The Most Massive Binary Black Hole Detections and the Identification of Population Outliers

Maya Fishbach[1] , Will M. Farr[2,3] , and Daniel E. Holz[4]
[1] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA; mfishbach@uchicago.edu
[2] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA; will.farr@stonybrook.edu
[3] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800, USA
[4] Enrico Fermi Institute, Department of Physics, Department of Astronomy and Astrophysics, and Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA; holz@uchicago.edu

Abstract

Advanced LIGO and Virgo detected 10 binary black holes (BBHs) in their first two observing runs (O1 and O2). Analysis of these events found evidence for a dearth of BBHs with component masses greater than $\sim45\,M_\odot$, as would be expected from pair-instability supernovae. Meanwhile, a standalone analysis of the merger GW170729 found its primary mass to be $m_1 = 51.2^{+16.2}_{-11.0}\,M_\odot$, which appears to be in contradiction with the existence of a limit at $\sim45\,M_\odot$. In this work, we argue that the masses of individual events can only be evaluated with reference to the full population. When GW170729 is analyzed jointly with the remaining detections, its inferred primary mass tightens considerably, to $m_1 = 38.9^{+7.3}_{-4.5}\,M_\odot$. In the presence of noise, apparent outliers in the detected distribution are inevitable. We discuss methods of distinguishing between statistical fluctuations and true population outliers using posterior predictive tests. Applying these tests to O1 and O2, we find that the 10 detections are consistent with even the simplest power-law plus maximum-mass model considered by the LIGO–Virgo Collaboration, supporting the claim that GW170729 is not a population outlier. We also provide non-parametric constraints on the rate of high-mass mergers and conservatively bound the rate of mergers with $m_1 > 45\,M_\odot$ at $2.8^{+5.4}_{-2.0}\%$ of the total merger rate. After 100 detections like those of O1 and O2 from a population with a maximum primary mass of $45\,M_\odot$, it would be common for the most massive system to have an observed maximum-likelihood mass $m_1 \gtrsim 70\,M_\odot$.

Unified Astronomy Thesaurus concepts: Astrophysical black holes (98); Gravitational wave astronomy (675); Astronomy data analysis (1858); Hierarchical models (1925)

1. Introduction

A major goal of gravitational-wave (GW) astronomy is to learn about the formation and evolutionary mechanisms of binary black hole (BBH) mergers, such as those detected by Advanced LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015). There are many proposed formation channels for BBHs, including isolated evolution (Dominik et al. 2015; Belczynski et al. 2016a, 2016b; Eldridge & Stanway 2016; Woosley 2016; Stevenson et al. 2017b; Kruckow et al. 2018; Spera et al. 2019), dynamical formation (Hurley et al. 2016; Mapelli 2016; Rodriguez et al. 2016a, 2018; Askar et al. 2017; Chatterjee et al. 2017; Samsing 2018; Di Carlo et al. 2019; Zevin et al. 2019), and primordial origin (Bird et al. 2016; García-Bellido 2017), and if several formation channels are active at once, the population of merging BBHs may consist of distinct sub-populations. These sub-populations may differ in their shape of the mass distribution and spin distribution, as well as the merger rate (and its evolution with redshift).

Previous studies have explored methods of distinguishing between different formation channels using GW observations of BBHs, including fitting for the mixture fraction (or branching ratio) between various sub-populations (Stevenson et al. 2015, 2017a; Rodriguez et al. 2016b; Vitale et al. 2017b; Zevin et al. 2017; Bouffanais et al. 2019). One proposed sub-population includes second-generation mergers, which occur when at least one of the component BHs in a binary is itself the product of a previous merger. Second-generation BHs are expected to have a characteristic distribution of dimensionless spin magnitudes that peaks at $a \sim 0.7$ and a mass distribution

that extends into the "pair-instability" mass gap starting at $\sim40\text{–}70\,M_\odot$ (Fishbach et al. 2017; Gerosa & Berti 2017; Rodriguez et al. 2018). On the other hand, "first-generation" BHs that form directly from stellar collapse are expected to avoid the mass range between ~50 and $120\,M_\odot$. When a stellar core lies in this mass range, it is predicted to explode as a (pulsational) pair-instability supernova, and either leave behind no remnant, or lose mass sufficiently so that the BH remnant lies below the gap (Fowler & Hoyle 1964; Rakavy et al. 1967; Heger & Woosley 2002). The spins of BHs formed from stellar collapse are less certain (Miller & Miller 2015; Belczynski et al. 2017; Fuller & Ma 2019).

Second-generation mergers are possible (and generally expected to occur) exclusively in dense stellar environments such as globular clusters; therefore, the existence of this population is an important discriminator between dynamical and isolated formation. An additional proposed sub-population consists of gravitationally lensed GW signals, for which the lensing magnification causes a bias in the inferred luminosity distance and the unredshifted, source-frame masses if not properly accounted for (Cao et al. 2014). Therefore, gravitationally lensed events, even if they originate from the same formation channel as the unlensed events, would appear as a sub-population of erroneously high-mass, low-distance events (Dai et al. 2017; Broadhurst et al. 2018; Ng et al. 2018; Oguri 2018; Hannuksela et al. 2019).

The LIGO–Virgo Collaboration (LVC) detected 10 BBHs in its first two observing runs (Abbott et al. 2019a). With these 10 detections, Abbott et al. (2019b) fit simple parameterized

models to the mass, spin, and redshift distribution of the BBH population, assuming that all detections belong to the same population.[5] The assumption of a single population was justified by a leave-one-out analysis, which shows explicitly that excluding GW170729, the most "unusual" event (in terms of having the highest mass, spin, and distance), from the analysis does not significantly impact the inferred mass, spin, and redshift distributions at a level beyond the statistical uncertainties. Kimball et al. (2020) and Chatziioannou et al. (2019) also found that there is insufficient evidence to claim GW170729 as a population outlier by specifically comparing the hypothesis that it belongs to a population of second-generation, as opposed to first-generation, mergers based on the expected mass and spin distributions under the two scenarios.

In this paper we examine in more detail whether the assumed single-population mass distribution is a good fit to the data from the first 10 detections, with a focus on the inferred "maximum mass," or lower edge of the pair-instability mass gap. We argue that there are no convincing outliers among the O1 and O2 detections, implying that a single-component population model is sufficient to fit the data. As the number of detections increases, we expect standard statistical fluctuations to produce individual events with significant posterior support inside the so-called upper mass gap; GW170729 may be an example of such a fluctuation. Additionally, we forecast the masses of future detections based on the population of BBHs from O1 and O2, and explore the masses that would be required to identify a BBH detection as a true population outlier. Such an outlier may belong to an alternative population consisting of, e.g., second-generation mergers, or otherwise indicate that the assumed population model provides an insufficient description of the data. The methods discussed here can be extended to classify any outliers, including in the spin, mass-ratio, or redshift distribution.

The parametric models discussed above and used in Abbott et al. (2019b) are designed to incorporate a high-mass feature inspired by the predicted pair-instability mass-gap, whether it is a sharp cut-off to the power law (Models A and B of Abbott et al. 2019b) or a Gaussian component (Model C). To constrain the rate of high-mass BBH mergers in a model-agnostic way, we apply the non-parametric method of Mandel et al. (2017) to the 10 BBHs from O1 and O2. This method models the mass distribution as a binned histogram in the m_1–m_2 plane, with a smoothing prior on the bin heights. In contrast to the parametric models, the non-parametric fit a priori prefers smoothness over sharp features, providing a conservative upper limit on the rate of high-mass mergers.

The remainder of the paper is organized as follows. In Section 2, we demonstrate how a population analysis provides updated inference on the event-level parameters, yielding a tighter measurement on the masses of GW170729 in particular. In Section 3, we use posterior predictive distributions to evaluate the goodness-of-fit of a model to observations, focusing on high-mass outliers. In Section 4, we apply the non-parametric histogram model of Mandel et al. (2017) to the 10 LVC events and compare the inferred mass distribution to the parametric inference. We conclude in Section 5. Additional analysis details are provided in the Appendix.

---

[5] We note that in addition to the LVC-published detections of Abbott et al. (2019a), new BBH detections in the O1 and O2 data have been reported by Venumadhav et al. (2019a, 2019b) and Nitz et al. (2019).

2. Population Prior on Individual Events

A population analysis is concerned with fitting features that are common across the population members; in this case, BBH merger events. We assume that the event-level parameters $\{\theta_i\}$ (e.g., the component masses of the BBH event i) follow a probability distribution function $p_{\rm pop}(\theta_i|\Lambda)$, where Λ are the population-level hyperparameters (e.g., the power-law slope of the primary mass distribution). A hierarchical Bayesian analysis simultaneously fits for the event-level parameters θ_i and the population-level hyperparameters Λ from the data (Hogg et al. 2010; Mandel 2010; Mandel et al. 2011). In the presence of selection effects and measurement uncertainty, the joint posterior probability distribution of the event-level parameters $\{\theta_i\}$ and the population parameters Λ given the data $\{d_i\}$ from N independent events is given by (Loredo 2004; Mandel et al. 2019)

$$p(\{\theta_i\}, \Lambda|\{d_i\}) \propto \pi(\Lambda) \prod_{i=1}^{N} \frac{\mathcal{L}(d_i|\theta_i)p_{\rm pop}(\theta_i|\Lambda)}{\int P_{\rm det}(\theta)p_{\rm pop}(\theta|\Lambda)d\theta}, \quad (1)$$

where $\pi(\Lambda)$ is the prior on the population hyperparameters, $\mathcal{L}(d_i|\theta_i)$ is the event-level likelihood, and $P_{\rm det}(\theta)$ is the probability of detecting a piece of data from a merger with true parameters θ, as discussed in detail below.

In a population analysis, we usually marginalize Equation (1) over the event-level parameters θ_i to recover the posterior of the population parameters Λ. Meanwhile, when analyzing data from an individual event ("parameter estimation," or PE), the posterior on the event-level parameters,
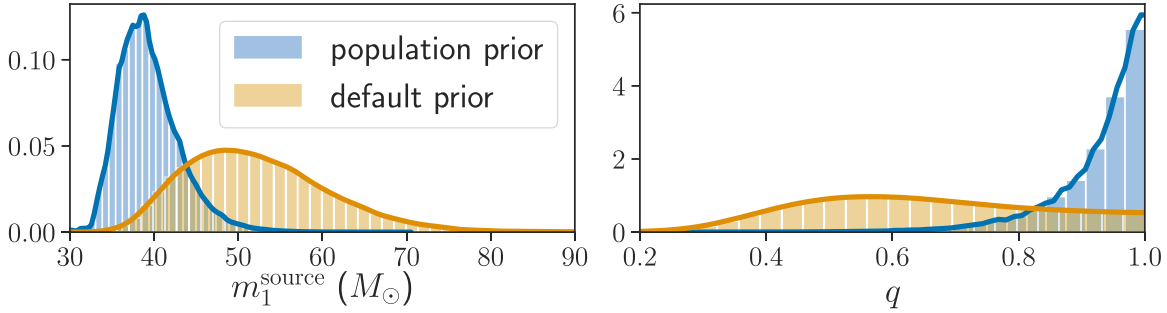
$$p(\theta_i|d_i) \propto \mathcal{L}(d_i|\theta_i)\pi(\theta_i), \quad (2)$$

is typically calculated using a default uninformative prior $\pi(\theta_i)$, rather than a population prior $p_{\rm pop}(\theta_i|\Lambda)$. The mass estimates in Abbott et al. (2019a), for instance, are obtained using priors that are flat in detector-frame component masses. Alternatively, one can marginalize Equation (1) over the population-level parameters Λ, and get a new posterior on the event-level parameters for each detection. Therefore, by calculating a joint posterior on the population-level and event-level parameters simultaneously, a hierarchical Bayesian analysis in effect replaces the default uninformative priors from PE with a population prior, yielding an informed posterior on the event-level parameters.

Since the first BBH detections, a variety of hierarchical analyses have been carried out to fit the BBH mass, spin, and redshift distributions to phenomenological parametric models (Abbott et al. 2016a; Fishbach & Holz 2017; Kovetz et al. 2017; Fishbach et al. 2018; Talbot & Thrane 2018; Roulet & Zaldarriaga 2019; Wysocki et al. 2019). In this work, we focus on the population analysis of Abbott et al. (2019b), specifically the mass distribution, which Abbott et al. fit according to three power-law-based models: Models A, B, and C. Model A has two free parameters: the power-law slope for the primary mass distribution and the maximum mass, and fixes the minimum mass to $5\,M_\odot$ and the conditional distribution for the secondary mass to be flat between the minimum mass and m_1. Model B is a generalization of Model A with two additional parameters: the minimum mass and the power-law slope of the secondary mass (or equivalently, mass ratio) distribution as conditioned on the primary mass. Model C generalizes Model B by tapering the low- and high-mass ends of the mass function (as opposed

GW170729

Figure 1. Primary mass (left panel) and mass ratio (right panel) of the BBH event GW170729 under the default (flat in detector-frame masses) prior (orange) vs. the prior implied by the Model B population analysis of Abbott et al. (2019b). The population analysis strongly constrains the maximum mass of the population to $\lesssim 50\,M_\odot$, and favors near-unity mass ratios, which implies that, if we believe the parametric model is a reasonable description of the population, the primary mass and mass ratio of GW170729 are relatively well-constrained. Alternatively, the analysis with the default prior, which ignores the rest of the detected population, ascribes a high probability for an outlier value to the mass.

to the sharp cutoffs of Models A and B) and allowing a high-mass Gaussian component on top of the power law in the primary mass. For each model, one can use the posterior samples for the population parameters Λ to recover the population-informed posteriors for individual event parameters θ_i according to Equation (1).[6]

In certain cases, the population analysis returns a noticeably different posterior for single-event parameters compared to the posterior returned from PE. For example, for the most massive event of O1 and O2, GW170729, the population analysis implies a much tighter prior on the masses compared to the uninformative priors of PE. The population-informed posteriors on m_1 and q for GW170729 under Model B are shown in Figure 1.[7] The Model B population analysis implies that the primary mass is $m_1 = 38.9^{+7.3}_{-4.5}\,M_\odot$ and the mass ratio is $q = 0.95^{+0.05}_{-0.18}$ compared to $m_1 = 51.2^{+16.2}_{-11.0}\,M_\odot$ and $q = 0.63^{+0.32}_{-0.26}$ under the default uninformative priors. Even under Model C, which has a high-mass Gaussian component instead of the sharp cutoff of Model B, we find that the population-informed primary mass of GW170729 is $m_1 = 37.9^{+7.5}_{-6.4}\,M_\odot$. Under either model, the population analysis implies that the primary mass of GW170729 is less than $50\,M_\odot$ at 99% credibility, in contrast to the results with uninformative priors.

3. Evaluating Tension between Model and Data

The above section assumes that the full set of observations can be adequately described by a given model, and it therefore makes sense to impose population priors and recalculate the posteriors on the parameters of each event. However, we often want to explicitly check whether the model fits the data sufficiently and whether the observations are consistent with one another under the model. In this section, we detail various methods of carrying out goodness-of-fit and outlier identification tests, and apply them to the BBH mass distribution fits from Abbott et al. (2019b).

3.1. Definitions and Assumptions

In a hierarchical population analysis, there are three levels at which we can perform a model goodness-of-fit/consistency test. The highest level consists of the population parameters Λ and their inferred values from the data. If fitting the population model separately on different subsets of events yields posteriors on the population parameters that are in significant tension with one another, one can conclude that the model does not fit all events as a single population. An example of this test was carried out for Model A of Abbott et al. (2019b) with a leave-one-out analysis, in which the population hyper-parameters were fit with and without GW170729. Comparing the posteriors on the hyper-parameters with and without GW170729 show that excluding GW170729 from the fit results in statistically consistent posteriors, leading to the conclusion that GW170729 is not a population outlier.

The second level of a hierarchical analysis consists of the (true) event-level parameters $\{\theta_i\}$. Following Abbott et al. (2019b), we define the posterior population distribution as a probability density function (pdf) on the true parameters θ of an (unobserved, and potentially unobservable) system that belongs to the population, given the data we have already observed:
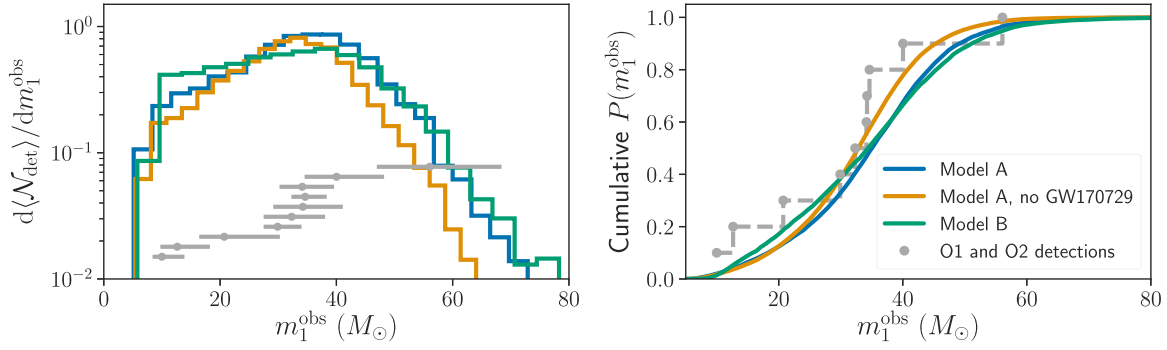
$$p(\theta|\{d_i\}) = \int p_{\rm pop}(\theta|\Lambda)p(\Lambda|\{d_i\})d\Lambda, \qquad (3)$$

where the posterior $p(\Lambda|\{d_i\})$ is given by marginalizing Equation (1) over the event-level parameters $\{\theta_i\}_{i=1}^N$ of the N detected events. The pdf on the true parameters θ of a *detection* must take into account selection effects by weighting each θ by the detectability of the data $\tilde{d}$ that it would produce in our detectors:

$$p(\theta, {\rm det}|\{d_i\}) = p(\theta|\{d_i\})P_{\rm det}(\theta)$$
$$= p(\theta|\{d_i\})\int P_{\rm det}(\tilde{d})p(\tilde{d}|\theta)d\tilde{d}. \qquad (4)$$

Note that the first term outside of the integral above is the posterior population distribution given by Equation (3). The term $p(\tilde{d}|\theta)$ takes into account the measurement uncertainty in going from the true parameters of the system θ to the data $\tilde{d}$, and the term $P_{\rm det}(\tilde{d})$ accounts for the fact that only some pieces of data are detectable. Throughout, we assume that the detectability of a piece of data, $P_{\rm det}(\tilde{d})$, is deterministic, meaning it is always 0 or 1,

---

[6] The posterior samples on the population hyper-parameters from Abbott et al. (2019b) are publicly available at https://dcc.ligo.org/LIGO-P1800324/public and the PE samples for individual events from Abbott et al. (2019a) are available at https://dcc.ligo.org/LIGO-P1800370/public.

[7] We use the parameter estimation results derived using the IMRPhenomPv2 (Khan et al. 2016) waveform approximant throughout. PE posteriors and population results with the SEOBNRv3 (Pan et al. 2014) approximant are available as well, and there are no significant differences between the two.

Figure 2. Left panel: posterior predictive distribution for the number of detections per *observed* primary mass bin during O1 and O2, based on the fits to Model A and Model B using all 10 BBHs (blue and green lines) and the nine BBHs excluding GW170729 (yellow). The gray error bars show the maximum-likelihood points and 90% credible intervals on the masses of each of the 10 BBHs; the vertical placement of these error bars is arbitrary. The observed mass is defined as the maximum likelihood estimate of m_1, and is predicted according to the synthetic detection and parameter estimation process described in the text. Right panel: cumulative posterior predictive distribution, or the probability that an observed mass is less than m_1^{obs}, inferred from the detections and the given population model, compared to the empirical distribution function from the 10 detections in gray (with the gray points denoting the maximum likelihood m_1 estimates). The agreement between the observations and each model can be quantified by the distances between the gray points and the colored curve of interest, as calculated in the text. The predictions of the population model match the observations fairly well, and GW170729 does not appear to be an outlier even when excluding it from the calculation of the posterior predictive distribution (yellow curve).

depending on whether the data pass a (known) detection threshold; this is discussed in more detail in Appendix A. The term $P_{det}(\theta)$ also appears in Equation (1). Sometimes $VT(\theta)$, the sensitive spacetime volume to a system with parameters θ, appears in place of $P_{det}(\theta)$, as these terms are proportional to each other assuming the merger rate is constant in comoving volume and source-frame time. Given a collection of events with data $\{d_i\}$ and a population model, we can calculate the above pdf for the true parameters of detected events (Equation (4)). Comparing the true parameters $\{\theta_i\}$ of the detected events (as inferred jointly with the population hyperparameters via Equation (1)) against the posterior predictive pdf $p(\theta, \text{det}|\{d_i\})$ provides another measurement of the consistency of the population model with the observations.

The third and final level of a hierarchical analysis consists of the data; this is the level on which we will evaluate the population fits for the remainder of this work. By folding in measurement uncertainty as well as the detection efficiency, we arrive at a probability distribution on the data $\tilde{d}$ from a future detection, rather than on the true parameters. We refer to this as the posterior predictive distribution:

$$p(\tilde{d}, \text{det}|\{d_i\}) = \int P_{det}(\tilde{d}) p(\tilde{d}|\theta) p(\theta|\{d_i\}) d\theta. \quad (5)$$

A detected piece of data d from a BBH merger is a timeseries consisting of the GW signal $h(t)$ which, assuming general relativity, is fully described by the source parameters θ and detector noise $n(t)$. Therefore, we shall identify the data d with the set of observed (maximum-likelihood) source parameters θ_{obs}, by which we mean the true source parameters offset by some measurement noise: $d = \{m_1^{obs}, m_2^{obs}, z^{obs}, \dots\} = \theta_{obs}$. Once an event is detected, the PE analysis returns a posterior pdf on the source parameters $p(\theta|\tilde{d})$ (Veitch et al. 2015). In referring to the observed parameters θ_{obs} in this work, we mean the maximum likelihood point-values as returned by PE.[8]

---

[8] The maximum-likelihood values do not correspond to the peak of the posteriors shown for event parameters in Abbott et al. (2019a) because a nonflat prior $\pi(m_1, m_2, z)$ is used. The default prior in the event-level analysis is proportional to the square of the luminosity distance, which places more weight at high z than a flat prior, so the maximum-likelihood source-frame mass values are larger than the maximum a posteriori values.

Details of how we simulate mock GW data sets with realistic measurement uncertainty in order to compute $P_{det}(\tilde{d})$ and $p(\tilde{d}|\theta)$ are found in Appendix A.

We stress that *noise fluctuations will generally cause us to estimate erroneously large source-frame masses.* A notable consequence of the detection threshold is that near-threshold sources are detectable only if a favorable noise fluctuation pulls the observed signal-to-noise ratio (S/N) above the detection threshold. Because the distance measurement correlates with the observed S/N, the systematic shift to larger S/N caused by the detection threshold implies that the observed redshift (as inferred from the distance) tends to be systematically shifted toward smaller values. This in turn implies that the observed source-frame masses, which are inferred by un-redshifting the detector-frame masses, are preferentially shifted toward larger values.

Although extreme noise fluctuations are intrinsically rare, they are statistically guaranteed to affect some fraction of detections. The larger the sample of detections, the larger the fluctuation it contains.

3.2. Application to the O1+O2 Population

3.2.1. Posterior Predictive Goodness-of-fit

In this subsection, we calculate the posterior predictive mass distribution for the BBH population from O1 and O2 and explore whether the population models provide an adequate fit to the 10 BBH observations. We focus on their primary masses, and more specifically, the primary mass of GW170729.

The posterior predictive distribution (Equation (5)) for m_1^{obs}, given the events from O1 and O2 and the assumed mass Models A and B from Abbott et al. (2019b), is shown in Figure 2. We focus on Models A and B, the simple power-law models, because they have fewer free parameters than Model C, and we wish to check whether these few parameters sufficiently fit the data. We show the posterior predictive distributions as inferred from all 10 BBHs, as well as under Model A excluding GW170729. The left panel in Figure 2 shows the expected number of detections per m_1^{obs} bin during O1 and O2 according to the model; this is based on the inferred merger rate together with the shape of the mass distribution.

We note that although Model A (B) predicts a sharp cutoff at $m_{\max} = 41.6^{+9.6}_{-4.3} \, M_\odot$ ($m_{\max} = 40.8^{+11.8}_{-4.4} \, M_\odot$), and all three models considered in Abbott et al. (2019b) predict that 99% of BBHs have $m_1 < 45 \, M_\odot$, Model A (B) predicts that 18% (20%) of detected systems have an *observed* primary mass $m_1^{\mathrm{obs}} > 45 \, M_\odot$. This is because out of the underlying population of BBHs, more massive systems are more likely to be detected, and out of those that are detected, statistical fluctuations can push the observed (maximum likelihood) m_1^{obs} to values that are oftentimes significantly larger than the true m_1. Recall that among detected sources, statistical fluctuations are more likely to push the observed source-frame masses to larger values than smaller values, because sources near threshold are only detectable if a fluctuation increases their observed S/N, leading to a smaller inferred redshift and a larger inferred source-frame mass. The full likelihood distribution should still have nonzero support at the true value, but for very large statistical fluctuations, the support at the true value may be very small and difficult to resolve.

From Figure 2, one can visually compare the number and observed masses of the O1 and O2 detections (gray errorbars) to the model predictions; this serves as a posterior predictive check that the model fits the data sensibly. The gray points at the center of the errorbars denote the maximum likelihood m_1^{obs} for each event; these points are the values that, according to the model, should be representative draws from the posterior predictive distributions. (The errorbars denote the 90% symmetric credible intervals from the full m_1 posteriors and are shown only for reference.)

The right panel of Figure 2 shows the cumulative, normalized versions of the posterior predictive distributions in the right panel (the colored curves) compared to the empirical distribution function (edf; the gray points). The edf is a cumulative histogram of the maximum likelihood m_1, $\hat{m}_{1,i}^{\mathrm{obs}}$, for each event i, defined as:

$$\hat{F}_n(m_1^{\mathrm{obs}}) = \frac{1}{n} \sum_i^n I[\hat{m}_{1,i}^{\mathrm{obs}} < m_1^{\mathrm{obs}}], \qquad (6)$$

where n is the number of events and I is the indicator function which evaluates to 1 if its argument is True and zero otherwise.

As seen in the right panel of Figure 2, the edf appears to follow the posterior predictive cumulative probability distributions (cdfs) for Models A and B, with perhaps slightly more low-mass and high-mass detections than predicted under the simple power-law models. The relative lack of intermediate-mass detections (and comparable abundance of detections at the low- and high-mass ends of the $m_{\min}$–$m_{\max}$ range) can also be seen in the fit to Model C (the tapered power law with a high-mass Gaussian) of Abbott et al. (2019b). Under Model C, the only model that allows for deviations from a pure power law in m_1, Abbott et al. (2019b) find that the data mildly prefer a merger rate that decreases at intermediate m_1 and then rises again at $m_1 \sim 30 \, M_\odot$ (see the top panel of their Figure 2). Abbott et al. (2019b) conclude that this preference for a power-law deviation is not statistically significant because all three models A, B, and C predict distributions of m_1 that overlap within the 90% statistical uncertainties. Abbott et al. (2019b) also report Bayes factors between all three models and find that a deviation from a power law is preferred by a factor of $e^{1.92} \approx 7$ although, as usual, the values of Bayes factors are sensitive to the priors chosen for the model hyper-parameters. Our posterior predictive checks for the pure power-law Models

A and B do not rely on an explicit comparison to an alternative model, but our conclusions agree with Abbott et al. (2019b) that the observed m_1 distribution is consistent with these power-law models well within the 90% level.

More quantitatively, the distance between the edf and the cdfs predicted from the models as fit to the data is a measure of "goodness-of-fit," or how well each model explains the data. This is the basis of the well-known Kolmogorov–Smirnov (KS) (Kolmogorov 1933; Smirnov 1948) and Anderson–Darling (AD) statistics (Anderson & Darling 1952).[9] Restricting ourselves only to the m_1 distribution, we compute the KS and AD statistics as follows. We first construct a null distribution by repeatedly drawing sets of 10 observations from the posterior predictive distribution of interest and calculating the KS and AD statistic for each set. Comparing the KS and AD statistics for our set of 10 observations against the statistics of the null distribution, we find that the KS statistic between the edf and the Model B posterior predictive cdf lies at 84% of the null distribution, while the AD statistic lies at 79% of its null distribution. For Model A, the KS (AD) statistic is at 87% (82%) of its null distribution. We also compute these statistics between the edf for all 10 events and the Model A posterior predictive cdf as inferred without GW170729. This does not affect the statistics significantly, with the KS statistic corresponding to 69% of the null distribution and the AD statistic at 71% in this case, which indicates that the observed primary mass of GW170729 is consistent with the m_1 distribution as inferred without it.[10] *In summary, Models A and B are all found to be consistent with the observed population.*
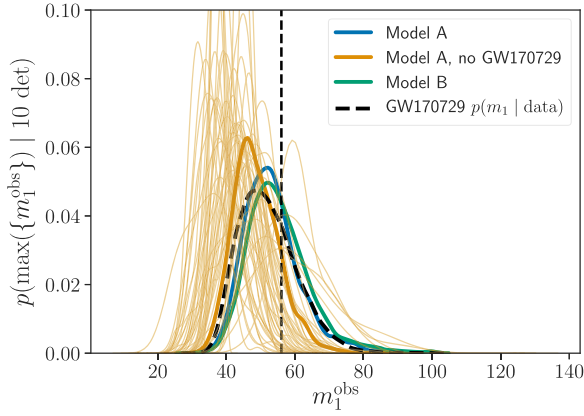
Note that the edf is not uniquely defined in two or more dimensions, and Figure 2 shows only how the model fits the observed m_1 distribution, ignoring all other parameters of BBH systems including the mass ratio, spin, and redshift. However, the edf-based test described above can be extended (non-uniquely) to an arbitrary number of dimensions as long as the algorithm for ordering the observations is fixed in advanced. The null distribution can still be calculated according to this fixed ordering algorithm and, if desired, a *p*-value can be calculated to evaluate goodness-of-fit.

3.2.2. Quantifying Outliers

We now turn to the more specific issue of quantifying whether a particular observation is an outlier with respect to other events that make up the population. For example, in order to investigate the degree of tension between a particularly massive observation, such as GW170729, and the maximum mass inferred from the population analysis, it is useful to consider the posterior predictive distribution of the *maximum* observed primary mass out of N detections, $\max(\{m_{1,\mathrm{obs}}^j\}_{j=1}^N)$, based on the data from the remaining

---

[9] The KS statistic is the maximum distance between the cdf of the model and the edf derived from the data (or two edfs for two different data sets) while the AD statistic is a weighted average of the distance between each point in the sample and the cdf. The KS statistic is mainly sensitive to differences between the centers of the two distributions while the AD statistic is more sensitive to differences in the tails of the distribution (Stephens 1974).

[10] Counterintuitively, the shift toward smaller KS and AD statistics seems to indicate that the m_1 posterior predictive distribution as inferred from only nine events actually provides a slightly better fit to the observed primary masses of all 10 events; however, the shift in these statistics is not significant, and illustrates that this is a simplified one-dimensional test that is only meant to check consistency between data and model and not to properly fit a model.

Figure 3. Posterior predictive distribution of the maximum *observed* mass out of 10 detections as inferred from the detections and the population model of interest (bold, solid colored curves). The observed mass refers to the maximum likelihood m_1 value of a detected event as predicted according to a synthetic detection and parameter estimation (PE) process (see the text). The thin orange curves show mock PE posteriors for 50 random events drawn from the bold orange curve, representing 50 examples of posteriors for the most massive m_1 that we expect to detect based on the fit to Model A from the nine detections excluding GW170729. For comparison, the posterior for the primary mass of GW170729 is shown (dashed black curve) with the maximum-likelihood value (vertical dashed line). Visually, the GW170729 m_1 posterior appears consistent with the thin orange curves. Quantitatively, comparing its maximum-likelihood value to the bold orange curve shows that the primary mass of GW170729 is consistent with the population as inferred from the other nine events at the 86% level.
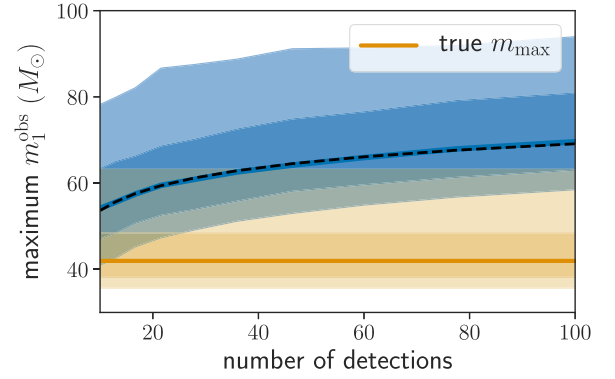
$N-1$ detections, $\{d_i\}_{i=1}^{N-1}$:

$$p(\max(\{m_{1,\text{obs}}^j\}_{j=1}^N)|\ \{d_i\}_{i=1}^{N-1}) = \frac{d}{dm}$$
$$\times \left(\left[\int_0^m p(m_{1,\text{obs}}|\{d_i\}_{i=1}^{N-1}) dm_{1,\text{obs}} \right]^{N-1} \right) \Bigg|_{m=\max(\{m_{1,\text{obs}}^j\}_{j=1}^N)},$$
$$(7)$$

where the pdf in the integral is simply the posterior predictive distribution. Equation (7) follows from the fact that given N random draws X_i from a fixed pdf $p(x)$ with corresponding cdf $P(X_i < x)$, the maximum draw Y will follow a cdf:

$$P(Y < x) = P(X_i < x)_{\forall i \in [1,N]} = P(X_i < x)^N. \qquad (8)$$

The posterior predictive distributions for the maximum observed mass out of 10 detections is shown in Figure 3. The maximum-likelihood primary mass m_1^{obs} for GW170729 (vertical dashed line) is consistent at the 86% level with the posterior predictive distribution on the maximum of 10 primary mass observations, as inferred from Model A and the remaining nine observations (orange solid curve). The light orange curves show mock posteriors on m_1 for 50 events drawn from the solid orange curve, and we can see that the m_1 posterior for GW170729 (dashed black curve) looks like a typical observation. When using GW170729 itself in calculating this posterior predictive distribution, its primary mass is consistent with being the maximum of 10 observations at 69% under Model A (blue curve) and 59% under Model B (green curve). Based on this analysis, we conclude that there is no evidence for tension between the primary mass of GW170729 and the remaining nine observations under Models A and B, in agreement with the leave-one-out analysis of Abbott et al. (2019b).
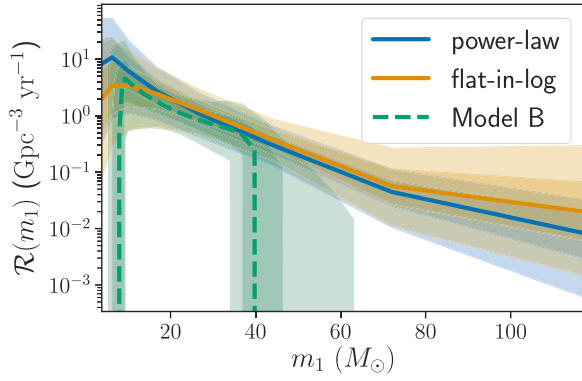

Figure 4. Maximum mass we expect to observe as a function of number of detections (blue), as inferred from the Model B fit to the 10 O1/O2 detections (which predicts the true $m_{\max} = 40.8^{+11.8}_{-4.4}\ M_\odot$, shown in orange). The observed mass m_1^{obs} represents the maximum likelihood value that would be inferred for the detected system, taking into account measurement uncertainty and selection effects. The solid line denotes the median and dark and light bands denote the 68% and 95% credible intervals of the posterior predictive distribution. As the number of detections increases, the largest noise fluctuation in the sample will become more extreme. Furthermore, because of the S/N threshold, these noise fluctuations statistically lead to larger inferred masses. The blue curves are well fit by $\max m_1^{\text{obs}} \simeq \max m_1 + \sigma\sqrt{2\log(fN)}$ with $\sigma \sim 17\ M_\odot$ (see the text); this fit is shown by the black dashed line.

As the sample of BBH detections grows, we expect to see more extreme statistical fluctuations, so that we observe primary masses m_1^{obs} that are much higher than the true maximum BH mass. In Figure 4, we show, as a function of the number of detections, the most massive m_1^{obs} that we expect in the sample, based on the fit to Model B with the 10 O1/O2 detections. We find that our predictions for the maximum observed mass after N detections matches expectations from analytic Gaussian scaling. Assuming Gaussian observational uncertainties of width σ, with N detections, some fraction fN of systems will have masses within 1σ of the true maximum mass, where f depends on the shape of the mass distribution convolved with selection effects. The largest observed mass will following a scaling $\max m_1^{\text{obs}} \sim \max m_1 + \sigma\sqrt{2\log(fN)}$; this scaling is observed empirically in Figure 4. In other words, the most extreme noise fluctuation in a sample of size N is typically observed at $\sqrt{2\log N}$ standard deviations above its true value. It would be common that by the time we have 100 detections, even if they are all described by Model B with a sharp cutoff at $m_{\max} = 40.8^{+11.8}_{-4.4}\ M_\odot$, we will observe a system with an apparent mass of $m_1^{\text{obs}} \sim 70\ M_\odot$.

4. Non-parametric Constraints on the Rate of Mass-gap Mergers

To explore the rate of high-mass mergers in a less parametric way, we follow the binned-histogram analyses of Foreman-Mackey et al. (2014) and Mandel et al. (2017). We model the rate of BBH mergers on the two-dimensional m_1–m_2 plane, $\mathcal{R}(m_1, m_2) = \frac{dN}{dm_1 dm_2 dV_c dt}$ as piecewise constant in 9×9 logarithmically spaced mass bins between 3 and 150 $M_\odot$. The height of each mass bin $\mathcal{R}_{ij}$ represents the merger rate in that bin. We take the prior on the bin heights to be a squared-exponential Gaussian process (GP) where the relative means μ_{ij} of the bin heights follow a fixed shape $p(m_{1,i}, m_{2,j})$ and the length scales in $\log(m_1)$ and $\log(m_2)$ are fit from the data (see Appendix B for more details). We consider two different shapes for the mean merger rate per bin in the GP prior: a

Figure 5. Non-parametric constraints on the merger rate distribution for m_1 from the binned histogram model under the two shape priors (flat-in-log prior in blue and power-law prior in orange). The solid lines denote the median rate and the dark and light shaded bands denote 68% and 95% symmetric credible intervals. In green we show the merger rate as a function of m_1 for the parametric mass distribution Model B; this model includes a maximum mass cutoff as one of the parameters. In the low-mass region $m_1 \lesssim 45\,M_\odot$, the nonparametric model under both priors agrees with the parametric model. Because of the lack of detections at high masses, the parametric Model B infers a tight constraint on $m_{\max}$ and the merger rate falls to zero, while the nonparametric model attempts to extrapolate smoothly to high masses under strong influence of the prior.

"power-law" shape prior and "flat-in-log" shape prior. These are motivated by the two fixed-parameter models that the LVC used to calculate BBH merger rates in O1 and O2 (Abbott et al. 2016a, 2019a). In the power-law shape prior, we have

$$p(m_1, m_2) \propto \frac{m_1^{-2.35}}{m_1 - M_{\min}}, \tag{9}$$

where $m_{\min} = 3\,M_\odot$. The flat-in-log shape prior is simply

$$p(m_1, m_2) \propto \frac{1}{m_1 m_2}. \tag{10}$$

Figure 5 shows the merger rate $\mathcal{R}(m_1)$ as inferred under the two different priors, where

$$\mathcal{R}(m_1) = \int_3^{m_1} \mathcal{R}(m_1, m_2)\,dm_2. \tag{11}$$

For comparison we also show the results from the parameterized Model B fit. We note that in the range $m_1 \lesssim 45\,M_\odot$, the inferred merger rates $\mathcal{R}(m_1)$ agree between all three models: the two shape priors in the nonparametric model as well as the parametric power-law model. (The difference for $m_1 < 5\,M_\odot$ is due to the prior on $m_{\min} > 5\,M_\odot$ for Model B, while the lowest-mass bin in the nonparametric model starts at $3\,M_\odot$.) Beyond $\sim 45\,M_\odot$, the merger rate inferred under Model B drops sharply due to the $m_{\max}$ parameter, whose inferred value closely follows the mass of the most massive observed system (Fishbach & Holz 2017). The binned model, on the other, does not have a $m_{\max}$ parameter that lets the rate fall to zero, and instead has a prior that strongly favors smooth variations of the merger rate from mass bin to mass bin. In the mass bins with $m_1 \gtrsim 45\,M_\odot$, where there are no detections, the posterior on the merger rate smoothly transitions to following the prior on the bin heights. This is visible in Figure 5 as the merger rates $\mathcal{R}(m_1)$ inferred under the two different priors (blue and orange bands) start to diverge from one another at $m_1 \gtrsim 45\,M_\odot$, and the uncertainty for each one grows as well.

This is a consequence of the GP smoothing prior. After enough detections, if the absence of high-mass events continues, the likelihood will overcome the smoothing prior and the posterior will reveal a sharp drop-off in the merger rate in the binned analysis, independently of the prior on the bin heights. This was demonstrated by Mandel et al. (2017) with simulated data in the context of the putative low-mass gap between the binary neutron star and BBH population. With only 10 detections, the binned model provides a conservative upper limit on the rate of mergers with $45\,M_\odot < m_1 < 150\,M_\odot$ under the prior that the merger rate should not vary sharply between neighboring mass bins.

Under the flat-in-log prior, we infer the merger rate in the mass range $45\,M_\odot < m_1 < 150\,M_\odot$ to be $3.02^{+12.97}_{-2.28}$ Gpc^{-3} yr^{-1} (90% equal-tailed credible interval), or $1.79^{+2.30}_{-1.23}$ Gpc^{-3} yr^{-1} under the power-law prior. We infer the total merger rate over the $3\,M_\odot < m_1 < 150\,M_\odot$ range to be $42.50^{+68.12}_{-24.89}$ Gpc^{-3} yr^{-1} with the flat-in-log shape prior or $65.58^{+102.34}_{-41.82}$ Gpc^{-3} yr^{-1} with the power-law shape prior, implying that the rate of mergers with $m_1 > 45\,M_\odot$ makes up $7.6\%^{+23.8}_{-6.0}$ (flat-in-log prior) or $2.8\%^{+5.4}_{-2.0}$ (power-law prior) of the total merger rate. This is to be compared to the parametric models of Abbott et al. (2019b), which all predict that less than 1% of mergers have $m_1 > 45\,M_\odot$. Unlike the parametric models with a maximum mass parameter, the binned-histogram model does not allow the high-mass merger rate to drop all the way to zero. The advantage of the nonparametric constraints is that, if there were a secondary population of BBH mergers that does not respect the maximum mass feature of the parameterized mass models (consisting, for example, of second-generation mergers that occupy the pair-instability mass gap), we still expect their merger rate to respect these nonparametric limits.

5. Conclusion

Focusing on the BBH mass distribution as inferred by Abbott et al. (2019b), we have explored how individual events fit into a population analysis, especially in the presence of measurement uncertainty and selection effects. We have presented simple posterior predictive/goodness-of-fit checks to show consistency between the O1/O2 events and the power-law mass distribution models of Abbott et al. (2019b). In particular, GW170729, the most massive event of O1/O2, is not an outlier with respect to even the simplest power-law model with a sharp high-mass cutoff. When folding in the full information about the population, the primary mass of GW170729 is inferred to be $m_1 = 38.9^{+7.3}_{-4.5}\,M_\odot$ as compared to the inferred value under the uninformative PE priors, $m_1 = 51.2^{+16.2}_{-11.0}\,M_\odot$. We argue that, as the number of detections increases, the sample will contain statistical fluctuations even more extreme than the one that likely gave rise to the observed mass of GW170729. These statistical fluctuations will lead us to preferentially overestimate the masses of individual events, unless we infer the individual events' masses jointly with the full population. In doing so, it will also be necessary to verify that the chosen population model fits the data sufficiently well. Although the simple parameterized models provide adequate fits to the BBH detections so far, we have presented nonparametric fits to the mass distribution based on the GP-regularized binned-histogram model of Mandel et al. (2017). Under this model, we place conservative upper limits on the

rate of mergers with $45\,M_\odot < m_1 < 150\,M_\odot$, and find that these high-mass mergers make up at most $7.6\%^{+23.8}_{-6.0}$ of the total merger rate in the range $3\,M_\odot < m_1 < 150\,M_\odot$ for a flat-in-log shape prior on the mass distribution, or $2.8\%^{+5.4}_{-2.0}$ of the total merger rate for a power-law shape prior.

Appendix A
Mock Observations

This section explains in greater detail how we calculate selection effects and simulate measurement uncertainty for mock observations. For a BBH with true parameters θ, we follow the simple prescription of Fishbach et al. (2018) and Farr et al. (2019) to assign realistic measurement uncertainty and compute θ_{obs}. Given the BBH's source-frame masses, spins, and redshift together with a power spectral density (PSD) describing the noise of the detector, we can calculate the optimal S/N ρ_{opt} of the source, which is the S/N that it would have if it were optimally oriented face-on and directly overhead of the detector (Chen et al. 2017). We assume a fixed cosmology described by the best-fit Planck 2015 parameters (Planck Collaboration et al. 2016) to interchange between redshift and luminosity distance (Astropy Collaboration et al. 2013). We fix the PSD to the "Advanced LIGO Early High Sensitivity" noise curve from Abbott et al. (2018), which is representative for O1 and O2. We also fix BBH spins to zero in this calculation, since they have a negligible effect on the S/N calculation for population studies (Abbott et al. 2019b). An isotropic distribution of sky positions and inclinations relative to a detector yields a distribution of true S/Ns ρ in that detector; this distribution can be summarized by the angular projection term $0 \leqslant \Theta = \frac{\rho}{\rho_{opt}} \leqslant 1$. The angular term Θ has a known distribution (Finn & Chernoff 1993). Therefore, for a BBH with intrinsic parameters $\{m_1, m_2, z\}$, we assign a single extrinsic parameter Θ drawn from this distribution. These four parameters together correspond to a true S/N ρ for the source.

Given $\{m_1, m_2, z, \Theta, \rho\}$ for a source, we assign measurement uncertainty as follows. We expect that in stationary, Gaussian noise, the matched-filter S/N ρ has unit variance (Allen et al. 2012). We first draw an observed S/N ρ_{obs} from

a normal distribution centered at the true S/N ρ:

$$\rho_{obs} = \rho + N(0,1) \qquad (12)$$

where $N(\mu, \sigma)$ is a random number drawn from a normal distribution with mean μ and standard deviation σ. We assume that sources are only detected if they pass an S/N threshold (in a single detector) $\rho_{obs} > 8$; this would be identical to the semi-analytic selection effects calculation of Abbott et al. (2016a, 2016b, 2019b) were it not for the inclusion of the noise term $N(0,1)$. To best approximate the mass measurement, we work with the detector-frame (redshifted) chirp mass

$$\mathcal{M}_z = (1+z)\frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \qquad (13)$$

and symmetric mass ratio

$$\eta = \frac{m_1 m_2}{(m_1 + m_2)^2}. \qquad (14)$$

The detector-frame chirp mass drives the leading-order GW frequency evolution during the inspiral and is thus the best-measured mass parameter for stellar-mass compact binary sources. We assume that the uncertainties $\{\sigma_{\mathcal{M}}, \sigma_\eta, \sigma_\Theta\}$ on the measured parameters scale inversely with ρ_{obs}, so that

$$\log \mathcal{M}_z^{obs} = \log[\mathcal{M}_z + N(0, \sigma_{\mathcal{M}}/\rho_{obs})], \qquad (15)$$

$$\eta_{obs} = N^{T[0,0.25]}(\eta, \sigma_\eta/\rho_{obs}), \qquad (16)$$

$$\Theta_{obs} = N^{T[0,1]}(\Theta, \sigma_\Theta/\rho_{obs}). \qquad (17)$$

In the above expressions, $N^{T[a,b]}(\mu, \sigma)$ denotes a random number drawn from the truncated normal distribution. From $\mathcal{M}_z^{obs}$ and η_{obs}, we calculate the detector-frame component masses $m_{1,z}^{obs}$ and $m_{2,z}^{obs}$:

$$m_{1/2,z}^{obs} = \frac{M \pm \sqrt{M^2 - 4\eta M^2}}{2} \qquad (18)$$

for $M = \mathcal{M}_z^{obs}/\eta_{obs}^{3/5}$. The observed redshift is inferred directly from the remaining parameters, via the observed luminosity distance d_{obs}:

$$d_{obs} = \frac{d_{fid}\,\rho_{opt}((1+z)m_1^{obs}, (1+z)m_2^{obs}, d_{fid})\Theta_{obs}}{\rho_{obs}}, \quad (19)$$

where d_{fid} is an arbitrary fiducial luminosity distance (for given detector-frame masses, the S/N of a source scales inversely with its luminosity distance). The observed redshift z_{obs} is derived from d_{obs} by the cosmological redshift–luminosity distance relation and, once this is known, we infer the source-frame masses:

$$m_{1/2}^{obs} = \frac{m_{1/2,z}^{obs}}{1 + z_{obs}}. \qquad (20)$$

The observed values m_1^{obs}, m_2^{obs}, z_{obs} denote the maximum likelihood values of the parameters as extracted from the GW signal. To simulate full posterior distributions on these parameters, we use Equations (12)–(19) as the likelihood for the observed parameters given true values ρ, $\mathcal{M}_z$, η, and Θ. We take samples on ρ, $\mathcal{M}_z$, η, and Θ from this likelihood and convert the samples to the space $m_{1,z}$, $m_{2,z}$, d_L, Θ, on which we

wish to set a prior:

$$\pi_{\mathrm{PE}}(m_{1,z}, m_{2,z}, d_L, \Theta) \propto p(\Theta) d_L^2. \qquad (21)$$

This matches the default PE prior used by LIGO/Virgo in the O1 and O2 event analysis (Abbott et al. 2019a), where $p(\Theta)$ is the true distribution from which the Θ values are drawn, representing an isotropic distribution on the sky and source inclination. The change of variables from $\rho, \mathcal{M}_z, \eta, \Theta$ to $m_{1,z}, m_{2,z}, d_L, \Theta$ means we also have to divide out by the induced prior, given by the Jacobian

$$\left| \frac{d\rho}{dd_L} \frac{d(\mathcal{M}_z, \eta)}{d(m_{1,z}, m_{2,z})} \right| = \frac{\Theta \rho_{\mathrm{opt}}(m_{1,z}, m_{2,z}, d_{\mathrm{fid}}) d_{\mathrm{fid}}}{d_L^2}$$
$$\times \frac{(m_{1,z} - m_{2,z}) \eta(m_{1,z}, m_{2,z})^{3/5}}{(m_{1,z} + m_{2,z})^2}. \qquad (22)$$

Once we reweight the $m_{1,z}, m_{2,z}, d_L, \Theta$ samples by Equation (21) divided by Equation (22), we can get posterior samples for the source-frame parameters by converting to m_1, m_2, z, Θ space.

We tune the σ parameters above to match the measurement uncertainties on masses and redshifts found by Vitale et al. (2017a) when simulating full PE on injected signals. Vitale et al. (2017a) found that for BBHs detected by Advanced LIGO/Virgo at design sensitivity, the relative uncertainty (at the 90% credible interval) on the detector-frame masses is typically $\sim$40% and the relative uncertainty on redshift is typically $\sim$50%. However, for the majority of O1 and O2, only the two (co-aligned) LIGO detectors were operational, implying a reduced ability for the network to constrain the polarization of a source and break the distance–inclination degeneracy, and worsened redshift constraints. We find that for the O1 and O2 events, a more typical relative redshift uncertainty is 70% (for a 90% credible interval relative to the median value). We find that choosing $\sigma_{\mathcal{M}} = 0.08\rho_{\mathrm{thresh}}$, $\sigma_\eta = 0.022\rho_{\mathrm{thresh}}$ and $\sigma_\Theta = 0.21\rho_{\mathrm{thresh}}$ yields measurement uncertainties that match the widths of the O1 and O2 credible intervals and the expectations from Vitale et al. (2017a). We use $\rho_{\mathrm{thresh}} = 8$ throughout, as discussed above. The measurement uncertainty on Θ controls the measurement uncertainty on z according to Equation (19). For simulating events for a three-detector network at design sensitivity, we would use $\sigma_\Theta = 0.15$ to reflect the improved distance constraints of a three-detector network with relative uncertainties of 50% rather than 70%. We note that the measurement uncertainty on the source-frame masses is a combination of the detector-frame mass uncertainty and the absolute redshift uncertainty, which is largest for sources at high redshift. Therefore, our predictions for distributions of m_1^{obs} based on distributions of m_1 are sensitive to assumptions regarding the underlying redshift distribution and the network sensitivity, which together determine the detected redshifts of the sources. The uncertainty in these assumptions is subdominant to uncertainties in the population model.

Appendix B
Binned Histogram Likelihood

In this appendix, we provide additional analysis details regarding the binned-histogram fit to the mass distribution (Section 4). The total posterior for the rate $\mathcal{R}_{ij}$ in each mass bin,

the parameters Γ governing the Gaussian process prior on the bin heights, and the true masses of the detected events are given by the inhomogeneous Poisson process likelihood:

$$p(\mathcal{R}_{ij}, \Gamma, \{m_1, m_2\} | d)$$
$$\propto \left[\prod_{k=1}^{N_{\mathrm{events}}} p(d^{(k)} | m_1^{(k)}, m_2^{(k)}) \mathcal{R}(m_1^{(i)}, m_2^{(i)}) \right]$$
$$\times \exp \left[-\sum_{ij} \lambda_{ij} \right] p(\mathcal{R}_{ij} | \Gamma) p(\Gamma), \qquad (23)$$

where $p(d^{(k)} | m_1^{(k)}, m_2^{(k)})$ denotes the likelihood of the data for event k given component masses, marginalized over all other parameters of the system, $p(\mathcal{R}_{ij} | \Gamma)$ represents the GP prior on the bin heights and λ_{ij} denotes the expected number of detections in bin ij, folding in selection effects.

The GP prior on the (log) of the bin heights takes the form

$$p(\log \mathcal{R}_{ij} | \Gamma) = N(\log \mathcal{R}_{ij} | \log(\mu + \mu_{ij}), \Sigma) \qquad (24)$$

where $N(x | \hat{\mu}, \hat{\Sigma})$ denotes the multivariate normal probability distribution function on x with mean $\hat{\mu}$ and covariance $\hat{\Sigma}$. Meanwhile, μ is an overall scale factor and $\mu_{ij} = \log p(\log m_{1,i}, \log m_{2,j})$ is fixed to either the power-law or flat-in-log shape discussed in Section 4. Note that the point $(\log m_{1,i}, \log m_{2,j})$ denotes the center of the ijth bin. Finally, Σ is a covariance matrix. We use a squared exponential kernel for the covariance matrix, and parameterize Σ as

$$\sum_{ijkl} = (1 - f)\sigma^2 \exp \left[-\frac{1}{2\tau^2} \Delta^2 \log m_{ijkl} \right] + f\sigma^2 \delta_{ijkl}. \qquad (25)$$

For numerical stability, the covariance Σ includes some fraction $f \ll 1$ that is white and uncorrelated; the precise value of f does not affect the results and is fixed to $f = 0.01$ throughout the analysis. The square of the Euclidean distance between the centers of the bins $(\log m_{1,i}, \log m_{2,j})$ and $(\log m_{1,k}, \log m_{2,l})$ is denoted $\Delta^2 \log m_{ijkl}$. At zero separation, the variance is σ^2. The correlation length scale is set by the parameter τ. Lastly, δ_{ijkl} is the Kronecker delta function.

In summary, the GP parameter set Γ consists of the free parameters μ, σ, and τ, and the fixed parameters μ_{ij} and f. For the prior on these hyper-parameters (written $p(\Gamma)$ in Equation (23)), we take a broad Gaussian prior for μ with mean 0 and standard deviation 10 and a half-Gaussian prior on $\sigma > 0$ with mean 0 and standard deviation 1.

We take a Gaussian prior for $\log \tau$. The mean and standard deviation of the Gaussian are chosen to place the width of the smallest mass bin 2σ below the mean and the width of the mass range considered (3 $M_\odot$–150 $M_\odot$) 2σ above the mean. Thus the prior constrains the correlation length of the rate in log-mass to be typically longer than one bin, but shorter than the entire mass range. The correlation length τ is poorly constrained by the data with only 10 detections, and the recovered posterior is very similar to the prior. We stress that, in the limit of a large number of detections, the likelihood will dominate the GP prior and the posterior on the bin heights will become independent of these prior choices.

To evaluate the posterior of Equation (23) we also need the expected number of detections λ_{ij} in bin ij. To do this we evaluate the sensitive spacetime volume in the ijth bin, $\langle VT \rangle_{ij}$,

so that

$$\lambda_{ij} = \mathcal{R}_{ij} \langle VT \rangle_{ij}. \qquad (26)$$

The sensitive volume was introduced in Section 3, and is calculated from the detection probability P_{det} by (Abbott et al. 2019b)

$$VT(m_1, m_2) = T_{\text{obs}} \int P_{\text{det}}(m_1, m_2, z) \frac{1}{1+z} \frac{dV_c}{dz} dz, \qquad (27)$$

where T_{obs} is the observing time. This assumes that the merger rate is constant (non-evolving with redshift) in comoving volume and source-frame time, and that the detector sensitivity is constant in time. The average sensitive volume in bin ij is

$$\langle VT \rangle_{ij} = \int_{m_{i-1/2}}^{m_{i+1/2}} \int_{m_{j-1/2}}^{m_{j+1/2}} VT(m_1, m_2) \frac{dm_1}{m_1} \frac{dm_2}{m_2}, \qquad (28)$$

where the integral boundaries are the edges of the bin.

Consistent with the other analyses in this work, we calculate $P_{\text{det}}(m_1, m_2, z)$ under the assumption of a single-detector S/N threshold. Abbott et al. (2019b) found this approximation to $VT(m_1, m_2)$ to overestimate the sensitive volume by a factor of ~1.6 relative to the volume calculated by injections into the detection pipeline. This factor of ~1.6 between the semi-analytic sensitive volume and the injection-estimated volume is fairly constant across the mass space. We therefore divide our value of $\langle VT \rangle_{ij}$ by 1.6 to match the results of injections. We further assume that the total $\langle VT \rangle$ has a (Gaussian) 1σ error of 18% (arising from an amplitude calibration error of $\sim6\%$) and marginalize over this uncertainty, split evenly between the mass bins (Abbott et al. 2016a).

ORCID iDs

Maya Fishbach https://orcid.org/0000-0002-1980-5293
Will M. Farr https://orcid.org/0000-0003-1540-8562
Daniel E. Holz https://orcid.org/0000-0002-0175-5064

References

Aasi, J., Abadie, J., Abbott, B. P., et al. 2015, CQGra, 32, 074001
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, PhRvX, 6, 041015
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, ApJL, 833, L1
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, LRR, 21, 3
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019a, PhRvX, 9, 031040
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, ApJL, 882, L24
Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, CQGra, 32, 024001
Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, PhRvD, 85, 122006
Anderson, T. W., & Darling, D. A. 1952, Ann. Math. Stat., 23, 193
Askar, A., Szkudlarek, M., Gondek-Rosińska, D., Giersz, M., & Bulik, T. 2017, MNRAS, 464, L36
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Belczynski, K., Holz, D. E., Bulik, T., & O'Shaughnessy, R. 2016a, Natur, 534, 512
Belczynski, K., Klencki, J., Fields, C. E., et al. 2017, arXiv:1706.07053
Belczynski, K., Repetto, S., Holz, D. E., et al. 2016b, ApJ, 819, 108
Bird, S., Cholis, I., Muñoz, J. B., et al. 2016, PhRvL, 116, 201301
Bouffanais, Y., Mapelli, M., Gerosa, D., et al. 2019, ApJ, 886, 25
Broadhurst, T., Diego, J. M., & Smoot, G., III 2018, arXiv:1802.05273
Cao, Z., Li, L.-F., & Wang, Y. 2014, PhRvD, 90, 062003
Chatterjee, S., Rodriguez, C. L., Kalogera, V., & Rasio, F. A. 2017, ApJL, 836, L26
Chatziioannou, K., Cotesta, R., Ghonge, S., et al. 2019, PhRvD, 100, 104015
Chen, H.-Y., Holz, D. E., Miller, J., et al. 2017, arXiv:1709.08079
Dai, L., Venumadhav, T., & Sigurdson, K. 2017, PhRvD, 95, 044011
Di Carlo, U. N., Giacobbo, N., Mapelli, M., et al. 2019, MNRAS, 487, 2947
Dominik, M., Berti, E., O'Shaughnessy, R., et al. 2015, ApJ, 806, 263
Eldridge, J. J., & Stanway, E. R. 2016, MNRAS, 462, 3302
Farr, W. M., Fishbach, M., Ye, J., & Holz, D. E. 2019, ApJL, 883, L42
Finn, L. S., & Chernoff, D. F. 1993, PhRvD, 47, 2198
Fishbach, M., & Holz, D. E. 2017, ApJL, 851, L25
Fishbach, M., Holz, D. E., & Farr, B. 2017, ApJL, 840, L24
Fishbach, M., Holz, D. E., & Farr, W. M. 2018, ApJL, 863, L41
Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64
Fowler, W. A., & Hoyle, F. 1964, ApJS, 9, 201
Fuller, J., & Ma, L. 2019, ApJL, 881, L1
García-Bellido, J. 2017, JPhCS, 840, 012032
Gerosa, D., & Berti, E. 2017, PhRvD, 95, 124046
Hannuksela, O. A., Haris, K., Ng, K. K. Y., et al. 2019, ApJL, 874, L2
Heger, A., & Woosley, S. E. 2002, ApJ, 567, 532
Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166
Hurley, J. R., Sippel, A. C., Tout, C. A., & Aarseth, S. J. 2016, PASA, 33, e036
Khan, S., Husa, S., Hannam, M., et al. 2016, PhRvD, 93, 044007
Kimball, C., Berry, C. P. L., & Kalogera, V. 2020, RNAAS, 4, 2
Kolmogorov, A. 1933, G. Ist. Ital. Attuari, 4, 83
Kovetz, E. D., Cholis, I., Breysse, P. C., & Kamionkowski, M. 2017, PhRvD, 95, 103010
Kruckow, M. U., Tauris, T. M., Langer, N., Kramer, M., & Izzard, R. G. 2018, MNRAS, 481, 1908
Loredo, T. J. 2004, in AIP Conf. Ser. 735, BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 195
Mandel, I. 2010, PhRvD, 81, 084029
Mandel, I., Farr, W. M., Colonna, A., et al. 2017, MNRAS, 465, 3254
Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086
Mandel, K. S., Narayan, G., & Kirshner, R. P. 2011, ApJ, 731, 120
Mapelli, M. 2016, MNRAS, 459, 3432
Miller, M. C., & Miller, J. M. 2015, PhR, 548, 1
Ng, K. K. Y., Wong, K. W. K., Broadhurst, T., & Li, T. G. F. 2018, PhRvD, 97, 023012
Nitz, A. H., Dent, T., Davies, G. S., et al. 2019, arXiv:1910.05331
Oguri, M. 2018, MNRAS, 480, 3842
Pan, Y., Buonanno, A., Taracchini, A., et al. 2014, PhRvD, 89, 084006
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Rakavy, G., Shaviv, G., & Zinamon, Z. 1967, ApJ, 150, 131
Rodriguez, C. L., Amaro-Seoane, P., Chatterjee, S., & Rasio, F. A. 2018, PhRvL, 120, 151101
Rodriguez, C. L., Haster, C.-J., Chatterjee, S., Kalogera, V., & Rasio, F. A. 2016a, ApJL, 824, L8
Rodriguez, C. L., Zevin, M., Pankow, C., Kalogera, V., & Rasio, F. A. 2016b, ApJL, 832, L2
Roulet, J., & Zaldarriaga, M. 2019, MNRAS, 484, 4216
Samsing, J. 2018, PhRvD, 97, 103014
Smirnov, N. 1948, Ann. Math. Stat., 19, 279
Spera, M., Mapelli, M., Giacobbo, N., et al. 2019, MNRAS, 485, 889
Stephens, M. A. 1974, J. Am. Stat. Assoc., 69, 730
Stevenson, S., Berry, C. P. L., & Mandel, I. 2017a, MNRAS, 471, 2801
Stevenson, S., Ohme, F., & Fairhurst, S. 2015, ApJ, 810, 58
Stevenson, S., Vigna-Gómez, A., Mandel, I., et al. 2017b, NatCo, 8, 14906
Talbot, C., & Thrane, E. 2018, ApJ, 856, 173
Veitch, J., Raymond, V., Farr, B., et al. 2015, PhRvD, 91, 042003
Venumadhav, T., Zackay, B., Roulet, J., Dai, L., & Zaldarriaga, M. 2019a, PhRvD, 100, 023011
Venumadhav, T., Zackay, B., Roulet, J., Dai, L., & Zaldarriaga, M. 2019b, arXiv:1904.07214
Vitale, S., Lynch, R., Raymond, V., et al. 2017a, PhRvD, 95, 064053
Vitale, S., Lynch, R., Sturani, R., & Graff, P. 2017b, CQGra, 34, 03LT01
Woosley, S. E. 2016, ApJL, 824, L10
Wysocki, D., Lange, J., & O'Shaughnessy, R. 2019, PhRvD, 100, 043012
Zevin, M., Pankow, C., Rodriguez, C. L., et al. 2017, ApJ, 846, 82
Zevin, M., Samsing, J., Rodriguez, C., Haster, C.-J., & Ramirez-Ruiz, E. 2019, ApJ, 871, 91