Matematika III
Mérnök hallgatók számára

Nagy Károly

2011

Nyíregyházi Főiskola
Matematika és Informatika Intézet

Tartalomjegyzék

1. Kombinatorika

1.1. Definíció. *Adott n darab egymástól különböző elem. Ezeknek egy meghatározott sorrendjét az n* ***elem egy permutációjának*** *nevezzük. Az n egymástól különböző elem összes permutációinak a számát P_n-el jelöljük.*

Ismert, hogy

$$P_n = n! = 1 \cdot ... \cdot n.$$

1.2. Definíció. *Legyen n számú elem adva, melyek között $k_1, ..., k_l$ számú egymás közt megegyező elem található. Ezeknek egy meghatározott sorrendjét az n* ***elem egy ismétléses permutációjának*** *nevezzük. Az összes ismétléses permutációk számát $P_n^{k_1,...,k_l}$-el jelöljük.*

Ismert, hogy

$$P_n^{k_1,...,k_l} = \frac{n!}{k_1! \cdot ... \cdot k_l!}.$$

Példa: 16 ember között kiosztunk 16 különböző levelet oly módon, hogy mindenki csak egyet kap. Hányféleképpen tehetjük ezt meg?

$P_{16} = 16!$

Példa: 16 ember között kiosztunk 16 levelet, melyek között 4 és 8 azonos tartalmú. Hányféleképpen tehetjük ezt meg, ha mindenki csak egyet kap?

$P_{16}^{4,8} = \frac{16!}{4! \cdot 8!}$

1.3. Definíció. *Adott n darab egymástól különböző elem. Ezek közül kiválasztunk egy k ($k \leq n$) elemből álló csoportot úgy, hogy a kiválasztott k elem sorrendjére nem vagyunk tekintettel. Egy ilyen csoportot az n* ***elem k-ad osztályú kombinációjának*** *nevezzük. Ezek számát C_n^k-val jelöljük.*

Ismert, hogy

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

1.4. Definíció. *Adott n darab egymástól különböző elem. Ezek közül kiválasztunk k darab elemet oly módon, hogy az egyes elemek többször is szerepelhetnek és az elemek sorrendjére nem vagyunk tekintettel. Az így kiválasztott k elemet az n* ***elem k-ad osztályú ismétléses kombinációjának*** *nevezzük. Ezek számát $C_n^{k,i}$-vel jelöljük.*

Ismert, hogy

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

Példa: Hányféleképpen helyezhetünk el 5 levelet 16 rekeszbe, ha a levelek között nem teszünk különbséget és egy rekeszbe a.) legfeljebb egy levelet tehetünk? $C_{16}^{5} = \binom{16}{5}$.

b.) több levelet is tehetünk? $C_{16}^{5,i} = \binom{16+5-1}{5} = \binom{20}{5}$.

1.5. Definíció. *Adott n darab egymástól különböző elem. Ezek közül kiválasztott k darab elem egy meghatározott sorrendjét az n* ***elem k-ad osztályú variációjának*** *nevezzük. Ezek számát V_n^k-val jelöljük.*

Ismert, hogy

$$V_n^k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1).$$

1.6. Definíció. *Adott n darab egymástól különböző elem. Ezek közül kiválasztunk k darab elemet oly módon, hogy a kiválasztott elemek sorrendjét is figyelembe vesszük és egy elem többször is szerepelhet. Az így kiválasztott k elemet az n* ***elem k-ad osztályú ismétléses variációjának*** *nevezzük. Ezek számát $V_n^{k,i}$-vel jelöljük.*

Ismert, hogy

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

Példa: Hányféleképpen helyezhetünk el 5 különböző levelet 16 rekeszbe, ha egy rekeszbe

a.) legfeljebb egy levelet tehetünk? $V_{16}^{5} = \frac{16!}{(16-5)!} = 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12$.

b.) több levelet is tehetünk? $V_{16}^{5,i} = 16^5$.

2. Valószínűségszámítás

2.1. Eseménytér, esemény algebra

2.1. Definíció. *Kisérlet (megfigyelés) lehetséges ,,legkisebb" kimeneteleit* ***elemi eseményeknek*** *nevezzük. Az összes elemi események halmazát* ***eseménytérnek*** *nevezzük és Ω-val jelöljük. Az $A \subset \Omega$ halmazt* ***eseménynek*** *nevezzük, amit úgy kell érteni, hogy ha az $\omega \in \Omega$ elemi esemény bekövetkezik, akkor*

- *ha $\omega \in A$ úgy az A esemény bekövetkezik.*
- *ha $\omega \notin A$ úgy az A esemény nem következik be.*

Azt az eseményt amely mindig bekövetkezik ***biztos eseménynek*** *nevezzük. A biztos eseményt beazonosítjuk az Ω-val (hiszen $\Omega \subset \Omega$). Azt az eseményt amely sohasem következik be* ***lehetetlen eseménynek*** *nevezzük. A lehetetlen eseményt beazonosítjuk az $\emptyset$-al (hiszen $\emptyset \subset \Omega$).*

Példa: Kockával dobunk.
Elemi események: $\omega_i = i$-t dobunk a kockával (i=1,...,6).
Eseménytér:
$$\Omega = \{\omega_1, ..., \omega_6\}$$
Biztos esemény: 10-nél kisebb számot dobunk $= \Omega$.
Lehetetlen esemény: 6-nál nagyobb számot dobunk $= \emptyset$.
Esemény: párosat dobunk $= \{\omega_2, \omega_4, \omega_6\}$, páratlant dobuk $= \{\omega_1, \omega_3, \omega_5\}$.

2.2. Definíció (műveletek eseményekkel). *Az A esemény* ***ellentett vagy komplementer eseményén*** *azt az eseményt értjük, amely pontosan akkor következik be, ha az A esemény nem következik be. Jele: $\overline{A}$.*

Két esemény A és B ***unióján (összegén)*** *azt az eseményt értjük, amely pontosan akkor következik be, ha az A és a B események közül legalább az egyik bekövetkezik. Jele: $A \cup B$, $A + B$.*

Két esemény A és B ***metszetén (szorzatán)*** *azt az eseményt értjük, amely pontosan akkor következik be, ha az A és a B események mindegyike bekövetkezik. Jele: $A \cap B$, $A \cdot B$.*

Két esemény A és B ***különbségén*** *azt az eseményt értjük, amely pontosan akkor következik be, amikor az A bekövetkezik és a B esemény pedig nem következik be. Jele: $A \backslash B$, $A - B$.*

2.3. Tétel (műveletek tulajdonságai).
Kommutativitás:

$$A \cup B = B \cup A,$$

$$A \cap B = B \cap A.$$

Asszociativitás:

$$A \cup (B \cup C) = (A \cup B) \cup C,$$

$$A \cap (B \cap C) = (A \cap B) \cap C.$$

Idempotencia:

$$A \cup A = A,$$

$$A \cap A = A.$$

Disztributivitás:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

de Morgan-féle azonosságok:

$$\overline{A \cup B} = \overline{A} \cap \overline{B},$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Különbség:

$$A \backslash B = A \cap \overline{B}.$$

2.4. Definíció. *Azt mondjuk, hogy az A és a B események* ***egymást kizáró (diszjunkt) események****, hogy ha egyszerre nem következhetnek be (azaz $A \cap B = \emptyset$). Azt mondjuk, hogy az A* ***esemény maga után vonja a B eseményt****, ha az A esemény minden bekövetkezésekor a B esemény is bekövetkezik. Jele: $A \Rightarrow B$.*

Példa: Kockával dobunk. Párosat illetve páratlant dobunk egymást kizáró (diszjunkt) események. 2-t illetve 3-t dobunk egymást kizáró (diszjunkt) események. 2-t dobunk maga után vonja azt, hogy párosat dobunk.

Megjegyzés: A következők ekvivalensek:

$A \Rightarrow B,$

$A \subset B,$

$\overline{B} \subset \overline{A},$

$\overline{B} \Rightarrow \overline{A}.$

2.5. Definíció (esemény algebra). *Az Ω eseménytér részhalmazainak (eseményeinek) egy $\mathcal{A}$ halmazrendszerét **esemény algebrának** nevezzük, ha $\mathcal{A}$ tartalmazza a biztos eseményt (Ω-t), és $\mathcal{A}$ zárt a komplementerképzésre és a véges unióképzésre.*

($\mathcal{A}$ zárt a komplementerképzésre: minden $A \in \mathcal{A}$ esetén $\overline{A} \in \mathcal{A}$ is teljesül. $\mathcal{A}$ zárt az unióképzésre: minden $A, B \in \mathcal{A}$ esetén $(A \cup B) \in \mathcal{A}$ is teljesül.)

Megjegyzés: Ha $\mathcal{A}$ esemény algebra, akkor

a.) $\emptyset \in \mathcal{A}$ (hiszen $\emptyset = \overline{\Omega}$).

b.) Ha $A, B \in \mathcal{A}$, úgy $(A \cap B) \in \mathcal{A}$ is teljesül.

c.) Ha $A_1, A_2, ..., A_n \in \mathcal{A}$, úgy $(A_1 \cup ... \cup A_n) \in \mathcal{A}$ és $(A_1 \cap ... \cap A_n) \in \mathcal{A}$ is teljesül.

2.6. Definíció (esemény σ-algebra). *Az $\mathcal{A}$ esemény algebrát **esemény σ-algebrának** nevezzük, ha az $\mathcal{A}$ halmazrendszer zárt a megszámlálható végtelen unióképzésre.*

($\mathcal{A}$ zárt a megszámlálható végtelen unióképzésre: Ha $A_1, A_2, ..., A_n, ... \in \mathcal{A}$, akkor $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ is teljesül. Az $\cup_{i=1}^{\infty} A_i$ eseményt úgy kell érteni, hogy az $A_1, A_2, ..., A_n, ...$ események közül legalább az egyik bekövetkezik.)

Megjegyzés: Ha $\mathcal{A}$ esemény σ-algebra, akkor $A_1, A_2, ..., A_n, ... \in \mathcal{A}$ esetén $\cap_{i=1}^{\infty} A_i \in \mathcal{A}$ is teljesül.

Példa: Legyen $\Omega = \{\omega_1, ..., \omega_n\}$ azaz Ω egy n elemű halmaz.

$2^{\Omega} = \{A : A \subseteq \Omega\}$, azaz 2^{Ω} az Ω halmaz összes részhalmazainak halmaza, tehát egy halmazrendszer. Ekkor 2^{Ω} egy σ-algebra.

Példa: Egy érmét feldobunk, ha fejet dobunk azt az elemi eseményt jelöljük F-el, ha írást dobunk azt jelöljük I-vel.

$\Omega = \{F, I\}, \quad 2^{\Omega} = \{\emptyset, \{F\}, \{I\}, \Omega\}$.

2.2. Relatív gyakoriság, valószínűség.

2.7. Definíció. *Egy kísérlet lehetséges kimenetele az A esemény. Az n kísérlet során az A esemény bekövetkezésének a számát az A **esemény gyakoriságának** nevezzük. Jele: $k_n(A)$*

($k_n(A)$ értéke 0, 1, ..., n lehet).

*Az A **esemény relatív gyakorisága** az $r_n(A) = \frac{k_n(A)}{n}$ szám.*

Megjegyzés: Ha a kísérletek számát n-t növeljük, úgy $r_n(A)$ egyre kisebb mértékben ingadozik egy szám körül, ez a szám $P(A)$, az A esemény valószínűsége. ($P(A)$-t majd később adjuk meg pontosabban.)

Felvetődik az a kérdés, hogy ezt honnét tudjuk?! Remélhetőleg a félév végéig erre a kérdésre is választ találunk.

2.8. Tétel (A relatív gyakoriság tulajdonságai).

a.) $0 \leq r_n(A) \leq 1,$

b.) $r_n(\Omega) = 1,\ r_n(\emptyset) = 0,$

c.) ha A *és* B *diszjunkt események, akkor* $r_n(A \cup B) = r_n(A) + r_n(B),$

d.) ha $A_1, A_2...$ *páronként diszjunkt események, akkor*

$$r_n\left(\bigcup_{i=1}^{\infty} A_i\right) = r_n(A_1) + r_n(A_2) + ... = \sum_{i=1}^{\infty} r_n(A_i),$$

e.) $r_n(A) = 1 - r_n(\overline{A})$ *és* $r_n(\overline{A}) = 1 - r_n(A),$

f.) ha $A \subset B$ *(azaz* $A \Rightarrow B$*) akkor* $r_n(A) \leq r_n(B).$

Megjegyzés: Ez a tétel előre vetíti azt, hogy a valószínűségtől milyen tulajdonságok teljesülését várjuk el.

2.9. Definíció (valószínűségi mező). *Az* $(\Omega, \mathcal{A}, P)$ *hármast* ***valószínűségi mezőnek*** *nevezzük. Ha* Ω *egy nem üres halmaz.* $\mathcal{A}$ *egy esemény* σ*-algebra (eseményrendszer) és* $P\colon \mathcal{A} \to [0,1]$ *függvény, amelyre*

a.) $P(\Omega) = 1.$

b.) A valószínűség σ*-additív: ha* $A_1, A_2, ...$ *páronként diszjunkt események, akkor*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Az $(\Omega, \mathcal{A}, P)$ *hármast valószínűségi mezőnek nevezzük. A* P *függvényt valószínűségnek, a* $P(A)$ *számot az* A *esemény valószínűségének nevezzük.*

2.10. Tétel (A valószínűség tulajdonságai). *Legyen* $(\Omega, \mathcal{A}, P)$ *egy valószínűségi mező. Ekkor* *a.)* $P(\emptyset) = 0$ *és* $0 \leq P(A) \leq 1$*,*

b.) *a valószínűség végesen additív: Ha* $A_1, ..., A_n$ *páronként diszjunkt* $\mathcal{A}$*-beli halmazok, akkor*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i),$$

c.) $P(A) = 1 - P(\overline{A})$ *és* $P(\overline{A}) = 1 - P(A)$*,*

d.) *ha* $A \subset B$ *(azaz* $A \Rightarrow B$*) akkor* $P(A) \leq P(B)$ *(valószínűség monotonitása),*

e.) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*,*

f.) *a valószínűség szubadditív:*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

2.11. Definíció (diszkrét valószínűségi mező, eloszlás). *Legyen* Ω *egy véges halmaz, azaz* $\Omega = \{\omega_1, \omega_2, .., \omega_n\}$ *vagy egy megszámlálható végtelen halmaz, azaz* $\Omega = \{\omega_1, \omega_2, ..., \omega_n, ...\}$*,* $\mathcal{A} = 2^{\Omega}$ *és* $P \colon \mathcal{A} \to [0, 1]$ *egy valószínűség. Az* $(\Omega, \mathcal{A}, P)$ *hármast* ***diszkrét valószínűségi mezőnek*** *nevezzük. A* $p_i = P(\{\omega_i\})$ *számokat valószínűségi eloszlásnak nevezzük.*

Példa: Addig dobunk kockával, amig 1-t vagy 6-t nem dobunk.

ω_i : az i-dik dobásra sikerül 1-t vagy 6-t dobni ($i = 1, 2, ...$). Ekkor az eseménytér megszámlálhatóan végtelen halmaz.

Megjegyzés: 1.) A valószínűség kiszámítása diszkrét valószínűségi mezőben: Legyen A egy esemény , $P(A) =?$

Mivel

$$A = \bigcup_{\omega_i \in A} \{\omega_i\}$$

így

$$P(A) = \sum_{\omega_i \in A} P(\{\omega_i\}) = \sum_{\omega_i \in A} p_i.$$

2.) A $p_1, ..., p_n, ...$ számok valószínűségi eloszlást alkotnak pontosan akkor, ha teljesülnek a következők:

a) $p_i \geq 0$ minden i-re,

b) $\sum_i p_i = 1$.

2.12. Definíció (Klasszikus valószínűségi mező). *Olyan diszkrét valószínűségi mező, ahol* Ω *véges halmaz, tehát* $\Omega = \{\omega_1, ..., \omega_n\}$ *és minden elemi esemény valószínűsége egyenlő, azaz* $p_i = P(\{\omega_i\}) = \frac{1}{n}$. *A* $p_i = \frac{1}{n}$ $(i = 1, ..., n)$ *eloszlást* ***egyenletes eloszlásnak*** *is nevezzük* Ω*-n.*

Megjegyzés:

$$P(A) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{A \text{ elemeinek száma}}{n} = \frac{\text{kedvező esetek száma}}{\text{összes esetek száma}}.$$

Példa: Kockával dobunk. ω_i: i-t dobunk a kockán $i = 1, \ldots ,6$.

$\Omega = \{\omega_1, ..., \omega_6\}$, $\mathcal{A} = 2^\Omega$.

P(2-t dobok) $= \frac{1}{6}$,

P(párosat dobunk) $= \frac{3}{6} = \frac{1}{2}$,

P(kevesebbet dobunk 5-nél) $= \frac{4}{6}$.

2.13. Definíció (Geometriai valószínűségi mező). *Legyen* $\Omega \subset \mathbf{R}^k$ *egy véges mértékű halmaz (ahol* k *= 1, 2, 3),* $\mathcal{A} = 2^\Omega$*-beli mérhető halmazok, A egy esemény*

$$P(A) = \frac{\mu(A)}{\mu(\Omega)},$$

ahol $\mu(A)$ *az* A *halmaz mértéke* $\mathbf{R}^k$*-ban.*

k *= 1 esetén* $\mu(A)$ *az* A *halmaz hossza,*

k *= 2 esetén* $\mu(A)$ *az* A *halmaz területe,*

k *= 3 esetén* $\mu(A)$ *az* A *halmaz térfogata.*

Példa: 1) Egy 10 méter hosszú és 2,5 méter magas épületen van 2 db 2mx1,5m ablak. Mekkora az esélye annak, hogy betörik az ablak, ha labdával rúgunk a ház falára? (A ház falának minden pontját egyenlő eséllyel találjuk el.)

$$P(\text{ betörik az ablak }) = \frac{\text{kedvező terület}}{\text{összes terület}} = \frac{6}{25}.$$

2.3. Feltételes valószínűség

Probléma felvetés:

Egy két gyermekes családban tudjuk, hogy az első gyermek fiú. Mi annak a valószínűsége, hogy a második is fiú? (1/2)

Egy kockát feldobva tudjuk, hogy párosat dobunk. Mi a valószínűsége annak, hogy

a) 2-t dobunk? (1/3)
b) 3-t dobunk? (0)
Ezeket így fogjuk írni:
$P($ 2-t dobunk | párosat dobunk $) = \frac{1}{3}$,
$P($ 3-t dobunk | párosat dobunk $) = 0$.

2.14. Definíció (feltételes valószínűség). *Legyen* $(\Omega, \mathcal{A}, P)$ *egy valószínűségi mező,* A, B *események (azaz* $A, B \in \mathcal{A}$*) és* $P(B) > 0$*. Ekkor*

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

az A ***esemény*** B ***feltétel melletti valószínűsége****.*

Példa: A motivációs feladataink ismét.

$$P(\text{2. gyermek fiú} \mid \text{1. gyerek fiú}) = \frac{P(\text{1. gyerek fiú és a 2. gyerek fiú})}{P(\text{ 1. gyerek fiú })} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

$$P(\text{2-t dobunk}|\text{párosat dobunk}) = \frac{P(\text{2-t dobunk})}{P(\text{párosat dobunk})} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}.$$

$$P(\text{3-t dobunk}|\text{párosat dobunk}) = \frac{P(\text{lehetetlen esemény})}{P(\text{párosat dobunk})} = \frac{0}{\frac{3}{6}} = 0.$$

Példa: Egy 2 gyerekes családban tudjuk, hogy az egyik gyerek lány. Mi a valószínűsége annak, hogy van legalább egy fiú?

$$\begin{aligned} P(\text{ van legalább egy fiú} \mid \text{van egy lány gyerek }) &= \frac{P(\text{ egy fiú és egy lány })}{P(\text{ van egy lány })} \\ &= \frac{\frac{2}{4}}{\frac{3}{4}} = \frac{2}{3}. \end{aligned}$$

2.15. Tétel (A feltételes valószínűség tulajdonságai). *Legyen* $(\Omega, \mathcal{A}, P)$ *egy valószínűségi mező,* B *egy esemény és* $P(B) > 0$

a) $0 \leq P(A|B) \leq 1$,

b) $P(B|B) = 1$,

c) végesen additiv: legyenek $A_1, ..., A_n$ páronként diszjunkt események, ekkor

$$P\left(\bigcup_{i=1}^{n} A_i|B\right) = \sum_{i=1}^{n} P(A_i|B),$$

d) σ-additiv: legyenek $A_1, ..., A_n, ...$ páronként diszjunkt események, ekkor

$$P\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B).$$

2.16. Tétel (Láncszabály). *$A_1, ..., A_n$ tetszőleges események és $P(A_1 \cap .... \cap A_{n-1}) > 0$, ekkor*

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

Példa: 32 lapos magyar kártyából 3 lapot kihúzunk. Mi annak a valószínűsége, hogy az első és a harmadik piros és második nem piros?

A_i: i-edik húzás piros ($i = 1, 2, 3$). A kérdést így írjuk: $P(A_1 \cap \overline{A_2} \cap A_3) =?$

$$P(A_1 \cap \overline{A_2} \cap A_3) = P(A_1)P(\overline{A_2}|A_1)P(A_3|A_1 \cap \overline{A_2}) = \frac{1}{4} \cdot \frac{24}{31} \cdot \frac{7}{30}.$$

2.17. Definíció. *Legyenek $B_1, ..., B_n$ események. A $B_1, ..., B_n$ események **teljes eseményrendszert** alkotnak, ha páronként diszjunkt események (azaz $B_i \cap B_j = \emptyset$ ha $i \neq j$) és uniójuk kiadja a biztos eseményt, Ω-t (azaz $\bigcup_{i=1}^{n} B_i = \Omega$).*

2.18. Tétel (Teljes valószínűség tétele). *Legyen $B_1, ..., B_n$ teljes eseményrendszer, A egy esemény és $P(B_1) > 0, ..., P(B_n) > 0$. Ekkor*

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

Példa: Egy céllövöldében három rekeszben vannak puskák. Az első rekeszben három puska van, ezekkel 0,5 a találat valószínűsége. A második rekeszben egy puska található, ezzel 0,7 valószínűségű a találat. A harmadik rekesz két puskájával 0,8 valószínűséggel találunk célba. Mennyi a találat valószínűsége, ha valaki találomra választ ki egy puskát?

$P($ találat $) =?$

A : találunk a puskával,

B_1 : az első rekeszből választunk,
B_2 : a második rekeszből választunk,
B_3 : a harmadik rekeszből választunk.

$$\begin{aligned}P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)\\ &= 0,5 \cdot \frac{3}{6} + 0,7 \cdot \frac{1}{6} + 0,8 \cdot \frac{2}{6} = \frac{3,8}{6} = 0,633.\end{aligned}$$

2.19. Tétel (Bayes tétel). *Legyen $B_1, ..., B_n$ egy teljes eseményrendszer, melyre $P(B_1) > 0, ..., P(B_n) > 0, P(A) > 0$. Ekkor*

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

minden j-re.

Példa: A férfiak 5%-a, nők 0,25%-a színvak. 20 nőből és 5 férfiból álló csoportból kiválasztunk egy embert. Megállapítjuk, hogy színvak. Mi annak a valószínűsége, hogy nőt választottunk? $P($ nő $|$ színvak $) =?$

Legyen B_1 : nőt választunk ki,
B_2 : férfit választunk ki,
A : színvakot választunk ki.

$$\begin{aligned}P(\text{ nő } | \text{ színvak }) = P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)}\\ &= \frac{0,0025 \cdot \frac{20}{25}}{0,0025 \cdot \frac{20}{25} + 0,05 \cdot \frac{5}{25}} = \frac{1}{6}.\end{aligned}$$

$$P(\text{ férfi } | \text{ színvak }) = \frac{5}{6}.$$

2.20. Definíció. *Legyen A és B két esemény.* ***A-t és B-t független eseményeknek*** *nevezzük, ha*

$$P(A \cap B) = P(A)P(B).$$

Megjegyzés: Legyenek A, B pozitív valószínűségű események. A következők ekvivalensek:

a) A és B független események,

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

c) $P(B|A) = P(B)$.

Példa: Két érmét feldobunk. Mi annak a valószínűsége, hogy 2 fejet dobunk?

A_i: az i-edik érmével fejet dobunk $i = 1, 2$.

$$P(2 \text{ darab fejet dobunk}) = P(A_1 \cap A_2) = P(A_1)P(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

$$P(2 \text{ darab írást dobunk}) = P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1})P(\overline{A_2}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

$$\begin{aligned} &P(1 \text{ írást és } 1 \text{ fejet dobunk}) = P((A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)) \\ &\quad = P(A_1 \cap \overline{A_2}) + P(\overline{A_1} \cap A_2) = P(A_1)P(\overline{A_2}) + P(\overline{A_1})P(A_2) \\ &\quad = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{2}{4}. \end{aligned}$$

2.21. Definíció. *Az $A_1, A_2, ..., A_n$ eseményeket* ***páronként függetlennek*** *nevezzük, ha bármely kettőt kiválasztva független eseményeket kapunk.*

Az $A_1, A_2, ..., A_n$ eseményeket ***teljesen függetlennek*** *nevezzük, ha bármely $i_1, ..., i_k$ indexre*

$$P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})...P(A_{i_k})$$

egyenlőség teljesül.

Megjegyzés: Megadható 3 esemény oly módon, hogy páronként függetlenek, de nem teljesen függetlenek. Azaz $P(A_1 \cap A_2) = P(A_1)P(A_2)$, $P(A_1 \cap A_3) = P(A_1)P(A_3)$, $P(A_2 \cap A_3) = P(A_2)P(A_3)$ teljesül, de $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$.

Példa: Most már meg tudjuk adni a 2.2 Példa eseményeinek a valószínűségét adni. Legyen A_j az az esemény, hogy az j-dik dobásra 1-t vagy 6-t dobunk (először). Ekkor

$$P(\omega_i) = P(\overline{A_1} \cap ... \cap \overline{A_{i-1}} \cap A_i) = P(\overline{A_1})...P(\overline{A_{i-1}})P(A_i) = \frac{4}{6}...\frac{4}{6}\frac{2}{6} = \left(\frac{2}{3}\right)^{i-1}\frac{1}{3}$$

2.4. Valószínűségi változók, eloszlásfüggvény

Motiváció: Valószínűségi változó olyan függvény, amely értékeit véletlenszerűen veszi fel.

Példa:

1. Egy kockát feldobunk. Legyen a valószínűségi változó a dobott szám.
2. 4 érmét feldobunk. A dobott fejek száma.
3. Addig dobunk kockával, amíg 6-t nem dobunk. A dobások száma.
4. Választunk egy számot [0, 200] intervallumban.

2.22. Definíció. *A $\xi : \Omega \to \mathbf{R}$ függvényt **valószínűségi változónak** nevezzük, ha minden x valós szám esetén $\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{A}$, azaz a $\{\omega \in \Omega : \xi(\omega) < x\}$ halmaz esemény (tehát meg tudjuk mondani a valószínűségét).*

Jelölés: $\{\xi < x\} = \{\omega \in \Omega : \xi(\omega) < x\}$

*A ξ valószínűségi változó **eloszlásfüggvénye** F_ξ, ahol*

$$F_\xi(x) := P(\xi < x).$$

Példa: A ξ valószínűségi változó a -1, 0, 1, 3 értékeket veszi fel, 1/2, 1/4, 1/8, 1/8 valószínűségekkel. Adjuk meg az F_ξ eloszlásfüggvényt!

Ha $x \leq -1$, akkor

$$F_\xi(x) = P(\xi < x) = P(\emptyset) = 0.$$

Ha $-1 < x \leq 0$, akkor

$$F_\xi(x) = P(\xi < x) = P(\xi = -1) = \frac{1}{2}.$$

Ha $0 < x \leq 1$, akkor

$$F_\xi(x) = P(\xi < x) = P(\xi = -1) + P(\xi = 0) = \frac{1}{2} + \frac{1}{4}.$$

Ha $1 < x \leq 3$, akkor

$$F_\xi(x) = P(\xi < x) = P(\xi = -1) + P(\xi = 0) + P(\xi = 1) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8}.$$

Ha $3 < x$, akkor

$$F_\xi(x) = P(\xi < x) = P(\Omega) = 1.$$

Összefoglalva:

$$F_\xi(x) = \begin{cases} 0, & \text{ha } x \leq -1, \\ \frac{1}{2}, & \text{ha } -1 < x \leq 0, \\ \frac{3}{4}, & \text{ha } 0 < x \leq 1, \\ \frac{7}{8}, & \text{ha } 1 < x \leq 3, \\ 1, & \text{ha } 3 < x. \end{cases}$$

2.23. Tétel (Az eloszlásfüggvény tulajdonságai). *Egy* $F\colon \mathbf{R} \to [0,1]$ *függvény valamely valószínűségi változó eloszlásfüggvénye pontosan akkor, ha teljesülnek a következők:*

a) F *monoton növekvő,*

b) $\lim_{x\to-\infty} F(x) = 0; \quad \lim_{x\to+\infty} F(x) = 1,$

c) F *balról folytonos minden pontban.*

Példa:

$$F(x) = \begin{cases} 0, & \text{ha } x < 0, \\ x/5, & \text{ha } 0 \leq x \leq 5, \\ 1, & \text{ha } 5 < x. \end{cases}$$

Rajzoljuk fel F képét, s olvassuk le a tulajdonságok teljesülését! Valóban eloszlásfüggvényt adtunk meg.

Megjegyzés: Mire jó az eloszlásfüggvény?

$$P(\xi < x) = F_\xi(x),$$

$$P(\xi \geq x) = 1 - P(\xi < x) = 1 - F_\xi(x),$$

$$P(x \leq \xi < y) = P(\xi < y) - P(\xi < x) = F_\xi(y) - F_\xi(x).$$

Tehát valószínűségeket lehet az eloszlásfüggvénnyel meghatározni.

Példa: Vegyük az előbb vizsgált eloszlásfüggvényt, azaz

$$F_\xi(x) = \begin{cases} 0, & \text{ha } x < 0, \\ x/5, & \text{ha } 0 \leq x \leq 5, \\ 1, & \text{ha } 5 < x. \end{cases}$$

Ekkor

$$P(\xi < 3) = F_\xi(3) = \frac{3}{5},$$
$$P(\xi \geq 2) = 1 - F_\xi(2) = 1 - \frac{2}{5} = \frac{3}{5},$$
$$P(2 \leq \xi < 3) = F_\xi(3) - F_\xi(2) = \frac{3}{5} - \frac{2}{5} = \frac{1}{5}.$$

2.24. Definíció. *Legyen $(\Omega, \mathcal{A}, P)$ egy valószínűségi mező a $\xi : \Omega \to \mathbf{R}$ valószínűségi változót* ***diszkrét valószínűségi változónak*** *nevezzük, ha ξ értékkészlete megszámlálható (véges vagy végtelen). Legyenek ξ lehetséges értékei $x_1, ..., x_n, ...$, ekkor a $p_i = P(\xi = x_i)$ számokat (i=1,2,...)* ***a*** *ξ* ***eloszlásának*** *nevezzük.*

Megjegyzés:

1. A p_i számok ($i = 1, 2, ...$) valamely diszkrét valószínűségi változó eloszlását alkotják pontosan akkor, ha teljesülnek a következők:
   a) $p_i \geq 0$ minden i-re,
   b) $\sum_i p_i = 1$.
2. Diszkrét valószínűségi változó eloszlásfüggvénye mindig olyan lépcsős függvény, amely a ξ lehetséges értékeinél x_i-nél ugrik és az ugrások nagysága a $P(\xi = x_i) = p_i$ szám.

Példa: Diszkrét valószínűségi változó: a fejezet elején szereplő 1-es, 2-es, 3-as példa. A 4. példa nem diszkrét valószínűségi változót ad meg.

2.25. Definíció. *A ξ valószínűségi változót* ***folytonos valószínűségi változónak*** *nevezzük, ha van olyan $f_\xi : \mathbf{R} \to \mathbf{R}$ függvény, hogy bármely $a < b$-re*

$$F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)dx$$

Ekkor a f_ξ függvényt a folytonos valószínűségi változó ***sűrűségfüggvényének*** *nevezzük.*

Megjegyzés:

1. A korábban tanultak szerint

$$P(a \leq \xi < b) = F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)dx$$

2. $n \to \infty$ határértéket véve

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

Tehát a valószínűségeket határozott integrálokkal kapjuk, azaz a valószínűségeket görbe alatti területekkel számoljuk ki.

3. Az előző pontokból következik, hogy $f_\xi(x) = F'_\xi(x)$ teljesül, azaz f_ξ-t szakaszonkénti deriválással kapjuk az F_ξ-ből.

2.26. Tétel (A sűrűségfüggvény tulajdonságai). *Egy $f\colon \mathbf{R} \to \mathbf{R}$ függvény valamely folytonos valószínűségi változó sűrűségfüggvénye pontosan akkor, ha teljesülnek a következők:*

a) $f(x) \geq 0$ *minden* x*-re,*

b) $\int_{-\infty}^{\infty} f(x)dx = 1.$

Példa: Tekintsük az előzőekben tárgyalt eloszlásfüggvényünket! Azaz

$$F_\xi(x) = \begin{cases} 0, & \text{ha } x < 0, \\ x/5, & \text{ha } 0 \leq x \leq 5, \\ 1, & \text{ha } 5 < x. \end{cases}$$

Mi lesz ξ sűrűségfüggvénye?

$$f_\xi(x) = F'_\xi(x) = \begin{cases} 0', & \text{ha } x < 0, \\ (x/5)', & \text{ha } 0 \leq x \leq 5, \\ (1)', & \text{ha } 5 < x. \end{cases} = \begin{cases} 1/5, & \text{ha } 0 \leq x \leq 5, \\ 0, & \text{máskor.} \end{cases}$$

Példa: Az $f(x) = \begin{cases} \sin x, & \text{ha } 0 < x < \pi/2, \\ 0, & \text{máskor} \end{cases}$

függvény sűrűségfüggvény-e?

Használjuk a tételünket!

$\sin x > 0$, ha $0 < x < \pi/2$, így az a) tulajdonság teljesül. A b) tulajdonsághoz

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\pi/2} \sin x dx = [-\cos x]_0^{\pi/2} = -\cos \pi/2 - (-\cos 0) = 0 + 1 = 1.$$

Tehát a b) tulajdonság is teljesül, azaz f valamely folytonos valószínűségi változó sűrűségfüggvénye.

2.5. Várható érték, szórás

Motiváció: 4 érmét feldobunk, ξ a dobott fejek száma. ξ átlagosan milyen értéket vesz fel?

Átlagosan 2 fejet dobunk. Ezt így fogjuk majd jelölni: $E\xi = 2$.

2.27. Definíció. *Legyen ξ egy diszkrét valószínűségi változó, melynek értékei $x_1, x_2, ..., x_n, ...$ és eloszlása $p_1, p_2, ..., p_n, ...$. A ξ* ***diszkrét valószínűségi változó várható értéke*** *az $E\xi$-vel jelölt szám, melyre*

$$E\xi := \sum_i x_i p_i,$$

ha a $\sum\limits_i |x_i| p_i < +\infty$. (Ha $\sum\limits_i |x_i| p_i = +\infty$, akkor azt mondjuk, hogy a várható érték nem létezik)

Példa: Tekintsük a motivációs feladatunkat! 4 érmét feldobunk, ξ a dobott fejek száma.

ξ értékei $(x_i\ )$: 0, 1, 2, 3, 4

ξ eloszlása (p_i): $\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}$.

$$E\xi = \sum_i x_i p_i = 0 \cdot \frac{1}{16} + 1 \cdot \frac{4}{16} + 2 \cdot \frac{6}{16} + 3 \cdot \frac{4}{16} + 4 \cdot \frac{1}{16} = \frac{0 + 4 + 12 + 12 + 4}{16} = 2.$$

2.28. Definíció. *Legyen ξ egy folytonos valószínűségi változó, melynek sűrűségfüggvénye f_ξ. A ξ* ***folytonos valószínűségi változó várható értéke*** *az $E\xi$-vel jelölt szám, melyre*

$$E\xi := \int_{-\infty}^{\infty} x f_\xi(x) dx,$$

ha $\int_{-\infty}^{\infty} |x| f_\xi(x) dx < \infty$. (Ha $\int_{-\infty}^{\infty} |x| f_\xi(x) dx = +\infty$, akkor azt mondjuk, hogy a várható érték nem létezik.)

Példa: A (0, 5) intervallumból választunk egy számot, legyen ez a szám ξ. Sejtésünk az, hogy átlagosan a 2,5-t választjuk ki, tehát $E\xi = 2,5$. ξ eloszlásfüggvénye a már jól ismert függvény lesz. Mivel minden számot azonos eséllyel választunk ki, ezért geometriai valószínűséggel számolunk.

$$F_\xi(x) = P(\xi < x) = \begin{cases} P(\emptyset) = 0, & \text{ha } x \leq 0, \\ \frac{\text{kedvező szakasz hossza}}{\text{összes szakasz hossza}} = \frac{x}{5}, & \text{ha } 0 < x < 5, \\ P(\Omega) = 1, & \text{ha } 5 \leq x. \end{cases}$$

f_ξ-t már kiszámoltuk korábban.

$$E\xi = \int_{-\infty}^{\infty} xf(x)dx = \int_0^5 x\frac{1}{5}dx = \int_0^5 \frac{1}{5}xdx = \left[\frac{1}{5}\frac{x^2}{2}\right]_0^5 = \frac{5^2}{10} - 0 = 2,5.$$

2.29. Tétel (A várható érték tulajdonságai). *Legyenek ξ, η valószínűségi változók, $E\xi$, $E\eta$ létezik. Ekkor*

a) $E(\xi+\eta) = E\xi + E\eta$, *(additivitás),*

b) $E(c\xi) = cE\xi$ *(homogenitás), ahol c egy valós szám,*

c) $E(a\xi + b\eta) = aE\xi + bE\eta$ *(linearitás), ahol a, b valós számok,*

d) ha $\xi \leq \eta$, *akkor* $E\xi \leq E\eta$ *(monotonitás),*

e) ha $\xi \geq 0$, *akkor* $E\xi \geq 0$ *(pozitivitás),*

f) ha ξ és η függetlenek, akkor $E(\xi\eta) = E\xi \cdot E\eta$.

Megjegyzés: $E(\xi\eta) = E\xi \cdot E\eta$ általában nem igaz.

2.30. Definíció. *Legyenek ξ, η valószínűségi változók, a ξ, η* ***valószínűségi változókat függetlennek*** *nevezzük, ha*

$$P(\xi < x,\ \eta < y) = P(\xi < x)P(\eta < y)$$

teljesül minden x, y-ra.

Megjegyzés:

1. A és B független, ha $P(A \cap B) = P(A) \cdot P(B)$ (ismert korábbról).
2. Ha $\zeta := \begin{pmatrix} \xi \\ \eta \end{pmatrix}$ egy új kétdimenziós valószínűségi változót jelöl, akkor $F_\zeta(x,y) := P(\xi < x, \eta < y)$ definícióval, a függetlenségből $F_\zeta(x,y) = F_\xi(x)F_\eta(y)$ adódik (feltéve, hogy ξ és η függetlenek).

2.31. Tétel (Függetlenség ekvivalens alakja diszkrét val. változókra). *Legyenek ξ, η diszkrét valószínűségi változók, ξ értékei $x_1, ..., x_n, ...$ és η értékei $y_1, ..., y_m, ...$.*

A ξ, η függetlenek pontosan akkor, ha

$$P(\xi = x_i,\ \eta = y_j) = P(\xi = x_i)P(\eta = y_j)$$

minden i, j-re

2.32. Definíció. *A ξ valószínűségi változó várható értéke $E\xi$ létezik,* ***a*** *ξ* ***szórásnégyzete*** *$D^2\xi$ a következő alakban megadott szám*

$$D^2\xi := E(\xi - E\xi)^2,$$

ha létezik.

Szórása: $D\xi = \sqrt{D^2\xi}$.

Megjegyzés:

1. Ha ξ egy diszkrét valószínűségi változó, melynek értékei $x_1, x_2, ..., x_n, ...$ és eloszlása $p_1, p_2, ..., p_n, ...$, úgy $Eg(\xi)$-t az
$$Eg(\xi) = \sum_i g(x_i)p_i$$
alakban kell számolni, ha a várható érték létezik.

2. Ha ξ egy folytonos valószínűségi változó f_ξ sűrűségfüggvénnyel, akkor $Eg(\xi)$-t az
$$Eg(\xi) = \int_{-\infty}^{\infty} g(x)f_\xi(x)dx$$
alakban kell számolni, ha a várható érték létezik.

Példa: Tekintsük korábbi feladatunkat! 4 érmét feldobunk, ξ a dobott fejek száma.

ξ értékei $(x_i\ )$: 0, 1, 2, 3, 4

ξ eloszlása (p_i):$\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}$.

Korábban már kiszámoltuk, hogy $E\xi = 2$. Megjegyzésünk szerint

$$\begin{aligned} D^2\xi &= \sum_i (x_i - E\xi)^2 p_i = \sum_i (x_i - 2)^2 p_i \\ &= (0-2)^2 \cdot \frac{1}{16} + (1-2)^2 \cdot \frac{4}{16} + (2-2)^2 \cdot \frac{6}{16} + (3-2)^2 \cdot \frac{4}{16} + (4-2)^2 \cdot \frac{1}{16} \\ &= \frac{4+4+0+4+4}{16} = 1. \end{aligned}$$

2.33. Tétel (A szórásnégyzet tulajdonságai). *ξ val. változó és $E\xi$, $D^2\xi$ létezik. Ekkor*

a) $D^2\xi = E(\xi^2) - (E\xi)^2$,

b) $D^2(\xi + b) = D^2\xi$ *ahol* b *egy adott valós szám,*

c) $D^2(a\xi) = a^2 D^2\xi$ *minden adott* a *valós szám esetén,*

d) Ha $\xi_1, ..., \xi_n$ *független valószínűségi változók, akkor*

$$D^2(\xi_1 + ... + \xi_n) = D^2\xi_1 + ... + D^2\xi_n.$$

Példa: A szórás kiszámolása ismét az előző példában, a tételünk a) pontját és a korábbi megjegyzésünket használva.

$$\begin{aligned} E(\xi^2) &= \sum_i x_i^2 p_i = 0^2 \cdot \frac{1}{16} + 1^2 \cdot \frac{4}{16} + 2^2 \cdot \frac{6}{16} + 3^2 \cdot \frac{4}{16} + 4^2 \cdot \frac{1}{16} \\ &= \frac{0 + 4 + 24 + 36 + 16}{16} = \frac{80}{16} = 5. \end{aligned}$$

$$D^2\xi = E(\xi^2) - (E\xi)^2 = 5 - 2^2 = 1.$$

Példa: A (0, 5) intervallumból választunk egy számot, legyen ez a szám ξ. Korábban már láttuk, hogy $E\xi = 5/2$. Az f_ξ sűrűségfüggvényt is kiszámoltuk korábban. $f_\xi(x) = \begin{cases} 1/5, & \text{ha } 0 \le x \le 5, \\ 0, & \text{máskor.} \end{cases}$

Megjegyzésünk 2. pontja szerint

$$E(\xi^2) = \int_{-\infty}^{\infty} x^2 f_\xi(x)dx = \int_0^5 x^2 \frac{1}{5} dx = \int_0^5 \frac{1}{5} x^2 dx = \left[\frac{1}{5}\frac{x^3}{3}\right]_0^5 = \frac{5^3}{15} - 0 = \frac{25}{3}.$$

$$D^2\xi = E(\xi^2) - (E\xi)^2 = \frac{25}{3} - \left(\frac{5}{2}\right)^2 = \frac{100 - 75}{12} = \frac{25}{12}.$$

2.6. Nevezetes diszkrét eloszlások

Diszkrét egyenletes eloszlás

Legyenek $x_1, ..., x_n$ számok. ξ értékei $x_1, ..., x_n$. A val. változó minden értéket azonos valószínűséggel vesz fel. Azaz

$$P(\xi = x_i) = \frac{1}{n} \qquad \text{minden } i\text{-re.}$$

Ekkor ξ-t **diszkrét egyenletes eloszlású** val. változónak nevezzük. (Lásd még a 2.12 Definíciót!)

Várható érték:
$$E\xi = \sum_i x_i p_i = \frac{\sum_i x_i}{n} = \overline{x}$$
az $x_1, ..., x_n$ számok számtani közepe.

Szórásnégyzet:
$$D^2\xi = E(\xi^2) - (E\xi)^2 = \sum_i x_i^2 \frac{1}{n} - (\overline{x})^2 = \sum_i \frac{x_i^2}{n} - (\overline{x})^2.$$

Példa: Kockával dobunk, legyen ξ a dobott szám.
$$P(\xi = k) = 1/6; \qquad k = 1, ..., 6$$
$$E\xi = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{1 + ... + 6}{6} = \frac{7}{2}.$$
$$E(\xi^2) = 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} = \frac{1 + 4 + ... + 36}{6} = \frac{91}{6}.$$
$$D^2\xi = E(\xi^2) - (E\xi)^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Binomiális eloszlás

2.34. Definíció. *Legyen n egy természetes szám ($n \geq 1$) és $0 < p < 1$. A ξ valószínűségi változót (n,p)-**paraméterű binomiális eloszlású** valószínűségi változónak nevezzük, ha ξ értékei: $0, 1, ..., n$ és eloszlása*
$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{ahol } k = 0, ..., n.$$

2.35. Tétel (A binomiális eloszlás jellemzői).
Várható értéke: $E\xi = np$.
Szórásnégyzete: $D^2\xi = np(1-p)$.

Megjegyzés: Mikor jelenik meg a binomiális eloszlás?

Egy kísérletet hajtunk végre, legyen A egy kitüntetett esemény $p = P(A)$. A kísérletet n-szer egymás után végrehajtjuk. Legyen ξ az A esemény bekövetkezésének a száma az n kísérlet során. Ekkor ξ binomiális eloszlású valószínűségi változó (n,p)-paraméterekkel.

Példa: Érmével dobunk 4-szer, legyen ξ a dobott fejek száma. $P(\xi = 0)$, $P(\xi = 1)$, $P(\xi = 2)$, $P(\xi = 3)$, $P(\xi = 4)$, $P(\xi < 3) =?$ $E\xi$, $D^2\xi =?$

a) az érme szabályos
Követjük az előbb megadott sémát.
Kísérlet: érmével dobunk.
Kitüntetett esemény(A): fejet dobunk.
$p = P(A) = 1/2$.
$n = 4$ kísérletet hajtunk végre.
ξ számolja a dobott fejek számát
$P(\overline{A}) = P(\text{írás}) = 1 - p = 1/2$,
Ekkor

$$P(\xi = 0) = P(\text{ 0 fej és 4 írás }) = \binom{4}{0}(1/2)^0(1/2)^4 = 1(1/2)^4 = 1/16,$$

$$P(\xi = 1) = P(\text{ 1 fej és 3 írás }) = \binom{4}{1}(1/2)^1(1/2)^3 = 4(1/2)^4 = 4/16,$$

$$P(\xi = 2) = P(\text{ 2 fej és 2 írás }) = \binom{4}{2}(1/2)^2(1/2)^2 = 6(1/2)^4 = 6/16,$$

$$P(\xi = 3) = P(\text{ 3 fej és 1 írás }) = \binom{4}{3}(1/2)^3(1/2)^1 = 4(1/2)^4 = 4/16,$$

$$P(\xi = 4) = P(\text{ 4 fej és 0 írás }) = \binom{4}{4}(1/2)^4(1/2)^0 = 1(1/2)4 = 1/16.$$

$$P(\xi < 3) = P(\xi = 0) + P(\xi = 1) + P(\xi = 2) = 1/16 + 4/16 + 6/16 = 11/16.$$

$$E\xi = np = 4\frac{1}{2} = \frac{4}{2} = 2, \qquad D^2\xi = np(1-p) = 4\frac{1}{2}\frac{1}{2} = 1.$$

Ezeket az eredményeket korábban már kiszámoltuk.

b) Az érme szabálytalan, 1/3 eséllyel ad fejet és 2/3 eséllyel ad írást.
A séma marad, csak $p = P(A) = 1/3$ lesz. $P(\overline{A}) = P(\text{írás}) = 2/3$.

$$P(\xi = 0) = P(\text{ 0 fej és 4 írás }) = \binom{4}{0}(1/3)^0(2/3)^4 = 16/81,$$

$$P(\xi = 1) = P(\text{ 1 fej és 3 írás }) = \binom{4}{1}(1/3)^1(2/3)^3 = 32/81,$$

$$P(\xi = 2) = P(\text{ 2 fej és 2 írás }) = \binom{4}{2}(1/3)^2(2/3)^2 = 24/81,$$

$$P(\xi = 3) = P(\text{ 3 fej és 1 írás }) = \binom{4}{3}(1/3)^3(2/3)^1 = 8/81,$$

$$P(\xi = 4) = P(\text{ 4 fej és 0 írás }) = \binom{4}{4}(1/3)^4(2/3)^0 = 1/81.$$

$$P(\xi < 3) = P(\xi = 0)+P(\xi = 1)+P(\xi = 2) = 16/81+32/81+24/81 = 72/81.$$

$$E\xi = np = 4\frac{1}{3} = \frac{4}{3}, \qquad D^2\xi = np(1-p) = 4\frac{1}{3}\frac{2}{3} = \frac{8}{9}.$$

Megjegyzés:

1. Kísérletek ismertető jegyei:
   - a.) tetszőlegesen sokszor megismételhető,
   - b.) egymástól függetlenül hajtjuk végre, azaz egy kísérlet kimenetele nem befolyásolja a következő kísérlet kimenetelét,
   - c.) lényegileg azonos feltételek mellett hajtjuk végre.
2. Fontos észrevenni, hogy visszatevéses mintavétel esetén binomiális eloszlással számolunk, hiszen a ,,kísérlet” feltételeit visszaállítjuk. Egyszerű (visszatevés nélküli) mintavétel esetén pedig hipergeometrikus eloszlással számolunk (lásd később).

2.36. Tétel (A binomiális eloszlás határeloszlása). *Legyen p kicsi és n nagy oly módon, hogy np = állandó=λ. Ekkor*

$$\lim_{n\to\infty}\binom{n}{k}p^k(1-p)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}.$$

Példa: Egy forgalmas postahivatalban 365 nap alatt 1017 db bélyeg nélküli levelet adnak fel. Mi annak a valószínűsége, hogy egy napon 2-nél több bélyeg nélküli levelet adnak fel?

Kísérlet: feladunk egy levelet,
kitüntetett esemény (A): a levél bélyeg nélküli,
$p = P(A)$ ismeretlen,
ξ: egy napon feladott bélyeg nélküli levelek száma,
$E\xi = 2,8 = np$, $n=$? $p =$?
Nem adható meg n és p!
$P(\xi > 2) =$?

$$P(\xi > 2) = 1 - P(\xi \le 2) = 1 - (P(\xi = 0) + P(\xi = 1) + P(\xi = 2))$$

n nagy, hisz sok levelet adnak fel, p kicsi.

Tételünk szerint $\lambda = np = 2,8$-al számolhatunk
$P(\xi = 0) = \binom{n}{0} p^0 (1-p)^n \approx \frac{\lambda^0}{0!} e^{-\lambda} = e^{-2,8}$,
$P(\xi = 1) = \binom{n}{1} p^1 (1-p)^{n-1} \approx \frac{\lambda^1}{1!} e^{-\lambda} = 2,8 e^{-2,8}$,
$P(\xi = 2) = \binom{n}{2} p^2 (1-p)^{n-2} \approx \frac{\lambda^2}{2!} e^{-\lambda} = \frac{2,8^2}{2} e^{-2,8}$,
és $P(\xi > 2) \approx 1 - e^{-2,8}(1 + 2,8 + \frac{2,8^2}{2})$.

Poisson-eloszlás

2.37. Definíció. *Legyen $\lambda > 0$ egy rögzített szám. Azt mondjuk, hogy a ξ valószínűségi változó **λ-paraméterű Poisson-eloszlású** valószínűségi változó, ha értékei: 0, 1, 2, ... és eloszlása*

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{ahol } k = 0, 1, 2, \ldots$$

2.38. Tétel (a Poisson eloszlás jellemzői).
Várható értéke: $E\xi = \lambda$.
Szórásnégyzete: $D^2\xi = \lambda$.

Példa: Kalács sütéskor 1 kg tésztába 30 szem mazsolát tesznek. Majd a kész süteményt felszeletelik 5 dkg-os szeletekre. Mi annak a valószínűsége, hogy egy szeletben 2-nél több mazsola van? Hány darab mazsola nélküli szelet várható?

Legyen ξ a mazsolák száma egy 5 dkg-os szeletben.
$\lambda = E\xi = 3/2$ lesz.

$$P(\xi > 2) = 1 - P(\xi \le 2) = 1 - (P(\xi = 0) + P(\xi = 1) + P(\xi = 2))$$
$$= 1 - \left(\frac{\lambda^0}{0!} e^{-\lambda} + \frac{\lambda^1}{1!} e^{-\lambda} + \frac{\lambda^2}{2!} e^{-\lambda} \right) = 1 - e^{-3/2} \left(1 + \frac{3}{2} + \frac{9}{8} \right).$$

$$P(\xi = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-3/2} = 0,223$$

Tehát az összes szelet 22,3%-a mazsola nélküli, azaz 4,46 szelet lesz várhatóan mazsolátlan.

Geometriai eloszlás (elsőrendű negatív binomiális eloszlás)

Egy kísérlet lehetséges kimenetele az A esemény, $p = P(A)$. A kísérletet annyiszor hajtjuk végre amíg A be nem következik. Legyen ξ a kísérletek száma. Ekkor ξ értékei: 1,2,... és eloszlása

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

ahol $k = 1, 2, ...$

Ekkor ξ eloszlását **geometriai eloszlásnak (elsőrendű negatív binomiális eloszlásnak)** nevezzük.

2.39. Tétel (a geometriai eloszlás jellemzői).

Várható érték: $E\xi = \frac{1}{p}$.

Szórásnégyzet: $D^2\xi = \frac{1-p}{p^2}$.

Példa: ,,Gazdálkodj okosan!" játékban akkor léphetünk tovább az egyik mezőről, ha 1-t vagy 6-t dobunk. Mi annak a valószínűsége, hogy 3-dik dobásra lépünk tovább? Átlagosan hányadikra lépünk tovább?

Kísérlet: kockadobás.

Kitüntetett esemény (A): 1-t vagy 6-t dobunk.

$p = P(A) = \frac{1}{3}$.

ξ: a dobások száma a tovább lépésig.

$$P(\xi = 3) = (1-p)^2 p = \left(\frac{2}{3}\right)^2 \frac{1}{3} = \frac{4}{27} = 0,148.$$

$$E\xi = \frac{1}{p} = 3.$$

Átlagosan 3 dobás alatt tovább lépünk.

Hipergeometrikus eloszlás

Legyen m elemünk, amelyből s darabot megkülönböztetünk, ezek a ,,jó elemek" a többi $m - s$ darab ,,rossz elem"-től. Ezután találomra kiválasztunk az m elemből n darabot visszatevés nélkül (ahol $n \leq s$ és $n \leq m - s$). Legyen ξ az a valószínűségi változó, amelynek értéke az n kiválasztott darab között levő ,,jó elemek" száma. ξ értékei $0, 1, ..., n$ a ξ eloszlása

$$P(\xi = k) = \frac{\binom{s}{k}\binom{m-s}{n-k}}{\binom{m}{n}} \qquad (k = 0, 1, ..., n).$$

Ekkor a ξ diszkrét valószínűségi változót **hipergeometrikus eloszlásúnak** nevezzük.

2.40. Tétel (a hipergeometrikus eloszlás jellemzői).

Legyen $p = s/m$=a ,,jók" aránya az összesből. Ekkor

Várható érték: $E\xi = np$.

Szórásnégyzet: $D^2\xi = np(1-p)\left(1 - \frac{n-1}{m-1}\right)$.

Példa: Egy rekeszben 3 jó és 2 hibás alkatrész van, 3 darabot kiválasztunk közülük találomra. Legyen ξ értéke a kivett jó alkatrészek száma. Adjuk meg ξ eloszlását, várható értékét és szórását!

ξ értékei (x_i): 1, 2, 3.

ξ eloszlása ($m = 5$, $s = 3$, $n = 3$ választással):

$$P(\xi = k) = \frac{\binom{3}{k}\binom{2}{3-k}}{\binom{5}{3}}, \qquad (k = 1, 2, 3).$$

Tehát

$$P(\xi = 1) = \frac{\binom{3}{1}\binom{2}{2}}{\binom{5}{3}} = \frac{3}{10},$$

$$P(\xi = 2) = \frac{\binom{3}{2}\binom{2}{1}}{\binom{5}{3}} = \frac{6}{10},$$

$$P(\xi = 3) = \frac{\binom{3}{3}\binom{2}{0}}{\binom{5}{3}} = \frac{1}{10}.$$

$$E\xi = np = 3\frac{3}{5} = \frac{9}{5}.$$

$$D^2\xi = np(1-p)\left(1 - \frac{n-1}{m-1}\right) = 3\frac{3}{5}\frac{2}{5}\left(1 - \frac{2}{4}\right) = \frac{18}{50}.$$

2.7. Nevezetes folytonos eloszlások

Egyenletes eloszlás az $[a, b]$ intervallumon:

2.41. Definíció. *Legyen adva egy $[a, b]$ intervallum. Azt mondjuk, hogy a $\xi : \Omega \to \mathbf{R}$ folytonos valószínűségi változó **egyenletes eloszlású az** $[a, b]$ **intervallumon**, ha sűrűségfüggvénye*

$$f_\xi(x) = \begin{cases} \frac{1}{b-a}, & \text{ha } a \le x \le b, \\ 0, & \text{máskor} \end{cases}$$

alakú.

2.42. Tétel (az egyenletes eloszlás tulajdonságai).

Eloszlásfüggvénye:

$$F_\xi(x) = \begin{cases} 0, & \text{ha } x < a, \\ \frac{x-a}{b-a}, & \text{ha } a \le x \le b, \\ 1, & \text{ha } b < x. \end{cases}$$

Várható értéke: $E\xi = \frac{a+b}{2}$.
Szórásnégyzete: $D^2\xi = \frac{(b-a)^2}{12}$.

Példa: Egy távbeszélő állomás és a központ közötti vezeték hossza 450m. Mi annak a valószínűsége, hogy az első hiba a vezetéknek a központtól 180m-nél távolabbi helyén lép fel, ha a vezeték mentén bárhol azonos a meghibásodás veszélye?

Mivel a vezeték mentén bárhol azonos a meghibásodás veszélye, ezért geometriai valószínűséggel számolunk.

$$P(180 \text{ m-nél távolabb lép fel a hiba}) = \frac{\text{kedvező szakasz hossza}}{\text{összes szakasz hossza}} = \frac{270}{450} = \frac{3}{5}.$$

Nézzük meg jobban ezt a feladatot!
Legyen ξ a hiba távolsága a központtól. Ekkor ξ értékkészlete: $[0, 450]$.
Ha $x \leq 0$ akkor

$$F_\xi(x) = P(\xi < x) = P(\emptyset) = 0.$$

Ha $0 < x \leq 450$ akkor

$$F_\xi(x) = P(\xi < x) = \frac{\text{kedvező szakasz hossza}}{\text{összes szakasz hossza}} = \frac{x}{450}.$$

Ha $450 \leq x$ akkor

$$F_\xi(x) = P(\xi < x) = P(\Omega) = 1.$$

Tehát az eloszlásfüggvény:

$$F_\xi(x) = \begin{cases} 0, & \text{ha } x \leq 0, \\ \frac{x}{450}, & \text{ha } 0 < x \leq 450, \\ 1, & \text{ha } 450 \leq x. \end{cases}$$

Azaz a ξ valószínűségi változó egyenletes eloszlású a $[0, 450]$ intervallumon.

2.23. tétel után következő megjegyzésünk szerint:

$$P(\xi \geq 180) = 1 - P(\xi < 180) = F_\xi(180) = \frac{180}{450}.$$

(Lásd még a 2.28. definíció után következő feladatot is!)

Megjegyzés: Adott az $[a, b]$ intervallum, az intervallum minden pontját ugyanolyan valószínűséggel választom ki, legyen a ξ a kiválasztott szám értéke. Ekkor annak valószínűsége, hogy a $\xi \in [a', b']$ ($\subseteq [a, b]$) úgy is számolható, hogy

$$P(a' \leq \xi \leq b') = \frac{\text{kedvező szakasz hossza}}{\text{összes szakasz hossza}} = \frac{b' - a'}{b - a},$$

illetve

$$P(a' \leq \xi \leq b') = F(b') - F(a').$$

Átlagosan az intervallum közepét választjuk ki, azaz $E\xi = \frac{a+b}{2}$.

λ-paraméterű exponenciális eloszlás

2.43. Definíció. *Legyen $\lambda > 0$ egy valós szám. Azt mondjuk, hogy a ξ folytonos valószínűségi változó* ***λ-paraméterű exponenciális eloszlású*** *valószínűségi változó, ha sűrűségfüggvénye:*

$$f_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{ha } x > 0, \\ 0, & \text{máskor.} \end{cases}$$

2.44. Tétel (az exponenciális eloszlás tulajdonságai).
Várható értéke: $E\xi = \frac{1}{\lambda}$.
Szórás négyzete: $D^2\xi = \frac{1}{\lambda^2}$.
Eloszlásfüggvénye:

$$F_\xi(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{ha } x > 0, \\ 0, & \text{máskor.} \end{cases}$$

Példa: Annak valószínűsége, hogy egy benzinkútnál 6 percnél többet kell várakoznunk, a tapasztalat szerint 0,1. A várakozási idő exponenciális eloszlású. Mi annak a valószínűsége, hogy egy gépkocsi 3 percen belül sorra kerül? Tudjuk, hogy egy gépkocsi már 3 percet várakozott, mi annak a valószínűsége, hogy további 3 percen belül sorra kerül?

Legyen ξ a várakozási idő (percben). Először határozzuk meg az exponenciális eloszlás λ paraméterét!

$$P(\xi \geq 6) = 0,1.$$

$$1 - P(\xi < 6) = 0,1.$$

$$1 - F_\xi(6) = 0,1.$$

$$1 - (1 - e^{-\lambda 6}) = 0,1.$$

$$e^{-\lambda 6} = 0,1.$$
$$\lambda = \frac{-\ln 0,1}{6}.$$

Most, válaszoljuk meg a kérdéseket.

$$P(\xi < 3) = F_\xi(3) = 1 - e^{-\lambda 3} = 1 - e^{\frac{3\ln 0,1}{6}} = 1 - \left(e^{\ln 0,1}\right)^{\frac{1}{2}} = 1 - \sqrt{0,1}.$$

$$P(\xi < 6 | \xi \geq 3) = \frac{P(3 \leq \xi < 6)}{P(\xi \geq 3)} = \frac{F_\xi(6) - F_\xi(3)}{1 - F_\xi(3)} = \frac{(1 - e^{-\lambda 6}) - (1 - e^{-\lambda 3})}{1 - (1 - e^{-\lambda 3})}$$
$$= \frac{e^{-\lambda 3} - e^{-\lambda 6}}{e^{-\lambda 3}} = \frac{e^{-3\lambda}(1 - e^{-3\lambda})}{e^{-3\lambda}} = 1 - e^{-3\lambda} = P(\xi < 3).$$

Általában kijelenthetjük a következő tételt is.

2.45. Tétel (Az exponenciális eloszlás örök ifjú tulajdonsága). *Legyen ξ exponenciális eloszlású valószínűségi változó, ekkor*

$$P(\xi < x + t | \xi \geq t) = P(\xi < x)$$

és

$$P(\xi \geq x + t | \xi \geq t) = P(\xi \geq x)$$

egyenlőségek teljesülnek.

Megjegyzés: Általában a várakozási idő és az élettartam exponenciális eloszlású szokott lenni.

(m, σ^2)-paraméterű normális eloszlás

2.46. Definíció. *Legyen m egy tetszőleges valós szám, $\sigma > 0$ valós szám. Azt mondjuk, hogy a ξ folytonos eloszlású valószínűségi változó (m, σ^2)**-paraméterű normális eloszlású** valószínűségi változó, ha sűrűségfüggvénye:*

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad \textit{minden } x \in \mathbf{R} \textit{ esetén.}$$

2.47. Tétel (a normális eloszlás tulajdonságai).

Várható értéke: $E\xi = m$.

Szórás négyzete: $D^2\xi = \sigma^2$.

Eloszlásfüggvénye:

$$F_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

2.48. Definíció. *Elnevezés:* $m = 0$, $\sigma = 1$ *esetén* ***standard normális eloszlásról*** *beszélünk, sűrűségfüggvényét* $\varphi(x)$*-el, eloszlásfüggvényét* $\Phi(x)$*-el jelöljük. Azaz*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \textit{és} \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Megjegyzés:

1. Mivel Φ értékei nem számolhatóak könnyen ezért Φ értékeit táblázatba foglalták (integrál közelítő összegek segítségével számoltak).

2. Ha x negatív akkor a $\Phi(x)$ nem szerepel a táblázatban, mert $\Phi(x) = 1 - \Phi(-x)$ illetve $\Phi(-x) = 1 - \Phi(x)$ teljesül.

3. ξ normális eloszlású (m, σ^2) paraméterekkel. Ekkor a $\frac{\xi - m}{\sigma}$-t a ξ standardizáltjának nevezzük, és a ξ standardizáltja standard normális eloszlású.

Példa: Egy gyár rádióalkatrészeket gyárt. Egy bizonyos alkatrész élettartama a vizsgálatok szerint normális eloszlású, 1170 óra várható értékkel és 100 óra szórással. A gyár az alkatrészre garanciát vállal. Hány óra működésre szóljon a garancia, ha a gyár legfeljebb 5% garancia igényt kíván kielégíteni?

Legyen ξ egy alkatrész élettartama (órában).

$m = 1170$, $\sigma = 100$, $t =$ garancia idő =?

$$P(\xi < t) = 0,05.$$
$$P\left(\frac{\xi - m}{\sigma} < \frac{t - m}{\sigma}\right) = 0,05.$$
$$\Phi\left(\frac{t - m}{\sigma}\right) = 0,05.$$

Ez nem szerepel a táblázatban! Miért? Mert $\frac{t-m}{\sigma} < 0$ kell legyen! Megjegyzésünk szerint

$$\Phi\left(-\frac{t - m}{\sigma}\right) = 1 - \Phi\left(\frac{t - m}{\sigma}\right) = 0,95.$$

Tehát

$$-\frac{t - 1170}{100} = 1,65 \qquad / \cdot (-100)$$
$$t - 1170 = -165 \qquad / + 1170$$
$$t = 1005 \text{ óra}.$$

2.8. A nagy számok törvényei

2.49. Tétel (Markov-egyenlőtlenség). *Legyen $\xi \geq 0$ és $E\xi$ létezik. Legyen $\varepsilon > 0$. Ekkor*

$$P(\xi \geq \varepsilon) \leq \frac{E\xi}{\varepsilon}.$$

2.50. Tétel (Csebisev-egyenlőtlenség). *ξ egy olyan valószínűségi változó, melynek várható értéke és szórása létezik. Legyen $\varepsilon > 0$. Ekkor*

$$P(|\xi - E\xi| \geq \varepsilon) \leq \frac{D^2\xi}{\varepsilon^2} \qquad \textit{illetve } P(|\xi - E\xi| < \varepsilon) \geq 1 - \frac{D^2\xi}{\varepsilon^2}.$$

Példa: Egy forgalmas útkereszteződésben egy óra alatt áthaladó gépkocsik száma legyen egy ξ valószínűségi változó. A felmérésekből ismert, hogy $E\xi = 500$, $D\xi = 25$. Legalább mekkora valószínűséggel esik 400 és 600 közzé az útkereszteződésen egy óra alatt áthaladó gépkocsik száma?

$$P(400 < \xi < 600) = P(-100 < \xi - E\xi < 100) = P(|\xi - E\xi| < 100) \geq$$
$$\geq 1 - \frac{D^2\xi}{\varepsilon^2} = 1 - \frac{25^2}{100^2}. = 1 - \left(\frac{1}{4}\right)^2 = \frac{15}{16}.$$

Megjegyzés: Ha ξ eloszlása ismert, akkor a pontosabb érték is kiszámolható, a feladat szövege erről nem tartalmaz információt.

2.51. Tétel (Bernoulli-féle nagy számok törvénye). *Legyen $p = P(A)$ az n kísérletből az A esemény k-szor következik be és legyen $\varepsilon > 0$. Ekkor*

$$P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{\varepsilon^2 n} \qquad \textit{illetve } P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{p(1-p)}{\varepsilon^2 n}$$

Megjegyzés:

1. Ha a p nem ismert, akkor használhatjuk még a $P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4\varepsilon^2 n}$ és a $P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{1}{4\varepsilon^2 n}$ alakot is

2. a $\frac{k}{n}$ számot az A esemény relatív gyakoriságának nevezzük (lásd 2.2. fejezetet).

Példa: 1.) Egy gyártmány 10%-a másodosztályú. A minőség ellenőrzés csak akkor találja elfogadhatónak a tételt, ha benne legfeljebb 12% másodosztályú.

Mekkora legyen a tételben a darabszám, hogy a hibás áruk relatív gyakorisága a megfelelő valószínűségtől legalább 0,95 valószínűséggel ne térjen el a megengedett mértékben? A

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)\geq 1-\frac{p(1-p)}{\varepsilon^2 n}$$

egyenlőtlenséget használjuk ($p=0,1$ és $\varepsilon=0,12-0,10$ választással).

$$P\left(\left|\frac{k}{n}-0,1\right|<0,02\right)\geq 1-\frac{0,1\cdot 0,9}{0,02^2 n}\geq 0,95.$$

Oldjuk meg a

$$1-\frac{0,09}{0,02^2 n}\geq 0,95$$

egyenlőtlenséget!

$$0,05\geq\frac{225}{n}$$

azaz

$$n\geq 4500.$$

2) Egy automatával meg szeretnénk határozni a selejt arányt. E célból megvizsgálunk 5000 terméket, összesen 80 selejtes terméket találunk. Határozzuk meg, hogy az ebből számított relatív gyakoriság ($\frac{80}{5000}=0,016$) az ismeretlen p valószínűséget 90% biztonsággal mennyire közelíti meg!

Az előző feladatban használt egyenlőtlenség nem használható, mert ε és p két ismeretlen. Használjuk a megjegyzésünket!

$$P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)\geq 1-\frac{1}{4\varepsilon^2 n},$$

behelyettesítjük az ismert számokat, így

$$P\left(\left|\frac{80}{5000}-p\right|<\varepsilon\right)\geq 1-\frac{1}{4\varepsilon^2 5000}\geq 0,90.$$

$$0,1\geq\frac{1}{20000\varepsilon^2},$$

Tehát $\varepsilon^2\geq 0,0005$ azaz $\varepsilon\geq 0,022$ Ebből a

$$-\varepsilon\leq p-\frac{k}{n}\leq\varepsilon$$

egyenlőtlenség így alakul:

$$-0,022+0,016\leq p\leq 0,022+0,016$$

s ez 90%-os biztonsággal teljesül.

2.52. Tétel (Centrális határeloszlás tétel). *Legyenek $\xi_1, ..., \xi_n$ független, azonos eloszlású valószínűségi változók és $S_n := \xi_1 + ... + \xi_n$. Tegyük fel, hogy $E\xi = m$ és $D^2\xi = \sigma^2$ létezik. Ekkor*

$$\lim_{n\to\infty} P\left(\frac{S_n - ES_n}{DS_n} < x\right) = \Phi(x) \qquad \textit{minden } x \in \mathbf{R} \textit{ esetén.}$$

Megjegyzés:

1. A képlet jelentése $F_{\frac{S_n - ES_n}{DS_n}}(x) \approx \Phi(x)$ ha az n nagy. Az $\frac{S_n - ES_n}{DS_n} = \frac{S_n - nm}{\sqrt{n}\sigma}$ az S_n valószínűségi változó standardizáltja.
2. A binomiális eloszlás előállítható n darab független azonos eloszlású valószínűségi változó összegeként. Ezért a binomiális eloszlás eloszlásfüggvénye jól közelíthető a normális eloszlás eloszlásfüggvényével, ha n nagy.

Példa: Egy tétel áru 40%-a hibátlan, a maradék másodosztályú. Egyenként választunk ki belőle véletlenszerűen n =200 darabot, amelyeket a kiválasztás után megvizsgáljuk és azonnal visszatesszük. Mi annak a valószínűsége, hogy 100-nál kevesebb esetben vettünk ki másodosztályút?

Legyen ξ a másodosztályú termékek száma a kiválasztott termékek között. Ekkor ξ binomiális eloszlású lesz $n = 200$ és $p = 0,6$ paraméterekkel (visszatevéses mintavétel).

$E\xi = np = 120$ és $D^2\xi = np(1-p) = 200 \cdot 0,6 \cdot 0,4 = 48$. A kérdés: $P(\xi < 100) =$? Mivel n nagy, így alkalmazzuk a centrális határeloszlás tételt.

$$P(\xi < 100) = P\left(\frac{\xi - E\xi}{D\xi} < \frac{100 - 120}{\sqrt{48}}\right) \approx \Phi\left(\frac{-20}{\sqrt{48}}\right) = \Phi(-2,88)$$
$$= 1 - \Phi(2,88) = 1 - 0,9980 = 0,002.$$

Megjegyzés: Ha a binomiális eloszlással számolnánk, akkor a

$$P(\xi < 100) = \sum_{k=0}^{99} \binom{200}{k} 0,6^k \; 0,4^{200-k}$$

összeget kellene kiszámolnunk.

2.9. Kovariancia és korrelációs együttható

2.53. Definíció. *A ξ és a η valósínűségi változók* ***kovarianciája*** *a*

$$cov(\xi, \eta) := E[(\xi - E\xi)(\eta - E\eta)]$$

szám, amennyiben a leírt várható érték létezik. Ha $cov(\xi, \eta) = 0$ úgy azt mondjuk, hogy a két valószínűségi változó korrelálatlan.

2.54. Tétel (a kovariancia tulajdonságai).

a) A ξ és η valószínűségi változók kovarianciája létezik, ekkor

$$cov(\xi, \eta) = E(\xi\eta) - (E\xi)(E\eta).$$

b) Legyen ξ és η független valószínűségi változók, ekkor $cov(\xi, \eta) = 0$.

c) Megadható két valószínűségi változó , úgy hogy $cov(\xi, \eta) = 0$ és a ξ és η nem függetlenek. (Tehát a b) tétel megfordítása nem igaz.)

2.55. Definíció. *Legyen ξ és η két olyan valószínűségi változó, hogy $D\xi, D\eta > 0$. A*

$$corr(\xi, \eta) := \frac{cov(\xi, \eta)}{D\xi D\eta}$$

számot a ξ és az η valószínűségi változók ***korrelációs együtthatójának*** *nevezzük.*

Megjegyzés: A korrelációs együtthatót a függőség ill. a függetlenség mérésére használják. Ezt mondja ki a következő tétel is.

2.56. Tétel (korrelációs együttható mint a függetlenség mérőszáma). *Legyen ξ és η olyan valószínűségi változók, hogy a korrelációs együtthatójuk létezik.*

a) Ha ξ és η függetlenek, akkor $corr(\xi, \eta) = 0$.

b) $|corr(\xi, \eta)| \leq 1$

c) A $|corr(\xi, \eta)| = 1$ pontosan akkor, ha $\xi = a\eta + b$ alakú egy valószínűséggel (a, b valós számok és $a \neq 0$). Ha $a > 0$, akkor $corr(\xi, \eta) = 1$, ha $a < 0$, akkor $corr(\xi, \eta) = -1$.

Példa: Számítsa ki a korrelációs együtthatót, ha ismert a következő kontingencia táblázat!

$\eta \setminus \xi$	1	2	3
-1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

Azaz $P(\xi = 1, \eta = -1) = \frac{1}{4}, ..., P(\xi = 3, \eta = 1) = \frac{1}{8}$.

Megjegyzés: Kovarianciát és korrelációs együtthatót mi csak diszkrét valószínűségi változók esetében számoljuk ki.

3. Matematikai statisztika

3.1. Minta és tapasztalati eloszlásfüggvény

3.1. Definíció. *A $\xi_1...,\xi_n$ független, azonos eloszlású valószínűségi változók együttesét n-**elemű mintának** nevezzük. A minta elemek közös eloszlását **alapeloszlásnak** nevezzük. Az n számú kísérlet (megfigyelés) során minden ξ_i mintaelem egy konkrét x_i számot vesz fel értékként. Az $x_1, x_2, ..., x_n$ számokat a **minta realizációjának** nevezzük.*

3.2. Definíció. ***Statisztikai függvény*** *(röviden **statisztika**) alatt olyan $T : \mathbf{R}^n \to \mathbf{R}$ függvényt értünk, amelyre teljesül az, hogy $T(\xi_1, \xi_2, ..., \xi_n)$ valószínűségi változó.*

3.3. Definíció.

*A **mintaátlag:***

$$\overline{\xi} := \frac{\xi_1 + \xi_2 + ... + \xi_n}{n},$$

tapasztalati (empirikus) szórásnégyzet:

$$S_n^2 := \frac{\sum_{i=1}^n (\xi_i - \overline{\xi})^2}{n} = \frac{(\xi_1 - \overline{\xi})^2 + (\xi_2 - \overline{\xi})^2 + ... + (\xi_n - \overline{\xi})^2}{n},$$

korrigált tapasztalati (empirikus) szórásnégyzet:

$$S_n^{*2} := \frac{\sum_{i=1}^n (\xi_i - \overline{\xi})^2}{n-1} = \frac{(\xi_1 - \overline{\xi})^2 + (\xi_2 - \overline{\xi})^2 + ... + (\xi_n - \overline{\xi})^2}{n-1}.$$

Megjegyzés:

1. A mintaátlag, (korrigált) tapasztalati szórásnégyzet statisztika.
2. $S_n^{*2} = \frac{n}{n-1} S_n^2$ teljesül.
3. Ha a mintabeli adatok gyakorisággal vannak megadva (tehát a minta realizáció ismert), azaz az x_i mintaelem f_i-szer fordul elő ($\sum_{i=1}^k f_i = n$), úgy

$$\overline{x} = \frac{\sum_{i=1}^k f_i x_i}{n}; \qquad S_n^2 = \frac{\sum_{i=1}^k f_i (x_i - \overline{x})^2}{n}; \qquad S_n^{*2} = \frac{\sum_{i=1}^k f_i (x_i - \overline{x})^2}{n-1}.$$

3.1. Lemma (Steiner formula). *Tetszőleges c valós szám esetén*

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - c)^2 - \left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i - c)\right)^2.$$

Speciálisan $c = 0$ esetén:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - (\ \overline{\xi}\ )^2.$$

3.4. Definíció. *Rendezzük a mintát nagyság szerint sorrendbe, ekkor úgynevezett **rendezett mintát** kapunk, ez $\xi_1^*, \xi_2^*, ..., \xi_n^*$. Ekkor a minta **mediánja** páratlan n-re a középső elem (azaz, ha $n = 2m+1$ úgy a medián ξ_{m+1}^*). Páros n-re a két középső elem átlaga (azaz ha $n = 2m$ úgy $\frac{\xi_m^* + \xi_{m+1}^*}{2}$).*

*A minta **módusza** a leggyakrabban előforduló elem, ha van olyan.*

*A minta **terjedelme:** $R := \xi_n^* - \xi_1^*$ azaz a legnagyobb és a legkisebb mintaelem különbsége.*

Megjegyzés: A rendezett minta elemei statisztikák. Fontos tudni, hogy a $\xi_1^*, \xi_2^*, ..., \xi_n^*$ valószínűségi változók nem függetlenek és nem azonos eloszlásúak.

3.5. Definíció. *$\xi_1, \xi_2, ..., \xi_n$ minta. **Tapasztalati eloszlásfüggvény** (n elemű mintából nyert empirikus eloszlásfüggvény):*

$$F_n(x) := \sum_{i:\xi_i < x} \frac{1}{n}.$$

Megjegyzés: Az elméleti eloszlásfüggvényt, azaz az alapeloszlás $F(x)$ eloszlásfüggvényét a mintából nyert tapasztalati eloszlásfüggvénnyel közelítjük. Ha n elég nagy, úgy $F_n(x)$ jól közelíti $F(x)$-t. Ezt állítja a következő tétel.

3.6. Tétel (Glivenkó-tétele, a matematikai statisztika alaptétele).
Adott egy n elemű minta, az alapeloszlás eloszlásfüggvénye $F(x)$. Ekkor

$$P\left(\lim_{n\to\infty}\sup_{x\in\mathbf{R}}|F_n(x) - F(x)| = 0\right) = 1.$$

3.7. Definíció. *A minta elemekből készített oszlopdiagramot **hisztogramnak** nevezzük. **Sűrűségi hisztogram** olyan hisztogram, amelynél az oszlopok összterülete 1.*

Eljárás sűrűségi hisztogram készítésére:

1) Meghatározzuk azt az $[a, b]$ intervallumot, ahová a mintabeli adatok esnek (a lehet a legkisebb mintaelem, vagy egy nála kisebb hozzá közeli kerek szám) (b hasonlóan)

2) A $[a, b]$-t felosztjuk részintervallumokra, többnyire egyenlő hosszúságúakra. Majd egy részintervallum fölé olyan téglalapot rajzolunk, melynek területe $\frac{k}{n} = \frac{\text{az intervallumba eső adatok száma}}{\text{összes adatok száma}}$. (Tehát egy oszlop magassága $m = \frac{k}{nl} = \frac{\text{az intervallumba eső adatok száma}}{\text{összes adatok száma} \cdot \text{részintervallum hossza}}$) Az $[a, b]$ -n kívül 0. Az így kapott függvényt $f_n(x)$-el jelöljük.

Megjegyzés:

1. Folytonos esetben az alapeloszlás sűrűségfüggvényét a sűrűségi hisztogrammal közelíthetem, melyet tapasztalati sűrűségfüggvénynek is nevezünk és $f_n(x)$-el jelölünk.

2. Ahhoz, hogy $f_n(x)$ jól mutassa az elméleti sűrűségfüggvény $f(x)$ alakját, jól el kell találni az osztópontok számát az $[a, b]$ intervallumban. (Általában: 6-14.)

3. Ha a hisztogram valamely jól ismert eloszlás sűrűségfüggvényét közelíti jól, úgy azt az eloszlást feltételezzük a mintáról.

4. Ha az adatok száma nagy, akkor az adatokat osztályokba soroljuk, s a mért adatokat az osztályközéppel helyettesítjük.

Példa: A -50-es (MSZ 500-as) acél szakítási szilárdság ellenőrzésére az egész sokaságból $n = 31$ mérést végeztek. A mért értékek (N/mm^2-ben): 470, 481, 483, 488, 489, 490, 491, 492, 493, 493, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 512, 514, 516, 519, 529, 530.

a) Adja meg a minta móduszát, mediánját, terjedelmét!

b) Mutassuk ki, hogy az adatok alapján a szakítószilárdság eloszlása jól közelíthető normális eloszlással (szerkesszük meg a tapasztalati sűrűségfüggvényt)!

c) Becsüljük meg az egész sokaság m várható értékét (a $\overline{\xi}$ mintaátlaggal), az egész sokaság σ szórását (az S_n^* korrigált tapasztalati szórással)!

d) Egy bizonyos munkánál selejtesnek tekintjük azt az acélt, amelynek szakítási szilárdsága 485 N/mm^2 alatt marad. Becsüljük meg a selejt valószínűségét!

e) Végezzük el az előző becslést az eloszlásfüggvény megszerkesztésével, grafikus úton!

f) Az adatokat osztályokba sorolva, helyettesítsünk minden adatot az osztályközéppel! Becsüljük meg így a várható értéket és a szórást, szerkesszük meg így a tapasztalati eloszlásfüggvényt, becsüljük meg újra a selejt-valószínűségét!

a) $n = 31$ elemű minta esetén a középső a 16-dik, tehát a medián 500. A módusz 493, a mita terjedelme R=530-470=60.

b) Kövessük az eljárásunkat! Legyen $a := 470$ és $b := 530$, a minta terjedelme $R = 60$ azt mutatja, hogy célszerű 6 db 10 egység hosszú vagy 10 db 6 egység hosszú osztályokba sorolnunk. Válasszuk az elsőt!

osztályhatárok	gyakoriság	oszlop magasság	osztály közép
$[470; 480)$	1	1/310	475
$[480; 490)$	4	4/310	485
$[490; 500)$	10	10/310	495
$[500; 510)$	10	10/310	505
$[510; 520)$	4	4/310	515
$[520; 530]$	2	2/310	525

Ez alapján elkészíthetjük az oszlopdiagramunkat! Lásd előadáson!

$$f_{31}(x) = \begin{cases} 1/310, & \text{ha } x \in [470; 480), \\ 4/310, & \text{ha } x \in [480; 490), \\ 10/310, & \text{ha } x \in [490; 500), \\ 10/310, & \text{ha } x \in [500; 510), \\ 4/310, & \text{ha } x \in [510; 520), \\ 2/310, & \text{ha } x \in [520; 530], \\ 0, & \text{máskor.} \end{cases}$$

c) az m várható érték becslése a minta átlaggal:

$$m \approx \bar{\xi} = \frac{470 + 481 + ... + 530}{31} = 500,645.$$

A σ szórásnégyzet becslése a korrigált tapasztalati szórásnégyzettel:

$$\sigma^2 \approx S_n^{*2} = \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}{n-1} = \frac{(470 - 511,645)^2 + ... + (530 - 500,645)^2}{30} = 172,76.$$

d) A selejt valószínűsége becsülhető a relatív gyakorisággal:

$$P(\text{ selejt }) = P(\xi < 485) = \frac{3}{31} = 0,0967.$$

Tehát 9,7%. De ez csak $1/31 = 0,0322 = 3,22\%$-nyi pontosságot biztosít!

Normális eloszlást feltételezve, számolhatunk m, σ becsült értékével:

$$\begin{aligned} P(\xi < 485) &= P\left(\frac{\xi - m}{\sigma} < \frac{485 - 500,645}{13,144}\right) = \Phi(-1,19) = 1 - \Phi(1,19) \\ &= 1 - 0,8830 = 0,1170. \end{aligned}$$

Azaz $11,7\%$-ra becsülhetjük a selejt valószínűségét.

e) Lásd előadáson!

Vegyük észre, hogy hosszadalmas az egyes mért értékekből megszerkeszteni az $F_n(x)$ tapasztalati eloszlásfüggvényt, ha az adatok száma nagy. Ilyenkor az adatokat osztályba sorolva, a mért adatokat az osztályközepekkel helyettesítjük, s így is megszerkeszthetjük az $F_n(x)$-t.

f)

$$\overline{\xi} = \frac{1 \cdot 475 + 4 \cdot 485 + 10 \cdot 505 + 4 \cdot 515 + 2 \cdot 525}{31} = 500,806.,$$

$$S_n^{*2} = \frac{1(475 - 500,809)^2 + 4(485 - 500,809)^2 + ... + 2(525 - 500,809)^2}{30} = 138,48.$$

A most kapott értékek természetesen durvább becslések, mint az előzők, viszont rövidebben megkaphatók, mint korábban. A selejt valószínűsége újra megbecsülhető ($m \approx 500,806,\ \sigma \approx \sqrt{138,48} = 11,768$):

$$\begin{aligned} P(\xi < 485) &= P\left(\frac{\xi - m}{\sigma} < \frac{485 - 500,806}{11,768}\right) = \Phi(-1,34) = 1 - \Phi(1,34) \\ &= 1 - 0,9099 = 0,0901. \end{aligned}$$

Tehát a selejt valószínűségét $9,01\%$-ra becsüljük.

3.2. Becslések tulajdonságai

Legyen a $\xi_1, \xi_2, \ldots, \xi_n$ mintaelemek alapeloszlásának várható értéke m, szórásnégyzete σ^2.

Ekkor könnyű megmutatni, hogy

$$E\overline{\xi} = m,$$

vagyis a mintaátlag várható értéke megegyezik az alapeloszlás várható értékével. Ekkor azt mondjuk, hogy a mintaátlag **torzítatlan becslése** a várható értéknek.

Teljesül még

$$P\left(\lim_{n\to\infty} \overline{\xi} = m\right) = 1.$$

egyenlőség is amit úgy nevezünk, hogy a mintaátlag **erősen konzisztens becslése** a várható értéknek.

A tapasztalati szórásnégyzet várható értéke

$$ES_n^2 = \frac{n-1}{n}\sigma^2,$$

Tehát a tapasztalati szórásnégyzet **nem torzítatlan becslése** a szórásnégyzetnek.

Emlékezzünk a következő összefüggésre: $S_n^{*2} = \frac{n}{n-1}S_n^2$. Így a korrigált tapasztalati szórásnégyzetre

$$ES_n^{*2} := \sigma^2$$

teljesül, tehát a korrigált tapasztalati szórásnégyzet **torzítatlan becslése** a szórásnégyzetnek. Továbbá a korrigált tapasztalati szórásnégyzet **erősen konzisztens** becslése a szórásnégyzetnek, azaz a

$$P\left(\lim_{n\to\infty} S_n^{*2} = \sigma^2\right) = 1$$

egyenlőség teljesül.

3.3. A statisztika néhány nevezetes eloszlása

3.8. Definíció. *Ha $\eta_1, \eta_2, \ldots, \eta_k$ független, standard normális eloszlású valószínűségi változók, akkor a*

$$\chi_k^2 := \eta_1^2 + \eta_2^2 + \cdots + \eta_k^2$$

valószínűségi változó eloszlását χ^2***-eloszlásnak*** *nevezzük, melynek szabadsági foka k.*

Megjegyzés: A χ_k^2 valószínűségi változó folytonos eloszlású, sűrűségfüggvényét meg lehet határozni, az eloszlásfüggvényének értékeit táblázatba foglalták.

3.9. Definíció. *Ha η és χ_k^2 független valószínűségi változók standard normális, illetve k szabadsági fokú χ^2-eloszlással, akkor a*

$$t_k := \frac{\eta}{\sqrt{\chi_k^2/k}}$$

*valószínűségi változó eloszlását t***–eloszlásnak (Student–eloszlásnak)** *nevezzük, melynek szabadsági foka k.*

Megjegyzés:

1. A t_k valószínűségi változó folytonos eloszlású, sűrűségfüggvényét meg lehet határozni, az eloszlásfüggvényének értékeit táblázatba foglalták.
2. A t_k valószínűségi változó fogalmából következik, hogy a sűrűségfüggvénye szimmetrikus a y-tengelyre.

3.10. Definíció. *Ha χ_k^2 és χ_l^2 független valószínűségi változók k szabadsági fokú χ^2-, illetve l szabadsági fokú χ^2–eloszlással, akkor a*

$$F_{k,l} := \frac{\chi_k^2/k}{\chi_l^2/l}$$

*valószínűségi változó eloszlását $F_{k,l}$–**eloszlásnak** nevezzük, melynek szabadsági fokai k és l.*

Megjegyzés: A $F_{k,l}$ valószínűségi változó folytonos eloszlású, sűrűségfüggvényét meg lehet határozni, az eloszlásfüggvényének értékeit táblázatba foglalták.

3.11. Tétel (normális eloszlású minta jellemzői). *Ha $\xi_1, \xi_2, \ldots, \xi_n$ minta, (m, σ^2)-paraméterű normális alapeloszlással, akkor*

a) $\overline{\xi}$ *is normális eloszlású $(m, \sigma^2/n)$ paraméterekkel,*

b) $\overline{\xi}_n$ *és S_n^2 függetlenek,*

c) nS_n^2/σ^2 *eloszlása χ^2-eloszlás $n-1$ szabadsági fokkal.*

3.4. Paraméteres statisztikai próbák

Egymintás u-próba

Tegyük fel, hogy $\xi_1, \xi_2, \ldots, \xi_n$ minta (m, σ^2)-paraméterű normális alapeloszlásra, ahol σ^2 ismert, de m ismeretlen. Szeretnénk arról dönteni, hogy a minta alapján elfogadható-e az a nullhipotézis (feltevés), hogy $E\xi = m_0$, vagy pedig az $E\xi \neq m_0$ alternatív hipotézis (ellenhipotézis) fogadható el? Röviden ezt így fogjuk írni a továbbiakban:

$$\begin{aligned} H_0: &\quad E\xi = m_0 \\ H_1: &\quad E\xi \neq m_0. \end{aligned}$$

Ha a H_0 nullhipotézis igaz, akkor a 3.11 Tétel alapján az

$$u := \frac{\overline{\xi} - m_0}{\sigma/\sqrt{n}}$$

valószínűségi változó standard normális eloszlású. Legyen $1 > \varepsilon > 0$ rögzített szignifikancia szint, általában 0,1; 0,05; 0,01 értékeket szokták választani. Válasszuk meg az $u_{\varepsilon/2}$ számot úgy, hogy $P\left(u < -u_{\varepsilon/2}\right) = \varepsilon/2$, és $P\left(u < u_{\varepsilon/2}\right) = 1 - \varepsilon/2$. Ekkor

$$P\left(u \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]\right) = 1 - \varepsilon$$

lesz. Döntsünk a következő módon:

a) Ha az $u \in [-u_{\varepsilon/2}, u_{\varepsilon/2}]$ akkor **elfogadjuk** a H_0 nullhipotézist $(1 - \varepsilon)100\%$ biztonsági szinten. Ilyenkor a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallumot **elfogadási tartománynak** is nevezzük.

b) Ha az $u \notin [-u_{\varepsilon/2}, u_{\varepsilon/2}]$, akkor **elvetjük** a H_0 nullhipotézist, azaz elfogadjuk az alternatív hipotézist $(1 - \varepsilon)100\%$ biztonsági szinten. a $[-u_{\varepsilon/2}, u_{\varepsilon/2}]$ intervallum komplementerét **kritikus tartománynak** is nevezzük.

3.12. Definíció. ***Elsőfajú hiba:*** *annak a valószínűsége, hogy a H_0 nullhipotézist elvetjük, pedig a H_0 igaz. (Ez a próba terjedelme.)*

Másodfajú hiba: *annak a valószínűsége, hogy a H_0 nullhipotézist elfogadjuk, pedig a H_0 nem igaz.*

Megjegyzés:

1. Az elsőfajú hiba pont az ε szignifikancia szint.
2. Meg lehet mutatni, hogy az elsőfajú hiba csökkenése esetén a másodfajú hiba növekszik. Tehát a két hiba ellentétes irányban mozog.

Példa: Egy automata csővágó gép 1200 mm-es darabok levágására van beállítva. A levágott cső hossza véletlentől függő változó. Előzetes adatfelvételből tudjuk, hogy normális eloszlású, melynek szórása 3 mm. Kiválasztunk n=16 legyártott csövet. A mintából kapott méretek:

1193, 1198, 1203, 1191, 1195, 1196, 1199, 1191, 1201, 1196, 1193, 1198, 1204, 1196, 1198, 1200.

Elfogadható-e, hogy az eltérés nem szignifikáns, vagyis az egész sokaságban a várható érték: $m_0 = 1200$?

Megoldás:
ξ : egy levágott cső hossza.

$$\begin{aligned} H_0: &\quad E\xi = 1200 \\ H_1: &\quad E\xi \neq 1200. \end{aligned}$$

$\varepsilon = 0,05$-t választunk.

$$\overline{\xi} = \frac{1193 + 1198 + ... + 1200}{16} = 1197.$$

$$u_{\text{számolt}} = \frac{1197 - 1200}{\frac{3}{\sqrt{16}}} = -3\frac{4}{3} = -4.$$

Most meghatározzuk az elfogadási tartományt.

$$P\left(u < u_{\varepsilon/2}\right) = 1 - \varepsilon/2$$

azaz

$$\Phi(u_{\varepsilon/2}) = 1 - \varepsilon/2 = 1 - 0,05/2 = 0,975.$$

A standard normális eloszlás táblázatából $u_{\varepsilon/2} = 1,96$.

$u_{\text{számolt}} \in [-u_{\varepsilon/2}, +u_{\varepsilon/2}]$ teljesül-e?

$-4 \notin [-1,96, +1,96]$, tehát a nullhipotézist 95%-os biztonsági szinten elvetjük, azaz a csővágógép nem a megfelelő méretet vágja. Beállítást kell végre hajtani rajta.

Kétmintás u-próba

Legyen a $\xi_1, \xi_2, \ldots, \xi_k$ az (m_ξ, σ_ξ^2)-paraméterű normális eloszlású minta, és az $\eta_1, \eta_2, \ldots, \eta_l$ pedig (m_η, σ_η^2)-paraméterű normális eloszlású minta. A két minta egymástól független. A σ_ξ^2 és σ_η^2 ismert, de m_ξ és m_η ismeretlen. Döntsünk a következő hipotézisekről:

$$\begin{aligned} H_0: &\quad E\xi = E\eta, \quad (\text{ azaz } m_\xi = m_\eta) \\ H_1: &\quad E\xi \neq E\eta. \end{aligned}$$

Ha a H_0 nullhipotézis igaz, akkor az

$$u := \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{\sigma_\xi^2}{k} + \frac{\sigma_\eta^2}{l}}}$$

statisztikai függvény standard normális eloszlású lesz. Adott ε szignifikancia szint esetén az eljárást az egymintás u-próbánál leírtak szerint folytatjuk.

Példa: Egy kiterjedt népegészségügyi vizsgálat során megállapították, hogy az egészséges felnőtt populáció esetén a diasztolés (alsó) vérnyomás értékek átlaga 84,8 higanymilliméter, szórása pedig 12,8 higanymilliméter. Egy sport klub hat véletlenszerűen kiválasztott versenyzőjénél a klub sportorvosa az alábbi diasztolés értékeket jegyezte fel:

79.2; 64.6; 86.8; 73.7; 74.9; 62.3.

Ugyanabban a városban működő sakk klub versenyzői szintén meglátogatták az említett doktort, hat véletlenszerűen kiválasztott sakkozó vérnyomás értékét megmérte és feljegyezte, melyek az alábbiak:

84.6; 93.2; 104.6; 106.7; 76.3; 78.2.

Hipotéziseit pontosan megfogalmazva döntsön 95%-os szinten arról, hogy a sakkozók átlagos diasztolés vérnyomása magasabb-e, mint a sportolóké! A sportolók és a sakkozók diasztolés vérnyomásáról feltehetjük, hogy normális eloszlást követ, szórása pedig megegyezik a teljes népesség körében mért értékkel.

Megoldás:

Legyen ξ a sportolók, míg η a sakkozók alsó diasztolés értéke.

Kétmintás u-próbát hajtunk végre, mivel a minták szórását ismerjük.

$$\begin{aligned} H_0: &\quad E\xi = E\eta, \\ H_1: &\quad E\xi < E\eta. \end{aligned}$$

Tehát egyoldali próbát hajtunk végre, azaz a teljes hibát ε-t csak egy oldalra rakjuk.

$$\begin{aligned} \overline{\xi} &= \frac{79,2 + ... + 62,3}{6} = 73.58, \\ \overline{\eta} &= \frac{84,6 + ... + 78,2}{6} = 90,6. \\ u_{\text{ számolt}} &= \frac{73,58 - 90,6}{\sqrt{\frac{12,8^2}{6} + \frac{12,8^2}{6}}} = -2,3. \end{aligned}$$

Most meghatározzuk az elfogadási tartományt.

$$P\left(u < u_\varepsilon\right) = 1 - \varepsilon$$

azaz

$$\Phi(u_\varepsilon) = 1 - \varepsilon = 1 - 0,05 = 0,95.$$

A standard normális eloszlás táblázatából $u_\varepsilon = 1,64$.

$u_{\text{számolt}} \in [-u_\varepsilon, +u_\varepsilon]$ teljesül-e?

$-2,3 \notin [-1,64; +1,64]$, tehát a nullhipotézist elvetjük 95%-os biztonsági szinten, azaz a sportolók alsó diasztolés értéke kisebb alternatív hipotézist fogadjuk el.

Egymintás t-próba

Tegyük fel, hogy $\xi_1, \xi_2, \ldots, \xi_n$ minta (m, σ^2)-paraméterű normális alapeloszlásra, ahol m és σ^2 ismeretlen. Szeretnénk dönteni a következő hipotézisekről:

$$\begin{aligned} H_0: &\quad E\xi = m_0 \\ H_1: &\quad E\xi \neq m_0. \end{aligned}$$

Ha a H_0 nullhipotézis igaz, akkor a

$$t := \frac{\overline{\xi} - m_0}{S_n^*/\sqrt{n}}$$

valószínűségi változó t-eloszlású $f = n - 1$ szabadsági fokkal.

Legyen $1 > \varepsilon > 0$ rögzített szignifikancia szint. Válasszuk a $t_{\varepsilon/2}$ számot úgy, hogy $P\left(t < -t_{\varepsilon/2}\right) = \varepsilon/2$ és $P\left(t < t_{\varepsilon/2}\right) = 1 - \varepsilon/2$. Ekkor

$$P\left(t \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]\right) = 1 - \varepsilon$$

lesz. Döntsünk a következő módon:

a) Ha az $t \in [-t_{\varepsilon/2}, t_{\varepsilon/2}]$ akkor telfogadjuk a H_0 nullhipotézist $(1-\varepsilon)100\%$ biztonsági szinten.

b) Ha az $t \notin [-t_{\varepsilon/2}, t_{\varepsilon/2}]$, akkor elvetjük a H_0 nullhipotézist, azaz elfogadjuk az alternatív hipotézist $(1 - \varepsilon)100\%$ biztonsági szinten.

Példa: Egy konzervgyárban adagolóautomata tölti a dobozokat. Az egy dobozba töltendő anyag tömegének várható értékére az előírás 500 gr. Mintavétel során az alábbi értékeket kapták:

483, 502, 498, 496, 502, 483, 494, 491, 505, 486.

Döntsünk 95%-os biztonsági szinten arról, hogy teljesül-e a várható értékre az $m_0 = 500$ gr előírás? (Feltételezzük a normális eloszlást!)

Megoldás: m és σ^2 nem ismert, így t-próbát hajtunk végre.

ξ: egy dobozba töltött anyag mennyisége.

$$\begin{aligned} H_0: &\quad E\xi = 500 \\ H_1: &\quad E\xi \neq 500. \end{aligned}$$

$\varepsilon = 0,05$ szinifikancia szint,

$$\overline{\xi} = \frac{483 + ... + 486}{10} = 494.$$

$$S_n^{*2} = \frac{(483-494)^2 + ... + (486-494)^2}{9} = 64,9.$$

$$t_{\text{ számolt }} := \frac{494-500}{8,056/\sqrt{10}} = -2,355.$$

Most határozzuk meg az elfogadási tartományt.

$$P\left(t < t_{\varepsilon/2}\right) = 1 - \varepsilon/2 = 1 - 0,05/2 = 0,975$$

szabadsági fok $f = n - 1 = 9$. Táblázatból $t_{\varepsilon/2} = 2,26$

$t_{\text{számolt}} \in [-t_\varepsilon, +t_\varepsilon]$ teljesül-e?

$-2,355 \notin [-2,26; +2,26]$ tehát 95%-os biztonsággal a nullhipotézist elvetjük, azaz az automatát újra be kell állítani.

Kétmintás t-próba

Legyen a $\xi_1, \xi_2, \ldots, \xi_k$ az (m_ξ, σ_ξ^2)-paraméterű normális eloszlású minta, és az $\eta_1, \eta_2, \ldots, \eta_l$ pedig (m_η, σ_η^2)-paraméterű normális eloszlású minta. A két minta egymástól független. A σ_ξ^2 és σ_η^2 ismeretlen, és m_ξ, m_η ismeretlen. De tudjuk azt, hogy $\sigma_\xi^2 = \sigma_\eta^2$, Döntsünk a következő hipotézisekről:

$$\begin{aligned} H_0: &\quad E\xi = E\eta, \quad (\text{ azaz } m_\xi = m_\eta) \\ H_1: &\quad E\xi \neq E\eta \quad (\text{ azaz } m_\xi \neq m_\eta). \end{aligned}$$

Ha a H_0 nullhipotézis igaz, akkor az

$$t := \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{((k-1)S_{\xi,k}^{*2} + (l-1)S_{\eta,l}^{*2}}{k+l-2}\left(\frac{1}{k} + \frac{1}{l}\right)}}$$

statisztikai függvény t-eloszlású lesz $f = k + l - 2$ szabadsági fokkal. Adott ε szignifikancia szint esetén az eljárást az egymintás t-próbánál leírtak szerint folytatjuk.

Példa: Vizsgáljuk meg, hogy egy új készítési eljárás növeli-e a beton normális eloszlású törőszilárdságát! Az egyik és a másik eljárással készített próbakockák törőszilárdságai:

I. eljárás: 300, 301, 303, 288, 294, 296.

II. eljárás: 305, 317, 308, 300, 314, 316.

Megoldás:

ξ egy az első eljárással készült kocka törőszilárdsága. ξ-re egy $k = 6$ elemű mintát vettünk.

η egy a második eljárással készült kocka törőszilárdsága. η-ra egy $l = 6$ elemű mintát vettünk.

m_ξ, m_η, σ_ξ és σ_η ismeretlen.

Kétmintás t-próbát végzünk a következő hipotézisekkel:

$$H_0 : m_\xi = m_\eta$$
$$H_1 : m_\xi \neq m_\eta.$$

Válasszunk $\varepsilon = 0,05$-t!

$$\overline{\xi} = \frac{300 + 301 + ... + 296}{6} = 297,$$

$$S^{*2}_{\xi,k} = \frac{(300 - 297)^2 + (301 - 297)^2 + ... + (296 - 297)^2}{5} = 30,4.$$

Hasonlóan

$$\overline{\eta} = \frac{305 + 317 + ... + 316}{6} = 310,$$

$$S^{*2}_{\eta,l} = \frac{(305 - 310)^2 + (317 - 310)^2 + ... + (316 - 310)^2}{5} = 46.$$

A szórások egyenlőségét majd F-próbával ellenőrizzük a következő példában.

$$t_{\text{ számolt }} = \frac{297 - 310}{\sqrt{\frac{5\cdot 30,4 + 5\cdot 46}{10}\left(\frac{1}{6} + \frac{1}{6}\right)}} = -3,33.$$

A t-eloszlás táblázatából kikeressük $t_{\varepsilon/2}$ értékét $f = 10$ szabadsági foknál figyelembe véve

$$P\left(t < t_{\varepsilon/2}\right) = 1 - \varepsilon/2 = 1 - 0,05/2 = 0,975$$

egyenlőséget. Tehát $t_{\varepsilon/2} = 2.23$. Az elfogadási tartomány $[-2,23; +2,23]$. $-3,3 \notin [-2,23; +2,23]$ ezért H_0 nullhipotézist 95%-os biztonsági szinten elvetjük. Az új eljárás javít a törőszilárdságon. (De célszerűbb nagyobb elemszámú mintát venni.)

F-próba

Tegyük fel, hogy a $\xi_1, \xi_2, \ldots, \xi_k$ (m_ξ, σ_ξ^2)-paraméterű normális eloszlású, és az $\eta_1, \eta_2, \ldots, \eta_l$ pedig (m_η, σ_η^2)-paraméterű normális eloszlású valószínűségi változók függetlenek, ahol σ_ξ^2, σ_η^2, m_ξ és m_η is ismeretlenek. Legyen

$$H_0 : D_\xi^2 = D^2\eta \qquad (\text{ azaz } \sigma_\xi^2 = \sigma_\eta^2),$$
$$H_1 : D_\xi^2 \neq D^2\eta \qquad (\text{ azaz } \sigma_\xi^2 \neq \sigma_\eta^2).$$

Az
$$F := \frac{S_{\eta,l}^{*2}}{S_{\xi,k}^{*2}}$$
statisztika, ha a H_0 nullhipotézis igaz, akkor F-eloszlású $f_1 = l-1,\ f_2 = k-1$ szabadsági fokokkal. Legyen $1 > \varepsilon > 0$ adott. Válasszunk olyan $0 < c_1 < c_2$ számokat úgy, hogy $P(F < c_1) = P(F > c_2) = \varepsilon/2$. Ekkor vehetjük elfogadási tartománynak a $[c_1, c_2]$ intervallumot, kritikus tartománynak pedig ennek a komplementerét. c_2 helyett F_ϵ-t fogunk írni.

Megjegyzés: Az F-próba táblázatában csak 1-nél nagyobb számok szerepelnek ezért a statisztikát úgy választjuk, hogy a $\frac{S_{\eta,l}^{*2}}{S_{\xi,k}^{*2}}$, $\frac{S_{\eta,k}^{*2}}{S_{\xi,l}^{*2}}$ hányadosok közül a nagyobbikat vesszük figyelembe, s ekkor f_1 a számláló f_2 a nevező szabadsági foka lesz

Példa: Az előző példában ellenőrizzük a szórások egyenlőségét. $\varepsilon = 0,1$-el fogunk számolni, mert a táblázatunk csak a 90%-os biztonsági szinthez készült. Tudjuk, hogy $S_{\xi,k}^{*2} = 30,4$ és $S_{\eta,l}^{*2} = 46$.

Hipotéziseink:
$$\begin{aligned} H_0 :& \sigma_\xi^2 = \sigma_\eta^2, \\ H_1 :& \sigma_\xi^2 \neq \sigma_\eta^2. \end{aligned}$$

$$F_{\text{számolt}} = \frac{46}{30,4} = 1,5,$$

Az F-eloszlás táblázatából $f_1 = l-1 = 5$ és $f_2 = k-1 = 5$ szabadsági fokoknál keressük ki F_ε értékét, $F_\varepsilon = 3,45$. $1,5 = F_{\text{számolt}} < F_\varepsilon = 3,45$, így a nullhipotézist elfogadjuk 90%-os biztonsági szinten, tehát a szórások egyenlőségét elfogadjuk.

Jogosan vetődik fel a kérdés, mit lehet csinálni akkor, ha az F-próba elveti a szórások egyenlőségét? A válasz a következő próbában van.

Welch próba

Legyen a $\xi_1, \xi_2, \ldots, \xi_k$ az (m_ξ, σ_ξ^2)-paraméterű normális eloszlású minta, és az $\eta_1, \eta_2, \ldots, \eta_l$ pedig (m_η, σ_η^2)-paraméterű normális eloszlású minta. A két minta egymástól független. A σ_ξ^2, σ_η^2, m_ξ és m_η ismeretlen. De tudjuk azt, hogy $\sigma_\xi^2 \neq \sigma_\eta^2$, Döntsünk a következő hipotézisekről:
$$\begin{aligned} H_0 : &\quad E\xi = E\eta, \quad (\text{ azaz } m_\xi = m_\eta) \\ H_1 : &\quad E\xi \neq E\eta \quad (\text{ azaz } m_\xi \neq m_\eta). \end{aligned}$$

Ha a H_0 nullhipotézis igaz, akkor az

$$t := \frac{\overline{\xi} - \overline{\eta}}{\sqrt{\frac{S_{\xi,k}^2}{k} + \frac{S_{\eta,l}^2}{l}}}$$

statisztika t eloszlású lesz. A szabadsági fokra pedig teljesül a következő:

$$\frac{1}{f} = \frac{1}{k-1}\left(\frac{\frac{S_{\xi,k}^2}{k}}{\frac{S_{\xi,k}^2}{k} + \frac{S_{\eta,l}^2}{l}}\right)^2 + \frac{1}{l-1}\left(\frac{\frac{S_{\eta,l}^2}{l}}{\frac{S_{\xi,k}^2}{k} + \frac{S_{\eta,l}^2}{l}}\right)^2.$$

Adott ε szignifikancia szint esetén az eljárást az egymintás t-próbánál leírtak szerint folytatjuk.

Példa: Kúpgörgős csapágy belső gyűrűjének kúpszögét mérjük egy hitelesített A műszerrel és egy hitelesítendő B műszeren. Az A műszeren végzett mérés eredménye normális eloszlású, a B műszeren végzett mérés eredménye , szintén normális eloszlású. A mérési eredmények:

$$\begin{aligned} A:&\quad k = 100, \quad \overline{\xi} = 0,625, \quad S_{\xi,k} = 0,754,\\ B:&\quad l = 100, \quad \overline{\eta} = 0,471, \quad S_{\eta,l} = 1,269. \end{aligned}$$

Hitelesnek tekinthető-e a B műszer? (Azaz az egész sokaságban a várható értékek egyenlők-e?)

Kétmintás t-próbával szeretnénk dönteni, de előbb ellenőrizzük a szórások egyenlőségét F-próbával!

Hipotéziseink:

$$\begin{aligned} H_0 :&\sigma_\xi^2 = \sigma_\eta^2,\\ H_1 :&\sigma_\xi^2 \neq \sigma_\eta^2. \end{aligned}$$

Az adatokból $S_{\xi,k}^{*2} = \frac{100 S_{\xi,k}^2}{99} = 0,57426$, és $S_{\eta,l}^{*2} = \frac{100 S_{\eta,l}^2}{99} = 1,6266$.

$$F_{\text{számolt}} = \frac{1,6266}{0,57426} = 2,83,$$

Az F-eloszlás táblázatából $f_1 = l-1 = 99$ és $f_2 = k-1 = 99$ szabadsági fokoknál keressük ki F_ε értékét, $F_\varepsilon = 1,39$. $2,83 = F_{\text{számolt}} > F_\varepsilon = 1,39$, így a nullhipotézist elvetjük 90%-os biztonsági szinten, tehát a szórások nem egyenlőek.

Így t-próba helyett Welch-próbát alkalmazunk. Hipotéziseink:

$$\begin{aligned} H_0 :&m_\xi = m_\eta\\ H_1 :&m_\xi \neq m_\eta. \end{aligned}$$

Ekkor

$$t_{\text{ számolt}} := \frac{0,625 - 0,471}{\sqrt{\frac{0,5685}{100} + \frac{1,6104}{100}}} = 1,043$$

$$\frac{1}{f} = \frac{1}{99}\left(\frac{\frac{0,5685}{100}}{\frac{0,5685}{100} + \frac{1,6104}{100}}\right)^2 + \frac{1}{99}\left(\frac{\frac{1.6104}{100}}{\frac{0,5685}{100} + \frac{1,6104}{100}}\right)^2 = 0,0062053,$$

$f = 161$. $\varepsilon = 0,05$-höz, $t_{\varepsilon/2} = 1,96$. Így az elfogadási tartomány $[-1,96; +1,96]$. $1,043 \in [-1,96; +1,96]$, tehát a várható értékek egyenlőségét 95%-os biztonsági szinten elfogadjuk.

3.5. χ^2-próbák

χ^2-illeszkedésvizsgálat

A minta tekinthető-e egy adott eloszlásból származónak?

$$\begin{aligned} H_0: &\quad \text{a minta az adott eloszlásból származik,} \\ H_1: &\quad \text{a minta nem az adott eloszlásból származik.} \end{aligned}$$

Pontosítva:

$A_1, A_2, ..., A_r$ teljes eseményrendszert alkotnak.

$$H_0: \quad P(A_1) = p_1, ..., P(A_r) = p_r.$$

Az n megfigyelés során A_i bekövetkezik k_i-szer (fontos: $\sum_{i=1}^{r} k_i = n$). Ha H_0 igaz, akkor

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - np_i)^2}{np_i}$$

χ^2 eloszlású $f = r - 1$ szabadsági fokkal.

χ^2_ε-t az $f = r - 1$ paraméterű χ^2 táblázatból keressük ki.

$$P(\chi^2 < \chi^2_\varepsilon) = 1 - \varepsilon$$

képletet alkalmazva. Azaz a próba egyoldali, így az elfogadási tartomány $[0, \chi^2_\varepsilon]$, a kritikus tartomány $(\chi^2_\varepsilon, +\infty)$.

Megjegyzés:

1. A fenti esetben tiszta illeszkedésvizsgálatról beszélünk. Előfordulhat, hogy néhány paramétert s darabot becsülnünk kell, ekkor becsléses illeszkedésvizsgálatról beszélünk és $f = r - 1 - s$ szabadsági fokú χ^2-eloszlás táblázatát használjuk.

2. A közelítés megfelelő, ha $np_i \geq 10$ (lehet, hogy n nagy kell legyen).

3. Folytonos eloszlások esetén az adatokat osztályokba soroljuk és $p_i = P(\xi$ az i-edik osztályba tartozik $) = P(a_i \leq \xi < b_i)$ valószínűségeket használjuk.

Példa: Tekinthető-e szabályosnak az a játékkocka, amelyet $n = 1200$-szor feldobva az egyes számok gyakoriságára az alábbi eredményeket kaptuk? 1-est 195-ször, 2-est 210-szer, 3-ast 190-szer 4-est 204-szer, 5-öst 205-ször, 6-ost 196-szor kaptunk.

Ha a kocka szabályos, akkor bármely szám dobása egyenlő valószínűségű. ξ a dobott szám (A_i : i-t dobunk).

$$H_0: \quad P(A_i) = P(\xi = i) = p_i = 1/6 \quad (i = 1, 2, ..., 6).$$
$$H_1: \quad \text{nem igaz a fenti eloszlás, azaz a kocka nem szabályos.}$$

$np_i = 1200 \cdot \frac{1}{6} = 200 \geq 10$, tehát a közelítés megfelelő lesz. A számolást táblázattal megkönnyíthetjük:

dobott szám	gyakoriság (k_i)	np_i	$(k_i - np_i)^2/np_i$
1-es	195	200	25/200
2-es	210	200	100/200
3-as	190	200	100/200
4-es	204	200	16/200
5-ös	205	200	25/200
6-os	196	200	16/200
		összeg:	$\chi^2_{\text{számolt}}$ =282/200=1,41

Tiszta illeszkedésvizsgálatról van szó, $f = 5$ szabadsági foknál $\varepsilon = 0,05$-höz $\chi^2_\varepsilon = 11,1$.

Mivel $1,41 = \chi^2_{\text{számolt}} < \chi^2_\varepsilon = 11,1$, így H_0-t 95%-os biztonsági szinten elfogadjuk, azaz a kocka szabályos.

χ^2-függetlenségvizsgálat

Két valószínűségi változó függetlennek tekinthető-e?

$$H_0: \quad \text{függetlenség van,}$$
$$H_1: \quad \text{nincs függetlenség.}$$

Az első valószínűségi változó értékkei az $A_1, ..., A_r$ teljes eseményrendszerbe esnek, a második valószínűségi változó értékei az $B_1, ..., B_s$ teljes eseményrendszerbe esnek. Az $A_i \cap B_j$ eseménybe $k_{i,j}$ darab minta elem esik.

Ismertek a következő valószínűségek (tiszta függetlenségvizsgálat): $p_1 = P(A_1), ..., p_r = P(A_r)$ illetve $q_1 = P(B_1), ..., q_s = P(B_s)$. Ha H_0 igaz, akkor

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(k_{ij} - np_i q_j)^2}{np_i q_j}$$

χ^2 eloszlású $f = rs - 1$ szabadsági fokkal.

Adott ε-hoz χ^2_ε-t táblázatból keressük ki.

$$P(\chi^2 < \chi^2_\varepsilon) = 1 - \varepsilon$$

képletet alkalmazva. Így az elfogadási tartomány $[0, \chi^2_\varepsilon]$, a kritikus tartomány $(\chi^2_\varepsilon, +\infty)$.

A gyakorlatban a p_i (i=1,...,r) és q_j (j=1,...,s) valószínűségek ismeretlenek, becsülni kell őket. Ekkor becsléses függetlenségvizsgálatról beszélünk.

Kontingencia táblázat:

$\xi \backslash \eta$	B_1	$\ldots$	B_j	$\ldots$	B_s	összesen
A_1	k_{11}	$\ldots$	k_{1j}	$\ldots$	k_{1s}	$f_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_i	k_{i1}	$\ldots$	k_{ij}	$\ldots$	k_{is}	$f_{i.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
A_r	k_{r1}	$\ldots$	k_{rj}	$\ldots$	k_{rs}	$f_{r.}$
összesen	$f_{.1}$	$\ldots$	$f_{.j}$	$\ldots$	$f_{.s}$	n

Ha H_0 igaz, akkor

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(k_{ij} - \frac{f_{i.} f_{.j}}{n})^2}{\frac{f_{i.} f_{.j}}{n}}$$

χ^2 eloszlású $f = (r-1)(s-1)$ szabadsági fokkal.

Példa: Csapággyűrűknél fontos minőségi jellemző a külső és belső átmérő (ξ és η). Az átmérő nagysága alapján az elkészült gyűrűket három kategóriába soroljuk: jó, javítható, selejtes. Találomra kiválasztunk $n = 200$ db-ot annak ellenőrzésére, hogy a külső és a belső átmérő független-e egymástól. Döntsünk a függetlenségről a következő kontingencia táblázat alapján!

belső átmérő \ külső átmérő	jó	javítható	selejtes	összeg
jó	169	8	1	178
javítható	9	4	1	14
selejtes	1	3	4	8
összeg	179	15	6	200

$$\begin{aligned} H_0: &\quad \text{függetlenség van,} \\ H_1: &\quad \text{nincs függetlenség.} \end{aligned}$$

$$\begin{aligned} \chi^2_{\text{számolt}} =& \frac{\left(169 - \frac{178\cdot179}{200}\right)^2}{\frac{178\cdot179}{200}} + \frac{\left(8 - \frac{178\cdot15}{200}\right)^2}{\frac{178\cdot15}{200}} + \frac{\left(1 - \frac{178\cdot6}{200}\right)^2}{\frac{178\cdot6}{200}} \\ &+ \frac{\left(9 - \frac{14\cdot179}{200}\right)^2}{\frac{14\cdot179}{200}} + \frac{\left(4 - \frac{14\cdot15}{200}\right)^2}{\frac{14\cdot15}{200}} + \frac{\left(1 - \frac{14\cdot6}{200}\right)^2}{\frac{14\cdot6}{200}} \\ &+ \frac{\left(1 - \frac{8\cdot179}{200}\right)^2}{\frac{8\cdot179}{200}} + \frac{\left(3 - \frac{8\cdot15}{200}\right)^2}{\frac{8\cdot15}{200}} + \frac{\left(4 - \frac{8\cdot6}{200}\right)^2}{\frac{8\cdot6}{200}} = 90,15. \end{aligned}$$

Függőlegesen $r = 3$ és vizszintesen $s = 3$ osztály van. Szabadsági fok $f = (r-1)(s-1) = 2 \cdot 2 = 4$. $\varepsilon = 0,05$-höz $\chi^2_\varepsilon = 9,49$ táblázatból. Elfogadási tartomány $[0; 9,49]$. $9,49 < 90,15$, tehát a nullhipotézist elvetjük 95%-os biztonsággal, azaz a két tulajdonság nem független.

Irodalomjegyzék

[1.] Solt György, Valószínűségszámítás, Műszaki Könyvkiadó, Budapest.

[2.] Lukács Ottó, Matematikai statisztika, Műszaki Könyvkiadó, Budapest.

[3.] Nagy Márta, Sztrik János, Tarr László, Valószínűségszámítás és matematikai statisztika feladatgyűjtemény, Kossuth Egyetemi Kiadó, Debrecen 2003.

[4.] Gát György, Valószínűségszámítás, (http://zeus.nyf.hu/˜gatgy/)

[5.] Baran Sándor, Valószínűségszámitás és matematikai statisztika feladatok (http://www.inf.unideb.hu/˜barans/)