Sequential Data - 1D, cont. 2

F N (Florence Nightingale) David (1909 - 1993)

Research Assistant with Karl Pearson 1931

Attended Fisher lectures 1932

1938 PhD in Statistics, UC London

1939-1945 Ministry of Home Security: statistical models to predict consequences of bombing in central London. Vital during the 1940-41 blitz; services kept running; and she updated the priors!

1962 Professorship, UCL

1969 Head of Dept of Statistics Univ Calif Riverside

1992 Elizabeth Scott award for "opening the door to women in statistics"

Famed and acclaimed as a wonderful teacher, researcher, outreacher, great generosity, charm, terrifyingly short fuse, ever-present cigars.

What was #18 all about?

Using 1D filtering:

Low-pass filtering (example: to cut noise and improve s/n)

High-pass filtering (example: "baseline assessment")

The correlation function:

What it is

Using it to assess the "useful" portion of the spectrum in the output of a physical system

The digital correlator

At radio frequencies, frequency resolution is achieved with a **correlator.**

We have stream of sampled data from a receiver, $\boldsymbol{f_{t1}}$, $\boldsymbol{f_{t2}}$, ***…***

1. Correlator takes short chunks of these data and forms the autocorrelation function (a fast operation in hardware).

2. Separate estimates of the correlation function are averaged, and

3. Fourier transformed to obtain (via the Wiener-Khinchine theorem) the power spectrum of the data.

The Digital Correlator, continued

Physically, our stream of data will consist of many wave packets, each corresponding to emission from a single atom or molecule. Thus the time series of, say, electric field amplitudes will be

$$f(t) = \sum_i w(t + \phi_i), \quad \text{with FT:} \quad F(\omega) = W(\omega) \sum_i \exp(\imath\omega\phi_i).$$

where $\boldsymbol{\phi_i}$ are the random phases of each wave packet $\boldsymbol{w}$. The average power spectrum will be

$$\mid W(\omega) \mid^2 E[\mid \sum_i \exp(\imath\omega\phi_i) \mid^2].$$

The exponential term, being an average of positive quantities, will converge to some positive value as more and more chunks of data are averaged - even though the phases are random. By contrast, the average Fourier transform will contain the term

$$E[\sum_i \exp(\imath\omega\phi_i)]$$

.... which will converge to zero.

The Digital Correlator - wrap-up

Key feature of the digital correlator - **quantization**

Little sensitivity is lost by **digitizing at the one-bit level**, ***f*** is positive or negative.

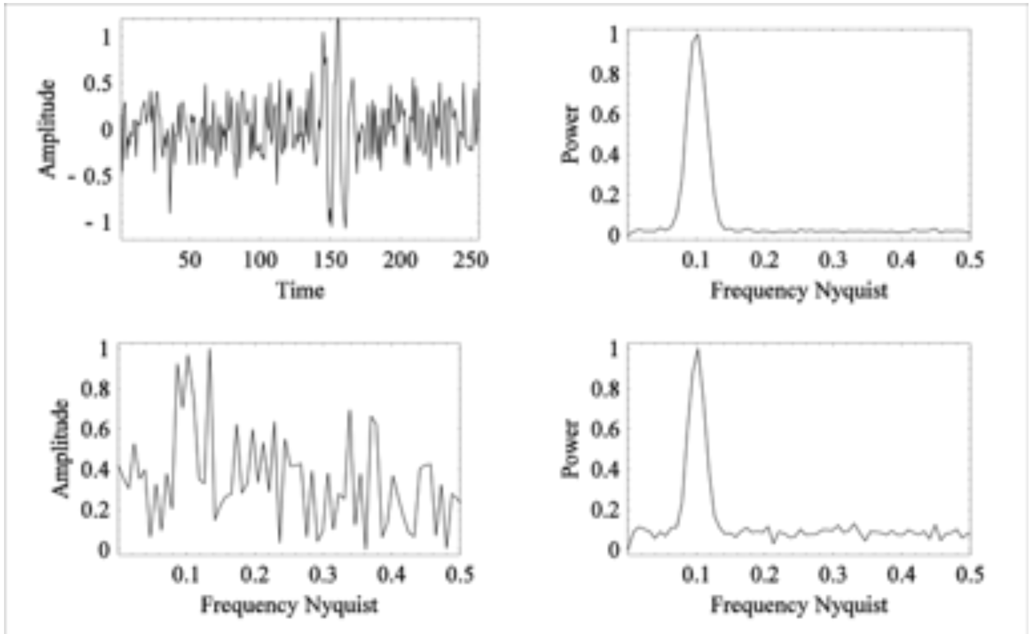
This **speeds up data-rates** and **reduces operations.**

Higher resolution is possible, this dependent on the number of channels in the shift-and-add of the correlator, rather than the sampling speed of the data (as long as this is high enough to exceed the Nyquist criterion).

Given the **correlation coefficient ρ** between the data values f_{ti} and f_{tj}, the **quantized correlation coefficient ρ_q** can be calculated by marginalization. The result is the **van Vleck equation**:

$$\rho_q = \frac{2}{\pi} \sin^{-1} \rho$$

The Digital Correlator - Example



Top left: part of the input data stream for the correlator, consisting of 64 wave packets, randomly located, with on average one per 128 units of time.

Top right: the derived power spectrum from forming the autocorrelation function over 128 time units.

Bottom left: the average Fourier transform of **1-bit** quantized data, again averaged in 128-long chunks.

Bottom right: the power spectrum derived from the quantized data with the same averaging. Almost the same!

Unevenly-Sampled Data - The Periodogram

E.g. the search for **periodicities in light curves of objects of variable luminosity**.

- **uneven sampling** from: daytime, bad weather, or bad time-assignment cttees.

- most modern analysis is based on the **Lomb-Scargle** method (see Num Rec).

- key features: (1) method **weights the data on a 'per point'** basis rather than on a 'per time interval' basis as does the FFT;

  (2) **null hypothesis** can be tested rigorously.

- If peak at frequency ***ω***, probability that height of peak ***Y(ω)*** lies between ***Y (>0)*** and ***Y + dY*** is ***exp(-Y)dY.*** If ***n* independent frequencies** are considered, then the probability that none gives a value **>*Y*** is $(1 - e^{-Y})^n$. Thus

$$P(>Y) = 1 - (1 - e^{-Y})^n$$

represents the significance level of any peak ***Y(ω).***

The Periodogram continued ....

BUT:

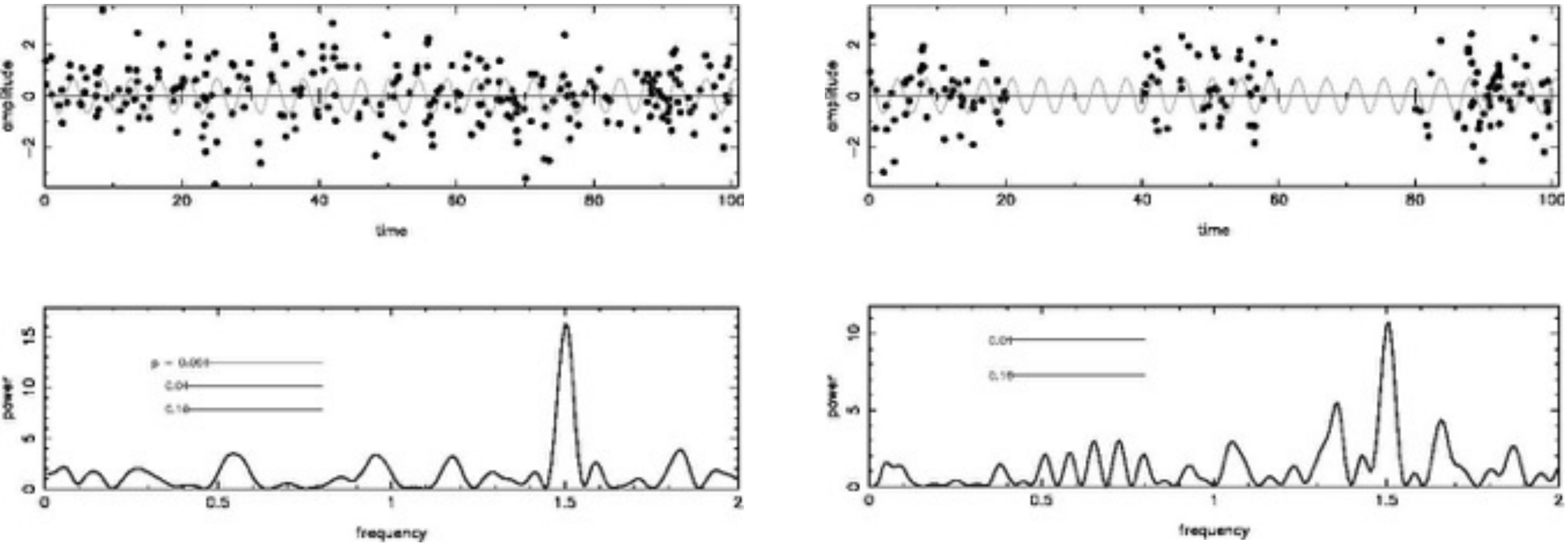
- **What is *n*** - how many independent frequencies have we looked at?

- In the limit of interest, when significance levels are << 1, $P(>Y) = ne^{-Y}$, scaling linearly with the estimate of ***n***, so ***n*** need not be estimated precisely.

- MC experiments: if ***N*** is the number of scattered but approximately evenly-spaced data points which oversample the range up to the Nyquist frequency, then $n \sim N$, and there is **little difference for *n*** between random spacing and equal spacing.

- When a **larger frequency range** is sampled, ***n*** increases proportionally.

The Periodogram - Example, Lomb-Scargle



Left - randomly-spaced data generated by a sine-wave of amplitude 0.5 and period 0.6 with noise of unit variance superposed.

Right - same rate of data but with gaps ~ night-to-night sampling of optical astronomy. e.g..

For the continuous data, even with the sine wave shown as a guide, the **eye cannot pick out the periodicity**. For gapped data, note **reduced significance** of the peak and serious **aliasing** resulting from windowing the data.

The Periodogram - Frequencies Beyond Nyquist?

Can we sample **frequencies beyond the Nyquist**?

- Recall that the Nyquist frequency refers to **equally spaced data;** with sampling interval Δt, it is $2\pi/\Delta t$.

- With randomly-spaced data evenly distributed through the sampling series, an equivalent (but non-physical) Nyquist frequency can be obtained from the mean time-interval.

- **The fundamental limitation of equally-spaced data is avoided by unequally-spaced data!** It is possible to sample well **above** the equivalent Nyquist frequency without significant aliasing.

- cf 2D, 3D - **clustering on scales much smaller** than the mean separation between objects can be determined if the objects are randomly sampled.

The Periodogram - Serious `Clumping'

What really happens with **serious clumping**, e.g. observations all stuffed into a few night-time hours?

- **MC again**: generate **synthetic data sets** of Gaussian noise by holding fixed the number of data points and their sampled locations, find the **largest values of $Y(\omega)$,** and find the best fit of the distribution to determine $\boldsymbol{n}$.

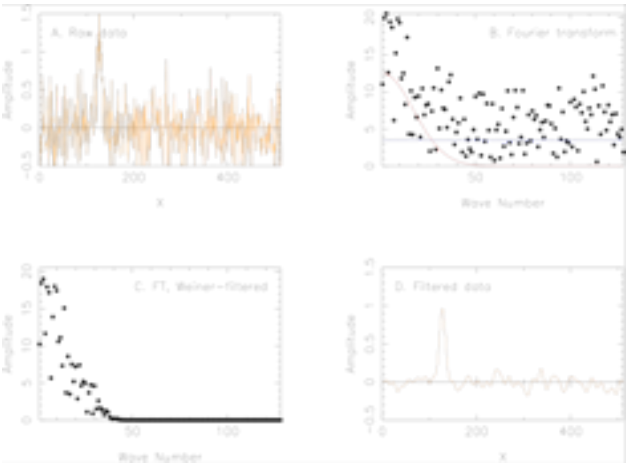
- Gappy data: **aliasing becomes serious**. With data of even poorer quality than that shown (no problem for astronomers), **choosing the right peak** is the issue.

- **Folding techniques**: **observing a similar data stream some time later** will enable a choice to be made. **Only one** of the frequencies will have the right phase to fit. The pulsar people are experts at this.

Wavelets - Why Isn't Fourier All We Need?

Fourier analysis (1) loses information about where in a scan things may be happening, and (2) can wipe some low-frequency information.

E.g. our low-pass filtering example:



Bad no. (3): The noise level might be different in the spectral line, but a Fourier filter applies the same degree of smoothing everywhere.

- These are results of the basis functions, the sin's and cos's being infinite in extent.

- They are the cause of many of the difficulties associated with transforms of finite-length data streams.

We would like a transform which picks out details of frequency content and preserves information about where in the scan those particular frequencies are prominent.

So What Are Wavelets?

A wavelet is a short function such that, being convolved with the data, it gives some frequency (or scale) information at a particular location in the scan.

Placing the wavelets at different places in the scan = “translating" + changing widths = “scaling" => frequency decomposition with location information.

Mathematical restrictions on what kind of function can be a wavelet.

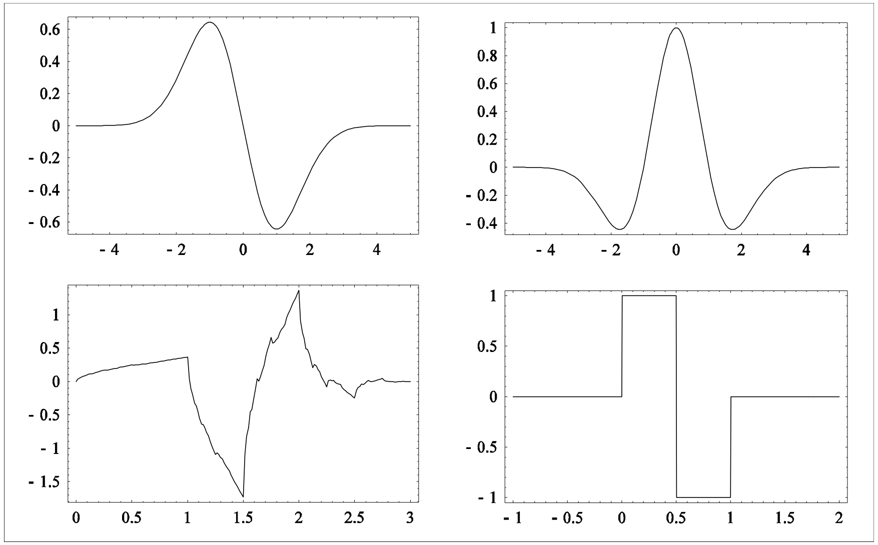
EXAMPLES of wavelets in current use

Top left, asymmetrical

Top right, Mexican Hat

Bottom left, Daubechies (a fractal)

Bottom right, Haar



Wavelet decomposition will be in terms of scaled and translated versions of each of these.

Wavelets continued

Different wavelets are to be sensitive to different aspects

- asymmetrical wavelet will be sensitive to local gradients

- Mexican hat will be good at picking out oscillations

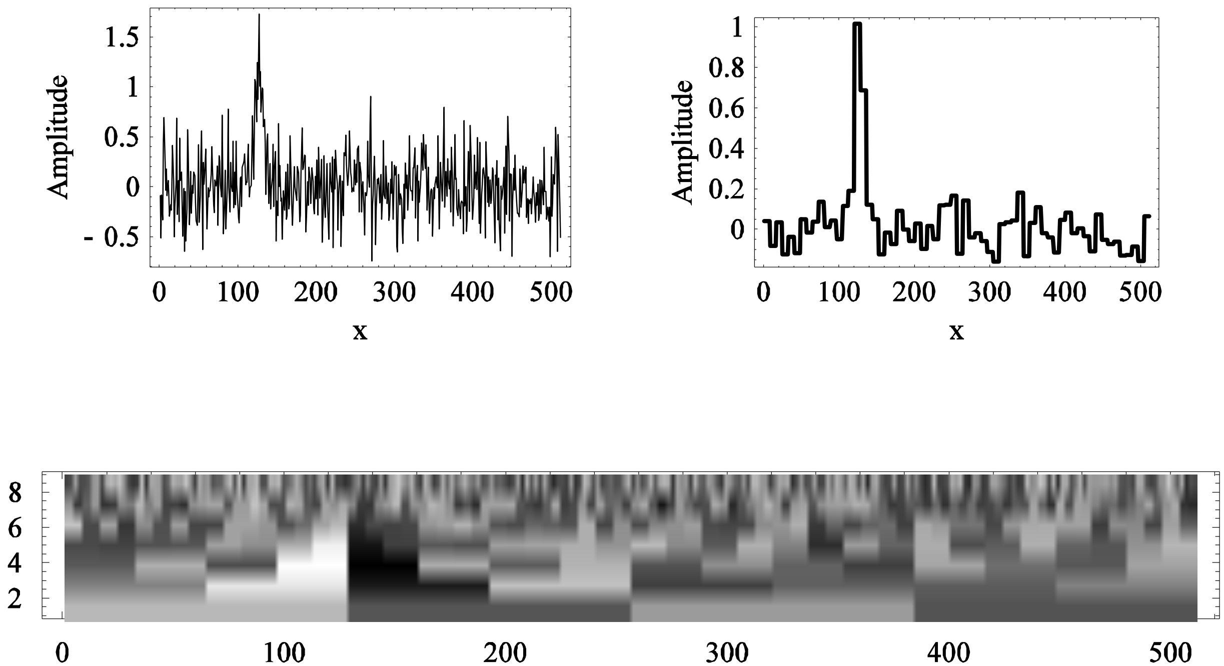
Very effective filtering and data compression.

- FBI uses a wavelet-based technique for the digitization and compression of their fingerprint database.

- In astronomy, we have scans with **important properties changing from place to place**, e.g. noisy regions in a spectrum, or when a light curve shows sudden change in behaviour such as quasi-periodicity.

- Wavelets offer new possibilities in data assessment, and a whole **new armoury of filtering techniques**, especially those where the filtering may be different in different parts of a scan.

Wavelets - Example Using Haar Wavelets



Top left, original spectrum.
Top right, the filtered spectrum.
Below, wavelet coefficients as a function of location and scale.

Dropping the three finest scales of wavelet coefficients is a suitable simple filter. The result : noise is much reduced **without loss of resolution** in the spectral line.

Detection Difficulties - 1/f Noise

***1/f* noise** has **a power spectrum inversely proportional to the Fourier variable** – frequency, if we are dealing with a time series.

Sometimes called **flicker noise**, it is a particular case of “pink” noise, **in which low frequencies** dominate.

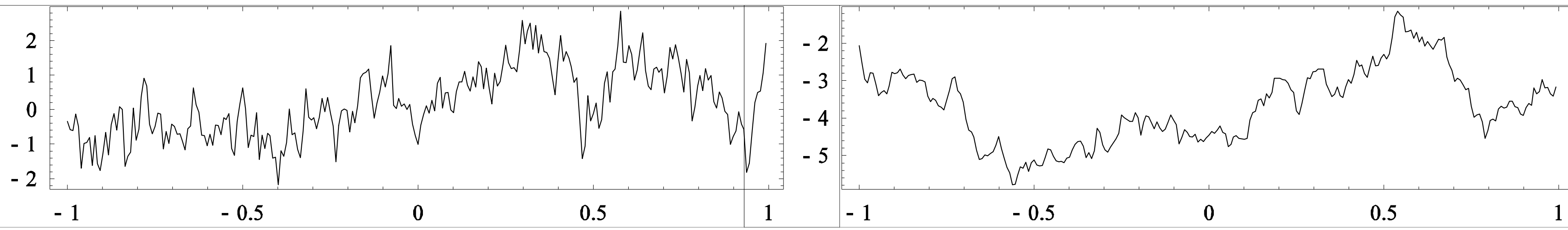
More extreme example - **Brownian or random-walk noise** : successive values of the noise are obtained by adding a random number to the previous value. Arises when we integrate a scan.

***1/f* occurs everywhere** (Beethoven symphonies or rock music or traffic flow….); this is hard to understand.

Its presence (or the presence of one of its near relatives) is **usually why averaging large amounts of data does not produce the improvement expected**.

1/f Noise continued

1/f noise is the curse of the experimenter's life



- it is why **filtering theory**, looking so good in simulations, **fails** to live up to promise;

- why **2σ results** are not results, even though the probability is >95%

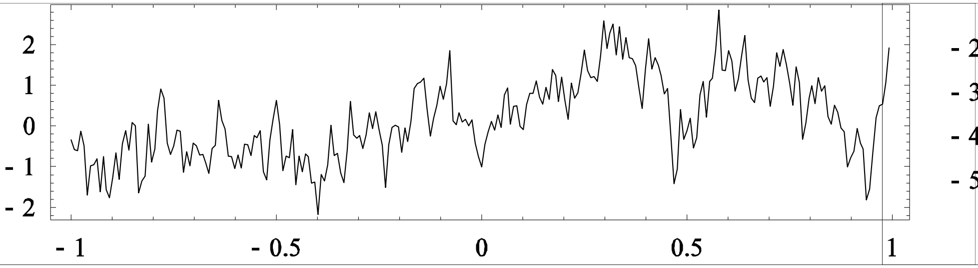
- why **increased integration time fails to improve s/n** according to ***1/√N*** which we naively expect from averaging ***N*** samples.

1/f Noise continued

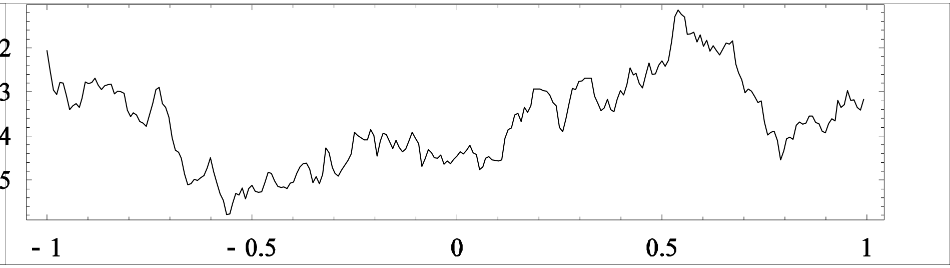
The **variance on a scan is the integral of its power spectrum**.

For finite length, variance ∝ the integral of power spectrum between Nyquist frequency and the first frequency above zero (1/L if L is the scan length).

So **variance of sampled *1/f* noise** will be $\propto \int_{1/L}^{f_{\mathrm{nyq}}} \frac{1}{f}\, df = \log(L f_{\mathrm{nyq}})$



Left: flicker (1/f) noise of unit variance



Right: a random walk of unit variance

So it **grows logarithmically** with scan length! ***1/f* noise** has infinite variance!

The noise is **highly correlated** from one sample to the next. Averaging? Useless?

Averaging **does not work at all for a noise spectrum of *1/f* or steeper.**

1/f Noise concluded

Variance on a mean derived from a scan *f* of length *L*, is $\mathrm{var}[\hat{\mu}] = \mid F(0) \mid^2 / L.$

Thus for white noise, we have $|F(0)|^2 = \sigma^2$, **and** the expected **$1/\sqrt{L}$ behaviour** follows.

But ***1/f* noise**? Best idea of power spectrum at zero is its **value at a frequency *1/L***. Now we can see that **the variance on the mean is independent of *L***!

Usually we will have **white noise dominating the power spectrum for frequencies greater than some value ω_0**, i.e. for **scans shorter than $1/\omega_0$**.

As scans lengthen, we uncover the ***1/f* noise below ω_0**, that's right, down where our signal is.

General model for the variance on the mean level of **a scan of length *L*** will be

$$\mathrm{var}[\hat{\mu}] \simeq \frac{a}{L} + b$$

where ***a*** and ***b*** describe the levels in **white noise** and ***1/f*** noise.

Note the analogy to low-pass and high-pass filtering: dealing with the slowly-varying component may be considered as a baseline issue.

END

Times of Arrival - the Rayleigh Test

Consider **small numbers of events**, times of arrival of pulses (pulsars) or photons (gamma-ray astronomy). **Do these times betray a period?**
The Rayleigh test is a classical test :

(1) Have a **period P** in mind.
(2) Call the times of arrival ***t1, t2, …***
(3) Assign a **phase** to each time by the algorithm $\phi_i = 2\pi\ [\mathrm{remainder\ of}(t_i/P)]$
(4) Form the statistic

$$R^2 = (\sum_{i=1}^{n} \cos\phi_i)^2 + (\sum_{i=1}^{n} \sin\phi_i)^2$$

and for ***n* > *10***, ***2R²/n*** is distributed as ***X²* with two degrees of freedom.** If ***R*** is large, it is unlikely that the phases are random - we have guessed the correct period, so we would then infer that **the period is indeed *P***.

We may also wish **to determine *P***, which we would do simply by searching for **a value of *P* that maximizes *R***. Having determined a parameter from the data, we **lose one degree of freedom** from ***X²*** in the significance test.

Details of this and more elaborate tests are in De Jager, Swanepol & Raubenheimer (1989)