Spring 2015 Statistics 153 (Time Series) : Lecture Seventeen

Aditya Guntuboyina

31 March 2015

1 ARIMA Models

ARIMA is essentially differencing plus ARMA. We have seen previously that differencing is commonly used on time series data to remove trends and seasonality.

For example, differencing can be used for

1. Removing polynomial trends: Suppose the data come from the model $Y_t = \mu_t + X_t$ where μ_t is a polynomial of order k and X_t is stationary, then differencing of order k: $\nabla^k Y_t = (I - B)^k Y_t$ results in stationary data to which an ARMA model can be fit.
2. Random walk models: Suppose that the data come from the random walk model: $Y_t = Y_{t-1} + X_t$ where X_t is stationary. Then clearly $\nabla Y_t = X_t$ is stationary and an ARMA model can be fit to this difference data.

Such models, which after appropriate differencing, reduce to ARMA models are called ARIMA models.

Definition 1.1 (ARIMA)**.** *A process Y_t is said to be ARIMA(p, d, q) with mean μ if $X_t = (I - B)^d Y_t$ is ARMA(p, q) with mean μ. In other words:*

$$\phi(B)(X_t - \mu) = \theta(B)Z_t,$$

where $\{Z_t\}$ is white noise.

2 Fitting ARIMA models

Just use the function *arima(dataset, order = c(p, d, q))*. I suggest you always use this function. If you know that you want to fit a pure AR model, you might consider the *ar()* function.

The arima function will give you the estimates of μ (under the name intercept), $\phi_1, \dots, \phi_p$ and $\theta_1, \dots, \theta_q$. It will also give you the estimated standard errors. An estimate of σ^2 is also provided.

3 ARIMA Forecasting

Forecasting for a future observation, x_{n+m}, is done using the best linear predictor of X_{n+m} in terms of $X_1, \dots, X_n$. The coefficients of the best linear predictor involve the parameters of the ARIMA model used for $x_1, \dots, x_n$. These paramters are estimated from data.

We have already seen how the best linear predictor of a random variable Y in terms of $W_1, \ldots, W_m$ is calculated.

Suppose that all the random variables $Y, W_1, \ldots, W_m$ have mean zero. Then the best linear predictor is $a_1 W_1 + \cdots + a_m W_m$ where $a_0, \ldots, a_m$ are characterized by the set of equations:

$$\mathrm{cov}(Y - a_1 W_1 - \cdots - a_m W_m, W_i) = 0 \qquad \text{for } i = 1, \ldots, m.$$

The above gives m equations in the m unknowns $a_1, \ldots, a_m$. The equations can be written in a compact form as $\Delta a = \zeta$ where $\Delta(i,j) = \mathrm{cov}(W_i, W_j)$ and $\zeta_i = \mathrm{cov}(Y, W_i)$.

If the random variables $Y, W_1, \ldots, W_m$ have different means: $\mathbb{E}Y = \mu_Y$ and $\mathbb{E}W_i = \mu_i$, then the best linear predictor of $Y - \mu_Y$ in terms of $W_1 - \mu_1, \ldots, W_m - \mu_m$ is given by $a_1(W_1 - \mu_1) + \cdots + a_m(W_m - \mu_m)$ where $a_1, \ldots, a_m$ are given by the same equation $\Delta a = \zeta$. Thus, in these non-zero mean case, the best linear predictor of Y in terms of $W_1, \ldots, W_m$ is

$$\mu_Y + a_1(W_1 - \mu_1) + \cdots + a_m(W_m - \mu_m).$$

The prediction error is measured by

$$\mathbb{E}\left(Y - \mu_Y - a_1(W_1 - \mu_1) - \cdots - a_m(W_m - \mu_m)\right)^2.$$

For ARMA models, there exist iterative algorithms for quickly calculating the best linear predictors of X_{n+m} based on $X_1, \ldots, X_n$ and the corresponding prediction errors recursively over n and m e.g., Durbin-Levinson and Innovations. These do not explicit inversion of the matrix Δ.

4 Time Series Data Analysis

1. Exploratory analysis.
2. Decide if it makes sense to transform the data (either for better interpretation or for stabilizing the variance).
3. Deal with trend or seasonality. Either by fitting deterministic models or by smoothing or differencing.
4. Fit an ARMA model to the residuals obtained after trend and seasonality are removed.
5. Check if the fitted ARMA model is adequate (Model Diagnostics).
6. Forecast.