When is a gravitational-wave signal stochastic?

Neil J. Cornish and Joseph D. Romano

Phys. Rev. D **92**, 042001 — Published 7 August 2015

DOI: 10.1103/PhysRevD.92.042001

When is a gravitational-wave signal stochastic?

Neil J. Cornish
Department of Physics, Montana State University, Bozeman, MT 59718, USA.

Joseph D. Romano
Department of Physics and Astronomy and Center for Gravitational-Wave Astronomy, University of Texas at Brownsville, Brownsville, TX 78520, USA.

We discuss the detection of gravitational-wave backgrounds in the context of Bayesian inference and suggest a practical definition of what it means for a signal to be considered stochastic—namely, that the Bayesian evidence favors a stochastic signal model over a deterministic signal model. A signal can further be classified as Gaussian-stochastic if a Gaussian signal model is favored. In our analysis we use Bayesian model selection to choose between several signal and noise models for simulated data consisting of uncorrelated Gaussian detector noise plus a superposition of sinusoidal signals from an astrophysical population of gravitational-wave sources. For simplicity, we consider co-located and co-aligned detectors with white detector noise, but the method can be extended to more realistic detector configurations and power spectra. The general trend we observe is that a deterministic model is favored for small source numbers, a non-Gaussian stochastic model is preferred for intermediate source numbers, and a Gaussian stochastic model is preferred for large source numbers. However, there is very large variation between individual signal realizations, leading to fuzzy boundaries between the three regimes. We find that a hybrid, trans-dimensional model comprised of a deterministic signal model for individual bright sources and a Gaussian-stochastic signal model for the remaining confusion background outperforms all other models in most instances.

PACS numbers: 04.80.Nn, 04.30.Db, 07.05.Kf, 95.55.Ym

I. INTRODUCTION

A stochastic background of gravitational radiation is usually defined as a random gravitational-wave signal produced by a *large number of weak, independent, and unresolved sources*. It can be of either astrophysical or cosmological origin. The signal is random in the sense that it can be characterized only statistically, in terms of expectation values of the Fourier components of a plane-wave expansion of the metric perturbations. For a sufficiently large number of independent sources, the background will be Gaussian by the central limit theorem. Knowledge of the first two moments of the distribution will then suffice to determine all higher-order moments, meaning that the quadratic expectation values (or *covariance* matrix) of the Fourier components completely define a Gaussian background of gravitational radiation. For non-Gaussian backgrounds, the only difference is that the probability distribution of the Fourier components is no longer Gaussian. Thus, third and/or higher-order moments of the distribution are now required.

Although there is general agreement with the above definition, there has been some confusion and/or disagreement about some of the defining properties of a stochastic background, in particular, related to the resolvability of a signal and its relationship to duty-cycle [1–4]. In order to avoid such confusion in this paper, we give operational definitions for these properties, framed in the context of Bayesian inference. For instance, we define a signal to be *stochastic* if it is more parsimonious (in a Bayesian model selection sense) to search for that signal using a stochastic signal model for the waveform than using a deterministic signal model. We also define a signal to be *resolvable* if it can be decomposed into *separate* (e.g. non-overlapping in either time or frequency) and *individually detectable* signals, again in a Bayesian model selection sense. Signals may be separable even when overlapping in time and frequency if the detector has good sky resolution, or the signals have additional complexities due to effects such as orbital evolution and precession.

This definition of resolvability is more restrictive than that of Rosado [1], who defines a signal to be resolvable if it is separable, independent of detectability. With our definition, it is possible to have separable signals that are not detectable (e.g., "subthreshold" low-duty cycle bursts [5] or non-overlapping low-SNR sinusoids), and signals that are detectable but not resolvable (e.g., a Gaussian stochastic background integrated over a large enough time or large enough frequency band).

In addition, for non-Gaussian backgrounds associated with the superposition of signals from many astrophysical sources, there will sometimes be cases where a few bright signals standout above the lower-amplitude "confusion" background. These resolvable deterministic signals should be 'subtracted' from the data, leaving a residual non-deterministic background whose statistical properties we would like to determine. In the context of Bayesian inference, this 'subtraction' is done by allowing *hybrid* signal models, which consist of both parametrized deterministic signals and non-deterministic backgrounds. By using such models we can investigate the statistical properties of the residual background without the influence of the resolvable signals. This is ultimately the prop-

erty of the stochastic background that we would like to determine.

The closely related question of whether a population of astrophysical signals is more likely to be first detected via a stochastic cross-correlation analysis or a template-based search for individual signals has recently been considered by Rosado *et al.* [6] in the context of pulsar timing detection of the low-frequency, slowly-evolving signals from binary supermassive black holes, and by Mandel [7] in the context of ground-based interferometer detections of chirping neutron star binaries and continuous waves from non-axisymmetric spinning neutron stars. Rosado *et al.* found that pulsar timing arrays are most likely to first detect a stochastic background, though in some cases a bright and nearby source may be detected first. Mandel concluded that individual signals will always be detected first for non-overlapping, evolving signals, while a stochastic background may be detected first for non-evolving, overlapping signals, and gave conditions for when this might occur. Our results are broadly in agreement with these studies, though it is difficult to directly compare our results since we frame the problem in different ways and have different definitions for what it means for a signal to be considered deterministic or stochastic.

In this paper, we apply Bayesian inference to non-Gaussian gravitational-wave backgrounds, which are produced whenever the overlap of the gravitational-wave signals in time-frequency space is sufficiently low that the central-limit theorem does not apply. Previous analyses for non-Gaussian backgrounds have typically been framed in the context of frequentist statistics, involving modifications [5, 8–11] of the standard optimally-filtered cross-correlation statistic [12] used to search for Gaussian backgrounds. Here we use Bayesian inference to address the same problem. We apply Bayesian model selection to compare several signal+noise models for simulated data consisting of uncorrelated Gaussian detector noise plus a superposition of sinusoidal signals from an astrophysical population of gravitational-wave sources. The analysis is done in the frequency domain since the signals we consider are well localized in frequency and spread out in time, but our results apply equally well to signals that are localized in time and spread out in frequency, such as a population of burst signals occurring with some Poisson rate. For simplicity, we consider a pair of co-located and co-aligned detectors with white detector noise, but the method can be extended to more realistic detector configurations and power spectra.

The general trend we observe from our simulations is that a deterministic signal model is favored whenever the number of sources contributing to the background is sufficiently small; a non-Gaussian stochastic model is preferred for an intermediate number of sources; and a Gaussian stochastic model is preferred for a large number of sources. However, due to large variations between individual signal realizations, the boundaries between the three regimes are not sharply defined. We find that a hybrid, trans-dimensional model comprised of a deterministic signal model for individual bright sources and a Gaussian-stochastic signal model for the remaining confusion background outperforms all other models in most instances.

The remainder of the paper is organized as follows: In Sec. II we give a brief overview of Bayesian inference, and apply it to the specific case of non-Gaussian gravitational-wave backgrounds in Sec. III. There we define the relevant noise and signal probability distributions, likelihood functions, prior and posterior probability distributions, etc. needed for our analysis. In Sec. IV we define the various signal+noise models that we use for the Bayesian model selection calculations, the results of which, for simulated data, are described in detail in Sec. V. Finally, in Sec. VI, we discuss the relevance of the results in the context of current searches for gravitational-wave backgrounds. Appendix A includes a discussion of different approaches for calculating Bayes factors.

II. BAYESIAN INFERENCE – OVERVIEW

Bayesian inference is a powerful tool for assessing the plausibility of hypotheses, given a set of observations and prior information [13]. It allows you to update your degree of belief in a particular hypothesis, based on how well the hypothesis (or model) fits the observed data. It also implements a quantitative version of Occam's razor [13], which says that given two models that fit the data equally well, the simpler model should be preferred. This result falls naturally out of a Bayesian model selection calculation, where one calculates the posterior odds ratio of one model against another. If two models fit the data equally well but have different parameter space volumes, then the model with the larger parameter space volume is penalized by the ratio of the larger parameter space volume to the smaller volume.

Using Bayesian inference to analyze a particular problem is very simple in principle—one applies Bayes' theorem,

$$p(\vec{\theta}|\mathbf{s}) = \frac{p(\mathbf{s}|\vec{\theta})\pi(\vec{\theta})}{\int d\vec{\theta}'\, p(\mathbf{s}|\vec{\theta}')\pi(\vec{\theta}')}, \tag{1}$$

to calculate posterior probability distributions given a likelihood function $p(\mathbf{s}|\vec{\theta})$ (which specifies the probabilty of the data given the model and the value of any parameters associated with it) and a prior probability distribution $\pi(\vec{\theta})$ for the model and its parameters.

In practice, however, these calculations can be extremely computationally-intensive, especially for models having a large number of parameters. But in recent years, thanks to advances in high-speed computing and the development of efficient sampling algorithms [14, 15], integrations over model parameter spaces having hundreds or even thousands of dimensions are now possible. Thus, the use of Bayesian inference to solve diverse problems

in the physical sciences has increased dramatically, given the ability to do numerical calculations, which, in the past, were not possible in practice.

In particular, in the field of gravitational-wave data analysis, it is now common to see Bayesian inference used for: (i) detector noise estimation and modeling [16, 17], (ii) sky-localization of signals from unmodeled gravitational-wave bursts, and (iii) parameter estimation for gravitational-wave signals associated with many different sources—binary inspiral events, continuous-wave sources (e.g., non-axisymmetric rotating neutron stars), and stochastic gravitational-wave backgrounds [18] of either astrophysical or cosmological origin.

Although Bayesian inference and frequentist optimally-filtered statistic methods give equivalent results for sufficiently simple signal and noise models and simple choices for the priors, the Bayesian formalism allows one to more easily handle problems involving more complicated models and/or non-trivial priors. This is the case when the model contains so-called *nuisance parameters* (such as non-negligible correlated noise), which are not of direct astrophysical interest, but nonetheless affect statistical statements about the signal parameters. For example, in the presence of correlated noise, the standard optimally-filtered cross-correlation statistic [12] for isotropic gravitational-wave backgrounds no longer corresponds to the optimal combination of the data from the two detectors. Calculating the maximum of the likelihood function is more complicated for this case, with no analytic closed-form solution in general. But a Bayesian approach to this problem, which numerically explores the likelihood function using e.g., Markov Chain Monte Carlo (MCMC) methods, is a viable alternative.

For gravitational-wave backgrounds generated by a superposition of signals from a population of astrophysical sources, Bayesian inference is particularly convenient since it allows one to compare several viable signal+noise models. Depending on the number of sources emitting gravitational waves in a particular time-frequency volume, the measured signal could be either: (i) stochastic and Gaussian distributed, (ii) stochastic but non-Gaussian, (iii) a superposition of individually resolvable signals, or (iv) some combination of both deterministic resolvable signals and a non-deterministic (i.e., stochastic) background. Using Bayesian model selection, we can rank these various models, and thus characterize the gravitational-wave component of the data. The following sections describe this procedure for the case of simulated data consisting of Gaussian white detector noise plus a superposition of sinusoidal signals from an astrophysical population of gravitational-wave sources.

III. BAYESIAN INFERENCE APPLIED TO NON-GAUSSIAN BACKGROUNDS

In this section we specify the various probability distributions, likehood functions, prior and posterior distributions, etc. that we will need in order to apply Bayesian inference to searches for non-Gaussian gravitational-wave backgrounds. Readers interested in more details regarding some of the calculations performed in this section should consult e.g., [19].

A. Noise and signal probability distributions

For simplicity, consider the simple case of N samples of data in a pair of co-located and co-aligned detectors:

$$\mathbf{s}_1 = \mathbf{n}_1 + \mathbf{h}\,, \quad \mathbf{s}_2 = \mathbf{n}_2 + \mathbf{h}\,, \tag{2}$$

where $\mathbf{s}_1 = [s_{11}, s_{12}, \cdots s_{1N}]^T$, etc. We will assume that the noise in each detector is Gaussian, white, and independent of one another, with zero mean and variance σ_1^2, σ_2^2:

$$p_n(\mathbf{n}|\vec{\theta}_n) = \frac{1}{\sqrt{\det(2\pi\mathbf{C}_n)}}\, e^{-\frac{1}{2}\,\mathbf{n}^T\mathbf{C}_n^{-1}\mathbf{n}}\,, \tag{3}$$

where

$$\mathbf{n} = \left[\begin{array}{c} \mathbf{n}_1 \\ \mathbf{n}_2 \end{array}\right], \quad \mathbf{C}_n = \left[\begin{array}{cc} \sigma_1^2\,\mathbb{1}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \sigma_2^2\,\mathbb{1}_{N\times N} \end{array}\right], \tag{4}$$

and $\vec{\theta}_n = \{\sigma_1, \sigma_2\}$. The signal $\mathbf{h}$, which is common to both detectors, is assumed to come from a probability distribution $p_h(\mathbf{h}|\vec{\theta}_h)$, which need not be Gaussian. The probability distribution $p_h(\mathbf{h}|\vec{\theta}_h)$ is called a *parameterized signal prior* and $\vec{\theta}_h$ are called *hyperparameters* [20, 21]. Examples of parameterized signal priors include:

(i) *Gaussian, white signal prior*:

$$p_h(\mathbf{h}|\vec{\theta}_h) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_h^2}} e^{-h_i^2/2\sigma_h^2}\,, \tag{5}$$

where $\vec{\theta}_h = \{\sigma_h\}$.

(ii) *Two-component Gaussian, white signal prior*:

$$p_h(\mathbf{h}|\vec{\theta}_h) = \prod_{i=1}^{N} \left[\xi\, \frac{1}{\sqrt{2\pi\alpha^2}}\, e^{-h_i^2/2\alpha^2} + (1-\xi)\, \frac{1}{\sqrt{2\pi\beta^2}}\, e^{-h_i^2/2\beta^2}\right], \tag{6}$$

where $\vec{\theta}_h = \{\xi, \alpha, \beta\}$. The two-component Gaussian signal prior reduces to the Gaussian signal prior in the limit $\xi \to 1$. It reduces to the Drasco and Flanagan signal prior [8] in the limit $\beta \to 0$. The Drasco and Flanagan signal prior corresponds to Gaussian bursts with root-mean-square (rms) amplitude α and probability $0 \le \xi \le 1$.

(iii) *Non-standardized Student's t-distribution signal prior*:

$$p_h(\mathbf{h}|\vec{\theta}_h) = \prod_{i=1}^{N} \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\alpha\,\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}} \left(1 + \frac{1}{\nu}\frac{h_i^2}{\alpha^2}\right)^{-\frac{\nu+1}{2}} \right] , \quad (7)$$

where

$$\Gamma(\nu) = \int_0^\infty dx\, x^{\nu-1} e^{-x} \quad (8)$$

is the Gamma function and $\vec{\theta}_h = \{\nu, \alpha\}$. The above distribution is an extension of the standard Student's t-distribution, which includes a scaling parameter α in addition to the number of degrees of freedom, $\nu > 0$ (real). (The t of Student's t-distribution is given by h_i/α.) The scaling parameter α is related to the variance of each h_i by

$$\sigma_h^2 = \alpha^2 \frac{\nu}{\nu - 2} , \quad \text{for } \nu > 2 . \quad (9)$$

For $\nu \to \infty$, the non-standardized Student's t-distribution becomes a Gaussian distribution with the above variance.

(iv) *Multi-sinusoid signal prior*:

$$p_h(\mathbf{h}|\vec{\theta}_h) = \delta\left(\mathbf{h} - \mathbf{h}(\vec{\theta}_h)\right) , \quad (10)$$

$$h_i(\vec{\theta}_h) = \sum_{I=1}^{M} A_I \cos(2\pi f_I t_i - \varphi_I) , \quad (11)$$

where $i = 1, 2, \cdots, N$ and $\vec{\theta}_h = \{A_I, f_I, \varphi_I | I = 1, 2, \cdots, M\}$. Here M can take on any value between 0 and $M_{\max}$, where $M_{\max}$ is the maximum number of allowed sinusoids (e.g., $M_{\max} = 100$). This is a *deterministic* signal model corresponding to the superposition of M individually resolvable sinusoids.

Although it is possible to write down more complicated non-Gaussian signal priors (since there are an infinite number of ways for a signal to be non-Gaussian), for the analysis considered in this paper, we will restrict ourselves to those given above.

B. Likelihood functions

To construct the likelihood function, we first adopt a waveform template $\mathbf{h}$ and form the residuals $\mathbf{r}_1 = \mathbf{s}_1 - \mathbf{h}$ and $\mathbf{r}_2 = \mathbf{s}_2 - \mathbf{h}$. We demand that the residuals be consistent with the probability distribution for the noise (cf. (3)), which gives rise to a multivariate Gaussian likelihood function for the data:

$$p(\mathbf{s}|\vec{\theta}_n, \mathbf{h}) \equiv p_n(\mathbf{r}|\vec{\theta}_n) = \frac{1}{\sqrt{\det(2\pi\mathbf{C}_n)}}\, e^{-\frac{1}{2}\,\mathbf{r}^T\mathbf{C}_n^{-1}\mathbf{r}} , \quad (12)$$

where

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \end{bmatrix} , \quad \mathbf{r} = \begin{bmatrix} \mathbf{s}_1 - \mathbf{h} \\ \mathbf{s}_2 - \mathbf{h} \end{bmatrix} . \quad (13)$$

But since $\mathbf{h}$ are random variables for stochastic signal models or specified functions of the parameters $\vec{\theta}_h$ for deterministic signal models, we are not interested in the particular values of $\mathbf{h}$, but rather in the values of the parameters $\vec{\theta}_h$ that define the signal prior $p_h(\mathbf{h}|\vec{\theta}_h)$. We thus *marginalize* over $\mathbf{h}$ by performing the integral:

$$p(\mathbf{s}|\vec{\theta}) \equiv p(\mathbf{s}|\vec{\theta}_n, \vec{\theta}_h) = \int d\mathbf{h}\, p(\mathbf{s}|\vec{\theta}_n, \mathbf{h})\, p_h(\mathbf{h}|\vec{\theta}_h) . \quad (14)$$

Here $\vec{\theta} \equiv \{\vec{\theta}_n, \vec{\theta}_h\}$ denotes the combined set of noise and signal parameters.

(i) For the Gaussian signal prior, we find:

$$p(\mathbf{s}|\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C}_s)}}\, e^{-\frac{1}{2}\,\mathbf{s}^T\mathbf{C}_s^{-1}\mathbf{s}} , \quad (15)$$

where

$$\mathbf{C}_s = \mathbf{C}_n + \sigma_h^2 \begin{bmatrix} \mathbb{1}_{N\times N} & \mathbb{1}_{N\times N} \\ \mathbb{1}_{N\times N} & \mathbb{1}_{N\times N} \end{bmatrix} . \quad (16)$$

The likelihood function given by (15) and (16) has the standard form used as the starting point for cross-correlation analyses for Gaussian stochastic backgrounds [18, 22].

(ii) For the two-component Gaussian signal prior, we obtain a two-component Gaussian distribution for the marginalized likelihood, with covariance matrices similar to (16), but with σ_h^2 replaced by α^2 and by β^2 for the two components, respectively:

$$p(\mathbf{s}|\vec{\theta}) = \xi \frac{1}{\sqrt{\det(2\pi\mathbf{C}_\alpha)}}\, e^{-\frac{1}{2}\,\mathbf{s}^T\mathbf{C}_\alpha^{-1}\mathbf{s}} + (1-\xi)\frac{1}{\sqrt{\det(2\pi\mathbf{C}_\beta)}}\, e^{-\frac{1}{2}\,\mathbf{s}^T\mathbf{C}_\beta^{-1}\mathbf{s}} , \quad (17)$$

where

$$\mathbf{C}_\alpha = \mathbf{C}_n + \alpha^2 \begin{bmatrix} \mathbb{1}_{N\times N} & \mathbb{1}_{N\times N} \\ \mathbb{1}_{N\times N} & \mathbb{1}_{N\times N} \end{bmatrix} \quad (18)$$

and similarly for $\mathbf{C}_\beta$.

(iii) For the non-standardized Student's t-distribution, the marginalization integrals are all of the form:

$$\int_{-\infty}^{\infty} dh_i\, e^{-\frac{1}{2}\frac{(s_{1i}-h_i)}{\sigma_1^2}} e^{-\frac{1}{2}\frac{(s_{2i}-h_i)}{\sigma_2^2}} \left(1 + \frac{1}{\nu}\frac{h_i^2}{\alpha^2}\right)^{-\frac{\nu+1}{2}} . \quad (19)$$

Unfortunately, we do not know how to analytically evaluate such an integral. It is possible to consider an Edgeworth expansion of the Student t-distribution in

terms of its non-zero cumulants, $c_2, c_4, \cdots$. But then truncating the expansion after a finite number of terms would produce a *different* non-Gaussian distribution, that would behave differently in model comparision tests from the full Student's t-distribution. Thus, if we want to use this distribution as one of our non-Gaussian signal models, we would need to evaluate the above integrals numerically.

(iv) For the deterministic multi-sinusoid signal model, the marginalized likelihood is simply

$$p(\mathbf{s}|\vec{\theta}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C}_n)}}\, e^{-\frac{1}{2}\,(\mathbf{s}-\mathbf{h}(\vec{\theta}_h))^T\mathbf{C}_n^{-1}(\mathbf{s}-\mathbf{h}(\vec{\theta}_h))}\,. \quad (20)$$

C. Posterior distributions and Bayesian model selection

The posterior distribution for a *subset* of the noise and signal parameters $\vec{\theta} \equiv \{\vec{\theta}_n, \vec{\theta}_h\}$ is obtained by marginalizing over the other noise and signal parameters. For example, for the Gaussian signal model, the posterior distribution for the signal parameter σ_h is obtained by evaluating the following integral:

$$p(\sigma_h|\mathbf{s}) = \int d\sigma_1 \int d\sigma_2\, p(\sigma_1, \sigma_2, \sigma_h|\mathbf{s})\,. \quad (21)$$

Similar integrals will give the posterior distributions for σ_1 and σ_2.

In a similar manner, we can calculate the posterior probability distribution for a signal+noise model $\mathcal{M}$ using Bayes' theorem in the form

$$p(\mathcal{M}|\mathbf{s}) = \frac{p(\mathbf{s}|\mathcal{M})\pi(\mathcal{M})}{\sum_I p(\mathbf{s}|\mathcal{M}_I)\pi(\mathcal{M}_I)}\,. \quad (22)$$

The quantity $p(\mathbf{s}|\mathcal{M})$ is called the *evidence* for model $\mathcal{M}$. It is just the likelihood function $p(\mathbf{s}|\vec{\theta}, \mathcal{M})$ marginalized over the parameter values

$$p(\mathbf{s}|\mathcal{M}) = \int d\vec{\theta}\, p(\mathbf{s}|\vec{\theta}, \mathcal{M})\pi(\vec{\theta}|\mathcal{M})\,, \quad (23)$$

where we have explicitly indicated the model dependence of both the prior and likelihood function.

To compare two models $\mathcal{M}_I$ and $\mathcal{M}_J$, we simply take the ratio of the posterior probability distributions for these two models:

$$\frac{p(\mathcal{M}_I|\mathbf{s})}{p(\mathcal{M}_J|\mathbf{s})} = \frac{p(\mathbf{s}|\mathcal{M}_I)\pi(\mathcal{M}_I)}{p(\mathbf{s}|\mathcal{M}_J)\pi(\mathcal{M}_J)}\,. \quad (24)$$

Note that the the common factor $\sum_I p(\mathbf{s}|\mathcal{M}_I)\pi(\mathcal{M}_I)$ has canceled out when forming the ratio. The left-hand side of the above equation is the *posterior odds ratio* for model $\mathcal{M}_I$ relative to $\mathcal{M}_J$; we see from this equation that it equals the prior odds ratio times the ratio of the evidences. This ratio of evidences is called the *Bayes factor* and is denoted by

$$\mathcal{B}_{IJ}(\mathbf{s}) \equiv \frac{p(\mathbf{s}|\mathcal{M}_I)}{p(\mathbf{s}|\mathcal{M}_J)}\,. \quad (25)$$

In many circumstances there is no a priori reason to prefer one model over another (i.e., the prior odds ratio is unity), so for these cases the posterior odds ratio is just the Bayes factor. If we fix some model, e.g., $\mathcal{M}_0$, and calculate the Bayes factors of all the other models relative to $\mathcal{M}_0$, the model with the largest Bayes' factor is the preferred model given the data.

$\mathcal{B}_{IJ}(\mathbf{s})$	$2\ln\mathcal{B}_{IJ}(\mathbf{s})$	Evidence for model $\mathcal{M}_I$ relative to $\mathcal{M}_J$
< 1	< 0	Negative (supports model $\mathcal{M}_J$)
1–3	0–2	Not worth more than a bare mention
3–12	2–5	Positive
12–150	5–10	Strong
> 150	> 10	Very strong

TABLE I: Bayes factors and their interpretation in terms of the evidence in favor of one model relative to the other.

Table I gives a list of possible Bayes factor values and their interpretation in terms of the evidence in favor of one model relative to another. The interpretation is based on betting odds, and the precise level at which one considers the evidence for a model to be strong or very strong is quiet subjective.

D. Comparison to maximum-likelihood analyses

It is interesting to compare the Bayesian model selection calculation discussed above to a maximum-likelihood frequentist analysis, e.g., that presented in [8]. There they construct a detection statistic by maximizing the likelihood ratio for a signal+noise model $\mathcal{M}_1$ to the noise-only model $\mathcal{M}_0$:

$$\Lambda_{\rm ML}(\mathbf{s}) \equiv \frac{\max_{\vec{\theta}_n}\max_{\vec{\theta}_h} p(\mathbf{s}|\vec{\theta}_n, \vec{\theta}_h, \mathcal{M}_1)}{\max_{\vec{\theta}'_n} p(\mathbf{s}|\vec{\theta}'_n, \mathcal{M}_0)}\,. \quad (26)$$

The Bayes factor calculation also involves a ratio of two quantities, but instead of *maximizing* over the parameters, we *marginalize* over the parameters:

$$\mathcal{B}_{10}(\mathbf{s}) \equiv \frac{p(\mathbf{s}|\mathcal{M}_1)}{p(\mathbf{s}|\mathcal{M}_0)} \quad (27)$$

$$= \frac{\int d\vec{\theta}_n \int d\vec{\theta}_h\, p(\mathbf{s}|\vec{\theta}_n, \vec{\theta}_h, \mathcal{M}_1)\pi(\vec{\theta}_n, \vec{\theta}_h|\mathcal{M}_1)}{\int d\vec{\theta}'_n\, p(\mathbf{s}|\vec{\theta}'_n, \mathcal{M}_0)\pi(\vec{\theta}'_n|\mathcal{M}_0)}\,. \quad (28)$$

These two expressions can be related to one another by using the Laplace approximation to individually approximate the evidences $p(\mathbf{s}|\mathcal{M}_1)$ and $p(\mathbf{s}|\mathcal{M}_0)$. As shown in

App. A,

$$p(\mathbf{s}|\mathcal{M}) \simeq p(\mathbf{s}|\vec{\theta}_{\rm ML})\frac{\Delta V_{\mathcal{M}}}{V_{\mathcal{M}}}\,, \tag{29}$$

where $\Delta V_{\mathcal{M}}/V_{\mathcal{M}}$ is the fraction of the parameter space volume for model $\mathcal{M}$ needed to fit the data, and $\vec{\theta}_{\rm ML}$ are the particular values of the model parameters that maximize the likelihood. Thus,

$$\mathcal{B}_{10}(\mathbf{s}) \simeq \Lambda_{\rm ML}(\mathbf{s})\frac{\Delta V_1/V_1}{\Delta V_0/V_0}\,, \tag{30}$$

which shows that the Bayes factor is proportional to the frequentist maximum-likelihood ratio. The proportionality constant is the Occam's factor mentioned in Sec. II, which penalizes a model if its parameter space volume is larger than necessary to fit the data.

E. Signal-to-noise ratios

One of the parameters that we will use to describe the simulations in Sec. V is the ratio of the power in the injected signals to that of the detector noise. For a stochastic gravitational-wave background described by the one-sided strain power spectral density $S_h(f)$, the expected signal-to-noise ratio of the optimally-filtered cross-correlation statistic in a pair of detectors I, J is given by [23]

$$\left.{\rm SNR}^2\right|_{\rm stoch} = \sqrt{2T}\left[\int_0^\infty df\,\frac{\Gamma_{IJ}^2(f)S_h^2(f)}{P_{n_I}(f)P_{n_J}(f)}\right]^{1/2}\,, \tag{31}$$

where $\Gamma_{IJ}(f)$ is the overlap function between detectors I and J (see, e.g., [24, 25]). For a pair of identical, co-located and co-aligned detectors the above expression simplifies to

$$\left.{\rm SNR}^2\right|_{\rm stoch} = \sqrt{2T}\left[\int_0^\infty df\,\frac{P_h^2(f)}{P_n^2(f)}\right]^{1/2}\,, \tag{32}$$

where $P_h(f) \equiv \Gamma_{II}(f)S_h(f)$ is the gravitational-wave power in a single detector. We are assuming here that the signal power is weak relative to the noise, so that the total power in detector I is given by $P_I(f) \equiv P_h(f)+P_{n_I}(f) \approx P_{n_I}(f)$.

For a deterministic signal described by the strain response $\tilde{h}(f)$, it is often more convenient to work with the *matched-filter* signal-to-noise ratio, which has expected value

$$\left.{\rm SNR}^2\right|_{\rm det} = 4\int_0^\infty df\,\frac{|\tilde{h}(f)|^2}{P_n(f)} = 2T\int_0^\infty df\,\frac{P_h(f)}{P_n(f)}\,. \tag{33}$$

For this case $P_h(f) = \frac{2}{T}|\tilde{h}(f)|^2$ is the one-sided power spectral density for the signal. Note that for deterministic signals, the squared signal-to-noise ratio scales with the number of frequency bins $N_{\rm bins}$, while for stochastic signals it scales like $\sqrt{N_{\rm bins}}$.

IV. SIGNAL+NOISE MODELS AND PRIORS

We consider the following five models for describing the signal and noise:

$\mathcal{M}_0$ - Noise-only model:

This is a noise-only model, which assumes uncorrelated, white Gaussian noise in the two detectors. There are only two parameters for this model, $\vec{\theta} = \{\sigma_1, \sigma_2\}$. For our simulations, the prior on the noise variances are flat between 0 and 10.

$\mathcal{M}_1$ - Noise plus Gaussian stochastic model:

White Gaussian detector noise plus a white Gaussian gravitational-wave background. There is one additional parameter corresponding to the variance σ_h^2 of the background, so $\vec{\theta} = \{\sigma_1, \sigma_2, \sigma_h\}$. The prior on σ_h^2 is also between 0 and 10, just like the detector noise variances.

$\mathcal{M}_2$ - Noise plus non-Gaussian (two-component) stochastic model:

White Gaussian detector noise plus a white two-component Gaussian model for the gravitational-wave background. There are three parameters for the two-component Gaussian model: the variances α^2 and β^2 for the two components, and the probability ξ of one of the components. (The probability of the other component necessarily equals $1-\xi$.) Thus, $\vec{\theta} = \{\sigma_1, \sigma_2, \alpha, \beta, \xi\}$. The prior on ξ is flat from 0 to 1. The prior on the variances are 0 to 10 for the wide component and 0 to 0.5 on the narrow, delta-function-like component.

$\mathcal{M}_3$ - Noise plus deterministic multi-sinusoid signal model:

White Gaussian detector noise plus up to 100 deterministic sinusoids. There are three parameters $\{A_I, f_I, \varphi_I\}$ corresponding to the amplitude, frequency, and phase for each sinusoid. Thus, for M sinusoids, there are $2+3M$ pameters for this particular model $\vec{\theta} = \{\sigma_1, \sigma_2, A_I, f_I, \varphi_I | I = 1, 2, \cdots M\}$. The prior on the amplitudes is uniform in the range $A \in [0, 1000]$, and the prior on the frequencies is uniform across the range spanned by the data. The prior on the phases is uniform between 0 and 2π.

$\mathcal{M}_4$ - Noise plus deterministic multi-sinusoid plus Gaussian background model:

White Gaussian detector noise plus a Gaussian gravitational-wave background plus up to 100 sinusoids. As for $\mathcal{M}_3$, there are three parameters (amplitude, frequency, and phase) for each sinusoid. Thus, for M sinusoids, there are $2+1+3M$ parameters for this model $\vec{\theta} = \{\sigma_1, \sigma_2, \sigma_h, A_I, f_I, \varphi_I | I = 1, 2, \cdots M\}$. The priors on the parameters are the same as in the previous models. This hybrid model

allows us to effectively 'subtract out' any sufficiently bright sinusoidal signals in the data.

Note that we do not consider a hybrid "noise plus deterministic multi-sinusoid plus *non-Gaussian* background" model in the above list, as we expect the subtraction of the bright sinusoids to remove most of the non-Gaussianity of the signal component. Also, as we shall discuss further in Sec. V, we do not consider a signal+noise model with a non-standardized Student's t-distribution for the non-Gaussian stochastic gravitational-wave component. This is because of the computational costs associated with the marginalized likelihood evaluations (see Eq. (19)), which are needed for the Bayesian model selection calculations.

V. SIMULATIONS

A. Astrophysical source populations

Simulated data is generated by co-adding sinusoidal signals with amplitudes drawn from one of three astrophysical models. The frequencies and phases of the sinusoids are drawn uniformly from the prior ranges defined in the previous section. Gaussian-distributed noise with a white power spectrum is then added to the signals. The amplitude of the signals is scaled so as to produce a pre-specified matched-filter signal-to-noise ratio (SNR) per frequency bin, calculated as an average across all frequency bins, using Eq. (33).

Figure 1 is a plot of the squared amplitude of the noise and signal components for a typical simulation using two detectors. For an SNR-per-bin of 1, the amplitudes of the astrophysicals signals are of the same order-of-magnitude as the noise in the two detectors, as can be seen in the figure.

We considered three astrophysical source models: Model $\mathcal{S}_0$ uniformly distributes standard sirens (sources with the same intrinsic amplitude) in space out to some cutoff radius $r = R$, after which the density falls-off exponential with an e-folding scale of $0.25R$. Model $\mathcal{S}_1$ distributes standard sirens with a Gaussian distribution in distance with density $\rho \propto e^{-r^2/2R^2}$. For model $\mathcal{S}_0$ the number of sources in a spherical shell of radius r is proportional to r^2 out to $r = R$. For model $\mathcal{S}_1$ the number of sources in a spherical shell of radius r is proportional to the product $r^2 e^{-r^2/2R^2}$, and thus has a larger number of sources at smaller r, as compared to the uniform distribution case. Model $\mathcal{S}_2$ is based on a population synthesis model for supermassive black hole binaries [26], where the amplitude of the sources depends on both the mass of the system and the distance. The usual frequency dependence of the amplitude was artificially suppressed so as to produce a white spectrum.

The amplitude distributions for the three models are shown in Fig. 2. Models $\mathcal{S}_0$ and $\mathcal{S}_1$ have similar amplitude distributions that are fairly tightly peaked, while model $\mathcal{S}_2$ has a large tail extending to high amplitude.

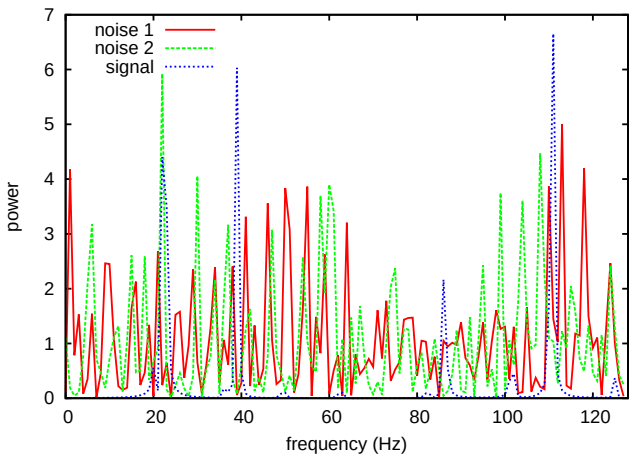


FIG. 1: The squared amplitude of the noise and signal components for data in two coincident and coaligned detectors consisting of white noise and a superposition of sinusoids drawn from astrophysical source population $\mathcal{S}_2$, with a source density of 0.1/bin in 128 frequency bins and an average SNR-per-bin of 1. The SNR is dominated here by the four brightest sources.

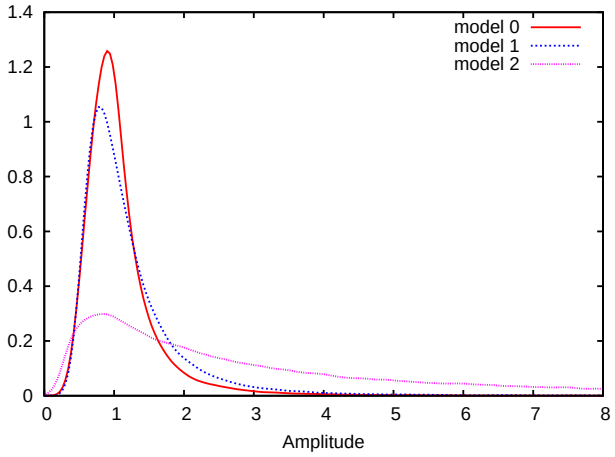


FIG. 2: Amplitude distributions for the three astrophysical source distributions considered in this study. The amplitude scale is arbitrary since the signal-to-noise ratios are set when producing simulated data sets drawn from these distributions. Models $\mathcal{S}_0$ and $\mathcal{S}_1$ are for "standard sirens" (equal intrinsic amplitude sources) with different spatial distributions, while model $\mathcal{S}_2$ is based on a population synthesis model where some sources have much higher intrinsic amplitudes than others.

B. Markov Chain Monte Carlo methods

We performed two types of analyses, both of which employed trans-dimensional Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithms. The first type of analysis looked at the signal in a single detector with no instrument noise. There the goal was to find which of three statistical models best described the intrinsic properties of the signal: a Gaussian distribution, a two-component Gaussian distribution; or a non-standardized Student's t-distribution. A RJMCMC analysis extends the usual MCMC exploration of the parameters of a single model to the exploration of a range of models and their parameters, thus allowing us to produce marginalized posteriors for both the model parameters *and* posterior distributions for the relative probability that each model is consistent with the data and our prior knowledge. In principle, a single RJMCMC routine could explore all three probability distributions at once, but we were able to achieve better mixing by performing pair-wise comparisons between the Gaussian and two-component Gaussian models and the Gaussian and the non-standardized Student's t model. The ratio of the number of iterations the Markov chain spends in each model yields Bayes factors between the Gaussian reference model and the two non-Gaussian alternatives.

The second type of analysis considered the detection and characterization of the astrophysical signals in the presence of detector noise in a two-detector network. Here we considered the five models described in Sec. IV. Once again, a single RJMCMC routine could simultaneously explore all five models, but achieving efficient mixing between models with different parameterizations and dimensionality is notoriously difficult. Instead, we again opted for a pairwise approach, comparing the noise-only model $\mathcal{M}_0$ to each of the four signal+noise models in turn. This yields a collection of Bayes factors between the reference noise model and the four signal models. It is important to note that models $\mathcal{M}_3$ and $\mathcal{M}_4$ are both complicated composite models that allow for a variable number of sinusoids to be used in the model. Model $\mathcal{M}_4$ further allows for, but does not require, a Gaussian signal component. Thus $\mathcal{M}_4$ contains models $\mathcal{M}_3$, $\mathcal{M}_1$ and $\mathcal{M}_0$ as sub-cases. If we had included a two-component Gaussian in $\mathcal{M}_4$, then we would have been able to explore all four signal models at once. We did check that the relative probabilities for the sub-models included in $\mathcal{M}_4$ were consistent with the relative probabilities found in the pair-wise comparisons, though the larger model space did lead to larger uncertainties on the Bayes factors. The uncertainties were computed from the variance of the running Bayes factors, and compared to analytic estimates based on the number of transitions between the models [17]. Both methods yielded consistent error estimates. We further checked that the error estimates were consistent with the spread seen when repeating the analysis dozens of times with different random number seeds.

C. Classifying the signals

We begin looking at the statistical properties of the signals themselves. While this is not something we can do with actual observations where the signals must be extracted from noisy data, it is interesting to compare the observed properties of the signals to the intrinsic properties of the signals.

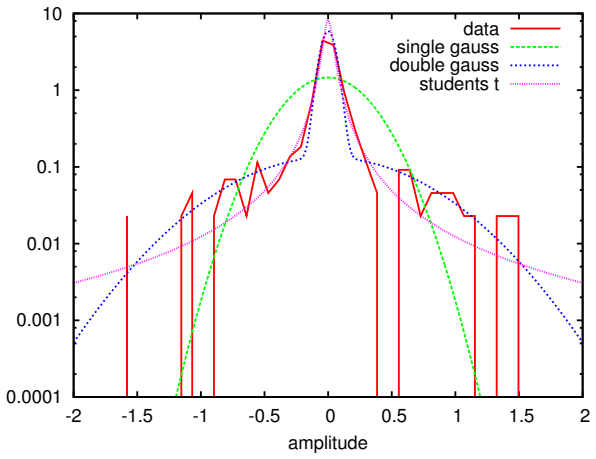


FIG. 3: Histogram of signal samples and the corresponding best fit single-Gaussian, double-Gaussian, and non-standardized Student's t-distribution for a signal consisting of a superposition of sinusoids drawn from an astrophysical population with a source density of 0.1/bin in 256 frequency bins and an SNR-per-bin of 1.

Figure 3 is a histogram of signal samples as well as the best fit Gaussian ("single gauss"), two-component Gaussian ("double gauss"), and non-standardized Student's t-distribution for a simulated signal with an average density of one source per ten frequency bins. As expected for such a sparse population, the single-Gaussian fit is extremely poor compared to the two-component Gaussian or non-standardized Student's t-distibution fit.

Figure 4 is a plot of Bayes factor quantile intervals as a function of the total number of frequency bins, comparing the two-component Gaussian and non-standardized Student's t-distribution models to the reference Gaussian model. The source density was set to 10/bin for these simulations. The two panels correspond to astrophysical source models $\mathcal{S}_0$ and $\mathcal{S}_1$. In the upper panel the astrophysical sources were drawn from source model $\mathcal{S}_0$. In the lower panel, the astrophysical sources were drawn from source model $\mathcal{S}_1$. There was no detector noise in these simulations. Note that the Bayes factors in the lower panel are shifted slightly higher relative to those in the upper panel, consistent with the expectation that the Gaussian-distributed astrophysical source population will tend to produce closer—and hence more-easily resolvable—sources. We see that at this relatively high source density, we need a large amount of data (many fre-

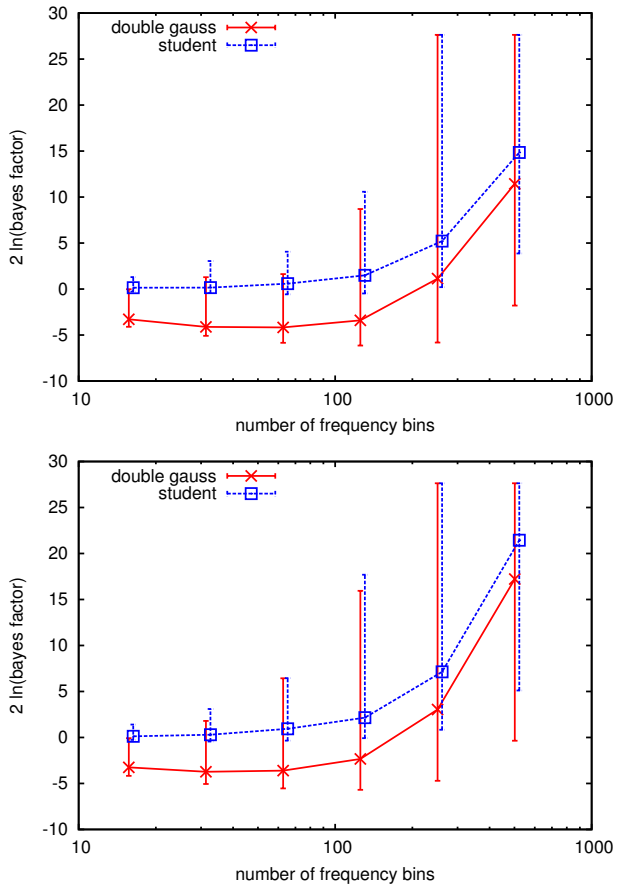


FIG. 4: Bayes factor 80% quantile intervals for the two-component Gaussian and non-standardized Student's t-distribution signal models as a function of the total number of frequency bins. The source density was set to 10/bin for all the simulations. In the upper panel the astrophysical sources were drawn from the source model $\mathcal{S}_0$. In the lower panel the astrophysical sources were drawn from source model $\mathcal{S}_1$.

quency bins) to detect the subtle departure from Gaussianity.

We should emphasize that the quantile intervals shown in Fig. 4 (and in several other figures to follow) define the probability distribution for the Bayes factor values as estimated from 256 independent realizations of the simulated signal and noise for each set of parameter values: *these are not error bars on the individual Bayes factors*. For a single realization of the simulated signal and noise, the uncertainty in the value of the Bayes factor as estimated from 128 independent Monte Carlo simulations is $\lesssim 10\%$, which we can ignore in the quantile plots.

Figure 5 is a similar plot of Bayes factor quantile intervals as a function of the number of sources per bin, comparing the two-component Gaussian and non-standardized Student's t-distribution. The total number of bins was set to 128 and the signals were drawn from astrophysical source model $\mathcal{S}_0$. As expected from the central limit theorem, the simulated data is consistent with a Gaussian probability distribution when the average number of signals per frequency bin is large. It is interesting to note the large spread in the Bayes factors in the transition region between 1 and 10 sources per bin. This tells us that some realizations look Gaussian, while others look highly non-Gaussian.

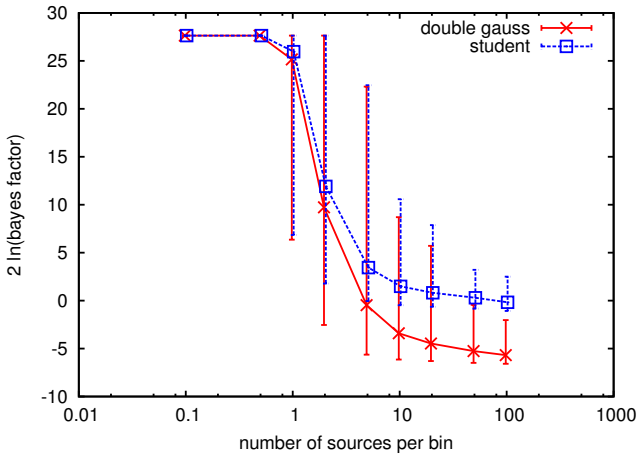


FIG. 5: Bayes factor 80% quantile intervals for the two-component Gaussian and non-standardized Student's t-distribution signal models as a function of the total number of sources per bin. The total number of bins was set to 128 and the astrophysical sources were drawn from source model $\mathcal{S}_0$.

Similar to the results shown in Fig. 4, the non-standardized Student's t-distribution model has consistently higher Bayes factors than the two-component Gaussian model. This suggests using it over the two-component Gaussian model when modeling non-Gaussian stochastic signals. However, the fact that we are not able to find an analytic expression for the corresponding marginalized likelihood function (see Eq. (19)) means that the Student's t-distribution model has a much higher computational cost than the two-component Gaussian model. As such, for all subsequent model comparison simulations that we do—which include simulated noise in a two-detector network—we use the two-component Gaussian stochastic model instead of the more expensive non-standardized Student's t-distribution model.

D. Detecting and characterizing signals in noisy detector data

Next we turn our attention to the observed properties of the signals in a more realistic setup that includes instrument noise and a network with two co-aligned and co-located detectors. A multi-detector analysis is needed to distinguish signals from noise. In this study we need to consider the dependence on SNR, in addition to the

dependence on source density and data volume (number of frequency bins). In the infinite SNR limit we recover the pure signal analysis described in the previous section. For more realistic SNRs, a signal that is best described as deterministic or non-Gaussian in the absence of noise may favor a stochastic or Gaussian description in the presence of noise when model simplicity wins out over model fidelity.

In addition to the Gaussian and two-component Gaussian signal models ($\mathcal{M}_1$ and $\mathcal{M}_2$), we additionally consider a deterministic model made up of the sum of sinusoids ($\mathcal{M}_3$), and a hybrid model with a Gaussian-stochastic component and a collection of sinusoids ($\mathcal{M}_4$). Figure 6 compares the cross-correlated data in two detectors to a marginalized posterior distribution for the frequencies used by the multi-component sinusoid model. In this instance, the brightest sinusoid in the data was confidently detected, as indicated by the large peak in the posterior distribution. The second and third brightest signals in the data were marginally detected, as indicated by the secondary peaks in the posterior distribution.

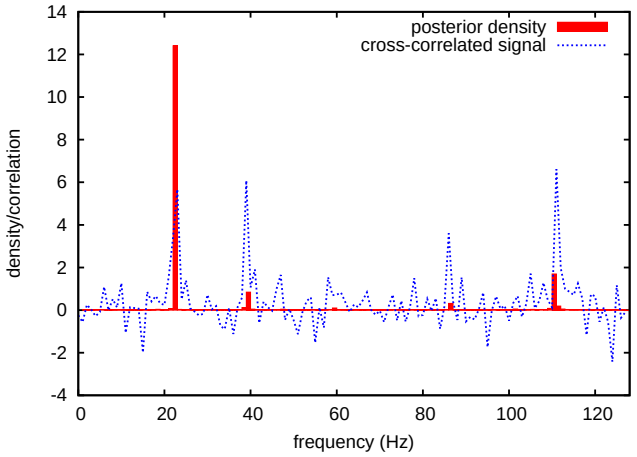


FIG. 6: The cross-correlated signal+noise in the two detectors for the simulated data shown in Fig. 1 is compared to the scaled posterior density for the frequencies of the sinusoids found by a trans-dimensional MCMC analysis of the data. The three brightest signals in the data had amplitudes and frequencies ($A = 4.46$, $f = 22.45$ Hz), ($A = 4.41$, $f = 110.66$ Hz) and ($A = 3.73$, $f = 39.23$ Hz). Only the brightest of these was a clear detection, though the analysis did occasionally lock onto the other signals.

The question of whether the data is best described by a deterministic model, a non-Gaussian stochastic model or a Gaussian-stochastic model depends on many factors, including the source density, the noise level, the number of frequency bins and the SNR-per-bin. In what follows we explore the impact of each of these factors.

A major challenge that we face is that the outcomes vary greatly from one simulation to the next, especially in the limit that there are few sources and/or few frequency bins. To counter this we performed a large number of simulations and aggregated the results. When showing Bayes factors between the various signal models and the noise model, we display the mean values along with the 80% quantile intervals derived from the ensemble of simulations. More directly, we also report the fraction of times each model had the highest Bayesian evidence on a realization-by-realization basis. While the general trend is that the Gaussian model is more likely to be favored as the number of sources per bin increases, the deterministic model can sometimes be preferred at high source density, and the Gaussian-stochastic model can sometimes be preferred at low source density. Note that while ground and space-based interferometric detectors and pulsar timing arrays nominally cover much larger frequency bands than we consider here, their "V" shaped sensitivity curves limit the effective number of

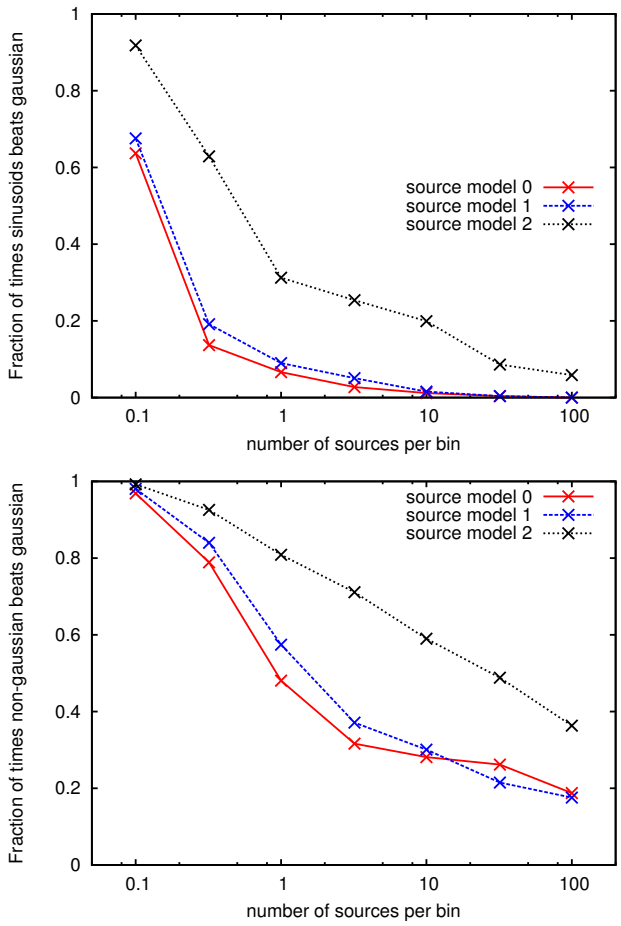


FIG. 7: The fraction of times a non-Gaussian model has higher evidence than the Gaussian model as a function of the source density for the three source models. For these simulations the number of bins was fixed at 32 and the average SNR-per-bin was fixed at 2. Upper panel: The fraction of times the deterministic, multi-sinusoid model $\mathcal{M}_3$ had higher evidence than the Gaussian-stochastic model $\mathcal{M}_1$. Lower panel: The fraction of times the stochastic two-component Gaussian model $\mathcal{M}_2$ had higher evidence than the Gaussian-stochastic model $\mathcal{M}_1$.

frequency bins to 100's for interferometers and 10's for pulsar timing arrays.

Figure 7 shows the fraction of times that a non-Gaussian model has higher evidence than the Gaussian model as a function of the source density for the three source models. Here the number of bins was fixed at 32 and the average SNR-per-bin was fixed at 2. The general trend is as expected from the central limit theorem—as the number of signals per frequency bin grows the data looks less deterministic and more Gaussian. The more realistic source model $\mathcal{S}_2$, which has a variety of intrinsic source luminosities, was consistently less Gaussian than source models $\mathcal{S}_0$ and $\mathcal{S}_1$, which assumed equal luminosity sources. Even at high source densities, model $\mathcal{S}_2$ could appear deterministic or non-Gaussian.

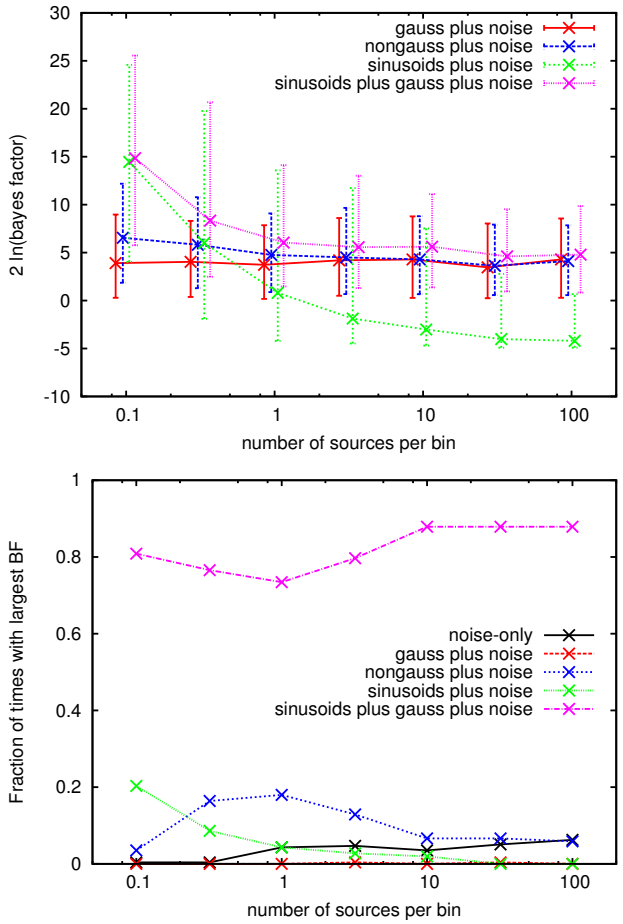
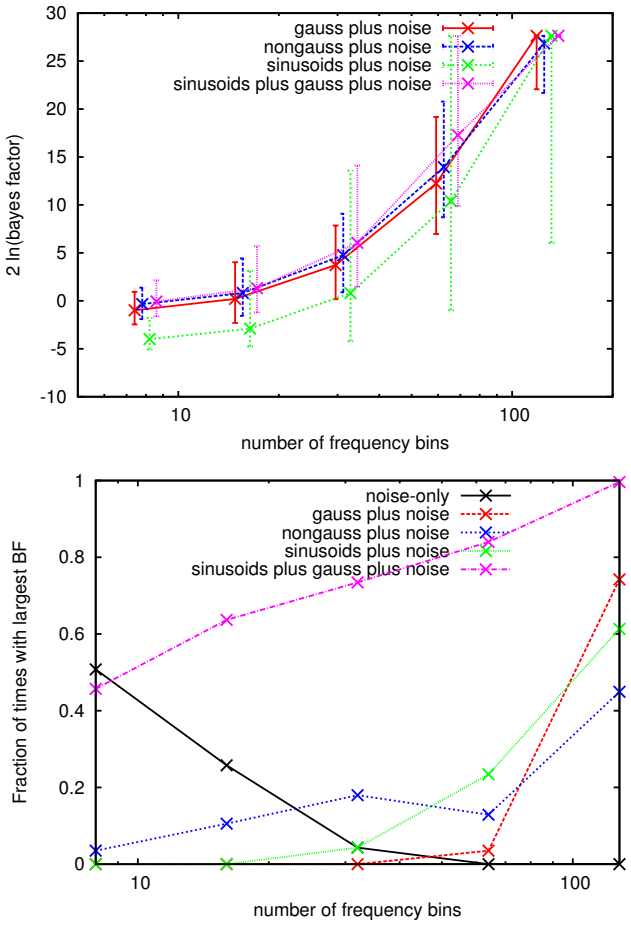


FIG. 8: Upper panel: Bayes factor 80% quantile intervals for the four different signal+noise models relative to the noise-only model as a function of the number of sources per bin. The total number of bins was set to 32 for these simulations, and the astrophysical sources were drawn from source model $\mathcal{S}_2$. The SNR-per-bin was fixed at 2, with different realizations of noise used for the different simulations. Lower panel: Fraction of time that the different models had the largest Bayes factor for the different simulations.

Figure 8 extends the study of the dependence on the source density to include the full set of signal models, and in addition to showing the fraction of time that each model is favored, also shows the Bayes factor quantile intervals for the four signal models. The total number of bins was set to 32 for these simulations, which included simulated noise in addition to the simulated astrophysical signals from source model $\mathcal{S}_2$. For these simulation the SNR-per-bin was fixed at 2, with different realizations of the noise used for the different simulations. Note that for low source densities, the models that include deterministic sinusoid signals are the preferred models. The effectiveness of models having Gaussian or non-Gaussian stochastic signal components improve as the source density increases. As expected, the hybrid model performs best for all source densities.



]

FIG. 9: Upper panel: Bayes factor 80% quantile intervals for the four different signal+noise models relative to the noise-only model as a function of the total number of frequency bins. The source density was set to 1/bin for all the simulations, and the astrophysical sources were drawn from source model $\mathcal{S}_2$. The SNR-per-bin was fixed at 2, with different realizations of noise used for the different simulations. Lower panel: Fraction of time that the different models had the largest Bayes factor for the different simulations.

Figure 9 shows how the model selection results are affected by the number of frequency bins, keeping the source density fixed at one-per-bin and the SNR-per-bin

fixed at 2. As the number of frequency bins increases, the chances of having one or two loud sources dominate the total signal increases, and consequently, the deterministic multi-sinudoid model and the two-component Gaussian stochastic models are more likely to outperform the Gaussian model.

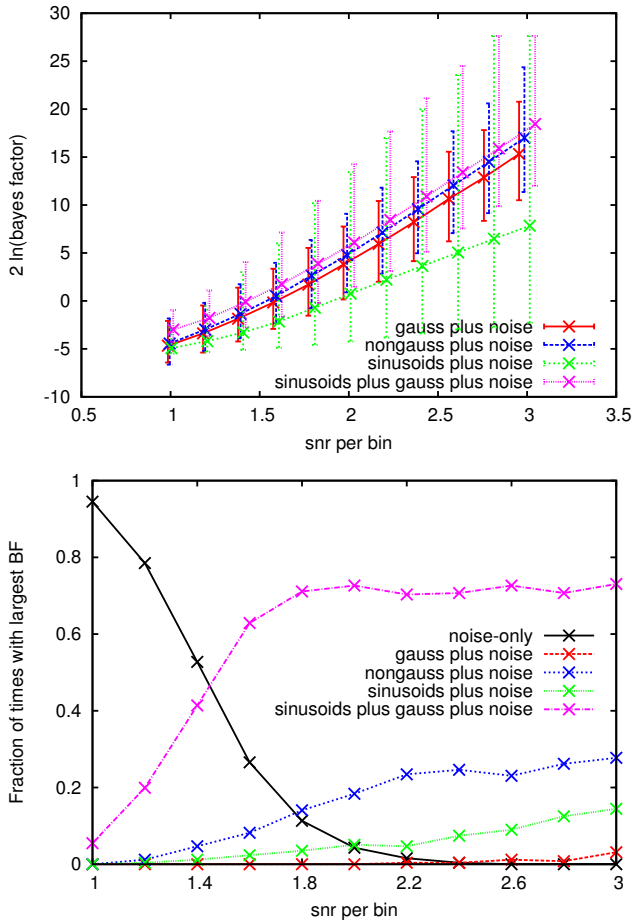


FIG. 10: Upper panel: Bayes factor 80% quantile intervals for the four different signal+noise models relative to the noise-only model as a function of SNR-per-bin. The total number of bins was set to 32 and the source density to 1/bin for all the simulations. The astrophysical sources were drawn from source model $\mathcal{S}_2$. Lower panel: Fraction of time that the different models had the largest Bayes factor for the different simulations.

Finally, Figure 10 shows how the model selection is affected by the SNR-per-bin, keeping the source density fixed at one-per-bin and the number of bins fixed at 32. The Bayes factors for all the signal+noise models increase quadratically with increasing SNR-per-bin, which is to be expected when compared to the noise-only model. For sufficiently large SNR-per-bin (so that the signal is detected by all the models), the relative performance of the various signal models is independent of the SNR-per-bin.

VI. DISCUSSION

We presented a Bayesian search for non-Gaussian gravitational-wave backgrounds. We found that a gravitational-wave signal comprised of the sum of discrete sources drawn from some astrophysical population may be best described as either deterministic, non-Gaussian stochastic, or Gaussian-stochastic depending on the number of sources, and the size of the data set. In our studies the simulated data were produced by adding together multiple sinusoids with amplitudes drawn from one of three astrophysical source distributions, to which was added an independent white noise realization in each detector. While the deterministic signal model $\mathcal{M}_3$, made up of a variable number of sinusoids, is able to precisely match the simulated data, the simpler Gaussian $\mathcal{M}_1$ and non-Gaussian $\mathcal{M}_2$ stochastic signal models are often preferred. The general trend follows our expectation that the signals appear increasingly stochastic and Gaussian as the number of sources per frequency bin increases. We found that departures from Gaussianity are more likely to be detected in large data sets (in our case, for large numbers of frequency bins), but that the ability to distinguish between the various signal models was independent of the SNR (once the signal became detectable). In all cases, a hybrid model, $\mathcal{M}_4$, that combines variable contributions from deterministic and stochastic signals that are determined by the data, outperformed each of the single element models most of the time. This finding may be of particular relevance to the detection of low frequency gravitational waves by pulsar timing arrays.

Although for simplicity we considered co-located and co-aligned detectors and white power spectra, the method can easily be extended to handle more realistic detector geometry (i.e., separated and misaligned detectors) as well as colored noise and signal spectra. For example, a black hole population typically has more bright signals at low frequencies. There are also many other signal and noise models that can be considered.

Acknowledgments

NJC acknowledges support from NSF Award PHY-1306702 and from the NANOGrav Physics Frontier Center, NSF PFC-1430284. JDR acknowledges support from NSF Awards PHY-1205585, CREST HRD-1242090 and the NANOGrav Physics Frontier Center, NSF PFC-1430284.

Appendix A: Bayes factor calculation

Suppose we have two models that we would like to compare, denoted $\mathcal{M}_1$ and $\mathcal{M}_0$, with parameters $\vec{\theta}_1$ and

$\vec{\theta}_0$, respectively. As described in Sec. III C, the Bayes factor $\mathcal{B}_{10}(\mathbf{s})$ for model $\mathcal{M}_1$ relative to model $\mathcal{M}_0$ given observed data $\mathbf{s}$ is defined by

$$\mathcal{B}_{10}(\mathbf{s}) = \frac{p(\mathbf{s}|\mathcal{M}_1)}{p(\mathbf{s}|\mathcal{M}_0)}\,, \tag{A1}$$

where

$$p(\mathbf{s}|\mathcal{M}) = \int d\vec{\theta}\, p(\mathbf{s}|\vec{\theta},\mathcal{M})\pi(\vec{\theta}|\mathcal{M}) \tag{A2}$$

for either model (i.e., $\mathcal{M} = \mathcal{M}_0$ or $\mathcal{M}_1$). The quantity $p(\mathbf{s}|\mathcal{M})$ is called the *evidence* for model $\mathcal{M}$. The Bayes factor is the ratio of the evidences for the two models; it equals the posterior odds ratio for the two models if they have equal a priori probabilities.

Since analytic or direct calculations of the evidence integrals is usually not possible, we need to estimate the Bayes factor numerically. In this appendix, we describe a few of the methods used for the study described in this paper. Readers interested in more details should see Sec. II of [27].

1. Laplace approximation

If we assume that the data is informative, so that the likelihood function is *peaked* relative to the prior probability distribution, then we can use the *Laplace approximation* to estimate the evidence integral (A2):

$$p(\mathbf{s}|\mathcal{M}) \simeq p(\mathbf{s}|\vec{\theta}_{\rm ML},\mathcal{M})\frac{\Delta V_{\mathcal{M}}}{V_{\mathcal{M}}}\,, \tag{A3}$$

where $\vec{\theta}_{\rm ML} \equiv \vec{\theta}_{\rm ML}(\mathbf{s})$ maximizes the likelihood $p(\mathbf{s}|\vec{\theta},\mathcal{M})$ with respect to variations of $\vec{\theta}$; $\Delta V_{\mathcal{M}}$ is the characteristic width of the likehood function around its maximum; and $V_{\mathcal{M}}$ is the total parameter space volume for the model parameters. The ratio $\Delta V_{\mathcal{M}}/V_{\mathcal{M}}$ can be thought of as an Occam's factor, which penalizes a model if its parameter space volume is larger than needed to fit the data. Doing this calculation for both $\mathcal{M}_0$ and $\mathcal{M}_1$, and then taking the ratio of the two results, we find

$$\mathcal{B}_{10}(\mathbf{s}) \simeq \frac{p(\mathbf{s}|\vec{\theta}_{1,\rm ML},\mathcal{M}_1)}{p(\mathbf{s}|\vec{\theta}_{0,\rm ML},\mathcal{M}_0)}\frac{\Delta V_1/V_1}{\Delta V_0/V_0} \tag{A4}$$

$$= \Lambda_{\rm ML}(\mathbf{s})\frac{\Delta V_1/V_1}{\Delta V_0/V_0}\,, \tag{A5}$$

where $\Lambda_{\rm ML}(\mathbf{s})$ is the *maximum-likelihood ratio*.

As a very simple example, consider the case of N samples of data $\mathbf{s}$, consisting of a unknown constant signal in additive white Gaussian-stationary noise with known variance σ^2. Let $\mathcal{M}_0$ denote the *noise-only* model with likelihood function

$$p(\mathbf{s}|\mathcal{M}_0) = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-s_i^2/2\sigma^2}\,, \tag{A6}$$

and let $\mathcal{M}_1$ be the *signal+noise* model defined by the likelihood function

$$p(\mathbf{s}|\theta,\mathcal{M}_1) = \prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(s_i-\theta)^2/2\sigma^2}\,, \tag{A7}$$

and prior $\pi(\theta) = 1/\theta_{\rm max}$, where $\theta \in [0,\theta_{\rm max}]$ is the unknown signal amplitude. Then one can easily show that the maximum-likelihood parameter value is the sample mean

$$\theta_{\rm ML}(\mathbf{s}) = \frac{1}{N}\sum_{i=1}^{N} s_i \equiv \bar{s}\,, \tag{A8}$$

and the Bayes factor for the signal+noise model $\mathcal{M}_1$ relative to the noise-only model $\mathcal{M}_0$ is:

$$\mathcal{B}_{10}(\mathbf{s}) \simeq \frac{\sigma/\sqrt{N}}{\theta_{\rm max}}\exp\left[\frac{1}{2}\frac{\bar{s}^2}{\sigma^2/N}\right]. \tag{A9}$$

It has logarithm

$$2\ln\mathcal{B}_{10}(s) \simeq 2\ln\left(\frac{\sigma/\sqrt{N}}{\theta_{\rm max}}\right) + \frac{\bar{s}^2}{\sigma^2/N}\,. \tag{A10}$$

Since $\bar{\sigma}^2 \equiv \sigma^2/N$ is the variance of the sample mean $\bar{s}$ (or, equivalently, it is the characteristic width of the likelihood function around its maximum value), we see that twice the log of the Bayes factor is effectively the squared SNR of the maximum-likelihood estimator $\theta_{\rm ML}(\mathbf{s})$. The first term on the right-hand side of Eq. (A10) is the Occam's penalty factor associated with the size of the parameter space volume $\theta_{\rm max}$. This term is negative and reduces the value of the log of the Bayes factors if increases N while $\theta_{\rm max}$ and σ are held fixed.

2. Savage-Dicke density ratio

The Savage-Dicke density ratio can be defined whenever model $\mathcal{M}_0$ is a subset of model $\mathcal{M}_1$, and the prior probabilities factorize. Both of these conditions hold, for example, if $\vec{\theta}_1 = \{\vec{\theta}_0, \vec{\theta}_{\rm extra}\}$, with

$$\pi(\vec{\theta}_1|\mathcal{M}_1) = \pi(\vec{\theta}_0|\mathcal{M}_0)\pi(\vec{\theta}_{\rm extra}|\mathcal{M}_1) \tag{A11}$$

and

$$p(\mathbf{s}|\vec{\theta}_0,\mathcal{M}_0) = p(\mathbf{s}|\vec{\theta}_1,\mathcal{M}_1)\big|_{\vec{\theta}_{\rm extra}=\vec{\theta}_{\rm extra,0}} \tag{A12}$$

for some fixed set of parameter values $\vec{\theta}_{\rm extra,0}$. The Savage-Dicke density ratio $r_{10}(\mathbf{s})$ is then defined as

$$r_{10}(\mathbf{s}) \equiv \frac{\pi(\vec{\theta}_{\rm extra,0}|\mathcal{M}_1)}{p(\vec{\theta}_{\rm extra,0}|\mathbf{s},\mathcal{M}_1)}\,, \tag{A13}$$

where $p(\vec{\theta}_{\rm extra,0}|\mathbf{s},\mathcal{M}_1)$ is the marginalized probability density function

$$p(\vec{\theta}_{\rm extra}|\mathbf{s},\mathcal{M}_1) = \int d\vec{\theta}_0\, p(\vec{\theta}_0,\vec{\theta}_{\rm extra}|\mathbf{s},\mathcal{M}_1) \tag{A14}$$

evaluated at $\vec{\theta}_{\rm extra} = \vec{\theta}_{\rm extra,0}$. Using Bayes' theorem in the form

$$p(\vec{\theta}_0, \vec{\theta}_{\rm extra}|\mathbf{s}, \mathcal{M}_1) = \frac{p(\mathbf{s}|\vec{\theta}_0, \vec{\theta}_{\rm extra}, \mathcal{M}_1)\pi(\vec{\theta}_0, \vec{\theta}_{\rm extra}|\mathcal{M}_1)}{p(\mathbf{s}|\mathcal{M}_1)} \tag{A15}$$

and Eqs. (A11) and (A12), one can show that

$$r_{10}(\mathbf{s}) = \mathcal{B}_{10}(s) \tag{A16}$$

exactly. The advantage of using the expression for the Savage-Dicke density ratio to estimate the Bayes factor is that it only requires exploration of the posterior distribution for model $\mathcal{M}_1$.

3. Reversible jump MCMC

Reversible jump, or trans-dimensional MCMC algorithms, explore the space of models in addition to the parameters of each model. The Bayes factor between two models $\mathcal{M}_0$ and $\mathcal{M}_1$ is simply estimated from the ratio of the number of iterations that the chain spends in each model:

$$\mathcal{B}_{10}(\mathbf{s}) = \frac{\text{number of interations in model } \mathcal{M}_1}{\text{number of interations in model } \mathcal{M}_0}\,. \tag{A17}$$

The accuracy of the estimate depends on the number of transitions between the two models—the more transitions, the more accurate the estimate. The problem with this simple approach is that it becomes difficult to compute Bayes factors smaller than 10^{-3} or larger than 10^3, since the chains spend very little time in the disfavoured model, and hence the exploration of that model can fail to converge to the stationary state. Ideally, we would like the chain to spend an equal amount of time in each model, so that all models are explored equally well (that is, assuming each model has a comparable dimensionality; if the model dimensions are significantly different, more time should be spent exploring in the higher-dimensional model).

To achieve good mixing within each model and between models, we introduce an artificial prior weighting on the models that compensates for the difference in the Bayes factors. For example, if the Bayes factor between two models is 1000, we introduce a prior that favors the low probability model by a factor of 1000, so the chains spend an equal number iterations in each model [28]. Since the appropriate weighting is not known in advance, an iterative scheme is used that adjusts the artificial prior weighting on the models until balance is achieved. The true Bayes factors are then found from the iteration ratio divided by the artificial prior odds ratio.

[1] P. Rosado, Phys. Rev. D **84**, 084004 (2011), arXiv:gr-qc/1106.5795.
[2] T. Regimbau and V. Mandic, Class. Quantum Grav. **25**, 184018 (2008), arXiv:astro-ph/0806.2794.
[3] T. Regimbau and S. A. Hughes, Phys. Rev. D **79**, 062002 (2009), URL `http://link.aps.org/doi/10.1103/PhysRevD.79.062002`.
[4] T. Regimbau, Research in Astronomy and Astrophysics **11**, 369 (2011), 1101.2762.
[5] E. Thrane, Phys. Rev. D **87**, 043009 (2013), arXiv:astro-ph/1301.0263.
[6] P. A. Rosado, A. Sesana, and J. Gair (2015), 1503.04803.
[7] I. Mandel, LIGO-P1400200 (2014).
[8] S. Drasco and E. E. Flanagan, Phys. Rev. D **67**, 082003 (2003), arXiv:gr-qc/0210032.
[9] D. M. Coward and R. R. Burman, MNRAS **361**, 362 (2005), astro-ph/0505181.
[10] N. Seto, Phys. Rev. D **80**, 043003 (2009), 0908.0228.
[11] L. Martellini and T. Regimbau, Phys. Rev. D **89**, 124009 (2014), 1405.5775.
[12] B. Allen and J. Romano, Phys. Rev. D **59**, 102001 (1999).
[13] E. T. Jaynes, *Probability theory: The logic of science* (Cambridge University Press, 2003).
[14] J. Skilling, Bayesian Analysis **1**, 833 (2006).
[15] F. Feroz, M. P. Hobson, E. Cameron, and A. N. Pettitt, ArXiv e-prints (2013).
[16] T. Littenberg and N. Cornish, LIGO-P1400187 (2014), arXiv:gr-qc/1410.3852.
[17] N. Cornish and T. Littenberg, LIGO-P1400185 (2014), arXiv:gr-qc/1410.3835.
[18] R. van Haasteren, Y. Levin, P. McDonald, and T. Lu, MNRAS **395**, 1005 (2009).
[19] N. J. Cornish and J. D. Romano, Phys. Rev. D **87**, 122003 (2013), 1305.2934.
[20] I. J. Good, *The Estimation of Probabilities: An Essay on Modern Bayesian Methods* (MIT Press, Cambridge, Massachusetts, 1965).
[21] C. N. Morris and S. L. Normand, in *Bayesian Statistics 4*, edited by A. P. D. J. M. Bernardo, J. O. Berger and A. F. M. Smith (Oxford University Press, Oxford, 1992), pp. 321–334.
[22] M. Adams and N. Cornish, Phys. Rev. **D82**, 022002 (2010), 1002.1291.
[23] E. Thrane and J. D. Romano, Phys. Rev. D **88**, 124032 (2013), URL `http://link.aps.org/doi/10.1103/PhysRevD.88.124032`.
[24] N. Christensen, Phys. Rev. D **46**, 5250 (1992).
[25] É. É. Flanagan, Phys. Rev. D **48**, 2389 (1993).
[26] A. Sesana, MNRAS **433**, L1 (2013), 1211.5375.
[27] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007), arXiv:gr-qc/0704.1808.
[28] C. Han and B. P. Carlin, Journal of the American Statistical Association **96**, 1122 (2001).