

Aegridade analüüs

Praktikum nr. 5, 2016

Käsitletavad teemad: aegridade sisselugemine failidest, sobiva mudelite tüübi esialgne hindamine autokorrelatsioonide ja osaautokorrelatsioonide põhjal.

Tänases praktikumis vaatleme lihtsamat tüüpi aegridade identifitseerimist. Kõigepealt, liikuva keskmisega mudelid $MA(q)$ on kujul

$$Z_t - \mu = A_t - \sum_{i=1}^q \theta_i A_{t-i},$$

kus A_t on vähemalt teist järku statsionaarsed, tsentreeritud ja mittekorreleeritud. Küllalt lihtne on veenduda, et sellise rea teoreetilised autokorrelatsioonid võrduvad nulliga alates järgust $q + 1$ (sest Z_{t-k} ja Z_t avaldises on $k \geq q + 1$ korral erinevate indeksitega A -d ning need on mittekorreleeritud).

Autoregressiivsed mudelid $AR(p)$ esituvad kujul (tähistades $\tilde{Z}_t = Z_t - EZ_t$)

$$\tilde{Z}_t = \sum_{i=1}^p \phi_i \tilde{Z}_{t-i} + A_t,$$

kus A_t rahuldavad samu nõudeid, mis $MA(q)$ mudeli korral ning on mittekorreleeritud ka kõikide varasemate Z_t väärtustega. Selliste ridade puhul annavad rohkem infot osaautokorrelatsioonid, mille leidmisel eemaldatakse Z_t -st ja Z_{t-k} -st vahepealsete Z väärtuste mõju ning leitakse korrelatsioon nii saadud suuruste vahel. Kui eemaldamisel kasutatakse vähemalt p vahepealset väärtust, siis nende mõju Z_t -le (ehk parim ennustus Z_t väärtusele vahepealseid Z väärtuseid kasutades) on $\sum_{i=1}^p \phi_i \tilde{Z}_{t-i}$, mille eemaldamisel jääb järele A_t . Et aga A_t on mittekorreleeritud kõikide varasemate Z väärtustega, siis teoreetiline osaautokorrelatsioon on alates järgust $p + 1$ võrdne nulliga.

Mudeli valikul veendutakse kõigepealt, et tegemist võib olla statsionaarse aegrega ning siis valitakse alustuseks sellist tüüpi mudel, millel on kõige vähem parameetreid. Statsionaarsuse osas ei pruugi olla väga lihtne lihtsalt vaatluse põhjal otsuseid teha. Kuna vähemalt teist järku nõrgalt statsionaarse rea puhul peaks vaatlused vastama juhuslikele suurustele, millel on sama keskvärtus ja ka mille erinevustel sellest keskmisest (või erinevuste ruutudel) on sama (ajas muutumatu) keskvärtus, siis loeme selgelt mittestatsionaarseks read, mille puhul on üleüldise ühise keskmise olemasolu väga väheusutav (nt kui väärtused on ajas selgelt kasvanud või kahanenud) või mille varieeruvus keskmisest on ajas selgelt muutuv. Käesolevas praktikumis tutvumegi esialgse mudeli valiku protsessiga. Failis praktikum5.csv on 7 erinevat aegrida. Nendest aegridadest osad on mittestatsionaarsed, osad on aga genereeritud mitmesuguste AR tüüpi või MA tüüpi mudelite põhjal. Ülesandeks on need andmed R -i sisse lugeda ning graafiku, autokorrelatsioonide ja osaautokorrelatsioonide põhjal püstitada hüpotees iga aegrea korral, mis tüüpi aegrega võib olla tegemist. R -is saab osaautokorrelatsioone arvutada funktsiooniga `pacf()`.