Identifiointiprosessi

Alustavia kokeita

- Koesuunnittelu, identifiointikoe
- Mittaustulosten / datan esikäsittely
- Ei-parametriset menetelmät:

  - Transientti-, korrelaatio-, taajuus-, Fourier- ja spektraalianalyysi
    => askel-, impulssi- ja taajuusvaste, ajatuksia esim. kertaluvuista
- Parametrien estimointi – mallirakenteen (rakenneparametrit) valinta
- **Mallin arviointi:**

  - Simulointi, navat, erilaiset vasteet, jne...
- **Mallin validointi:**

  - Parametrien merkitsevyys, residuaalianalyysi, jne...

Mallin validointi

- Koe suunniteltu, data kerätty, parametrit estimoitu, mallirakenne valittu
- Onko malli ”hyvä”?
  - Onko malli käyttötarkoitukseensa tarpeeksi hyvä? (tärkein pointti)
  - Onko malli sopusoinnussa havaitun datan kanssa? (validointi keskittyy tähän)
  - Kuvaako malli todellista systeemiä? (mahdotonta vastata pitävästi, validoinnilla tukea antavia tuloksia)
- Validointi = verrataan mallia mahdollisimman suureen todellisesta systeemistä peräisin olevaan tietomäärään

Validoinnin motivaatio

- Usein mahdotonta, kallista tai vaarallista kokeilla tosikäytössä kaikkia mahdollisia malleja
- Luottamus malliin luotava off-line
- Tärkeä mittatikku mallin käyttötarkoitus:

  Voidaanko mallinnusprosessin motivaationa ollut ongelma ratkaista mallin avulla?
- Näkökulmia:
  - parametriestimaattien ominaisuudet
  - sisäänmeno-ulostulokäyttäytyminen
  - mallin redusoinnin vaikutus
  - simulointi ja ennustaminen
  - residuaalianalyysi

Parametriestimaattien ominaisuudet

- Rakenteelliset mallit: ovatko parametriestimaatit järkeviä?
  – esim. kitkakertoimien merkit oikein
- Parametrien herkkyysanalyysi: simulointi hieman poikkeutetuin parametriarvoin (esim. luottamusvälin sisällä)
  – suuria poikkeamia => lisäidentifiointi tarpeen
- Black box-mallin parametriestimaattien luottamusvälit:
  – Estimaatit asymptoottisesti normaalijakautuneet => t-testattavissa
  – Nolla sisältyy luottamusväliin => parametri turha
  – Kaikkien estimaattien varianssit ja kovarianssit suuria => selittäjien kovarianssimatriisi on lähes singulaarinen (liian korkea kertaluku?)

Sisäänmeno-ulostulokäyttäytyminen

- Black box –malleilla tärkeä
- Lineaariset mallit => Boden diagrammi tärkeässä osassa
- Epälineaariset => simulointi
- Jos malli ei kuvaa systeemiä, saadaan estimoitaessa approksimaatio, joka riippuu
  - Koeolosuhteista
  - Esisuodatuksesta
  - Mallirakenteesta

=>Parametristen mallien Boden diagrammien vertailu erilaisilla datoilla, esisuodatuksilla ja mallirakenteilla datasta estimoituun taajuusvaste-estimaattiin hyvä työkalu

Simulointi ja ennustaminen

- Usein mallin käyttötarkoitus
  - tärkeä rooli
- Simulointi riippumattoman datan ohjauksella ja riippumattoman ulostulon ennustaminen
  - nähdään suoraan millaisia ilmiöitä malli kuvaa, millaisia ei
- Mallin redusointi:

  jos kertaluvuiltaan alennettu malli tuottaa oleellisesti saman sisäänmeno-ulostulokäytöksen, malli oli liian monimutkainen

Residuaalianalyysi

- Ennustevirheet kuvaavat sitä osuutta datassa, jota malli ei pysty kuvaamaan

  $\varepsilon(t)=y(t)-\hat{y}(t,\hat{\theta})$

  – ennustevirhe (”parametrien funktio”) = residuaali (”numeroita annetulla parametrillä”)

- Yksinkertaiset testit:

  Olkoon meillä N arvoa (estimointi- tai validointidataa)

  – $\max_N |\varepsilon(t)|$, $\frac{1}{N}\sum_{i=1}^{N}\varepsilon^2(t)$

  – ”arvio maksimi- tai keskivirheistä; jos asiat eivät muutu, mallin ei tulisi koskaan tuottaa suurempia poikkeamia”

Residuaalien valkoisuus

- Oleellista on, että
  - residuaalit eivät riipu ohjauksesta u(t) ja
  - residuaalit eivät riipu toisistaan

=> Ristikovarianssit oleellisesti nollia $\hat{R}_{\varepsilon u}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\varepsilon(t)u(t-\tau)$

- intuitio: jos ennustevirhe riippuu u(t)stä, ulostulo olisi voitu ennustaa paremmin ottamalla riippuvuus huomioon

=> Autokovarianssit oleellisesti nollia $\hat{R}_{\varepsilon}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\varepsilon(t)\varepsilon(t-\tau)$

- intuitio: jos ennustevirhe riippuu menneistä virheistä, se olisi voitu ennustaa paremmin

Tilastollinen testaaminen 1/2

- Jos ohjaukset ja residuaalit ovat riippumattomia, ovat ristikorrelaatiot
$$x_\tau = \hat{R}_{\varepsilon u}(\tau) / \sqrt{R_\varepsilon(0) R_u(0)}$$
asymptoottisesti normaalijakautuneita N(0,1/N)
- Suure
$$y = N \sum_{i=1}^{m} x_i^2 \xrightarrow{dist} \chi^2(m)$$
- Saadaan luottamusvälit
  - yksittäisille ristikorrelaatioille (95%: $|x_\tau| <= 1.96/N^{0.5}$
  - kaikille ristikorrelaatioille tiettyyn viiveeseen asti kerralla
- Huom! Korrelaatio negatiivisella τ:n arvolla voi olla merkki takaisinkytkennästä, ei väärästä mallirakenteesta

Tilastollinen testaaminen 2/2

- Vastaavasti autokorrelaatioille

$$x_\tau = \hat{R}_\varepsilon(\tau) / R_\varepsilon(0)$$

- Pätee oletuksella että residuaalit ovat valkoista kohinaa:
  - yksittäinen autokorrelaatio asymptoottisesti normaalijakautunut N(0,1/N) ja

$$y = N\sum_{i=1}^{m} x_i^2 \xrightarrow{dist} \chi^2(m)$$

- Saadaan luottamusvälit
  - yksittäisille autokorrelaatioille (95%: $|x_\tau| <= 1.96/N^{0.5}$
  - kaikille autokorrelaatioille tiettyyn viiveeseen asti kerralla

Muita testejä

- Residuaalien jakauma symmetrinen?
- H_0: $\varepsilon(t)$ riippumaton ja symmetrisesti jakautunut

  =>

  X_N= "$\varepsilon(t)$:n etumerkin muutosten lkm sarjassa $\varepsilon(1),\ldots,\varepsilon(N)$" asymptoottisesti normaalijakautunut $N(N/2,N/4)$
- Vinouden laskeminen residuaalien histogrammille
- Onko residuaalit normaalijakautuneet? (ks. TAP, luento nro. 4)
  - χ^2-yhteensopivuustesti
    - Otoksen vertaaminen annettuun (normaali) jakaumaan
  - Bowmanin ja Shentonin testi
    - Testisuure riippuu otoksen huipukkuudesta ja vinoudesta
  - Wilkin ja Shapiron testi
    - Rankit-plot –kuvio ja korrelaatikertoimien neliö

Identifiointiprosessi