Likelihood Analysis (A Theorist looks at Measurements)

David Spergel

Statistics of Signal

$$T(\hat{n}_i) = \sum_{lm} a_{lm} Y_{lm}(\hat{n}_i)$$

$$< a_{lm} a^*_{l'm'} >= c_l \delta_l^{l]} \delta_m^{m'}$$

$$\begin{aligned} S_{ij} \equiv =< T(\hat{n}_i) T(\hat{n}_j) > &= \sum_{lm} \sum_{l'm'} < a_{lm} a_{l'm'} > Y_{lm}(\hat{n}_i) Y_{lm}(\hat{n}_j) \\ &= \sum_l c_l \sum_m Y_{lm}(\hat{n}_i) Y^*_{lm}(\hat{n}_j) \\ &= \sum_l c_l \frac{(2l+1)}{4\pi} P_l(\hat{n}_i \cdot n_j) \end{aligned}$$

$$p(\vec{T}) = \frac{d\vec{T}}{\sqrt{(2\pi)^{n_{pix}} \det \mathbf{S}}} \exp\left[-\frac{1}{2} (\vec{T})^T \mathbf{S}^{-1} \vec{T} \right]$$

Spherical Harmonic Basis

$$\mathbf{Y} = Y_{lm}(\hat{n}_j)$$

$$n_{pix} \times n_{multipoles}$$

$$\mathbf{YSY}^* = \begin{pmatrix} c_0 & 0 & 0 \\ 0 & \cdot & 0 \\ 0 & 0 & c_l \end{pmatrix}$$

$$\det \mathbf{S} = \sum_l (2l+1) c_l$$

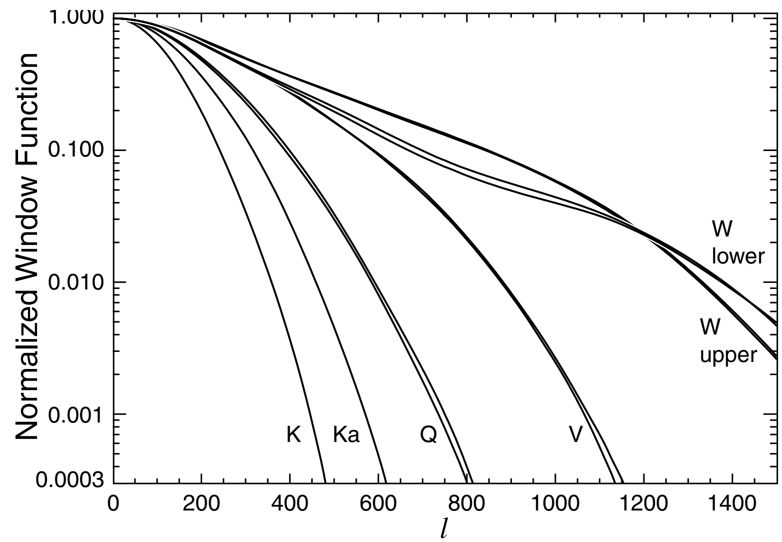
$$p(\vec{T}) = \Pi_{lm} \frac{\exp\left[-\frac{1}{2}\frac{a_{lm}a^*_{lm}}{c_l}\right]}{\sqrt{2\pi c_l}}$$

Beams

$$t_s(\hat{n}) = \int d^2 n T(\hat{n}') W(\hat{n} - \hat{n}')$$

$$t_s(\hat{n}) = \sum_{lm} a_{lm} w_l Y_{lm}(\hat{n}_i)$$

$$S_{ij} = \sum_l c_l w_l^2 \frac{(2l+1)}{4\pi} P_l(\hat{n}_i \cdot n_j)$$



3

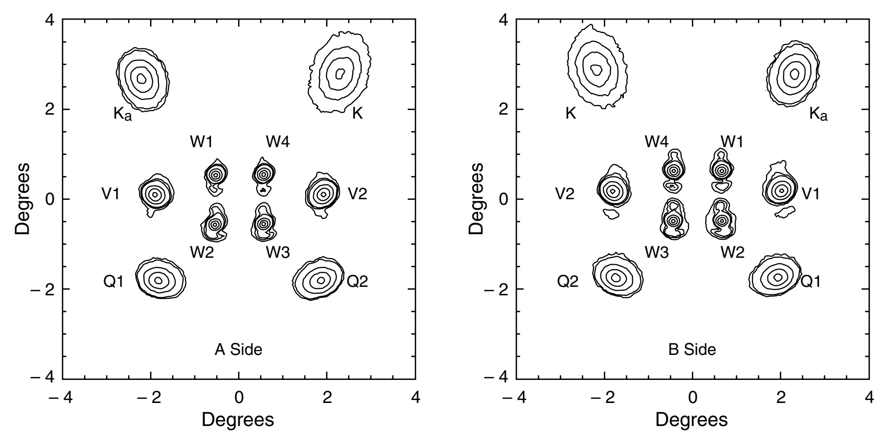


FIG. 1.— Jupiter maps of the A and B side focal planes (Bennett et al. 2003c) in the reference frame of the observatory. The contour levels are at 0.9, 0.6, 0.3, 0.09, 0.06, 0.03 of the peak value. W1 and W4 are the "upper" W-band radiometers. In W band, the lobes at the 0.09 contour level (≈ −10 dB) and lower are due to surface deformations.

Jupiter

Noise: Combining Two Gaussian Fields

$$\vec{t} = \vec{t}_s + \vec{n}$$

$$< n_i n_j >= N_{ij}$$

$$< n_i n_j >= \sigma_i^2 \delta_{ij}$$

$$p(n_i|N_{ij}) = p(\vec{t}|\vec{t}_s, \mathbf{N}) = \frac{d\vec{n}}{\sqrt{(2\pi)^{n_{pix}} \det \mathbf{N}}} \exp\left[-\frac{1}{2}(\vec{n})^T \mathbf{N}^{-1}(\vec{n})\right]$$

$$\begin{aligned} p(\vec{t}|c_l) &= p(\vec{t}|\vec{t}_s, \mathbf{N}) p(\vec{t}_s|c_l) \\ &= \int \frac{d\vec{t}_s}{\sqrt{(2\pi)^{n_{pix}} \det \mathbf{S}}} \exp\left[-\frac{1}{2}(\vec{t}_s)^T \mathbf{S}^{-1} \vec{t}_s\right] \\ &\quad \times \frac{1}{\sqrt{(2\pi)^{n_{pix}} \det \mathbf{N}}} \exp\left[-\frac{1}{2}(\vec{t}_s - \vec{t})^T \mathbf{N}^{-1}(\vec{t}_s - \vec{t})\right] \\ &= \frac{d\vec{t}}{\sqrt{(2\pi)^{n_{pix}} \det(\mathbf{S} + \mathbf{N})}} \exp\left[-\frac{1}{2}(\vec{t})^T (\mathbf{S} + \mathbf{N})^{-1} \vec{t}\right] \end{aligned}$$

CMB Map Making

$$y_i = M_{ij} t_j^s + n_j$$

$$< n_j n_{j'} >= \mathcal{N}(t_j - t_{j'})$$

$$\mathcal{L}(t_j^s | y_i) \propto \exp\left[-\frac{1}{2}(y_i - M_{ij} t_j^s)^T \mathcal{N}^{-1} (y_i - M_{ij} t_j^s)\right]$$

$$\frac{\partial \ln \mathcal{L}}{\partial t^s} = 0$$

$$\left(\mathbf{M} \mathcal{N}^{-1} \mathbf{M}\right) \vec{t_s} = \mathbf{M}^T \vec{y}$$

Pixel Noise Matrix

$$\mathbf{N}^{-1} = \frac{\partial^2 \ln \mathcal{L}}{\partial (t^s)^2} = \left(\mathbf{M}\mathcal{N}^{-1}\mathbf{M}\right)$$

$$\mathbf{N}^{-1}\vec{t}_s = \mathbf{M}^T \vec{y}$$

$$\begin{aligned} p(\vec{t}|c_l) &= \frac{d\vec{t}}{\sqrt{(2\pi)^{n_{pix}} \det(\mathbf{S}+\mathbf{N})}} \exp\left[-\frac{1}{2}(\vec{t})^T(\mathbf{S}+\mathbf{N})^{-1}\vec{t}\right] \\ &= \frac{d\vec{t}\exp\left[-\frac{1}{2}(\vec{t})^T\mathbf{N^{-1}}(\mathbf{N^{-1}SN^{-1}}+\mathbf{I})^{-\mathbf{1}}\mathbf{N^{-1}}\vec{\mathbf{t}}\right]}{\sqrt{(2\pi)^{n_{pix}} \det(\mathbf{N^{-1}}(\mathbf{N^{-1}SN^{-1}}+\mathbf{I})\mathbf{N^{-1}})}} \end{aligned}$$

Marginalization

$$y_i = M_{ij} t^s_j + n_j + g M_{ij} d_j$$

$$\mathcal{L}(t^s_j | y_i) \propto \int dg p(g) \exp\left[-\frac{1}{2}(y_i - M_{ij} t^s_j - g M'_{ij} d_j)^T \mathcal{N}^{-1} (y_i - M_{ij} t^s_j - g M'_{ij} d_j)\right]$$

$$\mathbf{N}^{-1} = \left(\mathbf{M} \mathcal{N}^{-1} \mathbf{M}\right) - \vec{q} \otimes \vec{q}$$

Null Tests

- Essential to combine data in ways that yield zero expected signal so to test noise model
  - year 1 - year 2
  - channels at common frequency
  - channels at different frequencies

Quadratic Estimator

$$c_l^{meas} = \frac{1}{2l+1}\sum_m a_{lm}a_{lm}^*$$

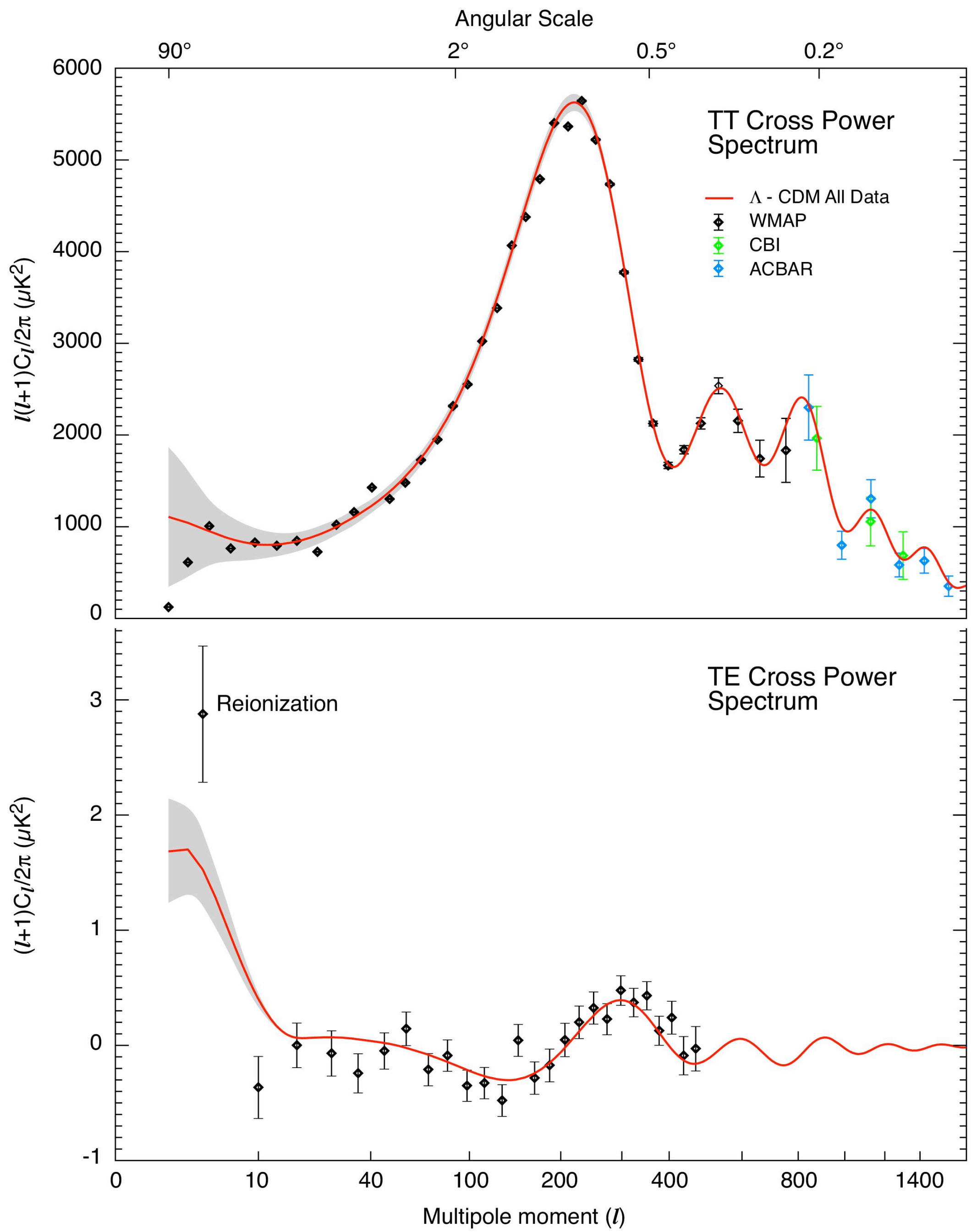
$$p(c_l^{meas}|c_l) = \Pi_m \int \frac{da_{lm}}{(2\pi)^{1/2}c_l} \exp\left(-\frac{a_{lm}a_{lm}^*}{2c_l}\right)\delta(\sum_m a_{lm}a_{lm}^* - (2l+1)c_l^{meas})$$

Exact Likelihood $$p(c_l^{meas}|c_l) = \frac{\exp\left[-\frac{(2l+1)c_l^{meas}}{2c_l}\right]}{(2\pi c_l)^{(2l+1)/2}}$$

Approximate Likelihood $$p(c_l^{meas}|c_l) = \frac{\exp\left[-(c_l - c_l^{meas})^2 2\sigma_l^2\right]}{\sqrt{2\pi}\sigma_l}$$

$$\sigma_{c_l}^2 = \frac{2}{2l+1}c_l^2$$

$$\sigma_{c_l}^2 = \frac{2}{2l+1}(c_l + n_0^2 w_l^{-2})^2$$


Angular Scale
90°
2°
0.5°
0.2°
6000
5000
4000
3000
2000
1000
0
l(l+1)C_l/2π (μK²)
TT Cross Power Spectrum
Λ - CDM All Data
WMAP
CBI
ACBAR
3
2
1
0
-1
(l+1)C_l/2π (μK²)
Reionization
TE Cross Power Spectrum
0
10
40
100
200
400
800
1400
Multipole moment (l)

Maximum Likelihood and Fisher Matrix

$$\mathcal{L}(T|S) = \frac{d\vec{t}}{\sqrt{(2\pi)^{n_{pix}} \det(\mathbf{S}+\mathbf{N})}} \exp\left[-\frac{1}{2}(\vec{t})^T(\mathbf{S}+\mathbf{N})^{-1}\vec{t}\right]$$

Oh, Spergel,Hinshaw

$$\frac{\partial \mathcal{L}}{\partial c_l} = 0$$

$$F_{ll'} = \frac{2}{(2l+1)(2l'+1)} \mathrm{tr}((S+N)^{-1}P_l(S+N)^{-1}P_{l'})$$

$$F_{ll} \simeq \frac{2}{(2l+1)f_{sky}^2}(c_l + n_0^2 w_l^{-2})^2$$

Evaluating Models

$$p(\text{parameters}|\text{data}) = p(\text{data}|\text{parameters})P_{prior}(\text{parameters})$$

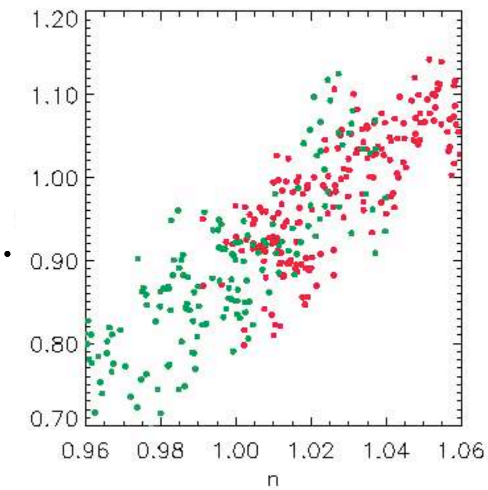
$$< A_i >= \int d\vec{A}A_i p(\vec{A}|\vec{T})$$

Relative Likelihood

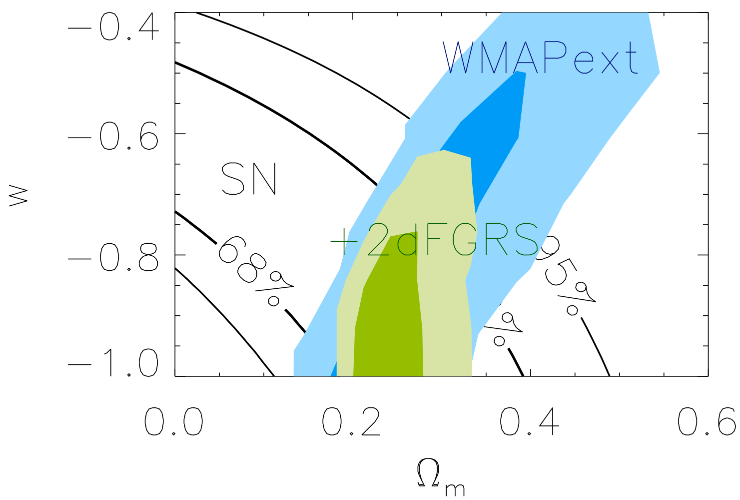
$$\frac{p_a(\text{data}|\text{MaximumLikelihoodparameters})}{p_b(\text{data}|\text{MaximumLikelihoodparameters})}$$

Markov Chains

- Need to evaluate the likelihood in a multi-dimensional parameter space
- General rule for integration: N < 3, use usual integration rules. N > 3, use Monte Carlo integration
- Markov Chain: start anywhere in the parameter space. Try taking a random step. If the step improves the likelihood, always take it. If it lowers the likelihood, select a random number, x, between 0 and 1. If x < exp(-(Lold - Lnew)), take the step. Otherwise, try a new step.
  - The likelihood chain will ***eventually*** fairly sample the likelihood surface
  - Run multiple chains long enough (100,000 steps) and compare the answers between the chains.



Cosmological Parameters



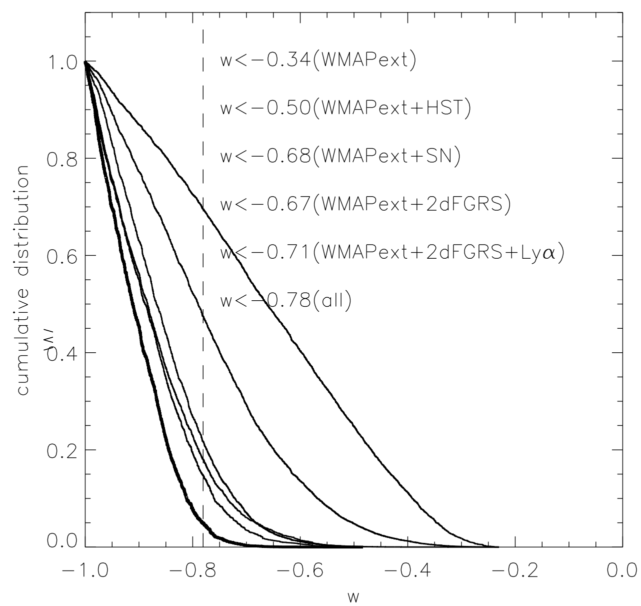
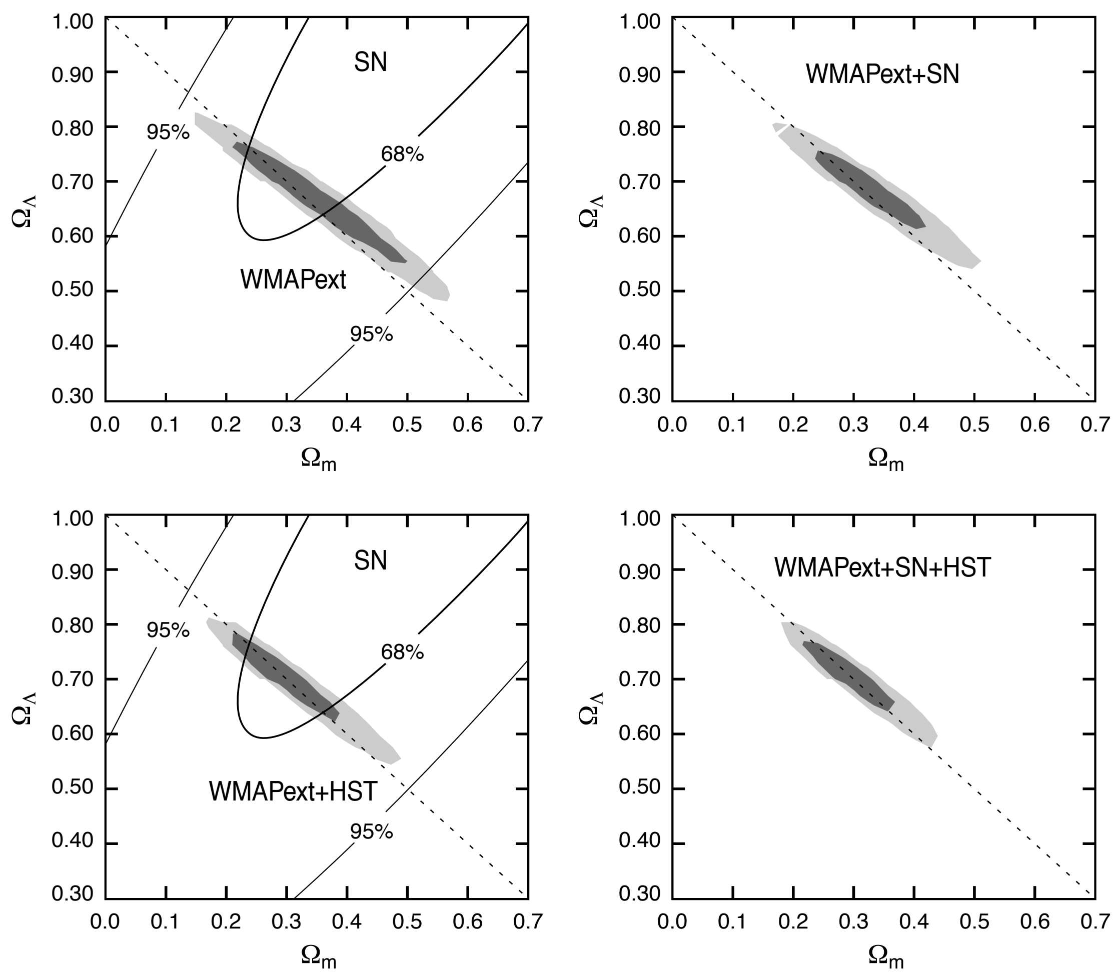


Fig. 12.— This figure shows the marginalized cumulative probability of w for several different combinations of data sets. The dashed line shows the 95% confidence upper limit on w based on combining all of the data sets. The values quoted in the captions are the 95% upper limit for various combinations of data sets. All combinations favor models where the dark energy behaves like a cosmological constant $w = -1$.

Priors, Degeneracies



Summary

- Bayesian likelihoods are a powerful language for describing experiments, prior knowledge and calculating the statistical implications of the results.
- They codify our assumption and uncertainties.

Problem

WMAP measured a quadruple of 125 $(uK)^2$. (1) What is the probability of measuring a value this small in a theory with $c_2 = 1000$? (2) Compare the likelihood of two cosmological models a running spectral index model (which predicts $c_2 = 700$) and a scale invariant model (which predicts $c_2 = 1100$) (3) Compare the likelihoods for a Gaussian model