Spring 2015 Statistics 153 (Time Series) : Lecture Thirteen

Aditya Guntuboyina

03 March 2015

1 Last Class: Best Linear Prediction

Let Y and $W_1, \ldots, W_m$ represent mean zero random variables which have finite variances. Let $cov(Y, W_i) = \zeta_i$ for $i = 1, \ldots, m$ and $Cov(W_i, W_j) = \Delta(i, j)$ for $i, j = 1, \ldots, m$.

The best linear predictor for Y in terms of $W_1, \ldots, W_m$ can be defined in two ways:

1. It is the linear combination $a_1 W_1 + \cdots + a_m W_m$ which minimizes $\mathbb{E}(Y - a_1 W_1 - \cdots - a_m W_m)^2$.
2. It is the linear combinator $a_1 W_1 + \cdots + a_m W_m$ for which

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

The coefficients $a_1, \ldots, a_m$ of the best linear predictor are determined by $a = \Delta^{-1}\zeta$.

2 The Partial Autocorrelation Function (pacf)

2.1 First Definition

Let $\{X_t\}$ be a mean zero stationary process. The Partial Autocorrelation at lag h, denoted by $pacf(h)$ is defined as the coefficient of X_{t-h} in the best linear predictor for X_t in terms of $X_{t-1}, \ldots, X_{t-h}$.

Check that $pacf(1)$ is the same as the autocorrelation at lag one, $\rho(1)$. But $pacf(h)$ for $h > 1$ can be quite different from $\rho(h)$.

An important fact is that for the AR(p) model: $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$, the partial autocorrelation function satisfies $pacf(p) = \phi_p$ and that $pacf(h) = 0$ for $h > p$.

2.2 Second Definition

From the first definition, it is not quite clear why this is called a correlation. This will be apparent from the second definition.

The pacf at lag h is defined as the correlation between X_t and X_{t-h} **with the effect of the intervening variables $X_{t-1}, X_{t-2}, \ldots, X_{t-h+1}$ removed**. Let $\beta_1 X_{t-1} + \cdots + \beta_{h-1} X_{t-h+1}$ denote the best linear predictor of X_t in terms of $X_{t-1}, \ldots, X_{t-h+1}$. By stationarity, the two sequences

$$X_t, X_{t-1}, \ldots, X_{t-h+1}$$

and

$$X_{t-h}, X_{t-h+1}, \ldots, X_{t-1}$$

have the same covariance matrix. Indeed, if $W_i = X_{t-i+1}$ and $\tilde{W}_i = X_{t-h+i-1}$ for $i = 1, \ldots, h$, then the covariance between W_i and W_j equals $\gamma_X(i-j)$ which is the same as the covariance between $\tilde{W}_i$ and $\tilde{W}_j$.

Therefore, the best linear prediction of X_{t-h} in terms of $X_{t-h+1}, \ldots, X_{t-1}$ equals $\beta_1 X_{t-h+1} + \cdots + \beta_{h-1} X_{t-1}$.

The pacf at lag h is defined as

$$pacf(h) = \text{corr}\left(X_t - \beta_1 X_{t-1} - \cdots - \beta_{h-1} X_{t-h+1}, X_{t-h} - \beta_1 X_{t-h+1} - \cdots - \beta_{h-1} X_{t-1}\right).$$

In other words, $pacf(h)$ is the correlation between the **errors in the best linear predictions** of X_t and X_{t-h} in terms of the intervening variables $X_{t-1}, \ldots, X_{t-h+1}$.

The equivalence between the two definitions of $pacf(h)$ can be proved by linear algebra. We will skip this derivation.

The fact that $pacf(h)$ equals zero for lags $h > p$ for an AR(p) model can also be deduced from this second definition of $pacf$. To see this: note that for $h > p$, the best linear predictor for X_t in terms of $X_{t-1}, \ldots, X_{t-h+1}$ equals $\phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p}$. In other words, $\beta_1 = \phi_1, \ldots, \beta_p = \phi_p$ and $\beta_i = 0$ for $i > p$.

Therefore for $h > p$, we have

$$\begin{aligned} pacf(h) &= \text{corr}\left(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}, X_{t-h} - \phi_1 X_{t-h+1} - \cdots - \phi_p X_{t-h+p}\right) \\ &= \text{corr}\left(Z_t, X_{t-h} - \phi_1 X_{t-h+1} - \cdots - \phi_p X_{t-h+p}\right) = 0, \end{aligned}$$

by causality.

3 Estimating $pacf$ from Data

How does one estimate $pacf(h)$ from data for different lags h? The coefficients $a_1, \ldots, a_h$ of $X_{t-1}, \ldots, X_{t-h}$ in the best linear predictor of X_t are obtained by solving an equation of the form $\Delta a = \zeta$.

Now all the elements of Δ and ζ are of the form $\gamma_X(i-j)$ for some i and j. Therefore, a natural method of estimating $pacf(h)$ is to estimate the entries in Δ and ζ by the respective sample autocorrelations to obtain $\hat{\Delta}$ and $\hat{\zeta}$ and then to solve the equation $\hat{\Delta}\hat{a} = \hat{\zeta}$ for $\hat{a}$. Note that $pacf(h)$ is precisely a_h.

It has been shown that when the data come from an AR(p) model, the sample partial autocorrelations at lags greater than p are approximately **independently normally** distributed with zero means and variances $1/n$. Thus for $h > p$, bands at $\pm 1.96 n^{-1/2}$ can be used for checking if an AR(p) model is appropriate.

4 Summary

For an $MA(q)$ model, the autocorrelation function $\rho_X(h)$ equals zero for $h > q$. Also for $h > q$, the sample autocorrelation functions r_h are approximately normal with mean 0 and variance w_{hh}/n where $w_{hh} := 1 + 2\rho^2(1) + \cdots + 2\rho^2(q)$.

For an $AR(p)$ model, the partial autocorrelation function $pacf(h)$ equals zero for $h > p$. Also for $h > p$, the sample partial autocorrelations are approximately normal with mean 0 and variance $1/n$.

If the sample acf for a data set cuts off at some lag, we use an MA model. If the sample pacf cuts off at some lag, we use an AR model.

What if neither of the above happens? How do we then choose an appropriate ARMA model? Here is a general strategy:

1. Try ARMA(p, q) for various choices of p and q.
2. For a fixed p and q, fit the ARMA(p, q) model to the data (we will soon learn how to do this).
3. See how good the fit is. Select p and q so that the fit is good **while making sure there is no overfitting**.

How to check if a model is fits the data well but does not overfit? This is a problem of model selection. Often automatic criteria like AIC, FPE, BIC are used. One should also use judgement.

Our plan is as follows:

1. How to fit an ARMA model to data?
2. How to assess goodness of fit?
3. Choosing p and q by an automatic Model selection technique.