$\boxed{\text{BEVEZETÉS}}$

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. $\boxed{\rightarrow \text{NÉPSZAVAZÁS}}$

④ A „mintavétel” nem akármilyen. Akárhányszor elvégezzük, *más és más eredményt kapunk*.

Ez a „mintavétel” lesz a dolog kulcsa. . $\boxed{\rightarrow \text{VENEREAL DISEASE}}$
↓
Ezért kell érteni a valószínűségszámításhoz.

Nevezzük a mintavételt $\boxed{\text{kísérletnek}}$.

$\boxed{\text{Kísérlet}}$: → *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?

$\boxed{\text{MATEMATIKA:}}$ halmazelmélet
algebra
mértékelmélet (differenciál- és integrálszámítás)
analízis

$\boxed{\text{Példa: NÉPSZAVAZÁS}}$ (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

$\boxed{\text{Kérdés:}}$ Van-e különbség Scotland és Northern Ireland véleménye között?

$\boxed{\text{Válasz:}}$ Annak a valószínűsége, hogy nincs, 10^{-8}.

MIK A VÉLETLEN TÖRVÉNYEI ?

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

Elemei: 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*

Ø: az üres halmaz (hogy 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 $A \cap B = 0$).

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

$\boxed{\text{Milyen esemény } A \cup B???}$

Ennyi axióma elég.

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

vagy $0 \le P(A) \le 1$ – de ezek már az előzőek következményei!

Néhány fontos következmény: ***valószínűség számítás tételek***

0. $P(\varnothing) = 0$

1. $P(A) \le 1$

2. $P(\overline{A}) \le 1 - P(A)$ ‖ Hány eseményt specifikál egy kísérlet kimenetele? ($\overline{A}$ 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

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

   (Ha $B \subseteq A \to P(A/B) = P(A) - P(B)$ ‖ Milyen esemény az A / B?

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

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

$$\boxed{P(D \cap E)} = P(D / (D \cap E)) + P(E / (D \cap E)) =$$
$$\boxed{= P(D) + P(E) - P(D \cap E)}$$

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

$\boxed{\text{Kiterjeszthetjük több eseményre}}$ $\to$ POINCARÉ tétele.

$\boxed{\text{Mit jelent } A \subseteq B?}$ (Ha B, akkor A is.)

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

$P(B / A) = P(B) - P(A)$ $\boxed{\text{Hogy állunk } P(A/B)\text{-vel?}}$

FELTÉTELES VALÓSZÍNŰSÉG

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

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

$\boxed{\text{Tétel:}}$ A és B események függetlenek, ha $P\,(\mathrm{A} \cap \mathrm{B}) = P\,(\mathrm{A}) \cdot P\,(\mathrm{B})$

$\boxed{\text{Bizonyítás:}}$ $\dfrac{P\,(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{A})} = P\,(\mathrm{B}) \rightarrow P\,(\mathrm{B}|\mathrm{A}) = P\,(\mathrm{B})$ (a B esemény valószínűsége független A-tól.)

Szimmetria okokból $\rightarrow P\,(\mathrm{A}|\mathrm{B}) = P\,(\mathrm{A})$ ■

A valószínűség gyakorlati értelmezése:

– Tapasztalati gyakoriság

– Klasszikus valószínűség (egyenletes, diszkrét)

– Geometriai valószínűség

$\boxed{\textit{Definíció:}}$ ***Függetlenek*** egymástól azok a ***kísérletek***, amelyek kimeneteleinek valószínűségét nem befolyásolják a többi kísérletek kimenetelei.

$\boxed{\textit{Elnevezés:}}$ ***Ismétlés***: ha az újabb kísérletek függetlenek a korábbiaktól.

Bernoulli tétele (sztochasztikus konvergencia):

$$p_{\mathrm{A},n} = \frac{h_{\mathrm{A}}}{n} \qquad \text{tapasztalati gyakoriság}$$

$$\lim_{n \to \infty} P\left(\left|P\,(\mathrm{A}) - p_{\mathrm{A},n}\right| < \varepsilon\right) = 1 \qquad \text{tetszőleges } \varepsilon\text{-ra}$$

(Az axiómák „igazolhatók” a $p_{\mathrm{A},n} = \dfrac{h_{\mathrm{A}}}{n}$ tapasztalati gyakoriságokból –

– nem kell hozzá az egyenletes valószínűség.)

| VALÓSZÍNŰSÉGI VÁLTOZÓK |

| nehéz fogalom! |

$\boxed{\textit{Elnevezés:}}$ ***Egyszerű*** (***elemi***) események: *diszkrét* eseménytér elemei.

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

MÉRÉSKOR NINCS FOLYTONOS ESEMÉNYTÉR!

Majdnem lehetetlen esemény	$\Delta x \rightarrow 0$ lehetetlensége:
Majdnem biztos esemény	$\left.\begin{array}{r}\text{két ember}\\ \text{két molekula}\end{array}\right\}$ távolsága

Definíció:

A ***valószínűségi változó*** *az eseménytéren 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??

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

vagy, ***ha elvégezhető a $\Delta x \to 0$ átmenet***: $P(x < X \le x + \mathrm{d}x)$

Mi a $\Delta x \to 0$ 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 $\mathrm{f}(x)$ függvény, amelynek az A intervallumon vett integrálja megadja annak a valószínűségét, hogy X realizációi az A intervallumon belül lesznek, azaz:

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

2. Az $X \in (X, x + \mathrm{d}x)$ elemi esemény valószínűsége $\mathrm{f}(x)\,\mathrm{d}x$, és $\mathrm{f}(x)\,\mathrm{d}x \ge 0$, $\forall x$

3. $\int_S \mathrm{f}(x)\, dx = 1$

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

$\boxed{\textit{\textbf{Definíció}}}$: 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. $P\left(x \in A\right) \overset{\text{jelölés}}{=} P\left(A\right) = \sum\limits_{x \in A} p\left(x\right)$
2. $0 \le p\,(x) \le 1$
3. $\sum\limits_{x \in A} p\left(x\right) = 1$

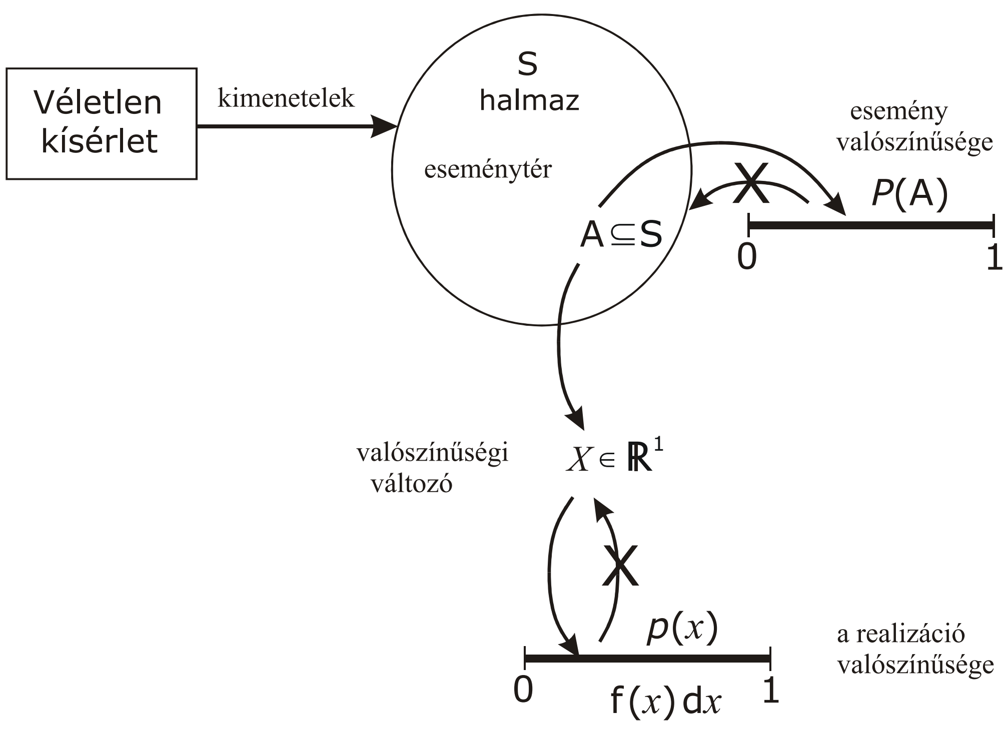
$\boxed{\text{Analógia:}}$ Tömegpontok / kontinuum mechanikája

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

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

(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***:

$\boxed{F(x) = P\,(y \le x)}$

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

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

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

Vegyük észre! $\lim\limits_{a\to b} P(a < X \le b) = 0$ $\boxed{\text{folytonos } X\text{-re}}$

$P(x = b) = 0$ $\forall\, b$ ***majdnem lehetetlen esemény***

$p\,(x \ne b) = 0$ ***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 középértéke)

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

Jelentése: ezt „szórják körül" a kísérlet eredményei.

M: mean (más jelölés: *E*: expectation)

X (eloszlásának) ***r-edik centrális momentuma***

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

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

ábra

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

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

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

D: deviation σ: scatter

Két valószínűségi változó esetén: KOVARIANCIA

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

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***:

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{D^2(x) \cdot D^2(Y)}}$$

„normált kovariancia"

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

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

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

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

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

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

$\frac{1}{N} g(x)$ 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 $\overline{A}$

$$P(A) = p \qquad P(\overline{A}) = q = 1 - p$$

Binomiális mintavétel

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

$$S = \{0, 1, 2, \ldots n\} \qquad k \in S$$

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***.

$\boxed{\mu = n\,p}$ $\boxed{\sigma^2 = n\,p\,q = n\,p(1 - p)}$

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, sejtszaporodás, születések száma)

Eseménytér: **N**

Jelölés: $K \sim Pn(m)$

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

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

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

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

Határeloszlás-tételek:

$$B(p,n) \to Pn(np) \quad \text{ha } {}^{p}\!/\!{}_{n} << 1$$

$$B(p,n) \to Pn(np) \quad \text{ha } \lim_{n\to\infty} np = m$$ (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!

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

$$\boxed{F(x) = 1 - e^{-ex}}$$

$$\boxed{\mu = \frac{1}{a}}$$ – átlagos élettartam,

ütközési gyakoriság,
szabad úthossz,
relaxációs idő

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

A Poisson rokona! – POISSON-folyamat

Normális eloszlás

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

***Pétervári 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: $Bn(0.5, n)$

de Moivre: $\lim\limits_{n\to\infty} P\left(\left|h_{n,fej} - h_{n,írás}\right| < x\sqrt{n}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{n^2}{2}}\, dn$

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

Jelölés: $X \sim N(\mu, \sigma)$

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

$Z \sim N(0, 1)$

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

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

Határeloszlások (matematikusok-fizikusok)

Központi határeloszlás tétele

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

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

Mérések!!

χ^2 eloszlás

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

$x \sim \chi^2_\nu$ $\qquad$ ν a ***szabadsági fokok száma***

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

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

Várható értéke: $\boxed{\mu = \nu}$

Mérések!

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

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

Kivételes: *t* kis betű, de valószínűségi változó!!

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

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

ábra

$\lim\limits_{\nu \to \infty} t_\nu = N(0, 1)$ $\qquad$ $\nu \approx 30$ fölött az eltérés kisebb mint 10 %

Jelentőség: mintavétel – ld. később

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

F-eloszlás $\qquad$ (Fisher-féle *F*-eloszlás)

$f(x)$ = igen bonyolult

Ha $U \sim \chi^2_\nu$ és $V \sim \chi^2_\nu$ függetlenek, akkor redukált hányadosaik eloszlása ilyen:

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

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

STATISZTIKAI MÓDSZEREK

Mintavétel: $\underline{x}$ $(x_1, x_2, x_3, \ldots x_n)$ elemek kiválasztása a sokaságból → $\boxed{\text{minta}}$

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

$\boxed{\text{mintastatisztika}}$ T függ a mintától !!

Statisztikai analízis:

$$\left.\begin{array}{l}\text{konfidencia}\\ \text{szingifikancia}\\ \text{hipotézis}\\ \text{modell}\\ \text{illeszkedés}\end{array}\right\}\ \text{vizsgálatok}$$

$\boxed{\text{Szükség van } T \text{ eloszlásának ismeretére!!}}$

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

A feladat leggyakrabban

$$\left.\begin{array}{l}P(T \le t)\\ P(T \ge t)\\ P(t_1 \le T \le t_2)\end{array}\right|\ \text{típusú valószínűségek számítása}$$

$\boxed{\text{Mintavétel}}$ – külön tudomány (pl. kísérlettervezés)

Idealizált: n ismétlés: mérések

$x_1, x_2, \ldots x_n$ azonos eloszlású kimentelek

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

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

Konkrét példák

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

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

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

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

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

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

A minta szórásnégyzete

Definíció: $S^2 = S^2(x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n-1}$

Ha $x_i \sim N(\mu, \sigma^2) \quad , \forall i$,

akkor $\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$ számláló: Z-szerű, nevező: redukált χ^2-szerű

Számolás: $$\boxed{S^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}}$$

$\boxed{M(S^2) = \sigma^2}$ σ^2 torzítatlan becslése
S^2 a ***minta szórásnégyzete***

A ***minta kovarianciája***:

$$\hat{C}(X, Y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$M(\hat{C}(X, Y)) = C(X, Y)$ a kovariancia torzítatlan becslése

BECSLÉS

A ***minta*** *T* statisztikáját úgy választjuk meg (no meg a mintát!),
hogy az ***eloszlás*** θ 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)	– *becslő függvény*
T egy $\hat{\vartheta}$ realizációja: ***becslés*** (estimate)	– *becsült érték*

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

Egy jó „becslő”

1. torzítatlan — $M(T) = \theta$
2. hatásos — („minimum variancia”)
3. elégséges — – ha a $T(\underline{x})$ minden szükséges információt tartalmaz θ-ról. (A hatásos becslés elégséges!!)
4. konzisztens — ha $\lim_{n\to\infty} P(|T - \vartheta| < \varepsilon) = 1$

   Feltétel: ha *T* torzítatlan, és $\lim_{n\to\infty} D^2(T) = 0$
5. kongruens — $M(f(t)) = f(M(t))$

Torzítatlan hatásos becslés: *Minimum Variance Unbaised* = 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

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

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

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

$D^2(\bar{x}) = \dfrac{\sigma^2}{n}$ a becsült várható érték relatív hibája: $\boxed{\dfrac{1}{\sqrt{n}}}$

(egy realizáció σ standard hibájához viszonyítva)

HIBATERJEDÉS

Legyen θ_1, θ_2, ...θ_r fizikai mennyiségek ϕ függvénye a becsülendő

① Becsüljük az egyedi θ_i -ket és szórásukat

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

Legyen a becslő függvény: $\phi(T_1, T_2, ...T_r)$

Fejtsük sorba $\boldsymbol{\theta}$ körül! (Taylor-sor)

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

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

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

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

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

A becslő statisztika szórásnégyzete:

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

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

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

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

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

Függvények várható értékének és szórásának becslése:

① $\vartheta_i^* = t_i$ a T_i statisztikák realizációja

② $\phi\ (\theta_1, \theta_2, ...\theta_r)$ becslése $\phi(t_1, t_2, ...t_r)$

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:

$$\boxed{S^2[\phi(T_1,T_2\ldots T_r)]=\sum_{i=1}^{r}\left(\frac{\partial\phi}{\partial\vartheta_i}\right)^2_{t_i}\cdot S^2(T_i)+\underbrace{2\sum_{j=1}^{r}\sum_{i<j}\left(\frac{\partial\phi}{\partial\vartheta_i}\right)_{t_i}\left(\frac{\partial\phi}{\partial\vartheta_j}\right)_{t_j}\hat{C}(T_i,T_j)}_{\text{ha } T_i\text{-k páronként függetlenek, ez a tag zérus!.}}}$$

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

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

KONFIDENCIA INTERVALLUMOK

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

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

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

***Ekvivalens megfogalmazás*:**

$$P\left[\underset{\substack{\downarrow\\ \text{valószínűségi változó}}}{T}\in\underbrace{(\vartheta-\delta,\ \vartheta+\delta)}_{\text{konstans intervallum}}\right]=P\left[\underset{\substack{\downarrow\\ \text{konstans}}}{\vartheta}\in\underbrace{(T-\delta,\ T+\delta)}_{\substack{\text{az intervallum}\\ \text{a valószínűségi változó!}}}\right]$$

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

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

$$P\left[\vartheta-\delta\le T\le\vartheta+\delta\right]=1-\alpha$$

$1-\alpha$: megbízhatósági szint

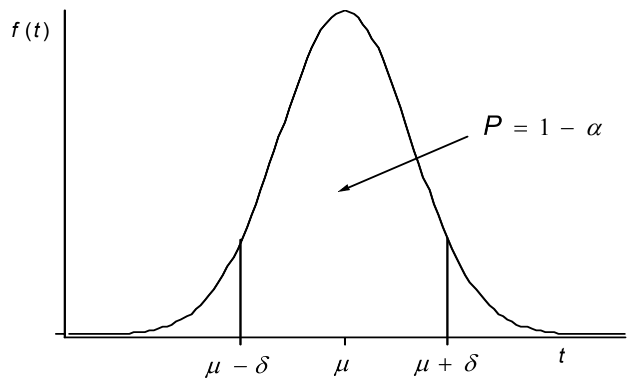
α: szingifikancia szint

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

Válasz: 0 vagy 1!

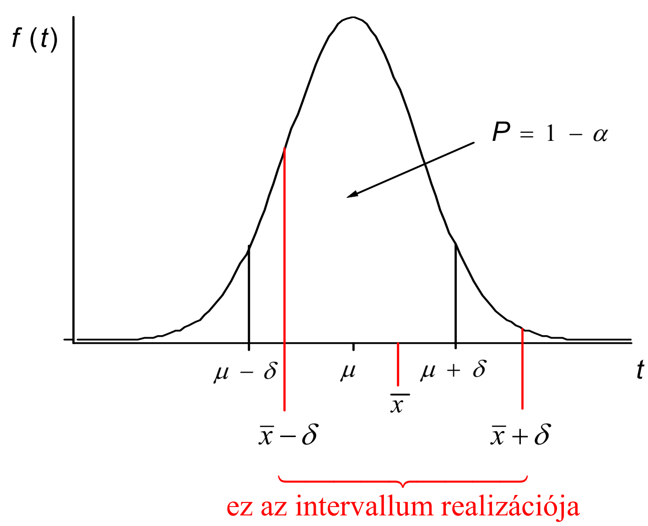
Ezért $\boxed{\text{MEGBÍZHATÓSÁG}}$

A konfidencia-intervallum számítása



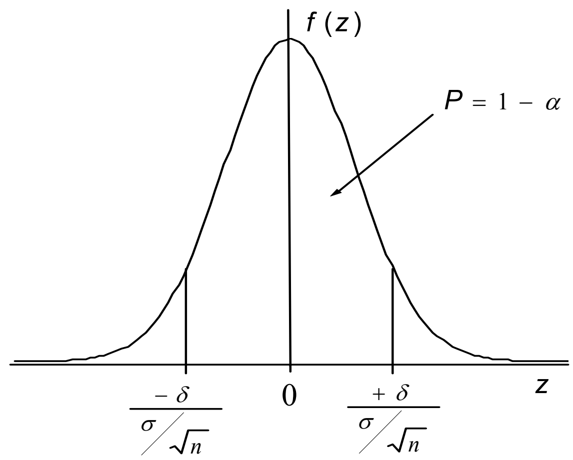
$\boxed{\textit{Példák}}$ σ^2 ismert, $\hat{\mu} = \bar{x}$ $x \sim N(\mu, \sigma^2)$

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



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

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



$\boxed{\text{Konkrét számítás:}}$

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

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

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

Miért jobb N(0,1)?

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

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

$\boxed{\text{Eljárás:}}$

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

$\boxed{\text{HIPOTÉZIS VIZSGÁLATOK}}$ (VIZSGÁLAT = TESZT)

Nullhipotézis – alternatív hipotézis

H_0 H_1

Lényeg: Rögzítsünk egy α szignifikanciaszintet,

ami egy igaz H_0 elvetése valószínűségének felső határa.

$\boxed{\textit{Példák}:}$ $\qquad H_0: \quad \vartheta = \vartheta_0 \qquad \left\{ \begin{array}{ll} H_1: \quad \vartheta \neq \vartheta_0 & \text{| kétoldali} \\ H_1^{'}: \quad \vartheta < \vartheta_0 & \text{| egyoldali} \\ H_1^{''}: \quad \vartheta > \vartheta_0 & \text{| egyoldali} \end{array} \right| \text{ alternatív hipotézis}$

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

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

A döntés alapja az

$$\alpha = P\left(T \geq c \,\middle|\, H_0\right) \qquad \text{vagy} \qquad \alpha = P\left(|T| \geq c \,\middle|\, H_0\right)$$

Elvetjük H_0-t, ha $\qquad t \geq c \qquad$ vagy $\qquad |t| \geq 0$

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 $|t| \cong 0$, érdemes új mintát (több adatot) vizsgálni.

Ha $t \ll$ c vagy $|t| \ll 0$, lehet kisebb az α.

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

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

ELFOGADJUK H_0-t, ha $\vartheta_0 \in (l_1, l_2)$, 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 ítélünk el ártatlant,

de mindig felmentjük a tettest, 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: $p_{\text{kezelt}} = p_{\text{nem kezelt}}$

$$\boxed{T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}} \qquad \cancel{H}_0 : \quad t \geq z(1-\alpha)$$

A VARIANCIA-ANALÍZIS (ANOVA) is *hipotézis-teszt* (ANOVA = **A**nalysis **O**f **VA**riance)

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 Függ-vagy-nem-függ-tőle 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 becsült értéke, $s^2(x)$ a σ^2 paraméter becsült értéke

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

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

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

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

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

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

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

20 x *mérésszám, 10* x *pontosság*

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

100 x *mérésszám, 10* x *pontosság*

$\boxed{\alpha = 0{,}05}$ 95 %-os konfidencia-intervallumok

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

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

Statisztikai modell: $\boxed{Y = f(x) + \varepsilon}$ – pl. REAKCIÓMECHANIZMUS

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

Y: valószínűségi változó

$f(x)$: determinisztikus függvény

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

$D^2(\varepsilon_i) = \sigma_i^2$ 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!)

$\boxed{\text{Cél:}}$ ***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: $\boxed{\text{legkisebb négyzetes}}$

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

$\boxed{w_i \text{ súlyok számítása}}$ – hogy az $f(x)$ paramétereire MVU-becslést kapjunk.

feltétel: $M(\hat{\vartheta}) = \vartheta$

és $D^2(\hat{\vartheta})$ minimális

$\boxed{\textit{Példa:}}$ Legyen $\boxed{Y = \alpha \cdot x}$ – modellfüggvény

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

Minta: $\{x_1, y_1, \ldots, x_i, y_i, \ldots, x_n, y_n\}$

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

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

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

$\boxed{\text{Feladat I.}}$ A minimalizálja a Q-t:

$$\frac{\partial Q}{\partial A} = 0$$

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

$$-2\sum_i \left(w_i\ x_i\ y_i - A\ w_i\ x_i^2\right) = 0$$

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

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

$\boxed{\text{Feladat II.}}$ 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: $C(X, Y) = 0$ (Mi a feltétele???)

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

Nem mindegy,
mit illesztünk minek a függvényében!!

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

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

↓ rendezzük:

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

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

$$\boxed{w_j = \frac{\sigma^2(A) \sum_i w_i\ x_i^2}{\sigma^2(y_j)}}$$ – 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:

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

egyszerűbb írásmóddal:

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

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

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

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

Ilyen esetben:

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

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

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

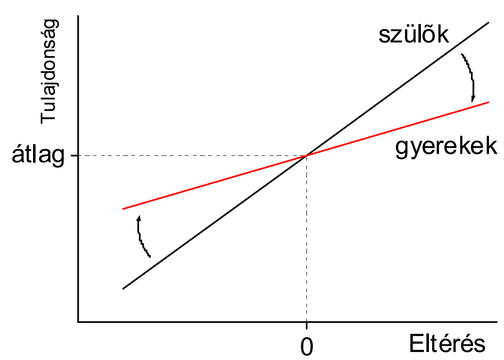
Mivel $\dfrac{\partial f}{\partial x_i} = A \quad \Rightarrow \quad \dfrac{1}{w_i}$ *függvénye A-nak!*

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 nevén 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: $\vartheta_i, \quad s^2(\vartheta_i)$

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

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

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

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

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

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

$\boxed{\alpha = 0{,}05}$ $\qquad$ 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. $\qquad$ $\boxed{\text{SÚLYRUTIN}}$

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

$\dfrac{1}{\sigma_i} = \dfrac{1}{\sigma_j} \qquad w_i = w_j \qquad \dfrac{w_i}{w_j} = 1$ $\qquad$ $\boxed{\text{SÚLYOZATLAN}}$

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

pl. relatív hiba $\quad w_i = \dfrac{1}{y^2}$ $\qquad$ Poisson-eloszlású minta $\quad w_i = \dfrac{1}{y}$, 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 $\left(\sum w_i = n\right)$-re normálás.

Ez meg itt a vége