Kiegyenlítő számítások MSc 2023/24

1. előadás

Áttekintés

- A tárggyal kapcsolatos információk
- Bevezetés
- Legjellemzőbb érték és mérési bizonytalanság
- Mérési eredmények eloszlásának vizsgálata

A tárggyal kapcsolatos információk

- Heti 3 óra (2ea + 1gy), jelenléti, 4 kredit
- TAD, ütemterv, anyagok: edu.epito.bme.hu
- Teljesítményértékelések (Moodle):

  1 ZH (9. hét, 10.31., 30 pont, nem kötelező, *nincs minimum*)

  2 HF ("kis házik": Kálmán szűrés, PSD becslés, 10-10 pont, *minimum*: 5-5 pont)

  vizsga (írásbeli, 50 pont, *minimum*: 25 pont)

  összesen: 100 pont, *minimum*: 50 pont

Tanulástámogató anyagok

- Detrekői Ákos: Kiegyenlítő számítások. Tankönyvkiadó, Budapest, 1991 (BME TKO, 22 példány)
- Steiner Ferenc: A geostatisztika alapjai. Tankönyvkiadó, Budapest, 1990 (BME KK 5 példány)
- Vincze István: Matematikai statisztika ipari alkalmazásokkal. Műszaki Könyvkiadó, 1975 (BME KK 6 példány)
- Szabó Norbert Péter: Bevezetés a geostatisztikába. Elektronikus jegyzet. Miskolci Egyetem, 2012 (online elérhető)
- edu.epito.bme.hu
- github.com/gyulat/kiegyenlito_szamitasok

Bevezetés

- “A kiegyenlítő számítások a geodéták által végzett **mérés**ek **matematika**i feldolgozásának alapvető **módszer**e” (Detrekői, 1991)
- **mérés**: „Műveletek összessége, amelyek célja egy mennyiség értékének a meghatározása” (Nemzetközi metrológiai értelmező szótár, 1998, OMH)

Mérések

- A geodéták által végzett mérések

egyszerű	**összetett**
távolság	helyzet (1D, 2D, 3D)
irány (szög)	tájékozás (1D, 2D)
sebesség (Doppler)	INS (gyorsulás, helyzet, elfordulás)
gyorsulás (g)	kép (fotó)
fényesség (pixel)	potenciálkülönbség

Mérések

- **Metrológia** (méréstan)

  „A mérésekkel kapcsolatos ismeretek teljes köre. A metrológia magában foglalja a méréseknek mind az elméleti, mind a gyakorlati szempontjait, függetlenül a pontossági szinttől és a tudományban vagy a műszaki életben való alkalmazás területétől” (http://mkeh.gov.hu/meresugy/metrologia)

Matematikai feldolgozás

- **Statisztika**
  „A valóság számszerű információinak megfigyelésére, összegzésére, elemzésére és modellezésére irányuló gyakorlati tevékenység és tudomány” (Wikipedia)
- **Matematika**
  „A matematika az egyetlen tökéletes módszere annak, hogy önmagunkat az orrunknál fogva vezessük” (Einstein)

Robusztus statisztika

- „A robusztus statisztika olyan elméleti keret, amelyen belül *gazdaságosan* oldható meg az a feladat, hogy egymástól jelentősen *eltérő eloszlástípusok* esetén is megbízható eredményt érjünk el, valamint hogy ne legyünk kitéve a *durva hibájú adatok* torzító (esetleg katasztrofális mértékben torzító) hatásának.” (Steiner, 1990)

Ismeretkörök

- Statisztika
- Metrológia
- Paraméterbecslés
- Idősorok elemzése

Statisztika

- Valószínűségeloszlás, modell-családok
- Adatrendszer legjellemzőbb értéke, adatrendszerben rejlő bizonytalanság
- Statisztikai próbák
- Cramer-Rao határ, becslés statisztikai hatásfoka
- Bevezetés a Bayes statisztikába
- Extrém érték eloszlások
- Monte Carlo eljárások
- Maximum likelihood, robusztus és rezisztens becslés

Metrológia

- Metrológiai alapok
- Mérési bizonytalanság meghatározása a „Guide to the Expression of Uncertainty in Measurement” (GUM) - „Útmutató a mérési bizonytalanság kifejezéséhez” - előírásai alapján

Paraméterbecslés

- GNSS mérések feldolgozása
- DLT, sugárnyaláb kiegyenlítés
- RANSAC
- Függvény meghatározás

Idősorok elemzése

- Kálmán szűrés elve, a szűrés végrehajtása lineáris és nem lineáris esetekben
- Teljesítménysűrűség spektrum (PSD) és becslése

Adatrendszerek ábrázolása

- Adatrendszer (minta)

  azonos mennyiségre, azonos műszerrel, azonos körülmények között végzett mérések eredménye
- Példa adatrendszer

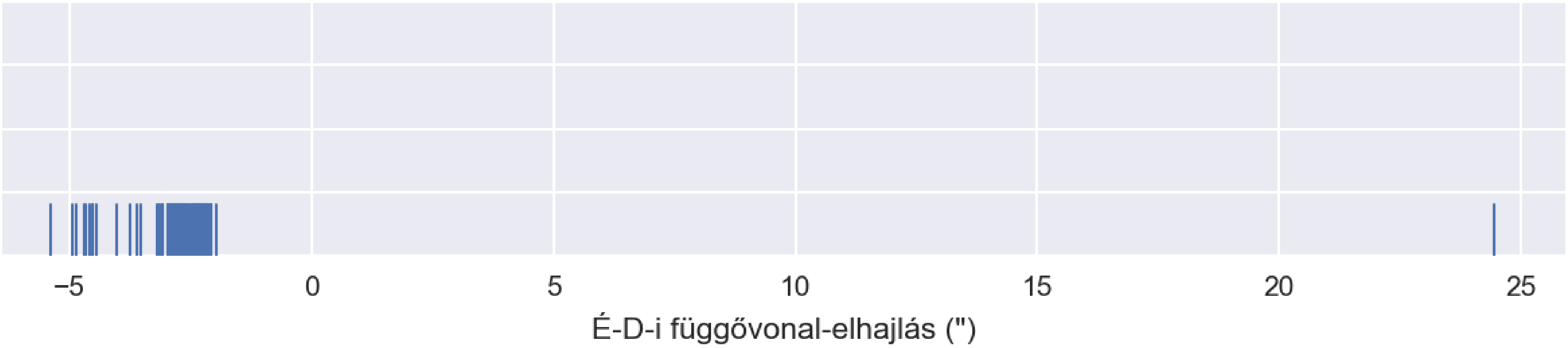
  A tanszéki QDaedalus rendszerrel mért É-D-i irányú függővonal-elhajlás összetevő értéke

  Mérési pont: Soroksár, Pistahegyi út

  Mérések száma: 125 (58 különböző éjszaka)

Ábrázolás számegyenesen

- angolul: rugplot, magyarul: vonalkázós ábra
- Python seaborn: rugplot()

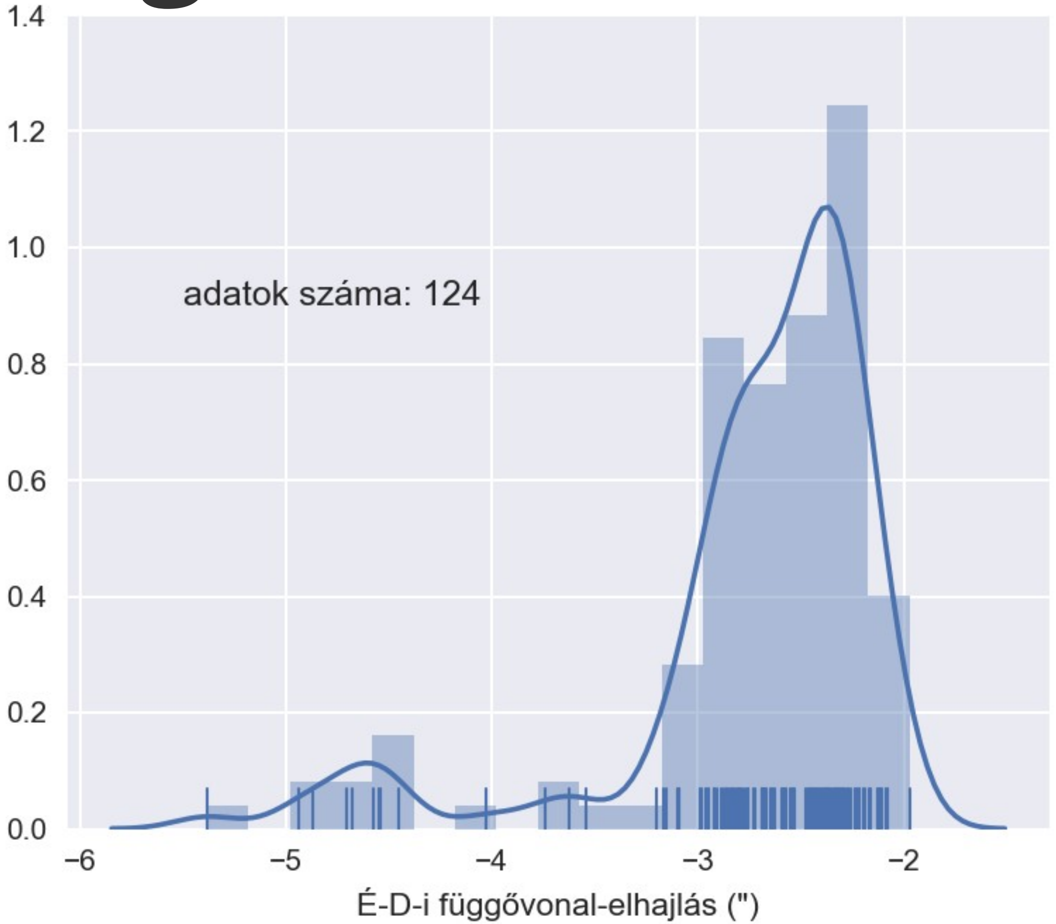


Ábrázolás számegyenesen

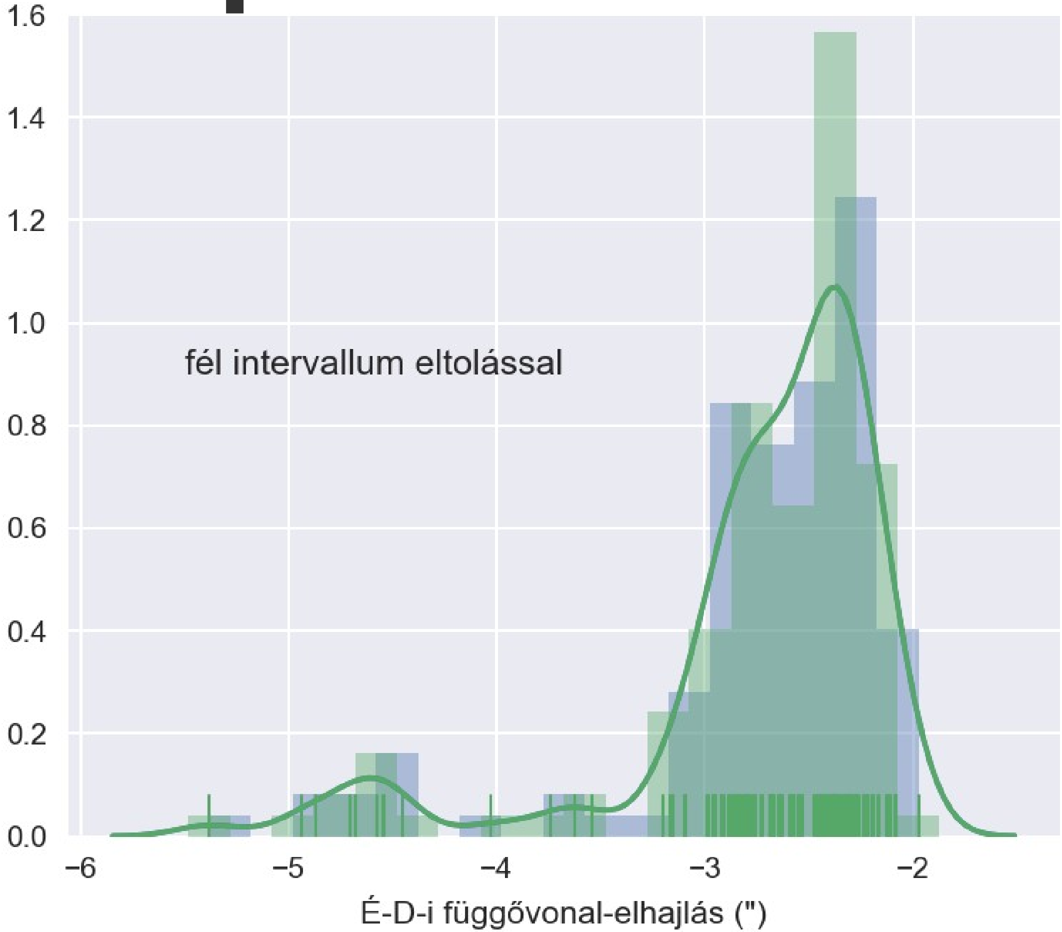
- angolul: rugplot, magyarul: vonalkázós ábra
- Python seaborn: rugplot()

Ábrázolás hisztogramon

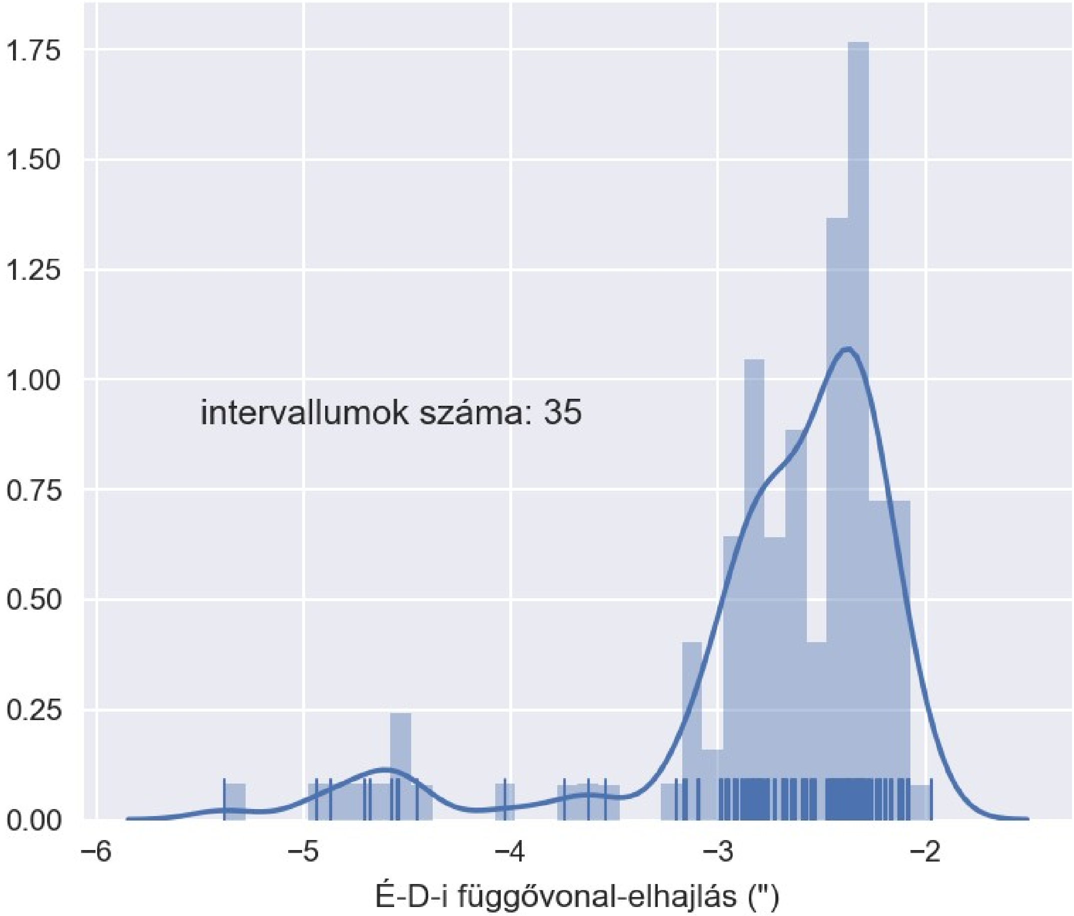
Python Seaborn:
distplot()



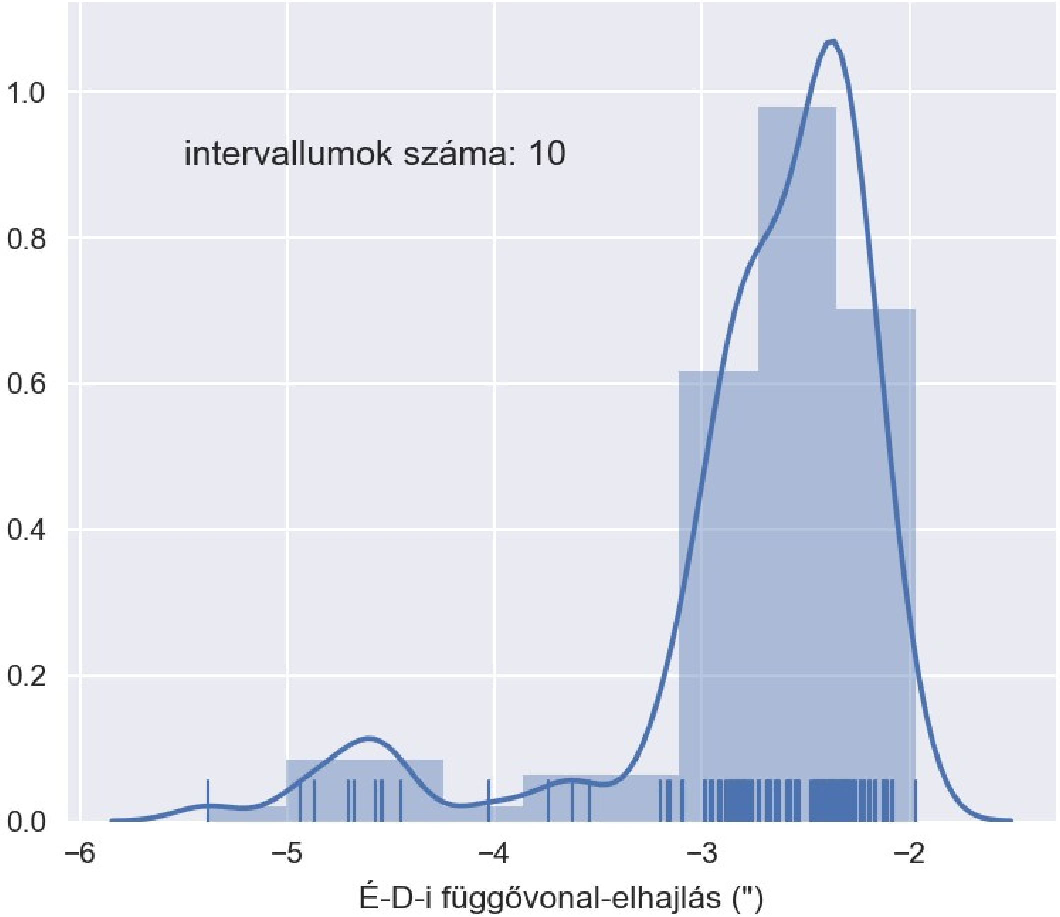
Hisztogram problémák



Hisztogram problémák



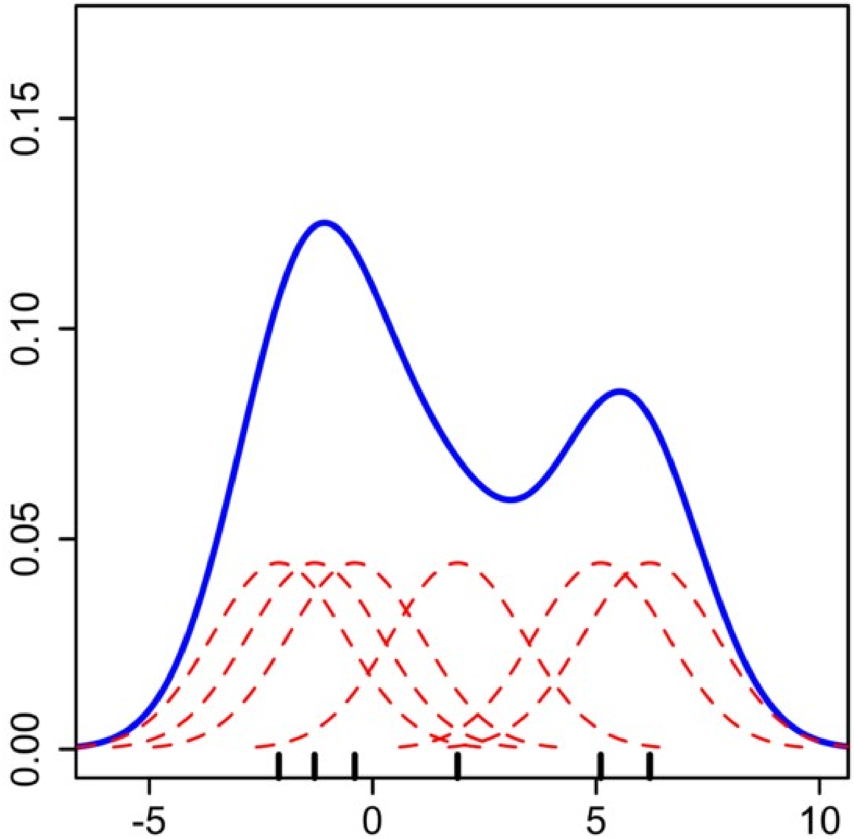
Hisztogram problémák



Hisztogram problémái

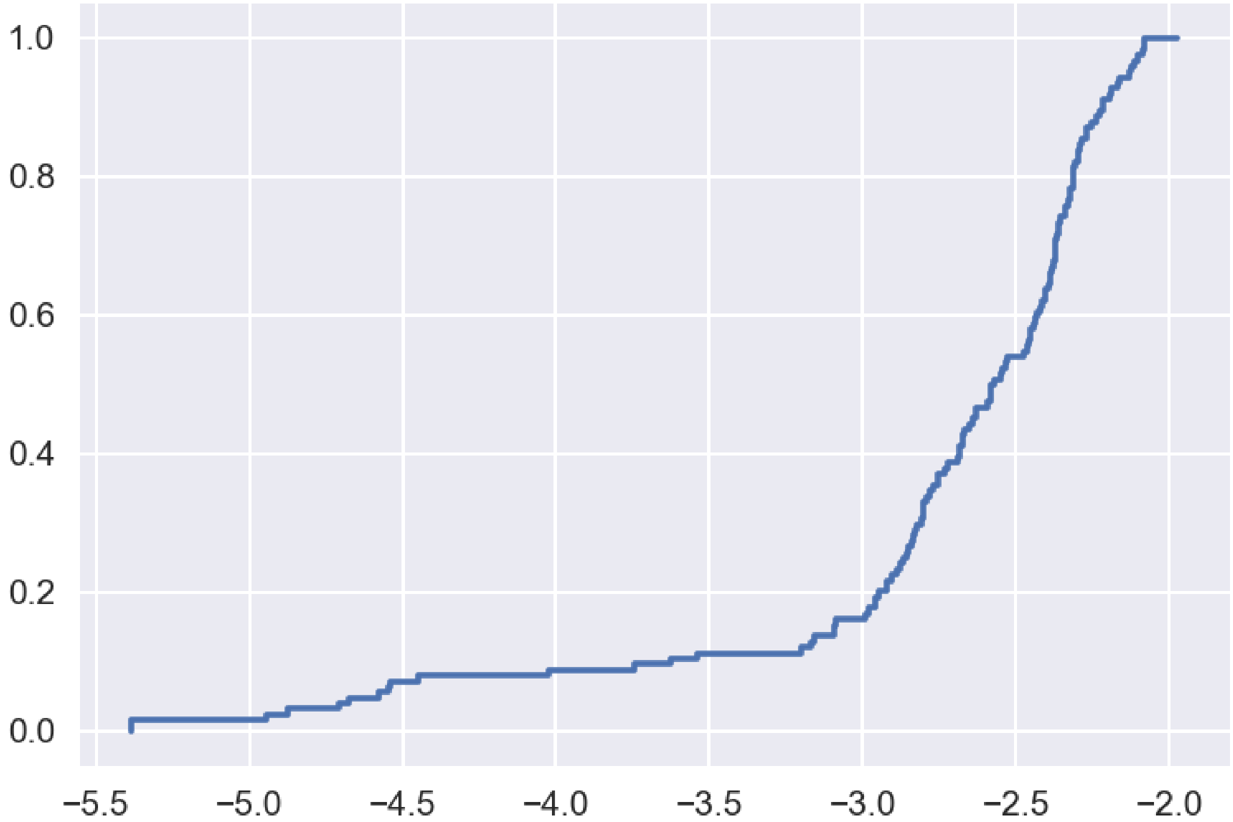
- erősen függ a részintervallumok
    - hosszától
    - elhelyezkedésétől
- megoldási lehetőségek
    - magfüggvényes sűrűségbecslés (kernel density estimate, KDE)
    - kumulált gyakorisági hisztogramok

Magfüggvényes sűrűségbecslés



Kumulált gyakoriságok hisztogramja

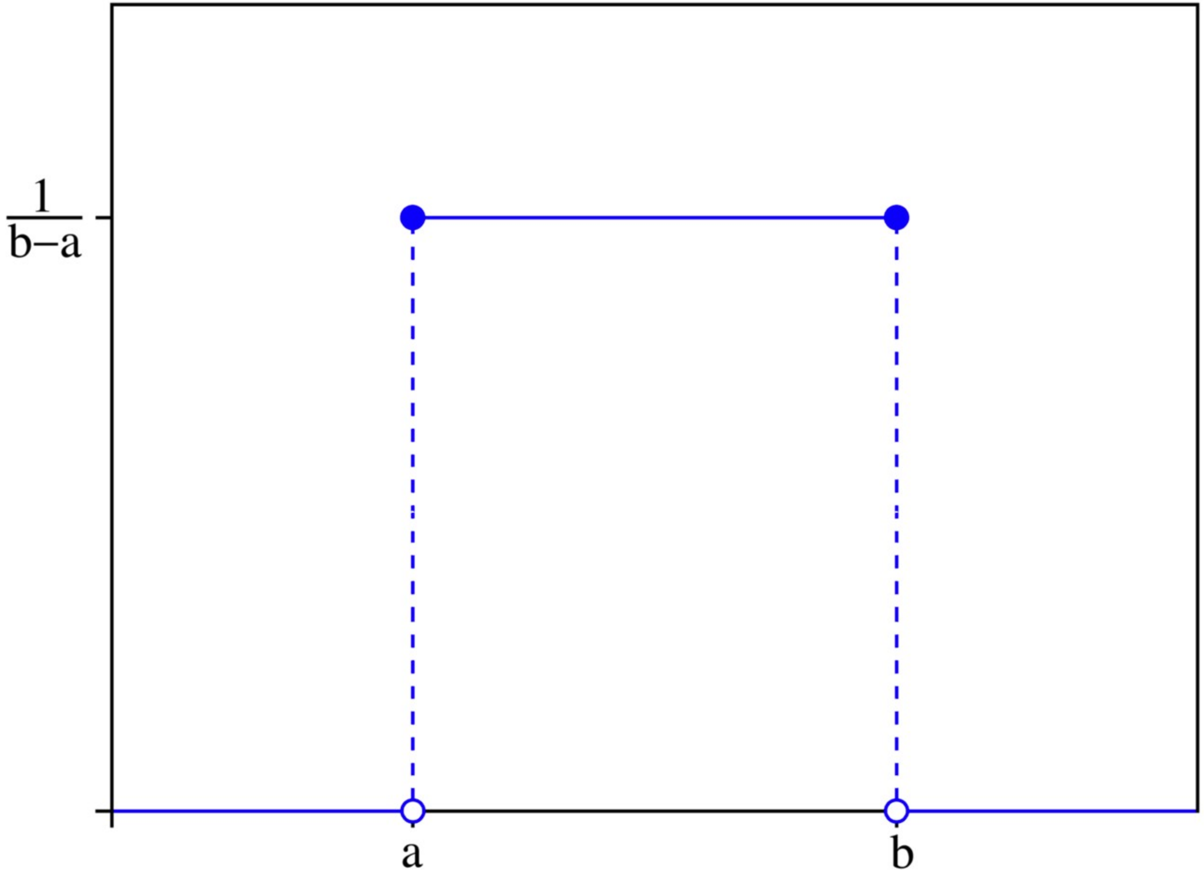
adatpontoknál
1/n-et „ugrik”



Sűrűségfüggvények

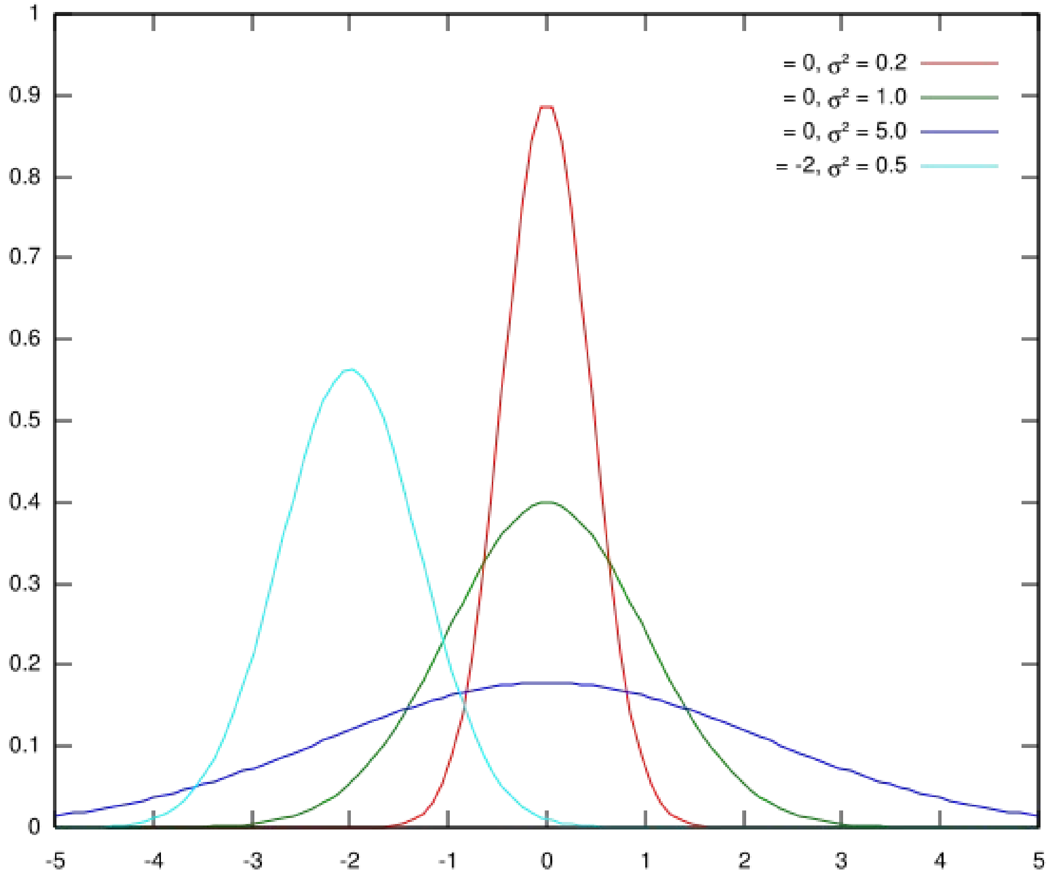
- Folytonos eloszlású valószínűségi változók
- Nevezetes adatsűrűség-modellek
  - egyenletes adateloszlás
  - Gauss-féle sűrűségfüggvény
  - Laplace-féle sűrűségfüggvény
  - Cauchy-féle sűrűségfüggvény
- Modell-családok (szupermodellek)

Egyenletes adateloszlás



1-D eset
két paraméter: a, b

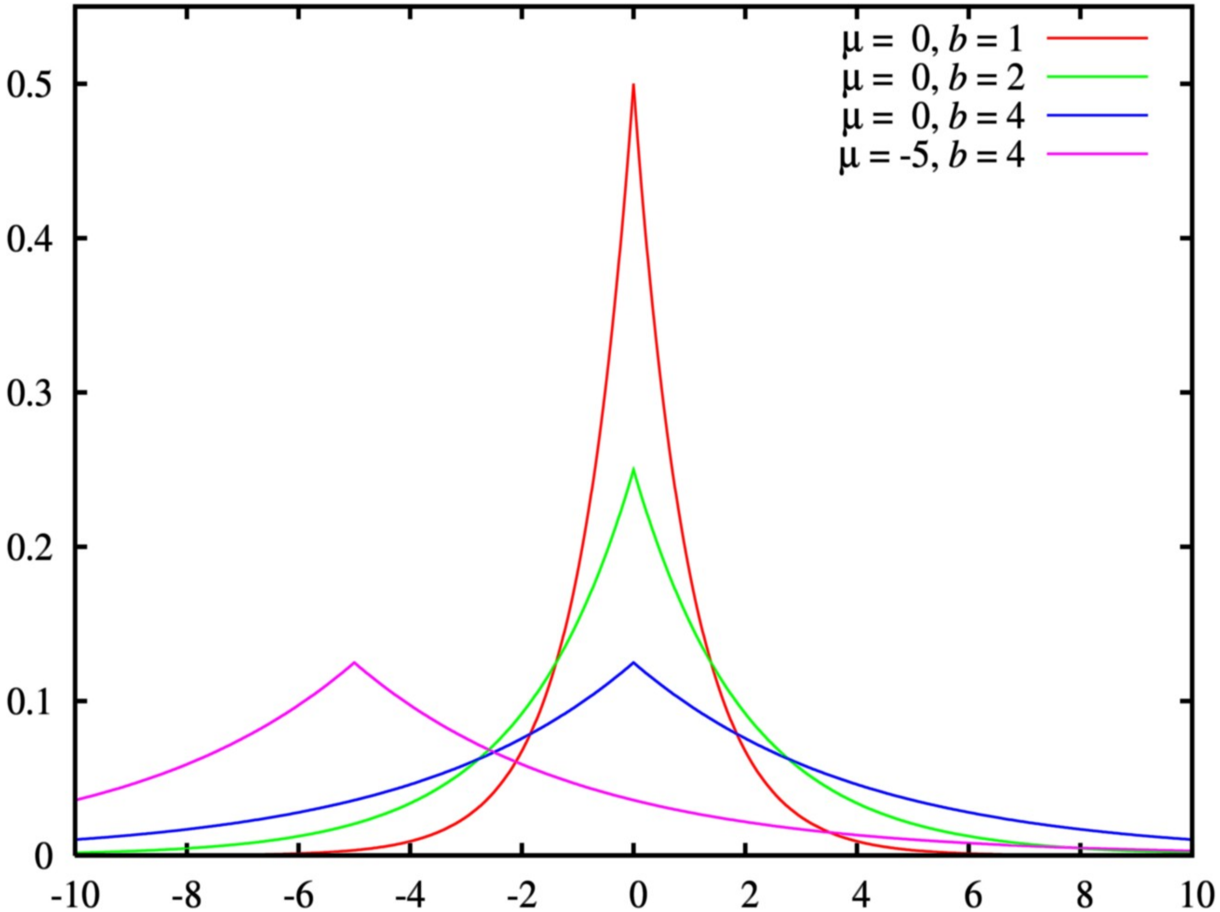
Gauss-féle sűrűségfüggvény



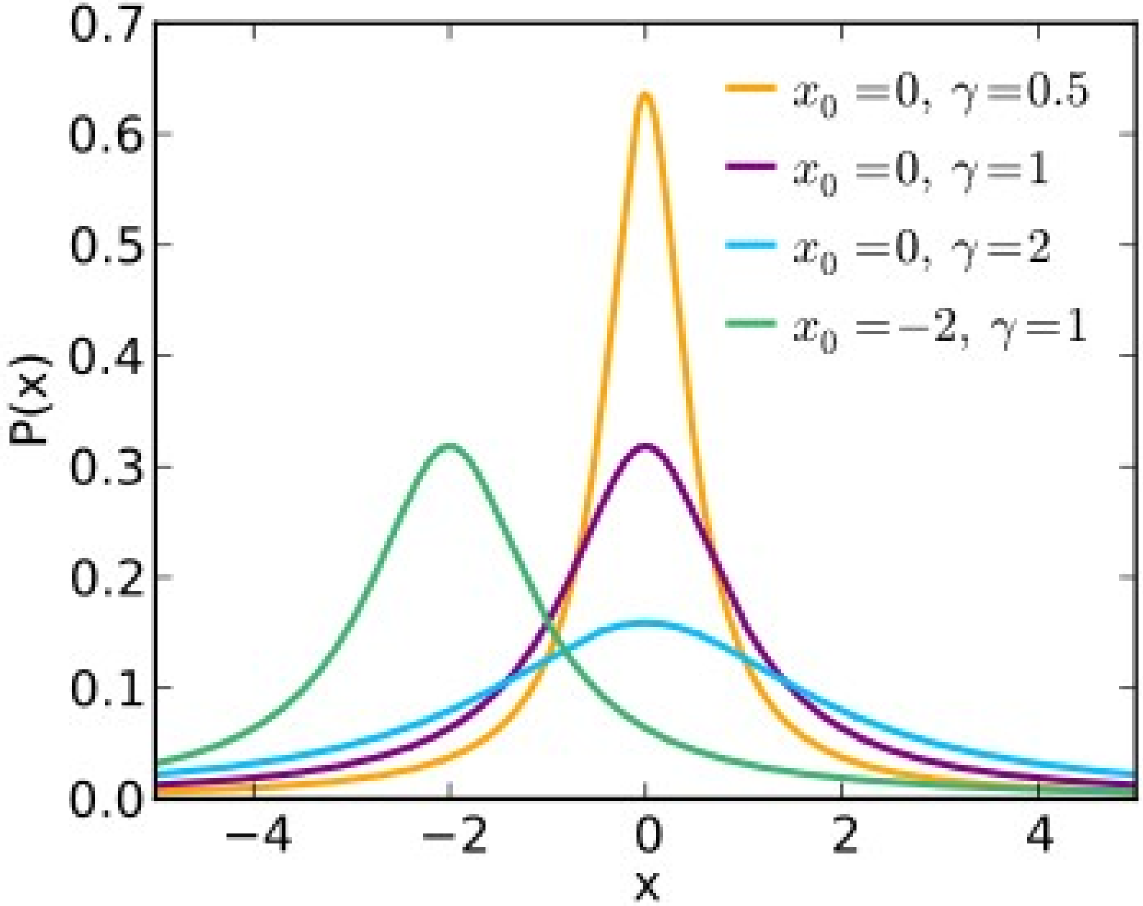
1-D eset
két paraméter:
átlag, szórás

Laplace-féle sűrűségfüggvény

1-D eset
két
paraméter:
μ, b



Cauchy-féle sűrűségfüggvény

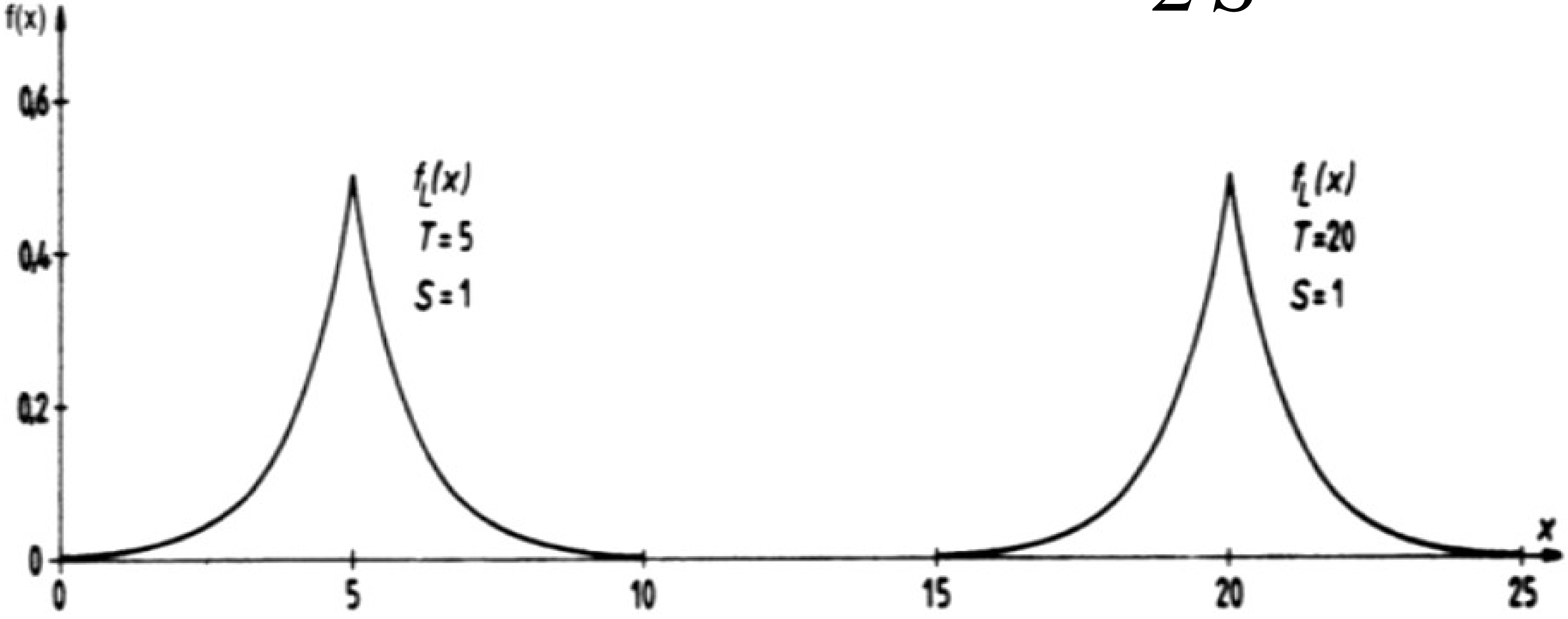


1-D eset
két paraméter:
x_0, γ

Helyparaméter

- T helyparaméter

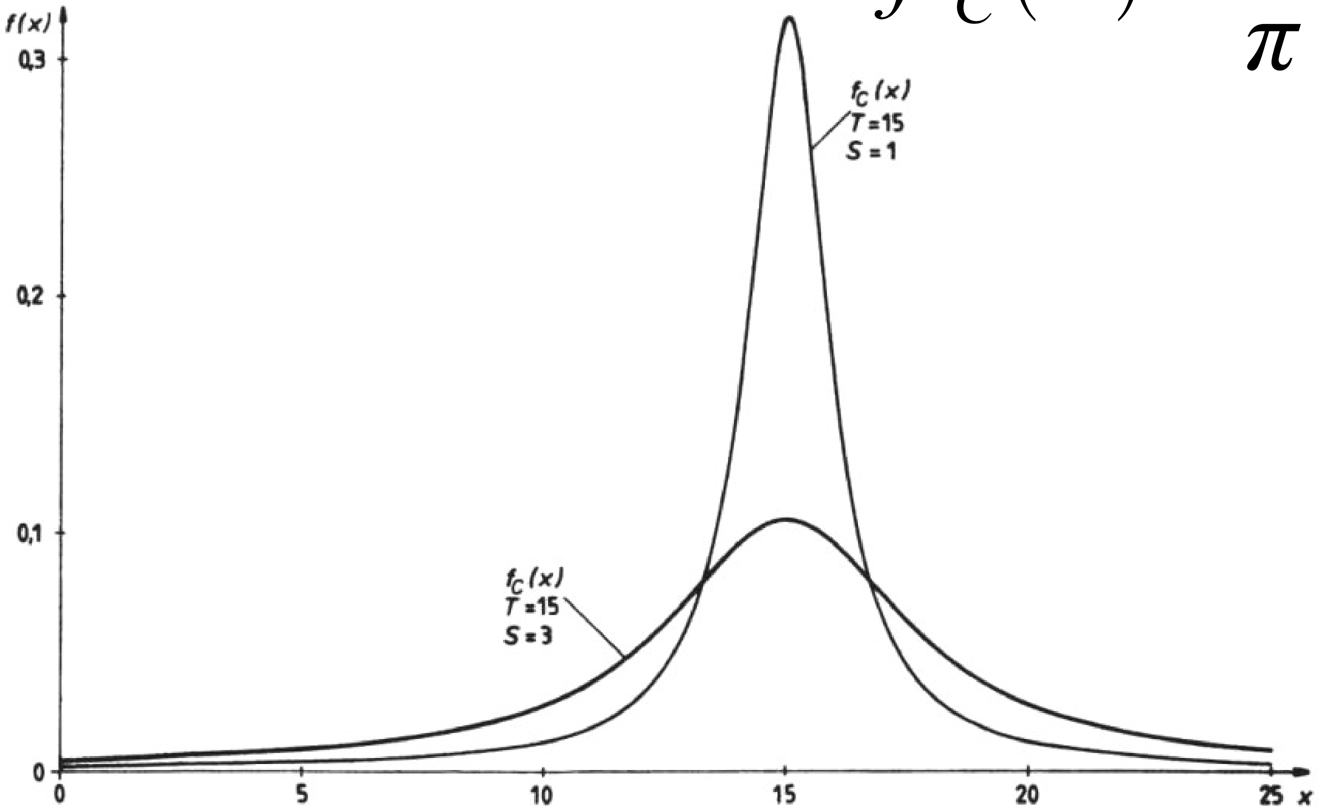
$$f_L(x) = \frac{1}{2S} e^{-\frac{|x-T|}{S}}$$



Skálaparaméter

- S skálaparaméter

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



Modell-családok (ún. szupermodellek)

- analitikusan megadott sűrűségmodell
- egy (vagy több) paraméter változtatásával más és más jellegű sűrűségmodellhez jutunk
- statisztikai eljárások vizsgálatához hasznos eszköz

    robusztusság, hatásfok

Az $f_a(x)$ szupermodell

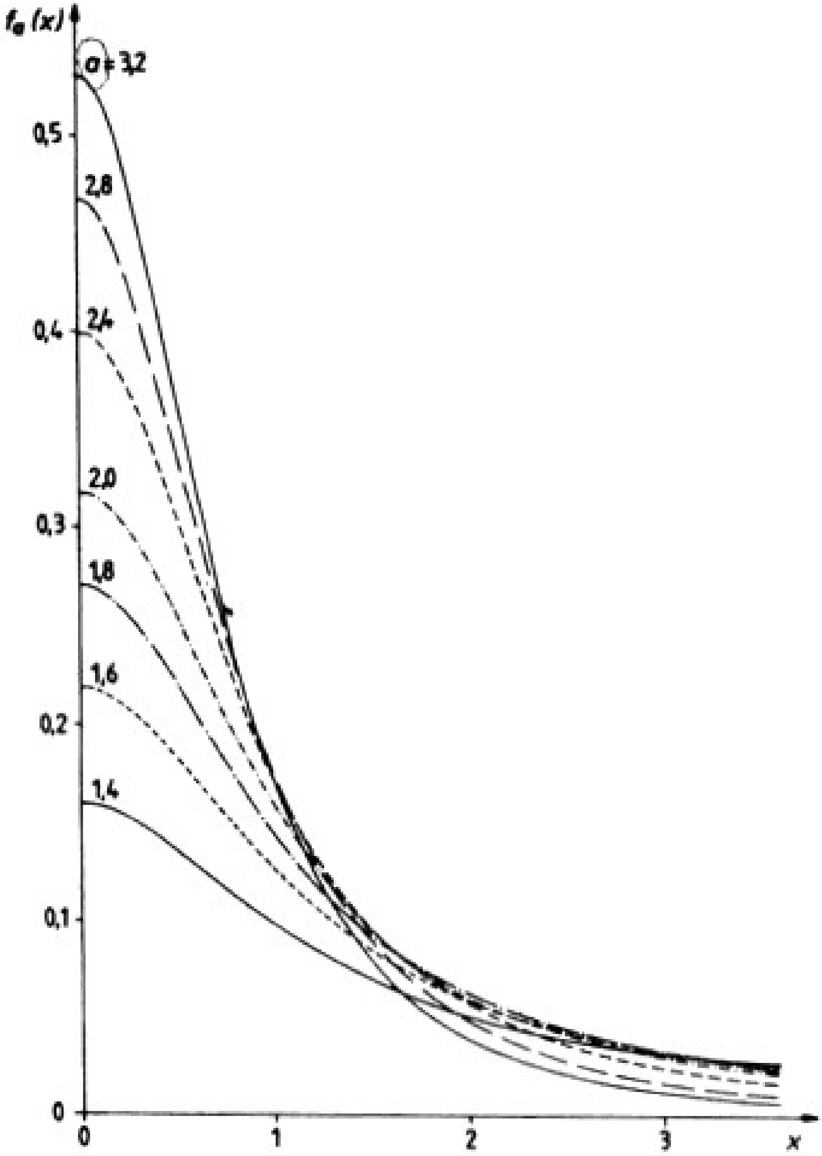
- szimmetrikus sűrűségfüggvény
- standard alak (T=0, S=1)
- általános alak mindig megkapható x helyére $(x-T)/S$-et írva és S-el osztva

$$f_a(x) = n(a) \frac{1}{[\sqrt{x^2+1}]^a}$$

$$n(a) = \frac{\Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{a-1}{2}\right)}$$

Az $f_a(x)$ szupermodell

- a : típusparaméter
- $a \longrightarrow \infty$: Gauss-eloszlás
- $a = 2$: Cauchy-eloszlás
- $a = N + 1$: N szabadságfokú Student-eloszlás

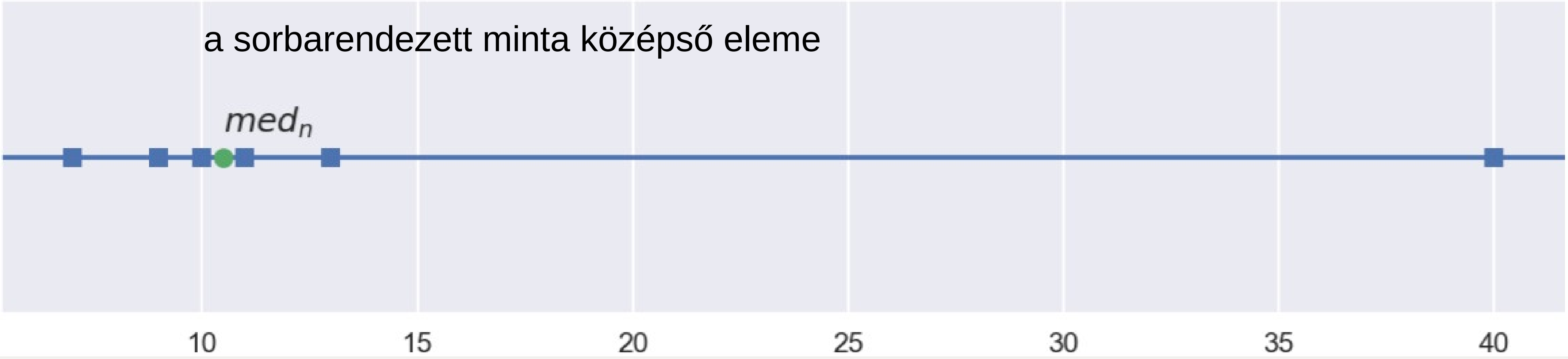


A legjellemzőbb érték meghatározása

- Minta: 7; 9; 10; 11; 13 és 40
  (6 elemű mini adatrendszer, 1 kieső értékkel)
- Mit fogadjunk el a minta *jellemző* értékének?
  kicsiny (néhányszor 10 elemszámú minta),
  aszimmetrikus adateloszlások esetén
- mintamedián (med_n)
- számtani átlag (E_n)
- leggyakoribb érték (M_n)

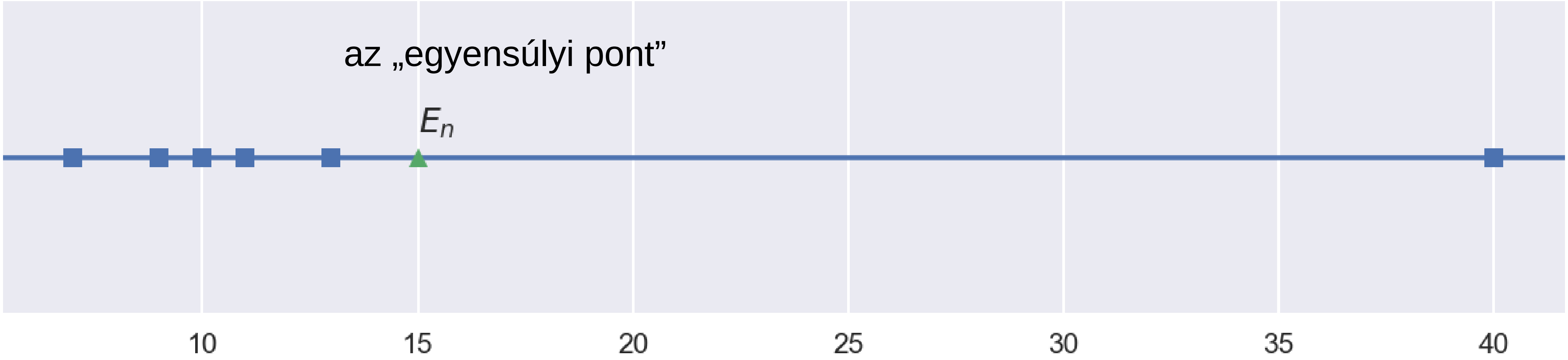
A mintamedián (med_n)

$$\text{med}_n = \begin{cases} x_{(n+1)/2} & \text{ha } n \text{ páratlan} \\ \dfrac{x_{n/2} + x_{(n+2)/2}}{2} & \text{ha } n \text{ páros} \end{cases}$$



A számtani átlag (E_n)

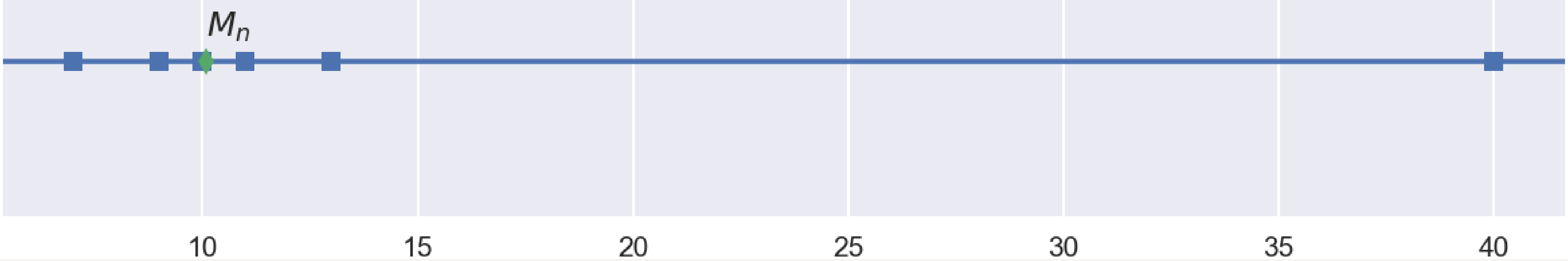
$$E_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$



A leggyakoribb érték (M_n)

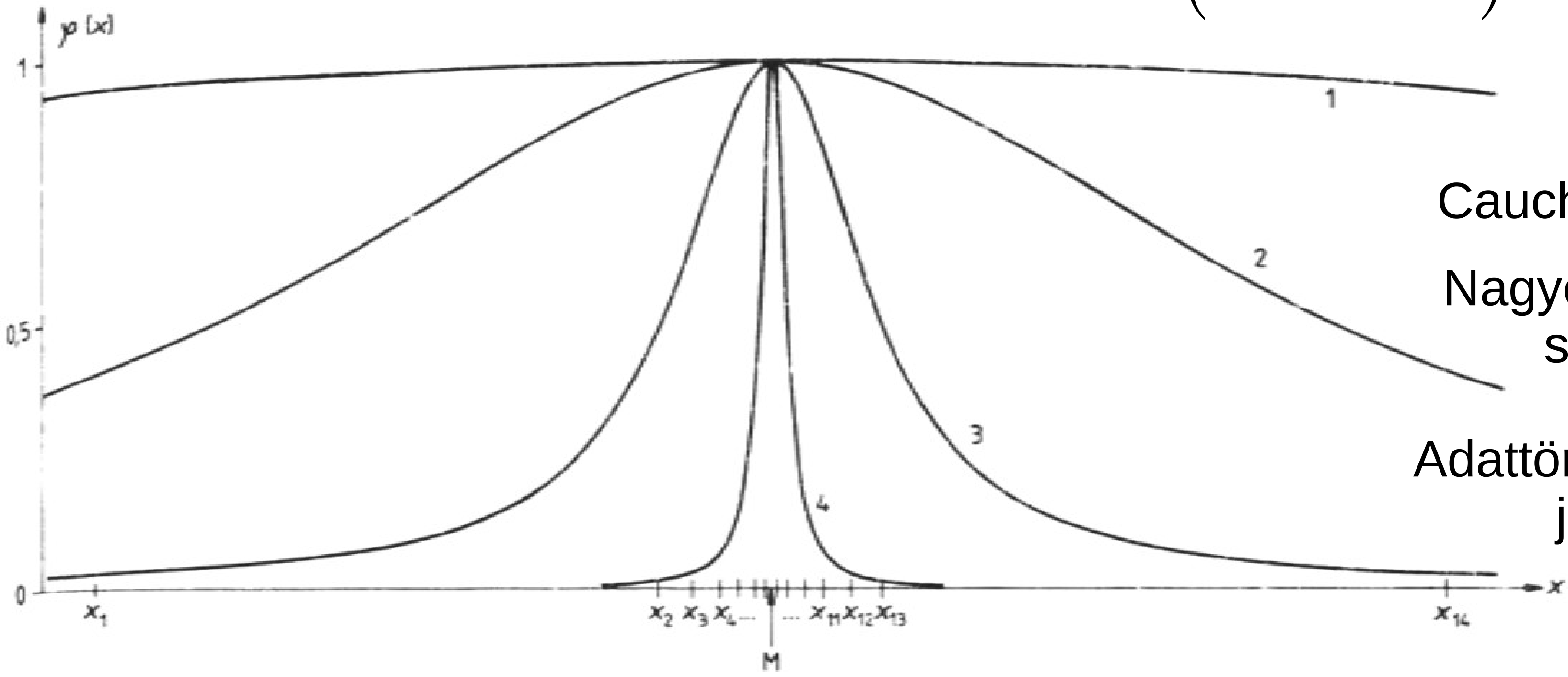
$$M_n = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i} = \frac{\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_n)^2} x_i}{\sum_{i=1}^{n} \frac{\varepsilon^2}{\varepsilon^2 + (x_i - M_n)^2}}$$

súlyozott átlag



A súlyfüggvény alakja különböző ε értékekre

$$p(x) = \frac{\varepsilon^2}{\varepsilon^2 + (x - M)^2}$$



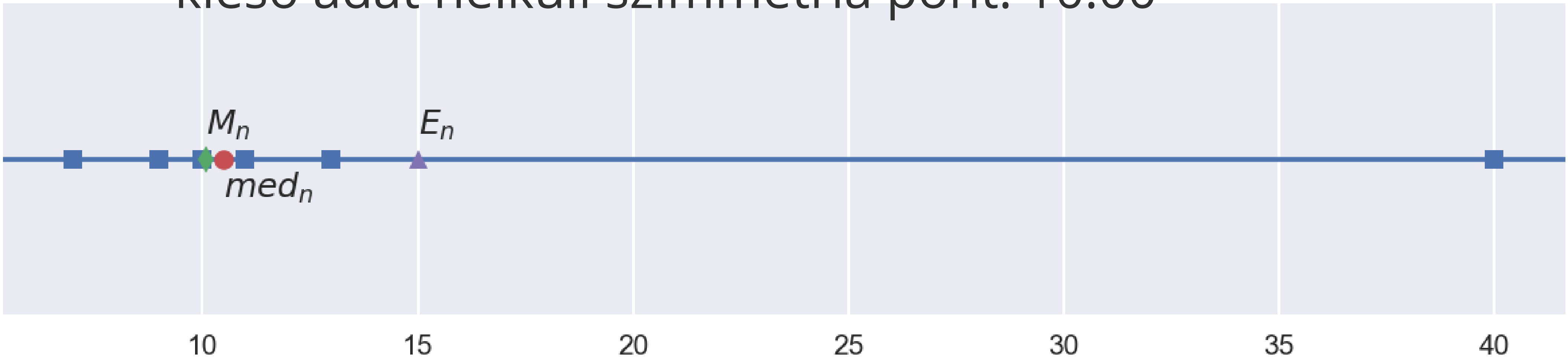
Cauchy-Steiner súlyok

Nagyobb ε, szélesebb súlyfüggvény

Adattömörödés mértékét jellemzi az ε

Összehasonlítás

- E_n = 15 (kieső adatra érzékeny)
- med_n = 10.5 (kieső adatra nem érzékeny)
- M_n = 10.07 (kieső adatra nem érzékeny)
- kieső adat nélküli szimmetria pont: 10.00



Dihézió (reciprok kohézió) (ε)

$$\varepsilon_0 = \frac{\sqrt{3}}{2}\left[\max(x_i) - \min(x_i)\right]$$

iteráció

$$\varepsilon_{k+1}^2 = \frac{3\sum_{i=1}^{n} \frac{(x_i - M_n)^2}{\left[\varepsilon_k^2 + (x_i - M_n)^2\right]^2}}{\sum_{i=1}^{n} \frac{1}{\left[\varepsilon_k^2 + (x_i - M_n)^2\right]^2}}$$

M_n és ε együttes számítása

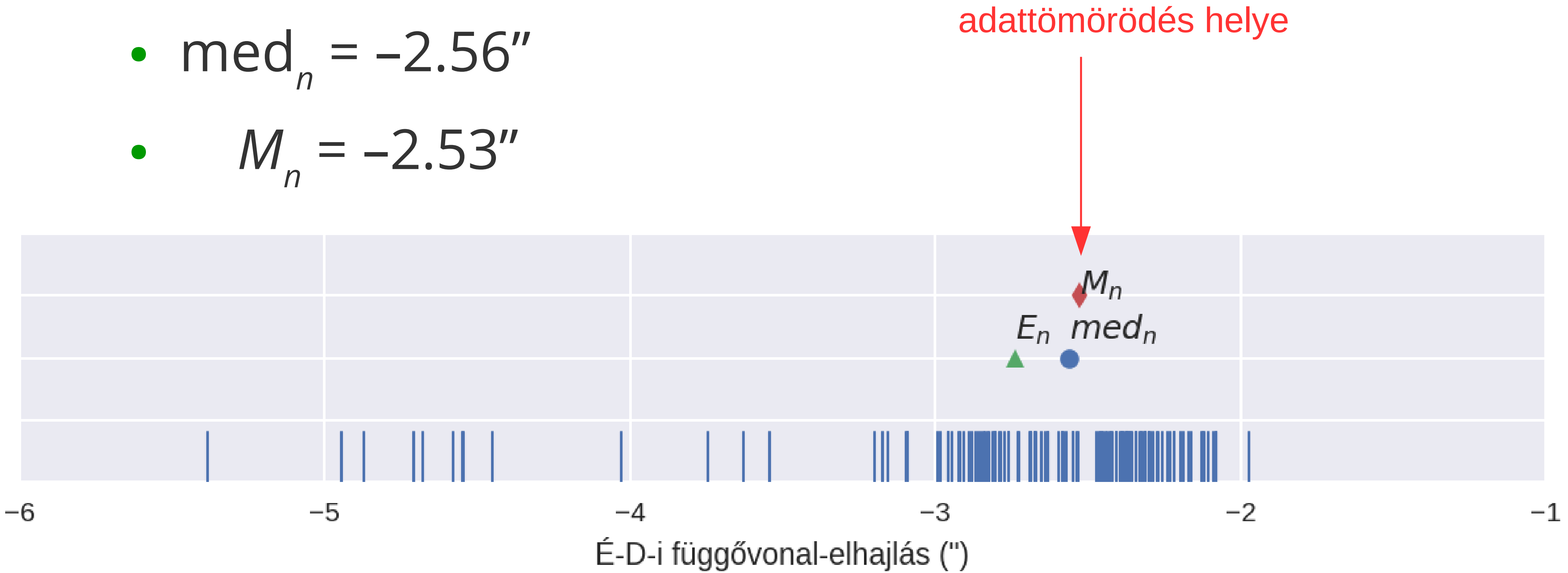
- ping-pong iteráció
- kiindulásként:

  $\varepsilon = \varepsilon_0$

  $M_n = E_n$ vagy med_n
- ε_1 számítása, majd M_n második közelítésének számítása
- stb.

Minta adatainkkal

- E_n = –2.74″
- med_n = –2.56″
- M_n = –2.53″



Az adatrendszerben rejlő bizonytalanság jellemzése

- adatrendszer távolsága egy x_0 értéktől
- távolságdefiníciók (ún. normák)
- adatrendszer legjellemzőbb értéke mint valamely norma minimumhelye
- adatrendszerben rejlő bizonytalanság mint a norma értéke a minimumhelyen
- bizonytalanság jellemzése különböző konfidenciaszintekhez tartozó intervallumokkal

Egyetlen adat távolsága x_0-tól

- geometriai távolság $$|x_i - x_0|$$
- általánosított távolság $$|x_i - x_0|^p$$
- statisztikai szempontból indokolható „távolság”

$$\varepsilon^2 + (x_i - x_0)^2$$

→ azért jó, mert nincs okunk különbséget tenni x_0-hoz nagyon közeli pontok „távolságai” között

Adatrendszer távolsága x_0-tól

- általánosított távolság
$$\sum_{i=1}^{n} |x_i - x_0|^p$$
- ha $p = 1$
$$\sum_{i=1}^{n} |x_i - x_0|$$
- statisztikai szempontból indokolható „távolság”
$$\prod_{i=1}^{n} \left[\varepsilon^2 + (x_i - x_0)^2\right]$$

  a szorzat alkalmazásával csökkentjük az igen nagy abszolút értékű $(x - x_0)$ eltérések hatását

Adatrendszer távolsága x_0-tól

- függetlenítjük *n*-től
- a távolság dimenziója egyezzen meg x_i-k dimenziójával
- általánosított távolság

$$\left[\frac{1}{n}\sum_{i=1}^{n}\left|x_i-x_0\right|^p\right]^{1/p}$$

- statisztikai szempontból indokolható „távolság”

$$\left\{\prod_{i=1}^{n}\left[\varepsilon^2+\left(x_i-x_0\right)^2\right]\right\}^{\frac{1}{2n}}=\varepsilon\cdot\left\{\prod_{i=1}^{n}\left[1+\left(\frac{x_i-x_0}{\varepsilon}\right)^2\right]\right\}^{\frac{1}{2n}}$$

Távolságdefiníciók (norma)

- n legyen igen nagy: adatrendszereinket igen pontosan jellemzi egy $f(x)$ sűrűségfüggvény
- L_p norma

$$L_p = \left[\int_{-\infty}^{\infty} |x - x_0|^p f(x)\, dx \right]^{1/p}$$

- P_1 norma

$$P_1 = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{x - x_0}{\varepsilon} \right)^2 \right] \cdot f(x)\, dx \right\}$$

Az adatrendszer legjellemzőbb értéke

- Az adatrendszer legjellemzőbb értékét az adott norma *minimumhelyeként* kapjuk meg

  (bizonyítás: Steiner (1990) 86-87. oldal)

- L_1 norma minimumhelye: medián
- L_2 norma minimumhelye: számtani átlag
- P_1 norma minimumhelye: leggyakoribb érték

Az adatrendszerben rejlő bizonytalanság

- Az adatrendszerben rejlő bizonytalanság értékét a norma minimumhelyen felvett értékeként kapjuk meg:

  L_1 norma: *közepes eltérés* (d)

  L_2 norma: *szórás* (σ)

  P_2 norma: *határozatlanság* (U)

Általánosított leggyakoribb érték ($M_{k;n}$)

$$M_{k;n} = \frac{\displaystyle\sum_{i=1}^{n} \frac{x_i}{(k\,\varepsilon)^2 + (x_i - M_{k;n})^2}}{\displaystyle\sum_{i=1}^{n} \frac{1}{(k\,\varepsilon)^2 + (x_i - M_{k;n})^2}}$$

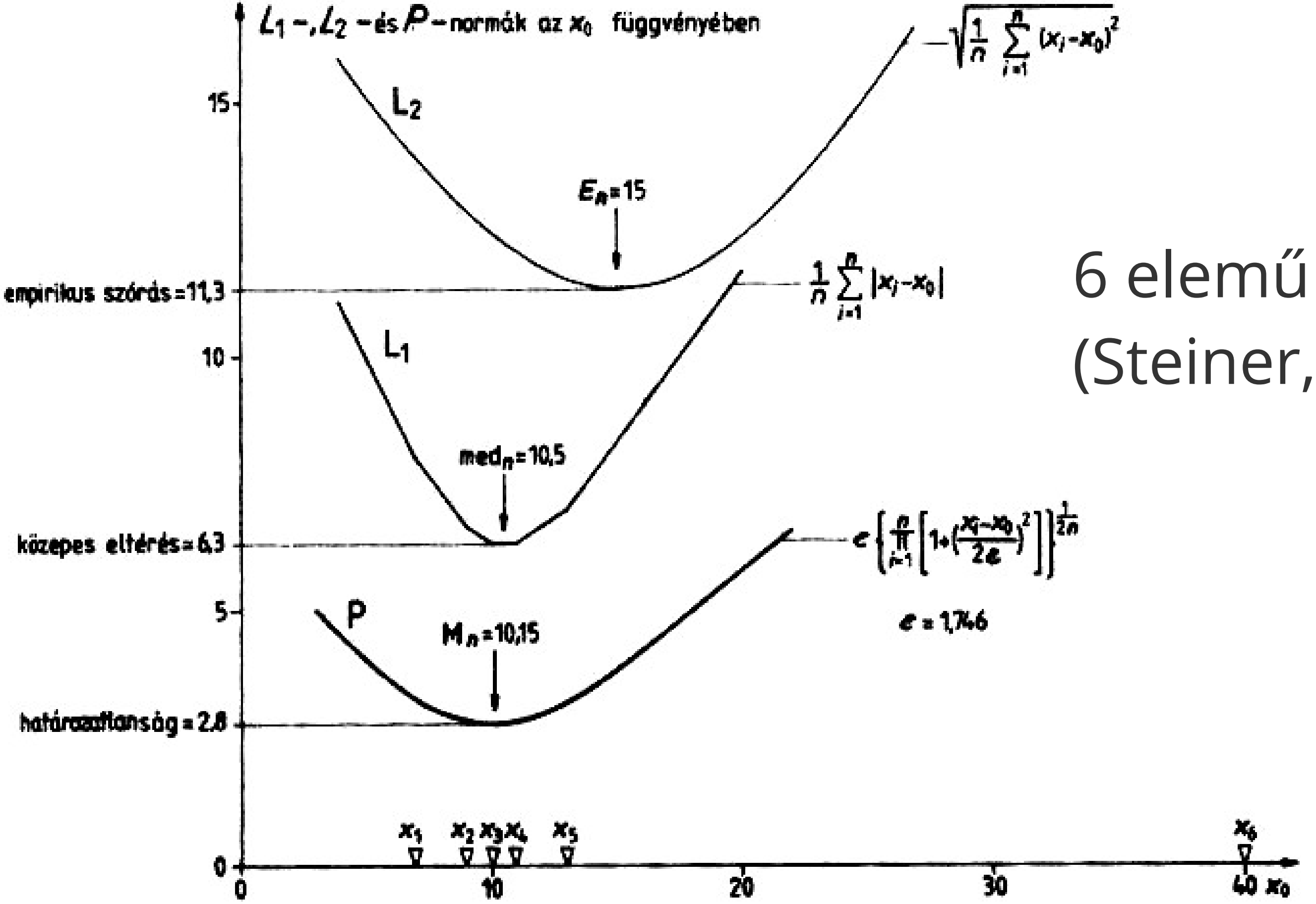
- leggyakrabban $k = 1, 2$

P_k norma

$$P_1 = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{x - x_0}{\varepsilon} \right)^2 \right] \cdot f(x)\, dx \right\}$$

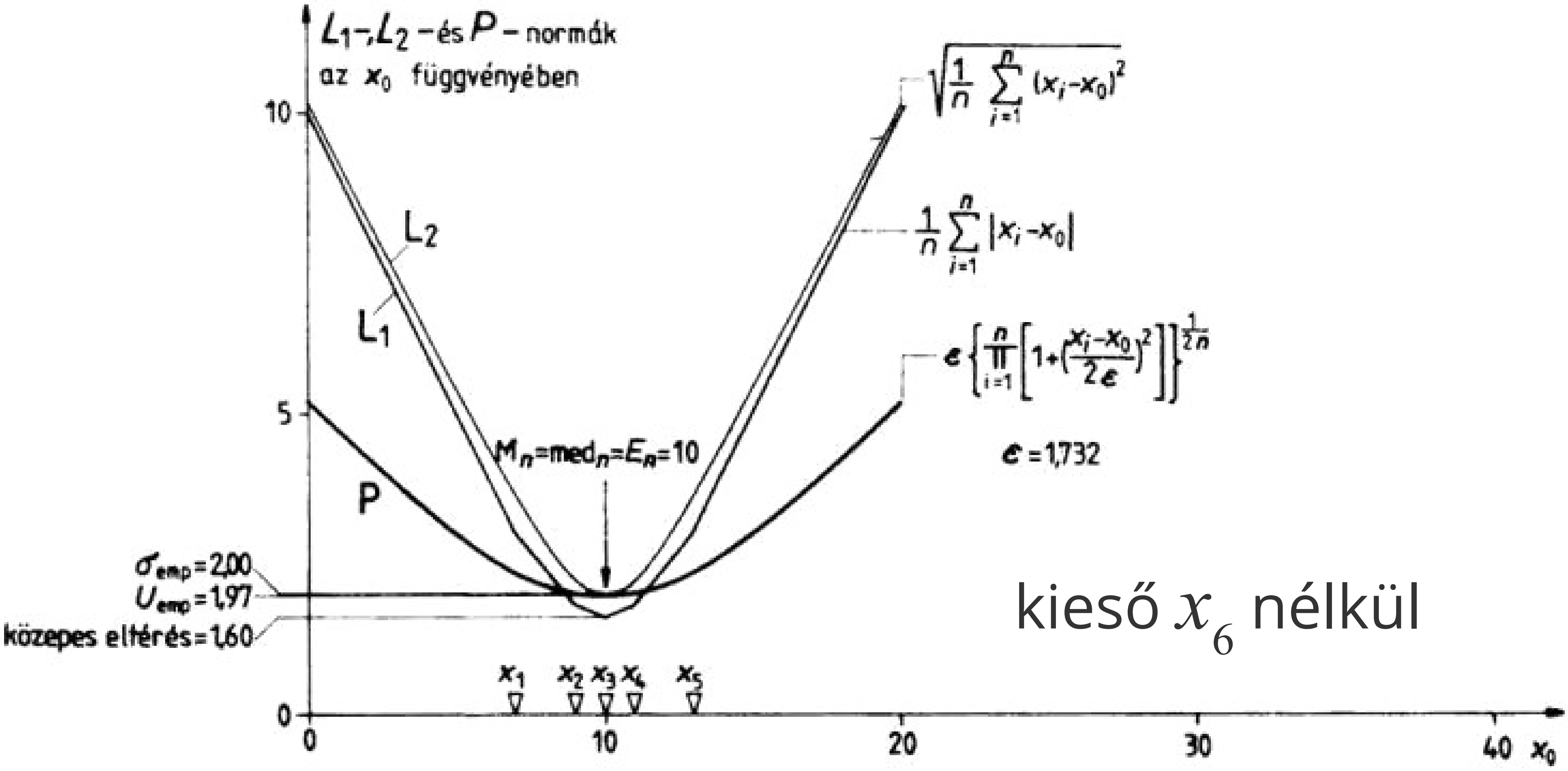
$$P_k = \varepsilon \cdot \exp\left\{ \frac{1}{2} \int_{-\infty}^{\infty} \ln\left[1 + \left(\frac{x - x_0}{k\,\varepsilon} \right)^2 \right] \cdot f(x)\, dx \right\}$$

Normák összehasonlítása



6 elemű mini adatrendszerre (Steiner, 1990)

Normák összehasonlítása



kieső x_6 nélkül

Konfidencia-intervallumok

- interkvartilis intervallum: $[-q, q]$

  az adatok 1/2-ede várható

  $-q$: alsó kvartilis, $+q$: felső kvartilis

  q: interkvartilis félterjedelem

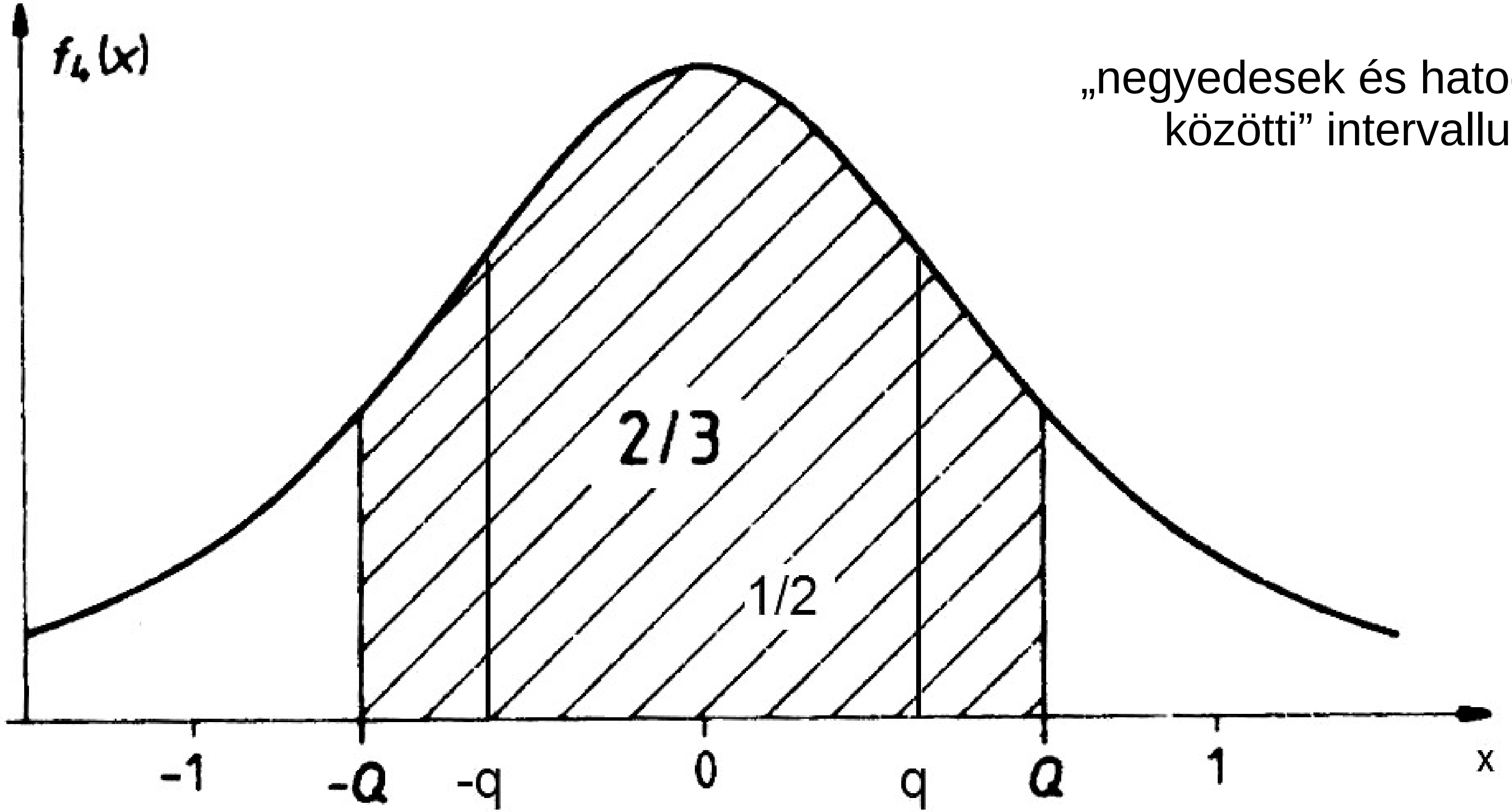
- interszeksztilis intervallum: $[-Q, Q]$

  az adatok 2/3-a várható

  $-Q$: alsó szeksztilis, $+Q$: felső szeksztilis

  Q: interszeksztilis félterjedelem

Interkvartilis és interszeksztilis intervallum



„negyedesek és hatodosok közötti” intervallum

Hibajellemzők kapcsolata – $f_a(x)$ szupermodell

a	$1/(a-1)$	q	Q	d	σ	U	ε
2 Cauchy	1.0000	1.0000	1.7320	∞	∞	1.5000	1.0000
3	0.5000	0.5774	0.8944	1.0000	∞	0.8605	0.6974
5	0.2500	0.3703	0.5497	0.5000	0.7071	0.5585	0.4818
∞ Gauss	0	0.6745	0.9674	0.7979	1	1.0327	0.9254

Mérési eredmények eloszlásának vizsgálata

- Illeszkedésvizsgálat: Kolmogorov-próba
- Kvantilis-kvantilis ábra

Nemparaméteres próba - illeszkedésvizsgálat

- A Kolmogorov-próba arra a kérdésre válaszol, hogy az $F(x)$ eloszlásfüggvénnyel jellemzett eloszlás, amelyből a minta származik, tekinthető-e valamely ***előre megadott*** folytonos $F_0(x)$ eloszlással azonosnak?

  apró betűs rész: Ha az eloszlás paramétereit *a mintából becsüljük*, akkor a Kolmogorov-próba többé nem eloszlás független, ezért a statisztika eloszlását külön meg kell határoznunk, például Monte Carlo szimulációval.

- Nullhipotézis H_0: $F(x) = F_0(x)$

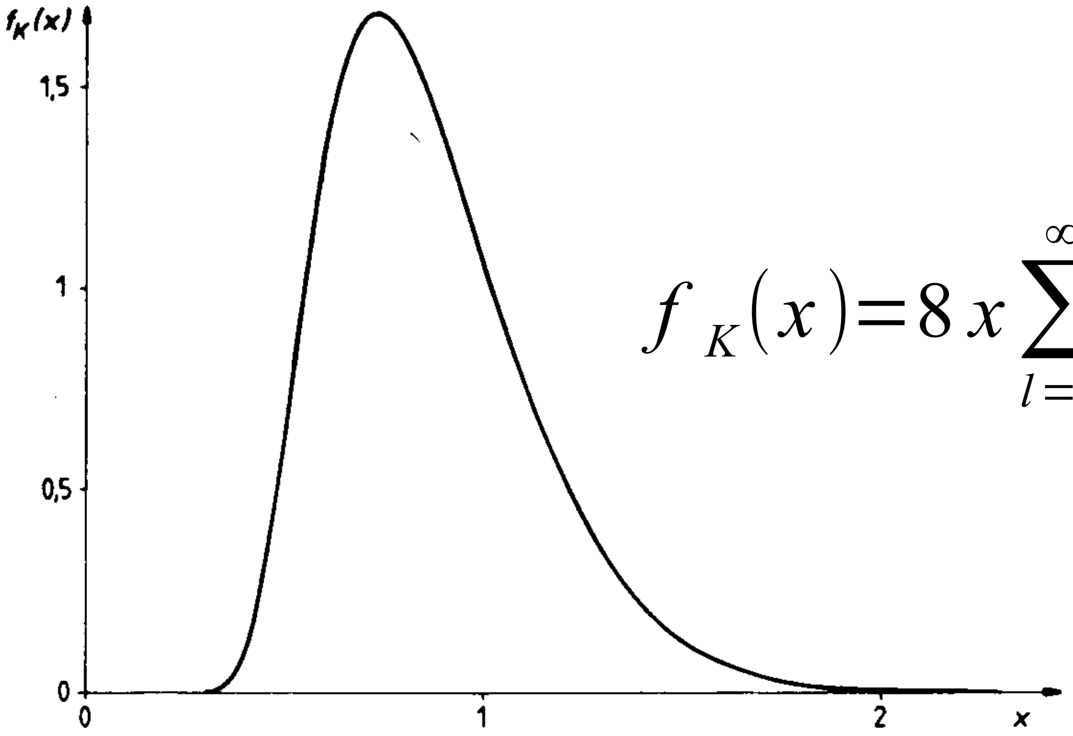
Kolmogorov-próba statisztikai függvénye

- A mintából meghatározzuk az $F_n(x)$ tapasztalati eloszlásfüggvényt (kumulált gyakoriságok hisztogramját)
- Meghatározzuk e függvény legnagyobb eltérését $F_0(x)$-tól
- Ez az eltérés $\sqrt{n}$-nel szorozva adja a K statisztikát:

$$K = \sqrt{n} \cdot \underset{-\infty < x < \infty}{\text{supremum}} \left\{ \left| F_n(x) - F_0(x) \right| \right\}$$

Kolmogorov-féle statisztika sűrűségfüggvénye

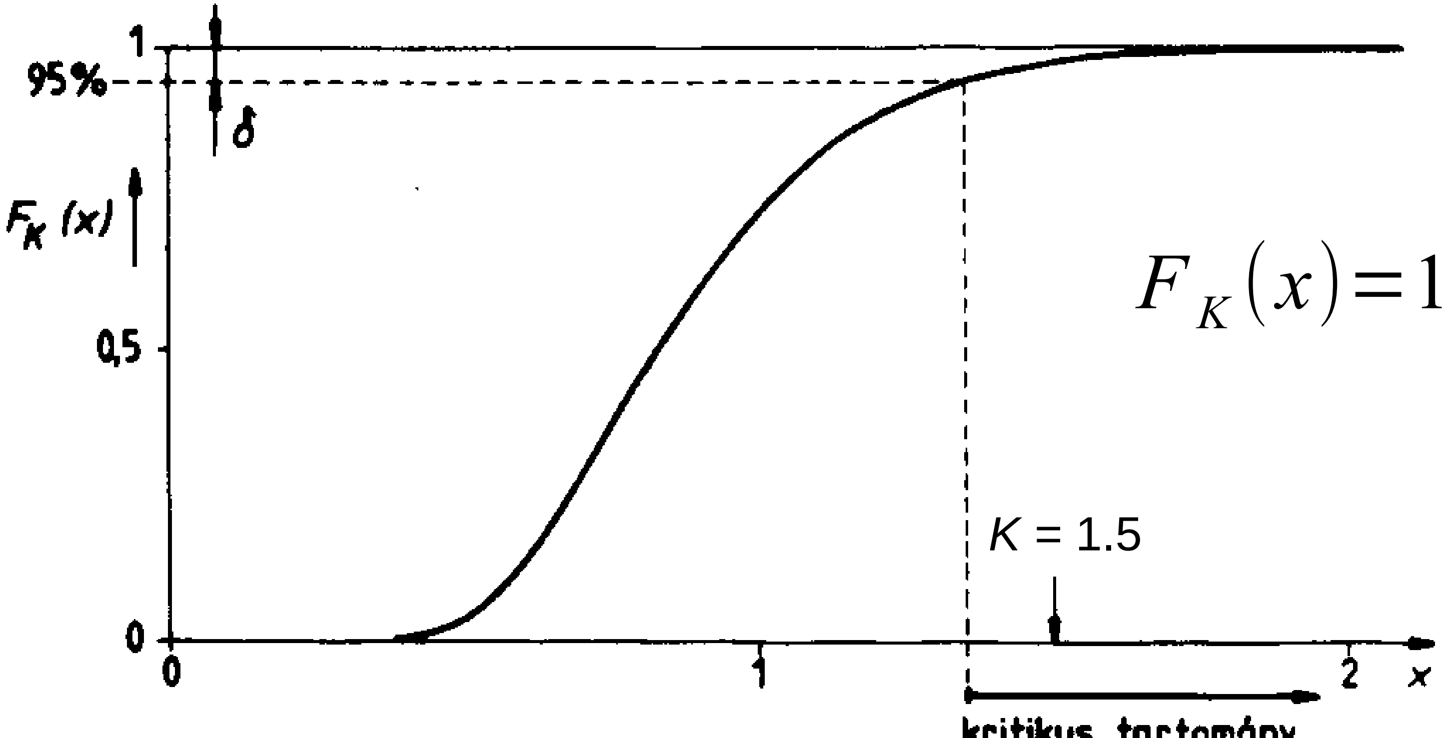
- A K statisztika $f_K(x)$ elméleti sűrűségfüggvénye $n \to \infty$ esetén:



$$f_K(x) = 8x \sum_{l=1}^{\infty} (-1)^{l+1} l^2 e^{-2l^2x^2} \quad (x \geq 0)$$

Kolmogorov-féle statisztika eloszlásfüggvénye

- A K statisztika $F_K(x)$ elméleti eloszlásfüggvénye $n \to \infty$ esetén:



$$F_K(x) = 1 + 2\sum_{l=1}^{\infty} (-1)^l \cdot e^{-2l^2x^2}$$

Kolmogorov-féle kétmintás próba

- Két minta azonos eloszlásból származik-e?
- A két, n és m elemű mintából meghatározott tapasztalati eloszlásfüggvény: $F_n(x)$ és $F_m(x)$
- Statisztikai függvény (Kolmogorov-eloszlású)

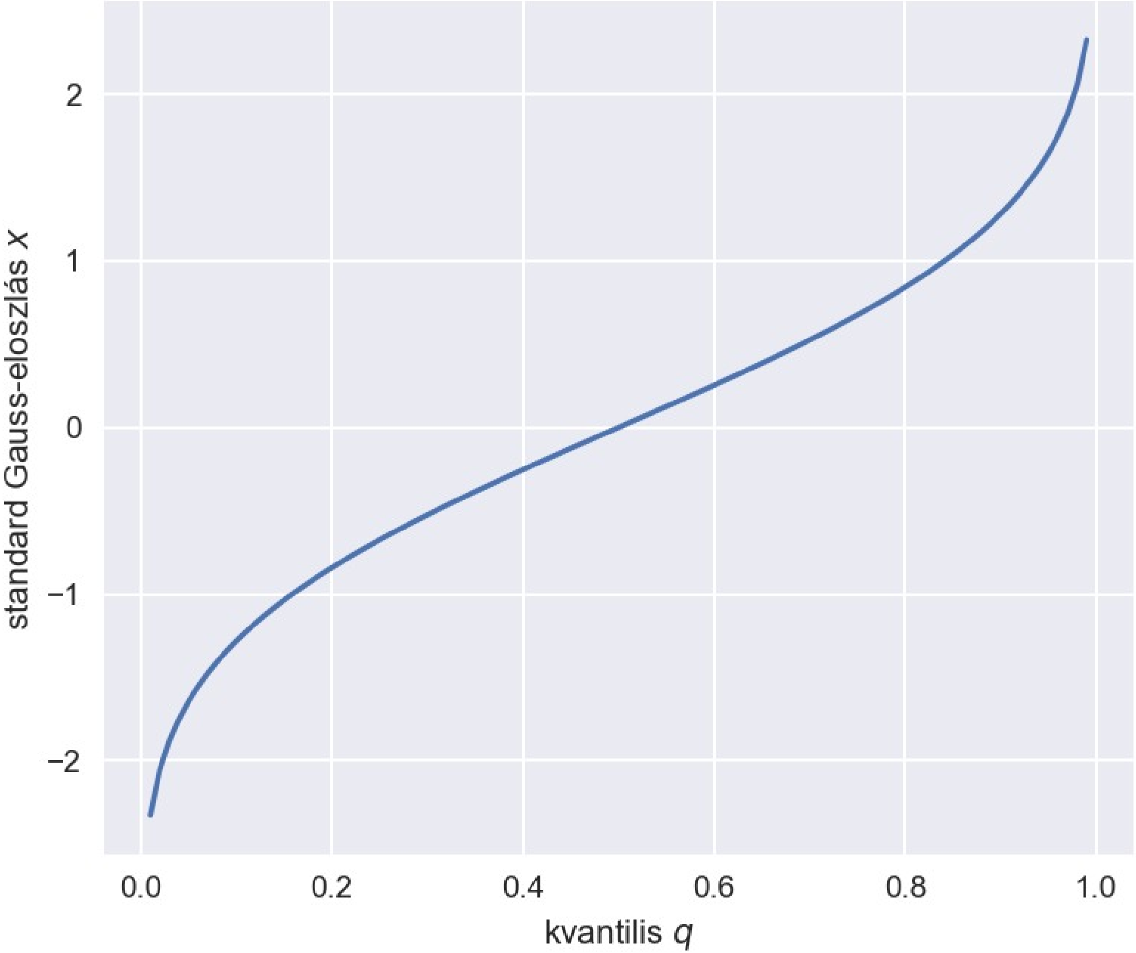
$$\sqrt{\frac{m \cdot n}{m+n} \cdot \max\left\{\left|F_n(x) - F_m(x)\right|\right\}}$$

Kvantilis-kvantilis ábra

- Minta $F_n(x)$ eloszlásának összehasonlítása valamilyen *elméleti* $F(x)$ eloszlásfüggvénnyel
- Két, n és m elemű mintából meghatározott *tapasztalati* eloszlásfüggvény összehasonlítása: $F_n(x)$ és $F_m(x)$
- Kvantilis ($F(x)$ eloszlásfüggvény $F^{-1}(q)$ inverze)

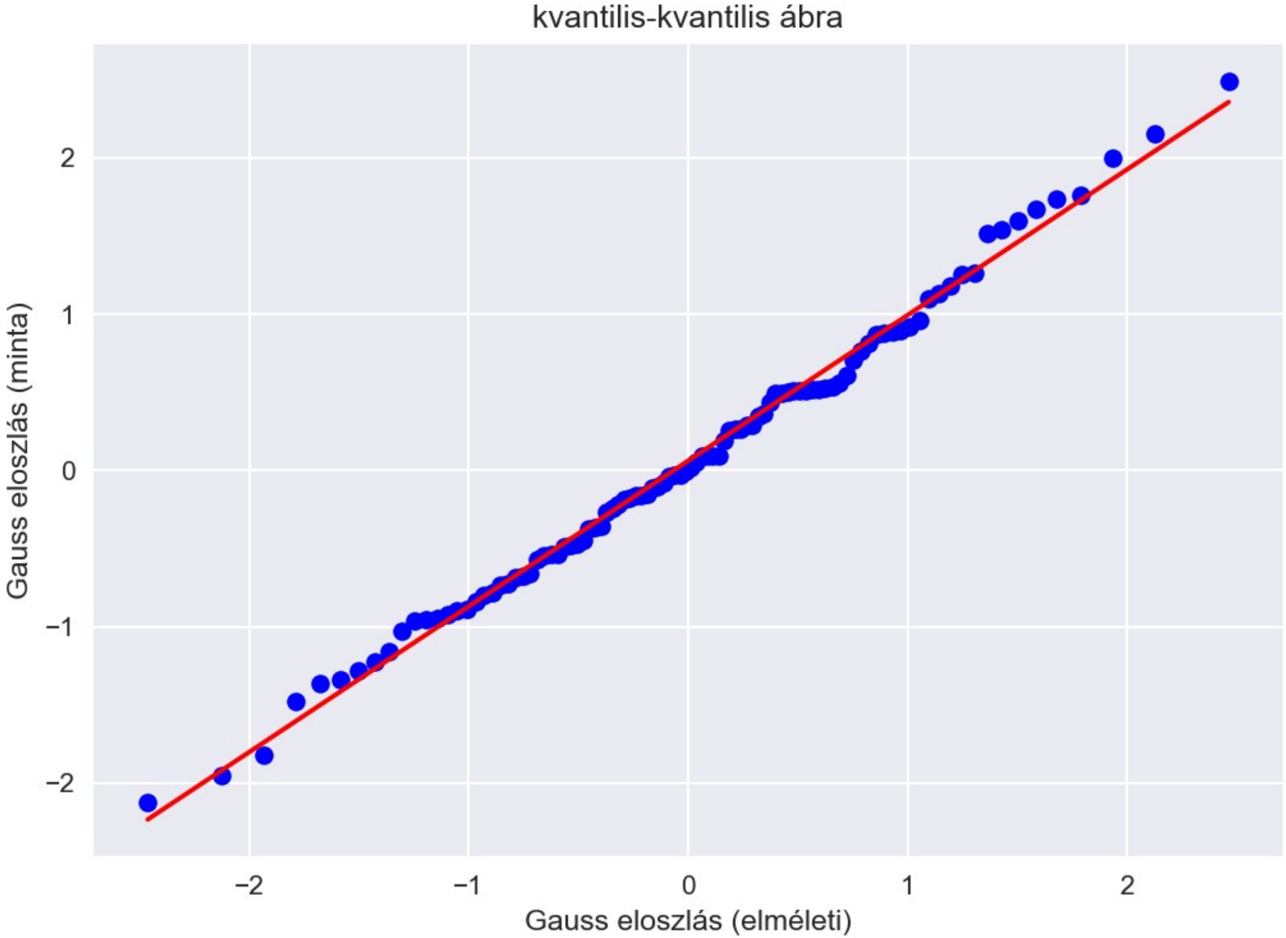
  Ha a sorba rendezett sokaságot egy x érték q:$(1 - q)$ arányban osztja ketté $(0 < q < 1)$, akkor ezt az x értéket q-ad rendű vagy q-adik kvantilisnek nevezzük

Gauss-eloszlás kvantilise

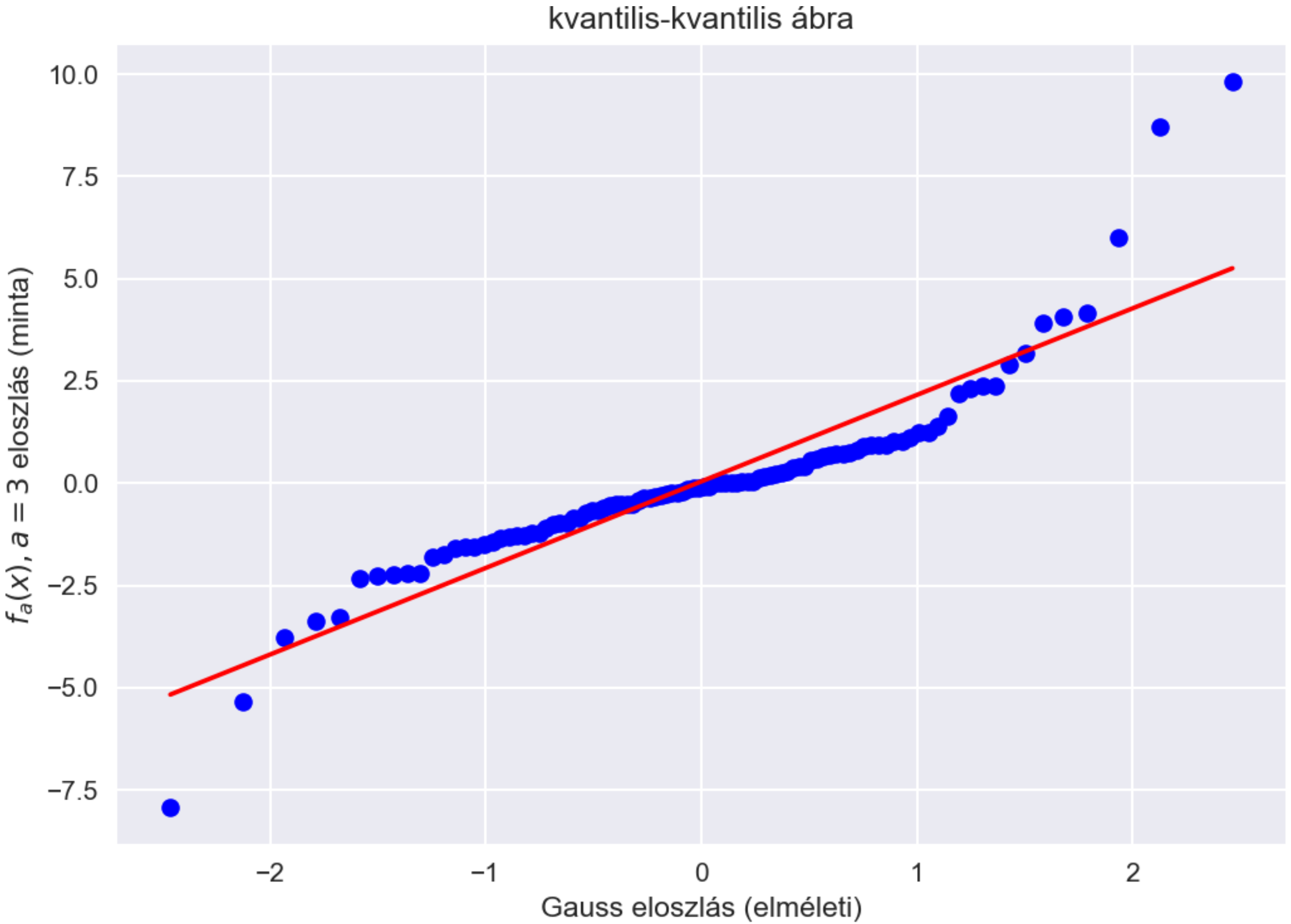


Gauss-eloszlású minta

Python:
scipy.stats.probplot()



$f_3(x)$ eloszlású minta



Tananyag

- Steiner (1990): 1-3, 6. fejezetek
- Detrekői (1991): 3.5, 3.6 alfejezetek
- Vincze (1975): 5. rész