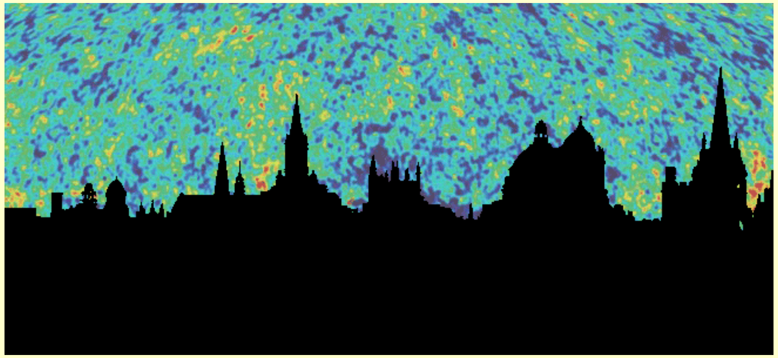


Which probability theory for cosmology?

From Bayes theorem to the anthropic principle

Roberto Trotta

Oxford Astrophysics & Royal Astronomical Society

A basic inference problem

θ : parameters

d : data

Hypothesis: M or F

Pregnant: Y/N

1) Select a random person
2) Gather data ("pregnant Y/N")
3) ... Don't get confused!

$$\mathcal{P}(\mathbf{d} = \mathbf{Y} | \boldsymbol{\theta} = \mathbf{F}) = \mathbf{0.03}$$

$$\mathcal{P}(\boldsymbol{\theta} = F | \mathbf{d} = \mathbf{Y}) \gg \mathbf{0.03}$$

$$\mathcal{P}(\text{data}|\text{hypothesis}) \neq \mathcal{P}(\text{hypothesis}|\text{data})$$

Bayesian parameter estimation

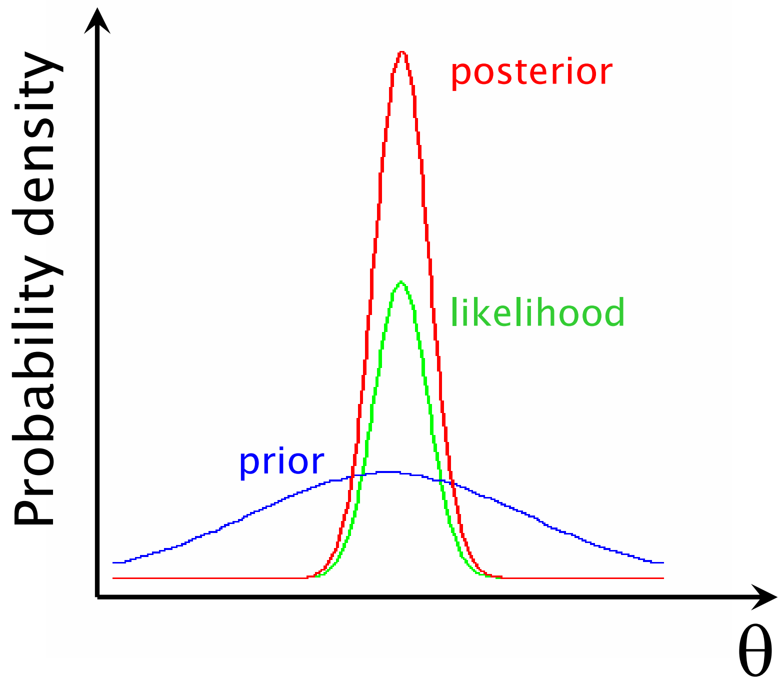
$\boldsymbol{\theta}$: parameters

$\mathbf{d}$: data

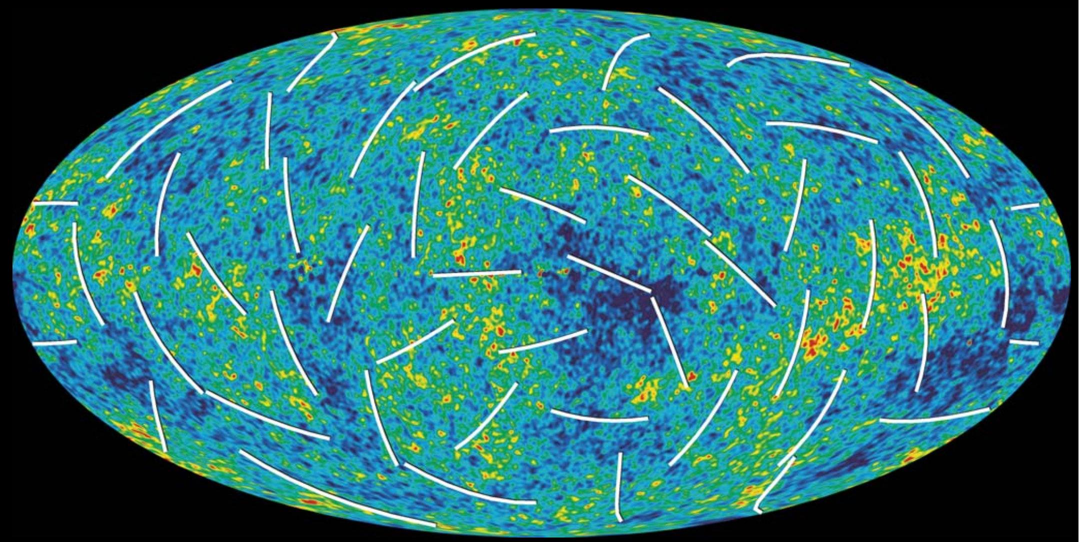
Bayes' Theorem

$$\mathcal{P}(\boldsymbol{\theta}|\mathrm{d}) = \frac{\mathrm{L}(\mathrm{d}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{P}(\mathrm{d})}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{model like}}$$



$$\mathcal{P}(\boldsymbol{\theta}, \mathrm{d}) = \underbrace{\mathrm{L}(\mathrm{d}|\boldsymbol{\theta})}_{\text{likelihood}}\underbrace{\pi(\boldsymbol{\theta})}_{\text{prior}} = \underbrace{\mathcal{P}(\boldsymbol{\theta}|\mathrm{d})}_{\text{inference}}\underbrace{\mathcal{P}(\mathrm{d})}_{\text{evidence}}$$



Probability as frequency

Repeatable sampling
Parent distribution
Asymptotically N $\to \infty$

Probability as state of knowledge

Only 1 sample
"Multiverse" approach ill-defined
N finite & limited

Two examples: hypothesis testing & anthropic reasoning

Physics of "random" experiments

Coin tossing: is the coin fair?

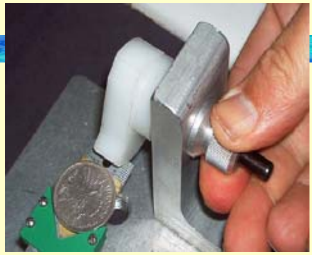
Test the null hypothesis $H_0: p = 0.5$

"The numbers p_r [the frequency with which a certain face comes up in die tossing] should, in fact, be regarded as physical constants of the particular die that we are using."

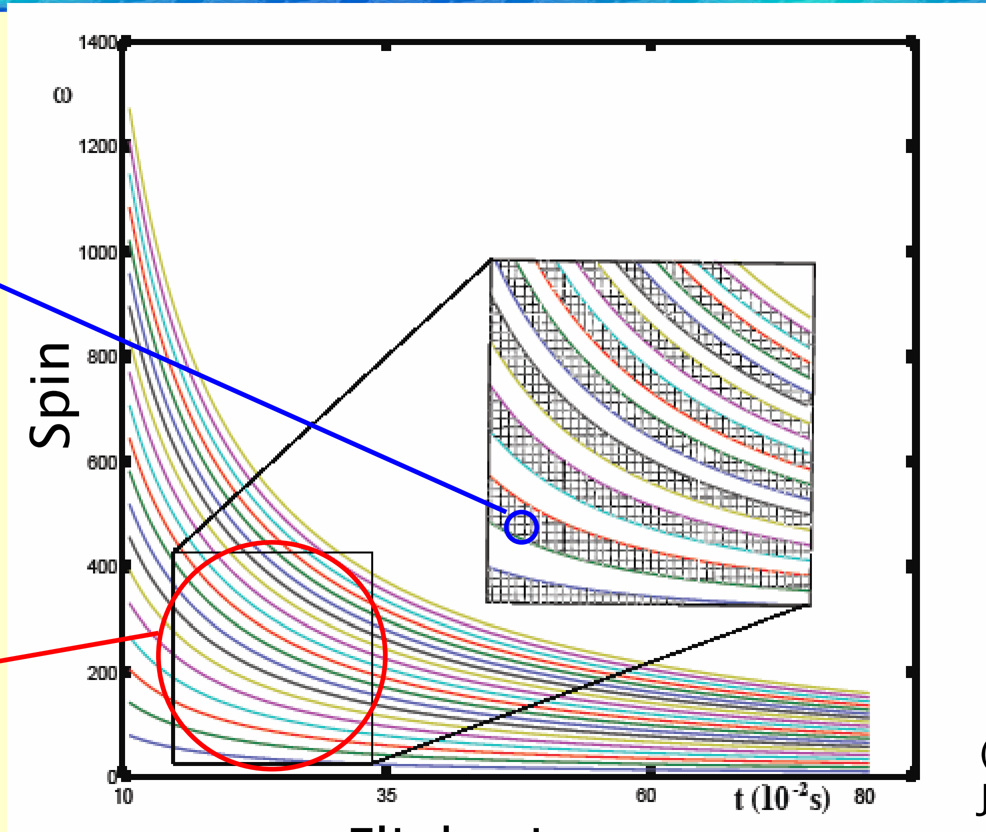
(Cramer, 1946)

Are physical probabilities meaningful?
What does it mean "to throw at random"?

Initial conditions space



p irrelevant!



Flight time

"Random" toss

With careful adjustment, the coin started heads up always lands heads up – 100% of the time. We conclude that coin-tossing is "physics" not "random".

(Diaconis et al 2004; Jaynes 1996)

Symmetric Lagrangian: $\Gamma_T = \Gamma_H$

p $\neq$ 0.5: Γ_T/Γ_H is NOT independent on location!

The nature of probability

- *Probabilistic nature of physical theories due to:*

  - 1) "Inherent" randomness

    QUANTUM MECHANICS
    (Copenhagen inter'on, collapse of the WF; consciousness?)

  - 2) Ignorance about initial conditions

    CLASSICAL (possibly chaotic) SYSTEMS

  - 3) Ignorance of our place in the cosmos

    QUANTUM MECHANICS
    (Many Worlds inter'on, all possible observations are made)

  - 4) Ignorance of relevant bits of the theory

    SCIENTIFIC PROCESS as gradual approximation to the Truth

Back to cosmology: parameters

- *Primordial fluctuations*

A, n_s, $dn/d\ln k$, features, ...

10x10 matrix M (isocurvature)

isocurvature tilts, running, ...

Planck scale (B, ω, ϕ, ...)

Inflation (V, V', V'', ...)

Gravity waves (r, n_T, ...)

- *Matter-energy budget*

Ω_κ, Ω_Λ, Ω_{cdm}, Ω_{wdm}, Ω_ν, Ω_b

neutrino sector (N_ν, m_ν, c^2_{vis}, ...)

dark energy sector ($w(z)$, c_s^2, ...)

baryons (Y_p, Ω_b)

dark matter sector (b, m_χ, σ, ...)

strings, monopoles, ...

- *Astrophysics*

Reionization (τ, x_e, history)

Cluster physics

Galaxy formation history

- *Exotica*

Branes, extra dimensions

Alignements, Bianchi VII models

Quintessence, axions, ...

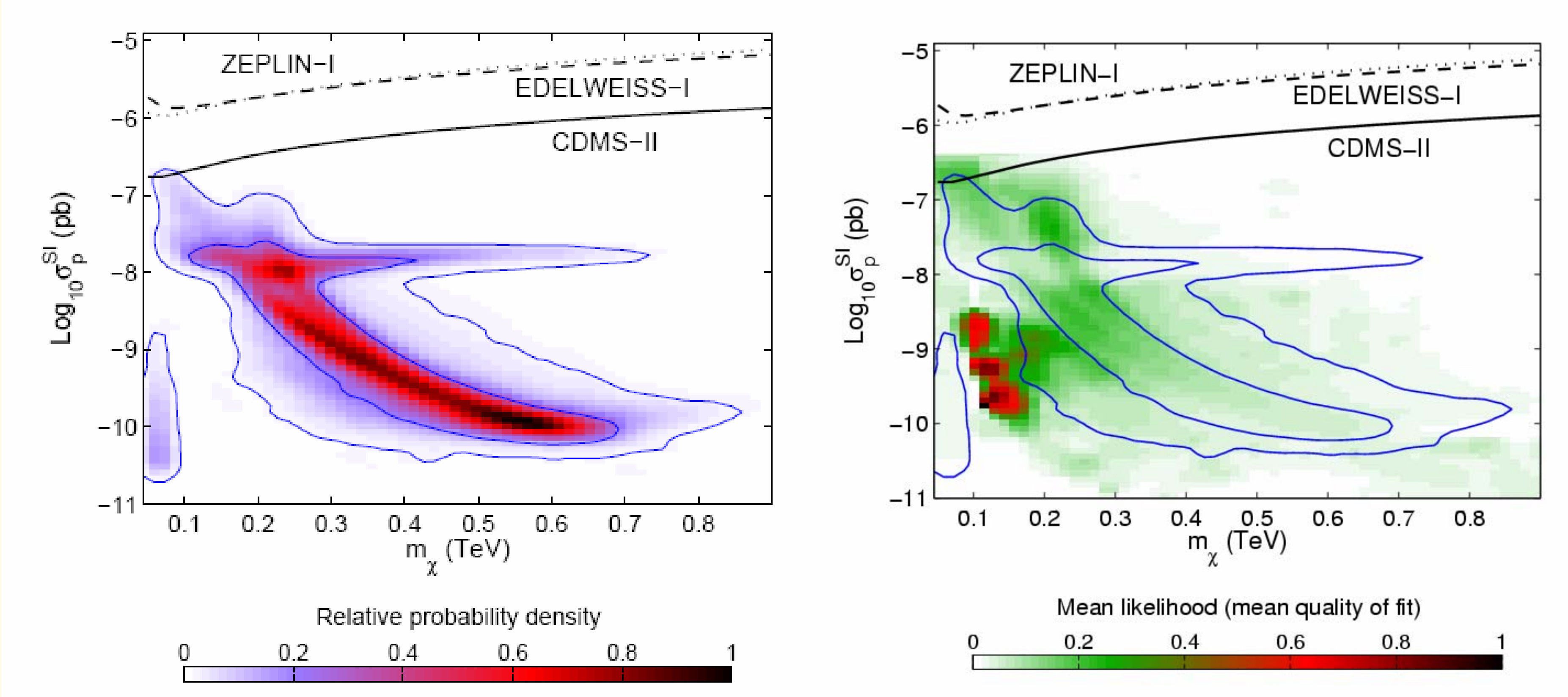
Bayes + Monte Carlo Markov Chain

- *MCMC: a procedure to draw samples from the posterior pdf*

	MCMC Bayesian	*Frequentist*
Efficiency	$\propto$ **N**	$\propto$ **k^{N}**
Nuisance params	**YES**	**undefined**
Marginalization	**trivial**	**close to impossible**
Derived params	**YES**	**need estimator**
Theoretical uncert'ies	**YES**	**only simplistic**
Prior information	**YES**	**undefined**
Model comparison	**YES**	**significance tests only**

Bayesian vs "Frequentist"

Ruiz, Trotta, Roszkowsky (2006)



Posterior pdf
Represents "state of knowledge"
High probability regions

Akin to "chi-square" statistics
Goodness of fit test
Quality of fit regions

Bayesian model comparison

Goal: to compare the "performance" of two models against the data

the model likelihood
("evidence")

$$\mathcal{P}(\mathbf{d}|\mathcal{M}) = \int_{\mathbf{\Omega}} \mathbf{L}(\mathbf{d}|\boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta}, \mathcal{M})\mathrm{d}\boldsymbol{\theta}$$

the posterior prob'ty
of the model given the data

$$\mathcal{P}(\mathcal{M}|\mathbf{d}) \propto \mathcal{P}(\mathbf{d}|\mathcal{M})\pi(\mathcal{M})$$

The Bayes factor
(model comparison)

$$B_{01} = \frac{\mathcal{P}(\mathcal{M}_0|\mathbf{d})}{\mathcal{P}(\mathcal{M}_1|\mathbf{d})}$$

$\|\ln B_{01}\|$	Odds	Interpretation
< 1	*< 3:1*	*not worth the mention*
< 2.5	*< 12:1*	*moderate*
< 5	*< 150:1*	*strong*
>5	*> 150:1*	*decisive*

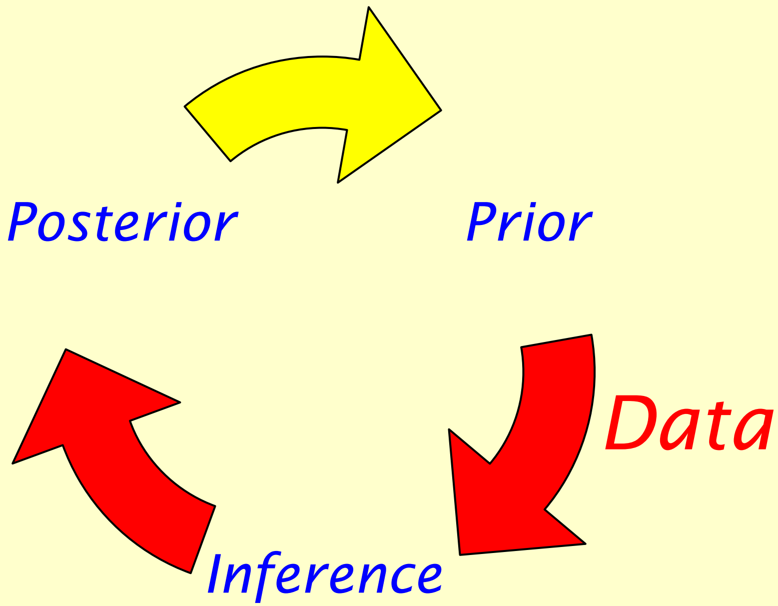
Jeffreys' scale for the strength of evidence

The role of the prior

- *Parameter inference*

Prior as "state of knowledge"

Updated to posterior through the data & Bayes Theorem



- *Model comparison*

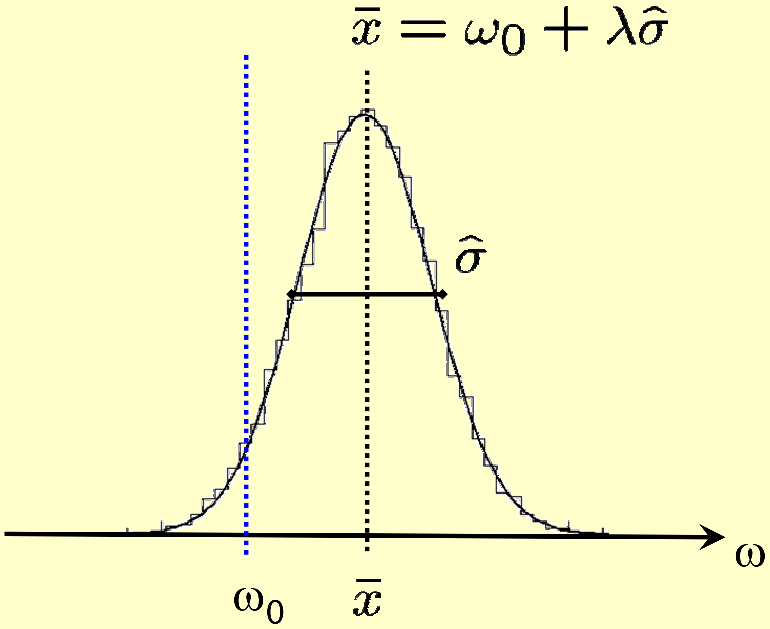
Prior inherent to model specification

Gives available model parameter space

$$\mathcal{M} \equiv \{\theta, \pi(\theta|\mathcal{M})\}$$

An automatic Occam's razor

- *The Bayes factor balances quality of fit vs extra model complexity:*



Model 0: $\omega = \omega_0$

Model 1: $\omega \neq \omega_0$ with $\pi(\omega)$

For "informative" data

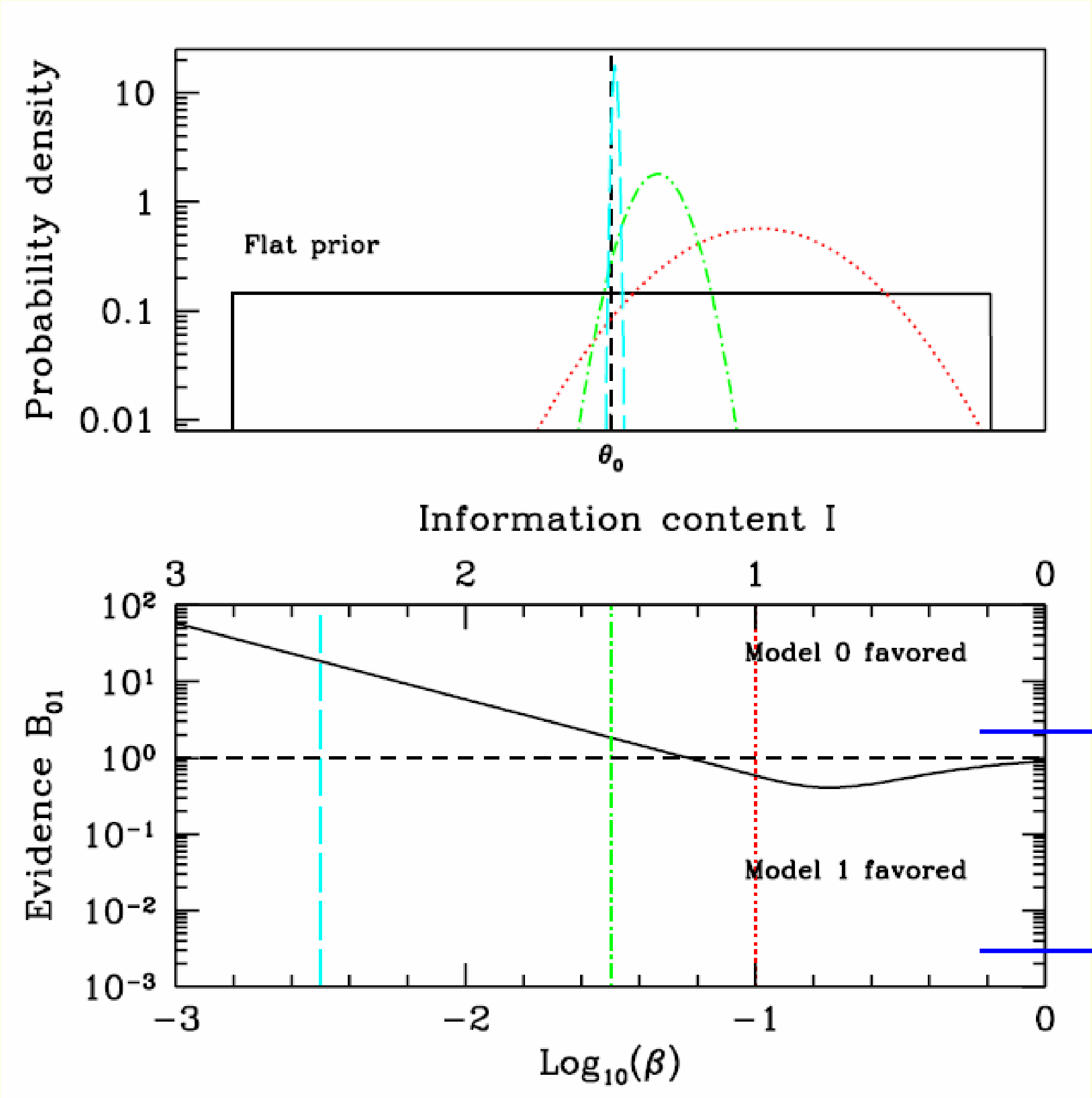
$$\ln B_{01} = I - \frac{\lambda^2}{2}$$

I = ln(prior width / likelihood width) ≥ 0

= "wasted" volume of parameter space

= amount by which our knowledge has increased

Lindley's paradox



Frequentist rejection test for H_0

λ = 1.96 for all 3 cases

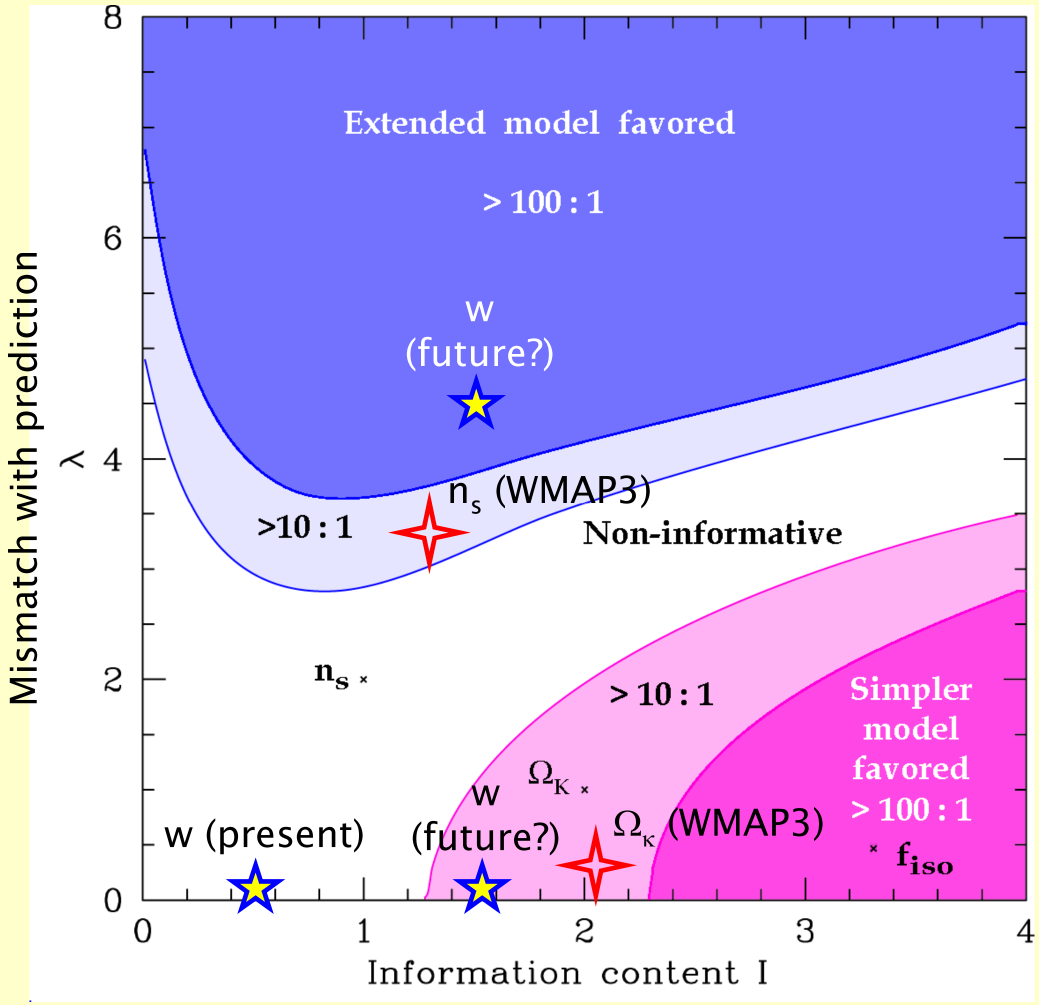
but different information content of the data

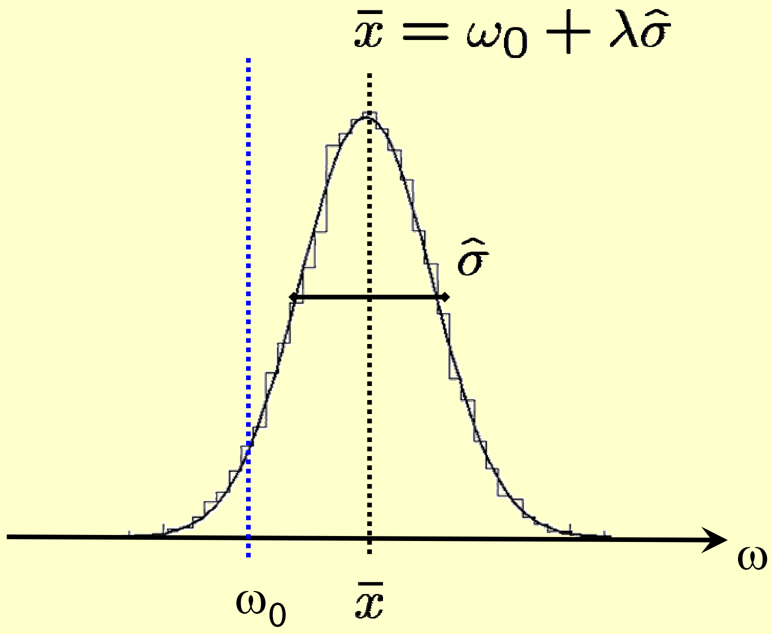
simpler model

model with 1 extra parameter

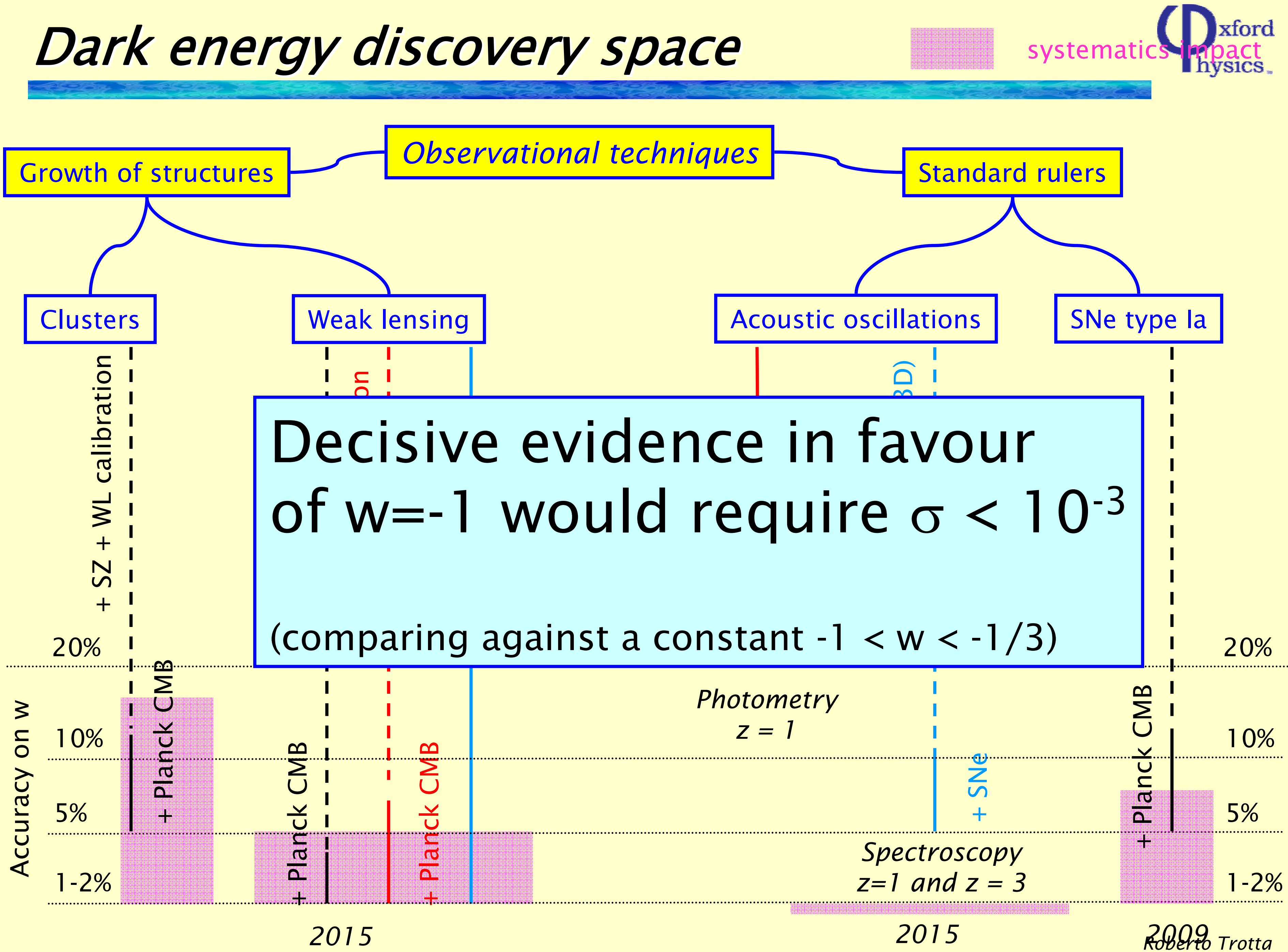
"Trust me, I'm a Bayesian!"

(Trotta 2005, 2006 in prep)





Dark energy discovery space



Decisive evidence in favour of w=-1 would require $\sigma < 10^{-3}$

(comparing against a constant $-1 < w < -1/3$)

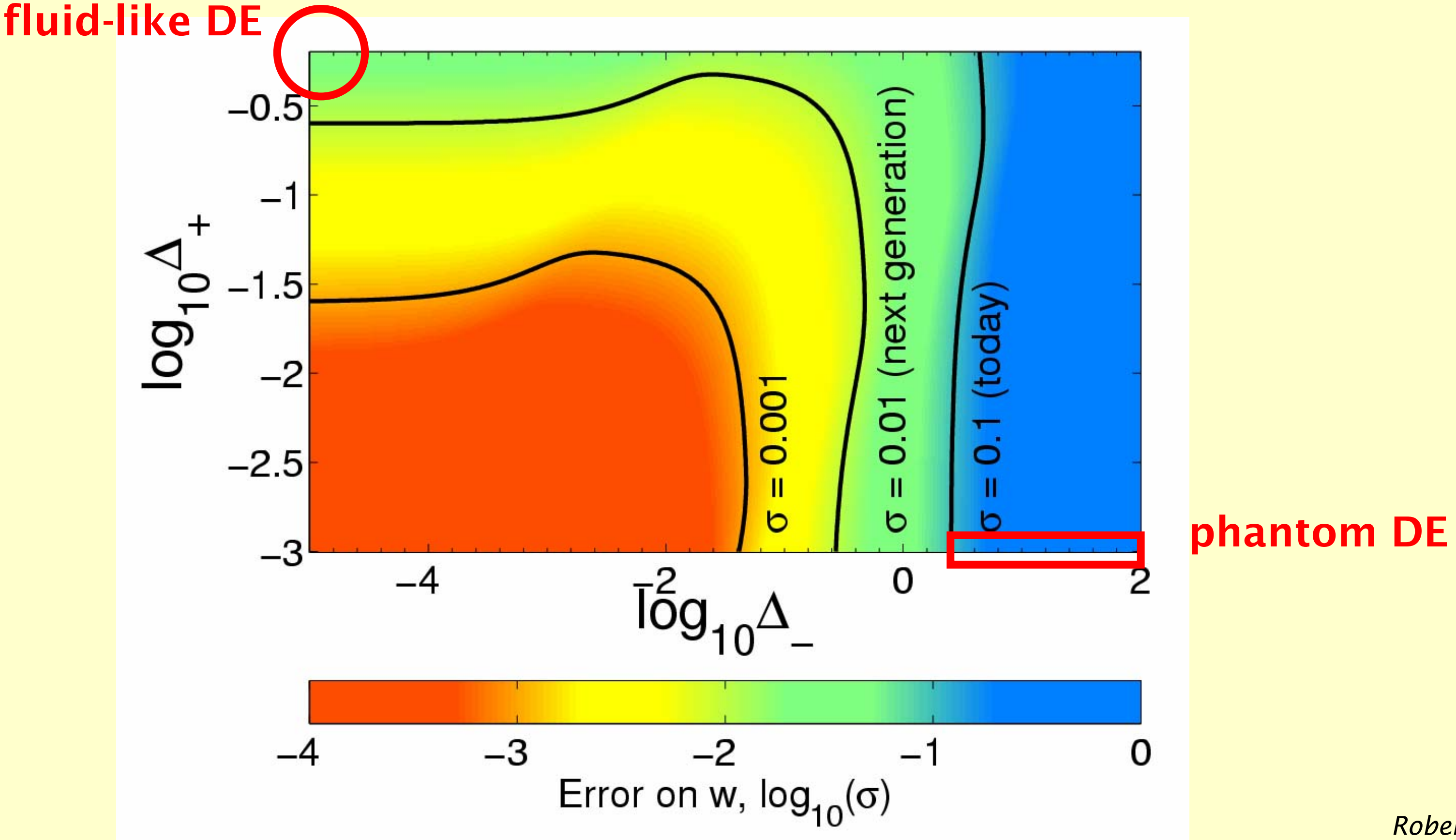
Roberto Trotta

Ruling in Λ

Trotta (2006)

- *Which dark energy models are strongly disfavoured against Λ for a given accuracy σ around $w = -1$?*

$$-1 - \Delta_- \leq w_{\rm eff} \leq -1 + \Delta_+$$



Roberto Trotta

Computing the Bayes factor

$$\mathcal{P}(\mathrm{d}|\mathcal{M}) = \int_{\Omega} \mathrm{L}(\mathrm{d}|\boldsymbol{\theta}, \mathcal{M}) \pi(\boldsymbol{\theta}, \mathcal{M}) \mathrm{d}\boldsymbol{\theta}$$

Multi-dimensional integral for the model likelihood

- *Thermodynamic integration: brute force, computationally intensive*
- *Laplace approximation (possibly + 3rd order corrections): inaccurate for non-Gaussian posteriors*
- *Nested sampling (Skilling, implemented by Mukherjee er al): neat algorithm, more efficient than TDI, needs to be rerun if priors changed*
- *Savage-Dickey density ratio (RT 2005): fast & economical for nested model, clarifies the role of prior*

The Savage-Dickey formula

How can we compute Bayes factors efficiently ?

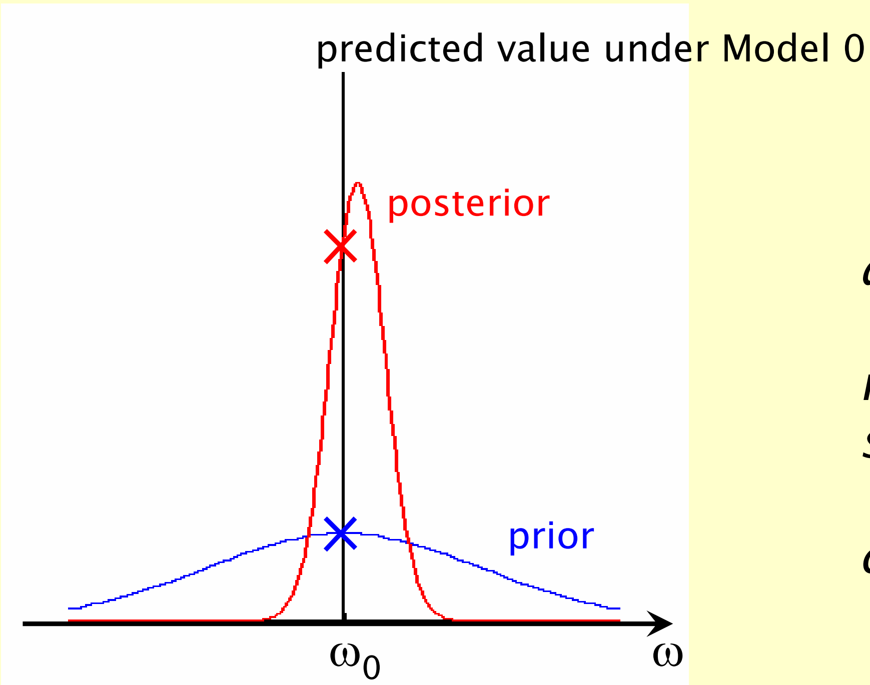
$$\mathcal{P}(\mathbf{d}|\mathcal{M}) = \int_{\Omega} \mathbf{L}(\mathbf{d}|\boldsymbol{\theta}, \mathcal{M}) \pi(\boldsymbol{\theta}, \mathcal{M}) d\boldsymbol{\theta}$$

For nested models and separable priors: use the Savage-Dickey density ratio

Model 1 has one extra param than Model 0

no correlations between priors

$$B_{01} = \frac{\mathcal{P}(\omega_0|\mathbf{d})}{\pi(\omega_0)}$$



- *Economical*
at no extra cost than MCMC
- *Exact*
no approximations (apart from sampling accuracy)
- *Intuitively easy*
clarifies role of prior

Introducing complexity

"For every complex problem, there is a solution that is simple, neat, and wrong"

Oscar Wilde

How many parameters can the data support, regardless of whether they have been measured or not?

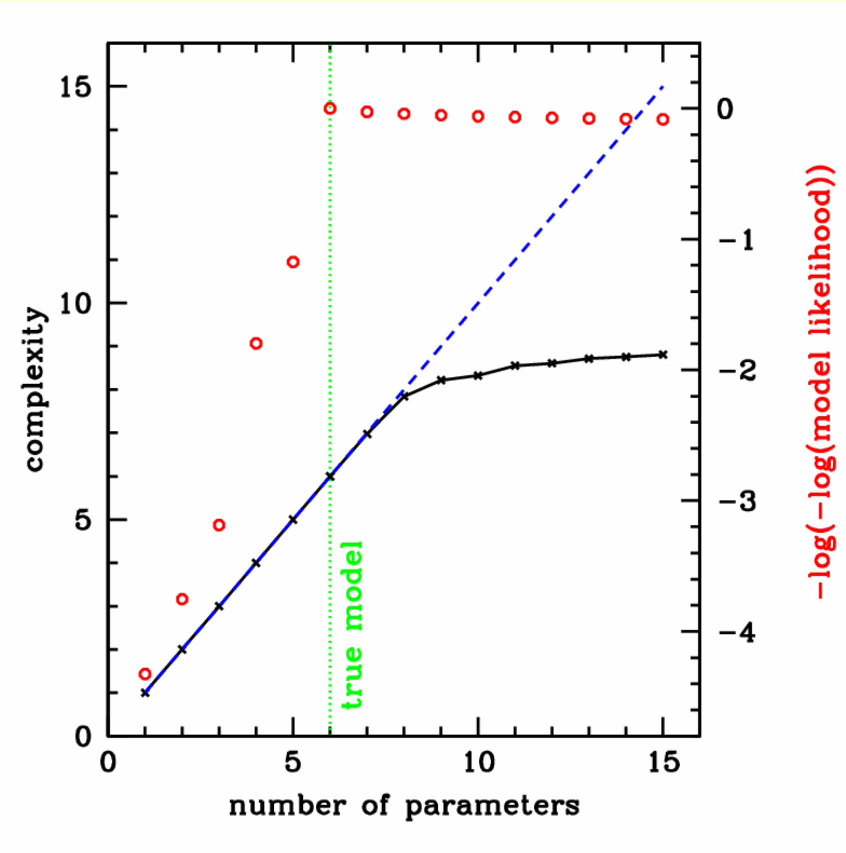
Bayesian complexity

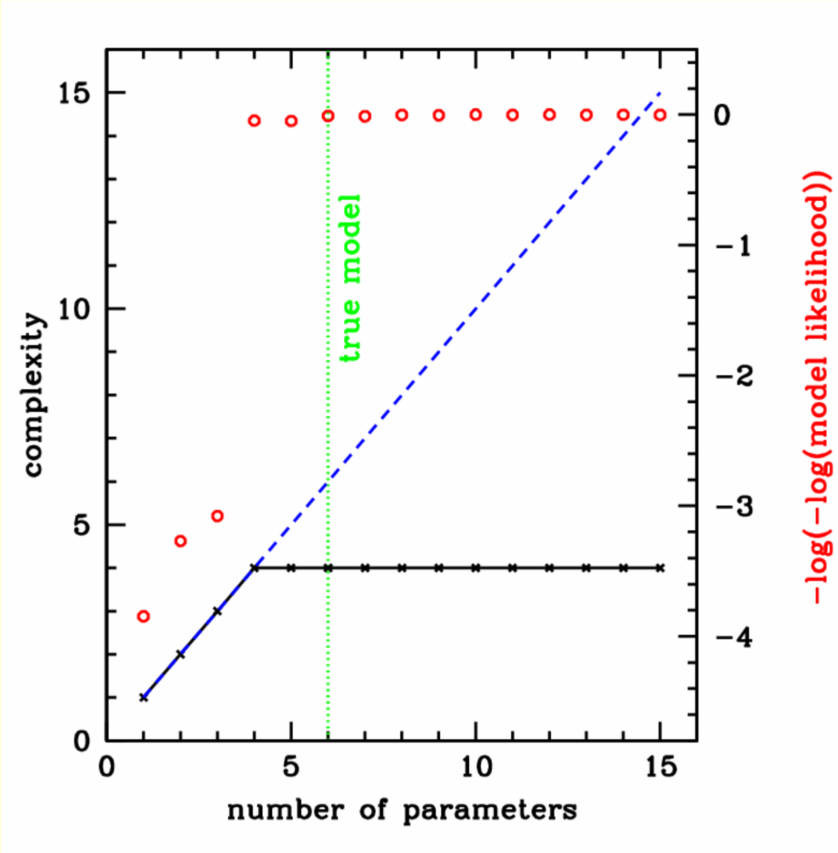
$$\mathcal{C}_b = -2\left(D_{KL}(p,\pi) - \widehat{D_{KL}}(p,\pi)\right)$$
$$= \overline{\chi^2(\theta)} - \chi^2(\hat{\theta})$$

(Kunz, RT & Parkinson, astro-ph/0602378, PRD accepted)

Example: polynomial fitting

Data generated from a model with N = 6

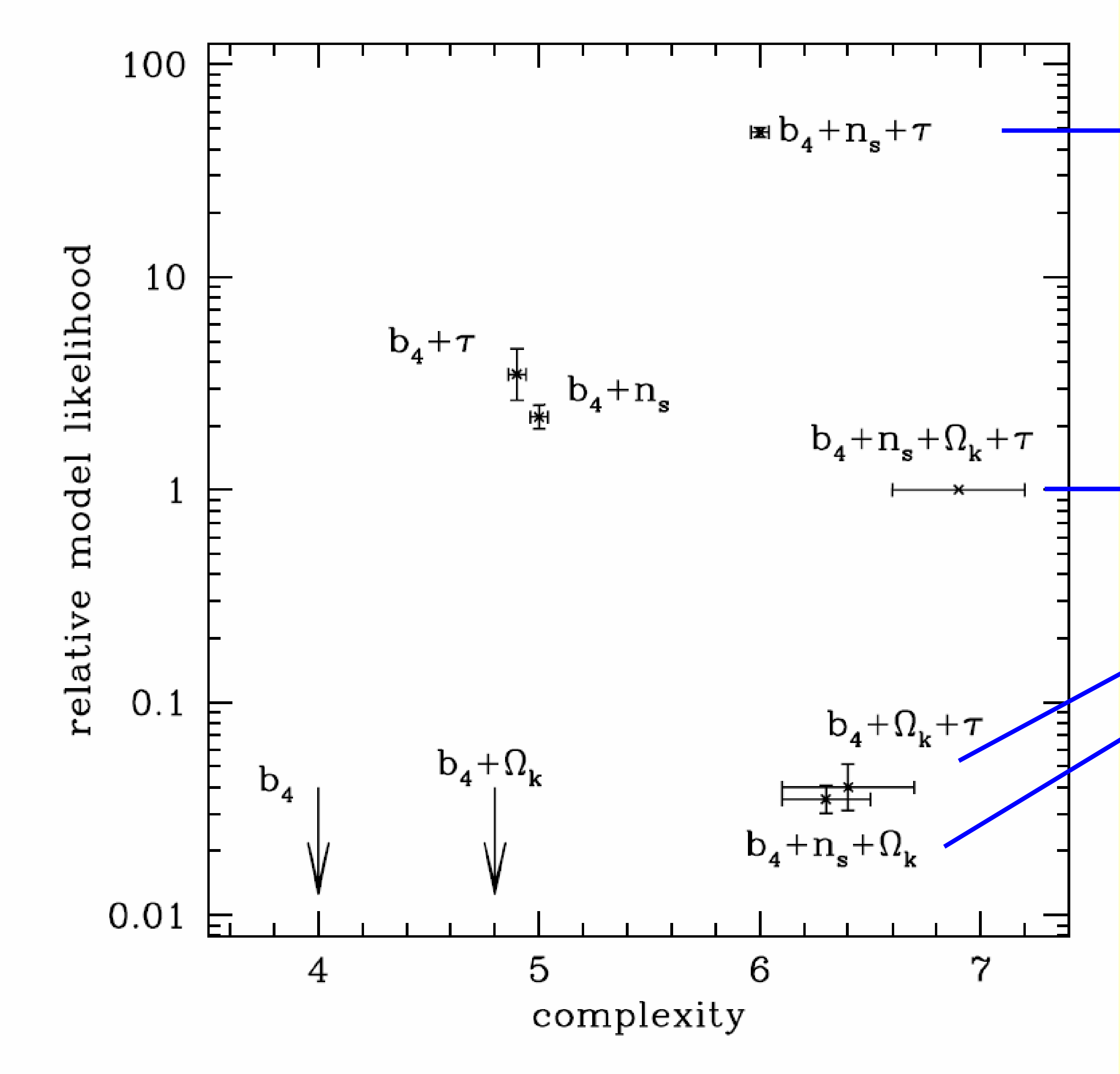




GOOD DATA
Max supported complexity $\approx$ 9

INSUFFICIENT DATA
Max supported complexity $\approx$ 4

How many parameters does the CMB need?



$b_4+n_s+\tau$ measured & favoured

Ω_κ measured & unnecessary

7 params measured only 6 sufficient

(Kunz, RT & Parkinson astro-ph/0602378)

Roberto Trotta

The many uses of model comparison

Bayesian model comparison tools provide a framework for new questions & approaches:

- *Model building: phenomenologically work out how many parameters we need. Needs model insight (prior).*
- *Experiment design: what is the best strategy to discriminate among models?*
- *Performance forecast: how well must we do to reach a certain level of evidence?*
- *Science return optimization: use present-day knowledge to optimize future searches (eg DES, WFMOS, SKA)*

Predictive Posterior Odds Distribution

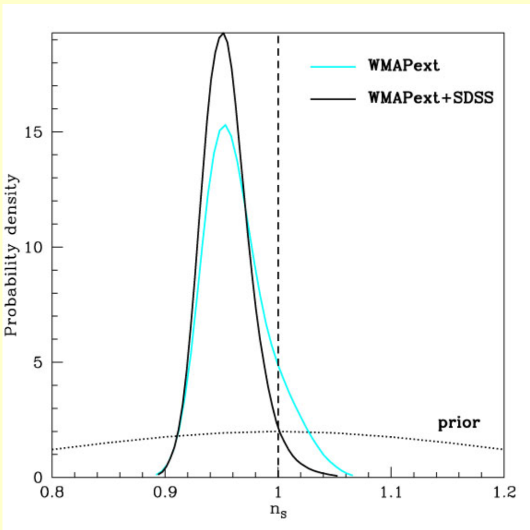
RT (2005)

PPOD: a new hybrid technique

(RT, astro-ph/0504022; see also Pahud et al, Parkinson et al (2006))

- *Gives the probability distribution for the model comparison result of a future measurement*
- *Conditional on our present knowledge*
- *Useful for experiment design & model building:*

Current data posterior



PPOD procedure

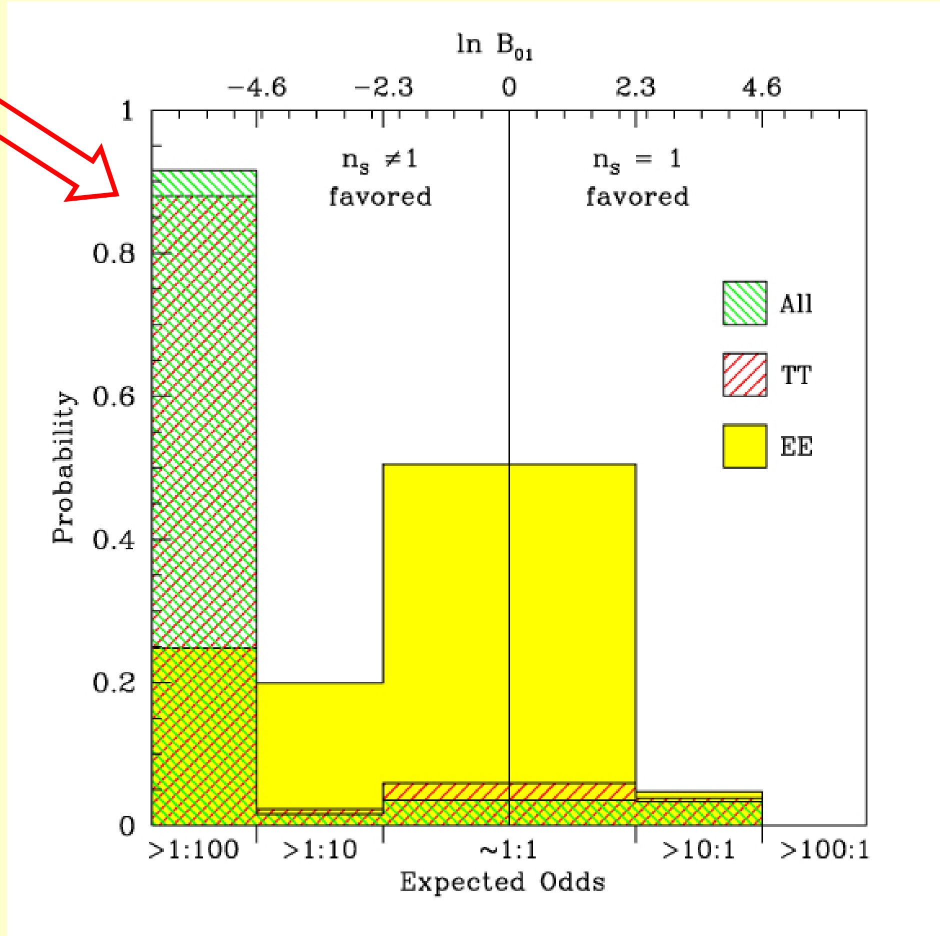
- Start from the posterior PDF from current data
- Fisher Matrix forecast at each sample
- Combine Laplace approximation & Savage-Dickey formula
- Compute Bayes factor probability distribution

Roberto Trotta

PPOD in action

Scale invariant vs $n_S \neq 1$:
PPOD for the Planck satellite (2008)
(Based on WMAP1 + SDDS data)

About 90% probability that Planck will disfavor $n_S = 1$ with odds of 1:100 or higher



Roberto Trotta

Anthropic coincidences?

Are physical constants tuned for life?

(Aguirre 2001, 2005; Weinberg, 2000; Tegmark et al 2005; Rees 1998, ...)

- *Primordial fluctuations amplitude Q*
- α_{EM}/G *and* α_s
- *Cosmological constant* Λ, ...

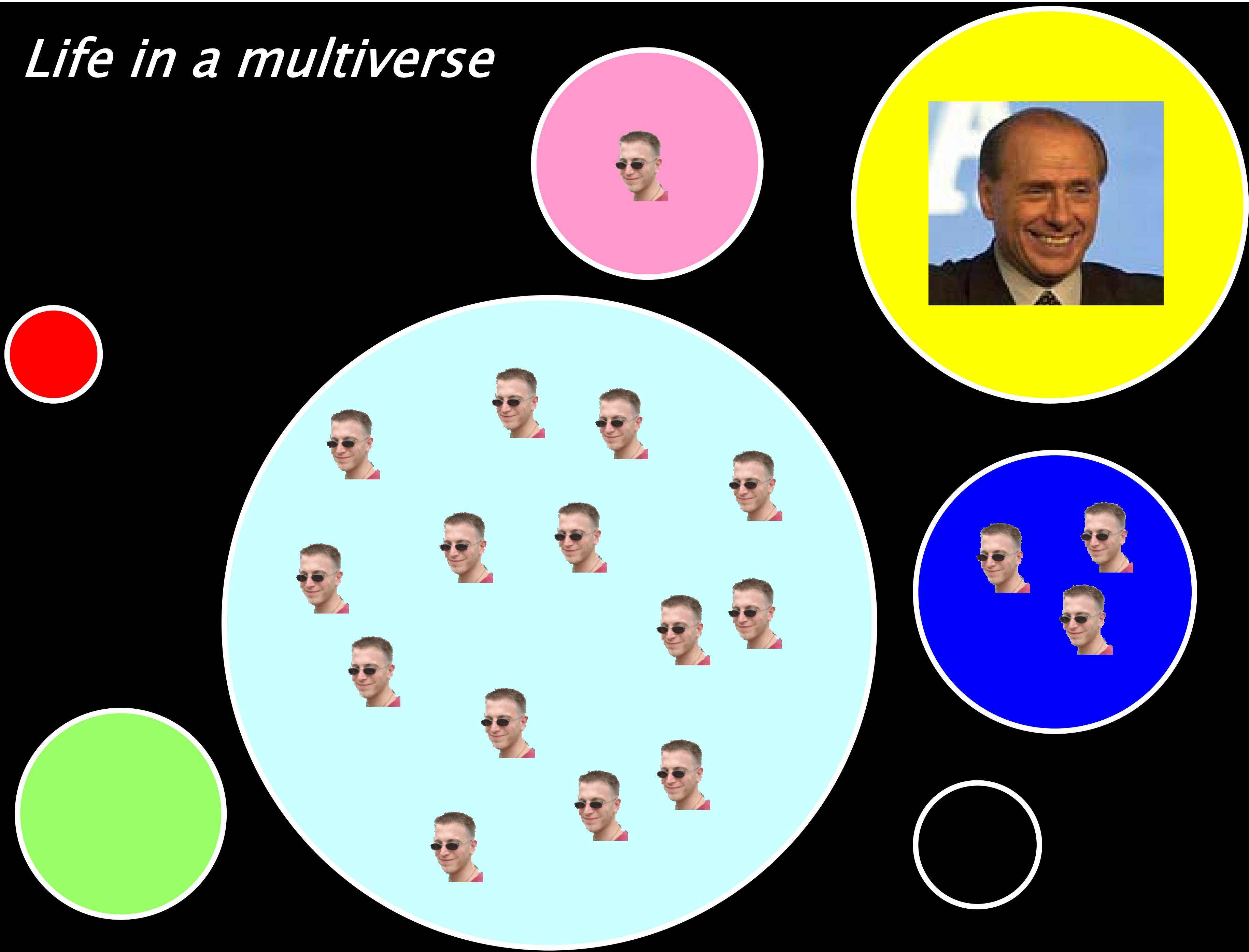
Possible viewpoints:

- *Deeper symmetry / laws of Nature*

  (but what determined THAT particular symmetry in the first place?)
- *Design or necessity*

  (outside the scope of scientific investigation)
- *Any parameters will do (no explanatory power)*
- *Multiverse: we must live in one "realization" favourable for life*

Life in a multiverse



Anthropic reasoning and Λ

The cosmological constant problem:

why is $\Lambda/M_{Pl} \approx 10^{-121}$?

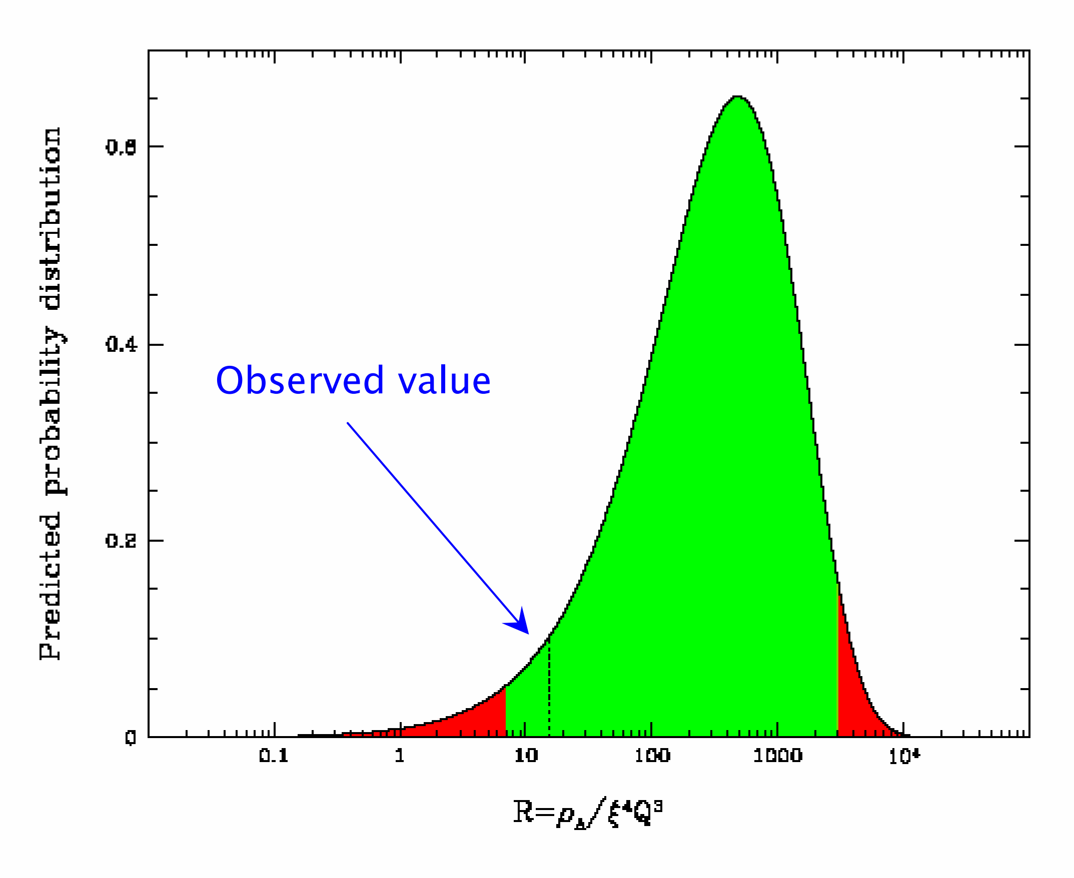
The anthropic “solution”:

if $\Lambda \gg 1$ galaxies cannot form
hence no observers (Weinberg, 1987)

Shortcuts & difficulties:

- What counts as observers?
- Which parameters are allowed to vary?
- Is the multiverse a scientific (ie testable) theory?

Which parameters should we vary?



(Tegmark at al 2005)

if Λ, Q and ξ varied:

$\Lambda = 10^{17}\,\Lambda_0$
perfectly “viable” !

(Aguirre 2001)

“Prediction” only successfull
conditional on ξ, Q = fixed
(AND that T_{CMB} = 2.73 K)

Probability theory and Λ

$$f_{obs}(\Lambda) = f(\Lambda)\, f_{sel}(\Lambda)$$

*prob of observing = sampling distribution * selection function*

"random sample" "typical observer"

The sampling distribution $f(\Lambda)$

As a frequency of outcomes? (untestable in cosmology)

Flat distribution (the "Weinberg conjecture") ? (assumed)

Ergodic arguments? (unclear in an infinite Universe)

No operational def'on of "random" sample: probabilities are NOT physical properties!

The selection function

$$f_{obs}(\Lambda) = f(\Lambda)\, f_{sel}(\Lambda)$$

The selection function $f_{sel}(\Lambda)$

What counts as "observers"? (it's the total number that counts!)

What if the Universe is infinite? (number density/Hubble volume?)

Do observers outside your causal horizon count?

Certainly important to integrate over time: we might not be "typical" in that we are early arrivals...

An explicit counter-example: MANO weighting

Maximum Number of Allowed Observations

MANO weighting of Universes

- *Integrate over lifetime of the Universe to obtain the total number of observations that can POTENTIALLY be carried out*
- *Universes that allow for more observations should weight more*
- *Gauge invariant, time independent quantity*
- *Maximum number of thermodynamic processes in a $\Lambda > 0$ Universe:*

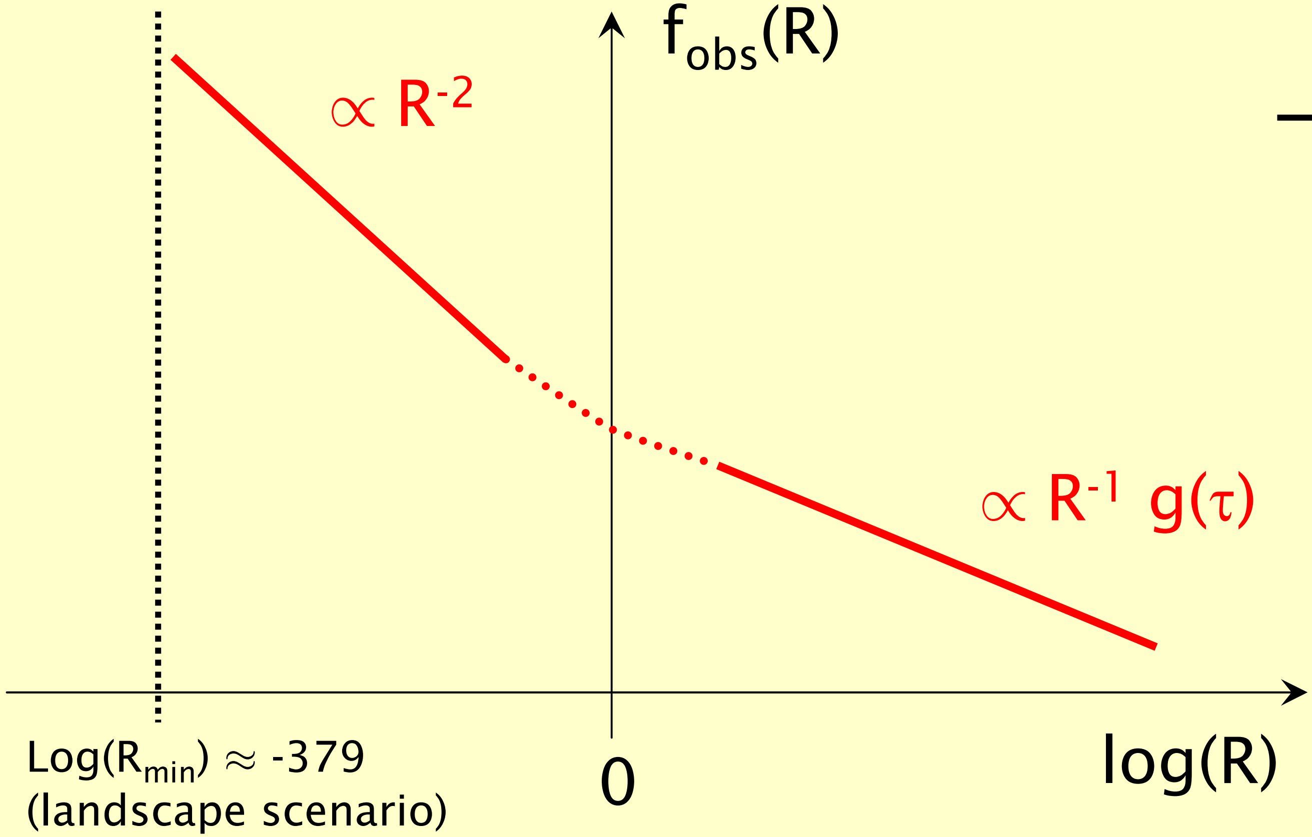
$$N_{max} < E_{coll}/k_B\, T_{ds}$$

- *This assumes "rare observers", otherwise density of observers sets the limit*
- *Still suffers from dependence of micro-physics + details of how civilizations arise & evolve*

Probability of observing Λ

- *2 parameters model:*

$$R = \Omega_\Lambda/\Omega_\Lambda^0 \qquad \tau = t_{obs}/t_0$$



τ	$f_{obs}(R>1)$
.1	8×10^{-3}
1	3×10^{-5}
3	5×10^{-8}
10	2×10^{-16}

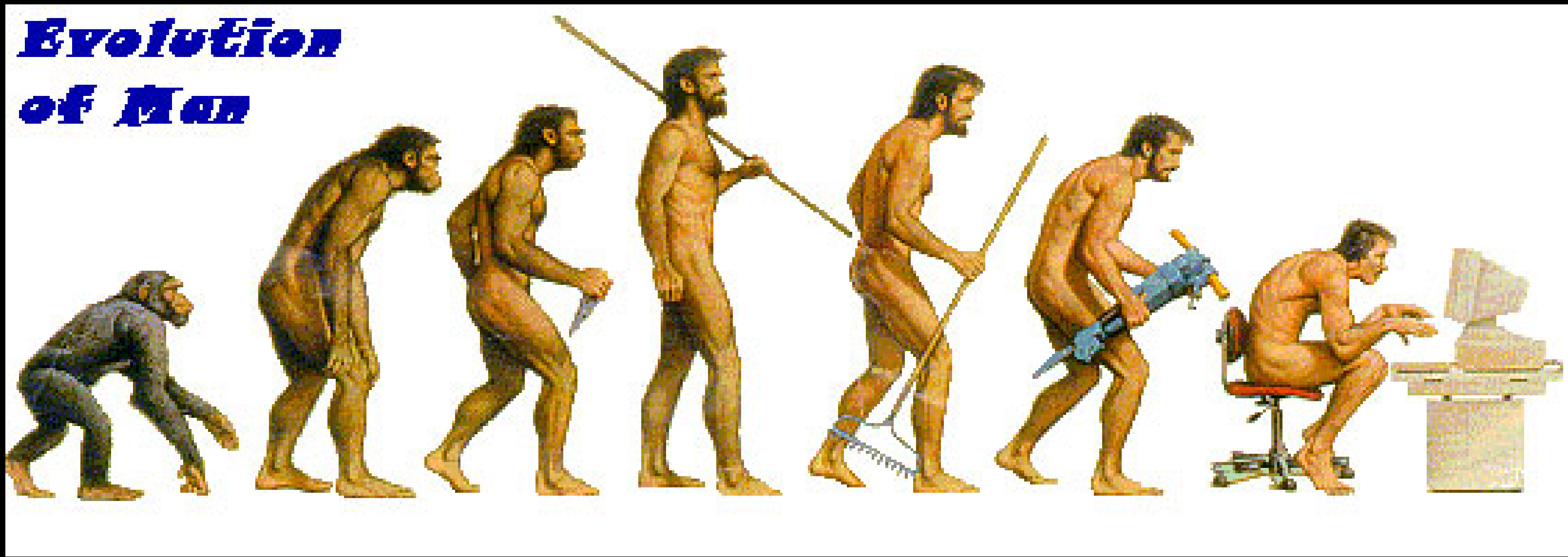
Final remarks

PROBABILITY THEORY AND COSMOLOGY

- *Probabilities are not physical properties but* ***states of knowledge***
- *Uniqueness of the Universe calls for a fully Bayesian approach*

ANTHROPIC REASONING AND SELECTION EFFECTS

- *Outcome depends on selection function*
- *Probability theory as logic at odds with multiverse approach*
- *Within "traditional" anthropic arguments: you should at least integrate over time*
- *MANO counterexample:* $P(\Lambda > 0.7) \sim 10^{-5}$
- *Anthropic "predictions" completely dependent on (many) assumptions*



Homo
a prioris

$$\pi(\theta)_{\pi(\theta)}$$

Homo
frequentistus

$$L(\mathrm{d}|\theta)$$

Homo
Bayesianus

$$\mathcal{P}(\theta|\mathrm{d})$$