Sztochasztikus rendszerek matematikája

4. gyakorlat

2017. október 24. kedd: 2-4; 4-6 és október 26. csütörtök: 2-4

[A] Főkomponenanalízis
[B] Idősorelemzés

[A] Főkomponenanalízis

A főkomponens analízis esetén a p dimenziós megfigyeléseket az első q score alapján közelítjük. Az ortogonális *score mátrix*ot a változónként standardizált adatokból a megfigyelések korreláció mátrixának egységhosszú sajátvektoraiból alkotott *loading mátrix*szal transzformálva kaphatjuk meg. A szkórok alapján vett közelítés jóságát a felhasznált sajátértékek aránya méri.

A főkomponens analízis lépései

- a minta legyen adott az $n \times p$ méretű X adatmátrix
- legyen a p dimenziós a vektor az adatok változónkénti átlaga
- legyen X_0 az adatok centrált változata, ekkor az X_0 minden oszlopának az átlaga 0
- állapítsuk meg a minta kovariancia mátrixát: $V = X_0^T X_0 / n$
- a kovariancia mátrix alapján az adatok variancia vektora $s^2 = diag(V)$
- vegyük az adatok korreláció mátrixát $R = diag(s^{-1})\ V\ diag(s^{-1})$

  legyen a p dimenziós λ vektor csökkenő sorrendben a korrelació mátrix sajátértékei, és $L = (L_1, ..., L_p)$ a megfelelő $n \times p$ méretű, egységhosszú sajátvektorokból álló mátrix.
- az $R = \sum_{k=1}^{p} \lambda_k L_k L_k^T$, az $L^T L = L L^T = I$ és az $||R||^2 = ||\lambda||^2$,
- az $R_{(1,...,q)} = \sum_{k=1}^{q} \lambda_k L_k L_k^T$ az R közelítése: $||R - R_q|| = \sum_{k=q+1}^{p} \lambda_k^2$
- a $n \times p$ méretű $S = X_0 \cdot diag(1/s) \cdot L$ mátrix a p ortogonális főkomponens szkór
- szkórok 0 átlagúak, ortogonálisak: $X_0^T X_0 = I$, a varianciáikból álló vektor a λ
- a szkórokból az $S \cdot L^T \cdot diag(s) = X_0$ transzformációval adódik az X_0 centrált adatmátrix
- ha $S_{(1,...,q)}$ az S-nek, $L_{(1,...,q)}$ pedig az L-nek az első q oszlopa,

  akkor az $n \times p$ méretű $X_{(1,...,q)} = S_{(1,...,q)} \cdot L_{(1,...,q)}^T \cdot diag(s)$ az X_0 közelítése, ahol

  a közelítés hibája $||(X_0 - X_{(1,...,q)}) \cdot diag(1/s)||^2 / n = \sum_{k=q+1}^{p} \lambda_k$
- vagyis az $X_{(1,...,q)}$ közelítés jóságát a $\sum_{k=1}^{q} \lambda_k / \sum_{k=1}^{p} \lambda_k$ arány méri

[B] Idősorelemzés

Legyenek $x_1, ..., x_T$ a $\xi_1, ..., \xi_T$ megfigyelt változósorozat értékei.
Tételezzük fel, hogy az adatok a

$$x_t = (a + bt) \cdot s_t \cdot e_t \qquad t = 1, ..., T$$

három komponensű multiplikatív modellnek megfelelően épülnek fel, ahol az e_t a független 1 várhatóértékű, azonos szórású ε_t, $t = 1, ..., T$ zajsorozat értékei. A modellben a és b két fix, ismeretlen valós szám, ezek határozzák meg a folyamat lineáris trendjét. Az s_t szezonális komponens p darab, szintén ismeretlen, $c_1, ..., c_p$ valós szám periódikus ismétlődése.
Vagyis a modell szerint a folyamat értékei az $a + bt$ lineáris trendtől
 determinisztikusan a p periódushosszú multiplikatív s_t szezonális trend szerint
 véletlenszerűen pedig szintén multiplikatívan a független ε_t folyamat szerint
térnek el. A modell additív változata szerint: $x_t = (a + bt) + s_t + e_t$, és az ε_t várhatóértéke 0.

A multiplikatív modell paramétereinek becslés módszere

A modell elemeit a következő sorrendben becsüljük meg:
 először készítünk hosszú simítással egy nyers becslést az adatok $a + bt$ trendjére;
 a nyers trend alapján előbecslést kapunk az s_t szezonális trendre;
 az s_t szezonális trend fázisonkénti értékét az előbecslés fázisonkénti átlagával becsüljük;
 az így nyert fázisátlagok periódikus ismétlése az s_t szezonális trend végső becslése: $\widehat{s}_t$;
 az $\widehat{s}_t$ alapján előbecslést készítünk a folyamat lineáris trendjének értékeire;
 az előbecsült trendértékek idő szerinti lineáris regressziója adja a folyamat lineáris trendjét;
 a $\widehat{a} + \widehat{b}t$ lineáris és $\widehat{s}_t$ szezonális trend alapján becslés adódik a folyamat hibatagjára;
 a becsült hibatagok alapján a véletlen hiba szórását becsülhetjük.

A modell illesztés és előrejelzés műveletei és paraméterei

Összesen 16 negyedéves adat áll rendelkezésre.
A folyamat periódushossza 4 és 4 havi előrejelzést kell készíteni.
A nyers trendet egy $\mathtt{MA(2)} \otimes \mathtt{MA(4)} = \mathtt{(1, 2, 2, 2, 1)/8}$ simítás állítja elő.
A megfigyelések és a nyers trend hányadosa a szezonális trend előbecslése.
A szezonális trend előbecslésének fázisonkénti átlagai adják a ciklikus trend periódusát.
Az eredeti megfigyelések és a ciklikus trend hányadosa a lineáris trend előbecslése.
A lineáris trend a lineáris trend előbecslésének az időpontok szerinti lineáris regressziója.
A folyamat előrejelzése a becsült lineáris trend és ciklikus kompones adott időpontbeli értéke.