TSDA lecture 8 - 2020

Non-Fourier methods for period detection:

- Epoch-folding — Leahy et al. (1983)
- Phase dispersion minimization (PDM) — Stellingwerf et al. (1978)
- Minimum string length — Dworetsky (1983)
- Rayleigh (Z^2) test — Leahy et al. (1983), Buccheri et al. 1983
- Analysis of Variance (AoV, ANOVA) — Schwarzenberg-Czerny (1989)
- Bayesian periodicity search — Gregory & Loredo (1992)

Why non-Fourier methods?

- Sensitive to non-sinusoidal variations, e.g. eclipses.
- Arrival time data vs binned data (lends itself to phasing)
- Can use error bars for weighting (although this is also possible in Lomb-Scargle - see `astropy`'s implementation of `LombScargle`)

Epoch folding, minimum string length and phase dispersion minimization

Aim: to find periodicity in time series that may be gappy and non-sinusoidal

Method:

1. Fold time series over a set of periods
2. Determine some test statistic for each folded lightcurve. Test statistic may be related to mean/variance within phase bins compared to sample mean/variance
3. Period of variability is where this test statistic differs significantly from the test statistic at other periods.

Epoch folding & PDM

Consider ***N*** observations $x_i, i = 1, 2, 3...N$ folded onto ***M*** trial phase bins.

x_{kj} is the *k*th observation in *j*th phase bin

n_j is the total number of data points in phase bin *j*

In each phase bin $\bar{x}_j$ is the sample mean and s_j^2 is the variance in the *jth* bin

Null hypothesis is that the $\bar{x}_j$ are uniform across phase.

The null hypothesis can be rejected if $\bar{x}_j$ are not uniform across phase.

Global mean is $\bar{x}$

how science works

1. Observe: *record the data*
2. Reduce: *clean-up the data*
3. Analyse: *get numbers from data, summary descriptors, statistics*
4. Conclude: *involves modelling, testing a hypothesis,*
5. Reflect: *what did we learn, what new data can check or refute our conclusion?*

what's a hypothesis?

a **proposed explanation** made on the basis of **limited evidence** as a **starting point** for further observation

- steady state hypothesis of the origin of the universe
- earth is at the centre of the solar system

what's the difference between an **hypothesis** and a **theory**?

hypothesis testing

- are our data consistent with someone else's data?
- are our data consistent with a model?
- are our data correlated with some variable?

hypothesis testing

1. Set up 2 possible and exclusive hypotheses, each with an associated terminal action

   1.1. H_0 (null hypothesis), usually formulated to be rejected

   1.2. H_1 (alternative hypothesis)

2. Specify a priori the significance level (α). Choose a test. Obtain sampling distribution and region of rejection whose area is a fraction α of the total area in the sampling distribution.

3. Run the test; reject H_0 if the test yields a value of the statistic whose probability of recurrence under H_0 is $\leq \alpha$.

4. Carry out the terminal action

Epoch folding statistic

$$Q^2 = \frac{1}{\sigma^2} \sum_{j=1}^{M} n_j (\bar{x}_j - \bar{x})^2$$

Note that here $\hat{\sigma}^2$ is the *sample* variance - not the variance within the bin.

Test statistic Q^2 has a χ^2 distribution and can be used with a χ^2 test.

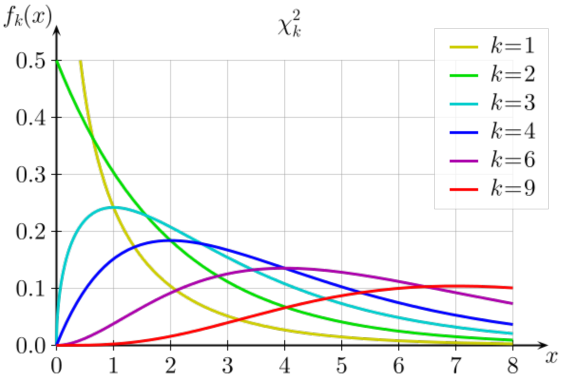
Leahy et al. 1983 see Reading folder on the Cloudcape

Aside: χ^2 distribution

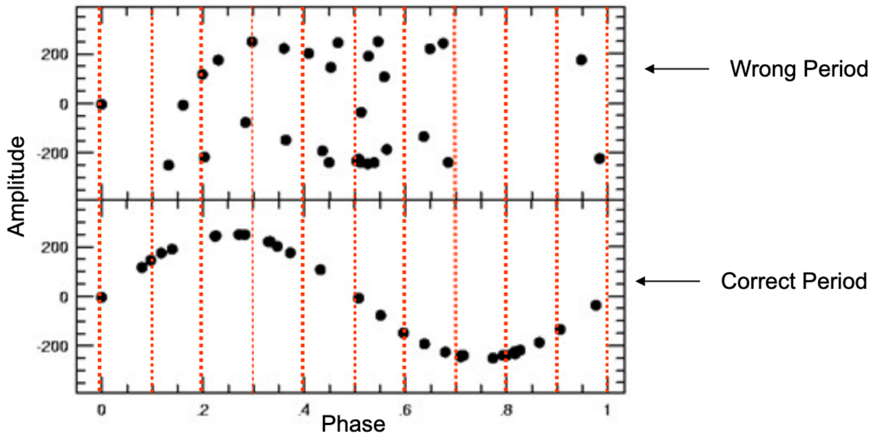
If $Z_1, .., Z_k$ are independent, normal random variables, then the sum of their squares

$$Q = \sum_{i=1}^{k} Z_i^2$$

is distributed according to the χ^2 distribution with k degrees of freedom.



Phase dispersion minimisation (PDM)



Choose a period and phase the data. Divide phased data into *M* bins and compute the standard deviation in each bin. If σ^2 is the variance of the time series data and s^2 the total variance of the *M* bin samples, the correct period has a minimum value of Θ :

$$\Theta = s^2/\sigma^2$$

Slide from Gavin Freer: http://slideplayer.com/slide/4212629/

PDM statistic

$$s^2 = \frac{\sum (n_j - 1)s_j^2}{\sum n_j - M} \qquad \Theta = \frac{s^2}{\sigma^2}$$

Note that here σ^2 is the *sample* variance and s_j^2 is the variance within the *j*th bin.

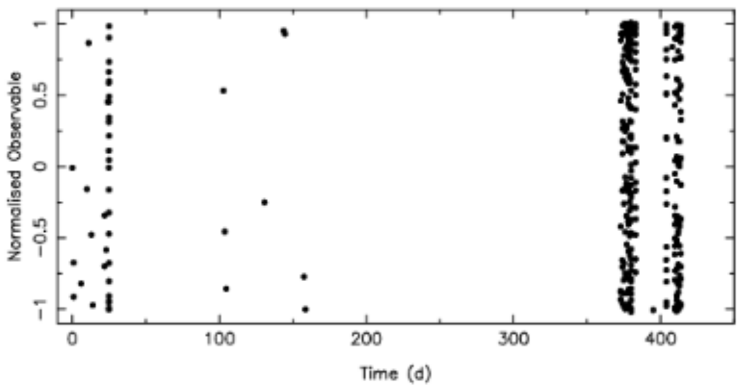
If there is a period in the data, $\mathbf{\Theta}^2$ will be minimized because there will be very small variance in the bins relative to the global light curve variance.

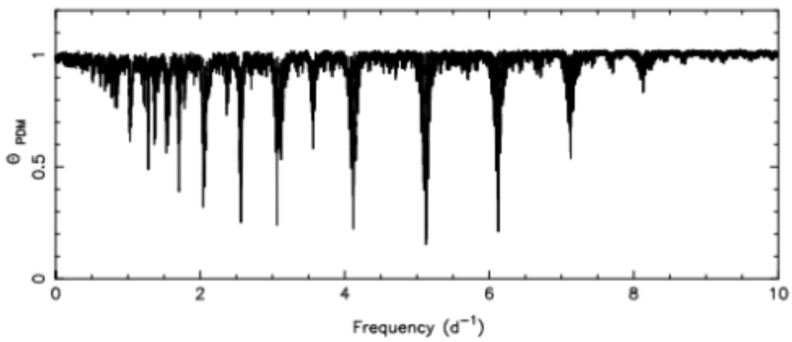
Example implementations:

1. in Python https://tinyurl.com/ybvwak2d

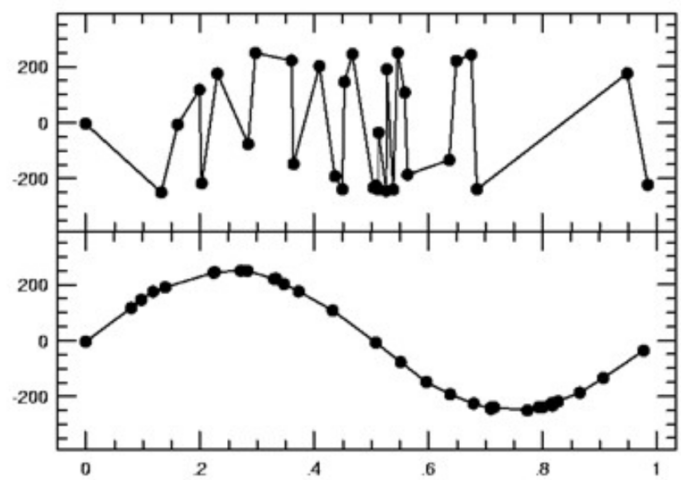
2. Period04 toolkit at www.univie.ac.at/tops/Period04/

PDM





Minimum String Length



Wrong period longer string length

Phase

Phase the data to a test period and minimize the distance between adjacent points

remember - the implementation is still in phase **bins**

Slide from Gavin Freer: http://slideplayer.com/slide/4212629/

Minimum string length

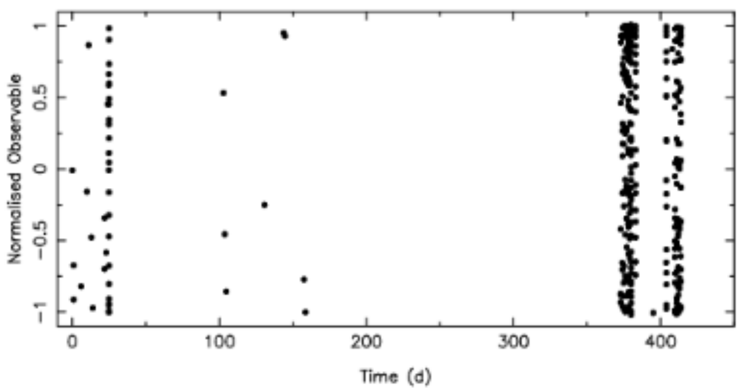
$$\Theta^2_{\mathrm{LK}} = \frac{\sum_{j=1}^{M}(\bar{x}_j - \bar{x}_{j+1})^2}{\sum_{j=1}^{M}(\bar{x}_j - \bar{x})^2}$$

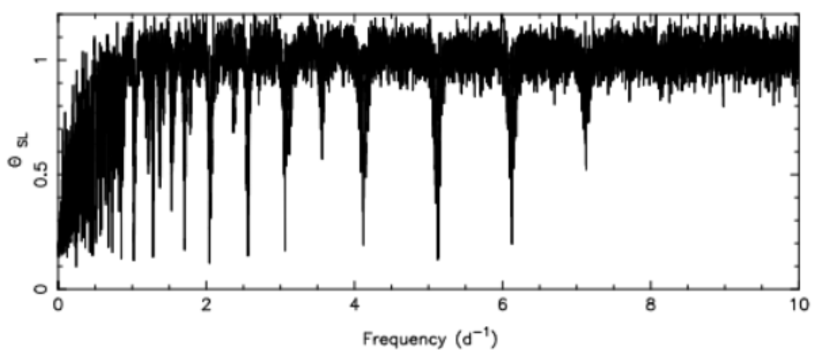
This statistic measures the difference in means between adjacent bins, relative to the difference between the bin means and the global mean.

If you fold the data on a true period, this statistic will be minimized.

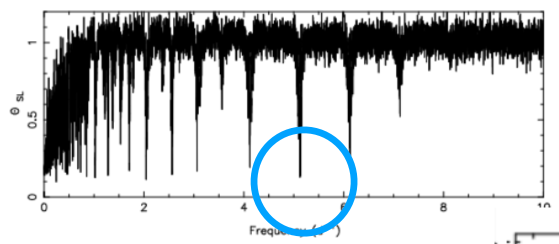
Dworetsky 1983; Lafler & Kinman 1965; Clarke 2002

Minimum string length



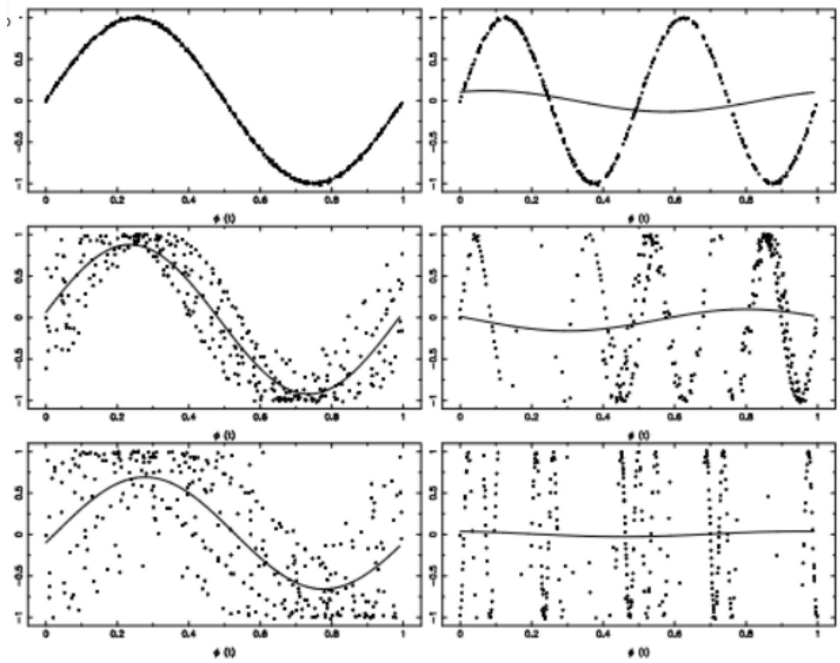


Minimum string length



What's the real period?

Looking at the folds for these dips:



Exercise

Python notebook on the cloudcape: PDM_Z2_2019_Tut3.ipynb

Class Test 1

16 March 2019
i.e. next week Monday
11am - 1pm

Open book
You can consult notes, google, stackoverflow, but **not your colleagues**
Test formatted as a Jupyter notebook

Content
Characterising time series
Fourier transforms
Lomb-Scargle
I.e. everything **before this** lecture