Time Series Analysis:

2. The power spectrum

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2009

Autocorrelation function

Suppose we have a random series $\{x_n\}_{n=0}^{N-1}$.

Question: Having measured a certain value x_j, what information do we have on x_{j+k}?

Answer: For a *stationary* series, calculate the autocorrelation function

$$C_k = \left\langle x_j \, x_{j+k} \right\rangle \tag{1a}$$

or the correlation coefficient

$$\rho_k = \frac{\left\langle x_j \, x_{j+k} \right\rangle}{\left\langle x_j^2 \right\rangle}. \tag{1b}$$

The braces $\langle \cdots \rangle$ stand for averaging over realizations of the random process.

Beware!

Usually, only a *single realization* of a time series is available!

Instead of (1a), calculate

$$C_k = \frac{1}{N-k} \sum_{j=0}^{N-k-1} x_j \, x_{j+k} \tag{2}$$

(similarly for ρ_k). (2) gives only an *estimate* of the “true” autocorrelation function.

Stationarity

A series is stationary if it does not change *qualitatively* over the time.

Definition: Divide a series into an arbitrary number of section (of arbitrary lengths). If statistical distributions of values in every section are identical, the series is stationary.

Another definition: A series is stationary if its autocorrelation function (1a) depends only on offsets k (in principle, it could depend on both k and j).

A majority of series are *not* stationary: for example, periodic series, series with trends or seasonal changes, series where noise parameters vary over the time, or random series superimposed on deterministic (not constant) signals are not stationary.

Remarks

- $C_k = C_{-k}$.
- $\rho_k = \pm 1$ — a deterministic relation.
- $|\rho_k| < 1$ — only *statistical* information.
  - Example: $\rho_k = 0.875$: $x_{j+k} = 0.875 x_j + 0.125 y_k$, $\left\langle x_j \, y_k \right\rangle = 0$.

The Wiener-Khinchin Theorem

Theorem: The Fourier transform of the autocorrelation of a *stationary* series is its power spectrum.

Why not non-stationary? Examples: music, speech, series with trends etc.

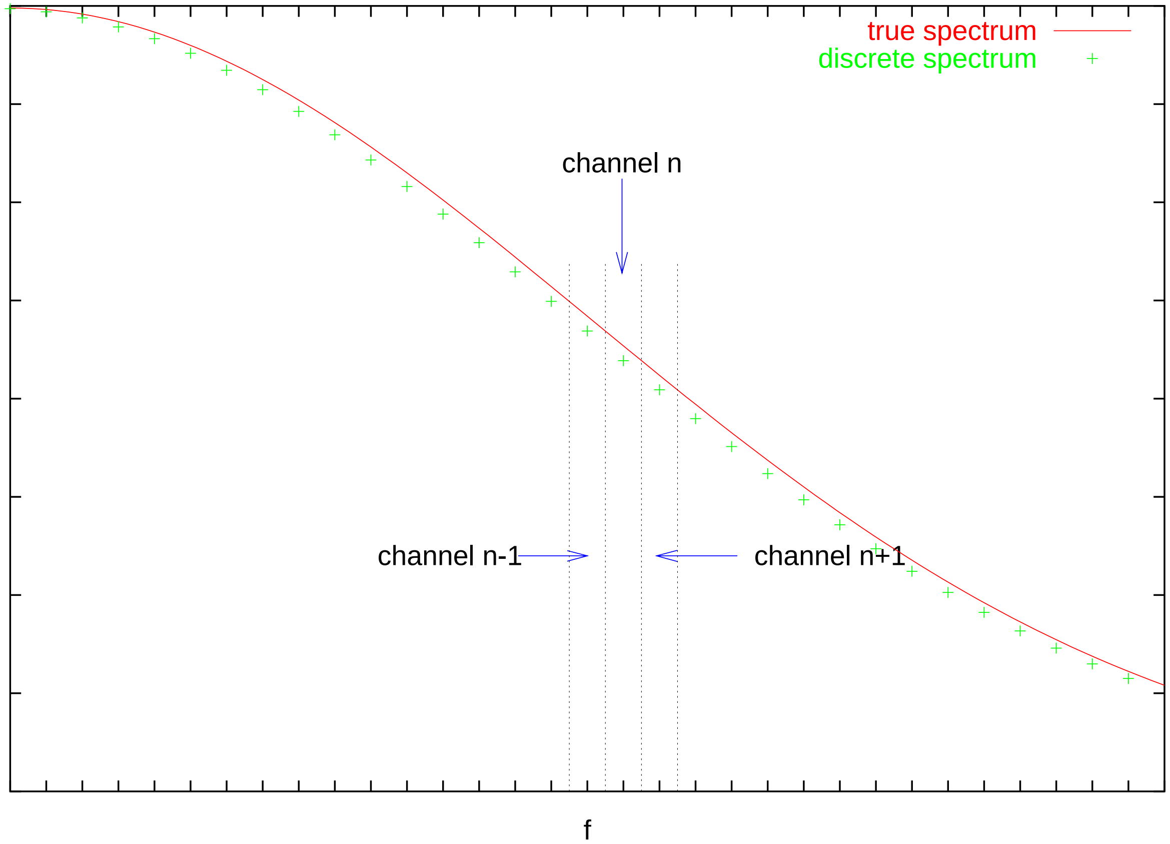
Given a stationary time series $\{x_n\}_{n=0}^{N-1}$, calculate the autocorrelation (2) and Fourier transform it to get the power spectrum. **Bad idea.**

The power spectrum

Continuous signals: $P(f)$ — the density of power in the interval $(f + df)$.

Discrete signals: $P(f_n)$ — an *estimate* of the power in the interval $(f_n - 1/(2N\Delta), f_n + 1/(2N\Delta))$.

The discrete power spectrum is only *an approximation* to the “true” power spectrum



Periodogram

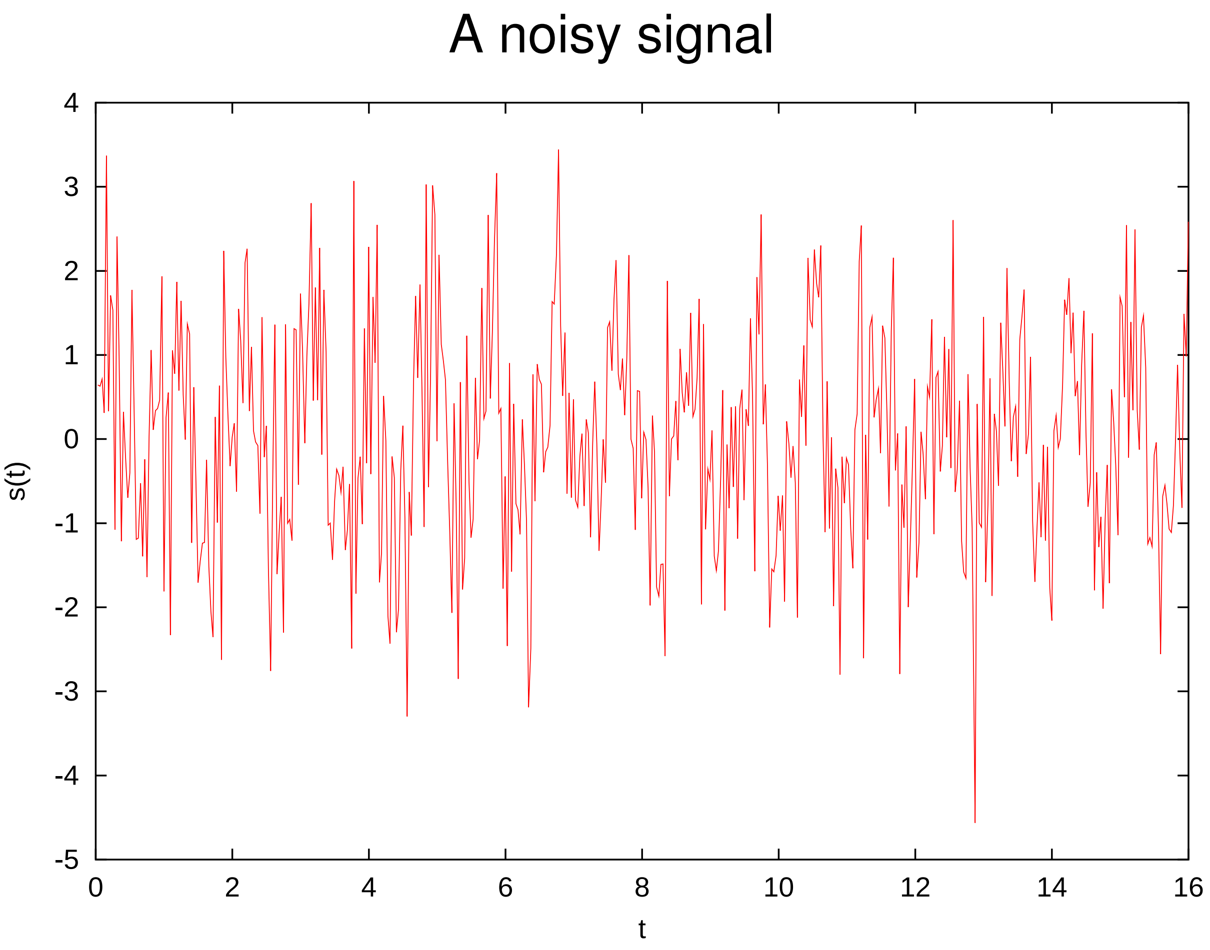
The most popular estimate of the power spectrum is called the periodogram. Because the power spectrum does not contain any information on the phase, we no longer distinguish the positive and negative frequencies for the purpose of calculating the periodogram.

$$P(0) = |G(0)|^2\,, \tag{3a}$$

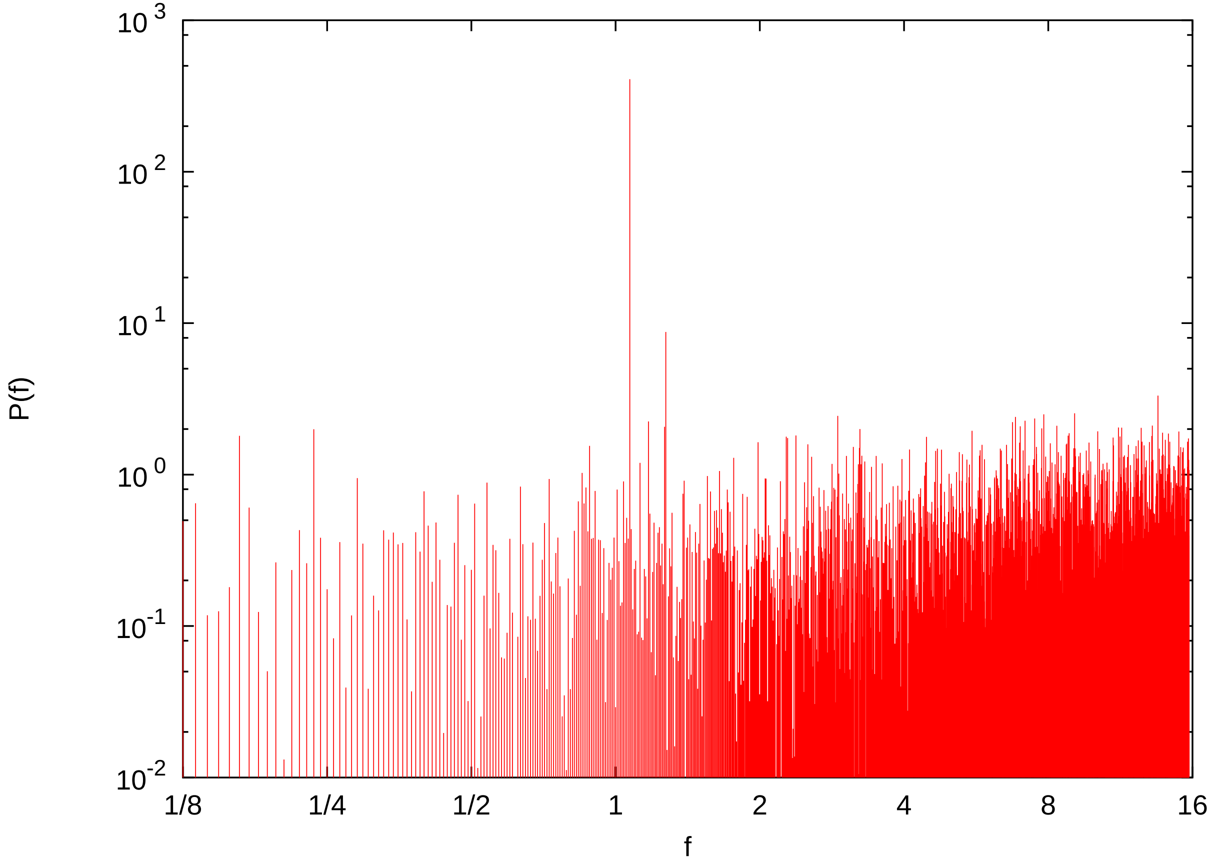
$$P(f_n) = \left[|G(f_n)|^2 + |G(f_{-n})|^2\right], \; n = 1, 2, \frac{N}{2} - 1 \tag{3b}$$

$$P(f_{N/2}) = |G(f_{N/2})|^2\,. \tag{3c}$$

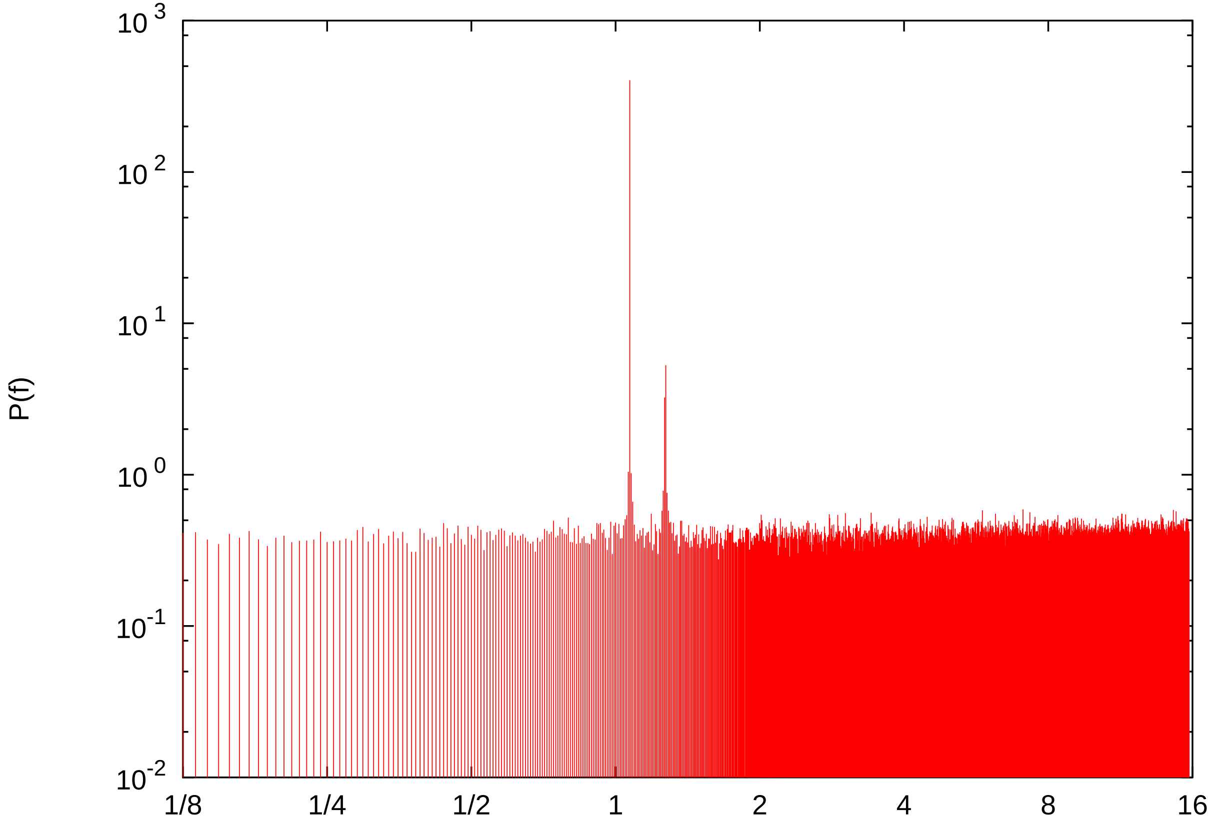
Caveat emptor! When calculating the periodogram, it is particularly important to take the proper care on what components are stored where, what the normalisation is etc. The point of the formula (3) is to bin together terms with the same "absolute" frequencies, but different pieces of software may store them differently. And they do.


A noisy signal
s(t)
t
4
3
2
1
0
-1
-2
-3
-4
-5
0
2
4
6
8
10
12
14
16

The power spectrum of the above signal



The power spectrum averaged over 64 realizations of the noise



When averaging over the realizations, *first* calculate the periodogram for each realization, and *then* take the average for each frequency.

A side note: Signal-to-Noise Ratio

Signal-to-Noise ratio measures the “goodness” of the signal, or how easily it can be distinguished from the noise. The general framework: calculate the ratio of the peak to the (local) background in the periodogram. Usually

$$\mathrm{SNR} = 10\,\log_{10}\left(\frac{\text{peak power}}{\text{background power}}\right)\ [\mathrm{dB}] \tag{4}$$

In the last example, SNR $\simeq 30$ dB.