The impact of astrophysical population model choices on post-Newtonian deviation tests of general relativity

Ruby Knudsen, Mentor: Ethan Payne

September 2024

Abstract

The Laser Interferometer Gravitational-wave Observatory detects gravitational waves and uses them to test the theory that predicts their existence: Einstein's theory of general relativity. Testing general relativity using gravitational waves can be done at the individual event level and the population level. Most current tests of general relativity are not inferred jointly with a population model to describe the astrophysical distribution of the sources. The omission of a population model describing the astrophysical population of binary black hole mergers inscribes an implied population model that may be inaccurate. This inaccurate population model could lead to biases in supposed deviations from general relativity. This investigation uses injected signals of simulated binary black hole systems to infer the probability distributions for population model parameters while simultaneously fitting the GR deviation population. These probability distributions are constructed using Markov Chain Monte Carlo analysis. A family of mass population models is injected and recovered using each of the different models as the population distribution to determine whether biases in deviations from general relativity and population misspecification can be observed. It is found that using a misspecified astrophysical population model leads to inaccurate recoveries of GR deviation parameters. Further research is needed as more gravitational wave events are detected and our understanding of the mass population distribution evolves.

1 Background

In the early 20th century, Albert Einstein put forward the theory of general relativity (GR), and summarized his theory mathematically through the Einstein field equations. Part of GR theorized the concept of spacetime, which spatially distorts in the vicinity of massive astrophysical objects. Building off of this work, two years later Einstein posited the existence of gravitational waves (GW) — spatial distortions as transverse waves that originate from non-spherically symmetric energy density distributions, before traveling away at the speed of light [2, 12]. Because GR theorizes the existence of GWs, it is possible to use GWs as a method to test GR.

Different tests of GR interrogate the theory to different levels, as summarized by Fig. 1. While all astrophysical objects have gravitational potential and cause spacetime curvature — a key insight of GR — the most extreme tests of GR exist in the strong-field regime. The gravitational potential is proportional to M/R^3, and the curvature is proportional to M/R. The perihelion procession of Mercury probes GR, but not strenuously, since it has potential between 10^{-7} and 10^{-8}, and curvature between 10^{-32} and 10^{-33} cm^{-2} [18]. Due to the Schwarzschild radius, the event horizon of a black hole yields potential approximately 0.5, and is the most extreme possible potential. While black holes Sagittarius A* and M87 both have potential near 0.5, their curvature is orders of magnitude lower than the spacetime curvature near merging stellar mass black hole binaries [9]. Curvatures between 10^{-14} and 10^{-10} cm^{-2} are the most extreme for realistic astrophysical sources [9]. Probing GR in this region requires the detection of GWs from black hole and neutron star mergers. Thus, GW tests probe GR at the most extreme curvatures and potentials, and as such are the best tests of GR.

The Laser Interferometer Gravitational-wave Observatory (LIGO) — a joint project between the California Institute of Technology, the Massachusetts Institute of Technology, and the National Science Foundation — was built to detect these GWs and aims to further our understanding of GWs and test GR [1, 6]. LIGO

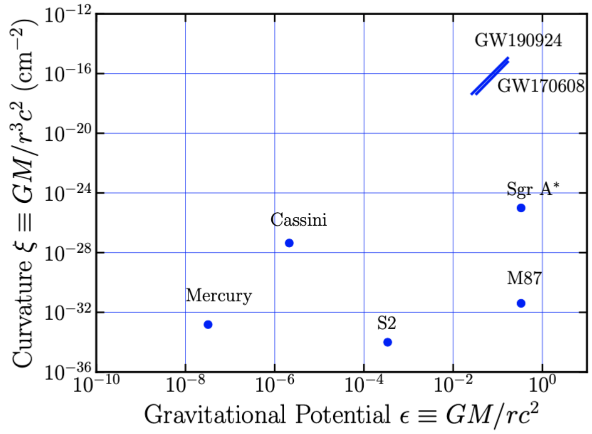

Figure 1: Plot of gravitational potential vs. curvature for various tests of GR from Ref. [18]. The theoretical limit for the gravitational potential exists between 0.5 and 1.

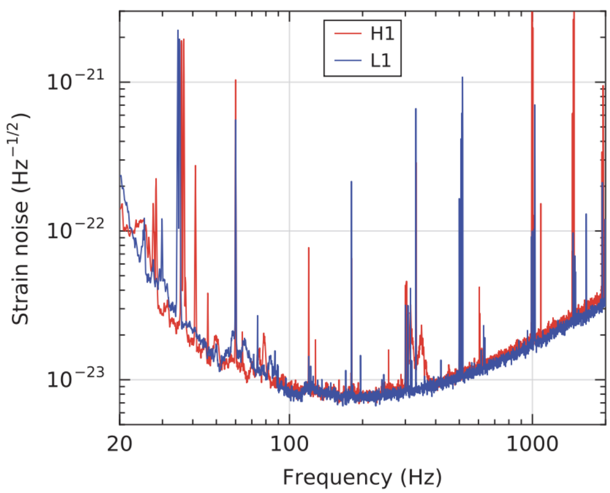

Figure 2: From Ref. [2]. This figure shows the sensitivity curve of LIGO, at different frequencies from the first observing run. LIGO is most sensitive to frequencies between 20 Hz and 1 kHz.

has perpendicular arms that detect compact binary mergers using the interference of light due to the difference in length of the arms that a passing GW causes. This change in length is directly proportional to GW strength [1, 6]. The compact binary mergers, mostly binary black hole mergers, are then analyzed to check for consistency with GR. LIGO is sensitive to GW frequencies between 20 and 1000 Hz, so black holes of stellar mass are easiest to detect [13]. Fig. 2 shows the sensitivity at each frequency; the lower strain noise in frequencies associated with orbital velocities near the merger of binary stellar mass black holes corresponds to better sensitivities at these frequencies.

2 Introduction to methods for testing GR

2.1 Modeling modifications to GR with individual observations

There are many different ways of modeling GR deviations by modeling the GW signal. One of these tests is inferring deviations from the post-Newtonian (PN) approximation. The post-Newtonian approximation at the lowest order is the quadrupole formula, which estimates the emitted radiation from a quadrupolar mass distribution [17]. The PN expansion for gravitational-wave emission is an expansion in the dimensionless parameter v/c, where v is the orbital velocity of the binary [20, 10]. PN tests are run by first constructing a post-Newtonian description of the GW inspiral in the frequency domain, before making modifications to individual parameters in the phase evolution [8, 19]. This is done with the phase, not the amplitude, because LIGO is more sensitive to phase deviations than those of amplitudes; over the course of an inspiral, phase shifts accumulate, while amplitude differences do not.

$$\Phi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128} \times \sum_{k=0}^{7} \frac{1}{\eta^{\frac{k}{5}}} (\varphi_k + \varphi_{k,l} \ln \tilde{f}) \tilde{f}^{\frac{k-5}{3}} \tag{1}$$

In this equation, $\Phi(f)$ is the frequency domain GW phase under the stationary phase approximation $\tilde{f} = \mathcal{M}\{/\rfloor^{\ni}$, where $\mathcal{M}$ is the red-shifted chirp-mass multiplied. Additionally, t_c and ϕ_c are the coalescence time and phase respectively, η is the symmetric mass ratio, and φ_k, $\varphi_{k,l}$ are PN coefficients for the $k/2$ PN order [8, 19].

After running these PN tests, a posterior is obtained by running Bayesian inference. Bayesian inference makes statements about the Universe from data by inferring the probable distribution for parameters of each event analyzed, known as the posterior distribution [22]. Calculating this posterior distribution requires a likelihood, to describe the measurement, and a prior, to encode prior beliefs about each event's parameters θ. Bayesian inference uses Bayes' Theorem,

$$p(\theta|d) = \frac{p(d|\theta)\pi(\theta)}{\mathcal{Z}} \tag{2}$$

In this equation, $p(\theta|d)$ is the posterior, $p(d|\theta)$ is the likelihood, and $\pi(\theta)$ is the prior, and the evidence, which normalizes the distribution, is denoted as $\mathcal{Z}$ [22]. This same formalism can be used to study multiple events, with a modified version of the likelihood function.

2.2 Testing GR with many observations

Hierarchical inference is a method that uses posterior distributions from individual events to model the overall population of the astrophysical distribution from which the events originated. The analysis returns a distribution for the hyper-parameters of the population model, which is called the hyper-posterior [7, 22]. In this hyper-posterior, the astrophysical population has been inferred under the assumption of the choice of population model. A population model is important because otherwise an incorrect astrophysical population may be implicitly assumed, which in turn leads to bias in supposed measured deviations from GR [16].

To look at the population properties of a collection of events, the prior for the parameters θ is made conditional on hyper-parameters Λ, which encodes the astrophysical distribution from which the θs are drawn [22]. These hyper-parameters are crucial because they parameterize the shape of the inferred astrophysical population, and one goal of population inference is estimating the posterior of these hyper-parameters. One key piece of estimating the hyper-posterior is the population likelihood [16, 15].

$$p(\{d\}|\Lambda) = \frac{1}{\xi(\Lambda)^N} \prod_{i=1}^{N} \int d\theta_i p(d_i|\theta_i)\pi(\theta_i|\Lambda) \tag{3}$$

Here, $\{d\}$ is the collection of N observations, $\xi(\Lambda)$ accounts for selection biases and is the detectable fraction of observations given the population hyper-parameters, $p(d_i|\theta_i)$ is the likelihood for each individual event, and $\pi(\theta_i|\Lambda)$ are the population distributions evaluated at each event [22]. Thus, $\pi(\theta_i|\Lambda)$ is where the population distribution is encoded.

This equation shows how individual observations are put together for hierarchical inference. As stated earlier, this hierarchical approach requires a population model, so we therefore must choose one population model to use. Currently, we are not sure in what manner the different choices of population model impact GR tests. This leads to my project.

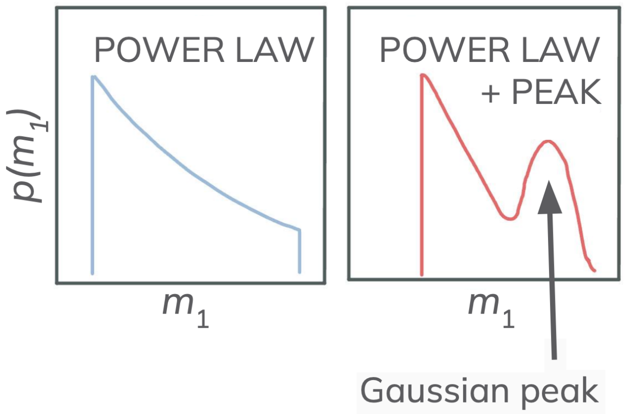

Figure 3: Visual representation of the Power law and Power law + Peak population distributions. Both are truncated, and thus do not have a smooth turn-on.

3 Project motivation

Population distributions depend in part on the choice of population model, thus the choice of population model must be carefully considered. So far, this approach of joint inference has, at the population level, yielded GR deviations more consistent with GR by about 0.4σ, when using a Power law + Peak population distribution for the mass of a black hole [16, 3]. However, it is possible that deviation from general relativity could be absorbed or hidden by an incorrect, assumed astrophysical population, so it is important to study the impact of different astrophysical population models on inferred GR deviation constraints. It is true that at some point that the incorrect choice will also lead to biases, the question is when, and how badly. Testing this using different population models is what I am working on.

4 Methods

Hierarchical inference of astrophysical populations jointly with general relativity deviation parameters relies in part on the astrophysical population model, and the GR deviation population model. For this project, the main astrophysical parameter examined (along with GR deviation) was primary mass, along with mass ratio. There are many ways to model the primary mass distribution, through various parametric models and other methods [21]. This project chose two relatively simple primary mass models: Power law and Power law + Peak, shown in Fig. 3 [3, 4, 5]. The Power law model can be described by the Power law + Peak model, with a peak fraction of 0. However, the Power law + Peak cannot be described by a Power law. The population of GR deviations were modelled as a Gaussian that approaches a delta function when GR is correct [14].

For each primary mass model, Power law and Power law + Peak, a catalog of events was generated following this model. This was accomplished by injecting a value for the Power law index, α. The injected value was passed into the log-likelihood function, along with the linear space of possible primary mass values — between 5 and 100 solar masses. The result for each point in space of the primary mass grid was stored, and then interpolated to create the inverse of the cumulative distribution function (CDF). Finally, random points along this inverse CDF were selected, and the primary mass associated was added to the event catalog.

We used GWFish to sample from the simulated event catalog in a way that was astrophysically realistic and waveform-motivated, as well as to correlate the deviation parameter to the primary mass distribution in a way that imitated post-Newtonian deviation tests [11]. GWFish creates simulated posteriors that are astrophysically motivated by Fisher matrices of futuristic detectors. The detectors simulated in this project were the Einstein Telescope and the Cosmic Explorer. It was assumed that Einstein Telescope and Cosmic Explorer could detect all events. The waveform model used was IMRPhenomD.

GWFish creates a Fisher matrix for a catalog of events. The inverse of that Fisher matrix is the covariance

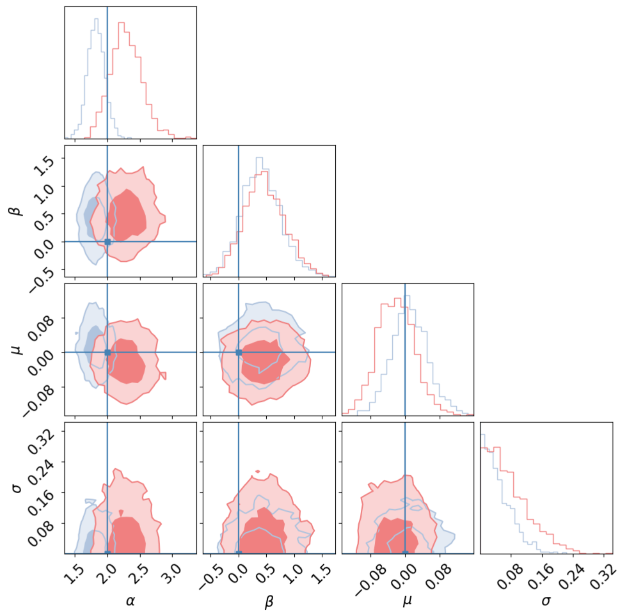

Figure 4: Corner plot of a MCMC where Power law + peak was injected (with TGR mean and standard deviation of 0), and then both Power law and Power law + peak were recovered. In the pink is recovery with the correct population model (Power law + peak) and in the blue is recovery with the misspecified population model (plain Power law).

matrix. GWFish also calculates the network SNR for each event. In order to correlate the deviation parameter to the mass, the covariance matrix was expanded from just chirp mass and mass ratio to chirp mass, mass ratio, and deviation parameter. This was done using the rotation of the covariant matrix and the network SNR. The deviation parameter variance was assumed to be 10/(network SNR). This covariance matrix was then used with a multivariate Gaussian that had means at the true value for each chirp mass, mass ratio, and deviation parameter. The Gaussian took $10,000$ samples from each event, and shifted them so that the true value was one of the observed values. Despite the goal of having more astrophysically realistic samples with GWFish, the sampling returned some astrophysically impossible values, and had to be constrained after the fact. Once each event had a posterior with $10,000$ astrophysically realistic samples, a Markov chain Monte Carlo was run. Each model (Power law, Power law + Peak) was injected, and then recovered with Power law and Power law + Peak.

5 Results and Analysis

Unless explicitly mentioned, each event catalog had $N = 100$ simulated events. This number was chosen to balance rigor with computational and time limits; further research should incorporate much great Ns, in additional to more than $10,000$ samples per event. Each corner plot displays the 50 percent and 90 percent confidence intervals.

In each corner plot, α refers to the Power law index for the primary mass, β refers to the Power law index for the mass ratio, μ refers to the mean for the testing general relativity deviation parameter, and σ refers to the standard deviation for the testing general relativity deviation parameter. In all cases, α was injected to be 2, β was injected to be 0, and μ was injected to be 0. There were σ injections of 0, 0.25, and 0.5.

5.1 Injecting consistency with GR

In Fig. 4, a Power law + peak population was injected, and was recovered with the Power law + peak (in pink) and Power law (in blue) models. Thus, the pink is a correctly specified population model,

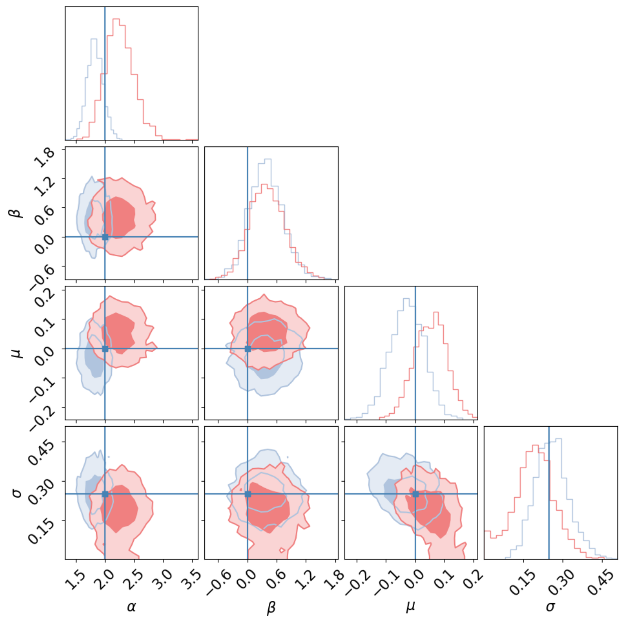

Figure 5: Corner plot of a MCMC where Power law + peak was injected, but this time with TGR mean of 0 and standard deviation of 0.25. This population was then recovered with both Power law and Power law + peak. In the pink is recovery with the correct population model (Power law + peak) and in the blue is recovery with the misspecified population model (plain Power law)

and the blue is a misspecified population model. It is evident that the misspecified population model does not recover α correctly — there is an obvious shift between the pink and blue histograms. This makes sense; when injecting with a Power law + peak model, and recovering with just a Power law, α must then account for the peak as well, and this skews the inferred hyper-posterior. Additionally, β mostly overlaps between correctly and misspecified population models. With the TGR parameters, μ shows a small but noticeable shift between pink and blue — while $\mu = 0$ is in the range of both, it is evident that the misspecified does not fully recover μ correctly. Furthermore, the σ value recovered by the misspecified population model is artificially constrained when compared with the correctly specified (pink) population model recovery. Therefore, both TGR parameters are not recovered correctly with a misspecified population model. However, it is important to note that the injected values are still present in the distributions for both parameters; without additional tests, it would not be possible to identify that the population model was misspecified.

5.2 Injecting sigma of 0.25 for the deviation parameter distribution

In Fig. 5, again a Power law + peak population was injected, and was recovered with the correctly specified Power law + peak (in pink) and the incorrectly specified Power law (in blue) models. Here, unlike above, σ is injected to be 0.25 — this corresponds to a measurable GR modification. It is again evident that the misspecified population model in blue does not recover α correctly — there is an obvious shift between the pink and blue histograms. Here, there is still mostly an overlap of β between correctly and misspecified population models.

With the TGR parameters, μ shows a larger shift between pink and blue than when $\sigma = 0$. Still, $\mu = 0$ is in the range of both. However, this larger μ shift shows that the misspecified population model does not recover μ correctly. Furthermore, the σ value recovered by the misspecified population model (blue) is now shifted when compared with the correctly specified (pink) population model recovery. Therefore, when $\sigma = 0.25$, both TGR parameters are not recovered correctly with a misspecified population model, and the differences in recovery of these two parameters when comparing population models is more pronounced than in the $\sigma = 0$ case. However, it is important to note that the injected values are still present in the

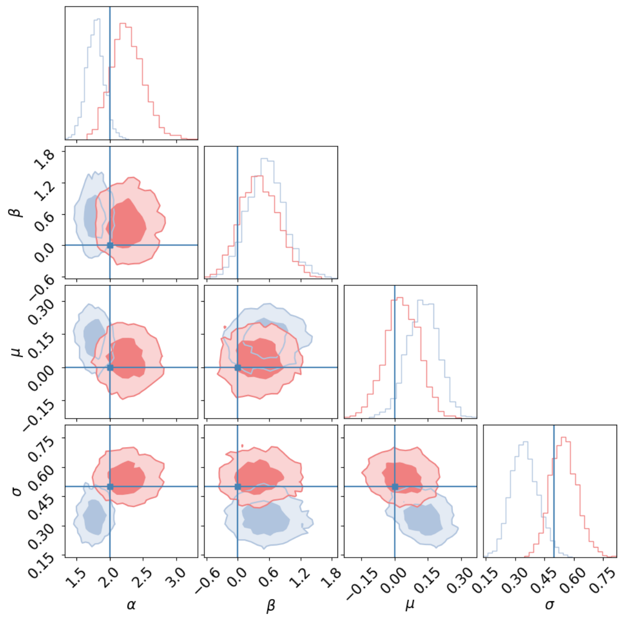

Figure 6: Corner plot of a MCMC where Power law + peak was injected, but this time with TGR mean of 0 and standard deviation of 0.5. This population was then recovered with both Power law and Power law + peak. In the pink is recovery with the correct population model (Power law + peak) and in the blue is recovery with the misspecified population model (plain Power law)

distributions for both parameters in these distributions.

5.3 Injecting sigma of 0.5 for the deviation parameter distribution

In Fig. 6, as above, a Power law + peak population was injected, and was recovered with the correctly specified Power law + peak (in pink) and the incorrectly specified Power law (in blue) models. This time σ is was injected to be 0.5 — GR is injected to be modified again. We see similar shifts in α and β as sigma increases.

With the TGR parameters, μ shows a much larger shift between pink and blue than when $\sigma = 0$ or 0.25. In the μ vs. σ plot, the point $(\mu, \sigma) = (0, 0.5)$ is not present when recovering with the misspecified population model (in blue). This time, $\mu = 0$ is barely in the range of the misspecified population model; the misspecified population model is recovering $\mu > 0$ even though this was not injected. Furthermore, the σ value recovered by the misspecified population model (blue) is now shifted too low when compared with the correctly specified (pink) population model recovery. Therefore, when $\sigma = 0.5$, both TGR parameters are obviously recovered incorrectly, with σ being recovered at lower than 0.5 as the 50 and 90 percent confidence intervals, and μ being recovered at a non-zero value. Unlike the above cases, it is important to note that the injected values are not present in the distribution for the misspecified population model.

5.4 Increasing event population size N and injecting consistency with GR

In Fig. 7 and 8, GR was injected to be *True* (μ and σ were both injected to be 0), but the number of events was $N = 250$ and $N = 500$ respectively. Again, while the injected values are present in the correctly and incorrectly recovered distributions, there is a shift. As N increases, so too does a shift in recovered μ value. Unlike the case where $N = 100$, however, σ does not seem to be artificially constrained, and seems to be recovered nearly identically between correctly specified and misspecified population models. Thus, recovery of μ becomes less accurate with an increase in event catalog size. However, much more work is needed with larger sizes to draw more permanent conclusions.

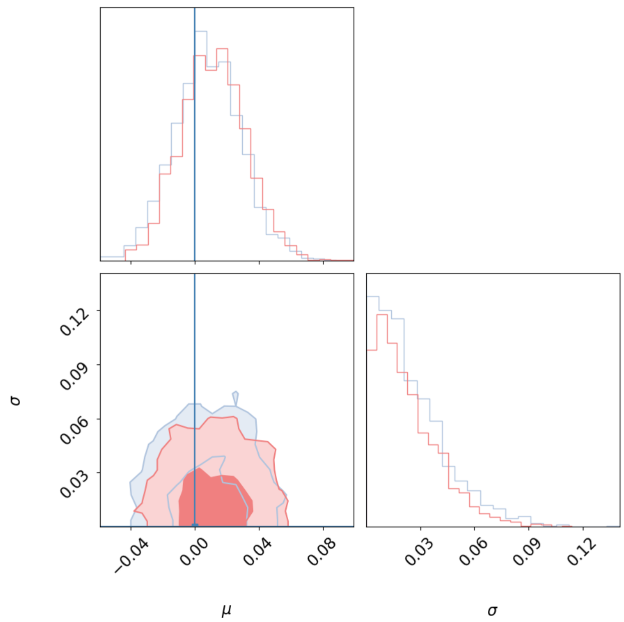

Figure 7: Corner plot of a MCMC where Power law + peak was injected (with TGR mean and standard deviation of 0), and then both Power law and Power law + peak were recovered. However, here the number of events is increased to $N = 250$. In the pink is recovery with the correct population model (Power law + peak) and in the blue is recovery with the misspecified population model (plain Power law).

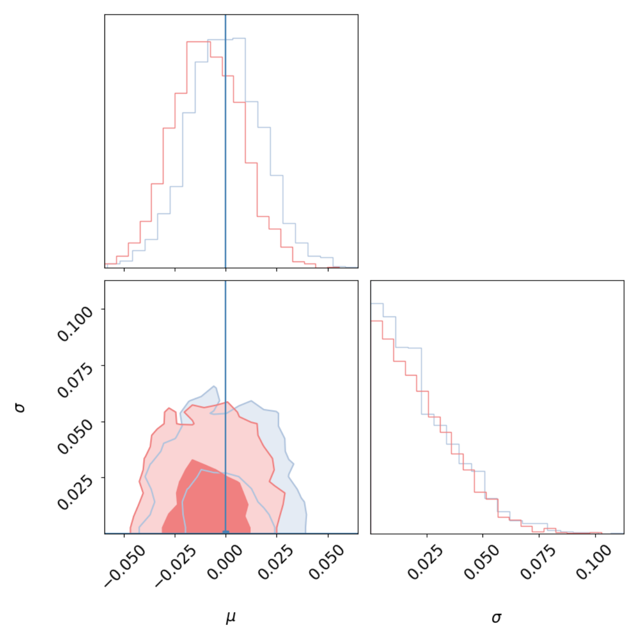

Figure 8: Corner plot of a MCMC where Power law + peak was injected (with TGR mean and standard deviation of 0), and then both Power law and Power law + peak were recovered. However, here the number of events is increased to $N = 500$. In the pink is recovery with the correct population model (Power law + peak) and in the blue is recovery with the misspecified population model (plain Power law).

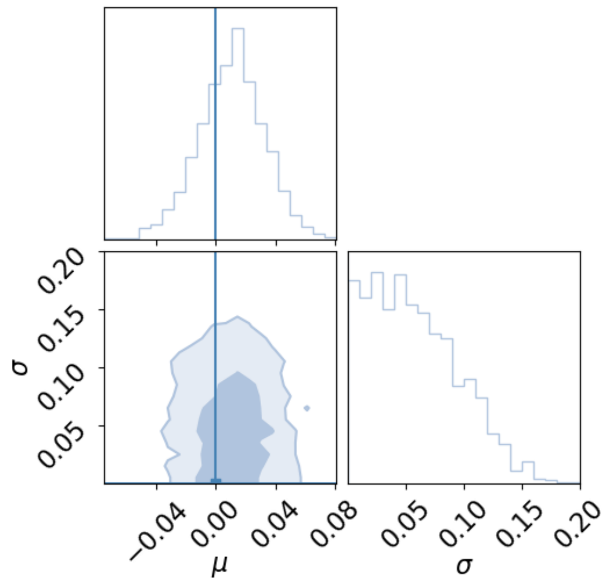

Figure 9: Corner plot of a MCMC where a Gaussian population was injected (with TGR mean and standard deviation of 0), and then Power law was recovered with. $N = 500$

5.5 Injecting extremely misspecified population models

These looked at the implications of using an extremely misspecified population model. First, in Fig. 9 a Gaussian population was injected and then recovered with a Power law. Next, responding to a question posed during a presentation of this project, in Fig. 10 a uniform population was injected and then recovered with a Power law. Since these populations were not recovered using the correct model, major meaningful conclusions and comparisons are not possible. However, it is noteworthy that in each case, μ and σ of 0 are recovered as possible values. This brings up an interesting case where population model does not matter as much *only when GR is correct*. Further research into this possibility should be completed.

6 Implications and Conclusions

Based on the results from this project, it is clear that using a misspecified astrophysical population model leads to inaccurate recoveries of GR deviations. Even in Fig 4, when recovered incorrectly, σ is artificially constrained. When σ is injected to be greater than zero, such as 0.5 in Fig. 6, is is obvious that σ and μ are recovered incorrectly. As a result of these incorrectly recovered TGR parameters, it could therefore be plausible that inconsistencies with GR are masked by using a misspecified astrophysical population model. Even if this is not the case, the TGR parameters are still not being recovered correctly.

In this project, a hierarchical test of general relativity was completed. Hierarchical TGR relies in part on an astrophysical population model. The model used for GR deviations was Gaussian, which goes to a delta function when GR is true. To achieve this, simulated catalogs of 100 events were created. These event catalogs followed Power law and Power law + peak primary black hole mass population models. The catalogs were generated by interpolating over a cumulative distribution function for the possible space of primary masses using the Power law and Power law + peak probability density functions. These event catalogs were then sampled with GWFish. GWFish was then used to manually correlate the primary mass and GR deviation parameter to simulate PN deviation test. Finally, Markov chain Monte Carlo analysis was used to recover GR deviation parameters (and population model parameters) using each model.

Further research is necessary into population model choices. This project focused on mass distribution models, which evolve as more gravitational wave events are detected. However, the properly specified population distribution for other parameters, such as spin and tilt, should also be investigated. More work

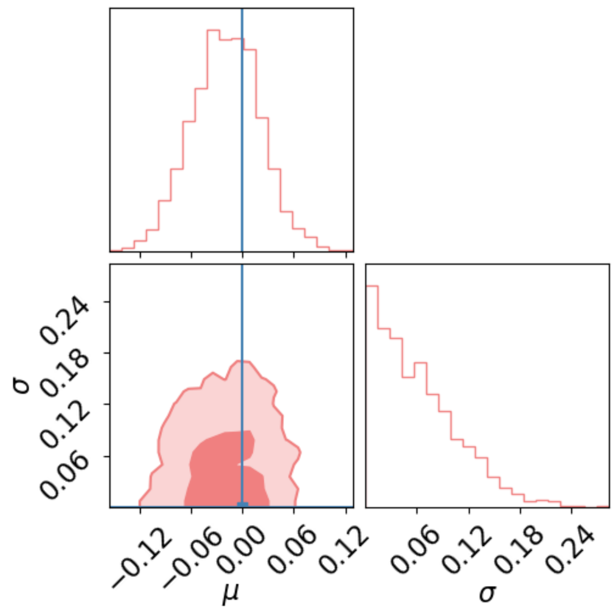

Figure 10: Corner plot of a MCMC where a UNIFORM population was injected (with TGR mean and standard deviation of 0), and then POWER LAW was recovered with. $N = 100$

also must be done to check for model misspecification using multiple techniques in conjunction, including posterior predictive checks. Finally, other models need to be used for similar simultaneous inference of TGR and astrophysical population distribution.

7 Acknowledgments

This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs. I would like to thank my mentor, Ethan Payne, my fellow LIGO SURF-ers, my family and friends, and all of the faculty and mentors at the LIGO SURF program, especially Alan Weinstein and Jonah Kanner.

We made use of the following code packages: Numpy, Jax, JaxNumpy, Numpyro, Matplotlib, and GW-Fish.

References

[1] Junaid Aasi, BP Abbott, Richard Abbott, Thomas Abbott, MR Abernathy, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Advanced ligo. *Classical and quantum gravity*, 32(7):074001, 2015.

[2] Benjamin P Abbott, Richard Abbott, TDe Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, Rana X Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Physical review letters*, 116(6):061102, 2016.

[3] R Abbott, TD Abbott, F Acernese, K Ackley, C Adams, N Adhikari, RX Adhikari, VB Adya, C Affeldt, D Agarwal, et al. Population of merging compact binaries inferred using gravitational waves through gwtc-3. *Physical Review X*, 13(1):011048, 2023.

[4] R Abbott, H Abe, F Acernese, K Ackley, N Adhikari, RX Adhikari, VK Adkins, VB Adya, C Affeldt, D Agarwal, et al. Tests of general relativity with gwtc-3. *arXiv preprint arXiv:2112.06861*, 2021.

[5] Rich Abbott, TD Abbott, S Abraham, Fausto Acernese, K Ackley, A Adams, C Adams, RX Adhikari, VB Adya, Christoph Affeldt, et al. Population properties of compact objects from the second ligo–virgo gravitational-wave transient catalog. *The Astrophysical journal letters*, 913(1):L7, 2021.

[6] Fet al Acernese, M Agathos, K Agatsuma, Damiano Aisa, N Allemandou, Aea Allocca, J Amarni, Pia Astone, G Balestri, G Ballardin, et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, 2014.

[7] Matthew R Adams, Neil J Cornish, and Tyson B Littenberg. Astrophysical model selection in gravitational wave astronomy. *Physical Review D*, 86(12):124032, 2012.

[8] KG Arun, Bala R Iyer, Bangalore Suryanarayana Sathyaprakash, and Pranesh A Sundararajan. Parameter estimation of inspiralling compact binaries using 3.5 post-newtonian gravitational wave phasing: The nonspinning case. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, 71(8):084008, 2005.

[9] Tessa Baker, Dimitrios Psaltis, and Constantinos Skordis. Linking tests of gravity on all scales: from the strong-field regime to cosmology. *The Astrophysical Journal*, 802(1):63, 2015.

[10] Luc Blanchet and Thibault Damour. Post-newtonian generation of gravitational waves. *Annales de l'IHP Physique théorique*, 1989.

[11] U. Dupletsa, J. Harms, B. Banerjee, M. Branchesi, B. Goncharov, A. Maselli, A. C. S. Oliveira, S. Ronchini, and J. Tissino. GWFISH: A simulation software to evaluate parameter-estimation capabilities of gravitational-wave detector networks. *Astronomy and Computing*, 42:100671, January 2023.

[12] Albert Einstein and Emil Warburg. *Die Relativitätstheorie*. Springer, 1911.

[13] Davide Gerosa, Geraint Pratten, and Alberto Vecchio. Gravitational-wave selection effects using neural-network classifiers. *Physical Review D*, 102(10):103020, 2020.

[14] Maximiliano Isi, Katerina Chatziioannou, and Will M Farr. Hierarchical test of general relativity with gravitational waves. *Physical Review Letters*, 123(12):121101, 2019.

[15] Ryan Magee, Maximiliano Isi, Ethan Payne, Katerina Chatziioannou, Will M. Farr, Geraint Pratten, and Salvatore Vitale. Impact of selection biases on tests of general relativity with gravitational-wave inspirals. *Phys. Rev. D*, 109(2):023014, 2024.

[16] Ethan Payne, Maximiliano Isi, Katerina Chatziioannou, and Will M. Farr. Fortifying gravitational-wave tests of general relativity against astrophysical assumptions. *Phys. Rev. D*, 108(12):124060, 2023.

[17] William H Press and Kip S Thorne. Gravitational-wave astronomy. *Annual Review of Astronomy and Astrophysics*, 10(1):335–374, 1972.

[18] Dimitrios Psaltis, Colm Talbot, Ethan Payne, and Ilya Mandel. Probing the black hole metric: Black hole shadows and binary black-hole inspirals. *Physical Review D*, 103(10):104036, 2021.

[19] Bangalore Suryanarayana Sathyaprakash and SV Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. *Physical Review D*, 44(12):3819, 1991.

[20] Hideyuki Tagoshi, Masaru Shibata, Takahiro Tanaka, and Misao Sasaki. Post-newtonian expansion of gravitational waves from a particle in circular orbit around a rotating black hole: Up to o (v 8) beyond the quadrupole formula. *Physical Review D*, 54(2):1439, 1996.

[21] Colm Talbot and Eric Thrane. Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization. *The Astrophysical Journal*, 856(2):173, 2018.

[22] Eric Thrane and Colm Talbot. An introduction to bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36:e010, 2019.