Valószínűségszámítás speci I. éves matematika tanárszakos hallgatóknak

Csiszár Villő

Ajánlott irodalom:
Székely J. Gábor: Paradoxonok a véletlen matematikájában (Műszaki könyvkiadó, 1982)
Székely J. Gábor: Paradoxonok a véletlen matematikájában, 2. átdolgozott kiadás (Typotex, 2004)
Warren Weaver: Szerencse kisasszony (Kairosz kiadó, 1997)
Denkinger Géza: Valószínűségszámítás (Nemzeti Tankönyvkiadó, 1997)
Számonkérés: A félévet írásbeli vizsga zárja.

1. A valószínűségszámítás tárgya, véletlen kísérlet matematikai modellje

Véletlen kísérlet: az eredménye nem jósolható meg előre, a véletlentől függ. A véletlent nem filozófiai oldalról közelítjük meg, hanem gyakorlati oldalról. Az életben tapasztaljuk, hogy vannak véletlen jelenségek, sőt, ezeknek bizonyos törvényszerűségei is vannak (pl. nagy számok törvénye).

Pl.: közvéleménykutatás: 2000 embert megkérdezve, az már jól fogja tükrözni a népesség jellemzőit; kaszinó hosszú távon nyereséges, mert ki tudja számítani, hogy az emberek átlagosan mennyit nyernek.

Akkor tanulmányozhatók jól *a véletlen törvényszerűségei*, ha a kísérletet sokszor meg tudjuk ismételni, azonos körülmények között: véletlen tömegjelenségek. Ilyenekre a legjobb példákat a szerencsejátékok szolgáltatják. (Bonyolultabb esetek: tőzsde-árfolyamingadozások, biztosítás-káresemények, biológia-járvány terjedése, stb.)

Célunk, hogy ezekre a véletlen jelenségekre *matematikai modellt* alkossunk, mégpedig úgy, hogy a modellből kiszámolt eredmények jól használhatóak legyenek a gyakorlatban.

1. Bár az eredményt nem tudjuk előre megjósolni, tudjuk, hogy mik a *lehetséges kimenetelek*. **Matematikailag** ezek egy (véges vagy végtelen) halmazt alkotnak. (Ennek neve: eseménytér, jel: Ω) Hogy mit tekintünk kimenetelnek, az rajtunk is múlik, hogy mit tartunk érdekesnek, fontosnak (pl. lottóhúzásnál csak az öt számot, vagy a sorrendjüket is).

2. Esemény: **matematikailag** kimenetelek egy halmaza (jel. pl. A). Ez vagy bekövetkezik, vagy nem. Az eseményekkel műveletek is végezhetők (A komplementere: pontosan akkor következik be, ha A nem; A és B uniója: akkor következik be, ha A és B legalább egyike bekövetkezik; A és B metszete: akkor következik be, ha A és B is bekövetkezik)

3. Minden kimenetelhez, eseményhez hozzárendelünk egy valószínűséget. Az A esemény valószínűségét jelölje P(A) (P, mint *probability*). **Matematikailag**?

A hétköznapi életben azt értjük egy esemény valószínűségén, hogy sok független kísérletet elvégezve, hány százalékban következik be: $P(A) \approx k_A/n$ ha n nagy. Ezért úgy szeretnénk definiálni a valószínűséget, hogy a relatív gyakoriság tulajdonságai érvényesüljenek → a valószínűség axiómái:

i, Bármely esemény valószínűsége egy 0 és 1 közötti szám. ii, A biztos esemény valószínűsége 1.

iii, Egymást kizáró eseményekre (A és B metszete üres) annak valószínűsége, hogy legalább egy bekövetkezik közülük, megegyezik az egyes események valószínűségeinek összegével.

Egy kísérlet **matematikai modelljén** tehát azt értjük, hogy megadjuk a lehetséges kimeneteleket (ill. az eseményeket), és ezek valószínűségeit.

Legegyszerűbb eset: minden kimenetelnek ugyanakkora az esélye (ha véges sok kimenetel van). Annak eldöntésénél, hogy egy kísérletben egyformán esélyesek-e a kimenetelek, szimmetria okokra lehet hivatkozni (pl. kockadobásnál, lottóhúzásnál).

Ha a lehetséges kimenetelek egyformán valószínűek (klasszikus eset), akkor egy esemény valószínűsége = kedvező kimenetelek száma / összes kimenetel száma.

Példák: Szabályos dobókockával dobunk: a 6 lehetséges érték egyformán valószínű.

Két kockával dobva 36 egyformán valószínű kimenetel van (fával ábrázolható). Ha csak a két számra vagyunk kíváncsiak (mindegy, hogy milyen sorrendben jöttek ki), akkor 21 kimenetel van, de azok nem egyformán esélyesek (KI LEHET PRÓBÁLNI!). A két kockát a természet mindenképp meg tudja különbözteti, még ha nekünk egyformának tűnnek is! A valóságot tehát az a modell írja le jól, amely a fizikailag különböző objektumokat különbözőnek tekinti, függetlenül attól, hogy mi, mint tökéletlen megfigyelők, meg tudjuk-e különböztetni őket egymástól. (Kivétel:

fizikai kísérletek kimutatták, hogy az elemi részecskék nem mindig így viselkednek, pl. a protonok és a fotonok.)

Események valószínűségének kiszámításához érdemes az egyforma vszgű kimenetelekből kiindulni, így használható a klasszikus képlet (pl. két kockánál melyik esélyesebb: összeg = 9, összeg = 10).

Lottó: 90 számból 5-öt húzunk visszatevés nélkül (sorrend számít-e?) – úgy is lehet venni, hogy számít, úgy is, hogy nem

90*89*88*87*86 = 5 273 912 160 120-val osztva: 43 949 268.
(mind egyformán valószínű kimenetelek)

2. A valószínűségszámítás születése (röviden)

A szerencsejátékok tanulmányozásából alakult ki a valószínűségszámítás, kezdetben nem volt egyéb, mint a kedvező/összes képlet alkalmazása.

Cardano: Liber de Ludo Aleae (A kockajátékok könyve), 1663-ban jelent meg, de több, mint száz évvel korábban íródott. (arabul *az-zar = kocka*, ebből származik a *hazárd* = kockázat szó)

1654: Antoine Gombauld, Méré lovagja (de Méré lovag), művelt, kíváncsi alkat, játéktermeket is látogatott. Ismerte a **Régi Szerencsejátékos-szabályt**, amely arra vonatkozott, hogy egy adott esemény bekövetkezésére nézve mi az a kritikus kísérletszám, amely esetén az esélyek kedvezőtlenről kedvezőre váltanak:

Ha a p_1 valószínűségű eseményre a kritikus szám n_1, akkor a p_2 valószínűségű eseményre $n_2 = n_1 * p_1 / p_2$.

Pl. a szabályt alkalmazva, ha tudjuk, hogy a "hatos dobás" kritikus száma 4, akkor a "dupla hatos"-é 24. (három dobásból 0,42 eséllyel lesz 6-os, négy dobásból 0,52 eséllyel) Igaz-e ez?

NEM! Mekkora az esélye, hogy n kísérletből legalább egyszer bekövetkezik a $p = k/m$ valószínűségű A esemény?

$1 - (m - k)^n / m^n = 1 - (1 - p)^n$. Ez akkor lesz legalább ½, ha n legalább $-\log 2/\log(1-p)$. $\log(1-p)$-t sorbafejtve, helyesen: n = 0,6931 / ($p + p^2/2 + p^3/3 + \dots$) felső egészrésze, ami elég kis p esetén megegyezik 0,6931 / p felső egészrészével. Azaz ekkor igaz, hogy

np közel konstans. (A dupla hatos kritikus száma egyébként 25.)

A lovag megkérdezte Blaise Pascalt is, aki megoldotta a feladatot. Később az osztozkodási problémát (mely 1494-ből származott) is megoldotta (Két játékos egy igazságos játékban hat győzelemig játszik, de 5:3-nál abbahagyják: hogyan osztozzanak meg a téten? A helyes válasz: 7:1 arányban kell osztozni.), majd Pierre de Fermat-val kezdett ezekről és más szerencsejátékokról levelezni – megszületett a valószínűségszámítás.

A modern valószínűségszámítás megalapítója A.N. Kolmogorov, aki az 1930-as években dolgozta ki a valószínűségszámítás mai felépítését.

3. Körbeverés: kockákra, számhúzásra (Székely I/13f)

- *A* és *B* azt a játékot játssza, hogy *A* megszámoz három kockát az 1-18 számokkal, *B* választ egyet, majd a maradék kettőből *A* egyet. Az nyer, aki nagyobb számot dob. Kinek előnyös a játék?
- Válasz: *A*-nak előnyös, mert legyen **I**: 1, 6, 11, 12, 13, 14 **II**: 2, 3, 4, 15, 16, 17 **III**: 5, 7, 8, 9, 10, 18
  P(**II** megveri **I**-t) = P(**III** megveri **II**-t) = P(**I** megveri **III**-t) = 21/36, azaz a kockák 21/36 valószínűséggel körbeverik egymást. Tehát akármelyik kockát választja is *B*, *A* tud az övénél jobbat választani. HF: Lehet-e ennél jobb számozást megadni (*A* szempontjából)?

Def: Egy véletlentől függő mennyiséget *valószínűségi változónak* nevezünk. Az X (diszkrét) valószínűségi változó megadása: felsoroljuk X lehetséges értékeit: $x_1, x_2, x_3, \dots$ és a hozzájuk tartozó valószínűségeket: $p_1, p_2, p_3, \dots$

Def: Az $X_1, X_2, \dots, X_n$ valószínűségi változók *körbeverik egymást*, ha $P(X_2 > X_1) = p_1$, $P(X_3 > X_2) = p_2, \dots, P(X_1 > X_n) = p_n$ jelöléssel $p_i > ½$ minden i-re.

A körbeverési valószínűség a p_i számok minimuma.

Az előző példában $X_{\mathbf{I}}$, $X_{\mathbf{II}}$, $X_{\mathbf{III}}$ a három kockával dobott érték, ezek nyilván függetlenek egymástól, nincsenek egymásra hatással, valamelyik ismerete nem mond semmit a többiről (a függetlenséget matematikailag is lehet definiálni, HF: vajon hogyan?).

- Mennyi a körbeverési valószínűség maximuma, ha n véletlen mennyiség nem feltétlenül független?
- Válasz: $(n–1)/n$.

  Megvalósítás: A és B játékosok játszanak, először A választ egy számot 1 és n között, majd B is, de nem választhatja ugyanazt, mint A. Legyenek a számaik k_A és k_B.

  Ezután a játékvezető kisorsol 0 és $(n–1)$ között egy X számot, véletlenszerűen. A játékosok ezt hozzáadják saját számukhoz, és az n-nel osztva kapott maradékot kapják meg, azaz az $Y_A = k_A+X \bmod n$ és az $Y_B = k_B+X \bmod n$ számokat. Itt Y_A és Y_B nyilván összefüggnek, hiszen ugyanaz a véletlen X mennyiség szerepel bennük. Ez a játék is B számára előnyös, hiszen az $Y_i = i+X \bmod n$ véletlen mennyiségek körbeverik egymást, amint ez az alábbi táblázatból látszik (n = 5-re):

X	Y_1	Y_2	Y_3	Y_4	Y_5
0	1	2	3	4	0
1	2	3	4	0	1
2	3	4	0	1	2
3	4	0	1	2	3
4	0	1	2	3	4

$P(Y_{i+1} > Y_i) = (n–1)/n$. Ez a játék hasonlít az előzőhöz, úgy képzelhetjük, hogy itt n db. “kocka” van, de ezek egy fonállal össze vannak kötve, úgy, hogy bármelyik kockán a dobott szám meghatározza az összes többi kockán dobott számot is.

Belátjuk, hogy ennél tényleg nem lehet nagyobb a körbeverési valószínűség. Tegyük fel, hogy n mennyiség körbeveri egymást. $A_k := \{X_{k+1} > X_k\}$, $P(A_k) := p_k$ jelöléssel ($k = 1, ..., n$), mivel ennek az n eseménynek a metszete üres, komplementereik uniója a biztos esemény. Mivel tehát ezek a komplementer események lefedik az összes lehetőséget, valószínűségeik összege legalább egy. Ezért van köztük olyan, amelyik legalább $1/n$ valószínűségű, azaz olyan k, hogy $1– p_k \geq 1/n$. Erre a k-ra $p_k \leq (n\text{-}1)/n$, azaz $\min_j p_j \leq (n–1)/n$.

Megmutatható, hogy független mennyiségek esetén a körbeverési valószínűség maximuma (monoton növekvő módon) ¾-hez tart. (Usiskin eredménye)

4. Névjegy-probléma (Székely I/6, Weaver 119)

- n ember esetén mekkora az esélye, hogy névjegyeiket véletlenszerűen összecserélve, senki sem a sajátját kapja vissza?
- Válasz: kb. $1/e$, azaz 37%. (ha legalább 4 ember játszik). – ez a szita-formulával számítható ki.

Pl. $n = 4$:

1-es a helyén van: 1342, 1423, *1243*, *1432*, *1324*, **1234**

2-es a helyén van: 3241, 4213, *1243*, *4231*, *3214*, **1234**

3-as a helyén van: 2431, 4132, *1432*, *4231*, *2134*, **1234**

4-es a helyén van: 3124, 2314, *1324*, *3214*, *2134*, **1234**

Összesen 15 olyan permutáció van, amelyben valamelyik szám a helyén van, tehát 9/24 = 0,375 a valószínűsége, hogy egyik sincs a helyén.

Szita-formula: n darab esemény uniójának valószínűségét – a logikai szitához hasonlóan – úgy számíthatjuk ki, hogy összeadjuk az n esemény valószínűségeit, majd kivonjuk belőle a kettes metszetek valószínűségeit, hozzáadjuk a hármas metszetek valószínűségeit, és így tovább az n-es metszetig. Ha most A_i az az esemény, hogy az i. ember a saját névjegyét kapja vissza, akkor ezek uniójára, azaz arra az eseményre, hogy legalább egy ember a sajátját kapja vissza, a

$\sum_{k=1}^{n}(-1)^{k-1}\frac{1}{k!}=1-\sum_{k=0}^{n}(-1)^{k}\frac{1}{k!}$ valószínűség adódik. Ez pedig n növekedésével gyorsan tart az $1 - 1/e$ számhoz.

- Átlagosan hány ember kapja a sajátját?
- Válasz: nagyon sokszor, mondjuk M-szer, kiosztva a névjegyeket, tegyük fel, hogy a j. kiosztásnál n_j ember kapta a saját névjegyét. Jelölje még k_j azt, hogy hány kísérletben kapta j ember a saját névjegyét. Az átlag tehát $(n_1 + \ldots + n_M)/M = 0*k_0/M + 1*k_1/M + \ldots + n*k_n/M$. Mivel a k_j / M relatív gyakoriság közel van a p_j = P(pontosan j ember kapja a saját névjegyét) valószínűséghez, mondhatjuk, hogy nagy átlagban $0*p_0 + 1*p_1 + \ldots + n*p_n$ ember kapja a saját névjegyét.

Def: Az X valószínűségi változó *várható értéke* $E(X) = x_1* p_1 + x_2* p_2 + x_3* p_3 + \ldots$

Ahhoz, hogy a várható értéket meghatározzuk, ismerni kellene a p_j valószínűségeket. Ezeket elég bonyolult kiszámolni.

A fenti példában n = 4-re leszámolhatjuk, hogy $p_0 = 9/24$, $p_1 = 8/24$, $p_2 = 6/24$, $p_3 = 0/24$, $p_4 = 1/24$. A várható érték tehát pontosan 1.

Az általános esetben a következő trükkhöz folyamodhatunk: az $n_1 + \ldots + n_M$ összeget úgy is megkaphatjuk, hogy minden emberről megszámoljuk, hogy hányszor kapta a saját névjegyét, és ezeket összeadjuk. Legyen u_j azon kísérletek száma, amikor a j. ember a saját névjegyét kapta. Ekkor tehát $n_1 + \ldots + n_M = u_1 + \ldots + u_n$. Így az átlag: $u_1/M + \ldots + u_n/M$. Viszont az u_j/M relatív gyakoriság közel van a P(a j. ember a saját névjegyét kapta) valószínűséghez, ami nem más, mint $1/n$. Így átlagosan $n*1/n = 1$ ember kapja a saját névjegyét, minden n-re.

A trükk lényege abban áll, hogy egy valószínűségi változót összegre bontunk, és várható értékét tagonként számoljuk ki. Pl.:

permutáció	hány van a helyén (X)	1-es helyén van-e (X_1)	2-es helyén van-e (X_2)	3-as helyén van-e (X_3)	4-es helyén van-e (X_4)
1234	4	1	1	1	1
3124	1	0	0	0	1
3214	2	0	1	0	1
4123	0	0	0	0	0
2431	1	0	0	1	0

5. A születésnap-paradoxon és az ajándékgyűjtő játék

Születésnap-paradoxon (Egy szimulációs program – angol nyelven – elérhető a következő helyen: http://www.mste.uiuc.edu/exner/java.f/birthday/default.html)

- n ember esetén mekkora az esély, hogy van köztük legalább kettő, akik ugyanazon hónap ugyanazon napján születtek, tehát ugyanakkor van a születésnapjuk? Tegyük fel az egyszerűség kedvéért, hogy az év 365 napból áll, és minden ember születésnapja, egymástól függetlenül, véletlenszerűen esik e 365 nap valamelyikére.
- Válasz: Pl. néhány n-re: 366: 100%, 68: 99,9%, 55: 90%, **23: 50%**. Képlettel is megadható: $1 - 365*364*\ldots*(366-n)/365^n$. Ebben a feladatban nincs semmi nehézség, az eredmény viszont meglepő, ezért nevezik paradoxonnak.
- Átlagosan hányadik embernél következik be az első ismétlődés?

- Válasz: Annak az esélye, hogy a *k*. embernél következik be az első ismétlődés, $p_k = (365*364*\ldots*(367–k)*(k–1))/365^k$. A várható érték tehát $2*p_2 + 3*p_3 + 4*p_4 +\ldots+ 366*p_{366} = 24{,}62$. HF: Vajon itt lehet-e összegre bontással egyszerűbb képletet kapni? HF: *n* ember esetén mekkora az esély, hogy valakinek ugyanakkor van a szülinapja, mint nekem?

Ajándékgyűjtő játék (Egy szimulációs program – angol nyelven – elérhető a következő helyen: http://www.mste.uiuc.edu/reese/cereal/intro.html)

- A reggelizőpelyhes dobozokban *n* különböző ajándékot lehet találni. Átlagosan hány dobozt kell vásárolni ahhoz, hogy mind az *n*-féle ajándékot összegyűjtsük? (vagy: dobókockával hányszor kell dobni, hogy mind a hat szám kijöjjön?)
- Válasz: Kellene annak a valószínűsége, hogy pontosan *k* dobozból gyűlik össze mindegyik ajándék. Ezt bonyolult kiszámítani. Összegre bontással egyszerűbb: X_1: hány doboz kell ahhoz, hogy egyféle ajándékunk már legyen, X_2: ezután még hány doboz kell ahhoz, hogy már kétféle ajándékunk legyen, stb. Pl. ha négy ajándék van, és az egymás utáni dobozokban a 34413132 ajándékokat találjuk, akkor $X_1 = 1$, $X_2 = 1$, $X_1 = 2$, $X_1 = 4$, és $X = 8$.

Mennyi X_k várható értéke? Ekkor (k–1)-féle ajándékunk már van, tehát minden további dobozban $(n–k+1)/n$ eséllyel találunk új ajándékot.

- Átlagosan mennyit kell várni egy *p* valószínűségű eseményre?
- Válasz: annak az esélye, hogy *k* kísérletet kell rá várni, $p_k = (1–p)^{k–1}p$. Tehát a várható érték $1*p_1 + 2*p_2 + 3*p_3 +\ldots = 1/p$. Máshogy: a várakozási időt fel lehet írni végtelen sok valószínűségi változó összegeként, ahol Y_k akkor 1, ha a *k*. kísérlet előtt nem következett be a várt esemény, egyébként pedig nulla. Y_k várható értéke $(1–p)^{k–1}$. Tehát az összegük várható értéke $1/p$.
- Az összes ajándék tehát átlagosan $n/n + n/(n-1) + n/(n-2) + \ldots + n/2 + n/1$ dobozból gyűlik össze.

Mekkora az esélye, hogy *k* doboz nem elegendő? Ez a szita-formulával írható fel. Pl. ha négyféle ajándékot kell összegyűjtenem, akkor 7 doboz megvásárlása után már 51% az esélyem, hogy mind megvan, és átlagosan 8,3 doboz kell.

HF: tegyük fel, hogy az ajándékok állatos kártyák, méghozzá hatféle állat van: agár, béka, cica, darázs, elefánt, fóka. Átlagosan hány dobozt kell vennem, ha csak négyféle kártyát szeretnék összegyűjteni? Átlagosan hány dobozt kell vennem, ha az agarat, békát, cicát, darázst szeretném összegyújteni?

6. Fej-írás sorozatok versenye, a Conway-algoritmus

- Tekintsük a következő játékot: Két (vagy több) játékos mindegyike választ magának egy fej-írás sorozatot, majd egy szabályos érmét elkezdenek dobálni. Az nyer, akinek a sorozata először felbukkan. Milyen sorozatot érdemes választani? (Székely I/13c)
- Pl: Anna sorozata: IFF, Bea sorozata: IIF. A dobássorozat: FFFIFIIIIF – Bea nyert, a játék 10 dobás hosszan tartott.
- Igaz-e, hogy egyenlőek a nyerési esélyek? Hiszen érvelhetünk azzal, hogy minden 3 hosszú sorozat ugyanolyan valószínűséggel jön ki! Mégsem igaz, pl. az IFF és FFI sorozatoknál, FFI csak akkor nyerhet, ha az első két dobás F. Ekkor ugyanis még egy ideig F fog jönni, de előbb-utóbb megjelenik egy I, és ezzel FFI nyert. Ha viszont az első két dobásban volt I is, akkor a játék addig folytatódik, amíg megjelenik egymás után két F, és ezzel IFF nyer. Tehát a két sorozat versenyében P(FFI nyer) = ¼, P(IFF nyer) = ¾. Nem minden esetben ilyen egyszerű a valószínűségeket látni.
- A játék menetét *gráffal* lehet ábrázolni, azaz felrajzoljuk a lehetséges állapotokat, és hogy melyikből melyikekbe léphetünk. Egy állapot azt fejezi ki, hogy a dobássorozat végén mi az a maximális részsorozat, ami még felhasználható valamelyik célsorozat felépítéséhez. Minden állapotból két nyíl vezet kifelé annak megfelelően, hogy a következő dobás fej-e vagy írás. Azok az állapotok, amikor valamelyik sorozat elkészült, a végállapotok, ekkor befejeződik a játék.
- Az egyes sorozatok nyerési esélyét egyenletrendszer felírásával és megoldásával számíthatjuk ki. Pl. az IFF és IIF versenyében 1/3 valószínűséggel nyer IFF (2/3 valószínűséggel pedig IIF): Legyenek az állapotok a következők: 0 – kezdő, 1 – I, 2 – IF, 3 – II, 4 – IFF (végállapot), 5 – IIF (végállapot). Jelölje x_k, hogy a k. állapotból indulva, mekkora eséllyel lesz IFF a nyertes. Ekkor nyilván $x_4 = 1$ (IFF már nyert is), és $x_5 = 0$ (IIF nyert). Ezt felhasználva a többi állapotra a következő egyenleteket írhatjuk fel:

$$x_0 = \frac{1}{2}x_0 + \frac{1}{2}x_1 \qquad x_1 = \frac{1}{2}x_2 + \frac{1}{2}x_3 \qquad x_2 = \frac{1}{2}\cdot 1 + \frac{1}{2}x_1 \qquad x_3 = \frac{1}{2}\cdot 0 + \frac{1}{2}x_3$$

Ennek megoldása $x_0 = 1/3$.

- Könnyen látszik, hogy a játék (dobásszámban mért) hosszának átlagos értéke felülről becsülhető. Ha *a* darab *n* hosszú sorozat versenyez egymással, akkor a dobássorozatot *n* dobásból álló blokkokra felosztva, minden blokkban $a/2^n$ eséllyel éppen a versenyző sorozatok valamelyike fog megjelenni. Korábban megmutattuk, hogy annak a blokknak az átlagos sorszáma, amelyben először megjelenik valamelyik versenyző, éppen $2^n/a$. Mivel egy blokk *n* dobásból áll, átlagosan $2^n n/a$ dobás kell ahhoz, hogy valamelyik sorozat *egy blokkban* megjelenjen. A játék legkésőbb ekkor véget ér, persze általában ennél hamarabb, hiszen az is elég, ha két blokkon átnyúlva jelenik meg egy sorozat. Tehát pl. egy *n* hosszú sorozat megjelenésére átlagosan biztosan kevesebb, mint $2^n n$ dobást kell várni.
- Pl. az IFF, FFI sorozatoknál, ha az első két dobás F, akkor ezután az első I megjelenésére átlagosan két dobást kell várni. Így ebben az esetben átlagosan 2+2 = 4 dobásra van szükség. Ha az első dobás I, akkor F-re átlagosan két dobást kell várni. Ezután vagy F, vagy I jön. Ha F, akkor vége a játéknak, ha I, akkor folytatódik. Tehát átlagosan 3 hosszú blokkjaink vannak, és ezekből átlagosan kettő kell, tehát 1 + 3*2 = 7 dobás kell. Ha az első két dobás FI, akkor nyilván átlagosan 1 + 7 = 8 dobás hosszat tart a játék. Tehát az átlagos játékhossz ¼ * 4 + ¼ * 8 + ½ * 7 = 6.5.
- Egyenletrendszerrel számítható ki az is, hogy átlagosan hány dobás (a gráfon lépés) hosszan tart a játék. A korábbi példánál maradva, IFF és IIF versenyében, jelölje m_k, hogy a *k*. állapotból indulva, még átlagosan hány lépés kell a játék befejezéséig. Nyilván $m_4 = m_5 = 0$, hiszen ekkor a játék már véget is ért. Ezt felhasználva a többi állapotra a következő egyenleteket írhatjuk fel:

$$m_0 = 1 + \frac{1}{2}m_0 + \frac{1}{2}m_1 \qquad m_1 = 1 + \frac{1}{2}m_2 + \frac{1}{2}m_3 \qquad m_2 = 1 + \frac{1}{2}\cdot 0 + \frac{1}{2}m_1$$

$$m_3 = 1 + \frac{1}{2}\cdot 0 + \frac{1}{2}m_3$$

Ennek megoldása $m_0 = 16/3$.

Ez a módszer nagyon fáradságos, ha több és hosszabb sorozat van. Egyszerűbb megoldást kínál a **Conway algoritmus** (csak vázlatosan bizonyítjuk).

Ehhez definiáljuk két sorozat átfedési számát:

Azt mondjuk, hogy két sorozat között van *k* hosszúságú átfedés, ha az első sorozat utolsó *k* tagja megegyezik a második sorozat első *k* tagjával. Az A és B sorozat átfedési száma a 2 azon hatványainak összege, melyekre van *k* hosszúságú átfedés. Jelölése: A*B. Pl. IFF*FFI = $2^1 + 2^2 = 6$. FFI*IFF = $2^1 = 2$.

Belátható, hogy az A sorozat megjelenésére átlagosan A*A dobást kell végezni, pl. az FIF sorozathoz 10-et.

Ebből következik, hogy az *n* hosszú sorozatok közül a legtöbbet azokra a sorozatokra kell várni, amelyekre minden átfedés megvan, azaz a csupa egyformából állókra, mégpedig $2^1 + 2^2 + \ldots + 2^n = 2^{n+1} - 2$ dobást. Legkevesebbet pedig azokra a sorozatokra, amelyekre csak *n* hosszúságú átfedés van, pl. a FF…FI sorozatra, 2^n dobást. HF: vajon hány ilyen sorozat van?

Ha több egyforma, mondjuk *n*, hosszú sorozat versenyez egymással, akkor az egyes sorozatok nyerési esélyét illetve a játék átlagos időtartamát egyetlen egyenletrendszer megoldásával kaphatjuk meg. Jelöljük a sorozatokat A, B, …, Z-vel, nyerési esélyüket p_A, p_B, …, p_Z –vel, az átlagos dobásszámot #-vel. Az egyenletek:

p_A A*A + p_B B*A + ... +p_Z Z*A = #

p_A A*B + p_B B*B + ... +p_Z Z*B = #

…

p_A A*Z + p_B B*Z + ... +p_Z Z*Z = #

$p_A + p_B + \ldots + p_Z = 1$

Speciálisan, ha csak egy sorozat van, akkor kapjuk, hogy # = A*A.

Speciálisan, ha két sorozat van, akkor az egyenletrendszer megoldása:

$$p_A = \frac{B*B - B*A}{A*A + B*B - A*B - B*A}, \quad \# = \frac{A*A \cdot B*B - A*B \cdot B*A}{A*A + B*B - A*B - B*A}$$

Az utolsó egyenlet azt fejezi ki, hogy valamelyik sorozat biztosan nyerni fog.

Nézzük az elsőt, mi indokolja az egyenletet? Egy játékvezető egy pénzdarabot dobál addig, amíg az A, B, …, Z sorozatok valamelyike megjelenik. Képzeljük el, hogy valaki a következő fogadásos játékot űzi: mikor beszáll a játékba, feltesz 1 Ft-ot – dupla vagy semmi alalpon – arra, hogy a játékvezető következő dobása éppen az A sorozat első eleme lesz. Ha veszít, kiszáll, viszont ha nyer, akkor a most már 2Ft-ját felteszi arra, hogy a következő dobás az A sorozat második elemét adja. És így tovább, ha nyer, akkor a megduplázott pénzét felteszi a sorozat következő elemére. Ha végül kijön A, akkor mindenképp kiszáll. Ez a játék igazságos, tehát a játékos átlagos nyereménye – minden dobás után – megegyezik a feltett 1Ft-tal. Tegyük most fel, hogy sok játékosunk van, a *k*. játékos a *k*. dobásnál száll be a játékba, egész addig, amíg a játék véget ér. A játék átlagosan # dobásig tart, tehát átlagosan ennyi a befizetett összeg. Nézzük, hogy mennyit nyertek a játékosaink, ha végül pl. a B sorozat jött ki? Akik az utolsó *n* dobás előtt szálltak be, azok mindenképp vesztettek, hiszen az A sorozat nem jött ki. Az pedig, aki csak az utolsó *k* dobásra szállt be ($k \leq n$), az akkor nyert 2^k Ft-ot, ha az A sorozat első *k* eleme – amikre fogadott – megegyezik a B sorozat utolsó *k* elemével – ami kijött. Így éppen B*A Ft a teljes nyeremény. Ezt az összes sorozatra végiggondolva, és nyerési esélyeikkel súlyozva kapjuk az első egyenletet, amely tehát azt fejezi ki, hogy az átlagos nyeremény megegyezik az átlagos befizetéssel. A többi egyenlet hasonlóan adódik, ha a játékosok egy másik sorozatra fogadnak.

Bővebben erről a témáról olvashatunk a www.math.elte.hu/~mori/fej12(1).rtf, a www.math.elte.hu/~mori/fej12(2).rtf, és a www.math.elte.hu/~mori/fej12(3).rtf írásokban, Móri Tamás tollából.

- Ez a témakör is szolgáltat meglepő jelenségeket. Itt is megfigyelhető a körbeverés. Vegyünk egy végtelen hosszú dobássorozatot, és tetszőleges A sorozatra jelölje X_A, hogy hányadik dobásra jelenik meg A először. A és B versenyében $P(X_B > X_A)$ éppen A nyerési esélyét jelenti. Pl. az FFI, FII, IIF, IFF sorozatok ebben az értelemben körbeverik egymást.
- Olyan sorozatok is vannak, melyeknél A átlagosan hamarabb jön ki, mint B, de B mégis 1/2-nél nagyobb valószínűséggel megelõzi A-t:
  tekintsük az A = IFII, B = FIFI sorozatokat. A Conway algoritmus szerint A-ra átlagosan 18 dobást kell várni, míg B-re 20-at. Viszont ha a két sorozat egymással versenyez, akkor 9/14, azaz kb 64% az esélye, hogy B jön ki előbb. (A paradoxon oka: általában egy kicsivel elõbb jön ki a B, de néha nagyon sokkal késõbb.)

7. Pétervári paradoxon

- Daniel Bernoulli az 1700-as évek elején írt cikket a Pétervári Tudományos Akadémia folyóiratában, innen származik az elnevezés.
- A paradox állítás a következő: annak a játéknak, amelynél 2^n forint a nyereség, ha *n*-edikre jön ki elõször fej, az "ára" végtelen.
- Valóban: a várható nyeremény: $2^1 * (½)^1 + 2^2 * (½)^2 + …= 1 + 1 + ….$ Azaz akármennyit is fizetünk a banknak azért, hogy ezt a játékot játszhassuk, a bank hosszú távon rosszul jár. Ez meglepő, hiszen nem nagyon találnánk olyan embert, aki hajlandó volna mondjuk egymillió forintot fizetni, hogy beszállhasson ebbe a játékba! Ennek oka az, hogy ugyan nagyon pici valószínűséggel nagyon nagy lesz a nyereményünk, de ha nem nyerünk az első néhány játékban, akkor tönkremegyünk.
- Hogyan lehetne megadni a játék "reális" árát?
- <u>Egy módosítás</u> (Buffon, Cramer): a gyakorlatban nem bízhatunk abban, hogy a bank akármilyen nagy nyereményt ki fog nekünk fizetni. Ésszerű feltenni, hogy a banknak csak 2^m forintja van, így ha többet kellene fizetnie a szabályok szerint, akkor is csak ennyit tud adni. Ekkor már csupán $m + 1$ Ft a játék méltányos ára, hiszen a várható nyereményünk most $2^1 * (½)^1 + 2^2 * (½)^2 + … + 2^m * (½)^m + 2^m * ((½)^{m+1} + …) = 1 + 1 + … + 1 + 1 = m + 1$.
- <u>Másik módosítás</u> (Feller): akkor méltányos a játék, ha nagy *m* esetén *m* játszma után a nyeremény és a befizetett részvételi díj hányadosa nagy valószínűséggel közel 1 lesz. Kicsit precízebben, tegyük fel, hogy *m* játszmáért R_m forintot kell befizetni, és N_m forintot nyerünk. A feltétel az, hogy tetszőleges előírt *c* esetén annak a valószínűsége, hogy $|N_m/R_m – 1|$ *c*-nél kisebb, egyhez tartson, ha *m* végtelenbe tart. Ezzel a definícióval méltányos a játék, ha *m* játszmáért $m*\log_2 m$ forintot fizetünk (ez összhangban van azzal, hogy egy játszma ára bármilyen *K* konstansnál nagyobb lehet, ha sok játszmára fizetünk be). Így pl. 2048 játszmáért játszmánként 11 forintot méltányos fizetni. (Buffon: 2084 játék alapján azt tapasztalta, hogy 10 forint a méltányos ár játszmánként.)
- <u>Analóg példa</u>: duplázó- (úgynevezett martingál-) stratégia igazságos játékban, pl szabályos érmét dobálva fejnél veszítek, írásnál nyerek. Ez csak végtelen nagy tőkénél (azaz a gyakorlatban nem) működik. A stratégia: az első játszmában 1 forint a tétem. Ha veszítek, akkor a következő játszmában már 2

forintot kockáztatok, majd minden vesztés után megduplázom a tétemet. Ennél a stratégiánál, mivel előbb-utóbb kijön a fej, egy forint a tiszta nyereségem: ha ugyanis n-edikre jön az első fej, akkor elvesztettem $1 + 2 + \ldots + 2^{n-2} = 2^{n-1} - 1$ forintot, viszont nyertem 2^{n-1} összeget. Viszont lehet, hogy már azelőtt tönkremegyek, hogy nyernék. (Székely I/7, Weaver 144)

8. A lóverseny játék

- Lóverseny-játék egyszerűsített változata: Érmét dobálva fej esetén 1-et, írás esetén 2-t lépünk előre. Tegyük fel, hogy a pálya végtelen hosszú ☺. Mennyi a valószínűsége, hogy egy adott, távoli mezőn elhelyezkedő akadályra rálépünk?
- Megoldás: rekurzióval, zárt alakból adódik a határérték. Azaz: jelölje p_n annak az esélyét, hogy az n-edik mezőre rálépünk. Ekkor igaz, hogy $p_n = p_{n-1} * ½ + (1 - p_{n-1}) * 1$. Ha ugyanis az $(n-1)$-dik mezőre ráléptünk, akkor onnan ½ eséllyel lépünk rá az n-edikre is (ha fejet dobunk), ha viszont az $(n-1)$-dik mezőre nem léptünk rá, akkor az n-edikre biztosan rálépünk, mert két szomszédos mezőt nem tudunk átugrani. Ebből rekurzióval kapjuk, hogy $p_n = 1 - ½ + ¼ - \ldots + (-1)^n (½)^n * p_0$. Felhasználva, hogy $p_0 = 1$, a mértani sorozat összegképlete alapján $p_n = 2/3 * (1 + (-1)^n/2^{n+1})$). Ebből látszik, hogy nagy n esetén a mezőre való rálépés esélye közelítőleg 2/3.
- Ezt a játékot úgy általánosíthatjuk, hogy többféle lehetséges lépéshosszal számolunk, amelyek nem feltétlenül egyformán valószínűek. Ekkor igaz, hogy az adott mezőre lépés valószínűségének határértéke (feltéve hogy lnko(lehetséges lépéshosszak)=1) éppen $1/m$, ahol m az egyes lépések átlagos hossza. (HF: mi a helyzet, ha lnko(lehetséges lépések)>1?) Ez heurisztikusan úgy fogadható el, ha arra gondolunk, hogy n lépés megtétele után közelítőleg az nm-dik mezőig jutottunk, és az addigi mezők közül éppen n darabot érintettünk. Így az érintett mezők aránya $n/nm = 1/m$, azaz egy-egy mezőre ilyen eséllyel léptünk rá. Az, hogy p_n egyre inkább stabilizálódik, onnan látszik, hogy mindegyik p_n néhány előzőnek a (súlyozott) átlaga. Például kockadobásnál $p_n = (p_{n-6} + p_{n-5} + \ldots + p_{n-1})/6$, hiszen azt az eseményt, hogy n-re rálépünk, hatfelé bonthatjuk aszerint, hogy melyik előző mezőről léptünk ide. (Itt $p_0 = 1$, negatív n-re pedig $p_n = 0$.) Ebben az esetben az első húsz mezőre az eredmények a következők:

0.167	0.194	0.227	0.265	0.309	0.360	0.254	0.268	0.280	0.289
0.293	0.291	0.279	0.284	0.286	0.287	0.287	0.286	0.285	0.286

9. A kezdőszámjegy eloszlása

- Feladat: képzeljük el, hogy Magyarország összes települése közül véletlenszerűen kiválasztottunk tizet, és megszámoltuk a lakosságukat. Írjuk le ennek a kísérletnek egy lehetséges végeredményét!

- Játék: képzeljük el, hogy valaki véletlenszerűen választ egy magyarországi települést. Anna arra fogad, hogy lakosságszáma 1–4-ig terjedő számmal kezdődik, Bea arra, hogy 5–9-ig terjedő számmal kezdődik. Kinek előnyös a játék?

- 1881-ben Simon Newcomb egy matematikai folyóiratban írt két oldalas cikkben megjegyezte, hogy a gyakorlatban előforduló adathalmazokban a kezdőszámjegyek megoszlása nem egyenletes, azaz nem ugyannyi adat kezdődik 1-gyel, 2-vel, stb. (American Journal of Mathematics)
  Később 1938-ban Frank Benford fizikus ugyanezt a megfigyelést tette, melyet húsz példán mutatott be (a példák egy kicsit túl jók voltak).

- **Benford törvénye**: A gyakorlatban előforduló legtöbb adathalmazban (népszámlálási adatok, tőzsdei árfolyamok, termelési mutatók, bevételek, stb.) a *k*-val kezdődő adatok aránya körülbelül $\log_{10}(1+1/k)$, illetve a legfeljebb *k*-val kezdődő adatok aránya körülbelül $\log_{10}(1+k)$ ($k = 1,\dots, 9$).

- Illusztráció: Somogy megye 244 településének lakossága az 1991-es népszámlálás szerint.

Kezdőszámjegy	Arány	Benford arány	Kezdőszámjegy	Arány	Benford arány
1	25,0%	30,1%	6	9,0%	6,7%
2	18,9%	17,6%	7	6,1%	5,8%
3	12,7%	12,5%	8	5,7%	5,1%
4	9,8%	9,7%	9	2,0%	4,6%
5	10,7%	7,9%			

- Mi lehet a jelenség magyarázata?

- <u>Magyarázat:</u> A természetben gyakoriak az exponenciálisan növekvő/csökkenő folyamatok, és ezekről megmutatható, hogy ilyen eloszláshoz vezetnek: Tegyük fel, hogy vagyonunk minden évben az *x*-szeresére nő (vagy csökken). Ha egy forintból indulunk ki, akkor az *n*-dik év végén x^n forintunk lesz. Ez akkor kezdődik 1-gyel (az első értékes számjegyet tekintve), ha van olyan egész s szám, hogy

$10^s \le x^n < 2*10^s$, azaz $n*\log_{10} x - \log_{10}2 < s \le n*\log_{10} x$. A kérdés tehát az, hogy ebben a $\log_{10}2$ hosszúságú intervallumban van-e egész szám. Ha pl. $\log_{10} x$ elég kicsi, akkor a „jó” n-ek aránya éppen $\log_{10}2$. (De pl. 10 hatványai mindig 1-gyel kezdődnek!) Ugyanígy számolható ki, hogy az exponenciálisan növő folyamat hány százalékban kezdődik legfeljebb k-val.

- Illusztráció: Legyen $x = 1{,}1$ (pl. a pénzem évente 10%–ot kamatozik). Erre $\log_{10} 1{,}1 = 0{,}04$, és $\log_{10}2 = 0{,}3$. Tehát azokat az n–eket keressük, melyekre a $(0{,}04n - 0{,}3\ ;\ 0{,}04n]$ intervallumban van egész szám.

- Ez a törvényszerűség akkor is érvényben marad, ha a pénzemet átváltom dollárba, hiszen ez csak annyit jelent, hogy az intervallumokat kicsit eltolom: ha a forintban kifejezett összeg f, akkor a dollárban kifejezett összeg $d = yf$, ahol y az árfolyam. Az n. év végén tehát yx^n dollárunk lesz, ez akkor kezdődik 1-gyel, ha van olyan egész s szám, hogy
$n*\log_{10} x - \log_{10}2 + \log_{10} y < s \le n*\log_{10} x + \log_{10} y$

- Megmutatható, hogy a kezdőszámjegyeknek ez az egyetlen olyan lehetséges eloszlása, ami a mértékegység megváltoztatása után is érvényben marad.

- Megmutatható, hogy a Benford törvény diktálta egyenetlenség a második számjegy eloszlásában már sokkal kevésbé jelenik meg, és ahogy egyre későbbi jegyek felé haladunk, az eloszlás egyre egyenletesebbé válik.

- <u>Hasonló eredményhez vezető valószínűségszámítási kérdés:</u> Mekkora a valószínűsége, hogy egy véletlenszerűen választott természetes szám 1-gyel kezdődik? A kérdés feltételezi, hogy az összes természetes szám közül lehetséges véletlenszerűen választani, ez azonban nincs így (mivel végtelen sok természetes szám van).

- Kérdezhetjük helyette a következőt: tf. hogy 1 és n között választunk egy véletlen természetes számot (mindegyiket $1/n$ eséllyel), jelölje p_n ezek között az 1-gyel kezdődőek arányát. Ha létezik a p_n sorozat p határértéke (amint n végtelenbe tart), akkor természetesnek tűnik azt mondani, hogy „egy véletlen természetes szám első számjegye p valószínűséggel 1-es”. Ilyen értelemben ½ az esélye, hogy egy véletlen természetes szám páros, vagy $6/\pi^2$ az esélye, hogy egy véletlen törtet nem lehet egyszerűsíteni. Azonban a p_n sorozat nem konvergál: egyre szélesebb „hullámokat” vet, ahol a hullámhegyek magassága 0,55, a hullámvölgyek mélysége pedig 0,11. Ha azonban ebből „átlagot” számolunk, akkor megint csak kb 30%-ot kapunk.

- Konkrétan: nyilván olyan n–ekre lesz p_n a lehető legnagyobb, melyekre $n = 2*10^s - 1$ (pl. 1, 19, 199, stb), ezekre $p_n = (10^{s+1} - 1) / (9*(2*10^s - 1)) \approx 5/9 = 0{,}555$. Ezután n növekedésével p_n számlálója változatlan, míg nevezője növekszik, egészen $n = 10^{s+1} - 1$ –ig (pl. 9, 99, 999), amikor $p_n = 1/9 = 0{,}111$. Innentől p_n számlálója és nevezője is egyesével növekszik a következő csúcsig.
- Játék: A játékvezető elrejtett két szomszédos, véletlen természetes számot két borítékba. A játékos véletlenszerűen választ egy borítékot. Ha a nagyobb szám van nála, akkor megnyeri a kisebb számot, ha viszont a kisebb szám van nála, akkor ennyit kell fizetnie a játékvezetőnek. A játékos számára nyereséges, veszteséges, vagy igazságos a játék?
- Egyrészt igazságos, hiszen ha a két borítékban az n és az n+1 számok vannak, akkor ½ eséllyel nyer n Ft-ot, ½ eséllyel pedig veszít n Ft-ot.
  Másrészt, tegyük fel, hogy a játékos a borítékban n-et talált. Ekkor a másik borítékban vagy n–1 van (ha a játékvezető az n–1, n számokat rejtette el), vagy n+1 (ha a játékvezető az n, n+1 számokat rejtette el). Mivel véletlenszerűen történt az elrejtés, ennek a két lehetőségnek egyforma az esélye. Tehát a játékos ½ eséllyel nyer n–1 Ft-ot, ½ eséllyel veszít n Ft-ot, azaz veszteséges számára a játék.
- Nyilván egy játék nem lehet egyszerre igazságos és veszteséges is. Az ellentmondás abból adódik, hogy feltettük, hogy lehet az összes természetes szám közül teljesen véletlenszerűen két szomszédosat kiválasztani.

10. Adatsorok, illetve valószínűségi változók jellemzése

- Korábban definiáltuk a valószínűségi változót, és megnéztük, hogy hogyan lehet leírni viselkedését a lehetséges értékek, illetve valószínűségeik felsorolásával:
  X lehetséges értékei: $x_1, x_2, x_3,\ldots$
  a hozzájuk tartozó valószínűségek: $p_1, p_2, p_3,\ldots$

- Ha egy adott valószínűségi változó értékét sokszor megfigyeljük, egymástól függetlenül, akkor egy adatsort kapunk. Pl:
  X: egy kockadobás eredménye
  X lehetséges értékei: 1, 2, 3, 4, 5, 6
  a hozzájuk tartozó valószínűségek: 1/6, 1/6, 1/6, 1/6, 1/6, 1/6
  12 megfigyelést végzünk, azaz 12-szer feldobunk egy kockát. Az adatsor pl.
  2, 5, 5, 1, 2, 3, 4, 3, 6, 2, 1, 1
  A "nagy számok törvénye" szerint az adatsorban az egyes értékek a valószínűségekhez közeli arányban fordulnak elő.
- Ez alapján a (hosszú) adatsorokat nagy eséllyel helyesen tudjuk összepárosítani a véletlen kísérletekkel. Tekintsük pl. a következő 4 kísérlettel előállított valószínűségi változót:
  a) Egy szabályos érmét 10-szer feldobunk. Jelölje X a fejek számát.
  b) Egy urnában 12 piros és 12 fehér golyó van. 10-et kihúzunk (visszatevés nélkül). Jelölje Y a kihúzottak között a pirosak számát.
  c) Egy urnában 20 piros és 30 fehér golyó van. 10-et kihúzunk (visszatevés nélkül). Jelölje Z a kihúzottak között a pirosak számát.
  d) Egy kalapba 11 cédulát teszünk, 0-tól 10-ig számozva. Véletlenszerűen húzunk egyet, jelölje ezt V.
- Tekintsük a következő négy, 30 elemű adatsort (nagyság szerint rendezve)!
  A: 0 0 0 1 1 2 2 2 3 4 4 5 5 5 6 6 6 6 6 6 7 7 8 8 8 8 8 9 10 10
  B: 2 2 3 3 3 3 4 4 4 4 5 5 5 5 5 5 5 5 5 6 6 6 6 7 7 7 7 8 8 9
  C: 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 6 6 7
  D: 2 2 3 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 7 7 7 7
- Hogyan tudnánk őket összepárosítani? (A–d, B–a, C–c, D–b)
  Értékkészlet: mindegyik kísérlet eredménye 0 és 10 között lehet.
  Várható érték: X, Y, és V várható értéke 5, Z–é ennél kevesebb.
  Szóródás: X, Y, és V közül nyilván V szóródik legjobban, legkevésbé pedig Y.

<u>Adatok jellemző értékei</u>

1. *Középértékek:*

- Számtani közép, azaz átlag: Legyen $y = (y_1, \ldots, y_n)$ egy adatsor. Jelölje az átlagát $\bar{y} = \frac{1}{n}\sum y_i$. Ha minden adatból kivonjuk az átlagot, akkor az így kapott adatsor átlaga már nulla. Tetszőleges *a* és *b* számokra jelölje $ay + b$ azt az adatsort, melynek *i*. eleme $ay_i + b$. Tetszőleges *y* és *z* azonos hosszúságú adatsorokra jelölje $y + z$ azt az adatsort, melynek *i*. eleme $y_i + z_i$. Ekkor $\overline{ay+b} = a\bar{y} + b$, $\overline{y+z} = \bar{y} + \bar{z}$
- Medián, azaz a nagyság szerint középső elem (vagy a két középső átlaga), jelölje ezt $m(y)$. Erre is igaz, hogy $m(ay + b) = am(y) + b$, de $m(y + z) = m(y) + m(z)$ általában nem teljesül.

A két középérték közül az átlag érzékeny a kiugróan nagy vagy kicsi értékekre, ellenben a medián nem.

2. *Szóródási mérőszámok:*

Az adatok szóródását az méri, hogy átlagosan milyen messze esnek a középértéktől. A használt "távolság" több féle is lehet.

- Hagyományos távolságfogalom, azaz két szám különbségének abszolút értéke. Megmutatható, hogy az adatok egy adott konstanstól vett átlagos távolságát a medián minimalizálja, azaz $e_c(y) = \frac{1}{n}\sum |y_i - c|$ akkor a legkisebb, ha $c = m(y)$. Az $e(y) = e_{m(y)}(y)$ érték tehát az adatok szóródásának egy fajta mérőszáma lehet. Teljesül, hogy $e(ay + b) = |a|e(y)$, azaz ha pl. az adatpontokat vízszintes irányban kétszeresen megnyújtjuk, akkor szóródásuk megduplázódik.
- A fenti állítás bizonyítása: tegyük fel, hogy *y* nagyság szerint növekvő sorrendbe van rakva. Válasszuk ki először a legkisebb és a legnagyobb adatot, ezek y_1 és y_n. Az $|y_1-c|+|y_n-c|$ érték akkor minimális, ha *c* az $[y_1, y_n]$ intervallumba esik. Haladjunk ezután az adatsoron „mindkét irányból befelé," látszik, hogy olyan *c* minimalizálja az eltérések összegét, amlely mindegyik $[y_k, y_{n+1-k}]$ intervallumba beleesik.
- A négyzetes távolság kevésbé tűnhet természetesnek, viszont könnyebb vele számolni, és az a jó tulajdonsága is megvan, hogy az átlagos

négyzetes távolságot az átlag minimalizálja: $d_c^2(y) = \frac{1}{n}\sum(y_i - c)^2$ akkor a legkisebb, ha $c = \overline{y}$. A hozzá tartozó $d^2(y) = d_{\overline{y}}^2(y)$ mennyiséget az adatok szórásnégyzetének, a $d(y)$ értéket pedig az adatok szórásának nevezzük. Teljesül, hogy $d(ay + b) = |a|d(y)$. A szórás kiszámítását gyakran megkönnyíti a következő képlet: $d^2(y) = d_c^2(y) - (c - \overline{y})^2$.

- Ezek a mennyiségek analóg módon a valószínűségi változókra is értelmezhetők. Az átlagnak a várható érték felel meg, az adatok szórásnégyzetének a valószínűségi változó szórásnégyzete: $D^2(X) = E[(X - E(X))^2]$.

11. Néhány nevezetes modell

- Mondjunk minél több, véletlen kísérlethez tartozó valószínűségi változót! Próbáljuk meg valahogy csoportosítani ezeket! Vannak-e közöttük hasonlóak?
- Rögzített számú *független* kísérletből megszámoljuk a sikeres kísérletek számát → binomiális eloszlás. Példák:
  - ✓ 10 kockadobásból hány 6-ost kapunk?
  - ✓ 20 kosárradobásból hány talál be?
  - ✓ 25 hallgató közül hányan születtek áprilisban?
  - ✓ Egy ötgyermekes családban hány fiú van?
- Azt a véletlen mennyiséget, amely a 0, 1, …, n értékeket veheti fel, méghozzá a k értéket $\binom{n}{k}p^k(1-p)^{n-k}$ eséllyel, n rendű, p paraméterű binomiális eloszlásúnak nevezzük. Egy ilyen eloszlású mennyiség várható értéke np, szórásnégyzete $np(1 - p)$. Pl. 36 érmedobásból átlagosan 18 fej lesz, a szórás 3.
- A binomiális eloszlást (p = ½-re) jól lehet szemléltetni a Galton-deszka elnevezésű eszközzel (http://www.stattucino.com/berrie/dsl/Galton.html). Azt, hogy a Galton-deszka egyes szintjein hogyan oszlanak meg a golyók,

illetve az egyes szögekhez hányféle úton lehet eljutni, a Pascal-háromszög írja le.

- Más eloszlást kapunk, ha a kísérletek nem függetlenek egymástól. Erre tipikus példa, ha urnából *visszatevés nélkül* húzunk golyókat → hipergeometrikus eloszlás.
- Számoljuk ki, hogy ha egy héten 22 millió, egymástó függetlenül és véletlenszerűen kitöltött szelvény játszik a lottón, akkor mennyi az esélye, hogy legalább két öttalálatos lesz? Binomiális eloszlással lehet számolni, de nagyon nagy számok szerepelnek.
- Ilyen esetben a binomiális eloszlást jól közelíti a Poisson eloszlás, amelyet 1837-ben Siméon Denis Poisson fedezett fel. Azt a véletlen mennyiséget, amely a 0, 1, 2, … értékeket veheti fel, méghozzá a k értéket $e^{-s}s^k/k!$ eséllyel, s paraméterű Poisson eloszlásúnak nevezzük. Ennek alapján a fenti kérdésre a válasz: kb. 9%.
- A Poisson eloszlás átlagos értéke és szórásnégyzete is s. Legvalószínűbb értéke az s egészrésze. Ilyen eloszlással modellezhető pl. a sajtóhibák száma egy szövegben, vagy egy ritka betegségben az évi halálozások száma.
- További példa: megnézzük, hogy hányadik kísérletre közvetkezett be az első siker → geometriai eloszlás.

12. Egyenletes eloszlás egy halmazon

- Feladat: Egy négyzetbe rajzoljunk be 20 “véletlen” pontot.
- Mit jelent itt a “véletlen”?

1. próbálkozás: minden egyes pontnak ugyanakkora a valószínűsége → igen ám, de mivel végtelen sok pont van, ez a közös valószínűség csak 0 lehet. Például annak a valószínűsége, hogy pont a négyzet középpontját választjuk ki, 0. Ugyanakkor ez nem lehetetlen. Ez paradox, de nem ellentmondás. Ezzel azonban még csak az egyes pontok valószínűségét definiáltuk, hogyan lehet halmazokra kiterjeszteni ezt?

2. próbálkozás: mekkora az esélye, hogy a véletlen pont a négyzet bal alsó negyedébe esik? A bal alsó negyed pontjainak valószínűségét kellene összeadni, de ez kontinuum sok nulla (ez különbözteti meg a természetes számoktól: ott láttuk, hogy nem lehet természetes számot “véletlenszerűen” választani, mert megszámlálhatóan végtelen sok nulla összege nulla. Viszont kontinuum sok nulla összege lehet pozitív)! Viszont nyilvánvaló, hogy ezt a valószínűséget ¼-nek szeretnénk definiálni. Ebből jön: a valószínűség legyen a területtel arányos!

- Az X véletlen pont egyenletes eloszlású az A (valahány dimenziós) halmazban, ha minden B részhalmaz esetén P(X B-be esik) = ter(B)/ter(A). Ez a definíció teljesíti a valószínűség axiómáit.
- Példa: 1 méter sugarú céltáblára lövünk messziről, tegyük fel, hogy eltaláljuk.
  Mekkora az esélye, hogy pont a közepét találjuk el? → 0
  Mekkora az esélye, hogy a lövedék a középponttól pont fél méterre csapódik be? → 0
  Mekkora az esélye, hogy a lövedék a középponthoz fél méternél közelebb csapódik be? → ¼
- Ilyen eloszlással modellezhető pl. a buszra való várakozási idő (1 dimenzió), a csillagok elhelyezkedése az égbolton (2 dimenzió), a mazsolák elhelyezkedése a kalácsban (3 dimenzió).
- Feladat: egy 1 méter hosszú botnak véletlenszerűen kiválasztjuk két pontját, majd ezekben a pontokban eltörjük. Adjuk meg, hogy milyen eloszlású lesz a 25 cm-nél rövidebb darabok száma!

- Megoldás: Jelölje a két töréspontnak a bot bal végpontjától vett távolságát X és Y. Ez két véletlen szám 0 és 1 között. Az (X, Y) kéttdimenziós pontnak tehát mindkét koordinátája véletlenszerű 0 és 1 között, és a két koordináta független. Ez azt jelenti, hogy (X, Y) egyenletes eloszlású az egységnégyzetben. Tehát azt kell megvizsgálni, hogy a négyzet mely pontjai adnak 0, 1, ill. 2 25cm-nél rövidebb darabot. A végeredmény: 1/16 eséllyel 0, 6/16 eséllyel 1, 9/16 eséllyel 2 rövid darab keletkezik.

- <u>Buffon-féle tűprobléma</u>. (Georges Buffon, 1733). Egy szoba padlóján a padlódeszkák egymással párhuzamosan futnak, szélességük egy egység.

Leejtünk egy szintén egy egység hosszú tűt. Mekkora a valószínűsége, hogy a tű metszi valamelyik két padlódeszka csatlakozási vonalát?

- Megoldás: a leesett tű helyzetét két adattal határozhatjuk meg: középpontjának a legközelebbi egyenestől vett X távolságával, és az egyenesek irányával bezárt Γ szögével. A véletlenszerű leejtés azt jelenti, hogy az (X, Γ) pár egyenletes eloszlású a $[0,1/2] \times [0, \pi/2]$ téglalapon. A tű pontosan akkor metszi a hozzá legközelebb eső egyenest, ha $X \leq (\sin \Gamma)/2$. Tehát a téglalapon belül az ezen pontok által meghatározott tartomány területét kell kiszámítani, ami $\int_0^{\pi/2} \frac{1}{2}\sin x\, dx = \frac{1}{2}$. Tehát a valószínűség: $(1/2)/(\pi/4) = 2/\pi \approx 0{,}637$. Ennek alapján, ha n kísérletből k esetben volt metszés, akkor π értékét közelíthetjük $2n/k$-val. (http://www.mste.uiuc.edu/reese/buffon/buffon.html)
- A vetület problémája. Vizsgáljuk meg, hogy átlagosan mekkora lesz a tűnek a deszkák irányára vett vetülete! A tű vetületének hossza $\cos \Gamma$, tehát ennek az átlagos értékét kell meghatározni, amint Γ 0 és $\pi/2$ között fut. Az átlagos érték $2/\pi$, hiszen a $[0, \pi/2] \times [0, 2/\pi]$ téglalapnak pont akkora a területe, mint a $\cos x$ függvény alatti terület, éppen 1.
  Ebből általánosítással kaphatjuk, hogy tetszőleges K kerületű konvex zárt görbét ledobva, vetülete átlagosan K/π hosszú lesz: egy a illetve b oldalhosszúságokkal rendelkező téglalapot (pl. radírt) véletlenszerűen leejtve, annak vetülete egyenlő lesz az a ill. b oldalak vetületeinek összegével. Az összeg átlagos értéke az átlagos értékek összege, így a vetület átlagosan $2a/\pi + 2b/\pi = K/\pi$. Ezt konvex sokszögre is elmondhatjuk: ha az összes oldal vetületét összeadjuk, akkor a sokszög vetületét kétszeresen kapjuk meg. Sokszögről határátmenettel tetszőleges görbére áttérhetünk.
- Bertrand paradoxon (Joseph Louis Bertrand, 1889). Egy körnek válasszuk ki véletlenszerűen egy húrját! Mekkora a valószínűsége, hogy a húr hosszabb lesz, mint a körbe írható szabályos háromszög oldala?
- Ez azért paradoxon, mert a “véletlen húrt” többféleképpen is definiálhatjuk, és a különböző definíciók más-más eredményt adnak. Például:

a) a körben véletlenszerűen választjuk a húr középpontját → ¼

b) a körvonalon véletlenszerűen választjuk a húr két végpontját → 1/3

c) véletlenszerűen választunk egy irányt, majd az ilyen irányú húrok közül választunk véletlenszerűen → ½

d) a körben véletlenszerűen választunk egy pontot, majd egy arra illeszkedő irányt → 0.609

e) a körben két véletlen pontot választunk → 0.745.

- A fenti kérdés azzal ekvivalens, hogy a húr metszi-e a fele akkora sugarú kisebb kört. Általánosításként belátható, hogy ha adott egy K konvex zárt görbe, és annak belsejében egy k másik konvex zárt görbe, akkor K egy véletlen húrja ker(k)/ker(K) valószínűséggel metszi a belső görbét, ha a véletlenszerű választás azt jelenti, hogy először egy véletlen irányt választunk, majd az adott irányú húrok közül választunk véletlenszerűen.

<u>Előadáson nem hangzott el, csak érdekességképp:</u>
<u>Monte-Carlo módszer</u>:tegyük fel, hogy valamilyen véletlen kísérlettel kapcsolatos valószínűségben vagy várható értékben fellép egy ismeretlen mennyiség. Ekkor az ismeretlen mennyiséget közelíthetjük úgy, hogy a kísérletet sokszor elvégezzük, illetve véletlen számsorozatok segítségével szimuláljuk (ebből ered a kissé félrevezető "Monte-Carlo módszer" elnevezés, hiszen pl. egy játékkaszinó ruletteredményei felhasználhatók lennének ilyen szimulációkhoz). A fenti Buffon probléma is egy Monte-Carlo módszert ír le a π meghatározására (ennek azonban csak elméleti jelentősége van, hiszen a π-t egymillió tizedesjegy pontossággal más módszerekkel is meg lehet határozni).
Más területeken azonban a módszer nagyon hasznosnak bizonyult. Egy egyszerű példa: egy érmét 1000-szer feldobva azt látjuk, hogy 600-szor mutat fejet. Vajon tekinthetjük-e az érmét szabályosnak? Még ha nem is tanultunk semmi valószínűségszámítást, szimulációval meghatározhatjuk, hogy egy szabályos érménél mekkora az esélye, hogy 1000 dobásból vagy a fej vagy az írás legalább 600-szor jön ki. Ha ennek esélye pl. csak 1 ezrelék, akkor ésszerű feltételezni, hogy ez az egyenlőtlenség nem írható a véletlen számlájára, hanem az érme nem szabályos.
Fontos alkalmazási terület még a numerikus integrálás (akár sok dimenzióban): kíváncsiak vagyunk pl. a (0,1) intervallumon az $f(x)$ nemnegatív függvény alatti területre. Ha tudjuk, hogy ezen az intervallumon az $f(x)$ maximuma legfeljebb C, akkor véletlenszerűen választva egy (0,1)-beli X-et és egy (0,C)-beli Y-t, ellenőrizhetjük, hogy $Y < f(X)$ teljesül-e. Ha n kísérletből ez k-szor teljesül, akkor a területet közelíthetjük a kC/n képlettel.
Ehhez természetesen kell, hogy a mai számítógépek nagyon jó véletlenszám-generátorokkal rendelkeznek, melyek ugyan csak pszeudovéletlen számokat állítanak elő, de a kapott sorozatok ugyanolyan tulajdonságokkal rendelkeznek, mint a valódi véletlen (pl. Mersenne twister, 1997).

13. A szimmetrikus bolyongás

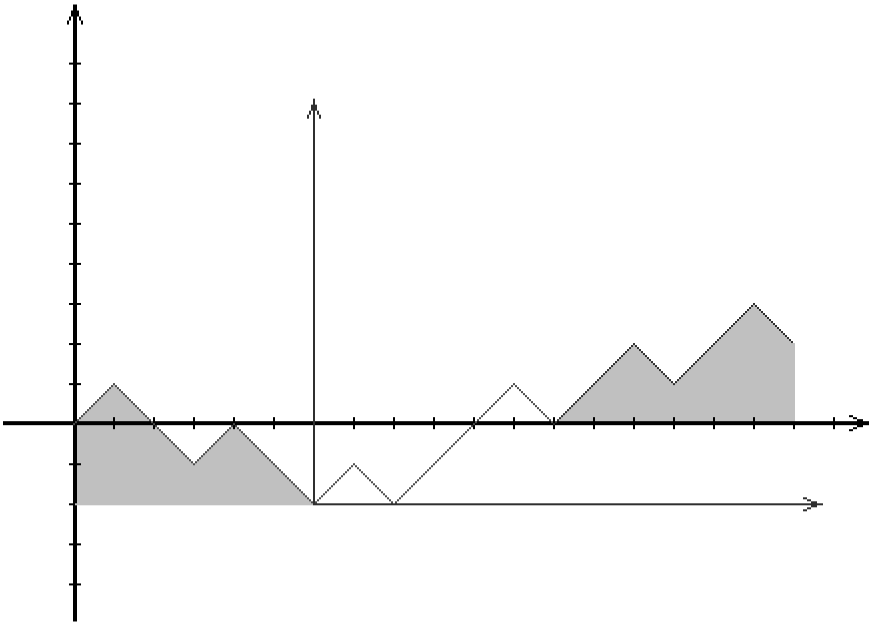
- **Az egyszerű szimmetrikus bolyongás 1 dimenzióban:** Tegyük fel, hogy egy részecske a (függőlegesen elhelyezett) valós számegyenes origójából kiindulva időegységenként lép egy egységnyit úgy, hogy minden alkalommal az előző lépésektől függetlenül, véletlenszerűen választ, hogy felfelé vagy lefelé lépjen.
  Jelölje a *k*. lépést X_k, ekkor $P(X_k = 1) = P(X_k = -1) = ½$.
- n lépés után a részecske az $S_n = X_1 + X_2 + \ldots + X_n$ helyen tartózkodik. Ha egy kétdimenziós koordinátarendszerben ábrázoljuk az (n, S_n) pontokat, képet kapunk a részecske által megtett útról.
- **Általánosítási lehetőségek:** a bolyongás
  – Nem egyszerű, ha a lépéshosszak egytől különböző értékeket is felvehetnek.
  – Nem szimmetrikus, ha a pozitív illetve negatív irányú lépések nem egyforma valószínűségűek. (Lásd "Cliff hanger," http://www.mste.uiuc.edu/activity/cliff/)
  – Több dimenziós, pl. m dimenzióban a részecske a $2m$ darab szomszédos rácspont valamelyikére lép.
- **Alkalmazások (pl):**

1. Modellezhetjük egyszerű szimmetrikus bolyongással egy szavazatszámlálás folyamatát (ilyenkor az út végpontja rögzített).
2. Növények magasságának növelésére irányuló kezelésnél, ha a kezelés hatástalan, akkor a kezelt illetve a kezeletlen növénycsoport egyedei véletlenszerűen követik egymást a nagyság szerinti sorrendben (az út végpontja most is rögzített).
3. Egy szabályos érmét dobálva, a dobások folyamatát szimmetrikus bolyongással írhatjuk le.

- **Alapok:** n hosszúságú útból 2^n darab van, ezek mindegyike egyformán valószínű. Az (n, a) pontba vezető utak száma $N_{n,a} = \binom{n}{p}$, ahol $p = (n+a)/2$ a felfelé tett lépések száma. (Vegyük észre, hogy ha n és a különböző paritású, akkor $N_{n,a} = 0$.)

- **A tükrözési elv:** Legyen $A = (m, a)$, $B = (n, b)$, $a, b > 0$, $n > m \geq 0$. Legyen A tükörképe az x-tengelyre $A' = (m, –a)$. Ekkor az A-ból B-be vezető, x-tengelyt érintő utak száma megegyezik az összes A'-ből B-be vezető út számával.
- Bizonyítás: A kétféle út között kölcsönösen egyértelmű megfeleltetés létesíthető úgy, hogy az x-tengely első érintéséig terjedő útszakaszt tükrözzük.
- **Következmény** (szavazási lemma): Tegyük fel, hogy egy választáson két jelölt indult, A és B. A összesen p szavazatot kapott, B pedig $q$szavazatot, és $p > q$. Ekkor annak a valószínűsége, hogy a szavazatszámlálás folyamán végig az A jelölt vezet (szavazategyenlőség sem megengedett), $(p – q)/(p + q)$.
- **Origóba való visszatérések:** legyen $u_{2n} = P(S_{2n} = 0) = N_{2n,0}/2^n$. ($u_0 = 1$)
  A Stirling formula alapján nagy n-re $u_{2n} \sim 1/\sqrt{\pi n}$. ($n! \approx \sqrt{2\pi} n^{n+0.5} e^{-n}$)
  Pl. $u_{20} \approx 0{,}18$, $u_{100} \approx 0{,}08$, $u_{500} \approx 0{,}035$, $u_{2000} \approx 0{,}018$.

1. Lemma: $P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0) = u_{2n}$.
   Bizonyítás: kölcsönösen egyértelmű megfeleltetéssel. Ha van egy olyan utunk, amely $2n$ lépés alatt az origóba tér vissza, abból az első minimumpontig vezető szakaszt vágjuk le, tükrözzük a függőleges tengelyre, és ragasszuk az út végéhez. Az origót a kapott út kezdőpontjához tolva, egy nemnegatív utat kapunk. Fordítva, ha van egy nemnegatív utunk, amelynek végpontja a $(2n, 2a)$ pont, akkor keressük meg az a szint utolsó elérését, az út ezutáni szakaszát vágjuk le, tükrözzük a függőleges tengelyre, és ragasszuk az út elejéhez. Az origót a kapott út kezdőpontjához tolva, egy origóba visszatérő utat kapunk.



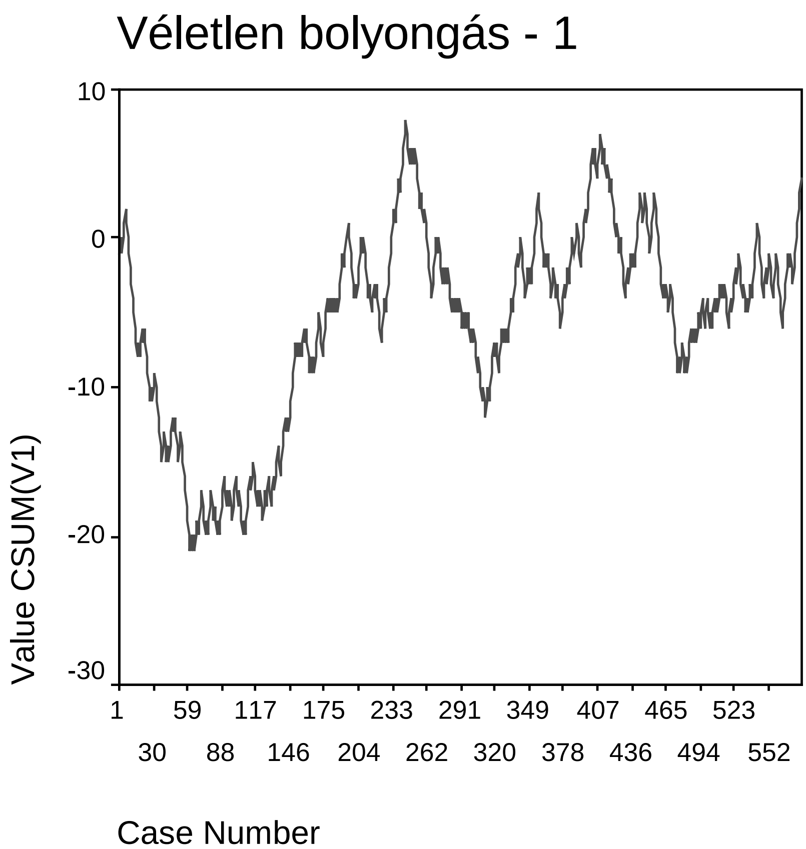
2. Lemma: $P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n} \neq 0) = u_{2n}$.

   Bizonyítás: $P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n} \neq 0) = 2P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0)$. Viszont $P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0) = 0.5\ P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n\text{-}1} \geq 0)$, hiszen az első lépés felfelé kell történjen, majd az origót az (1,1) pontba tolva, egy $2n - 1$ hosszú, nemnegatív útnak kell következnie. De $P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n\text{-}1} \geq 0) = P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n} \geq 0) = u_{2n}$, hiszen $2n - 1$ lépés alatt páratlan számhoz érünk, tehát $S_{2n\text{-}1} \geq 1$, így a következő lépés nem tud negatívba menni.

   - Példa: egy szabályos érmét 100-szor feldobunk. A következő három eseménynek egyaránt kb. 8% a valószínűsége:
     – a 100 dobásból pont 50 fej, 50 írás
     – a dobássorozatban végig legalább annyi fejdobás van, mint írás
     – a dobássorozatban egyszer sem fordul elő, hogy pont ugyanannyi fej legyen, mint írás
   - Következmény: jelölje f_{2n} annak valószínűségét, hogy a bolyongás először a $2n$. lépésben tér vissza az origóba, azaz $f_{2n} = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} = 0)$. Ekkor $f_{2n} = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-2} \neq 0) - P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n} \neq 0) = u_{2n-2} - u_{2n} = u_{2n}/(2n - 1)$.($f_0 = 0$) Ebből a felírásból látszik, hogy $f_2 + f_4 + f_6 + \ldots = 1$, azaz a bolyongás

előbb-utóbb visszatér az origóba (1 valószínűséggel). u_{2n} aszimptotikájából adódik azonban, hogy az első visszatéréshez szükséges lépések számának átlagos értéke minden előírt értéknél nagyobb, azaz végtelen: $f_{2n} \approx 1/(2n\sqrt{\pi n})$

- Egy (szimulált) példa, a lépések száma 580. A bolyongás 28-szor járt az origóban, 466 lépést töltött a negatív oldalon, és 86-ot a pozitív oldalon.



- Vizsgáljuk most azt a kérdést, hogy egy $2n$ hosszú bolyongás mikor járt utoljára az origóban!

Állítás: $\alpha_{2k,2n}$ **:=** $\mathrm{P}(S_{2k} = 0,\ S_{2k+1} \neq 0,\ S_{2k+2} \neq 0,\ \ldots,\ S_{2n} \neq 0) = u_{2k}\, u_{2n-2k}$

A bizonyítás az előző alkalom 2. lemmájából adódik, mivel

$P(S_{2k} = 0,\ S_{2k+1} \neq 0,\ S_{2k+2} \neq 0, \ldots, S_{2n} \neq 0) = P(S_{2k} = 0) * P(S_1 \neq 0,\ S_2 \neq 0, \ldots, S_{2n-2k} \neq 0)$. u_{2n} aszimptotikájából kapjuk, hogy $\alpha_{2k,2n} \sim \dfrac{1}{\pi\sqrt{k(n-k)}} = f(k/n)/n$, ahol

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \quad (0 < x < 1).$$

Vegyük észre, hogy $\alpha_{2k,2n} = \alpha_{2n-2k,2n}$, azaz pl. egy 1000 hosszú útnál ugyanolyan valószínű, hogy a 100. lépésben járt utoljára az origóban, mint hogy a 900.-ban.

Az $f(x)$ függvény integrálját ismerve ($\frac{2}{\pi}\arcsin\sqrt{x}$) felírható, hogy közelítőleg mekkora a p_y valószínűsége annak, hogy a bolyongás a teljes idő utolsó y százalékában nem járt az origóban:

y%	p_y
97%	10%
90%	20%
80%	30%
65%	40%
50%	50%

- Következőnek azt vizsgáljuk meg, hogy a $2n$ lépésből hány történt a pozitív oldalon!

Állítás: P(a $2n$ lépésből $2k$ történt a pozitív oldalon) = $\alpha_{2k,2n}$

Ezt az eredményt nem bizonyítjuk. Állapítsuk meg összegzésként, hogy a szimmetrikus bolyongás nem az intuíciónknak megfelelően viselkedik: ritkán tér vissza az origóba, és nem közel azonos számú lépést tesz a pozitív illetve a negatív oldalon.

- Összehasonlításul azonban álljon itt az az

Állítás: Feltéve, hogy az út végpontját az origóban rögzítjük, azaz feltesszük, hogy $S_{2n} = 0$, P(a $2n$ lépésből $2k$ történt a pozitív oldalon) = $1/(n + 1)$ ($k = 0, ..., n$).

Ehhez kapcsolódik a növények magasságának növelésére irányuló kezelés hatékonyságának vizsgálata. Legyen r darab kezelt növény $a_1 > a_2 > \ldots > a_r$ magasságokkal, és r darab kezeletlen növény $b_1 > b_2 > \ldots > b_r$ magasságokkal. Ha a $2r$ darab növény magasságát csökkenő sorrendbe állítjuk, majd az a-k és b-k váltakozását egy úttal ábrázoljuk (minden a-nál felfelé, minden b-nél lefelé lépünk), akkor hatástalan kezelés esetén egy egyszerű szimmetrikus bolyongás képét kell kapnunk. Nézzük meg, hogy hány olyan k index van, melyre $a_k > b_k$! Éppen

feleannyi, mint ahány lépést az út a pozitív oldalon tett. Legutóbbi állításunk szerint tehát minden lehetőség egyformán valószínű.
1876-ban Francis Galton Charles Darwintól kapott adatokat vizsgált, ahol $r = 15$ volt, és 13-szor voltak az a-k elől. Galton ez alapján hatásosnak minősítette a kezelést, pedig egy teljesen hatástalan kezelésnél is 3/16 = 18,75% az esélye, hogy az a-k legalább 13-szor vezetnek.

- **A tönkremenési probléma**. Van két játékos, az egyiknek 100, a másiknak 200 forintja. Fej vagy írás játékot játszanak, minden menetben a vesztes 1 forintot fizet a nyertesnek. Ezt addig folytatják, míg valamelyiküknek elfogy a pénze. Mik az egyes játékosok tönkremenési valószínűségei?
- A feladatot a szimmetrikus bolyongás segítségével lehet megoldani. Jelölje a két játékos össztőkéjét M, ez a játék folyamán változatlan. Legyen az első játékosnak k forintja, és jelölje p_k annak valószínűségét, hogy ő megy előbb tönkre. Ekkor
  $p_k = (p_{k+1} + p_{k-1})/2$ ($k = 1, 2, ..., M–1$), és természetesen $p_0 = 1$, $p_M = 0$. Ennek az egyenletrendszernek a $p_k = 1 – k/M$ lineáris függvény a megoldása.

- **Alkalmazás:** számítsuk ki, hogy a bolyongás az origóba való két visszatérés között átlagosan hányszor jár az a pontban!
  Annak valószínűsége, hogy egyáltalán eljut a-ba (mondjuk $a>0$): először felfelé kell lépni, majd onnan előbb jutni a-ba, mint 0-ba, azaz ½ * 1/a = 1/2a. Ha már eljutottunk a-ba, akkor annak valószínűsége, hogy többször már nem járunk ott az origóba való visszatérés előtt: ½ * 1/a = 1/2a. Minden a-ba való visszatérés után tehát 1/2a az esélye, hogy ez az utolsó látogatásunk az adott szinten az origóba való visszatérés előtt.
  Az átlagos érték tehát: $1/2a + 1/2a(1–1/2a) + 1/2a(1–1/2a)^2 + 1/2a(1–1/2a)^3 + ... = 1$. Meglepő, de ez független attól, hogy melyik szintet vizsgáljuk. Az átlag úgy lehet mindig 1, hogy a távoli szintekre kis valószínűséggel jutunk el, de ha egyszer már eljutottunk, akkor sokáig maradunk ott. Ebből az eredményből is következik, hogy a visszatéréshez szükséges lépések átlagos száma végtelen.

- **Pár szó a többdimenziós bolyongásról:** A kétdimenziós bolyongás két független egydimenziós bolyongással ekvivalens (nézem a bolyongás vetületét a két „átlóra”). Ebből látszik, hogy annak valószínűsége, hogy $2n$ lépésben visszatér a bolyongás az origóba, az egydimenziós valószínűség négyzetével egyezik meg, azaz $1/n$ nagyságrendű.
  Belátható, hogy a bolyongás akkor és csak akkor tér vissza előbb-utóbb biztosan az origóba, ha a visszatérési valószínűségek összege végtelen. A kétdimenziós bolyongás tehát visszatérő, a három- vagy többdimenziós azonban már nem!

14. A Galton-Watson-féle elágazó folyamat

- **Motiváció**: a XIX. században megfigyelték, hogy sok régi, szép családnév fokozatosan eltűnik, kihal. A nevek öröklődését modellező véletlen folyamatról 1874-ben Galton és Watson közölt alapvető cikket.
- **A modell**: Valamilyen egyedek (emberek, sejtek, stb.) egymást követő generációit vizsgáljuk. A 0. generáció egy tagú, ez az ős. Az ős véletlen számú utódot hoz létre, akik az 1. generációt alkotják. Az 1. generáció tagjai, egymástól függetlenül, újabb utódokat hoznak létre, ez lesz a 2. nemzedék. Az n. nemzedék az n – 1. nemzedék utódaiból áll.
  Feltesszük, hogy az utódok számának eloszlása mindig ugyanaz. Jelölje ezt a véletlen mennyiséget X, lehetséges értékei a nemnegatív számok: $k = 0, 1, 2,\ldots$ A lehetséges értékekhez tartozó valószínűségeket jelölje $p_k = P(X = k)$. Azt, hogy egy egyednek átlagosan hány utódja lesz, jelölje m, azaz $m = 0*p_0 + 1*p_1 + 2*p_2 + 3*p_3 +\ldots$
  Jelölje még az n. nemzedék tagjainak számát Z_n.
- **Kérdések**:

1. Átlagosan hány tagú az n. nemzedék, azaz mennyi $E(Z_n)$?
2. Mekkora a valószínűsége, hogy $Z_n = 0$, azaz az n. nemzedékre a “család” kihalt?
3. Mekkora a valószínűsége, hogy van olyan n, melyre $Z_n = 0$, azaz a család előbb–utóbb kihal?

4. Átlagosan hányadik nemzedéknél hal ki a család? (Jelölje a kihalás idejét T, azaz T a legkisebb olyan n, melyre $Z_n = 0$.)
    - Ezt a folyamatot a legkényelmesebben a **generátorfüggvény** segítségével lehet vizsgálni. A definíció a következő: A fenti X véletlen mennyiség generátorfüggvénye: $g(z) = z^0 * p_0 + z^1 * p_1 + z^2 * p_2 + z^3 * p_3 + \ldots$, ahol z egynél kisebb abszolút értékű szám (ilyenkor a sor biztosan konvergens, és $|g(z)| < 1$). Vegyük észre, hogy ez éppen a z^X mennyiség átlagos értéke. Könnyű látni, hogy $g(0) = p_0$, $g(1) = 1$. A generátorfüggvényt tagonként deriválhatjuk, azaz $g'(z) = 1 * z^0 * p_1 + 2 * z^1 * p_2 + 3 * z^2 * p_3 + \ldots$, amiből azt kapjuk, hogy $g'(1) = m$.
    - **Válaszok**:

1. A 0. nemzedék mindig 1 tagú. Az első nemzedék átlagosan m tagú. Mivel az 1. nemzedék átlagosan m tagú, és minden tagnak átlagosan m utóda lesz, a 2. nemzedék átlagosan m^2 tagú (ezt precízen is be lehet bizonyítani). A gondolatmenetet folytatva adódik, hogy az n. nemzedék átlagosan m^n tagú.
2. Jelölje az n. nemzedék tagszámának, Z_n–nek generátorfüggvényét $g_n(z)$. A fentiek szerint $P(Z_n = 0) = g_n(0) =: q_n$. Világos, hogy $g_0(z) = z$, $g_1(z) = g(z)$. Belátható, hogy általában $g_n(z) = g(g_{n-1}(z))$. Azaz $q_n = g(q_{n-1})$.
   A bizonyítás vázlata: indukcióval, $n = 1$-re már tudjuk az állítást. Az n. nemzedék az 1. nemzedék n–1-edfokú utódaiból áll. Ha az 1. nemzedék k tagú, akkor az n. nemzedék létszáma: $Z_n = Z^1{}_{n-1} + Z^2{}_{n-1} + \ldots + Z^k{}_{n-1}$, ahol az összeadandók függetlenek, és ugyanolyan eloszlásúak, mint Z_{n-1}. Ezt felemelve a z alapra, majd várható értéket véve – felhasználjuk, hogy függetlenek szorzatának a várható értéke megegyezik a várható értékek szorzatával – kapjuk, hogy ilyenkor $E(z^{Z_n}) = g^k_{n-1}(z)$. Az eddig leírt eseménynek a valószínűsége p_k. Ezekkel a valószínűségekkel súlyozva és összegezve, megkapjuk, hogy $g_n(z) = g(g_{n-1}(z))$.
3. Ha $Z_n = 0$, akkor természetesen minden $l > n$ esetén $Z_l = 0$ is igaz. Ezek az események tehát egyre bővebbek, és egyesítésük azt az eseményt adja ki, hogy a család előbb–utóbb kihal. Ha tehát a kérdéses valószínűséget q jelöli, akkor $q = \lim q_n$. Ha a $q_n = g(q_{n-1})$ egyenletben n végtelenhez tart, akkor azt kapjuk, hogy q kielégíti a $q = g(q)$ egyenletet. Láttuk, hogy az 1 mindig megoldás. Ha azonban $m > 1$, akkor van 1-nél kisebb pozitív megoldás is, és az adja meg a kihalás

valószínűségét (ábra: a $g(z)$ függvényt felrajzolhatjuk annak alapján, hogy $g(0) = p_0$, $g(1) = 1$, $g'(1) = m$, és a függvény a (0,1) intervallumon monoton növekvő és konvex).

Ha $m < 1$, akkor a folyamat szubkritikus. (1 valószínűséggel kihal)

Ha $m = 1$, akkor a folyamat kritikus. (1 valószínűséggel kihal)

Ha $m > 1$, akkor a folyamat szuperkritikus. (1-nél kisebb valószínűséggel hal ki)

4. Belátható, hogy a szubkritikus esetben $E(T) \leq 1/(1-m)$, a kritikus esetben viszont $E(T)$ végtelen.

- **Példa**: Ha az utódeloszlás binomiális: $X \sim \mathrm{Bin}(l,p)$, akkor $g(z) = (pz + 1 - p)^l$, tehát a kihalás valószínűségét egy l-edfokú egyenlet megoldása adja.