Spring 2015 Statistics 153 (Time Series) : Lecture Twelve

Aditya Guntuboyina

26 February 2015

The partial autocorrelation function is very useful for detecting autoregressive models. To understand partial autocorrelation function, we first need to learn about prediction.

1 Best Prediction

Suppose Y and $W_1, \dots, W_m$ are random variables. What is the best prediction for Y in terms of $W_1, \dots, W_m$? *Best* here is in terms of squared error loss: we want to find the function $f^*(W_1, \dots, W_m)$ which minimizes

$$\mathbb{E}(Y - f(W_1, \dots, W_m))^2$$

over all functions f. The answer is given by

$$f^*(W_1, \dots, W_n) := \mathbb{E}\left(Y | W_1, \dots, W_m\right).$$

A problem with this best prediction is that one needs to, in general, know the entire joint distribution of $Y, W_1, \dots, W_m$ in order to compute it. On the other hand, it is much easier to compute the best **linear** prediction of Y in terms of $W_1, \dots, W_m$.

2 Best Linear Prediction

Suppose that Y and $W_1, \dots, W_m$ are random variables with zero means and finite variances. Let $\text{cov}(Y, W_i) = \zeta_i, i = 1, \dots, m$ and

$$\text{cov}(W_i, W_j) = \Delta(i, j) \qquad \text{for } i, j = 1, \dots, m.$$

What is the best **linear predictor** of Y in terms of $W_1, \dots, W_m$?

The best linear predictor $a_1, \dots, a_m$ is characterized by the property that $Y - a_1 W_1 - \cdots - a_m W_m$ is uncorrelated with $W_1, \dots, W_m$. In other words:

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

Note that this gives m equations in the m unknowns $a_1, \dots, a_m$. The ith equation can be rewritten as

$$\zeta_i - \Delta(i, 1) a_1 - \cdots - \Delta(i, m) a_m = 0.$$

In other words, this means that ζ_i equals the ith row of Δ multiplied by the vector $a = (a_1, \dots, a_m)^T$ which is same as the ith element of the vector Δa. Thus these m equations can be written in one line as $\Delta a = \zeta$.

Another way to get this defining equation for the coefficients of the best linear predictor is to find values of $a_1, \dots, a_m$ that minimize

$$\begin{aligned}
F(\mathbf{a}) &:= \mathbb{E}\left(Y - a_1 W_1 - \cdots - a_m W_m\right)^2 \\
&= \mathbb{E}\left(Y - a^T W\right)^2 \\
&= \mathbb{E}Y^2 - 2\mathbb{E}((a^T W)Y) + \mathbb{E}(a^T W W^T a) \\
&= \mathbb{E}Y^2 - 2a^T \zeta + a^T \Delta a.
\end{aligned}$$

Differentiate with respect to a and set equal to zero to get

$$-2\zeta + 2\Delta a = 0$$

or $a = \Delta^{-1}\zeta$. Therefore the best linear predictor of Y in terms of $W_1, \dots, W_m$ equals $\zeta^T \Delta^{-1} W$.

The special case of this for $m = 1$ (when there is only one predictor W_1) may be more familiar. When $m = 1$, we have $\zeta_1 = \mathrm{cov}(Y, W_1)$ and $\Delta(1,1) = \mathrm{var}(W_1)$. Thus, the best predictor or Y in terms of W_1 is

$$\frac{\mathrm{cov}(Y, W_1)}{\mathrm{var}(W_1)} W_1.$$

Now consider a stationary mean zero time series $\{X_t\}$. Using the above with $Y = X_n$ and $W_1 = X_{n-1}$, we get that the best predictor of X_n in terms of X_{n-1} is

$$\frac{\mathrm{cov}(X_n, X_{n-1})}{\mathrm{var}(X_{n-1})} X_{n-1} = \frac{\gamma_X(1)}{\gamma_X(0)} X_{n-1} = \rho_X(1) X_{n-1}$$

What is the best predictor for X_n in terms of $X_{n-1}, X_{n-2}, \dots, X_{n-k}$? Here we take $Y = X_n$ and $W_i = X_{n-i}$ for $i = 1, \dots, k$. Therefore

$$\Delta(i,j) = \mathrm{cov}(W_i, W_j) = \mathrm{cov}(X_{n-i}, X_{n-j}) = \gamma_X(i-j)$$

and

$$\zeta_i = \mathrm{cov}(Y, W_i) = \mathrm{cov}(X_n, X_{n-i}) = \gamma_X(i).$$

With these Δ and ζ, solve for $\Delta a = \zeta$ to obtain the coefficients of $X_{n-1}, \dots, X_{n-k}$ in the best linear predictor of X_n.

Consider the special case of the AR(p) model: $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$. Directly from the defining equation and causality, it follows that $X_n - \phi_1 X_{n-1} - \cdots - \phi_p X_{n-p}$ is uncorrelated with $X_{n-1}, X_{n-2}, \dots$. We thus deduce that the best linear predictor of X_n in terms of $X_{n-1}, X_{n-2}, \dots$ equals $\phi_1 X_{n-1} + \phi_2 X_{n-2} + \cdots + \phi_p X_{n-p}$.

3 Best Prediction and Best Linear Prediction

The Best Linear Prediction is in general worse than the Best Prediction. However, it is much easier to compute because it only requires knowledge of the covariances between the variables while the best predictor requires knowledge of the entire joint distribution.

In the special case when $Y, W_1, \dots, W_m$ are jointly gaussian, the best prediction and best linear prediction coincide.