B.Sc./Grad. Dip.: Probability Models and Time Series
MAS programmes: Stochastic Models and Time Series

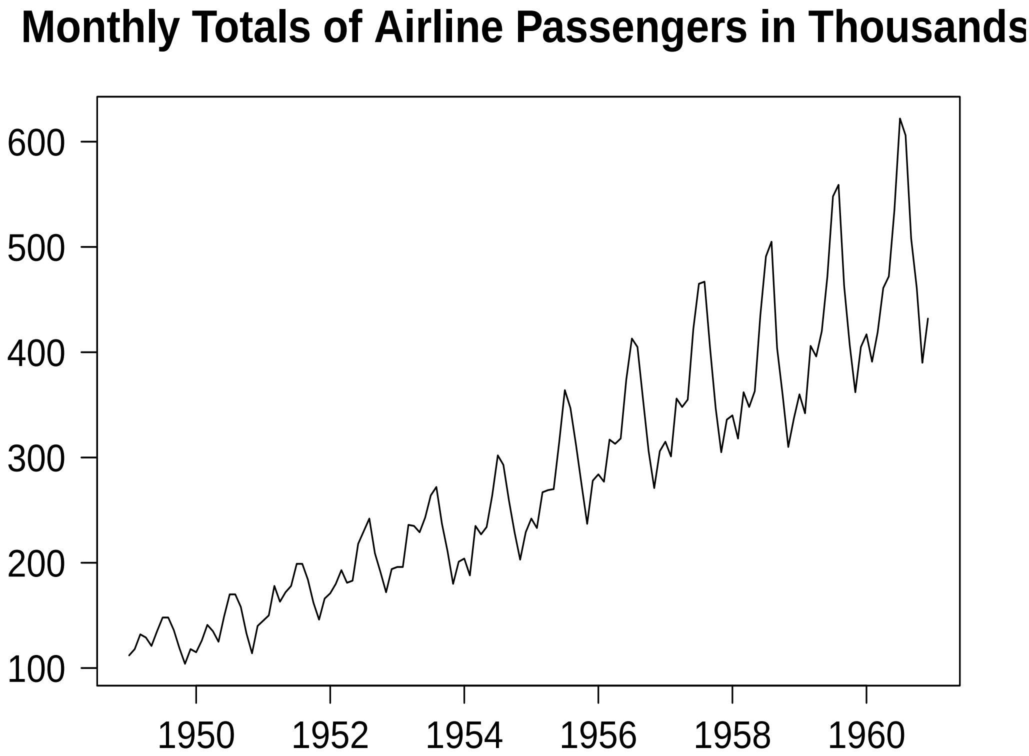
3 Stationary processes and autocorrelations

3.1 Introduction

A time series is a series of data, a sequence of observations recorded over time, usually at more or less equally spaced intervals of time. Many such series of data exhibit either a long-term *trend*, i.e., a systematic change of the mean level of the series over time, or *seasonality* (with period usually one year), or abrupt changes.

The following plot shows a series of monthly totals of international airline passengers in thousands, for the years 1949 - 1960, which in addition to an overall increasing trend has a period of length 12. (The data are taken from Box, Jenkins and Reinsel, *Time Series Analysis*)

```
> passengers<-c(112, 118, 132, 129, 121, 135, 148, 148, 136, 119, 104, 118, 115, 126,
+141, 135, 125, 149, 170, 170, 158, 133, 114, 140, 145, 150, 178, 163, 172, 178,
+199, 199, 184, 162, 146, 166, 171, 180, 193, 181, 183, 218, 230, 242, 209, 191,
+172, 194, 196, 196, 236, 235, 229, 243, 264, 272, 237, 211, 180, 201, 204, 188,
+235, 227, 234, 264, 302, 293, 259, 229, 203, 229, 242, 233, 267, 269, 270, 315,
+364, 347, 312, 274, 237, 278, 284, 277, 317, 313, 318, 374, 413, 405, 355, 306,
+271, 306, 315, 301, 356, 348, 355, 422, 465, 467, 404, 347, 305, 336, 340, 318,
+362, 348, 363, 435, 491, 505, 404, 359, 310, 337, 360, 342, 406, 396, 420, 472,
+548, 559, 463, 407, 362, 405, 417, 391, 419, 461, 472, 535, 622, 606, 508, 461,
+390, 432)
>
> air.ts <- ts(passengers, start = 1949, frequency = 12)
> air.ts
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
> plot(air.ts, xlab = "", ylab = "", main =
+"Monthly Totals of Airline Passengers in Thousands", las = 1)
```



However, some series appear to be *stationary*, i.e., their distributional properties do not change over time. In some approaches to the analysis of time series, when series with trend or seasonality are being analysed, they are first transformed into stationary series by some means or other. In any case, the study of stationary processes is fundamental to the analysis of time series.

In constructing models for time series, it turns out to be convenient to consider an underlying process that stretches back into the infinite past and forward into the infinite future. Thus we consider a doubly infinite sequence of random variables (r.v.s) $\{Y_t : t \in \mathbb{Z}\}$, a stochastic process in discrete time, where $\mathbb{Z}$ is the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$. We will also need to refer to sequences of random variables in which the subscript indexing runs over just a subset of $\mathbb{Z}$. Commonly used sets are listed below.

- $\mathbb{Z}^+ = \{1, 2, 3, \ldots\}$: the positive integers;
- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$: the natural numbers;
- $\{a, a+1, \ldots, b\}$: the set of integers running from a up to b (inclusive), where $a < b$, and
- $i \geq a$: refers to the set of integers that are greater than or equal to a, i.e. $\{a, a+1, a+2, \ldots\}$.

The $\{Y_t\}$ will usually be continuous r.v.s.

What we shall observe is a realization $y_1, y_2, \ldots, y_T$ of a finite section of the process, where y_t is the value that the process takes at time t. The observed data will be used to make inferences about the structure of the underlying process $\{Y_t\}$.

3.2 The definition of stationarity

Definition 3.2.1 (Strict Stationarity)
The process $\{Y_t\}$ is said to be *strictly stationary* if its probabilistic laws remain unchanged through shifts in time, i.e., if for every integer $n \geq 1$, every selection $t_1, t_2, \dots, t_n$ of distinct indices and every integer h, the joint distribution of $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$. □

Remarks 3.2.2
(i) In particular, taking $n = 1$, the marginal distribution of Y_t is the same for each t. Assuming existence, the mean and variance of Y_t is the same for each t.
(ii) Taking $n = 2$, for every pair t_1, t_2 and every h, the joint distribution of Y_{t_1}, Y_{t_2} is the same as the joint distribution of Y_{t_1+h}, Y_{t_2+h}. Another way of expressing this is that the joint distribution of $Y_t, Y_{t-\tau}$ depends only on the lag τ and not on t. (The *lag* is the difference in the subscripts, the time difference.)

Assume from now on that all first and second order moments of the process are finite.

Definition 3.2.3 (Moments)
(i) The first order moments of a stationary process are specified by the *process mean* μ, where

$$\mu = E[Y_t] \qquad t \in \mathbb{Z}.$$

(ii) The second order moments of a stationary process are specified by the *autocovariances* $\{\gamma_\tau\}$ at lag τ, where

$$\gamma_\tau = \mathrm{cov}(Y_t, Y_{t-\tau}) = E[(Y_t - \mu)(Y_{t-\tau} - \mu)] \qquad \tau \in \mathbb{Z}.$$

$$\gamma_\tau = Cov(Y_t, Y_{t-\tau}) = E[(Y_t - \mu)(Y_{t-\tau} - \mu)]$$
$$= E[(Y_{t-\tau} - \mu)(Y_t - \mu)] = Cov(Y_{t-\tau}, Y_t)$$
$$= \gamma_{-\tau}$$

□

Remarks 3.2.4
Note that $\gamma_0 = \mathrm{var}(Y_t)$ and that $\gamma_\tau = \gamma_{-\tau}$ for all τ, since, equivalently,

$\tau = 0$ ∴ no lag
$Cov(Y_t, Y_t) = Var(Y_t)$

$$\gamma_\tau = \mathrm{cov}(Y_t, Y_{t+\tau}).$$

The sequence $\{\gamma_\tau\}$ is called the *autocovariance function.* The autocovariance (or 'self' covariance) function represents the covariance of the process with previous values of itself.

Definition 3.2.5 (Autocorrelation)
The *autocorrelation* ρ_τ at lag τ is given by

$$\rho_\tau = \mathrm{cor}(Y_t, Y_{t-\tau}) = \gamma_\tau / \gamma_0 \qquad \tau \in \mathbb{Z}.$$

τ can be any lag in the integers,
$\tau \in \{\dots, -2, -1, 0, 1, 2, 3 \dots\}$

Note that $\rho_0 = 1$, $\rho_\tau = \rho_{-\tau}$ for all τ and $|\rho_\tau| \leq 1$ for all τ. ρ_τ is a measure of the linear dependency between values of the process at lag τ apart. The sequence $\{\rho_\tau\}$ is called the *autocorrelation function.* □

Recall: $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

Here: $\rho_\tau = \dfrac{Cov(Y_t, Y_{t-\tau})}{Cov(Y_t, Y_t)} = \dfrac{Cov(Y_t, Y_{t-\tau})}{Var(Y_t)}$

$\rho_{\tau=0} = \dfrac{\gamma_0}{\gamma_0} = 1$, $\rho_1 = \dfrac{\gamma_1}{\gamma_0}$

↳ $\{\rho_\tau\} = \left\{ \dfrac{\gamma_0}{\gamma_0}, \dfrac{\gamma_1}{\gamma_0}, \dfrac{\gamma_2}{\gamma_0}, \dfrac{\gamma_3}{\gamma_0}, \dots \right\}$

→ strong stationarity ↳ requires joint distributions of sequences of random numbers of the process to be the same as there separated by same integer, h. → joint dist for $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ same as $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$

Because many of the properties of stationary processes follow from the properties of their first and second moments, a weaker definition of stationarity is often used, which imposes conditions on the means and covariances of the process random variables.

↳ weaker stationarity just looks a PROCESS MEAN & AUTOCOVARIANCE

Definition 3.2.6 (Weak Stationarity)
The process is said to be *weakly stationary* (or *covariance stationary* or *wide-sense stationary*) if $E[Y_t]$ is constant over time and $\text{cov}(Y_t, Y_{t-\tau})$ depends only on the lag τ and not on t. □

$\mu = E[Y_t]$
μ = process mean.
→ constant $\forall t$

The first and second order moments of the process are then again specified by the process mean μ and the autocovariance function $\{\gamma_\tau\}$.

Given the assumption that all first and second order moments are finite, strict stationarity clearly implies weak stationarity. But not every weakly stationary process is also strictly stationary. However, any multivariate normal distribution is completely specified by its first and second order moments. Hence, if it is assumed that all the joint distributions of the Y_t are multivariate normal then weak stationarity does imply strict stationarity.

From now on, we shall refer to weakly stationary processes simply as stationary processes.

→ centre the data.

If $\{Y_t\}$ is a stationary process with process mean μ then we may work instead with the r.v.s $Y_t - \mu$, which does not alter the autocovariance function $\{\gamma_\tau\}$ but sets the process mean to zero. So in dealing with much of the theory of stationary processes we may **without any essential loss of generality assume that the process mean is zero**, in which case

random variable

$$\gamma_\tau = E[Y_t Y_{t-\tau}] \qquad \tau \in \mathbb{Z}.$$

$\gamma_\tau = Cov(Y_t, Y_{t-\tau}) = E[(Y_t-\mu)(Y_{t-\tau}-\mu)] = E[Y_t Y_{t-\tau}]$ when $\mu=0$.

$Y_t = Y_t - \mu$?

3.3 White noise processes

A simple but important type of stationary process is one in which the Y_t are independently and identically distributed (i.i.d.), with common variance σ^2, say, in which case

$$\gamma_0 = \sigma^2 \quad \text{and} \quad \gamma_\tau = 0 \quad \tau \neq 0,$$

$$\rho_0 = 1 \quad \text{and} \quad \rho_\tau = 0 \quad \tau \neq 0.$$

Y_t → i.i.d
$Var(Y_t) = Cov(Y_t, Y_t) = \sigma^2$
→ working with centred $\{Y_t\}$ so $\mu=0$.

$\gamma_0 = Cov(Y_t, Y_t) = Var(Y_t) = \sigma^2$
$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$

To use engineering/time series terminology, a stationary process whose autocovariances and autocorrelations satisfy the above two lines of conditions is known as a *white noise process.* A formal definition may be presented as follows.

If A and B are independent, then $P(A \cap B)P(B) = P(A)P(B)$, and $Cov(A,B)=0$, since independence implies 2 r.v's being uncorrelated.

Definition 3.3.1 (White Noise)
The sequence $\{Y_t\}$ is said to constitute a white noise process if it is stationary and the $\{Y_t\}$ are (pairwise) uncorrelated. □

↳ except for when $\tau=0$, and $\rho_0=1$.

This does not imply that the Y_t are i.i.d. → independence is a stronger condition than (un)correlation.

independence ⇒ uncorrelated → $Cov(X,Y)=0$
uncorrelated $\not\Rightarrow$ independence

$Cov(\varepsilon_t, \varepsilon_{t-\tau}) = E[(\varepsilon_t - \mu)(\varepsilon_{t-\tau} - \mu)] = E[\varepsilon_t \varepsilon_{t-\tau}]$ $\gamma_\tau = 0$ since ε_t uncorrelated for $\tau \neq 0$

$\gamma_0 = Cov(\varepsilon_t, \varepsilon_t) = \sigma^2$

$\rho_0 = 1$ $\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = 0.$

Remarks 3.3.2
(i) We shall use the notation $\{\epsilon_t\}$ for a white noise process, which, unless otherwise stated, will be assumed to have mean zero and variance σ^2. For such a process, then,

$$\gamma_0 = Cov(\varepsilon_t, \varepsilon_t) = E[\epsilon_t^2] = \sigma^2, \rho_0 = 1 \tag{1}$$
$$\gamma_\tau = Cov(\varepsilon_t, \varepsilon_{t-\tau}) = E[\epsilon_t \epsilon_{t-\tau}] = 0, \rho_\tau = 0 \quad \tau \neq 0. \tag{2}$$

(ii) In the context of forecasting, we shall make the stronger assumption that the ϵ_t are i.i.d. and, furthermore, that they are normally distributed. → i.i.d stronger than simply saying uncorrelated.

3.4 Sample autocovariances and autocorrelations

→ + white noise $\sim N(0, \sigma^2)$

Definition 3.4.1 (Sample Statistics)
Let $y_1, y_2, \ldots, y_T$ be a realization of part of a stationary process. → this is your actual, real data
(i) The sample mean, $\bar{y}$, is given by

T = history

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

(ii) The *sample autocovariance*, c_τ, at lag τ, is given by → but divide by T, not $T-\tau$

sum over $T-\tau$ times

γ_T = population autocovar

c_τ = sample autocovar AT LAG τ.

$$c_\tau = \frac{1}{T} \sum_{t=\tau+1}^{T} (y_t - \bar{y})(y_{t-\tau} - \bar{y}) \qquad \tau = 0, 1, \ldots, T-1.$$

(iii) The *sample autocorrelation*, r_τ, at lag τ, is given by

ρ_τ = population autocorrelation

r_τ = sample autocorrelation

$$r_\tau = \frac{c_\tau}{c_0} \qquad \tau = 0, 1, \ldots, T-1.$$

$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}$ $r_\tau = \frac{c_\tau}{c_0}$

□

$E[\bar{y}] = \mu$

First, note that $\bar{y}$ is an unbiased estimator of the process mean μ.

Also, the statistic c_τ is used as an estimator of γ_τ and r_τ as an estimator of ρ_τ. We also note the following:

$\hat{\gamma}_\tau = c_\tau$ $\hat{\rho}_\tau = r_\tau$

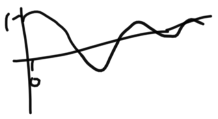
1. $r_0 = 1 = \rho_0$.

2. The range of t-values in the summation for c_τ depends on the lag τ. → b/c τ ≠ 1, and you realise that since $(y_{t-\tau} - \bar{y})$ needs to have y_1 initially then sum must start at $t = \tau + 1$

3. Although the divisor $T - \tau$ might seem more natural in the definition of c_τ, the divisor T is generally preferred. This ensures that the sample autocorrelations satisfy $|r_\tau| \leq 1$ for all τ.

4. We could not expect to estimate autocovariances and autocorrelations at lag T or greater if the observed time series is of length T, so the range of values of τ for which c_τ and r_τ are defined is in accordance with common sense.

The following table provides a summary of our notation for autocovariances and autocorrelations.



	mean	autocovariance	autocorrelation
process/population	μ	γ_τ	$\rho_\tau = \gamma_\tau/\gamma_0$
sample	$\bar{y}$	c_τ	$r_\tau = c_\tau/c_0$

The sequence $\{r_\tau : \tau = 0, 1, \ldots, T-1\}$ is the *sample autocorrelation function.* The plot of r_τ against τ is sometimes called the *correlogram.*

The statistic r_τ becomes less and less reliable as an estimator of ρ_τ as τ increases towards T. In practice, values of r_τ will be calculated and plotted over a limited range of τ values such as, for example, $0 \leq \tau \leq T/4$.

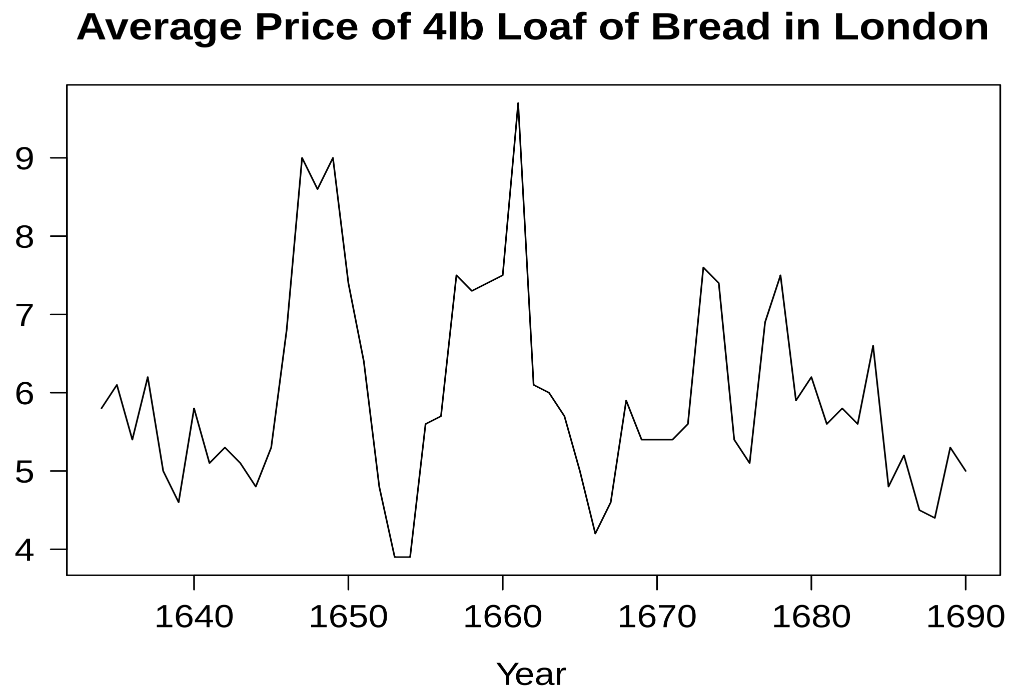
- The sample mean, sample autocovariances and sample autocorrelations may also be calculated for observed time series that are not necessarily assumed to come from an underlying stationary process. For example, if we have data with a highly regular seasonality of period s, we will observe large autocorrelations at lags $s, 2s, 3s, \ldots$

3.5 Bread price example

The following R output gives a preliminary analysis of the average annual price in pennies of a 4 lb loaf of bread in London for the years 1634 to 1690 (original series runs until 1757). The bread price data are plausibly a realization of a stationary process, but successive values are positively correlated.

```
> price<-c(5.8, 6.1, 5.4, 6.2, 5.0, 4.6, 5.8, 5.1, 5.3, 5.1, 4.8, 5.3, 6.8, 9.0, 8.6,
+ 9.0, 7.4, 6.4, 4.8, 3.9, 3.9, 5.6, 5.7, 7.5, 7.3, 7.4, 7.5, 9.7, 6.1, 6.0, 5.7, 5.0,
+ 4.2, 4.6, 5.9, 5.4, 5.4, 5.4, 5.6, 7.6, 7.4, 5.4, 5.1, 6.9, 7.5, 5.9, 6.2, 5.6, 5.8,
+ 5.6, 6.6, 4.8, 5.2, 4.5, 4.4, 5.3, 5.0, 6.4, 7.8, 8.5, 5.6, 7.1, 7.1, 8.0, 7.3, 5.7,
+ 4.8, 4.3, 4.4, 5.7, 4.7, 4.1, 4.1, 4.7, 7.0, 8.7, 6.2, 5.9, 5.4, 6.3, 4.9, 5.5, 5.4,
+ 4.7, 4.1, 4.6, 4.8, 4.5, 4.7, 4.8, 5.4, 6.0, 5.1, 6.5, 6.2, 4.6, 4.5, 4.0, 4.1, 4.7,
+ 5.1, 5.2, 5.3, 4.8, 5.0, 6.2, 6.4, 4.7, 4.1, 3.9, 4.0, 4.9, 4.9, 4.8, 5.0, 4.9, 4.9,
+ 5.4, 5.6, 5.0, 4.5, 5.0, 7.2, 6.1)
> BP.ts<-ts(price,start=1634)
> summary(BP.ts)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.900   4.800   5.400   5.652   6.200   9.700
> BP.short.ts<-ts(price[1:(1+(1690-1634))],start=1634)
> plot(BP.short.ts,xlab="Year",ylab="",
+      main="Average Price of 4lb Loaf of Bread in London",las=1)
> BP.acf<-acf(BP.ts,30)
> BP.acf
```

price in pennies for 4lb loaf of bread.

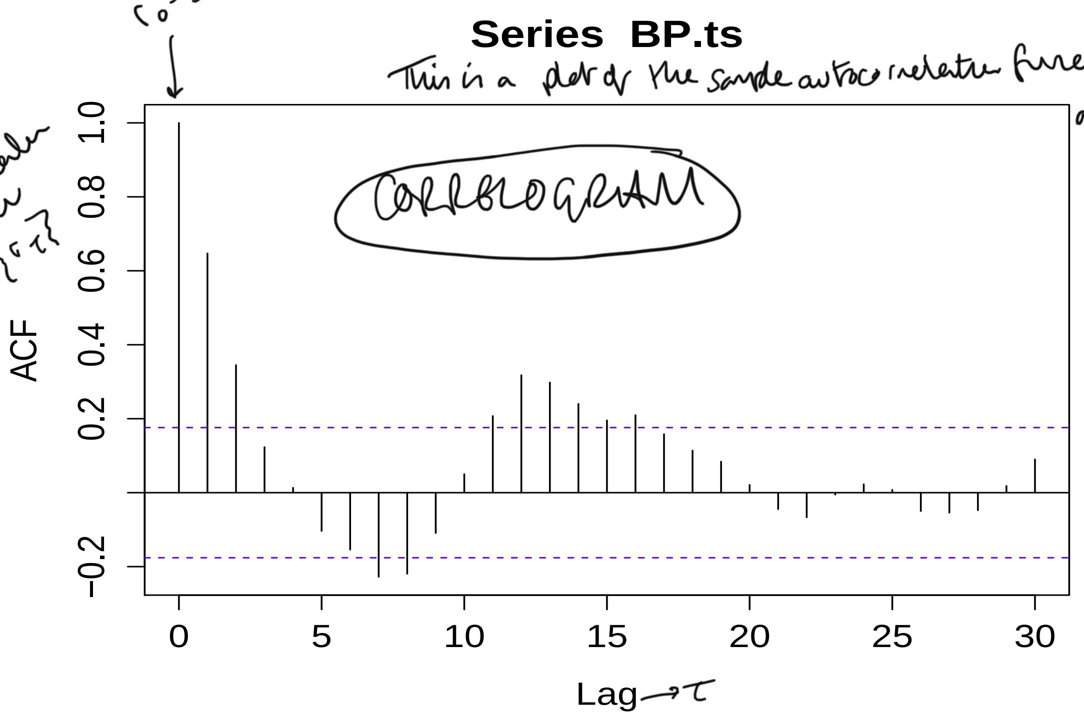


$r_0 = 1$

This is a plot of the sample autocorrelation function, r_τ, against τ.

Sample autocorrelation function $\{r_\tau\}$

$\boxed{r_\tau = \frac{c_\tau}{c_0}}$



White noise process $\{\epsilon_t\}$ characterised by: mean zero; variance $= \sigma^2$.

$\rightarrow \gamma_0 = Cov(\epsilon_t, \epsilon_t) = Var(\epsilon_t) = \sigma^2$ $\rightarrow \gamma_\tau = 0;\ \tau \neq 0$ → because white noise uncorrelated

$\rightarrow \rho_0 = \frac{\gamma_0}{\gamma_0} = 1$ $\rightarrow \rho_\tau = \frac{\gamma_\tau}{\gamma_0} = 0;\ \tau \neq 0$ ↳ so $\gamma_\tau = Cov(\epsilon_t, \epsilon_{t-\tau}) = 0$ since ϵ_t and $\epsilon_{t-\tau}$ uncorrelated

3.6 The model for a first order autoregressive process

Consider a simple model for the numbers of unemployed Y_t in successive months t. Ignoring seasonal effects and trend, to a first approximation, suppose that the number unemployed in any month is a fixed proportion ϕ of the unemployed in the previous month plus a number of newly unemployed workers seeking jobs. Assume that the newly unemployed form a white noise process with mean m, so that

fixed proportion of prev month → mean zero w.n. process

$$Y_t = \phi Y_{t-1} + m + \epsilon_t \qquad t \in \mathbb{Z}, \tag{3}$$

↳ mean of white noise process

where $\{\epsilon_t\}$ is a white noise process with mean zero. Assuming that the process $\{Y_t\}$ as defined by Equation (3) is stationary, let μ denote the process mean. Taking expectations in Equation (3) we find that

$E[Y_t] = E[\phi Y_{t-1}] + E[m] + E[\epsilon_t]$ (r.v; scalar; r.v; scalar; r.v)

$$\mu = \phi\mu + m,$$

$\mu = \phi\mu + m + 0$

$\therefore \mu = \phi\mu + m$

and hence, assuming that $\phi \neq 1$,

$$\mu = \frac{m}{1-\phi}.$$

$Y_t = \phi Y_{t-1} + (\mu - \phi\mu) + \epsilon_t$

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

The model of Equation (3) can be rewritten as

$\mu - \phi\mu = \mu(1-\phi) = m$

$\mu = \frac{m}{(1-\phi)}$

$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t, \qquad t \in \mathbb{Z}. \tag{4}$$

Replacing the r.v.s $\{Y_t - \mu\}$ by $\{Y_t\}$ in Equation (4), we obtain the simpler model

$\{Y_t - \mu\} \rightarrow \{Y_t\}$ $\boxed{Y_t = \phi Y_{t-1} + \epsilon_t}$

$$Y_t = \phi Y_{t-1} + \epsilon_t \qquad t \in \mathbb{Z}, \tag{5}$$

which has process mean zero. → $E[Y_t] = 0$ (because really it's $E[Y_t - \mu] = E[Y_t] - \mu = \mu - \mu = 0$

In general, a stationary process $\{Y_t\}$ which satisfies Equation (4) or Equation (5) is known as a *first order autoregressive process*, an AR(1) process, with autoregressive parameter ϕ. The term "autoregressive" is used, because Y_t is expressed as a regression on the previous value Y_{t-1} of the process itself. The autoregression is of "first order" because the regression is on only one previous process value.

This model, despite its simplicity, is a very useful one for modelling data, either on its own or as a component of more complex models; and in its simplicity it also has the attraction of being readily estimable. ↳ eg. ARMA / ARIMA.

For the present we shall use the version of the model specified by Equation (5), which is more convenient for the purposes of mathematical analysis. Both versions of the model have the same autocovariance and autocorrelation functions.

you don't see the white noise.

The process $\{Y_t\}$ generally represents a sequence of observable variables, whereas $\{\epsilon_t\}$ is a white noise process of unobservable random errors or innovations — each Y_t depends directly or indirectly on the previous values $\{Y_{t-i} : i \geq 1\}$ and $\{\epsilon_{t-i} : i \geq 1\}$, whereas each ϵ_t is a new input into the model, uncorrelated with the previous history of the process. Thus $\{\epsilon_t\}$ can be regarded as a process of mutually uncorrelated impulses, which drive the system. → really interesting

If the white noise term ϵ_t is removed from the model of Equation (5) then we are left with the deterministic model → no random variable

$$Y_t = \phi Y_{t-1} \qquad t \in \mathbb{Z}. \tag{6}$$

Will see ALOT more of this AR(1) notation.

AR = autoregressive → the RESPONSE Y_t is REGRESSED ONTO ITSELF, AT A PREVIOUS VALUE

1 = only one previous value!

↳ Would be AR(2) if $Y_t = \phi Y_{t-1} + \psi Y_{t-2} + \epsilon_t$.

ϕ here is the autoregressive parameter!

$Y_t = \phi Y_{t-1}$ general solution

This simple recurrence relation has the general solution $Y_t = A\phi^t$, where A is an arbitrary constant. Note that $Y_t \to 0$ as $t \to \infty$, whatever the value of A, if and only if $|\phi| < 1$. Thus $\{Y_t\}$ has the *stable limit point* 0 if and only if $|\phi| < 1$. → $Y_t \to 0$ as $t \to 0$ iff. $|\phi| < 1$

We have so far assumed that Equation (5) does indeed specify a stationary process. It turns out that the condition $|\phi| < 1$ is also the necessary and sufficient condition for there to exist a suitably defined stationary process $\{Y_t\}$ which satisfies Equation (5).

Condition to define stationary process: AUTOREGRESSIVE PARAM: $\boxed{|\phi| < 1}$

3.7 Infinite moving average representation of an AR(1) process

Using Equation (5) recursively, NICE! $Y_{t-1} = \phi Y_{t-2} + \epsilon_{t-1}$ ∴ $Y_t = \phi Y_{t-1} + \epsilon_t$

can express Y_t in terms of a realised value of the process n lags before.

$$
\begin{aligned}
Y_t &= \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \phi^2 Y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
&= \dots \\
&= \phi^n Y_{t-n} + \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i}. \qquad (7)
\end{aligned}
$$

$= \phi(\phi Y_{t-2} + \epsilon_{t-1}) + \epsilon_t$

$= \phi^2 Y_{t-2} + \phi \epsilon_{t-1} + \epsilon_t$

What we would like to do is to let $n \to \infty$ in Equation (7), state that $\phi^n Y_{t-n} \to 0$, and write

Y_t model contains two terms:

$Y_t = \phi^n Y_{t-n} + \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i}$ ② $\xrightarrow{n\to\infty}$ ① 0

$$Y_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}, \qquad (8)\ \Delta$$

which is the *infinite moving average* representation of the process $\{Y_t\}$. This would appear to be plausible if $|\phi| < 1$. However, we are dealing here with the convergence of sequences of random variables and have to be careful.

In the present context we use the concept of *convergence in mean square*:

Definition 3.7.1 (Convergence in Mean Square) "sequence of r.v.'s converge in mean square"

A sequence of r.v.s $\{\xi_n\}$ is said to converge *in mean square* to a r.v. ξ if $E[(\xi_n - \xi)^2] \to 0$ as $n \to \infty$. We may then use the notation that $\xi_n \xrightarrow{ms} \xi$. □

From Equation (7), if we assume that $\{Y_t\}$ is a stationary process and that $|\phi| < 1$,

$Y_t \xrightarrow{ms} Y$ if $E[(Y_t - Y)^2] \to 0$ as $t \to \infty$. ← the mean (i.e. the $E[\;]$ bit) square (i.e. the $(a-b)^2$ bit) → 0 (i.e. converges).

$$
\begin{aligned}
E\left[\left(\sum_{i=0}^{n-1} \phi^i \epsilon_{t-i} - Y_t\right)^2\right] &= E[\phi^{2n} Y_{t-n}^2] \\
&= \phi^{2n}\gamma_0 \to 0
\end{aligned}
$$

↳ $Y_t = \phi^n Y_{t-n} + \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i}$

↳ $\left(\sum_{i=0}^{n-1} \phi^i \epsilon_{t-i} - Y_t\right)^2$ becomes

$= \left(Y_t - \sum_{i=0}^{n-1} \phi^i \epsilon_{t-i}\right)^2$

$= (\phi^n Y_{t-n})^2 = \phi^{2n} Y_{t-n}^2$

as $n \to \infty$. Thus, under these assumptions,

infinite series of white noise terms, dampened by the AR param.

$$\sum_{i=0}^{n-1} \phi^i \epsilon_{t-i} \xrightarrow{ms} Y_t$$

and so the infinite moving average representation of Equation (8) is valid in the sense of convergence in mean square. The value of Y_t is represented as a linear function of the innovations ϵ_t going back into the infinite past. The geometric decline of the coefficients, ϕ^i, in the representation reflects the relative residual influence of the previous innovations ϵ_{t-i} on the current process value Y_t as we go back in time.

$E[\phi^{2n} Y_{t-n}^2] = \phi^{2n} E[Y_{t-n}^2]$

$= \phi^{2n}\gamma_0$

∴ when $n \to \infty$, $\phi^{2n}\gamma_0 \to 0$ when $|\phi| < 1$.

* Know $E[X^2] = Var(X) - (E[X])^2$

since $Var(X) = E[X^2] - (E[X])^2$

we know that $E[Y_{t-n}] = 0$ ∴ $(E[Y_{t-n}])^2 = 0$

so $E[Y_{t-n}^2] = Var(Y_{t-n}) = \sigma^2 = Cov(Y_{t-n}, Y_{t-n}) = \gamma_0$