

<http://math.uni-pannon.hu/~szalkai/walszam2021.pdf>

Valószínűesszámitás alapjai szemléletesen

/ Kézirat, 2021-08-07 /

dr.Szalkai István
Pannon Egyetem, Veszprém
Matematika Tanszék



<http://math.uni-pannon.hu/~szalkai>

szalkai@almos.uni-pannon.hu

Bevezetés

Ez csak egy nagyon vázlatos, félkész kézirat, segítség az előadáshoz .

A kisebb egységek végét □ jelzi.

További segédanyagok találhatóak még a honlapomon: <http://math.uni-pannon.hu/~szalkai>

dr. Szalkai István

Veszprém, 2021.08.07.

szalkai@almos.uni-pannon.hu

Tartalom

Bevezetés	2.
1. Események és eseménytér	3.
2. A relatív gyakoriság és a valószínűség	4.
3. A valószínűség kiszámítása	5.
4. Feltételes valószínűség, események függetlensége	6.
5. Valószínűségi változók és jellemzőik	8.
Valószínűségi változók függetlensége	12.
Valószínűségi változók függvényei	12.
6. Várható érték és szórás	13.
7. Nevezetes diszkrét eloszlású valószínűségi változók	15.
7.1. Diszkrét egyenletes eloszlású v.v.	15.
7.2. Hipergeometrikus eloszlású v.v.	15.
7.3. Binomiális- vagy Bernoulli- eloszlású v.v.	16.
7.4. Poisson eloszlású v.v.	18.
7.5. Geometriai eloszlású v.v.	19.
8. Nevezetes folytonos eloszlású valószínűségi változók	20.
8.1. Folytonos egyenletes eloszlású v.v.	20.
8.2. Exponenciális eloszlású v.v.	20.
9. Normális- és származtatott eloszlású v.v.	22.
9.1. Normális eloszlások	22.
9.2. Normális eloszlásból származtatott eloszlások	25.
Móduszok és mediánok összefoglalása	26
Összefoglaló táblázat	27.
10. Nagy számok törvényei	28.
Függelék	
Valószínűségszámítás - Matematika szótár	31.
A standard normális v.v. eloszlásfüggvény Φ táblázat	32.
Közelítő képletek a Φ és φ függvényekre	34.
Larson nomogram	35.

1. Események és eseménytér

1.1. Alapfogalmak (Definíciók):

Kísérlet: egy jelenség aktív vagy passzív megfigyelése

Determinisztikus (=meghatározott) kísérlet: mindenképpen csak egy lehetséges kimenetele (végeredménye) lehet.

Sztocasztikus (=véletlen) kísérlet: több kimenetele van

Példák: érme-, kockadobás, két kocka, izzó élettartama, fizikai mennyiségek mérése, vízállások, lottó-húzás, minőségellenőrzés, stb.

Elemi események: A kísérlet lehetséges, tovább már nem bontható kimenetelei. Az összes lehetséges elemi esemény *halmazát* (összességét) **eseménytérnek** nevezzük, mi Ω -val jelöljük. (Más könyvekben eltérő jelölés is lehet). A kísérlet egy elemi eseménye Ω -nak egy eleme: $\omega \in \Omega$ vagy $x \in \Omega$.

Matematikailag Ω egy *tetszőleges*, nem üres halmaz lehet!

Példák: "hatost dobtam", "3V", "2kg", ...

Esemény (általánosan): Ω bármely *részhalmaza*: $A \subseteq \Omega$.

Példák: "páros számot dobtam", "10V és 20V között van", ...

Az elemi eseményeket tehát Ω egyelemű részalmazainak kell (illik) tekintenünk: $\omega \in \Omega$ helyett inkább $\{\omega\} \subseteq \Omega$ írandó.

Az események összessége (halmaza) nyilvánvalóan Ω **hatványhalmaza** (power set), aminek jól ismert jele $P(\Omega)$.

A kísérlet (vég)**eredménye** tehát mindig Ω egy eleme: $\omega \in \Omega$. Azt mondjuk, hogy egy $A \subseteq \Omega$ esemény egy kísérlet (elvégzése) során **bekövetkezik**, ha $\omega \in A$.

Az alábbi fogalmak szemléletesek, de meg kell gondolnunk matematikai háttérüket. (Ezeket a fogalmakat a későbbiekben még általánosítjuk.)

Biztos esemény, ami mindig bekövetkezik tehát az egész alaphalmaz Ω .

Lehetetlen esemény, ami nem következhet be, tehát az üres halmaz \emptyset .

Egymást kizáró események csak diszjunktak lehetnek, vagyis $A \cap B = \emptyset$.

Az A eseményből **következik** a B esemény, vagy más szóval A **maga után vonja** B -t, ha minden $\omega \in \Omega$ esetén $\omega \in A$ -ből $\omega \in B$ következik. Ez pedig éppen azt jelenti, hogy $A \subseteq B$, vagyis A **részhalmaza** B -nek.

1.2. Műveletek eseményekkel (eseményalgebra)

Mivel az események valójában Ω részalmazai, ezért a szokásos halmazműveletekről van szó: unió, metszet, komplementer, különbség. A valószínűségszámítás elméletében más elnevezések és jelölések használatosak:

események összege	$A+B =$	$A \cup B =$	únió,
események szorzata	$A \cdot B =$	$A \cap B =$	metszet,
események különbsége	$A-B =$	$A \setminus B =$	különbség,
A esemény tagadása	$\bar{A} =$	$\bar{A} =$	komplementer.

Mi legtöbbször a hagyományos, halmaz- elnevezéseket használjuk.

1.3. A műveletek tulajdonságai

A fentiek alapján a halmazműveletek (Boole-algebrák) jól ismert tulajdonságairól van szó. Tanácsoljuk az Olvasónak az egyenlőségeket a valószínűségszámítás nyelvére lefordítani.

Boole-algebrák axiómái

$$\mathfrak{B} = (X, \cup, \cap, \bar{}, \emptyset, I)$$

kommutativitás	$A \cup B = B \cup A$	(BA1)
	$A \cap B = B \cap A$	(BA2)
asszociativitás	$A \cup (B \cap C) = (A \cup B) \cap C$	(BA3)
	$A \cap (B \cup C) = (A \cap B) \cup C$	(BA4)
disztributivitás	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	(BA5)
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	(BA6)
elnyelési tulajdonságok	$A \cup (A \cap B) = A$	(BA7)
	$A \cap (A \cup B) = A$	(BA8)
\emptyset és I tulajdonságai	$A \cup \bar{A} = I$	(BA9)
	$A \cap \bar{A} = \emptyset$	(BA10)
	$A \cup \emptyset = A$	(BA11)
	$A \cap \emptyset = \emptyset$	(BA12)
	$A \cup I = I$	(BA13)
	$A \cap I = A$	(BA14)

1. Táblázat: Eseményalgebra tulajdonságai

A fenti axiómák következményei például az alábbiak:

De Morgan-azonosságok: $\overline{A \cup B} = \bar{A} \cap \bar{B}$ és $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

2. A relatív gyakoriság és a valószínűség

2.0. Ha egy kísérletet a gyakorlatban n -szer elvégezzük, és egy A esemény k -szor következett be (**gyakorisága k**), akkor a k/n hányadost **relatív gyakoriságnak** hívjuk.

Tapasztalatok alapján hihető, hogy ha n vagy, akkor a k/n hányados egy bizonyos, az A eseményre jellemző p szám körül ingadozik. (Ezt az összefüggést a *Nagy számok törvényei* fejezetben Bernoulli tétele igazolja.) Ezt a p számot hívjuk az A esemény **valószínűségének**, és $P(A)$ -val jelöljük.

2.1. Definíció: A valószínűség Kolmogorov-féle axiómái:

P (egy) valószínűség az Ω eseménytéren, ha

- (o) $P: P(\Omega) \rightarrow \mathbf{R}$, vagyis $A \subseteq \Omega$ esetén $P(A) \in \mathbf{R}$ valós szám,
- (i) $0 \leq P(A) \leq 1$,
- (ii) $P(\emptyset) = 0$, $P(\Omega) = 1$,
- (iii) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ ha az A_i -k egymást páronként kizárják, vagyis ha $A_i \cap A_j = \emptyset$. \square

2.2. A fenti axiómák következményei:

1. tetszőleges $A, B \subseteq \Omega$ eseményekre $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
2. ha A és B kizárják egymást, akkor $P(A \cup B) = P(A) + P(B)$ (additivitás),

3. $P(\bar{A}) = 1 - P(A)$,
4. tetszőleges $A, B \subseteq \Omega$ eseményekre $P(A \setminus B) = P(A) - P(A \cap B)$,
5. ha $B \subset A$, akkor $P(A \setminus B) = P(A) - P(B)$,
6. ha $B \subset A$, akkor $P(B) \leq P(A)$ (monotonitás),
7. tetszőleges $A, B, C \subseteq \Omega$ eseményekre
 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$
(logikai szitaformula) ,
8. ... □

2.3. Vegyük észre, hogy P fenti tulajdonságai nagyon hasonlítanak a síkbeli halmazok *területé*-nek tulajdonságaival (ami nem meglepő, mert mindkettő *additív halmazfüggvény*). Tehát tanuláskor nyugodtan olvashatunk/gondolhatunk $P(A)$ helyett T_A -t !

Az 1.1. Definíció *általánosításaként* mondjuk az alábbiakat:

- 2.4. Definíció:** Tetszőleges $A, B \subseteq \Omega$ eseményekre:
 A **biztos esemény**, ha $P(A) = 1$,
 A **lehetetlen esemény**, ha $P(A) = 0$,
 A és B **egymást kizáró események**, ha $P(A \cap B) = 0$. □

2.5. Megjegyzés: Általánosan az "*esély*" és "*valószínűség*" szavak egymás szinonímái, de egyes statisztikai művek (pl. egészségügyben) az **esély** fogalom alatt mást értenek: a $p/q = p/(1-p)$ hányadost.

3. A valószínűség kiszámítása

Most csak a két legegyszerűbb módszert ismertetjük, hiszen az egész anyag célja végig a valószínűség kiszámítása !

3.1. Kombinatorikus (klasszikus) valószínűségi mező

Amennyiben Ω véges elemszámú halmaz, és minden $\omega \in \Omega$ elemi eseménynek *ugyanakkora* a valószínűsége (pl. szabályos kockával dobás), akkor

egyrészt $P(\{\omega\}) = \frac{1}{n}$ minden elemi esemény esetén,

másrészt tetszőleges $A \subseteq \Omega$ eseményre $P(A) := \frac{|A|}{|\Omega|} = \frac{\text{kedvező}}{\text{összes}}$. □

3.2. Geometriai valószínűség:

Amennyiben a Ω halmaz a számegegyenes, a sík vagy a tér valamely olyan részhalmazával reprezentálható (modelllezhető, megjeleníthető), amely esetén minden $\omega \in \Omega$ elemi eseménynek *ugyanakkora* a valószínűsége, akkor tetszőleges $A \subseteq \Omega$ eseményre

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} \quad (\text{arányos a halmaz hosszával/területével/térfogatával}) ,$$

ahol értelemszerűen $\mu(A)$ és $\mu(\Omega)$ a halmazok hosszát, területét vagy térfogatát jelöli. □

3.3. Például: buszra várok (de csak véletlenszerűen mentem ki a megállóba),

célbadobás (de csak véletlenszerűen célzok, nem vagyok profi, és minden lövésemmel legalább a céltáblát eltalálom).

Érdeemes még a "*Randevú a könyvtárban*" -típusú feladatokat is tanulmányozni.

A "*célbadobás*" -nál kihangsúlyozott feltételeket minden esetben alaposan ellenőrizni kell !

4. Feltételes valószínűség, események függetlensége

Ha már tudjuk, hogy valami bekövetkezett (egy B esemény), akkor ez mennyire befolyásolhatja az általunk vizsgált A esemény bekövetkezését? A relatív gyakoriság elemzésével kapjuk a következő képletet:

4.1. Definíció: Amennyiben a B esemény nem lehetetlen, vagyis $P(B) > 0$, akkor A -nak B bekövetkezése utáni valószínűsége $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Tehát $P(A|B)$ az A esemény valószínűsége, **feltéve**, hogy B már bekövetkezett, ezért hívjuk **feltételes valószínűségnek**. □

4.2. Megjegyzések: A *feltételes valószínűség* teljesíti a valószínűség axiómáit, ha a B feltétel rögzített. Ez azt jelenti, hogy a 2. Fejezet képletei mind igazak maradnak, ha $P(\dots)$ helyett mindenütt $P(\dots|B)$ -t írunk.

Természetesen merül fel a kérdés: B hogyan befolyásolja A -t (erősíti, gyengíti vagy nem befolyásolja). Ezt alább, a *Függetlenség* alfejezetben vizsgáljuk.

A 4.1. Definíció képletének átszorozása után kapjuk a következő egyszerű, de fontos összefüggést:

4.3. Szorzás Tétel: $P(A \cap B) = P(A|B) \cdot P(B)$. □

4.4. Definíció: A $B_1, B_2, B_3, \dots, B_n \subseteq \Omega$ események **teljes eseményrendszert** alkotnak, ha egymást páronként kizárják, és uniójuk a biztos esemény, vagyis $B_i \cap B_j = \emptyset$, és

$$B_1 \cup B_2 \cup \dots \cup B_n = \Omega. \quad \square$$

(Általánosabban: $P(B_i \cap B_j) = 0$ és $P(B_1 \cup B_2 \cup \dots \cup B_n) = 1$.)

A fenti tulajdonságú halmazrendszereket általában *partíciónak* (felosztásnak) nevezünk, lásd az alábbi ábrát.

4.5. Teljes valószínűség tétele:

Ha $\{B_1, B_2, B_3, \dots, B_n\}$ teljes eseményrendszer és $P(B_i) > 0$ minden i -re, akkor bármely A eseményre

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

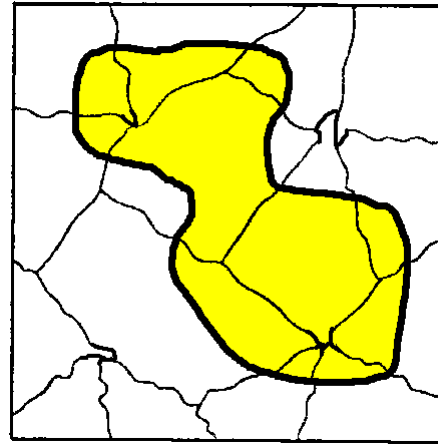
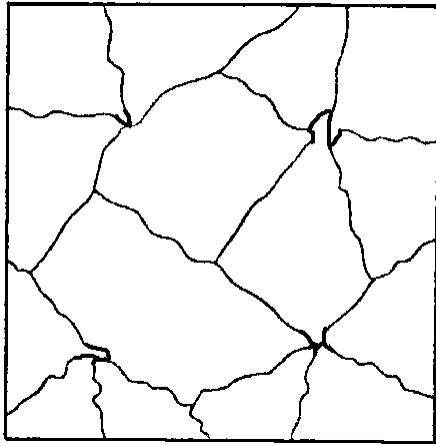
Bizonyítás: A Szorzás Tétel alapján a fenti képlet így írható:

$$P(A) = \sum_{i=1}^n P(A \cap B_i)$$

ami nyilván igaz, hiszen

$$A = \bigcup_{i=1}^n (A \cap B_i). \quad \square$$

A gondolatmenetet szemléltetik az alábbi ábrák (P helyett gondoljunk ismét a területre):



1. ábra: Teljes eseményrendszer (partíció)

Teljes valószínűség

4.6. Példa: Egy gyárban 3 műszakban gyártanak termékeket. Az I. műszak az összes termék 40%-át, a második az összes termék 35%-át, a harmadik műszak az összes termék 25%-át gyártja. Az első műszak által gyártott termékek mindegyike 0.05, a második műszak által gyártott termékek mindegyike 0.06, a harmadik műszak által gyártott termékek mindegyike 0.07 valószínűséggel selejtes. Kiválasztva egy akármilyen terméket, mennyi az esélye, hogy selejtes?

Megoldás: Legyenek B_1, B_2, B_3 rendre azon események, hogy a kiválasztott terméket melyik műszak gyártotta, és jelölje S azt, amikor a termék selejtes. A feltételek szerint $P(B_1)=0.4$, $P(B_2)=0.35$, $P(B_3)=0.25$, (ellenőrzés: $P(B_1)+P(B_2)+P(B_3)=0.4+0.35+0.25=1$). Továbbá $P(S|B_1)=0.05$, $P(S|B_2)=0.06$, $P(S|B_3)=0.07$. A teljes valószínűség tétele szerint

$$\begin{aligned} P(S) &= P(S|B_1) \cdot P(B_1) + P(S|B_2) \cdot P(B_2) + P(S|B_3) \cdot P(B_3) = \\ &= 0.05 \cdot 0.4 + 0.06 \cdot 0.35 + 0.07 \cdot 0.25 = 0.0585 . \end{aligned}$$

Megfordítás: Ha a kiválasztott termék selejtes, mennyi az esélye, hogy az első (vagy a második, vagy a harmadik) műszak gyártotta? Kit szidjunk (legnagyobb eséllyel)? Mert ugye a harmadik műszak termékei között van a legtöbb selejt, de ugyanekkor a harmadik műszak állítja elő a legkevesebb terméket Az alábbi tétel erre a kérdésre válaszol:

4.7. Bayes tétele (Megfordítási tétel): Ha $P(A)>0$ és $P(B)>0$, akkor

$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A)} .$$

Bizonyítás: könnyen következik a Szorzás Tételből, hiszen

$$P(B | A) \cdot P(A) = P(B \cap A) = P(A \cap B) = P(A | B) \cdot P(B)$$

és mindkét oldalt osztjuk $P(A)$ -val . □

Megjegyzés: A szakirodalomban sok helyen a fenti képlet nevezőjében $\sum_{i=1}^n P(A | B_i) \cdot P(B_i)$ szerepel, de ez a 4.5.Tétel szerint éppen $P(A)$.

A 4.6.Példa folytatása:

$$P(B_1|S) = P(S|B_1) \cdot P(B_1) / P(S) = 0.05 \cdot 0.4 / 0.0585 \approx 0.341880 ,$$

$$P(B_2|S) = P(S|B_2) \cdot P(B_2) / P(S) = 0.06 \cdot 0.35 / 0.0585 \approx 0.358974 ,$$

$$P(B_3|S) = P(S|B_3) \cdot P(B_3) / P(S) = 0.07 \cdot 0.25 / 0.0585 \approx 0.299145 ,$$

tehát (kis eltéréssel) a selejtes darabok legnagyobb része a második műszaktól származik.

(Ellenőrzés: $P(B_1|S)+P(B_2|S)+P(B_3|S)=1$.)

Események függetlensége

Felmerül a kérdés, hogy a már bekövetkezett B mennyire befolyásolja A -t? Nyilván három eset lehetséges:

- $P(A|B) < P(A)$, vagyis B gyengíti A -t,
- $P(A|B) > P(A)$, vagyis B erősíti A -t,
- $P(A|B) = P(A)$, vagyis B nem befolyásolja A -t.

4.8. Speciális esetek: Az 1. Fejezetben megismert speciális $B \subseteq A$ ("B -ből következik A") esetben a 4.1. Definíció képlete szerint

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \text{ vagyis ekkor } B \text{ -ből valóban következik } A.$$

Ha pedig kizárják egymást (a 2.4. Definíció szerint), akkor

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0, \text{ vagyis } B \text{ valóban kizárja } A \text{ -t.}$$

4.9. Nyilvánvaló követelmény, hogy az A és B események akkor lehetnek *függetlenek*, ha egyik sem befolyásolja a másikat, vagyis

$$P(A|B) = P(A) \text{ és } P(B|A) = P(B).$$

Kis számolás után (a Szorzás tétel segítségével) kapjuk, hogy a fenti két követelmény ekvivalens (azonos értékű) a következő, egyetlen összefüggéssel:

4.10. Definíció: Az A és B események **egymástól függetlenek**, ha $P(A \cap B) = P(A) \cdot P(B)$. □

Hangsúlyozzuk, hogy a fenti egyenlőség nem érvényes tetszőleges eseményekre !!!

4.11. Állítás: Ha $0 < P(A) < 1$, $0 < P(B) < 1$ és A és B függetlenek egymástól, akkor \bar{A} és B is, A és \bar{B} is, valamint \bar{A} és \bar{B} is függetlenek egymástól. □

5. Valószínűségi változók és jellemzőik

A legtöbb kísérletnél nem csak "esemény"-t észlelünk (piros, sikerült, stb.), hanem mérünk is (valamit). Általában többször megmérve "ugyanazt" a jelenséget (pl. testtömeg), mindig más eredményt kapunk, véletlenszerűen váltakozva.

5.1. Definíció: Az olyan ξ függvényeket, amelyek elemi eseményekhez rendelnek valós számokat, azaz $\xi: \Omega \rightarrow R$, **valószínűségi változóknak** (röviden: v.v.) nevezzük. □

(Könnyen megjegyezhető: "val.vált." = "a kísérlet számszerű végeredménye".)

A valószínűségi változók alábbi két típusának megkülönböztetése lényeges az anyag további részében: minden definíciónál, tételnél lényeges, hogy az milyen típusú val.vált.-ra érvényes!

5.2. Definíció: A ξ valószínűségi változó **diszkrét** (elkülönült), ha csak véges vagy megszámlálhatóan (felsorolhatóan) sok lehetséges értéke van, tehát értékészlete $Im(\xi) = \{x_1, x_2, \dots, x_n, \dots\}$ alakban írható.

A ξ valószínűségi változó **folytonos**, ha értékészlete tartalmaz egy intervallumot: $Im(\xi) \supseteq (a, b)$. □

5.3. Definíció: A ξ *diszkrét* valószínűségi változó **eloszlásán** értjük a lehetséges értékeinek halmazát a hozzájuk tartozó valószínűségekkel: $\{x_1, x_2, \dots, x_n, \dots\}$ és $\{p_1, p_2, \dots, p_n, \dots\}$ ahol $p_i := P(\xi = x_i)$. □

Nyilván ξ és η -nak akkor van **azonos eloszlása**, ha $\{x_1, x_2, \dots, x_n, \dots\}$ és $\{p_1, p_2, \dots, p_n, \dots\}$ ugyanazok.

5.4. Állítás: Egy $\{p_1, p_2, \dots, p_n, \dots\}$ számsorozat akkor és csak akkor egy diszkrét valószínűségi változó eloszlása, ha az alábbi **axiómákat** teljesíti:

- (i) $0 \leq p_i \leq 1$, (ii) $p_1 + p_2 + \dots + p_n + \dots = 1$. □

5.5. Definíció: Egy tetszőleges ξ (akár diszkrét, akár folytonos) valószínűségi változó (kumulatív) **eloszlásfüggvénye** $F: R \rightarrow R$, ahol $F(x) := P(\xi < x)$. □

Megjegyzések: Ne feledjük, hogy a definícióban szigorú $<$ van, ami diszkrét valószínűségi változók esetén lényeges!

A *kumulatív* ("összegyűjtött", lásd akkumulátor) jelző arra utal, hogy ξ valószínűségeit összegyűjtjük egészen x -ig. (Hasonlóan ahhoz, amikor a szeptemberi fizetési kimutatás tartalmazza az adott évi, azaz januártól szeptemberig összesített adatokat.)

5.6. Tétel: Az eloszlásfüggvény alaptulajdonságai (axiómák):

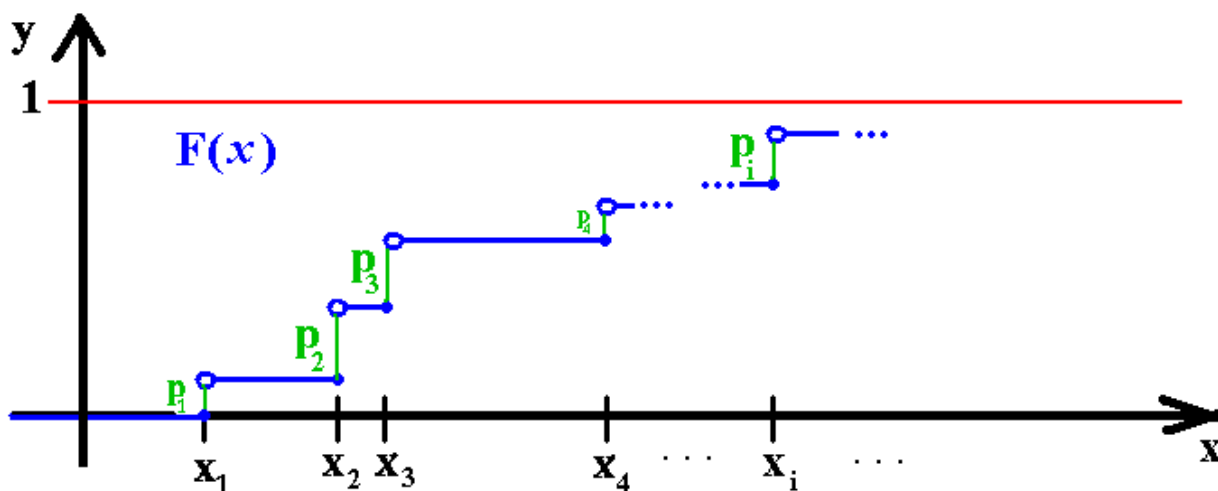
- 1) $0 \leq F(x) \leq 1$,
- 2) $F(x)$ monoton nő, azaz $x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$,
- 3) $F(x)$ balról folytonos (azaz szakadási helyeknél "teli karika" a görbe jobb végén),
- 4) $\lim_{x \rightarrow \infty} F(x) = 1$ és $\lim_{x \rightarrow -\infty} F(x) = 0$. □

Mire való az *eloszlásfüggvény*? Segítségével az alábbi kérdésekre tudunk gyors választ adni:

5.7. Tétel:

$P(\xi < a) = F(a)$	$P(\xi \geq a) = 1 - F(a)$
$P(\xi \leq a) = F(a) + P(\xi = a)$	$P(\xi > a) = 1 - F(a) - P(\xi = a)$
$P(a \leq \xi < b) = F(b) - F(a)$	$P(a \leq \xi \leq b) = F(b) - F(a) + P(\xi = b)$
$P(a < \xi < b) = F(b) - F(a) - P(\xi = a)$	$P(a < \xi \leq b) = F(b) - F(a) + P(\xi = b) - P(\xi = a)$
$P(\xi = a) = \lim_{x \rightarrow a^+} F(x) - F(a)$ □	

Érdekes átgondolnunk a *diszkrét* v.v. -k *eloszlásfüggvényének* ábrázolását. Legyen tehát ξ eloszlása $\{x_1, x_2, \dots, x_n, \dots\}$ és $\{p_1, p_2, \dots, p_n, \dots\}$. Ekkor $x \leq x_1$ esetén $F(x) = P(\xi < x_1) = 0$, és $x_{i-1} < x \leq x_i$ esetén $F(x) = P(\xi < x_i) = p_1 + \dots + p_{i-1}$. Tehát *diszkrét* v.v. $F(x)$ eloszlásfüggvénye vízszintes szakaszokból áll, ún. (emelkedő) *lépcsős függvény*. Amennyiben $\text{Im}(\xi)$ véges (azaz x_n és p_n az utolsó), akkor $x_n < x$ esetén $F(x) = 1$, $\text{Im}(\xi)$ *végtelen* esetén pedig $F(x)$ végtelen sok fokból álló lépcső, természetesen $F(x) < 1$.



2. ábra: Lépcsős függvény

Az 5.2. Definíció helyett matematikailag az alábbi követelmény pontosabb:

5.8. Definíció: ξ **folytonos**, ha létezik egy olyan $f: \mathbb{R} \rightarrow \mathbb{R}$ folytonos függvény, amelyre (legfeljebb véges sok pont kivételével)

$$F(t) = \int_{-\infty}^t f(x) dx .$$

Ekkor $f(x)$ -et ξ **sűrűségfüggvényének** nevezzük. □

A ξ és η v.v. **azonos eloszlásúak**, ha sűrűségfüggvényeikre $f_{\xi}(x)=f_{\eta}(x)$ véges sok $x \in \mathbb{R}$ pont kivételével. □

A "sűrűségfüggvény" elnevezést az 5.15. megjegyzésben világítjuk meg.

5.9. Állítás: A sűrűségfüggvény alaptulajdonságai (**axiómái**):

$$0) f: \mathbb{R} \rightarrow \mathbb{R} \qquad 1) f(x) \geq 0 \qquad 2) \int_{-\infty}^{\infty} f(x) = 1 . \qquad \square$$

5.10. Állítás: ha az $f: \mathbb{R} \rightarrow \mathbb{R}$ (legfeljebb véges sok pont kivételével) folytonos függvény rendelkezik a fenti tulajdonságokkal, akkor van olyan ξ valószínűségi változó, aminek f a sűrűségfüggvénye. □

5.11. Állítás: (i) Ha ξ folytonos, akkor eloszlásfüggvénye ($F(x)$) minden pontban folytonos.

(ii) Ha $F(x)$ folytonos és véges sok pont kivételével folytonosan differenciálható, akkor van ξ folytonos eloszlású valószínűségi változó, amelynek $F(x)$ eloszlásfüggvénye.

(iii) Ahol F deriválható, ott $F'(x) = f(x)$. □

5.12. Állítás: Ha $F(x)$ folytonos x -ben, akkor $P(\xi = x) = 0$. □

Mire való az $f(x)$ sűrűségfüggvény? Segítségével az alábbi kérdésekre tudunk gyors választ adni:

5.13. Állítás: ha ξ folytonos eloszlású, akkor

$$P(a \leq \xi < b) = P(a \leq \xi \leq b) = P(a < \xi \leq b) = P(a < \xi < b) = F(b) - F(a) . \quad \square$$

Most összefoglaljuk a fejezet legfontosabb képleteit:

5.14. Tétel: "Tipikus" kérdések (és a válaszok)

$$P(\xi < b) = \int_{-\infty}^b f(x) dx = F(b)$$

$$P(a \leq \xi) = \int_a^{\infty} f(x) dx = 1 - F(a) = 1 - P(\xi < a)$$

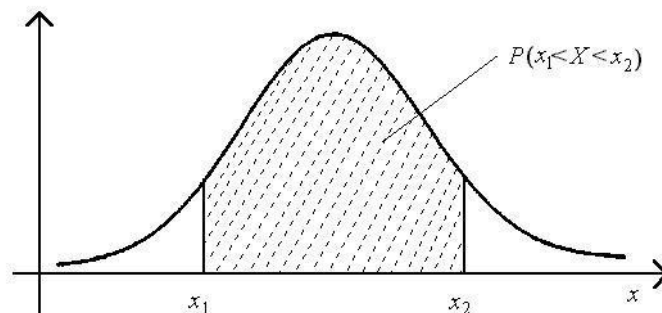
$$P(a \leq \xi < b) = \int_a^b f(x) dx = F(b) - F(a) \quad (\text{Newton-Leibniz szabály})$$

$$P(\xi = b) = 0 \quad (\text{ha } \xi \text{ folytonos v.v.})$$

$$P(\xi \approx c) = P(|\xi - c| < \varepsilon) = P(c - \varepsilon < \xi < c + \varepsilon) = F(c + \varepsilon) - F(c - \varepsilon) . \quad \square$$

5.15. Miért nevezzük $f(x)$ -et sűrűségfüggvénynek?

A Newton-Leibniz szabály alapján:

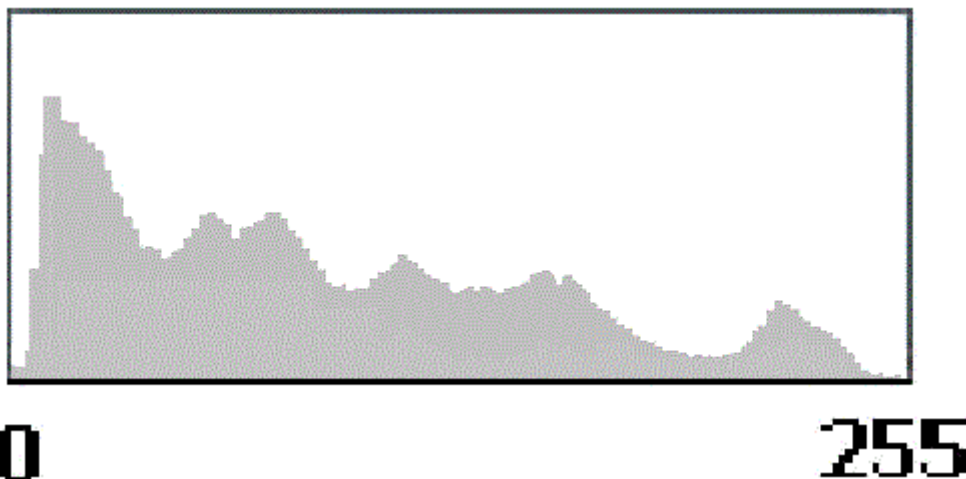


$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x) dx.$$

3. ábra: A sűrűségfüggvény

vagy másképpen: $\lim_{\Delta a \rightarrow 0} \frac{P(a \leq \xi < a + \Delta a)}{\Delta a} = f(a) . \quad \square$

5.16. **Megjegyzés:** Hisztogram (oszlopdiaagram) és sűrűségfüggvény kapcsolata:



4. ábra: Nagy felbontású oszlopdiaagram

5.17. **Megjegyzés:** Folytonos eloszlású v.v. és tetszőleges $b \in \mathbb{R}$ szám esetén nyilván $P(\xi=b) = \int_b^b f(x) dx = 0$. Azonban, ha b -nek csak "(nagyon) kis környezetét" vesszük, vagyis a " ξ közel van b -hez" állítás valószínűségét kérdezzük (azaz $P(\xi \approx b) = P(b-\varepsilon < \xi < b+\varepsilon) = ?$ ahol $0 < \varepsilon$ kicsi), akkor $P(\xi \approx b)$ arányos $f(x)$ -el, legalábbis a "nagy" vagy "kicsi" nagyságrendet illetően.

5.18. **Definíció:** Egy *diszkrét* v.v. **szimmetrikus eloszlású**, ha az 5.3. Definícióbeli $\{p_1, p_2, \dots, p_n\}$ sorozat szimmetrikus és $\{x_1, x_2, \dots, x_n\}$ számtani sorozat.

Egy *folytonos* v.v. **szimmetrikus eloszlású**, ha az $f(x)$ sűrűségfüggvénye szimmetrikus. \square

5.19. **Definíció:** *Diszkrét* valószínűségi változó **módusa(i)** a legvalószínűbb x_k érték(ek): **mod** := x_k ha $p_k = P(x_k)$ a legnagyobb(ak).

Folytonos valószínűségi változó **módusa(i)** az $f(x)$ sűrűségfüggvény lokális maximumhelye(i). \square

Tehát egy v.v.-nak több módusa is lehet. Ha ξ -nek csak egyetlen módusa van, akkor ξ -t **unimodális**-nak hívjuk.

5.20. Definíció: Diszkrét valószínűségi változó **mediánja** ("középen áll") a következő $\text{med} \in \mathbb{R}$ valós szám:

ha van olyan $x \in \mathbb{R}$ valós szám, amelyre $F(x) = \frac{1}{2}$, akkor ezen x számok egy intervallumot alkotnak, és legyen **med** ezen (α, β) intervallum *közep*: $\text{med} := (\alpha + \beta) / 2$.

ha az $F(x)$ függvény seholsem veszi fel az $\frac{1}{2}$ értéket, akkor **med** legyen a *legnagyobb* olyan $x \in \mathbb{R}$ amelyre $F(x) < \frac{1}{2}$ (ahol $F(x)$ "átlépi" $\frac{1}{2}$ -et).

Folytonos valószínűségi változó **mediánja** az $F(x) = \frac{1}{2}$ egyenlet megoldása, ha pedig több ilyen szám van, akkor m ennek az intervallumnak a közepe: $\text{med} := (\alpha + \beta) / 2$. □

A medián tehát lényegében azon x érték, aki a nagyság szerinti sorban "középen áll".

5.21. Tétel: Ha ξ szimmetrikus eloszlású, akkor $\text{med} = M(\xi)$. □

5.22. Tétel: Teszőlegesen ξ v.v. és $d \in \mathbb{R}$ esetén $M(|\xi - \text{med}|) \leq M(|\xi - d|)$. □

Valószínűségi változók függetlensége

5.23. Definíció: A ξ és η valószínűségi változókat **függetlennek** nevezzük, ha bármely $x, y \in \mathbb{R}$ valós számok esetén

$$P(\xi < x \text{ és } \eta < y) = P(\xi < x) \cdot P(\eta < y) . \quad \square$$

5.24. Állítás: Ha ξ és η *diszkrét* v.v., és lehetséges értékeik $\text{Im}(\xi) = \{x_1, x_2, \dots\}$ illetve $\text{Im}(\eta) = \{y_1, y_2, \dots\}$, akkor : ξ és η pontosan akkor függetlenek, ha

$$P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j) . \quad \square$$

5.25. Tétel: Ha : ξ és η folytonos eloszlásúak, sűrűségfüggvényük $f(x)$ illetve $g(y)$, akkor: ξ és η

pontosan akkor függetlenek, ha $\frac{\partial^2 P(\xi < x, \eta < y)}{\partial x \partial y} = f(x) \cdot g(y)$. □

Valószínűségi változók függvényei

A gyakorlatban néha előfordul, hogy a mérési eredményt számolással még át kell alakítanunk (transzformálnunk), pl. mértékegység vagy más átalakítás céljából.

5.26. Definíció: Ha a mért $\xi(\omega)$ értékre a $g(x)$ függvényt alkalmazzuk, akkor az $\eta(\omega) := g(\xi(\omega))$ (röviden: $\eta = g(\xi)$) mennyiség is egy valószínűségi változó (ugyanazon Ω eseménytéren), amit ξ -nak $g(x)$ szerinti transzformáltjának nevezünk. □

Nyilván ha ξ diszkrét v.v., akkor η is diszkrét v.v., ha pedig ξ folytonos v.v. és g folytonos függvény, akkor η is folytonos v.v.

5.27. Tétel: Ha a $g(x)$ függvény invertálható, akkor $\eta = g(\xi)$ eloszlása:

$$F_{\eta}(x) \equiv P(\eta < x) = P(g(\xi) < x) = P(\xi < g^{-1}(x)) = F_{\xi}(g^{-1}(x)) \quad (x \in \mathbb{R})$$

és ha ξ folytonos és g^{-1} deriválható, akkor

$$f_{\eta}(x) \equiv F_{\eta}'(x) = (F_{\xi}(g^{-1}(x)))' = F_{\xi}'(g^{-1}(x)) \cdot g^{-1}(x) = f_{\xi}(g^{-1}(x)) \cdot g^{-1}(x) . \quad (x \in \mathbb{R}) \quad \square$$

6. Várható érték és szórás

Egyetlen mérés helyett általában többször szoktunk mérni, és a méréseket utána átlagoljuk, ismétlődő értékek esetén súlyozzuk az értékeket.

6.1. Definíció: Ha ξ *diszkrét* v.v., lehetséges értékei $\{x_1, \dots\}$ és eloszlása $\{p_1, \dots\}$, akkor ξ **várható értéke** (expected value, $E(\xi)$) vagy **átlaga** (mean, $M(\xi)$)

- ha ξ -nek véges sok lehetséges értéke van, akkor

$$M(\xi) = E(\xi) = \sum_{i=1}^n x_i p_i,$$

- ha ξ -nek végtelen sok lehetséges értéke van, akkor

$$M(\xi) = E(\xi) = \sum_{i=1}^{\infty} x_i p_i$$

feltéve, hogy $\sum_{i=1}^{\infty} |x_i| p_i < \infty$. □

6.2. Definíció: Ha ξ *folytonos* v.v., sűrűségfüggvénye $f(x)$, akkor ξ **várható értéke** (expected value, $E(\xi)$) vagy **átlaga** (mean, $M(\xi)$)

$$M(\xi) = E(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

amennyiben az $\int_{-\infty}^{\infty} |x| \cdot f(x) dx$ improprius integrál konvergens. □

6.3. Megjegyzések: (i) $E(\xi)$ a régebben használt angol "*expected value*" (várt érték) kezdőbetűje, újabban inkább az $M(\xi)$ jelölést használjuk, ami az angol "*mean*" (átlag!) szót jelöli.

(ii) Úgy tapasztaljuk hogy a ξ mennyiség többszöri mérésekor kapott értékek az $M(\xi)$ átlag körül ingadoznak. Ezt a **Nagy számok gyenge törvénye (Csebisev-alak)** tétel igazolja a *Nagy számok* Fejezetben.

6.4. Tétel: A várható érték tulajdonságai:

1. Ha $\xi = c$, akkor $M(\xi) = c$ ("beragadt a mérőműszer").
2. Ha létezik $M(\xi)$, akkor $M(a\xi + b) = aM(\xi) + b$ (pl. °C helyett °F).
3. Ha $M(\xi)$ és $M(\eta)$ létezik, akkor $M(\xi \pm \eta) = M(\xi) \pm M(\eta)$.
4. Ha $a \leq \xi \leq b$, akkor $a \leq M(\xi) \leq b$.
5. Ha $0 \leq \xi$, akkor $0 \leq M(\xi)$.
6. Ha ξ és η függetlenek, akkor $M(\xi\eta) = M(\xi)M(\eta)$.
7. Ha $\xi_1, \xi_2, \dots, \xi_n$ független azonos eloszlású valószínűségi változók, akkor

$$M\left(\sum_{i=1}^n \xi_i\right) = nM(\xi_1) \text{ és } M\left(\frac{\sum_{i=1}^n \xi_i}{n}\right) = M(\xi_1)$$

8. Ha $g: H \subset R \rightarrow R$ folytonos függvény, akkor diszkrét eloszlás esetén

$$M(g(\xi)) = \sum_{i=1}^{\infty} g(x_i) p_i$$

folytonos eloszlás esetén $M(g(\xi)) = \int_{-\infty}^{\infty} g(x) f(x) dx$.

9. Speciálisan: $g(x) = x^2$ esetén $M(\xi^2) = \sum_{i=1}^{\infty} x_i^2 p_i$ illetve $M(\xi^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$. \square

Megjegyezzük, hogy $\xi_1 + \xi_2$ (3.pont) és $2 \cdot \xi$ (2.pont) messze *nem ugyanaz*! Annak ellenére, hogy ha ξ_1, ξ_2 és ξ azonos eloszlásúak ($f(x)$ mindegyiknél ugyanaz), amiből $M(\xi_1) = M(\xi_2) = M(\xi) = m$ következik, akkor $M(\xi_1 + \xi_2) = M(2 \cdot \xi) = 2m$.

Ugyanis: $\xi_1 + \xi_2 =$ két független ("megismételt") mérés összege, míg $2 \cdot \xi$ egyetlen mérés eredményének kettővel vett szorzata ("lusta pulzusz mérés"). A különbséget még jobban látjuk a 6.10. Tételben.

A szórás

Többször tapasztalhattuk (pl. iskolai dolgozatok osztályzatainál), hogy az átlagtól való *szóródás* néha kicsi, néha nagy. Ezeket az eltéréseket nyilván átlagolnunk kell.

Ha ξ -nek van várható értéke, akkor $M(\xi - M(\xi)) = 0$. $M(|\xi - M(\xi)|)$ sem jó, nem könnyű számolni.

De $(\xi - M(\xi))^2$ a kis eltéréseket még kisebbé, a nagyokat még nagyobbá teszi, ezért a $M((\xi - M(\xi))^2)$ mennyiséget használjuk az ingadozás mérésére.

6.6. Definíció: A ξ v.v. **szórásnégyzete** $D^2(\xi) = M((\xi - M(\xi))^2)$ (amennyiben ez a kifejezés véges), **szórása** $D(\xi) = \sqrt{M((\xi - M(\xi))^2)}$. \square

6.7. Állítás: A fenti, gyök alatti mennyiség nemnegativitása biztosított a várható érték 6.4. Tétel 5) tulajdonsága miatt, tehát gyököt lehet vonni. \square

6.8. Állítás: Ha $M(\xi^2)$ véges, akkor létezik a szórásnégyzet és $D^2(\xi) = M(\xi^2) - M^2(\xi)$.

Bizonyítás: A 6.4. Tétel pontjait alkalmazva

$$\begin{aligned} D^2(\xi) &= M((\xi - M(\xi))^2) = M(\xi^2 - 2\xi \cdot M(\xi) + M^2(\xi)) = M(\xi^2) - 2M(\xi)M(\xi) + M^2(\xi) = \\ &= M(\xi^2) - M^2(\xi) \end{aligned} \quad \square$$

6.9. Következmény: A várható érték (6.4. Tétel 9.) tulajdonsága miatt a szórásnégyzetre a következő számolási képleteink vannak:

diszkrét valószínűségi változó esetén: $D^2(\xi) = \sum_{i=1}^{\infty} x_i^2 p_i - \left(\sum_{i=1}^{\infty} x_i p_i \right)^2$,

folytonos valószínűségi változó esetén: $D^2(\xi) = \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\int_{-\infty}^{\infty} x f(x) dx \right)^2$. \square

6.10. Tétel: A szórásnégyzet és a szórás tulajdonságai:

1. $\xi = c$ akkor és csak akkor, ha $D^2(\xi) = D(\xi) = 0$. ("Beragadt a mérőműszer".)

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

3. CSAK Ha ξ és η függetlenek, akkor

$$D^2(\xi \pm \eta) = D^2(\xi) + D^2(\eta) \quad \text{és} \quad D(\xi \pm \eta) = \sqrt{D^2(\xi) + D^2(\eta)}.$$

4. Ha $\xi_1, \xi_2, \dots, \xi_n$ független, azonos eloszlású valószínűségi változók, akkor

$$D^2\left(\sum_{i=1}^n \xi_i\right) = n \cdot D^2(\xi_1) \quad , \quad D^2\left(\sum_{i=1}^n \xi_i / n\right) = \frac{D^2(\xi_1)}{n} \quad ,$$

$$D\left(\sum_{i=1}^n \xi_i\right) = \sqrt{n}D(\xi_1) \quad , \quad D\left(\sum_{i=1}^n \xi_i / n\right) = \frac{D(\xi_1)}{\sqrt{n}} \quad . \quad \square$$

Megjegyzések: felhívjuk a figyelmet arra, hogy a 3. pontban mind $\xi_+\eta$ mind $\xi_-\eta$ esetében $D^2(\xi)+D^2(\eta)$ szerepel ("Pitagoraszi tétel") !

A 6.4.Tétel után tett megjegyzést folytatva: ha ξ_1, ξ_2 és ξ azonos eloszlásúak ($f(x)$ mindegyiknél ugyanaz), akkor egyrészt $D(\xi_1)=D(\xi_2)=D(\xi)=\sigma$, másrészt (a fenti 3. és 2. pontok alapján)

$$D(\xi_1+\xi_2) = \sqrt{D^2(\xi_1) + D^2(\xi_2)} = \sqrt{2 \cdot \sigma^2} = \sqrt{2} \cdot \sigma \quad (\text{két független mérés összege})$$

míg

$$D(2 \cdot \xi) = 2 \cdot D(\xi) = 2 \cdot \sigma \quad (\text{egyetlen mérés "felszorozva" 2-vel}).$$

Látható, hogy két független mérés összegének szórása jóval kisebb (kb.30%), mint ha egyetlen mérést szoroznánk 2-vel, vagyis nem érdemes lustának lennünk: két mérés összege sokkal pontosabb !

6.11. Definíció: Tetszőleges ξ v.v. esetén a $\xi^* := \frac{\xi - M(\xi)}{D(\xi)}$ valószínűségi változót a ξ v.v. **standardizáltjának** (standard változata) nevezzük.

Ugyanis, általában ξ és ξ^* eloszlása ugyanaz, és $M(\xi^*)=0$, $D(\xi^*)=1$.

7. Nevezetes diszkrét eloszlású valószínűségi változók

Az alábbiakban $\text{Im}(\xi)=\{x_1, \dots, x_n\}$ vagy $\text{Im}(\xi)=\{x_1, \dots, x_n, \dots\}$ és használjuk a $p_k := P(\xi=x_k)$ rövidítést ahol $k \in \mathbb{N}$ tetszőleges természetes szám.

7.1. Diszkrét egyenletes eloszlású valószínűségi változók

7.1. Definíció: ξ **diszkrét egyenletes v.v.**, ha $\text{Im}(\xi)=\{x_1, \dots, x_n\}$ egy tetszőleges véges halmaz és ξ mindegyik x_i értéket ugyanakkora valószínűséggel veszi fel, azaz $p_i = P(\xi=x_i) = 1/n$ minden $i=1, \dots, n$ esetén. \square

7.2. Például szabályos dobókockával dobunk, a fizetendő összeg utolsó számjegye, véletlenül mondunk egy számot 1 és n között, a dobozban levő n db (különböző) tombolajegyek számai $\{x_1, \dots, x_n\}$, szabályos sokszög alapú egyenes hasáb (majdnem henger) alakú ceruza oldalaira különböző számokat írunk és a ceruzát elgurítjuk, stb.

7.3. Állítás: $M(\xi) = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$ és $D(\xi) = \sqrt{M(\xi^2) - M^2(\xi)} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$

(Ez lényegében csak a definíció felírása, tehát ellenőrzése nem nehéz házi feladat. \square)

7.2. Hipergeometrikus eloszlású valószínűségi változók

7.4. Definíció: Legyenek $N, S, n \in \mathbb{N}$ rögzített természetes számok, $N \geq 2$, $0 \leq S \leq N$, $1 \leq n \leq S$, $0 \leq N - S \leq n$. Ekkor ξ hipergeometrikus eloszlású v.v. az S, N, n paraméterekkel, ha ξ lehetséges értékei $\text{Im}(\xi) = \{0, 1, \dots, n\}$ és

$$p_k = P(\xi = k) = \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}} . \quad \square$$

ahol $\binom{\dots}{\dots}$ természetesen binomiális együtthatókat jelöl.

7.5. Állítás: Az alábbi típusú kísérletek, melyeket **visszatevés nélküli mintavételeknek** nevezünk, hipergeometrikus eloszlásúak. Tegyük fel, hogy egy halmazban N db elem van, amelyek között S db másnyen, mint a többi (például selejt). Visszatevés nélkül, akár egy marokkal kiveszünk közülük n -et. Legyen ξ a kivett elemek között a másnyenek (S -beliek) száma. Ekkor ξ hipergeometrikus v.v.

Az állítás belátása nem nehéz, házi feladat. □

7.6. Példák: öt kártyát húzva hány pirosat kapunk; hányas találatom lehet a lottón; 70 konzervből (20% selejt) hármát választva hány selejtet kapok; 13 tételből (4 algebra) kettőt húzva hány algebrát kapok,

7.7. Tétel: $M(\xi) = n \cdot \frac{S}{N}$ ($= np$) és $D(\xi) = \sqrt{np(1-p) \left(1 - \frac{n-1}{N-1}\right)}$ ahol $p = \frac{S}{N}$.

(A $\sqrt{1 - \frac{n-1}{N-1}}$ szorzótényezőt **korrekciós tényezőnek** hívjuk.)

A 27. oldalon levő táblázatban S helyett K szerepel.

7.8. Tétel: Ha n és k állandó, $S \rightarrow \infty$, $N \rightarrow \infty$, $\frac{S}{N} \rightarrow p = \text{állandó}$, akkor

$$\frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}} \rightarrow \binom{n}{k} p^k (1-p)^{n-k} \quad \square$$

7.9. Magyarázat: A fenti tétel szerint a hipergeometriai eloszlású v.v. hoz tartozó valószínűségek határértékben a binomiális eloszláshoz (lásd köv. fejezet) tartozó valószínűségekhez tartanak. Ez összhangban azzal a tapasztalati ténnyel és elméleti megfontolással, hogy nagy elemű sokaság (alaphalmaz) esetén majdnem mindegy, hogy visszatevéssel vagy visszatevés nélkül választunk ki elemeket.

7.3. Binomiális vagy Bernoulli eloszlású valószínűségi változók

7.10. Definíció ξ **binomiális** vagy **Bernoulli** eloszlású v.v. az n, p paraméterekkel, ha $n \in \mathbb{N}$ tetszőleges nemnulla természetes szám és $0 < p < 1$ tetszőleges szám, továbbá ha ξ lehetséges értékei (!)

$\text{Im}(\xi) = \{x_0, x_1, \dots, x_n\} = \{0, 1, 2, \dots, n\}$ és *eloszlása:* tetszőleges $k=0, 1, 2, \dots, n$ esetén

$$p_k = P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

ahol $\binom{n}{k}$ a binomiális együttható. □

7.11. Tétel: A következő típusú kísérletek, melyeket **visszatevéses mintavételeknek** nevezünk, binomiális eloszlásúak. Rögzítsünk egy kísérletet (Ω) és egy $A \subseteq \Omega$ eseményt, rögzítsünk egy tetszőleges $n \in \mathbb{N}$, $n \neq 0$ tetszőleges természetes számot, és legyen $p = P(A)$. Most végezzük el a kísérletet n -szer egymás után,

teljesen azonos körülmények között, egymástól függetlenül, és jelölje (számolja) ξ azt, hogy az n kísérlet alatt hányszor következett be az A esemény. Ekkor ξ binomiális eloszlású v.v. az n, p paraméterekkel. Ennek belátása nem nehéz, házi feladat. \square

Hangsúlyozzuk, hogy a fenti tételben említett feltételek ("teljesen azonos körülmények között, egymástól függetlenül") nem mindig teljesülnek maradéktalanul a gyakorlatban, az ilyen kísérletsorozatok tehát nem teljesen binomiális eloszlású v.v. -k.

7.12. Példák:

- 10-szer gurítunk egy kockával, ξ a dobott hatosok száma, ekkor ξ binomiális eloszlású az $n=10$ és $p=1/6$ paraméterekkel. Azonos feladat: egyszerre gurítunk 10 szabályos kockát, amik nem akadályozzák (befolyásolják) egymást.
- 7-szer feldobunk egy szabályos érmét, ξ a dobott fejek száma. ξ binomiális eloszlású az $n=7$, $p=0.5$ paraméterekkel.
- Visszatevéses mintavételek:* visszatevéssel (=egyesevel) választunk rögzített N elem közül n -szer. Az összes elem között S db selejtes van. Legyen ξ a kivett elemek közt a selejtes elemek száma. Ekkor ξ binomiális eloszlású az n és a $p=S/N$ paraméterekkel.
- Visszatevéssel* választunk a magyar kártyából 5-ször, ξ a kivett lapok közt a pirosak száma. Ekkor ξ binomiális eloszlású az $n=5$ és $p=8/32$ paraméterekkel.

7.13. Állítás: Ha ξ binomiális v.v., akkor $M(\xi) = np$ és $D(\xi) = \sqrt{np(1-p)}$. \square

7.14. Állítás: Ha ξ binomiális v.v., akkor ξ legvalószínűbb értéke (*modusza*):

$$m = [(n+1)p] \quad \text{akkor, ha } (n+1)p \text{ nem egész szám,}$$

$$m = (n+1)p \text{ és } m = (n+1)p - 1, \quad \text{akkor, ha } (n+1)p \text{ egész szám,}$$

ahol $[x]$ az x valós szám **egész részét** (lefelé kerekítés vagy csonkítás) jelenti. \square

7.15. Tétel: Ha ξ_1 binomiális eloszlású val. változó az n_1 és p paraméterekkel, ξ_2 binomiális eloszlású val. változó az n_2 és (ugyanazon!) p paraméterekkel, valamint ξ_1 és ξ_2 függetlenek, akkor $\xi_1 + \xi_2$ is binomiális eloszlású valószínűségi változó az $n = n_1 + n_2$ és p paraméterekkel. \square

7.16. Tétel: Ha $n \rightarrow \infty$ és $p \rightarrow 0$ úgy, hogy $np = \lambda = \text{állandó}$, akkor

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square$$

A legutolsó eredmény, többek között, segítségünkre van p_k kiszámolásában. Ugyanis nagy n , k és kis p esetén a binomiális együttható nagyon nagy míg p^k nagyon kicsi, és ezek szorzata nagy pontatlanságot okoz a számológépeken. A Poisson-eloszlások (köv. fejezet) alapja a fenti 7.16. Tétel.

Példa: $n=1000$, $p=0.002$, $k=10$, $\Rightarrow p^k = 0.002^{10} \approx 1.024 \times 10^{-27}$, $(1-p)^{n-k} = 0.998^{990} \approx 0.1378$,

$$\binom{n}{k} = \binom{1000}{10} = \frac{1000 \cdot 999 \cdot \dots \cdot 991}{10!} \approx 2.634 \times 10^{+23}, \text{ így}$$

$$P^{\text{Bem}}(\xi=k) \approx 2.634 \times 10^{+23} * 1.024 \times 10^{-27} * 0.1378 \approx \underline{3.7168 \times 10^{-5}}.$$

A 7.16. közelítés szerint: $\lambda = 1000 * 0.002 = 2$, ezért

$$P^{\text{Poiss}}(\xi=k) = \frac{2^{10}}{10!} e^{-2} \approx \underline{3.8190 \times 10^{-5}}, \text{ és a közelítés hibája:}$$

$$3.8190 \times 10^{-5} / 3.7168 \times 10^{-5} \approx 1.0275 \text{ miatt a hiba } \approx 2.75\%.$$

Vegyük észre, hogy a P^{Poiss} közelítő képlet kiszámítása sokkal egyszerűbb mint a P^{Bem} pontos képlet!

Tétel: Ha ξ_1 és ξ_2 független binomiális eloszlású v.v. az (n_1, p) illetve az (n_2, p) paraméterekkel (p ugyanaz!), akkor $\xi_1 + \xi_2$ is binomiális eloszlású v.v. az (n_1+n_2, p) paraméterekkel. \square

A fenti állítás szemléletesen azt mondja ki, hogy ha egy (p valószínűségű) kísérletet elvégezzünk n_1 -szer, majd folytatjuk egy másik, n_2 -ször megismételt kísérletsorozattal, akkor ez végeredményben egy n_1+n_2 hosszú binomiális eloszlású v.v., a p paraméterrel. Lényeges azonban, hogy a két kísérletsorozatban ugyanaz a p paraméter szerepel!

$P^{\text{Bern}}(\xi=k)$ összegeinek közelítő számítása a 10.5. (CHT) vagy a 10.6. (Moivre-Laplace) tételek, vagy a Függelékben található *Larson nomogram* használható.

7.4. Poisson eloszlású valószínűségi változók

7.17. Definíció: A ξ v.v. **Poisson** eloszlású a $\lambda > 0$ valós paraméterrel, ha lehetséges értékei $\text{Im}(\xi) = \mathbb{N} = \{0, 1, 2, \dots, n, \dots\}$ az összes természetes szám, és eloszlása ($e \approx 2,71828$ az Euler-féle szám)

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda} . \quad \square$$

7.18. Alkalmazás: A Poisson eloszlások egyik alkalmazása, mint az előző fejezetben említettük, a binomiális eloszlások közelítése nagy n és kis p esetén. Erre egy mintapéldát a 10. fejezet végén mutatunk.

7.19. Példák: "egységnyi" területen/hosszon/térfogatban keletkezett anyaghibák/aknák száma (textil/cső/anyag), adott térfogatban levő mazsolák/kavicsok/csillagok száma, adott időintervallumban a telefonhívások/ügyfelek/csillaghullások/féltéglák száma, adott oldalszámban levő sajtóhibák száma könyvben.

Általánosan ezeket a problémákat a következőképpen fogalmazhatjuk meg: *rögzítünk* egy "fizikai halmazt", ami lehet térbeli (vagy sík- vagy egyenesbeli) halmaz vagy időintervallum, ezt a fizikai halmazt a továbbiakban egységesen "**V térfogatnak**" hívjuk. Tegyük még fel, hogy ezen a halmazon belül sok, egymástól független jelenség léphet fel, melyek egyszerre történő megjelenése négyzetesen csökken. Ekkor bebizonyított tétel, hogy ξ , a fellépett jelenségek száma, Poisson v.v.

Ebből a megfogalmazásból ugye érezhető a Poisson és binomiális eloszlások hasonlósága, amit a 7.16. Tétel igazol. \square

7.20. Tétel: Ha ξ Poisson v.v. $\lambda > 0$ paraméterrel, akkor $M(\xi) = \lambda$ és $D(\xi) = \sqrt{\lambda}$. \square

7.21. Tétel: ξ legvalószínűbb értéke (*modusz*):

$$\begin{array}{ll} [\lambda] & \text{ha } \lambda \text{ nem egész szám,} \\ \lambda \text{ és } \lambda - 1 & \text{ha } \lambda \text{ egész szám.} \quad \square \end{array}$$

7.22. Tétel: Ha ξ_1 Poisson eloszlású val. változó a λ_1 paraméterrel, ξ_2 Poisson eloszlású val. változó a λ_2 paraméterrel, valamint ξ_1 és ξ_2 *függetlenek*, akkor $\xi_1 + \xi_2$ is Poisson eloszlású valószínűségi változó az $\lambda = \lambda_1 + \lambda_2$ paraméterrel.

Következmény: Ha ξ Poisson eloszlású a V térfogaton λ paraméterrel és $t \in \mathbb{R}^+$ tetszőleges pozitív szám, akkor ξ leszűkítése/kiterjesztése a V/t térfogatra szintén Poisson eloszlású λ/t paraméterrel ($t \in \mathbb{R}^+$). Ezt a jelenséget röviden úgy hívjuk, hogy a Poisson eloszlások **korlátlanul oszthatók**.

Példa: Évente kétféle féltégla esik a fejemre. Mekkora valószínűséggel úszom meg a mai napot?

Megoldás: $\xi =$ "évente hány féltégla" $\Rightarrow \lambda_\xi = M(\xi) = 2$, $\eta =$ "naponta hány féltégla" $\Rightarrow \eta = \xi/365$
 $\Rightarrow \lambda_\eta = \lambda_\xi/365 = 2/365 \Rightarrow P(\eta=0) = (\lambda_\eta)^0 / 0! * e^{-\lambda_\eta} = 1 * e^{-2/365} \approx 0.99454 \approx 0.99.5\%$.

Lásd még a 8.12. Tételt.

7.5. Geometriai eloszlású valószínűségi változók

7.23. Definíció: ξ geometriai eloszlású v.v. a p ($0 < p < 1$) paraméterrel, ha ξ lehetséges értékei $\text{Im}(\xi) = N = \{1, 2, \dots, n, \dots\}$ az összes pozitív természetes szám, és ξ eloszlása:

$$p_k = P(\xi = k) = p \cdot (1-p)^{k-1} . \quad \square$$

Megjegyzés: szokás a $q:=1-p$ rövidítést használni, ekkor p_k képlete: $p_k = p \cdot q^{k-1}$. Ezért hívják ξ -t mértani (geometriai) eloszlásnak, hiszen ekkor (p_1, p_2, \dots) egy mértani sorozat.

7.24. Tétel: A gyakorlatban a következő kísérletek mértani eloszlású v.v.-k: egymástól függetlenül végezzük ugyanazt a kísérletet, legyen A egy rögzített esemény, amelynek valószínűségét jelölje $p:=P(A)$ és legyen $0 < p < 1$. Ha a kísérletet addig ismételtjük, teljesen azonos körülmények között, egymástól függetlenül, amíg az A be nem következik, és ξ azt a legkisebb természetes számot jelöli, ahányadik kísérletnél A legelőször bekövetkezett ("addig, amíg"), akkor ξ eloszlása geometriai, a p paraméterrel. Ennek belátása nem nehéz, házi feladat.

7.25. Példák: addig jár a kórsó a kútra, amíg el nem törik,
addig járunk vizsgázni, amíg végre sikerül a vizsga,
addig ütünk a dióra kalapáccsal amíg meg nem reped,
addig lövünk az ellenségre amíg el nem találjuk,
addig ugrunk árkot, amíg végre száraz ruhával megússzuk,
addig stoppolunk, amíg felvesz egy autó,
addig vonulnak el a háremhölgyek Szindbád előtt, amíg Szindbád nem választ,

Megjegyezzük, hogy a valódi, gyakorlati életben nem mindegyik fenti példában illetve nem mindig teljesülnek maradéktalanul a feltételek ("teljesen azonos körülmények között, egymástól függetlenül"). Gondoljunk csak a vizsgákra vagy a diókra: a következő vizsgára/dióra már jobban készülünk/ütünk, vagy éppen idegesebbek vagyunk, jobban izgulunk vagy puskázunk, no és ki hallott már a "8 645 -dik vizsgáról"?
Ide kapcsolódik még a 8.14. Tétel is. □

7.26. Tétel: Ha ξ geometriai eloszlású v.v. p paraméterrel, akkor

$$M(\xi) = \frac{1}{p} \quad \text{és} \quad D(\xi) = \frac{\sqrt{1-p}}{p} . \quad \square$$

Gyakorlati feladatoknál sokszor felmerül a "hány kísérlet lesz elég?" kérdés:

7.27. Állítás: $P(\xi \leq k) = 1 - (1-p)^k = 1 - q^k$ (ahol $q=1-p$).

Bizonyítás: (ezt a számolást feladatnak is kitzehetnénk, tehát kivételesen hasznos lesz átolvasnunk):

$$P(\xi \leq k) = P(\xi=1) + P(\xi=2) + \dots + P(\xi=k) = p + p \cdot q^1 + p \cdot q^2 + \dots + p \cdot q^{k-1} = p \cdot \frac{q^k - 1}{q - 1} = p \cdot \frac{q^k - 1}{-p} = 1 - q^k . \quad \square$$

$$\text{Hasonlóan kapjuk: } F(k) = P(\xi < k) = 1 - q^{k-1} = 1 - (1-p)^{k-1} .$$

Példa: Ha k a kérdés, vagyis "hány vizsga elég α valószínűséggel", akkor nyilván $P(\xi \leq k) = 1 - q^k \geq \alpha$, ahonnan $1 - \alpha \geq q^k$, a megoldás pedig $\log_q(1 - \alpha) \leq k$ ($q < 1$ ezért \log_q szig.mon.csökken).
Zsebszámológéppel: $\log_q(1 - \alpha) = \lg(1 - \alpha) / \lg(q)$.

7.28. Tétel: Ha ξ geometriai eloszlású valószínűségi változó, akkor

$$P(\xi > m + n \mid \xi > m) = P(\xi > n) ,$$

vagyis, az eddigi kísérletek nem befolyásolják a továbbiakat (" ξ **nem öregedő**").

Bizonyítás: $P(\xi > k) = \sum_{i=k+1}^{\infty} P(\xi = i) = \sum_{i=k+1}^{\infty} p(1-p)^{i-1} = p(1-p)^k \frac{1}{1-(1-p)} = (1-p)^k$. $m+n > m$ miatt $P(\xi > m+n \cap \xi > m) = P(\xi > m+n) = (1-p)^{m+n}$, így $P(\xi > m+n | \xi > m) = \frac{P(\xi > m+n)}{P(\xi > m)} = (1-p)^n$. \square

8. Nevezetes folytonos eloszlású valószínűségi változók

8.1. Folytonos egyenletes eloszlású valószínűségi változók

8.1. Definíció Tetszőleges rögzített $a, b \in \mathbb{R}$ valós számok, $a < b$ esetén ξ **folytonos egyenletes eloszlású**

v.v. az a, b paraméterekkel, ha sűrűségfüggvénye $f(x) = \begin{cases} c & \text{ha } a \leq x \leq b \\ 0 & \text{különben} \end{cases}$. \square

Ebben az alfejezetben *nagyon egyszerűen* bizonyítható (kiszámolható) állítások vannak, a számolások elvégzése mindegyik esetben hasznos házi feladat.

8.2. Állítás: $c = \frac{1}{b-a}$ és $F(x) = \begin{cases} 0 & \text{ha } x \leq a \\ \frac{x-a}{b-a} & \text{ha } a \leq x \leq b \\ 1 & \text{ha } x \geq b \end{cases}$. \square

8.3. Állítás: $M(\xi) = \frac{a+b}{2}$ és $D(\xi) = \frac{b-a}{\sqrt{12}}$. \square

8.4. Tétel: A gyakorlatban a következő kísérletek *folytonos egyenletes eloszlású* v.v.-k: véletlenszerűen, minden befolyástól mentesen, "egyenletesen" választunk az $[a, b]$ intervallumban egy x valós számot. \square

A fenti és az alábbi állítások is igazolják a folytonos egyenletes eloszlások és a geometriai valószínűség (3.fejezet második része) azonosságát!

8.5. Példák: henger alakú ceruzát elgurítva mely pontján áll meg, papírszalagot hol vágunk ketté, pálcát hol törünk ketté, buszmegállóba megyünk ki "csak úgy", a buszok egyenletesen 15 percnként jönnek, stb.

8.6. Állítás: $P(c < \xi < d) = \frac{d-c}{b-a}$ arányos a $[c, d]$ intervallum hosszával, ha $[c, d] \subseteq [a, b]$. \square

8.2. Exponenciális eloszlású valószínűségi változók

8.7. Definíció Tetszőleges rögzített $\lambda \in \mathbb{R}^+$ pozitív valós szám esetén ξ **exponenciális eloszlású** v.v. a $\lambda > 0$ paraméterrel, ha sűrűségfüggvénye

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{ha } x \geq 0 \\ 0 & \text{különben} \end{cases} . \quad \square$$

Megjegyzés: A gyakorlatban a következő kísérletek *exponenciális eloszlású* v.v.-k: gépek, élettelen tárgyak (pl. izzók, gumiabroncsok, stb.) élettartama, azaz a tönkremenésig eltelt idő, radioaktív anyagok lebomlása (ez is élettartam). Továbbá: várakozási idők (pl. hivatalban, pénztárnál, benzinkútnál), telefonhívások *hossza*, vagy a következő eseményig (pl. csillaghullásig, telefonhívásig, alkatrész-cseréig) eltelt idők. Ezek nem elméleti tételek, hanem gyakorlati tapasztalat, pontosabban statisztikai módszerekkel ("*illeszkedésvizsgálat*") ellenőrzött tények.

Az exponenciális eloszlás a geometriai eloszlás *folytonos* analogonja. □

8.8. Tétel: Ha ξ exponenciális eloszlású v.v. a λ paraméterrel, akkor $F(x) = \begin{cases} 0 & \text{ha } x < 0 \\ 1 - e^{-\lambda x} & \text{ha } x \geq 0 \end{cases}$. □

8.9. Tétel: $M(\xi) = D(\xi) = \frac{1}{\lambda}$. □

8.10. Tétel: Ha ξ exponenciális eloszlású val. változó, és $x \geq 0, y \geq 0$ tetszőleges számok, akkor

$$P(\xi \geq x + y | \xi \geq y) = P(\xi \geq x) . \quad (\text{" } \xi \text{ örökifjú, nem öregedő" }) \quad \square$$

A fenti egyenlőség kiszámolása nem nehéz, javasolt házi feladat. Érdemes még összehasonlítani a 7.28. Tétellel is. A következő tétel kiemeli az örökifjú tulajdonság és az exponenciális eloszlás kapcsolatát.

8.11. Tétel: Ha ξ folytonos eloszlású, $F(0)=0, F(x)<1$, F minden nemnegatív x -re deriválható,

$\lim_{0+} F'(x) = \lambda > 0$ és ξ örökifjú, akkor ξ exponenciális eloszlású v.v. □

8.12. Tétel: Az exponenciális eloszlás és a *Poisson* eloszlás kapcsolata:

Legyenek ξ_1, ξ_2, \dots független exponenciális eloszlású v.v. azonos λ_ξ paraméterrel, legyen $T > 0$ rögzített, és legyen η a T "időpontig tönkrement/kicserélt alkatrészek" száma, azaz legyen $\text{Im}(\eta) = \mathbb{N}$, és legyen:

$$\eta = 0 \text{ ha } \xi_1 > T, \quad \eta = 1 \text{ ha } \xi_1 \leq T \text{ de } \xi_1 + \xi_2 > T, \quad \dots$$

és általában

$$\eta = k \text{ ha } \sum_{i=1}^k \xi_i \leq T \text{ de } \sum_{i=1}^{k+1} \xi_i > T .$$

Ekkor η Poisson eloszlású v.v. a $\lambda_\eta = \lambda_\xi \cdot T$ paraméterrel. □

8.13. Példa: Egy bizonyos típusú izzó élettartama exponenciális eloszlású valószínűségi változó (ξ) 1000 óra várható értékkel ($M(\xi)$). Ha egy izzó tönkremegy, azonnal kicseréljük egy ugyanolyan típusú másik izzóra, aminek az élettartama független az előző izzó élettartamától. Mennyi a valószínűsége annak, hogy 2500 óra alatt 2 izzócseré szükseges? (Azaz $P(\eta \geq 2) = ?$)

Megoldás:

$M(\xi) = 1000$, a 8.9. Tétel szerint $\lambda_\xi = 1/1000$, a fenti tétel szerint $\lambda_\eta = 2500/1000 = 2.5$, ezért

$$P(\eta \geq 2) = 1 - P(\eta < 2) = 1 - P(\eta = 1) - P(\eta = 0) = 1 - e^{-2.5} \cdot ((2.5)^0/0! + (2.5)^1/1!) \approx \underline{0.7127} .$$

8.14. Tétel: az exponenciális eloszlás és a *geometriai* eloszlás kapcsolata:

Ha ξ *exponenciális eloszlású* v.v. a λ_ξ paraméterrel, akkor $\eta = [\xi] + 1$ geometriai eloszlású v.v. a

$p_\eta = 1 - e^{-\lambda_\xi}$ paraméterrel.

Bizonyítás: η értékei lehetnek 1, 2, 3, ... , és

$$P(\eta = k) = P(k-1 \leq \xi < k) = F(k) - F(k-1) = 1 - e^{-\lambda k} - (1 - e^{-\lambda(k-1)}) = e^{-\lambda(k-1)} (1 - e^{-\lambda}) . \quad \square$$

Magyarázat: η jelentése: hány (egész) óráig működik ξ , meddig kell még várnom, hogy tönkremenjen, vagyis "egy-egy kockadobás" = "1 óráig működtetem". A percdíj alapú számlázás ("minden megkezdett perc egészben kifizetendő"), vagy parkolás is a fenti η -val számolható: $P(k \text{ percet fizetek}) = P(\eta = k) = (1 - p_\eta)^{k-1} * p_\eta$.

8.15. Példa: Egy telefontársaság folytonos (másodperc) alapon számláz, 20 Ft/perc díjjal. Egy másik társaság perc alapon számláz 15 Ft/ perc illetve percdíjjal. Melyik számlázási módot érdemes választani, ha egy hívás hossza exponenciális eloszlású valószínűségi változó 2 perc átlagos értékkel?

Megoldás: Legyen ξ a hívás hossza, θ_1 a folytonos alapon számlázott díj, θ_2 a perc alapon számlázott díj. Ekkor

$$M(\theta_1) = 20M(\xi) = 40 \text{ Ft} \quad \text{és} \quad M(\theta_2) = 15M([\xi]+1) = 15 \frac{1}{1 - e^{-0.5}} = 38.1 \text{ Ft.}$$

Sőt, ha **16 Ft** a percdíj, akkor $M(\theta_2) = 16 M([\xi]+1) \approx 40.67$, már 40 Ft fölé kerül a várható érték. \square

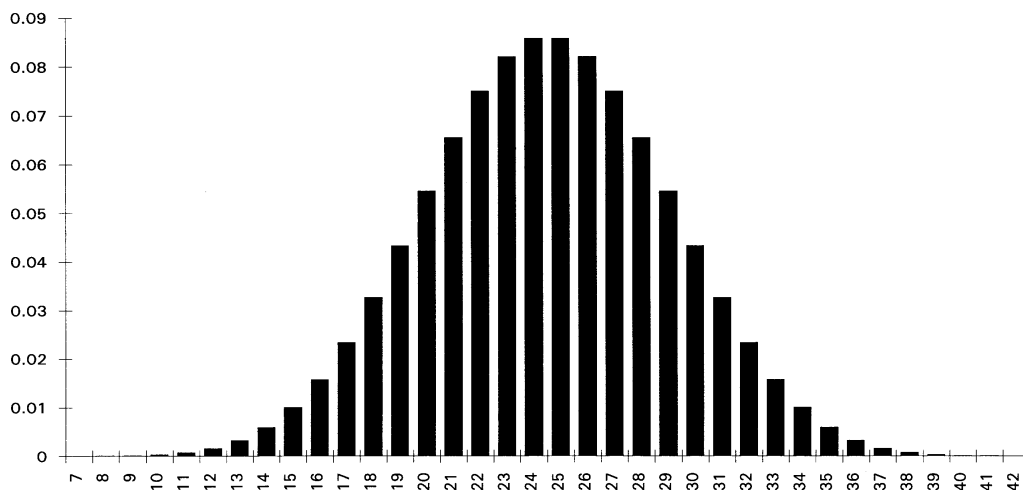
9. Normális- és származtatott eloszlású valószínűségi változók

9.1. Normális eloszlások

9.1. Bevezetés: Nagyon sok fizikai és egyéb mennyiség nagyon sok egyednél történt mérésének tanulmányozása után jutott Gauss arra a következtetésre, hogy a mennyiségek hisztogramjait (oszlopdiaagramjait) jól közelítik az e^{-x^2} függvény lineáris ("vízszintes és függőleges") transzformációi.

Általában olyan mennyiségekről van szó, amelyek nagyon sok, apró +/- hatás összegeződésékként jönnek létre (testmagasság, tömeg, térfogat, feszültség, stb., általában minden fizikai mennyiség).

Ezen feltételezés szemléltetésére sok példát láthatunk: több kocka dobásainak összege (lásd a honlapomon és az alábbi rajzon), Galton deszka (<https://hu.wikipedia.org/wiki/Galton-deszka>), stb. Ezt a jelenséget a 10.5. "Központi Határeloszlás Tétel" (CHT) és a Statisztika tárgy "Illeszkedésvizsgálat" módszere igazolja majd.



5. ábra: Hét kocka összegének eloszlása

<http://math.uni-pannon.hu/~szalkai/7kocka.gif>

9.2. Definíció: ξ standard normális eloszlású v.v., ha sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ ($x \in \mathbb{R}$). \square

9.3. Tétel: $M(\xi) = 0$ és $D(\xi) = 1$. \square

9.4. Megjegyzések: Érdekes alaposan tanulmányoznunk az e^{-x^2} függvény és lineáris transzformáltjainak képleteit és grafikonjait! Lásd még a 6. ábrát és az utána írt megjegyzéseket is!

9.5. Jelölés: $\Phi(x) := F(x) = \int_{-\infty}^x f(t)dt$ ahol $f(x)$ a 9.2.-ben szereplő sűrűségfüggvény ($x \in \mathbb{R}$) □

$\Phi(x)$ nyilván a **standard normális** eloszlású v.v. (kumulatív) **eloszlásfüggvénye**.

Tudjuk, hogy $\Phi(x)$ *szigorúan* növekvő (hiszen $f(x)$ pozitív), és így $0 < \Phi(x) < 1$ minden $x \in \mathbb{R}$ esetén.

9.6. Állítás: $\Phi(-x) = 1 - \Phi(x)$ és $\Phi^{-1}(y) = -\Phi^{-1}(1-y)$ tetszőleges $x, y \in \mathbb{R}$ esetén.

Bizonyítás: $f(x)$ nyilván szimmetrikus az origóra (páros függvény), azaz $f(-x) = f(x)$, ezért bármilyen $0 < x$ esetén

$$\Phi(-x) = F(-x) = P(\xi < -x) = P(x < \xi) = 1 - F(x) = 1 - \Phi(x). \quad \square$$

Következmény: $-\xi$ is standard normális eloszlású v.v.

Megjegyzések: *Liouville* tétele szerint a 9.5. integrál *nem számítható ki* pontosan, azaz *nincs pontos képlet* a $\Phi(x)$ függvényre. Csak közelítőleg lehet $\Phi(x)$ értékeit kiszámolni, egy egyszerűbb táblázatot és néhány közelítő képletet a Függelékben ismertetünk. A 9.6. összefüggés szerint elég csak nemnegatív x értékekre kiszámolnunk és táblázatba írunk $\Phi(x)$ értékeit!

Gyakoroljuk a (pl. a jegyzet végén levő) *táblázat használatát*, például:

$$\Phi(-1.2) = 1 - \Phi(1.2) \approx 1 - 0.8849 = 0.1151,$$

$$\Phi(3.9) \approx 0.9999 \text{ mert } \Phi(3.8) \approx 0.9999 \text{ és } \Phi(3.8) < \Phi(3.9) < 1.$$

$$\Phi^{-1}(0.99) \approx 2.33 \text{ mert } \Phi(2.33) \approx 0.99,$$

$$\Phi^{-1}(0.05) = -\Phi^{-1}(1 - 0.05) = -\Phi^{-1}(0.95) \approx -1.645 \text{ mert } \Phi(-1.645) = 1 - \Phi(1.645) \approx 1 - 0.95 = 0.05.$$

9.7. Definíció: Legyen ξ standard normális eloszlású v.v. és legyenek $m, \sigma \in \mathbb{R}$ tetszőleges valós számok, $\sigma > 0$. Ekkor az $\eta = \sigma \cdot \xi + m$ valószínűségi változót m, σ **paraméterű normális eloszlású v.v.**-nak nevezzük és erre a $\eta \sim N(m, \sigma)$ *jelölést* használjuk! □

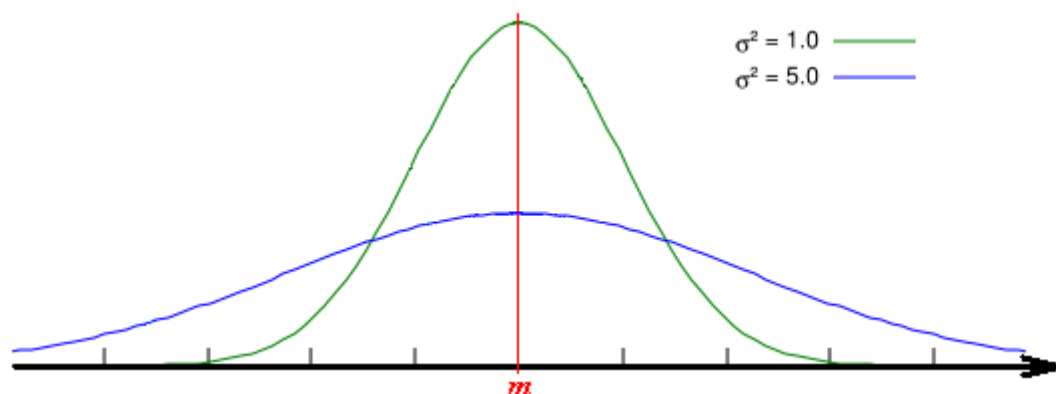
Az alábbi tétel ekvivalens definíciót jelent a normális eloszlások részére:

9.8. Tétel: η sűrűségfüggvénye $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ minden valós x -re. □

9.9. Állítás: $M(\eta) = m$ és $D(\eta) = \sigma$. □

9.10. Tétel: $\eta \sim N(m, \sigma)$ esetén $F_\eta(x) = F_{m,\sigma}(x) = \Phi\left(\frac{x-m}{\sigma}\right)$. □

Megjegyzés: A 9.10. Tétel szerint *bármely* normális eloszlás $F_{m,\sigma}(x)$ eloszlásfüggvénye $\Phi(x)$ -ből kiszámolható, vagyis elég *egyetlen* táblázat számításainkhoz.



6. ábra: Normális eloszlások sűrűségfüggvényei

Látható, hogy az $f(x)$ sűrűségfüggvények mindig *szimmetrikusak* az $x=m$ függőleges egyenesre, és kisebb szórás esetén karcsúbbak és magasabbak, nagyobb szórás esetén pedig szélesebbek és alacsonyabbak. Az 5.17. Megjegyzés és a fentiek alapján, ha m -nek csak "(nagyon) kis környezetét" vizsgáljuk,

akkor a $P(\xi \approx b)$ valószínűség kisebb szórás (σ) esetén nagyobb, nagyobb szórás esetén kisebb. Ezt a jelenséget az alábbi tétel is megerősíti.

9.11. Tétel: "k-szor szigma szabály" :

Ha $\xi \sim N(m, \sigma)$ és $k \in \mathbb{R}^+$ tetszőleges pozitív szám, akkor $P(m - k\sigma < \xi < m + k\sigma) = 2\Phi(k) - 1$.

Az egyenlőség könnyen kiszámolható a 9.10. és 9.6. összefüggések alapján, tanulságos házi feladat. \square

Jelentése: Annak a valószínűsége, hogy ξ értéke a várható értékének "k*szoꝛása" sugarú környezetébe esik, pontosan $2 \cdot \Phi(k) - 1$.

Ha $k \cdot \sigma$ helyett ε -t írunk, akkor a tétel így írható: $P(m - \varepsilon < \xi < m + \varepsilon) = 2 \cdot \Phi(\varepsilon/\sigma) - 1$

mert $P(m - \varepsilon < \xi < m + \varepsilon) = F(m + \varepsilon) - F(m - \varepsilon) = \Phi(\varepsilon/\sigma) - \Phi(-\varepsilon/\sigma) = \Phi(\varepsilon/\sigma) - (1 - \Phi(\varepsilon/\sigma)) = 2 \cdot \Phi(\varepsilon/\sigma) - 1$.

A tétel speciális eseteit érdemes külön is kiszámolnunk:

9.12. Tétel: Speciális esetek:

- $k=1 \Rightarrow P(m - \sigma < \xi < m + \sigma) = 0.68$,
- $k=2 \Rightarrow P(m - 2\sigma < \xi < m + 2\sigma) = 0.95$,
- $k=3 \Rightarrow P(m - 3\sigma < \xi < m + 3\sigma) = 0.997$. \square

9.13. Tétel: Ha $\xi \sim N(m, \sigma)$ és $\eta = a\xi + b$, $a \neq 0$, akkor $\eta \sim N(am + b, |a|\sigma)$. \square

9.14. Tétel: Ha $\xi \sim N(m_1, \sigma_1)$ és $\eta \sim N(m_2, \sigma_2)$ független valószínűségi változók, akkor

$$\xi \pm \eta \sim N\left(m_1 \pm m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) . \quad \square$$

Megjegyzés: Az általános 6.4. és 6.10. Tételekből már tudjuk, hogy $\xi \pm \eta$ várható értéke $m_1 \pm m_2$ és szórása $\sqrt{\sigma_1^2 + \sigma_2^2}$, a Tétel újdonsága az, hogy: "normális eloszlások összege is normális eloszlás" ! \square

Az alábbi, gyakorlatban is fontos összefüggés már könnyen következik az előző Tételből:

9.15. Tétel: Ha $\xi_1, \xi_2, \dots, \xi_n \sim N(m, \sigma)$ független, azonos eloszlású normális v.v.-k, akkor

$$\sum_{i=1}^n \xi_i \sim N(nm, \sigma\sqrt{n}) \quad \text{és} \quad \frac{\sum_{i=1}^n \xi_i}{n} \sim N\left(m, \frac{\sigma}{\sqrt{n}}\right) . \quad \square$$

Megjegyzés: Mint már többször említettük, a gyakorlatban egy mérés helyett többször mérünk és átlagolunk, ezáltal a szórás csökken, a pontosság nő. Jó tudnunk azt is, hogy az összeg és az átlag is normális eloszlású v.v. \square

9.16. Példa: egy felnőtt tömege normális eloszlású valószínűségi változó 75 kg várható értékkel és 15 kg szórással , egy iskolás gyerek tömege normális eloszlású 35 kg várható értékkel és 6 kg szórással. Ha a két személy tömegét független valószínűségi változónak tekintjük, akkor

- a) mekkora valószínűséggel lesz egy felnőtt tömege nagyobb, mint egy gyerek tömege,
- b) mennyi a valószínűsége annak, hogy az össztömegük 80 és 140 kg közé esik?

Megoldás: Legyen $\xi_1 =$ egy felnőtt tömege, $\xi_2 =$ egy gyermek tömege. Ekkor

a) $P(\xi_1 > \xi_2) = P(\xi_1 - \xi_2 > 0) = 1 - F_\eta(0)$ ahol $\eta = \xi_1 - \xi_2$. A tételek szerint $M(\eta) = M(\xi_1) - M(\xi_2) = 75 - 35 = 40$ és $D^2(\eta) = D^2(\xi_1) + D^2(\xi_2) = 15^2 + 6^2 = 261$ azaz $D(\eta) = \sqrt{261} \approx 16.155$, tehát $\eta \sim N(40, 16.155)$.

Tehát $P(\eta > 0) = 1 - F_\eta(0) = 1 - \Phi\left(\frac{0 - 40}{16.155}\right) \approx 1 - \Phi(-2.476) = 1 - (1 - \Phi(2.476)) = \Phi(2.476) \approx 0.9931$, vagyis

(a fenti adatokkal) 99% valószínűséggel lesz egy felnőtt tömege nagyobb, mint egy gyerek tömege.

b) $P(80 < \xi_1 + \xi_2 < 140) = P(80 < \theta < 140) = F_\theta(140) - F_\theta(80)$ ahol $\theta = \xi_1 + \xi_2$. A tételek szerint $M(\theta) = M(\xi_1) + M(\xi_2) = 75 + 35 = 110$ és $D^2(\theta) = D^2(\xi_1) + D^2(\xi_2) = 261$, azaz $D(\theta) = \sqrt{261} \approx 16.155$, tehát $\theta \sim N(110, 16.155)$. Ekkor pedig

$$F_{\theta}(140)-F_{\theta}(80) = \Phi\left(\frac{140-110}{16.155}\right) - \Phi\left(\frac{80-110}{16.155}\right) \approx \Phi(1.857)-\Phi(-1.857) = \Phi(1.857) - (1-\Phi(1.857)) =$$

$\approx 0,9686 - (1-0,9686) \approx 0,9372$. Tehát (a fenti adatokkal) 93.7% valószínűséggel esik egy felnőtt és egy gyermek össztömege 80 és 140 kg közé.

9.17. Példa: Egy liftet 8 felnőtt személyre méreteznek. A beszállók tömegét független normális eloszlású valószínűségi változónak tekintjük 75 kg várható értékkel és 15 kg szórással. Mennyi legyen a lift teherbíró képessége, ha azt szeretnénk, hogy 4 személy beszállása esetén 0.99 valószínűséggel ne gyulladjon ki a túlterheltséget jelző lámpa?

Megoldás: Tehát $\xi_i \sim N(75,15)$, $i=1,\dots,8$, $m=75$, $\sigma=15$, és legyen $\eta := \sum_{i=1}^8 \xi_i$.

Ekkor, a 9.15. Tétel szerint $\eta \sim N(8 \cdot m, \sqrt{8} \cdot \sigma) = N(8 \cdot 75, \sqrt{8} \cdot 15) = N(600, 42.43)$.

Olyan $x \in \mathbb{R}$ valós számot kell keresnünk, amelyre $P\left(\sum_{i=1}^8 \xi_i < x\right) = P(\eta < x) = 0.99$.

Mivel $P(\eta < x) = F_{\eta}(x) = \Phi\left(\frac{x-600}{42.43}\right) = 0.99$, és Φ táblázatából $\Phi^{-1}(0.99) \approx 2.32$ (mert $\Phi(2.32) \approx 0.99$),

ezért $\frac{x-600}{42.43} = 2.32$, ahonnan $x \approx 698.5 \approx 700 \text{ kg}$. □

9.2. Normális eloszlásból származtatott eloszlások

A következő összefüggésekre a Matematikai Statisztika tárgyban (következő félév) lesz majd szükségünk.

9.18. Tétel: Ha $\xi \sim N(0,1)$ akkor $\eta = \xi^2$ eloszlás- és sűrűségfüggvénye

$$F_{\eta}(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ 2\Phi(\sqrt{x}) - 1 & \text{ha } x > 0 \end{cases} \quad \text{és} \quad f_{\eta}(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} & \text{ha } x > 0 \\ 0 & \text{különben} \end{cases},$$

valamint várható értéke $M(\eta) = 1$.

Biz: $F_{\eta} = P(\eta < x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = \Phi(\sqrt{x}) - (1 - \Phi(\sqrt{x})) = 2\Phi(\sqrt{x}) - 1$, $f_{\eta} = F'_{\eta} = \dots$. □

9.19. Definíció: Tetszőleges független $\xi_1, \dots, \xi_n \sim N(0,1)$ v.v. ($n \in \mathbb{N}$) esetén az $\eta_n := (\xi_1)^2 + \dots + (\xi_n)^2$ v.v.-t n -szabadságfokú (vagy: n -paraméterű) χ^2 , vagyis **khí-négyzet** (chi-squared) eloszlásnak nevezzük. □

9.20. Tétel: A fenti η_n sűrűségfüggvénye $f_{\eta}(x) = c x^{\frac{n-1}{2}} e^{-\frac{x}{2}}$ ($0 < x$). □

9.21. Definíció: Tetszőleges független $\zeta, \xi_1, \dots, \xi_n \sim N(0,1)$ v.v. ($n \in \mathbb{N}$) esetén az

$$\theta_n = \frac{\zeta}{\sqrt{\frac{\sum_{i=1}^n \xi_i^2}{n}}}$$

v.v.-t n -szabadságfokú (vagy: n -paraméterű) **Student-** vagy **t-**eloszlásnak nevezzük □

9.22. Tétel: A fenti θ_n sűrűségfüggvénye:

$$f_{\theta}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right) \cdot \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad \text{ahol} \quad \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt ,$$

és $\Gamma(x)$ az úgynevezett **gamma** függvény ($n \in \mathbb{N}$ egész szám esetén $\Gamma(n) = (n-1)!$). □

9.23. Tétel: Ha $\xi \sim N(0,1)$ akkor $\eta = e^{\xi}$ (ún. "**lognormális eloszlás**") eloszlás- és sűrűségfüggvénye

$$F_{\eta}(x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ \Phi(\ln x) & \text{ha } x > 0 \end{cases} \quad \text{és} \quad f_{\eta}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{(\ln x)^2}{2}} & \text{ha } x > 0 \\ 0 & \text{különben} \end{cases} ,$$

valamint várható értéke $M(\eta) = \sqrt{e}$.

Biz: $F_{\eta} = P(\eta < x) = P(e^{\xi} < x) = P(\xi < \ln(x)) = \Phi(\ln(x))$, $f_{\eta} = F'_{\eta} = \dots$ □

Móduszok és mediánok összefoglalása

Valószínűségi változók móduszait és mediánjait általában az 5.19. és 5.20. pontokban definiáltuk. Most csak felsoroljuk a nevezetes eloszlások móduszait és mediánjait.

Diszkrét egyenletes mod = *mindegyik* x_k érték

med = $(x_{k-1} + x_k)/2$ ha n páros, ahol $k=1+n/2$,

med = x_k ha n páratlan, és k a legnagyobb olyan, amelyre $(k-1)/n < 1/2$,

Hipergeometriai mod = $\left[(n+1) \frac{K+1}{N+2} \right]$

Binomiális mod = $[(n+1)p]$ ha $(n+1)p \notin \mathbb{Z}$,

mod₁ = $(n+1)p-1$ és mod₂ = $(n+1)p$ ha $(n+1)p \in \mathbb{Z}$,

Poisson mod = $[\lambda]$ ha $\lambda \notin \mathbb{Z}$,

mod₁ = $\lambda-1$ és mod₂ = λ ha $\lambda \in \mathbb{Z}$,

Geometriai mod = 1 ("Szinbád"),

med = $(2k-1)/2$ ha $k = \log_q\left(\frac{1}{2}\right) + 1$ egész szám,

med = $\left[\log_q\left(\frac{1}{2}\right) + 1 \right]$ ha $\log_q\left(\frac{1}{2}\right) + 1$ nem egész szám,

Folyt. egyenletes mod = nincs, med = $(a+b)/2$,

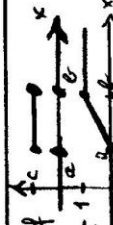
Exponenciális mod = 0, med = $\ln(2)/\lambda$,

Normális mod = med = $M(\xi)$ (inflexiós pontok $m \pm \sigma$),

A következő táblázat nagy méretben megtalálható a honlapomon:

[http://math.uni-pannon.hu/~szalkai/Eloszlasok\(pdf\)+kezjav+.gif](http://math.uni-pannon.hu/~szalkai/Eloszlasok(pdf)+kezjav+.gif)

Nevezetes eloszlások

Név	Paraméter	Definíció	$M = \mu$	σ^2	Alkalmazás
Indikátorváltozó	p	$p_1 = p, p_0 = 1 - p$	p	$p(1 - p)$	Egy p valószínűségű esemény bekövetkezik-e?
Binomiális eloszlás	n, p	$p_k = \binom{n}{k} p^k (1 - p)^{n-k} \quad (k = 0, 1, \dots, n)$	np	$np(1 - p)$	Viszatevéses mintavétel. Egy p valószínűségű csemény n független ismétlés során hányszor következik be?
Geometriai eloszlás	p	$p_k = (1 - p)^{k-1} \cdot p \quad (k = 0, 1, \dots)$	$\frac{1}{p}$	$\frac{(1 - p)}{p^2}$	
Hipergeometrikus eloszlás	N, K, n	$p_k = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n)$	$\frac{Kn}{N} = np$	$n \cdot \frac{K}{N} \left(1 - \frac{K}{N}\right) \cdot \frac{N-n}{N-1} = np \cdot q \cdot \left(\frac{N-n}{N-1}\right)$	Viszatevés nélküli mintavétel. Megjegyzés: A $p = \frac{K}{N}$ jelöléssel μ és σ egyszerűbb alakra írható.
Poisson-eloszlás	$\lambda (> 0)$	$p_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots)$	λ	λ	Egységnyi idő alatt megfigyelt események száma. Egységnyi területre eső(hibák) száma.
Standard normális eloszlás	—	$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	0	1	Normális eloszlás standardizáltja.
Normális eloszlás	$\mu, \sigma (> 0)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	Populáció egyedeinek méretei, tömegei.
Egyenletes eloszlás	a, b	 $c = \frac{1}{b-a}$ $F(x) = \frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\left(\frac{b-a}{\sqrt{12}}\right)^2$	Gyártási folyamatban fellépő mérteingadozások. Fizikai mennyiségek.
Exponenciális eloszlás	$\lambda (> 0)$	$f(x) = \lambda e^{-\lambda x} \quad ha \ x > 0$ $F(x) = 1 - e^{-\lambda x} \quad ha \ x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Élettartam, várakozási idő. Időintervallumok a Poisson-eloszláshoz kapcsolódóan.

10. Nagy számok törvényei

Az alábbi egyenlőtlenségek általános közelítéseket adnak bármely η valószínűségi változó értékeinek eloszlásáról. "Természetesen" a közelítések pontosságát a mérések számának növelésével javíthatjuk.

10.1. Tétel: Markov-egyenlőtlenség

Ha a ξ valószínűségi változónak létezik $M(\xi)$ várható értéke, akkor $\forall a \in \mathbb{R}^+$ esetén

$$P(\xi \geq a) \leq \frac{M(\xi)}{a} .$$

Magyarázat: A tétel szerint a mérés eredménye (ξ) egyre nagyobb értéket egyre kisebb valószínűséggel vesz fel. Pontosabban: ξ bármilyen nagy lehet ($\xi \geq a$), de ennek az eseménynek a valószínűsége legfeljebb M/a , amely korlát $a \rightarrow \infty$ esetén 0-hoz tart.

10.2. Tétel: Csebisev-egyenlőtlenség

Ha a ξ valószínűségi változónak létezik $M(\xi)$ várható értéke és $D(\xi)$ szórása, akkor $\forall k \in \mathbb{R}^+$ és $\forall \varepsilon \in \mathbb{R}^+$ esetén

$$P(|\xi - M(\xi)| \geq kD(\xi)) \leq \frac{1}{k^2}, \text{ vagy másként } P(|\xi - M(\xi)| \geq \varepsilon) \leq \frac{D^2(\xi)}{\varepsilon^2},$$

ill. az esemény tagadása:

$$P(|\xi - M(\xi)| \leq \varepsilon) \geq 1 - \frac{D^2(\xi)}{\varepsilon^2} .$$

Magyarázat: Azt várjuk, hogy a mérés (ξ) értékei az átlag ($M=M(\xi)$) körül ingadoznak, ezt igazolja a Csebisev-egyenlőtlenség: a mérés és az átlag eltérése (vagyis $|\xi - M|$) lehet ugyan nagy, vagyis lehetséges, hogy $|\xi - M| \geq \varepsilon$, de ennek a valószínűsége legfeljebb D^2/ε^2 . Ez a korlát pedig 0-hoz tart amennyiben $k \rightarrow \infty$. Természetesen ez a korlát is kisebb ha a D szórás is kicsi.

10.3. Tétel: Bernoulli-féle nagy számok törvénye

Legyen A egy tetszőleges esemény amelyre $P(A)=p$. Végezzünk n (független) kísérletet és jelölje ξ_n az A esemény gyakoriságát (hányszor "sikerült" A), ekkor a relatív gyakoriság $= \xi_n/n$. Ekkor tetszőleges $\varepsilon > 0$ és $n \in \mathbb{N}$ esetén ($q=1-p$):

$$P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{\varepsilon^2 n}, \text{ vagyis } P\left(\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{\varepsilon^2 n} .$$

Átfogalmazva: tetszőleges $\varepsilon > 0$ és $\delta > 0$ esetén $\exists n_0$ hogy minden $n > n_0$ esetén

$$P\left(\left|\frac{\xi_n}{n} - p\right| < \varepsilon\right) \geq 1 - \delta, \text{ vagy másként } P\left(\left|\frac{\xi_n}{n} - p\right| \geq \varepsilon\right) \leq \delta .$$

(ahol $\delta = \frac{pq}{\varepsilon^2 n} \rightarrow 0$ ha $n \rightarrow \infty$.)

Magyarázat: A tétel szerint az elméleti valószínűség (p) és a tapasztalati (gör: empirikus) relatív gyakoriság (azaz ξ_n/n) eltérése (vagyis $|\xi_n/n - p|$) **kicsi**, sőt tetszőleges ε számnál kisebb lehet - legalábbis majdnem 100% valószínűséggel igaz a mondat előző fele.

Másképpen fogalmazva (jobboldali képlet): az eltérés lehet nagy, de csak kis valószínűséggel. (Ez nem jelenti azt, hogy a relatív gyakoriság konvergál a valószínűséghez, hanem csak azt, hogy a nagy eltérés valószínűsége kicsi!) Ha p nem ismert, akkor a $p(1-p) \leq 1/4$ becslés alapján $\delta \leq 1/4\varepsilon^2 n$ is írható.

10.4. Tétel: A nagy számok gyenge törvénye (Csebisev-alak)

Legyenek $\xi_1, \xi_2, \dots, \xi_n$ független, azonos eloszlású valószínűségi változók, amelyeknek létezik (azonos) várható értékük és szórásuk: $m := M(\xi_i)$ és $\sigma := D(\xi_i)$, és legyen $S_n := \xi_1 + \xi_2 + \dots + \xi_n$. Ekkor minden $\varepsilon \in \mathbb{R}^+$ esetén

$$P\left(\left|\frac{S_n}{n} - m\right| \leq \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}, \text{ vagy másként } P\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Magyarázat: $\xi_1, \xi_2, \dots, \xi_n$ ugyanazon mennyiség többszöri független mérését jelölik (azonos eloszlásúak és függetlenek), így S_n/n éppen ezen mérések (tapasztalati) **átlaga**. A tétel pedig azt mondja ki, hogy a *tapasztalati* átlag (S_n/n) és az *elméleti* átlag (m) eltérése (azaz $|S_n/n - m|$) mekkora lehet. Például tetszőleges rögzített ε korlát esetén az első képletben szereplő hibátag $\sigma^2/n\varepsilon^2 \rightarrow 0$ midőn $n \rightarrow \infty$, vagyis: a kísérletek számának növelésével ($n \rightarrow \infty$) a tapasztalati és az elméleti átlag eltérés ε -nál kisebb (ill. ezen állítás 1-hez közeli valószínűséggel igaz).

E tételnek speciális esete a Bernoulli-féle törvény, ugyanis a 10.3.Tételben $m=p$ és $\sigma^2=pq$.

10.5. Tétel: Központi (=centrális) határeloszlás tétel (CHT) vagy Nagy számok erős törvénye

Legyenek $\xi_1, \xi_2, \dots, \xi_n, \dots$ független, azonos eloszlású valószínűségi változók, amelyeknek létezik (azonos) várható értékük és szórásuk: $M(\xi_i)=m$ és $D(\xi_i)=\sigma$ ($i=1,2,\dots,n,\dots$). Ekkor **összegük** (pontosabban: standardizált átlaguk) közelítőleg normális eloszlású: a

$$\zeta_n = \frac{\xi_1 + \dots + \xi_n - n \cdot m}{\sigma\sqrt{n}} \text{ jelöléssel } \lim_{n \rightarrow \infty} P(\zeta_n < y) = \Phi(y) .$$

Magyarázat: Mivel mindegyik $\xi_1, \xi_2, \dots, \xi_n \dots$ mérés átlagosan m , így összegük $+\infty$ -be tart: $\lim_{n \rightarrow \infty} (\xi_1 + \xi_2 + \dots + \xi_n) = +\infty$, ezért kell az összeg ζ_n (standardizált) változatát tekintenünk. Mivel a $P(\zeta_n \leq y)$ kifejezés éppen az ζ_n valószínűségi változó eloszlásfüggvénye és Φ a (standard) Normális eloszlás eloszlásfüggvénye, ezért a Tétel utolsó képlete valóban azt igazolja, hogy sok azonos (apró) hatás összeg valóban a Normális eloszláshoz tart.

10.6. Moivre-Laplace tétel:

Tetszőleges $0 \leq p \leq 1$ és $u, v \in \mathbb{R}$ valós számokra

$$\lim_{n \rightarrow \infty} \left(\sum_{u \leq k \leq v} \binom{n}{k} p^k q^{n-k} \right) = \Phi(v^*) - \Phi(u^*)$$

vagy egyszerűbben

$$\sum_{u \leq k \leq v} \binom{n}{k} p^k q^{n-k} \approx \Phi(v^*) - \Phi(u^*)$$

ahol

$$u^* = \frac{u - np}{\sqrt{npq}} = \frac{u - m}{\sigma} \quad \text{és} \quad v^* = \frac{v - np}{\sqrt{npq}} = \frac{v - m}{\sigma} .$$

Magyarázat: A tétel szerint nagyon sok azonos kísérlet elvégezve az, hogy a sikeres kísérletek száma u és v közé esik, normális eloszlással közelíthető.

Más szavakkal: a binomiális eloszlást nagy n esetén Normális eloszlással közelíthetjük.

Ötlet: A második (közelítő) képlet hasonlít a már "megszokott"

$$P(u < \xi < v) = F(v) - F(u) = \Phi(v^*) - \Phi(u^*) = \Phi\left(\frac{v - m}{\sigma}\right) - \Phi\left(\frac{u - m}{\sigma}\right) .$$

formulához! A tételt szokták a következő alakban is írni:

$$\lim_{n \rightarrow \infty} \left(\sum_{\substack{k - np \\ \sqrt{npq}} \leq v^*} \binom{n}{k} p^k q^{n-k} \right) = \Phi(v^*) - \Phi(u^*)$$

A Moivre-Laplace tétel a 10.5. Tétel speciális esete

Megjegyzés: Mint tanultuk: nagy n esetén a hipergeometrikus eloszlás közelíthető a binomiális eloszlással, ami pedig a Poisson eloszlással is közelíthető, és végül, a Poisson eloszlás is közelíthető a normális eloszlással.

10.7. Mintafeladat: $n=1000$ állat van egy farmon. Egy *nem* járványos betegség esetén (vagyis a megbetegedések függetlenek) $P(\text{meggyógyul})=0.7$. Mekkora eséllyel lesz a meggyógyult állatok száma $n/3$ és $2n/3$ között?

Megoldás:

1. rész: A feladat valójában binomiális (Bernoulli) eloszlású:

$n=1000$, $p=0,7$, $q=1-p=0,3$, $\xi=1000$ állat közül mennyi gyógyul meg.

Binomiális eloszlás: $P(\xi=k) = \binom{n}{k} p^k (1-p)^{n-k}$ ($k=0,1,\dots,n$),

$$P(334 \leq \xi \leq 666) = \sum_{k=334}^{k=666} \binom{1000}{k} 0,7^k 0,3^{1000-k}.$$

2. rész: Mivel n nagy, ezért a számolási nehézségek miatt közelíthetünk Poisson eloszlással.

Poisson-eloszlás: $\lambda = np = 1000 \cdot 0,7 = 700$, $P(\xi=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ($k=0,1,\dots$),

$$P(334 \leq \xi \leq 666) = \sum_{k=334}^{k=666} \frac{700^k}{k!} \cdot e^{-700}.$$

3. rész: Mivel n nagyon nagy, ezért a még mindig fennálló számolási nehézségek miatt a Moivre-Laplace tétellel kell közelítenünk és számolnunk.

Moivre - Laplace tételt használata

$$\sum_{u \leq k \leq v} \binom{n}{k} p^k q^{n-k} \approx \Phi(v^*) - \Phi(u^*) \quad , \text{ ahol: } u^* = \frac{u-np}{\sqrt{npq}} = \frac{u-m}{\sigma} \text{ és } v^* = \frac{v-np}{\sqrt{npq}} = \frac{v-m}{\sigma},$$

$M=m=np=1000 \cdot 0,7=700$, $D=\sigma=\sqrt{npq}=\sqrt{1000 \cdot 0,7 \cdot 0,3}=\sqrt{210} \approx 14,49$,

$$P(334 \leq \xi \leq 666) = \sum_{k=334}^{k=666} \binom{1000}{k} 0,7^k 0,3^{1000-k} \approx \Phi\left(\frac{666-700}{\sqrt{210}}\right) - \Phi\left(\frac{334-700}{\sqrt{210}}\right) =$$

$$= \Phi(-2,35) - \Phi(-25,26) = 1 - \Phi(2,35) - 0 = 1 - 0,99065 = \mathbf{0,00935}.$$

□

10.8. Mintafeladat: Két kockával gurítunk. Legyen nyereményünk annyiszor 100Ft, amekkora a dobott számok nagyobbika (maximuma).

a) Mekkora valószínűséggel lesz 200 játék után össznyereményünk legalább 85 000Ft ?

b) Ha egy játék ára 450 Ft, akkor 300 játék után mekkora valószínűséggel leszünk nyerőben (azaz nyereményünk pozitív) ?

Megoldás: Legyen $\xi :=$ nyereményünk *egy* játékban. Ekkor ξ diszkrét v.v., de nem nevezetes. Lehetséges értékei $\text{Im}(\xi) = \{100,200,300,400,500,600\}$, melyek valószínűségei $\{1/36, 3/36, 5/36, 7/36, 9/36, 11/36\}$ (házi feladat!). Így $M(\xi) = 100 \cdot 1/36 + 200 \cdot 3/36 + 300 \cdot 5/36 + 400 \cdot 7/36 + 500 \cdot 9/36 + 600 \cdot 11/36 \approx 447.22$,

$$M(\xi^2) = 100^2 \cdot 1/36 + 200^2 \cdot 3/36 + 300^2 \cdot 5/36 + 400^2 \cdot 7/36 + 500^2 \cdot 9/36 + 600^2 \cdot 11/36 \approx 219\,720,$$

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = 219\,720 - 447.22^2 \approx 19\,714, \quad D(\xi) \approx \sqrt{19714} \approx \underline{140.41}.$$

a) 200 független játék összege $\eta = \xi_1 + \xi_2 + \dots + \xi_{200}$, ekkor $m_\eta = M(\eta) = 200 \cdot M(\xi) \approx 89\,444$ és $\sigma_\eta = D(\eta) = \sqrt{200} \cdot D(\xi) \approx 1985.65$. A CHT tétel (10.5.) szerint η közelítőleg normális eloszlású, m_η és σ_η paraméterekkel, ezért $P(85\,000 < \eta) = 1 - F_\eta(85\,000) = 1 - \Phi\left(\frac{85\,000 - 89\,444}{1985.65}\right) \approx 1 - \Phi(2.24) = 1 - (1 - \Phi(2.24)) \approx \Phi(2.24) \approx 0,9875 \approx 99\%$.

b) Ha egy játék ára 450 Ft, akkor nyereményünk *egy* játékban $\gamma := \xi - 450$, amelynek eloszlása $\text{Im}(\gamma) = \{-350, -250, -150, -50, 50, 150\}$, melyek valószínűségei ismét $\{1/36, 3/36, 5/36, 7/36, 9/36, 11/36\}$. Tehát (a 6. fejezet tételei alapján) $M(\gamma) = M(\xi) - 450 \approx -2.78$ és $D(\gamma) = D(\xi) \approx \underline{140.41}$.

300 független játék összege $\delta = \gamma_1 + \gamma_2 + \dots + \gamma_{300}$, ekkor $m_\delta = M(\delta) = 300 \cdot M(\gamma) \approx -834$ és $\sigma_\delta = D(\delta) = \sqrt{300} \cdot D(\gamma) \approx 2432$. A CHT tétel szerint δ közelítőleg normális eloszlású, ezért $P(0 < \delta) = 1 - F_\delta(0) \approx 1 - \Phi\left(\frac{0 - (-834)}{2432}\right) \approx 1 - \Phi(0,343) \approx 1 - 0.6333 = 0,3667 \approx 37\%$.

Függelék

Valószínűségszámítás - Matematika szótár

Valószínűségszámítás	Matematika
Eseménytér (=kísérlet összes lehetséges kimenetele) kísérlet végeredménye esemény (=kísérlet aktuális kimenetele) elemi esemény A esemény bekövetkezik A esemény nem következik be biztos esemény lehetetlen esemény ellentett esemény (esemény tagadása) események összege: $A+B$ ("vagy") szorzata: $A \cdot B$ ("és") különbsége: $A-B$ kizáró események A -ból következik B (=A maga után vonja B-t)	$H \neq \emptyset$ (tetszőleges) alaphalmaz, inkább az Ω jelet használjuk $x \in \Omega$ tetszőleges elem, $A \subseteq \Omega$ (tetszőleges) részhalmaz, $\{x\} \subseteq \Omega$ egyelemű részhalmaz ("singleton"), $x \in A$ $x \notin A$ $\Omega \subseteq \Omega$ (az alaphalmaz), ill. ha $P(A)=1$, $\emptyset \subseteq \Omega$ (üres halmaz), ill. ha $P(A)=0$, A^c (komplementer halmaz), $A \cup B$ (únió), $A \cap B$ (metszet), $A \setminus B$ (különbség), $A \cap B = \emptyset$ (diszjunkt halmazok), ill. ha $P(A \cap B)=0$, $A \subseteq B$ (A részhalmaza B -nek),
valószínűség $P(A)$	$P: P(\Omega) \rightarrow \mathbb{R}$ tetsz. függvény ($P(\Omega)=\Omega$ hatványhalmaza) a Kolmogorov axiómákkal, (pl. terület),
független események A,B teljes eseményrendszer	$P(A \cap B) = P(A) \cdot P(B)$, Ω egy partíciója (felosztása),
valószínűségi változó (mérés számszerű végeredménye) diszkrét v.v. - " - eloszlása folytonos v.v.	$\xi: \Omega \rightarrow \mathbb{R}$ tetszőleges függvény, $\text{Im}(\xi) = \{x_1, x_2, \dots, x_n, \dots\}$ felsorolható, $\{p_1, p_2, \dots, p_n, \dots\}$ ahol $p_i = P(\xi=x_i)$ ha $i \in \mathbb{N}$ van $[a, b] \subseteq \text{Im}(\xi)$ intervallum,
eloszlásfüggvény	$F: \mathbb{R} \rightarrow \mathbb{R}$ tetszőleges függvény az axiómákkal, vagy: $F(t) = P(\xi < t)$ vagy: f primitív függvénye: $F(t) = \int_{-\infty}^t f(x) dx$
sűrűségfüggvény	$f: \mathbb{R} \rightarrow \mathbb{R}$ tetszőleges függvény az axiómákkal, vagy: F deriváltfüggvénye: $f(x) = F'(x)$ Newton-Leibniz szabály
$P(a \leq \xi < b) = \int_a^b f(x) dx = F(b) - F(a)$	
várható érték: $M(\xi)$ vagy $E(\xi)$ szórás: $D(\xi)$	átlag (=számítási közép, mean (angolul)) "szóródás" (=dispersion (latin))

A standard normális eloszlású változó eloszlásfüggvényének (Φ) táblázata

x	$\Phi(x)$	z	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,5000	0,34	0,6331	0,68	0,7517	1,02	0,8461
0,01	0,5040	0,35	0,6368	0,69	0,7549	1,03	0,8485
0,02	0,5080	0,36	0,6406	0,70	0,7580	1,04	0,8508
0,03	0,5120	0,37	0,6443	0,71	0,7611	1,05	0,8531
0,04	0,5160	0,38	0,6480	0,72	0,7642	1,06	0,8554
0,05	0,5199	0,39	0,6517	0,73	0,7673	1,07	0,8577
0,06	0,5239	0,40	0,6554	0,74	0,7704	1,08	0,8599
0,07	0,5279	0,41	0,6591	0,75	0,7734	1,09	0,8621
0,08	0,5319	0,42	0,6628	0,76	0,7764	1,10	0,8643
0,09	0,5359	0,43	0,6664	0,77	0,7794	1,11	0,8665
0,10	0,5398	0,44	0,6700	0,78	0,7823	1,12	0,8686
0,11	0,5438	0,45	0,6736	0,79	0,7852	1,13	0,8708
0,12	0,5478	0,46	0,6772	0,80	0,7881	1,14	0,8729
0,13	0,5517	0,47	0,6808	0,81	0,7910	1,15	0,8749
0,14	0,5557	0,48	0,6844	0,82	0,7939	1,16	0,8770
0,15	0,5596	0,49	0,6879	0,83	0,7967	1,17	0,8790
0,16	0,5636	0,50	0,6915	0,84	0,7995	1,18	0,8810
0,17	0,5675	0,51	0,6950	0,85	0,8023	1,19	0,8830
0,18	0,5714	0,52	0,6985	0,86	0,8051	1,20	0,8849
0,19	0,5753	0,53	0,7019	0,87	0,8078	1,21	0,8869
0,20	0,5793	0,54	0,7054	0,88	0,8106	1,22	0,8888
0,21	0,5832	0,55	0,7088	0,89	0,8133	1,23	0,8907
0,22	0,5871	0,56	0,7123	0,90	0,8159	1,24	0,8925
0,23	0,5910	0,57	0,7157	0,91	0,8186	1,25	0,8944
0,24	0,5948	0,58	0,7190	0,92	0,8212	1,26	0,8962
0,25	0,5987	0,59	0,7224	0,93	0,8238	1,27	0,8980
0,26	0,6026	0,60	0,7257	0,94	0,8264	1,28	0,8997
0,27	0,6064	0,61	0,7291	0,95	0,8289	1,29	0,9015
0,28	0,6103	0,62	0,7324	0,96	0,8315	1,30	0,9032
0,29	0,6141	0,63	0,7357	0,97	0,8340	1,31	0,9049
0,30	0,6179	0,64	0,7389	0,98	0,8365	1,32	0,9066
0,31	0,6217	0,65	0,7422	0,99	0,8389	1,33	0,9082
0,32	0,6255	0,66	0,7454	1,00	0,8413	1,34	0,9099
0,33	0,6293	0,67	0,7486	1,01	0,8438	1,35	0,9115

x	$\Phi(x)$
1,36	0,9131
1,37	0,9147
1,38	0,9162
1,39	0,9177
1,40	0,9192
1,41	0,9207
1,42	0,9222
1,43	0,9236
1,44	0,9251
1,45	0,9265
1,46	0,9279
1,47	0,9292
1,48	0,9306
1,49	0,9319
1,50	0,9332
1,51	0,9345
1,52	0,9357
1,53	0,9370
1,54	0,9382
1,55	0,9394
1,56	0,9406
1,57	0,9418
1,58	0,9429
1,59	0,9441
1,60	0,9452
1,61	0,9463
1,62	0,9474
1,63	0,9484
1,64	0,9495
1,65	0,9505
1,66	0,9515
1,67	0,9525
1,68	0,9535
1,69	0,9545

x	$\Phi(x)$
1,70	0,9554
1,71	0,9564
1,72	0,9573
1,73	0,9582
1,74	0,9591
1,75	0,9599
1,76	0,9608
1,77	0,9616
1,78	0,9625
1,79	0,9633
1,80	0,9641
1,81	0,9649
1,82	0,9656
1,83	0,9664
1,84	0,9671
1,85	0,9678
1,86	0,9686
1,87	0,9693
1,88	0,9699
1,89	0,9706
1,90	0,9713
1,91	0,9719
1,92	0,9726
1,93	0,9732
1,94	0,9738
1,95	0,9744
1,96	0,9750
1,97	0,9756
1,98	0,9761
1,99	0,9767
2,00	0,9772
2,02	0,9783
2,04	0,9793
2,06	0,9803

x	$\Phi(x)$
2,08	0,9812
2,10	0,9821
2,12	0,9830
2,14	0,9838
2,16	0,9846
2,18	0,9854
2,20	0,9861
2,22	0,9868
2,24	0,9875
2,26	0,9881
2,28	0,9887
2,30	0,9893
2,32	0,9898
2,34	0,9904
2,36	0,9909
2,38	0,9913
2,40	0,9918
2,42	0,9922
2,44	0,9927
2,46	0,9931
2,48	0,9934
2,50	0,9938
2,52	0,9941
2,54	0,9945
2,56	0,9948
2,58	0,9951
2,60	0,9953
2,62	0,9956
2,64	0,9959
2,66	0,9961
2,68	0,9963
2,70	0,9965
2,72	0,9967
2,74	0,9969

x	$\Phi(x)$
2,76	0,9971
2,78	0,9973
2,80	0,9974
2,82	0,9976
2,84	0,9977
2,86	0,9979
2,88	0,9980
2,90	0,9981
2,92	0,9982
2,94	0,9984
2,96	0,9985
2,98	0,9986
3,00	0,9987
3,05	0,9989
3,10	0,9990
3,15	0,9992
3,20	0,9993
3,25	0,9994
3,30	0,9995
3,35	0,9996
3,40	0,9997
3,45	0,9997
3,50	0,9998
3,55	0,9998
3,60	0,9998
3,65	0,9999
3,70	0,9999
3,75	0,9999
3,80	0,9999

3. Táblázat

Közelítő képletek a Φ és φ függvényekre

$$\Phi(x) \approx 1/2 + \text{TANH}(0.8 \cdot x)/2 \approx$$

$$\approx 0.5 + 0.4 \cdot x - 0.085333 \cdot x^3 + 0.0218453 \cdot x^5 - 0.0056589 \cdot x^7 + 0.0014676 \cdot x^9 - \dots$$

vagy

$$\Phi(x) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \left(x + \frac{x^3}{3} + \frac{x^5}{3 \cdot 5} + \frac{x^7}{3 \cdot 5 \cdot 7} + \frac{x^9}{3 \cdot 5 \cdot 7 \cdot 9} + \dots \right)$$

vagy

$$\Phi(x) \approx \frac{1}{2} \cdot \left(1 + \text{sign}(x) \cdot \sqrt{1 - e^{-\frac{2x^2}{\pi}}} \right)$$

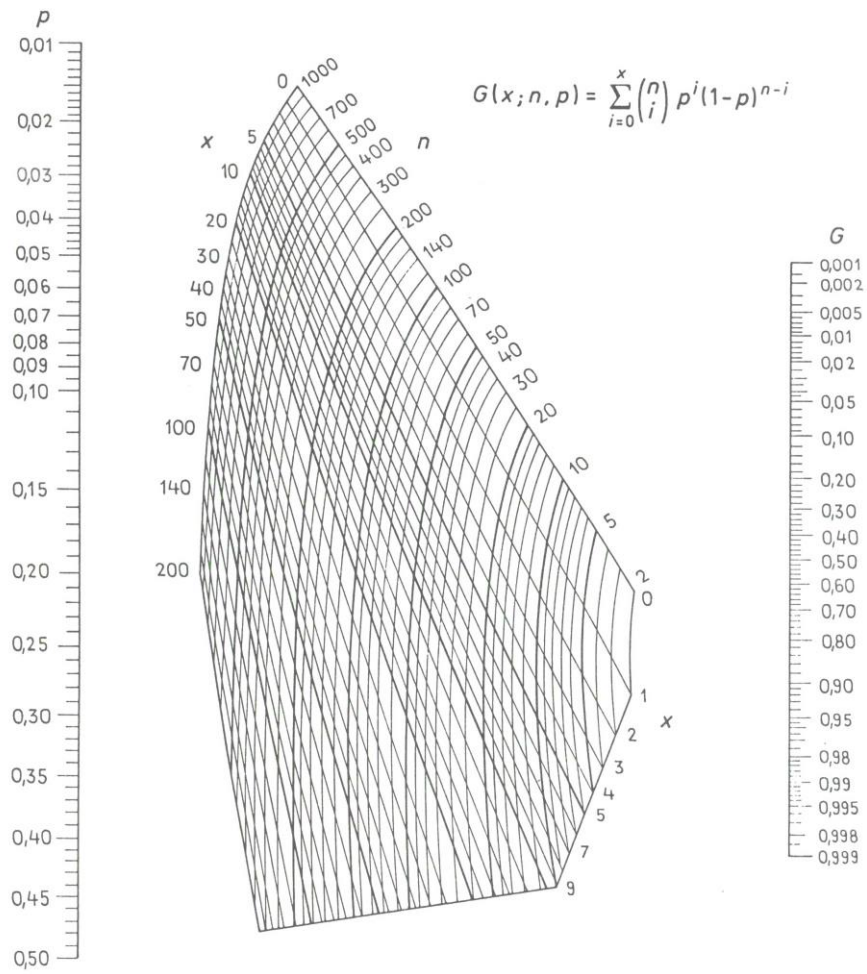
ahonnan Φ inverze

$$\Phi^{-1}(x) \approx \begin{cases} \sqrt{-1.57079632679 \cdot \ln(1 - (2x)^2)} & \text{ha } x \leq 0.5 \\ 1 - \sqrt{-1.57079632679 \cdot \ln(1 - (2x - 1)^2)} & \text{ha } 0.5 \leq x \end{cases}$$

Továbbá

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \approx 0.4 / \text{COSH}^2(0.8 \cdot x) \approx$$

$$\approx 0.4 - 0.256 \cdot x^2 + 0.1092267 \cdot x^4 - 0.03961287 \cdot x^6 + 0.0132087 \cdot x^8 - 0.0041874 \cdot x^{10} + \dots$$



6. ábra: Larson nomogram