Matematikai statisztika

Dr. Sándor Csaba

egyetemi docens, BME Sztochasztika Tanszék

Budapest, 2015 tavasz

1. Valószínűségszámítási alapok

A valószinűségszámítás tárgya véletlen tömegjelenségek vizsálata.
Véletlennek mondunk egy jelenséget, ha a létrejöttét befolyásoló összes tényezőt nem ismerjük, vagy ha a befolyásoló tényezők nagy száma miatt képtelenek vagyunk azok mindegyikére tekintettel lennünk.
Tömegjelenségről beszélünk, ha a jelenség bizonyos megadott feltételek mellett sokszor (elméletileg végtelen sokszor) megfigyelhető.
Példák véletlen tömegjelenségre:

1. Egy szabályos kockával dobunk. A véletlen a dobás eredménye.
2. A Duna vízállása Vácnál.
3. Egy óra alatt hány autó megy át a Petőfi hídon
4. Piros, sárga vagy zöld most a lámpa a Gellért téren a Gellért hegy felé menő autósoknak.

A valószínűségszámítás olyan véletlen tömegjelenségekkel foglalkozik, ahol a lehetséges kimenetelek számok. Az ilyen ilyen vélelen jelenséget valószínűségi változónak hívjuk.
A véletlen jelenség megismerése történhet absztrakt, elméleti módon és kisérleteken keresztül.
A véletlen jelenség lehetséges eredményei az <u>események</u>.
Azok az események, amik már nem bonthatók <u>további</u> eseményekre: <u>elemi események</u>.
Az elemi események összessége: <u>eseménytér</u> (Ω)
Matematikailag egy esemény az eseménytér egy részhalmaza.
Példák:

1. Kockadobás:
   - 1.1 Elemi események: 1,2,3,4,5,6
   - 1.2 Eseménytér: $\{1, 2, 3, 4, 5, 6\}$
   - 1.3 Nem elemi események: páros számot dobtam; háromnál nagyobbat dobtam
2. Duna vízállása Vácnál:
   - 2.1 Eseménytér: $\mathbb{R}^+$
   - 2.2 Egy esemény: $[2, 10]$ intervallumba esik

Hány autó megy át a Petőfi hídon:

1. Elemi események: 0,1,2,3,...
2. Eseménytér: $\mathbb{N}$
3. Esemény: 1000-nél kevesebb

Ha a véletlen jelenségeknél az elemi események száma véges vagy megszámlálhatóan végtelen, akkor diszkrét valószínűségi változóról, míg ha az eseménytér egy intervallum vagy egy félegyenes vagy minden valós szám, akkor folytonos valószínűségi változóról beszélünk.

Példák:

1. Diszkrét: Kockadobás eredménye; hány autó megy át egy óra alatt a Petőfi hídon
2. Folytonos: A Duna vízállása Vácnál; RND: $[0, 1]$ intervallumban egyenletes eloszlás

1.1 Eseményekkel kapcsolatos fogalmak, műveletek

Biztos esemény: ami mindig bekövetkezik
Lehetetlen esemény: ami sohasem következik be
Legyen A és B két esemény egy véletlen jelenséggel kapcsolatban.
Komplementer esemény: Az A esemény komplementere akkor következik be, ha az A esemény nem következik be ($\overline{A}$)

Események összege: Az A és B események összege akkor következik be, ha az A és B események közül legalább az egyik bekövetkezik, azaz ha A vagy B esemény bekövetkezik. ($A + B$)

Események szorzata: Az A és B eseméyek szorzata akkor következik be, ha az A és B események mindegyike bekövetkezik, azaz ha A és B események bekövetkeznek. (AB)

Kizáró események: Az A és B események kizáró események, ha egyszerre nem következhetnek be, azaz ha $AB = \emptyset$

Példák:

1. Kockadobás: Legyen A: páros szám (azaz $A = \{2, 4, 6\}$); B: 3-nál kisebb szám (azaz $B = \{1, 2\}$); C: 3-nál nagyobb szám (azaz $C = \{4, 5, 6\}$)
   1.1 $\overline{A} = \{1, 3, 5\}$ (azaz a páratlan számok), $\overline{B} = \{3, 4, 5, 6\}$, $\overline{C} = \{1, 2, 3\}$
   1.2 $A + B = \{1, 2, 4, 6\}$, $A + C = \{2, 4, 5, 6\}$, $B + C = \{1, 2, 4, 5, 6\}$
   1.3 $AB = \{2\}$, $AC = \{4, 6\}$, $BC = \emptyset$
2. RND: Legyen A: $RND \in [0; 0,5]$ és B: $RND \in [0,3; 0,8]$
   2.1 $\overline{A} =]0,5; 1]$, $\overline{B} = [0; 0,3[\cup]0,8; 1]$
   2.2 $A + B = [0; 0,8]$, $AB = [0,3; 0,5]$, $\overline{A}B =]0,5; 0,8]$

1.2 Események valószínűsége

Egy véletlen tömegjelenség esetén egy A esemény valószínűségét a következőképpen értelmezzük: n-szer elvégezzük a kisérletet és megszámoljuk, hogy hányszor következett be az esemény: k-szor.

k: az A esemény gyakorisága
$\frac{k}{n}$: az A esemény relatív gyakorisága

Az n növelésével azt tapasztaljuk, hogy a relatív gyakoriság stabilizálódik, egy értékhez tart (azaz létezik a $\lim\limits_{n\to\infty} \dfrac{k}{n}$). Az így kapott érték az A esemény valószínűsége. Jelölés: $P(A)$

Megjegyzések:

1. Mivel $0 \leq k \leq n$, ezért $0 \leq \frac{k}{n} \leq 1$, így $0 \leq P(A) \leq 1$.
2. Ha A egy biztos esemény, akkor $k = n$, így $\frac{k}{n} = 1$, ezért $P(A) = 1$.

1. Ha A egy lehetetlen esemény, akkor $k = 0$, így $\frac{k}{n} = 0$, ezért $P(A) = 0$.
2. Ha az A esemény gyakorisága k, akkor az $\overline{A}$ eseményé $n - k$, igy az A relatív gyakorisága $\frac{k}{n}$, míg az $\overline{A}$ eseményé $\frac{n-k}{n} = 1 - \frac{k}{n}$. Innen azt kapjuk, hogy $P(\overline{A}) = 1 - P(A)$.
3. Az A eseménynél jelölje k_A azt, hogy az A esemény hányszor következett be és a B eseménynél k_B, hogy a B esemény hányszor következett be. Ekkor $k_{A+B} = k_A + k_B - k_{AB}$, ezért $\frac{k_{A+B}}{n} = \frac{k_A}{n} + \frac{k_B}{n} - \frac{k_{AB}}{n}$, ahonnan azt kapjuk, hogy $P(A + B) = P(A) + P(B) - P(AB)$
4. Ha A és B kizáró események, akkor $k_{AB} = 0$, ezért $\frac{k_{AB}}{n} = 0$, ami a $P(AB) = 0$ valószínűséget adja. Innen az előző pont alapján azt kapjuk, hogy

$$P(A + B) = P(A) + P(B)$$

Példák:

1. Szabályos kockával dobás: A: 6-ost dobok, B: páratlan számot dobok
   1.1 $P(A) = \frac{1}{6}$,
   1.2 $P(\overline{A}) = P(1,2,3,4 \text{ vagy 5-öt dobok}) = \frac{5}{6} (= 1 - P(A))$
   1.3 $P(B) = P(1, 3 \text{ vagy 5-öt dobok}) = \frac{3}{6}$
   1.4 $P(A + B) = P(1,3,5 \text{ vagy 6-ot dobok}) = \frac{4}{6}$
2. Egy szabályos pénzérmével kétszer dobok. Eseménytér: $\{FF, FI, IF, II\}$. Legyen az A az, hogy dobunk fejet, a B esemény, hogy dobunk írást.
   2.1 $P(A) = P(\text{FF, FI vagy IF}) = \frac{3}{4}$
   2.2 $P(B) = P(\text{FI, IF vagy II}) = \frac{3}{4}$
   2.3 $P(\overline{A}) = P(\text{nem dobunk fejet}) = P(II) = \frac{1}{4}$
   2.4 $P(A + B) = P(\text{dobunk fejet vagy írást}) = 1$
   2.5 $P(AB) = P(\text{dobunk fejet és írást is}) = P(\text{FI vagy IF}) = \frac{2}{4}$

1. Tekintsük a kétgyerekes családokat. Eseménytér: $\{FF, FL, LF, LL\}$. Legyen az A az, hogy van fiú a családban, a B esemény, hogy van lány.
   1.1 $P(A) = P(\text{FF, FL vagy LF}) = \frac{3}{4}$
   1.2 $P(B) = P(\text{FL, LF vagy LL}) = \frac{3}{4}$
   1.3 $P(\overline{A}) = P(\text{nincs fiú}) = P(LL) = \frac{1}{4}$
   1.4 $P(A + B) = P(\text{van vagy fiú vagy lány}) = 1$
   1.5 $P(AB) = P(\text{van fiú és lány is}) = P(\text{FL vagy LF}) = \frac{2}{4}$
2. Legyen RND a $[0, 1]$-en egyenletes eloszlású valószínűségi változó. Ekkor
   2.1 $P(RND < \frac{1}{3}) = \frac{1}{3}$
   2.2 $P(RND \in [0,3; 0,9]) = 0,6$

1.3 Feltételes valószínűség

Feladat: Határozza meg az A esemény valószínűségét, feltéve, hogy egy másik B esemény már bekövetkezett!

Tegyük fel, hogy n kisérletet végeztünk. Ekkor jelölje k_B azon kisérletek számát, amikor a B esemény bekövetkezett. Minket csak ezek a kisérletek érdekelnek. Ezek közül k_{AB} azoknak a száma, amikor az A esemény is bekövetkezik. Így most a relatív gyakoriság: $\frac{k_{AB}}{k_B}$. De

$$\frac{k_{AB}}{k_B} = \frac{\frac{k_{AB}}{n}}{\frac{k_B}{n}} \approx \frac{P(AB)}{P(B)}.$$

Ez alapján természetes a következő definíció:

Ha A és B olyan események, hogy $P(B) \neq 0$, akkor az A esemény B feltétel melletti ún. <u>feltételes valószínűségét</u> a $\frac{P(AB)}{P(B)}$ képlettel értelmezzük. Jelölés: $P(A|B)$

Példák:

1. Kockadobás: Legyen A: 6-os, B: páros szám dobása. Ekkor

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(\text{6-os és páros})}{P(\text{páros})} =$$

$$\frac{P(6-os)}{P(\text{páros})} = \frac{1/6}{3/6} = \frac{1}{3}$$

2. Kétszer dobunk egy szabályos pénzérmével. Legyen A: dobunk fejet, B: dobunk írást. Ekkor

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(\text{fejet és írást is dobunk})}{P(\text{írást dobunk})} =$$

$$\frac{P(\text{FI vagy IF})}{P(\text{FI, IF vagy II})} = \frac{2/4}{3/4} = \frac{2}{3}$$

1. Legyen RND a $[0,1]$-en egyenletes eloszlású valószínűségi változó. Legyen A: $RND < 0,5$ és B: $0,4 < RND < 0,7$. Ekkor

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(RND < 0,5 \text{ és } 0,4 < RND < 0,7)}{P(0,4 < RND < 0,7)} =$$

$$\frac{P(0,4 < RND < 0,5)}{P(0,4 < RND < 0,7)} = \frac{0,5 - 0,4}{0,7 - 0,4} = \frac{1}{3}.$$

2. Egy gyárban két gépsoron készített termékek napi minősítése az alábbiak szerint alakult:

	1. gépsoron	2. gépsoron	összesen
I. osztály	420	210	630
II. osztály	265	455	720
összesen	685	665	1350

Ekkor:

$$P(\text{I. osztály}|\text{I. gépsor}) = \frac{420}{685}, \quad P(\text{I. gépsor}|\text{I. osztály}) = \frac{420}{630}$$

1.4 Független események

Az A és B események <u>függetlenek</u>, ha $P(A|B) = P(A)$.

Ez azt jelenti, hogy

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

Példák:

1. Kétszer dobok egy szabályos pénzérmével: A: első dobás fej, B: második dobás írás. Ekkor

   $$P(A) = P(FF \text{ vagy } FI) = \frac{2}{4}, \quad P(B) = P(FI \text{ vagy } II) = \frac{2}{4},$$

   ezért

   $$P(AB) = P(FI) = \frac{1}{4} = P(A)P(B),$$

   tehát A és B események függetlenek.

1. RND: A: $RND < 0,5$ és B: $0,4 < RND < 0,6$. Ekkor

$$P(A) = 0,5 \quad P(B) = 0,2$$

továbbá

$$P(AB) = P(RND < 0,5 \text{ és } 0,4 < RND < 0,6) =$$

$$P(0,4 < RND < 0,5) = 0,1 = P(A)P(B),$$

tehát A és B események függetlenek.

2. Kockával dobunk: A: párosat, B: 3-nál nagyobbat dobunk. Ekkor

$$P(A) = \frac{3}{6}, \quad P(B) = \frac{3}{6}$$

és

$$P(AB) = P(4 \text{ vagy } 6) = \frac{2}{6} \neq P(A)P(B),$$

tehát A és B események nem függetlenek.

1.5 Teljes valószínűség tétel, Bayes-tétel

A B_1, B_2,...,B_n események teljes eseményrendszert alkotnak, ha egy kisérlet során pontosan az egyik következik be, azaz formálisan

$$B_1 + B_2 + \cdots + B_n = \Omega$$

és

$$B_i B_j = \emptyset \quad ha \quad i \neq j$$

Feladat: Számítsuk ki a $P(A)$ valószínűséget a $P(A|B_i)$ feltételes és $P(B_i)$ valószínűségek ismeretében!

$$P(A) = P(A\Omega) = P(A(B_1+B_2+\ldots B_n)) = P(AB_1+AB_2+\cdots+AB_n) =$$

$$P(AB_1) + P(AB_2) + \cdots + P(AB_n)$$

Tudjuk, hogy $P(A|B_i) = \frac{P(AB_i)}{P(B_i)}$, ezért $P(AB_i) = P(A|B_i)P(B_i)$, innen kapjuk a Teljes valószínűség tételt:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$

Feladat: Fejezzük ki a $P(B_i|A)$ feltételes valószínűséget a $P(A|B_i)$ feltételes és $P(B_i)$ valószínűségek ismeretében!

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} =$$

$$\frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)},$$

amit Bayes-tételnek hívunk.

Feladat: Egy gyárban minden hétköznap reggeli, napközbeni és esti műszak dolgozik. A reggeli műszak készíti az aznapi termelés 25%-át, a napközbeni a 40%-át, az esti a 35%-át. A reggeli műszak 3%-ban, a napközbeni 4%-ban, az esti 2%-ban készít selejtet. Egy nap végén az ellenőrzés során véletlenszerűen kiválasztanak egy terméket az aznapi termelésből. Határozza meg, hogy

1. annak a valószínűséget, hogy selejtest választottak ki
2. ha selejtest választottak ki, akkor mi a valószínűsége, hogy a reggeli műszak készítette?

Megoldás: Legyen A: selejteset választottak ki, B_1: reggeli, B_2: napközbeni, B_3: esti műszakét választották ki. Ekkor

$$P(A|B_1) = 0,03, \quad P(A|B_2) = 0,04, \quad P(A|B_3) = 0,02$$

$$P(B_1) = 0,25, \quad P(B_2) = 0,4, \quad P(B_3) = 0,35$$

ezért

$$P(selejt) = P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) =$$

$$0,25 * 0,03 + 0,4 * 0,04 + 0,35 * 0,02$$

$$P(reggel|selejt) = P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} =$$

$$\frac{0,25 * 0,03}{0,25 * 0,03 + 0,4 * 0,04 + 0,35 * 0,02}$$

Feladat: Magyarországon minden tízezredik lakos HIV fertőzött. A fertőzöttség kiszűrésére AIDS-tesztet használnak, ami az esetek kis százalékában téved. Konkrétan az egészséges emberek 1%-ában pozitív, míg a fertőzötteknél 1,5%-ában negatív. Jancsi Bácsi tesztja sajnos pozitív. Mi a valószínűsége, hogy Jancsi Bácsi tényleg fertőzött?

Megoldás: A: Jancsi Bácsi tesztje pozitív
B_1: Jancsi Bácsi HIV fertőzött
B_2: Jancsi Bácsi nem HIV fertőzött
Ekkor $P(B_1) = 0,0001$, $P(B_2) = 0,9999$ és
$P(A|B_1) = 1 - 0,015 = 0,985$, $P(A|B_2) = 0,01$. Ezért

$$P(\text{fertőzött}|\text{teszt pozitív}) = P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)}$$

$$\frac{0,985 * 0,0001}{0,985 * 0,0001 + 0,01 * 0,9999} = 0,00975$$