About the Nyquist Frequency

F. Mignard,
Observatoire de la Côte d'Azur
Dpt. Cassiopée

GAIA_FM_022

11 April 2005

Abstract

This note addresses the question of the maximum frequency that can be retrieved without aliasing from an unevenly sampled time signal. A general condition is found which expresses the periodicity of the spectral window, from which one can generalise the Nyquist frequency. The standard value is recovered for regular sampling and extended to near-regular samplings, when the successive intervals between observations are integral multiples of a common term. The case of fully irregular signals is then considered in the last section in the light of simultaneous rational approximations of a set of real numbers. Applications are specifically provided as to the smallest periods that can be found in the Gaia photometric signal.

1 Introduction

It is stated very often that the frequency analysis of a time series has a limited range of efficiency because of the unavoidable aliasing effect for frequency higher than the Nyquist frequency $1/2\tau$ when τ is the sampling rate of the series. While this is true for regular samplings, it has been recognised for long that this does not extend to irregular samplings for which one sees, at least empirically, that one can extract from the signal periodic lines at frequencies much higher than the inverse of the mean time step, or even in some case of the inverse of the smallest step between two consecutive observations. This has been addressed many times in the context of observational series in Earth science or astronomy where the observation times are imposed by external constraints. Several rules were given as mentioned in [3], but never (as far as I know) with a rigourous foundation. In the same paper Eyer and Barholdi provide a first justification of some of these rules by showing as a sufficient condition that the highest frequency can be determined from the largest common term that can be found among the different observations times referred to the first. This result is confirmed here in a more general way as a necessary and sufficient condition involving the sequences of intervals between observations instead of their timing.

This allows to extend this result to more general samplings and provides practical rules for very irregular time series.

In the context of the Gaia mission it is important to determine for the very peculiar sampling brought about by the scanning law what kind of periods one can reach from the analysis of the epoch photometry, without being exposed to the risk of producing mirrored frequencies above the Nyquist limit. The theoretical results of this note are applied to various Gaia samplings and checked satisfactorily with simulations. It is shown that either with the astro field alone, or by combining data from the different fields of view, it will be possible to find periods much smaller than the revolution period.

2 Periodogram

Let $T(t)$ be a continuous signal sampled at $t_1, t_2, \ldots, t_n$. The samples will be noted $T(t_k)$ or T_k. The sampling can be regular with $t_{k+1} - t_k = \tau$, where τ is the constant time step, or irregular with $t_{k+1} - t_k = \tau_k$, $k = 1, \ldots, n-1$. There are only $n-1$ independent intervals, the n th variable being the origin of time, whose choice is irrelevant in this problem.

The frequency analysis of $T(t)$ consists in projecting $T(t)$ onto the frequency space by means of a linear rule $\mathcal{R}$ such that,

$$S(\nu) = \mathcal{R}\left(T(t), \exp[2i\pi\nu t]\right), \tag{1}$$

where the only important property needed in the following is the linearity of the operator $\mathcal{R}$. The Fourier transform, the standard projections on the basis functions by means of inner product (covariant coordinates on the basis functions) or least-squares fitting (contravariant coordinates on the basis functions) satisfy this requirement. $S(\nu)$ can be a complex function (meaning two real numbers per frequency), or in the case of a least square fitting with the model,

$$\Phi_\nu(t) = a\cos 2\pi\nu t + b\,\sin 2\pi\nu t + c, \tag{2}$$

there are three real numbers per frequency. The periodogram as a function of the frequency is a representation of the power spectrum and is defined by the real function,

$$P(\nu) = (S\bar{S})^{1/2}. \tag{3}$$

Some scaling factor can be used in Eq. 3 to give a more direct meaning to the periodogram. For example for the model of Eq. 2 a useful normalisation is

$$P(\nu) = \sqrt{2}(<\Phi_\nu(t), \Phi_\nu(t)>)^{1/2} \simeq (a_\nu^2 + b_\nu^2)^{1/2}, \tag{4}$$

so that it characterises the amplitude of the sine wave, which can be directly read out from the periodogram. In (4) the inner product is defined on the actual sampling as,

$$< f(t), g(t) >= \sum_{i=1}^{i=n} f(t_i)g(t_i)w(t_i), \tag{5}$$

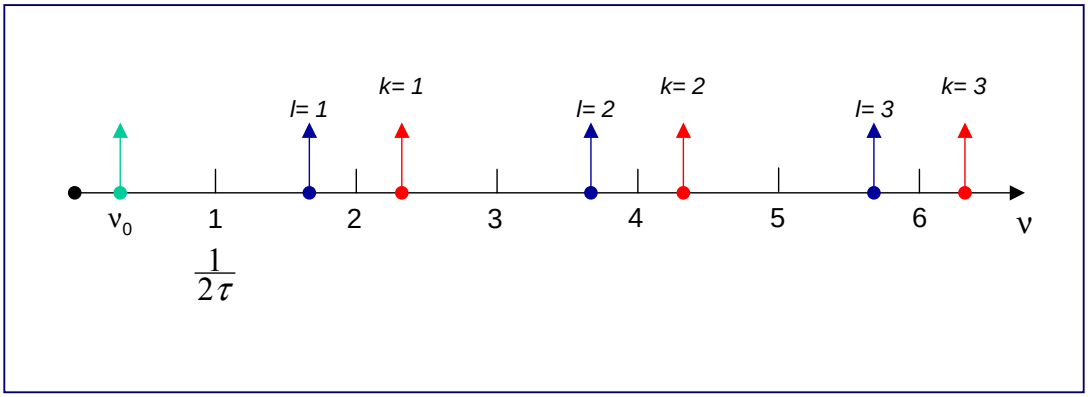


Figure 1: *Aliasing in the frequency domain for an uniformly sampled time series. The time step* $\tau = 0.5$*, giving the Nyquist frequency* $\eta = 1$*. The* ν_+ *and* ν_- *aliases appear respectively in red and blue.*

where $w(t)$ is a weight or window function that can be used to dampen the sidelobes. In all the applications I have used a Tuckey window $w(t) = [1 - \cos(2\pi(t - t_1)/T)]/2$ where $T = t_n - t_1$.

While $T(t_k)$ is the realisation at finitely many points $t_1, t_2, \ldots, t_n$ of an unknown continuous function, $S(\nu)$ is defined for any value of ν and is sampled in frequency for practical reasons. There are no constraints in this sampling as long as one does not use an FFT (although even in that case one can always zero-pad the data to increase the number of data points and then that of the sampling rate in the frequency domain).

In the case of regular sampling with a step τ it is well known that both $S(\nu)$ and $P(\nu)$ are periodic functions in the frequency domain with a period $\eta = 1/\tau$ such that,

$$\forall\ \nu: \quad S(\nu + \eta) = S(\nu), \tag{6}$$

$$\forall\ \nu: \quad P(\nu + \eta) = P(\nu). \tag{7}$$

But we have also from the definition of $P(\nu)$,

$$P(\nu) = P(-\nu), \tag{8}$$

and by combining (7)-(8), one gets the symmetry property,

$$P(\nu) = P(\eta - \nu), \tag{9}$$

meaning that not only is P periodic, but also symmetric with respect to $\eta/2$, the Nyquist frequency. Therefore any line of frequency $0 < \nu < \eta/2$ is mirrored into infinitely many lines at $\nu + k\eta$ and $-\nu + l\eta$ where k and l are positive integers. This gives rises to the two sets of aliases $\{\mathcal{A}_+\}$ and $\{\mathcal{A}_-\}$ at frequencies,

$$\nu_+ = \nu + k\eta, \tag{10}$$

$$\nu_- = -\nu + l\eta. \tag{11}$$

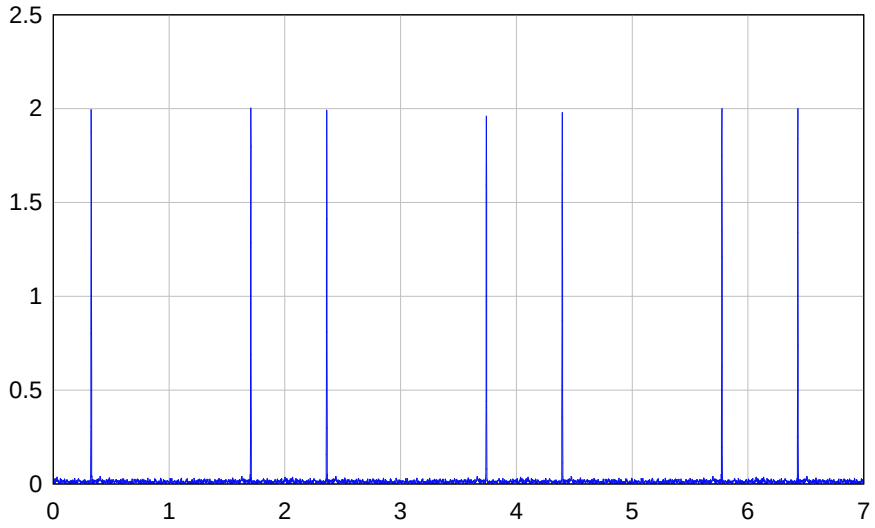


Figure 2: *Periodogram from simulated data with regular time sampling. The data is* $2\cos(2\pi/3\,t)$ *sampled over 1000 data points spaced every 0.5 unit of time. The Nyquist frequency is* $\nu_N = 1$ *in the frequency space and the two families of mirrored spectral lines are conspicuous. Note that the normalisation of the spectrum with (4) gives directly the amplitude of the line on the spectra.*

This well known result (it will be proved in the next section as a particular case of a more general result) is the basis of the common statement: in a regularly sampled time series with step τ one cannot retrieve a periodic signal with frequency higher than $1/2\tau$. This is better stated as follows: in a regularly sampled time series, one can retrieve any periodic signal without aliasing in a frequency range of length $1/2\tau$, e.g. $(1/\tau,\ 3/2\tau)$. In general the periodogram of evenly spaced data is not computed at frequency larger than the Nyquist frequency, since all the frequency information is available in this interval. If one has good reasons to search for a spectral information at higher frequency, it is mirrored into the interval 0, $\eta/2$, and the degeneracy cannot be lifted without an additional piece of information. This is illustrated in Fig. 1 where the two families of aliases are shown for one basic frequency less than the Nyquist frequency. When noise is added, the choice of the right frequency in the aliased domain will be most naturally searched with a Bayesian estimator as shown in [1].

The same pattern appears clearly on the results of a simulation in Fig. 2 with $S(t) = 2\cos(2\pi\nu\,t)$, with $1/\nu = 3$ and $\tau = 0.5$ using 1000 regularly distributed data points and gaussian random noise with $\sigma = 0.2$. The periodogram is computed well above the Nyquist frequency $\nu_N = 1$ to evidence the two sets of aliases (2.33, 4.33, $\cdots$ and 1.67, 3.67, $\cdots$), and the periodicity $\eta = 2$.

3 Aliasing for quasi regular sampling

One considers an arbitrary sampling t_1, t_2, $\ldots, t_n$ with $t_{k+1} - t_k = \tau_k$. One should notice that Eq. 8 still holds and aliasing will occur if and only if $S(\nu)$ is periodic. Given the linear nature of the projection operator, this happens only when the set of phases $\{2\pi\nu t_k, k = 1, \ldots, n\}$ reproduces identically (to a constant offset irrelevant in the power spectrum) for two frequencies ν and $\nu + \eta$. This translates into the fundamental equations,

$$\boxed{\nu\tau_k \equiv (\nu + \eta)\tau_k \pmod 1, \quad k = 1, \ldots, n-1} \tag{12}$$

or equivalently the period η in the frequency domain, if it exists, must satisfy the system of $n - 1$ equations,

$$\boxed{\eta\,\tau_k \equiv 0 \pmod 1, \qquad k = 1, \ldots, n-1} \tag{13}$$

Remark : If η_0 is a solution of (13) then any integral multiple $m\,\eta_0$ of η_0 is also a solution. By definition the period will be the smallest positive solution of (13).

3.1 Uniform sampling

In this case $\tau_k = \tau$ for $k = 1, \ldots, n-1$ and the system of equations degenerates into a single equation,

$$\eta\,\tau \equiv 0 \pmod 1, \tag{14}$$

whose least solution is not surprisingly ,

$$\eta = 1/\tau. \tag{15}$$

With the symmetry relation (8) one recovers the Nyquist frequency $\nu_N = \eta/2 = 1/2\tau$. Other solutions for the periods are $\eta = m/\tau$ with $m \in \mathbb{N}_+$, that is to say the multiples of the smallest period.

3.2 Regular sampling with gaps

There are many ways for a sampling to depart from a regular and uniform sampling. A mathematician will qualify them all as *irregular* samplings. However in practical situation there is a continuous range of samplings between the fully regular and the fully random, where the Nyquist frequency becomes infinitely large. Between this two extremes, there are near regular samplings with finite Nyquist frequency, but much larger than the inverse of the mean sampling interval $\bar{\tau} = (t_n - t_1)/(n-1)$, a very nice feature to determine the periods of variable stars with Gaia.

Consider the case when every sampling interval is an integral multiple of a common duration τ, which by a proper choice of units could be chosen equal to one. Therefore

$$\tau_k = p_k\,\tau, \qquad k = 1, \ldots, n-1, \quad p_k \in \mathbb{N}. \tag{16}$$

Let τ be a value satisfying (16) and r an integer; then $\tau' = \tau/r$ is also solution of (16) with $p'_k = rp_k$. Therefore τ will be uniquely defined if we select the largest acceptable τ, leading to the smallest set of $p_1, \ldots, p_{n-1}$. This set is such that $(p_1, \ldots, p_{n-1}) = 1$ where $(p, q) = \mathrm{GCD}(p, q)$ and $(p,\, q,\, r) = (p,\, (q,\, r))$.

Proof. The proof is by contradiction. Suppose that there is another value $\tau' = r\tau$ with $r \in \mathbb{N}$ satisfying (16). This would mean that there is a set $p'_1, \ldots, p'_{n-1}$ with $p'_k = p_k/r \;\; \forall k$, which would contradict the assumption that $(p_1, \ldots, p_{n-1}) = 1$. □

One should notice that in general for a truly irregular sampling there is no such τ, since this implies that the ratios τ_l/τ_m are rational numbers for any $(l,\, m)$, a very unlikely occurrence. In actual computation with finite arithmetic, or with truncated numbers read out in an input file, one can always find τ as small as the value of the last significant digit. In practice this is too small a number to be of practical interest in the context of the generalised Nyquist frequency.

Property 1. *For a pseudo-regular time sampling, such that any interval τ_k between two successive samples is an integral multiple of a certain τ, and provided τ is the largest such number, the power spectrum of a time series built on this sampling is periodic in the frequency domain with period η given by,*

$$\boxed{\eta = \frac{1}{\tau}} \tag{17}$$

Proof. With the largest integral common submultiple τ of the τ_k, (13) becomes,

$$\eta\tau\, p_k \equiv 0 \pmod 1, \qquad k = 1, \ldots, n-1, \tag{18}$$

whose solutions for each k are

$$\eta = \frac{m}{\tau p_k}. \tag{19}$$

Whence the solution of (18) is the least value generated by the multiples of the $1/p_k$:

$$\eta = \frac{1}{\tau}\left[\frac{l_1}{p_1},\, \frac{l_2}{p_2},\, \ldots,\, \frac{l_{n-1}}{p_{n-1}}\right] \tag{20}$$

where the l_i must be so selected as to obtain the less possible value of η. A trivial solution is $l_k = p_k, \;\; k = 1, \ldots, n-1$ leading to $\eta = 1/\tau$. Any smaller solution will be of the form η/p where p is an integer, meaning that p will be a common divisor of the p_k; hence $p = 1$ since $(p_1,\, p_2,\, \ldots, p_{n-1}) = 1$ and this completes the proof. □

Due to $P(\nu) = P(-\nu)$ the expression for the Nyquist frequency remains $1/2\tau$. As said by Bretthorst [1], the aliasing has not been removed with the irregular sampling, but has been pushed forward at higher frequency. When $\tau_k = \tau$ one recovers the regulars sampling as a particular case and the usual meaning of the Nyquist frequency.

If we had not selected the largest integral submultiple τ of the τ_k, $k = 1, \ldots, n-1$, but a smaller τ', the $l_k = p_k$ could have been be divided by their largest common factor to produce a smaller solution. With $p = (p_1, p_2, \ldots, p_{n-1})$, one would have had for every k : $p_k = p\, p'_k$ and then by taking $l_k = p'_k$ the least period would have been

$$\boxed{\eta = \frac{1}{p\,\tau'}} \tag{21}$$

a result already stated as a sufficient condition by Eyer and Bartholdi [3], but with a somewhat annoying statement about the GCD of real numbers, although what is meant is clear and right. The exact role played by the irrationality of the $t_k - t_1$ is pointed out but not formulated in a proper manner. This indeed plays no role at all; what matters is even not the rationality or irrationality of the τ_k, which can be made all irrational by a choice of the unit, but that of τ_k/τ_1 which is intrinsic as being independent of the origin of time and of a scaling factor related to the freedom in the choice of the units. It is obvious from this derivation that $\tau = p\tau'$ and (17) and (21) are in fact identical. A similar result is also given by Bretthorst in [1] in a signal processing paper probably not known from the astronomical community.

The property expressed by (17) is equivalent to saying that observations were planned at regular intervals τ and some were missed, in such a way that the actual intervals are multiples of the planned period and that these multiples are mutually relatively prime. So whatever the distributions of the gaps, the Nyquist frequency remains unchanged and keeps the value it would have had with a regular sampling with a step τ without gaps. What is nice in this sampling is that the gaps may exist from the start and none of the intervals τ_k needs to be as small as τ to benefit from the displacement of the Nyquist limit at higher frequency. So this can be used to plan efficiently observation runs so that the Nyquist frequency is large, without taking the burden of observing very often (I disregard here the statistical improvement provided by the increase of the number of observations).

4 Irregular sampling

As stated earlier, it is difficult to say what an irregular sampling is, as it is not just the negative of the regular case. To stress this point I have plotted the periodogram of a sine wave sampled with an artificial sampling (exponential waiting time between observations) and a real time series coming from 10 years of observation at the Lunar laser ranging station of the Observatoire de la Côte d'Azur [6]. In the first case the power spectrum is

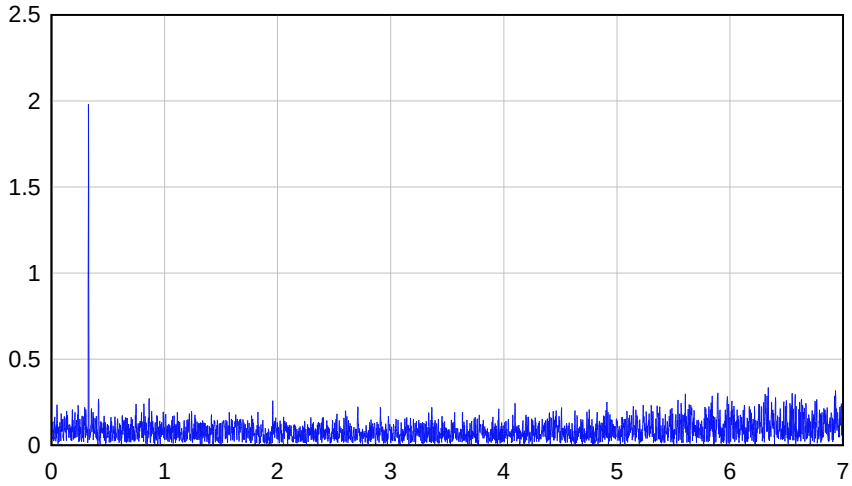


Figure 3: *Periodogram from simulated data with irregular time sampling. The data is* $2\cos(2\pi/3\,t)$ *sampled over 1000 data points spaced with a exponential waiting time of 0.5 unit of time as average, as in Fig. 2. The power spectrum is plotted up to seven times the Nyquist frequency of a regular sampling over the same duration. There is no sign of aliasing and this is true even at higher frequencies.*

plotted in Fig. 3 and must be compared to Fig. 2 with a regular sampling of time step equal to the average of the exponential probability distribution used for this simulation. There is no sign of aliasing and very high frequency signals could be retrieved with that random sampling.

Unfortunately in the real world we usually meet something between the regular and the purely random sampling as illustrated by the next example. For the Lunar observations nobody will qualify this sampling as regular or nearly regular mainly because of the randomness added by the occurrence of good or bad weather. This is basically an irregular sampling but with some repeating features, mainly due to the fact that one tries to observe the Moon every good night, excluding the few days bracketing the new and full Moon. The intervals between successive observations are primarily composed of the short durations of about 15 mn between data points within a night. These intervals account for about 90% of the τ_k. The next most frequent interval results from the daily attempts to range the Moon and is close to one day (on the average 1.035 day, the lunar day of the oceanic tides). As can be seen in Fig. 4 this feature dominates the power spectrum at low frequency and generates near aliases. So mathematically it is difficult to draw general conclusions as soon as the time sampling departs significantly from the two simple cases considered above and facing an actual sampling extreme care must be exercised.

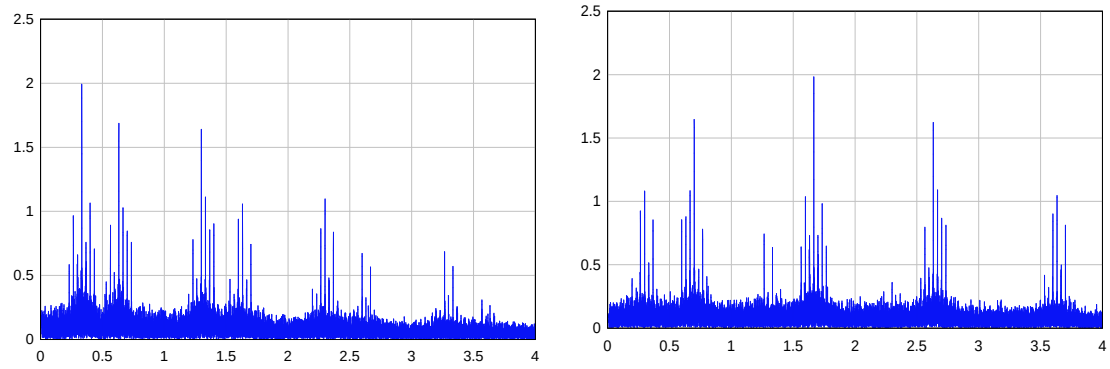


Figure 4: *Periodograms on simulated data with an irregular time sampling. The data is a sine wave of amplitude* 2 *with period* 3 *days on the left panel and* 0.6 *days on the right and the unit of frequency on the* x*-axis is cy/day. The time sampling comes from Lunar observations carried out every good night with the avoidance of the full and new Moon (few days around). Each night there are about 5 to 15 closely packed data points with a step of about 15 mn. Rigourously speaking the periodogram is not periodic and the Nyquist frequency is rejected at very high frequency, in principle removing any risk of aliasing. However both plots show a rather regular pattern with a quasi-period of about 1 days and the leftover of the Nyquist frequency of 0.5 cy/d. This means that the repetition of groups of observations with a planned recurrence of one day, albeit the big gaps, dominates the structure of the frequency analysis. However, although the lines are mirrored, they are not reproduced identically and there is no problem to identify the simulated line showing up with the largest power, even above the remnant Nyquist frequency on the right panel. One should also mention that all theses lines are not mutually orthogonal and after filtering the data for the main line, all the other ghost lines vanish from the subsequent periodogram. This feature illustrates again the traps that may result from (too) quick an interpretation of a periodogram : there is in the data only one periodic signal with no harmonics. Filtering is mandatory to analyse over several frequencies non-harmonic periodicity or multi-periodic signals.*

4.1 A numerical exemple

We go back again to the basic equation (16) whose solution determines essentially how far one can go in frequency without aliasing. As mentioned before, in general there is no exact solution for this system of equations, as the τ_k are mutually incommensurate. This results in practice into a very high Nyquist frequency compared to the inverse of the least interval. But as $\pi \simeq 355/113$ or $\sqrt{2} \simeq 239/169$, there may exist rather good simultaneous rational approximations of all the ratios τ_k/τ_1, $k = 2, \ldots, n-1$. In this case, despite a mathematically irregular sampling, there will be a near periodicity in the frequency space, and given the noisy nature of real data, this could make an effective Nyquist frequency within reach of the analysis. Mathematically this amounts to searching the simultaneous rational approximation of $\tau_2/\tau_1, \tau_3/\tau_1, \ldots, \tau_{n-1}/\tau_1$. The fact that one of the τ_k is not

directly involved comes from the degree of freedom left by the choice of unit. This is equivalent to setting for example $\tau_1 = 1$ and restoring the true units later. If there is a simple rational approximation (meaning with small denominators), the period η of the periodogram will be small and aliases will appear much earlier than expected.

We are now left with a formidable problem of finding the rational approximations of a set of real numbers. There is a considerable literature on this subject, with very few general results, let alone algorithms. As for the irrational numbers for which one finds rather good and unexpected simple rational approximations (see π above), in the real world the appearance of rational approximations is the rule rather than the exception (an irrational number without simple rational approximations should not be selected at random!). For the frequency analysis this means that even with a somewhat irregular sampling a near Nyquist frequency will appear much earlier than expected.

To illustrate this point before looking at the mathematics consider a numerical example (built on purpose). We have 6 observations at $t_1, \ldots, t_6$ giving for the intervals in arbitrary unit :

$$\begin{aligned} \tau_1 &= 3.12, \\ \tau_2 &= 14.60, \\ \tau_3 &= 20.05, \\ \tau_4 &= 8.62, \\ \tau_5 &= 22.95. \end{aligned}$$

The largest common measure between these numbers is $\tau = 0.01$ (multiplying each number by one hundred leaves five mutually prime integers) giving $\eta = 100$ and a Nyquist frequency of 50, much larger than the inverse of the smallest interval. Now it happens that one has approximately,

$$\begin{aligned} \tau_2/\tau_1 &\approx 50/11, \\ \tau_3/\tau_1 &\approx 70/11, \\ \tau_4/\tau_1 &\approx 30/11, \\ \tau_5/\tau_1 &\approx 80/11, \end{aligned}$$

and Eqs. 16 nearly hold with $\tau = \tau_1/11$ and $p_2 = 50$, $p_3 = 70$, $p_4 = 30$, $p_5 = 80$. Thus one finds $\eta = 11/\tau_1 = 3.52$ and a practical Nyquist frequency does exist at $\nu = 1.76$. Again this is higher than the inverse of the mean interval and even of the least interval, but much smaller than the mathematically defined Nyquist Frequency at $\nu = 50$. So even if the periodogram is not strictly periodic with period 3.52 (the true period is 100), the actual pattern is sufficiently close to repeat itself every 3.52 units of time, that in actual data treatment the risk of aliasing at $\nu = 1.76$ is high and cannot be overcome.

4.2 Simultaneous diophantine approximation

In this section I state the main known mathematical result about the rational approximation of a vector of real numbers to draw a general conclusion on the effective Nyquist frequency.

Definition 1. *Given n real numbers $\alpha_1, \ldots, \alpha_n$ a simultaneous diophantine $\epsilon-$approximation by rational numbers is a sequence of $n+1$ integers $p_1, \ldots, p_n$ and q such that for all $k \in \{1, \cdots, n\}$ one has $|q\alpha_k - p_k| < \epsilon$.*

For $n = 1$ one recovers the usual definition of the approximation of a real by the rational p/q and it is known that there are infinitely many solutions to the equation (see any textbook on the theory of number like [4], [5] or [2]),

$$|\alpha - \frac{p}{q}| < \frac{1}{q^2}, \tag{22}$$

with practical solutions following from the expansion of α in continued fractions. When $n > 1$ the only generalisation comes from the Dirichlet's theorem (based on the pigeonholes principle) which tells that:

Property 2. *There exists at least one simultaneous approximation satisfying the system of inequalities:*

$$|\alpha_k - \frac{p_k}{q}| < \frac{1}{q^{1+\frac{1}{n}}}, \quad k = 1, \cdots, n. \tag{23}$$

If one of the α_k is not rational, there are infinitely many solutions. This means that given an ϵ and the set of α_k, one can always find an integer q such that the $q\alpha_k$ differ from an integer by less than ϵ.

In the present context, we have $\alpha_k = \tau_k/\tau_1, \quad k = 2, \cdots, n-1$ and we know that to any desired degree of accuracy one can find q and $p_2, \ldots, p_{n-1}$ such that $q\tau_k/\tau_1 \approx p_k$. Inserting in (16) this yields,

$$\tau_k \approx p_k \frac{\tau_1}{q} \tag{24}$$

and then $1\tau \approx q/\tau_1$ for the quasi-period of the periodogram and $q/2\tau_1$ for the practical Nyquist frequency. While standard algorithms are easy to implement to obtain the best rational approximation of a real number (e.g. the set of convergents of the continued fraction), I know of no easily available method to find out the successive approximations of a vector of real numbers (The brute force is quickly overwhelmed by the running time).

Remark: Without any change one could have replaced τ_1 by $\tau_m = \min(\tau_1, \ldots, \tau_{n-1})$, which shows that the effective Nyquist frequency is in general of the order or larger (and can be much larger) than the inverse of the smallest interval. If the smallest interval τ_m is repeated many times and well separated in magnitude from the other intervals, we will

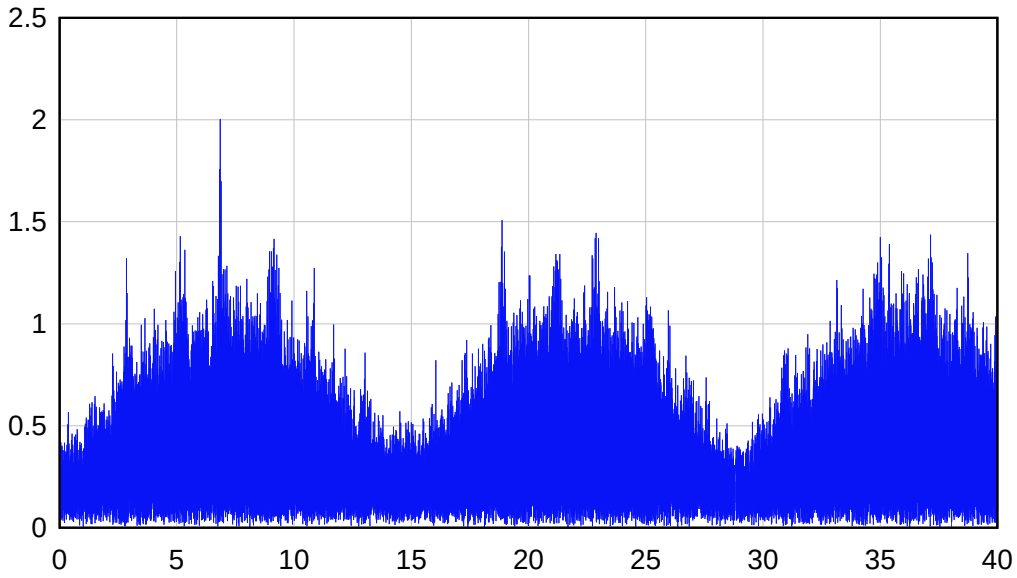


Figure 5: *Periodogram on simulated data with the Gaia time sampling for the Gaia astro field (x-axis in cy/day). The data is a sine wave of amplitude* 2 *with period* 3.5 *hours, i.e. 6.85 cy/d. The most conspicuous periodicity is linked to the smallest interval between successive observations, namely* $\sim$ 100 *mn* $\sim$ *1/14.5 days. So the effective Nyquist frequency is 7.2 cy/day, and in principle aliasing arises for periods less than* $\sim$ 3.4 *hours. The diagram shows however that the line with amplitude* 2 *at 3.5h period is recovered without ambiguity, and no large alias is present at higher frequencies.*

end up with $q = 1$ in the above and this will be virtually identical to the semi-regular sampling with gaps and the quasi-period will be τ_m.

For Gaia the hierarchy between the $\sim$ 100 mn interval between the preceding and following fields and the 6 h period is not very large and the effective Nyquist frequency will be larger than $1/\tau_m$. This indicates that, even by using only the astro photometric data, one should be able to obtain reliable periods of variable stars as short as $\sim$ 1h. Surprisingly, in case one combines the data from the astro and MBP fields of view (assuming that a colour equation can be applied to bring all the data on a common bandwidth) there will be many intervals with the revolution period of 6 h due the larger field height of the MBP field and quasi aliases may occur at lower frequency, although the smallest interval between the MBP and astro fields is smaller than 100 mn.

To evidence this situation I have simulated a time signal of period 3.5 or 0.9 hours with the Gaia sampling over 1800 days, (i) for the astrometric field only and (ii) for the combined data of BBP and MBP. The periodograms are shown in Figs. 5-7 and confirm the expected properties.

- For the BBP signal we see in Fig. 5 a pattern which repeats more or less identically

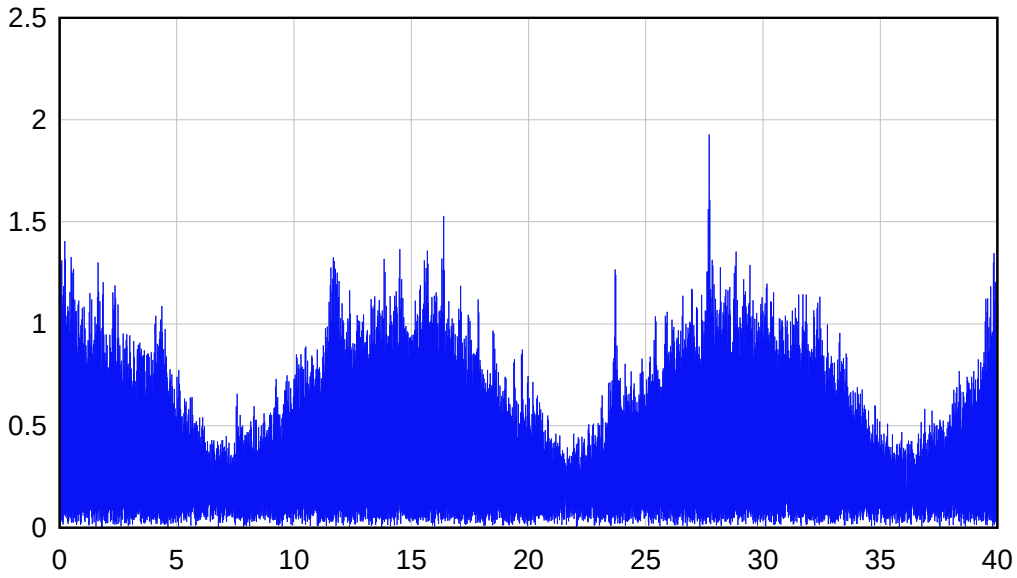


Figure 6: *Periodogram on simulated data with the Gaia time sampling for the astro field (x-axis in cy/day). The data is a sine wave of amplitude* 2 *with period* 52 *mn, i.e. 27.7 cy/d., well above the effective Nyquist frequency of 7.2 cy/d linked to the smallest interval between successive observations, namely 100 mn* $\sim$ *1/14.5 days. The diagram shows that the line with amplitude* 2 *is recovered without ambiguity at the correct period.*

every $\sim$ 14.5 cy/d (= 1/100 mn), yielding in principle aliased spectra for periods less than $\sim$ 3.3 hours. However the actual sampling departs enough from a pseudo-regular one so as to generate mirror lines with amplitudes significantly smaller than the main line, allowing to recognise the real line at a frequency higher than the effective Nyquist frequency.

- The case with a short period signal is shown in Fig. 6 with a signal of period = 52 mn simulated over the same time sampling. Although the period is located well into the aliased domain, its amplitude is larger than any other mirrored lines and is perfectly detected at the correct frequency with the correct amplitude.

- Finally Fig. 7 illustrates the first case (period 3.5h, 6.85 cy/d) when BBP and MBP samplings are combined into a single time series. The pattern in the frequency domain is more complex, but not at all surprising (the small white spikes are artefacts of the image compression) : (i) Small recurring blocs with frequency width of 4cy/day related to the revolution period and the repeated observation every 6h in the MBP field; (ii) the two families of aliases are visible at $\pm 6.85 + 4k$ cy/d; (iii) this structure repeats itself every 14.5 cy/day corresponding to the 100 mn interval between preceding and following field int he BBP band. However the 6h period is so common that the associated aliasing pervades the whole diagram but the aliased lines of the

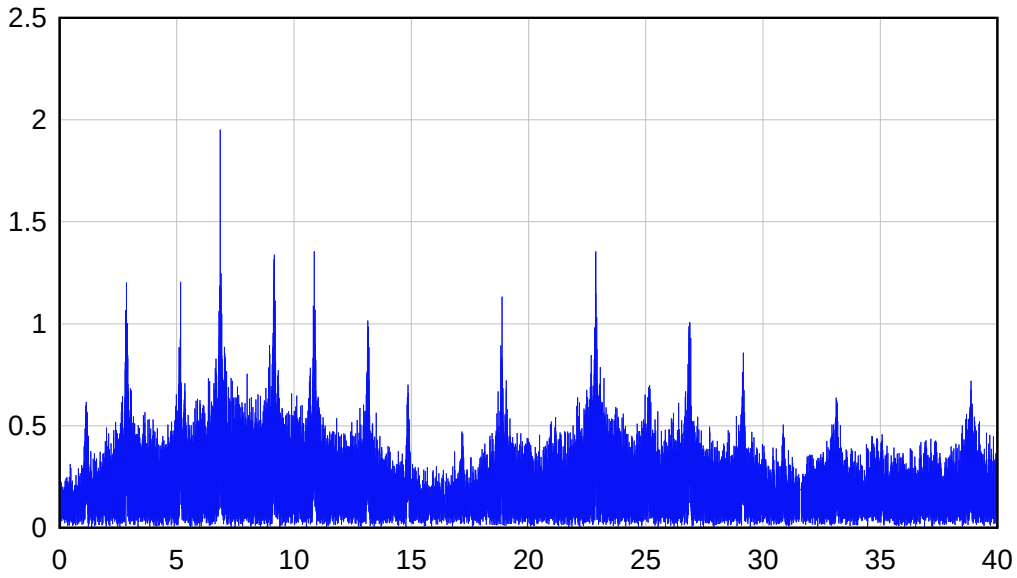


Figure 7: *Periodogram on simulated data with the Gaia time sampling combining the astro and the MBP fields (x-axis in cy/day). The data is a sine wave of amplitude* 2 *with period* 3.5 *h, i.e. 6.85 cy/d. The diagram structure is dominated by the repetition of observations every 6h (4 cy/d in frequency) in the MBP fields, producing quasi-aliases. However due to the irregularity introduced by the transits in the astro field and the long gaps between epochs, the power from the main spectral line remains higher than that of the mirrored lines and the simulated period and amplitude are perfectly recovered.*

actual period of 3.5h (6.8 cy/day)remain smaller than the main line, hence avoiding ambiguity in the detection. In these simulations I have also noticed that all the periodograms may exhibit surprising regularities when the period of the signal is in simple relation with one of the fundamental period of the observation window. This deserves a dedicated investigation to see the consequences in the detection of these particular periods.

5 Conclusion

The frequency analysis of an unevenly sampled time series may still suffer from aliasing, although not in the same way as uniform sampling. The bandwidth can be very high but remnants of subsets of data more or less regularly sampled can cause near aliasing, but fortunately the height of the spectral line allows in general to recover the correct frequency without ambiguity. For a near regular sampling, when all the intervals between successive observations have a common dwell time τ, the Nyquist frequency is $1/\tau$, at least as large as the inverse of the smallest interval. For irregular sampling, the existence of

near rational approximations between the intervals help understand the complex features in actual periodograms coming from real observations. But even in these cases, the true Nyquist frequency may be orders of magnitudes larger than the average sampling rate, making irregular sampling a desirable feature in observation planning. In the particular of Gaia we have shown that the period search could be performed successfully with duration significantly less than the smallest sampling interval.

Acknowledgment

The author thanks L. Eyer for his useful comments on a preliminary version of this note.

References

[1] Bretthorst, G. L., 2001, Nonuniform Sampling: Bandwidth and Aliasing, in Maximum Entropy and Bayesian Methods in Science and Engineering, Joshua Rychert, Gary Erickson and C. Ray Smith (eds.), American Institute of Physics, pp. 1-28.

[2] Davenport, W. J.,1982, The higher artithmetic, 5th edition, Cambridge, chap. 4.

[3] Eyer, L., Bartholdi, P., 1999, Astron. Astrophys. Sup. Ser., **135**, 1-3.

[4] Hardy, G.H., Wright, E.M, 1985, An introduction to the theory of numbers, 5th edition, Oxford UP, chap. 10.

[5] Leveque, W. J.,1977, Fundamentals of number theory, Addison-Wesley, chap. 9.

[6] LLR team, 2005, see `www.obs-azur.fr/cerga/laser/laslune/llr.htm`