Multistage Search for Hidden Patterns in Irregularly Observed Time Series

Jaan Pelt,
Institute of Astrophysics and Atmospheric Physics,
Tőravere 1-6, EE2444 Estonia

Abstract

The basic ideas of multistage search for hidden patterns in time series are discussed. A particularly effective realization of general ideas is outlined in the context of hidden period determination problems.

1 Introduction

Irregularities in data measurement timings appear in astronomy due to many different causes and unfortunately, most of them cannot be influenced by the observer. When stationary irregularity (for instance time point sequences with purely random distribution in time) can be handled with ease, the actual data sets also tend to contain gap sequences with nearly regular spacing. The interplay between the structure of the underlying process and regularities superimposed by observational schedule makes the data analysis of real observational sequences rather cumbersome. We will describe a rather general method to handle such complicated data sets. The method is implemented in a software package called ISDA (Irregularly Spaced Data Analysis [9]).

2 Least Squares Spectrum

In our approach [8, 9] the parametrised analytical process model \( M(t, c_1, c_2, \ldots, c_K) \) is fitted into observed data points \( f(t_i), i = 1, 2, \ldots, N \) in a standard manner. We seek the absolute minimum of a residual sum of squares

\[
RSS = \sum_{i=1}^{N} [f(t_i) - M(t_i, c_1, c_2, \ldots, c_K)]^2. \tag{1}
\]

Typically the model \( M \) can be written as a linear sum of model components \( M_r \):

\[
M(t, c_1, \ldots, c_K) = \sum_{r=1}^{R} c_r M_r(t, c_{R+1}, \ldots, c_K). \tag{2}
\]

For any fixed set of parameter values \( c_{R+1}, \ldots, c_K \) linear parameters \( c_1, c_2, \ldots, c_R \) can be estimated by ordinary regression procedures. The \( RSS \) can then be evaluated as a function of nonlinear parameters only. We get Least Squares Spectrum (LSS)

\[
LSS(c_{R+1}, \ldots, c_K) = \sum_{i=1}^{N} (f(t_i) - M(t_i, \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_R, c_{R+1}, \ldots, c_K))^2 \tag{3}
\]

where linear parameter estimates \( \hat{c}_1, \ldots, \hat{c}_R \) are also functions of nonlinear parameters. Minima of the \( LSS \) indicate most probable combinations of nonlinear parameters. In this way the model parameters estimation problem can be converted into the problem of seeking a global minimum of \( LSS \). For semi-randomly measured data sets, which are quite typical in astronomy, the \( LSS \) is often very erratic and its analysis is far from trivial.
3 Multistage Search

Global minimisation of the LSS proceeds in a number of steps: in the first step of the analysis we seek crude approximations to the nonlinear parameter values; then, in the second step we proceed with standard grid search on a fine mesh near the estimates sought in the first stage; and finally we use traditional parameter refinement procedures to get estimates with the highest precision obtainable. When grid search and refinement of model parameters are quite traditional techniques in astronomical data processing, the pilot search for crude approximations is not so well known.

Let us consider a particular case of our general problem. The simplest LSS can be computed for a periodic model

$$LSS(\nu) = \sum_{i=1}^{N} \left\{ f(t_i) - \hat{a}_0(\nu) - \sum_{r=1}^{R} [\hat{a}_r(\nu) \cos(2\pi rv_i) + \hat{b}_r(\nu) \sin(2\pi rv_i)] \right\}^2.$$  \hspace{1cm} (4)

where the frequency $\nu$ is the only nonlinear parameter in the model. The grid search for a full parameter space must be carried out with a fixed step in the frequency

$$\nu_l = (l - 1)\Delta \nu + \nu_{\text{min}}, \quad l = 1, 2, \ldots, L, \nu_L = \nu_{\text{max}},$$  \hspace{1cm} (5)

where $(\nu_{\text{min}}, \nu_{\text{max}})$ is the range of physically possible frequencies (or those of interest to the observer). The value for the step $\Delta \nu$ must be chosen carefully. For our period seeking procedure it is reasonable to bound maximum shift of any observed point on the phase-process diagram, when moving from one grid point to the next. This is attained when

$$\Delta \nu = \frac{\Delta \varphi}{t_N - t_1},$$  \hspace{1cm} (6)

where $\varphi$ is the upper limit for allowed shifts. The number of grid points needed can be now computed

$$L = \frac{(\nu_{\text{max}} - \nu_{\text{min}})(t_N - t_1)}{\Delta \varphi}.$$  \hspace{1cm} (7)

The number of grid points $L$ can be rather large, even for the modest data sets. For instance, when $t_N - t_1 = 2000$ (a typical value for astronomical photographic plate collections), $\Delta \varphi = 0.1$, $\nu_{\text{min}} = 2.0, \nu_{\text{min}} = 5.0$ we get $L = 60000$. (The time is expressed in days, frequencies are measured in cycles per day.) Solving least squares estimation equations for 60000 grid frequencies can take enormous time, and this is the reason why we must narrow the search range by some pilot search procedure.

From the formula for grid length we can conclude that the number $L$ can be reduced by shortening the search range $\nu_{\text{max}} - \nu_{\text{min}}$ or by reducing the data range $t_N - t_1$. The best results are obtained by doing both: reducing data range on the first stage of the search and reducing the search range on the second stage.

Here is the strategic plan for a simple multistage search:

- Divide the data range into $k$ subranges, with a maximal length of $\Delta t$

$$L_{\text{crude}} = \frac{(\nu_{\text{max}} - \nu_{\text{min}})\Delta t}{\Delta \varphi}.$$  \hspace{1cm} (8)

- Compute $LSS$ for every subrange.
- Compute the average $LSS$.
- Search the averaged $LSS$ to obtain crude estimates of probable frequencies.
• Compute fine spectra in the small ranges (with width) $\Delta \nu$ around the crude estimates

$$L_{\text{fine}} = \frac{M\Delta \nu(t_N - t_i)}{\Delta \varphi}$$

(9)

where $M$ is the number of crude estimates.

• Refine iteratively computed frequencies by the standard Levenberg-Marquardt iteration procedure [10] to get estimates with the maximum attainable precision.

By appropriate choice of the parameters $\Delta t$, $\Delta \nu$ and $M$, we can reduce the total amount of our computations significantly.

The simplest multistage search techniques were first described by Evans [4] but have afterwards been more or less neglected. It is much more popular to compute long spectra on powerful supercomputers, than to apply complex multistage algorithms.

4 Nonparametric Dispersion Estimation

The scheme, proposed in the last section, involves grid search for data subsets with further averaging of local spectra. But there is a much simpler procedure to get good preliminary estimates for significant frequencies. We start from the observation that dispersion around the mean curve on the phase-process diagram can be computed without exact estimation of the mean light curve itself. The simple nonparametric statistic

$$D(\nu) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} g(t_i, t_j, \nu)[f(t_i) - f(t_j)]^2}{2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} g(t_i, t_j, \nu)}$$

(10)

where $g(t_i, t_j, \nu)$ is significantly greater than zero only when

$$t_i - t_j \approx \frac{k}{\nu}, k = 0, \pm 1, \pm 2, \ldots,$$

(11)

well approximates $LSS(\nu)$ spectrum. The nice property of $D(\nu)$ spectrum is, that it can be smoothed by introducing an additional data window $W(t_i, t_j)$ into the general formula

$$D(\nu) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} g(t_i, t_j, \nu)W(t_i, t_j)[f(t_i) - f(t_j)]^2}{2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} g(t_i, t_j, \nu)W(t_i, t_j)}.$$  

(12)

The degree of smoothing obtained by the use of the data window

$$W(t_i, t_j) = \begin{cases} 1, & |t_i - t_j| \leq D_{\text{max}}, \\ 0, & \text{otherwise}. \end{cases}$$

(13)

depends on the algorithm parameter $D_{\text{max}}$. When $D_{\text{max}} = t_N - t_1$, then all the observation pairs are taken into account when computing $D(\nu)$, and maximal resolution is attained. For smaller values of $D_{\text{max}}$, only a subset of pairs is considered and the computed spectrum is correspondingly flatter. It allows a choice of longer frequency step for trials in the first search stage, and consequently makes computations faster. In standard Fourier spectrum analysis the spectral window is used analogically.

From the first sight the $D(\nu)$ statistic seems to be computationally complex, because it involves double summation for every frequency point. However, simple rearrangement of sums and use of trigonometric approximations allows computing $D(\nu)$ very effectively.
The values $t_i - t_j$ can be binned into a range of approximate $k \delta t$ values, for fixed $\delta t$ so that for every pair:

$$t_i - t_j \approx k \delta t, \text{ for some } k. \quad (14)$$

For every bin, the partial sums can then be computed:

$$C_k = \sum_{t_i - t_j = k \delta t} [f(t_i) - f(t_j)]^2 \quad (15)$$

and

$$S_k = \sum_{t_i - t_j = k \delta t} 1. \quad (16)$$

The sums $C_k$ and $S_k$ do not depend on the frequency, so we need compute them only once. The simple weights $g(t_i, t_j, \nu)$ can be chosen

$$g(t_i, t_j, \nu) = d((t_i - t_j)\nu) \approx d(k\delta t\nu), \quad (17)$$

where $d(x)$ is a periodic (with period 1) rectangular pulse sequence. Then, $D(\nu)$ can be computed from:

$$D(\nu) = \frac{\sum_{k=0}^{K} d(k\delta t\nu) C_k}{2 \sum_{k=0}^{K} d(k\delta t\nu) S_k}. \quad (18)$$

Some of the sums $C_k$ and $S_k$ are zeros, because there are no appropriate pairs of observations available in data set, and they can be omitted.

It is possible to use more complicated binning procedures which use so called extirpation [11] but our experience shows that for most of the cases there is no need for this, the errors in "nearest neighbour" binning procedure cancel out statistically and corresponding spectra does not differ from spectra computed with more complicated procedures.

Now the pulse sequence $d(k\delta t\nu)$ can be expanded into the Fourier series:

$$d(k\delta t\nu) = \sum_{r=0}^{\infty} d_r \cos (2\pi r k \delta t \nu), \quad (19)$$

and the computation of $D(\nu)$ can be reduced to the computation of appropriate trigonometric sums. The Fourier expansion can be truncated at some low number of $r$ (typically $r = 4$), and the corresponding sums can be computed via the Fast Fourier Transform. Possible shortcuts can be used to reduce the number of machine operations: all sums for denominator and divisor can be computed with one complex Fourier transform (see TOLFFT algorithm in [10]); the statistic values for intermediate grid points can be computed through interpolation and so on. There are many fine algorithmic details which can be followed best from actual program listings which can be obtained from the author.

As a result, the use of $D(\nu)$ statistic in the first stage of the period search speeds up the whole procedure significantly. In fact the computation of the rather complex statistic $D(\nu)$ for a full grid of frequencies takes no more time than simple Fast Fourier transforming of the input array whose length is comparable to the number of points in the frequency grid.

5 Generalizations

The above scheme can easily be generalised in many directions. For instance, if the underlying process is too complicated to be described by a single periodic process model, we can use multiple
periodic processes as models. For double periodic process the smoothed version of the $D(\nu_1,\nu_2)$ statistic

$$D(\nu_1,\nu_2) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} W(t_i,t_j)g(t_i,t_j,\nu_1)g(t_i,t_j,\nu_2)[f(t_i) - f(t_j)]^2}{2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} W(t_i,t_j)g(t_i,t_j,\nu_1)g(t_i,t_j,\nu_2)}$$ \hspace{1cm} (20)

can be used to seek for preliminary crude estimates for $(\nu_1,\nu_2)$, and then a grid search on small rectangular grids around a sample of the best frequency candidates can be carried out. It is also possible to introduce slight changes into model frequencies, the change rate is then the additional parameter in the model. In the ISDA package we have implemented many variations of this kind.

6 Significance estimation

The multistage search algorithm itself does not provide any methods to estimate significance of strongest minima in LSS. For that purpose any known technique (see [7, 3, 6] or any handbook on regression or dispersion analysis can be used). However, there is one interesting application of the $D(\nu)$ type smoothed statistics. It is possible, that the $D(\nu)$ spectrum shows strong minima, but a spectrum which is computed on the grid with full resolution does not contain any strong peaks. Consequently, the data contains cycles which are coherent only on short timescales. We can plot depth of the strongest minimum against a smoothing parameter $D_{\text{max}}$ and use this diagram to estimate typical coherence length. There is no need to say that certain resemblance with popular multiresolution analysis methods is apparent (see [2]).

7 Discussion

One of the main deficiencies of the data modeling by the least squares fitting is lack of criteria, when choosing model components. For nested models (e.g. trigonometric polynomials with increasing degree) it is possible to use various whiteness tests [5], or information criteria [1] and sometimes the so called cross validation schemes can be helpful. In general the subjective judgement of the researcher is decisive. Multistage schemes allow to compute alternative partial or full solutions very fast and consequently such schemes can be used in interactive environments. A researcher can now visually inspect different solutions to get an idea about the adequacy of models.

We strongly advocate for inclusion of multistage type least squares fitting algorithms into general astronomical data processing packages.

References


