Valószínűségelméleti és matematikai statisztikai alapfogalmak összefoglalása
(Kemény Sándor - Deák András: Mérések tervezése és eredményeik értékelése, kivonat)

Véletlen jelenség: okok rendszere hozza létre - nem ismerhetjük mind, ezért sztochasztikus.

Véletlen hiba: figyelembe nem vett, egyenként kis hatású változók ingadozásainak összessége, a hiba átlagértéke a mérési szám növelésével nullához tart.

Sokaság és minta: pl. sörösüvegek milliói... De ez véges sokaság - végtelen sokaság pl. egy tömegmérés eredménye.
Mérés: a sokaság egy elemének kiválasztása, véges számmal ismételve az lesz a minta, a mérés mintavétel.

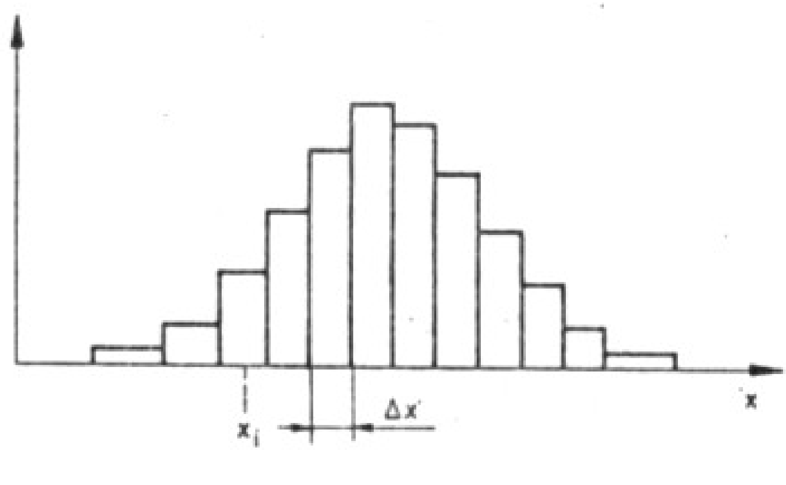
Valószínűségi változó: értéke nem állandó, bizonyos valószínűséggel esik adott határok közé.
(lehet diszkrét - folytonos)

Relatív gyakoriság és valószínűség

Vett minta $\longmapsto$ sokaság eloszlása.
végtelen sokaság: relatív gyakoriság megfelelője a valószínűség
(relatív gyakoriság: előfordulások száma osztva az összes érték előfordulásának számával)
végtelen sok mérésnél a minta azonos a sokasággal. Ekkor a relatív gyakoriság = valószínűség.
A valószínűség a relatív gyakoriság határértéke ∞ minta esetén.

Folytonos valószínűségi változó sűrűség- és eloszlásfüggvénye

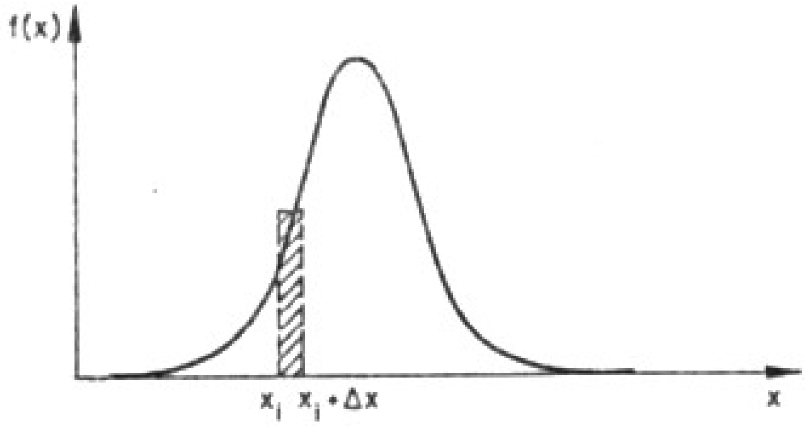
Osztályba sorolás: Δx az ossztály szélessége, x_i az osztály közepe, az osztályindex
Téglalapok területe $\sim$ intervallumba eső elemek számával n_i
Területek összege $\sim$ minta elemszámával (N)



1. ábra. A gyakorisági hisztogram

Relatív gyakorisági hisztogram: a téglalapok területe osztva az összterülettel $\frac{n_i}{N}$
A terület összege 1!
Első esetben a magasság $\sim \frac{n_i}{\Delta x}$, második esetben $\sim \frac{n_i}{(N\Delta x)}$.

Ha $N \to \infty$ (növeljük a minta elemszámát), és az osztályok szélességét csökkentjük ($\Delta x \to 0$) a valószínűség sűrűségfüggvényt kapjuk meg:



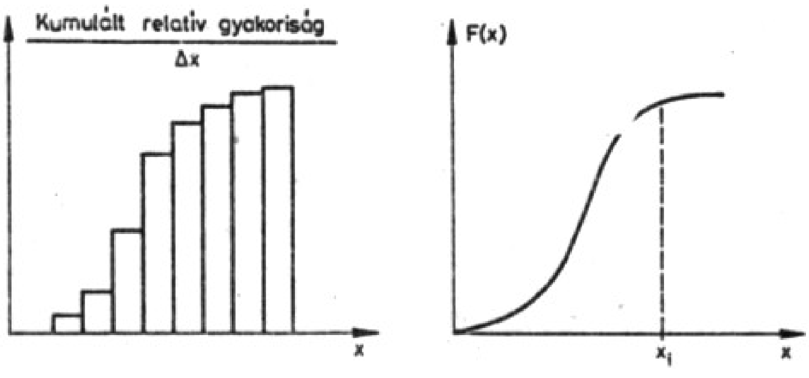
2. ábra. Valószínűség sűrűségfüggvény

$$\lim_{N\to\infty,\, \Delta x\to 0} \frac{rel.\ gyak.}{\Delta x} = f(x) \tag{1}$$

Értelmezése:

$$P(x_0 < x < x_0 + \Delta x) = \int_{x_0}^{x_0+\Delta x} f(x)dx \approx f(x_0)\Delta x \tag{2}$$

Kumulatív relatív gyakoriság: a valószínűségi változó x_i vagy annál kisebb értéket vesz fel.



3. ábra.

$$\lim_{N\to\infty} \frac{kum.\ rel.\ gyak.}{\Delta x} = F(x) \tag{3}$$

F(x) az eloszlásfüggvény!

$$F(x_i) = \int_{-\infty}^{x_i} f(x)dx \tag{4}$$

Értelmezése:

$$F(x_i) = P(x \leqq x_i) \tag{5}$$

$$F(\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1 \tag{6}$$

F(x), f(x) ismerete a sokaság ismeretét jelenti!

A várható érték ($E(x)$ is használatos)

$$M[x] = \int_{-\infty}^{+\infty} xf(x)dx = \mu \tag{7}$$

A sokaság várható értéke megegyezik a valódi értékkel - ha nincs rendszeres hiba.

$$M[\phi(x)] = \int_{-\infty}^{+\infty} \phi(x)f(x)dx \tag{8}$$

$$M[c \cdot x] = c \cdot M[x];\ M[c] = c \tag{9}$$

$$M[x_1 + x_2 + \ldots + x_k] = M[x_1] + M[x_2] + \ldots + M[x_k] \tag{10}$$

Minél inkább igaz, hogy a relatív gyakoriság egyenlő a valószínűséggel, annál jobban megközelíti $\overline{x}$ a várható értéket.

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{11}$$

A medián: az az érték, amelynél nagyobbat a valószínűségi változó ugyanolyan valószínűséggel vesz fel, mint kisebbet:

$$F(\mu_e) = \frac{1}{2} \tag{12}$$

Tapasztalati medián: sorba rendezve a középső érték, páratlan elemszám esetében a két középső átlaga!

Tekintsük példának okáért a következő számsorozatot, amelyek akár mérési sorozatot is reprezentálhatnak:

51, 49, 50, 52, 48, 51, 250, 49, 52,

ahol a hetedik érték szemmel láthatólag erősen kilóg a többi pont közül, amelyet nyugodtan tekinthetünk mérési hibának is. A különböző módon képzett középértékek a következő eredményeket adják:

aritmetikai:

$$A = \frac{1}{n}\sum_{i=1}^{n} a_i = 72,44$$

geometriai:

$$G = \sqrt[n]{\prod_{i=1}^{n} a_i} = 60,03$$

harmonikus:

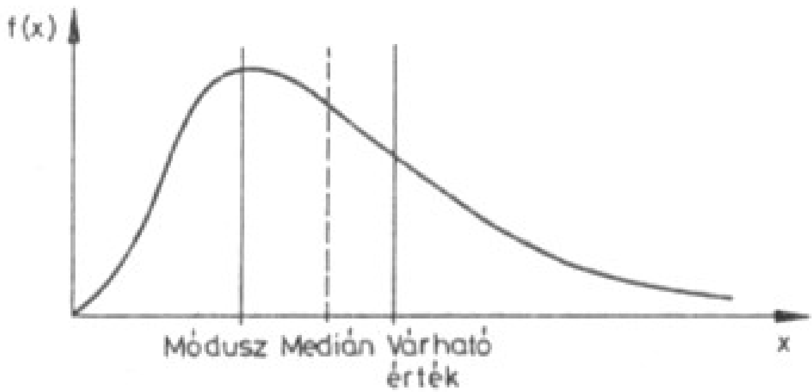
$$H = \frac{n}{\sum_{i=1}^{n}\frac{1}{a_i}} = 55,10$$

négyzetes:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} a_i^2}{n}} = 95,86$$

A median eljárás a „valódi” értékhez közelebb eső 51-et fogja szolgáltatni.

A módusz: a valószínűségi változók legnagyobb valószínűségű értéke, a sűrűségfüggvény maximumhelye (lehet több is, de ez ritka). A tapasztalati módusz a legnagyobb gyakorisági osztály osztályindexe. Egycsúcsos, szimmetrikus eloszlás esetén a módusz és a medián egybeesik a várható értékkel, egyébként nem feltétlenül.



4. ábra.

A variancia (szórásnégyzet)

$$Var[x] = \int_{-\infty}^{+\infty} (x - M[x])^2 f(x)dx = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx = M[(x-\mu)^2] \tag{13}$$

Azaz a várható értéktől való eltérés négyzetének várható értéke.
Jelölés még $D^2[x]$ és σ_x^2.

$$Var[x] = M[x^2] - \mu^2; \;\; \mu = M[x] \tag{14}$$

Számítástechnikailag jobb kifejezés!

$$Var[c \cdot x] = c^2 Var[x] \tag{15}$$

$$Var[x_1 + x_2 + \ldots + x_k] = Var[x_1] + Var[k_2] + \ldots + Var[x_k] \tag{16}$$

A szórásnégyzet:

$$s_x^{2^*} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \tag{17}$$

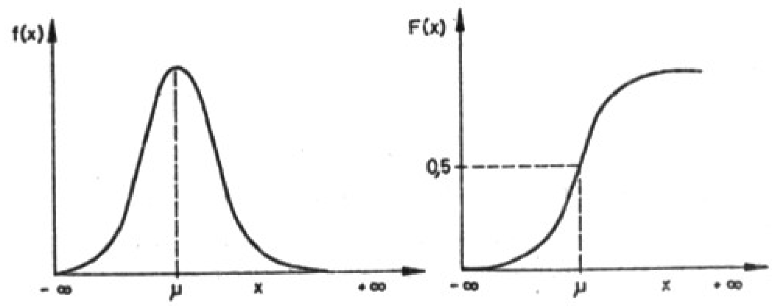
DOF (degrees of freedom): ha tudom a pontos értéket N, ha a méréseinkből származik inkább N-1!

Normális (Gauss) eloszlás: sok, egymástól független, egyenként kis hatású tényező hatása adódik össze.

Sűrűség- és eloszlásfüggvénye:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \, \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{18}$$

$$F(x) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \, \exp\left[-\frac{(\xi-\mu)^2}{2\sigma^2}\right] d\xi \tag{19}$$



5. ábra.

Várható értéke és varianciája:

$$M[x] = \mu; \; Var[x] = \sigma^2 \tag{20}$$

Szokásos jelölés még:

$$N(\mu, \; \sigma^2); \; pl. \; N(0,1) \tag{21}$$

Ha F(x) értékeit táblázatban akarjuk megadni, 3D-s kellene, mivel F(x) μ és σ^2 értékeit is tartalmazza! Transzformáció kellene!

A normalizált (standardizált) normális eloszlás (u-eloszlás)

Legyen u egy valószínűségi változó:

$$u = \frac{x-\mu}{\sigma} \tag{22}$$

Ekkor

$$M[u] = M\left[\frac{x-\mu}{\sigma}\right] = \frac{M[x]-\mu}{\sigma} = 0 \tag{23}$$

$$Var[u] = Var\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma^2}\,Var[x] = 1 \tag{24}$$

A sűrűségfüggvény

$$f(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) \tag{25}$$

$$P(x \leqq a) = F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right]\,dx \tag{26}$$

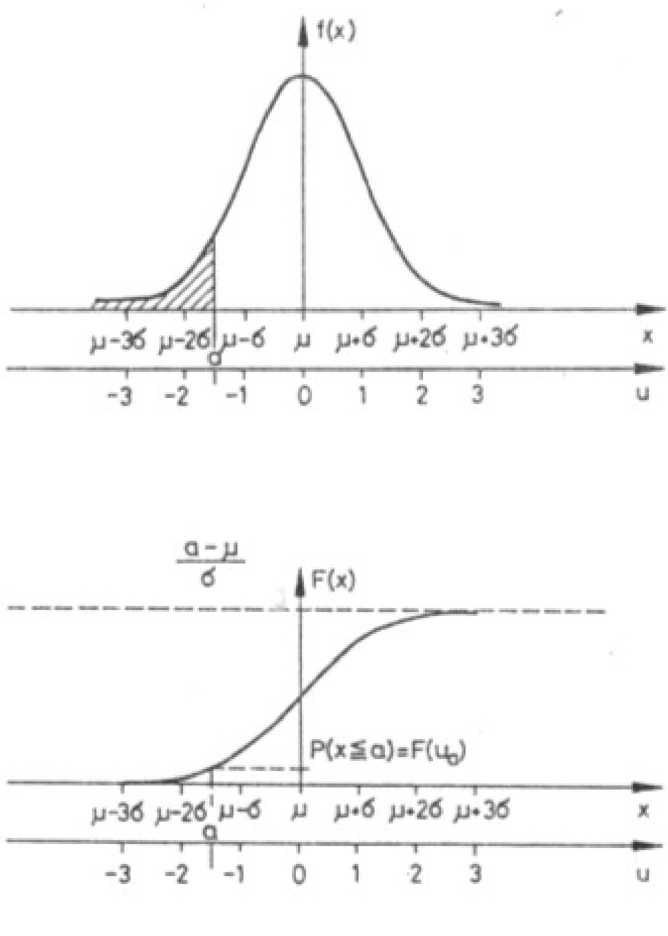
Ez a vonalkázott terület a 6. ábrán! u definíciójából és mivel $dx = \sigma du$

$$P(x \leqq a) = \int_{-\infty}^{u_a} \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{u^2}{2}\right]\,du = F(u_a) \tag{27}$$

ahol

$$u_a = \frac{a-\mu}{\sigma} \tag{28}$$

Tehát annak a valószínűsége, hogy $x \leqq a$ egyenlő $u \leqq u_a = (a-\mu)/\sigma$. Lásd 6. ábra b., része!



6. ábra.

1., példa

Mi annak a valószínűsége, hogy x normális eloszlású valószínűségi változó ebbe az intervallumba esik: $(\mu - \sigma,\ \mu + \sigma)$?

$$P(\mu - \sigma < x \leqq \mu + \sigma) = F(\mu + \sigma) - F(\mu - \sigma) \tag{29}$$

Ezt a 7. ábra szemlélteti. Ekkor igaz, hogy

$$u_{felso} = \frac{\mu + \sigma - \mu}{\sigma} = 1 \tag{30}$$

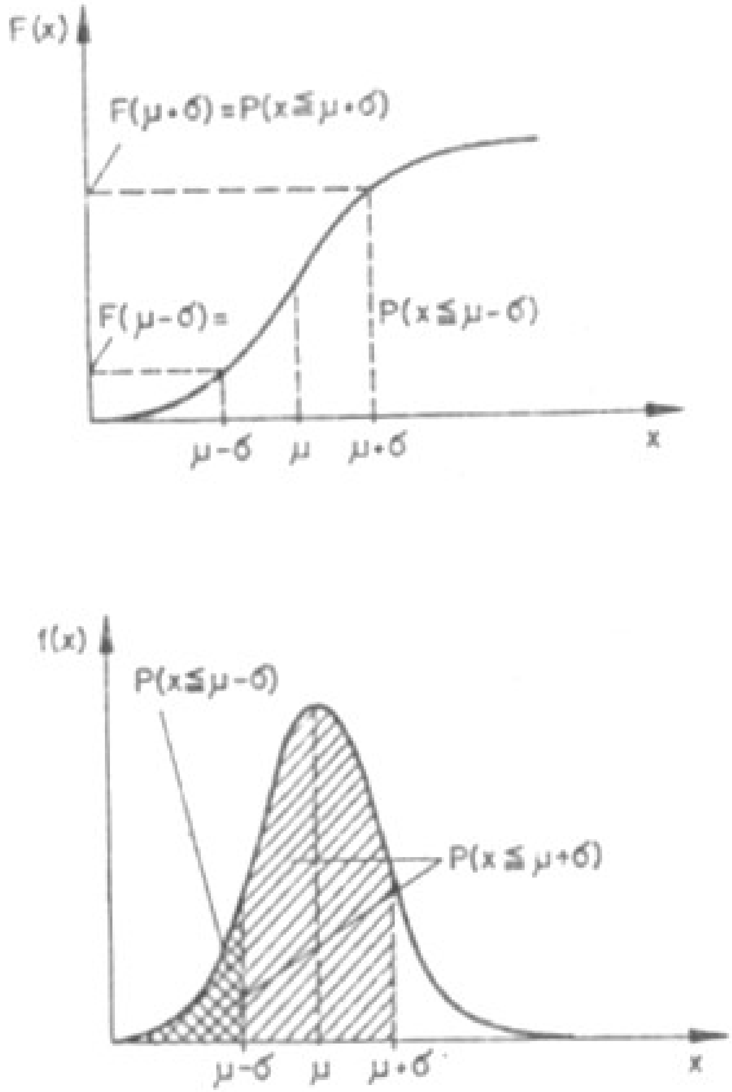
és

$$u_{also} = \frac{\mu - \sigma - \mu}{\sigma} = -1 \tag{31}$$

A táblázatból látjuk, hogy F(1)=0,84134. Mivel f(x) szimmetrikus függvény $F(\infty) = 1$ és $F(-a) = 1 - F(a)$. Így $F(-1) = 1 - F(1) = 0,15866$, ezzel

$$P(\mu - \sigma < x \leqq \mu + \sigma) = 0,68268 \tag{32}$$

Vagyis a valószínűség 68%. Hasonló számítással adódik:



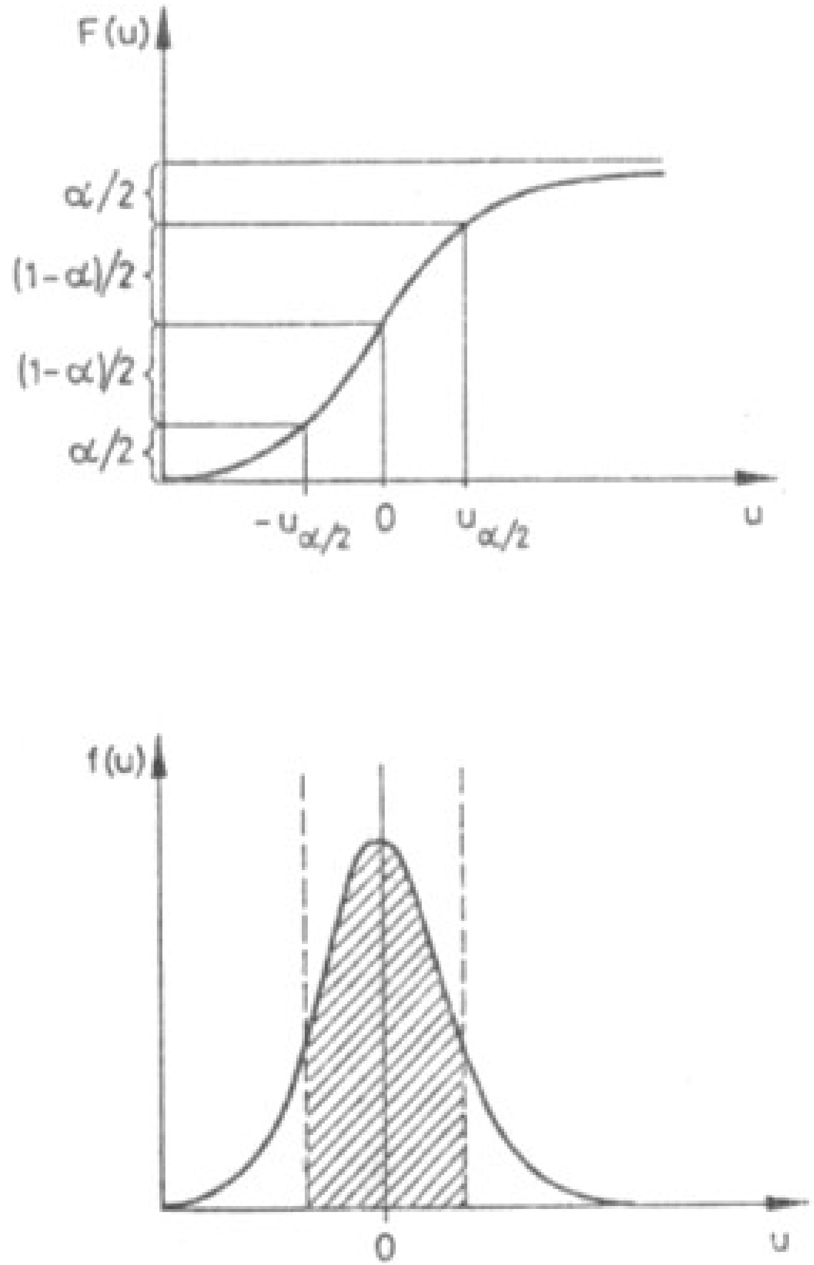
7. ábra.

Intervallum	$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
P	0,68268	0,9545	0,9973

1. táblázat.

2., példa

Egy x jelű, $N(\mu, \sigma^2)$ normális eloszlású valószínűségi változó értékei milyen szimmetrikus intervallumban vannak 95%-os, illetve 99%-os valószínűséggel? u alsó és felső értékének segítségével! Tudjuk, hogy α jelenti, hogy ezen kívül van, azaz jobb- és baloldalt $\alpha/2$ ennek a valószínűsége, lásd 8. ábra!



8. ábra.

A táblázatból látjuk:

α	0,05	0,01
$1-\alpha$	0,95	0,99
$1-\alpha/2$	0,975	0,995
u	1,96	2,58

2. táblázat.

Visszatérve az eredeti x változóra:

$$x_{also} = \mu - \sigma\, u_{\alpha/2}; \quad x_{felso} = \mu + \sigma\, u_{\alpha/2} \tag{33}$$

α	0,05	0,01
x_{also}	$\mu - 1,96\sigma$	$\mu - 2,58\sigma$
x_{felso}	$\mu + 1,96\sigma$	$\mu + 2,58\sigma$

3. táblázat.

3., példa

Analitikai mérlegen tömegmérés esete. Tudjuk, a tömeg $\mu = 5,00000\ g$, a variancia négyzetgyöke $\sigma = 10^{-4}\ g$ és a mérés normális eloszlású.

a., Egy mérés eredménye 99%-os biztonsággal (valószínűséggel) milyen szimmetrikus intervallumba esik?

$$P\left[-u_{\alpha/2} < u \leqq u_{\alpha/2}\right] = 1 - \alpha = 0,99 \tag{34}$$

$$u = \frac{x-\mu}{\sigma} = \frac{x - 5,00000}{10^{-4}} \tag{35}$$

$$P\left[\mu - u_{\alpha/2}\,\sigma < x \leqq \mu + u_{\alpha/2}\,\sigma\right] = 0,99 \tag{36}$$

A táblázatból látjuk, hogy $\alpha = 0,01$-hez $u_{\alpha/2} = 2,58$ tartozik, ezzel

$$P(5,00000 - 2,58 \cdot 10^{-4} < x \leqq 5,00000 + 2,58 \cdot 10^{-4}) = 0,99 \tag{37}$$

A mérés eredménye 99%-os valószínűséggel a (4,99974; 5,00026) tartományba esik, a 4. jegy már bizonytalan!

b., Több mérési eredmény átlagának hibája mekkora intervallumba esik 99%-os biztonsággal?

$$P\left[-u_{\alpha/2} < u \leqq u_{\alpha/2}\right] = 1 - \alpha = 0,99 \tag{38}$$

Itt

$$u = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \tag{39}$$

a hiba pedig $(\overline{x}-\mu)$.

$$P\left[-u_{\alpha/2}\,\sigma/\sqrt{n} < \overline{x}-\mu \leqq u_{\alpha/2}\,\sigma/\sqrt{n}\right] = 0,99 \tag{40}$$

$$P\left[-2,58\cdot 10^{-4}/\sqrt{n} < \overline{x}-\mu \leqq 2,58\cdot 10^{-4}/\sqrt{n}\right] = 0,99 \tag{41}$$

n	hibaintervallum fele	csökkenés mértéke
2	$1,82\cdot 10^{-4}$	0,707
3	$1,49\cdot 10^{-4}$	0,578
4	$1,29\cdot 10^{-4}$	0,500
5	$1,15\cdot 10^{-4}$	0,446
6	$1,05\cdot 10^{-4}$	0,408
7	$0,97\cdot 10^{-4}$	0,378

4. táblázat.

Látható, hogy 4 méréssel a hiba a felére, 7 méréssel majdnem a harmadára csökkenthető!

c., Hány mérés után csökken a középérték hibája 95%-os biztonsággal $\pm 10^{-4}$ alá?

$$P\left[-u_{\alpha/2} < u \leqq u_{\alpha/2}\right] = 1-\alpha \tag{42}$$

$$u = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \tag{43}$$

$$P\left[-u_{\alpha/2}\,\sigma/\sqrt{n} < \overline{x}-\mu \leqq u_{\alpha/2}\sigma/\sqrt{n}\right] = 1-\alpha \tag{44}$$

Tehát a követelmény: $|\overline{x}-\mu| < 10^{-4}$, azaz $u\,\sigma/\sqrt{n} < 10^{-4}$. Mivel

$$\sigma = 10^{-4};\;\; \frac{u_{\alpha/2}}{\sqrt{n}} < 1;\;\; n > \left[u_{\alpha/2}\right]^2 \tag{45}$$

A táblázatból $u_{\alpha/2} = 1,96$, ezért $n > 1,96^2 = 3,86$ vagyis n=4 elég.

Centrális határeloszlási tétel: bármilyen eloszlású sokaságból vett minták számtani középértékei - elég nagy minta esetén (min. 30-50) - közelítőleg normális eloszlást követnek az eredeti eloszlás várható értéke körül. Ezért van sok normális eloszlású változó!

A χ^2 eloszlás:

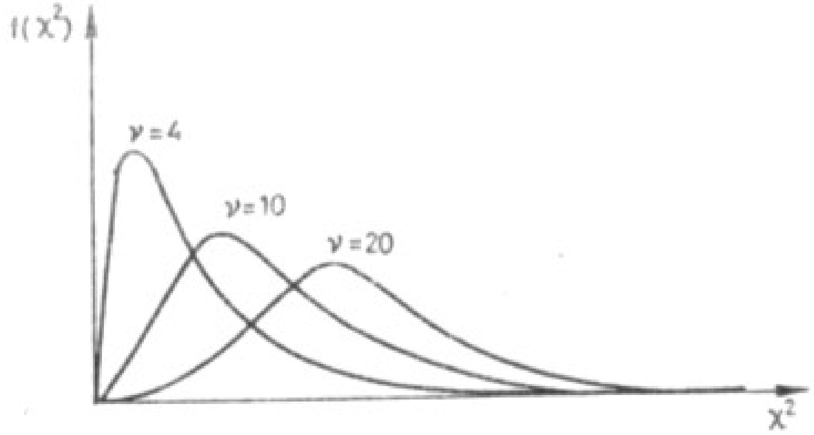
Vegyünk egy $N(\mu,\sigma^2)$ normális eloszlású sokaságból ν elemű mintát: $x_1, x_2, \ldots, x_i, x_\nu$. Ezekből

$$u_i = \frac{x_i-\mu}{\sigma} \tag{46}$$

normalizált normális eloszlású [N(0,1)] valószínűségi változók képezhetők. A χ^2 eloszlású valószínűségi változót így kapjuk:

$$\chi^2 = u_1^2 + u_2^2 + \ldots + u_i^2 + \ldots + u_\nu^2 = \sum_{i=1}^{\nu} u_i^2 \tag{47}$$

A négyzetösszeg szabadságifok-számán a lineáris rendszer szabadságifok-számát értjük. A lineáris rendszer szabadsági fokát megkapjuk, ha a változók számából levonjuk a köztük lévő lineáris összefüggések számát. Mivel itt a tagok egymástól függetlenek, az összeadandók száma (ν) megegyezik a szabadsági fokok számával.
Az eloszlás sűrűségfüggvénye csak a ν paramétert tartalmazza: $f_\nu(\chi^2)$



9. ábra.

Várható értéke és varianciája:

$$M[\chi^2] = \nu; \;\; Var[\chi^2] = 2\nu \tag{48}$$

A t-eloszlás (Student-eloszlás)

u és χ^2 egymástól független valószínűségi változó. Az első N(0,1) eloszlású, a második χ^2 eloszlású ν szabadsági fokkal. Legyen t egy új valószínűségi változó:

$$t = \frac{u}{\sqrt{\frac{\chi^2}{\nu}}} \tag{49}$$

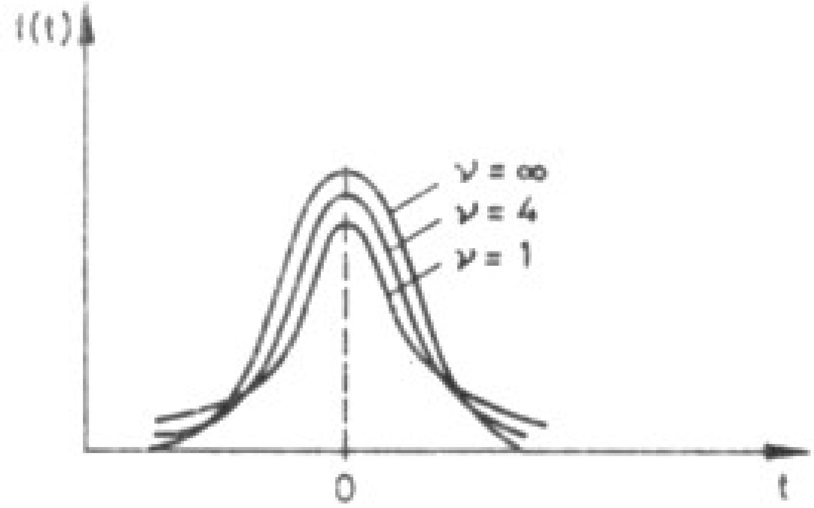
Ehhez az úgynevezett Student-eloszlás tartozik. Ennek egyetlen paramétere ν, a szabadsági fok. Ha $\nu \to \infty$ a t-eloszlás tart a normális eloszláshoz. A gyakorlatban $\nu > 30$ esetben ez igaz...
A középértékre vonatkozó

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \tag{50}$$

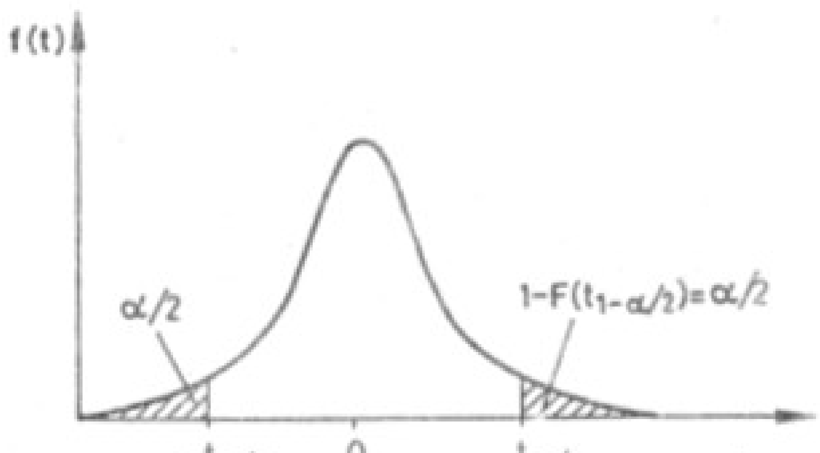
is t-eloszlású n-1 szabadsági fokkal, ahol s a szórás. A t-eloszlás szimmetrikus:

$$F\left[t_{\alpha/2}\right] = 1 - F\left[t_{1-\alpha/2}\right] \tag{51}$$

Nem túl elegáns jelöléssel: $t_{1-\alpha/2}$ helyett $-t_{\alpha/2}$



10. ábra.



11. ábra.