

4. A becslélmélet alapjai (A legkisebb négyzetes hibájú becslés ismételése)

Legkisebb négyzetes hibájú (LS) becslők: nincs előzetes ismeretünk sem a mérendő paramétréről, sem a csatorna karakterisztikáról (a zajról). Tegyük fel, hogy a megfigyelési egyenlet lineáris: $z = Ua + n$. Feltételezzük, hogy az a paraméter \hat{a} értéket vesz fel, és felállítjuk a megfigyelés modelljét: $U\hat{a}$. A megfigyelést ezzel összevetve keressük \hat{a} legjobb beállítását négyzetes hibafüggvény feltételezésével:

$$C(a, \hat{a}) = (z - U\hat{a})^T (z - U\hat{a}) = z^T z - z^T U\hat{a} - \hat{a}^T U^T z + \hat{a}^T U^T U\hat{a} = z^T z - 2 \hat{a}^T U^T z + \hat{a}^T U^T U\hat{a} \quad (68)$$

melynek szélsőértékét (minimumát) keressük: $\left. \frac{\partial C(a, \hat{a})}{\partial \hat{a}} \right|_{\hat{a}=\hat{a}_{LS}} = 0$ feltétel vizsgálatával. (68)

deriválásával $-2U^T z + 2U^T U\hat{a} = 0$, amivel:

$$\hat{a}_{LS} = [U^T U]^{-1} U^T z \quad (69)$$

Megjegyzések:

1. A derivált helyességét egyszerűen leellenőrizhetjük, ha a (68) összefüggésben kijelölt mátrix-, ill. vektor-szorzásokat kifejtjük, és azt követően a deriválást komponensenként végezzük el.
2. Általánosított/súlyozott négyzetes kritériumot is használhatunk, ha bevezetünk egy Q négyzetes súlyozó mátrixot: $C(a, \hat{a}) = (z - U\hat{a})^T Q(z - U\hat{a})$, (70)

amivel

$$\hat{a}_{LS} = [U^T Q U]^{-1} U^T Q z. \quad (71)$$

3. Ha $Q = \Sigma_n^{-1}$, akkor a (66) összefüggés szerinti Gauss-Markov becslőt kapjuk, tehát a GM becslő egy súlyozott legkisebb négyzetes hibájú becslő, ahol a súlyozást a megfigyelési zaj kovariancia mátrixának inverze adja.
4. Ha n nulla várható értékű, Gauss eloszlású fehér zaj vektor, akkor azt $n \sim N(0, \sigma^2 I)$ -vel jelöljük. Ezzel összhangban $\hat{a}_{LS} \sim N(a, \sigma^2 [U^T Q U]^{-1})$, mert z Gauss eloszlású vektor, és \hat{a}_{LS} pedig ennek a vektornak lineáris transzformáltja.

Példák:

1. Legyen az illesztendő modell a diszkrét időindex polinomja:

$$x_n = a_0 + a_1 n + a_2 n^2 + \dots + a_p n^p + w_n,$$

ahol w_n az additív megfigyelési zaj n -edik időpillanatbeli értéke. Mátrixos alakban:

$$z = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots & 2^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & N-1 & (N-1)^2 & \dots & (N-1)^p \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix} = Ua + n, \quad \hat{a}_{LS} = [U^T U]^{-1} U^T z$$

2. Legyen az illesztendő modell az időben diszkrét Fourier sorfejtés:

$$x_n = \sum_{k=1}^M a_k \cos \frac{2\pi kn}{N} + \sum_{k=1}^M b_k \sin \frac{2\pi kn}{N} + w_n$$

Megjegyzés: most DC komponenst nem illesztünk. Mátrixos alakban:

$$z = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} \dots & 1 & \dots & \dots & 1 & \dots \\ \dots & \cos \frac{2\pi k}{N} & \dots & \dots & \sin \frac{2\pi k}{N} & \dots \\ \dots & \cos \frac{4\pi k}{N} & \dots & \dots & \sin \frac{4\pi k}{N} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \cos \frac{2\pi}{N}(N-1) & \dots & \dots & \sin \frac{2\pi}{N}(N-1) & \dots \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \\ b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix} = Ua + n$$

$$\hat{a}_{LS} = [U^T U]^{-1} U^T z$$

Mivel most $(U^T U)^{-1} = \frac{2}{N} I$, ezért $\hat{a}_k = \frac{2}{N} \sum_{n=0}^{N-1} x_n \cos \frac{2\pi}{N} kn$, $\hat{b}_k = \frac{2}{N} \sum_{n=0}^{N-1} x_n \sin \frac{2\pi}{N} kn$.

Megjegyzés: A kovariancia mátrix, ha a zaj Gauss eloszlású és fehér:

$$C_{\hat{a}_{LS}} = \sigma_w^2 [U^T U]^{-1} = \frac{2}{N} \sigma_w^2 I$$

3. Kis módosítással a (72) összefüggéssel megadott (lásd még a 17. ábrát) mozgó átlagoló (FIR szűrő) arra az esetre, amikor a gerjesztés belépő függvény, tehát $x_n = 0$, ha $n < 0$:

$$y_n = \sum_{k=0}^{P-1} a_k x_{n-k} + w_k, \text{ amit a } z = Ua + n \text{ alakban felírva:}$$

$$z = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} x_0 & 0 & \dots & \dots & 0 \\ x_1 & x_0 & \dots & \dots & 0 \\ x_2 & x_1 & x_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & 0 \\ x_{N-1} & x_{N-2} & x_{N-3} & \dots & x_{N-P} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{P-1} \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix}, \quad \hat{a}_{LS} = [U^T U]^{-1} U^T z$$

Megjegyzés: A kovariancia mátrix, ha a zaj Gauss eloszlású és fehér: $C_{\hat{a}_{LS}} = \sigma_w^2 [U^T U]^{-1}$. A (73)-(76) összefüggések itt is felírható azzal, hogy $\{x_n\}$ belépő sorozat, amit az átlagolásoknál figyelembe kell venni.

4. Lineáris modell színes Gauss zaj esetén: $z = Ua + n$, $n \sim N(0, C)$, vagyis a zaj kovariancia mátrixa nem lesz diagonális mátrix. Az ún. fehéritési eljárást alkalmazzuk: mivel C pozitív definit, ezért létezik olyan invertálható D mátrix, amellyel: $C^{-1} = D^T D$. Ezzel a D mátrixszal megszorozva a megfigyelési zaj vektorát:

$$E[(Dn)(Dn)^T] = E[Dnn^T D^T] = DCD^T = DD^{-1}(D^T)^{-1}D^T = I,$$

vagyis a zaj kiféhéredik, és egységnyi varianciájú lesz. Ha megfigyelési egyenlet egészét transzformáljuk a D mátrixszal, akkor:

$$z' = Dz = DUa + Dz = U'a + n'. \text{ Itt } n' = Dn \sim N(0, I).$$

Ezzel a problémát visszavezettük a fehér zaj esetére:

$$\hat{a}_{LS} = [U'^T U']^{-1} U'^T z' = [U^T D^T D U]^{-1} U^T D^T D z = [U^T C^{-1} U]^{-1} U^T C^{-1} z, \quad C_{\hat{a}_{LS}} = [U^T C^{-1} U]^{-1}.$$

5. Lineáris modell ismert komponens esetén: az s jelkomponens ismert: $z = Ua + s + n$. $z' = z - s$ bevezetésével: $\hat{a}_{LS} = [U^T U]^{-1} U^T (z - s)$, fehér zaj esetén: $C_{\hat{a}_{LS}} = \sigma^2 [U^T U]^{-1}$.

Példa: $x_n = A + r^n + w_n$. Itt A egy ismeretlen konstans, r pedig egy ismert konstans:

$$z = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} A + \begin{bmatrix} 1 \\ r \\ \vdots \\ r^{N-1} \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix}, \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} (x_n - r^n), \text{var}\{\hat{A}\} = \frac{\sigma_w^2}{N},$$

ha a zaj Gauss, és fehér.

5. Modellillesztés

A legkisebb négyzetes hibájú becslők esetén nincs előzetes ismeretünk, valójában modellt illesztettünk. A becslés varianciáját akkor adtuk meg, amikor az additív megfigyelési zajról tudtuk, hogy Gauss eloszlású és fehér. Láttuk, hogy színes zaj esetén milyen módszerrel vezethetjük vissza az illesztés problémáját az alapesetre.

A modellillesztés problémája meglehetősen szerteágazó. Egyik klasszikus válfaja a regresszió számítás.

Regresszió-számítás: függő és független változók közötti közvetlen, determinisztikus kapcsolat meghatározása, a modellillesztés egy speciális esete. A 18. ábrán látható elrendezésben a modellezendő $y = g(u, n)$ függvény kétfajta független változóval rendelkezik: az egyiket $u(n)$ jelöli, amelyet ismerünk és „kézben tudunk tartani”, a másik, amelyiket $n(n)$ jelöli, amely ismeretlen, kézben nem tartható, tipikusan zajfolyamatnak elképzelt/modellezett folyamat.

Megjegyzések:

1. A továbbiakban az argumentumként szereplő kis „ n ” nagyon gyakran az iterációs lépést azonosítja vagy diszkrét időindex, amely ekvivalens módon tényleges indexként is megjelenik időnként. Ennek megfelelően $u(n) = u_n$, ill. $y(n) = y_n$ egyenértékűek.
2. A kis „ n ” kettős használata senkit se zavarjon, a különbség egyértelmű: argumentumként, ill. indexként diszkrét („idő”) index, önállóan pedig zajfolyamatként interpretáljuk.

A modellezéshez egy általunk kézben tartott, tipikusan paraméterek segítségével módosítható („hangolható”) $\hat{y} = \hat{g}(u)$ függvényt használunk. A cél egy olyan „beállítás” elérése, amely valamilyen értelemben optimális. Tipikusan négyzetes kritériumot használunk:

$$\varepsilon = E\{(y - \hat{y})^T (y - \hat{y})\} \quad (78)$$

Regresszió-számítás teljesen specifikált statisztikai jellemzőkkel: ha ismerjük u és y $f(u, y)$ együttes valószínűség sűrűség függvényét, akkor a feladat Bayes becslési probléma, melynek megoldása az a posteriori várható érték:

$$\hat{g}(u) = E\{y|u\} \quad (79)$$

Az $[u, \hat{g}(u)]$ görbe az y változó u -ra vonatkoztatott regressziós görbéje. Ha a bemenet vektor, akkor regressziós felület.

Regresszió-számítás részben specifikált statisztikai jellemzőkkel: nem ismerjük az együttes eloszlást, csak véges számú momentumát.

Lineáris regresszió: Az illesztendő függvény a $\hat{g}(u) = a_0 + a_1 u$ skalár lineáris függvény, melynek paraméterei úgy választandók meg, hogy $E\{(y - \hat{g}(u))^2\}$ minimális legyen. Legyen

ismert $\mu_u, \mu_y, \sigma_u, \sigma_y, \rho$, ahol az utóbbi a normalizált kereszt-kovariancia függvény:

$$\rho = \frac{E\{(u - \mu_u)(y - \mu_y)\}}{\sigma_u \sigma_y}. \text{ Minimalizálandó az}$$

$$\varepsilon = E\{(y - a_0 - a_1 u)^2\} = E\{y^2\} + a_0^2 + a_1^2 E\{u^2\} - 2a_0 E\{y\} - 2a_1 E\{uy\} - 2a_0 a_1 E\{u\} \quad (80)$$

összefüggés a_0 és a_1 szerint:

$$\frac{\partial \varepsilon}{\partial a_0} = 2a_0 - 2\mu_y + 2a_1 \mu_u = 0, \text{ ahonnan } a_0 = \mu_y - a_1 \mu_u, \text{ amit} \quad (81)$$

$$\frac{\partial \varepsilon}{\partial a_1} = 2a_1(\sigma_u^2 + \mu_u^2) - 2(\rho \sigma_u \sigma_y + \mu_u \mu_y) + 2a_0 \mu_u = 0 \text{ kifejezésbe behelyettesítve}$$

$$\boxed{\hat{a}_0 = \mu_y - \mu_u \rho \frac{\sigma_y}{\sigma_u}, \hat{a}_1 = \rho \frac{\sigma_y}{\sigma_u}} \quad (82)$$

Megjegyzések:

1. A (82) összefüggés származtatásakor felhasználtuk, hogy $E\{(u - \mu_u)^2\} = \sigma_u^2 = E\{u^2\} - \mu_u^2$, valamint $E\{(u - \mu_u)(y - \mu_y)\} = E\{uy\} - \mu_u \mu_y$.
2. A (80) összefüggésbe behelyettesítve $\text{var}\{y - \hat{g}(u)\} = \sigma_y^2(1 - \rho)$, ez az érték a hiba varianciája. Érdeemes megvizsgálni a viszonyokat a $0 \leq \rho \leq 1$ paraméter függvényében. Ha a kereszt-kovariancia nulla, akkor egy fekvő egyenest kapunk, a kimenet legjobb becslője a bemenettől függetlenül a mért értékek várható értéke. Ha a kereszt-kovariancia 1 (100%), akkor y csak u -tól függ, azaz a zajtól (n -től) nem.
3. A lineáris regresszió feladatának egyfajta általánosítása az ún. polinomiális regresszió:

$$\hat{g}(u) = \sum_{k=0}^N a_k u^k, \quad (83)$$

amelynek fontos tulajdonsága, hogy paramétereiben lineáris. A paramétereiben lineáris modelleket azért kedveljük, mert négyzetes hibakritérium esetén a szélsőérték-keresés lineáris egyenletrendszer megoldására vezet, mivel a négyzetes kifejezések paraméterek szerinti deriválása lineáris összefüggést eredményez.

Lineáris regresszió mérési adatok alapján: a fentieket végigvihetjük akkor is, ha nincsen előzetes információnk. Ilyenkor $y_n = a_0 + a_1 u_n + w_n$, $z = Ua + n$, mint eddig.

$$z = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & u_0 \\ 1 & u_1 \\ \vdots & \vdots \\ 1 & u_{N-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix}, [U^T U] = \begin{bmatrix} N & \sum_{n=0}^{N-1} u_n \\ \sum_{n=0}^{N-1} u_n & \sum_{n=0}^{N-1} u_n^2 \end{bmatrix}, U^T z = \begin{bmatrix} \sum_{n=0}^{N-1} y_n \\ \sum_{n=0}^{N-1} u_n y_n \end{bmatrix}.$$

$$\begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \end{bmatrix} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} u_n^2 - \left(\frac{1}{N} \sum_{n=0}^{N-1} u_n\right)^2} \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} u_n^2 & -\frac{1}{N} \sum_{n=0}^{N-1} u_n \\ -\frac{1}{N} \sum_{n=0}^{N-1} u_n & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} y_n \\ \frac{1}{N} \sum_{n=0}^{N-1} u_n y_n \end{bmatrix}.$$

Megjegyzés: Ebben az eredményben a (82) kifejezésben szereplő statisztikai jellemzők becslőinek összetevőit azonosíthatjuk, és átalakításokkal - tipikusan az átlagtól való eltérések felírásával - ezeket a kifejezéseket egymásnak teljesen megfeleltethetjük. Tegyük meg!

A regressziós séma általánosítása: A 19. ábrán a modell-illesztést a regressziós sémát követő módon mutatjuk be. Az u bemenetre adott y választ szeretnénk valamilyen kritérium szerint (az ábrán négyzetes értelemben) legjobban megközelíteni a modell \hat{y} válaszával. Érdekes összevetni ezt a sémát a megfigyelő sémával (lásd 2. ábra). Ehhez rajzoljuk át a 20. ábra szerinti formára. A nagyfokú hasonlóság egyértelmű: mindkét esetben modellillesztést végzünk. A megfigyelő séma esetén a paramétereket ismerjük, és az állapotokat becsüljük, míg a regressziós sémában a modellünk „állapotát” kézben tartjuk, és a paramétereket keressük. Mindkét séma „párhuzamos” abban az értelemben, hogy a „bemenő jelet” illetően párhuzamosan kapcsolódnak. A modell-illesztési probléma megragadható „soros” formában is, amikor tulajdonképpen az ún. inverz-modellt illesztjük (lásd 21. ábra), amikor a bemenetet az inverz-modell által becsüljük. Ennek a megközelítésnek hátránya dinamikus rendszerek esetében a soros „kapcsolás” eredő késleltetése, ezért az inverz-moddellel „jóslásra” kényszerülünk, ami sok nehézséggel jár.

Adaptív lineáris kombinátor: Az általánosított regressziós séma kapcsán az egyik gyakran használt modell-családot a 22. ábra mutatja be. Ebben az $u(n)$ diszkrét értéksorozatból egy $X^T(n) = [x_0(n) \ x_1(n) \ \dots \ x_{M-1}(n)]$ értéksorozatot állítunk elő először, majd ezen értékek lineáris kombinációjaként állítjuk elő az $\hat{y}(n)$ értéket. Az optimumkeresés során a $W^T(n) = [w_0(n) \ w_1(n) \ \dots \ w_{M-1}(n)]$ paraméterek legkedvezőbb, minimális négyzetes hibát eredményező beállítására törekszünk. Minimalizáljuk az

$$\begin{aligned} \varepsilon(n) &= E\{[y(n) - X^T(n)W(n)]^T [y(n) - X^T(n)W(n)]\} = \\ &= E\{y^T(n)y(n)\} - 2W^T(n)E\{X(n)y(n)\} + W^T(n)E\{X(n)X^T(n)\}W(n). \end{aligned} \quad (84)$$

Vezessük be a $E\{X(n)y(n)\} = P$, és a $E\{X(n)X^T(n)\} = R$ jelölést! Ezzel a szélsőérték keresés $\frac{\partial \varepsilon(n)}{\partial W(n)} = -2P + 2RW(n) = 0$, amiből az optimális beállítás: $W^* = R^{-1}P$ (85)

A (85) összefüggés az ún. Wiener-Hopf egyenlet.

Megjegyzések:

1. A (85) összefüggést visszahelyettesítve (84)-be:

$$\varepsilon_{\min} = E\{y^T(n)y(n)\} - P^T R P = E\{y^T(n)y(n)\} - P^T W^* \quad (86)$$

$$\varepsilon(n) = \varepsilon_{\min} + [W(n) - W^*]^T R [W(n) - W^*] = \varepsilon_{\min} + V^T(n) R V(n), \quad (87)$$

ahol $V(n) = [W(n) - W^*]$ az ún. paraméterhiba vektora.

2. A (87) összefüggés egyértelműen mutatja a négyzetes hiba alakulását a paraméterek, ill. a paraméterhiba függvényében. A viszonyok illusztrálására a 23. ábra szolgál. A hibafelület tetszőleges pontjában a hiba csökkenés fajlagos mértékét a felület meredekségével (gradiensével) mérhetjük:

$$\nabla(n) = \frac{\partial \varepsilon(n)}{\partial W(n)} = 2R[W(n) - W^*] = 2RV(n) = 2(RW(n) - P). \quad (88)$$

A (88) összefüggés kitüntetett szerepet kap az adaptív eljárásoknál, ahol a hibafelületen a gradiens mentén „ereszkedünk”.

Példa: Legyen $X^T(n) = [\sin(2\pi n/N) \ \sin(2\pi(n-1)/N)]$, azaz egy szinuszos hullámforma két egymás utáni mintája. $y(n) = 2\cos(2\pi n/N)$. Hogyan válasszuk meg a

$$W^T(n) = [w_0(n) \ w_1(n)] \quad (89)$$

paramétereket ahhoz, hogy a közelítés négyzetes hibája minimális legyen? A R és a P mátrixok a szinuszos, ill. koszinuszos hullámformák teljes ($N > 2$) periódusra történő átlagolásával származtathatók:

$$R = \begin{bmatrix} 0.5 & 0.5 \cos \frac{2\pi}{N} \\ 0.5 \cos \frac{2\pi}{N} & 0.5 \end{bmatrix}, P = \begin{bmatrix} 0 \\ -\sin \frac{2\pi}{N} \end{bmatrix}. \quad (90)$$

$$E\{\sin^2(2\pi/N)\} = E\{\sin^2(2\pi(n-1)/N)\} = 0.5, \quad E\{\sin(2\pi n/N) \sin(2\pi(n-1)/N)\} = \frac{\cos(\frac{2\pi}{N})}{2},$$

$$E\{2 \sin(2\pi/N) \cos(2\pi/N)\} = 0, \quad E\{2 \sin(2\pi(n-1)/N) \cos(2\pi/N)\} = -\sin \frac{2\pi}{N}.$$

$$R^{-1} = \frac{1}{0.25 \sin^2 \frac{2\pi}{N}} \begin{bmatrix} 0.5 & -0.5 \cos \frac{2\pi}{N} \\ -0.5 \cos \frac{2\pi}{N} & 0.5 \end{bmatrix}, \quad W^* = R^{-1}P = \begin{bmatrix} \frac{2}{\tan(2\pi/N)} \\ \frac{2}{\sin(2\pi/N)} \end{bmatrix} \quad (91)$$

Megjegyzések:

1. Mivel szinuszos minták lineáris kombinációjával hiba nélkül elő lehet állítani koszinuszos hullámformák mintáit, ezért a példa szerinti esetben $\varepsilon_{\min} = 0$, azaz a 23. ábrán a paraboloid legalsó pontja érinti a paraméterek síkját.
2. A példában a (72) összefüggéssel adott mozgó átlagot számítjuk (lásd 24. ábra) ismert hullámforma mintáiból, $P=1$ esetet (azaz az ablak szélesség kettő, az utóbbi, az adaptív lineáris kombinátornál használt jelöléssel az $M=2$ esetet) feltételezve. A teljes periódusra történő átlagolás megfelel annak, hogy mozgó átlag paramétereinek becslésekor $N \rightarrow \infty$.
3. $X^T(n)W^* = 2 \frac{\sin(2\pi n/N)}{\tan(2\pi/N)} - 2 \frac{\sin(2\pi(n-1)/N)}{\sin(2\pi/N)} = 2 \cos(2\pi n/N)$.

