2 Stationary Processes and Time Series I

Suppose that \( \{X_t\} \) is a ts such that \( E(X_t) = \mu < \infty \) and \( E(X_t^2) = C < \infty \), where \( \mu \) and \( C \) are constants independent of time \( t \). To achieve this in practice, it is necessary to apply suitable preliminary transformations as described before. Suppose that now we have a time series with following properties:

\[
E(X_t) = \text{Constant ind. of } t \text{ and } Var(X_t) = \text{Constant ind. of } t.
\]

It is known that the consecutive values of \( \{X_t\} \) are not independent and hence the variance function alone is not sufficient to specify the second order properties of a \( ts \). Since the consecutive observations are correlated, we define a measure called the autocovariance function in order to study the serial dependency of time series. The standardized version of this autocovariance function is called the autocorrelation function.

**Note:** Autocovariance and autocorrelation functions explain the serial dependence structure of neighbouring observations at different lags. The values can be compared with the ordinary covariance and correlation functions appear in bivariate data analysis.

**An Illustration**

**Note:** As the lag increases, it is clear from the above that autocovariance and autocorrelation functions decrease.
2.1 Autocovariance and Autocorrelation Functions

Suppose that \( \{X_t\} \) is a ts with mean \( \mu = E(X_t) \) and \( \sigma^2 = Var(X) \). The covariance between \( X_t \) and \( X_{t+k} \), \( Cov(X_t, X_{t+k}) \) or the autocovariance function of \( \{X_t\} \) at lag (separation) \( k \) and is denoted by \( \gamma(k) \) (or \( \gamma_k \)). This is defined as:

\[
\gamma(k) = Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = E(X_t X_{t+k}) - \mu^2, \text{ since } E(X_t) = \mu \text{ for all } t.
\]

Clearly, \( \gamma(0) = Var(X_t) = E(X_t^2) - \mu^2 \).

Let \( \rho(k) = Corr(X_t, X_{t+k}) \) be the autocorrelation function (acf) at lag \( k \). Since \( Var(X_t) = Var(X_{t=k}) \), we have

\[
\rho(k) = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t) \cdot Var(X_{t+k})}} = \frac{\gamma(k)}{\gamma(0)}.
\]

Notes:

(1) Some authors replace \( \gamma(k) \) by \( \gamma_k \) or \( \gamma_X(k) \) and \( \rho(k) \) by \( \rho_k \) or \( \rho_X(k) \).

(2) \( \rho(k) \) measures the dependence of observations with \( k \) time points apart of the series. That is, the correlation between the sets \( X_t, X_{t-1}, \ldots, X_{t+k}, X_{t+k-1}, \ldots \).

This is obviously a measure of the similarity between \( X_t \) and the same realization shifted to the right (or left) by \( k \) units.

3) It can be seen that \( |\rho(k)| \leq 1 \) for all \( k \).

4) Since \( \{X_t\} \) is time invariant statistical properties, \( Cov(X_t, X_{t-k}) = Cov(X_t, X_{t+k}) \).

Consequently,

\[
\gamma(k) = \gamma(-k).
\]

Estimation of \( \gamma(k) \) and \( \rho(k) \)

Suppose that we have a time series of size \( n, X_1, X_2, \ldots, X_n \). Let \( C_{n,k} \) and \( R_{n,k} \) be the sample autocovariance and autocorrelation estimators at lag \( k \) of \( \gamma(k) \) and \( \rho(k) \).

(1) An estimator of \( \gamma(k) \) based on \( n \) observations is given by

\[
C_{n,k} = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}); \quad k = 0, \ldots, n - 1.
\]
(See Chatfield p.20).

**Note:** Although the denominator of the rhs should be \( n - k \), many authors use \( n \) in order to simplify the distributional properties of \( C_{n,k} \). (for large \( n \), the difference between \( n \) and \( n - k \) is negligible.

(2) An estimator of \( \rho(k) \) based on \( n \) observations is given by

\[
R_{n,k} = \frac{C_{n,k}}{C_{n,0}}
\]

**Note:** For simplicity some authors write \( C_k \) for \( C_{n,k} \) and \( R_k \) for \( R_{n,k} \).

**Sample estimates:** Suppose that we have \( n \) observations \( x_1, x_2, \ldots, x_n \). Then the corresponding sample autocovariance and autocorrelation functions (or estimates) at lag \( k \) are:

- \( c_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{n} \) and
- \( r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \).

For example,

\[
c_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{n} \quad \text{and} \quad r_1 = \frac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}
\]

are used to estimate \( C_1 \) and \( R_1 \).

**Note:** It is clear that \( r_1 \) is approximately the ordinary sample correlation coefficient between the pairs of consecutive values \((x_1, x_2), (x_2, x_3), (x_3, x_4), \ldots, (x_{n-1}, x_n)\). This measures the internal association of the series \( x_1, \ldots, x_{n-1} \) and \( x_2, \ldots, x_n \) (at step 1 or lag 1) between successive readings.

Similarly, \( r_2 \) is approximately the ordinary sample correlation coefficient between the pairs of values \((x_1, x_3), (x_2, x_4), (x_3, x_5), \ldots, (x_{n-2}, x_n)\) and so measures the internal association of the series \( x_1, \ldots, x_n \) at step 2 (or lag 2) between readings.

**Example.** Compute \( r_1 \), \( r_2 \), and \( r_3 \) for the time series give below:

<table>
<thead>
<tr>
<th>( t )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_t )</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Solution:

\[
r_1 = \frac{\sum_{t=1}^{10-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{10}(x_t - \bar{x})^2} = \frac{c_1}{c_0}
\]
\[ \bar{x} = 4 \quad \text{and} \quad \sum (x_i - \bar{x})^2 = 48. \]

Therefore \[ r_1 = \frac{(1 - 4)(4 - 4) + (4 - 4)(5 - 4) + \cdots + (2 - 4)(1 - 4)}{48} = -0.0625 \]

\[ r_2 = \frac{(x_1 - \bar{x})(x_3 - \bar{x}) + (x_2 - \bar{x})(x_4 - \bar{x}) + \cdots + (x_8 - \bar{x})(x_{10} - \bar{x})}{\sum_{t=1}^{10} (x_t - \bar{x})^2} \]

\[ = \frac{(1 - 4)(5 - 4) + (4 - 4)(7 - 4) + \cdots + (1 - 4)(4 - 4)}{48} = -0.23 \]

Similarly, \( r_3 = 0.0625, r_4 = -0.1042, r_5 = -0.1875, \) etc.

**Note:** Many compute packages can be used to compute \( c_1 \) and \( r_1 \). For example, the **R command** is: \( r=acf(data) \). The numerical values of the acf can be viewed by just typing \( r \).

**Note:** For any time series (stationary or not), \( x_1, x_2, \ldots, x_n \), the autocorrelation function (acf) is a very useful diagnostic tool. If the

Now we study the sampling properties of \( \bar{X}_n, C_{n,k} \) and \( R_{n,k} \) and study their distributions.

### 2.2 Sampling properties of \( \bar{X}_n, C_{n,k} \) and \( R_{n,k} \)

#### Sampling Properties of \( \bar{X}_n \)

Let \( \bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \) be an estimator of \( \mu \).

**Exercise:** Show that \( E(\bar{X}_n) = \mu \) and \( Var(\bar{X}_n) = \frac{1}{n} \sum_{k=-n}^{n-1} (1 - \frac{|k|}{n}) \gamma(k) \).

Now suppose that \( \sum |\gamma(k)| < \infty \).

Then clearly, \( \sum_{k=-n+1}^{n-1} (1 - \frac{|k|}{n}) \gamma(k) \to \sum_{-\infty}^{\infty} \gamma(k) \) as \( n \to \infty \) and therefore we have

\[ Var(\bar{X}_n) \approx \frac{1}{n} \sum_{-\infty}^{\infty} \gamma(k) \]

for large \( n \).

When \( \{X_t\} \) is a Gaussian time series, we have the following approximate normal distribution for \( \bar{X}_n \) for large \( n \)

\[ \bar{X}_n \approx N(\mu, \frac{1}{n} \sum_{-\infty}^{\infty} |\gamma(k)|). \]
Sampling Properties of $C_{n,k}$

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

$$E(C_{n,k}) =$$

$$= ................................................................. \text{as } n \to \infty$$

i.e. $C_k$ or $C_{n,k}$ is asymptotically unbiased for $\gamma(k)$.

Sampling Properties of $R_{n,k}$

It can be seen that for large $n$, under the null hypothesis $H_0 : \rho(k) = 0$, $R_{n,k} \sim \mathcal{N}(-\frac{1}{n}, \frac{1}{n})$. Thus under the null hypothesis of $H_0 : \rho(k) = 0$, approximately 95% of the sample autocorrelations should fall between the bounds $-\frac{1}{n} \pm \frac{1.96}{\sqrt{n}}$. These bounds are further approximated by $\pm \frac{2}{\sqrt{n}}$ for large $n$.

i.e. Observed values of $r_k$ which fall outside $\pm \frac{2}{\sqrt{n}}$ are ‘significantly’ different from zero at 5% level.

Note: If $k$ is large relative to the sample size $n$, then the estimates $c_k$ and $r_k$ are very unreliable. The calculations are reasonable for $n \geq 50$ and a general rule of thumb is to restrict calculations of $c_k$ and $r_k$ for $n \geq 50$ and $k \leq n/4$.

The Distribution of $R_k$ under the Null Hypothesis of $H_0 : \rho_k = 0$

For a purely random process or under the null hypothesis of $H_0 : \rho_k = 0$, it can be shown that each $r_k$ is approximately independently and identically distributed normal random variables with mean 0 and variance $1/n$ for large $N$. Hence a simple test to see whether a given time series is a realization of purely random process (independently and identically distributed random variables) look at the proportion
of individual sample $r_k$’s satisfy

$$|r_k| > 2/\sqrt{n}, \quad k = 1, 2, ....$$

If the proportion is $1/20$ or less, the sequence is likely a random process.

**A portmantau test**

Suppose that all acf values are zero up to a maximum specified lag $\ell$. Then it is clear that the statistic

$$Q = n \sum_{k=1}^{\ell} r_k^2$$

has a $\chi^2$ distribution with $\ell$ degrees of freedom. *This test statistic can be used to test if all acf values are collectively zero up to a maximum lag $\ell$.*

There is another tool to recognise the serial dependence of time series data. This is known as the partial autocorrelation function (pacf).

**The Sample Correlogram**

The plot of $r_k$ against $k$ is called the sample correlogram of the data. This correlogram can be used to detect the long run behaviour of the autocorrelation structure of the series.

**Detection of Randomness, short-term and long-term correlations of a TS**

First obtain the correlogram of the data and mark $\pm 2/\sqrt{n}$ boundaries.

(1) **Purely Random Series or No Memory Series**

In this case $\rho(k) = 0$ for all $k$, i.e. the sample acf $\{r_k\}$ should be not significant at all lags

**Note:** However, a plot of the first 20 values of $r_k$, we expect to find one ‘slightly significant’ value on average even when the ts really is random (1 in 20 is 5%). If the significant values are at low lags (eg. one, two), then a care must be taken before proceeds. See Chatfield p.21 for details.
(2) **Short-term or Short Memory Series:** In this situation $r_k$’s are significant for small $k$ (i.e. Lags 1, 2, 3 etc.). $r_k \approx 0$ for moderate $k$. i.e. a.c.f. should die-out quickly.

(diagram)

Example: If $r_1$ is significant and $r_k \approx 0; k > 1$ tells us that there is a strong lag 1 dependence.

(3) **Long-term with a Linear Trend:** $r_k$’s are significant for many values of $k$. \{${r_k}$\} has many significant values. i.e. a.c.f. do not die out quickly.

(diagram)

(4) **Seasonal Fluctuation:** If a ts contains a seasonal component, then the correlogram will also exhibit an oscillation at the same frequency.

(diagram)
(5) Long Memory Time Series This type of ts exhibits a hyperbolic decay of the correlogram with a long tail behaviour.

The above arguments suggest that the sample correlogram can be used as a useful diagnostic tool to detect serial correlation structure of a time series.

2.3 The Autocorrelation Plot as Diagnostic Tool

We summarise the above information below:

(1) The acf, like most techniques of statistical inference for time series, requires the series to be ‘long’ (i.e. $n$ large, say $n \geq 40$).

(2) If in the series $x_1, \ldots, x_n$ a linear trend is present or the series is non-stationary, the plot of $r_k$ against $k$ decreases very slowly with increasing $k$.

(3) If a periodic component of period $d$ is present in the series $x_1, \ldots, x_n$ (and hence shows up as a periodic effect of period $d$ in a plot of $x_t$ vs. $t$), it also shows up as a periodic effect of period $d$ in the plot of $r_k$ vs. $k$.
(4) If the sequence $x_1, \ldots, x_N$ is stationary, the plot of $r_k$ against $k$ decreases rapidly as $k$ increases.

Note: The (sample) autocorrelation at lag $k$, $r_k$, is a measure of dependence of readings time-distance $k$ apart in a stationary time series, and is defined for $k \geq 1$. It is an estimate of the population quantity $\rho(k)$ or $\rho_k = \text{Corr}(x_t, x_{t+k})$.

Example: Comment on a time series of $N = 100$ with the following acf.

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_k$</td>
<td>0.7</td>
<td>-0.2</td>
<td>0.1</td>
<td>0.6</td>
<td>0.1</td>
<td>0.02</td>
<td>0.31</td>
<td>0.55</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Solution: The correlogram of the above series with boundaries $\pm \frac{2}{\sqrt{100}} = \pm 0.2$ is:
Comments: The \( r_k \)'s being correlation coefficients, it is clear that there is not a sharp drop-off of \( r_k \) with \( k \). That is, there are large values of \( r_k \) at lags \( k = 1, 4, 7, 8, 9 \) for the data set indicating non-stationarity. Alternatively, there could be a periodicity of the data set with period 3 or 4. However, this is very hard to justify as we have only 10 acf values.

The Partial Autocorrelation Function (PACF)

Given a stationary time series \( \{X_t\} \), the partial autocorrelation of lag \( k \) is defined as the additional autocorrelation between \( \{X_t\} \) and \( \{X_{t+k}\} \) when the linear dependence of \( \{X_{t+1}\} \) through to \( \{X_{t+k-1}\} \) removed. Equivalently, it is the correlation between \( X_t \) and \( X_{t+k} \) when the (linear) effect, on each of \( x_t \) and \( x_{t+k} \), of the intervening readings \( \{X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}\} \) has been removed. This quantity is denoted by \( \pi_k \) (or \( \pi(k) \) by some authors). Analogously, a sample estimate \( r_k \) (or \( \hat{\rho}_k \)) of \( \rho_k \), the corresponding sample estimate of \( \pi_k \) is denoted by \( \hat{\pi}_k \).

It is easy to see that, for a time series \( \{X_t\} \) with time invariant first and second order moments and autocovariances

\[
\pi_1 = \rho_1 \text{ and } \pi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.
\]

Therefore, the first two pacf estimates are

\[
\hat{\pi}_1 = r_1, \quad \hat{\pi}_2 = \frac{(r_2 - r_1^2)}{(1 - r_1^2)}.
\]

The values and plot of \( \hat{\pi}_k, k \geq 1 \) (the sample partial autocorrelation function: PACF) against \( k \) serve as a diagnostic tool complementing the role of the ACF. They are produced by many statistical packages with a time-series capability, such as SPlus, SAS, MINITAB and R.

The R command: acf (data, type = “partial”)

Now we give the formal definition of a stationary time series based on the moments of \( \{X_t\} \).