MNRAS **453,** 1793–1798 (2015) doi:10.1093/mnras/stv1744

The analysis of indexed astronomical time series – XII. The statistics of oversampled Fourier spectra of noise plus a single sinusoid

Chris Koen★

Department of Statistics, University of the Western Cape, Private Bag X17, Bellville, 7535 Cape Town, South Africa

Accepted 2015 July 28. Received 2015 May 30

ABSTRACT

With few exceptions, theoretical studies of periodogram properties focus on pure noise time series. This paper considers the case in which the time series consists of noise together with a single sinusoid, observed at regularly spaced time points. The distribution of the periodogram ordinates in this case is shown to be of exponentially modified Gaussian form. Simulations are used to demonstrate that if the periodogram is substantially oversampled (i.e. calculated in a dense grid of frequencies), then the distribution of the periodogram maxima can be accurately approximated by a simple form (at least at moderate signal-to-noise ratios). This result can be used to derive a calculation formula for the probability of correct signal frequency identification at given values of the time series length and (true) signal-to-noise ratio. A set of curves is presented which can be used to apply the theory to, for example, asteroseismic data. An illustrative application to *Kepler* data is given.

Key words: methods: statistical – stars: variables: general.

1 INTRODUCTION

A paper by Baran, Koen & Pokrzywka (2015) provides a different view of the hoary problem of testing periodogram peaks for significance. Standard procedure is to consider whether the largest peak is consistent with the time series consisting of only noise. If this null hypothesis is rejected, it is assumed that the frequency associated with the maximum peak is due to the presence of a sinusoidal component with that frequency. Baran et al. (2015), on the other hand, simulated time series consisting of noise together with a sinusoid. They presented probabilities of extracting the correct signal frequency, depending on the signal amplitude and the length N of the time series. Their results demonstrate that a surprisingly high signal-to-noise ratio $\mathcal{R}$ is required to find the correct frequency with high probability. They also show that the necessary $\mathcal{R}$ increases with increasing N.

The aim of this paper is to further explore the Baran et al. (2015) model. In particular, simulations are used to show that the distribution of signal peaks in the periodogram has a simple form if spectra are substantially oversampled, and the signal-to-noise ratio not too low. This result can be used to derive a calculation formula for the probability of correctly identifying the signal frequency from a noisy time series, in the case of measurements which are regularly spaced in time.

The reader should bear in mind that throughout the paper logarithms are to the base e.

2 THE PERIODOGRAM OF A SINUSOIDAL SIGNAL

Let

$$y_t = s_t + e_t \quad t = 1, 2, \ldots, N, \tag{1}$$

where

$$s_t = A\cos(\omega_* t + \phi) = A\cos(2\pi\nu_* t + \phi) \tag{2}$$

is the signal, and e_t is zero mean Gaussian noise with variance σ_e^2. The standard periodogram is

$$I(\omega) = \frac{1}{N}\left\{\left[\sum_t (y_t - \overline{y})\cos\omega t\right]^2 + \left[\sum_t (y_t - \overline{y})\sin\omega t\right]^2\right\} \tag{3}$$

Substituting (1) and (2) into (3) and multiplying out

$$I(\omega) = I_s(\omega) + I_e(\omega) + H(\omega), \tag{4}$$

where

$$I_s(\omega) = \frac{1}{N}\left\{\left[\sum_t (s_t - \overline{s})\cos\omega t\right]^2 + \left[\sum_t (s_t - \overline{s})\sin\omega t\right]^2\right\}$$

$$I_e(\omega) = \frac{1}{N}\left\{\left[\sum_t (e_t - \overline{e})\cos\omega t\right]^2 + \left[\sum_t (e_t - \overline{e})\sin\omega t\right]^2\right\}$$

$$= \frac{1}{N}\sum_{j,k}(e_j - \overline{e})(e_k - \overline{e})\cos(j-k)\omega \tag{5}$$

★ E-mail: ckoen@uwc.ac.za

$$H(\omega) \equiv \frac{2}{N}\sum_{j=1}^{N}\sum_{k=1}^{N}(s_j-\overline{s})(e_k-\overline{e})[\cos j\omega\cos k\omega+\sin j\omega\sin k\omega]. \tag{6}$$

The three spectra are now each examined in more detail.

It is well known that the signal periodogram has the sinc^2-function form

$$I_s(\omega) = \frac{NA^2}{4}\left[\frac{\sin\pi N(\nu-\nu_*)}{\pi N(\nu-\nu_*)}\right]^2 = \frac{NA^2}{4}\mathrm{sinc}^2[N(\nu-\nu_*)].$$

The signal-to-noise ratio is defined as

$$\mathcal{R} = \frac{A}{2\sigma_e/\sqrt{N}} \tag{7}$$

so that in terms of $\mathcal{R}$

$$I_s(\omega) = \mathcal{R}^2\sigma_e^2\ \mathrm{sinc}^2[N(\nu-\nu_*)]. \tag{8}$$

The presence of the sinc^2 factor in equation (8) implies that the signal spectrum is effectively zero outside a very narrow interval $(\nu_*-1/N, \nu_*+1/N)$, i.e. $I(\omega)\approx I_e(\omega)$ and hence

$$EI(\omega)\approx EI_e(\omega)=\sigma_e^2 \tag{9}$$

outside the narrow frequency interval. It follows that

$$I_s(\omega)\approx\mathcal{R}^2\overline{I}\ \mathrm{sinc}^2[N(\nu-\nu_*)]. \tag{10}$$

Values of I_e at different frequencies are all exponentially distributed with mean σ_e^2. Consequently, using equation (9),

$$\begin{aligned} I_e(\omega)/\overline{I_e} &\approx I_e(\omega)/\overline{I} \\ &\equiv I_e'(\omega)\sim\exp(1), \end{aligned} \tag{11}$$

i.e. an exponential distribution with unit mean. In what follows $\overline{I}$ will usually be treated as a constant: it is, of course, a random variable, with a variance of σ_e^2/N. Since the paper is concerned with large data sets, the variability of this mean value is negligible.

The distribution of the third spectrum in equation (4), namely $H(\omega)$, is dealt with next. Re-arranging equation (6),

$$\begin{aligned} H(\omega) &= \frac{2}{N}\sum_t[\alpha(\omega)\cos\omega t+\beta(\omega)\sin\omega t]\,e_t \\ &= \frac{2}{N}\sum_t\gamma_t(\omega)e_t, \end{aligned} \tag{12}$$

where

$$\alpha(\omega) = \sum_t(s_t-\overline{s})\cos\omega t$$

$$\beta(\omega) = \sum_t(s_t-\overline{s})\sin\omega t$$

$$\gamma_t(\omega) = \alpha(\omega)\cos\omega t+\beta(\omega)\sin\omega t. \tag{13}$$

It follows from equations (12) and (13) that

$$H(\omega)\sim N(0,v), \tag{14}$$

where $v(\omega)=4\sigma_e^2\sum_t\gamma_t^2(\omega)/N^2$.

From equation (13),

$$\begin{aligned} v(\omega) &= \frac{4\sigma_e^2}{N^2}\left[\alpha^2\sum_t\cos^2\omega t+\beta^2\sum_t\sin^2\omega t\right. \\ &\quad\left.+2\alpha\beta\sum_t\cos\omega t\sin\omega t\right] \\ &\approx \frac{2\sigma_e^2}{N}(\alpha^2+\beta^2) \\ &= 2\sigma_e^2 I_s(\omega). \end{aligned} \tag{15}$$

The noise spectrum I_e and the function H are uncorrelated, as can be seen from their definitions (5) and (6),

$$\begin{aligned} \mathrm{cov}\,[I_e(\omega),H(\omega)] &= \frac{2}{N^3}\sum_{i,\ell}\sum_{j,k}\cos(i-\ell)\omega\cos(j-k)\omega(s_j-\overline{s}) \\ &\quad\times E\,[(e_i-\overline{e})(e_\ell-\overline{e})(e_k-\overline{e})]=0 \end{aligned}$$

since the third moment of Gaussian noise is zero.

It is convenient to combine $I_s(\omega)$ and $H(\omega)$:

$$I_s(\omega)+H(\omega)\sim N\left[I_s(\omega),2\sigma_e^2I_s(\omega)\right].$$

If this is scaled by the mean periodogram,

$$I_s'(\omega)\equiv[I_s(\omega)+H(\omega)]/\overline{I}\sim N[I_s(\omega)/\overline{I},2I_s(\omega)/\overline{I}]. \tag{16}$$

The full scaled spectrum $I'(\omega)=I(\omega)/\overline{I}$ is thus the sum of the two uncorrelated spectra $I_e'(\omega)$ and $I_s'(\omega)$. It follows from equations (11) and (16) that

$$I'(\omega)\sim\exp(1)+N[I_s'(\omega),2I_s'(\omega)], \tag{17}$$

i.e. the scaled periodogram is distributed as the sum of uncorrelated exponentially distributed and normal variates.

An extensive discussion of the 'exponentially modified Gaussian distribution', of which equation (17) is a special case, can be found in Haney (2011). The probability density function (PDF) of this distribution at given ω is

$$f(x;\omega)=\exp\left(\mu+\frac{1}{2}\sigma^2-x\right)\Phi\left(\frac{x-\mu-\sigma^2}{\sigma}\right), \tag{18}$$

where $\mu=I_s'(\omega)$, $\sigma=\sqrt{2I_s'(\omega)}$ and $\Phi(\bullet)$ is the cumulative standard normal distribution. The mean and variance associated with this distribution are

$$\begin{aligned} E\,x(\omega) &= 1+\mu(\omega)=1+I_s'(\omega), \\ \mathrm{var}\,[x(\omega)] &= 1+\sigma^2(\omega)=1+2I_s'(\omega). \end{aligned} \tag{19}$$

The PDF (18) will serve as the basis for the investigation of the distribution of periodogram *maxima* which follows.

3 THE DISTRIBUTION OF THE SIGNAL PERIODOGRAM MAXIMUM

As pointed out above, for regularly spaced time points of observation, the signal spectrum I_s' is effectively non-zero only over the very narrow frequency range $B=(\nu_*-1/N,\nu_*+1/N)$. This means that

$$\begin{aligned} V=\max_\omega I'(\omega) &\approx \max_\omega[I_e'(\omega\backslash B),I'(B)] \\ &\approx \max_\omega[I_e'(\omega),I'(B)] \\ &= \max(W,X), \end{aligned} \tag{20}$$

where $\omega\backslash B$ is the set of all frequencies $0<\omega<\pi$ *excluding* those in the interval B. The random variables W and X are, respectively, the maxima of the noise and signal spectra. The distribution of W was discussed extensively in Koen (2015). If the periodogram is sampled only in the Fourier frequencies,

$$\nu_j=j/N,\quad j=1,2,\ldots N/2, \tag{21}$$

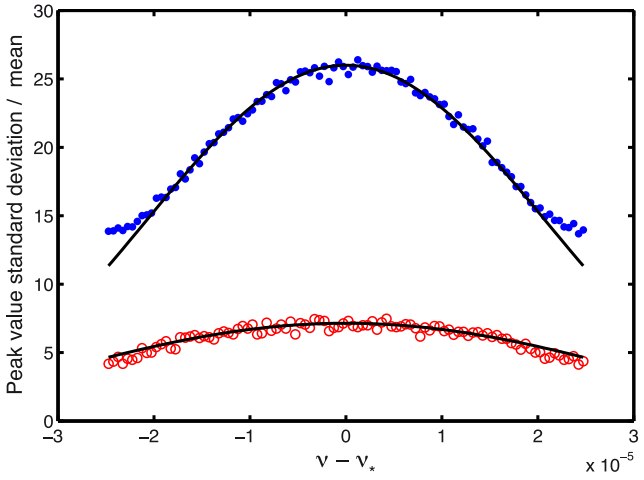

Figure 1. The means (solid dots) and standard deviations (open circles) of periodogram maxima, within 100 frequency bins. The plotted information is based on 30 000 simulated data sets with $N = 20\,000$ and $\mathcal{R} = 5$. Periodograms were only sampled in the Fourier frequencies (21). The lines show the theoretical results (19).

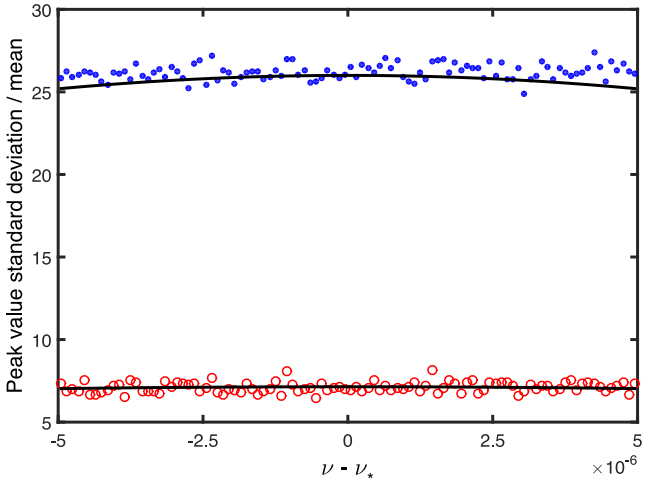

Figure 2. As for Fig. 1, but for periodograms oversampled by a factor $R = 4$.

then the cumulative distribution function (CDF) of W is

$$F_W(w) = [1 - \exp(-w)]^{N/2} \rightarrow \exp\left\{-\exp\left[-(w - \log N/2)\right]\right\}.$$

For noise spectra oversampled by a factor R,

$$\nu_j = j/[N(R+1)] \quad j = 1, 2, \ldots N(R+1)/2, \tag{22}$$

the distribution of the (scaled) periodogram maximum is of two-parameter Gumbel form. If $R \gtrsim 10$, then the limiting form

$$F_W(w) \approx \exp\left\{-\exp\left[-\frac{(w - 1.05\log N)}{1.04}\right]\right\} \tag{23}$$

is reached.

It remains to find $F_X(x)$, the distribution of the periodogram maximum over all frequencies in set B, i.e. in the neighbourhood of the signal frequency. Even in the absence of oversampling, two Fourier frequencies will typically lie in set B. For finer frequency sampling, the number of sampled values in B will increase, as will the correlation between them. This complicates the derivation of theoretical distributional results. In order to make progress with this challenging problem, we resort to a large simulation experiment.

Various combinations of sample size ($N = 10\,000$–$100\,000$), signal-to-noise ratio (3–7) and periodogram oversampling factor ($0 \leq R \leq 10$) were selected, and typically 30 000 simulated data sets per parameter combination were generated. For each simulation random values of the frequency and phase in equation (2) were generated, uniformly distributed over (0, 0.5) and $[0, 2\pi]$, respectively.

Results for the parameter combination $N = 20\,000$, signal-to-noise $\mathcal{R} = 5$, and no oversampling, are plotted in Fig. 1. The frequencies at which the signal periodogram maximum occurred differ by at most $0.5/N$ from the true value ν_*. The range $(\nu_* - 0.5/N,\ \nu_* + 0.5/N$ was therefore divided into 100 bins, and the mean and standard deviation of the ∼300 periodogram peak values in each bin calculated. The solid lines in Fig. 1 demonstrate that results conform to those predicted by equation (19), except near the extremes of $|\nu - \nu_*|$. This latter deviation can be explained as follows: if the true frequency lies about mid-way between two Fourier frequencies, then the position of the largest peak will be determined by the relative values of the noise spectra at the two Fourier frequencies. In other words, the periodogram peak value will not reflect the *mean* noise level, but the larger of two values. Clearly, on average, this will mean that peak values are inflated near the frequency extremes in Fig. 1. Simulation with $\mathcal{R} = 3$ confirm that the relative size of the effect increases with decreasing signal-to-noise.

The situation is rather different if the periodogram is oversampled. Since the frequency resolution is improved, the frequency range over which periodogram peak values occur is reduced – to be exact it becomes $|\nu - \nu_*| \leq 0.5/[N(R+1)]$. Fig. 2, for $R = 4$, shows that the variance of the peak heights is still accurately given by equation (19), whereas the simulated mean values are systematically slightly larger than those predicted by the equation. This can again be ascribed to boosting by noise.

An important point made by Fig. 2 is that there is very little variation with frequency of the means and variances. This suggests that the sampling distribution of the oversampled peaks may be largely independent of frequency, at least for $R \geq 4$ or so. Furthermore, experimentation showed that

$$U = \sqrt{X - 1} - \mathcal{R} \tag{24}$$

is a useful transformation of X to work with. Figs 3 and 4 show means and standard deviations of U calculated from periodogram peak values from a large number of simulation experiments. Lines connect results for fixed sample sizes N, and different lines are for different values of the signal-to-noise ratio $\mathcal{R}$. Lines lie in two groups, one corresponding to $\mathcal{R} = 3$, the other for all larger signal-to-noise. This is shown in more detail in Figs 5 and 6. For $R \geq 10$ and $\mathcal{R} \geq 4$ mean values of U are very close to zero, and values of σ_U cluster between 0.70 and 0.72. Kolmogorov–Smirnov goodness-of-fit tests show that the null hypothesis that

$$U \sim N(0, 0.715^2) \tag{25}$$

cannot be rejected for any of the parameter combinations tested (selected values in the ranges $4 \leq \mathcal{R} \leq 6$, $10\,000 \leq N \leq 100\,000$, $R = 10$–20). The two smallest p-values found were 0.06 and 0.09, amongst the 20 tests performed.

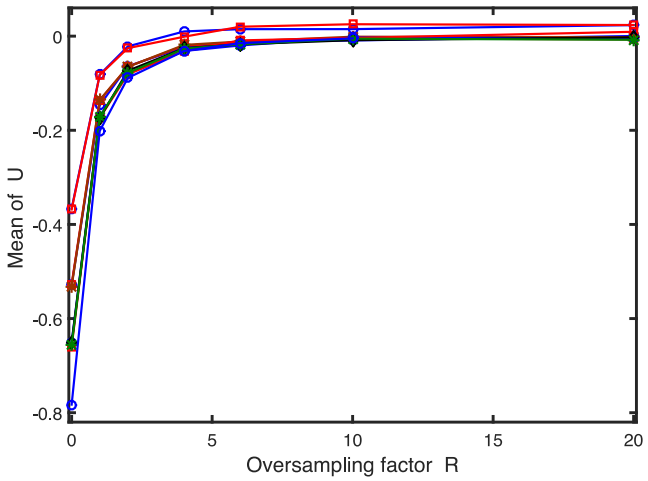

Figure 3. Mean values of the transformed maxima of signal spectra, from simulations. Each point is the mean over typically 30 000 simulations, for one combination of sample size, signal-to-noise ratio and periodogram oversampling factor. Lines connect results for the same values of N. Note the limiting behaviour as the oversampling factor is increased.

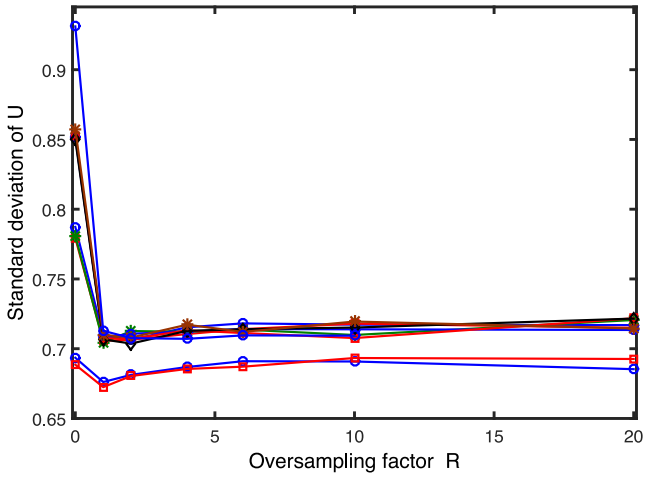

Figure 4. As for Fig. 3, but showing the standard deviations of the transformed periodogram peak values.

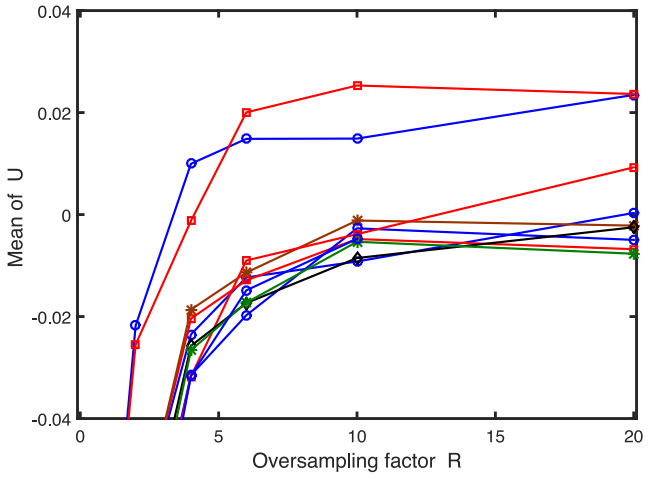

Figure 5. Detail of Fig. 3, at large R. The top two lines are for signal-to-noise ratios of 3; the collection of lower lines are for $\mathcal{R} = 4-6$.

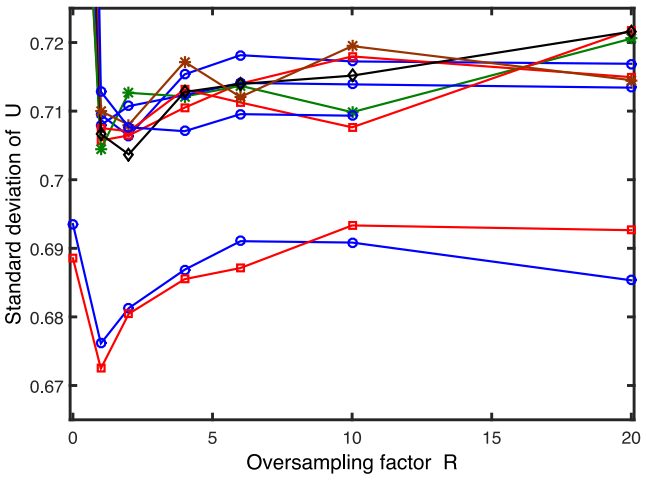

Figure 6. Detail of Fig. 4, at large R. The bottom two lines are for signal-to-noise ratios of 3; the collection of upper lines are for $\mathcal{R} = 4-6$.

Equations (24) and (25) imply that at high signal-to-noise ratios the distribution of maxima of I'_s is well described by the PDF,

$$f_X(x) = \frac{1}{1.43\sqrt{2\pi(x-1)}} \exp\left[-\frac{1}{2}\left(\frac{\sqrt{x-1}-\mathcal{R}}{0.715}\right)\right] \quad x > 1. \tag{26}$$

For low values of $\mathcal{R}$ the signal spectrum is no longer dominant and equation (26) no longer a good approximation.

Note that the transformation in equation (24) implies that U is a scaled and shifted version of the standardized *amplitude* spectrum

$$S(\omega) = 2\sqrt{I(\omega)/N}.$$

Equation (25) implies that for oversampled amplitude spectra of substantial signal-to-noise sinusoids, the spread of peak heights is a constant, independent of parameters such as sample size.

4 THE PROBABILITY OF CORRECTLY IDENTIFYING THE SIGNAL FREQUENCY

Clearly the probability of identifying the signal frequency is given by the probability that the largest peak value associated with the signal, is larger than the largest noise peak, i.e.

$$p = P(X > W).$$

This probability is given by

$$\begin{aligned} p &= \int_0^\infty f_X(x) \int_0^x f_W(w)\,\mathrm{d}w\,\mathrm{d}x \\ &= \int_0^\infty f_X(x) F_W(x)\,\mathrm{d}x. \end{aligned} \tag{27}$$

In the case of fully oversampled spectra, f_X is given by equation (26) (at least for $\mathcal{R} > 3$) and F_W by equation (23) so that

$$p = \frac{1}{1.43\sqrt{2\pi}} \int_1^\infty (x-1)^{-1/2} \exp\left\{ \left[-\frac{1}{2}\left(\frac{\sqrt{x-1}-\mathcal{R}}{0.715}\right)\right] - \exp\left[-\frac{(x-1.05\log N)}{1.04}\right] \right\} \mathrm{d}x. \tag{28}$$

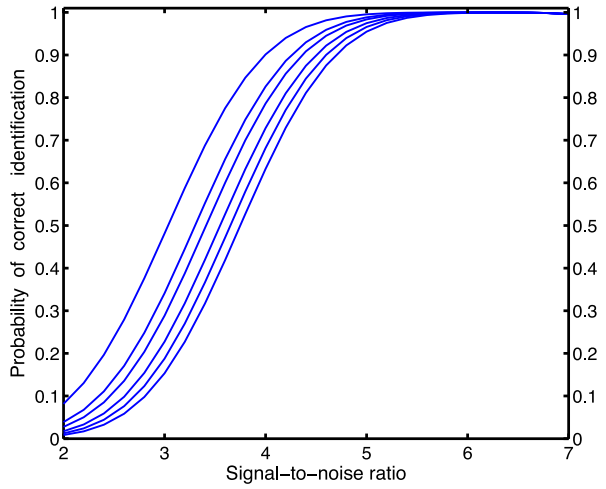

Figure 7. The probability of correctly identifying the signal frequency, for various signal-to-noise ratios $\mathcal{R}$ and data set sizes N, for fully oversampled spectra. From top to bottom the curves are for $N = 10\,000$, 50 000, 100 000, 250 000, 500 000 and 1000 000.

Values of p were simulated by generating 20 000–40 000 artificial data sets with various combinations of $10\,000 \leq N \leq 100\,000$, $3 \leq \mathcal{R} \leq 6$ and oversampling factors of $R = 10$, 20. Comparison with the predictions of equation (28) gave a maximum difference of $|\Delta p| = 0.008$. This suggests that equation (28) can also be used for $\mathcal{R}$ somewhat smaller than 4.

Perhaps more usefully, equation (28) can be used to ascertain, for a given value of the sample size, the signal-to-noise ratio required for secure signal frequency determinations – see Fig. 7.

5 AN ILLUSTRATIVE APPLICATION

The theory above is demonstrated by application to $N = 40\,000$ *Kepler* brightness measurements of the star KIC 8008067, taken during the tenth quarter of operation of the mission. The data were obtained in short cadence mode, i.e. a measurement was made every 58.84 s. The series mean was subtracted, and missing observations were replaced by zeros.

An *amplitude* spectrum of the data can be seen in the top panel of Fig. 8. In addition to prominent peaks at a few isolated frequencies this scaled version of the periodogram clearly shows a general excess of power at the lowest frequencies, visible as a broad bump. (Ordinary periodogram plots are dominated by the large peaks, hence the low-frequency power excess is less obvious.) A possible origin for the broad power excess is correlated noise. This is confirmed by autocorrelation and partial autocorrelation functions plots, which suggest that the noise e_t may be an autoregressive type of time series, of order 2,

$$e_t = \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \varepsilon_t,$$

where α_1 and α_2 are constants, and ε_t is white noise. Regressing y_t on y_{t-1} and y_{t-2} gives the estimates $\hat{\alpha}_1 = 0.195$ and $\hat{\alpha}_2 = 0.176$, both with standard errors of 0.0045.

The amplitude spectrum of the filtered series

$$r_t = y_t - \hat{\alpha}_1 y_{t-1} - \hat{\alpha}_2$$

is plotted in the bottom panel of Fig. 8. The general low-frequency power excess has clearly been efficiently whitened from the data.

We proceed to systematically pre-whiten frequencies from the spectrum in the bottom panel of Fig. 8. This is done by (i) identifying the frequency ω at which the spectrum is at a maximum; (ii) fitting, by least squares, a sinusoid with this frequency to the data; (iii) subtracting the fitted sinusoid; (iv) calculating the spectrum of the residuals, and repeating steps (i)–(iv).

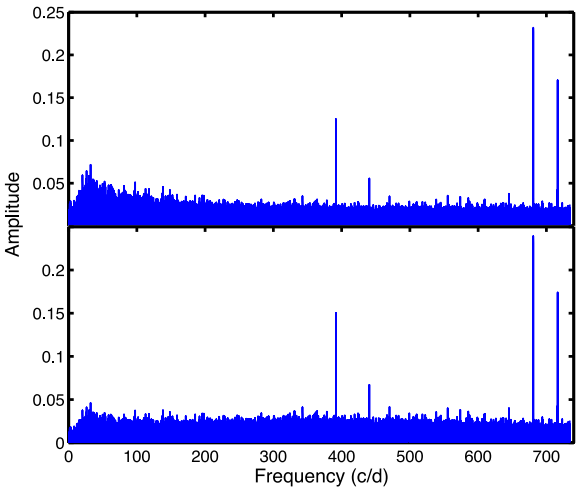

Figure 8. Amplitude spectra of 40 000 short cadence *Kepler* observations of the star KIC 8008067, obtained during quarter 10. The top panel spectrum is of the raw data; note the power excess over a wide low-frequency interval. The lower panel spectrum follows after application of a linear filter to the data.

Table 1. The results of successive pre-whitening of the filtered KIC 8008067 data. The columns headed 'Maximum' and 'Mean', respectively, contain the height of the largest spectral ordinate, and the mean over all spectral values. The estimated signal-to-noise ratio $\mathcal{R}$ for the largest spectral peak is $\sqrt{\pi}/2$ times the ratio of peak height to the spectrum mean – see the text for a full explanation. The last column is the probability that the peak position corresponds to a true signal frequency. Note that the removal of large spectral peaks has very little influence on the value of the spectrum mean.

	Frequency (d^{-1})	Maximum	Mean	Estimated $\mathcal{R}$	p
1	680.7609	0.2392	0.0104	20.5795	1.000
2	680.7242	0.1746	0.0104	15.0194	1.000
3	716.3978	0.1739	0.0104	14.9612	1.000
4	391.5798	0.1503	0.0103	12.9308	1.000
5	680.8013	0.1285	0.0103	11.0521	1.000
6	680.8343	0.0698	0.0103	6.0054	1.000
7	440.5286	0.0670	0.0103	5.7683	1.000
8	680.6948	0.0531	0.0103	4.5660	0.959
9	32.1529	0.0459	0.0103	3.9511	0.821

The results are summarized in Table 1. Note that for the periodogram

$$E\left[\max_{\omega} I(\omega)/\overline{I}\right] = \mathcal{R}^2,$$

whereas for the amplitude spectrum

$$E\left[\max_{\omega} S(\omega)/\overline{S}\right] = \frac{2}{\sqrt{\pi}}\mathcal{R} = 1.128\mathcal{R}$$

(where E is the expectation, i.e. ensemble average, operator). In order to interpret the results in Table 1 it is therefore necessary to read the probabilities in Fig. 7 which correspond to signal-to-noise ratios $0.89 \max(S)/\overline{S}$.

Inspection of the table shows that five of the eight reliable frequency detections have very similar values near 680.7 d^{-1},

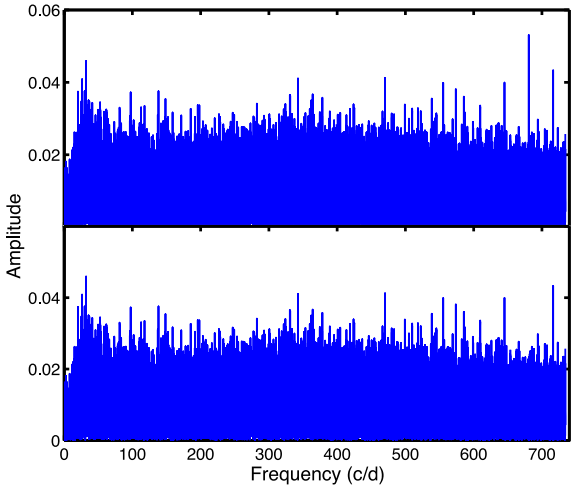

Figure 9. Spectra of the residuals, after pre-whitening 7 (top panel) or 8 (bottom panel) frequencies from the filtered observations of KIC 8008067.

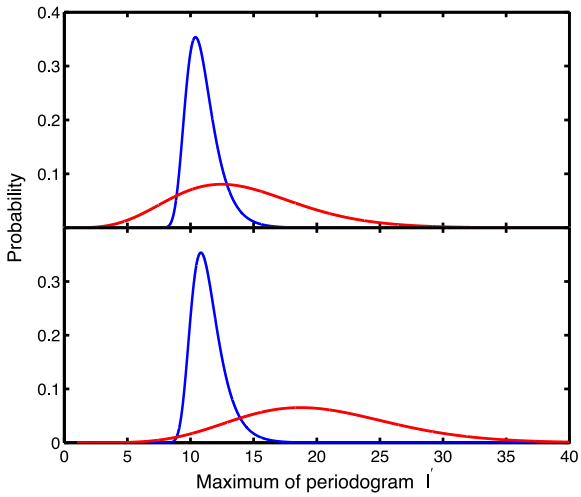

Figure 10. PDFs for the heights of the largest noise-induced peaks (sharply peaked blue curves) and signal peak maxima (broad red curves). In both panels the ratio $A/\sigma_e = 0.05$, but $N = 20\,000$ in the top panel, $N = 30\,000$ in the bottom panel.

suggesting a strong feature in the data with somewhat variable amplitude, frequency, and/or phase. Baran (2013) has in fact identified this frequency, as well as the other three 'significant' detections in Table 1 as being due to *Kepler* artefacts. In this sense, the features are truly present in the data. By contrast the spectrum in fig. 15 of Baran (2013), which covers the relevant frequency interval, shows no power excess at the last frequency in Table 1. Fig. 9 shows the spectra in which frequencies 8 and 9 are, respectively, the most prominent.

Note that in the above the influence of multiple tests on the overall significance level was not taken into account, in order to keep the discussion to the point and simple. In practice this issue should not be ignored (although it usually is).

6 A CONCLUDING REMARK

A look at Fig. 7 shows that the required signal-to-noise ratio for secure signal frequency detection increases markedly with increasing sample size. At first glance this is counter-intuitive, since it suggests that fewer data may be better. However, it should be borne in mind that the calculations above were done at fixed values of $\mathcal{R}$. In practice, for fixed A/σ_e, $\mathcal{R}$ increases with N as $N^{1/2}$. Given that the mean value of x associated with equation (26) is $Ex = 1 + 0.715^2 + \mathcal{R}^2$, this means that Ex is roughly proportional to N for even moderate $\mathcal{R}$. By contrast the mean of the maximal noise peak distribution (23) is given by $Ew = 1.05\log N + 1.04\gamma$ (γ being Euler's constant). The rate of increase of Ew with increasing N is therefore considerably slower than that of Ex.

The point is illustrated in Fig. 10, based on fixed $A/\sigma_e = 0.05$. In the top panel $N = 20\,000$, giving $\mathcal{R} = 3.54$, and leading to a probability $p = 0.70$ that the frequency of the sinusoid will be correctly identified in a spectrum. Increasing N to 30 000 (bottom panel) increases $\mathcal{R}$ to 4.33, and p to 0.93.

ACKNOWLEDGEMENTS

This research was partially funded by a South African National Research Foundation grant. The author thanks Dr Andy Baran for supplying the *Kepler* measurements of KIC 8008067 in a form suitable for easy analysis.

REFERENCES

Baran A. S., 2013, Acta Astron., 63, 203
Baran A. S., Koen C., Pokrzywka B., 2015, MNRAS, 448, L16
Haney S., 2011, PhD thesis, Drexel University
Koen C., 2015, MNRAS, 449, 1098

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