25 Spring 439/639 TSA: Lecture 12

Dr Sergey Kushnarev

Table of contents

Partial Autocorrelation Function (PACF) **1**
Definition 1 . . . 1
Definition 2 . . . 1
Definition 3 . . . 2

Durbin–Levinson Recursion **3**

Sample PACF **3**

Partial Autocorrelation Function (PACF)

Definition 1

Last time, we introduced one definition of the partial autocorrelation function (partial ACF, PACF) ϕ_{kk}.

Definition 1:

$$\phi_{kk} = \operatorname{corr}\left(Y_t,\ Y_{t-k} \mid Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}\right).$$

From this definition, it's easy to show that: For AR(1) process, $Y_t = \phi Y_{t-1} + e_t$, $\phi_{11} = \rho_1$, and $\phi_{kk} = 0$ for any lag $k \geq 2$.

Definition 2

Definition 2:

$$\phi_{kk} = \operatorname{corr}(\mathrm{Res}_t, \mathrm{Res}_{t-k}),$$

where Res_t = residual from (linear) regressing Y_t on $Y_{t-1}, \dots, Y_{t-k+1}$, and Res_{t-k} = residual from (linear) regressing Y_{t-k} on $Y_{t-1}, \dots, Y_{t-k+1}$. So Res_t and Res_{t-k} are the unexplained variation in Y_t, Y_{t-k} after "partialling out" the effect of $Y_{t-1}, \dots, Y_{t-k+1}$ (i.e., "controlling for" $Y_{t-1}, \dots, Y_{t-k+1}$).

Let's look at the AR(1) example $Y_t = \phi Y_{t-1} + e_t$ again. To find ϕ_{22}, we need to regress Y_t on Y_{t-1}, and regress Y_{t-2} on Y_{t-1}.

So we need to find a, b by minimizing

$$\min_a \mathbb{E}\left(Y_t - aY_{t-1}\right)^2, \quad \min_b \mathbb{E}\left(Y_{t-2} - bY_{t-1}\right)^2,$$

then $\mathrm{Res}_t = Y_t - \widehat{Y}_t$ and $\mathrm{Res}_{t-k} = Y_{t-2} - \widehat{Y}_{t-2}$, where $\widehat{Y}_t = aY_{t-1}$, $\widehat{Y}_{t-2} = bY_{t-1}$.

To solve $\min_a \mathbb{E}\left(Y_t - aY_{t-1}\right)^2$, we take the derivative of the objective function:

$$\frac{\partial}{\partial a} \mathbb{E}\left[\left(Y_t - aY_{t-1}\right)^2\right] = -2\mathbb{E}\left[\left(Y_t - aY_{t-1}\right) Y_{t-1}\right]$$

Set the derivative equal to 0 gives

$$\mathbb{E}\left[\left(Y_t - aY_{t-1}\right)Y_{t-1}\right] = 0 \implies \gamma_1 - a\gamma_0 = 0 \implies a = \frac{\gamma_1}{\gamma_0} = \rho_1 = \phi.$$

To solve $\min_b \mathbb{E}\left(Y_{t-2} - bY_{t-1}\right)^2$, we take the derivative of the objective function:

$$\frac{\partial}{\partial b}\mathbb{E}\left(Y_{t-2} - bY_{t-1}\right)^2 = -2\mathbb{E}\left[\left(Y_{t-2} - bY_{t-1}\right)Y_{t-1}\right]$$

Set the derivative equal to 0 gives

$$\mathbb{E}\left[\left(Y_{t-2} - bY_{t-1}\right)Y_{t-1}\right] = 0 \implies \gamma_1 - b\gamma_0 = 0 \implies b = \frac{\gamma_1}{\gamma_0} = \rho_1 = \phi.$$

Then we get

$$\phi_{22} = \mathrm{corr}(\mathrm{Res}_t, \mathrm{Res}_{t-2}) = \mathrm{corr}(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}) = \mathrm{corr}(e_t, Y_{t-2} - \phi Y_{t-1}) = 0.$$

In general, for AR(p), the PACF at lags 1 through p can be nonzero, and PACF at lags $k \geq p+1$ are all zero. And we also have the following results.

	MA(q)	AR(p)	ARMA(p, q)
ACF	cuts off after q	exponential decay	exponential decay
PACF	exponential decay	cuts off after p	exponential decay

Definition 3

We also have an alternative, "computational" definition of ϕ_{kk}. The idea is to fit an AR(k) model to the stationary time series (Y_t) of our interest, i.e, fit the linear regression to (Y_t):

$$Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots + \phi_{kk}Y_{t-k} + \epsilon,$$

or in other words, find $\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk}$ to minimize

$$\min_{\phi_{k1},\phi_{k2},\ldots,\phi_{kk}} \mathbb{E}\left(Y_t - \phi_{k1}Y_{t-1} - \phi_{k2}Y_{t-2} - \cdots - \phi_{kk}Y_{t-k}\right)^2.$$

Definition 3 (We claim without proof that:) For stationary (Y_t), its PACF at lag k, i.e. ϕ_{kk}, is same as the fitted value of ϕ_{kk} in the $(\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk})$ from the regression (namely, AR(k) fitting) above.

In other words, ϕ_{kk} **is the last** ϕ_{kj} **term in the AR(k) approximation to** (Y_t).

From this definition, we can show the following fact: the solution $(\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk})$ to the regression above must satisfy

$$\Gamma_k\,\vec{\phi}_k = \vec{\gamma}_k,$$

$$\text{where} \quad \Gamma_k = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix}, \quad \vec{\phi}_k = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \quad \vec{\gamma}_k = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}.$$

Note: the derivation for $\Gamma_k\,\vec{\phi}_k = \vec{\gamma}_k$ is very similar to the AR(1) example after Definition 2.

So now we know the PACF ϕ_{kk} is the last entry of $\Gamma_k^{-1}\,\vec{\gamma}_k$. Inverting a $k \times k$ matrix Γ_k can be computationally expensive. Instead, we can use Durbin–Levinson Recursion (we will get to it soon) to directly calculate the entries of $\Gamma_k^{-1}\,\vec{\gamma}_k$ without computing matrix inverse.

From this Definition 3 and the fact above, we can also intuitively see the PACF behavior for AR(p) that we mentioned earlier. Suppose (Y_n) is an AR(p), $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$ where $\phi_1, ..., \phi_p$ are the true parameters of the AR(p). To get PACF ϕ_{kk}, we need to fit an AR(k)-like linear regression for Y_t. If $k \geq p$, then intuitively,

$$\begin{aligned} \text{fitting a regression model} \quad & Y_t = \phi_{k1} Y_{t-1} + \phi_{k2} Y_{t-2} + \cdots + \phi_{kk} Y_{t-k} + \epsilon, \\ \text{for an AR}(p) \quad & Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t, \end{aligned}$$

should give us the fitted value $(\phi_{k1}, \phi_{k2}, \dots, \phi_{kk}) = (\phi_1, ..., \phi_p, 0, ..., 0)$. So we have the following results for AR(p).

- If $k = p$, then $\phi_{pp} = \phi_p$.
- If $k \geq p + 1$, i.e. $k > p$, then $\phi_{kk} = 0$.

Remark: For AR(p) and $k \geq p$, the statement $(\phi_{k1}, \phi_{k2}, \dots, \phi_{kk}) = (\phi_1, ..., \phi_p, 0, ..., 0)$ can be verified by showing that it is indeed a solution to $\Gamma_k \ \vec{\phi}_k = \vec{\gamma}_k$. The derivation reduces to YW equations for AR(p).

Durbin–Levinson Recursion

As we mentioned earlier, although the PACF ϕ_{kk} can be characterized as the last entry of $\Gamma_k^{-1} \ \vec{\gamma}_k$, it can be expensive to compute Γ_k^{-1}. We can use Durbin–Levinson Recursion to directly calculate the entries of $\Gamma_k^{-1} \ \vec{\gamma}_k$ without computing matrix inverse. (*Remark:* the basic idea is to utilize the special structure of the matrix Γ_k and the vector $\vec{\gamma}_k$ to compute $\Gamma_k^{-1} \ \vec{\gamma}_k$.)

Durbin–Levinson Recursion (DLR): Define $\phi_{00} = 1$. For $l \geq 0$, define

$$\begin{cases} \phi_{l+1,\ l+1} & = \frac{\gamma_{l+1} - \sum_{j=1}^{l} \phi_{l,\,j}\, \gamma_{l+1-j}}{\gamma_0 - \sum_{j=1}^{l} \phi_{l,\,j}\, \gamma_j} = \frac{\rho_{l+1} - \sum_{j=1}^{l} \phi_{l,\,j}\, \rho_{l+1-j}}{1 - \sum_{j=1}^{l} \phi_{l,\,j}\, \rho_j} \\ \phi_{l+1,\,j} & = \phi_{l,\,j} - \phi_{l+1,\,l+1}\, \phi_{l,\,l+1-j} \qquad \text{for } 1 \leq j \leq l \end{cases}$$

Note: For the first formula, we can use it either in terms of γ_k or ρ_k.

DLR is useful because we do not need to compute the inverse of a certain (large) matrix.

Example: Find PACF for an AR(1) via DLR. Consider the AR(1) process $Y_t = cY_{t-1} + \epsilon_t$ with $|c| < 1$, so we know $\rho_k = c^k$. To do DLR, we first let $l = 1$ to compute ϕ_{11}:

$$\phi_{11} = \frac{\rho_1 - 0}{1 - 0} = \rho_1 = c.$$

Then let $l = 1$ in DLR to compute ϕ_{22}, ϕ_{21}:

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{c^2 - c^2}{1 - c^2} = 0,$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \rho_1 - 0 \cdot \rho_1 = \rho_1 = c.$$

Letting $l = 1$ in DLR gives $\phi_{33}, \phi_{32}, \phi_{31}$. For example,

$$\phi_{33} = \frac{\rho_3 - (\phi_{21}\rho_2 + \phi_{22}\rho_1)}{1 - (\phi_{21}\rho_1 + \phi_{22}\rho_2)} = \frac{c^3 - (c \cdot c^2 + 0 \cdot c)}{1 - (c \cdot c + 0 \cdot c^2)} = 0.$$

Exercise: (1) Find ϕ_{44}. (2) Use induction to prove that $\phi_{kk} = 0$ for all $k \geq 2$, and $\phi_{k1} = c$ for all $k \geq 1$.

Sample PACF

Method 1

(Similar to the idea of sample ACF) We can get sample PACF $\hat{\phi}_{kk}$ by using the sample ACF $\hat{\rho}_k = r_k$ in place of the theoretical ACF ρ_k in the DLR formula.

Method 2

Alternatively, we can also modify the equation $\Gamma_k \, \vec{\phi}_k = \vec{\gamma}_k$ into an empirical version $\widehat{R}_k \, \hat{\vec{\phi}}_k = \hat{\vec{\rho}}_k$ to get the solution $\hat{\vec{\phi}}_k = \widehat{R}_k^{-1} \, \hat{\vec{\rho}}_k$, and take its last entry to get $\hat{\phi}_{kk}$. To be specific, recall that the theoretical PACF satisfies

$$\Gamma_k \, \vec{\phi}_k = \vec{\gamma}_k, \quad \text{where} \quad \Gamma_k = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{k-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{k-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{k-1} & \gamma_{k-2} & \cdots & \gamma_0 \end{bmatrix}, \quad \vec{\phi}_k = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \quad \vec{\gamma}_k = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}.$$

Dividing both sides by γ_0 transforms the ACVF into ACF:

$$\begin{bmatrix} \rho_0 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{k-2} \\ \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_0 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}.$$

Then we use the sample ACF $\hat{\rho}_k = r_k$ in place of the theoretical ACF ρ_k to get the following equation. Solving this equation gives the sample PACF.

$$\widehat{R}_k \, \hat{\vec{\phi}}_k = \hat{\vec{\rho}}_k, \quad \text{where} \quad \widehat{R}_k = \begin{bmatrix} r_0 & r_1 & \cdots & r_{k-1} \\ r_1 & r_0 & \cdots & r_{k-2} \\ \vdots & \vdots & & \vdots \\ r_{k-1} & r_{k-2} & \cdots & r_0 \end{bmatrix}, \quad \hat{\vec{\phi}}_k = \begin{bmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \vdots \\ \hat{\phi}_{kk} \end{bmatrix}, \quad \hat{\vec{\rho}}_k = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{bmatrix}.$$

Remark: We just mentioned that $\hat{\vec{\phi}}_k = \widehat{R}_k^{-1} \, \hat{\vec{\rho}}_k$. The matrix $\widehat{R}_k$ defined above is indeed invertible, which is guaranteed by the particular way (and details) we used to construct the sample ACF r_k in lecture 9. In fact, we remarked in lecture 9 that our construction of r_k makes the "sample ACF matrix" invertible, and this matrix is nothing but the $\widehat{R}_k$ we have just seen.

The theoretical PACF ϕ_{kk} of a given time series model is a non-random number (but it can be unknown). The sample PACF $\hat{\phi}_{kk}$ is a random variable determined by the observed data (and the observations are random). So the sample PACF $\hat{\phi}_{kk}$ follows a certain sampling distribution that depends on the underlying time series model.

For an AR(p) process, the sampling distribution of $\hat{\phi}_{kk}$ is

$$\hat{\phi}_{kk} \sim \mathcal{N}\left(0, \frac{1}{n}\right), \quad \text{for all } k \geq p+1.$$

So if we want to test whether a time series is an AR(p), we can use the following rejection region

$$\left|\hat{\phi}_{kk}\right| > \frac{2}{\sqrt{n}}$$

for sample PACFs with lag $k \geq p+1$.