Gravitational wave detection without boot straps: A Bayesian approach

Gregory Ashton [*], Eric Thrane,[†] and Rory J. E. Smith[‡]
Monash Centre for Astrophysics, School of Physics and Astronomy,
Monash University, Victoria 3800, Australia
and OzGrav: The ARC Centre of Excellence for Gravitational-Wave Discovery,
Clayton, Victoria 3800, Australia

In order to separate astrophysical gravitational-wave signals from instrumental noise, which often contains transient non-Gaussian artifacts, astronomers have traditionally relied on bootstrap methods such as time slides. Bootstrap methods sample with replacement, comparing single-observatory data to construct a background distribution, which is used to assign a false-alarm probability to candidate signals. While bootstrap methods have played an important role establishing the first gravitational-wave detections, there are limitations. First, as the number of detections increases, it makes increasingly less sense to treat single-observatory data as bootstrap-estimated noise, when we know that the data are filled with astrophysical signals, some resolved, some unresolved. Second, it has been known for a decade that background estimation from time slides eventually breaks down due to saturation effects, yielding incorrect estimates of significance. Third, the false alarm probability cannot be used to weight candidate significance, for example when performing population inference on a set of candidates. Given recent debate about marginally resolved gravitational-wave detection claims, the question of significance has practical consequences. We propose a Bayesian framework for calculating the odds that a signal is of astrophysical origin versus instrumental noise without bootstrap noise estimation. We show how the astrophysical odds can safely accommodate glitches. We argue that it is statistically optimal. We demonstrate the method with simulated noise and provide examples to build intuition about this new approach to significance.

I. INTRODUCTION

Recent breakthroughs in gravitational-wave astronomy have been facilitated by Bayesian inference [1]. Inference allows robust and unbiased determination of the source properties, model comparisons, population-level inferences for a set of observations, and optimal searches for the gravitational wave background (for examples, see [2–9]). However, the problem of detection for transient compact binary coalescence signals has typically been cast in frequentist terms.

Detection is the problem of identifying and quantifying the significance of astrophysical signals in noisy data. To first order, the noise in gravitational-wave observatories such as LIGO/Virgo [10,11] can be described as a colored Gaussian process. However, gravitational-wave observatories are subject to frequent transient artifacts known as *glitches*; see, for example, Refs. [12–14]. Glitches can arise from any number of reasons, for example, photodiode saturation, environmental influence, and scattered light [15]. In most cases, however,

the cause of a glitch is unknown. Most glitches are of local original and therefore not causally connected between sites. However, events such as lightning can potentially produce correlated glitches. For these correlated glitches, data from environmental sensing equipment are used as a veto. For a complete guide to the LIGO and Virgo detector noise, see, e.g., Refs. [16,17]. Here, we assume that glitches are not causally connected between sites (i.e., that any glitches occurring from global phenomena have been vetoed by the environmental sensing). In our model, glitches in different observatories are independent and any coincidences are due entirely to chance.

For the gravitational-wave transients detected thus far, a *p*-value approach (also known as null-hypothesis significance testing) is often applied to assign significance for detection claims (see, e.g., Refs. [18–21]). A detection statistic for the signal is compared against a background distribution. Without the ability to shield the observatory from astrophysical signals, the background distribution is generated instead using bootstrap methods such as time slides (see, e.g., Refs. [22–29]). The term "bootstrap" refers to the practice of using an empirical distribution (in this case, a detection statistic) to estimate the properties of an

[*]greg.ashton@monash.edu
[†]eric.thrane@monash.edu
[‡]rory.smith@monash.edu

123018-1

estimator (in this case, the p value); for more information, see, e.g., Ref. [30].

Time slides are a classic example. Data from independent observatories are time shifted by an amount greater than the coherence time of the signal. Each time shift provides an independent realization of bootstrap generated noise. By calculating the detection statistic for each time shifted dataset, a background distribution can be generated. Astrophysical signals are detected when their detection statistic is suitably large compared to this background distribution. Time slides are a highly successful and convenient way to detect signals in the presence of transient noise. However, they have limitations.

With O3 sensitivity, the rate of binary black hole detections is expected to average one per week [5,31]. If several days of data around the event are used to perform background estimation, there is non-negligible probability of including another event close to the detection threshold. Moreover, in addition to individually detectable signals, the data will also contain a stochastic background of unresolved signals [6,32,33]. Using data containing astrophysical signals violates the basic assumptions of the time-slide method and will eventually (with increasing sensitivity) result in false dismissal of astrophysical signals due to contamination from signals in the background estimation procedure [34]. In addition, it was shown by Wąs *et al.* [35] that the time-slide method suffers a saturation problem: increasing the number of time slides eventually fails to produce new realizations of the background data due to the limited variety of glitches in the original dataset. This means the p value estimated from time slides can be incorrect and will not be improved by more time slides [35].

An alternative bootstrap method to time slides is to model the distribution of noise as a Poisson process, which combines random draws from single-detector background distributions [36,37]. The background distribution is modeled using the empirical distribution of the single-detector detection statistic. Employing bootstrap techniques, the method is subject to the same limitations as time slides: saturation (from the limited noise realization available for the empirical distribution) and contamination from signals. See [38–40] for recent updates and applications of this method.

Large-scale mock data challenges comparing pipelines [41] have validated the p-value approach to significance for rare events (unambiguous detections) using various bootstrap methods. However, the authors of [41] point out that bootstrap methods are subject to limitations when applied to marginal events.[1]

These limitations motivate a fully Bayesian approach which eschews the bootstrap estimation of the background, a conclusion supported by [41]. We propose an approach, which unifies, in a single framework, the problems of significance and parameter estimation. We argue that the significance of a candidate event does not depend on the detection pipeline(s) used to identify it.

We anticipate that this Bayesian approach has broader applications for population modeling, robust multimessenger detection [42], and detector characterization.

There has been some work already toward this end. The first LIGO/Virgo gravitational-wave transient catalog [43] includes a table of p_{astro}, a Bayesian odds comparing the astrophysical/terrestrial hypotheses [44–46]. We regard this step as an important development. The method for calculating p_{astro} is fully Bayesian. However, it relies on the bootstrap noise models of the search pipelines used to identify the candidate events. Thus, the limitations of bootstrap methods discussed above affect the determination of p_{astro}. Moreover, using the bootstrap noise models of different search pipelines leads to other undesirable consequences which are highlighted in Table IV of [43]: different pipelines produce different values of p_{astro}. Consider, for example, GW170729: the p_{astro} values range from 52% to 98%. Recent claims from [47] have yielded new candidates—using an entirely different pipeline—which were not deemed significant by LIGO/Virgo, at least in the first gravitational-wave transient catalog [43]. This begs the question: what is the *actual* probability that GW170729 is of astrophysical origin? We argue that the question is best answered in a pipeline-independent way, using the same infrastructure used for parameter estimation.

The remainder of this paper is organized as follows. In Sec. II, we introduce the formalism for calculating the astrophysical odds, quantifying the probability that a data segment contains a signal of astrophysical origin versus noise. In Sec. III, we carry out a toy model demonstration to build intuition for the odds before providing a full-scale simulation study of binary black hole signals in modeled interferometer data in Sec. IV. We conclude in Sec. V with a discussion and future outlook.

II. TOWARD BAYESIAN DETECTION

In this section, we introduce a Bayesian odds for determining the significance of a gravitational-wave candidate without time slides. The details are somewhat technical, and it is necessary to introduce a bit of notation. Rather than build the odds from constituent parts, we opt to begin with the result and then explain the components one at a time.

This section is organized as follows. In Sec. II A, we provide a general expression for the astrophysical odds and define the component parts. In Sec. II B, we introduce a mixture model describing the data as a combination of signal and noise. In Sec. II C, we describe the noise model,

---

[1]As an aside, approaches based on data statistics hypotheses face a fundamental limitation: the associated p value cannot be used as a "score" to determine a candidate's astrophysical significance.

which includes a model of glitches. In Sec. II D, we describe the signal model for gravitational waves from compact binaries. In Sec. II E, we provide a practical expression for the astrophysical odds based on our signal and noise models. Finally, in Sec. II F, we place the method in the context of previous literature.

A. Formalism

We presuppose that the data are divided into segments, each of which may contain a single binary signal. (This assumption is valid provided the segment duration is less than 108 s, the implied 90% lower bound on the average waiting time between stellar-mass black hole coalescence events throughout the universe [48]). The following boxed equation is the *astrophysical odds*, which answers the question "what are the odds that the ith data segment contains a signal $\mathcal{S}_i$, given all the data $\boldsymbol{d}$?"

$$\boxed{\mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(\boldsymbol{d}) = \frac{\int d\Lambda \mathcal{L}(d_i|\mathcal{S}_i,\Lambda)\pi(\Lambda|\boldsymbol{d}_{k\neq i})\pi(\mathcal{S}_i|\boldsymbol{d}_{k\neq i},\Lambda)}{\int d\Lambda \mathcal{L}(d_i|\mathcal{N}_i,\Lambda)\pi(\Lambda|\boldsymbol{d}_{k\neq i})\pi(\mathcal{N}_i|\boldsymbol{d}_{k\neq i},\Lambda)}.}$$

(1)

Let us go through component pieces in turn:

(i) $\mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(\boldsymbol{d})$. The detection statistic is a *hyperparametrized astrophysical odds*, or for short, astrophysical odds, a single number, which we denote $\mathcal{O}$. See Appendix A for a derivation of the hyperodds in general. As all odds, it compares two hypotheses, in this case $\mathcal{S}$, the *signal hypothesis*, and $\mathcal{N}$, the *noise hypothesis*. According to the signal hypothesis, the segment contains an astrophysical signal in among noise. According to the noise hypothesis, there is no signal present, just noise. Noise, here, refers to both Gaussian noise and/or a glitch. The astrophysical odds depends on the *full dataset $\boldsymbol{d}$*.

(ii) Λ. The numerator and denominator include integrals over Λ, which is a set of *hyperparameters* modeling the distribution of *signal and noise parameters θ_i*. The models are described using *conditional priors* $\pi(\theta_i|\Lambda)$.

(iii) d_i and $\boldsymbol{d}_{k\neq i}$. We divide $\boldsymbol{d}$ into d_i, the ith data segment, and every other segment of data, $\boldsymbol{d}_{k\neq i}$, which we refer to as the *contextual data*. The contextual data provide context with which to understand the significance of d_i.

(iv) $\mathcal{L}(d_i|\mathcal{S}_i,\Lambda)$. The next term in the numerator is the likelihood of the data given the signal hypothesis, and given the hyperparameters. For short, we call it *the signal evidence*. The corresponding term in the denominator is $\mathcal{L}(d_i|\mathcal{N}_i,\Lambda)$, the likelihood of the data given the noise hypothesis and given the hyperparameters. For short, we call it *the noise evidence*. We note that these terms could also be

described equivalently as the fully marginalized signal and noise likelihoods.

(v) $\pi(\Lambda|\boldsymbol{d}_{k\neq i})$. The next term in the numerator is the posterior for the hyperparameters given the contextual data. We use this $k \neq i$ posterior as the prior for event i. For short, we call this *the hyperprior*.

(vi) $\pi(\mathcal{S}_i|\boldsymbol{d}_{k\neq i},\Lambda)$. The final term in the numerator is the prior for the signal hypothesis given the contextual data and given the hyperparameter Λ. For short, we call this *the signal prior*. The corresponding term in the denominator is the prior for the noise hypothesis, $\pi(\mathcal{N}_i|\boldsymbol{d}_{k\neq i},\Lambda)$. For short, we call this *the noise prior*.

Qualitatively, Eq. (1) is straightforward to understand now that we have introduced the necessary notation. Our signal and our noise are described by parameters θ_i. The prior distribution for θ_i has an uncertain shape, which we model using hyperparameters Λ, so that the signal and noise parameters are conditional on the hyperparameters: $\pi(\theta_i|\Lambda)$. In other words, we employ a hierarchical model [49]. The odds, defined in Eq. (1), is an example of a hyperodds (see Appendix A) which marginalizes over uncertainty in the hyperparameters using contextual data $\boldsymbol{d}_{k\neq i}$. The overall result is that Eq. (1) compares the probability for a signal and noise model evaluated using a hierarchical model conditional on the contextual data. In the following subsections, we will describe how each term in Eq. (1) is calculated in detail, resulting in the version used in practice, Eq. (15).

B. Mixture model

In Sec. II A we presented the astrophysical odds in a general form. It is now necessary to specialize further. First, we assume that the hyperparameters consist of three components:

$$\Lambda \rightarrow \{\Lambda_{\mathcal{S}}, \Lambda_{\cancel{\mathcal{S}}}, \xi\}$$

(2)

$$= \{\Lambda_{\cancel{\mathcal{S}}}, \xi\},$$

(3)

where $\xi \equiv \pi(\mathcal{S}|\xi)$ is the mixing fraction and the signal hyperparameters $\Lambda_{\mathcal{S}}$ determine the shape of the signal priors. For this analysis, we assume there are no signal-hyper parameters, and so this term can be neglected:

$$\mathcal{L}(d_i|\mathcal{S}_i,\Lambda) \rightarrow \mathcal{L}(d_i|\Lambda_{\mathcal{S}},\mathcal{S}_i)$$

(4)

$$= \mathcal{L}(d_i|\mathcal{S}_i).$$

(5)

However, one could straightforwardly extend the analysis to incorporate signal hyperparameters, for example, from [50].

The noise hyperparameters $\Lambda_{\cancel{\mathcal{S}}}$ determine the shape of the noise prior:

$$\mathcal{L}(d_i|\cancel{\mathcal{S}}_i, \Lambda) \rightarrow \mathcal{L}(d_i|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_i). \tag{6}$$

This is where the action happens: the problem of determining the significance of a candidate event is recast as a problem of determining a suitable hyperparametrization for the noise distribution.

Constructing a mixture model for the exhaustive hypothesis $\mathcal{S} \vee \mathcal{N}$, we obtain a *general hyperlikelihood*, conditional on the hyperparameters:

$$\mathcal{L}(d|\Lambda_{\mathcal{S}}, \Lambda_{\cancel{\mathcal{S}}}, \xi, \mathcal{S} \vee \mathcal{N}) = \xi \mathcal{L}(d|\mathcal{S}) + (1 - \xi)\mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \mathcal{N}). \tag{7}$$

The general likelihood is used to calculate $\pi(\Lambda|\boldsymbol{d}_{k \neq i})$. In Secs. II C and II D, we give details on how the noise and signal evidences are calculated in practice.

C. Noise model

We adopt a noise model consisting of multiple subhypotheses and for simplicity restrict the discussion to just two observatories, a and b. The entire noise hypothesis, which is the union of four subhypotheses, is denoted $\mathcal{N}$. Each subhypothesis accounts for the different kinds of noise. For example, $\cancel{\mathcal{S}}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b$ is the subhypothesis that no signal is present ($\cancel{\mathcal{S}}_0$); there is a glitch in observatory a ($\mathcal{G}^a$); there is no glitch present in observatory b ($\cancel{\mathcal{G}}^b$). The complete noise hypothesis is

$$\mathcal{N} \equiv \cancel{\mathcal{S}}_0 \mathcal{G}^a \mathcal{G}^b \vee \cancel{\mathcal{S}}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b \vee \cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^a \mathcal{G}^b \vee \cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^a \cancel{\mathcal{G}}^b. \tag{8}$$

The noise evidence from Eq. (7) can therefore be calculated using a mixture model

$$\begin{aligned}
\mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}_0}, \mathcal{N}) &= \xi_g^a \xi_g^b \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \mathcal{G}^a \mathcal{G}^b) \\
&\quad + \xi_a^g (1 - \xi_b^g) \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b) \\
&\quad + (1 - \xi_g^a) \xi_g^b \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^a \mathcal{G}^b) \\
&\quad + (1 - \xi_g^a)(1 - \xi_g^b) \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^a \cancel{\mathcal{G}}^b), \tag{9}
\end{aligned}$$

where $\xi_g^x \equiv \pi(\mathcal{G}^x|\Lambda_{\cancel{\mathcal{S}}})$ is the glitch mixing fraction for observatory x. The individual likelihoods in this expression are the evidence for glitches (or a lack thereof) in each observatory. Assuming the noise (including glitches) are independent between the two observatories, we can simplify the likelihoods, e.g.,

$$\mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b) = \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \mathcal{G}^a) \mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^b). \tag{10}$$

In the absence of a signal or glitch hypothesis, e.g., $\mathcal{L}(d|\cancel{\mathcal{S}}_0 \cancel{\mathcal{G}}^x)$, the evidence is the usual Gaussian-noise evidence [51]. On the other hand, given a glitch model, the evidence for a glitch in the x observatory is calculated from marginalizing over the glitch model parameters,

$$\mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}} \mathcal{G}^x) = \int d\theta \, \mathcal{L}(d|\theta, \Lambda_{\cancel{\mathcal{S}}}, \mathcal{G}^x) \pi(\theta|\Lambda_{\cancel{\mathcal{S}}}, \mathcal{G}^x). \tag{11}$$

In practice, this integration is performed numerically using nested sampling methods [52]. In order to allow rapid evaluation of $\mathcal{L}(d|\Lambda_{\cancel{\mathcal{S}}}, \cancel{\mathcal{S}} \mathcal{G}^x)$ while varying $\Lambda_{\cancel{\mathcal{S}}}$, we use the so-called recycling method; see Appendix B and Refs. [51,53,54].

However, it is difficult to build a glitch model to evaluate Eq. (11) from first principles since most glitches are poorly understood. In this work, we instead apply the conservative model first proposed by Veitch and Vecchio [55] in which glitches are modeled as compact binary signals with uncorrelated model parameters in each observatory.

This glitch model is founded on the principle of modeling the worst-case scenario: glitches are indistinguishable from signals except for the absence of coherence (in the model parameters θ) between observatories. Therefore for a signal to be preferred over this glitch model, it must not only match modeled waveforms, but must also look like the same binary system in multiple observatories with arrival times consistent with an astrophysical origin. We will apply this idea, further developed in Isi *et al.* [56], which extends the noise hypothesis to include Gaussian noise.

This glitch model is conservative in that we could better distinguish glitches and hence boost the significance of astrophysical signals by including more physically motivated models of glitches (see, e.g., Ref. [57]). However, in the absence of a trustworthy physical glitch model, our conservative approach ensures that we do not generate false positives due to a flaw in our glitch model. That said, this formalism can be extended to accommodate more sophisticated models. We return to this below when discussing possibilities for future work.

D. Signal model

We adopt a signal model consisting of multiple subhypotheses and for simplicity restrict the discussion to just two observatories a and b. The entire signal hypothesis, which is the union of four subhypotheses, is denoted $\mathcal{S}$. Each subhypothesis accounts for the different kinds of noise that the signal can be embedded within. For example, $\mathcal{S}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b$ is the subhypothesis that a signal is present ($\mathcal{S}_0$) with a glitch in observatory a ($\mathcal{G}^a$) with no glitch present in observatory b ($\cancel{\mathcal{G}}^b$). With these definitions, the complete signal hypothesis is

$$\mathcal{S} \equiv \mathcal{S}_0 \mathcal{G}^a \mathcal{G}^b \vee \mathcal{S}_0 \mathcal{G}^a \cancel{\mathcal{G}}^b \vee \mathcal{S}_0 \cancel{\mathcal{G}}^a \mathcal{G}^b \vee \mathcal{S}_0 \cancel{\mathcal{G}}^a \cancel{\mathcal{G}}^b. \tag{12}$$

Qualitatively, this expression captures the idea that a signal in the detector $\mathcal{S}_0$ may coincide with or without a glitch in either detector. Note that, while we refer to $\mathcal{S}$ as "the signal hypothesis" for brevity, it might be more aptly named "the set of all hypotheses that include a signal."

Signals from compact binaries are typically characterized by 15 parameters. In the previous subsection, we introduced a noise model where glitches modeled as incoherent binary signals, which introduces 15 parameters per observatory. Thus, this formulation will require marginalizing over 30 parameters (for the $\mathcal{S}\mathcal{G}^a\mathcal{G}^b$, $\mathcal{S}\mathcal{G}^a\mathcal{G}^b$ subhypotheses) and marginalizing over a 45-dimensional parameter space for the $\mathcal{S}\mathcal{G}^a\mathcal{G}^b$ subhypothesis. With current nested-sampling methods, these integrals remain challenging and robust methods to calculate them are an unsolved problem which has a number of applications beyond the current work.

However, if the rate of astrophysical signals is sufficiently small, then we can make the following approximation as in Smith and Thrane [6]:

$$\mathcal{S} \approx \mathcal{S}_0 \mathcal{G}^a \mathcal{G}^b. \qquad (13)$$

As such, the signal evidence in Eq. (7) is calculated by simply marginalizing over binary parameters

$$\mathcal{L}(d|\mathcal{S}) = \mathcal{L}(d|\mathcal{S}_0\mathcal{G}^a\mathcal{G}^b) = \int d\theta \mathcal{L}(d|\theta,\mathcal{S}_0)\pi(\theta|\mathcal{S}_0), \quad (14)$$

where we drop the conditional probability on the no-glitch hypotheses where they have no bearing in the signal likelihood, $\mathcal{L}(d_i|\theta,\mathcal{S}_0)$. The signal likelihood is constructed from a Gaussian noise likelihood (with the power spectral density estimated from the data), and standard stochastic sampling procedures are used to perform the marginalization (see, e.g., Ref. [51] for a detailed discussion).

E. The astrophysical odds: A practical implementation

The astrophysical odds has been given in the generic form of Eq. (1). Having defined the noise and signal models (see Secs. II C and II D), we obtain an expression for our specific signal and noise models:

$$\boxed{\mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(d) = \frac{\mathcal{L}(d_i|\mathcal{S}_i)\int d\xi\, \xi\, \pi(\xi|d_{k\neq i})}{\iint d\xi d\Lambda_{\mathcal{N}}\mathcal{L}(d_i|\Lambda_{\mathcal{N}},\mathcal{N}_i)(1-\xi)\pi(\xi,\Lambda_{\mathcal{N}}|d_{k\neq i})}.}$$
$$(15)$$

The odds of Eq. (15) is the Bayesian answer to the question: does data segment d_i contain an astrophysical signal (without a glitch) or is it noise? Where the noise hypothesis includes a conservative glitch model and the question is asked, not in the isolated case of a single data segment, but in the context of some wider set of data d.

That the odds is a fully Bayesian answer is important. There is no need to treat these odds as a frequentist detection statistic. There is no need to perform time slides to generate a background and then calculate a false alarm rate, a technique which runs into problems with saturation.

The odds is a statement about the probability that the data are signals rather than noise; for example, odds of $9\!:\!1$ will be a signal 9 times out of 10. We discuss this point further in Sec. III.

A point we have yet to discuss, and will be the core issue for practical applications of this method, is that the odds will be sensitive to the validity of the noise hypermodel. However, this is not a drawback to the method, but a feature. This method makes explicit the underlying noise model which is being applied. This will require careful checking to ensure the noise hypermodel is appropriate using posterior predictive checks (see, e.g., Gelman *et al.* [49]).

F. The astrophysical odds: Relation to previous results

The hyperparametrization step in the astrophysical odds is unique to this work. However, our work builds on ideas in the literature. Here, we show how these are related.

In Veitch and Vecchio [55], the coherence test was introduced, which is a Bayes factor between a coherent signal $\mathcal{S}$ and incoherent glitches $\mathcal{G}^a\mathcal{G}^b$. In the notation of this work, this is

$$B_{\mathrm{coh,inc}}(d_i) \equiv \frac{\mathcal{L}(d_i|\mathcal{S})}{\mathcal{L}(d_i|\mathcal{G}^a\mathcal{G}^b)} = \frac{\mathcal{L}(d_i|\mathcal{S})}{\mathcal{L}(d_i|\mathcal{G}^a)\mathcal{L}(d_i|\mathcal{G}^b)}, \quad (16)$$

with $\mathcal{S}$ as defined in Eq. (13). Note, in the original work this was generalized to an arbitrary number of observatories.

Later, Isi *et al.* [56] introduced the Bayesian coherence-ratio (BCR) statistic,

$$\mathrm{BCR} \equiv \frac{P(\mathcal{S}|d_i)}{P(\mathcal{N}|d_i)} = \frac{\alpha \mathcal{L}(d_i|\mathcal{S})}{\prod_{\ell=a,b}[\mathcal{L}(d_i|\mathcal{G}^\ell)\beta_\ell + (1-\beta_\ell)\mathcal{L}(d_i|\mathcal{G}^\ell)]},$$
$$(17)$$

which is an odds comparing the signal hypothesis $\mathcal{S}$, as defined in Eq. (13), with incoherent glitches or Gaussian noise. Comparing with the work herein, $\alpha = \xi/(1-\xi)$ and $\beta_\ell = \xi_g^\ell$. In the case where no contextual data are used, the astrophysical odds, Eq. (15), is a generalization of the BCR odds. Isi *et al.* [56] demonstrate how the rate of signals and glitches (α and β) can instead be tuned using time-slid data and injections to maximally separate signals from noise. This is analogous to how ξ and ξ_g are estimated in the astrophysical odds, but differs in that the work presented herein uses contextual non-time-slid data. The astrophysical odds also allows the glitch model-parameter hyperpriors to be inferred from the contextual data.

III. TOY MODEL DEMONSTRATION

In this section we demonstrate the Bayesian odds with a toy-model problem. The point of this exercise is to show the following: (1) given well-informed priors, the Bayesian odds has a clear and reliable interpretation; (2) if the prior is

misinformed (i.e., does not represent our true beliefs), this interpretation becomes unreliable. Later, in Sec. IV, we show that if the exact prior is unknown, it can be inferred using hierarchical modeling, reestablishing the reliability of the Bayesian odds.

In our toy model, each measurement consists of a single number x_i. The data are generated either by a noise model, consisting of a standard normal distribution (i.e., zero mean and unit variance), or by a signal mode, consisting of a normal distribution with some nonzero mean and unit variance. To summarize, our signal and noise hypotheses are as follows:

(i) $\mathcal{S}$: $x_i \sim \text{Normal}(\mu, \sigma = 1)$ with $\mu \sim \text{Uniform}(0, 1)$.
(ii) $\mathcal{N}$: $x_i \sim \text{Normal}(\mu = 0, \sigma = 1)$.

We fix the prior odds to unity, i.e., $\pi(\mathcal{S}) = \pi(\mathcal{N}) = 1/2$.

We simulate a dataset consisting of 10 data points and randomly assign each one to the $\mathcal{S}$ and $\mathcal{N}$ categories. Next, we draw random values of x_i based on the category of each event. Having generated a set of data $\{x_i\}$, we calculate the evidences $\mathcal{L}(\{x_i\}|\mathcal{N})$ and $\mathcal{L}(\{x_i\}|\mathcal{S})$. The first of these can be calculated directly, the second is estimated using a nested-sampling algorithm, marginalizing over μ. Initially, we apply the prior $\pi(\mu|\mathcal{S}) = \text{Uniform}(0, 1)$, i.e., the correct population prior used to generate the data. (Below we repeat the calculation using an intentionally misinformed prior for illustrative purposes.) Once these estimates of the evidence are calculated, we calculate the odds $\mathcal{O}$ for the dataset.

We demonstrate that if the odds is formulated correctly, they can be trusted at face value. That is, events with odds of $\mathcal{O} = 9$ will turn out to be signal 9 times out of 10. We illustrate this with a plot. Let $\delta M_{\mathcal{S}}(\mathcal{O})$ be the number of datasets simulated as signals with an odds on the interval $(\mathcal{O}, \mathcal{O} + \delta\mathcal{O})$. Then, if $\delta M_{\mathcal{N}}(\mathcal{O})$ is the number of simulated datasets simulated as noise on the same interval we can define

$$\mathcal{R}(\mathcal{O}) \equiv \frac{\delta M_{\mathcal{S}}(\mathcal{O})}{\delta M_{\mathcal{N}}(\mathcal{O})}. \qquad (18)$$

If the odds is what they say they it is, then this ratio should exactly equal the odds, i.e., $\mathcal{R} = \mathcal{O}$. In Fig. 1, we simulate 1000 datasets and plot $\log_{10}(\mathcal{R})$ against $\log_{10}(\mathcal{O})$ using the *correct prior*: i.e., the prior distribution from which the model parameters where drawn (blue line). The odds performs as expected: the diagonal, corresponding to $\mathcal{R} = \mathcal{O}$, is consistent with the diagonal $\mathcal{R} = \mathcal{O}$ line.

For comparison, we can purposefully repeat the calculation using a *misinformed prior* to understand what effect this has on the plot. We repeat the steps above, but when calculating the signal evidence, we do not use the correct prior distribution on μ, but instead use an intentionally misinformed prior $P(\mu|\mathcal{S}') \propto \mu^2$ with a minimum of 0 and a maximum of 1, i.e., a power-law distribution with the same support as the correct prior, but a different spectral index. In effect, we are comparing a new hypothesis $\mathcal{S}'$ with the noise instead of $\mathcal{S}$. The result is shown in Fig. 1 (orange line).
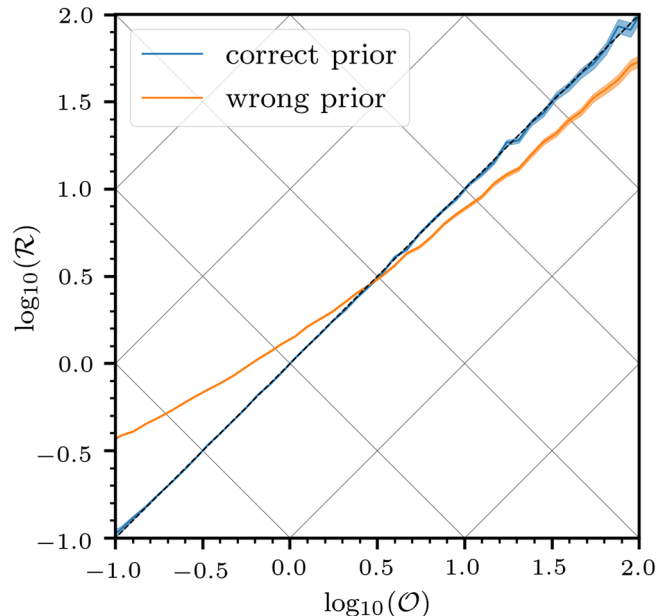

FIG. 1.   Performance plot of the odds using the correct and incorrect priors (see text). Shaded regions indicate the approximate $1 - \sigma$ uncertainty due to the finite number of realizations used. An optimally performing odds lies on the diagonal line.

Unlike the case of the correct prior, the odds now does not behave as expected: at times it is too liberal and other times too conservative.

IV. DEMONSTRATION WITH BINARY BLACK HOLE MERGERS

We now simulate the problem of binary coalescence significance estimation with data from two observatories (labeled H and L) containing binary black hole signals and glitches (as described in Secs. II C and II D). From this simulated data, we first infer the population

TABLE I.   Binary black hole coalescence simulation population parameters. The signal rate, ξ, and glitch rate for the H and L detectors, ξ^H and ξ^L, refer to the probability of a 4 s data segment containing a signal or glitch. The luminosity distance (chirp-mass) population distributions for signals d_L^s $(\mathcal{M}_s)$ and glitches d_L^g $(\mathcal{M}_g)$ have equal support, but obey differing scaling relations; the same glitch population is applied to both the H and L observatories.

Parameter	Distribution
$\xi = 0.001$	
$\xi_g^H = 0.6$	
$\xi_g^L = 0.4$	
$d_L^s \sim \text{PowerLaw}(\alpha = 2, 1\text{ Gpc}, 2\text{ Gpc})$	
$d_L^g \sim \text{PowerLaw}(\alpha = 0, 1\text{ Gpc}, 2\text{ Gpc})$	
$\mathcal{M}_s \sim \text{PowerLaw}(\alpha = 0, 25\text{ M}_\odot, 100\text{ M}_\odot)$	
$\mathcal{M}_g \sim \text{PowerLaw}(\alpha = 2, 25\text{ M}_\odot, 100\text{ M}_\odot)$	

hyperparameters, verifying that the values used in generating the simulation are properly recovered. Then, we demonstrate calculation of the astrophysical odds, showing the behavior as a function of the amount of contextual data. We use the BILBY [58] Bayesian inference package to generate and analyze the simulated data and the DYNESTY [59] nested-sampling package for parameter estimation and evidence evaluation.

We define the simulated population parameters in Table I. The signal and glitch rates (ξ, ξ_g^H, and ξ_g^L) refer to the probability of a 4 s data segment containing a signal or glitch. Their values are selected so that in a sample of a few hundred data segments, the expected number of segments containing a signal and glitch is less than one, but the absolute number of signals and glitches is sufficiently large enough for population-level inference. Meanwhile, Table I also outlines the population distribution of luminosity distances and chirp mass for signals and glitches; these have the same support, but obey different scaling laws. The population hyperparameters are chosen so that glitches are more frequent and tend to be louder and shorter in duration than signals.

Each segment of the contextual data is generated by first drawing parameters from the population level rate parameters (i.e., Table I) to determine what the segment should contain. If the segment is to contain a signal, a set of signal parameters are drawn from a standard set of priors, but with the luminosity distance and chirp mass as given in Table I. If the segment is to contain a glitch, a set of glitch parameters (for each observatory) is drawn from a standard set of priors identical to the standard signal priors, but with the luminosity distance and chirp mass as given in Table I. Afterwards, the simulated signals and glitches are added to Gaussian noise. The simulation will add both signals and glitches to the data simultaneously; however, we have chosen the rate parameters to ensure the probability of this occurring is small.

Having generated the contextual data segments, we proceed to recover the hyperparameters; this is done by applying Eq. (7), where a hyperparameter model for each observatory is used for both the luminosity distance and chirp mass. Each model has the same support as the true priors (see Table I), but an unknown spectral index α with subscript either H or L, labeling the observatory and superscript labeling if it is for the luminosity distance d_L or chirp mass $\mathcal{M}$. We repeat the hyperparameter inference for a variable number of segments randomly drawn from the prior. In Fig. 2, we show the marginalized one-dimensional posterior distributions for the hyperparameters as a function of the number of segments. This demonstrates the expected results that, as the number of segments increases, we correspondingly see the posteriors converge to the true values.

As the amount of contextual data increases, Fig. 2 demonstrates that inferences about the hyperparameters also become increasingly well informed. How does this
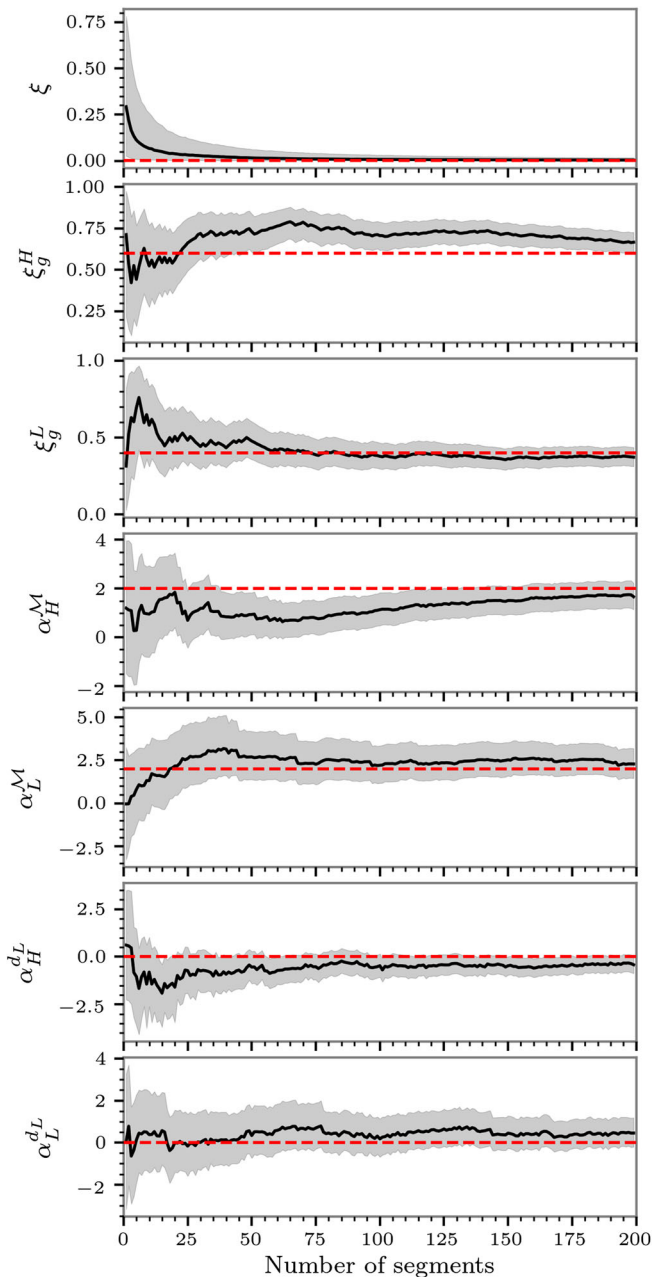

FIG. 2. Marginalized posterior median (solid line) and 90% credible interval (shaded region) for the hyperparameter of the binary black hole simulation. True values are shown as horizontal dashed lines.

affect the astrophysical odds? To study this, we simulate a glitch in Gaussian observatory noise and compute the astrophysical odds using Eq. (1) while varying the set of contextual data (using the same datasets used to produce Fig. 2). The glitch is the worst possible type of glitch: a quasicoherent glitch. Such a glitch consists of a binary black hole signal injected into both observatories with near-identical signal parameters to within the uncertainties afforded by the background Gaussian observatory noise. For false alarms, this is precisely the sort of glitch which is best at deceiving traditional detection statistics.
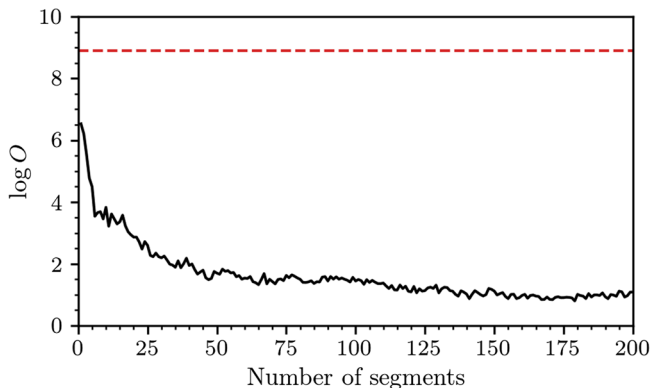
FIG. 3. The astrophysical odds, Eq. (15) as a function of the number of contextual data segments used in the calculation of the background prior. A dashed horizontal line marks the value of the odds without any contextual data.

The quasicoherent glitch has a network optimal signal-to-noise ratio of ~ 50 and log-Bayes factor comparing a signal versus Gaussian noise of ~ 1000; when compared to Gaussian noise it is distinctly signal-like. However, the simulated contextual data contain a multitude of non-Gaussian artifacts. For illustrative purposes, we calculate the significance several different ways. The coherence test from [55] yields a Bayes factor $\log B_{\rm coh,inc} = 8.9$ [see Eq. (16) for a definition]. This is unsurprising given we injected a quasicoherent glitch, which is essentially indistinguishable from a coherent signal. The BCR method [56] [see Eq. (17) for a definition] depends on the tuning parameters α and β. Using uniformed values of $\alpha = 1$ and $\beta = 0.5$, we have that $\log {\rm BCR} = 8.9$. In this case the alternative "Gaussian noise" hypothesis provided by the BCR [56] does not further distinguish the event. That these methods fail to veto this quasicoherent glitch is unsurprising and precisely the motivation for the astrophysical odds, which incorporates knowledge about the background. We note that tuning the BCR parameters would improve the performance of this metric, but is beyond the scope of this paper.

In Fig. 3, we plot the astrophysical odds, for this same pathological event, as a function of the number of background segments used. For small numbers of segments, the astrophysical odds agrees with the $\log B_{\rm coh,inc}$ and $\log {\rm BCR}$ numbers, which makes sense since the background is not yet well constrained. As the number of segments increases, however, the odds tends to decrease, eventually reaching a threshold of $\log \mathcal{O} \approx 0$. In this instance, the astrophysical odds is (appropriately) far more conservative than the other approaches. Given the nature of the background, the statistical significance of the event is marginal. It is worth noting that a similarly loud signal would also fail to pass the bar for detectability: fundamentally, the background in this region includes so many loud glitches that it decreases the operational sensitivity.

That the odds tends to a threshold is a direct result of the limited amount of information available from the background as seen in Fig. 2. Initially, each extra set of segments provides a substantial improvement to the constraint on the hyperpriors, but eventually once the background has been adequately sampled, extra data do not provide much new information about the nature of the background. The result is that one needs to make sure the number of background segments is sufficiently large before drawing a conclusion. By studying this behavior, gravitational-wave astronomers can ascertain how much data are required to determine the significance of a candidate before reaching the point of diminishing returns.

V. DISCUSSION AND OUTLOOK

The tools of Bayesian inference have been widely adopted in astrophysics due to their utility in parameter estimation and model selection. In this work, we introduce a complete formalism for calculating the astrophysical odds [Eq. (1)] for transient gravitational wave signals. Gravitational wave observatories are plagued by transient non-Gaussian artifacts often referred to as glitches. These glitches need to be carefully considered when evaluating the significance of a putative signal.

Building on the work of Smith and Thrane [6], Veitch and Vecchio [55], and Isi *et al.* [56], we present a complete framework for defining the likelihood of the data containing a signal, accounting for both Gaussian noise and glitches in the data. The conservative glitch model used herein relies on the ideas of pairs of incoherent compact binary signals as a conservative glitch model as proposed by Veitch and Vecchio [55]. However, in the future it may prove fruitful to extend this glitch model to include, for example, sine Gaussians [57] in addition to compact binary signals. If the data support the hypothesis that glitches look more like sine-Gaussians than compact binaries, the astrophysical odds will automatically adopt the less conservative assumption.

The key new ingredient developed in this work is the ability to marginalize over the glitch population properties using contextual data. This is done using hyperparameters, which allow the glitch population to differ from the priors for astrophysical signals. For example, astrophysical signals are expected to follow a prior in which more signals are found at a larger distance [in this paper, we model this as $P(d_L|I) \propto d_L^2$]. But for glitches in the observatory (as described by fitting incoherent compact binary coalescence signals), this is unlikely to be true. Equation (1) is the main result of this paper. It describes how the odds of an astrophysical signal can be calculated, marginalized over the contextual data which surround the event. This result is fully Bayesian: it does not require boot straps and hence does not suffer from issues such as saturation [35] or signal contamination. Moreover, the Bayesian approach (see also Refs. [44–46]) allows the significance to be easily integrated into downstream analyses, for example in population modeling.

One subtlety to this type of analysis is the characterization of the background. We demonstrated in this work how injected background with known power-law distributions can be recovered. But, for real interferometer data, more nuanced models may be needed. In future work, we aim to apply this method to data from the LIGO and Virgo interferometers. As part of this work, we aim to recompute the significance of previously published gravitational-wave detections and/or candidates [43,47,60]. We plan to carry out diagnostic tests, for example, with posterior predictive checking, in order to build confidence in our noise model.

ACKNOWLEDGMENTS

APPENDIX A: HYPERODDS

Typically, Bayesian model selection problems answer the question "given some data d, what are the relative probabilities of model A and model B?" by calculating the odds, $\mathcal{O}_{A/B}(d) \equiv P(A|d)/P(B|d)$. In this work, we answer the question "given a set of data $\boldsymbol{d}$, what are the relative probabilities of model A and model B for the ith data segment?" by calculating the generalized notion of a *hyperodds* which we define now. The important distinction here is that the data $\boldsymbol{d}_{k \neq i}$, which we refer to as *contextual data*, are used to inform the odds about the typical characteristics of the A and B models.

Let A_i be the hypothesis of model A for the ith segment. Then, the model evidence for A_i given all of the data can be calculated from

$$\mathcal{L}(A_i|\boldsymbol{d}) = \int d\Lambda \mathcal{L}(A_i|\boldsymbol{d}, \Lambda)\pi(\Lambda|\boldsymbol{d}). \tag{A1}$$

Applying the rules of conditional probability and explicitly writing the likelihood conditional on the data for segment i and for the contextual data, we find that

$$\mathcal{L}(A_i|\boldsymbol{d}) = \frac{1}{\mathcal{L}(\boldsymbol{d})}\int d\Lambda \mathcal{L}(d_i|A_i, \Lambda)\prod_{k \neq i}\mathcal{L}(d_k|\Lambda)\pi(\Lambda). \tag{A2}$$

Typically, the normalizing factor in this equation cannot easily be calculated since an exhaustive set of models is seldom known. Instead, the common alternative is to instead calculate an odds

$$\mathcal{O}_{A_i/B_i}(\boldsymbol{d}) = \frac{\mathcal{L}(A_i|\boldsymbol{d})}{\mathcal{L}(B_i|\boldsymbol{d})}. \tag{A3}$$

Applying Eq. (A2), we therefore have that

$$\mathcal{O}_{A_i/B_i}(\boldsymbol{d}) = \frac{\int d\Lambda \mathcal{L}(d_i|\Lambda, A_i)\mathcal{L}(A_i|\Lambda)\prod_{k \neq i}\mathcal{L}(d_k|\Lambda)\pi(\Lambda)}{\int d\Lambda \mathcal{L}(d_i|\Lambda, B_i)\mathcal{L}(B_i|\Lambda)\prod_{k \neq i}\mathcal{L}(d_k|\Lambda)\pi(\Lambda)}. \tag{A4}$$

Finally, we can simplify Eq. (A4) to

$$\mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(\boldsymbol{d}) = \frac{\int d\Lambda \mathcal{L}(d_i|\Lambda, \mathcal{S}_i)\mathcal{L}(\mathcal{S}_i|\Lambda)\pi(\Lambda|\boldsymbol{d}_{k \neq i})}{\int d\Lambda \mathcal{L}(d_i|\Lambda, \mathcal{N}_i)\mathcal{L}(\mathcal{N}_i|\Lambda)\pi(\Lambda|\boldsymbol{d}_{k \neq i})}, \tag{A5}$$

where $\pi(\Lambda|\boldsymbol{d}_{k \neq i})$ is the posterior distribution on the hyperparameters, conditional on the context data. This expression allows calculation of the odds marginalizing over uncertainty about the priors, at both the model parameters and the hypothesis level, using the contextual data. To provide some intuition and demonstrate the power of this general method we now discuss two specific cases.

First, let $P(A_i|\Lambda) \equiv \xi$ be a hyperparameter modeling the prior probability of model A for a segment. If ξ constitutes the only hyperparameter, e.g., $P(d_i|A_i, \Lambda) = P(d_i|A_i)$ (and similarly for the denominator), then

$$\mathcal{O}_{A_i/B_i}(\boldsymbol{d}) = \frac{\mathcal{L}(d_i|A_i)}{\mathcal{L}(d_i|B_i)}\frac{\int \xi\pi(\xi|\boldsymbol{d}_{k \neq i})d\xi}{\int (1-\xi)\pi(\xi|\boldsymbol{d}_{k \neq i})d\xi} \tag{A6}$$

$$= \frac{\mathcal{L}(d_i|A_i)}{\mathcal{L}(d_i|B_i)}\frac{E[\xi]}{1 - E[\xi]}, \tag{A7}$$

where $E[\xi]$ is the expectation value of ξ. The first factor here is the usual Bayes factor in support of signal in the ith segment, while the second factor is the prior odds conditioned on all other data. In the limit of small $E[\xi]$, this is approximately $E[\xi]$. This expression agrees with how the prior odds is typically defined.

Second, consider the case where model A is a signal in addition to white Gaussian noise with an unknown standard deviation σ. Then, we hyperparametrize by allowing $\sigma \sim \text{Normal}(\mu_\sigma, \sigma_\sigma)$. If the number of segments of data in $\boldsymbol{d}_{i \neq k}$ is suitably large, then σ can be precisely estimated from the contextual data. In the language of our hyperparameters, this would imply that $\sigma_\sigma/\mu_\sigma \ll 1$ and the term $\pi(\mu_\sigma, \sigma_\sigma|\boldsymbol{d}_{k \neq i}) \approx \delta(\sigma - \mu_\sigma)$. The consequence would be that $\mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(\boldsymbol{d}) \approx \mathcal{O}_{\mathcal{S}_i/\mathcal{N}_i}(d_i)$ with a prior $\sigma = \mu_\sigma$ (i.e., the precisely measured value taken from all the other data).

APPENDIX B: RECYCLING INFERENCE

In the case of a model $\mathcal{H}$ with model parameters θ, for some choice of hyperparameters Λ, the hypothesis evidence is given by

$$\mathcal{L}(d|\Lambda, \mathcal{H}) = \int d\theta \, \mathcal{L}(d|\theta, \Lambda, \mathcal{H}) \pi(\theta|\mathcal{H}). \quad \text{(B1)}$$

Typically, θ can be of high dimension, and this integration itself, done numerically, can take from a few minutes to many hours depending on the problem in hand. If one then wants to calculate as part of a posterior inference over Λ, say, this could be computationally challenging.

Instead, consider that we compute this marginalization once at a fixed value of $\Lambda = \Lambda'$ and we have a set of samples $\{\Theta_{i,k}\}$ (where i labels the data segment and k the sample number) and an evidence estimate $\mathcal{L}(d|\Lambda', \mathcal{H})$. Then, since Λ is a hyperparameter, $\mathcal{L}(d|\theta, \Lambda, \mathcal{H}) = \mathcal{L}(d|\theta, \Lambda', \mathcal{H})$ because θ is fully specified. Therefore

$$\mathcal{L}(d|\Lambda, \mathcal{H}) = \int d\theta \, \mathcal{L}(d|\theta, \Lambda', \mathcal{H}) \pi(\theta|\Lambda, \mathcal{H}). \quad \text{(B2)}$$

Applying Bayes rule to expand the likelihood dependent on Λ', noting that $\mathcal{L}(\Lambda'|d, \theta) = \mathcal{L}(\Lambda'|d)$ and applying Bayes rule again, we find that

$$\mathcal{L}(d|\Lambda, \mathcal{H}_j) = \mathcal{L}(d|\Lambda') \int d\theta \, \mathcal{L}(\theta|d) \frac{\pi(\theta|\Lambda)}{\pi(\theta|\Lambda')}. \quad \text{(B3)}$$

The first term here is the evidence at the fixed value of Λ' that we have already computed. The second term, given the samples generated in computing the evidence can be approximated by an average over the k samples

$$P(d|\Lambda, \mathcal{H}_j) = P(d|\Lambda') \left\langle \frac{P(\{\Theta_{i,k}\}|\Lambda)}{P(\{\Theta_{i,k}\}|\Lambda')} \right\rangle_k. \quad \text{(B4)}$$

This allows for rapid evaluation of the likelihood, without having to performed calculations of the high-dimensional nested integrals. One alternative to this method is to apply a kernel-density estimation method; see, e.g., Pitkin *et al.* [54].

We note that, while the above discussion was given assuming that the samples where computed at a fixed value of Λ, the same general principle applies given any prior choice (even one that did not arise from the same family) [51].

[1] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015).
[2] B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 241102 (2016).
[3] B. P. Abbott *et al.*, Phys. Rev. Lett. **121**, 161101 (2018).
[4] S. De, D. Finstad, J. M. Lattimer, D. A. Brown, E. Berger, and C. M. Biwer, Phys. Rev. Lett. **121**, 091102 (2018).
[5] LIGO Scientific and Virgo Collaborations, Astrophys. J. Lett. **882**, L24 (2019).
[6] R. Smith and E. Thrane, Phys. Rev. X **8**, 021019 (2018).
[7] B. P. Abbott *et al.*, Astrophys. J. **839**, 12 (2017).
[8] M. Pitkin *et al.*, arXiv:1705.08978.
[9] J. Powell, M. Szczepanczyk, and I. S. Heng, Phys. Rev. D **96**, 123013 (2017).
[10] J. Aasi *et al.*, Classical Quantum Gravity **32**, 115012 (2015).
[11] F. Acernese *et al.*, Classical Quantum Gravity **32**, 024001 (2015).
[12] L. Blackburn *et al.*, Classical Quantum Gravity **25**, 184004 (2008).
[13] M. Cabero, A. Lundgren, A. H. Nitz, T. Dent, D. Barker, E. Goetz, J. S. Kissel, L. K. Nuttall, P. Schale, R. Schofield, and D. Davis, Classical Quantum Gravity **36**, 155010 (2019).
[14] J. Powell, Classical Quantum Gravity **35**, 155017 (2018).
[15] B. P. Abbott *et al.*, Classical Quantum Gravity **33**, 134001 (2016).
[16] LIGO Scientific and Virgo Collaborations, arXiv:1908.11170.
[17] A. Effler, R. M. S. Schofield, V. V. Frolov, G. González, K. Kawabe, J. R. Smith, J. Birch, and R. McCarthy, Classical Quantum Gravity **32**, 035017 (2015).
[18] B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016).
[19] B. P. Abbott *et al.*, Phys. Rev. D **93**, 122003 (2016).
[20] B. P. Abbott *et al.*, Phys. Rev. D **93**, 122004 (2016).
[21] B. P. Abbott *et al.*, Classical Quantum Gravity **33**, 134001 (2016).
[22] B. P. Abbott *et al.*, Phys. Rev. D **77**, 062002 (2008).
[23] B. P. Abbott *et al.*, Classical Quantum Gravity **24**, 5343 (2007).
[24] S. A. Usman *et al.*, Classical Quantum Gravity **33**, 215004 (2016).
[25] S. Hooper, S. K. Chung, J. Luan, D. Blair, Y. Chen, and L. Wen, Phys. Rev. D **86**, 024012 (2012).
[26] Q. Chu, Low-latency detection and localization of gravitational waves from compact binary coalescences, Ph. D. thesis, University of Western Australia, 2017.
[27] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and M. Millhouse, Phys. Rev. D **94**, 044050 (2016).
[28] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Phys. Rev. D **95**, 104046 (2017).
[29] S. Klimenko *et al.*, Phys. Rev. D **93**, 042004 (2016).
[30] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, London, 1993).
[31] B. P. Abbott *et al.*, Living Rev. Relativity **21**, 3 (2018).
[32] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 131102 (2016).
[33] F. Hernandez Vivanco, R. Smith, E. Thrane, and P. D. Lasky, Phys. Rev. D **100**, 043023 (2019).
[34] J. Abadie *et al.* Phys. Rev. D **85**, 082002 (2012).
[35] M. Wąs *et al.*, Classical Quantum Gravity **27**, 015005 (2010).
[36] K. Cannon, C. Hanna, and D. Keppel, Phys. Rev. D **88**, 024025 (2013).

[37] K. Cannon, C. Hanna, and J. Peoples, arXiv:1504.04632.

[38] C. Messick *et al.*, Phys. Rev. D **95**, 042001 (2017).

[39] S. Sachdev *et al.*, arXiv:1901.08580.

[40] C. Hanna *et al.*, arXiv:1901.02227.

[41] C. Capano *et al.*, Phys. Rev. D **96**, 082002 (2017).

[42] G. Ashton *et al.*, Astrophys. J. **860**, 6 (2018).

[43] LIGO Scientific and Virgo Collaborations, Phys. Rev. X **9**, 031040 (2019).

[44] W. M. Farr *et al.*, Phys. Rev. D **91**, 023005 (2015).

[45] S. J. Kapadia *et al.*, arXiv:1903.06881.

[46] S. M. Gaebel *et al.*, Mon. Not. R. Astron. Soc. **484**, 4008 (2019).

[47] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, arXiv:1904.07214.

[48] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **120**, 091101 (2018).

[49] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, Boca Raton, FL, 2013).

[50] LIGO Scientific and Virgo Collaborations *et al.*, Astrophys. J. Lett. **882**, L24 (2019).

[51] E. Thrane and C. Talbot, Pub. Astron. Soc. Aust. **36**, e010 (2019).

[52] J. Skilling, in *American Institute of Physics Conference Series*, edited by R. Fischer, R. Preuss, and U. V. Toussaint, American Institute of Physics Conference Series Vol. 735 (AIP Publishing, Melville, NY, 2004) pp. 395–405.

[53] C. Talbot and E. Thrane, Astrophys. J. **856**, 173 (2018).

[54] M. Pitkin, C. Messenger, and X. Fan, Phys. Rev. D **98**, 063001 (2018).

[55] J. Veitch and A. Vecchio, Phys. Rev. D **81**, 062003 (2010).

[56] M. Isi *et al.*, Phys. Rev. D **98**, 042007 (2018).

[57] N. J. Cornish and T. B. Littenberg, Classical Quantum Gravity **32**, 135012 (2015).

[58] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).

[59] J. S. Speagle, arXiv:1904.02180.

[60] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, Astrophys. J. **872**, 195 (2019).