Time Series Analysis

Eric Feigelson
Penn State University

Outline

Time series in astronomy

- Periodic phenomena: binary orbits (stars, extrasolar planets); stellar rotation (radio pulsars); pulsation (helioseismology, Cepheids)
- Stochastic phenomena: accretion (CVs, X-ray binaries, Seyfert gals, quasars); scintillation (interplanetary & interstellar media); jet variations (blazars)
- Explosive phenomena: thermonuclear (novae, X-ray bursts), magnetic reconnection (solar/stellar flares), star death (supernovae, gamma-ray bursts)

Difficulties in astronomical time series

Gapped data streams:

Diurnal & monthly cycles; satellite orbital cycles; telescope allocations

Heteroscedastic measurement errors:

Signal-to-noise ratio differs from point to point

Poisson processes:

Individual photon/particle events in high-energy astronomy

Variety of temporal behaviors

Concepts of time series analysis

Stationarity The temporal behavior, whether deterministic (e.g. orbit) or stochastic, is statistically unchanged by shifts in time. Types of nonstationarity include: *trends* (secular changes in mean value), *heteroscedasticity* (changes in variance), and *change points* (different behaviors before and after t_0). GRS 1915 is very nonstationary.

Periodicity The measured levels repeat themselves deterministically with one or more periods. The signal becomes concentrated in frequency domain study: *spectral analysis, harmonic analysis, Fourier analysis*. These methods classically use trigonometric sine and cosine functions, but this is not required.

Autocorrelation The measured levels at time t_0 depend on levels measured at previous times. The autocorrelation can be deterministic (trend or periodicity) or can include a stochastic component. For an evenly spaced time series, the *autocorrelation function (ACF)* is the fraction of the total variance due to correlated values at lag k time steps:

$$\hat{\rho}(k) = ACF(k) = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}.$$

If the ACF has significant signal at small k, the time series has *short-term memory*. If the signal extends to large k, it has *long-term memory*. The latter includes *red noise* or $1/f^\alpha$-type processes that are often seen in (astro)physical systems. A stochastic time series with insignificant ACF values at all k exhibits *white noise*, often assumed to have a Gaussian (normal) distribution.

Other important concepts *Nonlinear* (in the parameters) time series (including *chaotic* systems); *multivariate* time series (including autoregressive with lags); *time-frequency* analysis (for nonstationary periodic behaviors); *wavelet* analysis (for multiscale aperiodic variations); *state space* models (hierarchical deterministic + stochastic models with MLE coefficients updated by the Kalman filter); *unevenly-spaced* time series (methods primarily developed by astronomers).

Nonparametric time domain methods

Autocorrelation function

This sample ACF is an estimator of the correlation between the x_t and x_{t-k} in an evenly-spaced time series. For zero mean and normal errors, the ACF is asymptotically normal with variance $Var\hat{\rho} = [n-k]/[n(n+2)]$. This allow probability statements to be made about the ACF.

The partial autocorrelation function (PACF) estimates the correlation with the linear effect of the intermediate observations, $x_{t-1}, ..., x_{t-k+1}$, removed. Calculate with the Durbin-Levinson algorithm based on an autoregressive model.

Density estimation

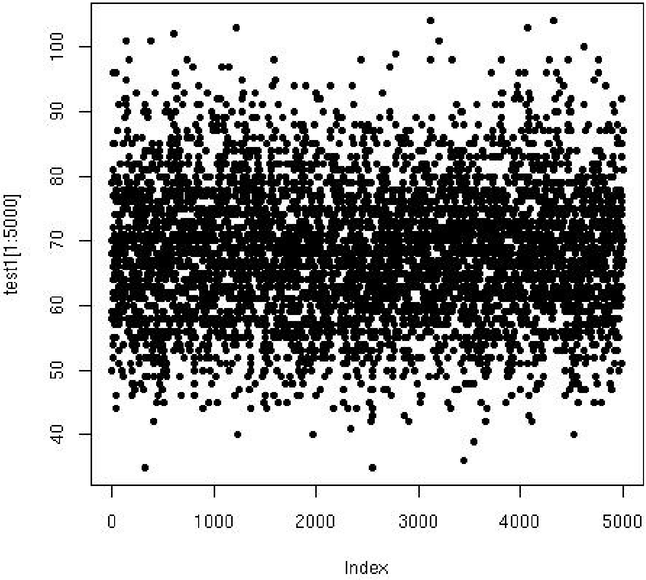
Standard methods of density estimation are often used on time series: kernel density estimation, local regressions, etc.

Ginga observations of X-ray binary GX 5-1

GX 5-1 is a binary star system with gas from a normal companion accreting onto a neutron star. Highly variable X-rays are produced in the inner accretion disk. X-ray binary time series often show 'red noise' and 'quasi-periodic oscillations', probably from inhomogeneities in the disk. We plot below the first 5000 of 65,536 count rates from Ginga satellite observations during the 1980s.

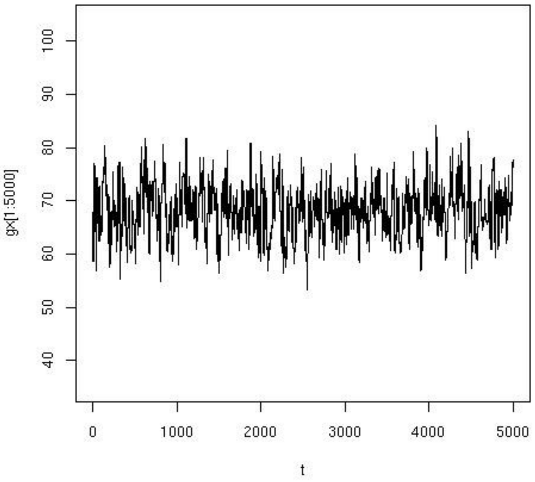
R script:

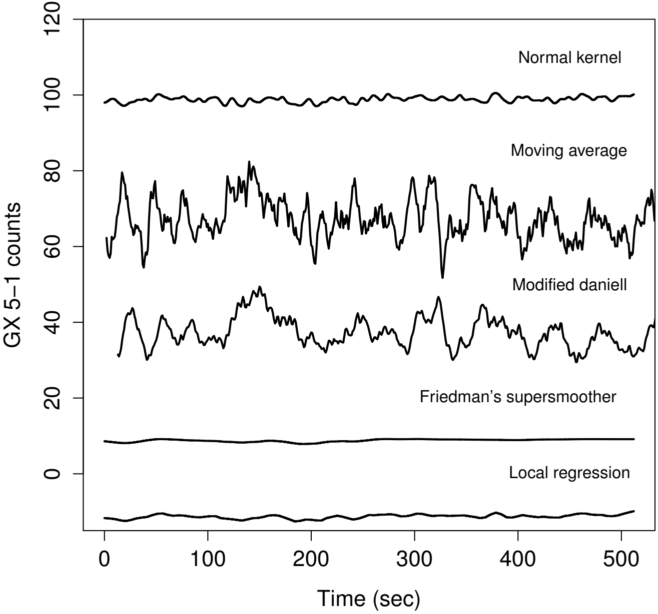
```
gx=scan("GX.dat")
t=1:5000
plot(t,gx[1:5000],pch=20)
```



100
90
80
70
60
50
40
test1[1:5000]
0
1000
2000
3000
4000
5000
Index

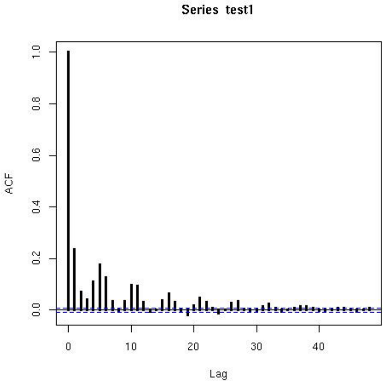
Kernel smoothing of GX 5+1 time series
Normal kernel, bandwidth = 7 bins

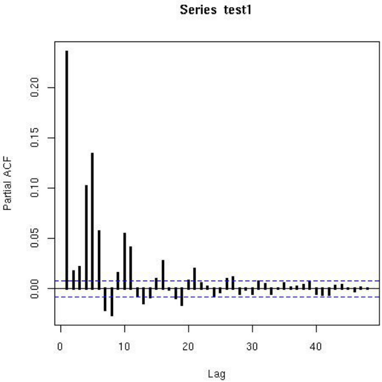




Normal kernel
Moving average
Modified daniell
Friedman's supersmoother
Local regression
GX 5–1 counts
Time (sec)
120
100
80
60
40
20
0
0
100
200
300
400
500

Autocorrelation functions





acf(GX, lwd=3)

pacf(GX, lwd=3)

Parametric time domain methods: ARMA models

Autoregressive moving average model

Very common model in human and engineering sciences, designed for aperiodic autocorrelated time series (e.g. 1/f-type ‘red noise’). Easily fit by maximum-likelihood. Disadvantage: parameter values are difficult to interpret physically.

$$\textbf{AR(p) model} \quad x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + w_t$$

$$\textbf{MA(q) model} \quad x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

The AR model is recursive with memory of past values. The MA model is the moving average across a window of size $q+1$. ARMA(p,q) combines these two characteristics.

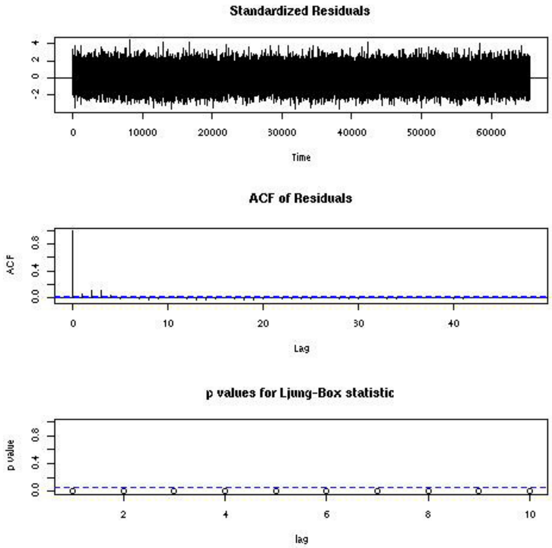
Many extensions to ARMA models:

- VAR (vector autoregressive)
- ARFIMA (ARIMA with long-memory component)
- GARCH (generalized autoregressive conditional heteroscedastic for stochastic volatility)
- Dozens of variants from econometrics: see ftp://ftp.econ.au.dk/creates/rp/08/rp08_49.pdf.

GX 5+1 modeling

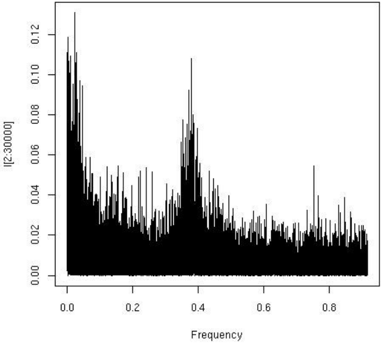
ar(x = GX, method = "mle")
Coefficients:
1 2 3 4 5 6 7 8
0.21 0.01 0.00 0.07 0.11 0.05 -0.02 -0.03

arima(x = GX, order = c(6, 2, 2))
Coefficients:
ar1 ar2 ar3 ar4 ar5 ar6 ma1 ma2
0.12 -0.13 -0.13 0.01 0.09 0.03 -1.93 0.93
Coeff s.e. = 0.004 $\sigma^2 = 102$ log L = -244446.5
AIC = 488911.1



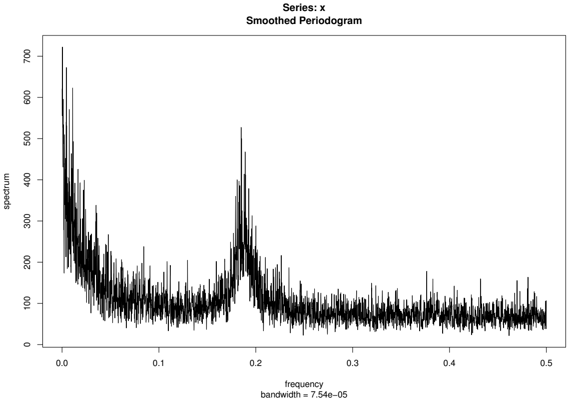
Although the scatter is reduced by a factor of 30, the chosen model is not adequate: Ljung-Box test shows significant correlation in the residuals. Use AIC for model selection.

Fast Fourier Transform of the GX 5-1 time series reveals the 'red noise' (high spectral amplitude at small frequencies), the QPO (broadened spectral peak around 0.35), and white noise.




```
f = 0:32768/65536
I = (4/65536)*abs(fft(gx)/sqrt(65536))^ 2
plot(f[2:60000],I[2:60000],type="l",xlab="Frequency")
```

Smoothed and tapered Fourier spectrum




```
postscript(file="~/Desktop/GX_sm_tap_fft.eps")
k = kernel("modified.daniell", c(7,7))
spec = spectrum(gx, k, method="pgram", taper=0.3, fast=TRUE, detrend=TRUE, log="no")
dev.off()
```

Spectral analysis

For challenging problems, smoothing, multitapering, linear filtering, (repeated) pre-whitening and Lomb-Scargle can be used together. Beware that aperiodic but autoregressive processes produce peaks in the spectral densities. Harmonic analysis is a complicated 'art' rather than a straightforward 'procedure'.

It is extremely difficult to derive the significance of a weak periodicity from harmonic analysis. Do not believe analytical estimates (e.g. exponential probability), as they rarely apply to real data. It is essential to make simulations, typically permuting or bootstrapping the data keeping the observing times fixed. Simulations of the final model with the observation times is also advised.

State space models

Often we cannot directly detect x_t, the system variable, but rather indirectly with an observed variable y_t. This commonly occurs in astronomy where y is observed with measurement error (errors-in-variable or EIV model). For AR(1) and errors $v_t = N(\mu, \sigma)$ and $w_t = N(\nu, \tau)$,

$$y_t = Ax_t + v_t \qquad x_t = \phi_1 x_{t-1} + w_t$$

This is a *state space model* where the goal is to estimate x_t from y_t, $p(x_t|y_t, \ldots, y_1)$. Parameters are estimated by maximum likelihood, Bayesian estimation, Kalman filtering, or prediction. Extended state space models: non-stationarity, hidden Markov chains, etc. MCMC evaluation of nonlinear and non-normal (e.g. Poisson) models

Statistical texts and monographs

D. Brillinger, Time Series: Data Analysis and Theory, 2001
C. Chatfield, The Analysis of Time Series: An Introduction, 6th ed., 2003
G. Kitagawa & W. Gersch, Smoothness Priors Analysis of Time Series, 1996
J. F. C. Kingman, Poisson Processes, 1993
J. K. Lindsey, Statistical Analysis of Stochastic Processes in Time, 2004
S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed, 1999
M. B. Priestley, Spectral Analysis and Time Series, 2 vol, 1981
R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications (with R examples), 2nd Ed., 2006

Astronomical references

Bretthorst 2003, "Frequency estimation and generalized Lomb-Scargle periodograms", in Statistical Challenges in Modern Astronomy

Collura et al 1987, "Variability analysis in low count rate sources", ApJ 315, 340

Dworetsky 1983, "A period-finding method for sparse randomly spaced observations ....", MNRAS 203, 917

Gregory & Loredo 1992, "A new method for the detection of a periodic signal of unknown shape and period", ApJ 398, 146

Kovacs et al. 2002, "A box-fitting algorithm in the search for periodic transits", A&A 391, 369

Leahy et al. 1983, "On searches for periodic pulsed emission: The Rayleigh test compared to epoch folding",. ApJ 272, 256

Roberts et al. 1987, "Time series analysis with CLEAN ...", AJ 93, 968

Scargle 1982, "Studies in astronomical time series, II. Statistical aspects of spectral analysis of unevenly spaced data", ApJ 263, 835

Scargle 1998, "Studies in astronomical time series, V. Bayesian Blocks, a new method to analyze structure in photon counting data, ApJ 504, 405

Stellingwerf 1972, "Period determination using phase dispersion measure", ApJ 224, 953

Vio et al. 2005, "Time series analysis in astronomy: Limits and potentialities, A&A 435, 773