

3. A döntéelmélet alapjai (folyt.)

Megjegyzések a kéthipotézises Bayes döntéshez:

1. Ha a költségeket úgy választjuk meg, hogy (17) $\eta = \frac{P_0}{P_1}$ legyen, akkor a döntési szabály

$$P_1 f(z|H_1) \underset{H_0}{>} P_0 f(z|H_0) \text{ alakban írható, ami megegyezik a } P(H_1|z) \underset{H_0}{>} P(H_0|z) \text{ kifejezéssel,}$$

vagyis a döntést az *a posteriori* valószínűségek alapján hozhatjuk meg. Ezt a speciális esetet maximum a posteriori (MAP) döntésnek nevezzük.

2. Szokás (19) helyett a $\lambda(z) = \ln \Lambda(z) \underset{H_0}{>} \ln \eta = \gamma$, ún. log-likelihood arányt használni.

1. Példa: Konstans jel detektálása zajos csatornán keresztül: jelen van-e a jel a megfigyelésekben vagy nincsen? Tegyük fel, hogy a megfigyelések független, Gauss eloszlású valószínűségi változók nulla várható értékkel, és σ_n^2 varianciával. A H_0 hipotézis az, hogy a megfigyelés $z_k = n_k$, azaz a jel nincsen jelen, csak a zaj aktuális mintáját kapjuk. A H_1 hipotézis az, hogy a megfigyelés $z_k = a + n_k$, azaz a jel jelen van, a jel és a zaj aktuális mintájának összegét figyeltük meg. $k = 1, 2, \dots, N$, azaz összesen N mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

$$f(z|H_0) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{z^2}{2\sigma_n^2}}, \text{ ill. } f(z|H_1) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(z-a)^2}{2\sigma_n^2}} \quad (20)$$

A megfigyelések függetlensége miatt a N megfigyelés együttes sűrűségfüggvénye az egyedi megfigyelések sűrűségfüggvényeinek szorzata. (19) aktuális alakja:

$$\Lambda(z) = \prod_{k=1}^N \frac{f(z_k|H_1)}{f(z_k|H_0)} = \prod_{k=1}^N \frac{e^{-\frac{(z_k-a)^2}{2\sigma_n^2}}}{e^{-\frac{z_k^2}{2\sigma_n^2}}} \underset{H_0}{>} \eta, \quad (21)$$

illetve a log-likelihood arány:

$$\ln \Lambda(z) = \lambda(z) = -\sum_{k=1}^N \frac{(z_k - a)^2}{2\sigma_n^2} + \sum_{k=1}^N \frac{z_k^2}{2\sigma_n^2} = \frac{a}{\sigma_n^2} \sum_{k=1}^N z_k - \frac{Na^2}{2\sigma_n^2} \underset{H_0}{>} \ln \eta = \gamma \quad (22)$$

Átrendezve:

$$\frac{1}{N} \sum_{k=1}^N z_k \underset{H_0}{>} \frac{\sigma_n^2}{Na} \ln \eta + \frac{a}{2} \quad (23)$$

A teszthez tehát a megfigyelések átlagát kell képeznünk, és összevetnünk egy küszöbértékkel. A „döntőkészülék” blokkvázlatát a 8. ábra mutatja be.

Megjegyzések:

1. Ha $\eta = 1$, akkor $\ln \eta = 0$, tehát $\frac{1}{N} \sum_{k=1}^N z_k \underset{H_0}{>} \frac{a}{2}$, azaz a döntési küszöb a konstans jelszint fele.

Ez például fennáll, ha $P_0 = P_1 = 0.5$, $C_{00} = C_{11}$ és $C_{10} = C_{01}$.

2. Ha $\eta < 1$, akkor $\ln \eta < 0$, amivel a döntési küszöb a konstans jelszint fele alá csökken. Ez például fennáll, ha $P_0 < P_1$, $C_{00} = C_{11}$ és $C_{10} = C_{01}$. Ilyenkor a konstans előfordulása gyakoribb, a küszöb lecsökken, ellenkező esetben nő.

3. Figyeljük meg, hogy a (23) összefüggésben milyen hatású a zaj varianciája, az együttes megfigyelések száma és maga konstans jel szintje.

2. Példa: Változó amplitúdójú jel detektálása zajos csatornán keresztül: jelen van-e a jel a megfigyelésekben vagy nincsen? Tegyük fel, hogy a megfigyelések független, Gauss eloszlású valószínűségi változók nulla várható értékkel, és σ_n^2 varianciával. A H_0 hipotézis az, hogy a megfigyelés $z_k = n_k$, azaz a jel nincsen jelen, csak a zaj aktuális mintáját kapjuk. A H_1 hipotézis az, hogy a megfigyelés $z_k = a_k + n_k$, azaz a jel jelen van, a jel és a zaj aktuális mintájának összegét figyeltük meg. $k = 1, 2, \dots, N$, azaz összesen N mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

$$f(z|H_0) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{z_k^2}{2\sigma_n^2}}, \text{ ill. } f(z|H_1) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(z_k - a_k)^2}{2\sigma_n^2}} \quad (24)$$

A megfigyelések függetlensége miatt a N megfigyelés együttes sűrűségfüggvénye az egyedi megfigyelések sűrűségfüggvényeinek szorzata. (19) aktuális alakja:

$$\Lambda(z) = \prod_{k=1}^N \frac{f(z_k|H_1)}{f(z_k|H_0)} = \prod_{k=1}^N e^{\frac{-(z_k - a_k)^2}{2\sigma_n^2} - \frac{-z_k^2}{2\sigma_n^2}} \underset{H_0}{>} \underset{H_1}{<} \eta, \quad (25)$$

illetve a log-likelihood arány:

$$\ln \Lambda(z) = \lambda(z) = -\sum_{k=1}^N \frac{(z_k - a_k)^2}{2\sigma_n^2} + \sum_{k=1}^N \frac{z_k^2}{2\sigma_n^2} = \frac{1}{\sigma_n^2} \sum_{k=1}^N z_k a_k - \frac{1}{2\sigma_n^2} \sum_{k=1}^N a_k^2 \underset{H_0}{>} \underset{H_1}{<} \ln \eta = \gamma \quad (26)$$

Átrendezve:

$$\boxed{\frac{1}{N} \sum_{k=1}^N a_k z_k \underset{H_0}{>} \underset{H_1}{<} \frac{\sigma_n^2}{N} \ln \eta + \frac{1}{2N} \sum_{k=1}^N a_k^2} \quad (27)$$

A teszthez tehát a megfigyelések – a jel mintáival – súlyozott átlagát kell képeznünk, és összevetnünk egy küszöbértékkel. A „döntőkészülék” blokkvázlatát a 9. ábra mutatja be.

Megjegyzés:

1. Vegyük észre, hogy a (27) összefüggésből (23) egyszerűen származtatható, ha a jel mintái rendre egyformák.
2. A (27) összefüggés szerinti jel-súlyozást (a jelhez „illesztett” szűrőnek nevezzük.

3. példa: Véletlen amplitúdójú jel detektálása zajos csatornán keresztül: jelen van-e a jel a megfigyelésekben vagy nincsen? Tegyük fel, hogy a jel és a zaj egyaránt időben diszkrét, stacionárius sztochasztikus fehér zaj folyamatok, Gauss eloszlással, nulla várható értékkel, és σ_a^2 , ill. σ_n^2 varianciával. A H_0 hipotézis az, hogy a megfigyelés $z_k = n_k$, azaz a jel nincsen jelen, csak a zaj aktuális mintáját kapjuk. A H_1 hipotézis az, hogy a megfigyelés $z_k = a_k + n_k$, azaz a jel jelen van, a jel és a zaj aktuális mintájának összegét figyeltük meg. $k = 1, 2, \dots, N$, azaz összesen N mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

$$f(z|H_0) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{z_k^2}{2\sigma_n^2}}, \text{ ill. } f(z|H_1) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_a^2 + \sigma_n^2}} e^{-\frac{z_k^2}{2(\sigma_a^2 + \sigma_n^2)}} \quad (28)$$

A megfigyelések függetlensége miatt a N megfigyelés együttes sűrűségfüggvénye az egyedi megfigyelések sűrűségfüggvényeinek szorzata. (19) aktuális alakja:

$$\Lambda(z) = \prod_{k=1}^N \frac{f(z_k|H_1)}{f(z_k|H_0)} = \prod_{k=1}^N \frac{\sigma_n}{\sqrt{\sigma_a^2 + \sigma_n^2}} e^{-\frac{z_k^2}{2(\sigma_a^2 + \sigma_n^2)} \underset{H_1}{>} \underset{H_0}{<} \eta}, \quad (25)$$

illetve a log-likelihood arány:

$$\ln \Lambda(z) = -\sum_{k=1}^N \frac{z_k^2}{2(\sigma_a^2 + \sigma_n^2)} + \sum_{k=1}^N \frac{z_k^2}{2\sigma_n^2} = \frac{\sigma_a^2}{2\sigma_n^2(\sigma_a^2 + \sigma_n^2)} \sum_{k=1}^N z_k^2 + \frac{N}{2} \ln \frac{\sigma_n^2}{\sigma_a^2 + \sigma_n^2} \underset{H_1}{>} \underset{H_0}{<} \ln \eta = \gamma \quad (29)$$

Átrendezve:

$$\frac{1}{N} \sum_{k=1}^N z_k^2 \underset{H_1}{>} \underset{H_0}{<} \frac{2\sigma_n^2(\sigma_a^2 + \sigma_n^2)}{\sigma_a^2} \left[\frac{1}{2} \ln \frac{\sigma_a^2 + \sigma_n^2}{\sigma_n^2} + \frac{1}{N} \ln \eta \right] \quad (30)$$

A teszthez tehát a megfigyelések négyzetének az átlagát kell képeznünk, és összevetnünk egy küszöbértékkel. A „döntőkészülék” blokkvázlatát a 10. ábra mutatja be.

Megjegyzés:

1. Ha $\eta = 1$, és $\sigma_a^2 = \sigma_n^2$, akkor $\frac{1}{N} \sum_{k=1}^N z_k^2 \underset{H_1}{>} \underset{H_0}{<} 2\sigma_n^2 \ln 2$.

2. (30) alternatív alakja: $\frac{1}{N} \sum_{k=1}^N z_k^2 \underset{H_1}{>} \underset{H_0}{<} 2\sigma_n^2 \left(1 + \frac{\sigma_a^2}{\sigma_n^2}\right) \left[\frac{1}{2} \ln \left(1 + \frac{\sigma_a^2}{\sigma_n^2}\right) + \frac{1}{N} \ln \eta \right]$, amelyben a

varianciák arányának hatását elemezhetjük.

3. A 11. ábra a döntési tartományok elhelyezkedését mutatja. Látható, hogy elhelyezkedésük szimmetrikus, mert az eredmény nem függ a megfigyelt érték előjelétől.

4. Példa: Bayes döntés diszkrét valószínűségekkel: Tárgyfelvétel eldöntése az első előadáson szerzett benyomások (*érdekes* vagy *unalmas*) alapján. Definiáljuk az előadások három csoportját: *jó*, *közepes*, *rossz*. A hallgató előzetes tapasztalatai alapján azt tudja, hogy

$$P(\text{jó})=0.2, P(\text{közepes})=0.4 \text{ és } P(\text{rossz})=0.4.$$

Ezek az ún. *apriori* valószínűségek. Ismeri továbbá, hogy az előadásokról szerzett benyomások (*érdekes* vagy *unalmas*) hogyan függenek össze az egyes előadás csoportokkal. Ezek a csatornakaraktisztikának megfelelő feltételes valószínűségek:

$$P(\text{érdekes}|\text{jó})=0.8, P(\text{érdekes}|\text{közepes})=0.5, P(\text{érdekes}|\text{rossz})=0.1$$

$$P(\text{unalmas}|\text{jó})=0.2, P(\text{unalmas}|\text{közepes})=0.5, P(\text{unalmas}|\text{rossz})=0.9$$

A költségek/kockázatok mérőszámai legyenek a következők:

$$C_{\text{felvesszük}|\text{jó}} = 0, C_{\text{felvesszük}|\text{közepes}} = 5, C_{\text{felvesszük}|\text{rossz}} = 10$$

$$C_{\text{nem_vesszük_fel}|\text{jó}} = 20, C_{\text{nem_vesszük_fel}|\text{közepes}} = 5, C_{\text{nem_vesszük_fel}|\text{rossz}} = 0$$

A hallgatónak a jó döntéshez minimalizálnia kell a feltételes kockázatot/költséget. (A feltétel az első előadáson szerzett tapasztalat.) A kockázati/költség függvények:

$$R(\text{felvesszük} | \text{érdekes}), R(\text{felvesszük} | \text{unalmas})$$

$$R(\text{nem_vesszük_fel} | \text{érdekes}), R(\text{nem_vesszük_fel} | \text{unalmas})$$

Ezek közül az elsőt kifejtve $R(\text{felvesszük} | \text{érdekes}) =$

$$= C_{\text{felvesszük} | \text{jó}} P(\text{jó} | \text{érdekes}) + C_{\text{felvesszük} | \text{közepes}} P(\text{közepes} | \text{érdekes}) + C_{\text{felvesszük} | \text{rossz}} P(\text{rossz} | \text{érdekes})$$

Az itt szereplő feltételes valószínűségek az ún. *a posteriori* valószínűségek a Bayes tétel segítségével számíthatók:

$$P(\text{jó} | \text{érdekes}) = \frac{P(\text{érdekes} | \text{jó})P(\text{jó})}{P(\text{érdekes})}, P(\text{közepes} | \text{érdekes}) = \frac{P(\text{érdekes} | \text{közepes})P(\text{közepes})}{P(\text{érdekes})},$$

$$P(\text{rossz} | \text{érdekes}) = \frac{P(\text{érdekes} | \text{rossz})P(\text{rossz})}{P(\text{érdekes})}$$

Ezekben a számláló tényezői ismertek, egyedül $P(\text{érdekes})$ számítandó a teljes valószínűség tétele felhasználásával:

$$P(\text{érdekes}) = P(\text{érdekes} | \text{jó})P(\text{jó}) + P(\text{érdekes} | \text{közepes})P(\text{közepes}) + P(\text{érdekes} | \text{rossz})P(\text{rossz}) =$$

$$= 0.8 * 0.2 + 0.5 * 0.4 + 0.1 * 0.4 = 0.4$$

Ezzel:

$$P(\text{unalmas}) = 0.6, P(\text{jó} | \text{érdekes}) = 0.4, P(\text{közepes} | \text{érdekes}) = 0.5, P(\text{rossz} | \text{érdekes}) = 0.1$$

Ha a hallgató az első előadás után azt tapasztalja, hogy az előadás érdekes, akkor a $R(\text{felvesszük} | \text{érdekes})$, és a $R(\text{nem_vesszük_fel} | \text{érdekes})$ kockázatokat/költségeket fogja összevetni.

$$R(\text{felvesszük} | \text{érdekes}) = 0 * 0.4 + 5 * 0.5 + 10 * 0.1 = 3.5,$$

$$R(\text{nem_vesszük_fel} | \text{érdekes}) = 20 * 0.4 + 5 * 0.5 + 0 * 0.1 = 10.5,$$

a hallgató tehát fel fogja venni a tárgyat. Határozzák meg a kockázati/költség értékeket arra az esetre is, amikor az első előadással kapcsolatos tapasztalat az, hogy unalmas.

4. A becslésmélet alapjai

A cél az a paraméter(vektor) \hat{a} becslőjének meghatározása a 12. ábrán látható előzetesen ismert (és az azokból származtatott) sűrűségfüggvények ismeretében. A továbbiak követhetősége érdekében érdemes felidézni: (1) a sűrűségfüggvény fogalmát és viszonyát a mérési sorozatokkal, (2) a feltételes sűrűségfüggvény fogalmát (pl. a csatorna karakterisztika jellemzésére), és (3) a várható érték fogalmát, valamint kiszámítását a sűrűségfüggvényre alapozva. A becslés jellemzésére és jóságának mértékére szolgáló jellemzők (lásd 13. ábra):

$$1. \text{ Feltételes várhatóérték: } E\{\hat{a} | a\} = \int_{-\infty}^{\infty} \hat{a} f(\hat{a} | a) d\hat{a} \quad (31)$$

$$2. \text{ Feltételes kovariancia mátrix: } \text{cov}\{\hat{a}, \hat{a} | a\} = E\{(\hat{a} - E(\hat{a} | a))(\hat{a} - E(\hat{a} | a))^T | a\} \quad (32)$$

3. Feltételes torzítás: $b(a) = E\{\hat{a}|a\} - a$ (33)

4. Átlagos négyzetes hiba: $E\{(\hat{a} - a)(\hat{a} - a)^T | a\} = \text{cov}\{\hat{a}, \hat{a}|a\} + b(a)b^T(a)$ (34)
(Mean Square Error (MSE))

Megjegyzés: Ha ismert az $f(a)$ valószínűség-sűrűségfüggvény, akkor definiálható:

5. Feltétel nélküli várható érték: $E(\hat{a}) = E\{E\{\hat{a}|a\}\}$ (35)

6. Feltétel nélküli kovariancia mátrix: $\text{cov}\{\hat{a}, \hat{a}\} = E\{\text{cov}\{\hat{a}, \hat{a}|a\}\}$ (36)

Bayes becslők: Ismert a megfigyelt paraméter sűrűségfüggvénye: $f(a)$, és a csatorna karakterisztikája: $f(z|a)$. A megfigyelés után, ezek alapján a Bayes szabály felhasználásával előállítható az $f(a|z)$ ún. a posteriori sűrűségfüggvény:

$$f(a|z) = \frac{f(z|a)f(a)}{f(z)}, \text{ ahol} \quad (37)$$

$$f(z) = \int_{-\infty}^{\infty} f(a)f(z|a)da. \quad (38)$$

A 14. ábra azt mutatja, hogy amennyiben a megfigyelések tartalmazzak a paraméterre vonatkozó információt, az a posteriori sűrűségfüggvény a konkrét paraméter-értékek egy szűkebb környezetére terjed ki. Az a posteriori sűrűségfüggvény segítségével határozzuk meg az ismeretlen paraméter (vektor) lehető legjobb becslőjét. Ehhez értelmezünk a „legjobb” fogalmát alkalmas költségfüggvényeken keresztül:

$$R(\hat{a}, a) = E\{C(\hat{a}, a)\} = \int_{-\infty}^{\infty} C(\hat{a}, a)f(a|z)da, \quad (39)$$

aminek minimumát Bayes költségnek nevezzük:

$$R_B = \min_{\hat{a}} R(\hat{a}, a) = \min_{\hat{a}} E\{C(\hat{a}, a)\}. \quad (40)$$

Itt $C(\hat{a}, a)$ az ún. kritériumfüggvény, vagy hibafüggvény, vagy illesztési függvény. Ennek széles körben használt változatait a 15. ábra mutatja be.

I. $C(\hat{a}, a) = \sum_{i=1}^m (\hat{a}_i - a_i)^2 = (\hat{a} - a)^T (\hat{a} - a)$ (41)

II. $C(\hat{a}, a) = \sum_{i=1}^m |\hat{a}_i - a_i|$ (42)

III. $C(\hat{a}, a) = \begin{cases} 0 & \text{ha } |\hat{a}_i - a_i| \leq \frac{\Delta}{2} \quad \forall i - re \\ 1 & \text{egyébként} \end{cases}$ (43)

Minimális átlagos négyzetes hibájú becslés:

$$R(\hat{a}, a) = E\{(\hat{a} - a)^T (\hat{a} - a)\} = \int_{-\infty}^{\infty} (\hat{a} - a)^T (\hat{a} - a)f(a|z)da \quad (44)$$

Ennek a minimumát keressük:

$$\frac{\partial}{\partial \hat{a}} \int_{-\infty}^{\infty} (\hat{a} - a)^T (\hat{a} - a) f(a|z) da \Big|_{\hat{a} = \hat{a}_{MS}} = 0 \quad (45)$$

Mivel $\frac{\partial}{\partial \hat{a}} (\hat{a} - a)^T (\hat{a} - a) = 2(\hat{a} - a)$, továbbá a (45) kifejezésben \hat{a} az integráljel elé kiemelhető, valamint a sűrűségfüggvény integrálja 1, ezért

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

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

Minimális átlagos abszolút hibájú becslés (skalár esetre bemutatva):

$$R(\hat{a}, a) = E\{|\hat{a} - a|\} = - \int_{-\infty}^{\hat{a}} (\hat{a} - a) f(a|z) da + \int_{\hat{a}}^{\infty} (\hat{a} - a) f(a|z) da \quad (47)$$

Ennek a minimumát keressük:

$$\frac{\partial}{\partial \hat{a}} \left[- \int_{-\infty}^{\hat{a}} (\hat{a} - a) f(a|z) da + \int_{\hat{a}}^{\infty} (\hat{a} - a) f(a|z) da \right] \Big|_{\hat{a} = \hat{a}_{ABS}} = 0 \quad (48)$$

Ebből

$$\int_{-\infty}^{\hat{a}_{ABS}} f(a|z) da = \int_{\hat{a}_{ABS}}^{\infty} f(a|z) da, \quad (49)$$

vagyis

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

Maximum a posteriori (MAP) becslés:

$$R(\hat{a}, a) = \int_{-\infty}^{\hat{a} - \frac{\Delta}{2}} f(a|z) da + \int_{\hat{a} + \frac{\Delta}{2}}^{\infty} f(a|z) da = 1 - \int_{\hat{a} - \frac{\Delta}{2}}^{\hat{a} + \frac{\Delta}{2}} f(a|z) da. \quad (51)$$

Ha Δ kicsi, de $\Delta \neq 0$, akkor az optimum az a posteriori sűrűségfüggvény maximumhelye, mivel az (51) összefüggés ekkor veszi fel a legkisebb értékét.

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

