Draft version November 11, 2022
Typeset using LaTeX **twocolumn** style in AASTeX631

Modeling the Mass Distribution of Binary Black Hole Mergers with GWTC-3

Jingyi Zhang,[1] Supervised by: Alan J. Weinstein,[2] and Jacob Golomb[2]

[1] *Smith College*
[2] *LIGO Laboratory, California Institute of Technology*

ABSTRACT

With 70 binary black hole merger (BBHs) events detected by the LIGO -Virgo Collaboration, it is possible to infer the overall character of the black hole population in the universe. Specifically, the mass distribution of BBHs provides us with valuable information on stellar evolution and binary formation channel. We here aim to test the current estimation of BBHs population mass distribution based on the third Gravitational-wave Transient Catalog (GWTC-3). The project involves: (1) Identifying outliers and examining the impact of non-conventional events with leave-one-out tests on population parameter estimation; (2) explore possible ways to adjust the model to better characterize astrophysical phenomena;(3) examine agreement between the observed data and the fitted mode by conducting goodness-of-fit tests. We did not found any statistically significant outlier event, though some events can help direct us with future observations. Our goodness-of-fit test shows agreement between current observations and fitted model.

1. INTRODUCTION: BLACK HOLE POPULATION IN THE UNIVERSE

Based on the third Gravitational-wave Transient Catalog (GWTC-3), The LIGO Scientific Collaboration et al. (2021) looked into the mass and spin distribution, overall merger rate, as well as the cosmological evolution of the merger rate of binary black holes. These information provides valuable information on high mass star formation and history of stellar evolution, as well as how compact objects fit in the evolution of the universe.

The BBHs population has been fitted with models and described with the associated population hyper-parameters. For example, Figure 1 shows the estimated primary mass distribution of BBHs using the most popular *POWER LAW + PEAK* model (PP) model described in Section 2.3.1.

Aiming to evaluate the population inference's performance, compare different black hole population models, discuss the effects of outliers, and improve the general modeling process; we want to conduct goodness-of-fit tests on each step of the hyper-parameter estimation. In addition to supplement the GWTC-3 estimations, the test can also help to better understand the expected population from the upcoming forth observation run (O4.) For this project, I focus on the mass distribution of BBHs.

This report is organized as follows. In Section II we introduce the parameters from the observations and how these parameters inform us about the BBHs events and population. We also introduce the models that we used to describe the black hole population and how are they characterized. In Section III, IV, and V we present three aspects of testing and current fitted model and possible ways of improvements. In each of the sections, we first introduce the methods, and then present the result. In Section III we discuss possible outliers in GWTC-3 and their impact on the inferred population; in Section VI we discuss the redshift evolution of BBHs population and how it may be included in the model fitting process;

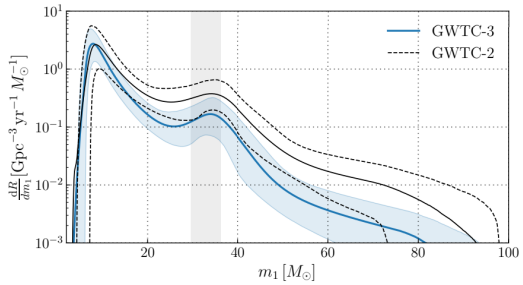

Figure 1. Probability density function of the primary mass distribution for the fiducial Power-Peak (PP) model. (For more on the model, please see Talbot & Thrane (2018)) The solid lines show the posterior population distribution (PPD) and the region shows the 90 percent credible interval of GWTC-3 (blue) and GWTC-2 (black), respectively. As shown in the figure, GWTC-3 suggests that the primary mass has more prominent peaks than we previously observed.(The LIGO Scientific Collaboration et al. 2021)

in Section V we present a Posterior Predictive Checks between the observed population and the fitted model predicted result. Finally in Section VI we conclude the results of this project.

2. BACKGROUND: STRAIN, EVENTS, AND POPULATION

The raw data LIGO obtained are strains that record the change of length of the interferometer arms. Merger events are identified by match filtering the strain data with a template bank of BBH waveforms. Once a possible BBHs event is found, its intrinsic parameters such and mass and spin and extrinsic parameters such as redshift and inclination of can be estimated from the strain data with Bayesian analysis; the process is called (single-event) parameter estimation (PE). There are many models that may describe the entire BBHs population, and these models are defined with hyper-parameters (power law index, height of Gaussian peak, etc.) By analyzing many BBHs events with Bayesian analysis, hyper-parameters can be estimated. The two-step approach, also known as the hierarchical Bayesian inference (Talbot et al. 2019) is the key of studying the black hole population.

2.1. *Single event parameters*

15 parameters describes an individual BBH events: 8 intrinsic parameters that characterized the spin and mass of the two black holes, 7 extrinsic parameters that describe the binary's position and orbit as seen by us. Packages such as BILBY (Ashton et al. 2019) and LALINFERENCE (Veitch et al. 2015) use Bayesian inference to estimate the parameters from the strain data. (Figure 2.) The result is given in the form of posterior samples. Each sample contain a possible combination of the parameters that may create the observed signal. The distribution of parameters in the posterior sample represent the probability density of each parameters, including how they may be correlated, as shown in Figure 3.

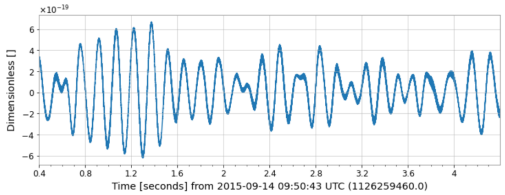

Figure 2. The strain data of BBH event GW150914 at the LIGO Hanford observatory. The strain contains 4 seconds of data around the event. (Abbott et al. 2021a)

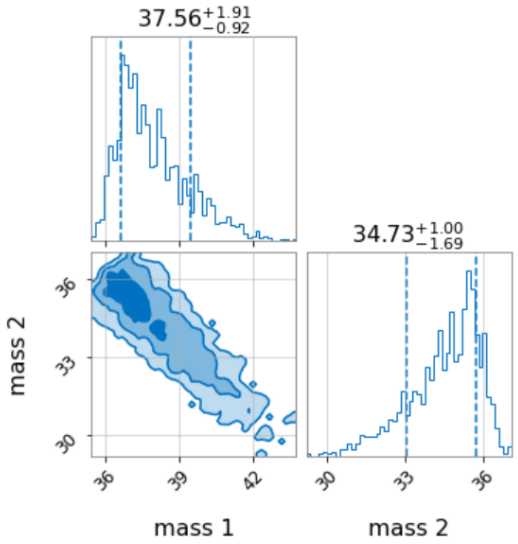

Figure 3. The posterior distribution of primary mass (m1), and secondary mass (m2) of GW150914 and their correlation. In this corner plot, all other parameters are marginalized. The dotted line marked the 90 percent credibility interval——90 percent of the posterior sample is in this mass range.

One thing to note is that the estimated parameters are in the detector frame, the properties we see on earth. However, many BBHs events happened far away enough that we must take cosmology into consideration. We can directly estimate the luminosity distance from the strain data. With the assumption of current date cosmological parameters from Planck 2015, we can calculate the redshift of the event. (Hogg 1999) The redshift can then be used to calculate the source frame parameters of the events.

$$M_{detector} = (1+z)M_{source} \tag{1}$$

The source frame parameter is independent of the observer and reflect the true astrophysical information. For studying the black hole population, we always use the source frame parameters.

2.2. *Population Hyper-Parameters*

With the single events parameters in hand, we would be able to describe the entire black hole population. We characterize the population with "population hyper-parameters," such as the index of the power law α of the *POWER LAW + PEAK* model, the lower mass cutoff m_{min}, or the position of the peak μ_m.We will discuss the specific parameters for each model and their physical motivation in Section 2.3.1.

In single event parameter estimation, we take in observed strain data as the likelihood to obtain the posterior distribution of black hole parameters. (Note that in this case, as we don't know much about the event,

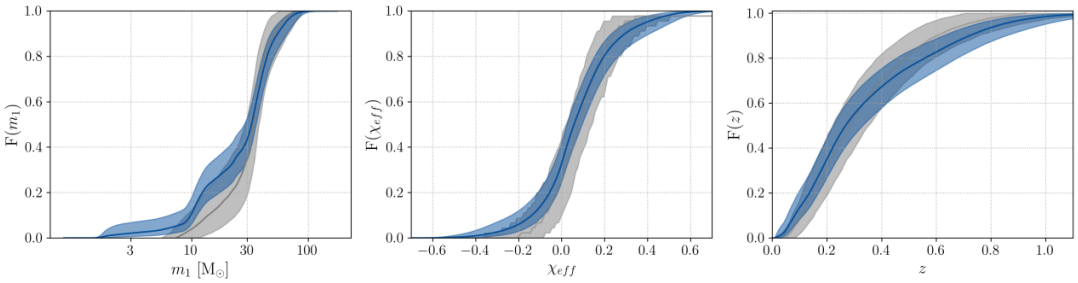

Figure 4. The empirical cumulative density function of primary mass, effective inspiral spin, and redshift. The blue region represent the distribution from GWTC-3, while the grey represent the distribution from GWTC-2. The solid lines show the median while the shaded area are 90 percent credible interval. (The LIGO Scientific Collaboration et al. 2021) For this project, we hope to conduct similar comparison between the observed GWTC-3 population and the theoretical modeled population.

we use uninformative prior in the Bayesian inference.) In population hyper-parameter estimation, the "data" we used as the likelihood is the single event parameters. Through similar Bayesian approach, we may obtain the posterior distribution of hyper-parameters. The process is called hierarchical Bayesian inference as it consist two "layers" of parameter estimation. (Talbot et al. 2019)

With the hyper-parameters defined by each population model, we can construct the likelihood for the Bayesian inference. For each BBH, we can calculate the probability of getting such events under a defined population model (the model is characterized with hyper-parameters). Given that all events are independent, the probability of getting all the events—-the likelihood–is the product of all individual probability.

$$\mathcal{L}(d|\Lambda) = \prod_i^N P(d_i|\Lambda) \tag{2}$$

Where Λ is the hyperparameters and θ_i is the parameter of the i^{th} event.

However, we do not have the exact value of the parameter describing each single events. We can only infer their probability distribution from the strain data.

$$\mathcal{L}(d|\Lambda) = \prod_i^N \int d\theta_i \mathcal{L}(d_i|\theta_i) p(\theta_i|\Lambda) \tag{3}$$

Here the $\mathcal{L}(d_i|\theta_i)$ term is the probability of getting the strain data d_i for the i^{th} event given a set of parameters θ_i, also referred to as the likelihood. We integrate over all θ_i that we can get from the strain data to get the total probability, hence the $\int d\theta_i \mathcal{L}(d_i|\theta_i)$ term. In reality, we do not have a smooth probability distribution, but rather discrete samples that represent the events. The relative number of samples per differential bin represent the probability (similar to how histogram is used to describe probability density.) Instead of integration, we take the sum of the probability of each sample to construct our hyper-parameter likelihood; this is a monte carlo approximation.

$$\mathcal{L}(d|\Lambda) = \prod_i^N \sum_{\theta_i \sim \mathcal{L}(d_i|\theta_i)} p(\theta_i|\Lambda) \tag{4}$$

$$\mathcal{L}(d|\Lambda) \propto \frac{\prod_i^N \sum_{\theta_i \sim \mathcal{L}(d_i|\theta_i)} p(\theta_i|\Lambda)}{p_{\rm det}(\Lambda)^N} \tag{5}$$

Finally, like all other astronomical observation, gravitational detection is subjected to selection effect: more massive black holes that create stronger signals are easier to detect. As we are interested in the astrophysical properties of black hole population, we correct this bias by including a "detection probability" term which is a function of the hyper-parameters.

$$\mathcal{L}(d|\Lambda) \propto \frac{\prod_i^N \int d\theta_i \mathcal{L}(d_i|\theta_i) p(\theta_i|\Lambda)}{p_{\rm det}(\Lambda)^N} \tag{6}$$

$$p_{\rm det}(\Lambda) = \int p(\theta|\Lambda) p_{\rm det}(\theta) d\theta \tag{7}$$

The "detection probability" term include the probability of detecting each parameter and the probability of detection such parameter under certain model. It can be thought of as the relative fraction of the astrophysical sources in the population which are detectable by our detectors.

With the likelihood defined, we pass it to python package GWPOPULATION do the Bayesian inference Talbot et al. (2019).

2.3. *Black hole population model*

2.3.1. *Mass Models*

The *POWER LAW + PEAK* model (PP model) is currently the most commonly used one that describe the mass distribution of black holes in the universe. It

is initially motivated by the stellar initial mass function (IMF) and pulsational pair-instability supernova (PPSN), which motivates the power law and the Gaussian peak in the model, respectively.

Stellar IMF suggested that the mass of the stars in the universe follows a power law, where more massive stars are rarer. As the mass of a stellar black hole is directly related to the mass of its progenitor star, it is reasonable for us to anticipate the mass distribution of stellar black holes also follows a similar power law.

PPSN occurs to massive stars (roughly between $100M_\odot$ and $150M_\odot$) at the end of their life. Stars of this size are large enough to create some electron-positron pairs but not massive enough that the electron-positron pairs would cause them to undergo complete pair-instability supernovae. These stars are very likely to lose significant amount of mass (until they reach around $100M_\odot$) until they can undergo "normal" core-collapse supernova. (Woosley 2017) This create a peak at the higher end of the black hole mass distribution which is expected to be followed by a cutoff. The peak is modeled by an Gaussian.

Though currently the *POWER LAW + PEAK* model seems to best match the observation, it does not seems to capture the full picture. In particular, the fitted result from both GWTC-2 and GWTC-3 show the peak to be centered around $35M_\odot$, not at the highest end of the mass distribution. There's currently no commonly agreed explanation on the issue. Some have suggested that this might be caused by hierarchical mergers: the black holes in the BBHs events might be a merged result from a previous merger event. This process may allow the mass distribution to have additional peaks and higher maximum mass.

Other parametric models characterize different aspects of the black hole mass population, including the different possible shape at the lower and higher mass cutoff, multiple peaks that could be caused by hierarchical mergers, broken power law that adjust for the different IMF, etc. (Figure) One of the goal of this project is to test whether which model have better agreement with the observations and how they might be improved.

2.3.2. *Spin Models*

The spin of black hole is harder to measure than the mass, and therefore the current fitted models have large uncertainties. Though we do not test the spin models in this project for their huge uncertainties, it is still helpful to introduce the model as all event parameters are estimated at the same time, and the spin of the black hole is often correlated with other properties.

The current most commonly used black hole spin model is motivated by two formation channel of BBHs. For common evolution BBHs, it is expected that the spin of individual black hole is in the same direction of the binary's angular momentum. This is represented as a normal distribution centered at $\theta = 0$. (The tilt angle θ is the angle between the direction of spin of individual black hole and the direction of angular momentum of the system.) For black holes that formed separately, there should be no preferred orientation, and thus a flat distribution of θ.

3. POTENTIAL OUTLIERS IN GWTC-3

In previous studies, two BBHs events (GW190814 and GW190917) are identified as outliers as they have very low secondary mass. The rest of the population can not predict the existence of these two black holes, as shown in Fig 5. The probability with and without these two events have almost no overlaps, and it is very unlikely that there is a value that can predict both scenario. We wonder if any other events have similar constraining power that with or without them in the population inference, the hyper-parameters estimated would not agree.

No new statistically significant outliers were found. However, several events have strong constraining power on certain hyper-parameters and may be potential source of interests. Future observations may help clarify whether other distinct sub-populations exist.

3.1. *Methods: Leave-one-out Tests*

To identify outliers, we conducted leave-one-out test for all 69 events (excluding the previously identified outliers.) First, all 69 events are used in the population inference. The result is used as a default comparison. We then exclude one event for population inference and generate a new set of population result. This "leave-one-out" trial is done for all 69 events. By comparing the leave-one-out results with the original result, we may see the impact of individual events when characterizing the whole population. Note, we used PP mass model and *Default* spin model for both the default and leave-one-out tests.

Two sample Kolmogorov–Smirnov tests (KS tests) were used to compare the probability distribution of hyper-parameters and quantify the differences. The two sample KS test measure the difference between two population's cumulative distribution function; higher KS test statistic implies that the two sample are more distinct. For all hyper-parameters the test runs, the events yields the highest statistic value in KS tests are picked out for closer analysis.

3.2. *Results: Potential Outliers*

3.2.1. *GW190924: Low Mass*

The secondary black hole of GW190924 has a mass of roughly 5 $M_\odot$, the lowest among all 69 events. Other lowest masses are around 7 $M_\odot$ (including uncertainties.) The posterior probability distribution of the leave-one out test in Figure 6 also shows this potential gap in the minimum mass. The KS test statistic value on minimum mass is over 0.9, much higher than the second highest statistic value of just over 0.1. It is possible that this black hole belongs to a separate population, or be more aligned with the two previously identified outliers. However, though the distributions of the minimum mass are quite different, they still have significant overlap and may agree with each other. Thus, we cannot conclude GW190924 s an outlier.

GW190924 also has the highest KS test statistic value on λ, a hyper-parameter in the PP model that characterize the normalized ratio between the black hole population falls into the Gaussian peak and the power law. Without GW190924, the inferred λ centered at around 0.07; with it, the he inferred λ centered at around 0.03. Excluding GW190924, less back holes are following the power law and more black holes are in the Gaussian peak. (Figure 7) This is highly correlated with the minimum mass: As the Gaussian peak is centered around 35 $M_\odot$, the lowest mass population is dominated by the power law. The power law also predicts low mass black holes are much more common than the high mass ones. Together, small decrease on the minimum mass would significantly impact λ as the it allows large portion of the events to be presented at the power law dominated, lowest mass range. Similar to minimum mass, we cannot conclude GW190924 to be an outlier based on λ as the two result overlap significantly.

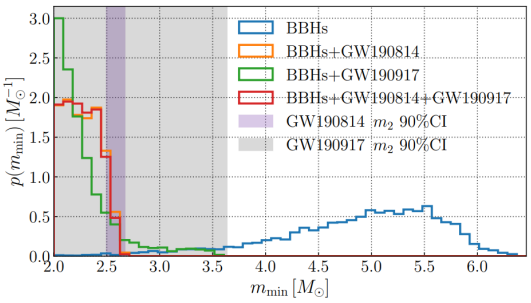

Figure 5. The posterior distribution of the minimum black hole mass inferred with the PP model. Posterior including and excluding GW190814 and GW190917 are shown. Inclusion of either event significantly impacted the distribution. Thus, we conclude that these two events may belong to a separate population that cannot be predicted by the rest of the events. (The LIGO Scientific Collaboration et al. 2021)

3.2.2. *GW190521: High Mass*

The primary mass of GW190521 is roughly 85 $M_\odot$, and has been flagged as a possible high-mass outlier in GWTC-2. As shown in Figure 8, the maximum mass predicted with or without GW190521 is centered around 85 $M_\odot$ and 70 $M_\odot$, respectively. The uncertainty of black hole mass on high mass events are often large, as the gravitational wave signal we detected are shorter and have lower frequencies. This is then reflected on the large spread on the probability distribution of maximum mass. The two distribution have significant overlap and we cannot conclude that GW190521 is an outlier. GW190521 also do not have significant impact on other hyper-parameters that defines the population as high mass events are, by nature, very rare in the total population.

It is worth noting that though we do not see GW190521 as an outlier, high mass events still lack an astrophysical explanation. GW190521, along with GW190602_175927 and GW190519_153544 (Abbott et al. 2021b), have primary mass above 45 $M_\odot$ at a $> 99\%$ confidence interval, above the mass range predicted by PPSN. They may involve with hierarchical mergers or other unknown astrophysical process.

3.2.3. *GW190412: Low mass ratio and well-measured tilt angle*

GW190412 is an outstanding event for both its unusually low mass ratio and its well-measured, positive aligned spin.

The two black holes of GW190412 have masses of around 30 $M_\odot$ and 8 $M_\odot$, making the mass ratio below 0.3, the lowest among all events. However, due to the high uncertainty on mass ratio and the scarcity of such unbalanced binaries, GW190412 does not significantly change the estimation of population mass ratio distribution, as shown in Fig 9.

GW190412 also has a relatively high constraining power on the component spin tilts ($cos\theta$) and effective inspiral spin distribution because this property is easier to measure for unbalanced pairs. The uncertainty of cos_t (tilt angle) of GW190412 is much smaller compared to the others. Here, σ_{spin} is the width of the aligned normal distribution; a lower σ_{spin} indicate that the population is more concentrated and more binaries are formed together. As shown in Figure 10, the inclusion of GW190412 shift the peak of σ_{spin} distribution to an lower value, and the uncertainty also dropped significantly. With future observation with lower uncertainty, we expect to have better estimation of hyper-parameters related with black hole spin and tilt.

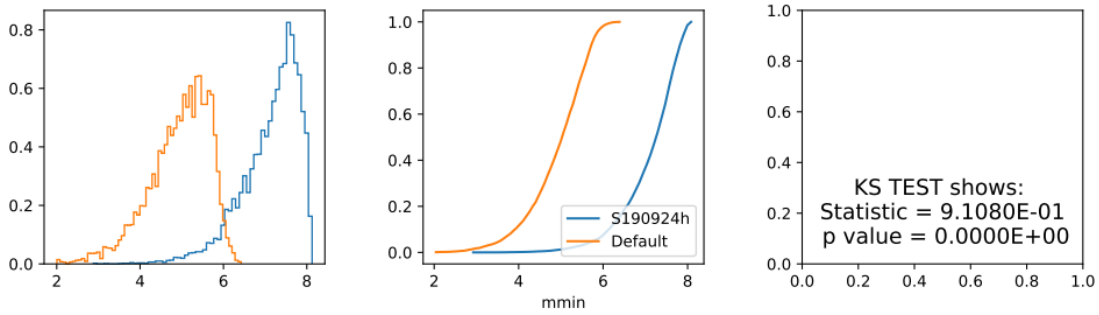

Figure 6. The plot on the left showed the posterior probability distribution of the minimum mass of black holes inferred from the default 69 events population and the population excluding GW190924, respectively. The middle plot shows the cumulative distribution that is used to conduct the KS test, and the KS test result is shown at the right.

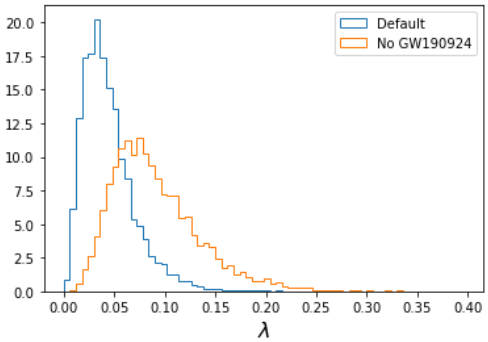

Figure 7. The figure shows posterior probability distribution of λ inferred by the default population and the leave-one-out test without GW190924. The default population predicts a much lower λ where smaller ratio of the total population are predicted to belong to the Gaussian peak.

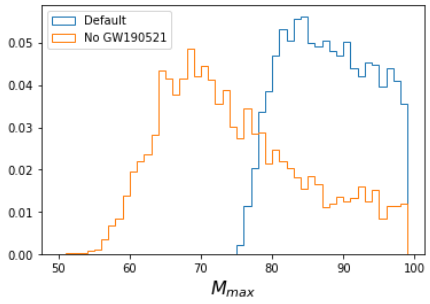

Figure 8. The figure shows posterior probability distribution of maximum mass inferred by the default population and the leave-one-out test without GW190521. The distribution stopped at 100 $M_\odot$ as the it is the highest allowed range of the prior.

4. REDSHIFT EVOLUTION OF BBHS MERGER RATE

In previous study of BBHs population, a simple power law is used to describe the redshift evolution of BBHs event rate, with more events at higher redshift. The most common redshift observed by LIGO is around 0.2.

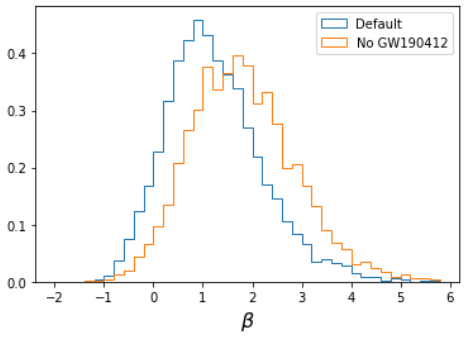

Figure 9. The mass ratio of BBHs primary and secondary mass is modeled with a power law where equal mass binaries are more common. The figure shows posterior probability distribution of mass ratio power law index inferred by the default population and the leave-one-out test without GW190412. The power law index is sightly higher when excluding the unbalanced binary GW190412, but the difference is small.

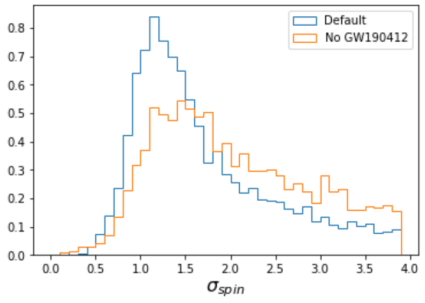

Figure 10. The figure shows posterior probability distribution of σ_{spin} inferred by the default population and the leave-one-out test without GW190412. Without GW190412, the estimation is more widespread and have higher uncertainties.

This is a combined result from the detection ability of the instrument and the black hole population with redshift evolution. GW190805W is the farthest event we have seen so far with an estimated redshift of 1. This

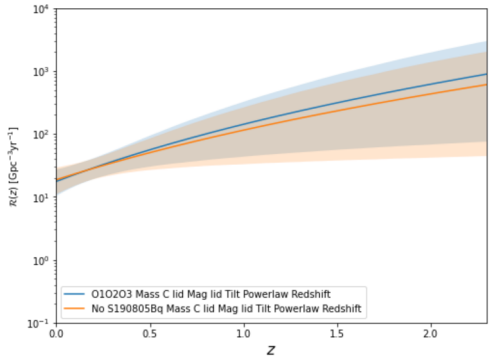

Figure 11. The plot shows how the inferred astrophysical BBH event rate change with the redshift. The shaded region marks the 90% confidence interval. The blue region shows the estimation including GW190805, and the orange region shows the estimation with it. The estimation has the lowest uncertainty at z 0.2 where most events happened.

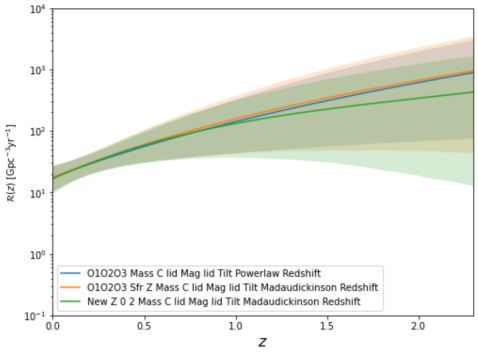

Figure 12. The plot shows BBHs rate evolution with redshift estimated with both simple power law and a broken power law motivated by cosmic stellar formation rate. The shaded area marks the 90% confidence interval. The blue region is predicted with a simple power law. The orange region is predicted by a broken power law with almost no constrain on peak position (at z between 0 to 10.) The uncertainty is slightly higher compare to the blue one due to more parameters presents in the broken power law model. The green region is also predicted by a broken power law, but with a more informative peak position (at z between 0 and 2.) We can see under this scenario, the rate may began to fall at higher redshift where observation is still technically possible.

event shed a light on the study of the redshift evolution of black hole population, yet with only one data point,it is hard to make any conclusion. However, in the upcoming fourth observing run (O4), with more advanced instruments, LIGO may expect several more of these high redshift events that can help constrain the redshift model. Therefore, it is worth to discuss the impact of high redshift events and a refined redshift model.

4.1. *GW190805W*

We use the same method of the leave-one-out test to see the impact of GW190805W on the estimation of the black hole population redshift evolution. Without GW190805W, the power law index of the redshift evolution is lower, but the uncertainty of the estimation does not change significantly (Figure 11). This is reasonable: the detection of one event would assure the abundance of BBHs at higher redshift, but there is not enough data to make a precise estimation. With more potential observations in O4, we hope to get a much better constraint.

4.2. *Madau-Dickinson Model*

Madau & Dickinson (2014) suggests that star formation rate does not grows indefinitely with redshift, but rather peaked at $z = 2$ 3 (the cosmological noon). The formation of stellar black holes should obey a similar trends with a time delay. The formation time of stellar black holes are relatively short, as their progenitors are massive stars with short lifespans. However, the in spiral time for BBHs can be very long can cause a significant time delay between the cosmology noon and the peak of BBHs mergers. Fishbach & Kalogera (2021) discusses this issue in great details. In practice, this time-delayed Madau-Dickenson model is describe by the broken power law, where the merger rate first goes up then drop down with growing redshift.

We run the population inference with all 69 events under three different model: first, a simple power law that does not have a turn; second, a broken power law with no constrain on the redshift of the "turning point" (In easier calculation purpose, we allow the turning point to be between z = 0 and z = 10. This is fine as z = 10 is way beyond the possible turning point.); and third, an adjusted broken power law that the turning point is only allowed to be happen later than the cosmological noon.(Between z = 0 and z = 2.) The result in Figure 12 shows that when their is no constrain, the power law and the Madau-Dickinson model have similar performance. However, if we only allow the turning point to be closer than z = 2, the distribution is more different than the default model: There would be much less events predicted at higher redshift.

5. POSTERIOR PREDICTIVE CHECKS

Under a predefined population model, population inference tell us the range of hyper-parameters that best describe the observations. However, the process lack a absolute measurement of how well the results describe the observation. We also cannot directly compare results that are based on different population model directly in the population inference.

Here, we use posterior predictive checks as a absolute measurement that characterize the agreement between the observation and the population posteriors. This allow us to compare different models, and potentially improve them.

5.1. *Methods: Sample Re-weighting*

Recall for single event parameter estimation, the parameters are estimated as follows:

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{P_{det}(\theta)} \tag{8}$$

Here, $\pi(\theta)$ describes our understanding of the probability of θ prior to the Bayesian inference. As we currently do not know much about the probability distribution of event parameters, uninformative priors are used during parameter estimation. This often means that all θ values in a certain range is assumed to have equal probability.

Recall that the posteriors from population inference describe the probability density function of the event parameter. These posteriors is now our "understanding of the probability of θ" and can substitute the uninformative priors. For instance, instead of assume that the mass of a black hole is evenly likely to be at anywhere between $2M_\odot$ to $100M_\odot$, we can say that its probability density follows a parameterized PP model, where the probability of having a smaller mass is larger than having a higher mass:

$$p_{re}(\theta|d) = \frac{p(\theta|d)p(\theta|\Lambda)}{\pi(\theta)} \tag{9}$$

$$p(\theta|\Lambda) = p_m.p_z.p_{mag}.p_{ori} \tag{10}$$

Here, $p(\theta|d)$ is the posterior probability of event parameters from the original event parameter inference, and $p_{re}(\theta|d)$ is the re-weighted posterior probability. $p(\theta|\Lambda)$ is the parameters' probability given a certain population model characterized by hyper-parameter Λ. It can be calculated as the product of the probability of the mass, the redshift, the spin magnitude, and the spin orientation.

We can perform similar sample re-weighting to an injected data set. We can test our population posterior by comparing the re-weighted observation event posteriors and the re-weighted injected event parameters. The population posterior is reasonable if the observation and the injected data set agrees with each other, and the underlying population model can describe the astrophysical results.

We compare the two data sets by comparing the re-weighted event parameters' cumulative distribution function(CDF). It is worth noting that generating CDFs for all event posterior samples is very computational costly. We also notice that the result change very little once the sample size is large enough. Thus, we only used a weight-selected sub-sample in our comparison.

In practice, we re-weighted all event posteriors based on a population posterior (a set of hyper-parameters.) We use the re-weighted probability as the weight to randomly select 10 posterior samples for each event, and 690 samples in total is selected. For the injection data set, we use the same false alarm rate and signal-to-noise threshold that are used for the observation signals. The injection events are then re-weighted, and 690 samples are selected from the entire injection set to match the sample size of the observation data. For all event parameter, a CDF is generated for each of the 690 samples. Then both the observation and the injection set is re-weighted based on a different population posterior, and two new CDFs are generated. Loop through all population posteriors, we obtained roughly 3000 CDFs for each parameters in both data sets. The 90% credible range is then plotted and compared.

5.2. *Result*

As shown in Figure 13 The observation and the injection data set agree with each other very well under the PP model. The 90% credible interval of the observation set completely overlap with the injection set. One thing to note is that there is a small peak around 15 $M_\odot$ that appears in the observation set. This peak is not statistically significant as it still overlaps with the injection set, but it might worth future investigations.

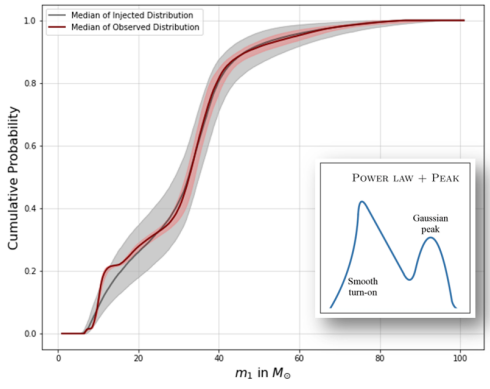

Figure 13. This figure shows the result of the posterior predictive check of the PP model. The 90% credible interval of the black hole mass CDF is shown. The narrow pink band is from the re-weighted observation samples, and the wide grey band represent the re-weighted injection sample. The Gaussian peak is well captured in this model and the two sets agree.

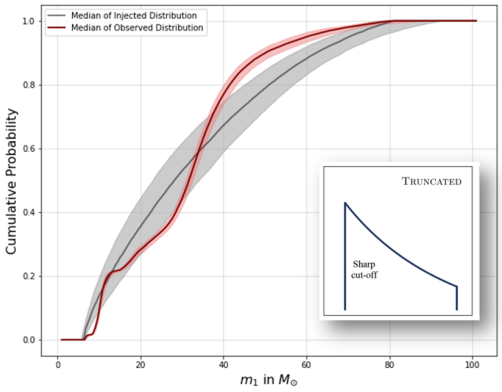

Figure 14. This figure shows the result of the posterior predictive check of the *TRUNCATED* model model. The 90% credible interval of the black hole mass CDF is shown. The narrow pink band is from the re-weighted observation samples, and the wide grey band represent the re-weighted injection sample. It is obvious that the simple power law does not capture the details that appeared in the observations.

A similar comparison is done by using the *TRUNCATED* model. This model is just a simple power law with no additional features and sharp cutoff at the minimum and maximum mass. As shown in Figure 13, the observation set does not agree with the injection set. This suggest that the *TRUNCATED* model cannot fully captures the astrophysical population of black holes. With many more observations expected in O4, other detailed features of the mass population may be revealed. (For instance, we may confirm the existence of the second peak at 15 $M_\odot$.) This method can be used to test new or improved models that describe such features.

6. CONCLUSION

This project aims to test the current model that describe the binary black hole population observed by LIGO.

We first conduct leave-one-out test to identify potential outliers that may represent different populations. We did not find any new statistically significant outliers. However, several events have special features that may bear further investigation. In particular, GW190924 has a very low secondary mass; GW190521 has a very high primary mass; and GW190412 is a very unbalanced pair and has very well defined spin measurements.

We also examine the effect of adding redshift evolution of black hole binary populations into the population model. We currently only have one event, GW190805W, that ha2 a redshift of roughly 1. However, more high redshift events are anticipated in the upcoming observations. We believe even a few number of these events can significantly help us better constrain the redshift model. We also suggest that the time-delayed Madau-Dickinson model can be use to better describe black hole population at higher redshift.

We finally conduct posterior predictive checks to absolute measurement of how well our population model describe the observations. The result shows that under PP model, the observed and inferred population agree.

REFERENCES

Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, SoftwareX, 13, 100658, doi: 10.1016/j.softx.2021.100658

—. 2021b, ApJL, 913, L7, doi: 10.3847/2041-8213/abe949

Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27, doi: 10.3847/1538-4365/ab06fc

Fishbach, M., & Kalogera, V. 2021, ApJL, 914, L30, doi: 10.3847/2041-8213/ac05c4

Hogg, D. W. 1999, arXiv e-prints, astro. https://arxiv.org/abs/astro-ph/9905116

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415, doi: 10.1146/annurev-astro-081811-125615

Talbot, C., Smith, R., Thrane, E., & Poole, G. B. 2019, PhRvD, 100, 043030, doi: 10.1103/PhysRevD.100.043030

Talbot, C., & Thrane, E. 2018, ApJ, 856, 173, doi: 10.3847/1538-4357/aab34c

The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2021, arXiv e-prints, arXiv:2111.03634. https://arxiv.org/abs/2111.03634

Veitch, J., Raymond, V., Farr, B., et al. 2015, PhRvD, 91, 042003, doi: 10.1103/PhysRevD.91.042003

Woosley, S. E. 2017, ApJ, 836, 244, doi: 10.3847/1538-4357/836/2/244