

Time Series Analysis:

2. Convolution

The Power Spectrum

P. F. Góra

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

2018

Convolution

$s(t)$, $r(t)$ — continuous signals such that $\forall t < 0: s(t) = r(t) = 0$. Their convolution:

$$z(t) = (s \star r)(t) = \int_0^t s(t')r(t - t') dt' \quad (1)$$

The Fourier transforms fulfil $Z(f) = S(f)R(f)$.

The discrete convolution

Likewise, we can define the discrete convolution. Let $\{s_n\}_{n=0}^{N_1-1}$, $\{r_n\}_{n=0}^{N_2-1}$ be *periodic* discrete signals. Their discrete convolution is:

$$z_j = (s \star r)_j = \sum_{k=-M/2+1}^{M/2} s_{j-k} r_k. \quad (2)$$

where $M = \max(N_1, N_2)$ and the shorter signal is padded with zeros. (In many situations, one of the “signals” is much shorter than the other.) Fourier transforms fulfil $Z_k = S_k R_k$.

Example

If we have a set of linear equations with constant coefficients with an external forcing

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \mathbf{s}(t) \quad (3)$$

the solution contains a convolution of the internal dynamics of the system and the forcing:

$$\mathbf{x}(t) = \exp(\mathbf{A}t) \mathbf{x}(0) + \int_0^t \exp(\mathbf{A}(t-t')) \mathbf{s}(t') dt' \quad (4)$$

(can be generalized to linear equations with time-varying coefficients).

More examples

If an incoming wave passes through a finite-size slit whose spatial profile is described by a function, the outgoing wave is a convolution of the incoming wave and the spatial profile of the slit.

If a beam of light passes through a lens whose optical density is described by a function, the outgoing beam is a convolution of the incoming beam and the optical density profile.

In **signal processing**, if a signal passes through a linear device, the output signal is a convolution of the input signal and the *response function* of the device.

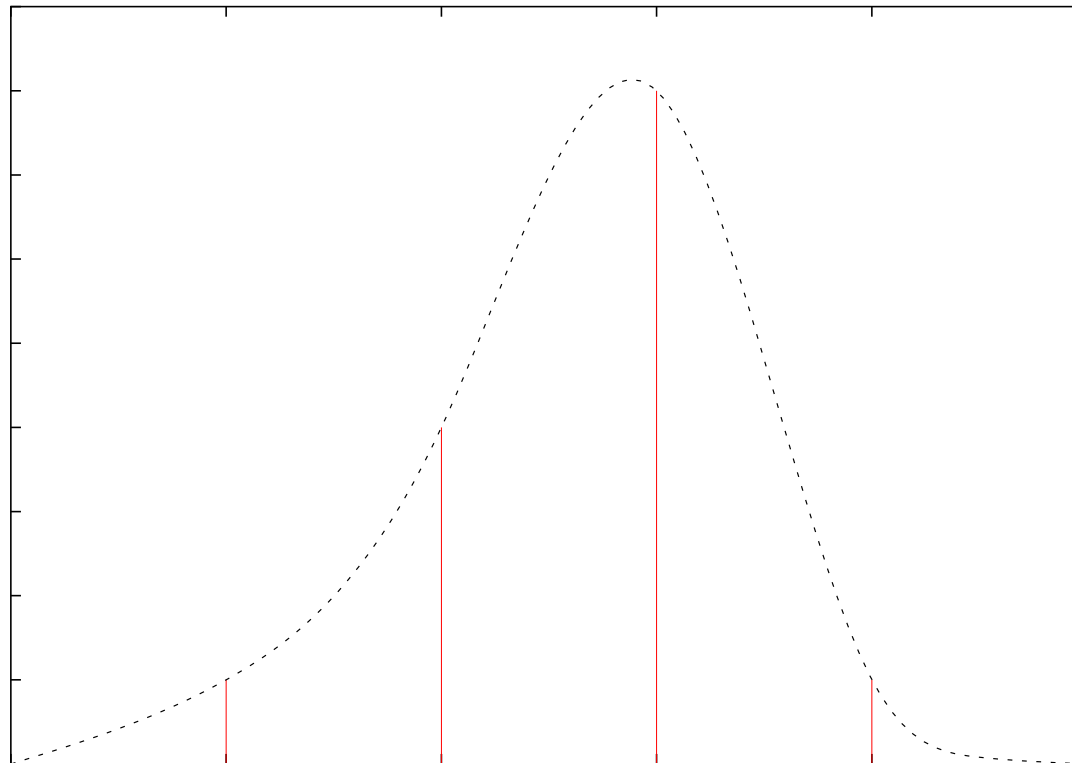
We call a linear device *a filter*.

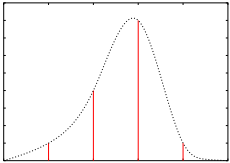
The response function?

For physical reasons, there are no devices that can respond *instantly* to an incoming signal: any physical device requires a finite time to “warm up” and start to respond to a signal. An instantaneous response would require an infinite power. Likewise, any physical device requires a finite time to stop creating output after the input has ceased.

The profile along which a filter starts and then stops responding to a signal is called *the response function* of the filter.

The response function of a filter

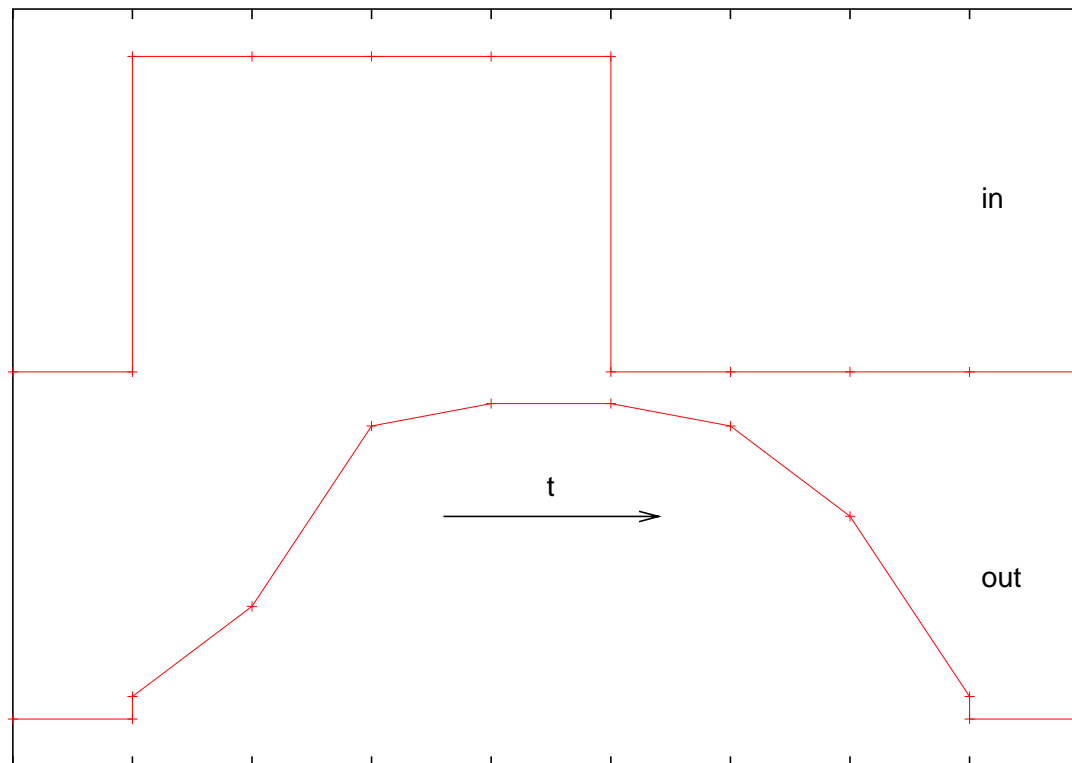




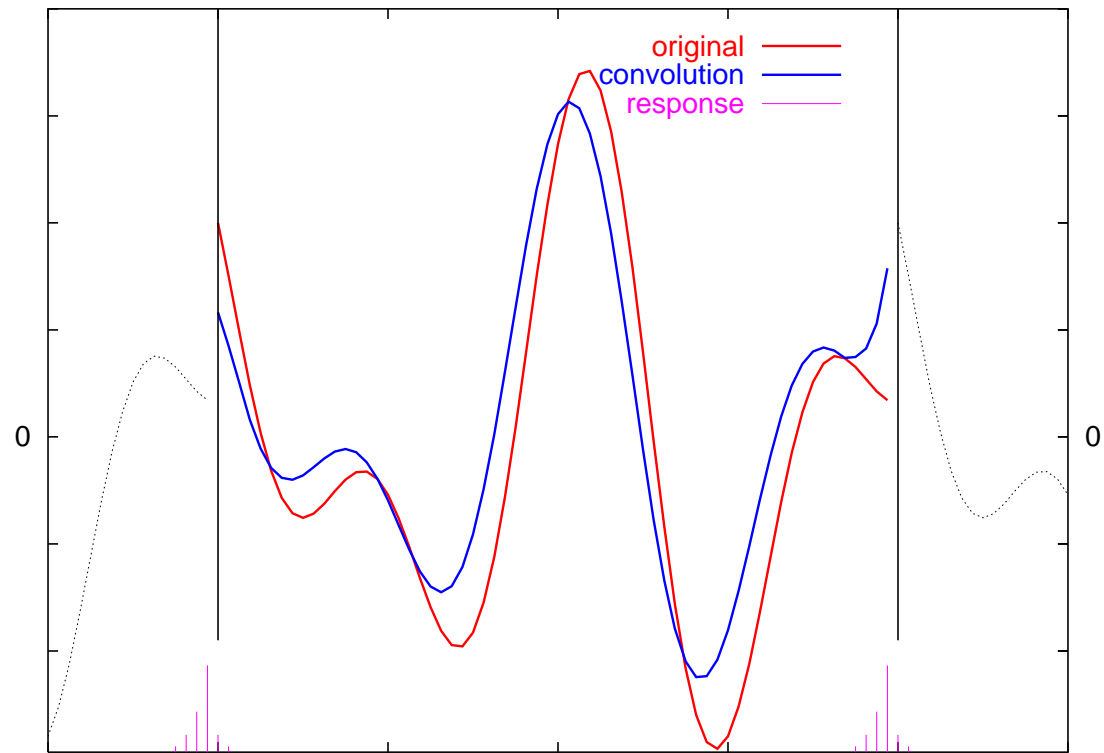
| | Input | | Output |
|--------|-------|-------|--------|
| 011111 | 0000 | 00000 | 0 |
| 001111 | 1000 | 00000 | 0.125 |
| 000111 | 1100 | 00000 | 0.625 |
| 000011 | 1110 | 00000 | 1.625 |
| 000001 | 1111 | 00000 | 1.75 |
| 000000 | 1111 | 10000 | 1.75 |
| 000000 | 0111 | 11000 | 1.625 |
| 000000 | 0011 | 11100 | 1.125 |
| 000000 | 0001 | 11110 | 0.125 |
| 000000 | 0000 | 11111 | 0 |

$$0.125A + 0.5B + C + 0.125D$$

A rectangular signal convoluted with the response function

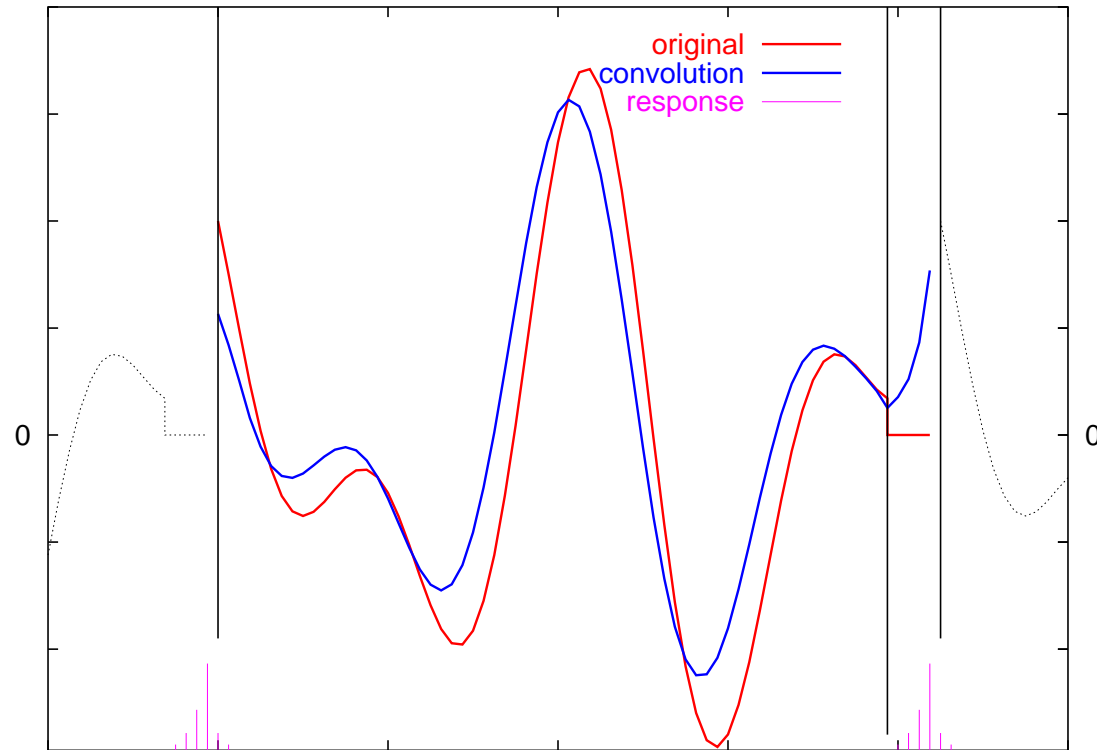


A convolution of a signal and a response function



**An unwelcome consequence of the assumed periodicity:
the ends are spoiled!**

A convolution of a signal and a response function



The ends are repaired by padding with zeros

The procedure

1. Pad the input signal with enough zeros to cover the longer tail of the response function (remember that the padded signal should have a “magic” length).
2. Calculate the FFT of both the signal and the response in one go.
3. Multiply the transforms termwise, in a linear time $O(N)$.
4. Calculate the inverse FFT.

The total cost $\sim O(2N \log N)$

An unexpected (?) application of the convolution

Given two polynomials $A(z)$, $B(z)$ of degrees not larger than n

$$A(z) = a_n z^n + a_{n-1} z^{n-1} + \dots + a_1 z + a_0, \quad (5a)$$

$$B(z) = b_n z^n + b_{n-1} z^{n-1} + \dots + b_1 z + b_0. \quad (5b)$$

Find the **coefficients** of their product $C(z) = A(z) B(z)$.

$$c_{2n} = a_n b_n \quad (6a)$$

$$c_{2n-1} = a_n b_{n-1} + a_{n-1} b_n \quad (6b)$$

$$c_{2n-2} = a_n b_{n-2} + a_{n-1} b_{n-1} + a_{n-2} b_n \quad (6c)$$

...

$$c_0 = a_0 b_0 \quad (6d)$$

The sum of indices in each term on rhs equals to the index on lhs.

We have

$$c_l = \sum_{k=0}^{2n} \tilde{a}_k \tilde{b}_{l-k}, \quad l = 0, 1, \dots, 2n, \quad (7a)$$

where

$$\tilde{a}_s, \tilde{b}_s = \begin{cases} a_s, b_s & s = 0, 1, \dots, n \\ 0 & s = n + 1, n + 2, \dots, 2n \end{cases} \quad (7b)$$

and periodicity (*sic!*) of the “signals” is used to calculate the coefficients with negative indices: $\tilde{b}_{-j} = \tilde{b}_{2n-j+1} = 0$ for $j = 1, 2, \dots, n$. The coefficients of the product are given by the convolution of the coefficients of the input polynomials. The numerical cost of calculating the coefficients of the product is $O(4n \log 2n)$, *not* $O(n^2)$, as it would appear.

Calculating products of large integers

$$x = \sum_{j=0}^n a_j 2^j, \quad (8a)$$

$$y = \sum_{j=0}^n b_j 2^j, \quad (8b)$$

where $a_j, b_j = \{0, 1\}$. The rhs are polynomials of the form (5), evaluated in $z = 2$. The product xy is also a polynomial, and its coefficients are given by the appropriate convolution. For a large n , calculating the product via FFT becomes efficient! In addition, calculating the FFT of a “signal” consisting of $\{0, 1\}$ is particularly simple.

The 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

$$C_k = \langle x_j x_{j+k} \rangle \quad (9a)$$

or the correlation coefficient

$$\rho_k = \frac{\langle x_j x_{j+k} \rangle}{\langle x_j^2 \rangle}. \quad (9b)$$

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

Beware!

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

Instead of (9a), calculate

$$C_k = \frac{1}{N - k} \sum_{j=0}^{N-k-1} x_j x_{j+k} \quad (10)$$

(similarly for ρ_k). (10) 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 (9a) 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: $\rho_k = 0.875$: $x_{j+k} = 0.875x_j + 0.125y_k$, $\langle x_j y_k \rangle = 0$.

The Wiener-Khinchin Theorem

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

$$P(f) = |G(f)|^2 . \quad (11)$$

Why not non-stationary? Examples: music, speech, series with trends etc. We shall return to the problem of non-stationary series later during this course.

Comment

If we have a finite time series then, under our assumption of periodicity (even if it is blatantly false), nothing prevents us from calculating its Fourier Transform and then from taking its square modulus. It wouldn't give us much information, though. The point of Wiener-Khinchin Theorem is that by calculating the power spectrum, we get information on the autocorrelation of the series.

For example, when we see a peak at a certain frequency f , we expect some form of periodicity, with the period $1/f$, in the autocorrelation.

Therefore, we do not use the autocorrelation to calculate the power spectrum, but we use the power spectrum to get insight on the autocorrelation.

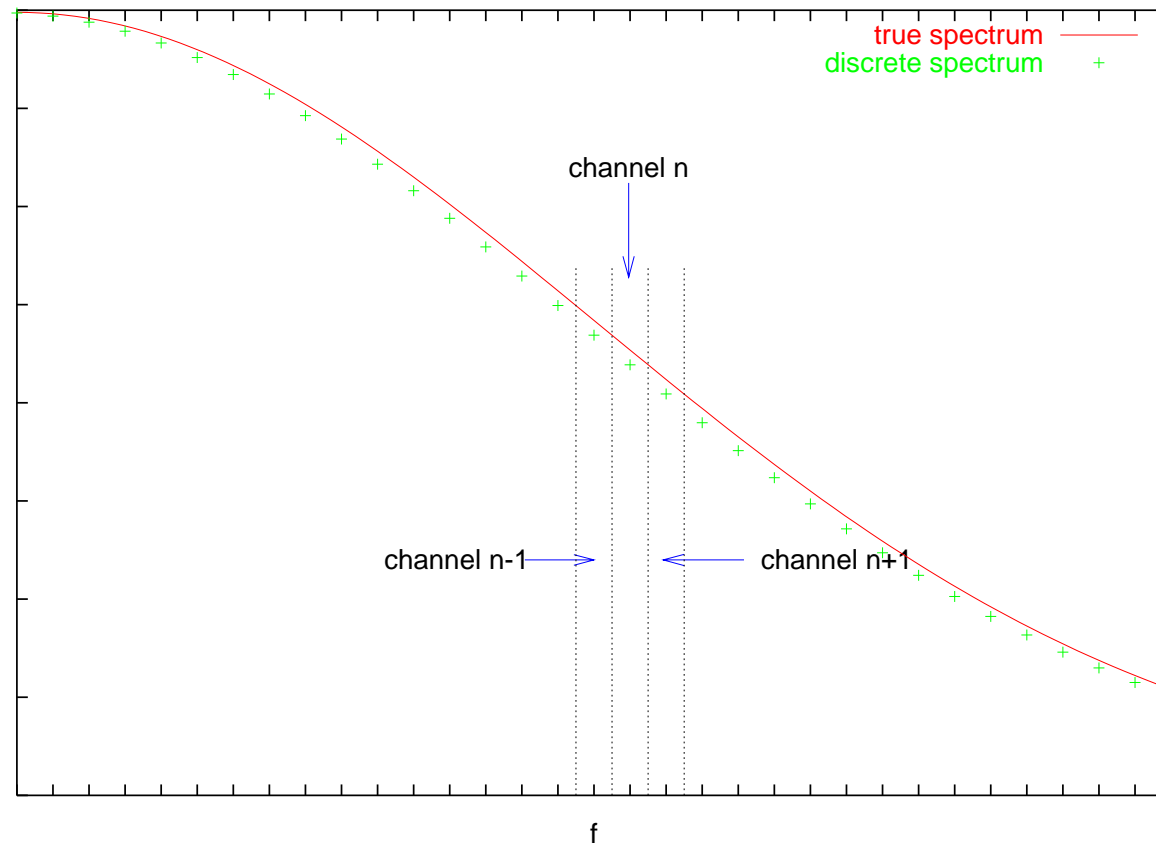
The power spectrum

Given a stationary time series $\{x_n\}_{n=0}^{N-1}$, calculate the autocorrelation (10) and Fourier transform it to get the power spectrum. **Bad idea.**

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.

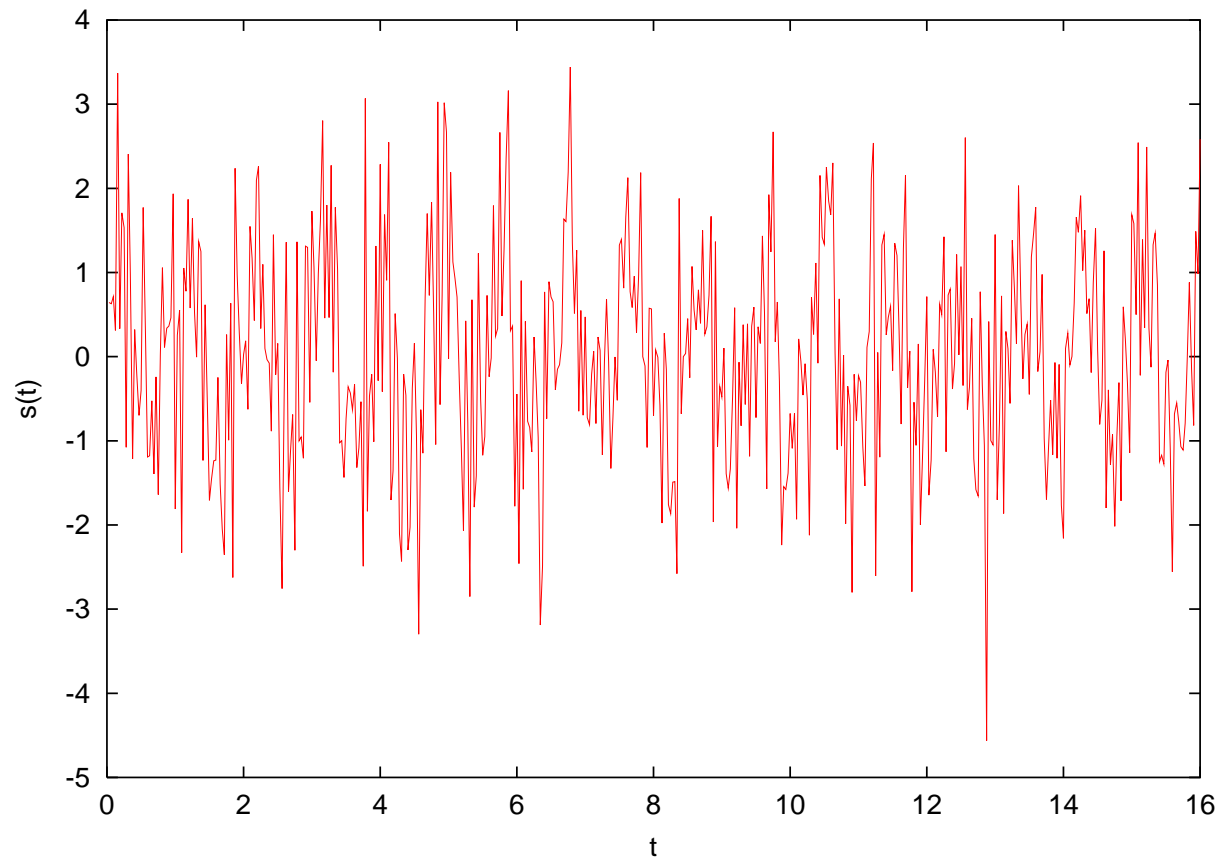
$$P(0) = |G(0)|^2, \quad (12a)$$

$$P(f_n) = \left[|G(f_n)|^2 + |G(f_{-n})|^2 \right], \quad n = 1, 2, \frac{N}{2} - 1 \quad (12b)$$

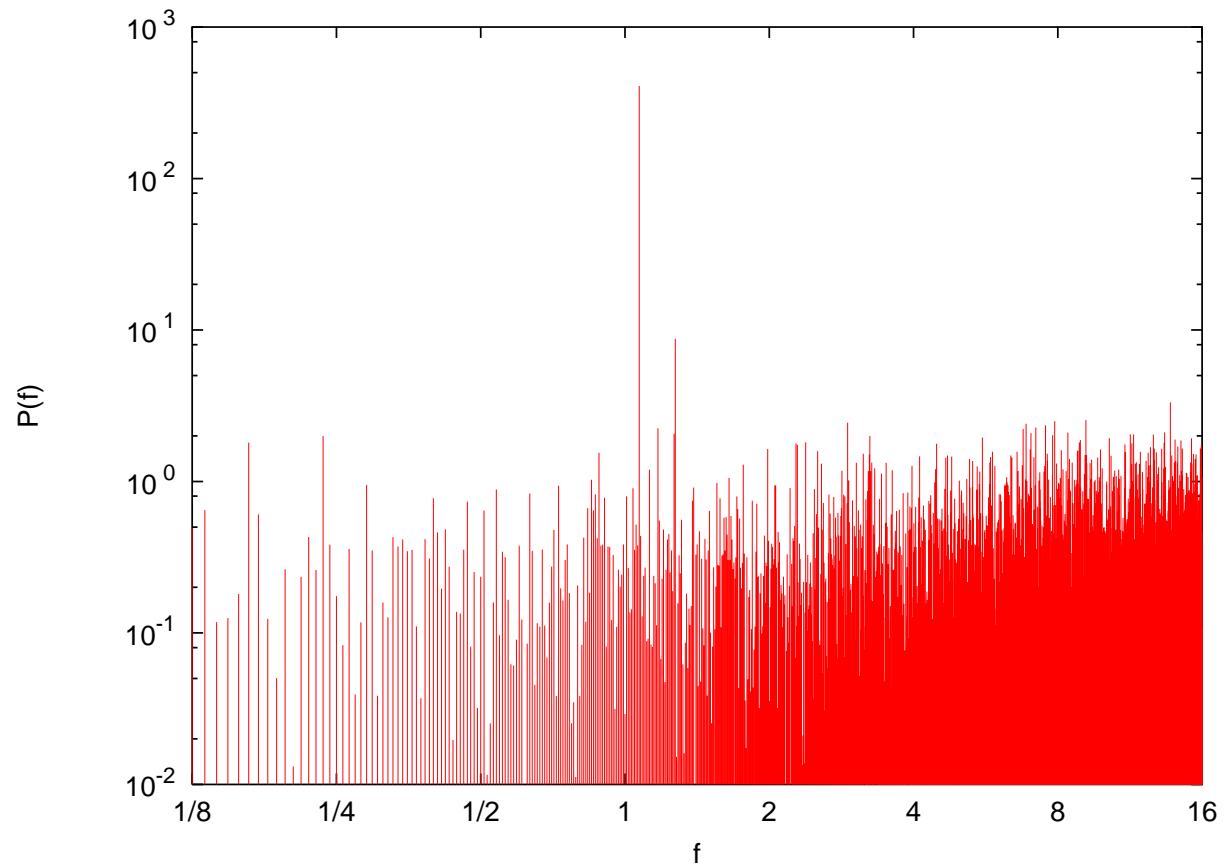
$$P(f_{N/2}) = |G(f_{N/2})|^2. \quad (12c)$$

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 (12) is to bin together terms with the same “absolute” frequencies, but different pieces of software may store them differently. And they do.

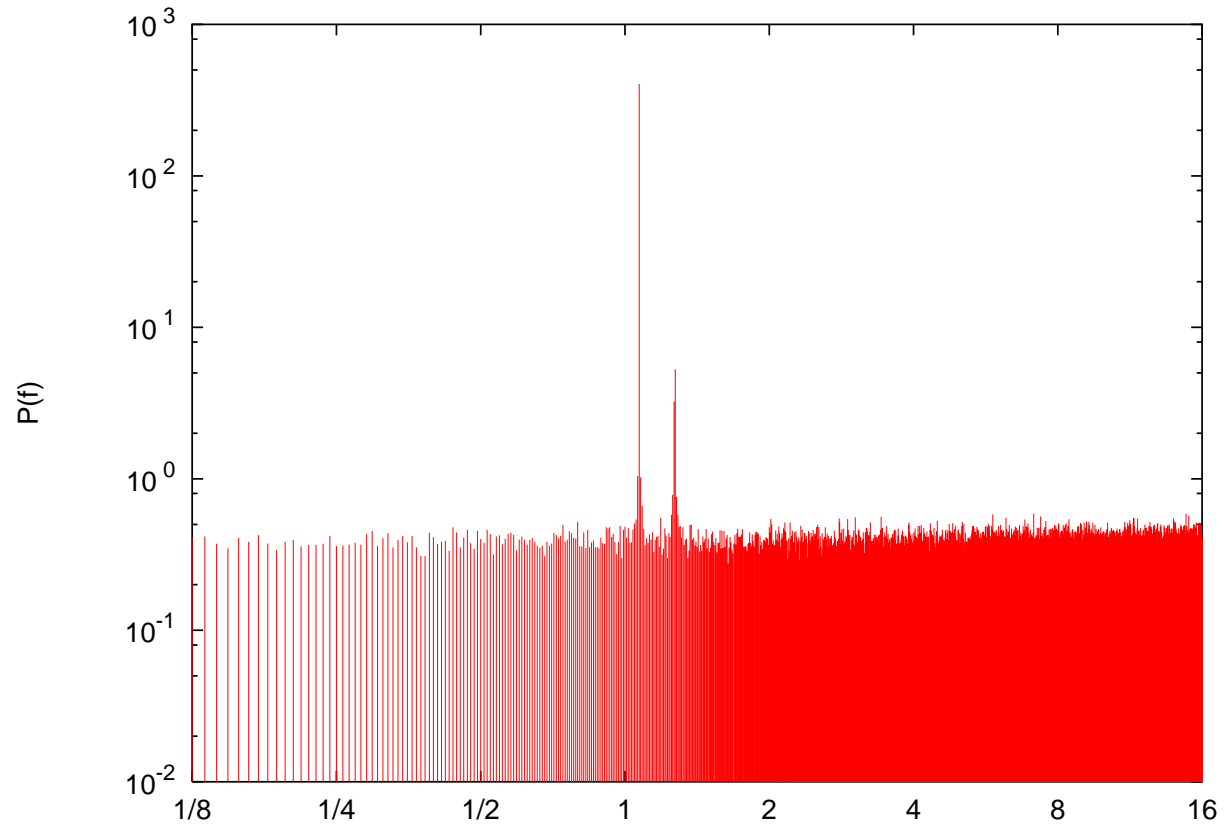
A noisy signal



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

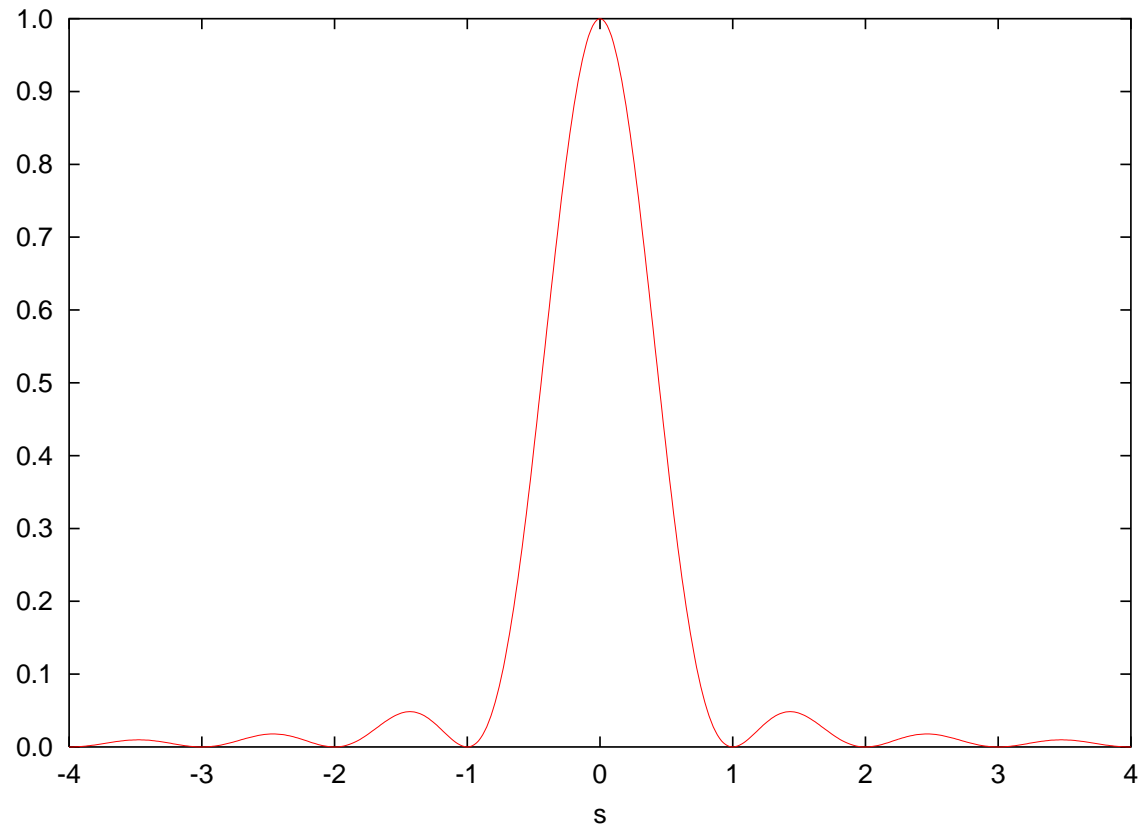
$$\text{SNR} = 10 \log_{10} \left(\frac{\text{peak power}}{\text{background power}} \right) \text{ [dB]} \quad (13)$$

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

Leakage

We expect the periodogram to project all the power from the interval $(f_n - 1/(2N\Delta), f_n + 1/(2N\Delta))$. However, as $f \neq f_n$ have non-zero projections onto $f_{m \neq n}$, the power spectrum leaks to other channels. The leakage to a channel offset by s is given by “the golden formula”:

$$P(k \rightarrow s) = \frac{1}{N^2} \left[\frac{\sin(\pi s)}{\sin(\pi s/N)} \right]^2. \quad (14)$$



Power leakage, formula (14)

Window function

To prevent the leakage, sometimes the series is multiplied by a function that kills its ends but is close to unity in the middle; this is supposed to smooth the series. Then the DFT of the modified series is calculated:

$$D_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} g_k w_k e^{2\pi i n k / N}. \quad (15)$$

$\{w_k\}$ is the window function.

Note that this is equivalent to taking a convolution of the signal and the window function *in the Fourier domain!*

The modified periodogram reads

$$P(0) = \frac{1}{W} |D_0|^2, \quad (16a)$$

$$P(f_n) = \frac{1}{W} [|D_n|^2 + |D_{-n}|^2], \quad (16b)$$

$$P(f_{Nyq}) = \frac{1}{W} |D_{N/2}|^2, \quad (16c)$$

$$W = \sum_{k=0}^{N-1} w_k^2. \quad (16d)$$

Most popular window functions

Barlett window:

$$w_k = 1 - \left| \frac{k - N/2}{N/2} \right|. \quad (17)$$

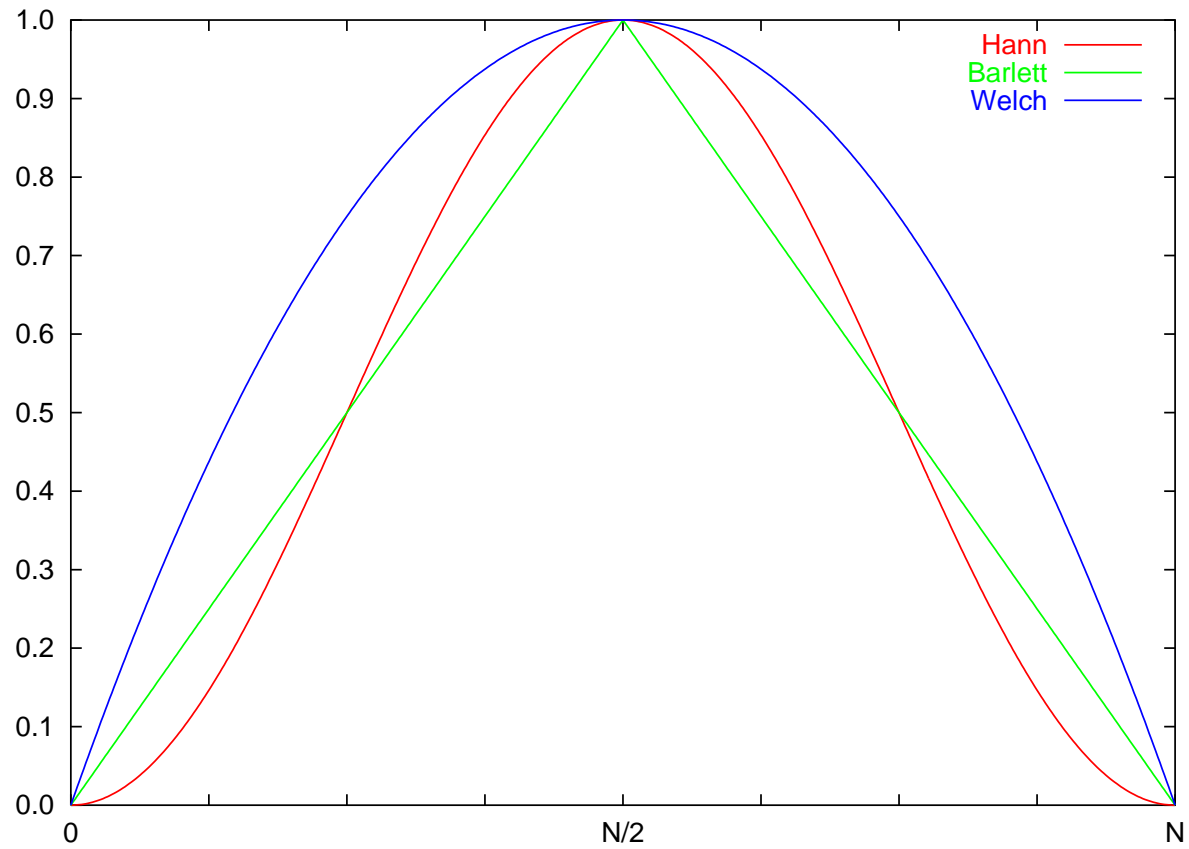
Hann window:

$$w_k = \frac{1}{2} \left[1 - \cos \left(\frac{2\pi k}{N} \right) \right]. \quad (18)$$

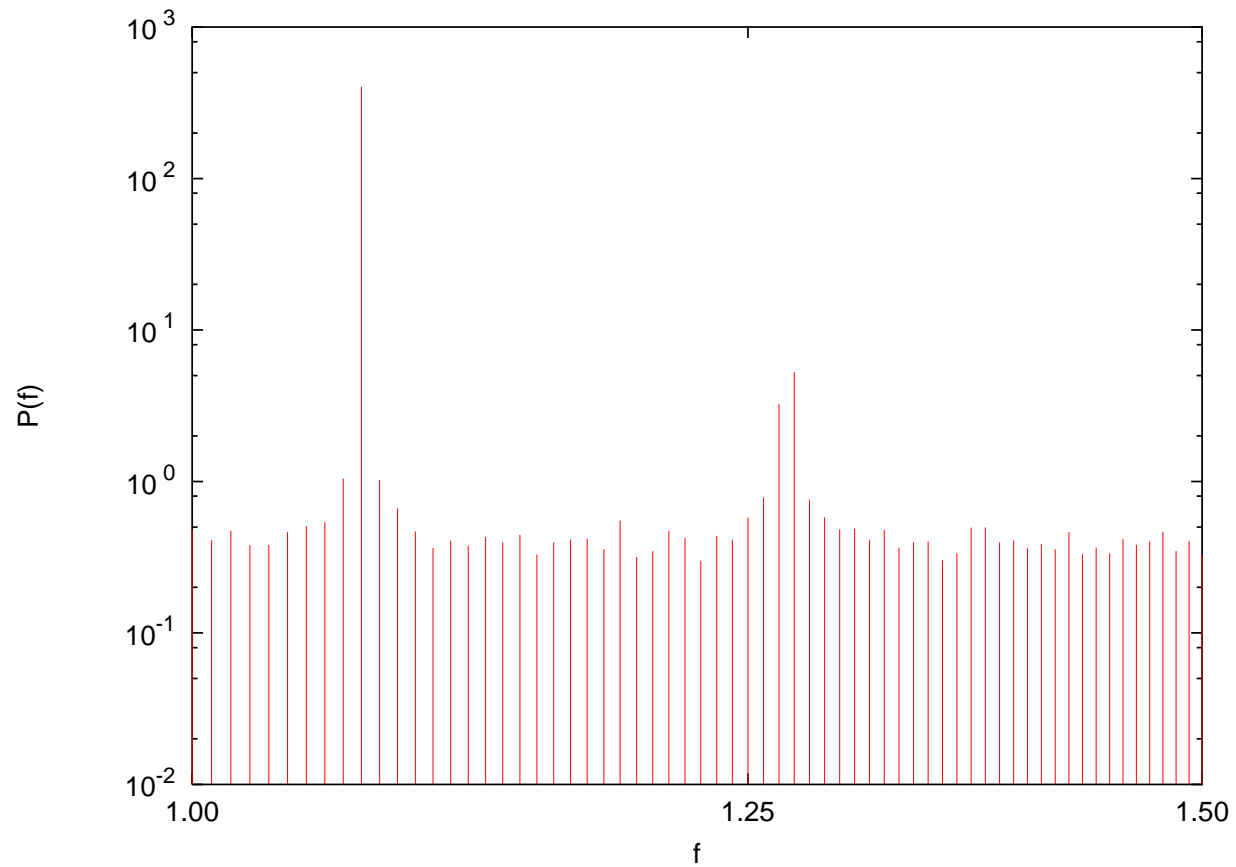
Welch window:

$$w_k = 1 - \left(\frac{k - N/2}{N/2} \right)^2. \quad (19)$$

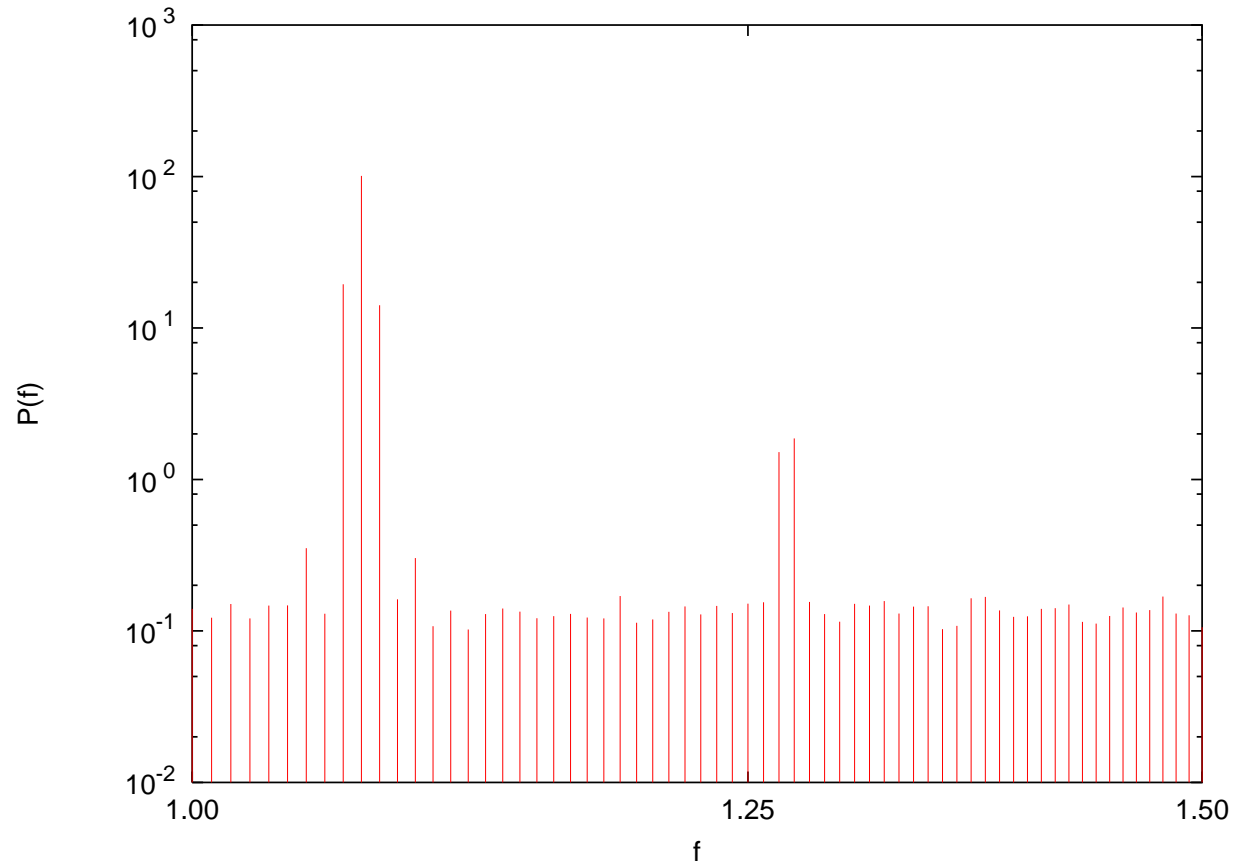
Windowing function



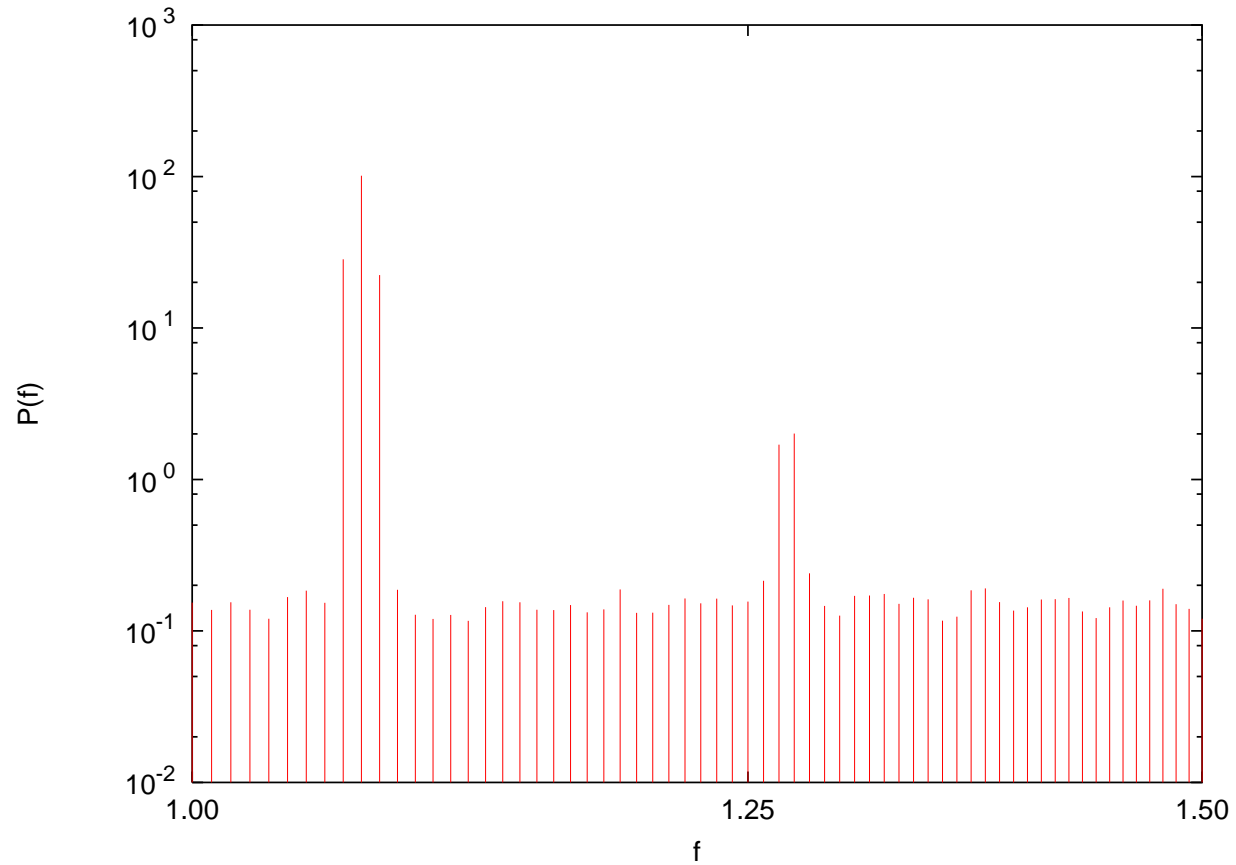
The averaged spectrum, square window



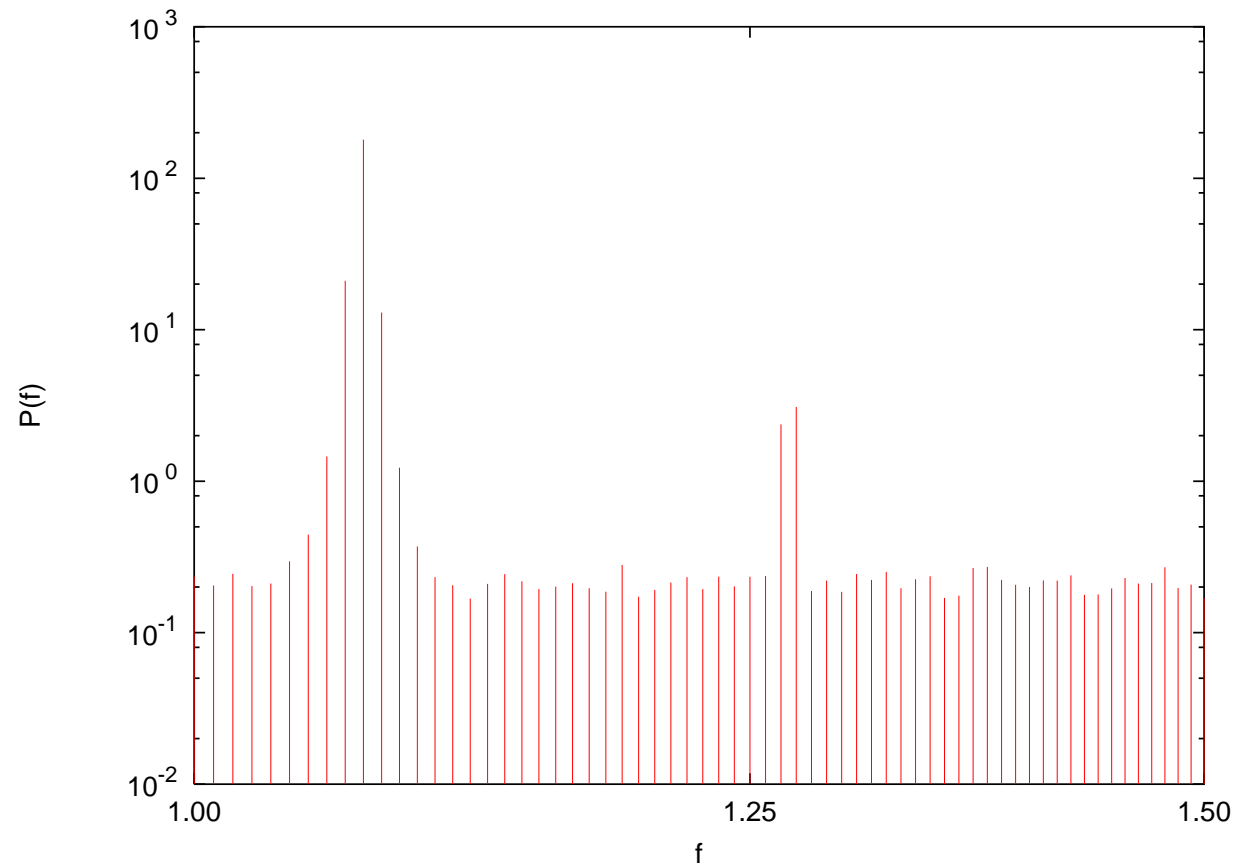
The averaged spectrum, Barlett window

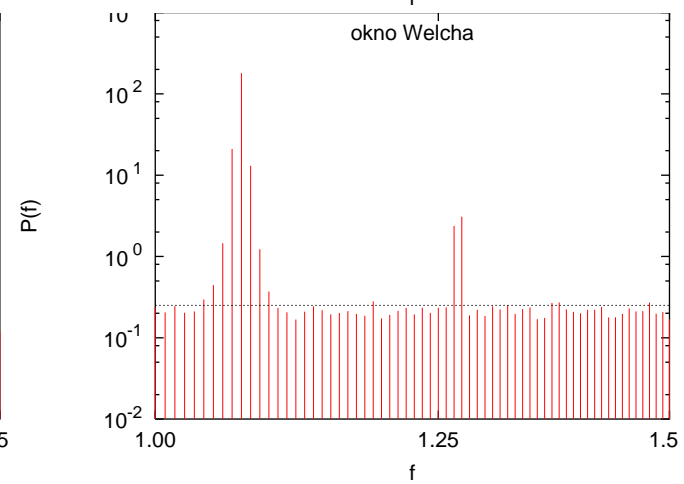
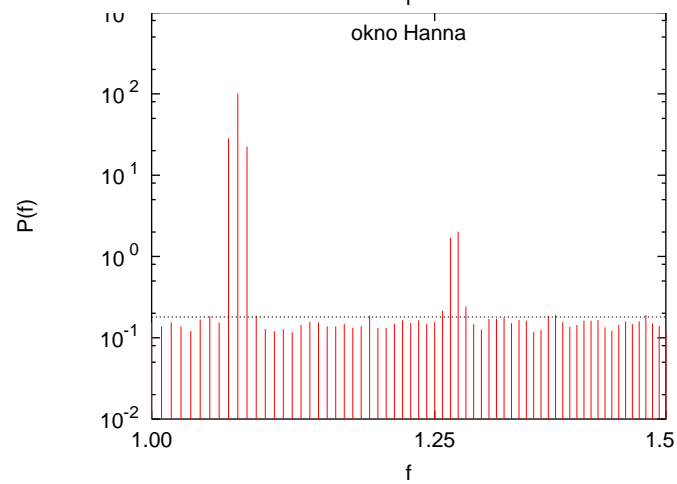
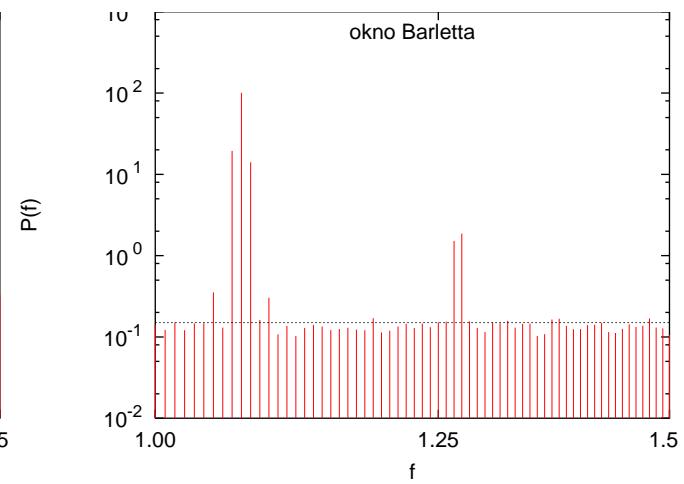
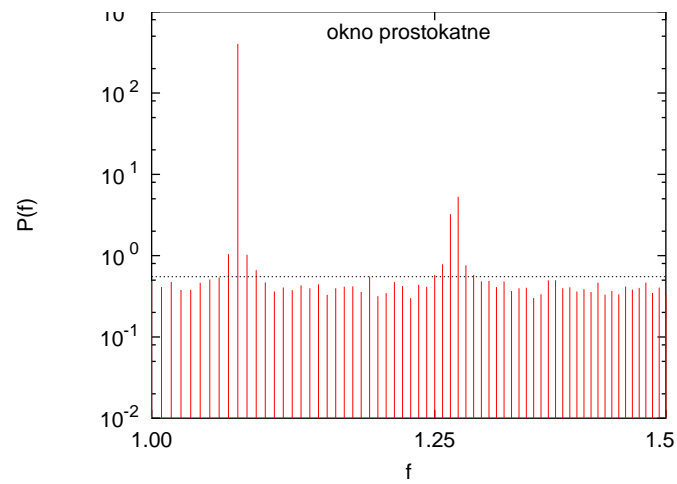


The averaged spectrum, Hann window



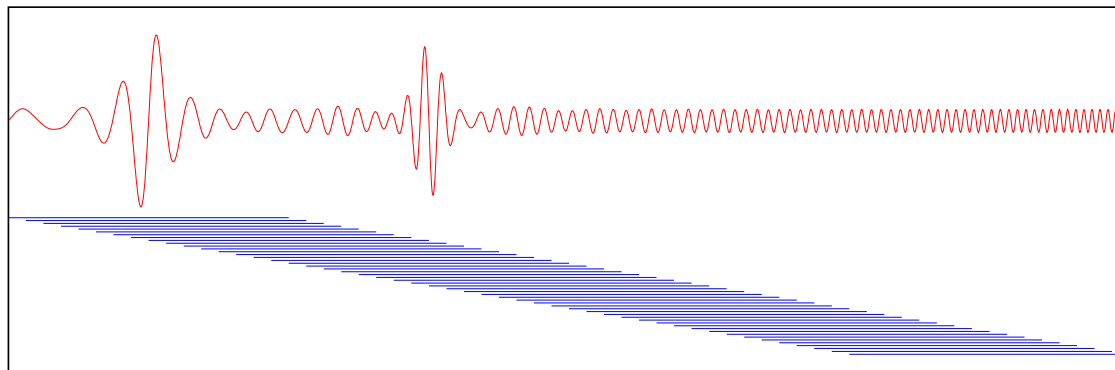
The averaged spectrum, Welch window

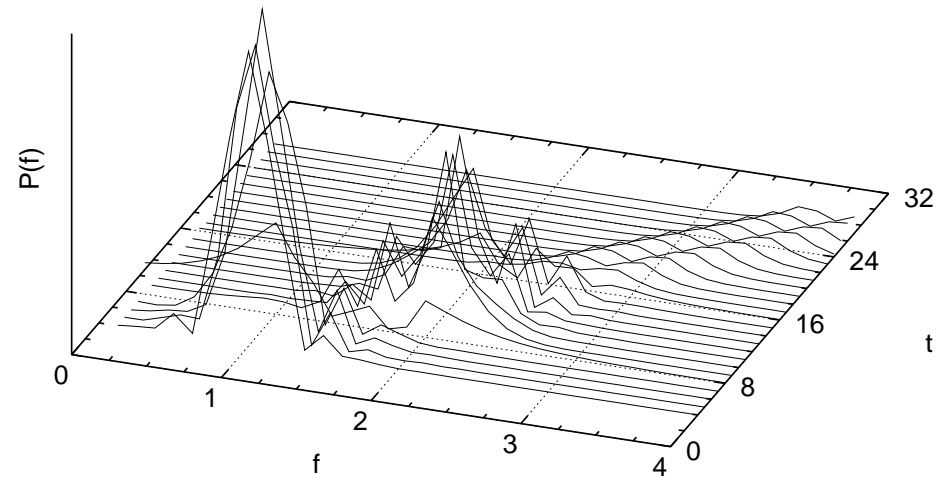
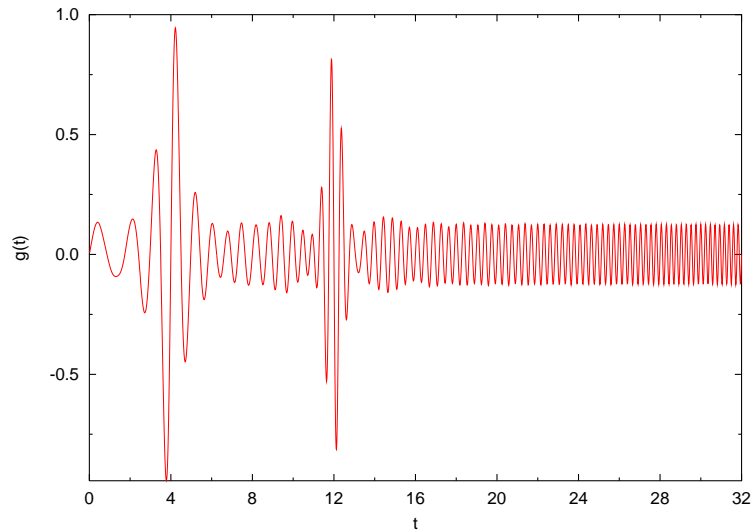




Power spectrum of nonstationary signals

To calculate the time-dependent power spectrum of nonstationary signals, divide the signal into overlapping segments; ideally, each segment should be approximately stationary. Then calculate the power spectrum for each segment.

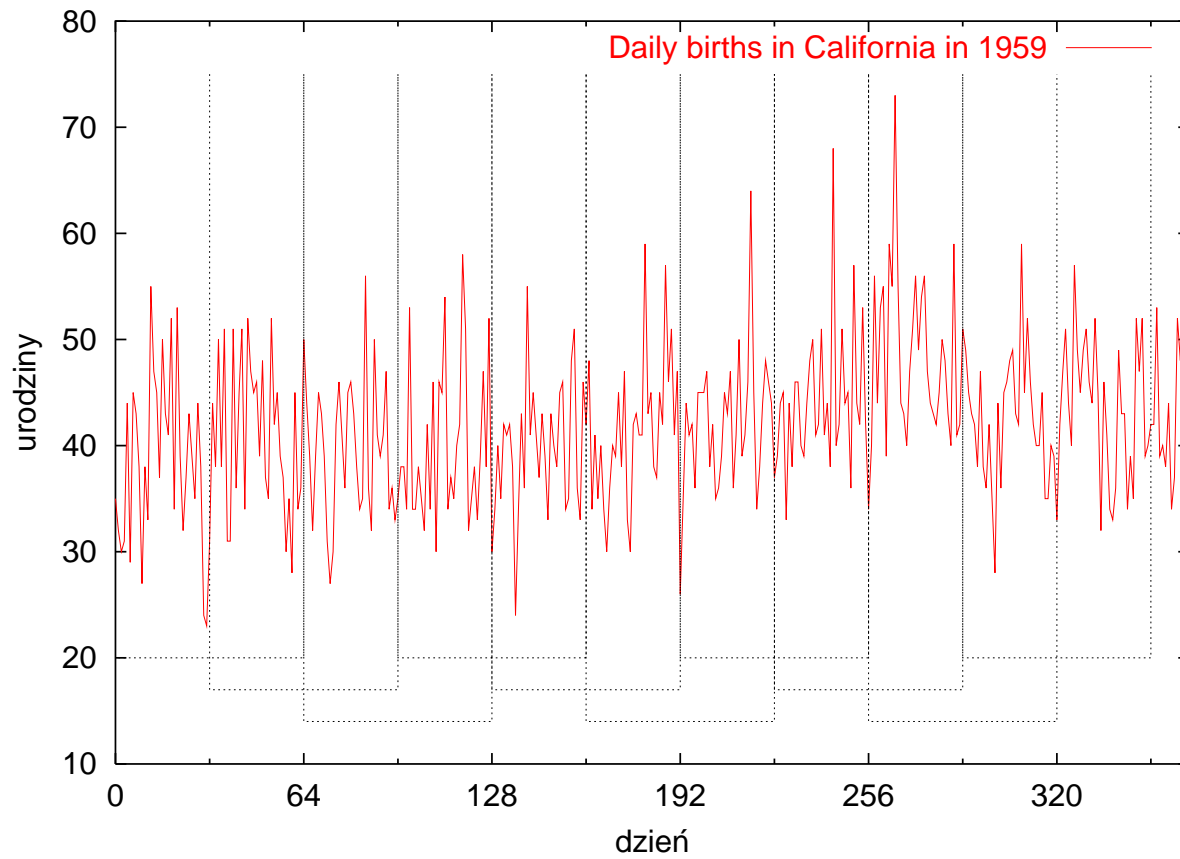




Left — a nonstationary signal. Right — its time dependent spectrum.

The power spectrum changes over the time — the autocorrelation depends not only on the offset, but also on the point around which it is evaluated.

Another example of nonstationary data



Time dependent spectrum

