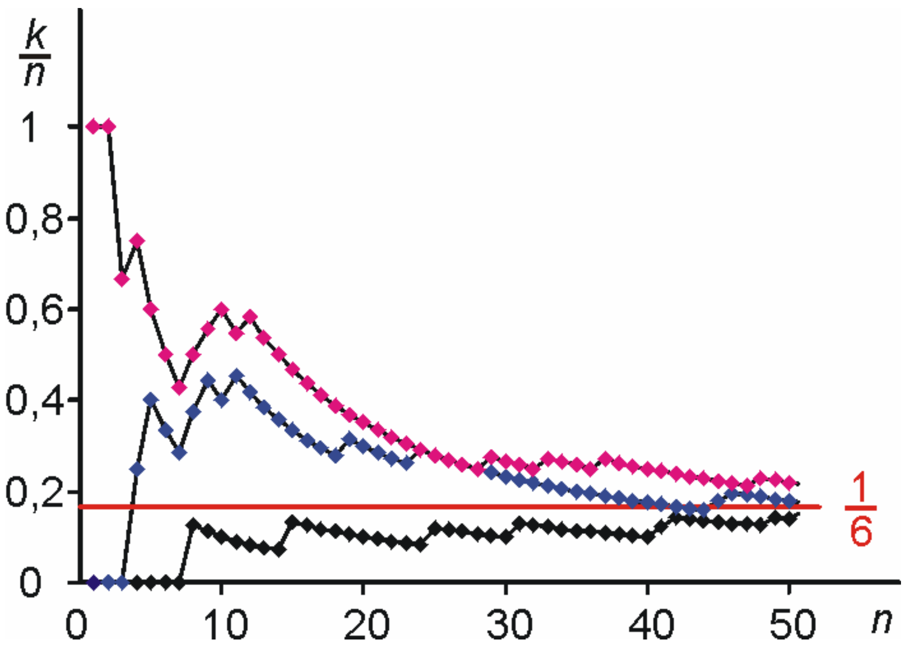
A valószínűségszámítás elemei

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.

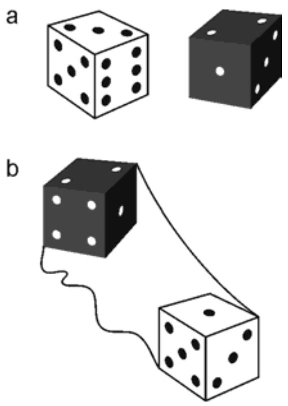
(Logikai úton bizonyítani nem lehet) (Karl Pearson 1857-1936)
Az eseményhez egy számot rendelhetünk: **valószínűség**

A valószínűség tulajdonságai:

1. Egy A esemény valószínűsége, $[P(A)]$ mindig $0 \le P(A) \le 1$.
2. A biztos esemény valószínűsége, $P(biztos) = 1$
3. Egymást kizáró események ($A \cap B = \varnothing$) egyesítésének valószínűsége: $P(A \cup B) \equiv P(A+B) = P(A) + P(B)$.

Függetlenség

1000 dobás



a

	1	2	3	4	5	6
1	30	25	30	29	28	25
2	24	27	31	27	24	27
3	28	30	39	32	24	29
4	28	28	22	26	27	33
5	27	24	26	21	31	27
6	30	25	32	30	29	25

b

	1	2	3	4	5	6
1	40	41	46	12	9	21
2	51	38	37	13	22	15
3	42	49	52	8	20	17
4	8	10	15	36	52	44
5	11	16	9	45	39	35
6	10	17	8	43	41	28

Feltételes valószínűség

Annak a valószínűsége, hogy a „fekete kockán 1 a dobott érték” (A esemény)
ha a „fehér kockán 1 a dobott érték” (B esemény)
$P(A|B)$: az A esemény B eseményre vonatkoztatott feltételes valószínűsége.

Ha $P(A|B) = P(A)$ akkor A esemény a B-től **független.**

Ha $P(A \cap B) \equiv P(AB)$ annak a valószínűsége, hogy A és B is bekövetkezik, akkor

$$P(A|B)\,P(B) = P(A \cap B) \qquad \text{(szorzási szabály)}$$

Függetlenség másképpen: $P(A)P(B) = P(A \cap B)$.

Feladat

Független-e az a két esemény, ha egy (ideális) kockával egyet dobunk, hogy az eredmény 3-nál kisebb (A esemény), illetve, hogy páros (B esemény)?

Nézzük meg, hogy annak a valószínűsége, hogy az A esemény és B esemény is bekövetkezik ugyanakkora-e, mint a két esemény külön-külön vett valószínűségeinek szorzata?

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:

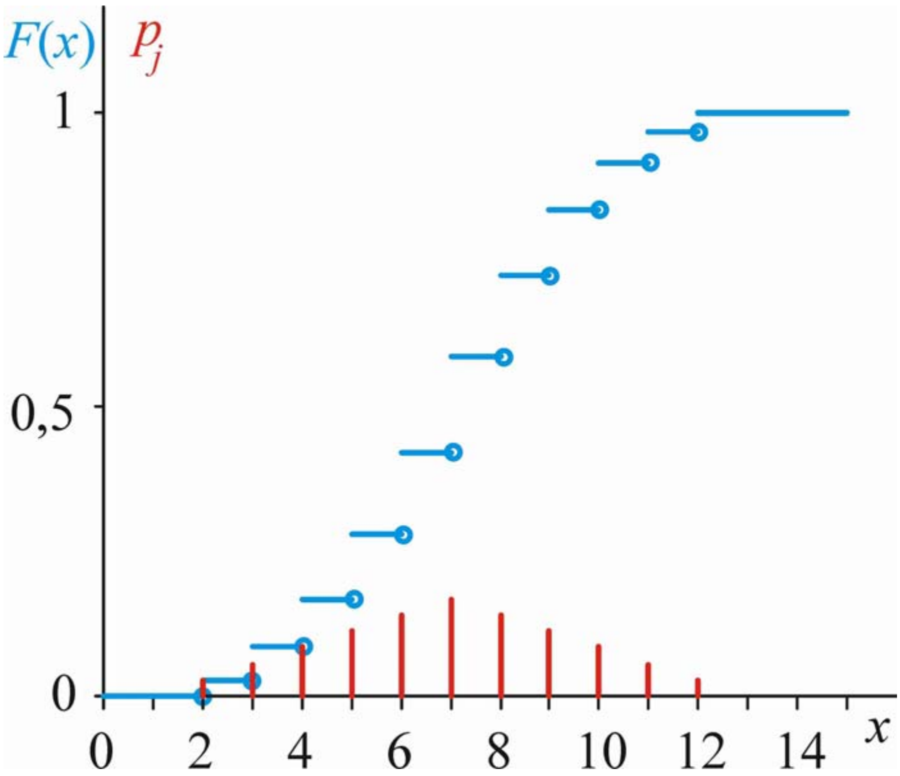
a lehetséges kimenetelek: $x_j = 2$ -től 12 -ig.

A dobás „eredménye” valamelyik lehetséges kimenetel.

Jellemzés:

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

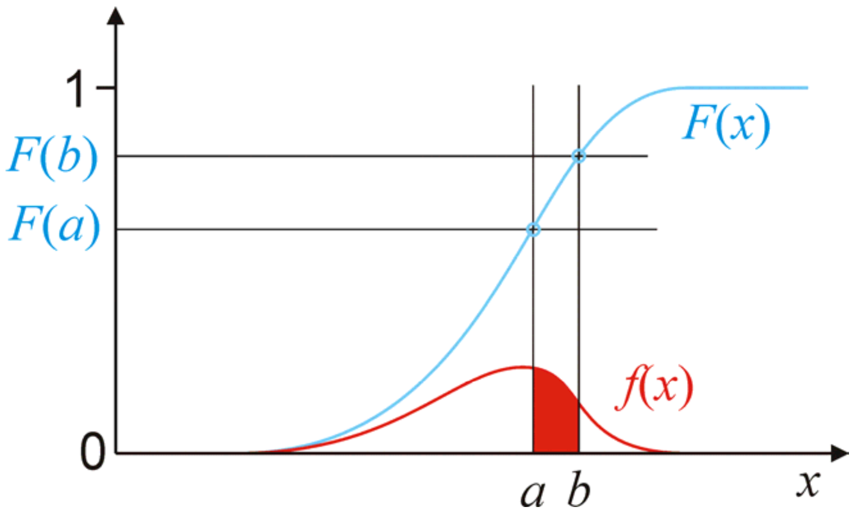
$$F(x) = p(\xi < x) = \sum_{x_j < x} p(\xi = x_j) \qquad p_j = p(\xi = x_j)$$



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 [$f(x)$]



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

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

Hol van az **eloszlás közepe**?

1a. **várható érték** [$M(\xi)$]

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

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

(kockadobás két kockával)

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

Kevés adat esetén nem látszanak az **adatrendszer** jellegzetességei. Számszerű jellemzők (mindig meghatározhatók):

Hol van az *n* elemű adatrendszer **közepe**? (középértékek)

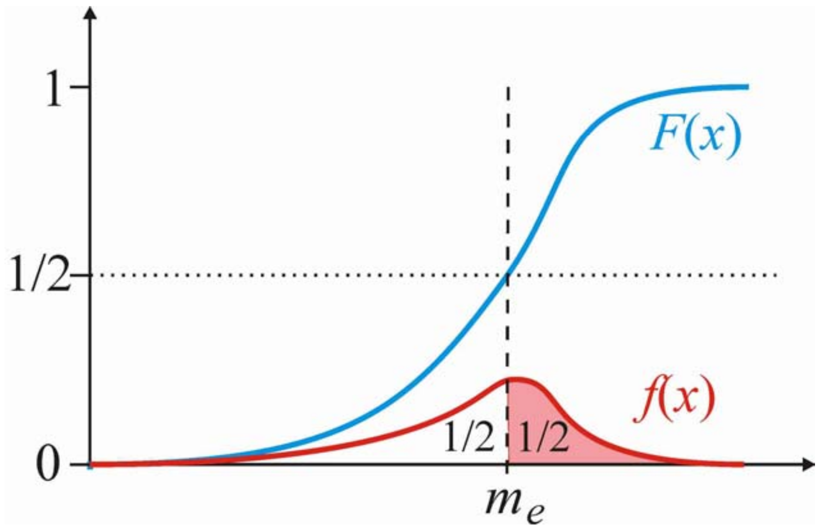
1b. az adatrendszer **átlaga** (számtani közép)

$$x_{\text{átlag}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{j=1}^{m} w_j x_j}{\sum_{j=1}^{m} w_j}$$

Érzékeny a kiugró értékekre!

2a. **medián** (m_e)

$F(m_e) = 1/2$



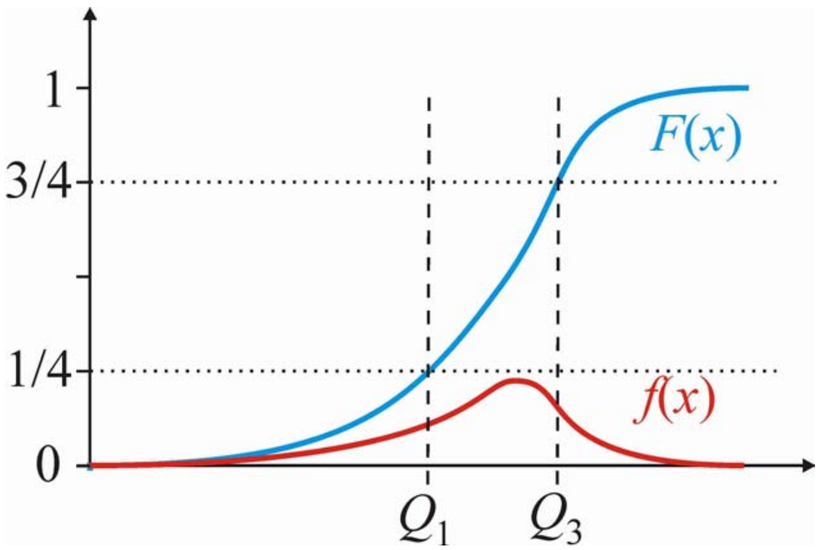
szemléltetése: két egyforma terület osztóértéke.

2b. az adatrendszer **mediánja** ($x_{medián}$)
nagyság szerint sorba rendezzük az adatokat és megkeressük a középsőt vagy középsőket

3a. **kvantilisek** (osztóértékek)
egyéb területarány osztóértékei (Q_1 alsó, Q_3 felső kvartilis)

$F(Q_1) = 1/4$

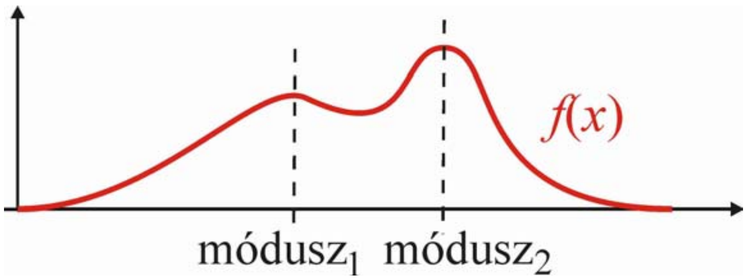
$F(Q_3) = 3/4$



3b. adatrendszerre Pl. Mekkora jövedelem esetén tartozik valaki a felső „tízezer”-be?

Az adatokat először itt is nagyság szerint sorba rendezzük.
Pl. alsó kvartilis, középső kvartilis = medián, felső kvartilis

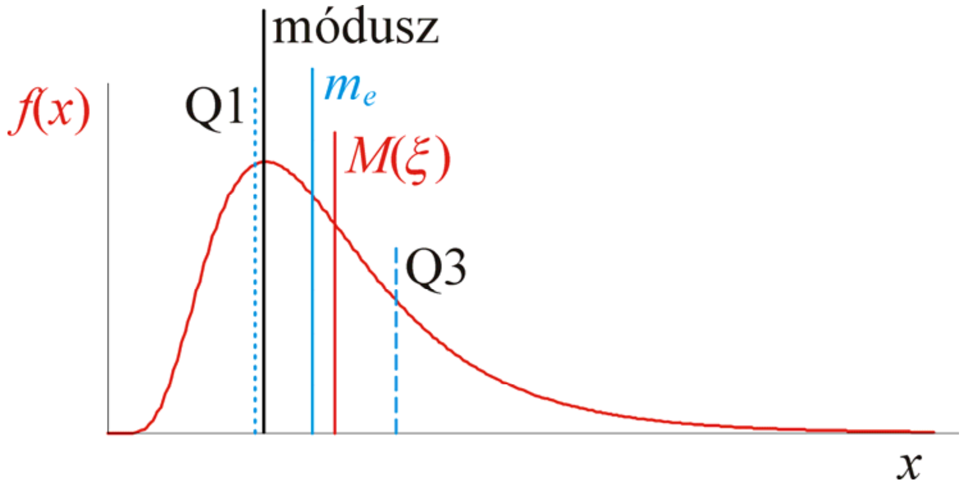
4a. **módusz(ok)**
a legvalószínűbb értéke(k),
a sűrűségfüggvény lokális maximum értéke(i)



4b. ha az adatrendszerben vannak azonosak, akkor azt, amelyikből a legtöbb van az adatrendszer **móduszának** ($x_{\text{módusz}}$) nevezzük (ebből lehet több is) („mode” divatos)

Nem érzékenyek a kiugró értékekre!

A „közép” számszerű jellemzőinek egymáshoz való viszonya:



Milyen **széles** az eloszlás?

1. **variancia** (szórásnégyzet).

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

Az adatrendszer szóródásának jellemzői

0. a legnagyobb és a legkisebb elem eltérése az adatrendszer **terjedelme**

1. az adatrendszer átlagától vett négyzetes eltérések átlagát az adatrendszer **szórásnégyzetének** vagy **varianciájának** nevezzük,

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

az adatrendszer **szórását** pedig az

$$s_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$ kifejezés adja meg

További jellemzők is megadhatók (ferdeségre, csúcsosságra).

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

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

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

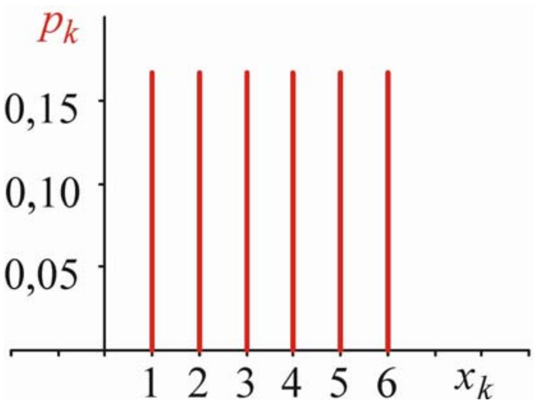
Ami az előadásból kimaradt, néhány példa eloszlásokra:

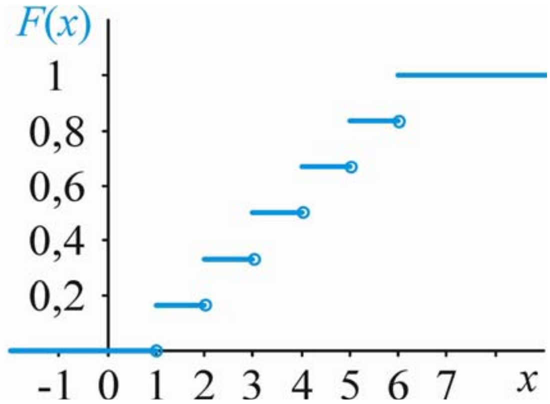
Nevezetes eloszlások (modellek)

1. Diszkrét eloszlások

Egyenletes eloszlás

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





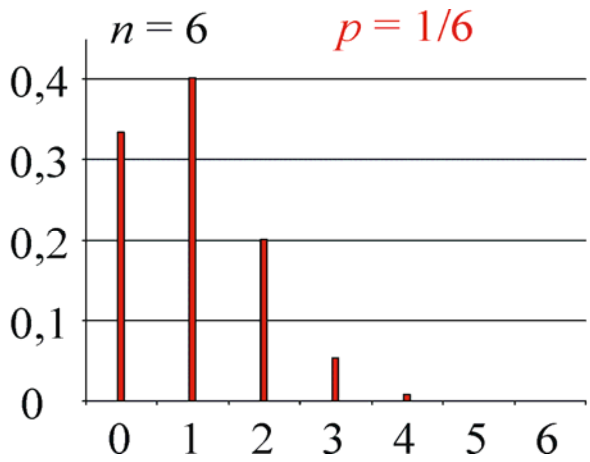
Binomiális eloszlás (Bernoulli-eloszlás)

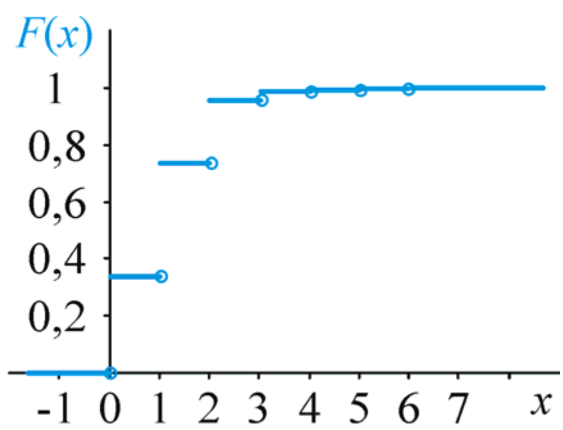
alternatíva p, $(1–p)$
n ismétlés $P(\xi = k) = B(n, k)$

Példa: dobókocka, 6 dobás, (n = 6)
Mi a valószínűsége annak, hogy k = 0-szor, 1-szer, 2-szer stb. dobok 6-ost? (p = 1/6)

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

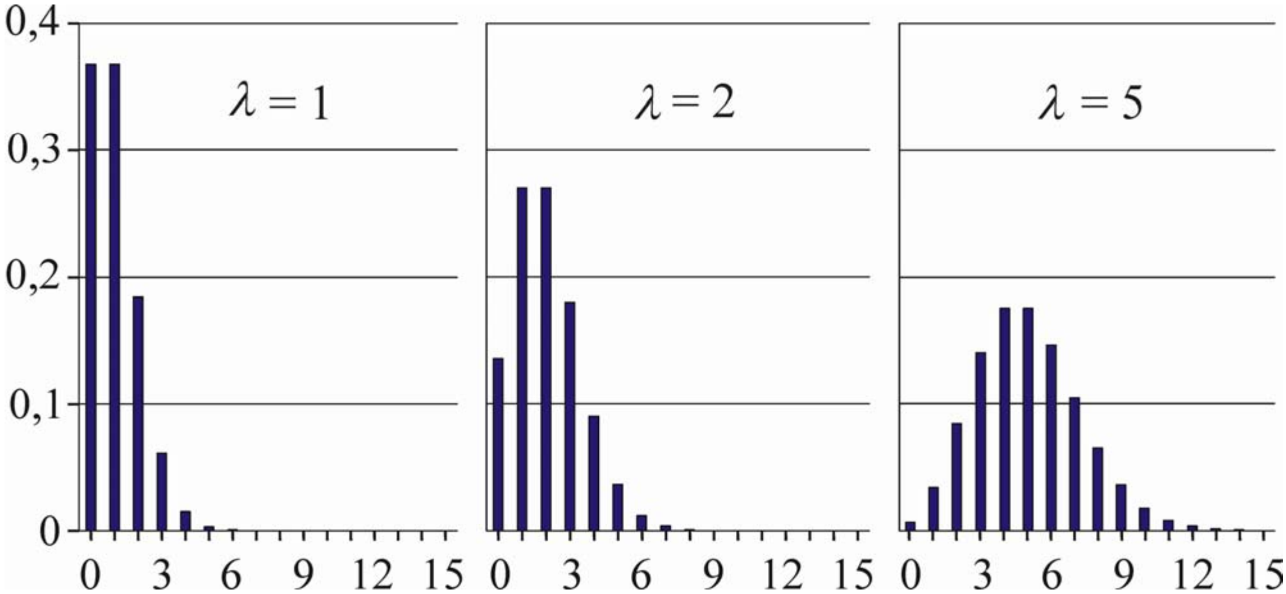
$M(\xi) = np$,
$D^2(\xi) = np(1–p)$





Poisson-eloszlás

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

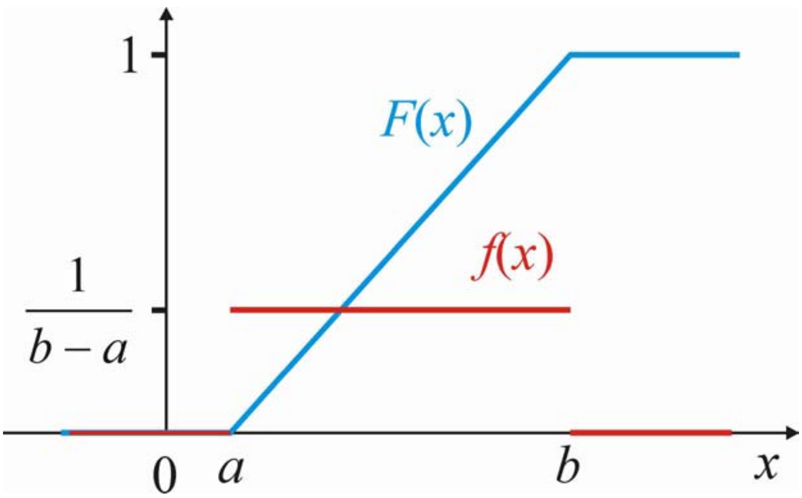


Példák:
Adott térfogatban lévő részecskék száma.
Radioaktív preparátumban adott idő alatt elbomló atomok száma.

2. Folytonos eloszlások

Egyenletes eloszlás

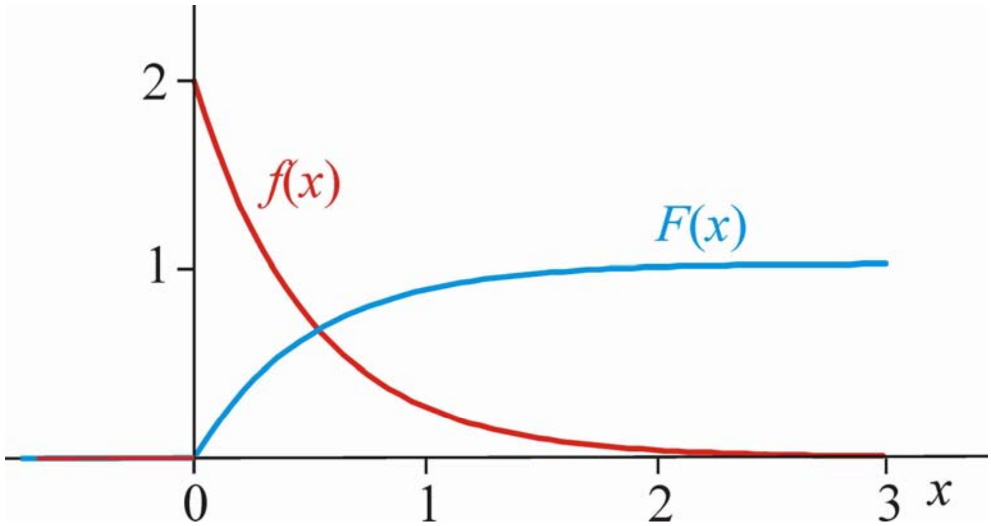
$M(\xi) = (a + b)/2$
$D^2(\xi) = (b - a)^2/12$



Példa: A teremben a levegő sűrűsége vagy hőmérséklete.

Exponenciális eloszlás

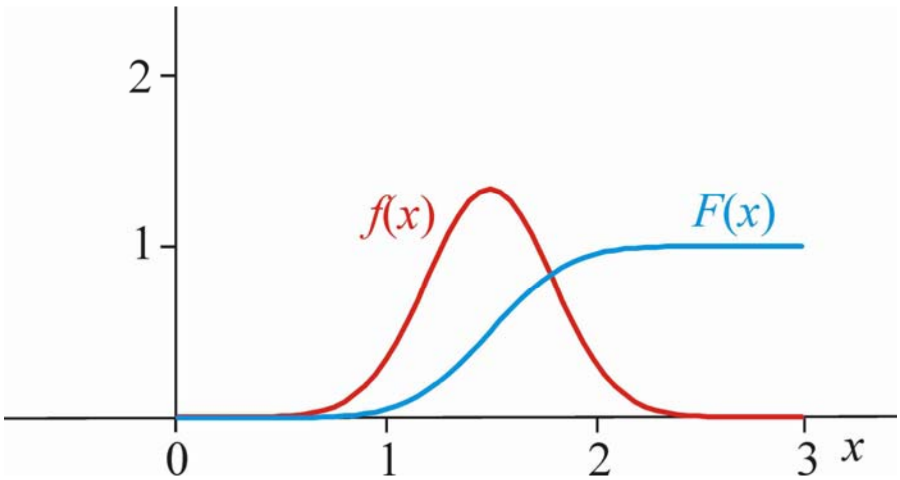
$M(\xi) = 1/\lambda$, $\quad (\lambda = 2)$
$D^2(\xi) = 1/\lambda^2$



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

Normális eloszlás (Gauss-eloszlás)

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



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(μ;σ)] $\rightarrow$ 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.

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!