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.

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. $0 \le P(A)$ ‖ $P(B)$-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. $P(\emptyset) = 0$
1. $P(A) \le 1$
2. $P(\overline{A}) = 1 - P(A)$ ‖ 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 $B \subseteq A \rightarrow P(A/B) = P(A) - P(B)$ ‖ Milyen esemény az A/B?
5. Ha két esemény nem diszjunkt, felbontható három diszjunkt eseményre. ***Legyen*** $D \cap E \ne \emptyset$

   Felbontás: $D \cap E$, $D / (D \cap E)$, $E / (D \cap E)$ → 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.

$\boxed{\text{Kiterjeszthetjük több eseményre}}$ → 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***.”

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

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

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):

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

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

(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!

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

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

MÉRÉSKOR NINCS FOLYTONOS ESEMÉNYTÉR!

Majdnem lehetetlen esemény	$\Delta x \rightarrow 0$ lehetetlensége:
Majdnem biztos esemény	két ember két molekula } távolsága

$\boxed{\textit{Definíció}}$:

$\boxed{\text{A } \textbf{\textit{valószínűségi változó}}}$ *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

$\boxed{\text{Mit jelent } P\,(X = x)?}$ Hogy van ez egy kocka dobásánál??

Mi a folytonos megfelelője a $\boxed{P\,(X = x)}$-nek??

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

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

$\boxed{\text{Mi a } \Delta x \to 0 \text{ feltétele???}}$

$\boxed{\text{N. B.}}$ – 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?

$\boxed{\textit{Definíció}}$: 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 $X \in (x, x + dx)$ elemi esemény valószínűsége $\mathrm{f}\,(x)\,dx$, és $\mathrm{f}\,(x)\,dx \ge 0$, $\forall x$

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

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

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

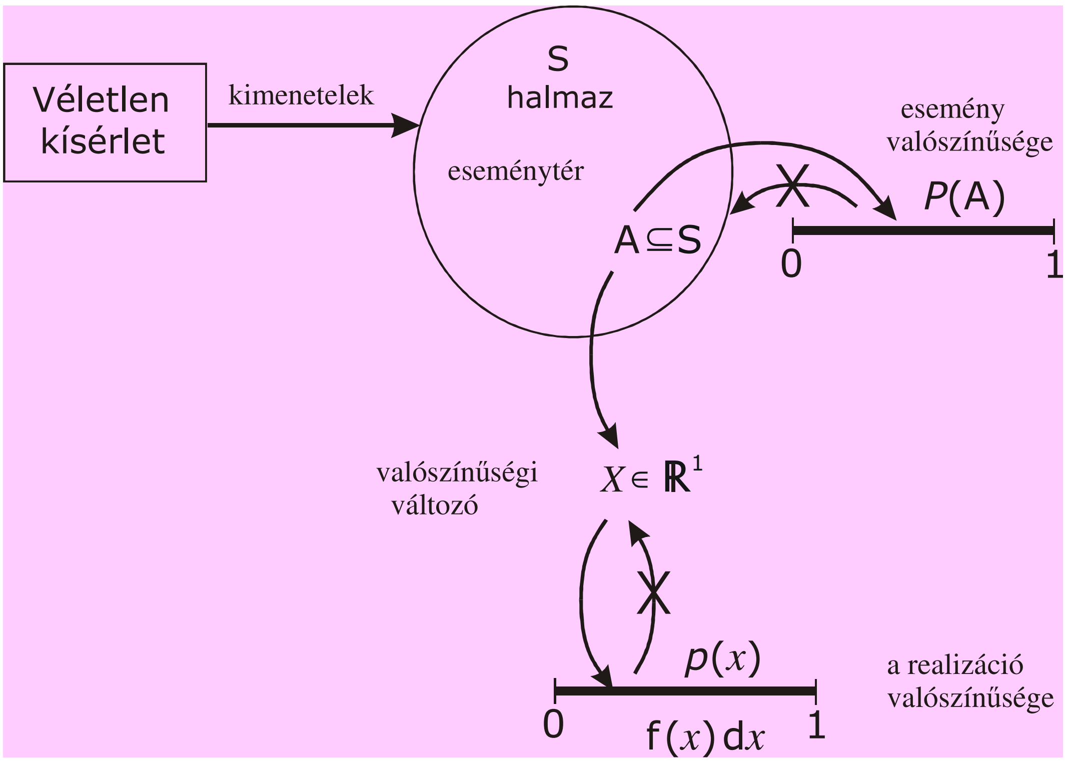
Analógia: Tömegpontok / kontinuum mechanikája

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

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

(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***:

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

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

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

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! $\lim\limits_{a\to b} P(a < X \leq b) = 0$ $\quad$ folytonos X-re

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

$p\,(x \neq b) = 0 \qquad$ ***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***,

illetve 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***.

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: $V(X_{ii})$ (= variancia)

Belőle származik a ***korrelációs együttható***:

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

$\boxed{\text{Tétel:}}$ Ha X és Y ***függetlenek*** $\quad M(XY) = M(X) \cdot M(Y)$

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

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

$\boxed{\text{Tétel:}}$ 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 $N = \int_{-\infty}^{\infty} g(x)\,dx$

$\boxed{N:\ \text{NORMA}}$

ELOSZLÁSFÜGGVÉNY TÍPUSOK

$\boxed{\text{Binomális eloszlás}}$ 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

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

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

$\boxed{\textit{Jelölés}:}$ $\quad K \sim B(p, n)$

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

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

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

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

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!

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

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

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

– 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: $B(0.5, n)$

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

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

Jelölés: $X \sim N(\mu, \sigma^2)$

Definíció $\quad$ 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

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

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

$\boxed{\text{Student-féle } t\text{-eloszlás}}$ (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

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

Jelentőség: mintavétel – ld. később

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

$\boxed{F\text{-eloszlás}}$ (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: $F_{\nu_1, \infty} = \frac{\chi^2_{\nu_1}}{\nu_1}$: redukált χ^2- eloszlás