Statistical techniques for data analysis in Cosmology

arXiv:0712.3028; arXiv:0911.3105

Licia Verde

ICREA & ICC UB-IEEC

http://icc.ub.edu/~liciaverde

outline

- Lecture 1: Introduction Bayes vs Frequentists, priors, the importance of being Gaussian, modeling and statistical inference, some useful tools.
- Lecture 2: Monte Carlo methods. Different type of errors. Going beyond parameter fitting.
- Lecture 3: forecasting: Fisher matrix approach. Introduction to model selection.

Handy explanations. Conclusions.

Statistical vs systematic errors

Statistics can tell you how to deal with statistical errors

As a data set grows, the statistical errors shrink; systematic errors do not shrink

You've got a problem.

Rumsfeld can help:

There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.

Donald Rumsfeld

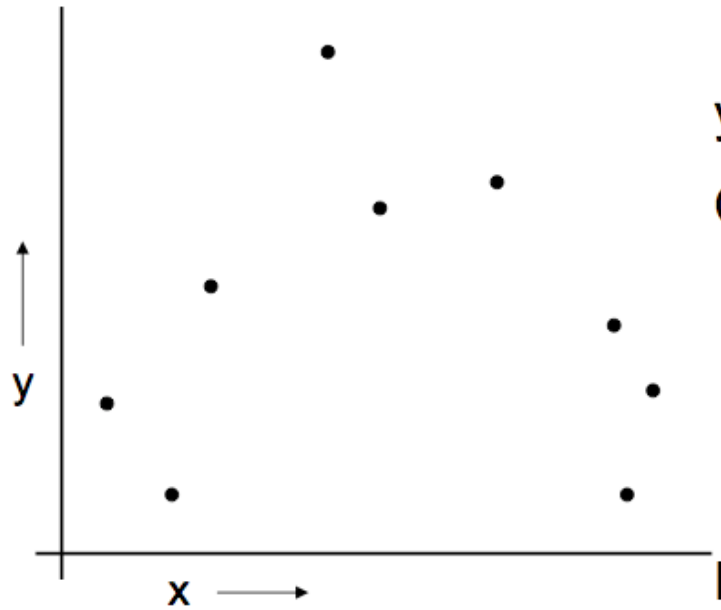
Jokes aside: some interesting literature has appeared in the past 2 yr…

How do you know you are not fitting the noise?

Example of powerful technique

Cross Validation is a powerful technique to make sure one is not fitting the noise

A Regression Problem



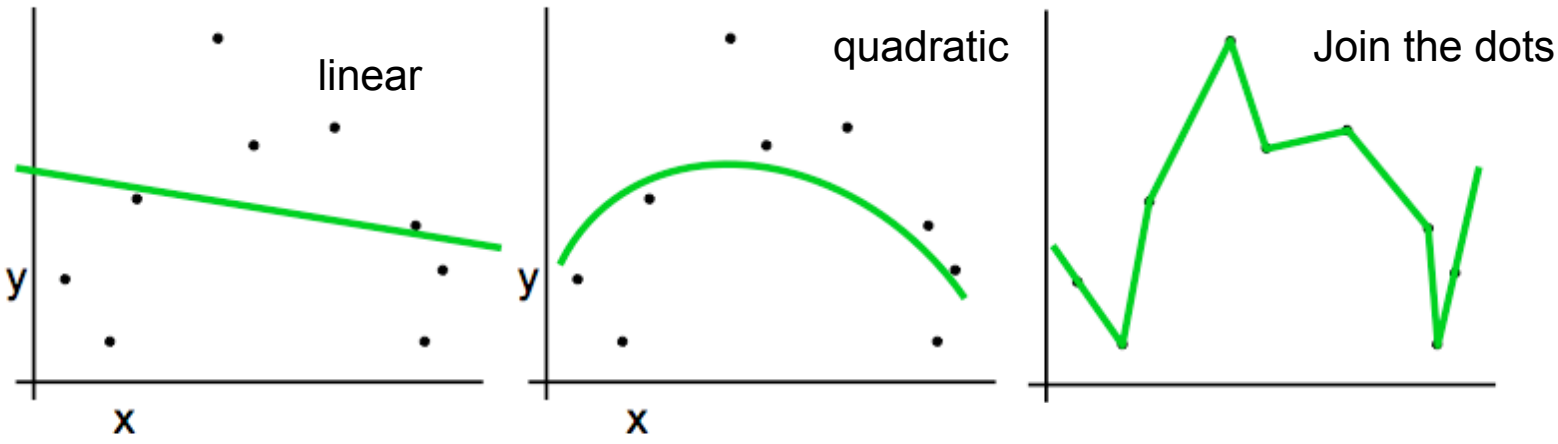
y = f(x) + noise

Can we learn f from this data?

Let's consider three methods…

Copyright ©Andrew W. Moore

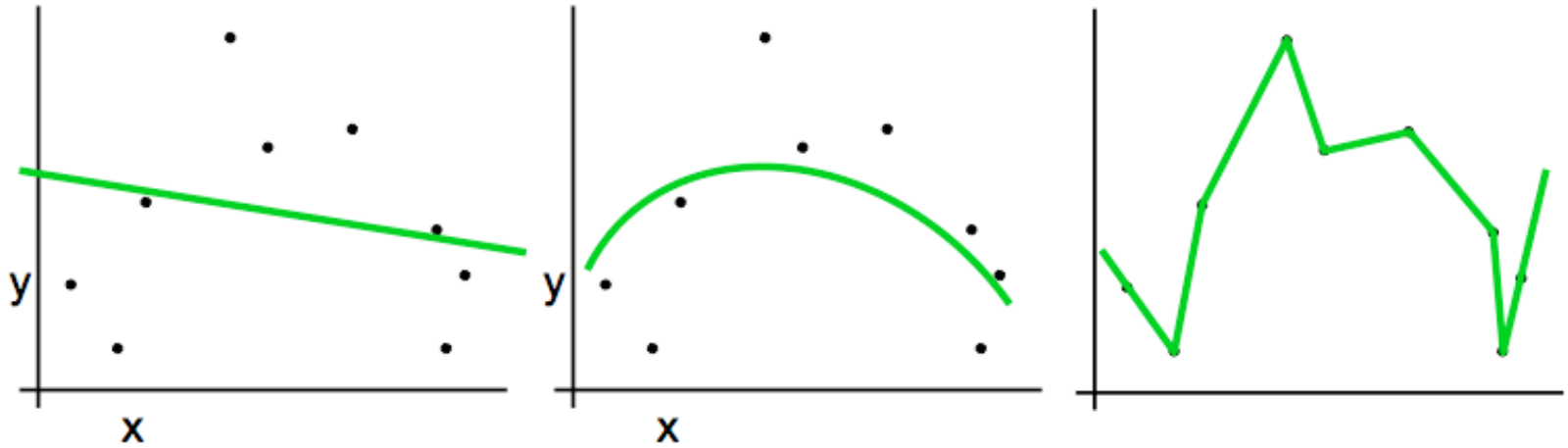
Which is best?



Why not choose the method with the best fit to the data?

Copyright ©Andrew W. Moore

What do we really want?



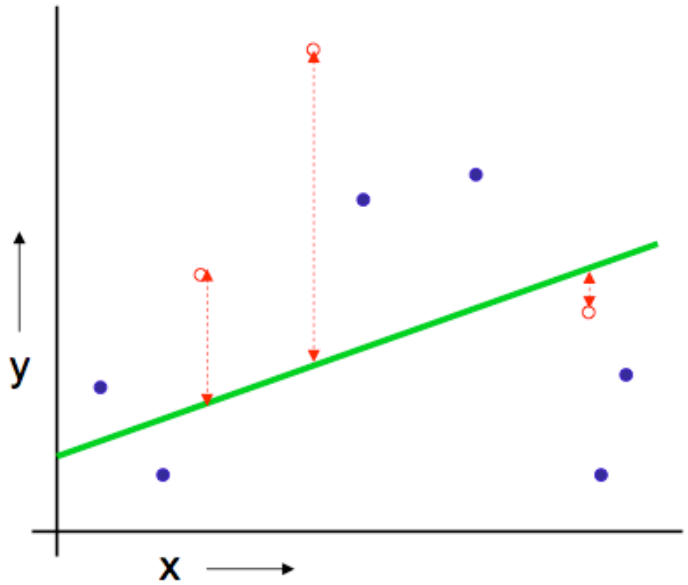
Why not choose the method with the best fit to the data?

"How well are you going to predict future data drawn from the same distribution?"

Copyright ©Andrew W. Moore

The training/test set approach is similar to leave one out CV.

Train on subset of the data



(Linear regression example)

Could be all-1

1 Randomly choose
x % of the data to be in a test set

2. The remainder is a training set

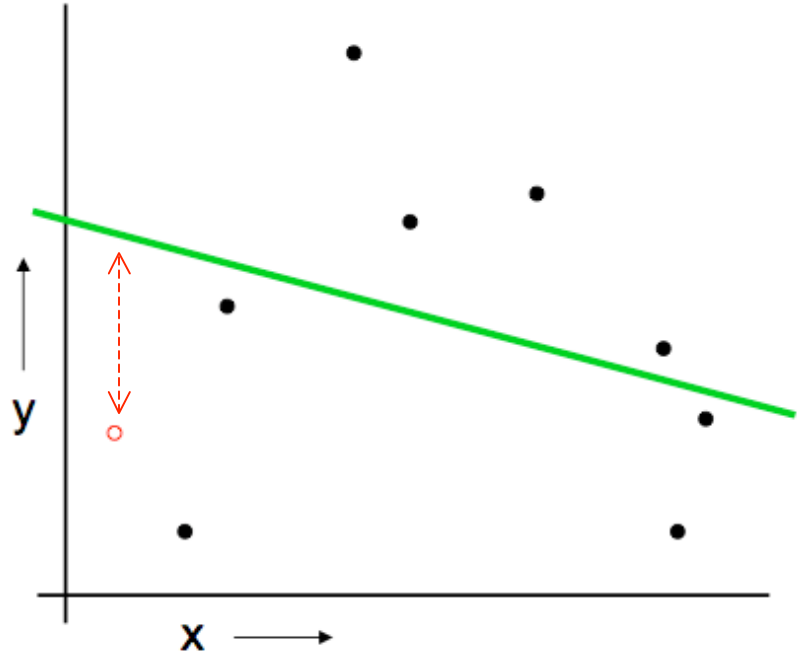
3. Perform your regression on the training set

4. Estimate your future performance with the test set

CV score

(this may remind you of training sets for photo-z)

Leave one out cross validation:



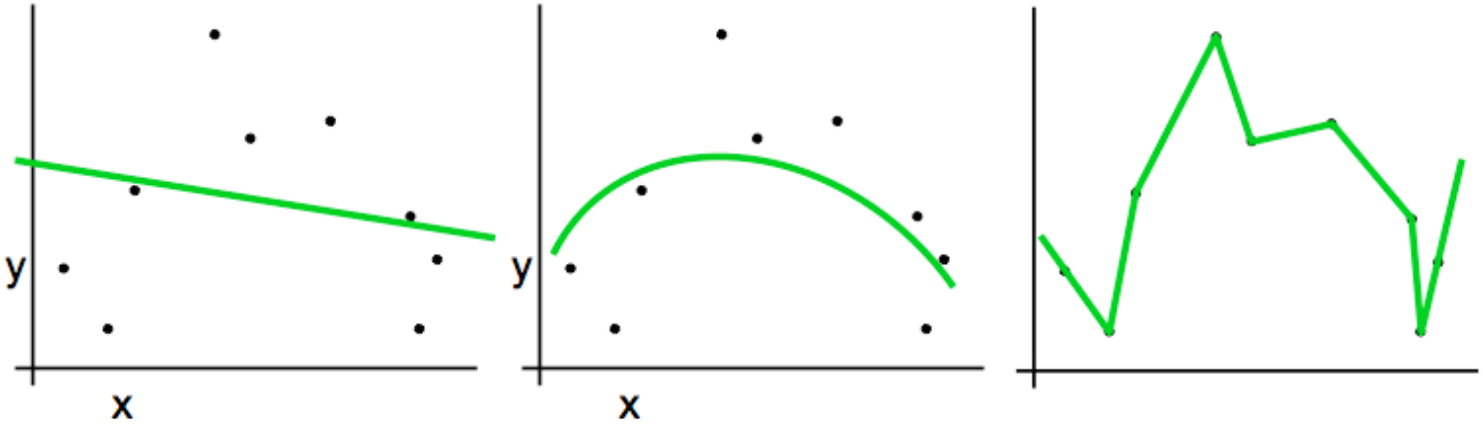
For k=1 to R

1. Let (x_k, y_k) be the kth record
2. Temporarily remove (x_k, y_k) from the dataset
3. Train on the remaining R-1 datapoints (example shown for linear model)
4. Note your error (x_k, y_k)

When you've done all points, report the mean error(CV score)

Copyright ©Andrew W. Moore

In this example:



MSE_{LOOCV} = 2.12

MSE_{LOOCV} =0.962

MSE_{LOOCV} =3.33

CV score

Leave one out CV is the ideal: does not waste much data
but it is very expensive; loong discussion about which CV is best…

Copyright ©Andrew W. Moore

Beyond parameter fitting

Going minimally parametric

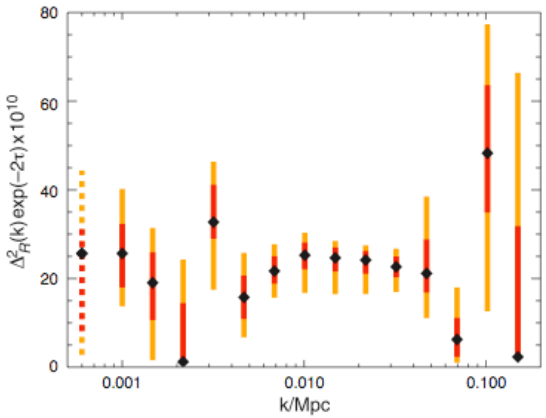
Working example: the shape of the primordial P(k)

Parameter fitting e.g., : $P(k)= A\ (k/k0)^{n-1}$

Other popular options are:

Instead of fitting a function to the data, use a basis function (wavelets, principal components etc…)

Use bins



Spergel et al 07

Piecewise linear

Beyond parameter fitting: model testing

Akaike Information criterion (Akaike 1974; Liddle 04)

$$\mathrm{AIC} \equiv -2 \ln \mathcal{L}_{\max} + 2k \,,$$

k=Number of parameters

Bayesian Information criterion (Schwarz 78, Liddle 04)

$$\mathrm{BIC} \equiv -2 \ln \mathcal{L}_{\max} + k \ln N$$

N=Number of data points

Bayesian Evidence

$$E = \int \mathcal{L}(\theta) \Pr(\theta) d\theta \,.$$

it does not focus on the best-fitting parameters of the model, but rather asks “of all the parameter values you thought were viable before the data came along, how well on average did they fit the data?”

Computationally expensive! (there are packages to help out there e.g. cosmonest)

$$\mathrm{P}(H \mid D) = \frac{\mathrm{P}(D \mid H)\mathrm{P}(H)}{\mathrm{P}(D)}$$

Bayes

$$\mathrm{P}(\boldsymbol{\theta} \mid D, H) = \frac{\mathrm{P}(D \mid \boldsymbol{\theta}, H)\mathrm{P}(\boldsymbol{\theta} \mid H)}{\mathrm{P}(D \mid H)}$$

Bayes, for parameter fitting

$$\mathcal{E} \equiv \mathrm{P}(D \mid H) = \int \mathrm{d}^n \boldsymbol{\theta} \mathrm{P}(D \mid \boldsymbol{\theta}, H)\mathrm{P}(\boldsymbol{\theta} \mid H)$$

Bayes for the MODEL itself

Introduction to Fisher

Cosmological examples of
hypothesis testing:

Are CMB data consistent with the hypothesis of Gaussian initial fluctuations?
Are CMB+LSS data consistent with the hypothesis of a spatially flat Universe?

Parameter estimation:

What is the value of the matter density parameter (in the LCDM model)?
And what it the value of the Hubble parameter today?

Model selection:

Is there evidence for a non-flat Universe?
Is there evidence for a non-constant dark energy?

Back to likelihoods

X data vector (random variable)

e.g., T value in pixels of CMB map, Fourier coefficients of density of survey etc.

$\boldsymbol{\Theta} = (\theta_1, \theta_2, ..., \theta_m).$ Vector of model parameters

$L(\boldsymbol{x}; \boldsymbol{\Theta}),$ Probability distribution of x

Since x is a random variable also Θ will be so … ideally:

$\langle \boldsymbol{\Theta} \rangle = \boldsymbol{\Theta}_0,$ UNBIASED

$\Delta\theta_i \equiv \left(\langle \theta_i^2 \rangle - \langle \theta_i \rangle^2 \right)^{1/2}$ Minimize this i.e. errors

i.e. we want the best unbiased estimator

Fisher information matrix

$$\boldsymbol{F}_{ij} \equiv \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle \qquad \mathcal{L} \equiv -\ln L$$

Fisher 1935

The maximum likelihood estimator is Θ_{ML} that maximizes $L(x;\Theta)$

a number of powerful theorems apply (e.g., Kendall & Stuart 1969):

For any unbiased estimator:

$$\Delta\theta_i \geq (\boldsymbol{F}^{-1})_{ii}^{1/2} \quad \text{or} \quad \Delta\theta_i \geq 1/\sqrt{\boldsymbol{F}_{ii}}.$$

Cramer-Rao inequality

If there is the best unbiased estimator it is the ML or function of thereof

The ML estimator is asymptotically the best unbiased estimator

Fisher matrix approach

(Fisher 1935)

How well can a future experiment do?
(quick and easy but not always accurate)

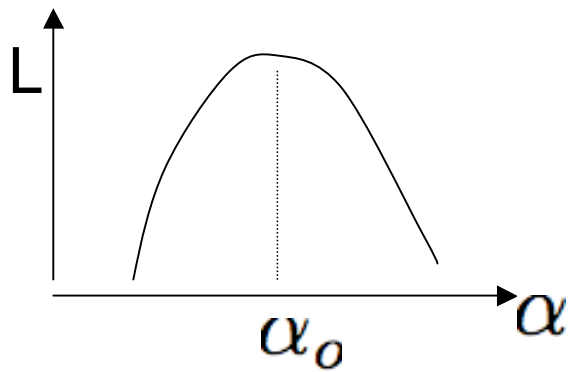
Fisher information matrix $$F_{ij} = \left\langle \frac{\partial^2 L}{\partial \alpha_i \partial \alpha_j} \right\rangle \qquad L = -\ln \mathcal{L}$$

e.g.
$$1/2(data - theory(\vec{\alpha}))C^{-1}(data - theory(\vec{\alpha})) \simeq$$
$$1/2(fiducial - theory(\vec{\alpha}))C^{-1}(fiducial - theory(\vec{\alpha}))$$

To develop intuition, one parameter case, Gaussian likelihood.

$$L = 1/2(\alpha - \alpha_0)/\sigma_\alpha^2$$

Second deriv. w.r.t. $\alpha \longrightarrow 1/\sigma_\alpha^2$



In general

Expand in Taylor series around α_0

$$\Delta L = \frac{1}{2}\frac{d^2 L}{d\alpha^2}(\alpha - \alpha_0)^2$$

First deriv 0 by construction

Should remind you of χ^2

All it is: Quadratic expansion around the max

$1/\sqrt{d^2L/d\alpha^2}$ is the 1 sigma displacement of α from α_0

Like a measure for the width of the peak…..

Multi dimensional case…

$$\sigma^2_{\alpha_i,\alpha_j} \geq (\mathbf{F}^{-1})_{ij}$$

Parameters covariance

$$\sigma_{\alpha_i} \geq \sqrt{\frac{1}{F_{ii}}}$$

If all other parameters are fixed

$$\sigma_\alpha = (\mathbf{F}^{-1})^{1/2}_{ii}$$

Marginalized errors

Matrix inversion performed

Conditional and marginal errors

Minumum error on α_i if all other parameters are known

$$\sigma_{\alpha_i} \geq \sqrt{\frac{1}{F_{ii}}}$$

ALMOST NEVER USED

The marginal distribution of α_i :integrate over other parameters

$$p(\alpha_1) = \int d\alpha_2..d\alpha_N p(\alpha)$$

$$\sigma_\alpha \geq (\mathbf{F}^{-1})_{ii}^{1/2}$$

What are we really saying?

$$\langle p(\mathbf{x}|\boldsymbol{\theta}, M)\rangle = L_0 \exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)_\alpha \mathsf{F}_{\alpha\beta}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)_\beta\right]$$

$$\propto \frac{1}{\sqrt{detC}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t\right]$$

This is sometimes called: Laplace approximation

Explicit calculation

$$2\mathcal{L} = \ln\det \boldsymbol{C} + (\boldsymbol{x}-\boldsymbol{\mu})\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^t$$

Dropped irrelevant constant
Assumed Gaussianity

$$\boldsymbol{C} = \langle(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^t\rangle$$

DATA covariance (can depend on the parameters)

You can show that:

$$\boxed{\mathsf{F}_{\alpha\beta} = \langle \mathcal{L}_{,\alpha\beta}\rangle = \frac{1}{2}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}\mathsf{C}_{,\beta} + \mathsf{C}^{-1}\mathsf{M}_{\alpha\beta}],}$$

This simplifies
in specific cases

where $\mathsf{M}_{\alpha\beta} \equiv \langle \mathsf{D}_{,\alpha\beta}\rangle = \boldsymbol{\mu}_{,\alpha}\,\boldsymbol{\mu}_{,\beta}^T + \boldsymbol{\mu}_{,\beta}\,\boldsymbol{\mu}_{,\alpha}^T.$ and $\mathsf{C}_{,\alpha} \equiv \frac{\partial}{\partial\theta_\alpha}\mathsf{C}.$

REQUIRES NO DATA!

Other option:

Compute explicitly:

$$-\frac{\partial^2 \ln L}{\partial\theta_\alpha \partial\theta_\beta}$$

Taking the data to be = a fiducial model

Numerical second derivative: beware!

ALWAYS TEST STABILITY OF DERIVATIVES!!!!

You can compute the Fisher matrix
BEFORE you do the experiment.

You can then use it as a tool
to design or optimize experiments

Within the assumptions made, now you know everything!

Say you have worked with 5 parameters but now you want to keep parameters 1 and 3 fixed at fiducial model….

Take the submatrix (1,3) of F_{ij}

Say you have a 5 parameters Fisher and you want to plot the joint 2D forecasted constraint for parameters 2 and 4 marginalized over the other parameters

Invert Fij, take the submatrix (2,4) invert this back. Call this Q. Q describes a Gaussian 2D likelihood i.e.

$$\tilde{\chi^2} = \sum_{kq} (\alpha_k - \alpha_k^{fid.}) Q_{jq} (\alpha_q - \alpha_q^{fid})$$

$$\Delta = \delta\vec{\alpha} Q \delta\vec{\alpha}$$

And look up Δ in the “famous” table

You can also draw the ellipses!

What if you want to reparameterize?

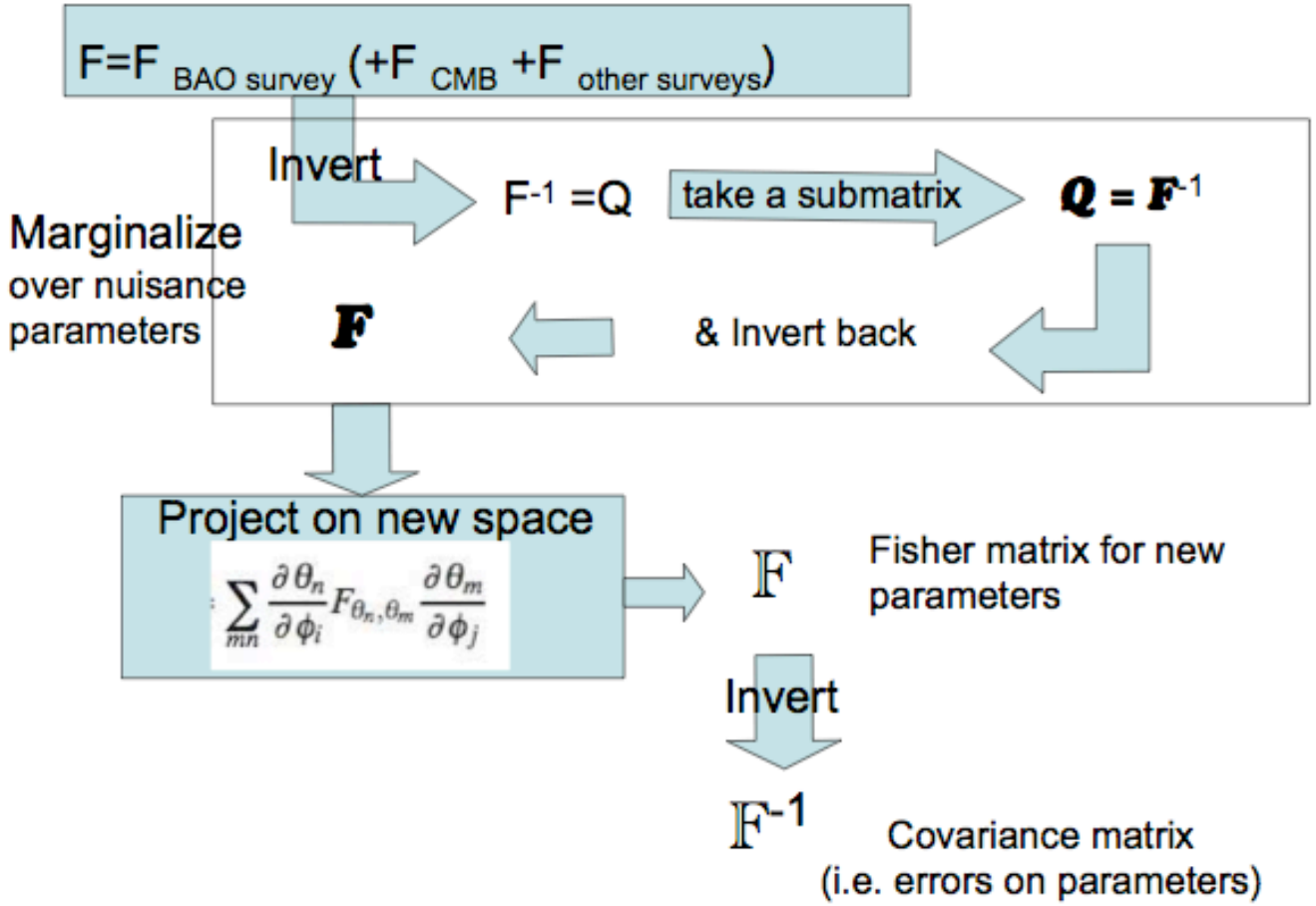
Typical example
CMB: parameters

Now you want to combine with BAO constraints

BAO parameters H(z), Da(z)

$$F_{\phi_i,\phi_j} = \sum_{mn} \frac{\partial \theta_n}{\partial \phi_i} F_{\theta_n,\theta_m} \frac{\partial \theta_m}{\partial \phi_j}$$

This is what you do….



Practical tools: icosmo

http://www.icosmo.org/Initiative_Web/Initiative.html

Notes from a tutorial course on icosmo link from:
http://icc.ub.edu/~liciaverde/ERCtraining.html

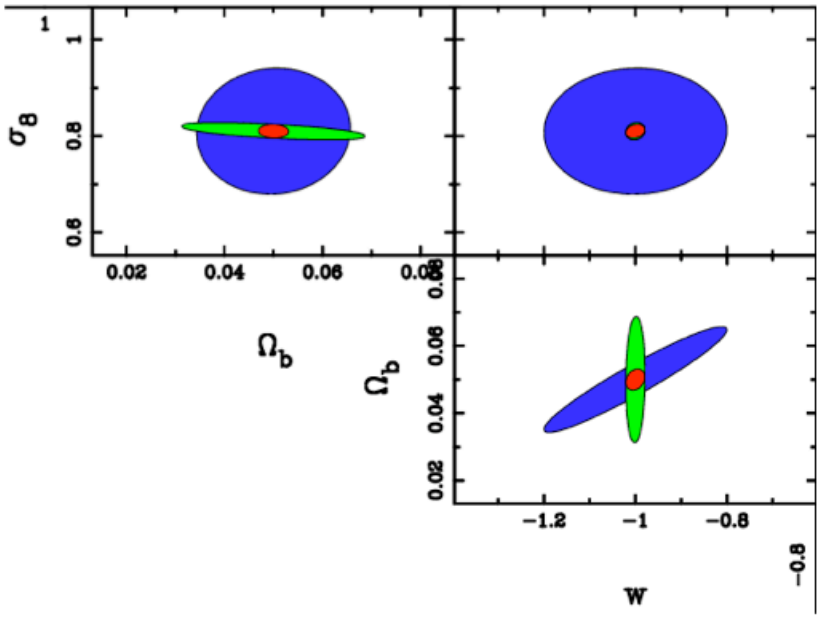


Fig. 1. The expected error ellipses for cosmological parameters (σ_8, baryon density parameter Ω_b, and dark energy equation of state $w \equiv p/\rho c^2$) from a 3D weak lensing survey of 1000 square degrees, with a median redshift of 1 and a photometric redshift error of 0.15. Probabilities are marginalised over all other parameters, except that $n = 1$ and a flat Universe are assumed. Dark ellipses represent a prior from WMAP, pale represents the 3D lensing survey alone, and the central ellipses show the combination (from Kitching T., priv. comm.).

aside

Popular CMB Fisher matrix

$$F_{ij} = \sum_{\ell} \frac{(2\ell+1)}{2} \frac{\frac{\partial C_\ell}{\partial \alpha_j}\frac{\partial C_\ell}{\partial \alpha_j}}{(C_\ell + \mathcal{N} e^{\sigma^2 \ell^2})^2}$$

Approximations?

Covariance?

Applicability: Fisher vs non Fisher

Fisher and systematic errors

Can the Fisher approach account for systematic errors?

In general NO

But there’s an exception

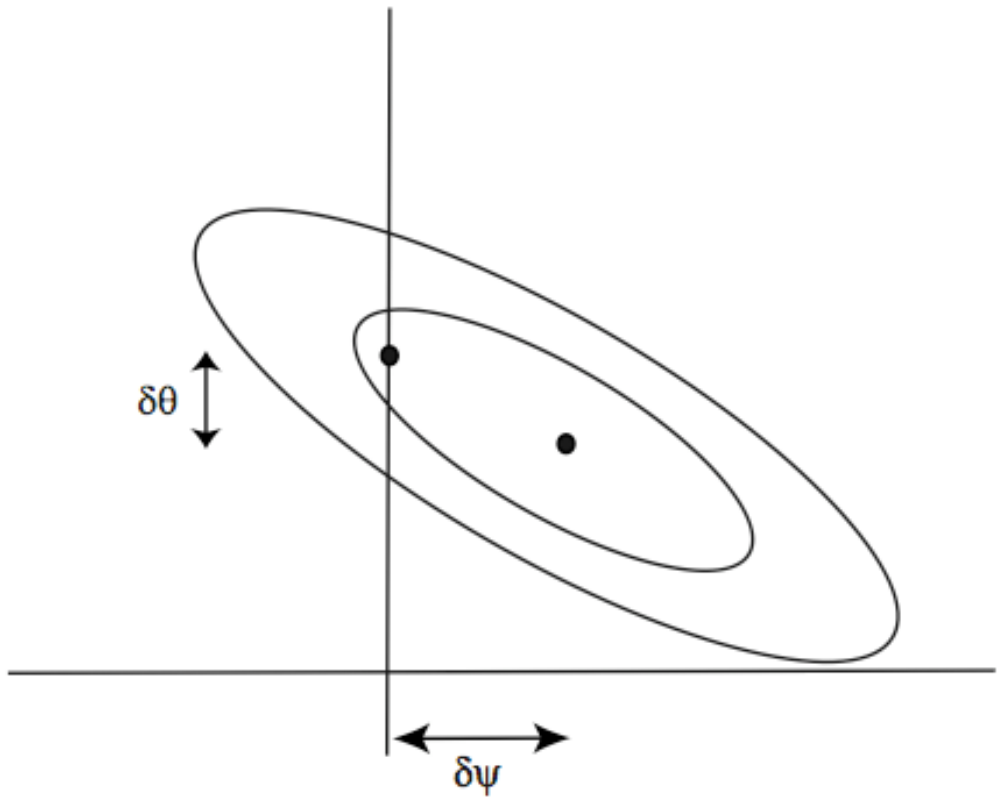
Imagine you have two competing models M and M': M with n parameters and M' with n', where n'<n. Say also that the two models are NESTED, i.e. M' is a particular case of M

If the true underlying model is M and you instead fit the data with M', the maximum expected likelihood will not be at the correct values of the parameters: if n-n'=p, the n' parametrs shift from their value to compensate for the fact that p parameters are kept fixed at "wrong" values.
If the p parameters differ by $\delta\psi_\alpha$ fro their true values, the other parameters are shifted by:

$$\delta\boldsymbol{\theta}'_\alpha = -(\mathsf{F}'^{-1})_{\alpha\beta}\mathsf{G}_{\beta\zeta}\delta\psi_\zeta \qquad \alpha,\beta = 1\ldots n', \zeta = 1\ldots p$$

$$\mathsf{G}_{\beta\zeta} = \frac{1}{2}\mathrm{Tr}\left[\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}\mathsf{C}_{,\zeta} + \mathsf{C}^{-1}(\mu_{,\zeta}\mu^T_{,\beta} + \mu_{,\beta}\mu^T_{,\zeta})\right],$$

This is what's going on:



This is very useful: e.g., isocurvatures, delayed recombination, neutrinos…

And there's more

Remember the evidence?

$$p(\mathbf{x}|M) = \int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}M)p(\boldsymbol{\theta}|M)$$

Back to models M and M':

$$\frac{p(M'|\mathbf{x})}{p(M|\mathbf{x})} = \frac{p(M')}{p(M)} \boxed{\frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}'M')p(\boldsymbol{\theta}'|M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}M)p(\boldsymbol{\theta}|M)}}.$$

Bayes factor B

For non-commital priors

$$p(M') = p(M).$$

What's the Bayes factor?

$$B \equiv \frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}' M')p(\boldsymbol{\theta}'|M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta} M)p(\boldsymbol{\theta}|M)}$$

M will have higher likelihood (or as high) but the evidence will favour the simpler model if the fit is nearly as good, through the smaller prior volume.

For uniform separable priors:

$$B = \frac{\int d\boldsymbol{\theta}'\, p(\mathbf{x}|\boldsymbol{\theta}', M')}{\int d\boldsymbol{\theta}\, p(\mathbf{x}|\boldsymbol{\theta}, M)} \frac{\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n}{\Delta\boldsymbol{\theta}'_1 \ldots \Delta\boldsymbol{\theta}'_{n'}}.$$

If prior is wide enough to encompass the"support" of the likelihood

$$\frac{\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n}{\Delta\boldsymbol{\theta}'_1 \ldots \Delta\boldsymbol{\theta}'_{n'}} = \Delta\boldsymbol{\theta}_{n'+1} \ldots \Delta\boldsymbol{\theta}_{n'+p}$$

Requires a painful multi-dimensional integration, but….

Laplace approximation and Fisher to the rescue!

$$\langle B \rangle = (2\pi)^{-p/2} \frac{\sqrt{\det \mathsf{F}}}{\sqrt{\det \mathsf{F}'}} \frac{L_0'}{L_0} \Delta\boldsymbol{\theta}_{n'+1} \ldots \Delta\boldsymbol{\theta}_{n'+p}.$$

Where you already know how to compute L'o and Lo

And you'll see that this simplifies to:

$$\langle B \rangle = (2\pi)^{-p/2} \frac{\sqrt{\det \mathsf{F}}}{\sqrt{\det \mathsf{F}'}} \exp\left(-\frac{1}{2}\delta\boldsymbol{\theta}_\alpha \mathsf{F}_{\alpha\beta}\delta\boldsymbol{\theta}_\beta\right) \prod_{q=1}^{p} \Delta\boldsymbol{\theta}_{n'+q}.$$

$$\delta\boldsymbol{\theta}_\alpha = \delta\boldsymbol{\theta}'_\alpha \text{ for } \alpha \leq n' \qquad \delta\boldsymbol{\theta}_\alpha = \delta\psi_{\alpha-n'} \quad \alpha > n'.$$

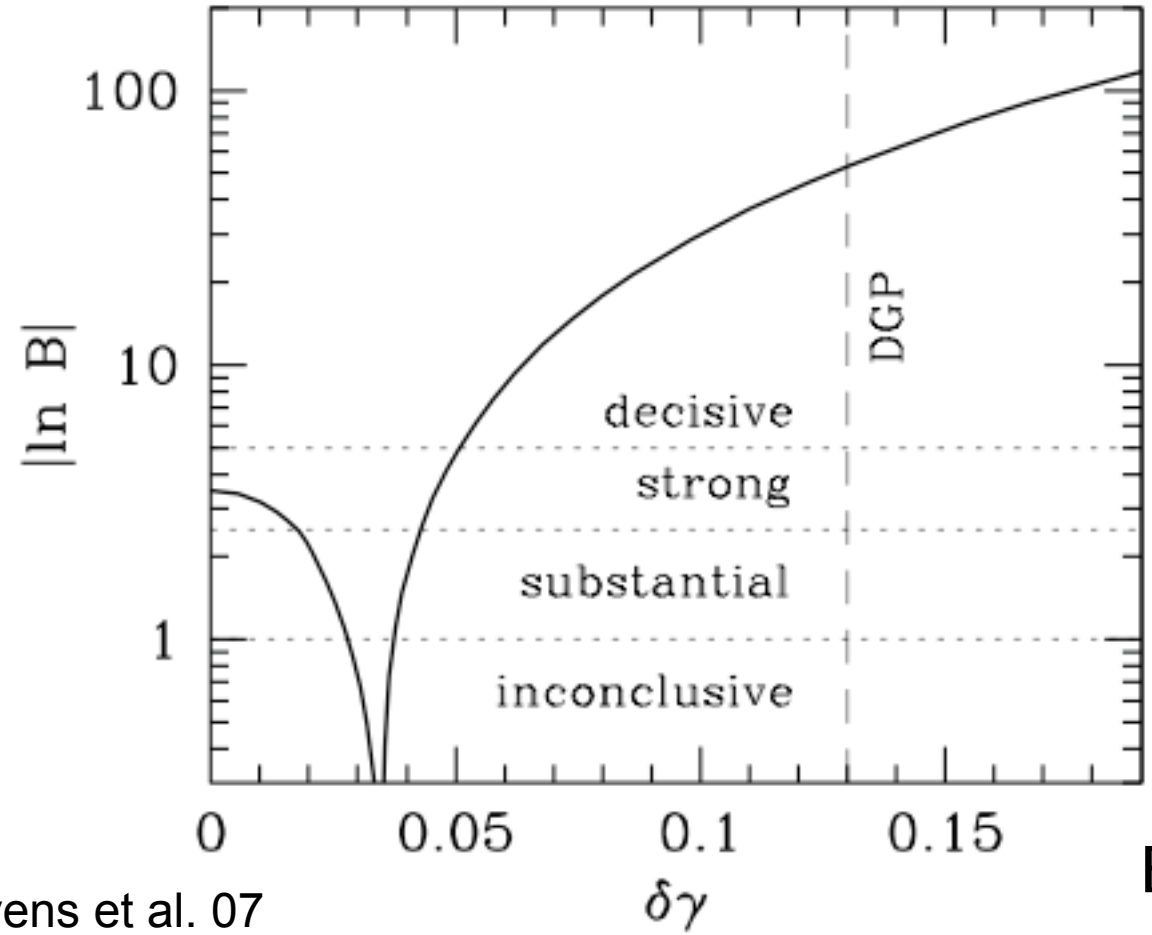
Example:

Is gravity described by General relativity?

$d \ln \delta / d \ln \Omega_m = \gamma$, where $\gamma = 0.55$ for GR

e.g., DGP:

$\gamma = 0.68$



Heavens et al. 07

Euclid+Planck

…..Back to likelihoods

CMB ΔT distribution is close to gaussian,

So the Cl's are NOT (and at low l CLT does not hold)

$$\mathcal{L}(T|C_\ell^{th}) \propto \frac{\exp[-(TS^{-1}T)/2]}{\sqrt{det(S)}}$$

$$S_{ij} = \sum_\ell \frac{(2\ell+1)}{4\pi} C_\ell^{th} P_\ell(\hat{n}_i \cdot \hat{n}_j)$$

Signal covariance

Given by a theoretical model

Legendre polynomials

Now in spherical harmonics

$$\mathcal{L}(T|C_\ell^{th}) \propto \frac{\exp[-1/2|a_{\ell m}|^2/C_\ell^{th}]}{\sqrt{C_\ell^{th}}}$$

Isotropy means that we can sum over $m's$

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\ln\left(\frac{C_\ell^{th}}{C_\ell^{data}}\right) + \left(\frac{C_\ell^{data}}{C_\ell^{th}}\right) - 1\right]$$

$$C_\ell^{data} = \sum_m |a_{\ell m}|^2/(2\ell+1).$$

Exercise:
show that

With noise	$C_\ell^{th} \longrightarrow C_\ell^{th} + \mathcal{N}_\ell$
Partial sky	$\ln\mathcal{L} \xrightarrow{\sim} f_{sky}\ln\mathcal{L}$

CMB light is polarized!

You can easily generalize the above to

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left\{\ln\left(\frac{C_\ell^{BB}}{\hat{C}_\ell^{BB}}\right) + \ln\left(\frac{C_\ell^{TT}C_\ell^{EE}-(C_\ell^{TE})^2}{\hat{C}_\ell^{TT}\hat{C}_\ell^{EE}-(\hat{C}_\ell^{TE})^2}\right)\right.$$
$$\left. + \frac{\hat{C}_\ell^{TT}C_\ell^{EE}+C_\ell^{TT}\hat{C}_\ell^{EE}-2\hat{C}_\ell^{TE}C_\ell^{TE}}{C_\ell^{TT}C_\ell^{EE}-(C_\ell^{TE})^2} + \frac{\hat{C}_\ell^{BB}}{C_\ell^{BB}} - 3\right\},$$

C_ℓ denotes C_ℓ^{th} and $\hat{C}_\ell$ denotes C_ℓ^{data}.

Exact TE,EE,BB Likelihood

D. EXACT LIKELIHOOD EVALUATION AT LOW MULTIPOLES

At low multipoles, $l \leq 23$, we evaluate the likelihood of the data for a given theoretical model exactly from the temperature and polarization maps. The standard likelihood is given by

$$L(\vec{m}|S)d\vec{m} = \frac{\exp\left[-\frac{1}{2}\vec{m}^t(S+N)^{-1}\vec{m}\right]}{|S+N|^{1/2}} \frac{d\vec{m}}{(2\pi)^{3n_p/2}}, \tag{D1}$$

where $\vec{m}$ is the data vector containing the temperature map, $\vec{T}$, as well as the polarization maps, $\vec{Q}$, and $\vec{U}$, n_p is the number of pixels of each map, and S and N are the signal and noise covariance matrix ($3n_p \times 3n_p$), respectively. As the temperature data are completely dominated by the signal at such low multipoles, noise in temperature may be ignored. This simplifies the form of likelihood as

$$L(\vec{m}|S)d\vec{m} = \frac{\exp\left[-\frac{1}{2}\vec{\tilde{m}}^t(\tilde{S}_P+N_P)^{-1}\vec{\tilde{m}}\right]}{|\tilde{S}_P+N_P|^{1/2}} \frac{d\vec{\tilde{m}}}{(2\pi)^{n_p}} \frac{\exp\left(-\frac{1}{2}\vec{T}^t S_T^{-1}\vec{T}\right)}{|S_T|^{1/2}} \frac{d\vec{T}}{(2\pi)^{n_p/2}}, \tag{D2}$$

where S_T is the temperature signal matrix ($n_p \times n_p$), the new polarization data vector, $\vec{\tilde{m}} = (\tilde{Q}_p,\ \tilde{U}_p)$, is given by

$$\tilde{Q}_p \equiv Q_p - \frac{1}{2}\sum_{l=2}^{23}\frac{S_l^{TE}}{S_l^{TT}}\sum_{m=-l}^{l} T_{lm}({}_{+2}Y_{lm,p} + {}_{-2}Y^*_{lm,p}), \tag{D3}$$

$$\tilde{U}_p \equiv U_p - \frac{i}{2}\sum_{l=2}^{23}\frac{S_l^{TE}}{S_l^{TT}}\sum_{m=-l}^{l} T_{lm}({}_{+2}Y_{lm,p} - {}_{-2}Y^*_{lm,p}), \tag{D4}$$

and $\tilde{S}_P$ is the signal matrix for the new polarization vector with the size of $2n_p \times 2n_p$. As T_{lm} is totally signal dominated, the noise matrix for $(\vec{\tilde{Q}},\ \vec{\tilde{U}})$ equals that for $(\vec{Q},\ \vec{U})$, n_p. To estimate T_{lm}, we used the full-sky internal linear combination (ILC) temperature map (Hinshaw et al. 2006).

One can show that equation (D1) and (D2) are mathematically equivalent when the temperature noise is ignored. The new form, equation (D2), allows us to factorize the likelihood of temperature and polarization, with the information in their cross-correlation, S_l^{TE}, fully retained. We further rewrite the polarization part of the likelihood as

Approximations

$$-2\ln\mathcal{L} = \sum_\ell (2\ell+1)\left[\ln\left(\frac{\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell}{\tilde{\mathcal{C}}_\ell}\right) + \frac{\tilde{\mathcal{C}}_\ell}{\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell} - 1\right]$$

$$\ln\mathcal{L}_{\rm Gauss} \propto -\frac{1}{2}\sum_{\ell\ell'}(\mathcal{C}_\ell^{\rm th}-\widehat{\mathcal{C}}_\ell)Q_{\ell\ell'}(\mathcal{C}_{\ell'}^{\rm th}-\widehat{\mathcal{C}}_{\ell'})$$

$$-2\ln\mathcal{L}_{\rm LN} = \sum_{\ell\ell'}(z_\ell^{\rm th}-\widehat{z}_\ell)\mathcal{Q}_{\ell\ell'}(z_{\ell'}^{\rm th}-\widehat{z}_{\ell'}) \qquad \left\{ \begin{array}{l} z_\ell^{\rm th} = \ln(\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell),\ \widehat{z}_\ell = \ln(\widehat{\mathcal{C}}_\ell+\mathcal{N}_\ell) \\ \mathcal{Q}_{\ell\ell'} = (\widehat{\mathcal{C}}_\ell+\mathcal{N}_\ell)Q_{\ell\ell'}(\widehat{\mathcal{C}}_{\ell'}+\mathcal{N}_{\ell'}). \end{array} \right.$$

Expand around the max $\quad ;\ \widehat{\mathcal{C}}_\ell = \mathcal{C}_\ell^{\rm th}(1+\epsilon).$

$$-2\ln\mathcal{L}_\ell = (2\ell+1)[\epsilon - \ln(1+\epsilon)] \simeq (2\ell+1)\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} + \mathcal{O}(\epsilon^4)\right)$$

Oh, look! $\quad \ln\mathcal{L} = \frac{1}{3}\ln\mathcal{L}_{\rm Gauss} + \frac{2}{3}\ln\mathcal{L}'_{\rm LN} \qquad \mathcal{Q}_{\ell\ell'} = (\mathcal{C}_\ell^{\rm th}+\mathcal{N}_\ell)Q_{\ell\ell'}(\mathcal{C}_{\ell'}^{\rm th}+\mathcal{N}_{\ell'}).$

For Fisher

$$F_{ij}^{CMB} = \sum_{XY}\sum_{\ell} \frac{\partial C_\ell^X}{\partial \theta_i} \left(\mathscr{C}_\ell^{XY}\right)^{-1} \frac{\partial C_\ell^Y}{\partial \theta_j}$$

X,Y=TT, TE, EE, BB etc.,

$$\mathscr{C}_\ell = \frac{2}{2\ell+1} \begin{pmatrix} (C_\ell^{TT})^2 & (C_\ell^{TE})^2 & C_\ell^{TT} C_\ell^{TE} & 0 \\ (C_\ell^{TE})^2 & (C_\ell^{EE})^2 & C_\ell^{EE} C_\ell^{TE} & 0 \\ C_\ell^{TT} C_\ell^{TE} & C_\ell^{EE} C_\ell^{TE} & 1/2[(C_\ell^{TE})^2 + C_\ell^{TT} C_\ell^{EE}] & 0 \\ 0 & 0 & 0 & (C_\ell^{BB})^2 \end{pmatrix}$$

Proof by intimidation: “just do it”

“If tortured sufficiently, data will confess to almost anything”

Fred Menger

Treat your data with respect

(Licia Verde)

Thank you!

SOG, LOC

references

http://arxiv.org/pdf/0712.3028

http://arxiv.org/pdf/0911.3105

http://arxiv.org/pdf/0906.0664

http://arxiv.org/pdf/astro-ph/9603021 (only sec 1 & 2)

http://arxiv.org/pdf/astro-ph/0703191

Real world issues

<u>Discrete Fourier transform</u> (routines available)

Box size L

even if $\delta(r)$ in the box is continuous, δ_k will be discrete.

$$\vec{k} = \left(\frac{2\pi}{L}\right)(i, j, k) \quad \text{where} \quad \Delta_k = \frac{2\pi}{L}$$

But you also pixelize $\delta(x)$

Cloud in cloud, Nearest grid point, triangular shaped cloud…

Remember to deconvolve at the end!!!!