I.

A most induló cikksorozatban szeretnénk ismertetni az iskolákban alig tanított statisztika és a valószínűségszámítás néhány válogatott, a gyakorlatban is gyakran használt alapgondolatát, és a hozzákapcsolódó elméleti matematikai megfontolásokat.

Először három helyzetet mutatunk be, amelyekben közös a probléma: mit lehet tenni, ha valamit nem tudunk biztosan. Mindhárom problémánál nagyon is aktuális kérdéseket teszünk föl, amivel az a fő célunk, hogy az olvasót egy kicsit bevezessük abba a gondolkodásmódba, amely nélkül, véleményünk szerint, a XX. század végén elképzelhetetlen a világ megértése.

Ha az olvasó már járatos ezekben a kérdésekben, akkor kísérelje meg a továbbolvasás előtt a válasz megadását. A cikk írásakor figyelembe vettük, hogy az „általános” iskolákban kevés szó esik ilyen problémákról. Emellett törekedtünk arra is, hogy a matematikában járatosabb olvasó is találjon érdekességet a cikkben. A kérdések megválaszolásához modelleket kell alkotnunk, majd ezeken belül bizonyos számításokat kell elvégezni. Ezek néhol igen nehezek, ezért fontosnak találtuk, hogy e bonyolultan bizonyítható, de a gyakorlatban a számítások elvégzésénél nagyon hasznos összefüggéseket bemutassuk, ezek bizonyítását azonban az olvasók nyugodtan mellőzhetik.

1. Probléma. (Nevezzük MALÉV problémának.) A MALÉV menetrend szerinti londoni járatán 250 hely van. A helyfoglalásokat folyamatosan veszik fel. Legfeljebb hány előjegyzést vegyenek föl, ha nagy (pl. 99%-os) eséllyel nem akarnak abba a kellemetlen helyzetbe kerülni, hogy az utasnak visszaigazolt megrendelés ellenére nem tudnak helyet biztosítani?

Tudjuk a korábbi évek tapasztalataiból, hogy a megrendeléseknek körülbelül a 10%-át az utazók különböző okokból (pl. betegség, vagy üzleti helyzet változása stb.) visszamondják.

Ez a probléma minden közlekedési céget, szállodát, egyéb szolgáltatást érint, tehát nem csak az idegenforgalmat, azaz nagyon általános kérdés. Mielőtt elemezni kezdenénk a helyzetet s valamilyen választ adnánk a kérdésre, nézzünk még két további problémát.

2. Probléma. (Ismerve a sokaság eloszlását, mit lehet mondani egy minta eloszlásról, amelyet ebből a sokaságból vesznek?)

Egy, az Országos Rendőrfőkapitányság és a Főügyészség által közölt statisztikából tudjuk, hogy Magyarországon 1990-ben 341 061 bűncselekmény vált ismertté (feljelentés, eljárás indítás), s ebből 187 655 esetben az elkövető ismeretlen maradt. Ha most kiválasztunk 1000 bűncselekményt a 341 061-ből, akkor vajon mekkora eséllyel lesz legalább 400 esetben ismeretlen az elkövető? Mit kell érteni kiválasztáson? Hogyan lehet megvalósítani a kiválasztás „véletlenszerűségét”, amelyre a modellünk mond valamit?

3. Probléma. (Hogyan lehet egy közvéleménykutatási eredményből, tehát egy minimodellből következtetni az egész sokaságra?)

A Wash&Go cég szeretne képet kapni, hogy legújabb termékük milyen fogadtatásban részesült a magyar piacon. Ezért valahogyan kiválasztanak 1000 családot, s megkérdezik a véleményüket. Ha 342 család nem észlelt javulást az új termék esetén, sőt 47-en még rosszabbnek is tartják, míg 611 család véli jobbnak, akkor vajon jobbnak tartsuk-e az új terméket? Milyen előfeltételeket használunk az okoskodásban?

Természetesen mindhárom problémakörben közös, hogy a „fekete-fehér” „igen-nem” gondolkodásról le kell mondanunk. A kérdés az, lehet-e másként gondolkozni, mit jelentenek az olyan misztikus fordulatok, mint: nagy eséllyel, valószínűleg, szinte kizárt stb. Lehet-e ezeknek a fogalmaknak valamilyen, a matematikában megszokott egzakt értelmet adni? Erre szeretnénk a következőkben választ keresni először az első probléma elemzésével.

Mindenki kapásból azt válaszolná, hogy 250 helyre legfeljebb 250 megrendelést lehet fölvenni, s akkor semmi gond. Ez igaz is, de ekkor a gép a tapasztalatok szerint csak 90%-os kihasználtsággal közlekedik, ami figyelembe véve az utasonkénti 30000 Ft körüli veszteséget, nem egészen elhanyagolható összeg. Vajon, ha a biztosból egy *kicsit engedek*, akkor mi adódik? Tehát több megrendelést veszek föl, mitn 250, remélve, hogy úgyis lesz lemondás. Először 251 megrendelést veszünk fül. Mikor kerülünk bajba? Egyetlen esetben, ha mind a 251 utas utazik is. Egy utasra vonatkoztatva a 10% lemondási rátát, azt mondhatjuk, hogy annak az esélye, hogy valaki, aki megrendelte a helyét, utazik is, 0,9. Hogyan lehet annak az esélyét kiszámítani, hogy minden utas utazni akar? A választ a következő modellel kíséreljük megadni.

Feltesszük, hogy egy urnában van 9 fehér és 1 fekete golyó. Ebből az urnából, húzunk az n utasszámnak megfelelően n-szer *visszatevéssel* (tehát minden húzás után visszatesszük a kihúzott golyót, miután feljegyeztük, fehér volt-e, vagy fekete) s minden fehér golyó utazást, a fekete pedig lemondást jelent. A kérdés az: milyen eséllyel lesz k számú fehér golyó az n között, ha $k = 0, 1, 2, \ldots, n$. Ez a modell biztosítja, hogy minden utas esetén 90% az esélye az utazásnak, valamint ha minden húzás után keverünk, akkor a függetlenség is teljesül. A modell előnye, hogy beleépíthető a függés, illetve az egyes emberek különböző esélyei azáltal, hogy az egyes húzások előtt lehet az urna összetételét változtatni, s ez tehető az előző húzások függvényében is. Ilyen bonyolultabb esettel azonban nem akarunk foglalkozni. Mindjárt kicsit általánosabban oldjuk meg a problémát, tehát a visszatevéses húzássorozatok esetét.

Legyen M fehér és $N - M$ fekete golyó egy urnában, amelyből n-szer húzunk visszatevéssel. Vajon hányszor húztunk fehér golyót?

Ez a szám lehet elvileg 0 és n között bármi, az esélyeket próbáljuk meg kiszámítani. Legyen a fehér jele I, a feketéé F. Ekkor tulajdonképpen n hosszúságú I, F sorozatokat kell vizsgálnunk. Feltesszük, hogy mivel minden húzás

„azonos" körülmények között zajlik, így az összes lehetőség egyenlő valószínű[1]. Ebből hányban lesz pontosan k darab fehér (I)? A fehérek helyét $\binom{n}{k}$-féleképpen választhatjuk ki, s ekkor, feltéve, hogy a golyók (például számozással) megkülönböztethetőek, $M^k \cdot (N-M)^{n-k}$-féleképpen valósulhat meg a k fehér golyó kihúzása. A k darab fehér golyó kihúzásának valószínűségét jelöljük $P(X=k)$-val.

Alkalmazva a klasszikus $\frac{\text{kedvező eset}}{\text{összes eset}}$ képletet[2]:

$$P(X=k)=\binom{n}{k}\frac{M^k\cdot(N-M)^{n-k}}{N^n}=$$
$$=\binom{n}{k}\cdot\left(\frac{M}{N}\right)^k\cdot\left(\frac{N-M}{B}\right)^{n-k}=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k},$$

ahol p-vel jelöltük a fehét golyók arányát az összes golyóhoz, ami egyben a fehér golyó hűzűsűnak az esélye is minden egyes húzásnál.

Ha a fenti esélyek szerint oszlik el egy véletlen mennyiség a $0,1,\ldots,n$ számok között, akkor *binomiális eloszlásúnak* nevezzük.

Néhány fontos tulajdonsága ennek az eloszlásnak leolvasható a alábbi *1. ábra* grafikonjáról, amelyek általánosan igazak a binomiális eloszlásra:

Az esélyek növekednek egészen a maximális np körüli értékig, majd onnan fogynak (az ilyen eloszlást hívják egycsúcsúnak is). Vajon ha sokszor ismételjük meg az n-húzásos kísérletet, áltagosan hány fehér lesz? Erre ad választ a *várható érték*, melyet a $\sum_{k=0}^{n} k\cdot P(X=k)$ összeggel definiálunk. Megmutatható, hogy ez binomiális eloszlás esetén pontosan np, épp az az érték, ahol az eloszlás csúcsa is van. Ez a tény, hogy a várható érték egyben a legvalószínűbb is, egy speciális tulajdonság, amely más eloszlásokra nem igaz. Ld. [1] (252. o.). Egy másik fontos jellemző az átlagtól való átlagos négyzetes eltérés, az ún. szórásnégyzet. Ez binomiális eloszlás esetén $np(1-p)$. Ld. [1] (253–254. o.).

Ha most feltesszük, hogy az utasok között nincs kapcsolat — ami szükséges feltétele annak, hogy a fenti modellt használjuk —, akkor egyszerűen számítható 251 utas esetén annak az esélye, hogy nem lesz visszamondás, mert mindenki utazni akar, nevezetesen $0,9^{251}\approx 3,27\ \cdot 10^{-12}$. Ez tehát nagyon valószínűtlen, hogy bekövetkezzen. Csak összehasonlításképpen, annak az esélye, hogy egy lottószelvény kitöltésével a jövő héten ötösöm lesz: $2,2\cdot10^{-8}$, tehát majdnem 10 000-szerese, és sajnos ez az esemény sem sűrűn következi be.

Ha 252 embert nézünk, akkor már két esetet kell figyelembe venni: ha senki, vagy ha csak egy ember mondja vissza. Ennek esélye: $0,9^{252}+252\cdot0,9^{251}\approx 8,53\cdot10^{-11}$, ami még mindig elég kicsi.

A számolás azonban modellünkben egyre nehezebb lesz, mert egyre több tagot kell összeadni, s a kiválasztás miatt egyre komplikáltabb binomiális együtthatókat kell számolni.

Ezt mutatja a következő számolás 260 megrendelés elfogadása esetén.

$$P(X\le 250)=1-\left[\sum_{k=251}^{260}P(X=k)\right]=1-\left[\binom{260}{251}\cdot0,9^{251}\cdot0,1^9+\right.$$
$$\left.+\binom{260}{252}\cdot0,9^{252}\cdot0,1^8+\cdots+\binom{260}{259}\cdot0,9^{259}\cdot0,1+0,9^{260}\right].$$

Ennek kiszámítása már nem kellemes, és még egy zsebszámológéppel is kb. 15–20 perc szükséges legalább az elvégzéséhez (ha elég ügyes, aki a számolást végzi!). Esetünkben a legnagyobb lesz az első kivonandó tag: $4,2566\cdot10^{-5}$, míg az utolsó már csak $1,27\cdot10^{-12}$. Ha az elsőt vennénk 10-szer, ami nagyon durva alsó becslés lesz a $P(X\le250)$-re — hiszen sokkal többet vonok le belőle, mint valójában kellene — akkor is 99,95% eséllyel nem lesz elutasított megrendelő. Valójában ez az esély még 99,99%-nál is kicsit nagyobb, tehát kb. minden tízezredik esetben lesz kellemetlenségünk a 260 megrendelés felvétele esetén.

A kérdésünk azonban úgy hangzott, hogy ha valaki megadja, mekkora biztonságra kell törekedni (0,9999 vagy 0,99 vagy 0,95 stb.), akkor adjuk meg ebben az esetben, hány megrendelést lehet felvenni. Ez a gyakorlatban nagyon fontos, hiszen a módszerrel sok pénz megtakarítható. Tehát kicsi kockázattal sokat lehet nyerni. Jó lenne, különösen még nagyobb esetszámnál, valahogyan becsülni. A binomiális eloszlást már Jakab Bernoulii ismerte, de egy nagyon fontos közelítőformulát először Abraham de Moivre és Pierre Laplace fedezett fel. Azt vették észre, hogy ha a binomiális eloszlás görbéjét eltoljuk úgy, hogy a csúcsa az y-tengelyre essen és ne nyúljon nagyon el, tehát még egy osztóval, ami éppen a szórása, *normáljuk*, akkor a Gauss-féle haranggörbét kapjuk. Ennek a jelenségnek egyelőre csak illusztrációja

[1] Ez az úgynevezett elégséges ok hiányának az elve, azaz ha nincs valamiért kitüntetett elem, akkor mindegyiket azonosan kezeljük.

[2] Ez lényegében következménye az elégséges ok elvének, s először Laplace fogalmazta meg 1812-ben kiadott könyvében, az „Essai philosophique des probabilités" címűben.

a következő, számítógéppel készített sorozat *(2. ábra a), b), c)*. Az a) ábrán különböző n-ekre a $p = 1/2$ paraméterű binomiális eloszlásokat ábrázoljuk, a b) ábrán egy binomiális eloszlás (a jobb oldali görbe) normálását látjuk, majd végül a c) ábrán az eloszlások normálásával keletkezett, az y-tengelyre szimmetrikus görbéket. A „határgörbe" a Gauss-féle haranggörbe, amelyet standard normális eloszlásnak hívnak. Arról, hogy ez milyen függvénynek a görbéje, következő számunkban olvashatnak.

Vancsó Ödön

Irodalom

[1] Hajnal–Nemetz–Pintér–Urbán: Matematika IV. (B fakt) Tankönyvkiadó 1982.