Bootstrap for multivariate time series and gravitational wave detection

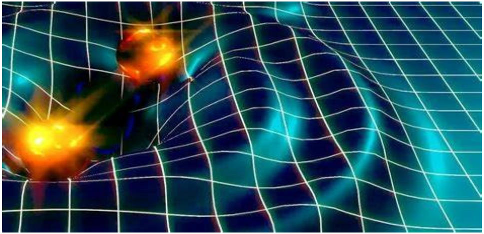
S.N. Lahiri

Department of Mathematics & Statistics
Washington University in St. Louis

s.lahiri@wustl.edu

GWB

- As described by general relativity, GW are freely propagating wave solutions to Einstein's equation, or "ripples" in the space-time metric.



- GW are expected to be generated by nearly any configuration of accelerating mass.
- However, due to the weakness of gravity, large masses/high accelerations (e.g., binary systems of neutron stars) are required to radiate significant GW.

GWB

- GWs can be indirectly inferred using precise measurements of timings of radio pulses from spinning, magnetized neutron stars (pulsars).
- The pulse times of arrival can be analyzed via models incorporating the GW component.
- One operational model for the observed pulsar time of arrivals (TOAs) can be written as

$$\tau = \tau^{\mathrm{TM}} + \tau^{\mathrm{DM}} + \tau^{\mathrm{GW}} + \tau^{\mathrm{other}} \tag{0.1}$$

  where
  - τ^{TM}: Physical model for TOAs taking into account spin period, proper motion, binary orbital dynamics, etc.
  - τ^{DM}: Model for time-varying dispersion measure variations.
  - τ^{GW}: Model for any GWs. This includes stochastic sources that have a unique correlation pattern across multiple pulsars, etc!

GWB

- Through some clever manipulation & approximations (cf. Demorest et al. (2007, 2012)), this leads to the following model for the *pre-fit* residuals for a single pulsar:

$$\mathbf{Y} = \mathbf{A}\beta + \boldsymbol{\epsilon}.$$

- Parameters β are estimated by WLS/GLS, giving

$$\widehat{\beta} = (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}\mathbf{Y}.$$

where $\mathbf{W}$ is (often) a diagonal matrix with inverse variances at each epoch (but can be more general).

GWB

- The *post-fit* residuals are given by $\mathbf{RY}$, where $\mathbf{R}$ is the projection operator:

$$\mathbf{R} = \mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{W}\mathbf{A})^{-\mathbf{1}}\mathbf{A}'\mathbf{W}.$$

- It is easy to check that $\mathbf{R}$ is idempotent (and singular).
- Further,

$$\mathbf{RY} = \mathbf{RA}\beta + \mathbf{R}\boldsymbol{\epsilon} = \mathbf{0} + \mathbf{R}\boldsymbol{\epsilon}.$$

- Thus, we can use the post-fit residuals to investigate the covariance structure of $\boldsymbol{\epsilon}$ that contains information about the GWB !!

GWB

- We write
$$\boldsymbol{\epsilon} = \boldsymbol{\epsilon}^{\text{gw}} + \boldsymbol{\epsilon}^{\text{other}},$$
where $\boldsymbol{\epsilon}^{\text{gw}}$ denotes the part of the noise due to GWB.
- Under an isotropic power law spectrum *assumption*, the **GWB covariances** are of the form:
$$\left((\mathbf{C}^{\text{gw}})\right)_{ij} = A_1^2 C^{\text{gw}}(t_i - t_j)$$
where
  - $C^{\text{gw}}(\cdot)$ is the covariance function corresponding to the spectrum and
  - A_1 is the unknown GW spectrum amplitude at the reference frequency $f_0 = 1\text{yr}^{-1}$ - **the parameter of interest!!!**

GWB

- The presence of GW component make the residual series of different pulsars correlated!!
- For a pair of pulsars (a, b), the noise variables $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\epsilon}_b$ can be written as

$$\boldsymbol{\epsilon}_a = \boldsymbol{\epsilon}_a^{\text{gw}} + \boldsymbol{\epsilon}_a^{\text{other}}, \quad \boldsymbol{\epsilon}_b = \boldsymbol{\epsilon}_b^{\text{gw}} + \boldsymbol{\epsilon}_b^{\text{other}}$$

where $\boldsymbol{\epsilon}_a^{\mathbf{gw}}$ and $\boldsymbol{\epsilon}_b^{\mathbf{gw}}$ are correlated, but $\boldsymbol{\epsilon}_a^{\text{other}}$ and $\boldsymbol{\epsilon}_b^{\text{other}}$ are NOT!

The GW component in residuals

- **The covariance between the noise variables $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\epsilon}_b$** for a pair of pulsars (a, b) is given by

$$\left((\mathbf{C}_{a,b})\right)_{ij} = A_1^2 C^{\mathrm{gw}}(t_i - t_j)\zeta(\theta_{ab})$$

  where

  - θ_{ab} = the angular separation between pulsars a and b and
  - $\zeta(\cdot)$ is the *Hellings-Downs function* !

- Thus, the cross-covariance matrix between $\boldsymbol{\epsilon}_a$ and $\boldsymbol{\epsilon}_b$ is determined by the GW power spectrum.

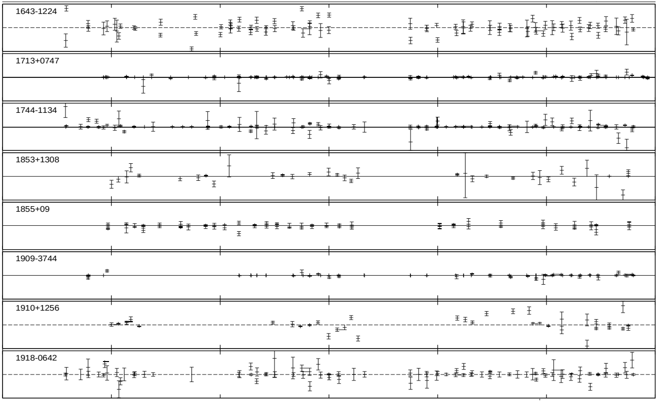
- Further, in $\mathbf{C}_{a,b}$, **the ONLY unknown parameter is A_1^2.**

The GW component in residuals

Q: How do we estimate A_1?

Cadence of TOAs/ residuals

- Here is a plot of the residuals for different pulars showing their cadence (cf. Demorest et al. (2012)):



Optimal Statistic

Note the following features:

- The time points are **irregularly** spaced!
- The **number of observations can be/are different** for any two distinct pulsars!
- The coverage is **NOT uniform** and there are **gaps appearing during 2007** in most of the series!
- The **densities and spans** of different series can be very **different**!
- There is **heteroskedsticity** (cf. size of the bars) !!

Optimal Statistic :

- Demorest et al. (2012) defined the following cross-correlation statistic (cf. eqn (9), p.13):

$$\rho_{ab} = \frac{\sum_{ijkl} r_i^{(a)} (C^{tot(a)})^-_{ij} C_{jk}(a,b) (C^{tot(b)})^-_{kl} r_l^{(b)}}{\sum_{ijkl} (C^{tot(a)})^-_{ij} C_{jk}(a,b) (C^{tot(b)})^-_{kl} C_{il}(a,b)}$$

where

- $i,j \in \{1, \ldots, N_a\}$ and $k, l \in \{1, \ldots, N_b\}$, with N_a denoting the number of TOAs for pulsar a, etc.
- $r_i^{(a)}$ and $r_l^{(b)}$ are the post-fit timing residuals for pulsars a and b, respectively.
- $\mathbf{C}^{tot(a)}$ is the (estimated) covariance matrix of the post-fit residuals for pulsar a, and $(\mathbf{C}^{tot(a)})^-$ is its generalized inverse!
- $C_{il}(a,b)$ is the (i,l) element of $R_a[\mathbf{C}_{a,b}]R_b'$.

GWB

- It can be shown that for all a, b,

$$E\rho_{ab} = A_1^2 \zeta(\theta_{ab}).$$

- Given a set of pulsars $\{1, \ldots, m\}$, we can set up the regression model:

$$\rho_{ab} = A_1^2 \zeta(\theta_{ab}) + \epsilon_{ab}$$

for pairs $(a, b) \in \Gamma$, where $\Gamma = \{(a, b) : 1 \leq a < b \leq m\}$.

- This leads to an estimator of A_1^2 of the form:

$$\hat{A}_1^2 = \frac{\sum_{(a,b)} \zeta(\theta_{ab}) \rho_{ab}}{\sum_{(a,b)} \zeta(\theta_{ab})^2}.$$

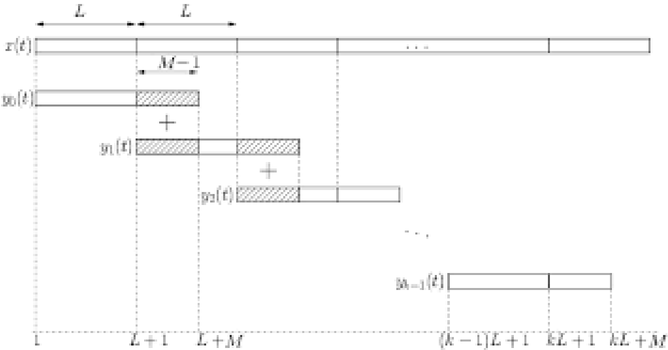
Optimal Statistic :

Q: How do we approximate the distribution of $\hat{A}_1^2$?

Bootstrap

- We can use the Bootstrap to approximate the distribution of $\hat{A}_1^2$.
- But what form of Bootstrap is appropriate?
    - Sampling with replacement / IID Bootstrap ?
    - Block Bootstrap ?

Construction of the Blocks for time series



- $M = 1$ gives the maximum overlapping version
- $M > 1$ can be used to reduce computational burden

Issues with the Block Bootstrap

- While this will work with a single time series, it may not be effective under the present scenario:
  - the number of pulsars $\approx 37+$
  - and the sample size is only around 500+.
- Thus, - **curse of dimensionality** will kick in!!
- Time-domain Block Bootstrap does not handle Red Noise very well (cf. Lahiri (1993)).
- There are also issues with irregularly spaced time-points! (cf. Lahiri and Zhu (2006)).

A new Bootstrap method

- Recall that

$$\rho_{ab} = \frac{\sum_{ijkl} r_i^{(a)} (C^{tot(a)})^{-}_{ij} C_{jk}(a,b) (C^{tot(b)})^{-}_{kl} r_l^{(b)}}{\sum_{ijkl} (C^{tot(a)})^{-}_{ij} C_{jk}(a,b) (C^{tot(b)})^{-}_{kl} \tilde{C}_{il}(a,b)}.$$

- Define

$$\mathbf{Z}^{(a)} \equiv (Z_1^{(a)}, \ldots, Z_{N_a}^{(a)})' = \left(\mathbf{C}^{tot(a)}\right)^{-1/2} \mathbf{r}^{(\mathbf{a})},$$

  the set of pre-whitened residuals for pulsar a.

- Note that ρ_{ab} can be written as

$$\rho_{ab} = \sum_{i=1}^{N_a} \sum_{l=1}^{N_b} w_{ab}(i,l) Z_i^{(a)} Z_i^{(b)}$$

  for some weights $w_{ab}(i,l)$.

A new Bootstrap method

- Thus, to define the Bootstrap version of

$$\hat{A}_1^2 = \frac{\sum_{(a,b)} \zeta(\theta_{ab})\rho_{ab}}{\sum_{(a,b)} \zeta(\theta_{ab})^2} = \sum_{(a,b)} \sum_{i=1}^{N_a} \sum_{l=1}^{N_b} \tilde{w}_{ab}(i,l) Z_i^{(a)} Z_i^{(b)},$$

  it is enough to be able to generate Bootstrap versions of $\mathbf{Z}^{(a)}$ for all a.
- Note that for each a, the variables $Z_1^{(a)}, \ldots, Z_{N_a}^{(a)}$ are approximately iid.
- So, we can resample with replacement to generate the (pre-)Bootstrap sample

$$Z_1^{\circ(a)}, \ldots, Z_{N_a}^{\circ(a)}.$$

A new Bootstrap method

- Note that this direct resampling only captures the **marginal** behavior of each set of residuals.
- It is important to also capture their interactions (correlations)!
- Thus, the Bootstrap version of $\{\mathbf{Z}^{(a)} : a = 1, \ldots, m\}$ is defined by

$$\begin{pmatrix} \mathbf{Z}^{*(a)} \\ \cdots \\ \mathbf{Z}^{*(m)} \end{pmatrix} = \Sigma^{1/2} \begin{pmatrix} \mathbf{Z}^{\circ(a)} \\ \cdots \\ \mathbf{Z}^{\circ(m)} \end{pmatrix}$$

where Σ is an $N \times N$ matrix (with $N = N_1 + \ldots + N_m$) consisting of $m \times m$ block matrices with (a, b)th block :

$$[\mathbf{C}^{tot(a)}]^{-1/2}\mathbf{C}_{a,b}[\mathbf{C}^{tot(b)}]^{-1/2}, \ a \neq b.$$

A new Bootstrap method

- Now define

$$[A_1^*]^2 = \sum_{(a,b)} \sum_{i=1}^{N_a} \sum_{l=1}^{N_b} \tilde{w}_{ab}(i,l) Z_i^{*(a)} Z_i^{*(b)}.$$

- We can use the Bootstrap distribution of $[A_1^*]^2$ to approximate the distribution of $\hat{A}_1^2$.
  - We can use the Monte-Carlo based Bootsrrap quantiles of $[A_1^*]^2$ to construct CIs for A_1^2.
  - This, in turn, can be used for testing $H_0 : A_1 = 0$, etc.

The END!!!

Thank you !!