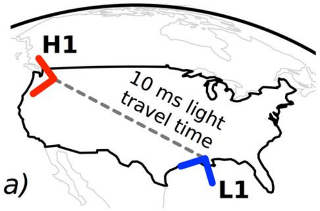
The Label Switching Problem
or
“How I turned a toy problem into my first paper”

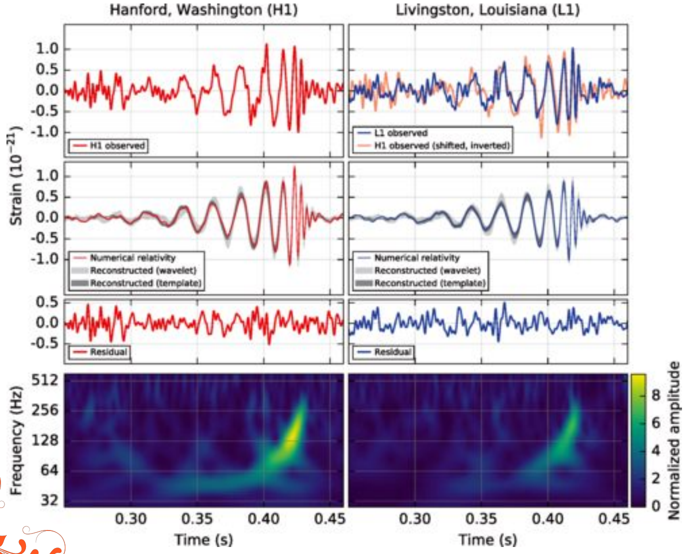
R. Buscicchio
Astrostatistics
Univ. Milano-Bicocca - 2022/04/27
riccardo.buscicchio@unimib.it
Physical Review D-100.084041 (2019)

Challenging astrostatistics with LISA

LIGO

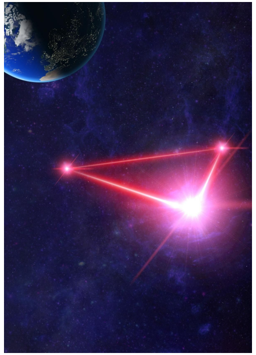


(+noise)

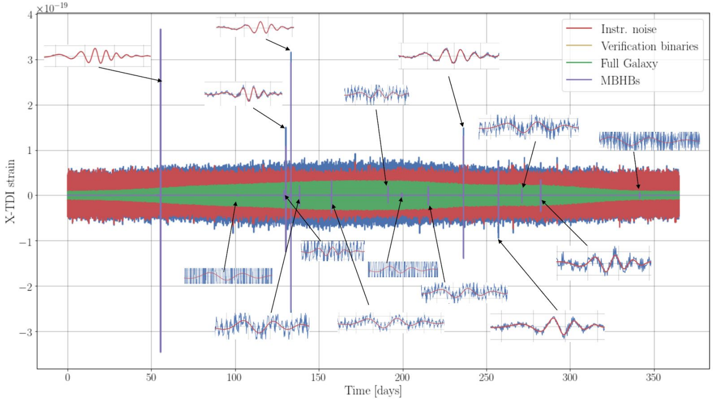


LISA

NASA/JPL-Caltech/NASAEA/ESA/
CXC/STScl/GSFCSVS/S.Barke



(+noise)

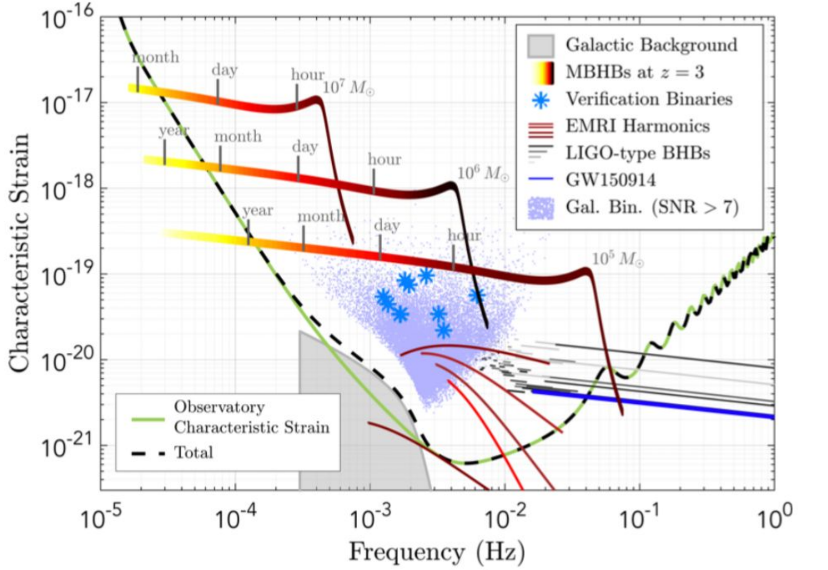


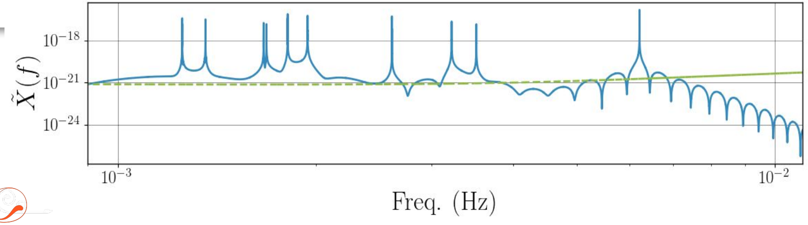
credits: courtesy of LISA Data Challenge working group

The Sources landscape

Switching labels

WHAT AN ASTROPHYSICIST SEES

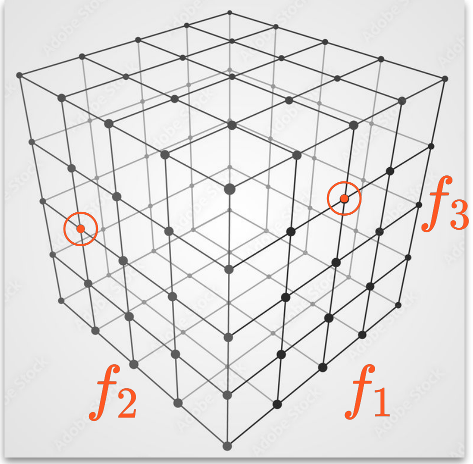




WHAT AN ASTROSTATISTICIAN SEES

$$\log \mathcal{L}(d \mid f_1, \ldots, f_N) \sim \sum_{i=1}^{N} \mathcal{N}(d \mid f_i)$$

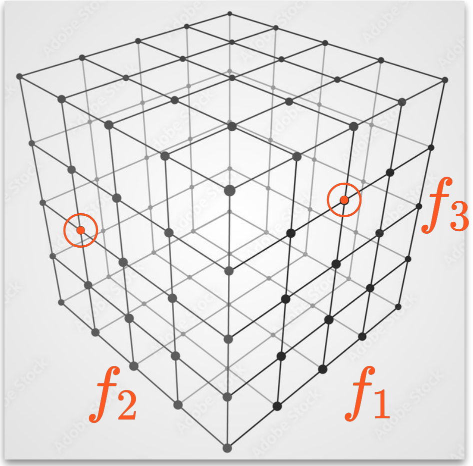
WHAT A STOCHASTIC SAMPLER SEES N=3



Are the two distinguishable?

What about 10^7 sources? How many points?

Possible solutions



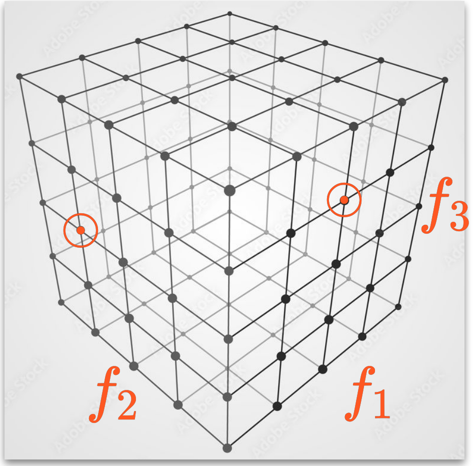
$\odot\ \odot\ \odot\ \odot\ \odot\ \odot$

$f_1 > f_2 > f_3\,?$

$\odot\ \otimes\ \otimes\ \otimes\ \otimes\ \otimes$

Viable, but very **inefficient:**
$1000! \sim 10^{2567}$

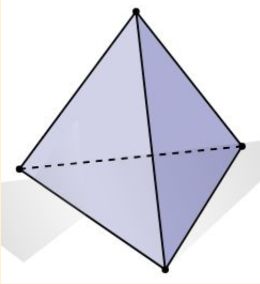
Possible solutions



$$\begin{array}{c}\odot\\\odot\\\odot\\\odot\\\odot\\\odot\end{array} \quad f_1 > f_2 > f_3\,? \quad \begin{array}{c}\odot\\\otimes\\\otimes\\\otimes\\\otimes\\\otimes\end{array}$$

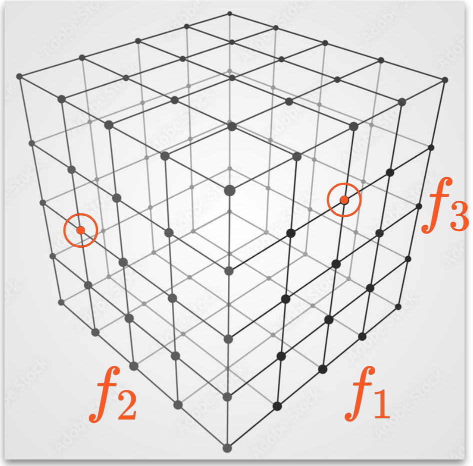
Viable, but very **inefficient:**
$1000! \sim 10^{2567}$

Restrict prior



- Samplers like **a lot** cubes
- Difficult to generalize to high N

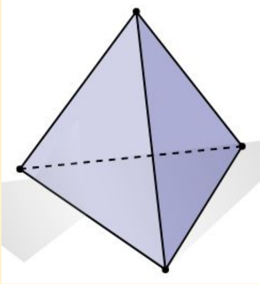
Possible solutions



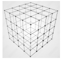
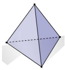
$\odot$ $\odot$ $\odot$ $\odot$ $\odot$ $\odot$ $\quad f_1 > f_2 > f_3$? $\quad \odot$ $\otimes$ $\otimes$ $\otimes$ $\otimes$ $\otimes$

Viable, but very **inefficient:**
$1000! \sim 10^{2567}$

Restrict prior



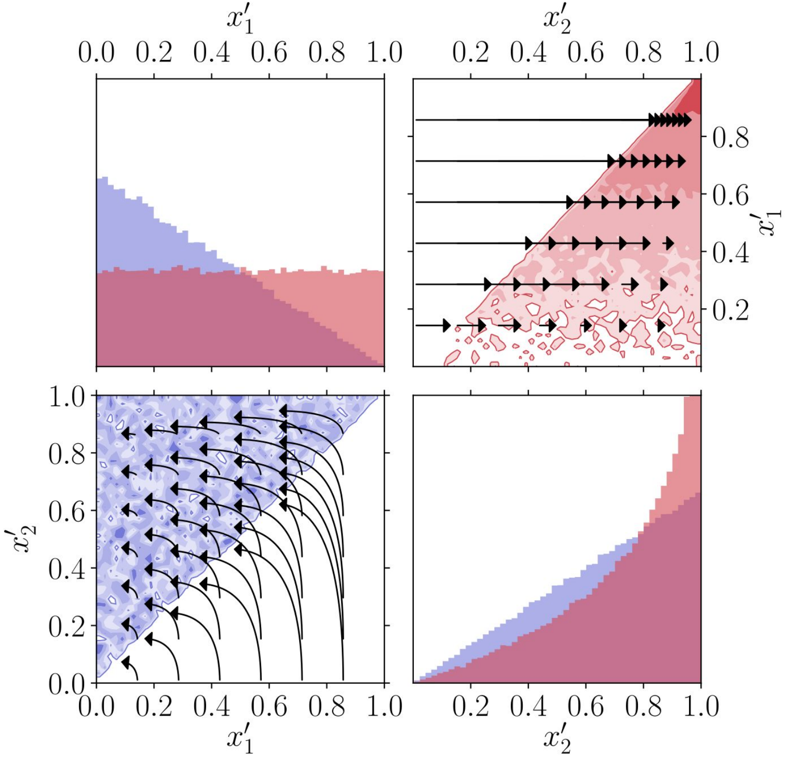
- Samplers like **a lot** cubes
- Difficult to generalize to high N

Question: can we let samplers explore  while we interpret them as points in  ? » Priors!!

Priors, priors, priors

$$x'_1 = 1 - \sqrt{1 - x_1}$$
$$x'_2 = x'_1 + (1 - x'_1)\, x_2$$

Why is this good?



$$x'_1 = x_1$$
$$x'_2 = x'_1 + (1 - x'_1)\, x_2$$

Why is this **not** good?

Priors, priors, priors

```python
def phi(x):
        K = len(x)
        i = numpy.arange(K)
        inner = numpy.power(1-x,1/(K-i))
        X = 1 - numpy.cumprod(inner)
        return X
```

$$x_1' = 1 - \sqrt{1 - x_1}$$
$$x_2' = x_1' + (1 - x_1')\, x_2$$

PHYSICAL REVIEW D **100,** 084041 (2019)

Label switching problem in Bayesian analysis for gravitational wave astronomy

Riccardo Buscicchio,* Elinore Roebber, Janna M. Goldstein, and Christopher J. Moore
School of Physics & Astronomy and Institute for Gravitational Wave Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom

(Received 26 July 2019; published 21 October 2019)

The label switching problem arises in the Bayesian analysis of models containing multiple indistinguishable parameters with arbitrary ordering. Any permutation of these parameters is equivalent, therefore models with many such parameters have extremely multimodal posterior distributions. It is difficult to sample efficiently from such posteriors. This paper discusses a solution to this problem which involves carefully mapping the input parameter space to a high dimensional *hypertriangle*. It is demonstrated that this solution is efficient even for large numbers of parameters and can be easily applied alongside any stochastic sampling algorithm. This method is illustrated using two example problems from the field of gravitational wave astronomy.

DOI: 10.1103/PhysRevD.100.084041

I. INTRODUCTION

It sometimes occurs in Bayesian inference problems that the target distribution depends on several parameters whose ordering is arbitrary. Three examples are immediately apparent from the field of gravitational wave (GW) astronomy alone. First, when describing a compact binary with component masses m_1 and m_2, the likelihood is symmetric under exchange of the labels 1 and 2 (provided all other relevant parameters are suitably adjusted simultaneously). Second, when analyzing GW time series data containing two or more overlapping sources of the same type, the likelihood is invariant under exchanging all of the parameters of any pair of sources. And third, when analyzing the parameters of a population of observed GW events, mixture models can be used to model the population and/or to infer the presence of distinct astrophysical populations. In this case the hyperlikelihood for the population parameters may be invariant under exchanging the parameters of the population components.

Sometimes a simple reparametrization and restricting the parameter range is enough to remove the degeneracy arising from the arbitrary ordering. In the first case of the binary with two component masses, it is possible to define, say, the total mass $M = m_1 + m_2$ and mass ratio $q = m_2/m_1$ and to sample these over the restricted ranges $M > 0$ and $q \leq 1$. This covers only the restricted portion of the parameter space $m_1 \geq m_2$, thereby removing the symmetry from the likelihood.

The second and third examples are more problematic as they are not restricted to just 2 degrees of freedom. In each case the target distribution has a high degree of symmetry and is invariant under permutations of some number of labels, K. A great deal of literature is devoted to this *label switching problem* in the context of mixture models [1–7]. The invariance of the target distribution under permutations means that if the posterior has a peak (or mode) at a particular point in parameter space it will necessarily have peaks at all $K!$ points related by symmetry. The extreme scaling of this multimodality poses a serious obstacle to any sampling algorithm in moderate or high dimensional problems.

The most natural way to solve the label switching problem is to impose an *artificial identifiability constraint*. Searching over the restricted region $m_1 \geq m_2$ of the binary component mass space is an example of such a constraint in 2 dimensions. In the K dimensional problem this can be generalized by demanding a certain ordering of the parameters; see, e.g., [2,4–6]. Restricting to this small region of parameter space avoids all symmetries and removes the excess multimodality. It is also obvious that if one can adequately explore the restricted parameter space satisfying the artificial identifiability constraint then, by symmetry, this is equivalent to exploring the full space.

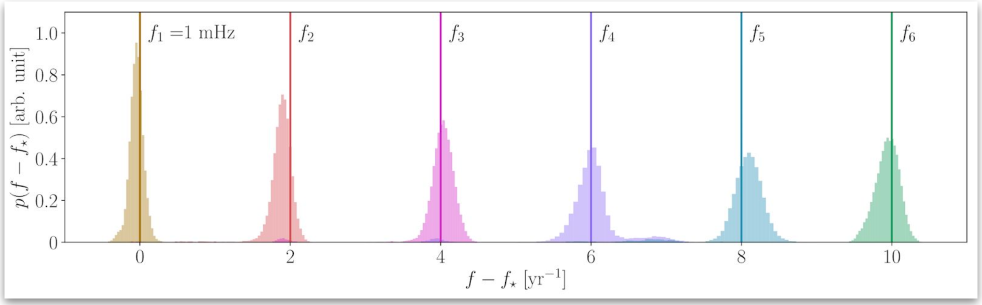
It remains to implement a suitable artificial identifiability constraint in practical inference problems. This problem can be approached in several ways. For example, when using an Markov chain Monte Carlo (MCMC) to explore the target distribution the proposal can be augmented by composing with a sorting function; i.e., propose a point then reorder the parameters such that the constraint is satisfied [7]. Alternatively, the log-prior distribution can be crudely modified so that it returns $-\infty$ for any point not satisfying the constraint. Either of these will ensure the chain never leaves the desired region of parameter space.

*riccardo@star.sr.bham.ac.uk

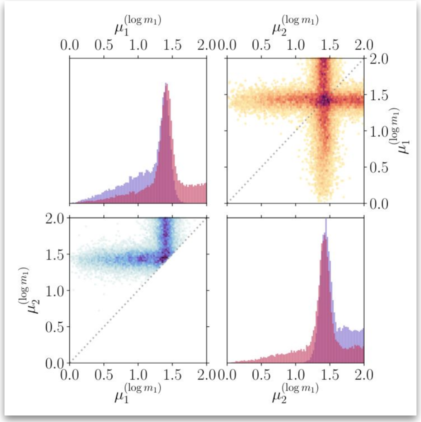
2470-0010/2019/100(8)/084041(13) 084041-1 © 2019 American Physical Society

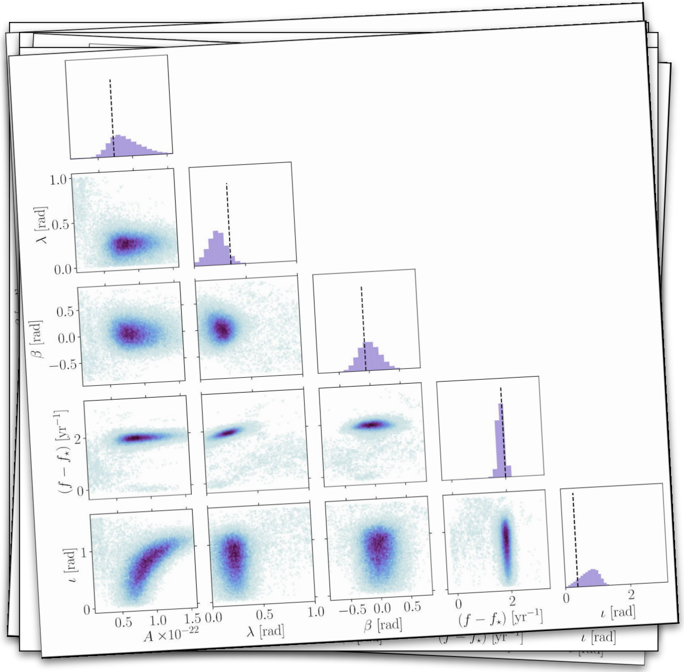
The label switching solution

Multiple sources inference



Buscicchio & al.
Physical Review D - 100.084041(2019)





$$\mathbf{\Lambda}_k = \{A_k, f_k, \dot{f}_k, \lambda_k, \sin\beta_k, \cos\iota_k, \psi_k, \Phi_k\}$$

$$h_+(t; \mathbf{\Lambda}) = A\left(1 + \cos^2\iota\right)\cos(2\pi f t - \Phi)$$

$$h_\times(t; \mathbf{\Lambda}) = -2A\cos\iota\sin(2\pi f t - \Phi)$$

Launch in the '30s!