Time Series Analysis

Eric Feigelson

Penn State University

http://astrostatistics.psu.edu

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

Poisson processes:

Individual photon/particle events in high-energy astronomy

Time domain methods: Nonparametric

Autocorrelation function

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad \text{where} \quad \hat{\gamma}(h) = \frac{\sum_{i=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x})}{n}$$

This sample ACF is an estimator of the correlation between the x_t and x_{t-h} in an evenly-spaced time series lags.

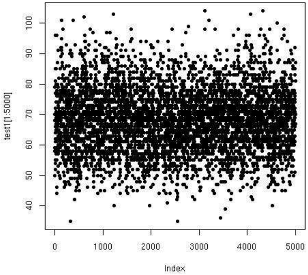
The partial autocorrelation function (PACF) estimates the correlation with the linear effect of the intermediate observations, $x_{t-1}, ..., x_{t-h+1}$, removed. Calculate with the Durbin-Levinson algorithm based on an autoregressive model.

Note that the error on the mean, or any other parameter, of an autocorrelated time series is different from the usual value:

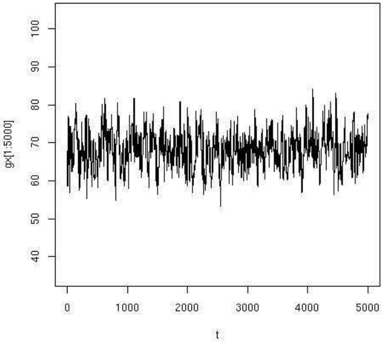
$$Var(\bar{x}) = \frac{\sigma^2}{N}[1 + \sum_{i=1}^{N-1}(1 - \frac{i}{N})\rho(i)]$$

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. XRB time series often show 'red noise' and 'quasi-periodic oscillations' (QPOs) from inhomogeneities in the disk and/or beating between the neutron star rotation and disk orbits. We plot below the first 5000 of 65,536 count rates from Ginga satellite observations. Superficially, it looks like white noise.

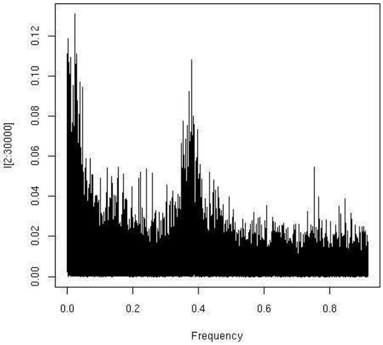


Nonparametric estimation: Kernel smoothing
Normal kernel, bandwidth = 7 bins



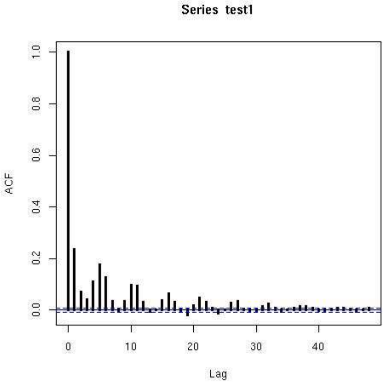
Smoothing is not very useful in this case, but it does reveal some correlated behavior.

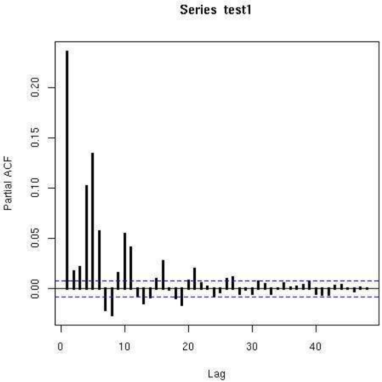
Fourier transform (FFT)



The Fourier power spectrum reveals three components: power at low frequency (1/f-type 'red noise'), the QPO around freq=0.4, and white noise. Most astronomical studies of XRB time series and the QPO phenomenon are based on FFT analysis.

Autocorrelation functions





acf(GX, lwd=3) pacf(GX, lwd=3)

The ACF and PACF provide quantitative measure of the short-term correlation, and show the periodic behavior.

Time domain models: ARMA models

Autoregressive moving average model

Very common model in human and engineering sciences, designed for stationary, Gaussian processes. Easily fit by maximum-likelihood. Disadvantage: parameter values are difficult to interpret physically.

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

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

(Note: model diverges if any $|\theta_i| > 1$)

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. ARIMA (I = integrated) models some types of non-stationary behaviors.

Time domain models: 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 $w_t = N(\nu, \tau)$,

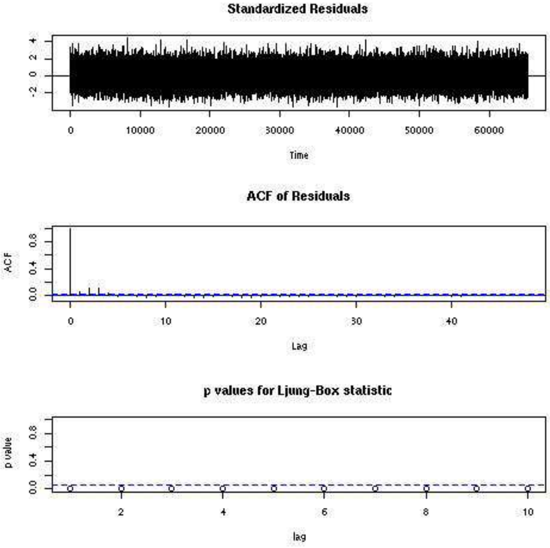
$$y_t = Ax_t + v_t \qquad x_t = \phi_1 x_{t-1} + w_t$$

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 can be fit by maximum likelihood methods, or a Bayesian framework by assuming priors for the parameters. The likelihood function is easily calculated and updated via Kalman filtering.

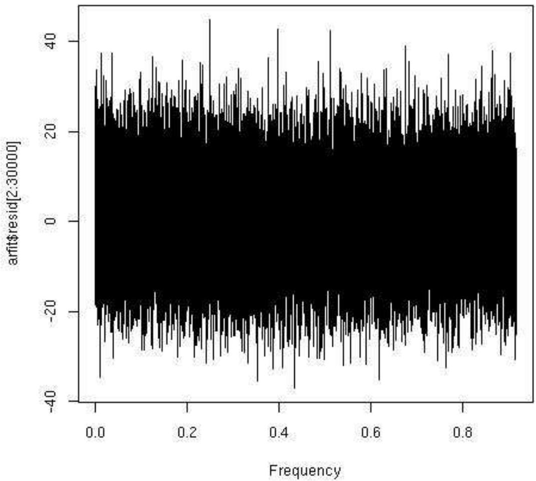
GX 5+1 autoregressive modeling

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 (use AIC for model selection)



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



Nonetheless, the FFT power spectrum of the ARIMA residuals shows that most of the red noise and QSO structure is removed by the model.

Other time domain models

- Extended ARMA models: VAR (vector autoregressive), SARIMA (S = seasonal for periodic behavior), ARFIMA (F = fractional for long-memory behavior), GARCH (generalized autoregressive conditional heteroscedastic for stochastic volatility)
- Extended state space models: non-stationarity, hidden Markov chains, etc. MCMC evaluation of nonlinear and non-normal (e.g. Poisson) models

References

C. Chatfield, The Analysis of Time Series: An Introduction, 6th Ed., 2003

R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications (with R examples), 2nd Ed., 2006

G. Kitagawa & W. Gersch, Smoothness Priors Analysis of Time Series, 1996

J. K. Lindsey, Statistical Analysis of Stochastic Processes in Time, 2004

S. M. Ross, Stochastic Processes, 2nd ed, 1996