<u>KOMBINATORIKA</u>

Permutáció: *n* egymástól különböző elem egy meghatározott sorrendben való elrendezése az *n* elem egy <u>permutációja</u>. Az összes permutáció (különböző sorrend) száma: $P^n = n!$ $0! := 1$

n*(n-1)*...*3*2*1 = n!

Ismétléses permutáció: Ha az *n* elem között $k_1, k_2, k_3, \ldots k_l$ darab egyező van - azaz *l* különböző féle elem van -, akkor az *n* elem ismétléses permutációinak száma: $P^n_{k_1,k_2,k_3..k_l} = \dfrac{n!}{k_1!k_2!...k_l!}$

Variáció: ha *n* különböző elemből *k* különbözőt kiválasztunk ($k<n$) és sorrendbe állítjuk azokat, az <u>*n* elem egy *k*-ad osztályú variációját</u> kapjuk. Az így keletkező összes lehetséges variáció száma: $V^n_k = \dfrac{n!}{(n-k)!}$

n*(n-1)*(n-2)*....*(n-k+1) = $\dfrac{n!}{(n-k)!}$

Ismétléses variáció: *n* különböző elemből *k*–t kiválasztunk (ált. $k \le n$, de lehet $k>n$) és sorrendbe állítjuk azokat. A kiválasztásnál ismétlődést is megengedünk, tehát ugyanazt az elemet többször is kiválaszthatjuk. Az így keletkező összes ismétléses variáció száma: $V^{n,i}_k = n^k$

n*n*..*n = n^k

Kombináció: Ha az *n* különböző elemből *k*-t kiválasztunk ($k<n$), de a kiválasztottakat nem rakjuk sorrendbe, az <u>*n* elem egy *k*-ad osztályú kombinációját</u> kapjuk. A keletkező összes lehetséges kombináció száma: $C^n_k = \binom{n}{k} = \dfrac{n!}{(n-k)!k!}$ $\binom{n}{0} := 1$

Ismétléses kombináció: Ha az *n* különböző elemből *k*-t kiválasztunk (ált. $k \le n$, de lehet $k>n$), de a kiválasztottakat nem rakjuk sorrendbe, és a kiválasztott elem megismétlődhet, akkor a keletkező ismétléses kombinációk száma: $C^{n,i}_k = \binom{n+k-1}{k} = \dfrac{(n+k-1)!}{(n-1)!k!}$

A képlet levezethető a kombinációk számának képletéből a következő elgondolás alapján: az ismétlés következtében a kivett elemeket mintegy "visszahelyezzük", tehát a 2., 3. … k. elem kiválasztásánál is a teljes elemhalmazból választunk. Így olyan, mintha egy n+(k-1) elemszámú halmazunk lenne, hisz minden visszahelyezéssel "növeljük" az eredeti elemszámot. Visszatevés pedig k-1 darab van, mivel az utolsó kivétel után nem helyezünk vissza semmit. Tehát egy n+k-1 elemű halmazból vesszünk ki *k* elemet „ismétlés nélkül”.

A kifejezésnél a nevező természetesen $([n+k-1]-k)!\,k! = (n-1)!\,k!$.

VALÓSZÍNŰSÉGSZÁMÍTÁS

1. ESEMÉNYALGEBRA - FOGALMAK ÉS ALAPOK

Elemi események: egy (véletlen jelenségre vonatkozó) kísérlet lehetséges kimenetei, amelyek egymást kizárják.
Eseménytér: az összes elemi esemény halmaza.
Véletlen esemény = esemény: az eseménytér egy részhalmaza (tehát elemi események halmaza). Egy esemény **bekövetkezett**, ha a kísérlet kimenetele olyan elemi esemény, ami eleme az adott eseménynek.

Az **eseményalgebra** tehát egyenértékű egy halmazalgebrával.
Az eseményalgebrában (és hasonlóan a halmazalgebrában)
- két műveletet definiálunk: összeadás és szorzás;
- minden eseménynek van ellentéte;
- egységelemnek a *H*-t, nullelemnek a -t tekinthetjük.

Az eseményeket ált. nagybetűkkel jelöljük: pl. A, B_1, B_2 esemény.

Két esemény **összege** (halmazelméletben *uniója*: $A \cup B$) az az esemény, amely akkor és csak akkor következik be, ha legalább az egyik esemény bekövetkezik (vagy A, vagy B, vagy mindkettő). Jelölése **A+B**.
Két esemény **szorzata** (halmazelméletben *metszete*: $A \cap B$) az az esemény, amely akkor és csak akkor következik be, ha mindkét esemény bekövetkezik. Jelölése **A·B = AB**.

Biztos esemény maga az eseménytér (= az összes elemi eseményt tartalmazó halmaz, H), **lehetetlen esemény** pedig az üreshalmaz ( ).
A **esemény ellentéte** (komplementere) az $\overline{A}$ esemény, amely akkor és csak akkor következik be, ha A nem következik be. (A biztos esemény ellentéte a lehetetlen esemény.)
A és B események **különbsége**: A - B = $A\overline{B}$.
A esemény **maga után vonja** B-t, ha az A eseménynek megfelelő halmaz részhalmaza B-nek, vagyis $A \subset B$.

Az A_1 ,A_2 ,...,A_n események **teljes eseményrendszert** alkotnak, ha

(1) egyikük biztosan bekövetkezik, vagyis $\sum_{i=1}^{n} A_i = H$ és

(2) egymást páronként kizárják (diszjunktak), vagyis A_iA_j = ha $i \neq j$. ➔ *egyszerre csak 1 következik be!*

Vagyis egy teljes eseményrendszer H egy *osztályozása*. (Az eseménytér összes elemi eseménye teljes eseményrendszert alkot.)

2. A VALÓSZÍNŰSÉG

Az *A* esemény relatív gyakorisága (f_A): az *A* esemény bekövetkezéseinek száma, arányítva az összes kísérlet számához.

Ha *n* kísérletből az *A* esemény k_A esetben következik be, akkor az *A* esemény relatív gyakorisága $f_A = \frac{k_A}{n}$ (ahol k_A - az *A* esemény *gyakorisága*).

Más megfogalmazásban: $f_A = \frac{\sum_i x_{Ai}}{n}$, ahol x_A az *A* esemény *indikátorváltozója*, amelynek értéke 1 - ha *A* bekövetkezik, és 0 - ha *A* nem következik be.

Az *A* esemény valószínűsége (*P(A)*) az a számérték, amely körül az esemény relatív gyakorisága ingadozik, ha egyre több kísérletet végzünk.

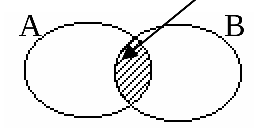
$$P(A) := \lim_{n\to\infty} f_A$$

P(A) tehát azt mutatja meg, hogy az *A* esemény az összes kísérlet *várhatóan (átlagosan) hány százalékában* következik be.

<u>A valószínűség tulajdonságai:</u>

1) Minden *A* eseményre $0 \le P(A) \le 1$ (ahol $A \subset H$);
2) $P(H) = \sum_i P(h_i) = 1$ ahol h_i - elemi események;
3) Minden $A \subset H$ eseményre $P(A) = \sum_{hi \in A} P(h_i)$
(vagyis *A* esemény vszg-e egyenlő az elemi eseményei vszg-nek összegével);
4) $P(\overline{A}) = 1 - P(A)$ (egy esemény ellentétének vszg-e = 1 - az esemény vszg-e);
5) ha $A \subset B$, akkor $P(A) \le P(B)$;
6) $P(A+B) = P(A) + P(B) - P(AB)$.

ha diszjunktak ez a tag 0



Ez általánosítható 3 ill. *n* (tetszőleges) eseményre is.

3. KLASSZIKUS VALÓSZÍNŰSÉG

<u>Az elemi események egyforma valószínűségűek</u>, vagyis:

$P(h_i) = \frac{1}{|H|} = \frac{1}{n}$ ahol $i = 1,2,3,...,n$ (egy elemi esemény vszg-e = 1/ összes elemi események száma)

Ilyen esetben egy *k* elemi eseményből álló *A* esemény vszg-e:

$$P(A) = \frac{|A|}{|H|} = \frac{k}{n} = \frac{kedvező\_esetek\_száma}{összes\_esetek\_száma}$$

<u>Vagyis a halmazok számosságával (elemszámával) felírható az események vszg-e.</u>

pl. szerencsejátékok

Feltételes vszg.

A $P(A|B)$ **feltételes valószínűség** az *A* esemény valószínűsége feltéve (azon feltétel mellett), hogy *B* esemény bekövetkezik. Akkor kapjuk meg, ha egy *A* esemény vszg-t nem a teljes *H* eseménytéren, hanem csak azok között az esetek között vizsgáljuk amikor *B* is bekövetkezett.

$P(A|B) = \frac{P(AB)}{P(B)}$ feltéve, hogy $P(B) \neq 0$.

Ebből: $P(AB) = P(A|B) \cdot P(B)$.

Általánosítva *n* esemény szorzatára:

$$P(A_1 A_2 ... A_n) = P(A_n | A_1 A_2 ... A_{n-1}) P(A_{n-1} | A_1 A_2 ... A_{n-2}) .... P(A_2 | A_1) P(A_1)$$

Ha *A* és *B* esemény **függetlenek**, akkor igaz, hogy

$P(AB) = P(A) \cdot P(B)$ és $P(A|B) = P(A)$.

Azaz független események bekövetkezésének vszg-e a két esemény szorzata.

Teljes valószínűség tétele:

Ismert *A* esemény vszg-e $B_1, B_2, ..., B_n$ feltételek mellett, ahol $B_1, B_2, ..., B_n$ teljes eseményrendszer. Ekkor az *A* esemény valószínűsége így számítható ki:

$P(A) = \sum_{i=1}^{n} P(A|B_i) P(B_i)$ ahol $P(B_i) \neq 0$.

(= a feltételes vszg-ei szorozva a feltétel vszg-ével és ezek összegezve)

Bayes-tétel:

Ha ismertek a $P(A|B_i)$ és a $P(B_i) \neq 0$ vszg-k, kiszámíthatók a $P(B_i|A)$ feltételes vszg-k - a következők alapján:

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

ahol $B_1, B_2, ..., B_n$ események teljes eseményrendszert alkotnak.

~~Valószínűségi változó~~

~~Várható érték, szórás~~

~~(külön file-ban~~)

Diszkrét és folytonos eloszlások

Diszkrét eloszlások

Diszkrét eloszlások azok, ahol a valószínűségi változó lehetséges értékei végesek, vagy megszámlálhatóan végtelenek.

Minden diszkrét eloszlást az x_k érték felvételének valószínűségével (röviden P_k), a várható értékkel [$M(\xi)$ vagy más jelöléssel $E(\xi)$] és a szórásnégyzettel (vagyis varianciával) [$D^2(\xi)=V(\xi)$] jellemzünk.

Egyenletes eloszlás

$$P(\xi = x_k) = \frac{1}{n}$$

$$M(\xi) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{ahol } k = 1,2,3,...,n$$

$$D^2(\xi) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M\right)^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$$

Binomiális eloszlás

A kísérletnek 2 lehetséges kimenetele van. Az egyedi valószínűségek nem változnak a kísérlet során (vagyis visszatevéses mintavételnek számít kombinatorikailag). Akkor használható, ha a populáció nagy vagy a mintavétel visszatevéses. Ilyenek pl. a jelölés-visszafogásos kísérletek.

Annak a valószínűsége, hogy n független kísérlet során a p valószínűségű A esemény k-szor következik be:

$$P(\xi = x_k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{ahol } k = 0,1,2,...,n$$

Az ilyen eloszlású ξ változó binomiális eloszlású.

$$M(\xi) = np$$

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

[Ha a függetlenség sérül, a következő két jelenség egyike figyelhető meg:

- Clumping = csoportosulás: az azonos típusú események csoportosulnak (ha az egyik bekövetkezett, nagyobb a vszg.-e, hogy legközelebb is az következik be)
- Repulsion = taszítás: ha az egyik kimenetel bekövetkezett, utána a másik kimenetel bekövetkezése a valószínűbb.]

A binomiális eloszlás általánosítása a **POLINOMIÁLIS ELOSZLÁS**. Egy kísérletnek legyen r számú kimenetele, az A_1, A_2, … A_r események. Annak a valószínűsége, hogy n független kísérlet során

a p_1 valószínűségű A_1 esemény k_1-szer,

a p_2 valószínűségű A_2 esemény k_2-szer,

…

a p_r valószínűségű A_r esemény k_r-szer

következett be:

$$P(\xi_1 = x_1, \xi_2 = x_2, ..., \xi_r = x_r) = \frac{n!}{k_1!k_2!...k_r!} p_1^{k_1} p_2^{k_2} ... p_r^{k_r} \quad \text{ahol } k_1+k_2+...+k_r=n$$

Hipergeometrikus eloszlás

A kísérletnek 2 lehetséges kimenetele van. Az egyedi valószínűségek változnak a kísérlet során (vagyis visszatevés nélküli mintavételnek számít kombinatorikailag). Akkor használható, ha a populáció/urna relatíve kicsi.

Ha N elég nagy és n kicsi, a visszatevés nélküli mintavétel valószínűsége jó közelítése a binomiális eloszlás megfelelő valószínűségének (vagyis a hipergeometrikus eloszlás tart a binomiálishoz).

Ha N darab elemből álló halmazban E elemből s darab van, és n elemű mintát veszünk visszatevés nélkül, akkor annak a valószínűsége, hogy a kivettek között k db E lesz, a következő:

$$P(\xi = x_k) = \frac{\binom{s}{k}\binom{N-s}{n-k}}{\binom{N}{n}}$$

Az ilyen eloszlású ξ változó hipergeometrikus eloszlású. Ha E elem relatív gyakorisága p = s/N, akkor

$$M(\xi) = np$$

$$D^2(\xi) = np(1-p)\left(1-\frac{n-1}{N-1}\right)$$

A hipergeometrikus eloszlás általánosítása a **POLIHIPERGEOMETRIKUS ELOSZLÁS**. Legyen a független, visszatevés nélküli mintavételnél r különböző kategóriánk: A_1, A_2, …, A_r, ahol egy-egy kategórián belül rendre s_1, s_2, … s_r elem van. Annak valószínűsége, hogy

az 1. kategóriába esők közül k_1-et,

az 2. kategóriába esők közül k_2-et,

…

az r. kategóriába esők közül k_r-et

tartalmaz az n elemű minta:

$$P(\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_r = x_r) = \frac{\binom{s_1}{k_1}\binom{s_2}{k_2}\cdots\binom{s_r}{k_r}}{\binom{N}{n}}$$

$$n = \sum_{i=1}^{r} k_i,\ N = \sum_{i=1}^{r} s_i$$

A polinomiális illetve polihipergeometrikus eloszlások alkalmasak a többállapotú tulajdonságok (karakterek) állapotainak jellemzésére.

Poisson eloszlás

A binomiális eloszlás, ha $n \to \infty$, $p \to 0$ (a minta nagy és az egyik vszg. kicsi), és np = λ állandó, akkor tart a Poisson eloszláshoz. A kísérletnek itt is 2 lehetséges kimenetele van: adott esemény bekövetkezik vagy nem (pl. a talajmintában találunk moha spórát vagy nem találunk).

Az előfordulások száma tér vagy időbeli egységre vonatkoztatva ismert. Az előfordulások egyenként véletlenszerűek. A bekövetkezések száma megadható, de a "nem bekövetkezések" száma nem adható meg. Pl. baktériumsejtek a mikroszkóp látóterében, villámcsapások száma óránként, stb.

$$P(\xi = x_k) = \frac{\lambda^k e^{-\lambda}}{k!}$$ ahol $k=0,1,2,\dots$.

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

A Poisson eloszlás viszonyítási alap a biológiában, mivel az előfordulások függetlenségét feltételezi!

Folytonos eloszlások

A folytonos eloszlások jellemzésére az *f(x)* sűrűségfüggvényt, az *F(x)* eloszlásfüggvényt, az *M(ξ)* várható értéket és a $D^2(\xi)$ szórásnégyzetet használjuk.

Normális eloszlás

A természetben akkor találkozunk normális eloszlással, ha sok, egymástól független, egyenként kis hatású tényező hatása összeadódik. Emiatt a közvetlenül mért, véletlenszerű ingadozásokat mutató adatok jó közelítéssel normális vagy Gauss-féle eloszlású sokaságból vett mintának tekinthetők.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\zeta-\mu)^2}{2\sigma^2}} d\zeta$$

(Az *f(x)* egy Gauss-féle haranggörbe, ld. 2.ábra.)

$$M(x) = \mu$$

$$D^2(x) = \sigma^2$$

Mint látható a normális eloszlás két paraméterrel adható meg. A normális eloszlás szokásos rövid jelölése az N(μ, σ). Például az N(5 , 2.4) egy 5 várható értékű, 2.4 szórású normális eloszlást jelöl.

A táblázatba foglalhatóság miatt és kényelmi szempontokból kitüntetett a μ = 0 várható értékű, σ = 1 szórású **standard normális eloszlás**. Ennek sűrűség-, és eloszlásfüggvénye:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{\zeta^2}{2}} d\zeta$$

Ezen speciális eloszlás szimmetrikus, azaz φ(-x) = φ(x), illetve **Φ(-x) = 1-Φ(x).**

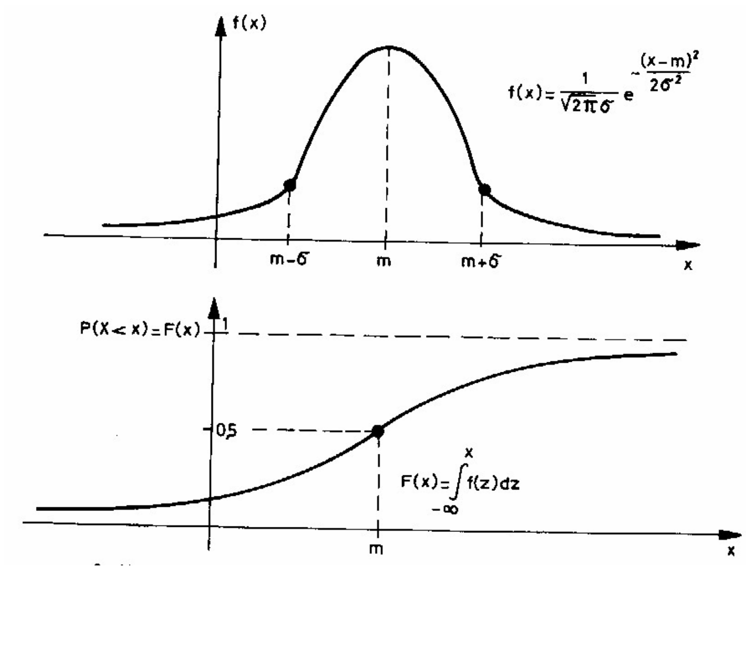
Bármely N(μ, σ) eloszlás levezethető az N(0,1) standard normális eloszlásból a következőképpen:

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Ezt nevezzük **standardizálás**nak.

Két vagy több független, normális eloszlású változó összege szintén normális eloszlású. A várható értékek és a szórásnégyzetek ez esetben összeadódnak.



1. ábra: A normális eloszlás sűrűség- (*f(x)*), és eloszlásfüggvénye (*F(x)*).

Egyenletes eloszlás az (a,b) intervallumon

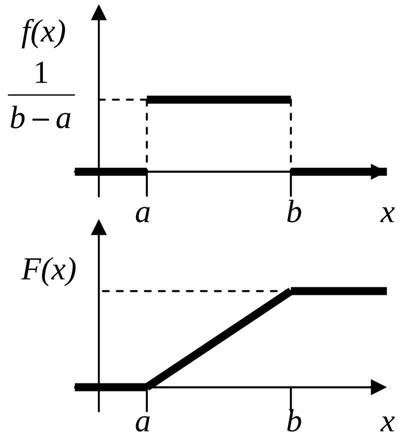
Az (a,b) intervallumon egyenletes eloszlás sűrűségfüggvényének értéke az (a,b) intervallumon állandó.

$f(x) = \dfrac{1}{b-a}$ ha $a < x < b$, különben $f(x) = 0$,

$$F(x) = \begin{cases} 0 & , \text{ha } x \le a \\ \dfrac{x - a}{b - a} & , \text{ha } a < x \le b \\ 1 & , \text{ha } x > b \end{cases}$$

$$M(\xi) = \frac{a+b}{2}$$

$$D^2(\xi) = \frac{(b-a)^2}{12}$$



2. ábra: Az egyenletes eloszlás sűrűség- (*f(x)*), és eloszlásfüggvénye (*F(x)*).

Egyenletes eloszlás esetén az egy-egy intervallumba esés valószínűsége egyenesen arányos az intervallum hosszával.

Exponenciális eloszlás

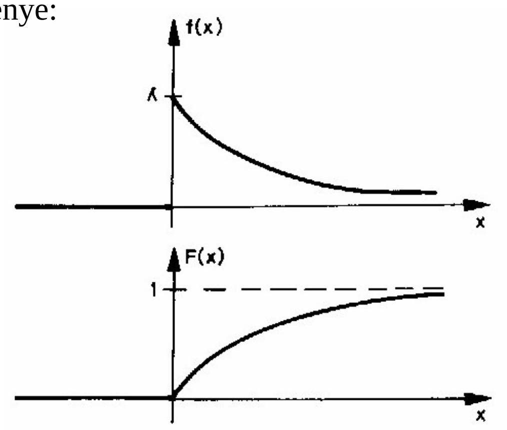
Az exponenciális eloszlás sűrűség-, és eloszlásfüggvénye:

$$f(x) = \begin{cases} 0, \text{ha } x \le 0 \\ \lambda e^{-\lambda x}, \text{ha } x > 0 \end{cases}$$

$$F(x) = \begin{cases} 0, \text{ha x} \le 0 \\ 1 - e^{-\lambda x}, \text{ha x} > 0 \end{cases}$$

A várható érték és a szórásnégyzet:

$$M(\xi) = D^2(\xi) = \frac{1}{\lambda}$$



3. ábra: Az exponenciális eloszlás sűrűség- (*f(x)*), és eloszlásfüggvénye (*F(x)*).

Az exponenciális valószínűségre igaz az "örökifjúság". Egy objektum (részecske vagy élőlény) populáció élettartamát exponenciális eloszlással modellezve igaz az, hogy ha egy objektum nem halt meg x ideig, akkor ezt követően a túlélési valószínűsége ugyan olyan, mintha a folyamat az x időpontban kezdődött volna. (A halál vszg-e minden időintervallumban azonos.)

Jó közelítéssel ilyen pár alacsonyabb rendű tengeri élőlény, pl. a korallpolipok élettartamának eloszlása.