Contents

Chapter 2

Stationary Processes

Some of the material in this chapter can be found in Fuller's *Introduction into Statistical Time Series (2nd Edition).*This material is just to provide a theretical foundation for what we constantly assum in most practical time series applications

2.1 Background and Notation

1. Let (Ω, A, P) be the probability space (Ω is the sample space, A is the sigma algebra or sigma-field satisfying complementarity and union, and P is the probability measure over A and hence Ω).

2. Let T be an index set (time) (typically the set of integers from $\pm\infty$).

3. A random variable Y is a real valued function (a mapping) defined on Ω such that the set $\{\omega : Y(\omega) \leq y\}$ is a member of A for every real number y.

4. The function $F_Y(y) = P\{\omega : Y(\omega) \leq y\}$ is called the **distribution function**.

5. A real valued time series (or stochastic process) is a real valued function $Y(t, \omega)$ defined on $T \times \Omega$ such that for each fixed $t, Y(t, \omega)$ is a random variable on (Ω, A, P). We write function $Y(t, \omega)$ as $Y_t(\omega)$ or more simply Y_t.

6. Outcome ω is assigned a real number $Y(\omega)$ and therefore for fixed $\omega \Rightarrow Y(t, \omega)$ is a real valued function of t (a realization).

7. While for applied purposes this distinction is irrelevant, we should note that if we plot data over time it is interpreted as for fixed ω.

8. The collection of all possible realizations is called the **ensemble of functions**.

9. Ideally we would want all Ω but the question is can we learn all the relevant information by having only a finite set of observations (will depend on the memory of the process)

2.1.1 Joint Distribution Function

For a finite set (say n observations) random variables the **joint cumulative function**:

$$F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_n}}(y_{t_1},y_{t_2},\ldots,y_{t_n}) = P\{\omega : Y(t_1,\omega) \leq y_{t_1},\ldots,Y(t_n,\omega) \leq y_{t_n}\}$$

2.1.2 Strict Stationarity

$$F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_n}}(y_{t_1},y_{t_2},\ldots,y_{t_n}) = F_{Y_{t_1+h},Y_{t_2+h},\ldots,Y_{t_n+h}}(y_{t_1},y_{t_2},\ldots,y_{t_n})$$

The distribution function of the random variable Y is the same at every point on the index set. The joint distribution depends only on the **distance** between the elements in the index set

2.1.3 Implications for Strict Stationarity

1. If Y_t is strictly stationary with $E[\ |Y_t|\ ] < \infty$ (finite first moment) , the expected value of Y_t is **constant** for all t since the distribution function is the same for all t.
2. The second moment is also constant if $E[Y_t^2] < \infty$ (finite second moment)

2.1.4 Weak Stationarity

This is typically what is assumed in most time series applications what Harvey calls stationarity. If Y_t satisfies:

1. (a) Expected value of Y_t is constant for all t

   $$E[Y_t] = \mu \qquad \text{say (this could be zero)}$$

   (b) **Autocovariance matrix** $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ is the same as $(Y_{t_1+h}, \ldots, Y_{t_n+h})$ for all t and h

   $$Cov[Y_t, Y_{t+h}] = E[(Y_t - \mu)(Y_{t+h} - \mu)] = \gamma(h) \qquad h = 1, 2, \ldots \quad (2.1)$$

2. The autocovariance depends only on the distance h and there is no time subscript on the function indicating it is time-invariant.
3. If a time series is weakly stationary and normally distributed then it is strictly stationary.
4. Weakly stationary series need not be strictly stationary, since it is possible for moments higher than 2 to depend on time t.
5. A process y_t is said to be Gaussian if the joint density is normally and weakly stationary. Moreover, since normal distributions are completely characterized by their first two moments, Gaussian processes are also strictly stationary.

2.1.5 Some Examples

White Noise

The simplest stationary stochastic process ϵ_t

1. $E[\epsilon_t] = 0$ mean zero
2. $$E[\epsilon_t \epsilon_{t+h}] = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$ constant variance $\sigma^2 < \infty$

Unit roots are not stationary

$$Y_t = Y_{t-1} + \epsilon_t \tag{2.2}$$

where ϵ_t is white noise. We may write this as a **partial sum**:

$$Y_t = \sum_{i=0}^{t} \epsilon_{t-i} \quad Y_0 = 0 \tag{2.3}$$

and given the $iid(0, \sigma^2)$ assumption

$$V[Y_t] = t\ \sigma^2 \tag{2.4}$$

and is clearly an *increasing function* of t and is *diverging* as $t \rightarrow \infty$

Trends are not stationary

$$Y_t = \alpha + \beta\ t + \epsilon_t,$$

since the population means $E[Y_t]$ are changing at each point in time.

2.1.6 Autocorrelation Functions (ACF)

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

The autocorrelation function is the autocovariance function normalized to be one at $h = 0$.

Results for stationary time series

1. The autocovariance function of a *vector stationary time series* is positive semidefinite. This does not imply than an estimated autocovariance matrix would be positive semidefinite. Heteroskedasticity Consistent Covariance Estimation (HAC) is discussed in Chapter 12.

2. For a single time series: $\mid \rho(h) \mid < 1$
3. The autocovariance and autocorrelation functions are **even** functions:

$$\gamma(h) = \gamma(-h) \quad \rho(h) = \rho(-h)$$

for *real- valued stationary time series.* This is not true for complex numbers.

4. If Y_t and X_t are jointly weak stationary mean zero processes then in general:

$$E[Y_t X_{t+h}] \neq E[X_t Y_{t+h}] \quad \text{for } h \neq 0$$

2.2 Consistency: Estimating the mean, variance, autocovariance and autocorrelations

2.2.1 Ergodic Process

1. If the mean square error (MSE) of the sample mean as an estimator of the population mean approaches zero as the number of observations gets large, the times series is said to be **ergodic** for the mean
2. The idea is to take sample means (time averaging) over the T consecutive observations

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

and obtain **consistent** estimates of μ ($p\lim \bar{Y} = \mu$). We say $\bar{Y}$ is *ergodic* for the mean.

3. Similarly if (MSE) of sample autocovariance as an estimator of the population autocovariance function approaches zero asymtotically, the time series is said to be ergodic for the autocovariance.
4. If the process Y_t is weakly stationary, then Y_t is ergodic for the mean if:

$$\lim_{T\to\infty}\left(\frac{1}{T}\sum_{h=1}^{T}\gamma(h)\right) = 0 \tag{2.5}$$

5. Ergodicity implies that the memory of the process as measured by $\gamma(h) \to 0$ weakens by averaging over time.
6. Absolute summability is a stronger condition that implies ergodicity for the mean:

$$\sum_{h=0}^{\infty} \mid \gamma(h) \mid < \infty \Longrightarrow \gamma(h) \to 0 \text{ as } h \longrightarrow \infty$$

7. To see that covariance stationarity does not necessarily give ergodicity for the mean, consider the process
$$y_t^{(i)} = \mu^{(i)} + \epsilon_t$$
with
$$\mu^{(i)} \sim N(0, \lambda^2) \text{ and } \epsilon_t \text{ Gaussian white noise } (0, \sigma^2)$$
Clearly we have weak stationarity since
$$\begin{aligned} E[y_t^{(i)}] &= E\left[\left(\mu^{(i)} + \epsilon_t\right)\right] = E[\mu^{(i)}] + E[\epsilon_t] = 0 \\ \gamma_t^i(h) &= E\left[\left(\mu^{(i)} + \epsilon_t\right)\left(\mu^{(i)} + \epsilon_{t-h}\right)\right] = \lambda^2 \quad \forall j \neq 0, t \end{aligned}$$
but
$$\begin{aligned} \bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t^{(i)} = \frac{1}{T}\sum_{t=1}^{T}\left(\mu^{(i)} + \epsilon_t\right) \\ &= \mu^{(i)} + \frac{1}{T}\sum_{t=1}^{T}\epsilon_t = \mu^{(i)} \neq 0 \end{aligned}$$

8. For the statistical buffs weak stationarity plus strong mixing $\Rightarrow$ ergodicity

2.2.2 Statistical Ergodic Theorem (Hannan -Time Series Analysis)

1. While the idea seems pretty obvious keep in mind that we really have a single observation on a vector random variable $\{Y(1,\omega), Y(2,\omega)\ldots Y(T,\omega)\}$.
2. The justification for this averaging is from the statistical ergodic theorem.

Application of the Theorem

1. Sample mean
$$\hat{\mu} = \bar{y} = 1/T\sum Y_t$$
2. Sample Variance
$$\hat{\gamma}(0) = c(0) = 1/T\sum(Y_t - \hat{\mu})^2$$
3. Sample Autocovariance of order h
$$\hat{\gamma}(h) = c(h) = \sum_{t=h+1}^{T}(Y_t - \hat{\mu})(Y_{t-h} - \hat{\mu}) \quad h = 1, \ldots$$
4. Correlogram (ACF)
$$\hat{\rho}(h) = r(h) = \hat{\gamma}(h)/\hat{\gamma}(0) = c(h)/c(0)$$

- When we speak of stationarity we will assume that sufficient conditions hold so that all of these sample moments converge to the corresponding population moments as $t \to \infty$ (i.e. are consistent).

2.3 Two Useful Operators

2.3.1 The Lag Operator L

$$
\begin{aligned}
Ly_t &= y_{t-1} \qquad (2.6)\\
L^2 y_t &= L(Ly_t) = Ly_{t-1} = y_{t-2}\\
L^h y_t &= y_{t-h} \qquad h = 1, \ldots \qquad (L^0 y_t = y_t)
\end{aligned}
$$

- Polynomials in the lag operator can be manipulated like any other algebraic polynomial.

Example

- Consider the following specific infinite moving average process with a coefficient of ϕ^j on the ϵ_{t-j} and $\mid \phi \mid < 1$

$$y_t = \sum_{j=0}^{\infty} (\phi L)^j \epsilon_t \tag{2.7}$$

We regard $\mid L \mid \leq 1$ so that $\mid L\phi \mid < 1 \Rightarrow$ the series $1, \phi L, (\phi L)^2, \ldots$ may be summed as an *infinite geometric progression* and hence equation (2.7) becomes:

$$y_t = \frac{\epsilon_t}{(1 - \phi L)} \tag{2.8}$$

and this may be rewritten as:

$$y_t = \phi y_{t-1} + \epsilon_t \tag{2.9}$$

which is the first order autoregressive process ($AR(1)$).

- We may write out polynomials in the lag operator for example:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$$

which is a p^{th} order polynomial. With some slightly loose notation we have the infinite

$$\phi_\infty(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots$$

Sign Convention

- There is a *sign* convention that may confuse you. Consider the autoregressive equation with $p = 1$ (an $AR(1)$ model)

$$
\begin{aligned}
\phi_1(L) y_t &= \varepsilon_t \qquad (2.10)\\
(1 - \phi_1 L) y_t &= \varepsilon_t\\
y_t &= \phi_1 y_{t-1} + \varepsilon_t
\end{aligned}
$$

We have the negative sign convention in $\phi_p(L)$, so that when we take the lag $y's$ to the right-hand-side the $\phi's$ enter with a "+". The positive convention holds for moving average processes.

$$\theta_q(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \tag{2.11}$$

- So that a pure moving-average process of order 1, can be succinctly written :

$$\begin{aligned} y_t &= \theta_1(L)\varepsilon_t \\ &= \varepsilon_t + \theta_1\varepsilon_{t-1} \end{aligned} \tag{2.12}$$

Inverse of a Polynomial

- Most of the manipulations with standard polynomials can be done with polynomials in the lag operator. So that if

$$\phi_p(L)\psi_q(L) = 1 \tag{2.13}$$

- This implies that $\psi_q(L)$ is the inverse of $\phi_p(L)$

$$\psi_q(L) = \phi_p(L)^{-1} \tag{2.14}$$

We have already seen an application of this result

$$(1 + \phi L + \phi^2 L^2 + \ldots) \times (1 - \phi L) = 1 \tag{2.15}$$

.

2.3.2 The First difference Operator $\Delta = (1 - L)$

$$\begin{aligned} \Delta y_t &= (1 - L)y_t = y_t - y_{t-1} \\ \Delta^2 y_t &= (1 - L)^2 y_t = (1 - 2L + L^2)y_t = y_t - 2y_{t-1} + y_{t-2}. \end{aligned} \tag{2.16}$$