2 Models of stationary processes

2.1 Purely indeterministic processes

Suppose $\{X_t\}$ is a second order stationary process, with mean 0. Its **autocovariance function** is

$$\gamma_k = \mathbb{E}(X_t X_{t+k}) = \mathrm{cov}(X_t, X_{t+k}), \quad k \in \mathbb{Z}.$$

1. As $\{X_t\}$ is stationary, γ_k does not depend on t.

2. A process is said to be **purely-indeterministic** if the regression of X_t on $X_{t-q}, X_{t-q-1}, \ldots$ has explanatory power tending to 0 as $q \to \infty$. That is, the residual variance tends to $\mathrm{var}(X_t)$.

   An important theorem due to Wold (1938) states that every purely-indeterministic second order stationary process $\{X_t\}$ can be written in the form

   $$X_t = \mu + \theta_0 Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots$$

   where $\{Z_t\}$ is a sequence of uncorrelated random variables.

3. A **Gaussian process** is one for which $X_{t_1}, \ldots, X_{t_n}$ has a joint normal distribution for all $t_1, \ldots, t_n$. No two distinct Gaussian processes have the same autocovariance function.

2.2 ARMA processes

The **autoregressive moving average process**, ARMA(p, q), is defined by

$$X_t - \sum_{r=1}^{p} \phi_r X_{t-r} = \sum_{s=0}^{q} \theta_s \epsilon_{t-s}$$

where again $\{\epsilon_t\}$ is white noise. This process is stationary for appropriate ϕ, θ.

Example 2.1
Consider the **state space model**

$$X_t = \phi X_{t-1} + \epsilon_t\,,$$
$$Y_t = X_t + \eta_t\,.$$

Suppose $\{X_t\}$ is unobserved, $\{Y_t\}$ is observed and $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent white noise sequences. Note that $\{X_t\}$ is AR(1). We can write

$$\begin{aligned}\xi_t &= Y_t - \phi Y_{t-1}\\ &= (X_t + \eta_t) - \phi(X_{t-1} + \eta_{t-1})\\ &= (X_t - \phi X_{t-1}) + (\eta_t - \phi\eta_{t-1})\\ &= \epsilon_t + \eta_t - \phi\eta_{t-1}\end{aligned}$$

Now ξ_t is stationary and $\mathrm{cov}(\xi_t, \xi_{t+k}) = 0$, $k \geq 2$. As such, ξ_t can be modelled as a MA(1) process and $\{Y_t\}$ as ARMA$(1, 1)$.

2.3 ARIMA processes

If the original process $\{Y_t\}$ is not stationary, we can look at the first order difference process

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

or the second order differences

$$X_t = \nabla^2 Y_t = \nabla(\nabla Y)_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

and so on. If we ever find that the differenced process is a stationary process we can look for a ARMA model of that.

The process $\{Y_t\}$ is said to be an **autoregressive integrated moving average process**, ARIMA(p, d, q), if $X_t = \nabla^d Y_t$ is an ARMA(p, q) process.

AR, MA, ARMA and ARIMA processes can be used to model many time series. A key tool in identifying a model is an estimate of the autocovariance function.

2.4 Estimation of the autocovariance function

Suppose we have data $(X_1, \ldots, X_T)$ from a stationary time series. We can estimate

- the mean by $\bar{X} = (1/T)\sum_1^T X_t$,
- the autocovariance by $c_k = \hat{\gamma}_k = (1/T)\sum_{t=k+1}^{T}(X_t - \bar{X})(X_{t-k} - \bar{X})$, and
- the autocorrelation by $r_k = \hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$.

The plot of r_k against k is known as the **correlogram**. If it is known that μ is 0 there is no need to correct for the mean and γ_k can be estimated by

$$\hat{\gamma}_k = (1/T)\textstyle\sum_{t=k+1}^{T} X_t X_{t-k}\,.$$

Notice that in defining $\hat{\gamma}_k$ we divide by T rather than by $(T - k)$. When T is large relative to k it does not much matter which divisor we use. However, for mathematical simplicity and other reasons there are advantages in dividing by T.

Suppose the stationary process $\{X_t\}$ has autocovariance function $\{\gamma_k\}$. Then

$$\mathrm{var}\left(\sum_{t=1}^{T} a_t X_t\right) = \sum_{t=1}^{T}\sum_{s=1}^{T} a_t a_s \,\mathrm{cov}(X_t, X_s) = \sum_{t=1}^{T}\sum_{s=1}^{T} a_t a_s \gamma_{|t-s|} \geq 0.$$

A sequence $\{\gamma_k\}$ for which this holds for every $T \geq 1$ and set of constants $(a_1, \ldots, a_T)$ is called a **nonnegative definite sequence**. The following theorem states that $\{\gamma_k\}$ is a valid autocovariance function if and only if it is nonnegative definite.

THEOREM 2.2 (Blochner) The following are equivalent.

1. There exists a stationary sequence with autocovariance function $\{\gamma_k\}$.
2. $\{\gamma_k\}$ is nonnegative definite.
3. The spectral density function,
$$f(\omega) = \frac{1}{\pi}\sum_{k=-\infty}^{\infty} \gamma_k e^{ik\omega} = \frac{1}{\pi}\gamma_0 + \frac{2}{\pi}\sum_{k=1}^{\infty}\gamma_k \cos(\omega k)\,,$$
is positive if it exists.

Dividing by T rather than by $(T-k)$ in the definition of $\hat{\gamma}_k$

- ensures that $\{\hat{\gamma}_k\}$ is nonnegative definite (and thus that it could be the autocovariance function of a stationary process), and
- can reduce the L^2-error of r_k.

2.5 Identifying a MA(q) process

In a later lecture we consider the problem of identifying an ARMA or ARIMA model for a given time series. A key tool in doing this is the correlogram.

The MA(q) process X_t has $\rho_k = 0$ for all k, $|k| > q$. So a diagnostic for MA(q) is that $|r_k|$ drops to near zero beyond some threshold.

2.6 Identifying an AR(p) process

The AR(p) process has ρ_k decaying exponentially. This can be difficult to recognise in the correlogram. Suppose we have a process X_t which we believe is AR(k) with
$$X_t = \sum_{j=1}^{k} \phi_{j,k} X_{t-j} + \epsilon_t$$
with ϵ_t independent of $X_1, \dots, X_{t-1}$.

Given the data $X_1, \dots, X_T$, the least squares estimates of $(\phi_{1,k}, \dots, \phi_{k,k})$ are obtained by minimizing
$$\frac{1}{T}\sum_{t=k+1}^{T}\left(X_t - \sum_{j=1}^{k}\phi_{j,k}X_{t-j}\right)^2 .$$
This is approximately equivalent to solving equations similar to the Yule-Walker equations,
$$\hat{\gamma}_j = \sum_{\ell=1}^{k}\hat{\phi}_{\ell,k}\hat{\gamma}_{|j-\ell|}, \quad j = 1, \dots, k$$
These can be solved by the **Levinson-Durbin recursion**:

Step 0. $\sigma_0^2 := \hat{\gamma}_0, \quad \hat{\phi}_{1,1} = \hat{\gamma}_1/\hat{\gamma}_0, \quad k := 0$

Step 1. Repeat until $\hat{\phi}_{k,k}$ near 0:
$$k := k+1$$
$$\hat{\phi}_{k,k} := \left(\hat{\gamma}_k - \sum_{j=1}^{k-1}\hat{\phi}_{j,k-1}\hat{\gamma}_{k-j}\right)\Big/\ \sigma_{k-1}^2$$
$$\hat{\phi}_{j,k} := \hat{\phi}_{j,k-1} - \hat{\phi}_{k,k}\hat{\phi}_{k-j,k-1}, \text{ for } j = 1, \dots, k-1$$
$$\sigma_k^2 := \sigma_{k-1}^2(1 - \hat{\phi}_{k,k}^2)$$

We test whether the order k fit is an improvement over the order $k-1$ fit by looking to see if $\hat{\phi}_{k,k}$ is far from zero.

The statistic $\hat{\phi}_{k,k}$ is called the kth **sample partial autocorrelation coefficient** (PACF). If the process X_t is genuinely AR(p) then the population PACF, $\phi_{k,k}$, is exactly zero for all $k > p$. Thus a diagnostic for AR(p) is that the sample PACFs are close to zero for $k > p$.

2.7 Distributions of the ACF and PACF

Both the sample ACF and PACF are approximately normally distributed about their population values, and have standard deviation of about $1/\sqrt{T}$, where T is the length of the series. A rule of thumb for negligibility of ρ_k (and similarly for $\phi_{k,k}$) is that r_k (similarly $\hat{\phi}_{k,k}$) should lie between $\pm 2/\sqrt{T}$. (2 is an approximation to 1.96. Recall that if $Z_1, \dots, Z_n \sim N(\mu, 1)$, a test of size 0.05 of the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ rejects H_0 if and only if $\bar{Z}$ lies outside $\pm 1.96/\sqrt{n}$).

Care is needed in applying this rule of thumb. It is important to realize that the sample autocorrelations, $r_1, r_2, \dots$, (and sample partial autocorrelations, $\hat{\phi}_{1,1}, \hat{\phi}_{2,2}, \dots$) are not independently distributed. The probability that any one r_k should lie outside $\pm 2/\sqrt{T}$ depends on the values of the other r_k.

A 'portmanteau' test of white noise (due to Box & Pierce and Ljung & Box) can be based on the fact that approximately
$$Q'_m = T(T+2)\sum_{k=1}^{m}(T-k)^{-1}r_k^2 \sim \chi_m^2 .$$
The sensitivity of the test to departure from white noise depends on the choice of m. If the true model is ARMA(p,q) then greatest power is obtained (rejection of the white noise hypothesis is most probable) when m is about $p+q$.