Time series

Christiana Kartsonaki

January 20, 2017

Introduction

Time series data gathered sequentially in time

different types:

- one or a few long series
- a large number of short series

focusing on the first

Introduction

Time series data gathered sequentially in time

different types:

- one or a few long series
- a large number of short series

focusing on the first

usually at equally spaced intervals

dependence between nearby time points
trend
seasonality

Applications

- economics
- engineering
- environmental statistics
- physics, including meteorology
- medical statistics
- ...

Objectives of time series analysis

- description (plots, summaries)
- estimation (model to describe the data structure, interpretation)
- prediction/forecasting (given a series of observations, predict future values)
- adjustment for time dependence (time dependence as a nuisance)

Example: Sleep EEG

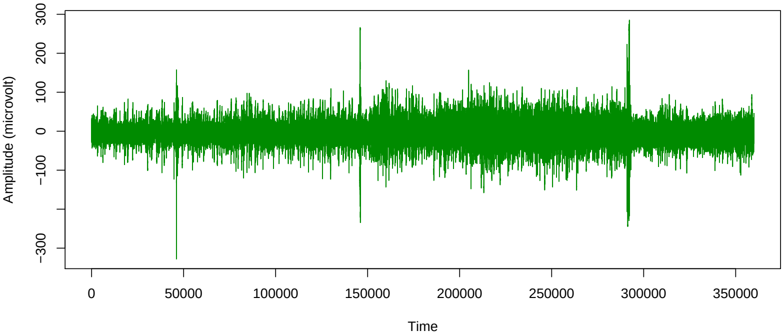


Figure: A 30-minute segment of a recording from a electroencephalogram (EEG) channel from polysomnography recording of a 40 year old man.

University of Mons, Belgium, TCTS Laboratory (Stphanie Devuyst, Thierry Dutoit) and Universit Libre de Bruxelles CHU de Charleroi

Sleep Laboratory (Myriam Kerkhofs)

Example: fMRI

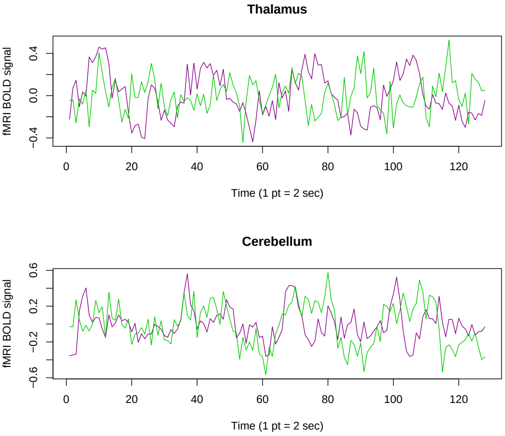


Figure: Multiple time series of fMRI BOLD signals at different brain locations, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its application*, Third edition, Springer.

Other types of series

many features of time series apply to other types of series, e.g. spatial

however time has a unique before $\rightarrow$ after direction whereas space does not, which affects some aspects of interpretation

Example: Soil surface temperatures

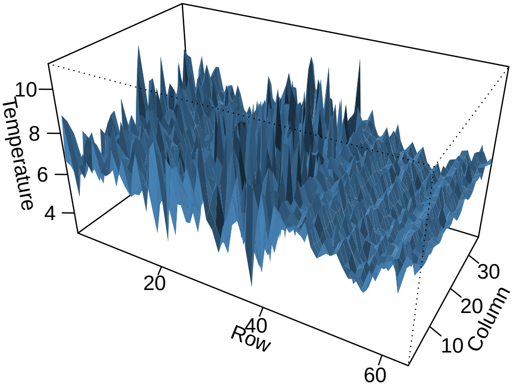


Figure: Two-dimensional spatial series of temperature measurements taken on a rectangular field (64×36 with 17-foot spacing). Data from Bazza et al. (1988).

Shumway, R. H. and Stoffer, D. S. (2011). *Time series analysis and its application*, Third edition, Springer.

Stationarity

A process $\{X_t\}$ is **weakly stationary** or **second-order stationary** if for all integers t, τ,

$$\begin{aligned} \mathbb{E}(X_t) &= \mu \\ \mathrm{cov}(X_t, X_{t+s}) &= \gamma_s \end{aligned}$$

where μ is constant and γ_s does not depend on t.

Stationarity

A process $\{X_t\}$ is **weakly stationary** or **second-order stationary** if for all integers t, τ,

$$\begin{aligned} \mathbb{E}(X_t) &= \mu \\ \text{cov}(X_t, X_{t+s}) &= \gamma_s \end{aligned}$$

where μ is constant and γ_s does not depend on t.

It is **strictly stationary** (or **strongly stationary**) if

$(X_{t_1}, \ldots, X_{t_k})$ and $(X_{t_1+s}, \ldots, X_{t_k+s})$

have the same distribution for all sets of time points $t_1, \ldots, t_k$ and all integers s.

Autocovariance and autocorrelation

For a stationary time series:

Autocovariance function $\gamma_s = \text{cov}(X_t, X_{t+s})$

Autocovariance and autocorrelation

For a stationary time series:

Autocovariance function $\gamma_s = \mathrm{cov}(X_t, X_{t+s})$

Autocorrelation function (ACF) $\rho_s = \mathrm{cor}(X_t, X_{t+s}) = \frac{\gamma_s}{\gamma_0}$

$\mathrm{var}(X_t) = \gamma_0$

ACF – estimation

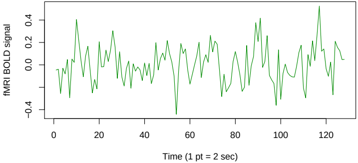


Figure: Time series of a fMRI BOLD signal from one location in the thalamus, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

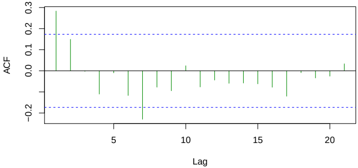


Figure: ACF of a fMRI BOLD signal from one location in the thalamus, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

Partial autocorrelation function (PACF)

Partial autocorrelation at lag s: the correlation between X_t and X_{t+s} after regression on $X_{t+1}, \ldots, X_{t+s-1}$

estimated using **Levison–Durbin recursion**

PACF – estimation

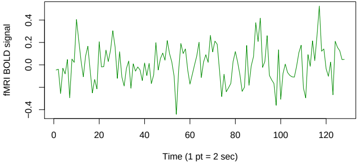


Figure: Time series of a fMRI BOLD signal from one location in the thalamus, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

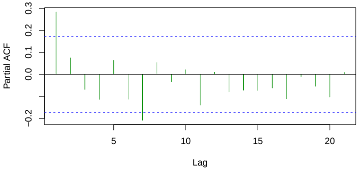


Figure: PACF of a fMRI BOLD signal from one location in the thalamus, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

Example: white noise

white noise: uncorrelated variables w_t with mean 0 and variance σ_w^2

$w_t \sim wn(0, \sigma_w^2)$

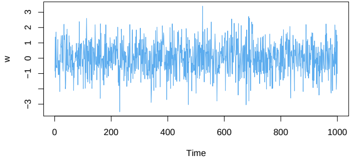


Figure: White noise.

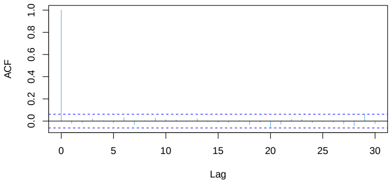


Figure: ACF of white noise.

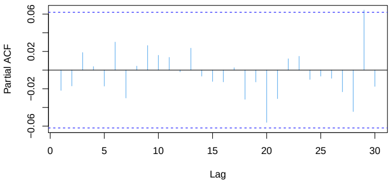


Figure: PACF of white noise.

Cross-covariance and cross-correlation

For stationary time series:

Cross-covariance function between two series x, y:

$$\gamma_{xy}(t, t+s) = \text{cov}(x_t, y_{t+s})$$

Cross-covariance and cross-correlation

For stationary time series:

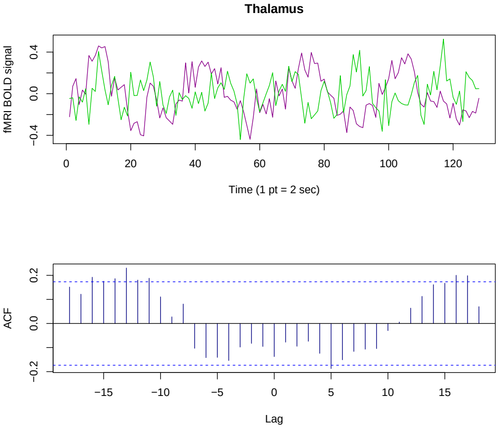
Cross-covariance function between two series x, y:

$$\gamma_{xy}(t, t+s) = \mathrm{cov}(x_t, y_{t+s})$$

Cross-correlation function

$$\rho_{xy}(t, t+s) = \frac{\gamma_{xy}(t, t+s)}{\sqrt{\gamma_x(0)\gamma_y(0)}}$$

Example: Cross-correlation



cross-correlation of X and Y at lag $s \to$ not the same as lag $-s$

Figure: Cross-correlation of two time series of fMRI BOLD signals at different brain locations, when a stimulus was applied for 32 seconds and then stopped for 32 seconds.

Linear process

$\{X_t\}$ is a **linear process** if it has a representation of the form

$$X_t = \mu + \sum_{r=-\infty}^{\infty} c_r \epsilon_{t-r}$$

where

μ: common mean
$\{c_r\}$: sequence of fixed constants
$\{\epsilon_t\}$: independent random variables with mean zero and common variance

We assume that $\sum c_r^2 < \infty$ to ensure that the variance of X_t is finite.

Autoregressive processes

- the current value of the series depends linearly on its previous values with some error

AR(1) (autoregressive process of order (lag) 1)

$$X_t = \phi X_{t-1} + \epsilon_t$$

ϵ_t: **white noise** (a series of uncorrelated random variables with mean 0 and variance σ^2)

AR(p) (autoregressive process of order (lag) p)

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t$$

AR(1) stationary if $|\phi| < 1$.

Example: AR(1)

$$x_t = 0.9x_{t-1} + \epsilon_t$$

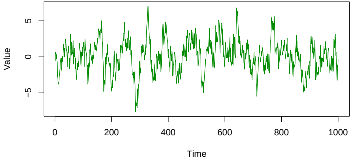


Figure: AR(1).

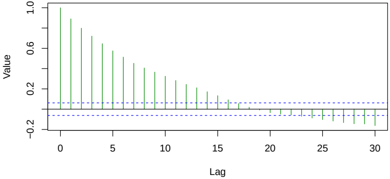


Figure: ACF of AR(1).

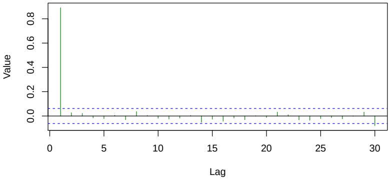


Figure: PACF of AR(1).

Moving average processes

- the current value of the series is a weighted sum of past white noise terms

MA(1) (moving average process of order (lag) 1)

$$X_t = \epsilon_t + \theta\epsilon_{t-1}$$

ϵ_t: **white noise** (a series of uncorrelated random variables with mean 0 and variance σ^2)

MA(q) (moving average process of order (lag) q)

$$X_t = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

MA(q) stationary for any θ

Example: MA(1)

$$x_t = \epsilon_t + 0.9\epsilon_{t-1}$$

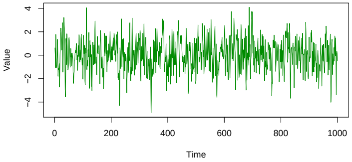


Figure: MA(1).

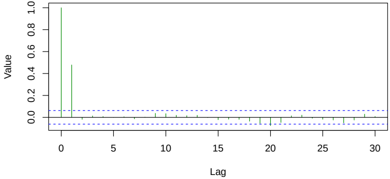


Figure: ACF of MA(1).

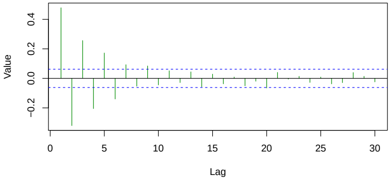


Figure: PACF of MA(1).

ARMA processes

ARMA(*p*, *q*) (autoregressive moving average process)

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=0}^{q} \theta_j \epsilon_{t-j}$$

AR(p) and ARMA(p, q) models $\rightarrow$ not necessarily stationary

autoregressive polynomial: $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$, defined for any complex number z

stationarity condition for an AR(p) process:
all the zeros of the function $\phi(z)$ lie outside the unit circle in the complex plane

The backshift operator

Backshift operator B

$$BX_t = X_{t-1}, \quad B^2X_t = B(BX_t) = X_{t-2}, \quad \ldots$$

identity operator: $IX_t = B^0X_t = X_t$

ARMA(p, q) in terms of the backshift operator:

$$\phi(B)X = \theta(B)\epsilon$$

$\phi(\cdot), \theta(\cdot)$: generating functions of the autoregressive and moving average operators

Differencing

Difference operator ∇

$$\nabla X_t = X_t - X_{t-1}, \quad \nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2}, \ldots$$

If a series is not stationary we can look at the differenced series and look for an ARMA model for the differenced series.

ARIMA processes

ARIMA(p, d, q) (autoregressive integrated moving average process) if its dth difference $\nabla^d X$ is an ARMA(p, q) process

$$\phi(B)\nabla(B)^d X = \theta(B)\epsilon$$

$\phi(\cdot), \theta(\cdot)$: generating functions of the autoregressive and moving average operators

Fitting ARIMA models – The Box–Jenkins approach

Focused on prediction/forecasting.

- identification
- estimation
- verification

Iterated until a suitable model is identified.

Box–Jenkins approach – identification

Identification initial processing of the data to make the series stationary, preliminary identification of suitable orders p and q for the ARMA components of the model

Box–Jenkins approach – identification

Identification initial processing of the data to make the series stationary, preliminary identification of suitable orders p and q for the ARMA components of the model

	AR(p)	**MA(q)**	**ARMA(p, q)**
ACF	decay	cuts off after lag q	decay
PACF	cuts off after lag p	decay	decay

Table: ACF and PACF for ARMA models.

Box–Jenkins approach – estimation

Estimation of model parameters

for AR processes, solving the **Yule–Walker equations**

Box–Jenkins approach – verification

Verification

check for overfitting

check that residuals are consistent with white noise

Seasonal ARIMA (SARIMA) models

Seasonal ARIMA

special case of ARIMA with seasonal component with period M

extension $\rightarrow$ periodically correlated processes

Time series regression – short-term associations of exposures

investigate whether some of the short-term variation in the outcome is associated with variation in an exposure

associations with a few lags

Time series regression – short-term associations of exposures

investigate whether some of the short-term variation in the outcome is associated with variation in an exposure

associations with a few lags

- Time-stratified model
- Periodic functions (Fourier terms)
- Flexible spline functions

confounding by other time-varying factors
short-term displacement, or 'harvesting'

model checking and sensitivity analyses

(Bhaskaran et al., 2013)

Interrupted time-series

Assessing the effect of an 'interruption' on some outcome measured in time.

e.g. effect of public health intervention on disease counts

- Using ARIMA modelling.
- Using a segmented linear regression, adjusting the standard errors for autocorrelation (e.g. Newey–West standard errors).

(Lagarde, 2011)

More...

- Analysis in the frequency domain.
  Spectrum, periodogram, smoothing, filters.
- State space models.
  Linear models, Kalman filters.
- Nonlinear models.
  ARCH and stochastic volatility models, chaos.
- Multivariate time series.

References

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L. and Armstrong, B. (2013). Time series regression studies in environmental epidemiology. *International Journal of Epidemiology*, **42**, 1187–1195.

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, Second Edition. Springer.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments (with discussion). *Scandinavian Journal of Statistics*, **8**, 93–115.

Cox, D. R., Hinkley, D. V. and Barndorff-Nielsen, O. E. (Eds) (1996). *Time Series Models: In econometrics, finance and other fields*. CRC Press.

References – continued

Diggle, P. (1990). *Time Series: A Biostatistical Introduction*. Clarendon Press.

Lagarde, M. (2012). How to do (or not to do)... Assessing the impact of a policy change with routine longitudinal data. *Health Policy and Planning*, **27** (1), 76–83.

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

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.