Including unknown unknowns in Bayesian model selection

Glenn Starkman Roberto Trotta PMV

CASE WESTERN RESERVE UNIVERSITY EST. 1826

26.08.2008

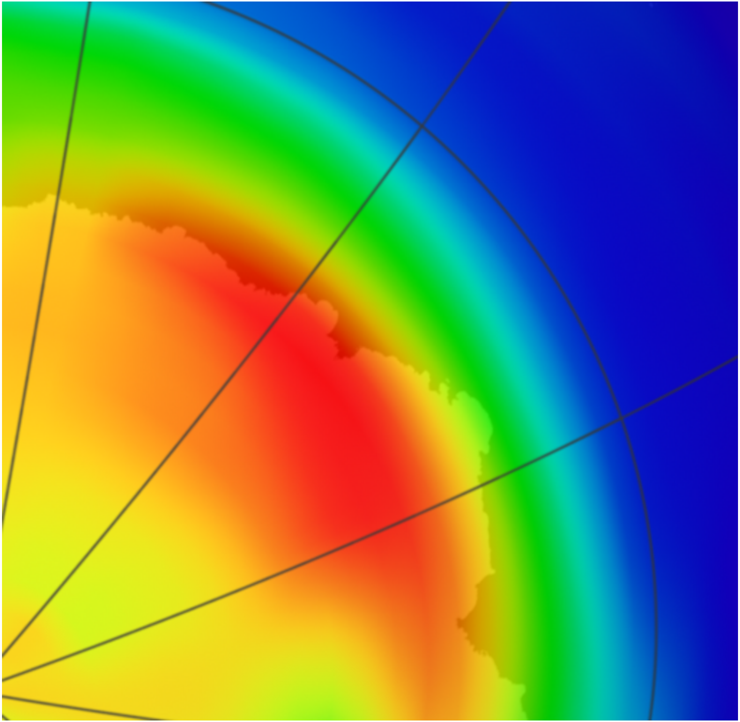
The Unknown

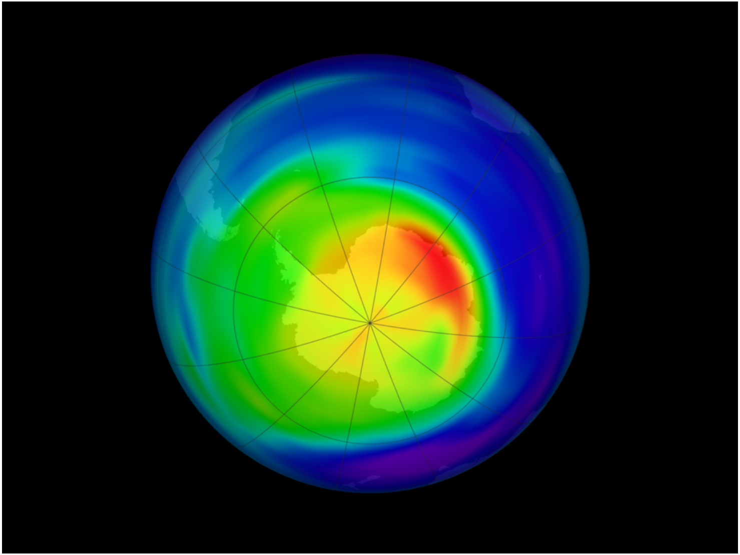
As we know,
There are known knowns.
There are things we know we know.
We also know
There are known unknowns.
That is to say
We know there are some things
We do not know.
But there are also unknown unknowns,
The ones we don't know
We don't know.

The Unknown

As we know,
There are known knowns.
There are things we know we know.
We also know
There are known unknowns.
That is to say
We know there are some things
We do not know.
But there are also unknown unknowns,
The ones we don’t know
We don’t know.

(D. Rumsfeld)

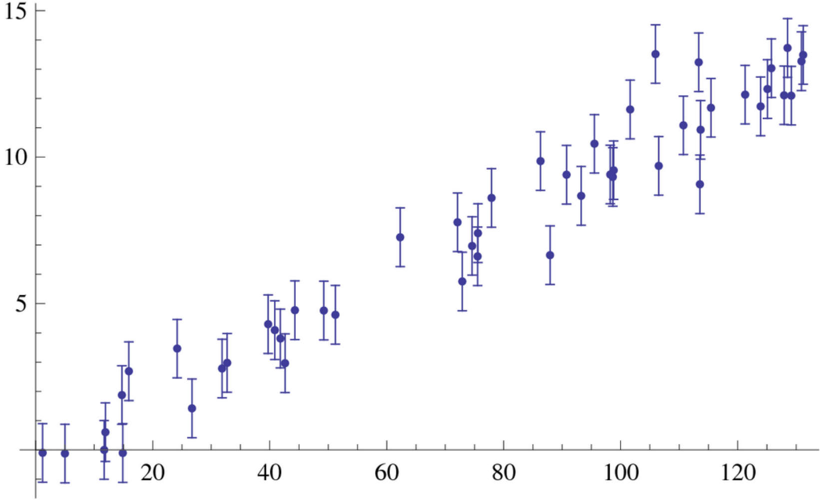




(Microwave Limb Sounder, NASA)

Outline

How to detect features?



- visual inspection
- χ^2/dof

Bayesian model selection

$$p(\mathcal{M}|d) = \frac{p(d|\mathcal{M})p(\mathcal{M})}{p(d)},$$

- evidence $p(d|\mathcal{M}) = \int d\theta \, p(d|\theta, \mathcal{M})p(\theta)$
- normalization constant $p(d) = \sum_i p(d|\mathcal{M}_i)p(\mathcal{M}_i)$ (hard to compute, normally ignored)
- priors $p(\mathcal{M}), p(\theta)$
- Bayes factor $B_{01} = \frac{p(d|\mathcal{M}_0)}{p(d|\mathcal{M}_1)}$

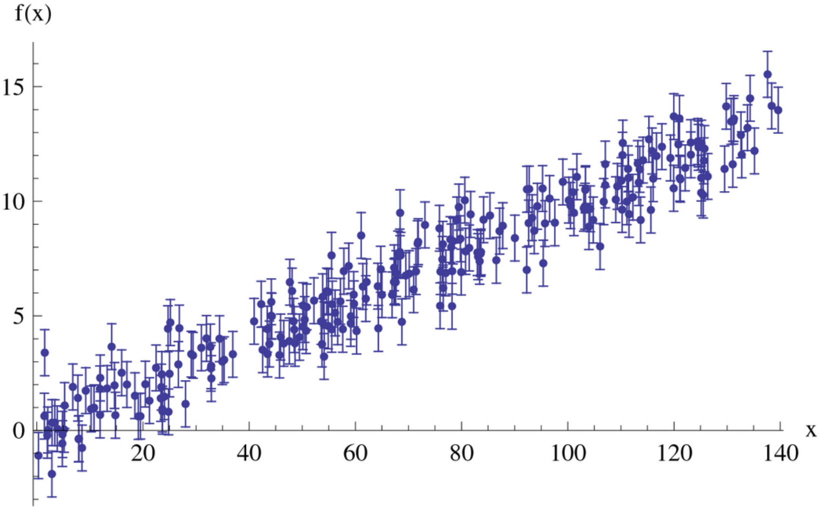
Doubt

$$p(\mathcal{X}|d) = \frac{p(d|\mathcal{X})p(\mathcal{X})}{p(d)}$$

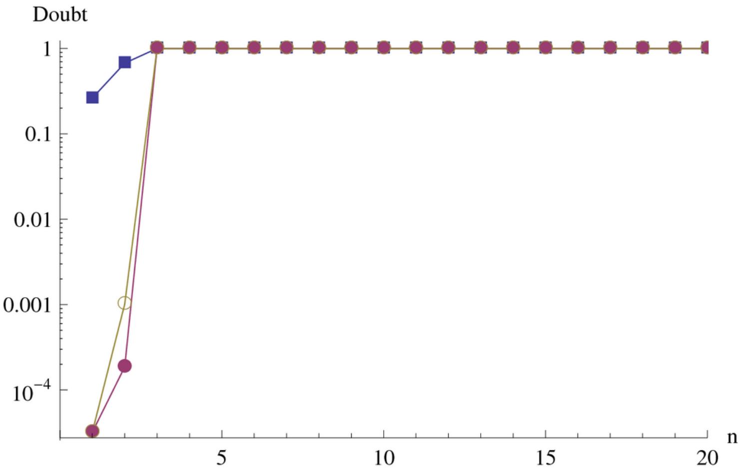
- unknown model $\mathcal{X}$
- estimate evidence
  $p(d|\mathcal{X}) = e^{-\frac{1}{2}\text{BIC}} = \mathcal{L}_{\text{max}} n^{-\frac{1}{2}k}$
- normalization constant computable
  $p(d) = p(d|\mathcal{M})p(\mathcal{M}) + p(d|\mathcal{X})p(\mathcal{X})$
- number of data points n
- number of parameters of the model k
- $\mathcal{L}_{\text{max}}$ from χ^2/dof (sort of)

Example

$$f(x) = \frac{1}{10} x$$



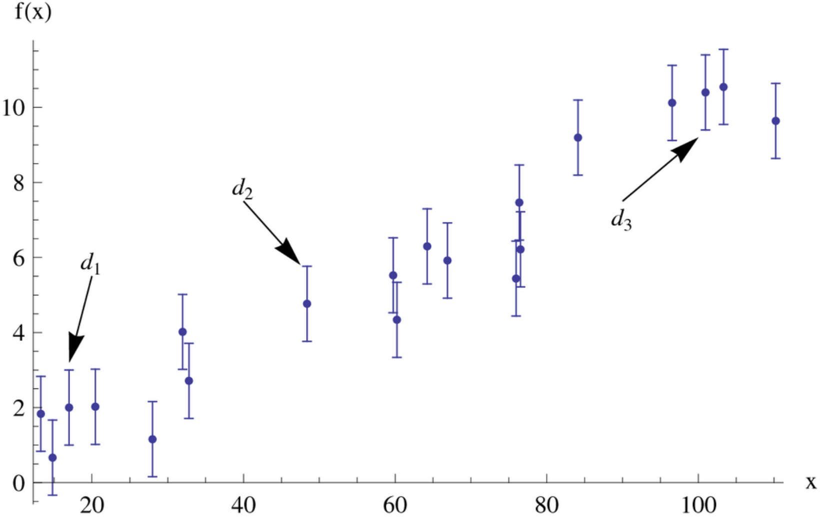
Using the wrong model distribution



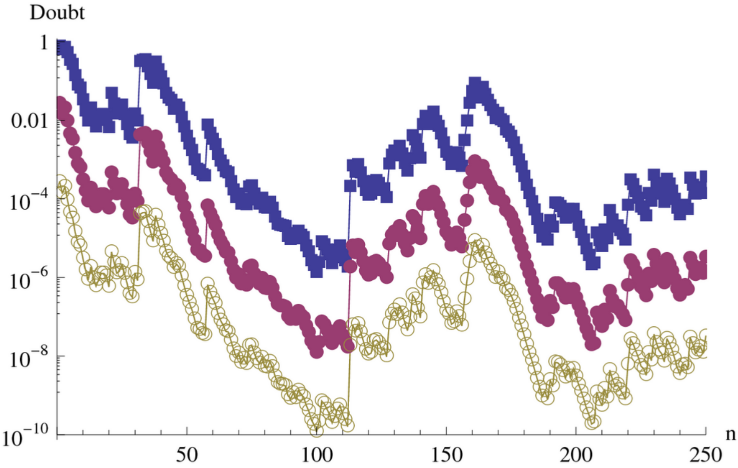
- $p(\mathcal{X}) = 10^{-1}$ (solid blue box)
- $p(\mathcal{X}) = 10^{-5}$ (solid purple circle)

Using the wrong model distribution

$$f(x) = \frac{1}{10} x$$

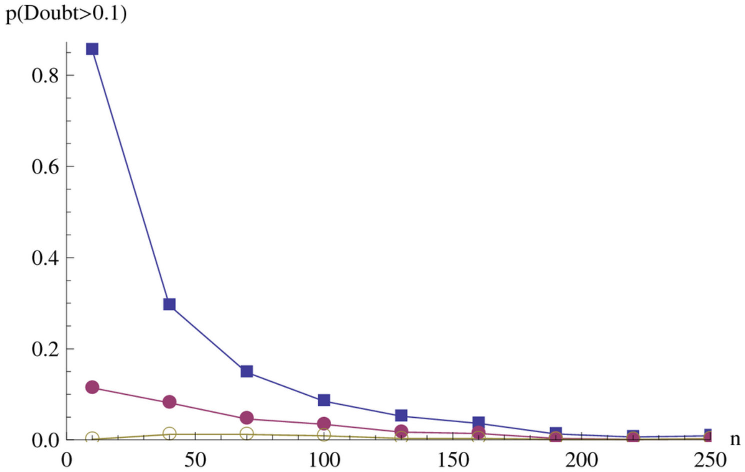


Using the correct model distribution



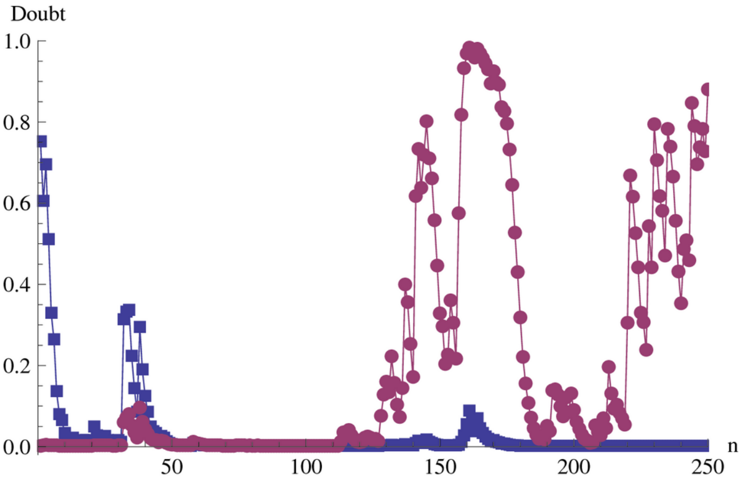
- $p(\mathcal{X}) = 10^{-1}$ (solid blue box)
- $p(\mathcal{X}) = 10^{-3}$ (solid purple circle)
- $p(\mathcal{X}) = 10^{-5}$ (open olive circle)

Tail distribution of doubt on correct model distribution



- $p(\mathcal{X}) = 10^{-1}$ (solid blue box)
- $p(\mathcal{X}) = 10^{-3}$ (solid purple circle)
- $p(\mathcal{X}) = 10^{-5}$ (open olive circle)

Estimating $\mathcal{L}_{\mathrm{max}}$



- $\chi^2/\mathrm{dof} = 1$: purple circle, too restrictive
- want $\widehat{\mathcal{L}}_{\mathrm{max}} < \mathcal{L}^{\mathrm{true}}_{\mathrm{max}}$ in e.g. 95% of cases
- correct estimator $\widehat{\mathcal{L}}_{\mathrm{max}}$: blue boxes

Conclusions – Doubt

- single number to quantify degree of (dis)belief
- works well for linear toy model
- many applications especially in cosmology