Mon. Not. R. Astron. Soc. **371,** 1390–1394 (2006) doi:10.1111/j.1365-2966.2006.10762.x

The Nyquist frequency for irregularly spaced time-series: a calculation formula

Chris Koen★
Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape, South Africa

Accepted 2006 June 29. Received 2006 June 29; in original form 2006 May 19

ABSTRACT
Eyer & Bartholdi showed that the Nyquist frequency of irregularly sampled time-series can be very high. In this paper, a calculation formula for the Nyquist frequency is presented. In practice there is an upper limit of $0.5/\Delta$ on this frequency, where Δ is the best accuracy with which time is recorded.

Key words: methods: statistical.

1 INTRODUCTION

Consider a time-series of measurements $y_1, y_2, \ldots, y_N$ taken at times $t_1, t_2, \ldots, t_N$; the t_i are, in general, not regularly spaced, that is,

$$\delta_j = t_{j+1} - t_j \tag{1}$$

is not constant. The notation is simplified if it assumed that the mean of the y_i is zero, hence this non-essential assumption is made. The periodogram of the y_i is defined as

$$I(\omega) = \frac{1}{N}\left[\left(\sum_{k=1}^{N} y_k \cos\omega t_k\right)^2 + \left(\sum_{k=1}^{N} y_k \sin\omega t_k\right)^2\right], \tag{2}$$

where $\omega = 2\pi\nu$ is the angular frequency.

It is instructive to first consider regularly spaced data, with the time-unit defined such that $\delta_j \equiv 1$ ($j = 1, 2, \ldots, N - 1$); it is well known that the Nyquist frequency in this case is 0.5. Fig. 1 shows the periodogram of the function $y_j = \cos(0.6\pi j) + e_j (j = 1, 2, \ldots, 100)$, the e_j being Gaussian noise with standard deviation 1.5. The spectrum has been plotted at frequencies up to twice the Nyquist frequency. A striking feature of the periodogram is its symmetry with respect to the Nyquist frequency. The entire pattern of Fig. 1 repeats at even higher frequencies, that is, $I(\omega + 2J\pi) = I(\omega)$, for integer J, as can be verified from (1). The Nyquist frequency thus constitutes an upper limit to the frequency range over which the periodogram can be uniquely calculated. (Note though the choices of frequency intervals other than [0, 0.5] are formally acceptable – and may in fact be preferred if there is information pointing to frequencies of interest *not* being in [0, 0.5].)

We follow Eyer & Bartholdi (1999) in using this property of the periodogram – in particular, the symmetry about the Nyquist frequency – as a basis for defining the Nyquist frequency in the case of irregularly spaced time-series. In the next section of this paper, a formula suitable for calculation of the symmetry frequency is derived. Section 3 contains a few simulated and real-life examples.

2 A FORMULA FOR THE CALCULATION OF THE NYQUIST FREQUENCY

Equation (2) can be written in the equivalent form

$$I(\omega) = \frac{1}{N}\sum_{k,\ell=1}^{N} y_k y_\ell \cos\omega(t_k - t_\ell) \tag{3}$$

which saves some algebra.

Setting

$$I(\omega_0 + \psi) = I(\omega_0 - \psi),$$

★E-mail: ckoen@uwc.ac.za

it follows from (3) that

$$
\begin{aligned}
I(\omega_0+\psi)-I(\omega_0-\psi) &= \frac{1}{N}\sum_{k,\ell=1}^{N} y_k y_\ell[\cos(\omega_0+\psi)(t_k-t_\ell)-\cos(\omega_0-\psi)(t_k-t_\ell)] \\
&= \frac{-2}{N}\sum_{k,\ell=1}^{N} y_k y_\ell \sin\psi(t_k-t_\ell)\sin\omega_0(t_k-t_\ell) \\
&= \frac{-4}{N}\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N} y_k y_\ell \sin\psi(t_k-t_\ell)[\sin\omega_0(t_k-t_\ell)] \\
&= 0.
\end{aligned}
$$

In order to have the last sum equal to zero for arbitrary y_i, t_i and ψ, it is sufficient that all terms of the form in square brackets vanish, that is,

$$\sin\omega_0(t_k-t_\ell)=0 \quad \text{for all} \quad k,\ell. \tag{4}$$

The last requirement can be written as a single equation for the unknown $\nu_0=\omega_0/2\pi$:

$$SS(\nu_0)=\sum_{\ell=1}^{N-1}\sum_{k=\ell+1}^{N}[\sin 2\pi\nu_0(t_k-t_\ell)]^2=0. \tag{5}$$

A few remarks:

(i) Let

$$\delta_{\min}=\min_{k,\ell}(t_k-t_\ell)=\min_j(\delta_j)$$

(where δ_j is defined in equation 1). Eyer & Bartholdi (1999) showed that the effect of irregular sampling is to increase the Nyquist frequency over the value expected by analogy with the case of regular time-spacing, namely $0.5/\delta_{\min}$. In fact, in theory there may be no *finite* solution for ν_0.

(ii) In the statistics and signal processing literatures the effect is referred to under the heading ‘alias free spectrum estimation’–a recent reference is Tarczyński & Qu (2005). Note that ‘aliasing’ here refers to the non-uniqueness of the spectrum discussed in Section 1, and not to its usual meaning in the astronomy literature, which is cycle count ambiguity induced by data gaps; see Kurtz (1983).

(iii) If the observations are separated by integer multiples of some underlying time-unit δ_*, that is,

$$t_k-t_\ell=M_{k\ell}\delta_*,$$

where $M_{k\ell}$ is an integer, then

$$\nu_0=\frac{0.5}{\delta_*} \tag{6}$$

satisfies (4).

(iv) Note that (6) will always apply in practice, since time is recorded to some finite precision Δ, hence implicitly defining δ_*. As an example, if the Julian Day is recorded to five decimal places at best, and measurements are made at random times, then $\nu_0=5\times10^4\ \mathrm{d}^{-1}$ or 0.58 Hz. Of course, it may be possible to find smaller ν_0 satisfying (5). In other words, equation (6) provides the upper limit to the Nyquist frequency.

(v) In general it is only necessary to scrutinize (5) in a limited number of frequencies: this follows from (4). Focus on a particular pair t_k, t_ℓ of time-points. It is then required that ω_0 be such that (4) is satisfied; this implies that a *necessary* condition for ω to be the Nyquist frequency ω_0 is

$$\omega=\frac{J\pi}{t_k-t_\ell}, \tag{7}$$

where J is any integer. It is therefore only necessary to evaluate (5) for frequencies given by (7). In practice, it is advantageous to choose the indices k and ℓ so that $\pi/(t_k-t_\ell)$ is as large as possible, that is to have $t_k-t_\ell=\delta_{\min}$. Equation (5) then only needs to be evaluated in integer multiples of $0.5/\delta_{\min}$.

(vi) As pointed out by Eyer & Bartholdi (1999) the effect of finite exposure times will also place a practical upper limit on highest observable frequency.

3 EXAMPLES

We return to the example of Section 1, and consider the same sinusoid defined on the time-interval (1, 100). It is again sampled 100 times, but now at points randomly selected from a 0.0213 time-unit spacing over (0, 100). The periodogram of these data is plotted in Fig. 2, for the same frequency interval as in Fig. 1. The symmetry properties are clearly different (see also the discussion in Deeming 1975).

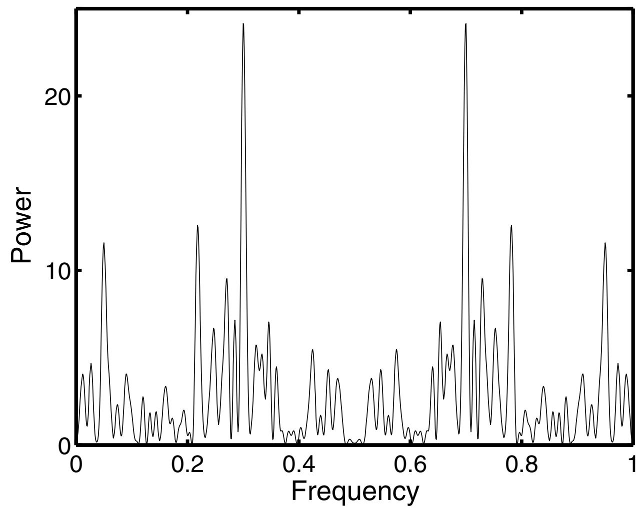

Figure 1. The periodogram of a sinusoid with frequency 0.3, with simulated noise superimposed. The 100 data elements are regularly spaced, one time-unit apart. The plot extends over a frequency interval equal to twice the Nyquist frequency.

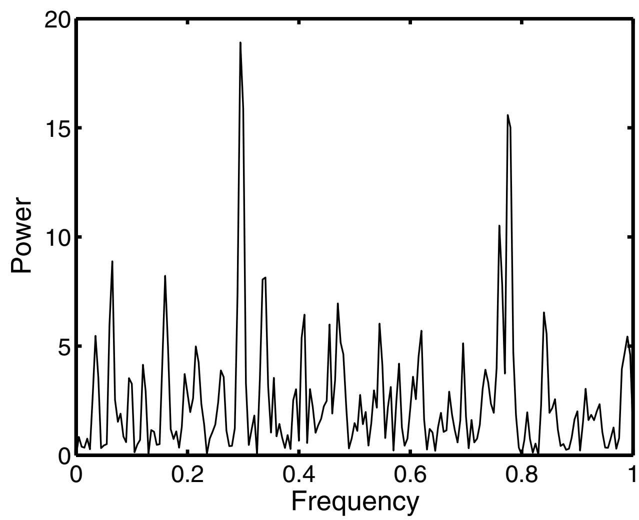

Figure 2. The periodogram of a sinusoid with frequency 0.3, randomly sampled 100 times over a time interval 100 units long. Although the only difference with Fig. 1 is that the time-sampling is no longer regular, the periodogram is no longer symmetrical about a frequency of 0.5.

The results of calculating the sum of squares (5) in the trial frequencies

$$\omega_0 = \frac{j\pi}{\delta_{\min}} = \frac{j\pi}{0.1277} \quad j = 1, 2, \ldots, 20 \tag{8}$$

are plotted in Fig. 3. The solutions are at multiples of $\nu_0 = 23.5$; confirmation that this is correct can be seen in Fig. 4 in which the periodogram is plotted over the interval [0, 47]. By contrast, $0.5/\delta_{\min} = 3.9154$.

Fig. 5 shows plots of the function *SS* in (5) for two disparate data sets. Koen & Ishihara (2006) monitored a few fields in the young open cluster IC 2391 for a few hours each. The intervals between observations for one of these runs, made up of 53 observations spanning 0.28 d (6.8 h), is plotted in Fig. 6. There are slight irregularities in the intervals; these are spread over the interval $[5.31, 5.61] \times 10^{-3}$ d. This scatter is sufficient to produce an alias-free (in the sense of Section 2, point ii) periodogram. The other data set consists of 177 *Hipparcos* brightness measurements of the Cepheid pulsator V473 Lyr, taken over 1150 days. In both cases, the Nyquist frequency is equal to the upper limit of 50 000 d^{-1} (see points iii and iv of Section 2).

We close with a simulation aimed at persuading the disbelieving: a sinusoid with unit amplitude and frequency 17 280 d^{-1} (i.e. a period of 5 s) is sampled with the time-spacing of the IC 2391 observations. For veracity Gaussian noise with a standard deviation of 1.5 is added to each of the 53 simulated observations. The resulting periodogram can be seen in Fig. 7.

ACKNOWLEDGMENT

The author is grateful for fruitful exchanges of ideas with Prof Fred Lombard and Dr Laurent Eyer.

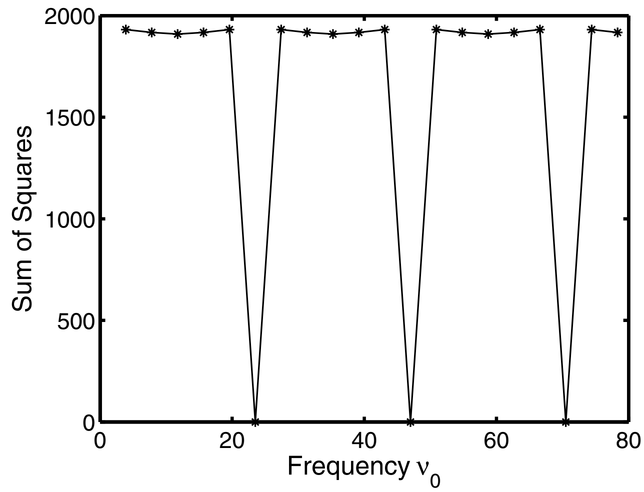

Figure 3. The sum of squares *SS* (equation 5) for the simulated data giving rise to Fig. 2, calculated at multiples of $0.5/\delta_*$ (see text). The zeros occur at multiples of 23.5, the Nyquist frequency.

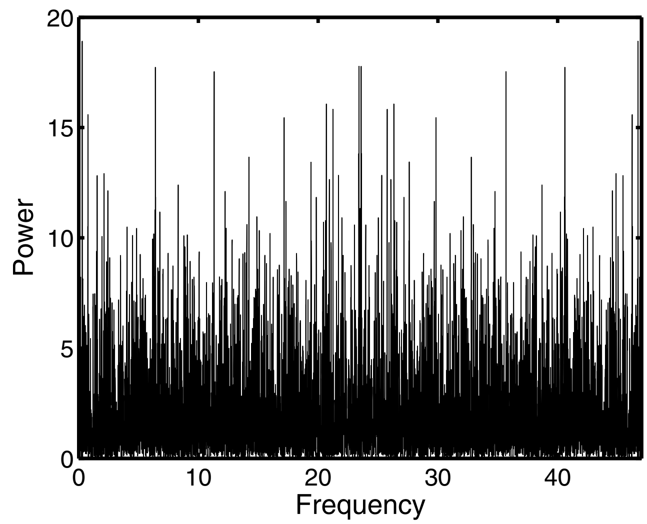

Figure 4. The periodogram of the data described in the caption of Fig. 2. Note the symmetry about the Nyquist frequency 23.5.

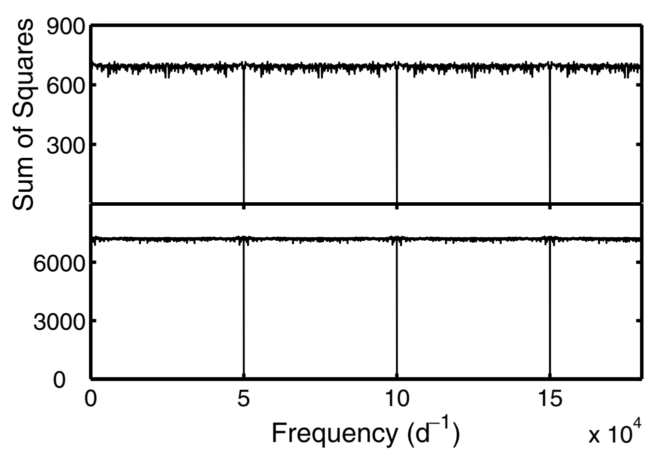

Figure 5. The sum *SS* (equation 5) for one night of observations of a field in IC 2391 (top panel) and for 3 years' *Hipparcos* coverage of V473 Lyr (bottom panel).

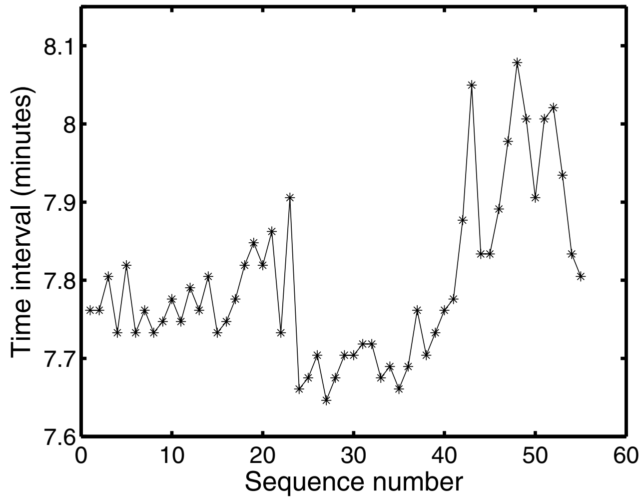

Figure 6. The lengths of the time-intervals between successive observations of a field in the cluster IC 2391.

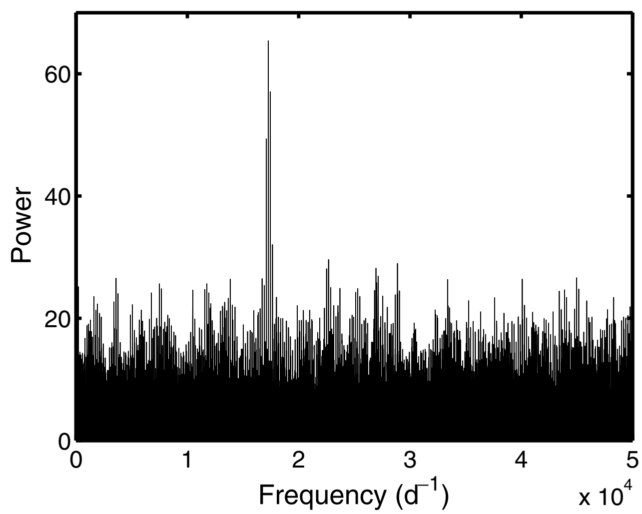

Figure 7. The periodogram of 53 simulated observations of a sinusoid with a period of 5 s (frequency 17 280 d^{-1}). The 'measurements' are spread somewhat unevenly (see Fig. 5) over 6.8 h.

REFERENCES

Deeming T. J., 1975, Ap&SS, 36, 137
Eyer L., Bartholdi P., 1999, A&AS, 135, 1
Koen C., Ishihara A., 2006, MNRAS, 369, 846
Kurtz D., 1983, IBVS, 2285
Tarczyński A., Qu D., 2005, Int. J. Appl. Math. Comput. Sci., 15, 463

This paper has been typeset from a TEX/LATEX file prepared by the author.