The Periodogram

Sample covariance between X and $\sin(2\pi\omega t + \phi)$ is

$$\frac{1}{T}\sum_{t=0}^{T-1} X_t \sin(2\pi\omega t+\phi) - \bar{X}\frac{1}{T}\sum_{t=0}^{T-1}\sin(2\pi\omega t+\phi)$$

Use identity $\sin(\theta) = (e^{i\theta} - e^{-i\theta})/(2i)$ and formulas for geometric sums to compute mean.

When $\omega = k/T$ for an integer k, not 0, we find that $\sum_{t=0}^{T-1}\sin(2\pi\omega t+\phi) = 0$.

So sample covariance is simply

$$\frac{1}{T}\sum_{t=0}^{T-1} X_t \sin(2\pi\omega t+\phi)\,.$$

For these special ω we can also compute

$$\sum_{t=0}^{T-1}\sin^2(2\pi\omega t+\phi) = T/2.$$

So correlation between X and $\sin(2\pi\omega t+\phi)$ is

$$\frac{\frac{1}{T}\sum_{t=0}^{T-1} X_t \sin(2\pi\omega t+\phi)}{s_x\sqrt{1/2}}$$

where s_x^2 is sample variance $\sum(X_t-\bar{X})^2/T$.

Adjust ϕ to maximize this correlation.

The sine can be rewritten as

$$\cos(\phi)\sin(2\pi\omega t)+\sin(\phi)\cos(2\pi\omega t)$$

so choose coefficients a and b to maximize correlation between X and

$$a\sin(2\pi\omega t)+b\cos(2\pi\omega t)$$

subject to the condition $a^2+b^2=1$.

Correlations are scale invariant so drop condition on a and b and maximize the correlation between X and the linear combination of sine and cosine.

Problem solved by linear regression. Coefficients given by $(M^TM)^{-1}M^TX$:

M is T by 2 design matrix full of sines and cosines.

Get $M^T M = \frac{T}{2} I_{T\times T}$; regression coefficients are

$$a = \frac{2}{T} \sum_{t=0}^{T-1} X_t \sin(2\pi\omega t)$$

and

$$b = \frac{2}{T} \sum_{t=0}^{T-1} X_t \cos(2\pi\omega t)\,.$$

Covariance between X and best linear combination is

$$\frac{1}{T}\left\{a \sum_{t=0}^{T-1} X_t \sin(2\pi\omega t) + b \sum_{t=0}^{T-1} X_t \cos(2\pi\omega t)\right\} = (a^2 + b^2)/2.$$

But in fact

$$a^2 + b^2 = \left|\frac{1}{T} \sum_{t=0}^{T-1} X_t \exp(2\pi\omega t i)\right|^2$$

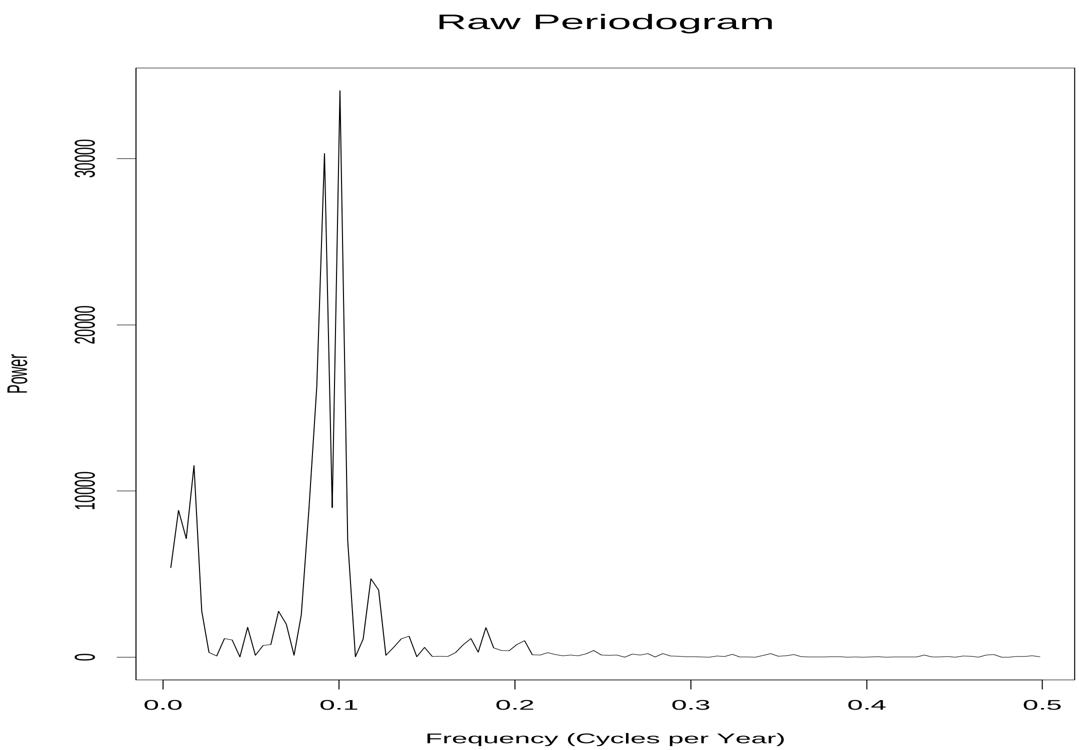
is modulus of DFT $\widehat{X}(\omega)$ divided by T.

Defn: Periodogram is function

$$|\hat{X}(\omega)|^2$$

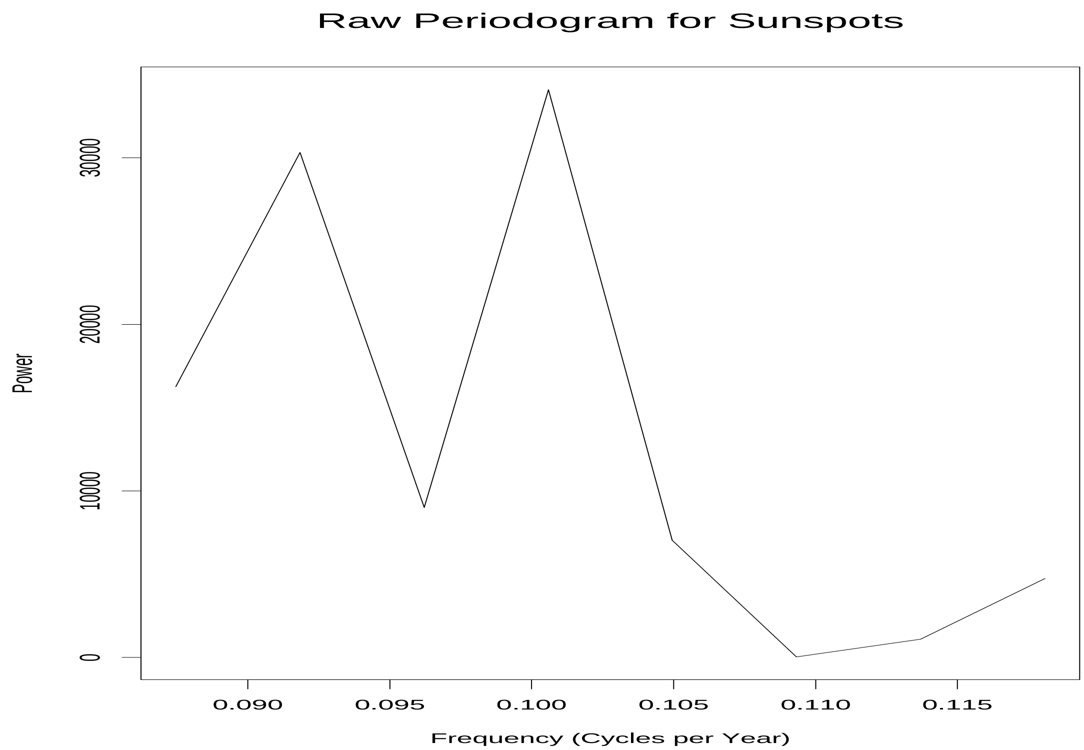
Some periodogram plots:

- $|\hat{X}|$ vs frequency for sunspots minus mean



Notice peak at frequency slightly below 0.1 cycles per year as well as peak at frequency close to 0.03.

Plot only for frequencies from 1/12 to 1/8 which should include the largest peak.

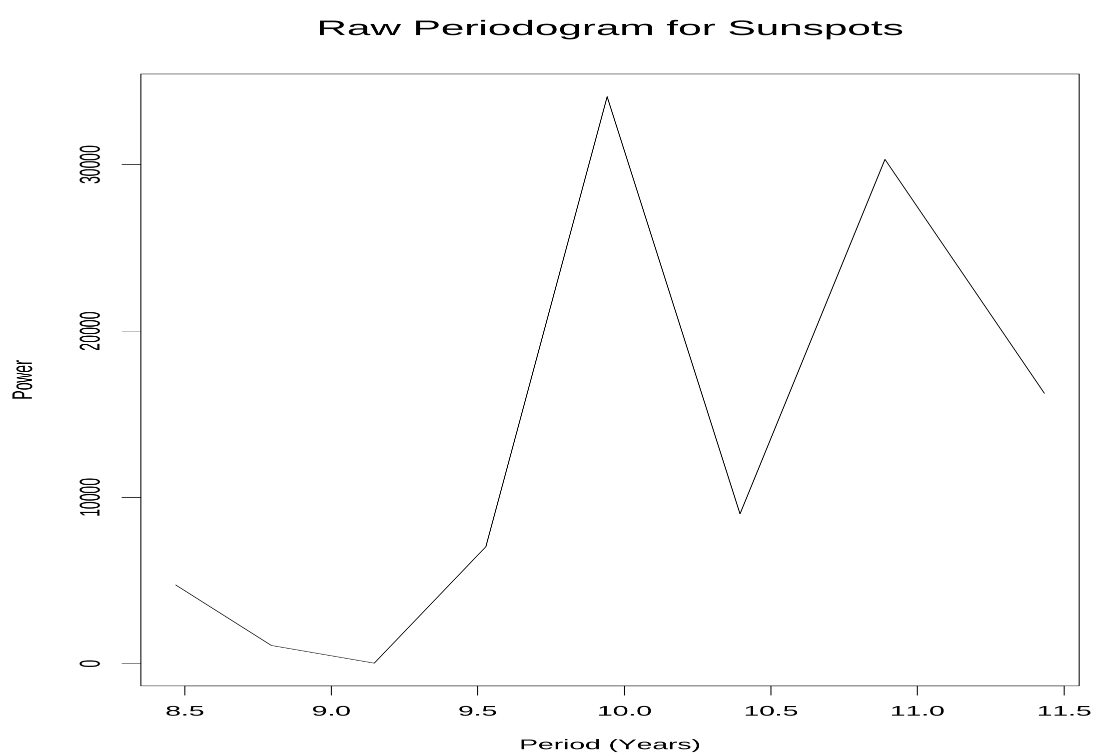


Notice: picture clearly piecewise linear.

Actually using DFT: computes sample spectrum only at frequencies of form k/T (in cycles per point) for integer values of K.

There are only about 10 points on this plot.

Same plot against period ($= 1/\omega$) shows peaks just below 10 years and just below 11.



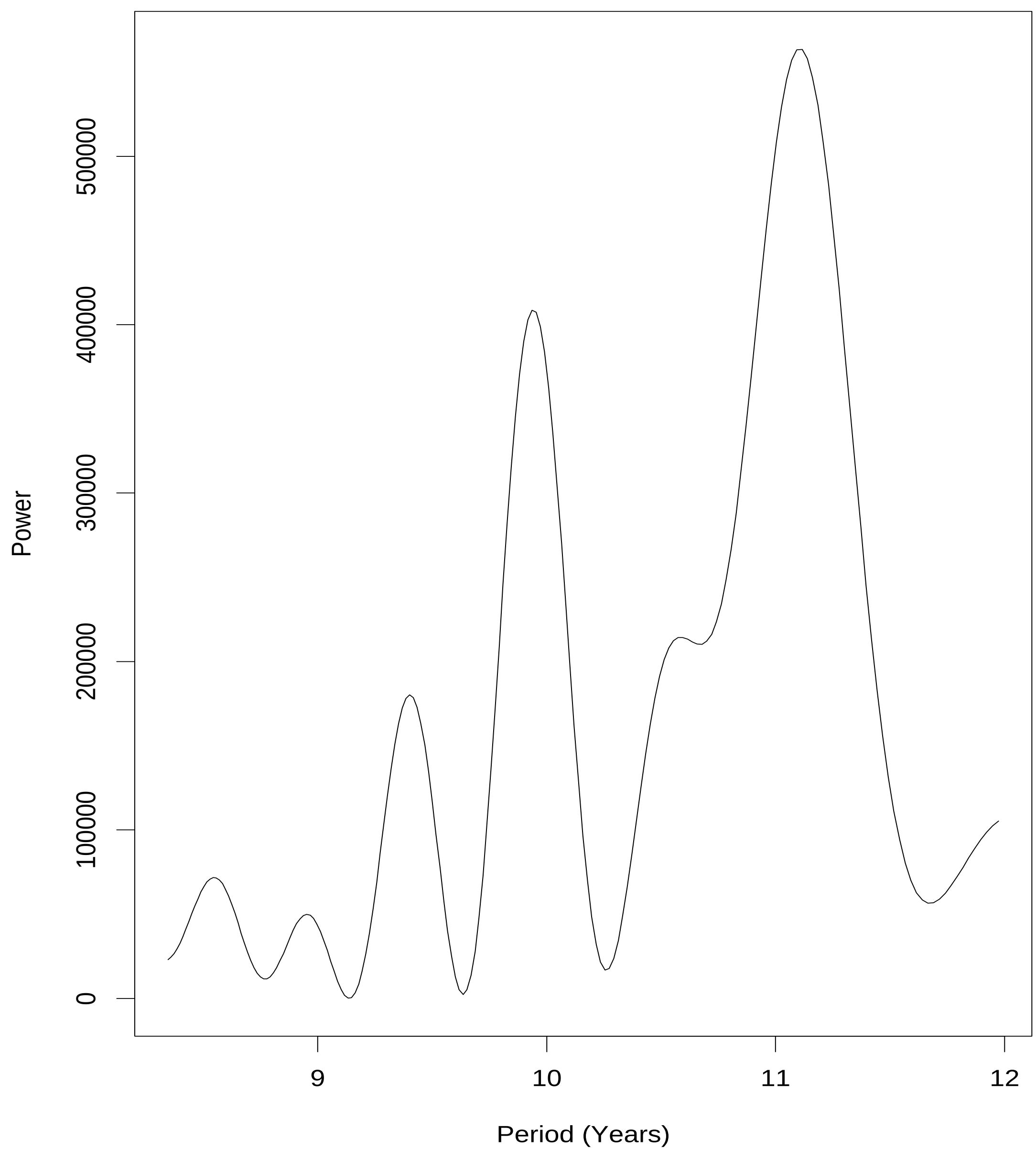
DFT can be computed very quickly at special frequencies but to see structure clearly near a peak need to compute $\hat{X}(\omega)$ for a denser grid of ω.

Use S-Plus function

```
transform<- function(x, a, b, n = 100)
{
    f <- seq(a, b, length = n)
    nn <- 1:length(x)
    args <- outer(f, nn, "*") * 2 * pi
    cosines <- t(cos(t(args)) * x)
    sines <- t(sin(t(args)) * x)
    one <- rep(1, length(x))
    ((cosines %*% one)^2
       + (sines %*% one)^2)/length(x)
}
```

to compute lots of values for periods between 8 and 12 years.

Plot of Spectrum vs Period for Sunspots

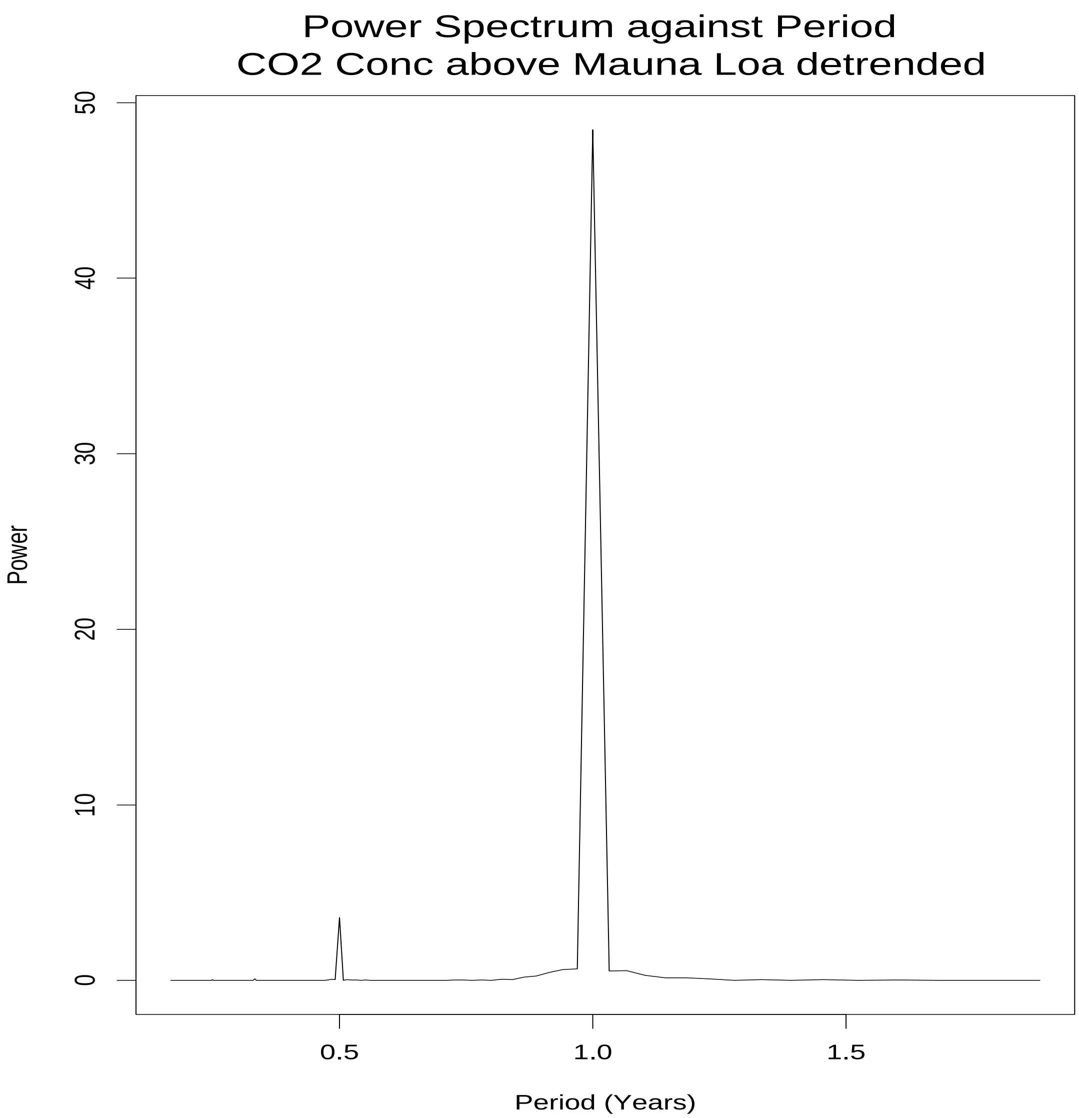


Periodogram for CO_2 above Mauna Loa: Linear trend removed by linear regression.

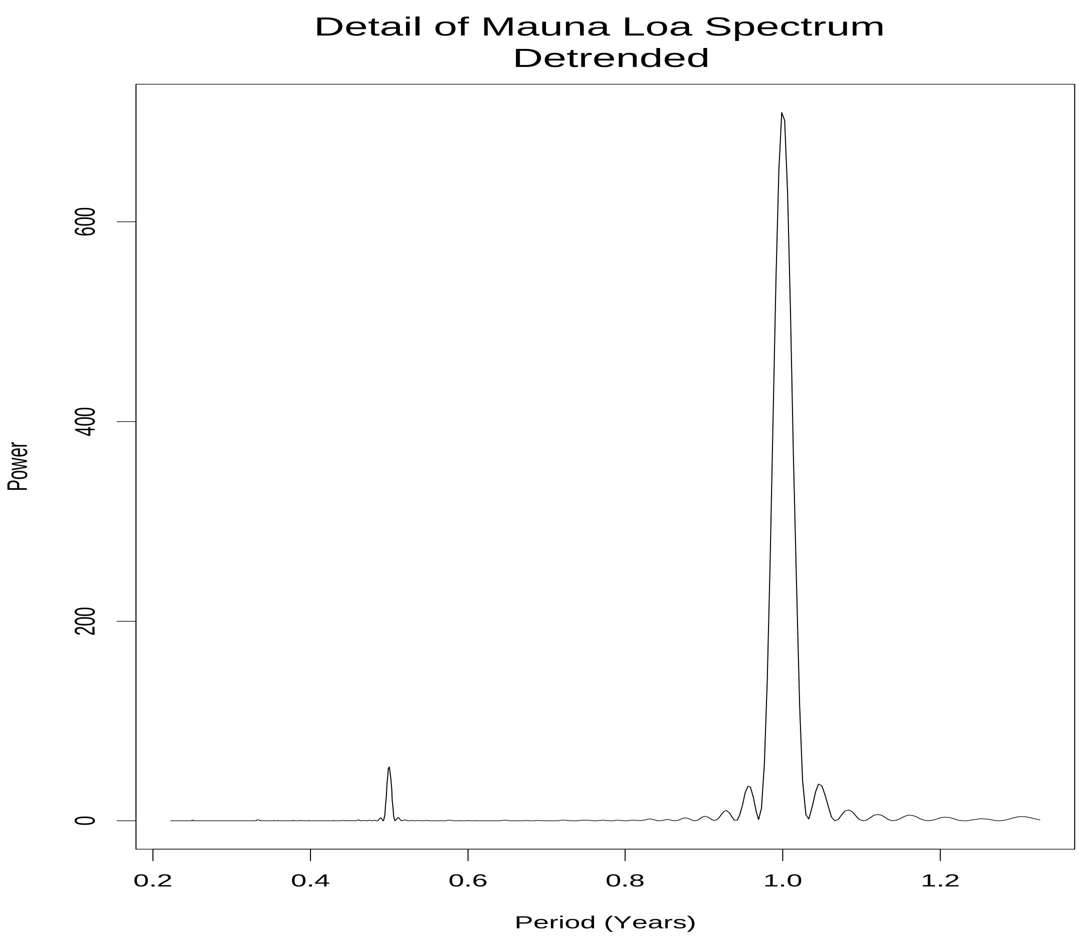
Note peaks at periods of 1 year and 6 months.

Peaks show clear annual cycle.

Annual cycle not simple sine wave – contains overtones: components whose frequency is integer multiple of basic frequency of 1 cycle per year.


Power Spectrum against Period
CO2 Conc above Mauna Loa detrended
Power
0
10
20
30
40
50
0.5
1.0
1.5
Period (Years)

Now a detail of this image:



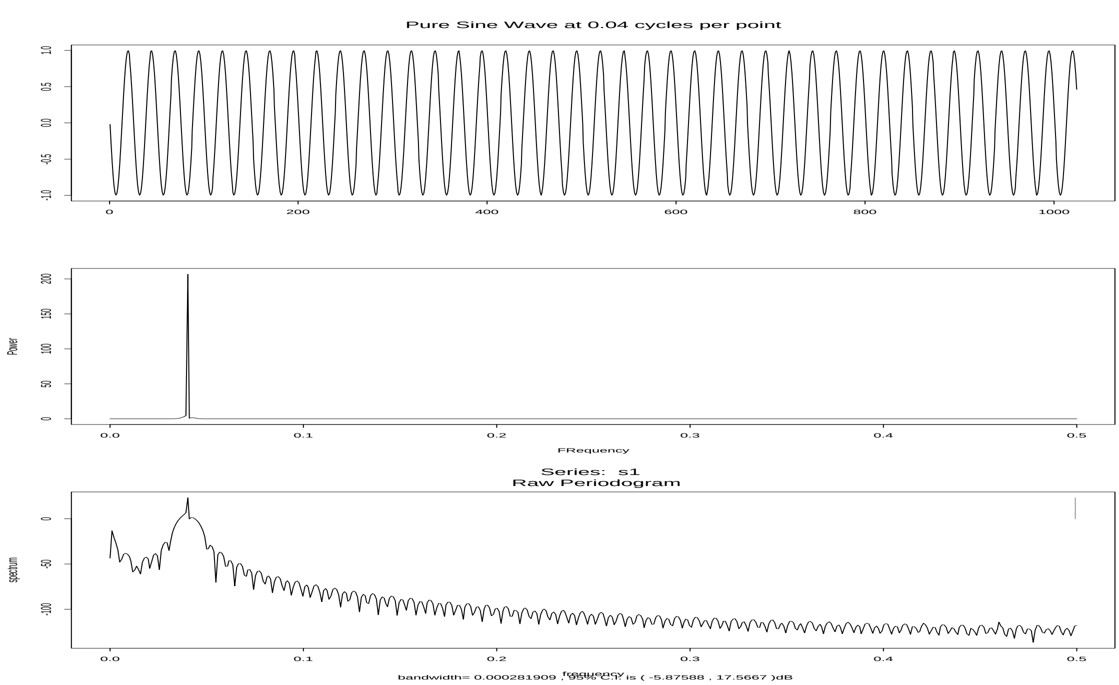
Periodogram of various generated series which have exact sinusoidal components.

First a pure sine wave with no noise.

Middle panel: periodogram.

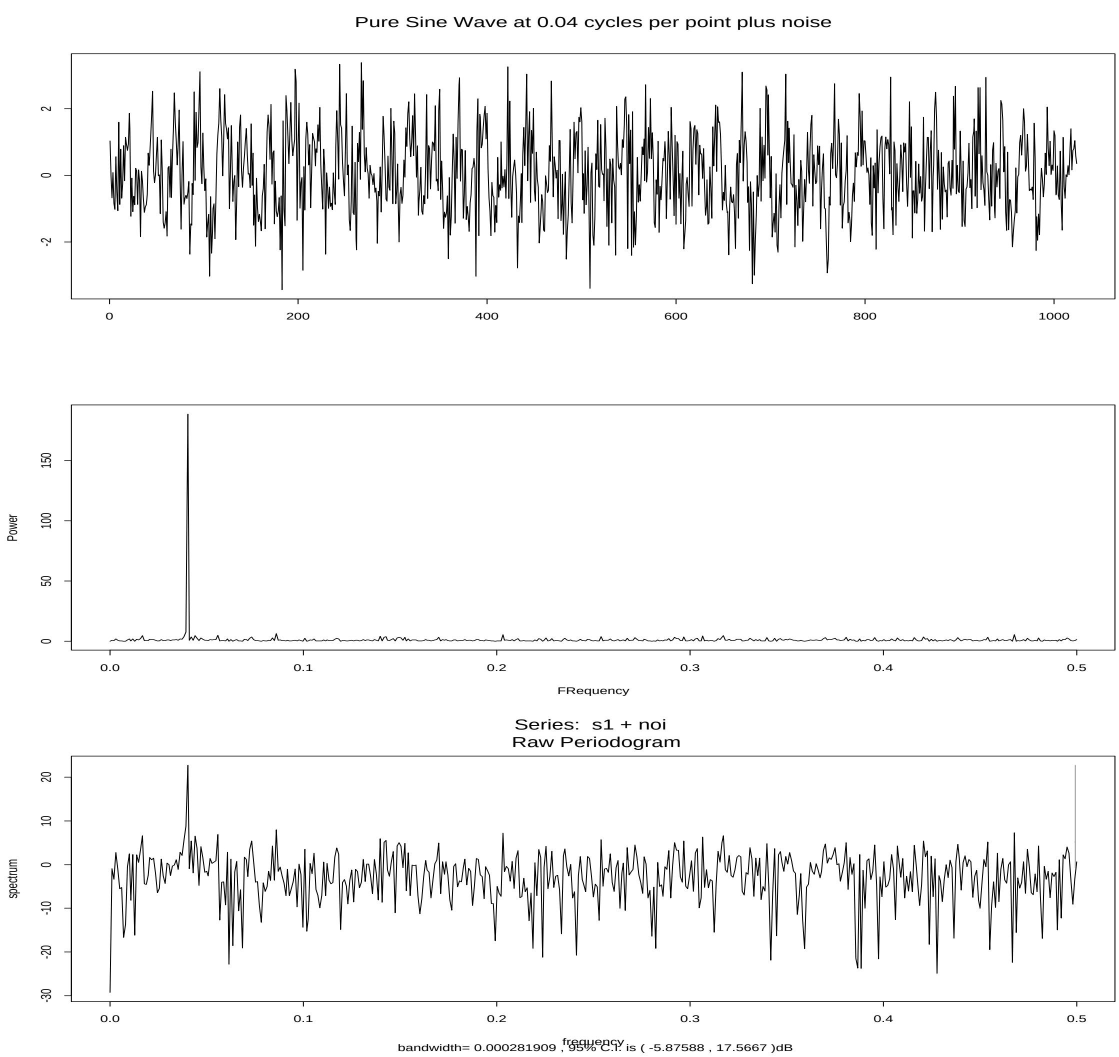
Lower panel: $\log_{10}(|\hat{X}(\omega)|) * 10$.

Apparent waves: round off error $\log(\approx 0)$.

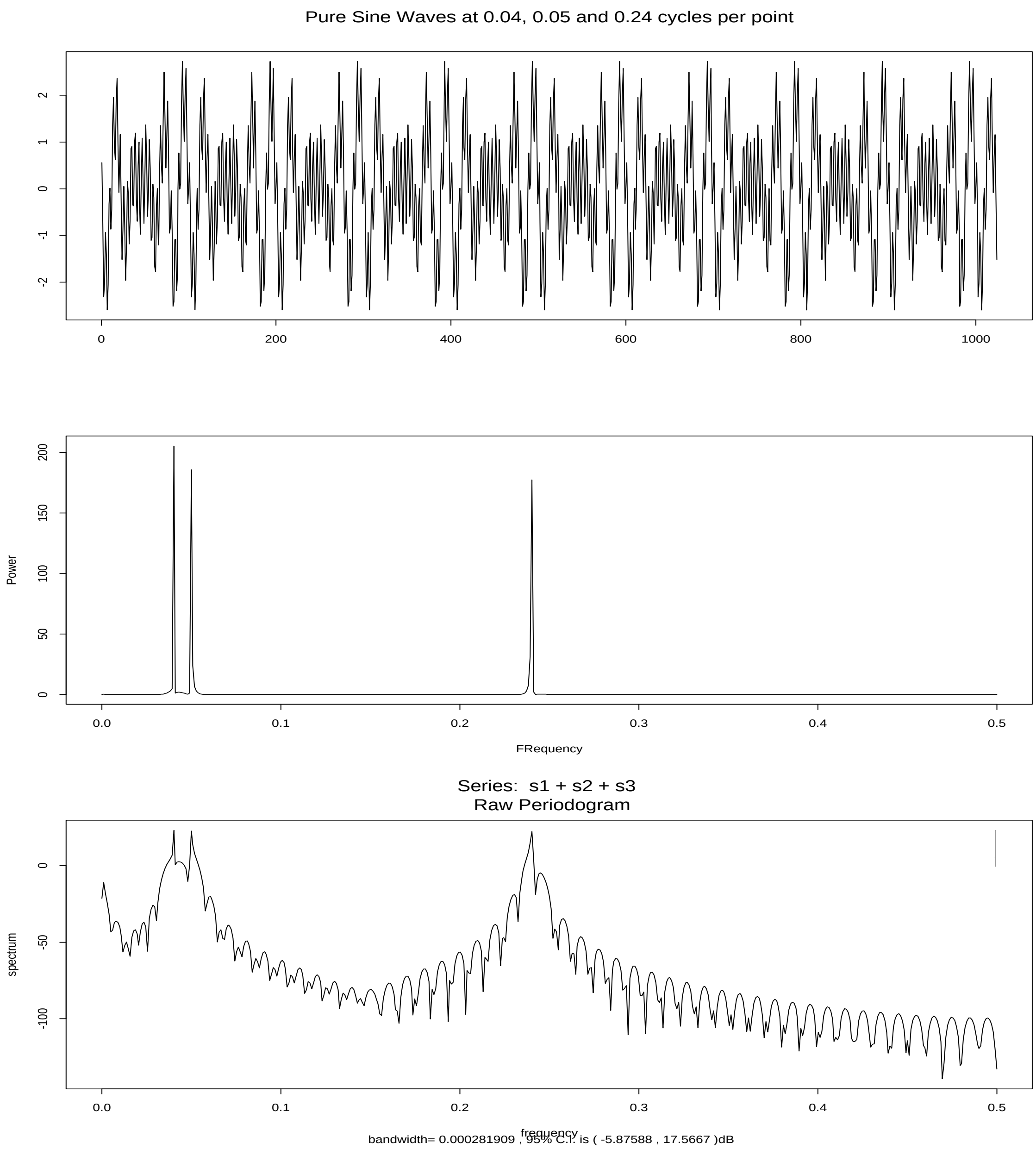


Same series plus N(0,1) white noise.

Note: much harder to see perfect sine wave in data but periodogram shows presence of sine wave quite clearly.

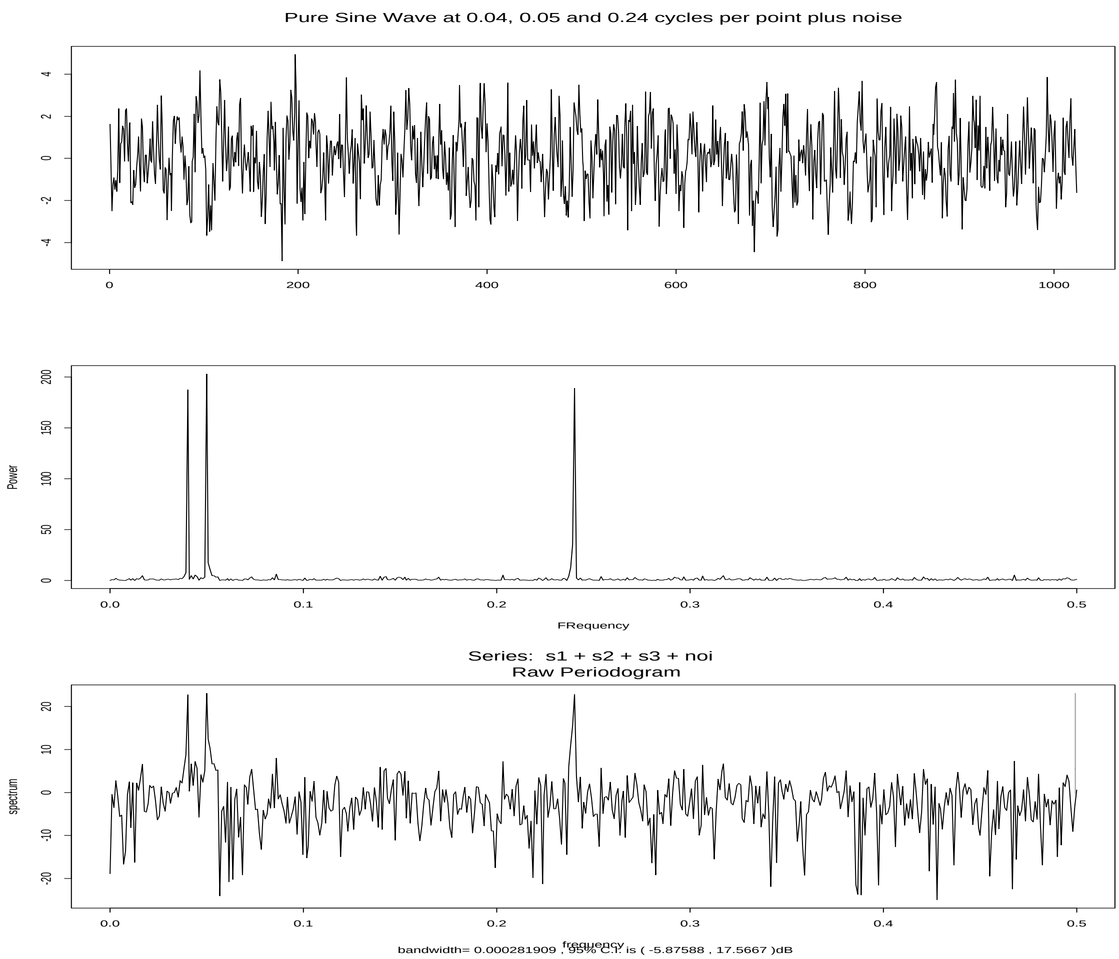


The sum of three sine waves.



Now add N(0,1) white noise. Periodogram still picks out each of component.

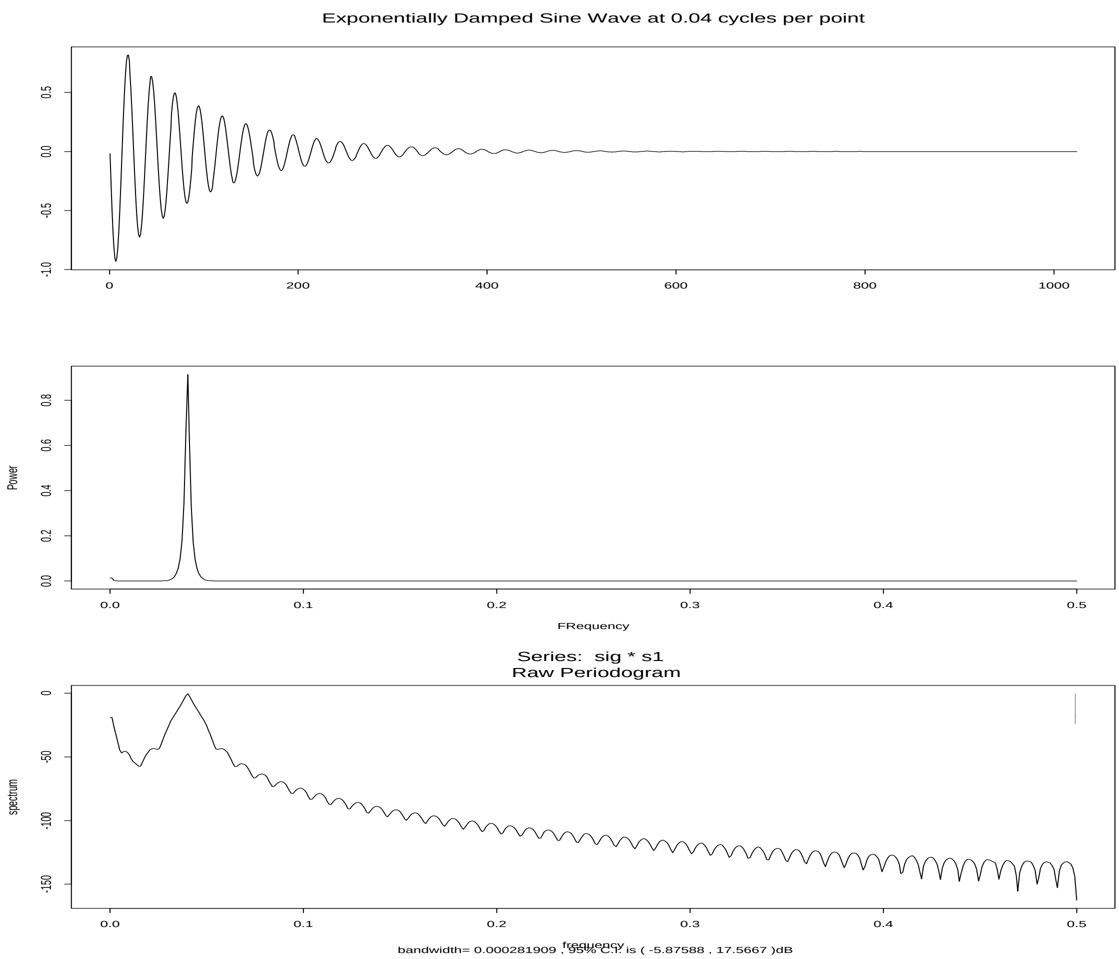
The sum of three sine waves.



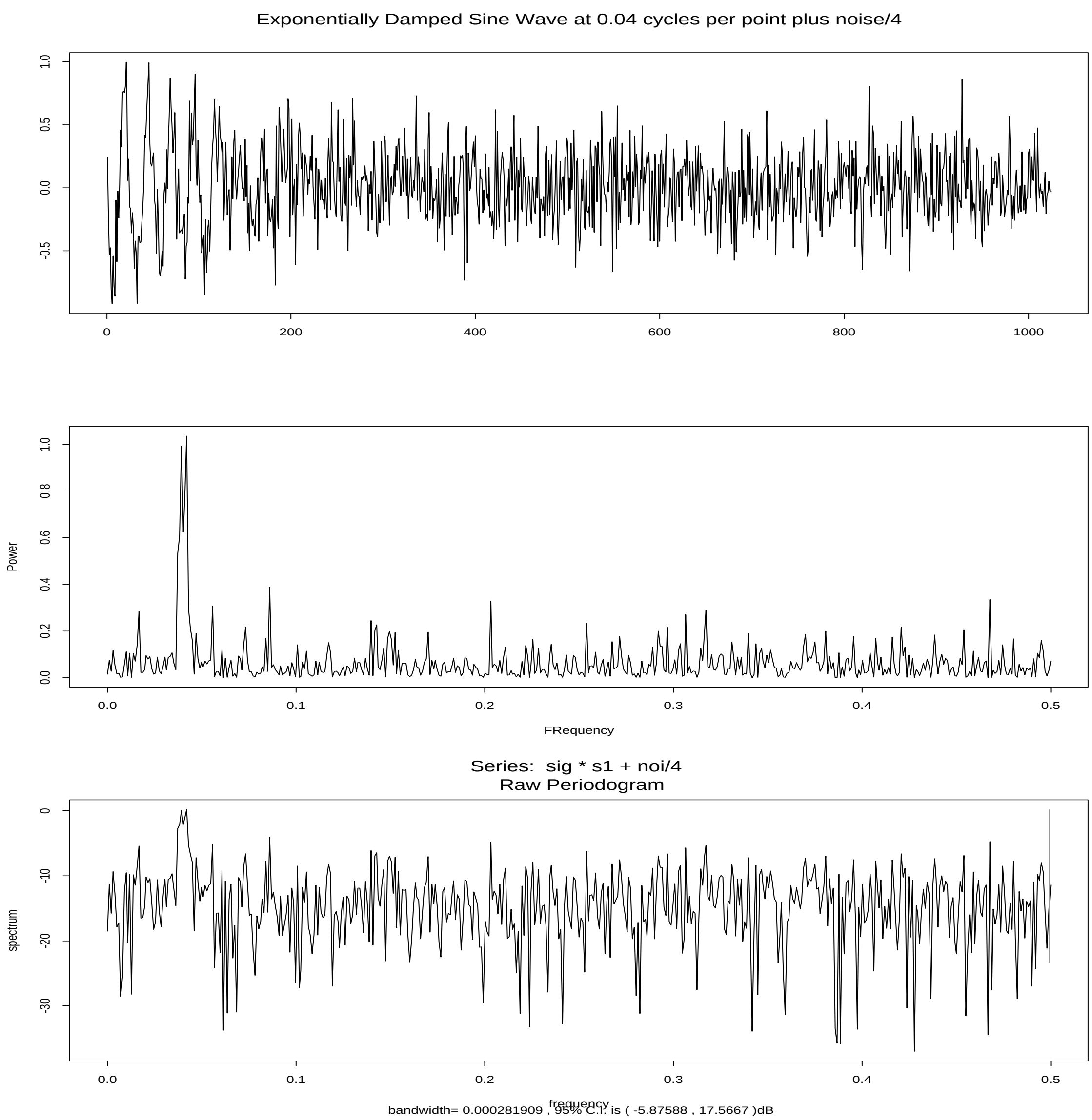
Multiply sine wave by damping exponential.

Signal gone ≈ quarter of way through series.

Periodogram peak still at 0.04 cycles per point.



With noise added can still see effect. But compare the scales on the middle plots between all these series.

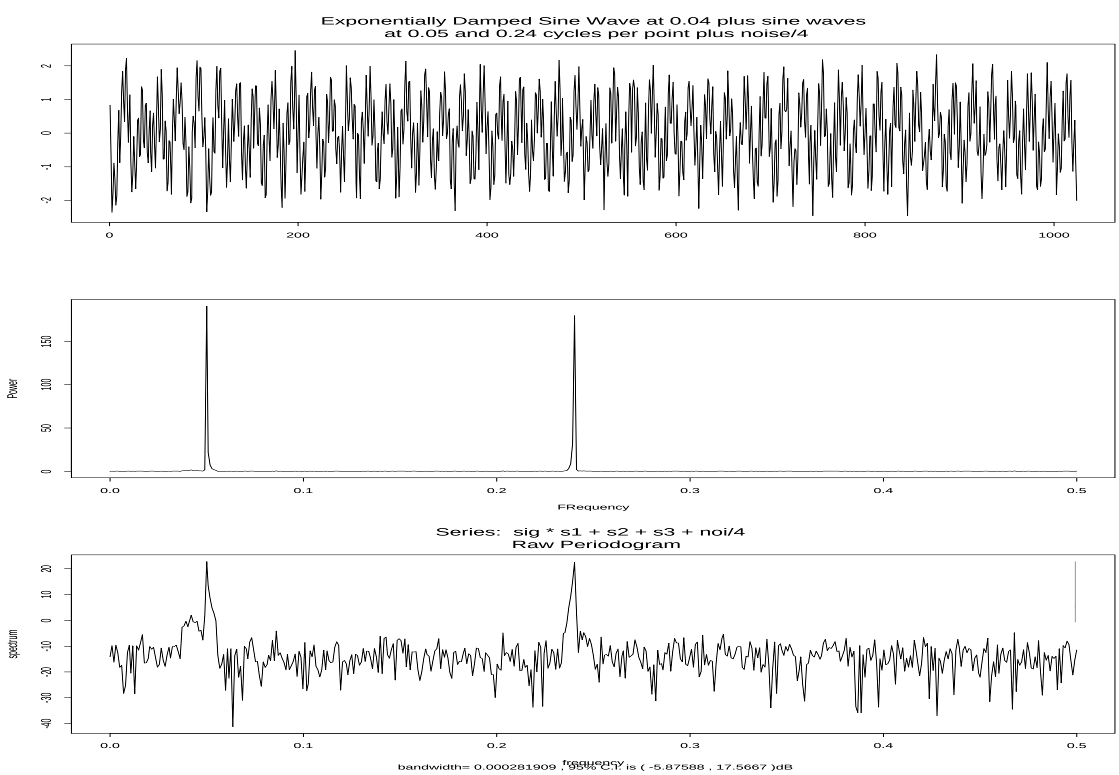


Exponentially damped sine wave plus two sine waves with N(0,1/16) noise.

Only two peaks visible in raw periodogram.

On logarithmic scale: hump on left of peak at 0.05 which is peak at 0.04.

Raw scale can make small secondary peaks invisible.



Behaviour of DFT when sinusoid present.

$$X_t = A\cos(2\pi\theta t + \phi) + Y_t$$

where Y is mean 0 stationary series with spectrum f_Y.

$$\hat{X}(\omega) = \hat{Y}(\omega) + A\sum\cos(2\pi\theta t+\phi)\exp(2\pi\omega t i)/\sqrt{T}$$

Use complex exponentials to do sum.

$$\begin{aligned}\sum\cos(2\pi\theta t + \phi)\exp(2\pi\omega t i) \\ &= \sum\exp(2\pi i((\omega+\theta)t+\phi)) \\ &\quad + \sum\exp(2\pi i((\omega-\theta)t-\phi))\end{aligned}$$

For α not an integer:

$$\sum \exp(2\pi\alpha t i) = \frac{1 - \exp(2\pi\alpha T i)}{1 - \exp(2\pi\alpha i)}$$

while for α an integer the sum is T.

So: at $\omega = \theta$ periodogram gets bigger as T grows:

$$|\hat{X}(\omega)|^2 \sim T^2/T = T$$

For other ω not too close to θ periodogram does not grow with T.

Properties of the Periodogram

The discrete Fourier transform

$$\hat{X}(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X_t \exp(2\pi\omega t i)$$

is periodic with period 1 because all the exponentials have period 1. Moreover,

$$\hat{X}(1-\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X_t \exp(-2\pi\omega t i)\exp(2\pi t i) = \overline{\hat{X}(\omega)}$$

so periodogram satisfies

$$|\hat{X}(1-\omega)|^2 = |\hat{X}(\omega)|^2.$$

So periodogram symmetric around $\omega = 1/2$.

Called *Nyquist* or *folding* frequency.

(Value is always 1/2 in cycles per point; usually converted to cycles per time unit like year or day.)

Similarly power spectral density f_X given by

$$f_X(\omega) = \sum_{-\infty}^{\infty} C_X(h) \exp(2\pi h \omega i)$$

is periodic with period 1 and satisfies

$$f_X(-\omega) = f_X(\omega)$$

which is equivalent to

$$f_X(1-\omega) = f_X(\omega) \, .$$