Idősorok - jegyzet kiegészítés

Az AR(p) folyamatok paraméterbecslése

2016. május 1.

Első a paraméterbecslés és csak utána következik a rendszelekció. (Épp ezért szelekció és nem becslés.)
Az autoregressziós egyenlet:

$$X(t) = \mu + \alpha_1(X(t-1) - \mu) + \ldots + \alpha_p(X(t-p) - \mu) + \varepsilon(t)$$

Így most a várható érték sem feltétlen 0, hanem ismeretlen konstans:

$$EX(t) = \mu$$

Ezzel a modell paraméterek: $(\mu, \alpha_1, \ldots, \alpha_p, \sigma_\varepsilon)$.

Tegyük fel, hogy $P(z) = 1 - (\alpha_1 z + \ldots + \alpha_p z^p)$ gyökei az egységkörön belül vannak, és így létezik az $X(t)$ stacionárius megoldás.

- Az autoregresszió regresszióra hasonlít, de a függő változója $X(t)$ és a független változói ugyanazon $X(t)$ eltoltjai, késleltetettjei, így nem csak a változók, hanem a megfigyelések "esetek", "sorok" is korreláltak.
- Ez a becslést nem, de a tulajdonságait, "jóságát", eloszlását befolyásolja, a szokásos teszteket, konfidencia intervallumokat, szórásnégyzetet érvényteleníti.
- A közönséges legkisebb négyzetek (OLS) módszere a regressziónak megfelelően alkalmazható.

A következő Q kvadratikus funkcionált kell minimalizálni:

$$Q(\mu, \alpha_1, \ldots, \alpha_p) = \sum_{t=p+1}^{N} \varepsilon_t^2 = \qquad (1)$$

$$= \sum_{t=p+1}^{N} [X(t) - \mu - \alpha_1(X(t-1) - \mu) - \ldots - \alpha_p(X(t-p) - \mu)]^2 \longrightarrow min$$

Az 1-beli szummából az $\varepsilon^2(1), \ldots, \varepsilon^2(p)$ tagokat ki kell hagyni, mert $X(0), \ldots, X(-p+1)$ nem megfigyelhető.

Nézzük most a likelihoodot **Gauss** generáló zaj esetén.
Először AR(1)-re:

$$(X(t)-\mu)-\alpha(X(t-1)-\mu)=\varepsilon(t)$$

és $\varepsilon(t)$-k függetlenek egymástól valamint $t \geq 2$-re X(1)-től is, így

$$f(x_1,\ldots,x_T)=f(x_1)\cdot f(\varepsilon_2)\cdot\ldots\cdot f(\varepsilon_T)$$

és ez proporcionálisan:

$$f(x_1,\ldots,x_T)\propto f(x_1)\cdot\frac{1}{\sigma_\varepsilon^{T-1}}\cdot exp\{-\frac{1}{2\sigma_\varepsilon^2}\cdot\sum_{t=2}^{T}\varepsilon_t^2\}$$

Ebből az adott kezdőérték melletti feltételes likelihood, arányosan:

$$f(X(2),\ldots,X(T)\mid X(1))\propto$$

$$\frac{1}{\sigma_\varepsilon^{T-1}}\cdot exp\{-\frac{1}{2\sigma_\varepsilon^2}\cdot\sum_{t=2}^{T}\left[X(t)-\mu-\alpha(X(t-1)-\mu)\right]^2\}$$

AR(p)-re a szummában p+1-től megyünk, és a szumma mögötti szögletes zárójelben $X(t-1),\ldots,X(t-p)$ megfelelő lineáris kombinációja áll.

Ez ugyanannak a $Q(\mu, \alpha_1, \ldots, \alpha_p)$ funkcionálnak a minimalizására vezet, mint az OLS.
Ez ugyanaz a kapcsolat, mint a közönséges regresszió esetén a OLS és a Gauss ML között.
Ha a teljes (nem feltételes) likelihoodot nézzük, akkor mivel $X(1) \sim N(\mu, \frac{\sigma_\varepsilon^2}{1-\alpha^2})$

$$f(x_1, ..., x_T) \propto \frac{\sqrt{(1-\alpha^2)}}{\sigma_\varepsilon} \cdot exp\left\{-\frac{1-\alpha^2}{2\sigma_\varepsilon^2}(x_1-\mu)^2\right\} f(x_2, \ldots, x_T \mid X_1 = x_1)$$

ami már nemlineáris legkisebb négyzetes minimalizálásra vezet, viszont az eltérés a lineáristól kicsi és nagyon gyorsan 0-hoz tart.

Térjünk vissza Q-hoz:

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

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

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

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

$$\cdot (X(t-j) - \mu) = 0 \ (**)$$

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

Az előző egyenletekből:

$$\hat{\mu} = \frac{\overline{X}_1 - (\hat{\alpha_1}\overline{X}_2 + \ldots + \hat{\alpha}_p\overline{X}_p)}{1 - (\hat{\alpha_1} + \ldots + \hat{\alpha_p})},$$

ahol

$$\overline{X}_{j+1} = \frac{1}{T-p}\sum_{t=p+1-j}^{T-j} X(t), \qquad j = 0, 1, \ldots, p-1$$

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

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

A $**$ egyenletekből kapjuk az együtthatók becslését, ezt azonban zárt alakban nem adjuk itt meg.

Vizsgáljuk most az AR(1) egyenletre vonatkozó maximum likelihood becslést. Tegyük fel, hogy μ ismert és $\mu = 0$. Ebben az esetben a ** 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:

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

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 hibatagok lineáris kombinációja állna, ami maga is normális lenne, és a szórást számítva:

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

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.

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:

$$E\varepsilon(t) \cdot X(t-1) \cdot \varepsilon(t-1) \cdot X(t-2) =$$

$$E\varepsilon(t) \cdot EX(t-1) \cdot \varepsilon(t-1)X(t-2) = 0$$

ezért az összeg szórásnégyzete a szórásnégyzetek összege, így

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

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

$$D^2\hat{\alpha} \cong D^2\frac{\sum_{t=2}^{T}\varepsilon(t)\cdot X(t-1)}{(T-1)\sigma_X^2} = \frac{1}{(T-1)^2\cdot\sigma_X^4}\cdot D^2\sum_{t=2}^{T}\varepsilon(t)X(t-1) =$$

$$\frac{(T-1)\sigma_\varepsilon^2\sigma_X^2}{(T-1)^2\sigma_X^4} = \frac{\sigma_\varepsilon^2}{(T-1)\sigma_X^2} = (*)$$

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

Ez alapján

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

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

Térjünk vissza az AR(p) modellhez és a Q funkcionálhoz.
A μ szerinti deriváltból (*) kapott $\hat{\mu}$ a "véghatás" elhanyagolásával az $\overline{X}$ -ra egyszerűsödött.
Írjuk ezt be az α_j szerinti deriváltból adódó (**) egyenletekbe. Így

$$\sum_{t=p+1}^{T} \left\{X(t) - \overline{X} - (\widehat{\alpha_1}(X(t-1) - \overline{X}) + \cdots + \widehat{\alpha_p}(X(t-p) - \overline{X}))\right\} \times$$

$$\times (X(t-j) - \overline{X}) = 0$$

Megint elhanyagolva a kezdeti hatást felhasználjuk, hogy

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

Ezt beírva, egyenletünk

$$\widehat{R}(j) = \widehat{\alpha_1}\widehat{R}(j-1) + \cdots + \widehat{\alpha_p}\widehat{R}(j-p)$$

$$j = 1, \cdots, p$$

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

$$\widehat{R}_p = \left(\widehat{R}(1), \cdots, \widehat{R}(p)\right)^T ,$$

$$\widehat{\alpha} = (\widehat{\alpha}(1), \cdots, \widehat{\alpha}(p))^T \text{ és}$$

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

Ezek az egyenletek azonosak a Yule-Walker egyenletekkel, amelyek az elméleti autokovariacia függvény összefüggését adják az elméleti paraméterekkel.

$$(*) \quad \widehat{\widehat{\sigma}_\varepsilon}^{\,2} = \widehat{R}(0) - \widehat{R}_p^T \widehat{\mathfrak{R}}_p^{-1} R_p = \widehat{R}(0) + \widehat{R}_p \cdot \widehat{a}$$

(Brockwell-Davis könyv 138. oldal)

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{T-2p-1} Q(\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_p) =$$

$$= \frac{T-p}{T-2p-1} \left\{ \hat{R}(0) + \hat{\alpha}_1 \hat{R}(1) + \cdots + \hat{\alpha}_p \hat{R}(p) \right\}$$

A nevező: megfigyelések száma - becsült paraméterek száma, mivel T-p megfigyelés(az első p elvész) és $p+1$ becsült paraméter van.

Ha σ_ε^2-et is ismeretlen paraméternek tekintjük, a variancia **ML becslése** nem azonos a legkisebb négyzetessel:

$$\widehat{\hat{\sigma}_\varepsilon}^2 = \frac{1}{T-p} Q(\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_p) = R(0) + \hat{\alpha}_1 \hat{R}(1) + \cdots + \hat{\alpha}_p \hat{R}(p)$$

Aszimptotikus eloszlás: Gyakran a momentum becslés sokkal jobban szór, mint a ML. Itt nem így van, az OLS és a ML közelítőleg ugyanolyan eloszlású:

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

Ez Mann és Wald (1943) eredménye.

Tapasztalati autokorreláció

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 három próbát gyakran összefoglaló néven Portmanteau próbákként említik. Egy T elemű idősor megfigyelése mellett a null-hipotézis mindháromban az, hogy a megfigyelt folyamat fehér zaj.

Klasszikus Portmanteau próba, vagy Box-Pierce teszt:

- Az első h "lag"-re véve a becsült 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.
- Ha $Q > \chi_h^2(1-\alpha) \Rightarrow$ 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-szezonalitású adatra. Patrick Burns $h \leq 0.05 \cdot T$-t javasol.
- Az eddigi két próba **nem csak** normális fehér zaj mellett jó.

McLeod – Li próba

McLeod és Li tesztje (1983) csak Gauss folyamatra alkalmazható.

- Ha igaz a nullhipotézis, akkor a folyamat Gauss fehér zaj, ami független értékű is, de akkor a négyzete is az, így az is korrelálatlan értékű, azaz fehér zaj, és akkor erre a Ljung-Box teszt használható.
- Az adatok négyzetét veszi: $Y(t) = X(t)^2$ és ennek autokorrelációfüggvény becslését, $\hat{r}_Y(\tau)$-t használja:

$$\tilde{Q} = T \cdot (T+2) \sum_{\tau=1}^{h} \hat{r}_Y^2(\tau)/(T-\tau)$$

- Ez a tesztstatisztika is χ_h^2 eloszlású, de érzékenyebb az előzőeknél, így a próba erősebb.

Függetlenséget ellenőrző tesztek

- Az alábbi három próba megfigyelések függetlenségét teszteli, tetszőleges (azaz nem csak idősor) mintarealizáció esetén.
- Így elvileg a zaj független értékűségét is tesztelhetjük.
- Azonban nem szabad elfeledni, hogy a zajra nincs megfigyelésünk, azt is csak becsültük, általában valamilyen illesztett modell reziduálisaként.
- Önmagában egy elutasító teszt alapján még nem biztos, hogy el kell utasítani a nullhipotézist.

Fordulópont próba

Fordulópont próba: Legyen $y_1, \cdots, y_n$ egy megfigyelt mintarealizáció.
Definíció.: i-ben fordulópont van, ha

$$\begin{cases} y_{i-1} < y_i & \text{és } y_i > y_{i+1} \\ y_{i-1} > y_i & \text{és } y_i < y_{i+1}, \qquad 2 < i < n-1 \end{cases}$$

Mivel a fordulás valószínűsége i-ben $2/3$ ezért a várható fordulatok száma: $(n-2)\frac{2}{3} = ET = \mu_T$

Megmutatható, hogy $D^2(T) = (16n - 29)/90 = \sigma_T^2$ és hogy T közelítőleg normális eloszlású, hiszen indikátorok összegeként írva a centrális határeloszlás tételre hivatkozhatunk. Tehát $T \sim N(\mu_T, \sigma_T^2)$.

Megjegyzés: Nagy $T - \mu_T$ érték azt jelenti, hogy hevesebben fluktuál a sorozat mint egy $i.i.d.$=> negatív korreláció van a szomszédos tagok közt.
Nagy negatív $T - \mu_T$ érték kis fluktuációra utal => pozitív korreláció van a szomszédos tagok közt.

Differencia-előjel próba:
Az előző felépítésben most számoljuk azon i-ket, amelyre $y_i > y_{i-1}$, $i = 2, ..., n$.

Ez ugyananannyi, mint a differenciált sorozat pozitív tagjainak S száma.
$i.i.d$ sorozatra:

$$\mu_S = ES = \frac{1}{2} \cdot (n-1)$$

$$\sigma_S^2 = D^2 S = \frac{n+1}{12}$$

és nagy n-re $S \sim N(\mu_S, \sigma_S^2)$

Ha $S - \mu_S$ nagy abszolút értékű pozitív vagy negatív szám, akkor valószínűleg növekvő, ill. csökkenő trend van a sorozatban.

Rang próba

Rang próba. Hasznos például, ha lineáris trend ellenében kell tesztelni a nullhipotézist.
Legyen P azon (i,j) párok számra a fentebbi mintarealizációban, amelyre

$$y_j > y_i \text{ és } j > i, i = 1, \cdots, n-1$$

Mivel $\binom{n}{2}$ olyan pár van, amelyben $j > i$, és mindegyikre $\frac{1}{2}$ a valószínűsége, hogy $y_j > y_i$, ezért

$$\mu_p = EP = \frac{1}{4} \cdot n \cdot (n-1)$$

valamint belátható, hogy:

$$\sigma_p^2 = D^2 P = n \cdot (n-1)(2n+5) \cdot \frac{1}{72}$$

és akárcsak az előzőekben, nagy n-re $P \sim N(\mu_p, \sigma_p^2)$. Ezt teszteljük.

Nagy abszolút értékű pozitív vagy negatív $P - \mu_p$ érték növekvő, ill. csökkenő trendre utal.

Nemparametrikus regresszió:

Running line (futó egyenes):

Simítás, általában a trend torzított becslése.

N(0,1)-es *i.i.d* mintát, ha simítjuk, akár periodikus görbét is kaphatunk belőle.

A simítás során 2 alapkérdés van:

1. Hogyan "átlagoljunk" egy bizonyos környezetben
2. Hogyan válasszuk meg a környezetet

- Legközelebbi szomszéd - a legközelebbi k pont
- Szimmetrikus legközelebbi szomszéd - az egyik és másik oldalon is $\frac{k}{2}$, $\frac{k}{2}$ pont. Egyfajta mozgó ablak.

Legyen:

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

ahol μ sima függvény, és legyen mintánk Y-ra 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),...(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.

Mag regresszió (Kernel regression):
Ekkor is környezeteket választunk, de ezen belül nem egyenlő súllyal vesszük figyelembe a pontokat.
Ha x_0-ban vagyunk kíváncsiak a simított predikcióra, akkor a megfigyelési "helyeket" (a regresszor értékeit) súlyozzuk az x_0-tól való távolságuk függvényében

$$w_{0,i} = \frac{C_0}{\lambda} \cdot K \cdot \left(\left| \frac{x_0 - x_i}{\lambda} \right| \right)$$

ahol K egy magfüggvény
λ a sávszélesség
(Egy lehetőség pl. x_i-t a szórásával osztani.)
Ezekkel a súlyokkal egy súlyozott regressziót csinálunk, vagyis a minimalizálandó legkisebb négyzetes kifejezést súlyozva állítjuk elő.

$$\hat{\mu}(x_0) = \frac{\sum K \cdot \left(\frac{x_0 - x_i}{\lambda}\right) \cdot y_i}{\sum K \cdot \left(\frac{x_0 - x_i}{\lambda}\right)}$$

Magfüggvények:

Gauss mag:

a Gauss eloszlás sűrűségfüggvénye

Minimális variancia mag:

$K(t) = \frac{3}{8}(3 - 5t^2) \quad |t| \leq 1$

Epanechnikov mag:

$K(t) = \frac{3}{4}(1 - t^2) \quad |t| \leq 1$

Lokális regresszió: LOESS

A Running line és a Kernel regresszió kombinációja.
Minden környezetben súlyozott legkisebb négyzetes illesztés.
A lokális regresszió célja: modell identifikálás (lehetőleg polinomiális tagokkal)
regresszorok száma Mallow féle C_p statisztika?????

$$W = \left[\frac{x_0 - x_i}{\Delta x_0} \right]$$

Δx_0 az adott környezet legnagyobb távolsága x_0-tól.

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

$$0 \le t \le 1$$

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