Improving Posterior Predictive Checks for Gravitational Wave Population Analyses

Sophia Winney

(Dated: August 2, 2024)

In population analyses of gravitational waves emitted from binary black holes (BBH), spin magnitude, polar angle, and azimuthal angle distributions provide critical insights regarding BBH evolutionary histories and formation channels, yet studies have presented conflicting conclusions about the nature of these distributions. For this reason, developing reliable models of BBH spin component distributions continues to be an essential problem. However, the effects of spin magnitude and directions on gravitational wave signals are subdominant compared to the influence of the effective aligned spin and the effective precessing spin parameters. The resulting degeneracy of spin components poses challenges for constraining models of their astrophysical distributions and determining model accuracy. Posterior predictive checks, a widely used model-checking method in gravitational wave science, especially fall short when applied to models based on data with high uncertainties. In this project, we implement two alternative methods for predictive checking: partial posterior predictive checks and split predictive checks. We aim to explore the efficacy of these methods by applying them to models of varying accuracy of simulated astrophysical populations with the same effective-spin distribution and different spin component distributions.

I. INTRODUCTION

The continued success of gravitational wave observation has permitted population analyses of binary black hole (BBH) mergers [1–3]. Such events can be described by the spin components of the primary and the secondary, entailing each of their spin magnitudes (χ_i), azimuthal angles (ϕ_i), and polar angles (θ_i) [4]. Spin magnitudes provide a probe into the angular momentum processes within stellar cores, and large magnitudes can indicate hierarchical black hole formation through previous mergers. Previous studies have favored small but non-zero spins, as well as a wide range of polar angles; conflicting work has favored a majority of spin zero black holes and remaining nonzero spins that are primarily aligned with orbital angular momentum [5].

These opposing conclusions have significant implications for the prevalence of different BBH formation channels. The isolated binary formation channel consists of binary star systems in which one star becomes a black hole, resulting in mass transfer and an eventual merger in an observable amount of time. These systems tend to exist for long periods of time, which is conducive to a majority of spins that are aligned with orbital angular momentum. Conversely, dynamic formation occurs in dense environments where black holes of similar mass congregate and become gravitationally bound into binary systems. Such dynamically formed systems lack tidal effects and mass transfer, thus favoring a random distribution of spin alignment [6]. Hierarchical mergers—which often occur in galactic nuclei, active galactic nuclei, and extremely low-mass ultrafaint dwarf galaxies—may increase BBH eccentricities and quicken inspiral, increasing the likelihood of a merger within Hubble time and further complicating the picture of BBH formation [6].

Although highly significant, the parameters χ_i, ϕ_i, and θ_i induce subdominant effects on gravitational wave signals, which are instead primarily influenced by the effective aligned spin (χ_{eff}), containing the spin components that are aligned with the orbital angular momentum, and effective precessing parameter (χ_p), encompassing the anti-aligned components. The lower dimensionality of informative parameters makes it difficult to ascertain the underlying distributions of spin components. Attempting to assess the accuracy of population models for these subdominant parameters poses further challenges.

Posterior predictive checks (PPCs) [7], a common test of model accuracy, evaluate the performance of predictive models by checking the consistency between data predicted by the model and current observations. Although widely used in gravitational wave population analyses, PPCs demonstrate significant limitations [8].

II. OBJECTIVES

The objective of this project is to determine whether partial predictive checks and/or split predictive checks are more discerning tools for model criticism than PPCs. Our approach is described further in Section III.

III. APPROACH

To ascertain the efficacy of these methods, we obtain different spin component distributions from a simulated astrophysical population of binary black hole systems with identical χ_{eff} distributions. We attempt to recover the known component distributions using predictive models of varying accuracy and then determine whether partial predictive checks and split predictive checks properly reflect the performance of the models.

A. Spin Parameterization

The effective aligned spin is defined as

$$\chi_{\rm eff} = \frac{\chi_1 \cos\theta_1 + q\chi_2 \cos\theta_2}{1+q}, \tag{1}$$

and the effective precessing parameter as

$$\chi_{\rm p} = \max\left[\ \chi_1 \sin\theta_1\ ,\ \left(\frac{3+4q}{4+3q}\right) q\chi_2 \sin\theta_2\ \right], \tag{2}$$

where $q = \frac{m_2}{m_1}$, χ_1 and χ_2 are the dimensionless component spins, and θ_1 and θ_2 are the angles between the component spins and the orbital angular momentum.

B. Posterior Predictive Checks

In posterior predictive checking, the distribution for a chosen diagnostic statistic is assumed under the null model $f(x|\mathbf{P})$. The likelihood of this diagnostic statistic, along with the prior for the parameters ($\mathbf{P}$), is used to integrate out the parameters from its posterior,

$$m_{\rm post}(t|x_{\rm obs}) = \int f(t|\mathbf{P})\pi(\mathbf{P}|x_{\rm obs})d\mathbf{P}, \tag{3}$$

resulting in the distribution $m_{\rm post}$ for the diagnostic statistic under the null model. Then, the p-value can be calculated according to

$$p = {\rm Pr}^{m_{\rm post}(t|x_{\rm obs})}(t(x) \geq t(x_{\rm obs})). \tag{4}$$

By this definition, smaller p-values signify greater conflict between the model and the observed data.

PPCs fall short if the observed data has high uncertainties. In this scenario, sampling the proposed posterior of a poor model with the draws weighted according to high uncertainties will produce a nearly identical population to the observed data, making posterior predictive checking unhelpful when dealing with uninformative data.

During the summer, we plan to address these issues by implementing two alternatives to posterior predictive checks: partial posterior predictive checks [7] and split predictive checks [9].

C. Partial Posterior Predictive Checks

Training the improper prior into a distribution that integrates to 1 and evaluating measures of surprise with the same data can lead to a non-representative p-value. Partial PPCs address this shortcoming of posterior predictive checking by avoiding the double-use of data. The partial PPC method does use the statistic data $t_{\rm obs}$ to compute measures of surprise, but only uses information not present in $t_{\rm obs}$ when training the prior. The conditional distribution $f(x_{\rm obs}|t_{\rm obs}, \mathbf{P})$ is used as the likelihood to determine the posterior distribution $\pi_{\rm ppp}$ of parameters $\mathbf{P}$,

$$\pi_{\rm ppp}(\mathbf{P}|x_{\rm obs}\backslash t_{\rm obs}) \propto f(x_{\rm obs}|t_{\rm obs}, \mathbf{P})\pi(\mathbf{P}) \tag{5}$$

$$\propto \frac{f(x_{\rm obs}|\mathbf{P})\pi(\mathbf{P})}{f(t_{\rm obs}|\mathbf{P})}, \tag{6}$$

which is then used as a prior to determine posterior of t. With the contribution of $t_{\rm obs}$ already eliminated, $\mathbf{P}$ is integrated out of the posterior:

$$m_{\rm ppp}(t|x_{\rm obs}\backslash t_{\rm obs}) = \int f(t|\mathbf{P})\pi(\mathbf{P}|x_{\rm obs}\backslash t_{\rm obs})d\mathbf{P} \tag{7}$$

The new p-value then takes the form

$$p = {\rm Pr}^{m_{\rm ppp}(t|x_{\rm obs}\backslash t_{\rm obs})}(t(\mathbf{x}) \geq t(\mathbf{x}_{\rm obs})). \tag{8}$$

The use of $f(x_{\rm obs}|t_{\rm obs}, \mathbf{P})$ to define the likelihood instead of $f(x_{\rm obs}|\mathbf{P})$ evades the double-use of data that occurs in PPCs when training the prior into a proper distribution and then evaluating the p-value [10, 11].

D. Split Predictive Checks

The split predictive check (SPC) similarly aims to avoid repeated use of data. Rather than conditioning the prior for $\mathbf{P}$ on the influence of $t_{\rm obs}$ (as in partial PPCs), however, split predictive checks partition the data into two disjoint subsets from the start. With a single split of data $x_{\rm obs} = x_a + x_b$, the method uses different subsets when training the posterior and when determining the p-value. The posterior distribution for x under the null model

$$m_{\rm SPC}(x_{\rm pred}|x_a) = \int f(x_{\rm pred}|\mathbf{P})\pi(\mathbf{P}|x_a)d\mathbf{P} \tag{9}$$

integrates out the parameters and is used to define a new p-value

$$p = {\rm Pr}^{m_{\rm SPC}(x_{\rm pred}|x_a)}(t(x_b) \geq t(x_{\rm pred})). \tag{10}$$

The divided split predictive check (divided SPC) extends this method. Data is divided into N equal subsets, and the single SPC p-value is calculated for each individual subset. The divided SPC p-value is defined as the p-value obtained by performing the Kolmogorov–Smirnov test for uniformity on the collection of single SPC p-values [12].

IV. TIMELINE

- Weeks 1-2: Attend gravitational wave workshops. Register an account for the cluster, download the data, and set up a Github. Begin to replicate Figure 4 from Miller *et al.* [8] to gain familiarity with the posterior sampling.
- Week 3: Finish reproducing Figure 4 from Miller *et al.* [8].
- Week 4: Design and write the Python algorithms for the alternative posterior predictive checks.
- Weeks 5-6: Implement the code for partial predictive posterior checking on the simulated binary black hole component distributions.
- Weeks 7-8: Implement the code for the split predictive posterior checking on the simulated binary black hole component distributions.
- Weeks 9-10: Finish remaining aspects of the implementation and apply the algorithms to real data.

[1] R. Abbott, **882**, L24 (2019).

[2] R. Abbott, The Astrophysical Journal Letters **913**, 10.3847/2041-8213/abe949 ().

[3] R. Abbott, The population of merging compact binaries inferred using gravitational waves through gwtc-3 ().

[4] J. Roulet and T. Venumadhav, Inferring binary properties from gravitational wave signals (2024), arXiv:2402.11439 [stat.ME].

[5] T. A. Callister, S. J. Miller, K. Chatziioannou, and W. M. Farr, The Astrophysical Journal Letters **937**, L13 (2022).

[6] M. Zevin, S. S. Bavera, C. P. L. Berry, V. Kalogera, T. Fragos, P. Marchant, C. L. Rodriguez, F. Antonini, D. E. Holz, and C. Pankow, The Astrophysical Journal **910**, 152 (2021).

[7] M. J. Bayarri and M. E. Castellanos, Statistical Science **22**, 322 (2007).

[8] S. J. Miller, Z. Ko, T. A. Callister, and K. Chatziioannou, Gravitational waves carry information beyond effective spin parameters but it is hard to extract (2024), arXiv:2401.05613 [gr-qc].

[9] J. Li and J. H. Huggins, Calibrated model criticism using split predictive checks (2022), arXiv:2203.15897 [stat.ME].

[10] M. J. Bayarri and J. O. Berger, Journal of the American Statistical Association **95**, 1127 (2000).

[11] J. M. Robins, A. van der Vaart, and V. Ventura, Journal of the American Statistical Association **95**, 1143 (2000).

[12] F. J. Massey Jr., Journal of the American Statistical Association **46**, 68 (1951).