Using what we know:
Inference with Physical Constraints

Chad M. Schafer & Philip B. Stark
Department of Statistics
University of California, Berkeley
{cschafer,stark}@stat.berkeley.edu

PhyStat 2003
8 September 2003

Some work joint w/
Steve Evans (Berkeley)
Ben Hansen (Michigan)

Examples of Physical Constraints

Nonnegativity:

Counts, energies, densities; monotonicity

Bounds on functionals:

Energy in geomagnetic field,
rotation rate in solar interior (w/ C. Genovese)

Parametrizations:

Power law for CMB spectrum (w/ L. Tenorio, C. Lineweaver)

How can we use such constraints to reduce uncertainty?

Most texts don't treat constraints.

Examples.

Transparencies.

Seismic velocity in Earth's core (w/ Bob Parker)
aftershock probability density (w/ N. Hengartner)
SH coefficients of CMB and of geomagnetic field

Models and Parameters

Model specifies prob. distribution of data X.
Models indexed by θ; call θ and $\Pr_\theta$ "the model."

Physical Constraint: Know *a priori* that $\theta \in \Theta$

Parameter: image $f(\theta)$ of θ under a mapping f.

Building block: Bounded Normal Mean
Model $X \sim N(\theta, 1)$, with $\theta \in \Theta = [-\tau, \tau]$
Seek to estimate $f(\theta) = \theta$.

Confi dence Sets and Coverage

Confidence Set:
Random set S of parameter values: depends on X

Coverage probability $(1-\alpha)$:
Minimum chance $S(X)$ contains true $f(\theta)$.

$$\Pr_{\theta}\{S(X) \ni f(\theta)\} \geq 1-\alpha, \quad \forall \theta \in \Theta. \tag{1}$$

Duality between testing and confidence sets:
Invert family of level α tests $\Rightarrow$ $1-\alpha$ conf. set–
all $\theta' \in \Theta$ that aren't rejected.

Standard 95% conf. interval for normal mean:
$\mathcal{I} = [X - 1.96, X + 1.96]$.
Doesn't use the constraint $\theta \in [-\tau, \tau]$.

Procrustean (truncated) interval:
$\mathcal{I}_T = [X - 1.96, X + 1.96] \cap [-\tau, \tau]$.
Uses constraint, but is it best?

Common ways to add constraints

Ignore them. (Prof. Barlow claimed that's the only frequentist option)

Ad hoc; procrustean: Make unconstrained estimate; force it to be in the constraint set.

Bayesian: Use prior π that assigns probability 1 to the constraint set.

$$\Pr_{\pi}\{\theta \in \Theta\} = 1. \tag{2}$$

Frequentist minimax: Use estimator that (within some class) is minimax for some loss over the constraint set

Shortcomings

Ad hoc; procrustean: Can do “better.”

Bayesian: Where does the prior come from? No such thing as uninformative prior.
Sensitive to loss and assumptions

Frequentist minimax: Driven by worst case
Sensitive to loss and assumptions

Defi ning “best:” small expected size

Want a precise answer:

Minimize expected size of the confidence set.

Expected size depends on true value of θ: tradeoff.

Bayesian: minimize average (for prior π) expected size

Frequentist Minimax: minimize max expected size for $\theta \in \Theta$

Bayes/Minimax duality:

Minimax is Bayes for *least favorable prior*

Minimax expected size CI for bdd Normal mean

τ	standard $\mathcal{I}$		truncated standard $\mathcal{I} \cap [-\tau, \tau]$		Best meas. fixed-width[a] $\mathcal{I}_{\mathrm{N}}$		Opt. meas.[b] $\mathcal{I}_{\mathrm{OPT}}$
1.75	3.9	49%	2.9	10%	3.3	25%	2.6
2.00	3.9	38%	3.2	11%	3.3	16%	2.8
2.25	3.9	31%	3.4	13%	3.3	10%	3.0
2.50	3.9	26%	3.6	14%	3.3	6%	3.1
2.75	3.9	22%	3.7	15%	3.3	3%	3.2
3.00	3.9	21%	3.8	16%	3.3	1%	3.3
3.25	3.9	19%	3.8	16%	3.3	0%	3.3
3.50	3.9	18%	3.9	16%	3.5	5%	3.3 [c]
3.75	3.9	16%	3.9	15%	3.6	6%	3.4
4.00	3.9	14%	3.9	13%	3.6	5%	3.4

[a] These have form $[\hat{\theta}(X) - e, \hat{\theta}(X) + e]$, with $\hat{\theta}(\cdot)$ measurable and e constant.

[b] These have form $\{\theta \in \Theta : (\theta, X) \in S\}$, where $S \subseteq \Theta \times \mathcal{X}$ is product-measurable.

[c] Truncated Pratt interval $\mathcal{I}_{\mathrm{TP}}$ is optimal for $\tau \leq 3.29$. The entries in the rightmost column for $\tau = 3.50, 3.75$, and 4.00 are approximated numerically.

Size isn't everything: how you use it matters.

Cost of including $\theta' \neq \theta$ can depend on θ:

E.g., might sacrifice length to include values with one sign only.

Curvature of the universe?

Can allow size measure to depend on θ.

Numerics:

Approximate least-favorable π by Monte Carlo

Well suited to distributed/parallel computing.

Uses importance sampling.

Don't need closed-form likelihood.

Solve 2-player matrix game iteratively.

Example: Microwave cosmology–relic of the Big Bang

Model for Power Spectrum

$$\mathbf{X} \sim N\left(\mathbf{0},\ \mathbf{N} + \sum_{\ell=1}^{\infty} \left(\frac{2\ell+1}{4\pi}\right) C_\ell(\theta) B_\ell^2\, \mathbf{P}_\ell\right)$$

$\mathbf{P}_\ell$: the ℓ^{th} Legendre polynomial matrix

$\mathbf{N}$: noise covariance matrix

$\{C_\ell(\theta)\}$: power spectrum for the set of cosmological parameters θ

B_ℓ: transfer function of the observing filter.

Complicated relationship between cosmological "parameters" and spectrum $\{C_\ell\}$: nonlinear PDE.

Constraints on Parameters of Cosmological Model

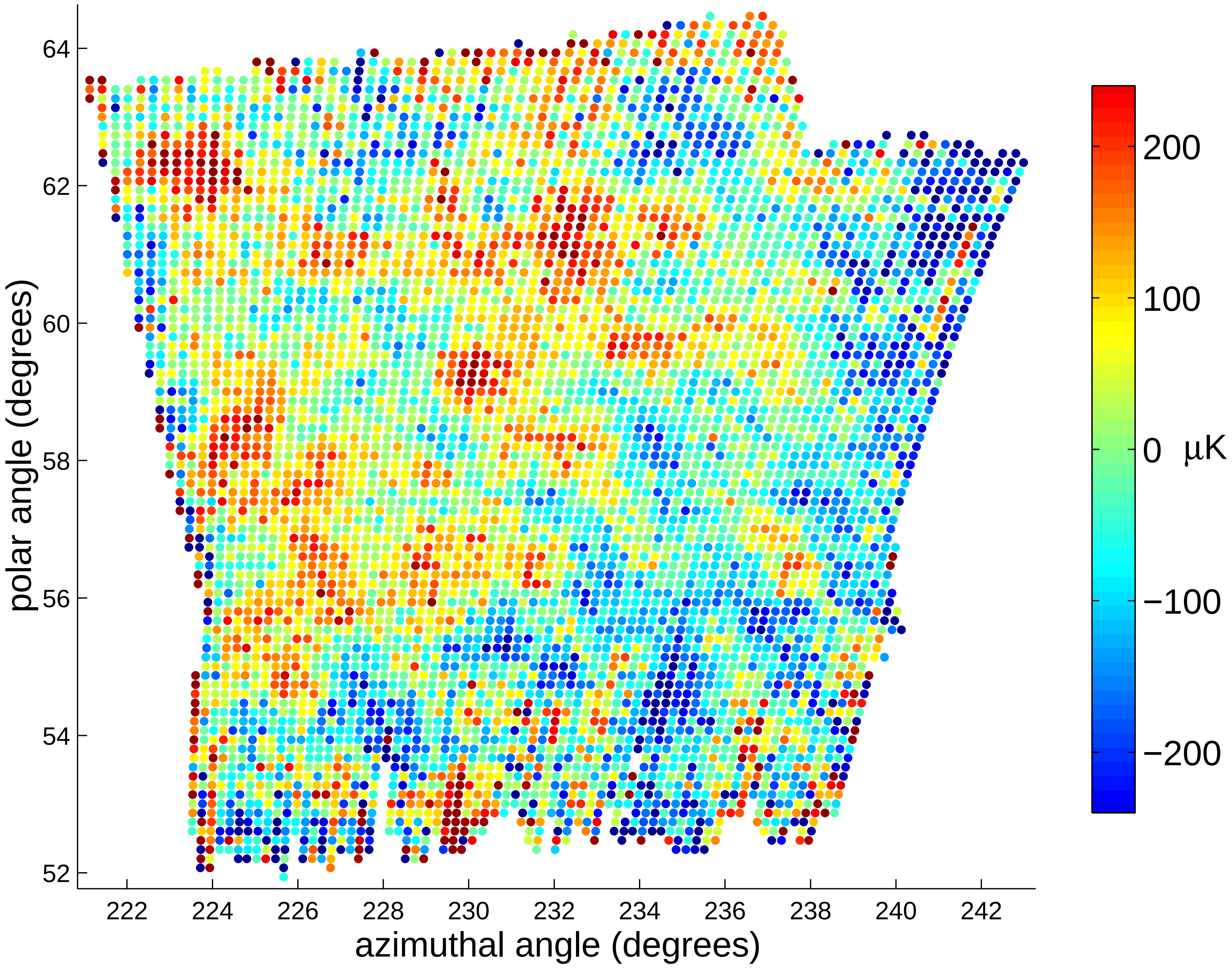
Parameter		Lower	Upper
Total Matter†	Ω_m	0.05	1.00
Baryonic Matter†	Ω_b	0.005	0.15
Cosmological Constant†	Ω_Λ	0.0	1.0
Hubble Constant (km s^{-1} Mpc^{-1})	H_0	40.0	90.0
Scalar Spectral Index	n_s	0.6	1.5
Optical Depth	τ	0.0	0.5

† Relative to critical density.

Monte-Carlo choses model uniformly, subject to

$$\Omega_b \leq \Omega_m \text{ and } 0.6 \leq \Omega_m + \Omega_\Lambda \leq 1.4. \quad (3)$$

MAXIMA-1 Data



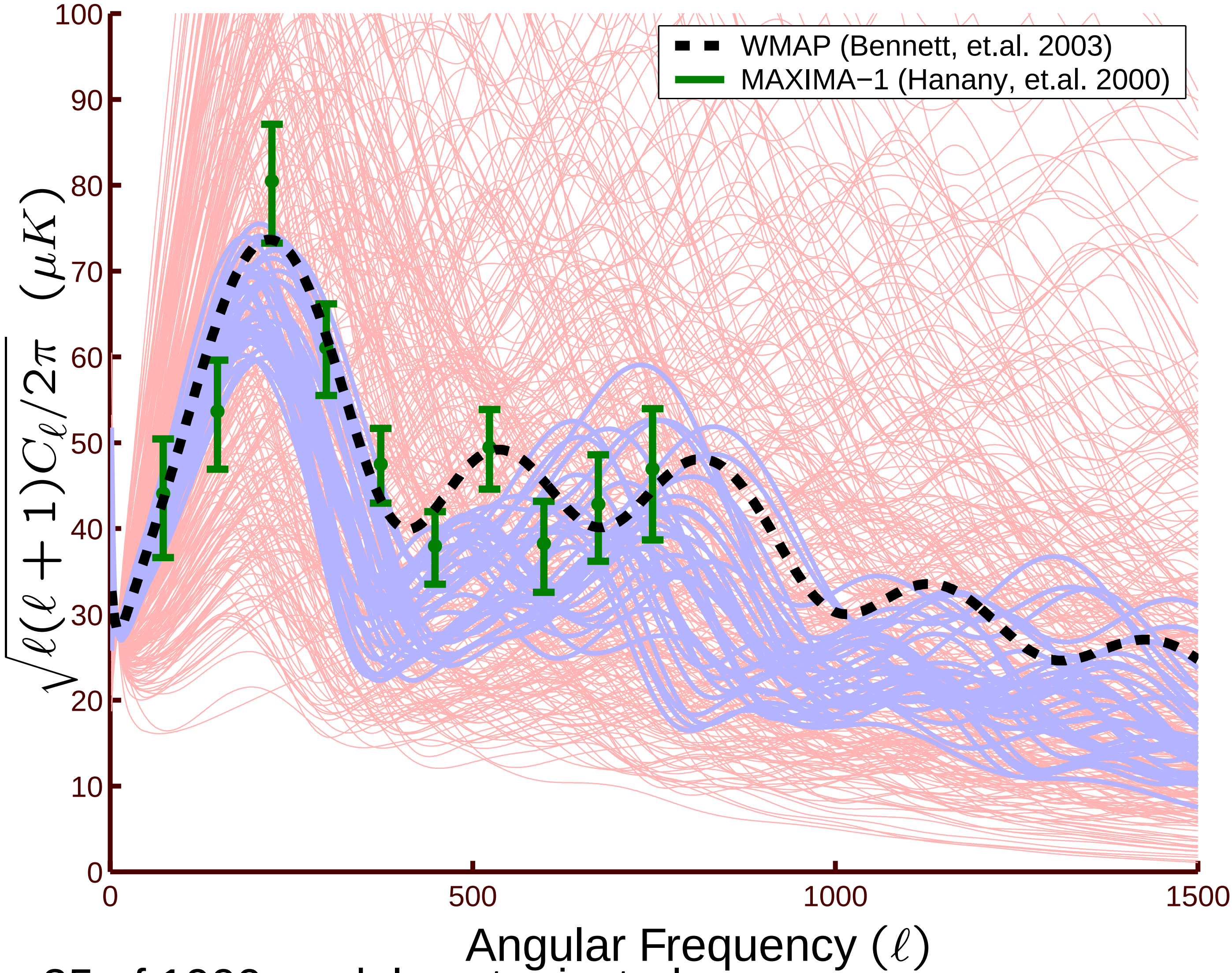
Data reduction/compression

Form 20,000 linear combinations of 5972 pixel temps.

Select 5972 linear combinations–eigenvectors of co-variance matrix for reference model.

Select 2000 of those using estimated Kullbach-Leibler divergence to approximate false coverage prob.

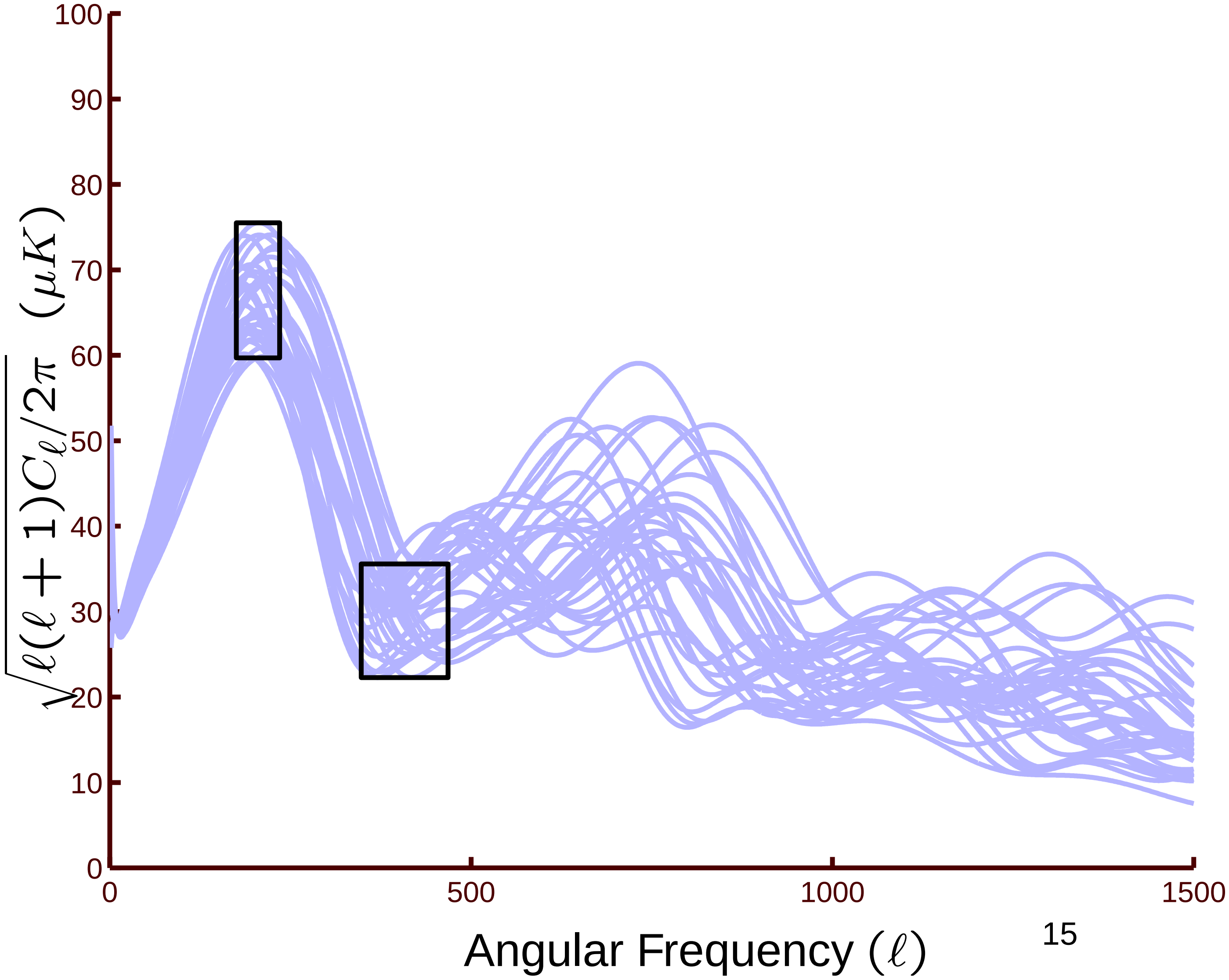
Results



35 of 1000 models not rejected.

Spectra of 300 rejected models shown.

Results



100
90
80
70
60
50
40
30
20
10
0
0
500
1000
1500
$\sqrt{\ell(\ell+1)C_\ell/2\pi}\ \ (\mu K)$
Angular Frequency (ℓ)

Results (contd)

Location of main peak (ℓ) : $(175, 235)$.
Height of main peak (μK) : $(59.7, 75.5)$.

Location of first valley (ℓ) : $(348, 468)$.
Height of first valley (μK) : $(22.3, 35.6)$.

Results (contd)

Some Accepted Models:

	Ω_b	Ω_m	Ω_Λ	τ	H_0	n_s	Ω
	0.042	0.674	0.241	0.317	77.00	1.117	0.915
	0.091	0.620	0.319	0.346	65.93	1.095	0.939
	0.104	0.684	0.316	0.152	53.00	0.960	1.000
	0.081	0.540	0.526	0.000	77.15	0.809	1.066
	0.072	0.754	0.328	0.116	86.55	0.970	1.082
	0.134	0.940	0.161	0.466	66.83	1.038	1.101
	0.085	0.301	0.815	0.058	62.21	0.729	1.116
	0.096	0.479	0.730	0.428	87.64	1.043	1.209
	0.096	0.555	0.693	0.260	81.28	0.923	1.248
	0.093	0.708	0.551	0.000	76.96	0.855	1.259
	0.139	0.667	0.623	0.269	61.15	0.954	1.290
	0.133	0.692	0.642	0.068	41.47	0.846	1.334
Min incl	0.011	0.058	0.082	0.000	41.47	0.729	0.915
Max incl	0.139	0.994	0.988	0.466	89.24	1.151	1.334
Min poss	0.005	0.000	0.000	0.000	40.00	0.600	0.600
Max poss	0.150	1.000	1.000	0.500	90.00	1.500	1.400

References

Balbi, A., et al., 2000. Constraints on Cosmological Parameters from MAXIMA-1, *Astrophys. J.*, *545*, L1–L4

Bennett, C.L., et al., 2003. First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Maps and Basic Results, *Ap. J.*, in press.

Evans, S.N., B. Hansen, and P.B. Stark, 2002. Minimax Expected Measure Confidence Sets for Restricted Location Parameters. Tech. Rept. 617, Dept. Statistics Univ. Calif Berkeley.

Hanany, et al., 2000. MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on angular scales of 10 arcminutes to 5 degrees, *Astrophys. J.*, *545*, L5.

Jaffe, et al., 2001. Cosmology from Maxima-1, Boomerang and COBE/DMR CMB Observations, *Phys. Rev. Lett.*, *86*, 3475–3479.

Kempthorne, P.J., 1987. Numerical specification of discrete least favorable prior distributions, *SIAM J. sci. statis. Comp.*, *8*, 171–184.

Nelson, W., 1966. Minimax solution of statistical decision problems by iteration, *Ann. math. Statis.*, *37*, 1643–1657.

Pratt, J.W., 1961. Length of Confidence Intervals, *J. amer. statis. Assoc.*, *56*, 549–567.

Rao, C.R., 1962. Efficient estimates and optimum inference procedures, *J. roy. statis. Soc.*, *24 (B)*, 46–72.

Robinson, J., 1951. An iterative method for solving a game, *Ann. Math.*, *54*, 296–301.