MEM-205 Περιγραφική Στατιστική

Τμήμα Μαθηματικών και Εφ. Μαθηματικών, Πανεπιστήμιο Κρήτης

Κώστας Σμαραγδάκης (kesmarag@gmail.com)

Θεωρία 11ης εβδομάδας

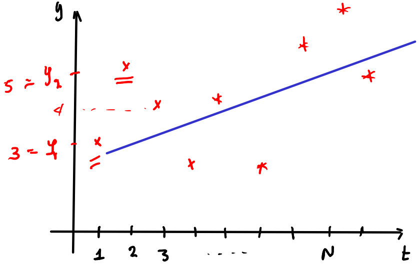
- Θέλουμε να προβλέψουμε μελοντικές τιμές μιας χρονολογικής σειράς

$t_1=1 \quad t_2=2, \ldots, t_N=N$

$$\{Y_1, Y_2, \ldots, Y_N\}$$

Στόχος είναι η πρόβλεψη των μελλοντικών τιμών $Y_j,\ j>N$

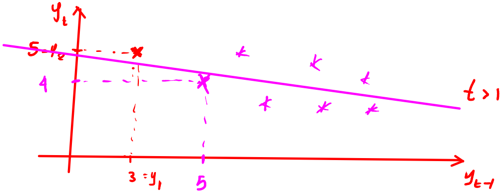
- Θα μελετήσουμε τη γραμμική συσχέτιση μεταξύ των τυχαίων μεταβλητών Y_t



Πρώτη σκέψη: $Y_t = A + Bt + \varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma^2)$

Δευτερη σκέψη: $Y_t = A + BY_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma^2)$

$y_{t-1} \rightarrow \hat{y}_t \qquad \hat{y}_t = \alpha + b y_{t-1}$



- Αρχικά θεωρούμε το πιθανοθεωρητικό μοντέλο

$$Y_t = A + BY_{t-1} + \epsilon_t \leftarrow$$

Στην πραγματικότητα είναι ένα μοντέλο απλής γραμμικής παλινδρόμησης!

Νέο Dataset

$$\{(1, y_1), (2, y_2), \dots, (N, y_N)\} \longrightarrow \{(\overset{x}{y_1}, \overset{y}{y_2}), (y_2, y_3), \dots, (y_{N-1}, y_N)\}$$

N ζεύγη

N-1 ζεύγη!

Συντελεστής γραμμικής συσχέτισης (Pearson)

$$ACF(1) = r = \frac{SS_{Y_t, Y_{t-1}}}{\sqrt{SS_{Y_t, Y_t} SS_{Y_{t-1}, Y_{t-1}}}}$$

Στην γραμμική παλινδρ

$$r = \frac{SS_{XY}}{\sqrt{SS_{XX} SS_{YY}}}$$

- Ανάλογα για k μη αρνητικό ακέραιο θεωρούμε το μοντέλο

$$Y_t = A + BY_{t-k} + \epsilon_t, \quad k > 0$$

Συνάρτηση Αυτόσυσχέτισης (Auto-Correlation Function)

$$\mathrm{ACF}(k) = \frac{SS_{Y_t,Y_{t-k}}}{\sqrt{SS_{Y_t,Y_t}SS_{Y_{t-k},Y_{t-k}}}}, \quad k > 0$$

Δειγματικό μοντέλο $\rightarrow \hat{y}_t = \alpha + b^{(1)} y_{t-1} + b^{(2)} y_{t-2} + \dots + b^{(k)} y_{t-k}$

Αυτοπαλινδρομικό μοντέλο k τάξης (Auto-Regressive model of order k)

$$\mathrm{AR}(k) \quad : \quad Y_t = A + \sum_{j=1}^{k} B^{(j)} Y_{t-j} + \epsilon_t, \quad k > 0$$

Πολλαπλή γραμμική παλινδρόμηση

Νέο Dataset $\{(y_1, \dots, y_k, y_{k+1}), (y_2, \dots, y_{k+1}, y_{k+2}), \dots, (y_{N-k}, \dots y_{N-1}, y_N)\}$

Συνολικά $N-k$ σημεία του $\mathbb{R}^{k+1}$

$$\underset{\sim}{P} = \begin{bmatrix} \alpha \\ b^{(1)} \\ \vdots \\ b^{(k)} \end{bmatrix} \in \mathbb{R}^{k+1} \qquad \underset{\approx}{X} = \begin{bmatrix} 1 & y_1 & y_2 & \dots & y_k \\ 1 & y_2 & y_3 & & y_{k+1} \\ \vdots & & & & \\ 1 & y_{N-k} & y_{N-k+1} & \dots & y_{N-1} \end{bmatrix} \in \mathbb{R}^{N-k,\, k+1} \qquad \underset{\sim}{y} = \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^{N-k}$$

Λύνω το γραμμικό σύστημα

$(X^T X) P = X^T y$ (μορφή : $Ax = b$)

$P = (X^T X)^{-1} X^T y$

Συνάρτηση Μερικής Αυτόσυσχέτισης (Partial Auto-Correlation Function)

- Ποσοτικοποιεί την άμεση γραμμική επίδραση του Y_{t-k} στο Y_t

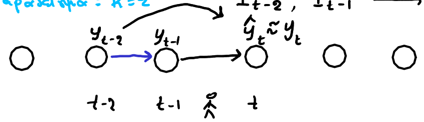
? ACF(2)? ACF(1)

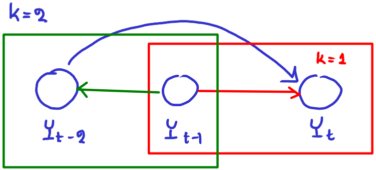
$$\mathrm{PACF}(k) = \ldots$$

Παράδειγμα: k=2

$Y_{t-2}, Y_{t-1} \longrightarrow Y_t$

$$\hat{y}_t = \alpha + b^{(1)} y_{t-1} + b^{(2)} y_{t-2}$$





I: $Y_t = A_1 + B_1 Y_{t-1} + (\varepsilon_1)_t$

II: $Y_{t-2} = A_2 + B_2 Y_{t-1} + (\varepsilon_2)_{t-2}$

χρονολογική σειρά

I: $\hat{y}_t = \alpha_1 + b_1 y_{t-1} \Rightarrow (e_1)_t = y_t - \hat{y}_t$

II: $\hat{y}_{t-2} = \alpha_2 + b_2 y_{t-1} \Rightarrow \underbrace{(e_2)_{t-2} = y_{t-2} - \hat{y}_{t-2}, \; t = 3, \dots, N+2}_{\text{χρονολογική σειρά.}}$

III: $(E_1)_t = A_3 + B_3 (E_2)_{t-2} + (\varepsilon_3)_t \quad \rightsquigarrow \quad (\hat{e}_1)_t = \alpha_3 + b_3 (e_2)_{t-2}$

$$r = \frac{SS_{e_1 e_2}}{\sqrt{SS_{e_1 e_1} SS_{e_2 e_2}}} = PACF(2)$$

Την πραγματική πληροφορία
που δίνει η χρονική στιγμή $t-2$