MCMC and the Fisher Matrix

NASSP Observational Cosmology Course
Bruce Bassett

Overview

- **The Monte Carlo Markov Chain (MCMC) method for parameter estimation**
- Designed to answer the question:

  "How do I estimate 1000 parameters from my dataset?"
- **The Fisher Matrix**
- Designed to answer the question:

  "How do I estimate the power of a survey or experiment that hasn't taken any data yet?"

Monte Carlo Markov Chain (MCMC)

- If we are searching through a high-dimensional parameter space we cant use the grid approach (why?)
- Need a clever way to find the best-fits and to estimate error-bars
- MCMC is such a method and is based on jumping around randomly

Basic idea

Write the parameters we want to estimate as a vector

$$\vec{\theta}$$

We want to find the parameters which minimise

$$\chi^2(\vec{\theta})$$

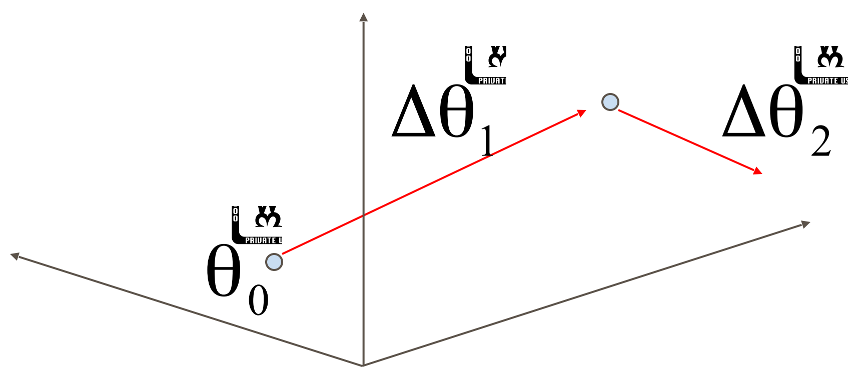
Step 1: choose an initial vector randomly

$$\vec{\theta}_0$$

Subscript is step number not Which parameter!

The MCMC Algorithm

- Then loop doing the following N times:
- Choose a jump $\Delta\theta_i$ from a Gaussian distribution with mean zero and standard deviation vector σ (controls how big the characteristic jumps are in each parameter)



The Metropolis-Hastings Algorithm in MCMC

- We construct $u_{i+1} = \theta_i + \Delta\theta_i$
- And compute

$$\boxed{R = \exp((\chi^2(\theta_i) - \chi^2(u_{i+1}))/2)}$$

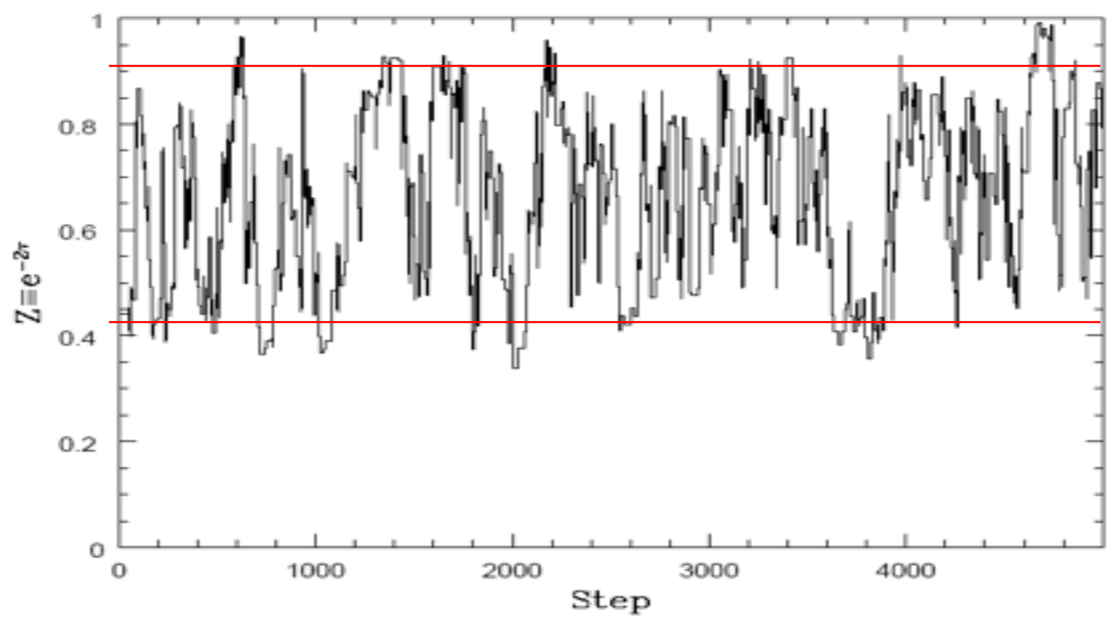
- If we have got to a better place, R > 1.
- The metropolis-Hastings algorithm then accepts this jump with probability:

$$\boxed{P(accept) = \min(1, R)}$$

The Metropolis-Hastings Algorithm

- If we accept the step, set $\theta_{i+1} = u_{i+1}$ (move to the new point)
- Otherwise set $\theta_{i+1} = \theta_i$ (don't move!)
- Iterate this process (the result is called a **chain**)

- E.g.: a chain for reionisation, τ



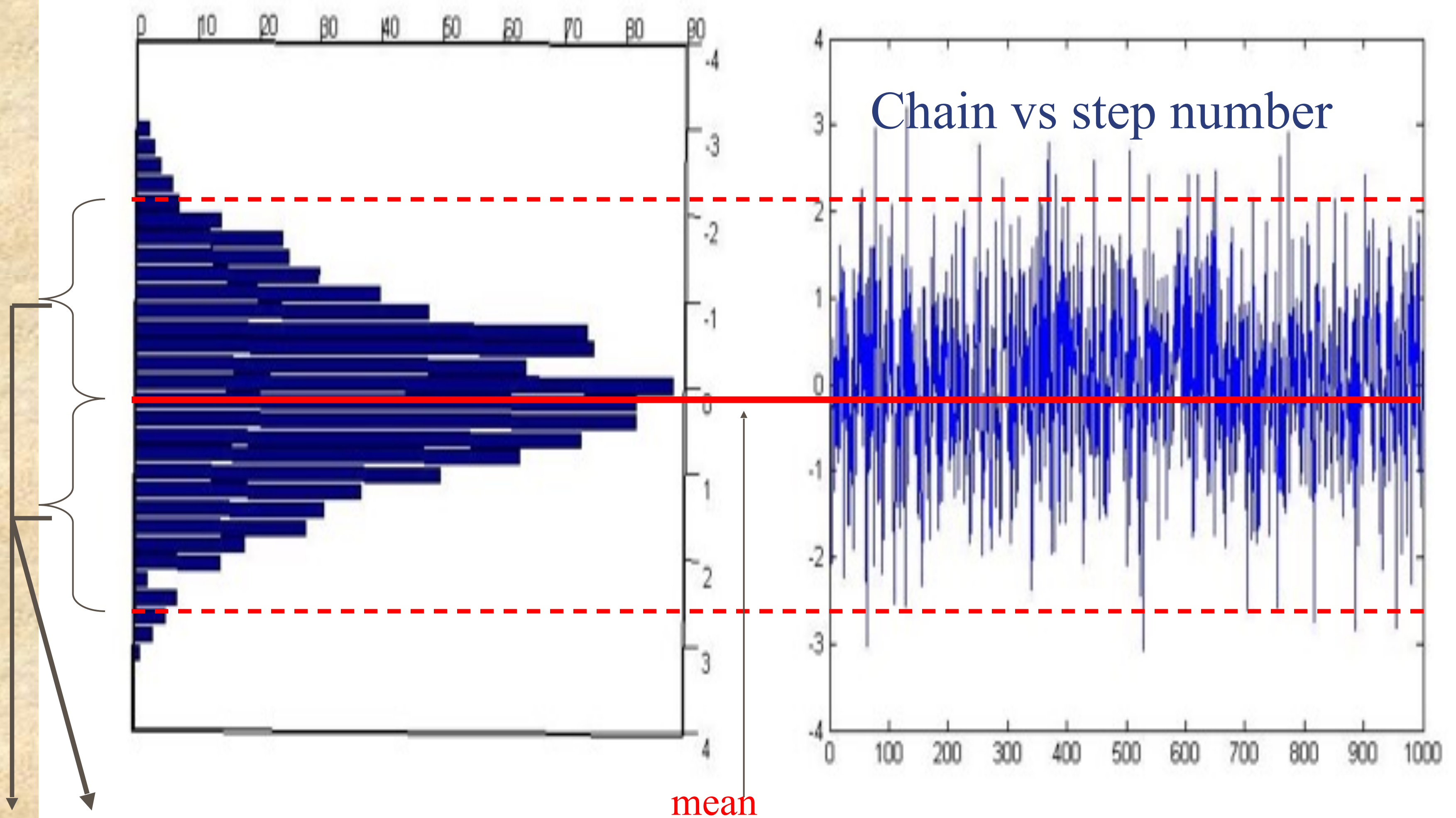
Sometimes going the wrong way is good…

- While MCMC always accepts a good step it sometimes accepts a bad step (since the probability for acceptance when $R < 1$ is not zero). This *stops* it from being trapped in local minima.

Why is MCMC useful?

1. It is a theorem that eventually MCMC will find the global minimum (usually pretty quick)
2. The computational power needed to do a good job does **not** grow exponentially with the number of parameters (unlike other methods) – computationally efficient
3. We are typically interested in computing not only the best-fit parameters, but also the error bars on those parameters. This means integrating the likelihood over all parameters other than the one we are interested in (a process called marginalisation). **With MCMC this is trivial – all we do is histogram the component of the chain corresponding to that parameter.**

Histograming to find the best-fit and errors



(upper/lower) error bars (1, 2 sigma bars containing 68%, 95% of chain)

Burn-in

- The first part of the chain is usually not relevant and is removed (burn-in)
- There is no precise burn-in criterion. One way is to remove the initial part of the chain before the first time the chain reached 50% of the best likelihood achieved in the full chain

Convergence

- MCMC only assures you of the right answers as N → infinity
- How big is infinity in practise?
- One way to check is to either run a **single** long chain, divide it into sub-chains and check that they give the same answers…
- Or, run **multiple** chains starting at **random** points in the parameter space and check that they give the same answers…

Exercise: Code your own MCMC

- Write pseudo-code for your own MCMC algorithm

Fisher Matrix

The Fisher Matrix

- How do we estimate how well a survey will do before it gets any data? E.g. how well will a future SNIa survey constrain w?
- One way is to do a simulation, produce fake data and then run it through MCMC like you would the real data. This is the best, but is time-consuming.
- The Fisher matrix provides a quick and easy answer which is often good enough.

The Fisher Matrix

- Assumes the Likelihood *(L*~ exp(-χ^2/2)) is close to a Gaussian function of the parameters…
- The Fisher matrix is then:

$$F_{AB} = \sum_i \left(\frac{\partial X}{\partial \theta_A} \frac{\partial X}{\partial \theta_B} \right)_i \epsilon_i^{-2}$$

Where the quantity being measured is X (e.g. d_L for a SNIa survey) and ε_i is the *i*-th projected error on X and the sum is over all the measurements which will be made (e.g sum over *N* SNIa).

But what does it do?

Well, if we know all the other parameters then the error on the μ–th parameter is:

$$\boxed{\Delta\theta_{\mu} \geq (F_{\mu\mu})^{-1/2}}$$

If we must estimate **al** usly then the error on the μ-th parameter is:

$$\boxed{\Delta\theta_{\mu} \geq (F^{-1/2})_{\mu\mu}}$$

(i.e. compute the inverse square-root and take the μμ element)

Convince yourself using a 2 always greater than the first…why does this make sense fr oods?

A linear example

1. A linear parameter dependence (here z is the independent variable, e.g. redshift)

$$X(z,\theta_1) = z \cdot \theta + z^2$$

$$\Rightarrow F_{11} = \sum_i \varepsilon_i^{-2} (z_i)^2$$

$$\Rightarrow \Delta\theta_1 \approx \frac{1}{\sqrt{\sum_i \varepsilon_i^{-2} (z_i)^2}} \approx \frac{\varepsilon}{\sqrt{N}\sqrt{\sum_i z_i^2}}$$

If all the N measurements have the same error, ε

In general…

- In general the measurement are nonlinear functions of the parameters…e.g.
- If X = d_L and the parameters are w, H_0 and $\Omega_{m.}$ *Then:*

$$d_L \propto \frac{c}{H_0} \int \frac{dz'}{\sqrt{\Omega_m (1+z)^3 + \Omega_{DE} (1+z)^{3(1+w)}}}$$

- Compute the Fisher matrix in this case…

Why bother?

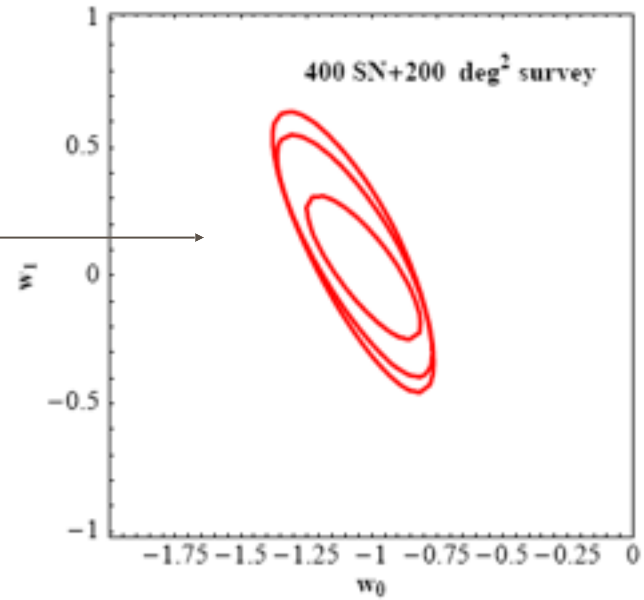
- Allows predicted performance estimates, e.g…

> We find that a 200 deg^2 spectroscopic survey reaching $z \approx 3$ can constrain w_0, w_1 to within $\Delta w_0 = 0.21, \Delta w_1 = 0.26$ and to $\Delta w_0 = 0.39, \Delta w_1 = 0.54$

- It also allows easy detection of degeneracies:

$$\boxed{\Delta\theta^T \cdot F \cdot \Delta\theta = \alpha}$$

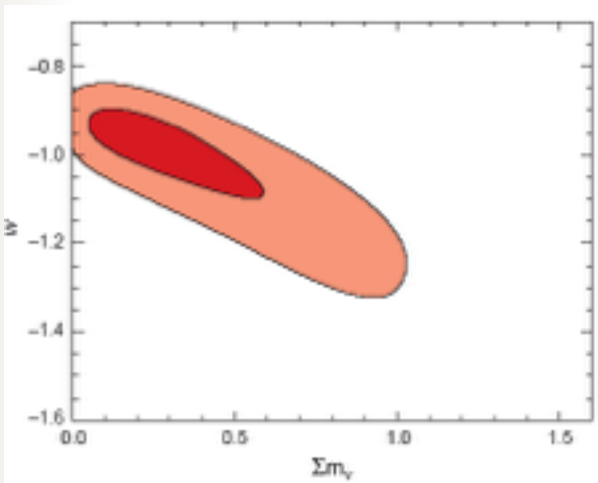
Here F is the Fisher matrix - this is the equation for an ellipse in matrix forn (the error-ellipse)



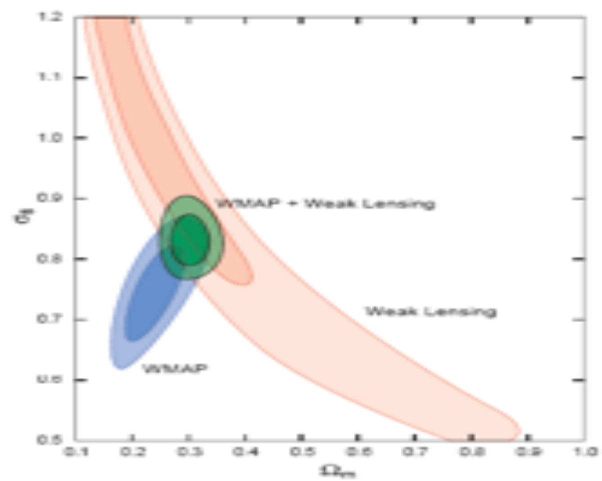
Here α **= 2.31, 6.2** for 2 parameters at 1 and 2 sigma contours respectively.

Limitations of the Fisher matrix

- It is only accurate if the Likelihood is nearly Gaussian. Put another way…
- Error-contours can be very non-Gaussian in general and not shaped like ellipses!

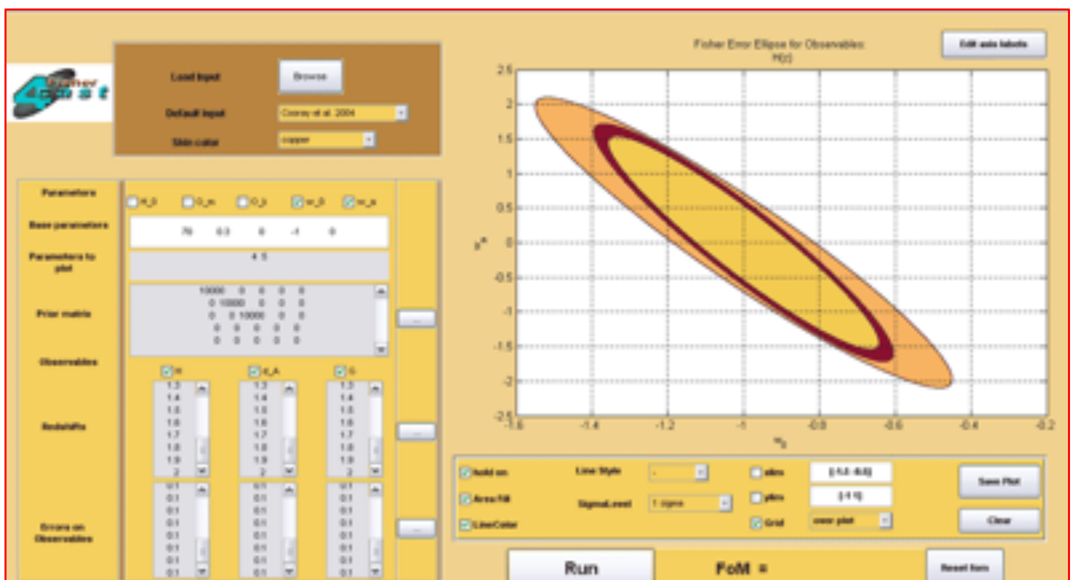


WMAP +
Other data



Fisher4Cast

- Available at http://www.cosmology.org.za
- General GUI-based Fisher matrix analysis
- Produced by UCT students
- Publicly released code under GNU license



Summary

- MCMC provides a toolbox to estimate best-fit parameters and their errors from any data-set even with large numbers of parameters (> 1000)
- The Fisher matrix provides an easy way to forecast the power of an experiment or survey to constrain a parameter of interest.
- Both of these are very powerful tools one can use anywhere (not just cosmology)