

# FEJEZETEK A HOMOGEN FEJSOROZATOKRÓL

SZAKDOLGOZAT

**Készítette: Kovács Balázs**

Matematika BSc, tanári szakirány



**Témavezető: dr. Wintsche Gergely, adjunktus**

ELTE TTK, Matematikatanítási és Módszertani Központ

Eötvös Loránd Tudományegyetem

Természettudományi Kar

Budapest, 2011

# Tartalomjegyzék

Bevezetés .....	3
1. A leghosszabb homogén sorozat hosszának eloszlásfüggvénye .....	6
1.1. $R_n$ eloszlásfüggvénye – a szabályos pénzérme esete .....	7
1.2. $R_n$ eloszlásfüggvénye – a nem szabályos pénzérme esete.....	10
2. A leghosszabb homogén fejsorozat hosszának várható értéke .....	13
2.1. A várható érték közelítése.....	13
2.2. Illusztrációk.....	15
3. Dobáljunk $n$ darab érmét, míg az összes érme fejet nem mutat .....	18
3.1. $Y$ eloszlásfüggvénye .....	18
3.2. A $p_m = P(Y = m)$ függvény.....	20
3.3. $Y$ várható értéke.....	21
4. Kitekintés: Erdős Pál és Rényi Alfréd eredményei .....	23
4.1. Erdős és Rényi vonatkozó tétele .....	23
4.2. A tétel bizonyításához szükséges fogalmak és állítások.....	24
Irodalomjegyzék.....	26

## Bevezetés

[Rosencrantz pénzérmét pöccint a magasba, az aláhulló érmét elkapja, majd a kézfejére helyezi. Guildenstein közelről figyeli.]

ROSENCRANTZ: „*Fej. Fej. Fej... Fej.*” [Tizennyolcszor dobja fel a pénzérmét, és valamennyi dobás eredménye fej.]

GUILDENSTEIN [A tizenkilencedszer feldobott érmét röptében elkapja. Meggyőződik arról, hogy a pénzérme szabályos, majd visszahajítja Rosencrantzknak.]

ROSENCRANTZ [Elkapja a visszadobott érmét, majd a kézfejére helyezi.] „*Fej.*”

(...)

[Rosencrantz száznál is több fejet dob közvetlenül egymás után. Guildenstein megpróbálja megmagyarázni a jelenséget.]

GUILDENSTEIN: „*(...) azon elv lenyűgöző érvényre jutása, hogy minden egyes érme minden feldobásánál mind a fejre, mind az írásra egyenlő az esély. Ekképp nem okozhat meglepetést, ha minden esetben így tesz.*”

(Részlet a *Rosencrantz és Guildenstein halott* c. fimből)

- o -

A *Valószínűségszámítás gyakorlat* című kurzus első órája szokatlanul, ám annál érdekesebben kezdődött. Wintsche Gergely tanár úr – a kurzus oktatója – a következő kéréssel fordult a hallgatósághoz: *Jegyezzetek le két papírlapra egy-egy 100 hosszú, „F” és „Í” jelekből álló sorozatot. Az első sorozatot „találomra” készítsétek el annak alapján, amit egy pénzfeldobás-sorozatból gondoltok; míg a második egy valós pénzfeldobás-sorozat alapján készüljön. Miután ezt megtettétek, bármelyikőtök ideadhatja az általa lejegyzett két sorozatot, én pedig – jó eséllyel – meg fogom tudni mondani, hogy közülük melyik reprezentálja a valós pénzfeldobások kimeneteleit.*

Megdöbbenésemre Wintsche tanár úr minden egyes hallgató esetén másodpercek alatt meg tudta különböztetni a valós pénzfeldobás-sorozatnak megfelelő eredménykészletet

a „találomra” elkészítettől; és azt is megosztotta a hallgatósággal, hogy mi a megkülönböztetés kulcsa: Elmondása szerint a leghosszabb *homogén sorozat* (azaz az egymást követő fejek vagy írások sorozatai közül a leghosszabb sorozat) hossza lényegesen rövidebb azon eredménykészletekben, melyeket saját elképzelésünk alapján állítunk elő.

Az érdekességek sora nem ért véget az előbbi kísérlettel: Wintsche tanár úr ezután a fent idézett filmrészletet vetítette le a hallgatóság számára. Az epizódot magától értetődő szkepticizmussal szemléltem; de kíváncsiságom arra sarkallt, hogy ki is számoljam annak valószínűségét, hogy egymást követő száz pénzfeldobás kimenetele mindannyiszor fej legyen. A számológépem másodpercek alatt megadta a választ, a kérdéses valószínűség:  $\frac{1}{2^{100}} \approx 7,89 \cdot 10^{-31}$ .

A valószínűségszámítás tárgyköre iránti rokonszenveimből adódóan a kurzus elvégzése után is megannyi, a valószínűségelmélet különböző diszciplínáit tárgyaló szakirodalmat olvastam; és később ez vezetett arra az elhatározásra, hogy hajdani gyakorlatvezetőmet, Wintsche Gergely tanár urat kérjem fel szakdolgozati témavezetőmnek. Ezúton szeretném megköszönni Wintsche tanár úrnak a briliáns témajavaslatot, munkám szakszerű és lelkiismeretes irányítását, és a felmerülő kérdéseim megválaszolásában nyújtott alapos segítséget.

Szakdolgozatom célja a pénzfeldobás-sorozatokban előforduló homogén sorozatok egyes figyelemre méltó tulajdonságainak elemzése volt, melyhez néhány matematikus ezen problémakörrel kapcsolatban kifejtett értekezését vettem alapul. Fő törekvésem volt továbbá, hogy a forrásművek szemléletesen tárgyalt jelenségeihez, állításaihoz önállóan konstruáljak formalizált matematikai leírásokat, bizonyításokat.

A dolgozat első fejezetének tárgya az  $n$  hosszú pénzfeldobás-sorozatban előforduló leghosszabb homogén fejsorozat hosszának eloszlásfüggvénye mind szabályos, mind nem szabályos pénzérme esetén. Az eloszlásfüggvény tulajdonságainak elemzése mellett kitérek arra is, hogy a tárgyalt attribútumok hogyan hozhatók kapcsolatba a leghosszabb – fejekből vagy írásokból álló – homogén sorozat általános esetével.

A második fejezetben közelítő eredményt fogalmazok meg az  $n$  hosszú pénzfeldobás-sorozatban előforduló leghosszabb homogén fejsorozat hosszának várható értékéről,

Microsoft Office Excel™-ben elkészített illusztrációkkal is szemléltetve a közelítés helyességét.

Az elkövetkező fejezetben egy összetett jelenség elemzését végzem el az előbb bemutatott fejezetek szempontjai szerint, végül az utolsó fejezetben rövid betekintést nyújtok Erdős Pál és Rényi Alfréd homogén sorozatokkal kapcsolatban megfogalmazott eredményeibe.

# 1. A leghosszabb homogén sorozat hosszának eloszlásfüggvénye

Tekintsünk egy  $n$  darab pénzfeldobásból álló dobássorozatot; és tegyük fel, hogy a pénzérme szabályos, azaz annak valószínűsége, hogy egy pénzfeldobás kimenetele fej (avagy írás):  $p = \frac{1}{2}$ . Jelölje  $R_n$  a dobássorozatban előforduló leghosszabb homogén fejsorozat hosszát.  $n = 3$  esetén könnyen felírható az összes lehetséges dobássorozat, és rövid számolás útján megkaphatók  $R_3$  lehetséges értékeihez tartozó valószínűségek:

*FFF, FFÍ, FÍF, FÍÍ,*  
*ÍFF, ÍFÍ, ÍÍF, ÍÍÍ.*

$R_3$ értékei ( $x$ )	konkrét esetek	$P(R_3 = x)$
<b>0</b>	<i>ÍÍÍ</i>	1/8
<b>1</b>	<i>FÍF, FÍÍ, ÍFÍ, ÍÍF</i>	4/8
<b>2</b>	<i>FFÍ, ÍFF</i>	2/8
<b>3</b>	<i>FFF</i>	1/8

Ily módon pontosan kiszámíthatjuk  $R_3$  várható értékét:

$$E(R_3) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{4}{8} + 2 \cdot \frac{2}{8} + 3 \cdot \frac{1}{8} = \frac{11}{8}.$$

Kisebb  $n$ -ek – például  $n = 5$  – esetén  $R_n$  várható értéke könnyen megkapható ugyanezen módszerrel. Megjegyzendő azonban; hogy minél nagyobb  $n$  értéke, annál több időt vesz igénybe az összes lehetséges dobássorozat lejegyzése;  $2^{200}$  darab dobássorozat felírása pedig még a megfelelő számítógépprogrammal is meglehetősen nehéz. A probléma tovább bonyolódik, ha a pénzérme nem szabályos – azaz, ha annak valószínűsége, hogy

egy pénzfeldobás kimenetele fej:  $p$ , míg annak, hogy írás:  $q := 1 - p$ , ahol  $p \neq \frac{1}{2}$  –; hiszen ekkor az egyes dobássorozatok eltérő valószínűséggel fordulnak elő. Láthatjuk, hogy egy *általános* módszerre van szükség, ennek megkonstruálásához pedig elsőként  $R_n$  eloszlását kell részletesen megvizsgálnunk. Legyen  $F_n(x)$  az  $R_n$  valószínűségi változó eloszlásfüggvénye, azaz:

$$F_n(x) := P(R_n \leq x).$$

### 1.1. $R_n$ eloszlásfüggvénye – a szabályos pénzérme esete

Tekintsünk egy  $n$  darab pénzfeldobásból álló dobássorozatot; és tegyük fel, hogy a pénzérme szabályos. Jelölje  $A_n(x)$  azon  $n$  hosszú dobássorozatok számát, melyekben a leghosszabb homogén fejsorozat hossza legfeljebb  $x$ . Megállapítható, hogy  $F_n(x) = \frac{A_n(x)}{2^n}$ ; hiszen a lehetséges  $n$  hosszúságú dobássorozatok száma  $2^n$ .  $F_n(x)$  esetén meghatározásához az  $A_n(x)$  sorozat értékeire van szükségünk, ezért az elkövetkezőkben erre a sorozatra konstruálunk rekurzív formulát.

$$\mathbf{1.1.1. \text{ Állítás: }} A_n(x) = \begin{cases} \sum_{j=0}^x A_{n-j-1}(x), & \text{ha } n > x; \\ 2^n, & \text{ha } n \leq x. \end{cases}$$

**Bizonyítás:** Először belátjuk, hogy  $x = 3$  esetén igaz az állítás; majd a következtetéseink alapján megfogalmazzuk az általánosítást.

$x = 3$  esetén a következők teljesülnek:

- Ha  $n \leq 3$ , akkor:  $A_n(3) = 2^n$ ; hiszen triviális, hogy bármely legfeljebb 3 hosszú sorozatra teljesül, hogy a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb 3.

- Ha  $n > 3$ ; akkor minden olyan sorozat, mely legfeljebb 3 hosszú homogén fejsorozatot tartalmaz,  $I$ ,  $FI$ ,  $FFI$ , vagy  $FFFI$  kezdetű; és olyan „láncban” folytatódik, melyben a leghosszabb homogén fejsorozat hossza legfeljebb 3. Ily módon a következő rekurzió adódik:

$$A_n(3) = A_{n-1}(3) + A_{n-2}(3) + A_{n-3}(3) + A_{n-4}(3) = \sum_{j=0}^3 A_{n-j-1}(3), \text{ ha } n > 3.$$

Az előbbieken bizonyítást nyert, hogy  $x = 3$  esetén igaz az állítás. Hasonló gondolatmenetet követve könnyen igazolható, hogy az állítás bármely  $x$ -re teljesül:

- Ha  $n \leq x$ , akkor  $A_n(x) = 2^n$ ; hiszen triviális, hogy bármely legfeljebb  $x$  hosszú sorozatra teljesül, hogy a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb  $x$ .
- Ha  $n > x$ ; akkor minden olyan sorozat, mely legfeljebb  $x$  hosszú homogén fejsorozatot tartalmaz,  $I$ ,  $FI$ ,  $FFI$ , ... vagy  $FFF \dots FI$  kezdetű (utóbbi kezdőelem-sorozatban  $x$  darab  $F$  követi egymást), és olyan „láncban” folytatódik, melyben a leghosszabb homogén fejsorozat hossza legfeljebb  $x$ .

Ily módon a következő rekurzió adódik:

$$A_n(x) = A_{n-1}(x) + A_{n-2}(x) + A_{n-3}(x) + \dots + A_{n-(x+1)}(x) = \sum_{j=0}^x A_{n-j-1}(x), \text{ ha } n > x.$$

□

Észrevehetjük, hogy  $A_n(1)$  a *Fibonacci-sorozat* eltoltja:

$$A_0(1) = 2^0 = 1;$$

$$A_1(1) = 2^1 = 2;$$

$$A_2(1) = \sum_{j=0}^1 A_{2-j-1}(1) = \sum_{j=0}^1 A_{2-(j+1)}(1) = A_{2-(0+1)} + A_{2-(1+1)} = A_1 + A_0 = 2 + 1 = 3; \dots$$

$$A_n(1) = \sum_{j=0}^1 A_{n-j-1}(1) = \sum_{j=0}^1 A_{n-(j+1)}(1) = A_{n-(0+1)} + A_{n-(1+1)} = A_{n-1} + A_{n-2}.$$



Joggal megülhet fel a kérdés: Mennyiben változik az eloszlás abban az esetben; ha a leghosszabb homogén – fejből vagy írásokból álló – sorozat hosszát vizsgáljuk?

Jelölje  $R_n^*$  az  $n$  hosszú dobássorozatban előforduló leghosszabb homogén sorozat hosszát, és  $B_n(x)$  azon  $n$  hosszú dobássorozatok számát, melyekben a leghosszabb homogén sorozat hossza legfeljebb  $x$ . Legyen  $F_n^*(x)$  az  $R_n^*$  valószínűségi változó eloszlásfüggvénye, azaz:

$$F_n^*(x) := P(R_n^* \leq x).$$

Triviális, hogy  $F_n^*(x) = \frac{B_n(x)}{2^n}$ ; azaz ebben az esetben a  $B_n(x)$  sorozatra kell rekurziót konstruálnunk.

**1.1.2. Állítás:** 
$$B_n(x) = \begin{cases} 2 \cdot A_{n-1}(x-1), & \text{ha } x \geq 1; \\ 0, & \text{ha } x = 0. \end{cases}$$

**Bizonyítás:** Jegyezzük le az egyes pénzfeldobások kimeneteleit. Így egy  $n$  hosszú,  $F$  és  $Í$  jelekből álló sorozathoz jutunk, melynek  $i$ -edik elemét jelölje  $w_i$ . Jegyezzünk le egy másik sorozatot a következőképp: Két szomszédos jel alá írjunk egy  $K$  jelet, ha azok különbözőek (azaz  $w_i \neq w_{i+1}$ ); illetve  $A$  jelet, ha a két jel azonos (azaz  $w_i = w_{i+1}$ ). Az így kapott –  $K$  és  $A$  jelekből álló – sorozatra a következők igazak:

- $n - 1$  hosszú,
- egy  $x$  hosszú homogén –  $F$  vagy  $Í$  jelekből álló – sorozatnak megfelel egy  $x - 1$  hosszú  $A$ -sorozat – lásd az alábbi példát:

$$\begin{array}{cccccccccc} \acute{I} & F & \acute{I} & \acute{I} & \acute{I} & F & \acute{I} & \acute{I} & F & F & \dots \\ K & K & A & A & K & K & A & K & A & \dots \end{array}$$

Ily módon:

$$F_n^*(x) = P(R_n^* \leq x) = F_{n-1}(x-1) = P(R_{n-1} \leq x-1).$$

Korábbi megállapításaink alapján a következők igazak:

$$P(R_{n-1} \leq x-1) = \frac{A_{n-1}(x-1)}{2^{n-1}}; \text{ és } F_n^*(x) = \frac{B_n(x)}{2^n}.$$

Mindezekből következik, hogy  $x \geq 1$  esetén:

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

$$\Downarrow$$

$$B_n(x) = \frac{2^n}{2^{n-1}} \cdot A_{n-1}(x-1) = 2 \cdot A_{n-1}(x-1).$$

□

## 1.2. $R_n$ eloszlásfüggvénye – a nem szabályos pénzérme esete

Tekintsünk egy  $n$  darab pénzfeldobásból álló dobássorozatot; és tegyük fel, hogy a pénzérme nem szabályos, azaz annak valószínűsége, hogy egy pénzfeldobás kimenetele fej:  $p$ , míg annak, hogy írás:  $q := 1 - p$ , ahol  $p \neq \frac{1}{2}$ . Jelölje  $C_n^{(k)}(x)$  azon  $n$  hosszú dobássorozatok számát, melyekben pontosan  $k$  darab fej fordul elő, és melyekre  $R_n \leq x$ .

**1.2.1. Állítás:**  $F_n(x) = \sum_{k=0}^n C_n^{(k)}(x) \cdot p^k \cdot q^{n-k}$ .

**Bizonyítás:** Esetszétválasztást alkalmazunk aszerint, hogy azon pénzfeldobássorozatokban, melyekre  $R_n \leq x$ , pontosan hány darab fej fordul elő.

Fejek száma ( $k$ )	#(konkrét esetek)	$P(R_n \leq x)$
<b>0</b>	$C_n^{(0)}(x)$	$C_n^{(0)}(x) \cdot p^0 \cdot q^{n-0}$
<b>1</b>	$C_n^{(1)}(x)$	$C_n^{(1)}(x) \cdot p^1 \cdot q^{n-1}$
$\vdots$	$\vdots$	$\vdots$
<b><math>n</math></b>	$C_n^{(n)}(x)$	$C_n^{(n)}(x) \cdot p^n \cdot q^{n-n}$

Nyilvánvaló, hogy:

$$(R_n \leq x) = [(R_n \leq x) \wedge (k = 0)] \vee [(R_n \leq x) \wedge (k = 1)] \vee \dots \vee [(R_n \leq x) \wedge (k = n)],$$

és mivel az egyes események kizáróak:

$$\begin{aligned} P(R_n \leq x) &= P[(R_n \leq x) \wedge (k = 0)] + P[(R_n \leq x) \wedge (k = 1)] + \dots + P[(R_n \leq x) \wedge (k = n)] = \\ &= C_n^{(0)}(x) \cdot p^0 \cdot q^{n-0} + C_n^{(1)}(x) \cdot p^1 \cdot q^{n-1} + \dots + C_n^{(n)}(x) \cdot p^n \cdot q^{n-n} = \sum_{k=0}^n C_n^{(k)}(x) \cdot p^k \cdot q^{n-k}. \end{aligned}$$

□

$F_n(x)$  esetén meghatározásához a  $C_n^{(k)}(x)$  sorozat értékeire van szükségünk, ezért az elkövetkezőkben erre a sorozatra konstruálunk rekurzív formulát.

$$\mathbf{1.2.2. \text{ Állítás: }} C_n^{(k)}(x) = \begin{cases} \binom{n}{k}, & \text{ha } k \leq x; \\ 0, & \text{ha } x < k = n; \\ \sum_{j=0}^x C_{n-(j+1)}^{(k-j)}(x), & \text{ha } x < k < n. \end{cases}$$

**Bizonyítás:** Először belátjuk, hogy  $x = 3$  esetén igaz az állítás; majd a következtetéseink alapján megfogalmazzuk az általánosítást.

$x = 3$  esetén a következők teljesülnek:

- Ha  $k \leq 3$ , akkor  $C_n^{(k)}(3) = \binom{n}{k}$ ; hiszen triviális, hogy bármely legfeljebb 3 fejet tartalmazó sorozatra teljesül, hogy a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb 3. A pontosan  $k$  darab fejet tartalmazó,  $n$  hosszú dobássorozatok száma pedig  $\binom{n}{k}$ .
- Ha  $3 < k = n$ , akkor  $C_n^{(k)}(3) = 0$ .

- Ha  $3 < k < n$ , akkor minden olyan sorozat, mely pontosan  $k$  darab fejet és legfeljebb  $x$  hosszú homogén fejsorozatot tartalmaz,  $I, FI, FFI, \dots$  vagy  $FFFI$  kezdetű, és olyan „láncban” folytatódik, mely – az előző esetek sorrendje szerint – pontosan  $k, k-1, k-2$ , vagy  $k-3$  darab fejet tartalmaz; valamint a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb  $x$ .

Ily módon a következő rekurzió adódik:

$$C_n^{(k)}(3) = C_{n-1}^{(k)}(3) + C_{n-2}^{(k-1)}(3) + C_{n-3}^{(k-2)}(3) + C_{n-4}^{(k-3)}(3) = \sum_{j=0}^3 C_{n-(j+1)}^{(k-j)}(3), \text{ ha } 3 < k < n.$$

Az előbbieken bizonyítást nyert, hogy  $x = 3$  esetén igaz az állítás. Hasonló gondolatmenetet követve könnyen igazolható, hogy az állítás bármely  $x$ -re teljesül:

- Ha  $k \leq x$ , akkor  $C_n^{(k)}(x) = \binom{n}{k}$ ; hiszen triviális, hogy bármely legfeljebb  $x$  fejet tartalmazó sorozatra teljesül, hogy a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb  $x$ . A pontosan  $k$  darab fejet tartalmazó,  $n$  hosszú dobássorozatok száma pedig  $\binom{n}{k}$ .
- Ha  $x < k = n$ , akkor  $C_n^{(k)}(x) = 0$ .
- Ha  $x < k < n$ , akkor minden olyan sorozat, mely pontosan  $k$  darab fejet és legfeljebb  $x$  hosszú homogén fejsorozatot tartalmaz,  $I, FI, FFI, \dots$  vagy  $FFF\dots FI$  kezdetű (utóbbi kezdőelem-sorozatban  $x$  darab  $F$  követi egymást), és olyan „láncban” folytatódik, mely – az előző esetek sorrendje szerint – pontosan  $k, k-1, k-2, \dots$  vagy  $k-x$  darab fejet tartalmaz; valamint a benne előforduló leghosszabb homogén fejsorozat hossza legfeljebb  $x$ .

Ily módon a következő rekurzió adódik:

$$C_n^{(k)}(x) = C_{n-1}^{(k)}(x) + C_{n-2}^{(k-1)}(x) + \dots + C_{n-(x+1)}^{(k-x)}(x) = \sum_{j=0}^x C_{n-(j+1)}^{(k-j)}(x), \text{ ha } x < k < n.$$

□

## 2. A leghosszabb homogén fejsorozat hosszának várható értéke

### 2.1. A várható érték közelítése

Tekintsünk egy  $n$  darab pénzfeldobásból álló dobássorozatot; és tegyük fel, hogy annak valószínűsége, hogy egy pénzfeldobás kimenetele fej:  $p$ , míg annak, hogy írás:  $q := 1 - p$ . A továbbiakban megengedjük a nulla hosszúságú homogén fejsorozatok meglétét is. Szemléltetésképp tekinthetjük az alábbi 7 hosszúságú dobássorozatot, melyben a „ $\wedge$ ” szimbólumok jelölik a nulla hosszú homogén fejsorozatokat:

$$\wedge \quad I \quad \wedge \quad I \quad \wedge \quad I \quad F \quad F \quad I \quad \wedge \quad I.$$

Látható, hogy  $h$  darab írás pontosan  $h$  darab homogén fejsorozatot választ el.

**2.1.1. Állítás:**  $E(R_n) \approx \log_{\frac{1}{p}}(n \cdot q)$ .

**Heurisztikus érvelés:** Jelöljük  $K_0; K_1; K_2 \dots$  a következő eseményeket:

$K_0 :=$  (írást dobok);

$K_1 :=$  (egy fejet és egy írást dobok ilyen sorrendben);

$K_2 :=$  (két fejet és egy írást dobok ilyen sorrendben); ...

$K_j :=$  ( $j$  darab fejet és egy írást dobok ilyen sorrendben).

Jelölje  $k_j$  a  $K_j$  esemény gyakoriságát. Evidens, hogy  $k_0$  *binomiális eloszlású*, így megállapítható, hogy az  $n$  hosszú pénzfeldobás-sorozatban várhatóan  $n \cdot q$  darab írás fordul elő. Ebből – korábbi megállapításunk szerint – adódik, hogy a dobássorozat várhatóan  $n \cdot q$  darab homogén fejsorozatot fog tartalmazni.

Szintén könnyű látni, hogy várhatóan a homogén fejsorozatok  $p$ -edrészében (azaz  $n \cdot q \cdot p$  darabban) fog legalább egy fej előfordulni; valamint, hogy várhatóan ezen sorozatok szintén  $p$ -edrésze (azaz  $n \cdot q \cdot p^2$  darab) tartalmaz legalább két fejet stb. Ennélfogva:

$$E(k_0) = n \cdot q;$$

$$E(k_1) = n \cdot q \cdot p;$$

$$E(k_2) = n \cdot q \cdot p^2; \dots$$

$$E(k_j) = n \cdot q \cdot p^j.$$

Ily módon megállapíthatjuk, hogy az  $n$  hosszú pénzfeldobás-sorozatban várhatóan  $n \cdot q$  darab homogén fejsorozat fordul elő; ezek közül pedig várhatóan  $n \cdot q \cdot p^j$  darab lesz legalább  $j$  hosszú. Az eddigiek alapján szemléletesen látható; hogy  $n \cdot q \cdot p^j \geq 1$  esetén a dobássorozatban várhatóan előfordul legalább egy  $j$  hosszú – vagy akár hosszabb – homogén fejsorozat; illetve, hogy  $n \cdot q \cdot p^j < 1$  esetén a  $j$  hosszú fejsorozatok száma várhatóan elenyésző. Ennélfogva az  $n \cdot q \cdot p^{E(R_n)} \approx 1$  reláció  $E(R_n)$ -re történő megoldásával jó becslést kaphatunk  $R_n$  várható értékére:

$$\begin{aligned} n \cdot q \cdot p^{E(R_n)} &\approx 1 \\ \Downarrow \\ (p^{-1})^{E(R_n)} &\approx n \cdot q \\ \Downarrow \\ E(R_n) &\approx \log_{\frac{1}{p}}(n \cdot q). \end{aligned}$$

□

A fenti közelítésből a szabályos pénzérme  $\left(p = \frac{1}{2}\right)$  esetére a következőt nyerjük:

$$E(R_n) \approx \log_{1:(1/2)}\left(n \cdot \frac{1}{2}\right) = \log_2\left(\frac{n}{2}\right) = \log_2(n) - \log_2(2) = \log_2(n) - 1.$$

## 2.2. Illusztrációk

A leghosszabb homogén fejsorozat hosszának várható értékére kapott közelítő eredmény alapján az alábbi következtetést fogalmazhatjuk meg:

Hajtsunk végre egy másik kísérletet, melynek során az érmét nem  $n$ -szer, hanem  $2n$ -szer dobjuk fel. Ekkor a megfelelő logaritmusazonosság alkalmazásával a következőt kaphatjuk:

$$E(R_{2n}) \approx \log_2(2n) - 1 = \log_2(2) + \log_2(n) - 1 = 1 + (\log_2(n) - 1) = 1 + E(R_n);$$

azaz ha a dobássorozat hosszát a kétszeresére növeljük, akkor (közelítőleg) eggyel nő a leghosszabb homogén fejsorozat várható hossza. A jelenség ábrázolására kiváló illusztráció adható a Microsoft Office Excel™ segítségével: A megfelelő függvényeket használva több száz hosszúságú fej-írás sorozatokat generálhatunk, megkereshetjük a bennük előforduló leghosszabb homogén fejsorozatok hosszait; és megvizsgálhatjuk, hogy mely hossz milyen gyakran fordul elő.

Az 1. ábrán látható grafikon a következő – Excel™-ben modellezett – kísérletsorozatban megfigyelt jelenségeket szemlélteti: Egy kísérlet során egy szabályos pénzérmét 200-szor feldobtunk, a dobások eredményeit lejegyeztük, majd ezt újabb 199 alkalommal megismételtük; így 200 darab 200 hosszú fej-írás sorozathoz jutottunk. Megvizsgáltuk, hogy az egyes dobássorozatokban mekkora volt a leghosszabb homogén fejsorozat hossza az ötvenedik, századik, illetve kétszázadik pénzfeldobásig; ezt követően pedig megszámláltuk, hogy ezen hosszak közül hány egyezett meg az 1, 2, ..., 20 értékekkel (20-nál nagyobb érték a kísérletek során nem fordult elő).

### **Az első 50 pénzfeldobás vizsgálata**

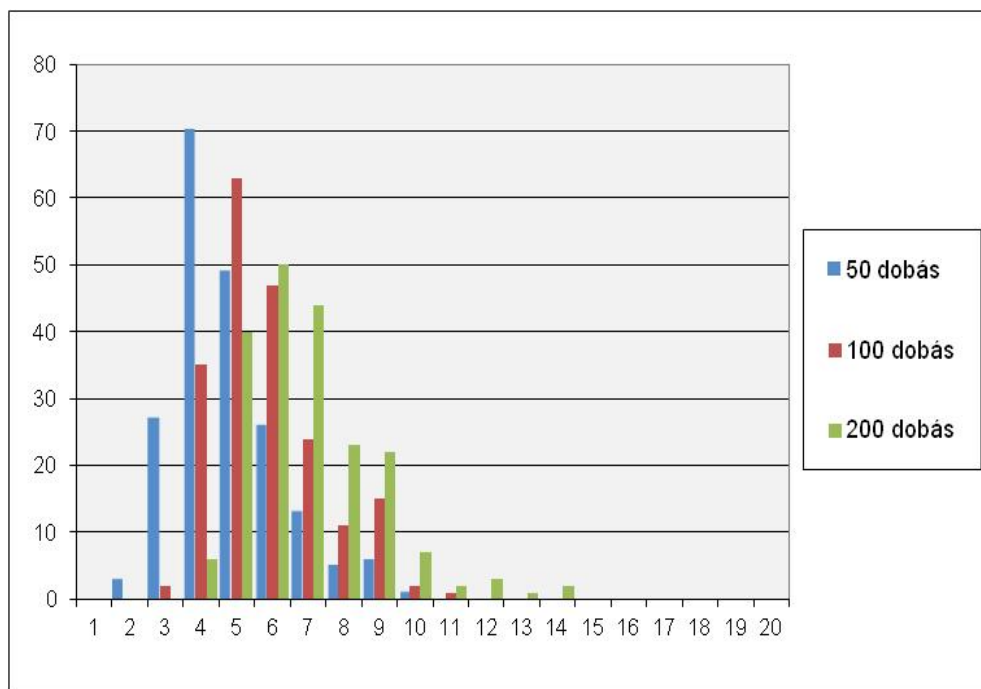
- A legtöbb esetben 4 volt a leghosszabb homogén fejsorozat hossza az ötvenedik pénzfeldobásig.
- A tapasztalati átlagra vonatkozóan a program néhány futás (cella-újraszámoltatás) után a következő eredményeket szolgáltatta: 4,9; 5; 5,1; 5,3.

### Az első 100 pénzfeldobás vizsgálata

- A legtöbb esetben 5 volt a leghosszabb homogén fejsorozat hossza a századik pénzfeldobásig (vegyük észre, hogy ez az érték épp eggyel nagyobb az előzőnél).
- A tapasztalati átlagra vonatkozóan a program néhány futás után a következő eredményeket szolgáltatta: 5,8; 5,9; 6; 6,1.

### Az első 200 pénzfeldobás vizsgálata

- A legtöbb esetben 6 volt a leghosszabb homogén fejsorozat hossza a kétszázadik pénzfeldobásig (vegyük észre, hogy ez az érték épp eggyel nagyobb az előzőnél).
- A tapasztalati átlagra vonatkozóan a program néhány futás után a következő eredményeket szolgáltatta: 6,7; 6,9; 7; 7,1.



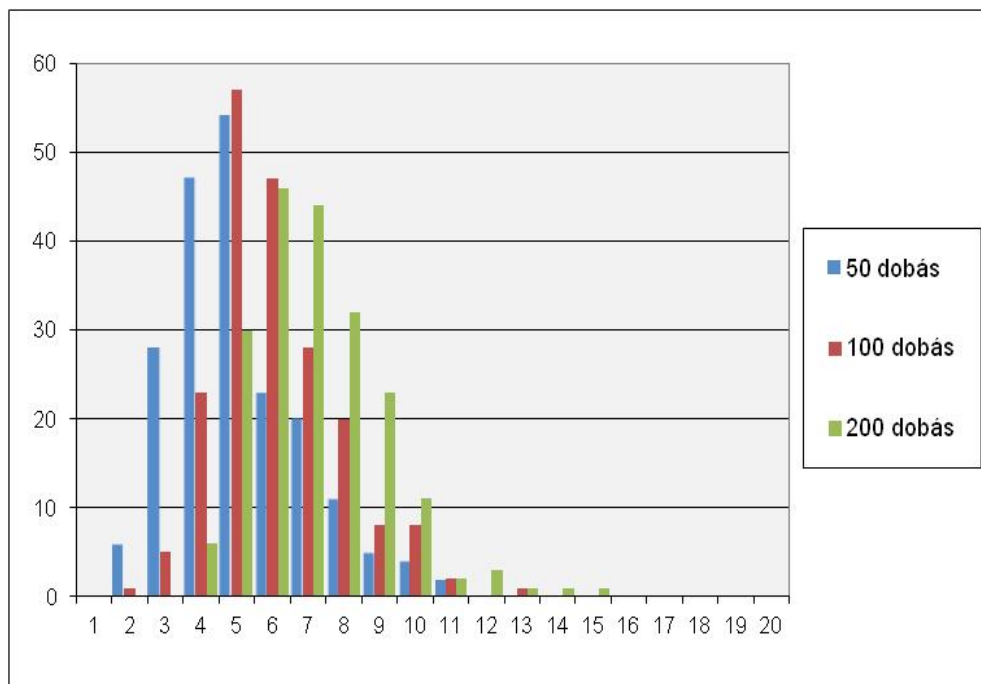
1. Ábra



Megállapíthatjuk, hogy a fenti tapasztalatok híven illusztrálják a fentiekben felvázolt jelenséget, és az előző alfejezet végén kapott közelítő eredménnyel is összeegyeztethetők:

- $E(R_{50}) = \log_2(50) - 1 \approx 4,64$ ;
- $E(R_{100}) = \log_2(100) - 1 \approx 5,64$ ;
- $E(R_{200}) = \log_2(200) - 1 \approx 6,64$ .

Nem meglepő, hogy a program kellően sokszor történő futtatása után – elenyésző számban ugyan, de – olyan eredmény is nyerhető, mely nem tükrözi a fenti jelenséget. A 2. ábrán látható grafikon például egy olyan kísérletsorozatból származik, melyben az első 50 és az első 100 pénzfeldobás esetén is 5 volt a leghosszabb homogén fejsorozatok hosszaiból álló sorozat módusza.



2. Ábra

### 3. Dobáljunk $n$ darab érmét, míg az összes érme fejet nem mutat

Tekintsük a következő kísérletet:

1. Dobjunk fel **egyszerre**  $n$  darab – nem feltétlenül szabályos – pénzérmét; majd azokat, amelyek fejet mutatnak, tegyük félre.

2. Dobjuk fel az így megmaradó pénzérméket újra (egyszerre); majd azokat, amelyek fejet mutatnak, tegyük félre.

3. Folytassuk az eljárást mindaddig, amíg minden érme a félretettek közé nem kerül; azaz addig, amíg az összes érme fejet nem mutat.

Jelen fejezetben a következő kérdésre keressük a választ: *Hány ilyen – egyszerre történő – feldobásra van átlagosan szükségünk a kísérlet befejezéséhez?*

Jelölje az  $Y$  valószínűségi változó a kísérlet befejezéséhez szükséges dobások számát.

Ismeretes, hogy  $E(Y) = \sum_{m=1}^{\infty} m \cdot P(Y = m)$ ; így célunk a  $p_m = P(Y = m)$  függvény

meghatározása. Ehhez szükségünk lesz az  $Y$  valószínűségi változó  $G(y) := P(Y \leq y)$  eloszlásfüggvényére.

#### 3.1. $Y$ eloszlásfüggvénye

Tekintsünk egy  $n$  darab számozott (1, 2, ...,  $n$  sorszámokkal ellátott) pénzérmét használó kísérletet; és tegyük fel, hogy annak valószínűsége, hogy egy pénzfeldobás kimenetele fej:  $p$ , míg annak, hogy írás:  $q := 1 - p$ . Jelölje az  $X_i$  valószínűségi változó azon dobásszámot, mely ahhoz szükséges, hogy az  $i$  sorszámú pénzérmével először fejet dobjunk. A *Pascal-eloszlás* ismert tulajdonságai alapján nyilvánvaló, hogy:

$$P(X_i = k) = p \cdot q^{k-1}; k \geq 1.$$

Szintén evidens, hogy pontosan annyi – egyszerre történő – feldobás szükséges a kísérlet befejezéséhez, amennyi az egyes pénzérmék esetén az első fejdobás bekövetkezéséhez szükséges dobásszámok közül a legnagyobb. Ily módon:

$$Y = \max(X_1, X_2, \dots, X_n) = \max_{1 \leq i \leq n} (X_i).$$

**3.1.1. Állítás:**  $G(y) = \left[ \sum_{k=1}^y q^{k-1} \cdot p \right]^n$ .

**Bizonyítás:** Az előbbi megállapítás alapján igaz, hogy:

$$\begin{aligned} G(y) &= P(Y \leq y) = P[\max(X_1, X_2, \dots, X_n) \leq y] = \\ &= P[(X_1 \leq y) \cap (X_2 \leq y) \cap \dots \cap (X_n \leq y)] \end{aligned}$$

Mivel az egyes  $(X_i \leq y \mid i = 1, 2, \dots, n)$  események függetlenek:

$$[(X_1 \leq y) \cap (X_2 \leq y) \cap \dots \cap (X_n \leq y)] = P(X_1 \leq y) \cdot P(X_2 \leq y) \cdot \dots \cdot P(X_n \leq y);$$

ahol az  $X_1, X_2, \dots, X_n$  valószínűségi változók Pascal-eloszlásúak azonos  $(p)$  paraméterrel. Ennélfogva:

$$\begin{aligned} G(y) &= P(X_1 \leq y) \cdot P(X_2 \leq y) \cdot \dots \cdot P(X_n \leq y) = P(X_i \leq y) \cdot P(X_i \leq y) \cdot \dots \cdot P(X_i \leq y) = \\ &= [P(X_i \leq y)]^n. \end{aligned}$$

Továbbá megállapítható, hogy:

$$\begin{aligned} P(X_i \leq y) &= P[(X_i = 1) \vee (X_i = 2) \vee \dots \vee (X_i = y)] = \\ &= P(X_i = 1) + P(X_i = 2) + \dots + P(X_i = y) = q^0 \cdot p + q^1 \cdot p + \dots + q^{y-1} \cdot p = \sum_{k=1}^y q^{k-1} \cdot p; \end{aligned}$$

így a  $G(y)$  eloszlásfüggvény a következőképpen írható fel:

$$G(y) = P(Y \leq y) = [P(X_i \leq y)]^n = \left[ \sum_{k=1}^y q^{k-1} \cdot p \right]^n.$$

□

### 3.2. A $p_m = P(Y = m)$ függvény

**3.2.1. Állítás:** 
$$p_m = \sum_{k=0}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot q^{km-k} \cdot (1-q^k).$$

**Bizonyítás:** Vegyük észre, hogy:

$$P(Y = m) = P(Y \leq m) - P(Y \leq m-1).$$

Ily módon – a 3.1.1. állítás szerint –  $p_m$  könnyen formalizálható:

$$p_m = P(Y = m) = P(Y \leq m) - P(Y \leq m-1) = \left[ \sum_{k=1}^m q^{k-1} \cdot p \right]^n - \left[ \sum_{k=1}^{m-1} q^{k-1} \cdot p \right]^n.$$

A véges mértani sor összegére vonatkozó összefüggés ismeretében a formulában szereplő két összegre a következőképp adhatunk zárt képleteket:

$$\sum_{k=1}^m q^{k-1} \cdot p = p \cdot (1 + q + q^2 + \dots + q^{m-1}) = p \cdot \frac{1-q^m}{1-q} = 1 - q^m;$$

$$\sum_{k=1}^{m-1} q^{k-1} \cdot p = p \cdot (1 + q + q^2 + \dots + q^{m-2}) = p \cdot \frac{1-q^{m-1}}{1-q} = 1 - q^{m-1}.$$

A fentiek és a *binomiális tétel* alapján:

$$\begin{aligned} p_m &= (1 - q^m)^n - (1 - q^{m-1})^n = \sum_{k=0}^n (-1)^k \cdot \binom{n}{k} \cdot q^{mk} - \sum_{k=0}^n (-1)^k \cdot \binom{n}{k} \cdot q^{(m-1)k} = \\ &= \sum_{k=0}^n (-1)^k \cdot \binom{n}{k} \cdot q^{km-k} \cdot (q^k - 1) = \sum_{k=0}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot q^{km-k} \cdot (1 - q^k). \end{aligned}$$

□

### 3.3. $Y$ várható értéke

**3.3.1. Lemma:** 
$$\sum_{m=1}^{\infty} km \cdot q^{km} = \frac{k \cdot q^k}{(1 - q^k)^2}.$$

**Bizonyítás:** 
$$\sum_{m=1}^{\infty} km \cdot q^{km} = k \cdot 1 \cdot q^k + k \cdot 2 \cdot q^{2k} + k \cdot 3 \cdot q^{3k} + \dots$$

Vezessük be az  $a := q^k$  jelölést. Így:

$$\begin{aligned} \sum_{m=1}^{\infty} km \cdot q^{km} &= k \cdot 1 \cdot a + k \cdot 2 \cdot a^2 + k \cdot 3 \cdot a^3 + \dots = k \cdot (a + 2 \cdot a^2 + 3 \cdot a^3 + \dots) = \\ &= a \cdot k \cdot (1 + 2 \cdot a + 3 \cdot a^2 + \dots) = a \cdot k \cdot (a + a^2 + a^3 + \dots)'. \end{aligned}$$

A véges mértani sor összegére vonatkozó összefüggés és a deriválási szabályok ismeretében:

$$\begin{aligned} a \cdot k \cdot (a + a^2 + a^3 + \dots)' &= a \cdot k \cdot \left( \frac{a}{1-a} \right)' = a \cdot k \cdot \frac{a' \cdot (1-a) - a \cdot (1-a)'}{(1-a)^2} = \\ &= a \cdot k \cdot \frac{1-a-a \cdot (0-1)}{(1-a)^2} = \frac{a \cdot k}{(1-a)^2}. \end{aligned}$$

Ily módon:

$$\sum_{m=1}^{\infty} km \cdot q^{km} = \frac{q^k \cdot k}{(1 - q^k)^2}.$$

□

**3.3.2. Állítás:** 
$$E(Y) = \sum_{k=1}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{1}{1 - q^k}.$$

**Bizonyítás:** Korábbi megállapításunk szerint  $E(Y) = \sum_{m=1}^{\infty} m \cdot p_m$ . Így a 3.2.1. állítás

alapján:

$$E(Y) = \sum_{m=1}^{\infty} m \cdot \left[ \sum_{k=1}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot q^{km-k} \cdot (1 - q^k) \right].$$

A megfelelő átrendezések után a következőt nyerjük:

$$E(Y) = \sum_{k=1}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{1}{q^k} \cdot (1-q^k) \cdot \sum_{m=1}^{\infty} m \cdot q^{km}.$$

Ahhoz, hogy a 3.3.1. lemmát alkalmazni tudjuk; a kapott kifejezést szorozzuk meg, és osszuk is el  $k$ -val:

$$\begin{aligned} E(Y) &= \sum_{k=1}^n \frac{1}{k} \cdot (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{1}{q^k} \cdot (1-q^k) \cdot \sum_{m=1}^{\infty} km \cdot q^{km} = \\ &= \sum_{k=1}^n \frac{1}{k} (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{1}{q^k} \cdot (1-q^k) \cdot \frac{q^k \cdot k}{(1-q^k)^2} = \sum_{k=1}^n (-1)^{k+1} \cdot \binom{n}{k} \cdot \frac{1}{1-q^k}. \end{aligned}$$

□

## 4. Kitekintés: Erdős Pál és Rényi Alfréd eredményei

### 4.1. Erdős és Rényi vonatkozó tétele

Erdős Pál és Rényi Alfréd kiemelkedő eredményeket fogalmaztak meg hasonló játékok leghosszabb homogén nyeréssorozataival kapcsolatban. Kezdeti vizsgálódásaik széles kutatási területet nyitottak: Publikációik nyomán számos, ezzel a témával foglalkozó cikk és tanulmány jelent meg.

Az eredményeik alapjául szolgáló tétel a 4.1.1. pontban olvasható; a bizonyításához szükséges ismeretek azonban túlmutatnak a Matematika BSc képzés ismeretanyagának spektrumán, így ennek tárgyalásától eltekintek. A tétel kimondását követően azon fogalmakat és állításokat vázolom fel, melyek szintén nem mind részei a tanári szakirány tananyagának, de ismeretük elengedhetetlen a bizonyítás megkonstruálásához.

Legyenek adottak az  $X_1, X_2, \dots, X_n$  független, azonos eloszlású valószínűségi változók, ahol:

$$X_i = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Látható, hogy egy  $X_i$  valószínűségi változó megfeleltethető az  $n$  hosszú dobássorozat egy konkrét pénzfeldobásának. Reprézntálja a fejdobás bekövetkezését az „1” érték felvétele.

Legyenek  $S_n$  és  $I(N, K)$  a következőképpen definiálva:

$$S_0 = 0; S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i ;$$
$$I(N, K) = \max_{0 \leq n \leq N-K; N \geq K} (S_{n+K} - S_n).$$

Vegyük észre, hogy igaz a következő:

$$\exists K [I(N, K) = K]$$



(a dobássorozatban van  $K$  hosszúságú homogén fejsorozat).

Jelölje  $Z_N$  ( $N = 1, 2, \dots$ ) azon  $K$  értékek közül a legnagyobbat, melyekre teljesül, hogy  $I(N, K) = K$ . Megállapítható, hogy  $Z_N$  értéke megegyezik az  $n$  hosszú pénzfeldobás-sorozatban előforduló leghosszabb homogén fejsorozat hosszával.

**4.1.1. Tétel (Erdős, Rényi):** Legyenek adottak a  $C_1, C_2$  konstansok, melyekre igaz, hogy:  $0 < C_1 < 1 < C_2$ . Jelölje  $\Omega$  az összes esemény halmazát.

Ekkor minden nem 0 valószínűségű  $\omega \in \Omega$ -hoz található olyan  $N_0 = N_0(\omega, C_1, C_2)$  véges szám, hogy minden  $N \geq N_0$  esetén teljesül a következő:

$$[C_1 \cdot \log_2(N)] \leq Z_N \leq [C_2 \cdot \log_2(N)].$$

## 4.2. A tétel bizonyításához szükséges fogalmak és állítások

**4.2.2. Definíció:** Az  $M_s = \sum_{k=1}^{\infty} k^s \cdot p_k$  számot, azaz az

$$X = \begin{pmatrix} 1 & 2 & \dots & k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$$

valószínűségi változó  $s$ -edik hatványának várható értékét az  $X$  valószínűségi változó  **$s$ -edik momentumának** nevezzük.

**4.2.3. Definíció:** A  $\phi(t) = \sum_{s=0}^{\infty} \frac{M_s}{s!} \cdot t^s$  függvényt az

$$X = \begin{pmatrix} 1 & 2 & \dots & k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$$

valószínűségi változó **momentumgeneráló függvényének** nevezzük.



**4.2.4. Definíció:** A  $h(X) = \sum_{i=1}^k -p_i \cdot \log_2(p_i)$  számot az

$$X = \begin{pmatrix} 1 & 2 & \dots & k \\ p_1 & p_2 & \dots & p_k \end{pmatrix}$$

valószínűségi változó  $\{p_1, p_2, \dots, p_k\}$  valószínűségeloszlásához tartozó **entrópiának** nevezzük.

**4.2.5. A Stirling-formula:**

$$n! \sim \left(\frac{n}{e}\right)^n \cdot \sqrt{2 \cdot n \cdot \pi}.$$

(A fenti képletben a „ $\sim$ ” szimbólum *aszimptotikus egyenlőséget* jelöl.)

**4.2.6. A Borel-Cantelli lemma:**

Legyenek adottak az  $A_1, A_2, \dots, A_n, \dots$  események.

(1) Ha  $\sum_{n=1}^{\infty} P(A_n) < \infty$ , akkor 1 a valószínűsége annak, hogy az  $A_1, A_2, \dots, A_n, \dots$

események közül egyidejűleg legfeljebb véges sok következik be.

(2) Ha az  $A_1, A_2, \dots, A_n, \dots$  események *teljesen függetlenek*, és  $\sum_{n=1}^{\infty} P(A_n) = \infty$ , akkor

1 a valószínűsége annak, hogy közülük egyidejűleg végtelen sok bekövetkezik.

## Irodalomjegyzék

ERDŐS P., RÉNYI A.

On a new law of large numbers, *Journal d'Analyse Mathématique*, Vol. 23, (1970), 103–111.

ERDŐS P., RÉVÉSZ P.

On the length of the longest head-run, *Colloquia Mathematica Societatis János Bolyai*, No. 16, (1977), 219–228.

GORDON, L.; SCHILLING, M. F; WATERMAN, M. S.

An extreme value theory for long head runs, *Probability Theory and Related Fields*, Vol. 72, (1986), 279–281.

KINNEY, J.

Tossing coins until all are heads, *Mathematics Magazine*, Vol. 51, No. 3, (1978), 184–186.

RÉNYI A.

*Valószínűségszámítás*, Tankönyvkiadó, Budapest, 1966, 130–131., 323–324.

SCHILLING, M. F.

The longest run of heads, *The College Mathematics Journal*, Vol. 21, No. 3, (1990), 196–207.

SHANNON, C. E.; WEAVER, W.

*A kommunikáció matematikai elmélete*, Országos Műszaki Információs Központ és Könyvtár, Budapest, 1986.