Spring 2015 Statistics 153 (Time Series) : Lecture Eleven

Aditya Guntuboyina

24 February 2015

The ARMA models provide a reasonably versatile collection for modelling stationary time series data. We shall now study how to choose and fit an approprite ARMA model to a given data set. This is done in two parts:

1. How to choose the order of the ARMA model i.e., how to choose p and q?
2. After choosing p and q, how to fit an ARMA(p, q) model to the data i.e., how to estimate the parameters $\phi_1, \dots, \phi_p$ and $\theta_1, \dots, \theta_q$ and σ^2.

For choosing the ARMA order, the main idea is to look at the sample autocorrelation function (or the correlogram) of the data set and find the ARMA model whose acf best resembles this sample autocorrelation function. The important thing to note here is that the sample autocorrelation function is a function of the data and is thus subject to noise. In other words, for example, even if the data is from a true AR(1) model, the sample autocorrelation function would not exactly equal the acf of AR(1). The hope is that it will be close to the AR(1) however especially if the sample size is large.

1 Approximate Distribution of the Sample Autocorrelations

It is important to understand the distribution of the sample autocorrelations and to know how close they get to the true acf. Recall that the sample autocorrelation function for a given data set $x_1, \dots, x_n$ is defined by

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \qquad \text{for } k = 1, 2, \dots$$

and $\bar{x} = (x_1 + \cdots + x_n)/n$.

For a not too small class of models which includes ARMA models with i.i.d noise $\{Z_t\}$, it can be shown that, for every fixed value of k, the random vector $(r_1, \dots, r_k)$ is approximately distibuted according to the multivariate normal distribution with mean $(\rho_X(1), \dots, \rho_X(k))$ and covariance matrix W/n where the (i, j)th entry of the matrix W is denoted by w_{ij} and is given by the following formula (known as Bartlett's formula):

$$\sum_{m=1}^{\infty} \left(\rho_X(m+i) + \rho_X(m-i) - 2\rho_X(i)\rho_X(m)\right) \left(\rho_X(m+j) + \rho_X(m-j) - 2\rho_X(j)\rho_X(m)\right).$$

In particular the approximate distribution of the ith sample autocorrelation r_i is $N(\rho_X(i), w_{ii}/n)$. Also the approximate correlation between r_i and r_j equals

$$\text{corr}(r_i, r_j) = \frac{\text{cov}(r_i, r_j)}{\sqrt{\text{var}(r_i)\text{var}(r_j)}} \approx \frac{w_{ij}/n}{\sqrt{w_{ii}/n \; w_{jj}/n}} = \frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}}.$$

Example 1.1 (i.i.d Noise)**.** *Suppose $\{X_t\}$ is i.i.d with mean zero and variance σ^2. Then the acf $\rho_X(h)$ equals 1 if $h = 0$ and 0 for all other h. Therefore, by Bartlett's formula, the entries w_{ij} of the matrix W equal $w_{ij} = 1$ if $i = j$ and 0 otherwise. In other words, for large n, the sample autocorrelations $(r_1, \dots, r_k)$ are approximately independent and identically distributed normal random variables with mean 0 and variance $1/n$. We have already seen this before. The horizontal blue lines in R in an acf plot are at $\pm 1.96 n^{-1/2}$ and are based on this fact.*

Example 1.2 (MA(1) Process)**.** *Suppose $X_t = Z_t + \theta Z_{t-1}$. We have seen that $\rho_X(1) = \theta/(1+\theta^2)$ and $\rho_X(h) = 0$ for higher lags h. Bartlett's formula says that the variance of r_i is approximately w_{ii}/n where*

$$w_{ii} = \sum_{m=1}^{\infty} \left(\rho(m+i) + \rho(m-i) - 2\rho(i)\rho(m)\right)^2 .$$

For $i = 1$ i.e., when we consider the first order sample autocorrelation, this formula gives $1 - 3\rho^2(1) + 4\rho^4(1)$. Because $\rho^2 \leq 1/4$ for an MA(1) process, we have $w_{11} < 1$. In other words, r_1 for MA(1) is less variable than r_1 for white noise.

For higher values of i, the formula gives $w_{ii} = \sum_m \rho^2(m-i) = 1 + 2\rho^2(1)$ and note that this is strictly larger 1. In other words, r_k for $k \geq 2$ are more variable for MA(1) than for white noise. Thus we can expect to see more r_ks sticking out the horizontal blue lines for MA(1).

What about the correlation between r_1 and r_2? To answer this, we need to find out w_{12}:

$$w_{12} = (1 - 2\rho^2(1))(\rho(1)) + \rho(1) = 2\rho(1)(1 - \rho^2(1)).$$

Example 1.3 (AR(1) Process)**.** *Suppose $X_t - \phi X_{t-1} = Z_t$ where $|\phi| < 1$. Here the formula for w_{ii} becomes*

$$\begin{aligned}
w_{ii} &= \sum_{m=1}^{\infty} \left(\phi^{m+i} + \phi^{|m-i|} - 2\phi^{m+i}\right)^2 \\
&= \sum_{m=1}^{\infty} \left(\phi^{|m-i|} - \phi^{m+i}\right)^2 \\
&= \sum_{m=1}^{i} \left(\phi^{i-m} - \phi^{m+i}\right)^2 + \sum_{m=i+1}^{\infty} \left(\phi^{m-i} - \phi^{m+i}\right)^2 \\
&= \frac{(1+\phi^2)(1-\phi^{2i})}{1-\phi^2} - 2i\phi^{2i}.
\end{aligned}$$

Thus the variance of r_1 is approximately $(1-\phi^2)/n$. For a larger lag i, the terms ϕ^{2i} and $i\phi^{2i}$ will be small and can be ignored. In that case, the approximate variance of r_i becomes:

$$var(r_i) \approx \frac{1}{n}\frac{1+\phi^2}{1-\phi^2} \qquad \text{when } i \text{ is large.}$$

Note that, in contrast to the variance of r_1, the term $1-\phi^2$ is in the denominator in the above expression.

Here is one technique for finding a suitable ARMA model for the data (assuming of course that the data look stationary). Plot the sample autocorrelation function and compare it with the theoretical acf of an ARMA **keeping in mind the variability of the sample autocovariance function that is given by Bartlett's formula**. For example, if the sample autocorrelations after lag q drop off and lie between the bands for MA(q) given by Bartlett's formula, the MA(q) can be used for the data.

It is not easy to choose the order of an appropriate AR process for the data in this way looking at the sample autocorrealtion function. Here, the **Partial Autocorrelation Function** will be much more helpful. The Partial Autocorrelation Function or **pacf** is a function of the lag which has the following property: For an AR(p) model, the pacf equals zero for lags strictly larger than p. In order to define the pacf, we need to understand the concept of **best linear prediction**. This will be the subject of the next class.

Book Readings: Theorem A.7 (Appendix A).