smfl6.tex
Lecture 6. 28.1.2011

II. 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:

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

say (note that $\rho(0) = 1$).

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 the *second-order properties* of the TS or stochastic process.
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 (Gaussian) - so distinct X_t are independent, because uncorrelated.

The term *white noise* (WN) is used for some/all such cases, or $WN(\sigma^2)$ if the variance σ^2 needs mention.
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).

2. THE CORRELOGRAM

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

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

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

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

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),

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

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

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

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

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: it is traditional in Statistics to use Roman letters for sample quantities, and Greek letters for the corresponding 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?