ST304 Week 7

Kaixin Liu[1]

[1]PhD Student in Statistics, LSE

Mar 5, 2026

Table of Contents

Table of Contents

Why partial autocorrelation?

Motivation

For a causal AR(p) process, the ACF typically **decays geometrically** and can be noisy in finite samples, so it is not always easy to read off p from an ACF plot.

Key idea

Measure the **direct** dependence between X_t and X_{t-h} after removing (regressing out) the linear effect of the intervening variables.

Outcome

The **PACF** is designed so that for a causal AR(p) model, it **cuts off** after lag p (Proposition 2.14).

Best linear predictor

Definition (Best linear predictor)

Let $Y, X_1, \ldots, X_m$ be **zero-mean** random variables with finite variance. The best linear predictor of Y in terms of $X = (X_1, \ldots, X_m)$ is

$$\Pi(Y \,|\, X_1, \ldots, X_m) \;=\; \beta_1^* X_1 + \cdots + \beta_m^* X_m \;=\; X^\top \beta^*,$$

where $\beta^* = (\beta_1^*, \ldots, \beta_m^*)$ minimises

$$\mathbb{E}(Y - \beta_1 X_1 - \cdots - \beta_m X_m)^2 \quad \text{over } (\beta_1, \ldots, \beta_m) \in \mathbb{R}^m.$$

Past/future best linear predictors

For $t \in \mathbb{Z}$ and $h \in \mathbb{N}$, define

$$\check{X}_t^h := \Pi(X_t \,|\, X_{t-1}, \ldots, X_{t-h}), \qquad \widehat{X}_t^h := \Pi(X_t \,|\, X_{t+1}, \ldots, X_{t+h}).$$

PACF via best linear prediction

Definition (Partial autocorrelation)

Let $(X_t)_{t\in\mathbb{Z}}$ be a zero-mean weakly stationary process. For $h \in \mathbb{N}$, define the partial autocorrelation at lag h by

$$\alpha(h) := \begin{cases} \mathrm{Corr}(X_1, X_0) = \rho(1), & h = 1, \\ \mathrm{Corr}\big(X_h - \check{X}_h^{(h-1)},\ X_0 - \widehat{X}_0^{(h-1)}\big), & h \geq 2. \end{cases}$$

By weak stationarity, for $h \geq 2$ this can be written (for any $t \in \mathbb{Z}$) as

$$\alpha(h) = \mathrm{Corr}\big(X_{t+h} - \check{X}_{t+h}^{(h-1)},\ X_t - \widehat{X}_t^{(h-1)}\big).$$

PACF cutoff for AR models

Proposition 2.14 (AR(p) cutoff)

If (X_t) is a **causal** AR(p) process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2),$$

then

$$\alpha(p) = \phi_p, \qquad \alpha(h) = 0 \text{ for all } h \geq p+1.$$

Equivalent definition of PACF (Yule–Walker form)

Equivalent definition (Equation (2.20))

For each $h \in \mathbb{N}$, let $\beta^*_{h,1}, \ldots, \beta^*_{h,h}$ satisfy

$$\underbrace{\begin{pmatrix} \gamma(0) & \gamma(-1) & \cdots & \gamma(-h+1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(-h+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \cdots & \gamma(0) \end{pmatrix}}_{\Gamma[h]} \begin{pmatrix} \beta^*_{h,1} \\ \beta^*_{h,2} \\ \vdots \\ \beta^*_{h,h} \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(h) \end{pmatrix}.$$

If $\Gamma[h]$ is invertible, then

$$\alpha(h) = \beta^*_{h,h}.$$

Table of Contents

ARIMA(p, d, q)

Differencing operator

Let

$$\nabla^d := (I - B)^d, \qquad d \geq 0,$$

so that $(\nabla^d X_t)$ is the d-th differenced series (with $\nabla X_t = X_t - X_{t-1}$).

Definition 2.15 (ARIMA(p, d, q))

We say that (X_t) is an ARIMA(p, d, q) process (orders $p, d, q \geq 0$) if $(\nabla^d X_t)$ is a **causal and invertible** ARMA(p, q) process, i.e.

$$(I - B)^d \Phi(B)(X_t - \mu) = \Theta(B)\varepsilon_t, \qquad \varepsilon_t \sim WN(0, \sigma^2),$$

where the AR and MA polynomials Φ, Θ have degrees p, q and have no common roots.