

## 3. Valószínűségszámítás

### 3. Valószínűségszámítás 1

#### 3.1. Bevezetés 1

#### 3.2. Kombinatorika 1

##### 3.2.1. Permutációk 1

##### 3.2.2. Variációk 2

##### 3.2.3. Kombinációk 3

##### 3.3. Binomiális együtthatók tulajdonságai 3

#### 3.4. Kísérlet és esemény 4

##### 3.4.1. Eseményalgebra 4

##### 3.5. A valószínűség fogalma 5

##### 3.5.1. Kolmogorov-axiómák és következményei 5

##### 3.5.2. Klasszikus valószínűségi modell 6

##### 3.5.3. Feltételes valószínűség 6

##### 3.5.4. Nagyszámok gyenge törvénye 7

##### 3.5.5. Függetlenség 7

##### 3.5.6. Teljes valószínűség tétele 7

##### 3.5.7. Bayes-tétel 8

##### 3.6. Valószínűségi változók jellemzése 9

##### 3.6.1. Diszkrét valószínűségi változók 9

##### 3.6.2. Folytonos valószínűségi változók 9

##### 3.6.3. Valószínűségi változók várható értéke 10

##### 3.6.4. Valószínűségi változók szórása 10

##### 3.6.5. Nevezetes diszkrét eloszlások 10

#### 3.7. Nevezetes folytonos eloszlások 12

##### 3.7.1. Egyenletes eloszlás 12

##### 3.7.2. Exponenciális eloszlás 12

##### 3.7.3. Normális eloszlás 13

##### 3.8. Centrális határeloszlás tétele 15

##### 3.9. Szabadságfok fogalma 15

### 3.1. Bevezetés

A valószínűségszámítás a matematikának egy önállóan fejlődő ága és erre a fogalomrendszerre épül a matematikai statisztika (biometria) is. A modern valószínűségszámítás kidolgozása Kolmogorov orosz matematikus nevéhez fűződik, aki ebben 1930-ban fektette le a valószínűségszámítás alapjait.

A valószínűségszámítás csak egy eszköz a döntéseinkhez. Egy olyan eszköz, mely számszerűsíti egy esemény bekövetkezésének az esélyét, és ezen érték alapján dönteni lehet az eseményre bekövetkezésére vagy be nem következésére vonatkozóan.

A valószínűségszámítás a következő fogalmakra épül.

### 3.2. Kombinatorika

A kombinatorika (kapcsolóstan) az elemek csoportosításával foglalkozik. Elsődleges feladata az elemek csoportjainak előállítás, valamint a csoportok számának meghatározása. Az elemek egy elrendezését komplexiónak nevezzük.

Az elemek elrendezésének három legfontosabb fogalma a permutáció, a variáció és a kombináció témaköréhez tartozik.

#### 3.2.1. Permutációk

Ha az elrendezendő ( $n$  db) elemek mind különbözők, akkor ismétlés nélküli, ha az elemek között

azonosak is vannak, akkor ismétléses permutációról beszélhetünk. Megegyezés szerint az azonos elemek felcserélését nem tekintjük különböző sorrendnek.

Az ismétlés nélküli permutációk száma:

$$P_n = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n! \quad \text{vagy röviden } P_n = n!$$

Jelölésben  $n!$  (ejtsd:  $n$  faktoriális), ami az  $n$  elem faktoriális értékét jelöli. Megállapodás szerint  $0! = 1$ .

Ismétléses permutációk száma:

$$p_n^{k_1, k_2, k_3, \dots, k_n} = \frac{n!}{k_1! \cdot k_2! \cdot k_3! \cdot \dots \cdot k_n!}$$

ahol  $k_1, k_2, k_3, \dots, k_n$  az egymás közt megegyező elemek számát jelöli.

Példák

1. Hány hatjegyű szám állítható elő a 0, 1, 4, 5, 6, 8 számjegyekből?

*Megoldás:*

$$P_6 - P_5 = 6! - 5! = 720 - 120 = 600$$

2. Hány hatjegyű számot képezhetünk az 1, 1, 3, 3, 3, 6 számokból?

*Megoldás:*

A számok ismétléses permutáció adja a megoldást

$$P_6^{2,3} = \frac{6!}{2! \cdot 3!} = 60$$

### 3.2.2. Variációk

Ha  $n$  számú különböző elemből kiválasztunk  $k$  ( $k \leq n$ ) számú elemet úgy, hogy figyelembe vesszük ezek sorrendjét is, akkor  $n$  elem  $k$ -ad osztályú variációjáról beszélünk. Az összes variáció számát a

$$V_n^k = \frac{n!}{(n-k)!} = n(n-1) \cdot (n-2) \cdot (n-3) \cdot \dots \cdot (n-k+1)$$

kifejezés adja.

Ha az  $n$  elemből úgy választunk  $k$  elemet tartalmazó csoportokat, hogy a csoportban egy elem többször is szerepelhet és az elemek sorrendje is fontos, akkor az  $n$  elem  $k$ -ad osztályú ismétléses variációját határozzuk meg:

$$V_n^{k,i} = n^k$$

A felső indexben az  $i$  betű jelöli az ismétléses variációt.

Példák

1. Négy sebész kettesével, felváltva használja a műtőt úgy, hogy az egyikük a vezető sebész legyen. Adjuk meg a lehetséges beosztást.

Megoldás

Legyen A, B, C, D a négy sebész. A vezető mindig az első helyre kerül, amire 4 lehetőség adódik. A "beosztott" sebészt a maradó 3 fő közül választhatjuk ki. Így a lehetőségek száma

$$4 \cdot 3 = 12$$

A párok tehát:

AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC

2. A METRO aluljárójában 4 ablaknál lehet bérletet, jegyet venni. Az egyszerre odaérkező 8 fő, hányféle módon kereshet ablakot magának?

Megoldás

Ugyanannál az ablaknál való elhelyezkedés is megengedett, tehát 4 elemből (ablakból) kell 8-as csoportokat képezni a sorrend beszámításával. A csoportok számát az ismétléses variáció adja:

$$V_4^{8,i} = 4^8 = 65536$$

**3.2.3. Kombinációk**

Ha az  $n$  számú különböző elemből úgy választunk ki  $k$  ( $k \leq n$ ) számút minden lehetséges módon, hogy a kiválasztás során a csoportokon belül az elemek sorrendje nem fontos, akkor  $n$  elem  $k$ -ad osztályú kombinációjáról beszélünk. Az összes lehetséges kiválasztás száma:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1) \cdot (n-2) \cdot (n-3) \cdot \dots \cdot (n-k+1)}{k(k-1) \cdot \dots \cdot 1}$$

Az  $\binom{n}{k}$  jelölést úgy olvassuk, hogy "n alatt a k".

Ha a  $k$  elem között egy elem többször is előfordulhat, akkor  $n$  elem  $k$ -ad osztályú ismétléses kombinációjáról beszélünk. Az összes kiválasztási lehetőségek száma:

$$C_n^{k,i} = \binom{n+k-1}{k}$$

**Példák**

3. Az egyetemi menza két ételkiadó ablakához 6 hallgató érkezik. Hányféle módon választhatják ki maguk közül az első két hallgatót?

Megoldás

A kiválasztásnál a sorrend nem fontos. A csoportok számát a 6 fő 2-od osztályú kombinációja adja

$$C_6^2 = \binom{6}{2} = \frac{6!}{2!(6-2)!} = 15$$

4. Egy öttagú családnál a telefon 4-szer szólalt meg TV nézés közben. Egy személy 3-szor is odamehetett a készülékhez. A sorrendet nem figyelve, hányféle módon vehették fel a kagylót?

Megoldás

A csoportok száma 5 elem (fő) 3-ad osztályú ismétléses kombinációja

$$C_5^{3,i} = \binom{5+3-1}{3} = \binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

**3.3. Binomiális együtthatók tulajdonságai**

Az olyan kifejezéseket amelyek két tagból állnak binomiális kifejezéseknek nevezzük, pl.  $(a+b)$  vagy  $(a-b)$ . Vegyük az  $(a+b)$  binom hatványait sorba egészen a 3. hatványig ( $n = 0, 1, 2, 3$ ):

$$(a+b)^0 = 1$$

$$(a+b)^1 = a+b$$

$$(a+b)^2 = a^2 + 2ab + b^2$$

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

Ha az egyes tagok együtthatóit egymás alá írjuk, akkor az ún. Pascal háromszöget kapjuk, ahol a külső szárok mentén csak 1-es áll. A háromszög belsejében álló bármely szám a közvetlen felette lévő és attól balra álló két szám összege:

Vezessük be az  $\binom{n}{0} = 1$ ,  $\binom{n}{n} = 1$  jelöléseket és írjuk fel a Newton-féle binomiális tételt:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \dots + \binom{n}{n-1} a b^{n-1} + \binom{n}{n} b^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

ahol az  $\binom{n}{k}$  együtthatókat binomiális együtthatóknak nevezzük.

A tételt a kifejtett binomiális együtthatókkal is felírhatjuk:

$$(a+b)^n = a^n + \frac{n}{1!} a^{n-1} b + \frac{n(n-1)}{2!} a^{n-2} b^2 + \dots + x^n$$

A tételnek egy következménye az alábbi kifejezés:

$$(1+x)^n \approx 1+nx \quad (nx \text{ közel van a } 0\text{-hoz})$$

### 3.4. Kísérlet és esemény

Kísérletnek lehet tekinteni egyrészt minden olyan tevékenységet, amit valamilyen cél érdekében hajtunk végre.

A kísérlet egyes lehetséges kimeneteleit elemi eseményeknek nevezzük. Az eseményeket az ABC nyomtatott nagybetűivel jelöljük. Két eseményt azonosnak tekintünk, ha egy kísérlet minden lehetséges kimenetelét figyelembe véve vagy mindkettő bekövetkezik, vagy egyik sem. Ha két esemény A és B olyan kapcsolatban van egymással, hogy A csak akkor következhet be, ha B is bekövetkezik, akkor azt mondjuk, az A esemény maga után vonja a B eseményt. Az ilyen eseményeket a következő módon jelöljük:

$$A \subset B$$

Egy kísérlet összes elemi eseményeinek a halmaza az eseménytér ( $\Omega$ ).

Az elemi eseményekkel kapcsolatos három további fogalom:

- lehetetlen esemény (O): sohasem következhet be a kísérlet folyamán,
- biztos esemény (I): mindig bekövetkezik,
- ellentett (komplementer) esemény ( $\bar{A}$ ) csak akkor következhet be, ha az A esemény nem következik be.

#### 3.4.1. Eseményalgebra

##### 3.4.1.1. Összeadás

Az A és B események összege az a C esemény, amely akkor következik be, ha az A és B események közül legalább az egyik bekövetkezik:

$$A + B = C$$

##### 3.4.1.2. Kivonás

Az A és B események különbsége az a  $A-B$  esemény, amely akkor következik be, amikor az A esemény teljesül, de a B esemény nem:

$$A - B = F = A\bar{B}$$

### 3.4.1.3. Szorzás

A G és H események szorzatán azt az eseményt (jelölésben AG) értjük, amely csak akkor következik be, ha a G és H esemény is bekövetkezik:

$$K = GH$$

Ha a B és C eseményre igaz, hogy szorzatuk a lehetetlen eseményt adja akkor a két esemény kizárja egymást:

$$BC = O$$

Egy A eseményre vonatkozóan az alábbi műveletek végezhetők el:

Összeadás	Szorzás	Komplementer művelet
$A + A = A$	$A \cdot O = O$	$I^c = O$ $A + A^c = I$
$A + I = I$	$A \cdot A = A$	$O^c = I$ $A \cdot A^c = O$
$A + O = A$	$A \cdot I = A$	$(A^c)^c = A$

Az eseményekkel végezhető műveleteket összefoglalóan Boole–algebrának hívják. A gyakorlatban főleg a logikai áramkörökben fontosak az ún. de Morgan–azonosságok ( az események fölött a vonás a komplementer jele):

$$\overline{A + B} = \bar{A} \cdot \bar{B} \quad \text{és} \quad \overline{A \cdot B} = \bar{A} + \bar{B}$$

Ezek a kifejezések több tagra is érvényesek és kiterjeszthetők.

### 3.4.1.4. Összetett esemény

Egy A esemény összetett vagy felbontható esemény, ha legalább két, tőle különböző esemény összegeként egyértelműen előállítható.

$$K = G + H \quad K \neq G \quad \text{és} \quad K \neq H$$

Egy elemi esemény nem állítható elő ilyen alakban.

### 3.4.1.5. Teljes eseményrendszer

Az  $A_1, A_2, A_3, \dots, A_n$  események teljes eseményrendszert képeznek ha igazak rájuk az alábbi feltételek:

$$a) \quad A_1 + A_2 + A_3 + \dots + A_n = I$$

$$b) \quad A_i A_j = O \quad \text{ha } i \neq j \quad (i = 1, 2, 3, \dots, n \quad \text{és} \quad j = 1, 2, 3, \dots, n)$$

## 3.5. A valószínűség fogalma

A mindennapi életben igen gyakran használjuk ezt a fogalmat, amikor egy esemény bekövetkezési esélyét próbáljuk számszerűen meghatározni. A lehetetlen esemény valószínűsége 0, a biztos esemény valószínűsége 1, és a két szélső érték között a valószínűségi skála egyéb értékei szerepelnek. Minél nagyobb egy esemény bekövetkezésének az esélye, valószínűsége annál inkább közelíti az 1 értéket. A valószínűségi értékeket  $p$ -vel jelöljük.

A valószínűség másik ismert megadási módja a százalékos forma, amikor pl.  $p = 0.5$  helyett 50 %-os esélyt mondunk egy esemény bekövetkezésére. Ha magát az A eseményt is jelöljük a valószínűségével együtt, akkor a  $P(A)$  jelölést használjuk.

### 3.5.1. Kolmogorov–axiómák és következményei

Egy esemény valószínűségére az alábbiak érvényesek:

$$1) \quad 0 \leq P(A) \leq 1$$

Egy esemény valószínűsége csak 0 és 1 közötti szám lehet.

- 2)  $P(0) = 0$   
A lehetetlen esemény valószínűsége 0.
- 3)  $P(1) = 1$   
A biztos esemény valószínűsége 1.
- 4) Ha az A és B egymást kizáró események (vagyis  $AB = 0$ ) akkor az A és B eseményekre igaz:  
 $P(A+B) = P(A) + P(B)$

Az axiómák következményei:

a.) Ha az A esemény maga után vonja a B eseményt, akkor a valószínűségeikre teljesül, hogy:

$$P(A) \leq P(B)$$

b) Az A eseményre és ellentétjére az  $\bar{A}$ -ra igaz, hogy:

$$P(A) + P(\bar{A}) = 1$$

c) Két esemény független egymástól, ha szorzatukra igaz, hogy

$$P(A \bullet B) = P(A) \bullet P(B)$$

d) Ha az  $A_1, A_2, A_3, \dots, A_n$  események páronként kizárják egymást, akkor igaz az alábbi felbontás:

$$P(A_1 + A_2 + A_3 + \dots + A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n)$$

Ennek az additivitásnak egy fontos esete az, ha a  $A_1 + A_2 + A_3 + \dots + A_n$  események teljes eseményrendszert alkotnak, akkor:

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$$

### 3.5.2. Klasszikus valószínűségi modell

A valószínűséget az egyes események relatív gyakorisága alapján határoztuk meg, amit úgy számítottunk, hogy:

$$P(A) = \frac{k}{n}$$

vagy

$$P(A) = \frac{\text{kedvező esetek száma}}{\text{összes eset száma}}$$

Példa

1. Egy dobozban 5 piros, 3 fehér, 2 kék tableta van.

Mi a valószínűsége a kék tableta húzásának?

Megoldás

Az összes lehetőségek száma  $n = 10$ . A kedvező lehetőségek száma  $k = 2$

$$P(A) = \frac{k}{n} = \frac{2}{10} = \frac{1}{5} = 0.20$$

### 3.5.3. Feltételes valószínűség

Legyen A és B két esemény és  $P(B) \neq 0$ . Az A eseménynek a B esemény melletti feltételes valószínűsége az A esemény bekövetkezésének a valószínűségét jelenti, ha a B esemény mint

feltétele az A eseménynek bekövetkezett:

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

Következmény:

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

Ezt az egyenlőséget felhasználva, az  $A_1, A_2, A_3, \dots, A_n$  események szorzatára kapjuk, hogy:

$$P(A_1 \cdot A_2 \cdot A_3 \cdot \dots \cdot A_n) = P(A_n | A_1 A_2 A_3 \dots A_{n-1}) \cdot P(A_{n-1} | A_1 A_2 A_3 \dots A_{n-2}) \cdot \dots \cdot P(A_2 | A_1) \cdot P(A_1)$$

Példa

Mennyi annak a valószínűsége, hogy egy kétgyermekes családban mindkét gyermek fiú, ha

a) az idősebb gyermek fiú

b) legalább az egyikük fiú

(A fiú és leány születésének valószínűsége azonos.)

Megoldás

Legyen A az az esemény, hogy az idősebb gyermek fiú, B a fiatalabb gyermek fiú. Ekkor a keresett feltételes valószínűségek

$$a) P(AB|A) = \frac{P(ABA)}{P(A)} = \frac{P(AB)}{P(A)} = \frac{1}{2}$$

$$b) P(AB|A+B) = \frac{P(ABA + ABB)}{P(A+B)} = \frac{P(AB)}{P(A+B)} = \frac{1}{4}$$

### 3.5.4. Nagyszámok gyenge törvénye

A kísérlet során az A esemény bekövetkezési valószínűsége legyen

$$P(A) = p$$

és az ellentett esemény ( $\bar{A}$ ) valószínűsége:

$$P(\bar{A}) = 1 - p = q$$

Legyen  $\varepsilon > 0$  tetszőleges valós szám, ekkor a nagy számok gyenge törvénye szerint:

$$P\left(\left|\frac{k}{N} - p\right| \geq \varepsilon\right) \leq \frac{p \cdot q}{\varepsilon^2 \cdot N}$$

A törvényt Bernoulli-féle törvénynek is nevezik. A törvény azt fejezi ki, hogy a kísérletszám (N) növelésével egyre kisebb lesz annak valószínűsége, hogy valamely esemény relatív gyakorisága és valószínűsége között nagy a különbség.

### 3.5.5. Függetlenség

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

### 3.5.6. Teljes valószínűség tétele

Ha a  $B_1, B_2, B_3, \dots, B_n$  események teljes eseményrendszert alkotnak és igaz továbbá, hogy  $P(B_i) \neq 0$ , akkor tetszőleges A esemény valószínűségére igaz az alábbi kifejezés:

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

vagyis az A esemény valószínűsége a  $B_i$  események feltétele mellett meghatározható.

Példa

Az anatómia vizsgán az A csoport hallgatóinak 60% -a, a B csoport hallgatóinak 80%-a sikerrel szerepel. Az A csoport az évfolyam 15%-át teszi ki. Mi a valószínűsége annak, hogy egy véletlenül kiválasztott hallgató sikeresen vizsgázik?

Megoldás

Események:

a) Legyen A a vizsgált esemény.

b) Legyen  $C_1$  az az esemény, hogy a kiválasztott egyén A csoport belülről. Ennek kiválasztására az esély

$$P(C_1) = \frac{15}{100} = 0.15\%$$

c) Legyen  $C_2$  az az esemény, hogy a B csoportból választottunk. Erre az esély

$$P(C_2) = \frac{85}{100} = 0.85\%$$

A sikeres vizsgázás valószínűsége csoportonként

A csoport:

$$P(A|C_1) = \frac{60}{100} = 0.6\%$$

B csoport:

$$P(A|C_2) = \frac{80}{100} = 0.8\%$$

A teljes valószínűség tétele szerint

$$P(A) = P(A|C_1)P(C_1) + P(A|C_2)P(C_2) = 0.6 \cdot 0.15 + 0.8 \cdot 0.85 = 0.77$$

### 3.5.7. Bayes-tétel

Ha a  $B_1, B_2, B_3, \dots, B_n$  események teljes eseményrendszert alkotnak és igaz továbbá, hogy  $P(B_i) \neq 0$  és egy tetszőleges A eseményre  $P(A) \neq 0$ , akkor a  $B_i$  eseményekre igaz az alábbi kifejezés:

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

Tehát a  $B_i$  események valószínűsége az A esemény bekövetkezése esetén mint feltétel mellett a formula segítségével meghatározható. A kifejezésben a  $P(B_i)$  valószínűséget priori valószínűségeknek nevezzük. A Bayes-tétel fontos alkalmazási területe a szakértői rendszerek világa. Pl. egy diagnosztikus folyamat leírása ezen az úton valósulhat meg.

Példa

Atomrobbanás környezetében három zónát különböztetnek meg. Ezekben a túlélő lakosságnak 15, 40, 45 százaléka lakik. Az első zónában minden túlélő sugársérülést szenved, a második illetve a harmadik zónában 60 illetve 25 százalék ez az arány.

Mi a valószínűsége annak, hogy véletlenszerűen kiválasztva egy sugársérülést szenvedett egyént, az az első zónából való?

Megoldás

Események

$A_1$ : első zónából való

$A_2$ : második zónából való

$A_3$ : harmadik zónából való

B : az illető sugársérült

Bayes tétele szerint



$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{\sum_{i=1}^3 P(B|A_i)P(A_i)} = 0.299 \approx 0.3$$

### 3.6. Valószínűségi változók jellemzése

A biometriai vizsgálatok során megfigyelt vagy mért értékek véletlentől függő mennyiségek, amelyekhez számértékeket rendelünk. Ezeket a véletlen által befolyásolt értékeket közös néven valószínűségi változóknak (random variable) nevezzük. A változó név onnan származik, hogy az értéke megfigyelési egyedenként más és más értéket vehet fel, vagyis az érték egyedenként változik. Ezeket az értékeket bizonyos valószínűségek mellett veszik fel a változók, ezért használhatjuk a valószínűségi változó elnevezést.

A valószínűségi változóknak két formáját ismerjük: diszkrét és folytonos valószínűségi változókat.

#### 3.6.1. Diszkrét valószínűségi változók

Ha a  $\xi$  valószínűségi változó értékkészlete véges vagy megszámlálhatóan végtelen  $x_k$  számsorozat, akkor magát a  $\xi$ -t diszkrét valószínűségi változónak nevezzük. Ha az  $A_k \subset \Omega$  olyan részhalmaz, amelynek elemi eseményeihez a  $\xi$  hozzárendeli az  $x_k$  számsorozat értékeit, akkor az egyes események valószínűségeit ( $p_k$ ) a:

$$p_k = P(A_k) = P(\xi = x_k)$$

formulával lehet megadni. Az így meghatározott valószínűségeket a  $\xi$  változó eloszlásának nevezzük. A képlet azt fejezi ki, hogy a  $\xi$  valószínűségi változó az egyes  $x_k$  értékeket milyen valószínűséggel veszi fel.

Egy  $\xi$  valószínűségi változó eloszlásfüggvényét (distribution function)  $F(x)$  jelöljük és annak valószínűségét adja meg, hogy a  $\xi$  milyen valószínűséggel vesz fel egy tetszőleges  $x$  értéknél kisebb értéket. Jelölésben:

$$F(x) = P(\xi < x)$$

Megjegyzendő, hogy a diszkrét valószínűségi változó  $F(x)$  eloszlásfüggvénye lépcsős alakú függvény.

Az  $F(x)$  eloszlásfüggvény tulajdonságai az ábráról is leolvashatók:

- balról folytonos,
- monoton növekedő,
- értéke 0 és 1 közötti.

#### 3.6.2. Folytonos valószínűségi változók

A valószínűségi változók azon csoportját, amelyek értékkészlete véges vagy nem megszámlálhatóan végtelen, folytonos valószínűségi változóknak nevezzük.

Az ilyen típusú változó eloszlásfüggvényének meghatározása éppen a végtelen értékkészlete miatt nehezebb mint diszkrét változó esetében. Az egyes tartományok (szakaszok) valószínűségének megadása ugyanis közvetlenül nem lehetséges. Ezért került bevezetésre a sűrűségfüggvény ( $f(x)$ ) használata, amelynek révén minden szakasz valószínűsége megadható a szakaszhoz tartozó függvénygörbe alatti terület (integráljának) nagyságával.

Az is mondható, hogy az eloszlásfüggvény ( $F(x)$ ) a sűrűségfüggvény  $f(x)$  integrálja. Folytonos valószínűségi változók esetében mindig létezik a  $\xi$  valószínűségi változónak sűrűségfüggvénye. A sűrűségfüggvény tulajdonsága, hogy

- értéke  $\geq 0$  (hiszen a valószínűség nem lehet negatív értékű),
- a függvény görbe alatti területe = 1 (a valószínűség max. értéke csak 1 lehet).

Némely esetben a sűrűségfüggvény meghatározása nem egyszerű, mert ha ismerjük is, nem könnyű elvégezni a függvény integrálását. Ezért a biometriában leggyakrabban használt folytonos függvényekre mint pl.  $\chi^2$  eloszlás, normális eloszlás, F eloszlás, t eloszlás, stb. eloszlástáblázatokat készítettek éppen a gyakorlati munka megkönnyítése miatt. Ezekből a táblázatokból a kívánt valószínűségeket egyszerű módon ki lehet olvasni.

### 3.6.3. Valószínűségi változók várható értéke

Ha egy kísérletet sokszor megismétlünk és mindegyik kísérletet egymástól függetlenül hajtjuk végre, akkor a valószínűségi változónak az egyes kísérletek során felvett értékei egy jól meghatározott érték körül ingadoznak. Ezt az értéket várható értéknek nevezzük. Diszkrét valószínűségi változó esetén a várható érték véges  $k$  esetén:

$$M(\xi) = \sum_{k=1}^n p_k x_k$$

Folytonos eloszlású valószínűségi változó esetén az  $f(x)$  függvény  $-\infty$ -tól  $+\infty$ -ig integrálja adja a várható értéket. Ennek meghatározása az esetek többségében nem könnyű feladat.

### 3.6.4. Valószínűségi változók szórása

Egy valószínűségi változó értékeinek a várható értéke körüli elhelyezkedését, szóródását nevezzük a változó szórásának. Jelölve  $D(\xi)$ . Ennek négyzete a variancia ami a  $\xi$  változó és várható értéke különbségének a négyzete, illetve ennek várható értéke:

$$\text{Var}(\xi) = D^2(\xi) = M[(\xi - M(\xi))^2] = M(\xi) - [M(\xi)^2]$$

A szórás nyilván csak akkor van értelmezve, ha a várható érték is létezik.

Diszkrét valószínűségi változó esetén a szórásnégyzet (variancia):

$$\text{Var}(\xi) = D^2(\xi) = \sum_{k=1}^n p_k x_k^2 - \left( \sum_{k=1}^n p_k x_k \right)^2$$

Folytonos valószínűségi változó esetén a  $\text{Var}(\xi)$  kétszeri integrálással határozható meg.

### 3.6.5. Nevezetes diszkrét eloszlások

#### 3.6.5.1. Binomiális eloszlás

Végezzünk el egy kísérletet  $n$ -szer egymástól függetlenül. A kísérlet során egy  $A$  esemény bekövetkezésének valószínűsége legyen  $P(A) = p$  és az ellentett esemény valószínűsége pedig  $P(\bar{A}) = q = 1-p$ . A  $p$ -ről feltesszük, hogy konstans a kísérlet folyamán. A  $\xi$  valószínűségi változó az  $A$  esemény bekövetkezéseinek a számát jelenti. Ekkor annak valószínűsége, hogy a kísérlet során az  $A$  esemény  $k$ -szor következik be a következő alakban adható meg:

$$p_k = P(\xi = k) = \binom{n}{k} p^k \cdot q^{n-k} \quad (k = 0, 1, 2, \dots, n)$$

A  $\xi$  valószínűségi változó eloszlását binomiális eloszlásnak nevezzük, amelynek várható értéke:

$$M(\xi) = n \cdot p$$

és szórása:

$$D(\xi) = \sqrt{n \cdot p \cdot q}$$

formában határozható meg.

### Példa

1. Egy bizonyos betegség a hagyományos terápiával az esetek egynegyed részében gyógyítható. Új kezelést akarnak bevezetni, melyet előzőleg 10 betegen kipróbálnak. Ha legalább heten meggyógyulnak, akkor az új kezelést bevezetik. Ha legfeljebb hárman gyógyulnak meg, akkor az új eljárást elvetik. Ha 4, 5, vagy 6 beteg gyógyul meg, akkor az eljárást tovább vizsgálják. A kezelés hatása a régi terápiás eljárással azonos. Határozzuk meg a három esethez tartozó valószínűségeket.

#### Megoldás

Jelöljük a vizsgált eseményeket A, B, C betűkkel. Az események binomiális eloszlást követnek, így

$$P(A) = \sum_{k=7}^{10} \binom{10}{k} 0.25^k 0.75^{10-k} = 0.0035$$

$$P(B) = \sum_{k=0}^3 \binom{10}{k} 0.25^k 0.75^{10-k} = 0.7759$$

$$P(C) = 1 - (P(A) + P(B)) = 1 - (0.0035 + 0.7759) = 0.2206$$

### 3.6.5.2. Poisson–eloszlás

A

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

eloszlást a  $\xi$  valószínűségi változó Poisson–eloszlásának nevezzük, ahol  $\lambda > 0$  egy tetszőleges valós szám.

Poisson eloszlást követnek pl. a kalácsban egy adott területre eső mazsolák száma, a lehulló hópelyhek száma egy adott tartományon, baktériumok, sejtek száma egy adott térfogatban, balesetek száma egy időintervallumban, stb.

A Poisson–eloszlás és a binomiális eloszlás között szoros a kapcsolat. Ha a binomiális eloszlásban  $n$  nagy és a vizsgált esemény valószínűsége a  $p$  értéke 0–hoz közeli érték (az  $n \cdot p$  szorzat értéke  $< 5$ ), ilyenkor a  $\lambda = n \cdot p$  választással a binomiális eloszlás jól közelíthető a Poisson–eloszlással:

$$\binom{n}{k} p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

A Poisson–eloszlás várható értéke:

$$M(\xi) = \lambda$$

szórása:

$$D(\xi) = \sqrt{\lambda}$$

### Példa

Egy vizsgálat kimutatta, hogy egy adott tóban a baktériumok  $2$  baktérium/cm<sup>3</sup> sűrűséggel fordulnak elő, és Poisson-típusú eloszlást követnek. Mi a valószínűsége, hogy egy  $2$  cm<sup>3</sup> nagyságú minta

- baktériummentes
- legalább két baktériumot tartalmaz?

#### Megoldás

A mintában 4 baktérium van, így  $\lambda=4$  paraméterű Poisson-eloszlással van dolgunk.

a)  $P(k=0)=e^{-4}=0.0183$

b)  $1-(P(k=0)+P(k=1))=1-5e^{-4}=0.9080$

### 3.7. Nevezetes folytonos eloszlások

#### 3.7.1. Egyenletes eloszlás

Az egyenletes eloszlás sűrűségfüggvénye és grafikonja:

$$f(x) = \begin{cases} 0 & \text{ha } x \leq a \\ \frac{1}{b-a} & \text{ha } a < x \leq b \\ 0 & \text{ha } x > b \end{cases}$$

Eloszlásfüggvénye:

$$F(x) = P(\xi < x) = \begin{cases} 0 & \text{ha } x \leq a \\ \frac{x-a}{b-a} & \text{ha } a < x \leq b \\ 1 & \text{ha } x > b \end{cases}$$

A várható érték és szórás:

$$M(\xi) = \frac{a+b}{2} \quad \text{és} \quad D(\xi) = \frac{b-a}{\sqrt{12}}$$

#### 3.7.2. Exponenciális eloszlás

Az exponenciális eloszlás sűrűségfüggvénye:

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

ahol  $x > 0$  tetszőleges pozitív szám.

Az exponenciális eloszlásfüggvény alakja

$$F(x) = P(\xi < x) = \begin{cases} 0 & \text{ha } x \leq 0 \\ 1 - e^{-\lambda x} & \text{ha } x > 0 \end{cases}$$

A várható érték és szórás:

$$M(\xi) = \frac{1}{\lambda} \quad \text{és} \quad D(\xi) = \frac{1}{\lambda}$$

Exponenciális eloszlást követnek pl. a radioaktív bomlási folyamatok, az alkatrészek élettartamai stb.

Az exponenciális eloszlás általánosított alakja a Weibull-eloszlás, amelynek sűrűségfüggvénye ( $c > 0$  és  $\alpha > 0$  állandók):

$$f(x) = \begin{cases} c \cdot \alpha \cdot x^{\alpha-1} \cdot e^{-c \cdot x^\alpha} & \text{ha } x \geq 0 \\ 0 & \text{ha } x < 0 \end{cases}$$

Eloszlásfüggvénye:

$$F(x) = \begin{cases} 1 - e^{-c \cdot x^\alpha} & \text{ha } x \geq 0 \\ 0 & \text{ha } x < 0 \end{cases}$$

A Weibull-eloszlás egyik sajátos felhasználási területe a gyógyszerkinetikai vizsgálatok.

### 3.7.3. Normális eloszlás

A statisztikai vizsgálatok szempontjából az egyik legfontosabb eloszlás a normális eloszlás. Központi helyet foglal el a vizsgálatok között mivel számos statisztikai eljárás ezen az eloszlástípuson alapszik. Maga az elnevezés is arra utal, hogy a mért adatainktól az várjuk, hogy ilyen módon viselkedjenek, mert az a természetes, a normális viselkedése az adatoknak. Az eloszlás többféle elnevezéssel is használatos: Gauss-eloszlás, harang-görbe elnevezések szinonimái a normális jelzőnek.

Egy tetszőleges  $\xi$  valószínűségi változó normális eloszlású, ha sűrűségfüggvényére igaz az alábbi kifejezés:

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

A kifejezésben a  $\mu$  és  $\sigma$  az eloszlás két paramétere, ahol  $\mu$  tetszőleges valós szám, a  $\sigma$  tetszőleges pozitív szám. Ez a két paraméter határozza meg, hogy a végtelen sok eloszlást tartalmazó ún. normális eloszlás családnak éppen melyik tagját vizsgáljuk.

Az ilyen típusú eloszlások szimmetrikus, egycsúcsú eloszlások, amelynek szárjai a  $-\infty$  és  $+\infty$ -hez tartoznak. A függvények az X-tengelyt csak aszimptótikusan közelítik, de azt soha nem érintik. A görbe maximum helye az X-tengelyen a  $\mu$  értéknél van. A  $\sigma$  paraméter a görbe szélességét, vagyis az adatok elhelyezkedését határozza meg.

Az eloszlás várható értéke és szórása:

$$M(\xi) = \mu \quad \text{és} \quad D(\xi) = \sigma$$

A harang-görbe csúcsa az eloszlás várható értékénél a  $\mu$  értéknél található.

Bármely normális eloszlásra igaz, hogy az adatok 68 %-a a várható értéktől a  $\mu - \sigma$  és  $\mu + \sigma$  távolságon belül helyezkednek el, vagyis az adatok a várható érték körül tömörülnek. További jellegzetessége az eloszlásnak, hogy az adatok 95 %-a a  $\mu - 2\sigma$  és  $\mu + 2\sigma$  értékek közt van és az adatok 5 %-a helyezkedik el ezen távolságokon kívül. Ez a rész az ún. farok rész (tail) a szignifikancia vizsgálatokban kap igen fontos szerepet. Ebbe a részbe csak kis valószínűséggel esnek adatok, s ezt a tulajdonságot használjuk fel döntéseinkhez.

Mivel a normális eloszlások átszámolhatók az egyikből a másikba, minden eloszlás azonos alakra hozható az ún. standardizálási eljárással. Az így kapott normális eloszlást standard normális

eloszlásnak nevezzük, és igaz rá, hogy az eloszlás várható értéke a  $\mu = 0$ , szórása  $\sigma = 1$ . A standardizálási formula, amellyel bármelyik  $\xi$  normális eloszlású változót egy új  $z$  változóba standardizálhatjuk:

$$z_i = \frac{x_i - \mu}{\sigma}$$

A kifejezés azt jelenti, hogy minden mért  $x_i$  értékből levonjuk az eredeti normális eloszlás várható értékét és a különbséget osztjuk a szórással. Az így kapott  $z_i$  értékek eloszlása standard normális eloszlású lesz. Az eljárás eredményeképpen az eloszlás szimmetria-tengelye az  $Y$ -tengely lesz és a szórások egységnyi távolságban helyezkednek el az origó körül. A standard normális eloszlás sűrűségfüggvénye:

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

A standard normális eloszlás sűrűségfüggvényére a  $\varphi(x)$ , az eloszlás függvényére a  $\Phi(x)$  jelöléseket használjuk.

A függvény tulajdonságai az alábbiak szerint foglalható össze:

a) szimmetrikus függvény az  $y$ -tengelyre (az  $y$  tengely a szimmetria tengelye)

$$\varphi(x) = \varphi(-x) \quad \text{és} \quad \Phi(-x) = 1 - \Phi(x)$$

b) a függvény legmagasabb pontjának koordinátái:

$$\left(0, \frac{1}{\sqrt{2\pi}} \approx 0.4\right) \text{ értékek}$$

c) a függvény görbe alatti területe = 1, ami azt jelenti, hogy egy standard normál eloszlású valószínűségi változó értékei 1 valószínűséggel a  $(-\infty, +\infty)$  tartományból származnak

f) az a) és e) pontok értelmében az  $y$  tengelytől jobbra és balra első területek nagysága:  $\frac{1}{2}$

g) egy tetszőleges  $(\mu, \sigma)$  paraméterű normális eloszlású valószínűségi változó sűrűség és eloszlásfüggvénye kifejezhető a standard normális eloszlás hasonló függvényeivel:

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

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

h) a binomiális eloszlás tagjait jó megközelítéssel meghatározhatjuk a standard normális eloszlás segítségével, ha az  $n$  nagy és a  $p, q$  értékek nincsenek szorosan a 0 közelében, akkor:

$$\binom{n}{k} p^k \cdot q^{n-k} \approx \frac{1}{\sqrt{n \cdot p \cdot q}} \varphi\left(\frac{k - n \cdot p}{\sqrt{n \cdot p \cdot q}}\right)$$

a közelítés akkor jó, ha az  $n \cdot p > 5$  és  $n \cdot q > 5$  egyenlőtlenség teljesül.

Hasonló kapcsolat van a Poisson-eloszlás és standard normális eloszlás között is, ha a  $\lambda$  elég nagy, akkor a Poisson-eloszlás jól közelíthető a standard normális eloszlással:

$$\frac{\lambda^k}{k!} e^{-\lambda} \approx \frac{1}{\sqrt{\lambda}} \phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right)$$

**Példa.**

Tegyük fel, hogy a sorozáson megjelenő férfiak körében a systoles vérnyomásérték várható értéke 130 Hgmm és a szórása 12 Hgmm. Várhatóan a férfiaknak hány %-a esik a 140–150 Hgmm tartományba, ha a vérnyomás értékek eloszlása normális eloszlást követ?

**Megoldás.** A feladat értelmében a  $\mu = 130$  és a  $\sigma = 12$ . Transzformáljuk át az értékeket z eloszlásba, hogy a standard normális eloszlás táblázatát tudjuk használni.

$$z_1 = \frac{x - \mu}{\sigma} = \frac{140 - 130}{12} = \frac{10}{12} = 0.83$$

$$z_2 = \frac{x - \mu}{\sigma} = \frac{150 - 130}{12} = \frac{20}{12} = 1.67$$

A keresett arányt a  $z_1$  és  $z_2$  értékek közötti terület nagysága adja meg:

A terület megállapításához használjuk az I. táblázatot:

$$T = Z_{1.67} - Z_{0.83} = 0.4554 - 0.2881 = 0.1673$$

vagyis

$$P(140 \leq x \leq 150) = 0.1673$$

Tehát várhatóan a férfiaknak 16.7 %-a esik az enyhe hipertóniás kategóriába.

**3.8. Centrális határeloszlás tétele**

A statisztikában oly fontos normális eloszlást a valószínűség számítás egyik alapvető tétele a központi (centrális) határeloszlás tétele biztosítja. A tétel szerint – szabad megfogalmazásban – egymástól független sok apró hatás együttes eredményeként keletkezett értékek eloszlása normális eloszlást követ függetlenül az összetevők eloszlásától.

Különösen fontos a tétel alkalmazhatósága az élettani folyamatok esetén, hiszen itt egy-egy jelenség számos független hatás eredőjeként alakul ki.

**3.9. Szabadságfok fogalma**

A szabadságfok fogalmát Sir R.A. Fisher vezette be. Egy statisztika szabadságfokát – amelyet df-el (degrees of freedom) jelölünk a továbbiakban –, úgy definiáljuk, hogy az N mintaszámból levonjuk az adott statisztika kiszámításához szükséges, az adatokból már meghatározott paraméterek k számát.

$$df = N - k$$

A példa kedvéért az alább bemutatott statisztikák a későbbi fejezetekben részletesen tárgyalásra kerülnek.

**Példa.**

Az n számú minta adatból számított számtani átlag szabadságfoka n, mivel az átlag kiszámításához csak a minta adatokat használjuk fel, a képletben nincs olyan paraméter, amit az adatokból számolnánk ki:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

A számlálóban csak a minta adatai, a nevezőben a minta száma szerepel.