

BEVEZETÉS

A statisztika teljesen laikusoknak: nagy munkával gyűjtött *adatok vizsgálata*, abból *következtetések levonása* („statistical inference”)

(Egy kicsit sok hűhó semmiért – azaz *Much ado about nothing*.)

Mi is a statisztika?

- ① Egy *populációból* veszünk *mintát*. (A *szavakat* a KSH találta ki.)
- ② A minta alapján akarunk valamit mondani, de *az egész populációról*.
- ③ Állítsunk *megbízhatóságáról* is nyilatkozunk. → NÉPSZAVAZÁS
- ④ A „mintavétel” nem akármilyen. Akárhányszor elvégezzük, *más és más eredményt kapunk*.

Ez a „mintavétel” lesz a dolog kulcsa. → VENEREAL DISEASE



Ezért kell érteni a valószínűségszámításhoz.

Nevezzük a mintavételt **kísérletnek**.

Kísérlet: → *determinisztikus* : előre meghatározható eredményhez vezet

→ *véletlen* : statisztikai törvényeknek engedelmeskedik

(Mi az ami közös a népszavazásban, a betegségek gyógyulásában és a fiz. kém. laborban?)

Mi kell a statisztika tanulásához?

MATEMATIKA:

halmazelmélet

algebra

mértékelmélet (differenciál- és integrálszámítás)

analízis

Példa: NÉPSZAVAZÁS (Belépjen-e az Egyesült Királyság az Európai Unióba?)

	YES	NO	SUM
Scotland	1 332 186	947 769	2 279 355
Northern Ireland	259 251	237 311	497 162

Kérdés: Van-e különbség Scotland és Northern Ireland véleménye között?

Válasz: Annak a valószínűsége, hogy nincs, 10^{-8} .

MIK A VÉLETLEN TÖRVÉNYEI ?

Definíció: **Eseménytér:** a véletlen kísérlet *összes* lehetséges „kimenetelének” halmaza.

Elemi: az egyes kísérletek *kimenetelei*.

Az eseménytér lehet: – korlátos folytonos: pl. testmagasság

– végtelen diszkrét: pl. radioaktív bomlás

– véges diszkrét: pl. látósejtek száma a retinán,

kockadobás,

urna

(MI A BAJ A KLASSZIKUS ELMÉLETTEL?? (Kombinatorika))

– végtelen folytonos: ha így *definiáljuk!*

– egyváltozós

– többváltozós

Definíció: **Esemény:** Az eseménytér tetszőleges részhalmaza.

Elnevezés: *Bekövetkezik* egy esemény, ha a kísérlet olyan kimenetele fordul elő, amelynek valódi része az esemény.

HF. Hány lehetséges esemény van *egy* kocka dobásánál (és kettőnél)?

Egy kocka: *ábra*

\emptyset : az üres halmaz (hogyan az eseménytér zárt legyen, ne vezessen ki belőle semmilyen művelet.)

Definíció: **Diszjunkt (egymást kizáró) események:**

Ha (tetszőleges párra) nincsen páronként közös részük. (A metszetük üres.)

Példák: Páratlan / páros kocka – 2 vagy kisebb / 2-nél nagyobb

A VALÓSZÍNŰSÉGSZÁMÍTÁS AXIÓMÁI

Legyen: A és B egy eseménytér két (diszjunkt) eseménye (azaz $A \cap B = \emptyset$).

Jelölés: $P(A)$ az A, $P(B)$ a B esemény valószínűségeit jelölő *számok*, ha teljesül 3 axióma.

1. $0 \leq P(A)$ $\|$ $P(B)$ -re természetesen ugyanez igaz

2. $P(A \cup B) = P(A) + P(B)$

3. $P(S) = 1$

S: A teljes eseménytér

Milyen esemény $A \cup B$???

Ennyi axióma elég.

$$\text{Szokás még: } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

$$\text{vagy } 0 \leq P(A) \leq 1$$

– de ezek már az előzőek következményei!

Néhány fontos következmény: **valószínűség számítás tételek**

0. $P(\emptyset) = 0$

1. $P(A) \leq 1$

2. $P(\bar{A}) \leq 1 - P(A)$ || Hány eseményt specifikál egy kísérlet kimenetele? (\bar{A} az A komplementere.)

3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ – kiterjesztés több (páronként független) eseményre

4. A 2. axióma következménye: események különbségének valószínűsége

$$P(A/B) = P(A) - P(A \cap B)$$

$$(\text{Ha } B \subseteq A \rightarrow P(A/B) = P(A) - P(B) \quad || \text{ Milyen esemény az } A/B?)$$

5. Ha két esemény nem diszjunkt, felbontható három diszjunkt eseményre. **Legyen** $D \cap E \neq \emptyset$

Felbontás: $D \cap E, D / (D \cap E), E / (D \cap E) \rightarrow$ uniójuk: $D \cup E$

$$P(D \cup E) = P(D / (D \cap E)) + P(E / (D \cap E)) + P(D \cap E) =$$

$$= P(D) + P(E) - P(D \cap E)$$

Vegyük észre: ha D és E diszjunktak, visszakapjuk a 2. axiómát.

Kiterjeszthetjük több eseményre \rightarrow POINCARÉ tétele.

Mit jelent $A \subseteq B$? (Ha B, akkor A is.)

Ekkor: $P(A) \leq P(B)$

$$P(B/A) = P(B) - P(A)$$

Hogy állunk $P(A/B)$ -vel?

FELTÉTELES VALÓSZÍNŰSÉG

Jelölés: $A|B$: A, feltéve, hogy B bekövetkezett.

Definíció: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ „az A esemény B-re vonatkoztatott **feltételes valószínűsége.**”

Tétel: A és B események függetlenek, ha $P(A \cap B) = P(A) \cdot P(B)$

Bizonyítás: $\frac{P(A \cap B)}{P(A)} = P(B) \rightarrow P(B|A) = P(B)$ (a B esemény valószínűsége független A-tól.)

Szimmetria okokból $\rightarrow P(A|B) = P(A)$ ■

A valószínűség gyakorlati értelmezése:

- Tapasztalati gyakoriság
- Klasszikus valószínűség (egyenletes, diszkrét)
- Geometriai valószínűség

Definíció: **Függetlenek** egymástól azok a **kísérletek**, amelyek kimeneteleinek valószínűségét nem befolyásolják a többi kísérletek kimenetelei.

Elnevezés: **Ismétlés:** ha az újabb kísérletek függetlenek a korábbiaktól.

Bernoulli tétele (sztochasztikus konvergencia):

$$p_{A,n} = \frac{h_A}{n} \quad \text{tapasztalati gyakoriság}$$

$$\lim_{n \rightarrow \infty} P(|P(A) - p_{A,n}| < \varepsilon) = 1 \quad \text{tetszőleges } \varepsilon\text{-ra}$$

(Az axiómák „igazolhatók” a $p_{A,n} = \frac{h_A}{n}$ tapasztalati gyakoriságokból –

– nem kell hozzá az egyenletes valószínűség.)

VALÓSZÍNŰSÉGI VÁLTOZÓK

nehéz fogalom!

Elnevezés: **Egyszerű (elemi)** események: *diszkrét* eseménytér elemei.

Folytonos eseménytérben: $X \in (x + \Delta x)$

MÉRÉSKOR NINCS FOLYTONOS ESEMÉNYTÉR!

Majdnem lehetetlen esemény

Majdnem biztos esemény

$\Delta x \rightarrow 0$ lehetetlensége:

két ember }
két molekula } távolsága

Definíció:

A **valószínűségi változó** az eseménytérre értelmezett függvény. A kísérlet minden egyes kimenetelének megfelelően felvesz egy értéket, ez az ő **realizációja**.
Értékkészlete alkotja a valószínűségi változó **eseményterét**.
Más neve: **statisztika**.

Változó: **NAGY** latin betű, realizáció: **kis** latin betű

Mit jelent $P(X = x)$? Hogy van ez egy kocka dobásánál??

Mi a folytonos megfelelője a $P(X = x)$ -nek??

$$P(x < X \leq x + \Delta x)$$

vagy, ha elvégezhető a $\Delta x \rightarrow 0$ átmenet: $P(x < X \leq x + dx)$

Mi a $\Delta x \rightarrow 0$ feltétele???

- N. B.** – Valószínűségi változók bármely függvénye is valószínűségi változó! (Miért?)
– Bármely függvény, amely érvényes valószínűségi változók között,
érvényes ugyanúgy a realizációk között is. (Miért?)

VALÓSZÍNŰSÉGI SŰRŰSÉGFÜGGVÉNY

Legyen X egy valószínűségi változó,
 S az ő eseménytere.

Kérdés: Hogyan oszlanak el S fölött a valószínűségek?

Definíció: Ha X folytonos valószínűségi változó, akkor **valószínűségi sűrűségfüggvénye** az az $f(x)$ függvény, amelynek az A intervallumon vett integrálja megadja annak a valószínűségét, hogy X realizációi az A intervallumon belül lesznek, azaz:

$$1. P(X \in A) \stackrel{jel}{=} P(A) = \int_A f(x) dx$$

2. Az $X \in (x, x + dx)$ elemi esemény valószínűsége $f(x) dx$, és $f(x) dx \geq 0$, $\forall x$

$$3. \int_S f(x) dx = 1$$

$A(-\infty, \infty)$ -beli definíció esetén: $\int_{-\infty}^{\infty} f(x) dx = 1$ – Hogy lehet ezt így kiterjeszteni?

Definíció: Ha X diszkrét valószínűségi változó, akkor minden egyes x értéke (realizációja) elemi esemény, $p(x)$ valószínűséggel.

Ekkor a $p(x) = P(X = x)$ az X **valószínűségi sűrűségfüggvénye**.

Erre igaz

1. $P(x \in A) \stackrel{\text{jelölés}}{=} P(A) = \sum_{x \in A} p(x)$
2. $0 \leq p(x) \leq 1$
3. $\sum_{x \in A} p(x) = 1$

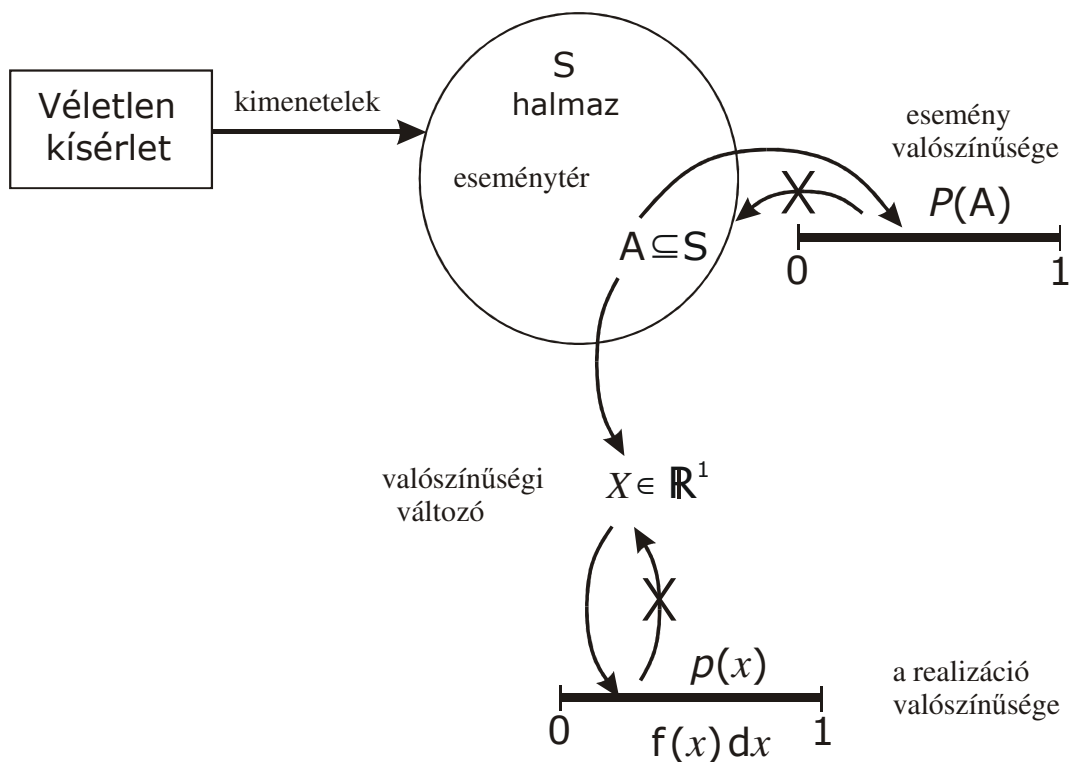
Analógia: Tömegpontok / kontinuum mechanikája

$$\sum_{test} m_i = \int_{test} \rho dV$$

$$\sum_{test} f(m_i) = \int_{test} f(\rho) dV \quad \rho : \text{tömegsűrűség}$$

(Stieltjes integrál)

Mostanra épült fel teljesen a használható matematikai apparátus:



(A matematikus nem az S halmazt tekinti alapként,

hanem annak **összes részhalmazából** álló H halmazt!)

Definíció: Az Y valószínűségi változó **eloszlásfüggvénye**:

$$F(x) = P(y \leq x)$$

$$F(x) = \sum_{y < x} p(y) \quad \text{diszkrét}$$

$$F\left(x = \int_{-\infty}^x f(y) dy\right) \quad \text{folytonos}$$

Fogalmak áttekintése \ eloszlás típusa	folytonos	diszkrét
sűrűségfüggvény	$f(x)$	$p(x)$
elemi esemény valószínűsége	$f(x) dx$	$p(x)$
adott A esemény valószínűsége	$\int_A f(x) dx$	$\sum_A p(x)$
eloszlásfüggvény	$F(x)$	$F(x)$
$P(X \leq x)$	$F(x)$	$F(x)$
$P(x_1 \leq X \leq x_2)$	$F(x_2) - F(x_1)$ $\int_{x_1}^{x_2} f(x) dx$	$F(x_2) - F(x_1)$ $\sum_{X=x_1}^{X=x_2} p(x)$

Vegyük észre!

$$\lim_{a \rightarrow b} P(a < X \leq b) = 0$$

folytonos X-re

$$P(x = b) = 0$$

$\forall b$ **majdnem lehetetlen esemény**

$$p(x \neq b) = 0$$

majdnem biztos esemény

VÁRHATÓ ÉRTÉK

Definíció: X valószínűségi változó **bármely** $g(x)$ függvényének várható értéke:

$$M(g(x)) = \begin{cases} \int_{-\infty}^{\infty} g(x) f(x) dx & \text{– folytonos} \\ \sum_{x \in S} g(x) p(x) & \text{– diszkrét} \end{cases}$$

(Stieltjes integrállal: $M(g(x)) = \int_0^1 g(x) dF(x)$)

Feltételek: Ha a $\sum g(x) p(x)$ sor *konvergens*.

vagy a $\int_{-\infty}^{\infty} g(x) f(x) dx$ integrál *létezik* és *véges*.

Speciális várható értékek:

X várható értéke (X átlaga, X eloszlásának középpértéke)

$$\mu_x = \mu = M(x) = \begin{cases} \int_{-\infty}^{\infty} x f(x) dx \\ \sum_{x \in S} x p(x) \end{cases}$$

Jelentése: ezt „szórják körül” a kísérlet eredményei.

M : mean (más jelölés: E : expectation)

X (eloszlásának) *r-edik centrális momentuma*

$$\mu_r = M[(x - \mu)^r] = M[(x - M(x))^r]$$

N.B.: Ha az eloszlás *szimmetrikus*, minden *páratlan* centrális momentuma *zérus*.

ábra

2. centrális momentum: X (eloszlásának) *szórásnégyzete / varianciája*

$$D^2(x) = \sigma^2 = V(x) = \mu_2 = M[(x - \mu)^2] = M[(x - M(x))^2]$$

Elnevezés: Standard deviáció (hiba): $\sigma = \sqrt{D^2(x)}$

D : deviation

σ : scatter

Két valószínűségi változó esetén: KOVARIANCIA

$$C(X, Y) = M[(x - \mu_x)(Y - \mu_r)]$$

Vegyük észre a határesetet: $C(X, X) = D^2(X) = V(X)$ (szórásnégyzet, variancia)

Kovariancia mátrix: elemei: $C(X_i, X_j)$

főátló: $V(X_i)$ (= variancia)

Belőle származik a *korrelációs együttható*:

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{D^2(x) \cdot D^2(Y)}} \quad \text{„normált kovariancia”}$$

Tétel: Ha X és Y **függetlenek** $M(XY) = M(X) \cdot M(Y)$

akkor $C(X, Y) = 0$ és $\rho(X, Y) = 0$

MEGFORDÍTVÁ CSAK AKKOR IGAZ, ha X és Y együttes eloszlása normális.

Tétel: Minden nemnegatív $f(x)$, ha integrálható a $(-\infty, \infty)$ intervallumon, és

$$\int_{-\infty}^{\infty} f(x) dx = 1, \text{ valószínűségi sűrűségfüggvény lehet.}$$

Ha $\int_{-\infty}^{\infty} g(x) dx \neq 1$, de véges, akkor

$$\frac{1}{N} g(x) \text{ is lehet sűrűségfüggvény, ahol } N = \int_{-\infty}^{\infty} g(x) dx$$

N : NORMA

ELOSZLÁSFÜGGVÉNY TÍPUSOK

Binomális eloszlás Legyen: tetszőlegesen ismételhető kísérlet két kimenetellel: A és \bar{A}

$$P(A) = p \quad P(\bar{A}) = q = 1-p$$

Binomiális mintavétel

Legyen n **ismétlésből** K az A események száma

$$S = \{0, 1, 2, \dots, n\} \quad k \in S$$

Definíció: $P(K = k) = \binom{n}{k} p^k q^{n-k}$ ez a sűrűségfüggvény

Jelölés: $K \sim B(p, n)$

A **név** eredete: $P(K = k)$ kifejezés a $(p + q)^n$ **binomális sorból való**.

$$\mu = n p$$

$$\sigma^2 = n p q = n p (1 - p)$$

Más név: Bernoulli-eloszlás
ismételt alternatívák eloszlása

Alakalmazás: Népszavazás, feleletválasztás, stb.....

Poisson eloszlás

Diszkrét Gyakran használható.

Időben: egyenletes valószínűséggel bekövetkező események *száma adott időintervallumban*.

Térben: egyenletes valószínűséggel bekövetkező események (véletlen elhelyezkedése) *száma adott felületen*. (Esőcsepp, radioaktív bomlás, gépelési hiba, LÓRÚGÁS, forgalom, gólok focimeccsen, telefonhívások, sejtziporodás, születések száma)

Eseménytér: \mathbf{N}

Jelölés: $K \sim Pn(m)$

Definíció $P(K = k) = P(b) = m^k \frac{e^{-m}}{k!} \quad k \in \mathbf{N}$

$\mu = m$ $\sigma^2 = m$ $\sigma = \sqrt{m}$

Tétel: c -szeres intervallum: $K \sim Pn(c \cdot m)$

ha $K_1 \sim Pn(m_1)$ és $K_2 \sim Pn(m_2)$ függetlenek, akkor $K_1 + K_2 \sim Pn(m_1 + m_2)$

Határeloszlás-tételek:

$B(p, n) \rightarrow Pn(np)$ ha $\frac{p}{n} \ll 1$

$B(p, n) \rightarrow Pn(np)$ ha $\lim_{n \rightarrow \infty} np = m$ (azaz, ha n nő, p csökken)

Exponenciális eloszlás

Folytonos **Időben:** (egyenletes eloszlású) véletlen események bekövetkezésének idejéig eltelt idő – ÉLETTARTAM-eloszlás

Térben: (egyenletes eloszlású) véletlen események helyének távolsága egy adott (tetszőleges) helytől

Várakozás!, ütközések távolsága /ideje, élettartam.

REAKCIÓKINETIKA!

$f(x) \begin{cases} = a \cdot e^{-ax}, & \text{ha } x \geq 0 \\ = 0, & \text{ha } x < 0 \end{cases} \parallel a > 0$

$F(x) = 1 - e^{-ax}$

$\mu = \frac{1}{a}$

– átlagos élettartam,
ütközési gyakoriság,
szabad úthossz,
relaxációs idő

$$\sigma^2 = \frac{1}{a^2}$$

$$\sigma = \frac{1}{a}$$

A Poisson rokona! – POISSON-folyamat

Normális eloszlás

Felfedezője: Abraham de Moivre – ezért hívják még Gauss-eloszlásnak.

Pétevári játék: Addig dobunk, míg fej nem jön ki. Ha n -edikre dobunk fejet, 2^n rubelt kapunk. Mennyit kell befizetni a banknak, hogy ne menjen tönkre?

Dobások: $Bn(0.5, n)$

$$\text{de Moivre: } \lim_{n \rightarrow \infty} P(|h_{n,fej} - h_{n,írás}| < x\sqrt{n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{n^2}{2}} dn$$

Definíció

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

Jelölés: $X \sim N(\mu, \sigma)$

Definíció a $Z = \frac{X - \mu}{\sigma}$ **STANDARD NORMÁLIS** eloszlású

$Z \sim N(0, 1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$x = z \cdot \sigma + \mu$ – táblázatok, belső függvények

Határeloszlások

(matematikusok-fizikusok)

Központi határeloszlás tétele

Legyenek x_1, x_2, \dots, x_n azonos eloszlású valószínűségi változók, μ és σ^2 (véges) paraméterekkel, akkor $n \rightarrow \infty$ esetén a $\sum_{i=1}^n x_i \sim N(n\mu, n\sigma^2)$, továbbá

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} : \quad \bar{x} \sim \lim_{n \rightarrow \infty} N\left(\mu, \frac{\sigma^2}{n}\right), \text{ amiből } \sigma(\bar{x}, n) = \frac{\sigma}{\sqrt{n}}$$

Mérések!!

χ^2 eloszlás

$$f(x) = \frac{x^{\left(\frac{\nu}{2}-1\right)} \cdot e^{-\frac{x}{2}}}{\Gamma\left(\frac{\nu}{2}\right) \cdot 2^{\frac{\nu}{2}}} \quad 0 \leq x < \infty, \quad \nu > 0$$

$$x \sim \chi^2_\nu \quad \nu \text{ a szabadsági fokok száma}$$

Miért fontos? Ha $x_1, x_2, x_3, \dots, x_n$ függetlenek és $N(\mu, \sigma^2)$ eloszlásúak:

$$W = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \quad W \sim \chi^2_n$$

Várható értéke: $\mu = \nu$

Mérések!

Elnevezés: $\frac{W}{n} \sim \frac{\chi^2_m}{n}$ redukált χ^2 - eloszlás: $\mu = 1$

Student-féle t -eloszlás (Student: angol úr álneve, ezen a néven írta matematikai cikkeit)

Kivétel: t kis betű, de valószínűségi változó!!

$$f(t) = \frac{1}{\sqrt{\nu} \cdot \beta\left(\frac{1}{2}, \frac{\nu}{2}\right) \cdot \left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} \quad -\infty < t < \infty, \nu > 0$$

$$\beta\left(\frac{1}{2}, \frac{\nu}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right) \cdot \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{1}{2} + \frac{\nu}{2}\right)}$$

ábra

$$\lim_{\nu \rightarrow \infty} t_\nu = N(0, 1) \quad \nu \approx 30 \text{ fölött az eltérés kisebb mint } 10 \%$$

Jelentőség: mintavétel – ld. később

Ha $Z \sim N(0, 1)$ és $U \sim \chi^2_\nu$ függetlenek, akkor $T = \frac{Z}{\sqrt{\frac{U}{\nu}}} \sim t_\nu$

F-eloszlás (Fisher-féle F -eloszlás)

$f(x) =$ igen bonyolult

Ha $U \sim \chi^2_{\nu_1}$ és $V \sim \chi^2_{\nu_2}$ függetlenek, akkor redukált hányadosaik eloszlása ilyen:

$$\frac{\frac{\mu}{\nu_1}}{\frac{\nu}{\nu_2}} \sim F_{\nu_1, \nu_2} \quad \text{és} \quad F_{\nu_1, \nu_2} = \frac{1}{F_{\nu_2, \nu_1}}$$

Számolás: $F_{\nu_1, \infty} = \frac{\chi^2_{\nu_1}}{\nu_1}$

STATISZTIKAI MÓDSZEREK

Mintavétel: \underline{x} ($x_1, x_2, x_3, \dots, x_n$) elemek kiválasztása a sokaságból \rightarrow minta

Becslés: $T = f(\underline{x})$ statisztikák számítása

mintastatisztika T függ a mintától !!

Statisztikai analízis:

konfidencia	}	vizsgálatok
szingifikancia		
hipotézis		
modell		
illeszkedés		

Szükség van T eloszlásának ismeretére!!

(Az x eloszlás ismeretére nem mindig: NEMPARAMÉTERES ROBUSZTUS módszerek)

$P(T \leq t)$		típusú valószínűségek számítása
$P(T \geq t)$		
$P(t_1 \leq T \leq t_2)$		

Mintavétel – külön tudomány (pl. kísérlettervezés)

Idealizált: n ismétlés: mérések

x_1, x_2, \dots, x_n azonos eloszlású kimentelek

$T(x_1, \dots, x_n)^{\text{jel}} = T(\underline{x})$ a megfigyelések valamely függvénye: *mintastatisztika*

$T(\underline{x})$ eloszlása a **minta eloszlása**, amely az x_i -k eloszlásától függ.

Konkrét példák

A minta középértéke: $T(\underline{x}) = \frac{\sum_{i=1}^n x_i}{n}$

Jelölés: \bar{x} – eloszlása általában nem ismert!

ha $x_i \sim N(\mu, \sigma^2)$, akkor $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$M(\bar{x}) = M(x) = \mu$ μ **torzítatlan becslése**

$D^2(\bar{x}) = \frac{\sigma^2}{n}$ ezért a **mint**a középértéke

N.B. n növelésével csak $\frac{1}{\sqrt{n}}$ -szeresére csökken a szórás!

A minta szórásnégyzete

Definíció: S^2 = S^2(x) = sum_{i=1}^n (x_i - x_bar)^2 / (n-1)

Ha x_i ~ N(mu, sigma^2), for all i,

akkor (x_bar - mu) / (S / sqrt(n)) ~ t_{n-1} számlálóról: Z-szerű, nevező: redukált chi^2-szerű

Számolás:

S^2 = (sum x_i^2 - (sum x_i)^2 / n) / (n-1)

M(S^2) = sigma^2

sigma^2 torzítatlan becslése S^2 a minta szórásnégyzete

A minta kovarianciája:

C-hat(X, Y) = sum_{i=1}^n (x_i - x_bar)(y_i - y_bar) / (n-1)

M(C-hat(X, Y)) = C(X, Y) a kovariancia torzítatlan becslése

BECSLÉS

A minta T statisztikáját úgy választjuk meg (no meg a mintát!), hogy az eloszlás theta paraméteréhez közel legyen.

(Szovjet mondás: A hazugságnak három fokozata van: 1. hazugság 2. arcátlan hazugság 3. statisztika)

A magyar nyelv sem kutya:

Az eljárás: becslés (estimation) - becslési eljárás

A T valószínűségi változó: becslés (estimator) - becselő függvény

T egy v-hat realizációja: becslés (estimate) - becsült érték

N. B. T egy valószínűségi változó. Realizációja a konkrét mintától függ. Általában exists eloszlása, várható értéke, szórása.

Egy jó „becslő”

1. torzítatlan

M(T) = theta

2. hatásos

(„minimum variancia”)

3. elégséges

- ha a T(x) minden szükséges információt tartalmaz theta-ról. (A hatásos becslés elégséges!!)

4. konzisztens

ha lim_{n to infinity} P(|T - theta| < epsilon) = 1

Feltétel: ha T torzítatlan, és lim_{n to infinity} D^2(T) = 0

5. kongruens

M(f(t)) = f(M(t))

Torzítatlan hatásos becslés: Minimum Variance Unbiased = MVU

Módszerek

Maximum likelihood (ML)	–MVU, elégséges, konzisztens
Legkisebb négyzetes	– azonos normális eloszlású mintaelemek esetén maximum likelihood
Momentumok módszere	– nem foglalkozunk vele
Minimax	– ezzel sem

$$\hat{\mu} = \bar{x} \quad \text{ML becslés (MVU)}$$

$$\hat{\sigma}^2 = S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{ML becslés (MVU)}$$

$$\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad \text{aszimptotikusan hatásos, konzisztens.}$$

$$D^2(\bar{x}) = \frac{\sigma^2}{n} \quad \text{a becslt várható érték relatív hibája: } \boxed{\frac{1}{\sqrt{n}}}$$

(egy realizáció σ standard hibájához viszonyítva)

HIBATERJEDÉS

Legyen $\theta_1, \theta_2, \dots, \theta_r$ fizikai mennyiségek ϕ függvénye a becslendő

① Becsüljük az egyedi θ_i -ket és szórásukat

② Ebből becsüljük $\phi(\theta)$ -t és $D^2(\phi(\theta))$ -t

Legyen a becsülő függvény: $\phi(T_1, T_2, \dots, T_r)$

Fejtsük sorba θ körül! (Taylor-sor)

$$\phi(T_1, T_2, \dots, T_r) = \phi(\vartheta_1, \vartheta_2, \dots, \vartheta_r) + \sum_{i=1}^r \frac{\partial \phi}{\partial \vartheta_i} (T_i - \vartheta_i) + \dots \text{(magasabb rendű tagok)}$$

① Ha $D^2(\theta_i)$ kicsi θ_i -hez képest, akkor $T_i - \theta_i$ is kicsi. Így elegendő a

$(T_i - \theta_i)$ elsőfokú tagok figyelembevételével, a $(T_i - \theta_i)^2$ már elhanyagolható. (**Közelítés!**)

② Tegyük fel: T_i torzítatlan becslő $\Rightarrow M(T_i - \theta_i) = 0$

$$\Rightarrow M(\phi(T_1, T_2, \dots, T_r)) = \phi(\theta_1, \theta_2, \dots, \theta_r)$$

így ϕ becslése is torzítatlan. **Ez nem mindig közelítés!**

A becsülő statisztika szórásnégyzete:

$$D^2[\phi(T_1, T_2, \dots, T_r)] = M\left\{[\phi(T_1, T_2, \dots, T_r) - \phi(\vartheta_1, \vartheta_2, \dots, \vartheta_r)]^2\right\}$$

A Taylor-sorból a jobb oldalon []-ben lévő különbség éppen $\sum_{i=1}^r \frac{\partial \phi}{\partial \vartheta_i} (T_i - \vartheta_i)$:

$$D^2[\phi(T_1, T_2, \dots, T_r)] \cong M\left\{\left[\sum_{i=1}^r \frac{\partial \phi}{\partial \vartheta_i} (T_i - \vartheta_i)\right]^2\right\}$$

A jobb oldal egy r tagú összeg négyzete, amely kifejtve:

$$\sum_{i=1}^r \left(\frac{\partial \phi}{\partial \vartheta_i} \right)^2 \cdot D^2(T_i) + 2 \sum_{i < j}^r \sum_{j=1}^r \frac{\partial \phi}{\partial \vartheta_i} \frac{\partial \phi}{\partial \vartheta_j} C(T_i - T_j)$$

Függvények várható értékének és szórásának becslése:

- ① $\vartheta_i^* = t_i$ a T_i statisztikák realizációja
- ② $\phi(\theta_1, \theta_2, \dots, \theta_r)$ becslése $\phi(t_1, t_2, \dots, t_r)$
- ③ $D^2(T_i)$ becslése $S^2(T_i)$
 $C(T_i, T_j)$ becslése $\hat{C}(T_i, T_j)$
- ④ $D^2(\phi)$ becslése:

$$S^2[\phi(T_1, T_2, \dots, T_r)] = \sum_{i=1}^r \left(\frac{\partial \phi}{\partial \vartheta_i} \right)_{t_i}^2 \cdot S^2(T_i) + 2 \sum_{j=1}^r \sum_{i < j} \left(\frac{\partial \phi}{\partial \vartheta_i} \right)_{t_i} \left(\frac{\partial \phi}{\partial \vartheta_j} \right)_{t_j} \hat{C}(T_i, T_j)$$

ha T_i -k páronként függetlenek, ez a tag zérus!

- ⑤ $S^2(\phi)$ szabadsági fokainak száma közelítőleg:

$$\bar{\nu} = \frac{s^4(\phi)}{\sum_i \left(\frac{\partial \phi}{\partial \vartheta_i} \right)^4 \cdot S^4(T_i) \cdot \frac{1}{\nu_i}} \quad \boxed{\nu_i = n_i - 1}$$

KONFIDENCIA INTERVALLUMOK

A T becslő függvény (mint valószínűségi változó) θ -hoz való közelségének mértéke:

$$P[T \in (\vartheta - \delta, \vartheta + \delta)]$$

Baj van! θ -t nem ismerjük! (Ha ismernénk, nem becsülnénk!)

Ekvivalens megfogalmazás:

$$\begin{array}{ccccccc}
 P[T \in (\vartheta - \delta, \vartheta + \delta)] & = & P[\vartheta \in (T - \delta, T + \delta)] \\
 \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 \text{valószínűségi változó} & & \underbrace{\hspace{2cm}} & \text{konstans intervallum} & & \underbrace{\hspace{2cm}} & \text{konstans az intervallum} \\
 & & & & & & \text{a valószínűségi változó!}
 \end{array}$$

Konfidencia intervallum: a $(T - \delta, T + \delta)$ intervallum **REALIZÁCIÓJA**

Konfidencia valószínűség: $P[T \in (\vartheta \pm \delta)]$

$$P[\vartheta - \delta \leq T \leq \vartheta + \delta] = 1 - \alpha$$

$1 - \alpha$: megbízhatósági szint

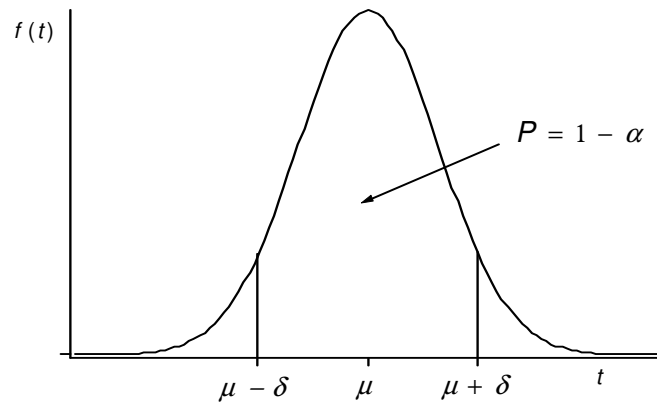
α : szignifikancia szint

Ha (l_1, l_2) a konfidenciaintervallum, mekkora a $P[\vartheta \in (l_1, l_2)]$ valószínűség??

Válasz: 0 vagy 1!

Ezért MEGBÍZHATÓSÁG

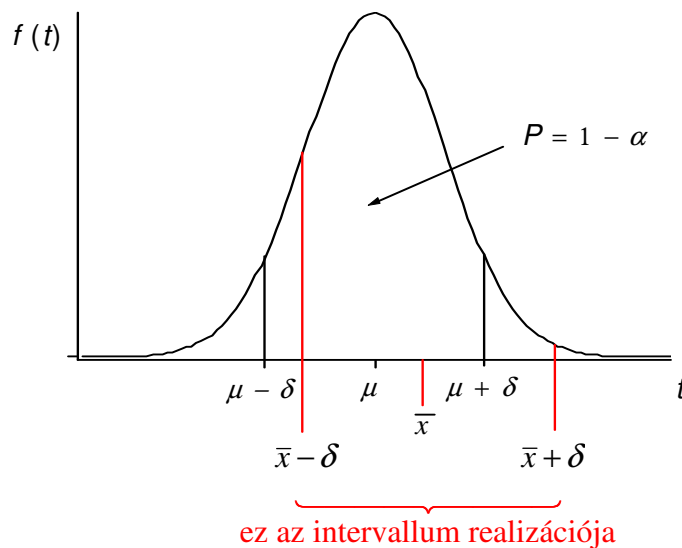
A konfidencia-intervallum számítása



Példák

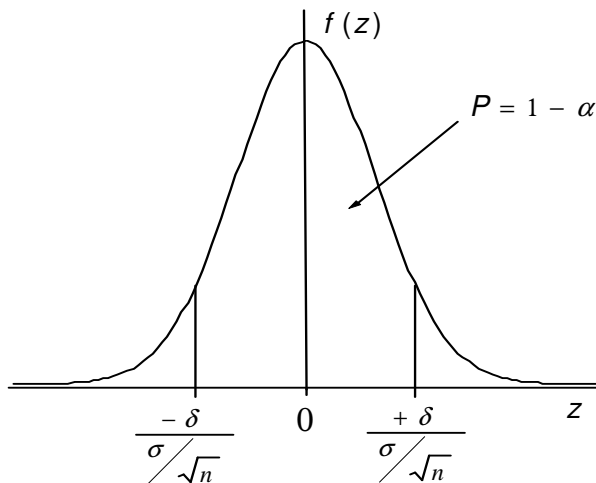
σ^2 ismert, $\hat{\mu} = \bar{x}$ $x \sim N(\mu, \sigma^2)$

$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, de μ nem ismert



Legyen: $Y = \bar{x} - \mu \rightarrow Y \sim N\left(0, \frac{\sigma^2}{n}\right)$

még jobb: $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow Z \sim N(0,1)$ – ezt a legkönnyebb számítani is.



Konkrét számítás:

$$P(\mu - \delta \leq \bar{x} \leq \mu + \delta) = \int_{\mu - \delta}^{\mu + \delta} f(\bar{x}) d\bar{x} = F(\mu + \delta) - F(\mu - \delta) = 1 - \alpha$$

Használjuk ki a standard normális transzformációt (vegyük észre: ekkor **eltűnik a μ** – ez volt a cél):

$$1 - \alpha = \int_{-\frac{\delta}{\frac{\sigma}{\sqrt{n}}}}^{\frac{\delta}{\frac{\sigma}{\sqrt{n}}}} f(z) dz = F\left(\frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right) - F\left(-\frac{\delta}{\frac{\sigma}{\sqrt{n}}}\right)$$

Miért jobb $N(0,1)$?

1. Egyszerűbb
2. Könyvtári szubrutinok ezt számolják
3. Táblázatokban ez szerepel

(Manapság már nincs jelentősége; a számítógép $N(0, \sigma^2/n)$ -t is tudna számolni.)

Eljárás:

1. θ becslése $\hat{\vartheta} = t$
2. T eloszlásának meghatározása
3. T célszerű transzformációja
4. $1 - \alpha = P(\vartheta - \delta \leq T \leq \vartheta + \delta)$ valószínűség **kiszámítható** megfogalmazása (a **kiszámítható** azt jelenti, ne szerepeljen benne az ismeretlen θ)
ez eddig általában kész recept (a matematikusok már megcsinálták)
5. δ meghatározása az *adott mintára* ez a feladat

HIPOTÉZIS VIZSGÁLATOK

(VIZSGÁLAT = TESZT)

Nullhipotézis – alternatív hipotézis

H_0

H_1

Lényeg: Rögzítsünk egy α szignifikanciaszintet,

ami egy igaz H_0 elvetése valószínűségének felső határa.

Példák: $H_0: \vartheta = \vartheta_0$ $\left\{ \begin{array}{l} H_1: \vartheta \neq \vartheta_0 \\ H_1': \vartheta < \vartheta_0 \\ H_1'': \vartheta > \vartheta_0 \end{array} \right. \begin{array}{l} \text{kétoldali} \\ \text{egyoldali} \\ \text{egyoldali} \end{array} \text{ alternatív hipotézis}$

Lehet: $H_0: \theta \leq \theta_0$ vagy $\theta \geq \theta_0$ is.

H_0 -t megvédjük vagy elvetjük α szignifikanciaszinten ($\alpha: 0,1; 0,05; 0,01$)

A döntés alapja az

$$\alpha = P(T \geq c | H_0) \quad \text{vagy} \quad \alpha = P(|T| \geq c | H_0)$$

Elvetjük H_0 -t, ha $t \geq c$ vagy $|t| \geq c$

c : kritikus érték

Honnan tudjuk 1. c értékét?

2. Az α valószínűséget?

Válasz: Ismerjük (vagy azt hisszük, hogy ismerjük!!) T eloszlását,

és abból kiszámíthatjuk fix α -hoz a c -t.

Ezt a c -t hasonlítjuk a mintából számított t realizációhoz.

Az α szignifikancia-szint értéke a minta elemeinek számától (is) függ.

Ha $t \cong c$ vagy $|t| \cong c$, érdemes új mintát (több adatot) vizsgálni.

Ha $t \ll c$ vagy $|t| \ll c$, lehet kisebb az α .

Kiszámítható közvetlenül az $\alpha = P(T \geq c | H_0)$ szignifikancia-valószínűség is!

Ekvivalencia a konfidencia-intervallummal (kétoldali alternatív hipotézis esetén):

$$\text{ELFOGADJUK } H_0\text{-t, ha } \vartheta_0 \in (l_1, l_2), \text{ elvetjük, ha } \vartheta_0 \notin (l_1, l_2).$$

N. B. 1. Olyan statisztikai teszt *nincs*, amely mindig elveti H_0 -t, ha hamis,

és mindig elfogadja, ha igaz. (Bécsi nyelvfilozófusok.)

2. Mekkora legyen α ?? Ha $1 - \alpha = 0$, akkor soha nem ítélnék el ártatlant,

de mindig felmentjük a tettet, ha gyanúsított.

Ha $1 - \alpha \gg 0$, akkor a bűnöst elítéljük, ha a gyanúsítottak között

van, de ha a gyanúsított ártatlan, akkor is kénytelenek vagyunk

elítélni, a bűnös pedig a markába nevet.

KOCKÁZATI FÜGGVÉNYEK

1. fajú hiba: Az igaz H_0 elvetése / (a csalfa H_1 elfogadása)

2. fajú hiba: A hamis H_0 elfogadása / (az igaz H_1 elvetése)

A **VENEREAL DISEASES** példában: használ a gyógyszer: 5-5 % szignifikancia

Σ nem használ: 5 % szignifikancia

Legyen K a javult esetek száma

Tegyük fel: $K \sim B(p, n)$

$$H_0: p_{\text{kezelt}} > p_{\text{nem kezelt}}$$

$$H_1: p_{\text{kezelt}} = p_{\text{nem kezelt}}$$

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

$$X_0 : t \geq z(1 - \alpha)$$

A VARIANCIA-ANALÍZIS (ANOVA) is *hipotézis-teszt* (ANOVA = Analysis Of VAriance)

$$H_0: x_{ij} = \mu + \nu_j + \varepsilon_{ij}$$

$$H_1: x_{ij} = \mu + \mu_i + \nu_j + \varepsilon_{ij}$$

μ : alaphatás

μ_i : Az i -edik „kezelés” hatása (pl. adag mennyisége)

ν_j : A j -edik blokk hatása (pl. életkor, nemek)

ε_{ij} : A mérési hiba. ($M(\varepsilon_{ij}) = 0$, $D^2(\varepsilon_{ij}) = \sigma^2$)

MIRE HASZNÁLJUK MI A BECSLÉSEKET

– ÁLTALÁBAN

– pl. A REAKCIÓKINETIKÁBAN

- ① Független-vagy-nem-független kérdések eldöntésére (hipotézisvizsgálatok)
- ② Az összefüggés módjának eldöntésére (függvényillesztések)
- ① Adott eltérések oka lehet-e a véletlen ingadozás, vagy szisztematikus függésről van-e szó??
 - Mekkora a valószínűsége egy adott eltérésnek?? (szignifikancia valószínűség)
 - Kiszór-e egy pont, vagy szabad neki akkorát ingadozni??
(Utóbbiak csúnya, bonyolult, megbízhatatlan tesztek.)

HOGYAN ADJUNK MEG EGY BECSÜLT EREDMÉNYT?

① n darab mérés átlagolása esetén $\rightarrow \bar{x}, s^2(x)$

\bar{x} a μ paraméter becslt értéke, $s^2(x)$ a σ^2 paraméter becslt értéke

$s(\bar{x}) = \frac{s(x)}{\sqrt{n}}$ a becslt \bar{x} szórása μ körül. (Innen marad benne az $\frac{1}{\sqrt{n}}$.)

Mi az $s(\bar{x})$ információtartalma?

Legyen $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$; $T \sim t_{n-1}$ ($n - 1$ szabadsági fokú Student-eloszlás)

Ekkor: $P[t_{n-1}(\alpha/2) \leq T \leq t_{n-1}(1 - \alpha/2)] = 1 - \alpha$

Az $1 - \alpha$ megbízhatósági szintű konfidencia-intervallum:

$$\bar{x} \pm t_{n-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}, \text{ átrendezve:}$$

$$\bar{x} \pm s \cdot \frac{t_{n-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}$$

n	2	3	5	10	20	30	40	60	120	1000
$\frac{t_{n-1}\left(1 - \frac{\alpha}{2}\right)}{\sqrt{n}}$	9,0	2,5	1,15	0,70	0,47	0,37	0,32	0,26	0,18	0,06

\downarrow 20 x mérésszám, 10 x pontosság
 \uparrow 100 x mérésszám, 10 x pontosság

$\alpha = 0,05$

95 %-os konfidencia-intervallumok

② (Feltételezett!!) függvény (= modell) paramétereinek becslése esetén

(Maga az **eredmény megadása** a 24. oldal végén található.)

Statisztikai modell:

$$Y = f(x) + \varepsilon$$

– pl. REAKCIÓMECHANIZMUS

(realizációk: $y_i = f(x_i) + \varepsilon_i$)

Y : valószínűségi változó

$f(x)$: determinisztikus függvény

ε : valószínűségi változó: $M(\varepsilon) = 0 \quad \forall x$

$$D^2(\varepsilon_i) = \sigma_i^2 \text{ ha } x = x_i$$

(Határeset, amit szeretünk *feltételezni*: $\sigma_i^2 \equiv \sigma^2, \quad \forall i$)

\Rightarrow Csak VÉLETLEN hiba esetén használható!! (Egyébként pl. ANOVA!)

Cél: *nem* ε eloszlásának jellemzése, azok paramétereivel,

hanem az $f(x)$ **modellfüggvény paramétereinek becslése**, lehetőleg **MVU!**

Vegyük észre: ez egy feltételes valószínűség!

$P(y_i | x_i)$ – ez adja a fenti modellt.

A leggyakrabban használt becslő módszer: legkisebb négyzetes

$$Q = \sum_{i=1}^n w_i (Y_i - f(x_i))^2 \quad \text{legyen minimális}$$

w_i súlyok számítása – hogy az $f(x)$ paramétereire MVU-becslést kapjunk.

$$\text{feltétel: } M(\hat{\vartheta}) = \vartheta \\ \text{és } D^2(\hat{\vartheta}) \text{ minimális}$$

Példa: Legyen $Y = \alpha \cdot x$ – modelfüggvény

Statisztikai modell: $Y = \alpha \cdot x + \varepsilon$ – $D^2(\varepsilon) = \sigma^2$

Minta: $\{x_1, y_1, \dots, x_n, y_n\}$

$$y_i = \alpha \cdot x_i + \varepsilon \quad - D^2(\varepsilon_i) = \sigma_i^2$$

Legyen A az α becslő függvénye: $y_i = A \cdot x_i + \varepsilon$ (A : mintastatisztika)

$$Q = \sum_{i=1}^n w_i (y_i - A x_i)^2$$

Feladat I. A minimalizálja a Q -t: $\frac{\partial Q}{\partial A} = 0$

$$\frac{\partial Q}{\partial A} = -\sum_i 2w_i (y_i - A x_i) x_i = 0 \\ -2 \sum_i (w_i x_i y_i - A w_i x_i^2) = 0$$

$$\sum_i w_i x_i y_i = A \sum_i w_i x_i^2$$

$$A = \frac{\sum_i w_i x_i y_i}{\sum_i w_i x_i^2}$$

Feladat II. A láthatóan nem csak az $\{x_i, y_i\}$ mintától, hanem a w_i súlyoktól is függ.

Eredeti feltételünk: legyen $\sigma^2(A)$ minimális

$$\sigma^2(A) = \sum_i \left(\frac{\partial A}{\partial x_i} \right)^2 \sigma^2(x_i) + \sum_i \left(\frac{\partial A}{\partial y_i} \right)^2 \sigma^2(y_i)$$

① Tegyük fel: $C(X_i, Y_i) = 0$ (Mi a feltétele???)

② Tegyük fel: $\sigma^2(y_i) \gg \sigma^2(x_i), \forall i$ (KÖVETKEZMÉNY!!!)

Nem mindegy,
mit illesztünk minek a függvényében!!

$$\frac{\partial A}{\partial y_i} = \frac{w_i x_i}{\sum_i w_i x_i^2}$$

$$\sigma^2(A) = \frac{\sum_i (w_i^2 x_i^2 \sigma^2(y_i))}{\left(\sum_i w_i x_i^2\right)^2} \quad \left\| \quad \text{minimumfeltétel: } \frac{\partial \sigma^2(A)}{\partial w_j} = 0 \quad , \quad \forall j \right.$$

↓ rendezzük:

$$\sigma^2(A) \cdot \left(\sum_i w_i x_i^2\right)^2 = \sum_i w_i^2 x_i^2 \sigma^2(y_i) \quad \left\| \quad \frac{\partial}{\partial w_j}$$

$$\left(2\sigma^2(A) \cdot \sum_i w_i x_i^2\right) x_j^2 = 2w_j x_j^2 \sigma^2(y_j)$$

$$\boxed{w_j = \frac{\sigma^2(A) \sum_i w_i x_i^2}{\sigma^2(y_j)}} \quad - \text{ ez a jó súly}$$

A $\frac{\partial Q}{\partial A} = 0$ egyenlet elosztható $\sigma^2(A) \cdot \sum_i w_i x_i^2$ -tel: $\frac{w_j}{w_k} = \frac{\sigma^2(y_k)}{\sigma^2(y_j)}$

Tehát elegendő a $\boxed{w_i = \frac{1}{\sigma^2(y_i)}}$ választás.

A feladat megoldása:

$$\boxed{A = \frac{\sum_{i=1}^n \frac{1}{\sigma^2(y_i)} x_i y_i}{\sum_{i=1}^n \frac{1}{\sigma^2(y_i)} x_i^2}} \quad \text{MVU becslés}$$

egyszerűbb írásmóddal:

$$\boxed{A = \frac{\sum_i w_i x_i y_i}{\sum_i w_i x_i^2}}$$

$$\boxed{S^2(A) = \frac{\sum_i w_i^2 x_i^2 \cdot S^2(y_i)}{\left(\sum_i w_i x_i^2\right)^2}} \quad w_i = \frac{1}{\sigma^2(y_i)}$$

Megjegyzések: 1. $S^2(y_i)$ becsülhető az adatokból.

2. Ha $w_i = w_j$, azaz $\sigma^2(y_i) \equiv \sigma^2(y_j) = \sigma^2 \quad \forall i, j$,
akkor írható $w = 1 \equiv w_i$.

Ilyen esetben:

(„súlyozatlan becslés”:) $A = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$ és $S^2(A) = \sum_i \left(\frac{x_i}{\sum_i x_i^2} \right)^2 \cdot S^2(y) = \frac{\sigma^2}{\sum_i x_i^2}$

Mi a helyzet, ha nem igaz $\sigma^2(y) \gg \sigma^2(x)$?

Ekkor $\frac{1}{w_i} = \sigma^2(y_i) + \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma^2(x_i)$

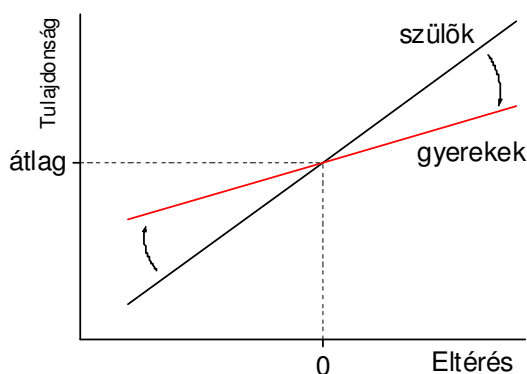
Mivel $\frac{\partial f}{\partial x_i} = A \Rightarrow \frac{1}{w_i}$ függvénye A -nak!

Következmény: Az A **becsült értéke függ w_i -ktől, a w_i -k pedig A -tól!**

Ilyenkor csak **iteratív módszerek használhatók!**
(„Implicit legkisebb négyzetes becslés”)

Egy elnevezés története: Regressziós analízis (regresszió!)

latin: *regressio* = visszafejlődés, visszatérés az egyszerűbb/régi formához



(Sir Francis Galton)

Regressziószámítás – itt alkalmazták (publikálva) először függvény (egyenes) paramétereinek becslésére a legkisebb négyzetes módszert. (Galton úr jó statisztikus volt.)

Azért én a **legkisebb négyzetes becslés** nevet jobban szeretem. Az nevével nevezi az eljárást.

(Nevezett Galton úrnak nagy szerepe volt a statisztikai módszerek széleskörű elterjedésében.)

Most térhetünk rá az **eredmény megadásának** problémájára:

HOGYAN ADJUK MEG A BECSÜLT PARAMÉTEREKET??

n db mérés \longrightarrow r paraméterre: $\vartheta_i, s^2(\vartheta_i)$

Mi az $s^2(\vartheta_i)$ információtartalma??

Legyen: $T_i = \frac{\hat{\vartheta}_i - \vartheta_i}{S(\vartheta_i)}$; $T \sim t_{n-r}$ (a T_i $n - r$ szabadsági fokú Student eloszlású.)

újfent: $P[t_{n-r}(\alpha/2) \leq T \leq t_{n-r}(1 - \alpha/2)] = 1 - \alpha$

Ennek alapján az $1 - \alpha$ megbízhatósági szintű konfidencia-intervallum:

$$\hat{\vartheta}_i \pm t_{n-r}(1 - \alpha/2) \cdot S(\vartheta_i)$$

n	1	2	3	5	10	20	40	100	1000
$t_{n-r}(1 - \alpha/2)$	12,7	4,3	3,2	2,6	2,23	2,09	2,02	1,98	1,96

$$\alpha = 0,05$$

95 %-os konfidencia-intervallumok

Mi a helyzet a súlyozással?

Milyen a mért x -ek és $f(x)$ -ek hibája?

1. Ha a kettő jelentősen eltér, **legyen x a kisebb, y a nagyobb hibájú.**
2. Ha egyik hibája a másikhoz képest nem elhanyagolható, akkor **implicit LSQ becslés** kell. SÚLYRUTIN

3. Ha a hibák azonosak: $\sigma_i \equiv \sigma_j \quad \forall i, j$

$$\frac{1}{\sigma_i} = \frac{1}{\sigma_j} \quad w_i = w_j \quad \frac{w_i}{w_j} = 1 \quad \text{SÚLYOZATLAN}$$

5. Ha a hibák azonosak, de transzformálunk: **a hibák a mért érték függvényei** lesznek (ld. hibaterjedés)

pl. relatív hiba $w_i = \frac{1}{y^2}$ Poisson-eloszlású minta $w_i = \frac{1}{y}$, stb.....

6. Ha a hibák nem azonosak: az MVU becsléshez **meg kell adni a hibákat is**: $w_i = 1/\sigma^2(y_i)$

7. Ha különböző súlyozású becslési eredményeket hasonlítunk össze, célszerű a $w_i \equiv 1$ -re ($\sum w_i = n$)-re normálás.

Ez meg itt a vége