Fourier Analysis with R

Why time series in astronomy?

- Periodic phenomena: binary orbits; stellar rotation; pulsation (helioseismology, Cepheids)
- Stochastic phenomena: accretion (CVs, X-ray binaries, Seyfert gals, quasars); scintillation (interplanetary interstellar media); jet variations (blazars)
- Explosive phenomena: thermonuclear (novae, X-ray bursts), magnetic reconnection (solar/stellar flares), star death (supernovae, gamma-ray bursts)

Why time series in astronomy?

- Let us consider the dataset `COUP_var`. This dataset consists of three time series representing (in)homogeneous Poisson processes. They are tables of arrival times of individual X-ray photons of some young stars More details `https://rdrr.io/cran/astrodatR/man/COUP_var.html`

Why time series in astronomy?

```
library(astrodatR)
```

```
data("COUP_var")
plot(1:15145,COUP_var[,2])
```

- Randomness may appear in the data and evolve with time. It is such the case for financial or environmental time series
- Randomness aspect could be
  - separated from the evolution with time : we add an error measurement ε_t to a deterministic time series
  - be intrinsic (financial time series) and evolving with time

  or
- We need relevant stochastic models to describe how randomness evolves with time or not

- Autoregressive time series : (x_t) is AR(p) if

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t \text{ with } (\varepsilon_t) \text{ white noise}$$

- Moving Average time series (x_t) is MA(q) if

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \text{ with } (\varepsilon_t) \text{ white noise}$$

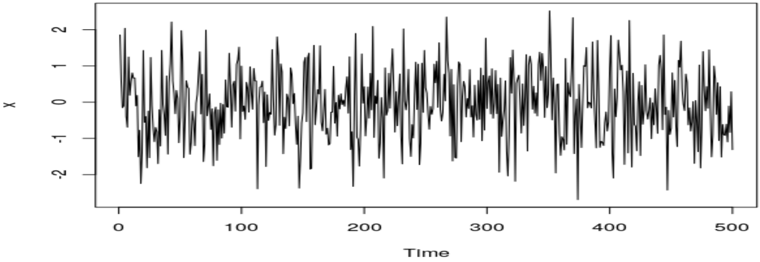
- ARMA models are the combination of the two previous ones : (x_t) is ARMA(p,q) if

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \text{ with } (\varepsilon_t) \text{ white noise}$$

If we integrate d times a ARMA time series, we obtain a ARIMA(p,d,q) time series if d is an integer and a FARIMA(p,d,q) if we apply fractional integration The difference between the different models lies in their dependence propertiesDenote γ the autocovariance function defined as $\gamma(h) = \mathbb{E}[(x_{t+h} - \mu)(x_t - \mu)]$ with $\mu = \mathbb{E}[x_t]$.ARMA are short range dependent, tht is $\sum_h \gamma(h) < \infty$ whereas FARIMA are long-range dependent, that is $\sum_h \gamma(h) = \infty$

Time series
Classical Gaussian time series models

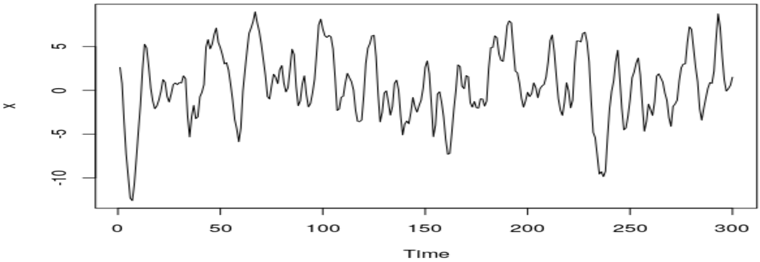
```
#simulate a sample of length 500 from a white noise
x <- rnorm(500)
plot.ts(x)
```



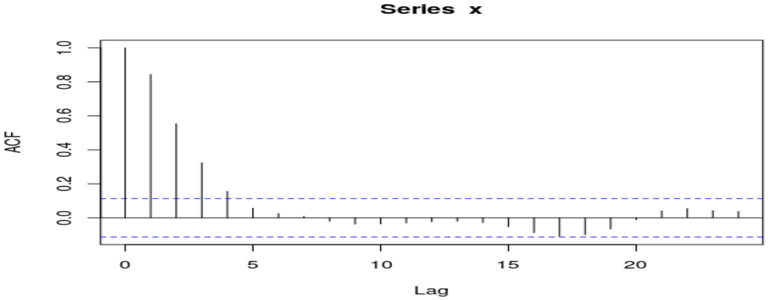
Time series

#Simulate ARMA

```
x=arima.sim(model=list(ar=c(1,-0.25),ma=1),300,rand.gen=rnorm)
plot.ts(x)
```

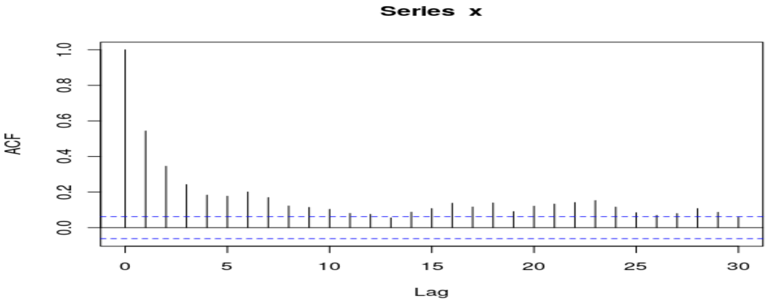


Time series
`acf(x)`



# Time series
library(arfima)

```
x <- arfima.sim(1000, model = list(phi = .2, dfrac =
.3))
acf(x)
```



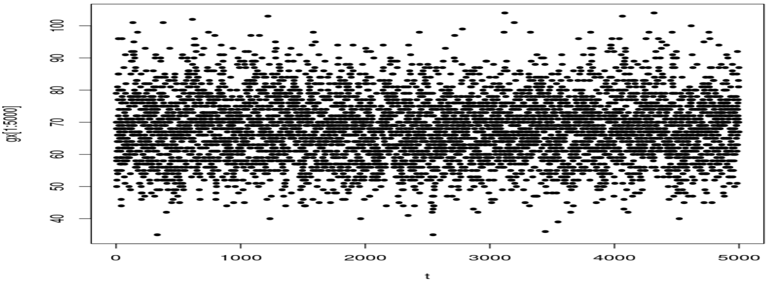
Why Fourier analysis in astronomy?

- Spectral analysis of a phenomena
- Identification of periodicities in the signal
- Restrictive assumptions of Fourier analysis : infinitely long dataset of equally-spaced observations; homoscedastic Gaussian noise; sinusoidal shape
- Challenges : deal with noise, unequal sampling....

- We consider the X-ray Binary GX 5-1 Time Series which can be downloaded from the website `https://www.iiap.res.in//astrostat/`
- More details about this dataset `http://astrostatistics.psu.edu/datasets`

Why Fourier analysis in astronomy?

```
gx=scan("/home/marianne/Desktop/gx.dat")
t=1:5000
plot(t,gx[1:5000],pch=20)
```



Challenges : identify quasi periodic oscillations and red noise using Fourier analysis

Fourier analysis for time series : the theory

Spectral density

- Let (x_t) a stationary second order time series, that is let us assume that $\mathbb{E}[x_t^2] < \infty$.
- There exists a function f_x called the spectral density of x such that

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2i\pi\omega h} f(\omega) d\omega$$

- Short and long range dependence can also be characterized in the Fourier domain, considering the behavior at $\omega = 0$ of the spectral density

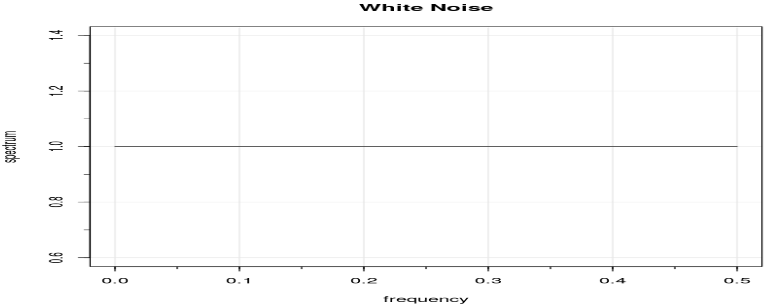
Basics Fourier analysis

Spectral density

- One has explicit expressions of the spectral density of classical time series model as White noise, ARMA,....
- For white noise with variance σ^2, $f(\omega) = \sigma^2$
- If (x_t) is ARMA(p,q), $\phi(B)x_t = \theta(B)w_t$, its spectral density is given by

$$f(\omega) = \sigma^2 \frac{|\theta(e^{-2i\pi\omega)}|^2}{|\phi(e^{-2i\pi\omega}|}$$

```
freq <- seq.int(0, 0.5, length.out = 500)
spec <- rep(1,500)
plot(freq, spec,'l')
```

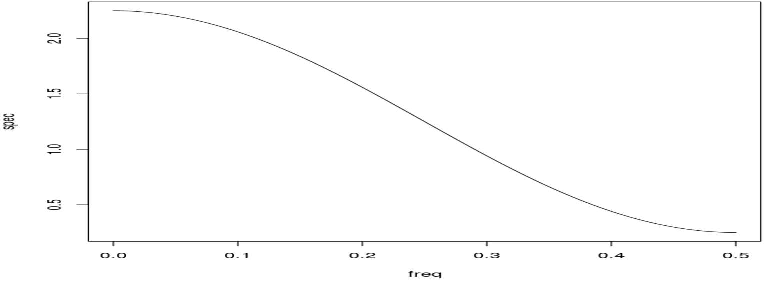


Spectral density

Basics Fourier analysis

Spectral density of of MA(1) with $\theta_0 = 1$, $\theta_1 = 0.5$. Spectral density $f(\omega) = \sigma^2 |1 + 0.5e^{-2i\pi\omega}|^2$

```
freq <- seq.int(0, 0.5, length.out = 500)
spec <- ((1 + 0.5*cos(2 * pi * freq ))∧2 + (0.5*sin(2
* pi * freq ))∧2)
plot(freq, spec)
```

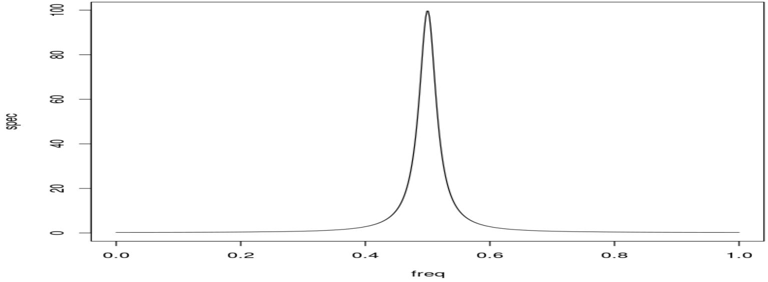


Spectral density

Basics Fourier analysis

Spectral density of of AR(1) with $\phi_0 = 1$ and $\phi_1 = -0.9$. Spectral density $f(\omega) = |1 + 0.9e^{-2i\pi\omega}|^{-2}$

```
freq <- seq.int(0, 0.5, length.out = 500)
spec <- ((1 + 0.9*cos(2 * pi * freq ))∧2 + (0.9*sin(2
* pi * freq ))∧2)
plot(freq, spec)
```



Spectral density

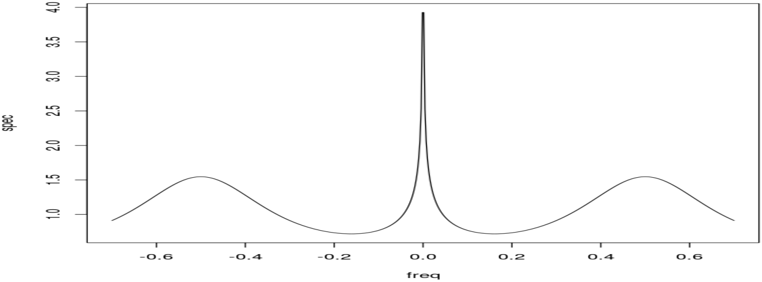
- Let (x_t) a stationary second order time series. When $\sum_{h=-\infty}^{\infty} \gamma(h) = \infty$, the time series is said to be long range dependent
- If at 0, $f(\omega) \sim |\omega|^{-2d} f^*(\omega)$ with $0 < d < 1/2$, f^* bounded, the time series is LRD
- The spectral densiy of a FARIMA(p,d,q) is

$$f(\omega) = |1 - e^{-2i\pi\omega}|^{-2d} f_{ARMA(p,q)}(\omega) = \sigma^2 |1 - e^{-2i\pi\omega}|^{-2d} \frac{|\theta(e^{-2i\pi\omega)}|^2}{|\phi(e^{-2i\pi\omega}|}$$

Basics Fourier analysis

Spectral density of of FARIMA(1,0.2,0) with $\phi_0 = 1$ and $\phi_1 = -0.3$. Spectral density $f(\omega) = |1 - e^{-2i\pi\omega}|^{-0.4}|1 + 0.3e^{-2i\pi\omega}|^{-2}$

```
freq <- seq.int(-0.7, 0.7, length.out = 500)
spec <- ((1-cos(2 * pi * freq))^2 + (sin(2*pi*freq))^2)^( -
0.2)/((1+0.3*cos(2*pi*freq))^2 + (0.3*sin(2*pi*freq))^2)
plot(freq, spec,'l')
```



Spectral density

Fourier analysis for time series : practical implementation

- Discrete Fourier Transform of a discrete time series $x_0 = x(0), \cdots, x_{n-1} = x((n-1)T)$ observed on a finite window $[0, T]$

$$X[\omega_j] = n^{-1/2} \sum_{t=0}^{n-1} x(t) e^{-2i\pi t \omega_j}$$

  where $\omega_j = j/n$ for $j = 0, \cdots, n-1$
- In practice : efficient FFT algorihm
- Classical periodogramm approximating spectral density

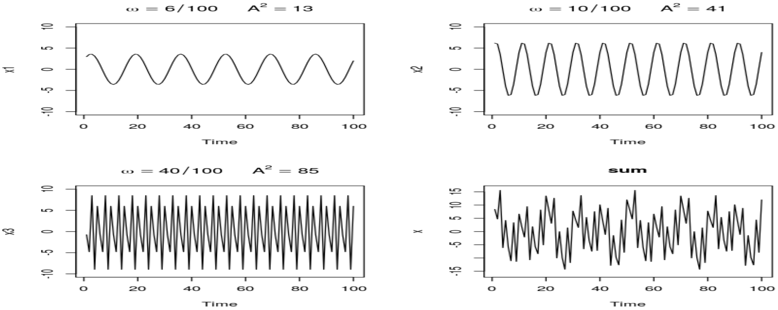
$$I(\omega_j) = |X[\omega_j]|^2$$

Fourier analysis for time series : practical implementation

```
t = 1:100

x1 = 2*cos(2*pi*t*6/100) + 3*sin(2*pi*t*6/100)
x2 = 4*cos(2*pi*t*10/100) + 5*sin(2*pi*t*10/100)
x3 = 6*cos(2*pi*t*40/100) + 7*sin(2*pi*t*40/100)
x = x1+x2+ x3
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-10,10), main =
expression(omega==6/100   A^2 == 13))
plot.ts(x2, ylim=c(-10,10), main =
expression(omega==10/100   A^2 == 41))
plot.ts(x3, ylim=c(-10,10), main =
expression(omega==40/100   A^2 == 85))
plot.ts(x, ylim=c(-16,16),main="sum")
```

Fourier analysis for time series : practical implementation

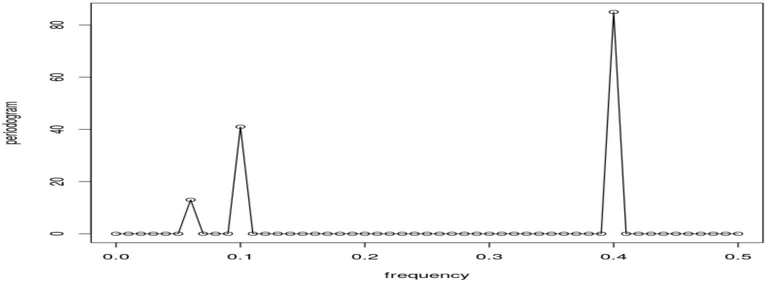


Fourier analysis for time series : practical implementation

```
P = abs(2*fft(x)/100)^2
```

```
f = 0:50/100
plot(f, P[1:51], type="o", xlab="frequency",
ylab="periodogram")
```

Fourier analysis for time series : practical implementation



- The first sample $X(0)$ of the transformed series is the average of the input series.
- The DFT of a real series results in a symmetric series about the Nyquist frequency $f_N = n/(2T)$. The negative frequency samples are also the inverse of the positive frequency samples.
- The Nyquist frequency is the highest frequency component that should exist in the input series for the DFT to yield uncorrupted results.
- More specifically if there are no frequencies above Nyquist the original signal can be exactly reconstructed from the samples.

- Spectral analysis can also be thought of as an analysis of variance
- Code R

```
c1=cos(2*pi*t*6/100)
s1=sin(2*pi*t*6/100)
c2=cos(2*pi*t*10/100)
s2=sin(2*pi*t*10/100)
c3=cos(2*pi*t*40/100)
s3=sin(2*pi*t*40/100)
summary(lm(x~I(c1)+I(c2)+I(c3)+I(s1)+I(s2)+I(s3)))
anova(lm(x~I(c1)+I(c2)+I(c3)+I(s1)+I(s2)+I(s3)))
```

Fourier analysis for time series : beyond classical implementation

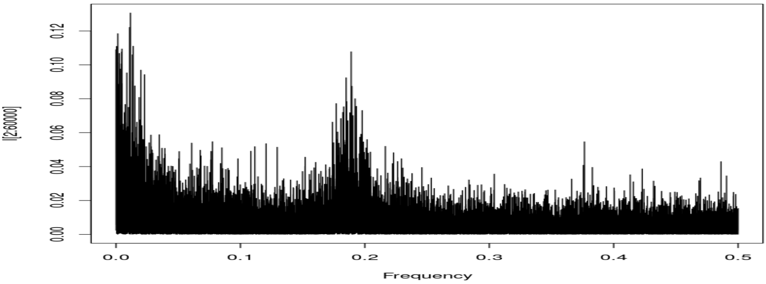
- The classical periodogram is not a good estimator!
- It is formally ’inconsistent’ because the number of parameters grows with the number of datapoints.
- In addition the DFT depends on several strong assumptions which are rarely achieved in real astronomical data :
    - evenly spaced data of infinite duration with a high sampling rate (Nyquist frequency)
    - Gaussian noise
    - single frequency periodicity with sinusoidal shape
    - centered time series and stationary behavior

How deal with real-life data and improve spectral density estimation?

Fourier analysis for time series : beyond classical implementation

Back to GX astronomical data

```
freq = 0:32768/65536
I = (4/65536)*abs(fft(gx)/sqrt(65536))∧2
plot(freq[2:60000],I[2:60000],type='l',xlab='Frequency')
```



- We introduce a frequency band, B of length $L = 2m + 1$ with Ln contiguous fundamental frequencies, centered around a frequency of interest $\omega_j = j/n$
- The basic idea is that $f(\omega_j) \approx f(\omega_j + k/n)$ for $k \in B$
- One then defines the average periodoram as

$$\bar{f}(\omega_j) = \frac{1}{L} \sum_{\ell=-m}^{m} I(\omega_j + \ell/n)$$

- Under the assumption that the spectral density is fairly constant in the band it can be proved that for large n

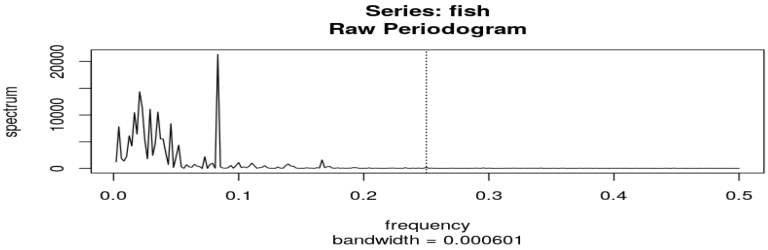
$$\frac{L\bar{f}(\omega)}{f(\omega)} \sim \chi^2_{2L}$$

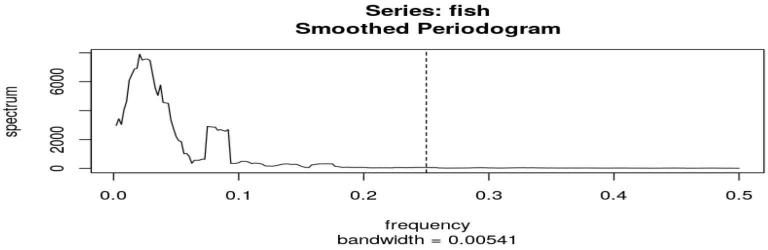
- The width of the frequency interval L/n is called the bandwidth
- To compute averaged periodograms one can also use convolution with kernels

- We consider an environmental series giving the number of new fish in Pacific Ocean
- We want to analyze the spectral content of this time series
- It can be downloaded on the website `http://anson.ucdavis.edu/~shumway/recruit.dat`

```
fish=scan("/home/marianne/Desktop/recruit.dat")
par(mfrow=c(2,1))
fish.per = spec.pgram(fish, taper=0, log="no")
abline(v=1/4, lty="dotted")
k = kernel("daniell", 4)
fish.ave = spec.pgram(fish, k, taper=0, log="no")
abline(v=c(.25,1,2,3), lty=2)
fish.ave$bandwidth
fish.ave$bandwidth*(1/12)*sqrt(12)
```

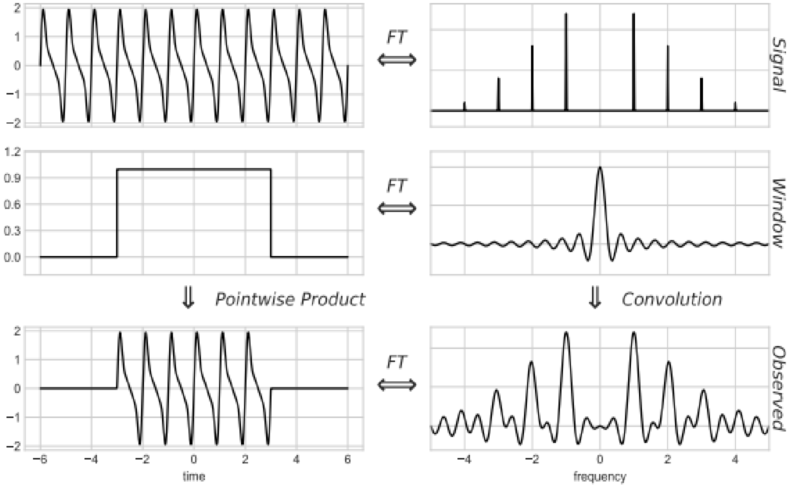
Fourier analysis for real-life time series





Fourier analysis for real-life time series

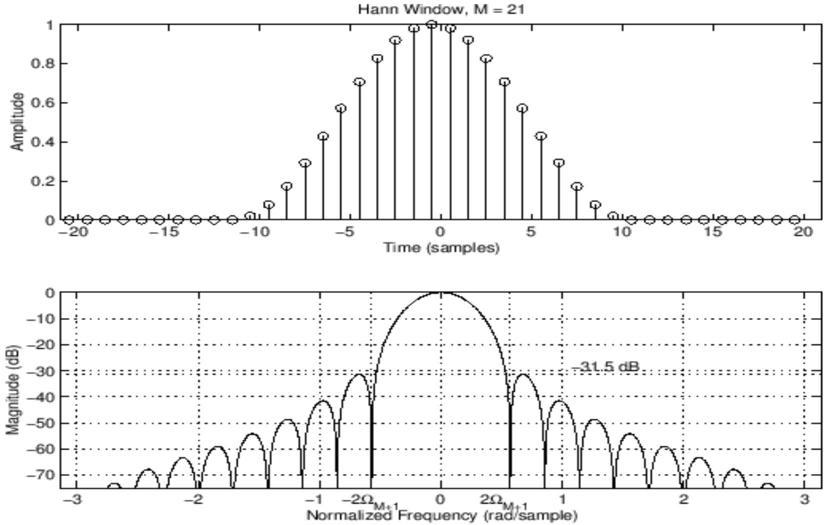
The Fourier Transform of a product of two continuous signals $h \cdot x$ is the convolution of the Fourier transform of the two signals



Fourier analysis for real-life time series

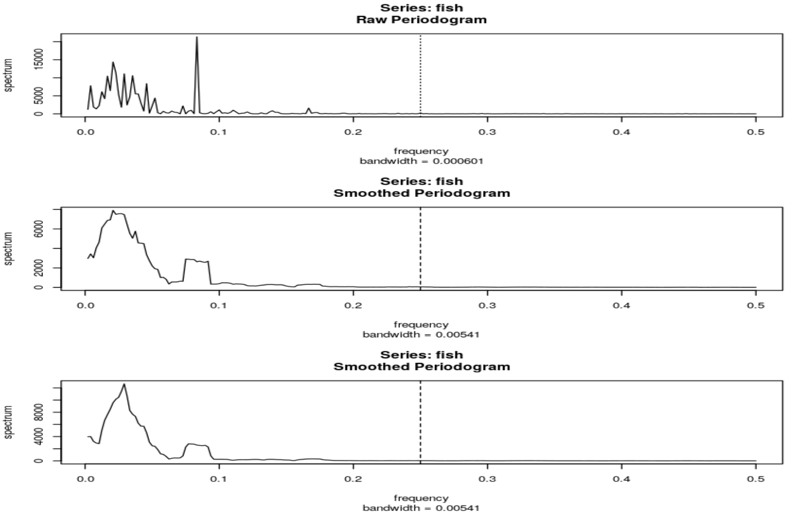
- The aim of tapering is to take into account this windowing effect
- We shall replace (x_t) with $y_t h_t \cdot x_t$, where h_t is a given window called the taper whose goal is to minimize the distorsion of the signal in the frequency domain
- Tapers generally have a shape that enhances the center of the data relative to the extremities

Fourier analysis for real-life time series



Fourier analysis for real-life time series

```
fish.taper = spec.pgram(fish, k, taper=0.5, log="no")
```



- In the case of a second order multivariate time series $(x_t^{(1)}, \cdots, x_t^{(m)})$ the spectral density is a matrix f of size $m \times m$
- It is related to the cross covariance $\gamma_{\ell_1,\ell_2}(h) = \mathbb{E}[(x_{t+h}^{(\ell_1)} - \mu^{(\ell_1)})((x_t^{(\ell_2)} - \mu^{(\ell_2)})]$ as follows

$$\gamma_{\ell_1,\ell_2}(h) = \int_{-1/2}^{1/2} f_{\ell_1,\ell_2}(\omega) e^{-2i\pi h\omega} d\omega$$

- If $\sum_h |\gamma_{\ell_1,\ell_2}(h)| < \infty$, one has

$$f_{\ell_1,\ell_2}(\omega) = \sum \gamma_{\ell_1,\ell_2}(h) e^{-2i\pi\omega h}$$

- It may be a complex number!

- Each component $f_{\ell,\ell}$ of this matrix is the spectral density of the univariate time series $(x_t^{(\ell)})$
- If we are given $\ell_1 \neq \ell_2$, the component f_{ℓ_1,ℓ_2} are the cross spectral density and is related to cross covariance

- The coherence between two time series $x_t^{(1)}$ and $x_t^{(1)}$ is defined as:

$$C(\omega) = \frac{|f_{1,2}(\omega)|^2}{f_{11}(\omega)f_{22}(\omega)}$$

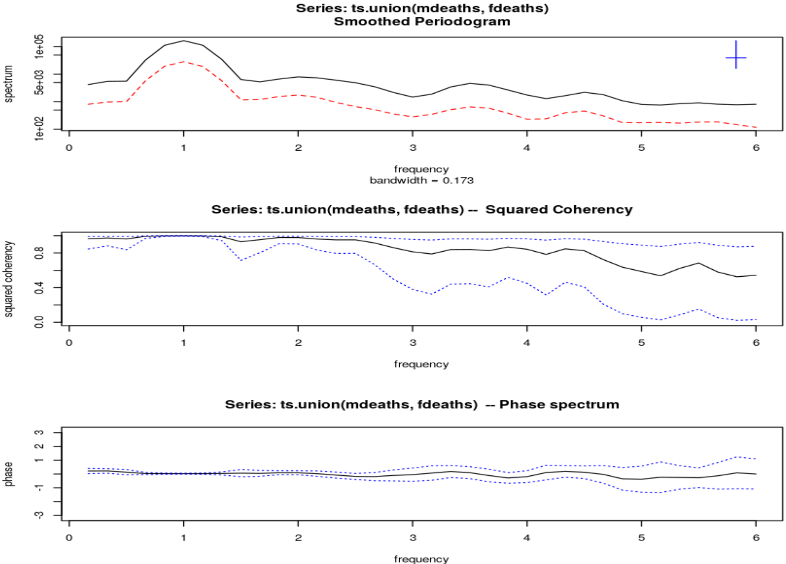
- Note that f_{ℓ_1,ℓ_2} may take complex values. The phase of f_{ℓ_1,ℓ_2} is called the phase spectrum

- We use the dataset `deaths` giving monthly deaths in the UK from a set of common lung diseases for the years 1974 to 1979
- It is the sum of two series `mdeaths` and `fdeaths` for males and females that we can analse simultaneously

```
par(mfrow=c(3,1))
mfdeaths.spc <- spec.pgram(ts.union(mdeaths,
fdeaths), spans = c(3,3))
plot(mfdeaths.spc, plot.type = "coherency")
plot(mfdeaths.spc, plot.type = "phase")
```

Fourier analysis for real-life time series

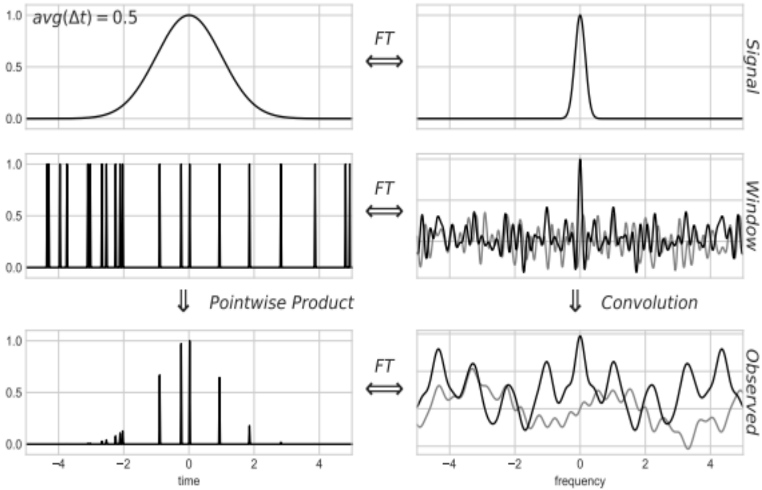
Multivariate time series



- In astronomy, it is common to have incomplete or unevenly sampled time series for a given variable.
- Determining cycles in such series is not directly possible with methods such as Fast Fourier Transform (FFT) and may require some degree of interpolation to fill in gaps
- An alternative method is to se the Lomb Scargle periodogram

Fourier analysis for real-life time series

Effect of irregular sampling on a signal



Fourier analysis for real-life time series

Principle of Lomb Scargle peridogram

- Let $\widetilde{x}_\ell := x(t_\ell)$ irregularly sampled data at times (t_n)
- Naive approach to compute the periodogram

$$
\begin{aligned}
I(\omega) &= \frac{1}{n}\left|\sum \widetilde{x}_\ell e^{-2i\pi t_\ell \omega}\right|^2 \\
&= \frac{1}{n}\left(\sum_\ell \widetilde{x}_\ell \cos(2\pi t_\ell \omega)\right)^2 + \frac{1}{n}\left((\sum_\ell \widetilde{x}_\ell \sin(2\pi t_\ell \omega)\right)^2
\end{aligned}
$$

- The approach consist in considering a periodogram of the form

$$
\begin{aligned}
I_{LS}(\omega) &= \frac{A^2(\omega)}{2}\left(\sum_\ell \widetilde{x}_\ell \cos(2\pi(t_\ell - \tau(\omega))\omega)\right)^2 \\
&\quad + \frac{B^2(\omega)}{2}\left((\sum_\ell \widetilde{x}_\ell \sin(2\pi(t_\ell - \tau(\omega))\omega)\right)^2
\end{aligned}
$$

A, B, τ are functions depending on the frequency such that

- The periodogram reduces to the classical form in the case of equally-spaced observations,
- The periodogram's statistics are analytically computable,
- The periodogram is insensitive to global time-shifts in the data.

Fourier analysis for real-life time series

Principle of Lomb Scargle peridogram

- It has been prove by Lomb and Scargle tha the optimal choice is

$$\tau(\omega) = \frac{1}{4\pi\omega} arctan\left(\frac{\sum \sin(4\pi\omega t_\ell}{\sum \cos(4\pi\omega t_\ell}\right)$$

$$A(\omega) = \sum_\ell \widetilde{x}_\ell \cos^2(2\pi(t_\ell - \tau(\omega))\omega)$$

$$B(\omega) = \sum_\ell \widetilde{x}_\ell \sin^2(2\pi(t_\ell - \tau(\omega))\omega)$$

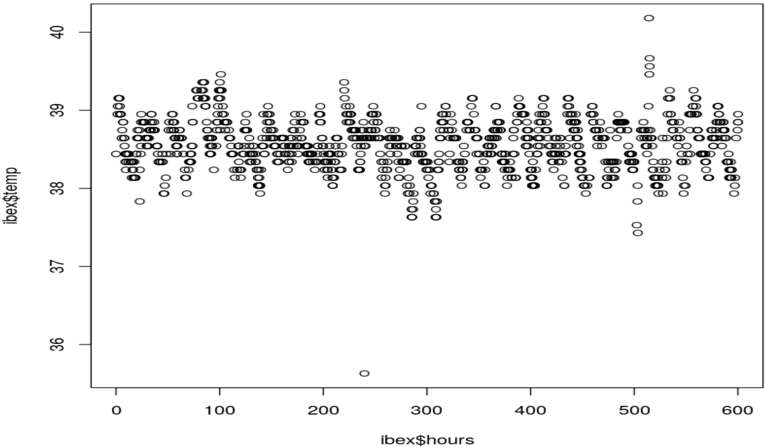
- These formula are designed such that any individual frequency ω, this method gives the same power as does a least-squares fit to sinusoids of that frequency

- One can add an offset depending on the frequency to the model to take into account the fact that the data are not centered
- One can also take into account sparsity of the data penalizing the least square method

- We consider the dataset ibex.
- This dataset is an object of class ”ltraj” (regular trajectory, relocations every 4 hours) containing the GPS relocations of four ibex during 15 days in the Belledonne mountain (French Alps).

# Fourier analysis for real-life time series
data(ibex)

```
plot(ibex$hours,ibex$temp)
```



Fourier analysis for real-life time series

```
library(lomb
lsp(ibex[2:3],)
```