4. A becsléselmélet alapjai (folyt.)

Minimális átlagos négyzetes hibájú becslés:

$$\boxed{\hat{a}_{MS} = \int_{-\infty}^{\infty} a f(a|z) da}, \tag{46}$$

azaz ezzel a kritériumfüggvénnyel a legjobb becslés az a posteriori várható érték.

Minimális átlagos abszolút hibájú becslés (skalár esetre bemutatva)**:**

$$\boxed{\hat{a}_{ABS} = f(a|z) \text{ mediánja.}} \tag{50}$$

Maximum a posteriori (MAP) becslés:

$$\boxed{\hat{a}_{MAP} = f(a|z) \text{ maximumhelye.}} \tag{52}$$

Megjegyzések:

1. A Bayes becslések mindig az a posteriori sűrűségfüggvények alapján történnek.
2. Az MS becslés lineáris abban az alábbi értelemben:

Ha $b = Aa + c$, akkor $\hat{b}_{MS} = A\hat{a}_{MS} + c$, továbbá $E\{a+b|z\} = E\{a|z\} + E\{b|z\} = \hat{a}_{MS} + \hat{b}_{MS}$.

Bayes becslő Gauss eloszlások esetén: Tegyük fel, hogy a keresett a paraméter és a megfigyelési zaj Gauss eloszlásúak. Adott $E\{a\} = \mu_a$, $\text{cov}\{a,a\} = \Sigma_{aa}$, $E\{n\} = 0$, $\text{cov}\{n,n\} = \Sigma_{nn}$. Ha "minden" Gauss eloszlású, akkor az a posteriori sűrűségfüggvény momentumai explicit formában megadhatók. Tegyük fel, hogy az additív zajjal terhelt megfigyelés az alábbi összefüggéssel írható le:

$$z = Ua + n, \tag{53}$$

ahol dim a=p, dim z=q, dim U=$q*p$. Az U az ún. megfigyelési mátrix. Az explicit forma az a posteriori várható értékre:

$$\hat{a}_{MS} = \mu_{a|z} = \underbrace{\mu_a}_{apriori\_ismeret} + \underbrace{[U^T \Sigma_{nn}^{-1} U + \Sigma_{aa}^{-1}]^{-1} U^T \Sigma_{nn}^{-1} (z - U\mu_a)}_{korrekció\_z-U\mu_a\_függvényében} \tag{54}$$

Megjegyzés: $\hat{a}_{MS} = \hat{a}_{ABS} = \hat{a}_{MAP}$, mert az a posteriori sűrűségfüggvény Gauss.

1. Példa: Mérendő egy ellenállás értékét úgy, hogy ismert áram által ejtett feszültséget mérünk. N megfigyelést végzünk. $U = IR + n$. A megfigyelt értékek: $z_k = a + n_k$, k=1, 2, ..., N, a az ismeretlen paraméter (ellenállás), n_k az additív zaj mintája. Tegyük fel, hogy az ellenállás (a gyártási sorozat egy eleme), és a megfigyelési zaj egyaránt Gauss eloszlású valószínűségi változóknak tekinthetők. A zaj megfigyelési értékei korrelálatlanok. Tegyük fel, hogy ismert μ_a és σ_a^2, a zaj várható értéke $\mu_n = 0$, $\text{cov}\{n_k, n_j\} = \sigma_n^2 \delta_{kj}$, ahol $\delta_{kj} = \begin{cases} 1\, ha\, k = j \\ 0\, ha\, k \neq j \end{cases}$.

Vektoros alakban: $z = Ua + n$, $z^T = [z_1, z_2, ..., z_N]$, $n^T = [n_1, n_2, ..., n_N]$, $U^T = [11...1]$, $\Sigma_{nn} = \sigma_n^2 I$.

$$\boxed{\hat{a}_{MS} = \mu_{a|z} = \mu_a + [\frac{N}{\sigma_n^2} + \frac{1}{\sigma_a^2}]^{-1} (\frac{1}{\sigma_n^2} \sum_{k=1}^{N} z_k - \frac{N}{\sigma_n^2} \mu_a) = \mu_a + \frac{N \frac{\sigma_a^2}{\sigma_n^2}}{1 + N \frac{\sigma_a^2}{\sigma_n^2}} (\frac{1}{N} \sum_{k=1}^{N} z_k - \mu_a) = \frac{\mu_a + \frac{\sigma_a^2}{\sigma_n^2} \sum_{k=1}^{N} z_k}{1 + N \frac{\sigma_a^2}{\sigma_n^2}}} \tag{55}$$

Megjegyzések:

1. Az (55) kifejezés alapján a becslés úgy interpretálható, mint egy predikciós-korrekciós formula, amelynek első tagja az a priori ismeret alapján egy jóslás: mekkora lehet a paraméter az előzetes ismeretek alapján, amit a többletinformáció (mérési eredmények) birtokában egy korrekciós tag egészít ki. Ez utóbbi a mért értékek átlaga és a paraméter várható értéke különbségével arányos. Az arányossági tényező $N\dfrac{\sigma_a^2}{\sigma_n^2}$ értékének függvényében nulla és egy közötti érték. Ha $\sigma_a << \sigma_n$, akkor $\hat{a}_{MS} \cong \mu_a$, ha $\sigma_a >> \sigma_n$ vagy $N \to \infty$, akkor $\hat{a}_{MS} \cong \dfrac{1}{N}\sum_{k=1}^{N} z_k$.

2. **A becslési hiba varianciája:** az a posteriori kovariancia:

$$\mathrm{cov}\{a, a|z\} = \Sigma_{aa|z} = [U^T \Sigma_{nn}^{-1} U + \Sigma_{aa}^{-1}]^{-1} = \mathrm{var}\{\tilde{a}\} = \frac{\sigma_a^2}{1 + N\dfrac{\sigma_a^2}{\sigma_n^2}}, \tag{56}$$

ahol $\tilde{a} = \hat{a} - a$. Az $N\dfrac{\sigma_a^2}{\sigma_n^2}$ értékének függvényében σ_a^2 és $\dfrac{1}{N}\sigma_n^2$ közötti érték. Ha $\sigma_a << \sigma_n$, akkor $\mathrm{var}\{\tilde{a}\} \cong \sigma_a^2$, ha $\sigma_a >> \sigma_n$ vagy $N \to \infty$, akkor $\mathrm{var}\{\tilde{a}\} \cong \dfrac{1}{N}\sigma_n^2$.

3. **A becslés feltételesen torzított:** *(nem hangzott el, de az anyag része!)*

$$b(a) = E\{\hat{a}_{MS}|a\} - a = \frac{\mu_a + \dfrac{\sigma_a^2}{\sigma_n^2}\sum_{k=1}^{N} z_k}{1 + N\dfrac{\sigma_a^2}{\sigma_n^2}} - a = \frac{\mu_a - a}{1 + N\dfrac{\sigma_a^2}{\sigma_n^2}}. \tag{57}$$

2. Példa: Mérendő egy ismert jel ismeretlen amplitúdója. $z_k = as_k + n_k$, k = 1, 2, … , N. Az ismeretlen a amplitúdó és n_k Gauss eloszlású, ismert várható értékkel és varianciával. $E\{a\} = \mu_a$, $\mathrm{var}\{a\} = \sigma_a^2$, $E\{n_k\} = 0$, $\mathrm{cov}\{n_i, n_j\} = \sigma_n^2 \delta_{ij}$, $\mathrm{cov}\{a, n_i\} = 0$ $\forall i$ -re, $\forall j$ -re.

Most használjuk a maximum a posteriori (MAP) becslés technikáját!

$$\left.\frac{\partial f(a|z)}{\partial a}\right|_{a=\hat{a}_{MAP}} = 0, \text{ illetve (37) felhasználásával } \left.\frac{\partial \ln f(z|a)}{\partial a} + \frac{\partial \ln f(a)}{\partial a}\right|_{a=\hat{a}_{MAP}} = 0. \tag{58}$$

Most $f(a) = \dfrac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{(a-\mu_a)^2}{2\sigma_a^2}}$, $f(z|a) = \dfrac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma_n^2}\sum_{k=1}^{N}(z_k - as_k)^2}$, amit az (58) összefüggés felhasználásával

$$\frac{\partial \ln f(a)}{\partial a} = -\frac{a - \mu_a}{\sigma_a^2}, \quad \frac{\partial \ln f(z|a)}{\partial a} = \frac{1}{\sigma_n^2}\sum_{k=1}^{N} s_k (z_k - as_k). \tag{59}$$

A két egyenletet egymással összeadva, és behelyettesítve az (58) összefüggésbe:

$$\left.\frac{1}{\sigma_n^2}\sum_{k=1}^{N} s_k(z_k - as_k) - \frac{a-\mu_a}{\sigma_a^2}\right|_{a=\hat{a}_{MAP}} = 0 \text{, amiből } \boxed{\hat{a}_{MAP} = \frac{\mu_a + \frac{\sigma_a^2}{\sigma_n^2}\sum_{k=1}^{N} s_k z_k}{1+\frac{\sigma_a^2}{\sigma_n^2}\sum_{k=1}^{N} s_k^2}} \qquad (60)$$

Megjegyzések:

1. A MAP becslés alkalmazásával megspóroltuk az (54) mátrix-összefüggés használatát a becslő értékének meghatározásakor.

2. **A becslési hiba varianciája:**

$$\text{var}\{\tilde{a}\} = \frac{\sigma_a^2}{1+\frac{\sigma_a^2}{\sigma_n^2}\sum_{k=1}^{N} s_k^2}, \qquad (61)$$

3. Ha s_k=1, $\forall k$, akkor megkapjuk az előző példa eredményét.

4. Itt is azonosítható a döntéselméleti rész 2. feladatában említett „illesztett” szűrő.

5. Természetesen $\hat{a}_{MAP} = \hat{a}_{MS}$.

Maximum likelihood (ML) becslő: Nem ismerjük a mérendő mennyiség a priori valószínűség sűrűségfüggvényét. Ilyenkor azt feltételezzük, hogy ez a függvény „szélesen elterülő”, és ebből adódóan az a posteriori sűrűségfüggvény megegyezik a csatorna-karakterisztikával. Az optimális becslőt a csatorna-karakterisztika maximumához rendeljük:

$$\left.\frac{\partial f(z|a)}{\partial a}\right|_{a=\hat{a}_{ML}} = 0 \text{, ill. } \left.\frac{\partial \ln f(z|a)}{\partial a}\right|_{a=\hat{a}_{ML}} = 0 \qquad (62)$$

Megjegyzés: A 16. ábrán látható viszonyokat a (37) összefüggés „szemszögéből” célszerű elemezni: az a posteriori sűrűségfüggvényt megadó összefüggés számlálója az ábrán látható két függvény szorzata. A szorzat maximumhelye láthatóan a (62) összefüggéssel megadható.

Gauss-Markov (GM) becslő: A maximum likelihood becslő speciális esete, amikor a megfigyelési zaj Gauss eloszlású, a megfigyelési egyenlet pedig lineáris. N dimenziós megfigyeléseket végzünk, n az N dimenziós zaj-vektor:

$$E\{n\} = 0,\ \text{cov}\{n,n\} = \Sigma_{nn},\ f(n) = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma_{nn}|^{\frac{1}{2}}} e^{-\frac{1}{2}n^T\Sigma_{nn}^{-1}n}, \qquad (63)$$

ahol $|\Sigma_{nn}|$ a Σ_{nn} mátrix determinánsát jelöli.

A megfigyelési egyenlet: $z = Ua + n$, amellyel a csatorna karakterisztika

$$f(z|a) = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma_{nn}|^{\frac{1}{2}}} e^{-\frac{1}{2}(z-Ua)^T\Sigma_{nn}^{-1}(z-Ua)}, \qquad (64)$$

melynek szélsőérték helye a

$$\frac{\partial \ln f(z|a)}{\partial a} = \left.\frac{\partial}{\partial a}[(z-Ua)^T\Sigma_{nn}^{-1}(z-Ua)]\right|_{a=\hat{a}_{GM}} = 0 \qquad (65)$$

összefüggés alapján határozható meg. (65) alapján:

$U^T\Sigma_{nn}^{-1}Ua - U^T\Sigma_{nn}^{-1}z = 0$, amiből

$$\boxed{\hat{a}_{GM} = [U^T\Sigma_{nn}^{-1}U]^{-1}U^T\Sigma_{nn}^{-1}z}, \qquad (66)$$

ha $[U^T \Sigma_{nn}^{-1} U]^{-1}$ létezik.

Megjegyzések: (1) A (66) összefüggés $\Sigma_{aa}^{-1} = 0$ (a „szórás végtelen") helyettesítés mellett a Bayes esetből származtatható. (2) A Gauss-Markov becslő torzítatlan.

Példa: A $z_k = a + n_k$, k=1,2,…,N, független megfigyelések. $E\{n_k\} = 0$; $\operatorname{cov}\{n_k, n_j\} = \sigma_n^2 \delta_{kj}$. A csatorna karakterisztika: $f(z|a) = \dfrac{1}{(2\pi\sigma_n^2)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma_n^2}\sum_{k=1}^{N}(z_k - a)^2}$, amelynek maximumhelyénél kapjuk az a paraméter Gauss-Markov becslője: $\left.\dfrac{\partial \ln f(z|a)}{\partial a}\right|_{\hat{a}=\hat{a}_{ML}=\hat{a}_{GM}} = \dfrac{N}{\sigma_n^2}[\dfrac{1}{N}\sum_{k=1}^{N} z_k - a] = 0$;

$$\boxed{\hat{a}_{ML} = \hat{a}_{GM} = \frac{1}{N}\sum_{k=1}^{N} z_k}, \tag{67}$$

azaz lineáris megfigyelési egyenlet, és Gauss eloszlású csatorna zaj esetén a legjobb (Gauss-Markov) becslő az egyszerű átlagolás.

Legkisebb négyzetes hibájú (LS) becslők: nincs előzetes ismeretünk sem a mérendő paraméterről, 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:

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

melynek szélsőértékét (minimumát) keressük: $\left.\dfrac{\partial C(a,\hat{a})}{\partial \hat{a}}\right|_{\hat{a}=\hat{a}_{LS}} = 0$ feltétel vizsgálatával. (68) deriválásával $-2U^T z + 2U^T U\hat{a} = 0$, amivel:

$$\boxed{\hat{a}_{LS} = [U^T U]^{-1} U^T z} \tag{69}$$

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
   $$\boxed{\hat{a}_{LS} = [U^T Q U]^{-1} U^T Q z}. \tag{71}$$
3. Ha $Q = \Sigma_{nn}^{-1}$, 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.

Példa: Mozgó-átlag (Moving Average: MA) paraméterek becslése: Az $\{y_n\}$, n=1,2,…,N, diszkrét értéksorozat („kimenőjel") elemei egy csúszó ablakon keresztül látható $\{x_n\}$ diszkrét értékek („bemenőjel") lineáris kombinációjaként állnak elő:

$$y_n = \sum_{k=0}^{P} a_k x_{n-k} \ . \qquad (72)$$

Keresettek az $\{a_k\}$ súlyozó együtthatók. A megfigyelés lineáris: $z = Ua + n$; $z_n = y_n + n_n$, n=1,2,…,N, $z^T = [z_1, z_2, ..., z_N]$, $a^T = [a_0, a_1, ..., a_P]$. A legkisebb négyzetes értelemben optimális paramétereket a (69) összefüggés adja.

Érdemes megvizsgálni, hogy mi az információtartalma az $[U^T U]^{-1}$ mátrixnak, illetve az $U^T z$ vektornak. Ehhez írjuk fel az $U^T U$ mátrixot diadikus szorzatok összegként:

$$\begin{bmatrix} x_1 & x_2 & \dots & x_N \\ x_0 & x_1 & \dots & x_{N-1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1-P} & x_{2-P} & \dots & x_{N-P} \end{bmatrix} \begin{bmatrix} x_1 & x_0 & \dots & x_{1-P} \\ x_2 & x_1 & \dots & x_{2-P} \\ \vdots & \vdots & \vdots & \vdots \\ x_N & X_{N-1} & \dots & x_{N-P} \end{bmatrix} = \sum_{n=1}^{N} x_n x_n^T = \sum_{n=1}^{N} \begin{bmatrix} x_n^2 & x_n x_{n-1} & \dots & x_n x_{n-P} \\ x_{n-1} x_n & x_{n-1}^2 & \dots & x_{n-1} x_{n-P} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-P} x_n & x_{n-P} x_{n-1} & \dots & x_{n-P}^2 \end{bmatrix} \qquad (73)$$

Figyeljük meg, hogy (73) egy olyan mátrix, amelynek elemei (alkalmas normálással kiegészítve) az $\{x_n\}$ diszkrét értéksorozat autokorreláció értékeit becslik:

$$\hat{R}_{xx}(k-p) = \frac{1}{N} \sum_{n=1}^{N} x_{n-k} x_{n-p} \ , \ k,p = 0,1,\dots,P. \qquad (74)$$

Hasonlóképpen felírva az $U^T z$ vektort:

$$U^T z = \sum_{k=1}^{N} \begin{bmatrix} x_n z_n \\ x_{n-1} z_n \\ \vdots \\ x_{n-P} z_n \end{bmatrix} \qquad (75)$$

(75) egy olyan vektor, amelynek elemei (alkalmas normálással kiegészítve) az $\{x_n\}$ és $\{z_n\}$ diszkrét értéksorozatok keresztkorreláció értékeit becslik:

$$\hat{R}_{xz}(k) = \frac{1}{N} \sum_{n=1}^{N} x_{n-k} z_n, \ k = 0,1, \dots, P. \qquad (76)$$

Jelölje $\hat{R}_{xx}$ a (74) összefüggés szerinti elemekből álló mátrixot, $\hat{R}_{xz}$ pedig a (76) összefüggés szerinti elemekből álló vektort:

$$\boxed{\hat{a}_{LS} = \hat{R}_{xx}^{-1} \hat{R}_{xz}} \qquad (77)$$

Megjegyzés:

1. A (77) összefüggés természetesen mindössze egy formális átírás a példa szerinti viszonyok esetére, de a későbbiekben látni fogjuk ennek létjogosultságát.
2. Ezen a ponton formálisan befejezzük a becsléselmélet alapjainak tárgyalását, de a folytatás, mint látni fogjuk, rendre „mérésekről”, azaz jelenségek, ill. azok paramétereinek és állapotainak megragadásáról szól, aminek konkrét eszközei tipikusan becslési eljárások lesznek.

5. Modellillesztés

Regresszió-számítás: függő és független változók közötti közvetlen, determinisztikus kapcsolat meghatározása, a modellillesztés egy speciális esete. A 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:

$$\varepsilon = E\left\{(y - \hat{y})^T (y - \hat{y})\right\} \tag{78}$$