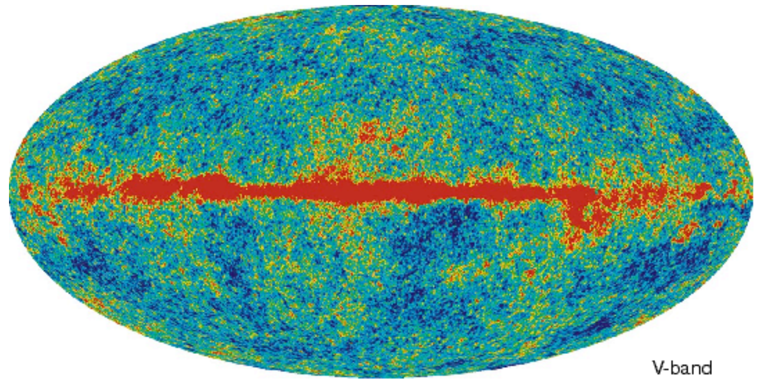
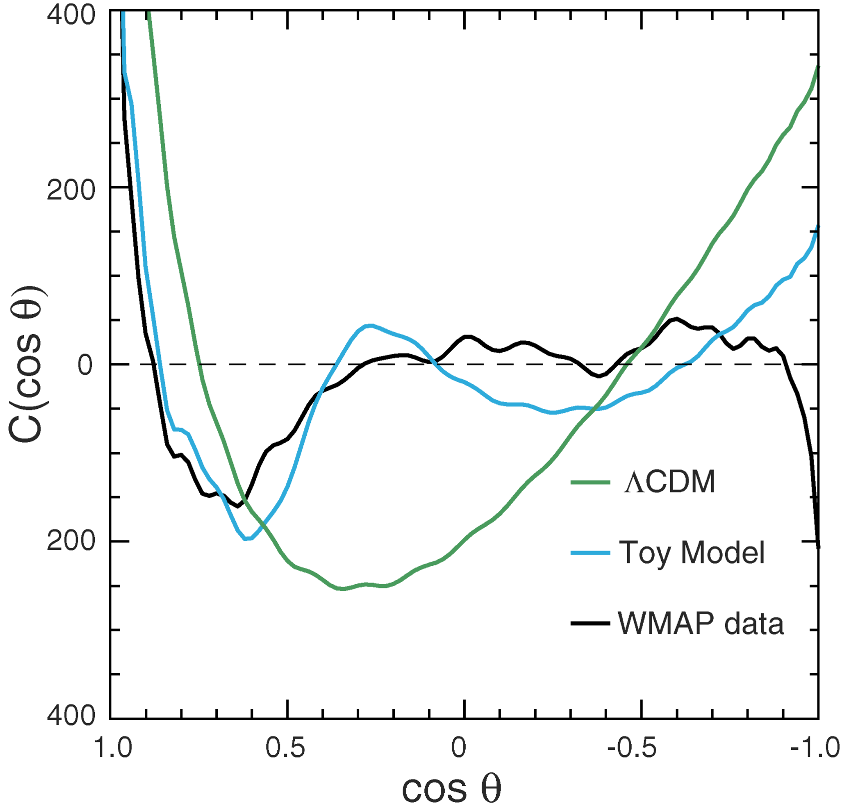
Outline

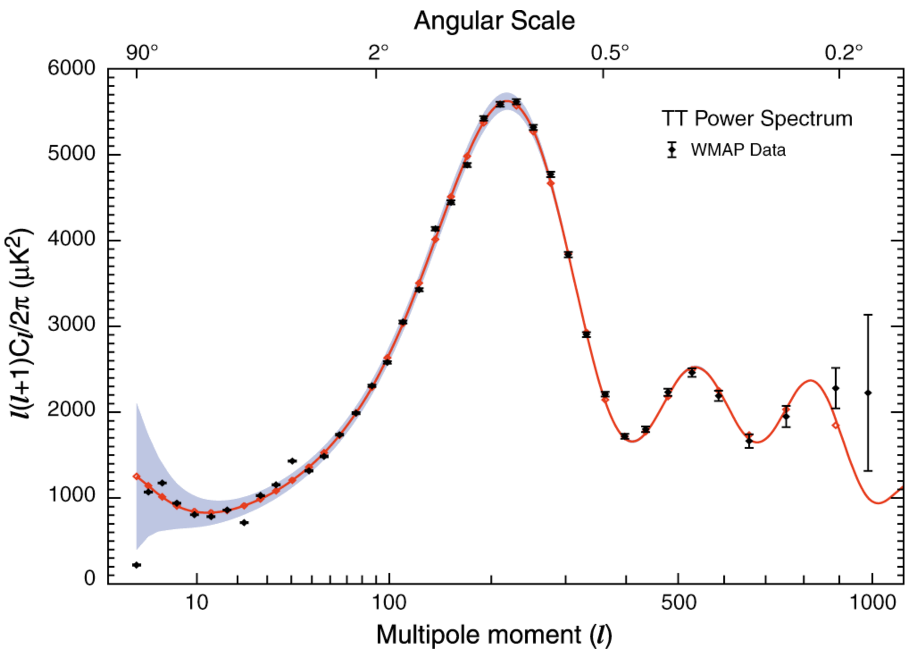
- Parameter Estimation
- Fisher Matrix analysis
- Numerical methods for Parameter Estimation
- Model Selection
- Data Compression

Alan Heavens, University of Edinburgh
afh@roe.ac.uk

LCDM fits the WMAP data well.







Inverse problems

Most data analysis problems are *inverse problems*. You have a set of data $\mathbf{x}$, and you wish to interpret the data in some way. Typical classifications are:

- Hypothesis testing
- Parameter estimation
- Model selection

Cosmological examples of the first type include

- Are CMB data consistent with the hypothesis that the initial fluctuations were gaussian, as predicted by inflation theory? (see C. Porciani lectures)
- Are large-scale structure observations consistent with the hypothesis that the Universe is spatially flat?

Cosmological examples of the second type include

- In the Big Bang model, what is the value of the matter density parameter?
- What is the value of the Hubble constant?

Model selection can include slightly different types (but are mostly concerned with larger, often more qualitative questions):

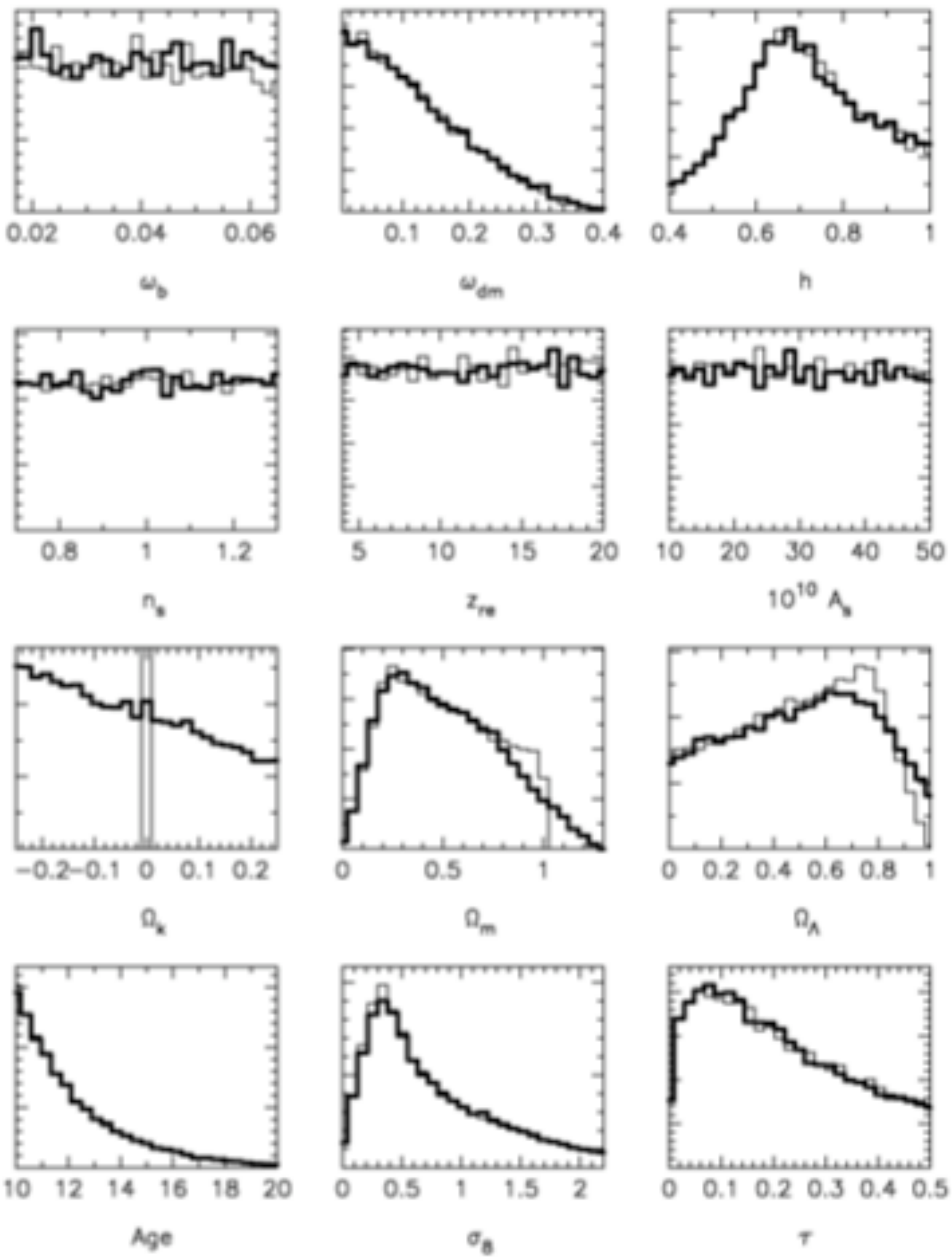
- Do cosmological data favour the Big Bang theory or the Steady State theory?
- Is the gravity law General Relativity or higher-dimensional?
- Is there evidence for a non-flat Universe?

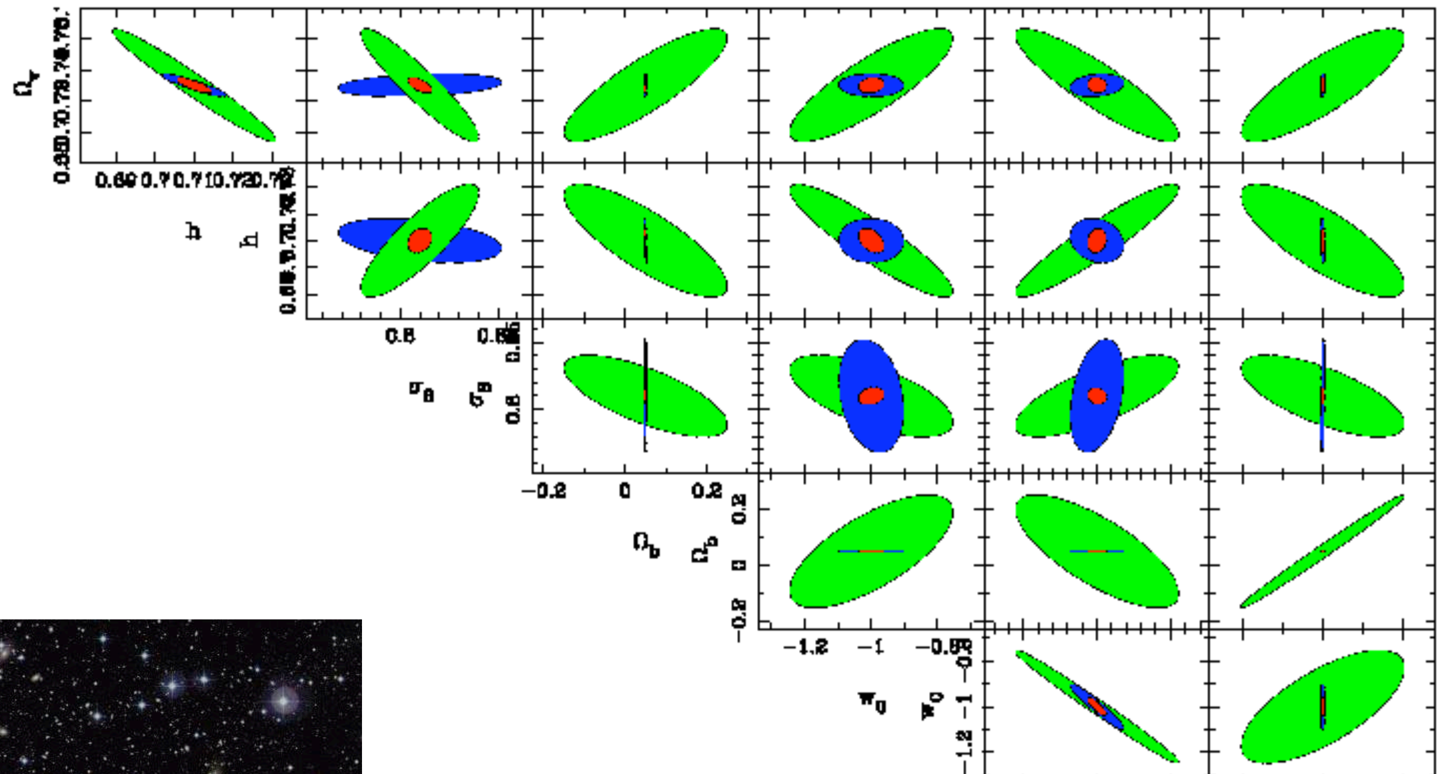
- VSA CMB experiment

(Slosar et al 2003)

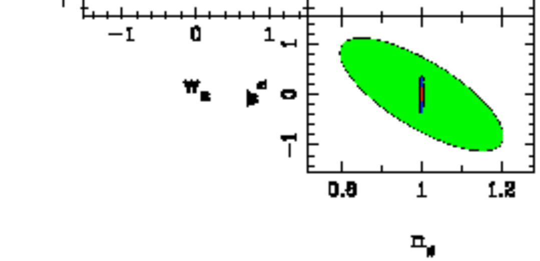
H ≈ 0.7 ± 0.1

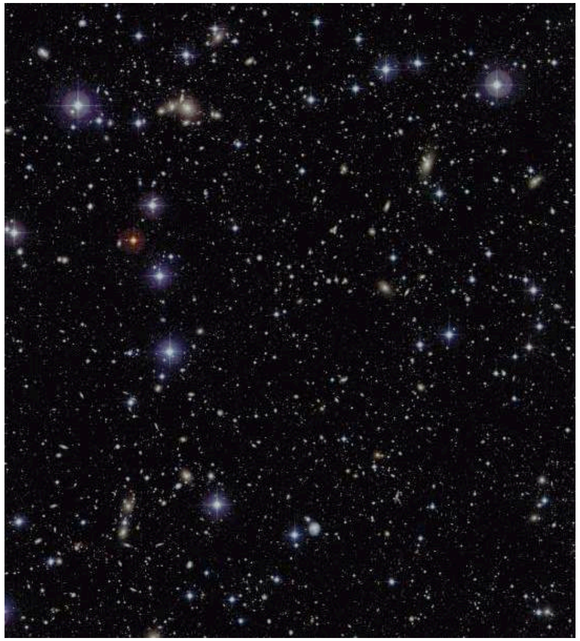
Priors: Λ≥0
10 ≤ age ≤ 20 Gyr

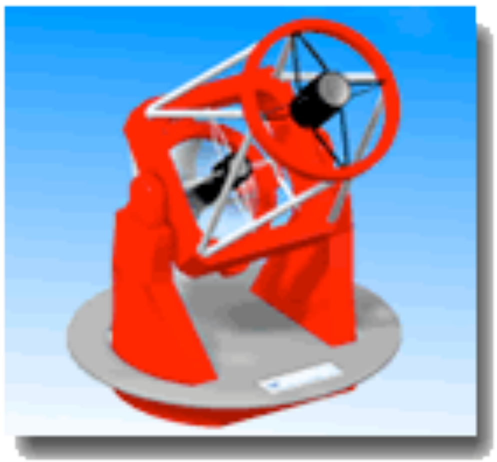




h
σ_8
Ω_b
w_0
w_a
n_s







✔

?

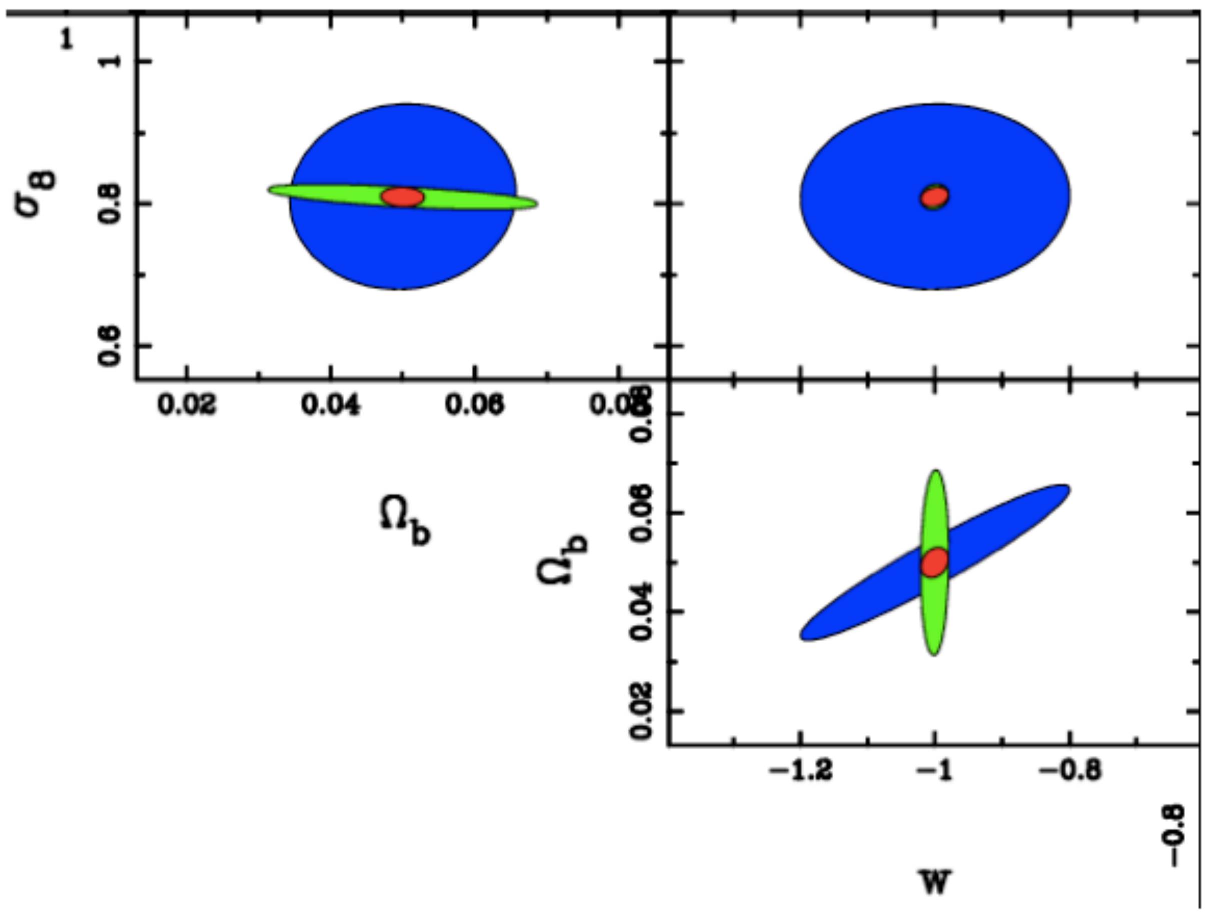
Do you change your choice?

This is the Monty Hall problem

Questions:

1. Prove that $(\mathsf{C}^{-1})_{,\alpha} = -\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\,\mathsf{C}^{-1}$
2. Prove that $(\ln \mathsf{C})_{,\alpha} = \mathsf{C}^{-1}\mathsf{C}_{,\alpha}$.
3. Prove that $\ln \det \mathsf{C} = \mathrm{Tr} \ln \mathsf{C}$.

Combining datasets



iCOSMO.org

Initiative Tools Resources Help Contact Us FAQs

Initiative for Cosmology

iCosmo

Welcome!

This site is designed to make cosmology calculations easy and pain-free. Here, you will find a host of tools and resources for performing calculations, ranging from distance calculations to cosmological error predictions for future surveys.

The site also contains a set of tutorials and links that are useful whether you are a newbie to cosmology or a seasoned professional. These resources have been made available in an easy-to-access format and will be continually updated and expanded.

Cosmology Tools:

You can perform a calculation either by using your web browser or by downloading the source code. To get started you can either go to tools, and you will be guided through each step. Alternatively, you can use the QuickStart Calculator to the right.

Cosmology Resources:

Here you will find general cosmology support materials, such as tutorials and links to external sites. To find the material you need go to resources or use the QuickStart Tutorial to the right. If you wish to create your own interactive web pages you can use the templates available here. A discussion forum for the tools and resources is provided at Cosmocoffee.

QuickStart Calculator

Ω_m	0.3	Ω_{DE}	0.7
Ω_B	0.045	w_0	-0.95
h	0.7	w_a	0.0
σ_8	0.8	n_s	1.0

QuickStart Cosmology

QuickStart Tutorial

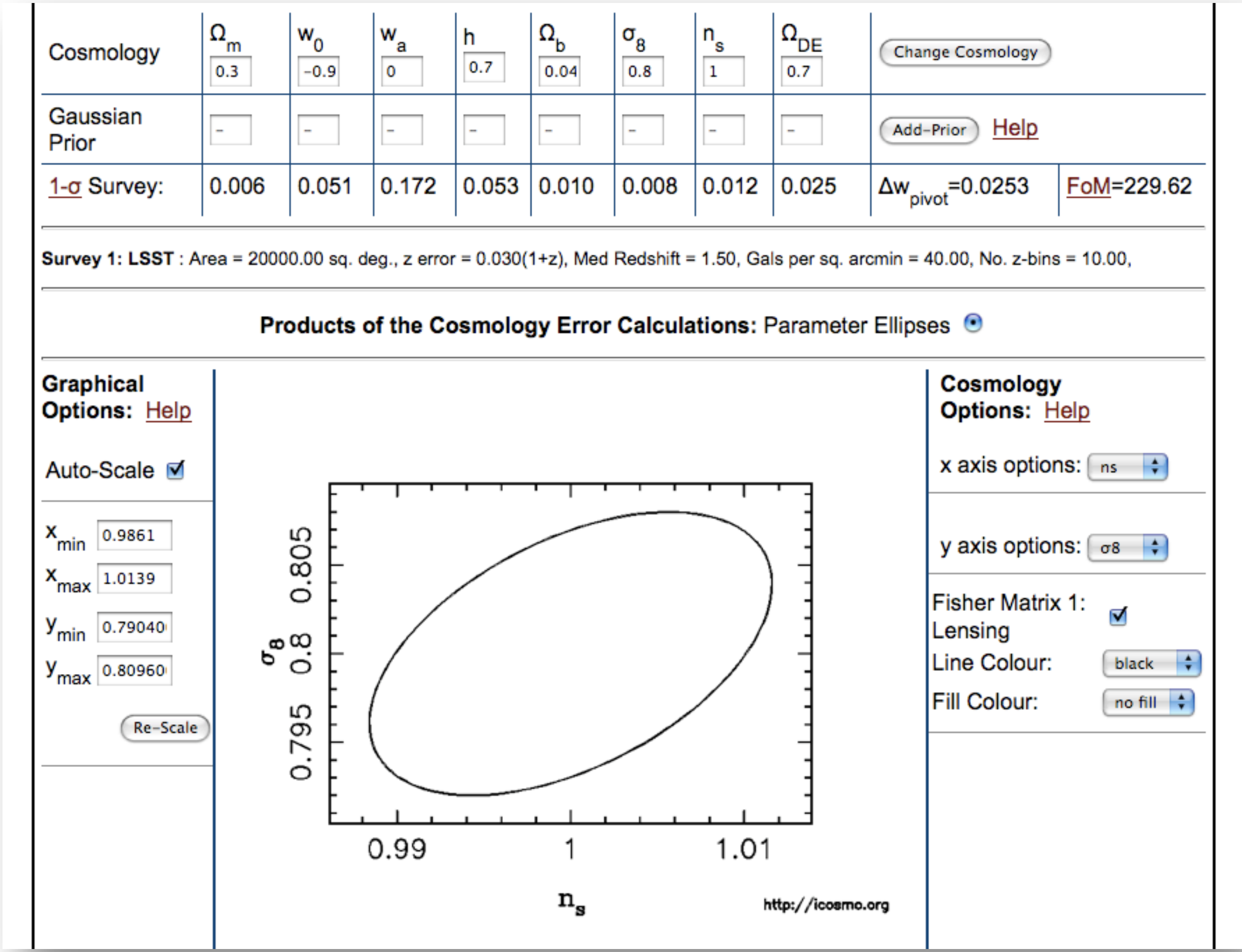
- Gravitational Lensing
- Galaxy Correlations
- CMB

QuickStart Tutorial

News:

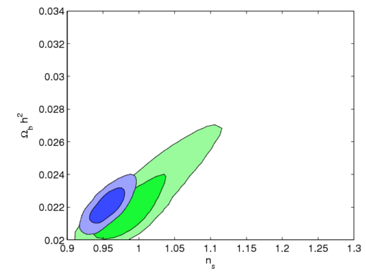
21/05/2009 - **w(z) eigenfunctions**. Module for astro-ph/0905.3383 to be included in iCosmo v1.2.
20/05/2009 - **Hardware-Software balance**. Code for astro-ph/0905.3176 can be downloaded here iCosmo PublicAstroCodes.
11/02/2009 - **Redshift Distortion & ISW**. Module for astro-ph/0902.1759 to be included in iCosmo v1.2.
21/01/2009 - **Cloud Cosmology**. Article available here. Template web pages available here.

Open source Fisher matrices



Computing posteriors

- For 2 parameters, a grid is usually possible
  - Marginalise by numerically integrating along each axis of the grid



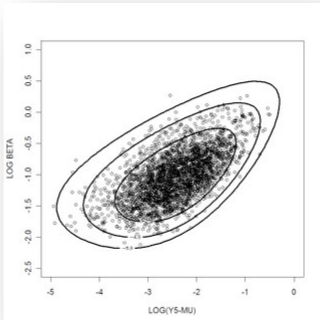
- For >>2 parameters it is not feasible to have a grid (e.g. 10 points in each parameter direction, 12 parameters = 10^{12} likelihood evaluations)

Numerical Sampling methods

- MCMC (Markov Chain Monte Carlo)
- HMC (Hamiltonian Monte Carlo)

MCMC

Aim of MCMC: generate a set of points in the parameter space whose distribution function is the same as the target density.



MCMC follows a Markov process - i.e. the next sample depends on the present one, but not on previous ones.

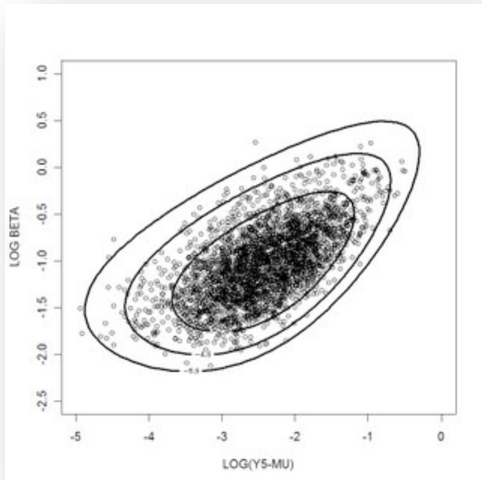
Target density

The *target density* is approximated by a set of delta functions (you may need to normalise):

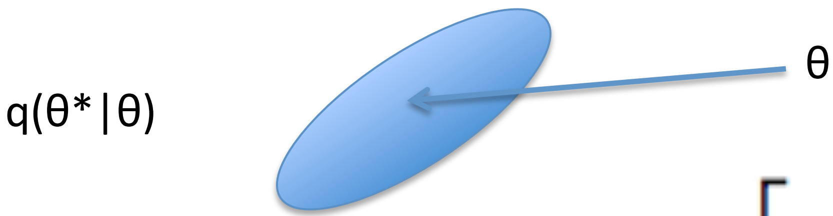
$$p(\boldsymbol{\theta}) \simeq \frac{1}{N}\sum_{i=1}^{N}\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_i)$$

and we can estimate any function f by

$$f(\boldsymbol{\theta}) \simeq \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\theta}_i).$$



Metropolis-Hastings algorithm



$$p(acceptance) = min\left[1, \frac{p(\theta^*)q(\theta^*|\theta)}{p(\theta)q(\theta|\theta^*)}\right]$$

Metropolis algorithm (special case):

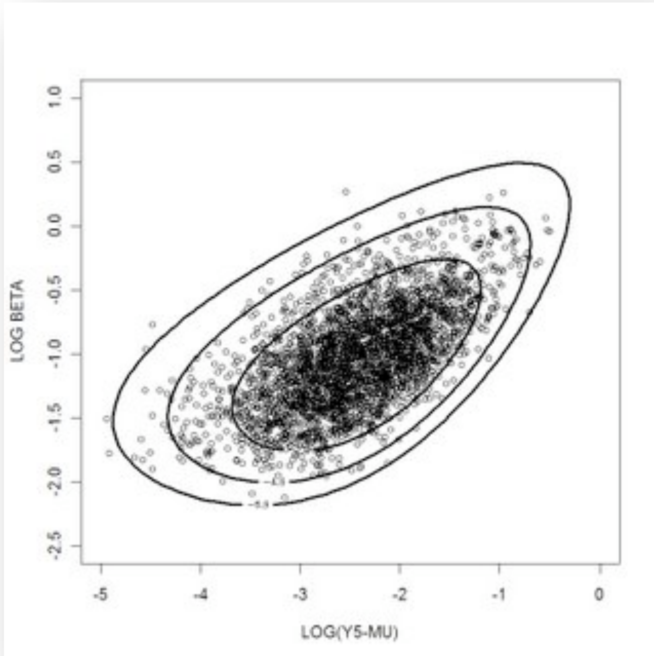
$$min\left[1, \frac{p(\theta^*)}{p(\theta)}\right]$$

MCMC Algorithm

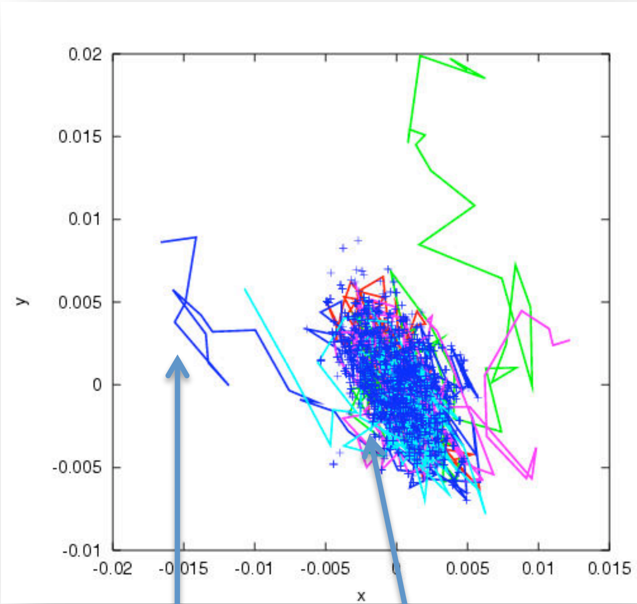
- Choose a random initial starting point in parameter space, and compute the target density.}
- Repeat:
  - Generate a step in parameter space from a proposal distribution, generating a new trial point for the chain.
  - Compute the target density at the new point, and accept it (or not) with the Metropolis-Hastings algorithm.
  - If the point is not accepted, the previous point is repeated in the chain.
- End Repeat:

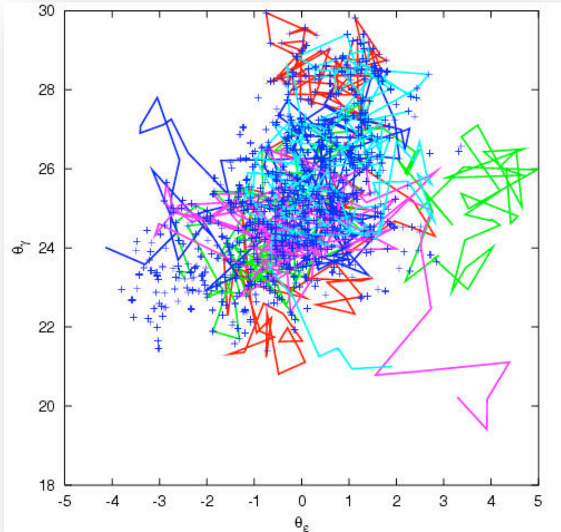
The proposal distribution

- Too small, and it takes a long time to explore the target
- Too large and almost all trials are rejected
- q ~ `Fisher size' is good.



Burn-in and convergence



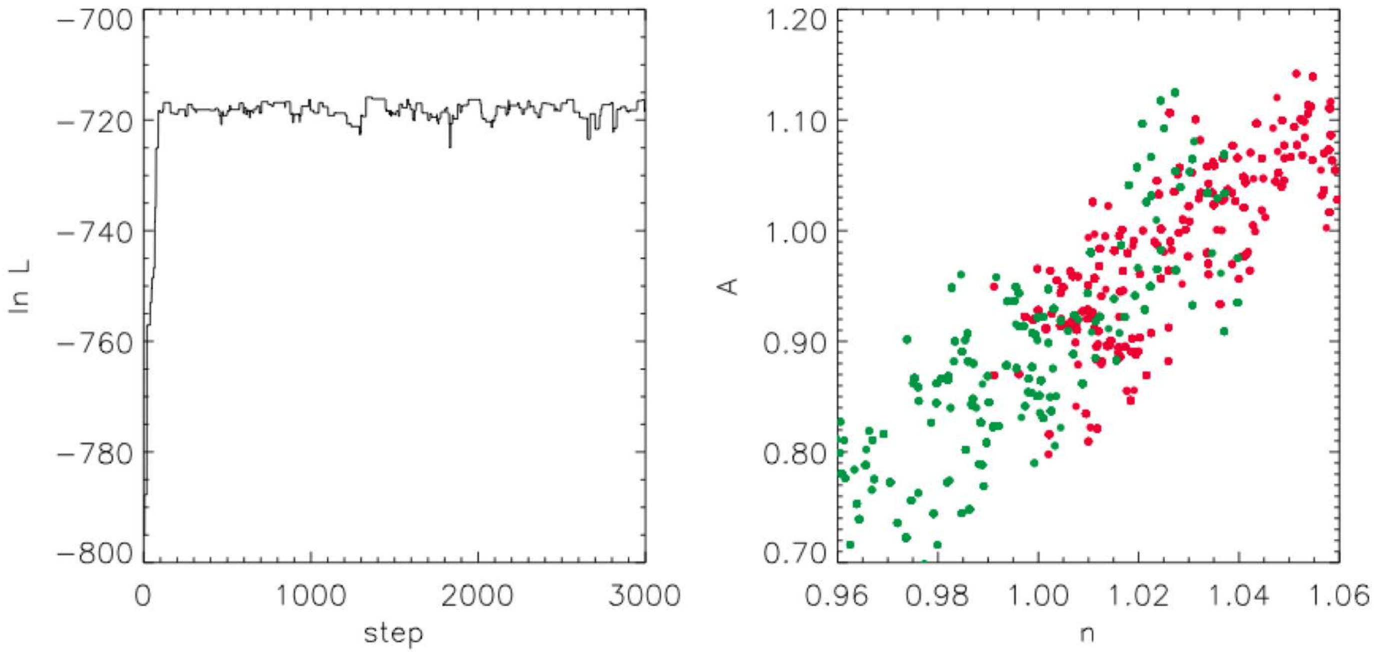


You **must** use a convergence test.
Gelman-Rubin test is most common (see notes)

"Burn-in"

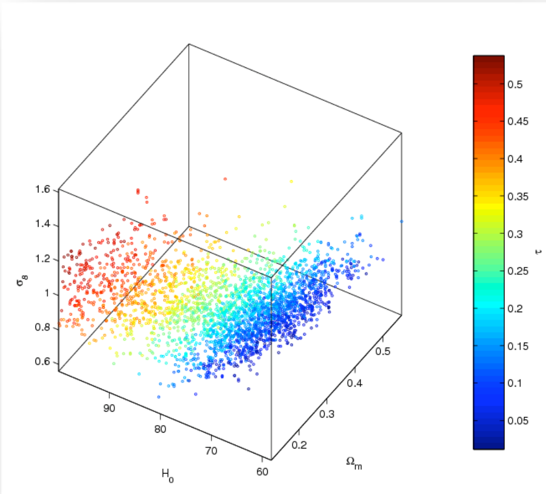
Points are correlated

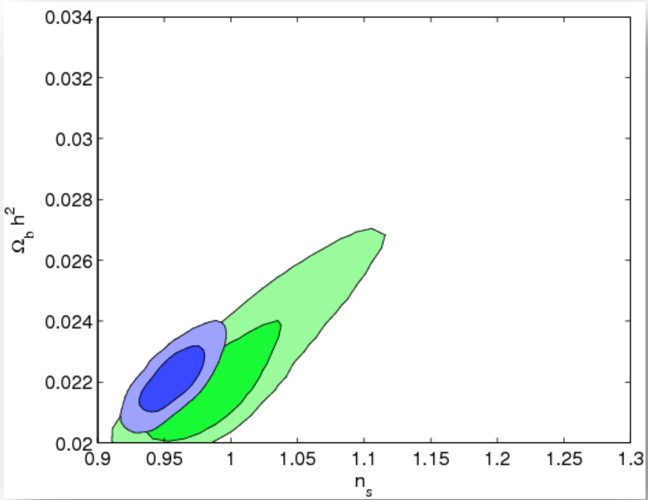
Unconverged chains



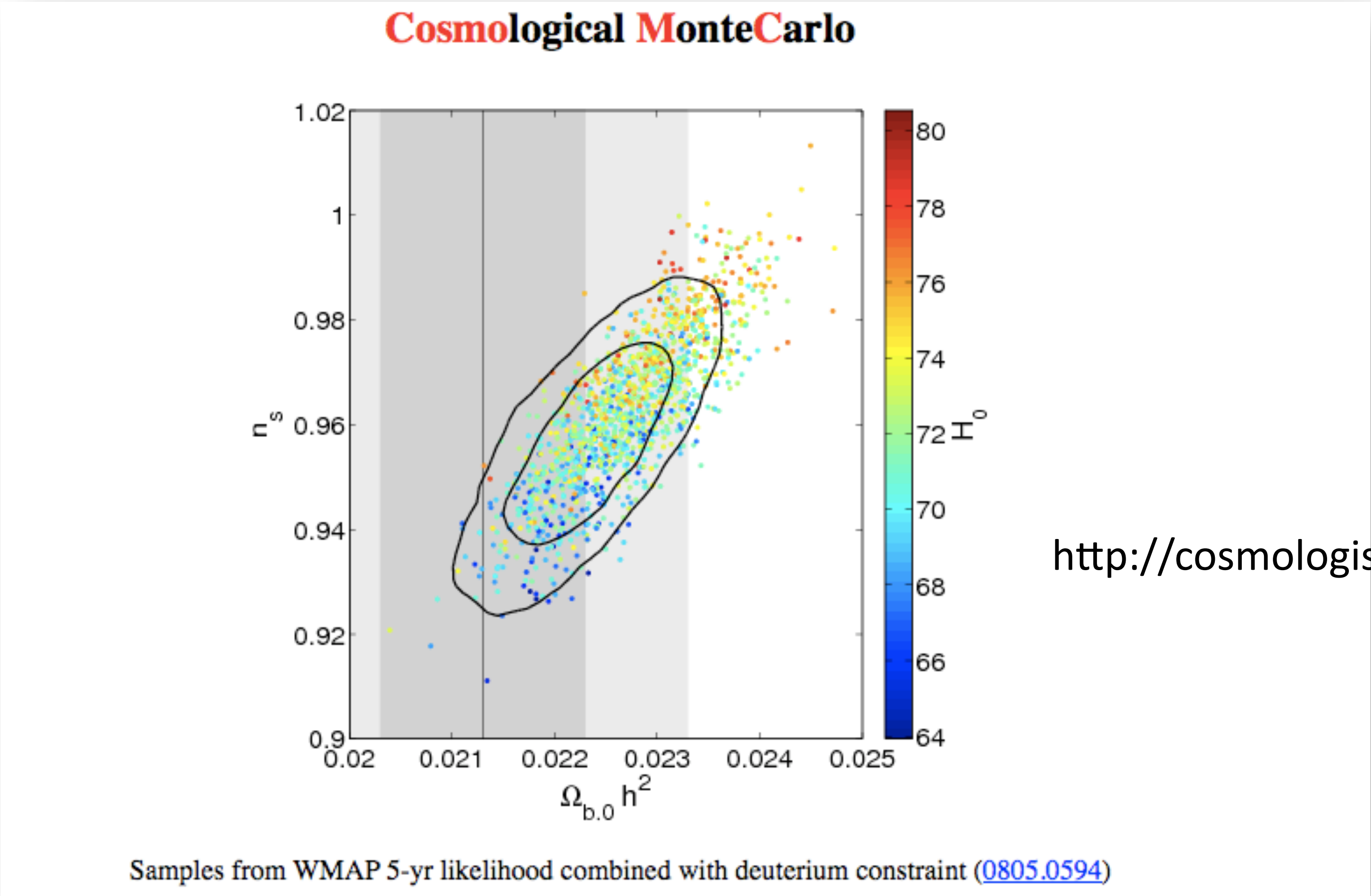
Marginalisation

- Marginalisation is trivial
  - Each point in the chain is labelled by all the parameters
  - To marginalise, just ignore the labels you don't want





CosmoMC



http://cosmologist.info/cosmomc/

Hamiltonian Monte Carlo

- We would like to increase the acceptance rate to improve efficiency

- HMC works by sampling from a *larger* parameter space:
- M auxiliary variables, one for each parameter in the model.
- Imagine each of the parameters in the problem as a coordinate.
- Target distribution = effective potential
- For each coordinate HMC generates a generalised momentum.
- It then samples from the extended target distribution in 2M dimensions.
- It explores this space by treating the problem as a dynamical system, and evolving the phase space coordinates by solving the dynamical equations.
- Finally, it ignores the momenta (marginalising, as in MCMC), and this gives a sample of the original target distribution.

Theory

- Potential $U(\theta) = - \ln p(\theta)$
- For each θ_α, generate a momentum u_α.
- K.E. $K = u^T u/2$
- Define a Hamiltonian

$$H(\boldsymbol{\theta}, \mathbf{u}) \equiv U(\boldsymbol{\theta}) + K(\mathbf{u})$$

- and define an extended target density

$$p(\boldsymbol{\theta}, \mathbf{u}) = \exp\left[-H(\boldsymbol{\theta}, \mathbf{u})\right]$$

Magic of HMC

William Rowan Hamilton

- Evolve as a dynamical system

$$\dot{\theta}_\alpha = u_\alpha$$
$$\dot{u}_\alpha = -\frac{\partial H}{\partial \theta_\alpha}$$

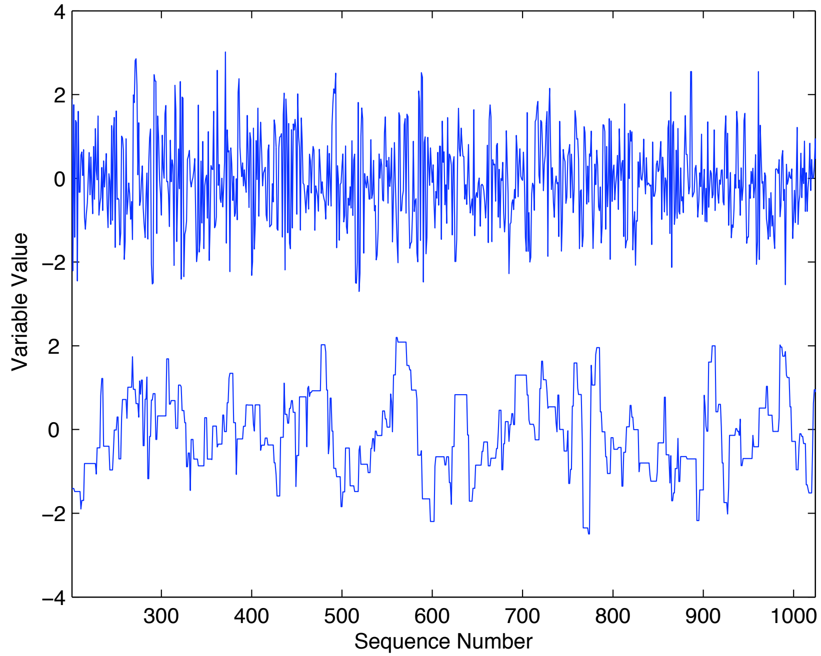
- H remains constant, so extended target density is uniform – all points get accepted!
- Also, you can make big jumps – good mixing, if you generate a new u each time a point is accepted

Complications

- Evolving the system takes time. Take big steps.
- We don't know U = - ln p (it's what we are looking for)
- We approximate U (from a short MCMC)
- H is therefore not constant
- Use Metropolis-Hastings. Accept new point with probability

$$min\left\{1, \exp\left[-H(\boldsymbol{\theta}^*, \mathbf{u}^*) + H(\boldsymbol{\theta}, \mathbf{u})\right]\right\}$$

HMC vs MCMC



HMC should be ~M times as fast as MCMC.
Typical speed-ups: factor 4.

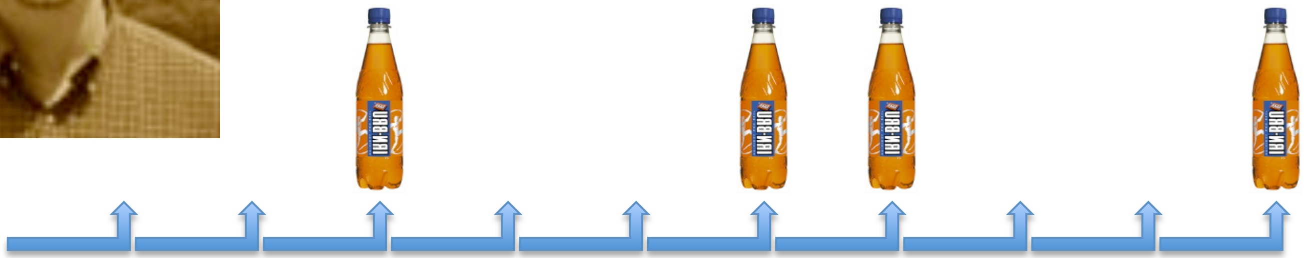
✔

P=2/3

Do you change your choice?

This is the Monty Hall problem

Binomial drinking*



Cristiano thinks about drinking a bottle of Irn Bru once every minute. He decides to drink one with probability p = 0.1.

The distribution tells us that the mean time between drinks is 1/p = 10 minutes

You check at random times and note the time of the last drink, and the next drink, and record the time between drinks

The expectation value of the time you measure between drinks is 2/p-1 =19 minutes

WHY IS IT > 10 minutes?

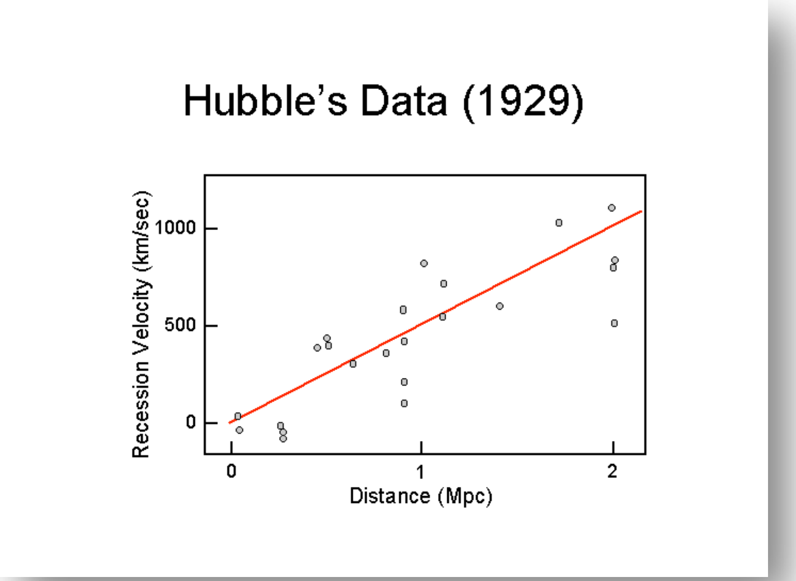
Not to be confused with Poisson drinking, which is Drinking Like a Fish

Model Selection

- Model selection: in a sense a higher-level question than parameter estimation
- Is the theoretical framework OK, or do we need to consider something else?
- We can compare widely different models, or want to decide whether we need to introduce an additional parameter into our model (e.g. curvature)
- In the latter case, using likelihood alone is dangerous: the new model will always be at least as good a fit, and virtually always better, so naïve maximum likelihood won't work.

Mr A and Mr B

- Mr A has a theory that v=0 for all galaxies.
- Mr B has a theory that v = Hr for all galaxies, where H is a free parameter.
- Who should we believe?



Bayesian approach

- Let models be M, M'
- Apply Rule 1: Write down what you want to know. Here it is $p(M|x)$ - the probability of the model, given the data.

More Bayes:

$$p(M|\mathbf{x}) = \frac{p(\mathbf{x}|M)p(M)}{p(\mathbf{x})}$$

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

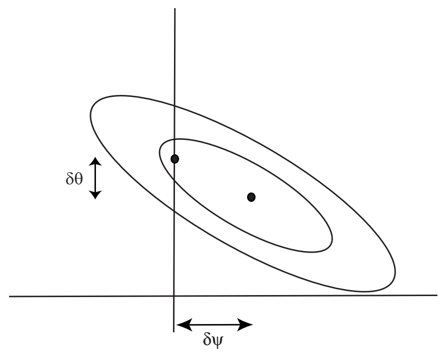
Define the Bayes factor as the ratio of evidences:

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

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

Let us assume flat priors:

$$p(\boldsymbol{\theta}|M) = (\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n)^{-1}$$



So priors do not entirely cancel:

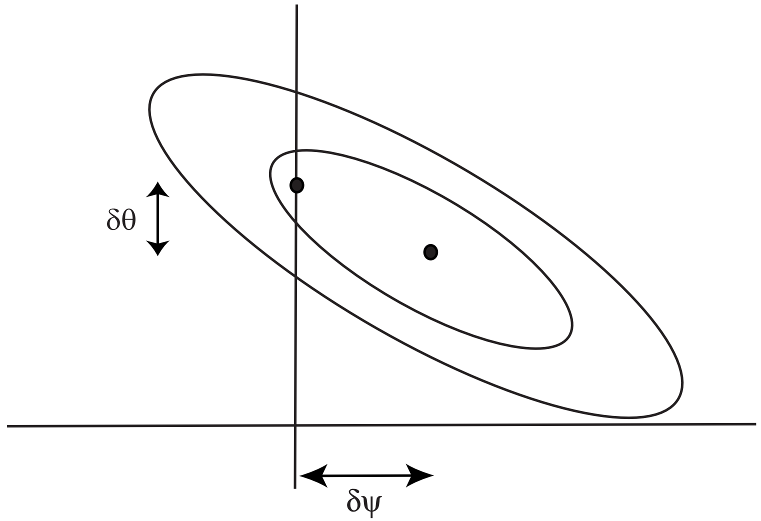
$$\frac{\Delta\boldsymbol{\theta}_1 \ldots \Delta\boldsymbol{\theta}_n}{\Delta\boldsymbol{\theta}'_1 \ldots \Delta\boldsymbol{\theta}'_{n'}} = \Delta\boldsymbol{\theta}_{n'+1} \ldots \Delta\boldsymbol{\theta}_{n'+p}$$

Laplace approximation (analogue of Fisher matrix approach, but for model selection)

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

$$\langle B\rangle = (2\pi)^{-p/2} \frac{\sqrt{\det \mathsf{F}}}{\sqrt{\det \mathsf{F}'}} \frac{L'_0}{L_0} \Delta\boldsymbol{\theta}_{n'+1} \ldots \Delta\boldsymbol{\theta}_{n'+p}$$

$$L_0' = L_0 \exp\left(-\frac{1}{2}\delta\boldsymbol{\theta}_\alpha \mathsf{F}_{\alpha\beta}\delta\boldsymbol{\theta}_\beta\right)$$



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

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

Occam’s razor… tends to favour simpler model.

Occam's razor

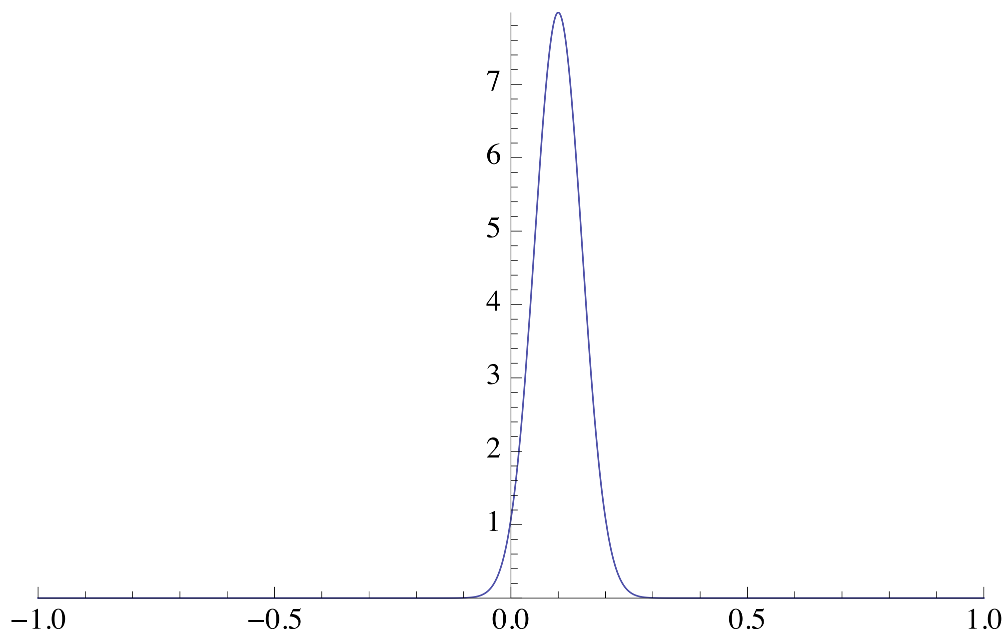
Ockham chooses a razor

- "entities should not be multiplied unnecessarily."
- "The simplest explanation for a phenomenon is most likely the correct explanation."
- "Make everything as simple as possible, but not simpler."

  \- Einstein

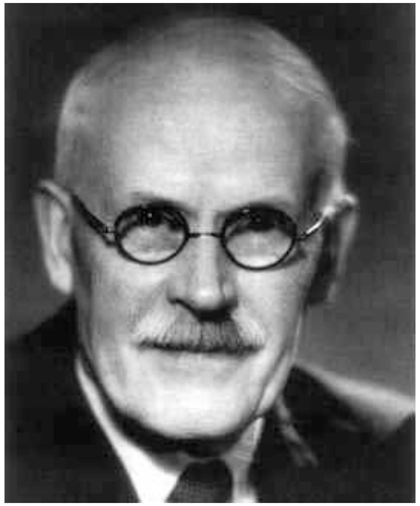
Which model is more likely?

Mr A and Mr B



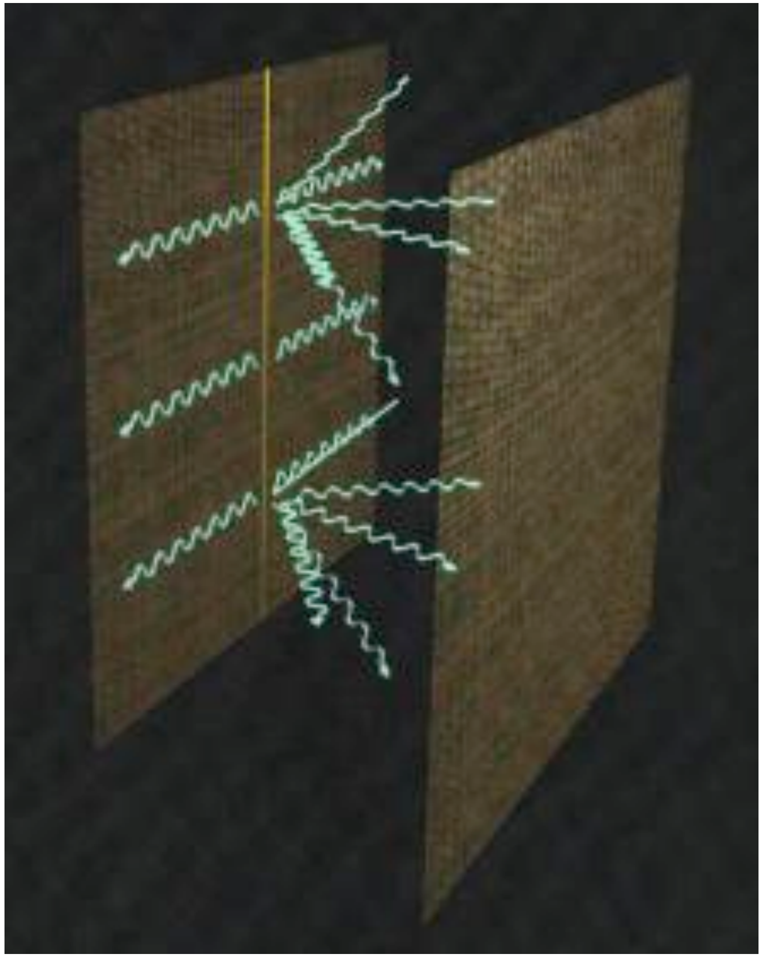
p(model with mean zero)/p(model with extra parameter))=2.2

Jeffreys' criteria



- Evidence:
- 1 < ln B < 2.5 'substantial'
- 2.5 < ln B < 5 'strong'
- ln B > 5 'decisive'
- These descriptions seem too aggressive:
  - ln B=1 corresponds to a posterior probability for the less-favoured model which is 0.37 of the favoured model

Extra-dimensional gravity?



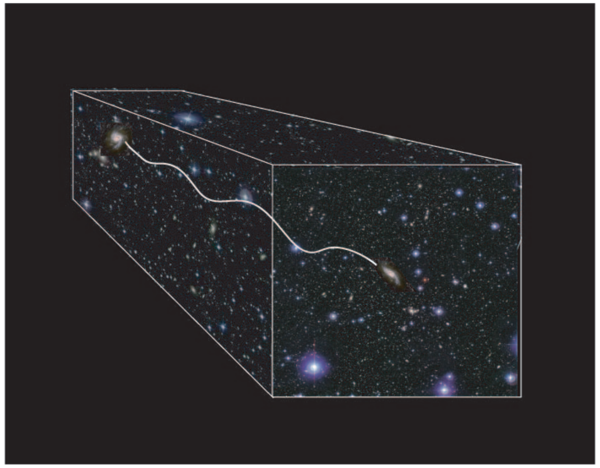
Evidence for beyond-Einstein gravity

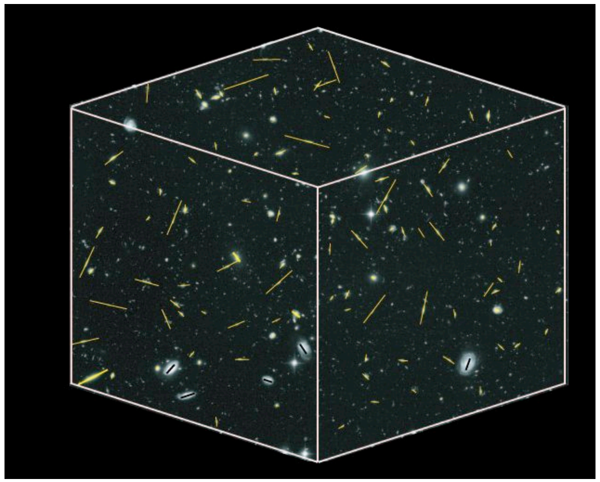
- How would we tell? Different growth rate

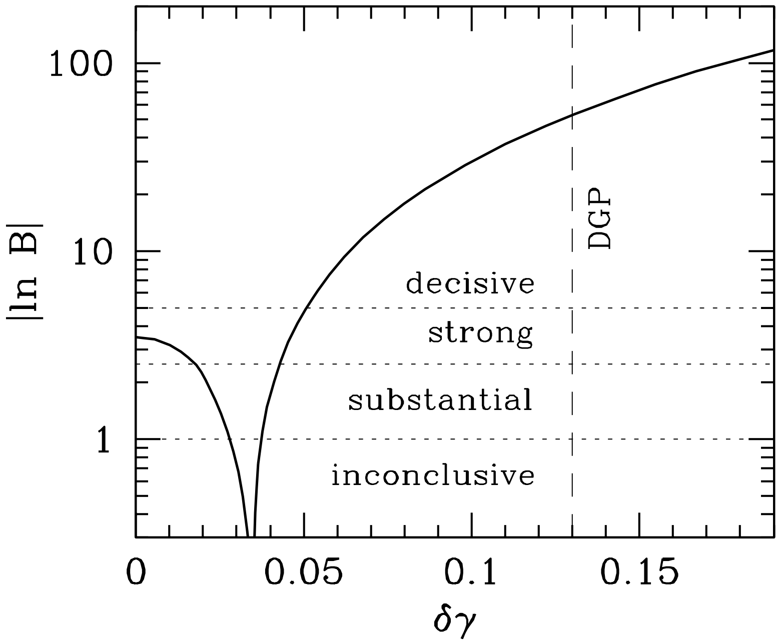
$$\frac{\delta_m}{a} \equiv g(a) = \exp\left\{\int_0^a \frac{da'}{a'}\left[\Omega_m(a')^\gamma - 1\right]\right\}$$

- γ = 0.55 (GR) 0.68 (Flat DGP model)
- Do the data demand an additional parameter, γ?

Expected Evidence: braneworld gravity?

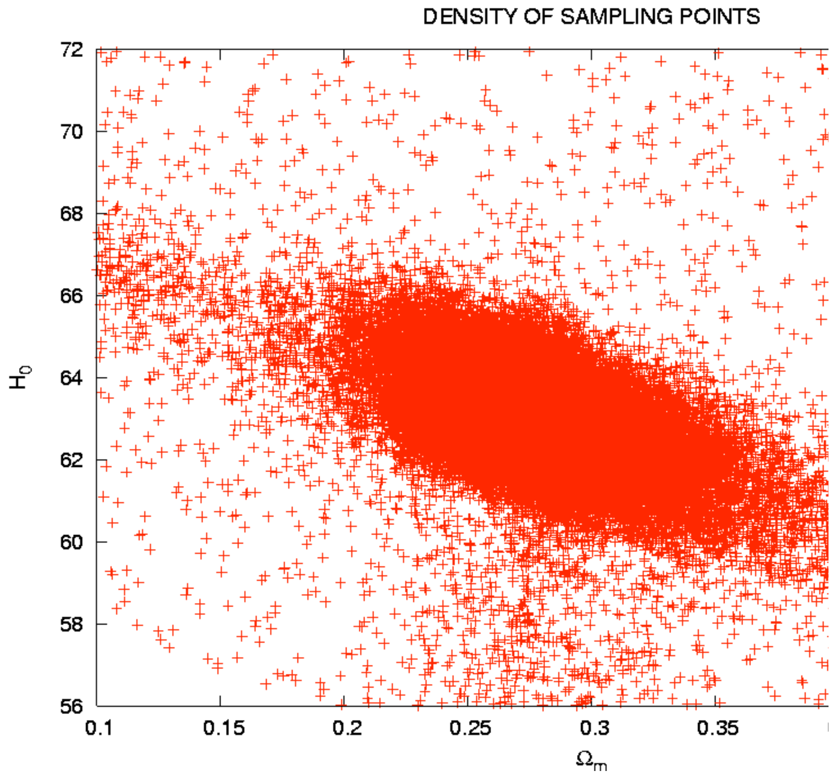






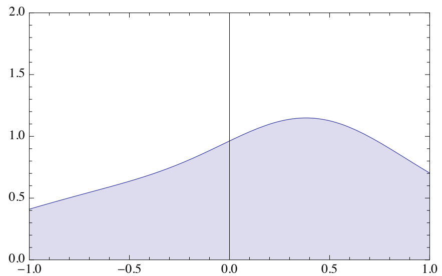
VEGAS sampling

Want a way to sample from a *separable* function of the parameters.
$P(\theta) = p_1(\theta_1)\, p_2(\theta_2) \ldots p_N(\theta_N)$

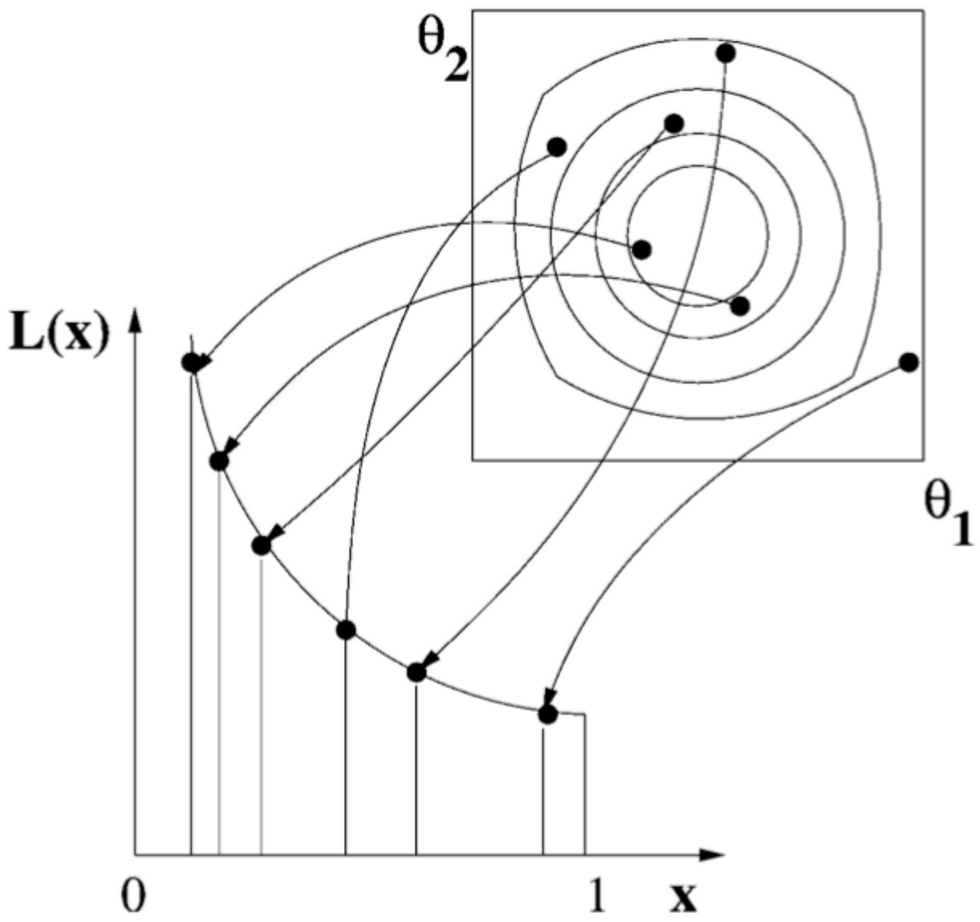


For correlated parameters, rotate the axes first (do a preliminary MCMC).

Use rejection method in 1D, for example



Nested Sampling

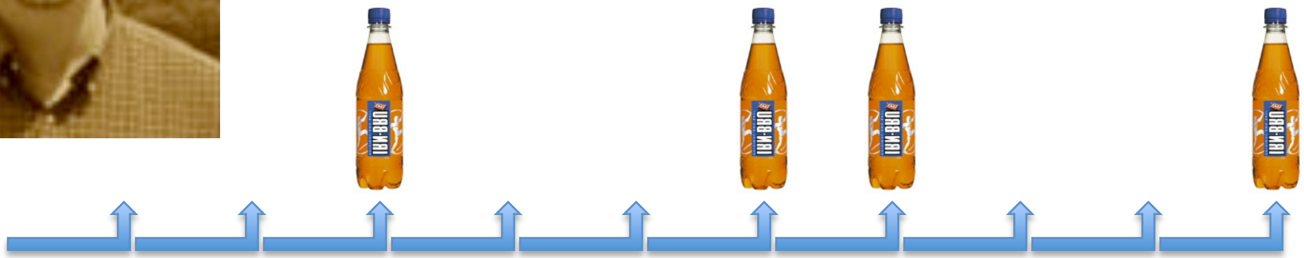


Skilling (2004)
Sample from the prior volume, replacing the lowest point with one from a higher target density.

See: CosmoNEST (add-on for CosmoMC)

Multimodal? MultiNEST

Binomial drinking*



Cristiano thinks about drinking a bottle of Irn Bru once every minute. He decides to drink one with probability p = 0.1.

The distribution tells us that the mean time between drinks is 1/p = 10 minutes

You check at random times and note the time of the last drink, and the next drink, and record the time between drinks

The expectation value of the time you measure between drinks is 2/p-1 =19 minutes

WHY IS IT > 10 minutes?

Not to be confused with Poisson drinking, which is Drinking Like a Fish

Event B: box b opened

✔

P=2/3

a b c

Do you change your choice?

This is the Monty Hall problem

Monty Hall solution

- Rule 1: write down what you want
- It is $p(a|B)$
- Now $p(a|B) = p(B|a)p(a)/p(B)$
- $p(B) = p(B,a)+p(B,b)+p(B,c)$ (marginalisation)
  - $p(B) = p(B|a)p(a) + p(B|b)p(b) + p(B|c)p(c)$
  - $p(B) = \frac{1}{2} \times \frac{1}{3} + 0 + 1 \times \frac{1}{3} = \frac{1}{2}$
- $p(a|B) = \frac{1}{2} \times \frac{1}{3} / \frac{1}{2} = \frac{1}{3}$