Ch1. Introduction

1. Time series; examples and objectives
2. Time series models
3. Stationary process
4. Preliminary analysis: classical decomposition

Time series (TS) - Definition

- Time series = a set of observations made sequentially in time
- Notation: $\{x_t,\ t \in T_0\}$, t is a **time index** and T_0 is a **index set** (the set of all possible time points)
- If $T_0 \in \mathbb{Z}$, $\{x_t\}$ is said to be a discrete TS.
- If $T_0 \in \mathbb{R}$, $\{x_t\}$ is said to be a continuous TS.
- Purpose of TS analysis: want to understand underlying physical dynamics and predict (forecasting) future values.
- TS analysis is the area of statistics which deals with the analysis of dependency between different observations in time.
- Why model dependency? If we ignore the dependencies that we observe in time series data, then we can be led to incorrect statistical inferences.

TS examples

Almost everywhere in our real life applications.

- Economics: unemployment rate, annual inflation rate, average salary, GDP, GNP etc. Important in economic policy making
- Finance: stock price, return, volatility, exchange rate.
- Demography: planning and control of population, tax collection, military service
- Sales/Marketing: Forecasting future sales
- Environmental statistics: global warming, climate changes, air pollution, rain fall.
- Hydrology: water level of a lake, dam control to prevent flooding and drought.
- Physics: sunspots, electro-magnetic field, star light
- Engineering: internet traffic, signal denoising
- Medical science: ECG (Electro-Cardiography)

Time series models

- A time series model is a probabilistic model that describes the different ways that the series data $\{x_t\}$ could have been generated.
- Not just a model for one observation at one time point. Rather a model for the entire set of observations in time. That is, find the joint distribution of observations.
- More formally, a time series model is usually a probability model for $\{X_t, t \in T_0\}$, a collection of random variables (RVs) indexed in time (this is the population).
- Also, for the forecasting purpose, we want to include future values.

Time series models

- Recall that the joint distribution of $X = (X_1, \ldots, X_n)'$ is given by

$$F_X(x_1, \ldots x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n)$$

- However in the TS analysis, because of future prediction, we need to estimate the joint distribution of infinite dimension such as

$$(X_1, X_2, \ldots, X_n, X_{n+1}, \ldots)'$$

- The joint distribution of infinite dimension is called the finite dimensional distribution (FDD) defined as the finite joint distribution for any finite selection of random variables. (e.g) X_1, X_1, X_3, X_3, X_6, X_7 etc.
- FDD is a comprehensive modeling of TS, but it is too **complex**. It is comprehensive but of no use in practice since we cannot estimate them from a finite sample.

Time series models

- Instead assume simple structure to the population.
- Plausible assumption is called the stationarity. The underlying system do not change a lot over time in the sense that
  - Graphs over two equal-length TS exhibit similar feature.
  - If you shift a time series by k time points, that characteristic of the distribution will not change.
  - Stationarity means that some characteristic of the distribution of a time series does not depend on the time points, only on the distance between time points.

Strict stationarity

Definition

$\{X_t, t \in \mathbb{Z}\}$ *is strictly stationary if for all* n *and* h*,*

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$$

- If $n = 1$, it means that $X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} X_3 \ldots$.
- If $n = 2$, then

$$(X_1, X_2) \stackrel{d}{=} (X_2, X_3) \stackrel{d}{=} (X_5, X_6) \stackrel{d}{=} \ldots$$

$$(X_1, X_3) \stackrel{d}{=} (X_2, X_4) \stackrel{d}{=} (X_3, X_5) \stackrel{d}{=} \ldots$$

- It is much simpler than fdd, but still hard. What about assuming parametric family of distribution such as Gaussianity?

Strict stationarity

If we further assume Gaussianity on strict stationarity

$$(X_{t_1}, \ldots, X_{t_n}) \sim MVN(\mu, \Sigma),$$

then we only need to find the mean vector μ and covariance matrix Σ to estimate underlying distribution.

- From $X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} X_3 \ldots$

$$EX_1 = EX_2 = \ldots$$

  the mean is constant over time

- Also, $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) \ldots$. Furthermore, from $(X_1, X_2) \stackrel{d}{=} (X_2, X_3) \stackrel{d}{=} (X_5, X_6) \stackrel{d}{=} \ldots$

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_3) = \ldots$$

  the covariance only depends on time difference $|t_1 - t_2|$.

Strict stationarity with Gaussianity

- Therefore, with Gaussian assumption, strictly stationarity becomes

$$
\begin{aligned}
&i) EX_t = m, \quad \forall t \in \mathbb{Z} \\
&ii) \operatorname{Cov}(X_r, X_s) = \operatorname{Cov}(X_{r+h}, X_{s+h}), \quad \forall r, s, h \in \mathbb{Z}
\end{aligned}
$$

- Is Gaussianity really needed? Normal distribution provides good approximation to bell-shaped curve. Central limit theorem holds without Gaussianity etc. We can broaden the scope of model by relaxing Gaussian assumption.
- It leads to weakly stationarity.

Weakly stationary TS

Definition (Weakly stationarity)

The TS $\{X_t, t \in \mathbb{Z}\}$ *is said to be weakly stationary iff*

$$
\begin{aligned}
&i)\ E|X_t|^2 < \infty \quad \forall t \in \mathbb{Z} \\
&ii)\ EX_t = m, \quad \forall t \in \mathbb{Z} \\
&iii)\ \gamma_X(r, s) = \gamma_X(r + h, s + h), \quad \forall r, s, h \in \mathbb{Z} \\
&(or, iii)'\ Cov(X_t, X_{t+h}) \text{ does not depend on } t)
\end{aligned}
$$

- The first condition guarantees the existence of the 1st and 2nd moments from Cauchy-Schwartz ineqality.
- If $\{X_t\}$ is strictly stationary then it is also weakly stationary.
- Converse is not true. But if a weakly stationary TS $\{X_t\}$ is Gaussian then it is strictly stationary.
- It is also called covariance stationary, 2nd order stationary, stationarity in wide sense. If otherwise specified in the book, it refers to weakly stationarity.

ACVF/ACF

Since iii) implies that the covariance function does not depend on t but only a function of lag (time difference) h, we define ACVF/ACF as

- $\{X_t\}$ be a stationary TS. The autocovariance function (ACVF) of $\{X_t\}$ at lag h is

$$\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h})$$

- The autocorrelation (ACF) of $\{X_t\}$ at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \mathrm{Corr}(X_t, X_{t_h})$$

Thus, TS analysis with weakly stationary assumption means that we only need to estimate

$$\boxed{EX_t, \quad \gamma(h)}$$

Nonstationary TS

If $\{X_t\}$ is not stationary, then it is called nonstationary TS. Major sources of nonstationarity comes from

- Non-constant mean: May have some trend (linear, quadratic). Mean could shifts (abruptly changes). Some regular trend may repeat over time (this is called seasonality)
- Non-constant varaince: Variance may shifts or increasing/decreasing (heteroscedascity)
- Time dependent covariance: Covariance structure may depend on time

Most real time series are not stationary, but don't worry! We will remove or model the non-stationary parts (the components that depend on the time index), so that we are only left with a stationary component.

Decompositon of TS

Our general strategy is to decompose X_t by non-stationary parts and stationary part.

$$\boxed{X_t = m_t + s_t + Y_t}$$

m_t : trend

s_t : seasonality with period d in the sense that $s_t = s_{t+d}$

Y_t : weakly stationary errors

- Thus, before estimating mean and covariance of Y_t, we will first model/remove trend and seasonality. Three major methods are
  1. Regression
  2. Smoothing (local regression)
  3. Differencing

Exploratory TS analysis - time plot

- Start with a time series plot of x_t versus t.
  - The axes of a time series plot should be carefully labeled.
  - Also, think about the time scale. For example, in examining monthly sales figures over twenty years, consider making year, not month number, the time variable.
- Look for the following:
  1. Are there any trend? (e.g., linear, quadratic or exponentially increasing trend?)
  2. Are there abrupt changes in behavior? (e.g., are there shifts in mean and/or variance?)
  3. Are there outliers? (unusual values with respect to the rest of the data).

Time plot

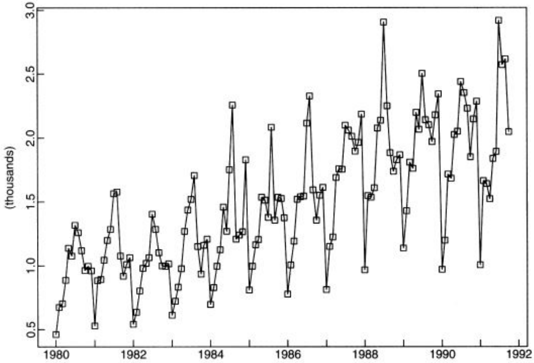


Figure 1-1
The Australian red wine sales, Jan. '80 – Oct. '91.

Features:

Time plot

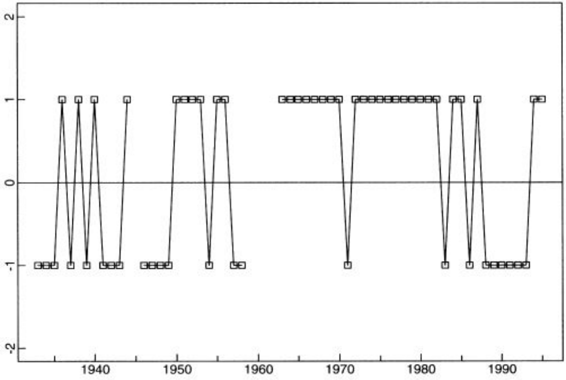


Figure 1-2
Results of the all-star baseball games, 1933–1995.

Features:

Time plot

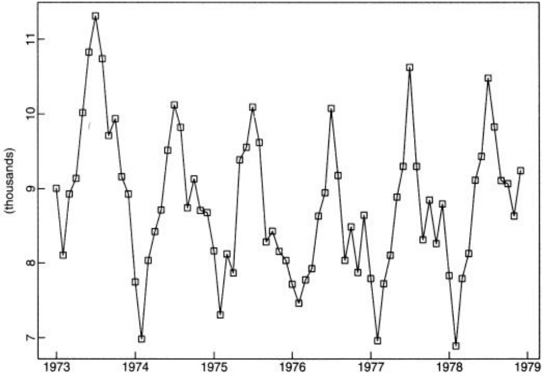


Figure 1-3
The monthly accidental deaths data, 1973–1978.

Features:

Time plot

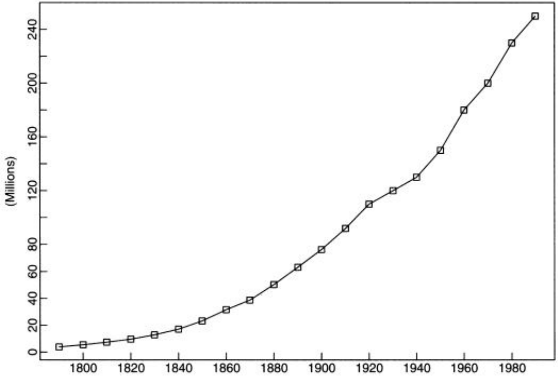


Figure 1-5
Population of the U.S.A. at ten-year intervals, 1790–1990.

Features:

Time plot

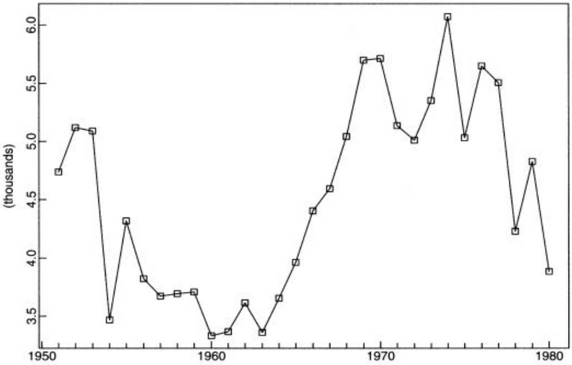


Figure 1-6
Strikes in the U.S.A., 1951–1980.

Features:

Ch1.5 Estimation and Elimination of Trend and Seasonal components

Keep in mind that three major tools are

1. Regression
2. Local regression/Moving average/smoothing
3. Differencing

Estimating trend only

First consider model without seasonality

$$X_t = m_t + Y_t, \quad E(Y_t) = 0$$

- Method 1: Polynomial Regression

  $$m_t = c_0 + c_1 t + \ldots + c_p t^p$$

  Parameters are estimated by OLS

  $$(\hat{c}_0, \ldots, \hat{c}_p) = \underset{\mathbf{c}}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - m_t)^2 \, .$$

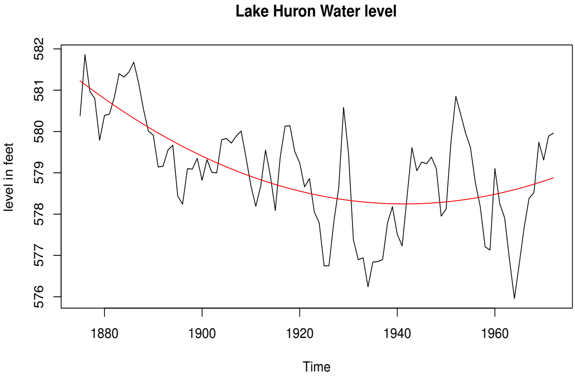
  In a vector notation,

  $$\hat{\boldsymbol{c}} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{X},$$

  where

  $$\mathbf{T} = \begin{pmatrix} 1 & 1^1 & \ldots & 1^p \\ 1 & 2^1 & \ldots & 2^p \\ \vdots & \vdots & \ldots & \vdots \\ 1 & n^1 & \ldots & n^p \end{pmatrix}$$

Polynomial regression - OLS




```
x = seq(from=1875, to = 1972, by=1);
x2 = x^2;
out.lm = lm(data ~ 1 + x + x2);
plot.ts(data, ylab="level in feet");
title("Lake Huron Water level")
lines(x,out.lm$fitted.values, col="red")
```

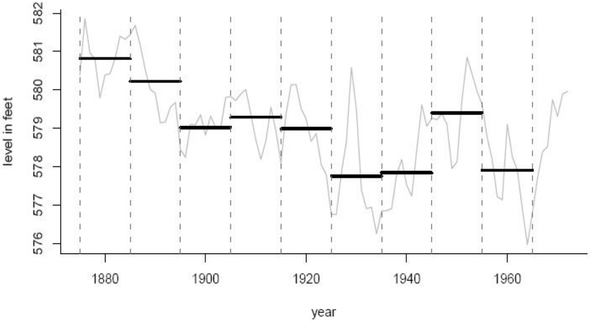
Polynomial regression - OLS

Very important remark! The OLS estimator has the following problems:

- Since $\{Y_t\}$ is typically not an IID process, the statistical properties of the LSEs will be different from the standard results cited in Ch0.
- If $\{Y_t\}$ is a well-behaved mean zero stationary process, the LSEs are unbiased estimates of $(c_0, c_1, \ldots, c_p)$
- However, the variance of the LSEs calculated assuming IID errors will be wrong.
- Hence, $\hat{\boldsymbol{c}}$ is an unbiased estimator, but not suitable for providing confidence interval if $\{Y_t\}$ are correlated.
- We will revisit this later in Ch6.6.

Estimating trend only - Smoothing

- Method2: Smoothing
  Suppose we break the time series up into small blocks and average each block. For the Lake Huron series we average every ten years of data:

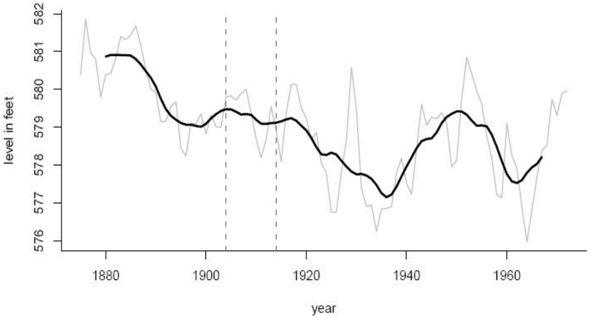


This is a very rough estimate of the trend. The key idea is a local averaging. We will generalize this idea as follows.

Smoothing 1 - Moving Average filter

Consider the simple local averaging

$$\boxed{W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j}}$$



Smoothing 1 - Moving Average filter

Then, observe

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^{q} (m_{t+j} + Y_{t+j})$$

$$= \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t+j}.$$

If the true trend m_t is linear, that is $m_t = c_0 + c_1 t$, then

$$\frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} = c_0 + c_1 t = m_t, \quad t \in [q+1, n-q]$$

$$\frac{1}{2q+1} \sum_{j=-q}^{q} Y_{t+j} \approx E(Y_t) = 0.$$

Thus, it preserves linear trend and filters noise.

Smoothing 1 - Moving Average filter

We can further write this smoothing operation as

$$\boxed{\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}}$$

- $\{a_j\}$ determines filter. It is a weighted average.
- For example, MA filter is given by $a_j = 1/(2q+1)$, $-q \leq j \leq q$ and 0 elsewhere.
- MA filter is low-pass filter since it filters out high frequency variation.
- There are lots of other choices of filter.

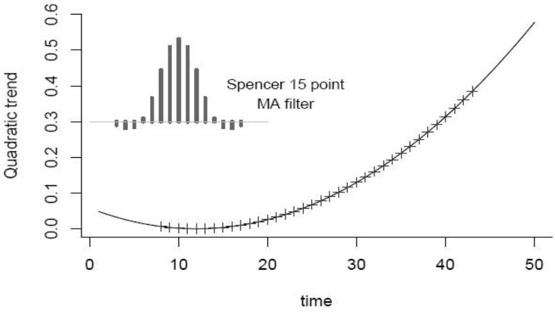
Smoothing 1 - Spencer's 15 point MA filter

Define filter as

$$a_i = a_{-i} \quad |j| \leq 7$$

$$[a_0, a_1, \ldots, a_7] = \frac{1}{320}(74, 67, 46, 21, 3, -5, -6, -3)$$

Then, it preserves cubic trend and filter out noise.



Smoothing 2 - Exponential smoothing

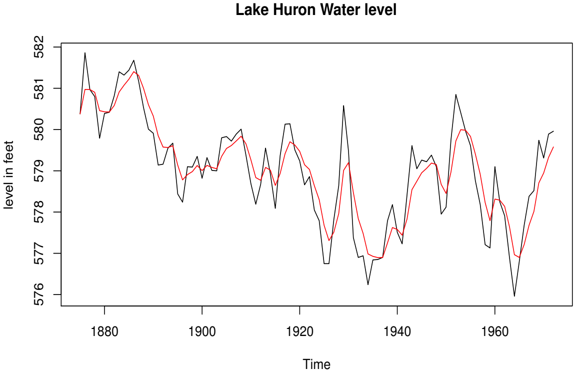
Consider the filter only depends on the past data. That is, $\hat{m}_t$ is estimated by using observations up to time t. This is called on-line/real time smoother. For $a \in [0, 1]$,

$$
\begin{aligned}
\hat{m}_1 &= X_1 \\
\hat{m}_2 &= aX_2 + (1-a)\hat{m}_1 = aX_2 + (1-a)X_1 \\
\hat{m}_3 &= aX_3 + (1-a)\hat{m}_2 = \\
&\vdots \\
\hat{m}_t &= aX_t + (1-a)\hat{m}_{t-1} = \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^{t-1} X_1
\end{aligned}
$$

Hence, weights are exponentially decaying except the last one.

Smoothing 2 - Exponential smoothing

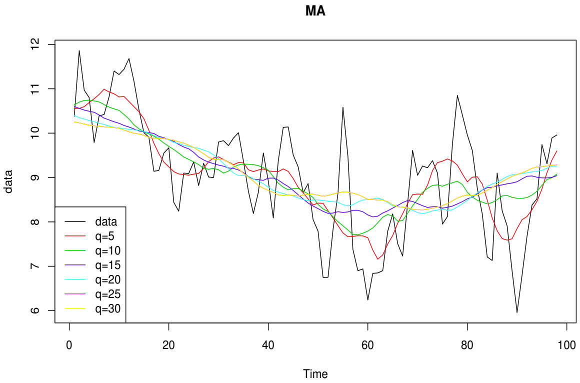
Exponential smoothing with $a = .4$




```
library(itsmr)
ex4 = smooth.exp(data, .4)
plot.ts(data, ylab="level in feet");
title("Lake Huron Water level")
lines(x,ex4, col="red")
```

Weakness of Smoothing - Bandwidth selection

Smoothing method is very appealing, but it has serious disadvantage - a tuning parameter selection. Filter length q in MA or weight constant a in Exponential smoothing plays central role in smoothing.



Which one is the best?

Smoothing - Bandwidth selection

It is explained by the so-called bias-variance trade off. For MA(q), note that

$$W_t = \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j}}_{\approx m_t \text{ if } q \text{ is small}} + \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}}_{\approx 0 \text{ if } q \text{ is large}}$$

- smaller q: reduce bias, but increase variance
- larger q: high bias, but smaller variance

General ways of selecting bandwidth are cross-validation (CV)

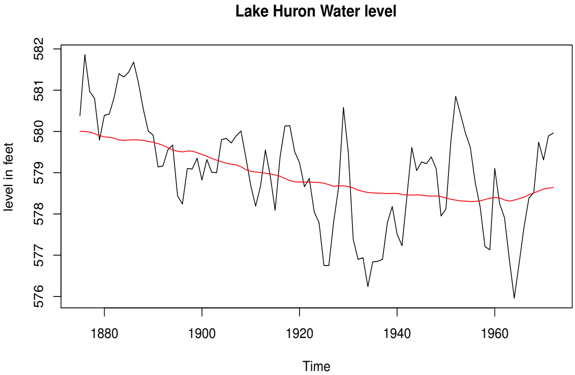
- Cross-validation to estimate MSE

$$\hat{q} = \operatorname*{argmin}_q \sum_{t=1}^{n} (X_t - \hat{m}_t^{-(t)})^2,$$

where $\hat{m}_t^{-(t)}$ is an estimate of $\hat{m}_t$ without using t-th observation.

Smoothing - MA Bandwidth selection

If we apply MA filter with CV bandwidth selection, the optimal $q = 33$ and it gives the following result.




```
library(itsmr)
smooth.ma(data, q=33)
```

Estimating trend only - Differencing

Definition (Backshift operator)

$$BX_t = X_{t-1}$$

Definition (Lag-1 Differencing)

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t$$

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) =$$

Thus, if $m_t = c_0 + c_1 t$, then

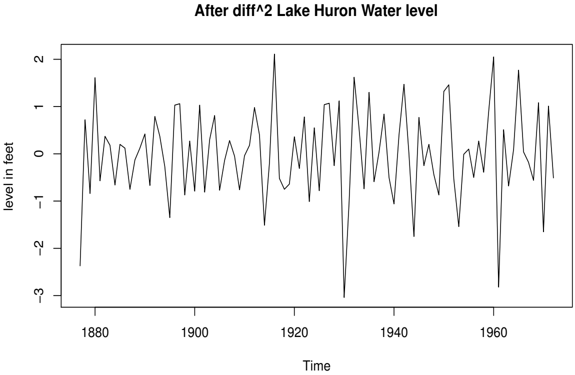
$$\nabla m_t = (c_0 + c_1 t) - (c_0 + c_1(t-1)) = c_1$$

In general, if you apply k-th differencing, then it kiils k-th order polynomial trend. Discrete version of differenciation.

$$\nabla^k X_t = k! c_k + \nabla^k Y_t = \text{const.} + error$$

Estimating trend only - Differencing

Once you apply ∇^2, then the error term look like:




```
y = diff(diff(data));
```

Which model/methods to use to detrend?

So far, we have seen lots of ways to estimate trend. Hence, a natural question is which model is the best? Philosophical question in statistics. My perspective is

- In statistics, there is no correct model, but the approximation of true model.
- This leads to the study of model selection methods.
- Hence, you should have some reasoning support your model. For example, it could be MSE, BIC, forecasting error, simplicity of model, handy calculation etc.
- If various methods indicate that your model is better than others, your model will gain more rationality.
- However, keep in mind that nobody knows the true model from the real data!
- DO NOT ASK WHETHER YOUR MODEL IS CORRECT, but ask whether your model selection is reasonable.

Estimating seasonality only

Consider that the process only has seasonal non-stationary part

$$X_t = s_t + Y_t, \quad EY_t = 0,$$

where seasonality with period d

$$\boxed{s_{t+d} = s_t = s_{t-d}}$$

- We will also assume that the period d is known.
- Three ways to estimate seasonality:
  - Harmonic regression
  - Seasonal smoothing
  - Seasonal differencing

Harmonic regression

Joseph Fourier (1768-1830) showed that

$$\{1, \cos x, \cos 2x, \cos 3x, \ldots, \sin x, \sin 2x, \ldots\}$$

forms a basis for $L^2(-\pi, \pi]$, hence f in $L^2(-\pi, \pi]$ can be represented as

$$f(x) = a_0 + \sum_{n=1}^{\infty}(a_n \cos(nx) + b_n \sin(nx)).$$

Based on this theory, we will consider finite order approximation of s_t (also extend to whole real line)

$$\boxed{s_t = a_0 + \sum_{j=1}^{k}(a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t))}$$

$$\lambda_j = \text{ fixed frequency}$$

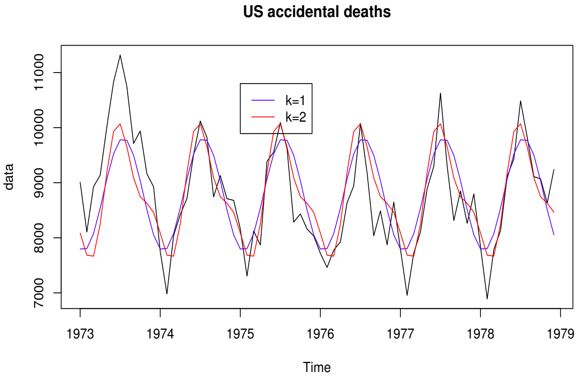
Harmonic regression

- Once k, the number of basis, and corresponding λ_j is selected, we can simply apply OLS to get estimates of coefficients.
- We will assume that k is known. Otherwise, as in the regression, you can apply variable selection to choose k. In practice $k = 1 \sim 4$.
- How to choose λ_j?
  1. Set $f_1 = [n/d]$. This is a number of cycles that s_t repeated in the data. Take $f_j = j f_1$.
  2. $\lambda_j = f_j(2\pi/n)$
- For example if $n = 72$ and $d = 12$,

$$f_1 = [72/12] = 6, \quad \lambda_j = j \times 6 \times 2\pi/72$$

Harmonic regression

Take $k = 2$ will gives the following result.




```
t=1:n; f1 = 6; f2 = 12;
costerm1 = cos(f1*2*pi/n*t); sinterm1 = sin(f1*2*pi/n*t);
costerm2 = cos(f2*2*pi/n*t); sinterm2 = sin(f2*2*pi/n*t);
out.lm2 = lm(data ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2)
```

Seasonal soothing

Basic idea is to overlay observations with period d in one cycle.

1. For $k = 1, \ldots d$,

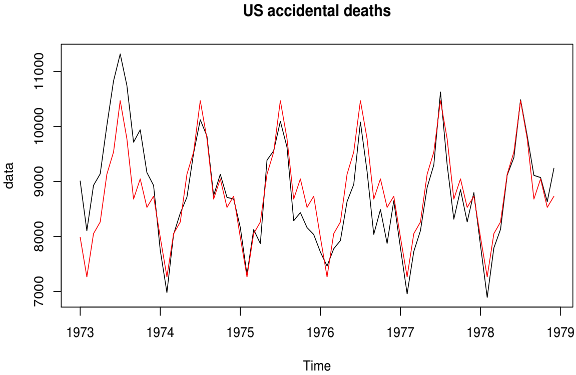
$$\hat{s_k} = \frac{1}{m}(x_k + x_{k+d} + \ldots + x_{k+(m-1)d}) = \frac{1}{m}\sum_{j=0}^{m-1} x_{k+jd},$$

   where m is the number of observations in the k-th seasonal component.

2. $\hat{s}_k = \hat{s}_{k-d}$, if $k > d$.

Graphically:

Seasonal soothing




```
library(itsmr)
season.avg = season(data, d=12);
```

Seasonal differencing

We can also eliminate seasonal trend by applying lag-d differencing.

- The lag-d difference operator, ∇_d, defined by

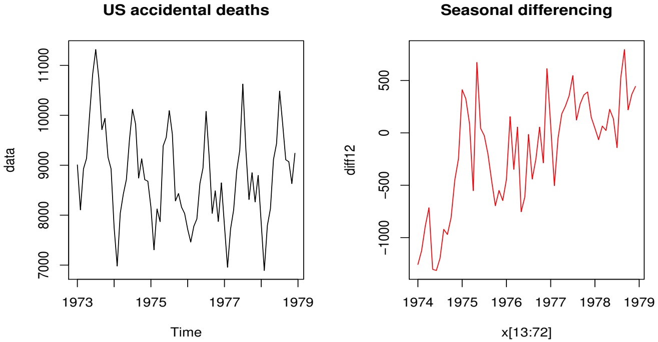
$$\nabla_d X_t = (1 - B^d) X_t, \quad t = 1, \ldots, n$$

- Beware: This is not ∇^d. It means $(1 - B)^d$ (d-th order differencing)
- If $s_t = s_{t+d}$, then

$$\nabla_d X_t = s_t - s_{t-d} + Y_t - Y_{t-d} = 0 + error$$

- Note that you cannot apply seasonal differencing to the first d observation in the real data!

Seasonal differencing




```
diff12 = diff(data, lag=12);
```

Estimating both trend and seasonality

Consider model having both trend and seasonality

$$\boxed{X_t = m_t + s_t + Y_t, \quad t = 1, \dots, n}$$

$$EY_t = 0, \quad s_{t+d} = s_t, \quad \sum_{j=1}^{d} s_j = 0$$

- We added one more condition $\sum_{j=1}^{d} s_j = 0$ so that constant becomes a part of m_t.
- Similarly, three methods can be applied here.
- Method 1: Regression
  Use polynomial regression to estimate trend and harmonic regression for seasonal component.

Estimating both trend and seasonality - Differencing

- Method 2: Differencing

  If we apply seasonal differencing, then

  $$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d},$$

  so apply trend differencing to remove trend $m_t - m_{t-d}$. Also, be aware that

  $$\nabla_d = (1 - B^d) = (1 - B)(1 + B + \ldots + B^{d-1})$$

  implies that seasonal differencing include trend differencing of order 1. Therefore, in practice, if the trend

  $$m_t = c_0 + c_1 t + \ldots c_p t^p$$

  then applying

  $$\boxed{\nabla^{p-1} \nabla_d X_t}$$

Estimating both trend and seasonality - Smoothing

- Method 3: Smoothing based classical decomposition algorithm
  Multistage algorithm to estimate trend, seasonality and noise.
  1. Obtain a rough estimate of the trend using a MA filter (we smooth over each season).
  2. Remove the estimate of trend, and estimate the season- ality by averaging over the seasons.
  3. Re-estimate the trend from the deseasonalized series via least squares.
  4. Take away the estimate of the trend and seasonality in step 2 and 3 to obtain an estimate of the noise.

Classical decomposition algorithm

(STEP1) Preliminary estimation of trend by MA filter.

If $d = 2q$,

$$\hat{m}_t = \frac{.5X_{t-q} + X_{t-q+1} + \ldots + X_{t+q-1} + .5X_{t+q}}{2q}$$

If $d = 2q + 1$,

$$\hat{m}_t = \frac{X_{t-q} + X_{t-q+1} + \ldots + X_{t+q-1} + X_{t+q}}{2q+1}$$

- Why this works to estimate trend? Note that $\sum_{j=1}^{d} s_j = 0$ (hence $\sum_{j=k}^{d+k-1} s_j = 0$ for all k) implies that MA filter vanishes seasonal terms. For example $d = 3$,

$$\frac{X_{t-1}+X_t+X_{t+1}}{3} = \frac{m_{t-1}+m_t+m_{t+1}}{3} + \frac{s_{t-1}+s_t+s_{t+1}}{3} + err = \frac{m_{t-1}+m_t+m_{t+1}}{3} + err$$

- Be aware of downweight at boundary when $d = 2q$. MA filter lenghth is $2q + 1$ (odd), so need some adjustment for even d.

Classical decomposition algorithm

(STEP2) Remove trend part, and estimate seasonal component by seasonal averaging.

$$z_t = X_t - \hat{m}_t \approx s_t + Y_t.$$

Hence, do seasonal averaging

$$\hat{s}_t^* = \frac{1}{m} \sum_{k=0}^{m-1} z_{t+kd}, \quad k = 1, \ldots, d.$$

Now, we do centering due to $\sum_j s_j = 0$.

$$\hat{s}_t = \hat{s}_t^* - \overline{\hat{s}^*}$$

Classical decomposition algorithm

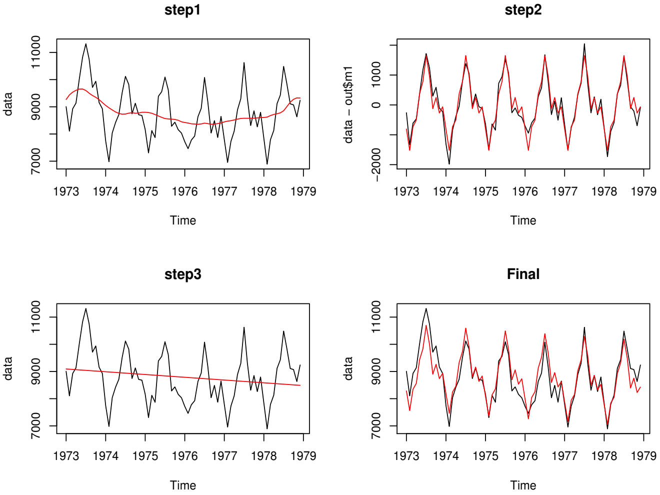
(STEP3) Reestimte the trend from the deseasonalized series via OLS.

$$\hat{m}_t^{new} = \underset{\mathbf{c}}{\operatorname{argmin}} \sum_{t=1}^{n} (X_t - \hat{s}_t - c_0 - c_1 t - c_2 t^2 - \ldots - c_p t^p)^2.$$

(STEP4) Estimate errors by

$$\hat{e}_t = X_t - \hat{m}_t^{new} - \hat{s}_t.$$

Accidental deaths - classical decomposition




```
out = classical(data, d=12, order=1);
```

Ch1.4 & Ch 2.1 The Autocovariance function (ACVF) of a stationary processes

- After successfully remove (model) trend and seasonality, it is only left with stationary errors.
- Here, we will learn how to deal with such stationary errors.
- The ACVF is the key quantity describes dependency between observations. In other words, dependency can be captures by studying ACVF.
- Math background: non-negative definiteness (n.n.d.)

Weakly stationary process

- Recall that weakly stationary TS $\{X_t\}$ is covariance stationary in the sense that

$$
\begin{aligned}
&i)\ E|X_t|^2 < \infty \quad \forall t \in \mathbb{Z} \\
&ii)\ EX_t = m, \quad \forall t \in \mathbb{Z} \\
&iii)\ \gamma_X(h) := \mathrm{Cov}(X_t, X_{t+h}) \text{ is independent of } t
\end{aligned}
$$

- Condition $iii)$ is the key quantity in analyzing a stationary TS. It means that

$$\gamma_X(h) := \mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_0, X_h) = \mathrm{Cov}(X_{t+h}, X_t).$$

- Thus, it means that to successfully understand a stationary TS, we need to understand its mean and ACVF.

Properties of ACVF

Key properties of ACVF of a stationary TS

1. $\gamma(0) = \mathrm{Var}(X_t) \geq 0$. Thus, $\rho(0) = 1$.
2. $|\gamma(h)| \leq \gamma(0)$ for all $h \in \mathbb{Z}$. Hence $|\rho(h)| \leq 1$.
3. (even function) $\quad \gamma(h) = \gamma(-h)$
4. (non-negative definiteness) For any integer $n \geq 1$ and vector $\mathbf{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$,

$$\sum_{i,j=1}^{n} a_i \gamma(i-j) a_j \geq 0$$

Matrix version:

Properties of ACVF: Bochner's Theorem

In fact, even and n.n.d determines essential feature of ACVF.

Theorem (Bochner's Theorem*)

A real-valued function defined on the integers is ACVF of a stationary process iff it is even and non-negative definite.

Now, we will introduce some examples of a stationary TS. Here we frequently use linear property of covariance

$$\mathrm{Cov}(aX + bY + c, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$$

Examples of a stationary TS

- IID process: $\{X_t\}$ are i.i.d sequence of random variables with mean μ and variance σ^2.

- WN sequence: $\{X_t\} \sim WN(0, \sigma^2)$. Relaxing the condition of independence, but assume that they are uncorrelated.

Difference between independence and uncorrelated

Only consider r.v X and Y for simplicity:

- Independence

- Uncorrelated

- Thus, IID sequence is WN, but not conversely.

Examples of a stationary TS

- (Random Walk) Let $\{X_t\}$ be WN(0, σ^2). Consider the trace of cumulate sum

$$S_0 = 0, \quad S_t = X_1 + X_2 + \ldots + X_t$$

- What about the increment $\nabla S_t = S_t - S_{t-1}$?

Examples of a stationary TS

- Consider sinusoidal

$$X_t = A\cos(\theta t) + B\sin(\theta t),$$

where $\theta \in (-\pi, \pi]$ is a fixed frequency and A and B are uncorrelated with zero means and unit varinaces. This is a Fourier series with random coefficients.

Estimation of ACVF - sample ACVF (SACVF)

- To assess the degree of dependence in data and to select a model for the data reflects this, we need to estimate ACVF from the observed data $\{x_t\}$
- SACF may suggest which of the many possible stationary TS models is a suitable candidate for representing the dependence in the data.
- For example, SACF shows $\hat{\gamma}(h) = 0$ for all $h \geq 1$, then suitable model will be
- The key idea is to use the method of moment.

Definition (SACVF)

The *sample autocovariance function* *is*

$$\boxed{\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_j - \bar{x})(x_{j+h} - \bar{x}), \quad 0 \leq h < n}$$

$$\hat{\gamma}(-h) = \hat{\gamma}(h), \quad -n < h \leq 0.$$

Sample autocorrelation function (SACF)

- It is divided by n, not by $n-h$ to achieve that $\hat{\gamma}$ is n.n.d.
- Sample autocorrelation function (SACF) is defined by

$$\boxed{\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad |h| < n.}$$

- Thus, $\hat{\rho}(h)$ measures linearity between observations

$$(x_1, x_2, \ldots, x_{n-h}) \quad \text{and} \quad (x_{1+h}, x_{2+h}, \ldots, x_n)$$

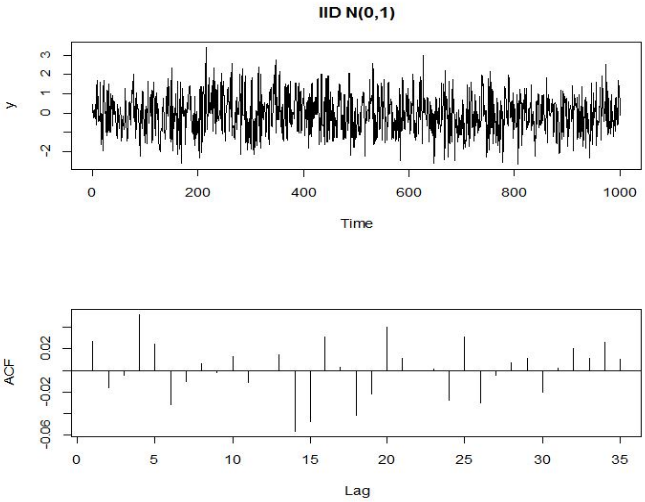
- If $\{X_t\}$ are WN(0,1), then for $h \neq 0$,

$$\hat{\rho}(h) \approx \mathcal{N}\left(0, \frac{1}{n}\right)$$

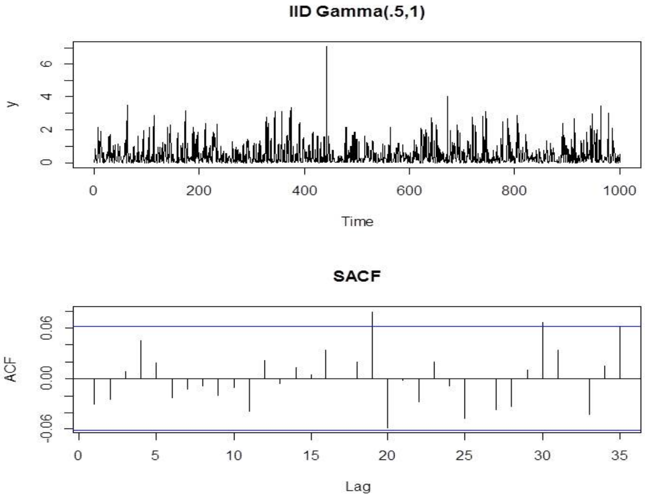
Thus, we will reject $H_0 : \hat{\rho}(h) = 0$ if $|\hat{\rho}(h)| \geq 2/\sqrt{n}$.

Examples of SACF: IID $\mathcal{N}(0,1)$

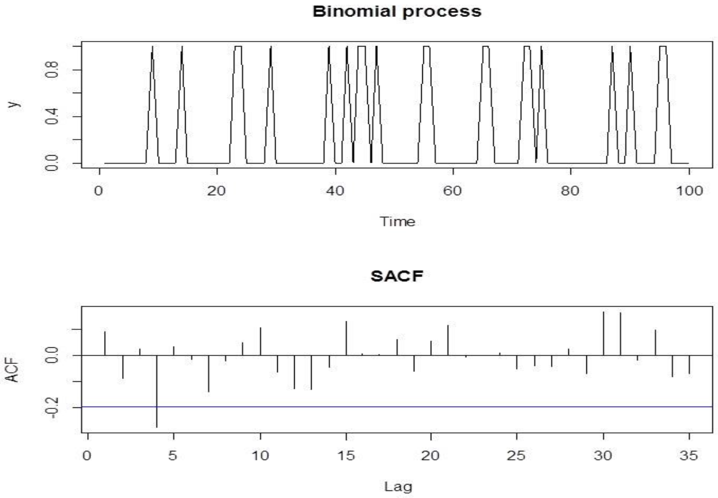
```
acf(data, lag=35);
```



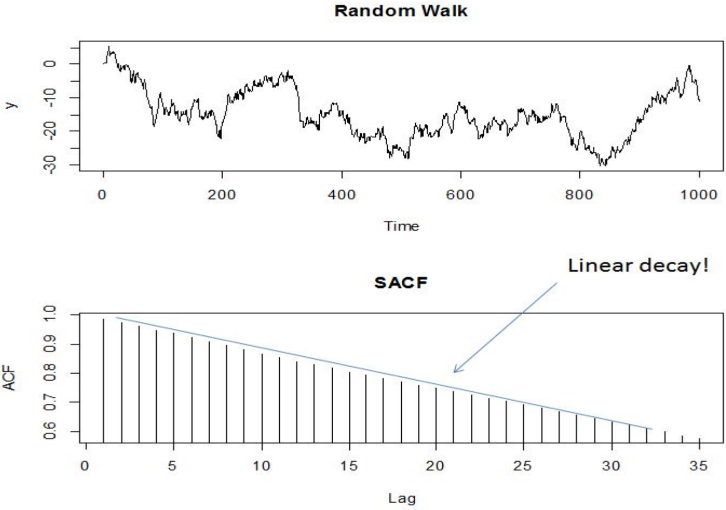
Examples of SACF: IID $\Gamma(.5, 1)$



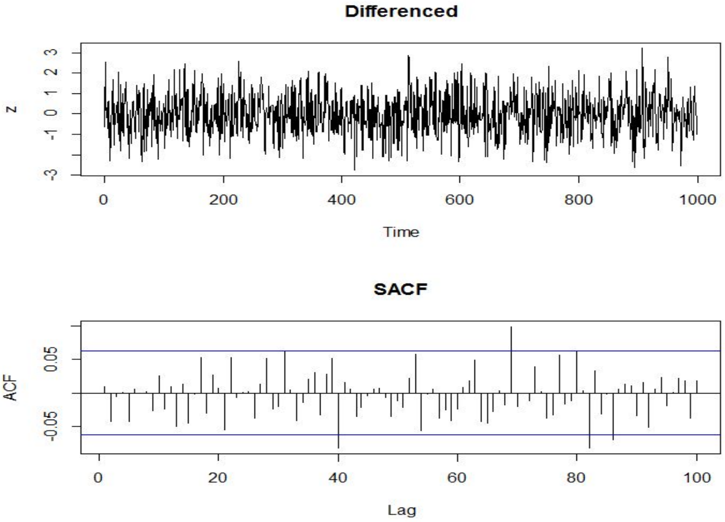
Examples of SACF: Binomial process



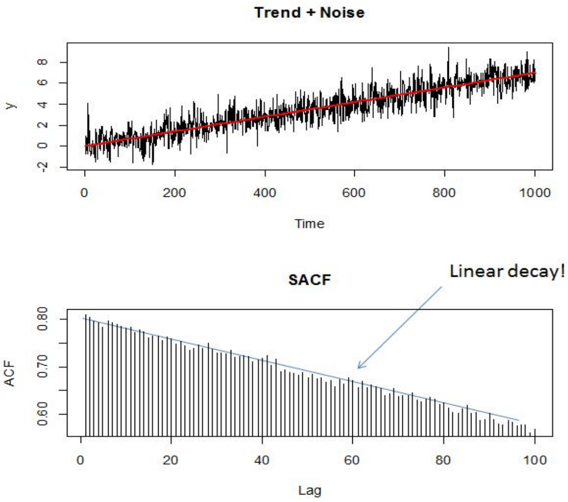
Examples of SACF: Random Walk



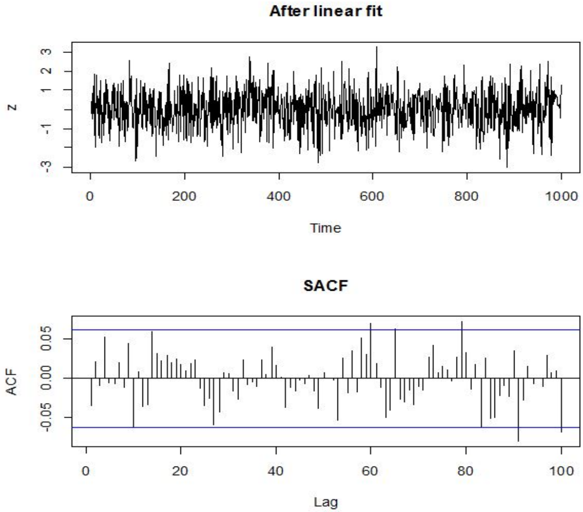
Examples of SACF: Random Walk (Differenced)



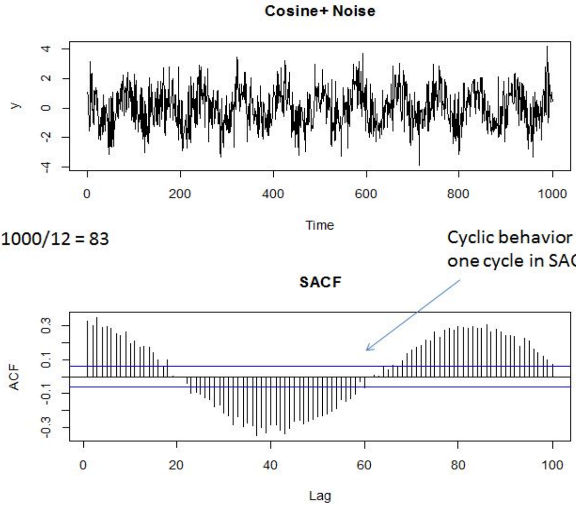
Examples of SACF: Linear trend + Noise



Examples of SACF: Linear trend + Noise



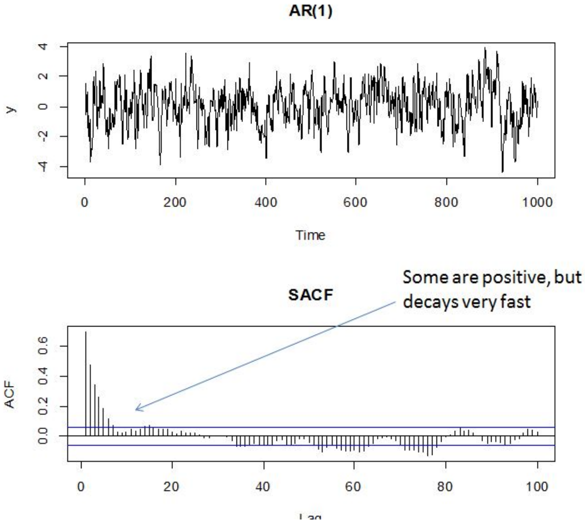
Examples of SACF: Cosine + Noise



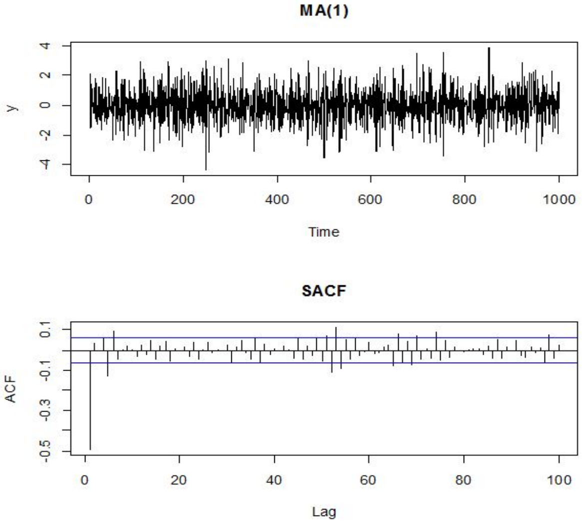
Period = 1000/12 = 83

Cyclic behavior remains & one cycle in SACF is period

Examples of SACF: AR(1)



Examples of SACF: MA(1)



Ch1.6 Testing the estimated noise sequence

- We have successfully removed trend/seasonality and the remaining residuals are stationary process.
- In particular, we further want to know whether $\{\hat{Y}_t\}$ is IID/(WN) or not.
- This is because if statioanry errors are WN/IID$(0, \sigma^2)$, you only need to estimate $\sigma^2 = \gamma(0)$. However, if there is significant dependence among the residuals, then we need to look for a more complex stationary time series model for the noise that accounts for the dependence. That is, we need to estimate ACVF $\gamma(h)$ from the residuals $\{\widehat{Y}_t\}$ to explain dependence structure.

Correlogram

SACF $\hat{\rho}(h)$ plot. If errors are WN, then

$$\hat{\rho}(h) \approx \mathcal{N}\left(0, \frac{1}{n}\right).$$

Thus, we perform testing of

$$H_0 : \rho(h) = 0 \quad \text{vs} \quad H_1 : \rho(h) \neq 0$$

Rejection rule:

- Note that $\hat{\rho}(0) = 1$.
- If $\hat{\rho}(h)$ is inside $1.96/\sqrt{n}$ bound, then we can say that the errors are uncorrelated.
- Suggest to take h upto $n/4$, but first few lags are much more important than larger lags.

Portmanteau test: IID against correlated errors

Orginal idea: [Box-Pierce] Recall from

$$\hat{\rho}(j) \approx \mathcal{N}\left(0, \frac{1}{n}\right) \Rightarrow \sqrt{n}\hat{\rho}(j) \sim \mathcal{N}(0,1)$$

so that

$$Q = n \sum_{j=1}^{H} \hat{\rho}^2(j) \approx$$

Thus, we reject

$$H_0\text{: errors are i.i.d} \quad \text{vs} \quad H_1\text{: Not } H_0$$

if

$$Q > \chi^2_H(1-\alpha)$$

Portmanteau test: IID against correlated errors

Some other refinements:

- Ljung-Box (1978): For IID sequence,

$$Q_{LB} = n(n+2)\sum_{j=1}^{H} \hat{\rho}(j)^2/(n-j) \approx \chi^2(H)$$

- McLeod and Li (1983): For IID Normal sequence,

$$\widetilde{Q} = n(n+2)\sum_{j=1}^{H} \hat{\rho}_{ww}^2(j)/(n-j) \approx \chi^2(H),$$

  where $\hat{\rho}_{ww}^2(j)$ is the SACF of the squared errors.
- We typically take $H \approx 20$, but can be taken arbitrarily.

Other tests:

More tests are introduced in the textbook. It includes

- Turning point test
- Difference sign test
- Rank test

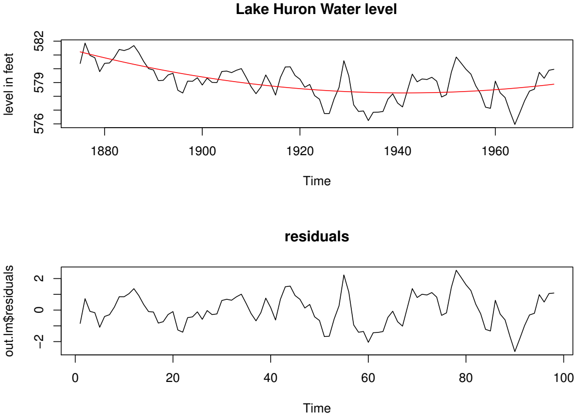
They are based upon properties of IID random variables and CLT. DIfference sign/Rank tests are in particular powerful for detecting linear trends.

Normality check

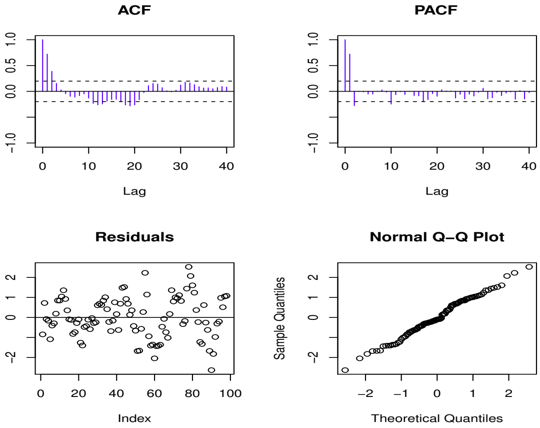
- QQ plot: Sample quantile vs Normal quantile. Plot $x_{(i)}$ versus $\Phi^{-1}((i - .5/n))$.
- Kolmogorov-Smirnov test and variants (Anderson-Darling, Cramér-von Mises test): Empirical CDF vs Theoretical CDF
- Jarque-Bera test: Based on r-th central moment
- What if errors are not Normal? We can make a transformation (log, Box-Cox transformation) to make the errors close to normal. Or, work with other family of distributions such as t-dist!

Test of randomness: Lake Huron

We want to test whether erros after eliminating quadratic trend is IID/Normal etc.



Test of randomness: Lake Huron



Observations:

Test of randomness: Lake Huron

```
>library(itsmr)
>test(out.lm$residuals)
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)    138.67         0 *
McLeod-Li Q                Q ~ chisq(20)     56.45         0 *
Turning points T      (T-64)/4.1 ~ N(0,1)       40         0 *
Diff signs S        (S-48.5)/2.9 ~ N(0,1)       50    0.6015
Rank P          (P-2376.5)/162.9 ~ N(0,1)     2406    0.8563

>library(nortest)
>lillie.test(out.lm$residuals)
 Lilliefors (Kolmogorov-Smirnov) normality test

data:  out.lm$residuals
D = 0.0724, p-value = 0.2335

>library(tseries)
>jarque.bera.test(out.lm$residuals)
 Jarque Bera Test

data:  out.lm$residuals
X-squared = 0.5376, df = 2, p-value = 0.7643
```

Chapter summary

- Population and Sample, statistical inference, probability distribution, parametric modelling.
- Time series analysis
- Stationarity concepts: weakly stationary and strictly stationary.
- ACF and SACF
- Decomposing trend, seasonal and stationary errors. Three major tools are i) regression, ii) smoothing and iii) lag differencing.
- Once we remove trend and seasonal component, we perform the test of randomness to check IID errors. If IID errors, only need to estimate $\gamma(0)$, otherwise need to estimate $\gamma(h)$ for all lags h.