10. előadás

Kiegyenlítő számítások MSc 2018/19

Áttekintés

- Paraméterbecslések

  Valószínűségeloszlások paramétereinek maximum likelihood becslése

  Maximum likelihood típusú becslések (M-becslések)

Paraméterbecslés

- Statisztikai minta eloszlásfüggvénye, ismeretlen paraméterek, sűrűségfüggvény
- Likelihood függvény
- Torzítatlan, hatékony, konzisztens becslés
- A maximális valószínűség (maximum likelihood) módszere
- M-becslések

Statisztikai becslés

- Azt a statisztikai eljárást, mely a minta ismeretében valamely mintajellemzőt állít elő, *statisztikai becslés*nek nevezzük
- Először tegyük fel, hogy ismerjük az adateloszlást jellemző $f(x)$ sűrűségfüggvény típusát

  maximum likelihood (ML) becslés
- Ha nem ismerjük az $f(x)$ sűrűségfüggvény típusát

  M-becslés

Statisztikai minta eloszlásfüggvénye

- Statisztikai *minta* mért értékek (valószínűségi változók) együttese: $\xi_1, \xi_2, \ldots, \xi_n$
- Minta *eloszlásfüggvényei*: $F_1(x), F_2(x), \ldots, F_n(x)$
- Minta *együttes* eloszlásfüggvénye

  független mintavétel: $$F(x_{1,} x_{2,} \ldots, x_n) = \prod_{i=1}^{n} F_i(x_i)$$

  azonos eloszlású mintaelemek: $$F(x_{1,} x_{2,} \ldots, x_n) = \prod_{i=1}^{n} F(x_i)$$

  ismeretlen *paraméterek* $(a_1, a_1, \ldots, a_m)$

$$F(\boldsymbol{x}, \boldsymbol{a}) = F(x_{1,} x_{2,} \ldots, x_n; a_{1,} a_{2,} \ldots, a_m)$$

Minta sűrűségfüggvénye, likelihood függvény

- Statisztikai minta *sűrűségfüggvénye* az $x_1, x_2, \ldots, x_n$ változók szerinti derivált

$$L(\boldsymbol{x}, \boldsymbol{a}) = L(x_{1,} x_{2,} \ldots, x_n; a_{1,} a_{2,} \ldots, a_m) = \frac{\partial^n F(\boldsymbol{x}, \boldsymbol{a})}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

- Ez a *likelihood függvény*
- logaritmusa a *loglikelihood függvény*

Paraméterbecslés kívánatos tulajdonságai

- Statisztikai minta bármely függvénye a *statisztika*

$$\tilde{a}_k = t_k(\xi_1, \xi_2, \ldots, \xi_n) = t_k(\boldsymbol{\xi})$$

- *Torzítatlan* becslés

$$M(\tilde{a}_k) = a_k \qquad k = 1,2, \ldots, m$$

- A becsült paraméter szórása legyen a lehető legkisebb

  Nem csökkenthető minden határon túl (*Cramér-Rao határ*)

  Érje el ezt az alsó határt, vagyis legyen *hatékony* (*efficiens*) becslés

Paraméterbecslés kívánatos tulajdonságai

- A becslés legyen *konzisztens*

  Egy becslési eljárás *konzisztens*, ha a mérések *n* számának növekedésével a paraméterek becsült értékei a valódi értékhez tartanak

  $$\lim_{n \to \infty} P\left\{ \left| \tilde{a}_k - a_k \right| > \varepsilon \right\} = 0$$

  minden k-ra és tetszőleges pozitív ε-ra

- Gyakori eset, hogy a paraméterek becsült értékeinek a szórása növekvő n-el 0-hoz tart:

  $$\lim_{n \to \infty} D\left(\tilde{a}_k\right) = 0, \qquad k = 1,2\,,\,\ldots\,,\,m$$

  Ekkor a becslés konzisztens

Maximum likelihood módszer

- Elv: Keressük azt az a paramétert, amelyre a minta $p(\boldsymbol{x}, a)$ *valószínűsége maximális* lesz

    ismernünk kell a minta eloszlását

- Fisher vezette be ezen a néven a módszert 1912-ben
- A maximum likelihood elv konzisztens becslésekhez vezet, mely elég nagy n-re közelítőleg Gauss eloszlású
- Példa

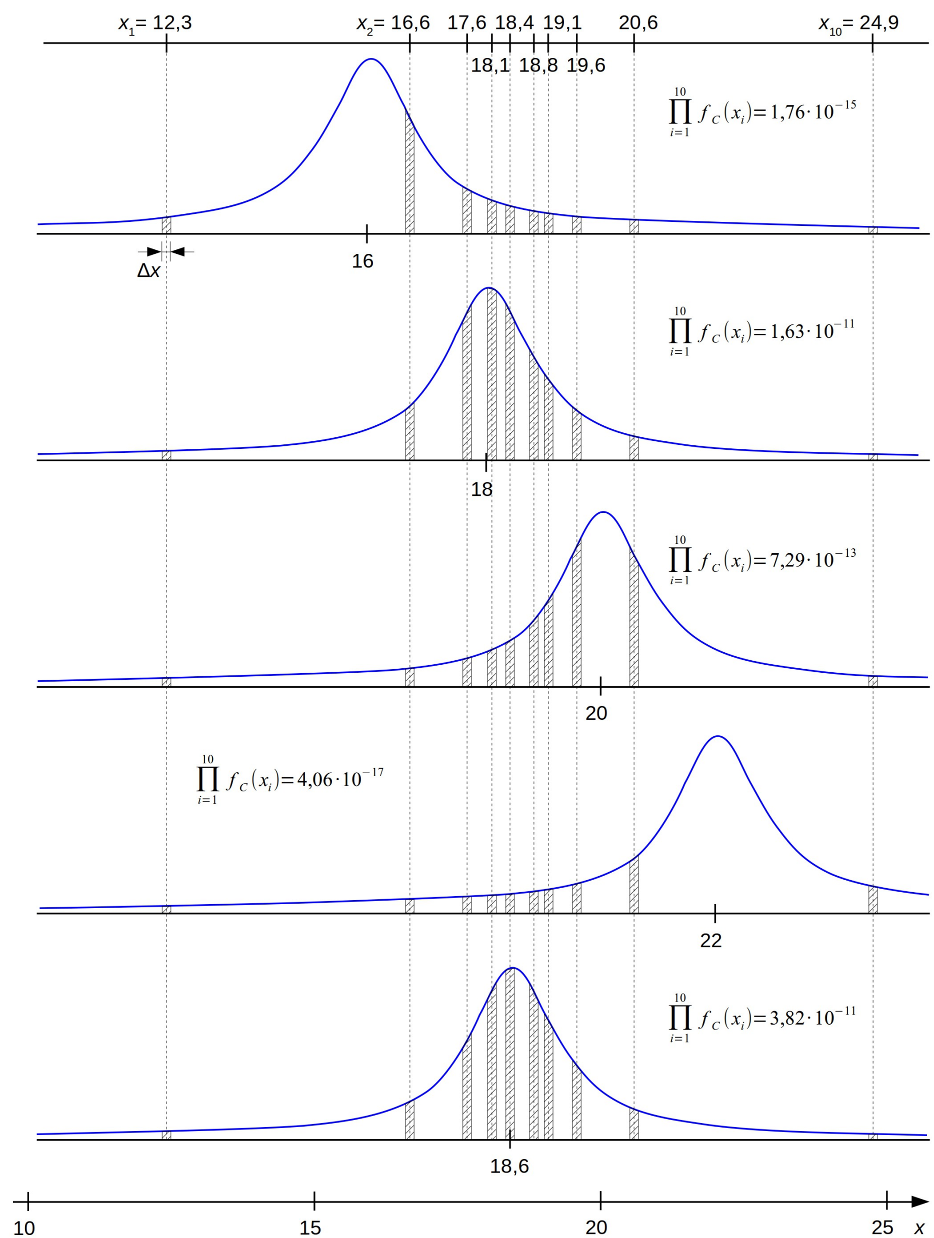
    Cauchy eloszlású mintánk van, egységnyi skála paraméterrel ($S = 1$)

    T helyparamétert becsüljük

Maximum likelihood módszer

- $f(x)$ típusát ismerjük: Cauchy eloszlás ($f_C(x)$)
- 10 elemű mintánk van [12.3, 16.6, 17.6, 18.1, 18.4, 18.8, 19.1, 19.6, 20.6, 24.9]
- A T helyparaméter maximum likelihood becslését keressük

$$\prod_{i=1}^{10} f_C(x_i, T) = \text{maximum}$$

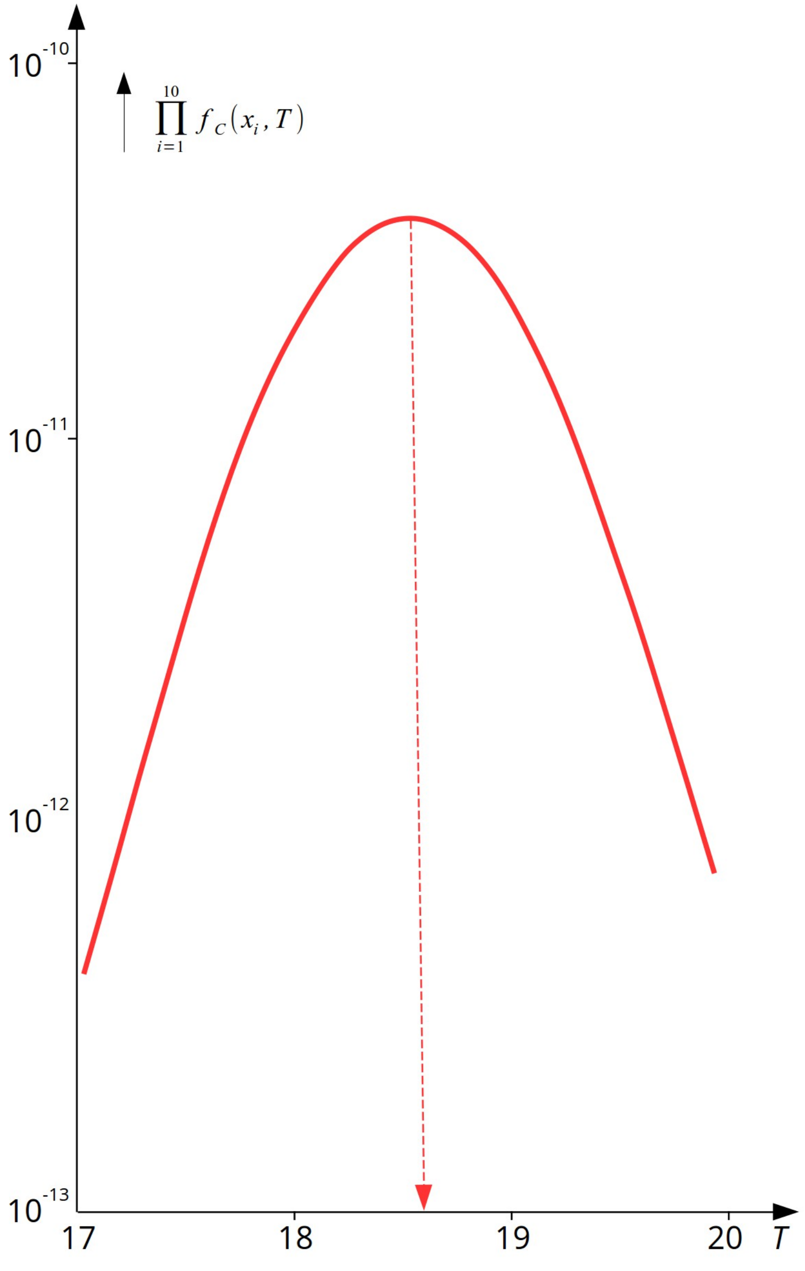


Maximum likelihood módszer

- szemilogaritmikus skálán

$$\prod_{i=1}^{10} f_C(x_i, T)$$

- A T helyparaméter maximum likelihood becslése: 18.6



ML becslés algoritmusa

- Milyen algoritmus alkalmazásával számíthatjuk ki a maximum likelihood (ML) elvnek eleget tevő T értéket?

attól függ, milyen az a priori ismert $f(x)$ analitikus alakja:

$$p(\boldsymbol{x}, T) = \prod_{i=1}^{n} f(x_i, T) \cdot (\Delta x)^n = \text{maximum}$$

ezzel ekvivalens alakok

$$\prod_{i=1}^{n} f(x_i, T) = \text{maximum} \qquad \sum_{i=1}^{n} \ln f(x_i, T) = \text{maximum}$$

ML becslés – Cauchy eloszlás

- Az ismert eloszlásfüggvény Cauchy-féle ($f_C(x)$)

$$f(x) = f_C(x) = \frac{1}{\pi} \frac{1}{1+(x-T)^2}$$

- a ML becslés $\sum_{i=1}^{n} \ln \frac{1}{1+(x-T)^2} = \text{maximum}$

- differenciálás után $\sum_{i=1}^{n} \frac{x_i - T}{1+(x_i-T)^2} = 0$

- megoldás T-re: $T = \dfrac{\sum_{i=1}^{n} \frac{x_i}{1+(x_i-T)^2}}{\sum_{i=1}^{n} \frac{1}{1+(x_i-T)^2}}$

ML becslés – Cauchy eloszlás

- Az egységnyi helyett ε szélesség paraméterű eloszlásra

$$T = \frac{\sum_{i=1}^{n} \frac{x_i}{\varepsilon^2 + (x_i - T)^2}}{\sum_{i=1}^{n} \frac{1}{\varepsilon^2 + (x_i - T)^2}} = M_n = \frac{\sum_{i=1}^{n} \frac{x_i}{\varepsilon^2 + (x_i - M_n)^2}}{\sum_{i=1}^{n} \frac{1}{\varepsilon^2 + (x_i - M_n)^2}}$$

A ML becslés algoritmusa azonos a *leggyakoribb érték* meghatározásának algoritmusával

ML becslés – Gauss eloszlás

- Az ismert eloszlásfüggvény Gauss-féle ($f_G(x)$)

$$f(x) = f_G(x) = \mathrm{const} \cdot e^{-\frac{(x-T)^2}{2}}$$

a ML becslés $$-\sum_{i=1}^{n} (x_i - T)^2 = \text{maximum}$$

differenciálás után $$\sum_{i=1}^{n} (x_i - T) = 0$$

megoldás T-re: $$T = \frac{1}{n} \sum_{i=1}^{n} x_i$$

ML becslés – Gauss eloszlás

- Az egységnyi szélesség paraméterű eloszlásra

$$T = \frac{1}{n} \sum_{i=1}^{n} x_i = E_n$$

  A ML becslés algoritmusa azonos a *számtani átlag* meghatározásának algoritmusával

- C. F. Gauss fordított gondolatmenete:

  Melyik az a sűrűségfüggvény, amelyhez optimális szimmetriapont meghatározási algoritmusként a számtani átlagképzés tartozik?
  A válasz: a *Gauss eloszlás* („normális” eloszlás) sűrűségfüggvénye

Gauss eloszlás S skála paraméterének ML becslése

- likelihood függvény

$$L=\prod_{i=1}^{n} f(x_i, S, T)=\prod_{i=1}^{n} \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2S^2}\cdot(x_i-T)^2}=\frac{1}{\left(S\sqrt{2\pi}\right)^n} e^{-\frac{1}{2S^2}\sum_{i=1}^{n}(x_i-T)^2}$$

- loglikelihood függvény

$$L^*=\ln L=\left(-n\ln S-\frac{n}{2}\ln 2\pi\right)-\left(\frac{1}{2S^2}\sum_{i=1}^{n}(x_i-T)^2\right)=\max$$

- S szerinti parciális derivált

$$\frac{\partial L^*}{\partial S}=-\frac{n}{S}+\frac{1}{S^3}\sum_{i=1}^{n}(x_i-T)^2=0$$

- S skála paraméter ML becslése az empirikus szórás

$$S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-T)^2}=\sigma_n$$

ML becslés és M-becslés

- Az ML becslés *minimum*feltételként megfogalmazva:

$$-\sum_{i=1}^{n} \ln f\left(x_i, T\right) = \text{minimum}$$

A gyakorlatban sokszor nem ismerjük az anyaeloszlás, $f(x)$ típusát

- *Huber* javaslata: helyettesítsük az ismeretlen negatív logaritmizált $f(x)$ eloszlást egy *ismert* $\rho(x)$ függvénnyel

$$\sum_{i=1}^{n} \rho\left(x_i, T\right) = \text{minimum}$$

ekkor is kapunk becslést a T helyparaméterre

ez az **M-becslés**

$\rho(\boldsymbol{x}, T)$ a *célfüggvény*

M-becslés hatásfüggvénye

- A célfüggvény minimalizálása helyett a differenciálással kapható

$$\sum_{i=1}^{n} \psi(x_i, T) = 0$$

megoldása célszerűbb

- Az M-becslést a célfüggvény helyett közvetlenül a $\psi(\boldsymbol{x}, T)$ függvénnyel, a *hatásfüggvénnyel* is megadhatjuk:

$$\psi(\boldsymbol{x}, T) = \frac{\partial \rho(\boldsymbol{x}, T)}{\partial T}$$

a T helyparaméter becslés esetén

$$\sum_{i=1}^{n} \rho(x_i - T) = \text{minimum} \qquad \sum_{i=1}^{n} \psi(x_i - T) = 0$$

M-becslés súlyfüggvénye

- a T helyparaméter becslés esetén, ha bevezetjük a

$$\varphi(x_i)=\frac{\psi(x_i-T)}{x_i-T}$$

*súlyfüggvény*t,

az M-becslést súlyozott átlag képzésére vezethetjük vissza:

$$\sum_{i=1}^{n}\varphi(x_i)\cdot(x_i-T)=0$$

a megoldása

$$T=\frac{\sum_{i=1}^{n}\varphi(x_i)\cdot x_i}{\sum_{i=1}^{n}\varphi(x_i)}$$

M-becslés skálaparaméterre

- az S skálaparaméter becslés esetén a ψ-vel analóg függvényt χ-vel szokás jelölni
- az M-becslést az alábbi egyenlet megoldása adja:

$$\sum_{i=1}^{n} \chi\left(\frac{x_i - T}{S}\right) = 0$$

feltételezve, hogy a T helyparaméter már ismert

M-becslés az ismertebb eloszlástípusokra

1. Gauss eloszlás

$$f(x) = e^{-\frac{x^2}{2}} \qquad \rho(x) = \frac{x^2}{2} \qquad \psi(x) = x \qquad \varphi(x) = 1$$

célfüggvény — hatásfüggvény — súlyfüggvény

az M-becslés az *átlag*gal egyezik meg

a hibák négyzetösszegét minimalizáljuk (ez a LKN vagy L2 norma szerinti becslés)

a hatásfüggvény *lineáris*, ezért

a megoldandó egyenletrendszer is *lineáris*

M-becslés az ismertebb eloszlástípusokra

2. Laplace eloszlás

$$f(x)=e^{-|x|} \qquad \rho(x)=|x| \qquad \psi(x)=\operatorname{sign} x \qquad \varphi(x)=\frac{1}{|x|}$$

célfüggvény — hatásfüggvény — súlyfüggvény

az M-becslés a *medián*nal egyezik meg

ahhoz, hogy az előjelek összege zérus legyen, az adatok felének előjele pozitív, a másik feléé negatív kell, hogy legyen – ez a tapasztalati medián (sign *y*: *y* előjele: +1/-1)

a hibák abszolút értékének összegét minimalizáljuk (L1 norma szerinti becslés, Boscovich, 1790, Laplace, 1799)

M-becslés az ismertebb eloszlástípusokra

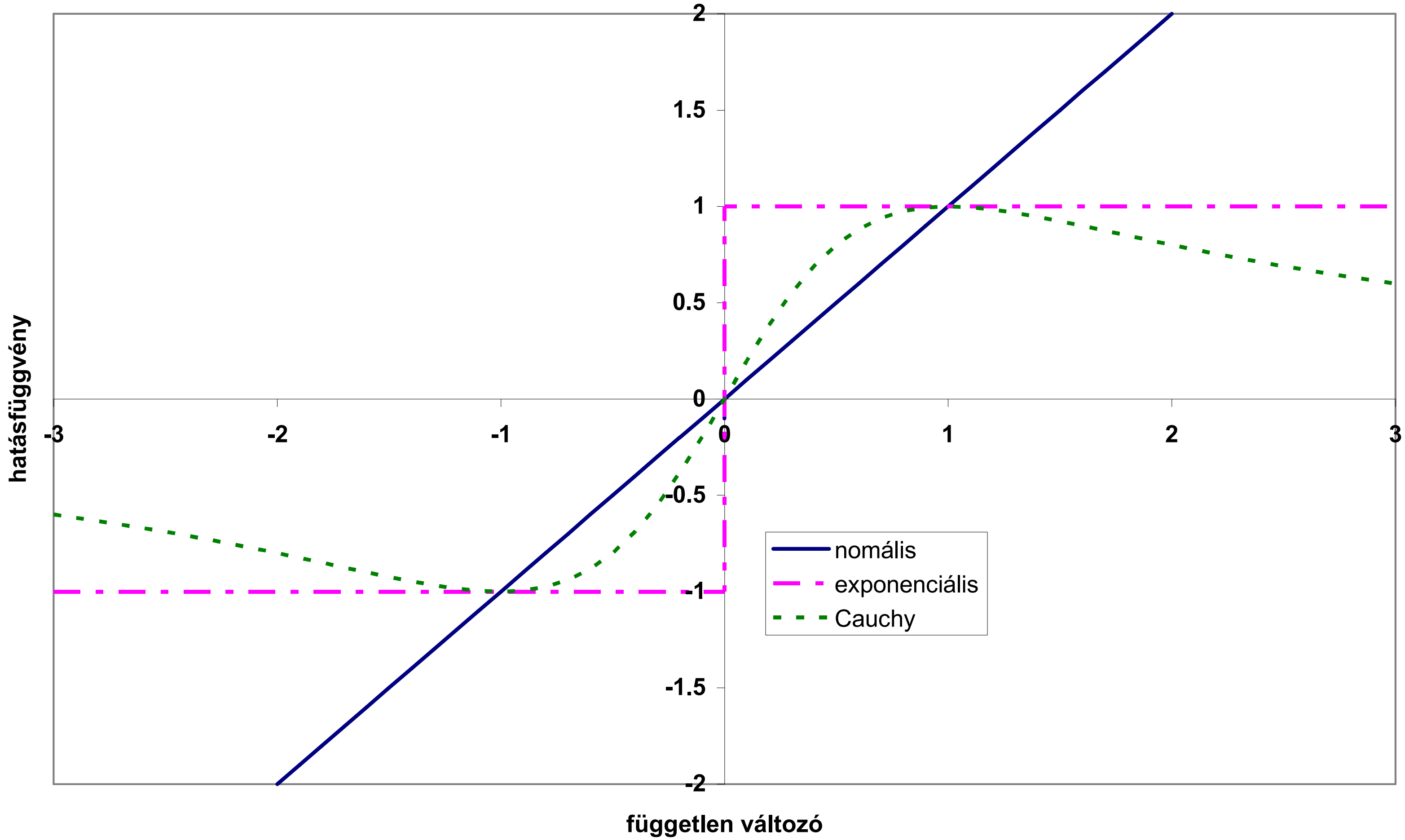
3. Cauchy eloszlás

$$f(x)=\frac{1}{1+x^2} \quad \rho(x)=\ln\left(1+x^2\right) \quad \psi(x)=\frac{2x}{1+x^2} \quad \varphi(x)=\frac{2}{1+x^2}$$

célfüggvény hatásfüggvény súlyfüggvény

az M-becslés a *leggyakoribb érték*kel egyezik meg (Steiner, 1990)
csak iterációval határozható meg

Az M-becslők hatásfüggvényei



- Ha a hatásfüggvény nem korlátos, a becslés nem rezisztens (vagyis érzékeny a kivágó értékekre):

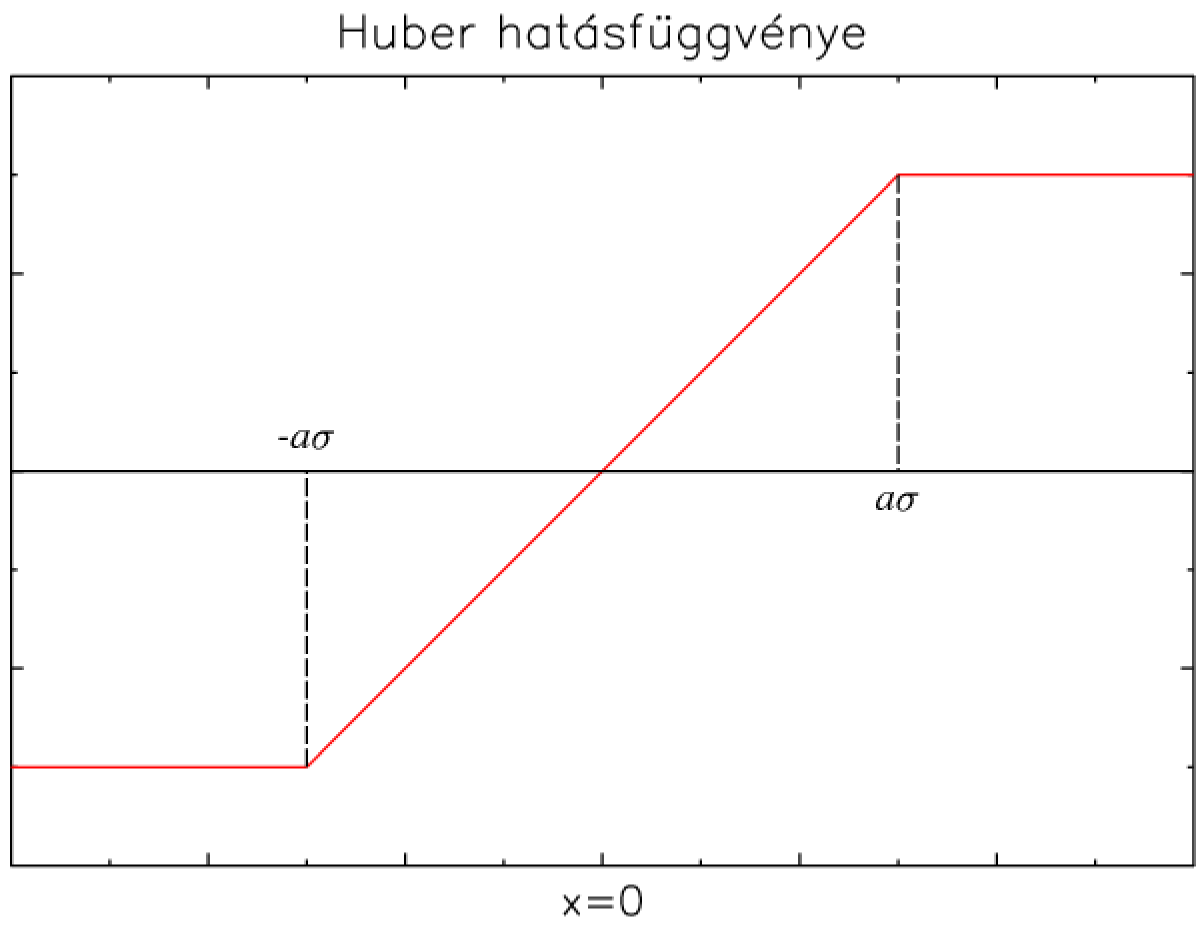
  A Gauss eloszláson alapuló M-becslés nem rezisztens

Huber hatásfüggvénye

- A tapasztalatok szerint a geodéziai mérési eredmények hibaeloszlása középen gyakran jó közelítéssel Gauss eloszlású, de a széleken bizonytalan
- Huber az alábbi hatásfüggvényt javasolta (például az $a = 1.5\sigma$ paraméterrel)

$$\Psi(x) = \begin{cases} x, & \text{ha } |x| \leq a \\ a \operatorname{sign} x & \text{egyébként} \end{cases}$$

A Huber hatásfüggvény alakja

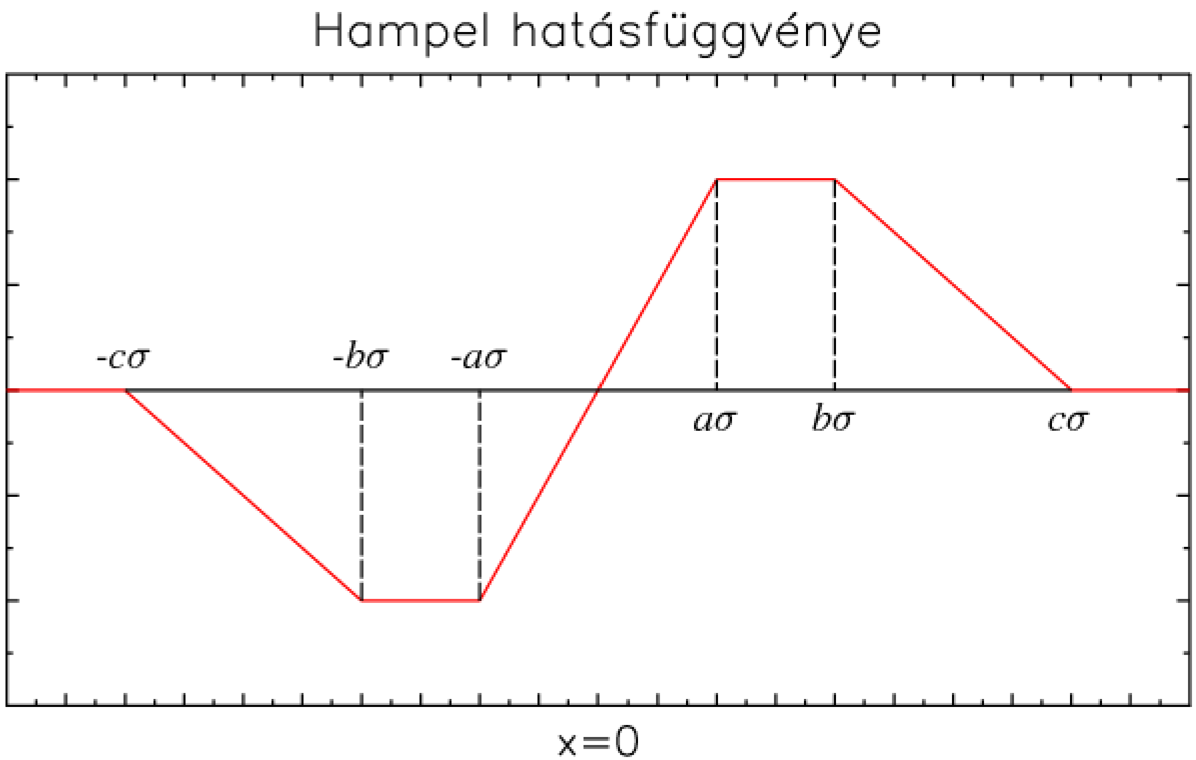


Hampel hatásfüggvénye

- A durva hibák (kivágó értékek) biztos kizárása érdekében Hampel a hatásfüggvényt a széleken csökkentette

$$\Psi(x)=\begin{cases} x, & \text{ha } |x|\le a \\ a \operatorname{sign} x & \text{ha } a<|x|\le b \\ \dfrac{a(c \operatorname{sign} x - x)}{c-b} & \text{ha } b<|x|\le c \\ 0 & \text{egyébként} \end{cases}$$

A Hampel hatásfüggvény alakja



$x > 0$ értékekre 4 részből áll a hatásfüggvény

pl. a dán geodéziai alaphálózat kiegyenlítésekor alkalmazták

Rezisztens kiegyenlítés M-becslőkkel

- *Hampel* hatásfüggvényét alkalmazzuk a közvetett mérések rezisztens kiegyenlítésére

  Huber hatásfüggvénye Hampel speciális esete, ha $b = c = \infty$

- r elemű $\boldsymbol{x}$ paramétervektor
- n elemű $\boldsymbol{v}$ hibavektor

$$\boldsymbol{v} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{l} \qquad \boldsymbol{A} = \left[a_{i,j}\right]_{i=1,n}^{j=1,r}$$

A célfüggvény minimalizálása

- az $\boldsymbol{x}$ paramétervektor legjobb becslésének feltételi egyenlete:

$$\sum_{i=1}^{n} \rho(v_i, \boldsymbol{x}) \Rightarrow \min$$

- az aktuális szélsőérték-feladatnál:

$$v_i = \sum_{j=1}^{r} a_{i,j} x_j - l_i \quad i = 1, \ldots, n$$

ezért

$$\frac{dv_i}{dx_k} = a_{i,k} \quad k = 1, \ldots, r$$

A szélsőérték megkeresése

- az $\boldsymbol{x}$ paramétervektor legjobb becslése a célfüggvény $\boldsymbol{x}$ szerinti deriválásával:

$$\frac{d}{dx_k}\sum_{i=1}^{n}\rho(v_i(\boldsymbol{x}))=\sum_{i=1}^{n}\frac{d\rho(v_i(\boldsymbol{x}))}{dv_i}\frac{dv_i}{dx_k}=\sum_{i=1}^{n}a_{i,k}\psi(v_i)=0 \quad k=1,\ldots,r$$

tömörített írásmódban

$$\boldsymbol{A}^T\psi(\boldsymbol{v})=0 \qquad \psi(\boldsymbol{v})=\begin{bmatrix}\psi(v_1)\\ \psi(v_2)\\ \vdots\end{bmatrix}$$

és $\psi(\boldsymbol{v})$ a Hampel-féle hatásfüggvény

A mérési javítások Gauss eloszlásúak

- $\psi(\boldsymbol{v}) = \boldsymbol{v}$ (a hatásfüggvény lineáris)
- az $\boldsymbol{x}$ paramétervektor „rezisztens” becslése megegyezik a LKN kiegyenlítéssel:

$$\boldsymbol{A}^T \boldsymbol{v} = \boldsymbol{A}^T \left(\mathbf{Ax} - \boldsymbol{l} \right) = 0$$

ebből

$$\boldsymbol{x} = \left(\boldsymbol{A}^T \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T \boldsymbol{l}$$

A Hampel-módszer javítási egyenletei

- az $\boldsymbol{A}$ mátrixot a v_i hibák nagysága alapján függőlegesen 4 almátrixra bonthatjuk

$\boldsymbol{A}_1$ $(|v_i| \leq a)$, $\boldsymbol{A}_2$ $(a < |v_i| \leq b)$,

$\boldsymbol{A}_3$ $(b < |v_i| \leq c)$ és $\boldsymbol{A}_4$ $(c < |v_i|)$

- particionáljuk ennek megfelelően a javítási egyenleteket:

$$\begin{bmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \\ \boldsymbol{v}^3 \\ \boldsymbol{v}^4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \boldsymbol{A}_3 \\ \boldsymbol{A}_4 \end{bmatrix} \cdot \boldsymbol{x} - \begin{bmatrix} \boldsymbol{l}_1 \\ \boldsymbol{l}_2 \\ \boldsymbol{l}_3 \\ \boldsymbol{l}_4 \end{bmatrix}$$

A Hampel-módszer feltételi egyenletei

- a feltételi egyenleteket is 4 almátrixra bonthatjuk

$$\begin{bmatrix} \boldsymbol{A}_1^T & \boldsymbol{A}_2^T & \boldsymbol{A}_3^T & \boldsymbol{A}_4^T \end{bmatrix} \cdot \begin{bmatrix} \psi\left(\boldsymbol{v}^1\right) \\ \psi\left(\boldsymbol{v}^2\right) \\ \psi\left(\boldsymbol{v}^3\right) \\ \psi\left(\boldsymbol{v}^4\right) \end{bmatrix} = 0$$

amelyből

$$\boldsymbol{A}_1^T \psi\left(\boldsymbol{v}^1\right) + \boldsymbol{A}_2^T \psi\left(\boldsymbol{v}^2\right) + \boldsymbol{A}_3^T \psi\left(\boldsymbol{v}^3\right) + \boldsymbol{A}_4^T \psi\left(\boldsymbol{v}^4\right) = 0$$

A megoldandó egyenletrendszer

- A $\psi(\boldsymbol{v})$ a Hampel-féle hatásfüggvény behelyettesítése után

$$\boldsymbol{A}_1^T \boldsymbol{v}^1 + \boldsymbol{A}_2^T a \operatorname{sign} \boldsymbol{v}^2 + \boldsymbol{A}_3^T \frac{a}{c-b}\left(c \operatorname{sign} \boldsymbol{v}^3 - \boldsymbol{v}^3\right) = 0$$

a javításokat beírva

$$\boldsymbol{A}_1^T\left(\boldsymbol{A}_1 x - \boldsymbol{l}_1\right) + \boldsymbol{A}_2^T a \operatorname{sign} \boldsymbol{v}^2 + \boldsymbol{A}_3^T \frac{a}{c-b}\left(c \operatorname{sign} \boldsymbol{v}^3 + \boldsymbol{l}_3 - \boldsymbol{A}_3 \boldsymbol{x}\right) = 0$$

A megoldandó egyenletrendszer

- Az előző egyenletet rendezve

$$\left[\boldsymbol{A}_1^T\boldsymbol{A}_1-\boldsymbol{A}_3^T\boldsymbol{A}_3\frac{a}{c-b}\right]\boldsymbol{x}=\boldsymbol{A}_1^T\boldsymbol{l}_1-\boldsymbol{A}_2^T a\,\mathrm{sign}\,\boldsymbol{v}^2+\boldsymbol{A}_3^T\frac{a}{c-b}\left(c\,\mathrm{sign}\,\boldsymbol{v}^3+\boldsymbol{l}_3\right)$$

- Végeredményben egy $\boldsymbol{Bx}=\boldsymbol{d}$ alakú egyenletrendszerhez jutunk, ahol

$$\boldsymbol{B}=\boldsymbol{A}^T\boldsymbol{PA} \qquad \boldsymbol{d}=\boldsymbol{A}^T\boldsymbol{Pl}-a\,\boldsymbol{A}^T\boldsymbol{Q}\,\mathrm{sign}\,\boldsymbol{v}$$

A súlymátrixok

- az előző egyenletekben $\boldsymbol{P}$ és $\boldsymbol{Q}$ átlós súlymátrixok, melyeknek főátlóelemei:

$$p_{i,j}=\begin{cases} 1 & \text{ha } |v_i| \le a \\ \dfrac{a}{b-c} & \text{ha } b < |v_i| \le c \\ 0 & \text{egyébként} \end{cases} \qquad q_{i,j}=\begin{cases} 1 & \text{ha } a < |v_i| \le b \\ \dfrac{b}{c-b} & \text{ha } b < |v_i| \le c \\ 0 & \text{egyébként} \end{cases}$$

- a megoldás *iterációban* történik
- a $\boldsymbol{v}$ ellentmondás-vektor alapján a kiegyenlítést mindig az aktuális *osztályba sorolás*nak megfelelően ismételjük meg

A dán módszer

- durvahibaszűrésre kifejlesztett eljárás *Krarup* (1967) elgondolása nyomán
- a mérések *iteratív újrasúlyozás*a a kiegyenlítésből kapott mérési javítások függvényében (a gyakorlatban $a = 3$)

$$p_{i+1} = \begin{cases} 1 & \text{ha } |v_i| < a \cdot \sigma \\ e^{-\frac{|v_i|}{a \cdot \sigma}} & \text{egyébként} \end{cases}$$

Példa rezisztens M-kiegyenlítésre: regressziós egyenes számítása

- LKN paraméter becslés:

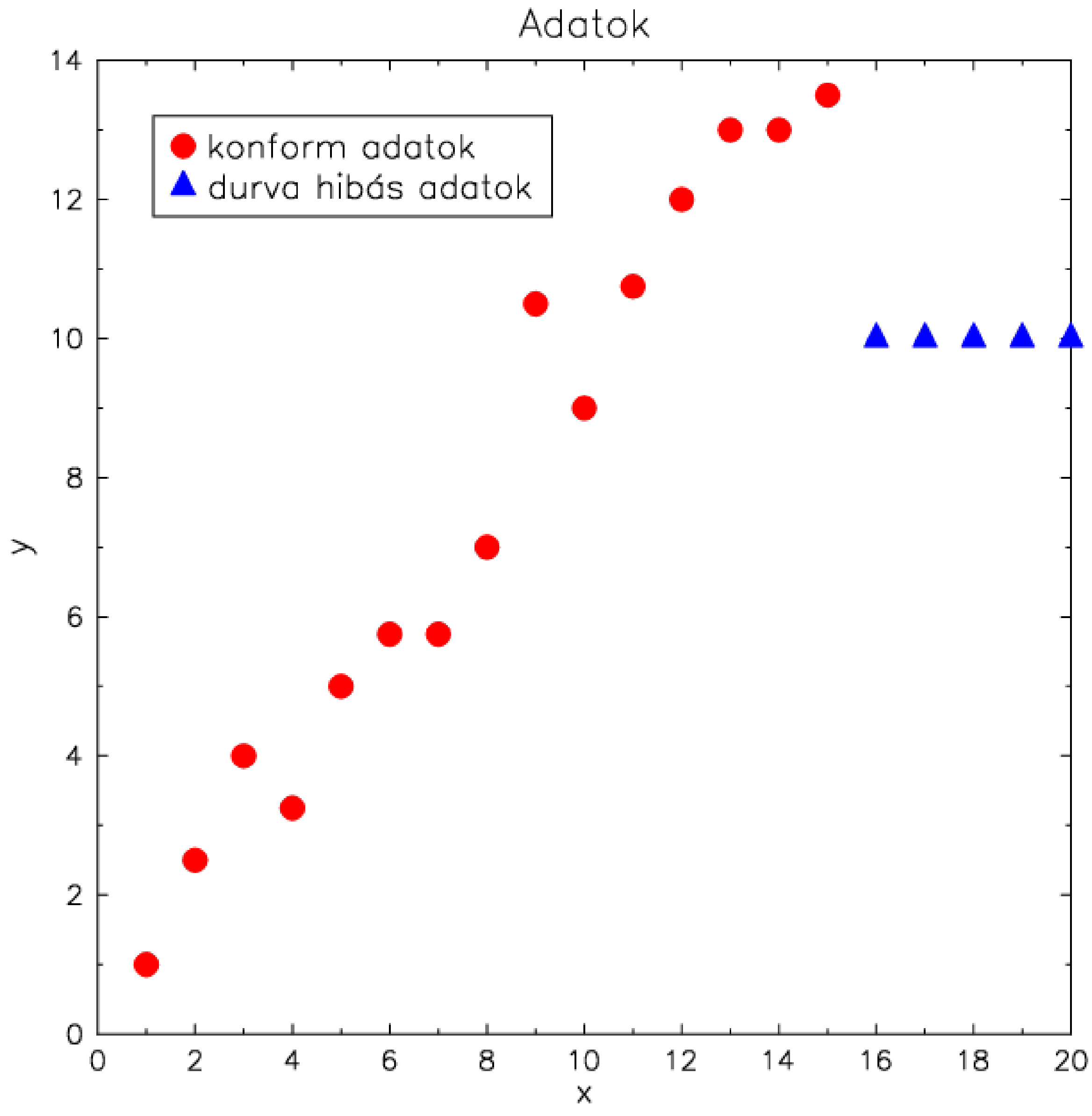
egyenes egyenlete: $y = ax + b$

paraméter vektor: $\boldsymbol{x} = \begin{bmatrix} a & b \end{bmatrix}^T$

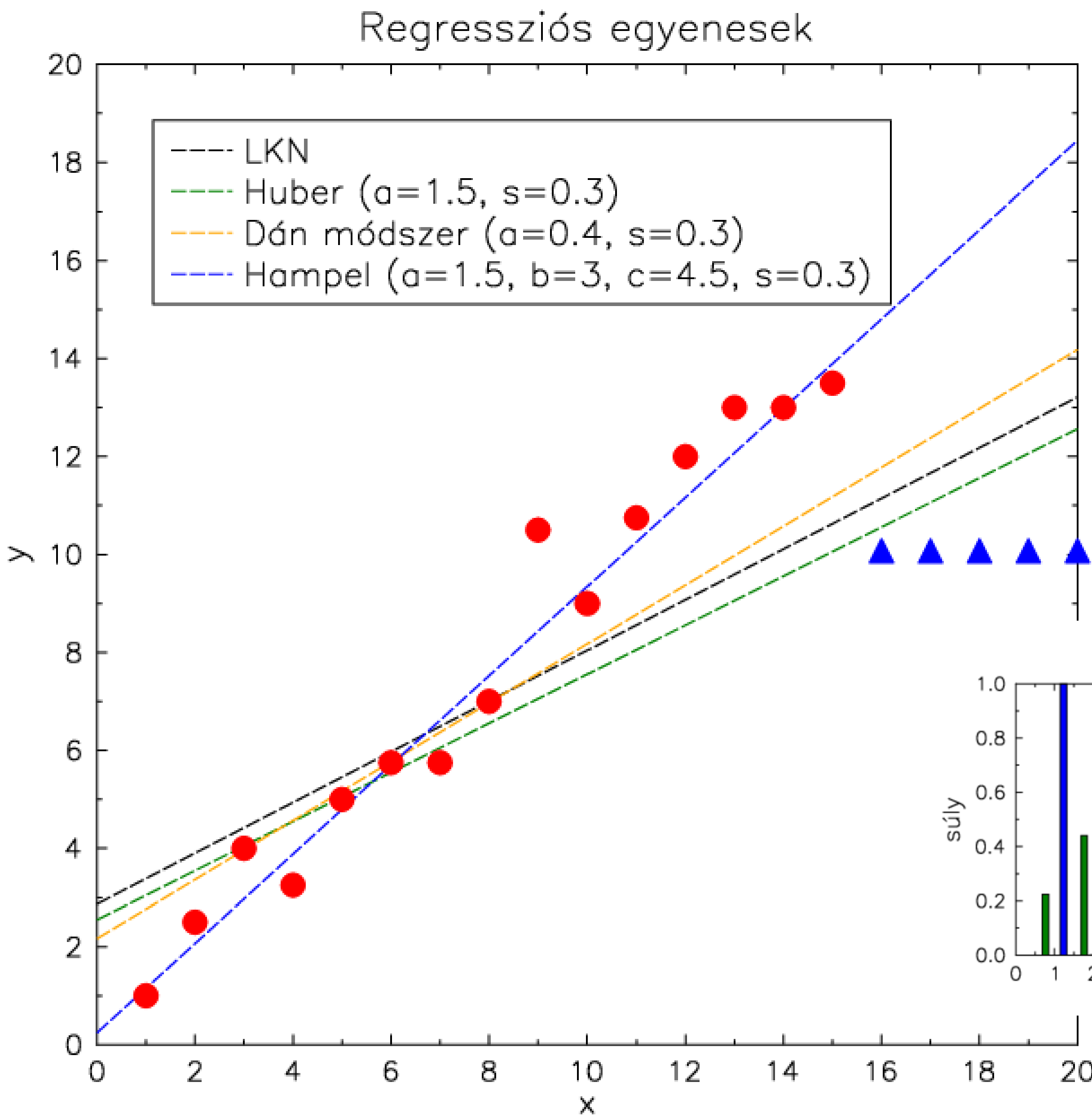
alakmátrix: $\boldsymbol{A} = \begin{bmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{bmatrix}^T$

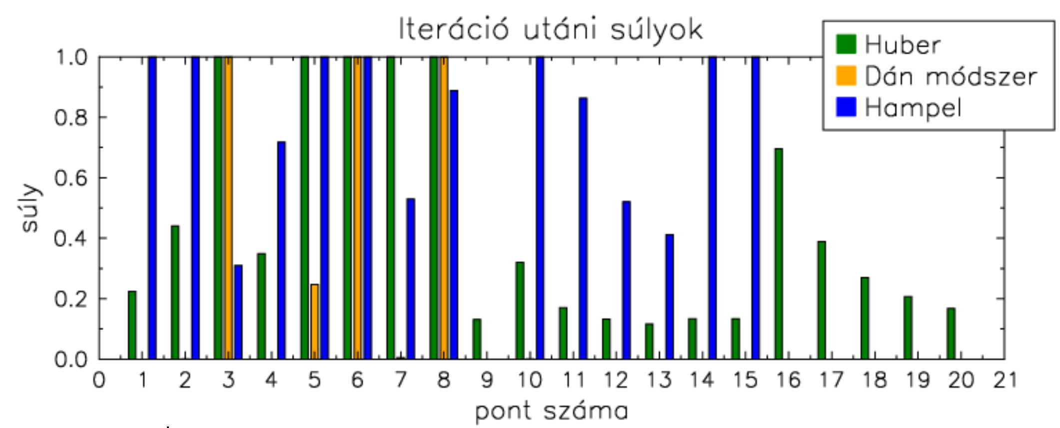
tisztatag vektor: $\boldsymbol{l} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^T$

Mérések (5/20 durva hibás)



Eredmények


Regressziós egyenesek
LKN
Huber (a=1.5, s=0.3)
Dán módszer (a=0.4, s=0.3)
Hampel (a=1.5, b=3, c=4.5, s=0.3)
y
x
0
2
4
6
8
10
12
14
16
18
20


Iteráció utáni súlyok
Huber
Dán módszer
Hampel
súly
pont száma
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

Tananyag, szakirodalom

- Steiner (1990): 5.2
- Vincze (1968): 4.2-4.4, 4.7.1
- Szabó Norbert Péter (2012): Bevezetés a geostatisztikába. Jegyzet. Miskolci Egyetem
- Huber P.J. (1981): Robust Statistics. Wiley & Sons 3.2 (M-Estimates)