# Inferring Binary Pulsar Population Statistics Using the NANOGrav 11-year Data Set

Steven Stetzler[1,2], Kevin Stovall[2], Paul Demorest[2]
1: University of Virginia, Charlottesville, VA – sgs7cr@virginia.edu
2: National Radio Astronomy Observatory, Socorro, NM

## Abstract

The North American Nanohertz Observatory for Gravitational Waves (NANOGrav) monitors a set of millisecond pulsars to search for the effects of gravitational waves on pulsar signals. Containing the timing solutions for 31 millisecond pulsars in binary orbits with white dwarf companions, this data set provides us with access to a large number of binary pulsars which have been observed with a unique level of consistency over multi-year time scales. We use this data set to examine the binary pulsar orbital inclination angle distribution to see if the results are consistent with the standard assumption that the angles are uniformly distributed over the cosine of the inclination. While only applied to orbital inclinations, the statistical analysis tools we develop can be used to infer and test the population statistics of any of the parameters present in a pulsar's timing solution.

## Pulsar Timing

Pulsar timing is the process of constructing a model for when a pulsar's next pulse of radiation will arrive at earth. By only including physical parameters in the model, we can construct an extremely precise model of the pulsar's environment. Through modelling of the Shapiro Delay (Fig. 1), we can extract measurements of a binary pulsar's mass, its companion's mass, and its orbital inclination. For our work, we focus on the cosine of the inclination, called COSI.

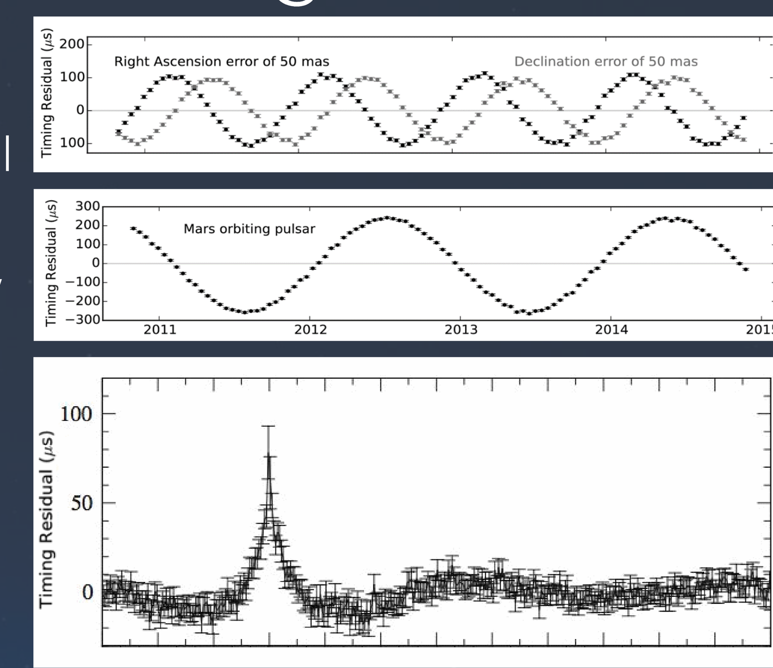


Fig. 1: Precise observation of a pulsar over a long period of time reveals systematic changes in the time of arrival of the pulsar's pulsed radiation due to various effects.

## Bayesian Model Fitting

We employ Markov chain Monte Carlo (MCMC) techniques and the PAL2 inference package to find a pulsar's timing solution for each of our pulsars in a Bayesian manner. This method lets us construct probability distributions for COSI (Fig. 3).

Posterior — Likelihood — Prior

$$\mathbf{Pr}(\mathrm{COSI}|\mathrm{data}) = \frac{\mathbf{Pr}(\mathrm{data}|\mathrm{COSI})\mathbf{Pr}(\mathrm{COSI})}{\mathbf{Pr}(\mathrm{data})}$$

Fig. 2: Sampling from the posterior is often intractable; however, MCMC techniques allow us to sample from the posterior, Pr(COSI | data), (Fig. 3) given only a likelihood, Pr(data | COSI), and a prior, Pr(COSI).

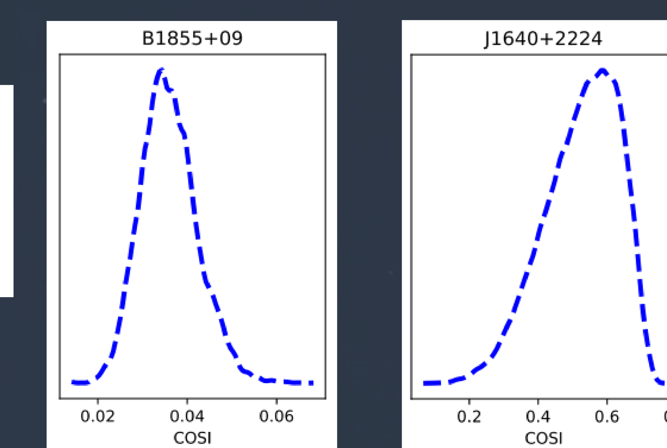


Fig. 3: The marginalized COSI probability distributions, Pr(COSI | data), for pulsars B1855+09 and J1640+2224.

## Are Binary Orbital Inclinations Uniform Over COSI?

### Distribution Summing

We can simply add all of the individual probability distributions for COSI together to get a snapshot of how the population distribution looks. We can then eyeball this distribution and make a rough comparison to uniform (Fig. 4). We notice from Fig. 4 a lack of low-inclination binary pulsar orbits, which makes sense considering the Shapiro Delay is more easily measured for highly inclined orbits.

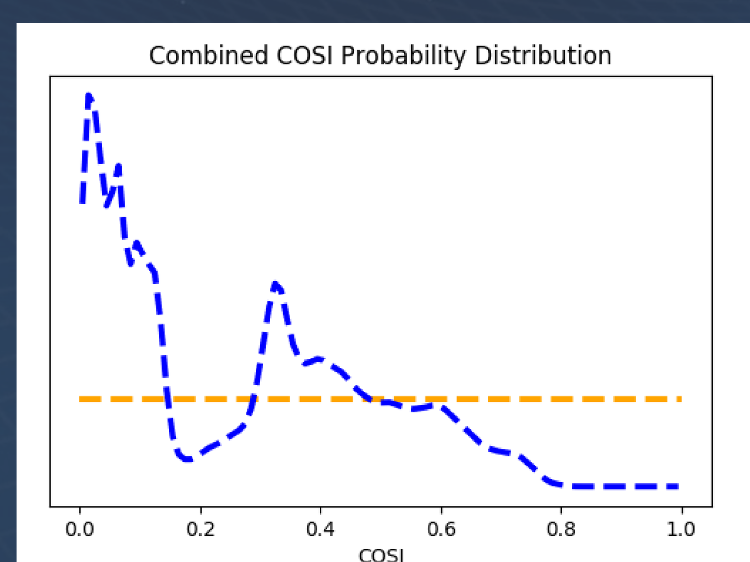


Fig. 4: The combination of all COSI PDFs into one distribution (blue) and its comparison to uniform (orange).

### Statistical Tests with Representative Values

Summary statistics provide us with a single number that represents our probability distribution. Three commonly used summary statistics are the mean, median, and mode (Fig. 5). We can then use the Anderson-Darling (AD) test to test whether our set of measurements of COSI are drawn from a uniform distribution. The AD test produces a p-value which simply indicates that, if the distribution of COSI values is uniform, we would only get samples at least this extreme p fraction of the time. The results shown in Table 1 indicate that COSI is not uniformly distributed.

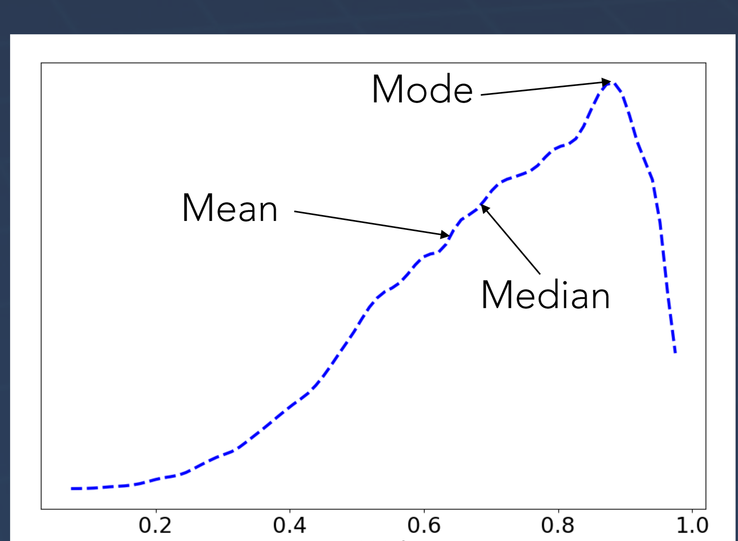


Fig. 5: An example of the mean, median, and mode of a COSI probability distribution.

Table 1: P-values from running the AD test on a set of summary statistics of different types.

| | |
|---|---|
| Mean | $4.1\times10^{-2}$ |
| Median | $4.4\times10^{-2}$ |
| Mode | $4.5\times10^{-5}$ |

### Monte Carlo Methods

Choosing a single representative value is a waste when we have entire COSI probability distributions available to us. We propose a new, simple algorithm that allows us to use traditional statistical tests with representative values without imposing which representative value to use. The algorithm, outlined in Fig. 6, produces a distribution of p-values which tells us how often we can expect a certain p-value given our data and its associated noise.


Fig. 6: Our algorithm to perform statistical tests with representative samples without imposing which representative sample to use.

Our algorithm mixes a Bayesian and frequentist approach to statistics in a dangerous way. This algorithm is not a robust way to account for noise in the COSI observations since the AD test compares our data to a uniform distribution, not a uniform + noise distribution. The p-value distribution is useful however in demonstrating whether noise has a significant effect on our conclusion about uniformity. As shown in Fig. 7, we see a bias toward low p-values, indicating that the noise on each pulsar's COSI measurement does not affect our conclusion that COSI is not uniformly distributed.

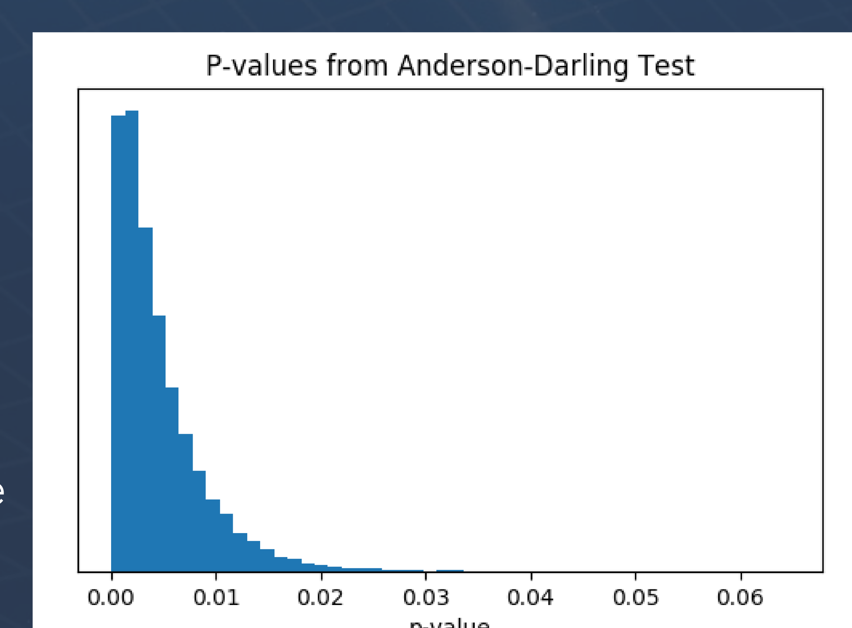


Fig. 7: The p-value distribution resulting from using our outlined algorithm (Fig. 6). The overwhelming bias towards p-values less than 0.05 tells us that we would always reject that COSI is uniformly distributed, regardless of which representative sample we use.

## Future Work

### Hierarchical Bayesian Modelling

Hierarchical Bayesian modelling allows us to incorporate extra parameters into our timing model to account for missing information about how a parameter is distributed. We can modify the equation in Fig. 2 to make a weaker assumption on how COSI is distributed, by imposing instead that COSI is distributed according to some distribution parameterized by some number of hyper parameters, each of which is distributed according to some hyper prior. This weaker assumption lets us directly model the distribution from which COSI is drawn, Pr(population | data).

Posterior — Hyper Prior

$$\mathbf{Pr}(\mathrm{population}|\mathrm{data}) = \frac{\mathbf{Pr}(\mathrm{data}|\mathrm{COSI})\mathbf{Pr}(\mathrm{COSI}|\mathrm{population})\mathbf{Pr}(\mathrm{population})}{\mathbf{Pr}(\mathrm{data})}$$

Fig. 8: Adding another layer to our model allows us to model how COSI is distributed.

### Expanding to Other Parameters

Orbital inclinations were a good first test of our methods; however, we can expand our analysis to other parameters of interest. For example, when using a Hierarchical Bayesian modelling scheme we can robustly infer the mass distribution of recycled pulsars while accurately accounting for all uncertainties in a pulsar timing data set.

## Acknowledgements

I would like to thank my advisor, Kevin Stovall, for his tremendous support and assistance in completing this project. I would also like to thank Josh Speagle for his current assistance in completing this project. This project was funded by the NSF through the REU program at NRAO. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. Background image courtesy the European Southern Observatory.