

4. SEGÉDESZKÖZÖK MATEMATIKAI STATISZTIKÁBÓL

4.1. Alapfogalmak

A matematikai statisztika alapfogalma a *statisztikai minta*, ami nem más, mint mért (vagy megfigyelt) értékek együttese: $\xi_1, \xi_2, \dots, \xi_n$. Mint a 3. fejezetben tettük, a mérést vagy megfigyelést most is *kísérletnek* fogjuk nevezni. Mivel a kísérletben kapott mennyiségek értéke a kísérlet kimenetelétől függ, a statisztikai minta valószínűségi változók együttese. Jelöljük eloszlásfüggvényüket rendre $F_1(x)$ -, $F_2(x)$ -, ... $F_n(x)$ -szel. A matematikai statisztikában bevett szóhasználat szerint a kísérlet elvégzésével *mintát vettünk ezekből az eloszlásokból*. Ennyiben nevezhetjük a kísérletet *mintavételnek* is.

Ahhoz, hogy a kapott mintából gyakorlati következtetéseket vonhassunk le, ismernünk kell a minta (együttes) eloszlásfüggvényét: $F(x_1, x_2, \dots, x_n)$. Fontos speciális esetek a következők. A leggyakoribb, hogy az egyes mintaelemek mérése egymástól független:

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i). \quad (4.1a)$$

Ekkor *független mintavételről* beszélünk. Ebben a jegyzetben többnyire ilyen kísérletek kiértékelésével foglalkozunk. A dolog tovább egyszerűsödik, amikor az egyes mintaelemek eloszlásfüggvénye azonos: $F_i(x) \equiv F(x)$. Ilyenkor azt mondjuk, hogy a mintavétel ebből a közös eloszlásból történt. A minta eloszlásfüggvénye ezzel tovább egyszerűsödik:

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i). \quad (4.1b)$$

A gyakorlatban ritkán tudjuk az eloszlásfüggvényt pontosan megadni. Legtöbbször csak az eloszlásfüggvény *matematikai alakját* tudjuk felírni, és ebben általában szerepelnek ismeretlen paraméterek is (a_1, a_2, \dots, a_m). A kísérlet célja éppen ez utóbbiak meghatározása. Ebben a függelékben ennek matematikai részleteiről lesz szó. Végeredményben a statisztikai minta függvényében kapjuk meg a keresett paraméterek értékét. Ennek hangsúlyo-

zására szoktuk az eloszlásfüggvényt $F(x_1, x_2, \dots, x_n; a_1, a_2, \dots, a_m)$ alakban felírni. A rövideg kedvéért általában a vektori írásmódot használjuk: $F(\mathbf{x}, \mathbf{a})$. A levezetésekben többnyire ennek az x_1, x_2, \dots, x_n változók szerinti deriváltja, a sűrűségfüggvény jelenik meg. A sűrűségfüggvényt a 3. fejezetben – a szokásos módon – f -fel jelöltük. A mérések kiértékelésében mást jelölünk f -fel, így helyette az

$$L(x_1, x_2, \dots, x_n; a_1, a_2, \dots, a_m) = L(\mathbf{x}, \mathbf{a}) = \frac{\partial^n F(\mathbf{x}, \mathbf{a})}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (4.2)$$

jelölést használjuk. Ezt a függvényt *valószínűség-függvénynek* nevezzük. Gyakran lehet a *likelihood-függvény* elnevezéssel is találkozni.¹

A matematikai statisztika másik alapfogalma: a statisztikai minta bármilyen függvényét *statisztikának* nevezzük. Példák:

- mintaátlag: $\bar{\xi} = \sum_{i=1}^n \xi_i$,
- a mintaelemek maximuma: ξ_{\max} ,
- a mintaelemek minimuma: ξ_{\min} ,
- a minta terjedelme: $\xi_{\max} - \xi_{\min}$,
- empirikus szórásnégyzet: $s^2 = \frac{\sum_{i=1}^n (\xi_i - \bar{\xi})^2}{n-1}$,

és a sort lehetne folytatni.

A mérések kiértékelése szempontjából a legfontosabbak azok a statisztikák, amelyek az ismeretlen a_1, a_2, \dots, a_m paraméterek értékét megadják. Az ilyen statisztikákat a *paraméterek becsült értékének* nevezzük, amit a következő alakban írunk fel ($k = 1, 2, \dots, m$):

$$\tilde{a}_k = t_k(\xi_1, \xi_2, \dots, \xi_n) = t_k(\vec{\xi}). \quad (4.3)$$

Ezek a mennyiségek valószínűségi változók függvényei, így maguk is valószínűségi változók.

4.2. Paraméterbecslés

A becsült paraméterek kívánatos tulajdonságai

Egy becslési eljárástól az alábbi tulajdonságokat várjuk el. Mindenekelőtt megköveteljük, hogy *torzítatlan* legyen, vagyis fennálljanak a

¹ A likelihood angol szó jelentése: valószínűség. Az angol név kezdőbetűjéből ered az $L(\mathbf{x}, \mathbf{a})$ jelölés. A jelölésen kívül nem lenne más okunk az angol elnevezés megtartására.

$$M(\tilde{a}_k) = a_k, \quad k = 1, 2, \dots, m \quad (4.4)$$

egyenlőségek. Ez nincs mindig így. Ezért hasznosak a

$$\delta(a_k) = M(\tilde{a}_k) - a_k \quad (4.4a)$$

mennyiségek, amelyeket *torzításnak* nevezzük.

A becslt paraméter szórása a lehető legkisebb, lehetőleg zérushoz közel legyen. Ilyenkor ugyanis a Csebisev-egyenlőtlenségből következik, hogy a paraméter becslt értéke nagy valószínűséggel megegyezik a paraméter valódi értékével [vö. (3.15)]. Az alábbiakban megmutatjuk, hogy a szórás nem csökkenthető minden határon túl. Azt azonban mindenképpen elvárjuk, hogy a paraméterek becslésére használt eljárásunk az alsó határt elérje. Az ilyen becsléseket *hatékony (efficiens)* becsléseknek nevezzük.

További fontos jellemző a *konzisztencia*. Egy becslési eljárást *konzisztensnek* mondunk, ha a mérések n számának növekedésével a paraméterek becslt értékei a valódi értékekhez tartanak (sztochasztikus értelemben):

$$\lim_{n \rightarrow \infty} P\{|\tilde{a}_k - a_k| > \varepsilon\} = 0 \quad (4.5)$$

minden k -ra és tetszőleges pozitív ε -ra.

Az első két tulajdonság tetszőleges n -re vonatkozik, az utóbbi viszont csak az aszimptotikus viselkedést szabja meg abban az esetben, amikor a mérések száma minden határon túl nő. Nagyon gyakori eset, hogy a paraméterek becslt értékeinek a szórása növekvő n -nel 0-hoz tart:

$$\lim_{n \rightarrow \infty} D(\tilde{a}_k) = 0, \quad k = 1, 2, \dots, m. \quad (4.6)$$

Ekkor a (3.15a) Csebisev-egyenlőtlenségből következik a becslés konzisztenciája, hiszen

$$\lim_{n \rightarrow \infty} P\{|\tilde{a}_k - a_k| > \varepsilon\} \leq \lim_{n \rightarrow \infty} \frac{D^2(\tilde{a}_k)}{\varepsilon^2} = 0$$

minden k -ra. Ilyen esetekben elég azt vizsgálni, hogyan viselkednek a szórások nagy n -re. A leggyakrabban $1/\sqrt{n}$ rendben tartanak 0-hoz, de ennél gyorsabb csökkenésre is látunk majd példát.

A paraméterbecslés alapvető összefüggése a Cramér-Rao egyenlőtlenség², amelynek a megfogalmazásához szükség van a (4.2) képlettel definiált *valószínűség-függvényre*. Lássunk erre az eddig említettek köréből példákat! Legyen

² H. Cramér svéd, C. R. Rao indiai matematikus. Francia szerzők *Fréchet*-nek tulajdonítják az eredményt.

$$M(\xi_i) = f_i(a_1, a_2, \dots, a_m) = f_i(\mathbf{a}), \quad i = 1, 2, \dots, n. \quad (4.7a)$$

továbbá

$$Q = \sum_{i=1}^n \frac{[x_i - f_i(\mathbf{a})]^2}{\sigma_i^2}. \quad (4.7b)$$

Ha a $\vec{\xi}$ minta független, *Gauss-eloszlású* valószínűségi változókból áll, a valószínűség-függvény (3.37c) alapján

$$L(\mathbf{x}, \mathbf{a}) = \frac{\exp(-Q/2)}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}}. \quad (4.8)$$

A nukleáris mérés technikában gyakran találkozunk a *Poisson-eloszlással*, amely diszkrét eloszlás. Ekkor (3.35) szerint a valószínűség-függvény

$$L(\mathbf{x}, \mathbf{a}) = \prod_{i=1}^n e^{-f_i(\mathbf{a})} \frac{[f_i(\mathbf{a})]^{x_i}}{x_i!}. \quad (4.9)$$

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

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

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 $(T_1, \xi_1; T_2, \xi_2; \dots; T_n, \xi_n)$ statisztikai minta valószínűség-függvényét (azzal a feltételezéssel, hogy a T_1, T_2, \dots, T_n időpontok nem valószínűségi változók).

Minden, amit a keresett paraméterekről tudunk, az a $\vec{\xi}$ minta és a valószínűség-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 becslt paraméterek szórása legyen a lehető legkisebb*.

Mind (4.8), mind (4.9) esetében feltettük, hogy a statisztikai minta elemei egymástól függetlenek. Ezeknél bonyolultabb alakú valószínűség-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

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

A folytonos változók esetére való általánosíthatóság kedvéért itt a

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

jelölést használjuk. Erre az eloszlásra (3.11a) szerint fennáll:

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

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

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

és

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

Az előbbi egyenletet szorozzuk be a -val, majd az eredményt vonjuk ki az utóbbi egyenletből:

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

$$= \mathbb{M} \left[\left(t(\bar{\xi}) - a \right) \left(\frac{1}{L(\bar{\xi}, a)} \frac{\partial L(\bar{\xi}, a)}{\partial a} \right) \right]. \quad (4.14)$$

Az itt szereplő $(t(\bar{\xi}) - a)$ tényező várható értéke a torzítatlanság miatt zérus [vö. (4.11)]. A másik tényező várható értéke szintén 0 [vö. (4.13a)].³ Alkalmazhatjuk tehát a Schwarz-féle egyenlőtlenséget [vö. (3.22)]:

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

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

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

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

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

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

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

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

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

Erre való tekintettel a (4.16) jobb oldalának a nevezőjében álló mennyiséget *Fisher-féle információnak* nevezzük.⁴

³ A $t(\bar{\xi})$ és $L(\bar{\xi}, a)$ mennyiségek azért valószínűségi változók, mert a $\bar{\xi}$ valószínűségi változótól függenek. Ebben az értelemben beszélhetünk várható értékükről, szórásukról, kovarianciájukról stb.

⁴ Nem tévesztendő össze a *Shannon-féle* információval.

Abban az esetben, amikor a becslés torzított, egyenlőtlenségünk módosul. A fenti levezetést megismételve⁵ (4.16) helyett a

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

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: $\tilde{a} = 30$ s. Tekintve, hogy ez konstans, szórása zérus. Nézzük meg, mit kapunk (4.17) szerint. A becslés torzítása ekkor

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

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 érdektelenek. A gyakorlatban csak a torzítatlan vagy csak elfogadhatóan kis mértékben torzított becslések jönnek szóba. Ezekre pedig a (4.17) egyenlőtlenség nemzérus 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. Az ilyen becsléseket *hatékony (efficiens)* becsléseknek nevezzük. A (3.22) Schwarz-féle egyenlőtlenség akkor egyenlőség, ha a benne szereplő valószínűségi változók egymásnak lineáris függvényei, vagyis esetünkben fennáll a

$$\frac{\partial \ln L(\tilde{\xi}, a)}{\partial a} = K(a) \left(t(\tilde{\xi}) - a \right) \quad (4.18)$$

egyenlőség, amelyben K nem függ $\tilde{\xi}$ -től (de a -tól még függhet). Levezetéseinkben gondosan ügyeltünk a képletekben szereplő függvények argumentumaira. (4.18) a jobb és a bal oldalon szereplő *valószínűségi változók* között állapít meg összefüggést. Ha bennük a $\tilde{\xi}$ valószínűségi változó helyébe az

⁵ Ajánljuk az Olvasónak, hogy – gyakorlásképpen – végezze el a módosított levezetést.

\mathbf{x} változót írjuk, az egyenlet két oldalán szereplő függvények *alakjára* kapunk összefüggést:

$$\ln L(\mathbf{x}, a) = \varphi_1(a)t(\mathbf{x}) + \varphi_2(a) + \varphi_3(\mathbf{x}) \quad (4.19a)$$

Végeredményben tehát azt kaptuk, hogy a becslés szórása akkor és csak akkor veheti fel a minimumát, ha a valószínűség-függvény ilyen alakú. Fenti gondolatmenetünkéből következik, hogy az a becslés pedig, amelyik ezt biztosítja, a valószínűség-függvényben fellépő $t(\bar{\xi})$ függvény:

$$\tilde{a} = t(\bar{\xi}). \quad (4.20)$$

Ha a valószínűség-függvény alakja a (4.19a) képlet szerinti, a benne ilyen módon szereplő $t(\bar{\xi})$ -t *elégéses statisztikának* nevezzük. Az elnevezés értelme az, hogy ez az a statisztika, amely a $\bar{\xi}$ mintából az a paraméterre vonatkozóan minden mérési információt magába sűrít.⁶ (4.19a) szerint ugyanis a valószínűség-függvényt ilyen alakban írhatjuk fel:

$$L(\mathbf{x}, a) = \exp\{\varphi_3(\mathbf{x})\} \cdot \exp\{\varphi_1(a)t(\mathbf{x}) + \varphi_2(a)\}. \quad (4.19b)$$

Következésképpen L két tényező szorzatára bontható úgy, hogy csak az egyik függ a keresett paramétertől, továbbá ebben a tényezőben az \mathbf{x} változó csak a $t(\mathbf{x})$ szerinti kombinációban fordul elő. Ezt jelentette az a fenti kijelentés, hogy $t(\mathbf{x})$ “az a paraméterre vonatkozóan minden mérési információt magába sűrít”.

Ha tehát a hatékony becslést (4.20) szolgáltatja, akkor (4.18) szerint ezt úgy is kifejezhetjük, hogy a becslési eljárás során meg kell keresni a

$$\frac{\partial \ln L(\bar{\xi}, a)}{\partial a} = 0 \quad (4.21)$$

egyenletnek a -ra vonatkozó megoldását. Mivel az egyenlet bal oldalán szereplő derivált – (4.18) szerint – a megoldástól balra ($a < t$) pozitív, tőle jobbra ($a > t$) negatív, (4.21) megoldása a $\ln L$ függvény maximumának a keresését jelenti. Erre vezethető vissza az

4.1. DEFINÍCIÓ. A (4.21) egyenletnek a keresett a paraméterre való megoldását a *maximális valószínűség elvének* (vagy módszerének) nevezzük.

Az elmondottak lényegét pedig tétel formájában is kimondjuk:

⁶ Az *elégéses statisztika* a matematikai statisztika egyik legnehezebben érthető fogalma. Ha az Olvasónak első olvasáskor nehézségei vannak a megértéssel, a 4.1. DEFINÍCIÓ után ugorjon a (4.24) képletekhez. A későbbiekben és az eddigiek újraolvasásakor azonban el kell jutnia a fogalom megértéséig.

4.2. TÉTEL. A maximális valószínűség módszere hatékony becslést ad, amikor létezik elégséges statisztika.

Ha a valószínűség-függvény alakja nem felel meg a (4.19b) képletnek (és így nem létezik elégséges statisztika), a maximális valószínűség módszere továbbra is alkalmazható ugyan, de a segítségével kapott becslés nem lesz hatékony, hiszen ebben az esetben nem létezik hatékony becslés.

A további jelölésekben nem különböztetjük meg a diszkrét és folytonos valószínűségi változók eseteit. Az elégséges statisztikára vonatkozó példaként tekintsük az azonos várható értékű, Gauss-eloszlású változók esetét:

$$M(\xi_i) = a \quad \text{és} \quad D^2(\xi_i) = \sigma_i^2,$$

$i = 1, 2, \dots, n$. Együttes sűrűségfüggvényük

$$L(\mathbf{x}, a) = \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} \exp \left\{ -\sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma_i^2} \right\},$$

aminek a logaritmusa a (4.19a) szerinti alakra hozható:

$$\ln L(\mathbf{x}, a) = -\sum_{i=1}^n \ln(\sqrt{2\pi\sigma_i^2}) - \sum_{i=1}^n \frac{x_i^2}{2\sigma_i^2} + a \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{a^2/2}{\sigma_i^2},$$

vagyis

$$\frac{\partial \ln L}{\partial a} = \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - a \sum_{i=1}^n \frac{1}{\sigma_i^2} = \left(\frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} - a \right) \sum_{i=1}^n \frac{1}{\sigma_i^2}.$$

Ez éppen (4.19a) szerinti alakú, és látszik, hogy

$$t(\mathbf{x}) = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}. \quad (4.22)$$

Nyilvánvaló, hogy $M[t(\tilde{\xi})] = a$, tehát ez torzítatlan becslés. A

$$K = \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

tényező itt a -tól függetlennek adódott, de ez nincs mindig így.

Már jóval a maximális valószínűség módszerének a felfedezése előtt ismert volt Gauss-tétele, amely szerint a szórásnégyzetek reciprokával súlyozott (4.22) átlag a közös várható értéknek minimális szórású becslése a *lineáris* becslések között (5.2. alfejezet). Most az is kiderült, hogy ennek szórása az *összes becslések körében is minimális*.

A maximális valószínűség módszerét Fisher javasolta. A 4.3. alfejezet végén megadjuk az általa követett gondolatmenetet.

Példa nemreguláris becslési problémára

Reguláris becslési problémák esetében a szórásnak Cramér-Rao szerinti alsó korlátja általában \sqrt{n} -nel arányosan csökken. (n a mérési adatok száma.) A (4.1b) szerinti esetben ugyanis a Cramér-Rao egyenlőtlenség így írható [vö. (4.16)]:

$$D^2(t) \geq \frac{1}{n M \left(-\frac{\partial^2 \ln L}{\partial a^2} \right)}. \quad (4.23)$$

A gyakorlatban nagyon ritka, hogy a valószínűség-függvény ilyen alakú legyen. Ennek ellenére az $1/\sqrt{n}$ -nel arányos csökkenés nagyon jó közelítéssel fenn szokott állni.

Nemreguláris becslési problémákra általában nem alkalmazható a Cramér-Rao egyenlőtlenség. Nevezetes kivétel például az *egyenletes eloszlás* [vö. (3.32)], amely azért *nem reguláris*, mert nem lehet a fenti deriválást az integrálással felcserélni. Legyen θ az eloszlás terjedelme, vagyis

$$0 \leq \xi_i \leq \theta, \quad i = 1, 2, \dots, n.$$

θ becslésére felhasználható – például – a mintaátlag, amelynek a várható értéke $\theta/2$. Egyetlen változó szórása $\theta^2/12$, tehát a $\tilde{\theta} = 2\bar{\xi}$ becslés szórásnégyzete

$$D^2(2\bar{\xi}) = \frac{\theta^2}{3n},$$

vagyis ennek a becslésnek a szórása $1/\sqrt{n}$ rendben csökken, ami megegyezik a (4.23)-ból levont következtetéssel. Megmutatjuk azonban, hogy a maximális elemből kiindulva ennél jobb becslést is lehet kapni. ξ_{\max} sűrűségfüggvénye

$$f(x) = \frac{nx^{n-1}}{\theta^n},$$

várható értéke pedig $M(\xi_{\max}) = \frac{n\Theta}{n+1}$. Így a $\tilde{\Theta} = \frac{n+1}{n}\xi_{\max}$ torzítatlan becslés szórásnégyzete

$$D^2(\tilde{\Theta}) = \left(\frac{n+1}{n}\right)^2 D^2(\xi_{\max}) = \frac{\Theta^2}{n(n+2)},$$

vagyis ennek a becslésnek a szórása nagy n -re $1/n$ rendben csökken.

Ezzel az ellenpéldával kapcsolatban érdemes megjegyezni, hogy az átlag révén kapott becslés szinte reflexszerűen jut az eszünkbe, hiszen megszoktuk, hogy azonos eloszlású változók esetében az átlag jelenti a legjobb becslést. Ez így is van, amikor a becslési probléma reguláris. Az egyenletes eloszlás példája mutatja, hogy a reflexek nem mindig működnek jól, amivel mindig számolnunk kell, amikor a valószínűség-függvény nem reguláris.

*Több ismeretlen paraméter esete

Abban az esetben, amikor egyszerre több paramétert kell becsülnünk, az eddigi megfontolások lényege érvényben marad: továbbra is igaz, hogy minden paraméter becslült értékének a szórásnégyzete alulról korlátos. Ennek pontosabb megfogalmazásához bevezetjük a következő jelöléseket. \mathbf{I} -vel jelöljük az ún. *információs mátrixot*, amelynek (k, l) eleme

$$I_{kl} = M\left(-\frac{\partial^2 \ln L}{\partial a_k \partial a_l}\right), \quad (4.24a)$$

ahol – mint korábban – $L(\mathbf{x}, \mathbf{a})$ a valószínűség-függvény. Ezt a következő alakban is felírhatjuk:

$$I_{kl} = M\left(\frac{1}{L} \frac{\partial L}{\partial a_k} \cdot \frac{1}{L} \frac{\partial L}{\partial a_l}\right). \quad (4.24b)$$

Számítsuk ki ugyanis a (4.24a)-ban szereplő kétszeres deriváltat:

$$\frac{\partial^2 \ln L}{\partial a_k \partial a_l} = -\frac{1}{L^2} \frac{\partial L}{\partial a_k} \frac{\partial L}{\partial a_l} + \frac{1}{L} \frac{\partial^2 L}{\partial a_k \partial a_l}.$$

Könnyen belátható, hogy a második tag várható értéke 0:

$$M\left(\frac{1}{L} \frac{\partial^2 L}{\partial a_k \partial a_l}\right) = \int \frac{\partial^2 L(\mathbf{x}, \mathbf{a})}{\partial a_k \partial a_l} d\mathbf{x} = \frac{\partial^2}{\partial a_k \partial a_l} \int L(\mathbf{x}, \mathbf{a}) d\mathbf{x} = 0,$$

hiszen a kétszer derivált integrál értéke azonosan 1.⁷ Ebből viszont következik a (4.24) egyenletek jobb oldalán álló várható értékek azonossága.

⁷ A változatosság kedvéért a további levezetéseket nem a diszkrét, hanem a folytonos valószínűségi változókra vonatkoztatjuk.

Feltesszük, hogy az információs mátrix nem szinguláris.⁸ Az a_k paraméter becslésére szolgáljon a $t_k(\bar{\xi})$ függvény ($k = 1, 2, \dots, m$). \mathbf{B}_t a $\mathbf{t}(\bar{\xi})$ vektor kovarianciamátrixa. Ezekkel a jelölésekkel a Cramér-Rao-egyenlőtlenség többváltozós alakja az

4.3. TÉTEL. Tetszőleges \mathbf{z} vektorra igaz, hogy

$$\mathbf{z}^T \mathbf{B}_t \mathbf{z} \geq \mathbf{z}^T \mathbf{I}^{-1} \mathbf{z}, \quad (4.25a)$$

A bizonyítandó egyenlőtlenséget átírjuk a

$$\mathbf{z}^T (\mathbf{B}_t - \mathbf{I}^{-1}) \mathbf{z} \geq 0 \quad (4.25b)$$

alakba, ami szerint a zárójelben lévő mátrixról meg kell mutatnunk, hogy pozitív szemidefinit. Képezzük a $(t_k(\bar{\xi}) - a_k)$ és a $\partial \ln L(\bar{\xi}, \mathbf{a}) / \partial a_l$ zérus várható értékű valószínűségi változók kovarianciáját ($k, l = 1, 2, \dots, m$):

$$\begin{aligned} M \left[(t_k(\bar{\xi}) - a_k) \left(\frac{1}{L(\bar{\xi}, \mathbf{a})} \frac{\partial L(\bar{\xi}, \mathbf{a})}{\partial a_l} \right) \right] &= \int [t_k(\mathbf{x}) - a_k] \frac{\partial L(\mathbf{x}, \mathbf{a})}{\partial a_l} d\mathbf{x} = \\ &= \frac{\partial}{\partial a_l} \int t_k(\mathbf{x}) L(\mathbf{x}, \mathbf{a}) d\mathbf{x} - a_k \int \frac{\partial L(\mathbf{x}, \mathbf{a})}{\partial a_l} d\mathbf{x} = \frac{\partial}{\partial a_l} M(t_k(\bar{\xi})) = \frac{\partial a_k}{\partial a_l} = \delta_{kl}. \end{aligned}$$

Így tehát a $2m$ elemű

$$\left(t_1(\bar{\xi}), t_2(\bar{\xi}), \dots, t_m(\bar{\xi}), \frac{1}{L} \frac{\partial L(\bar{\xi}, \mathbf{a})}{\partial a_1}, \frac{1}{L} \frac{\partial L(\bar{\xi}, \mathbf{a})}{\partial a_2}, \dots, \frac{1}{L} \frac{\partial L(\bar{\xi}, \mathbf{a})}{\partial a_m} \right)$$

vektor kovarianciamátrixa a

$$\begin{bmatrix} \mathbf{B} & \mathbf{E} \\ \mathbf{E} & \mathbf{I} \end{bmatrix} \quad (4.26)$$

alakban írható, ahol mindegyik blokk $m \times m$ -es mátrix.⁹ (Az egyszerűbb írásmód kedvéért \mathbf{B}_t mellől elhagytuk a “t” indexet.) A (4.26) mátrixról tudjuk, hogy pozitív szemidefinit, hiszen minden kovarianciamátrix ilyen, vagyis tetszőleges \mathbf{z}_1 és \mathbf{z}_2 vektorokkal fennáll a

⁸ Ha szinguláris lenne, akkor ez azt jelentené, hogy a becslt paraméterek nem lineárisan függetlenek. Ezt az esetet kizárhatjuk, mert ez annak lenne a jele, hogy a paraméterbecslési problémát rosszul fogalmaztuk meg.

⁹ A hipermátrixokkal a 2.4. alfejezetben foglalkozunk.

$$\begin{bmatrix} \mathbf{z}_1^T & \mathbf{z}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{B} & \mathbf{E} \\ \mathbf{E} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \geq 0 \quad (4.27)$$

egyenlőtlenség. Ha a beszorzást elvégezzük, a

$$\mathbf{z}_1^T \mathbf{B} \mathbf{z}_1 + \mathbf{z}_2^T \mathbf{I} \mathbf{z}_2 + 2 \mathbf{z}_1^T \mathbf{z}_2 \geq 0 \quad (4.28)$$

egyenlőtlenséget kapjuk. A \mathbf{z}_2 vektort a

$$\mathbf{z}_2 = -\mathbf{I}^{-1} \mathbf{z}_1$$

egyenlet szerint választjuk meg. Mivel \mathbf{I} nem szinguláris, tetszőleges \mathbf{z}_1 mellett létezik ilyen \mathbf{z}_2 vektor. Ekkor a (4.28) egyenlőtlenség átmegy a

$$\mathbf{z}_1^T (\mathbf{B} - \mathbf{I}^{-1}) \mathbf{z}_1 \geq 0$$

alakba. Ezzel az állítást bebizonyítottuk.

A 3.4. alfejezetben írtak alapján beláthatjuk, hogy a (4.25a) egyenlőtlenség bal oldalán éppen a $\mathbf{z}^T \mathbf{t}(\tilde{\xi})$ skalárszorzat szórásnégyzete szerepel. A most bizonyított tétel azt jelenti, hogy az \mathbf{I} információs mátrixból kiindulva a becült paraméterek bármilyen lineáris kombinációjának a szórása számára lehet alsó korlátot levezetni. Ennek speciális esete a következő. Legyen \mathbf{z} az \mathbf{e}_k egységvektor, amelynek minden eleme zérus, kivéve a k -adikat, amely 1. Ekkor a (4.25a) egyenlőtlenségből következik, hogy

$$[\mathbf{B}_t]_{kk} = D^2 \left[t_k(\tilde{\xi}) \right] \geq [\mathbf{I}^{-1}]_{kk}, \quad (4.29)$$

vagyis a k -adik becült paraméter szórásnégyzete nem lehet kisebb, mint az információs mátrix inverzének a főátlójában álló k -adik elem ($k = 1, 2, \dots, m$). Tehát amit egyetlen becült paraméter esetében találtunk, érvényes több becült paraméter esetében is, legfeljebb az alsó korlát meghatározása nem egy szám reciprokának, hanem egy mátrix inverzének a kiszámítását igényli.

A maximális valószínűség módszerével kapott becslés tulajdonságai

A maximális valószínűség módszerével kapott becslés tulajdonságait illetően néhány tételt fogalmazunk meg, amelyek mind aszimptotikusan érvényesek $n \rightarrow \infty$ esetén:

4.4. TÉTEL. A (4.21) egyenlet megoldhatóságának a valószínűsége 1-hez tart, amikor $n \rightarrow \infty$.

4.5. TÉTEL. A (4.21) egyenlet megoldása $n \rightarrow \infty$ esetén 1-hez tartó valószínűséggel a valószínűség-függvény maximumát adja.

4.6. TÉTEL. A kapott becslés konzisztens, vagyis $\varepsilon > 0$ -ra

$$\lim_{n \rightarrow \infty} P\{|\tilde{a} - a| > \varepsilon\} = 0.$$

4.7. TÉTEL. A (4.21) egyenlet megoldása aszimptotikusan Gauss-eloszlású, amelynek a várható értéke a .

4.8. TÉTEL. A (4.21) egyenlet megoldása az a paraméternek aszimptotikusan hatékony becslése.

A felsorolt tételek teljesüléséhez szükséges feltételeket nem adjuk meg, mert az általunk tárgyalt kísérletek esetében mindig teljesülnek. Részletes megfogalmazásuk megtalálható például Linnyik könyvében [1].

Végül megjegyezzük, hogy a kimondott tételek több becslt paraméter esetében is igazak. Ebben a jegyzetben a tételek közül a 4.7. TÉTEL két állítására hivatkozunk a leggyakrabban. Fontos kijelentés ugyanis, hogy a becslt paraméterek általában Gauss-eloszlásúak. Ennek a jelentősége abban áll, hogy a paraméterbecslési eljárások erre az eloszlásra vannak a legjobban kidolgozva. A másik kijelentésnek elsősorban a negatív tartalmára kell felhívni a figyelmet: előfordulhat, hogy a maximális valószínűség módszerével *csak aszimptotikusan* kapunk torzítatlan becslést. Mivel nem végezhetünk végtelen számú kísérletet, n mindig véges, tehát minden esetben ajánlatos ellenőrizni a kapott becslések torzítatlanságát. Ha torzítást találunk, ki kell dolgoznunk a megfelelő korrekciót, és a becslést úgy módosítani, hogy végül torzítatlan legyen.

4.3. Hipotézisek vizsgálata

Ismeretlen paraméterek becslése mellett a matematikai statisztika másik fő feladata elméleti hipotézisek helyességének kísérleti ellenőrzése. Ennek ugyanolyan részletesen kidolgozott elmélete és módszertana van, mint a paraméterbecslésnek. Nincs lehetőségünk mindennek akár csak vázlatos ismertetésére sem. Ezért csak a jegyzet témája szempontjából legfontosabb dolgok magyarázatára szorítkozunk. Mindenekelőtt néhány példát hozunk hipotézisekre, amelyeknek a vizsgálata a gyakorlatban felmerül:

- Mennyiségek egyenlősége: kísérletileg meghatározzuk egy paraméter értékét, és ezt összehasonlítjuk ugyanennek valamilyen számított értékével. A vizsgálandó hipotézis ekkor a számított paraméterértéknek a mért paraméterérték várható értékével való egyenlősége.
- Illeszkedésvizsgálat: a közvetlenül mért $\xi_1, \xi_2, \dots, \xi_n$ mennyiségek eloszlásfüggvényére vonatkozóan szükségünk van valamilyen (többnyire elméleti megfontolásokkal kapott) feltevésre. A vizsgálandó hipotézis ekkor az, hogy tekinthetők-e a $\xi_1, \xi_2, \dots, \xi_n$ mennyiségek a feltételezett eloszlásból vett mintának.
- Mennyiségek összehasonlítása: tegyük fel, hogy két független mérési módszerrel meghatároztuk ugyanazt a mennyiséget, és azt vizsgáljuk, me-

lyik módszer a pontosabb. Ilyenkor a vizsgálandó hipotézis abban áll, hogy az első módszer szórása kisebb, mint a másodiké.

A példák sorát folytathatnánk. Fogalmazzuk meg a problémát általánosan! A vizsgálandó hipotézist H_0 -al jelöljük, és *null-hipotézis*nek nevezzük, amellyel szemben áll a H_1 *alternatív hipotézis*. Az utóbbi lényegében a null-hipotézis tagadása, de nem mindegy, hogyan fogalmazzuk ezt meg. Amikor például a mért és számított paraméterértékek, a_m és a_s egyenlőségét vizsgáljuk, ennek ellentettjét több módon is kimondhatjuk: $a_m \neq a_s$, vagy $a_m < a_s$, vagy $a_m > a_s$, és így tovább. H_0 vizsgálata abban áll, hogy a mért értékek $\vec{\xi}$ vektora számára az n -dimenziós térben kijelölünk egy ún. *elfogadási tartományt*, és H_0 -at igaznak fogadjuk el, ha $\vec{\xi}$ ebbe esik. Ellenkező esetben a H_1 *alternatív hipotézis* javára döntünk.

A gyakorlatban természetesen az a legtrikább eset, hogy az elfogadási tartományt közvetlenül a $\vec{\xi}$ vektor számára jelöljük ki az n -dimenziós térben. Általában redukáljuk a közvetlenül mért mennyiségeket, és az elfogadási tartományt a redukált mennyiségekre adjuk meg. Ennek legközönségebb módja bizonyos paraméterek becslése. Vegyük ismét a fenti példánkat. Az a_m "mért érték" a kérdéses a paraméter becsült értéke, tehát valószínűségi változó. A H_0 hipotézist tehát így fogalmazhatjuk:

$$H_0: M(a_m) = a_s. \quad (4.30)$$

Legyen az a_m becslés szórása σ . Ha igaz a null-hipotézis, a

$$\zeta = \frac{a_m - a_s}{\sigma} \quad (4.31)$$

mennyiség 0 várható értékű, 1 szórású, Gauss-eloszlású valószínűségi változó, amelyre vonatkozóan már egyszerűen szerkeszthetünk elfogadási tartományt. Ennek érdekében a következő gondolatmenetet alkalmazzuk. Választunk egy kis ε számot (például 0,01 vagy 0,05) úgy, hogy gyakorlatilag kizártnak tartjuk azokat az eseményeket, amelyek valószínűsége ennél kisebb. Keresünk egy olyan γ számot, amelyre fennáll, hogy

$$P\{|\zeta| < \gamma\} = F(\gamma) - F(-\gamma) = 1 - \varepsilon, \quad (4.32)$$

ahol $F(x)$ ζ eloszlásfüggvénye. Esetünkben ez a (3.36) sűrűségfüggvény integrálja $a = 0$ és $\sigma = 1$ mellett. Mivel a kapcsos zárójelen belül levő esemény – fenti kijelentésünk szerint – gyakorlatilag biztos esemény, ha igaz a null-hipotézis, azt mondhatjuk, hogy a null-hipotézist csak akkor fogadjuk el, ha a kapcsos zárójelen belüli egyenlőtlenség fennáll. ζ helyére (4.31) jobb oldalát helyettesítve a

$$-\gamma < \frac{a_m - a_s}{\sigma} < \gamma$$

elfogadási tartomány adódik, vagy átrendezve:

$$a_m - \gamma\sigma < a_s < a_m + \gamma\sigma. \quad (4.33)$$

A null-hipotézist tehát akkor fogadjuk el, ha a_s az itt szereplő két határ közé esik.

Az eddig bevezetett mennyiségeket a matematikai statisztikában a következőképpen nevezzük:

ε . *konfidencia-valószínűség* vagy *konfidenciaszint*. Az elnevezés logikus: a valószínűségeknek ez az a szintje, amely alatti valószínűségű eseményeket kizártnak tartjuk. A szóhasználat nem egészen egyértelmű, mert néha ε -t, néha $(1 - \varepsilon)$ -t nevezzük konfidenciaszintnek, viszont majdnem mindig százalékban adjuk meg. Az $\varepsilon = 0,05$ értéket például egyaránt mondjuk 5%-os és 95%-os konfidenciaszintnek. Félreértés ebből nem származhat, ugyanis ε ritkán nagyobb 0,1-nél.

γ . *kvantilis*. Mindig (4.32) alakú egyenletek megoldásaként számítjuk ki. Értéke függ mindenekelőtt a választott konfidenciaszinttől, de függ attól is, ahogy a ζ valószínűségi változót előállítottuk. Ezen keresztül függ a közvetlenül mért adatok n számától és a paraméterbecslés módjától. A kvantilisokat a leggyakrabban előforduló eloszlásokra (lásd 3.2. alfejezet) a statisztikai szakkönyvek n és ε szerint szerkesztett táblázatokban közlik. Jegyzetünk 2. függeléke tartalmazza a legfontosabb eloszlások kvantiliseit.

A (4.33)-ban szereplő $(a_m - \gamma\sigma, a_m + \gamma\sigma)$ intervallumot *konfidenciaintervallumnak* nevezzük. Kísérletünk alapján tulajdonképpen ez a legtöbb, amit a keresett a paraméterről mondani tudunk: értéke $(1 - \varepsilon)$ valószínűséggel a konfidenciaintervallum belsejébe esik. Erre való tekintettel ezt *intervallumbecslésnek* is nevezzük – szemben az a_m *pontbecsléssel*. Az intervallum $\gamma\sigma$ fél szélességét *mérési bizonytalanságnak* nevezzük.

Befejezésül megbeszéljük a hipotézisek vizsgálatában elkövethető kétféle hibát. Előfordulhat, hogy a null-hipotézis valójában igaz, de mi mégis elvetjük. A fenti példában az akkor következik be, amikor ζ abszolút értéke nagyobb, mint γ . Ez az *elsőfajú hiba*. Elkövetésének a valószínűsége definíció szerint ε . Hiába gondoltuk fentebb, hogy az ennél kisebb valószínűségű eseményeket kizárhatónak tekintjük, a véletlen játéka következtében mégis előfordulhatnak.

Az elsőfajú hiba csökkentése kedvéért az az érdekünk, hogy ε értékét minél kisebbre válasszuk. Van azonban más szempont is. Minél kisebb ε , annál szélesebb a konfidenciaintervallum, így annál valószínűbb, hogy elkövetjük

az ún. *másodfajú hibát*: jöllehet a null-hipotézis nem igaz, mégis elfogadjuk. A fenti példánkban ez azt jelenti, hogy ugyan az a paraméter a_s számított értéke rossz, mi mégis a mért értékkel egyezőnek találjuk, tehát a számítási módszert kísérletileg igazoltnak tekintjük. A másodfajú hiba csökkentése érdekében célszerű minél nagyobb ε -t választani.

A mi feladatunk a kétfajta hiba csökkentéséből adódó, egymásnak ellentmondó követelmények között az egyensúlyt megtalálni. Elemeznünk kell mindkettő hatását, meg kell határoznunk, melyiket tartjuk veszélyesebbnek, és annak megfelelően kell a konfidenciaszintet megválasztanunk. A másodfajú hiba valószínűsége függ a H_1 alternatív hipotézistől is, tehát ezt is körültekintően kell megfogalmazni.

A maximális valószínűség elvének heurisztikus levezetése

A maximális valószínűség alapelvét a fentiek alapján az alábbi megfontolással is megvilágíthatjuk. Mivel feltettük, hogy a becslés torzítatlan, a $t(\bar{\xi})$ statisztika nagy valószínűséggel az a paraméter valódi értéke közelében lesz. Legyen a (4.16)-ban szereplő szórásnégyzet négyzetgyöke σ_t , és tegyük fel, hogy $t(\bar{\xi})$ Gauss-eloszlású. Ekkor 95% annak a valószínűsége (lásd a 4.3. alfejezetben), hogy

$$a - 2\sigma_t < t(\bar{\xi}) < a + 2\sigma_t \quad (4.34)$$

teljesüljön. Legyen $\Xi(a)$ a $\bar{\xi}$ véletlen vektoroknak az a tartománya, amelyekre a (4.34) egyenlőtlenségek teljesülnek. (Ez a tartomány nyilván függ a értékétől.) Helyettesítsük $\bar{\xi}$ mért értékét a valószínűség-függvénybe, és vizsgáljuk $L(\bar{\xi}, a)$ -t a függvényében. Ha a értéke olyan, hogy $\bar{\xi}$ kívül esik a $\Xi(a)$ tartományon, akkor nagyon valószínű, hogy ez az a érték távol esik a paraméter valószínűségi értékétől, hiszen ebben az esetben a mérés eredménye egy kis valószínűségű tartományba esik. Ezért logikus az ismeretlen paraméter számára olyan értéket választani, amelyre a megfigyelt $\bar{\xi}$ vektor mellett a (4.34) egyenlőtlenségek kielégülnek. Ennek a valószínűsége akkor a legnagyobb, amikor a valószínűség-függvény a függvényében éppen felveszi a maximumát. Heurisztikus gondolatmenetünk logikája tehát így összegezhető: válasszuk meg a keresett paraméter értékét úgy, hogy a kapott $\bar{\xi}$ mérési eredmény a lehető legvalószínűbb legyen. Ezt a paraméterbecslési alapelvet alaposan megvizsgálták (lásd például [1], Cramér), és azt találták, hogy nagyon kedvező tulajdonságai vannak. Ezért terjedt el a gyakorlatban.

*4.4. Konfidenciaellipszoid

Amikor csak egy paramétert vizsgálunk, konfidenciaintervallumot jelölünk ki a keresett paraméter valódi értéke számára. Amikor azonban több

paraméter *egyszerre* érdekel bennünket, ez nem elég, mert egy *konfidencia-tartományt* kell kijelölnünk. Nos, ezt a 3.4. alfejezetben tárgyalt korrelációs ellipszoid és a χ^2 -eloszlás segítségével tehetjük meg.

A korrelációs ellipszoidot úgy kaptuk, hogy kerestük azokat az \mathbf{x} vektorokat, amelyek a (3.43) egyenletet kielégítik. Ha ezt γ -szorosára megnyújtjuk, olyan felületet kapunk, amelynek az egyenlete

$$(\mathbf{x} - \mathbf{a})^T \mathbf{B}_{\bar{\xi}}^{-1} (\mathbf{x} - \mathbf{a}) = \gamma^2, \quad (4.35)$$

ahol – mint korábban – az \mathbf{a} vektor $\bar{\xi}$ várható értéke. Bevezetjük az $\mathbf{x}' = \mathbf{x} - \mathbf{a}$ jelölést. Tegyük fel a következő kérdést: hogyan kell a γ kvantilist megválasztani, hogy a $\bar{\xi}$ vektor $(1 - \varepsilon)$ valószínűséggel essen a fentiekben definiált ellipszoid belsejébe? Az így definiált γ -hoz tartozó, a (4.35) egyenlettel definiált felületet nevezzük *konfidenciaellipszoidnak*.

A fenti feltételeknek eleget tevő γ értékét az alábbi egyenletből kaphatjuk meg:

$$P(\gamma) = \int_{\mathbf{x}'^T \mathbf{B}_{\bar{\xi}}^{-1} \mathbf{x}' < \gamma^2} \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{B}_{\bar{\xi}}}} \exp\left\{-\frac{1}{2} \mathbf{x}'^T \mathbf{B}_{\bar{\xi}}^{-1} \mathbf{x}'\right\} d\mathbf{x}' = 1 - \varepsilon.$$

A 3.4. alfejezet mintájára alkalmazzuk a

$$\mathbf{z} = \text{diag}\left(\frac{1}{\sigma}\right) \mathbf{U} \mathbf{x}' \quad (4.36)$$

transzformációt. Könnyű belátni, hogy ennek a Jacobi-determinánsa éppen $\sqrt{\det \mathbf{B}_{\bar{\xi}}}$, továbbá

$$\mathbf{x}'^T \mathbf{B}_{\bar{\xi}}^{-1} \mathbf{x}' = \mathbf{z}^T \mathbf{z} = \sum_{i=1}^n z_i^2 = \chi_n^2,$$

hiszen a (4.36) transzformáció a $(\bar{\xi} - \mathbf{a})$ vektort olyan $\bar{\zeta}$ vektorba viszi át, amelynek a komponensei egymástól függetlenek, várható értékük 0, és szóráruk 1. Fenti egyenletünk tehát a következőbe megy át:

$$P(\gamma) = \int_{\chi_n^2 < \gamma^2} \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \chi_n^2\right\} d\mathbf{z} = 1 - \varepsilon,$$

hiszen az itteni integrál éppen annak a valószínűségét adja meg, hogy $\chi_n^2 < \gamma^2$. Ennek megfelelően γ^2 az n szabadsági fokú χ^2 eloszlás kvantilise.