Lecture 14

Partial Autocorrelation Function (PACF)
Textbook Sections: 3.2

The first part of the lecture will be devoted to forecasting $ARMA(p, q)$ models.

Motivation

Suppose we want to measure the strength of the linear relationship of X_t and X_{t-2}. We would first look at $\gamma(2)$, or a sample estimate $\hat{\gamma}(2)$. However, consider this. If X_t and X_{t-2} are both correlated with X_{t-1}, then some of the linear dependence between X_t and X_{t-2} is due to that relationship with X_{t-1}. It's useful to be able to filter that out.

In particular, if our goal is to make a predictive model of X_t in terms of the previous variables, then this knowledge would tell us whether it is useful to include X_{t-2} in the model once X_{t-1} is already included in the model.

The **partial autocorrelation (PACF)** function at lag h ($h = 2, 3, \cdots$) measures the linear relationship between X_t and X_{t-h} while accounting for the presence of the intermediate terms $X_{t-1} \cdots X_{t-(h-1)}$.

Definition

Suppose we are dealing with a sequence $\{X_t\}$. This can be any random sequence, not necessarily $ARMA(p, q)$. The following steps outline how the PACF is defined.

Suppose $h = 2, 3, \cdots$.

1. Define $\hat{X}^{(f)}_{t-h,t}$ to be the best linear predictor (forecast) of X_t based on $X_{t-(h-1)}, \cdots, X_{t-1}$.
2. Similarly, define $\hat{X}^{(b)}_{t-h,t}$ to be the best linear predictor (sometimes called a backcast) of X_{t-h} based on $X_{t-(h-1)}, \cdots, X_{t-1}$.
3. Consider the error of the forecast, $X_t - \hat{X}^{(f)}_{t-h,t}$. This is the difference between the actual X_t, and the best we can do in predicting it linearly using terms between X_{t-h} and X_t.
4. Consider the error of the backcast, $X_t - \hat{X}^{(b)}_{t-h,t}$. This is the difference between the actual X_{t-h}, and the best we can do in predicting it linearly using terms between X_{t-h} and X_t.
5. The PACF at lag $h = 2, 3, \cdots$ is defined as the correlation of the two errors,

$$Corr(X_t - \hat{X}^{(b)}_{t-h,t}, X_t - \hat{X}^{(f)}_{t-h,t})$$

By convention, the PACF at lag $h = 1$ is set equal to the ACF at lag $h = 1$.

The term "best linear predictor" is defined with respect to expected squared error. Thus,

$$\hat{X}^{(f)}_{t-h,t} = c_0 + c_1 X_{t-1} + \cdots c_{h-1} X_{t-(h-1)},$$

where $c_0, c_1, \cdots, c_{h-1}$ are real constants computed to minimize $\mathrm{E}[(X_t - \hat{X}^{(f)}_{t-h,t})^2]$.
The backcast $\hat{X}^{(b)}_{t-h,t}$ is defined similarly.

PACF of an Autoregressive Process

In general, the computation of PACF values is complicated. However, there is one important case. If we are dealing with an $AR(p)$ process, then the PACF values are equal to 0 for lag $h > p$. Here's a rough explanation.

Let's consider a zero-mean $AR(p)$ process. An $AR(p)$ process with mean $\mu \neq 0$ will have the same ACF and PACF structure as a zero-mean $AR(p)$ process if the other parameters are the same. Thus let's work with

$$X_t = \phi_1 X_{t-1} + \cdots \phi_p X_{t-p} + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is $WN(0, \sigma^2)$. Let's follow the steps in the previous section for any $h > p$.

1. Define $\hat{X}^{(f)}_{t-h,t}$ to be the best linear predictor (forecast) of X_t based on $X_{t-(h-1)}, \cdots, X_{t-1}$.
   Since $h > p$, we can predict X_t up to the noise term ε_t using the intermediate terms. In other words, $\hat{X}^{(f)}_{t-h,t} = \phi_1 X_{t-1} + \cdots \phi_p X_{t-p}$.

2. Similarly, define $\hat{X}^{(b)}_{t-h,t}$ to be the best linear predictor (sometimes called a backcast) of X_{t-h} based on $X_{t-(h-1)}, \cdots, X_{t-1}$.
   The exact form of $\hat{X}^{(b)}_{t-h,t}$ isn't as important. But it is important to note that it is composed only of $X_{t-(h-1)}, \cdots, X_{t-1}$, all terms that came before time t.

3. Consider the error of the forecast, $X_t - \hat{X}^{(f)}_{t-h,t}$. This is the difference between the actual X_t, and the best we can do in predicting it linearly using terms between X_{t-h} and X_t.
   Here the forecast error is precisely the noise term at time t, $X_t - \hat{X}^{(f)}_{t-h,t} = \varepsilon_t$.

4. Consider the error of the backcast, $X_t - \hat{X}^{(b)}_{t-h,t}$. This is the difference between the actual X_{t-h}, and the best we can do in predicting it linearly using terms between X_{t-h} and X_t.
   Again, we won't derive the exact form of the backcast error $X_{t-h} - \hat{X}^{(b)}_{t-h,t}$. It is enough to note that it only depends on X_{t-2} and $X_{t-(h-1)}, \cdots, X_{t-1}$, again all terms that came before time t.

5. The PACF at lag $h = 2, 3, \cdots$ is defined as the correlation of the two errors,

   $$Corr(X_t - \hat{X}^{(b)}_{t-h,t}, X_t - \hat{X}^{(f)}_{t-h,t})$$

   The first term inside the parentheses depends on variables with indices less than t. The right term is ε_t, which is uncorrelated with anything with index other than t. Thus the correlation is equal to 0.

ACF and PACF as Diagnostic tools

Let's recall the different ways in which the sample ACF can be used as a diagnostic tool.

- The sample ACF plot can be used to study the correlation structure of the data. The $1.96/\sqrt{n}$ bars can be used to determine whether there is significant correlation at any lag.
- The sample ACF values can be used in the Box-Ljung test, or other similar tests.

The PACF is used in a similar way. It's conventional to plot the sample PACF values for lags $h = 1, 2, 3 \cdots$ in the same fashion as the sample ACF values. The two plots are usually made at the same time (when checking if data is i.i.d., analyzing the dependence structure, or choosing a model.

The $1.96/\sqrt{n}$ bars can be used in the same way to determine significance, as there is a similar result that states that if the data is i.i.d., then the sample PACF values are approximately standard normal random variables.

Finally, here's an important result:

- The ACF of an $MA(q)$ process is equal to 0 if lag is $h > q$. The PACF of an $MA(q)$ process decays exponentially with h (possibly changing sign).
- The PACF of an $AR(p)$ process is equal to 0 if lag is $h > p$. The ACF of an $AR(p)$ process decays exponentially with h (possibly changing sign).

You can see this by making plots of true (theoretical) PACF and ACF values in R. The function `ARMAacf()` can be used to obtain the true (not sample-based) values of the ACF and PACF for a specified $ARMA(p, q)$ process. Plot these values for a few $MA(q)$ and $AR(p)$ processes. See what happens when you make some coefficients (θs or ϕs) positive, and some negative.

Thus we may sometimes be able to use the ACF and PACF plots to make a preliminary model choice. If the sample plots exhibit any of the above characteristics, then we can consider an $MA(q)$ or $AR(p)$ model, and we have a guess for q or p. However, it is always best to select the final model using some systematic approach, such as model selection criteria (typically AICc or BIC).

Keep in mind, that the sample ACF and PACF plots won't reflect the patterns perfectly. You can try this in R and see for yourself. Simply simulate some $AR(p)$ or$MA(q)$ data using `sim.arima()`, then use `acf()` or `pacf()` to produce the sample ACF and PACF plots. Try this several times for each choice of model to see how things vary from sample to sample.

You can find more information about interpreting the sample ACF and PACF plots here:
`https://en.wikipedia.org/wiki/Box%E2%80%93Jenkins_method#Autocorrelation_and_partial_autocorrelation_plots`.

Further Info

For some further information about the PACF, along with some technical details about the computation, I'll post Dr. Burman's handout on the topic on Smartsite. It's labeled Handout 9, and will be available in the Supplementary section of Resources. Check it out if you're interested, but keep in mind that you won't be tested on any further PACF information.