Introduction to Time Series Analysis

Dr David A. Stephens

Department of Mathematics & Statistics
McGill University
Room 1225, Burnside Hall

david.stephens@mcgill.ca
www.math.mcgill.ca/dstephens/

Time Series: Prerequisites

- Basic distribution theory
- Expectations and other moments
- Basics of statistical theory (ie estimation, testing, large sample theory)
- Linear algebra (matrix manipulations and decomposition)
- Basic knowledge of complex numbers
- Computing an advantage (knowledge of R).

Time Series: Objectives

The basic objective is to construct models for observed data

$$x_1, x_2, \ldots, x_t, \ldots$$

with t indexing time, with the goals of

- explaining the underlying data generating mechanism
- forecasting future values of the underlying process.

We will mainly focus on discrete time modelling, and denote the underlying stochastic (random) process $\{X_t\}$.

Time Series: Objectives

The stochastic process can be (partly) characterized via its

- finite dimensional distributions, that is the joint distributions of

  $$(X_{t_1}, \ldots, X_{t_n})$$

  for each index set $\{t_1, \ldots, t_n\}$
- moment properties (expectation, variance, covariance)

Typically $\{X_t\}$ will not be independent, and inference about the parameters of the joint distribution will be of central interest.

Time Series: Forecasting

Given observed values $X_1 = x_1, X_2 = x_2, \ldots, X_T = x_T$, we aim to predict or forecast future values of the process

$$X_{T+1}, \ldots, X_{T+K}$$

say.

For example, could use a simple linear predictor

$$\hat{X}_{T+1} = \sum_{t=1}^{T} a_t x_t$$

that is, a linear combination of past values of the process. The objective then is to choose the optimal values for the a_t.

Chapter 1
Introduction

1.1 Introduction

A time series is a sequence of random variables

$$\{X_t\}_{t=1,\ldots,T} = (X_1, \ldots, X_T) = X_{1:T}$$

or a realization

$$\{x_t\}_{t=1,\ldots,T} = (x_1, \ldots, x_T) = x_{1:T}$$

collected over time index t representing some unit time period (hours, days, months etc.).

1.2 Simple Stochastic Models

A time series model is a probability model specifying

- the joint distribution
- the second-order moment structure (mean, variance, covariance)

of $\{X_t\}$.

The joint distribution specifies the moment structure.

1.2 Simple Stochastic Models

Recall that for any two random variables, the covariance is defined as

$$\text{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

where $\mu_i = \mathbb{E}[X_i]$ and $\mu_j = \mathbb{E}[X_j]$, and the correlation is

$$\text{Corr}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i]\text{Var}[X_j]}}$$

1.2 Simple Stochastic Models

For a d-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_d)^\top$, we form the covariance matrix, Σ, so that the (i,j)th element of Σ is

$$[\Sigma]_{i,j} = \mathrm{Cov}[X_i, X_j].$$

This matrix is symmetric, and non-negative definite. We may write

$$\Sigma = \mathbf{V}\mathbf{R}\mathbf{V}^\top$$

where

$$\mathbf{V} = \mathrm{diag}(\sqrt{\mathrm{Var}[X_1]}, \ldots, \sqrt{\mathrm{Var}[X_d]})$$

and $\mathbf{R}$ is the matrix of correlations.

1.2 Simple Stochastic Models

Special Case: Independence

$$F_{X_{1:T}}(x_{1:T}) = \prod_{t=1}^{T} F_X(x_t) = \prod_{t=1}^{T} \Pr[X_t \leqslant x_t]$$

or equivalently, for each t,

$$\mathbb{P}[X_t \leqslant x_t | X_{1:(t-1)} = x_{1:(t-1)}] = \Pr[X_t \leqslant x_t]$$

where the past values of X contain no information about future X.

Note: If two variables are independent, then they are also uncorrelated; however, in general uncorrelatedness does not imply independence.

1.2 Simple Stochastic Models

For the specification via moments, interest focusses on

$$\begin{aligned} \text{Expectation} \quad & \mu_t = \mathbb{E}[X_t] \\ \text{Variance} \quad & \sigma^2_{X_t} = \text{Var}[X_t] \\ \text{Covariance} \quad & \gamma_X(t,s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)] \end{aligned}$$

for any indices t, s. As the X_t usually represent repeated measurements of the same phenomenon over time, the term autocovariance is used.

1.2 Simple Stochastic Models

Example: IID Process

Let $0 < p < 1$, and suppose

$$\mathbb{P}[X_t = 1] = p \qquad \mathbb{P}[X_t = -1] = 1 - p$$

with the X_t mutually independent. Then $\{X_t\}$ is an IID (independent, identically distributed) process.

1.2 Simple Stochastic Models

Example: IID Process

We have that

$$\mu_t = p + (1-p)(-1) = 2p - 1$$

$$\sigma_t^2 = [p(1)^2 + (1-p)(-1)^2] - (2p-1)^2 = 1 - (1-2p)^2$$

$$\gamma(t,s) = 0 \qquad t \neq s$$

and these quantities do not depend non t or s.

1.2 Simple Stochastic Models

Example: Random Walk

Suppose $\{X_t\}$ is an IID process, and let

$$S_t = \sum_{i=1}^{t} X_i = S_{t-1} + X_t$$

Then $\{S_t\}$ is a random walk. Also, setting $S_0 = 0$, we have

$$X_t = S_t - S_{t-1}$$

that is, $\{X_t\}$ can be obtained from $\{S_t\}$ by differencing.

1.2 Simple Stochastic Models

Example: Random walk

Suppose $\{X_t\}$ is an IID process with

$$\mathbb{P}[X_t = 1] = \mathbb{P}[X_t = -1] = \frac{1}{2}$$

and zero otherwise, and let

$$S_t = \sum_{i=1}^{t} X_i.$$

Then both $\{X_t\}$ and $\{S_t\}$ are zero-mean processes.

If $p \neq 1/2$, then $\mathbb{E}[X_t], \mathbb{E}[S_t]$ are non-zero; if $p < 1/2$, then $\mathbb{E}[S_t] < 0$ and we have a downward drift.

Example: Random walk

Note that for any (time-homogeneous) random walk

$$\mathrm{Var}[S_t] = \mathrm{Var}\left[\sum_{i=1}^{t} X_i\right] = \sum_{i=1}^{t} \mathrm{Var}\left[X_i\right] = t\mathrm{Var}\left[X_1\right]$$

by the independence of the X_t, so the variance grows linearly with t.

Thus even if $\mathbb{E}[S_t] = 0$, the probability distribution of S_t has increasing variability around zero, and will (almost surely) diverge to an infinite value.

1.3 Gaussian processes

Suppose, for all finite collections $X_1, \ldots, X_n$, $n \geqslant 1$, the joint distribution is multivariate Gaussian (Normal) with

$$\mathbb{E}[X_t] = \mu_X$$

and covariance defined for X_t, X_s as

$$\mathrm{Cov}[X_t, X_s] = \gamma_X(|t - s|) \qquad t, s \in \{1, \ldots, n\}.$$

for some function of a single argument γ_X. We have that

$$\mathrm{Var}[X_t] = \gamma_X(0).$$

1.3 Gaussian processes

Note that the autocovariance only depends on the value of

$$|t - s|$$

This version imposes extra conditions, which corresponds to a structured covariance matrix which has fewer than $n(n+1)/2$ different elements.

1.3 Gaussian processes

Let $\mathbf{\Gamma}_X(n)$ denote the $(n \times n)$ matrix with

$$[\mathbf{\Gamma}_X(n)]_{t,s} = \gamma_X(|t-s|).$$

Then $\mathbf{\Gamma}_X(n)$ is a symmetric, positive definite matrix so that

$$\mathbf{x}^\top \mathbf{\Gamma}_X(n)\mathbf{x} > 0 \qquad \mathbf{x} \in \mathbb{R}^n.$$

1.3 Gaussian processes

The matrix $\mathbf{\Gamma}_X(n)$ has Toeplitz structure, constant among diagonals,

$$\mathbf{\Gamma}_X(n) = \begin{bmatrix} \gamma_X(0) & \gamma_X(1) & \cdots & \cdots & \gamma_X(n-1) \\ \gamma_X(1) & \gamma_X(0) & \gamma_X(1) & \cdots & \vdots \\ \gamma_X(2) & \gamma_X(1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \gamma_X(1) \\ \gamma_X(n-1) & \cdots & \cdots & \gamma_X(1) & \gamma_X(0) \end{bmatrix}$$

We write the vector

$$\mathbf{X}_n \equiv X_{1:n} \sim \mathcal{N}_n(\mu_X \mathbf{1}_n, \mathbf{\Gamma}_X(n)).$$

where $\mathbf{1} = (1, 1, \ldots, 1)^\top$ is an $n \times 1$ vector.

1.3 Gaussian processes

We have the decomposition

$$\mathbf{X}_n = \mu_X \mathbf{1}_n + \mathbf{Z}_n \qquad (n \times 1)$$

where

$$\mathbf{Z}_n \sim \mathcal{N}_n(\mathbf{0}_n, \mathbf{\Gamma}_X(n)).$$

1.3 Gaussian processes

In general, if

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

with $\mathbf{X}_1$ and $\mathbf{X}_2$ being $k_1 \times 1$ and $k_2 \times$ respectively, where

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

then

$$\mathbf{X}_1 \sim \mathcal{N}_{k_1}\left(\boldsymbol{\mu}_1, \Sigma_{11}\right)$$

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim \mathcal{N}_{k_2}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right).$$

1.3 Gaussian processes

Specifically, if we consider the one-step-ahead (conditional) probability distribution

$$X_{n+1} | X_{1:n} = x_{1:n}$$

it follows by standard properties that

$$X_{n+1} | X_{1:n} = x_{1:n} \sim \mathcal{N}\left(\mu_{(n+1)|n}, \sigma^2_{(n+1)|n}\right)$$

1.3 Gaussian processes

Write $\gamma_k = \gamma_X(k)$ for each $k \geqslant 0$. The covariance matrix of $X_{1:(n+1)}$ is

$$\boldsymbol{\Gamma}_X(n+1) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-2} & \gamma_{n-1} & \gamma_n \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-3} & \gamma_{n-2} & \gamma_{n-1} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{n-4} & \gamma_{n-3} & \gamma_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{n-2} & \gamma_{n-3} & \gamma_{n-4} & \cdots & \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_n & \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Gamma}_X(n) & \boldsymbol{\gamma}_n^{\mathrm{R}} \\ \{\boldsymbol{\gamma}_n^{\mathrm{R}}\}^\top & \gamma_0 \end{bmatrix}$$

where

$$\boldsymbol{\gamma}_n^{\mathrm{R}} = [\gamma_n, \gamma_{n-1}, \gamma_{n-1}, \ldots, \gamma_2, \gamma_1]^\top$$

1.3 Gaussian processes

Applying the previous formula, we have

$$\mu_{(n+1)|n} = \mu_X + \{\boldsymbol{\gamma}_n^{\mathrm{R}}\}^\top \{\boldsymbol{\Gamma}_X(n)\}^{-1}(\mathbf{x}_{1:n} - \mu_X \mathbf{1}_n),$$

and

$$\sigma^2_{(n+1)|n} = \gamma_0 - \{\boldsymbol{\gamma}_n^{\mathrm{R}}\}^\top \{\boldsymbol{\Gamma}_X(n)\}^{-1} \boldsymbol{\gamma}_n^{\mathrm{R}}.$$

Note therefore that

$$\sigma^2_{(n+1)|n} < \gamma_0$$

as $\boldsymbol{\Gamma}_X(n)$ is positive definite.

1.4 Stationarity

Suppose $\{X_t\}$ is a time series process, where for all t

$$\mathbb{E}[X_t^2] < \infty.$$

Define

$$\text{Mean function :} \quad \mu_X(t) = \mathbb{E}[X_t]$$

$$\text{Covariance function :} \ \gamma_X(t,s) = \mathbb{E}[(X_t - \mu_X(t))(X_s - \mu_X(s))]$$

for integers t, s. Note that

$$\gamma_X(t,s) = \gamma_X(s,t).$$

1.4 Stationarity

The process $\{X_t\}$ is weakly stationary if

(i) $\mu_X(t)$ does not depend on t

(ii) $\gamma_X(t+h,t)$ does not depend on t for each integer h, so that we may write

$$\gamma_X(t,t+h) \equiv \gamma_X(h).$$

Note also that we must have

$$\gamma_X(-h) = \gamma_X(t,t-h) = \gamma_X(t-h,t) = \gamma_X(h).$$

for $h \geqslant 1$.

Note: The quantity h is often referred to as the lag.

1.4 Stationarity

$\{X_t\}$ is strongly stationary if

$$X_{t:(t+n)} = (X_{t+1}, \ldots, X_{t+n})$$

and

$$X_{(t+h+1):(t+h+n)} = (X_{t+h+1}, \ldots, X_{t+h+n})$$

have the same joint distribution for all t, h, n.

In practice, the (multivariate) distributions required to model such strongly stationary processes are limited in number.

1.4 Stationarity

If $\{X_t\}$ is weakly stationary process define

- the autocovariance function (ACVF), $\gamma_X(h)$, by

$$\gamma_X(h) \equiv \gamma_X(h,0) = \gamma_X(t+h,t)$$

- the autocorrelation function (ACF), $\rho_X(h)$, by

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

That is

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t] \qquad \rho_X(h) = \text{Corr}[X_{t+h}, X_t].$$

The fundamental concept of stationarity is that the stochastic structure does not change though time

1.4 Stationarity

Example: IID Process

An IID process $\{X_t\} \sim \text{IID}(0, \sigma_X^2)$, is characterized by

$$\begin{aligned} \mathbb{E}[X_t] &= 0 \\ \text{Var}[X_t] = \mathbb{E}[X_t^2] &= \sigma_X^2 < \infty. \end{aligned}$$

with the X_t mutually independent. This process is stationary:

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t] = \begin{cases} \sigma_X^2 & h = 0 \\ 0 & h > 0 \end{cases}$$

$$\rho_X(h) = \text{Corr}[X_{t+h}, X_t] = \begin{cases} 1 & h = 0 \\ 0 & h > 0 \end{cases}$$

1.4 Stationarity

Example: White Noise Process

A white noise process $\{X_t\} \sim \text{WN}(0, \sigma_X^2)$ is characterized by

$$\begin{aligned} \mathbb{E}[X_t] &= 0 \\ \text{Var}[X_t] = \mathbb{E}[X_t^2] &= \sigma_X^2 < \infty. \end{aligned}$$

with the X_t uncorrelated. This process is stationary:

$$\gamma_X(h) = \text{Cov}[X_{t+h}, X_t] = \begin{cases} \sigma_X^2 & h = 0 \\ 0 & h > 0 \end{cases}$$

$$\rho_X(h) = \text{Corr}[X_{t+h}, X_t] = \begin{cases} 1 & h = 0 \\ 0 & h > 0 \end{cases}$$

1.4 Stationarity

Example: Random Walk

If $\{X_t\} \sim \text{WN}(0, \sigma_X^2)$, then the process $\{S_t\}$ defined by

$$S_t = \sum_{i=1}^{t} X_i$$

is non-stationary as, by properties of uncorrelated rvs,

$$\mathbb{E}[S_t] = 0$$

$$\text{Var}[S_t] = t\sigma_X^2$$

which does depend on t.

Example: Random Walk

For the ACVF:

$$\begin{aligned}
\gamma_S(t+h,t) &= \mathrm{Cov}[S_{t+h}, S_t] \\
&= \mathrm{Cov}[S_t + X_{t+1} + \cdots + X_{t+h}, S_t] \\
&= \mathrm{Cov}[S_t, S_t] + \sum_{i=1}^{h} \mathrm{Cov}[X_{t+i}, S_t] \\
&= \mathrm{Cov}[S_t, S_t] \\
&= \mathrm{Var}[S_t] \\
&= t\sigma_X^2.
\end{aligned}$$

Example: Moving Average

Let $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and define process $\{X_t\}$ by

$$X_t = Z_t + \theta_1 Z_{t-1} \qquad t \in \mathbb{Z}$$

where θ_1 is a real-valued parameter. As $\mathbb{E}[Z_t] = 0$, we also have $\mathbb{E}[X_t] = 0$, and by uncorrelatedness of the $\{Z_t\}$ components

$$\text{Var}[X_t] = \text{Var}[Z_t] + \theta_1^2 \text{Var}[Z_{t-1}] = (1 + \theta_1^2)\sigma_Z^2$$

which does not depend on t.

Example: Moving Average

We have

$$\begin{aligned}\gamma_X(t+h,t) &= \mathrm{Cov}[X_{t+h}, X_t] = \mathbb{E}[X_t X_{t+h}] \\ &= \mathbb{E}[(Z_t + \theta_1 Z_{t-1})(Z_{t+h} + \theta_1 Z_{t+h-1})] \\ &= \mathbb{E}[Z_{t+h} Z_t] + \theta_1 \left(\mathbb{E}[Z_{t+h} Z_{t-1}] + \mathbb{E}[Z_{t+h-1} Z_t]\right) \\ &\quad + \theta_1^2 \mathbb{E}[Z_{t-1} Z_{t+h-1}]\end{aligned}$$

Now,

$$\mathbb{E}[Z_j Z_k] = \begin{cases} \sigma_Z^2 & j = k \\ 0 & j \neq k \end{cases}.$$

Example: Moving Average

Hence

$$\gamma_X(t+h,t) = \begin{cases} \sigma_Z^2(1+\theta_1^2) & h=0 \\ \sigma_Z^2\theta_1 & h=\pm 1 \\ 0 & \text{otherwise} \end{cases}.$$

$$\rho_X(t+h,t) = \begin{cases} 1 & h=0 \\ \theta_1/(1+\theta_1^2) & h=\pm 1 \\ 0 & \text{otherwise} \end{cases}$$

These functions do not depend on t, so $\{X_t\}$ is a stationary process

$$\{X_t\} \sim \text{MA}(1)$$

Example: Autoregression

Let $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and define process $\{X_t\}$ by

$$X_t = \phi_1 X_{t-1} + Z_t \qquad t \in \mathbb{Z}$$

where ϕ_1 is a real-valued parameter with $|\phi_1| < 1$.

Assume for the moment that $\{X_t\}$ is stationary, and that

$$\text{Cov}[Z_t, X_s] = 0 \qquad t > s.$$

that is, future values of the Z series are uncorrelated with past values of the X series.

1.4 Stationarity

Example: Autoregression

Then

$$\mathbb{E}[X_t] = \phi_1 \mathbb{E}[X_{t-1}] + \mathbb{E}[Z_t] = \phi_1 \mathbb{E}[X_{t-1}]$$

so therefore by the stationarity assumption $\mathbb{E}[X_t] = 0$.

Example: Autoregression

We have for all $h > 0$

$$\begin{aligned}
\gamma_X(h) &= \mathrm{Cov}[X_{t+h}, X_t] = \mathrm{Cov}[X_{t-h}, X_t] \\
&= \mathrm{Cov}[\phi_1 X_{t-1}, X_{t-h}] + \mathrm{Cov}[Z_t, X_{t-h}] \\
&= \phi_1 \mathrm{Cov}[X_{t-1}, X_{t-h}] + 0 \\
&= \phi_1 \gamma_X(h-1) \\
&= \phi_1^2 \gamma_X(h-2) \\
&\vdots \\
&= \phi_1^h \gamma_X(0)
\end{aligned}$$

by recursion.

Example: Autoregression

Note that

$$\mathbb{E}[X_t X_{t-h}] = \mathbb{E}[X_{t+h} X_t]$$

by the stationarity assumption. Therefore

$$\gamma_X(-h) = \gamma_X(h)$$

and so

$$\rho_X(h) = \phi_1^{|h|} \qquad h = 0, \pm 1, \pm 2$$

that is, a geometrically decaying ACF.

Example: Autoregression

Note also that

$$\begin{aligned}\gamma_X(0) &= \mathrm{Cov}[X_t, X_t] \\ &= \mathrm{Cov}[(\phi_1 X_{t-1} + Z_t)(\phi_1 X_{t-1} + Z_t)] \\ &= \phi_1^2 \mathrm{Cov}[X_{t-1}, X_{t-1}] + \mathrm{Cov}[Z_t, Z_t] \\ &= \phi_1^2 \gamma_X(0) + \sigma_Z^2\end{aligned}$$

so therefore

$$\gamma_X(0) = \frac{\sigma_Z^2}{1 - \phi_1^2}$$

$\{X_t\}$ is the autoregressive process of order 1: $\{X_t\} \sim \mathrm{AR}(1)$.

Example: Autoregression

From the definition, we have

$$
\begin{aligned}
X_t &= \phi_1 X_{t-1} + Z_t \\
&= \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \phi_1^2 X_{t-2} + (Z_t + \phi_1 Z_{t-1}) \\
&= \phi_1^2(\phi_1 X_{t-3} + Z_{t-2}) + (Z_t + \phi_1 Z_{t-1}) \\
&= \phi_1^3 X_{t-3} + (Z_t + \phi_1 Z_{t-1} + \phi_1^2 Z_{t-2}) \\
&\vdots \\
&= \phi_1^h X_{t-h} + \sum_{j=0}^{h-1} \phi_1^j Z_{t-j}
\end{aligned}
$$

Example: Autoregression

If we allow $h \longrightarrow \infty$, provided $\left|\phi_1\right| < 1$, we can continue the recursion indefinitely to obtain

$$X_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$$

as the first term converges to zero.

1.4 Stationarity

Note: when we write

$$X_t = Z_t + \theta_1 Z_{t-1}$$

or

$$X_t = \phi_1 X_{t-1} + Z_t$$

we need to be precise about the meaning of the symbol '='.

1.4 Stationarity

For random variables we have (for example)

- equality in distribution

$$X_t \stackrel{d}{=} Y_t$$

(i.e. the distribution of X_t is the same as the distribution of Y_t)

- equality in probability

$$X_t \stackrel{p}{=} Y_t$$

(i.e. $\Pr\left[|X_t - Y_t| < \epsilon\right] = 1$ for all $\epsilon > 0$)

1.4 Stationarity

- equality with probability 1 (almost sure equality)

$$X_t \overset{a.s.}{=} Y_t$$

(i.e. For each ω, $\Pr[|X_t(\omega) - Y_t(\omega)| < \epsilon] = 1$ for all $\epsilon > 0$.)

- mean-square equality

$$X_t \overset{m.s.}{=} Y_t$$

(i.e. $\mathbb{E}[(X_t - Y_t)^2] = 0$)

1.4 Stationarity

In the case of an explicit formula, we can interpret the equality as a definition

$$X_t = Z_t + \theta_1 Z_{t-1}$$

but for an implicit formula

$$X_t = \phi_1 X_{t-1} + Z_t$$

we may need to use the other definitions.

1.4 Stationarity

For a stationary process, moment estimators can be used to estimate the mean/covariance structure. The estimates are

$$\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n}(x_t - \widehat{\mu})(x_{t+|h|} - \widehat{\mu}) \qquad -n < h < n$$

These are consistent estimators of the two functions. The $n \times n$ matrix estimate

$$\widehat{\mathbf{\Gamma}}_n = [\widehat{\gamma}(i-j)]_{ij}$$

is non-negative definite.

1.5 Trends and Seasonality

The models from the previous section are purely stochastic. It is also possible to incorporate deterministic components.

We focus on

- trends,
- seasonality.

1.5 Trends and Seasonality

Trends: Suppose that for each t,

$$X_t = m_t + E_t$$

where

- m_t is a deterministic function of t,
- E_t is a purely stochastic, zero mean time series.

For example, might have

$$m_t = a_0 + a_1 t$$

that is, a linear trend in time. Constants a_0, a_1 are in general unknown and must be estimated.

1.5 Trends and Seasonality

Ordinary Least-Squares (OLS): estimate a_0, a_1 by minimizing the sum of squared errors

$$\hat{a} = \arg\min \left\{ \sum_{t=1}^{T} (x_t - a_0 - a_1 t)^2 \right\}.$$

Provided $\mathbb{E}[E_t] = 0$, OLS estimators are consistent estimators of the true parameters (if the trend truly is linear)

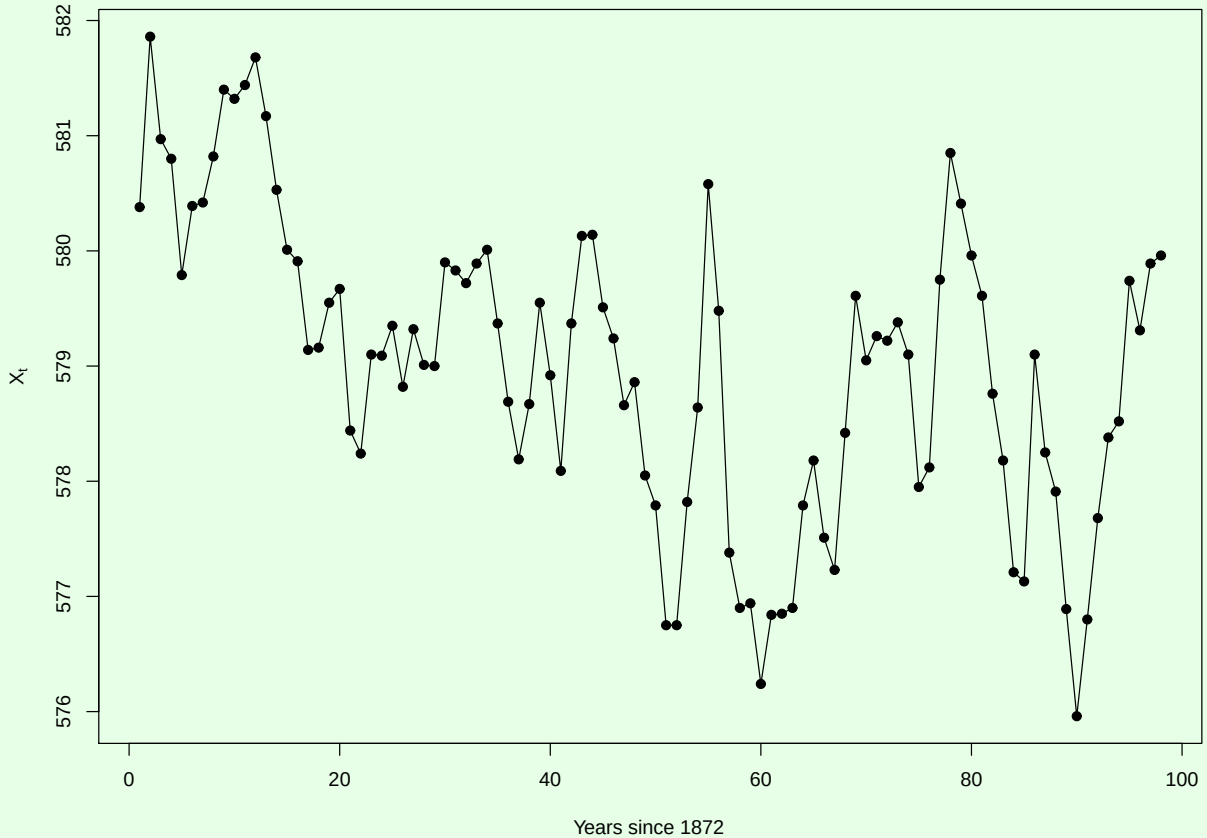
$$\hat{a} \xrightarrow{p} a$$

as $n \longrightarrow \infty$. OLS is easily applied for many types of deterministic trend (e.g. polynomial trend).

1.5 Trends and Seasonality

Example: Lake Huron data

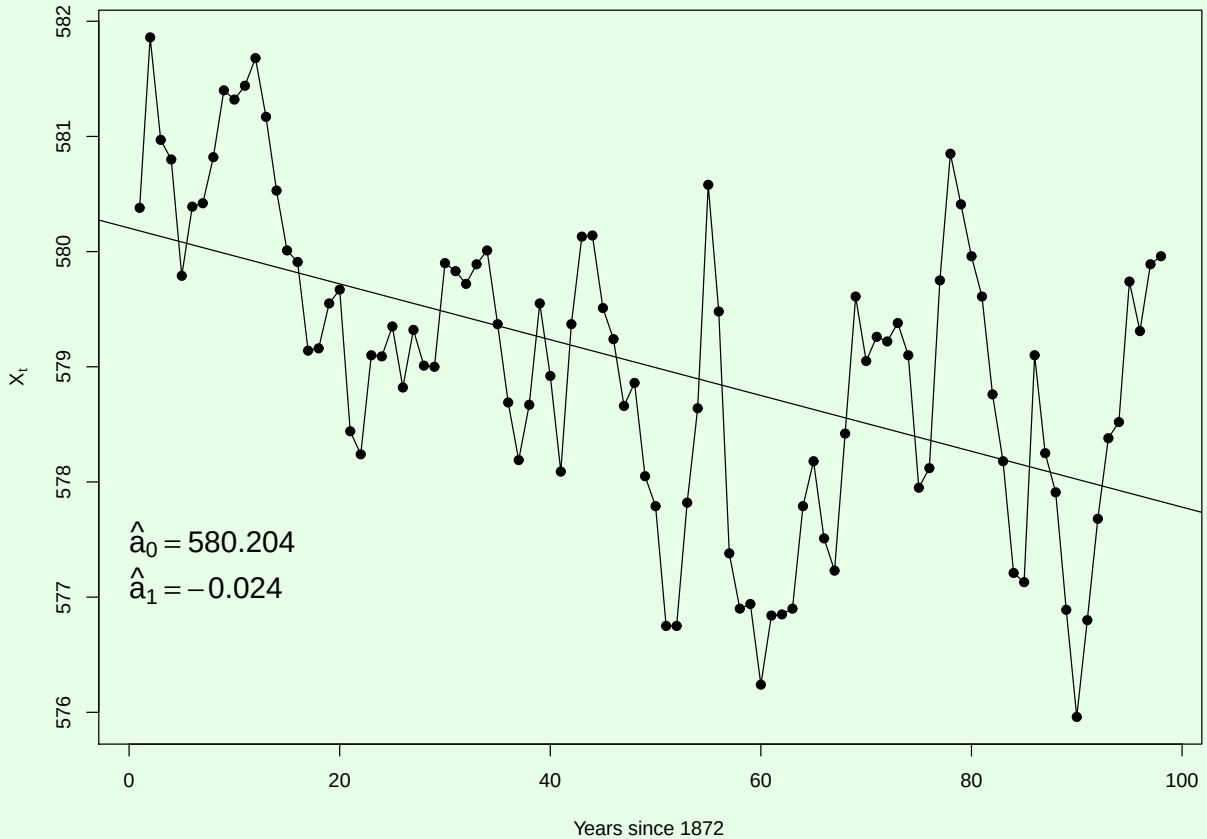
Annual level (in ft) of Lake Huron 1875-1972 ($T = 98$).



1.5 Trends and Seasonality

Example: Lake Huron data

Fitted linear trend: $\hat{a}_0 = 580.204$, $\hat{a}_1 = -0.024$.



1.5 Trends and Seasonality

Example: Lake Huron data

After fitting the linear trend model

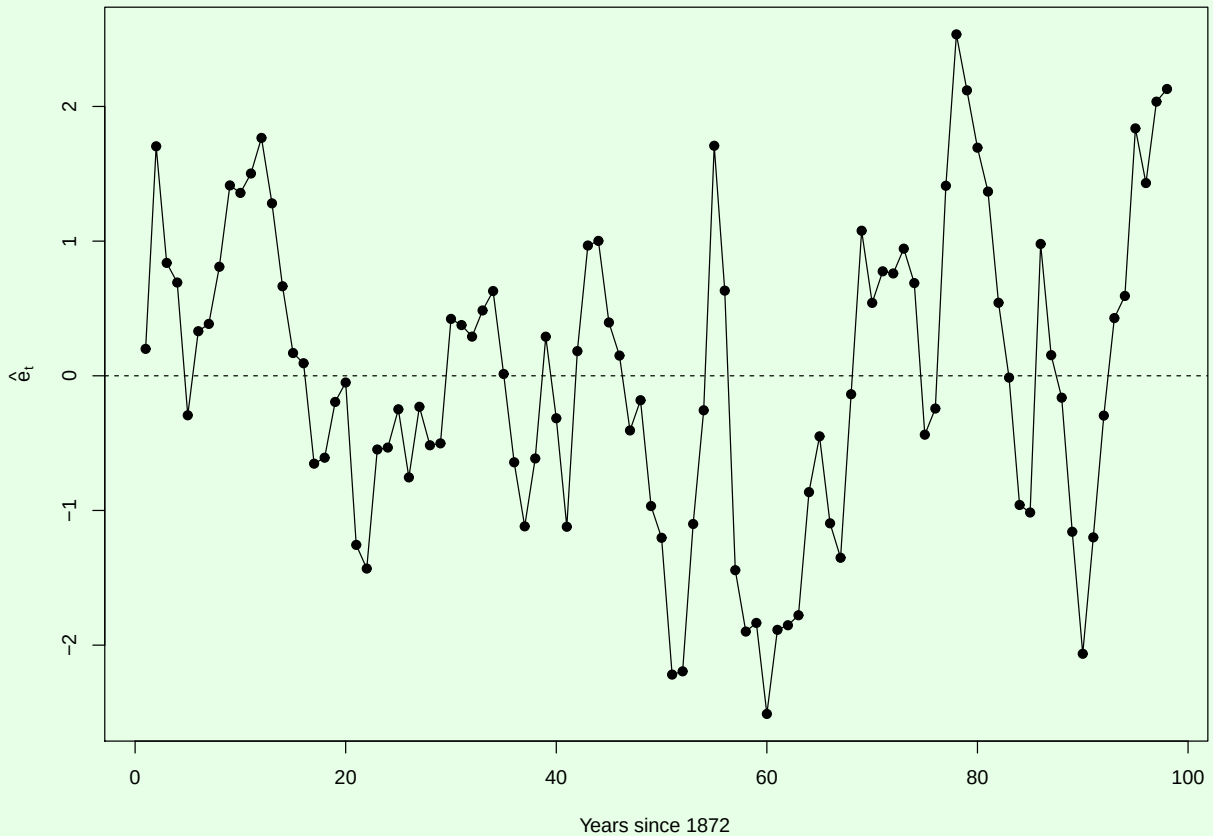
$$X_t = a_0 + a_1 t + E_t$$

by OLS to obtain $\hat{a}_0, \hat{a}_1$, we may examine the residual series

$$\hat{e}_t = x_t - \hat{a}_0 - \hat{a}_1 t$$

Example: Lake Huron data

Residual series: $\hat{e}_t = x_t - \hat{a}_0 - \hat{a}_1 t.$



1.5 Trends and Seasonality

Example: Lake Huron data

The residual series is approximately zero mean. However, there is some structure:

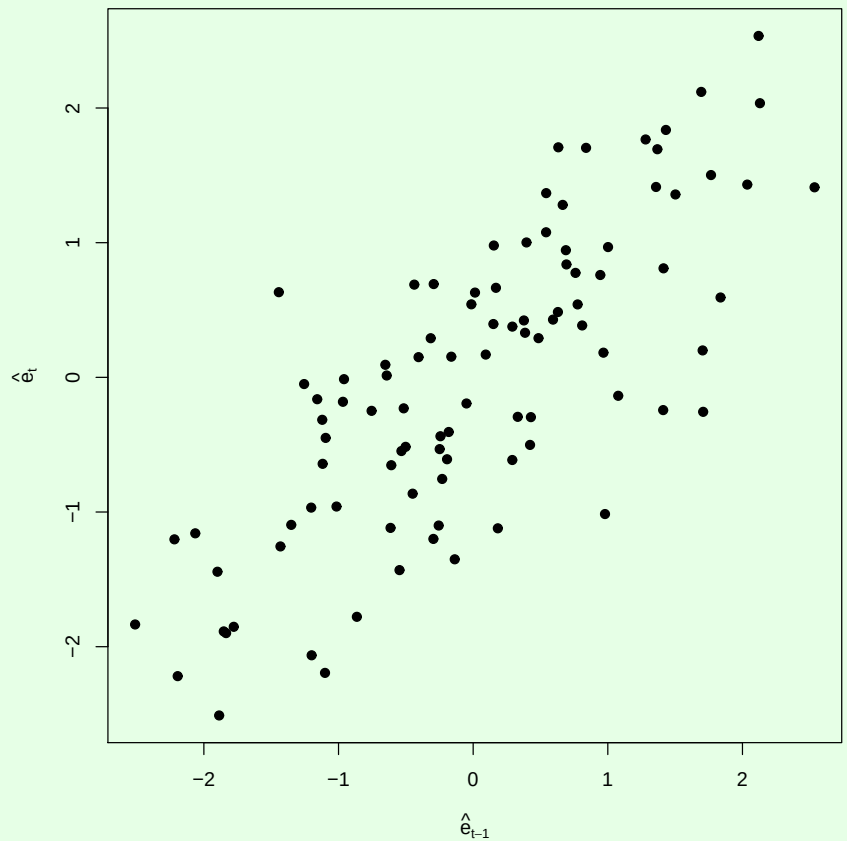
- the variance is not constant over time ?
- successive residuals $(\hat{e}_{t-1}, \hat{e}_t)$ are positively correlated. The sample correlation is

$$\text{corr}[\hat{e}_{t-1}, \hat{e}_t] \approx 0.775$$

That is, the residual series **does not** appear to be a realization of an IID process.

1.5 Trends and Seasonality

Example: Lake Huron data: $\widehat{e}_t$ vs $\widehat{e}_{t-1}$.



Example: Lake Huron data

The positive dependence is potentially useful for forecasting; if

$$\hat{e}_{93}, \hat{e}_{94}, \ldots, \hat{e}_{98}$$

are positive, it is likely that

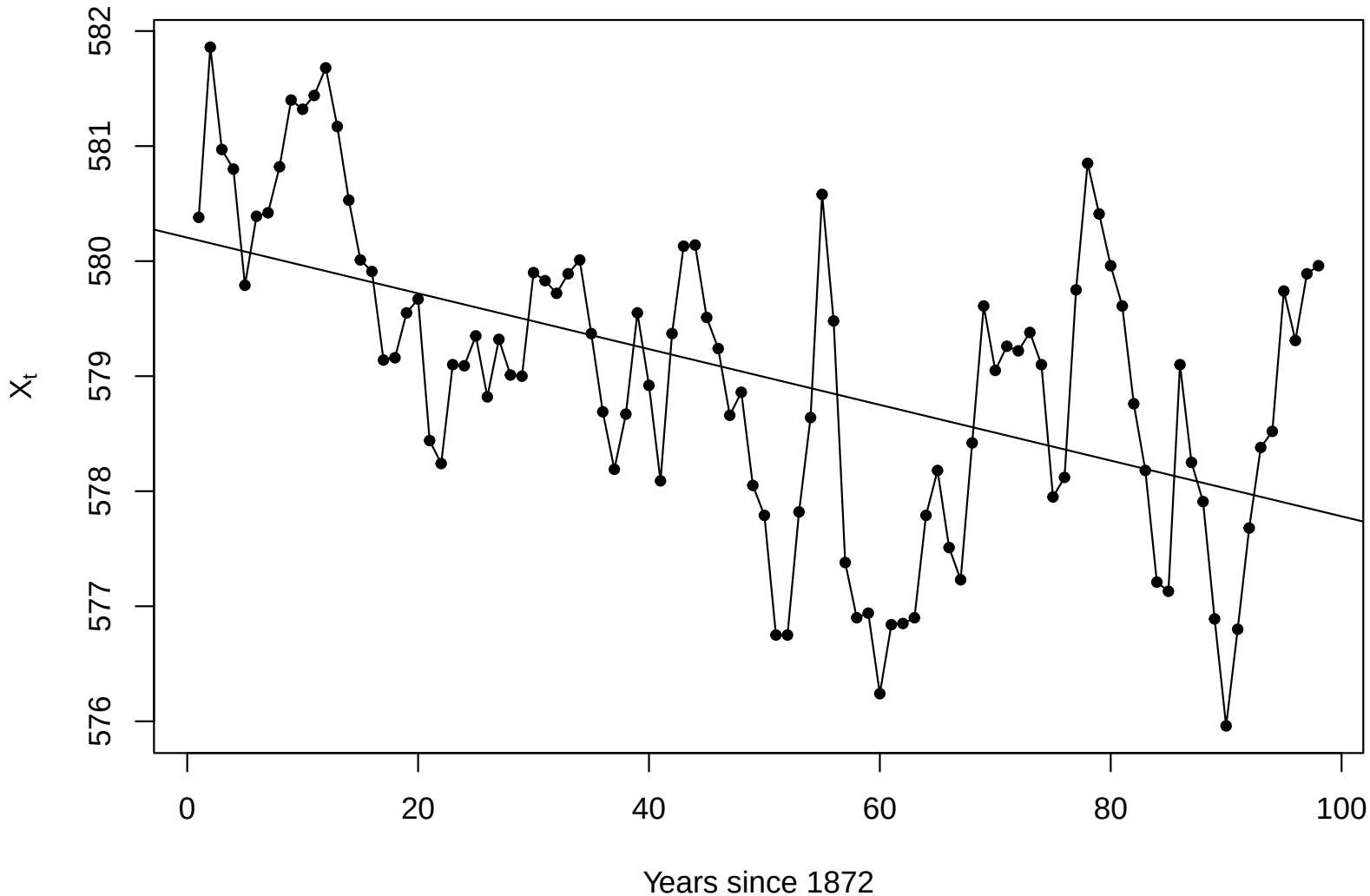
$$\hat{e}_{99}$$

will also be positive. That is, conditional on the available data, the prediction for the next data point X_{99} will be different if the residuals apparently exhibit correlation.

R Code

```r
Huron<-scan("https://www.math.mcgill.ca//dstephens//545//Data//huron.txt")
Huron<-Huron+570
n.H<-length(Huron); t.H<-c(1:n.H); x<-c(1875:1972)
par(mar=c(4,4,2,2))
plot(t.H,Huron,type="l",
     xlab="Years since 1872",ylab=expression(X[t]))
points(t.H,Huron,pch=19,cex=0.7)
abline(lm(Huron~t.H))
```

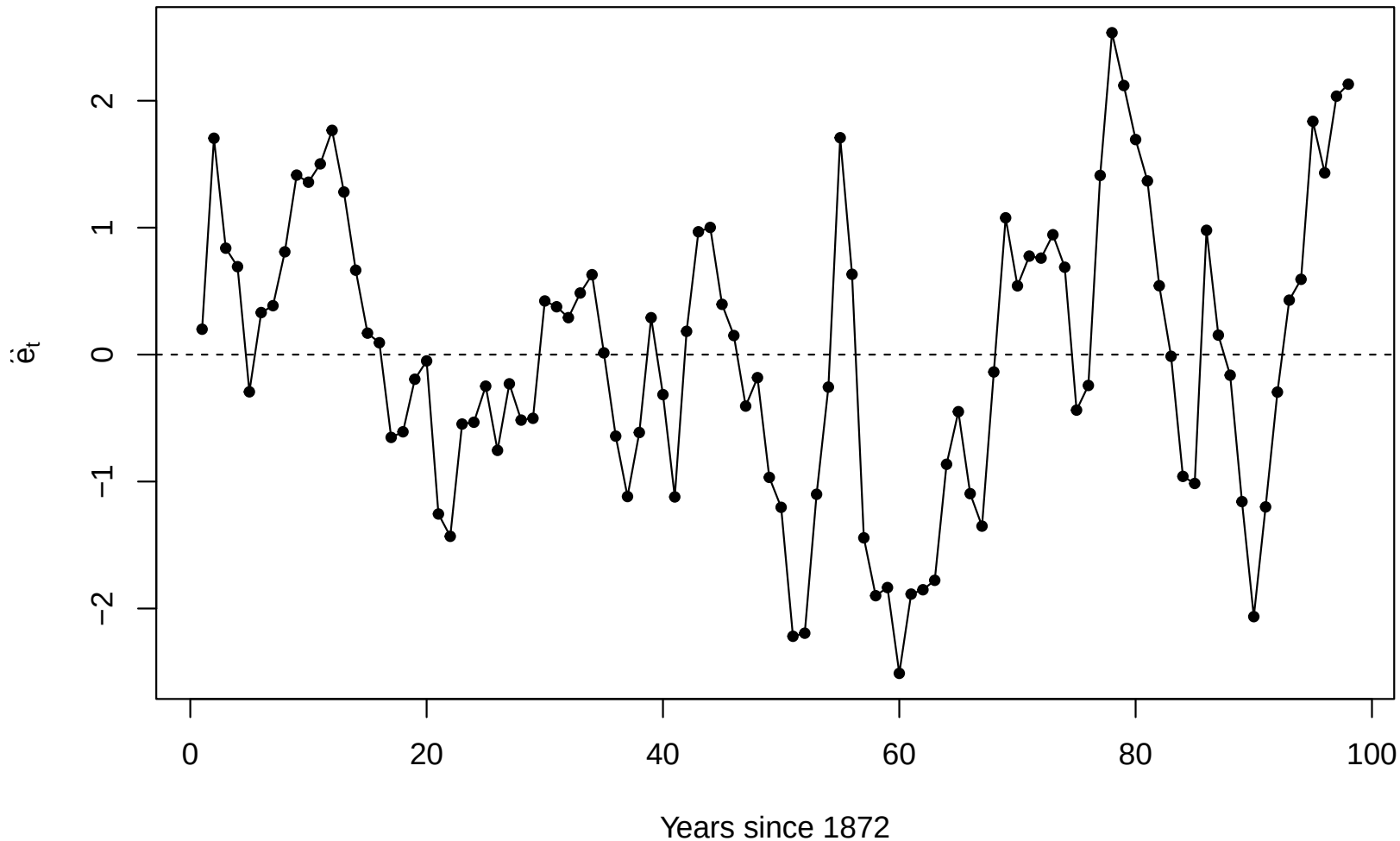
R Code



R Code

```r
a0<-lm(Huron~t.H)$coeff[1]; a1<-lm(Huron~t.H)$coeff[2]
mt.H<-a0+a1*t.H; et.H<-Huron-mt.H
par(mar=c(4,4,2,2))
plot(t.H,et.H,type="l",xlab="Years since 1872",
                       ylab=expression(hat(e)[t]))
points(t.H,et.H,pch=19,cex=0.7); abline(h=0,lty=2)
```

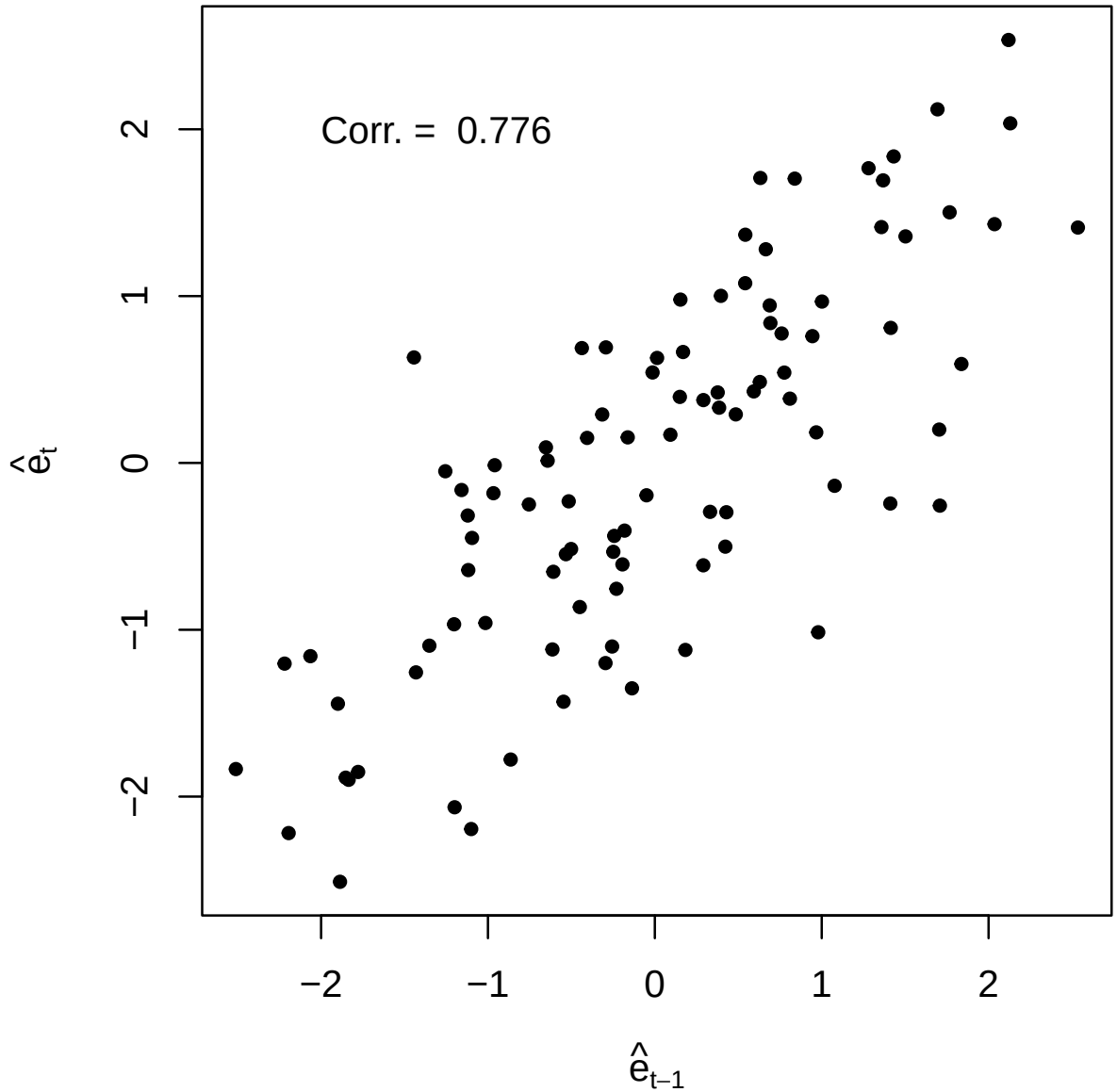
R Code



R Code

```
y1<-et.H[2:98];y2<-et.H[1:97];cor(y1,y2)
par(mar=c(5,4,2,2),pty='s')
plot(y1,y2,xlab=expression(hat(e)[t-1]),
           ylab=expression(hat(e)[t]),pch=19,cex=0.7)
```

R Code



Example: Lake Huron data

Recall that the model

$$X_t = m_t + E_t$$

was fitted using a linear trend $m_t = a_0 + a_1 t$ leaving residuals that were positively correlated (see plot on page 58). We may also model

$$E_t \sim \text{AR}(1),$$

that is, assume

$$E_t = \phi_1 E_{t-1} + Z_t$$

and estimate ϕ_1 using OLS from the fitted residuals e_t. This assumes that $\{Z_t\} \sim WN(0, \sigma_Z^2)$.

Example: Lake Huron data

This procedure yields $\widehat{\phi}_1 = 0.791$. By inspecting the fitted values, we may construct the residual quantities

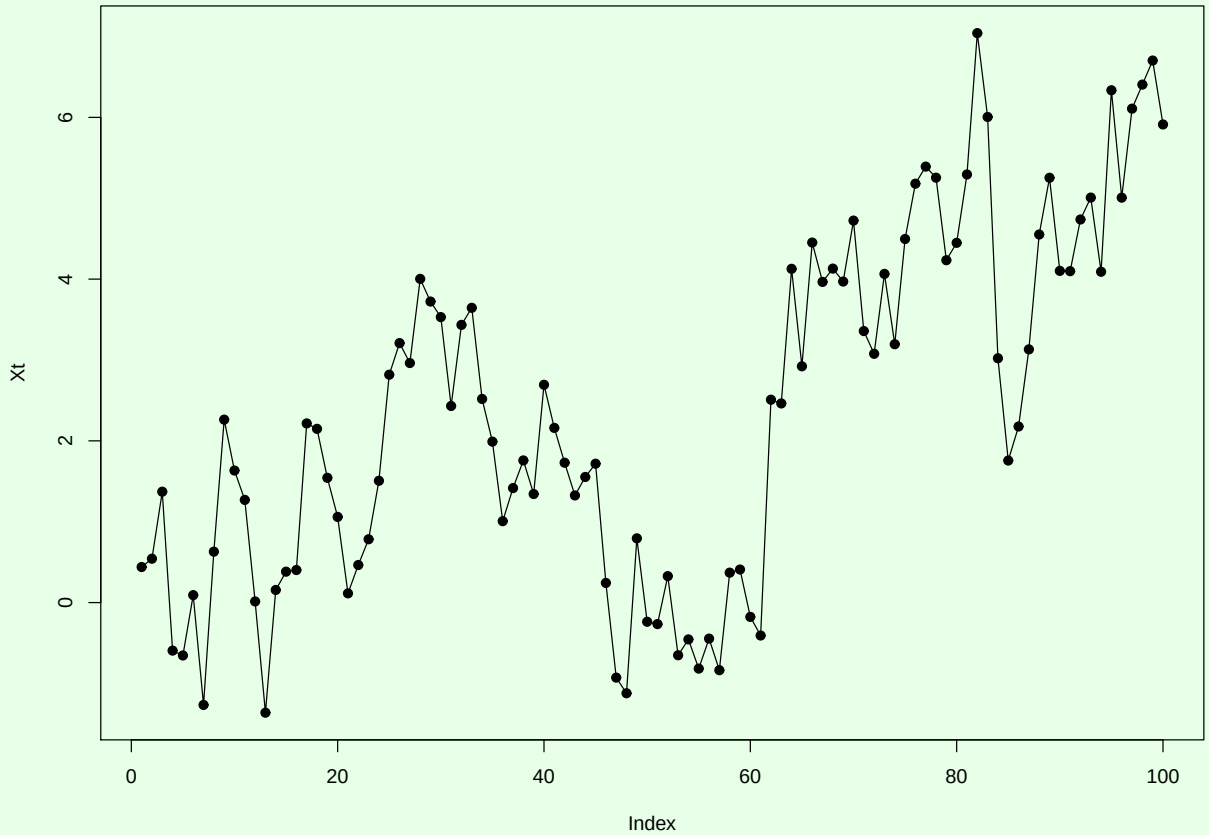
$$\widehat{z}_t = e_t - \widehat{\phi}_1 e_{t-1}$$

we may estimate σ_Z^2 via the sample variance of the $\widehat{z}_t$. This yields $\widehat{\sigma}^2 = 0.502$.

If the AR(1) model were correct, then the residual quantities $\widehat{z}_t$ should resemble a realization of a white noise process; however, they do not - there is still positive correlation between successive $\widehat{z}_t$s.

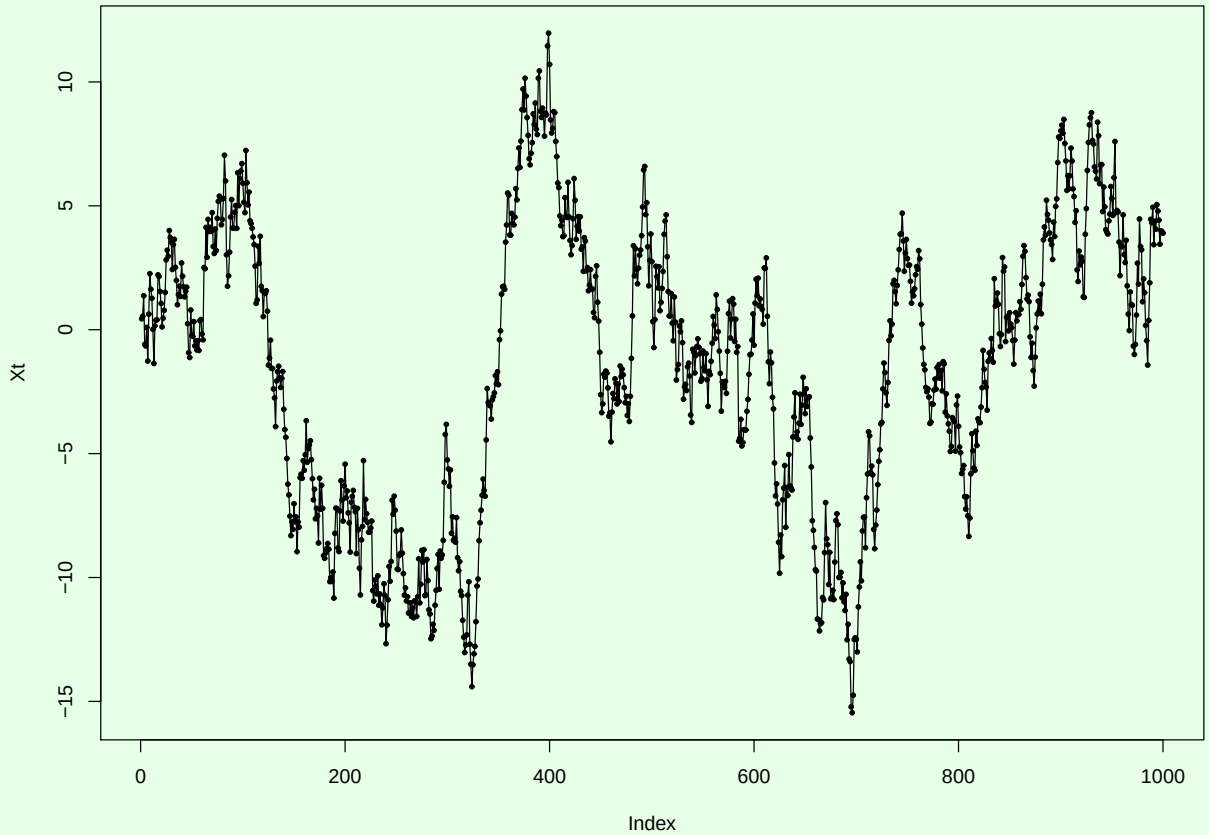
Example: Trend ?

100 observations



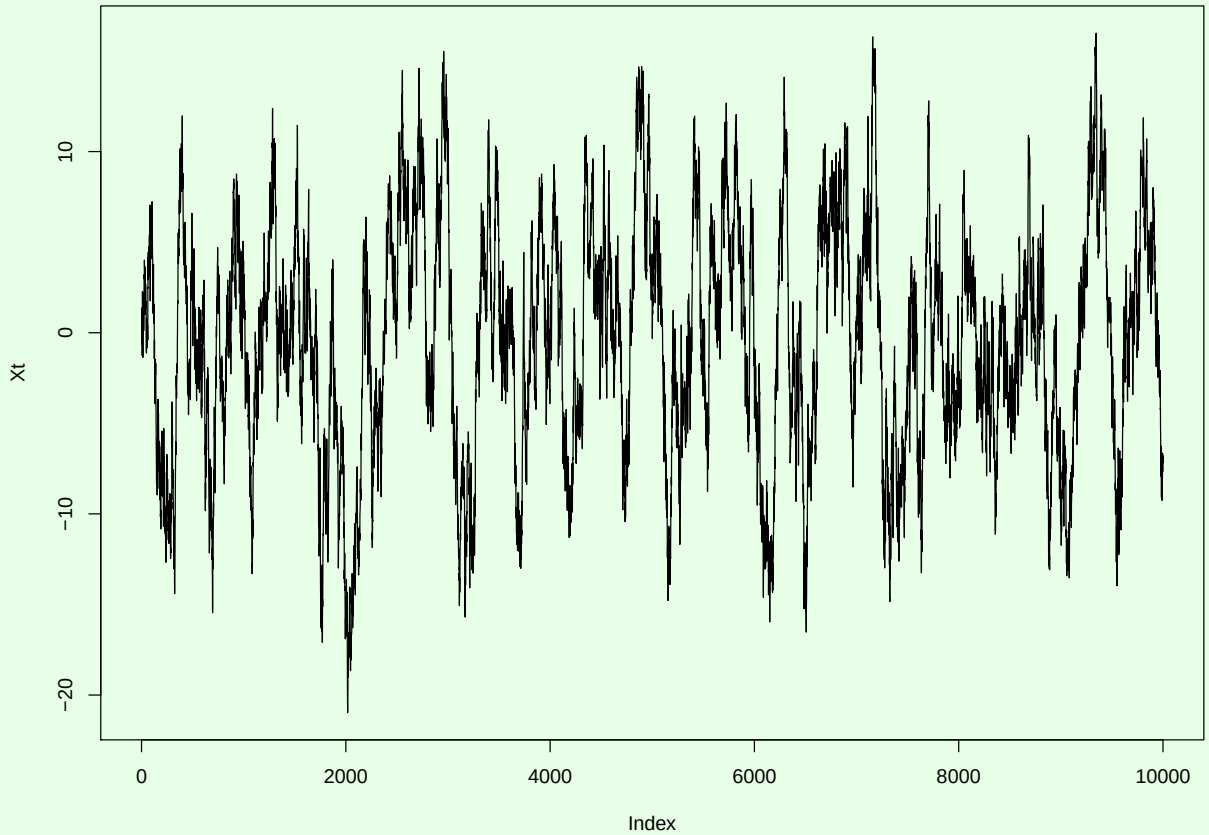
Example: Trend ?

1000 observations



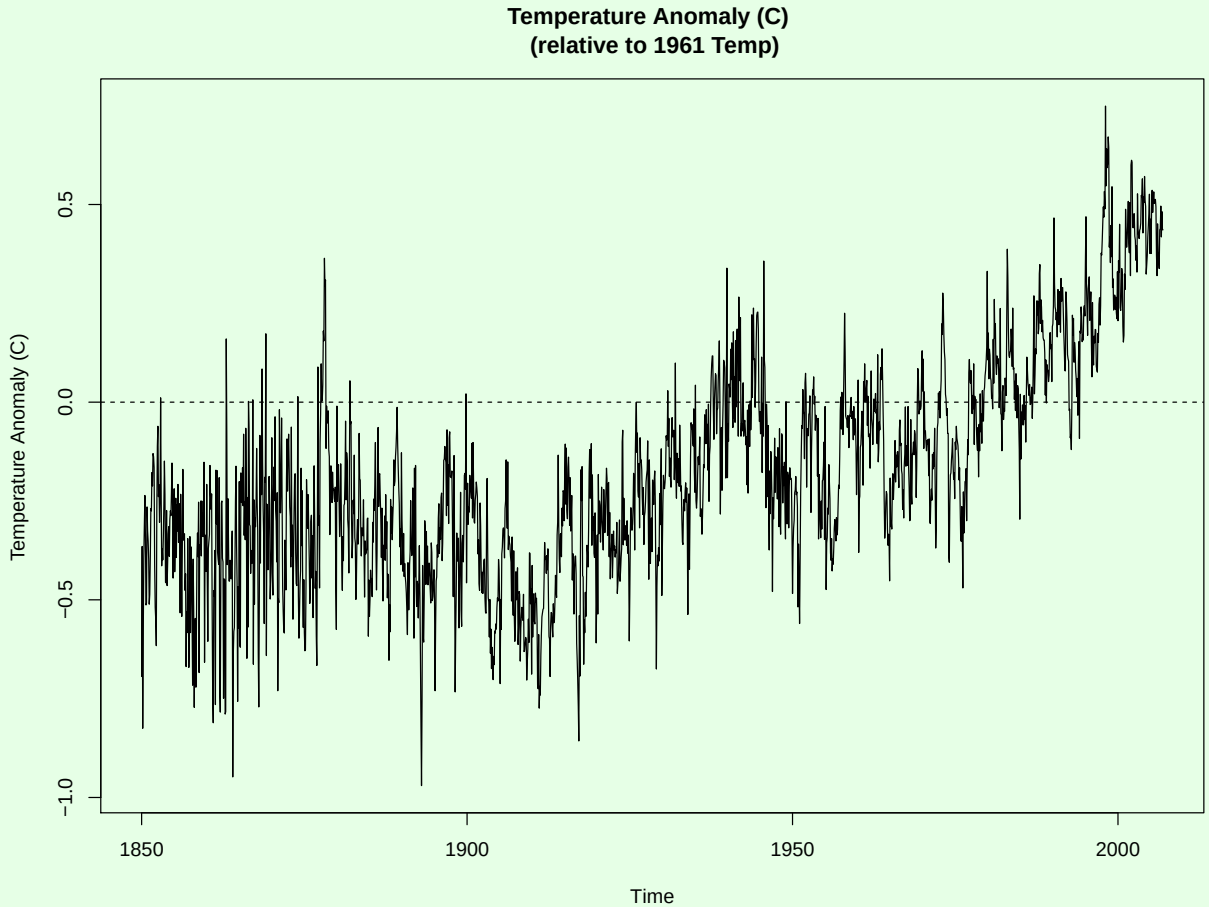
Example: Trend ?

10000 observations



1.5 Trends and Seasonality

Example: Temperature Anomaly: 1850-2006



1.5 Trends and Seasonality

Seasonality: Many time series are influenced by seasonally-varying factors

- calendar
- climate
- economic cycles

1.5 Trends and Seasonality

A deterministic seasonal model can be constructed. Suppose that for each t,

$$X_t = s_t + E_t$$

where s_t is a periodic function with period d say, so that

$$s_{t-d} = s_t$$

for all t.

1.5 Trends and Seasonality

For example, could use a harmonic regression

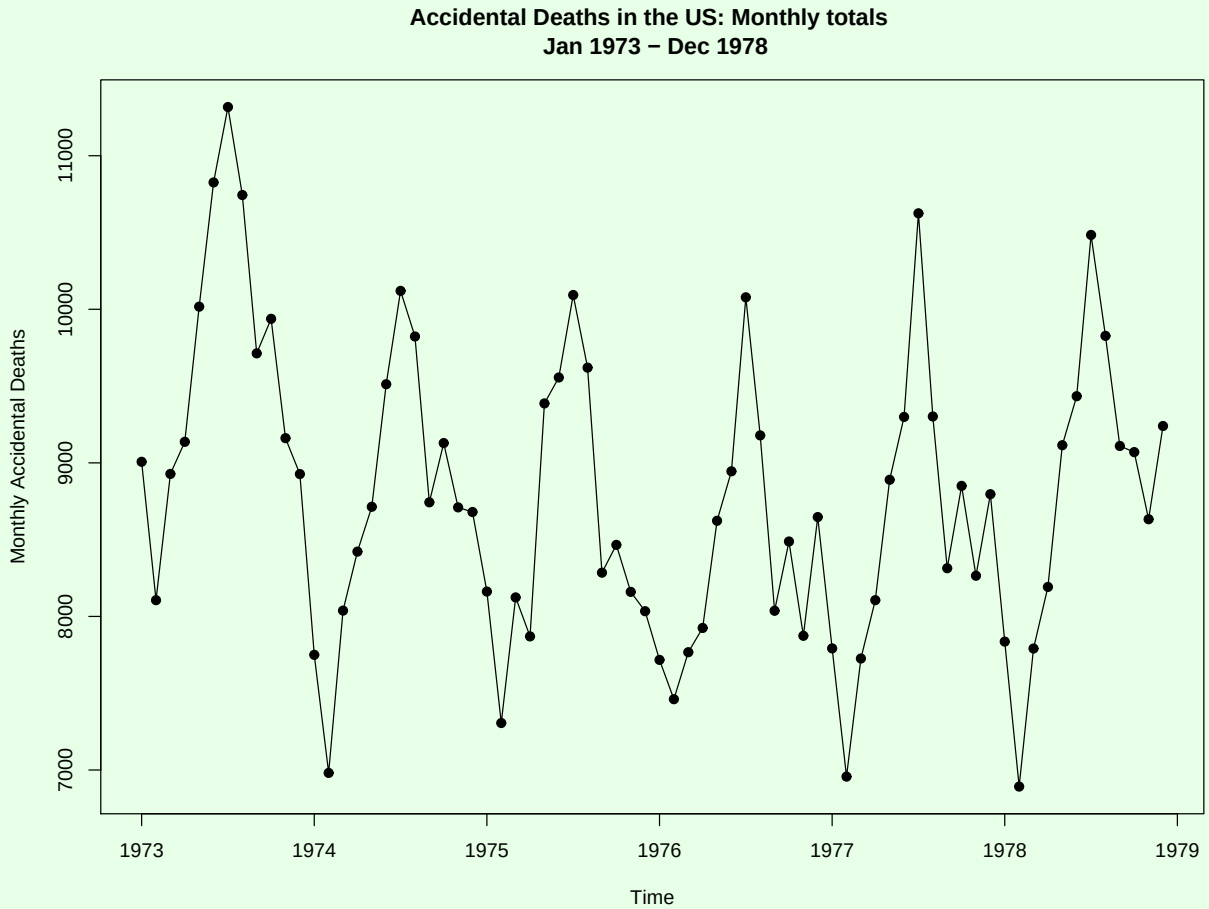
$$s_t = a_0 + \sum_{j=1}^{k} [a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)]$$

where $a_0, a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ are unknown coefficients to be estimated, and

$$\lambda_1, \ldots, \lambda_k$$

are known constants that define the periodic nature of the function.

Example: Accidental Deaths in the US



1.5 Trends and Seasonality

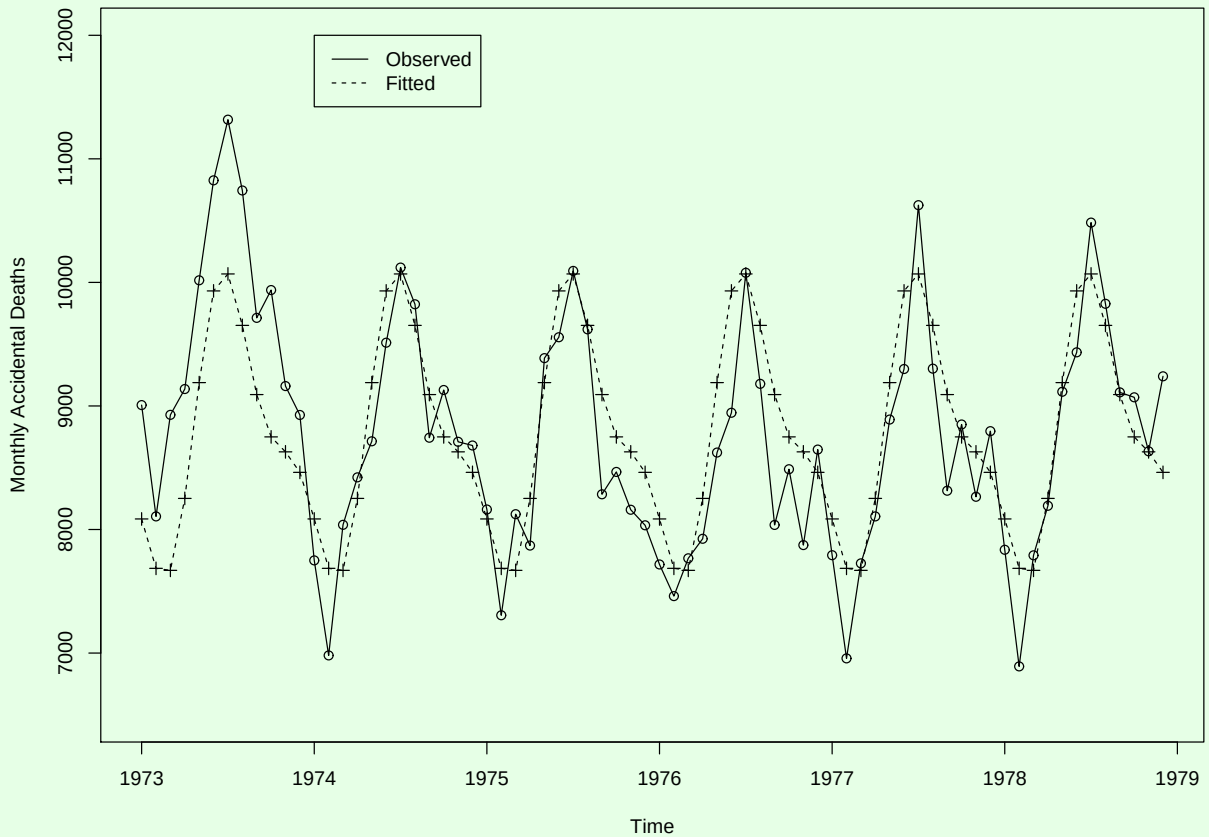
Example: Accidental Deaths in the US

For this data set $T = 72$, for monthly data. To fit a model with $k = 2$

- 12 month cycle
- 6 month cycle

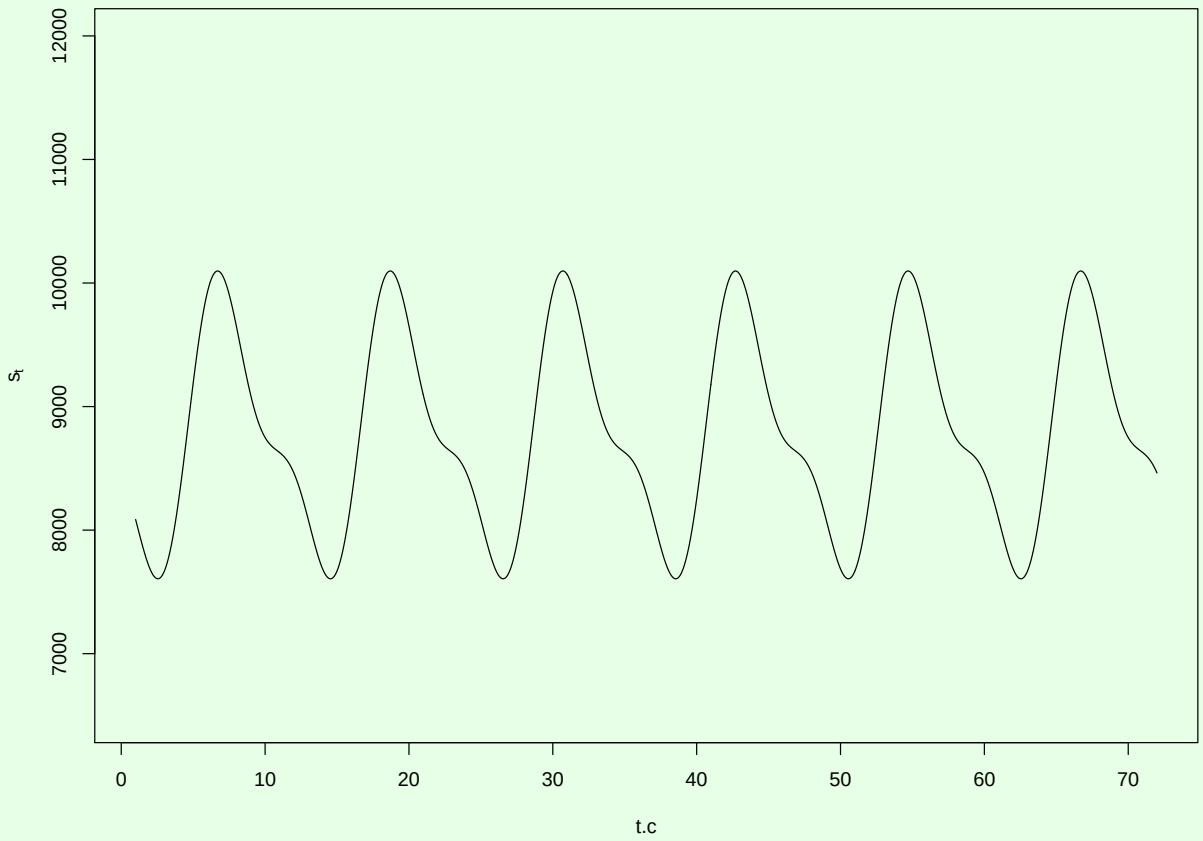
set $\lambda_1 = 2\pi/12$, $\lambda_2 = 2\pi/6$. Then fit using OLS.

Example: Accidental Deaths in the US

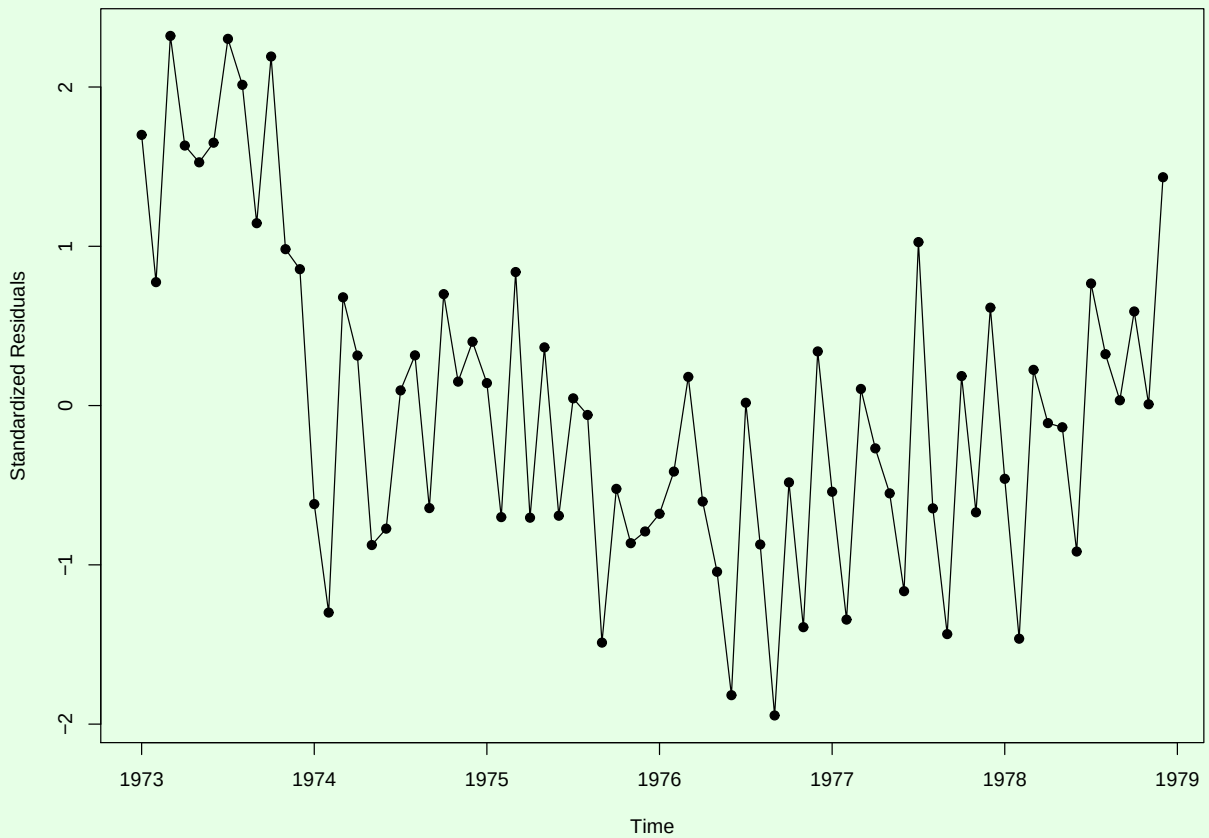


1.5 Trends and Seasonality

Example: Accidental Deaths in the US: Harmonics



Example: Accidental Deaths in the US: Residuals



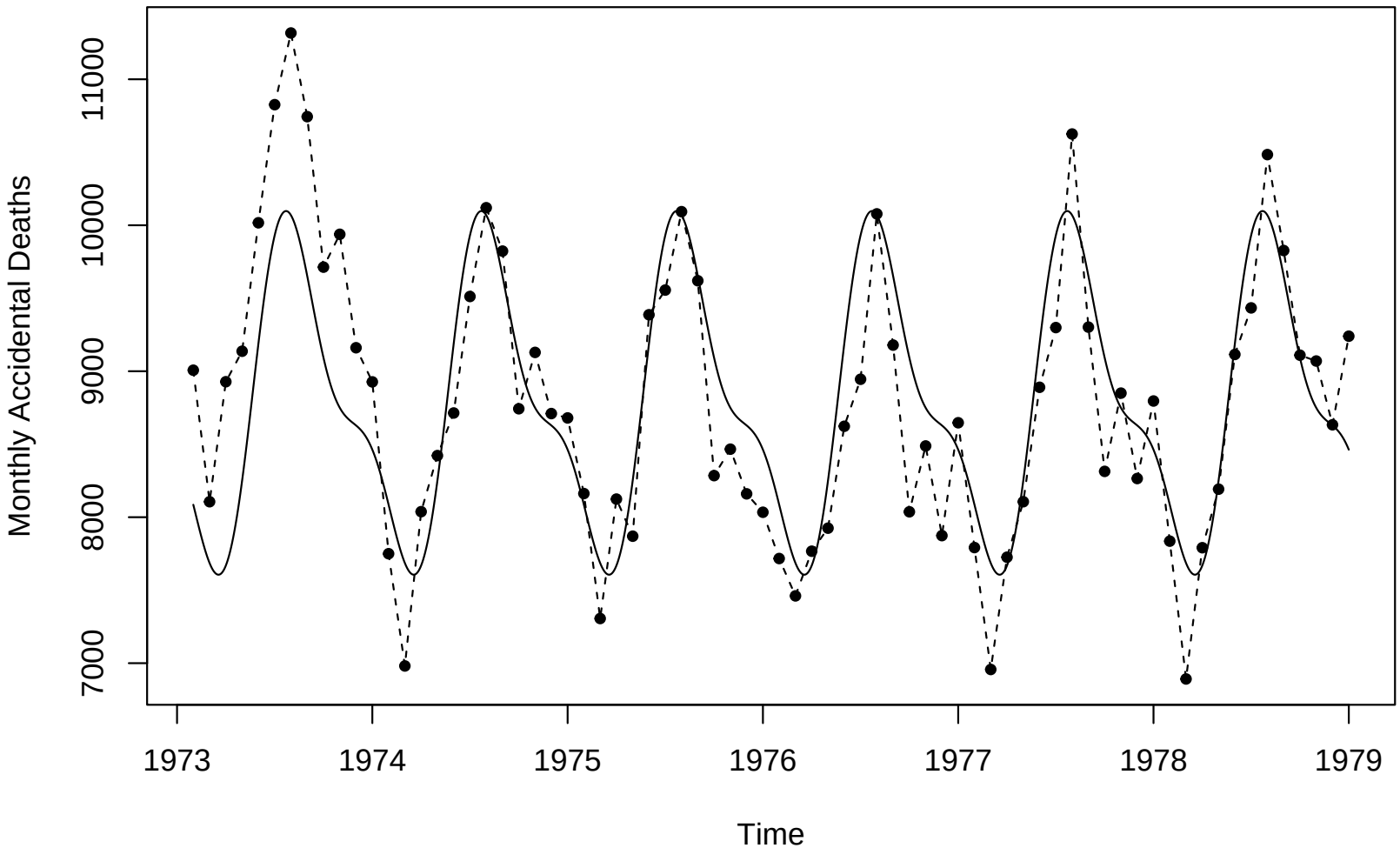
R Code

```
library(datasets)
t.v<-c(1:72)
C1<-cos(2*pi*t.v/12); S1<-sin(2*pi*t.v/12)
C2<-cos(2*pi*t.v/6); S2<-sin(2*pi*t.v/6)
USA.ACC<-data.frame(as.numeric(USAccDeaths),S1,C1,S2,C2)
USA.ACC.fit<-lm(USAccDeaths~S1+C1+S2+C2, data=USA.ACC)$fitted
USA.ACC.coef<-lm(USAccDeaths~S1+C1+S2+C2,data=USA.ACC)$coef
t.c<-seq(from=1,to=72,by=0.01)
C1.c<-cos(2*pi*t.c/12); S1.c<-sin(2*pi*t.c/12)
C2.c<-cos(2*pi*t.c/6); S2.c<-sin(2*pi*t.c/6)
y.c<-USA.ACC.coef[1]+USA.ACC.coef[2]*S1.c+
     USA.ACC.coef[3]*C1.c +USA.ACC.coef[4]*S2.c+
     USA.ACC.coef[5]*C2.c
```

R Code

```
par(mar=c(4,4,2,2))
plot(1:72,as.numeric(USAccDeaths),pch=19,cex=0.7,xaxt='n',
    xlab='Time',ylab='Monthly Accidental Deaths')
lines(1:72,as.numeric(USAccDeaths),lty=2)
lines(t.c,y.c)
axis(1,at=12*c(0:6),c(1973:1979))
```

R Code


Monthly Accidental Deaths
7000
8000
9000
10000
11000
1973
1974
1975
1976
1977
1978
1979
Time

1.6 A Decomposition Strategy

Recap: We seek to decompose an observed series $\{Y_t\}$ as

$$Y_t = m_t + s_t + X_t$$

where

- m_t is a deterministic trend
- s_t is a seasonal component with period d
- X_t is a zero mean stationary process.

It may be necessary to replace the global model for trend (linear, polynomial etc.) by a local model

1.6 A Decomposition Strategy

Smoothing by averaging: Suppose

$$\hat{m}_t = \frac{1}{2q+1} \sum_{i=-q}^{q} x_{t+i}$$

with $x_s = 1$ for $s < 1$ and $s > T$. More generally,

$$\hat{m}_t = \sum_{i=-q}^{q} a_i x_{t+i}$$

for constants $a_{-q}, \ldots, a_q$.

This is a form of low-pass filter.

1.6 A Decomposition Strategy

Exponential Smoothing: Suppose $\hat{m}_1 = x_1$, and for $t \geqslant 2$

$$\hat{m}_t = \alpha x_t + (1 - \alpha)\hat{m}_{t-1}$$

for $0 < \alpha < 1$. It follows that

$$\hat{m}_t = (1 - \alpha)^{t-1} x_1 + \sum_{j=1}^{t-2} \alpha(1 - \alpha)^j x_{t-j} + \alpha x_t$$

If α is near one, there is strong dependence on recent x values.

1.6 A Decomposition Strategy

General Strategy: For series $\{Y_t\}$

$$Y_t = m_t + s_t + X_t$$

where s_t has period d.

1. If d is even, set $q = d/2$, else set $q = (d-1)/2$.
2. Use moving average for detrending respecting the seasonality

$$\widehat{m}_t = \begin{cases} \dfrac{y_{t-q} + y_{t+q}}{2d} + \dfrac{1}{d} \displaystyle\sum_{j=-q+1}^{q-1} y_{t+j} & d \text{ even} \\ \dfrac{1}{d} \displaystyle\sum_{j=-q}^{q} y_{t+j} & d \text{ odd} \end{cases}$$

3. Remove seasonality: for $k = 1, \ldots, d$, compute

$$w_k = \frac{1}{n_d} \sum_j (y_{k+jd} - \hat{m}_{k+jd})$$

where n_d is the number of terms in the sum (essentially the number of cycles in the series). For $k = 1, \ldots, d$, set

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{j=1}^{d} w_j$$

and set $\hat{s}_k = \hat{s}_{k-d}$ for $k = d+1, d+2, \ldots, T$. The deseasonalized data are $y_t^\star = y_t - \hat{s}_t$.

1.6 A Decomposition Strategy

4. Recompute the deseasonalized trend $\widehat{m}_t^\star$ from the deseasonalized data $y_t^\star$ using the procedure in Step 2.
5. Compute the residual series

$$\widehat{x}_t = y_t - \widehat{m}_t^\star - \widehat{s}_t.$$

The resulting process $\widehat{x}_t$ should be a zero mean, but potentially autocorrelated process.

1.7 Differencing

Differencing: Another method for removing non-stationarity is differencing. The lag-1 difference operator, ∇, acts on X_t as follows

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

where B is the backshift operator

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

Note that

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}$$

etc.

1.7 Differencing

Similarly

$$\begin{aligned}
\nabla^2 X_t &= (1-B)^2 X_t \\
&= (1-B)((1-B)X_t) \\
&= (1-B)(X_t - X_{t-1}) \\
&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\
&= X_t - 2X_{t-1} + X_{t-2}
\end{aligned}$$

Note that k times differencing removes a polynomial trend of order k.

1.7 Differencing

Seasonal Differencing: The lag-d difference operator, ∇_d, acts on X_t as follows

$$\nabla_d X_t = (1 - B^d) X_t = X_t - X_{t-d}$$

This operator removes a seasonality with period d.

∇ and ∇_d can be applied consecutively; they commute

$$\nabla \nabla_d X_t = \nabla_d \nabla X_t$$

1.7 Differencing

Elementary algebraic properties:

(i)

$$B^j X_t = X_{t-j} = BX_{t-j+1} = B^2 X_{t-j+2} = \cdots$$

(ii) $\nabla^j X_t = \nabla(\nabla^{j-1} X_t)$. So for example,

$$\begin{aligned}
\nabla^2 X_t &= (1-B)(1-B)X_t \\
&= (1 - 2B + B^2)X_t \\
&= X_t - 2X_{t-1} + X_{t-2} \\
&= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})
\end{aligned}$$

(iii) If $X_t = m_t + Y_t$, applying the first-difference operator to X_t, then

$$\nabla X_t = X_t - X_{t-1} = (m_t - m_{t-1}) + (Y_t - Y_{t-1})$$

so if $m_t = \beta_0 + \beta_1 t$, then $m_t - m_{t-1} = \beta_1$.

That is, ∇ removes a linear trend.

1.7 Differencing

The stochastic term

$$Y_t^* = (Y_t - Y_{t-1})$$

is obtained by differencing.

This is different from our first approach to detrending, which used a linear parametric formulation for m_t.

- If Y_t is white-noise, then Y_t^* is no longer white noise, but MA(1).

1.7 Differencing

(iv) To remove a polynomial trend of order k, one may apply k^{th} order differencing, that is we look at $\nabla^k X_t$.

Note that if $\{Y_t\} \sim \text{WN}(0, \sigma_Y^2)$, then $\nabla^k Y_t$ is not white-noise, but is still stationary.

1.7 Differencing

(v) Commutativity

$$\nabla\nabla_d X_t = (1-B)(1-B^d)X_t = (1-B^d)(1-B)X_t = \nabla_d \nabla X_t$$

(vi) If

$$X_t = m_t + s_t + Y_t,$$

then

$$\nabla^d X_t = (m_t - m_{t-d}) + (s_t - s_{t-d}) + (Y_t - Y_{t-d})$$

and

$$s_t - s_{t-d} = 0.$$

That is, ∇_d removes a seasonality with period d.

Chapter 2
Stationary processes

2.1 Stationary Processes

The objective of the previous section was to remove (systematic) deterministic components from an observed series to leave only the stochastic part. In this section, we study the basic properties of stationary processes: such processes are inherently stable (in the long run), and form natural models for the stochastic component of observed series.

In the main, we will focus on weakly stationary processes, where the mean and covariance structure is stable over time; in general we will not make distributional assumptions.

Initially, we focus on forecasting, and demonstrate the utility of utilizing the covariance structure of the process.

2.2 Properties of Stationary Processes

Consider first a stationary Gaussian process $\{X_t\}$, where for all t and n,

$$(X_{t+1}, \ldots, X_{t+n})$$

is jointly normally distributed, with $\mathbb{E}[X_t] = \mu$, $\text{Var}[X_t] = \sigma_X^2$, and for $s > t$.

$$\text{Cov}[X_t, X_s] = \gamma_X(s-t) \qquad \text{Corr}[X_t, X_s] = \rho_X(s-t) = \frac{\gamma_X(s-t)}{\sigma_X^2}$$

Then it follows that

$$X_{n+h}|X_n = x_n \sim \mathcal{N}(m_h(x_n), v_h(x_n))$$

where

$$\begin{aligned} m_h(x) &= \mu + \rho_X(h)(x - \mu) \\ v_h(x) &= \sigma_X^2(1 - \rho_X^2(h)) \end{aligned}$$

2.2 Properties of Stationary Processes

To define “optimal" forecasting, we need to define a criterion via which to assess the quality of predictions.

We use a minimum mean-square error (minimum MSE) criterion, and attempt to to make the prediction $\widehat{X}_{n+h} = \widehat{x}_{n+h}$ such that

$$\mathbb{E}\left[\left(X_{n+h} - \widehat{x}_{n+h}\right)^2 \middle| X_n = x_n\right]$$

is minimized. Here the expectation is taken with respect to the conditional distribution $X_{n+h} \middle| X_n = x_n$.

This minimization can be carried out analytically to yield

$$\widehat{x}_{n+h} = \mathbb{E}[X_{n+h}|X_n = x_n] = m_h(x) = \mu + \rho_X(h)(x_n - \mu)$$

that is, the best prediction (measured in minimum prediction MSE terms) is the conditional expectation. The prediction MSE corresponding to this prediction is

$$v_h(x) = \sigma_X^2(1 - \rho_X^2(h))$$

In this case, we could deduce the appropriate form for the optimal predictor from properties of the Gaussian process.

2.2 Properties of Stationary Processes

Best Linear Prediction: For a non-Gaussian process, we seek the again seek the minimum MSE predictor. We restrict attention to linear predictors, that is, predictors of the form

$$\widehat{X}_{n+h} = l_h(X_n) = aX_n + b$$

for some a and b. Under this restriction, and stationarity, it again transpires that the optimal predictor takes the form

$$\widehat{x}_{n+h} = \mathbb{E}[X_{n+h}|X_n = x_n] = m_h(x) = \mu + \rho_X(h)(x_n - \mu)$$

with the same prediction MSE as before.

2.2 Properties of Stationary Processes

Non-negative definite functions: A scalar function g is non-negative definite if, for all n, and vectors $a = (a_1, \ldots, a_n)$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i g(i-j) a_j \geqslant 0.$$

If the summation is strictly greater than zero, the function is termed positive definite.

If G is the $n \times n$ matrix formed as $G = [g(i-j)]_{ij}$, then the above definition becomes that

$$a^\top G a \geqslant 0.$$

and G is a non-negative/positive definite matrix.

2.2 Properties of Stationary Processes

Theorem:

A function g is the autocovariance function (ACVF) for a stationary stochastic process if and only if g is **even**, that is

$$g(h) = g(-h) \qquad \text{for all } h \in \mathbb{Z}$$

and g is **non-negative definite**.

Note: by elementary properties of covariance, for all h,

$$\begin{aligned} \gamma(0) &\geqslant 0 \\ |\gamma(h)| &\leqslant \gamma(0) \\ \gamma(-h) &= \gamma(h) \end{aligned}$$

Characterizing the ACVF

Proof. Suppose that g is the ACVF of a stationary stochastic process, $\{X_t\}$ say. By definition of covariance, g is an even function, as

$$g(-h) = \mathbb{E}[X_t X_{t-h}] = \mathbb{E}[X_{t+h} X_t] = \mathbb{E}[X_t X_{t+h}] = g(h).$$

Let $a \in \mathbb{R}^n$ be an $n \times 1$ vector, and let $X_{n:1} = (x_n, \ldots, x_1)^\top$. Then $a^\top X_{n:1}$ is a scalar random variable, and

$$\mathrm{Var}[a^\top X_{n:1}] = a^\top \mathrm{Var}[X_{n:1}] a = \sum_{i=1}^{n} \sum_{j=1} a_i g(i-j) a_j \geqslant 0$$

so g is non-negative definite.

Characterizing the ACVF

Now suppose that g is non-negative definite and even. For each $n \geqslant 1$, define the $n \times n$ matrix Σ by

$$[\Sigma]_{ij} = g(i-j)$$

and consider the multivariate Normal distribution with mean zero and variance-covariance matrix Σ. This distribution is the finite dimensional distribution arising from a stationary Gaussian process with the specified covariance structure. $\square$

Note that an autocorrelation function has the same properties, except that $\rho(0) = 1$.

2.2 Properties of Stationary Processes

Example: Simple harmonic process

Suppose U and V are zero mean and uncorrelated random variables with variance 1. Let $\{X_t\}$ be defined by

$$X_t = U\cos(\omega t) + V\sin(\omega t)$$

where ω is a fixed constant.

U and V define random amplitudes.

Example: Simple harmonic process

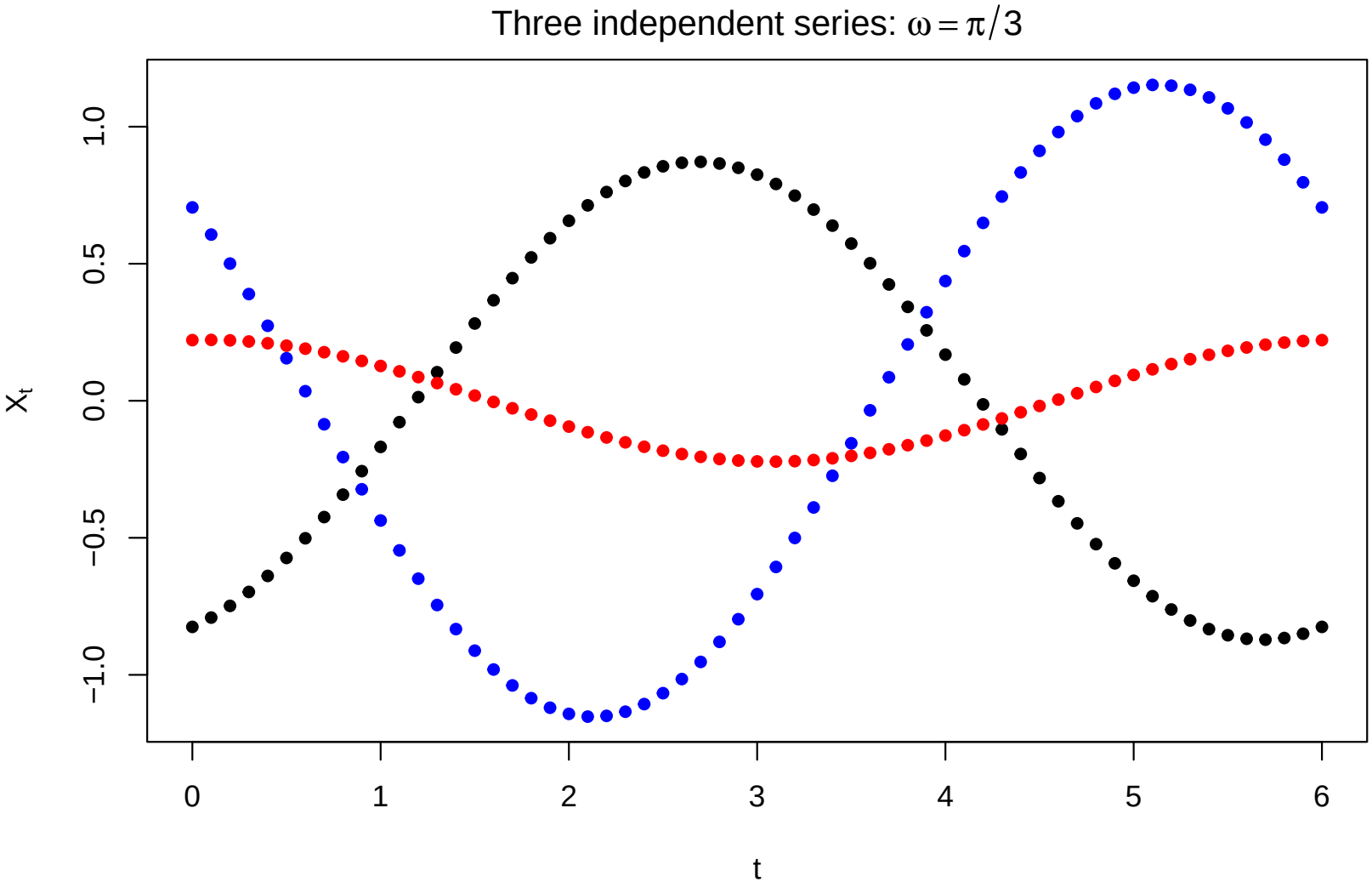
Then $\mathbb{E}[X_t] = 0$ for all t, and

$$\begin{aligned}
\operatorname{Cov}[X_t, X_{t+h}] &= \mathbb{E}[X_t X_{t+h}] \\
&= \mathbb{E}\left[(U\cos(\omega t) + V\sin(\omega t))\right. \\
&\qquad \left.(U\cos(\omega(t+h)) + V\sin(\omega(t+h)))\right] \\
&= \mathbb{E}[U^2]\cos(\omega t)\cos(\omega(t+h)) + \\
&\qquad \mathbb{E}[V^2]\sin(\omega t)\sin(\omega(t+h)) \\
&= \cos(\omega t)\cos(\omega(t+h)) + \sin(\omega t)\sin(\omega(t+h)) \\
&= \cos(\omega h) = \cos(\omega(-h))
\end{aligned}$$

2.2 Properties of Stationary Processes

```
omega<-pi/3
set.seed(2343)
xt<-seq(0,6,by=0.1)
y1<-rnorm(1)*cos(omega*xt)+rnorm(1)*sin(omega*xt)
y2<-rnorm(1)*cos(omega*xt)+rnorm(1)*sin(omega*xt)
y3<-rnorm(1)*cos(omega*xt)+rnorm(1)*sin(omega*xt)
ymin<-min(y1,y2,y3)
ymax<-max(y1,y2,y3)
par(mar=c(4,4,2,2))
plot(xt,y1,xlab='t',ylab=expression(X[t]),
        main=expression(paste('Three independent series: ',omega==pi/3)),
        ylim=range(ymin,ymax),pch=19,cex=0.8)
points(xt,y2,col='red',pch=19,cex=0.8)
points(xt,y3,col='blue',pch=19,cex=0.8)
```

2.2 Properties of Stationary Processes



2.3 Construction of stationary processes

Let $\{Z_t\}$ be an IID process, and let $\{X_t\}$ be defined by

$$X_t = f(Z_t, Z_{t-1}, \ldots, Z_{t-q})$$

for some $q \geqslant 0$. $\{X_t\}$ is a filtered version of $\{Z_t\}$.

- $\{X_t\}$ is strictly stationary as $\{Z_t\}$ is strictly stationary.
- X_t and X_s are independent if $|t-s| > q$; $\{X_t\}$ is termed q-dependent, and $\gamma_X(h) = 0$ if $|h| > q$.

2.3 Construction of stationary processes

Example: Moving average processes

Let $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$ be a white noise process, and let $\{X_t\}$ be defined by

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

for some $q \geqslant 0$, where $\theta_1, \theta_2, \ldots, \theta_q$ are real-valued constants.

$\{X_t\}$ is zero mean, stationary and q-dependent; it is termed a moving average process of order q (MA(q)).

2.4 Linear Processes

$\{X_t\}$ is a linear process if for all t

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where

- $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$
- $\{\psi_j\}$ is a real sequence with

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

2.4 Linear Processes

Using the backshift operator notation

$$X_t = \psi(B)Z_t$$

where

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j.$$

The process is termed non-anticipating or causal if $\psi_j = 0$ for all $j < 0$, so that

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

2.4 Linear Processes

Note that for the linear process to be sensible we need the sum

$$\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

to converge in some appropriate fashion; in this context we require mean-square convergence, denoted

$$X_t \overset{\text{m.s.}}{=} \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

2.4 Linear Processes

We require for the definition

$$\lim_{n \longrightarrow \infty} \mathbb{E}\left[\left(X_t - \sum_{j=-n}^{n} \psi_j Z_{t-j}\right)^2\right] = 0$$

For finite $n \geqslant 0$,

$$\left(\sum_{j=-n}^{n} \psi_j Z_{t-j}\right)^2 = \sum_{j=-n}^{n} \psi_j^2 Z_{t-j}^2 + \sum_{j=-n}^{n} \sum_{k \neq j} \psi_j \psi_k Z_{t-j} Z_{t-k}$$

Taking expectations, as $\{Z_t\}$ is a white noise process so that $\mathbb{E}[Z_{t-j}Z_{t-k}] = 0$ if $j \neq k$

$$\mathbb{E}\left[\left(\sum_{j=-n}^{n} \psi_j Z_{t-j}\right)^2\right] = \sigma_Z^2 \sum_{j=-n}^{n} \psi_j^2.$$

2.4 Linear Processes

Now

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad \Longrightarrow \quad \sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$$

so therefore the partial sums (with limits $\pm n$) are also finite for all n.

If

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

then we term the sequence $\{\psi_j\}$ absolutely summable.

2.4 Linear Processes

Note also that by the Cauchy-Schwarz inequality

$$\mathbb{E}[|Z_t|] \leqslant \sqrt{\mathbb{E}[Z_t^2]} = \sigma,$$

therefore

$$\mathbb{E}[|X_t|] \leqslant \sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}[|Z_{t-j}|] \leqslant \sigma \sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

so that

$$\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

is convergent (almost surely).

2.4 Linear Processes

The operator $\psi(B)$ is termed a linear filter; it acts on the stationary (white noise) process $\{Z_t\}$ to produce the stationary (autocorrelated) process $\{X_t\}$.

The operator can also be applied to any stationary process.

2.4 Linear Processes

Suppose that $\{Y_t\}$ is a weakly stationary, zero mean process with ACVF γ_Y, and $\psi(B)$ is the above linear filter with absolutely summable coefficients. Then the process

$$X_t = \psi(B)Y_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}$$

is also weakly stationary with mean zero. The ACVF of $\{X_t\}$ is

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(h + j - k).$$

2.4 Linear Processes

If $\{Y_t\}$ is a white noise process, so that $\{X_t\}$ is a linear process, then

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma_Y^2.$$

To see this, note that

$$\mathbb{E}[X_t] \leqslant \mathbb{E}[|X_t|] \leqslant \sum_{j=-\infty}^{\infty} |\psi_j| |\mathbb{E}[Y_t]| < \infty$$

so that $\mathbb{E}[X_t] = 0$ for all t. Then

$$\mathbb{E}[X_t X_{t+h}] = \mathbb{E}\left[\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \right) \left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t+h-k} \right) \right]$$

Multiplying out yields

$$\begin{aligned}\mathbb{E}[X_t X_{t+h}] &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}\left[Y_{t-j} Y_{t+h-k}\right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(j+h-k).\end{aligned}$$

If $\{Y_t\}$ is white noise, then $\gamma_Y(h) = \sigma_Y^2$ if $h = 0$ and zero otherwise, in which case

$$\gamma_X(h) = \mathbb{E}[X_t X_{t+h}] = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma_Y^2.$$

2.4 Linear Processes

If $\alpha(B)$ and $\beta(B)$ are two linear filters, each of which has absolutely summable coefficients, and

$$\psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B)$$

then $\psi(B)$ is also a linear filter, with

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \beta_k \alpha_{j-k}$$

and if $\{Y_t\}$ is a stationary process, then

$$X_t = \psi(B)Y_t = \alpha(B)\beta(B)Y_t$$

defines a stationary process $\{X_t\}$. Note that $W_t = \beta(B)Y_t$ is also stationary, and $X_t = \alpha(B)W_t$.

Example: AR(1) process

Let $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, $|\phi| < 1$, and

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

with $\mathbb{E}[Z_s X_t] = 0$ for $s > t$. That is, $\{X_t\}$ is the solution of the equation

$$X_t - \phi X_{t-1} = Z_t$$

or

$$(1 - \phi B) X_t = Z_t.$$

Example: AR(1) process

Consider the linear process

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

Note that

$$\sum_{j=0}^{\infty} |\phi|^j = \frac{1}{1-|\phi|} < \infty$$

hence the conditions of absolute summability hold.

Example: AR(1) process

This is a solution to the equation above as

$$
\begin{aligned}
X_t &= \phi X_{t-1} + Z_t \\
&= \phi\left(\phi X_{t-2} + Z_{t-1}\right) + Z_t && \text{recursion} \\
&\vdots \\
&= \lim_{n \longrightarrow \infty} \left\{ \phi^n X_{t-n} + \sum_{j=0}^{n-1} \phi^j Z_{t-j} \right\} \\
&= \sum_{j=0}^{\infty} \phi^j Z_{t-j}
\end{aligned}
$$

Example: AR(1) process

Also, for $h \geqslant 0$,

$$\gamma_X(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma_Z^2 = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma_Z^2 = \sigma_Z^2 \sum_{j=0}^{\infty} \phi^{2j+h}$$

that is

$$\gamma_X(h) = \frac{\sigma_Z^2 \phi^h}{1 - \phi^2}$$

Example: AR(1) process

If $\{Y_t\}$ is another stationary solution to the AR(1) equation, then, for any $k \geqslant 0$

$$\begin{aligned}
\mathbb{E}\left[\left(Y_t - \sum_{j=0}^{k} \phi^j Z_{t-j}\right)^2\right] &= \mathbb{E}\left[\left(\sum_{j=k+1}^{\infty} \phi^j Z_{t-j}\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{l=0}^{\infty} \phi^{l+k+1} Z_{t-k-1-l}\right)^2\right] \\
&= \phi^{2(k+1)} \mathbb{E}\left[\left(\sum_{l=0}^{\infty} \phi^{l} Z_{t-k-1-l}\right)^2\right] \\
&= \phi^{2(k+1)} \mathbb{E}\left[\left(Y_{t-k-1}\right)^2\right]
\end{aligned}$$

Example: AR(1) process

The right hand side of this equation converges to zero as $k \longrightarrow \infty$, as $|\phi| < 1$, and $\mathbb{E}\left[Y_{t-k-1}^2\right] < \infty$.

Therefore Y_t and X_t are equal in mean square

$$Y_t \overset{\text{m.s.}}{=} X_t$$

and therefore $\{X_t\}$ is essentially the unique stationary solution to the AR(1) equation

$$X_t - \phi X_{t-1} = Z_t$$

with $|\phi| < 1$ and $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$.

Example: AR(1) process

Therefore for any stationary AR process there is an essentially unique MA(∞) representation, with $\psi_j = \phi^j$ for $j \geqslant 0$, and $\psi_j = 0$ for $j < 0$.

Example: Non-stationary AR(1) process

If $|\phi| = 1$, say

$$X_t = X_{t-1} + Z_t \qquad \text{or} \qquad X_t = -X_{t-1} + Z_t$$

then X_t is not stationary. Clearly

$$\text{Var}[X_t] = \text{Var}[\pm X_{t-1}] + \text{Var}[Z_t] > \text{Var}[X_{t-1}]$$

for all t.

Non-causal AR(1): In the AR(1) formulation, suppose $|\phi| > 1$, that is

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

so that

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

or

$$X_{t-1} = \phi_* X_t + Z_t^*$$

where $\{Z_t^*\} \sim \text{WN}(0, \sigma_Z^2/\phi^2)$ is defined by

$$Z_t^* = -\frac{1}{\phi} Z_t$$

and $\phi_* = 1/\phi$, so that $|\phi_*| < 1$.

Therefore, by considering a reverse time formulation, we can construct a stationary solution to the AR(1) equation. It takes the linear process form

$$X_t = \sum_{j=0}^{\infty} \phi_*^j Z_{t+1+j}^* = -\sum_{j=0}^{\infty} \phi^{-(j+1)} Z_{t+1+j}$$

that is, X_t is defined in terms of future values of $\{Z_t\}$.

The ACVF can be computed by the usual methods. Such a process is termed **non-causal**.

2.4 Linear Processes

Note: we have for the AR(1) process that

$$(1 - \phi B)X_t = Z_t \qquad \therefore \qquad X_t = (1 - \phi B)^{-1} Z_t$$

and by a geometric series expansion,

$$X_t = \left\{ \sum_{j=0}^{\infty} \phi^j B^j \right\} Z_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

as before.

2.5 Autoregressive Processes

Suppose that $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and that the equation

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$$

is used to define (implicitly) $\{X_t\}$. This definition does not ensure (or even suggest) that the solution can be found, or is unique or stationary.

This process is termed the autoregressive process of order p (AR(p)). We have that

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

Example: AR(2) process

If $p = 2$ we have

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t \qquad \text{or} \qquad \Phi(B) X_t = Z_t.$$

where $\Phi(B) = (1 - \phi_1 B - \phi_2 B^2)$, and the following linear process representation obtained by noting that

$$X_t = (1 - \phi_1 B - \phi_2 B^2)^{-1} Z_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where ψ_j is the coefficient of B^j in the series expansion of

$$(1 - \phi_1 B - \phi_2 B^2)^{-1}$$

Example: AR(2) process

Consider the factorization

$$\Phi(z) = (1 - \phi_1 z - \phi_2 z^2) = (1 - \xi_1 z)(1 - \xi_2 z).$$

where z is an arbitrary complex number, and (ξ_1, ξ_2) are a complex conjugate pair. In general,

$$(1 - \xi z)^{-1}$$

has radius of convergence $|\xi| < 1$ when $|z| = 1$.

2.5 Autoregressive Processes

Example: AR(2) process

(ξ_1, ξ_2) are the reciprocal roots (that is, the reciprocals of the roots) of the equation

$$\Phi(z) = 0$$

then the series expansion of

$$\{(1 - \xi_1 z)(1 - \xi_2 z)\}^{-1}$$

is convergent for $|z| = 1$ if and only if

$$|\xi_1| = |\xi_2| < 1.$$

2.5 Autoregressive Processes

Example: AR(2) process

We have that

$$\Phi(B)X_t = Z_t \qquad \therefore \qquad X_t = \{\Phi(B)\}^{-1} Z_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

where ψ_j is the coefficient of z^j in the series expansion of

$$\{\Phi(z)\}^{-1} = \{(1-\xi_1 z)(1-\xi_2 z)\}^{-1} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \xi_1^j \xi_2^k z^{j+k}$$

that is

$$\psi_j = \sum_{k=0}^{j} \xi_1^k \xi_2^{j-k}$$

Example: AR(2) process

Note that

$$|\psi_j| \leqslant \sum_{k=0}^{j} |\xi_1|^k |\xi_2|^{j-k} = (j+1)M^j$$

where $M = \max\{|\xi_1|, |\xi_2|\} < 1$. Therefore

$$\sum_{j=0}^{\infty} |\psi_j| \leqslant \sum_{j=0}^{\infty} (j+1)M^j = \frac{1}{(1-M)^2} < \infty.$$

Hence $\{X_t\}$ is a linear process with absolutely summable coefficients.

2.5 Autoregressive Processes

The extension to the AR(p) case is straightforward: write

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = \prod_{j=1}^{p} (1 - \xi_j z)$$

and by an identical argument to above, to ensure stationarity, we require that the reciprocal roots

$$\xi_1, \xi_2, \ldots, \xi_p$$

all satisfy $|\xi_j| < 1$.

Equivalently, if we let $\eta_j = \xi_j^{-1}$ denote the roots of $\Phi(z) = 0$, then we have

$$\Phi(z) = \prod_{j=1}^{p} (\eta_j - z)$$

and require $|\eta_j| > 1$ for each j.

2.5 Autoregressive Processes

If $|\xi_j| > 1$ for at least one j, then no causal stationary solution to the AR(p) equation exists, but a similar approach to the AR(1) case allows us to define a (unique) non-causal stationary solution.

Hence the only case that leads to non-stationary solutions is when $|\xi_j| = 1$ for at least one j. In the simplest case $\xi_j = 1$, which corresponds to the factor $(1 - B)$ in the AR polynomial; however any complex root with modulus 1 yields a non-stationary solution, that is

$$\xi_j = e^{i\omega}$$

for $-\pi < \omega \leqslant \pi$, where $i = \sqrt{-1}$.

Example: AR(2) process

Suppose

$$X_t - \frac{13}{4}X_{t-1} + \frac{3}{4}X_{t-2} = Z_t.$$

Then

$$\Phi(z) = 1 - \frac{13}{4}z + \frac{3}{4}z^2 = (1 - 3z)(1 - z/4)$$

Therefore the reciprocal roots are 3 and $1/4$. Therefore no causal stationary solution exists.

Example: AR(2) process

Suppose

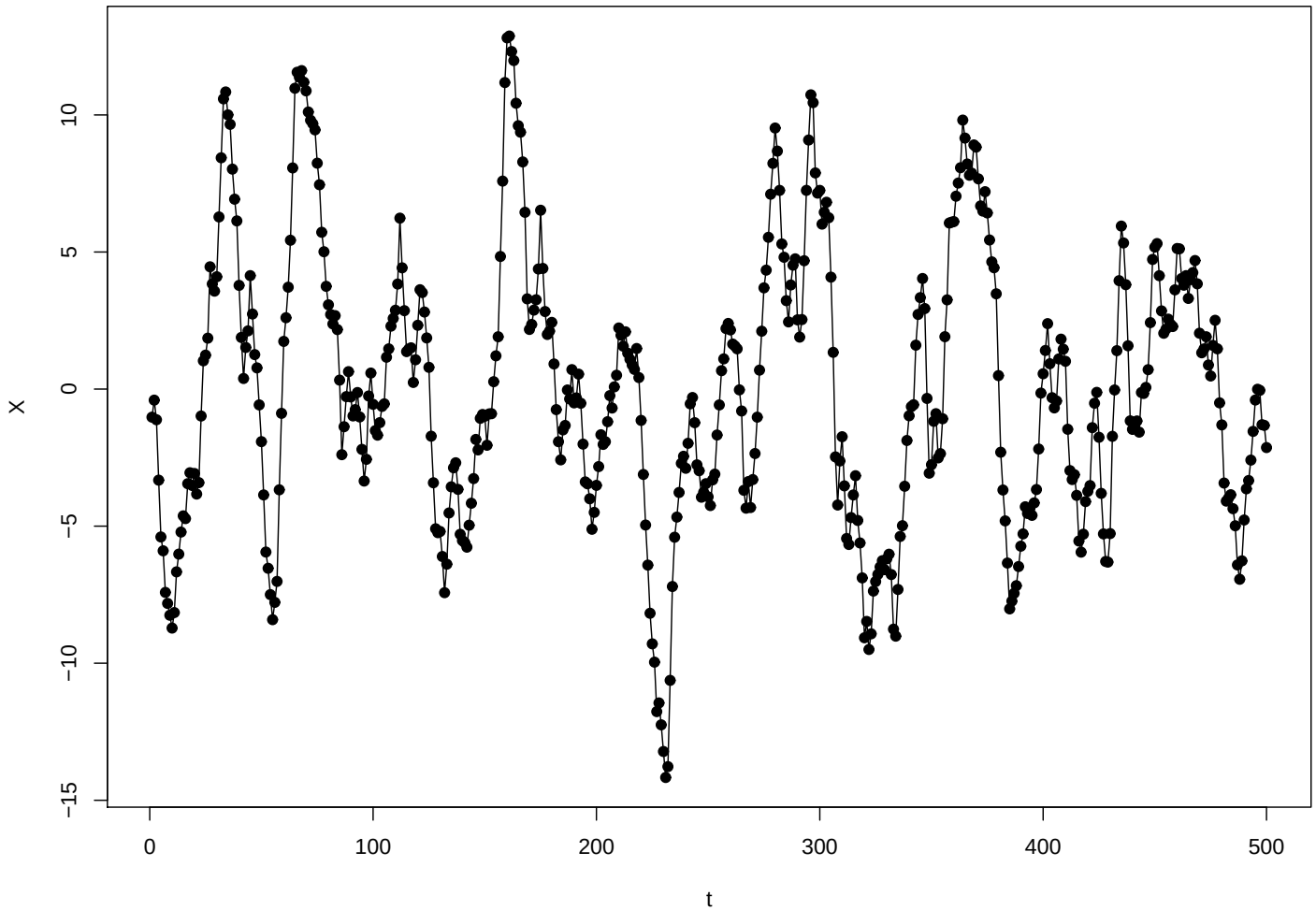
$$X_t - \frac{31}{20}X_{t-1} + \frac{3}{5}X_{t-2} = Z_t.$$

Then

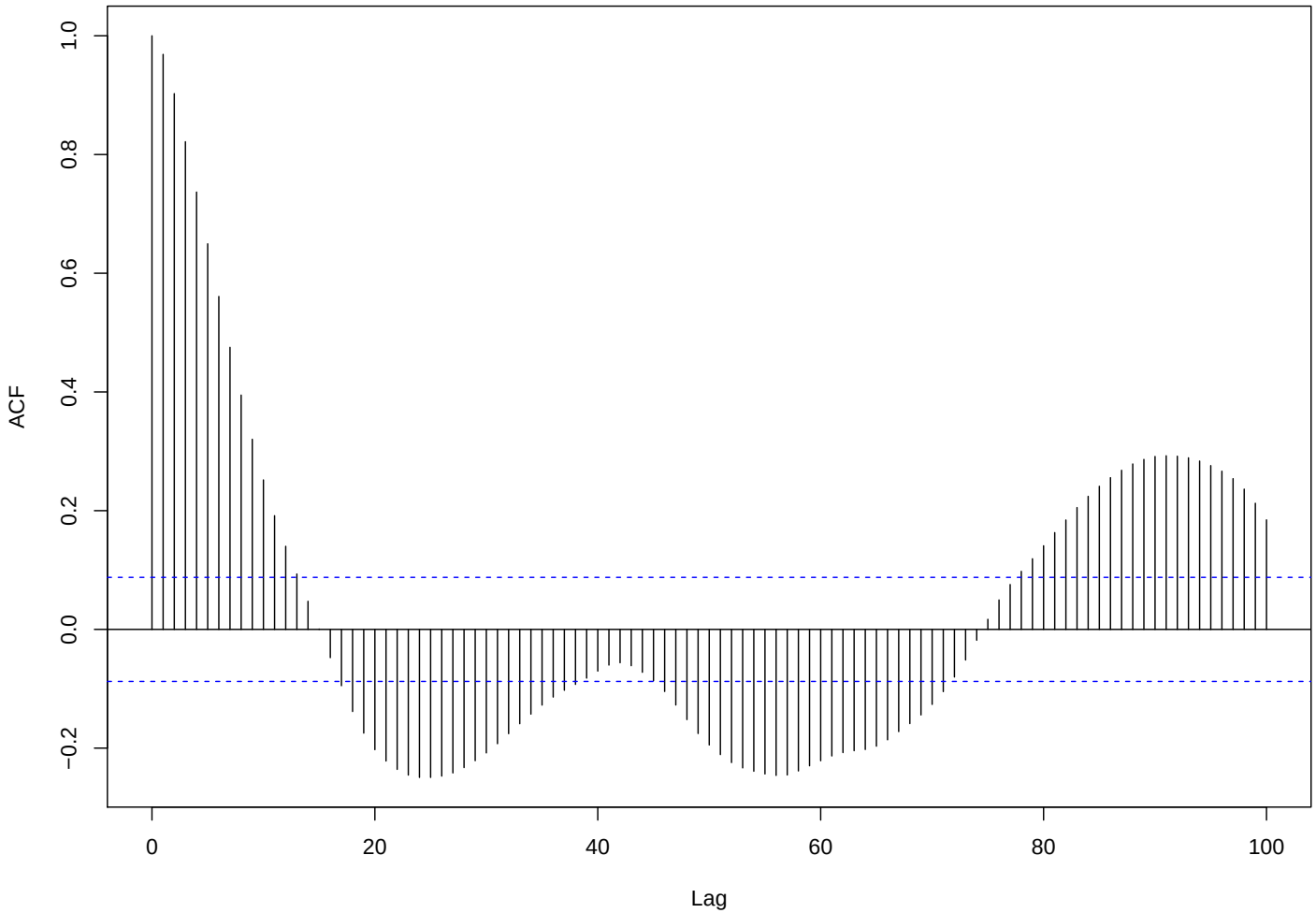
$$\Phi(z) = 1 - \frac{31}{20}z + \frac{3}{5}z^2 = (1 - 4z/5)(1 - 3z/4)$$

Therefore the reciprocal roots are $4/5$ and $3/4$. Therefore a causal stationary solution does exist.

2.5 Autoregressive Processes



2.5 Autoregressive Processes



2.5 Autoregressive Processes

Code: we can use the functions in `R` to perform key calculations

- `polyroot` to compute the roots of $\Phi(z) = 0$.
- `arima.sim` to simulate a time series.
- `ARMAtoMA` to compute the ψ_j values in the MA(∞) linear process formulation
- `ARMAacf` to compute the theoretical ACF.

In these functions, the representation

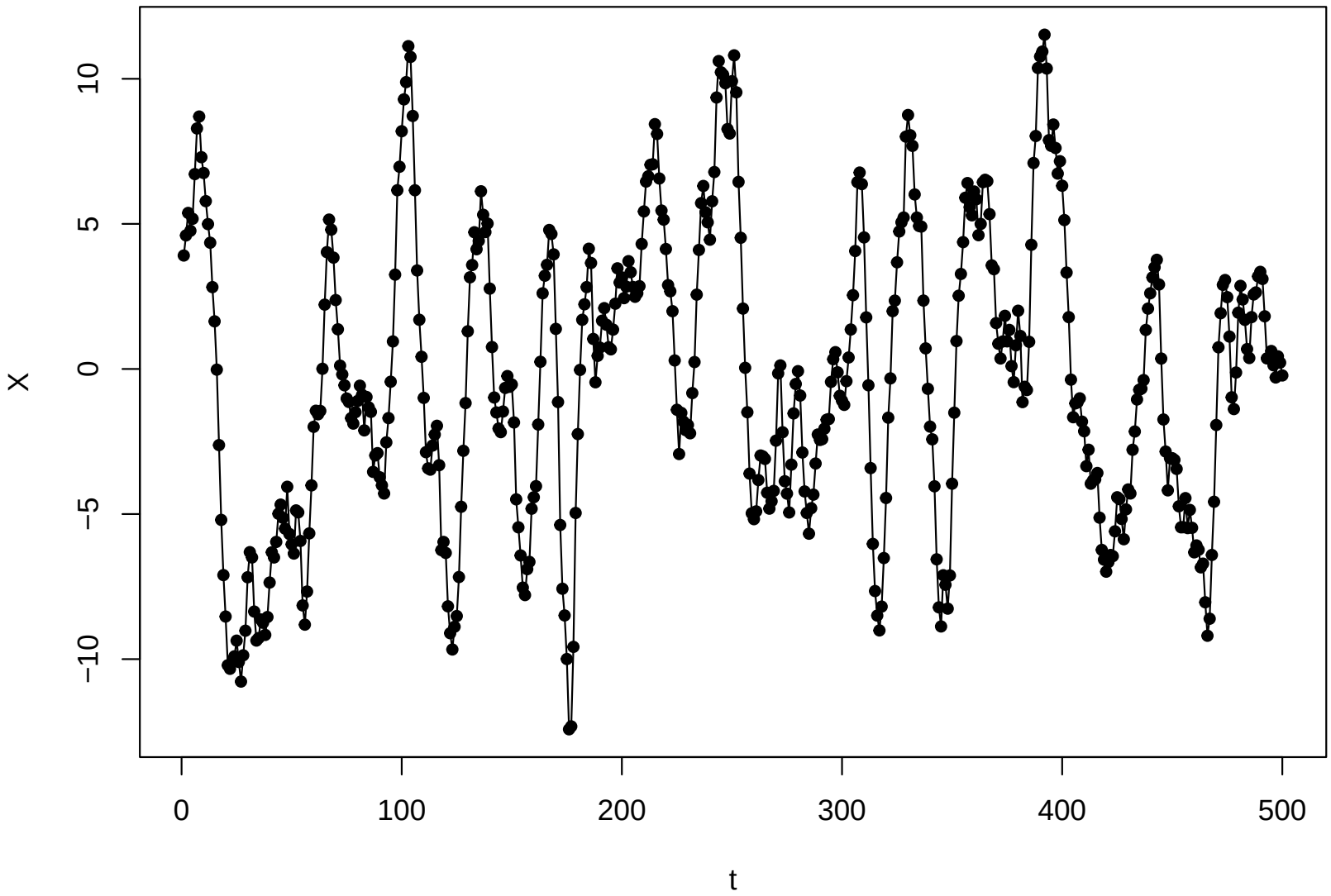
$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

is used, so that the coefficients in the above model are $\phi_1 = 31/20$ and $\phi_2 = -3/5$.

2.5 Autoregressive Processes

```r
polyroot(c(3/5,-31/20,1))  #Finding the roots (coefs in increasing term order)

## [1] 0.75+2.041443e-14i 0.80-2.041443e-14i

#Numerical solutions for roots are in complex number form, but imaginary part is negligible.

#Using the roots of a quadratic formula
aval<-1
bval<--31/20
cval<-3/5
root.vals<-(-bval+c(-1,1)*sqrt(bval^2-4*aval*cval))/(2*aval)
root.vals

## [1] 0.75 0.80

1/root.vals

## [1] 1.333333 1.250000

set.seed(32)
X<-arima.sim(n=500,model=list(ar=c(31/20,-3/5)))
par(mar=c(4,4,2,2))
plot(1:500,X,pch=19,cex=0.8,xlab='t')
lines(1:500,X)
```

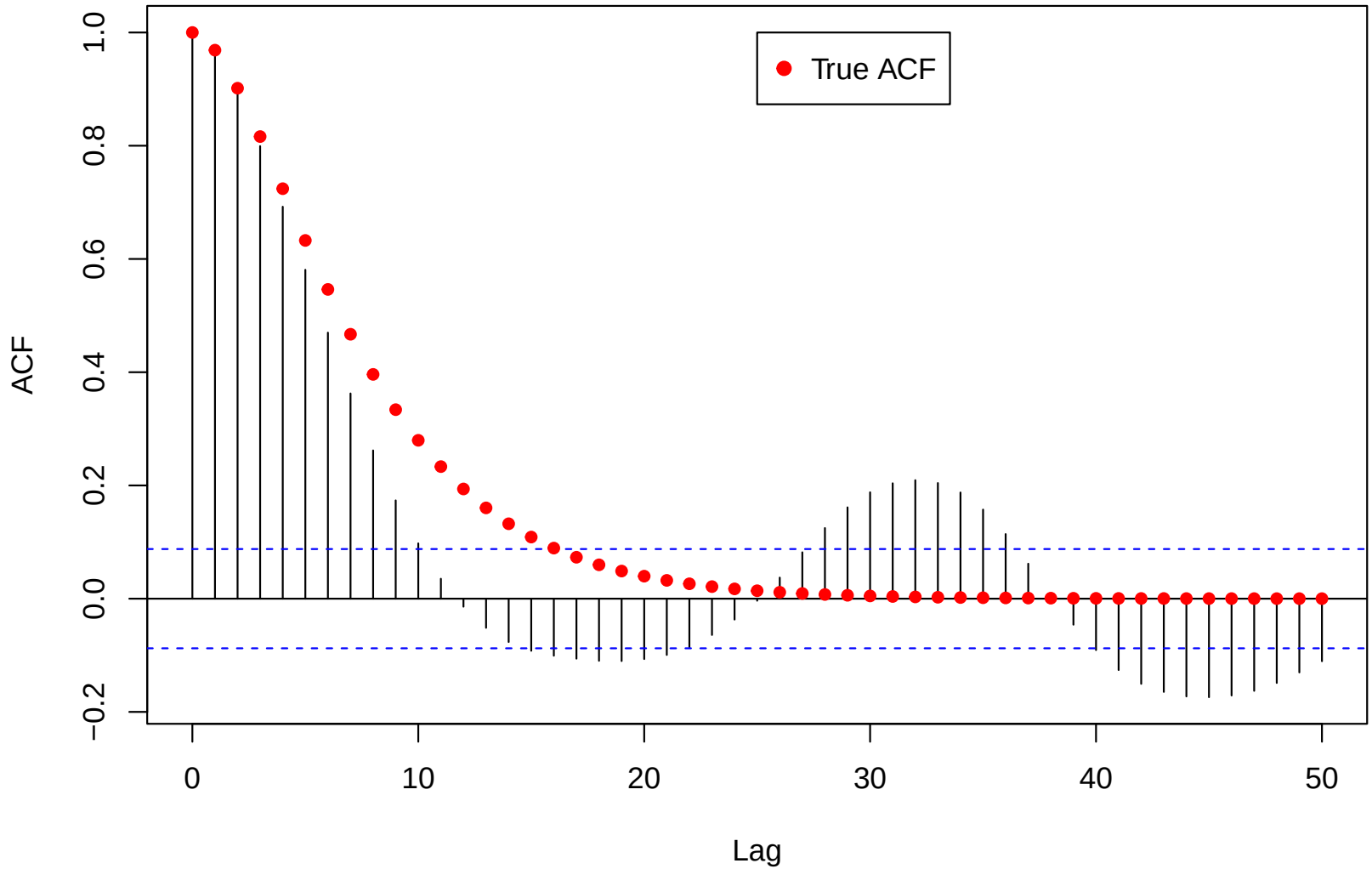
2.5 Autoregressive Processes



2.5 Autoregressive Processes

```r
ARMAtoMA(ar=c(31/20,-3/5),lag.max=20)  #Linear Process representation (first 20 coefs)

##  [1] 1.5500000 1.8025000 1.8638750 1.8075063 1.6833097 1.5246263 1.3531849
##  [8] 1.1826608 1.0212134 0.8732842 0.7408625 0.6243663 0.5232503 0.4364182
## [15] 0.3624980 0.3000210 0.2475338 0.2036647 0.1671601 0.1368993

par(mar=c(4,4,1,2))
true.acf<-ARMAacf(ar=c(31/20,-3/5),lag.max=50)
acf(X,main='ACF of X',lag.max=50)
points(0:50,true.acf,col='red',pch=19,cex=0.8)
legend(25,1,c('True ACF'),pch=19,col='red')
```

2.5 Autoregressive Processes

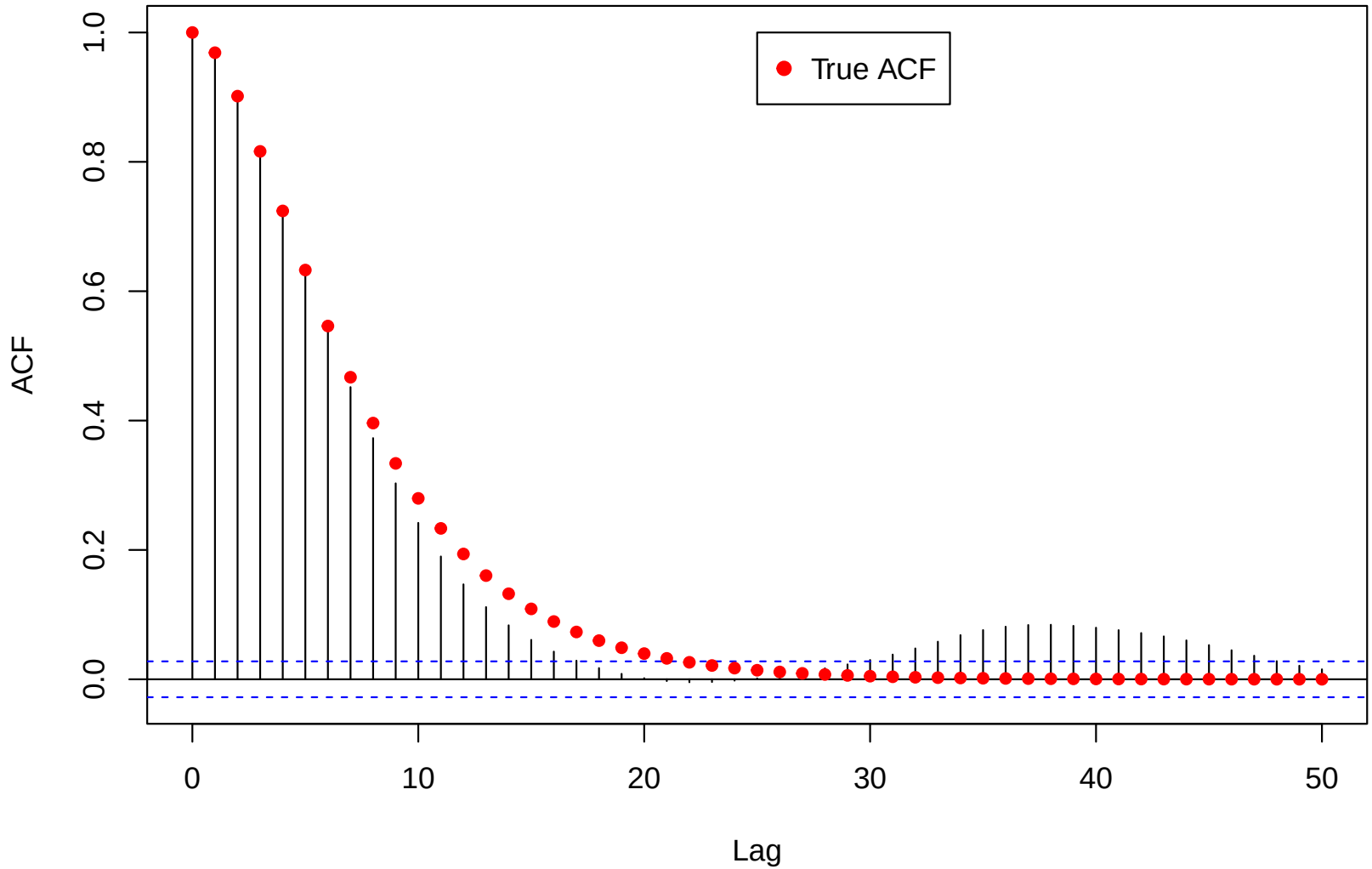


2.5 Autoregressive Processes

Larger sample size: $n = 5000$.

```
set.seed(32)
X<-arima.sim(n=5000,model=list(ar=c(31/20,-3/5)))
par(mar=c(4,4,1,2))
true.acf<-ARMAacf(ar=c(31/20,-3/5),lag.max=50)
acf(X,main='ACF of X',lag.max=50)
points(0:50,true.acf,col='red',pch=19,cex=0.8)
legend(25,1,c('True ACF'),pch=19,col='red')
```

2.5 Autoregressive Processes



2.6 ARMA Processes

If $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, we have

- **Moving Average:**

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

  - stationary
  - q-dependent; $\gamma_X(h) = 0$ for $h > q$

2.6 ARMA Processes

- **Autoregression:**

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$$

  - stationary provided the roots of the AR polynomial

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p = \prod_{j=1}^{p} (1 - \xi_j z)$$

    satisfy $|\xi_j| \neq 1$ for each j; causal if $|\xi_j| < 1$ for each j.
  - yields an MA(∞) representation.

2.6 ARMA Processes

$\{X_t\}$ is an ARMA(p, q) process if it is stationary and satisfies

$$\Phi(B)X_t = \Theta(B)Z_t$$

where $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

$$\begin{aligned}
\Phi(z) &= 1 - \phi_1 z - \cdots - \phi_p z^p = \prod_{j=1}^{p}(1 - \xi_j z) \\
\Theta(z) &= 1 + \theta_1 z + \cdots + \theta_q z^q = \prod_{j=1}^{q}(1 - \omega_j z)
\end{aligned}$$

Example: ARMA$(1,1)$

Suppose $\{X_t\}$ satisfies

$$(1-\phi B)X_t = (1+\theta B)Z_t$$

This is the ARMA$(1,1)$ process. Suppose that $|\phi| < 1$, and let

$$\alpha(B) = \{\Phi(B)\}^{-1} = \sum_{j=0}^{\infty} \phi^j B^j$$

Example: ARMA$(1,1)$

Then

$$X_t = \alpha(B)\Theta(B)Z_t = \Psi(B)Z_t$$

where

$$\Psi(z) = \alpha(z)\Theta(z) = (1 + \phi z + \phi^2 z^2 + \cdots)(1 + \theta z)$$

Example: ARMA$(1,1)$

Therefore

$$\begin{aligned} \psi_0 &= 1 \\ \psi_j &= (\theta+\phi)\phi^{j-1},\ j \geqslant 1. \end{aligned}$$

so

$$X_t = Z_t + (\theta+\phi)\sum_{j=1}^{\infty}\phi^{j-1}Z_{t-j}$$

Note: if $|\phi| > 1$, can construct a stationary non-causal version.

Invertibility: If

$$\Phi(B)X_t = \Theta(B)Z_t$$

then we may also write

$$\{\Theta(B)\}^{-1}\Phi(B)X_t = Z_t$$

provided the expansion

$$\beta(B) = \{\Theta(B)\}^{-1} = \sum_{j=0}^{\infty} \beta_j B^j$$

converges.

From above, we have

$$\Theta(z) = \prod_{j=1}^{q} (1 - \omega_j z)$$

so therefore expansion exists provided $|\omega_j| < 1$ converges. Such a process is termed **invertible.**

Example: ARMA$(1,1)$

In this case

$$\Theta(B) = 1 + \theta B \qquad \beta(B) = 1 - \theta B + \theta^2 B^2 - \cdots$$

and we require $|\theta| < 1$. Let $\pi(B) = \beta(B)\Phi(B)$, where

$$\pi_0 = 1$$

$$\pi_j = -(\theta + \phi)(-\theta)^{j-1},\ j \geqslant 1.$$

We then have an AR(∞) representation

$$\pi(B)X_t = Z_t.$$

2.7 ARMA processes: Estimation

Estimation for ARMA processes, and indeed all stationary processes, can be achieved in a non-parametric fashion using moment-based estimation.

For example, we may estimate expectation $\mu = \mathbb{E}[X_t]$ using estimator

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_t = \overline{X}$$

2.7 ARMA processes: Estimation

Recall that we require an estimator to have good statistical properties. In the non IID case, properties of this estimator are more difficult to study; it is clear that, under stationarity,

$$\mathbb{E}[\overline{X}] = \mu$$

so we have an unbiased estimator, but the variance of this estimator must be computed with care.

2.7 ARMA processes: Estimation

$$\begin{aligned}
\mathrm{Var}[\overline{X}] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}[X_i, X_j] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_X(|i-j|) \\
&= \frac{1}{n^2}\sum_{h=-n}^{n}(n-|h|)\gamma_X(h) \\
&= \frac{1}{n}\sum_{h=-n}^{n}\left(1-\frac{|h|}{n}\right)\gamma_X(h)
\end{aligned}$$

As $n \longrightarrow \infty$, this variance should stay finite, otherwise the performance of the estimator will be poor.

2.7 ARMA processes: Estimation

A necessary condition is that

$$\gamma_X(h) \longrightarrow 0 \qquad \text{as } h \longrightarrow \infty,$$

but this is not sufficient, for example if $\gamma_X(h) \sim 1/h$ as $h \longrightarrow \infty$, the sum diverges.

2.7 ARMA processes: Estimation

If $\text{Var}[\overline{X}] \longrightarrow 0$, then $\overline{X} \longrightarrow \mu$ in mean square. If

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$$

the asymptotic variance is the probability limit of $n\text{Var}[\overline{X}]$,

$$n\text{Var}[\overline{X}] \xrightarrow{p} \sum_{h=-\infty}^{\infty} \gamma_X(h)$$

Note that it is not necessarily the case that

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty.$$

2.7 ARMA processes: Estimation

For a Gaussian series $\{X_t\}$

$$\sqrt{n}(\overline{X} - \mu) \sim \mathcal{N}\left(0, \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma_X(h)\right)$$

whereas for non-Gaussian stationary series this result holds as an asymptotic approximation, provided the autocovariance series is absolutely summable.

2.7 ARMA processes: Estimation

In practice, $\gamma_X(h)$ must be estimated from the data, and in the asymptotic case, the infinite sum must be truncated;

$$\sum_{h=-l}^{l} \left(1 - \frac{|h|}{n}\right) \widehat{\gamma}(h)$$

commonly this is done at $l = \sqrt{n}$.

2.7 ARMA processes: Estimation

Example: AR(1) process

Suppose that $\{Z_t\} \sim \text{WN}(0, \sigma_Z^2)$, and

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

with $|\phi| < 1$. Then

$$\gamma_X(h) = \frac{\sigma_Z^2 \phi^{|h|}}{1 - \phi^2}$$

Example: AR(1) process

Then

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| = \sum_{h=-\infty}^{\infty} \gamma_X(h)$$

$$= \frac{\sigma_Z^2}{1-\phi^2}\left[1 + 2\sum_{h=1}^{\infty}\phi^h\right]$$

$$= \frac{\sigma_Z^2}{(1-\phi)^2}$$

2.7 ARMA processes: Estimation

Estimation of the autocovariance function $\gamma_X(h)$ can also be achieved by moment-based methods

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+h} - \overline{x})(x_t - \overline{x}).$$

This is a biased estimator of $\gamma_X(h)$, but the bias is low (order $o(1/n)$).

Note that the $k \times k$ symmetrix matrix $\widehat{\mathbf{\Gamma}}_k$ with

$$[\widehat{\mathbf{\Gamma}}_k]_{ij} = \widehat{\gamma}(|i-j|)$$

is non-negative definite for each $k \geqslant 1$.

2.7 ARMA processes: Estimation

The autocorrelation function $\rho_X(h)$ can be estimated by

$$\widehat{\mathbf{R}}_k = \frac{1}{\widehat{\gamma}(0)}\widehat{\mathbf{\Gamma}}_k$$

The properties of this estimator can be established, and can be used to construct hypothesis tests for each h that $\rho_X(h)$ is zero based on a z-score approach.

2.7 ARMA processes: Estimation

In particular, under the null hypothesis that the process is a white-noise process, we have approximately that

$$\sqrt{n}[\mathbf{R}_k]_{1,h} = \sqrt{n}\hat{\rho}(h) \sim \mathcal{N}(0,1) \qquad h \geqslant 1$$

Global (or portmanteau) tests, that assess multiple h can be constructed based on a Chi-squared distribution approximation. See for example the Box-Ljung test (`Box.test` in R).

2.7 ARMA processes: Estimation

Note: The matrices $\mathbf{\Gamma}_k, \widehat{\mathbf{\Gamma}}_k$ and other related matrices are symmetric Toeplitz matrices (that is, they are constant along diagonals).

$$\Gamma_k = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(k-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(k-3) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(k-4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \gamma(k-3) & \gamma(k-4) & \cdots & \gamma(0) \end{bmatrix}$$

2.7 ARMA processes: Estimation

Note: The matrices $\boldsymbol{\Gamma}_k, \widehat{\boldsymbol{\Gamma}}_k$ and other related matrices are symmetric Toeplitz matrices (that is, they are constant along diagonals).

$$\Gamma_k = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(k-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(k-3) \\ \gamma(3) & \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(k-4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \gamma(k-3) & \gamma(k-4) & \cdots & \gamma(0) \end{bmatrix}$$

2.8 Forecasting for Stationary Processes

Forecasting for stationary processes utilizes the autocovariance structure of the process to optimize predictions. Suppose we wish to forecast X_{n+h} given the observed data

$$x_{1:n} = (x_1, \ldots, x_n).$$

We focus on linear predictors, that is, linear combinations of past values.

We will focus on model-free procedures, and utilize moment properties rather distributional properties of $\{X_t\}$.

2.8 Forecasting for Stationary Processes

We utilize the minimum Mean-Square Error (MSE) criterion, and choose constants $a_{0:n} = (a_0, a_1, \ldots, a_n)^\top$ to minimize

$$\mathbb{E}\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n-i+1}\right)^2\right]$$

where the expectation is over the joint distribution of all the random quantities.

We denote the linear predictor using the n-data prediction operator notation

$$P_n X_{n+h} = a_0 + \sum_{i=1}^{n} a_i X_{n-i+1}$$

2.8 Forecasting for Stationary Processes

By differentiation, we need to solve simultaneously

$$\mathbb{E}\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n-i+1}\right)\right] = 0$$

and, for $j = 1, \ldots, n$

$$\mathbb{E}\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n-i+1}\right) X_{n-j+1}\right] = 0$$

From the first equation, we see that we must choose

$$a_0 = \mu\left(1 - \sum_{i=1}^{n} a_i\right)$$

where $\mu = \mathbb{E}[X_t]$.

2.8 Forecasting for Stationary Processes

In vector notation, the resulting system of n simultaneous equations is written

$$\mathbf{\Gamma}_n a = \gamma_{h:(h+n-1)} \tag{1}$$

where

$$\begin{aligned} \mathbf{\Gamma}_n &= [\gamma_X(i-j)]_{ij} \qquad (n \times n) \\ a &= (a_1, \ldots, a_n)^\top \qquad (n \times 1) \\ \gamma_{h:(h+n-1)} &= (\gamma_X(h), \ldots, \gamma_X(h+n-1))^\top \qquad (n \times 1) \end{aligned}$$

with $a_0 \equiv a_0(a_1, \ldots, a_n)$ given by the previous expression. This is a linear system in $(a_1, \ldots, a_n)$; in general the solution will also depend on h.

2.8 Forecasting for Stationary Processes

Denote the solution to (1) by

$$a(n,h) = (a(1,h),\ldots,a(n,h))^\top$$

so that

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a(i,h)(X_{n-i+1} - \mu).$$

Note that, by construction

$$\begin{aligned}
\mathbb{E}[(X_{n+h} - P_n X_{n+h})] &= 0 \\
\mathbb{E}[(X_{n+h} - P_n X_{n+h})X_j] &= 0 \qquad j = 1,\ldots,n
\end{aligned}$$

The minimum value of the MSE achieved by $a(n,h)$ is

$$\mathbb{E}[(X_{n+h} - P_n X_{n+h})^2] = \gamma_X(0) - a(n,h)^\top \boldsymbol{\gamma}_{h:(h+n-1)}$$

2.8 Forecasting for Stationary Processes

The formula

$$\mathbf{\Gamma}_n a = \gamma_{h:(h+n-1)}$$

depends only on the autocovariance function. Direct solution of this equation is possible for most practical situations

$$a(n,h) = \mathbf{\Gamma}_n^{-1}\gamma_{h:(h+n-1)}$$

can be computed using `solve` in R. However, the form of $\mathbf{\Gamma}_n$ (symmetric, Toeplitz) can make it relatively easy to invert.

2.8 Forecasting for Stationary Processes

Example: AR(1)

Recall that

$$[\mathbf{\Gamma}_n]_{ij} = \phi^{|i-j|}$$

so that

$$\boldsymbol{\gamma}_{h:(h+n-1)} = (\phi^h, \phi^{h+1}, \ldots, \phi^{h+n-1})^\top$$

Then the matrix inversion can be carried out analytically, yielding

$$a(n,h) = (\phi^h, 0, 0, \ldots, 0)^\top$$

which is an intuitively reasonable result. It follows that

$$P_n X_{n+1} = \phi X_n \qquad\qquad P_n X_{n+2} = \phi^2 X_n$$

and so on.

2.8 Forecasting for Stationary Processes

Let $Y, W_1, \ldots, W_n$ be random variables with

$$\mathbb{E}[Y] = \mu_Y \quad \mathbb{E}[W_i] = \mu_i$$

and

$$\mathrm{Var}[Y], \mathrm{Var}[W_i], \mathrm{Cov}[Y, W_i], \mathrm{Cov}[W_i, W_j] < \infty.$$

Write $W_{n:1} = (W_n, \ldots, W_1)^\top$, and

$$\boldsymbol{\gamma} = \mathrm{Cov}(Y, W_{n:1}) \qquad (n \times 1)$$
$$\boldsymbol{\Gamma} = \mathrm{Cov}(W_{n:1}, W_{n:1}) \qquad (n \times n)$$

2.8 Forecasting for Stationary Processes

Denote the best linear predictor of Y, given $W_1, \ldots, W_n$ by

$$\widehat{P}(Y|W) = \mu_Y + a^\top (W_{n:1} - \mu_{n:1})$$

where a solves $\mathbf{\Gamma} a = \boldsymbol{\gamma}$. The minimum MSE is then

$$\mathbb{E}[(Y - \widehat{P}(Y|W))^2] = \mathrm{Var}[Y] - a^\top \gamma.$$

2.8 Forecasting for Stationary Processes

Let U and V be random variables with $\mathbb{E}[U^2], \mathbb{E}[V^2] < \infty$, and $\beta, \alpha_1, \ldots, \alpha_n$ be constants. Then

1. $\widehat{P}(U|W_{n:1}) = \mathbb{E}[U] + a^\top (W_{n:1} - \mu_{n:1})$.
2. $\mathbb{E}[U - \widehat{P}(U|W_{n:1})] = \mathbb{E}[(U - \widehat{P}(U|W_{n:1}))W_{n:1}] = \mathbf{0}$.
3. $\mathbb{E}[(U - \widehat{P}(U|W_{n:1}))^2] = \mathrm{Var}[U] - a^\top \gamma$.
4. $\widehat{P}$ is a linear operator

$$\widehat{P}(\alpha_1 U + \alpha_2 V + \beta|W_{n:1}) = \alpha_1 \widehat{P}(U|W_{n:1}) + \alpha_2 \widehat{P}(V|W_{n:1}) + \beta.$$

5. $\widehat{P}(\alpha^\top W_{n:1} + \beta | W_{n:1}) = \alpha^\top W_{n:1} + \beta$.
6. If $\gamma = 0$, $\widehat{P}(U|W_{n:1}) = \mathbb{E}[U]$.
7. Coherent iterated prediction: If $W = (W_{n:1}, W^*_{n:1})$, then

$$\widehat{P}(U|W) = \widehat{P}(\widehat{P}(U|W_{n:1}, W^*_{n:1})|W_{n:1}).$$

2.8 Forecasting for Stationary Processes

Example: AR(p) process

If $\{Z_t\} \sim \text{WN}(0, \sigma^2)$, and

$$\Phi(B)X_t = Z_t$$

such that $\{X_t\}$ is stationary and causal. Then for each n

$$\begin{aligned} P_{n-1}X_n &= P_n[\phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} + Z_n] \\ &= \phi_1 X_{n-1} + \cdots + \phi_p X_{n-p} \end{aligned}$$

by 4., 5. and 6.

2.9 Levinson-Durbin Algorithm

For a stationary process $\{X_t\}$, suppose we write

$$P_n X_{n+1} = \sum_{j=1}^{n} \phi_{n,j} X_{n-j+1} = \Phi_n^\top X_{1:n}$$

say as the general form of a linear predictor, where

$$\Phi_n = (\phi_{n,1}, \ldots, \phi_{n,n})^\top \qquad (n \times 1)$$

is the coefficient vector resulting from n data.

2.9 Levinson-Durbin Algorithm

By the arguments above, the optimal MSE choice for ϕ_n is

$$\phi_n = \mathbf{\Gamma}_n^{-1}\boldsymbol{\gamma}_n$$

where

$$[\mathbf{\Gamma}_n]_{ij} = \gamma_X(|i-j|) \qquad \boldsymbol{\gamma}_n = (\gamma_X(1), \ldots, \gamma_X(n))^\top.$$

Let the minimum MSE be denoted

$$v_n = \mathbb{E}\left[\left(X_{n+1} - P_n X_{n+1}\right)^2\right] = \gamma_X(0) - \phi_n^\top \boldsymbol{\gamma}_n.$$

2.9 Levinson-Durbin Algorithm

We seek a recursion for ϕ_n; suppose ϕ_{n-1} and v_{n-1} have been computed, and define ϕ_n by setting the nth component as

$$\phi_{n,n} = \frac{1}{v_{n-1}} \left[\gamma_X(n) - \Phi_{n-1}^\top \gamma_{(n-1):1} \right]$$

and the first $n-1$ components via

$$\Phi_{n,1:(n-1)} = \Phi_{n-1} - \phi_{n,n} \Phi_{n-1}^{\mathrm{R}}$$

where

$$\Phi_{n-1}^{\mathrm{R}} = (\phi_{n-1,n-1}, \ldots, \phi_{n-1,1})^\top$$

is the reversed version of ϕ_{n-1}. Finally, set

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

2.9 Levinson-Durbin Algorithm

To initialize the recursion, set

$$\phi_{1,1} = \rho_X(1) = \frac{\gamma_X(1)}{\gamma_X(0)} \qquad v_0 = \gamma_X(0) \qquad v_1 = \gamma_X(0)(1 - \phi_{1,1}^2)$$

To see that the recursion produces a valid solution, first set $\mathbf{R}_n$ to be the autocorrelation matrix $\mathbf{R}_n = \mathbf{\Gamma}_n / \gamma_X(0)$, and define

$$\boldsymbol{\rho}_n = \frac{1}{\gamma_X(0)} \boldsymbol{\gamma}_n = (\rho_X(1), \ldots, \rho_X(n))^\top.$$

We verify the inductive step

$$\mathbf{R}_n \boldsymbol{\phi}_n = \boldsymbol{\rho}_n \qquad \Longrightarrow \qquad \mathbf{R}_{n+1} \boldsymbol{\phi}_{n+1} = \boldsymbol{\rho}_{n+1}$$

2.9 Levinson-Durbin Algorithm

Clearly the relation holds for $n = 1$, as

$$\mathbf{R}_1 = 1 \qquad \boldsymbol{\phi}_1 = \phi_{1,1} = \rho_X(1) \qquad \boldsymbol{\rho}_1 = \rho_X(1).$$

We assume the relation holds for $n = k$, and verify that it holds for $n = k + 1$. Now

$$\mathbf{R}_{k+1} = \begin{bmatrix} \mathbf{R}_k & \boldsymbol{\rho}_k^{\mathrm{R}} \\ \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top & 1 \end{bmatrix}$$

where

$$\boldsymbol{\rho}_k^{\mathrm{R}} = (\rho_X(k), \ldots, \rho_X(1))^\top = \frac{1}{\gamma_X(0)} \boldsymbol{\gamma}_k^{\mathrm{R}}$$

2.9 Levinson-Durbin Algorithm

Recall that

$$\mathbf{R}_{k+1} = \frac{1}{\gamma_X(0)} \begin{bmatrix} \gamma_X(0) & \gamma_X(1) & \gamma_X(2) & \cdots & \gamma_X(k) \\ \gamma_X(1) & \gamma_X(0) & \gamma_X(1) & \cdots & \gamma_X(k-1) \\ \gamma_X(2) & \gamma_X(1) & \gamma_X(0) & \cdots & \gamma_X(k-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_X(k) & \gamma_X(k-1) & \gamma_X(k-2) & \cdots & \gamma_X(0) \end{bmatrix}$$

and note that

$$\mathbf{R}_k \phi_k = \boldsymbol{\rho}_k \qquad \Longrightarrow \qquad \mathbf{R}_k \phi_k^{\mathrm{R}} = \boldsymbol{\rho}_k^{\mathrm{R}}.$$

2.9 Levinson-Durbin Algorithm

$$
\begin{aligned}
\mathbf{R}_{k+1}\boldsymbol{\phi}_{k+1} &= \begin{bmatrix} \mathbf{R}_k & \boldsymbol{\rho}_k^{\mathrm{R}} \\ \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_k - \phi_{k+1,k+1}\boldsymbol{\phi}_k^{\mathrm{R}} \\ \phi_{k+1,k+1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}_k\boldsymbol{\phi}_k - \phi_{k+1,k+1}\mathbf{R}_k\boldsymbol{\phi}_k^{\mathrm{R}} + \phi_{k+1,k+1}\boldsymbol{\rho}_k^{\mathrm{R}} \\ \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top\boldsymbol{\phi}_k - \phi_{k+1,k+1}\{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top\boldsymbol{\phi}_k^{\mathrm{R}} + \phi_{k+1,k+1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}_k\boldsymbol{\phi}_k \\ \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top\boldsymbol{\phi}_k + \phi_{k+1,k+1}\big(1 - \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top\boldsymbol{\phi}_k^{\mathrm{R}}\big) \end{bmatrix}
\end{aligned}
$$

as $\mathbf{R}_k\boldsymbol{\phi}_k^{\mathrm{R}} = \boldsymbol{\rho}_k^{\mathrm{R}}$.

2.9 Levinson-Durbin Algorithm

Now by definition

$$\phi_{k+1,k+1} = \frac{1}{v_k}\left[\gamma_X(k+1) - \phi_k^\top \gamma_k^{\mathrm{R}}\right]$$

and by definition

$$\begin{aligned} v_k = \gamma_X(0) - \phi_k^\top \gamma_k \quad &= \quad \gamma_X(0) - \{\phi_k^{\mathrm{R}}\}^\top \gamma_k^{\mathrm{R}} \\ &= \quad \gamma_X(0)(1 - \{\rho_k^{\mathrm{R}}\}^\top \phi_k^{\mathrm{R}}) \end{aligned}$$

so therefore

$$\frac{1}{v_k}(1 - \{\rho_k^{\mathrm{R}}\}^\top \phi_k^{\mathrm{R}}) = \frac{1}{\gamma_X(0)}$$

2.9 Levinson-Durbin Algorithm

Therefore

$$\phi_{k+1,k+1} = \frac{1}{\gamma_X(0)}\left[\gamma_X(k+1) - \boldsymbol{\phi}_k^\top \boldsymbol{\gamma}_k^{\mathrm{R}}\right] = \rho_X(k+1) - \boldsymbol{\phi}_k^\top \boldsymbol{\rho}_k^{\mathrm{R}}$$

and hence

$$\begin{aligned}
\mathbf{R}_{k+1}\boldsymbol{\phi}_{k+1} &= \begin{bmatrix} \mathbf{R}_k\boldsymbol{\phi}_k \\ \{\boldsymbol{\rho}_k^{\mathrm{R}}\}^\top \boldsymbol{\phi}_k + \rho_X(k+1) - \boldsymbol{\phi}_k^\top \boldsymbol{\rho}_k^{\mathrm{R}} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\rho}_k \\ \rho_X(k+1) \end{bmatrix} \\
&= \boldsymbol{\rho}_{k+1}
\end{aligned}$$

and the recursion holds.

2.9 Levinson-Durbin Algorithm

For the prediction MSE recursion: using the recursion for ϕ_n,

$$\begin{aligned}
v_n &= \mathbb{E}[(X_{n+1} - \phi_n^\top X_{n:1})^2] = \gamma_X(0) - \phi_n^\top \gamma_n \\
&= \gamma_X(0) - \phi_{n-1}^\top \gamma_{n-1} + \phi_{n,n} \left\{\phi_{n-1}^{\mathrm{R}}\right\}^\top \gamma_{n-1} - \phi_{n,n}\gamma_X(n) \\
&= v_{n-1} + \phi_{n,n} \left(\left\{\phi_{n-1}^{\mathrm{R}}\right\}^\top \gamma_{n-1} - \gamma_X(n) \right) \\
&= v_{n-1} - \phi_{n,n}^2 (\gamma_X(0) - \phi_{n-1}^\top \gamma_{n-1}) \\
&= v_{n-1}(1 - \phi_{n,n}^2).
\end{aligned}$$

where line 2 follows from line 1 by substituting in the formula for ϕ_n given by the recursion.

2.10 Partial Autocorrelation Function (PACF)

Define the function $\alpha_X(h)$ by $\alpha_X(0) = 1$ and

$$\alpha_X(h) = \phi_{h,h} \qquad h = 1, 2, \ldots, n$$

where $\phi_{h,h}$ is the value returned as the coefficient on the hth step of the Levinson-Durbin recursion.

This function is termed the partial autocorrelation function, or PACF, which is uniquely determined by the ACF γ_X.

2.10 Partial Autocorrelation Function (PACF)

Note also that as v_n is non-negative, and

$$v_n = v_{n-1}(1 - \phi_{n,n}^2)$$

we must have that $\phi_{n,n}^2 \leqslant 1$ for all n, and hence that

$$v_n \leqslant v_{n-1}$$

so the optimal prediction MSEs form a decreasing sequence.

2.10 Partial Autocorrelation Function (PACF)

For random variables X, Y, Z, the partial correlation of X and Y, given Z is defined by

$$\text{Corr}[(X - \mathbb{E}[X|Z]), (Y - \mathbb{E}[Y|Z])]$$

that is, the correlation between the residuals from a regression of X on Z with those from a regression of Y on Z.

2.10 Partial Autocorrelation Function (PACF)

For process $\{X_t\}$, if we let $X \equiv X_t$, $Y = X_{t+h}$, and $Z = X_{(t+1):(t+h-1)}$.

$$\begin{aligned}\alpha_X(h) &= \phi_{h,h} \\ &= \text{Corr}\big[\big(X_t - \widehat{P}(X_t|X_{(t+1):(t+h-1)})\big), \\ &\qquad \big(X_{t+h} - \widehat{P}(X_{t+h}|X_{(t+1):(t+h-1)})\big)\big]\end{aligned}$$

The PACF computes the correlation between prediction residuals for X_t and X_{t+h}, using $X_{(t+1):(t+h-1)}$ to make the predictions.

2.10 Partial Autocorrelation Function (PACF)

Example: AR(p)

Suppose $\{X_t\} \sim \text{AR}(p)$, with parameters $\phi_1, \ldots, \phi_p$, and consider $h > p$. Then

$$\widehat{P}(X_{t+h} | X_{(t+1):(t+h-1)}) = \phi_1 X_{t+h-1} + \cdots + \phi_p X_{t+h-p}$$

and

$$\widehat{P}(X_t | X_{(t+1):(t+h-1)}) = \widehat{X}_t = \phi_1 X_{t+1} + \cdots + \phi_p X_{t+p}$$

so therefore

$$\begin{aligned} X_{t+h} - \widehat{P}(X_{t+h} | X_{(t+1):(t+h-1)}) &= Z_{t+h} \\ X_t - \widehat{P}(X_t | X_{(t+1):(t+h-1)}) &= Z_t \end{aligned}$$

and as $\text{Corr}[Z_t, Z_{t+h}] = 0$, the PACF is zero for $h > p$.

2.10 Partial Autocorrelation Function (PACF)

The Levinson-Durbin algorithm provides a method for solving the linear system

$$\mathbf{R}_n \boldsymbol{\phi}_n = \boldsymbol{\rho}_n$$

by recursion, for arbitrary n, and given ACF $\gamma_X(h)$ without the need for large matrix inversion. It exploits the symmetric Toeplitz nature of an autocovariance matrix, and is an order n^2 algorithm.

Other algorithms for inverting this type of matrix are typically higher order; the Cholesky decomposition

$$\mathbf{R}_n = \mathbf{L}_n \mathbf{L}_n^\top$$

with $\mathbf{L}_n$ lower triangular is order n^3. It does not exploit the Toeplitz nature of $\mathbf{R}_n$.

2.11 The Innovations Algorithm

Another approach to forecasting proves feasible even for the non-stationary case. Suppose that $\{X_t\}$ is zero mean process with $\mathbb{E}[|X_t|^2]$, and $\mathbb{E}[X_i X_j] = \kappa(i,j)$. Denote

$$\widehat{X}_n = \begin{cases} 0 & n = 1 \\ P_{n-1} X_n & n = 2, 3, \ldots. \end{cases}$$

and let

$$U_n = X_n - \widehat{X}_n$$

denote the one step prediction error. Then

$$U_{1:n} = \mathbf{A}_n X_{1:n}$$

where $\mathbf{A}_n$ contains the coefficients $a_{i,j}$ that appear in the optimal linear predictor.

2.11 The Innovations Algorithm

$$\mathbf{A}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{1,1} & 1 & 0 & \cdots & 0 \\ a_{2,2} & a_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

This lower triangular matrix is non-singular and let

$$\mathbf{A}_n^{-1} = \mathbf{C}_n$$

where $\mathbf{C}_n$ is also lower triangular.

2.11 The Innovations Algorithm

Let

$$\mathbf{C}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 1 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}$$

Then, recalling that $X_{1:n} = \mathbf{C}_n U_{1:n}$,

$$X_{1:n} = \mathbf{C}_n U_{1:n} = \mathbf{C}_n (X_{1:n} - \widehat{X}_{1:n}).$$

Also

$$\widehat{X}_{1:n} = X_{1:n} - U_{1:n} = \mathbf{C}_n U_{1:n} - U_{1:n} = \mathbf{\Theta}_n (X_{1:n} - \widehat{X}_{1:n})$$

2.11 The Innovations Algorithm

$$\mathbf{\Theta}_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{1,1} & 0 & 0 & \cdots & 0 \\ \theta_{2,2} & \theta_{2,1} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

Therefore

$$\widehat{X}_{n+1} = \begin{cases} 0 & n = 0 \\ \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & n = 1, 2, \ldots. \end{cases}$$

2.11 The Innovations Algorithm

Therefore the one-step predictions

$$\widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_n, \ldots$$

can be computed as weighted sums of previous prediction errors

$$X_1 - \widehat{X}_1, X_2 - \widehat{X}_2, \ldots, X_n - \widehat{X}_n, \ldots$$

provided the entries in the matrix $\boldsymbol{\Theta}_n$ are known. Note that by the previous results, the sequence $\{D_n\}$ defined by

$$D_n = X_n - \widehat{X}_n$$

is an **uncorrelated** sequence.

Another recursive approach avoids the need to carry out matrix inversion.

2.11 The Innovations Algorithm

Initialize $v_0 = \kappa(1,1) = \mathbb{E}[X_1^2]$, then for each n set

$$\theta_{n,n-k} = \frac{1}{v_k}\left[\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\,v_j\right]$$

for $0 \leqslant k < n$, and

$$v_n = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2\,v_j$$

That is, we compute

$$\begin{array}{cccc} v_0 & & & \\ \theta_{1,1} & v_1 & & \\ \theta_{2,2} & \theta_{2,1} & v_2 & \\ \theta_{3,3} & \theta_{3,2} & \theta_{3,1} & v_3 \end{array}$$

and so on.

2.11 The Innovations Algorithm

For h-step ahead prediction

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+j-1,j}\left(X_{n+h-j} - \hat{X}_{n+j-j}\right)$$

and

$$v_{n+h} = \kappa(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta^2_{n+j-1,j} v_{n+h-j-1}$$

In this case, the innovations algorithm must be run **forward** $h-1$ steps.

2.11 The Innovations Algorithm

Note: It is possible to extend the idea of optimal linear prediction from the finite case to the infinite, that is, to find constants $\{\beta_j\}$ such that

$$\widetilde{P}X_{n+h} = \sum_{j=1}^{\infty} \beta_j X_{n+1-j}.$$

The infinite system of moment constraints

$$\mathbb{E}[(X_{n+h} - \widetilde{P}X_{n+h})X_{n+1-i}] = 0 \qquad i = 1, 2, \ldots$$

yield the infinite system of equations

$$\sum_{j=1}^{\infty} \gamma_X(i-j)\beta_j = \gamma_X(h+i-1) \qquad i = 1, 2,$$

2.12 The Wold Decomposition

Any process $\{X_t\}$ where

$$X_n - \widetilde{P}_{n-1} X_n = 0$$

in mean square, that is

$$\mathbb{E}\left[(X_n - \widetilde{P}_{n-1} X_n)^2\right] = 0$$

for all n is termed deterministic

2.12 The Wold Decomposition

Example: Simple Deterministic Process

Suppose U, V are zero mean and uncorrelated variables with variance σ^2, and let

$$X_t = U\cos(\omega t) + V\sin(\omega t)$$

for some $\omega \in (0, \pi)$. Then for each integer n,

$$X_n = 2\cos(\omega)X_{n-1} - X_{n-2} = \widetilde{P}_{n-1}X_n$$

say, and hence

$$X_n - \widetilde{P}_{n-1}X_n = 0$$

2.12 The Wold Decomposition

The Wold Decomposition gives a fundamental representation of stationary processes.

If $\{X_t\}$ is stationary and non-deterministic, then $\{X_t\}$ has the representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t \qquad \text{(Wold)}$$

where

- $\{Z_t\} \sim \text{WN}(0, \sigma^2)$
- $\{V_t\}$ is deterministic
- $\{Z_t\}$ and $\{V_t\}$ are uncorrelated
- $\psi_0 = 1$, $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.

2.12 The Wold Decomposition

We also have that

- $Z_t = \widetilde{P}_t Z_t$ for all t
- $V_t = \widetilde{P}_s V_t$ for all s, t.

and that the representation in (Wold) is unique. Furthermore

- for each t, $Z_t = X_t - \widetilde{P}_{t-1} X_t$;
- for each j, ψ_j satisfies

$$\psi_j = \frac{\mathbb{E}[X_t Z_{t-j}]}{\mathbb{E}[Z_t^2]};$$

- for each t, $V_t = X_t - \sum_{j=0}^{\infty} \psi_j Z_{t-j}$.