

*9. ASZIMPTOTIKUS TARTOMÁNY KERESÉSE

*9.1. A probléma felvetése

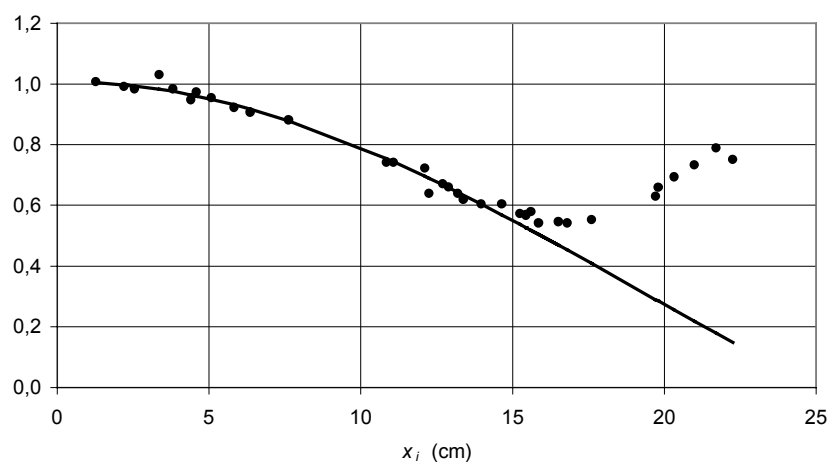
Gyakran választunk olyan illesztőfüggvényt, amelyről *tudjuk*, hogy az x_i változónak van olyan tartománya, ahol

$$M(\xi_i) \neq f(x_i, \mathbf{a}). \quad (9.1)$$

Ez emlékeztet a kiszóró pontok definíciójára használt (8.1) relációra, de már a 8. fejezet elején is hangsúlyoztuk, hogy az x_i változó említett tartományába eső mért értékeket nem tekintjük kiszóró pontnak, hiszen a (9.1) reláció fennállásának az oka az illesztőfüggvény tudatos egyszerűsítése. A definíciót egy példával világítjuk meg. A 9.1. ábrán látható pontok egy mérés eredményei, amelyről tudjuk, hogy x_i elegendően kis értékeire

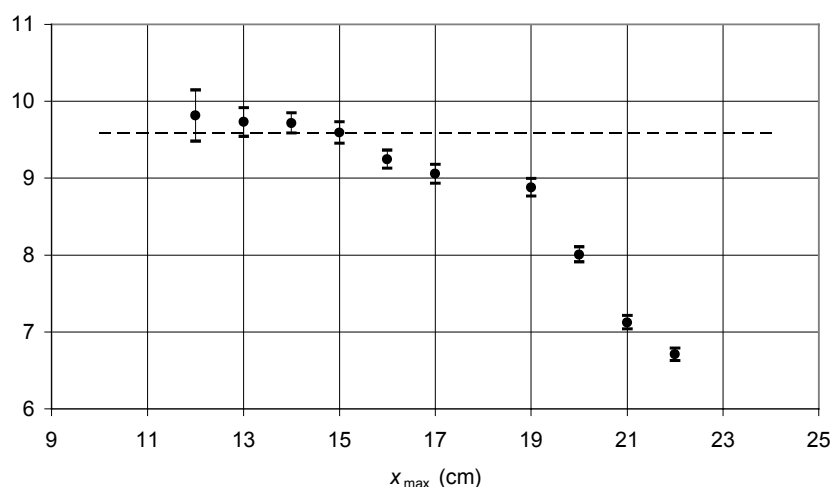
$$M(\xi_i) = a_1 J_0(a_2 x_i), \quad (9.2)$$

ahol J_0 a nulladrendű Bessel-függvény, amelyet folytonos görbével be is rajzoltunk az ábrára. Látható, hogy körülbelül 15 cm-nél nagyobb x_i értékekre ez a függvény nem írja le a méréseket, hiszen a mért pontok 16 cm után emelkedni kezdenek, viszont a Bessel-függvény szigorúan monoton csökkenő. Azt persze nem tudjuk, hol van az a határ, amelynél kisebb x_i értékekre (9.2) fennáll, a probléma éppen ennek a megkeresése. Tekintve, hogy (9.2) az elegendően kis x_i értékekre csak aszimptotikusan teljesül, a keresett tartományt *aszimptotikus tartománynak* nevezzük.



9.1. ábra. Példa a (9.1) relációra

Az aszimptotikus tartomány keresésére általánosan elterjedt módszer a *pontelhagyás módszere*, amely abban áll, hogy az illesztést többször megismételjük úgy, hogy fokozatosan elhagyjuk azokat a pontokat, ahol sejtésünk szerint fennáll a (9.1) reláció. Mindaddig, amíg az illesztésben van olyan pont, amelyre (9.1) érvényes, az illesztett paraméterekre torzított becslést kapunk, és a torzítás mértéke függ attól, hány ilyen pont van. Miután azonban az ilyen pontokat mind elhagytuk, a becslült paraméterek torzítatlanok lesznek, tehát várható értékük azonos lesz, legfeljebb szórásuk fog növekedni, hiszen a pontok elhagyása információvesztést jelent, ami a szórás növekedésében jut kifejezésre. A pontelhagyásos módszer szerint az aszimptotikus tartomány keresése így annak a határnak a keresését jelenti, amelytől kezdve a becslült paraméterek várható értéke azonos.



9.2. ábra. A aszimptotikus tartomány keresése a pontelhagyás módszerével

A módszert a 9.2. ábrán világítjuk meg. A grafikonon a (9.2) függvény a_2 paramétereit ábrázoltuk x_{\max} függvényében. Mindegyik illesztésben figyelembe vettük a 9.1. ábra mindazon pontjait, amelyekre $x_i \leq x_{\max}$. Az első illesztésben $x_{\max} = 22$ cm, majd fokozatosan csökkentettük 1 cm-rel. Mindegyik lépésben több pontot hagyunk el, mint az előzőben. Mint fentebb mondtuk, a szórások eközben monoton nőnek. Az a_2 paraméter becslült értékei pedig kezdetben nőnek, majd lassan stabilizálódnak. Az $x_{\max} = 22$ cm, 21 cm, 20 cm és 19 cm-nél végzett illesztések megítélése nem jelent problémát: ezek majdnem biztosan tartalmazznak még az aszimptotikus tartományon kívüli pontokat. Az $x_{\max} = 17$ cm-nél végzett illesztéstől kezdve azonban már nem lehetünk ilyen biztosak a dolgunkban. Az ábra szerint az $x_{\max} = 12$ cm, 13 cm és 14 cm-nél végzett illesztések eredményei között nyilvánvalóan nincs szignifikáns különbség. A többire vonatkozóan ugyanez már nem evidens, de az ellenkezője sem az. Például az $x_{\max} = 12$ cm-hez tartozó illesztett paraméter szórása már olyan nagy, hogy láthatóan nem tér el szignifikánsan az $x_{\max} = 16$ cm-hez tartozó illesztett paramétertől. Ezeket a megállapításokat csak sejtés alapján tehetjük, hiszen nyilvánvaló, hogy az egymás után kapott paraméterbecslések között erős korreláció van, tehát két érték közötti különbség szignifikáns voltát

nem olyan egyszerű eldönteni (lásd a “Mért mennyiségek egyenlősége” című részt az 5.2. alfejezetben).

A 9.2. ábrán mutatott helyzet az egyszerűbbek közé tartozik, mert a kapott paraméterek az első lépésekben határozottan emelkednek, majd egy bizonyos lépéstől kezdve szemmel is láthatóan stabilizálódnak. Nem szokott ez mindig így kijönni, vannak nehezebben áttekinthető esetek is. Mindenképpen beláthatjuk, hogy valamilyen statisztikai próbára van szükségünk ahhoz, hogy az aszimptotikus tartományt megnyugtató módon megtaláljuk, illetve kimutassuk, hogy nincs ilyen tartomány. A jelen fejezet célja egy, a gyakorlatban jól bevált próba ismertetése. A 9.2. ábrára berajzolt szaggatott vonal azt az értéket jelzi, amelyet ezzel a próbával kaptunk. Megjegyezzük, hogy – a kiszóró pontok kereséséhez hasonlóan – két próba egymást követő alkalmazására lesz szükség, de ennek most más oka van, mint a kiszóró pontok esetében.

*9.2. Definíciók és jelölések

Tudjuk, hogy az $f(x, \mathbf{a})$ függvény csak egy $x \in X_\infty$ tartományban (az ún. *aszimptotikus tartományban*) írja le jó közelítéssel a mért ξ_i -k várható értékét. A pontelhagyásos módszer értelmében választunk K tartományt

$$X_1 \supset X_2 \supset \dots \supset X_l \supset \dots \supset X_K, \quad (9.3a)$$

és reméljük, hogy van köztük olyan, amelyre $X_\infty \supset X_l$. (9.3a)-ból következik, hogy ez minden további tartományra is igaz. Jelöljük az X_l -re vonatkozóan becsült paramétervektort $\tilde{\mathbf{a}}_l$ -lel. További jelölések:

$$Q_l = \sum_{i \in I_l} w_i (\xi_i - f(x_i, \tilde{\mathbf{a}}_l))^2, \quad (9.4)$$

ahol I_l az X_l -tartománynak megfelelő i indexek halmaza:

$$I_1 \supset I_2 \supset \dots \supset I_l \supset \dots \supset I_K. \quad (9.3b)$$

σ^2 -nek az l -edik lépésben kapott becslése

$$s_l^2 = \frac{Q_l}{n_l - m}; \quad (9.5)$$

továbbá

n_l : az I_l -hez tartozó mérési pontok száma;

$\bar{\xi}_l$: a $\bar{\xi}$ vektornak az a vetülete, amelyben csak az I_l -hez tartozó i indexek szerepelnek;

$$\Delta \mathbf{a}_l = \tilde{\mathbf{a}}_l - \mathbf{a};$$

\mathbf{F}_l : az \mathbf{F} mátrixnak az az almátrixa, amelyben csak az I_l -hez tartozó i indexek szerepelnek;

\mathbf{W}_l : a \mathbf{W} mátrixnak az az almátrixa, amelyben csak az I_l -hez tartozó i indexek szerepelnek;

$$\mathbf{R}_l = \mathbf{F}_l^T \mathbf{W}_l \mathbf{F}_l. \quad (9.6)$$

Az ellenőrzendő hipotézis:

$$H_0: \text{létezik olyan } l, \text{ amelyre } M(\tilde{\mathbf{a}}_l) = \mathbf{a} \quad (9.7a)$$

és

$$M(Q_l) = \sigma^2(n_l - m). \quad (9.7b)$$

Az l -edik lépésben végrehajtott illesztés a következő eredményeket adja. Feltesszük, hogy H_0 igaz az l -edik lépésre. Ekkor

$$\tilde{y}_i = f(x_i, \tilde{\mathbf{a}}_l) = f(x_i, \mathbf{a}) + \sum_{k=1}^m F_{ik} \Delta a_{lk}, \quad i \in I_l,$$

amit vektoralakba írva a

$$\Delta \mathbf{y}_l = \tilde{\mathbf{y}}_l - M(\tilde{\boldsymbol{\xi}}_l) = \mathbf{F}_l \Delta \mathbf{a}_l$$

képlet adódik. (9.4)-ben a

$$\tilde{\boldsymbol{\xi}}_l - \tilde{\mathbf{y}}_l = \Delta \tilde{\boldsymbol{\xi}}_l - \mathbf{F}_l \Delta \mathbf{a}_l$$

vektor komponenseit emeljük négyzetre, vagyis

$$\begin{aligned} Q_l &= (\tilde{\boldsymbol{\xi}}_l - \tilde{\mathbf{y}}_l)^T \mathbf{W}_l (\tilde{\boldsymbol{\xi}}_l - \tilde{\mathbf{y}}_l) = \\ &= (\Delta \tilde{\boldsymbol{\xi}}_l - \mathbf{F}_l \Delta \mathbf{a}_l)^T \mathbf{W}_l (\Delta \tilde{\boldsymbol{\xi}}_l - \mathbf{F}_l \Delta \mathbf{a}_l). \end{aligned}$$

Közvetlen behelyettesítéssel ellenőrizhetjük a következő átalakítás helyességét:

$$\begin{aligned} Q_l &= \Delta \tilde{\boldsymbol{\xi}}_l^T (\mathbf{W}_l - \mathbf{W}_l \mathbf{F}_l \mathbf{R}_l^{-1} \mathbf{F}_l^T \mathbf{W}_l) \Delta \tilde{\boldsymbol{\xi}}_l + \\ &+ (\Delta \tilde{\boldsymbol{\xi}}_l^T \mathbf{W}_l \mathbf{F}_l \mathbf{R}_l^{-1} - \Delta \mathbf{a}_l^T) \mathbf{R}_l (\mathbf{R}_l^{-1} \mathbf{F}_l^T \mathbf{W}_l \Delta \tilde{\boldsymbol{\xi}}_l - \Delta \mathbf{a}_l). \end{aligned}$$

Mivel mindkét tag pozitív definit, ez akkor veszi fel a minimumát, amikor

$$\Delta \mathbf{a}_l = \mathbf{R}_l^{-1} \mathbf{F}_l^T \mathbf{W}_l \Delta \tilde{\boldsymbol{\xi}}_l, \quad (9.8)$$

és a minimum értéke

$$Q_l = \Delta \tilde{\boldsymbol{\xi}}_l^T (\mathbf{W}_l - \mathbf{W}_l \mathbf{F}_l \mathbf{R}_l^{-1} \mathbf{F}_l^T \mathbf{W}_l) \Delta \tilde{\boldsymbol{\xi}}_l.$$

Ezeket a képleteket korábban más úton már levezettük, lásd például a (6.12b) képletet. A 6.2. TÉTEL levezetését az I_l halmazra alkalmazva kapjuk, hogy

$$Q_l = \sigma^2 \chi_{n_l - m}^2. \quad (9.9)$$

*9.3. Kovariancia az l -edik és az l' -edik lépések között

Mivel a pontelhagyás módszerében az egymást követő lépésekben kapott paraméterbecslések azonosságát vagy eltéréseit vizsgáljuk, szükségünk van az egyes lépésekben becsült paraméterek közötti kovarianciára. Ennek kiszámításában a (9.8) képletből indulunk ki, amelynek alapján

$$\mathbf{M}(\Delta \mathbf{a}_l \Delta \mathbf{a}_{l'}^T) = \mathbf{R}_l^{-1} \mathbf{F}_l^T \mathbf{W}_l \mathbf{M}(\Delta \bar{\boldsymbol{\xi}}_l \Delta \bar{\boldsymbol{\xi}}_{l'}^T) \mathbf{W}_{l'} \mathbf{F}_{l'} \mathbf{R}_{l'}^{-1}.$$

A jobb oldalon szereplő várható érték egy diagonális mátrix almátrixa:

$$\left[\mathbf{M}(\Delta \bar{\boldsymbol{\xi}}_l \Delta \bar{\boldsymbol{\xi}}_{l'}^T) \right]_{ii'} = \delta_{ii'} \frac{\sigma^2}{w_i},$$

ahol

$$i \in I_l \quad \text{és} \quad i' \in I_{l'},$$

vagyis

$$\begin{aligned} & \left[\mathbf{F}_l^T \mathbf{W}_l \mathbf{M}(\Delta \bar{\boldsymbol{\xi}}_l \Delta \bar{\boldsymbol{\xi}}_{l'}^T) \mathbf{W}_{l'} \mathbf{F}_{l'} \right]_{kk'} = \\ & = \sigma^2 \sum_{i \in I_l} \sum_{i' \in I_{l'}} \delta_{ii'} F_{ik} F_{i'k'} \frac{w_{i'} w_i}{w_i} = \\ & = \sigma^2 \sum_{i' \in I_{l'}} w_{i'} F_{i'k} F_{i'k'} = \sigma^2 \left[\mathbf{R}_{l'}^{-1} \right]_{kk'}, \end{aligned}$$

ahol kihasználtuk, hogy $I_{l'} \subset I_l$. Ezzel

$$\mathbf{M}(\Delta \mathbf{a}_l \Delta \mathbf{a}_{l'}^T) = \sigma^2 \mathbf{R}_l^{-1} \mathbf{R}_{l'} \mathbf{R}_{l'}^{-1} = \sigma^2 \mathbf{R}_l^{-1}.$$

Ezt az első látásra meghökkentő eredményt tétel formájában is kimondjuk:

9.1 TÉTEL. Ha $I_{l'} \subset I_l$, akkor az ezeknek az intervallumoknak megfelelő illesztésekben becsült paraméterek kovarianciamátrixa

$$\mathbf{M}(\Delta \mathbf{a}_l \Delta \mathbf{a}_{l'}^T) = \sigma^2 \mathbf{R}_l^{-1}. \quad (9.10)$$

Ez nagyon erős korrelációt jelent. (9.10)-ből ugyanis az következik, hogy bármelyik illesztett a_k paraméternek ($k = 1, 2, \dots, m$) az l és l' indexű illesztésekben becsült értékei közötti kovariancia az I_l intervallumban kapott érték szórásnégyzete:

$$\text{cov}(\tilde{a}_{lk}, \tilde{a}_{l'k}) = \sigma^2 \left[\mathbf{R}_l^{-1} \right]_{kk} = D^2(\tilde{a}_{lk}). \quad (9.11)$$

Ebből egyszerűen igazolhatjuk azt az állítást, amelyet a 9.1. alfejezetben csak heurisztikusan sejtettünk:

9.2. TÉTEL. Ha $I_{l'} \subset I_l$, akkor az ezeknek az intervallumoknak megfelelő illesztésekben becsült paraméterek szórásai közül az előbbié a nagyobb:

$$D(\tilde{a}_{l'k}) \geq D(\tilde{a}_{lk}), \quad k = 1, 2, \dots, m. \quad (9.12)$$

Egyenlőség csak akkor állhat fenn, ha $I_{l'} = I_l$.

A bizonyításhoz a Schwarz-féle egyenlőtlenségből (3.5. TÉTEL) indulunk ki. Ha ezt a (9.11) képletben adott kovarianciára alkalmazzuk, akkor

$$D(\tilde{a}_{lk}) \cdot D(\tilde{a}_{l'k}) \geq |\text{cov}(\tilde{a}_{lk}, \tilde{a}_{l'k})| = D^2(\tilde{a}_{lk}),$$

amiből következik (9.12). A Schwarz-féle egyenlőtlenség bizonyításából következik, hogy itt csak akkor állhat egyenlőség, amikor $\tilde{a}_{lk} = \tilde{a}_{l'k}$, aminek a feltétele pedig $I_{l'} = I_l$.

A gyakorlatban bármelyik illesztett paraméter becsült értékeinek az azonoságát vizsgálhatjuk, és a tapasztalat szerint mindegyik esetben ugyanazt az aszimptotikus tartományt kapjuk eredményül. Ezért elég az egyik paramétert vizsgálni. Mindegyik illesztőfüggvény paraméterei között van egy olyan, amely fizikailag a legérdekesebb. Legyen ez az, amelyiknek a becsült értékeit vizsgáljuk. Az egyszerűség kedvéért ezt p_l -vel jelöljük. Ha ennek az indexe k , akkor

$$\text{cov}(p_l, p_{l'}) = \sigma^2 [\mathbf{R}_l^{-1}]_{kk} = \sigma^2 c_l^2,$$

ha $l \leq l'$ (vagyis $I_{l'} \subset I_l$).

A 9.1 TÉTEL szerint a p_1, p_2, \dots, p_K valószínűségi változók $\sigma^2 \mathbf{C}$ kovarianciamátrixát a

$$\mathbf{C} = \begin{bmatrix} c_1^2 & c_1^2 & c_1^2 & \dots & c_1^2 \\ c_1^2 & c_2^2 & c_2^2 & \dots & c_2^2 \\ c_1^2 & c_2^2 & c_3^2 & \dots & c_3^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ c_1^2 & c_2^2 & c_3^2 & \dots & c_K^2 \end{bmatrix}$$

alakban írhatjuk fel.

*9.4. p_0 becslése

A (9.7) alatti H_0 hipotézis helyett a következő, szigorúbb hipotézist fogjuk vizsgálni:

$$H_1: M(p_l) = p_0, \quad l = 1, 2, \dots, K. \quad (9.13)$$

Ha ez igaz, akkor a p_1, p_2, \dots, p_K valószínűségi változók együttes sűrűségfüggvénye:

$$f(p_1, p_2, \dots, p_K) = \frac{1}{(2\pi\sigma^2)^{K/2} \sqrt{\det \mathbf{C}}} \exp\left\{-\frac{S}{2\sigma^2}\right\},$$

ahol

$$S = \sum_{l=1}^K \sum_{l'=1}^K \omega_{ll'} (p_l - p_0)(p_{l'} - p_0), \quad \omega_{ll'} = [\mathbf{C}^{-1}]_{ll'}. \quad (9.14)$$

p_0 becslésére a maximális valószínűség módszerét használjuk: úgy választjuk meg, hogy a sűrűségfüggvény maximális, vagyis S minimális legyen.

9.3. TÉTEL. S akkor minimális, ha $\tilde{p}_0 = p_1$, és ekkor

$$S_1 = S_{\min} = \sum_{l=1}^{K-1} \frac{(p_{l+1} - p_l)^2}{c_{l+1}^2 - c_l^2}. \quad (9.15)$$

A H_1 ellenőrzésére szolgáló statisztikai próba részben ezen a tételel alapul.

A bizonyítás érdekében bevezetjük be az

$$u_l = p_l - p_0, \quad l = 1, 2, \dots, K$$

jelöléseket. Ha ezeket az \mathbf{u} vektor komponenseinek tekintjük, akkor

$$S = \mathbf{u}^T \mathbf{C}^{-1} \mathbf{u}.$$

Először kiszámítjuk a

$$\mathbf{z} = \mathbf{C}^{-1} \mathbf{u}$$

vektort, amely a

$$\mathbf{C} \mathbf{z} = \mathbf{u}$$

egyenlet megoldása. A \mathbf{C} mátrix fenti alakját figyelembe véve ennek egyes egyenletei a következők:

$k > 1$ esetében:

$$\sum_{l=1}^{k-1} c_l^2 z_l + c_k^2 \sum_{l=k}^K z_l = u_k. \quad (9.16a)$$

$k = 1$ esetében:

$$c_1^2 \sum_{l=1}^K z_l = u_1. \quad (9.16b)$$

Írjuk fel a (9.16a) egyenletet k helyett $(k-1)$ -re, és vonjuk ki a két egyenletet egymásból. Rövid számolás után azt kapjuk, hogy

$$\sum_{l=k}^K z_l = \frac{u_k - u_{k-1}}{c_k^2 - c_{k-1}^2}, \quad k > 1 \quad (9.16c)$$

és (9.16b) alapján pedig

$$\sum_{l=1}^K z_l = \frac{u_1}{c_1^2}.$$

Ennek alapján

$$\begin{aligned} S = \mathbf{u}^T \mathbf{z} &= \sum_{k=1}^K u_k z_k = u_K z_K + \sum_{k=1}^{K-1} u_k \left\{ \sum_{l=k}^K z_l - \sum_{l=k+1}^K z_l \right\} = \\ &= u_K z_K + u_1 \sum_{l=1}^K z_l + \sum_{k=2}^{K-1} u_k \sum_{l=k}^K z_l - \sum_{k=1}^{K-1} u_k \sum_{l=k+1}^K z_l = \\ &= u_1 \sum_{l=1}^K z_l + \sum_{k=2}^{K-1} u_k \sum_{l=k}^K z_l - \sum_{k=1}^{K-1} u_k \sum_{l=k+1}^K z_l = \\ &= u_1 \sum_{l=1}^K z_l + \sum_{k=1}^{K-1} u_{k+1} \sum_{l=k+1}^K z_l - \sum_{k=1}^{K-1} u_k \sum_{l=k+1}^K z_l = \\ &= u_1 \sum_{l=1}^K z_l + \sum_{k=1}^{K-1} (u_{k+1} - u_k) \sum_{l=k+1}^K z_l = \frac{u_1^2}{c_1^2} + \sum_{k=1}^{K-1} \frac{(u_{k+1} - u_k)^2}{c_{k+1}^2 - c_k^2} = \\ &= \frac{(p_1 - p_0)^2}{c_1^2} + \sum_{l=1}^{K-1} \frac{(p_{l+1} - p_l)^2}{c_{l+1}^2 - c_l^2}. \end{aligned}$$

A második tagban álló összeg független p_0 -tól. Az első tag mindig pozitív kivéve, ha

$$\tilde{p}_0 = p_1.$$

Ennélfogva p_0 becslése p_1 a maximális valószínűség elve szerint. Ezért használtuk a \sim jelet a p felett. Ezzel a tételt igazoltuk.

A H_1 hipotézis ellenőrzésére szolgáló próbához még egy tételre van szükségünk:

9.4. TÉTEL. Ha igaz a H_1 hipotézis, akkor a $\tilde{p}_0 = p_1$ becslés torzítatlan, továbbá

$$S_1 = \sigma^2 \chi_{K-1}^2. \quad (9.17)$$

A tételnek a becslés torzítatlanságára vonatkozó része triviális:

$$M(\tilde{p}_0) = M(p_1) = p_0.$$

A χ^2 -eloszlásra vonatkozó állítást három lépésben látjuk be.

a) A (9.15) alatti összegben minden tag 0 várható értékű Gauss-eloszlású valószínűségi változó négyzete, ugyanis,

$$M(p_{l+1} - p_l) = M(p_{l+1}) - M(p_l) = p_0 - p_0 = 0.$$

b) Mindegyik tag várható értéke σ^2 , hiszen

$$\begin{aligned} M\{(p_{l+1} - p_l)^2\} &= M\{[(p_{l+1} - p_0) - (p_l - p_0)]^2\} = \\ &= D^2(p_{l+1}) + D^2(p_l) - 2 \operatorname{cov}(p_{l+1}, p_l) = \sigma^2(c_{l+1}^2 - c_l^2). \end{aligned}$$

c) $l < l'$ esetén a különböző tagok kovarianciája 0, ami Gauss-eloszlás esetében függetlenséget jelent:

$$\begin{aligned} M\{(p_{l+1} - p_l)(p_{l'+1} - p_{l'})\} &= \operatorname{cov}(p_{l+1}, p_{l'+1}) - \operatorname{cov}(p_{l+1}, p_{l'}) - \\ &- \operatorname{cov}(p_l, p_{l'+1}) + \operatorname{cov}(p_l, p_{l'}) = \sigma^2(c_{l+1}^2 - c_{l'+1}^2 - c_l^2 + c_{l'}^2) = 0. \end{aligned}$$

Így tehát S_1 valóban $(L - 1)$ darab független, zérus várható értékű és σ szórású, Gauss-eloszlású valószínűségi változó összege. Ezzel a tételt beláttuk.

*9.5. χ^2 - vagy F-próba H_1 vizsgálatára

A H_1 hipotézis vizsgálatában két esetet különböztetünk meg: σ^2 ismert, illetve σ^2 nem ismert. Az előbbi esetben egy χ^2 -próbát fogalmazunk meg, az utóbbiban pedig egy F-próbát.

χ^2 -próba, amikor σ^2 ismert

Ha σ^2 ismert, akkor a fentiek alapján egy χ^2 -próbát lehet csinálni H_1 vizsgálatára. Ebben az esetben a próba 9.4. TÉTELEN alapul. A 2. függelék táblázataiból kikeressük azt a γ kvantilist, amelyre

$$P\{\chi_n^2 \leq \gamma\} = 1 - \varepsilon,$$

és a H_1 hipotézist elfogadjuk, amikor

$$S_1 < \sigma^2 \gamma \tag{9.18}$$

ahol S_1 -et a (9.15) képlettel számítjuk ki.

F-próba, amikor σ^2 nem ismert

Ha σ^2 nem ismert (ami a gyakoribb eset), akkor σ^2 -et becsülni kell. Ekkor egy F-próbát lehet definiálni a következő tétel alapján:

9.5. TÉTEL. $Q_{l'}$ statisztikailag független p_l -től, ha $l \leq l'$.

Ehhez elég belátni, hogy

$$\left(\Delta\bar{\xi}_{l'} - \mathbf{F}_{l'}\Delta\mathbf{a}_{l'}\right) \quad \text{és} \quad \Delta\mathbf{a}_l$$

független, ha $l \leq l'$, hiszen $Q_{l'}$ az előbbi függvénye.

$$\begin{aligned} & \mathbf{M}\left[\left(\Delta\bar{\xi}_{l'} - \mathbf{F}_{l'}\Delta\mathbf{a}_{l'}\right)\Delta\mathbf{a}_l^T\right] = \\ & = \left(\mathbf{E}_{l'} - \mathbf{F}_{l'}\mathbf{R}_{l'}^{-1}\mathbf{F}_{l'}^T\mathbf{W}_{l'}\right)\mathbf{M}\left(\Delta\bar{\xi}_{l'}\Delta\bar{\xi}_l^T\right)\mathbf{W}_l\mathbf{F}_l\mathbf{R}_l^{-1}. \end{aligned}$$

A korábbiak szerint ($i' \in I_{l'}$ esetén)

$$\left[\mathbf{M}\left(\Delta\bar{\xi}_{l'}\Delta\bar{\xi}_l^T\right)\mathbf{W}_l\mathbf{F}_l\right]_{i'k} = \sigma^2 \sum_{i \in I_l} \delta_{ii'} \frac{w_i F_{ik}}{w_{i'}} = \sigma^2 F_{i'k} = \sigma^2 [\mathbf{F}_{l'}]_{i'k},$$

vagyis

$$\mathbf{M}\left(\Delta\bar{\xi}_{l'}\Delta\mathbf{a}_l^T\right) = \mathbf{M}\left(\Delta\bar{\xi}_{l'}\Delta\bar{\xi}_l^T\right)\mathbf{W}_l\mathbf{F}_l\mathbf{R}_l^{-1} = \sigma^2\mathbf{F}_{l'}\mathbf{R}_l^{-1},$$

amiből – ha (9.10)-et is figyelembe vesszük –

$$\mathbf{M}\left[\left(\Delta\bar{\xi}_{l'} - \mathbf{F}_{l'}\Delta\mathbf{a}_{l'}\right)\Delta\mathbf{a}_l^T\right] = \sigma^2\left(\mathbf{F}_{l'}\mathbf{R}_l^{-1} - \mathbf{F}_{l'}\mathbf{R}_l^{-1}\right) = 0.$$

Ezzel a tételt beláttuk.

A tétel alapján csak Q_K független S_1 -től. Így tehát

$$\frac{S_1}{Q_K/(n_K - m)} = \frac{\chi_{K-1}^2}{\chi_{n_K - m}^2/(n_K - m)} = (K-1)f_{K-1, n_K - m}.$$

A kvantilist a következő egyenlettel definiáljuk:

$$P\{f_{K-1, n_K - m} < \gamma_f\} = 1 - \varepsilon_2.$$

Ebből kapjuk az F-próbát: a H_1 hipotézist elfogadjuk, ha

$$\frac{S_1}{Q_K/(n_K - m)} < (K-1)\gamma_f. \quad (9.19)$$

*9.6. Próbák sorozata

Ha a (9.18) vagy (9.19) próba eredménye pozitív, akkor – az adott konfidenciaszinten – az I_1 tartomány aszimptotikus. Mi van, ha a próba eredménye negatív, hiszen ekkor az I_1 tartomány nem aszimptotikus? Ebben az esetben elhagyandók az $(I_1 - I_2)$ tartományban levő nem aszimptotikus pontok. Tehát a próbát meg kell ismételni az I_2 tartományra vonatkozóan. Ha a próba ezután is negatív, akkor az I_3 tartományra vonatkozóan végezzük el, és így tovább, amíg végül pozitív eredményre nem jutunk. Valójában tehát a statisztikai próbáknak

egy sorozatát kell végrehajtani. A (9.13) hipotézis helyett most a hipotézisek sorozatát kell definiálnunk:

$$H_k: \quad M(p_l) = p_0, \quad l = k, k+1, \dots, K, \quad (9.20)$$

($k = 1, 2, \dots, K-1$). A próbák sorozatának az alapja a fentiek általánosítása:

$$S_k = \sum_{l=k}^{K-1} \frac{(p_{l+1} - p_l)^2}{c_{l+1}^2 - c_l^2} = \sigma^2 \chi_{K-k}^2. \quad (9.21a)$$

Ezt használjuk, amikor σ^2 ismert. Ellenkező esetben az F-próbák sorozatát hajtjuk végre, amelyek a

$$\frac{S_k}{Q_K/(n_K - m)} = \frac{\chi_{K-k}^2}{\chi_{n_K - m}^2/(n_K - m)} = (K - k) f_{K-k, n_K - m} \quad (9.21b)$$

statisztikákon alapulnak.

A próbák sorozata egy sor pozitív és negatív választ produkál. Nem lebecsülendő probléma, hogyan válasszuk ki ezek alapján az aszimptotikus tartományt. A sok lehetséges stratégia közül kettőt emelünk ki. Kezdetben a próbák többnyire csupa “nem” eredményt adnak. A döntési probléma csak akkor válik komollyá, amikor valamelyik próba eredményeként végre kijön egy “igen”. Legyen ennek a próbának az indexe k_1 . Kétféleképpen gondolkodhatunk:

1. Mondhatjuk azt, hogy a k_1 -edik lépés már az aszimptotikus tartomány indexe, tehát, ha valamelyik $k > k_1$ -re “nem” jön ki, akkor ez az elsőfajú hiba következménye, és emiatt figyelmen kívül hagyjuk, vagyis az I_{k_1} intervallumot tekintjük az aszimptotikus tartománynak. Ezt a gondolkodást nevezhetjük “az első igen stratégiájá”-nak. Ez megengedő stratégia, mert nem nagyon törődik a másodfajú hibával.
2. Gondolkozhatunk azonban kevésbé megengedő módon is. Mondhatjuk, hogy a k_1 -edik lépésben kijött “igen” lehetett a másodfajú hiba következménye, amit csak alátámaszt, ha valamelyik $k > k_1$ -re “nem” jön ki. Ezért csak olyan próbának “hiszünk”, amely után csupa “igen” következik. Ezt a gondolkodást nevezhetjük “az utolsó nem stratégiájá”-nak.

A két stratégia között csak a másodfajú hiba analízise alapján lehet dönteni (lásd alább).

A próbáknak ez a sorozata végeredményben megfelel a (9.7) képletekben definiált H_0 hipotézis ellenőrzésének.

*9.7. ϕ -próba

Az F-próba előtt meg kell győződni arról, hogy helyesen becsüljük a paraméterek szórását, pontosabban a c_l^2 mennyiségeket. Ha ugyanis a paraméterek becsült értéke torzított, akkor ezek becslése is torzított lesz. Gyakran még az

sem teljesül, hogy ezek monoton növekvő sorozatot alkossanak. Egy külön próba, a φ -próba szolgál arra, hogy megtaláljuk azt a k_2 indexet, amelynek megfelelő lépésben becsült paraméterek biztosítják, hogy a c_l^2 mennyiségek számítása közelítőleg torzítatlan becslést adjon.

Bizonyítható az alábbi két tétel:

9.6. TÉTEL. Fennáll a következő egyenlőség:

$$\text{cov}\left(\frac{Q_l}{n_l - m}, \frac{Q_{l'}}{n_{l'} - m}\right) = \frac{2\sigma^4}{n_l - m}, \quad l \leq l', \quad (9.22)$$

továbbá

9.7. TÉTEL. A

$$\frac{Q_l}{n_l - m} \quad \text{és} \quad \frac{Q_{l'+1}}{n_{l'+1} - m} - \frac{Q_{l'}}{n_{l'} - m}$$

valószínűségi változók korrelálatlanok, ha $l < l'$.

E két tétel bizonyítása hosszadalmas, ezért elhagyjuk. Megtalálható [3]-ban (1991). Ha a χ^2 -eloszlást Gauss-eloszlással közelítjük belátható [3] a

9.8. TÉTEL. $k = 1, 2, \dots, (K - 1)$ -re

$$T_k = \sum_{l=k}^{K-1} \frac{\left[\frac{Q_{l+1}}{n_{l+1} - m} - \frac{Q_l}{n_l - m} \right]^2}{\frac{2}{n_{l+1} - m} - \frac{2}{n_l - m}} = \sigma^4 \chi_{K-k}^2, \quad (9.23)$$

továbbá a

$$\varphi_{K-k, n_k - m} = \frac{\sqrt{T_k / (K - k)}}{Q_k / (n_k - m)} = \frac{\chi_k / \sqrt{K - k}}{\chi_{n_k - m}^2 / (n_k - m)} \quad (9.24)$$

hányados számlálója és nevezője statisztikailag független egymástól.

Ennek valószínűségi változónak a statisztikai jellemzői nem ismertek az irodalomban. Vizsgálata [3]-ban, kvantilisei pedig a 2. függelékben található meg. Ezekből ki lehet keresni az alábbi kvantilist:

$$P\{\varphi_{K-k, n_k - m} < \gamma_\varphi\} = 1 - \varepsilon_1. \quad (9.25)$$

Ezen alapul a φ -próba, amely az alábbi hipotézis ellenőrzésére szolgál:

$$H'_k: \quad M\left(\frac{Q_l}{n_l - m}\right) = \sigma^2, \quad l = k, k+1, \dots, K. \quad (9.26)$$

A φ -próba szerepét a fejezet elején mutatott példa révén mutatjuk be. Először a φ -próbát hajtjuk végre. Az eredmények a 9.1. táblázatban láthatók. A kvantiliseket a 2. függelék $\varepsilon = 0,05$ -höz tartozó táblázataiból vettük. Az első “igen” az 5. lépésben adódik. Mivel ennek a próbának csak kiegészítő szerepe van, itt elegendő “az első igen stratégiáját” alkalmazni, tehát az I_5 intervallumot alkalmasnak tartjuk arra, hogy a becült paraméterek szórását az ebben a lépésben kapott paraméterértékek mellett számoljuk ki.

9.1. táblázat. A φ -próba eredményei a 9.1. ábrán látható példára

k	$Q_k/(n_k - m)$	$n_k - m$	φ	Kvantilis	Próba
1	60,21	32	2,825	1,735	nem
2	49,67	31	3,225	1,761	nem
3	27,58	29	4,238	1,804	nem
4	5,672	27	2,931	1,854	nem
5	3,663	26	18,97	1,898	igen
6	2,625	24	1,632	1,970	igen
7	1,815	19	2,097	2,142	igen
8	1,402	18	1,903	2,257	igen
9	1,624	15	2,132	2,532	igen
10	0,8617	11			

A becült paraméterek szórását újraszámoljuk az I_5 intervallumra kapott paraméterértékek mellett. Így ezekre érvényes lesz a 9.2. TÉTEL, vagyis ezekkel képezhetjük S_k -t, végrehajthatjuk a (9.21b) szerinti F-próbák sorozatát $k = 1, 2, \dots, (K - 1)$ -re. Eredményük a 9.2. táblázatban látható. A kvantiliseket a 2. függelék $\varepsilon = 0,05$ -höz tartozó táblázataiból vettük ki. Az “utolsó nem stratégiája” alapján a $k = 7$ lépéshez tartozó I_7 intervallumot minősíthetjük aszimptotikusnak – 95% konfidenciaszinten. A 9.2. táblázattal kapcsolatban megjegyezzük, hogy harmadik oszlopban levő szórások nem növekszenek monoton módon, ami ellentmondani látszik a 9.2. TÉTELnek. A dolognak az a magyarázata, hogy ezek a szórások a tétel szerint valóban növekvő c_k^2 -ek és a 9.1. táblázatban levő $Q_k/(n_k - m)$ tényezők szorzatából vont négyzetgyökök. Ezek a monotonitáshoz képest mutathatnak kis eltéréseket.

9.2. táblázat. Az F-próba eredményei a 9.1. ábrán látható példára

k	\bar{a}_2	Szórás	$(K - k)f$	Kvantilis	Próba
1	6,708	0,082	3693	26,07	nem
2	7,125	0,086	2577	23,58	nem
3	8,006	0,098	1050	21,09	nem
4	8,879	0,115	137,8	18,57	nem
5	9,055	0,123	71,55	16,02	nem
6	9,245	0,118	36,38	13,43	nem
7	9,593	0,138	9,202	10,76	igen
8	9,717	0,132	0,194	7,965	igen
9	9,730	0,185	0,174	4,844	igen
10	9,813	0,332			

*9.8. A másodfajú hiba

Az alábbiakban az ε konfidenciaszint megválasztásával foglalkozunk. Nyilván az az érdekünk, hogy ε -t minél kisebbre válasszuk, hiszen nagy ε -ra a végül elfogadott paraméterbecslés szórása túlságosan nagyra nő. Ennek azonban korlátot szab a másodfajú hiba. Mint általában, most is meg kell határozunk az alternatív hipotézist, amellyel szemben ezt vizsgáljuk. A 9.2. ábrával kapcsolatban mondottakból, továbbá a 9.2. táblázatból látszik, hogy a nullhipotézis elfogadása vagy elvetése csak azokban a lépésekben jelent problémát, amelyeknek megfelelő I_l intervallumok közel vannak az aszimptotikus tartományhoz. Ezért alternatív hipotézisként azt tesszük fel, hogy egy kivételével mindegyik lépés aszimptotikus:

$$M(p_l) = p_0, \quad l > k, \quad (9.27a)$$

$$M(p_k) = p'_0 \neq p_0. \quad (9.27b)$$

(9.21a) alapján ezzel

$$S_k = \sum_{l=k}^{K-1} \zeta_l^2 = S_{k+1} + \zeta_k^2, \quad (9.28a)$$

ahol

$$\zeta_l = \frac{p_{l+1} - p_l}{\sqrt{c_{l+1}^2 - c_l^2}}. \quad (9.28b)$$

Az alternatív hipotézis szerint

$$M(\zeta_l) = 0, \quad l > k, \quad (9.29a)$$

$$M(\zeta_k) = \frac{p_0 - p'_0}{\sqrt{c_{k+1}^2 - c_k^2}} = h \neq 0, \quad (9.29b)$$

továbbá

$$D^2(\zeta_l) = 1 \quad (9.29c)$$

minden l -re.

A vizsgálatban – az egyszerűség kedvéért – feltételezzük, hogy σ^2 -et ismerjük, és értéke $\sigma^2 = 1$, vagyis a (9.18) próbát alkalmazzuk. A másodfajú hiba valószínűsége ekkor

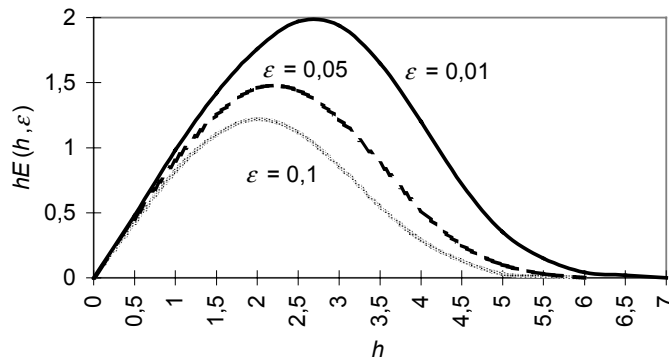
$$\begin{aligned} E(h, \varepsilon) &= P\{S_k < \gamma_k\} = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P\{S_{k+1} < \gamma_k - x^2\} \exp\left[-\frac{(x-h)^2}{2}\right] dx, \end{aligned}$$

ugyanis ilyen valószínűséggel fogadja el a próba a valójában nem igaz H_k hipotézist. Az integrál alatti valószínűséget a χ^2 -eloszlás alapján lehet számolni. $(1 - E)$ a próba ereje, annak a valószínűsége, hogy a H_k hipotézist a próba el-

veti. Tehát a próba E , illetve $(1 - E)$ valószínűséggel választ egy rendre p'_0 és p_0 várható értékű paramétert. Így a kiválasztott érték várható értéke

$$\begin{aligned} p'_0 E(h, \varepsilon) + p_0 [1 - E(h, \varepsilon)] &= p_0 + (p'_0 - p_0) E(h, \varepsilon) = \\ &= p_0 - h E(h, \varepsilon) \sqrt{c_{k+1}^2 - c_k^2}. \end{aligned}$$

Ennek p_0 -tól való eltérése tulajdonképpen tekinthető a próba *szisztematikus hibájának*. Két dolgot azonban figyelembe kell még venni. Először is a hE szorzat erősen függ h -tól, mint a 9.3. ábrán látható. Kis h -ra ugyan nagy az E valószínűség, de ez várható értékben kis hibát okoz. Nagy h esetén a hiba nagy, de E kicsi, tehát várható értékben ennek sem jelentős a hatása. Minden ε -ra van egy közbenső h -érték, ahol a hiba várható értéke maximális. Ha ezt a maximumot tekintjük szisztematikus hibának, akkor a tényleges hiba h bármilyen értéke mellett ennél nem lehet nagyobb. A 9.3. ábráról leolvasható, hogy ez a maximum ε -nak monoton csökkenő függvénye. Gyengén függ $(K - k)$ -től is [3].



9.3. ábra. A $hE(h, \varepsilon)$ hiba függése a (9.29b)-ben definiált h -tól

A másik dolog, amit figyelembe kell venni, az alkalmazott stratégia. Legyen P_{K-k} annak a valószínűsége, hogy a k -adik lépést ténylegesen elfogadjuk, feltéve, hogy a próba ezt választotta ki. Nyilvánvaló, hogy az “első igen stratégiája” mellett $P_{K-k} = 1$. Ugyanakkor az “utolsó nem stratégiája” mellett általában $P_{K-k} < 1$. Ennek az az oka, hogy e mellett a stratégia mellett akkor fogadjuk el a k -adik lépést, ha az $l > k$ lépésekre csupa “igen” jön ki. Ennek P_{K-k} valószínűsége pedig lehet kisebb 1-nél. Legegyszerűbb Monte Carlo módszerrel kiszámítani: a leírt döntési folyamatot számítógéppel szimulálva nagy pontossággal becsülhetjük a P_{K-k} valószínűséget. Végeredményben tehát a

$$\delta p_0 = P_{K-k} \sqrt{c_{k+1}^2 - c_k^2} \cdot \max_h [hE(h, \varepsilon)] \quad (9.30)$$

szorzatot tekintjük az alkalmazott statisztikai próba szisztematikus hibájának. Annál kisebb, minél nagyobbra választjuk ε -t.

Van azonban egy másik szempont is: a *szórás*. Minél nagyobb ε , annál nagyobb a végeredményben kiválasztott lépéshez tartozó szórás. Ez az elsőfajú hiba következménye. A jelen alfejezet befejezésekképpen ezt vizsgáljuk meg. Legyen $P\{I_l\}$ annak a valószínűsége, hogy a próbasorozat az I_l intervallumot nyilvánítja aszimptotikusnak – feltéve, hogy H_k igaz. Két eset lehetséges: a próbasorozat mindegyik I_l intervallumot elveti, vagy valamelyiket kiválasztja. Az előbbi esetben az *egész mérés* elvetendő. Ennek a valószínűsége általában kicsi, de ε nagy értékeire az “utolsó nem stratégiája” mellett előfordul néha. Például $\varepsilon = 0,1$ mellett körülbelül 5% ennek a valószínűsége. Ez már önmagában is arra utal, hogy ε -t nem célszerű nagyra választani. Feltéve, hogy ez a kedvezőtlen eset nem következik be, a végeredményben kiválasztott I_l intervallumhoz tartozó paraméterbecslés szórásnégyzetének átlagos értéke¹

$$\overline{D^2(\tilde{p}_0)} = \frac{\sum_{l=k}^K P\{I_l\} D^2(p_l)}{\sum_{l=k}^K P\{I_l\}}. \quad (9.31)$$

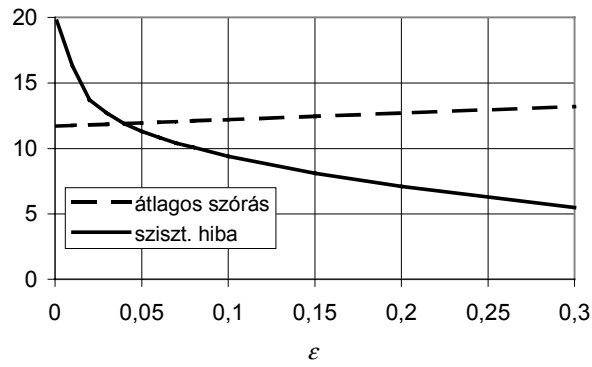
Ennek az átlagnak a négyzetgyökét nevezzük a továbbiakban *átlagos szórás*-nak. A két stratégiára a 9.4a. és 9.4b. ábrákon hasonlítjuk ezt össze a (9.30) alatt definiált szisztematikus hibával.² Az ábrákról a következő tanulságokat szűrhetjük le:

- Ahogy vártuk, a szisztematikus hiba ε -nal gyorsabban csökken az “utolsó nem stratégiájára”, mint a másokra.
- Az átlagos szórás *lassan* emelkedik ε -nal, tehát az elsőfajú hiba hatása nem növeli meg jelentősen a szórást.
- Mindkét stratégiára a szisztematikus hiba egy $\varepsilon < 0,05$ konfidenciaszintnél kisebbé válik, mint az átlagos szórás, és utána lassan csökken.

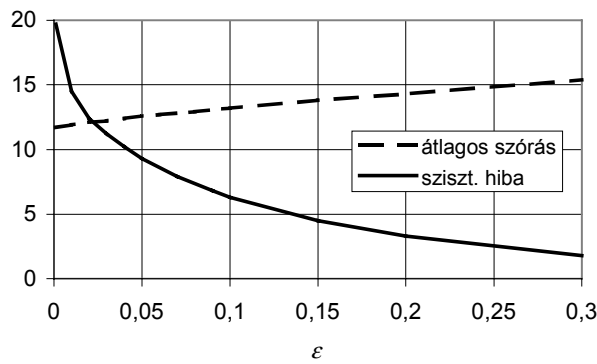
Ezekből következik, hogy ε értékét 0,05 és 0,1 között célszerű megválasztani, mert ez biztosítja, hogy a szisztematikus hiba kisebb lesz, mint az átlagos szórás.

¹ Az átlagot itt a következőképpen értjük. Az egész mérést (gondolatban) végtelen sokszor megismételjük azonos körülmények között, majd a fent leírt módon mindegyiket külön kiértékeljük. A próba által kiválasztott lépéshez tartozó szórásnégyzeteket átlagoljuk. Ennek az átlagnak a határértéke (6.31), amikor az ismétlések száma tart a végtelenhez.

² Az ábrákon mutatott számok egy koszinusz-illesztésnek felel meg.



9.4a. ábra. A szisztematikus hiba és az átlagos szórás összevetése.
Az "első igen stratégiája"; $K - k = 9$



9.4b. ábra. A szisztematikus hiba és az átlagos szórás összevetése.
Az "utolsó nem stratégiája"; $K - k = 9$

IRODALOM

- [1] Ajánlott kézikönyvek:
H. CRAMÉR, *Mathematical methods of statistics*, Princeton (1946)
Denkinger Géza, *Valószínűség-számítás*, Nemzeti Tankönyvkiadó (1978)
Sh. Dowdy & S. Warden, *Statistics for Research*, John Wiley & Sons (1982)
B. V. Gnyegyenko, A. J. Hincsin, *Bevezetés a valószínűség-számításba*, Művelt Nép Könyvkiadó, (1954)
JÁNOSSY LAJOS, *Theory and Practica of the Evaluation of Measurements*, Oxford University Press (1965) – létezik magyar és orosz nyelven is.
Pál Lénárd, *A valószínűség-számítás és a statisztikai alapjai*, Akadémiai Kiadó (1995)
JU. V. LINNYIK, *A legkisebb négyzetek módszere és a megfigyelések kiértékelésének matematikai-statisztikai alapjai*, Moszkva (1958) – oroszul és németül
Reimann József, *Valószínűség-elmélet és matematikai statisztika mérnököknek*, Tankönyvkiadó (1992)
D. J. Saville & G. R. Wood, *Statistical Methods: The Geometric Approach*, Springer Verlag New York, Inc. (1991)
B. L. Van der Waerden, *Mathematische Statistik (Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen, Band 87)*, Berlin–Göttingen–Heidelberg (1957)
Vetier András, *Szemléletes mérték- és valószínűség-elmélet*, Tankönyvkiadó (1991)
Vincze István, Varbanova Mária, *Nemparaméteres matematikai statisztika, elmélet és alkalmazások*, Akadémiai Kiadó, Budapest (1993)
Vincze István, *Matematikai statisztika ipari alkalmazásokkal*, Műszaki Könyvkiadó, Budapest (1975)
- [2] Kemény Sándor, Deák András, *Mérések tervezése és eredményeik kiértékelése*, Műszaki Könyvkiadó, Budapest (1993)
- [3] Z. Szatmáry, *Data Evaluation Methods in Reacytor Physics. Theory of Program RFIT*, report KFKI–1977–43 (1977)
Z. Szatmáry, *User’s Manual of Program RFIT*,
Part 1. General Description (report KFKI–1991–13/G)
Part 2. Parameter Estimation (report KFKI–1991–14/G)
Part 3. The Data Files (report KFKI–1991–15/G)
Part 4. Statistical Analysis (report KFKI–1991–16/G)
- [4] W. R. Thompson, *On a Criterion for Testing Outlying Observations and the Distribution of the Ratio of Deviation to Sample Standard Deviation*, *Annals of Math. Stat.* **6**, 214–219 (1935)
E. S. Pearson and C. ChandraSekar, *The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations*, *Biometrika*, **28**, 308–320 (1936)
- [5] Rózsa Pál, *Lineáris algebra és alkalmazásai*, Műszaki Könyvkiadó (1764)
- [6] F. E. Grubbs, *Sample Criteria for Testing Outlying Observations in Samples*, *Technometrics*, Vol. 11, 1–21 (1969)
J. R. Green and D. Margerison, *Statistical Treatment of Experimental Data*, Elsevier Scientific Publ. Comp. (1978)
- [7] R. von Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Íphysik*, Leipzig, Deuticke (1931)

- T. W. Anderson and D. A. Darling, Asymptotic Theory of Certain “Goodness of Fit” Criteria based on Stochastic Processes, *Annals of Math. Statistics* **23**, 193 (1952)
- K. Sarkadi, On Testing for Normality, Reports of the Research Institute for Mathematics V.A/3, Budapest (1960)
- [8] Pierre, Giacomo (International Bureau of Weights and Measures), *International Vocabulary of Basic and General Terms in Metrology*, (1993)
- [9] Fizika laboratóriumi gyakorlatok I.–V., egyetemi jegyzet, BME (1998)