

# A maximum likelihood becslésről

## Definíció

Parametrikus becsléssel foglalkozunk. Adott egy modell, mellyel elképzeléseink szerint jól leírható a meghatározni kívánt rendszer. (A modell típusának és rendszámának megválasztásával most nem foglalkozunk, adottnak tekintjük.) A modellnek vannak szabad paraméterei, melyeket mérésel kívánunk meghatározni. A mérési eredményeink zajjal terheltek. Azaz, magukat a paramétereket nem tudjuk mérni, csak egy valószínűségi változót, mely több-kevesebb összefüggést mutat a modell paraméterrel. Kérdés, hogy a mért értékekből hogyan becsüljük meg a modell paramétereit, hogy a lehető legpontosabb becslést kapjuk. Egyáltalán mit értsünk "legpontosabban"?

Az egyik legáltalánosabb becslési stratégia, amit parametrikus becslésnél használunk, az ún. maximum likelihood eljárás. (A magyar szakirodalomban is az angol kifejezést használják, nincs meghonosodott magyar kifejezés rá, talán a legnagyobb valószínűség elvének fordíthatánk.) A módszert abban az esetben alkalmazzuk, mikor a modell paraméterek sűrűségfüggvényei ismeretlenek (hiszen a modellben szereplő paraméterek is valószínűségi változók a mérés szempontjából), viszont a mérést terhelő zaj eloszlása ismert. Amennyiben egy eloszlásról semmit sem tudunk, legkézenfekvőbb megoldás egyenletesnek tételezni föl.

A maximum likelihood becslés tehát a következőt jelenti: maximalizálni kell a

$$P \{ \text{ezt mértem} \mid \text{a paraméter ennyi és ennyi} \}$$

feltételes valószínűséget. Formálisan a bayes-döntésből vezethetjük le. A bayes döntést az alábbi formula írja le:

$$P \{ \mathbf{p} \mid \mathbf{y}_m \} = \frac{P \{ \mathbf{y}_m \mid \mathbf{p} \} P \{ \mathbf{p} \}}{P \{ \mathbf{y}_m \}} \quad (1)$$

(ahol  $\mathbf{p}$  a paramétervektort,  $\mathbf{y}_m$  a mért vektort jelenti) ML döntés esetén ez a képlet leegyszerűsödik:

$$P \{ \mathbf{p} \mid \mathbf{y}_m \} = C P \{ \mathbf{y}_m \mid \mathbf{p} \} \quad (2)$$

hiszen a paraméterek egyenletes eloszlásúak, a  $P \{ \mathbf{y}_m \}$  pedig csak súlyozó tényezőként működik.

Az  $L(\mathbf{y}_m | \mathbf{p}) = \mathbb{P}\{\mathbf{y} = \mathbf{y}_m | \mathbf{p}\}$  függvényt likelihood függvénynek nevezük, és mindig meghatározható kizárólag a mérési zaj eloszlásának ismeretében. A paramétervektor maximum likelihood becslése ( $\mathbf{p}_{ML}$ ) pedig a likelihood függvény  $\mathbf{p}$  szerinti maximalizálásával adódik.

Megjegyzés: A logaritmus függvény monotonitása miatt a maximalizálás szempontjából ekvivalens a likelihood függvény helyett annak logaritmusát maximalizálni, ami sokszor – számítástechnikai okból – célszerű lehet.

## 1. Példa

Adjunk becslést  $N$  db zajos mérésből egy vekni súlyára. A modellünk a következő:

$$y = g + n \quad (3)$$

ahol  $y$  a mért érték,  $g$  a valódi súly,  $n$  pedig a mérést terhelő zaj. Tudjuk, hogy a mérési zaj Gauss-eloszlású, az egyes mérések zaja egymástól független. Ezek alapján a likelihood függvény kiszámolható:

$$L(\mathbf{y}_m | g) = \prod_{i=1}^N \mathbb{P}\{y_i | g\} = \prod_{i=1}^N f_n(y_i | g) \quad (4)$$

ahol  $f_n$  jelöli a zaj sűrűségfüggvényét. Mivel a zaj normális eloszlású, így a likelihood függvény:

$$L(\mathbf{y}_m | g) = \frac{1}{\sqrt{2\pi\sigma_n^2}^N} \exp\left(-\sum_{i=1}^N \frac{(y_i - g)^2}{2\sigma_n^2}\right) \quad (5)$$

Esetünkben a likelihood függvény logaritmusát egyszerűbb lesz maximalizálni, így felírjuk az ún. log-likelihood függvényt:

$$\ln L = C - \frac{1}{2\sigma_n^2} \sum_{i=1}^N ((y_i - g)^2) \quad (6)$$

ahol  $C$  egy konstans. A vekni súlyának ML becslése ezek alapján.

$$g_{ML} = \frac{1}{N} \sum_{i=1}^N y_i \quad (7)$$

Vagyis eredményül azt kaptuk, hogy a vekni súlyának ML becslését úgy kapjuk, hogy a mért értékek számtani átlagát képezzük.

## 2. Példa

Legyen  $z_1, \dots, z_m$  egy normális valószínűségi változó független megfigyelései. A valószínűségi változó várható értéke legyen  $\mu$ , a szórása  $\sigma$ . Határozzuk meg ezen paraméterek ML becslését.

Első lépésként a likelihood függvényt ell felírunk.

$$L(z_1, \dots, z_m \mid \mu, \sigma) = L(\mathbf{z} \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}^m} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^m (z_i - \mu)^2\right) \quad (8)$$

A log-likelihood függvény:

$$\ln L = -\frac{m}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m (z_i - \mu)^2 \quad (9)$$

Most  $\mu$  és  $\sigma$  szerint külön-külön kell maximalizálni (9)-t, hogy rendre megkapjuk  $\mu$  és  $\sigma$  ML becslését:

$$\mu_{ML} = \frac{1}{m} \sum_{i=1}^m z_i \quad (10)$$

$$\sigma_{ML}^2 = \frac{1}{m} \sum_{i=1}^m (z_i - \mu_{ML})^2 \quad (11)$$

A várható érték becslésére a jól ismert mintaátlag adódott. Jegyezzük meg, hogy a várható értékre torzítatlan (lásd később), míg a szórásnégyzetre torzított becslését kaptuk a valódi paraméternek.

## Az ML becselő tulajdonságai

Az alábbiakban az ML becselő tulajdonságait összegezzük. Ez azért fontos, mert ha sikerül általános esetre belátni az alábbiakat, akkor egyedi esetekben nem kell végigszámolni a levezetéseket, hanem a becselő tulajdonságai "zsebből előhúzzhatóak". Általában a levezetések a következő feltételezésekkel élnek:

- a mérési zaj mérésről mérésre független, s ugyanolyan eloszlású (i.i.d);
- és a log-likelihood függvény kétszer differenciálható;

Egyedi esetekben előfordulhat, hogy kevésbé szoros feltevések mellett is bizonyítható némelyik tulajdonság.

### Egyértelműség

Bizonyítható, hogy a ML becslés egyértelmű a fenti feltevések mellett.

### Konzisztencia

Az ML becselő konzisztens. Vagyis igaz, hogy

$$\lim_{m \rightarrow \infty} P\{|\mathbf{p}_{ML} - \mathbf{p}| > \delta\} = 0 \quad \forall \delta > 0 \quad (12)$$

## Aszimptotikusan torzítatlan

Bizonyítható, hogy az ML becslő aszimptotikusan torzítatlan. Ez azt jelenti, hogy ha a mérések száma a végtelenbe nő, akkor a becslés torzítatlan lesz. (Torzítatlanságon a következőt értjük:

$$E[\mathbf{p}_{ML}] = \mathbf{p}$$

vagyis a becslő várható értéke megegyezik a valódi paraméterrel.) Bizonyítás helyett csak az előző példa kapcsán teszünk megjegyzést: a várható érték becslése (10) torzítatlan, hiszen (10) várható értéke éppen  $\mu$ . Ezzel szemben a szórásra kapott becslő (11) torzított, a torzítás mértéke  $\sigma^2/m$ . Azaz nagy ( $m \rightarrow \infty$ ) esetben a torzítás eltűnik: a becslő aszimptotikusan torzítatlan.

## Hatásosság

Az ML becslő kovarianciamátrixa aszimptotikusan tart a Fischer-információs mátrix inverzéhez, ami azt jelenti, hogy aszimptotikus értelemben a lehető legjobb becslő:

$$\mathbf{C}_{\mathbf{p}} = \mathbf{F}^{-1} \quad (13)$$

ahol  $\mathbf{F}$  a Fischer információs mátrix, melynek definíciója:

$$\mathbf{F} = E \left[ \left( \frac{\partial}{\partial \mathbf{p}} \ln L \right)^T \left( \frac{\partial}{\partial \mathbf{p}} \ln L \right) \mid \mathbf{p} \right] \quad (14)$$

A Fischer mátrix azt írja le, hogy mennyi a mérésekben jelenlevő információ-mennyiség a paraméterekre nézve. A (14) azt fejezi ki, hogy annál kisebb a becslés bizonytalansága, minél több információ van a mérési adatokban. Ezt az elvet lehet arra felhasználni, hogy olyan kísérleteket tervezzünk, melyek során a lehető legtöbb információt tartalmazó mérési eredmények születhetnek. Bizonyítható, hogy a Fischer információs mátrix inverzénél kisebb kovariancia mátrixú torzítatlan becslő nem létezik. Ez azt jelenti, hogy a becslő kovarianciájára létezik egy alsó határ, a mérési adatok függvényében. Ezt nevezzük Cramér-Rao korlátnak. (Létezés természetesen független az ML becslőtől.) Az ML becslő aszimptotikus értelemben megközelíti ezt a korlátot, ezért nevezzük *aszimptotikusan hatásosnak*.

## Aszimptotikusan normális eloszlású

A  $\mathbf{p}_{ML}$  becslő zajos mérési adatok függvénye, így önmaga is valószínűségi változó, amit a sűrűségfüggvényével írhatunk le. Ha a kísérletek száma nagy, akkor az ML becslő normális eloszlású lesz.

### **Az invariancia elv**

Ha  $\mathbf{p}_{ML}$  a  $K$ -dimenziós  $\mathbf{p}$  ML becslője, akkor  $\mathbf{g}(\mathbf{p}_{ML})$  az  $L$ -dimenziós  $\mathbf{g}(\mathbf{p})$  ML becslője,  $L \leq K$  esetén. A gyakorlatban ez egy nagyon fontos tulajdonság, hiszen például az előző példában a szórásnégyzet becslőjének kiszámolásából nem következtethetnénk a szórás becslőjére, ha ez az elv nem lenne érvényes.

### **Összefoglalás**

A maximum likelihood-becslő tulajdonságainak felsorolásából látható, hogy az ideális becslő minden tulajdonsága érvényes rá, bár csak aszimptotikus értelemben. Ezért a mérések számát nagynak kell választani, a jó minőségű becslés érdekében. Mindezeknek köszönhetően a parametrikus mérések világában a legelterjedtebb megközelítés a maximum likelihood becslés.