A valószínűségszámítás elemei

Alapfogalmak

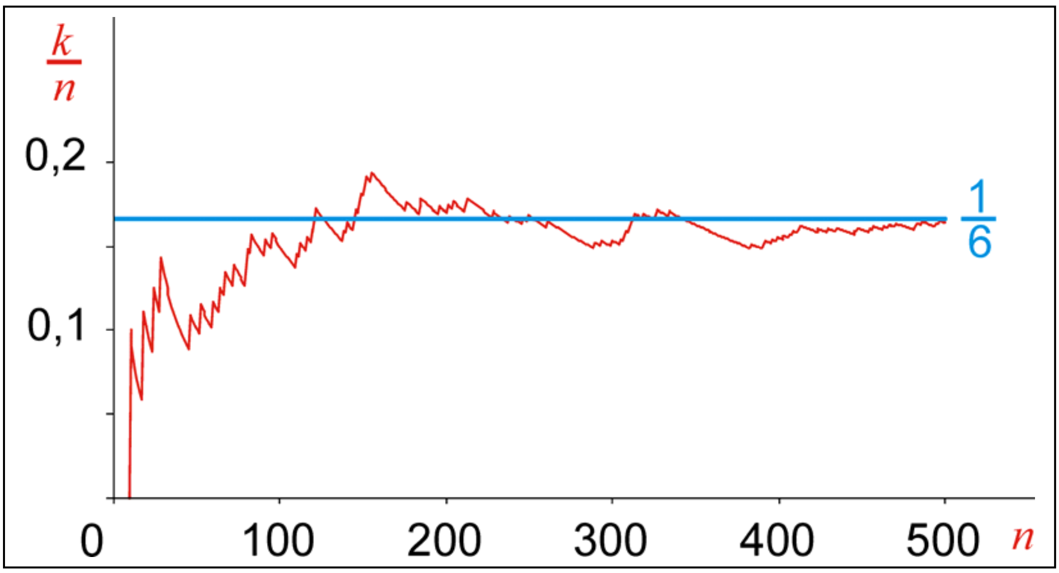
Jelenség: minden, ami lényegében azonos feltételek mellett megismételhető, amivel kapcsolatban megfigyeléseket lehet végezni, lehet vele „kísérletezni”.
Pl. orvosi vizsgálat | pénzfeldobás | várakozás a villamosra

Megfigyelés: megadjuk, hogy a jelenséggel kapcsolatban mire vagyunk kíváncsiak és, hogy azt hogyan érzékeljük, hogyan mérjük.
Pl. bőr szín | az érme repülési ideje | hányan várakoznak

Esemény: egy állítás, ami vagy bekövetkezik, vagy nem.
Pl. sárga | 0,5 s és 1,5 s között van | 10-en

Kísérletsorozatban az esemény relatív gyakorisága:
k/n, ahol k az esemény bekövetkezésének abszolút gyakorisága, n a kísérletek száma.

Pl. Jelenség: kockadobás
Megfigyelés: hányast dobunk
Esemény: 6-ost dobunk



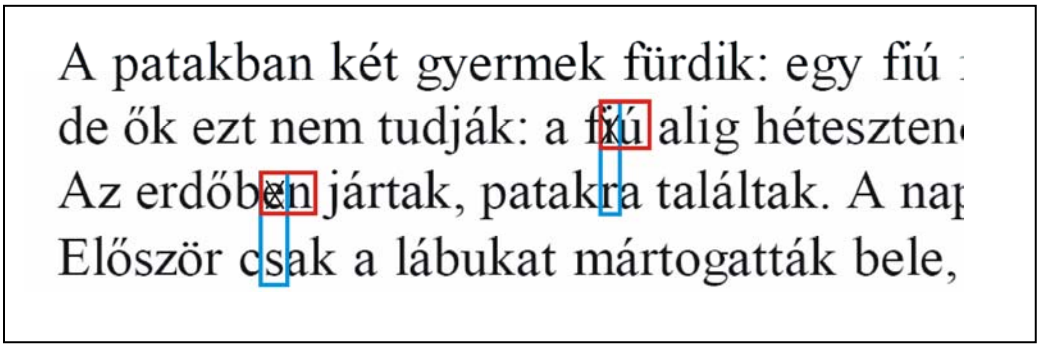
A **nagy számok** (relatív gyakoriságokra vonatkozó) tapasztalati **törvénye**:
n növekedtével k/n stabilizálódik valamilyen érték körül. Ez a szám nem függ az aktuális kísérletsorozattól.
(Logikai úton bizonyítani nem lehet) (Karl Pearson 1857-1936)
Az eseményhez egy számot rendelhetünk: **valószínűség**

A valószínűség tulajdonságai:

1. Egy A esemény valószínűsége, $[P(A)]$ mindig $0 \le P(A) \le 1$.
2. A biztos esemény valószínűsége, $P(biztos) = 1$,
   a lehetetlen esemény valószínűsége, $P(lehetetlen) = 0$.
3. Egymást kizáró események (pl. A és B) egyesítésének valószínűsége: $P(A+B) = P(A) + P(B)$.

Összetett megfigyelés

Példa:
Egy magyar nyelven írt szövegből „rábökéssel” kiválasztunk egy betűt, majd megfigyeljük a (tőle jobbra) mellette és az alatta lévőt. (Az üres helyeket nem vesszük figyelembe.)


A patakban két gyermek fürdik: egy fiú
de ők ezt nem tudják: a fiú alig hétesztend
Az erdőben jártak, patakra találtak. A na
Először csak a lábukat mártogatták bele,

A esemény: „a második betű magánhangzó”
B esemény: „az első betű magánhangzó”
$\overline{A}$ esemény (A ellentettje): „a második betű mássalhangzó”
$\overline{B}$ esemény (B ellentettje): „az első betű mássalhangzó”

A 100 ismételt „rábökésből” álló kísérletsorozat eredménye:

Abszolút gyakoriságok az egymás melletti betűkre.

	B	$\overline{B}$	összesen
A	**6**	34	40
$\overline{A}$	36	24	60
összesen	42	58	100

Abszolút gyakoriságok az egymás alatti betűkre.

	B	$\overline{B}$	összesen
A	**16**	23	39
$\overline{A}$	26	35	61
összesen	42	58	100

Számítsuk ki a ciánkék részre vonatkozó feltételes relatív gyakoriságokat:

	B
A	**0,14**
$\overline{A}$	0,86
összesen	1

	B
A	**0,38**
$\overline{A}$	0,62
összesen	1

Ennek alapján bevezethetjük a **feltételes valószínűséget.** Annak a valószínűsége, hogy „a második betű magánhangzó” ha „az első betű magánhangzó”, $P(A|B)$: az A esemény B eseményre vonatkoztatott feltételes valószínűsége.

$P(AB)$: annak a valószínűsége, hogy A és B is bekövetkezik.
Ez ≈0,06 az egymás melletti, és ≈0,16 az egymás alatti betűkre.
$P(B)$: B bekövetkezésének valószínűsége, ez ≈0,42.

Így:

$$P(A|B) = \frac{P(AB)}{P(B)} \qquad \text{(szorzási szabály)}$$

A esemény a B-től **független**, ha

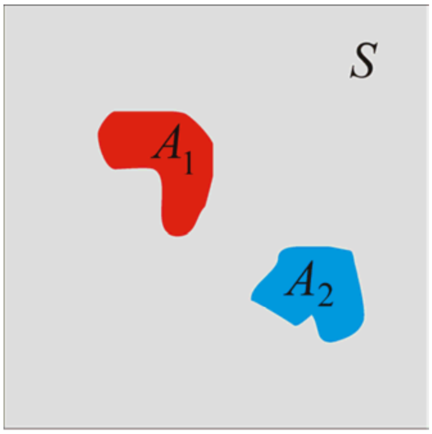
$$P(A|B) = P(A),$$

ami ekvivalens azzal, hogy

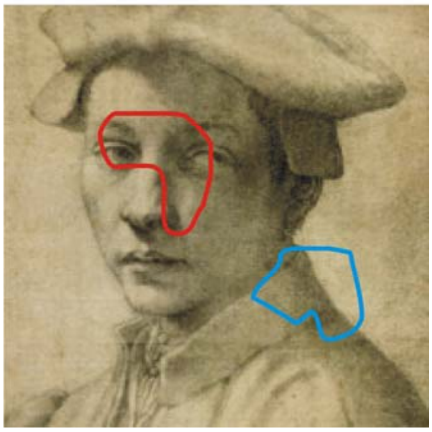
$$P(AB) = P(A)P(B).$$

Az eloszlás hétköznapi fogalma

Vegyünk egy példát: grafika.



Kell egy darab papír, S alaphalmaz, amin eloszlik a grafit.
Ennek A részhalmazaihoz számokat rendelünk (pl. annak alapján, hogy mennyi a grafit tömege (súlya) az adott halmazon).



Additív tulajdonság:

$$P(A) = \sum_{k=1}^{n} P(A_k)$$

ha A_k-knak nincs közös részük.

Normálás: az összes grafit mennyisége legyen egységnyi tömegű (súlyú),

$$P(S) = 1.$$

Valószínűségi változó

Egy jelenséggel kapcsolatban kvantitatív dolgot figyelünk meg.

1. Megadjuk, hogy mit és hogyan „mérünk”.
2. A valószínűségi változót az eloszlásával,
illetve, (ha vannak) annak paramétereivel jellemezzük.

Ezeket általában nem ismerjük.

Gyakorlatilag minden olyan megfigyelésen, tehát nem kizárólag absztrakción alapuló „változás”, amihez számokat rendelhetünk, ilyen. Értéke számba nem vehető tényezőktől, tehát a „véletlentől” is függ.

Diszkrét valószínűségi változó jellemzése

Pl. kockadobás két kockával (függetlenek), 36 lehetséges dobás.

Legyen a valószínűségi változó $\xi = i + k$;
$i = 1, 2, 3, 4, 5, 6$ és $k = 1, 2, 3, 4, 5, 6$, így
ξ 11 különböző értéket vehet föl:

a lehetséges kimenetelek: $x_j = 2$ -től 12 -ig.

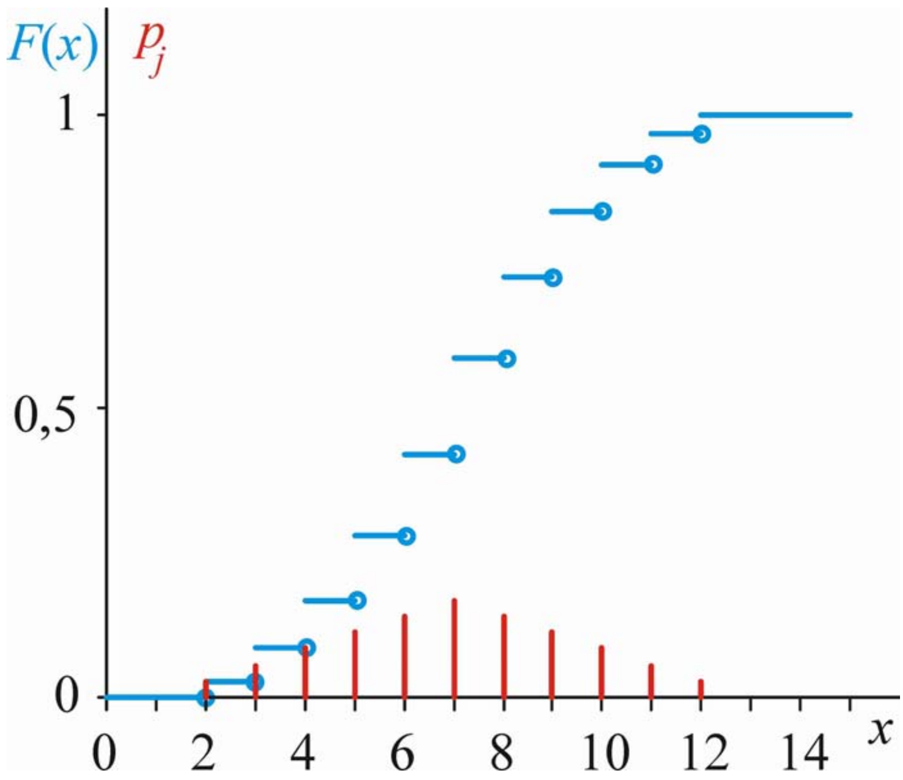
A dobás „eredménye” valamelyik lehetséges kimenetel.

Jellemzés:

Eloszlásfüggvénnyel [$F(x)$] és **Valószínűségekkel** [p_j]

$$F(x) = p(\xi < x) = \sum_{x_j < x} p(\xi = x_j)$$

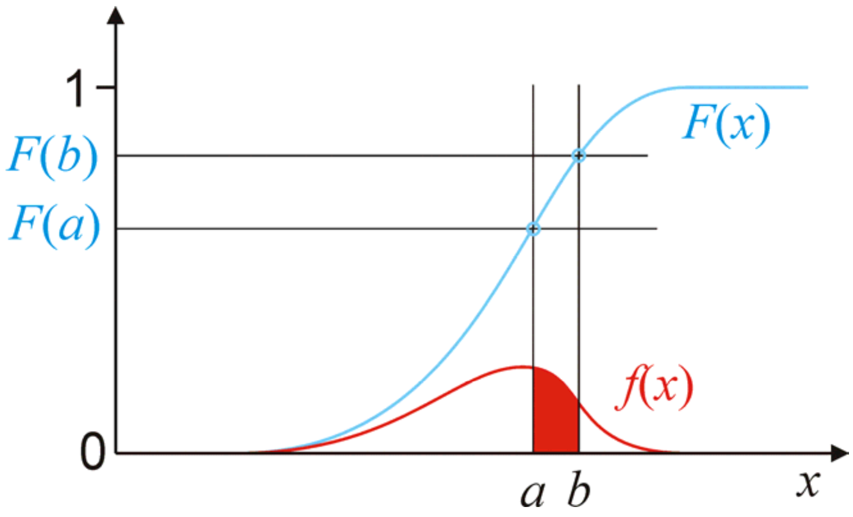
$$p_j = p(\xi = x_j)$$



x_j	p_j
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Folytonos valószínűségi változó jellemzése

(Kumulatív)
Eloszlásfüggvénnyel [$F(x)$] és Sűrűségfüggvénnyel [$f(x)$]



$$F(b) - F(a) =$$
$$= p(a < \xi < b) =$$
$$= \int_a^b f(x)dx =$$
= [piros terület]

A valószínűségi változóra ill. annak eloszlására vonatkozó **számszerű jellemzők** (paraméterek)

Hol van az eloszlás közepe?

várható érték [$M(\xi)$] ()

Diszkrét eset: $M(\xi) = \sum_i x_i p_i$

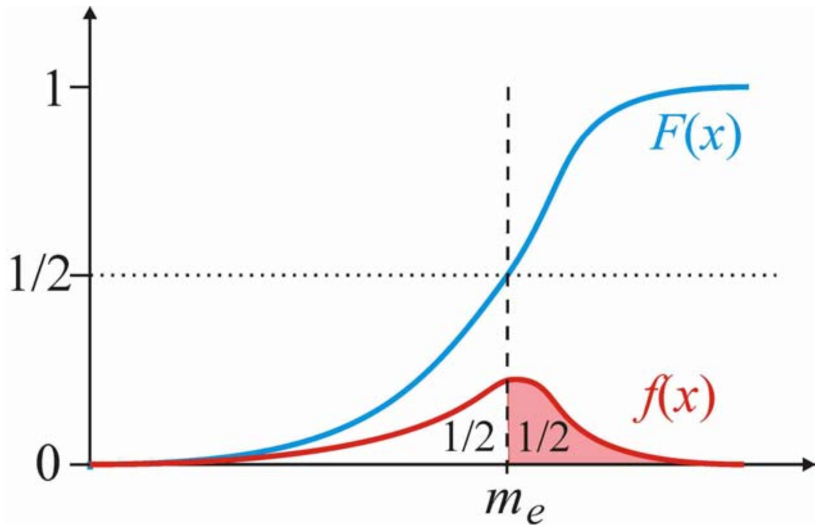
Folytonos eset: $M(\xi) = \int_{-\infty}^{\infty} x f(x)dx$

x_i	p_i	$x_i p_i$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
		252/36 = 7

szemléltetése: tömegközéppont (súlypont) helyzete

medián (m_e)

$F(m_e) = 1/2$



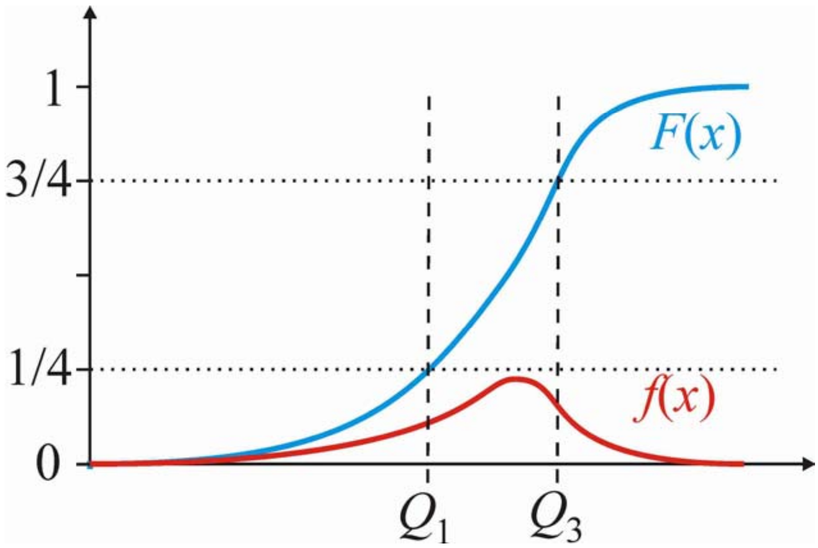
szemléltetése: két egyforma valószínűségű (1/2) tömeg, (súly) ill. terület osztóértéke.

kvantilisek

egyéb valószínűségarány vagy, tömegarány (súlyarány), ill. területarány osztóértékei (Q_1 alsó, Q_3 felső kvartilis)

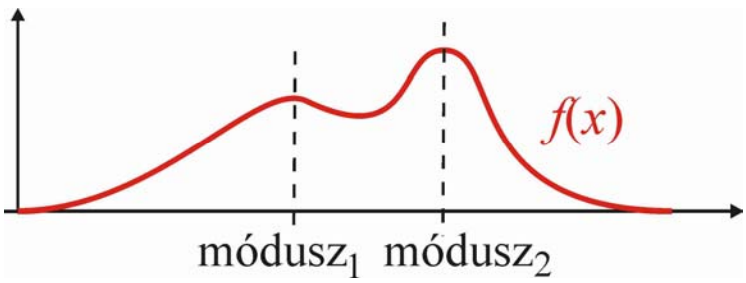
$F(Q_1) = 1/4$

$F(Q_3) = 3/4$

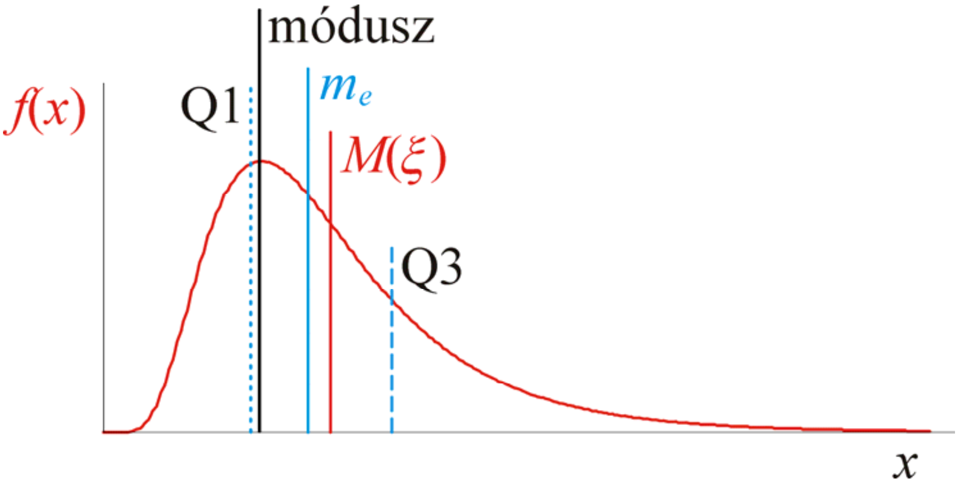


módusz(ok)

a legvalószínűbb értéke(k),
a sűrűségfüggvény lokális maximum értéke(i)



A „közép" számszerű jellemzőinek egymáshoz való viszonya:



Milyen széles az eloszlás?

variancia (szórásnégyzet)

$$D^2(\xi) = M[(\xi - M(\xi))^2]$$

További általánosított jellemzők:

Momentumok | **Centrális momentumok**

$M(\xi^k)$ | $M[(\xi - M(\xi))^k]$ | $k = 1, 2, 3, 4, \ldots$

Mennyire ferde az eloszlás?
A 3. centrális momentummal jellemezhető

Mennyire csúcsos (vagy lapos) az eloszlás?
A 4. centrális momentummal jellemezhető

A várható érték néhány tulajdonsága

$$M(k\xi) = kM(\xi)$$

$$M(\xi + \eta) = M(\xi) + M(\eta)$$

ha ξ és η független valószínűségi változók, akkor
$M(\xi\eta) = M(\xi)M(\eta)$,

A variancia néhány tulajdonsága

$$D^2(a\xi + b) = a^2 D^2(\xi)$$

ha ξ és η független valószínűségi változók, akkor
$D^2(\xi + \eta) = D^2(\xi) + D^2(\eta)$

Ebből következik
ha $D(\xi_i) = \sigma$, $i = 1, 2, \ldots, n$, akkor $D^2(\xi_1 + \xi_2 + \ldots + \xi_n) = n\sigma^2$

Nevezetes eloszlások

1. Diszkrét eloszlások

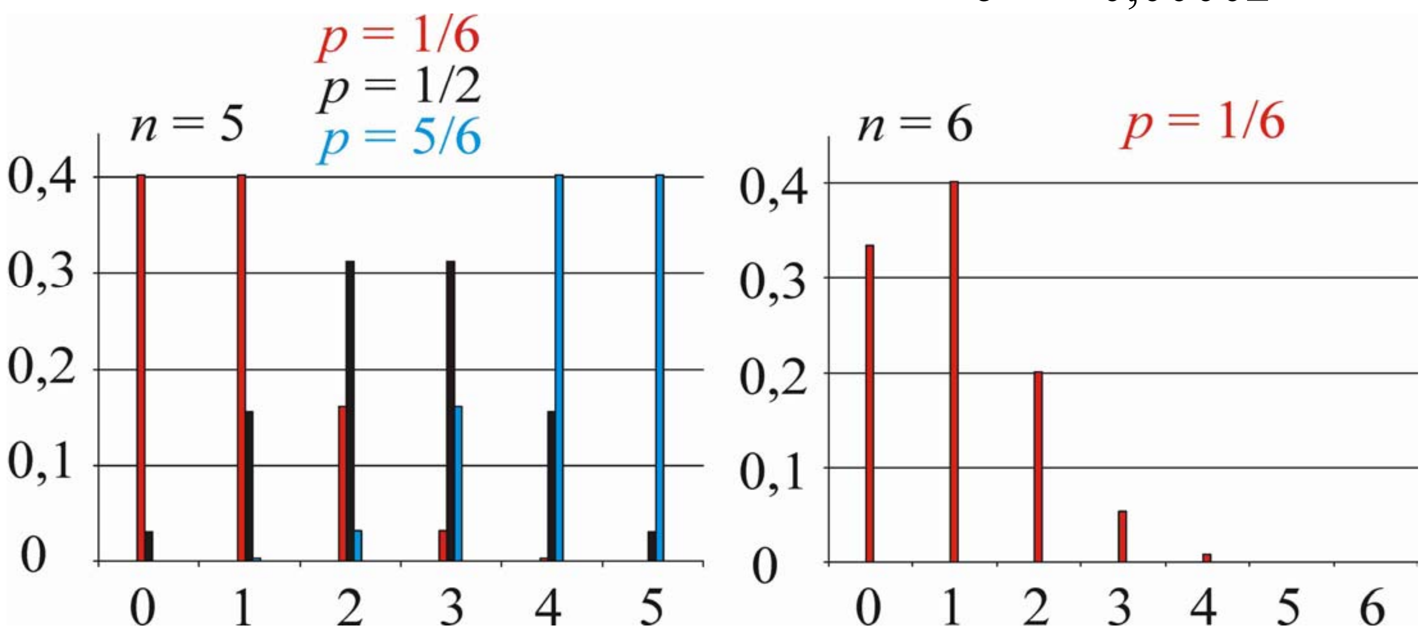
Binomiális eloszlás (Bernoulli-eloszlás)

alternatíva p, $(1-p)$
n ismétlés $P(\xi = k) = B(n, k)$

$M(\xi) = np$,
$D^2(\xi) = np(1-p)$

Példa dobókocka, 6 dobás, $n = 6$
hányszor dobok 6-ost?

k	P
0	0,33
1	0,4
2	0,2
3	0,05
4	0,008
5	0,0006
6	0,00002

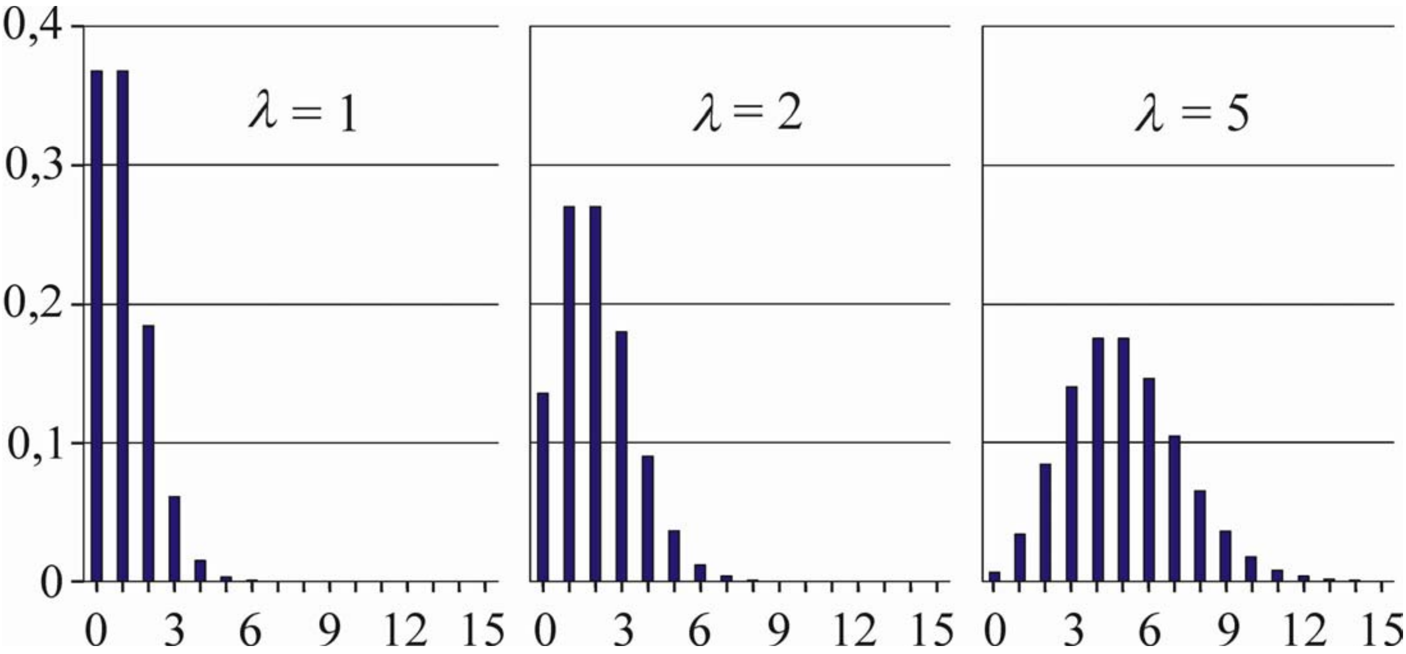


Poisson-eloszlás

$M(\xi) = \lambda$, $\quad\quad D^2(\xi) = \lambda$

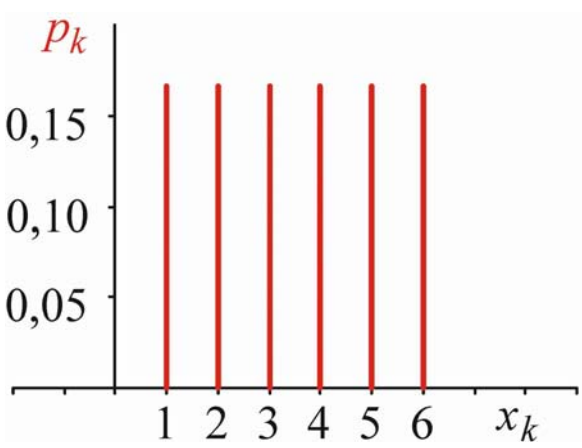
Példák:

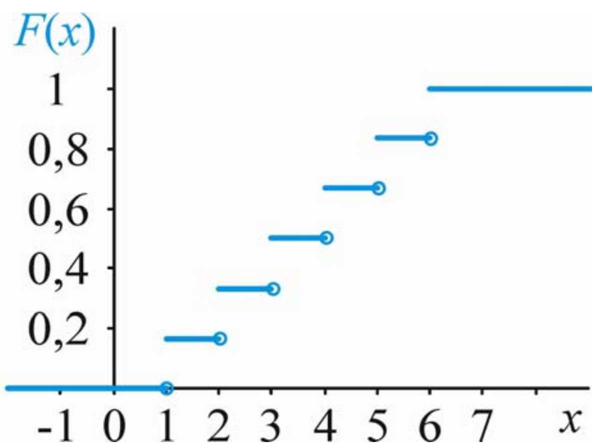
Egy telefonközpontba adott idő alatt befutó hívások száma.
Adott térfogatban lévő részecskék száma.
Radioaktív preparátumban adott idő alatt elbomló atomok száma.



Egyenletes eloszlás

Egy konkrét esetben:
Példa dobókocka, az egyes dobások valószínűsége $p = 1/6$.
Lehetséges értékek 1, 2, 3, 4, 5, 6.

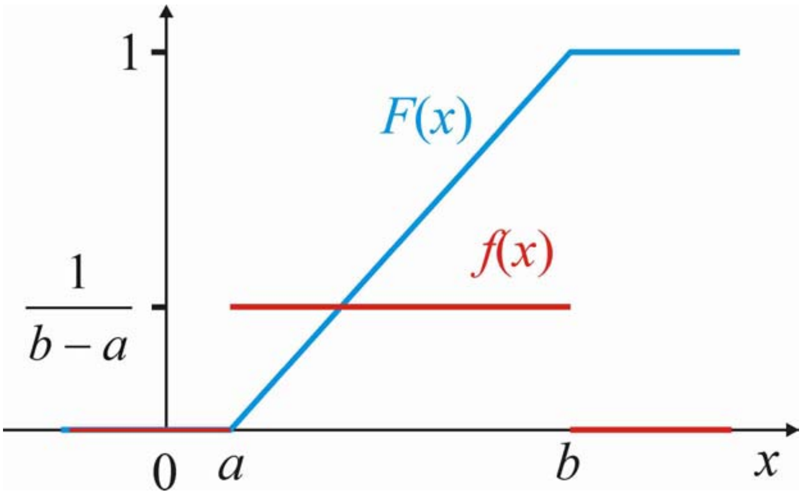




2. Folytonos eloszlások

Egyenletes eloszlás

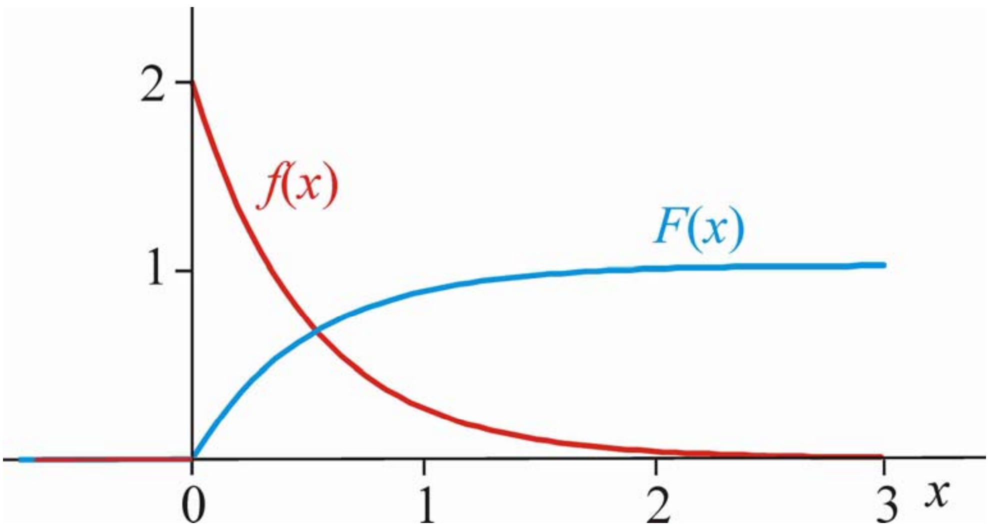
$M(\xi) = (a + b)/2$
$D^2(\xi) = (b - a)^2/12$



Példa: A teremben a levegő sűrűsége vagy hőmérséklete.

Exponenciális eloszlás

$M(\xi) = 1/\lambda$, $(\lambda = 2)$
$D^2(\xi) = 1/\lambda^2$

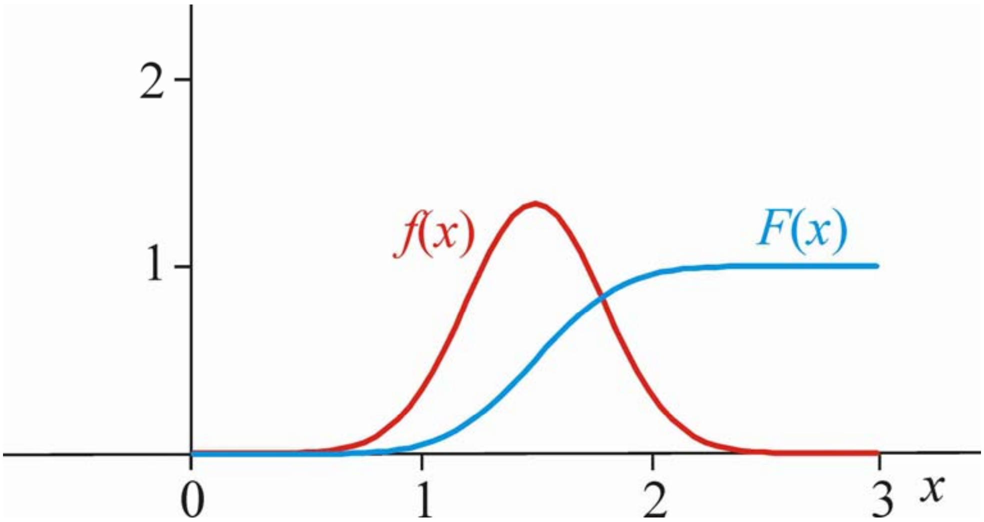


Példák: Várakozási idők.
Radioaktív bomlás során az egyes atomok élettartama.
Egy adott berendezés működési ideje (az első hibáig).

Normális eloszlás (Gauss-eloszlás)

$M(\xi) = \mu$, $N(\mu;\sigma)$

$D^2(\xi) = \sigma^2$ $N(1{,}5;0{,}3)$



Példák:

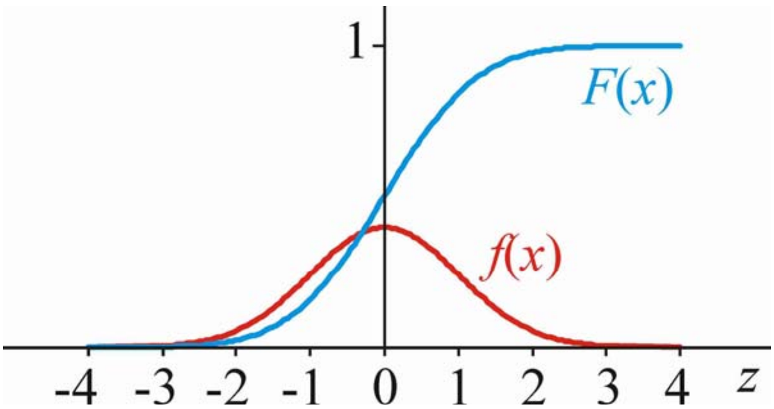
Magyarországon a felnőtt férfiak testmagassága cm-ben $N(171;7)$

Iskoláskorú fiúk diasztolés vérnyomása Hgmm-ben: $N(58;8)$

Standard normális eloszlás

$M(\xi) = 0$

$D^2(\xi) = 1$

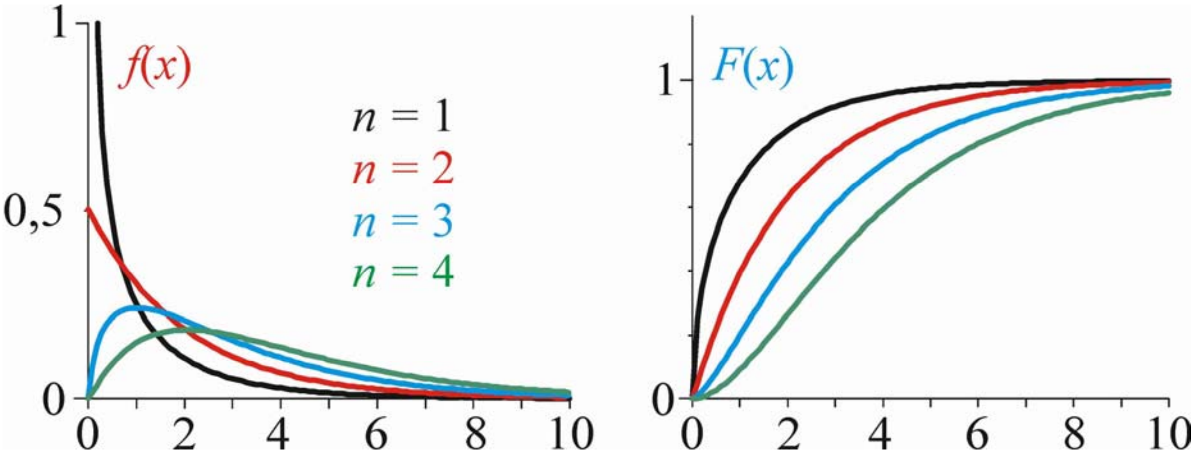


Transzformáció: x [N($\mu;\sigma$)] $\rightarrow$ z [N(0;1)] $$z = \frac{x-\mu}{\sigma}$$

Standard normális eloszlású változók (ξ_n) adott transzformáltjai eredményezik a χ^2-eloszlást és a t-eloszlást is.

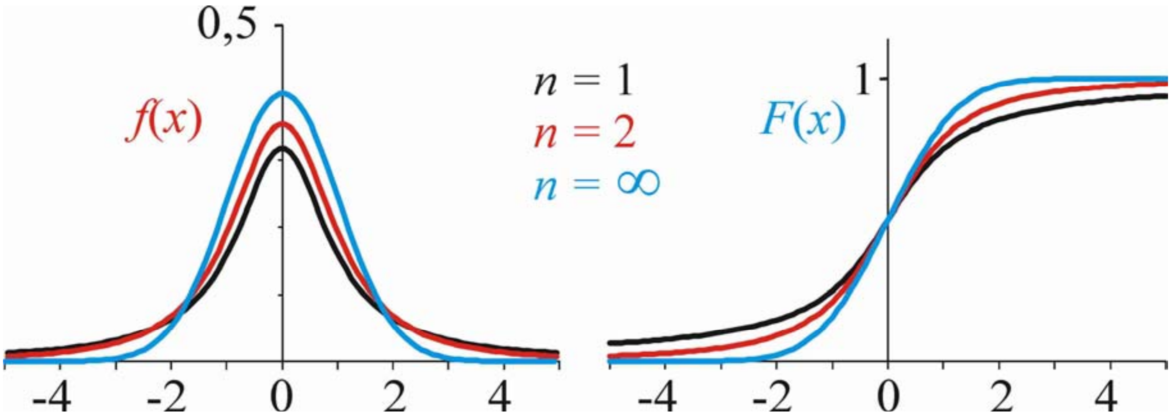
χ^2-eloszlás

$M(\eta_n) = n$ $\quad$ $D^2(\eta_n) = 2n$ $\quad$ $\eta_n = \xi_1^2 + \ldots + \xi_n^2$



t-eloszlás

$M(\zeta_n) = 0$ $\quad$ $D^2(\zeta_n) = n/(n-2)$ $\quad$ $\zeta_n = \dfrac{\sqrt{n}\,\xi}{\sqrt{\eta_n}}$

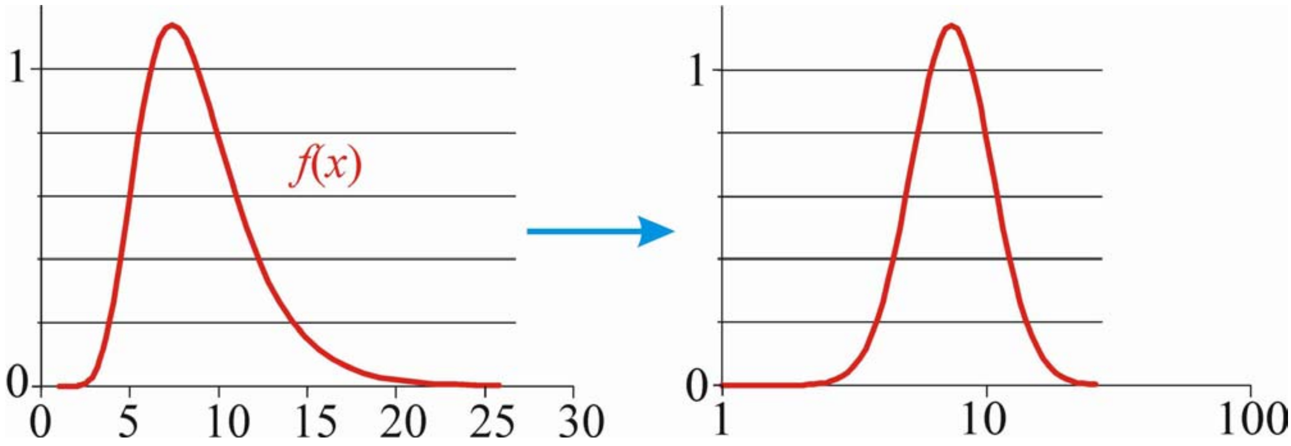


Ezekben az eloszlásokban n az ún. **szabadsági fok.** Ez a paraméter azzal van összefüggésben, hogy hány elemű a tanulmányozandó adathalmaz.

A t-eloszlás $n \to \infty$ határesetben a standard normális eloszlással megegyezik.

Lognormális eloszlás

ξ lognormális eloszlású, ha $\varphi = \ln\xi$ normális eloszlású



Miért kitüntetett a normális eloszlás?

Centrális határeloszlás-tétel

Ha egy valószínűségi változó sok egymástól független kis hatás összegződéseként áll elő, akkor az jó közelítéssel normális eloszlású.

Ki lehet próbálni!