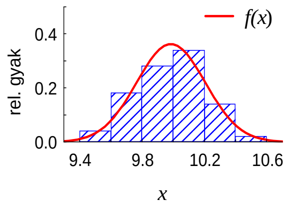
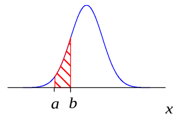
STATISZTIKAI ALAPOK

Folytonos valószínűségi változó

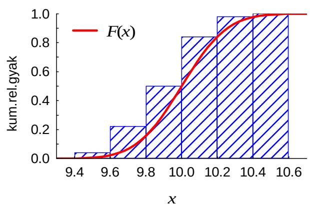


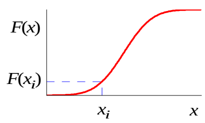


$$P(a < x \le b) = \int_a^b f(x)dx$$

sűrűségfüggvény

Folytonos valószínűségi változó





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

eloszlásfüggvény

Paraméter és statisztika

sokaság

- várható érték:

$$E(x) = \mu$$

- variancia

$$Var(x) = \sigma_x^2$$

minta

- számtani átlag:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

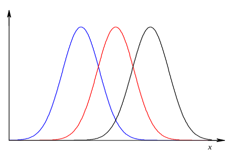
- szórásnégyzet (korrigált)

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

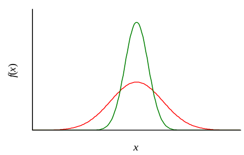
A legfontosabb folytonos eloszlás: normális eloszlás

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

Két paramétere van: μ és σ^2



μ különböző



σ különböző

Várható értéke és varianciája:

$$E(x) = \mu \qquad , \qquad Var(x) = \sigma^2$$

Rövid jelölése:

$$N(\mu, \sigma^2) \qquad \text{pl.} \qquad N(0,1)$$

Célszerű transzformációt keresnünk

Normalizált (standardizált) normális eloszlás

$$u = \frac{x-\mu}{\sigma} \qquad f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

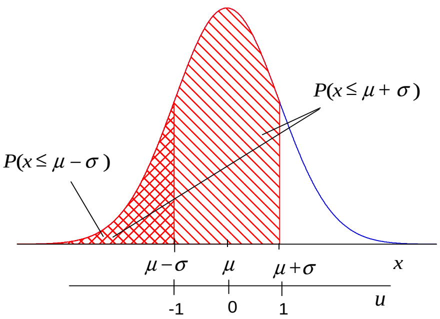
$$E(u) = 0 \qquad Var(u) = 1$$

1. példa

Határozzuk meg annak valószínűségét, hogy az x normális eloszlású valószínűségi változó a (μ-σ , μ+σ) intervallumba eső értéket vesz fel!

(Pl. azt kérdezzük, hogy milyen valószínűséggel esik a 10±0.5 intervallumba, ha μ=10, σ=0.5)

$$P(\mu - \sigma < x \le \mu + \sigma) = F(\mu + \sigma) - F(\mu - \sigma)$$



$$u_{alsó} = \frac{\mu - \sigma - \mu}{\sigma} = -1 \qquad u_{fölső} = \frac{\mu + \sigma - \mu}{\sigma} = 1 \qquad u = \frac{x-\mu}{\sigma}$$

intervallum szélessége	$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
P			

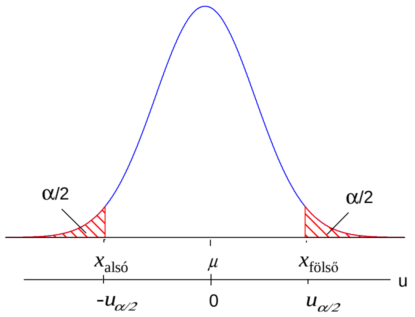
2. Példa

Határozzuk meg, hogy egy $N(\mu,\sigma^2)$ normális eloszlású valószínűségi változó értékei milyen szimmetrikus intervallumban vannak 95 %-os, ill. 99 %-os valószínűséggel! (Pl. μ=10, σ=0.5)

α	0.05	0.01
1-α	0.95	0.99
1-α/2	0.975	0.995
u		

α	0.05	0.01
$x_{alsó}$		
$x_{fölső}$		

$$u = \frac{x-\mu}{\sigma} \longrightarrow x = u\sigma + \mu$$



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

A számtani középérték

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum x_i$$

$$E(\bar{x}) = \frac{1}{n}[nE(x)] = E(x) = \mu$$

$$\sigma_{\bar{x}}^2 = Var(\bar{x}) = \frac{Var(x)}{n} = \frac{\sigma_x^2}{n}$$

Centrális határeloszlási tétel

Bármilyen eloszlású sokaságból vett minták számtani középértéke közelítőleg normális eloszlást követ az eredeti eloszlás várható értéke körül, varianciájuk pedig σ^2/n; tehát $N(\mu, \sigma^2/n)$ eloszlású.

3. példa

Ha egy μ=10 várható értékű és σ^2 =0.25 varianciájú sokaságból n = 5 elemű mintát veszünk, milyen intervallumban lesz a mintaelemek átlaga 95% valószínűséggel?

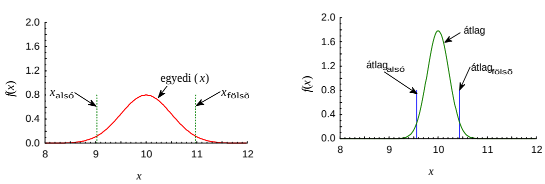
$$P\left(\mu - u_{\alpha/2}\,\sigma/\sqrt{n} < \bar{x} \leq \mu + u_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1-\alpha$$

$$u_{\alpha/2} =$$

$$\frac{\sigma}{\sqrt{n}} =$$

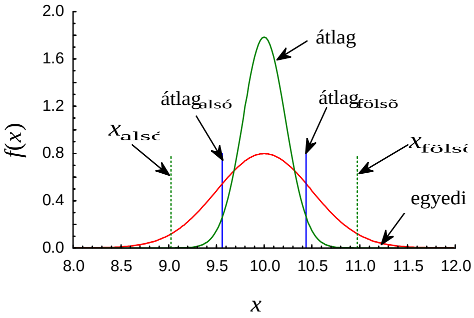
$$\mu + u_{\alpha/2}\frac{\sigma}{\sqrt{n}} =$$

$$\mu - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} =$$



$$P(\qquad < x \leq \qquad) = 0.95$$

$$P(\qquad < \bar{x} \leq \qquad) = 0.95$$

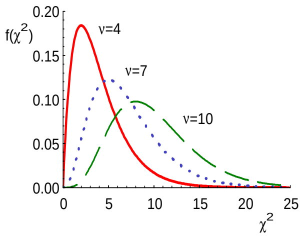


$$u = \frac{x-\mu}{\sigma} \qquad u = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

(az átlag is normális eloszlású valószínűségi változó)

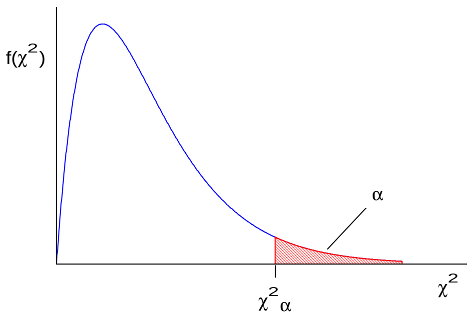
$$P\left(-u_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le u_{\alpha/2}\right) = 1-\alpha$$

χ^2- (khi-négyzet-) eloszlás



$$\chi^2 = \sum_{i=1}^{n} u_i^2$$

$$E(\chi^2) = \nu$$



A normális eloszlású sokaságból vett minta tapasztalati szórásnégyzetének eloszlása

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

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \chi^2\sigma^2, \qquad \nu = n-1$$

$$s^2 = \frac{\chi^2\sigma^2}{\nu} \longrightarrow \chi^2 = \frac{s^2\nu}{\sigma^2}$$

4. példa

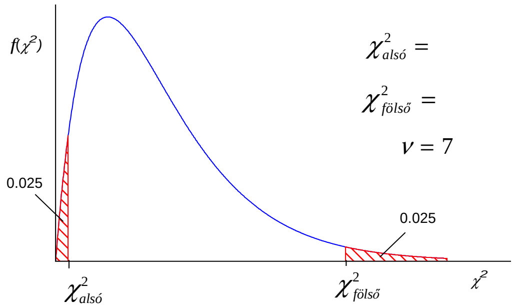
Egy $\sigma^2 = 0.08$ varianciájú normális eloszlású sokaságból 8 elemű mintát veszünk.

a) Határozzuk meg azt a szimmetrikus valószínűségű intervallumot, amelyben az eredmények s^2 korrigált tapasztalati szórásnégyzete 95%-os valószínűséggel megtalálható!

$$P\left(s^2_{alsó} < s^2 \le s^2_{fölső}\right) = 0.95$$

$$P\left(\frac{s^2_{alsó}\nu}{\sigma^2} < \chi^2 \le \frac{s^2_{fölső}\nu}{\sigma^2}\right) = P\left(\chi^2_{alsó} < \chi^2 \le \chi^2_{fölső}\right) = 0.95$$

A χ^2-eloszlás kritikus értékei



$\chi^2_{alsó} =$

$\chi^2_{fölső} =$

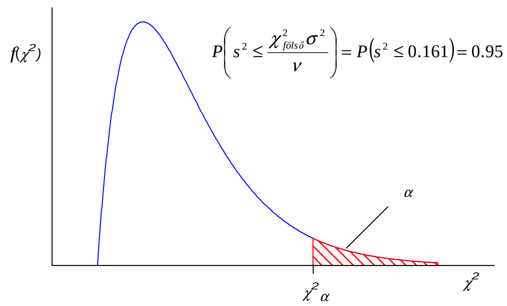
$\nu = 7$

$$P\left(\chi^2_{alsó} < \chi^2 \le \chi^2_{fölső}\right) = P\left(\frac{\chi^2_{alsó}\sigma^2}{\nu} < s^2 \le \frac{\chi^2_{fölső}\sigma^2}{\nu}\right) =$$

b) Határozzuk meg azt az értéket, amelyet s^2 95%-os valószínűséggel nem halad meg!

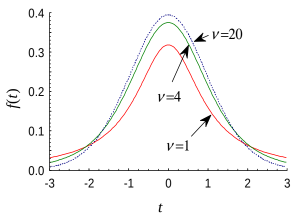
$$P\left(s^2 \le s^2_{fölső}\right) = 0.95$$

A χ^2 eloszlás fölső kritikus értéke: $\chi^2_{fölső} = 14.1$



$$P\left(s^2 \le \frac{\chi^2_{fölső}\sigma^2}{\nu}\right) = P\left(s^2 \le 0.161\right) = 0.95$$

t-eloszlás (Student-eloszlás)



$$t = \frac{u}{\sqrt{\frac{\chi^2}{\nu}}} = \frac{\xi - E(\xi)}{s_\xi}$$

$$E(t) = 0$$

pl. $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

5. példa

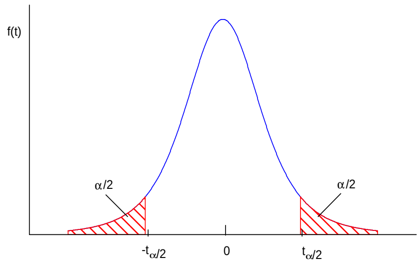
10 mérés eredménye:
24.46; 23.93; 25.79; 25.17; 23.82; 25.39; 26.54; 23.85; 24.19; 25.50.
Kérdés: a valódi érték 95%-os valószínűséggel milyen intervallumban van? (Adjunk 95%-os konfidencia-intervallumot a várható értékre!)

$\bar{x} = 24,9$ $\quad s^2 = 0,894$ $\quad s =$

$\nu =$ $\quad t_{\alpha/2} =$

$$P\left(\bar{x} - t_{\alpha/2}\, s/\sqrt{n} < \mu \le \bar{x} + t_{\alpha/2}\, s/\sqrt{n}\right) = 1 - \alpha$$

konfidencia-intervallum:



$\nu = n - 1 = 9$

$\alpha =$

$t_{\alpha/2} =$

F-eloszlás

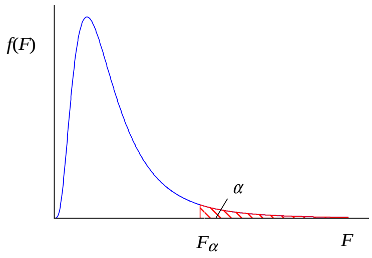
Legyen χ_1^2 és χ_2^2 két, egymástól független, χ^2 -eloszlású valószínűségi változó ν_1 , ill. ν_2 szabadsági fokkal. A következő kifejezés *F*-eloszlású, a számláló szabadsági fokainak száma ν_1 , a nevezőé ν_2 :

$$F = \frac{\chi_1^2}{\nu_1} \bigg/ \frac{\chi_2^2}{\nu_2} , \qquad F = \frac{s_1^2 \,/\, \sigma_1^2}{s_2^2 \,/\, \sigma_2^2} ,$$

ha $\sigma_1^2 = \sigma_2^2$

$$F = s_1^2 / s_2^2$$

Az *F*-eloszlás kritikus értékei



$$F_\alpha(\nu_1,\nu_2) = \frac{1}{F_{1-\alpha}(\nu_2,\nu_1)}$$

6. példa

Azonos analitikai módszerrel két méréssorozatot kaptunk, amelyek 4 ill. 7 mérésből állnak. Milyen intervallumban lehet 90 % valószínűséggel a két minta szórásnégyzetének aránya?

Minthogy azonos módszerről van szó, a variancia változatlan:

$$\sigma_1^2 = \sigma_2^2$$

$$P\left(F_{alsó} < s_1^2 / s_2^2 \le F_{fölső}\right) = 0.90$$

$F_{fölső} =$

$F_{alsó} =$