Frequency estimation uncertainties of periodic signals

Laurent Eyer¹, Jean-Marc Nicoletti², S. Morgenthaler³

(1) Department of Astronomy, University of Geneva, Sauverny, Switzerland
(2) Swiss Statistics Federal Office, Neuchâtel, Switzerland
(3) Applied statistics, EPFL, Lausanne, Switzerland

Statistics and exoplanets
Focus Meeting 8, IAU
Hawaii, USA
Tuesday August 4, 2015
Motivation: The catalogue of the Gaia mission

• Cornerstone mission of the European Space Agency

• Observations of all the objects brighter than a certain limit G=20.5 (>1.5 billion objects)

• Multi-epoch measurements of:
  - positions (astrometry)
  - brightnesses, colors (spectro-photometry)
  - radial velocities (spectroscopy)

• Launch (Soyuz rocket): December 19 2013

• Length: 5 (+1) years (70 times all sky)

• Final Catalogue: 2021-2022
Motivation: The catalogue of the Gaia mission

The content of the ESA Gaia catalogue

<table>
<thead>
<tr>
<th>Objects</th>
<th>1.5 billion objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>— Variable objects</td>
<td>150 million?</td>
</tr>
<tr>
<td>- Periodic</td>
<td>20 - 30 million?</td>
</tr>
<tr>
<td>see next talk by Süveges</td>
<td></td>
</tr>
<tr>
<td>· Exoplanet (from transits)</td>
<td>100s to 1000s?</td>
</tr>
<tr>
<td></td>
<td>Dzigan &amp; Zucker 2012</td>
</tr>
</tbody>
</table>

General principle

In the catalogue if there is an estimated quantity, it is advisable, to give its uncertainty for example: frequency of the periodic signals.
Introduction: the problem

Time series: Time, Magnitude, Uncertainty

\[ t_i, m_i, \sigma_i, \quad i = 1, \ldots, n \]

Unevenly sampled data
Heteroscedastic data, with correlations in the noise

Period Search Method
(“Fourier”, ANOVA, String, etc…)

Estimation of frequency

Question: what is the standard deviation of this quantity?
Introduction: an example

Time series: Time, Magnitude, Uncertainty

Simulated data from Gaia mission over (partial data) time span, $\Delta T$, of 3 years

Sinusoid with Period $P = 1/\nu$

and Amplitude $A$

Gaussian noise $N(0, \sigma^2)$

$n = 37$

$\nu \in [0, 25]$ 1/day

$A/\sigma = 6$

e.g. 6 mmag

1 mmag

Period Search Method
("Fourier")

Estimation of frequency
Result

The image shows a scatter plot with two distinct linear trends. The x-axis represents the Input Frequency [1/day], and the y-axis represents the Output Frequency [1/day]. The figure illustrates a strong linear relationship between the input and output frequencies, with data points forming two distinct lines, indicating a possible correlation or pattern in the data.
Well known problem of aliasing

Alias and Nyquist frequency (Eyer & Bartholdi 1999) of unevenly sampled data as seen from the spectral window.
Estimation of uncertainty on the frequency

For a sinusoid independent and identically distributed noise
Regular sampling

\[ \delta \nu = \frac{\sqrt{6}}{\pi} \frac{\sigma}{A\sqrt{n} \Delta T} \]

Cuypers 1987

Number of measurements: \( n \)
Amplitude: \( A \)
Time span: \( \Delta T = t_n - t_1 \)
Standard deviation of Noise: \( \sigma \)

Independent of the frequency
In astronomy literature

Kovacs (1981) derived a similar result and also noted that in a periodogram, shifts may occur in frequency:

\[
\frac{0.11}{\nu \Delta T^2}
\]

Baliunas et al. (1985), Baliunas and Fisher (1985) used these formulae, and many others after I contacted Prof. Morgenthaler at EPFL (Chair of Applied Statistics), Lausanne
From statistics perspective

B. Quinn 2012 (Handbook of Statistics: Time Series Analysis: Methods and Applications):

Problem was studied by Whittle 1952 and solved rigorously in 1971 by Walker. According to Quinn, these results remained unknown to engineering literature for 20 years

Generalisation of Walker (1973) with correlated errors
The effect of light curve shape

From common sense: If there are sharp features in the light curve, the frequency will be better determined.

RR Lyrae star

Transit

Gaia data
Image of the Week on Gaia ESA website: http://www.cosmos.esa.int/web/gaia

Kepler data
from Shporer et al. 2014
More general estimation

From Hall et al. (2000, biometrika):

\[
\delta \omega = \sqrt{24\pi} \frac{\sigma \sqrt{\omega_0}}{\eta \sqrt{n^3 \int_0^{P_0} g'(u)^2 du}}
\]

where

- Signal: \( g(u) \)
- Number of measurements: \( n \)
- Standard deviation of Noise: \( \sigma \)
- Angular frequency: \( \omega = 2\pi \nu \)

and \( \eta \) is the expectation value of \( t_{i+1} - t_i \).
Generalisation of the result by Nicoletti

Two approaches to period search

Harmonic analysis ("Fourier")

Non parametric regression

Uncertainty on frequency:

Estimation of the variance of the asymptotic distributions

With relaxing assumptions

From regular to irregular sampling

From i.i.d. to correlated errors
Generalisation of the result

submitted to Journal of Time Series Analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Signal</th>
<th>Sampling</th>
<th>Errors</th>
<th>Asymptotic variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker 1971</td>
<td>AH</td>
<td>$A \cos(\omega_0 t) + \frac{B}{2} \sin(\omega_0 t) + \mu$</td>
<td>regular</td>
<td>iid</td>
<td>$\frac{1}{n^3} \frac{24}{A^2 + B^2} \sigma^2$</td>
</tr>
<tr>
<td>Walker 1973</td>
<td>AH</td>
<td>$A \cos(\omega_0 t) + \frac{B}{2} \sin(\omega_0 t) + \mu$</td>
<td>regular</td>
<td>correlated</td>
<td>$\frac{1}{n^3} \frac{24}{A^2 + B^2} f_{sp}(\omega_0)$</td>
</tr>
<tr>
<td>Hall, Reimann, Rice 2000</td>
<td>NP</td>
<td>“well behaved”</td>
<td>irregular</td>
<td>iid</td>
<td>$\frac{1}{n^3} 24\pi \sigma^2 \omega_0 \eta^{-2} \left{ \int_0^P g'(u)^2 du \right}^{-1}$</td>
</tr>
<tr>
<td>Nicoletti</td>
<td>AH</td>
<td>“well behaved”</td>
<td>irregular</td>
<td>correlated</td>
<td>$\frac{1}{n^3} \frac{24}{A_g^2 + B_g^2} f_{sp}(\omega_0)$</td>
</tr>
<tr>
<td>Nicoletti</td>
<td>NP</td>
<td>“well behaved”</td>
<td>irregular</td>
<td>correlated</td>
<td>$\frac{12}{n^3} \omega_0 \frac{1}{\eta^2} \frac{\int_{-\infty}^{\infty}</td>
</tr>
<tr>
<td>Nicoletti</td>
<td>regression with known g</td>
<td>“well behaved”</td>
<td>irregular</td>
<td>correlated</td>
<td>$\frac{3}{n^3} \omega_0 \frac{1}{\eta^2} \frac{\int_{-\infty}^{\infty}</td>
</tr>
<tr>
<td>Nicoletti</td>
<td>NP</td>
<td>“well behaved”</td>
<td>irregular</td>
<td>iid</td>
<td>$2\pi \frac{L}{n^3} \omega_0 \frac{1}{\delta^2} 24 \beta \frac{1}{\text{Int}_g} \sigma^2$</td>
</tr>
</tbody>
</table>
A simple case: a box signal

Estimation of frequency uncertainty

\[ \delta \nu \sim \frac{\sigma}{D n \Delta T} \]

Time span: \( \Delta T \)
Depth of transit: \( D \)
Number of measurements: \( n \)
Standard deviation of Noise: \( \sigma \)

A bit surprising, it does not depend on length of transit, or number of points in transit…
Standard deviation of the asymptotic distribution…
Conclusions

Clear view how to handle the estimation of the standard deviation of frequency

- Determine the “boundaries” of the validity of the asymptotic behaviour
- Tests should be done to determine for some possible corrections
- Study of the distribution of noise in Gaia (correlations)
Thank you for your attention