FALSE ALARM PROBABILITY IN THE MULTIPERIODICITY SEARCH

J. Molenda-Żakowicz[1]

[1] Institute of Astronomy, Wrocław University
ul. Kopernika 11, 51-622 Wrocław, Poland, *molenda@astro.uni.wroc.pl*

ABSTRACT. We apply the False Alarm Probability analysis, (FAP), to the multiperiodicity search. Then we show the necessity of using the FAP method in the analysis of the astronomical time-series. We present the results obtained for 153 stars supposed or known to be pulsating variables. We examine the statistical properties of the excited frequencies and find a relation between the parameters of the fitted sine-curves and the FAP. Finally we show the application of our results to the individual stars and large samples of stars.

Key words: Stars: variables; statistics: numerical methods.

1. Introduction

Testing significance of periodigram peaks is very important. Identyfying a frequency we should be sure that we do not take stochastical fluctuations of noise as an intrinsic variability. In our analysis we examined the Hp magnitudes of 153 stars observed by the *Hipparcos* satellite (ESA 1997). All of the stars are situated in the domain of instability of Slowly Pulsating B stars. We computed the least–squares power spectra as described in Jerzykiewicz & Pamyatnykh (2000), using the Lomb (1976) method modified by adding a floating mean and weights calculated for each Hp magnitude. Then we prewhitend data fitting simultaneously all frequencies found so far. The False Alarm Probability method described by Cumming et al. 1999 was used by us to test significance of periodogram peaks. We generated hundrets sets of gaussian noise and compared the height of the highest peaks produced by the noise with the height of the highest peak, $z_{\rm max}$, obtained for the stellar data.

Having derived all excited frequencies, we examined the distributions of A_i/σ_{A_i} obtained for all the stars. We analysed the statistical properties of the distributions and compared the results obtained for the *Hipparcos* observations with the results obtained for pure gaussian noise and for gaussian noise with the sinusoidal signal added.

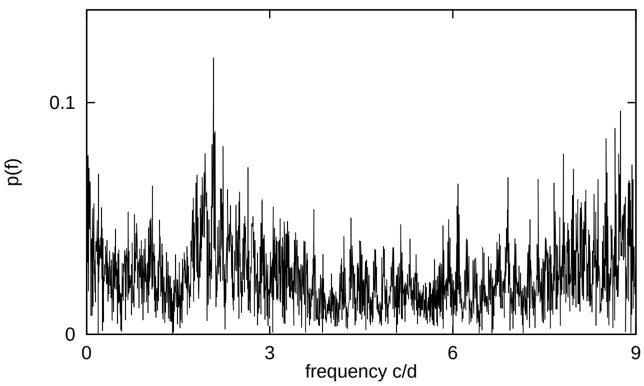

Figure 1: A periodogram of pure gaussian noise.

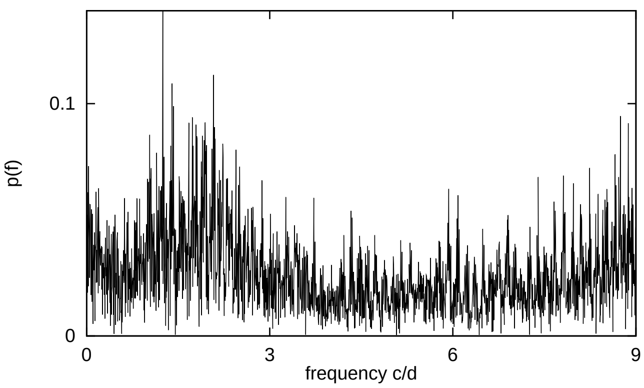

Figure 2: A periodogram of the pure gaussian noise from Fig 1 with a sine–curve added. The highest peak indicates the hidden periodicity.

2. Testing significance of periodogram peaks.

It is well–known that noise can give rise to high periodogram peaks and fluctuations of noise can mimic periodic variations. To show how high such noise–originated peaks can be, we show two periodograms in Figs 1 and 2. In Fig 1 we can see a periodogram of pure gaussian noise while in Fig 2 we can see a periodogram of the same noise with a sine–curve added. In both figures the highest peaks, $z_{\max}$, have comparable height. When we analyse data which consists of little number of observations scattered over hundrets of days, as in the case of *Hipparcos* observations, we often face situation when fluctuations of noise are comparable with the supposed variability. Therefore for stars with small–amplitude variations it is easy to miss a hidden periodicity taking it as fluctuations of the noise.

Most of the statistical tests are not applicable in astronomy. The majority of the tests are devised for equally spaced observations containing noise or noise with a *known* number of periodic signal present. While the stellar observations are usually unequally spaced and contain huge gaps. The False Alarm Probability (FAP) method used by us gives the probability that the highest periodigram peak, $z_{\max}$, is due to the noise–fluctuations. The adventage of the FAP method is that FAP does not require equally–spaced observations and can be applied to all data with the know distribution.

In case of gaussian distribution, which is the most common in astronomy, we are able to generate sets of gaussian noise with the same mean and variance as the original observations. Computing the power–spectra we can compare the height of peaks computed for the noise, with the height of $z_{\max}$ computed for the original observations. When we express in percents the number of peaks which are higher than $z_{\max}$, we obtain the probability (FAP) that we deal with pure noise. Using the FAP method we obtain information of presence of not periodic variations which produce more power than the pure noice could do. We can also know whether the high peaks in the low frequency region indicate a changing mean of the time–series or are rather due to fluctuations of the noise.

3. Analysis of individual stars.

As the FAP method requires knowledge of the distribution of observations we used the Lilieforse–Kolmogorov and the Shapiro–Wilk statistical tests to find stars with gaussian distribution of observations. We used the STATISTICA software package which provides both tests. We found that over 50% of the distributions were accepted by both tests as gaussian at the $\alpha = 0.05$ level, where α is the probability that we uncorrectly classify a distribution as non–gaussian. For these stars we performed the periodogram analysis and, having found the frequency of the hidden periodicity, we fitted the data with a sine–curve and obtained the amplitude of the variation, $A_1 \pm \sigma_{A_1}$. Then we performed the FAP analysis accepting a periodicity as real only if the FAP was less then 1%. Seeking for the next frequencies we repeated all the described procedure.

Then we found that FAP is strongly correlated to the value of A_i/σ_{A_i}, where i is the number of a derived frequency (Molenda–Żakowicz 2000) and found this correlation as a usefull tool for testing significance peaks for stars with non–gaussian distribution of data. We show this correlation in Fig. 3 where we denoted A_1/σ_{A_1} with dots, A_2/σ_{A_2} with squares, A_3/σ_{A_3} with triangles and A_4/σ_{A_4} with crosses. One can see that for each next frequency the threshold where FAP becomes greater than 1% is higher. We list in Table 1 the thresholds accepted by us for the first four periodicities detectable in the *Hipparcos* data.

Table 1: The accepted thresholds for A_i/σ_{A_i}.

number of frequency, i	threshold for A_i/σ_{A_i}
1	5.8
2	6.2
3	7.0
4	7.8

When dealing with stars with non–gaussian distribution of observations, we assume that the general population of the observations was gaussian and the observed irregularities were due to sampling. Then we performed the periodogram analysis and accepted only these frequencies for which the values of A_i/σ_{A_i} were greater than the accepted thresholds. Because there was some scatter present around each thershold, we set the thersholds high enough to protect us from considering fluctuations of noise as a hidden periodicity.

2. Analysis of large samples of stars

In this section we pay attention to large samples of stars. We defined a sapmle as large when the number of stars exceeds 30 and ananysed the distribution of the values of A_i/σ_{A_i} of constant, mono– and multiperiodic stars in order to examine the statistical properties of these distributions.

We studied samples of stars denoted as constant in the *Hipparcos Catalogue* and mono– and multiperiodic stars extracted from the 153 stars examined by us. In all cases we obtained skewed distributions of A_i/σ_{A_i} (in case of constant stars we analysed distribution of A_i/σ_{A_i} obtained by fitting data with a sine–curve with the frequency indicated by the highest peak of the periodogram). This result is in a full agreement with the theoretically predicted distribution of extreme values (Sachs 1984 and references therein). The new result

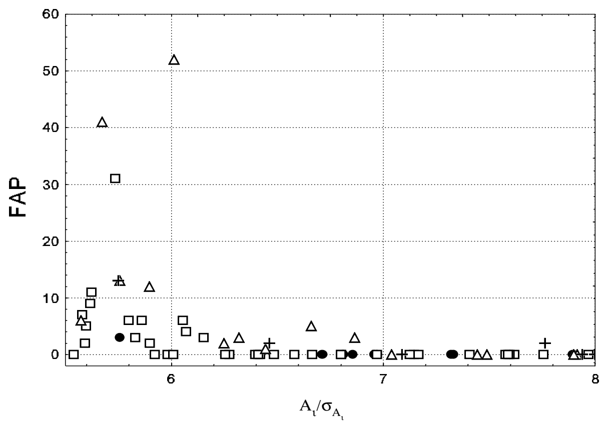

Figure 3: The correlation of A_i/σ_{A_i} and FAP for the first four frequencies detectable in the *Hipparcos* observations. We denoted $i = 1$ with asterisks, $i = 2$ with triangles, $i = 3$ with squares and $i = 4$ with dots.

is a dependence of the skewness coefficient, γ, on the amount of multiperiodic stars in the sample. The skewness coefficient, γ, is defined as

$$\gamma = \frac{\mu_3}{\sigma^3},$$

with

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

where μ_3 is the third central moment, x_i is the observed magnitude, $\bar{x}$ is the mean magnitude derived from the data, n is the number of observations, and σ is the standard deviation of the analysed time–series.

We found that the distribution of A/σ_A obtained for constant stars was steep on the left and flat to the right as we expected. For all the other distributions we observed cutting of the left tail. This cutting arises because of the observational selection: the histograms obtained for variable stars were limited from the left side to these values of A_i/σ_{A_i} for which FAP was less than 1%. This result is an important indication that we still miss many low–amplitude periodicities looking for multiperiodic stars.

Analysing the dependence of γ on the amount of multiperiodic stars in the sample we found that the higher is the value of γ, the higher is the amount of multiperiodic stars. For samples containing only constant stars we obtained values of γ from the range 0.6–0.9 while for samples containing a mixture of mono– and multiperiodic stars, γ was much higher. We preformed the same analysis for pure gaussian noise and noise with sinusoidal signal added and obtained a full agreement with the results for *Hipparcos* observations. In this way we showed that not only the cutting of the distribution of A_i/σ_{A_i} but also its skewness gives information of the non–detected periodicities hidden in data.

3. Conclusions

Basing on our results we reccomend the FAP analysis as a tool that should be used to testing periodogram peaks of mono– and multiperiodic stars. Our work gave two important results. The first was the relation of A_i/σ_{A_i}–FAP which should be used for testing significance of periodogram peaks when it is not possible to compute FAP directly. The second result was the relation between γ and the amount of multiperiodic stars in the sample. This relation, however not so strict as the first one, can be used for estimating the number of not–detected variables present in large data-bases.

Acknowledgements. We thank Prof. M. Jerzykiewicz and Dr R. Getko for their help. We acknowledge the the KBN grant 5 *P*03*D* 014 20.

References

Cumming A., Marcy G.W., Butler R.P.: 1999, *ApJ*, **526**, 890

ESA 1997 The *Hipparcos* and *Tycho* Catalogues ESA SP-1200, Nordvijk: ESA

Jerzykiewicz M., Pamyatnykh A.A.: 2000, *PASP*, **112**, 1341

Koen C.: 1990, *Ap.J.*, **348**, 700

Lomb N.R.: 1976, *Ap.Space Sci.*, **39**, 447

Molenda–Żakowicz J.: 2000, *Hipparcos and the Luminosity Calibration of the Nearer Stars*, XXIV IAU Symp., Manchester, England

Sachs L.: 1984, *Applied Statistics. A Handbook of Techniques. Sec. Ed.*, Springer-Verlag, New York Inc.

STATISTICA For Windows 95/98/2000 Version 5.5, 1984-2000 StatSoft Inc.