Frequentist GWB Searches with the PTA Optimal Statistic

Kyle Gersbach - VIPER PTA Summer School
July 10th 2024

The Plan

- A reminder of frequentist statistics
- The PTA optimal statistic (OS)
  - Dissecting the matrix products
- The OS as a parameter estimator
  - Pair covariance
- The OS as a detection statistic

A reminder of frequentist statistics

How frequently do events happen?

- The key is that it assumes many events happen!

What is a statistic? — Just a number!

- The importance of a statistic is its utility
- Make sure that it does what you want it to do

What makes a static ‘optimal’?

If we have an estimator for the GWB Amplitude

- $\hat{A}^2$

The variance on that estimator is (i.e. spread on the fit)

- $(\sigma_{\hat{A}2})^2 = < (\hat{A}^2)^2 > - < (\hat{A}^2) >^2$

An ‘optimal’ estimator is one which minimizes its own variance!

The PTA optimal statistic (OS)

Here is the actual OS derived for the full PTA

- a and b are pulsar indices

$$P_a = \left\langle \delta t_a \delta t_a^T \right\rangle$$

$$S_{ab} = \left\langle \delta t_a \delta t_b^T \right\rangle$$

$$\tilde{S}_{ab} = A_{\mathrm{gw}}^2 \Gamma_{ab} S_{ab}$$

Hellings and Downs!

$$\hat{A}^2 = \frac{\sum_{a<b} \delta t_a^T P_a^{-1} \tilde{S}_{ab} P_b^{-1} \delta t_b}{\sum_{a<b} \mathrm{tr}\left(\mathrm{P}_a^{-1} \tilde{\mathrm{S}}_{\mathrm{ab}} \mathrm{P}_b^{-1} \tilde{\mathrm{S}}_{\mathrm{ba}}\right)}$$

$$\sigma_{\hat{A}^2} = \left(\sum_{a<b} \mathrm{tr}\left(\mathrm{P}_a^{-1} \tilde{\mathrm{S}}_{\mathrm{ab}} \mathrm{P}_b^{-1} \tilde{\mathrm{S}}_{\mathrm{ba}}\right) \right)^{-1/2}$$

[Anholm et al 2009], [Chamberlin et al 2015]

The better OS

Correlated power in a pair

$$\rho_{ab} = \frac{\boldsymbol{X}_a^T\,\hat{\phi}\,\boldsymbol{X}_b}{\mathrm{tr}\left[\boldsymbol{Z}_b\,\hat{\phi}\,\boldsymbol{Z}_a\,\hat{\phi}\right]},$$

$$\sigma_{ab,0} = \mathrm{tr}\left[\boldsymbol{Z}_b\,\hat{\phi}\,\boldsymbol{Z}_a\,\hat{\phi}\right]^{-1/2}$$

Apply both pairwise values and rank-reduction strategies to simplify!

- $\hat{\phi}$ Is a unit-power-law spectrum

$$\hat{A}^2_{\mathrm{gw}} = \frac{\sum_{a<b} \rho_{ab}\Gamma_{ab}/\sigma^2_{ab,0}}{\sum_{a<b} \Gamma^2_{ab}/\sigma^2_{ab,0}}.$$

Sort of like unweighted fourier coefficients

$$\boldsymbol{X}_a = \boldsymbol{F}_a^T \boldsymbol{P}_a^{-1} \boldsymbol{\delta t}_a,$$
$$\boldsymbol{Z}_a = \boldsymbol{F}_a^T \boldsymbol{P}_a^{-1} \boldsymbol{F}_a.$$

$$\sigma_{\hat{A}^2,0} = \left(\sum_{a<b} \frac{\Gamma^2_{ab}}{\sigma^2_{ab,0}}\right)^{-1/2}.$$

[Pol, Taylor, Romano 2022], [Gersbach et al 2024]

Why's it better?

Working in fourier domain instead of time domain

- Matrices shrink! (N_toa x N_toa) → (2N_freq x 2N_freq)
- Inversions are more stable!

Pairwise values are handy with visualization!

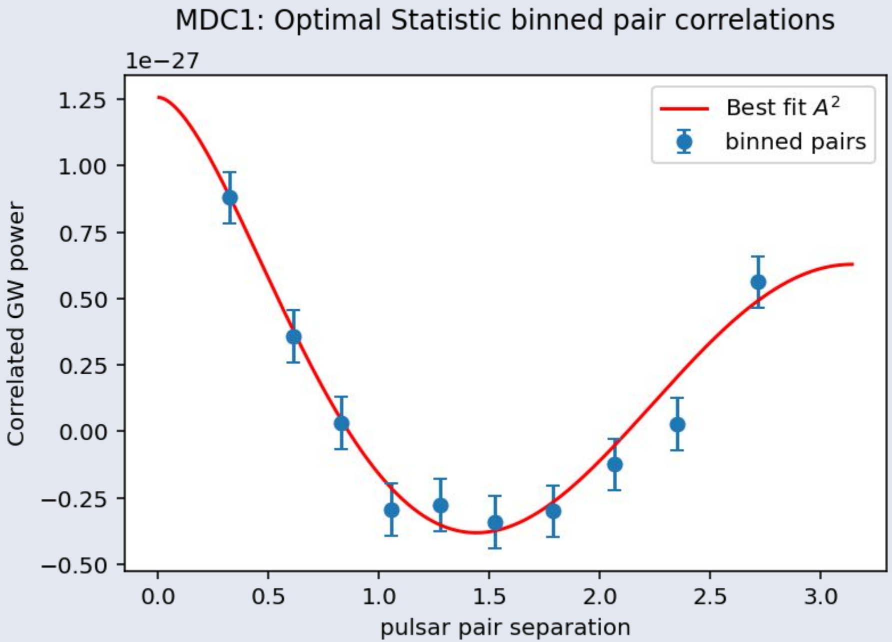
- In this form we linearly fit for an HD curve

What does the OS actually do?

Fitting data to a model

The ‘data’ is the pairwise correlated estimators

The ‘model’ is the amplitude modulated HD curve

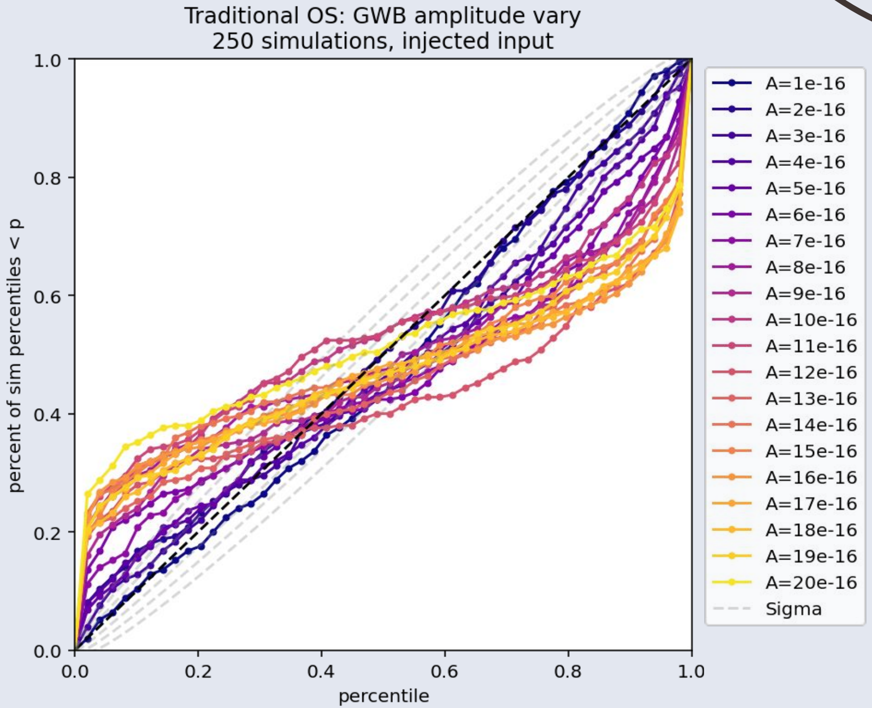


The OS as a parameter estimator

We use the OS for its estimated amplitude!

- There's a bit of a problem with higher signals

Strong signal regime breaks the OS!



PP plots with increasing amplitude
(more diagonal is better)

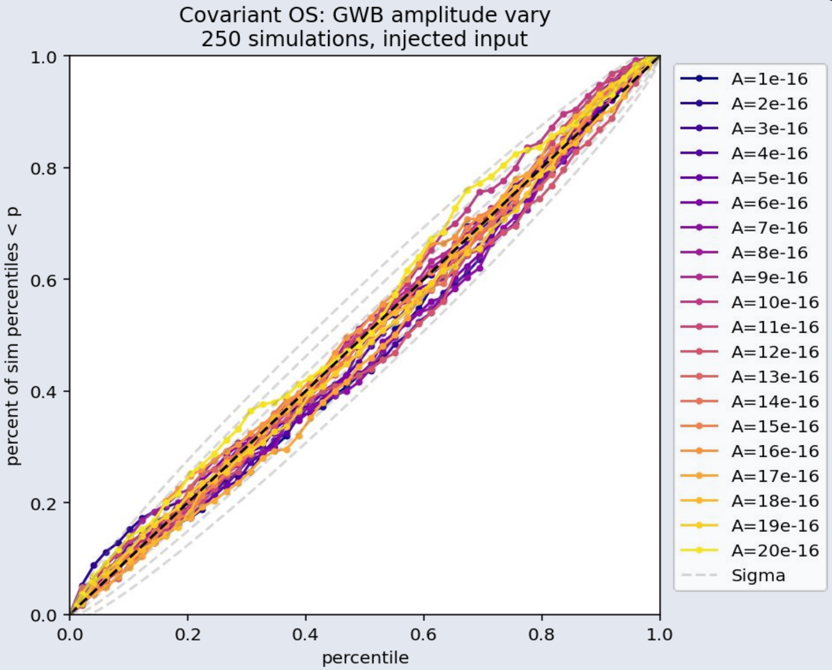
Fear not! For Pair Covariance is here!

We need to remove the pair independence assumption!

- Correlations between different pulsar pairs!

More computationally expensive but MUCH more accurate!

[Allen, Romano 2023]
This is a 71 page paper...



PP plots with increasing amplitude (more diagonal is better)

As a detection statistic

SNR is only approximately number of sigma away from 0

- Our distribution isn't gaussian

We need to calibrate our SNR by estimating the null distribution

- Recall yesterday's notebook?
- Destroy correlations with processes like phase shifts
  - [Taylor et al 2016]

$$\mathrm{SNR} = \frac{\hat{A}^2_{\mathrm{gw}}}{\sigma_{\hat{A}^2,0}}.$$

PP plots with increasing amplitude (more diagonal is better)

[Hazboun et al 2023]

What about those matrix products?

We need to calculate our X and Z matrices:

$$\boldsymbol{X}_a = \boldsymbol{F}_a^T \boldsymbol{P}_a^{-1} \boldsymbol{\delta t}_a,$$
$$\boldsymbol{Z}_a = \boldsymbol{F}_a^T \boldsymbol{P}_a^{-1} \boldsymbol{F}_a.$$

The problem is, the autocorrelation terms (P_a) need estimates of the total power. So how do we get these?

A Bayesian CURN Search

Why don't we use a Bayesian search!

Instead of searching for full correlations (VERY SLOW)

Use a Common Uncorrelated Red Noise (CURN) search!

- Accurate estimates for everything you need
- At a rate >100x faster than a correlated search

An additional snag

We have distributions of posteriors, what now?

Which vector of parameters do we choose?

1. Select the maximum likelihood value
   a. Biased in PTAs with significant intrinsic red noise

2. Marginalize over many parameter vectors!
   a. Don’t choose one, choose a random sample of them!

[Vigeland et al 2018]

The Noise Marginalized OS (NMOS)

Use 1,000 - 10,000 random samples from the CURN search

- Run each through the OS
- Get a distribution of means, uncertainties, and SNRs

NOTE: Be careful, the distribution on $\hat{A}^2_{\rm gw}$ is NOT the full distribution on A^2, its only the distribution of means!

- Can use Uncertainty Sampling to get full amplitude distribution

[Gersbach et al 2018]

Some additional expansions

The OS can be further generalized in many ways

- Multiple Correlations (MCOS) [Sardesai et al 2024]
- Spectrally agnostic SNRS [Verma et al in prep.]
- Individual frequency searches (PFOS) [Gersbach et al 2024]
- Correlations for each telescope (MCOS)
- Anisotropic GWB searches [Pol, Taylor, Romano 2022]
- Many others

The OS is expansive. Ask me about how the OS can fit your project!