Spectral analysis of unevenly sampled signals: an effective alternative to the Lomb-Scargle periodogram

R. Vio\textsuperscript{1} and P. Andreani\textsuperscript{2}

\textsuperscript{1} Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy
e-mail: robertovio@tin.it
\textsuperscript{2} ESO, Karl Schwarzschild strasse 2, 85748 Garching, Germany
e-mail: pandrean@eso.org

Received .............; accepted ................

ABSTRACT

The detection of signals hidden in noise is one of the oldest and common problems in astronomy. Various solutions have been proposed such as the parametric approaches based on the least-squares fit of theoretical templates or the non-parametric techniques as the phase-folding method. Most of them, however, are suited only for signals with specific time evolution. For generic signals the spectral approach, based on the periodogram, is potentially the most effective. In astronomy the main problem in adopting such an approach is that often the sampling time grid of the signals is irregular. This complicates the efficient computation of the periodogram (the fast Fourier transform cannot be used) and the determination of its statistical characteristics. The Lomb-Scargle periodogram (LSP) solves this last important issue. However, the weakness of this technique is that it is neither intuitive nor transparent. Moreover, the LSP provides a distorted version of the true periodogram. This results in theoretical as well practical issues with no easy solution. In this paper, we propose an alternative approach which has the advantage to work with the true periodograms and hence it is easier to deal with from both the theoretical and the practical point of view.

Key words. Methods: data analysis – Methods: statistical

1. Introduction

The detection of weak signals in noise requires a careful analysis of the data with appropriate statistical tools. Given the simplicity of its use, one of the most popular approaches is the spectral analysis of the time series by means of the periodogram technique. In astronomical applications the use of this technique is not straightforward. In fact, this tool exhibits its optimal properties only when the signals are sampled on a regular time grid, an uncommon situation in astronomical observations. The analysis of a periodogram in the case of irregular sampling is often limited by the possibility to fix completely its statistical properties. This problem is not new (see e.g. Gottlieb et al. 1975) and there have been many attempts to solve it. The solution found in the Lomb–Scargle periodogram (LSP) approach (Lomb 1976; Scargle 1982) comes at the price of theoretical difficulties that make its use unclear and not obvious. This is because the Lomb–Scargle method provides a modified version of the true periodogram. As a consequence, over the years a lot of papers have been dedicated to understand and solve the resulting shortcomings (some recent works are Reegen 2007; Zecharmeier & Kürster 2009; Vio et al. [2010], [2013]; VanderPlas 2018 and reference therein). In this paper, we provide an alternative approach to the LSP which works with the unmodified periodogram and hence it is easier to deal with from both theoretical and practical point of view.

In Sect. 2 we discuss the limitations of the LSP and in Sect. 3 the approach proposed by us is detailed. In Sect. 4 a couple of numerical experiments are presented. Finally, Sect. 5 derives our conclusions.

2. Mathematics of the problem

When a continuous signal $x(t)$ is sampled on an irregular time grid $[t_0, t_1, \ldots, t_{M-1}]$, a time series $[x_{t_0}, x_{t_1}, \ldots, x_{t_{M-1}}]$ is obtained. The periodogram of $[x_t]$ is defined as

$$p_\nu = a_c^2 + b_c^2$$

where

$$a_c = \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} x_i \cos 2\pi \nu t_i,$$

$$b_c = \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} x_i \sin 2\pi \nu t_i.$$
\( \sigma_n \), it can be readily verified that, independently of \( k \), \( p_k/\sigma_n^2 \) is given by the sum of two squared independent, zero-mean, unit-variance, Gaussian random quantities. As a consequence, the corresponding probability density function (PDF) is the exponential distribution,

\[
g_{p_k}(p_k) = \exp(-p_k).
\]

(4)

Moreover, whenever \( k \neq k' \) with \( k, k' = 0, 1, \ldots, M/2 \), \( p_k \) is independent of \( p_{k'} \). Hence, the probability of false alarm (PFA) \( \alpha \), that is the probability \( \alpha \) that, at a preselected frequency \( k \), \( p_k \) is expected to exceed a level \( z \), is

\[
\alpha = 1 - \left[ G_{p_k}(z/\sigma_n^2) \right]^{N^*} ;
\]

\[
= 1 - \left[ 1 - e^{-z/\sigma_n^2} \right]^{N^*},
\]

(5)

(6)

with \( G_{p_k}(p_k) \) the cumulative distribution function (CDF) of the exponential distribution and \( N^* \) the number of statistically independent frequencies in the periodogram. Through this quantity it is possible to fix a detection threshold,

\[
L_{E_k} = -\sigma_n^2 \ln \left[ 1 - (1 - \alpha)^{1/N^*} \right],
\]

(7)

corresponding to the level that one or more peaks due to the noise would exceed with a prefixed probability \( \alpha \) when a number \( N^* \) of (statistically independent) frequencies are inspected. Typically \( N^* = M/2 \). Threshold \( L_{E_k} \) is called the "level of false alarm".

When the sampling is irregular, the situation becomes more complex given that \( a_i/\sigma_n \) and \( b_i/\sigma_n \) are still zero-mean, Gaussian quantities, but they are not correlated and no longer of unitary variance. Hence, for a fixed frequency, the PDF of \( p_\nu \) is not \( g_{p_\nu}(p_\nu) \). Lomb (1976) and Scargle (1982) bypass this problem introducing a modified version of the periodogram

\[
\hat{p}_\nu = \frac{1}{2} (\hat{a}_\nu^2 + \hat{b}_\nu^2)
\]

(8)

where

\[
\hat{a}_\nu = \sum_{m=0}^{M-1} x_\nu \cos \left[ 2\pi n(t_m - \tau) \right],
\]

\[
\sqrt{\sum_{m=0}^{M-1} \cos^2 \left[ 2\pi n(t_m - \tau) \right]},
\]

\[
\hat{b}_\nu = \sum_{m=0}^{M-1} x_\nu \sin \left[ 2\pi n(t_m - \tau) \right],
\]

\[
\sqrt{\sum_{m=0}^{M-1} \sin^2 \left[ 2\pi n(t_m - \tau) \right]},
\]

(9)

(10)

with \( \tau \) defined by

\[
\tan(2\pi \nu \tau) = \frac{\sum_{m=0}^{M-1} \sin \left( 4\pi n(t_m) \right)}{\sum_{m=0}^{M-1} \cos \left( 4\pi n(t_m) \right)}.
\]

(11)

The reason for such a modification is that, under the hypothesis \( x_\nu = n_\nu \), \( \hat{a}_\nu \) and \( \hat{b}_\nu \) are zero-mean, unit variance, uncorrelated Gaussian random quantities. Therefore, the PDF of \( \hat{p}_\nu \) is again of exponential type.

A further problem which raises in the case of irregular sampling is that it is no longer possible to define a set of natural frequencies (such as the Fourier frequencies) for which to compute the periodogram. Hence, there is no reason why the number \( N \) of frequencies where to compute \( \hat{p}_\nu \) is equal to the number \( M \) of the sampling time instants. Indeed, often \( N \gg M \). Here, the point is that the number \( N^* \) of independent frequencies is not defined, hence the threshold \( L_{E_k} \) (see eq. 7) is not computable. However, as discussed in Press et al. (1992), the exact value of \( N^* \) is not critical and in many situations \( N^* = M/2 \) represents a reasonable choice (e.g. see Vio et al. 2010, 2013).

Although with \( \hat{p}_\nu \) the issue of the PDF of periodogram under the hypothesis \( x_\nu = n_\nu \) has been solved, the proposed solution is a bit tortuous and not immediately obvious. Moreover, \( \hat{p}_\nu \) constitutes a distorted version of \( p_\nu \). This is not desirable from the theoretical as well the practical point of view. In the following, it is shown as it is possible to work directly with \( p_\nu \), avoiding these problems.

3. An alternative to the Lomb-Scargle periodogram

As explained above, the main issue in the computation of the PDF of \( p_\nu \) lies in the correlation between \( a_\nu \) and \( b_\nu \). However, although not well known, the PDF \( f_\nu(z) \) of \( Z = Y_1^2 + Y_2^2 \) with \( Y_1 \) and \( Y_2 \) zero-mean, Gaussian random variables with standard deviation \( \sigma_1 \) and \( \sigma_2 \), respectively, and correlation coefficient \( \rho \) is available (Simon 2006)

\[
f_\nu(z) = \frac{1}{2\sigma_1\sigma_2 \sqrt{1-\rho^2}} \exp \left[ -\frac{1}{4} (\beta^* - \gamma)^2 \right] I_0 \left( \frac{1}{4} \gamma^* z \right), \quad z \geq 0
\]

(12)

with

\[
\gamma = \frac{[\sigma_1^2 + \sigma_2^2]^3 - 4\sigma_1^2 \sigma_2^2 (1 - \rho^2)]^{1/2},
\]

\[
\beta^* = \gamma + \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)},
\]

(13)

(14)

and \( I_0(.) \) the modified Bessel function of the first kind of zero order. The corresponding central moments are given by

\[
E \left[ (Z - Z) \right] = \sum_{k=0}^{k} k! \left[ \frac{k}{k+1} \right] \left[ \frac{k}{k+1} \right] E[Z].
\]

(15)

where \( E[Z] \)

\[
E[Z] = \frac{2^{2k+1} k!}{\sigma_1 \sigma_2 \sqrt{1 - \rho^2} (\beta^* - \gamma)^{k+1}}
\]

\[
2F1 \left[ \frac{k+1}{2}, \frac{k}{2} + 1; 1; \left( \frac{\gamma}{\beta^* - \gamma} \right)^2 \right],
\]

(16)

with \( 2F1(...) \) the Gauss hypergeometric function. Cases of interest are \( k = 1 \) and \( 2 \) for which

\[
2F1 \left[ \frac{k+1}{2}, \frac{k}{2} + 1; 1; \frac{\gamma}{\beta^* - \gamma} \right] = \frac{1}{\left( \frac{1}{2} + \gamma \right)^{k+1}} \left( \frac{1}{2} + \gamma \right)^{k+1} \left( \frac{\gamma}{\beta^* - \gamma} \right)^{k+1}
\]

(17)

The cumulative distribution function \( F_\nu(z) \) is also available,

\[
F\nu(z) = 1 + \exp \left[ -\frac{1}{4} (\beta^* - \gamma)^2 \right] I_0 \left( \frac{1}{4} \gamma^* z \right) - 2Q_1(A, B),
\]

(18)

where \( Q_1(...) \) is the Marcum Q-function with

\[
A = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \sqrt{\gamma},
\]

\[
2\sigma_1 \sigma_2 \sqrt{1 - \rho^2}
\]

(19)

(20)

1 Symbol \( E[.] \) denotes the expectation operator.
Although the functions which appear in these equations could look a bit exotic, actually they are available in all the main software packages.

In the present case, it is \( z = p_v, \sigma_1 = \sigma_{a_1}, \sigma_2 = \sigma_{b_1}, \rho = \rho_v \) with

\[
\sigma_{a_1}^2 = \frac{\sigma^2_n}{M} \sum_{i=0}^{M-1} \cos^2 (2\pi vt_i),
\]

\[
\sigma_{b_1}^2 = \frac{\sigma^2_n}{M} \sum_{i=0}^{M-1} \sin^2 (2\pi vt_i),
\]

and \([\text{Vio et al. 2013}]\)

\[
\rho_v = \frac{E[a_i b_i]}{\sigma_{a_i} \sigma_{b_i}} = \frac{\sum_{i=0}^{M-1} \sin (4\pi vt_i)}{2 \left( \sum_{i=0}^{M-1} \cos^2 (2\pi vt_i) \right) \left( \sum_{i=0}^{M-1} \sin^2 (2\pi vt_i) \right)}.
\]

The main advantage of working with \( f_p(p_v) \) and \( F_p(p_v) \) is that the computation of the statistical significance of a peak in the periodogram does not require that this last be modified frequency by frequency as it happens if Eqs. \((9)\) and \((10)\) are used. This makes the detection procedure more transparent and intuitive. Another benefit concerns the spectral window \( w(\nu) \) which, strictly speaking, cannot be computed in the context of the LSP method. We recall that the spectral window, defined as

\[
w(\nu) = \sum_{i=0}^{M-1} e^{-2\pi i \nu t_i},
\]

with \( \nu = \sqrt{\nu_1 \nu_2} \), is key since all spectral leakage effects (aliasing, sidelobes, interference phenomena, ghosts, etc.) are manifested directly in it and can be easily evaluated quantitatively \([\text{Deeming 1975}]\) \([\text{Scargle 1982}]\).

The proposed method is inferior to the LSP in only one aspect: the detection threshold is not available in an analytical form and hence it must be numerically computed by solving the integral equation

\[
\int_{p_{\nu}}^{\infty} f_p(p_v) dp_v = \alpha
\]

for \( L_p \). However, an effective alternative is to fix a threshold value \( \alpha \) for the PFA, to compute the quantities \( \{a_i\} \) for the highest peaks \( \{p_{\nu_i}\} \) in the periodogram,

\[
a_i' = 1 - F_{p_{\nu_i}}(p_{\nu_i}),
\]

and then to claim the \( i \)-th peak statistically significant if \( a_i' \leq \alpha \).

4. An example

Figures \((12)\) show the results concerning two numerical experiments where \( 10^5 \) realizations of a discrete zero-mean, unit-variance, Gaussian white-noise process are simulated with \( t_i = (m \times 5 + j)/205, i = j + m/2 \) with \( j = 1,2,\ldots,5, \) and \( m = 0,2,\ldots,40 \). Each time series contains 105 data with sampling time instants in the range \([0,1]\). Moreover, every sequence of five time instants where an observation is available is followed by a sequence of five time instants with no data. In this way, a time series with periodic gaps is simulated. Two frequencies are considered, \( \nu = 0.05 \) and \( \nu = 1.00 \). As it is visible in the top-left panel of these figures, for \( \nu = 0.05 \) the correlation between the coefficients \( a_i \) and \( b_i \) is rather strong contrary to what happens for \( \nu = 1.00 \). In the first case, the top-right and the bottom-left panels of Fig. \((1)\) show that \( f_p(p_v) \) and \( F_p(p_v) \) are different from the corresponding exponential counterpart \( g_p(p_v) \) and \( G_p(p_v) \), but only the first two are in good agreement with the respective empirical estimators of the PDF and the CDF of the simulated \( p_v \). The bottom-right panel in the same figure shows that the relative difference between the PFA’s provided by the two CDF’s is rather strong. As the corresponding panels in Fig. \((2)\) show, these differences disappear for the frequency \( \nu = 1.00 \).

These results are related to the distribution of the angles \( \theta_i = 4\pi vt_i \) of a unit circle. Indeed, from Eq. \((25)\) it results that the correlation \( \rho_v \) between the coefficients \( a_i \) and \( b_i \) at a given frequency \( \nu \) is proportional to \( \sum_{i=0}^{M-1} \sin (4\pi vt_i) \). Since the sinus function is an odd function, one may expect that \( \rho_v \approx 0 \) if the angles \( \{\theta_i\} \) of a unit circle are uniformly and/or symmetrically distributed. The different distributions on the unit circle of the angles \( \{\alpha_i\} \) corresponding to the two frequencies \( \nu = 0.05 \) and \( \nu = 1.00 \) are visible in Fig. \((3)\). It is clear that for the frequency \( \nu = 0.05 \) the distribution is asymmetric contrary to that corresponding to the frequency \( \nu = 1.00 \). Similar experiments with other kinds of sampling time grids provide identical results.

5. Conclusions

The Lomb-Scargle periodogram (LSP) has the great merit to have provided a rigorous statistical solution to the problem of the detection of signals embedded in noise when the sampling is irregular. However, the adopted procedure is tortuous, not immediately obvious and introduces some theoretical difficulties which makes its use difficult in certain steps of the detection procedure (e.g. the computation of the spectral windows) as well as in the development of fast algorithms to compute the periodogram itself. In this paper, we propose an alternative approach that is superior to the LSP both theoretically and computationally. For this reason, the use of LSP in the case of unevenly sampled signal is no longer necessary.

References

Deeming, T.J. 1975, Ap. Space. Sci., 36, 137
Gottlieb, E.W., Write, E.L., & Liller, W. 1975, AJL 195, L33
Lomb, N.R. 1976, Ap. Space. Sci., 39, 447
Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. 1992, Numerical Recipes in Fortran (Cambridge: Cambridge University Press)
Reegen, P., 2007, A&A, 467, 1353
Scargle, J.D. 1982, Ap.J., 263, 835
VanderPlas, J.T. 2018, ApJ Supp. Series, 236, 16
Simon, M.K. 2006, Probability Distributions Involving Gaussian Random Variables (New York: Springer)
Vio, R., Andreani, & P., Biggs, A., 2010. A&A, 519, A85
Vio, R., Diaz-Trigo, M., & Andreani, P. 2013. Astronomy and Computing, 1, 5
Zechmeister, M., & Kürster, M., 2009. A&A, 496, 577

Acknowledgements.
Fig. 1. Top-left panel: coefficients $a_\nu$ vs. $b_\nu$, when $\nu = 0.05$, for the numerical experiment where $10^5$ zero-mean, unit-variance, Gaussian white-noise processes are simulated on a time sampling grid containing periodic gaps (see the text). To notice the strong correlation between these two quantities. Top-right and bottom-left panel: the probability density function (PDF) $f_{p_\nu}(p_\nu)$ and the cumulative distribution function (CDF) $F_{p_\nu}(p_\nu)$ of the values of the periodogram $p_\nu = a_\nu^2 + b_\nu^2$. For reference, the exponential PDF $g_{p_\nu}(p_\nu)$ and CDF $G_{p_\nu}(p_\nu)$ are plotted as well as the histogram and the empirical CDF. The almost perfect overlapping of $f_{p_\nu}(p_\nu)$ and $F_{p_\nu}(p_\nu)$ with the corresponding empirical functions is clear. The same is not true for $g_{p_\nu}(p_\nu)$ and $G_{p_\nu}(p_\nu)$. Bottom-right panel: relative difference between the probabilities of false alarm (or false detection) $(\text{PFA} - \text{PFA}_\text{exp})/\text{PFA}$, where PFA = $1 - F_{p_\nu}(p_\nu)$ and PFA$_\text{exp}$ = $1 - G_{p_\nu}(p_\nu)$. 

Article number, page 4 of 6
Fig. 2. The same as in Fig. 1 but for $\nu = 1.00$. Here, $g_{P_\nu}(p_\nu)$ and $G_{P_\nu}(p_\nu)$ are almost indistinguishable from $f_{P_\nu}(p_\nu)$ and $F_{P_\nu}(p_\nu)$, respectively.
Fig. 3. Distribution of the angles $\theta_i = 4\pi \nu t_i$ on the unit circle corresponding to a set of different frequencies $\nu$ for the numerical experiments described in the text. The circles corresponding to the frequencies $\nu = 0.05$ and 1.00 are to be related to the results shown in Figs. 1 and 2.