C}_n \ni f \right\} - (1-\alpha) \right| \to 0$).
- Constrain the model post hoc and search the confidence set to make inferences about any features of interest.

This enables model checking and comparison, follow up analyses, and complicated functionals, all with post-hoc protection.

The confidence set induces a map from a set of assumptions to a set of inferences.

Two good approaches based on the Pivot-Ball (Beran and Dümbgen 1998) and Subspace Pretesting (Baraud 2004) methods.

Pivot-Ball Confidence Sets

Expand unknown f in orthonormal basis (ϕ_j): $f = \sum_j \theta_j \phi_j$.

Method yields confidence sets $\mathcal{C}_n$ satisfying

$$\sup_{f \in \mathcal{F}} \left| \mathsf{P}\left\{ \mathcal{C}_n \ni f_n \right\} - (1-\alpha) \right| \to 0,$$

where f_n is projection onto first n coefficients.

With extra assumptions, can dilate $\mathcal{C}_n$ to cover f similarly.

Define

- Coefficient estimator $\widehat{\theta}_j(\lambda)$ for (possibly vector-valued) tuning parameter λ.
- Loss function $L_n(\lambda) = \sum_{j=1}^{n} (\widehat{\theta}_j(\lambda) - \theta_j)^2$.
- Unbiased risk estimate $S_n(\lambda)$.

Pivot-Ball Method

1. Choose $\widehat{\lambda}_n$ to minimize $S_n(\lambda)$.
2. Show that for an appropriate statistic $\widehat{\tau}_n$,

$$\frac{\sqrt{n}(L_n(\widehat{\lambda}_n) - S_n(\widehat{\lambda}_n))}{\widehat{\tau}_n} \rightsquigarrow \mathsf{N}(0,1).$$

3. Conclude that $\mathcal{D}_n$ is an asymptotic $1-\alpha$ confidence set for $\boldsymbol{\theta}$:

$$\mathcal{D}_n = \left\{ \theta\colon \sum_{\ell=1}^{n} (\widehat{\theta}_n(\widehat{\lambda}_n) - \theta_\ell)^2 \le \frac{\widehat{\tau}_n\, z_\alpha}{\sqrt{n}} + S_n(\widehat{\lambda}_n) \right\}.$$

4. Hence $\mathcal{C}_n = \left\{ f_n\colon \int (f_n - \widehat{f}_n)^2 \le \frac{\widehat{\tau}_n z_\alpha}{\sqrt{n}} + S_n(\widehat{\lambda}_n) \right\}$ has desired properties.

Extensions and Alternatives

- Extensions of the pivot ball method
  - Weighted-Loss and non-constant variance (Genovese et al. 2004), as with the CMB spectrum.
  - Wavelet bases (Genovese and Wasserman 2003).
  - Density estimation (Jang, Genovese, and Wasserman 2004).
  - Using better risk estimates allow limited adaptation (Van der Vaart and Robins, 2004)
- Alternatives: Finite-sample and uniform asymptotic confidence *bands* (Baraud 2004, Genovese and Wasserman 2004).

Road Map

1. Cosmology and the Cosmic Microwave Background
   - A Brief History
   - Can you hear the shape of the Universe?

2. Constructing Confidence Sets for Unknown Functions
   - Simultaneity, Bias, and Relevance
   - The Pivot-Ball Method (Beran and Dümbgen 1998)
   - Extensions and Alternatives

3. Case Study: The Cosmic Microwave Background
   - The WMAP Spectrum
   - Keeping Our Eyes on the Ball: Parametric Probes and Model Checking

Inference for the CMB Power Spectrum

After correcting for our motion and other artifacts, expand sky temperature map into spherical harmonics

$$T = \sum_{\ell=1}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}.$$

The spectrum C_ℓ is the variance of the $a_{\ell m}$s for each ℓ.

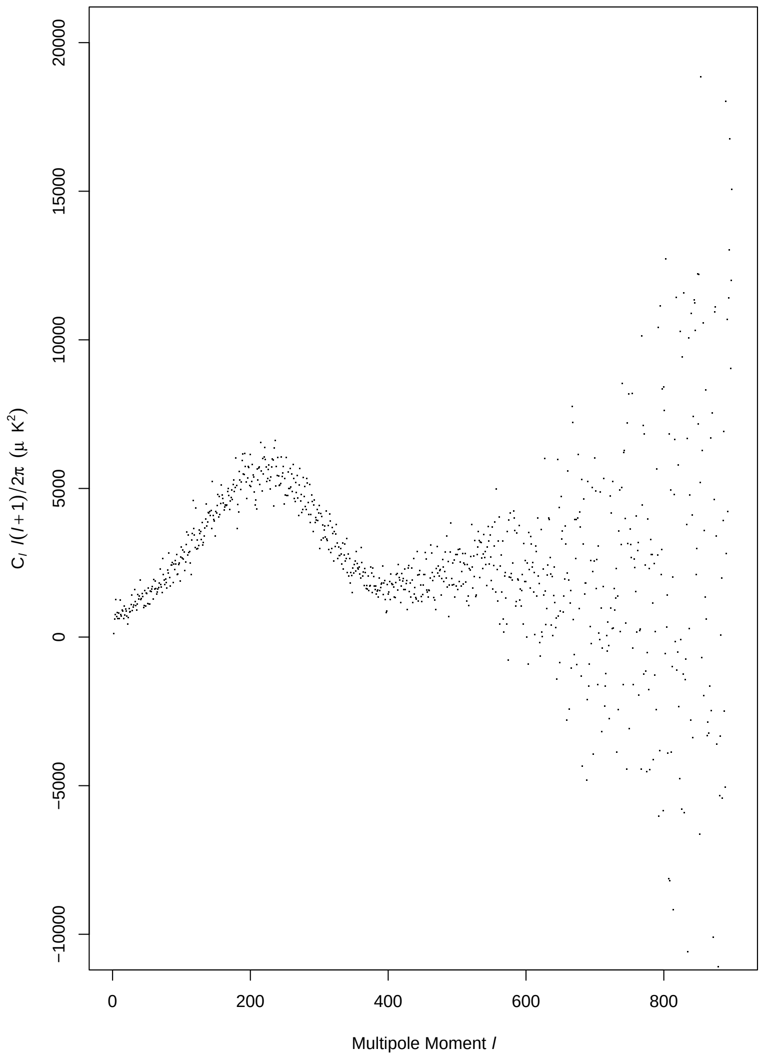
From the data, get noisy version of C_ℓ, which can be cast as a nonparametric regression problem.

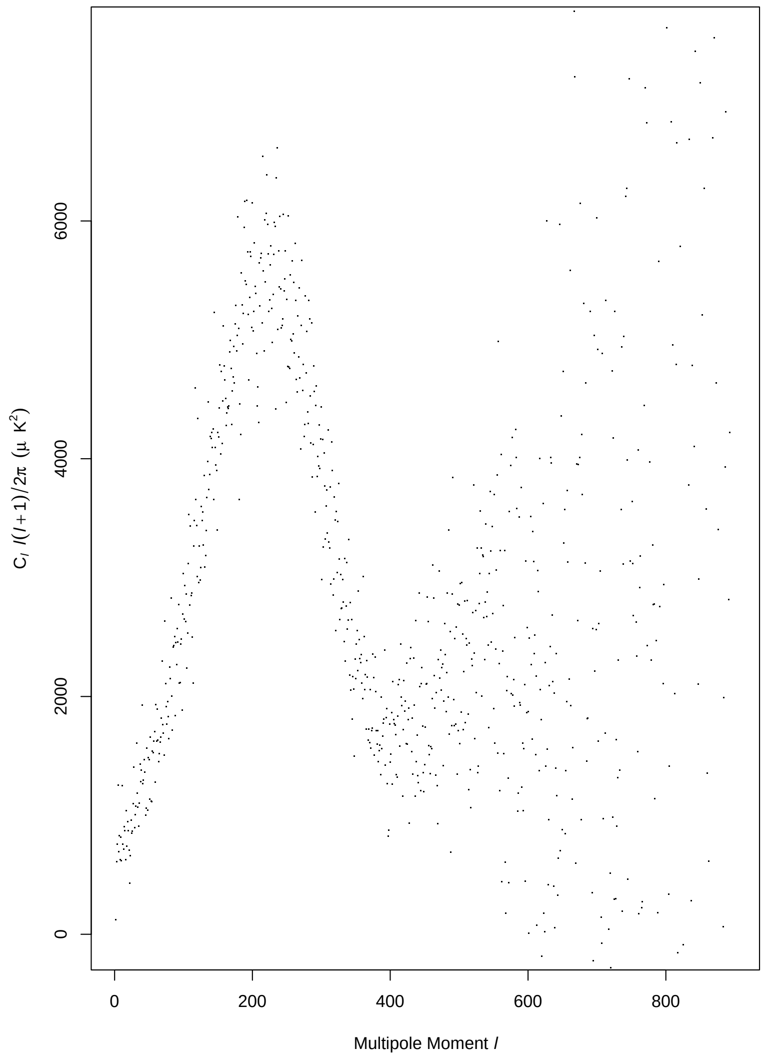
Apply the pivot-ball method, generalized for non-constant variance.

Constrain $\mathcal{F}$ to functions with bounded number of local extremes.

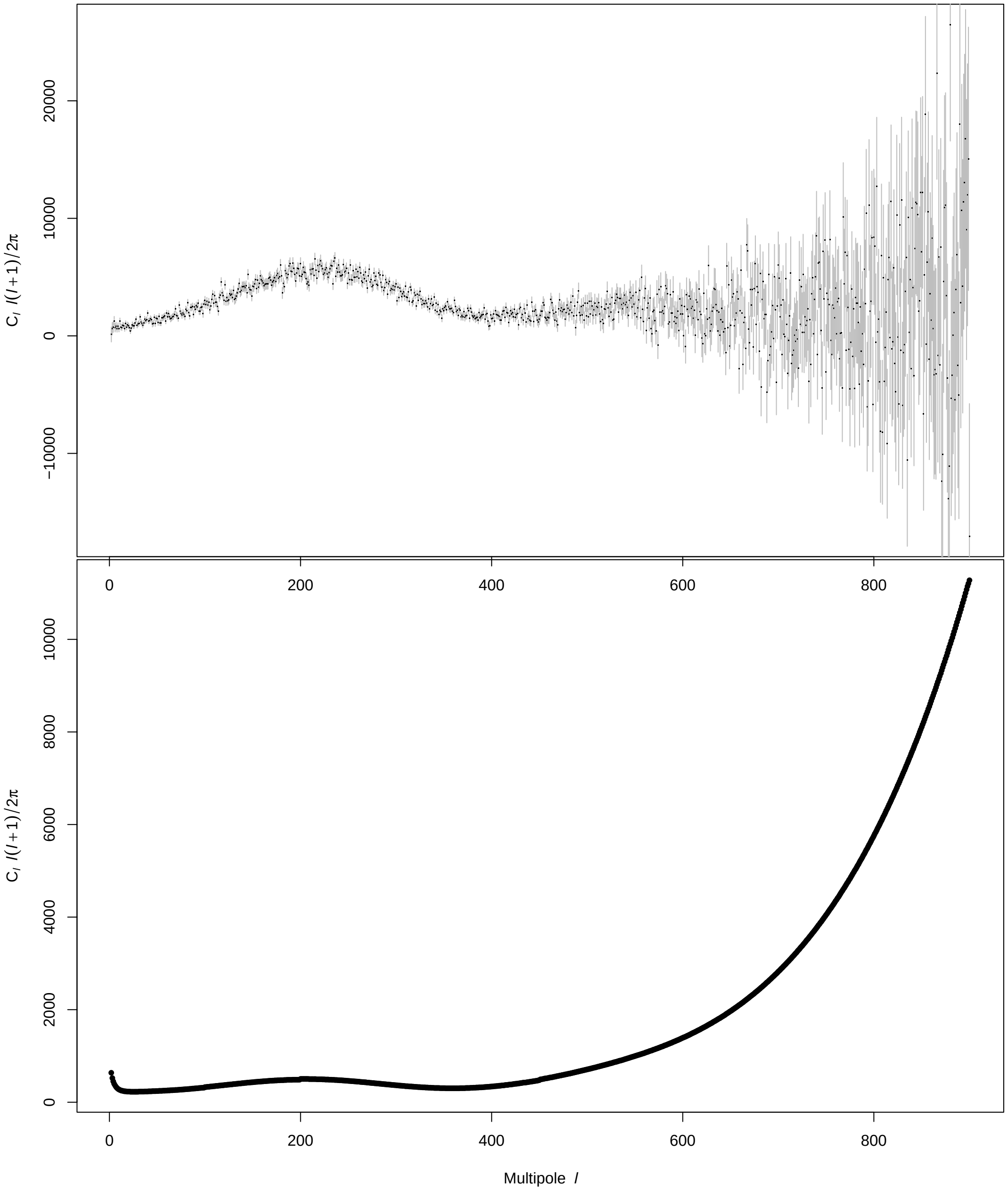
Obtain confidence set $\mathcal{C}_n$ for the unknown spectrum; use this to check the models and make inferences about the peaks.

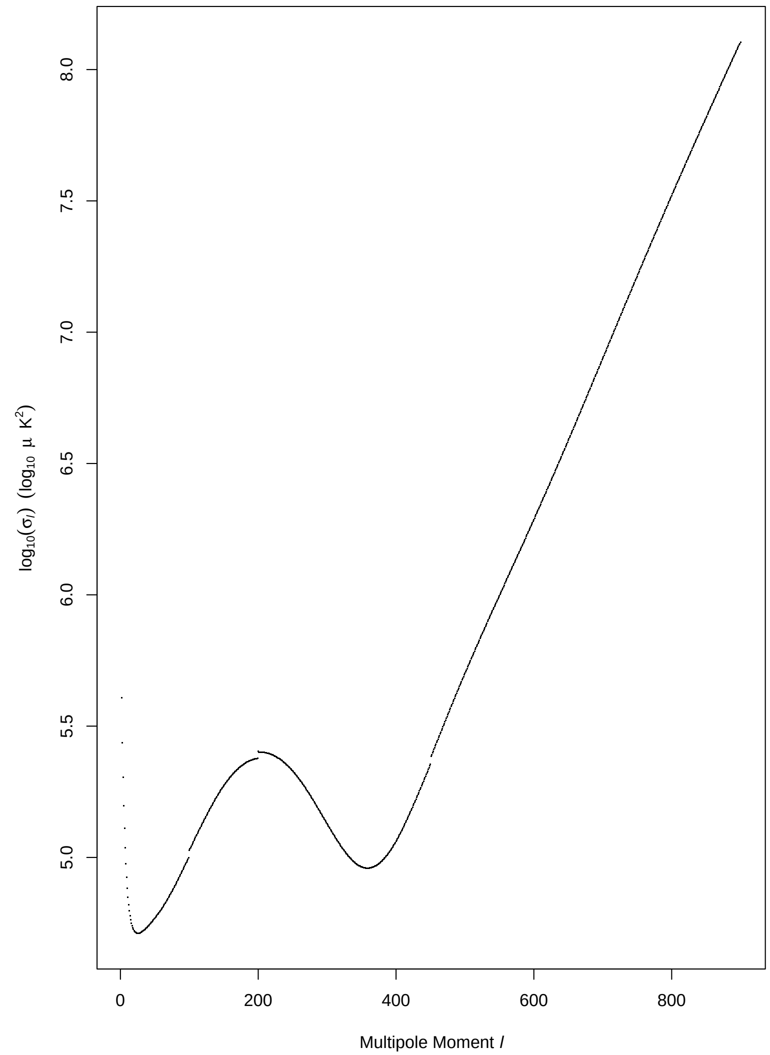
CMB Power Spectrum: WMAP Data





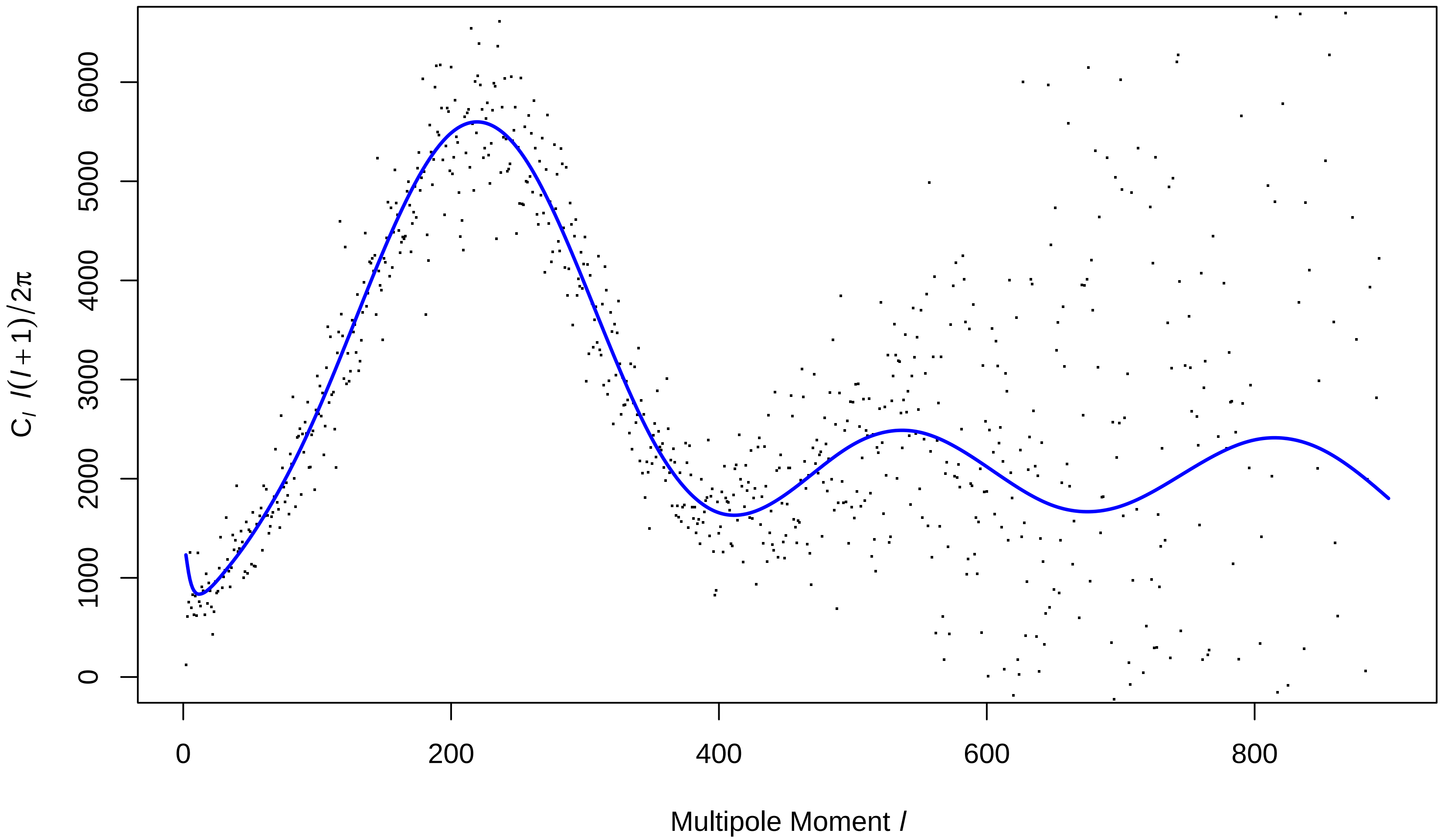
CMB Power Spectrum: WMAP Variances





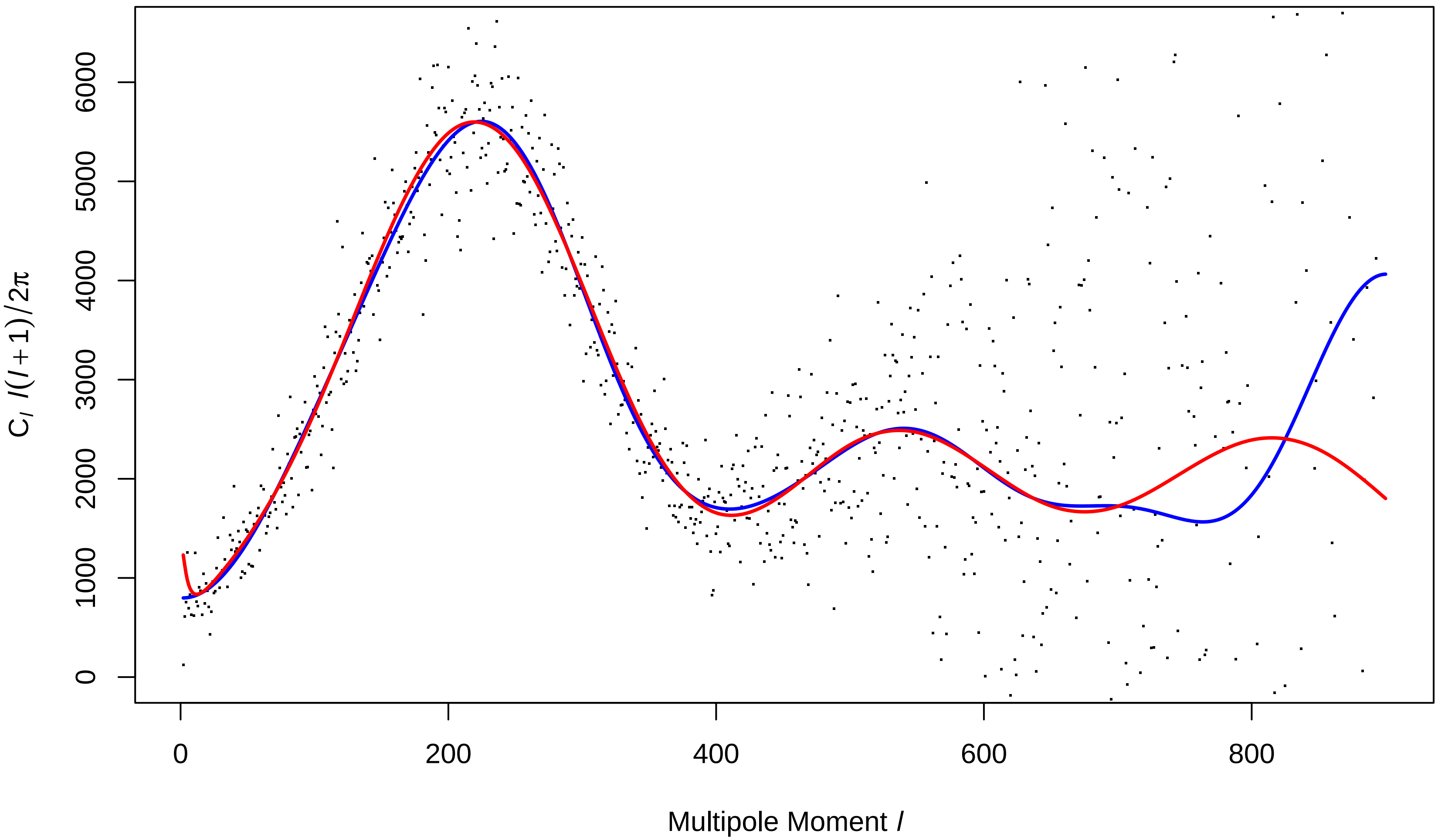
Noise correlated and heteroskedastic

Cosmological Models



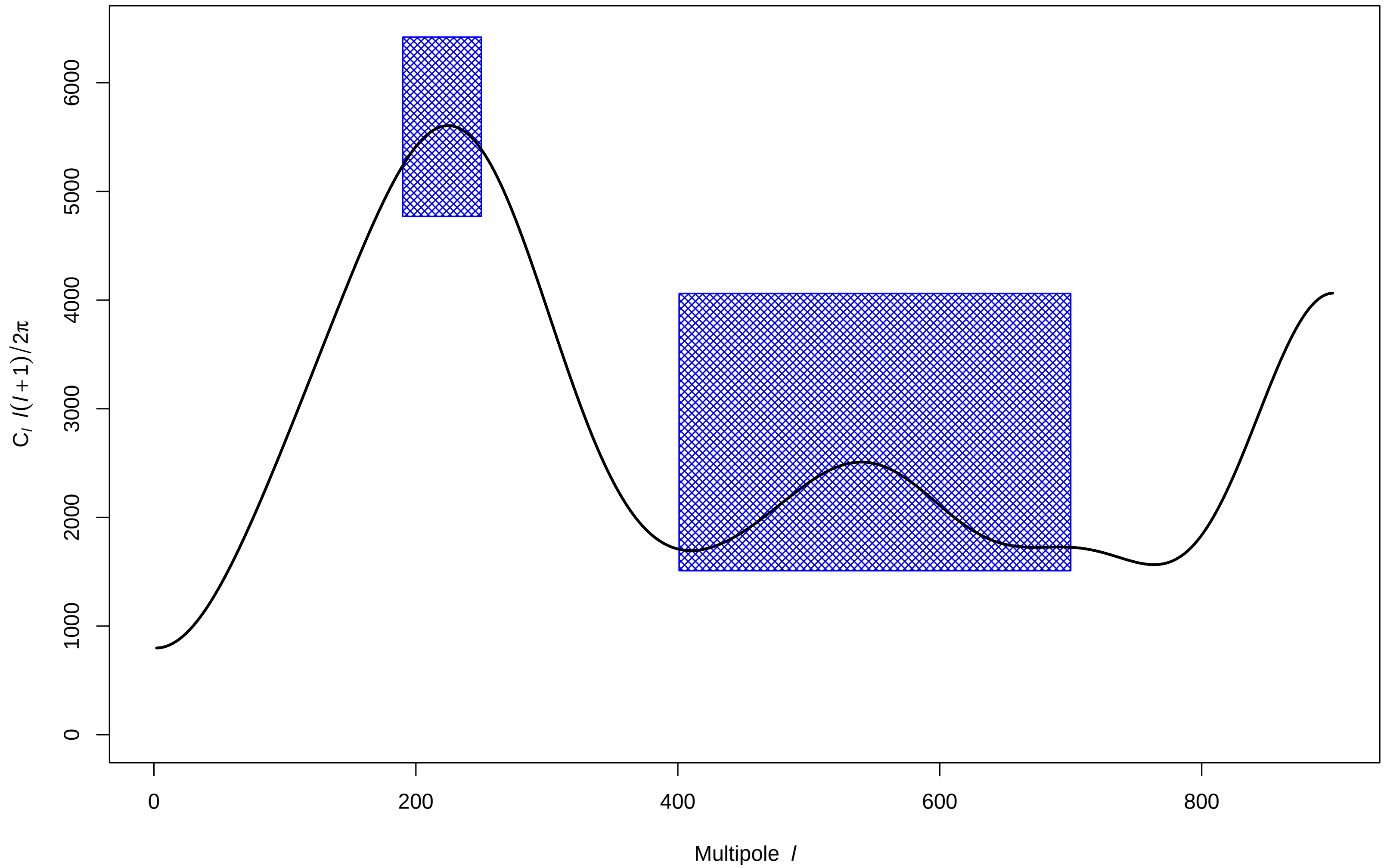
- 6–13 dimensional model maps cosmological parameters to spectra.
- Ultimate goal: inferences about basic cosmological quantities.
- Subsidiary goal: identify location, height, widths of peaks

Confidence Ball Center vs Concordance Model



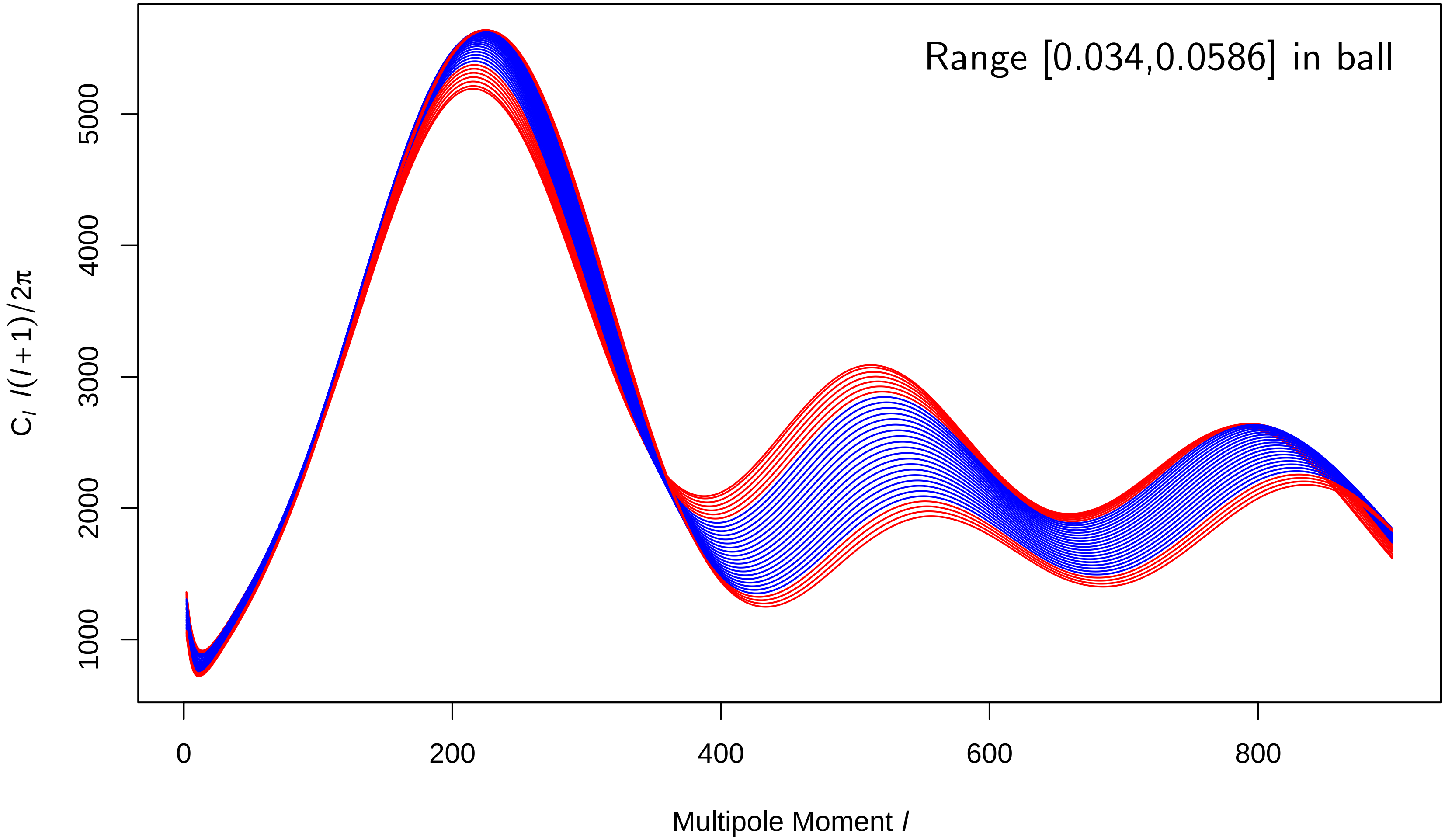
Eyes on the Ball I: Parametric Probes

Peak Heights, Peak Locations, Ratios of Peak Heights



Eyes on the Ball I: Parametric Probes (cont'd)

Varied baryon fraction in CMBFAST keeping $\Omega_{\rm total} = 1$

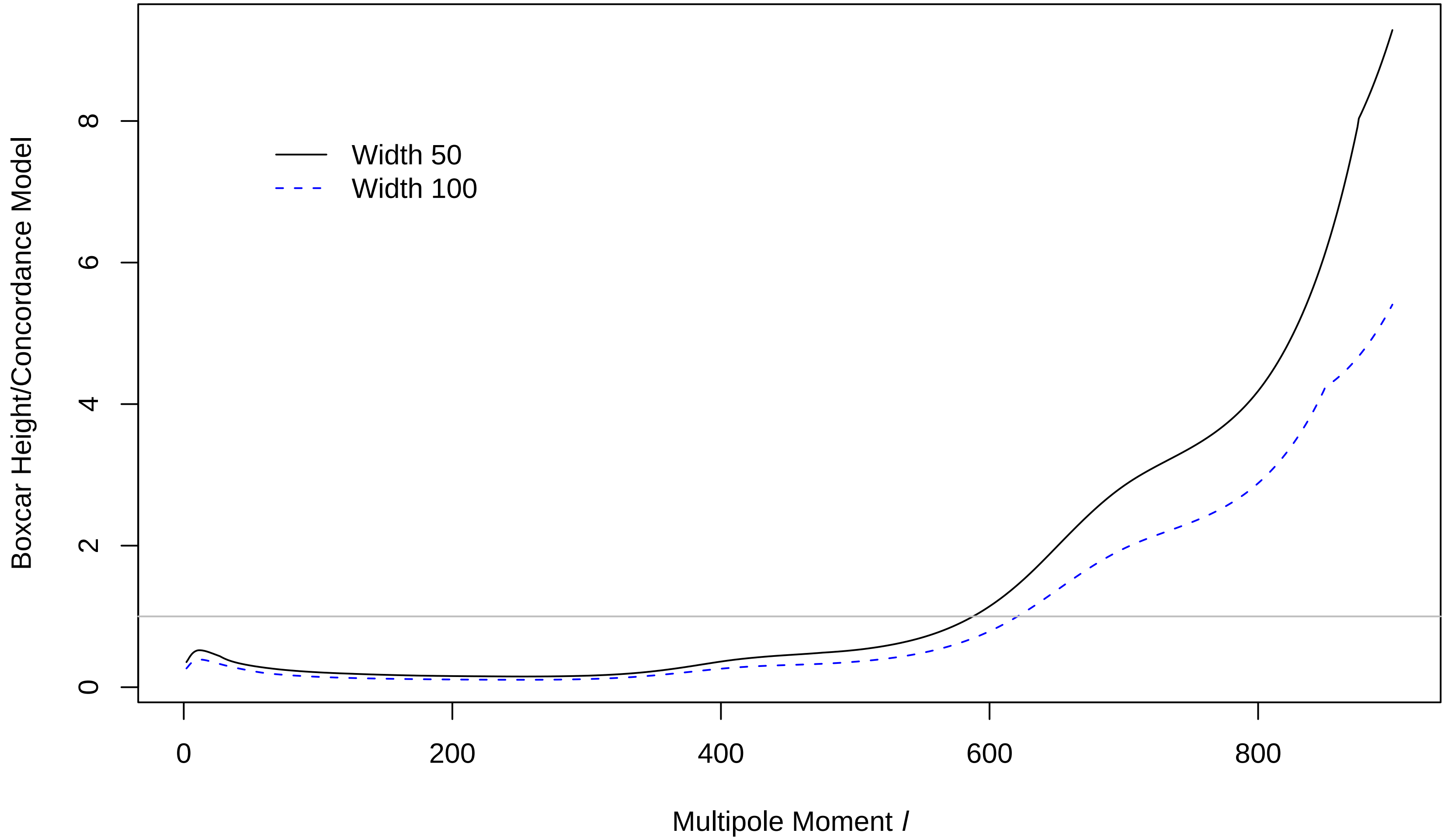


Extended search, over millions of spectra, in progress.

Eyes on the Ball I: Parametric Probes (cont'd)

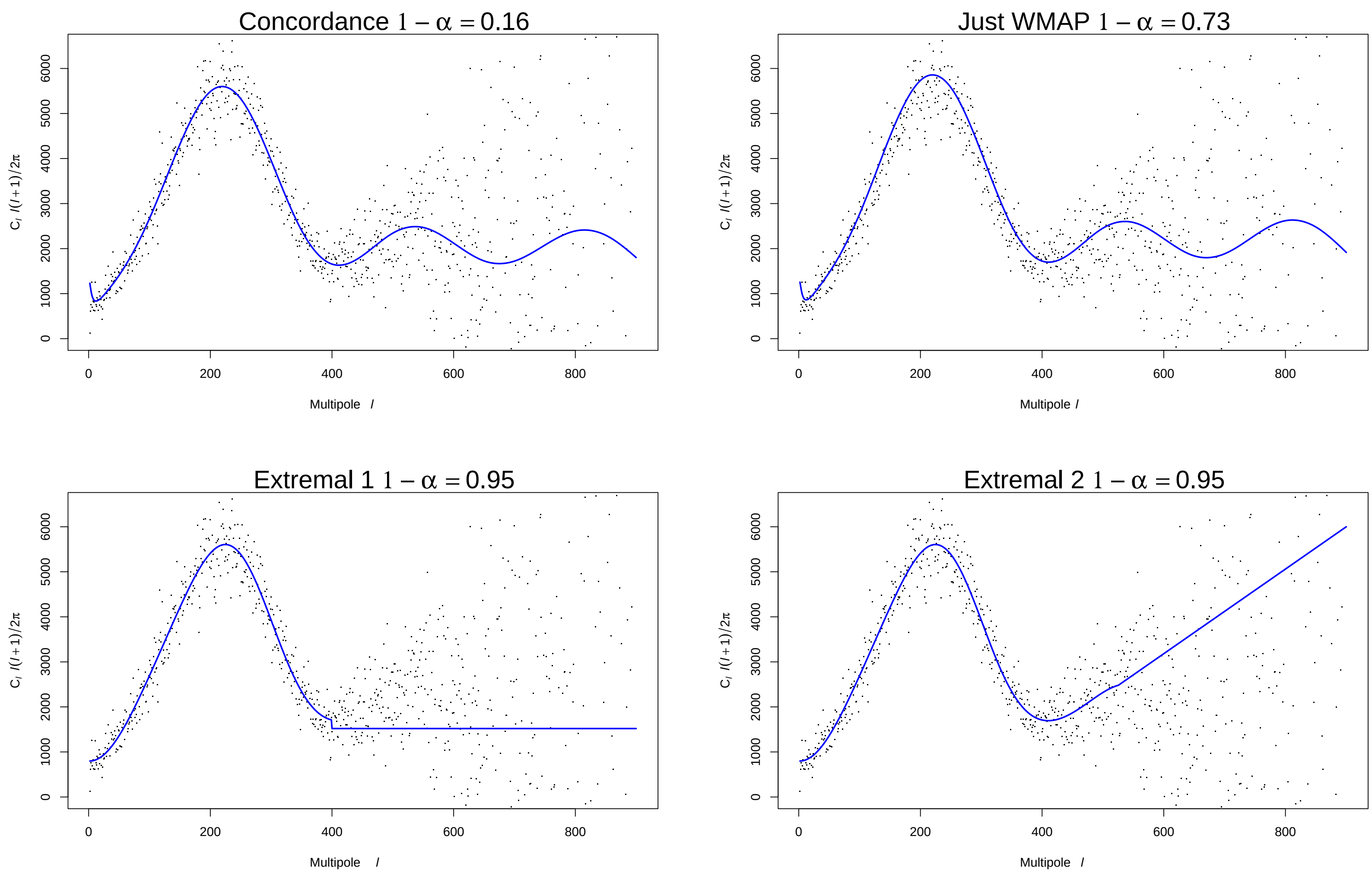
Probe from center with boxcars of given width centered at each ℓ.

Maximum boxcar height in 95% ball, relative to Concordance Model



Eyes on the Ball II: Model Checking

Inclusion in the confidence ball provides simultaneous goodness-of-fit tests for parametric (or other) models.



Take-Home Points

- A general strategy for nonparametric inference
  - Build confidence sets for functions with uniform asymptotic (or finite-sample) coverage and post-hoc protection.
  - Constrain and search post-hoc to make inferences

- Advantages:
  - Supports simultaneous inferences on any functionals.
  - Makes it easy to work with complicated spaces or features.
  - Straightforward to vary and compare assumptions.

- With CMB, nonparametric analysis strikingly confirms many WMAP findings but also shows how some depend strongly on the model.

  Also provides guidance for subsequent analyses and data collection.