

Idősorok és többdimenziós statisztika

9. előadás

Zempléni András

2019.11.12

Az ML becslés levezetése

Térjünk vissza Q-hoz:

$$\frac{\partial}{\partial \mu} Q(\mu, \alpha_1, \dots, \alpha_p) = 0 \Rightarrow$$

$$\sum_{t=p+1}^T [X(t) - \mu - \alpha_1(X(t-1) - \mu) - \dots - \alpha_p(X(t-p) - \mu)] = 0 (*)$$

$$\frac{\partial}{\partial \alpha_j} Q(\mu, \alpha_1, \dots, \alpha_p) = 0 \Rightarrow$$

$$\sum_{t=p+1}^T [X(t) - \mu - \alpha_1(X(t-1) - \mu) - \dots - \alpha_p(X(t-p) - \mu)] \cdot (X(t-j) - \mu) = 0 \quad (0.1)$$

$$j = 1, \dots, p$$

ML becslés μ -re

Az előző egyenletekből:

$$\hat{\mu} = \frac{\bar{X}_1 - (\hat{\alpha}_1 \bar{X}_2 + \dots + \hat{\alpha}_p \bar{X}_p)}{1 - (\hat{\alpha}_1 + \dots + \hat{\alpha}_p)},$$

ahol

$$\bar{X}_{j+1} = \frac{1}{T-p} \sum_{t=p+1-j}^{T-j} X(t), \quad j = 0, 1, \dots, p-1$$

Ha itt $p \ll T \Rightarrow \bar{X}_{j+1} \cong \bar{X} \forall j$ és így

$$\hat{\mu} = \bar{X}$$

A (0.1) egyenletekből kapjuk az együtthatók becslését, ezt azonban zárt alakban nem adjuk itt meg.

ML becslés az AR(1) esetben

Tegyük fel, hogy μ ismert és $\mu = 0$. Ebben az esetben a (0.1) egyenlet megoldása egyszerűen felírható

$$\hat{\alpha} = \frac{\sum_{t=2}^T X(t) \cdot X(t-1)}{\sum_{t=2}^T X^2(t-1)}$$

Ide $X(t)$ előállítását beírva:

$$\hat{\alpha} = \alpha + \frac{\sum_{t=2}^T \varepsilon(t) \cdot X(t-1)}{\sum_{t=2}^T X^2(t-1)}$$

A becslés szórása a regressziónál

Az AR(1) esetben a paraméterbecslés szórásának meghatározása kapcsán próbáljuk megérteni, hogy miért kell másként gondolkodni az autoregresszió, mint egy szokásos regresszió esetén. Ha ez egy közönséges regresszió lenne, akkor az $X(t-1)$ magyarázó változó megfigyelt értékeit konstansokként, és α -t is ismertként kezelve (azaz $\hat{\alpha}$ feltételes eloszlását vizsgálva) itt a regressziós hibatarok lineáris kombinációja állna, ami maga is normális lenne, és a szórást számítva:

$$\hat{\alpha} = N\left(\alpha, \frac{\sigma_\varepsilon^2}{\sum_{t=2}^T X^2(t-1)}\right)$$

adódna. Azonban az AR(1) modellben, ha a magyarázó változó, és a paraméter adott, akkor a függő változó is $\Rightarrow \varepsilon(t)$ is számolható és így $\hat{\alpha}$ egyáltalán nem véletlen!

A fenti érvelés tehát nem használható.

A becslés szórása az AR(1) modellnél

Nézzük hát másképp: $\hat{\alpha}$ -ban a számlálóbeli szorzat tagjai függetlenek, ezért a számláló 0 várható értékű.

Bár az összeg tagjai nem függetlenek, de korrelálatlanok:

$$\begin{aligned} E(\varepsilon(t) \cdot X(t-1) \cdot \varepsilon(t-1) \cdot X(t-2)) &= \\ &= E\varepsilon(t) \cdot E(X(t-1) \cdot \varepsilon(t-1)X(t-2)) = 0 \end{aligned}$$

(mert $\varepsilon(t)$ független a többi tényezőtől)

ezért az összeg szórásnégyzete a szórásnégyzetek összege, így

$$D^2 \sum_{t=2}^T \varepsilon(t) \cdot X(t-1) = (T-1) \cdot \sigma_\varepsilon^2 \cdot \sigma_X^2$$

Közelítés

A nevező $T-1$ -gyel osztva a tapasztalati szórásnégyzet, "nagy" mintára tehát $(T-1) \cdot \sigma_X^2$ -tel becsülhető, így

$$\begin{aligned} D^2 \hat{\alpha} &\cong D^2 \left[\frac{\sum_{t=2}^T \varepsilon(t) \cdot X(t-1)}{(T-1)\sigma_X^2} \right] = \\ &= \frac{1}{(T-1)^2 \cdot \sigma_X^4} \cdot D^2 \left[\sum_{t=2}^T \varepsilon(t)X(t-1) \right] = \frac{(T-1)\sigma_\varepsilon^2\sigma_X^2}{(T-1)^2\sigma_X^4} = (*) \end{aligned}$$

$$\text{itt } \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1-\alpha^2} \Rightarrow (*) = \frac{1-\alpha^2}{T-1}$$

Ez alapján

$$\hat{\alpha} \sim N\left(\alpha, \frac{1-\alpha^2}{T-1}\right)$$

Fontos kivétel, ha $\alpha = 1$, illetve ha α u.n. közel nemstacionárius, azaz α közel van, vagy valamilyen értelemben tart 1-hez.

Paraméterbecslés az AR(p) modellben

Térjünk vissza az AR(p) modellhez és a Q funkcionálhoz.

A μ szerinti deriváltból kapott $\hat{\mu}$ a "véghatás" elhanyagolásával az \bar{X} -ra egyszerűsödött. Írjuk ezt be az α_j szerinti deriváltból adódó (0.1) egyenletekbe. Így

$$\begin{aligned} \sum_{t=p+1}^T \left\{ X(t) - \bar{X} - \left(\hat{\alpha}_1(X(t-1) - \bar{X}) + \dots + \hat{\alpha}_p(X(t-p) - \bar{X}) \right) \right\} \times \\ \times (X(t-j) - \bar{X}) = 0 \end{aligned}$$

Megint elhanyagolva a kezdeti hatást felhasználjuk, hogy

$$\sum_{t=p+1}^T (X(t-k) - \bar{X}) \cdot (X(t-j) - \bar{X}) \cong T \cdot \hat{R}(j-k)$$

Ezt beírva, egyenletünk így alakul:

$$\hat{R}(j) = \hat{\alpha}_1 \hat{R}(j-1) + \dots + \hat{\alpha}_p \hat{R}(j-p), \quad j = 1, \dots, p$$

Az autokovarianciák becslése

Azaz a likelihood egyenletekből a véghatások elhanyagolásával a Yule-Walker egyenleteket kapjuk, a becsült autokovarianciákkal. Mátrix alakban:

$$\widehat{R}_p = \widehat{\mathfrak{R}}_p \widehat{\alpha}$$

ahol

$$\widehat{R}_p = (\widehat{R}(1), \dots, \widehat{R}(p))^T, \quad \widehat{\alpha} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)^T$$

és

$$\widehat{\mathfrak{R}}_p = \begin{bmatrix} \widehat{R}(0) & \widehat{R}(1) & \dots & \widehat{R}(p-1) \\ \widehat{R}(1) & \widehat{R}(0) & \dots & \widehat{R}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{R}(p-1) & \widehat{R}(p-2) & \dots & \widehat{R}(0) \end{bmatrix}$$

A szórás becslése

Azaz

$$\widehat{\alpha} = \widehat{\mathfrak{R}}_p^{-1} \widehat{R}_p.$$

A Yule-Walker egyenletekből a zaj szórásnégyzetének becslése:

$$\widehat{\sigma}_\varepsilon^2 = \widehat{R}(0) - \widehat{R}_p^T \widehat{\mathfrak{R}}_p^{-1} \widehat{R}_p = \widehat{R}(0) - \widehat{R}_p \cdot \widehat{\alpha}.$$

Aszimptotikus eloszlás: Gyakran a legkisebb négyzetes becslés sokkal nagyobb szórású, mint az ML becslés. Itt nem így van, az OLS és a ML közelítőleg ugyanolyan eloszlású:

$$\widehat{\alpha}_p \approx N(\alpha, \frac{\sigma_\varepsilon^2}{T} \mathfrak{R}_p^{-1})$$

Ez Mann és Wald (1943) eredménye.

A spektrálsűrűség becslése

Periodogram:

$$I_T(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^T x_t e^{-it\lambda} \right|^2.$$

- Ha az ún. Fourier-frekvenciákra számoljuk ki ($\lambda = 2\pi k/T$), akkor éppen a megfigyelt idősorra kiszámolt empirikus spektrálsűrűségfüggvényt kapjuk
- A kapott becslés gyorsan lecsengő autokovarianciájú stacionárius Gauss folyamatra aszimptotikusan torzítatlan
- Szükséges lehet a kapott I_T függvény simítása

A tapasztalati autokorreláció eloszlása

Hogyan ellenőrizhető a fehér zaj tulajdonság?

- **A tapasztalati autokorreláció függvénynek** "szignifikánsan" 0-nak kell lennie
- Mivel tudjuk, hogy ezek a becslések gyakorlatilag függetlenek (Nem csak normális fehér zajra!) és eloszlásuk közel $N(0, \frac{1}{T})$, ha igaz a nullhipotézis így a 95 %-uknak a $\pm \frac{1.96}{\sqrt{T}}$ -s határokon belül kell lennie - az 1,96 a standard normális eloszlás 95 %-os kvantilise. Így például 40 "lag"-re elkészítve, ha kettő-három **kicsit**, vagy ha egy **nagyon** kilóg, akkor elutasítjuk a fehér zaj hipotézist.

Portmanteau próbák

Az alábbi próbákat gyakran összefoglaló néven Portmanteau próbákként említik. Egy T elemű idősor megfigyelése mellett a nullhipotézis mindvégig az, hogy a megfigyelt folyamat fehér zaj.

Klasszikus Portmanteau próba, vagy Box-Pierce teszt:

- Az első h "lag"-re véve a becült autokorrelációk négyzetösszegét, ennek T -szerese:

$$Q = T \cdot \sum_{\tau=1}^h \hat{r}^2(\tau)$$

úgy kell viselkedjen, mint h db független $N(0,1)$ -es négyzetösszege, ezért ez a statisztika χ_h^2 eloszlású. Ezt teszteljük $1 - q$ szignifikancia szinten.

- Ha $Q > \chi_{h,q}^2$ a χ_h^2 eloszlás q -kvantilisével, akkor elutasítjuk a 0 hipotézist.
- Ez egy elég gyenge próba.

Ljung – Box próba

Ljung és Box finomítja az előző eljárást.

- A nagyobb "lag"-ekhez tartozó autokorrelációk súlyát kissé megnöveljük:

$$Q_{LB} = T \cdot (T + 2) \sum_{\tau=1}^h r^2(\tau) / (T - \tau)$$

ez érzékenyebb és ezt jobban is közelíti a χ_h^2 eloszlás.

- h megválasztására Hyndman és Athanasopoulos $h = 10$ -et javasol nem szezonális adatra, és $2m$ -et m -szezonális adatra. Patrick Burns $h \leq 0.05 \cdot T$ -t javasol.
- Ez a két próba **nem csak** normális fehér zaj mellett jó.

Nemparaméteres regresszió

Running line (futó egyenes):

Simítás, általában a trend torzított becslése.

A simítás során 2 alapkérdés van:

- Hogyan "átlagoljunk" egy bizonyos környezetben
- Hogyan válasszuk meg a környezetet
- k -legközelebbi szomszéd - a legközelebbi k pont
- Szimmetrikus k -legközelebbi szomszéd - az egyik és másik oldalon is $\frac{k}{2}$, $\frac{k}{2}$ pont. Egyfajta mozgó ablak.

Mozgó ablakolás

Legyen:

$$Y = \mu(X) + \varepsilon,$$

ahol μ sima függvény, és legyen mintánk Y -ra és X -re.

- Running line: egy mozgó ablakot választunk, és az ablakon belül egy egyszerű lineáris regressziót alkalmazunk Y -ra X -szel.
- Y_i -t az X_i alapján abból az ablakból becsüljük, amelynek ő van a közepén.
- Pl. $k=11$ -re Y_{14} -et az $(X_9, Y_9), \dots, (X_{19}, Y_{19})$ ablakból, azaz ezen párokra végzünk regressziót, és ennek együtthatóival predikáljuk Y_{14} -et X_{14} -ből.
- Ez az eljárás jó irregulárisan megfigyelt idősorra is. Ekkor X_i az idő, t , ami "véletlenszerű", vagyis regresszorként is felfogható.

Lokális regresszió: LOESS

- A Running line és a múlt félévben látott Nadarajah-regresszió kombinációja.
- Minden környezetben súlyozott legkisebb négyzetes illesztés.
- A lokális regresszió célja pl. trend kiszűrése
- A leggyakrabban használt súlyfüggvény:

$$W(t) = \begin{cases} (1 - t^3)^3 & 0 \leq t \leq 1 \\ 0 & \text{különben} \end{cases}$$

ahol t az adott pont távolsága az illesztendő ponttól (0;1)-re skálázva.

- Robusztus regressziót is lehet használni, ha szimmetrikusnak tételezzük fel a zajt normális helyett.

A modellszelekció

- A modellszelekció azt jelenti, hogy a megfigyelt adatainkból nem csak egy 'a priori' adott paraméteres modell ismeretlen paramétereit becsüljük, hanem magát a modellt is.
- Ehhez elsőként meg kell határozni az alkalmas modellek halmazát, melyből aztán valamely kritérium szerint a legjobbat választjuk.
- Hogy mit tekintünk legjobbnak, az a vizsgálat céljától függ.

A rendszelekció mint modellszelekció

- Idősorok esetén a rendbecslés a modellszelekció egy speciális fajtája.
- Tekintsük például alkalmas modelleknek az autoregressziós folyamatok családját. Ekkor a modellválasztás problémája leszűkül az autoregressziós egyenletben szereplő legnagyobb késleltetés meghatározására. Másszóval a paraméterek dimenzióját, a folyamat rendjét becsüljük.
- A rendbecslési eljárások rendkívül sokfélék, a témával foglalkozó szakirodalom rohamosan bővül. A továbbiakban az információelméleti módszerekre helyezjük a hangsúlyt.
- Egy rövid bevezetés után bemutatjuk az információs kritériumok családját, köztük az Akaike és a Bayes-i kritériumokat.

Modellszelekció hipotézisvizsgálattal

- Modellválasztás céljára alkalmazhatunk egyszerű hipotézisvizsgálatot.
- Tekintsünk elsőként egy két elemű \mathcal{M} halmazt és tegyük fel, hogy M_1 és M_2 egymásbaágyazott, azaz az M_1 modell az M_2 egy speciális eseteként fogható fel.
- Ekkor a modellszelekció annak a H_0 nullhipotézisnek a tesztje, hogy a valódi eloszlás az M_1 szerinti. Amennyiben H_0 -t elvetjük, az M_2 bővebb modellt érdemes választanunk.
- Ha például a vizsgálatunk célja, hogy minél jobb előrejelzést adjunk, a nullhipotézist módosíthatjuk eszerint: H_0 jelentse, hogy az előrejelzés átlagos négyzetes hibája az M_1 modellben kisebb.
- Több egymásba ágyazott modell esetén a fenti eljárás kiterjeszhető tesztek láncolatává, az általánostól az egyre specifikusabb modellek felé.

Modellszelekció kritériumok alapján

- A modellszelekciós kritériumok megjelenésével a modellválasztás kérdése hipotézisvizsgálatból becslési problémává alakult át.
- Az illeszthető modellek halmazából azt az eloszlást választjuk, amelyre nézve az adott hibafüggvény a legkisebb értéket veszi fel.
- A legtöbb kritérium a becslült és a valódi folyamat közötti eltérést tekinti kockázatnak. Lineáris modellek esetén a végső előrejelzési hiba (final prediction error) kritérium az előrejelzés átlagos négyzetes hibáját nézi.
- Szintén elterjedt módszer a Parzen-féle CAT kritérium, ami a spektrálsűrűségfüggvények eltérését, az ún. integrált relatív négyzetes hibát minimalizálja autoregressziós folyamatok esetén.

Modellszelekció információs kritériumokkal

- Az információelméleti módszerek a fentiekkel szemben általánosan kiterjeszthetők bármely modell típusra.
- Jelölje $\hat{\theta}_M$ az M modell becslült paramétereit. Hibafüggvényként tekintünk a $P_{\hat{\theta}_M}$ modelleloszlás és az Y adatok tényleges eloszlása közötti Kullback-Leibler távolságot. Ekkor a minimalizálandó függvény az alábbi általános alakba írható:

$$IC(M) = -2 \log f(y, \hat{\theta}_M) + C(n)k_M$$

- ahol $f(y, \hat{\theta}_M)$ az adott modellnek megfelelő likelihood-függvény, $k_M \in \mathbb{Z}$ a modellbeli paraméterek száma, $C(n)$ pedig a megfigyelések számától függő ún. büntetőfaktor. Az általunk vizsgált információs kritériumok mindössze a $C(n)$ faktorban különböznek egymástól.
- A két legelterjedtebb módszer az Akaike információs kritérium (AIC), ahol $C(n) = 2$, valamint a Schwarz-féle bayesi információs kritérium (BIC), melyre $C(n) = \log n$.

MDL és MML kritériumok

- Szintén információelméleti alapokon nyugszik a különböző struktúrájú modellek összehasonlítására is alkalmas minimális leírás hossz (minimum description length), valamint annak bayesi alternatívája a minimális üzenethossz (minimum message length) kritérium. A módszer koncepciója, hogy azt a modellt választja, ami a megfigyelt adatokat a legrövidebb kóddal írja le.
- Érdekes megemlíteni még a David Spiegelhalter által kidolgozott deviancia információs kritériumot, mely bonyolult hierarchikus modellek esetén az optimális paraméterszám meghatározására szolgál. Ez az eljárás elsősorban akkor hasznos, amikor a paraméterek száma igen nagy és összemérhető a mintaelemszámmal.