Mutatunk példát a (4.7)-ben definiált függvényre is. Amikor egy radioaktív anyag $T_{1/2}$ felezési idejét keressük, különböző $T_1$, $T_2$, ..., $T_n$ időpontokban mérünk beütésszámokat, amelyek várható értékét az

$$f_i(\mathbf{a}) = f_i(a_1, a_2) = a_1 \mathrm{e}^{-a_2 T_i} \tag{4.10}$$

függvénnyel írhatjuk le, ahol a keresett felezési időt a

$$T_{1/2} = \frac{\ln 2}{a_2}$$

összefüggésből határozhatjuk meg. Ha (4.10)-et (4.9)-be helyettesítjük, kapjuk a ($T_1$, $\xi_1$; $T_2$, $\xi_2$; ...; $T_n$, $\xi_n$) statisztikai minta likelihood-függvényét (azzal a feltételezéssel, hogy a $T_1$, $T_2$, ..., $T_n$ időpontok nem valószínűségi változók).

Minden, amit a keresett paraméterekről tudunk, az a $\vec{\xi}$ minta és a likelihood-függvény alakja. Ebből kell a keresett paramétereket a lehető legpontosabban meghatározni. Azt a célt tűzzük ki magunk elé, hogy megkeressük a számunkra legkedvezőbb (4.3) becslési eljárást, amin azt értjük, hogy *a becsült paraméterek szórása legyen a lehető legkisebb.*

Mind (4.8), mind (4.9) esetében feltettük, hogy a statisztikai minta elemei egymástól függetlenek. Ezeknél bonyolultabb alakú likelihood-függvényekre jutunk, ha ezt a feltevést elejtjük. A levezetendő Cramér-Rao egyenlőtlenség azonban ezekben az esetekben is igaz marad. A dolog lényegének a megértését megkönnyíti, ha először azt az esetet tekintjük, amikor csak egyetlen paramétert kell becsülnünk. A több paraméter esetére csak ezt követően térünk át.

## Egyetlen paraméter becslése. A Cramér-Rao egyenlőtlenség

Tegyük fel, hogy a $\vec{\xi}$ vektor komponensei diszkrét valószínűségi változók. (Folytonos eloszlás esetében a szumma helyett integrál áll. Egyébként az alábbi levezetések azonosak.) Legyen adva egy $t(\vec{\xi})$ *torzítatlan* becslési eljárás, tehát

$$\mathrm{M}\left[t\left(\vec{\xi}\right)\right] = \sum_{k=1}^{\infty} t(\mathbf{x}_k) L(\mathbf{x}_k, a) = a\,. \tag{4.11}$$

A folytonos változók esetére való általánosíthatóság $p_k$ helyett kedvéért itt a

$$p_k = L(\mathbf{x}_k, a)$$

jelölést használjuk, ami annak a valószínűségét adja meg, hogy $\xi = x_k$. Erre az eloszlásra (3.11a) szerint fennáll:

$$\sum_{k=1}^{\infty} L(\mathbf{x}_k, a) = 1\,. \tag{4.12}$$

Az általánosság megszorítása nélkül feltehetjük, hogy mindkét összegzésből kihagytuk azokat a tagokat, amelyekre $L(\mathbf{x}_k, a) = 0$. A (4.11) által kifejezett torzítatlanság fontos megszorítás, következményeire még visszatérünk. Feltesszük továbbá, hogy az összegzés (folytonos változó esetében az integrálás) és az $a$ szerint való differenciálás felcserélhető, továbbá hogy $a$-tól független azoknak az $\mathbf{x}_k$-értékeknek a halmaza, amelyekre $L(\mathbf{x}_k, a) \neq 0$. Ha e feltételek teljesülnek, azt mondjuk, hogy a becslési probléma *reguláris*. Ekkor egyszerűen deriválhatjuk (4.12)-t és (4.11)-et $a$ szerint:

$$\sum_{k=1}^{\infty} \frac{\partial L(\mathbf{x}_k, a)}{\partial a} = 0 \tag{4.13a}$$

és

$$\sum_{k=1}^{\infty} t(\mathbf{x}_k) \frac{\partial L(\mathbf{x}_k, a)}{\partial a} = 1 \,. \tag{4.13b}$$

Az előbbi egyenletet beszorozzuk $a$-val, majd az eredményt kivonjuk az utóbbi egyenletből:

$$\begin{aligned} 1 &= \sum_{k=1}^{\infty} \left[ t(\mathbf{x}_k) - a \right] \frac{\partial L(\mathbf{x}_k, a)}{\partial a} = \\ &= \sum_{k=1}^{\infty} \left[ t(\mathbf{x}_k) - a \right] \left[ \frac{1}{L(\mathbf{x}_k, a)} \frac{\partial L(\mathbf{x}_k, a)}{\partial a} \right] L(\mathbf{x}_k, a) = \\ &= \mathrm{M}\left[ \left( t(\vec{\xi}) - a \right) \left( \frac{1}{L(\vec{\xi}, a)} \frac{\partial L(\vec{\xi}, a)}{\partial a} \right) \right]. \end{aligned} \tag{4.14}$$

Az itt szereplő $(t(\vec{\xi}) - a)$ tényező várható értéke a torzítatlanság miatt zérus [vö. (4.11)]. A másik tényező várható értéke szintén 0 [vö. (4.13a)].[29] (4.14) szerint tehát a két valószínűségi változó kovarianciája 1-gyel egyenlő. Alkalmazzuk a Schwarz-féle egyenlőtlenséget [vö. (3.22)]:

$$\mathrm{D}^2\left[ t(\vec{\xi}) - a \right] \cdot \mathrm{D}^2 \left( \frac{1}{L(\vec{\xi}, a)} \frac{\partial L(\vec{\xi}, a)}{\partial a} \right) \geq 1 \,. \tag{4.15}$$

A további képletekben $t(...)$ és $L(...,...)$ argumentuma ugyanaz, mint itt, így az egyszerűbb írásmód kedvéért a továbbiakban elhagyjuk, de mindig beleértjük a képletekbe. Könnyen belátható, hogy

$$\frac{\partial^2 \ln L}{\partial a^2} = \frac{\partial}{\partial a} \left( \frac{1}{L} \frac{\partial L}{\partial a} \right) = -\left( \frac{1}{L} \frac{\partial L}{\partial a} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial a^2} \,,$$

továbbá, hogy a jobb oldal második tagjának a várható értéke zérus [vö. (4.13a)]. Ezzel

$$\mathrm{D}^2 \left( \frac{1}{L} \frac{\partial L}{\partial a} \right) = \mathrm{M}\left[ \left( \frac{1}{L} \frac{\partial L}{\partial a} \right)^2 \right] = \mathrm{M}\left( -\frac{\partial^2 \ln L}{\partial a^2} \right).$$

(4.15) szerint azt kaptuk tehát, hogy a becslés szórása alulról korlátos:

---

[29] A $t(\vec{\xi})$ és $L(\vec{\xi}, a)$ mennyiségek azért valószínűségi változók, mert a $\vec{\xi}$ valószínűségi változótól függnek. Ebben az értelemben beszélhetünk várható értékükről, szórásukról, kovarianciájukról stb.

$$\mathrm{D}^2(t-a) = \mathrm{D}^2(t) \geq \frac{1}{\mathrm{M}\left(-\frac{\partial^2 \ln L}{\partial a^2}\right)}. \tag{4.16}$$

Ez a *Cramér-Rao egyenlőtlenség* szokásos felírása. Nagy jelentősége van az ismeretlen paraméterek becslése szempontjából. Kimondjuk tétel formájában is:

4.1. TÉTEL. A becsült érték szórása – a torzítatlan becslések körében – alulról korlátos. Az alsó korlátot (4.16) adja meg.

Erre való tekintettel (4.16) jobb oldalának a nevezőjében álló mennyiséget *Fisher-féle információnak* nevezzük.[30]

Abban az esetben, amikor a becslés torzított, egyenlőtlenségünk módosul. A fenti levezetést megismételve[31] (4.16) helyett a

$$\mathrm{D}^2(t) \geq \frac{\left[1+\delta'(a)\right]^2}{\mathrm{M}\left(-\frac{\partial^2 \ln L}{\partial a^2}\right)} \tag{4.17}$$

korlátot kapjuk, ahol $\delta(a)$ a becslés torzítása [vö. (4.4a)]. A torzítástól függően tehát az alsó korlát módosul. A $\delta'(a) = -1$ szélső esetben az alsó korlát akár el is tűnhet. Erre triviális példa a következő. Tegyük fel, hogy a keresett paraméter a felezési idő, és a következő „becslést” alkalmazzuk: $\widetilde{a} = 30\ \mathrm{s}$. Tekintve, hogy ez konstans, szórása zérus. Nézzük meg, mit kapunk (4.17) szerint. A becslés torzítása ekkor

$$\delta(a) = 30\ \mathrm{s} - a,$$

amiből

$$\delta'(a) = -1,$$

vagyis az alsó korlát (4.17) szerint szintén eltűnik. Fenti eredményünk tehát érvényben marad. Ezt a szélsőséges példát a torzítatlansági feltétel fontosságának az illusztrálására mutattuk be: a *tetszőlegesen torzított* becslések körében akármilyen kis szórások elképzelhetők, de ezek mint becslési eljárások általában érdektelenek. A gyakorlatban csak a torzítatlan vagy csak elfogadhatóan kis mértékben torzított becslések jönnek szóba. Ezekre pedig a (4.17) egyenlőtlenség nem-zérus alsó korlátot jelent.

### A maximális valószínűség (maximum likelihood) módszere

E kis kitérő után térjünk vissza a torzítatlan becslésekhez. Mikor van (4.16)-ban egyenlőség? Amikor ez fennáll, becslési eljárásunkkal elértük a lehető legkisebb szórást, vagyis becslésünk *hatékony (efficiens)*. A (3.22) Schwarz-féle egyenlőtlenség akkor egyenlőség, amikor a benne szereplő valószínűségi változók egymásnak lineáris függvényei, vagyis esetünkben fennáll a

$$\frac{\partial \ln L(\xi, a)}{\partial a} = K(a)(t(\xi) - a) \tag{4.18}$$

[30] Nem tévesztendő össze a *Shannon-féle* információval.

[31] Ajánljuk az Olvasónak, hogy – gyakorlásképpen – végezze el a módosított levezetést.