smfd5(13).tex
Day 7. 29.5.2013

VI. TIME SERIES (TS).

1. Stationary processes and autocorrelation

A TS - a sequence of observations indexed by time - may well exhibit, on visual inspection after plotting, a *trend* - a tendency to increase or decrease with time, or *seasonality*, or both. However, the simplest case is where trend and seasonality are absent, and we begin with this. Furthermore, even if they are present, our first task may well be to remove them, by *detrending* and/or *seasonal adjustment*.

Definition. A TS, or stochastic process, is *strictly stationary* if its finite-dimensional distributions are invariant under time-shifts - that is, if for all n, $t_1, \cdots, t_n$ and h, $(X_{t_1}, \cdots, X_{t_n})$ and $(X_{t_1+h}, \cdots, X_{t_n+h})$ have the same distribution. In particular, for a stationary TS:

(i) taking $n = 1$, the marginal distribution of X_t is the same for all t, so the mean of X_t (if it is defined, as we shall assume) is constant, $= \mu$ say, and so is its variance (if defined, as we shall also assume), $= \sigma^2$ say:

$$EX_t = \mu, \qquad varX_t = \sigma^2 \qquad \text{for all } t.$$

(ii) Taking $n = 2$, the distributions of (X_{t_1}, X_{t_2}) is the same as that of (X_{t_1+h}, X_{t_2+h}), and so depends only on the *time-difference* $t_2 - t_1$, called the *lag*. With lag τ, it thus suffices to consider the distribution of $(X_t, X_{t+\tau})$, which depends only on the lag τ, not the time t. In particular, the covariance $cov(X_t, X_{t+\tau})$ is a function of τ only, $\gamma(\tau)$ say:

$$cov(X_t, X_{t+\tau}) = \gamma(\tau) \qquad \text{for all } t$$

(note that $\gamma(0) = varX_t = \sigma^2$, for all t). Similarly for the correlation:

$$corr(X_t, X_{t+\tau}) = \gamma(\tau)/\gamma(0) = \rho(\tau).$$

Definition. The function

$$\rho(\tau) := corr(X_t, X_{t+\tau})$$

is called the *autocorrelation function* of the (strictly) stationary process (X_t).

Note. 1. If X_t is normal (Gaussian), its distribution (that is, the set of its finite-dimensional distributions) is completely determined by its means and covariances (equivalently, variances and correlations), μ and $\gamma(\tau)$ or $\rho(\tau)$. Sometimes, however, we do not want to make the very strong assumption of normality, but only need to specify the distribution of the process as far as its means and covariances/correlations. As these involve only the one- and two-dimensional distributions, they are called *second-order properties*.
2. Since covariance and correlation are commutative – $cov(X, Y) = cov(Y, X)$ and $corr(X, Y) = corr(Y, X)$ –

$$\gamma(-\tau) = \gamma(\tau), \qquad \rho(-\tau) = \rho(\tau).$$

So we can think of the lag just as a time-difference – it does not matter whether we think forwards in time or backwards in time.

Definition. A process (X_t) whose means and variances exist is called *weakly stationary* (covariance stationary, second-order stationary, wide-sense stationary) if its mean EX_t is constant over time and its covariance $cov(X_t, X_{t+\tau})$ depends only on the lag τ and not on the time t. We then use the notation $EX_t = \mu$, $cov(X_t, X_{t+\tau}) = \gamma(\tau)$, $corr(X_t, X_{t+\tau}) = \rho(\tau)$ as above.
Note. 1. A strictly stationary process is always weakly stationary. The converse is false in general but true for the normal (Gaussian) case.
2. For brevity, we now abbreviate 'weakly stationary' to 'stationary'. We will continue to say 'strictly stationary', unless the process is normal (Gaussian), when the strictness is automatic (by above), so can be understood.
White Noise. The simplest possible case of stationarity is $\mu = EX_t = 0$, $\gamma(\tau) = \sigma^2\rho(\tau)$, where $\rho(\tau) = corr(X_t, X_{t+\tau})$ is 1 for $\tau = 0$ and 0 otherwise. Such processes exist in three levels of generality:
(i) no further restriction (distinct X_t uncorrelated, but may be dependent);
(ii) distinct X_t independent;
(iii) (X_t) normal – so distinct X_t are independent, because uncorrelated.

The term *white noise* (WN) is used for some/all such cases, or $WN(\sigma^2)$.
Note. The term shows clearly its engineering origins. The word 'noise' derives from radio engineering (for instance, spontaneous thermal fluctuations, or 'shot noise', in thermionic valves), and telephone engineering. It is also used in telecommunications, where the 'noise' – random error or disturbances – may be visual rather than aural (recall that optical fibres are used nowadays in cables for long-distance communication, with photons playing the role of

electrons in the traditional telephone cables). The term 'white' is by analogy with white rather than coloured light. In the language of spectral theory, white noise has a *flat spectrum* (a 'uniform mixture' of frequencies - just as white light is a mixture of the colours of the rainbow).
3. We shall use definition (ii) of white noise for convenience. Independence will allow us to use LLN and CLT.
4. White noise is specific to *discrete* time. A process with correlation

$$\rho(\tau) = \begin{cases} 1 & (\tau = 0) \\ 0 & (\tau \neq 0) \end{cases}$$

is realistic in discrete time (such as the white noise above), but would be pathological (and physically unrealisable) in *continuous* time, because of the discontinuity in the correlation function. However, the process corresponding to the integrated version of white noise in continuous time does exist and is extremely important: *Brownian motion* (SP, Ch. III).
5. The $\rho(.)$ above (1 at 0, 0 elsewhere) is the 'Dirac delta'. To treat it mathematically, we need Functional Analysis – *generalised functions*, or *Schwartz distributions* (Laurent SCHWARTZ (1915-2002) in 1948). This can then be applied to develop *white noise analysis*, an extensive and useful field.

2. The correlogram

If $(X_1, \cdots, X_n)$ is a section of a TS observed over a finite time-interval,

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

is the *sample mean*. If $\mu = EX_t$ is the population mean, by LLN

$$\bar{X} \to \mu = EX_t \qquad (n \to \infty) :$$

$\bar{X}$ is a consistent estimator of $\mu = EX_t$.

The *sample autocorrelation* at *lag* τ is

$$c(\tau), c_\tau := \frac{1}{n}\sum_{1}^{n-\tau}(X_t - \bar{X})(X_{t+\tau} - \bar{X}).$$

Proposition. $c(\tau) \to \gamma(\tau) \qquad (n \to \infty)$.
Proof. Expanding out the brackets in the definition above,

$$c(\tau) = \frac{1}{n}\sum(X_t X_{t+\tau}) - \bar{X}.\frac{1}{n}\sum X_{t+\tau} - \bar{X}.\frac{1}{n}\sum X_t + \frac{(n-\tau)}{n}(\bar{X})^2.$$

By LLN (applied to stationary, rather than independent, sequences – the Birkhoff-Khintchine Ergodic Theorem, which we quote),

$$\frac{1}{n}\sum X_t X_{t+\tau} \to E(X_t X_{t+\tau}), \quad \frac{1}{n}\sum X_{t+\tau} \to EX_{t+\tau} = \mu,$$

$$\frac{1}{n}\sum X_t \to EX_0 = \mu.$$

So

$$c(\tau) \to E(X_t X_{t+\tau}) - \mu^2 - \mu^2 + \mu^2 = E(X_t X_{t+\tau}) - \mu^2.$$

But

$$\gamma(\tau) = E[(X_{t+\tau} - \mu)(X_t - \mu)] = E(X_{t+\tau}X_t) - \mu EX_t - \mu EX_{t+\tau} + \mu^2$$

$$= E(X_{t+\tau}X_t) - \mu^2 - \mu^2 + \mu^2 = E(X_t X_{t+\tau}) - \mu^2,$$

the limit obtained above. So $c(\tau) \to \gamma(\tau)$. //

Note. 1. Thus the sample autocovariance $c(\tau)$ is a consistent estimator of the population autocovariance $\gamma(\tau)$.
2. To help remember this: in Statistics we use Roman letters for sample quantities, Greek letters for population quantities or parameters.
Definition. The *sample autocorrelation* at *lag* τ is

$$r_\tau, r(\tau) := \rho(\tau)/c(0).$$

Corollary. $r(\tau) \to c(\tau) \qquad (n \to \infty)$:
the sample autocorrelation $r(\tau)$ is a consistent estimator of the population autocorrelation $\rho(\tau)$.
Definition. A plot of $r(\tau)$ against τ is called the *correlogram*.

The correlogram is the principal tool for dealing with Time Series in the *time domain* - that is, looking at time-dependence directly. This is in contrast to the *frequency domain* (spectral properties and Fourier analysis).
Large-Sample Behaviour.

The simplest case is where (X_t) is itself white noise, WN. Then $\rho(0) = 1$, $\rho(\tau) = 0$ for all non-zero lags τ, by definition of WN, and $r(0) = c(0)/c(0) = 1$ also. For τ non-zero and n large, one expects $r(\tau)$ to be small (as $r(\tau) \to c(\tau) = 0$) – but how small? It was shown by M. S. BARTLETT in 1946 (see

e.g. Diggle [D] 2.5) that for large n and τ non-zero, $r(\tau) \sim N(0, 1/n)$. So as $\sqrt{n}r(\tau) \sim \Phi := N(0,1)$, the *standard normal* distribution, which takes values $> 1.96 \sim 2$ in modulus with probability 5%, only values of $r(\tau)$ with

$$|r(\tau)| \geq 1.96/\sqrt{n} \sim 2/\sqrt{n}$$

differ significantly from zero.

3. Autoregressive processes, AR(1)

Recall that in a linear regression model, the dependent variable Y depends in a linear way on an independent variable X (or $X_1, X_2, X_3, \cdots$, or $X, X^2, X^3, \cdots$), with an error structure or noise process also present.

In a TS model, the current value X_t depends in a linear way on the previous value X_{t-1} (or on the p previous values $X_{t-1}, X_{t-2}, \cdots, X_{t-p}$), again plus noise.

First-order case: $AR(1)$. Suppose that our model is

$$X_t = \phi X_{t-1} + m + \epsilon_t, \qquad ((\epsilon_t) \quad WN)$$

for t an integer (positive, negative or zero), where (ϵ_t) is a white noise process $WN(\sigma^2)$. Take means and use $EX_t = \mu$, $E\epsilon_t = 0$:

$$\mu = \phi\mu + m.$$

So if $\phi \neq 1$,

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

and if $\phi = 1$, then $m = 0$.

For simplicity, centre at means:

$$\begin{aligned} X_t - \mu &= \phi(X_{t-1} - \mu) + m - \mu + \phi\mu + \epsilon_t \\ &= \phi(X_{t-1} - \mu) + m - \mu(1-\phi) + \epsilon_t \\ &= \phi(X_{t-1} - \mu) + \epsilon_t, \end{aligned}$$

by above. Centring at means (i.e. replacing $X_t - \mu$ by X_t) for simplicity, we have

$$X_t = \phi X_{t-1} + \epsilon_t, \qquad (*)$$

a simpler model, with all means zero. This is called an *autoregressive model of order one*, $AR(1)$. For, it has the form of a regression model, with X_{t-1} as

the 'dependent variable' and X_t as the 'independent variable': X_t is regressed on the previous X-value (earlier in time), so the *process* (X_t) is *regressed on itself* (Greek: autos = self).

Using $(*)$ recursively,

$$\begin{aligned} X_t &= \phi(\phi X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi^2 X_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\ &= \cdots \\ &= \phi^n X_{t-n} + \textstyle\sum_{i=0}^{n-1} \phi^i \epsilon_{t-1}. \end{aligned}$$

If $|\phi| < 1$, this suggests that the first term on the RHS $\to 0$ as $n \to \infty$, giving $X_t = \sum_0^\infty \phi^i \epsilon_{t-i}$. This is true, provided we interpret the convergence of the infinite series on RHS suitably. We have

$$E[(X_t - \sum_1^{n-1} \phi^i \epsilon_{t-i})^2] = E[(\phi^n X_{t-n})^2] = \phi^{2n} E[X_{t-n}^2] = \phi^{2n}\gamma_0,$$

where $\gamma_0 = varX_t$ for all t. Since $|\phi| < 1$, $\phi^{2n} \to 0$ as $n \to \infty$, so RHS $\to 0$ as $n \to \infty$. So LHS $\to 0$ as $n \to \infty$. This says that

$$\textstyle\sum_0^n \phi^i \epsilon_{t-i} \to X_t \qquad (n \to \infty),$$

or

$$\textstyle\sum_0^\infty \phi^i \epsilon_{t-i} = X_t,$$

in mean square (or, in L_2).

Interpreting convergence in this mean-square sense,

$$X_t = \textstyle\sum_0^\infty \phi^i \epsilon_{t-i} \qquad (**)$$

expresses X_t on LHS as a *weighted sum* of $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots$ on RHS. This weighted sum resembles an *average* (although the weights sum to $1/(1-\phi)$, not 1 as is usual for an average), and the set $(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots)$ of white-noise variables being averaged over *moves* with t; there are *infinitely many* of them. Hence $(**)$ is called the *infinite moving-average representation* of the $AR(1)$ process $(*)$. Note that the further we go back in time, the more the ϵ_{t-i} are down-weighted by the geometrically decreasing weights ϕ^i.

Autocovariance of $AR(1)$. Since ϵ_{t+1} is independent of (or, using the weaker definition of white noise, uncorrelated with) $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \cdots$, it is independent of (or uncorrelated with) the linear combination $X_t = \sum_0^\infty \phi^i \epsilon_{t-i}$ of

them. So ϵ_{t+1} is uncorrelated with $X_t, X_{t-1}, \cdots$. This says that X_s and ϵ_t are uncorrelated for $s < t$. Since all means are zero:

$$E(X_s \epsilon_t) = 0 \qquad (s < t).$$

Square both sides of $(*)$ and take expectations:

$$E[X_t^2] = \phi^2 E[X_{t-1}^2] + 2\phi E[X_{t-1}\epsilon_{t-1}] + E[\epsilon_t^2].$$

The second term on RHS is zero by above; $E[X_t^2] = varX_t = \gamma_0$ for all t, and $E[\epsilon_t^2] = var\epsilon_t = \sigma^2$ for all t. So

$$\gamma_0 = \phi^2 \gamma_0 + \sigma^2 : \qquad \gamma_0 = \sigma^2/(1-\phi^2),$$

identifying γ_0 in terms of the WN variance σ^2 and the weight ϕ.

Multiply $(*)$ by $X_{t-\tau}$ $(\tau \geq 1)$ and take expectations:

$$\gamma_\tau = \phi \gamma_{\tau-1}$$

(since ϵ_t on RHS is uncorrelated with $X_{t-\tau}$). Using this repeatedly,

$$\gamma_\tau = \phi\gamma_{\tau-1} = \phi^2\gamma_{\tau-2} = \cdots = \phi^\tau \gamma_0 = \phi^\tau \sigma^2/(1-\phi^2) :$$

$$\gamma_\tau = \sigma^2.\phi^\tau/(1-\phi^2) \qquad (\tau \geq 0),$$

giving the autocovariance of an $AR(1)$ process as geometrically decreasing. Passing to the autocorrelation $\rho_\tau = \gamma_\tau/\gamma_0$: $\rho_\tau = \phi^\tau$ for $\tau \geq 0$). Note that $\rho_\tau = \rho_{-\tau}$ (since two random variables have the same covariance and correlation either way round), so we can re-write this as

$$\rho_\tau = \phi^{|\tau|}.$$

Recall $|\phi| < 1$ here. Two cases are worth distinguishing.

Case 1: $0 \leq \phi < 1$. Here the graph of ρ_τ is a geometric series with non-negative common ratio. Since the sample autocorrelation r_τ is an approximation to ρ_τ, the correlogram (graph of r_τ) is an approximation to this. Successive values of X_t are positively correlated: positive values of X_t tend to be succeeded by positive values, and similarly negative by negative.

Case 2: $-1 < \phi < 0$. Here the graph is again a geometric series, but one that *oscillates in sign*, as well as damping down geometrically. Successive values of X_t are negatively correlated: positive values tend to be succeeded

by negative values, and vice versa.

To summarise: the signature of an $AR(1)$ process is a correlogram that looks like an approximation to a geometric series.

The Lag Operator.

Before proceeding, we introduce some useful notation and terminology. The *lag operator*, or *backward shift operator*, operates on sequences by shifting the index back in time by one. We write it as B:

$$BX_t = X_{t-1},$$

(though L - L for lag - is also used). Repeating this, B^2 shifts back in time by two, $B^2X_t = X_{t-2}$, and generally

$$B^nX_t = X_{t-n} \qquad (n = 0, 1, 2, \cdots)$$

($B^0 = I$ is the identity operator: $B^0X_t = IX_t = X_t$).

We can re-write $(*)$ in this notation as

$$X_t = \phi BX_t + \epsilon_t : \qquad (1 - \phi B)X_t = \epsilon_t.$$

Formally, this suggests

$$\begin{aligned} X_t = (1 - \phi B)^{-1}\epsilon_t &= (1 + \phi B + \phi^2B^2 + \cdots + \phi^iB^i + \cdots)\epsilon_t \\ &= 1 + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots + \phi^i\epsilon_{t-i} + \cdots \\ &= \textstyle\sum_0^\infty \phi^i\epsilon_{t-i}, \end{aligned}$$

which is $(**)$ as above, *provided* that the operator equation

$$(1 - \phi B)^{-1} = \textstyle\sum_{i=0}^\infty \phi^iB^i$$

makes sense. It does make sense, with convergence on the RHS interpreted in the *mean-square sense* as above, if $|\phi| < 1$.

4. General autoregressive processes, AR(p).

Again working with the zero-mean case for simplicity, the extension of the above to p parameters is the model

$$X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + \cdots + \phi_pX_{t-p} + \epsilon_t, \qquad (*)$$

with (ϵ_t) WN as before. Since $X_{t-i} = B^iX_t$, we may re-write this as

$$X_t - \phi_1BX_t - \cdots - \phi_pB^pX_t = \epsilon_t.$$

Write

$$\phi(\lambda) := 1 - \phi_1\lambda - \cdots - \phi_p\lambda^p$$

for the pth order polynomial here. Then formally,

$$\phi(B)X_t = \epsilon_t : \qquad X_t = \phi(B)^{-1}\epsilon_t,$$

so if we expand $1/\phi(\lambda)$ in a power series as

$$1/\phi(\lambda) \equiv 1 + \beta_1\lambda + \cdots + \beta_n\lambda^n + \cdots,$$

$$X_t = \sum_{i=0}^{\infty}\beta_i B^i\epsilon_t = \sum_0^{\infty}\beta_i\epsilon_{t-i}.$$

This is the analogue of $X_t = \sum_0^{\infty}\phi^i\epsilon_{t-i}$ for $AR(1)$, and shows that X_t can again be represented as an infinite moving-average process – or *linear process* (X_t is an (infinite) *linear combination* of the ϵ_{t-i}).

Multiply $(*)$ through by X_{t-k} and take expectations. Since $E[X_{t-k}X_{t-i}] = \rho(|k-i|) = \rho(k-i)$, this gives

$$\rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p) \qquad (k>0). \qquad (YW)$$

These are the *Yule-Walker equations*, due to G. Udny YULE (1871-1951) in 1926 and Sir Gilbert WALKER (1868-1958) in 1931.

The Yule-Walker equations (YW) have the form of a *difference equation* of *order* p. The *characteristic polynomial* of this difference equation is

$$\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p = 0,$$

which by above is

$$\phi(1/\lambda) = 0.$$

If $\lambda_1, \cdots, \lambda_p$ are the roots of this characteristic polynomial, the trial solution $\rho(k) = \lambda^k$ is a solution if and only if λ is one of the roots λ_i. Since the equation is linear,

$$\rho(k) = c_1\lambda_1^k + \cdots + c_p\lambda_p^k$$

(for $k \geq 0$, and use $\rho(-k) = \rho(k)$ for $k < 0$) is a solution for all choices of constants $c_1, \cdots, c_p$. This is the *general solution* of (YW) if all the roots λ_i are distinct, with appropriate modifications for repeated roots (if $\lambda_1 = \lambda_2$, use $c_1\lambda_1^k + c_2k\lambda_1^k$, etc.).

Now $|\rho(k)| \leq 1$ for all k (as $\rho(.)$ is a correlation coefficient), and this is *only possible* if

$$|\lambda_i| \leq 1 \qquad (i = 1, \cdots, p)$$

– that is, all the roots lie inside (or on) the unit circle. This happens (as our polynomial is $\phi(1/\lambda)$) if and only if *all the roots of the polynomial* $\phi(\lambda)$ *lie outside (or on) the unit circle.* Then $|\rho(k)| \leq 1$ for all k, and when there are no roots of unit modulus, also $\rho(k) \to 0$ as $k \to \infty$ – that is, the influence of the remote past tends to zero, as it should. We shall see below that this is also the condition for the $AR(p)$ process above to be *stationary.*
Example of an $AR(2)$ *process.*

$$X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + \epsilon_t, \qquad (\epsilon_t) \quad WN. \tag{1}$$

Moving-average representation. Let the infinite moving-average representation of (X_t) be

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}. \tag{2}$$

Substitute (2) into (1):

$$\begin{aligned} \sum_0^{\infty} \psi_i \epsilon_{t-i} &= \frac{1}{3}\sum_0^{\infty} \psi_i \epsilon_{t-i-1} + \frac{2}{9}\sum_0^{\infty} \psi_i \epsilon_{t-2-i} + \epsilon_t \\ &= \frac{1}{3}\sum_1^{\infty} \psi_{i-1} \epsilon_{t-i} + \frac{2}{9}\sum_2^{\infty} \psi_{i-2} \epsilon_{t-i} + \epsilon_t. \end{aligned}$$

Equate coefficients of ϵ_{t-i}:
$i = 0$ gives $\psi_0 = 1$; $i = 1$ gives $\psi_1 = \frac{1}{3}\psi_0 = 1/3$; $i \geq 2$ gives

$$\psi_i = \frac{1}{3}\psi_{i-1} + \frac{2}{9}\psi_{i-2}.$$

This is again a difference equation, which we solve as above. The characteristic polynomial is

$$\lambda^2 - \frac{1}{3}\lambda - \frac{2}{9} = 0, \qquad \text{or} \qquad (\lambda - \frac{2}{3})(\lambda + \frac{1}{3}) = 0,$$

with roots $\lambda_1 = 2/3$ and $\lambda_2 = -l/3$. The general solution of the difference equation is thus $\psi_i = c_1\lambda_1^i + c_2\lambda_2^i = c_1(2/3)^i + c_2(-1/3)^i$. We can find c_1, c_2 from the values of ψ_0, ψ_1, found above:
$i = 0$ gives $c_1 + c_2 = 0$, or $c_2 = 1 - c_1$.
$i = 1$ gives $c_1.(2/3)+(1-c_1)(-1/3) = \psi_1 = 1/3$: $2c_1-(1-c_1) = 1$: $c_1 = 2/3$, $c_2 = 1/3$. So

$$\psi_i = \frac{2}{3}(\frac{2}{3})^i + \frac{1}{3}(\frac{-1}{3})^i = (\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1},$$

and

$$X_t = \sum_0^\infty [(\frac{2}{3})^{i+1} - (\frac{-1}{3})^{i+1}]\epsilon_{t-i},$$

giving the moving-average representation, as required.
Autocovariance. Recall the Yule-Walker equations

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2)$$

for $AR(2)$. As before,

$$\rho(k) = a\lambda_1^k + b\lambda_2^k$$

for some constants a, b. Taking $k = 0$ and using $\rho(0) = 1$ gives $a + b = 1$: $b = 1 - a$. So here,

$$\rho(k) = a(2/3)^k + (1-a)(-1/3)^k.$$

Taking $k = 1$ in the Yule-Walker equations gives

$$\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1),$$

which as $\rho(0) = 1$ and $\rho(-1) = \rho(1)$ gives

$$\rho(1) = \phi_1/(1-\phi_2).$$

As here $\phi_1 = 1/3$ and $\phi_2 = 2/9$, this gives $\rho(1) = 3/7$. We can now use this and the above expression for $\rho(k)$ to find a: taking $k = 1$ and equating,

$$\rho(1) = 3/7 = a.(2/3) + (1-a).(-1/3).$$

That is,

$$(\frac{3}{7} + \frac{1}{3}) = a.(\frac{2}{3} + \frac{1}{3}) = a :$$

$a = (9+7)/21 = 16/21$. Thus

$$\rho(k) = \frac{16}{21}(\frac{2}{3})^k + \frac{5}{21}(\frac{-1}{3})^k.$$

Note. For large k, the first term dominates, and

$$\rho^k \sim \frac{16}{21}.(\frac{2}{3})^k \qquad (k \to \infty).$$

AR(p) processes (continued). We return to the general case. Just as in the $AR(2)$ example above, if the $AR(p)$ process has a moving-average representation

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

then if $\sigma^2 = var\epsilon_t$,

$$varX_t = \sigma^2 . \sum_{i=0}^{\infty} \psi_i^2.$$

The condition

$$\sum_{i=0}^{\infty} \psi_i^2 < \infty$$

(in words: (ψ_i) is *square-summable*, or is *in* L_2) is necessary and sufficient for
(i) $varX_t < \infty$;
(ii) the series $\sum \psi_i \epsilon_{t-i}$ in the moving-average representation to be convergent in mean square – or, in L_2.
So for convergence in L_2, $\sum \psi_i^2 < \infty$ is the necessary and sufficient condition (NASC) for the moving-average representation of X_t to *exist*. Since $\sum \psi_i \epsilon_{t-i}$ is (when convergent) stationary (because (ϵ_t) is stationary: if $\sum \psi_i^2 < \infty$, then X_t is stationary. The converse is also true; see Section 5 below.

5. Condition for stationarity
We return to the general case. Just as in the $AR(2)$ example above, if the $AR(p)$ process has a moving-average representation

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

then if $\sigma^2 = var\epsilon_t$,

$$varX_t = \sigma^2 \sum_{i=0}^{\infty} \phi_i^2.$$

The condition

$$\sum_{i=0}^{\infty} \phi_i^2 < \infty$$

((ϕ_i) is *square-summable*, or is *in* L_2) is necessary and sufficient for
(i) $varX_t < \infty$;
(ii) the series $\sum \phi_i \epsilon_{t-i}$ in the moving-average representation to be convergent in mean square – or, in L_2.