## 3. A döntéselmélet 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 $P_1 f(z|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P_0 f(z|H_0)$ alakban írható, ami megegyezik a $P(H_1|z) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0|z)$ 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 $\lambda(z) = \ln \Lambda(z) \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta = \gamma$, 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 $\sigma_n^2$ 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,… ,$N$, azaz összesen $N$ mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

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

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

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

illetve a log-likelihood arány:

$$\ln \Lambda(z) = \lambda(z) = -\sum_{k=1}^{N} \frac{(z_k - a)^2}{2\sigma_n^2} + \sum_{k=1}^{N} \frac{z_k^2}{2\sigma_n^2} = \frac{a}{\sigma_n^2} \sum_{k=1}^{N} z_k - \frac{Na^2}{2\sigma_n^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta = \gamma \tag{22}$$

Átrendezve:

$$\boxed{\frac{1}{N} \sum_{k=1}^{N} z_k \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma_n^2}{Na} \ln \eta + \frac{a}{2}} \tag{23}$$

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 $\frac{1}{N} \sum_{k=1}^{N} z_k \underset{H_0}{\overset{H_1}{\gtrless}} \frac{a}{2}$, 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 $\sigma_n^2$ 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,… ,$N$, azaz összesen $N$ mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

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

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

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

illetve a log-likelihood arány:

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

Átrendezve:

$$\boxed{\frac{1}{N} \sum_{k=1}^{N} a_k z_k \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\sigma_n^2}{N} \ln \eta + \frac{1}{2N} \sum_{k=1}^{N} a_k^2} \tag{27}$$

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 $\sigma_a^2$, ill. $\sigma_n^2$ 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,… ,$N$, azaz összesen $N$ mintát figyelünk meg egyszerre. A döntés során az együttes feltételes sűrűségfüggvényeket fogjuk használni. Egyetlen megfigyelés esetén a feltételes sűrűségfüggvények:

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

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

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

(25)

illetve a log-likelihood arány:

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

Átrendezve:

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

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

$P(jó)$=0.2, $P(közepes)$=0.4 és $P(rossz)$=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 csatornakarakterisztikának megfelelő feltételes valószínűségek:

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

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

A költségek/kockázatok mérőszámai legyenek a következők:

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

$$C_{nem\_vesszük\_fel|jó}=20,\ C_{nem\_vesszük\_fel|közepes}=5,\ C_{nem\_vesszük\_fel|rossz}=0$$

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

$$R(felvesszük \,|\, érdekes),\ R(felvesszük \,|\, unalmas)$$
$$R(nem\_vesszük\_fel \,|\, érdekes),\ R(nem\_vesszük\_fel \,|\, unalmas)$$

Ezek közül az elsőt kifejtve $R(felvesszük \,|\, érdekes) =$

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

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

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

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

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

$$P(érdekes) = P(érdekes \,|\, jó) P(jó) + P(érdekes \,|\, közepes) P(közepes) + P(érdekes \,|\, rossz) P(rossz) =$$
$$= 0.8 * 0.2 + 0.5 * 0.4 + 0.1 * 0.4 = 0.4$$

Ezzel:

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

Ha a hallgató az első előadás után azt tapasztalja, hogy az előadás érdekes, akkor a $R(felvesszük \,|\, érdekes)$, és a $R(nem\_vesszük\_fel \,|\, érdekes)$ kockázatokat/költségeket fogja összevetni.

$$R(felvesszük \,|\, érdekes) = 0 * 0.4 + 5 * 0.5 + 10 * 0.1 = 3.5,$$
$$R(nem\_vesszük\_fel \,|\, érdekes) = 20 * 0.4 + 5 * 0.5 + 0 * 0.1 = 10.5,$$

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

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

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

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

A 14. ábra azt mutatja, hogy amennyiben a megfigyelések tartalmaznak 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:

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

aminek minimumát Bayes költségnek nevezzük:

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

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

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

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

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

**Minimális átlagos négyzetes hibájú becslés:**

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

Ennek a minimumát keressük:

$$\left.\frac{\partial}{\partial \hat{a}} \int_{-\infty}^{\infty} (\hat{a}-a)^T (\hat{a}-a)\, f(a|z)\,da \right|_{\hat{a}=\hat{a}_{MS}} = 0 \tag{45}$$

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

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

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

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

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

Ennek a minimumát keressük:

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

Ebből

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

vagyis

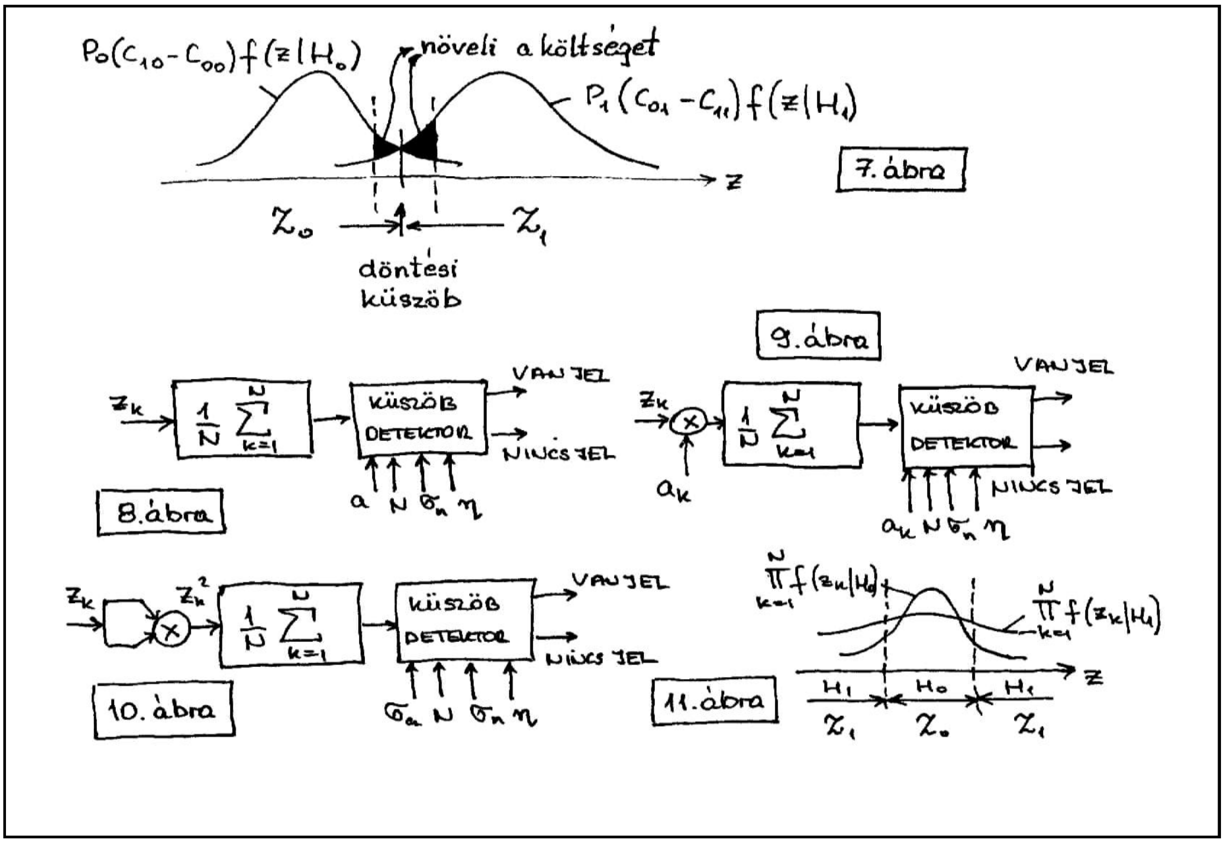
$$\boxed{\hat{a}_{ABS} = f(a|z) \text{ mediánja.}} \tag{50}$$

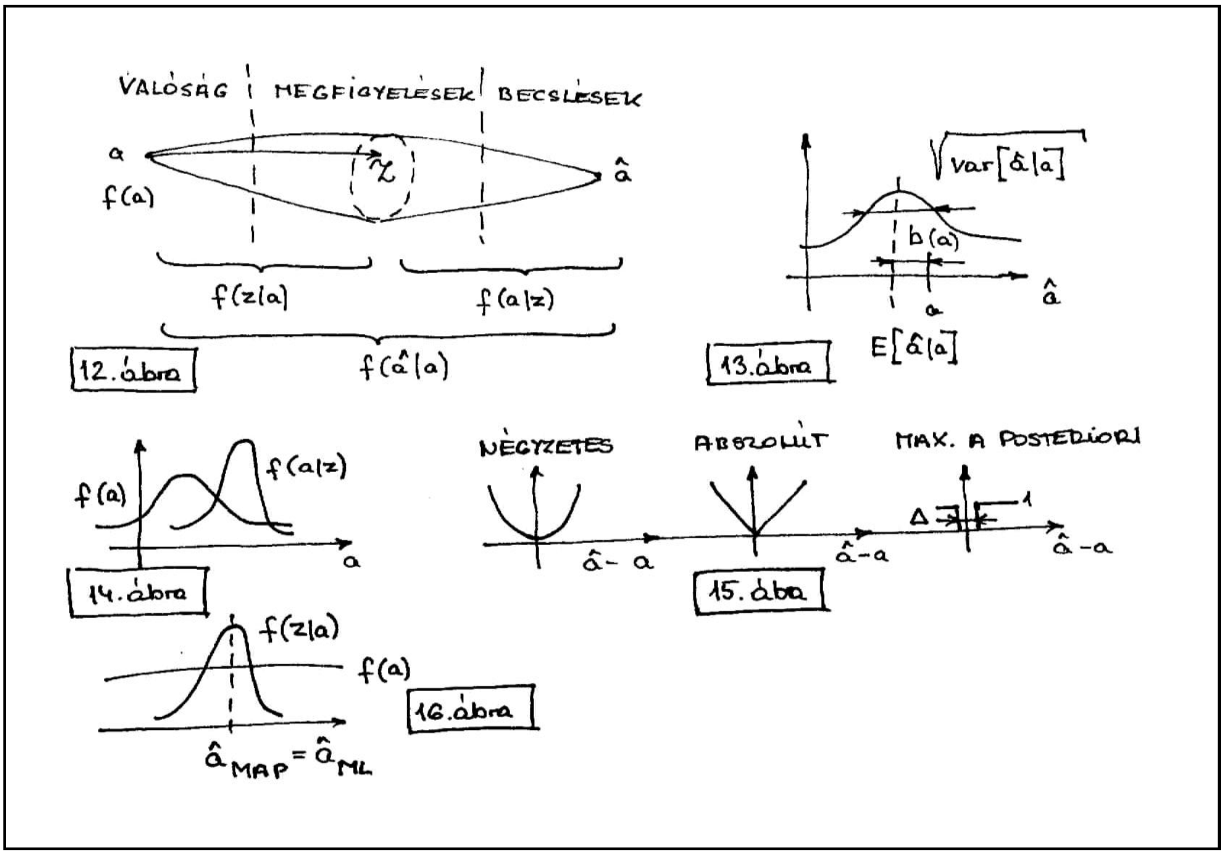
**Maximum a posteriori (MAP) becslés:**

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

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

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


P0(C10-C00)f(z|H0)
növeli a költséget
P1(C01-C11)f(z|H1)
z
7. ábra
Z0
Z1
döntési
küszöb
9. ábra
zk
1/N Σ k=1 N
KÜSZÖB
DETEKTOR
VAN JEL
NINCS JEL
a N σn η
8. ábra
ak
ak N σn η
zk²
σa N σn η
10. ábra
Π k=1 N f(zk|H0)
Π k=1 N f(zk|H1)
H1
H0
H1
Z1
Z0
Z1
11. ábra


VALÓSÁG
MEGFIGYELÉSEK
BECSLÉSEK
a
f(a)
z
â
f(z|a)
f(a|z)
f(â|a)
12. ábra
√var[â|a]
b(a)
â
a
E[â|a]
13. ábra
f(a)
f(a|z)
a
14. ábra
NÉGYZETES
ABSZOLÚT
MAX. A POSTERIORI
â - a
â-a
Δ
1
â -a
15. ábra
f(z|a)
f(a)
16. ábra
âMAP = âML