STATS 3005 Time Series III
Lecture notes

Lecturer: Patty Solomon

School of Mathematical Sciences, University of Adelaide

Semester 2, 2018

Time series

In the simplest statistical problems, we usually consider data

$$y_1, y_2, \ldots, y_n$$

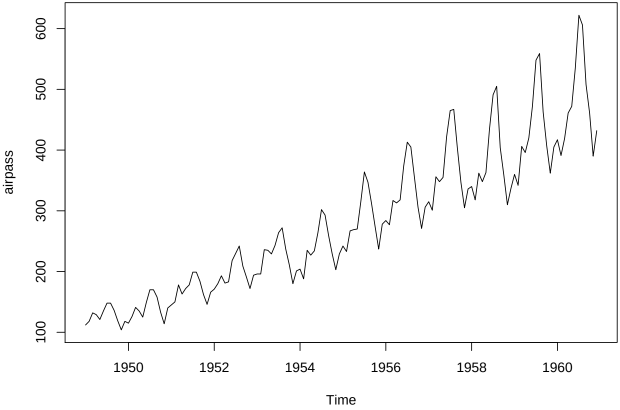
assumed to be realisations of independent, identically distributed random variables.

A univariate time series consists of observations, y_t, of a single variable made repeatedly over time. Most often, it is assumed that

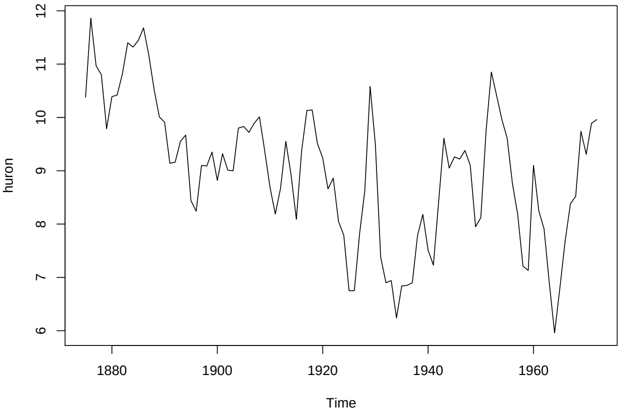
- The observations are equally spaced in time so that $t = 1, 2, 3, \ldots$
- A single measurement is made at each time.

Unlike in the IID case, time series data cannot be assumed independent.

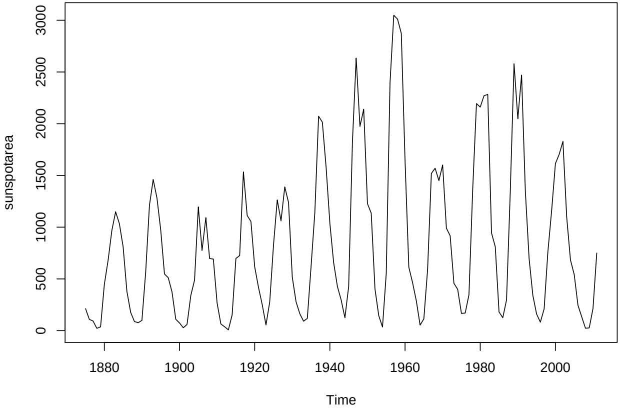
Example: the air passengers data



Example: the level of Lake Huron 1875-1972



Example: area of sunspots 1875-2011



Common features in time series data

Trend an apparent systematic component;

Seasonality cycles corresponding to known period;

Cyclical Fluctuation cycles of period not known in advance;

Residual effects unexplained variability.

Objectives of time series analysis

- Describing the series:
  - graphically;
  - summary statistics.
- Forecasting;
- Comparing two or more series;
- Assessing interventions;
- Comparing treatments;
- Modelling the data.

Notation

- Use upper case Y to denote random variables and lower case y to denote realisations.
- A time series $\{y_t : \ t = 1, 2, \ldots, n\}$ is the set of realised values of $\{Y_t : \ t = 1, 2, \ldots, n\}$.
- For continuous time processes, represent as $y(t)$ and $Y(t)$ respectively.
- Although observed series are finite, it is mathematically convenient to treat them as infinitely extendable.

Times series data

- Discrete time series often arise from an underlying continuous time process.
- This may be simply through sampling:
  - $Y_t = X(t)$ for $t = 1, 2, \ldots, n$;
  - For example, the Lake Huron data.
- It may also be through aggregation:
  - $Y_t = \int_{t-1}^{t} X(s) ds$;
  - For example, the air passenger data.
- The Y_t's may also be discrete or continuous variables.

Moments

- We use μ to denote the mean of a time series.
  - $\mu(t) = E(Y(t))$ for continuous time processes;
  - $\mu_t = E(Y_t)$ for discrete time.
- The autocovariance function is defined by

$$\gamma(s, t) = E[(Y(s) - \mu(s))(Y(t) - \mu(t))]$$

  - $\gamma(t, t) = \text{var}(Y(t))$;
  - $\gamma(s, t) = \text{cov}(Y(s), Y(t))$.

Stationarity

Strict Stationarity

A random process $\{Y(t)\}$ is said to be strictly stationary if for any set of times $t_1, t_2, \ldots, t_k$, the joint distribution of

$$(Y(t_1 + \tau), Y(t_2 + \tau), \ldots, Y(t_k + \tau))$$

is the same for all τ.

The property of stationarity implies that, in terms of its probability distribution, the process $\{Y(t)\}$ looks the same at all times.

For a stationary process:

- $\mu(t) = \mu$ for all t;
- $\gamma(s, t)$ depends on s, t only through $|t - s|$;
- The auto covariance function is usually written $\gamma(|t - s|)$.

Marginal Stationarity

A random process $\{Y(t)\}$ is said to satisfy marginal stationarity if the distribution $f(y(t))$ is the same for all t.

A strictly stationary process also satisfies marginal stationarity but not vice versa.

Marginal stationarity implies $\mu(t) = \mu$ for all t.

Exercise

Construct an example of a discrete time process that satisfies marginal stationarity but not strict stationarity.

Stationarity

Second order stationarity

A random process $\{Y(t)\}$ is said to satisfy second order stationarity if the mean $\mu(t+\tau)$ and the auto covariance function, $\gamma(s+\tau, t+\tau)$ do not depend τ.

For a second order stationary process, we write

- $\mu(t) = \mu$;
- $\gamma(\tau) = E[(Y(t) - \mu)(Y(t+\tau) - \mu)]$.

Strict stationarity implies second order stationarity but not vice versa.

In practice, second order stationarity is the most commonly considered assumption.

For this reason, some texts use the term “stationary” to mean “second order stationary”.

Smoothing Time Series

A common way to improve the appearance of a time series plot is to apply smoothing.

A rationale for smoothing is as follows.

Suppose the series is of the form

$$Y(t) = \mu(t) + U(t)$$

where $\mu(t)$ is the deterministic signal and $U(t)$ is a noise process.

Typically we assume $U(t)$ to be stationary with $E(U(t)) = 0$ and also that $\mu(t)$ varies smoothly with time.

Smoothing (continued)

Assuming also that the noise process is “rough” we might then argue:

- Applying smoothing to the noise process $U(t)$ could be expected to reduce the amplitude of the noise process.
- Applying appropriate smoothing to the (already) smooth signal $\mu(t)$ should not degrade the signal.
- Assuming the effect of the smoother to be linear, applying it to the observed series $y(t)$ could be expected to preserve the signal but reduce the noise.
  - A high level of smoothing could be expected to reduce the noise further;
  - Over-smoothing will also degrade the signal;
  - The desired level of smoothing is often chosen subjectively.

Linear Filters

Consider a discrete time series $\{y_t : t = 1, 2, \ldots, n\}$.

A simple way to smooth the series is to apply a three-point moving average.

That is, we construct the smoothed series
$\{s_t : \; t = 2, 3, \ldots, n-1\}$ defined by

$$s_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}.$$

More generally, a linear filter of order $2m+1$ is defined by

$$s_t = \sum_{j=-m}^{m} w_j y_{t+j}$$

where $\sum_{j=-m}^{m} w_j = 1$.

Linear Filters (continued)

- Linear filters of odd order are generally used as they allow for the alignment of s_t with y_t.
- In most cases, the weights are chosen symmetrically. That is, $w_j = w_{-j}$.
- With this definition, the smoothed series s_t is defined only for $m < t < n - m$.
  - One solution is to not calculate those points (as is the default for the `filter` function in `R`).
  - A simple alternative is to renormalise the weights near the end points.

Moving Average Smoothers

The moving average of order $2m+1$ is defined by taking

$$w_j = \frac{1}{2m+1}.$$

To illustrate how this smoother might improve the signal to noise ratio, consider a time series

$$Y_t = \mu_t + U_t$$

where

$$U_1, U_2, \ldots, U_n$$

are independent, identically distributed $N(0, \sigma^2)$ variables.

Note

The error process $\{U_t : \; t = 1, 2, \ldots, n\}$ is referred to as Gaussian white noise.

Moving Average Smoothers (continued)

If the moving average of order $2m+1$ is applied, the smoothed series is

$$S_t = \bar{\mu}_t + \bar{U}_t$$

where

$$\bar{\mu}_t = \frac{1}{2m+1} \sum_{j=-m}^{m} \mu_{t+j} \text{ and } \bar{U}_t = \frac{1}{2m+1} \sum_{j=-m}^{m} U_{t+j}.$$

The Smoothed Signal

Suppose μ_t was linear in t so that

$$\mu_t = a + bt.$$

It follows that

$$\bar{\mu}_t = \frac{1}{2m+1}\sum_{j=-m}^{m}(a + b(t+j)) = a + bt$$

so a purely linear signal would be preserved.

In general, the signal μ_t would be assumed "smooth" rather than linear.

If μ_t is approximately linear over the smoothing window $(t-m, t+m)$ the effect of smoothing should be minimal.

The Smoothed Signal (continued)

For a differentiable function, $\mu(t)$, we would expect this to be the case provided the sampling frequency is adequate.

Even if the signal is not approximately linear in the smoothing window, the degradation may be minor provided the variation within the smoothing window is small compared to the overall variation.

Mean Squared Error

Under the assumption of Gaussian white noise, it follows that

$$\bar{U}_t \sim N\left(0, \frac{\sigma^2}{2m+1}\right).$$

The mean squared error is

$$E\left(\frac{1}{n}\sum_t (Y_t - \mu_t)^2\right) = \sigma^2$$

for the raw data and

$$E\left(\frac{1}{n-2m}\sum_t (S_t - \mu_t)^2\right) = \frac{1}{n-2m}\sum_t (\bar{\mu}_t - \mu_t)^2 + \sigma^2/(2m+1)$$

for the smoothed series.

Smoothed Gaussian White Noise

Under the assumption of Gaussian white noise, it follows that

$$\bar{U}_t \sim N\left(0, \frac{\sigma^2}{2m+1}\right)$$

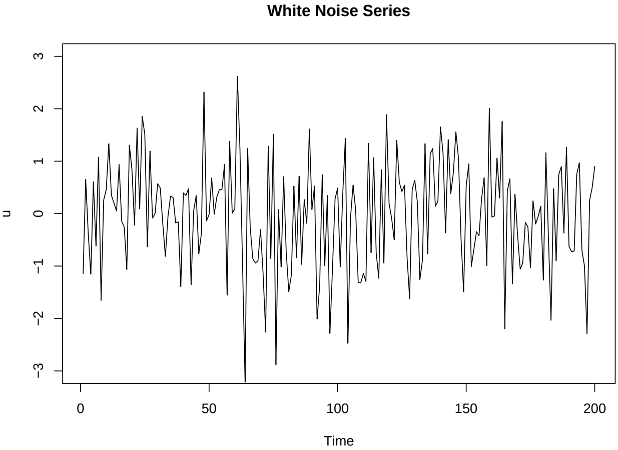
but the smoothed series is no longer white noise.

Shown below is:

- A white noise series;
- The white noise series, smoothed with a five-point moving average smoother.

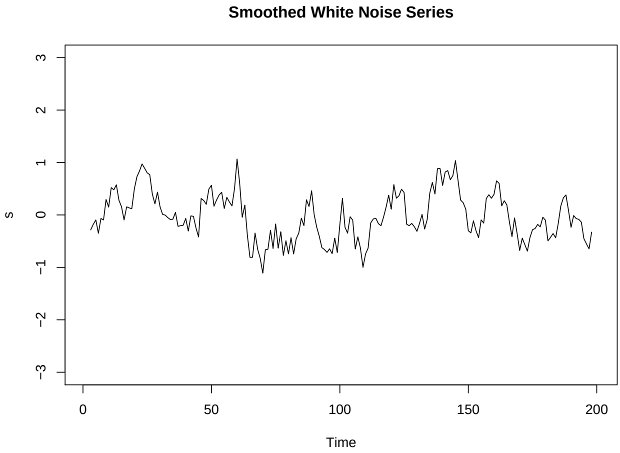
Gaussian White Noise Series

```
u=rnorm(200)
u=ts(u)
plot(u,main="White Noise Series",ylim=c(-3,3))
```



Smoothed White Noise Series

```
s=filter(u,filter=rep(1/5,5))
plot(s,main="Smoothed White Noise Series",ylim=c(-3,3))
```



Smoothed Gaussian White Noise

The smoothed series differs from the original white noise series in two respects.

- The amplitude is reduced, corresponding to the variance $\sigma^2/(2m+1)$.
- The smoothed series also appears more regular in that there are longer runs of increasing and decreasing values than occur for white noise.

Serial Correlation

The moving average filter induces serial correlation in the smoothed series $\bar{U}_t$.

Suppose

$$\bar{U}_t = \sum_{j=-m}^{m} w_j U_{t+j}.$$

It can be checked that

$$\operatorname{cor}(\bar{U}_t, \bar{U}_{t+k}) = \left(\sum_{j=-m+k}^{m} w_j w_{j-k} \right) \bigg/ \left(\sum_{j=-m}^{m} w_j^2 \right)$$

for $k = 0, 1, 2, \ldots, 2m$ and 0 otherwise.

Serial Correlation (continued)

For the five-point moving average smoother we have

$$\mathrm{cor}(\bar{U}_t, \bar{U}_{t+k}) = \begin{cases} 1 & \text{if } k = 0 \\ 0.8 & \text{if } k = 1 \\ 0.6 & \text{if } k = 2 \\ 0.4 & \text{if } k = 3 \\ 0.2 & \text{if } k = 4 \\ 0 & \text{if } k \geq 5 \end{cases}$$

Iterated Smoothing

The degree of smoothing with a moving average filter is controlled by choosing the width of the averaging window.

For a given data set, this can be explored by trial and error.

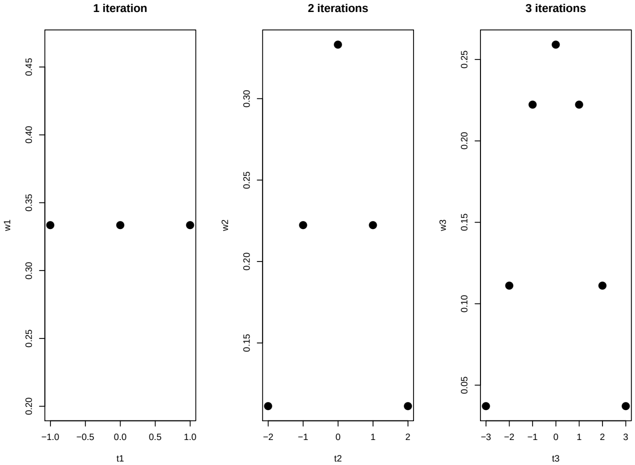
An alternative approach is to iterate the three-point moving average process.

That is, smooth the series once, using the three-point moving average.

If further smoothing is required, apply a three-point moving average to the smoothed series and iterate until the desired level of smoothing is achieved.

Smoothers produced by this method are also linear filters and the weights are shown below.

Weights from iterated smoothing



Spline Smoothers

An alternative to the smoothing with a linear filter is to fit a smoothing spline to the data.

Conside a set of knots,

$$t_1, t_2, \ldots, t_n.$$

A natural cubic spline is a function, $\mu(t)$ with the following properties.

- $\mu(t)$ is linear for $t < t_1$ and $t > t_n$;
- $\mu(t)$ is a cubic polynomial on every interval (t_i, t_{i+1});
- $\mu(t)$ is twice continuously differentiable for all t.

Spline Smoothers (continued)

A smoothing spline is fitted to data $\{y_{t_i}\}$ by choosing knots corresponding to the observed time points and choosing a natural cubic spline to minimise

$$\sum_{i=1}^{n}(y_{t_i} - \mu(t_i))^2 + \lambda \int_{-\infty}^{\infty} \mu''(t)^2 dt.$$

The parameter $\lambda \geq 0$ is called the smoothing parameter.

- For $\lambda = 0$ the solution will have $\mu(t_i) = y_{t_i}$.
- As λ increases the solution becomes increasingly smooth.
- In the limit as $\lambda \to \infty$ the solution is a straight line.

Spline Smoothers (continued)

In most implementations of smoothing splines, the smoothing parameter λ can be chosen automatically by cross validation.

The simplest method is leave one out cross validation.

- For a given value of λ, fit a smoothing spline to the subset of data consisting of all points except (t_i, y_{t_i}).
- Let $\mu_{(-i)}(t_i)$ be the predicted value for y_{t_i}.
- The cross-validated error is

$$\text{cv}(\lambda) = \sum_{i=1}^{n} \left(y_{t_i} - \mu_{(-i)}(t_i) \right)^2 .$$

- The smoothing parameter is chosen to minimize $\text{cv}(\lambda)$.

Cross validation works very well when the error process is white noise but can be misleading otherwise.

Spline Smoothers (continued)

For given λ, the smoothed series obtained from a smoothing spline can be expressed in the form

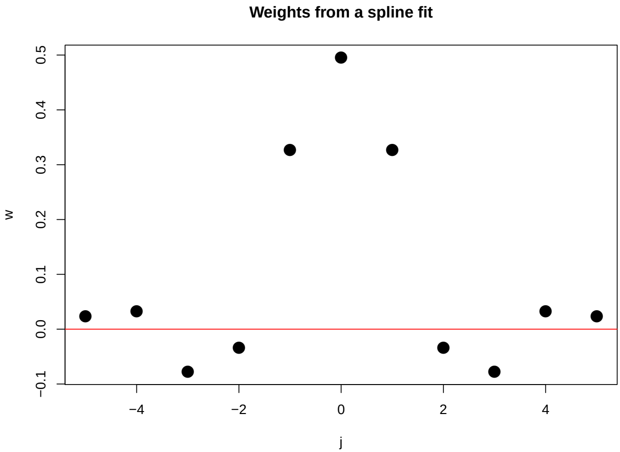
$$\hat{\boldsymbol{\mu}} = M\boldsymbol{y}$$

where M is an $n \times n$ matrix.

Thus, despite the more complicated derivation, the smoothing spline fit is essentially a linear combination of the observed time series.

An example of the coefficients for a spline smoother are shown below.

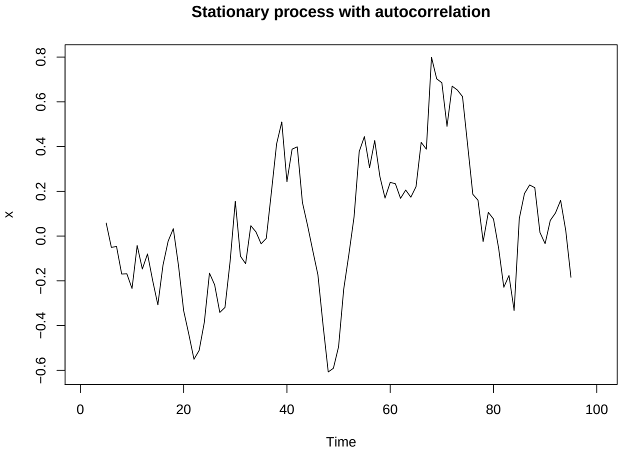
Spline Smoothers (continued)



Autocorrelation and Trend

Strong patterns can occur in stationary data when autocorrelation is present.

```
x=ts(rnorm(100)); x=filter(x,filter=rep(0.1,10))
plot(x,main="Stationary process with autocorrelation")
```



Autocorrelation and Trend (continued)

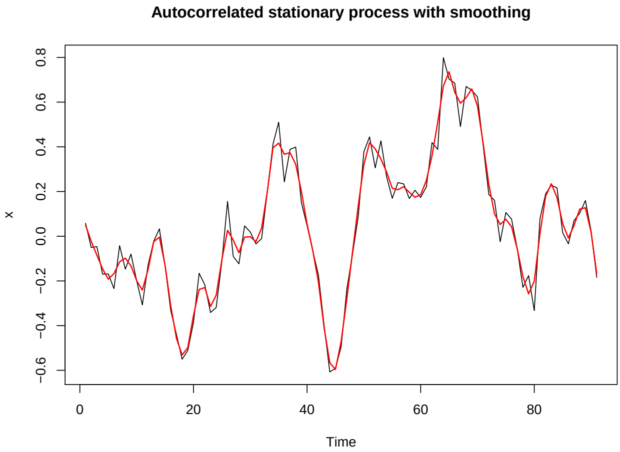
It can be difficult to distinguish visually between a pure noise process that is

- stationary
- with strong autocorrelation

and deterministic signal plus noise.

Smoothing in the former case may recover an apparently smooth signal even though the input was just noise.

Example (continued)



Removing Seasonality

Seasonal effects can be removed by filtering.

To illustrate, consider a time series with 12 observations per year and an annual cycle.

Applying a 12-point moving average will average out any seasonal effects to produce a smoothed version of the trend in the series.

To obtain a smoothed series that aligns with the original series, a filter of odd order is needed.

In this case the filter weights are modified so that

$$w_{-(2m+1)} = w_{2m+1} = \frac{1}{4m}$$

and

$$w_{-2m} = w_{-2m+1} = \cdots = w_{2m} = \frac{1}{2m}.$$

Removing Seasonality

For example, for monthly data,the weights would be

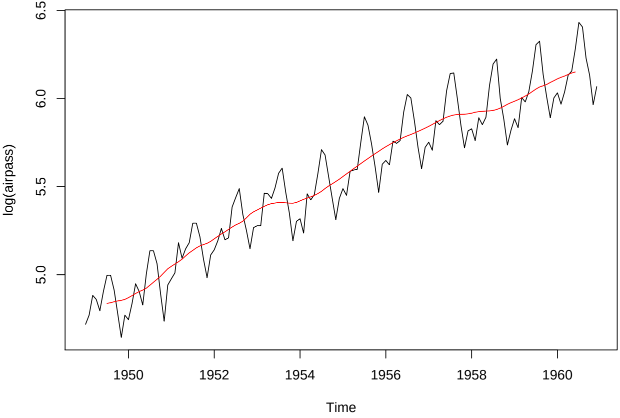
$$\frac{1}{24}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{12}\ \frac{1}{24}.$$

For quarterly data, the weights would be

$$\frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{8}.$$

Example

```
plot(log(airpass))
trend=filter(log(airpass),filter=c(1/24,rep(1/12,11),1/24))
lines(trend,col="red")
```



The decompose function in R

The decompose function in R can be applied to decompose a series in to

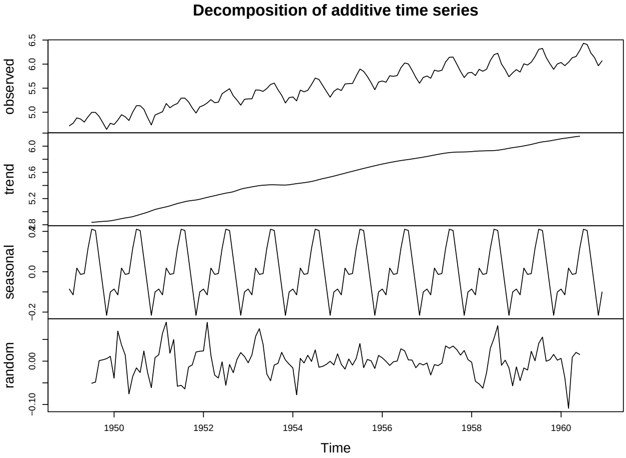
$$\text{trend}+\text{seasonality}+\text{noise}$$

This is a purely descriptive tool that is sometimes useful for looking at data.

It should be noted that applying the decompose function does not imply an underlying the fitting of an underlying statistical model to the data.

Example

```
plot(decompose(log(airpass)))
```



The decompose function in R

The algorithm for the decompose function is as follows:

1. Calculate the trend using a linear filter determined by the `period` of the series;
2. Subtract the trend to obtain a residual series;
3. Calculate the seasonal effect in two stages
   - Calculate the mean for each month (or quarter) of the residual series;
   - Subtract the overall mean to center the seasonal estimate;
4. Calculate the random component by subtracting the estimated trend and seasonal effect from the original series.

Differencing Time Series

Differencing is another method traditionally used to help understand the structure of time series data.

The purpose is (informally) to remove any trends from the data to emphasize other features of the data.

The first order differencing operator D is defined by

$$Dy_t = y_t - y_{t-1}$$

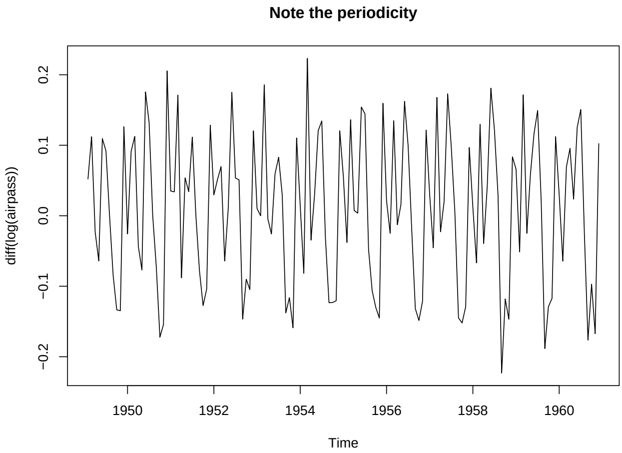
Higher order differences are defined by successive application of the differencing operator.

For example,

$$D^2 y_t = D(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}$$

Example - the air passenger data

```
plot(diff(log(airpass)),main="Note the periodicity")
```



The effect of differencing

If y_t is a polynomial of degree k, then applying the k^{th} order difference will annihilate the trend in the data.

If y_t is a stationary process, then the differenced series $D\boldsymbol{y}$ is also stationary.

Hence if y_t consists of a trend that can be approximated by a polynomial and additive stationary noise, the suitably differenced series will be a stationary process.

Choosing the order of the difference

Differencing is an informal descriptive method and the order is typically chosen by trial and error.

Typically differences of order 1,2 or 3 may be useful.

If excessive differencing is applied, the effect will be to increase the amplitude of the noise.

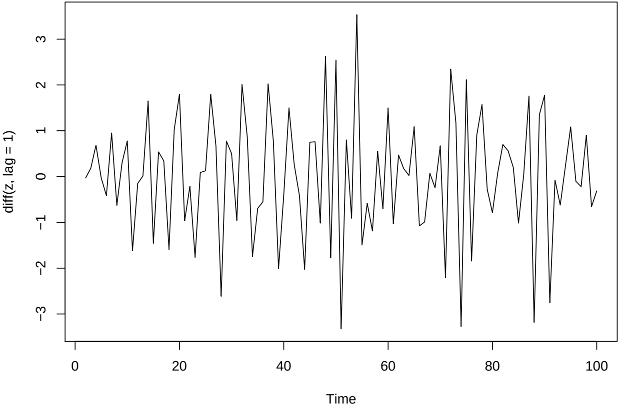
For example, consider a Gaussian white noise process $U_t \sim N(0, 1)$.

It can be checked that the differenced series, $D(U_t)$ has:

- $\text{var}(D(U_t)) = 2$
- $\text{cor}(D(U_t), D(U_{t-1})) = -0.5$

Example

```
z=ts(rnorm(100))
plot(diff(z,lag=1))
```



Differencing to remove seasonality

Differencing can also be used to remove seasonality from a series by choosing an appropriate lag.

For example, with the air passenger data, the lag 12 differences

$$y_t - y_{t-12}$$

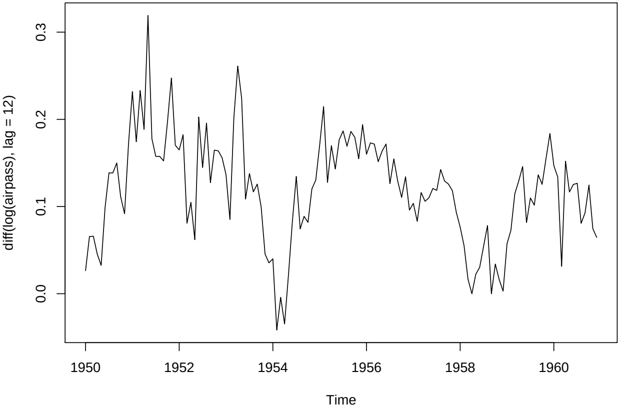
can be calculated.

This plot shows:

- An overall positive level corresponding to the increasing trend;
- It also highlights some short term departures from the trend.

Example - the air passenger data

```
plot(diff(log(airpass),lag=12))
```



The Autocovariance Function

Consider a second order stationary function $Y(t)$.

The autocovariance function is defined by

$$\gamma(s) = \mathrm{cov}(Y(t), Y(t+s)).$$

Note that

$$\gamma(0) = \mathrm{var}(Y(t))$$

and $\gamma(-s) = \gamma(s)$ for all s.

Autocorrelation Function

The autocorrelation function is defined by

$$\rho(s) = \mathrm{cor}(Y(t), Y(t+s)) = \frac{\gamma(s)}{\gamma(0)}.$$

The autocorrelation function is a correlation coefficient for each value of s, so it must satisfy

$$-1 \leq \rho(s) \leq 1$$

for all s.

It also satisfies

$$\rho(0) = 1 \text{ and } \rho(-s) = \rho(s).$$

Discrete processes

For a discrete second order stationary process, $\{Y_t\}$, we can only define the autocovariance and autocorrelation at integer lags.

In this case, we write

$$\gamma_k = \mathrm{cov}(Y_t, Y_{t+k})$$

and

$$\rho_k = \mathrm{cor}(Y_t, Y_{t+k}).$$

Example - The Moving Average Process

Suppose $\{U_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a doubly infinite sequence of independent random variables with

$$E(U_t) = 0 \text{ and } \operatorname{var}(U_t) = \sigma^2.$$

For a set of coefficients $\beta_1, \beta_2, \ldots, \beta_q$ let

$$Y_t = U_t + \sum_{j=1}^{q} \beta_j U_{t-j} = \sum_{j=0}^{q} \beta_j U_{t-j}$$

where β_0=1.

Can check that

$$\gamma_0 = \operatorname{var}(Y_t) = \sigma^2 \sum_{j=0}^{q} \beta_j^2.$$

Example (continued)

More generally,

$$\gamma_k = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} & \text{for } k = 0, 1, 2, \ldots, q \\ 0 & \text{for } k > q \end{cases}$$

and, hence,

$$\rho_k = \begin{cases} \sum_{j=0}^{q-k} \beta_j \beta_{j+k} \Big/ \sum_{j=0}^{q} \beta_j^2 & \text{for } k = 0, 1, 2, \ldots, q \\ 0 & \text{for } k > q. \end{cases}$$

Example - Autoregressive Process

Suppose $\{U_t\}$ is as previously and consider the process defined by

$$Y_t = U_t + \alpha Y_{t-1}$$

where $|\alpha| < 1$.

We will show later that $\{Y_t\}$ is a stationary process.

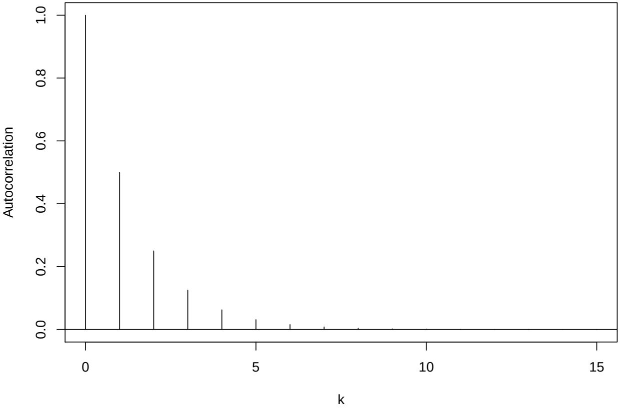
It follows from the assumption of stationarity that $E(Y_t) = 0$ and it can also be shown that

$$\gamma_0 = \frac{\sigma^2}{1 - \alpha^2}$$

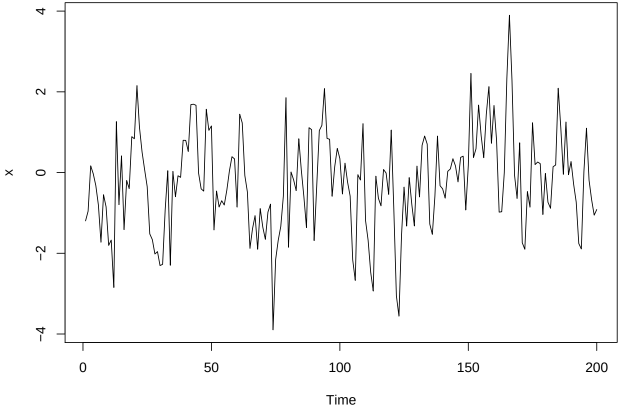
and

$$\gamma_k = \alpha^k \gamma_0.$$

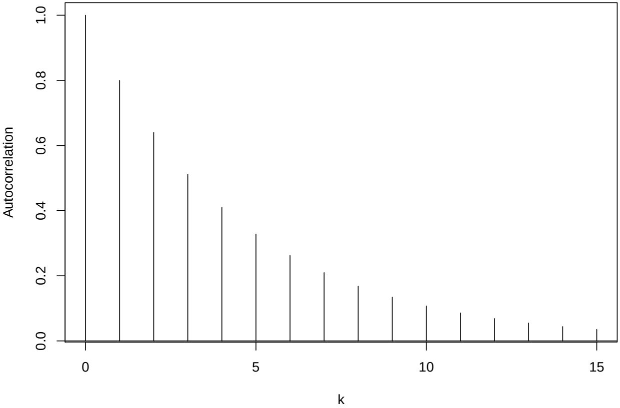
Example - Autoregressive Order 1, $\alpha = 0.5$



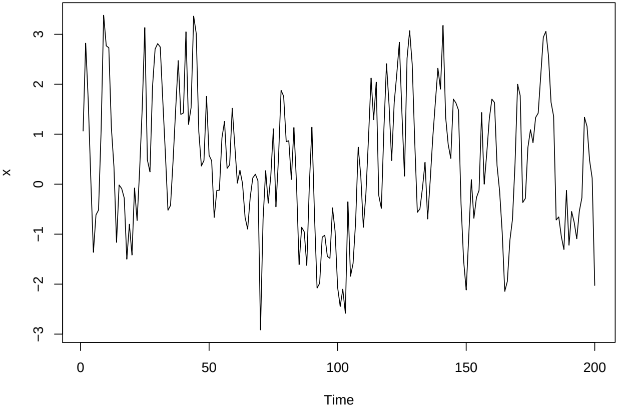
Example - Gaussian Autoregressive Order 1, $\alpha = 0.5$



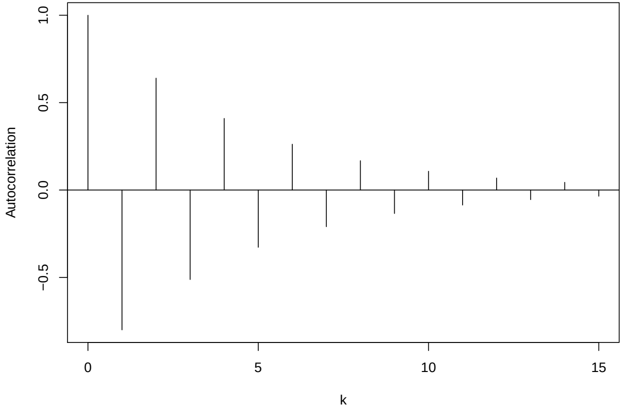
Example - Autoregressive Order 1, $\alpha = 0.8$



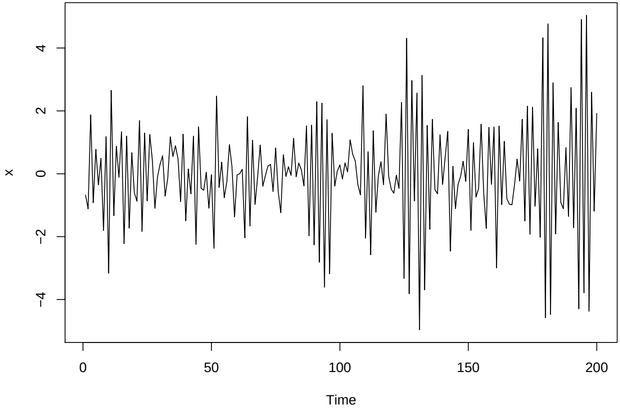
Example - Gaussian Autoregressive Order 1, $\alpha = 0.8$



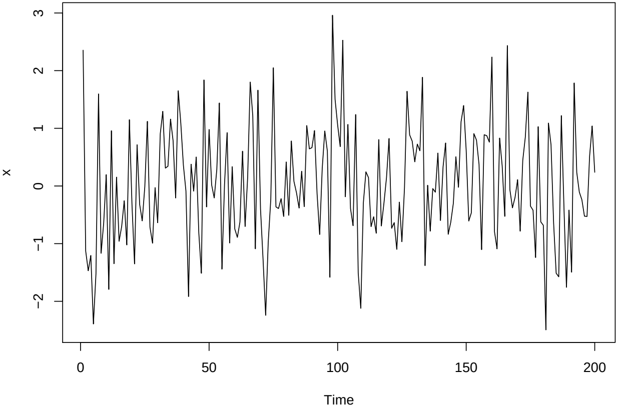
Example - Autoregressive Order 1, $\alpha = -0.8$



Example - Gaussian Autoregressive Order 1, $\alpha = -0.8$



Example Gaussian White Noise



The Sample Autocorrelation Function

In practice, the autocovariance and autocorrelations are estimated from data.

Consider an observed series $\{y_t : \; t = 1, 2, \ldots, n\}$.

The lag k sample autocovariance is defined by

$$g_k = \frac{1}{n} \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})$$

where

$$\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t.$$

The lag k sample autocorrelation is

$$r_k = g_k / g_0$$

Statistical Properties of Sample Autocovariance

The exact statistical properties of the sample autocovariance and autocorrelation functions are generally very complicated.

Consider a second order stationary time series $\{Y_t\}$ with $E(Y_t) = \mu$ autocovariance function γ_k.

$$
\begin{aligned}
ng_k &= \sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}) \\
&= \sum_{t=k+1}^{n} ((Y_t - \mu) - (\bar{Y} - \mu))((Y_{t-k} - \mu) - (\bar{Y} - \mu)) \\
&= \sum_{t=k+1}^{n} (Y_t - \mu)(Y_{t-k} - \mu) - \sum_{t=k+1}^{n} (Y_t - \mu)(\bar{Y} - \mu) \\
&\quad - \sum_{t=k+1}^{n} (Y_{t-k} - \mu)(\bar{Y} - \mu) + (n-k)(\bar{Y} - \mu)^2
\end{aligned}
$$

Statistical Properties of Sample Autocovariance

Observe

$$E((Y_t - \mu)(Y_{t-k} - \mu)) = \gamma_k$$

so that

$$E\left(\sum_{t=k+1}^{n} (Y_t - \mu)(Y_{t-k} - \mu)\right) = (n-k)\gamma_k. \qquad (1)$$

Also

$$E((Y_t - \mu)(\bar{Y} - \mu)) = \frac{1}{n}\sum_{s=1}^{n} \gamma_{t-s}$$

so that

$$E\left(\sum_{t=k+1}^{n} (Y_t - \mu)(\bar{Y} - \mu)\right) = \frac{1}{n}\sum_{t=k+1}^{n}\sum_{s=1}^{n} \gamma_{t-s}. \qquad (2)$$

Statistical Properties of Sample Autocovariance

Similarly,

$$E\left(\sum_{t=k+1}^{n}(Y_{t-k}-\mu)(\bar{Y}-\mu)\right)=\frac{1}{n}\sum_{t=k+1}^{n}\sum_{s=1}^{n}\gamma_{t-k-s}. \qquad (3)$$

Finally,

$$E\left((n-k)(\bar{Y}-\mu)^2\right)=\frac{(n-k)}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma_{t-s}. \qquad (4)$$

Statistical Properties of Sample Autocovariance

Combining equations 1, 2, 3, 4, we obtain

$$E(g_k) = \frac{n-k}{n}\gamma_k + \frac{1}{n^2}R$$

where

$$\begin{aligned} R &= \frac{n-k}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma_{t-s} - \sum_{t=k+1}^{n}\sum_{s=1}^{n}(\gamma_{t-s} + \gamma_{t-k-s}) \\ &= -\sum_{t=k+1}^{n-k}\sum_{s=1}^{n}\gamma_{t-s} - \frac{k}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\gamma_{t-s}, \end{aligned}$$

for $k < n/2$.

Statistical Properties of Sample Autocovariance

It follows that g_k is generally a biased estimator for γ_k but is asymptotically unbiased provided

$$\lim_{n\to\infty} \frac{1}{n^2} R = 0.$$

Mean Ergodicity

A stationary process $\{Y_t\}$ with $E(Y_t) = \mu$ is said to be ergodic in the mean if

$$\bar{Y}_n = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

converges in a probabilistic sense to μ.

For example, if

$$\lim_{n \to \infty} E\left((\bar{Y}_n - \mu)^2\right) = \lim_{n \to \infty} \text{var}(\bar{Y}_n) = 0.$$

Mean Ergodicity

From equation 4,

$$\begin{aligned} \mathrm{var}(\bar{Y}_n) &= \frac{1}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\gamma_{t-s} \\ &= \gamma_0/n + \frac{2}{n^2}\sum_{t=1}^{n-1}(n-t)\gamma_t \end{aligned}$$

For a white noise process, $\gamma_t = 0$ for $t \geq 1$, in which case

$$\mathrm{var}(\bar{Y}_n) = \gamma_0/n \to 0 \text{ as } n \to \infty$$

Mean Ergodicity

More generally, to obtain ergodicity in the mean, we require

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{t=1}^{n-1} (n-t)\gamma_t = 0.$$

Sufficient conditions can be found by observing that

$$\left| \frac{1}{n^2} \sum_{t=1}^{n-1} (n-t)\gamma_t \right| \le \frac{1}{n} \sum_{t=1}^{n} |\gamma_t|.$$

For example, it follows that

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty \text{ and } \lim_{k\to\infty} \gamma_k = 0$$

are both sufficient conditions for ergodicity in the mean.

In this case, g_k is asymptotically unbiased for γ_k as $n \to \infty$ for each fixed k.

Gaussian White Noise

For a Gaussian white noise process the autocovariance function, γ_k, satisfies $\gamma_k = 0$ for $k \geq 1$ and hence the autocorrelation function is

$$\rho_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

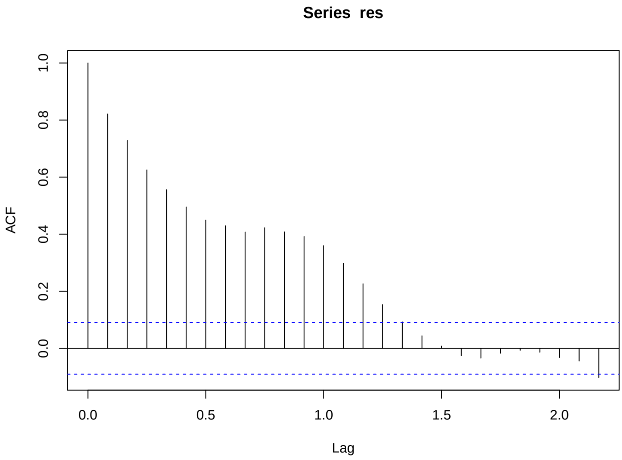
The exact statistical properties of the sample autocorrelation function, $r_k = g_k/g_0$, are very complicated but it has been proved for large n that

$$r_k \mathrel{\dot\sim} N(0, 1/n)$$

The `acf` function in `R` shows limits at $\pm 2/\sqrt{n}$ to identify significant correlations.

Example the Mauna Loa data

```r
acf(res)
```



Tests for serial correlation

The autocorrelation function (or correlogram) can be used to test for serial correlations.

For a Gaussian white noise process, the asymptotic distribution of r_k is $N(0, 1/n)$.

Hence, an approximate 5% level test of $H_k : \rho_k = 0$ is obtained by the rule,

$$\textit{Reject } H_k \textit{ if } |r_k| > 1.96/\sqrt{n}.$$

A natural hypothesis to consider is that the process is white noise. That is,

$$H_0 : \; 0 = \rho_1 = \rho_2 = \rho_3 = \ldots$$

This hypothesis cannot be tested by testing the individual hypotheses separately because it fails to account for multiple comparisons.

Tests for serial correlation

A test of

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0$$

can be obtained from the test statistic

$$Q_m = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} r_k^2.$$

Provided $n >> m$, it can be shown that the null distribution of Q_m is approximately χ^2_m.

Hence a level α test of H_0 is defined by the rule,

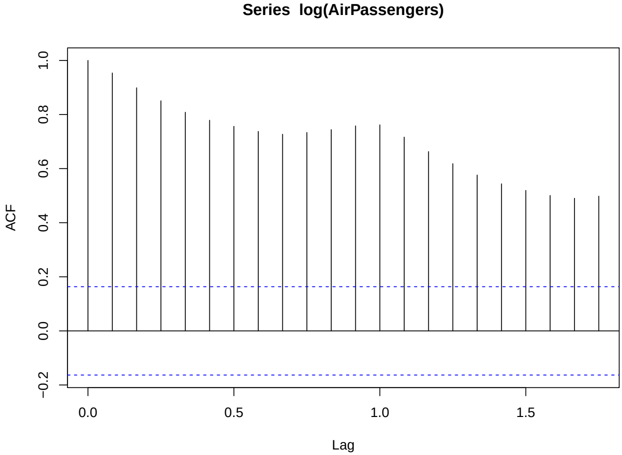
Reject H_0 if $Q_m \geq \chi^2_{m,\alpha}$.

Remarks

1. In practice the correlogram (acf) is most useful for understanding the dependence in the series.
2. The pattern in the acf can give indications about plausible models for the structure of the correlation by comparison to known acfs for certain theoretical models.
3. Slow linear decay of the correlations is an indication of non-stationarity, in which case the true acf does not exist.
4. Trend and periodicity are usually found by plotting the original series but will also be evident in the acf.
   - Trend results in slow linear decay of the acf;
   - Oscillation in the acf is the result of periodicity.

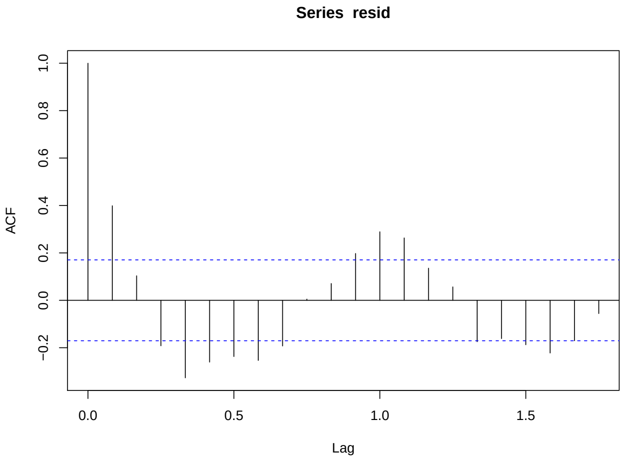
Example: the Air Passenger Data

```r
acf(log(AirPassengers))
```



Example: the Air Passenger Data

```
resid=decompose(log(AirPassengers))[["random"]]
resid=resid[!is.na(resid)]
resid=ts(resid,start=c(1950,7),frequency=12)
acf(resid)
```



The variogram

For unequally spaced data, the auto correlation function can be estimated through the variogram.

Consider a stationary process $Y(t)$.

The variogram is defined by

$$V(k) = \frac{1}{2} E\left((Y(t) - Y(t-k))^2\right).$$

Observe that

$$\begin{aligned} V(k) &= \frac{1}{2} E\left(((Y(t)-\mu) - (Y(t-k)-\mu))^2\right) \\ &= \frac{1}{2}\left\{E\left((Y(t)-\mu)^2\right) + E\left((Y(t-k)-\mu)^2\right)\right. \\ &\quad \left. -2E\left((Y(t)-\mu)(Y(t-k)-\mu)\right)\right\} \\ &= \gamma(0) - \gamma(k) = \gamma(0)(1-\rho(k)). \end{aligned}$$

Variogram estimation

Consider an unequally spaced time series $\{y(t_i)\}$
$i = 1, 2, \dots, n$.

The variogram can be estimated by calculating

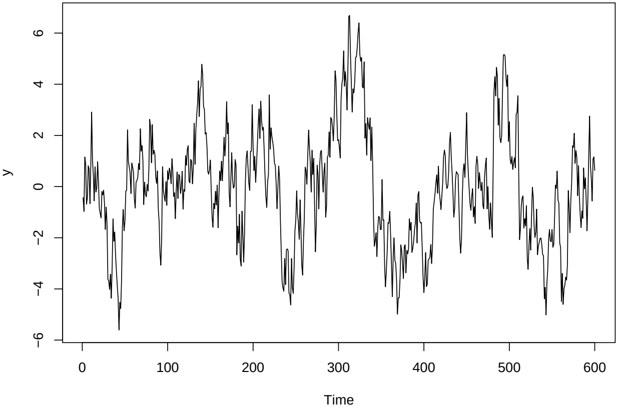
$$\begin{aligned} v_{ij} &= \frac{1}{2}(y(t_i) - y(t_j))^2 \\ k_{ij} &= |t_i - t_j| \end{aligned}$$

for all $n(n-1)/2$ distinct pairs of times t_i, t_j.

The estimate at lag k, $\bar{v}(k)$, is the average over all values of v_{ij} with $k_{ij} = k$.

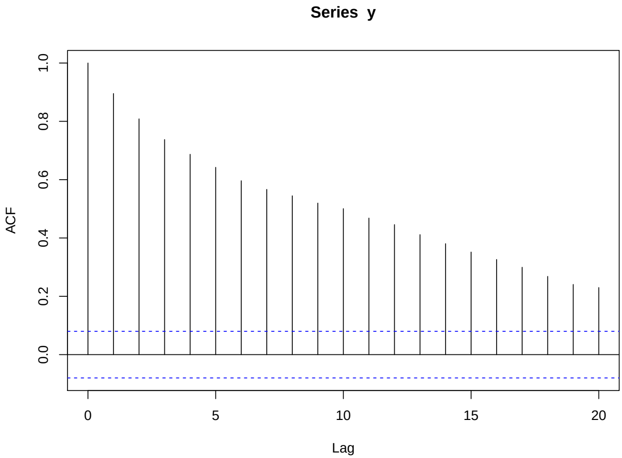
Example - simulated time series of length 600

```r
plot(y)
```



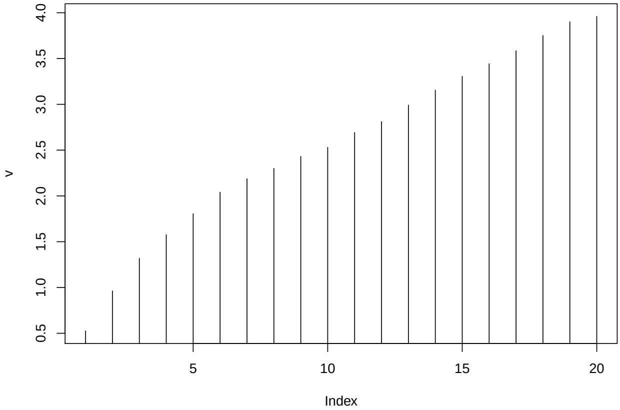
Example continued - acf

```
acf(y,lag.max=20)
```



Example continued - variogram

```
v=rep(0,20); n=600
for(i in 1:20) v[i]=0.5*mean((y[1:(n-i)]-y[(i+1):n])^2)
plot(v,type="h")
```



The variogram

For a stationary process with negligible correlations beyond some lag, ℓ, we expect the variogram to increase until lag ℓ and then flatten off.

The level at which it flattens off can be used to estimate the variance, $\gamma(0) = v(\infty)$.

The autocorrelation function can be estimated via the relation

$$v(k) = \gamma(0)(1 - \rho(k))$$

In applications where there are only a small number of pairs at each lag, the variogram is sometimes estimated by smoothing or fitting a parametric model to the scatter plot of v_{ij} vs k_{ij}.

The variogram

The variogram is well defined for second order stationary processes.

However, it should also be noted the the variogram may be well defined from some non-stationary processes.

For example, consider a random walk defined by

$$U_t \sim N(0, \sigma^2) \text{ independently for } t = 0, 1, 2, 3, \ldots$$

and

$$Y_t = \sum_{s=0}^{t} U_s.$$

In this case $\{Y_t\}$ is not stationary but

$$v(k) = k\sigma^2$$

is well defined for all k.

A simple model for periodicity

Consider a time series $\{y_t : y = 1, 2, \ldots, n\}$ and the model

$$M: \; y_t = \alpha \cos(2\pi jt/n) + \beta \sin(2\pi jt/n) + z_t$$

for some positive integer $j < n/2$ and where $\{z_t\}$ is a white noise process.

Note that frequency is $2\pi j/n$ and the period is n/j.

The amplitude is $\sqrt{\alpha^2 + \beta^2}$.

Least squares estimates

Least squares estimates for $\boldsymbol{\theta} = (\alpha, \beta)^T$ can be obtained by expressing M as linear model,

$$\boldsymbol{y} = X\boldsymbol{\theta} + \boldsymbol{z}$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} \cos(2\pi j/n) & \sin(2\pi j/n) \\ \cos(2\pi 2j/n) & \sin(2\pi 2j/n) \\ \vdots & \vdots \\ \cos(2\pi j) & \sin(2\pi j) \end{pmatrix}, \ \boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$

Least squares estimates

The least squares estimates are obtained in the usual way,

$$\hat{\boldsymbol{\theta}} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

In this case, it can be shown that

$$(X^T X) = \begin{pmatrix} n/2 & 0 \\ 0 & n/2 \end{pmatrix}$$

so

$$\hat{\alpha} = \frac{2}{n} \sum_{t=1}^{n} \cos(2\pi jt/n) y_t, \text{ and } \hat{\beta} = \frac{2}{n} \sum_{t=1}^{n} \sin(2\pi jt/n) y_t.$$

Trigonometric Identities

The fact that $(X^T X)$ is a diagonal matrix follows from the following.

Lemma For any positive integer $j < n/2$,

$$\begin{aligned}
\sum_{t=1}^{n} \cos(2\pi jt/n) = \sum_{t=1}^{n} \sin(2\pi jt/n) &= 0 \\
\sum_{t=1}^{n} \cos(2\pi jt/n) \sin(2\pi jt/n) &= 0 \\
\sum_{t=1}^{n} \cos^2(2\pi jt/n) = \sum_{t=1}^{n} \sin^2(2\pi jt/n) &= n/2
\end{aligned}$$

□

Trigonometric Identities

Corollary For positive integers $j < n/2$ and $k < n/2$ with $j \neq k$,

$$
\begin{aligned}
\sum_{t=1}^{n} \cos(2\pi jt/n) \sin(2\pi kt/n) &= 0 \\
\sum_{t=1}^{n} \cos(2\pi jt/n) \cos(2\pi kt/n) &= 0 \\
\sum_{t=1}^{n} \sin(2\pi jt/n) \sin(2\pi kt/n) &= 0
\end{aligned}
$$

$\square$

Decomposition of Time Series

More generally, the time series $\{y_t : \; t = 1, 2, \ldots, n\}$ can be completely decomposed into trigonometric functions.

For odd n, we can consider the decomposition

$$y_t = \mu + \sum_{j=1}^{(n-1)/2} \{\alpha_j \cos(2\pi jt/n) + \beta_j \sin(2\pi jt/n)\}.$$

For even n, we have

$$y_t = \mu + \sum_{j=1}^{n/2-1} \{\alpha_j \cos(2\pi jt/n) + \beta_j \sin(2\pi jt/n)\} + \alpha_{n/2} \cos(\pi t).$$

Note that $cos(\pi t) = (-1)^t$ for $t = 1, 2, \ldots, n$.

The periodogram

The decomposition of a time series into periodic components is an orthogonal decomposition.

The least squares estimates are then

$$
\begin{aligned}
\hat{\mu} &= \frac{1}{n}\sum_{t=1}^{n} y_t \\
\hat{\alpha}_j &= \frac{2}{n}\sum_{t=1}^{n} y_t \cos(2\pi jt/n) \\
\hat{\beta}_j &= \frac{2}{n}\sum_{t=1}^{n} y_t \sin(2\pi jt/n)
\end{aligned}
$$

and, for n even,

$$\hat{\alpha}_{n/2} = \frac{1}{n}\sum_{t=1}^{n} y_t \cos(\pi t).$$

The periodogram

This fact motivates the following definition.

The periodogram $I(\omega)$ is defined by

$$I(\omega) = \frac{1}{n} \left\{ \left(\sum_{t=1}^{n} y_t \cos(\omega t) \right)^2 + \left(\sum_{t=1}^{n} y_t \sin \omega t) \right)^2 \right\}.$$

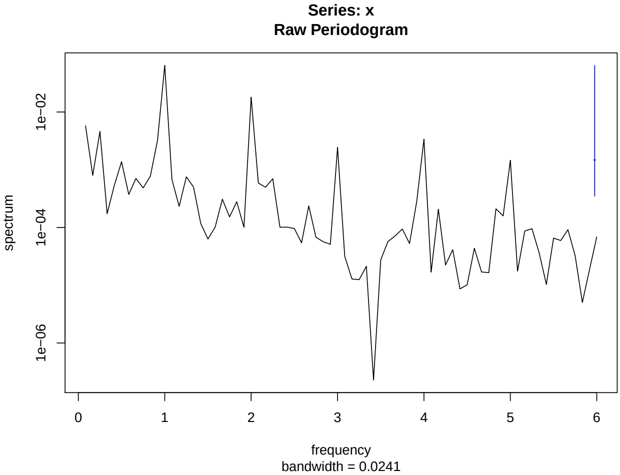
Taking $\omega = 2\pi j/n$,

$$I(2\pi j/n) = \frac{n}{4} \left(\hat{\alpha}_j^2 + \hat{\beta}_j^2 \right).$$

The periodogram is obtained by plotting $I(\omega)$ vs ω.

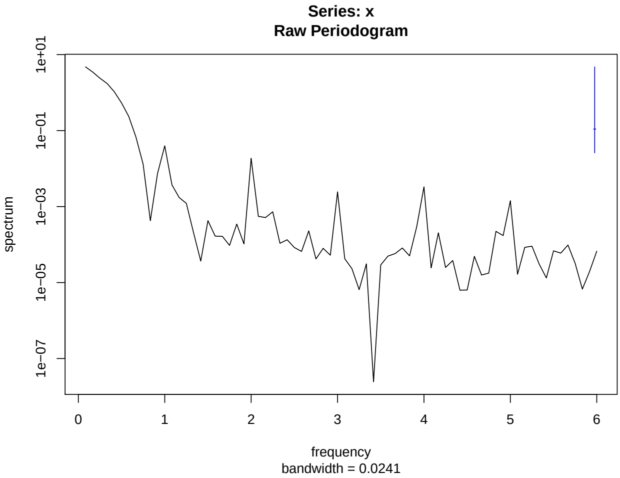
Example - the Air Passenger Data

```r
spectrum(log(AirPassengers))
```



Example - the Air Passenger Data

```
spectrum(log(AirPassengers),detrend=FALSE)
```



Note: by default the `spectrum` function removes any linear trend from the data.

Example - continued

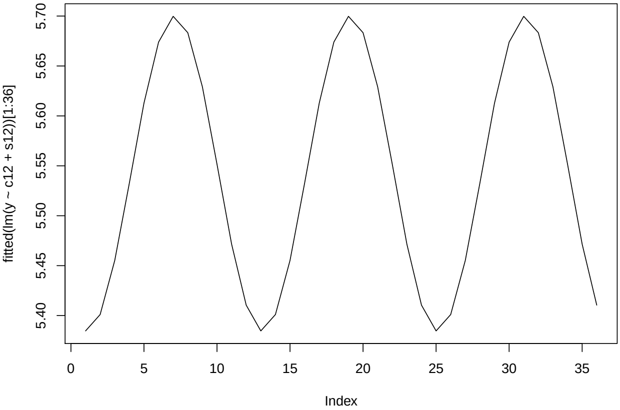
The periodogram of the air passenger data shows a series of well defined peaks at frequencies of 1,2,3,4,5 cycles per year.

The interpretation of this is that there is an annual cycle in the data.

The peaks at higher frequencies arise because the cyclic pattern is not sinusoidal.

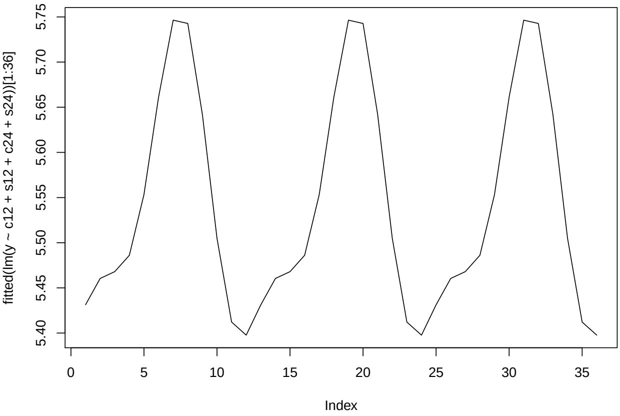
Example continued

```
t=c(1:144); y=log(AirPassengers)
c12=cos(2*pi*12*t/144); s12=sin(2*pi*12*t/144)
plot(fitted(lm(y~c12+s12))[1:36],type="l")
```



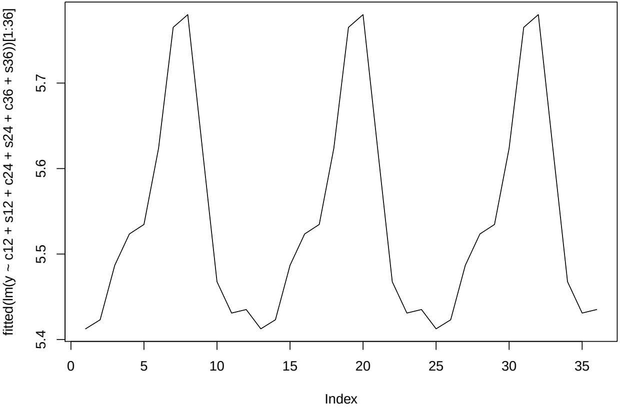
Example continued

```
c24=cos(2*pi*24*t/144); s24=sin(2*pi*24*t/144)
plot(fitted(lm(y~c12+s12+c24+s24))[1:36],type="l")
```



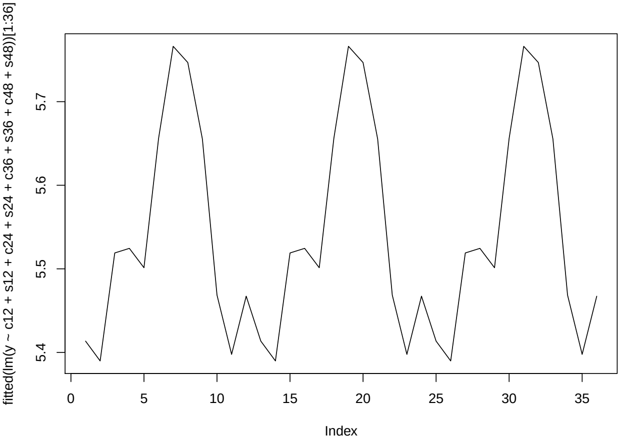
Example continued

```
c36=cos(2*pi*36*t/144); s36=sin(2*pi*36*t/144)
plot(fitted(lm(y~c12+s12+c24+s24+c36+s36))[1:36],type="l")
```



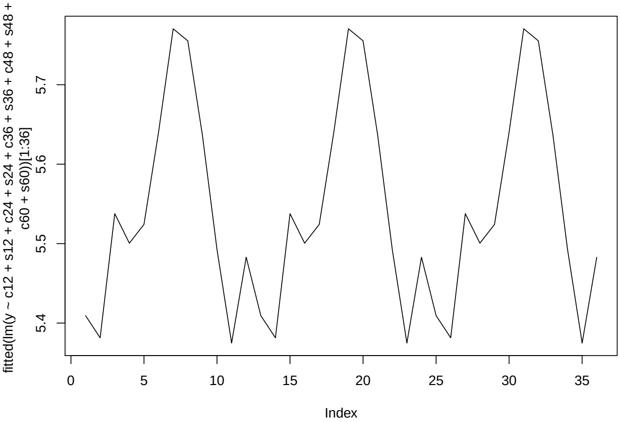
Example continued

```
c48=cos(2*pi*48*t/144); s48=sin(2*pi*48*t/144)
plot(fitted(lm(y~c12+s12+c24+s24+c36+s36+c48+s48))[1:36],
type="l")
```



Example continued

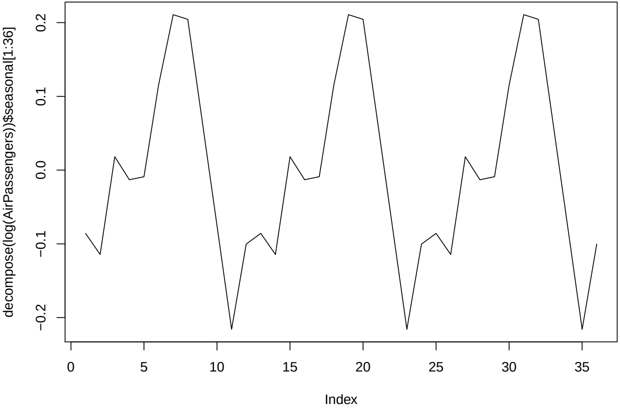
```
c60=cos(2*pi*60*t/144); s60=sin(2*pi*60*t/144)
plot(fitted(lm(y~c12+s12+c24+s24+c36+s36+c48+s48
+c60+s60))[1:36], type="l")
```



Example

For comparison, we can also plot the seasonal effect estimated by filtering.

```
plot(decompose(log(AirPassengers))$seasonal[1:36],type="l")
```



The periodogram and correlogram

For $\omega = 2\pi j/n$, where $j < n/2$ is a positive integer, we can write

$$I(\omega) = \frac{1}{n}\left\{\left(\sum_{t=1}^{n}(y_t - \bar{y})\cos(\omega t)\right)^2 + \left(\sum_{t=1}^{n}(y_t - \bar{y})\sin(\omega t)\right)^2\right\}.$$

since

$$\sum_{t=1}^{n}\cos\omega t = \sum_{t=1}^{n}\sin\omega t = 0.$$

It can then be shown that

$$I(\omega) = g_0 + 2\sum_{k=1}^{n-1} g_k \cos(k\omega)$$

or equivalently

$$I(\omega)/g_0 = 1 + 2\sum_{k=1}^{n-1} r_k \cos(k\omega)$$

The periodogram and correlogram (continued)

To see why, consider the problem of adding all the elements of an $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & & \dots & & a1n \\ a_{21} & a_{22} & a_{23} & \dots & & a_{2n} \\ \vdots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ a_{n1} & & & \dots & a_{n,n-1} & a_{nn} \end{pmatrix}$$

Summing row by row, the total is

$$T = \sum_{s=1}^{n} \sum_{t=1}^{n} a_{st}.$$

The periodogram and correlogram (continued)

On the other hand, the elements can be added by summing on down each of the diagonals. $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & & \dots & & a1n \\ a_{21} & a_{22} & a_{23} & \dots & & a_{2n} \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ a_{n1} & & & \dots & a_{n,n-1} & a_{nn} \end{pmatrix}$$

In this case, the total is

$$T = \sum_{t=1}^{n} a_{tt} + \sum_{k=1}^{n-1} \sum_{t=k+1}^{n} (a_{t-k,t} + a_{t,t-k}).$$

The periodogram and correlogram (continued)

Now observe

$$
\begin{aligned}
I(\omega) &= \frac{1}{n}\left\{\left(\sum_{t=1}^{n}(y_t-\bar{y})\cos(\omega t)\right)^2+\left(\sum_{t=1}^{n}(y_t-\bar{y})\sin(\omega t)\right)^2\right\} \\
&= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\{(y_s-\bar{y})(y_t-\bar{y})(\cos(\omega s)\cos(\omega t)+\sin(\omega s)\sin(\omega t))\} \\
&= \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}(y_s-\bar{y})(y_t-\bar{y})\cos(\omega(t-s)) \\
&= \frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2+\frac{1}{n}\sum_{k=1}^{n-1}\sum_{t=k+1}^{n}\{(y_{t-k}-\bar{y})(y_t-\bar{y})\cos(-k\omega)+ \\
&\qquad (y_t-\bar{y})(y_{t-k}-\bar{y})\cos(k\omega)\} \\
&= \frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2+\frac{2}{n}\sum_{k=1}^{n-1}\cos(k\omega)\sum_{t=k+1}^{n}(y_t-\bar{y})(y_{t-k}-\bar{y}).
\end{aligned}
$$

The cumulative periodogram

Recall the periodogram for data $y_1, y_2, \ldots, y_n$ is defined by

$$I(\omega) = \frac{1}{n}\left\{\left(\sum_{t=1}^{n} y_t \cos(\omega t)\right)^2 + \left(\sum_{t=1}^{n} y_t \sin(\omega t)\right)^2\right\}.$$

for $\omega = 2\pi k/n$ where $k < n/2$ is a positive integer.

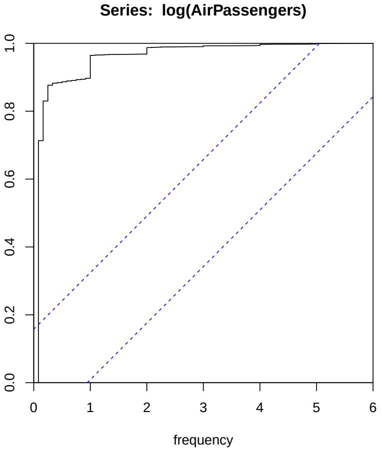
Let $p = [n/2]$ and for $j = 1, 2, \ldots, p$ let

$$C_j = \sum_{k=1}^{j} I(2\pi k/n).$$

A test for white noise can be obtained from the cumulative periodogram by taking $U_j = C_j/C_p$ for $j = 1, 2, \ldots, p-1$ and plotting U_j against $j/(p-1)$.

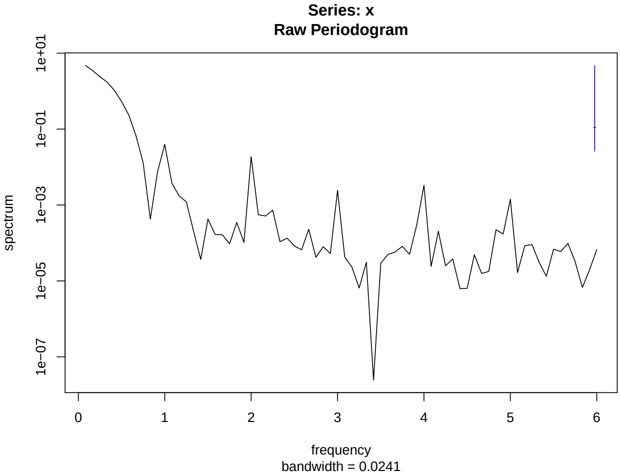
Air Passenger Example

```r
cpgram(log(AirPassengers))
```



Air Passenger Example (continued)

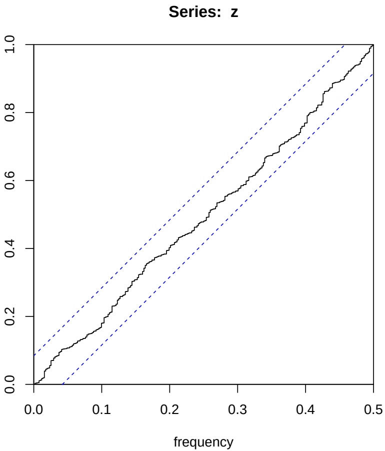
```
spectrum(log(AirPassengers),detrend=FALSE)
```



The cumulative periodogram

In general the hypothesis of white noise is accepted if the plot of the cumulative periodogram is contained wholly within the blue limits and rejected otherwise.

```
z=rnorm(512); z=ts(z); cpgram(z)
```



The cumulative periodogram

The rationale for this test can be explained as follows.

- Suppose $Y_1, Y_2, \ldots, Y_n$ are independent with $Y_t \sim N(0, \sigma^2)$.
- In this case the true values of the coefficients are, $\alpha_j = \beta_j = 0$ for $j = 1, 2, \ldots, [n/2]$.
- Recall that

$$I(2\pi j/n) = \frac{n}{4}\left(\hat{\alpha}_j^2 + \hat{\beta}_j^2\right).$$

- From the standard theory of linear models and using the fact that

$$X^T X = \text{diag}(n/2, n/2, \ldots, n/2)$$

it follows that

$$\frac{\sqrt{n}}{\sigma\sqrt{2}}\hat{\alpha}_j \sim N(0,1) \text{ and } \frac{\sqrt{n}}{\sigma\sqrt{2}}\hat{\beta}_j \sim N(0,1)$$

independently.

The cumulative periodogram

- Hence
$$\frac{2}{\sigma^2} I(2\pi k/n) \sim \chi_2^2$$
independently for $k = 1, 2, \ldots, p$.
- Recall also that χ_2^2 is also the exponential distribution with parameter $\lambda = 1/2$.
- It follows that the joint distribution of $U_1, U_2, \ldots, U_{p-1}$ is the same as the order statistics from a sample of independent uniform $U(0, 1)$ variables.
- The test based on the cumulative periodogram can be seen to be Kolmogorov-Smirnov test for the uniform distribution.

Stationary Random Processes

In what follows we consider doubly infinite sequences of random variables.

We will take $\{Y_t\}$ to mean $\{Y_t : t = 0, \pm 1, \pm 2, \ldots\}$.

A **linear filter** is a transformation from $\{U_t\}$ to $\{Y_t\}$ defined by

$$Y_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}.$$

If $\{U_t\}$ is stationary and only a finite number of the a_j are non-zero, then $\{Y_t\}$ is also stationary.

Otherwise the $\{Y_t\}$ may or may not be stationary depending on the form of the a_j.

The General Linear Process

The general linear linear process is defined by

$$Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}$$

where $\{Z_t\}$ is a white noise process and a_j are constants.

It follows that $\{Y_t\}$ is stationary provided $\sum_{j=1}^{\infty} a_j^2 < \infty$ and

$$\begin{aligned}
E(Y_t) &= 0 \\
\mathrm{var}(Y_t) &= \sigma^2 \sum_{j=0}^{\infty} a_j^2 \\
\gamma(k) &= \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+k}.
\end{aligned}$$

Autoregressive and moving average processes

Let $\{Z_t\}$ be a white noise process.

The **moving average** process, $\{Y_t\} \sim \mathrm{MA}(q)$, is defined by

$$Y_t = Z_t + \sum_{j=1}^{q} \beta_j Z_{t-j}.$$

The **autoregressive process**, $\{Y_t\} \sim \mathrm{AR}(p)$, is defined by

$$Y_t = Z_t + \sum_{j=1}^{p} \alpha_j Y_{t-j}.$$

The **autoregressive moving average** process, $\{Y_t\} \sim \mathrm{ARMA}(p,q)$, is defined by

$$Y_t = Z_t + \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{j=1}^{q} \beta_j Z_{t-j}.$$

The backward shift operator

Define the backward shift operator by

$$BY_t = Y_{t-1}$$

so that

$$B^j Y_t = Y_{t-j} \text{ and } 1Y_t = B^0 Y_t = Y_t.$$

The MA(q) process can be written,

$$Y_t = (1 + \beta_1 B + \ldots + \beta_q B^q) Z_t = \theta(B) Z_t$$

where $\theta(B) = 1 + \beta_1 B + \ldots + \beta_q B^q$.

The AR(q) process can be written,

$$Z_t = (1 - \alpha_1 B - \ldots - \alpha_p B^p) Y_t = \phi(B) Y_t$$

where $\phi(B) = 1 - \alpha_1 B - \ldots - \alpha_p B^p$.

The ARMA(p, q) model is then

$$\phi(B) Y_t = \theta(B) Z_t.$$

The backward shift operator

Note in the specification of AR, MA and ARMA models:

- We require $\phi(0) = \theta(0) = 1$ to eliminate redundancy in the specification.
- In the ARMA model,

$$\phi(B)Y_t = \theta(B)Z_t$$

  we assume ϕ and θ to have no common factors.

Observe also that functions of B can be interpreted in terms of their formal series expansions.

For example,

$$\begin{aligned}\left(1 - \frac{1}{2}B\right)^{-1} Y_t &= \left(\sum_{k=0}^{\infty} \frac{1}{2^k} B^k\right) Y_t \\ &= \sum_{k=0}^{\infty} 2^{-k} Y_{t-k}.\end{aligned}$$

The Moving Average Model

Consider the moving average model,

$$Y_t = \theta(B)Z_t = \sum_{j=0}^{q} \beta_j Z_{t-j}$$

where we assume $\beta_0 = 1$.

The moving average process is always stationary with

$$\text{var}(Y_t) = \sigma^2 \sum_{j=0}^{q} \beta_j^2.$$

The autocovariance is

$$\gamma_k = \begin{cases} \sigma^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} & \text{for } k = 0, 1, 2, \ldots, q \\ 0 & \text{otherwise} \end{cases}$$

and the autocorrelation function is

$$\rho_k = \frac{\sum_{j=0}^{q-k} \beta_j \beta_{j+k}}{\sum_{j=0}^{q} \beta_j^2} \text{ for } k = 0, 1, 2, \ldots, q.$$

Invertibility

Consider the MA(1) process,

$$Y_t = Z_t + \beta Z_{t-1}.$$

The autocorrelation function is defined by the sole non-zero autocorrelation

$$\rho_1 = \frac{\beta}{1+\beta^2}.$$

Consider now the MA(1) process,

$$Y_t = Z_t + \frac{1}{\beta} Z_{t-1}.$$

It can be seen that the autocorrelation is also

$$\rho_1 = \frac{\beta}{1+\beta^2}.$$

Invertibility

In the case of the MA(1) process, the ambiguity can be resolved by requiring $|\beta| < 1$.

In this case, we can also invert the specification of the model to express Z_t as a linear filter of Y_t.

$$Y_t = (1 + \beta B) Z_t \Rightarrow Z_t = (1 + \beta B)^{-1} Y_t = \sum_{j=0}^{\infty} (-1)^j \beta^j Y_{t-j}.$$

Note that this inversion is possible if and only if $|\beta| < 1$.

In general, the MA(q) process

$$Y_t = \theta(B) Z_t$$

is said to be invertible provided the (complex) roots of $\theta(u)$ lie outside the unit circle.

The Autoregressive Process

Consider now the AR(p) process,

$$\phi(B)Y_t = Z_t$$

where $\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j$.

It can be proved that the AR(p) process is stationary if and only if all roots of $\phi(u)$ lie outside the unit circle.

Example

Consider the AR(1) process

$$(1 - \alpha B)Y_t = Z_t.$$

Multiplying both sides by $(1 - \alpha B)^{-1}$, it follows that

$$Y_t = (1 - \alpha B)^{-1} Z_t = \sum_{j=0}^{\infty} \alpha^j Z_{t-j}.$$

That is, Y_t is expressible as a general linear process. The necessary and sufficient condition for stationarity is then

$$\sum_{j=0}^{\infty} \alpha^{2j} < \infty \Rightarrow |\alpha| < 1.$$

Since $\phi(u) = 1 - \alpha u$ has root $1/\alpha$, the process is stationary as all roots lie outside the unit circle.

Proof of stationarity and invertibility

The conditions for stationarity of the AR process and invertibility of an MA process are both that the polynomial $\phi(u)$ (or $\theta(u)$ has all roots outside the unit circle.

To prove the result, consider the polynomial $\phi(u)$.

We will show that $1/\phi(u)$ is expressible in the form

$$1/\phi(u) = \sum_{j=0}^{\infty} a_j u^j \text{ with } \sum_{j=0}^{\infty} a_j^2 < \infty$$

if and only if all roots of $\phi(u)$ lie outside the unit circle.

Proof of stationarity and invertibility

- Assume for convenience that all roots of $\phi(u)$ are distinct, so that
$$\phi(u) = \prod_{j=1}^{p}(1 - b_j u).$$
- It follows that
$$1/\phi(u) = \sum_{j=1}^{p} \frac{c_j}{1 - b_j u}$$
for some constants $c_1, c_2, \ldots, c_p$.

Proof of stationarity and invertibility

- Using the fact that

$$\frac{1}{1 - b_j u} = \sum_{k=0}^{\infty} b_j^k u^k$$

we obtain

$$1/\phi(u) = \sum_{k=0}^{\infty} a_k u^k$$

where $a_k = \sum_{j=1}^{p} c_j b_j^k$.

Proof of stationarity and invertibility

- Suppose all roots lie outside the unit circle, so that each $|b_j| < b$ for some real number $b < 1$.
- It follows that
$$a_j^2 < \left(\sum_{j=1}^{p} |c_j| \right)^2 b^{2j}$$
whereby the series $\sum_{j=0}^{\infty} a_j^2$ is convergent.
- On the other hand, if $\max |b_j| > 1$ it can be shown that the series will diverge.

Causality

A time series Y_t is said to be causal, if it is expressible in form

$$Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}$$

with

$$\sum_{j=0}^{\infty} |a_j| < \infty.$$

That is, if Y_t depends on Z_t only through the values up until time t.

For example, the AR(1) process with $|\alpha| < 1$ is causal.

Autoregressive processes

Consider an AR(p) process

$$\phi(B)Y_t = Z_t$$

where

$$\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j$$

and all roots of $\phi(u)$ lie outside the unit circle.

The problem at hand is to determine the autocorrelation function from the coefficients,

$$\alpha_1, \alpha_2, \ldots, \alpha_p.$$

The Yule-Walker equations

The autocorrelation function is the solution to the system of equations

$$\rho_k = \sum_{j=1}^{p} \alpha_j \rho_{k-j}.$$

To see this, observe

$$Y_t = \sum_{j=1}^{p} \alpha_j Y_{t-j} + Z_t$$

$$\Rightarrow \quad \operatorname{cov}(Y_{t-k}, Y_t) = \sum_{j=1}^{p} \alpha_j \operatorname{cov}(Y_{t-k}, Y_{t-j)} + \operatorname{cov}(Y_{t-k}, Z_t)$$

$$\Rightarrow \quad \gamma_k = \sum_{j=1}^{p} \alpha_j \gamma_{k-j} + 0$$

$$\Rightarrow \quad \rho_k = \sum_{j=1}^{p} \alpha_j \rho_{k-j}$$

Difference Equations

The Yule-Walker equations are an example of a difference equation.

In general, consider the p-th order difference equation

$$\lambda_{k+p} + a_1\lambda_{k+p-1} + \ldots + a_p\lambda_k = b_k$$

where $a_1, a_2, \ldots, a_p$ and $\{b_k\}$ are given.

When $b_k = 0$, we obtain the homogeneous equations

$$\lambda_{k+p} + a_1\lambda_{k+p-1} + \ldots + a_p\lambda_k = 0$$

The problem is to solve for the sequence $\{\lambda_k\}$.

The auxiliary equation

It can be shown that $\lambda_k = \lambda^k$ is a solution to the homogeneous equation if and only if λ is a solution to the auxiliary equation

$$\lambda^p + a_1\lambda^{p-1} + \ldots + a_{p-1}\lambda + a_p = 0.$$

Suppose the auxiliary equation has distinct roots,

$$r_1, r_2, \ldots, r_p.$$

The general solution is

$$\lambda_k = c_1 r_1^k + c_2 r_2^k + \ldots + c_p r_p^k$$

where the constants $c_1, c_2, \ldots, c_k$ are determined from the initial conditions.

Auxiliary equation

Suppose now the auxiliary equation has roots of multiplicity greater than 1.

For example, suppose $r_1 = r_2 = \ldots = r_m$ and all other roots have multiplicity 1. The general solution is then

$$\lambda_k = (c_1 + c_2 k + c_3 k^2 + \ldots + c_m k^{m-1}) r_1^k + c_{m+1} r_{m+1}^k + \ldots + c_p r_p^k.$$

Yule-Walker equations

Consider again the Yule-Walker equations

$$\rho_k = \sum_{j=1}^{p} \alpha_j \rho_{k-j}.$$

This is expressible as the homogeneous difference equation with auxiliary equation

$$\lambda^p - \alpha_1 \lambda^{p-1} - \ldots - \alpha_{p-1}\lambda - \alpha_p = 0. \tag{5}$$

The general solution is then of the form

$$\rho_k = c_1 r_1^k + \ldots + c_p r_p^k$$

if $r_1, r_2, \ldots, r_p$ are the distinct roots of equation (5).

Example

Consider the AR(2) process

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + Z_t.$$

The Yule-Walker equations are

$$\rho_k - \alpha_1 \rho_{k-1} - \alpha_2 \rho_{k-2} = 0.$$

The auxiliary equation is

$$\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0. \tag{6}$$

Solution to auxiliary equation

The solutions to equation (6) are

$$r_1 = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \text{ and } r_2 = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

There are three cases:

- $r_1 \neq r_2$ both real;
- $r_1 = r_2$ real;
- $r_1 = \bar{r}_2$ complex.

Example distinct real roots

Consider the AR(2) process,

$$Y_t = 0.3Y_{t-1} + 0.1Y_{t-2} + Z_t.$$

Before attempting to solve the Yule-Walker equations, check that the process is stationary.

In this case,

$$\phi(u) = 1 - 0.3u - 0.1u^2.$$

The roots are $u = -5$ and $u = 2$.

Both lie outside the unit circle so the process is stationary.

Example - continued

Next consider the auxiliary equation,

$$\lambda^2 - 0.3\lambda - 0.1 = 0.$$

Note this is not the same as $\phi(\lambda) = 0$.

The solutions are

$$r_1 = -\frac{1}{5} \text{ and } r_2 = \frac{1}{2}.$$

The general solution is this

$$\rho_k = c_1 \left(-\frac{1}{5}\right)^k + c_2 \left(\frac{1}{2}\right)^k.$$

Example - continued

To solve c_1 and c_2, initial conditions are required.

Observe first that $\rho_0 = 1$.

Now evaluate the Yule-Walker equations directly for ρ_1.

$$\rho_1 - 0.3\rho_0 - 0.1\rho_{-1} = 0.$$

Using the fact that, $\rho_{-1} = \rho_1$, it follows that $\rho_1 = 1/3$.

Example - continued

Finally solve for c_1 and c_2 in the system of linear equations

$$\begin{aligned} 1 &= c_1 + c_2 \\ \frac{1}{3} &= -c_1\frac{1}{5} + c_2\frac{1}{2} \end{aligned}$$

It can be checked that the solutions are

$$c_1 = \frac{5}{21} \text{ and } c_2 = \frac{16}{21}.$$

Example distinct real roots

Consider the AR(2) process,

$$Y_t = 0.3Y_{t-1} + 0.1Y_{t-2} + Z_t.$$

Last time, showed that the autocorrelation function was

$$\rho_k = \frac{5}{21}\left(-\frac{1}{5}\right)^k + \frac{16}{21}\left(\frac{1}{2}\right)^k$$

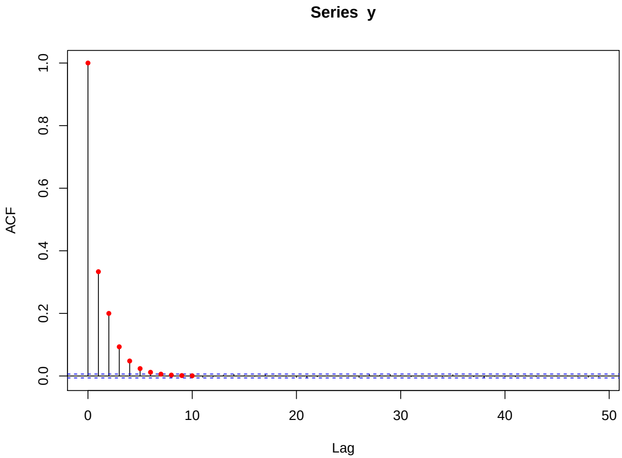
Example - continued

Can check our answer using simulation.

```
z=rnorm(100000)
y=rep(0,100000)
for(i in 3:100000) {y[i]=0.3*y[i-1]+0.1*y[i-2]+z[i]}
y=y[-c(1:1000)]
y=ts(y)
k=c(0:10)
r=(5*(-0.2)^k+16*0.5^k)/21
```

Example - continued

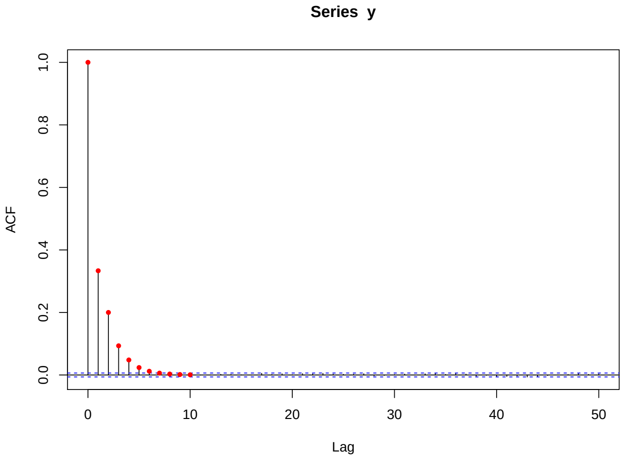
```
acf(y); points(k,r,pch=20,col="red")
```



Example - continued

Note we can also do the simulation using the inbuilt `arima.sim` function.

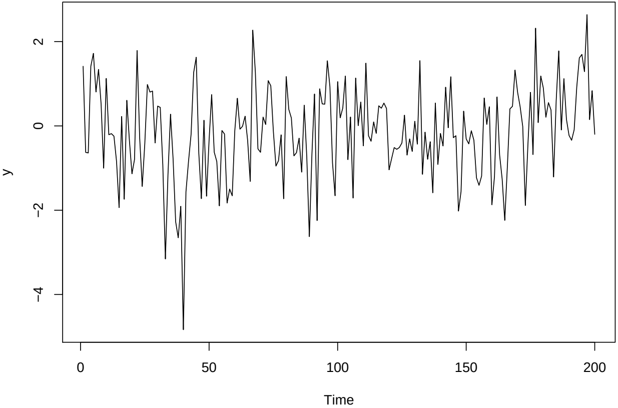
```
y=arima.sim(model=list(ar=c(0.3,0.1)),n=100000)
acf(y); points(k,r,pch=20,col="red")
```



Example - continued

Can also examine behaviour of the time series.

```
y=ts(y[1:200])
plot(y)
```



Example - repeated root

Consider the AR(2) process,

$$Y_t = Y_{t-1} - 0.25Y_{t-2} + Z_t.$$

Before attempting to solve the Yule-Walker equations, check that the process is stationary.

In this case,

$$\phi(u) = 1 - u + 0.25u^2.$$

There is a single root, $u = 2$, which lies outside the unit circle so the process is stationary.

Example - continued

Next consider the auxiliary equation,

$$\lambda^2 - \lambda + 0.25 = 0.$$

The single solution is $r_1 = 0.5$.

The general solution is then

$$\rho_k = (c_1 + c_2 k)0.5^k.$$

Example - continued

To solve for c_1 and c_2, initial conditions are required.

Observe first that $\rho_0 = 1$.

Now evaluate the Yule-Walker equations directly for ρ_1.

$$\rho_1 - \rho_0 + 0.25\rho_{-1} = 0.$$

Using the fact that, $\rho_{-1} = \rho_1$, it follows that $\rho_1 = 4/5$.

Example - continued

Finally solve for c_1 and c_2 in the system of linear equations

$$\begin{aligned} 1 &= c_1 \\ \frac{4}{5} &= (c_1 + c_2)0.5 \end{aligned}$$

It can be checked that the solutions are

$$c_1 = 1 \text{ and } c_2 = \frac{3}{5}.$$

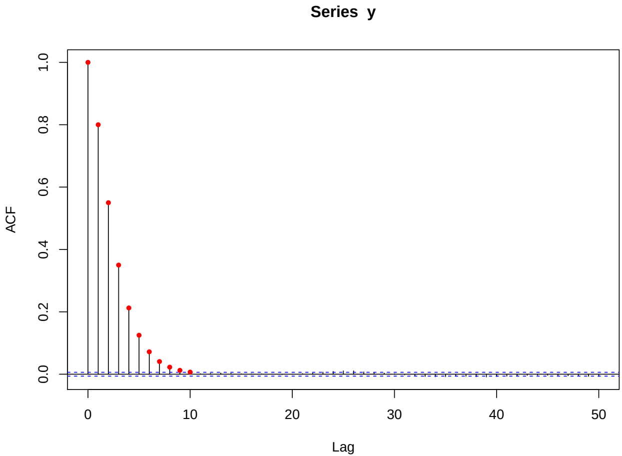
Hence the autocorrelation function is

$$\rho_k = (1 + 0.6k)0.5^k$$

Example - continued

Can check our answer using simulation.

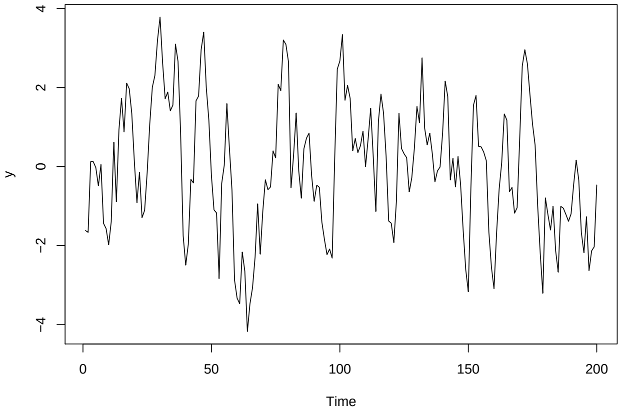
```
y=arima.sim(model=list(ar=c(1,-0.25)),n=100000)
k=c(0:10); r=(1+0.6*k)*0.5^k
acf(y); points(k,r,pch=20,col="red")
```



Example - continued

Can also examine behaviour of the time series.

```
y=ts(y[1:200])
plot(y)
```



Complex roots

When the auxiliary equation has two complex roots, $r_1,\ r_2$, the general solution is of the form

$$\rho_k = c_1 r_1^k + c_2 r_2^k$$

where c_1 and c_2 are complex coefficients.

Recall that $r_2 = \bar{r}_1$, so we must have $c_2 = \bar{c}_1$ to ensure that ρ_k is real.

Now let

$$r_1 = re^{i\theta} \text{ and } c_1 = ce^{i\phi}.$$

Complex roots

$$\begin{aligned} c_1 r_1^k + \bar{c}_1 \bar{r}_1^k &= c r^k \left(e^{i(k\theta+\phi)} + e^{-i(k\theta+\phi)} \right) \\ &= 2cr^k \cos(k\theta + \phi) \\ &= 2cr^k (\cos(\phi)\cos(k\theta) - \sin(\phi)\sin(k\theta)) \\ &= r^k (a\cos(k\theta) + b\sin(k\theta)) \end{aligned}$$

where

$$a = 2c\cos(\phi) \text{ and } b = -2c\sin(\phi).$$

Hence if, $r_1 = re^{i\theta}$, the general solution is

$$\rho_k = r^k (a\cos(k\theta) + b\sin(k\theta)).$$

Example - complex roots

Consider the AR(2) process

$$Y_t = 0.5Y_{t-1} - 0.25Y_{t-2} + Z_t.$$

Check first that the process is stationary. In this case,

$$\phi(u) = 1 - 0.5u + 0.25u^2$$

which has roots $r_1 = 1 - i\sqrt{3}$ and $r_2 = 1 + i\sqrt{3}$ which lie outside the unit circle, so the process is stationary.

Example - continued

The auxiliary equation is

$$\lambda^2 - 0.5\lambda + 0.25 = 0$$

which has solutions

$$r_1 = \frac{1}{4}(1 + i\sqrt{3}) \text{ and } r_2 = \frac{1}{4}(1 - i\sqrt{3}).$$

That is,

$$r_1 = 0.5e^{i\pi/3} \text{ and } r_2 = 0.5e^{-i\pi/3}$$

so the general solution has the form

$$\rho_k = 0.5^k(a\cos(k\pi/3) + b\sin(k\pi/3)).$$

Example - continued

To solve for a and b, initial conditions are required.

Observe first that $\rho_0 = 1$.

Now evaluate the Yule-Walker equations directly for ρ_1.

$$\rho_1 - 0.5\rho_0 + 0.25\rho_{-1} = 0.$$

Using the fact that, $\rho_{-1} = \rho_1$, it follows that $\rho_1 = 0.4$.

Example - continued

Finally solve for a and b in the system of linear equations

$$\begin{aligned} 1 &= a \\ 0.4 &= 0.25(a + b\sqrt{3}) \end{aligned}$$

It can be checked that the solutions are

$$a = 1 \text{ and } b = \frac{\sqrt{3}}{5} = 0.346.$$

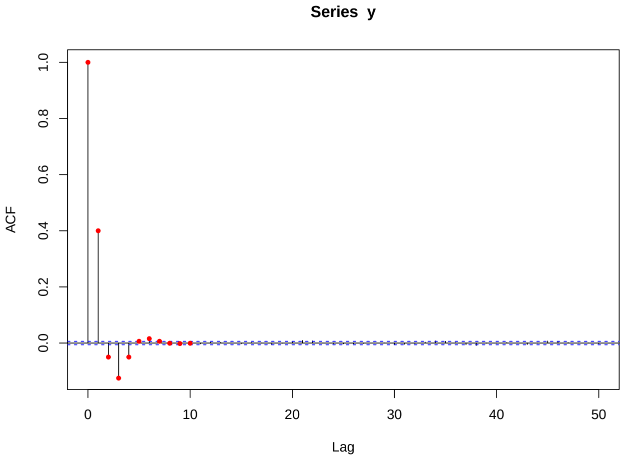
Hence the autocorrelation function is

$$\rho_k = 0.5^k(\cos(k\pi/3) + \sqrt{3}\sin(k\pi/3)/5)$$

Example - continued

Can check our answer using simulation.

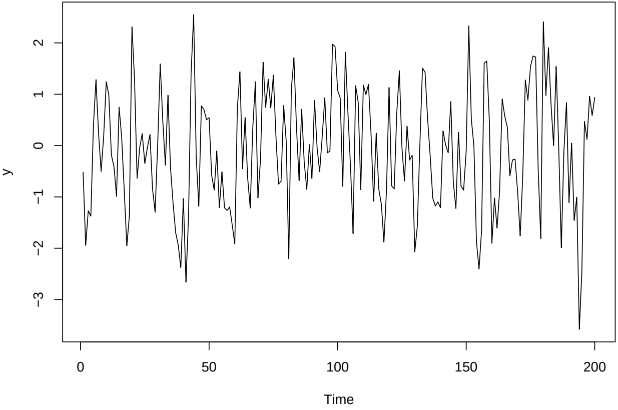
```
y=arima.sim(model=list(ar=c(0.5,-0.25)),n=100000)
k=c(0:10); r=0.5^k*(cos(k*pi/3)+sqrt(3)*sin(k*pi/3)/5)
acf(y); points(k,r,pch=20,col="red")
```



Example - continued

Can also examine behaviour of the time series.

```
y=ts(y[1:200])
plot(y)
```



ARMA Processes

The ARMA(p, q) model is specified by

$$\phi(B)Y_t = \theta(B)Z_t$$

where $\phi(B)$ and $\theta(B)$ are polynomials of degree p and q respectively with no common factors and $\phi(0) = \theta(0) = 1$.

For the process to be stationary, all roots of ϕ must lie outside the unit circle.

For the process to be invertible, all roots of θ must lie outside the unit circle.

The autocorrelation function can be found by expressing Y_t as a general linear process

$$Y_t = \{\phi(B)\}^{-1}\theta(B)Z_t.$$

Example

Consider the ARMA$(1,1)$ process

$$Y_t = \alpha Y_{t-1} + Z_t + \beta Z_{t-1}$$

with $|\alpha| < 1$ and $|\beta| < 1$.

To express this as a general linear process, observe

$$\phi(B) Y_t = \theta(B) Z_t$$

where

$$\phi(B) = 1 - \alpha B \text{ and } \theta(B) = 1 + \beta B$$

Example - continued

Hence,

$$\begin{aligned}
Y_t &= \{\phi(B)\}^{-1}\theta(B)Z_t \\
&= (1+\alpha B+\alpha^2B^2+\dots)(1+\beta B)Z_t \\
&= \{1+(\alpha+\beta)B+(\alpha^2+\alpha\beta)B^2+(\alpha^3+\alpha^2\beta)B^3+\dots\}Z_t \\
&= \{1+(\alpha+\beta)(B+\alpha B^2+\alpha^2B^3+\dots)\}Z_t \\
&= Z_t+(\alpha+\beta)\{Z_{t-1}+\alpha Z_{t-2}+\alpha^2Z_{t-3}+\dots\}
\end{aligned}$$

so the series is expressible in the general form

$$Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}.$$

The autocovariances can be found from the result

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} a_j a_{k+j}.$$

Example - continued

For example,

$$\begin{aligned}
\gamma_0 &= \sigma^2 \left\{ 1 + (\alpha+\beta)^2 (1 + \alpha^2 + \alpha^4 + \ldots) \right\} \\
&= \sigma^2 \left\{ 1 + \frac{(\alpha+\beta)^2}{1-\alpha^2} \right\} \\
&= \sigma^2 \frac{1 + \beta^2 + 2\alpha\beta}{1-\alpha^2}
\end{aligned}$$

It can also be shown that

$$\begin{aligned}
\rho_1 &= \frac{(1+\alpha\beta)(\alpha+\beta)}{(1+\beta^2+2\alpha\beta)} \\
\rho_k &= \alpha\rho_{k-1} \quad \text{for} \quad k = 2, 3, \ldots
\end{aligned}$$

Spectral Analysis

Recall that the periodogram was defined by

$$I(\omega) = \frac{1}{n}\left\{\left(\sum_{t=1}^{n} y_t \cos(\omega t)\right)^2 + \left(\sum_{t=1}^{n} y_t \sin \omega t)\right)^2\right\}.$$

and we showed

$$I(\omega) = g_0 + 2\sum_{k=1}^{n-1} g_k \cos(k\omega)$$

so that or equivalently

$$I(\omega)/g_0 = 1 + 2\sum_{k=1}^{n-1} r_k \cos(k\omega).$$

The spectrum

Consider now a stationary process with autocovariance function γ_k and autocorrelation function ρ_k.

The theoretical counterpart to the periodogram is the spectrum, defined by,

$$
\begin{aligned}
f(\omega) &= \sum_{j=-\infty}^{\infty} \gamma_j e^{i\omega j} \\
&= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \left(e^{i\omega k} + e^{-i\omega k}\right) \\
&= \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos(k\omega)
\end{aligned}
$$

The normalized spectrum is defined by

$$f^*(\omega) = \frac{f(\omega)}{\gamma_0} = 1 + 2\sum_{k=1}^{\infty} \rho_k \cos(k\omega).$$

The periodogram

In this light, the periodogram can be seen to be an estimate of the spectrum.

However, the naively calculated periodogam is not a consistent estimate of the spectrum.

Assuming that the spectrum is smooth, the estimate can be improved by smoothing.

For this reason, the periodogram estimates calculated in R are always smoothed.

The inverse transformation

The autocovariance function can be recovered from the spectrum via the inverse transformation

$$\gamma_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} f(\omega) d\omega.$$

This result follows from the fact that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} e^{ik'\omega} d\omega = \begin{cases} 1 & \text{if } k = k' \\ 0 & \text{otherwise.} \end{cases}$$

The since $f(\omega) = f(-\omega)$, the inverse transformation can also be expressed as

$$\gamma_k = \frac{1}{\pi} \int_{0}^{\pi} \cos(k\omega) f(\omega) d\omega.$$

Wold's Theorem

The inversion formula can also be shown to characterise the autocovariance function.

It can be proved that any integrable $f(\omega)$ is a legitimate spectrum and consequently every legitimate autocovariance function is expressible in the form

$$\gamma_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\omega} f(\omega) d\omega.$$

This proof is omitted.

However, it is an important result because the conditions required to ensure γ_k is a legitimate autocovariance function are very complicated.

Example

Consider the AR(1) process,

$$Y_t = \alpha Y_{t-1} + Z_t \text{ with } |\alpha| < 1.$$

We have shown previously that

$$\rho_k = \alpha^{|k|}$$

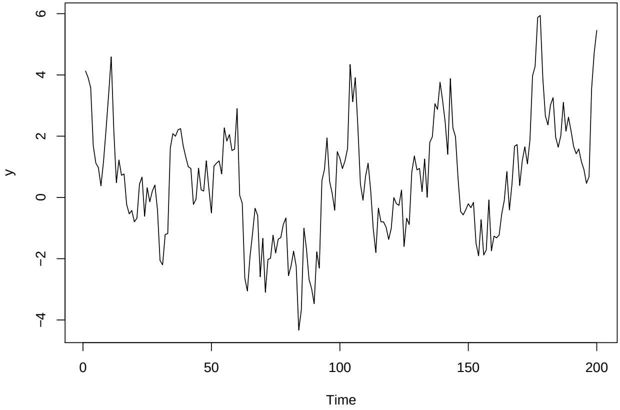
for $k = 0, \pm 1, \pm 2 \ldots$.

It can be shown that

$$f^*(\omega) = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos \omega}.$$

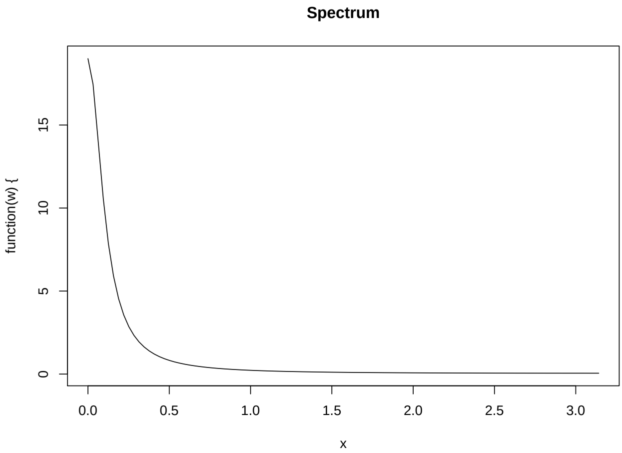
Example $\alpha = 0.9$

```
a=0.9
y=arima.sim(model=list(ar=a),n=200); plot(y)
```



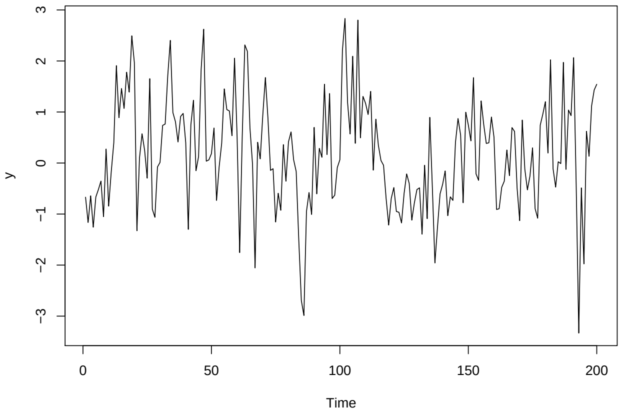
Example $\alpha = 0.9$

```
plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



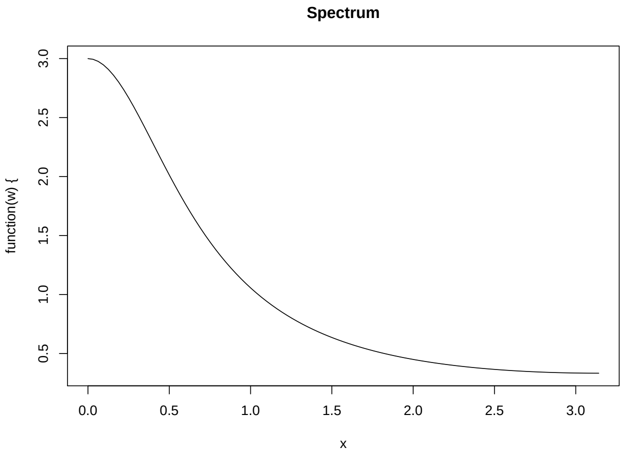
Example $\alpha = 0.5$

```
a=0.5;
y=arima.sim(model=list(ar=a),n=200); plot(y)
```



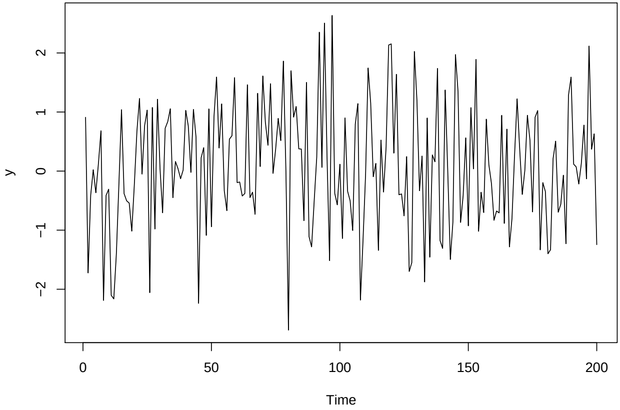
Example $\alpha = 0.5$

```
plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



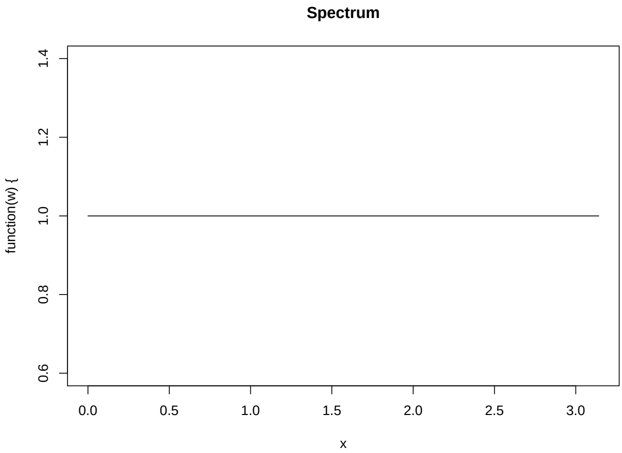
Example $\alpha = 0$ (white noise)

```
a=0;
y=ts(rnorm(200)); plot(y)
```



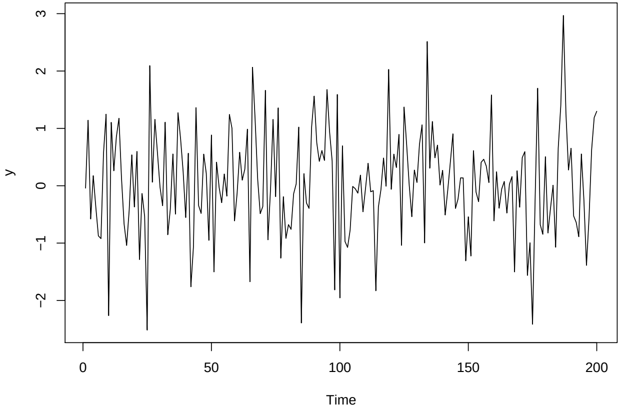
Example $\alpha = 0$ (white noise)

```
plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



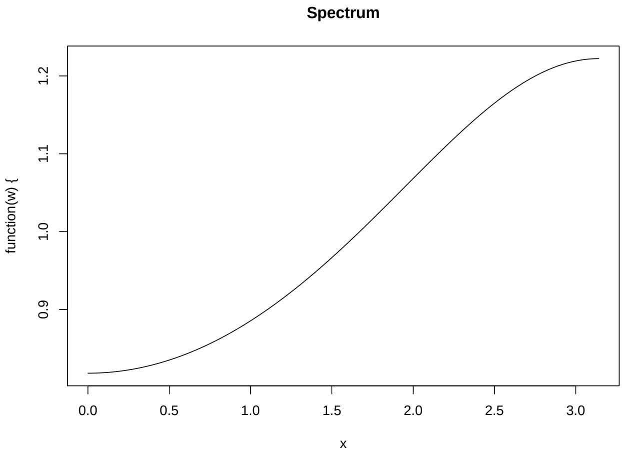
Example $\alpha = -0.1$

```
a=-0.1;
y=arima.sim(model=list(ar=a),n=200); plot(y)
```



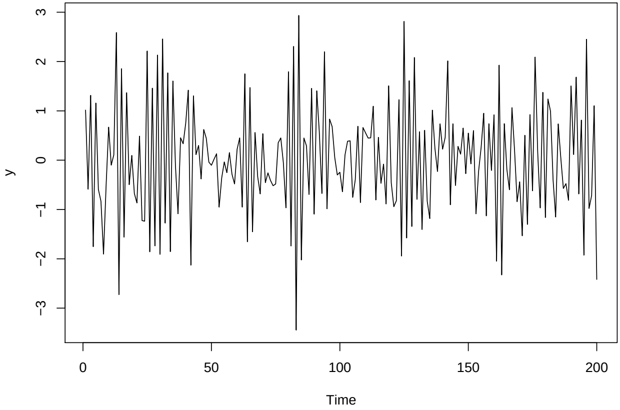
Example $\alpha = -0.1$

```
plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



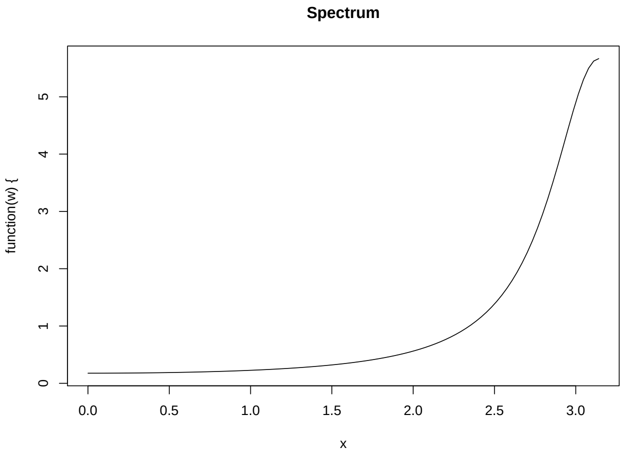
Example $\alpha = -0.7$

```
a=-0.7;
y=arima.sim(model=list(ar=a),n=200); plot(y)
```



Example $\alpha = -0.7$

```
plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



Spectral Analysis - Aliasing

For a discrete time process $\{Y_t\}$, interpretation of the spectrum is subject to the phenomenon called aliasing.

Consider a purely deterministic signal,

$$y_t = \cos(\omega t) \text{ for } t = 0, \pm 1, \pm 2, \ldots$$

where $0 < \omega < \pi$.

Let

$$y_t' = \cos((\omega + 2k\pi)t)$$

for any integer k and observe

$$y_t' = y_t.$$

Spectral Analysis - Aliasing

Similarly if

$$y'_t = \cos((-\omega + 2k\pi)t)$$

It follows from the fact that $\cos(x) = \cos(-x)$ that

$$y'_t = y_t$$

Hence we cannot distinguish between the frequencies $0 < \omega < \pi$ and frequencies outside this range for data recorded integer time points.

Convergence of the spectrum

Recall that the spectrum, $f(\omega)$, is defined by

$$f(\omega) = \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos(k\omega).$$

It is relevant to ask whether the spectrum always converges for a stationary process.

The following example shows that this is not always the case.

Divergent spectrum example

Consider a white noise process Z_t with $var(Z_t) = \sigma^2$ and let A and B be independent random variables such that

$$E(A) = E(B) = 0 \text{ and } \operatorname{var}(A) = \operatorname{var}(B) = \tau^2.$$

Let

$$Y_t = A\cos(\theta t) + B\sin(\theta t) + Z_t.$$

It can be checked that

$$\begin{aligned} E(Y_t) &= 0 \\ \operatorname{var}(Y_t) &= \sigma^2 + \tau^2 \\ \operatorname{cov}(Y_t, Y_s) &= \tau^2\cos(\theta(t-s)). \end{aligned}$$

Example - continued

Hence, the series $\{Y_t\}$ is second order stationary and the spectral density is

$$f(\omega) = \sigma^2 + \tau^2 + 2\tau^2 \sum_{k=1}^{\infty} \cos(k\theta)\cos(k\omega).$$

If $\theta = \omega$ then the summation diverges since

$$\sum_{k=1}^{\infty} \cos(k\theta)^2 = \infty.$$

- This example, shows that strictly speaking, spectrum need not always converge.
- The practical significance of this example is questionable.
  - For a single realisation, it would be more natural to treat A and B as constants that define a non-stationary process.

Linear Filters

Consider a stationary process U_t and recall that a linear filter is defined by

$$Y_t = \sum_{j=-\infty}^{\infty} a_j U_{t-j}.$$

Assuming convergence, the autocovariance functions can be seen to satisfy

$$\gamma_Y(k) = \sum_{\ell=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_\ell a_j \gamma_U(k + \ell - j).$$

Consider now, the spectrum

$$f_Y(\omega) = \sum_{k=-\infty}^{\infty} e^{ik\omega} \gamma_Y(k).$$

Linear Filters - continued

It can then be shown that

$$
\begin{aligned}
f_Y(\omega) &= \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} e^{ik\omega} a_\ell a_j \gamma_U(k+\ell-j) \\
&= f_U(\omega) \left(\sum_{\ell=-\infty}^{\infty} a_\ell e^{-i\ell\omega} \right) \left(\sum_{j=-\infty}^{\infty} a_j e^{ij\omega} \right) \\
&= |a(\omega)|^2 f_U(\omega)
\end{aligned}
$$

where

$$a(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{ij\omega}.$$

Example

Consider the three point moving average filter, defined by

$$a_{-1} = a_0 = a_1 = \frac{1}{3}.$$

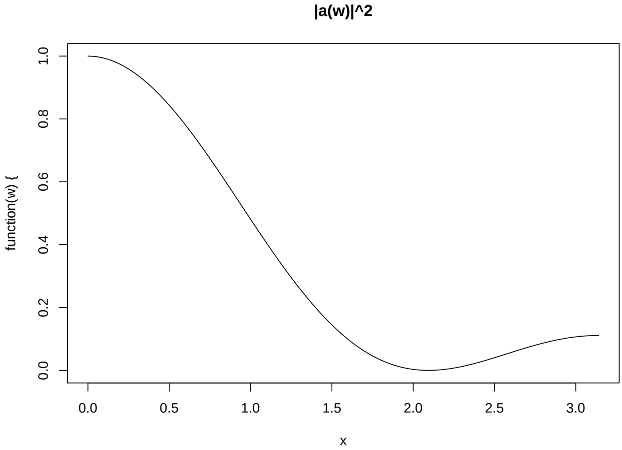
$$\begin{aligned} a(\omega) &= \sum_{j=-1}^{1} a_j e^{ij\omega} \\ &= \frac{1}{3}\left(e^{-i\omega} + 1 + e^{i\omega}\right) \\ &= \frac{1}{3}(1 + 2\cos\omega) \end{aligned}$$

and, hence,

$$|a(\omega)|^2 = (1 + 2\cos\omega)^2/9.$$

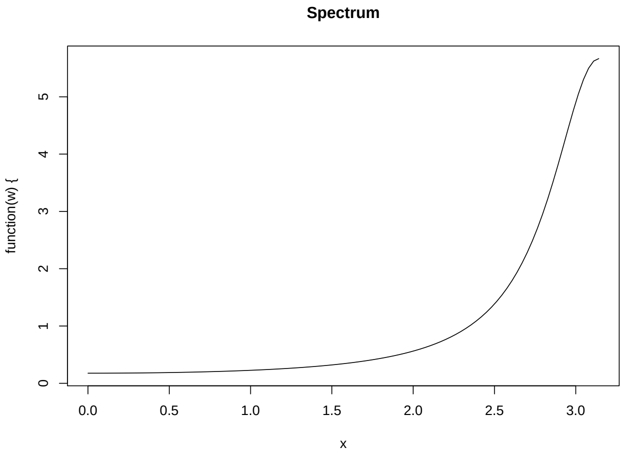
Example - 3 point moving average

```
plot(function(w){(1+2*cos(w))^2/9},0,pi)
title(main="|a(w)|^2")
```



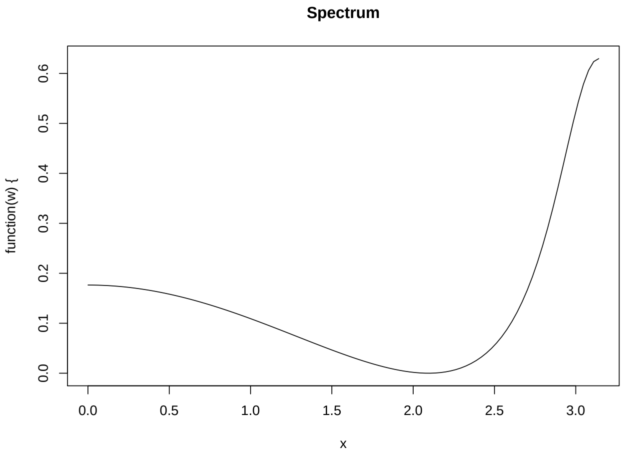
Example AR(1) $\alpha = -0.7$

```
a=-0.7; plot(function(w){(1-a^2)/(1+a^2-2*a*cos(w))},0,pi)
title(main="Spectrum")
```



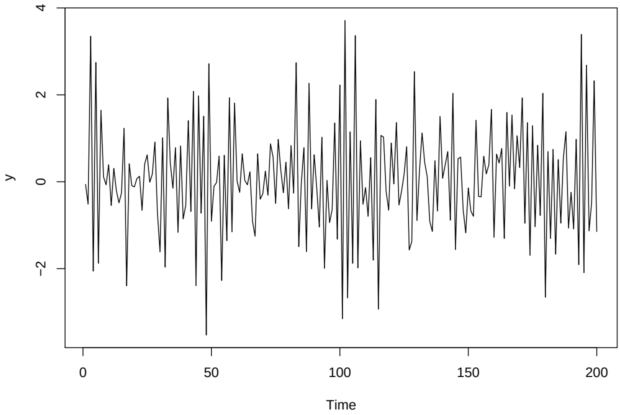
Example Filtered AR(1) $\alpha = -0.7$

```
plot(function(w){(1+2*cos(w))^2*(1-a^2)/((1+a^2-2*a*cos(w))
title(main="Spectrum")
```



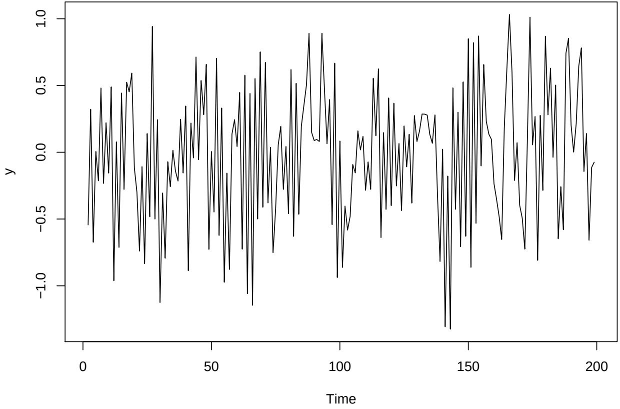
Example AR(1) $\alpha = -0.7$

```
a=-0.7;
y=arima.sim(model=list(ar=a),n=200); plot(y)
```



Example Filtered AR(1) $\alpha = -0.7$

```
a=-0.7;
u=arima.sim(model=list(ar=a),n=200);
y=filter(u,filter=c(1,1,1)/3)
plot(y)
```



Spectra for ARMA processes

Recall that a linear filter is defined by

$$U_t = \sum_{j=-\infty}^{\infty} a_j Y_{t-j}.$$

Last time we proved that

$$f_u(\omega) = |a(\omega)|^2 f_y(\omega) \tag{7}$$

where

$$a(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{ij\omega}$$

is called the *transfer function*.

This fact allows us to calculate the spectrum for all MA, AR and ARMA processes.

Moving Average Processes

Consider an MA(q) process,

$$Y_t = Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} = \theta(B) Z_t$$

where Z_t is a white noise process with $\text{var}(Z_t) = \sigma^2$.

Since Y_t is a linear filter of the white noise process Z_t, it follows immediately that

$$f_y(\omega) = |\theta(e^{i\omega})|^2 f_z(\omega).$$

Since

$$\theta(e^{i\omega}) = 1 + \sum_{j=1}^{q} \beta_j \cos(\omega j) + i \sum_{j=1}^{q} \beta_j \sin(\omega j) \text{ and } f_z(\omega) = \sigma^2$$

it follows that

$$f_y(\omega) = \sigma^2 \left(\left(1 + \sum_{j=1}^{q} \beta_j \cos(\omega j) \right)^2 + \left(\sum_{j=1}^{q} \beta_j \sin(\omega j) \right)^2 \right).$$

Example

Consider the MA(3) process

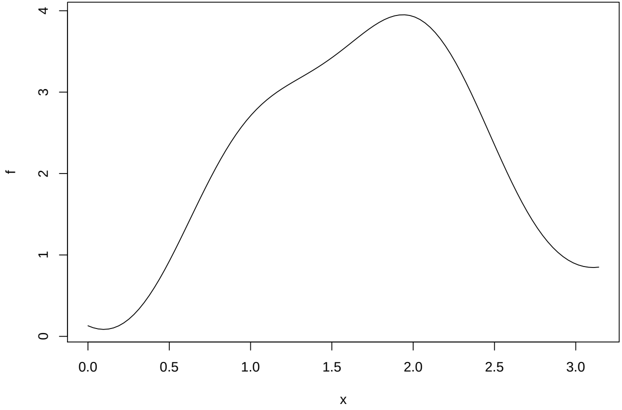
$$Y_t = Z_t - 0.8Z_{t-1} - 0.7Z_{t-2} + 0.2Z_{t-3}.$$

The spectrum is

$$\begin{aligned} f(\omega) &= \sigma^2 \Big((1 - 0.8\cos(\omega) - 0.7\cos(2\omega) + 0.2\cos(3\omega))^2 \\ &\quad + (-0.8\sin(\omega) - 0.7\sin(2\omega) + 0.2\sin(3\omega))^2 \Big). \end{aligned}$$

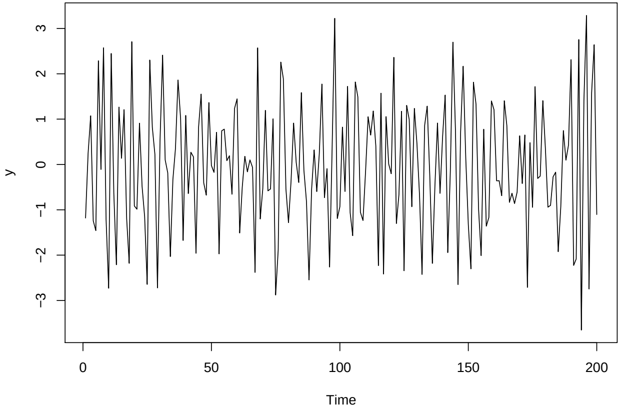
Example - continued

```
f=function(w){(1-0.8*cos(w)-0.7*cos(2*w)+0.2*cos(3*w))^2+
(-0.8*sin(w)-0.7*sin(2*w)+0.2*cos(3*w))^2}
plot(f,0,pi)
```



Example - continued

```
y=arima.sim(model=list(ma=c(-0.8,-0.7,0.2)),n=200); plot(y)
```



Autoregressive processes

Consider the AR(p) process,

$$\phi(B)Y_t = Z_t$$

where Z_t is a white noise process with $\mathrm{var}(Z_t) = \sigma^2$.

It follows that

$$|\phi(e^{i\omega})|^2 f_y(\omega) = f_z(\omega) = \sigma^2$$

and hence

$$f_y(\omega) = \frac{\sigma^2}{|\phi(e^{i\omega})|^2}.$$

Note that for a stationary AR process, all roots of ϕ lie outside the unit circle so the spectrum is defined for all ω.

Example

Consider the AR(2) process

$$Y_t = 0.8Y_{t-1} - 0.7Y_{t-2} + Z_t.$$

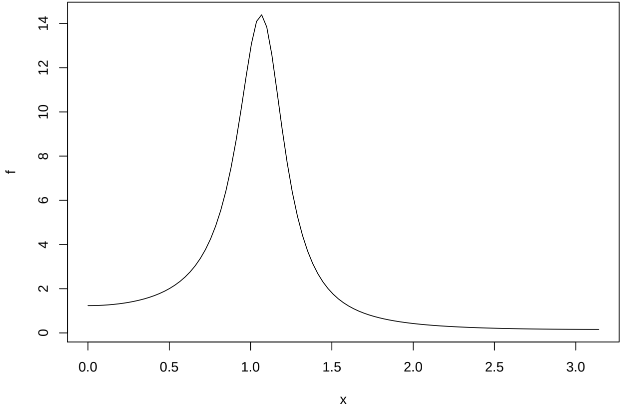
In this case, $\phi(u) = 1 - 0.8u + 0.7u^2$ has roots $r_1 = (4 - 3i\sqrt{6})/7$ and $r_2 = (4 + 3i\sqrt{6})/7$, so the process is stationary.

The spectrum is

$$f(\omega) = \frac{\sigma^2}{(1 - 0.8\cos(\omega) + 0.7\cos(2\omega))^2 + (-0.8\sin(\omega) + 0.7\sin(2\omega))^2}$$

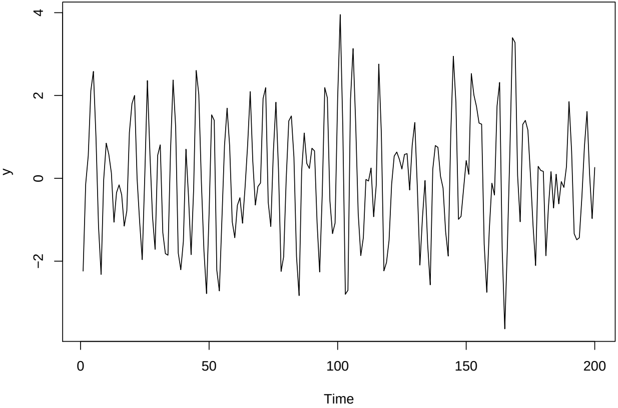
Example - continued

```
f=function(w){1/((1-0.8*cos(w)+0.7*cos(2*w))^2
+(-0.8*sin(w)+0.7*sin(2*w))^2)}
plot(f,0,pi)
```



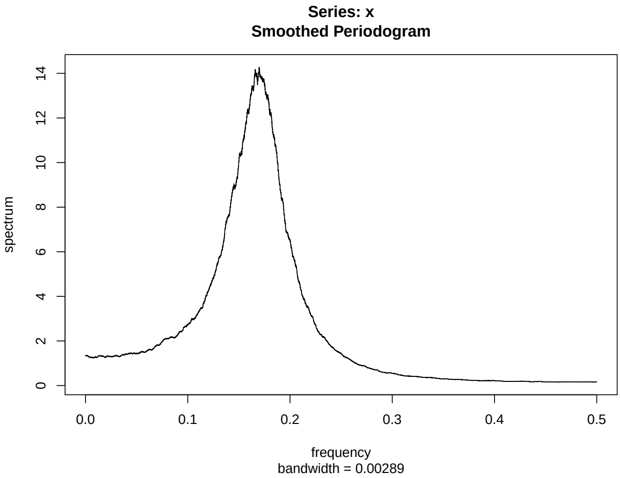
Example - continued

```
y=arima.sim(model=list(ar=c(0.8,-0.7)),n=200)
plot(y)
```



Example - continued

```r
y=arima.sim(model=list(ar=c(0.8,-0.7)),n=100000)
spectrum(y,log="no",spans=1000)
```



Autoregressive moving average processes

Consider the ARMA(p, q) process,

$$\phi(B)Y_t = \theta(B)Z_t$$

where

$$\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j B^j \text{ and } \theta(B) = 1 + \sum_{j=1}^{q} \beta_j B^j$$

such that ϕ and θ have no common factors and all roots of ϕ lie outside the unit circle.

Calculating the spectrum for both sides, yields

$$|\phi(e^{i\omega})|^2 f_y(\omega) = |\theta(e^{i\omega})|^2 f_z(\omega)$$

and, hence,

$$f_y(\omega) = \sigma^2 \frac{|\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2}.$$

Example

Consider the ARMA$(2,2)$ process

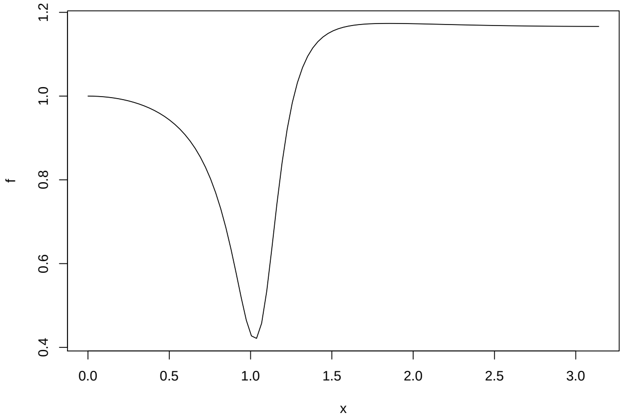
$$Y_t = 0.8Y_{t-1} - 0.7Y_{t-2} + Z_t - 0.9Z_{t-1} + 0.8Z_{t-2}.$$

In this case, $\phi(u) = 1 - 0.8u + 0.7u^2$ and $\theta(u) = 1 - 0.9u + 0.8u^2$ so the spectrum is

$$f(\omega) = \sigma^2 \frac{(1 - 0.9\cos(\omega) + 0.8\cos(2\omega))^2 + (-0.9\sin(\omega) + 0.8\sin(2\omega))^2}{(1 - 0.8\cos(\omega) + 0.7\cos(2\omega))^2 + (-0.8\sin(\omega) + 0.7\sin(2\omega))^2}$$

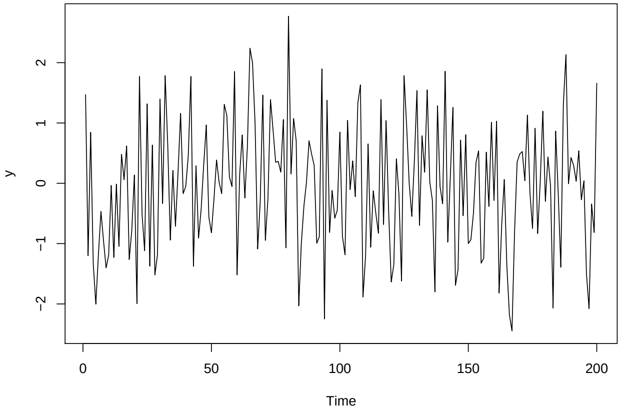
Example - continued

```
f=function(w){((1-0.9*cos(w)+0.8*cos(2*w))^2+
(-0.9*sin(w)+0.8*sin(2*w))^2)/((1-0.8*cos(w)+0.7*cos(2*w))^
(-0.8*sin(w)+0.7*sin(2*w))^2)}
plot(f,0,pi)
```



Example - continued

```
y=arima.sim(model=list(ar=c(0.8,-0.7),
ma=c(-0.9,0.8)),n=200)
plot(y)
```



Processes with continuous spectra

A further consequence of equation (7) is that every process with a continuous spectrum is expressible as a general linear process.

Recall that the general linear process has the form

$$Y_t = \sum_{j=0}^{\infty} a_j Z_{t-j}$$

where Z_t is a white noise process.

The spectrum of Y_t is

$$\begin{aligned} \sigma^2 |a(\omega)|^2 &= \sigma^2 \left\{ \left(\sum_{j=0}^{\infty} a_j e^{ij\omega} \right) \left(\sum_{k=0}^{\infty} a_k e^{-ik\omega} \right) \right\} \\ &= \sigma^2 \left(b_0 + \sum_{m=1}^{\infty} b_m \cos(\omega) \right) \end{aligned}$$

Processes with continuous spectra

A key result in Fourier analysis is that every real valued, continuous function, f, such that

$$f(-\omega) = f(\omega)$$

and such that

$$f(\omega + 2m\pi) = f(\omega) \text{ for all integers } m,$$

is expressible in the form

$$f(\omega) = \sigma^2 \left(b_0 + \sum_{m=1}^{\infty} b_m \cos(m\omega) \right).$$

Processes with continuous spectra

On the other hand, the spectrum of a second-order stationary process is a real valued function such that

$$f(\omega) = f(-\omega) \text{ and } f(\omega + 2m\pi) = f(\omega)$$

Consequently, any stationary process with continuous spectrum can be represented as a general linear process, for the purposes of its second order properties.

ARIMA Models

ARMA models provide a flexible family of time series models, all of which are stationary.

The AutoRegressive Integrated Moving Average (ARIMA) models, provide a framework in which non-stationarity can be accommodated.

The ARIMA(p, d, q) model is defined by

$$\phi(B)(1-B)^d Y_t = \theta(B) Z_t$$

where

$$\phi(B) \text{ and } \theta(B)$$

are polynomials of degree p and q respectively such that all roots lie outside the unit circle.

ARIMA Models

The introduction of the difference term, $(1-B)^d$ allows for non-stationarity.

For example, the process

$$(1-B)Y_t = Z_t$$

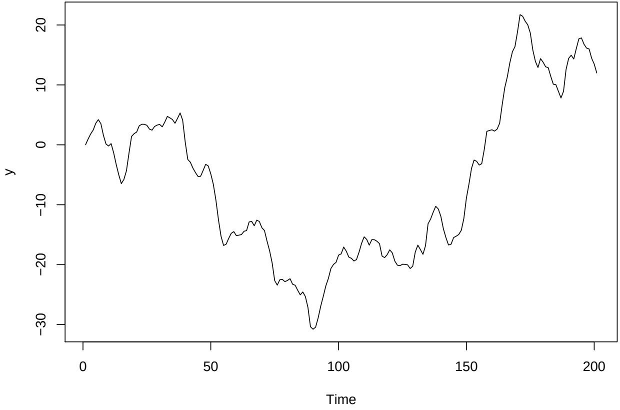
is a random walk, which is known to be non-stationary.

Put another way, the formulation of the ARIMA(p, d, q) process can be interpreted equivalently as the d^{th} order difference being an ARMA(p, q) process.

The ARIMA model provides a more formal structure in which differencing is applied to remove non-stationarity.

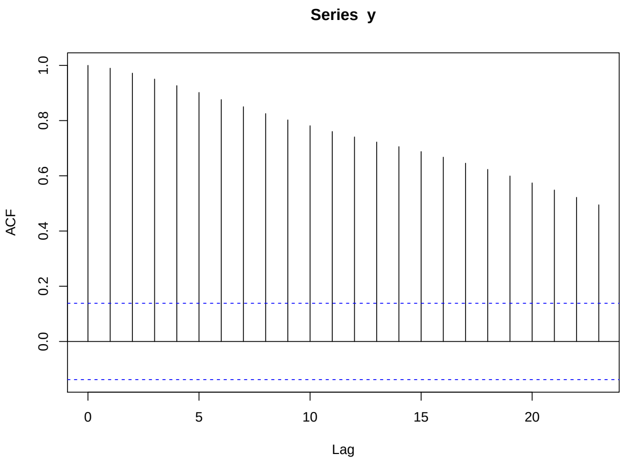
Example *arima*$(1,1,1), \alpha = 0.4, \beta = 0.5$

```
y=arima.sim(list(order=c(1,1,1),ar=0.4,ma=0.5),n=200)
plot(y)
```



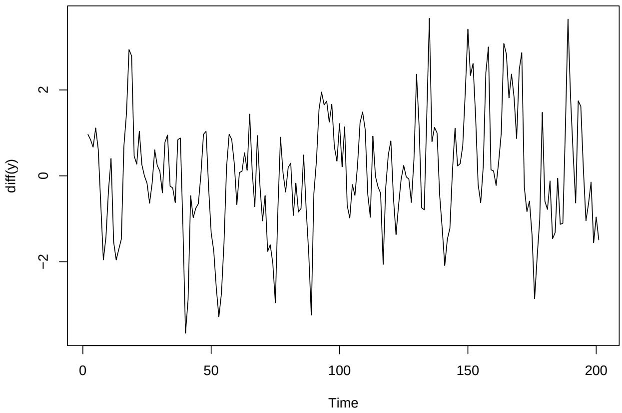
Example (continued)

```
acf(y)
```



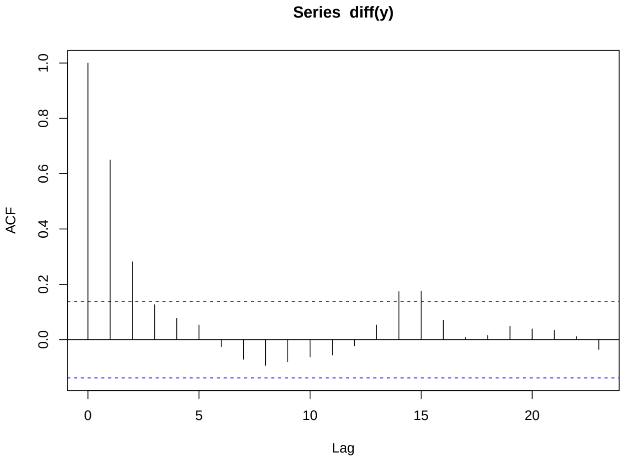
Example (continued)

```
plot(diff(y))
```



Example (continued)

```
acf(diff(y))
```



Identification of ARIMA Models

Given an observed time series, the problem at hand is to identify a suitable ARIMA model by choosing p, d, q appropriately and to estimate the unknown coefficients.

The main steps are

- Choose an appropriate value d.
- Find p and q to minimize the AIC.

Choice of d

The order of difference d is often chosen informally by trial and error.

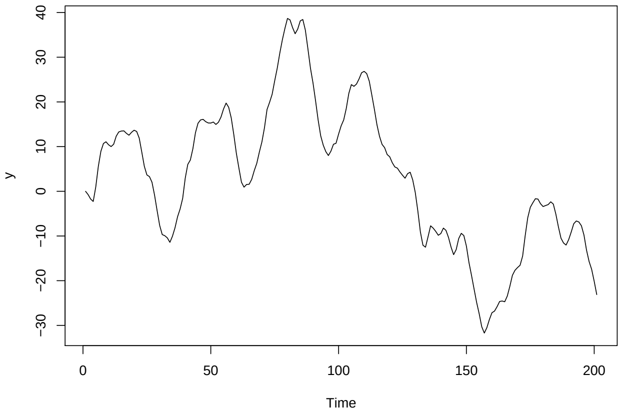
The purpose of differencing is to reduce a non-stationary series to one that is stationary.

This step is often achieved by inspection of the time series plot and the correlogram.

Recall that a typical signature of non-stationarity is a correlogram in which the correlations decrease very slowly and roughly linearly.

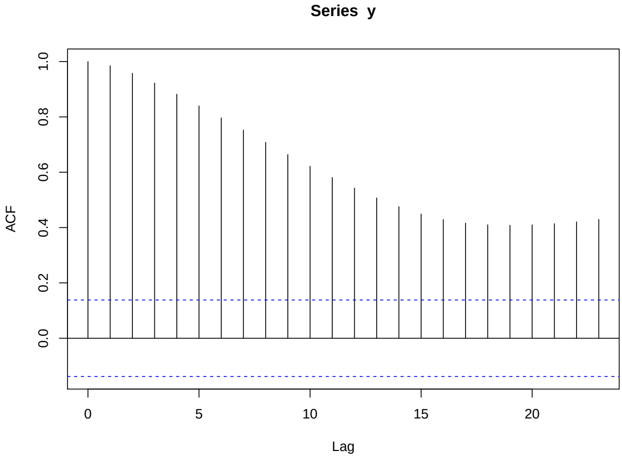
Example

```r
plot(y)
```



Example (continued)

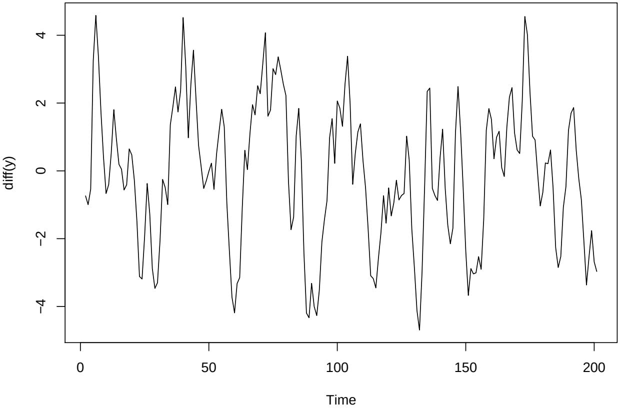
```
acf(y)
```



Plot shows typical non-stationarity.

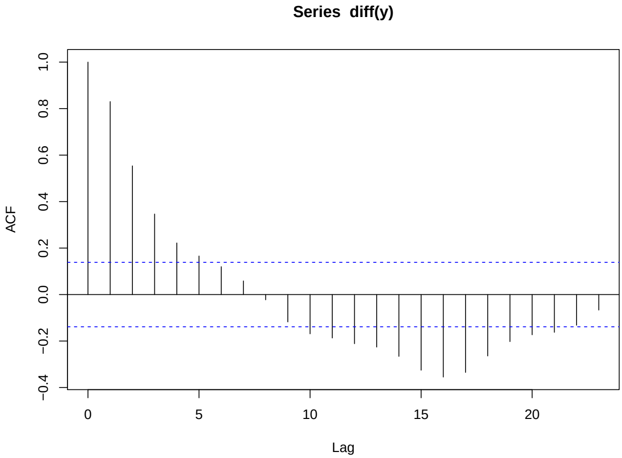
Example (continued)

```
plot(diff(y))
```



Example (continued)

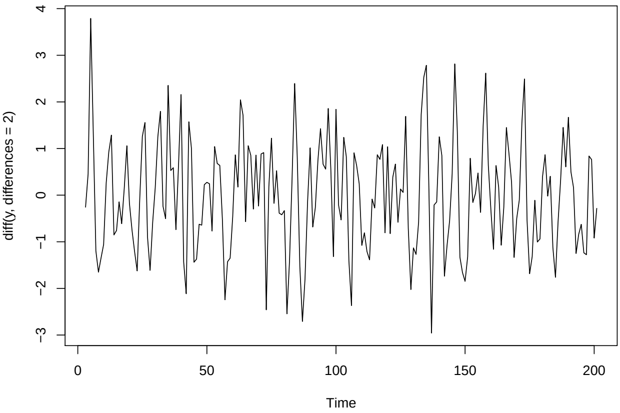
```
acf(diff(y))
```



Taking first order differences appears to have made the series stationary.

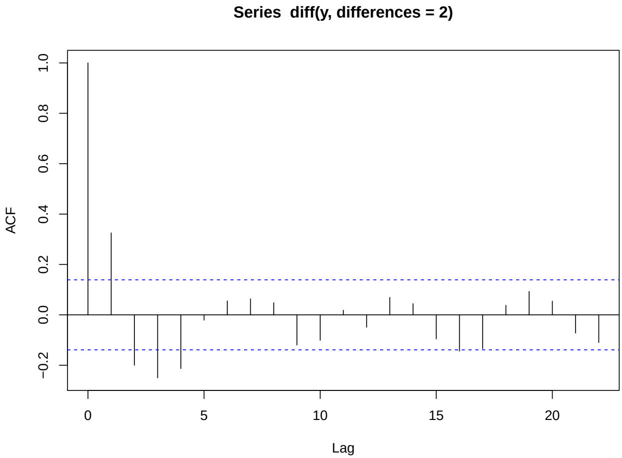
Example (continued)

```
plot(diff(y,differences=2))
```



Example (continued)

```
acf(diff(y,differences=2))
```



Taking second order differences doesn't improve the situation further, so assume $d = 1$.

The partial autocorrelation function

Consider an AR(p) process.

We have seen that the autocorrelations generally do not have a sharp cut-off at lag p but may die away exponentially.

For example, we have seen for the AR(1) process

$$Y_t = \alpha Y_{t-1} + Z_t$$

that

$$\rho_k = \alpha^k \text{ for } k = 0, 1, 2, \ldots$$

The partial autocorrelation function

For a stationary time series, Y_t, the k^{th} order partial correlation, $\alpha(k)$ is defined by

$$\alpha(k) = \mathrm{cor}(Y_k - \hat{Y}_{k|k-1,k-2,\ldots,1}, Y_0 - \hat{Y}_{0|k-1,k-2,\ldots,1})$$

where $\hat{Y}_{t|k-1,k-2,\ldots,1}$ is the linear combination of

$$Y_1, Y_2, \ldots, Y_{k-1}$$

that minimises

$$E((Y_t - \hat{Y})^2).$$

Example AR(1)

Consider the $AR(1)$ process,

$$Y_t = \alpha Y_{t-1} + Z_t.$$

To find $\alpha(1)$, observe that $\hat{Y} = 0$ by default, and

$$\alpha(1) = \alpha = \mathrm{cor}(Y_1, Y_0).$$

To find $\alpha(2)$, first need to find $\hat{Y}_{2|1}$. That is, need to choose β to minimise

$$\begin{aligned} E((Y_2 - \beta Y_1)^2) &= \mathrm{var}((\alpha - \beta) Y_1 + Z_2) \\ &= \sigma^2 \left\{ \frac{(\alpha-\beta)^2}{1-\alpha^2} + 1 \right\} \end{aligned}$$

which is minimised when $\alpha = \beta$.

Example AR(1)

Hence it follows that

$$\begin{aligned}\alpha(2) &= \operatorname{cor}(Y_2 - \alpha Y_1, Y_0 - \alpha Y_1)\\ &= \operatorname{cor}(Z_2, Y_0 - \alpha Y_1)\\ &= 0.\end{aligned}$$

If can be shown similarly that $\alpha(k) = 0$ for all $k > 2$.

Hence, the PACF is

$$\alpha(k) = \begin{cases}\alpha & \text{for } k = 1\\ 0 & \text{otherwise.}\end{cases}$$

PACF for the AR(p) process

Consider the AR(p) process

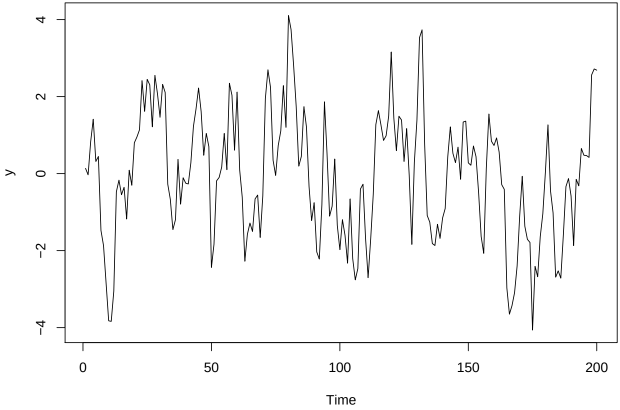
$$Y_t = \sum_{j=1}^{p} \alpha_j Y_{t-j} + Z_t.$$

It can be shown that the PACF satisfies $\alpha(k) = 0$ for $k > p$.

Hence the PACF is useful for the identification of pure AR(p) processes.

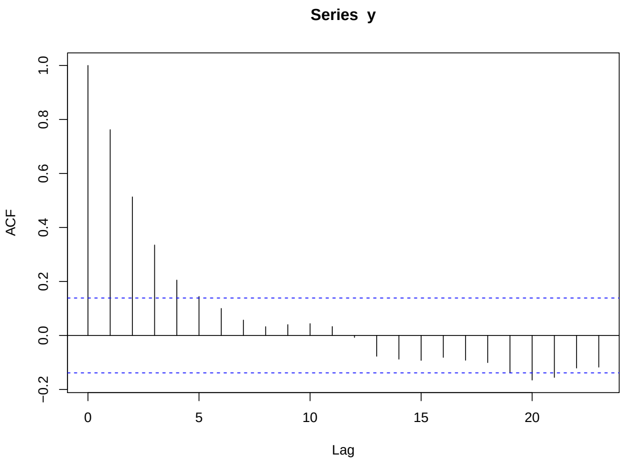
Example AR(2) process

```r
plot(y)
```



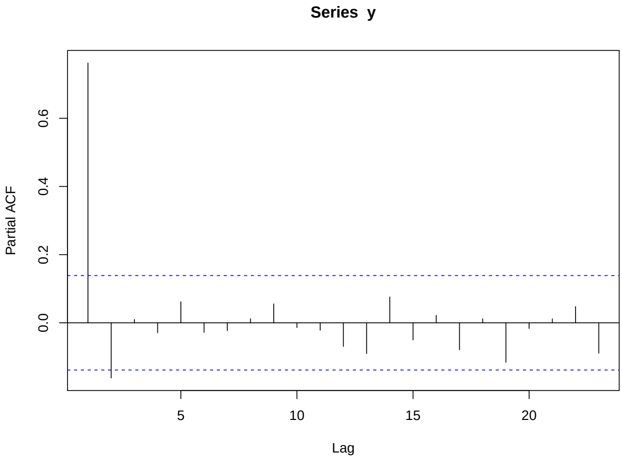
Example AR(2) process

```r
acf(y)
```



Example AR(2) process

```r
acf(y,type="partial")
```



Estimation for ARMA Models

Consider the ARMA model

$$\phi(B)(Y_t - \mu) = \theta(B)Z_t$$

where $Z_t \sim N(0, \sigma^2)$ independently for $t = 1, 2, \ldots, n$.

The method of maximum likelihood is used to estimate the unknown parameters.

The most direct specification is to let

$$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^T$$

and define $\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$ to be the $n \times n$ variance matrix

$$\mathrm{Var}(\boldsymbol{Y}) = \Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2).$$

The log likelihood function

The assuming normality, it follows that

$$\boldsymbol{Y} \sim N_n(\mu \mathbf{1}, \Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2))$$

so the log likelihood function is

$$\begin{aligned} & \ell(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2; \boldsymbol{y}) \\ = \; & -\frac{1}{2} \log |\Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)| - \frac{1}{2}(\boldsymbol{y} - \mu \mathbf{1})^T \Sigma(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)^{-1}(\boldsymbol{y} - \mu \mathbf{1}). \end{aligned}$$

The maximum likelihood estimates are then defined to be the values

$$(\hat{\mu}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$$

that maximize ℓ.

The computational methods for obtaining the maximum likelihood estimates are beyond the scope of the course.

Example the AR(1) process

In the special case of the AR(1) process, it is possible to evaluate the likelihood explicitly.

Consider the factorisation of the PDF,

$$f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2|y_1)f(y_3|y_2, y_1)\ldots f(y_n|y_{n-1}\ldots y_1).$$

For the AR(1) process, it follows that the conditional independence relations,

$$Y_t \perp\!\!\!\perp Y_{t-2}, Y_{t-3}, \ldots, Y_1|Y_{t-1}$$

hold, and hence, the factorisation becomes

$$f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2|y_1)f(y_3|y_2)\ldots f(y_n|y_{n-1}).$$

Example - continued

Under the AR(1) model with normal errors,

$$Y_t | Y_{t-1} \sim N(\mu + \alpha(Y_{t-1} - \mu), \sigma^2)$$

and, assuming stationarity,

$$Y_1 \sim N\left(\mu, \frac{\sigma^2}{1-\alpha^2}\right).$$

The log-likelihood is therefore

$$\begin{aligned}\ell(\mu, \alpha, \sigma^2; \boldsymbol{y}) &= -\frac{1}{2}\log\frac{\sigma^2}{1-\alpha^2} - \frac{1}{2}(y_1 - \mu)^2 / \left(\frac{\sigma^2}{1-\alpha^2}\right) \\ &\quad -\frac{n-1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}((y_t - \mu) - \alpha(y_{t-1} - \mu))^2.\end{aligned}$$

Example - continued

The log likelihood cannot be maximized explicitly, but some insight may be obtained by considering

$$\ell(\mu, \alpha, \sigma^2; \boldsymbol{y}|y_1) = -\frac{n-1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}((y_t - \mu) - \alpha(y_{t-1} - \mu))^2.$$

Maximisation with respect to μ and α corresponds to minimising the sum of squares

$$Q(\mu, \alpha) = \sum_{t=2}^{n}((y_t - \mu) - \alpha(y_{t-1} - \mu))^2.$$

Example - continued

By arguments similar to those used in least squares regression, it follows that

$$\hat{\mu} = \frac{1}{n-1}\sum_{t=2}^{n-1} y_t + \frac{1}{(n-1)(1-\hat{\alpha})}(y_n - \hat{\alpha} y_1)$$

and

$$\hat{\alpha} = \frac{\sum_{t=2}^{n}(y_t - \hat{\mu})(y_{t-1} - \hat{\mu})}{\sum_{t=2}^{n}(y_{t-1} - \hat{\mu})^2}.$$

The maximum likelihood estimate for σ^2 can be seen to be

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{t=2}^{n}((y_t - \hat{\mu}) - \hat{\alpha}(y_{t-1} - \hat{\mu}))^2.$$

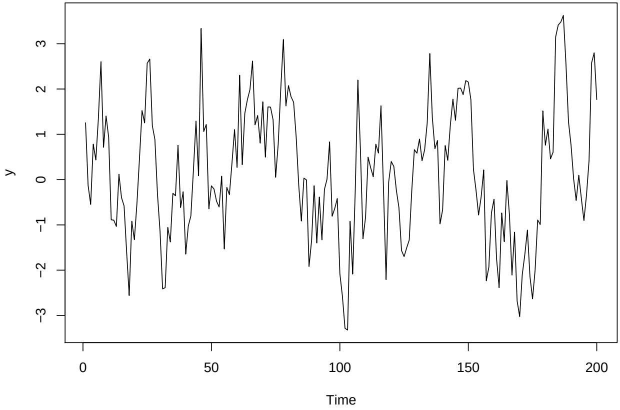
Example - continued

The solutions $\hat{\alpha}$ and $\hat{\mu}$ are given as a pair of coupled non-linear equations. However, for large n the solutions will be approximately

$$\hat{\mu} = \bar{y} \text{ and } \hat{\alpha} = \frac{\sum_{t=2}^{n}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=2}^{n}(y_t - \bar{y})^2}.$$

Example - simulated AR(1) data, $\alpha = 0.8$, $n = 200$

```r
plot(y)
```



Example - continued

```
arima(y,c(1,0,0))

##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.7385     0.1687
## s.e.  0.0474     0.2624
##
## sigma^2 estimated as 0.9674:  log likelihood = -280.87,
```

Example - continued

```r
ybar=mean(y)
a=y[-1]-ybar;b=y[-200]-ybar
alpha=sum(a*b)/sum(b^2)
ybar
```

```
## [1] 0.1307355
```

```r
alpha
```

```
## [1] 0.7399175
```

```r
sigma2.hat=sum((a-alpha*b)^2)/(length(y)-1)
sigma2.hat
```

```
## [1] 0.9694619
```

Identification of ARIMA Models

The problem we are considering is to find an appropriate ARIMA(p, d, q) model for an observed time series y_t.

The first step is to select the differencing parameter d.

- This is usually achieved by trial and error.
- The criterion is to choose the smallest d for which the series $D^d y_t$ appears to be stationary.
- Usually only need to check $d = 0, 1, 2$.
- Judge stationarity from the ACF and the time series plot.

Identification of ARIMA Models

If (p, d, q) are given, it is easy to fit the corresponding ARIMA model by maximum likelihood in software packages such as `R`.

In what what follows, assume that d has been chosen.

The problem is to determine appropriate values of p and q.

Broad terms, we seek the smallest values of p and q that provide an adequate fit to the data.

The Akaike information criterion (AIC)

Consider the ARMA(p, q) model

$$Y_t = \alpha_1 Y_{t-1} + \ldots \alpha_p Y_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}.$$

The AIC statistic is

$$\text{AIC} = -2\ell(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2) + 2(p + q + 1)$$

The approach is to choose p and q that result in the minimum AIC.

Interpretation of (AIC)

Consider the analogy of finding a well fitting regression model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

by choosing the "best" subset from a set of predictor variables.

- Measure the fit of the model by the accuracy of predictions.
- If the model is too simple, predictions will be poor because of bias.
  - That is, $E(\boldsymbol{Y}) \neq X\boldsymbol{\beta}$.
- If the model is correct but contains extra unnecessary terms, the model fit will overfit the data and the predictions will also be poor.

Interpretation of (AIC)

For the linear regression model with p predictors, it can be shown that the maximized log likelihood is

$$\ell(\hat{\beta}, \hat{\sigma}^2; \boldsymbol{y}) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2}$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

The AIC statistic is then

$$\mathrm{AIC} = n \log 2\pi + n \log(\hat{\sigma}^2) + n + 2(p+1).$$

Minimising AIC with respect to p corresponds to maximising the log-likelihood but imposes a penalty for adding extra parameters.

AIC and likelihood ratio tests

Consider a general likelihood $\ell(\theta)$ and the model

$$M : \theta \in \Theta$$

and the hypothesis

$$H_0 : \theta \in \Theta_0$$

where

$$\Theta_0 \subset \Theta \text{ and } \dim(\Theta_0) = p_0 < p = \dim(\Theta).$$

The log likelihood ratio test statistic is

$$G^2 = 2(\ell(\hat{\theta}) - \ell(\hat{\theta}_0)).$$

where $\hat{\theta}$ and $\hat{\theta}_0$ are the maximum likelihood estimates under M and H_0 respectively.

AIC and likelihood ratio tests

If H_0 is true then,

$$G^2 \sim \chi^2_{p-p_0}$$

(approximately) for large samples.

It is also easy to see that

$$G^2 = \mathrm{AIC}(\hat{\theta}_0) - \mathrm{AIC}(\hat{\theta}) + 2(p - p_0).$$

Suppose $p - p_0 = 1$.

- The critical value for the likelihood ratio test is $\chi^2_{1,0.05} = 3.84$.
- $2(p - p_0) = 2$.
- For this reason it is sometimes given as a rule of thumb that models for which AIC is within 2 (≈ 1.84) of the optimal AIC should also be considered.

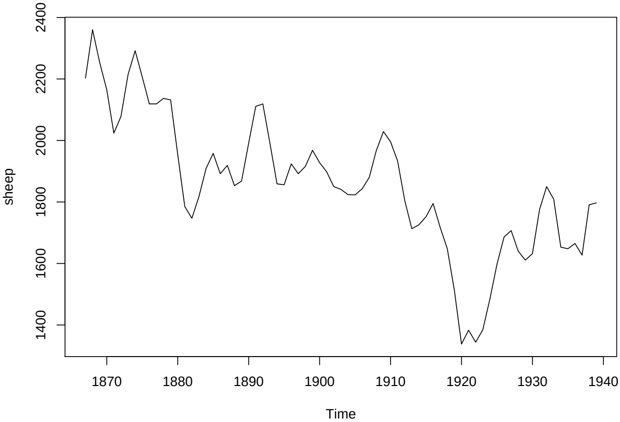
Model Identification Procedure

1. Inspect the original series for stationarity and select a suitable d.
2. Use the ACF and PACF to suggest suitable starting values for p and q.
3. Select suitable candidate models on the basis of AIC.
4. Check the residuals from the model to determine whether they are consistent with white noise.

Example

The data set sheep contains annual measurements of the population of sheep in England and Wales from 1867 to 1939.

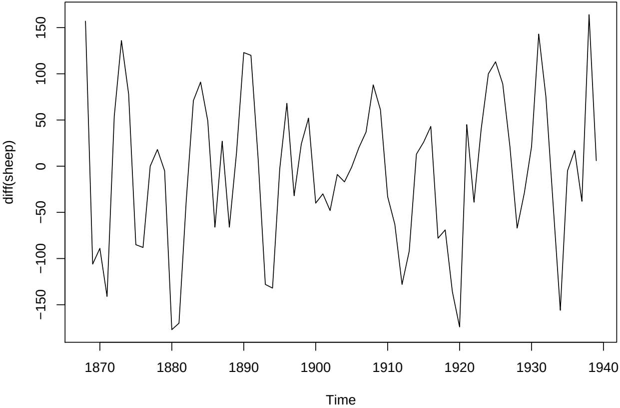
```r
require(fma); data(sheep); plot(sheep)
```



Example

There is some suggestion of a trend so try taking first differences.

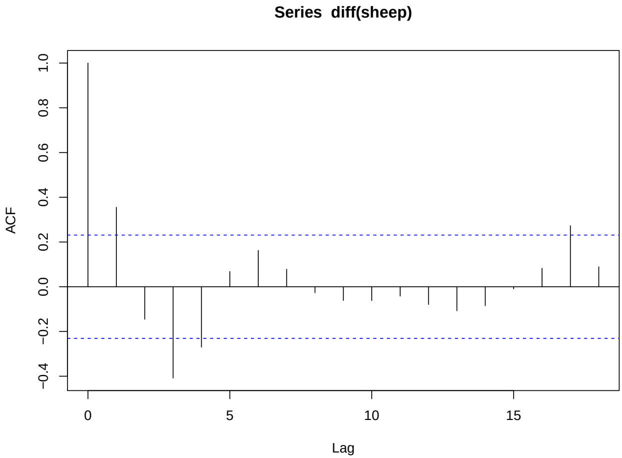
```
plot(diff(sheep))
```



Example

Examine the ACF of the differenced data.

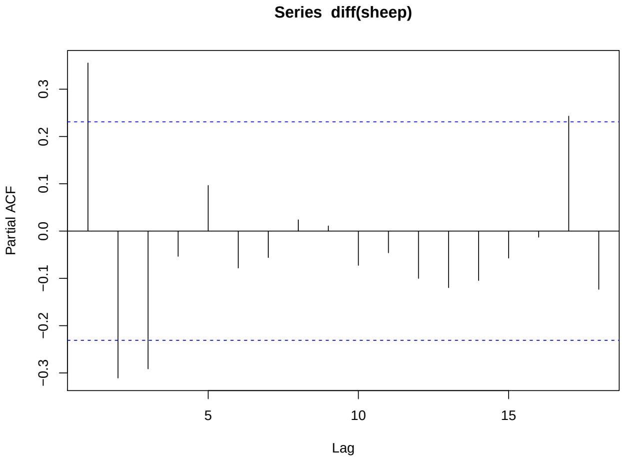
```
acf(diff(sheep))
```



Example

Examine the PACF of the differenced data.

```
pacf(diff(sheep))
```



Example

ACF and PACF suggest we should try ARMA models.

```
arima(sheep,order=c(2,1,2))

##
## Call:
## arima(x = sheep, order = c(2, 1, 2))
##
## Coefficients:
##          ar1      ar2      ma1     ma2
##       0.8739  -0.7648  -0.4341  0.3891
## s.e.  0.1736   0.1272   0.2120  0.2188
##
## sigma^2 estimated as 4721:  log likelihood = -407.14,  aic = 824.27

arima(sheep,order=c(1,1,2))

##
## Call:
## arima(x = sheep, order = c(1, 1, 2))
##
## Coefficients:
##           ar1     ma1     ma2
##       -0.1796  0.6912  0.2452
## s.e.   0.3774  0.3506  0.2202
##
## sigma^2 estimated as 5615:  log likelihood = -413.1,  aic = 834.2
```

Example

```
arima(sheep,order=c(2,1,1))

##
## Call:
## arima(x = sheep, order = c(2, 1, 1))
##
## Coefficients:
##          ar1      ar2      ma1
##       0.9150  -0.5454  -0.4553
## s.e.  0.1725   0.1066   0.1885
##
## sigma^2 estimated as 4896:  log likelihood = -408.35,  aic = 824.7

arima(sheep,order=c(3,1,2))

##
## Call:
## arima(x = sheep, order = c(3, 1, 2))
##
## Coefficients:
##          ar1      ar2      ar3      ma1     ma2
##       0.7070  -0.5960  -0.1139  -0.2853  0.3158
## s.e.  0.4388   0.4338   0.2746   0.4182  0.2844
##
## sigma^2 estimated as 4711:  log likelihood = -407.06,  aic = 826.12
```

Example

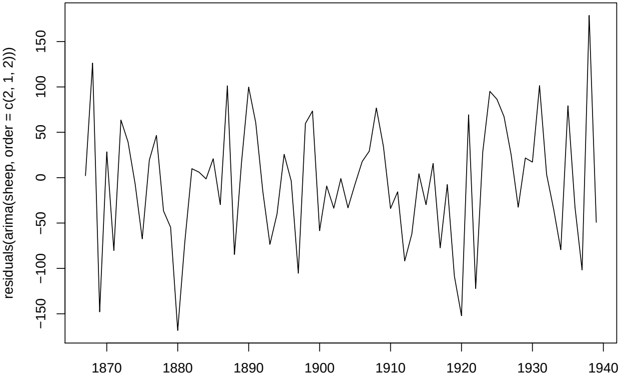
```
arima(sheep,order=c(2,1,3))

##
## Call:
## arima(x = sheep, order = c(2, 1, 3))
##
## Coefficients:
##          ar1      ar2      ma1     ma2      ma3
##       0.8322  -0.6857  -0.4228  0.3689  -0.1109
## s.e.  0.2044   0.2051   0.2152  0.1967   0.2031
##
## sigma^2 estimated as 4702:  log likelihood = -406.99,  aic = 825.98

arima(sheep,order=c(3,1,3))

##
## Call:
## arima(x = sheep, order = c(3, 1, 3))
##
## Coefficients:
##           ar1     ar2      ar3     ma1      ma2     ma3
##       -0.0699  0.0325  -0.7252  0.5575  -0.0153  0.4272
## s.e.   0.1787  0.1909   0.1189  0.2251   0.3260  0.2139
##
## sigma^2 estimated as 4551:  log likelihood = -406.44,  aic = 826.88
```

Example

Based on AIC ARIMA$(2,1,1)$ and ARIMA$(2,1,2)$ appear the best choices.

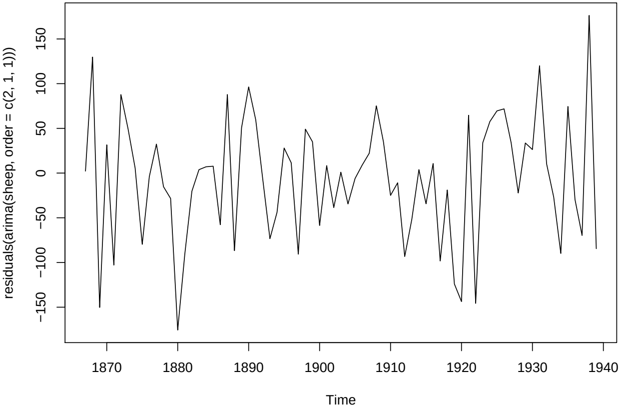
Now check the residuals.

```r
plot(residuals(arima(sheep,order=c(2,1,2))))
```



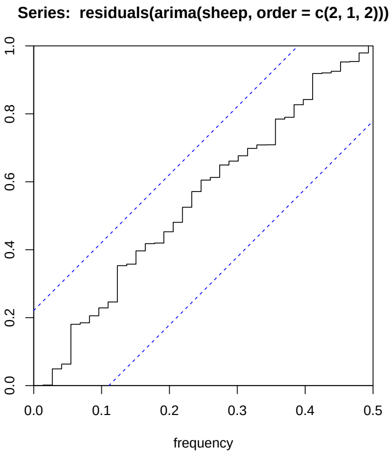
Example

```
plot(residuals(arima(sheep,order=c(2,1,1))))
```



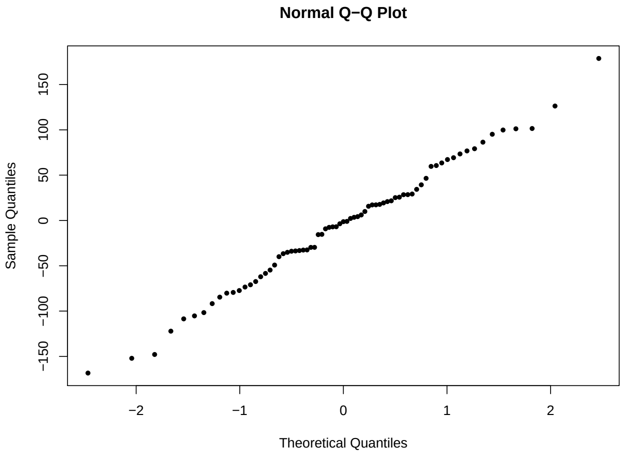
Example

```
cpgram(residuals(arima(sheep,order=c(2,1,2))))
```



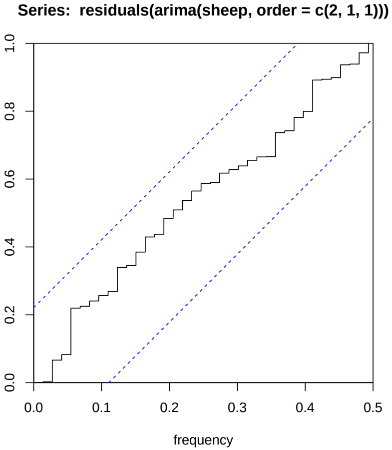
Example

```
qqnorm(residuals(arima(sheep,order=c(2,1,2))),pch=20)
```



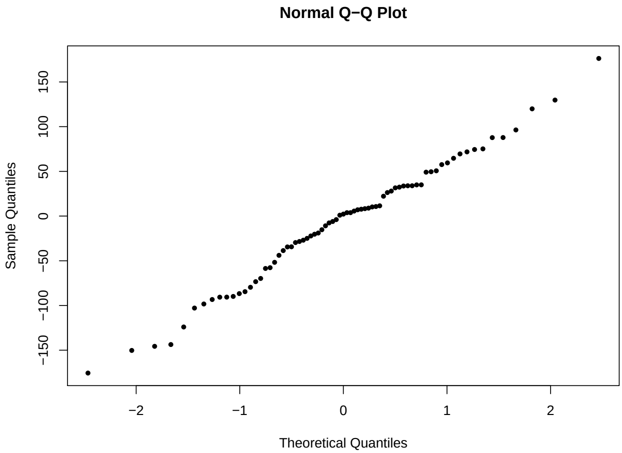
Example

```
cpgram(residuals(arima(sheep,order=c(2,1,1))))
```



Example

```
qqnorm(residuals(arima(sheep,order=c(2,1,1))),pch=20)
```



Example

Both models appear comparable so choose the simpler ARIMA$(2, 1, 1)$ as a reasonable model.

```
arima(sheep,order=c(2,1,1))

##
## Call:
## arima(x = sheep, order = c(2, 1, 1))
##
## Coefficients:
##          ar1      ar2      ma1
##       0.9150  -0.5454  -0.4553
## s.e.  0.1725   0.1066   0.1885
##
## sigma^2 estimated as 4896:  log likelihood = -408.35,  aic =
```

SARIMA Models

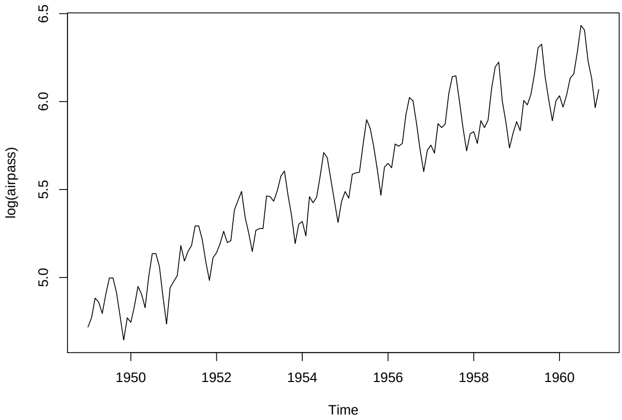
Last time we illustrated the steps for finding a suitable ARIMA model for an observed time series y_t.

However, data with a seasonal component are not well suited to the this treatment.

Consider the Air Passenger Data.

Example - The Air Passenger Data

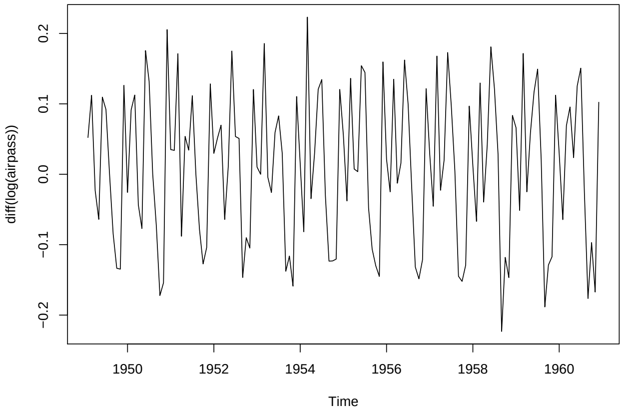
```r
require(fma); data(airpass); plot(log(airpass))
```



Example

Taking differences does not eliminate the periodicity in the data.

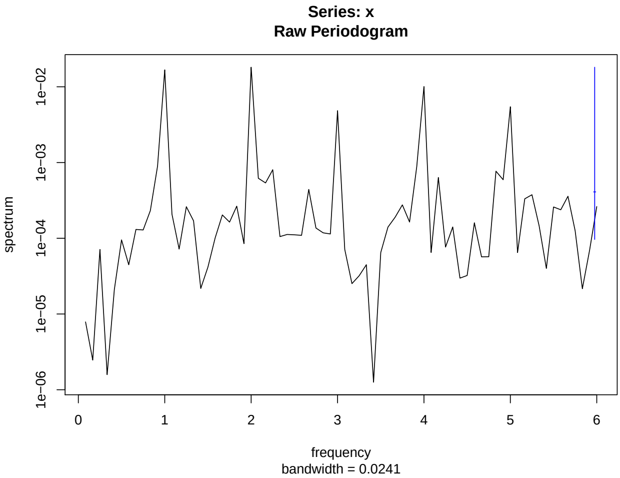
```
plot(diff(log(airpass)))
```



Example

Can see this in the spectrum of the differenced data. Also taking second differences does not help.

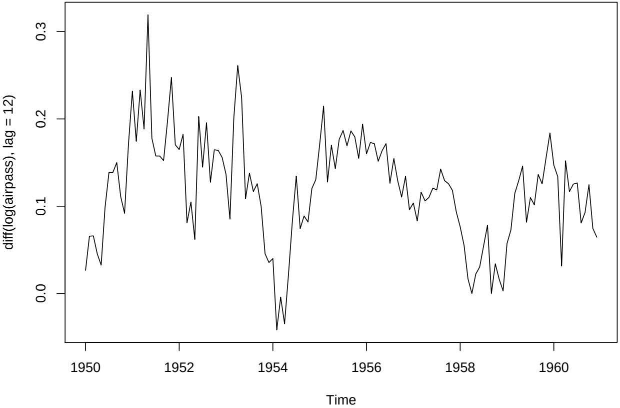
```
spectrum(diff(log(airpass)))
```



Example

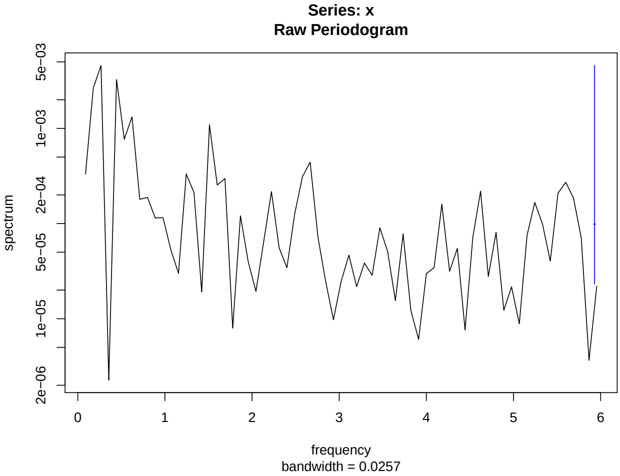
On the other hand, taking differences at lag 12 eliminates the seasonal effects. But the series may still be non-stationary or have very complex structure.

```
plot(diff(log(airpass),lag=12))
```



Example

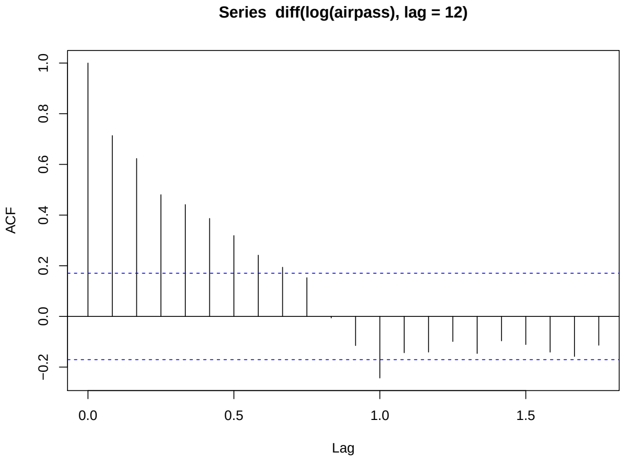
```
spectrum(diff(log(airpass),lag=12))
```



Example

Examine the ACF of the differenced data.

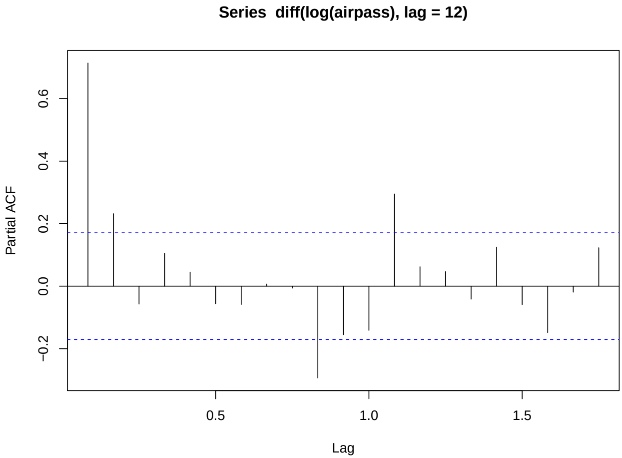
```
acf(diff(log(airpass),lag=12))
```



Example

Examine the PACF of the differenced data.

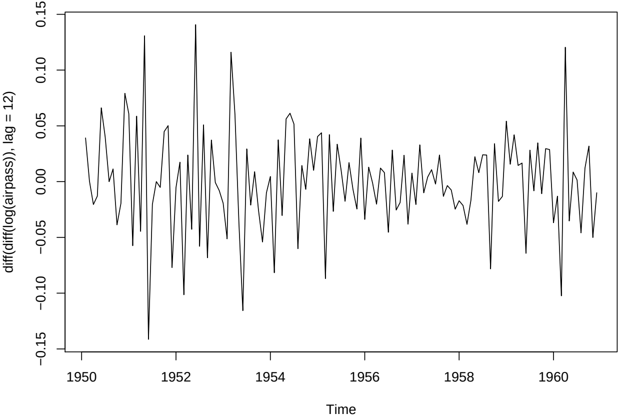
```
pacf(diff(log(airpass),lag=12))
```



Example

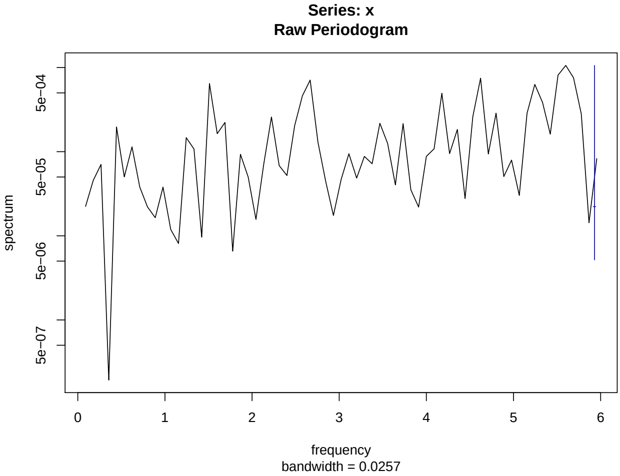
Applying differencing at both lag 1 and lag 12 simplifies things somewhat.

```r
plot(diff(diff(log(airpass)),lag=12))
```



Example

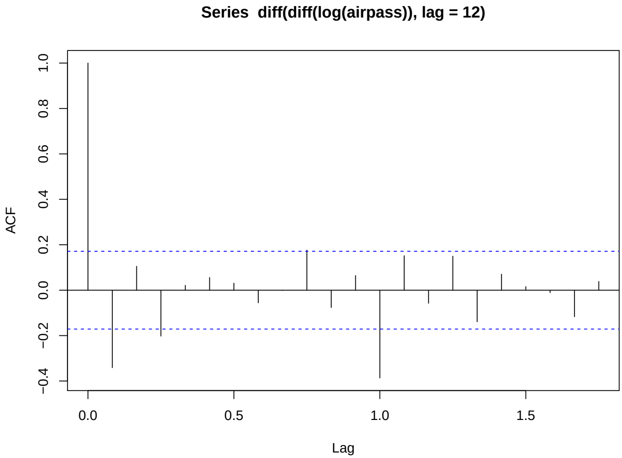
```
spectrum(diff(diff(log(airpass)),lag=12))
```



Example

Examine the ACF of the differenced data.

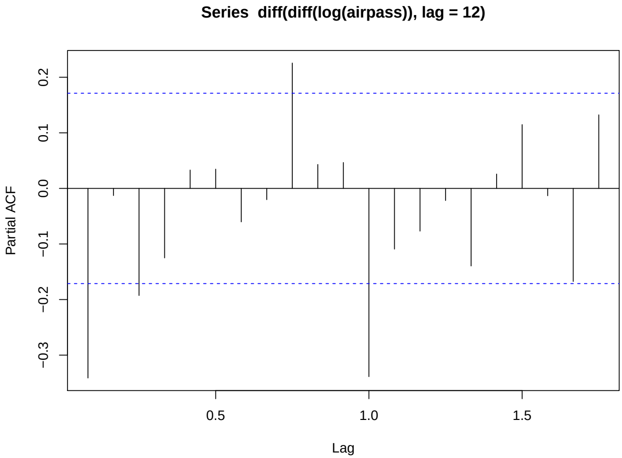
```
acf(diff(diff(log(airpass)),lag=12))
```



Example

Examine the PACF of the differenced data.

```
pacf(diff(diff(log(airpass)),lag=12))
```



SARIMA Models

The Air Passengers data shows that after applying differencing twice, once at lag 12 and once at lag 1, there are significant autocorrelations at short time scales (months) and also at longer times scales (years).

This motivates the Seasonal ARIMA model

$$\phi_s(B^k)\phi(B)(1-B^k)^D(1-B)^d Y_t = \theta_s(B^k)\theta(B)Z_t$$

where

- k is the number of observations per year;
- D is the order of annual differencing;
- d is order of differencing;
- ϕ_s, ϕ, θ_s, θ are polynomials of degree P, p, Q, q respectively, all having roots outside the unit circle.

SARIMA Models

Mathematically, SARIMA models are identical in structure to ARIMA models. In particular both models are of the form

$$\Phi(B)\nabla(B)Y_t = \Theta(B)Z_t$$

where Φ and Θ are both polynomials with roots outside the unit circle and ∇ is a fixed polynomial with unit roots.

SARIMA models are also implemented in `R` and the procedure for model identification is analogous to that for ARIMA models.

Example

```
arima(log(airpass),order=c(1,1,1),seasonal=list(order=c(1,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 1), seasonal = list(order = c(1, 1,
##     1)))
##
## Coefficients:
##          ar1      ma1    sar1     sma1
##       0.1666  -0.5615  -0.099  -0.4973
## s.e.  0.2459   0.2115   0.154   0.1360
##
## sigma^2 estimated as 0.001336:  log likelihood = 245.16,  aic = -480.31

arima(log(airpass),order=c(1,1,1),seasonal=list(order=c(1,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 1), seasonal = list(order = c(1, 1,
##     0)))
##
## Coefficients:
##          ar1      ma1     sar1
##       0.0547  -0.4886  -0.4731
## s.e.  0.2161   0.1933   0.0800
##
## sigma^2 estimated as 0.001425:  log likelihood = 241.73,  aic = -475.47
```

Example

```
arima(log(airpass),order=c(1,1,1),seasonal=list(order=c(0,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 1), seasonal = list(order = c(0, 1,
##     1)))
##
## Coefficients:
##          ar1      ma1     sma1
##       0.1960  -0.5784  -0.5643
## s.e.  0.2475   0.2132   0.0747
##
## sigma^2 estimated as 0.001341:  log likelihood = 244.95,  aic = -481.9

arima(log(airpass),order=c(1,1,1),seasonal=list(order=c(0,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 1), seasonal = list(order = c(0, 1,
##     0)))
##
## Coefficients:
##          ar1      ma1
##       0.1449  -0.5190
## s.e.  0.2455   0.2179
##
## sigma^2 estimated as 0.001824:  log likelihood = 227.13,  aic = -448.25
```

Example

```
arima(log(airpass),order=c(1,1,0),seasonal=list(order=c(1,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 0), seasonal = list(order = c(1, 1,
##     1)))
##
## Coefficients:
##           ar1     sar1     sma1
##       -0.3451  -0.0760  -0.5108
## s.e.   0.0828   0.1548   0.1347
##
## sigma^2 estimated as 0.001364:  log likelihood = 243.86,  aic = -479.73

arima(log(airpass),order=c(1,1,0),seasonal=list(order=c(1,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 0), seasonal = list(order = c(1, 1,
##     0)))
##
## Coefficients:
##           ar1     sar1
##       -0.3745  -0.4637
## s.e.   0.0808   0.0808
##
## sigma^2 estimated as 0.001457:  log likelihood = 240.41,  aic = -474.82
```

Example

```
arima(log(airpass),order=c(1,1,0),seasonal=list(order=c(0,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 0), seasonal = list(order = c(0, 1,
##     1)))
##
## Coefficients:
##           ar1     sma1
##       -0.3395  -0.5619
## s.e.   0.0822   0.0748
##
## sigma^2 estimated as 0.001367:  log likelihood = 243.74,  aic = -481.49

arima(log(airpass),order=c(1,1,0),seasonal=list(order=c(0,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(1, 1, 0), seasonal = list(order = c(0, 1,
##     0)))
##
## Coefficients:
##           ar1
##       -0.3405
## s.e.   0.0820
##
## sigma^2 estimated as 0.001842:  log likelihood = 226.51,  aic = -449.01
```

Example

```
arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(1,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(1, 1,
##     1)))
##
## Coefficients:
##           ma1     sar1     sma1
##       -0.4143  -0.1116  -0.4817
## s.e.   0.0899   0.1547   0.1363
##
## sigma^2 estimated as 0.001341:  log likelihood = 244.96,  aic = -481.91

arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(1,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(1, 1,
##     0)))
##
## Coefficients:
##           ma1     sar1
##       -0.4423  -0.4743
## s.e.   0.0832   0.0798
##
## sigma^2 estimated as 0.001426:  log likelihood = 241.7,  aic = -477.41
```

Example

```
arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(0,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(0, 1,
##     1)))
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -483.4

arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(0,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(0, 1,
##     0)))
##
## Coefficients:
##           ma1
##       -0.3870
## s.e.   0.0887
##
## sigma^2 estimated as 0.001828:  log likelihood = 226.99,  aic = -449.98
```

Example

```
arima(log(airpass),order=c(0,1,0),seasonal=list(order=c(1,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 0), seasonal = list(order = c(1, 1,
##     1)))
##
## Coefficients:
##         sar1     sma1
##       0.0109  -0.6090
## s.e.  0.1447   0.1199
##
## sigma^2 estimated as 0.001536:  log likelihood = 235.78,  aic = -465.56

arima(log(airpass),order=c(0,1,0),seasonal=list(order=c(1,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 0), seasonal = list(order = c(1, 1,
##     0)))
##
## Coefficients:
##          sar1
##       -0.4320
## s.e.   0.0817
##
## sigma^2 estimated as 0.001702:  log likelihood = 230.51,  aic = -457.02
```

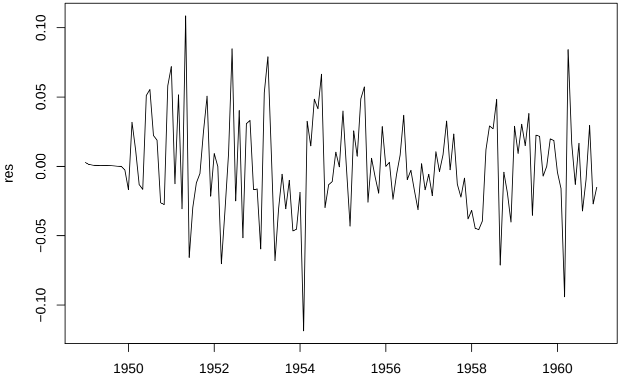
Example

```
arima(log(airpass),order=c(0,1,0),seasonal=list(order=c(0,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 0), seasonal = list(order = c(0, 1,
##     1)))
##
## Coefficients:
##          sma1
##       -0.6021
## s.e.   0.0784
##
## sigma^2 estimated as 0.001536:  log likelihood = 235.78,  aic = -467.56

arima(log(airpass),order=c(0,1,0),seasonal=list(order=c(0,1,0)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 0), seasonal = list(order = c(0, 1,
##     0)))
##
##
## sigma^2 estimated as 0.002086:  log likelihood = 218.41,  aic = -434.83
```

Example

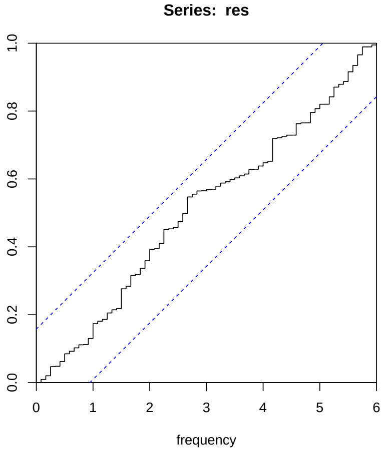
Based on AIC, the SARIMA model with $(p = 0, d = 1, q = 1)$ and $(P = 0, D = 1, Q = 1)$ appears the most promising. Now check the residuals.

```
res=residuals(arima(log(airpass),order=c(0,1,1),seasonal=li
plot(res)
```



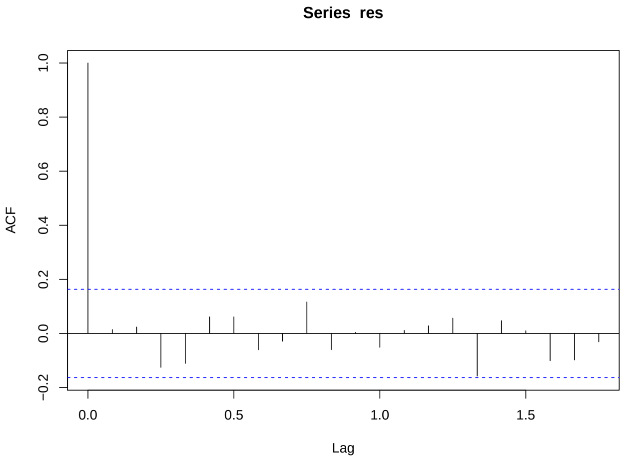
Example

```
cpgram(res)
```



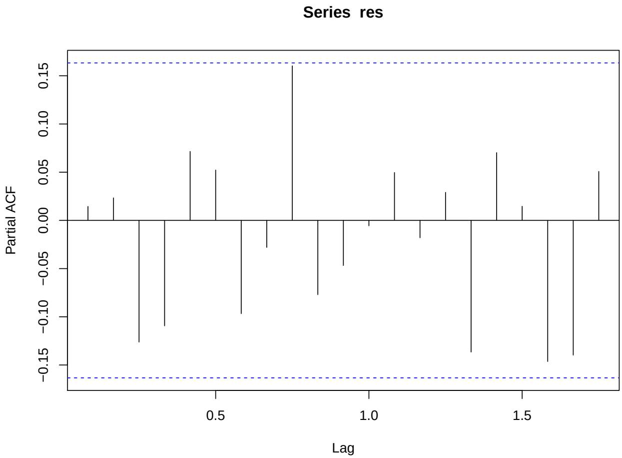
Example

```r
acf(res)
```



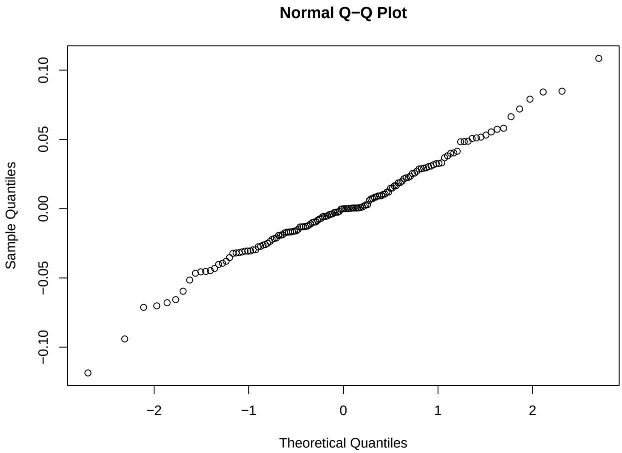
Example

```r
pacf(res)
```



Example

```r
qqnorm(res)
```



Example

```
arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(0,1,1)))
```

```
##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(o
##     1)))
##
## Coefficients:
##           ma1     sma1
##       -0.4018  -0.5569
## s.e.   0.0896   0.0731
##
## sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic
```

Forecasting

A key practical problem in time series analysis is to use an observed series to predict future values.

As such, forecasting time series is a form of extrapolation and carries the same risks inherent in extrapolating in linear regression.

In particular, a strong assumption that the processes driving the series do not change is needed.

For this reason we consider only short term forecasts.

ARMA Processes

Consider the ARMA(p, q) process,

$$\phi(B)Y_t = \theta(B)Z_t$$

where:

- $\phi(B)$ and $\theta(B)$ are polynomials of degrees p and q respectively, with no common factors and all roots outside the unit circle;
- $Z_t \sim N(0, \sigma^2)$ independently is a Gaussian white noise process.

Recall that Y_t can be expressed as a general linear process

$$Y_t = \phi(B)^{-1}\theta(B)Z_t = \sum_{j=0}^{\infty} \theta_j Z_{t-j}.$$

ARMA Processes

Consider now an observed series,

$$Y_1, Y_2, \ldots, Y_n$$

and suppose we wish to forecast the future observation Y_{n+k} for $k = 1, 2, 3, \ldots$.

Let $Y_n(k)$ denote the prediction based on $Y_1, Y_2, \ldots, Y_n$.

We seek a prediction of the form

$$Y_n(k) = \sum_{i=0}^{n-1} w_i Y_{n-i}.$$

ARMA Processes MSE

Substituting for Y_t we obtain

$$Y_n(k) = \sum_{i=0}^{n-1} w_i \left(\sum_{j=0}^{\infty} \theta_j Z_{n-i-j} \right) = \sum_{j=0}^{\infty} W_j Z_{n-j}.$$

The mean squared error is thus

$$\begin{aligned} & E((Y_{n+k} - Y_n(k))^2) \\ = \quad & E\left(\left(\sum_{j=0}^{\infty} \theta_j Z_{n+k-j} - \sum_{j=0}^{\infty} W_j Z_{n-j} \right)^2 \right) \\ = \quad & E\left(\left(\sum_{j=0}^{k-1} \theta_j Z_{n+k-j} - \sum_{j=k}^{\infty} (\theta_j - W_{j-k}) Z_{n+k-j} \right)^2 \right) \end{aligned}$$

ARMA Processes MSE

Since Z_t is a white noise process, it follows that

$$E((Y_{n+k} - Y_n(k))^2) = \sigma^2 \left(\sum_{j=0}^{k-1} \theta_j^2 + \sum_{j=k}^{\infty} (\theta_j - W_{j-k})^2 \right).$$

The minimum mean squared error prediction is then to choose $W_j = \theta_{j+k}$ so that

$$\begin{aligned} Y_n(k) &= \sum_{j=0}^{\infty} W_j Z_{n-j} \\ &= \sum_{j=0}^{\infty} \theta_{j+k} Z_{n-j} \\ &= \sum_{i=k}^{\infty} \theta_i Z_{n+k-i} \end{aligned}$$

.

ARMA Processes

On the other hand

$$Y_{n+k} = \sum_{j=0}^{\infty} \theta_j Z_{n+k-j}$$

so the optimal forecast corresponds to replacing all future values,

$$Z_{n+1}, Z_{n+2}, \ldots, Z_{n+k}$$

by $0 = E(Z_j)$.

It also follows that the minimum mean squared error is

$$E((Y_{n+k} - Y_n(k))^2) = \sigma^2 \sum_{j=0}^{k-1} \theta_j^2.$$

Example AR(1)

Consider the AR(1) process,

$$Y_t = \alpha Y_{t-1} + Z_t$$

with $|\alpha| < 1$.

Intuitively, the best prediction for Y_{n+1} should be

$$Y_n(1) = \alpha Y_n.$$

We can verify this directly by observing that the AR(1) process may be expressed as

$$Y_t = \sum_{j=0}^{\infty} \alpha^j Z_{t-j}.$$

Example AR(1)

Hence, the optimal one-step prediction is

$$Y_n(1) = \sum_{i=1}^{\infty} \alpha^i Z_{n+1-i} = \sum_{j=0}^{\infty} \alpha^{j+1} Z_{n-j} = \alpha \sum_{j=0}^{\infty} \alpha^j Z_{n-j} = \alpha Y_n.$$

The prediction MSE is σ^2.

More generally, the k-step prediction is

$$Y_n(k) = \alpha^k Y_n.$$

The prediction MSE can can be seen to be

$$\sigma^2 \frac{1-\alpha^{2k}}{1-\alpha^2}.$$

Example MA(1) process

Consider the MA(1) process,

$$Y_t = Z_t + \beta Z_{t-1}$$

with $|\beta| < 1$.

The minimum mean squared error prediction, in terms of Z is

$$Y_n(k) = \begin{cases} \beta Z_n & \text{if } k = 1 \\ 0 & \text{otherwise} \end{cases}$$

If the model includes a non-zero mean, $Y_t = \mu + Z_t + \beta Z_{t-1}$, the prediction is

$$Y_n(k) = \begin{cases} \mu + \beta Z_n & \text{if } k = 1 \\ \mu & \text{otherwise} \end{cases}$$

Example MA(1) process

To express the forecast in terms of the Y_t, observe that the model can be written

$$Y_t = (1 + \beta B)Z_t$$

so that

$$\begin{aligned} Z_t &= (1 + \beta B)^{-1} Y_t \\ &= (1 - \beta B + \beta^2 B^2 - \dots) Y_t \\ &= \sum_{j=0}^{\infty} (-1)^j \beta^j Y_{t-j}. \end{aligned}$$

The forecast, $Y_n(1)$ is therefore

$$Y_n(1) = \sum_{j=0}^{\infty} (-1)^j \beta^{j+1} Y_{n-j}.$$

Example MA(1) process

In practice, Y_t, is only oberved for $t = 1, 2, \ldots, n$ and and $Y_n(1)$ can be approximated by

$$Y_n(1) = \sum_{j=0}^{n-1} (-1)^j \beta^{j+1} Y_{n-j}.$$

The minimum mean squared error prediction can be calculated exactly if it is assumed further that the errors are Gaussian.

Example MA(1) process

In this case, it follows that

$$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{n+1})^T$$

has a $n+1$ dimensional multivariate normal distribution with mean vector and variance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \Sigma = \sigma^2 \begin{pmatrix} 1+\beta^2 & \beta & 0 & 0 & \ldots & 0 & 0 \\ \beta & 1+\beta^2 & \beta & 0 & \ldots & 0 & 0 \\ 0 & \beta & 1+\beta^2 & \beta & \ldots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\beta^2 & \beta \\ 0 & 0 & 0 & 0 & \ldots & \beta & 1+\beta^2 \end{pmatrix}$$

The multivariate normal distribution

In general, consider the $p_1 + p_2$ dimensional normal distribution partitioned as

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim N_{p_1+p_2}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

where $\boldsymbol{Y}_1, \ \boldsymbol{\mu}_1$ are both of dimension $p_1 \times 1, \ \boldsymbol{Y}_2, \ \boldsymbol{\mu}_2$ are both of dimension $p_2 \times 1$, Σ_{11} is $p_1 \times p_1$, Σ_{12} is $p_1 \times p_2$, Σ_{21} is $p_2 \times p_1$ and Σ_{22} is $p_2 \times p_2$.

The conditional distribution of $\boldsymbol{Y}_2 | \boldsymbol{Y}_1$ is

$$\boldsymbol{Y}_2 | \boldsymbol{Y}_1 \sim N_{p_2}(\boldsymbol{\mu_2} + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

Minimum MSE prediction

Consider the problem of finding a function $f(Y_1, Y_2, \ldots, Y_n)$ that minimises the mean squared error

$$E((Y_{n+1} - f(Y_1, Y_2, \ldots, Y_n))^2).$$

It can be shown that the solution is to choose

$$f(Y_1, Y_2, \ldots, Y_n) = E(Y_{n+1} | Y_1, Y_2, \ldots, Y_n).$$

In the Gaussian case, we have

$$E(\boldsymbol{Y}_2 | \boldsymbol{Y}_1) = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1)$$

Example MA(1) process

To apply this result to the MA(1) process, we consider

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \\ Y_{n+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_1 \\ Y_{n+1} \end{pmatrix}$$

so that the multivariate normal distribution is partitioned as

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ Y_{n+1} \end{pmatrix} \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mu_{n+1} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{1,n+1} \\ \Sigma_{n+1,1} & \sigma_{n+1,n+1} \end{pmatrix}.$$

Example MA(1) process

To illustrate, consider the MA(1) process with $\beta = 0.5$ and $n = 10$.

```
S11

##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
##  [1,] 1.25 0.50 0.00 0.00 0.00 0.00 0.00 0.00 0.00  0.00
##  [2,] 0.50 1.25 0.50 0.00 0.00 0.00 0.00 0.00 0.00  0.00
##  [3,] 0.00 0.50 1.25 0.50 0.00 0.00 0.00 0.00 0.00  0.00
##  [4,] 0.00 0.00 0.50 1.25 0.50 0.00 0.00 0.00 0.00  0.00
##  [5,] 0.00 0.00 0.00 0.50 1.25 0.50 0.00 0.00 0.00  0.00
##  [6,] 0.00 0.00 0.00 0.00 0.50 1.25 0.50 0.00 0.00  0.00
##  [7,] 0.00 0.00 0.00 0.00 0.00 0.50 1.25 0.50 0.00  0.00
##  [8,] 0.00 0.00 0.00 0.00 0.00 0.00 0.50 1.25 0.50  0.00
##  [9,] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.50 1.25  0.50
## [10,] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.50  1.25

S21

##  [1] 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.5
```

Example MA(1) process

```
S21%*%solve(S11)

##             [,1]      [,2]       [,3]     [,4]     [,5]
## [1,] -0.00073242 0.0018311 -0.0038452 0.007782 -0.01561
##          [,6]      [,7]  [,8]  [,9] [,10]
## [1,] 0.031242 -0.062496 0.125 -0.25   0.5
```

Note that this is not *identical* to the approximate formula

$$Y_{10}(1) = \sum_{j=0}^{10} (-1)^j 0.5^{j+1} Y_{n-j}.$$

but for practical purposes the difference would be negligible.

Prediction in R

```
y=3+arima.sim(n=100,list(ar=0.8))
fit.y=arima(y,c(1,0,0))
fit.y
```

```
##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##        0.809      3.384
## s.e.  0.060      0.512
##
## sigma^2 estimated as 1.03:  log likelihood = -143.97,  aic =
```

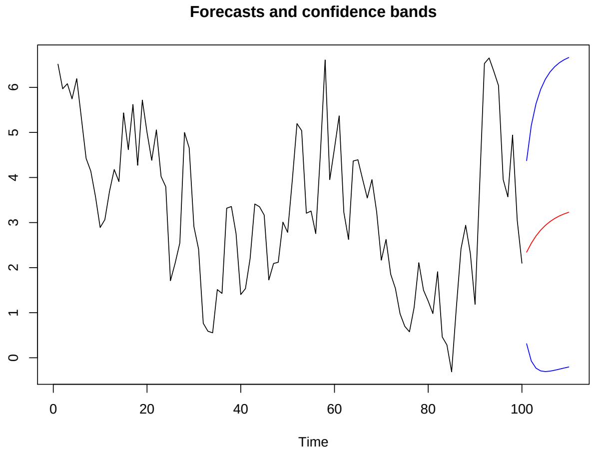
Example AR(1) process

```
pred.y=predict(fit.y,n.ahead=10)
pred.y
```

```
## $pred
## Time Series:
## Start = 101
## End = 110
## Frequency = 1
##  [1] 2.3431 2.5416 2.7022 2.8322 2.9375 3.0227 3.0916 3.1474
##  [9] 3.1926 3.2292
##
## $se
## Time Series:
## Start = 101
## End = 110
## Frequency = 1
##  [1] 1.0156 1.3066 1.4662 1.5620 1.6217 1.6596 1.6840 1.6998
##  [9] 1.7100 1.7167
```

Example AR(1) process

```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```



Example AR(1) process

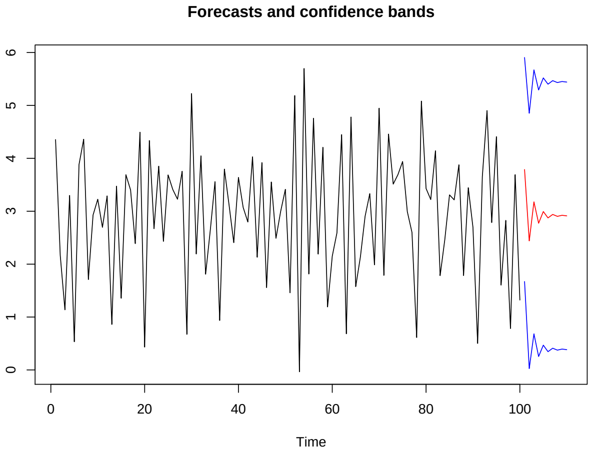
```
y=3+arima.sim(n=100,list(ar=-0.5))
fit.y=arima(y,c(1,0,0))
fit.y
```

```
##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##           ar1  intercept
##        -0.547      2.916
## s.e.    0.084      0.069
##
## sigma^2 estimated as 1.12:  log likelihood = -147.79,  aic =
```

```
pred.y=predict(fit.y,n.ahead=10)
```

Example AR(1) process

```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```

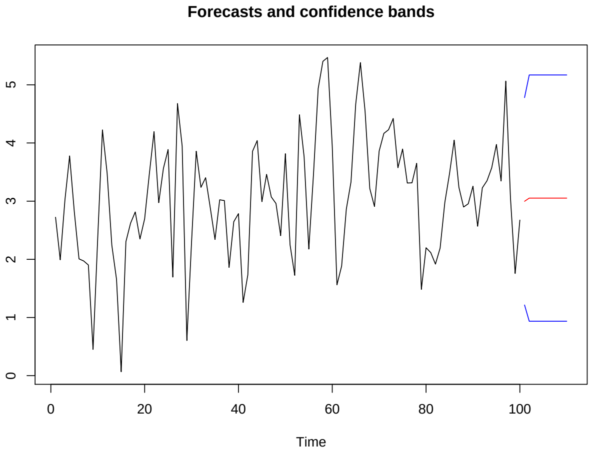


Example MA(1) process

```
y=3+arima.sim(n=100,list(ma=0.8))
fit.y=arima(y,c(0,0,1))
fit.y

##
## Call:
## arima(x = y, order = c(0, 0, 1))
##
## Coefficients:
##          ma1  intercept
##        0.640      3.052
## s.e.  0.079      0.146
##
## sigma^2 estimated as 0.794:  log likelihood = -130.64,  aic =

pred.y=predict(fit.y,n.ahead=10)
```

Example MA(1) process

```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```

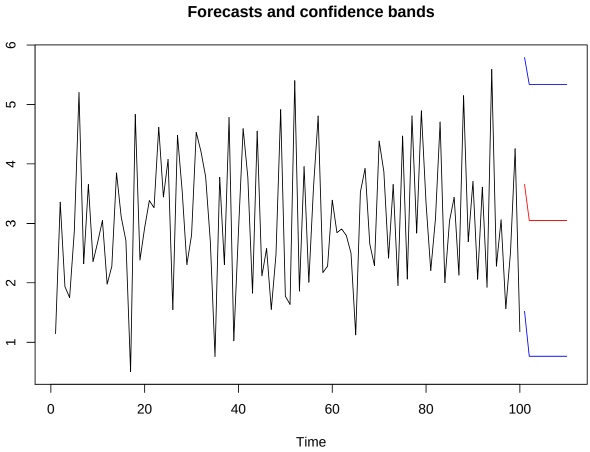


Example MA(1) process

```
y=3+arima.sim(n=100,list(ma=c(-0.5)))
fit.y=arima(y,c(0,0,1))
fit.y

##
## Call:
## arima(x = y, order = c(0, 0, 1))
##
## Coefficients:
##           ma1  intercept
##       -0.383      3.052
## s.e.   0.093      0.066
##
## sigma^2 estimated as 1.14:  log likelihood = -148.52,  aic =

pred.y=predict(fit.y,n.ahead=10)
```

Example MA(1) process

```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```

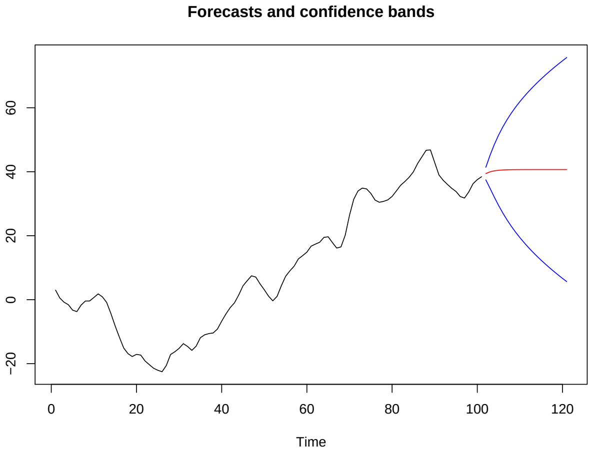


Example ARIMA$(1, 1, 1)$ process

```
y=3+arima.sim(n=100,list(order=c(1,1,1),ar=0.5, ma=0.8))
fit.y=arima(y,c(1,1,1))
fit.y

##
## Call:
## arima(x = y, order = c(1, 1, 1))
##
## Coefficients:
##          ar1    ma1
##        0.557  0.898
## s.e.  0.085  0.043
##
## sigma^2 estimated as 0.959:  log likelihood = -141.23,  aic =

pred.y=predict(fit.y,n.ahead=20)
```

Example ARIMA(1, 1, 1) process

```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```



Example ARIMA$(1, 2, 1)$ process

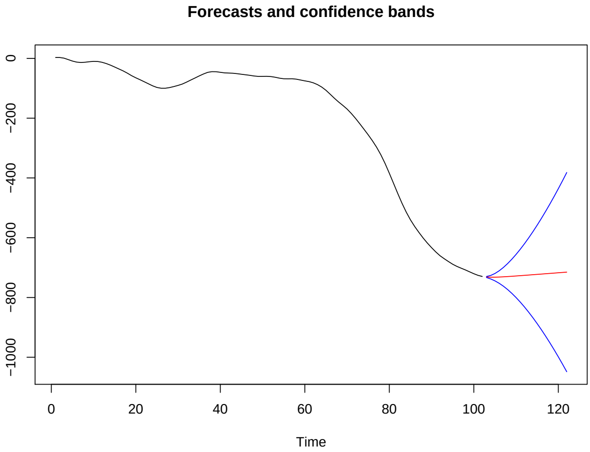
```
y=3+arima.sim(n=100,list(order=c(1,2,1),ar=0.5, ma=0.8))
fit.y=arima(y,c(1,2,1))
fit.y
```

```
##
## Call:
## arima(x = y, order = c(1, 2, 1))
##
## Coefficients:
##          ar1    ma1
##        0.597  0.708
## s.e.  0.089  0.087
##
## sigma^2 estimated as 0.718:  log likelihood = -126.27,  aic =
```

```
pred.y=predict(fit.y,n.ahead=20)
```

Example ARIMA$(1, 2, 1)$ process

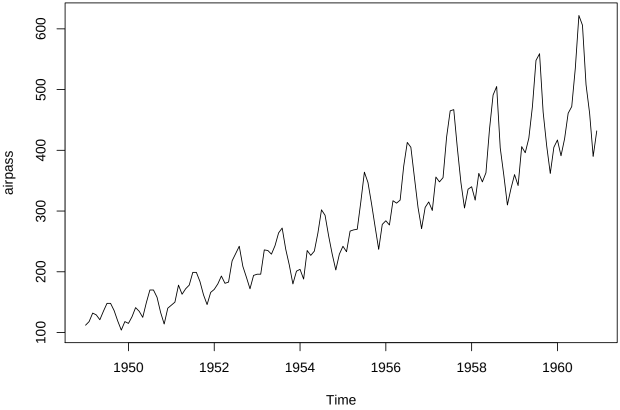
```
ts.plot(y, pred.y$pred, pred.y$pred + 2*pred.y$se,
pred.y$pred - 2*pred.y$se, col=c("black", "red", "blue", "blue"))
title("Forecasts and confidence bands")
```



Forecasting with SARIMA models

Consider the Air Passengers data.

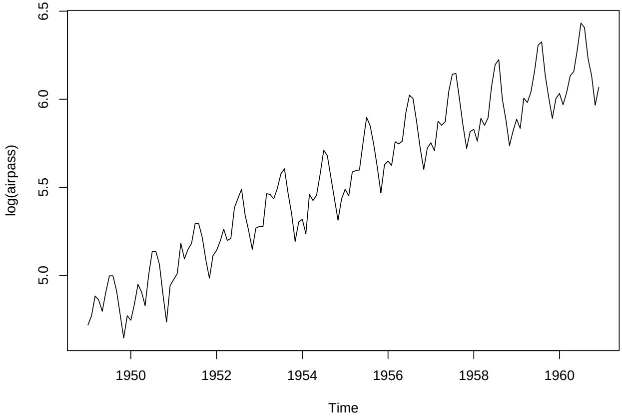
```
plot(airpass)
```



Non-constant variance

Take logs to stabilise increasing variance.

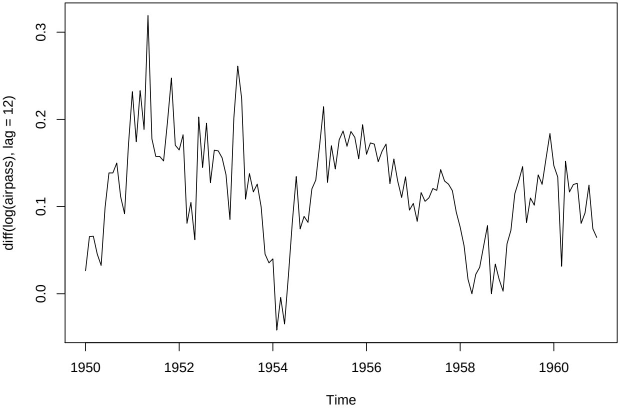
```
plot(log(airpass))
```



Differencing 1

Take lag 12 differences to remove seasonality

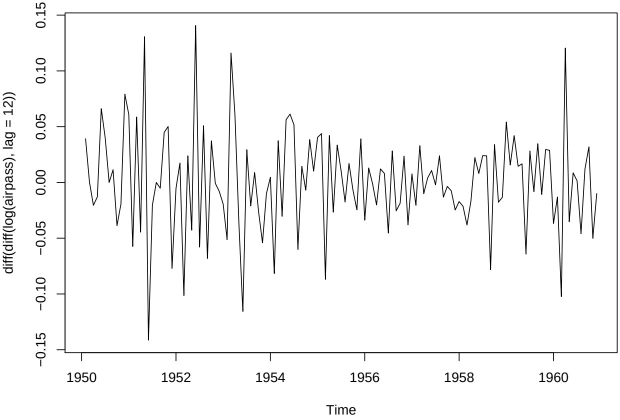
```
plot(diff(log(airpass),lag=12))
```



Differencing 2

Take lag 1 differences to remove non-stationarity

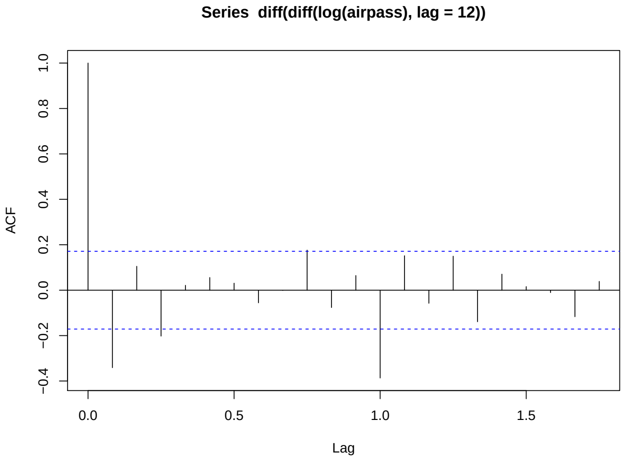
```
plot(diff(diff(log(airpass),lag=12)))
```



Model identification 1

Examine ACF

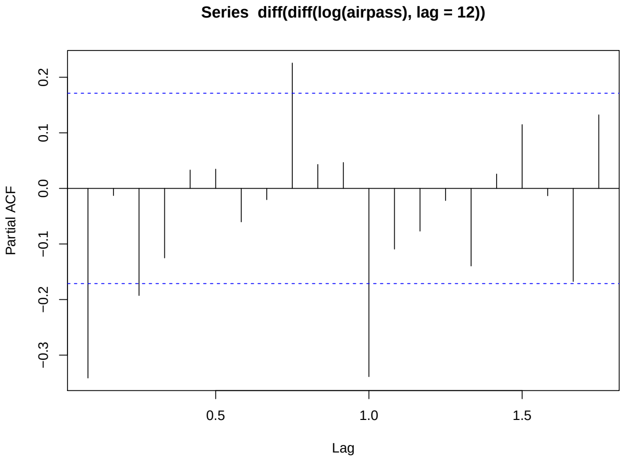
```
acf(diff(diff(log(airpass),lag=12)))
```



Model identification 2

Examine PACF also

```
pacf(diff(diff(log(airpass),lag=12)))
```



Model identification 3

Based on ACF, try SARIMA$(0,1,1) \times (0,1,1)$ as a starting model.

```
arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(0,1,1)))

##
## Call:
## arima(x = log(airpass), order = c(0, 1, 1), seasonal = list(order = c(0, 1,
##     1)))
##
## Coefficients:
##           ma1     sma1
##       -0.402  -0.557
## s.e.   0.090   0.073
##
## sigma^2 estimated as 0.00135:  log likelihood = 244.7,  aic = -483.4
```

Model identification 3

Try a range of models and choose the best based on simplicity and AIC.

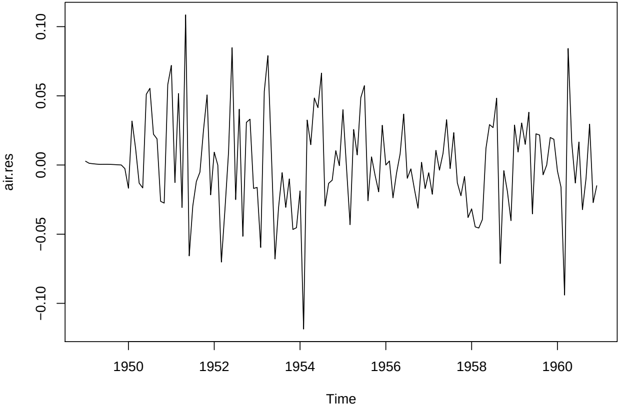
```
Models=NULL
for(p in 0:1) for(q in 0:2) for (P in 0:1) for (Q in 0:1)
Models=rbind(Models,c(p,q,P,Q,
arima(log(airpass),order=c(p,1,q),
seasonal=list(order=c(P,1,Q)))$aic))}
Models=Models[order(Models[,5]),]
head(Models)
```

```
##      [,1] [,2] [,3] [,4]    [,5]
## [1,]    0    1    0    1 -483.40
## [2,]    1    2    0    1 -482.04
## [3,]    0    1    1    1 -481.91
## [4,]    1    1    0    1 -481.90
## [5,]    0    2    0    1 -481.62
## [6,]    1    0    0    1 -481.49
```

Diagnostics 1

Take SARIMA$(0,1,1) \times (0,1,1)$ as best model and examine residuals.

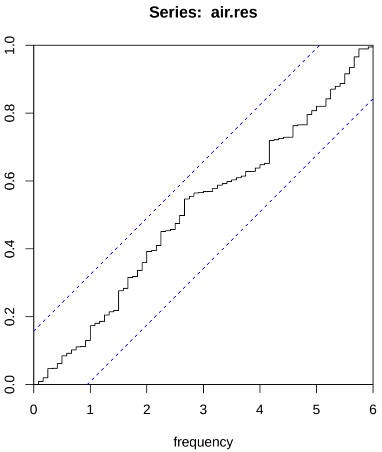
```
air.fit=arima(log(airpass),order=c(0,1,1),seasonal=list(order=c(0,1,1)))
air.res=residuals(air.fit)
plot(air.res)
```



Diagnostics 2

Use cumulative periodogram to check independence.

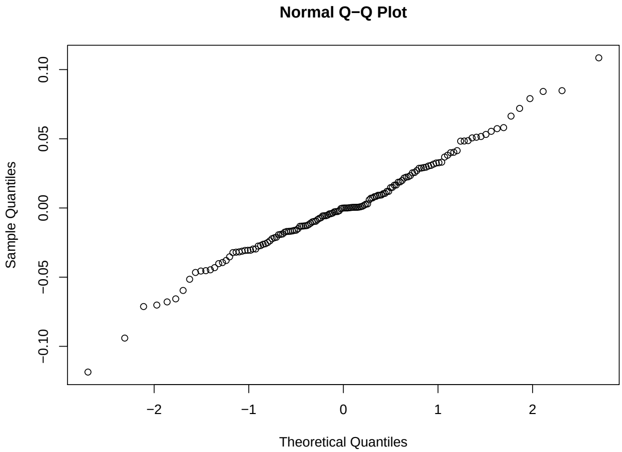
```r
cpgram(air.res)
```



Diagnostics 3

Use normal quantile plot to check normality.

```
qqnorm(air.res)
```



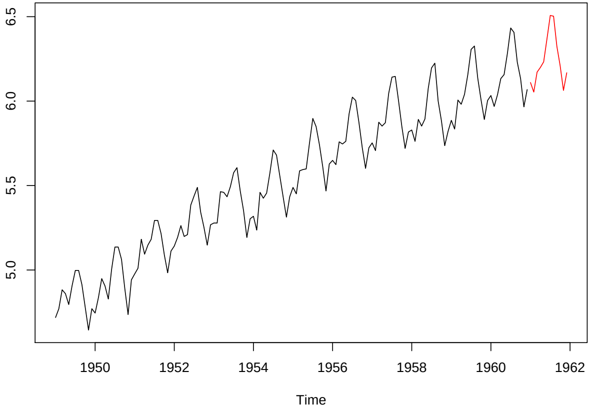
Prediction 1

Can obtain predictions using the generic `predict` function.

```
air.pred=predict(air.fit,n.ahead=12)
air.pred

## $pred
##         Jan    Feb    Mar    Apr    May    Jun    Jul
## 1961 6.1102 6.0538 6.1717 6.1993 6.2326 6.3688 6.5073
##         Aug    Sep    Oct    Nov    Dec
## 1961 6.5029 6.3247 6.2090 6.0635 6.1680
##
## $se
##           Jan      Feb      Mar      Apr      May      Jun
## 1961 0.036716 0.042783 0.048091 0.052868 0.057249 0.061317
##           Jul      Aug      Sep      Oct      Nov      Dec
## 1961 0.065131 0.068734 0.072158 0.075426 0.078559 0.081571
```

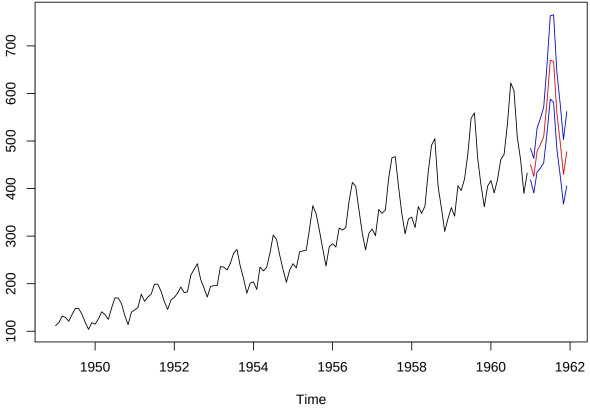
Prediction 2

```
ts.plot(log(airpass),air.pred$pred,col=c("black","red"))
```



Prediction 3

```
pred.values=exp(air.pred$pred)
lower=exp(air.pred$pred-2*air.pred$se)
upper=exp(air.pred$pred+2*air.pred$se)
ts.plot(airpass,pred.values,lower,upper,col=c("black","red'
```



Prediction 4

```
ts.plot(pred.values,lower,upper,col=c("red","blue","blue"))
```