Randomization Inference of Periodicity with Unequally Spaced Time Series

Panagiotis (Panos) Toulis
panos.toulis@chicagobooth.edu

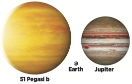
Econometrics and Statistics
University of Chicago, Booth School of Business

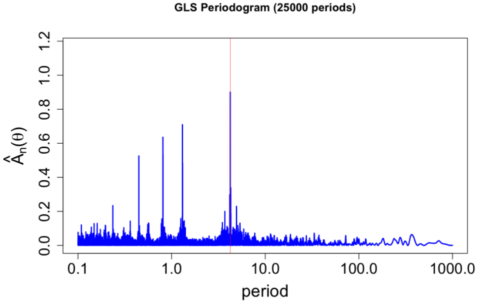
Joint work with Jacob Bean (U Chicago, Astrophysics)

Outline

1. Identification.
2. Failures with (asymptotic) frequentist or Bayes.
3. Method + examples.
4. (if time) Improving observation designs.

Detecting periodicity (51 Pegasi b)





Detection of periodicity poses no challenges as it is a well-studied problem (Fisher, 1929) and (Siegel, 1980; Bolviken, 1983; Chiu, 1989). Details

Most methods rely on the periodogram peak, $\hat{\theta}_n = \arg\max_{\theta \in \Theta} A_n(\theta)$.

How to use $\hat{\theta}_n$ for inference on θ^* (true period)?

Identification

A common mistake is to interpret detection of periodicity as θ^* being "near $\hat{\theta}_n$".

This implicitly relies on standard asymptotics of the form $\sqrt{n}(\hat{\theta}_n - \theta^*) \to \mathcal{N}(0, ..)$

However, these asymptotics break down even in the simple harmonic model.

1. Likelihood is irregular, non-smooth and multimodal $\Rightarrow$ no normality (of $\hat{\theta}_n$).
2. Observation times are not iid $\Rightarrow$ no consistency.
3. Pernicious effects from "hyperparameters" (e.g., granularity of period space, Θ).

As such, "$\pm$" statistical statements for period estimators can be meaningless.

(Bayes could resolve these issues? I think they make things worse.. Details)

Example 1: Synthetic data

Let $t_i = i + 0.05U_i$, $i = 1, \ldots, 100$, and $y_i = 1.5\cos(2\pi t_i/\sqrt{2}) + \varepsilon_i$, where $U_i \sim \text{Unif}[-1, 1]$ and $\varepsilon_i \sim N(0, 1)$ i.i.d. So, $\theta^* = \sqrt{2} \approx 1.414$.

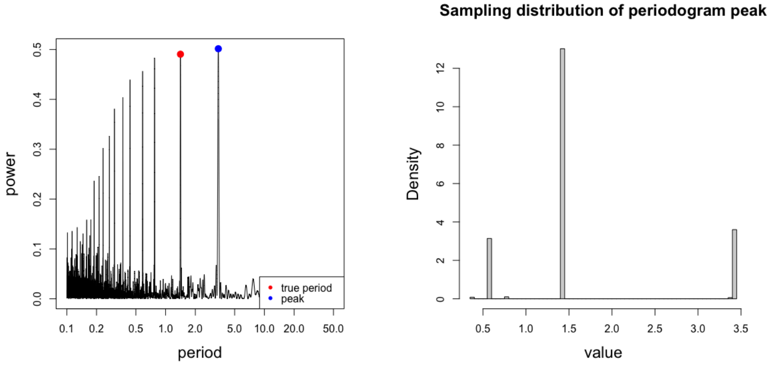


Figure: *Left:* Periodogram from one problematic dataset. *Right:* Sampling distribution of the periodogram peak from the same model over 1,000 replications.

Our method 1/2

Start with the test

$$H_0 : \theta_\star = \theta_0.$$

Fit model $y(t) = f_{\theta_0}(t) + e(t)$ and obtain $\hat{f}_{\theta_0}$ and $\hat{e}$ (residuals). Compute our test statistic $\hat{s}$ (e.g., periodogram peak, but we use a variation).

Then, simulate data:

$$y^{(1)}(t) = \hat{f}_{\theta_0}(t) + g^{(1)} * \hat{e}(t).$$
$$y^{(2)}(t) = \hat{f}_{\theta_0}(t) + g^{(2)} * \hat{e}(t).$$
$$....$$

Here,

- $g^{(i)}$ is a random transformation of the residuals (e.g., randomly flipping signs).
- Each dataset produces a new value for the test statistic, $s^{(i)}$.

Then, the p-value for H_0 is:

$$\text{pval}(\theta_0) = \text{E}(s^{(i)} \geq \hat{s}).$$

Our method 2/2

1. Choose Θ, a grid of values that contains θ^* w.p. 1. Pick a test statistic, s_n.
2. For all $\theta_0 \in \Theta$ do: Set $\hat{\Theta}_{1-\alpha} \leftarrow \hat{\Theta}_{1-\alpha} \cup \{\theta_0\}$ if $\text{pval}(\theta_0) > \alpha$.
3. Return $\hat{\Theta}_{1-\alpha}$ as the $100(1-\alpha)\%$ confidence set of θ^*.

Our method 2/2

1. Choose Θ, a grid of values that contains θ^* w.p. 1. Pick a test statistic, s_n.
2. For all $\theta_0 \in \Theta$ do: Set $\hat{\Theta}_{1-\alpha} \leftarrow \hat{\Theta}_{1-\alpha} \cup \{\theta_0\}$ if $\text{pval}(\theta_0) > \alpha$.
3. Return $\hat{\Theta}_{1-\alpha}$ as the $100(1-\alpha)\%$ confidence set of θ^*.

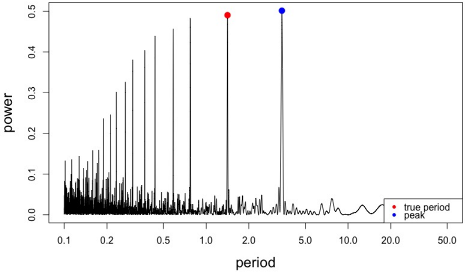
Advantages

- Confidence *set*, not interval. Appropriate for identification.
- Can accommodate complex error structure (through $g^{(i)}$). Distribution-free.
- No assumption on the test statistic. Not necessary to be "well-behaved" (e.g., consistent or normal).
- No assumption on the observation design or spacings.
- Inference conditional on hyperparameters (e.g., Θ).

Challenges

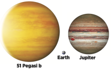
- Power: Is the method conservative?
- Choice of test statistic. Details
- Computational challenges (requires computation over entire Θ). Details

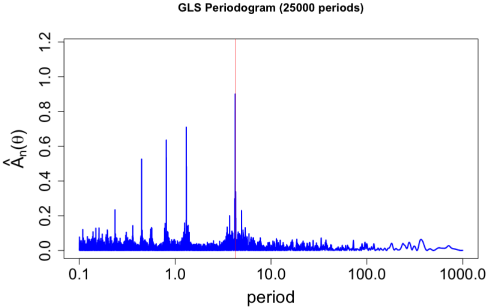
Example 1: Synthetic data — What does our method produce?



θ_0	p-value	$\hat{\Theta}_{0.95}$	$\hat{\Theta}_{0.99}$
0.1752	0.00	no	no
0.1890	0.00	no	no
0.2124	0.00	no	no
0.2330	0.00	no	no
0.2696	0.00	no	no
0.3036	0.00	no	no
0.3693	0.00	no	no
0.4362	0.00	no	no
0.5857	0.03	no	yes
0.7737	0.17	yes	yes
1.4130	0.48	yes	yes
3.4175	1.00	yes	yes

51 Pegasi b (Mayor and Queloz, 1995)

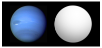


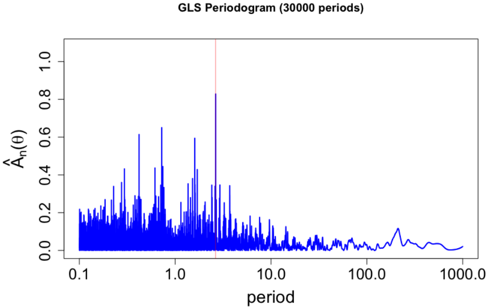


θ_0	p-value	$\hat{\Theta}_{0.95}$	$\hat{\Theta}_{0.99}$
0.3085	0.00	no	no
0.5662	0.00	no	no
0.8069	0.00	no	no
0.8089	0.00	no	no
0.8295	0.00	no	no
1.3047	0.00	no	no
1.3095	0.00	no	no
3.7033	0.00	no	no
4.1807	0.00	no	no
4.2311	1.00	yes	yes
4.2821	0.00	no	no
4.9331	0.00	no	no

Left: Periodogram of radial velocity on exoplanet "51Pegb".

We see that there are no identification issues as the 4.23-day signal is the only one accepted in the confidence sets.

Gliese 436 b (Butler et al., 2004) 

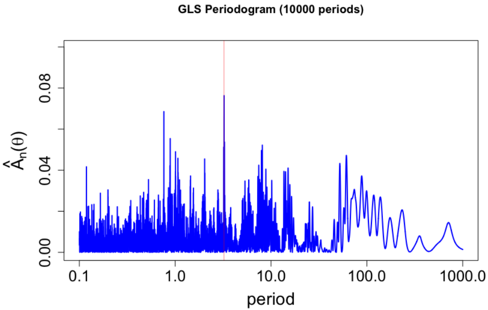


θ_0	p-value	$\hat{\Theta}_{0.95}$	$\hat{\Theta}_{0.99}$
0.4200	0.00	no	no
0.6155	0.00	no	no
0.7067	0.00	no	no
0.7438	0.00	no	no
1.3641	0.00	no	no
1.5187	0.00	no	no
1.6013	0.00	no	no
1.6086	0.00	no	no
1.7008	0.00	no	no
2.4103	0.00	no	no
2.6441	1.00	yes	yes
3.7092	0.0000	no	no

Left: Periodogram of radial velocity on exoplanet GJ436b.

We see that there are no identification issues as the 2.64-day signal is the only one accepted in the confidence sets.

α Centauri B (Dumusque et.al., 2012)

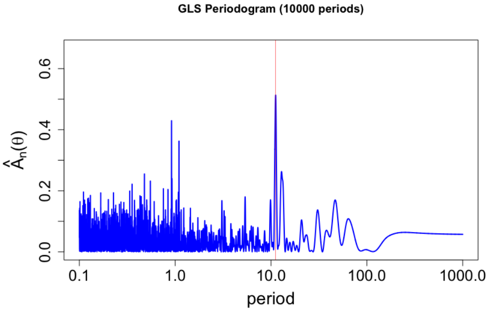


θ_0	p-value	$\hat{\Theta}_{0.95}$	$\hat{\Theta}_{0.99}$
0.7622	0.0705	yes	yes
0.8882	0.0271	no	yes
1.0086	0.0174	no	yes
1.0678	0.0079	no	no
2.0292	0.0122	no	yes
3.2074	0.0163	no	yes
3.2371	1.0000	yes	yes
3.2670	0.0178	no	yes
7.9394	0.0116	no	yes
8.1169	0.0175	no	yes
52.2242	0.0121	no	yes
61.1334	0.0226	no	yes

Left: Periodogram of radial velocity on candidate exoplanet orbiting α Centauri B.

We see that there are severe identification issues as several signals other than the periodogram peak are accepted in the confidence sets.

Proxima Centauri (Anglada-Escude et.al., 2016)



θ_0	p-value	$\hat{\Theta}_{0.95}$	$\hat{\Theta}_{0.99}$
0.1106	0.0007	no	no
0.3355	0.0022	no	no
0.3552	0.0055	no	no
0.4778	0.0025	no	no
0.5512	0.0047	no	no
0.7532	0.0052	no	no
0.8412	0.0059	no	no
0.9164	0.0173	no	yes
0.9266	0.0005	no	no
1.0957	0.0080	no	no
11.1739	1.0000	yes	yes
12.8769	0.0006	no	no

Left: Periodogram of radial velocity on candidate exoplanet Proxima Centauri b.

We see that there are no severe identification issues. The detection appears to be robust except for a nuisance signal at 0.9164 days.

Could optimize observation design to get rid of the nuisance signal.

Thank You.

Toulis, P. and Bean, J. (2021). Randomization Inference of Periodicity in Unequally Spaced Time Series with Application to Exoplanet Detection (*submitted*)

Observation designs

The importance of observation times in identifying a periodic signal is well understood (Feigelson and Babu, 2012; VanderPlas, 2018; Ivezic et al., 2014).

Surprisingly, there is little (to none) work in the statistical aspects of careful observation design.

We propose to synthesize data under alternative designs, and then pick the design that yields "ϵ-identification"; i.e., $\hat{\Theta}_{1-\alpha}$ only contains values ϵ-away to a candidate signal θ_*^{cand}.

We address two questions:

1. How much to randomize observation times for ϵ-identification?
2. How many more observations to make for ϵ-identification?

(Candidate) Exoplanet	Design (A) randomness needed for identification (best δ)	$\pm$ hrs.	Design (B) #additional obs. needed for identification (best $n' - n$)
51 Pegasi b	0	0	0
Gliese 436 b	0	0	0
α Centauri B	0.18	4.32	137
Proxima Centauri	0.06	1.44	17

Table: Observation designs (A) and (B) to achieve identification in the exoplanet applications. Design (A) introduces randomness in the observation times, while design (B) introduces additional observations.

We see that 51Pegb and GJ436b require no improvement in the observation times.

For α Centauri B: We need an additional variation of ± 0.18 days around the actual observation times (i.e., ± 4.32 hrs./observation). Alternatively, we need 137 new observations with a random variation of ± 15 mins./observation.

For Proxima Centauri: We need an additional variation of ± 0.06 days (i.e., ± 1.44 hrs./observation) on the actual observation times. Alternatively, we only need an 17

Detecting periodicity — Periodogram peak

Main method developed by Fisher (1929). Power refined by (Siegel, 1980; Bolviken, 1983; Chiu, 1989), and extended to more general hypotheses (Juditsky et al., 2015) and sparse alternatives (Cai et al., 2016).

Most methods rely on the periodogram peak, $\hat{\theta}_n = \arg\max_{\theta \in \Theta} A_n(\theta)$.

Idea is to reject the null of no periodicity when the peak exceeds a threshold ("false alarm probability"). See also (Baluev, 2008, 2013; Delisle et al., 2020; Nemec and Nemec, 1985) for adaptations in astronomy.

Under normality assumptions, each $A_n(\theta)$ is associated to a χ^2_2, and so the distribution of $\hat{\theta}_n$ (under the null) can be approximated via extreme value theory.

Detection of periodicity is generally robust and poses no major challenges.

Go back

Bayesian methods?

We might expect that a Bayesian approach could address these issues.

However, a Bayesian approach also faces problems.

(i) Prior specification: uniform priors give preference to parameter regions that not only have high likelihood but are also wide. This sweeps the identification problem "under the rug"; see also (Hall and Yin, 2003, Section 1).

(ii) Posterior summarization is challenging when the likelihood is multimodal and non-smooth. Also affected by hyperparameters (e.g., Θ.)

(iii) Model selection: Bayes factors may strongly depend on features that are esoteric to the specified models. See also (Gelman and Yao, 2020, Sections 3 and 6).

Go back

Structured inference

Suppose we want to estimate parameter $\theta^* \in \Theta$ through a statistic S.

Typical asymptotic approach for inference is to derive a law $\sqrt{n}(S - \theta^*) \to \ldots$ and then pivot to CIs. Relies on asymptotics and usually normality.

However, we can do finite-sample valid inference if we know that

$$\mathrm{g}S \overset{\mathrm{d}}{=} S,$$

for some transformation g, via inversion of randomization tests.

The simplest case is when we have access to $f(S \mid \theta)$, the distribution of S. Then, we can build a finite-sample valid confidence set for θ^* (cf. Neyman construction):

Construct 95% *confidence set*:

$$\widehat{\Theta} = \left\{ \theta \in [0,1]^3 : \sum_{s \in \mathbb{S}} \mathbb{I}\{f(s|\theta) \le f(s_{\text{obs}}|\theta)\} f(s|\theta) > 0.05 \right\}.$$

In words: "accept all θ for which there is at least 5% of the density mass of $f(S|\theta)$

Comparison with standard methods

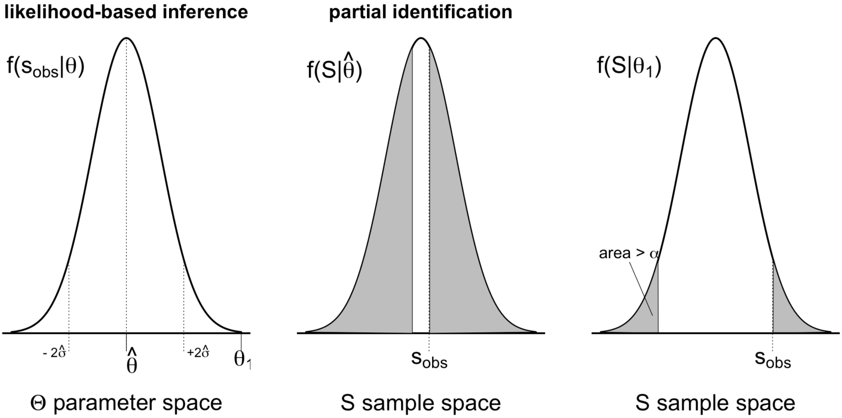
For standard methods:

- Focus is on $f(s_{\text{obs}}|\theta)$ as a *function of* θ (likelihood-centric).
- Inference "happens around the mode", $\hat{\theta} = \arg\max_\theta f(s_{\text{obs}}|\theta)$. Tails of likelihood are ignored.
- The "hope" is that $\hat{\theta}$ is near θ_0. Asymptotics and approximations are necessary.
- Many problems (usually undetected) when #samples is small, likelihood is multimodal, nonsmooth, modes are not separable, etc. (think of exoplanet detection!).

For structured inference methods:

- Focus is on $f(S|\theta)$ as a *function of* S or on invariances $gS \overset{d}{=} S$.
- Inference "happens everywhere" in the parameter space. The likelihood value of $f(s_{\text{obs}}|\theta)$ only matters *relatively* to other values $f(S|\theta)$.
- No asymptotics or approximations are necessary.
- Finite sample guarantee: Works even when #samples is small, likelihood is multimodal, nonsmooth etc.
- Downside: requires computation over entire Θ and possible over $\mathbb{S}$ (sample

Illustrative comparison



Go back or Covid-19 application

Covid-19 serology model

We have two calibration studies and one main study:

$$\begin{aligned}
& & & \text{observed values} \\
S_c^- &= \#\text{positives in calibration study out of 401 true negatives} & s_c^- &= 2; \\
S_c^+ &= \#\text{positives in calibration study out of 197 true positives} & s_c^+ &= 178; \\
S_m &= \#\text{positives in main study out of 3,330 trials} & s_m &= 50.
\end{aligned}$$

<u>**Assume**</u>:

$$\begin{aligned}
\text{pr}(\text{positive result}|\text{actual negative}) &= p \quad [\text{false positive rate}] \\
\text{pr}(\text{positive result}|\text{actual positive}) &= q \quad [\text{true positive rate}] \\
\frac{\#\text{ actual positives in main study}}{3,330} &= \pi \quad [\text{prevalence}].
\end{aligned} \tag{1}$$

Parameter $\theta = (p, q, \pi) =\in [0, 1]^3$, and statistic $S = (S_c^-, S_c^+, S_m) \in \mathbb{S}$.

Key observation: We can calculate the density, $f(S|\theta)$, of the statistic <u>exactly</u>.

Covid-19 serology model

Setup: $\theta = (p, q, \pi) =$ (FPR, TPR, prevalence), data $S = (S_c^-, S_c^+, S_m)$.

Density of data statistic.

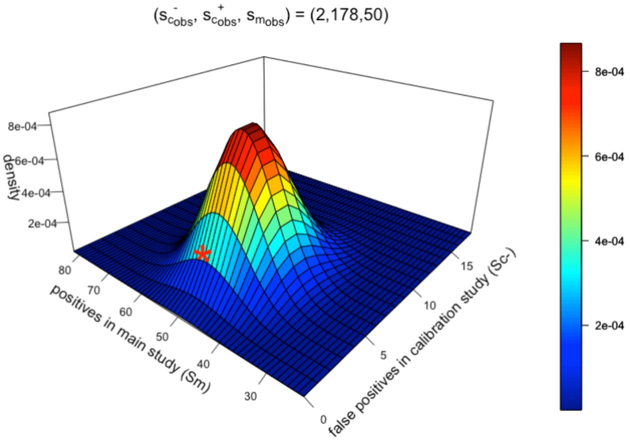
$$f(S|\theta) = \underbrace{\text{Bin}(S_c^-; 401, p)}_{\text{FP in calibration}} \cdot \underbrace{\text{Bin}(S_c^+; 197, q)}_{\text{TP in calibration}} \cdot \underbrace{\sum_i \text{Bin}(i; N_\pi, q) \cdot \text{Bin}(S_m - i; N - N_\pi, p)}_{\text{prob of } S_m \text{ positives out of } N_\pi \text{ actual positives in main study}},$$

where $N_\pi = 3300\pi =$ #actual positives in main study.

- In the sample, we observe $s_{\text{obs}} = (2, 178, 50)$. How to do inference on θ?

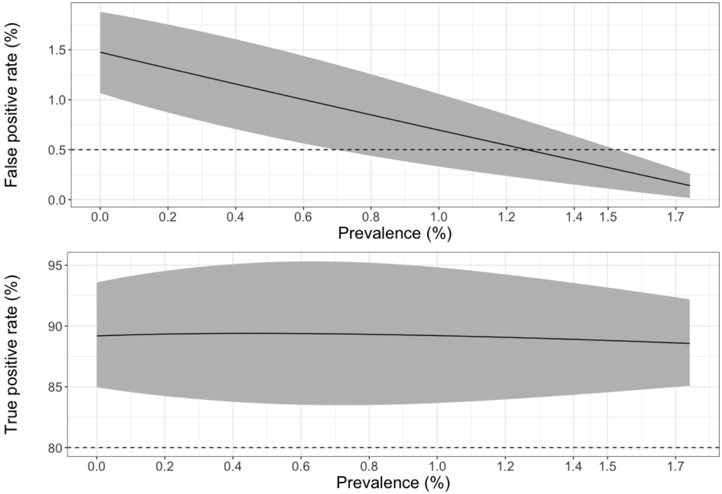
Illustration

Suppose $\theta_0 = (p, q, \pi) = (1.5\%, 100\%, 0\%)$. Then, $f(S|\theta_0)$ looks as follows:



- We have to decide: Is θ_0 plausible?

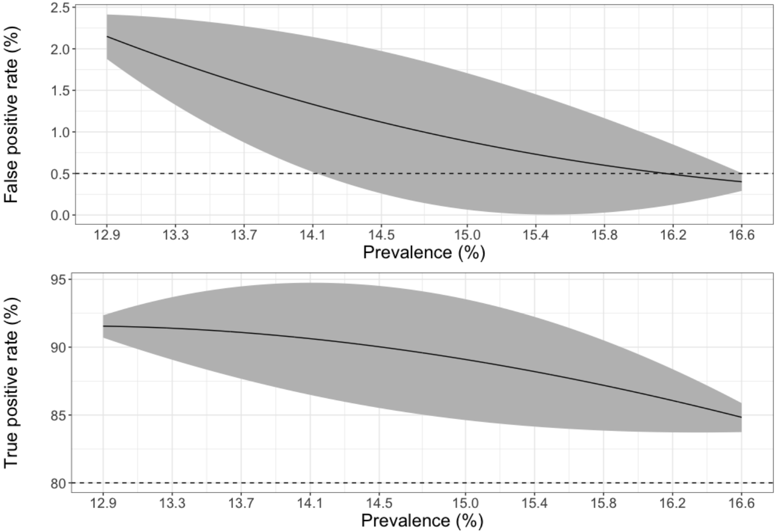
Application: Santa Clara study



Visualization of (p, q, π) in $\widehat{\Theta}$; dashed lines = empirical estimates of FPR, TPR;

Results: $\pi = 0\%$ is included; but [0.7-1.5%] is arguably more plausible.

Application: New York study



Discussion: Choice of test statistic

Procedure 1 is valid for any choice of the test statistic, s_n.

However, power depends on how sensitive s_n is in detecting violations of the null hypothesis.

We choose $s_n(Y^n, T^n) = A_n(\hat{\theta}_n) - A_n(\theta_0)$, the difference between periodogram values at the global peak peak and the null, θ_0.

Fisher's classical statistic is $s_n = \max_{\theta \in \Theta} \hat{A}_n(\theta)/\bar{A}_n$, where $\bar{A}_n = |\Theta|^{-1} \sum_\theta A_n(\theta)$.

Improvements using a trimmed mean in place of $\bar{A}_n$ have also been suggested (Bolviken, 1983; Siegel, 1980; Damsleth and Spjotvoll, 1982). See also (McSweeney, 2006) for numerical comparisons. Go back

Discussion: Computation

The complexity of our method is, prima facie, $O(|\Theta|^2 \cdot R \cdot C)$, where $C =$ time for weighted least-squares.

e.g., for $|\Theta| = 10^4$, $R = 10^3$, and $C = 50\mu$s an analysis on a conventional laptop of a time series with 200 observation times takes a total of 1,388 hrs. of wall clock time (approx. 58 days).

However, several reductions of computation time are possible.

1. Procedure 1 can be fully parallelized in step 3; e.g., with 100 nodes the wall clock time thus drops to 14 hrs.
2. Again in step 3, there is no need to consider all values in Θ but only a proportion; e.g., consider local peaks that are at least 20% as high as the global peak. This leads to a complexity $O(\gamma|\Theta|^2 \cdot R \cdot C)$ with $\gamma \sim 0.1\%$-3%.

As such, the computation in the above example drops dramatically to approximately 30 mins. of wall clock time. Indeed, in our application, get up to $R = 100,000$ and still finish all analyses in a few hours using a cluster with 400 nodes. Go back

Randomization Tests (Lehman and Romano, 2005)

Let $D \in \mathbb{R}^n$ be the data, and $\mathcal{G}^n$ a group of $\mathbb{R}^n \times \mathbb{R}^n$ transformations.
We are testing some H_0 under which:

$$D \overset{d}{=} \mathrm{g}D, \text{ for all } \mathrm{g} \in \mathcal{G}^n.$$

Define a test statistic $T_n = t_n(D)$ and $\mathrm{T}_D = \{t_n(\mathrm{g}D) : \mathrm{g} \in \mathcal{G}^n\}$. Then,

$$T_n \mid \mathrm{T}_D = \text{ Uniform}.$$

To test H_0, we could take the p-value of T_n wrt to T_D.

$*$ This test is (i) exact in finite samples and (ii) works for any choice of T_n.

Go back

Error invariance

Assumption: For any observation times $T^n = \{t_1, \ldots, t_n\}$, with n finite, there exists an algebraic group $\mathcal{G}^n$ of $n \times n$ matrices such that

$$\mathrm{g} \cdot \varepsilon^n \overset{\mathrm{d}}{=} \varepsilon^n \mid T^n \quad (\mathrm{g} \in \mathcal{G}^n). \tag{A2}$$

To keep things simple, we assume that $\mathcal{G}^n = [\pm]^{n \times n}$, the set of $n \times n$ diagonal matrices with ± 1 in the diagonal.

As such, our inference works with any symmetric distribution of independent errors beyond just normal (Gaussian) as frequently assumed in practice.

This formulation follows the framework of randomization tests (Lehmann and Romano, 2006) where testing is based on structural rather than analytical assumptions.

Example of "structured inference". Details Go back

Non-parametric approach (1/2)

Define

$$\Pi(T^n;\theta) = \{\pi \in \mathsf{S}_n : \pi(t_i) \equiv t_i (\text{mod}\,\theta),\ i = 1,\dots,n\}.$$

In words, $\Pi(T^n;\theta)$ is the set of permutations of $(t_1,\dots,t_n)$ such that any time t_i is mapped only to an observation time that is equivalent to t_i modulo θ.

We wish to test the following nonparametric null hypothesis of periodicity θ_0:

$$H_0^{\text{np}} :\ y^{\text{p}}(t') = y^{\text{p}}(t),\ \text{for all } t', t \text{ such that } t' \equiv t (\text{mod}\,\theta_0). \tag{2}$$

To test H_0^{np} we can adapt Procedure 1 as follows.

1. **For** all $r = 1,\dots,R$ **do:**
   - (i) Sample $\pi \sim \text{Unif}\big(\Pi(T^n;\theta_0)\big)$.
   - (ii) Generate synthetic outcome data $Y^{n,(r)} = \pi \cdot Y^n$ obtained by permuting the data Y^n according to π while observation times, T^n, are fixed.
2. Using the samples from 2(ii), calculate the p-value, say $\text{pval}(\theta_0)$, as in (**??**), and reject if the p-value is less than α.

Theorem

Suppose that Assumptions (A1) and (A2) hold with $\mathcal{G}^n = \Pi(T^n;\theta_0)$. Then, the p-value from Procedure 2 is exact under H_0^{np} conditionally on the observation times, that is,

Non-parametric approach (2/2)

An alternative approach would be to use the nonparametric estimators of θ^* developed by (Hall et al., 2000); (Hall and Li, 2006); (Hall, 2008) together with a variation of Procedure 1 or Procedure 2.

Both these procedures do not require regularity conditions on the observation times but only a consistent estimator for the periodic component, y^{p}. We leave these directions for future work. Go back