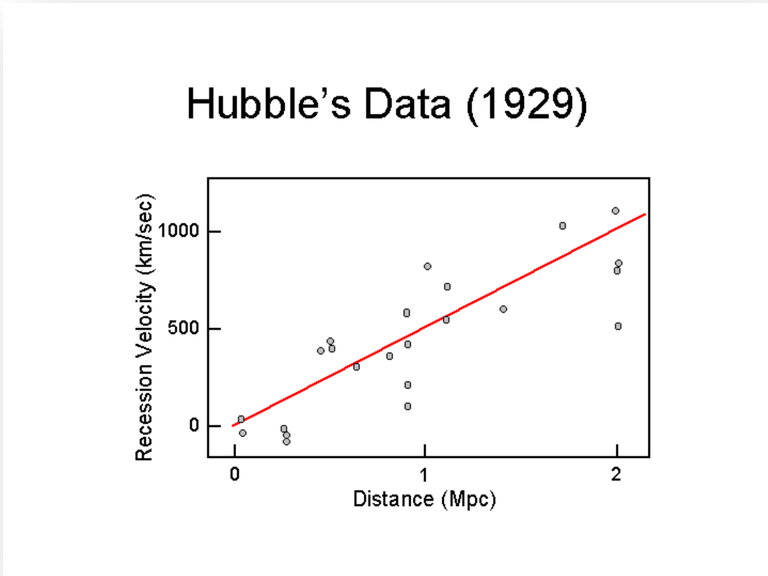
Model Selection

- 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:

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

The relative posterior probabilities of the models is:

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

Define the BAYES FACTOR as the ratio of evidences:

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

This is the relative probability of the models given the data, assuming the model priors are the same.

Model probability

- If there are N models, then the posterior probability of model M_i is

$$p(M_i|d, N) = \frac{p(M_i|N)p(d|M_i)}{\sum_{j=1}^{N} p(M_j|N)p(d|M_j)}$$

- This is a properly normalised probability
- It assumes that the models are exhaustive – one of them is correct

Occam’s razor

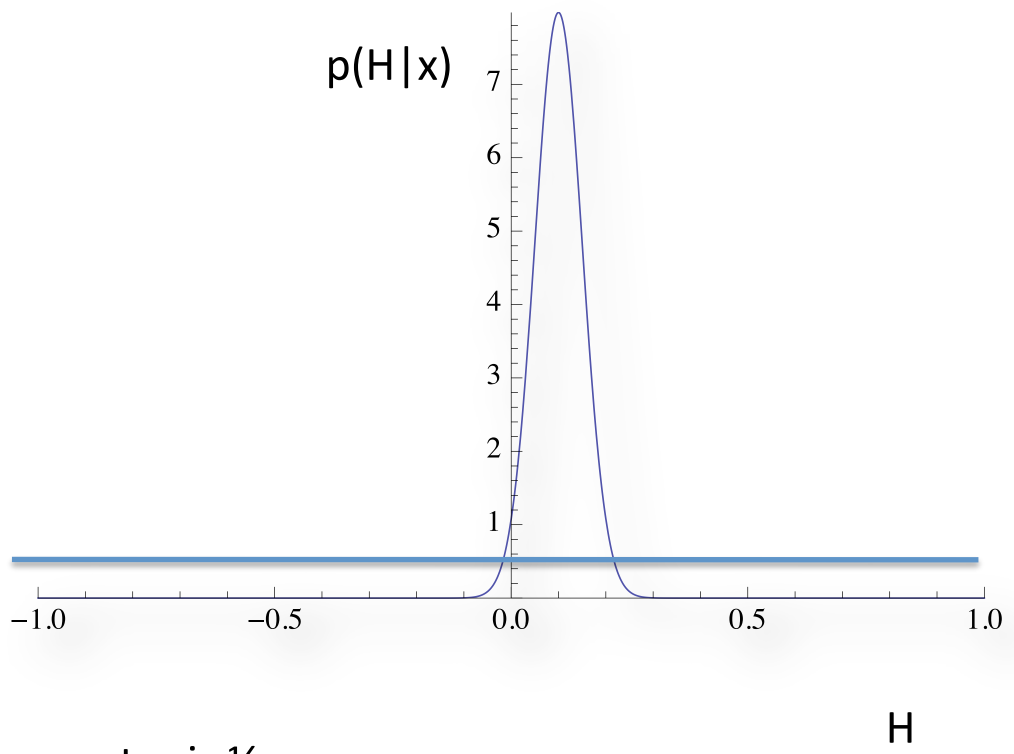
Ockham chooses a razor

- "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?

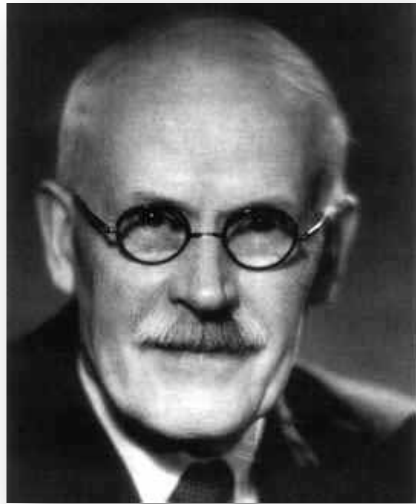
Mr A and Mr B



Prior of extra parameter is ½

$$\frac{p(\text{Model A})}{p(\text{Model B})} = \frac{1.1}{0.5} = 2.2$$

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

Higgs Boson

- Frequentist: Data consistent with 0+. 0- ruled out at 97.8% C.L.
- Bayesian analysis:

$$\frac{p(0^+|q)}{p(0^-|q)} = \frac{0.13}{0.012} = 11$$

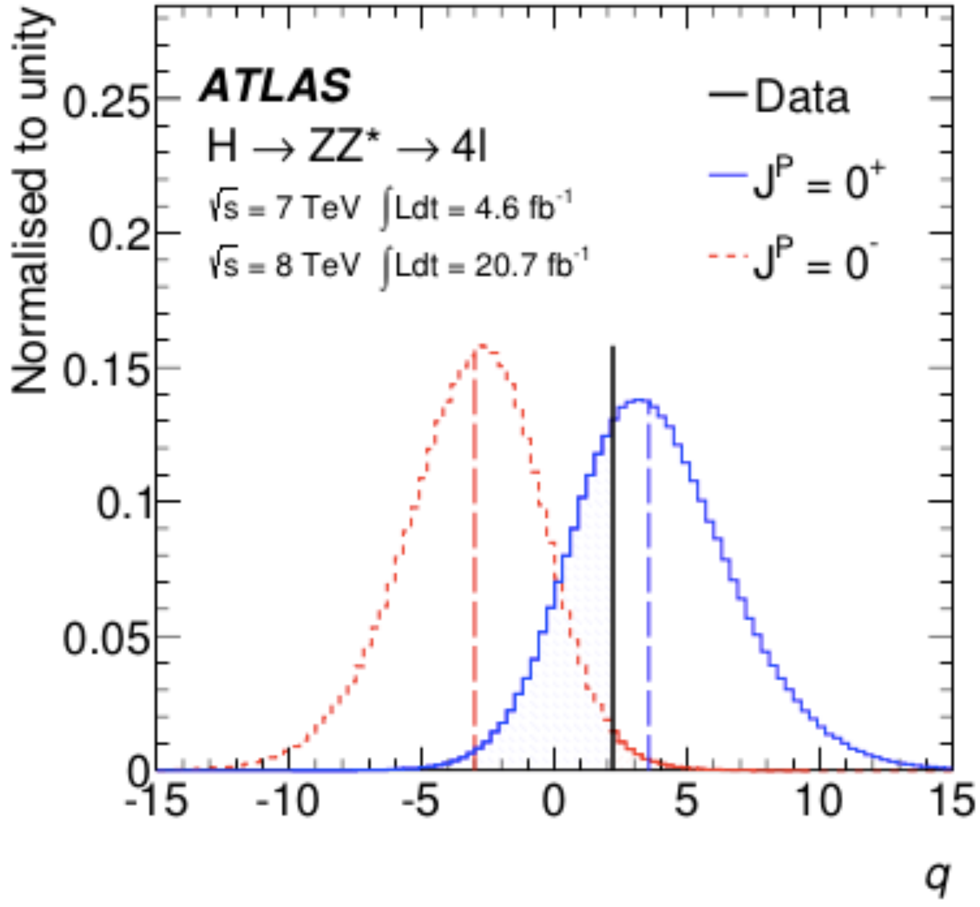
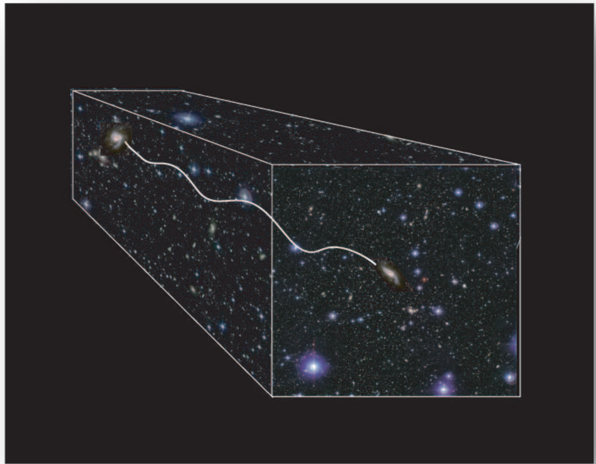
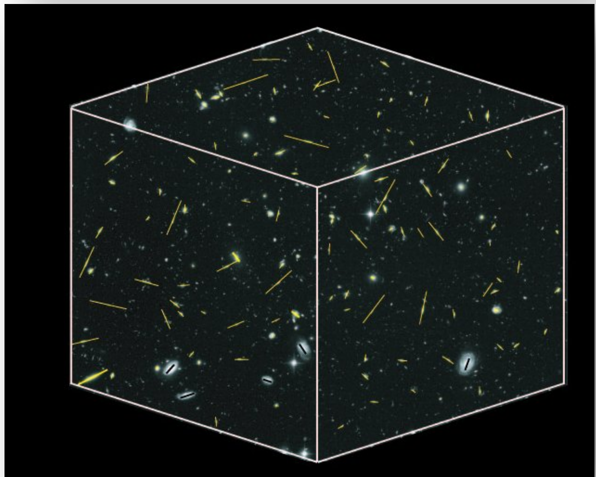
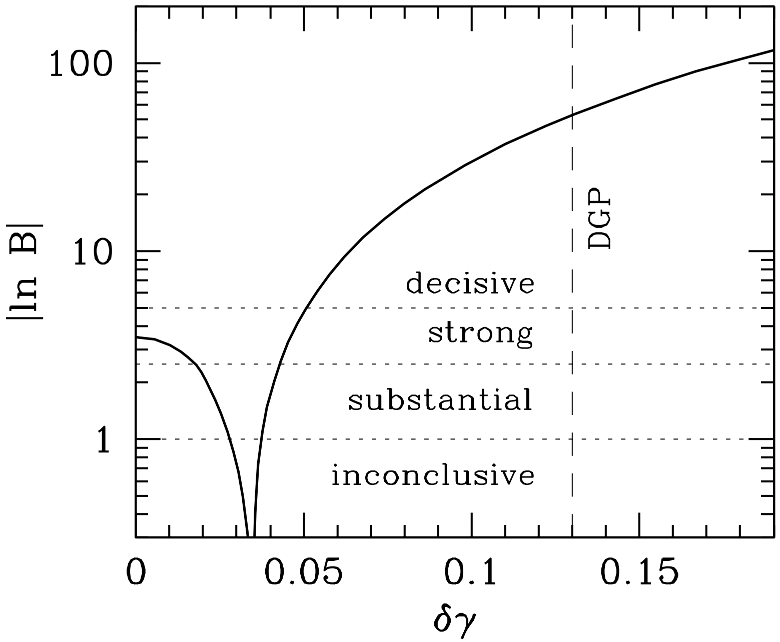


Figure 7: Expected distributions of $q = \log(\mathcal{L}(J^P = 0^+)/\mathcal{L}(J^P = 0^-))$, the logarithm of the ratio of profiled likelihoods, under the $J^P = 0^+$ and 0^- hypotheses for the Standard Model $J^P = 0^+$ (blue/solid line distribution) or 0^- (red/dashed line distribution) signals. The observed value is indicated by the vertical solid line and the expected medians by the dashed lines. The coloured areas correspond to the integrals of the expected distributions up to the observed value and are used to compute the p_0-values for the rejection of each hypothesis.

Expected Evidence: braneworld gravity?

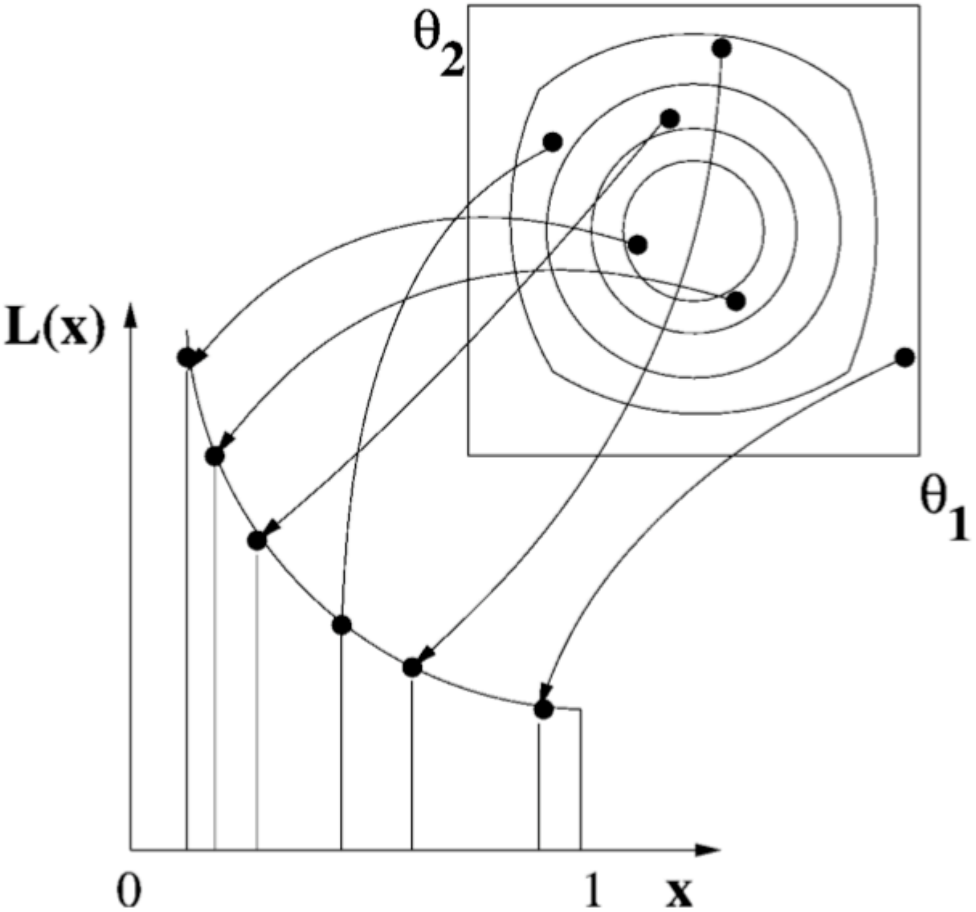






Heavens, Kitching & Verde 2007

Sampling for Model Testing: 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

Can estimate the Bayesian Evidence (which MCMC struggles with)

Hypothesis testing

- Comparing two hypotheses is fine in a Bayesian framework (it is model testing)
- Rejecting a model because the data 'don't fit' is problematic
- The formalism assumes the model set contains the true model, so if only one is tested, it has prior probability 1, and posterior probability 1 (unlike the likelihood is zero)

Hypothesis testing

- Generally frequentist: p-values
- What is p(data are 'more extreme' than observed)?
- If p<0.05 (or whatever), then 'reject the null hypothesis at the 95% level'
- But the p-value has no link to the probability of the model given the data
- Other problems – 'more extreme'; in what way?
- Can only compare alternative hypotheses

Bayesian Hierarchical Models

- Bayesian analysis of more complex, multilevel systems
- E.g., supernova Hubble diagram with errors in z as well as μ