Mathematics & Statistics | Lancaster University

Introduction to Time Series Data and Analysis

Simon Taylor

Department of Mathematics and Statistics

20 April 2016

Contents

What is Time Series Data?

A **time series** is a set of observations made sequentially through time.

Examples:

- Changes in execution time, RAM or bandwidth usage.
- Times a software has run in consecutive periods of time.
- Financial, geophysical, marketing, demographic, etc.

The objectives in time series analysis are:

Description	How does the data vary over time?
Explanation	What causes the observed variation?
Prediction	What are the future values of the series?
Control	Aim to improve control over the process.

Common Questions

Q: How important is preserving data order?

A: Very! Changing data order breaks the dependence between measurements.

Q: How frequent do I need to take measurements?

A: It depends:

- Too sparse, risk missing the dependence structure.
- Too frequent, swamped with noise.

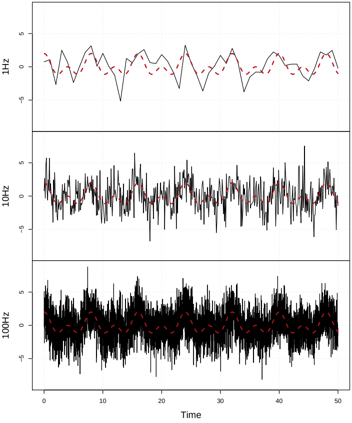

Figure: Sampling Frequency

Why is time series important in benchmarking?

Q: Can I use simple summary statistics?

A: You can, but they only describe overall properties.

Q: Can't I just interpolate between data points?

A: Signals are often subject to uncontrollable random noise. Error from interpolation may be large if noise is large.

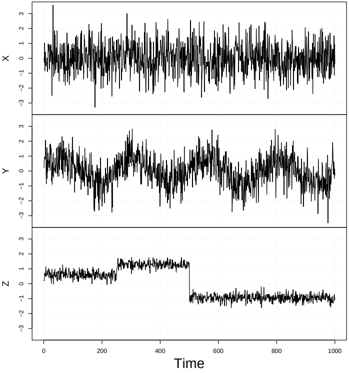

Figure: Three times series with $\bar{x} = 0$ and $s^2 = 1$.

Analysis Tools – Trace Plot

A **trace plot** is a graph of the measurements against time.

Easy to visually identify key features:

- Trends – Long-term trend in the mean level.
- Seasonality – Regular peaks & falls in the measurements.
- Outliers – Unusual measurements that are inconsistent with the rest of the data.
- Discontinuities – Abrupt change to the underlying process.

Analysis Tools – Auto-correlation function

Correlation measures the linear dependence between two data sets.

Auto-correlation measures the correlation between all data pairs at lag *k* apart.

$$r_k = \frac{\sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{(T-1)s^2},$$

where $\bar{x}$ and s^2 is the sample mean and variance.

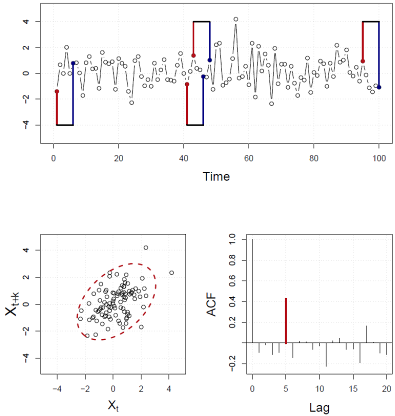

Figure: Lag 5 ACF calculation.

Analysis Tools – Spectrum

The **spectrum** describes how the power in a time series varies across frequencies.

$$I(\omega) = \frac{1}{\pi T} \left| \sum_{t=1}^{T} x_t e^{i2\pi t\omega} \right|^2,$$

for $\omega \in (0, 1/2]$.

Identifies prominent seasonal and cyclic variation.

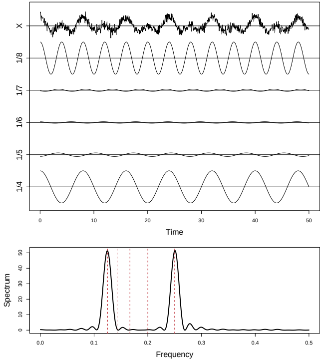

Figure: Fourier decomposition and spectrum of time series X_t.

Time series models

Let $X_{1:T} = \{X_1, \ldots, X_T\}$ denote a sequence of T measurements.

A time series is **stationary** if the distribution of any pair of subset separated by lag k, $X_{1:t}$ and $X_{1+k,t+k}$, are the same.

A time series is **weakly stationary** if the first two moments are constant over time:

$$\mathbb{E}[X_t] = \mu \quad \text{and} \quad \mathrm{Cov}(X_t, X_{t+k}) = \gamma(k).$$

Gaussian White Noise Process, GWNP

The time series $\{Z_t\}$ follows a Gaussian white noise process if:

$$Z_t \sim \mathcal{N}(0, \sigma^2), \quad t = 1, \ldots, T$$

Gaussian White Noise Process

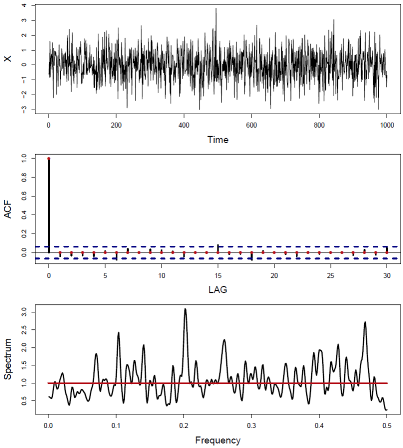

Figure: Gaussian white noise process.

MA(q) process

Moving Average Process of Order q, MA(q)

The process $\{X_t\}$ is a moving average process of order q if:

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

where $\{Z_t\}$ is a GWNP and $\beta_0, \ldots, \beta_q$ are constants ($\beta_0 = 1$).

Expectation: $\mathbb{E}[X_t] = 0$.

Auto-covariance:

$$\mathrm{Cov}(X_t, X_{t+k}) = \begin{cases} \sigma^2 \sum_{i=0}^{q-|k|} \beta_i \beta_{i+|k|}, & |k| = 0, \ldots, q; \\ 0, & \text{otherwise.} \end{cases}$$

MA(q) process

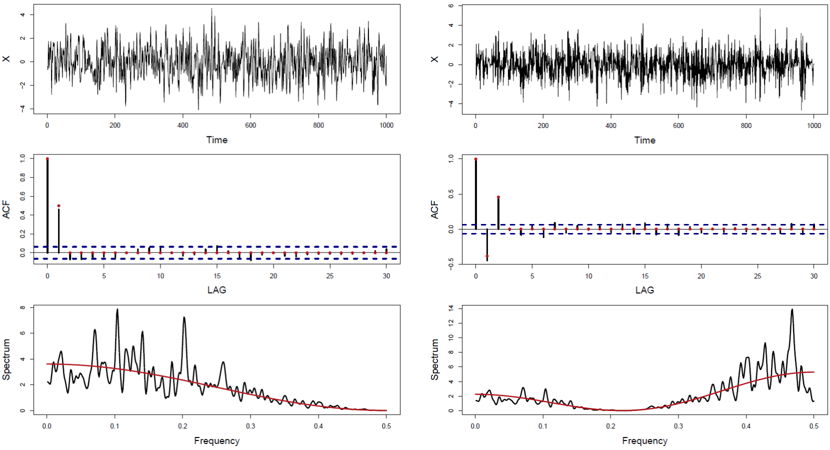

Figure: Left: MA(1), $\beta_1 = 0.9$. Right: MA(2), $(\beta_1, \beta_2) = (-0.4, 0.9)$.

AR(p) process

Autoregressive Process of Order p, AR(p)

The process $\{Y_t\}$ is an autoregressive process of order p if:

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + Z_t$$

where $\{Z_t\}$ is a GWNP and $\alpha_1, \ldots, \alpha_p$ are constants.

Expectation: $\mathbb{E}[X_t] = 0$.

Auto-covariance for AR(1):

$$\mathrm{Cov}(X_t, X_{t+k}) = \sigma^2 \frac{\alpha_1^{|k|}}{1 - \alpha_1^2}, \quad \text{provided } |\alpha_1| < 1.$$

AR(p) process

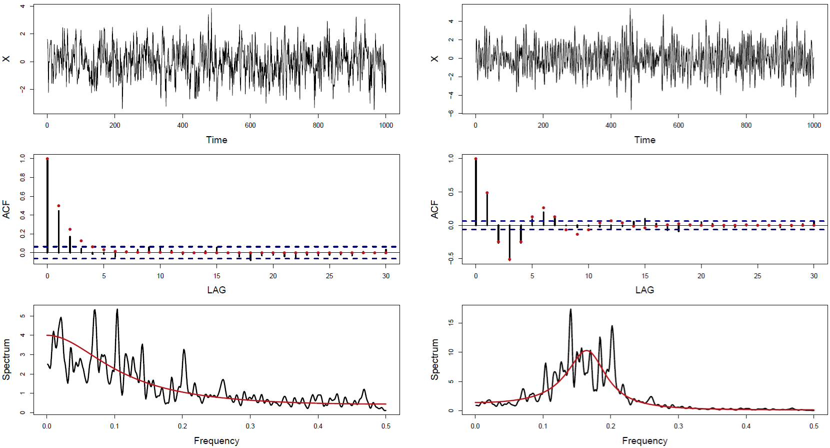

Figure: Left: AR(1), $\alpha_1 = 0.9$. Right: AR(2), $(\alpha_1, \alpha_2) = (0.8, -0.64)$.

Non-stationary process

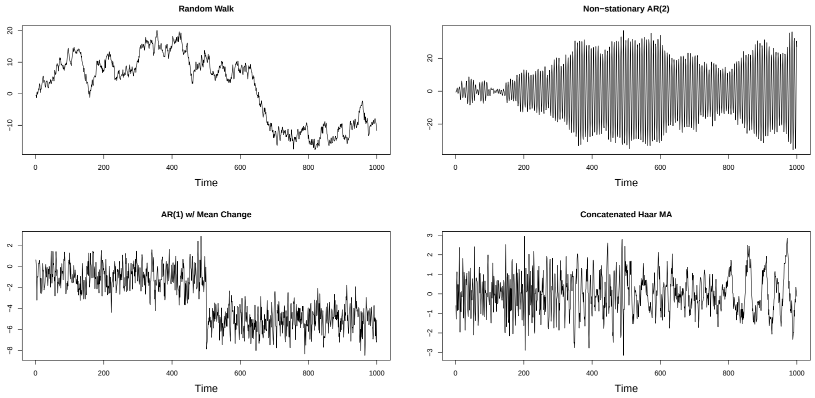

Figure: Examples of non-stationary processes.

On-going Research in Time Series

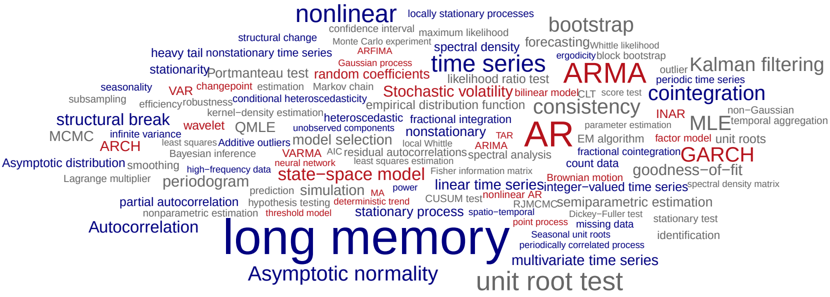

Figure: Keyword cloud from the Journal of Time Series Analysis, 2002–2015. Red: Models, Navy: Properties, Grey: Inference & Methods

Further Reading

- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008) **Time series analysis: Forecasting and control**. 4th ed., John Wiley & Sons.
- Chatfield, C. (2004) **The Analysis of Time Series: An Introduction**. 6th ed., CRC press.
- Signal processing toolbox, MATLAB® (`http://uk.mathworks.com/products/signal/`)
- Statsmodels, python (`http://statsmodels.sourceforge.net/`)

Simon Taylor

Department of Mathematics and Statistics

20 April 2016