

MS-E2129 Systemien identifiointi

11. kotitehtävä

Verrataan kahta vaihtoehtoista mallia systeemille, jonka kohinaisesta ulostulosta $y(t)$ on kerätty N näytettä. Mallissa M_1 on d_1 parametria ja mallissa M_2 on d_2 parametria, $d_2 > d_1$. Mallin M_1 ennustevirheen neliösumma on V_N^1 ja mallin M_2 on V_N^2 . Tässä

$$V_N^i = \frac{1}{N} \sum_{t=1}^N \varepsilon(t, \theta^i)^2$$
$$\varepsilon(t, \theta^i) = y(t) - \hat{y}(t, \theta^i).$$

Mallinvalintaan käytetyssä F-testissä käytetään testisuuretta

$$x = N \frac{V_N^1 - V_N^2}{V_N^2}.$$

Testisuureen voidaan osoittaa noudattavan χ^2 -jakaumaa vapausastein $d_2 - d_1$, i.e., $x \sim \chi^2(d_2 - d_1)$. Näin ollen vähemmän parametreja sisältävä malli M_1 valitaan, jos

$$N \frac{V_N^1 - V_N^2}{V_N^2} < \chi_\alpha^2(d_2 - d_1)$$

jossa $\chi_\alpha^2(d_2 - d_1)$ on määritelty seuraavan yhteyden kautta:

$$\alpha = P(x > \chi_\alpha^2(d_2 - d_1))$$

Tämä voidaan myös kirjoittaa muodossa

$$V_N^1 < V_N^2 \left(1 + \frac{1}{N} \chi_\alpha^2(d_2 - d_1) \right). \quad (1)$$

a) Mallinvalintaan käytettävät informaatiokriteerit voidaan kirjoittaa muodossa (huom: poikkeaa hieman luentokalvoilla esitetystä)

$$f_N^i = V_N^i (1 + \beta(N, d_i)) \quad (2)$$

jossa funktion $\beta(N, d_i)$ arvo kasvaa parametrien lukumäärän kasvaessa ja lähestyy nollaa kun $N \rightarrow \infty$.

Osoita, että informaatiokriteerin (2) käyttö mallinvalintaan voidaan tulkita F-testin (1) käyttämiseksi mallinvalintaan tietyllä α :lla, jolle pätee

$$\chi_\alpha^2(d_2 - d_1) = N \frac{\beta(N, d_2) - \beta(N, d_1)}{1 + \beta(N, d_1)}$$

b) Final prediction error (FPE)-informaatiokriteeri voidaan kirjoittaa muodossa

$$f_N^i = V_N^i \frac{1 + d_i/N}{1 - d_i/N}.$$

Osoita, että FPE-informaatiokriteerille $\beta(N, d_i) = 2d_i/(N - d_i)$.

c) Oletetaan, että $d_2 - d_1 = 1$ ja N on hyvin suuri. Oletetaan lisäksi, että oikea systeemiä kuvaava malli on M_1 . Millä todennäköisyydellä valitaan FPE-kriteeriä käytettäessä virheellisesti malli M_2 ?