STAT302: Time Series Analysis

Chapter 2. Time Series Basics

Sangbum Choi, Ph.D

Department of Statistics, Korea University

Outline

Fundamental features of time series

- A fundamental feature of time series is that values X_t at different times tend to be related in certain ways.
- That is, they tend not to be independent.
- Time series analysis is aimed at studying and characterizing the nature of relationship in X_t's over time.

Time series as stochastic process

- Univariate time series - single time series.
- Bivariate, multivariate time series - two or more time series.
- A discrete time series $\{X_1, X_2, \ldots, X_T\}$ is a sequence of random variables (RVs), which has a joint probability distribution.
- The joint probability distribution may be represented by a joint distribution function

$$F(x_1, x_2, \ldots, x_T) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_T \leq x_T)$$

Time series as stochastic process

- In a statistical setting, we obtain a sample realization (denoted by $x_1, x_2, \ldots, x_T$) from the stochastic process (denoted by $X_1, X_2, \ldots, X_T$) and then use the sample to estimate/infer some of the probability characteristics of the stochastic process.
- It is impossible to obtain "accurate" estimate of the complete joint probability distribution of $X_1, X_2, \ldots, X_T$, unless we make very strong assumptions.
- One approach is to restrict interest to certain summary measures of the joint probability distribution.

Summary measures

- Mean function:

$$\mu_t = E(X_t) = \int_{-\infty}^{\infty} x_t f_t(x_t)\, dx_t$$

  for $t = \ldots, -1, 0, 1, \ldots$
- Variance function:

$$\sigma_t^2 = E[(X_t - \mu_t)^2] = \int_{-\infty}^{\infty} (x_t - \mu_t)^2 f_t(x_t)\, dx_t$$

Summary measures

- Covariance function:

$$\gamma(t_1, t_2) = \text{Cov}(X_{t_1}, X_{t_2}) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_{t_1} - \mu_{t_1})(x_{t_2} - \mu_{t_2}) f(x_{t_1}, x_{t_2})\, dx_{t_1} dx_{t_2}$$

- Correlation function:

$$\rho(t_1, t_2) = \text{Cor}(X_{t_1}, X_{t_2}) = \frac{\text{Cov}(X_{t_1}, X_{t_2})}{\sqrt{\text{Var}(X_{t_1})}\sqrt{\text{Var}(X_{t_2})}}$$
$$= \frac{\gamma(t_1, t_2)}{\sigma_{t_1} \sigma_{t_2}}$$

is always between –1 and 1 and provides a measure of extent of linear relation between the RVs X_{t_1} and X_{t_2}.

Weak stationarity

A process $\{X_t\}$ is weakly (or second-order) stationary if it satisfies,

(1) $E(X_t) = \mu$ does not depend on t.

(2) $\mathrm{Var}(X_t) = \gamma(0)$ does not depend on t.

(3) $\mathrm{Cov}(X_t, X_{t+k}) = \gamma(k)$ depends only on lag k, not on t.

Strict stationarity

- Consider finite time indices

$$t_1, t_2, \cdots t_k \quad \text{and} \quad t_1 + h, t_2 + h, \cdots, t_k + h, \; (h > 0).$$

- For all $(t_1, t_2, \cdots t_k)$ and for all $h > 0$, if

$$\begin{aligned} P(X_{t_1} \le x_1, X_{t_2} \le x_2, \cdots, X_{t_k} \le x_k) \\ = P(X_{t_1+h} \le x_1, X_{t_2+h} \le x_2, \cdots, X_{t_k+h} \le x_k) \end{aligned}$$

then $\{X_t\}$ is said to be **strictly (or strongly) stationary**.

Gaussian processes

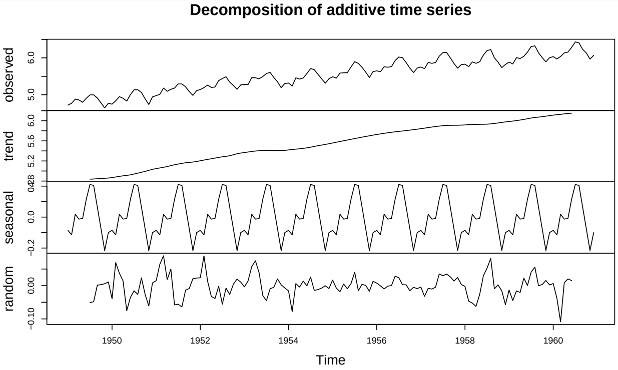
- A process $\{X_t\}$ is Gaussian, if the joint probability distributions of all finite sets of RV $\boldsymbol{X}_{n\times 1} = (X_{t_1}, X_{t_2}, \ldots, X_{t_n})'$ are mulitvariate normal distributions.
- Recall that a random vector $\boldsymbol{X}_{n\times 1}$ has multivariate normal distribution with mean vector $\boldsymbol{\mu}_{n\times 1}$, covariance matrix $\boldsymbol{\Gamma}_{n\times n}$ if its joint pdf is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Gamma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

- Note that if the process $\{X_t\}$ is weakly stationary and is Gaussian, then it is strictly stationary. [Why?]

Stationary and nonstationary processes

```r
library(tidyverse)
AirPassengers %>% log %>% decompose %>% plot
```



Autocorrelation function (ACF)

- For a stationary process $\{X_t\}$,

$$\rho(k) = \mathrm{Cor}(X_t, X_{t+k}) = \frac{\gamma(k)}{\gamma(0)}, k = 0, \pm 1, \pm 2, \ldots$$

  is the autocorrelation function (ACF).
- The ACF $\rho(k)$ is of prime interest in the study of stationary process, because it gives a summary of relations between values at different time lags.

Basic properties of autocorrelation function

1. $-1 \leq \rho(k) \leq 1$ for all k with $\rho(0) = 1$.
2. $\gamma(k)$ and $\rho(k)$ are even functions. That is, $\gamma(-k) = \gamma(k)$, $\rho(-k) = \rho(k)$. [Why?]
3. $\{\gamma(k)\}$ and $\{\rho(k)\}$ are positive-(semi)definite sequences in that they satisfy, for every n, times $t_1, t_2, \ldots, t_n$, and constants $c_1, c_2, \ldots, c_n$,
$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \gamma(i - j) > 0.$$
[Why?]

Outline

Mean and covariance formulas

Let $X_1, X_2, \ldots, X_n$ be RVs, $c_1, c_2, \ldots, c_n$ be constants. Then

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i E(X_i)$$

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^n c_i X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n c_i c_j \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n c_i^2 \mathrm{Var}(X_i) + 2 \sum_{i<j} c_i c_j \mathrm{Cov}(X_i, X_j)\end{aligned}$$

$$\mathrm{Cov}\left(\sum_{i=1}^n c_i X_i, \sum_{j=1}^n d_j X_j\right) = \sum_{i=1}^n \sum_{j=1}^n c_i d_j \mathrm{Cov}(X_i, X_j)$$

Mean and covariance formulas

- If X_i are independent RVs, then $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$. Thus

$$\text{Var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 \text{Var}(X_i)$$

$$\text{Cov}\left(\sum_{i=1}^{n} c_i X_i, \sum_{j=1}^{n} d_j X_j\right) = \sum_{i=1}^{n} c_i d_i \text{Var}(X_i)$$

Stationary process: Example 1

White noise (WN):

- Suppose $\epsilon_0, \epsilon_1, \ldots, \epsilon_t, \ldots$ is a sequence of independent RVs with common mean $E(\epsilon_t) = 0$, common variance $\text{Var}(\epsilon_t) = \sigma^2$ for all t.
- Then $\{\epsilon_t\}$ is a (weakly) stationary process and is called a white noise process.

Stationary process: Example 1

- $E(\epsilon_t) = 0$ does not depend on t.
- $\text{Var}(\epsilon_t) = \sigma^2$ does not depend on t.
- The autocovariance function is

$$\text{Cov}(\epsilon_t, \epsilon_{t+k}) = \gamma(k) = \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

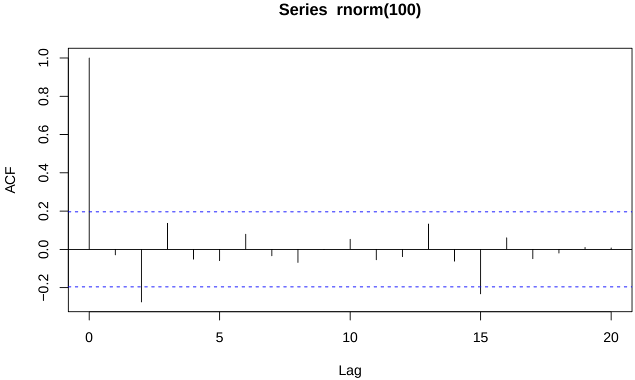
  depends only on k, not on t.
- The ACF of a white noise process is

$$\rho(k) = \text{Cor}(\epsilon_t, \epsilon_{t+k}) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

Stationary process: Example 1

- How does the ACF plot of a white noise process look like?

```
acf(rnorm(100))
```



Stationary process: Example 2

Let $\{\epsilon_t\}$ denote a white noise process. Define a process

$$X_t = \mu + \epsilon_t + \epsilon_{t-1},\ t = \ldots, -1, 0, 1, \ldots,$$

where μ is a constant.

- $\{X_t\}$ is a stationary process.
- $E(X_t) = \mu$

Stationary process: Example 2

- The autocovariance function is

$$\mathrm{Cov}(X_t, X_{t+k}) = \gamma(k) = \begin{cases} 2\sigma^2 & \text{if } k = 0 \\ \sigma^2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

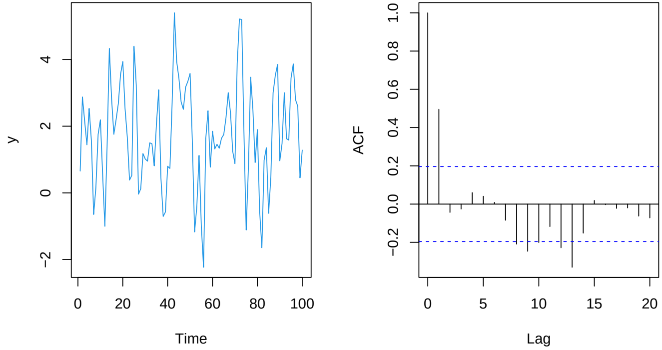
  depends only on k, not on t.
- The ACF is

$$\rho(k) = \mathrm{Cor}(X_t, X_{t+k}) = \begin{cases} 1 & \text{if } k = 0 \\ 1/2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

Stationary process: Example 2

- Now how does ACF plot look like?

```
y = eps = rnorm(100)
for (i in 2:100) y[i] = 2 + eps[i] + eps[i-1]
par(mfrow=c(1,2))
ts.plot(y, col=4); acf(y,main="")
```



Stationary process: Example 2'

Let $\{\epsilon_t\}$ denote a white noise process. Define a process

$$X_t = \mu + \epsilon_t - \epsilon_{t-1}, \quad t = \dots, -1, 0, 1, \dots,$$

where μ is a constant.

- $\{X_t\}$ is a stationary process.
- $E(X_t) = \mu$

Stationary process: Example 2'

- The autocovariance function is

$$\mathrm{Cov}(X_t, X_{t+k}) = \gamma(k) = \begin{cases} 2\sigma^2 & \text{if } k = 0 \\ -\sigma^2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

depends only on k, not on t.

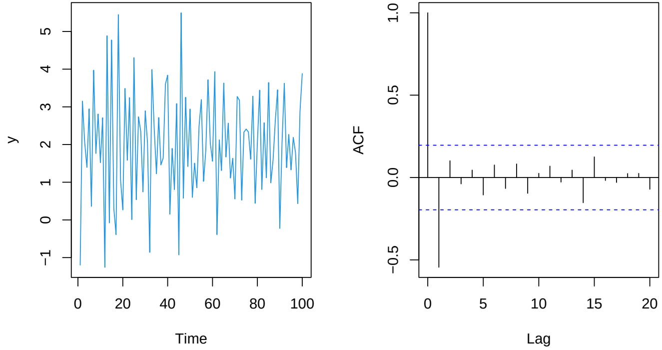
- The ACF is

$$\rho(k) = \mathrm{Cor}(X_t, X_{t+k}) = \begin{cases} 1 & \text{if } k = 0 \\ -1/2 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

Stationary process: Example 2

- Now how does ACF plot look like?

```
y = eps = rnorm(100)
for (i in 2:100) y[i] = 2 + eps[i] - eps[i-1]
par(mfrow=c(1,2))
ts.plot(y, col=4); acf(y,main="")
```



Stationary process: Example 3

- Suppose $\{\epsilon_t\}$ denote a white noise process. Define a process

$$X_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \ldots + \psi_q \epsilon_{t-q},$$

  where $\mu, \psi_1, \psi_2, \ldots, \psi_q$ are constants.
- Note that X_t is called an MA(q) process.
- For example, let $q = 2$, i.e., when the model is MA(2),

$$X_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2}, \ \epsilon_t \sim \mathrm{WN}(0, \sigma_\epsilon^2)$$

- Show that $\{X_t\}$ is weakly stationary.

Nonstationary process: Random walk

Let $\{\epsilon_t\}$ denote a white noise process. Define a process $\{X_t\}$ by recursive equation

$$X_t = X_{t-1} + \delta + \epsilon_t, \quad t = 1, 2, \ldots,$$

where

- δ is a constant.
- $X_0 = 0$ (or some other assumption is needed)

Then $\{X_t\}$ is nonstationary.

Nonstationary process: Random walk

What are the means, variances, covariances, etc of the process $\{X_t\}$?
Express X_t explicitly in terms of the white noise process ϵ_t

$$
\begin{aligned}
X_t &= X_{t-1} + \delta + \epsilon_t \\
&= X_{t-2} + \delta + \epsilon_{t-1} + \delta + \epsilon_t \\
&= X_{t-2} + 2\delta + \epsilon_{t-1} + \epsilon_t \\
&= X_{t-3} + \delta + \epsilon_{t-2} + 2\delta + \epsilon_{t-1} + \epsilon_t \\
&= X_{t-3} + 3\delta + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\
&= \ldots \\
&= X_0 + t\delta + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_{t-1} + \epsilon_t \\
&= X_0 + t\delta + \sum_{i=0}^{t-1} \epsilon_{t-i} \\
&= X_0 + t\delta + \sum_{j=1}^{t} \epsilon_j
\end{aligned}
$$

Nonstationary process: Random walk

- Thus,

$$
\begin{aligned}
E(X_t) &= E\left(X_0 + t\delta + \sum_{i=0}^{t-1} \epsilon_{t-i}\right) \\
&= t\delta + \sum_{i=0}^{t-1} E(\epsilon_{t-i}) = t\delta \\
\mathrm{Var}(X_t) &= \mathrm{Var}\left(X_0 + t\delta + \sum_{i=0}^{t-1} \epsilon_{t-i}\right) \\
&= \sum_{i=0}^{t-1} \mathrm{Var}(\epsilon_{t-i}) = t\sigma^2
\end{aligned}
$$

- Note that both mean and variance depend on time t. This means that the mean and variance become explosive as $t \to \infty$.

Nonstationary process: Random walk

- Also for $k > 0$,

$$X_{t+k} = X_0 + (t+k)\delta + \sum_{i=0}^{t+k-1} \epsilon_{t+k-i}$$

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t+k}) &= \mathrm{Cov}\left(\sum_{i=0}^{t-1} \epsilon_{t-i}, \sum_{i=0}^{t+k-1} \epsilon_{t+k-i}\right) \\ &= \sum_{i=0}^{t-1} \mathrm{Cov}(\epsilon_{t-i}, \epsilon_{t-i}) = t\sigma^2\end{aligned}$$

Nonstationary process: Random walk

- Thus, note that, for $k > 0$,

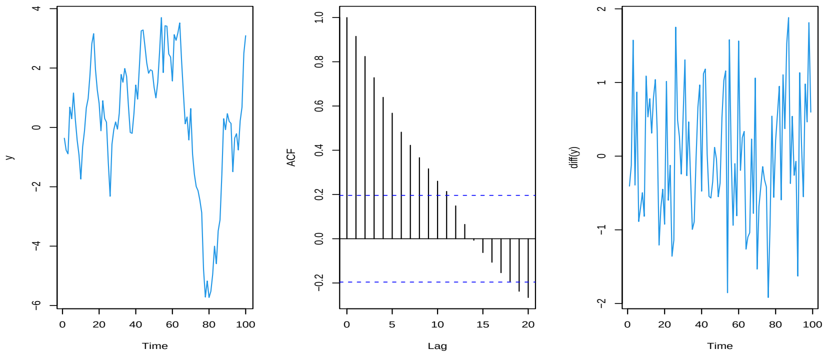
$$
\begin{aligned}
\mathrm{Cor}(X_t, X_{t+k}) &= \frac{\mathrm{Cov}(X_t, X_{t+k})}{\sqrt{\mathrm{Var}(X_t)}\sqrt{\mathrm{Var}(X_{t+k})}} \\
&= \frac{t\sigma^2}{\sqrt{t\sigma^2}\sqrt{(t+k)\sigma^2}} \\
&= \frac{t}{\sqrt{t(t+k)}},
\end{aligned}
$$

  which is approximately 1 when k is small relative to t.
- Notice that the correlation tends to 1 as $t \to \infty$ as long as k is finite.

Nonstationary process: Random walk

- A RW process is not stationary. Now how does ACF plot look like? Notice that the difference of RW becomes stationary.

```
y = cumsum(rnorm(100)) # RW
par(mfrow=c(1,3))
ts.plot(y, col=4); acf(y,main=""); ts.plot(diff(y), col=4)
```



Nonstationary process: Random walk

- Nearby values of a random walk process tend to be very highly positively correlated.
- This leads to very smooth behavior over time in the process.
- If $\delta = 0$, $E(X_t) = 0$. If $\delta > 0$, there is a linear trend $E(X_t) = \delta t$.
- The constant δ is a drift parameter.
- There is a simple relation to a stationary process by considering the first differences of $\{X_t\}$ with

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

  which is stationary.
- Thus the first differences are a stationary process.

Homogeneous nonstationary process

- The random walk process above will be later referred to as a **homogeneous nonstationary process.**
- Such processes are also referred to as **integrated processes**.
- Define a more general integrated process as

$$W_t = X_t - X_{t-1}$$

  where W_t is a stationary process (i.e. not necessarily white noise).
- Thus we can represent X_t as

$$X_t = X_0 + W_1 + \cdots + W_t$$

  where $W_1 + \cdots + W_t$ is integration (or sum) of a stationary process $\{W_t\}$

Deterministic mean/trend function

- Suppose $\{Z_t\}$ is a stationary process with mean 0.
- Define

$$X_t = \mu_t + Z_t$$

  where μ_t is a nonrandom function of t.
- For example,

$$\mu_t = \alpha + \delta t$$

  is a linear trend function of t.
- Since $E(X_t) = \mu_t$ depends on t, the process $\{X_t\}$ is not stationary.
- However, by assumption, the deviations $Z_t = X_t - \mu_t$ is stationary.

Analysis of nonstationary processes

- Thus analysis of such a series $\{X_t\}$ would typically involve both methods of regression analysis to model the linear function and (stationary) time series methods to model the "noise" $\{Z_t\}$.
- For the linear trend example, consider the first differences of $\{X_t\}$.

$$\begin{aligned} W_T &= X_t - X_{t-1} = [\alpha + \delta t + Z_t] - [\alpha + \delta(t-1) + Z_{t-1}] \\ &= \delta + Z_t - Z_{t-1} \end{aligned}$$

- Here the process $\{W_t\}$ forms a stationary time series. [Why?]
- We may say that the first differencing "removes" a linear trend component or "reduces" the series to stationarity.

Analysis of nonstationary processes

- We also will study processes that have nonstationary behavior in a “seasonal sense”.
- For example, for a monthly time series that exhibits annual seasonal behavior, we may analyze the series by considering the seasonal differences.
- Consider

$$X_t = \mu + \beta_1 \cos(2\pi t/12) + \beta_2 \sin(2\pi t/12) + Z_t$$

  which has a period of 12.
- Take the seasonal difference

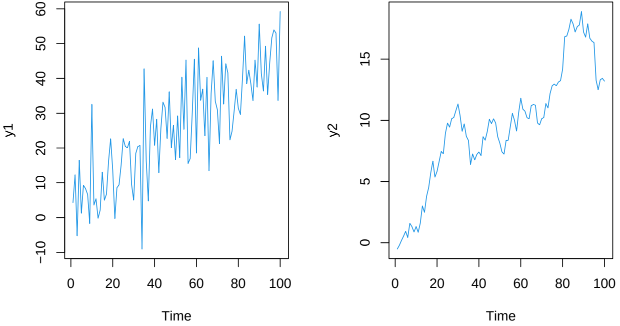
$$W_t = X_t - X_{t-12} = Z_t - Z_{t-12}$$

  which forms a stationary time series. [Why?]

Deterministic vs. stochastic trends

Deterministic and stochastic trends are different in data generation.

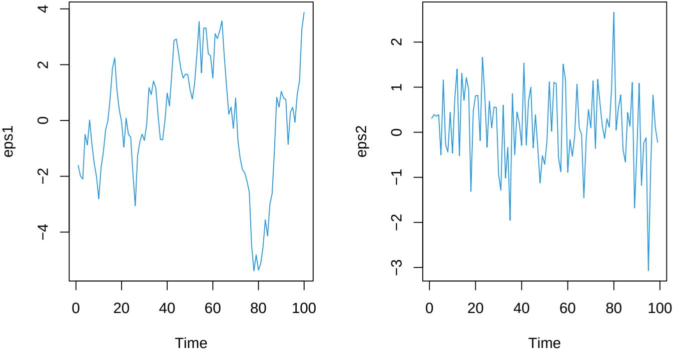
```
time = 1:100
y1 = 0.5 * time + 10*rnorm(100) # deterministic trend
y2 = cumsum(rnorm(100)) # stochastic trend
par(mfrow=c(1,2))
ts.plot(y1, col=4); ts.plot(y2, col=4)
```



Deterministic vs. stochastic trends

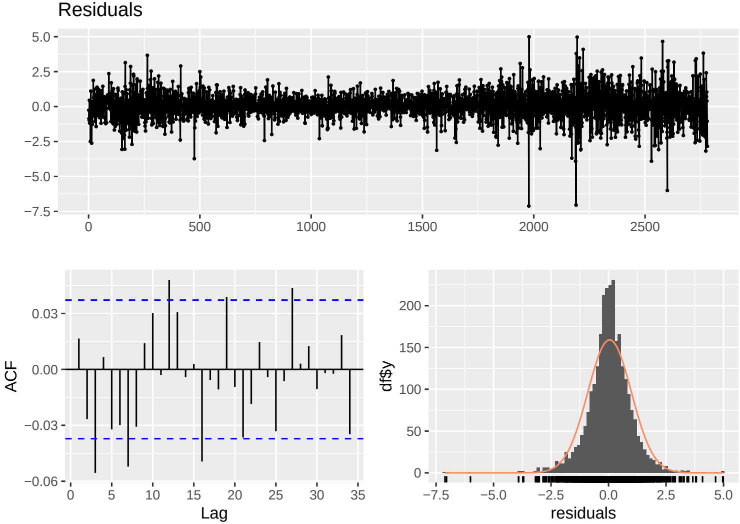
How we can detend deterministic and stochastic trends?

```r
eps1 = lm(y ~ time)$residual
eps2 = diff(y2)
par(mfrow=c(1,2))
ts.plot(eps1, col=4); ts.plot(eps2, col=4)
```



Example: S&P 500 index

```r
library(MASS); library(forecast)
checkresiduals(SP500)
```




```
##
```

Outline

Summary measures for stationary process

Let $X_1, X_2, \ldots, X_T$ denote a stationary process with

- Mean

$$\mu = E(X_t)$$

- Autocovariance

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

- ACF

$$\rho(k) = \text{Cor}(X_t, X_{t+k}) = \frac{\gamma(k)}{\gamma(0)}$$

Estimation of summary measures

- Sample mean

$$\hat{\mu} = \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} X_t$$

- Sample autocovariance

$$\hat{\gamma}(k) = \frac{1}{T}\sum_{t=1}^{T-k} (X_t - \bar{Y})(X_{t+k} - \bar{Y})$$

is an estimate of $\gamma(k)$ for $k = 0, 1, 2, \ldots, K$, and usually K is small relative to T.

Estimation of summary measures

- In particular, sample variance

$$\hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{Y})^2$$

is an estimate of $\gamma(0)$.

- Sample ACF

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{T-k}(X_t - \bar{Y})(X_{t+k} - \bar{Y})}{\sum_{t=1}^{T}(X_t - \bar{Y})^2}$$

is an estimate of $\rho(k)$ for $k = 0, 1, 2, \ldots, K$, and usually K is small relative to T.

Sample ACF

- Asymptotic properties of $\hat{\rho}(k)$ can be derived. In general they are too complicated for practical use, but special cases are useful.
- Consider a white noise process. Then it can be shown that

$$E(\hat{\rho}(k)) \approx \rho(k) = 0 \quad \text{and} \quad \mathrm{Var}(\hat{\rho}(k)) \approx \frac{1}{T}$$

$$\hat{\rho}(k) \approx N\left(0, \frac{1}{T}\right)$$

for all $k = 1, 2, \ldots$.

- In fact,

$$\hat{\rho}(1), \ldots, \hat{\rho}(K) \sim \text{approx iid } N\left(0, \frac{1}{T}\right)$$

Sample ACF

- An approximate 95% confidence interval for $\rho(k)$ is

$$\hat{\rho}(k) \pm 1.96 \frac{1}{\sqrt{T}}$$

- It is a common practice to use $\pm \frac{2}{\sqrt{T}}$ as limits to assess the significance of $\rho(k)$ in terms of deviation from 0.
- For example, when $T = 100$, the limits are ± 0.2. Suppose $\hat{\rho}(1) = 0.45$. Then the time series is not compatible with the white noise assumption.

Jointly stationary processes

- Two time series, X_t and Y_t, are said to be **jointly stationary** if they are each stationary, and the **cross-covariance function**

$$\gamma_{XY}(h) = \mathrm{Cov}(X_{t+h}, Y_t) = E[(X_{t+h} - \mu_X)(Y_t - \mu_Y)]$$

  is a function only of lag h.
- The **cross-correlation function (CCF)** of jointly stationary time series X_t and Y_t is defined as

$$\rho_{XY}(h) = \frac{\gamma_{XY}(h)}{\sqrt{\gamma_X(0)\gamma_Y(0)}}$$

- Note that $\rho_{XY}(h) = \rho_{YX}(-h)$.

Jointly stationary processes

- For example, let

$$X_t = \epsilon_t + \epsilon_{t-1}, \quad Y_t = \epsilon_t - \epsilon_{t-1},$$

where $\epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2)$.

- It can be shown that

$$\rho_{XY}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

- Clearly, the CCF depends only on the lag separation h, so the series are jointly stationary.

Outline

Linear filter

- A **(time invariant) linear filter** is a linear operation applied to a series $\{X_t\}$ to produce a new series $\{Y_t\}$, as

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j},$$

  where $t = ..., -1, 0, 1, ...$
  - $\{\psi_j\}$ are coefficients of the linear filter.
  - X_t and Y_t can be thought of as the input and output, respectively.

Linear filter

- Note that the filter is time-invariant, in that the coefficients $\{\psi_j\}$ do not depend on t.
- The filter is stable if $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ (i.e. $\{\psi\}$ are absolutely summable).
- The filter is one-sided, phisically realizable, or causable, or causal if $\psi_j = 0$ for $j < 0$ so that

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}.$$

Linear filter: Two main contexts

1. Given a single time series $\{X_t\}$, we choose a linear filter and apply $\{X_t\}$ to obtain a new series $\{Y_t\}$. The purpose is to obtain a new series with desired characteristics. For example,
   - $Y_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$ can be thought of as an averaging filter or smoothing filter.
   - $Y_t = X_t - X_{t-1}$ is the first difference.
2. Two distinct series $\{X_t\}$ and $\{Y_t\}$ are in a dynamic system where X_t is input and Y_t is output. The linear filter could be viewed as a model to represent the relationship between Y_t and X_t. Then the filter is unknown and can be modeled as

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j} + noise$$

Backshift and linear filter operator

- A backshift operator B operates on $\{Y_t\}$ such that

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

- Thus for $t = \ldots, -1, 0, 1, \ldots,$

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j} = \sum_{j=0}^{\infty} \psi_j B^j X_t = \left(\sum_{j=0}^{\infty} \psi_j B^j \right) X_t = \psi(B) X_t$$

where

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$$

is the linear filter operator.

Theorem (on linear filter and stationarity)

If $\{X_t\}$ is a stationary series with mean μ_x and autocovariance γ_x and $\{Y_t\}$ is the output of a time-invariant, absolutely summable, and causal linear filter with $\{X_t\}$ as input, i.e.,

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j},$$

where $\sum_{j=0}^{\infty} |\psi_j| < \infty$, then the output $\{Y_t\}$ is a **stationary** process with

$$\mu_Y = E(Y_t) = \mu_X \sum_{j=0}^{\infty} \psi_j$$

$$\gamma_Y(k) = \mathrm{Cov}(Y_t, Y_{t+k}) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_X(i - j + k)$$

Linear processes

- A process $\{Y_t\}$ is a **linear process** if it is representable as

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \mu + \psi(B)\epsilon_t$$

where $\{\epsilon_t\}$ is a white noise process (i.e. ϵ_t are independent RVs with mean 0 and variance σ^2 for all t) and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Linear processes

- FACT: If $\{Y_t\}$ is a linear process, then $\{Y_t\}$ is stationary.
- The autocovariances of $\{Y_t\}$ are

$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t+k}) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{k+i},\ k = 0, 1, 2, ...$$

- In particular,

$$\gamma(0) = \mathrm{Var}(Y_t) = \sigma^2 \sum_{i=0}^{\infty} \psi^2$$

Example of linear processes

- Consider a process $\{Y_t\}$:

$$Y_t = \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$$

  where ϕ is a constant.
- If $|\phi| < 1$, then the process $\{Y_t\}$ is stationary.
- The condition $|\phi| < 1$ will be called the stationarity condition.
- If $\phi = 1$, what is the process $\{Y_t\}$?

Example of linear processes

- Note that
$$\sum_{j=0}^{\infty} |\phi^j| = \frac{1}{1-|\phi|} < \infty$$
if and only if $|\phi| < 1$.
- If $|\phi| < 1$, the autocovariances of $\{Y_t\}$ are
$$\gamma(0) = \mathrm{Var}(Y_t) = \sigma^2 \left(\frac{1}{1-\phi^2}\right)$$
$$\gamma(k) = \mathrm{Cov}(Y_t, Y_{t+k}) = \sigma^2 \left(\frac{\phi^k}{1-\phi^2}\right), \quad k = 0, 1, 2, \ldots$$

Example of linear processes

- Thus, if $|\phi| < 1$, the ACF of $\{Y_t\}$ is

$$\rho(k) = \mathrm{Cor}(Y_t, Y_{t+k}) = \frac{\gamma(k)}{\gamma(0)} = \phi^k,\ k = 0, 1, 2, ...$$

- That is, the ACF has the form of simple exponential decay.
- The process $\{Y_t\}$ as defined satisfies the equation

$$Y_t = \phi Y_{t-1} + \delta + \epsilon_t$$

  where $\delta = \mu(1 - \phi)$.
- The process $\{Y_t\}$ is called a first-order autoregressive (i.e., AR(1)) process and is stationary if $|\phi| < 1$.