False-Alarm Probabilities in Period Searches: Can Extreme-Value Distributions be of Use?

Jan Cuypers

Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium
email: `jan.cuypers@oma.be`

Abstract. Results of simulating false-alarm probabilities in irregularly sampled time series are presented. Relations to well-known expressions and earlier-used criteria are shown and tested for applicability. The use of an extreme-values distribution in this context is investigated.

Keywords. methods: statistical

1. Introduction

False-alarm probability (FAP) in period (frequency) search is defined as the probability that a value resulting from a time-series analysis based on some form of periodogram is caused by noise. Since knowing the significance of a peak in periodogram analysis is essential, this probability has to be determined accurately.

False-alarm probabilities for many periodograms can be described by β distributions (Schwarzenberg-Czerny, 1998; Frescura *et al.*, 2008; Baluev, 2008) having the form:

$$1 - (1 - (1 - \frac{z}{[N/2]\sigma_X^2})^{N'})^M \qquad (1.1)$$

where z is the power in the periodogram, N the number and σ_X^2 the variance of the data points. In the literature the inner power term N' varies between $(N-3)/2$ and $N/2$.

On the other hand, simulations (Frescura *et al.*, 2008) suggest that formulæ like Eq.(1.1), or some variants, do not fully describe the cumulative distribution functions (CDFs), especially when only a relatively small number of observations is available. For larger numbers the exponential limit expression can be used as well as for the fully equidistant cases. If more accuracy is needed in the tail of the distribution, it would require a huge number of simulations, and ways of avoiding that are always welcome. There is therefore justification in searching for other ways to describe the CDFs and determine the FAP.

2. Extreme-Value Distributions

One option is extreme-value distributions (Coles, 2004; de Haan, 2006)), as was done by Baluev (2008), extending the results of Davies (2002). We can explore a similar but somehow simpler approach: since the peaks of the periodograms are extreme values of a well-known (or not so well-known) distribution, the distribution of the peaks is an extreme-value one. The generalized extreme-value distribution has the form

$$F(x; \mu, \sigma, \xi) = e^{-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}} \qquad (2.1)$$

and the Gumbel version, for example ($\xi = 0$) is

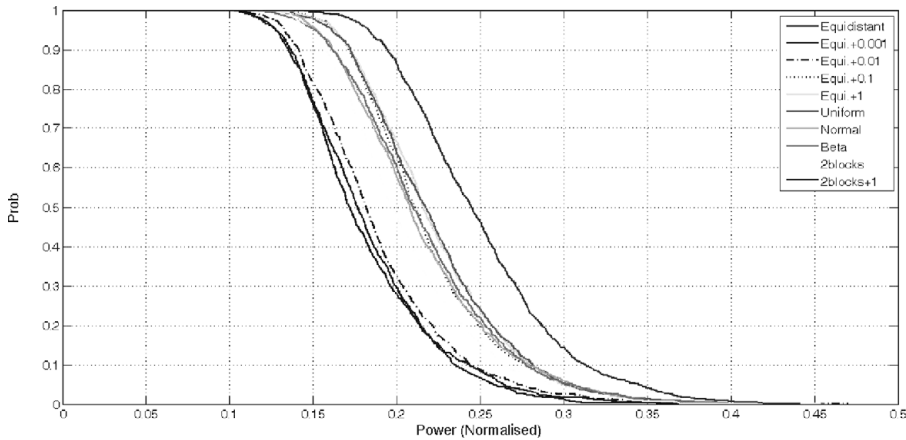$$G(x; \mu, \beta) = e^{-e^{-\frac{x-\mu}{\beta}}} \qquad (2.2)$$

299

Figure 1. Empirical cumulative distributions for simulated data (50 points, Gaussian noise, time-points distributed as indicated over the same interval; only 1000 simulations are shown).

To construct the CDFs we followed the method outlined by Frescura et al. (2008): we simulated a large number of pseudo-Gaussian random time-series with the sampling times of the actual data. We performed a period analysis for each time series after choosing the appropriate frequency range, constructed the empirical CDF of the highest values of each periodogram and fitted the probability function. How the CDF changes as a function of the distribution of the points in time is illustrated in Fig.1. Remark how large the spread can be for an equal number of data points distributed in the same time interval.

3. Tentative Conclusions

Extreme-value distributions seem in general no better at describing the empirical CDFs than fitted β distributions (Eq. 1.1, with M between $N_f/5$ and N_f, where N_f is the number of scanned frequencies corrected for the oversampling in the periodogram). Only for a few distributions of the time-points are there hints that the tail is closer to an extreme-value distribution.

This is only a first exploration of the field. More research is necessary, and many caveats and questions remain.

Acknowledgements

References

Baluev, R. V. 2008, *MNRAS*, 385, 1279
Coles, S. 2001, *An introduction to Statistical Modelling of Extreme values* (London: Springer-Verlag), p. 46
Davies, R. B. 2002, *Biometrika*, 89, 484
Frescura, F. A. M., *et al.* 2008, *MNRAS*, 388, 1693
de Haan, L. & Ferreira, A. 2006, *Extreme Value Theory: An introduction* (New York: Springer-Verlag), p. 9
Lomb, N. R. 1976, *Ap&SS*, 39, 447
Scargle, J. D. 1982, *ApJ*, 263, 835
Schwarzenberg-Czerny, A. 1998, *MNRAS*, 301, 831