Time Series Analysis – Lecture 2

Fundamental Concepts

Dr. Kevin Burke

1 Stochastic Process

We will model a time series as a sequence of random variables

$$\{Y_1, Y_2, Y_3, \ldots\} = \{Y_t \,|\, t \in \mathbb{N}\}$$

called a **stochastic process** where Y_t is a random variable indexed by the time-point t.

We will consider only discrete, equally-spaced time-points which arises in daily, weekly, monthly, quarterly or annual data.

As is standard in probability theory, lower case letters, e.g., y_t, will be used to denote observations and an observed time-series is denoted by the vector $\mathbf{y} = (y_1, \ldots, y_n)$, i.e., we observe n values of the process.

Note: We will break the above rule by using lower-case e_t to denote a zero-mean random variable (with variance σ_e^2) which represents an error term. However, there is usually no confusion here as we do not *observe* errors (although we can estimate them). The use of a lower-case random error should be familiar from regression analysis: $Y = \beta_0 + \beta_1 x + e$.

A time series $\{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}\}$ is completely determined by its joint distribution which has density function

$$f(y_{t_1}, \ldots, y_{t_n})$$

and distribution function

$$\begin{aligned} F(y_{t_1}, \ldots, y_{t_n}) &= \Pr(Y_{t_1} \leq y_{t_1}, \ldots, Y_{t_n} \leq y_{t_n}) \\ &= \int_{-\infty}^{y_{t_1}} \cdots \int_{-\infty}^{y_{t_n}} f(y_{t_1}, \ldots, y_{t_n})\, dy_{t_1} \cdots dy_{t_n}. \end{aligned}$$

However, it is virtually impossible to understand multi-dimensional distributions (for example by plotting f or F) due to their complexity. For this reason, we typically focus on the first and second moments (i.e., mean and covariance).

It is worth noting that if the joint-distribution is (approximately) a multivariate normal distribution then the density function is given by

$$\begin{aligned} &f(y_{t_1}, \ldots, y_{t_n}) \\ &\quad = [(2\pi)^n \det(\Sigma)]^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})\right] \end{aligned}$$

which is fully characterised by the first two moments via the mean vector $\boldsymbol{\mu} = [E(Y_{t_1}), \ldots, E(Y_{t_n})]$ and the covariance matrix, Σ, whose elements are given by $\mathrm{Cov}(Y_{t_i}, Y_{t_j})$.

2 Moments

The **mean function** is given by

$$\mu_t = E(Y_t)$$

which may or may not depend on the time-point t.

The **autocovariance function** (acvf) is given by

$$\gamma_{t,s} = E[(Y_t - \mu_t)(Y_s - \mu_s)] = E(Y_t Y_s) - \mu_t \mu_s$$

where $\gamma_{t,s} = \gamma_{s,t}$.

The **variance function** is

$$\gamma_{t,t} = \mathrm{Cov}(Y_t, Y_t) = \mathrm{Var}(Y_t).$$

Thus, the **autocorrelation function** (acf) is

$$\rho_{t,s} = \frac{\mathrm{Cov}(Y_t, Y_t)}{\sqrt{\mathrm{Var}(Y_t)\mathrm{Var}(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} \in [-1, 1]$$

where $\rho_{t,s} = \rho_{s,t}$. The acf is a *key quantity* in the analysis of time series models and data.

Note: The random variables Y_t and Y_s come from the *same* time series at different time-points. Hence, we use the phrase *auto*-correlation (where "auto" means with itself) as this is a measure of linear dependence between different points on the *same series*.

Example 2.1. *Random Walk*
Let $e_1, e_2, \ldots$ be a sequence of independent, identically distributed (iid) random variables with zero mean and variance σ_e^2.

Now construct the following time series:

$$\begin{aligned} Y_1 &= e_1 \\ Y_2 &= e_1 + e_2 &&= Y_1 + e_2 \\ Y_3 &= e_1 + e_2 + e_3 &&= Y_2 + e_3 \\ &\vdots &&\vdots \\ Y_t &= e_1 + \cdots + e_t &&= Y_{t-1} + e_t \end{aligned}$$

Thus, the mean function is given by

$$\begin{aligned} \mu_t = E(Y_t) &= E(e_1) + \cdots + E(e_t) \\ &= 0 + \cdots + 0 \\ &= 0. \end{aligned}$$

The variance is

$$\begin{aligned}\gamma_{t,t} = \text{Var}\,(Y_t) &= \text{Cov}(Y_t, Y_t)\\ &= \text{Var}\left(\sum_{i=1}^{t} e_i\right)\\ &= \sum_{i=1}^{t} \text{Var}\,(e_i) + 2\sum_{i\neq j} \underbrace{\text{Cov}\,(e_i, e_j)}_{=0}\\ &(\text{Cov}(e_i, e_j) = 0 \text{ when } i \neq j \text{ due to independence})\\ &= \sum_{i=1}^{t} \sigma_e^2 = t\,\sigma_e^2.\end{aligned}$$

Thus, the variance is increasing with time.

We will now calculate the covariance between Y_t and Y_{t-k} where $t > k \geq 0$:

$$\begin{aligned}\gamma_{t,t-k} &= \text{Cov}(Y_t, Y_{t-k})\\ &= \text{Cov}\left(\sum_{i=1}^{t} e_i, \sum_{j=1}^{t-k} e_j\right)\\ &= \sum_{i=1}^{t}\sum_{j=1}^{t-k} \text{Cov}\,(e_i, e_j)\,.\end{aligned}$$

$$\text{Note:}\qquad \text{Cov}\,(e_i, e_j) = \begin{cases} 0 & \text{when } i \neq j\\ \sigma_e^2 & \text{when } i = j\end{cases}$$

$$\begin{aligned}\Rightarrow \gamma_{t,t-k} &= \sum_{i=1}^{t-k} \text{Cov}\,(e_i, e_i)\\ &= \sum_{i=1}^{t-k} \text{Var}\,(e_i)\\ &= \sum_{i=1}^{t-k} \sigma_e^2 = (t-k)\,\sigma_e^2.\end{aligned}$$

Note that the covariance formula includes the variance formula when $k = 0$: $\gamma_{t,t} = \gamma_{t,t-0} = (t-0)\sigma_e^2 = t\sigma_e^2$.

The autocorrelation function is then

$$\begin{aligned}\rho_{t,t-k} &= \frac{\gamma_{t,t-k}}{\sqrt{\gamma_{t,t}\;\gamma_{t-k,t-k}}}\\ &= \frac{(t-k)\,\sigma_e^2}{\sqrt{t\,\sigma_e^2\,(t-k)\,\sigma_e^2}}\\ &= \frac{(t-k)\,\sigma_e^2}{\sqrt{t}\,\sqrt{t-k}\,\sigma_e^2}\\ &= \sqrt{\frac{t-k}{t}} = \sqrt{1-\frac{k}{t}}.\end{aligned}$$

Note that when t is very large, $\rho_{t,t-k} \approx 1$ which means that neighbouring values become highly positively correlated as time goes by. However, for values which are far apart (i.e., $k \approx t$), $\rho_{t,t-k} \approx 0$ which means distant points are less and less correlated.

Example 2.2. *Moving Average*
Let $e_1, e_2, \ldots$ be a sequence of independent, identically distributed (iid) random variables with zero mean and variance σ_e^2.

Now define a moving average process:

$$Y_t = e_t + \tfrac{1}{2}e_{t-1}$$

The mean function is given by

$$\begin{aligned}\mu_t = E(Y_t) &= E(e_t + \tfrac{1}{2}e_{t-1})\\ &= E(e_t) + \tfrac{1}{2}E(e_{t-1})\\ &= 0 + \tfrac{1}{2}(0) = 0\end{aligned}$$

and the variance function is

$$\begin{aligned}\gamma_{t,t} = \text{Var}\,(Y_t) &= \text{Var}(e_t + \tfrac{1}{2}e_{t-1})\\ &= \text{Var}(e_t) + \text{Var}(\tfrac{1}{2}e_{t-1}) + 2\text{Cov}(e_t, \tfrac{1}{2}e_{t-1})\\ &= \text{Var}(e_t) + \tfrac{1}{4}\text{Var}(e_{t-1}) + 2(\tfrac{1}{2})\text{Cov}(e_t, e_{t-1})\\ &= \sigma_e^2 + \tfrac{1}{4}\sigma_e^2 + (1)(0) = \tfrac{5}{4}\sigma_e^2.\end{aligned}$$

The autocovariance is (for $t > k \geq 0$)

$$\begin{aligned}\gamma_{t,t-k} &= \text{Cov}\,(Y_t, Y_{t-k})\\ &= \text{Cov}(e_t + \tfrac{1}{2}e_{t-1}, e_{t-k} + \tfrac{1}{2}e_{t-k-1})\\ &= \text{Cov}(e_t, e_{t-k}) + \tfrac{1}{2}\text{Cov}(e_t, e_{t-k-1})\\ &\quad + \tfrac{1}{2}\text{Cov}(e_{t-1}, e_{t-k}) + \tfrac{1}{4}\text{Cov}(e_{t-1}, e_{t-k-1})\end{aligned}$$

Now note the following:

- $\text{Cov}(e_t, e_{t-k}) = \sigma_e^2$ for $k = 0$ and zero otherwise.
- $\text{Cov}(e_t, e_{t-k-1})$ is zero for all values of k.
- $\text{Cov}(e_{t-1}, e_{t-k}) = \sigma_e^2$ for $k = 1$ and zero otherwise.
- $\text{Cov}(e_{t-1}, e_{t-k-1}) = \sigma_e^2$ for $k = 0$ and zero otherwise.

Putting the above together we get

$$\gamma_{t,t-k} = \begin{cases} \sigma_e^2 + \frac{1}{4}\sigma_e^2 = \frac{5}{4}\sigma_e^2 & \text{for } k = 0\\ \frac{1}{2}\sigma_e^2 & \text{for } k = 1\\ 0 & \text{for } k \geq 2\end{cases}$$

The autocorrelation function is given by

$$\begin{aligned}\rho_{t,t-k} &= \frac{\gamma_{t,t-k}}{\sqrt{\gamma_{t,t}\ \gamma_{t-k,t-k}}} \\ &= \frac{\gamma_{t,t-k}}{\sqrt{\frac{5}{4}\sigma_e^2\ \frac{5}{4}\sigma_e^2}} \\ &= \frac{\gamma_{t,t-k}}{\sqrt{(\frac{5}{4}\sigma_e^2)^2}} \\ &= \frac{\gamma_{t,t-k}}{\frac{5}{4}\sigma_e^2} \\ &= \begin{cases} 1 & \text{for } k=0 \\ \frac{2}{5} & \text{for } k=1 \\ 0 & \text{for } k \geq 2 \end{cases}\end{aligned}$$

Note that, for this process, the mean and variance are constant with time, i.e., do not depend on t and, similarly, the covariance depends only on the *lag*, k, but not t. This brings us to the concept of *stationarity*.

---

3 Stationarity

In practice, we require that a time series is **stationary** (or we first transform it into a stationary series) in order to make inferences about its structure.

3.1 Strictly Stationary

Stationarity means that the probability rules governing the process do not depend on time, i.e., they do not depend on t.

A process $\{Y_t\}$ is said to be **strictly stationary** if the if the joint distribution of $Y_{t_1}, \ldots, Y_{t_n}$ is the same as the joint distribution of $Y_{t_1-k}, \ldots, Y_{t_n-k}$. In other words, the distribution of one section of the time series is the same as that of another (potentially overlapping) section. Formally,

$$f(y_{t_1}, \ldots, y_{t_n}) = f(y_{t_1-k}, \ldots, y_{t_n-k}).$$

As mentioned in Section 1, we usually focus on the first and second moments rather than investigating the multivariate distribution.

Setting $n = 1$ gives $f(y_t) = f(y_{t-k})$ for all t and k which implies

$$\begin{aligned}\mu_t = E(Y_t) = E(Y_{t-k}) &= \mu_{t-k} \\ &= \mu_{t-t} \qquad \text{(setting } k=t\text{)} \\ &= \mu_0\end{aligned}$$

i.e., the mean function is a constant - it does not depend on time. The same argument can be used to show that the variance is constant.

Setting $n = 2$ gives $f(y_t, y_s) = f(y_{t-k}, y_{s-k})$ for all t, s and k which implies

$$\begin{aligned}\gamma_{t,s} = \text{Cov}(Y_t, Y_s) &= \text{Cov}(Y_{t-k}, Y_{s-k}) \\ &= \gamma_{t-k,s-k} \\ &= \gamma_{t-s,0} \qquad \text{(setting } k=s\text{)} \\ &= \gamma_{0,s-t} \qquad \text{(setting } k=t\text{)}\end{aligned}$$

Thus, the covariance depends only on the time difference $|t-s|$ and not on the actual times.

3.2 Simplified Notation

For stationary series we can simplify our notation. Since the autocovariance depends only on the time difference we can write

$$\gamma_{t,s} = \gamma_{t-s}$$

i.e., dropping the unnecessary "0" in the subscript.

Now the lag-k autocovariance is given by

$$\gamma_{t,t-k} = \gamma_{t-(t-k)} = \gamma_{t-t+k} = \gamma_k.$$

Also note that

$$\gamma_{t-k,t} = \gamma_{t-k-t} = \gamma_{-k}.$$

and, hence, by symmetry of covariance (i.e., $\gamma_{t,t-k} = \gamma_{t-k,t}$), we have that $\gamma_k = \gamma_{-k}$.

Thus we have the following:

$$\begin{aligned}\gamma_k &= \text{Cov}(Y_t, Y_{t-k}) \\ \gamma_0 &= \text{Var}(Y_t) \\ \rho_k &= \text{Corr}(Y_t, Y_{t-k}) = \frac{\gamma_k}{\sqrt{\gamma_0\gamma_0}} = \frac{\gamma_k}{\gamma_0}\end{aligned}$$

3.3 Weakly Stationary

Typically we do not require that time series are strictly stationary. **Throughout this course "stationary" is intended to mean weakly stationary** which pertains only to the first two moments:

1. The mean function is constant with time (and finite): $\mu_t = \mu < \infty$.
2. The autocovariance function depends only on the time-lag (and is finite) $\gamma_{t,t-k} = \gamma_k < \infty$.

Example 3.1. *White Noise* The sequence $\{e_1, e_2, \ldots\}$ of independent, identically distributed (iid) random variables with zero mean and variance σ_e^2 is referred to as **White noise**.

White noise is an example of a (strictly) stationary series which can be used to construct other series such as the random walk or the moving average process.

Note that the random walk is non-stationary since its autocovariance depends on time. However, the moving average process is stationary since the mean is constant and the autocovaraince depends only on the lag, k.

4 Sample ACF

4.1 Background

For a two vectors of data $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, constituting n pairs (x_i, y_i), the sample covariance is given by

$$\frac{1}{n}\sum_{i=1}^{n}(x_t - \bar{x})(y_t - \bar{y})$$

where we could have divided by $n-1$ for an unbiased estimator (we will not worry about that here) and $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ and $\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$ are the sample means.

In the case of a time series, we can calculate the lag-1 autocovariance based on $\mathbf{y}_1 = (y_1, \ldots, y_{n-1})$ and $\mathbf{y}_2 = (y_2, \ldots, y_n)$ where there are $n-1$ pairs. Applying the usual formula above suggests the estimate:

$$\frac{1}{n-1}\sum_{t=1}^{n-1}(y_t - \bar{y}_1)(y_{t+1} - \bar{y}_2)$$

For the lag-2 autocovariance, we can use $\mathbf{y}_1 = (y_1, \ldots, y_{n-2})$ and $\mathbf{y}_2 = (y_3, \ldots, y_n)$ where there are $n-2$ pairs. This leads to the estimate

$$\frac{1}{n-2}\sum_{t=1}^{n-2}(y_t - \bar{y}_1)(y_{t+2} - \bar{y}_2)$$

and, more generally, the lag-k autocovariance could be estimated by

$$\frac{1}{n-k}\sum_{t=1}^{n-k}(y_t - \bar{y}_1)(y_{t+k} - \bar{y}_2).$$

4.2 Estimate

Stationarity is typically assumed and the above formula is simplified to produce the *standard* **lag-k autocovariance estimate**:

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

This formula differs to the previous one in two ways:

1. $\bar{y}_1$ and $\bar{y}_2$ are both replaced by the overall mean from the whole series $\bar{y} = \sum_{t=1}^{n} y_t$. This is due to assuming stationarity, i.e., the mean should be equal in different sections of the series.
2. We divide by n instead of $n-k$. From a practical perspective this will not change things much in time series data where n is usually quite large. However, dividing by n also ensures that the sequence $\{\hat{\gamma}_k\}$ is *non-negative definite* – a property required for the autocovariance function of a stationary process (Bochner's Theorem – details omitted).

Thus, the estimated lag-k autocorrelation is

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

$$= \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

4.3 Distribution

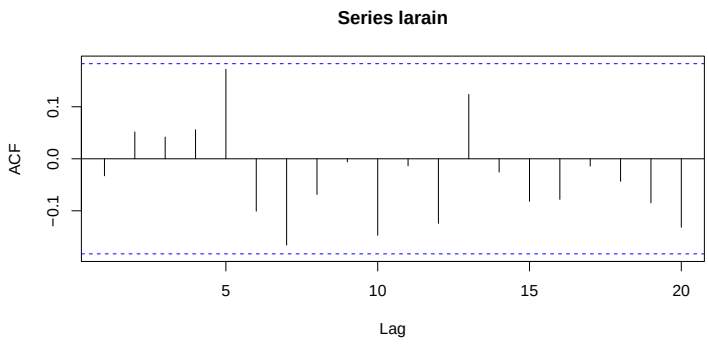
It can be shown that if Y_t is white noise (i.e., a purely random process with no autocorrelation) then the sample ACF is approximately normally distributed as follows:

$$\hat{\rho}_k \to N\left(0,\ \sigma^2_{\hat{\rho}_k} = \tfrac{1}{n}\right) \text{ when } n \text{ is large.}$$

This implies that 95% of $\hat{\rho}_k$ values will lie within $\pm 1.96\frac{1}{\sqrt{n}}$.

A **correlogram** is a plot of the ACF over a range of different k values (lags) with $\pm 1.96\frac{1}{\sqrt{n}}$ bands which is equivalent to a series of hypothesis tests. If the series is white noise then the ACF values should be within the bands with no obvious pattern (however, we would expect one in twenty to be outside of these bands by chance).

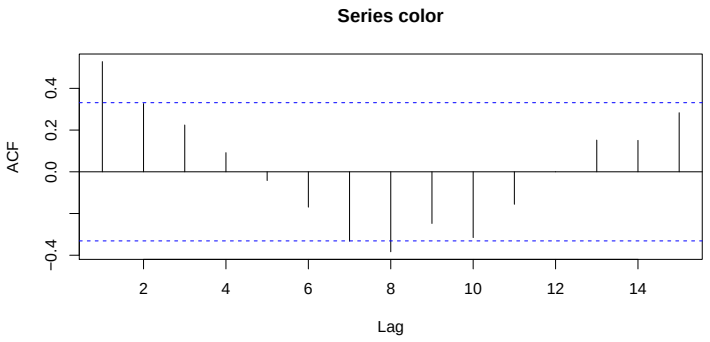
Example 4.1. *ACF for LA Rainfall Data*




```
dev.new(width=8, height=4)
data(larain)
acf(larain)
```

The above correlogram shows that the LA rainfall series is approximately white noise, i.e., no pattern.

Example 4.2. *ACF for Colour Data*




```
dev.new(width=8, height=4)
data(color)
acf(color)
```

Here the lag-1 autocorrelation is clearly significant, i.e., the values in one batch are related to values in a previous batch. To a lesser extent the lag-8 autocorrelation is significant but this may just be due to random chance.

Important: When modelling time series data, we typically aim to reduce the series to a white noise process. This means that we have captured the correlation structure observed in the data with the chosen model and all that is left (i.e., the set of residuals) is purely random noise.

5 Differencing

Often some simple transformations of the series can eliminate correlations and trend in the data. One of the simplest transformations, which is a core technique in time series analysis, is **differencing**:

$$\nabla Y_t = Y_t - Y_{t-1}.$$

Rather than working with the original series, we work with the differenced series:

$$\begin{aligned}\nabla Y_2 &= Y_2 - Y_1\\ \nabla Y_3 &= Y_3 - Y_2\\ &\vdots\\ \nabla Y_n &= Y_n - Y_{n-1}.\end{aligned}$$

Clearly, the differenced series has one less value than the original series.

A **second order difference** is the difference of a differenced series:

$$\begin{aligned}\nabla^2 Y_t = \nabla(\nabla Y_t) &= \nabla(Y_t - Y_{t-1})\\ &= \nabla Y_t - \nabla Y_{t-1}\\ &= Y_t - Y_{t-1} - (Y_{t-1} - Y_{t-2})\\ &= Y_t - 2Y_{t-1} + Y_{t-2}.\end{aligned}$$

Uses of differencing:

- Differencing can eliminate linear trend and second order differencing can eliminate quadratic trend.
- Differencing can convert a non-stationary series into a stationary series. For example, we have seen that a random walk, $Y_t = Y_{t-1} + e_t$ is non-stationary; however, subtracting Y_{t-1} from both sides gives $Y_t - Y_{t-1} = \nabla Y_t = e_t$, i.e., white noise, which is stationary.
- Often we can eliminate seasonal (say, monthly) trend by carrying out **seasonal differences**:

  $$\nabla_{12} Y_t = Y_t - Y_{t-12}.$$

  Note: This is *not* the same as a 12th order difference $\nabla^{12} Y_t$ which is a succession of 12 differences.

5.1 Backshift Operator

The above transformations can, alternatively be summarised using the **backshift operator**

$$B^m Y_t = Y_{t-m}$$

as follows

$$\begin{aligned}\text{1st difference:} &\quad (1-B)Y_t\\ \text{2nd difference:} &\quad (1-B)^2 Y_t\\ &\vdots\\ d\text{th difference:} &\quad (1-B)^d Y_t\end{aligned}$$

and

$$\begin{aligned}\text{1st seasonal difference:} &\quad (1-B^{12})Y_t\\ \text{2nd seasonal difference:} &\quad (1-B^{12})^2 Y_t\\ &\vdots\\ d\text{th seasonal difference:} &\quad (1-B^{12})^d Y_t\end{aligned}$$

5.2 R Code

The `diff` function in `R` carries out differencing, e.g., `diff(x,lag=1,diff=1)` carries out a first difference, `diff(x,lag=1,diff=2)` carries out a second difference, and `diff(x,lag=12,diff=1)` carries out a seasonal difference (for monthly data).