Alkalmazott Statisztika

Előadás 2

Dr. Barta Attila[1], Dr. Pecsora Sándor[2]

PTI BSc

Áttekintés

A valószínűségszámítás témája

A véletlen tömegjelenségekre vonatkozó törvényszerűségek megállapítása. Véletlen jelenség az, aminek a kimenetelét a tekintetbevett (rendelkezésre álló) feltételek nem határozzák meg egyértelműen. Tömegjelenségen pedig olyan jelenséget értünk, amely nagy számban megy végbe egyszerre(pl. atomi bomlás), vagy sokszor megismételhető (pl. szerencsejátékok). A levonható törvényszerűségek statisztikai jellegűek, azaz nagy számú végrehajtás során átlagosan érvényes törvények.

Dr. Fazekas István: *Valószínűségszámítás és statisztika*

Példa

Az 1650-es évek Franciaországában Chevalier de Mérének szerencsejáték-problémája volt. Nem arról volt szó, hogy szerencsejátékot játszott túl sokat (bár megtette), de tudni akarta, hogy a két játék közül melyikkel van a nagyobb esély nyerni:
1. játék: Dobj egy szabályos kockát legfeljebb négyszer, és nyersz, ha hatost kapsz.
2. játék: Dobj két szabályos kockát legfeljebb huszonnégyszer, és nyersz, ha dupla-hatost kapsz.

Melyik volt a jobb fogadás?

Példa

Az 1650-es évek Franciaországában Chevalier de Mérének szerencsejáték-problémája volt. Nem arról volt szó, hogy szerencsejátékot játszott túl sokat (bár megtette), de tudni akarta, hogy a két játék közül melyikkel van a nagyobb esély nyerni:
1. játék: Dobj egy szabályos kockát legfeljebb négyszer, és nyersz, ha hatost kapsz.
2. játék: Dobj két szabályos kockát legfeljebb huszonnégyszer, és nyersz, ha dupla-hatost kapsz.

Melyik volt a jobb fogadás?

Megoldás *Blaise Pascal* és *Pierre de Fermat* által.

Definíció

Vegyünk egy rögzített kísérletet K. A kísérlet azon eredményei, amelyeket nem tudunk felosztani a kisebb részeket elemi eseményeknek nevezzük. Az elemi eseményeket általában ω-val (görög omega) jelölik.
Az összes elemi esemény halmazát mintatérnek nevezzük (valószínűségi tér).
Jelölés: Ω (görög Omega)

Példák

1. Érmedobás. $\Omega = \{Fej, Írás\}$
2. Két érme felobása. $\Omega = \{FF, FÍ, ÍF, ÍÍ\}$
3. Kockadobás. $\Omega = 1, 2, 3, 4, 5, 6$
4. Egy pont válsztása a (0, 1) intervallumról. Ekkor $\Omega = (0, 1)$

Events

Az Ω részhalmazait eseményeknek nevezzük.

Példa

1. Dobjunk fel két érmét. Legyen A az az esemény, hogy legalább egy Fejet dobunk. Ekkor $A = \{FF, FÍ, ÍF\}$.
2. Dobjunk fel egy érmét. Legyen A az az esemény, hogy páros számot dobunk. Ekkor $A = \{2, 4, 6\}$.

- Ω a biztos esemény. $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$
- $\emptyset$ a lehetetlen esemény. $\emptyset = \{\ \}$

Műveletek eseményekkel

- A vagy $B = A \cup B = A + B$
- A és $B = A \cap B = A \cdot B$
- A ellentetje $=$ nem $A = \overline{A}$
- A és B különbsége $= A \setminus B$
- pontosan egyik teljesül A és B között $=$ A és B szimmetrikus különbsége $= \mathrm{A} \circ \mathrm{B}$
- A-ból B következik $= A \subseteq B$

Relatív gyakoriság

Definíció

Ismételje meg a kísérletet n alkalommal. Az A esemény k_A alkalommal fordul elő. Ekkor

$$\frac{k_A}{n}$$

A relatív gyakoriságának nevezzük.

Tétel

A relatív gyakoriság stabil, ha n nagy:

$$\frac{k_A}{n} \sim P(A),$$

ahol $P(A)$ egy fix szám.

Kolmogorov valószínűségi axiómái

Axiómák

1. Nem-negatív: $P(A) \geq 0$, bármely A eseményre.
2. Normált: $P(\Omega) = 1$.
3. Megszámlálhatóan additív:

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots,$$

ha $A_1, A_2, \ldots$ páronként kizáróak.

Megjegyzés: $(\Omega, \mathscr{A}, P)$ a Kolmogorov féle valószínűségi mező, ahol
$\Omega \neq \emptyset$ eseménytér,
$\mathscr{A}$ események halmaza (pl. kocka: $\mathscr{A} = \{\{1\}, \{1,2,3\}, \{2,4,6\}, ..., \{1,2,3,4,5,6\}\}$,
P valószínűség.

A valószínűség tulajdonságai

Végesen additív: ha $A_1, A_2, \ldots, A_n$ egymást páronként kizáróak, ekkor

$$P(A_1 + A_2 + \cdots + A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

Tipp: alkalmazza a matematikai indukciót és az additív axiómát.

$$P(\emptyset) = 0.$$

Tipp: $\Omega = \Omega + \emptyset$

$$P(A - B) = P(A) - P(A \cdot B).$$

Tipp: $A = A \cdot B + (A - B)$

$$P(A) = 1 - P(\overline{A}).$$

Tipp: $\Omega = A + \overline{A}$

Véges valószínűségi mező

Definíció

A $p_1,\dots,p_N$ véges (valószínűségi) eloszlásnak nevezzük, ha a p_i számok nem-negatívak és

$$\sum_{i=1}^{N} p_i = 1.$$

Definíció

Bármely N elemű valószínűségi tér a következőképpen írható le.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\},$$

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

bármely A eseményre, ahol $p_1,\dots,p_N$ egy adott eloszlás.

Definíció

Feltételezzük, hogy egy kísérlet lehetséges kimenetelei egyenlően valószínűek. Ekkor $p_1 = p_2 = \cdots = p_N$. Mivel $\sum_{i=1}^{N} pi = 1$, ezért

$$p_1 = p_2 = \cdots = p_N = \frac{1}{N},$$

és

$$P(A) = \sum_{\omega_i \in A} p_i = \sum_{\omega_i \in A} \frac{1}{N}.$$

Így megfigyelhetjük a valószínűség kiszámítására a klasszikus szabályt

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

Ez a klasszikus szabély, amely sok, de nem minden esetben alkalmazható.

Előadás kérdés

Tíz ember, 5 nő és 5 férfi ül egy kerek asztal körül. Határozza meg annak a valószínűségét, hogy sem két nő, sem két férfi nem ül egymás mellett!

Előadás kérdés

Tíz ember, 5 nő és 5 férfi ül egy kerek asztal körül. Határozza meg annak a valószínűségét, hogy sem két nő, sem két férfi nem ül egymás mellett!

$$P(A) = \frac{4! \cdot 5!}{9!}$$

Előadás kérdés

Egy kártyapakliból három lapot osztanak ki. Határozza meg annak valószínűségét, hogy nincs közöttük pikk.

Klasszikus valószínűségi mező

Előadás kérdés

Egy kártyapakliból három lapot osztanak ki. Határozza meg annak valószínűségét, hogy nincs közöttük pikk.

$$P(A) = \frac{C_{39}^3}{C_{52}^3}$$

Klasszikus valószínűségi mező

Előadás kérdés

Egy kártyapakliból három lapot osztanak ki. Határozza meg annak valószínűségét, hogy nincs közöttük pikk.

$$P(A) = \frac{C_{39}^3}{C_{52}^3}$$

Előadás kérdés

Határozza meg annak valószínűségét, hogy az 5-ös lottón 90-ből legalább három nyerőszámot találunk.

Előadás kérdés

Határozza meg annak valószínűségét, hogy az 5-ös lottón 90-ből legalább három nyerőszámot találunk.

$$P(A) = \frac{C_5^3 C_{85}^2 + C_5^4 C_{85}^1 + C_5^5 C_{85}^0}{C_{95}^5}$$

Diszkrét eloszlás

Definíció

$p_1, p_2, \ldots$ diszkrét (valószínűségi) eloszlás, ha a p_i számok nem-negatívak és

$$\sum_{i=1}^{\infty} p_i = 1.$$

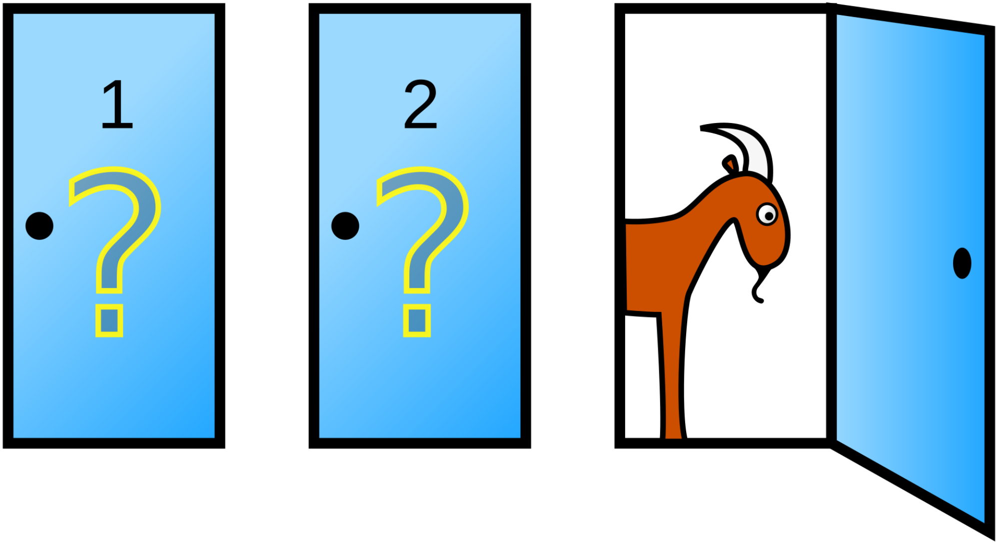
A valószínűség geometriai kiszámítási módja

Legyen G egy részhalmaza a következőnek: $\mathbb{R}^n (n = 1,2,3)$. Válasszunk egy véletlenszerű pontot G-ből. Legyen $A \subseteq G$. Ekkor annak a valószínűsége, hogy a kiválasztott pont A része

$$P(A) = \lambda(A)/\lambda(G),$$

ahol λ a hossz, terület vagy térfogat, ha egyenesen, síkon vagy térben vagyunk (feltételezzük, hogy $0 < \lambda(G) < \infty$).

Monty Hall probléma



Definíció

Adott két esemény A és B $(P(B) > 0)$. A feltételes valószínűség $P(A \mid B)$ megmutatja A valószínűségét, ha B bekövetkezett. Ekkor

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

Teljes valószínűség tétele

Definíció

Események egy $A_1, A_2, \ldots$ sorozatát teljes eseményrendszernek nevezzük, ha egymást páronként kizárják és összegük az egész eseménytér.

$$P(A_1) + P(A_2) + \cdots = 1.$$

$$A_i \cap A_j = \emptyset, \quad i \neq j$$

Tétel (Teljes valószínűség)

Legyen $B_1, B_2, \ldots$ teljes eseményrendszer. Tegyük fel, hogy $P(B_i) > 0$, bármely i esetén. Ekkor bármely A esemény esetén

$$P(A) = P(A|B1)P(B1) + P(A|B2)P(B2) + \cdots.$$

Bayes formula

Legyenek A és B pozitív valószínűségűek. Ekkor

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

Bayes tétel

Legyen A esemény, $B_1, B_2, \ldots$ egy teljes eseményrendszer, $P(A) > 0, P(B_i) > 0, i = 1, 2, \ldots$. Ekkor bármely i esetén

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}.$$

Példa

Egy bizonyos típusú COVID gyorstesztnél betegség esetén a pozitív eredmény valószínűsége 85% (true positive), míg egészségesen ugyanez 2% (false negative).
(a) Ha a lakosság 5%-a fertőzött, mi annak a valószínűsége, hogy egy véletlenszerűen kiválasztott ember pozitív tesztet ad?

Példa

Egy bizonyos típusú COVID gyorstesztnél betegség esetén a pozitív eredmény valószínűsége 85% (true positive), míg egészségesen ugyanez 2% (false negative).

(a) Ha a lakosság 5%-a fertőzött, mi annak a valószínűsége, hogy egy véletlenszerűen kiválasztott ember pozitív tesztet ad?

$P(\text{pozitív}) =$

$P(\text{pozitív} \mid \text{fertőzött})P(\text{fertőzött}) + P(\text{pozitív} \mid \text{egészséges})P(\text{egészséges}) =$

$0.85 \cdot 0.05 + 0.02 \cdot 0.95 = 0.0615 \approx 6.15\%$

Példa

Egy bizonyos típusú COVID gyorstesztnél betegség esetén a pozitív eredmény valószínűsége 85% (true positive), míg egészségesen ugyanez 2% (false negative).
(a) Ha a lakosság 5%-a fertőzött, mi annak a valószínűsége, hogy egy véletlenszerűen kiválasztott ember pozitív tesztet ad?
$P(\text{pozitív}) =$
$P(\text{pozitív} \mid \text{fertőzött})P(\text{fertőzött}) + P(\text{pozitív} \mid \text{egészséges})P(\text{egészséges}) =$
$0.85 \cdot 0.05 + 0.02 \cdot 0.95 = 0.0615 \approx 6.15\%$
(b) Mi annak a valószínsége, hogy negatív teszt esetén valaki fertőzött?

Példa

Egy bizonyos típusú COVID gyorstesztnél betegség esetén a pozitív eredmény valószínűsége 85% (true positive), míg egészségesen ugyanez 2% (false negative).

(a) Ha a lakosság 5%-a fertőzött, mi annak a valószínűsége, hogy egy véletlenszerűen kiválasztott ember pozitív tesztet ad?

$P(\text{pozitív}) =$

$P(\text{pozitív} \mid \text{fertőzött})P(\text{fertőzött}) + P(\text{pozitív} \mid \text{egészséges})P(\text{egészséges}) =$

$0.85 \cdot 0.05 + 0.02 \cdot 0.95 = 0.0615 \approx 6.15\%$

(b) Mi annak a valószínsége, hogy negatív teszt esetén valaki fertőzött?

$P(\text{fertőzött} \mid \text{negatív}) = \frac{P(\text{negatív} \mid \text{fertőzött})P(\text{fertőzött})}{P(\text{negatív})} = 0.008 \approx 0.8\%$

Monty Hall

A: a játék elején a kiválaszott ajtó mögött autó van
K: Monty egy olyan ajtót mutat ami mögött kecske van

$$P(A \mid K) = \frac{P(K \mid A) \cdot P(A)}{P(K)} = \frac{P(K \mid A) \cdot P(A)}{P(K \mid A) \cdot P(A) + P(K \mid \overline{A}) \cdot P(\overline{A})}$$

- Ha Monty szándékosan választja a kecskés ajtót:

$$P(A \mid K) = \frac{1}{3}$$

- Ha Monty véletlenül választja a kecskés ajtót:

$$P(A \mid K) = \frac{1}{2}$$

Két esemény függetlensége

Definíció

Azt mondjuk, hogy az A esemény független a B eseménytől, ha B bekövetkezése nincs hatással A valószínűségére. Azaz

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

Ez a meghatározás azonban nem szimmetrikus, és feltéteneznünk kell hogy $P(B) > 0$. A következő definícióban az A és a B szerepe szimmetrikus.

Definíció

Azt mondjuk, hogy A és B független, ha

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

Több esemény függetlensége

Definíció

Azt mondjuk, hogy az $A_1, A_2, \ldots$ események páronként függetlenek, ha közülük bármely kettő független, azaz

$$P(A_i A_j) = P(A_i)P(A_j), i \neq j.$$

Definíció

Azt mondjuk, hogy az $A_1, A_2, \ldots, A_n$ események (teljesen) függetlenek, ha bármely $k = 1, 2, \ldots, n$ eseményszámra és bármely $i_1, \ldots, i_k \in \{1, \ldots, n\}$ indexű eseményekre

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

Páronkénti függetlenség és függetlenség

Állítás A függetlenség egyúttal páronkénti függetlenséget is jelent, de fordítva nem igaz az állítás.

Példa. Válasszunk egy pontot véletlenszerűen az egységnégyzetből. Legyenek A, B és C azok az események, amelyeknél a pont az alábbi részhalmazokból lett kiválasztva.

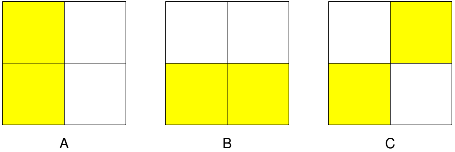


Fig. 2: A, B és C események.

Ekkor $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(AB) = P(AC) = P(BC) = \frac{1}{4}$, $P(ABC) = \frac{1}{4}$. Páronként függetlenek, de nem teljesen függetlenek.

Chevalier de Méré

- 1.játék:
  $P(\text{legalább 1 hatos 4 dobás esetén}) = 1 - P(\text{nincs hatos 4 dobás esetén}) =$
  $1 - P(D_1 \cap D_2 \cap D_3 \cap D_4) = 1 - P(D_1) \cdot P(D_2) \cdot P(D_3) \cdot P(D_4) = 1 - P(D_1)^4 =$
  $1 - \left(\frac{5}{6}\right)^4 = 0.5177$
- 2. játék:
  $P(\text{legalább egy dupla-hatos, 2 kocka 24-szer}) =$
  $1 - P(\text{nincs dupla-hatos, 2 kocka 24-szer}) = 1 - P(D_1 \cap D_2 \cap D_3 \cap \cdots \cap D_{24}) =$
  $1 - P(D_1) \cdot P(D_2) \cdot P(D_3) \cdot \ldots \cdot P(D_{24}) = 1 - P(D_1)^{24} = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914$

D_i: i-edik dobás sikertelen a játék szempontjából