Confidence Sets for Nonparametric Regression with Application to Cosmology

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

Larry Wasserman
Department of Statistics
Carnegie Mellon University

This work partially supported by NIH Grant 1 R01 NS047493-01 and NSF Grant ITR 0104016.

WMAP: The Cosmic Microwave Background

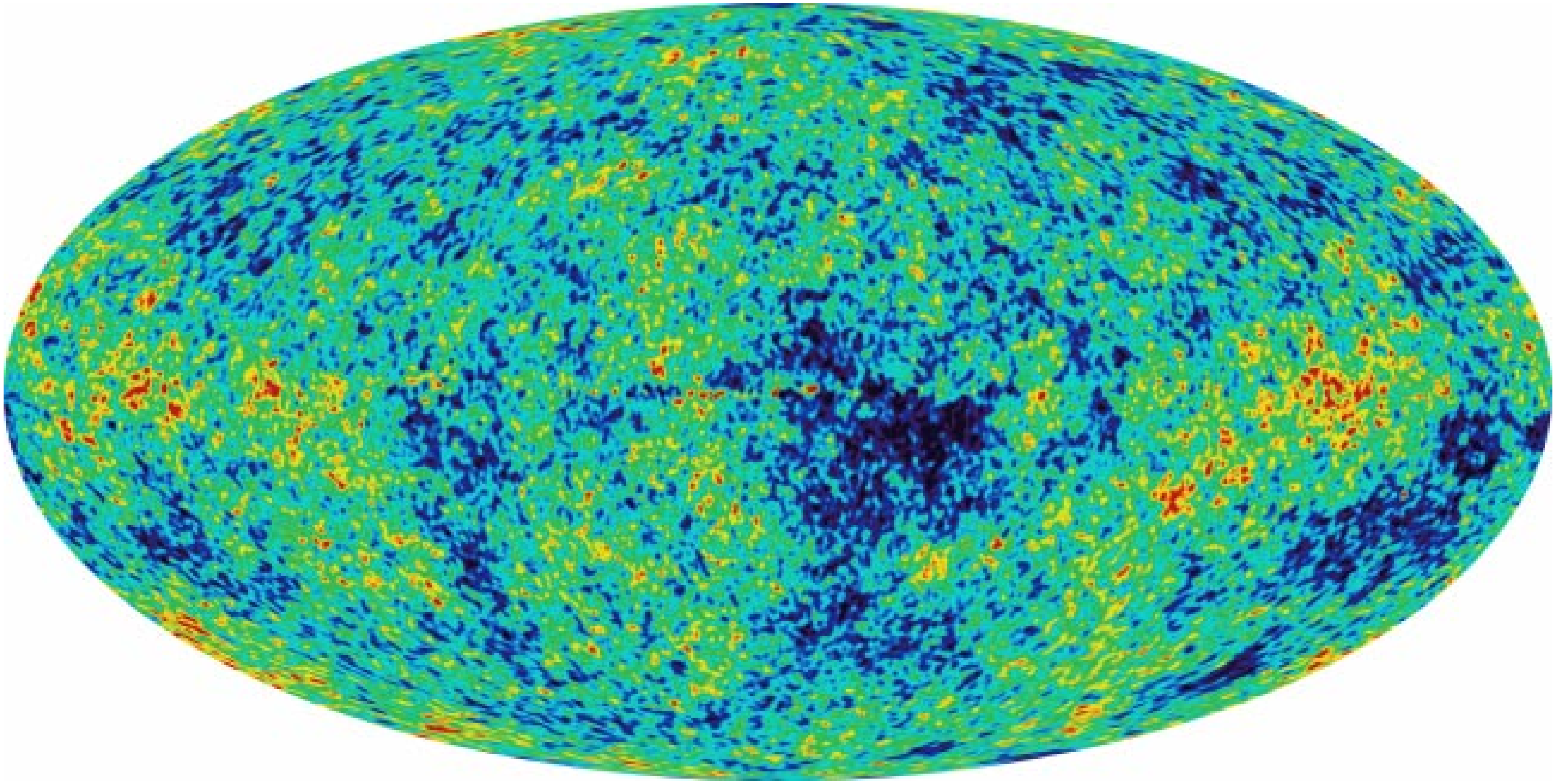


Image: NASA/WMAP Science Team

WMAP in the News

``Breakthrough of the Year, 2003'' — *Science*

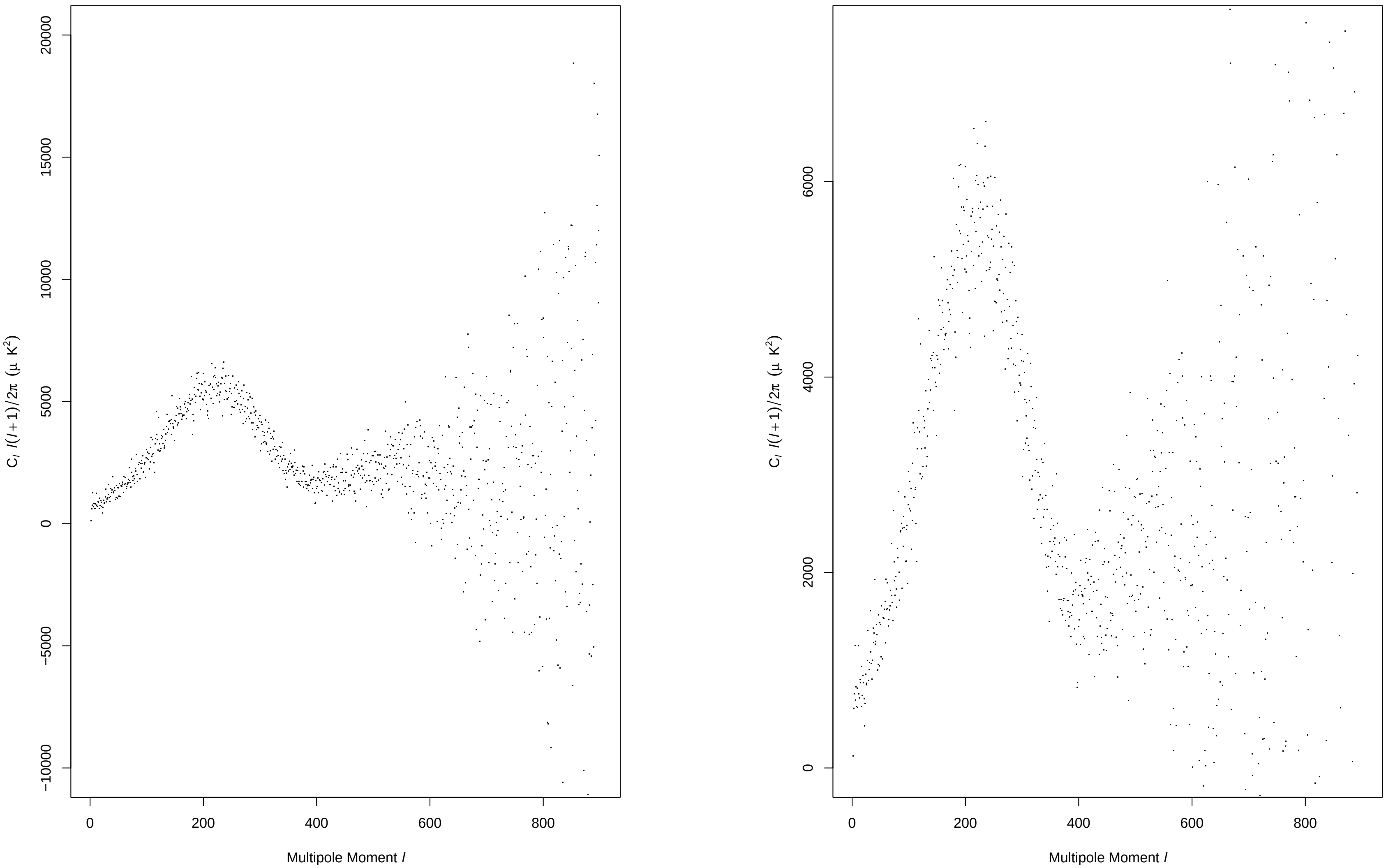
``Most precise, detailed map yet produced of universe just after its birth ... confirms Big Bang theory'' — *New York Times*

``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



Road Map

1. Constructing Confidence Sets for Unknown Functions
   – Simultaneity, Bias, and Relevance
   – The Pivot-Ball and Subspace-Pretesting Methods
2. Extensions to the Pivot-Ball Method
   – Wavelet Bases
   – Weighted Loss and Non-constant variance
3. Extensions to Subspace Pretesting
   – Confidence Bands
   – Double-Confidence Tube
4. Case Study: The Cosmic Microwave Background
   – The Physics of the Early Universe
   – WMAP Data
   – Keeping Our Eyes on the Ball: Parametric Probes and Confidence Catalogs

Road Map

1. Constructing Confidence Sets for Unknown Functions
   - Simultaneity, Bias, and Relevance
   - The Pivot-Ball and Subspace-Pretesting Methods

2. Extensions to the Pivot-Ball Method
   - Wavelet Bases
   - Weighted Loss and Non-constant variance

3. Extensions to Subspace Pretesting
   - Confidence Bands
   - Double-Confidence Tube

4. Case Study: The Cosmic Microwave Background
   - The Physics of the Early Universe
   - WMAP Data
   - Keeping Our Eyes on the Ball: Parametric Probes and Confidence Catalogs

The Nonparametric Regression Problem

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

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

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

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

$$\mathcal{F}_p(C) = \left\{ f \in \mathcal{L}^2 \colon \int |f^{(p)}|^2 \leq C^2 \right\}.$$

Other examples: Besov space, Lipschitz class

Smoothing

To obtain consistent estimators in this problem, it is necessary to smooth.

Common methods include kernels, local polynomials, splines, wavelet shrinkage, and other orthogonal basis expansion.

All of these methods have a tuning parameter which is usually chosen to balance bias and variance.

This tuning parameter can be selected using the data (e.g., CV, GCV, SURE).

Rate-Optimal Estimators

Define risk $R(\widehat{f}, f) = \mathsf{E}L(\widehat{f}, f)$ for specified loss $L(\widehat{f}, f)$.

Typical choice: $L(\widehat{f}, f) = \int (\widehat{f} - f)^2$

Optimize rate of convergence r_n for minimax risk. That is,

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

In infinite-dimensional problems, $r_n \sqrt{n} \to \infty$.

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.

More generally, Donoho and Johnstone (1998), Cai (1999) show that certain wavelet shrinkage estimators adapt to unknown smoothness in Besov spaces.

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 bias$^2 \approx$ var.

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

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

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

Two common solutions in the literature:

– Bias Correction: Shift confidence set by estimated bias.

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

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 $\alpha = K\phi(c) + 2(1 - \Phi(c))$.

The constants K and c depend on $T(x) = \ell(x)/\|\ell(x)\|$.

Simultaneity (cont'd)

Special case: $f(x) = \langle \ell(x), \theta \rangle$. Then

$$\begin{aligned}\alpha &= \mathsf{P}\left\{ \sup_x \left| \frac{\widehat{f}(x) - f(x)}{\|\ell(x)\|} \right| > c\sigma \right\} \\ &= \mathsf{P}\left\{ \sup_x |\langle T(x), \epsilon \rangle| > c\sigma \right\} \\ &= \mathsf{P}\left\{ \sup_x \left| \langle T(x), \frac{\epsilon}{\|\epsilon\|} \rangle \right| > c\sigma/\|\epsilon\| \right\}.\end{aligned}$$

By conditioning on $\|\epsilon\|$ this reduces to finding the volume of a tube on the sphere S^{n-1} around the image of T.

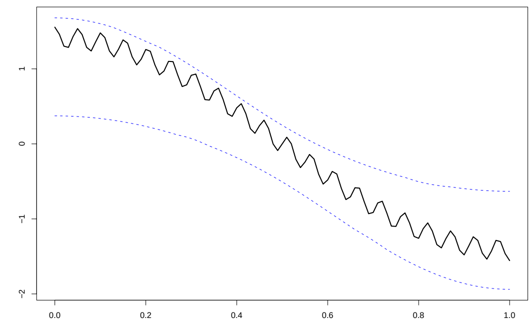
Approximation found by Hotelling (1939), generalized by Weyl (1939), well studied since then.

Must account for bias in general.

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.

Remark: Uniform Coverage

For asymptotic confidence procedures, we prefer uniform coverage:

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

This ensures that the coverage error depends only on n, not on f.

Li (1989) showed that with no prior smoothness bound in $\mathcal{F}$, any $1-\alpha$ confidence sets for $f_n = (f(x_1), \ldots, f(x_n))$ of the form

$$\mathcal{C}_n = \left\{ f_n \in \mathbb{R}^n \colon\ n^{-1/2} \| f_n - \widehat{f}_n \| \leq s_n(y) \right\}$$

that are "asymptotically honest" in the sense of

$$\lim_{n \to \infty} \inf_{f \in \mathcal{F}} \mathsf{P}_f \left\{ \mathcal{C}_n \ni f \right\} \geq 1 - \alpha$$

must have $s_n \geq n^{-1/4}$.

Possible Approaches

X Estimate bias pointwise

Often increases variance more than it reduces bias.

X Undersmoothing

Requires additional calibration; typically non-uniform coverage.

$\surd$ Pivot-Ball Method (Beran and Dümbgen 1998)

Uniform asymptotic coverage for $\mathcal{L}^2$ confidence balls.

Supports functional search.

$\surd$ Subspace Pretesting (Baraud 2004)

Finite-sample coverage for ℓ^2 confidence balls.

Supports functional search.

Other approaches include bounding global bias (Sun and Loader) and scale-space methods (Chaudhuri and Marron).

Pivot-Ball (Beran and Dümbgen 1998)

Let $\phi_1, \phi_2, \ldots$ be an orthonormal basis and write $f = \sum_j \theta_j \phi_j$.

Let $\widehat{\theta}_j(\lambda)$ for (possibly vector-valued) tuning parameter λ.

0. Define loss $L_n(\lambda) = \sum_{j=1}^n (\widehat{\theta}_j(\lambda) - \theta_j)^2$.

   Let $S_n(\lambda)$ be an unbiased estimate of $\mathsf{E} L_n(\lambda)$.

   Choose $\widehat{\lambda}_n$ to minimize $S_n(\lambda)$.

1. Show that *pivot process* $B_n(\lambda) = \sqrt{n}(L_n(\lambda) - S_n(\lambda))$ converges weakly to Gaussian process with mean 0, cov. $K(s,t)$.

2. Find an estimator $\widehat{\tau}_n^2$ of $K(\widehat{\lambda}_n, \widehat{\lambda}_n)$ so that

$$\frac{B_n(\widehat{\lambda}_n)}{\widehat{\tau}_n} \rightsquigarrow \mathsf{N}(0,1).$$

Pivot-Ball (cont'd)

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

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

is

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

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

for projection f_n onto first n coefficients.

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

Pivot-Ball (cont'd)

Beran and Dümbgen considered modulators

$$\widehat{\theta} = (\lambda_1 \widetilde{\theta}_1, \ldots, \lambda_n \widetilde{\theta}_n),$$

where

$$1 \geq \lambda_1 \geq \cdots \geq \lambda_n \geq 0,$$

and

$$\widetilde{\theta}_j \approx \frac{1}{n} \sum_{i=1}^{n} Y_i \phi_j(x_i).$$

Remark: Beran and Dümbgen (1998) stated the main result in ℓ^2, but in practice need Sobolev assumptions to (i) estimation σ and (ii) move from sequence to function space.

Subspace Pretesting (Baraud 2004)

Write $\boldsymbol{Y} = \boldsymbol{f} + \sigma\boldsymbol{\epsilon}$ where $\boldsymbol{f} = (f(x_1), \ldots, f(x_n))$.

Baraud procedure constructs finite-sample confidence ball for $\boldsymbol{f}$.

Ideal Construction: Control $\|\boldsymbol{f} - \widehat{\boldsymbol{f}}\|$ uniformly over $\boldsymbol{f}$ for good $\widehat{\boldsymbol{f}}$.

But this doesn't work.

- Let $S \subset \mathbb{R}^n$ be a subspace of dimension $< n$ and define π_S to be orthogonal projection onto S.
- If $\widehat{\boldsymbol{f}} \equiv \pi_S \boldsymbol{Y}$, then $\|\boldsymbol{f} - \widehat{\boldsymbol{f}}\|^2 = \|(I - \pi_S)\boldsymbol{f}\|^2 + \sigma^2 \|\pi_S \boldsymbol{\epsilon}\|^2$.
- We usually cannot bound $\|(I - \pi_S)\boldsymbol{f}\|^2$ a priori.

Instead: Use pretest to control $\|(I - \pi_S)\boldsymbol{f}\|^2$. Specifically, use $(I - \pi_S)\boldsymbol{Y}$ to test $\boldsymbol{f} \in S$ versus $\boldsymbol{f} \notin S$.

- When don't reject $\boldsymbol{f} \in S$, then $\|(I - \pi_S)\boldsymbol{f}\|^2$ is small with high-probability.

Subspace Pretesting (cont'd)

Let $\mathcal{S}$ be a collection of subspaces S such that $\mathbb{R}^n \in \mathcal{S}$.

Example: $f(x) = \sum_j \theta_j \phi_j(x)$ and S_j corresponds to j-term partial sums.

For $S \in \mathcal{S}$, let $\widehat{\boldsymbol{f}}_S = \pi_S \boldsymbol{Y}$ and choose α_S such that $\sum_S \alpha_S \leq \alpha$.

Choose tests and radii ρ_S for $S \in \mathcal{S}$ so that

$$\mathsf{P}\Big\{\text{Ball}(\widehat{\boldsymbol{f}}_S, \rho_S) \not\ni \boldsymbol{f} \text{ and Don't reject } H_0 : \boldsymbol{f} \in S\Big\} \leq \alpha_S.$$

If $\widehat{S} = \text{argmin}_{S \in \mathcal{S}}\, \rho_S$, then

$$\begin{aligned}\mathsf{P}\Big\{\text{Ball}(\widehat{\boldsymbol{f}}_{\widehat{S}}, \rho_{\widehat{S}}) \not\ni \boldsymbol{f}\,\Big\} &\leq \sum \mathsf{P}\Big\{\boldsymbol{f} \notin \text{Ball}(\widehat{\boldsymbol{f}}_S, \rho_S) \text{ and Don't reject S}\Big\} \\ &\leq \sum \alpha_S \leq \alpha.\end{aligned}$$

Can get a smaller set by taking intersection of balls.

Road Map

1. Constructing Confidence Sets for Unknown Functions
   – Simultaneity, Bias, and Relevance
   – The Pivot-Ball and Subspace-Pretesting Methods
2. Extensions to the Pivot-Ball Method
   – Wavelet Bases
   – Weighted Loss and Non-constant variance
3. Extensions to Subspace Pretesting
   – Confidence Bands
   – Double-Confidence Tube
4. Case Study: The Cosmic Microwave Background
   – The Physics of the Early Universe
   – WMAP Data
   – Keeping Our Eyes on the Ball: Parametric Probes and Confidence Catalogs

Extensions to the Pivot-Ball Method

- Wavelet bases (Genovese and Wasserman 2003)
  [next slide]

- Weighted-Loss/Nonconstant Variance (Genovese et al. 2004)
  [see CMB example]

- Density Estimation (Jang, Genovese, and Wasserman 2004)
  [not discussed]

- Other loss functions (in progress)

Pivot-Ball Method: Wavelets

Write $f = \sum_k \alpha_{J_0,k}\varphi_j + \sum_{j=J_0}^{\infty} \sum_k \beta_{jk}\psi_{jk}$.

Wavelet coefficients characterize f within family of Besov spaces.

Donoho and Johnstone considered several shrinkage schemes that lead to rate-optimal (adaptive) estimators over Besov spaces: Universal thresholding, global SureShrink, levelwise SureShrink.

Because wavelet basis functions unbounded in general, pivot process is not asymptotically equicontinuous near zero.

Thus we need to restrict thresholds to $[\varrho\widehat{\sigma}\rho_n, \widehat{\sigma}\rho_n]$, where $\rho_n = \sqrt{2\log n/n}$ and $1/\sqrt{2} < \varrho < 1$.

Conjecture that SureShrink result holds for $\varrho > 0$ but it doesn't hold for $\varrho = 0$.

Pivot-Ball Method: Wavelets (cont'd)

But pivot-ball method does carry over to wavelet shrinkage with ball $\mathcal{C}_n$ as before but with radius:

$$s_n^2 = \frac{\widehat{\sigma}^2 z_\alpha}{\sqrt{n/2}} + S_n(\widehat{\lambda}_n).$$

Confidence set radius $O(n^{-1/4})$ over Besov balls consistent with results of Li (1989) and Baraud (2004).

Have the adaptive estimators really helped here?

Pivot-Ball Method: Simple Simulation

Test functions:

$$f_0(x) = 0$$
$$f_1(x) = 2(6.75)^3 x^6 (1-x)^3$$
$$f_2(x) = \begin{cases} 1.5 & \text{if } 0 \le x < 0.3 \\ 0.5 & \text{if } 0.3 \le x < 0.6 \\ 2.0 & \text{if } 0.6 \le x < 0.8 \\ 0.0 & \text{otherwise.} \end{cases}$$

Let $\alpha = .05$, $n = 1024$, $\sigma = 1$, and use 5000 iterations.

For comparison, χ^2 radius is 1.074.

Pivot-Ball Method: Simple Simulation (cont'd)

σ known:

Method	Function	Coverage	Average Radius
SureShrink (levelwise)	f_0	0.944	0.268
	f_1	0.940	0.289
	f_2	0.927	0.395
Modulator (cosine)	f_0	0.931	0.253
	f_1	0.930	0.259
	f_2	0.905	0.318

σ unknown:

Method	Function	Coverage
SureShrink (levelwise)	f_0	0.954
	f_1	0.953
	f_2	0.929
Modulator (cosine)	f_0	0.999
	f_1	0.999
	f_2	0.997

Pivot-Ball Method: Functionals

To make inferences for functionals of f, we can search $\mathcal{C}_n$:

$$\left(\inf_{f\in\mathcal{C}_n} T(f),\ \sup_{f\in\mathcal{C}_n} T(f)\right)$$

is a confidence set for $T(f)$.

If $\mathcal{T}$ is a set of functionals, then

$$\left\{\left(\inf_{f\in\mathcal{C}_n} T(f),\ \sup_{f\in\mathcal{C}_n} T(f)\right) : T \in \mathcal{T}\right\}$$

gives simultaneous intervals for all the functionals in $\mathcal{T}$.

This is useful for post-hoc exploration.

Pivot-Ball Method: Functionals (cont'd)

Fix a decreasing sequence $\Delta_n > 0$ and consider block-averages

$$\mathcal{T}_n = \left\{T \colon T(f) = \frac{1}{b-a}\int_a^b f\,dx,\ 0 \le a < b \le 1,\ |b-a| \ge \Delta_n\right\}.$$

For $\eta, c > 0$, define

$$\mathcal{F}_{\eta,c} = \bigcup_{p,q\ge 1}\ \bigcup_{\gamma \ge 1/2+\eta} \mathcal{B}_{p,q}^{\varsigma(\gamma)}(c),$$

with $\varsigma(\gamma) = \gamma + (1/p - 1/2)_+$. The parameter η is an increment of smoothness required only in the non-sparse case ($p \ge 2$).

Let

$$\kappa = \sup\left\{\#\{\psi_{jk}(x) \neq 0 : 0 \le k < 2^j\}\colon\ 0 \le x \le 1, j \ge J_0\right\}$$

be the maximal number of $\psi_{j.}$ that "hit" a given point.

Pivot-Ball Method: Functionals (cont'd)

Theorem. If the mother and father wavelets are compactly supported with $\kappa < \infty$ and $\|\psi\|_1 < \infty$ and if $\Delta_n^{-1} = o(n^\zeta/(\log n)^{\lfloor \zeta \rfloor})$ for some $0 \le \zeta \le 1$, then for any sequence $w_n \ge 0$ that satisfies

$$w_n \to 0 \qquad \text{and} \qquad \liminf_{n\to\infty} w_n n^{1-\zeta} (\log n)^{\lfloor \zeta \rfloor} > 0,$$

we have

$$\liminf_{n\to\infty} \inf_{f \in \mathcal{F}_{\eta,c}} \mathsf{P}\Big\{ T(f) \in J_n(T) \text{ for all } T \in \mathcal{T}_n \Big\} \ge 1 - \alpha.$$

where

$$J_n(T) = \left(\inf_{f_n \in \mathcal{C}_n} T(f_n) - w_n, \sup_{f_n \in \mathcal{C}_n} T(f_n) + w_n \right).$$

Conditions are satisfied by most standard wavelet functions.

Road Map

1. Constructing Confidence Sets for Unknown Functions
   - Simultaneity, Bias, and Relevance
   - The Pivot-Ball and Subspace-Pretesting Methods
2. Extensions to the Pivot-Ball Method
   - Wavelet Bases
   - Weighted Loss and Non-constant variance
3. Extensions to Subspace Pretesting
   - Confidence Bands
   - Double-Confidence Tube
4. Case Study: The Cosmic Microwave Background
   - The Physics of the Early Universe
   - WMAP Data
   - Keeping Our Eyes on the Ball: Parametric Probes and Confidence Catalogs

Extensions to Subspace Pretesting: Bands

Let $\phi_1, \phi_2, \ldots$ be bounded, ortho. basis on $[0, 1]$, e.g., cosine basis.

Assume $f = \sum_{j=1}^{J} \theta_j \phi_j$ for specified $J \equiv J_n$. Often take $J_n = n$.

Consider subspaces S_d of partial sums, and let $\widehat{f}_d = \sum_{j=1}^{d} \widehat{\theta}_j \phi_j$, where $\widehat{\theta}_j \approx$ ind $N(\theta_j, \sigma^2/n)$.

Define $a_d = \sqrt{\sum_{j=1}^{d} \phi_j^2}$ and $b_d = \sqrt{\sum_{j=d+1}^{J} \phi_j^2}$.

We consider confidence bands

$$\mathcal{B} = \cap_d \mathcal{B}_d \quad \text{where} \quad \mathcal{B}_d = \left\{ f\colon\, |f(x) - \widehat{f}_d(x)| \le \frac{\sigma}{\sqrt{n}} \delta_d a_d(x) \right\}.$$

Following Baraud, choose tests and define δ_ds so that

$$\mathsf{P}\left\{ \mathcal{B}_d \not\ni f \text{ and } S_d \text{ not rejected} \right\} \le \alpha_d,$$

then $\mathcal{B}$ is a $1 - \alpha$ confidence set that accounts for the bias.

Subspace Pretesting: Bands (cont'd)

Define the following:

1. Normalized basis functions $T_{dj} = \phi_j / a_d$ and $\widetilde{T}_{dj} = \phi_j / b_d$.
2. Bias function $z_{\theta,d} = \frac{b_d}{a_d} \sum_{j=d+1}^{J} \theta_j \widetilde{T}_{dj}$.
3. Maximum bias

$$W_d = \left\| \frac{\widehat{f}_d - f}{a_d} \right\|_\infty = \sup_x \left| \frac{\sigma}{\sqrt{n}} \sum_{j=1}^{d} Z_j T_{dj}(x) + z_{\theta,d}(x) \right|$$

4. Test statistic

$$U_d = \left\| \frac{\widehat{f}_d - \widehat{f}_J}{a_d} \right\|_\infty = \sup_x \left| \frac{\sigma}{\sqrt{n}} \frac{b_d(x)}{a_d(x)} \sum_{j=1}^{d} Z_j \widetilde{T}_{dj}(x) + z_{\theta,d}(x) \right|.$$

5. Distributions $\overline{G}_{\theta,d}(w) = \mathsf{P}\left\{ W_d > w \right\}$ and $H_{\theta,d}(u) = \mathsf{P}\left\{ U_d \le u \right\}$.
6. Critical value for test $c_d = H_{0,d}^{-1}(1 - \gamma_d)$ where $0 < \gamma_d < 1 - \alpha_d$.

Subspace Pretesting: Bands (cont'd)

Set

$$\delta_d = \sup_\theta \frac{\sqrt{n}}{\sigma} \overline{G}_{\theta,d}^{-1}\left(\frac{\alpha_d}{H_{\theta,d}(c_d)}\right).$$

We have

$$\begin{aligned}
\mathsf{P}\Big\{\mathcal{B}_d \not\ni f, U_d \le c_d\Big\} &= \mathsf{P}\Big\{\mathcal{B}_d \not\ni f\Big\}\,\mathsf{P}\Big\{U_d \le c_d\Big\} \\
&= \mathsf{P}\left\{\left\|\frac{\widehat{f}_d - f}{a_d}\right\|_\infty > \frac{\sigma}{\sqrt{n}}\delta_d\right\} H_{\theta,d}(c_d) \\
&= \overline{G}_{\theta,d}\left(\frac{\sigma}{\sqrt{n}}\delta_d\right) H_{\theta,d}(c_d) \\
&\le \alpha_d.
\end{aligned}$$

By suitable approximation, we can reduce the δ_d computation to a smooth (but largish) optimization problem.

Still working on computational side of the problem.

Double-Confidence Tube

Let $I_d(x)$ be tube $1-\alpha_d/2$ confidence band centered on $\widehat{f}_d$.

Let $B_d = \|z_{\theta,d}(x)\|_\infty \le \sup_x \left|\frac{b_d(x)}{a_d(x)}\right| \sqrt{\sum_{j=d+1}^J \theta_j^2}$.

1. Find $\overline{B}_d$ such that $\mathsf{P}\left\{B_d \le \overline{B}_d\right\} \ge 1-\alpha_d/2$.
2. Let $\overline{I}_d$ be I_d inflated by $\overline{B}_d$.
3. Then $I = \cap I_d$ is a $1-\alpha$ confidence set for f.

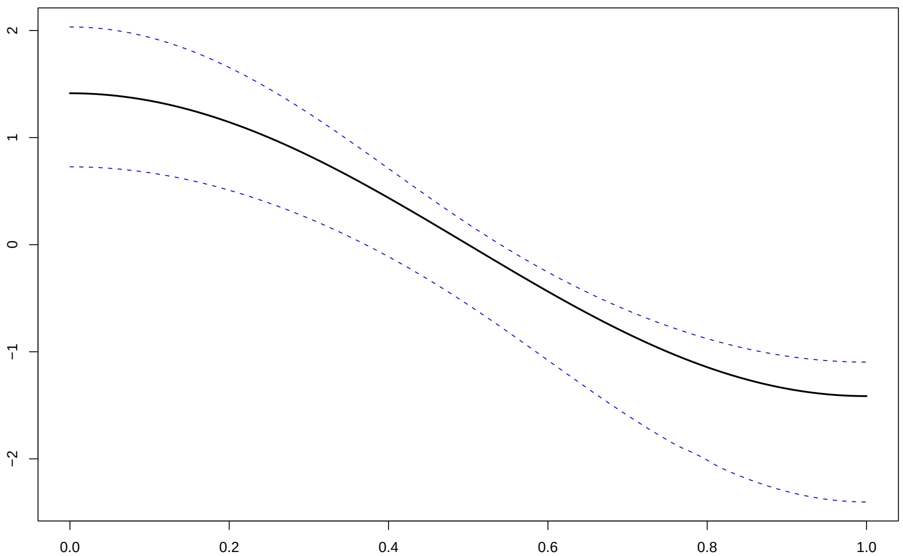
Note that with $\psi^2 = \sum_{j=d+1}^J \theta_j^2$,

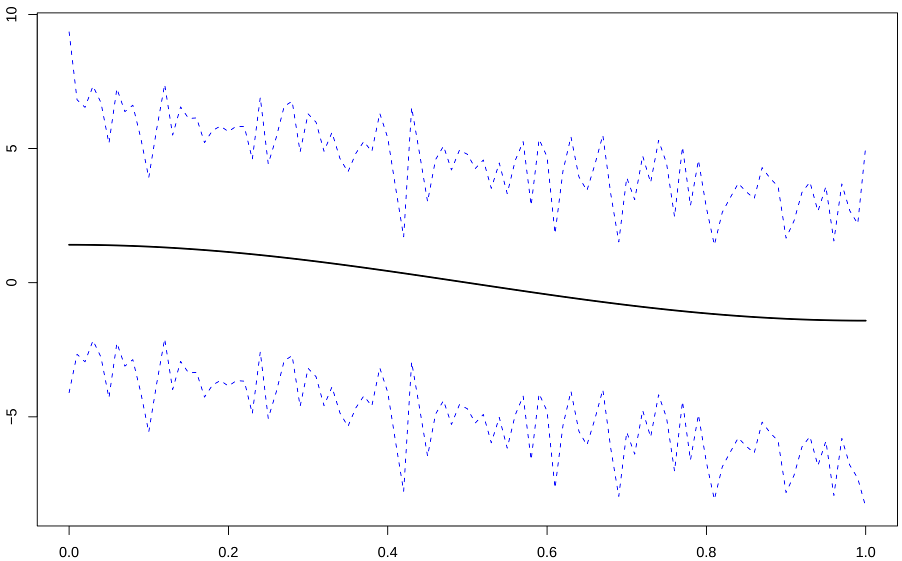
$$\widehat{\psi}^2 = \sum_{j=d+1}^{J} \widehat{\theta}_j^2 \overset{d}{=} \frac{\sigma^2}{n} \chi^2_{J-d}(n\psi^2/\sigma^2).$$

Then, set

$$\overline{B}_d = \sup_x \left|\frac{b_d(x)}{a_d(x)}\right| \sup\{\psi^2\colon \text{qchisq}(\alpha_d/2, J-d, n\psi^2/\sigma^2) \le n\widehat{\psi}^2/\sigma^2\}.$$

Double-Confidence Tube (cont'd)



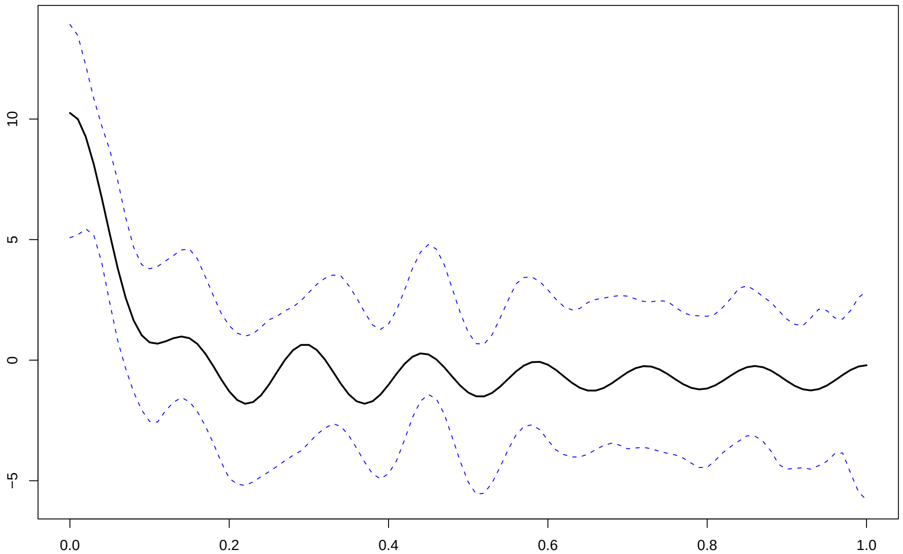


Double Confidence Tube

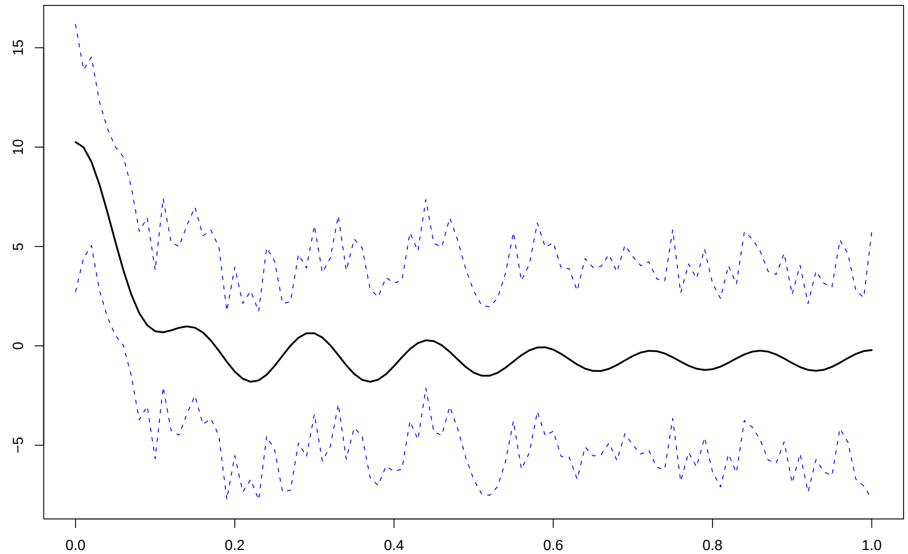
Full basis

$(n = 100,\ \alpha = 0.05)$

Double-Confidence Tube (cont'd)



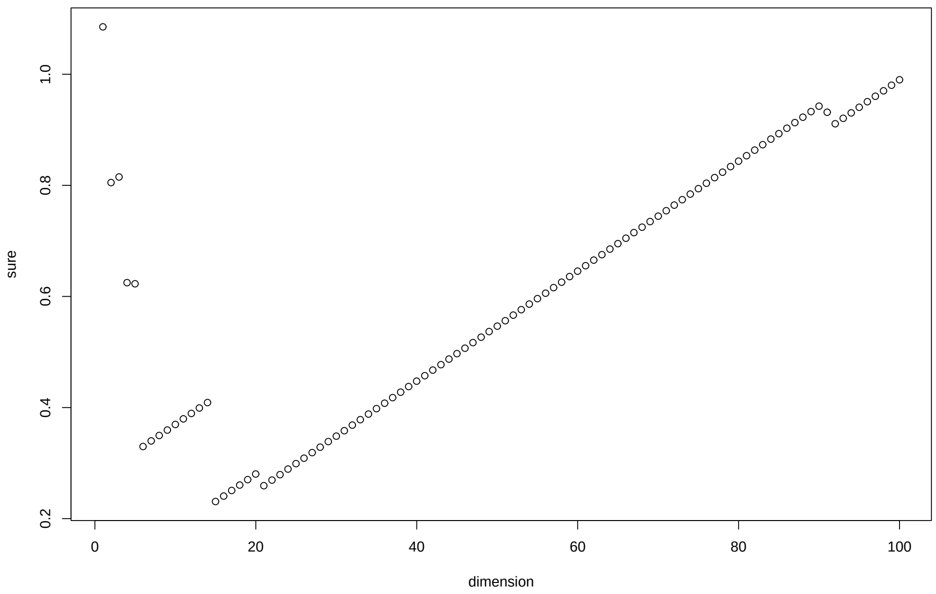
Double Confidence Tube



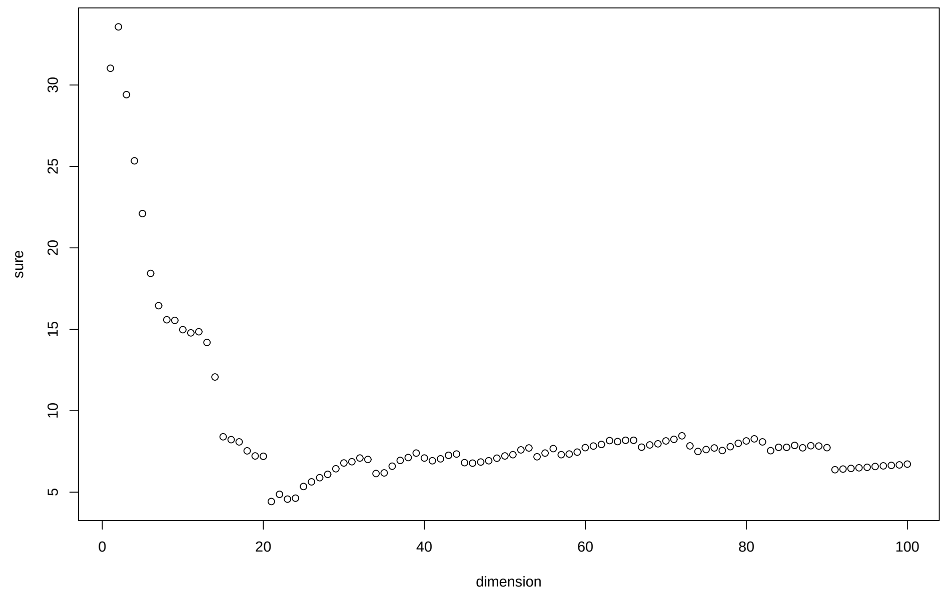
Full basis

($n = 100$, $\alpha = 0.05$)

Double-Confidence Tube (cont'd)



SURE



Maximum Width

Road Map

1. Constructing Confidence Sets for Unknown Functions
   – Simultaneity, Bias, and Relevance
   – The Pivot-Ball and Subspace-Pretesting Methods
2. Extensions to the Pivot-Ball Method
   – Wavelet Bases
   – Weighted Loss and Non-constant variance
3. Extensions to Subspace Pretesting
   – Confidence Bands
   – Double-Confidence Tube
4. Case Study: The Cosmic Microwave Background
   – The Physics of the Early Universe
   – WMAP Data
   – Keeping Our Eyes on the Ball: Parametric Probes and Confidence Catalogs

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

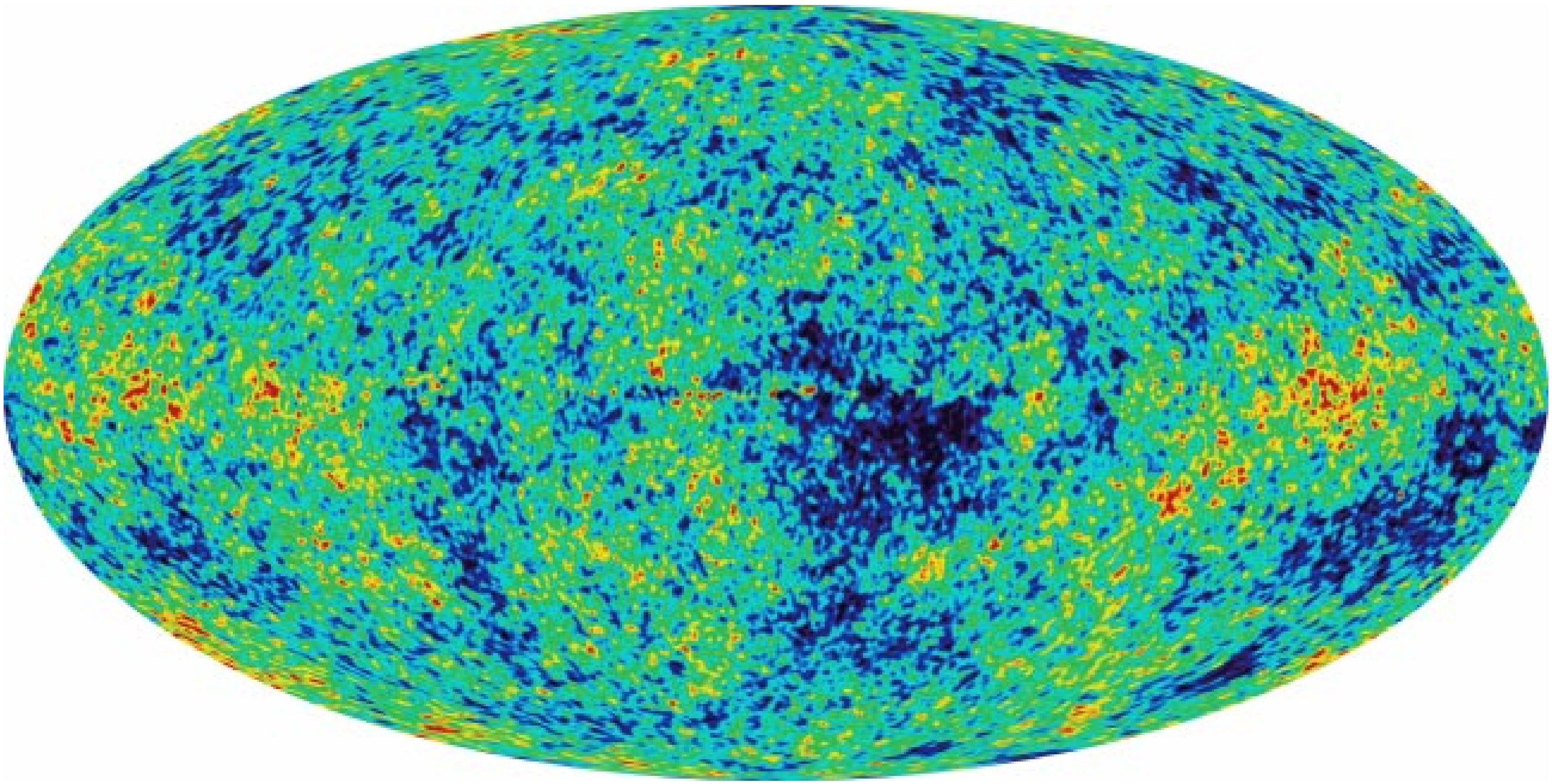


Image: NASA/WMAP Science Team

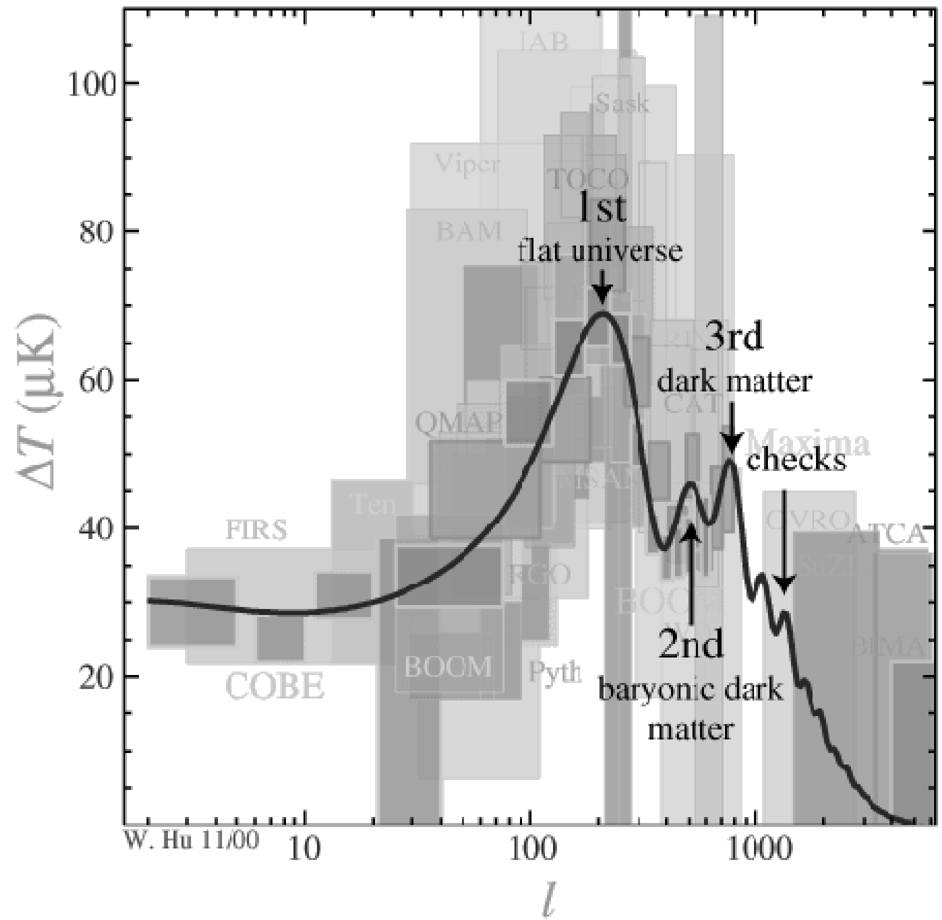
Physics of the Early Universe (cont'd)

The acoustic oscillations before recombination carry information about the geometry and composition of the early universe. (Can you hear the shape of the universe?)

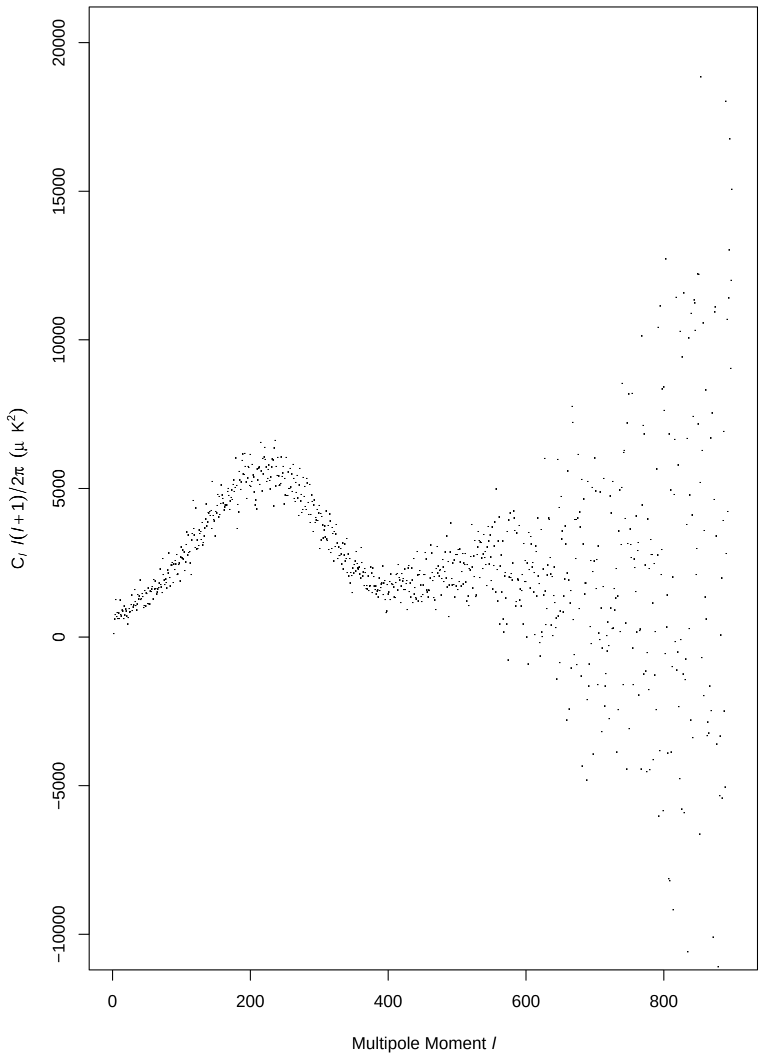
They produce a pattern of hot and cold spots on the sky map.

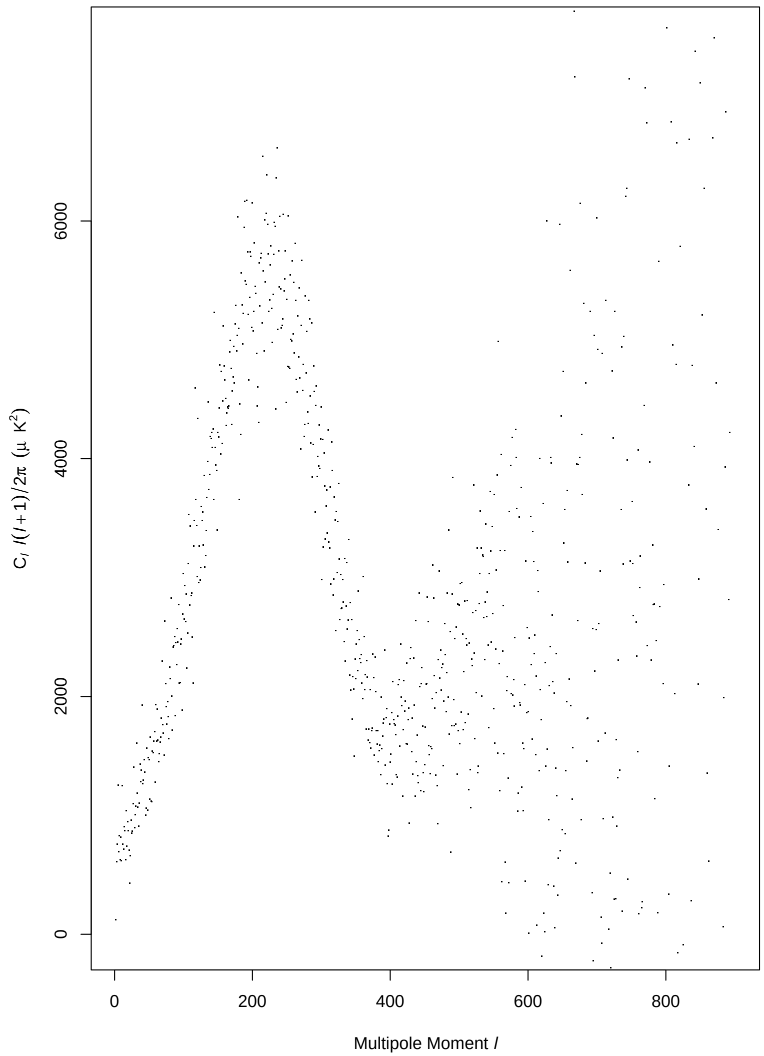
Cosmologists decompose the sky map into spherical harmonics and compute the coefficient variance at each angular scale ℓ.

This is the raw estimated CMB "power spectrum" $f(\ell) = \widehat{C}_\ell$.

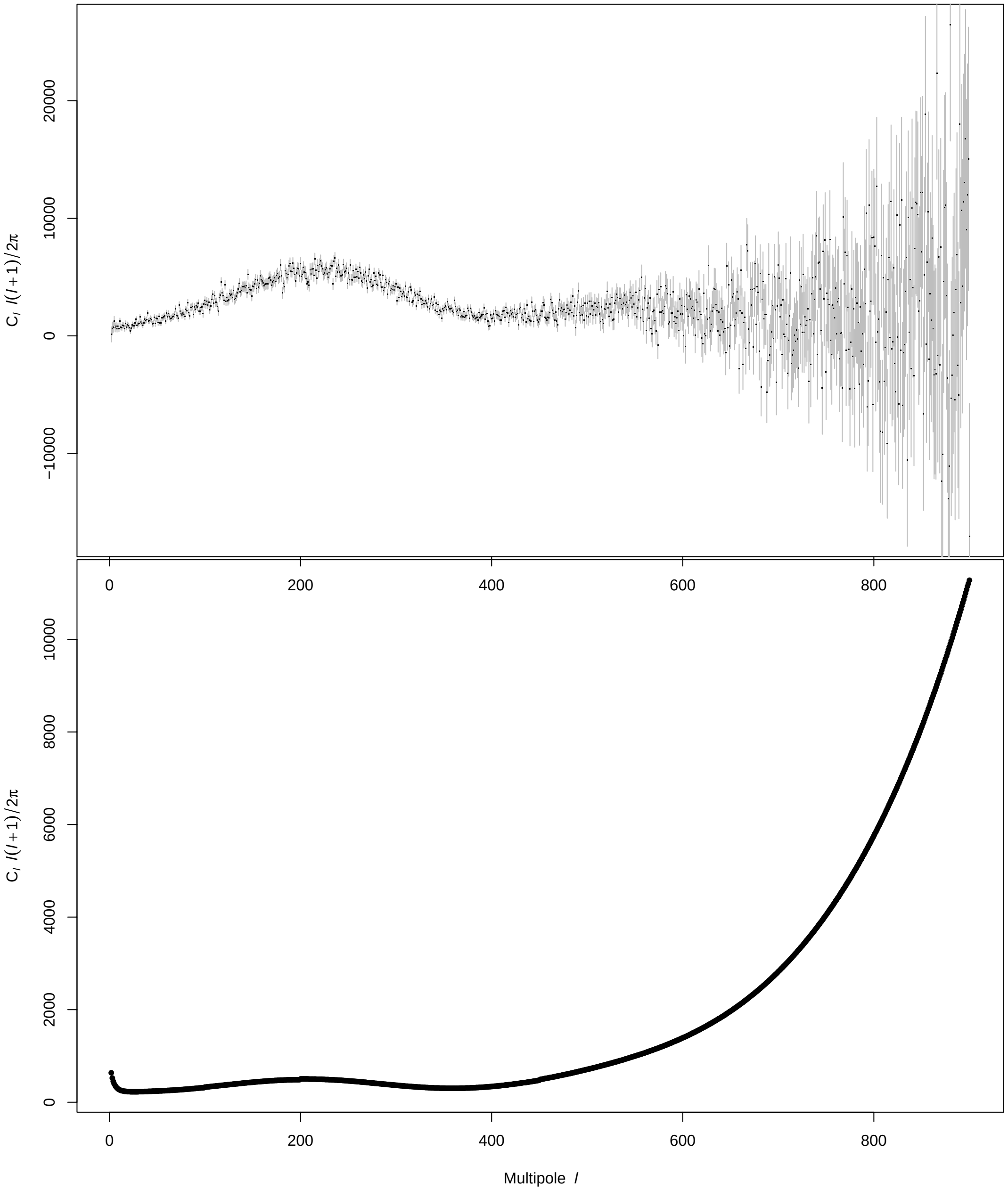


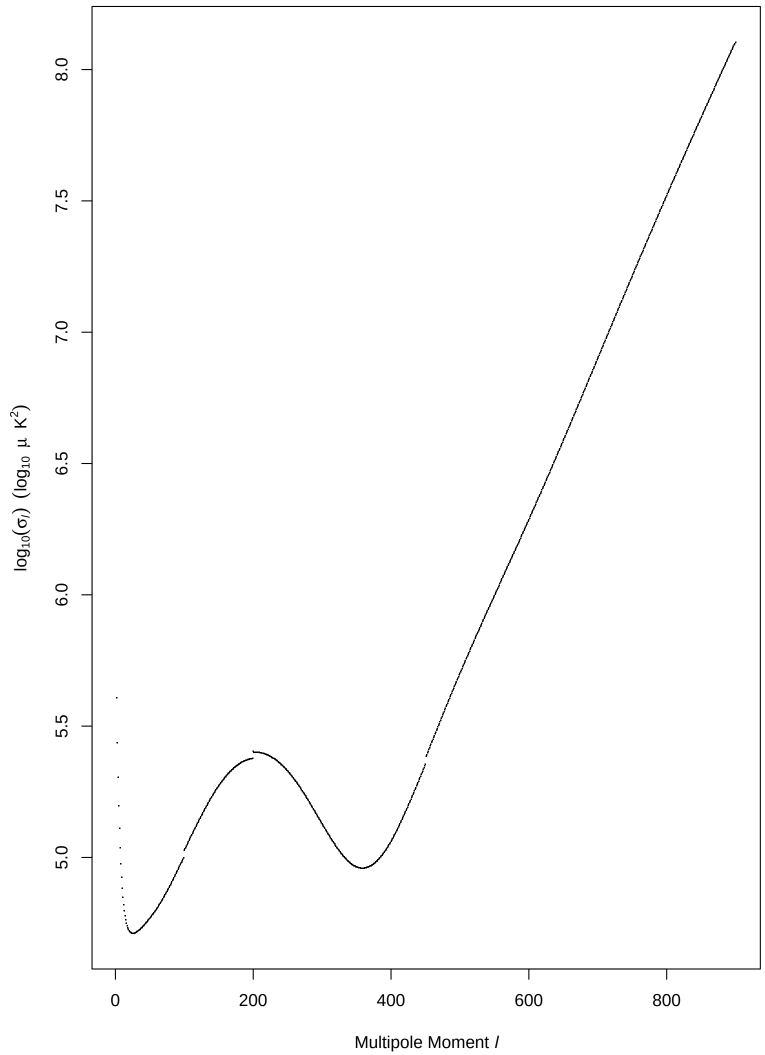
CMB Power Spectrum: WMAP Data





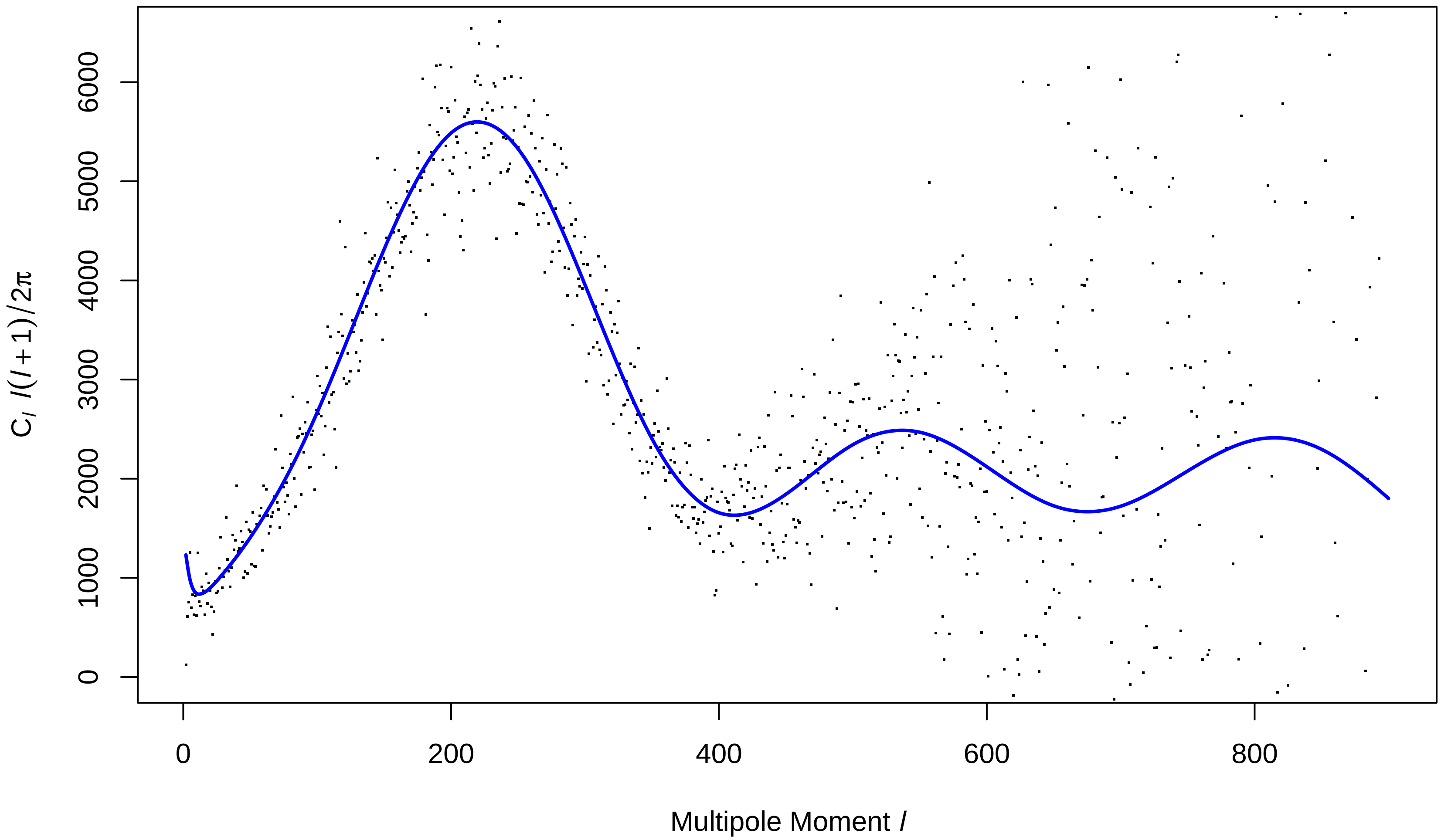
CMB Power Spectrum: WMAP Variances





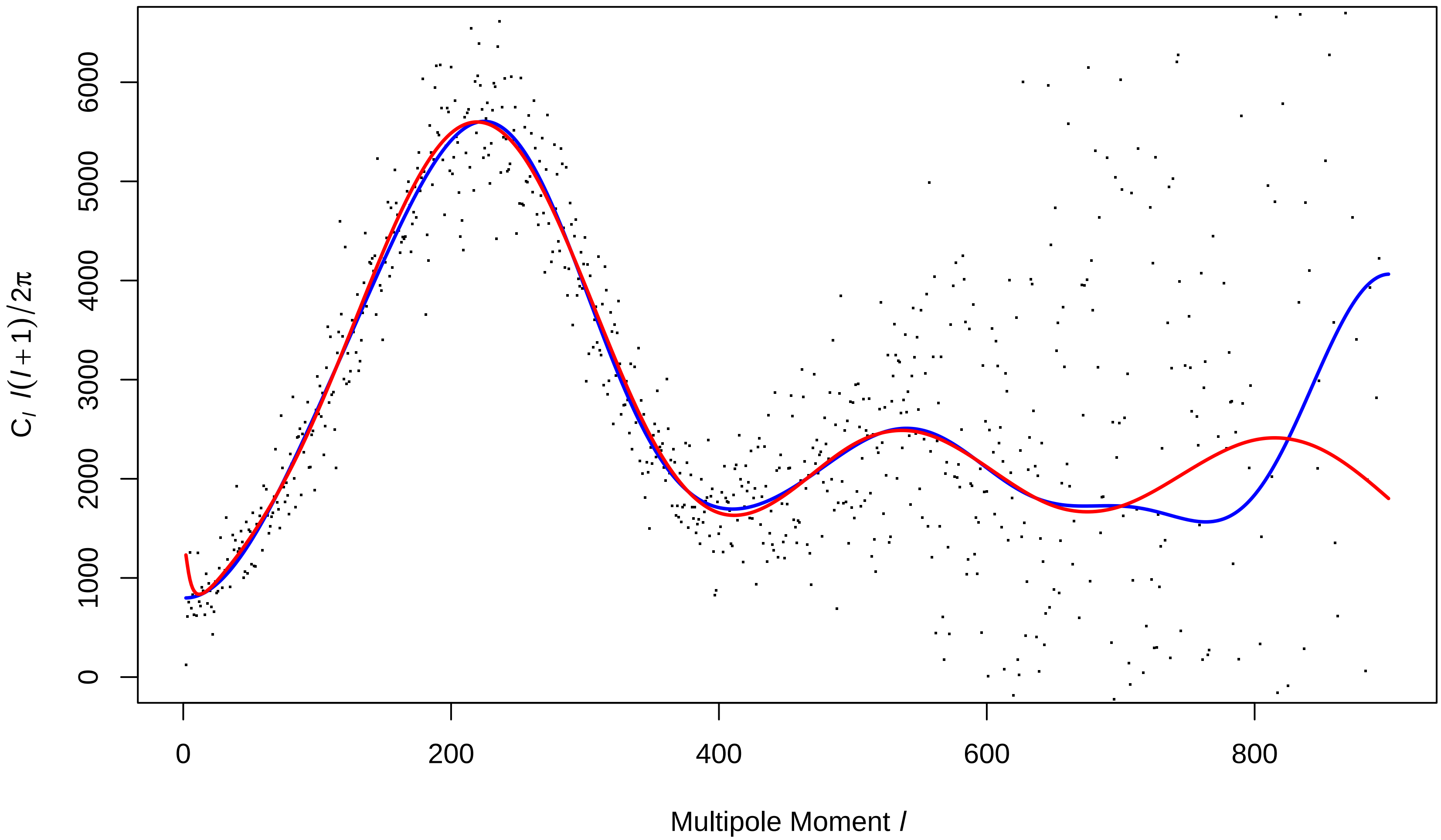
Noise correlated and heteroskedastic

Cosmological Models



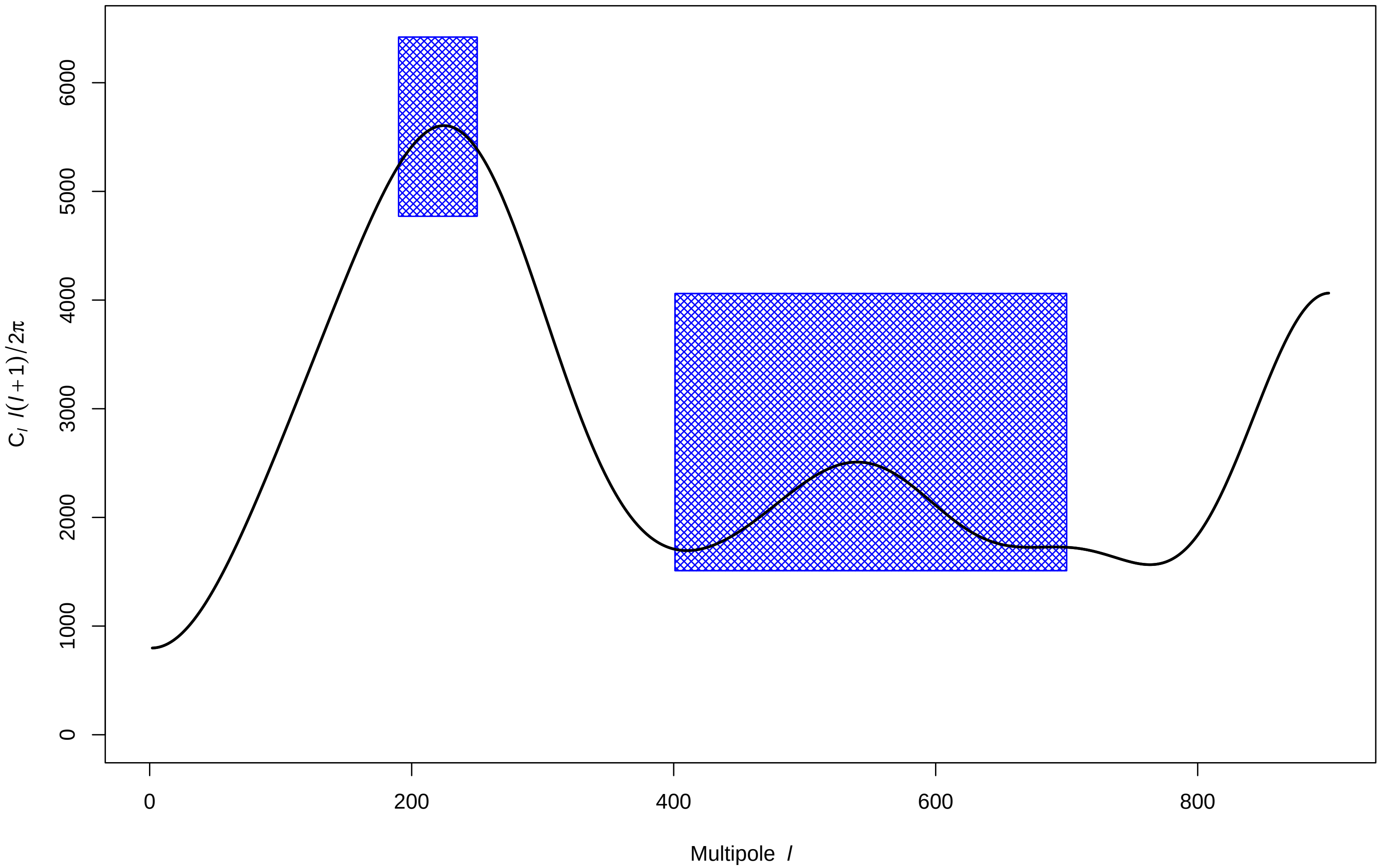
- 13/11/7-dimensional model maps cosmological parameters to spectra.
- Ultimate goal: inferences about these cosmological parameters.
- Subsidiary goal: identify location, height, widths of peaks

Confidence Ball Center vs Concordance Model



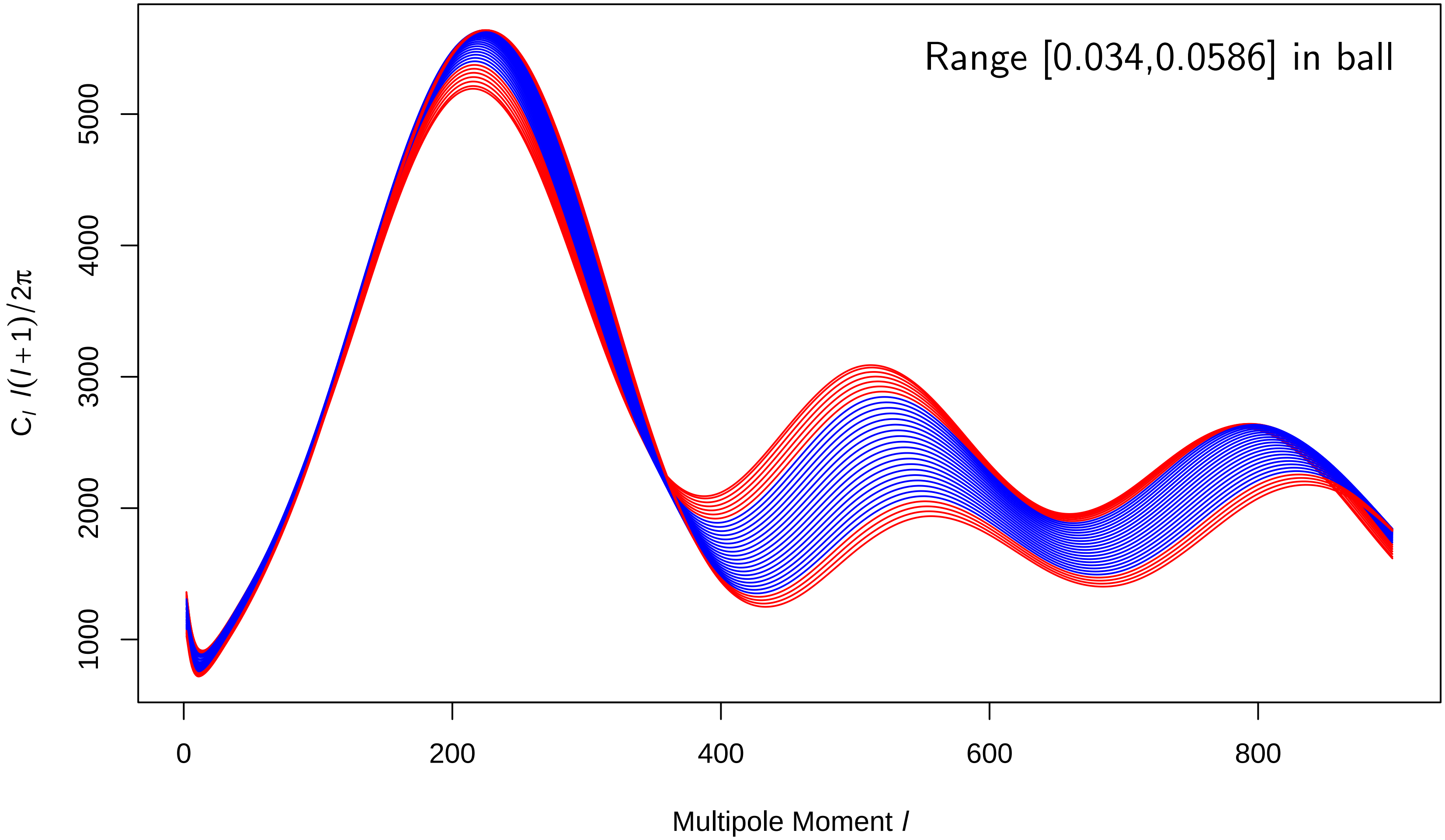
Eyes on the Ball I: Parametric Probes

Peak Heights, Peak Locations, Ratios of Peak Heights



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

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

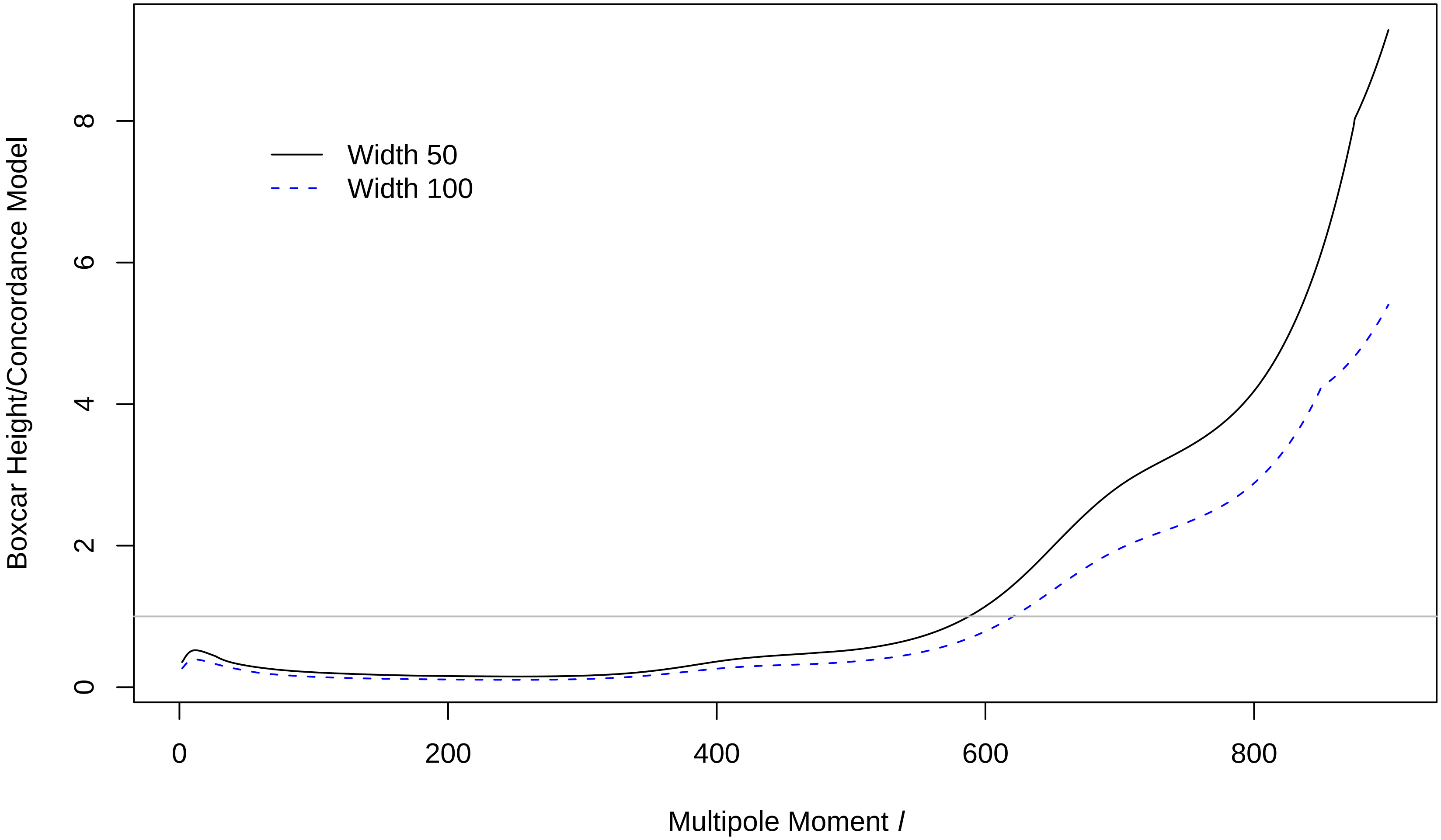


Extended search, over millions of spectra, in progress.

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

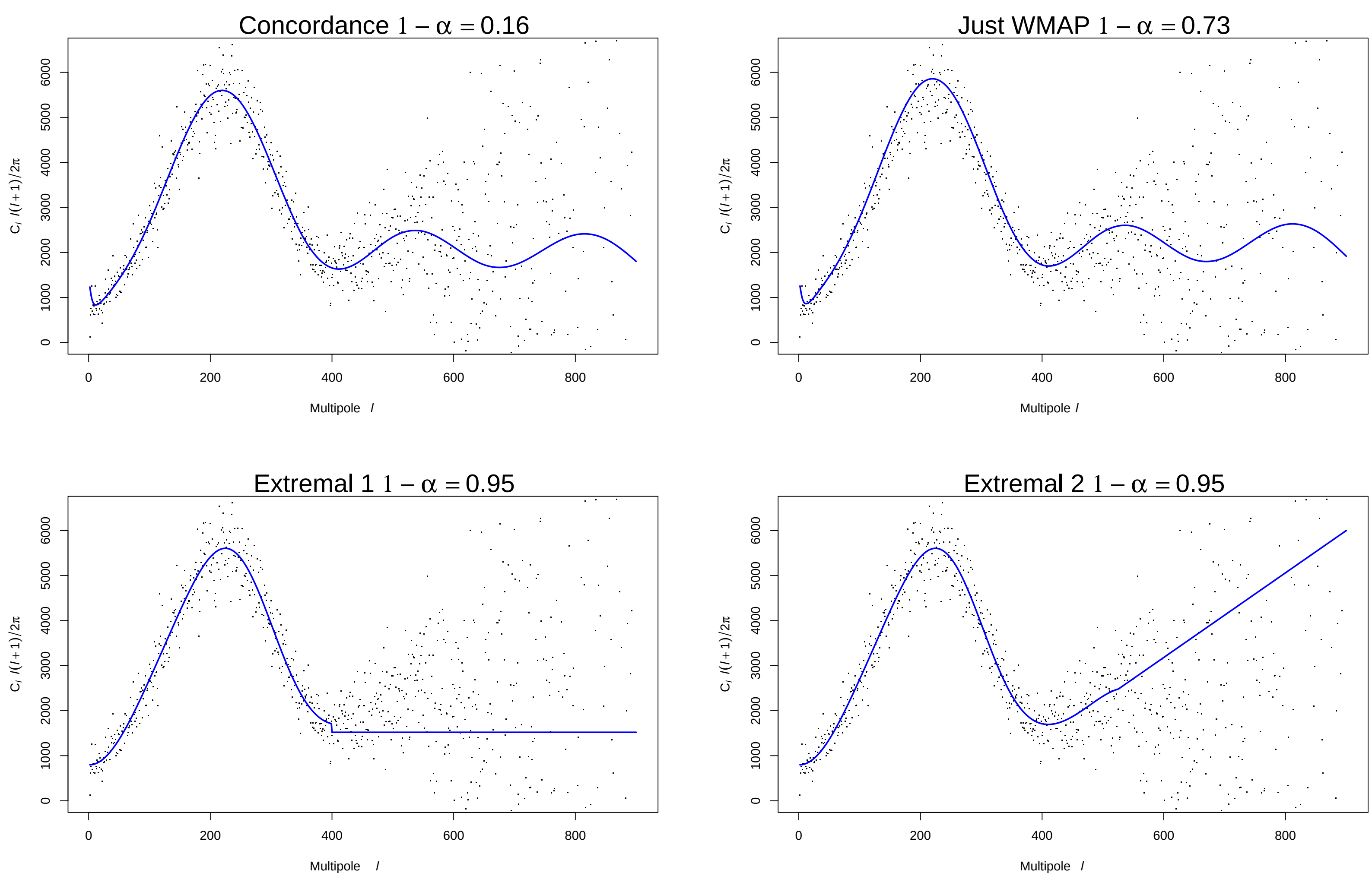
Probe from center with boxcars of given width centered at each ℓ.

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.



Eyes on the Ball III: Confidence Catalogs

- Our confidence set construction does not impose constraints based on prior knowledge.

  Instead: form ball first and impose constraints at will.

- Raises the possibility of viewing inferences *as a function* of prior assumptions.

  The confidence ball creates a mapping from prior assumptions to inferences; we call this a confidence catalog.

- Ex: Constraints on peak curvature over range defined by reasonable parametric models.

Take-Home Points

- Strategy
  - Build confidence sets for functions with uniform asymptotic (or finite-sample) coverage and post-hoc protection.
  - Constrain and search post-hoc

- Supports inferences on any set of functionals with simultaneous validity.

- Embedding parametric model in constrained nonparametric model gives flexibility when model is uncertain. Model checking.

- Balls and bands can both be useful, alone or in combination. Considering other metrics as well.