Vorlesung "Astronomische Datenanalyse"
Laboratory 3: Time Series Analysis
Version of July 29, 2004

1 Time Series Analysis

Even if the sky appears to be static to many humans, time varying phenomena in astronomy are rather the rule than the exception. These phenomena can be observed on almost all astronomical objects, such as in variable stars, in radio pulsars, in accreting neutron stars in X-rays, or in the lightcurves of active galactic nuclei.

The aim of this laboratory is to work on the most used tool for the characterization of lightcurves, the periodogram. All data needed can be found at `http://astro.uni-tuebingen.de/~wilms/teach/data/prakt3/`.

1.1 Definitions

Time series analysis is often performed using techniques from Fourier analysis. Here we will limit ourselves to the analysis of an equidistant time series, i.e., of a time series where the temporal information x_i was measured at times $t_i = t_0 + i\Delta t$. Here i is a natural number and t_0 is the zero point of the time series. In Fourier analysis it is shown that a time series can be expressed under fairly wide conditions as

$$x_k = \frac{1}{N}\sum_{k=0}^{N-1} X_j \mathrm{e}^{-2\pi i jk/N} \tag{1}$$

where

$$X_j = \sum_{k=0}^{N-1} x_k \mathrm{e}^{2\pi i jk/N} \tag{2}$$

The amplitudes X_j are called the (discrete) Fourier transform of x_i. Die frequency at which X_j is computed,

$$\omega_j = 2\pi\nu_j = 2\pi j/(N\Delta t) \tag{3}$$

is the Fourier frequency. The highest frequency, for which X_j can be computed, $\nu_{N/2} = 1/(2\Delta t)$, is called the Nyquist frequency. The term belonging to frequency 0 is proportional to the mean of the time series. Often this term is set explicitly to zero by subtracting the mean from the time series before Fourier transforming.

In the practical analysis of time series it is often not necessary, to use all of the information present in the trignonometrical sum (1). For many applications it is sufficient to concentrate on the amplitude of the series at a certain frequency, which is given by

$$P_j = A|X_j|^2 \quad \text{wo } j \in [0\ldots N/2] \tag{4}$$

These values are called the periodogram. Other common names are PSD for *power spectral density* or PDS for *power density spectrum*. Since astronomical measurements are real numbers, $X_j = X^*_{-j}$ and therefore $|X_j|^2 = |X_{-j}|^2$, such that the PSD only has to be defined for positive frequencies (the star *, denotes complex conjugation).

Unfortunately, the normalization constant A of the periodogram is not standardized. In the following we will be using the so-called Leahy normalization, where

$$A = \frac{2\Delta t}{X_0} = \frac{2\Delta t^2}{N_{\text{ph}}} \tag{5}$$

In the second expression, N_{ph} is the total number of detected photons, and one has to compute the PSD from the measured count rate. The Leahy normalization is used in many areas of astronomy since the value of the periodogram from a time series, where the variability is completely dominated by Poisson noise, equals 2. Because of this property, the PSD in Leahy normalization can be easily used to determine deviations from the Poisson statistics, i.e., the PSD in Leahy norm can be used to separate the variability inherent to the source from the statistical noise introduced by the measurement.

1.2 Computation of the periodogram

In IDL it is possible to compute the periodogram in several normalizations using the routine `PSD` from `aitlib`. The calling sequence of `PSD` is

```
psd,time,rate,freq,psd,/leahy
```

where `time` is the time and `rate` die measured count rate. The arrays `freq` and `psd` return the Fourier frequency and the value of the periodogram. The keyword `/leahy` tells `PSD` to use the Leahy normalization. Apart from these arguments, `PSD` has several other keywords which we will not consider further.

To study the properties of the PSD we will first generate simple time series with IDL and study their periodograms. In a second step we will then analyze the time series from two typical astronomical objects.

1.3 Simple periodograms

The simplest case of a time series is a purely sinusoidal variation. This can be studied in IDL with the program `zeitsinus`:

```
;;
;; Periodogramm einer sinusfoermigen Zeitreihe
;;

tmin=0  ;; Anfangszeit
tmax=100.;; Endzeit
f=2.    ;; Frequenz
r0=1    ;; Mittlere Rate
ampl=2. ;; Amplitude

npt=600 ;; Zahl der Punkte in der Zeitreihe
time=tmin+(tmax-tmin)*findgen(npt)/(npt-1)
rate=r0+ampl*sin(2.*!pi*f*time)
psd,time,rate,fre,psd,/leahy
plot,fre,psd

END
```

Change `npt` to smaller and larger values and observe how the Fourier peak is changed. Modify `zeitsinus` and compute the periodograms of more complicated time series:

Multiple sinusoids:

$$x(t) = r_0 + A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t)$$

what changes if you vary A_1 und A_2? What frequencies f_1 und f_2 can you still separate with the PSD?

Time series with exponentially decreasing amplitude:

$$x(t) = r_0 + A_1 \exp(-0.69\, t/t_{1/2}) \sin(2\pi f_1 t)$$

Study the PSD with different half-life times $t_{1/2}$ and vary the ratio between f_1 and $t_{1/2}$.

Such periodic functions, however, are rare in astronomy. In astronomy the measurement is often severely disturbed by noise. This effect can be simulated by adding normally distributed random numbers to the simulated light curve:

```
seed=13244L ;; variieren!
rate=rate+a1*randomu(seed,n_elements(rate),/normal)
```

where `a1` is the standard deviation of the noise. Repeat the above experiments, including such a noise term. Up to what amplitude is it possible to see the period in the simulated data? Can you still see the variation by eye?

1.4 Astronomical Periodograms

1.4.1 4U 0115+62

In spring 1999 the Be-X-ray binary 4U 0115+62 had one of its massive outbursts. In such systems a neutronstar moves on an highly elliptic orbit around a strongly rotating B-star. The Rotation of the B-Star is so fast that the rotational speed at the equator reaches the escape velocity. A gas disk is formed around the star, which is observable in the optical spectrum through the emission of lines (designated by the "e" in the name of the system). When the neutron star smashes through the disk it can accrete material and thus cause an outburst in the X-rays.

In collaboration with colleagues from San Diego we have observed the outburst in 1999 with the Rossi X-ray Timing explorer. Further information can be found at `http://astro.uni-tuebingen.de/publications/ppC_99.shtml`. You should search for this paper on the Astrophysical Data System (`http://adsabs.harvard.edu/default_service.html`) and print it in its published form. In the paper you will find further hints as to what you are supposed to do in the following...

The file `src-0.125s-2nd-long.lc` contains the lightcurve of 4U 0115+62 as measured on 12 March 1999. This lightcurve has a resolution of 0.125 s and was measured during two of RXTE's revolutions around Earth.

You read the file using the IDL-routine `readlc`

```
readlc,time,counts,'src-0.125s-2nd-long.lc',/mjd,/counts
```

where `time` is measured in *days*.

First convert the time into seconds since the start of the measurement:

```
time=(time-time[0])*86400.
```

and plot it. The light curve has a very high resolution. Use the xrange-parameter of the IDL plot command to look at part of the lightcurve. Good observing times are pieces of 1000 s duration or even pieces that are much shorter. What temporal phenomena can you see?

The observed data contain a gap, caused by the time where the satellite was behind the Earth. In the following we will only look at the longer time segment measured after the gap. For this you need to cut out part of the data. In order to do so, determine the duration of the individual time bins:

```
dt=shift(time,-1)-time
```

The gap is where dt is significantly larger than the bin time of the time series. Determine the position of the gap and cut it out. Since the gap is the longes interval, it is easiest to find it using IDL's max-function:

```
dummy=max(dt,ndx)
ndx=ndx+1
time=time[ndx:n_elements(time)-1]
counts=counts[ndx:n_elements(counts)-1]
dt=dt[ndx:n_elements(dt)-1]
```

finally, convert the number of photons into a count rate:

```
rate=counts/dt
```

Compute the periodogram of the time series. It is recommended to plot such diagrams in a log-log way, using the /xlog and /ylog keywords of the plot command.

A disadvantage of the periodogram is that it covers a very large frequency range. It can be shown that the statistical uncertainty of the PSD is of the magnitude of the PSD values. Due to this property it is often not very useful to look at individual periodograms. Rather, one computes several periodograms from pieces of the lightcurve and then averages these periodograms. With this approach the statistical uncertainty of the periodogram is reduced, even if the frequency range is reduced as well. The length of such segments (in units of dt) is defined with the keyword dseg of psd. Compute PSDs for different values of dseg. Due to computational time restrictions, it is recommended to choose dseg as powers of two, since the Fast Fourier Transform algorithm is much faster for these values ($\mathcal{O}(N \log N)$, compared to $\mathcal{O}(N^2)$ for random lengths).

One of the frequencies which are apparent in the PSD is the rotational frequency of the neutron star. Determine the frequency from the periodogram and compute the pulse profile. For the latter you will have to read the description of the pfold routine of aitlib.

1.5 The Sun

A very nice example for a quasi-periodic phenomenon is the variation of the sunspot number of the sun. Such data are contained in the file spot_num.txt, which you should analyze as well using the techniques described above.

1.6 A variable star

As a final example, look at the light curve in the file HLTau_all and analyze it using the periodogram.