

Inferring Binary Pulsar Population Statistics Using the NANOGrav 11 Year Dataset

Steven Stetzler
Advisor: Kevin Stovall

September 15, 2017

1 Introduction

The NANOGrav 11 Year Dataset offers a unique opportunity to explore the statistics of binary pulsar populations. Containing the timing solutions for 31 millisecond pulsars in binary orbits with a white dwarf companion, this dataset provides us with access to large number of binary pulsars which have been observed with a unique level of consistency and frequency. All pulsars were observed the same way, at the same frequency bands, using either the Green Bank Telescope (GBT) or the Arecibo Observatory (AO), and all timing solutions are constructed using observations over a long period of time, using similar and consistent processes for constructing the timing solution. This self-consistency in the observation and timing of the pulsars along with the fact that binary pulsars in the dataset were chosen based on their ability to be timed rather than their nature as binary pulsars allows us to trust the validity of the timing solutions provided and remove from consideration a selection bias in the way we chose the binary pulsars to study.

What follows is a brief introduction to pulsars and the process of pulsar timing, a review of our approach to model fitting in a Bayesian context using Markov Chain Monte Carlo, and the leveraging of the resulting distributions of model parameters to extract relevant population statistics. Specifically, we examine whether we can reject the commonly stated fact that the inclinations of binary pulsar orbits are uniformly distributed over the cosine of the inclination. We present these results and propose other population models that the tools we have developed can be used to tackle.

1.1 Pulsars

A pulsar, shown in Fig. 1 (Left), is a spinning neutron star that emits beamed radio emission. The neutron star is a possible leftover remnant from a supernova. When a star collapses, its angular momentum is conserved throughout the collapse, frequently producing a rapidly spinning remnant. The magnetic field of the star also strengthens during the collapse, leaving the remnant highly magnetized. The intense magnetic fields surrounding the neutron star cause electrons around the neutron star to accelerate in their motion around the star, producing synchrotron emission resulting in a beam of radio emission protruding from the poles of the magnetic field of the neutron star. This beam of emission need not align with the rotation axis of the neutron star, and it is extremely unlikely that it does. The off-axis alignment of the beamed radio emission causes the emission to

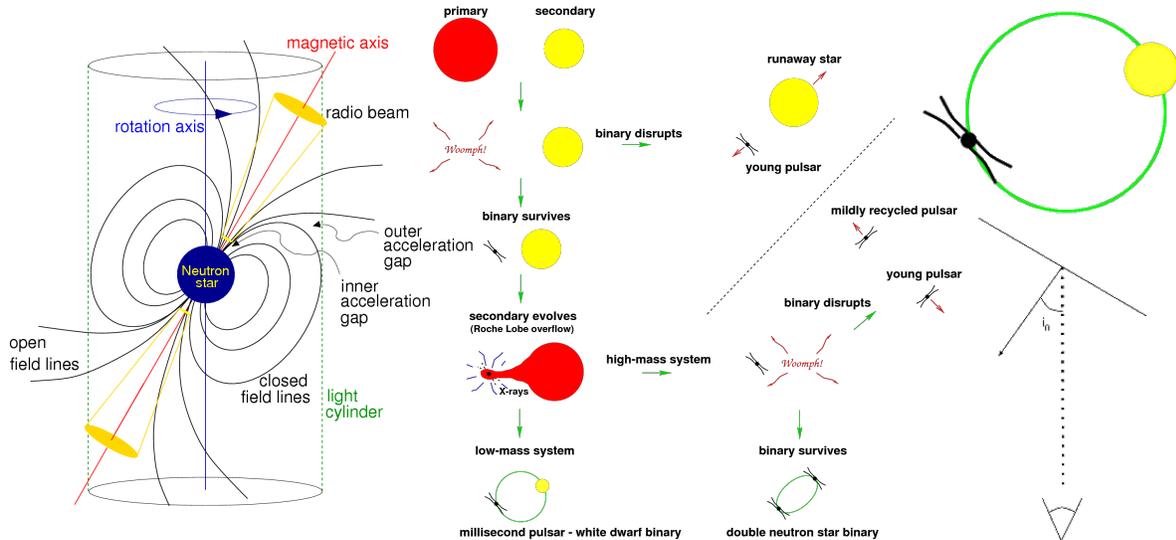


Figure 1: **Left:** A diagram of a pulsar. Note the beamed radio emission which is not aligned with the rotation axis of the neutron star. **Middle:** The current theory of the evolution of binary pulsar systems. **Right:** A millisecond pulsar along with its orbital inclination (i_0) relative to the line of sight of the observer. The filled line represents the orbit, the dashed line represents the line-of-sight of the observer, and the arrow represents the orbit's angular momentum vector.

sweep through the sky. If an object (e.g. the Earth) lies along the line of sight of the sweeping path of the radio beam, the object will observe radio emission washing over it in pulses - the radio emission will appear to turn on for a brief period once per rotation of the neutron star, with a lack of emission during the rest of the rotation as the object lies outside of the path of the radio beam.

Binary systems with millisecond pulsars, shown in Fig. 1 (Middle), are especially interesting to us. Binary pulsar systems are formed from stellar binaries, where one or both of the stars in the binary undergo supernova, with at least one of the remnants being a pulsar. Millisecond pulsars, according to current pulsar evolutionary theories, are formed via an accretion process between a pulsar and its companion. First, a single star in a stellar binary goes supernova, leaving behind a pulsar. When the second star enters its red giant phase, and if the outer shell of the red giant enters into the Roche Lobe of its pulsar companion, the pulsar will start to accrete it. This process causes the pulsar to pull in material from the companion, and since the two stars are in an orbit, there is a transfer of orbital angular momentum from the binary pulsar orbit to rotational angular momentum of the pulsar, effectively spinning up the pulsar, shifting its rotational period from ≈ 1 sec to ≈ 1 ms, hence the name millisecond pulsar. The typical final state of this process is a millisecond pulsar in a binary pulsar orbit with a white dwarf, however there are other possible paths in the binary evolution.

There is a special parameter associated with the resulting orbits which we would like to measure - the inclination, as shown in Fig. 1 (Right). The inclination is defined as the angle between the angular momentum vector of the binary pulsar orbit and the line-of-sight of the observer.

1.2 Pulsar Timing

Everything we know about pulsars comes from a process called pulsar timing. The main goal of pulsar timing is to create a model that predicts when the next pulse of radio emission will arrive

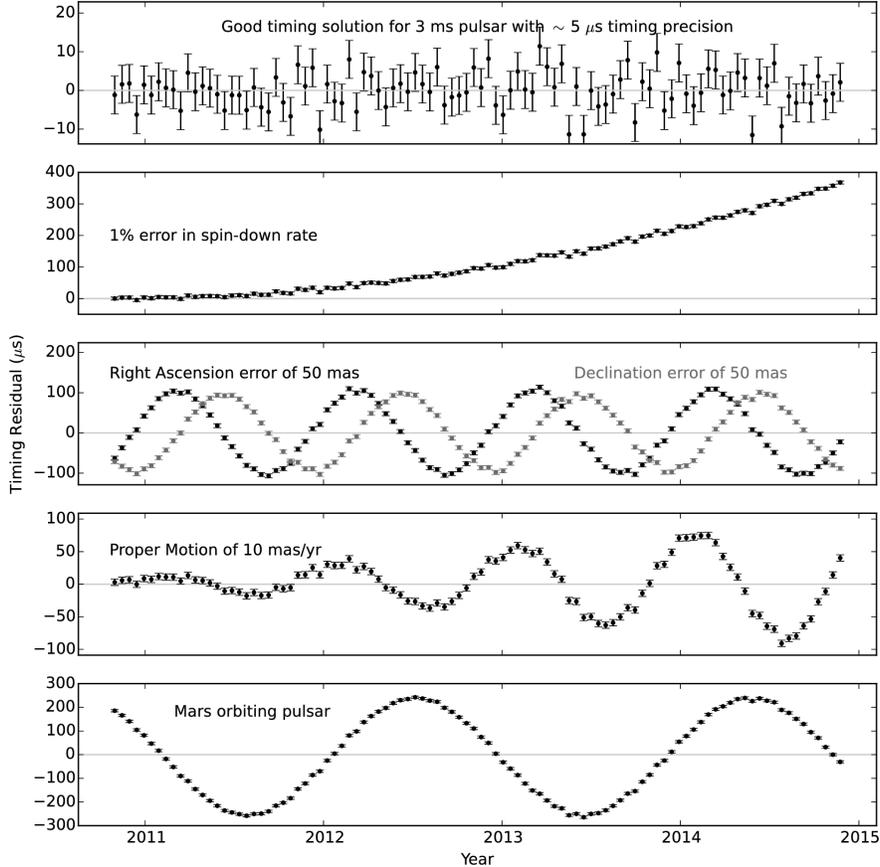


Figure 2: Examples of timing models for a pulsar when various parameters are and are not taken account of. **Top:** An example of a good timing solution, when all physical parameters that can be seen in our data have been accounted for. **Bottom:** A timing solution when a planet the mass of mars is orbiting the pulsar and this isn't accounting for in our model. A sinusoid with period the period of the planet's orbit and amplitude related to the mass of the planet and the pulsar appears in the data.

at Earth. Each of the parameters in this model is physically motivated, meaning our model will account for physical attributes of the pulsar and its binary system. As an example of what this model might look like, the simplest and most naive model would only account for the rotation rate of the pulsar - we expect that the next pulse of radiation will arrive at Earth after 1 rotation period of the pulsar. This is a fine model, however observing a pulsar for a duration even as short as a few hours will reveal that this model is incomplete: if we don't perfectly account for the rate of change of the pulsar's spin, the times of arrival of the pulses will drift away from our model quadratically. This example, as well as examples of other physical parameters and their effects on our timing model are shown and discussed in Fig. 2.

The importance of pulsar timing comes from the fact that it allows us to make a direct measurement of the reality of the pulsar and its surroundings. As shown in Fig. 2 (Bottom), we can model for the effects of binary pulsar orbits on our timing solution. We can encode the inclination of the binary pulsar orbit into our model and thus make a direct measurement of the parameters of the binary orbit (e.g. the inclination).

1.3 The Shapiro Delay

There is a specific physical process we can model for called the Shapiro delay. The Shapiro delay is present in all binary pulsar orbits, and in the case of a pulsar, it is the delay caused to the beamed radio emission as it travels through the gravitational field of the pulsar’s companion star. The Shapiro delay will take on the shape shown in Fig. 3. In the TEMPO2 pulsar timing package, the Shapiro delay can be modeled for using the parameters (with the TEMPO2 names in parentheses) m_2 (M2) and $sini$ (SINI), the ”traditional” parameterization of the delay. The m_2 parameter accounts for the mass of the pulsar’s companion. The $sini$ parameter accounts for the binary pulsar orbit inclination.

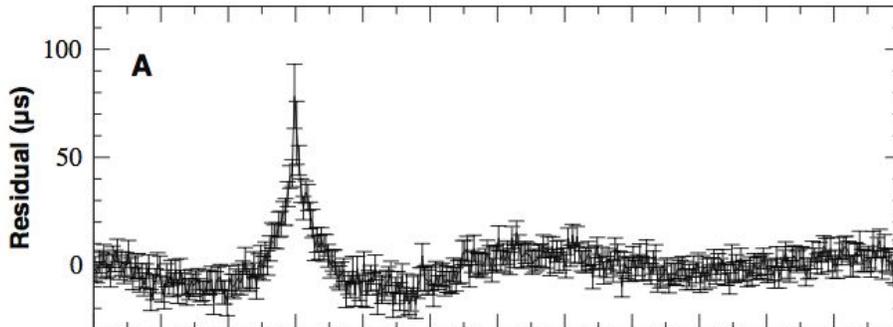


Figure 3: The expected shape that the Shapiro delay would take on in our timing model when unaccounted for.

There is an alternative parameterization, called the ”orthometric” parameterization which uses as parameters the third and fourth harmonics, h_3 (H3) and h_4 (H4), or their ratio $\zeta = \frac{h_4}{h_3}$ (STIG), of the Fourier decomposition of the Shapiro delay delay signal. Here,

$$\zeta = \sqrt{\frac{1 - \cos i}{1 + \cos i}}$$

encodes the information about the inclination of the binary pulsar orbit and

$$h_3 = T_{\odot} m_2 \zeta^3,$$

with T_{\odot} the solar mass represented in units of time, encodes the information about the mass of the pulsar’s companion. Finally,

$$h_4 = h_3 \zeta$$

encodes a mixture of both of these.

This orthometric parameterization is preferred, i.e. the two parameters used are more statistically significant and less covariant in our model, for certain types of binary pulsar orbits. Here, statistically significant means that the inclusion of the parameter in our model producing a statistically significant change in the χ^2 value of the resulting fit of the model to the timing data. The covariance of two parameters means how similar the effect of the two parameters is when they are included in the timing model. For example, from Fig. 2, a misalignment of the position of the pulsar in Right Ascension (RA) and Declination (DEC) produce similar signals in our data, thus when they are both included as model parameters while we are early in the observation process

for a pulsar, the two parameters will be highly covariant. When modeling the Shapiro delay using the orthometric parameterization, our two parameters of interest are h_3 and h_4 when we have low-inclination and low-eccentricity binary pulsar orbits. The combination h_3 and ζ is used when we have high-inclination orbits of all eccentricities.

It is through a measurement of the Shapiro delay that we can make direct measurements of the inclinations of the binary pulsar orbits we are interested in. When using the traditional parameterization, we are able to make a direct measurement of the cosine of the inclination using the *sini* parameter and the transformation

$$\cos(i) = \sqrt{1 - \sin^2(i)}.$$

When using the orthometric parameterization, our model parameters can be transformed to a corresponding *cosi* value as well.

1.4 Bayesian Inference with Markov Chain Monte Carlo

The traditional approach to model fitting is to use a least-squares method. In this method, we analytically find the parameter values which minimize the square distance between our data and the model. This method is usually preferred because there is usually a single, easily found analytic solution to our parameter values. An alternative approach to model fitting which we use is called Bayesian Inference with Markov Chain Monte Carlo (MCMC). The benefits of this method is that the end result of the model fitting isn't a single value for each parameter, but rather an entire probability distribution for what our parameter values could be.

The goal of model fitting in the traditional sense is to find the model such that $P(\text{data}|\text{model})$ (read as the probability that our data would come from the given model) is maximized. In the least-squares approach, we construct a model that does exactly this - maximizes the probability by minimizing the square distance between the model and the data. The Bayesian approach to model fitting, and to probability in general, flips this approach around. Rather, we become interested in $P(\text{model}|\text{data})$ (read as the probability that our model is correct given the data we have). This is expressed using Bayes' Theorem:

$$P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}.$$

To any person of an empirical mindset, this is the only logical approach. However, computing this probability is often intractable or impossible due to the pesky $P(\text{data})$ in the denominator. The Markov Chain Monte Carlo algorithm gives us an iterative way to draw samples from $P(\text{data}|\text{model})$ and $P(\text{model})$, the numerator alone, and at the end construct the probability distribution $P(\text{model}|\text{data})$, ignoring the $P(\text{data})$ factor entirely.

To understand the algorithm for constructing the parameter probability distribution, first consider a vector of parameters, \vec{p} . To make it more clear what this represents, let's consider a model which only accounts for the period (T) and position (RA and DEC) of the pulsar: $\vec{p} = \langle T, RA, DEC \rangle$. In the Bayesian Inference algorithm, we assume some *prior* information about what values our parameters could possibly be. For example, we might make the assumption that $T = 1 \pm 0.5$ (ms), since we are working with millisecond pulsars. However, our priors take the form of probability distributions themselves, so we might make the assumption that T is normally distributed with a mean of 1 (ms) and a standard deviation of 0.5 (ms). We assign prior distributions for each of the parameters. The MCMC algorithm proceeds as follows:

- Draw $\vec{p}_0 = \langle T_0, RA_0, DEC_0, \dots \rangle$ from our prior distributions
- for n iterations, at the i_{th} iteration:
 - Choose $\vec{p}_i = \langle T_i, RA_i, DEC_i, \dots \rangle$ from our prior distributions
 - If \vec{p}_i provides a **better** fit of our model to the data, **always** accept \vec{p}_i as the current best solution.
 - If \vec{p}_i provides a **worse** fit of our model to the data, **sometimes** accept \vec{p}_i as the current best solution and the rest of the time, keep \vec{p}_{i-1} as the current best solution

The end result of this algorithm is a list of parameter vectors \vec{p}_i with $i \in [0, n]$, which is called a Markov Chain. The beauty and relevance of this algorithm is that this list of parameter vectors has the property that they were effectively drawn from the *posterior* probability distribution $P(model|data)$, the quantity we were originally after (I know, it’s magic). The magic comes from the fact that we *sometimes* reject worse parameter proposals. That “sometimes” is related directly to how well the current model fits our data compared to the previous proposal. This causes our simulation to spend more time in areas of high probability and less time in areas of low probability, in a way that exactly models the posterior distribution we were looking for.

We can then *marginalize* our resulting probability distribution to find the probability distribution of any single parameter. If we want to find the probability distribution of the cosine of the inclination of a binary pulsar orbit, for example, we marginalize our probability distribution by integrating over all of our other parameters:

$$P(cosi|data) = \int P(\vec{p}|data) dp_i \quad p_i \neq cosi.$$

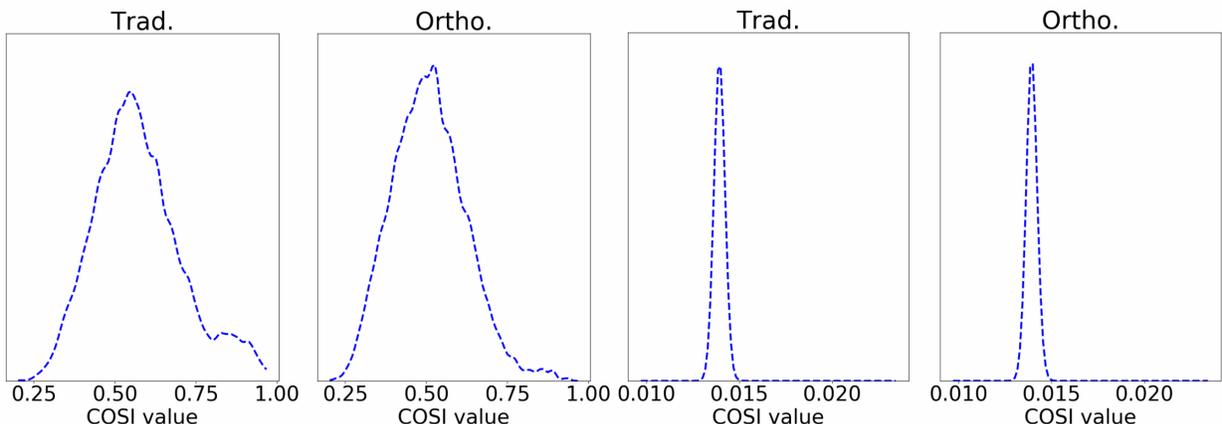


Figure 4: The marginalized distributions for $cosi$ for J1455+3330 (left) and J1614-2230 (right) using both the traditional (Trad.) and orthometric (Ortho.) parameterizations of the Shapiro delay.

Histogramming our $cosi$ values over the course of the simulation gives us in effect this marginalized distribution. An example of the marginalized distribution, a smoothed histogram, of $cosi$ values for the pulsar J1455+3330 and J1614-2230 is shown in Fig. 4. Observing these smooth histograms reveals that there is general agreement in the $cosi$ distributions between the traditional and orthometric parameterizations of the Shapiro delay. Comparing J1455+3330 and J1614-2230, it is clear that the value of $cosi$ is much better known for J1614-2230, which has a probability distribution that is very narrowly peaked around a value of 0.014, than it is for J1455+3330, which

has a probability distribution that runs from 0.25 to 1.00 in $\cos i$. Generating the probability distributions in this way lets us see the sometimes large and sometimes asymmetric uncertainties in the values of the $\cos i$ parameter value in our model.

2 The Question: Are binary pulsar orbit inclinations uniformly distributed over the cosine of the inclination?

The ultimate question we would like to answer is: Are binary pulsar orbit inclinations uniformly distributed over the cosine of the inclination? We will provide a short proof for why this is believed to be true, as well as motivate our pursuit for the answer.

The true problem statement is: A collection of randomly oriented isometric binary pulsar orbits is uniformly distributed over the cosine of the inclination. The first key word to understand is *randomly oriented*, which tells us that there is no preferred orbit, which is equivalent to saying that there is no preference in the orientation of the orbit's angular momentum vector. It is convenient to think of how the angular momentum vector points, since it is within that context that we defined inclination. The second important keyword *isometric* though tells us that this *are* preferred orbital inclinations. Isometric means that the only information we receive about the orbit comes from its projection onto the plane of the sky. This means that any orbit rotated through the line-of-sight (or through the plane of the sky) will by the isometric assumption be an equivalent orbit, as described in Fig. 5.

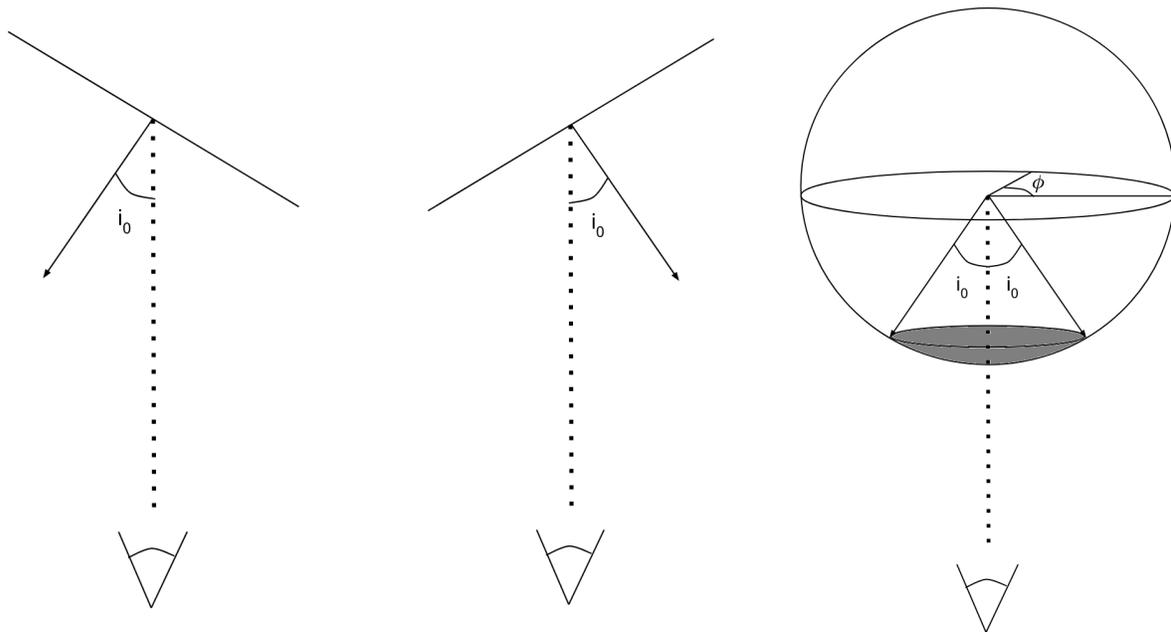


Figure 5: **Left:** The orbital inclination as defined. **Middle:** An orbit with the equivalent orbital inclination given the isometric condition. **Right:** As we consider all possible directions of the orbital angular momentum, and all possible rotations through the plane of the sky, a sphere is traced out. A particular inclination as it is rotated through $\phi = 0$ to $\phi = 2\pi$ traces out a circle on this sphere.

Thinking of this problem in terms of the sphere shown in Fig. 5 makes it straightforward to compute the probability distribution of the inclination of the randomly oriented binary orbits. The

probability to find an orbit with inclination less than some i_0 is encoded in the gray shaded region of Fig. 5 (Right). We can compute this probability by finding the surface area of the shaded area:

$$P(i < i_0) = \frac{1}{Z} \int_0^{2\pi} \int_0^{i_0} \sin(i) di d\phi = \frac{1}{Z} 2\pi [-\cos i(i)]_0^{i_0} = \frac{2\pi}{Z} (1 - \cos(i_0)),$$

and computing Z , our normalization constant, as the total surface area of the lower hemisphere:

$$Z = P(i < \frac{\pi}{2}) = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \sin(i) di d\phi = 2\pi [-\cos i(i)]_0^{\frac{\pi}{2}} = 2\pi,$$

we have that

$$P(i < i_0) = 1 - \cos(i_0).$$

Which is reminiscent of the definition of a uniform distribution. A probability distribution is uniform over x if $P(x < x_0) = 1 - x_0$. Therefore we have that our collection of randomly orientated isometric binary pulsar orbits will be uniformly distributed over the cosine of the inclination.

The motivation for testing whether this is true is two-fold: first, it is mathematically proven that this is true, making this problem well defined, and making it clear that at least one of our assumptions is wrong if our data show this fact is not true. We also can easily guess which of these assumptions is likely to be wrong: The Shapiro delay is stronger for highly inclined systems, meaning that if we find a lack of low inclination systems in our dataset, leading to a rejection of our proven fact, then a likely reason why is because higher inclined systems are more likely to be found given the limited observation time we have - stronger Shapiro delay signals are inherently preferred over weaker ones. Second, this is an easy and well defined first test for the general tools we would like to develop for testing other binary pulsar population models.

3 Results

In the following sections, we elaborate on several approaches to answering our core question: Are binary pulsar orbit inclinations uniformly distributed over the cosine of the inclination? We fit a timing model to each of the 31 binary pulsars in our data set using the MCMC algorithm described in Sec. 2, producing 31 probability distributions for the cosine of the inclination of each of those binary pulsar orbits. We take on three approaches to comparing these distributions to uniformity: summing together our collection of 31 $\cos i$ parameter distributions, performing statistical tests on representative values of each $\cos i$ distribution, and a new proposed method which leverages simulations to perform statistical tests on representative values without choosing *which* representative value to use.

3.1 Distribution Summing

If we had a collection of 31 individual measurements of $\cos i$, then a straightforward approach to test for uniformity in $\cos i$ is to histogram our data and try to see if it is relatively flat. We can use this same logic and apply it to our 31 probability distributions of $\cos i$ instead, plotting all 31 of them on the same axes, then summing them together. If the underlying probability distribution of the population of binary pulsars is uniform over $\cos i$, then we would expect that if we collect enough probability distributions, then when summed together we should arrive at a flat distribution. Figure 6 (Left) shows the $\cos i$ parameter distribution, derived using the MCMC algorithm outlined in Sec. 1.4, for each of the binary systems we have, and for both the traditional

and the orthometric parameterizations of the Shapiro delay. In Fig. 6 (Right), the *cosi* distributions are summed together and smoothed, to show the shape of the underlying distribution.

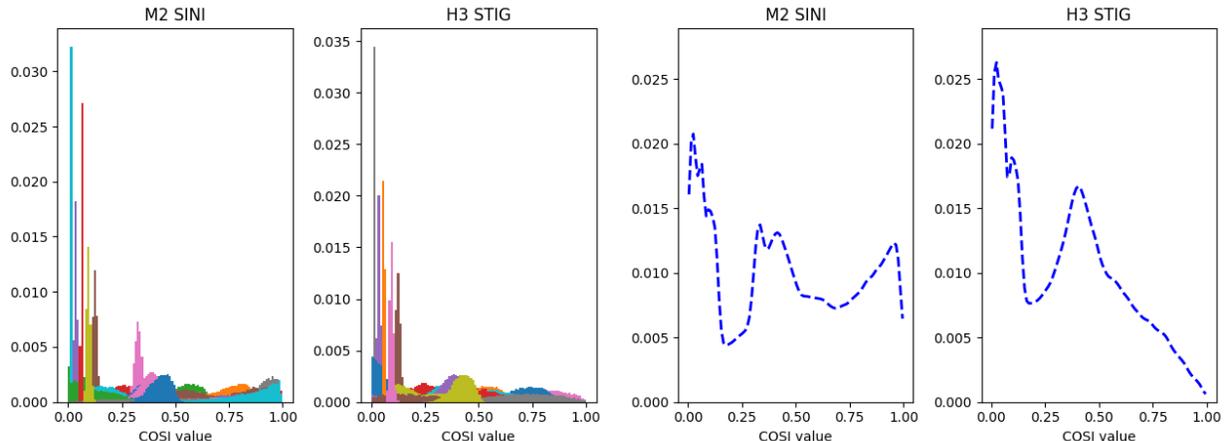


Figure 6: **Left:** The collection of 31 *cosi* parameter distributions derived from our 31 timing models of binary systems. We found distributions for each of the parameterizations of the Shapiro delay: the traditional (M2 SINI) and orthometric (H3 STIG) versions. Each color represents a single distribution. **Right:** The distributions shown on the left were summed together and smoothed to observe the underlying population distribution.

By a visual inspection alone, we can see that our distributions found using the traditional (M2 SINI) parameterization admit an underlying distribution in *cosi* that looks relatively flat. In particular, there is a clustering of probability for small, medium, and large *cosi* values. Looking at the summed distributions found using the orthometric (H3 STIG) parameterization, we can see a deviation from uniformity and from agreement with the traditional parameterization. Specifically, it appears that it is very unlikely for binary pulsar systems to have inclinations corresponding to large (> 0.80) values of *cosi*.

We can perform analysis beyond the visual, if desired. For example, one can perform a χ^2 analysis to find the summed distribution’s departure from uniformity. However, we did not follow this route of analysis since this method would ultimately be sensitive to our arbitrary choice of how we bin the data. It would be worthwhile in the future to revisit this method to confirm the results shown in the following sections. For example, a χ^2 analysis could be performed with a variety of binnings.

3.2 Statistical Tests with Representative Values - I

An alternative to performing a statistical analysis using the entirety of each of the probability distributions is to simply use a representative value from each distribution. Reducing each *cosi* probability distribution to a single value gives us 31 single values of *cosi* which we can use as our data to compare against uniformity. Of course reducing each probability distribution to a single value throws away an incredible wealth of information about our binary pulsar orbits, and we will introduce later on a way to use representative values without losing the information encoded in the probability distributions.

Our statistical test of choice for this case is the Anderson-Darling (AD) test. The Anderson-Darling test answers the question “does the data provided come from the distribution specified?” It

is a modification of the more familiar Kolmogorov-Smirnov test. In both of these tests, we measure the deviation between our data’s empirical cumulative distribution function and our specified model cumulative distribution function. An example showing how the Kolmogorov-Smirnov test (and the AD test) works is shown in Fig. 7. We chose the Anderson-Darling test because it tends to be a more sensitive test than the Kolmogorov-Smirnov test.

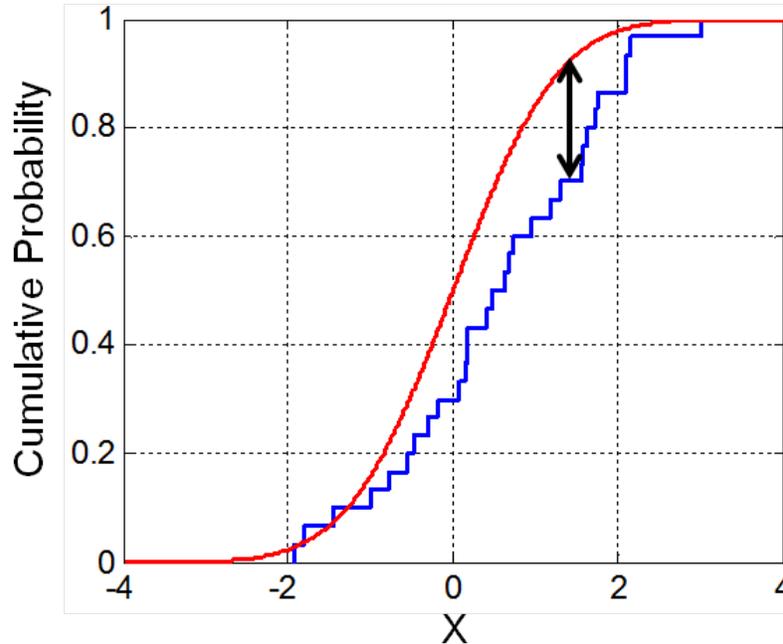


Figure 7: An example showing how the Kolmogorov-Smirnov test operates. The Kolmogorov-Smirnov test measures the maximum deviation between the data’s empirical cumulative distribution function (ECDF), shown in blue, and the provided model’s cumulative distribution function (CDF), shown in red. The Anderson-Darling test works similarly, however differences between the ECDF and the CDF are taken into account along their entire extent.

More specifically, the Anderson-Darling test answers the question “given that the data is drawn from the distribution specified, with what probability would one actually observe the data provided?” This “probability” of observation is known as a p-value. The way we use p-values in this context is as a judgment of how confident one can be in saying the data does not come from the distribution specified. For example, if we provide the AD test with some data and specify that it came from a uniform distribution and the AD test spits back a p-value of 0.05, what this means is that I would only expect to see the data I have 5% of the time given the data was drawn from a uniform distribution. Considering the rarity of this event, I might equivalently say that I reject that the data was drawn from a uniform distribution with 95% confidence.

The final question to answer then is “what representative values do we use”. There are many possible answers to this question, but the most obvious and useful values to represent a probability distribution are the mean, median, and mode. An example of where these three different values lie on one of our probability distributions is shown in Fig. 8. The mean is the average, or expectation, value of our distribution. The median is defined as the value below which 50% of the probability in our distribution lies. If we repeatedly drew values of $\cos i$ from a distribution, then we would expect that half of the values fall below the median and half lie above the median. The mode is the

most likely value of *cosi* that we can measure - it is the most likely value to show up in repeated measurements of *cosi*.

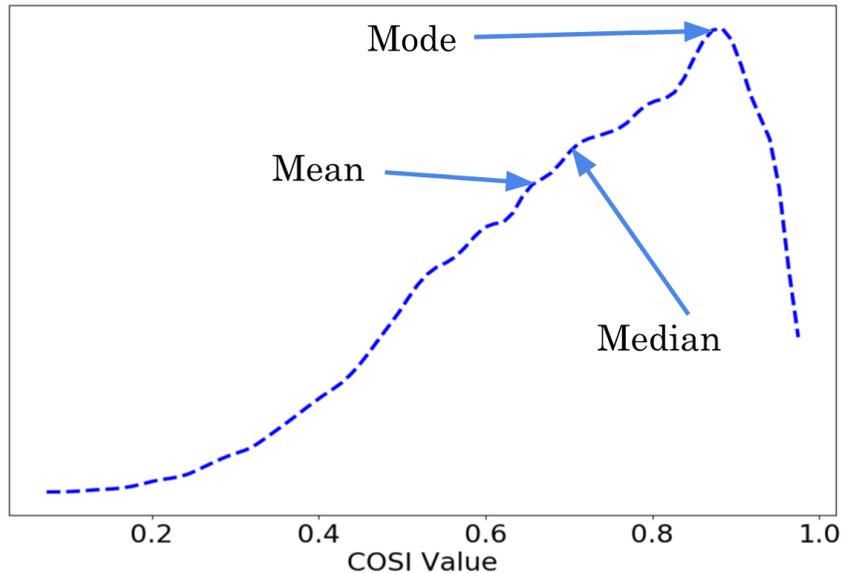


Figure 8: An example of the three different values we can use to represent a probability distribution.

We decided to see what results we would get from the Anderson-Darling test if we used all three of these reasonable representative values. So to review, we took the mean, median, and mode values from our 31 *cosi* probability distributions, resulting in 3 sets of 31 values of *cosi*. We then performed an Anderson-Darling test on each of these 3 sets of data, specifying a uniform distribution, and obtained a p-value which we can use to accept or reject our hypothesis of uniformity of the data. The resulting p-values are shown in Table 1 and a visualization of the cumulative distribution functions used in the test are shown in Fig. 9.

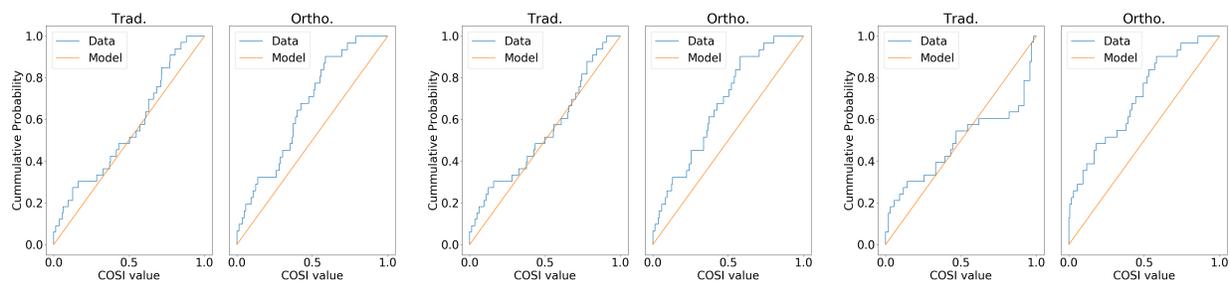


Figure 9: A visualization of the cumulative distributions of our collections of means (left), medians (middle), and modes (right), all shown in blue, compared to a uniform distribution, shown in orange. This allows us to visualize the results of the Anderson-Darling test. We performed tests for both the traditional (Trad.) and orthometric (Ortho.) versions of our *cosi* probability distributions.

Looking at the p-values in Table 1, using the mean or the median value, we see that we cannot reject the hypothesis that the data is uniformly distribution when using the traditional parameterization of the Shapiro delay, however we can for the orthometric parameterization. And we can reject the hypothesis of uniformity for both parameterizations when using the mode as the repre-

Table 1: P-Values from the Anderson-Darling Test

	Mean	Median	Mode
Trad.	0.42	0.62	0.003
Ortho.	0.010	0.007	0.0002

representative value. This may be enough to reject our hypothesis that binary pulsar orbit inclinations are uniformly distributed over the cosine of the inclination! However, the fact we can get vastly different results when using a different representative value is concerning. Each of the representative values we chose is a reasonable representation of a probability distribution. Since it is unclear which representation is preferred, and our choice of representation gives us conflicting results, then we cannot be sure about any choice of representation. This leads us to our final approach in a test for uniformity: performing statistical tests like these without imposing a certain representation for each distribution, allowing us to leverage the strengths of these tests along with the entirety of the information encoded in our *cosi* probability distributions.

3.3 Statistical Tests with Representative Values - II

We propose a new, simple algorithm that allows us to use the methods developed in the previous section for statistical tests with representative values without ever imposing which representative value to use. The algorithm is developed with the fundamentals of experimentation in mind: when we make a measurement of anything, we produce a value that does not represent reality. The truth is that we must make that measurement over and over again to uncover the distribution from which the measurement was drawn. This distribution is the true representation of our measurable reality and it accounts for uncertainties inherent to our measurement techniques and to the system we are trying to measure. What we would like to do is generate this distribution for the p-values from the Anderson-Darling test.

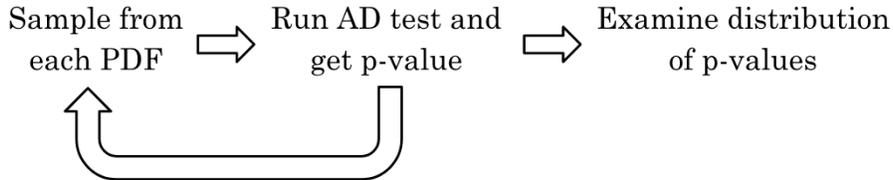


Figure 10: The proposed algorithm to test for uniformity in *cosi* while leveraging the entirety of each *cosi* probability distribution we have for each binary pulsar orbit.

In our algorithm, outlined in Fig. 10, a single experiment consists of making a measurement of *cosi* for each binary pulsar orbit, then running the AD test on the resulting 31 values of *cosi* that we have measured, recording the p-value produced. The way we make a measurement of *cosi* for a single binary pulsar orbit is by sampling from its probability distribution. Then we repeat this experiment many, many times, generating a long list of p-values. The beauty of this algorithm is that it allows us to construct a p-value from the AD test without ever imposing which representative value to use from each of the probability distributions. Instead, the choice is made naturally for us: the most likely value that we will draw in an experiment is the mode of the distribution, and the average value we draw is the mean of the distribution. Sampling in this way allows us to incorporate the entirety of the our probability distributions for *cosi*. The end result of

this algorithm is a collection of p-values, which we can then histogram to visually represent their underlying probability distribution. We can then analyze these probability distributions to draw a conclusion about the uniformity of our data in *cosi*. After running this algorithm, we produced the p-value distributions shown in Fig. 11.

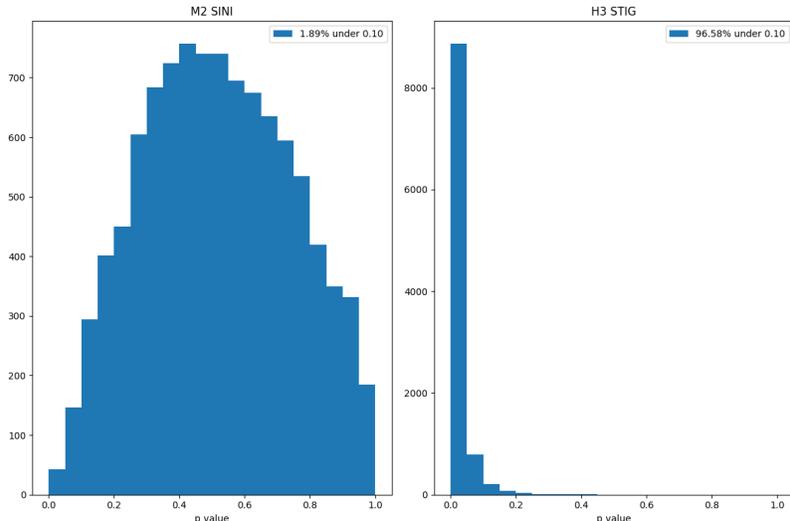


Figure 11: The p-value distribution generated using the algorithm described in Fig. 10 for both the traditional (M2 SINI) and orthometric (H3 STIG) parameterizations.

We examine the resulting p-value distribution by looking at its shape. If the hypothesis that binary pulsar orbit inclinations are drawn from a distribution uniform over the cosine of the inclination is true, then we would expect our p-value distribution to be uniform (flat). This is how p-values behave when the null hypothesis of a test is in fact true. If the null hypothesis is incorrect, then we would expect a bias in the p-value distribution toward small p-values. What do we see in the distributions in Fig. 11? Something a little different. The distribution found using the orthometric parameterization is biased toward small p-values, therefore we can reject the null hypothesis. However, for the traditional parameterization p-value distribution, there is neither a bias in p-values toward small p-values nor a flat spread of them. Unfortunately, this is indicative of a mistake in our methodology. Right now, it is unclear what the mistake is.

There was one nuance in our methodology that was not explained thus far: we derived *cosi* probability distributions for binary pulsar orbits which did not have a statistically significant measurement of the Shapiro delay. The reason we did this is that the elegance and versatility of the MCMC algorithm allows us to find the preferred values of *cosi* in a timing model anyway, even if the inclusion of the parameter in the model is not significant. If the Shapiro delay is completely undetected, then we expect that the MCMC will give us back the prior assumption we made about what the distribution could be. In this case, we used a uniform prior for *cosi*. None of the *cosi* distributions that we found through simulation however matched our prior, so we accepted these as valid distributions of the *cosi* parameter. The *cosi* distributions for these binary pulsar orbits showed vast disagreements between their traditional and orthometric versions, as shown in Fig. 12, with the traditional version frequently reporting that large values of *cosi* were the most probable for these binary pulsar orbits.

If we remove all of the binary pulsar orbits without a significant measurement of the Shapiro delay from consideration, we find a firm agreement between the p-value distributions corresponding

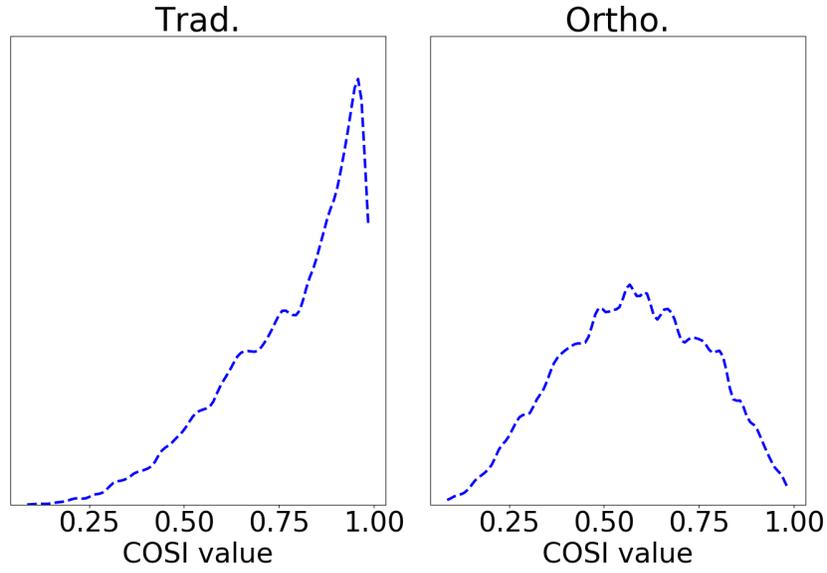


Figure 12: The *cosi* distribution for B1953+29 shows vast disagreement between its traditional (Trad.) and orthometric (Ortho.) versions.

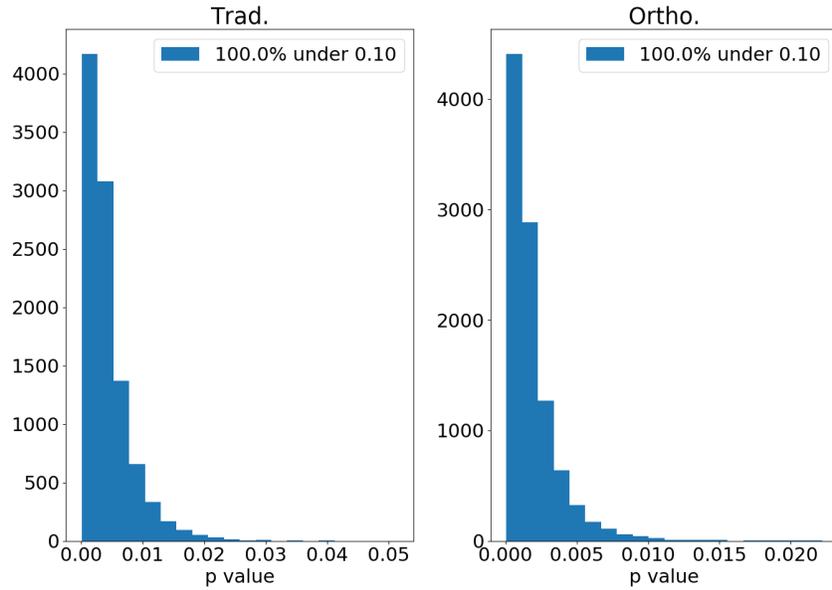


Figure 13: The p-value distribution generated using the algorithm described in Fig. 10 for both the traditional (M2 SINI) and orthometric (H3 STIG) parameterizations only for binary pulsar orbits with a statistically significant measurement of the Shapiro delay.

to the traditional and orthometric parameterizations, shown in Fig. 13. Both of these distributions show a bias towards small p-values, and thus a rejection of the null hypothesis. With this analysis we could conclude that binary pulsar orbit inclinations are not uniformly distributed over the cosine

of the inclination.

However, the reason we drew this conclusion is obvious: when we removed the orbits without a statistically significant measurement of the Shapiro delay, we removed all orbits with a weak Shapiro delay, likely due to the orbit having a small inclination, and thus a large cosine of the inclination. Of course our distribution of measurements of $\cos i$ is not uniform over $[0, 1]$, for we removed the possibility of measuring $\cos i$ close to 1. For now, our result is inconclusive until we can reconcile the disagreement among all of the distributions with weak measurements of the Shapiro delay, like the one shown in Fig. 12.

4 Conclusions and Future Work

We set out to answer the question: Are binary pulsar orbit inclinations uniformly distributed over the cosine of the inclination? We did not succeed in providing a conclusive result. However, we have developed several techniques, outlined in Sec. 3 to approach this problem. We identified that our methodology is flawed, however it is not immediately clear where the flaw lies. To resolve this, we are currently in discussion with various statisticians who can hopefully verify our results and our methodology. We also must look into verifying which, if any, of the $\cos i$ probability distributions for our binary pulsar orbits without a statistically significant measurement of the Shapiro delay are correct. One way to do this is to generate $\cos i$ probability distributions using another equivalent parameterization. Throughout, we used the H3 STIG orthometric parameterization, but we could have just as easily used the H3 H4 orthometric parameterization. Generation and analysis of distributions generated using such a parameterization is currently under way.

The techniques we developed can also be applied generally when testing any proposed model for the probability distribution of the population of one of the parameters in a timing model. For example, through measurement of the Shapiro delay, we can obtain a measurement of the pulsar's mass and also its companion mass. Applying the tools we've developed should be straightforward to test various proposed models for the population distribution of the mass of pulsars as well as white dwarfs. Finally, we can expand our data set beyond the NANOGrav 11-year dataset. For example, we can search the International Pulsar Timing Array dataset to find more well-timed binary pulsar systems to include in our dataset. Expanding our dataset will allow us to determine if the lack of low-inclination binary pulsar orbits is unique to our dataset, is a by-product of some bias in pulsar timing towards large Shapiro delay signals, or if it is truly representative of the binary pulsar population at large.