Lecture 14: Variability and Periodicity

Outline

1. Variable and Periodic Signals in Astronomy
2. Lomb-Scarle diagrams
3. Phase dispersion minimisation
4. Kolmogorov-Smirnov tests
5. Fourier Analysis

Variable and Periodic Signals in Astronomy

Examples

- variable stars (Cepheids, eclipsing/interacting binaries)
- magnetic activity (spots, flares, activity cycles)
- exoplanets (Doppler, transients, micro-lensing)
- pulsars, neutron star QPO
- gravitational lensing
- transients (flare stars, novae, supernovae, GRB)
- new synoptic telescopes: LSST, Pan-STARRS, VST

Finding Variability and Periodicity

Problems:

- uneven sampling
- data gaps, sometimes periodic
- variable noise
- variability of Earth atmosphere, instrument, detector

Testing for Constant Signal

- N measurements y_i with errors σ_i at times t_i
- best guess for constant with Gaussian errors

$$\overline{y} \equiv a_{min} = \frac{\sum_{i=1}^{N} \frac{y_i}{\sigma_i{}^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i{}^2}}$$

- minimizes

$$\chi^2 \equiv \sum_{i=1}^{N} \chi_i{}^2 \equiv \sum_{i=1}^{N} \frac{(y_i - y_m)^2}{\sigma_i{}^2}$$

- probability that chi-squared by chance

$$P(\chi^2_{obs}) = \mathtt{gammq}((N-1)/2, \chi^2_{obs}/2)$$

- but test is often insufficient

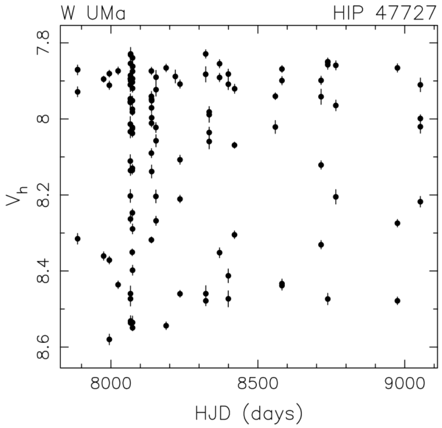
Counter Example 1

- N measurements y_i, Gaussian distribution of errors around constant value with constant error σ
- observed chi-squared due to chance
- re-order measurements such that $y_N \geq y_{N-1} \geq \dots y_2 \geq y_1$
- new time series has same chi-squared, but cannot be obtained by chance
- significant increase of y_i with time not uncovered by chi-squared test

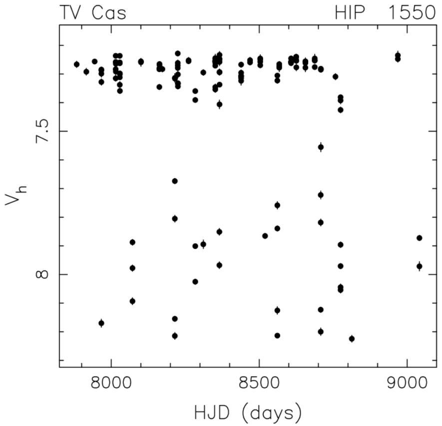
Counter Example 2

- same series y_i with measurements at equidistant time intervals $t_i = i \times \Delta t$
- order y_i so that higher values are assigned to t_i with even i and lower values to t_i with odd i
- significant periodicity present in re-ordered data not uncovered by chi-squared test
- if time series is long enough, can uncover significant variability from other tests

W UMa (pulsation variable)



TV Cas



- data obtained by Hipparcos
- source is significantly variable (variations large compared to error bars
- due to observing method, data taken at irregular intervals

Fitting sine-functions: Lomb-Scargle

- fit (co)sine curve

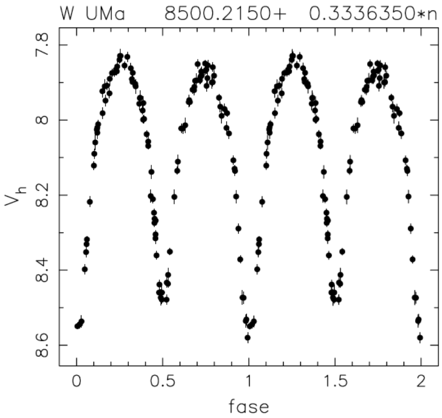
$$V_h = a\cos(\omega t - \phi_o) = A\cos\omega t + B\sin\omega t$$

- A, B related to a, ϕ_o by

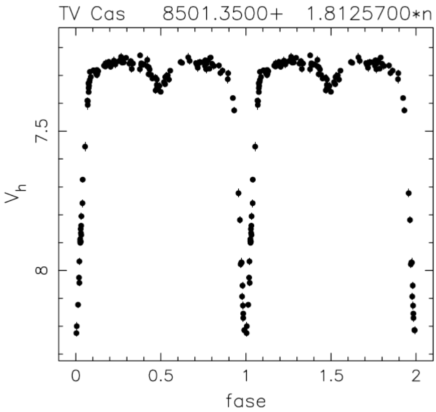
$$a^2 = A^2 + B^2; \qquad \tan\phi_o = \frac{B}{A}$$

- fit a, ϕ_o, $\omega \equiv 2\pi/P$ by minimizing sum of chi-squares
- specialized method developed by Lomb (1967), improved by Scargle (1982), Horne & Baliunas (1986), and Press & Rybicki (1989)
- see Numerical Recipes, Ch. 13.8

W UMa



TV Cas



Folded Light-Curve

- W UMa roughly sinusoidal: Lomb-Scargle works well
- note two maxima and minima in each period

Period Folding

- TV Cas: folded light curve very different from sine
- Lomb-Scargle may not be optimally efficient in finding period
- Stellingwerf (1978, ApJ 224, 953) developed method working for lightcurves of arbitrary forms
- Fold data on trial period to produce folded lightcurve
- divide folded lightcurve into M bins
- if period is (almost) correct, variance ${s_j}^2$ inside each bin $j \in 1, M$ is small
- if period is wrong, variance in each bin is almost the same as the total variance
- best period has lowest value for $\sum_{j=1}^{M} {s_j}^2$

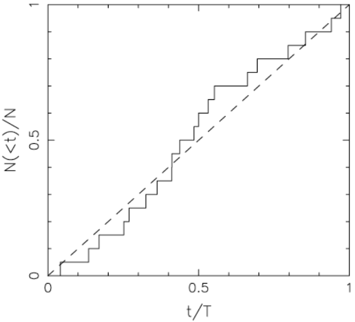
False Alarm Probability

- probability that result is due to chance
- analytic derivation of this probability is difficult
- often safest estimate obtained by simulations
- N measurements y_i at t_i
- scramble data and apply Lomb-Scargle or Stellingwerf method
- scrambled data should not have periodicity
- many scrambles $\Rightarrow$ distribution of significances that arises due to chance, probability that period obtained from actual data is due to chance
- this probability is often called *false-alarm probability*

Variability through Kolmogorov-Smirnov (KS) tests

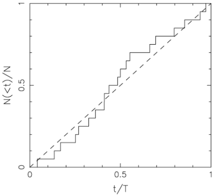
- data may be variable without strict periodicity
- consider detector exposing for T seconds detecting N photons
- M bins of equal length $T/(M-1)$
- constant source: $n = N/(M-1)$ photons per bin
- test with chi-squared or maximum-likelihood test
- loss of information by binning
- result depends on the number of bins chosen
- Kolmogorov-Smirnov test (KS-test) test avoids these problems
- KS test computes probability that two distributions are the same
- computes probability that two distributions have been drawn from the same parent
- one-sided KS-test compares theoretical distribution without errors with observed distribution
- two-sided KS-test compares two observed distributions, each of which has errors

Kolmogorov-Smirnov Test Example



- number of photons from constant source increase linearly with time
- normalize total number N of detected photons to 1
- theoretical expectation: normalized number of photons $N(< t)$ arriving before time t increases linearly with t from 0 at $t = 0$ to 1 at $t = T$

Kolmogorov-Smirnov Test Example (continued)



- observed distribution is a histogram which starts at 0 for $t = 0$, and increases with $1/N$ at each time $t_i, i \in 1, N$ that a photon arrives
- determine largest difference d between theoretical curve and observed curve
- KS-test gives probability that a difference d or larger arises in a sample of N photons due to chance
- KS-test takes into account that for large N one expects any d arising due to chance to be smaller than in a small sample

Fourier transforms

Introduction

- periodic signal "builds up" with time
- discover periodic signal in long time series, even if signal is small with respect to noise level
- best for un-interrupted series at equidistant intervals
- data gaps lead to spurious periodicities
- can remove spurious periodicities ('cleaning')
- continuous and discrete Fourier transforms
- observations $\Rightarrow$ only discrete transform

Continuous Fourier Transform

- continuous transform $a(\nu)$ of signal $x(t)$

$$a(\nu) = \int_{-\infty}^{\infty} x(t) e^{i2\pi\nu t} dt \qquad \mathrm{for} \; -\infty < \nu < \infty$$

- reverse transform

$$x(t) = \int_{-\infty}^{\infty} a(\nu) e^{-i2\pi\nu t} d\nu \qquad rmfor \; -\infty < t < \infty$$

- therefore Parseval theorem

$$\int_{-\infty}^{\infty} x(t)^2 dt = \int_{-\infty}^{\infty} a(\nu)^2 d\nu$$

- occasionally written with the cyclic frequency $\omega \equiv 2\pi\nu$
- write $e^{i2\pi\nu t}$ as $\cos(2\pi\nu t) + i \sin(2\pi\nu t) \Rightarrow$ Fourier transform gives the correlation between the time series $x(t)$ and a sine or cosine function, in terms of amplitude and phase at each frequency ν.

Discrete Fourier Transform

- series of measurements $x(t_k) \equiv x_k$ taken at times t_k
- $t_k \equiv kT/N$, T is total time for N measurements
- time step $\delta t = T/N$
- discrete Fourier transform defined at N frequencies ν_j, for $j = -N/2, \ldots, N/2 - 1$, frequency step $\delta\nu = 1/T$
- discrete versions of continuous transforms

$$a_j = \sum_{k=0}^{N-1} x_k e^{i2\pi jk/N} \qquad j = -\frac{N}{2}, -\frac{N}{2}+1, \ldots \frac{N}{2}-2, \frac{N}{2}-1$$

$$x_k = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} a_j e^{-i2\pi jk/N} \qquad k = 0, 1, 2, \ldots, N-1$$

Discrete Fourier Transform (continued)

- discrete Parseval theorem

$$\sum_{k=0}^{N-1} |x_k|^2 = \frac{1}{N} \sum_{j=-N/2}^{N/2-1} |a_j|^2$$

- occasionally also in terms of cyclic frequencies $\omega_j \equiv 2\pi\nu_j$
- $1/N$-normalization is matter of convention
- other conventions: $1/N$-term in forward transform or $1/\sqrt{N}$-term in both forward and backward transforms
- in general: both x and a are complex numbers
- x_j real $\Rightarrow a_{-j} = a_j{}^*$

Nyquist and DC Frequencies

- highest frequency is $\nu_{N/2} = 0.5N/T$ (*Nyquist frequency*)
- with $a_{-N/2} = a_{N/2}$:

$$a_{-N/2} = \sum_{k=0}^{N-1} x_k e^{-i\pi k} = \sum_{k=0}^{N-1} x_k (-1)^k = a_{N/2}$$

- may list the amplitude at the Nyquist frequency either at the positive or negative end of the series of a_j
- amplitude at zero frequency is the total number of photons:

$$a_o = \sum_{k=0}^{N-1} x_k \equiv N_{tot}$$

Parseval's Theorem

- Parseval's theorem: express variance of signal in terms of Fourier amplitudes a_j:

$$\sum_{k=0}^{N-1} (x_k - \overline{x})^2 = \sum_{k=0}^{N-1} {x_k}^2 - \frac{1}{N}\left(\sum_{k=0}^{N-1} x_k\right)^2 = \frac{1}{N}\sum_{j=-N/2}^{N/2-1} |a_j|^2 - \frac{1}{N}{a_o}^2$$

- discrete Fourier transform converts N measurements x_k into $N/2$ complex Fourier amplitudes $a_j = {a_{-j}}^*$
- each Fourier amplitude has amplitude and phase

$$a_j = |a_j| e^{i\phi_j}$$

- if the N measurements are uncorrelated, the N numbers (amplitudes and phases) associated with the $N/2$ Fourier amplitudes are uncorrelated as well

Correlations in Real and Fourier Spaces

$$\sum_{k=0}^{N-1} \sin \omega_j k = 0, \qquad \sum_{k=0}^{N-1} \cos \omega_j k = 0 \quad (j \neq 0)$$

$$\sum_{k=0}^{N-1} \cos \omega_j k \cos \omega_m k = \begin{cases} N/2, & j = m \neq 0 \quad \text{or} \quad N/2 \\ N, & j = m = 0 \quad \text{or} N/2 \\ 0, & j \neq m \end{cases}$$

$$\sum_{k=0}^{N-1} \cos \omega_j k \sin \omega_m k = 0$$

$$\sum_{k=0}^{N-1} \sin \omega_j k \sin \omega_m k = \begin{cases} N/2, & j = m \neq 0 \quad \text{or} \quad N/2 \\ 0, & \text{otherwise} \end{cases}$$

Period Searching with Fourier Transform

- phase often less important than period
- period search often based on power of Fourier coefficients
- defined as a series of $N/2$ numbers P_j

$$P_j \equiv \frac{2}{a_o}|a_j|^2 = \frac{2}{N_{tot}}|a_j|^2 \qquad j = 0, 1, 2, \ldots, \frac{N}{2}$$

- series P_j is called the *power spectrum*
- does not contain information on phases
- normalization of power spectrum is convention
- Fourier coefficients a_j follow super-position theorem
- Fourier power spectrum coefficients P_j do not: a_j Fourier amplitude of x_k, b_j Fourier amplitude of y_k
- Fourier amplitude c_j of $z_k = x_k + y_k$ given by $c_j = a_j + b_j$
- power spectrum of z_k is $|c_j|^2 = |a_j + b_j|^2 \neq |a_j|^2 + |b_j|^2$
- difference being due to correlation term $a_j b_j$

Variance and Fourier Transform

- Only if x_k and y_k are not correlated, then the power of the combined signal may be approximated with the sum of the powers of the separate signals.
- variance expressed in terms of powers

$$\sum_{k=0}^{N-1}(x_k-\overline{x})^2=\frac{N_{tot}}{N}\left(\sum_{j=1}^{N/2-1}P_j+\frac{1}{2}P_{N/2}\right)$$

- In characterizing the variation of a signal one also uses the *fractional root-mean-square variation*,

$$r\equiv\frac{\sqrt{\frac{1}{N}\sum_k(x_k-\overline{x})^2}}{\overline{x}}=\sqrt{\frac{\sum_{j=1}^{N/2-1}P_j+0.5P_{N/2}}{N_{tot}}}$$

From continuous to discrete

- measurements x_k taken between $t = 0$ and $t = T$ at equidistant times t_k.
- describe as continuous time series $x(t)$ multiplied with window function

$$w(t) = \left\{ \begin{array}{ll} 1, & 0 \leq t < T \\ 0, & \text{otherwise} \end{array} \right\}$$

- and then multipled with sampling function ('Dirac comb')

$$s(t) = \sum_{k=-\infty}^{\infty} \delta(t - \frac{kT}{N})$$

From continuous to discrete (continued)

- $a(\nu)$ is continuous Fourier transform of $x(t)$
- $W(\nu)$ and $S(\nu)$ Fourier transforms of $w(t)$ and $s(t)$
- then

$$|W(\nu)|^2 \equiv \left| \int_{-\infty}^{\infty} w(t) e^{-i2\pi\nu t} dt \right|^2 = \left| \frac{sin(\pi\nu T)}{\pi\nu} \right|^2 = |T \mathrm{sinc}(\pi\nu T)|^2$$

- Fourier transform of a window function is (the absolute value of) a sinc-function, and

$$S(\nu) = \int_{-\infty}^{\infty} s(t) e^{-i2\pi\nu t} dt = \frac{N}{T} \sum_{m=-\infty}^{\infty} \delta\left(\nu - m\frac{N}{T}\right)$$

- Fourier Transform of Dirac comb is also a Dirac comb
- all these transforms are symmetric around $\nu = 0$ by definition

From continuous to discrete (continued)

- Fourier Transform of product is convolution of Fourier Transforms
- convolution of $a(\nu)$ and $b(\nu)$ is

$$a(\nu) * b(\nu) \equiv \int_{-\infty}^{\infty} a(\nu')b(\nu - \nu')d\nu'$$

- $x(t)w(t)$: window function $w(t)$ convolves each component with a sinc-function
- widening $d\nu$ inversely proportional to length of time series: $d\nu = 1/T$.
- $[x(t)w(t)]s(t)$: multiplication of signal by Dirac comb corresponds to convolution of its transform with Dirac comb, i.e. by an infinite repeat of the convolution.

From continuous to discrete (continued)

- from continuous $a(\nu)$ to discontinuous $a_d(\nu)$:

$$
\begin{aligned}
a_d(\nu) \equiv \quad & a(\nu) * W(\nu) * S(\nu) = \int_{-\infty}^{\infty} x(t) w(t) s(t) dt \\
= \quad & \int_{-\infty}^{\infty} x(t) \sum_{k=0}^{N-1} \delta\left(t - \frac{kT}{N}\right) e^{i2\pi\nu t} dt = \sum_{k=0}^{N-1} x\left(\frac{kT}{N}\right) e^{i2\pi\nu kT/N}
\end{aligned}
$$

- finite length of time series $\Rightarrow$ broadening of Fourier transform with width $d\nu = 1/T$ with sidelobes
- discreteness of sampling causes aliasing (reflection of periods beyond Nyquist frequency into range $0, \nu_{N/2}$)
- sample often integration over finite exposure time
- convolution of time series $x(t)$ with window function

$$
b(t) = \begin{cases} N/T, & -\frac{T}{2N} < t < \frac{T}{2N} \\ 0, & \text{otherwise} \end{cases}
$$

From continuous to discrete (continued)

- Fourier transform $a_d(\nu)$ is multiplied with Fourier transform of $b(t)$

$$B(\nu) = \frac{\sin \pi\nu T/N}{\pi\nu T/N}$$

- at frequency zero, $B(0) = 1$, at the Nyquist frequency $B(\nu_{N/2} = T/(2N)) = 2/\pi$, and at double the Nyquist frequency $B(\nu = N/T) = 0$.
- frequencies beyond Nyquist frequency are aliased into window $(0, \nu_{N/2})$ with reduced amplitude
- integration of the exposure time corresponds to an averaging over a time interval T/N, and this reduces the variations at frequencies near N/T

Power Spectra

- time series $x(t)$ consists of uncorrelated noise and signal

$$P_j = P_{j,\mathrm{noise}} + P_{j,\mathrm{signal}}$$

- power $P_{j,\mathrm{noise}}$ often approximately follows chi-squared distribution with 2 degrees of freedom
- normalization of powers ensures that power of Poissonian noise is exactly distributed as the chi-squares with two degrees of freedom
- probability of finding a power $P_{j,\mathrm{noise}}$ larger than an observed value P_j:

$$Q(P_j) = \mathtt{gammq}(0.5 * 2, 0.5P_j)$$

- standard deviation of noise power equal to their mean value: $\sigma_P = \overline{P_j} = 2$.
- fairly high values of P_j are possible du to chance

Power Spectra (continued)

- reduce noise of power spectrum by averaging:
- method 1: bin the power spectrum
- method 2: divide time series into M subseries and average their power spectra
- loss of frequency resolution in both cases
- but binned/averaged power spectrum is less noisy
- chi-squared distribution of power spectrum divided into M intervals, and in which W successive powers in each spectrum are averaged, is given by the chi-squared distribution with $2MW$ degrees of freedom, scaled by $1/(MW)$
- average of distribution is 2, variance $4/(MW)$
- probability that binned/averaged power $>$ observed power $P_{j,\mathrm{b}}$:

$$Q(P_{j,\mathrm{b}} = \mathtt{gammq}(0.5[2MW], 0.5[MWP_{j,\mathrm{b}}])$$

- for sufficiently large MW this approaches the Gauss function

Detecting and quantifying a signal

- can decide whether at given frequency observed signal exceeds noise level significantly, for any significance level
- 90% significance $\Rightarrow$ first compute P_j for which $Q = 0.1$
- in words: probability which is exceeded by chance in only 10% of the cases
- check whether the observed power is bigger than this P_j
- decided on frequency *before* we did the statistics, i.e. if we first select *one single frequency* ν_j
- good for known period, e.g. orbital period of a binary, or pulse period of pulsar
- in general: searching for a period, i.e. we do not know which frequency is important $\Rightarrow$ apply recipe many times, once for each frequency
- corresponds to many trials, and thus our probability level has te be set accordingly

Unknown Frequency and Amplitude

- consider one frequency, $P_{\rm detect}$ has probability $1-\epsilon'$ *not* to be due to chance
- try $N_{\rm trials}$ frequencies
- probability that the value $P_{\rm detect}$ is not due to chance *at any of these frequencies* is given by $(1-\epsilon')^{N_{\rm trials}}$, which for small ϵ' equals $1-N_{\rm trials}\epsilon'$.
- probability that value $P_{\rm detect}$ *is* due to chance *at any of these frequencies* is given by $\epsilon = N_{\rm trials}\epsilon'$.
- if we wish to set an overall chance of ϵ, we must take the chance per trial as $\epsilon' = \epsilon/N_{\rm trials}$, i.e.

$$\epsilon' = \frac{\epsilon}{N_{\rm trials}} = \mathtt{gammq}(0.5[2MW], 0.5[MWP_{\rm detect}])$$

Upper Limit

- observed power $P_{j,\mathrm{b}}$ higher than detection power P_{detect} for given chance ϵ'
- observed power is sum of noise power and signal power

$$P_{j,\mathrm{signal}} > P_{j,\mathrm{b}} - P_{j,\mathrm{noise}} \qquad (1-\epsilon') \quad \text{confidence}$$

- no observed power exceeds detection level $\Rightarrow$ upper limit
- determine level P_{exceed}, exceeded by noise alone with high probability $(1-\delta)$, from

$$1-\delta = \mathtt{gammq}(0.5[2MW], 0.5[MWP_{\mathrm{exceed}}])$$

- highest observed power $P_{\mathrm{max}} \Rightarrow$ upper limit P_{UL} to power is

$$P_{\mathrm{UL}} = P_{\mathrm{max}} - P_{\mathrm{exceed}}$$

- if there were signal power higher than P_{UL}, the highest observed power would be higher than P_{max} with a $(1-\delta)$ probability