What Are The Odds?

An alternative approach to the detection problem

Neil Cornish
Montana State University

LIGO-G080311-00-Z

Outline

Parameter Estimation

- Application to simulated LIGO data

Model Selection

- Bayesian
- Frequentist
- Comparison

Computing the Odds

- Application to simulated LISA data
- Application to simulated LIGO data

Can this be used by LIGO?

Parameter Estimation

Data: $$s(t) = h(t) + n(t)$$

Includes all detectors in the network

Includes time delays and antenna patterns

Given some model M for h, want to compute posterior PDF $p(\vec{\lambda}|s)$ for the parameters $\vec{\lambda}$ that describe $h(\vec{\lambda}, t)$.

Parameter Estimation

Data: $$s(t) = h(t) + n(t)$$

Includes all detectors in the network

Includes time delays and antenna patterns

Given some model M for h, want to compute posterior PDF $p(\vec{\lambda}|s)$ for the parameters $\vec{\lambda}$ that describe $h(\vec{\lambda}, t)$.

$$p(\vec{\lambda}|s) = \frac{p(\vec{\lambda})\; p(s|\vec{\lambda})}{p(s)}$$

Parameter Estimation

Data: $$s(t) = h(t) + n(t)$$

Includes all detectors in the network

Includes time delays and antenna patterns

Given some model M for h, want to compute posterior PDF $p(\vec{\lambda}|s)$ for the parameters $\vec{\lambda}$ that describe $h(\vec{\lambda},t)$.

Prior, Likelihood, Evidence:

$$p(\vec{\lambda}|s) = \frac{p(\vec{\lambda})\, p(s|\vec{\lambda})}{p(s)}$$

Computing the Posterior Distribution

Prior
(informed by theory, EM observations)

$$\text{eg.} \quad p(\theta) = \frac{1}{2}\sin(\theta)$$

Likelihood
(Stationary, Gaussian Noise)

$$p(s|\vec{\lambda}) = \prod_{k=1}^{N} \frac{1}{2\pi\sigma_k^2} \exp\left(-\frac{|\tilde{s}_k - \tilde{h}_k(\vec{\lambda})|^2}{2\sigma_k^2}\right)$$

$$= C \exp\left(-\frac{(s-h|s-h)}{2}\right)$$

Evidence
(expensive to compute for large dimension models)

$$p(s) = \int d\vec{\lambda}\, p(\vec{\lambda})\, p(s|\vec{\lambda})$$

Bayesian Learning

$$\text{Posterior Belief} \propto (\text{Prior Belief}) \times (\text{Likelihood of data})$$

Bits of Information Obtained From Data

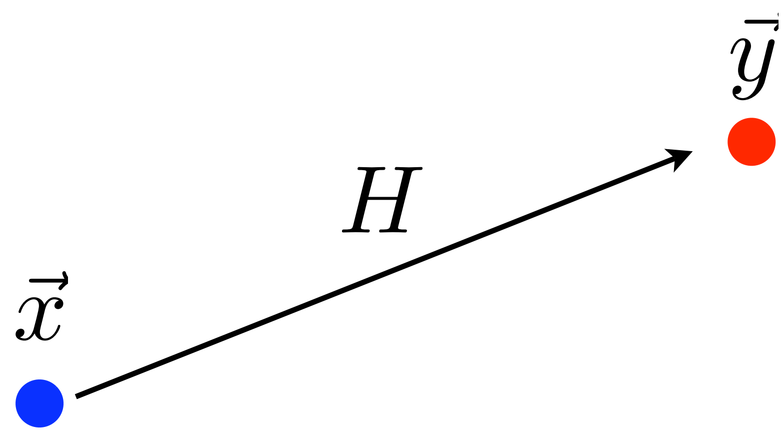
$$I = \int d\vec{\lambda}\, p(\vec{\lambda}|s)\, \log_2 \left(\frac{p(\vec{\lambda}|s)}{p(\vec{\lambda})} \right)$$

Example:

$$p(x) = U[\frac{\Delta x}{2}, \frac{\Delta x}{2}] \qquad p(x|s) = C \exp\left(-\frac{x^2}{2\delta x^2} \right)$$

$$\Rightarrow \quad I \simeq \log_2 \left(\frac{\Delta x}{\delta x} \right) - \log_2(\sqrt{2\pi e}) \qquad (\delta x \ll \Delta x)$$

Markov Chain Monte Carlo



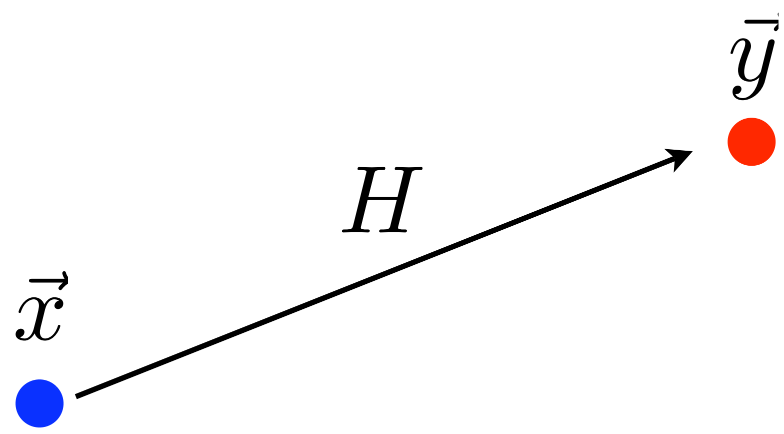
Yields $p(\vec{\lambda}|s)$ for parameters $\vec{\lambda}$ given data s for any non-trivial q

Avoids the need to compute the Evidence

$$H = \min\left(1, \frac{p(\vec{y})p(s|\vec{y})q(\vec{x}|\vec{y})}{p(\vec{x})p(s|\vec{x})q(\vec{y}|\vec{x})}\right)$$

Transition Probability
(Metropolis-Hastings)

Markov Chain Monte Carlo



Yields $p(\vec{\lambda}|s)$ for parameters $\vec{\lambda}$ given data s for any non-trivial q

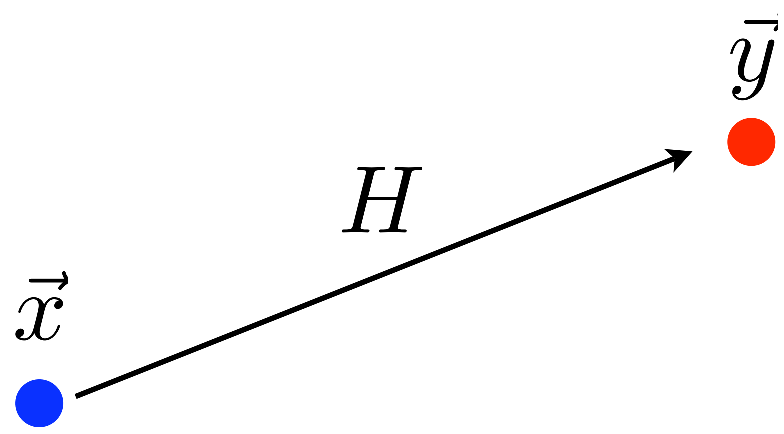
Avoids the need to compute the Evidence

$$H = \min\left(1, \frac{p(\vec{y})p(s|\vec{y})q(\vec{x}|\vec{y})}{p(\vec{x})p(s|\vec{x})q(\vec{y}|\vec{x})}\right)$$

Prior

Transition Probability
(Metropolis-Hastings)

Markov Chain Monte Carlo



Yields $p(\vec{\lambda}|s)$ for parameters $\vec{\lambda}$ given data s for any non-trivial q

Avoids the need to compute the Evidence

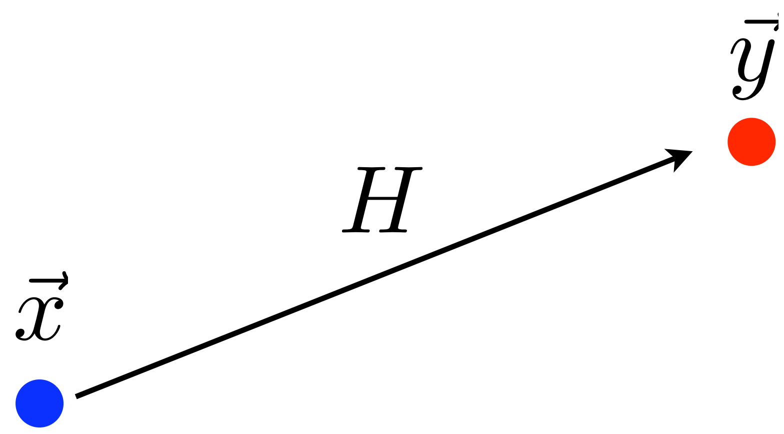
$$H = \min\left(1, \frac{p(\vec{y})p(s|\vec{y})q(\vec{x}|\vec{y})}{p(\vec{x})p(s|\vec{x})q(\vec{y}|\vec{x})}\right)$$

Transition Probability
(Metropolis-Hastings)

Prior

Likelihood $\sim e^{-\chi^2/2}$

Markov Chain Monte Carlo



Yields $p(\vec{\lambda}|s)$ for parameters $\vec{\lambda}$ given data s for any non-trivial q

Avoids the need to compute the Evidence

$$H = \min\left(1, \frac{p(\vec{y})\,p(s|\vec{y})\,q(\vec{x}|\vec{y})}{p(\vec{x})\,p(s|\vec{x})\,q(\vec{y}|\vec{x})}\right)$$

Prior

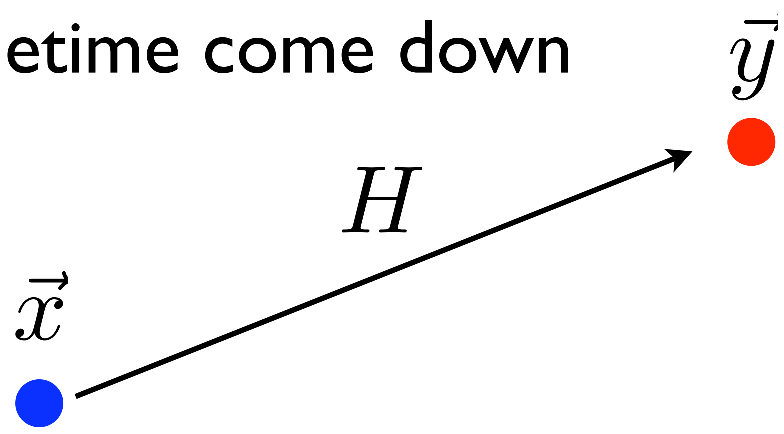
Likelihood $\sim e^{-\chi^2/2}$

Proposal

Transition Probability
(Metropolis-Hastings)

Markov Chain Monte Carlo

Always go up,
Sometime come down



Yields $p(\vec{\lambda}|s)$ for parameters $\vec{\lambda}$ given data s for any non-trivial q

Avoids the need to compute the Evidence

$$H = \min\left(1, \frac{p(\vec{y})\,p(s|\vec{y})\,q(\vec{x}|\vec{y})}{p(\vec{x})\,p(s|\vec{x})\,q(\vec{y}|\vec{x})}\right)$$

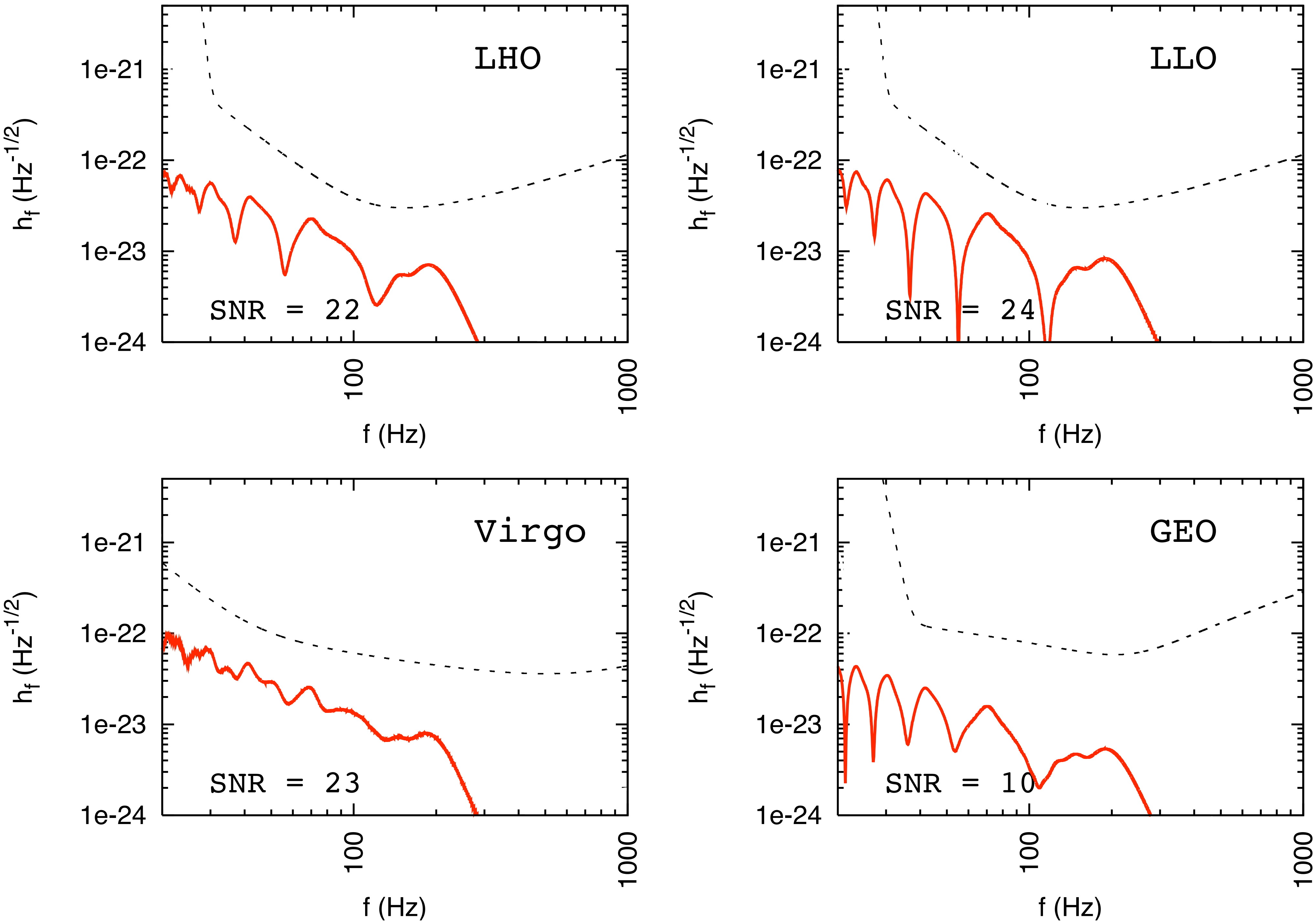
Transition Probability
(Metropolis-Hastings)

Prior

Proposal

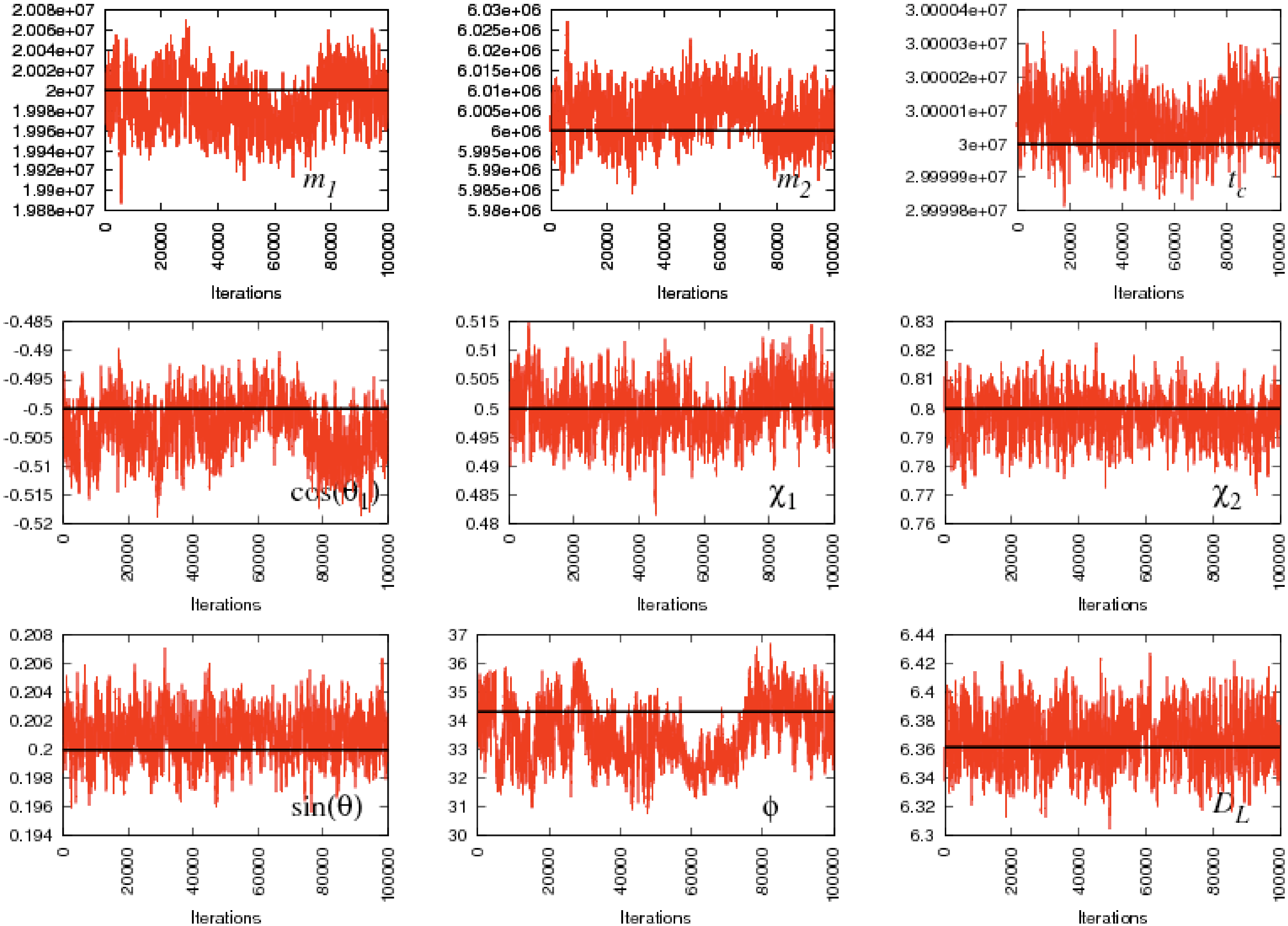
Likelihood $\sim e^{-\chi^2/2}$

Example: Spinning MBH, LVG Network



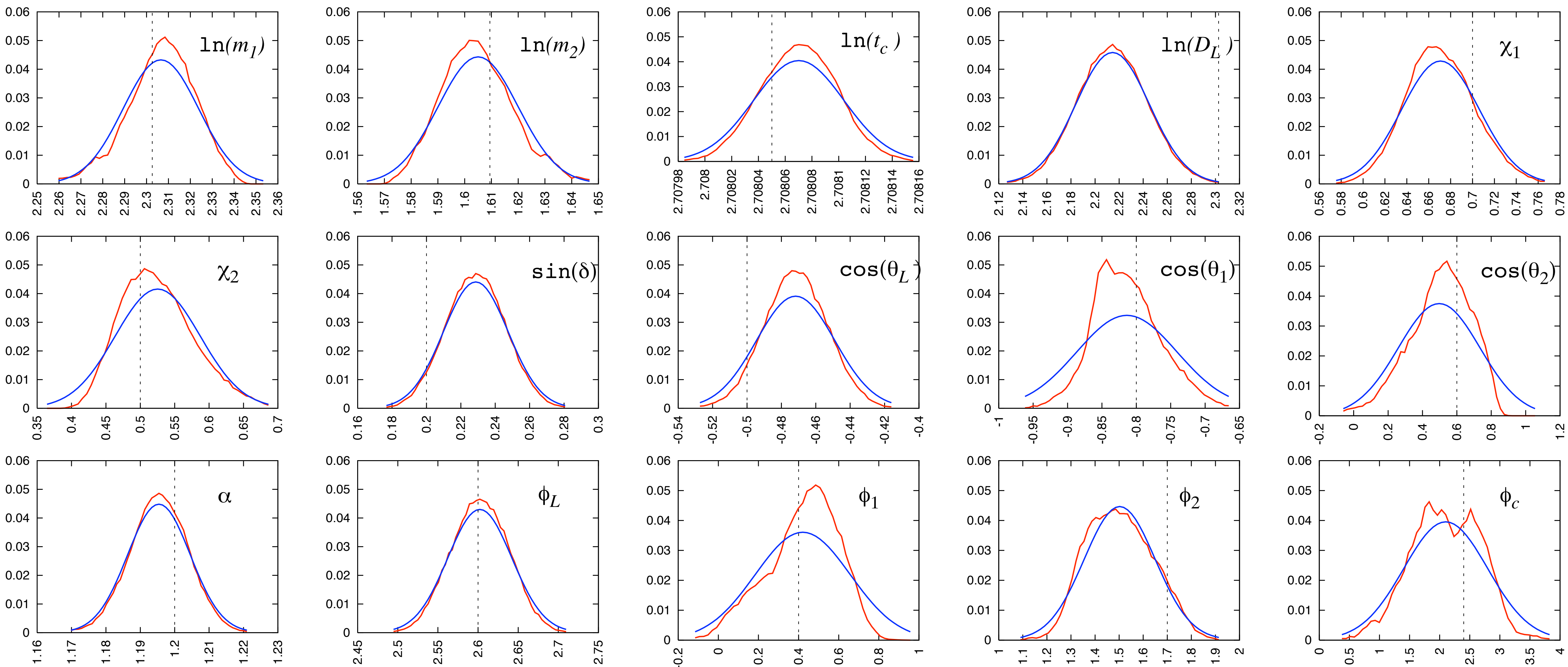
$$m_1 = 10M_\odot \quad m_2 = 5M_\odot \quad s_1/m_1^2 = 0.7 \quad s_2/m_2^2 = 0.5 \quad D_L = 10 \text{ Mpc}$$

Example: Spinning MBH, LVG Network



Marginalized PDFs

(compared to Fisher Matrix Estimates)



(Cornish, Hughes, Lang & Nissanke, 2008)

15 signal parameters, 4 x 26 =104 noise parameters

Bayesian Model Selection

Probability of Model M: $p(M|s) \propto p(M)\, p(s|M)$

Odds Ratio:

$$O_{ij} = \frac{p(M_i|s)}{p(M_j|s)}$$

$$= \frac{p(M_i)}{p(M_j)} \frac{p(s|M_i)}{p(s|M_j)}$$

$$= \text{Prior Odds} \times \text{Bayes Factor}$$

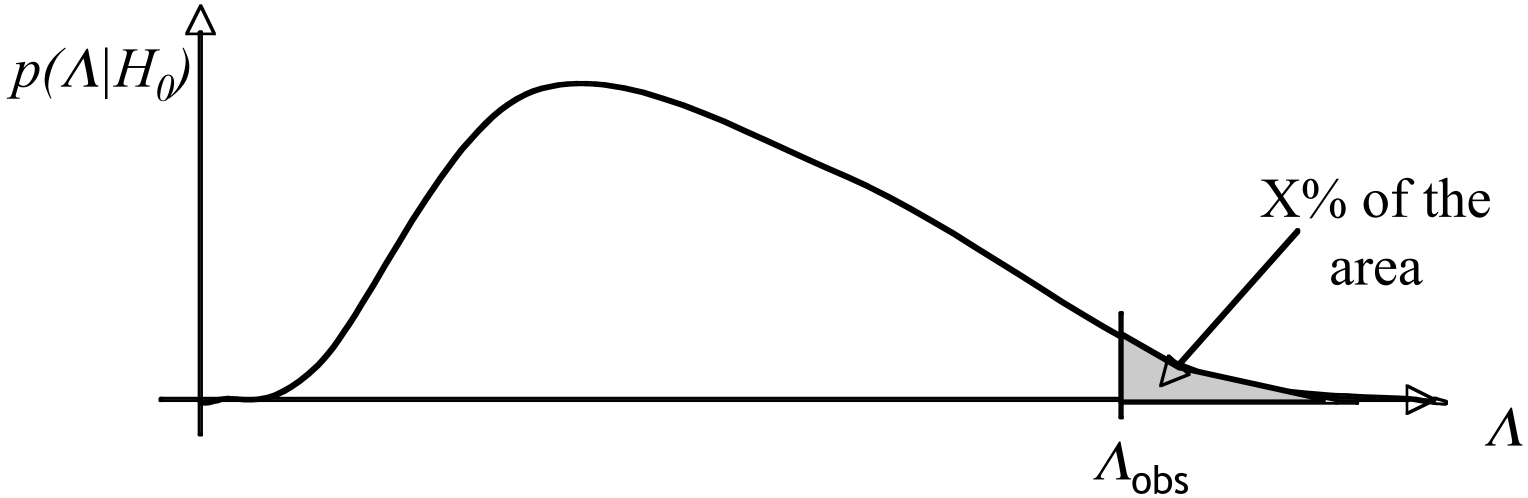
How do we compute the Bayes Factor (Evidence Ratio)?

Frequentist Model Selection

To test a hypothesis H_1 consider another hypothesis, called the *null hypothesis*, H_0, the truth of which would deny H_1. Then argue *against* H_0…

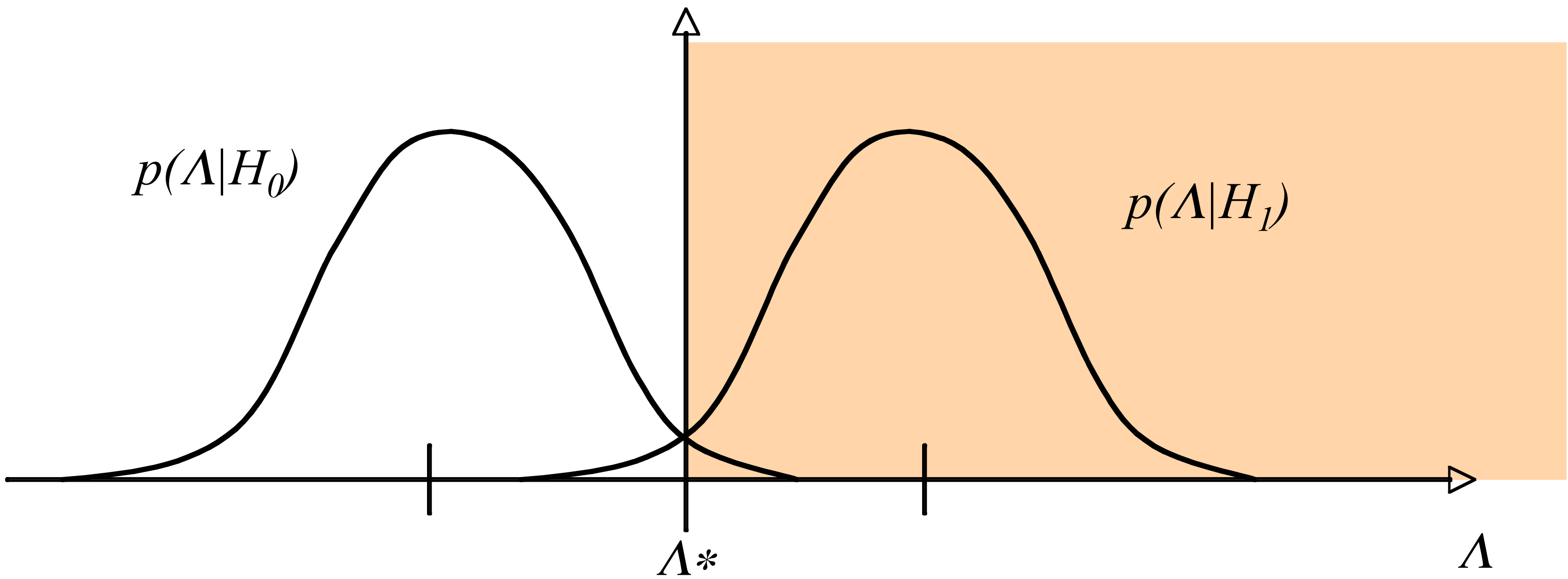
Use the data you have gathered to compute a test statistic $\Lambda_{\rm obs}$ which has a calculable pdf if H_0 is true. This can be calculated analytically or by Monte Carlo methods.

Look where your observed value of the statistic lies in the pdf, and reject H_0 based on how far in the wings of the distribution you have fallen (but make no statement about how unlikely under any other scenario, including H_1).

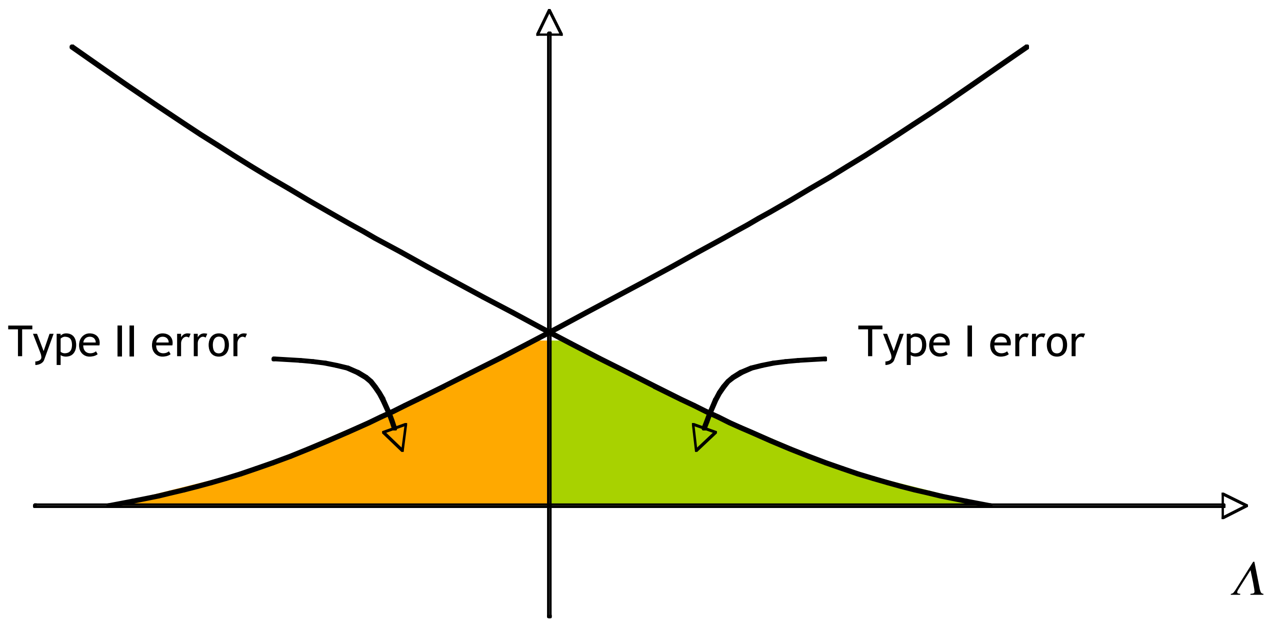


(taken from lecture notes by Alicia Sintes)

Frequentist Model Selection



Set threshold Λ_* such that $\Lambda > \Lambda_*$ favors hypothesis H_1



Type I error - False Alarm

Type II error - False Dismissal

Neyman-Pearson

For fixed false alarm rate, the false dismissal rate is minimized by the likelihood ratio statistic

$$\Lambda(\vec{\lambda}) = \frac{p(s|h, M_1)}{p(s|0, M_0)} = e^{-(s|h)+\frac{1}{2}(h|h)}$$

This quantity is maximized over the signal parameters

c.f. Bayesian alternative where the evidence is marginalized (integrated) over the signal parameters

Standard Detection Procedure

1. Set a threshold on the (network) search statistic using time slides of the data to give acceptable false alarm rate.
2. For each candidate detection look in more detail at the monitoring channels to see if anything might have been missed by the vetos. (R. Gouaty, arXiv:0805.2412)

Standard Detection Procedure

1. Set a threshold on the (network) search statistic using time slides of the data to give acceptable false alarm rate.
2. For each candidate detection look in more detail at the monitoring channels to see if anything might have been missed by the vetos. (R. Gouaty, arXiv:0805.2412)
3. Put champagne in fridge.

Comparison of Approaches

- Bayesian approach has explicit priors - forces assumptions into the open
- Frequentist approach has implicit priors - can be unphysical (Searle, Sutton, Tinto & Woan, **arXiv:0712.0196** [gr-qc])
- Bayesian approach is mathematically rigorous
- Frequentist approach has funny rules about playgrounds and boxes. What do you do if you find a bug after the box is opened - throw away all the data?

The dead astronomer paradox

A theorist calculates that 10% of nearby stars are G-type

The dead astronomer paradox

A theorist calculates that 10% of nearby stars are G-type

An astronomer sets out to tests this claim, and observes that 5 stars out of a sample of 102 are G-type

The dead astronomer paradox

A theorist calculates that 10% of nearby stars are G-type

An astronomer sets out to tests this claim, and observes that 5 stars out of a sample of 102 are G-type

The astronomer dies later that night

The dead astronomer paradox

A theorist calculates that 10% of nearby stars are G-type

An astronomer sets out to tests this claim, and observes that 5 stars out of a sample of 102 are G-type

The astronomer dies later that night

Two of the astronomer's students decide to analyze the data and publish a memorial paper

The dead astronomer paradox

A theorist calculates that 10% of nearby stars are G-type

An astronomer sets out to tests this claim, and observes that 5 stars out of a sample of 102 are G-type

The astronomer dies later that night

Two of the astronomer's students decide to analyze the data and publish a memorial paper

One finds a P-value of 4.3% and rejects the hypothesis at 95% confidence. The other finds a P-value of 10% and is unable to rule out the hypothesis. Both calculations were found to be free of mathematical errors.

The dead astronomer paradox

The P-value of 4.3% assumes that the dead astronomer planned to observe until 5 G-stars were found, so the total number of stars is the data, d.

The P-value of 10% assumes that the dead astronomer planned to observe a total of 102 stars, so the number of G-type stars is the data, d.

The dead astronomer paradox

The P-value of 4.3% assumes that the dead astronomer planned to observe until 5 G-stars were found, so the total number of stars is the data, d.

The P-value of 10% assumes that the dead astronomer planned to observe a total of 102 stars, so the number of G-type stars is the data, d.

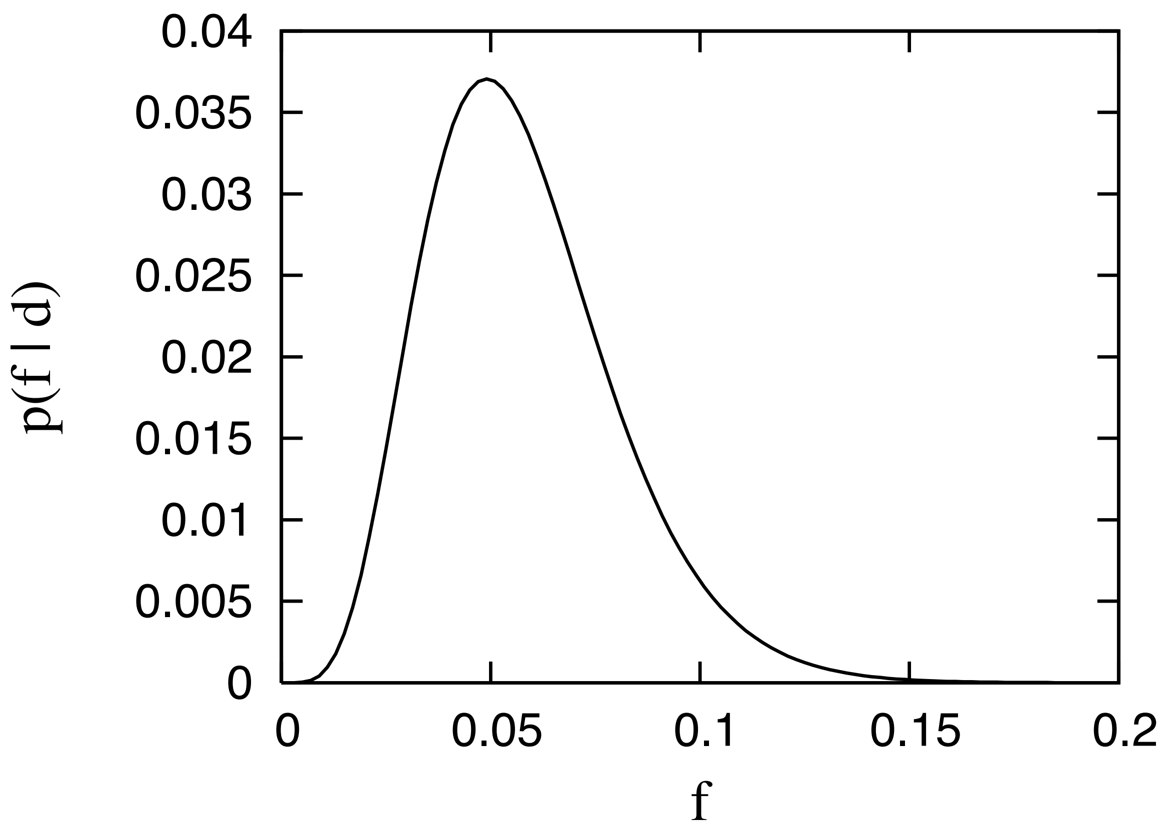
A Bayesian analysis returns a PDF for the fraction, f, of G-type stars that is *independent* of the dead astronomer's intentions:

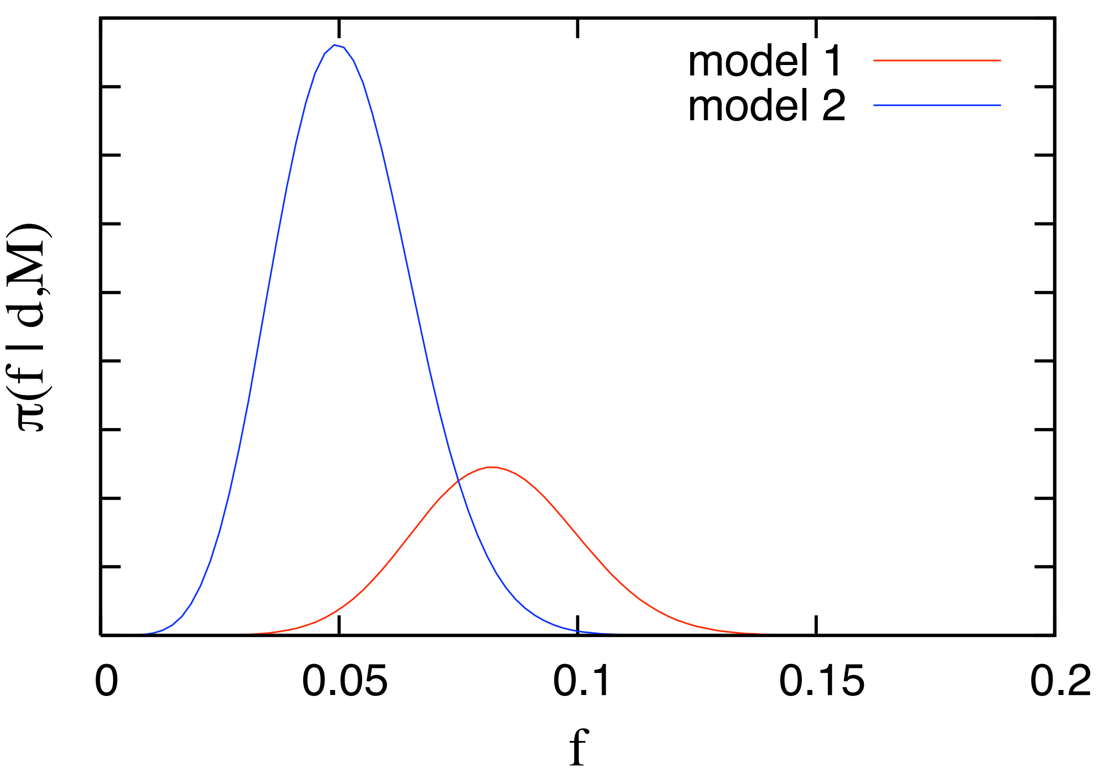
$$\pi(f|d,M) = p(f|d,M)f^n(1-f)^{N-n}$$

(un-normalized posterior)

The dead astronomer paradox

Uniform prior on f





Model selection:

model 1 $f = 0.1 \pm 0.02$

model 2 $f = 0.05 \pm 0.02$

$O_{21} = 3.8$ Data inconclusive

Bayesian Model Selection: Computing the Evidence

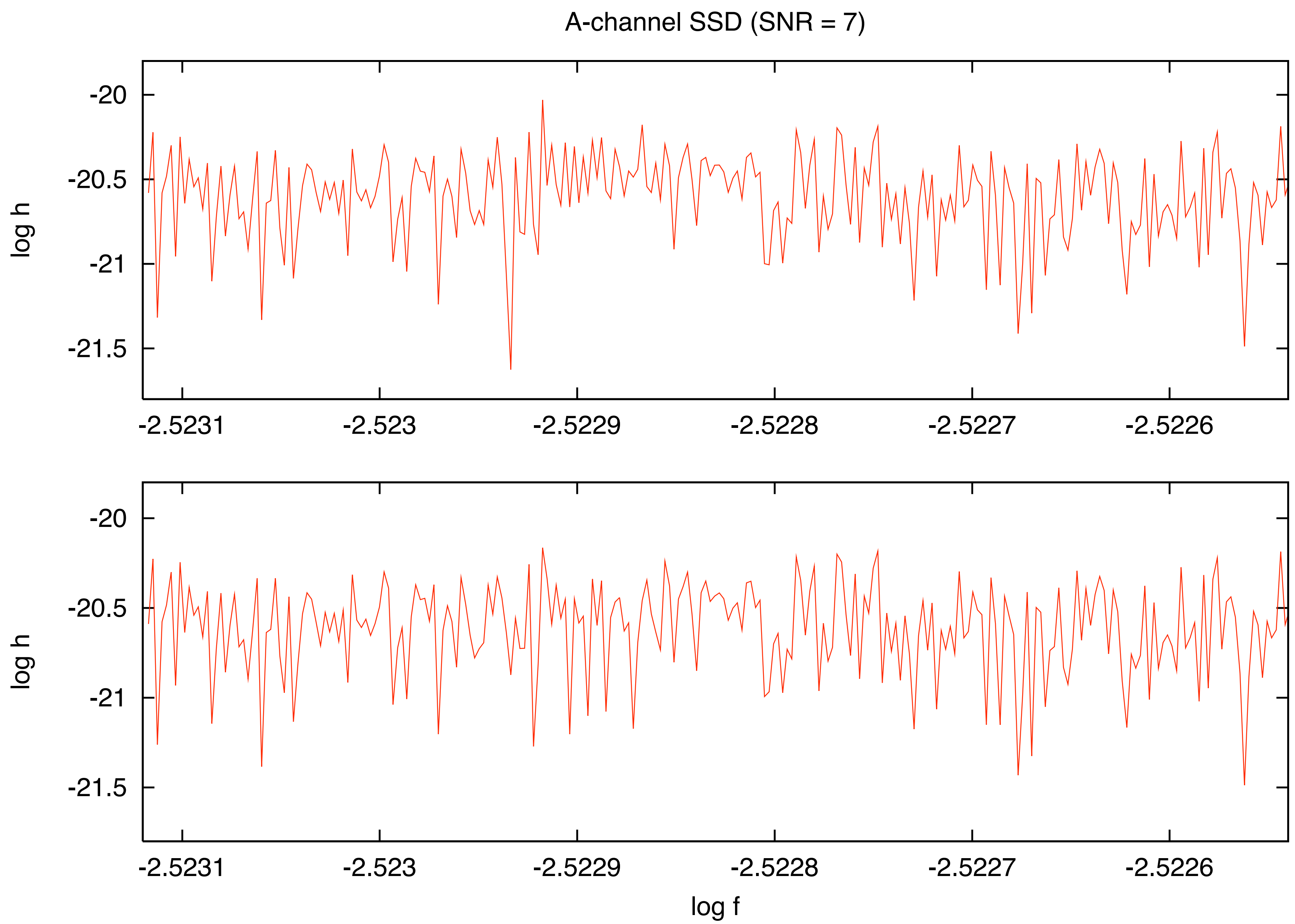
$$p(s|M) = \int d\vec{\lambda}\, p(\vec{\lambda}|M)\, p(s|\vec{\lambda}, M)$$

Expensive to compute for large dimension models

- Brute force grid or Monte Carlo integration
  (e.g. NS f-mode ringdown search, Clark, Heng, Pitkin & Woan, arXiv:0711.4039 [gr-qc])
- Reverse Jump Markov Chain Monte Carlo
- Parallel Tempering + Thermodynamic integration
- Vegas integration algorithm
- Nested Sampling

Model Selection Example: WD-WD binary in simulated LISA

$$\vec{\lambda}_1 \rightarrow \{A, f, \dot{f}, \theta, \phi, \psi, \iota, \varphi_0, \vec{\lambda}_0\} \qquad \vec{\lambda}_0 \rightarrow \{S_A^1, S_A^2, S_A^3, S_A^4, S_E^1, S_E^2, S_E^3, S_E^4\}$$

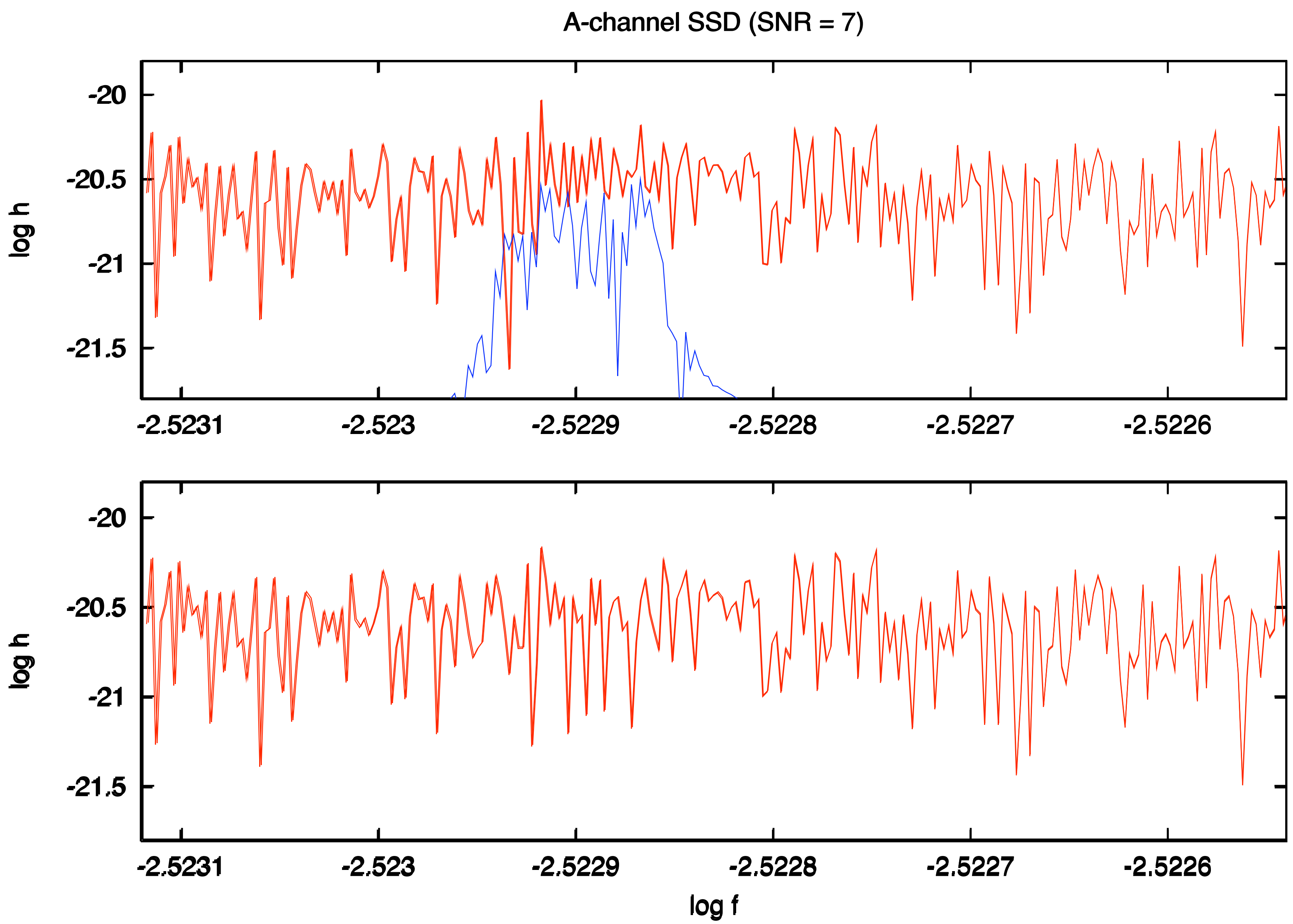


Is there a signal present?

Tyson Littenberg & NJC (2008)

Model Selection Example: WD-WD binary in simulated LISA

$$\vec{\lambda}_1 \rightarrow \{A, f, \dot{f}, \theta, \phi, \psi, \iota, \varphi_0, \vec{\lambda}_0\} \qquad \vec{\lambda}_0 \rightarrow \{S_A^1, S_A^2, S_A^3, S_A^4, S_E^1, S_E^2, S_E^3, S_E^4\}$$



Is there a signal present?

Tyson Littenberg & NJC (2008)

Reverse Jump Markov Chain Monte Carlo (Green, 1995)

$$p(\vec{\lambda}_M, M|s)$$

Make the model one of the parameters and allow transitions between models as well as between parameters

Propose a transition by drawing a random vector $\vec{u}$, then use a deterministic dimension matching function:

$$\vec{\lambda}_{M'} = f(\vec{\lambda}_M, \vec{u})$$

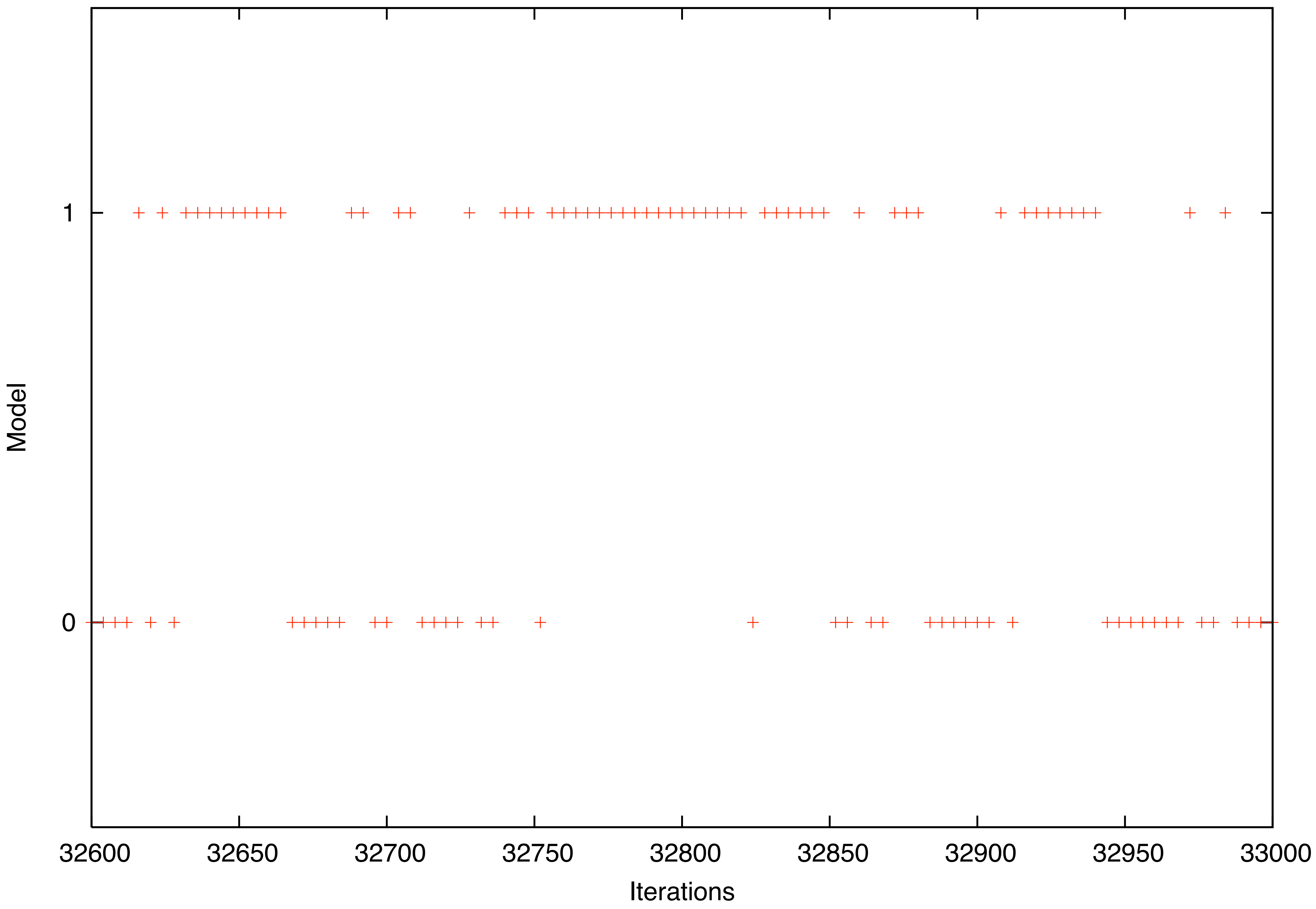
where $\dim(\vec{\lambda}_{M'}) = \dim(\vec{\lambda}_M) + \dim(\vec{u})$

Transition Probability

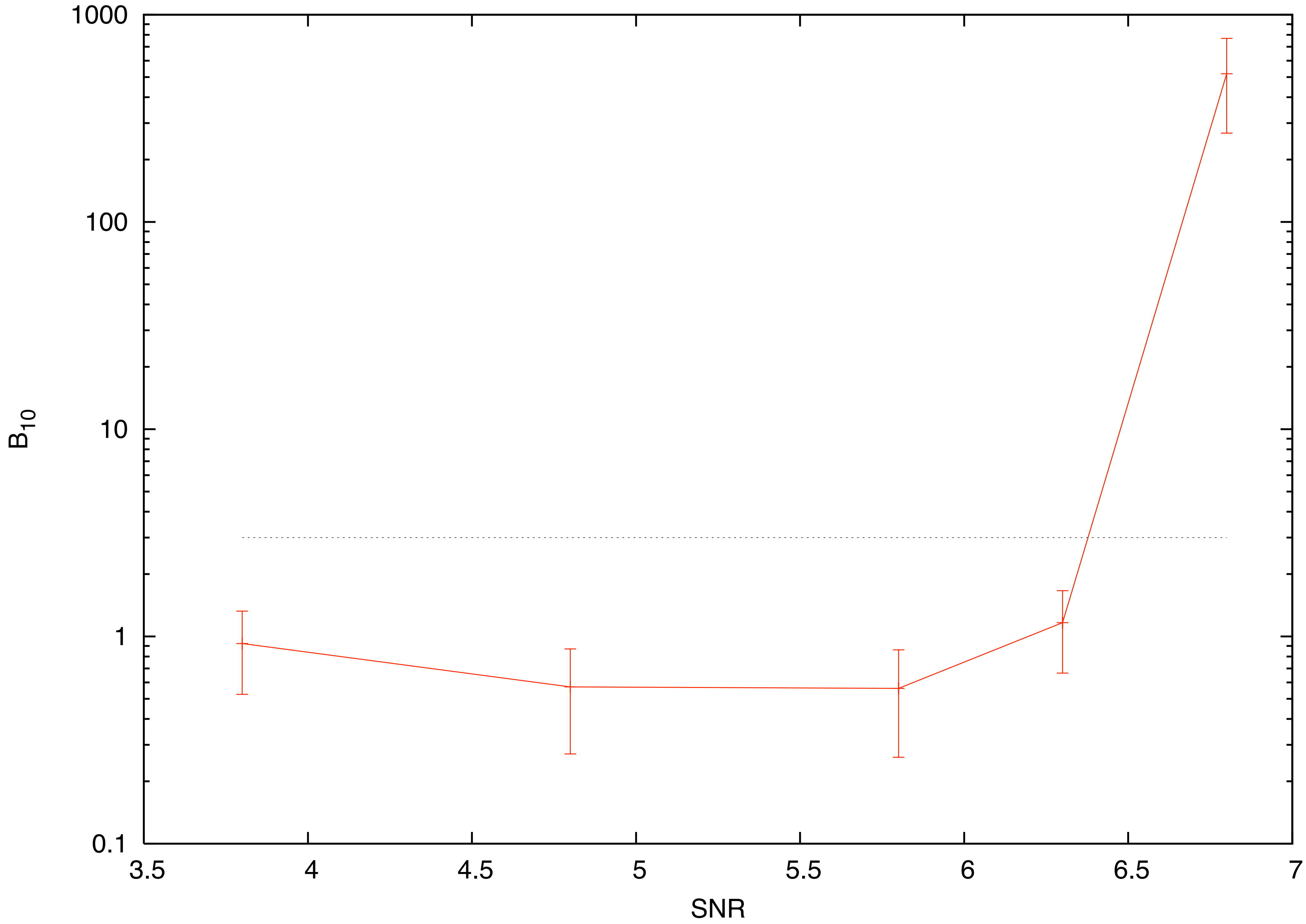
$$H = \min\left(1, \frac{\pi(s|\vec{\lambda}_{M'}, M')q(\vec{u}')}{\pi(s|\vec{\lambda}_M, M)q(\vec{u})}|J|\right)$$

Bayes factor given by ratio of time spent in each model

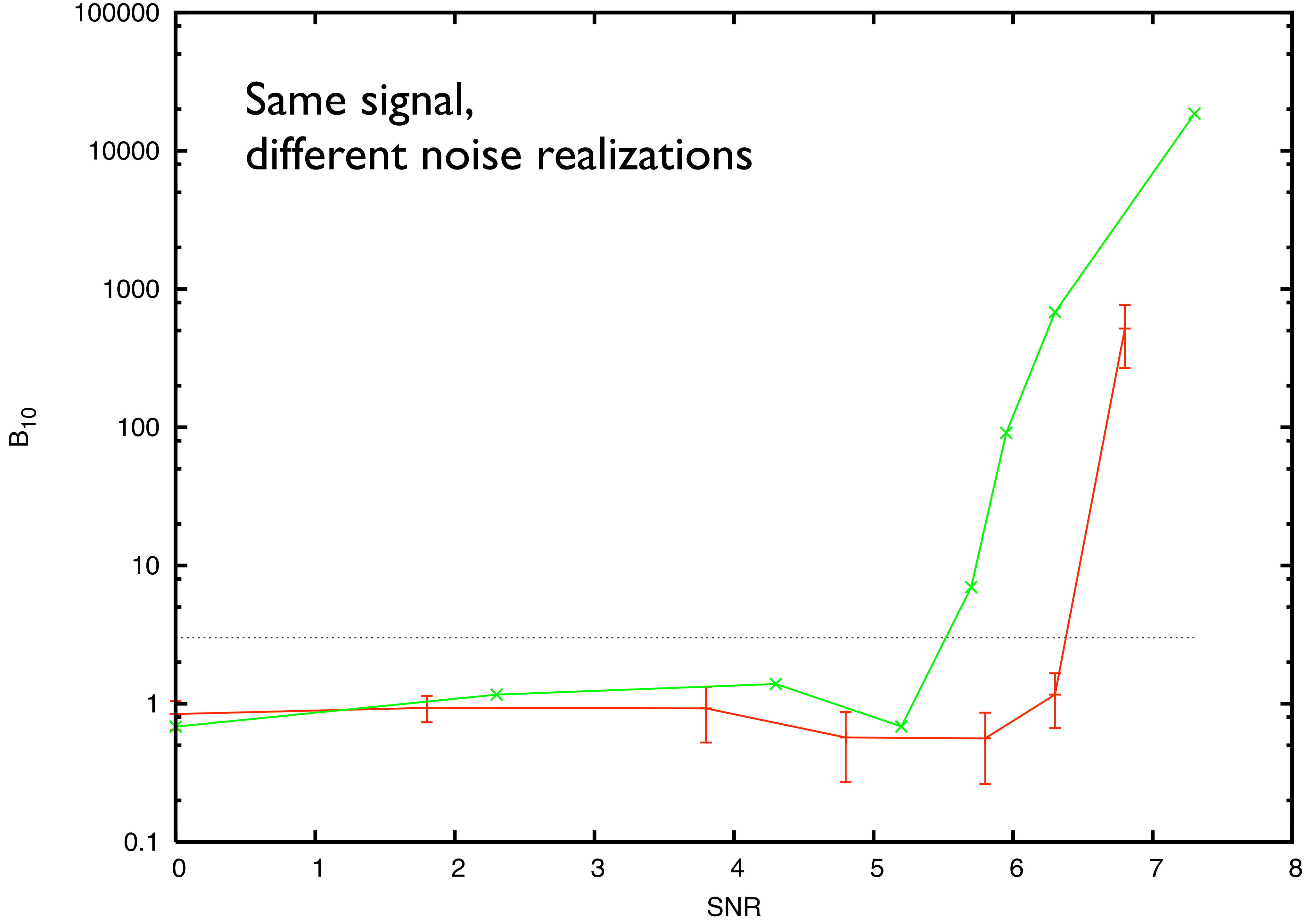
Reverse Jump Markov Chain Monte Carlo



Reverse Jump Markov Chain Monte Carlo



Reverse Jump Markov Chain Monte Carlo



Parallel Tempering (Swendsen & Wang, 1986)

Ordinary MCMC techniques side-step the need to compute the evidence. PT uses multiple, coupled chains to improve mixing, and also allows the evidence to be computed.

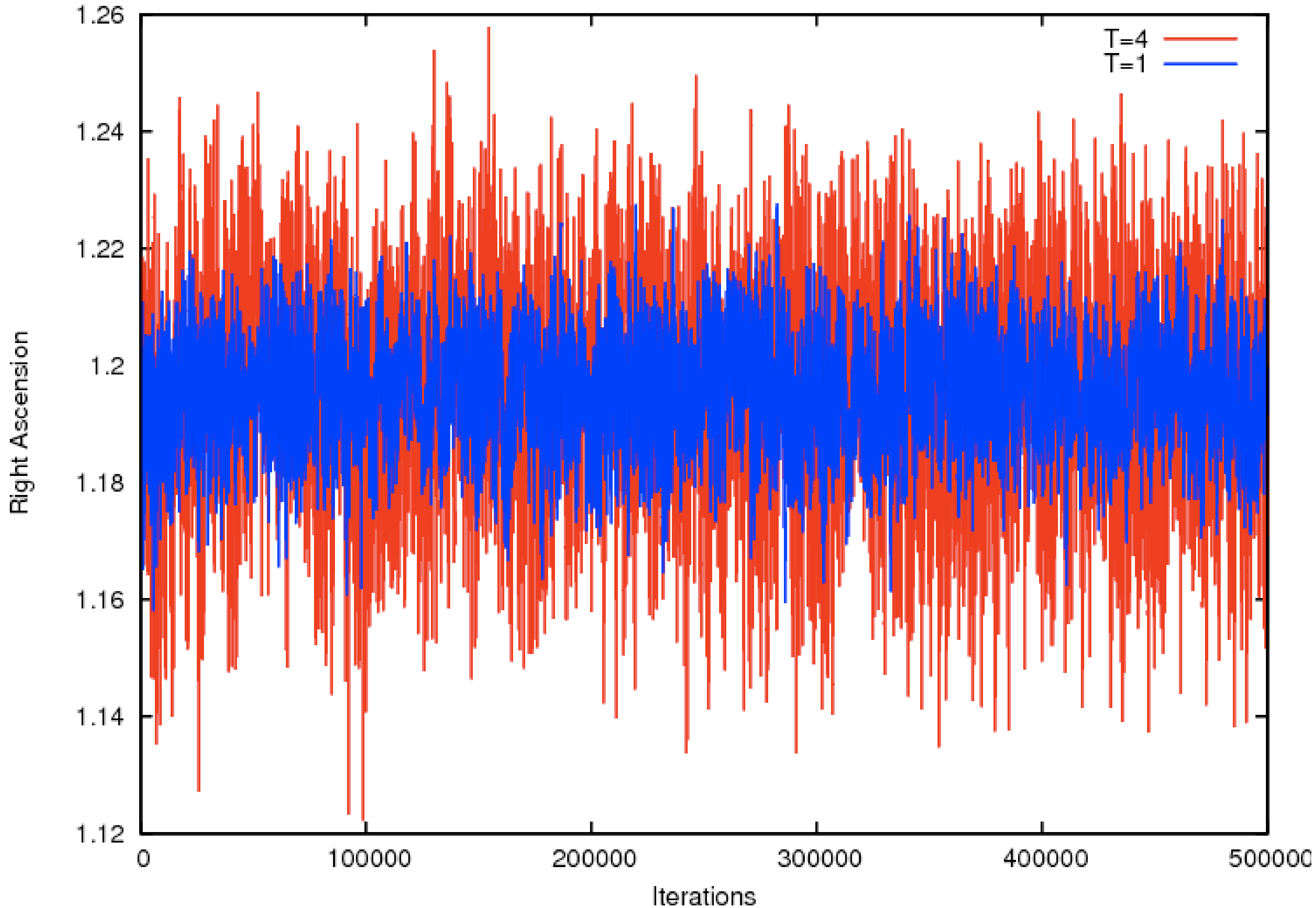
$$\pi(\vec{\lambda}|s,\beta) = p(\vec{\lambda})p(s|\vec{\lambda})^{\beta} \qquad \beta = \frac{1}{T}$$

Inter-chain transition probability

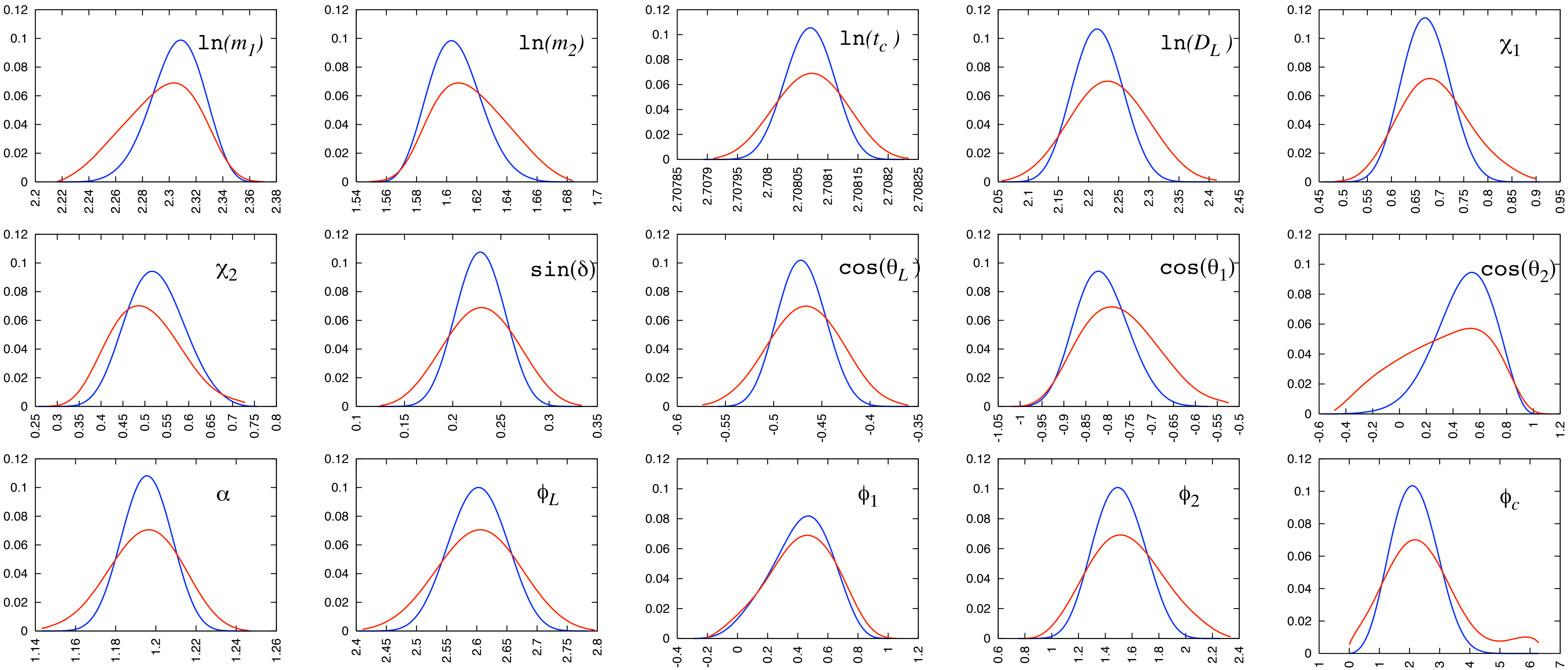
$$H = \min\left(1, \frac{\pi(\vec{\lambda}_{i+1}|s,\beta_i)\pi(\vec{\lambda}_i|s,\beta_{i+1})}{\pi(\vec{\lambda}_i|s,\beta_i)\pi(\vec{\lambda}_{i+1}|s,\beta_{i+1})}\right)$$

$\beta = 1$ chain yields the usual pdf, while the hotter chains improve the mixing

Parallel Tempering



Parallel Tempering



Parallel Tempering + Thermodynamic Integration (Goggans & Chi, 2004)

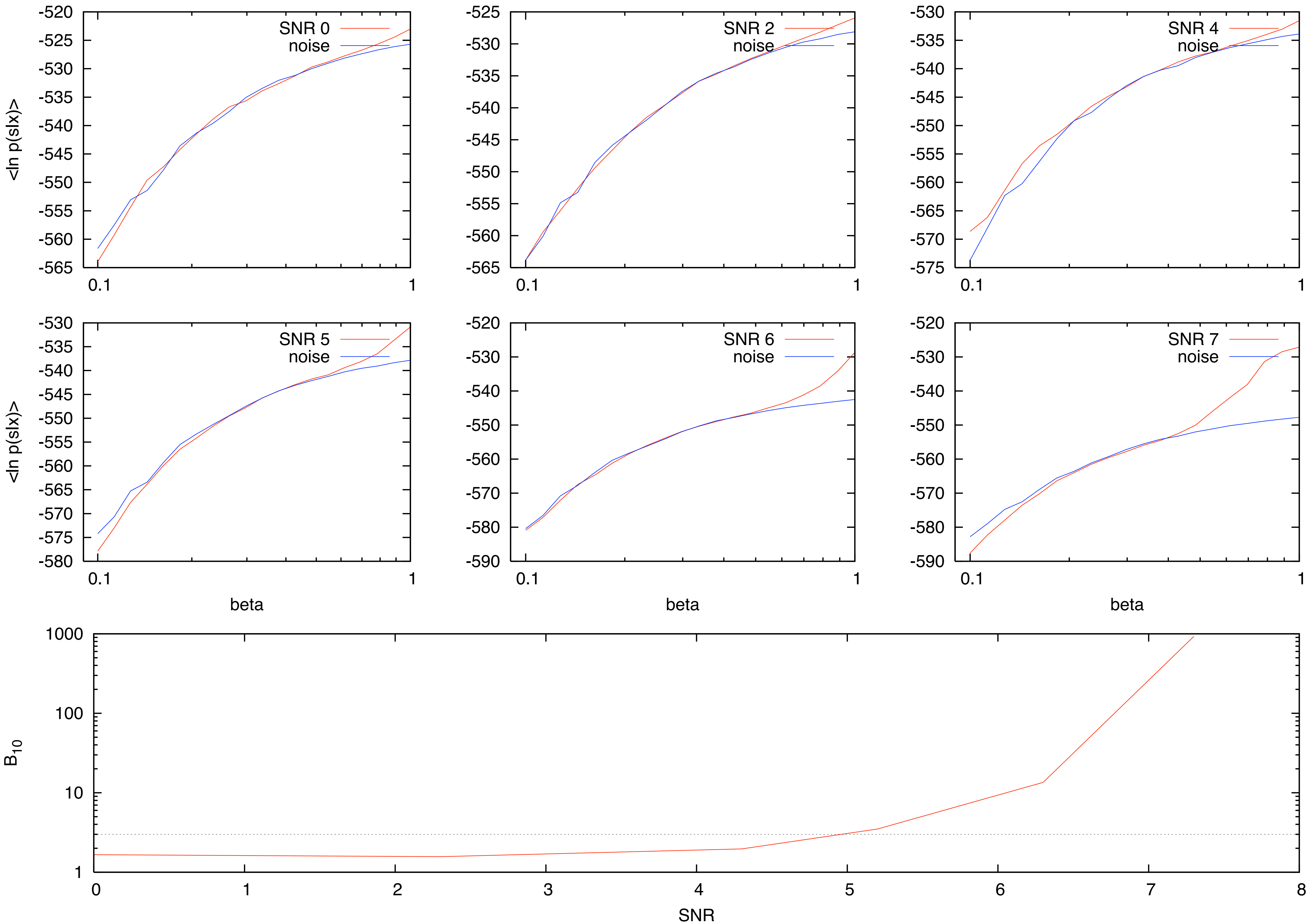
Define partition function

$$Z(\beta) = p(s, \beta) = \int d\vec{\lambda}\, \pi(\vec{\lambda}|s, \beta)$$

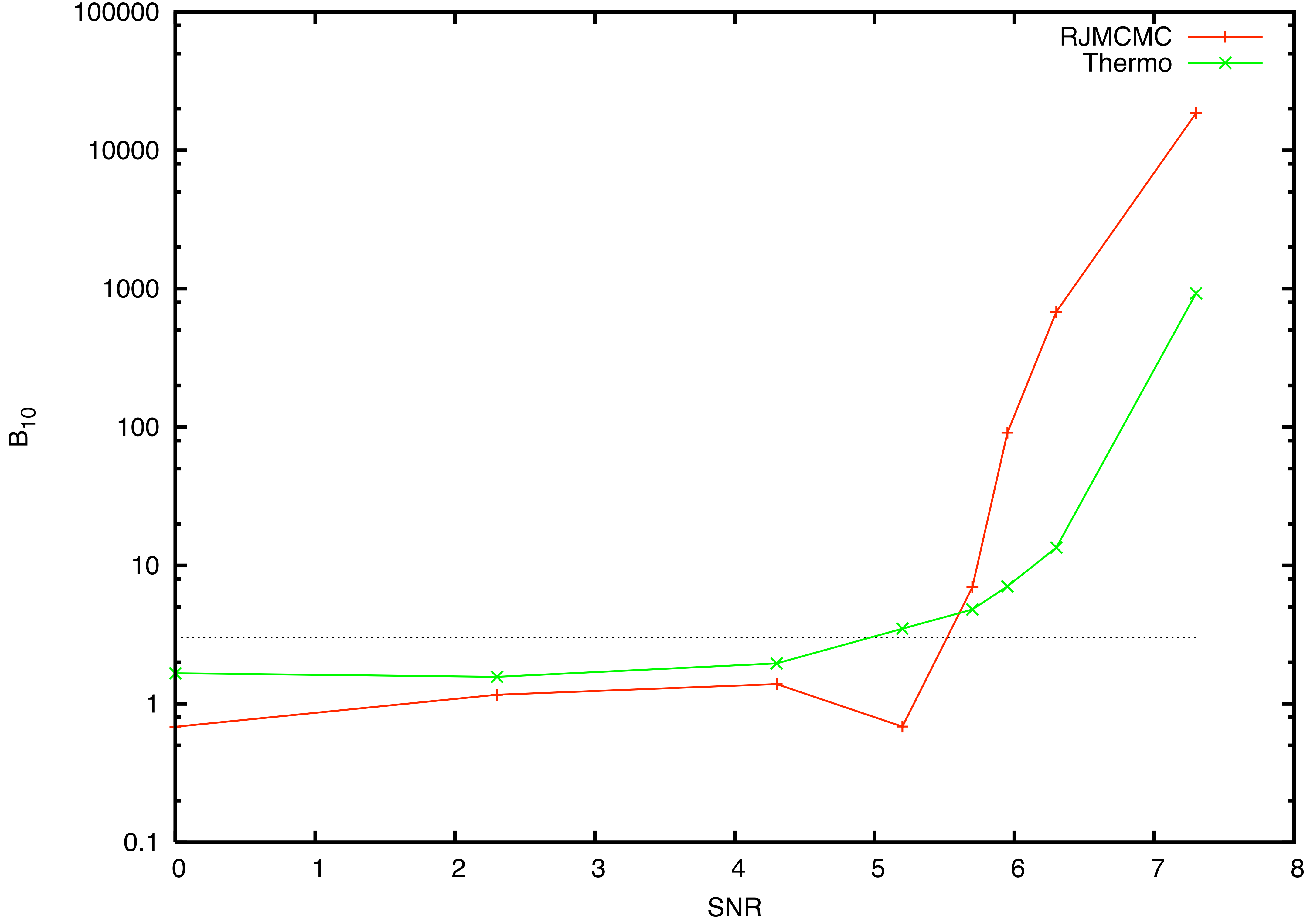
$$\Rightarrow \quad \frac{d}{d\beta} \ln Z(\beta) = \langle \ln p(s|\vec{\lambda}) \rangle_\beta$$

$$\Rightarrow \ln p(s) = \int_0^1 d\beta \, \langle \ln p(s|\vec{\lambda}) \rangle_\beta$$

Thermodynamic Integration



Thermodynamic Integration



Vegas Algorithm (Lepage 1978)

$$\int d\vec{\lambda}\, f(\vec{\lambda}) = \left\langle \frac{f}{g} \right\rangle_g$$

Sampling function g approximates f, but is easier to sample from (e.g. use Fisher Matrix approximation to the posterior)

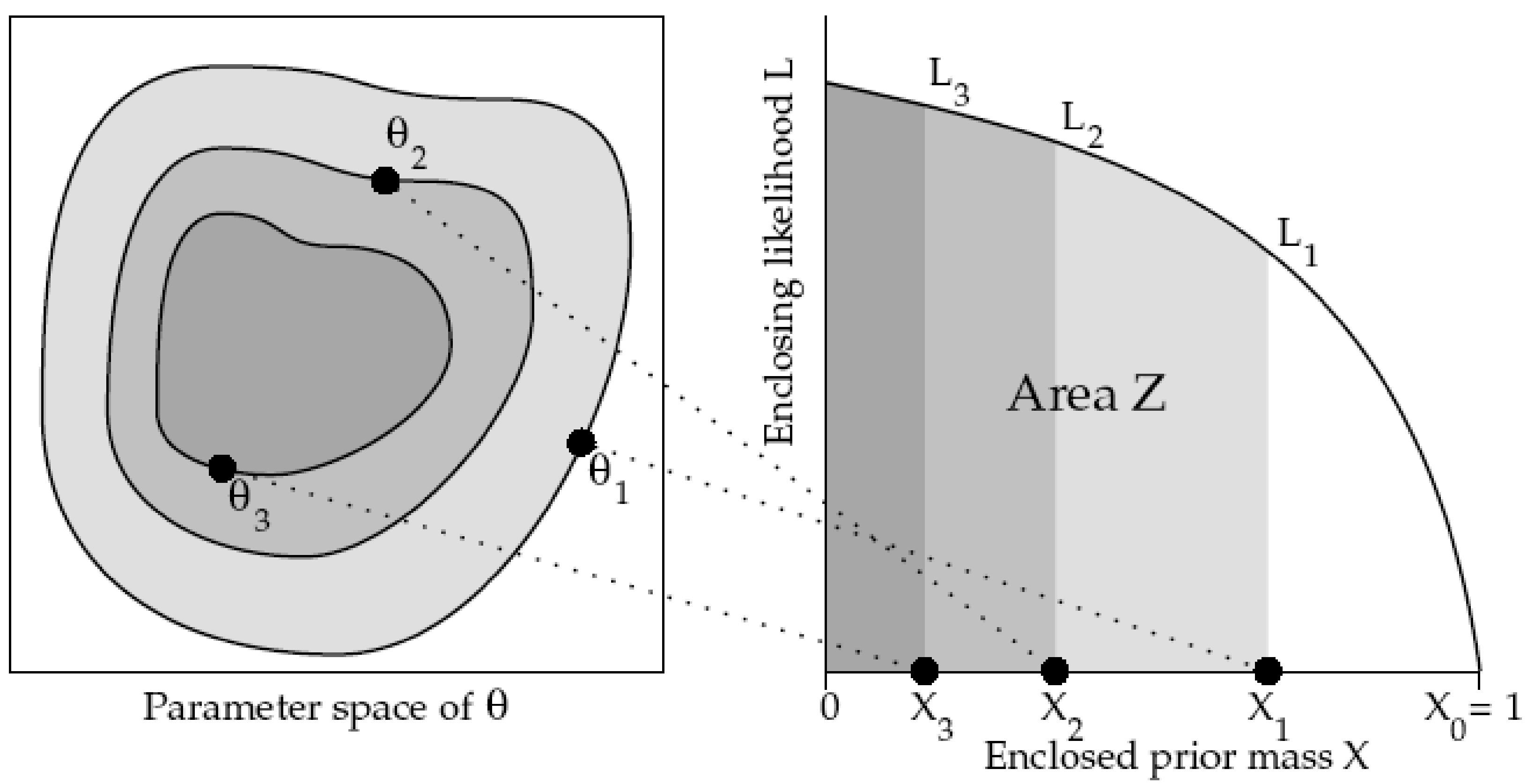
Nested Sampling (Skilling, 2004)

$$X(\mathcal{L}) = \int_{\mathcal{L}(\vec{\lambda})>\mathcal{L}} p(\vec{\lambda})\, d\vec{\lambda}$$

Define prior mass $X(\mathcal{L})$ enclosed by level sets of the likelihood $\mathcal{L} = p(s|\vec{\lambda})$

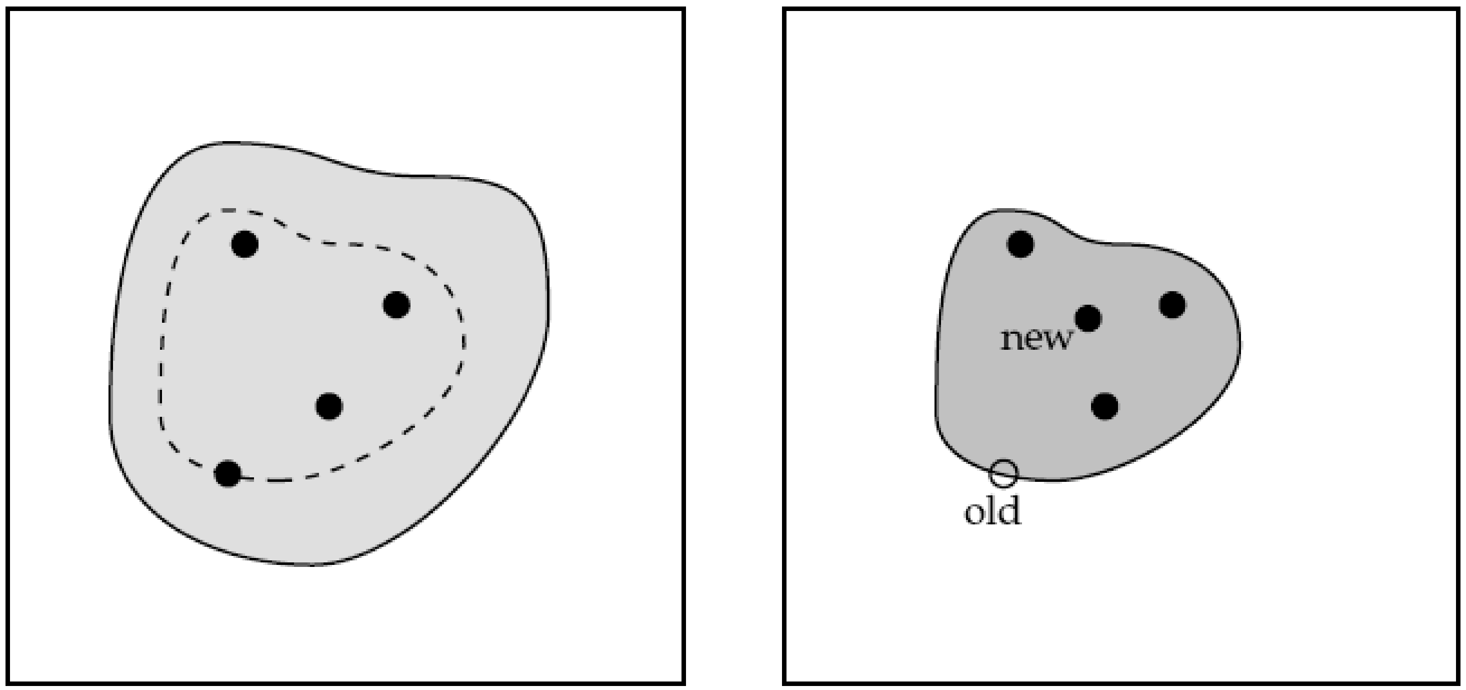
Has a well defined inverse $\mathcal{L}(X)$

Evidence: $p(s) = Z = \int_0^1 \mathcal{L}(X)\, dX$



Nested Sampling

Start by drawing N points from the prior. By construction has $X_0 = 1$.



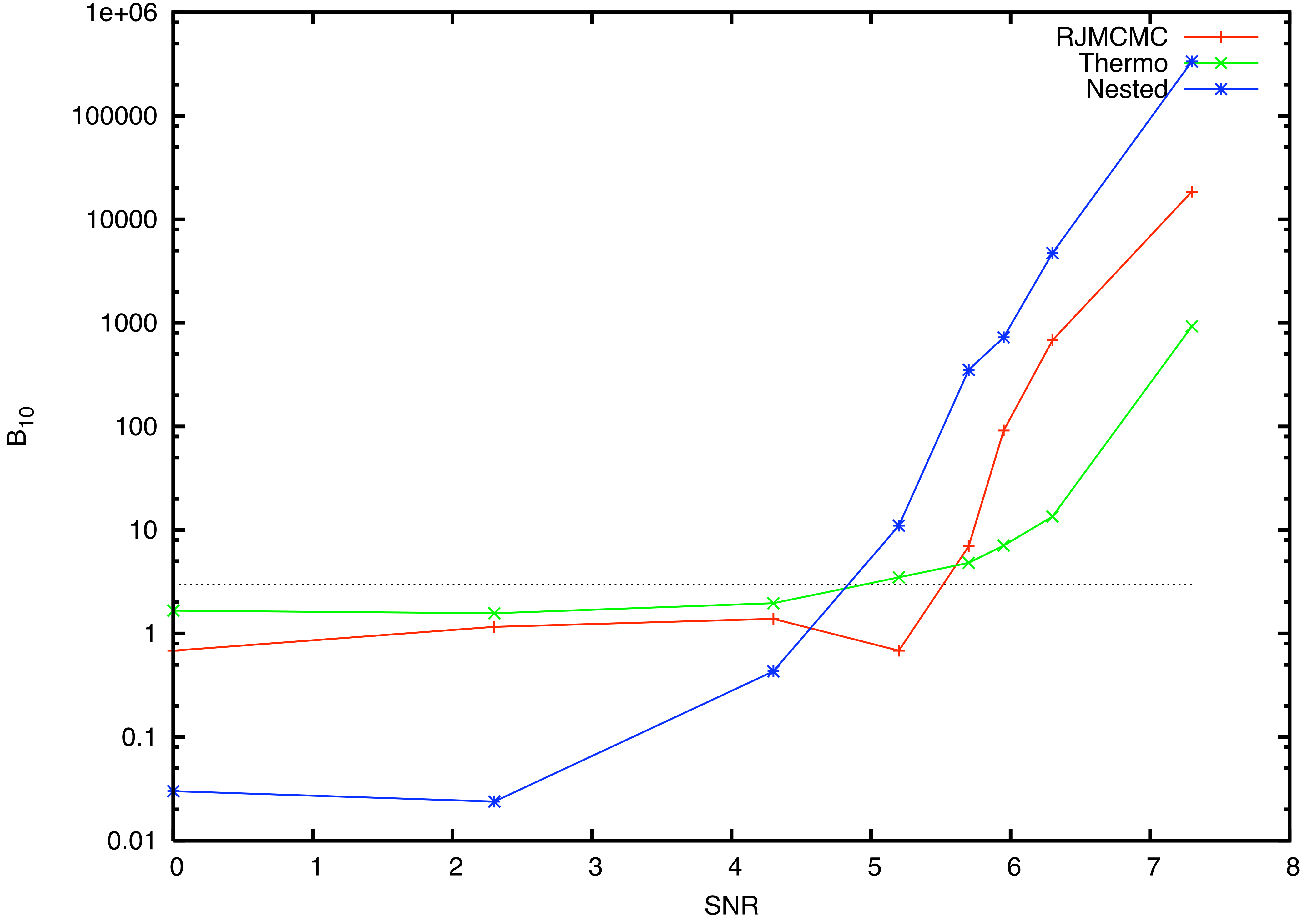
Iterate i -1   Iterate i

"Tightening the noose"

★ Delete point with lowest likelihood and draw a new point with the constraint that its likelihood exceeds that of the one deleted. We now have $X_i < X_{i-1}$ and $\mathcal{L}_i > \mathcal{L}_{i-1}$

Repeat ★ until a certain convergence condition is met and integrate Z using trapezoid rule.

Nested Sampling

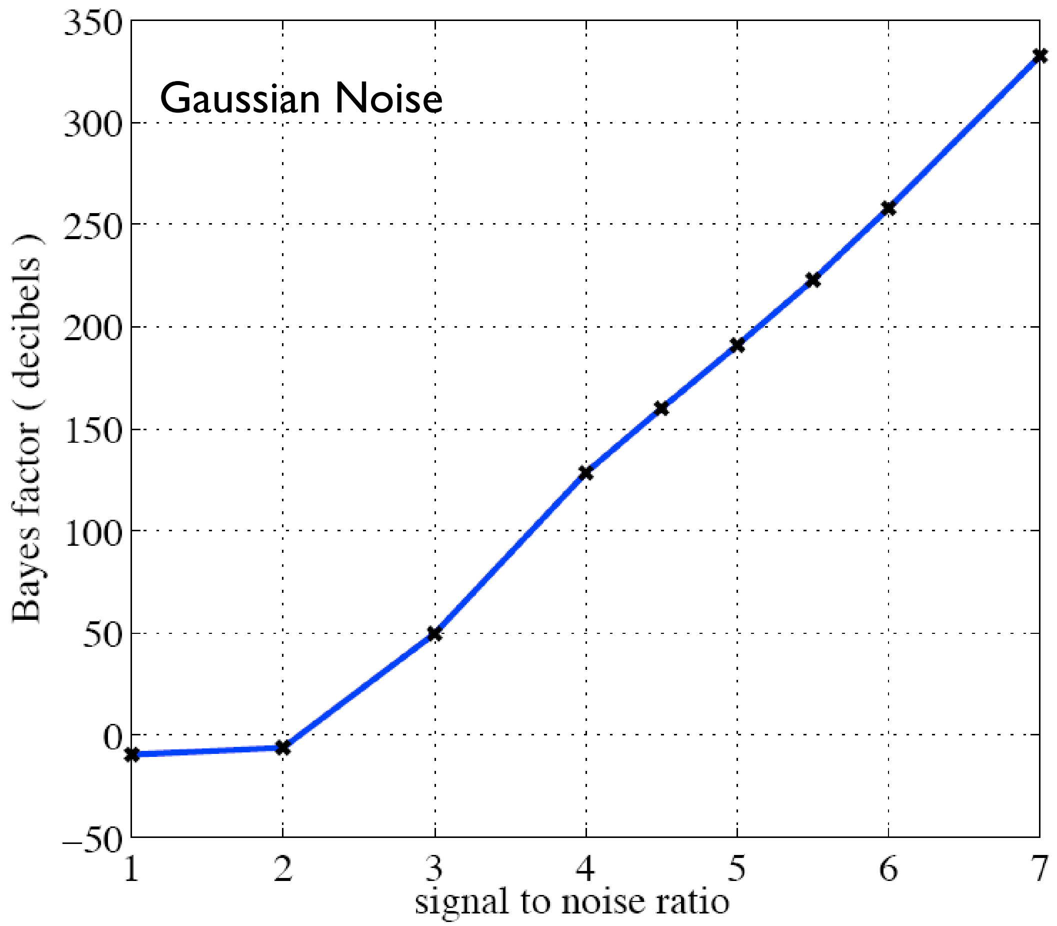


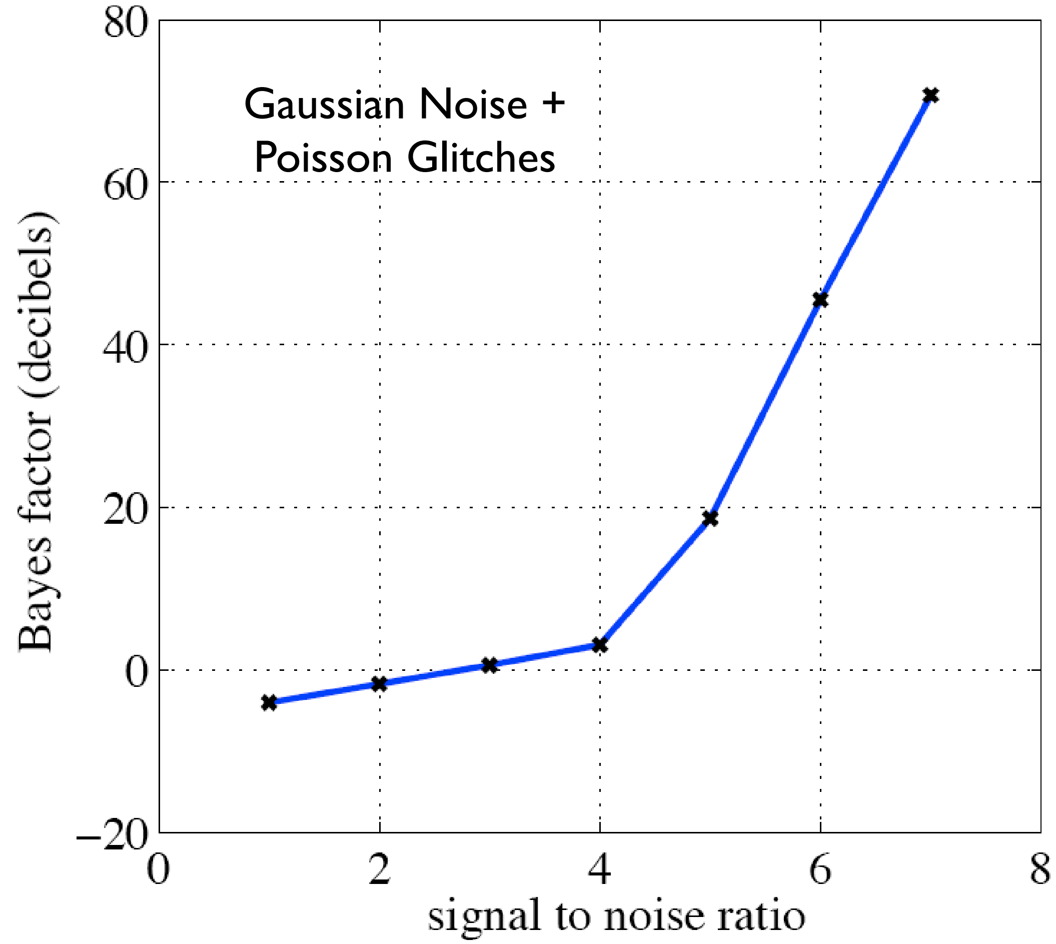
Nested Sampling: LIGO Inspiral

Veitch & Vecchio, arXiv:0801.4313 [gr-qc]

Signal Model: Single Interferometer 0PN inspiral, 4 parameters $\vec{\lambda} \rightarrow \{A, \mathcal{M}, t_c, \phi_c\}$

Noise model: Stationary Gaussian Noise, used no free parameters





Tested on simulated gaussian noise and noise with simulated glitches

Can LIGO use this?

- Already are to some extent when looking at consistency of MCMC parameter estimation chains - might as well do it properly

- Would be nice to have a likelihood function that includes all the data

$$p(s, a|\vec{\lambda})$$

auxiliary channels

Can LIGO use this?

- Already are to some extent when looking at consistency of MCMC parameter estimation chains - might as well do it properly

- Would be nice to have a likelihood function that includes all the data

$$p(s, a|\vec{\lambda})$$

auxiliary channels

Can LIGO use this?

- Already are to some extent when looking at consistency of MCMC parameter estimation chains - might as well do it properly

- Would be nice to have a likelihood function that includes all the data

$$p(s, a|\vec{\lambda})$$

↖ auxiliary channels

$$\text{e.g.} \quad p(s, a|\vec{\lambda}) = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-a^2/2\sigma_a^2}\, p(s|\vec{\lambda})$$

Can LIGO use this?

Usual form of likelihood assumes Gaussian noise

$$p(s|\vec{\lambda}) = C\, e^{-\chi^2/2}\,, \qquad \chi^2 = (s-h|s-h)$$

Performance will be more robust if the noise model allows for larger tails, and if the noise parameters are part of the fit

$$p(s|\vec{\lambda}) = C\, e^{-\chi^2/2} + C'\, e^{-\chi^2/2\alpha}$$

Small additional component with larger variance

Allen, Creighton, Flanagan & Romano (PRD 67, 122002, 2003)

Next Steps

- Get all approaches to agree
- Apply to simulated LIGO inspirals
- Work with Vecchio, Christensen etc to implement in LIGO software for candidate follow-ups