2 Time Series Analysis and Forecasting

2.4 More Examples of Time Series

Los Angeles Annual Rainfall

## The package TSA should be installed before running this code
## Annual Rainfall in Los Angeles
## Time Series Plot of LA annual rainfall

library(TSA)

win.graph(width=4.875, height = 2.5, pointsize=8)

# Load the LA rainfall data
data(larain);
print(larain);

# Plot the time series
plot(larain, main="", ylab='Inches', xlab='Year', type='o');
title(main = list("Time Series Plot of LA Annual Rainfall",
                  cex=1.5, col="red", font=2))

Los Angeles Annual Rainfall

![Time Series Plot of LA Annual Rainfall](image1.png)

Figure 1: Time series of annual rainfall in Los Angeles.

Chemical Process

## An Industrial Chemical Process
## Time Series Plot of Color Property from a Chemical Process

win.graph(width=4.875, height = 2.5, pointsize=8)

data(color);
print(color);

plot(color, main='Time Series Plot of Color Property from a Chemical Process',
     ylab='Color Property', xlab='Batch', type='o');
Chemical Process

Figure 2: Color property from a chemical process.

Abundance of Canadian Hare

```r
## Annual Abundance of Canadian Hare
## Time Series Plot of Abundance of Canadian Hare
win.graph(width=4.875, height = 2.5, pointsize=8)
data(hare);
print(hare);
plot(hare, main='Time Series Plot of Annual Abundance of Canadian Hare',
     ylab='Abundance', xlab='Year', type='o');
```

Abundance of Canadian Hare

Figure 3: Abundance of Canadian hare.
2.5 Residuals from Smoothed Data

Residual from Smoothed Value

\[ \text{Res}(X_t) = \text{residual of the smoothed time series } X_t. \]

\[ \text{Res}(X_t) = X_t - Y_t = X_t - \sum_{r=-q}^{q} b_r \cdot X_{t+r}. \]

\( b_0 = 1 - a_0, \quad b_r = -a_r \quad \text{for } r \neq 0. \) This is a linear filter. If \( \sum a_r = 1 \) then \( \sum b_r = 0 \) and the filter \( \text{Res}(X_t) \) is a trend remover.

Exercise

Let \( q = 3, Y_t = X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3} \), find \( \text{Res}(X_t) \).

Filters in time series

- A smoothing procedure may be carried out in two or more stages.
- A series of linear operation is still a linear filter.

\[ Z_t = \sum_{j} b_j \cdot Y_{t+j} = \sum_{j} b_j \sum_{r} a_r \cdot X_{t+j+r} = \sum_{k} c_k \cdot X_{t+k}, \]

for \( k = j + r \rightarrow j = k - r \) and \( c_k = \sum_{r} a_r \cdot b_{(k-r)} \). This procedure is called convolution, \( \{c_k\} = \{a_r\} * \{b_j\} \).

Filters in time series

Example 1. Consider the following filter:

\[ \left( \begin{array}{c} 1 \\ 4 \\ 1 \\ 2 \\ 1 \\ 4 \end{array} \right) = \left( \begin{array}{c} 1 \\ 2 \\ 1 \\ 2 \end{array} \right) * \left( \begin{array}{c} 1 \\ 2 \\ 1 \\ 2 \end{array} \right). \]

\( \{c_1, c_2, c_3, c_4\} = \{a_1, a_2\} * \{b_1, b_2\} \).

\[ c_1 = \sum_{r} a_r b_{1-r} = \]

\[ c_2 = \sum_{r} a_r b_{2-r} = \]

\[ c_3 = \sum_{r} a_r b_{3-r} = \]

\[ c_4 = \sum_{r} a_r b_{4-r} = \]
Time Series with a Trend

Approaches to Describe Trend

3 Difference to remove trend.

- Box-Jenkins procedure.
- Difference the series until it becomes stationary.
- First order differencing (widely used).

\[ Y_t = \nabla X_t = X_t - X_{t-1}, \quad t = 2, 3, \cdots, N. \]

- Second order differencing (not popular).

\[ Z_t = \nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}, \quad t = 3, 4, \cdots, N. \]

Example 2. Consider the time series \( X_1, X_2, X_3, X_4, X_5, X_6, X_7 \).

\[ Y_2 = X_2 - X_1. \]
\[ Y_3 = X_3 - X_2. \]
\[ Y_4 = X_4 - X_3. \]
\[ Y_5 = X_5 - X_4. \]
\[ Y_6 = X_6 - X_5. \]
\[ Y_7 = X_7 - X_6. \]

The new time series is \((Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)\).

Summary

Trend

1. Curve-fitting.

2. Filtering – moving average \((Y_t = \sum_{r=-q}^{s} a_r X_{t+r} \text{ s.t. } \sum a_r = 1)\), also convolution.

3. Differencing

- 1st order \( Y_t = \nabla X_t = X_t - X_{t-1} \).
- 2nd order \( \nabla^2 X_t = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2} \).
2.6 Time Series with Seasonal Variation

Modeling Seasonal Variation

Three Seasonal Models

\[ A: \quad X_t = \mu_t + S_t + \epsilon_t. \]
\[ B: \quad X_t = \mu_t \cdot S_t + \epsilon_t. \]
\[ C: \quad X_t = \mu_t \cdot S_t \cdot \epsilon_t. \]

\( \mu_t \) is the deseasonalized mean level at time \( t \). \( S_t \) is the seasonal effect at time \( t \). \( \epsilon_t \) is the random error.

- Generally speaking, it’s annual in period.
- Model A represents additive seasonality.
- Models B and C represent multiplicative seasonality.
- Model C: The log transformation removes the multiplicative seasonality and turns it into additive (linear).

\[ \log X_t = \log \mu_t + \log S_t + \log \epsilon_t. \]

Seasonality

Figure 4: Additive seasonality with constant additive noise (model A).

Seasonality

Seasonality
Figure 5: Multiplicative seasonality with constant additive noise (model B).

Figure 6: Multiplicative seasonality with large multiplicative noise (model C).
Seasonality

Question: measure seasonal effect or eliminate seasonality?

- little trend \(\rightarrow\) estimate seasonal effect for a particular period.
- substantial trend \(\rightarrow\) eliminate seasonal effect (filtering).

Eliminating Seasonal Effect

1 Filtering seasonal data.

**Monthly Data from January to January**

\[ Y_t = Sm(X_t) = \frac{\frac{1}{2}X_{t-6} + X_{t-5} + \cdots + X_{t+5} + \frac{1}{2}X_{t+6}}{12}. \]

Moving average over 13 successive observations. Two end coefficients are different but they sum up to unity.

*Example 3.* From April to April, on the same month.

\[ Y_{10} = Sm(X_{10}) = \frac{\frac{1}{2}X_{4} + X_{5} + X_{6} + \cdots + X_{15} + \frac{1}{2}X_{16}}{12}. \]

**Seasonality**

**Quarterly Data**

\[ Sm(X_t) = \frac{\frac{1}{2}X_{t-2} + X_{t-1} + X_t + X_{t+1} + \frac{1}{2}X_{t+2}}{4}. \]

2 Differencing to remove seasonal effect.

For the seasonal times series with month data define

\[ \nabla_{12}X_t = X_t - X_{t-12}. \]

**X-12 Method**

Used to estimate or remove both trend and seasonal variation.

2.7 Autocorrelation and the Correlogram

**Autocorrelation and Correlogram**

**Sample Correlation Coefficient**

\[
 r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \bar{Y}}{s_Y} \right) = \frac{\sum x_iy_i - n\bar{x}\bar{y}}{(n-1)sxsy}. 
\]

- If the random variables \(X\) and \(Y\) are independent then \(r = 0\).
- \(-1 \leq r \leq 1\).
- Autocorrelation: measure if successive observations are correlated.
Autocorrelation and Correlogram

Given $X_1, \cdots, X_n$ on a time series. Let 

$$(X_1, X_2), \ (X_2, X_3), \ \cdots, \ (X_{n-1}, X_n).$$

be the $(n - 1)$ pairs at lag 1. The autocorrelation at lag 1 is given by

$$r_1 = \frac{\sum_{t=1}^{n-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}.$$

Consider the $(n - 2)$ pairs at lag 2.

$$(X_1, X_3), \ (X_2, X_4), \ \cdots, \ (X_{n-2}, X_n).$$

The autocorrelation at lag 2 is given by

$$r_2 = \frac{\sum_{t=1}^{n-2} (X_t - \bar{X})(X_{t+2} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}.$$

Autocorrelation at lag $k$

The autocorrelation at lag $k$ is given by

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}.$$

The auto-covariance at lag $k$ is given by

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) = \text{cov}(X_t, X_{t+k}).$$

$$c_0 = \frac{1}{n} \sum_{t=1}^{n} (X_t - \bar{X})^2 = \text{var}(X_t).$$

The autocorrelation at lag $k$ may be written as

$$r_k = \frac{c_k}{c_0} = \frac{\text{cov}(X_t, X_{t+k})}{\text{var}(X_t)}.$$

Use the sample to compute $\text{cov}(X_t, X_{t+1}), \text{cov}(X_t, X_{t+2}), \cdots, \text{cov}(X_t, X_{t+k}).$

The Correlogram

Definition

A graph of the sample autocorrelation coefficients, also called the sample autocorrelation function (ac.f.).

- $-1 \leq r_k \leq 1.$
- $r_k$ is plotted against the lag $k$ for $k = 0, 1, \cdots, m,$ where $m \leq n.$
Figure 7: Correlogram ($r_k$ vs $k$).

**The Correlogram**

R code for generating a random time series and plotting the correlogram.

```r
# Generate 100 normally distributed random numbers plus sin
x <- rnorm(100) + sin(2*pi*1:100/10)

win.graph(width=3, height = 4, pointsize=8)
par(mfrow=c(2,1), mar=c(3.5, 3.5, 0.5, 0.5), mgp=c(2, 0.8, 0), cex=.7)
# Start the PNG writing device
png(filename="H:/System/Desktop/Stat420/Lecture Notes/autocorr.png")
# Plot the series
plot(x, type="o", col="blue", main=NA)
# Plot the autocorrelation function
acf(x, 100, main=NA)
# Turn off the device
dev.off()
```

Correlogram

Figure 8: Correlogram of a time series with confidence limits.

2.8 Interpreting the Correlogram

Random Series
- Interpreting autocorrelation coefficients is not always easy.
- A series is called random if it consists of independent observations having the same distribution.
- For large $n$ we expect $r_k \approx 0 \forall k \neq 0$
- For a completely random time series
  $$r_k \sim N \left( 0, \frac{1}{n} \right).$$
- 95% C.I. for $r_k$ is $\left( -\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}} \right)$.
  - We expect 19 out of 20 of the values of $r_k$ between $\pm \frac{2}{\sqrt{n}}$.
- If many $r_k$ fall beyond the C.I. it is not random.

**Short-Term Correlation**

Figure 9: A time series showing short-term correlation together with its correlogram.

- Stationary series exhibit short-term correlation.
- Values of $r_k$ for longer lags tend to be approximately zero.
- An observation above the mean tends to be followed by one or more further observations above the mean and similarly for observations below the mean.

**Alternating Series**

Figure 10: An alternating time series together with its correlogram.
• Successive observations on different sides of the overall mean.

• Series tends to alternate $\rightarrow$ correlogram tends to alternate.

**Seasonal Series**

![Correlogram of a seasonal time series](image)

Figure 11: Correlogram of a seasonal time series.

• If series contains seasonal variation then the correlogram will exhibit oscillation at the same frequency.

• If $X_t = a \cos t\omega$, then $r_k \approx \cos k\omega$ for large $n$. $X_t$ shows sinusoidal pattern.