Alkalmazott Matematikai Lapok **27** (2010), 135-156.

A LEGHOSSZABB SZÉRIÁK VIZSGÁLATA [1] [2]

FAZEKAS ISTVÁN, KARÁCSONY ZSOLT[3], LIBOR JÓZSEFNÉ

1. Bevezetés

Számos mű foglalkozott már a címben megjelölt témával, bár a szóhasználatban léteznek eltérések. Vannak szerzők, akik leghosszabb futamnak, vagy siker sorozatnak, vagy egyszerűen leghosszabb sorozatnak nevezik egy adott kísérletsorozatban az egymást követő azonos jelek leghosszabb szériáját. Például az érmedobás kísérletben az egymás után következő – vagyis írással meg nem szakított – fejdobások számának a maximumát leghosszabb fejszériának fogjuk nevezni. Ebben a munkában ismert rekurziós és aszimptotikus tételeket, valamint a szimuláció szolgáltatta eredményeket fogjuk összehasonlítani. Így Erdős-Révész [4] és Földes [5] aszimptotikus eredményeit fogjuk összevetni Schilling [15], Bloom [3] és Kopocinski [8] általunk esetenként kiegészített és bizonyított rekurzív formuláival, valamint a szimulációs eredményeinkkel. Megjegyezzük, hogy Binswanger-Embrechts [2] az aszimptotikus tételeket vetette össze szimulációs eredményekkel. Ebben a cikkben a rekurziós eljárásokat hangsúlyozzuk, ezért ezeket részletesen bizonyítjuk. A rekurziós eljárások adják a pontos eredményt. Az aszimptotikus eredmények csak hosszú dobássorozat esetén adnak jó közelítést. A szimulációs eredmények pedig véletlenszerűek, és a dobássorozat nagyon sokszori számítógépes legenerálása esetén közelítik a pontos értéket. Tanulmányunk kiterjed a szabályos és szabálytalan érme esetére is, valamint mindkétféle érménél vizsgáljuk a leghosszabb fejszéria és a leghosszabb bármilyen (tiszta fej vagy tiszta írás) széria hosszát is. A szimulációkat a MATLAB programmal végeztük 20000 ismétlést alkalmazva, rövid ($n = 50$) és hosszú ($n = 1000$) sorozatok esetén. Munkánk során vizsgáltuk a visszatevés nélküli húzásokat is, melyet például a kártyalap-húzás kísérlettel tudunk szemléltetni. A kérdés itt is ugyanaz, hogyan alakul a leghosszabb azonos jelsorozat hossza (például a leghosszabb tiszta piros lapszéria hossza a francia kártya csomagból történt húzások során). Írásunkban megemlítünk néhány matematikatörténeti, didaktikai érdekességet is, melyek segítenek az egyetemi, főiskolai hallgatók érdeklődését felkelteni a téma iránt.

[1] *2000 Mathematics Subject Classification:* 97K50, 60C05, 60F05.
[2] *Key words and phrases:* széria, rekurzió, szimuláció
[3] A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

2. Független kísérletsorozat (visszatevéses mintavétel)

Bevezetésül tekintsük Varga Tamás egy érdekes kísérletét, melyet Révész Pál 1978-ban ismertetett Helsinkiben egy nemzetközi matematikai konferencián [12], majd Schilling [15] cikkének bevezetéseként is szolgált. Varga a tanulócsoportját két részre osztotta, majd az egyik csoportnak azt adta feladatul, hogy mindenki dobjon fel 200-szor egy pénzérmét, és jegyezze le a kapott fej-írás eredményeket. A csoport másik részének pedig a kísérletet csak gondolatban kellett elvégezni, és a gondolati eredményeket lejegyezni. Vagyis nekik olyan 200 elemű fej-írás sorozatot kellett írniuk, amilyen szerintük egy 200 elemű dobássorozat. A munka végeztével összekeverték a lejegyzett eredményeket tartalmazó lapokat, majd átadták Vargának, aki majdnem 100%-os biztonsággal megmondta, hogy az adott lap valós eredményt tükröz-e, vagy kitaláltat. Hiszen míg a valós sorozatokban nem volt ritka a 7 (esetleg 8) egymást követő fej – Rényi Alfréd jól ismert $\log_2 200$-as eredményével összhangban –, addig a képzelt sorozatokban maximum 5 egyforma elemet mertek a tanulók egymás után leírni.

Amikor volt szerencsénk Révész professzor úrral találkozni és beszélgetni erről, elmondta a kísérlet továbbvitelét is. Ő, miután ismertette a hallgatókkal a Varga-féle kísérletet, és az eredményt is megbeszélték, újra elvégeztette az eredeti kísérletet. Vagyis a hallgatók fele újra valós kísérletet végzett az érme 200-szori feldobásával, míg a csoport másik fele a gondolati eredményeit írta le. Az összegyűjtött papírlapokat újra sikerült majdnem teljes pontossággal szétválogatnia Révésznek. A magyarázat nagyon egyszerű. A hallgatók többsége csak az egyikféle, például a leghosszabb fejszériára koncentrált, de már nem figyelt a leghosszabb írásszériára. Felvetődik tehát az alábbi két kérdés.

i. Egy n hosszúságú sorozat esetén hogyan alakul a *leghosszabb fejszéria hossza*?

ii. Egy n hosszúságú sorozat esetén mekkora a *leghosszabb bármilyen széria (akár fej, akár írás) hossza*?

2.1. Szabályos pénzérme esete

2.1.1. Leghosszabb fejszéria vizsgálata

Schilling [15] cikke nyomán vizsgáljuk a következőt. Dobjunk fel egy szabályos pénzérmét n-szer. Fejszériának nevezzük az egymást követő (tehát írással meg nem szakított) fejek sorozatát. Jelölje R_n a leghosszabb fejszéria nagyságát. Az eloszlásfüggvényünk: $F_n(x) = P(R_n \leq x)$. $F_n(x)$-et elegendő nemnegatív egész x-ekre megadni (hiszen $F_n(x) = 0$, ha $x < 0$; $F_n(x) = F_n([x])$, ha $x \geq 0$, ahol $[x]$ jelöli x egészrészét). Legyen $A_n(x)$ azon n hosszúságú sorozatok száma, amelyekben a leghosszabb fejszéria nem haladja meg x-et. Szabályos érme esetén egy n elemű sorozatot vizsgálva kapjuk tehát:

$$F_n(x) = P(R_n \leq x) = \frac{A_n(x)}{2^n}.$$

De hogyan tudnánk meghatározni az $A_n(x)$ értékét?

Vegyük először azt az esetet, amikor a leghosszabb fejszéria legfeljebb 3 elemű ($x = 3$). Ha $n \leq 3$, akkor $A_n(3) = 2^n$, hiszen minden lehetséges eset megfelel annak a kritériumnak, hogy az egymás utáni fejek száma maximum 3. Ha viszont az $n > 3$, akkor a számunkra kedvező sorozatok kezdődhetnek a következőképpen (ahol F jelöli a fej, I pedig az írás dobást): I, FI, FFI, FFFI, és utánuk csak olyan jelsorozat van, amelyben nincs háromnál hosszabb fejszéria. Kapjuk tehát a következőt:

$$A_n(3) = A_{n-1}(3) + A_{n-2}(3) + A_{n-3}(3) + A_{n-4}(3), \quad \text{ha } n > 3.$$

Ugyanezzel a módszerrel adódik az általános rekurziós képlet.

2.1. Állítás. *(Schilling [15], 198. o.)*

$$A_n(x) = \begin{cases} \sum_{j=0}^{x} A_{n-1-j}(x), & \text{ha } n > x, \\ 2^n, & \text{ha } 0 \leq n \leq x. \end{cases}$$

Megjegyzés. Ha megnézzük $A_n(1)$ értékeit, vagyis azon n elemű sorozatok számát, melyekben legfeljebb 1 hosszúságú fejszéria van, éppen a Fibonacci-sorozat (azaz $a_0 = 0$, $a_1 = 1$, és $a_n = a_{n-1} + a_{n-2}$, ha $n \geq 2$) 2-vel eltolt elemeit kapjuk.

n	0	1	2	3	4	5	6	7	8	...
$A_n(1)$	1	2	3	5	8	13	21	34	55	...

A k-rendű Fibonacci-számok segítségével kifejezhető $A_n(k)$, sőt a k-rendű Fibonacci-polinomok felhasználásával a szabálytalan pénzérme esete is kezelhető, lásd [10].

R_n aszimptotikus viselkedését Földes Antónia [5] alábbi tétele alapján írhatjuk le.

2.1. Tétel. *(Földes [5].) Valamennyi egész k esetén*

$$P\left(R_n - \left[\frac{\log n}{\log 2}\right] < k\right) = \exp\left(-2^{-\left(k+1-\left\{\frac{\log n}{\log 2}\right\}\right)}\right) + o(1), \tag{1}$$

ahol $[a]$ jelöli az egészrészét a-nak és $\{a\} = a - [a]$, a törtrésze.

Itt log *a természetes alapú logaritmust jelöli.*

2.1.2. Leghosszabb bármilyen széria vizsgálata

Ismét alapul vesszük Schilling [15] cikkét. Dobjunk fel egy szabályos pénzérmét n-szer, és jelölje R'_n a leghosszabb széria (akár a tiszta fej, akár a tiszta írás) nagyságát. Legyen $B_n(x)$ azon n hosszúságú sorozatok száma, amelyekben a leghosszabb (tetszőleges) széria nem haladja meg x-et. Szabályos érme esetén egy n elemű sorozatot vizsgálva, kapjuk az eloszlásfüggvényt:

$$F'_n(x) = P(R'_n \leq x) = \frac{B_n(x)}{2^n}.$$

Most Schilling [15], 199. oldalon leírt ötletét használjuk. A fej-írás sorozat minden eleme alatt jelölje A azt, hogy az utána következő vele azonos, illetve K azt, hogy különböző. Például:

F F F I F I F I I I I F F
A A K K K K K A A A K A

Az alsó A, K elemekből álló sorozatban a leghosszabb tiszta A sorozat akkor és csak akkor $k-1$ hosszúságú, ha fölötte a leghosszabb tiszta széria (fej vagy írás) k hosszúságú. Ha a felső sorozat n hosszú, és ebben a leghosszabb széria k elemű, akkor az alsó sorozat $n-1$ hosszú, és a leghosszabb A széria $k-1$ elemű. Vagyis

$$B_n(x) = 2A_{n-1}(x-1)$$

(hiszen minden alsó sorozat pontosan 2 felső sorozathoz tartozhat). Ennek felhasználásával kapjuk:

$$\begin{aligned} F'_n(x) = P(R'_n \le x) = \frac{B_n(x)}{2^n} = \frac{2A_{n-1}(x-1)}{2^n} = \frac{A_{n-1}(x-1)}{2^{n-1}} = \\ = P(R_{n-1} \le x-1) = F_{n-1}(x-1). \end{aligned}$$

Beláttuk tehát, hogy

$$F'_n(x) = F_{n-1}(x-1), \tag{2}$$

vagyis visszavezettük esetünket a tiszta fejszéria esetére.

Most vizsgáljuk azt, hogy a leghosszabb széria pontosan k hosszúságú. Jelölések:

$b_n(k)$: n dobásból hányszor lesz a leghosszabb széria (akár fej, akár írás) *pontosan* k hosszúságú,

$a_n(k)$: n dobásból hányszor lesz a leghosszabb fejszéria *pontosan* k hosszúságú.

A $b_n(k)$-ra Szászné Simon Judit is ad [18] doktori értekezésében rekurzív képletet, melyet mi kétféleképpen bizonyítunk. Először a (7) képletből vezetjük le, majd teljes eseményrendszerre bontással kapjuk meg az adott képletet. A (7) formulát a következő szakaszban igazoljuk csak, de természetesen ehhez nem fogunk támaszkodni a jelen szakaszra.

2.2. Állítás. *(Szászné [18].) Minden $n = 1, 2, \dots$ esetén $b_n(1) = b_n(n) = 2$, $b_n(r) = 0$, ha $r > n$ vagy $r \le 0$*

$$b_n(r) = \sum_{h=1}^{r-1} b_{n-h}(r) + \sum_{i=1}^{r} b_{n-r}(i), \quad \text{ha} \quad 1 < r < n. \tag{3}$$

Első bizonyítás. Tekintsük a (7) összefüggésen a $p = q = \frac{1}{2}$ esetet. Mivel az összes elemi esemény száma 2^n, ezért a $p(n, k)$ kifejezés 2^n-nel való beszorzásával adódik a számunkra kedvező esetek száma, vagyis $a_n(k)$. Azaz (7)-ből

$$\underbrace{2^n p(n, k)}_{a_n(k)} = \sum_{j=0}^{k-1} \underbrace{2^{n-j-1} p(n-j-1, k)}_{a_{n-j-1}(k)} + \underbrace{2^{n-k-1} F_{n-k-1}(k)}_{A_{n-k-1}(k)}. \tag{4}$$

Alkalmazzuk újra Schilling ötletét, vagyis: a fej-írás sorozat minden eleme alatt jelölje A azt, hogy az utána következő vele azonos, és K pedig azt, hogy különböző. Ahogyan adódott $B_n(k) = 2A_{n-1}(k-1)$, ugyanúgy adódik $b_n(k) = 2a_{n-1}(k-1)$. Ennek felhasználásával (4) képletünk a következőképpen alakul:

$$\frac{b_{n+1}(k+1)}{2} = \sum_{j=0}^{k-1} \frac{b_{n-j}(k+1)}{2} + \frac{B_{n-k}(k+1)}{2}.$$

Végezzük el a 2-vel való beszorzást és alkalmazzuk a következő átindexelést:
$k+1 \to r, \quad n+1 \to n, \quad j+1 \to h$. Kapjuk a bizonyítandó formulát.

Második bizonyítás. Nézzük meg, hogy mit is fejez ki a (3) jobb oldalán lévő két összeg. Bontsuk fel az eseményterünket aszerint, hogy milyen hosszú széria van elől.

Ha a sorozatunk h ($h = 1,\ 2,\ \ldots,\ r-1$) hosszúságú szériával kezdődik, akkor a maradék $n - h$ dobásból kell a leghosszabb szériának r hosszúságúnak lenni. Ha a kezdő h széria fej, akkor a $(h+1)$-edik elem írás kell hogy legyen. Ha a kezdő h széria írás, akkor a $(h+1)$-edik elem fej kell hogy legyen.

$$\underbrace{\mathrm{F}\ldots\ \mathrm{F}}_{\substack{h \text{ db fej} \\ (1 \le h < r)}}\ \underbrace{\mathrm{I}\ldots\ \mathrm{F}\ \ldots\ \mathrm{F}\ \ldots}_{\substack{n-h \text{ elem között} \\ r \text{ hosszú széria}}}$$

vagy:

$$\underbrace{\mathrm{I}\ldots\ \mathrm{I}}_{\substack{h \text{ db írás} \\ (1 \le h < r)}}\ \underbrace{\mathrm{F}\ldots\ \mathrm{I}\ \ldots\ \mathrm{I}\ \ldots}_{\substack{n-h \text{ elem között} \\ r \text{ hosszú széria}}}$$

De ennek a kettőnek a száma megegyezik, és az összegük éppen $b_{n-h}(r)$. Ha r hosszúságú szériával kezdődik a sorozat, akkor a maradék $n - r$ elemből $i = 1,\ 2,\ \ldots,\ r$ hosszú széria lehet. Ha az első r elem fej, akkor az $(r+1)$-edik írás kell, hogy legyen, míg ha az első r elem írás, akkor az $(r+1)$-edik fej kell hogy legyen.

$$\underbrace{\mathrm{F}\ldots\ \mathrm{F}}_{r \text{ db fej}}\ \underbrace{\mathrm{I}\ldots\ \mathrm{F}\ \ldots\ \mathrm{F}\ \ldots}_{\substack{n-r \text{ elem között} \\ \text{legfeljebb } r \text{ hosszú széria}}}$$

vagy:

$$\underbrace{\text{I} \ldots \text{ I}}_{r \text{ db írás}} \quad \underbrace{\text{F} \ldots \text{I} \ldots \text{I} \ldots}_{\substack{n-r \text{ elem között} \\ \text{legfeljebb } r \text{ hosszú széria}}}$$

Ezen két részeset száma egyenlő, összegük pedig $\sum_{i=1}^{r} b_{n-r}(i)$. □

A $b_n(r)$-rel tehát megadtuk azon sorozatok számát, melyekben az n dobás során a leghosszabb bármilyen széria (akár fej, akár írás) pontosan r hosszúságú.

R'_n aszimptotikus viselkedésének vizsgálatához Földes [5] már idézett eredményét használjuk. (1) és (2) alapján kapjuk az alábbit.

2.2. Tétel. *Valamennyi egész k esetén*

$$P\left(R'_n - \left[\frac{\log(n-1)}{\log 2}\right] < k\right) = \exp\left(-2^{-\left(k-\left\{\frac{\log(n-1)}{\log 2}\right\}\right)}\right) + o(1),$$

ahol - mint korábban - $[a]$ az a egészrészét, $\{a\}$ pedig az a törtrészét jelöli.

2.1.3. Numerikus eredmények

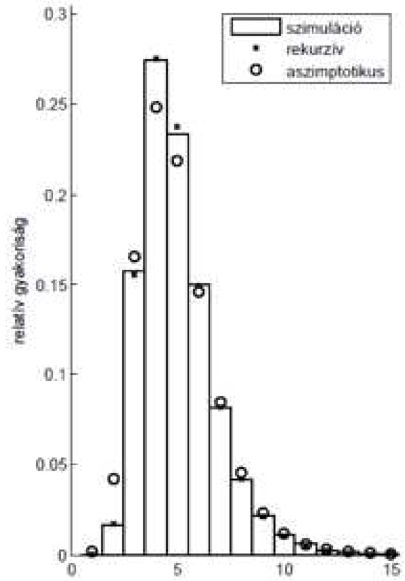
Vizsgálatunkhoz a MATLAB programot használtuk 20.000 ismétlésszámmal. Az alkalmazott számítógép paraméterei pedig a következők: INTEL Core Quad Q9550 processzor, 4Gb, DDR3 memória.

A következő ábrákon „x" jelöli a rekurzióval kapott eredményeket, „o" az aszimptotikus tételek eredményeit, az oszlopdiagram pedig a szimulációval kapott eredményeket mutatja. Az első grafikonpár a rövid ($n = 50$) sorozat eredményeit mutatja a leghosszabb fej (bal oldali) és a leghosszabb bármilyen (jobb oldali) széria esetén, majd a további grafikonpár ugyanezt a két esetet mutatja hosszú ($n = 1000$) dobáshosszra vonatkozóan.

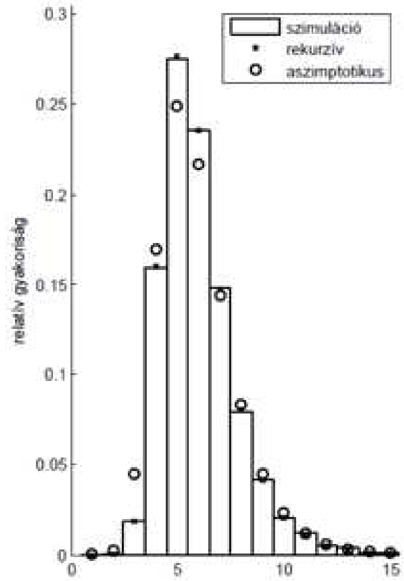
Mindkét esetre (leghosszabb fej, illetve leghosszabb bármilyen széria vizsgálatára) elmondható, hogy kis n esetén a szimulációs eredmények vannak közelebb a rekurzív eredményekhez, az aszimptotikus tételek n növelésével adják a rekurzióhoz közeli eredményeket. $n \geq 3000$ esetén a rekurziós, a szimulációs és az aszimptotikus értékek gyakorlatilag egybeesnek. Míg kis n esetén a rekurziós algoritmus gyors, n növelésével rohamosan lassul a számítási eljárás. A futási időket tekintve csak példaként néhányat megemlítve:

n	ism	futási idő
10.000	20.000	31.795.056 s.
1.000	20.000	3.984.981 s.
50	20.000	2.092.010 s.

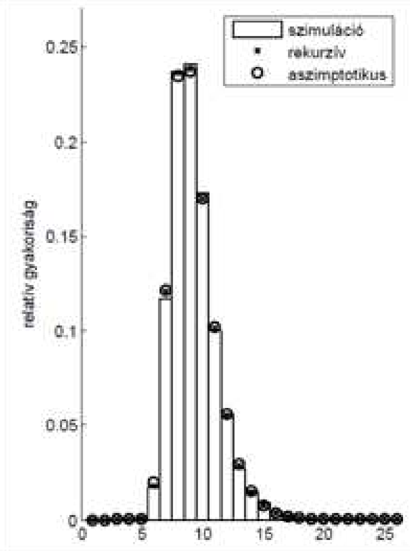
A párbaállított grafikonokon jól látszik a 2.1.2-ben leírt eredmény, miszerint R'_n eloszlása (közelítőleg) az R_n eloszlásából 1 egységgel jobbra való eltolással adódik.



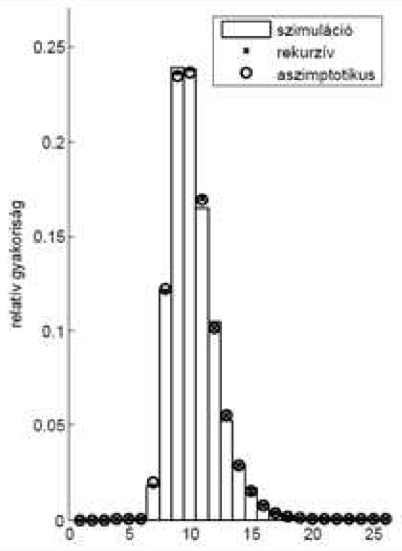
Leghosszabb fejszéria
$p = 0,5; n = 50.$



Leghosszabb széria
$p = 0,5; n = 50.$

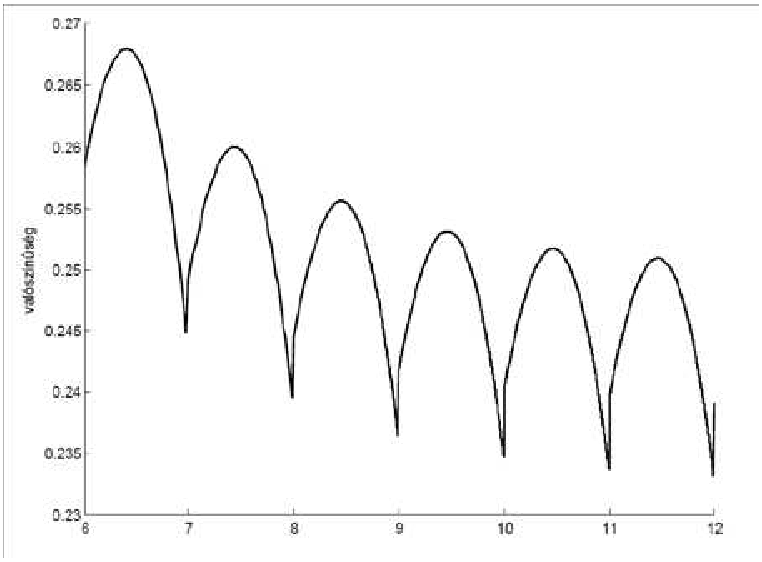


Leghosszabb fejszéria
$p = 0,5; n = 1000.$

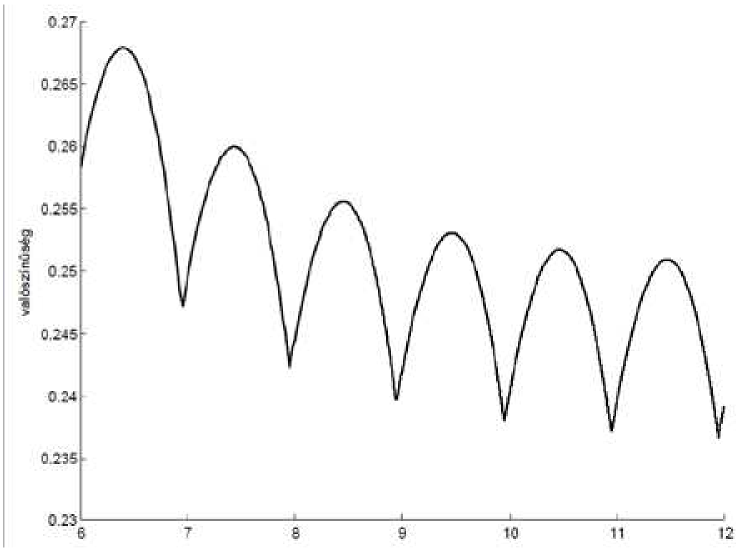


Leghosszabb széria
$p = 0,5; n = 1000.$

A következő ábrákról R_n periodikus jellegű viselkedése olvasható le, ahol a vízszintes tengelyen logaritmikus skálát alkalmaztunk, a függőleges tengelyen pedig a $P(R_n = [\log_2 n] - 1)$ értékek szerepelnek. Látható, hogy a P értékek ábrázolásakor 2 minden egész kitevős hatványa helyeken ugrás van (a rekurziós képlettel számolva).



$$P\left(R_n = [\log_2 n] - 1\right), \; n = 2^6, \ldots, 2^{12}, \; p = 0,5$$



$$\max_k P\left(R_n = k\right), \; n = 2^6, \ldots, 2^{12}, \; p = 0,5$$

A fenti két ábra alapján $\max_k P(R_n = k)$ értékei eltérnek $P(R_n = [\log_2 n] - 1)$-től, azaz R_n módusza nem minden n-re lesz $[\log_2 n] - 1$.

2.2. Szabálytalan pénzérme esete

Ebben az esetben a fejdobás valószínűsége, azaz p értéke bármilyen valós szám lehet a $(0, 1)$ intervallumból. Kérdés, hogy ez a tény milyen hatással van a leghosszabb fej-, illetve leghosszabb bármilyen széria alakulására. Most nyilván nem számolhatunk a klasszikus (kedvező/összes) képlettel.

Jelölje p a fejdobás valószínűségét és $q(= 1 - p)$ az írás valószínűségét.

2.2.1. Leghosszabb fejszéria vizsgálata

Schilling [15] alapján tekintsük azon n hosszúságú fej-írás sorozatokat, amelyekben k db fej van. Ezek közül jelentse $C_n^{(k)}(x)$ azon sorozatok számát, amelyekben legfeljebb x fej következik egymás után. (Azaz a leghosszabb fejszéria legfeljebb x hosszúságú.) Az adott jelölésekkel az eloszlásfüggvényünk a következő lesz:

$$F_n(x) = P(R_n \le x) = \sum_{k=0}^{n} C_n^{(k)}(x) p^k q^{n-k}. \tag{5}$$

2.3. Állítás. *(Schilling [15], 200.o.)*

$$C_n^{(k)}(x) = \begin{cases} \displaystyle\sum_{j=0}^{x} C_{n-1-j}^{(k-j)}(x), & \textit{ha} \;\; x < k < n, \\ \binom{n}{k}, & \textit{ha} \;\; 0 \le k \le x, \\ 0, & \textit{ha} \;\; x < k = n. \end{cases} \tag{6}$$

Bizonyítás. Ha $x < k = n$, akkor nyilvánvalóan $C_n^k(x) = 0$, hiszen ekkor az összes (x-nél több) elem fej, így nincs olyan sorozat, ahol legfeljebb x fej van egymás után. Ha $0 \le k \le x$, akkor $C_n^k(x)$ éppen a binomiális együtthatókat adja, hiszen ez az az eset, amikor az n elem között legfeljebb x fej van, és azon eseteket kell összeszámlálni, amikor a leghosszabb fejszéria legfeljebb x. Márpedig ekkor az összes sorrend ilyen tulajdonságú. Ha $x < k < n$, akkor a (6) képlet helyességének belátásához (melyet Schilling [15]-ban nem közöl) elegendő átgondolnunk a következőket. A sorozatunk kezdődhet $j = 0, 1, 2, \ldots, x$ fejjel, utána biztosan van 1 írás, majd olyan sorozat következik, ahol a maradék $n - j - 1$ elem között $k - j$ db fej van úgy, hogy a leghosszabb fejszéria legfeljebb x hosszúságú.

$$\underbrace{\text{F} \;\ldots\; \text{F}}_{j \text{ db fej}} \quad \text{I} \quad \underbrace{\;\ldots\quad \text{F} \quad\ldots\quad \text{I} \quad\ldots\;}_{\substack{n-j-1 \text{ elem, melyben } k-j \text{ db fej van úgy,} \\ \text{hogy legfeljebb } x \text{ hosszú a fejszéria}}}$$

Ezek száma pedig éppen $C_{n-1-j}^{(k-j)}(x)$. Ezzel (6) helyességét beláttuk. □

Kiszámolva $C_n^{(k)}(3)$ értékeit $n \leq 8$ esetre az alábbi értékek adódnak:

8									0
7								0	0
6							0	1	10
5						0	***2***	***12***	40
4					0	***3***	12	31	65
3				1	***4***	10	20	35	56
2			1	***3***	6	10	15	21	28
1		1	2	3	4	5	6	7	8
0	1	1	1	1	1	1	1	1	1
k/n	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

Észrevehetjük, hogy az alsó négy sor $\big(k = 0, 1, 2, 3$ esetben $\binom{n}{k}$ értékek$\big)$ a Pascal-háromszög részletét adja, $3 < k = n$ esetben pedig beírva a 0-kat, a táblázat többi adata a rekurziós képlet alapján számítható. Látható, hogy a fenti (6) képletet jól mutatja az ún. „hokiütő” forma. Hiszen pl. $C_7^{(5)}(3) = 2 + 3 + 4 + 3 = 12$. (Táblázatban dőlt, vastag számmal jelölve.)

Nézzük most, hogy mi lesz annak a valószínűsége, hogy n dobásból a leghosszabb fejszéria pontosan k hosszúságú? Jelöljük ezt $p(n, k)$-val, melyre Kopocinski [8]-ban két formulát is ad. Az alábbi (7) és (8) képlet alkalmazhatóságához megjegyezzük, hogy az (5)-beli F függvényre:

$$F_n(k) = \begin{cases} 0, & \text{ha} \quad k < 0, \\ 1, & \text{ha} \quad k \geq n, \\ \sum_{i=0}^{k} p(n, i) & \text{egyébként.} \end{cases}$$

2.3. Tétel. *(Kopocinski [8], Theorem 1, (2), (3).) Legyen $p(n, k) = 0$, ha $k < 0$, vagy $k > n$, $p(k, k) = p^k$, ha $k = 0, 1, 2, \ldots$ Ekkor*

$$p(n, k) = \sum_{j=0}^{k-1} p^j q p(n - j - 1, k) + p^k q F_{n-k-1}(k), \tag{7}$$

$$\begin{aligned} p(n, k) = p^k q F_{n-k-1}(k) + &\sum_{j=0}^{n-k-2} F_j(k-1) q^2 p^k F_{n-j-k-2}(k) + \\ &+ F_{n-k-1}(k-1) q p^k, \end{aligned} \tag{8}$$

ha $n = k+1, k+2, \ldots$, $k = 0, 1, 2, \ldots$, és $F_n(k)$ jelöli annak a valószínűségét – az eloszlásfüggvény definíciójának megfelelően –, hogy n esetből legfeljebb k hosszúságú fejszéria adódik.

Bizonyítás. A (7) képlet helyességét vizsgáljuk teljes eseményrendszer szerinti részekre bontással aszerint, hogy az első helyeken lehet $0, 1, 2, \ldots, k$ db fej. Lehet olyan, hogy az első j helyen fejet kapunk, $(0 \leq j < k)$, majd jön 1 írás és az utána lévőkben van pontosan k hosszú fejszéria:

$$\underbrace{\mathrm{F} \ldots \mathrm{F}}_{\substack{j \text{ db fej} \\ 0 \leq j \leq k-1}} \quad \mathrm{I} \quad \underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{\substack{\text{az } n-j-1 \text{ elem között} \\ \text{pontosan } k \text{ hosszú fejszéria van}}}$$

Ebből az esetből adódik a (7) összeg első k tagja, azaz $\sum_{j=0}^{k-1} p^j qp(n-j-1,k)$. Vagy pedig lehet az az eset, hogy rögtön az elején adódik k hosszúságú fejszéria, utána 1 írás, majd legfeljebb k hosszúságú fejszéria:

$$\underbrace{\mathrm{F} \ldots \mathrm{F}}_{k \text{ db fej}} \quad \mathrm{I} \quad \underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{\substack{n-k-1 \text{ elem melyben} \\ \text{legfeljebb } k \text{ hosszú fejszéria van}}}$$

Ez pedig éppen a (7) összefüggés utolsó tagja, amivel a képlet helyességét beláttuk.

A (8) bizonyításához bontsuk fel az eseményterünket aszerint, hogy az első k hosszúságú fejszéria hol kezdődik. Lehet olyan eset, hogy rögtön az első k dobás fej, utána 1 írás, majd legfeljebb k hosszú fejszéria következik:

$$\underbrace{\mathrm{F} \ldots \mathrm{F}}_{k \text{ db fej}} \quad \mathrm{I} \quad \underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{n-k-1 \text{ elem között legfeljebb } k \text{ hosszú fejszéria van}}$$

Ez éppen az összeg első tagja: $p^k q F_{n-k-1}(k)$. Lehet még olyan, hogy legfeljebb $k-1$ fej van az elején, utána 1 írás kell, hogy legyen, majd következik a k db fej, utána megint 1 írás, és végül legfeljebb k db fej:

$$\underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{\text{legfeljebb } k-1 \text{ db fej}} \quad \mathrm{I} \quad \underbrace{\mathrm{F} \ldots \mathrm{F}}_{k \text{ db fej}} \quad \mathrm{I} \quad \underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{\substack{\text{az } n-j-k-2 \text{ elem között} \\ \text{legfeljebb } k \text{ db fej van}}}$$

Ez éppen az összeg második része: $\sum_{j=0}^{n-k-2} F_j(k-1) q^2 p^k F_{n-j-k-2}(k)$. Vagy lehet az az eset, amikor az utolsó k db lesz fej, előtte 1 írás és a kezdőszériában legfeljebb $k-1$ fej van:

$$\underbrace{\ldots \quad \mathrm{F} \quad \ldots \quad \mathrm{I} \quad \ldots}_{\text{legfeljebb } k-1 \text{ fej}} \quad \mathrm{I} \quad \underbrace{\mathrm{F} \ldots \mathrm{F}}_{k \text{ db fej}}$$

Ez éppen az összeg utolsó tagját adja: $F_{n-k-1}(k-1) qp^k$.

Így a (8) formula helyességét is beláttuk. (A (7) és (8) bizonyítását Kopocinski nem végzi el, csak [8] 5. oldalán útmutatást ad.) $\square$

Az aszimptotikus viselkedés leírását Gordon-Schilling-Waterman [7] adja a következő tétellel.

2.4. TÉTEL. *(Gordon-Schilling-Waterman [7].) Legyen* $\mu(n) = -\frac{\log n}{\log p}, q = 1-p$ *és legyen* W *olyan, hogy teljesüljön rá:* $(P(W \leq t) = \exp(-\exp(-t)))$, *ekkor* t*-ben egyenletesen:*

$$P\left(R_n - \mu(qn) \leq t\right) - P\left(\left[\frac{W}{-\log p} + \{\mu(qn)\}\right] - \{\mu(qn)\} \leq t\right) \to 0,$$

ahol $n \to \infty$. *(Az előzőekhez hasonlóan* [ ] *az egészrész,* { } *pedig a törtrész jele.)*

2.2.2. Leghosszabb bármilyen széria vizsgálata

Egy szabálytalan pénzérmét feldobva n-szer, jelölje R'_n a leghosszabb bármilyen széria (akár fej, akár írás) nagyságát. Az (5) képlethez hasonlóan adódik az alábbi (lásd Schilling [15], 200. oldal).

2.4. ÁLLÍTÁS. *(Schilling [15].)*

$$F'_n(x) = P(R'_n \leq x) = \sum_{k=0}^{n} \overline{C}_n^{(k)}(x) p^k q^{n-k},$$

ahol $\overline{C}_n^{(k)}(x)$ *jelenti azon* n *hosszúságú sorozatok számát, amelyben* k *db fej van, a leghosszabb bármilyen széria hossza legfeljebb* x, *és ahol*

$$\overline{C}_{m+k}^{(k)}(x) = C_{x+1}(m,k),$$

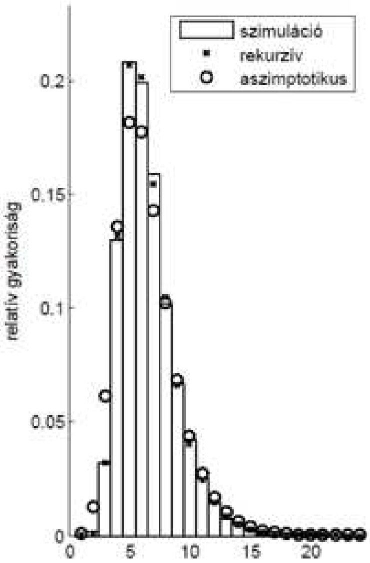
valamint a $C_x(m,k)$ *mennyiségek kielégítik a* (9) *és* (10) *rekurziókat.*

Itt $C_x(m,k)$ jelöli, hogy m piros és k fekete golyót visszatevés nélkül kihúzva nem lesz x hosszúságú széria. $C_t(m,k)$ értékeire Bloom [3] adott két rekurzív képletet, melyek bizonyításait a 3. fejezetben adjuk meg ((9) és (10)).

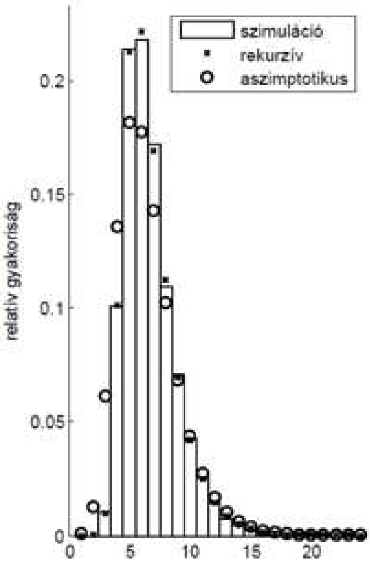
Megjegyzés. R'_n aszimptotikus viselkedését vizsgálva Muselli [9] tételét használhatjuk, melyben $V_n(p)$ jelöli annak a valószínűségét, hogy a leghosszabb széria n dobás esetén a fejekből adódik:

$$\lim_{n\to\infty} V_n(p) = \begin{cases} 0, & \text{ha } 0 \leq p < 1/2, \\ 1, & \text{ha } 1/2 < p \leq 1. \end{cases}$$

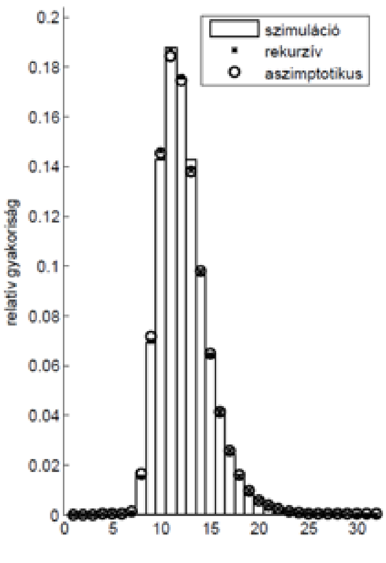
Azaz 1-hez tart annak a valószínűsége, hogy a leghosszabb széria az esélyesebb kimenetelből, azaz jelen esetben fejekből áll.



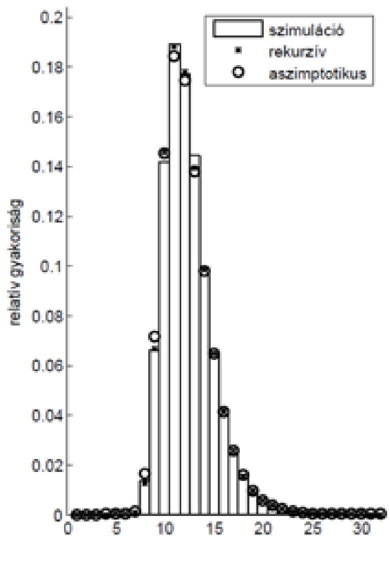
Leghosszabb fejszéria
$p = 0,6; n = 50.$



Leghosszabb széria
$p = 0,6; n = 50.$



Leghosszabb fejszéria
$p = 0,6; n = 1000.$



Leghosszabb széria
$p = 0,6; n = 1000.$

2.2.3. Numerikus eredmények

A szimulációkat ugyanolyan tárgyi és szoftver eszközökkel végeztük, mint azt a 2.1.3-ban leírtuk. Ugyanazokat a jelöléseket alkalmaztuk, és ismét 20.000 ismétlést végeztünk. Az előző grafikonokon szabálytalan érme esetén mutatjuk a leghosszabb fej (bal oldali) és a leghosszabb bármilyen (jobb oldali) szériák esetét, sorban rövid ($n = 50$) és hosszú ($n = 1000$) dobássorozat esetén. Kis n esetén az aszimptotikus eredmények távol esnek a pontos (rekurziós) értékektől, nagy n esetén közel kerülnek hozzájuk. A 2.4. állítást követő megjegyzés numerikus alátámasztását adja, hogy nagy n esetén R_n és R'_n eloszlása közel azonos.

3. Nem független kísérletsorozat (visszatevés nélküli mintavétel)

Gardner [6] könyvében szerepel az alábbi állítás: "Az 52 lapos összekevert kártyacsomagban majdnem mindig lesz 6 egyforma színű egymás után." Ahogyan [3]-ben is szerepel, elvégezve többször is ezt a kísérletet nem tapasztaltuk ezt az eredményt. 6-nál általában kevesebb elemű szériák adódtak. Csak nem volt szerencsénk, vagy a szerző tévedett?

Felvetődik az alábbi kérdés. Ha egy halmaz kétféle tulajdonságú elemet tartalmaz, az egyikből m, a másikból k db-ot, mi a valószínűsége annak, hogy az $m + k$ elemet sorban egymás után kihúzva visszatevés nélkül, lesz t hosszúságú széria, vagyis akármelyik tulajdonságú elemből legalább t következik egymás után?

Az egyszerűség kedvéért a két tulajdonságú elem legyen piros (m db) és fekete (k db), és jelöljük a keresett valószínűséget $P_t(m, k)$-val. Meghatározásához vizsgáljuk az esemény komplementerének, vagyis annak a valószínűségét, hogy nincs t hosszúságú széria az $m + k$ elem sorozatában. Ez legyen: $\overline{P_t}(m, k)$, aminek a segítségével a $P_t(m, k) = 1 - \overline{P_t}(m, k)$ képlet alapján adódik kérdésünkre a válasz. $\overline{P_t}(m, k)$-nek klasszikus képlettel való kiszámításához vizsgáljuk először az összes elemi esemény számát: az $m + k$ elemet kell sorbarendezni, melyek között az m db és a k db azonos típusúak. Az ilyen sorozatok száma nem más mint: $\binom{m+k}{m}$.

Ezek után a keresett hányados számlálójának meghatározásához össze kell számlálnunk azon sorozatok számát, amelyben nincs t hosszúságú széria. Jelöljük ezt $C_t(m, k)$-val. Először számoljuk ki $C_t(m, k)$ értékeit $t = 3$ és nem túl nagy (10-nél kisebb) m és k esetén:

m\k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	1	1	1	0	0	0	0	0	0	0
1	1	2	3	2	1	0	0	0	0	0
2	1	3	6	7	6	3	1	0	0	0
3	0	2	7	***14***	***18***	16	10	4	1	0
4	0	1	6	18	34	45	43	30	15	5
5	0	0	3	***16***	***45***	***84***	113	114	87	50
6	0	0	1	10	43	***113***	208	285	300	246
7	0	0	0	4	30	114	285	518	720	786
8	0	0	0	1	15	87	300	720	1296	1823
9	0	0	0	0	5	50	246	786	1823	3254

3.1. ÁLLÍTÁS. *(Bloom [3], 369. o.) Ha $m = k = 0$, akkor definíció szerint legyen $C_t(0,0) = 1$. Ha m vagy k negatív, akkor pedig definíció szerint $C_t(m,k) = 0$.*

$$C_t(m,k) = \sum_{i=0}^{t-1} C_t(m-1,k-i) - \sum_{i=1}^{t-1} C_t(m-t,k-i) + e_t(m,k), \tag{9}$$

ahol tehát $C_t(m,k)$ jelenti az m db piros és k db fekete elem olyan sorbarendezéseinek a számát, ahol nincs t hosszúságú széria ($t \geq 2$), valamint

$$e_t(m,k) = \begin{cases} 1, & \text{ha } m = 0 \text{ és } 0 \leq k < t, \\ -1, & \text{ha } m = t \text{ és } 0 \leq k < t, \\ 0, & \text{különben.} \end{cases}$$

Igazoljuk (9)-et, melyet Bloom [3]-ben nem végzett el.

Bizonyítás.

$\underline{m = 0}$ eset.
Ha $0 \leq k < t$, akkor $C_t(0,k) = 1$, hiszen ez azt jelenti, hogy csak egyféle elem van, de kevesebb, mint a szériahossz. Ezeket akárhogyan is húzom, nem lehet t hosszúságú széria, és mivel az azonos elemek egymás között nem megkülönböztethetők, ezért ez egyféle sorrendet jelent. Így a (9) képletünk a következő alakú: $1 = 0 - 0 + 1$. Például:

$$C_3(0,2) = C_3(-1,2) + C_3(-1,1) + C_3(-1,0) - [C_3(-3,1) + C_3(-3,0)] + 1 =$$
$$= 0 + 0 + 0 - [0 + 0] + 1 = 1.$$

Ha $k \geq t$, akkor $C_t(0,k) = 0$, hiszen ekkor is csak egyféle elem van, de legalább annyi, mint a szériahossz. Ekkor nyilván egyetlen olyan sorozat sincs, melyben ne lenne t hosszúságú széria. A (9) a következő: $0 = 0 - 0 + 0$. Például:

$$C_3(0,4) = C_3(-1,4) + C_3(-1,3) + C_3(-1,2) - [C_3(-3,3) + C_3(-3,2)] + 0 =$$
$$= 0 + 0 + 0 - [0 + 0] + 0 = 0.$$

$\underline{0 < m < t}$ esetén (9) alakja:

$$C_t(m,k) = \sum_{i=0}^{t-1} C_t(m-1,k-i) - 0 + 0.$$

Hiszen ez a következő eset:

$$\underbrace{\text{F} \ldots \text{F}}_{i \text{ db fekete}} \quad \text{P} \quad \underbrace{\ldots \quad \text{F} \quad \ldots \quad \text{P} \quad \ldots}_{\substack{k-i \text{ db fekete, } m-1 \text{ db piros,} \\ \text{és közöttük nincs } t \text{ széria}}}$$

Ezek száma $C_t(m-1,k-i)$. Mivel $m < t$, így pirosból t széria nem lehet, azaz a fenti összegből nem kell levonni semmit. Példaként tekintsük:

$$\begin{aligned} C_3(2,2) &= C_3(1,2) + C_3(1,1) + C_3(1,0) - [C_3(-1,1) + C_3(-1,0)] + 0 = \\ &= 3 + 2 + 1 - [0 + 0] + 0 = 6. \end{aligned}$$

Ha $\underline{m = t}$ és $0 \leq k < t$, ez azt jelenti, hogy az egyik fajta elemből (feketéből) kevesebb, a másikból (pirosból) pedig pontosan annyi áll rendelkezésre, mint a szériahossz. Ekkor (9) alakja a következő:

$$C_t(m,k) = \sum_{i=0}^{t-1} C_t(m-1,k-i) - \sum_{i=1}^{t-1} C_t(0,k-i) - 1.$$

Az első szumma nem t, hanem csupán $k+1$ részesetet jelent, melyek i-edik tagja olyan, hogy i db feketével kezdődik, utána 1 piros, majd $m-1$ piros és $k-i$ fekete úgy, hogy nincs t széria.

$$\underbrace{\text{F} \ldots \text{F}}_{i \text{ db fekete}} \quad \text{P} \quad \underbrace{\ldots \quad \text{F} \quad \ldots \quad \text{P} \quad \ldots}_{\substack{k-i \text{ db fekete, } m-1 \text{ db piros,} \\ \text{és közöttük nincs } t \text{ széria}}}$$

Ebben egy „rossz” elem van, amikor az $m = t$ piros egymás után helyezkedik el. Mivel a második szumma értéke k, így összesen $k+1$ levonása történik. Például:

$$\begin{aligned} C_3(3,2) &= C_3(2,2) + C_3(2,1) + C_3(2,0) - [C_3(0,1) + C_3(0,0)] - 1 = \\ &= 6 + 3 + 1 - [1 + 1] - 1 = 7. \end{aligned}$$

Ha $m = t$ és $k \geq t$, akkor (9) a következő:

$$C_t(m,k) = \sum_{i=0}^{t-1} C_t(m-1,k-i) - \sum_{i=1}^{t-1} C_t(0,k-i) + 0.$$

Ebben $i = 0$ esetén:

$$\text{P} \quad \underbrace{\ldots\ \text{P}\ \ldots\ \text{F}\ \ldots}_{\substack{k \text{ db fekete, } m-1 \text{ piros} \\ \text{és nincs közöttük } t \text{ széria}}}$$

Ezek száma $C_t(m-1,k)$. Ebben lehetne egy „rossz" is, amikor az $m-1$ piros van elől, és a legelső pirossal t szériát alkot. De ekkor a k fekete a végén állva t szériát alkotna, vagyis ez a rossz szituáció már nincs benne $C_t(m-1,k)$-ban.
Illetve $i = 1, 2, \ldots, t-1$ esetén:

$$\underbrace{\text{F}\ \ldots\ \text{F}}_{i \text{ db fekete}} \quad \text{P} \quad \underbrace{\ldots \quad \text{F} \quad \ldots \quad \text{P} \quad \ldots}_{\substack{k-i \text{ db fekete, } m-1 \text{ db piros,} \\ \text{és közöttük nincs } t \text{ széria}}}$$

Ezek száma $C_t(m-1,k-i)$. De ebben lehet egy „rossz" eset, amikor mind a $t = m$ piros egymás mellett van, így le kell vonni $C_t(0,k-i)$ mennyiséget, ami lehet 0 is. Nézzük például:

$$\begin{aligned} C_3(3,4) &= C_3(2,4) + C_3(2,3) + C_3(2,2) - [C_3(0,3) + C_3(0,2)] + 0 = \\ &= 6 + 7 + 6 - [0 + 1] + 0 = 18. \end{aligned}$$

Ha $\underline{m > 0 \text{ és } m > t}$, akkor a következőképpen kaphatunk olyan sorozatokat, melyekben nincs t hosszúságú széria. Kezdődhet i db (t-nél kevesebb, vagyis $1 \le i \le t-1$) azonos elemmel (feketével), utána egy másmilyen (piros), majd ezután sincs t széria:

$$\underbrace{\text{F}\ \ldots\ \text{F}}_{i \text{ db fekete}} \quad \text{P} \quad \underbrace{\ldots \quad \text{F} \quad \ldots \quad \text{P} \quad \ldots}_{\substack{k-i \text{ db fekete, } m-1 \text{ db piros,} \\ \text{és közöttük nincs } t \text{ széria}}}$$

Ezen sorozatok száma: $\sum_{i=1}^{t-1} C_t(m-1,k-i)$.

De ebben lehetnek olyanok is, ahol az 1 db piros után is pirosak vannak úgy, hogy t db piros van egymás után, és utána nincs t széria:

$$\underbrace{\text{F}\ \ldots\ \text{F}}_{i \text{ db fekete}} \ \underbrace{\text{P}\ \ldots\ \text{P}}_{t \text{ db piros}} \underbrace{\ \ldots \quad \text{F} \quad \ldots \quad \text{P} \quad \ldots}_{\substack{k-i \text{ db fekete, } m-t \text{ db piros és} \\ \text{nincs közöttük } t \text{ széria}}}$$

Ezek száma: $\sum_{i=1}^{t-1} C_t(m-t,k-i)$, amit az előző összegből le kell vonni. De ezekben lehet olyan is, hogy a t hosszú piros után is piros következik, vagyis „eltoltan" is lehet t széria. Ezek száma $\sum_{i=1}^{t-1} C_t^*(m-t,k-i)$, ahol $C_t^*(m-t,k-i)$ a pirossal kezdődő, $m-t$ pirosat és $k-i$ feketét tartalmazó olyan sorozatok száma, melyekben nincs t széria.

Nem számítottuk még be azokat az eseteket, amikor $i = 0$, vagyis pirossal kezdődik a sorozat. Ekkor az első elem piros és utána nincs t széria:

$$\text{P} \quad \underbrace{\ldots\ \text{P}\ \ldots\ \text{F}\ \ldots}_{\substack{k \text{ db fekete, } m-1 \text{ piros} \\ \text{és nincs közöttük } t \text{ széria}}}$$

Ezek száma: $C_t(m-1, k)$. De ebben lehetnek olyanok is, melyek t szériával kezdődnek és utána nincs t széria:

$$\underbrace{\text{P}\ \ldots\ \text{P}}_{t \text{ db piros,}}\underbrace{\text{F}\ \ldots\ \text{F}}_{i \text{ db fekete}} \qquad \underbrace{\text{P}\ \ldots\ \ \text{F}\ldots\ \text{P}\ \ldots}_{\substack{m-t \text{ piros, } k-i \text{ db fekete, pirossal kezdődik és} \\ \text{nincs közöttük } t \text{ széria } (1 \le i \le t-1)}}$$

Ezek száma: $\sum_{i=1}^{t-1} C_t^*(m-t, k-i)$, amit előző mennyiségből ($C_t(m-1,k)$-ból) le kell vonni.
Összesítve tehát kapjuk a következőt:

$$C_t(m,k) = \sum_{i=1}^{t-1} C_t(m-1, k-i) - \left\{ \sum_{i=1}^{t-1} C_t(m-t, k-i) - \sum_{i=1}^{t-1} C_t^*(m-t, k-i) \right\} + \\ + \left\{ C_t(m-1,k) - \sum_{i=1}^{t-1} C_t^*(m-t, k-i) \right\} + e_t(m,k).$$

Ami az eredeti (9) képletünket adja. Példaként vegyük:

$$C_3(5,2) = C_3(4,2) + C_3(4,1) + C_3(4,0) - [C_3(2,1) + C_3(2,0)] + 0 = \\ = 6 + 1 + 0 - [3+1] + 0 = 3.$$

Így tehát azon $m+k$ elemű sorozatok számát, melyben m db piros és k db fekete elemet rakunk visszatevés nélkül sorba úgy, hogy nincs t széria, valóban a (9)-es képlet szolgáltatja. □

Az alapkérdésünkre tehát a választ a $P_t(m,k) = 1 - \frac{C_t(m,k)}{\binom{m+k}{k}}$ képletbe való behelyettesítéssel kapjuk.

A $C_t(m,k)$ értékére Bloom [3] 371. oldalán egy olyan formulát is ad, mely az m, k és t értékétől függetlenül mindig 6 tagból áll.

3.2. Állítás. *(Bloom [3], 371. o.) $t \geq 2$ esetén*

$$C_t(m,k) = C_t(m-1,k) + C_t(m,k-1) - C_t(m-t,k-1) - \\ - C_t(m-1,k-t) + C_t(m-t,k-t) + e_t^*(m,k), \tag{10}$$

ahol

$$e_t^*(m,k) = \begin{cases} 1, & \text{ha } m = k = 0, \text{ vagy } m = k = t, \\ -1, & \text{ha } m = 0 \text{ és } k = t \text{ vagy } m = t \text{ és } k = 0, \\ 0, & \text{különben.} \end{cases}$$

Peremfeltételeink pedig ugyanazok, mint a 3.1. állításnál, vagyis, ha $m = k = 0$, *akkor definíció szerint legyen* $C_t(0,0) = 1$, *ha* m *vagy* k *negatív, akkor pedig definíció szerint* $C_t(m,k) = 0$.

A (10) képlet helyességét Bloom [3] 371. oldalon közölttől eltérő módon az alábbiakból látjuk be.

Bizonyítás. Kezdődhet a sorozatunk 1 piros vagy 1 fekete elemmel:

$$\left.\text{P}\ \underbrace{\ldots\ \text{P}\ldots\text{F}\ldots}_{m-1\text{ piros és }k\text{ fekete}}\right\}\text{ezek száma: } C_t(m-1,k);$$

$$\left.\text{F}\ \underbrace{\ldots\ \text{P}\ldots\text{F}\ldots}_{m\text{ piros és }k-1\text{ fekete}}\right\}\text{ezek száma: } C_t(m,k-1).$$

Ezekből le kell vonni azokat, melyekben az első jel után is ugyanolyan következik összesen t hosszon, utána 1 másmilyen és utána nincs t széria:

$$\left.\underbrace{\text{P}\ldots\text{P}}_{t\text{ piros}}\quad\text{F}\quad\underbrace{\ldots\ \text{P}\ldots\text{F}\ldots}_{m-t\text{ piros és }k-1\text{ fekete}}\right\}\text{ezek száma: } C_t(m-t,k-1);$$

$$\left.\underbrace{\text{F}\ldots\text{F}}_{t\text{ fekete}}\quad\text{P}\quad\underbrace{\ldots\ \text{P}\ldots\text{F}\ldots}_{m-1\text{ piros és }k-t\text{ fekete}}\right\}\text{ezek száma: } C_t(m-1,k-t).$$

De ezekben benne szerepelnek a következő sorozatok is.
A t piros, t fekete, utána pirossal kezdődően olyan $m-t$ piros és $k-t$ fekete elemű sorozat, melyben nincs t széria. Ezek száma $C_t^{(p)}(m-t,k-t)$. Illetve – fordítva – t fekete, t piros, utána feketével kezdődő olyan sorozat, melyben nincs t széria. Ezek száma $C_t^{(f)}(m-t,k-t)$. De összegük pedig éppen

$$C_t^{(p)}(m-t,k-t) + C_t^{(f)}(m-t,k-t) = C_t(m-t,k-t).$$

Összegezve tehát:

$$C_t(m,k) = C_t(m-1,k) + C_t(m,k-1) -$$
$$-\{C_t(m-t,k-1) + C_t(m-1,k-t) - C_t(m-t,k-t)\} + e_t^*(m,k).$$

Vagyis valóban a (10) képletet kapjuk. □

A szakasz elején említett kártyás példát tekintve azt találjuk, hogy míg $P_6(26,26) = 0,464$, (annak a valószínűsége, hogy az 52 lapos kártyacsomag lapjait egymás után kirakva lesz benne egymás után 6 egyforma színű csak 0,464), addig a $P_4(26, 26) = 0,974$. Vagyis sokkal valószínűbb a 4-es széria, így Gardner tévedett [6]-ban, amikor azt állította, hogy majdnem mindig kapunk 6-os szériát.

4. Néhány alkalmazás

Az eredmények felhasználhatók a különböző gazdasági, sorbanállási problémák, Markov-láncok (lásd pl. Samarova [14], Schuster [16] vagy Schwager [17] cikkét), tőzsdei eseménysorok vizsgálatában (lásd pl. Binswanger és Embrechts [2] cikkének 4.2. fejezetét). Az alkalmazásnak jelentős szerepe van a műszaki, pl. hidrológiai folyamatokban (lásd pl. Sen [19] cikkét) és a biológiai, pl. molekuláris biológia, DNS sorozatok vizsgálatában (lásd pl. Schilling [15] cikkét). Sok egyéb területen, például grafológiában is felhasználhatóak az eredmények (lásd pl. Arazi [1] cikkét). Révész [13] a számítástechnikában felmerülő, véletlen számsorozat számítógép általi létrehozását említi hasznos alkalmazási területként. Már több módszert is kidolgoztak, melyek véletlen fej-írás sorozatokat generálnak valamilyen módon. Nehéz azonban eldönteni, hogy a kapott véletlen sorozat valóban olyan-e, mintha igazi dobássorozatot írtunk volna. A legtöbbször attól kell tartani, hogy valamilyen – számunkra esetleg ismeretlen – periódus van a felírt sorozatban. A legtöbb statisztikai módszer nem alkalmas ennek a hibának a kiszűrésére. Egy, a leghosszabb fejszéria hosszán alapuló próba alkalmas arra, hogy a sorozatnak a periodicitás okozta nem véletlen voltát kimutassa.

Hivatkozások

[1] Arazi, B.: *Handwriting identification by means of run-length measurements.* IEEE Transactions on Systems, Man and Cybernetics, **7** (12), 1977, 878–881.

[2] Binswanger, K. – Embrechts, P.: *Longest runs in coin tossing.* Insurance Math. Econom. **15**, No. **2-3**, 1994, 139–149.

[3] Bloom, D. M.: *Probabilities of Clumps in a Binary Sequence.* Mathematics Magazine, **69**, No. **5**, 1996.

[4] Erdős Pál - Révész Pál: *On the length of the longest head-run.* Topics in information theory (Second Colloq., Keszthely, 1975.) Colloq. Math. Soc. János Bolyai, Vol. **16**, North-Holland, Amsterdam, 1977, 219–228.

[5] Földes Antónia: *The limit distribution of the length of the longest head-run.* Period. Math. Hungar. **10**, No. **4**, 1979, 301–310.

[6] Gardner, M.: *aha! Gotcha*, Freeman, New York, 1982.

[7] Gordon, L. – Schilling, M. F. – Waterman, M. S.: *An extreme value theory for long head runs.* Probab. Theory Relat. Fields, **72**, No. **2**, 1986, 279–287.

[8] Kopocinski, B.: *On the distribution of the longest succes-run in Bernoulli trials.* Roczniki Polskiego Towarzystwa Matematycznego, Seria III, Matematyka Stosowana XXXIV, 1991.

[9] Muselli, M.: *Useful inequalities for the longest run distribution.* Statist. Probab. Lett. **46**, No. **3**, 2000, 239–249.

[10] Philippou, A. N. – Makri, F. S.: *Longest success runs and Fibonacci-type polynomials*, The Fibonacci Quarterly **23**, Nov. 1985.

[11] Rényi Alfréd: *Probability Theory*, Akad. Kiadó, Budapest, 1970.

[12] Révész Pál: *Strong theorems on coin tossing.* Proceedings of the International Congress of Mathematicians, Helsinki, 1978.

[13] Révész Pál: *Mennyire véletlen a véletlen?* Akadémiai székfoglaló, Akadémiai Kiadó, Budapest, 1982.

[14] Samarova, S. S.: *On the asymptotic behaviour of the maximal sojourn time of an ergodic Markov chain in a fixed state.* Russian Math Surveys **35** (6), 1980, 103–104.

[15] Schilling, M. F.: *The Longest Run of Heads.* The College Mathematics Journal, 1990. Vol. **21**, No. **3**

[16] Schuster, E. F.: *On overwhelming numerical evidence in the settling of Kinney's waiting time conjecture.* SIAM Journal of Statistical Computing, **6** (4), 1985, 977–982.

[17] Schwager, S. J.: *Run probabilities in sequences of Markov-dependent trials.* Journal of the American Statistical Association, **78**, 1983, 168–175.

[18] Szászné Simon Judit: *A sztochasztika középiskolai oktatása.* PhD értekezés, Debreceni Egyetem, 2005.

[19] Sen, Z.: *Statistical analysis of hydrologic critical droughts.* Journal of the Hydraulics Division **106** (HY1), 1980, 99–115.

(Beérkezett: 2009. december 7.)

FAZEKAS ISTVÁN
Debreceni Egyetem
Alkalmazott Matematika és Valószínűségszámítás Tanszék
H-4010 Debrecen, Pf. 12
e-mail: fazekasi@inf.unideb.hu

KARÁCSONY ZSOLT
Miskolci Egyetem
Alkalmazott Matematikai Tanszék
H-3515 Miskolc-Egyetemváros
e-mail: matkzs@uni-miskolc.hu

LIBOR JÓZSEFNÉ
Szolnoki Főiskola
Gazdaságelemzési Módszertani Tanszék
H-5000 Szolnok, Tiszaligeti sétány
e-mail: liborne@szolf.hu

ON LONGEST RUNS

István Fazekas, Zsolt Karácsony and Józsefné Libor

The coin tossing experiment is considered. The length of the longest head run can be studied by asymptotic theorems (Erdős-Révész [4], Földes [5]), by recursive formulae (Schilling [15], Kopocinski [8], Bloom [3]) or by computer simulations (Binswanger-Embrechts [2]). The aim of the paper is to compare numerically the asymptotic results, the recursive formulae, and the simulation results. Moreover, we consider also the longest run (i.e. the longest pure heads or pure tails). We compare the distribution of the longest head run and that of the longest run. We consider both fair and biased coins. We also study the draw of cards without replacement. We give detailed proofs for the recursive formulae. We also touch upon a little history and applications of this topic.