Pólya-féle urnamodell I.

1. Bevezetés

Hogyan magyarázhatók azok a jelenségek, mikor több, egymással versengő, hasonló esélyekkel induló alternatíva közül valamelyik egy rövid időn belül átveszi a vezetést, és ha az előnyt megkaparintotta, onnantól kezdve a többinek már nincs esélye őt behozni? Miért használja a világ elsöprő többsége az egyébként kevésbé hatékonynak tartott QWERTY billentyűzetet? A sok, alapvetően hasonló szolgáltatást nyújtó közösségi hálózat létezése ellenére miért vált elsöprő népszerűségűvé a Facebook?

A valószínűségszámítás egyik alapvető gondolatkísérlete *Pólya György* (1887–1985) nevéhez fűződik, és épp ezt a jelenségkört modellezi. Egy időben véletlenül növekedő rendszerben a kezdeti időszakban megvalósuló véletlen lépések olykor sokkal meghatározóbbá válnak a távoli jövőbeli véletlen kimenetel szempontjából, mint a későbbi lépések. A Pólya-féle urna folyamat után, a cikk második felében egyéb kapcsolódó modelleket is vázolunk, melyek mind erre a szálra fűzhetők fel. Ezeket a kombinatorikus megközelítéssel jól megérthető modelleket részletesen tárgyalja [2] és [3]. Magyar nyelven is elérhető, klasszikus leírásért lásd [1]-et.

A szóhasználatról megjegyezzük, hogy bár az „urna" kifejezés mai füllel kissé bizarrul hangozhat, a matematikus társadalom mégis ezt a szót használja történelmi okokból[1], amikor dobozokban elhelyezett színes golyók véletlen kihúzásairól beszél. A magyar nyelvű matematikai irodalom is megőrzi ezt a szóhasználatot, és ennek megfelelően mi is így teszünk.

A megértéshez a kombinatorikus valószínűség valamint a határérték fogalmának ismerete szükséges. Az egyéb felmerülő fogalmakat a szemléletességet szem előtt tartva igyekszem bemutatni.

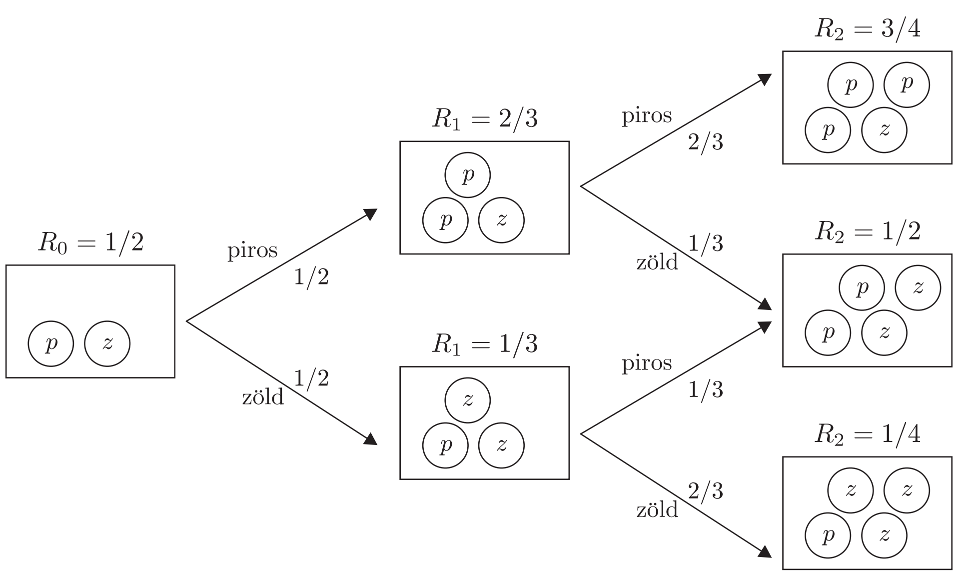
2. A Pólya-féle urnamodell alapesete

2.1. A modell

Egy dobozban kezdetben két golyó van, egy piros és egy zöld. A dobozban lévő piros golyók arányát jelöljük R_0-lal, jelen esetben tehát $R_0 = 1/2$.

Az első lépés abból áll, hogy véletlenszerűen (minden golyónak azonos esélyt adva, azaz *egyenletes eloszlással*) kihúzzuk az egyik golyót a dobozból és megnézzük a színét, majd visszatesszük. Ezen kívül teszünk a dobozba még egy új golyót is, mégpedig olyan színűt, mint amilyet az imént húztunk. A dobozban tehát most három golyó van: ha pirosat húztunk, akkor egy zöld és két piros, ha pedig zöldet húztunk, akkor egy piros és két zöld. A dobozban lévő piros golyók arányát jelöljük R_1-gyel, ami tehát lehet $2/3$ (pirosat húztunk) vagy $1/3$ (zöldet húztunk).

Jöhet a második lépés: a most bent lévő három golyó közül húzunk véletlenül, egyenletes eloszlással, megnézzük a színét, visszatesszük a dobozba egy új, ugyanolyan színű golyóval együtt. A most ott lévő piros golyók aránya, R_2, tehát lehet $3/4$, $1/2$ vagy $1/4$, attól függően, hogy mennyi volt R_1, és hogy milyen színű golyót húztunk most (lásd *1. ábra*).



1. ábra. Pólya-féle urnamodell, alapeset

[1] Már Jacob Bernoulli 1713-as *Ars Conjectandi* c. munkájában is a latin *urna* szó jelenik meg, ő ebben az agyag edényben helyez el gondolatban színes kavicsokat, ezek közül húz ki gondolatban egy pár darabot, és az így az urnában maradó véletlen kavicshalmazban vizsgálja a különböző színek arányának eloszlását.

Ugyanígy haladva tovább, ezt az elemi lépést ismételjük meg egymás után sokszor. Hogyha a dobozban n lépés után p_n darab piros és z_n darab zöld golyó található, azaz a piros golyók aránya

$$R_n = \frac{p_n}{p_n + z_n},$$

akkor a következő állapotot úgy kapjuk, hogy húzunk véletlenszerűen a dobozból egy golyót (R_n valószínűséggel pirosat, $1 - R_n$ valószínűséggel zöldet), és a visszatételével egyidőben teszünk még egy ugyanilyen színű golyót a dobozba. Az R_{n+1} érték ennek megfelelően kétféle lehet,

a) ha pirosat húzunk, akkor $R_{n+1} = \dfrac{p_n + 1}{p_n + z_n + 1}$, ennek valószínűsége R_n,

b) ha zöldet húzunk, akkor $R_{n+1} = \dfrac{p_n}{p_n + z_n + 1}$, ennek valószínűsége $1 - R_n$.

Ez a Pólya-féle urnamodell legegyszerűbb esete.

2.2. A véletlen állapot rögzített számú lépés megtétele után

Milyen szabályszerűséget fedezhetünk fel, rögzített n mellett, az R_n szám *véletlen* értékével kapcsolatban? Könnyen látható, hogy R_n lehetséges értékei a $(0, 1)$ intervallumban a $k/(n + 2)$ alakú törtek, ahol $1 \le k \le (n + 1)$, azaz ahogy n egyre nagyobb, az R_n értékkészlete kezdi egyenletesen sűrűn kitölteni a teljes $(0, 1)$ intervallumot. De milyen eséllyel lesz R_n értéke éppen $k/(n + 2)$, valamilyen adott k-ra?

Világos, hogy R_n, a piros golyók aránya az n-edik lépés után épp annak az aktuális (feltételes) valószínűségét adja meg, hogy a következő, $(n+1)$-edik lépésnél piros színű golyót húzunk, feltéve, hogy már eljutottunk az n-edik lépésig.

Az első pár lépésben a megvalósulások valószínűségeit könnyen kiszámolhatjuk, például az *1. ábrát* követve. R_1 értéke kétféle lehet,

a) $R_1 = 2/3$, ha pirosat húzunk elsőre, ennek esélye $1/2$, vagy

b) $R_1 = 1/3$, ha zöldet húzunk, ennek esélye ugyanúgy $1/2$.

R_2 értékének kiszámolásakor ezt már figyelembe vesszük. Az *1. ábráról* leolvashatók a következő esetek valószínűségei:

a) az $R_2 = 3/4$ érték úgy érhető el, ha $R_1 = 2/3$, és újra pirosat húzunk, ez az eseménysorozat $1/2 \cdot 2/3 = 1/3$ eséllyel történik.

b) Az $R_2 = 1/2$ érték kétféleképp érhető el:

  i) vagy $R_1 = 2/3$ és most zöldet húzunk, aminek az esélye $1/2 \cdot 1/3 = 1/6$,

  ii) vagy pedig $R_1 = 1/3$ és most pirosat húzunk, ennek esélye szintén $1/2 \cdot 1/3 = 1/6$.

  Összességében tehát az $R_2 = 1/2$ értéket is $1/6 + 1/6 = 1/3$ eséllyel érjük el.

b) Végül az $R_2 = 1/4$ érték csak úgy lehetséges, ha $R_1 = 1/3$ és újra zöldet húzunk, ennek esélye ismét $1/2 \cdot 2/3 = 1/3$.

Azaz R_1 és R_2 esetében is teljesül az, hogy az összes lehetséges érték között egyenletesen oszlik el a valószínűség, más szóval a változó az értékkészletén egyenletes eloszlású. Véletlen-e, hogy $n = 1$ és $n = 2$ esetében ez így van?

2.1. feladat. *A merész olvasónak javasoljuk, ugorja át a felvezetésként megadott a) és b) részfeladatokat, kezdje rögtön az utolsó, c) jelű feladattal.*

a) *Mi a valószínűsége annak, hogy az urnából kihúzott golyósorozat első k eleme mind piros, az azutáni $l = n - k$ eleme pedig mind zöld?*

b) *Rögzítsünk egy $A \subset \{1, 2, \ldots, n\}$, $|A| = k$ elemszámú részhalmazt. Mi a valószínűsége annak, hogy az urnából kihúzott golyósorozat első n eleme közül pontosan az A-beli sorszámú golyók pirosak?*

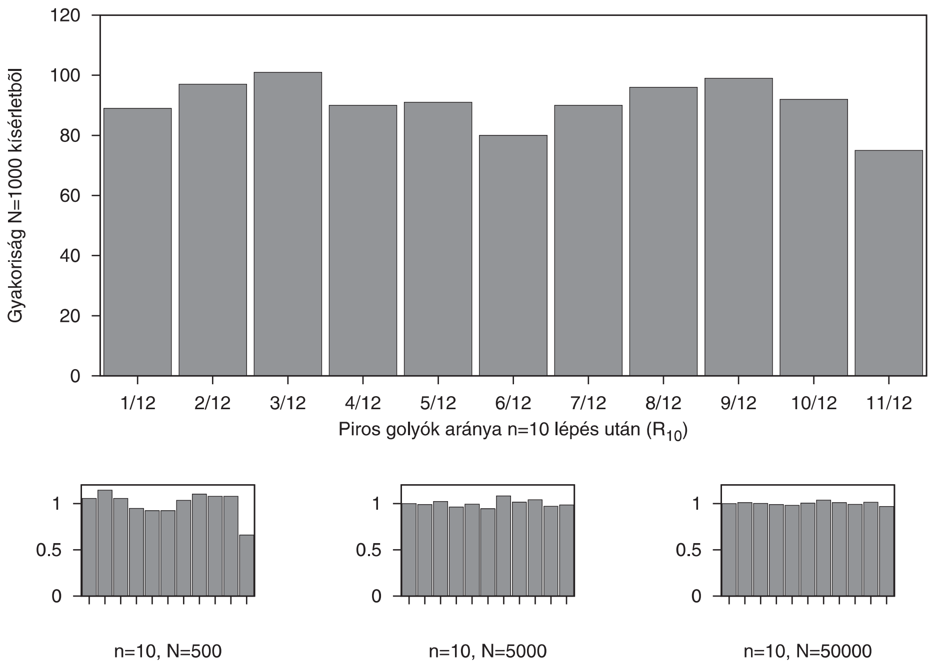
c) *Mi a valószínűsége annak, hogy az urnából kihúzott golyósorozat első n eleme közül pontosan k darab piros?*

A 2.1. feladatban felfedezett tulajdonságot „felcserélhetőségnek” nevezik. Azt jelenti, hogy egy adott sorozat megvalósulásának valószínűsége invariáns a permutációra nézve: a valószínűség csak a sorozatban szereplő arányoktól függ, a sorrendtől nem. A felcserélhetőségnek a Pólya-féle urnamodell az egyik tankönyvi alappéldája.

2.2. feladat. *Bizonyítsuk be tetszőleges $n \in \mathbb{N}$-re, hogy R_n azonos valószínűséggel veszi fel az összes lehetséges értékét! (Az értékkészlet természetesen n-től függ.)*

2.3. Számítógépes kísérlet

Vegyünk egy rögzített n értéket ($n = 10$), és indítsuk el ugyanannak a Pólya-urna kísérletnek sok ($N = 1000$) független példányát, mindig egy piros és egy zöld golyóval kezdve, és futtassuk n lépésig. Ezekben a független kísérletekben mindig jegyezzük fel, mi jött ki $R_n = R_{10}$ értékére, és rajzoljuk fel a gyakorisági ábrát, más néven hisztogramot, ebből az $N = 1000$ kísérletből. Az általunk írt program segítségével előállított példa-hisztogramot lásd a *2. ábra* tetején.



2. ábra. Pólya-féle urnamodell alapesetének hisztogramja, illetve normált hisztogramjai 10 lépés után, különböző kísérletszámok mellett

Tegyük fel, hogy az $n = 10$ választást rögzítjük, de N értékét, azaz a kísérletek számát változtatni akarjuk, és az eredményül kapott hisztogramokat össze akarjuk vetni egymással. Hogyha a különböző N-ekre előállított hisztogramokat úgy normáljuk, hogy az oszlopok által lefedett tartomány területe mindig azonos legyen, akkor az összevetés kényelmes lesz, mert a lépték minden ábránkon azonos, lásd a *2. ábra* alsó sorozatát. Rögzített n mellett N növelésével az ábra egyre inkább hasonlít a feladatban bizonyított elméleti eloszláshoz tartozó, egyenletes hisztogramhoz. Azt, hogy a hisztogramok N növelésével valóban az elméletihez tartanak, a statisztika alaptétele mondja ki.

2.3. feladat. *Írjuk meg a Pólya-féle urnamodell alapesetének algoritmusát egy tetszőleges, általunk elérhető programnyelven*[2] *és állítsunk elő a* 2. ábrákhoz *hasonlókat!*

A 2.2. feladat állításából kiolvasható, hogy n növelésével R_n elméleti eloszlása a $(0,1)$ intervallumon értelmezett úgynevezett folytonos egyenletes eloszláshoz tart (ennek úgynevezett „sűrűségfüggvénye" a vízszintes vonal, ahol a vonal alatti terület 1-re van normálva). Ez már önmagában is érdekes tény, ámde ennél egy sokkal erősebb állítás is igaz. A következő fejezetben e felé az állítás felé haladunk.

3. A véletlen sorozat tulajdonságai

3.1. A Pólya-féle urnamodell általánosabb esetei

Mielőtt továbblépünk további megfigyelések felé, tágítsuk ki egy kissé a vizsgált modellek körét. A Pólya-féle urnamodell 2.1. fejezetben vázolt alapesete többféleképpen is általánosítható, például a kezdeti feltétel variálásával. A golyók számának növelésére szolgáló algoritmust változatlanul hagyjuk, viszont ahelyett, hogy kezdetben 1 piros és 1 zöld golyót tennénk az urnába, indítsunk p_0 piros és z_0 zöld golyóval ($p_0 \geq 1$, $z_0 \geq 1$).

Általánosíthatjuk ennél jobban is az urna modellt. Például úgy, hogy nem csak két színt engedünk meg, hanem egy előre rögzített, K elemszámú színhalmazból válogatva teszünk kezdetben az urnába r_1 darab c_1 színű, r_2 darab c_2 színű, ..., r_K darab c_K színű golyót. Ha az olvasóban felmerül, meg tudunk-e engedni végtelen sok színt is, még egy kis türelmet kérünk tőle, egy későbbi fejezetből ez is ki fog derülni. Egyelőre maradjunk a két szín használata melletti általánosításnál, mert ez is rejteget még érdekességet.

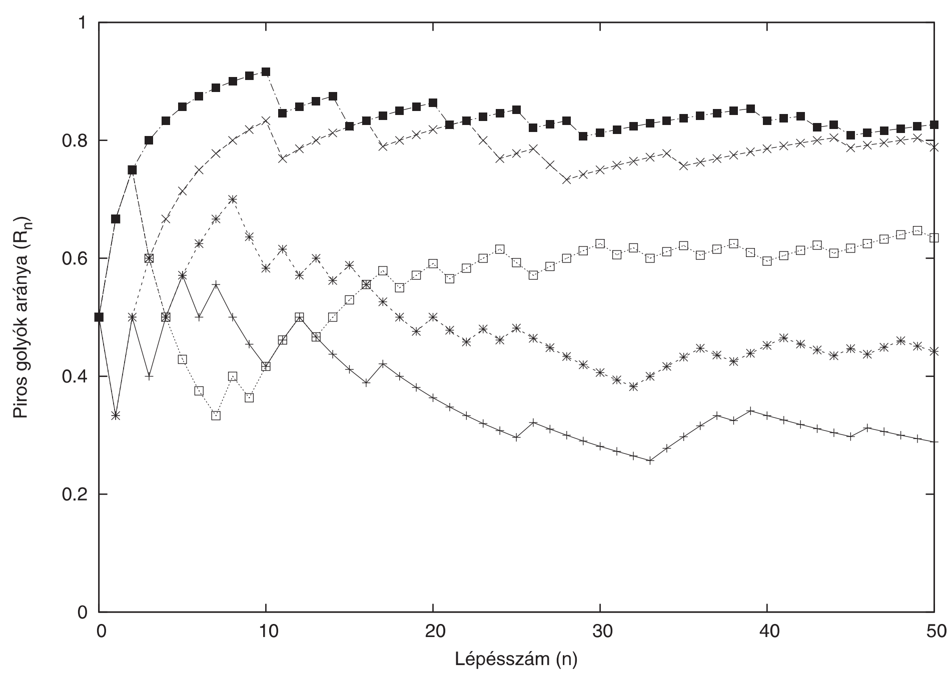
3.2. Mi a helyzet a véletlen sorozattal?

A 2.2. feladatban bizonyított állítás rögzített n mellett mond valamit az R_n véletlen szám (valószínűségi változó) lehetséges értékeiről, és az értékek megvalósulásának valószínűségeiről (a valószínűségi változó eloszlásáról). Olyan típusú eredmény ez, mint mikor a 2.3. feladatban megírt programunk futásakor nem írattuk ki a futás közben érintett R_j értékeket ($j < n$), hanem minden egyes futáskor csak az általunk előre kiszemelt n mellett voltunk kíváncsiak R_n kimeneteleire. Hogy éppen hogyan jutottunk el addig a véletlen számig, azt nem vizsgáltuk.

Természetes módon felmerülhet bennünk ugyanakkor a kérdés, megállapíthatunk-e valamit a véletlen számsorozat egészéről (véletlen folyamatról) is, ami egy-egy ilyen szimuláció során előáll. Feljegyezhetnénk sok, egymástól függetlenül előálló ilyen sorozatot, és ezekről a véletlen sorozatokról is megkérdezhetnénk, vajon milyen tulajdonságokkal rendelkeznek.

3.1. feladat. *A korábban írt számítógépes programunkkal állítsunk elő független kísérletekből előálló véletlen pályákat, más néven trajektóriákat, a Pólya-urna alapesetében* ($p_0 = 1$, $z_0 = 1$).

Az általunk írt programmal készített trajektóriákat lásd a *3. ábrán.* Ezek megfigyelésével máris tehetünk néhány kezdeti megállapítást.



3. ábra. Trajektóriák a Pólya-féle

[2] A cikkben szereplő szimulációkat Python nyelven írtam, a kód szerkesztéséhez és futtatásához Eclipse-t használtam (Pydev modullal), a képek megjelenítéséhez pedig Gnuplot-ot. Ezek mindegyike ingyen hozzáférhető.

urnamodellben ($p_0 = 1$, $z_0 = 1$)

Az algoritmusról általánosságban megállapíthatjuk azt, hogy a piros golyó húzásának éppen aktuális valószínűsége időben véletlenszerűen változik, ahogyan az urnában a piros golyók aránya az összes golyóhoz képest mindig más és más. Ugyanakkor egyetlen golyó betételének hatása erre az arányra, ahogy haladunk előre az időben, egyre kisebb és kisebb, hiszen az új golyó megjelenése az éppen aktuális arányt már egyre kevésbé tudja megváltoztatni.

3.2. feladat. *Az* 3. ábrán *mintha hiperbolikus ívdarabok jelennének meg. Miért van ez?*

3.3. feladat. *Ha visszatekintünk a 2.2. feladatban bizonyított eredményre, látjuk, hogy a programunkat ($p_0 = 1$, $z_0 = 1$)-ből sokszor elindítva az $R_n = R_{10}$ értékek átlaga sok kísérlet után $1/2$. Jegyezzünk fel most minden trajektórián két értéket, R_1-et és R_{10}-et. Mi a megvalósult R_{10} értékek átlaga azokon a trajektóriákon, amelyeken $R_1 = 1/3$? És mennyi azokon, amelyeken $R_1 = 2/3$?*

3.3. A megfigyelések matematikai háttere

Ez a számunkra egyelőre kísérleteken alapuló megfigyelés pontossá is tehető, és a valószínűségszámításnak egy különleges fejezetéhez, az úgynevezett martingálelmélethez kalauzol bennünket. Anélkül, hogy a precíz definícióhoz elengedhetetlenül szükséges elméleti hátteret megkísérelnénk itt megadni, a lényeget mégis vázoljuk.

Azok a véletlen folyamatok, amelyek mindig az éppen aktuális (véletlen) pozíciójukat „szeretnék megtartani”, különös jelentőségük van, és a martingál elnevezést kapták. Ezek, ha egy adott lépéskor megfigyeljük őket, a véletlen érték ismeretében a következő lépésben „várhatóan” pontosan ugyanott maradnak. Ez nem azt jelenti, hogy a következő lépésben nem mozdulnak el, hanem azt, hogy a valószínűségekkel súlyozott átlagos elmozdulásuk nulla. A témakörben az egyik leggyakrabban használt tétel szerint pedig, ismét nem teljesen precízen megfogalmazva, hogyha egy ilyen véletlen folyamat értékkészlete korlátos, akkor annak a véletlen folyamatnak bármilyen tipikus megvalósulása olyan számsorozat lesz, amelynek létezik a határértéke. (A „tipikus megvalósulás” fogalmát jelen cikkben nem definiáljuk.)

3.4. Alkalmazható ez az erős eszköz ebben az esetben? Számoljunk!

Ahhoz, hogy lássuk, ez az erős tétel valóban alkalmazható a Pólya-féle urnamodell fenti általánosításainak esetében, a két feltétel fennállását vizsgáljuk. Az értékkészlet nyilvánvalóan korlátos, R_n csak $(0,1)$-ben mozog, tehát a tétel másik feltétele rögtön teljesül. A másik feltétel igazolása érdekében ellenőrizzük, hogy a martingálokat definiáló tulajdonság igaz az R_n folyamatra.

Képzeljük el, hogy n lépést már megtettünk, és a szokásos módon jelöljük p_n-nel az n-edik lépés után a piros golyók számát, z_n-nel a zöldek számát, R_n-nel pedig a piros golyók arányát az összeshez képest. Ekkor

a) ha a következő lépésben pirosat húzunk, akkor a piros golyók száma $p_n + 1$ lesz, tehát R_{n+1} értéke $\dfrac{p_n + 1}{p_n + z_n + 1}$, ennek a valószínűsége
$$R_n = \frac{p_n}{p_n + z_n};$$

b) ha viszont zöldet húzunk, akkor a piros golyók száma p_n marad, tehát az új arány, R_{n+1} értéke $\dfrac{p_n}{p_n + z_n + 1}$ lesz, ennek valószínűsége pedig
$$1 - R_n = \frac{z_n}{p_n + z_n}.$$

Hogy mennyi az R_n szám ismeretében R_{n+1} feltételes várható értéke, úgy kapjuk meg, hogy mindkét szóba jövő értéket megszorozzuk a valószínűségükkel, majd összeadjuk őket:

$$\frac{p_n + 1}{p_n + z_n + 1} \frac{p_n}{p_n + z_n} + \frac{p_n}{p_n + z_n + 1} \frac{z_n}{p_n + z_n} = \frac{p_n}{p_n + z_n} = R_n.$$

Ennek a számolásnak az eredménye épp R_n, azaz a folyamat n-edik lépésében kapott érték ismeretében a következő lépés által okozott elmozdulás *feltételes várható értéke* nulla. Ez a számolás p_0 és z_0 bármely értéke mellett érvényes. A Pólya-féle urnamodellben a vizsgált véletlen folyamat tehát valóban egy korlátos martingál, aminek a 3.3. szakaszban vázolt tétel szerint tehát létezik a határértéke (ami azonban egy véletlen szám!). A határérték valószínűségi eloszlása a folytonos egyenletes eloszlás a $(0,1)$ intervallumon.

Ez összefoglalva tehát azt jelenti, hogy a véletlen trajektóriák, melyeknek véges szeleteit a számítógépes programunkkal meg is figyelhetjük (*3. ábra*), „végtelen hosszú ideig futtatva” mind beállnak egy-egy (véletlen) értékre.

3.5. Trajektóriák és hamis érmék

Amikor a Pólya-féle modellt először ismerjük meg, lehet az az érzésünk, hogy egy-egy húzás véletlenségében a folyamat teljes múltja közrejátszik. Hamar rájövünk azonban arra, hogy ez nem egészen van így: az $(n+1)$-edik húzás során nem lényeges, hogy az urnában lévő golyók milyen sorrendben kerültek oda, csak a piros golyók aktuális aránya, R_n számít. Azt azonban egy pillanatig sem gondoljuk (és nem is volna igaz), hogy az egymás utáni húzások függetlenek volnának egymástól. De akkor mi a helyzet a következő gondolatmenettel?

Kódoljuk az n lépésből álló kísérleteket egy-egy n-hosszú, P és Z betűkből álló sorozattal, azaz a trajektóriákat kódoljuk a $B_n = \{P, Z\}^n$ elemeivel, az egymás után kihúzott golyók színének megfelelően. E halmaz azon részhalmazát, mely sorozatokban pontosan p_n darab P betű szerepel, jelölje $B_n^{(p_n)}$, ennek elemszáma

$$\left|B_n^{(p_n)}\right| = \binom{n}{p_n}.$$

A 2.1. feladat c) szakaszának megoldásakor kapott kombinatorikus formulából kiolvasható, hogy adott n mellett egy rögzített R_n arány, azaz rögzített p_n számú piros golyó összes lehetséges előállási sorrendje mind azonos valószínűséggel történik. Ez azt jelenti, hogy az R_n arányhoz vezető n-hosszú trajektóriák, a $B_n^{(p_n)}$ halmaz elemei mind egyforma valószínűséggel állnak elő.

Vegyünk egy hamis érmét, mégpedig olyat, ami rögzített $R_n = \dfrac{p_n}{p_n + z_n}$ valószínűséggel esik a piros oldalával felfelé, $(1 - R_n)$ valószínűséggel pedig a zöld oldalával felfelé. Egy kísérlet álljon abból, hogy feldobjuk ezt az érmét n-szer, a dobások egymástól legyenek függetlenek, és jegyezzük fel a dobások eredményeit P és Z betűkkel.

3.4. feladat. *Adjunk módszert a $B_n^{(p_n)}$ trajektória halmazon egyenletes eloszlás generálására ennek a kísérletnek a kis módosításával!*

Ha n-et egyre inkább növeljük, akkor a fenti, R_n hamis érmés kísérlet eredményeképp előálló sorozatban egyre valószínűbb, hogy a P betűk aránya közel van R_n-hez. Ezt úgy hívják, hogy a nagy számok gyenge törvénye.

A nagy számok erős törvénye is igaz, ami pedig azt mondja ki, hogy ha előállítunk egy „végtelen hosszú” sorozatot egy hamis érmével, legyen a hamis érme piros oldallal felfelé esésének valószínűsége θ, akkor a véletlen sorozat egyre hosszabb véges szeleteit nézve a benne szereplő P betűk aránya θ-hoz tart.

Ha tehát rögzítünk egy θ értéket, akkor azon feltétel mellett, hogy a Pólya-féle urnamodell trajektóriája épp ehhez a θ-hoz tart, magát a trajektóriát előállíthatjuk úgy is, hogy veszünk egy θ-hamis érmét, és azt végtelen sokszor feldobjuk. A 3.3. szakaszban pedig láttuk, hogy a határérték, Θ egyenletes eloszlású valószínűségi változó a $[0, 1]$ intervallumon.

Az eddigi megfigyeléseket összerakva tehát a következőt kaptuk. Vegyünk egy Θ valószínűségi változót, amely egyenletes eloszlású a $[0, 1]$ intervallumon. A kisorsolt érték legyen θ, ez tehát most már egy ismert érték a $[0, 1]$ intervallumból. Vegyünk egy hamis érmét, amely ezzel a θ valószínűséggel esik a piros oldalával felfelé, $(1 - \theta)$ valószínűséggel pedig a zöld oldalával felfelé. Dobjuk fel ezt az érmét végtelen sokszor, és jegyezzük fel a dobások eredményeit P és Z betűkkel. Az így kapott végtelen sorozat ugyanolyan eloszlású, mint amit a Pólya-féle urnamodell alapesetében, a kihúzott golyók színét feljegyezve kapunk.

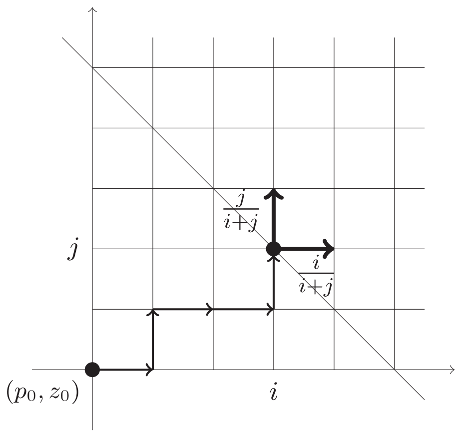
3.5. feladat. *Azok a kedves olvasók, akik az integrálás szabályaival már tisztában vannak, bizonyítsák be a fenti állítást a nagy számok erős törvényére való hivatkozás nélkül: számolják ki a*

$$\int_0^1 \theta^k (1-\theta)^{n-k} \, \mathrm{d}\theta$$

értékét, és vessék össze a 2.1. *feladat* b) *szakaszának megoldásakor kapott kombinatorikus formulával!*

3.6. Bolyongás a negyedsíkon

A Pólya-féle urnamodell két színt használó eseteinek trajektóriáit kézenfekvő módon tekinthetjük egy negyedsík egész koordinátájú rácspontjain való bolyongásnak is. Az x tengelyen jelöljük a piros, az y tengelyen pedig a zöld golyók számát. A pályák a (p_0, z_0) pontból indulnak, és az (i, j) pontból kétfelé léphetnek: az $(i+1, j)$ pontba $\dfrac{i}{i+j}$ valószínűséggel, az $(i, j+1)$ pontba pedig $\dfrac{j}{i+j}$ valószínűséggel (lásd *4. ábra*).



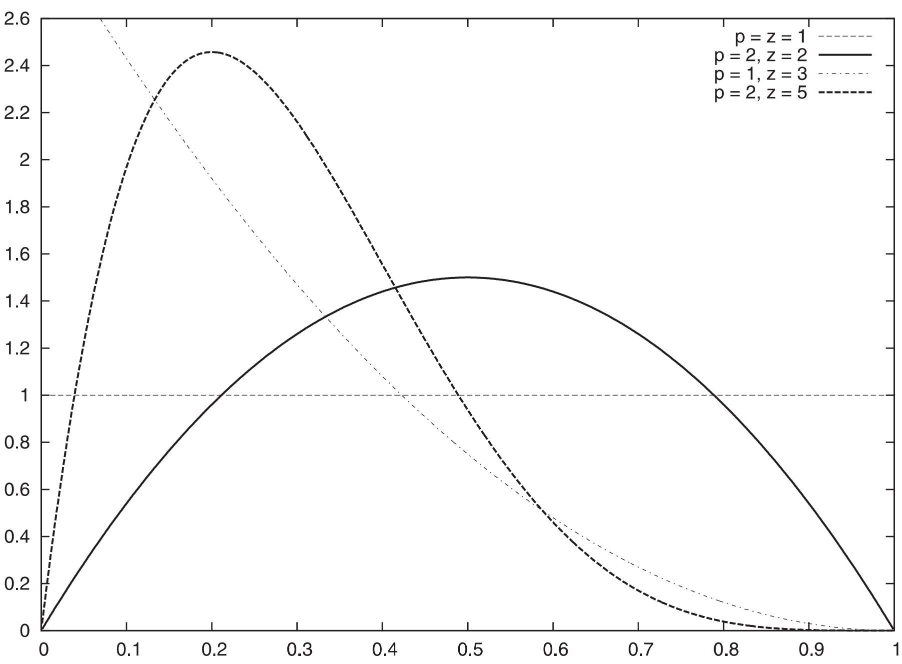
4. ábra. A Pólya-féle urnamodell trajektóriája mint bolyongás a (p_0, z_0) feletti negyedsík rácspontjain

Ebben a reprezentációban R_n egyenletes eloszlása a (p_0+n, z_0) és (p_0, z_0+n) pontokat összekötő átlós egyenes által tartalmazott rácspontokon valósul meg. A véletlen határérték pedig felfogható egy véletlen iránynak a $(0,\pi)$ szögtartományban.

3.7. Milyen az eloszlás, ha nem egyenletes?

Anélkül, hogy folytonos eloszlásokról precízen szót ejtenénk, az említés szintjén álljon itt, hogy míg két szín esetén a legegyszerűbb (egy piros, egy zöld golyóval való indítás) esetben egyenletes eloszlást kapunk a limeszben, addig általános (p_0 piros, z_0 zöld kezdeti állapot) esetben az úgynevezett Béta eloszláscsalád p_0-tól és z_0-tól függő tagjához jutunk.

3.6. feladat. *A korábbi feladat során megírt programot is tegyük alkalmassá p_0 és z_0 különböző értékeinek kezelésére! A Béta eloszláscsalád tagjainak hozzávetőleges képét így fel is rajzolhatjuk. Milyen változást figyelhetünk meg a hisztogramon, ahogy p_0 és z_0 értékét változtatjuk?* (*Néhány eset elméleti képét lásd az* 5. ábrán.)



5. ábra. Béta eloszlás képei különböző paraméterek mellett

Érdekes megfigyelni, milyen hatása van annak, ha kezdetben ugyanannyi piros és zöld golyót teszünk a dobozba, de nem egyet-egyet, hanem például ötöt-ötöt, vagy húszat-húszat. A határérték-hisztogram egyre csúcsosabbá válik. Ez is azt sejteti, hogy konvergál a folyamat: minél nagyobb $p+z$, annál kevésbé távolodik el R_n a kezdeti $R_0 = \dfrac{p}{p+z}$ értéktől, még akkor is, ha n már nagyon nagy.

Az elnevezés említése kedvéért álljon itt az is, hogy ha több, előre rögzített számú (K db) színnel indítunk, akkor az arányok vektora, amint n-nel a végtelenbe tartunk, az úgynevezett *Dirichlet-eloszláscsalád* kezdeti állapottól függően paraméterezett tagjához tart.

Hivatkozások

[1] William Feller, *Bevezetés a valószínűségszámításba és alkalmazásaiba.* Műszaki Könyvkiadó (Budapest, 1978).

[2] Robin Pemantle, A survey of random processes with reinforcement, *Probability Surveys,* **4** (2007), 1–79.

[3] Jim Pitman, Combinatorial Stochastic Processes, *Saint Flour Lectures* (2002), Lecture Notes in Mathematics, Vol. 1875. Springer-Verlag (Berlin).

Rudas Anna
Budapesti Műszaki és Gazdaságtudományi Egyetem
Matematika- és Számítástudományok Doktori Iskola
rudasa@math.bme.hu