# Population-informed priors in gravitational-wave astronomy

Christopher J. Moore[1, *] and Davide Gerosa[1, 2, 3]

[1]*Institute for Gravitational Wave Astronomy & School of Physics and
Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[2]*Dipartimento di Fisica "G. Occhialini", Universitá degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*
[3]*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*
(Dated: November 11, 2021)

We describe a Bayesian formalism for analyzing individual gravitational-wave events in light of the rest of an observed population. This analysis reveals how the idea of a "population-informed prior" arises naturally from a suitable marginalization of an underlying hierarchical Bayesian model which consistently accounts for selection effects. Our formalism naturally leads to the presence of "leave-one-out" distributions which include subsets of events. This differs from other approximations, also known as empirical Bayes methods, which effectively double count one or more events. We design a double-reweighting post-processing strategy that uses only existing data products to reconstruct the resulting population-informed posterior distributions. Although the correction we highlight is an important conceptual point, we find it has a limited impact on the current catalog of gravitational-wave events. Our approach further allows us to study, for the first time in the gravitational-wave literature, correlations between the parameters of individual events and those of the population.

## I. INTRODUCTION

Bayesian statistics plays a prominent role in gravitational-wave (GW) astronomy, where it is routinely used to infer the properties of individual binary black hole (BH) events [1, 2]. Bayesian statistics is also used to infer the properties of the underlying distribution of sources [3, 4], assuming all events come from the same modeled population (which can be a mixture of several channels; for reviews see [5, 6]). Although individual-event and population inferences are often treated separately for practical and computational purposes, they can be viewed as two sides of the same coin: namely a full, hierarchical Bayesian model.

Hierarchical Bayesian models have been successfully applied to many astronomical data sets, including spectroscopic data for the determination of stellar ages [7], light curve [8] and radial velocity [9] data for the determination of exoplanet obliquities and eccentricities respectively, and astroseismic data for the determination of stellar inclinations [10] and helium enrichment [11]. One benefit of hierarchical Bayesian models is that one can obtain improved measurements of the parameters of individual events by exploiting the fact that they are part of a large catalog—an approach that can be described as using a *population-informed prior*. Taking the first GW event as an example, this line of reasoning is equivalent to asking:

> What can be learned about GW150914 using not only the $\sim 0.2\,\mathrm{s}$ of data from September 14th, 2015, but rather from all of the LIGO/Virgo observations to date?

We now describe several different, but related analyses. The interplay between these procedures can be visualized using the probabilistic graphical models (PGMs) in Fig. 1. In these diagrams, data and parameters are indicated with circles while arrows represent conditional probabilistic dependencies between quantities. Figure 1(a) illustrates the standard single-event parameter estimation (e.g. [1, 2]): one selects an event (index $i$) from the catalog and infers its parameters $\theta_i$ (masses, spins, etc.) using only the data $d_i$ for that event and an uninformative prior. Figure 1(b) illustrates the standard population analysis (e.g. [3, 4]): hyperparameters $\lambda$ describing the source population (e.g. the slope of the mass spectrum) are inferred using the results of all of the single-event analyses, taking care to properly account for selection effects. In this approach, the single-event parameters are marginalized over and, therefore, cannot be sampled. Figure 1(c) illustrates a population-informed single-event analysis, which is the main topic of this paper. Targeting event $j$, this analysis uses the data for all other events, $\{d_{i\neq j}\}$, to infer the population parameters $\lambda$ which are then marginalized over, while incorporating the data $d_j$, to infer the parameters $\theta_j$. As is shown below, the analyses depicted in panels (b) and (c) are appropriate marginalizations of the full hierarchical Bayesian model indicated in Fig. 1(d).

This paper describes a complete formalism for performing population-informed single-event analyses in the presence of selection effects, showing in particular how it follows from a full hierarchical Bayesian model (Sec. II). Our solution differs from some previous studies that rely on heuristic derivations and that sometimes mistreat selection effects and/or implicitly apply an empirical Bayes method [12]—a known approximation of a hierarchical Bayesian analysis. We present a practical implementation of a population-informed single-event analysis which relies on a double-reweighting procedure and makes use of existing data products (Sec. III). Finally, we apply our formalism to the current LIGO/Virgo catalog (Sec. IV). We obtain population-informed posterior distributions for all the events in the catalog and consider, for the first

* moorecj@bham.ac.uk

**(a)** Single-event parameter estimation

**(b)** Population inference

**(c)** Population-informed single-event inference
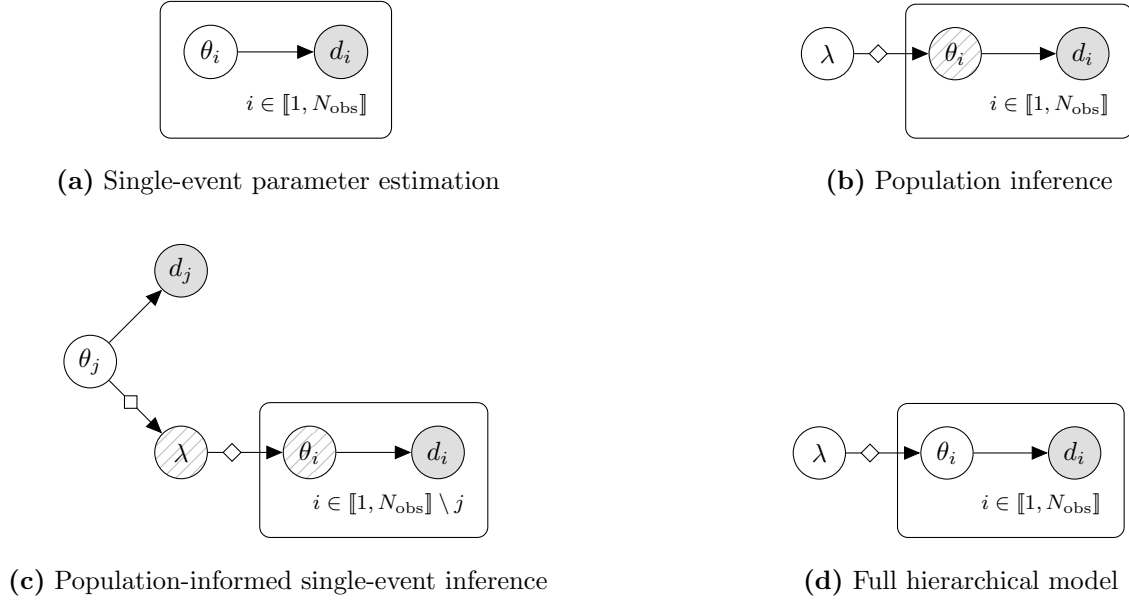
**(d)** Full hierarchical model

FIG. 1. Probabilistic graphical models (PGMs) for the analyses described in the introduction. Observed data are indicated with gray circles; empty circles indicate model parameters one wishes to sample; hatched circles indicate model parameters that are marginalized over in the analysis; boxes indicate parts of the analysis that are repeated independently for multiple events; diamonds indicate steps in the analysis where selection effects must be taken into account. *Panel (a): Single-event parameter estimation.* It is common practice for every event to be analyzed individually under an uninformative choice of prior, $\pi_{\mathrm{PE}}(\theta)$; the target distribution for this analysis is given in Eq. (15). *Panel (b): Population inference.* When interested solely in the population parameters, these can be inferred using posterior samples from the single-event analyses; the target distribution for this analysis is given in Eq. (8). The $\theta_i$ parameters are analytically marginalized over and therefore cannot be inferred from this analysis. Note how the PGM from panel (a) is contained within this diagram. *Panel (c): Population-informed single-event inference.* In this case, single events are analyzed in light of the whole population; the target distribution for this analysis is given in Eq. (11). Note how the PGM from panel (b) is contained within this diagram but with one event omitted and $\lambda$ marginalised out. *Panel (d): Full hierarchical model.* In principle, the population can be analyzed simultaneously with all of the events in a full hierarchical Bayesian model; the target distribution is given in Eq. (2). The analyses in panels (b) and (c) can be obtained by marginalising this over $\{\theta_i\}$ (see Sec. II C) and $(\{\theta_{i\neq j}\}, \lambda)$ (see Sec. II D), respectively.

time in GW astronomy, the correlations between event and population parameters. We discuss the prospects of our work in Sec. V and present some generalizations in Appendices A, B and C. Additional results are provided as supplemental material.

Although we restrict ourselves to GW astronomy, our statistical methods are very general and can be applied to any scenario where individual observations need to be analyzed as part of a larger set while consistently accounting for their intrinsic detectability.

## II. HIERARCHICAL BAYESIAN INFERENCE

### A. Notation

Let $i \in [\![1, N_{\mathrm{obs}}]\!]$ label observed events in a catalog and $d_i$ denote the strain for event $i$. Events are described by parameters $\theta_i$ (e.g. BH masses and spins). The collection of all $N_{\mathrm{obs}}$ parameters and strain data are denoted $\{\theta_i\}$ and $\{d_i\}$ respectively. It will be necessary to use sets with one specific event, say $j$, is omitted; these are denoted

$\{\theta_{i\neq j}\}$ and $\{d_{i\neq j}\}$.

We assume that we have an astrophysical population model ("pop") which depends on parameters $\lambda$ and predicts an expected number of sources $N(\lambda)$ distributed such that the number with parameters in volume $\mathrm{d}\theta$ is given by

$$\frac{\mathrm{d}N}{\mathrm{d}\theta} = N(\lambda)p_{\mathrm{pop}}(\theta|\lambda), \quad (1)$$

with $\int \mathrm{d}\theta\, p_{\mathrm{pop}}(\theta|\lambda) = 1$. It is convenient to reparameterize $\lambda = (N, \lambda)$, separating out one parameter $N$ that describes the rate from the remaining $\lambda$ that describe the shape of the population (e.g. the slope of the mass function and locations of any mass gaps). With this change of variables, one has $N(\lambda) = N$ and $p_{\mathrm{pop}}(\theta|\lambda) = p_{\mathrm{pop}}(\theta|\lambda)$.

### B. Full hierarchical model

Given all the observed data $\{d_i\}$, what we would like to do is to simultaneously infer the properties of all the individual events $\{\theta_i\}$ and the population parameters

$\lambda$. The posterior on these parameters is given by the following hierarchical Bayesian model,

$$P_{\text{pop}}(\{\theta_i\}, \lambda | \{d_i\}) = \frac{\mathcal{L}(\{d_i\}|\{\theta_i\})p(\{\theta_i\}|\lambda)\pi_{\text{pop}}(\lambda)}{\mathcal{Z}_{\text{pop}}}. \quad (2)$$

where $\mathcal{Z}_{\text{pop}} = P_{\text{pop}}(\{d_i\})$ is the evidence.[1]

If events are independent and non-overlapping, then the first term in the numerator of Eq. (2) is the product of the individual event likelihoods,

$$\mathcal{L}(\{d_i\}|\{\theta_i\}) = \prod_{i=1}^{N_{\text{obs}}} \mathcal{L}(d_i|\theta_i), \quad (3)$$

where $\mathcal{L}(d_i|\theta_i)$ is the usual single-event GW likelihood (e.g. [13]).

The second term describe an in-homogeneous Poisson process and involves a factor of the population model for each event (cf. [14])

$$p(\{\theta_i\}|\lambda) \propto e^{-N\alpha(\lambda)} N^{N_{\text{obs}}} \prod_{i=1}^{N_{\text{obs}}} p_{\text{pop}}(\theta_i|\lambda), \quad (4)$$

where we again split $\lambda = (N, \lambda)$ and discard normalization terms that do not depend on $\{\theta_i\}$ or $\lambda$. The *efficiency*

$$\alpha(\lambda) = \int d\theta \, P_\theta(\det)p_{\text{pop}}(\theta|\lambda). \quad (5)$$

is the fraction of events in the population that are detectable and accounts for selection effects via the inclusion of a detection probability for a given event.[2]

The final term in Eq. (2) is the Bayesian prior on the population parameters, $\pi_{\text{pop}}(\lambda)$.

As the focus of this study is mainly on the shape of the population, not on the event rate, here we will restrict to the case where the (improper) prior on $N$ is scale invariant, i.e. $\pi_{\text{pop}}(\lambda) \propto \pi_{\text{pop}}(\lambda)/N$, and marginalize over $N$. A generalized derivation that includes $N$ and is suitable for inferring on the rate of events is presented in Appendix A. Marginalizing over $N$, the posterior in Eq. (2) becomes

$$P_{\text{pop}}(\{\theta_i\}, \lambda | \{d_i\}) \propto \mathcal{L}(\{d_i\}|\{\theta_i\})p(\{\theta_i\}|\lambda)\pi_{\text{pop}}(\lambda). \quad (6)$$

where

$$p(\{\theta_i\}|\lambda) \propto \alpha(\lambda)^{-N_{\text{obs}}} \prod_{i=1}^{N_{\text{obs}}} p_{\text{pop}}(\theta_i|\lambda). \quad (7)$$

The logical structure of this full hierarchical model is indicated graphically in Fig. 1(d). This model is impractical to sample from directly due to its high dimensionality and the computational cost of its evaluation. In practice, two of its marginalized distributions are used, as described in the following subsections.

———

[1] Hereafter, a subscript is used to denote a conditional probability. E.g. $P_{\text{pop}}(\{d_i\}) \equiv P(\{d_i\}|\text{pop})$ is the probability of observing the data given a particular astrophysical model for the population.

[2] The detection probability $0 \le P_\theta(\det) \le 1$ is the conditional probability of detecting an event given its parameters; i.e. $P(\det|\theta)$, hence the subscript $\theta$. Other authors write this as $p_{\det}(\theta)$.

## C. Population inference

First, consider marginalizing Eq. (2) over all the $\{\theta_i\}$. This is what is typically done GW population inference [3, 4] and leaves a posterior on just the population parameters

$$P_{\text{pop}}(\lambda|\{d_i\}) \propto \int d\{\theta_i\} \, P_{\text{pop}}(\{\theta_i\}, \lambda|\{d_i\}), \quad (8)$$

$$\propto \frac{\pi_{\text{pop}}(\lambda)}{\alpha(\lambda)^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} \int d\theta_i \, \mathcal{L}(d_i|\theta_i)p_{\text{pop}}(\theta_i|\lambda).$$

In order to sample this distribution, it is necessary to evaluate the $\theta_i$-integrals; in practice this is done using parameter estimation samples, see Sec. III B. The logical structure of this analysis is indicated graphically in Fig. 1(b).

## D. Population-informed single-event inference

Second, consider marginalizing Eq. (2) over $\lambda$ and $\{\theta_{i\neq j}\}$, resulting in a posterior on just the parameters for event $j$. Using Eqs. (3), (6) and (7) and bringing the terms that depend on $\theta_j$ in front of the product on $i$ yields

$$P_{\text{pop}}(\theta_j|\{d_i\}) = \int d\lambda \int d\{\theta_{i\neq j}\} \, P_{\text{pop}}(\{\theta_i\}, \lambda|\{d_i\}) \quad (9)$$

$$= \mathcal{L}(d_j|\theta_j) \int d\lambda \, \frac{p_{\text{pop}}(\theta_j|\lambda)}{\alpha(\lambda)} P_{\text{pop}}(\lambda|\{d_{i\neq j}\}),$$

where $P_{\text{pop}}(\lambda|\{d_{i\neq j}\})$ is given by Eq. (8) but with event $j$ omitted from the catalog; i.e.

$$P_{\text{pop}}(\lambda|\{d_{i\neq j}\}) \propto \frac{\pi_{\text{pop}}(\lambda)}{\alpha(\lambda)^{N_{\text{obs}}-1}} \prod_{i\neq j} \int d\theta_i \, \mathcal{L}(d_i|\theta_i)p_{\text{pop}}(\theta_i|\lambda). \quad (10)$$

This is the typical end-product of the so-called "leave-one-out" analyses, where individual events are excluded from the catalog. These analyses can be used as part of a posterior-predictive test for the presence of outlying events from the main population. For instance, some of the analyses in Refs. [3, 4] were performed excluding either GW170729, GW190521, or GW190814.

Equation (9) can be rewritten in a suggestive manner:

$$P_{\text{pop}}(\theta_j|\{d_i\}) \propto \mathcal{L}(d_j|\theta_j)\varpi(\theta_j|\{d_{i\neq j}\}), \quad (11)$$

where

$$\varpi(\theta_j|\{d_{i\neq j}\}) = \int d\lambda \, \frac{p_{\text{pop}}(\theta_j|\lambda)}{\alpha(\lambda)} P_{\text{pop}}(\lambda|\{d_{i\neq j}\}). \quad (12)$$

This now resembles Bayes' theorem where $\varpi$ plays the role of a *population-informed prior* (which also incorporates selection effects) for the parameters of event $j$. Crucially, this expression relies on the leave-one-out posterior $P_{\text{pop}}(\lambda|\{d_{i\neq j}\})$, thus avoiding double-counting the

event $j$ in Eq. (11). Given the properties of the $N_{\mathrm{obs}} - 1$ events we have observed so far, and what we know about the sensitivity of my instrument, the population-informed prior $\varpi$ quantifies what we expect the $N_{\mathrm{obs}}^{\mathrm{th}}$ event to look like. Although we informally refer to $\varpi$ as a prior distribution, this analogy must be used carefully because the quantity $\varpi(\theta)$ is not normalized.

The logical structure of this analysis is indicated graphically in Fig. 1(c). First $N_{\mathrm{obs}} - 1$ events are used to learn about the population, this information is encoded in the distribution $P_{\mathrm{pop}}(\lambda|\{d_{i\neq j}\})$. This is used, along with information about selection effects, to build the (pseudo) prior distribution $\varpi(\theta_j|\{d_{i\neq j}\})$ for the parameters of the final event. Finally, this prior is combined with the likelihood for the final event to obtain the posterior distribution $P_{\mathrm{pop}}(\theta_j|\{d_i\})$ on those event parameters.

It is also important to remember that population-informed reanalysis of event $j$ is necessarily conditioned on a particular model for the population.

### E.  Empirical Bayes method

A important comment is due about the distribution $P_{\mathrm{pop}}(\theta_j|\{d_i\})$ in Eq. (11): care must be taken not to include event $j$ twice in the inference; it is for this reason that the integrals in Eqs. (10) and (12) must be taken over the leave-one-out distribution $P_{\mathrm{pop}}(\lambda|\{d_{i\neq j}\})$ which only use information from the *other* events in the catalog.

One could attempt to use population-informed priors *without* omitting the event from the prior. This is a known approach to approximate the outcome of a hierarchical analysis and is often referred to as "empirical Bayes" [12]. In our context, the empirical Bayes approximation of the single-event parameters reduces to using

$$
P_{\mathrm{pop}}(\theta_j|\{d_i\}) \approx \mathcal{L}(d_j|\theta_j) \int \mathrm{d}\lambda \, p_{\mathrm{pop}}(\theta_j|\lambda) P_{\mathrm{pop}}(\lambda|\{d_i\})
\tag{13}
$$

instead of Eq. (9). The PGM for an analysis based on this expression looks identical to that of Fig. 1(c) with the exception that the event $j$ is not excluded from the box when iterating over $i$. We stress that Eq. (13) does not follow from the hierarchical Bayesian model in Eq. (2) and, compared to Eq. (9), double counts event number $j$.

Assuming event $j$ is not a population outlier, we would expect the empirical Bayes approximation to become increasingly accurate as the size of the GW catalog increases. As we will see below in Sec. IV, it seems that we are already in this regime and the empirical Bayes method generally provides a good approximation to the full hierarchical analysis. However, it is an important conceptual point that event $j$ is being double-counted in this analysis (it enters once in the population-informed prior and once in the likelihood) and that this is only an approximation to the full hierarchical Bayesian model. This conceptual point is in some ways analogous to Bessel's $N/(N-1)$ cor-

rection factor in the frequentist estimation of the variance in a population.

There have been some previous attempts at using population-informed priors in the GW context. The literature on this topic is rather opaque and it is often unclear which expressions are actually being used. Careful reading of Refs. [4, 15, 16] suggests that the empirical Bayes is implicitly being used and that one event is erroneously being double-counted, although private discussions with some of the authors indicate that this is in fact not the case. Ref. [17] considers a population of GW events modeled using a restricted set of parameters but does explicitly use the correct leave-one-out expression in Eq. (9). The heuristic expressions for the population-informed prior reported in Refs. [18, 19] do not double-count any events, however they differ in the treatment of selection effects by not including the extra factor of $\alpha(\lambda)$ in Eq. (9); therefore, those treatments do not follow from a hierarchical Bayesian analysis, although in practice similar numerical results are obtained. Finally there are some unpublished technical documents [20–22] which also describe the leave-one-out population-informed posterior in Eq. (9) together with a condition-reweighting approach to sampling this distribution which differs from the double-reweighting strategy put forward in this paper.

### F.  Correlations between event parameters and population parameters

The full hierarchical model of Eq. (2) contains information on the correlations between the event parameters $\{\theta_i\}$ and the population parameters $\lambda$. These correlations are lost if one pursues the common population approach described in Sec. II C where the event parameters are marginalized over. The correlations are also lost when one pursues the population-informed single-event analysis described in Sec. II D (or the empirical Bayes approximation described in Sec. II E) as this analysis marginalizes over the population parameters.

In order to study correlations between individual event parameters $\theta_j$ and the population parameters $\lambda$ we can marginalize the full hierarchical model of Eq. (2) over $\{\theta_{i\neq j}\}$. This gives

$$
P_{\mathrm{pop}}(\theta_j, \lambda|\{d_i\}) = \mathcal{L}(d_j|\theta_j) \frac{p_{\mathrm{pop}}(\theta_j|\lambda)}{\alpha(\lambda)} P_{\mathrm{pop}}(\lambda|\{d_{i\neq j}\}).
\tag{14}
$$

This distribution can be used to assess the extent to which certain parameters from specific events might be affecting the population parameters, see Sec. IV D.

## III.  SAMPLING AND REWEIGHTING

A straightforward implementation of the population-informed individual-event analysis in Sec. II D would require $N_{\mathrm{obs}}$ dedicated population inference runs to be

performed, one leaving out each event in turn. This is clearly impractical, especially as $N_{\text{obs}}$ becomes large. In this section we present a double-reweighting strategy that avoids the needs for these expensive calculations.

First, we briefly describe the two analyses that are commonly performed in GW astronomy: single-event parameter estimation (Sec. III A) and population inference (Sec. III B). We then show how the results of these can be used to obtain weighted samples from the full population-informed prior analysis in Eq. (11) via a double-reweighting strategy (Sec. III C).

### A. Parameter estimation

Parameter estimation (PE) is routinely performed on individual GW events using an uninformative[3] prior, $\pi_{\text{PE}}(\theta)$; for instance, common choices include priors that are uniform in masses and isotropic in spins. The target posterior for this analysis is given by,

$$P_{\text{PE}}(\theta_i|d_i) \propto \mathcal{L}(d_i|\theta_i)\pi_{\text{PE}}(\theta_i). \quad (15)$$

A subscript PE indicates that a probability is conditioned on the assumptions in the uninformative prior used in single-event PE. The simple logical structure of this analysis indicated graphically in Fig. 1(a).

An output of the PE analysis is a set of (equally weighted;[4] $w_i^k = 1$) posterior samples,

$$(\theta_i^k, w_i^k = 1) \sim P_{\text{PE}}(\theta_i|d_i), \quad \text{for } k = 1, 2, \ldots, S_i. \quad (16)$$

### B. Population inference

Higher-level population analyses are performed with the target posterior distribution $P_{\text{pop}}(\lambda|\{d_i\})$ described in Sec. II C. Again, the subscript pop is there to remind us that this is a probability conditioned on the assumptions in a particular population model.

In practice, the samples from the PE analyses of individual events are used to efficiently evaluate the population likelihood. The integrals in Eq. (8) are usually approximated by the following Monte Carlo sums using Eq. (15) and the samples in Eq. (16),

$$\int \mathrm{d}\theta_i \, \mathcal{L}(d_i|\theta) p_{\text{pop}}(\theta_i|\lambda) \propto \int \mathrm{d}\theta_i \, P_{\text{PE}}(\theta_i|d_i) \frac{p_{\text{pop}}(\theta_i|\lambda)}{\pi_{\text{PE}}(\theta_i)}$$

$$\approx \frac{1}{S_i} \sum_{k=1}^{S_i} \frac{p_{\text{pop}}(\theta_i^k|\lambda)}{\pi_{\text{PE}}(\theta_i^k)}. \quad (17)$$

This is a reweighting procedure, where the ratio in the integrand divides by the original PE prior and multiplies by the desired population model.

An output of this analysis (e.g. [4]), is a set posterior samples (again, equally weighted, $\omega^l = 1$) drawn from the target distribution;

$$(\lambda^l, \omega^l = 1) \sim P_{\text{pop}}(\lambda|\{d_i\}), \quad \text{for } l = 1, 2, \ldots, \mathcal{S}. \quad (18)$$

### C. Reweighting samples for population-informed single-event inference

We take the samples $\theta_j^k$ from Eq. (16) and reweight them to the population-informed target distribution in Eq. (11) for a single event;

$$(\theta_j^k, W^k) \sim P_{\text{pop}}(\theta_j|\{d_i\}). \quad (19)$$

The necessary weights are given by the ratio of the target distribution to the current PE distribution,

$$W^k = \frac{P_{\text{pop}}(\theta_j^k|\{d_i\})}{P_{\text{PE}}(\theta_j^k|d_j)}. \quad (20)$$

Using Eqs. (10), (11) and (12) for the numerator, and Eq. (15) for the denominator, this expression simplifies to

$$W^k = \frac{1}{\pi_{\text{PE}}(\theta_j^k)} \int \mathrm{d}\lambda \, \frac{P_{\text{pop}}(\lambda|\{d_i\})p_{\text{pop}}(\theta_j^k|\lambda)}{\int \mathrm{d}\theta_j \, \mathcal{L}(d_j|\theta_j)p_{\text{pop}}(\theta_j|\lambda)}. \quad (21)$$

The integral in the denominator is the same as that in Eq. (8) and can be approximated by the Monte Carlo sum in Eq. (17). The integral in the numerator can also be be approximated by a Monte Carlo sum using the population samples in Eq. (18). Doing so gives

$$W^k = \frac{1}{\pi_{\text{PE}}(\theta_j^k)\mathcal{S}} \sum_{l=1}^{\mathcal{S}} \frac{p_{\text{pop}}(\theta_j^k|\lambda^l)S_j}{\sum_{k'=1}^{S_j} \frac{p_{\text{pop}}(\theta_j^{k'}|\lambda^l)}{\pi_{\text{PE}}(\theta_j^{k'})}}. \quad (22)$$

The samples $\lambda^l$ were obtained from a likelihood that involves a reweighting of the PE samples, $\theta_i^k$. We are now using the $\lambda^l$ samples to reweight the original PE samples $\theta_i^k$ to target the population-informed single-event posterior. In this sense this procedure is a double reweighting of existing posterior samples.

Via an entirely analogous calculation, an additional set of weights can be obtained to approximate $P_{\text{pop}}(\theta_j, \lambda|\{d_i\})$ of Eq. (14) and thus capture correlations between parameters and hyperparameters.

Finally, one can also attempt a similar reweighting targeting the full hierarchical model of Sec. II B (this involves reweighting samples on $\lambda$ and $\{\theta_i\}$ simultaneously). We have attempted this and found that it is not possible to do accurately with given existing LIGO/Virgo data products. The much higher dimensionality of the distribution means that results are dominated by errors due to the finite number of samples available. Although we

---

[3] All priors introduce some information in the analysis. In this context "uninformative" indicate a prior choice that does not make use of data from the other events in the catalog.

[4] We use the notation $(x^a, w^a) \sim P(x)$, $a = 1, 2, \ldots, N$, to indicate weighted samples from distribution $P(x)$. The importance sampling estimate for the expectation of a function of $x$ is given by $\mathrm{E}[f(x)] = \sum_{a=1}^{N}(w^a f(x^a))/\sum_{a=1}^{N}(w^a)$.

are reporting a null result, we provide the expressions for the full Hierarchical reweighting in Appendix B and plan to re-investigate this point in the future with different techniques.

## IV. APPLICATION TO THE LIGO/VIRGO CATALOG

We refer to the different analyses as follows:

- "Uninformative": the posterior $P_{PE}(\theta_j|d_j)$ in Eq. (15) obtained from single-event parameter-estimation that does not use the rest of the catalog.

- "Hierarchical": the population-informed posterior $P_{pop}(\theta_j|\{d_i\})$ of Eq. (9), which was obtained by marginalizing the full hierarchical model. In practice, we sample this using the double-reweighting procedure described in Sec. III C.

- "Empirical": the empirical Bayes approximation to the population-informed posterior $P_{pop}(\theta_j|\{d_i\})$ given in Eq. (13), where the analyzed event enter twice. This is also sampled using a double-reweighting procedure.

We use both the parameter estimation samples for $\theta$ and population inference samples for $\lambda$ provided with Ref. [4]. This analysis includes the $N_{obs} = 44$ confident binary BH events detected with false-alarm rate $< 1\text{yr}^{-1}$ during the first two and a half observing runs of LIGO/Virgo. We consider individual events to be described by the following 6 parameters: $\theta = \{m_1, m_2, \chi_1, \chi_2, \theta_1, \theta_2\}$ where $m_i$ are the BH masses, $\chi_i$ are the dimensionless BH spin magnitudes, and $\theta_i$ are the tilt angles between the BH spins and the orbital angular momentum (subscripts $i = 1, 2$ refer to the heavier and the lighter BH, respectively).

For our population model, we use the POWER LAW + PEAK model from Ref. [4] as an example; however, we stress that our method is general and can be applied to any population model. This POWER LAW + PEAK model depends on $\dim(\lambda) = 12$ population parameters which are briefly described here. The primary mass is assumed to follow a power-law distribution between $m_{max}$ and $m_{min}$ with a spectral index $\alpha$ and a Gaussian peak component designed to model a possible pileup of events below the pair-instability supernova mass gap (the peak has fractional strength $\lambda_m$ and is located at $\mu_m$ with a width $\sigma_m$). The mass ratio $q = m_2/m_1 \leq 1$ is assumed to follow a power-law distribution with index $\beta_q$. Both the primary mass and mass ratio distributions are smoothed over a range of masses $\delta_m$ at the low mass end. The dimensionless spin magnitudes are assumed to follow a beta distributions parameterised by a mean $\mu_\chi$ and a variance $\sigma_\chi^2$. The distributions of the spin tilt cosines have a uniform component and a (truncated) Gaussian component with fractional strength $\zeta$ and width $\sigma_t$ that is designed to model field formation channels that might preferentially form binaries with aligned spins. The redshift $z$ is assumed to be distributed uniformly in comoving volume and source frame time independently of $\lambda$.

### A. Catalog summary

We compute the three posterior distributions highlighted above ("uninformative", "hierarchical", and "empirical") for each of the 44 events. The complete set of posteriors for all 44 events is provided as a supplementary file. Inspecting these plots reveals that, in general, the most noticeable effect of using a population-informed prior is to reduce the error on the mass ratio $q$ measurement and shift the posterior towards $q = 1$; this is expected as most of the events in the catalog are consistent with $q = 1$ and this trend is captured by the population model. It is also noticeable that in most cases there is excellent agreement between the "Hierarchical" posterior and the "Empirical" approximation. Here we attempt to summarize the results for all these events by computing the Hellinger distance $d_H^2(p, q) \equiv 1 - \int \sqrt{p(x)q(x)}dx \in [0, 1]$ [23] between the "Uniformative" and the other two population-informed distributions. A short summary of the key properties of the Hellinger distance is presented in Appendix C. Although we find the Hellinger distance to be a useful indicator, we caution against overinterpreting of these results; it is not a substitute for visually inspecting the distributions.

Figure 2 shows the distances between the uninformative distribution and the (i) hierarchical and (ii) empirical distributions. The resulting values of $d_H$ are in the range $\sim 0.4$–0.7, indicating that the population-informed reweighting has a relative major impact on the interpretation of the systems. The events with largest $d_H$ have either large masses (GW190521, GW170729, GW190929_012149) or large effective spin (GW151226, GW190517_055101, GW190620_030421) compared to the rest of the catalog [1, 2]. In general, the events with low $d_H$ (i.e. those for which the population-informed reweighting is less impactful) have either parameters that are well in the bulk of the population or large signal-to-noise ratios such that the posterior is mostly driven by the likelihood and not the prior. GW150914 is an excellent example of this.

We find that the two distances we computed (uninformative vs hierarchical and uninformative vs empirical) are, in general, very similar to each other with differences as small as $\Delta d_H \lesssim 0.01$. This is within the errors on $d_H$, which we estimate by splitting the samples weights into 10 subsets and computing the resulting standard deviations. Overall we find that, despite being an important conceptual point, the Bessel-like correction highlighted in this paper is subdominant. Some exceptions include GW190517_055101 ($\Delta d_H \sim 0.07$, see Sec. IV C below), GW190728_064510, GW190720_000836, and GW190924_021846 ($\Delta d_H \sim 0.02$)

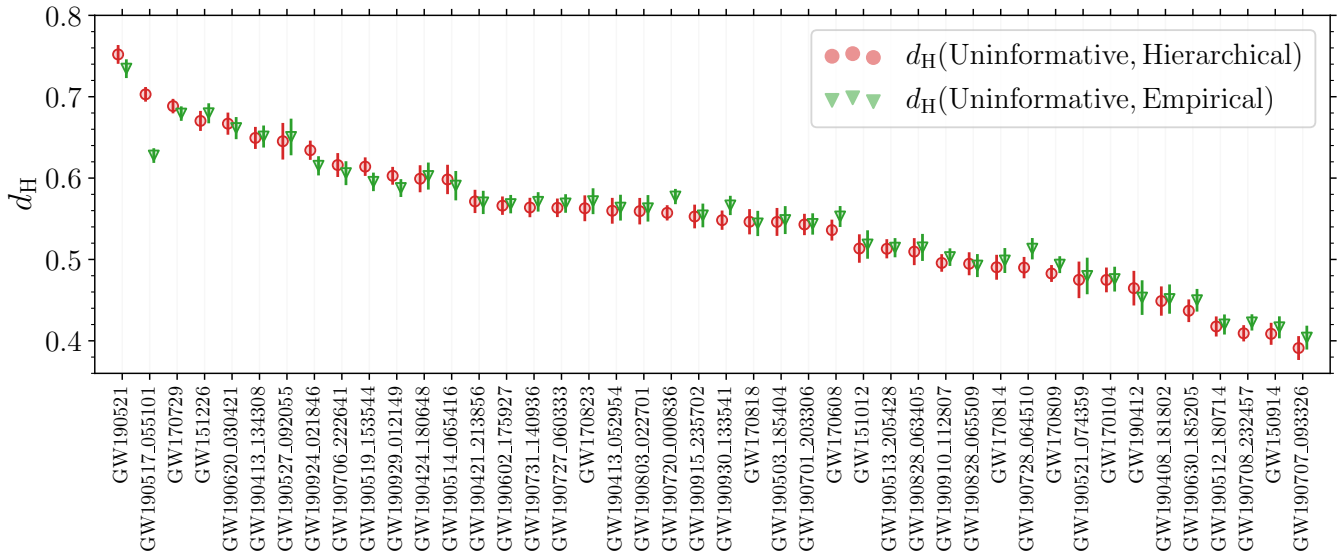Let us now focus on two cases of particular interest.

FIG. 2. The Hellinger distances between the standard parameter-estimation 6-dimensional posterior distributions on $\theta_i$ (uninformative prior) and the population-informed posteriors. Results are shown for all 44 BH binary events detected with false-alarm rate $< 1\mathrm{yr}^{-1}$ during the first two and a half LIGO/Virgo observing runs. "Hierarchical" (red circles) refers to the population-informed expression derived in this paper, cf Eq. (9); "Empirical" (green triangles) refers to the approximation of Eq. (13) which double-counts the event considered. Errors are estimated by splitting posterior samples into 10 sets and computing the standard deviation of $d_\mathrm{H}$ computed from each set.

## B. The masses of GW190521

GW190521 is the heaviest binary in the catalog. It is also the event for which the use of a population-informed prior makes the biggest difference to the posterior in the sense that it has the largest value of $d_\mathrm{H}$ in Fig. 2. Figure 3 shows the posterior distributions on its component masses $m_1$ and $m_2$. Priors that are inferred from the (rest of the) population shift the posterior distribution toward lower masses compared to the uninformative prior adopted in the single-event parameter-estimation analysis (cf. Ref. [4] for an analogous discussion). This is unsurprising because all of the other events have masses that are likely to be lower than those of GW190521. We find that the hierarchical and empirical estimate of the population-informed posterior return essentially the same result: it appears that, already with a catalog of $\sim 50$ events, the double counting of one event involved in the empirical Bayes approximation has a small effect on the result. This is compatible with the results the posterior-predictive checks reported in Ref. [4], which concluded that GW190521 is not a population outlier.

## C. The spins of GW190517_055101

GW190517_055101 is the event with the largest effective spin $\chi_\mathrm{eff}$ [24] in the catalog (this is the spin combination parameter that is currently best measured) and also present a moderate value of $\chi_\mathrm{p}$ [25] (another measured spin combination that captures the precession of the or-
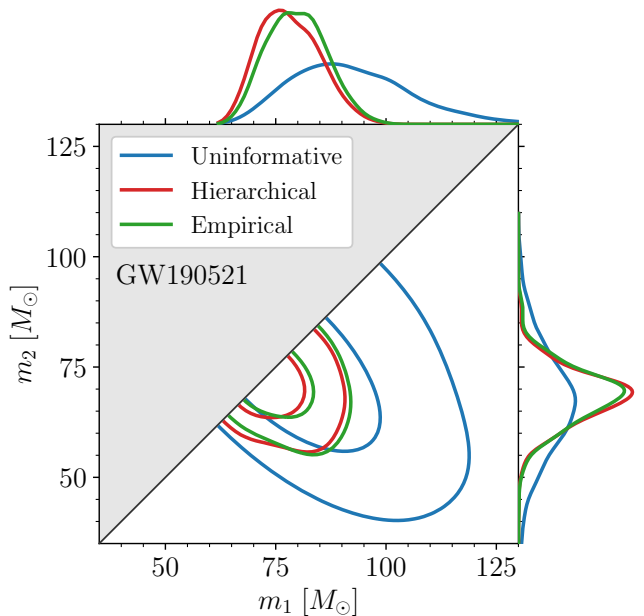


FIG. 3. The posteriors for the source-frame component masses of the heaviest binary observed to date, GW190521. The blue contours show results of the single-event parameter estimation, ["Uninformative", Eq. (15)]. The red contours show the population-informed single-event inference developed in this paper ["Hierarchical", Eq. (9)]. The green contours show the empirical Bayes approximation to the population-informed single-event distribution ["Empirical", Eq. (13)]. We show 50% and 90% contours.
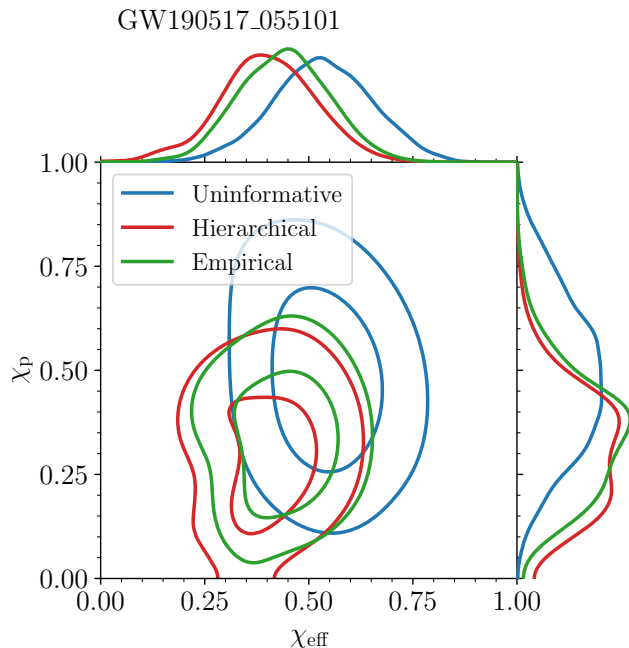
GW190517_055101

FIG. 4. The posteriors for the spin combinations $\chi_{\text{eff}}$ and $\chi_{\text{p}}$ of the binary with the largest $\chi_{\text{eff}}$ in the catalog, GW190517_055101. The blue, red and green contours show the results of the "Uninformative" [Eq. (15)], "Hierarchical" [Eq. (9)], and "Empirical" [Eq. (13)] inferences respectively. We show 50% and 90% contours.

bital plane; see Ref. [26] for issues and refinements on its definition). Figure 4 illustrates the joint distribution of these two parameters for GW190517_055101. Much like for the masses of GW190521, GW190517_055101 has higher effective spins compared to the rest of the catalog and thus the population-informed posteriors peak at lower values compared to the uninformative ones. In this case, the difference between the full hierarchical estimate and the empirical approximation is more pronounced. When the event is double counted (empirical Bayes method), the high-spin value of GW190517_055101 contaminates the population-informed prior and produces a distribution with $\chi_{\text{eff}} = 0.44^{+0.17}_{-0.18}$ (median and 90% credible interval). The result obtained from marginalizing the full hierarchical model instead returns $\chi_{\text{eff}} = 0.40^{+0.17}_{-0.19}$, which is further from the uninformative results $\chi_{\text{eff}} = 0.53^{+0.19}_{-0.20}$. However, although the maximum-a-posteriori is affected by the erroneous double counting, the difference between the peaks of the hierarchical and empirical of $\chi_{\text{eff}}$ posteriors is still smaller than the widths of the distributions. This suggests that, at least at the present signal-to-noise-ratio, the empirical Bayes approximation is appropriate. The population-informed estimates of $\chi_{\text{p}}$ peak at lower values compared to the uninformative case; this is driven by the bulk of the observed catalog which shows relatively little evidence for spin precession [4].

## D. Correlations between event and population parameters

Beside re-investigating single events in light of the population, our double reweighting approach allows us to study the correlations between the parameters of single events (i.e. the $\theta$'s) and those of the population (i.e. the $\lambda$'s). This is the first time that such $\theta$–$\lambda$ correlations are studied in the published GW literature (although we note some unpublished results are present at Ref. [27]). Studying these correlations requires a suitable marginalization of the full hierarchical model as described in Sec. II F and cannot be done with the usual approach of, e.g., Refs. [3, 4].

Figure 5 shows the joint distribution of the primary mass $m_1$ of GW190521 (which is part of $\{\theta_i\}$) and the upper mass cutoff in the population model $m_{\text{max}}$ (which is instead part of $\lambda$). Among the many combinations between parameters and hyperparameters, this choice is particularly interesting because the value of the upper mass cutoff in the population fit is largely driven by the need to accommodate the largest mass in the catalog (although, the POWER LAW + PEAK population model we are using can accommodate a fraction of events above $m_{\text{max}}$, depending on the values of the "peak" parameters $\lambda_m$, $\mu_m$, and $\sigma_m$). Indeed, we find that the two quantities are moderately strongly correlated. This result can be read both ways: (i) if the mass of GW190521 turns out to be on the lower edge of its credible interval, one infers a lower mass cutoff in the entire population; (ii) if statistical and systematic errors in the population inference cause an overestimate of the mass cutoff (i.e. $m_{\text{max}}$ is smaller than what comes out of the inference), then the population-informed estimate of GW190521 points to lighter BHs.

This is just one of the possible analyses that can be performed with the method we presented. Going forward, we argue these correlations should be investigated in greater detail as they might shed further light on the GW data-analysis procedure and the related astrophysical interpretation of the sources.

## V. DISCUSSION

Inference on single events and inference on the population are part of a common Bayesian hierarchical model. These two aspects are, in practice, tackled individually for practical and computational constraints. In the LIGO context, this is made possible by the fact that sources are not overlapping, such that one can take a short time segment of data containing an event and analyze it independently on the rest of the data stream. This will not be possible with next-generation GW detectors. Source overlap will require the development of the so-called "global fit", tackling individual events, populations, and noise all at the same time (this is especially true for LISA, but 3rd generation ground-based detectors will also be affected).
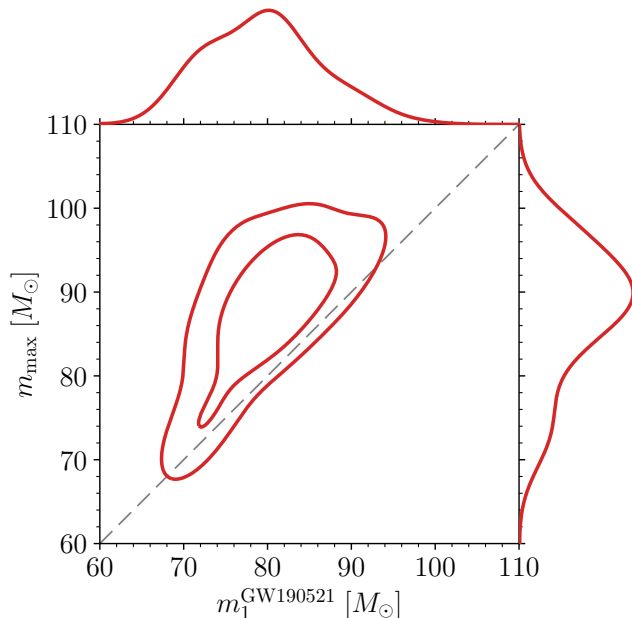
But even for current detector networks, developing a full

FIG. 5. Correlations between the primary mass of GW190521 (one of the event parameters $\theta$) and the maximum mass cutoff in the source distribution (one of the population parameters $\lambda$). Investigating $\theta$–$\lambda$ correlations requires considering the full hierarchical model (cf. Sec. II F) and cannot be tackled with the most common population approach which marginalizes over the event parameters (cf. Sec. II C and III B).

hierarchical model might uncover new features in the data. This paper tackles a particular aspect of this problem, namely the characterization of individual events in light of the (rest of the) detected population, an approach which is often referred to as using a "population-informed prior". The key result is presented in Eqs. (11) and (12).

We presented a thorough derivation of population-informed single-event statistics in the presence of selection effects, which we hope clarifies implicit assumptions and inconsistencies present in previous treatments. In particular, we highlighted the conceptual difference between the correct marginalization stemming from the full hierarchical model and the so-called empirical Bayes method, where the targeted event is effectively double counted.

Our formalism has several applications. Firstly, one can constrain the properties of GW events under the plausible assumption that they belong to a common population of sources. While we find that the conceptual correction pointed out here is subdominant, we stress it can be applied "for free", using data products that are routinely produced and made available. We thus argue that this correction should be applied to all GW analyses that make use of population-informed priors. Secondly, this approach further allows us to study, for the first time in the context of GW astronomy, the correlations between the parameters of individual events and those of the population. The line of investigation that is put forward in this paper has the potential to unveil new details on the

(astro)physics of GW sources from existing and future data.

Selection effects —analogous to the so-called Malmquist bias in observational astronomy— play an important role in GW population inference [5, 6]. For population-informed single-event analyses like those presented here, selection effects turn out to be a nuisance. In this context, one is interested solely in the *observed* population of sources, not the *observable* one. Our formalism contains selection effects because we wish to rely on current pipelines (e.g. Refs. [3, 4]) where $\lambda$ parameterizes the intrinsic population of BHs. This is also indicated in the PGM of Fig. 1(c), where selection effects are first introduced when inferring $\lambda$ from $\{\theta_{i \neq j}\}$ and then removed when reconstructing $\theta_j$ from $\lambda$. We stress that developing a simpler population analysis without selection effects —targeting the observed population and not the intrinsic one—, can also have applications in GW astronomy, an example of which is presented here. If the observed population has parameters $\tilde{\lambda}$, our equations can be immediately applied by substituting $\lambda \to \tilde{\lambda}$ and setting $\alpha(\lambda) = 1$.

We hope that the formalism presented in this paper can set the stage for deeper explorations of GW data exploiting the interplay between single events and populations.

## Appendix A: Rates analysis

In this appendix, we present the more general expressions one obtains without marginalizing over the expected number of events $N$.

The starting point is given by Eq. (4). Marginalizing over the individual-event parameters as in Sec. II C yields

$$P_{\mathrm{pop}}(\lambda|\{d_i\}) \propto \pi_{\mathrm{pop}}(\lambda)e^{-N\alpha(\lambda)}N^{N_{\mathrm{obs}}}$$
$$\times \prod_{i=1}^{N_{\mathrm{obs}}} \int \mathrm{d}\theta_i \, \mathcal{L}(d_i|\theta_i)p_{\mathrm{pop}}(\theta_i|\lambda). \quad \text{(A1)}$$

With a calculation analogous to that of Sec. II D, the population-informed single-event posterior can be obtained by instead marginalizing Eq. (4) over $\theta_{i \neq j}$ and $\lambda$. We obtain

$$P_{\mathrm{pop}}(\theta_j|\{d_i\}) \propto \mathcal{L}(d_j|\theta_j)\varpi(\theta_j|\{d_{i \neq j}\}, \quad \text{(A2)}$$

where

$$\varpi(\theta_j|\{d_{i\neq j}\}) = \int d\lambda\, N\, p_{\text{pop}}(\theta_j|\lambda)P_{\text{pop}}(\lambda|\{d_{i\neq j}\}) \quad \text{(A3)}$$

and

$$P_{\text{pop}}(\lambda|\{d_{i\neq j}\}) \propto \pi_{\text{pop}}(\lambda)e^{-N\alpha(\lambda)}N^{N_{\text{obs}}-1} \quad \text{(A4)}$$
$$\times \prod_{i\neq j}^{N_{\text{obs}}-1} \int d\theta_i\, \mathcal{L}(d_i|\theta_i)p_{\text{pop}}(\theta_i|\lambda)\,.$$

## Appendix B: Full hierarchical reweighting

Here we highlight a possible strategy to combine samples from single-event and population analysis to target the full hierarchical posterior of Eq. (2). We have attempted to implement this strategy and report here that this approach fails due to the limited numbers ($\sim 10^3$-$10^4$) of samples provided in public LIGO/Virgo data products.

Let us take one sample from each of the individual PE posterior chains (Eq. [16]) and a sample from the population inference posterior chain (Eq. [18]) and denote the result $X^m = (\{\theta_i\}^m, \lambda^m)$. This combined sample follows a distribution $Q(\{\theta_i\}, \lambda)$ that is the product of $N_{\text{obs}} + 1$ target distributions of the individual analyses, i.e.

$$Q(\{\theta_i\}, \lambda) \propto P_{\text{pop}}(\lambda|\{d_i\}) \prod_{i=1}^{N_{\text{obs}}} P_{\text{PE}}(\theta_i|d_i). \quad \text{(B1)}$$

Posterior samples from the full hierarchical model in Eq. (2) can be obtained by a reweighting of $X^m$;

$$(X^m, W^m) \sim P_{\text{pop}}(\{\theta_i\}, \lambda|\{d_i\}), \quad \text{for } m = 1, 2, \ldots, s, \quad \text{(B2)}$$

where $s = \min(\mathcal{S}, \{S_i\})$. The necessary weights are given by the ratio of the target distribution in Eq. (2) to the

$Q$-distribution evaluated at the $X^m$ samples:

$$W^m = \frac{P_{\text{pop}}(\{\theta_i\}^m, \lambda^m|\{d_i\})}{Q(\{\theta_i\}^m, \lambda^m)}, \quad \text{(B3)}$$
$$= \prod_{i=1}^{N_{\text{obs}}} \frac{p_{\text{pop}}(\theta_i^m|\lambda^m)}{\pi_{\text{PE}}(\theta_i^m)} \left[ \int d\theta_i\, \mathcal{L}(d_i|\theta_i)p_{\text{pop}}(\theta_i|\lambda^m) \right]^{-1}$$
$$\approx \prod_{i=1}^{N_{\text{obs}}} \frac{p_{\text{pop}}(\theta_i^m|\lambda^m)S_i}{\pi_{\text{PE}}(\theta_i^m)} \left[ \sum_{k=1}^{S_i} \frac{p_{\text{pop}}(\theta_i^k|\lambda^m)}{\pi_{\text{PE}}(\theta_i^k)} \right]^{-1}.$$

## Appendix C: Properties of the Hellinger distance

There are many ways to quantify the difference between probability distributions (see, for example, [28]). There is no particularly natural choice for our problem. Nevertheless, we find it convenient to quote a single number to quantify the *distance* between two distributions; we use the Hellinger distance [23] for this purpose.

The Hellinger distance between distributions $p(x)$ and $q(x)$ is defined as

$$d_{\text{H}}^2(p, q) = 1 - \int dx\, \sqrt{p(x)q(x)} \quad \text{(C1)}$$

and has a few convenient properties. First, it is symmetric: $d_{\text{H}}(p, q) = d_{\text{H}}(q, p)$. It also lies the range $0 \leq d_{\text{H}} \leq 1$ with the equalities occurring when $p$ and $q$ are identical or disjoint (mutually exclusive), respectively. To gain intuition for what a value of $d_{\text{H}}$ "means", one can compute the distance between offset Gaussians: if $p = \mathcal{N}(\mu, \Sigma)$ and $q = \mathcal{N}(\mu + c, \Sigma)$ then $d_{\text{H}}^2(p, q) = 1 - \exp(-c \cdot \Sigma^{-1} \cdot c/8)$ in any dimension.

Finally, we present a toy Bayesian calculation that is useful for gaining further intuition about $d_{\text{H}}$. Suppose we infer the value of parameters $x \in \mathbb{A}$. We have a prior $\pi(x)$, normalized such that $\int_{\mathbb{A}} dx\, \pi(x) = 1$. We make observations with a likelihood function $\mathcal{L}(\text{obs}|x) \propto \mathbb{1}_{\mathbb{B}}(x)$, where $\mathbb{1}$ is the indicator function and $\mathbb{B} \subset \mathbb{A}$. That is, our observations exclude with certainty some values of $x$ while allowing all others with equal likelihood. Using Bayes' theorem, the posterior distribution is $P(x|\text{obs}) = \mathbb{1}_{\mathbb{B}}(x)\pi(x)/F$, where the normalization $F = \int_{\mathbb{B}} dx\, \pi(x)$ is the fraction of the prior allowed by our observations. We find that the squared Hellinger distance between our prior and posterior is related to this fraction by $d_{\text{H}}^2(P, \pi) = 1 - \sqrt{F}$.

[1] B. P. Abbott *et al.* (LIGO and Virgo Collaboration), PRX **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[2] R. Abbott *et al.* (LIGO and Virgo Collaboration), PRX **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].

[3] B. P. Abbott *et al.* (LIGO and Virgo Collaboration), ApJ **882**, L24 (2019), arXiv:1811.12940 [astro-ph.HE].

[4] R. Abbott *et al.* (LIGO and Virgo Collaboration), ApJ **913**, L7 (2021), arXiv:2010.14533 [astro-ph.HE].

[5] E. Thrane and C. Talbot, PASA **36**, e010 (2019), arXiv:1809.02293 [astro-ph.IM].

[6] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, in *Handbook of Gravitational Wave Astronomy*, In press

(Springer, 2021) arXiv:2007.05579 [astro-ph.IM].

[7] D. K. Feuillet, J. Bovy, J. Holtzman, L. Girardi, N. MacDonald, S. R. Majewski, and D. L. Nidever, ApJ **817**, 40 (2016), arXiv:1511.04088 [astro-ph.GA].

[8] T. D. Morton and J. N. Winn, ApJ **796**, 47 (2014), arXiv:1408.6606 [astro-ph.EP].

[9] D. W. Hogg, A. D. Myers, and J. Bovy, ApJ **725**, 2166 (2010), arXiv:1008.4146 [astro-ph.SR].

[10] T. L. Campante, M. N. Lund, J. S. Kuszlewicz, *et al.*, ApJ **819**, 85 (2016), arXiv:1601.06052 [astro-ph.EP].

[11] A. J. Lyttle, G. R. Davies, T. Li, *et al.*, MNRAS 10.1093/mnras/stab1368 (2021), arXiv:2105.04482 [astro-ph.SR].

[12] G. Casella, The American Statistician **39**, 83 (1985).

[13] K. S. Thorne, in *Three Hundred Years of Gravitation* (Cambridge University Press, 1987) pp. 330–458.

[14] I. Mandel, W. M. Farr, and J. R. Gair, MNRAS **486**, 1086 (2019), arXiv:1809.02063 [physics.data-an].

[15] M. Fishbach, W. M. Farr, and D. E. Holz, ApJ **891**, L31 (2020), arXiv:1911.05882 [astro-ph.HE].

[16] C. Kimball, C. Talbot, C. P. L. Berry, M. Carney, M. Zevin, E. Thrane, and V. Kalogera, ApJ **900**, 177 (2020), arXiv:2005.00023 [astro-ph.HE].

[17] S. Miller, T. A. Callister, and W. M. Farr, ApJ **895**, 128 (2020), arXiv:2001.06051 [astro-ph.HE].

[18] S. Galaudage, C. Talbot, and E. Thrane, PRD **102**, 083026 (2020), arXiv:1912.09708 [astro-ph.HE].

[19] M. Fishbach and D. E. Holz, ApJ **904**, L26 (2020), arXiv:2009.05472 [astro-ph.HE].

[20] T. Callister, LIGO Document **T2100301** (2021), dcc.ligo.org/LIGO-T2100301.

[21] R. Essick and M. Fishbach, LIGO Document **T1900895** (2021), dcc.ligo.org/LIGO-T1900895.

[22] W. Farr, (2019), github.com/farr/Reweighting.

[23] E. Hellinger, Journal für die reine und angewandte Mathematik **1909**, 210 (1909).

[24] É. Racine, PRD **78**, 044021 (2008), arXiv:0803.1820 [gr-qc].

[25] P. Schmidt, F. Ohme, and M. Hannam, PRD **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].

[26] D. Gerosa, M. Mould, D. Gangardt, P. Schmidt, G. Pratten, and L. M. Thomas, PRD **103**, 064067 (2021), arXiv:2011.11948 [gr-qc].

[27] R. Abbott *et al.* (LIGO and Virgo Collaboration), LIGO Document **P2000434** (2021), dcc.ligo.org/LIGO-P2000434.

[28] G. E. Crooks, On Measures of Information and Entropy , v0.7 (2018), threeplusone.com.

[29] See Supplemental Material at for the complete set of "uninformative", "hierarchical", and "empirical" posterior distributions for each of the 44 GW events considered.