



M Ű E G Y E T E M 1 7 8 2

# Geometriai mintázatok felbukkanása véletlen kavicspopuláció tengelyarányainak fázisterében

BME Szilárdságtani és Tartószerkezeti Tanszék  
BME Morfológia és Geometriai Modellezés Tanszék

Szondi Máté Álmos  
harmadéves építészmérnök hallgató

Konzulensek:  
Dr. Domokos Gábor  
Dr. Sipos András Árpád

MTA-BME Morfodinamikai Kutatócsoport

TUDOMÁNYOS DIÁKKÖRI KONFERENCIA 2021

## Tartalomjegyzék

1	Absztrakt .....	3
2	Bevezetés .....	5
3	Osztályozási rendszerek.....	7
3.1	Krumbein osztályozás.....	7
3.2	Zingg-féle osztályozás.....	7
4	Alapfeltevések.....	9
4.1	Modell 1: Véletlen modell.....	9
4.2	Modell 2: Determinisztikus modell.....	12
4.3	A ponthalmaz tárolása .....	14
5	Vizsgálat eszközei.....	16
6	Adatelemzés .....	18
6.1	Alapdiagrammok .....	18
6.2	Egyenesen lévő pontok számának normálása és átskálázása .....	20
7	Összegzés és további lehetőségek.....	24

## 1 Absztrakt

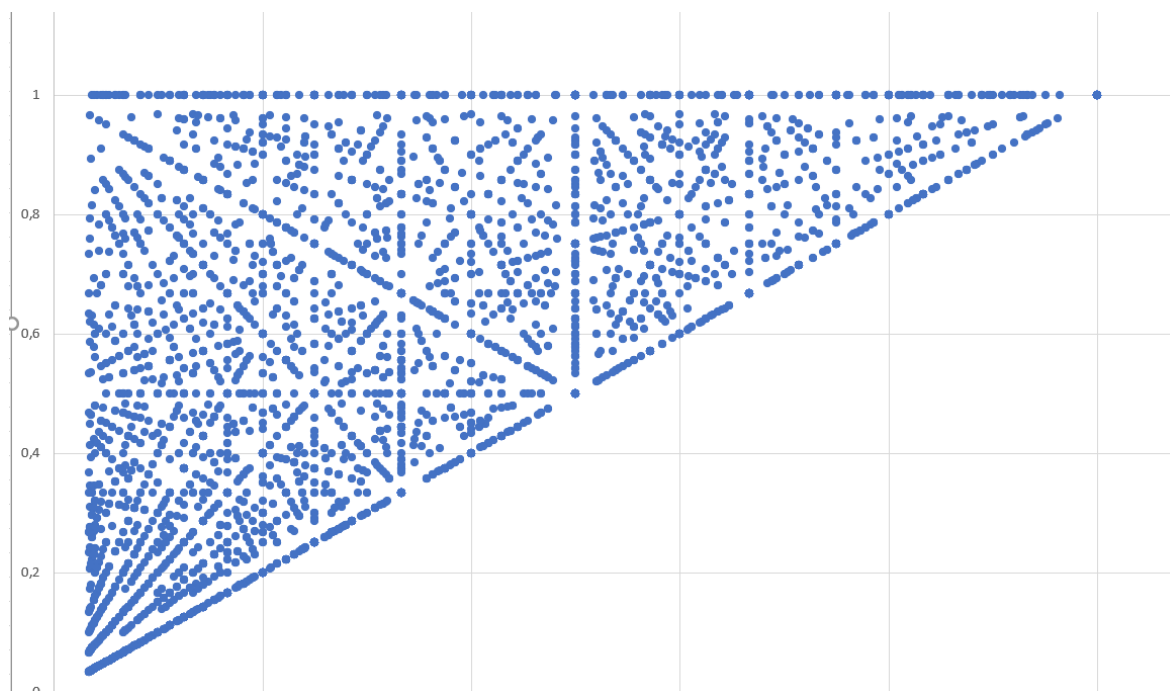
A kavicsok alakja sok információt hordoz a természetben lejátszódó eróziós folyamatokról, így például az üledékképződésről és a geofizikai transzport folyamatokról, emiatt a téma ma is a geomorfológiai kutatások homlokterében áll.

A geológiában számos osztályozási rendszer létezik a kavicsok morfológiájának leírására. Ezek közül a legrégebbi és legelterjedtebb a Zingg-féle osztályozás (1935)[1], valamint a Sneed és Folk-osztályozás (1958)[2], mely utóbbi a Zingg osztályozás továbbfejlesztésének tekinthető. Bár máig vitatott, hogy melyik a leginformatívabb rendszer, az azonban tény, hogy manapság is széles körben alkalmazzák ezeket a leírásokat.

A Zingg-osztályozás a kavicsok alakját egy  $a \geq b \geq c$  tengelyhosszokkal rendelkező háromtengelyű ellipszoiddal közelíti, és ezt az  $(y_1, y_2) = (c/a, b/a)$  tengelyarányokra megadott intervallumok segítségével 4 osztályba sorolja. A Zingg-féle osztályozás alkalmazásakor a geológusok a terepen megméri az  $a > b > c$  tengelyhosszakokat és ezekből számítják a tengelyarányokat. Ilyen mérések akár több százezer kavics adatait is tartalmazhatják, például Carr 1969-ben az egyesült királyságbeli Chesil Beachen 100000 db kavics mérését végezte el.[3][4] Dolgozatomban a tengelyarányok terepen mért adatainak értelmezésével foglalkozom.

Mivel a mérés pontossága véges (a geológus a mért értékeket kerekíti), ezért feltesszük, hogy az  $a, b, c$  tengelyhosszok természetes számok. A terepi méréseket véletlen adathalmazzal közelíttem, vagyis feltettem, hogy  $a, b, c$  diszkrét, egyenletes eloszlású valószínűségi változók, rendre a  $[1, 2, \dots, K]$ ,  $[1, 2, \dots, a]$ ,  $[1, 2, \dots, b]$  halmazokon. Egy így generált adathalmazhoz tartozó tengelyarányokat az  $[y_1, y_2]$  koordináta-rendszerben ábrázolva azt várnánk, hogy a pontok véletlenszerűen helyezkednek el. Ezzel szemben kirajzolódik egy mintázat, mintha a véletlen nem is lenne teljesen véletlen. Megfigyelhető, hogy az ábrázolt pontok egy része egyenesekre illeszthető. Dolgozatomban megmutatom, hogy ezen egyenesek természetes módon vannak jelen, ha ezek nincsenek ott, akkor lehet okunk gyanakodni arra, hogy a minta esetleg manipulált.

Dolgozatomban definiálom az  $[y_1, y_2]$  térben ábrázolt ponthalmaz egyeneseinek *láthatóságát*, továbbá ezen egyenesek meredekségét  $p/q$  relatív prím alakban veszem fel. Excel és a Matlab programok segítségével eltérő számosságú ponthalmazokat elemezve megállapítottam, hogy minél kisebb a  $(p, q)$  vektor normája, az egyenes annál inkább látható.



1.1. ábra 1000 darab véletlen kísérlet ábrázolása

## 2 Bevezetés

A kavicsok változatos formavilága iránt manapság is nagy az érdeklődés hiszen a kavicsok alakja segíthet megérteni a természetben lejátszódó eróziós folyamatokat. A kavicsok alakja számos információt hordoz az üledékképződésről és a múltbéli, illetve jelenkori transzportfolyamatokról. Nem véletlen, hogy a kavicsok alakjának vizsgálata ma frekventált kutatási téma a geomorfológia területén.

A kavicsok alakjának vizsgálatára a történelem során számos osztályozási rendszert dolgoztak ki. A klasszikus rendszerek általában hossz mértéken alapulnak. A Wentworth-osztályozás a kavicsokat és szemcséket méretük alapján sorolja különféle osztályokba. Krumbein a Wentworth-osztályokhoz rendelt hozzá egy logaritmikus skálát ( $\phi$ -skála).

Dolgozatomban a Zingg-féle osztályozással fogok foglalkozni, mely az egyik legrégebbi és legelterjedtebb hossz mértéken alapuló osztályozási rendszer. A Zingg-osztályozás [1] a kavicsok alakját egy  $a \geq b \geq c$  tengelyhosszú ellipszoiddal közelíti, melyeket a  $\frac{b}{a}, \frac{c}{a}$  tengelyarányok szerint négy osztályba sorol. (3.1. ábra)

Dolgozatomban felhasználva a Zingg-osztályozás tengelyarányait definiálok két modellt (véletlen és determinisztikus), melyekkel a geológusok által terepen mért adatokat közelítem. Mivel a terepen mért eredmények pontossága véges így felteszem, hogy  $a, b, c$  számok pozitív egész számok. A véletlen modellben  $a, b, c \in \mathbb{N}$  számokat sorsoljuk rendre az  $[1, 2, \dots, K]$ ;  $[1, 2, \dots, a]$ ;  $[1, 2, \dots, b]$  intervallumokból, ahol  $K \in \mathbb{N}$  a maximális tengelyhossz. Egy derékszögű koordináta rendszerben ábrázolva a tengelyarányainkat  $(x = \frac{b}{a}, y = \frac{c}{a})$  megfigyelhető, hogy kirajzolódik egy mintázat, a pontok egy része egyenesekre illeszthető. Ha ezek az egyenesek nincsenek jelen a mintázatban, akkor lehet okunk gyanakodni arra, hogy manipulált az adathalmazunk.

A véletlen modell szerinti kiválasztás során a pontok eloszlása nem egyenletes, valamint nem jelenik meg az összes lehetséges pont  $(0,0)$ ;  $(0,1)$ ;  $(1,1)$  tartományon belül. Mivel célunk az egyenes láthatóságának vizsgálata, így célszerű egy olyan modellt bevezetni, amely tudja az előbb leírt tulajdonságokat.

A determinisztikus modellben a pontjainkat  $[a, b, c]$  térben, a T tetraéderen belülről választjuk ki az egész koordinátájú pontokat. T csúcspontjai rendre:  $(0, 0, 0)$ ,  $(K+1, 0, 0)$ ,  $(K+1, K+1, 0)$ ,

Geometriai mintázatok felbukkanása véletlen kavicspopuláció tengelyarányainak fázissterében

$(K+1, K+1, K+1)$ . Vegyük észre, hogy a véletlen modell kiválasztási szabálya megfelel annak, hogy  $T$  tetraéderen belül egyenletes valószínűséggel választunk egy egész koordinátájú pontot.

Dolgozatomban definiálom a  $[\frac{b}{a}, \frac{c}{a}]$  síkon található egyenesek láthatóságát, továbbá ezen egyenesek meredekségét  $p_1/p_2$  relatív prím alakban veszem fel. Matlab programban a determinisztikus modellt használva többféle mérőeszközt kipróbálva eltérő számosságú adathalmazokon megállapítom, hogy minél kisebb a  $(p_1, p_2)$  vektor normája, a vizsgált egyenes annál inkább látható.

### 3 Osztályozási rendszerek

A geológiában számos osztályozási rendszer létezik a kavicsok morfológiájának leírására, mivel a kavicsok alakja sok információt hordoz a természetben lejátszódó folyamatokról. A klasszikus rendszerek általában hossz mértéken alapulnak.

#### 3.1 Krumbein osztályozás

**Definíció:** A *szemcseméret* a törmelékes üledék, üledékes kőzetek, hordalék, talajok, valamint a mesterségesen előállított, szilárd, szemcsés-darabos anyagok szemcséinek térbeli kiterjedését jellemző érték.

Az általánosan elfogadott szemcseméret-tartományok: kolloid, agyag, kőzetliszt, homok, kavics, görgeteg, tömb. A természetesen kialakuló anyagok nagyon ritkán tartalmaznak csak azonos méretű szemcséket.

**Definíció:** A *szemcseeloszlás* adott anyagban a szemcsék méret szerinti eloszlása, tehát az egyes szemcseméret-tartományok tömeg, térfogat vagy darabszám szerinti részaránya.

Krumbein egy logaritmikus skálát hozott létre ( $\phi$ -skála), amely a  $\phi$ -érték és a szemcseméret között a következő összefüggést definiálja:

$$\phi = -\log_2 \frac{D}{D_0}, \text{ ahol}$$

$\phi$  a Krumbein-féle  $\phi$ -érték

$D$  szemcse átmérője milliméterben

$D_0$  az egységátmérő (1 mm)

A Krumbein-féle osztályozás a  $\phi$ -értéken alapul, ezek alapján határozott mérettartományokat, melyeket az Egyesült Államokban már létező Wentworth-osztályokhoz rendelt hozzá. A Wentworth-osztályozás (1922) a kavicsokat és szemcséket méretek alapján kategorizálja.

#### 3.2 Zingg-féle osztályozás

A hossz mértéken alapuló osztályozási rendszerek közül napjainkban az egyik legelterjedtebb a Zingg-féle osztályozás (1935)[1].




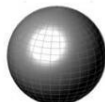
Geometriai mintázatok felbukkanása véletlen kavicspopuláció tengelyarányainak fázissterében

A Zingg-féle osztályozás a kavicsok alakját egy háromtengelyű ellipszoiddal közelíti, melyek tengelyhosszai rendre  $a \geq b \geq c$ . Ezeket a  $\frac{b}{a}, \frac{c}{a}$  tengelyarányok szerint sorolja négy osztályba. (3.1. ábra)

A Zingg-féle osztályozás alkalmazásakor a geológusok terepen megméri a tengelyhosszokat és ezekből számolják az arányokat. Az ilyen mérések akár több százezer kavics adatait tartalmazhatják, például Carr 1969-ben az egyesült királyságbeli Chesil Beachen 100000 darab kavics mérését végezte el. [3] [4]

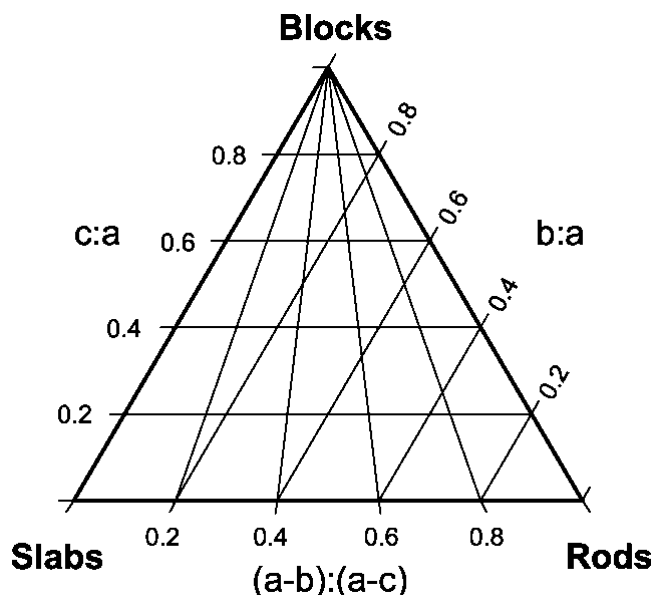
A Sneed és Folk osztályozás (1958) [2] tulajdonképpen a Zingg-féle osztályozás továbbfejlesztése, ami a fenti tengelyarányokon kívül figyelembe veszi az  $\frac{a-b}{a-c}$  arányt is. (2.2. ábra)

Ezek az osztályozási rendszerek hasznosnak bizonyultak a gyakorlatban, azonban a mérések pontatlansága és a tengelyirányok nehéz beazonosítása miatt bizonytalanságot hordoznak magukban. Ezen felül a Zingg-osztályozás az osztályok közötti hatást önkényesen a  $\frac{2}{3}$  tengelyarányánál veszi fel. Ez viszont általánosítható, ha nem egy konstans tekintünk határnak, hanem felvesszünk egy  $0 < p \leq 1$  paramétert, ami az osztályokat elválasztó tengelyarányt jelöli

ZINGG CLASSES	$b/a < 2/3$	$b/a > 2/3$
$c/b < 2/3$	<b>BLADE</b> 	<b>DISC</b> 
$c/b > 2/3$	<b>ROD</b> 	<b>SPHERE</b> 

**Zingg classes**  
 $a > b > c$ : axis lengths of the ellipsoid

3.2. ábra Zingg osztályozás



3.1. ábra Sneed and Folk osztályozás



## 4 Alapfeltevések

Mivel a mérés pontossága véges, geológusok a mért értéket kerekítik. Dolgozatomban felteszem, hogy az  $a, b, c$  tengelyhosszok természetes számok.

A terepi méréseket véletlen adathalmazokkal közelítem, tehát felteszem, hogy az  $a, b, c$  számok diszkrét, egyenletes eloszlású valószínűségi változók rendre a  $[1, 2, \dots, K]$ ,  $[1, 2, \dots, a]$ ,  $[1, 2, \dots, b]$  halmazokon.

### 4.1 Modell 1: Véletlen modell

Excel program segítségével generáltam egy véletlen adathalmazt, melynek változói  $K, M, a, b, c$ , ahol  $K, M \in \mathbb{N}$  és  $a \in [1; K]$ ,  $b \in [1; a]$ ,  $c \in [1; b]$ .

$K$ : a maximális tengelyhossz

$M$ : a kavicspopuláció nagysága

Bevezetve az  $x = \frac{b}{a}$  és az  $y = \frac{c}{a}$  számokat, ábrázoljuk Descartes féle koordináta rendszerben pontfelhő formájában a kapott számokat. (4.1. ábra)

Többszöri futtatásokat figyelembe véve a következő megfigyeléseket tehetjük:

-A pontok a  $(0,0)$ ;  $(0,1)$ ;  $(1,1)$  derékszögű háromszögön belül helyezkednek el. A továbbiakban a háromszög alakú tartományt  $\Omega$ -val jelöljük. Az első megfigyelés természetes módon következik a tengelyarányokból és az  $a, b, c$  számok közötti nagyságbeli különbségekből.

- A pontok eloszlása a nem egyenletes. Ennek oka, hogy  $a \in [1; K]$ ,  $b \in [1; a]$ ,  $c \in [1; b]$ , így  $b$  értéke függ  $a$ -tól,  $c$  értéke pedig  $a$ -tól és  $b$ -től. Tehát az  $a, b, c$  diszkrét egyenletes eloszlású valószínűségi változók egymástól nem függetlenek. Ennek egyik következménye, hogy amennyiben  $a$  értéke kicsiny, akkor  $b$  is és  $c$  is kicsi lesz.

-Vegyük észre, hogy kicsiny  $b$  és kicsiny  $c$  értékhez egyaránt tartozhat kicsiny vagy nagy  $a$  érték. Ebből következően az origó környezetében a pontok sűrűsödnek.

-Feltűnő, hogy a generált ábrákon a pontokból egyenesek rajzolódnak ki.

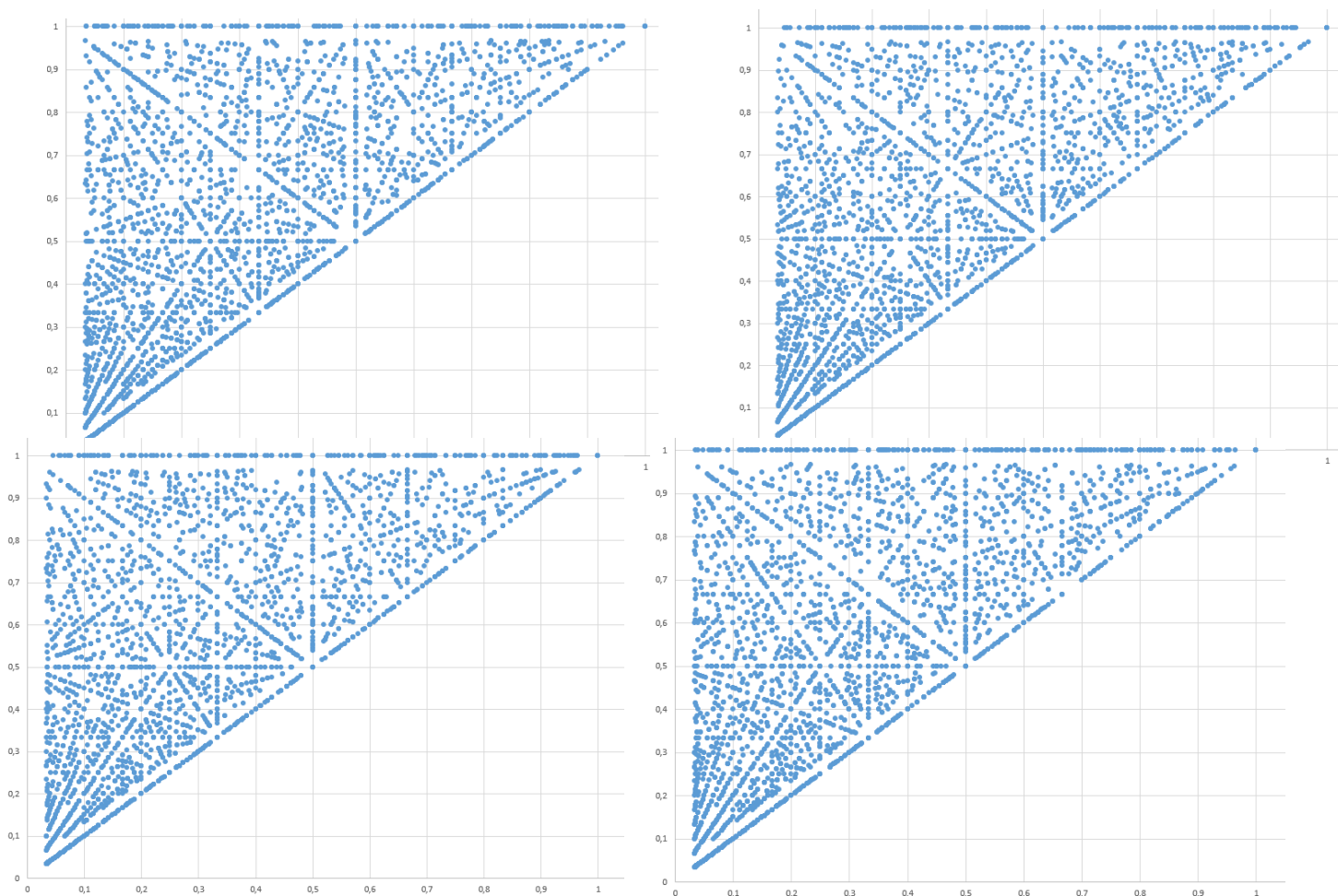
Továbbiakban ezen egyenesekkel fogok foglalkozni:

$$(1) \quad x = \frac{b}{a}, y = \frac{c}{a}$$

$$(2) \quad y = px + q$$

$$\frac{c}{a} = p \frac{b}{a} + q$$

ahol  $p$  jelöli az egyenes meredekségét és  $q$  az  $y$  tengellyel vett tengelymetszetet.



4.1. ábra A véletlen modell futtatásai közti különbségek

**Definíció:**

Rács egyenesnek nevezzük azokat az  $y=px+q$  egyeneseket, amelyek az  $\Omega$  tartományon belül legalább két rácsponton átmennek.

**Definíció:**

Egy egyenes vizuális detektálhatósága azon rácspontok száma, amelyeken az egyenes átmegy.  
jele:  $N(p,q)$

Feltételezzük, hogy  $a=d$ ,  $d \in \mathbb{R}$  esetén  $N(b)$  kvázikonkáv, tehát egyetlen maximummal és minimummal rendelkezik. Azaz, ha egy egyenest a háromszög tartományon belül önmagával párhuzamosan eltolom, akkor a tartományon belül  $N$ -nek lesz egy maximuma.

Mivel a célunk az egyenesek láthatóságának vizsgálata, célszerűnek tűnt egy olyan modell bevezetése, amelyben minden lehetséges pont megjelenik az  $\Omega$  tartományon belül. Ezt a megközelítést *determinisztikus modellnek* nevezem.

## 4.2 Modell 2: Determinisztikus modell

Az új modellben a pontokat nem véletlenszerűen sorsoljuk, hanem az  $[a, b, c]$  térben egy  $T$  tetraéderen belül választjuk az egész koordinátájú pontokat.  $T$  csúcspontjai rendre  $(0, 0, 0)$ ,  $(K+1, 0, 0)$ ,  $(K+1, K+1, 0)$ ,  $(K+1, K+1, K+1)$ . Vegyük észre, hogy a véletlen modell (Modell 1) kiválasztási szabálya megfelel annak, hogy  $T$  tetraéderen belül egyenletes valószínűséggel választunk egy egész koordinátájú pontot. A determinisztikus modellben  $T$  belsejének minden egész koordinátájú pontját figyelembe vesszük.

A következő ábra szemlélteti a különbséget a kétféle modell között.  $K=18$ -ig képeztük a pontokat mindkét esetben. A 10. oldalon szereplő állítás miatt 1140 db pontot generáltam mindkét modell szerint.

Látható, hogy a véletlen modell nem tartalmazza az összes egész koordinátájú pontból képzett számokat, valamint, hogy a pontok eloszlása nem egyenletes, míg a determinisztikus modellben az összes pont látható, ezek eloszlása pedig egyenletes.

**Állítás:**  $S_K = \frac{(K+1)^3 - (K+1)}{6}$  db egész koordinátájú pont van.

**Bizonyítás:** Mivel számhármassok létrehozásánál fennáll az alábbi egyenlőtlenség  $a \geq b \geq c$  így ezek száma megkapható a következő összegzésből:

$$\sum_{a=1}^K \sum_{b=1}^a \sum_{c=1}^b 1 = \sum_{a=1}^K \sum_{b=1}^a b = \sum_{a=1}^K \frac{a * (a + 1)}{2} = \sum_{a=1}^K \frac{a^2}{2} + \frac{a}{2}$$

felhasználva a négyzetszámok és a természetes számok 1-től  $K$ -ig szummázására szolgáló képleteket, a következő összefüggést kapjuk:

$$\begin{aligned} \sum_{a=1}^K \frac{a^2}{2} + \frac{a}{2} &= \frac{K * (K + 1) * (2K + 1)}{6 * 2} + \frac{K * (K + 1)}{2 * 2} = \\ &= \frac{K * (K + 1) * (2K + 1)}{12} + \frac{3 * K * (K + 1)}{12} = \\ &= \frac{(K + 1) * (K * (2K + 1) + 3K)}{12} = \\ &= \frac{2K^3 + K^2 + 2K^2 + K + 3K^2 + 3K}{12} = \frac{2K^3 + 6K^2 + 4K}{12} = \frac{K^3}{6} + \frac{K^2}{2} + \frac{K}{3} \end{aligned}$$

Most pedig belátom, hogy

$$\frac{(K + 1)^3 - (K + 1)}{6} = \frac{K^3}{6} + \frac{K^2}{2} + \frac{K}{3}$$

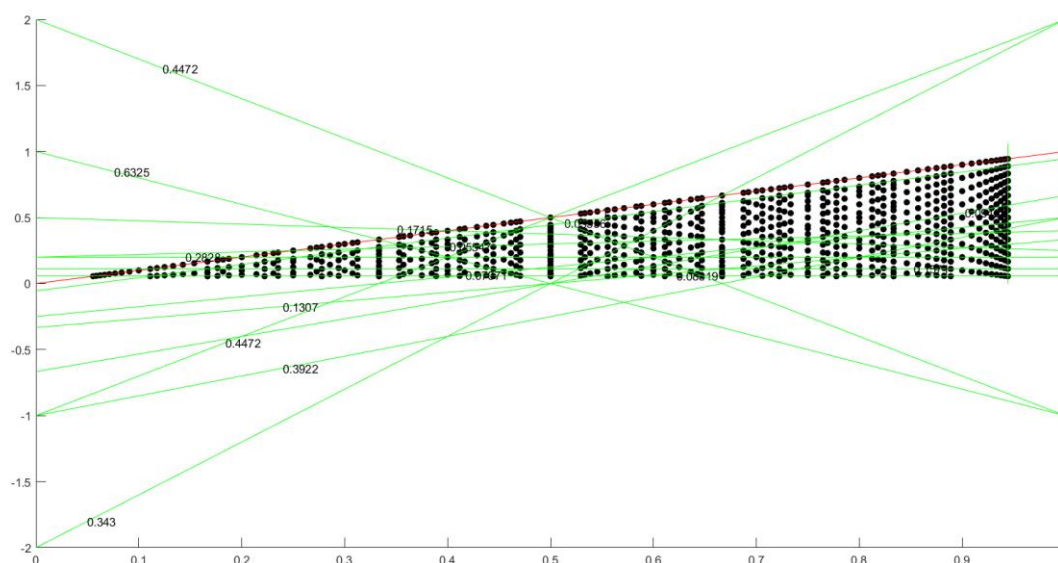
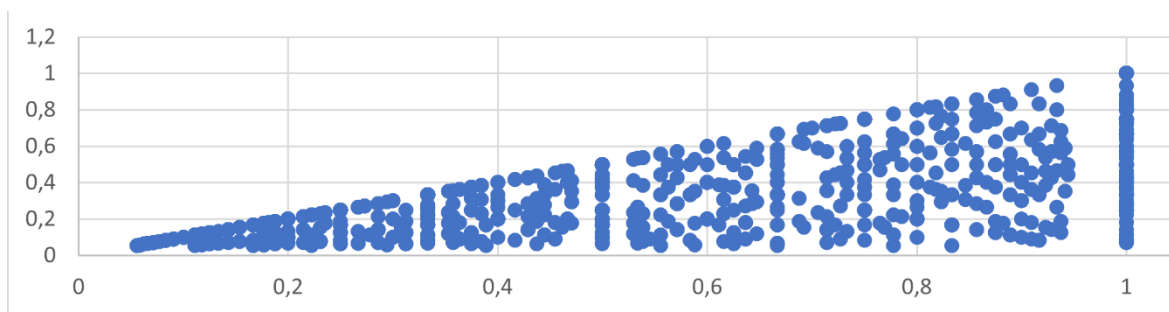
$$(K + 1)^3 - (K + 1) = K^3 + 3K^2 + 3K + 1 - K - 1 = K^3 + 3K^2 + 2K$$

Ezt elosztva 6-tal megkapjuk a bizonyítani kívánt összefüggést:

$$S_K = \frac{K^3}{6} + \frac{K^2}{2} + \frac{K}{3} = \sum_{a=1}^K \sum_{b=1}^a \sum_{c=1}^b 1 = \frac{(K + 1)^3 - (K + 1)}{6}$$

A 4.2. ábra szemlélteti a különbséget a véletlenszerű és a determinisztikus modell között: K=18-ig képeztük a pontokat mindkét esetben. Az imént bizonyított állítás miatt 1140 db pontot generáltam mindkét modell szerint.

Látható, hogy a véletlen modell nem tartalmazza az összes egész koordinátájú pontból képzett számokat, míg a determinisztikus modellben az összes pont látható



4.2. ábra A véletlen és a determinisztikus modell összehasonlítása

### 4.3 A ponthalmaz tárolása

Matlab programban egy for ciklus segítségével meghatároztam a, b, c számok lehetséges értékeit és ezeket eltároltam egy tömbben (P) melynek sorai rendre  $a_i, b_i, c_i$  számok. Majd töröltem belőle az ábrán egymást fedő pontokat.

Vizsgáljuk meg a két modell közti kapcsolatot:

(2)-es egyenletből következik, hogy  $[x, y]$  síkban egy egyenes az  $[a, b, c]$  térben egy  $pb + qa - c = 0$  egyenletű síknak felel meg. Vegyünk két pontot  $(P_1, P_2) \in T$  tetraéderen belül:

$$P_1: q * a_1 + p * b_1 - c_1 = 0$$

$$P_2: q * a_2 + p * b_2 - c_2 = 0$$

Az alábbi egyenletrendszernek létezik egyértelmű megoldása, ha az együtthatómátrixa nem szinguláris.

Megoldás: Cramer szabály felhasználásával

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 * b_2 - a_2 * b_1$$

$$\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} = c_1 * b_2 - c_2 * b_1$$

$$\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = a_1 * c_2 - a_2 * c_1$$

A pontosság érdekében a nevezőt és a számlálót külön tároltam, így ugyanis egész számokkal számolhatok tovább. Ezért legyen  $p = \frac{p_1}{p_2}, q = \frac{q_1}{q_2}$ .

$$p = \frac{a_1 * c_2 - a_2 * c_1}{a_1 * b_2 - a_2 * b_1}$$

$$p_1 = a_1 * c_2 - a_2 * c_1$$

$$p_2 = a_1 * b_2 - a_2 * b_1$$

$$q = \frac{c_1 * b_2 - c_2 * b_1}{a_1 * b_2 - a_2 * b_1}$$

$$q_1 = c_1 * b_2 - c_2 * b_1$$

$$q_2 = a_1 * b_2 - a_2 * b_1$$

Az összes lehetséges pontpárra kiszámoltam  $p_1, p_2, q_1, q_2$  számokat. Mivel  $p(i) = \frac{p_1(i)}{p_2(i)}$  és  $q(i) = \frac{q_1(i)}{q_2(i)}$  így a könnyebb átláthatóság érdekében leegyszerűsítettem minden  $p_1(i), p_2(i)$  és  $q_1(i), q_2(i)$  számpárt a hozzájuk tartozó legnagyobb közös osztóval.

Legyen  $p_1$  oszlopvektor, melyben  $p_1(i)$  értékei szerepelnek egyszerűsítés után. Hasonlóan képezzük  $p_2, q_1, q_2$  oszlopvektorokat.

Legyen  $ppqq$  egy tömb, amelynek oszlopai rendre  $p_1, p_2, q_1, q_2$  vektorok.

## 5 Vizsgálat eszközei

Az adathalmazunk vizsgálatunkhoz vezessünk be néhány fogalmat és mérési eszközt.

Legyen  $\Omega$  egy konvex tartomány egy  $[x_1, x_2]$  négyzetrácson.

### Definíció:

$R_a(p)=0$ , ha  $p \in \mathbb{R} \setminus \mathbb{Q}$

$R_a(p) = \frac{\sqrt{2}}{\sqrt{p_1^2 + p_2^2}}$ , ha  $a = \frac{p_1(p)}{p_2(p)}$ , ahol  $p_1(p), p_2(p) \in \mathbb{N}$  valamint  $p_1(p)$  és  $p_2(p)$  relatív prímelek

Definíció:

$R_b(q)=0$ , ha  $q \in \mathbb{R} \setminus \mathbb{Q}$

$R_b(q) = \frac{1}{q_2(q)}$ , ha  $b = \frac{q_1(q)}{q_2(q)}$ , ahol  $q_1(q), q_2(q) \in \mathbb{N}$  valamint  $q_1(q)$  és  $q_2(q)$  relatív prímelek

Definíció:

$R(p,q) = R_a(p) * R_b(q)$

Az  $R_a(p)$  függvény azt méri, hogy rögzített oldalhosszúságú négyzetrács esetén mennyire látható az egyenes, az  $R_b(q)$  pedig, hogy sok eltérő négyzetrács esetén mennyire látható.

Alkalmazzuk a fent szereplő függvényeket a mi adathalmazunkon.

$$R_a(k) = \left| \frac{\sqrt{2}}{\sqrt{(p_1(p))^2 + (p_2(p))^2}} \right|$$

$$R_b(q) = \left| \frac{1}{q_2(q)} \right|$$

$$R(k) = R_a(k) * R_b(k)$$

A ppqq tömb elemeinek felhasználásával kiszámoljuk  $R_a(p)$ ,  $R_b(q)$ ,  $R(p,q)$  értékeket és ezeket eltároljuk rendre az  $R_a$ ,  $R_b$  és  $R$  oszlopvektorokban.

Válasszunk ki két pontot ( $P_i, P_j$ ). Jelölje  $N(i, j)$  a két kiválasztott ponttal egy egyenesre eső pontok számát. Látható, hogy  $2 \leq N(i, j) \leq M$ .

Egy egyenes láthatóságát meghatározza, hogy hány pont található rajta. Emiatt meg kell számolni az egy egyenesre eső pontokat. Vegyük észre, hogy a ppqq tömb egy sora egyértelműen meghatároz egy egyenest. Számoljuk meg, hogy hány olyan egyenes van,



melyeket ugyanazon  $p_1, p_2, q_1, q_2$  paraméterek jellemzik. Ezeket a számokat rendeljük hozzá a ppqq mátrix megfelelő soraihoz. Így kapunk egy öt oszlopból álló mátrixot, melynek első négy oszlopa  $p_1, p_2, q_1, q_2$  vektorok ötödik oszlopa pedig az előbbieken kiszámolt adatokból álló oszlopvektor, aminek legyen a neve  $maph$ .

Állítás:

$$N(k) = \frac{1 + \sqrt{1 + 8 * maph(k)}}{2}$$

Bizonyítás:

A ppqq mátrixban szerepelnek a tetszőleges két pont között rajzolható egyenes paraméterei ( $p_1, p_2, q_1, q_2$ ). A  $maph$  változó megmutatja, hogy hány olyan egyenes van, amelynek  $p_1, p_2, q_1, q_2$  paraméterei megegyeznek, másként fogalmazva, hogy hányszor két pontja közé húzható be egy egyenes.

Válasszunk ki egy egyenest és annak egy pontját. Az egyenesen van  $N$  darab pont. A kiválasztott pontból ( $N-1$ ) pontba húzható egyenes.  $N$  darab pontunk van, tehát az összesen húzható egyenesek száma  $N*(N-1)$ , így azonban minden egyenest kétszer számoltunk, ezért ezt a számot el kell osztani kettővel. Ez a szám azt jelöli, hogy az egyenes pontjai között hány szakasz található, tehát megegyezik a  $maph$  változóval. Így:

$$maph = \frac{N * (N - 1)}{2}$$

$$2 * maph = N * (N - 1)$$

$$N^2 - N - 2 * maph = 0$$

Ezt a másodfokú egyenletet  $N$ -re megoldva kapjuk:

$$N_{1,2} = \frac{1 \pm \sqrt{1 + 8 * maph}}{2}$$

Mivel  $2 \leq N(i, j) \leq M$  így az egyenletünknek csak egy megoldása lesz:

$$N = \frac{1 + \sqrt{1 + 8 * maph}}{2}$$

$N$  értékeit tároljuk az  $N$  vektorban.

Sejtésünk az, hogy  $N$  és  $R$  között monoton az összefüggés.

## 6 Adatelemzés

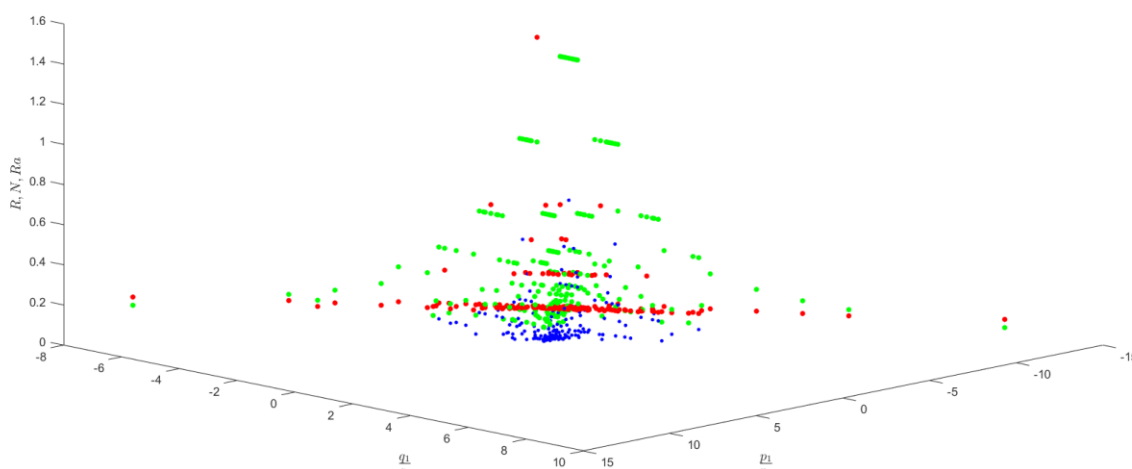
### 6.1 Alapdiagrammok

A következőkben a Matlab program által generált adathalmazokat fogom elemezni.

Az 6.1. ábra egy egyesített 3-dimenziós grafikon, melynek alapsíkra eső tengelyei:

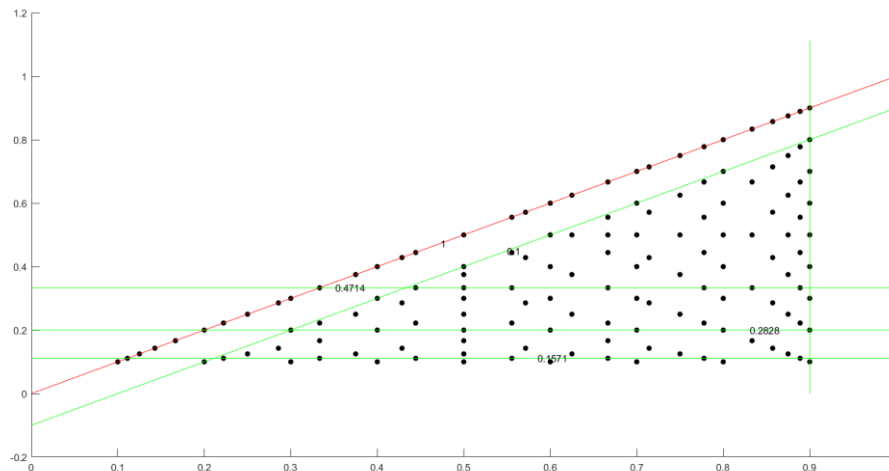
$$\frac{p1}{p2} \text{ és } \frac{q1}{q2}$$

a z tengely pedig  $R(p,q)$ ,  $N(p,q)$ , és  $R_a(p)$ , értékeit tartalmazza, melyeknek színei rendre: kék, piros, zöld. (N értékei egy skalárral vannak normálva a jobb láthatóság miatt)



6.1. ábra  $R(p,q)$ ,  $N(p,q)$ ,  $R_a(p)$   $p1/p2$ ,  $q1/q2$  függvényében

A grafikonon az látszik, hogy a pontjaink az alapsíkon a  $[-1;1] \times [-1;1]$  intervallumban sűrűsödnek. Ennek oka az, hogy a tartományunk, ahonnan a pontokat választjuk lehatárolható a  $(0,0)$ ,  $(0,1)$ ,  $(1,1)$  csúcspontú háromszöggel. Megfigyelések alapján látható, hogy sem az egyenesek meredeksége, sem az eltolásuk y tengelyen nem nagy érték így természetes, hogy a fenti ábrán  $[-1;1] \times [-1;1]$  intervallumban sűrűbben vannak a pontok.



6.2. ábra Determinisztikus modellen a pontok elhelyezkedése, ha  $K=10$

Az 6.1.-es ábrán már meg lehet állapítani némi monotonitást az adataink között. Vegyük például az alapsíkon fekvő  $(3; -2,2)$  és  $(1; -0,5)$  pontokat és vizsgáljuk meg a függvényeink értékeit.

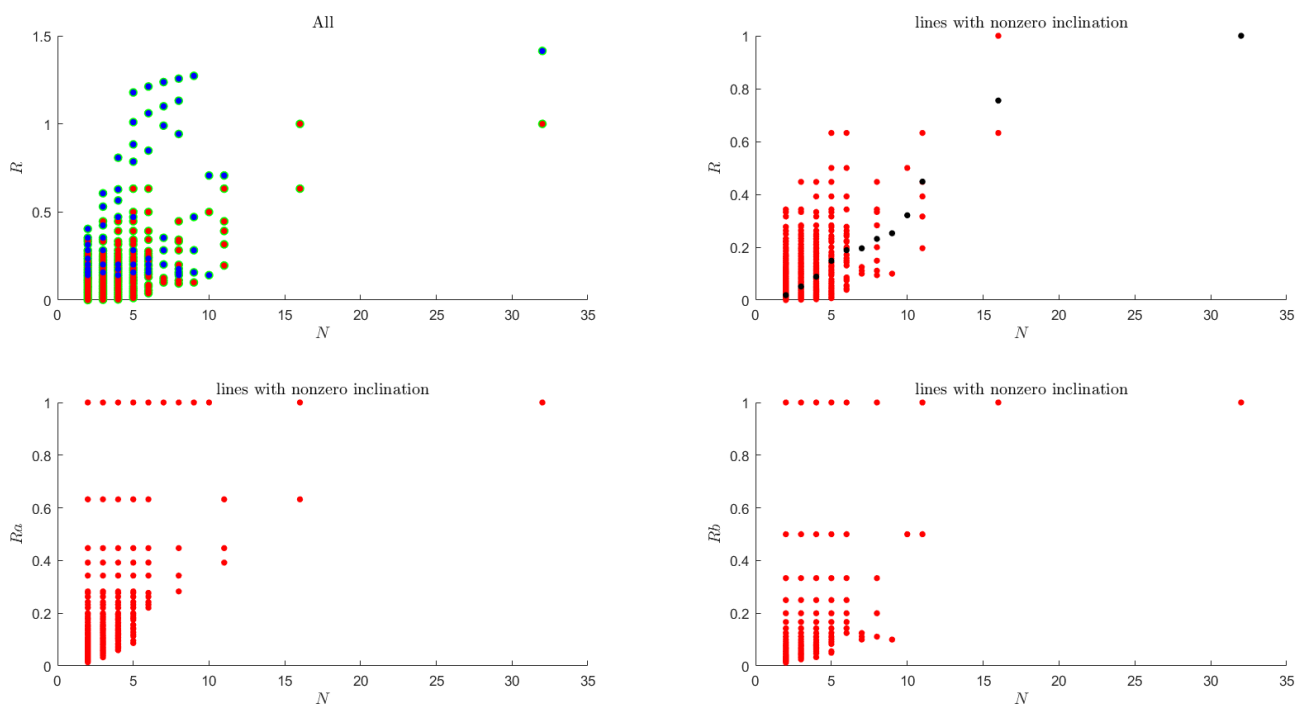
$$R(3; -2,2) = 0,0894427 < R(1; -0,5) = 0,5$$

$$N(3; -2,2) = 0,0169031 < N(1; -0,5) = 0,169031$$

$$R_a(3; -2,2) = 0,447 < R_a(1; -0,5) = 1$$

További pontok vizsgálatával már észrevehető a monoton összefüggés az adataink között, azonban ezt az ábrát nehézkes használni, így bevezettünk új ábrázolásmódokat.

Ábrázoljuk az adatainkat skálázás vagy normálás nélkül és térjünk át a síkbeli koordináta rendszerre.



6.3. ábra  $R(p,q)$ ,  $R_a(p)$ ,  $R_b(q)$  ábrázolása  $N(p,q)$  függvényében skálázás és normálás nélkül

Ábrázoljuk  $R(p,q)$ ,  $R_a(p)$ ,  $R_b(q)$  értékeit  $N(p,q)$  függvényében. ( $N$  a  $(p,q)$  paraméterű egyenesre eső pontok száma).

A 6.3. ábra bal alsó diagrammján  $R_a(p)$  értékeit ábrázoltuk, még hozzá azon egyenesek esetében, melyeknek nem 0 a meredeksége. Figyeljük meg, hogy  $N(p,q)$  értékének növekedésével láthatunk egy hasonlóan növekvő tendenciát  $R_a(p)$  értékeire vonatkozólag, azonban az is észrevehető, hogy  $R_a(p)$  kis  $N(p,q)$ -ra is vesz fel maximális értéket ez pedig sérti a monotonitást

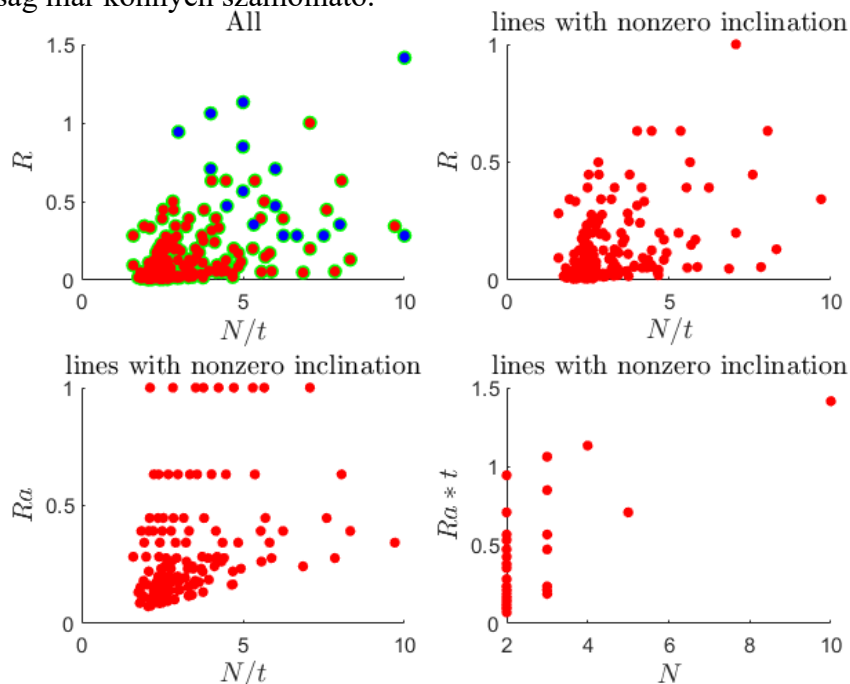
feltételét.

A jobb alsó ábrán  $R_b(q)$  értékeit láthatjuk. A probléma hasonló, mint  $R_a(p)$  értékeinél. Itt viszont nem figyelhető meg  $R_a(p)$ -nél látható monotonitás, mert itt  $N(p,q)$  növekedésével nem lesz nagyobb  $R_b(q)$  értéke minden esetben. Vegyük például  $N=7$  és  $N=9$  eseteket.  $R_b(7)=R_b(9)=0,1$ . A jobb felső diagrammon pirossal jelöljük  $R(p,q)$  megfelelő értékeit, feketével pedig a 12 legnagyobb értékű  $N(p,q)$ -hoz tartozó  $R(p,q)$ -k átlagos értékét. Mivel  $R(p,q)=R_a(p)*R_b(q)$  így hordozza az előző kettőnek a tulajdonságait, tehát valamiféle monotonitás látható, de ez messze van még egy elfogadható eredménytől. A bal felső ábrán ugyanaz látható, mint a jobb felsőn csak kiegészítve a 0 és a  $\infty$  meredekségű egyenesek adataival, melyek kékkel vannak jelölve.

Vizsgáljuk meg  $(N(p,q), R_a(p))$  diagrammot még egyszer.  $N(p,q)$  növekedésével  $R_a(p)$  legkisebb értékei is nőni fognak, tehát ezek között monoton a kapcsolat. Valamint kis  $N(p,q)$  és kis  $R_a(p)$  értékek esetén a pontok besűrűsödnek. Ha  $N(p,q)$  értékeit lenormáljuk, akkor az általunk várt monotonitás  $N$  és  $R_a(p)$  között látható lehet.

## 6.2 Egyenesen lévő pontok számának normalizálása és átskálázása

A következőkben egy normalizálást és egy átskálázást fogunk végrehajtani. Az elsőben  $N(p,q)$  értékét leosztjuk a hozzájuk tartozó egyenesnek a háromszög tartományon belüli hosszával. Ehhez nyilvánvalóan ki kell számolnunk minden  $N(p,q)$  paraméter párhoz az egyenes hosszát. Ennek az algoritmus a következő: megkeressük az egyenes metszéspontját az átfogóval, a függőleges befogóval, valamint a vízszintes befogóval. Ezután a megfelelő metszéspontok közötti távolság már könnyen számolható.



6.4. ábra  $R(p,q)$ ,  $R_a(p)$  ábrázolása  $N(p,q)$  normált verziójával

Az 6.4.-es ábrán ugyanazokat az adatokat ábrázoljuk, mint az 6.3.-ason lévőn, annyi különbséggel, hogy  $N(p,q)$  helyett  $N(p,q)$  normált értékeit szerepeltetjük, valamint a jobb alsó ábrán  $R_b(q)$  helyett  $R_a(p)*t$  értéket vizsgáljuk  $N$  függvényében, ahol  $t$  az egyenes háromszögön belüli hossza.

$R_b(q)$  kihagyásának az az oka, hogy ennél a függvénynél nem volt megfigyelhető az a monotonitás, amit vártunk. Megvizsgálva az ábrákat, levonhatjuk azt a következtetést, hogy  $N(p,q)$  normálása annak az egyenesnek a hosszával, melyre a kiválasztott pontunk ráesik nem járt eredménnyel. Ennek az oka az, hogy hiába esik relatív sok pont egy egyenesre, ha a háromszög területen belüli hossza nagy, akkor az ezzel történő normálás torzítja az eredményeinket. Ezen felül vannak olyan egyenesek is, amelyek láthatóak annak ellenére, hogy kevés pont van rajtuk. Ha ez az egyenes relatív hosszabb, akkor a normálás nagyon lecsökkenti  $N(p,q)$  értékét,  $R_a(p)$  értékei viszont lehetnek nagyok, mert azok csak az egyenes meredekségétől függenek.

A 6.2.-es ábrán észrevehetjük, hogy a háromszög átfogója alatt megtalálható egy ugyanolyan meredekségű egyenes, amelyre azonban kevesebb pont esik. Így adódott a gondolat, hogy  $N$  értékeit ne normáljuk az egyenes hosszával, hanem skálázzuk át oly módon, hogy az azonos meredekségű egyenesekhez rendeljük hozzá a legmagasabb  $N(p,q)$  értéket.

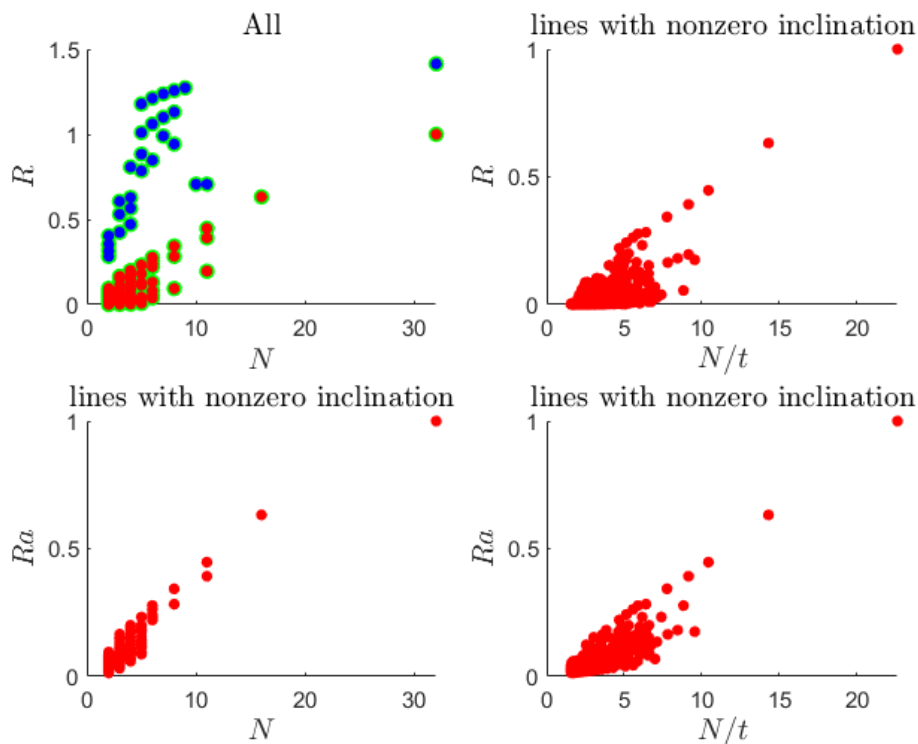
A skálázás során  $N(p,q)$  értékeit meg fogjuk változtatni. A  $p_1, p_2, q_1, q_2$  paraméterek egyértelműen meghatározzák az egyenesünket. Válasszunk ki egy egyenest és vegyük az összes olyan egyenest, aminek a meredeksége (tehát  $\frac{p_1}{p_2}$ ) megegyezik az általunk választott egyenesével. Ezután hozzunk létre egy az egyes megfeleltetést úgy, hogy az azonos meredekségű egyenesekhez hozzárendeljük azt a számot, ami a legtöbb pontot tartalmazó egyenesen lévő pontok száma.

Az előbb említett skálázást végrehajtva a 6.5. ábra diagrammjain megfigyelhető a várt monotonitás.

Vizsgáljuk meg az ábráinkat. Az első ábrán az imént említett átskálázott  $N$  értékek láthatók az  $x$  tengelyen, az  $y$  tengelyen pedig  $R(p,q)$  értékei. Kékkel szerepelnek az egyenesek, amelyek függőlegesek vagy vízszintesek, pirossal pedig a nem 0, de véges meredekségűek. A bal alsó ábrán az  $x$  tengelyen szintén az átskálázott  $N(p,q)$  értékek szerepelnek. A másik két ábrán ennek az átskálázott  $N(p,q)$  értéknek az egyenes hosszával normált verzióját figyelhetjük meg. Mindegyik ábráról már leolvasható az a monotonitás, amelyet vártunk. Mégis látható, hogy az

Geometriai mintázatok felbukkanása véletlen kavicspopuláció tengelyarányainak fázisterében

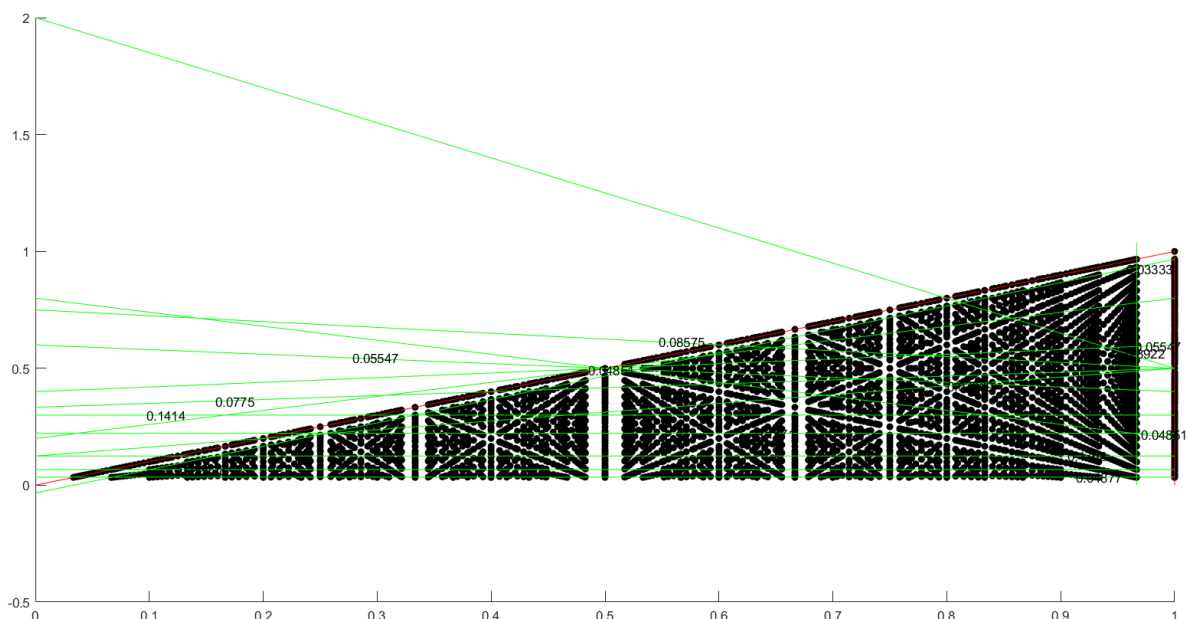
$R_b(q)$  függvényünk, valamilyen szinten elrontja az összefüggésünket. Ezért azt gondolom, hogy egy egyenes vizuális detektálhatóságára  $R_b(q)$  függvény nem jó mérőeszköz,  $R_a(p)$  viszont igen.



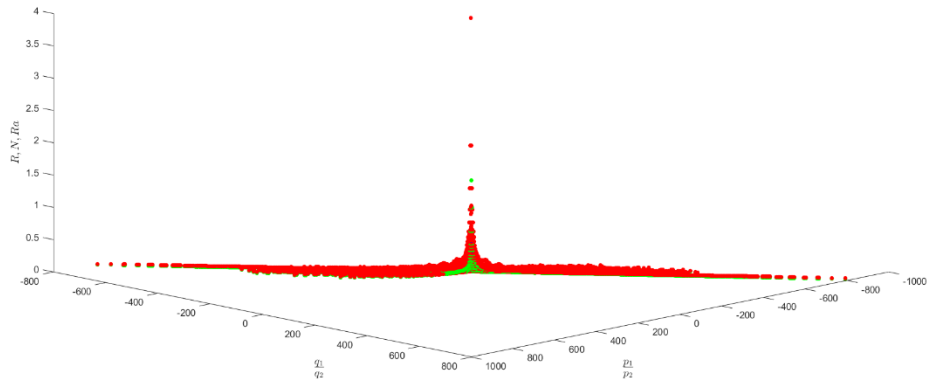
6.5. ábra  $R(p,q)$ ,  $R_a(p)$  ábrázolása a skálázott  $N(p,q)$  függvényében,  $K=10$

Így megállapítható, hogy minél kisebb a  $(p_1, p_2)$  vektor normája, az egyenes annál jobban látható.

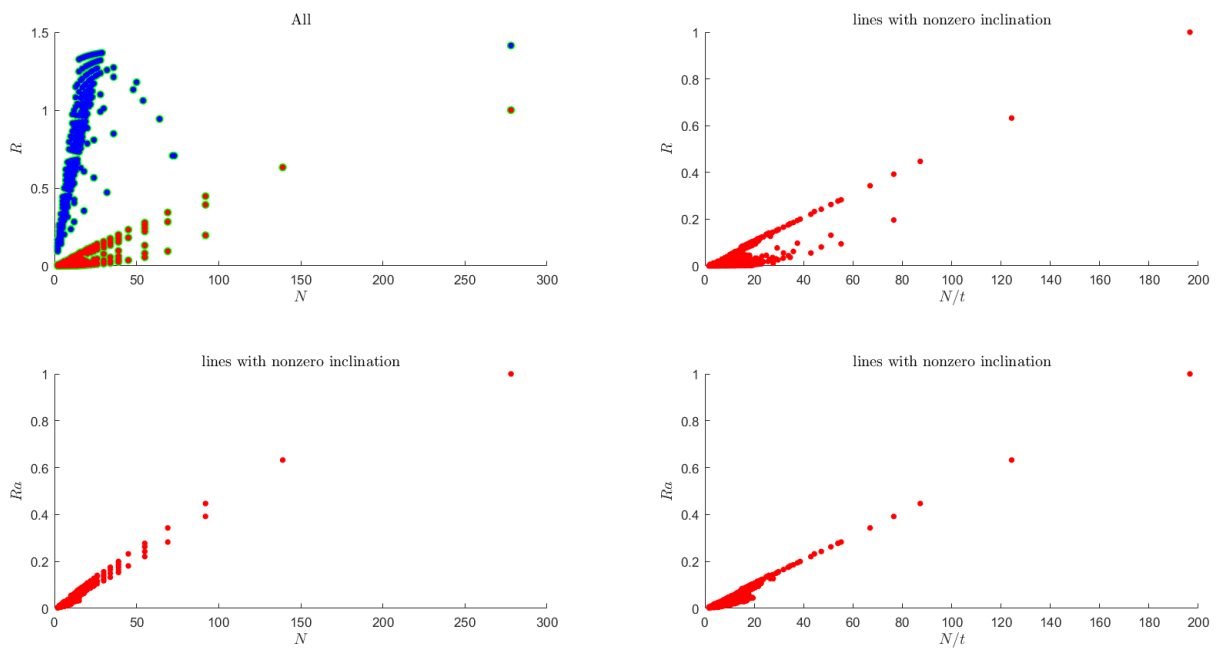
A fent szereplő ábrák az átláthatóság kedvéért  $K=5$ , és  $K=10$  eseteket mutatnak be. Azonban  $K$  növelésével az ábrázolt értékek jobban egy görbére illeszthetők. Vegyük például az  $K=30$  esetet.



6.6. ábra Determinisztikus modellen a pontok eloszlása, ha  $K=30$



6.7. ábra  $R(p,q)$ ,  $N(p,q)$ ,  $R_a(p)$   $p1/p2$ ,  $q1/q2$  függvényében, ha  $K=30$



6.8. ábra  $R(p,q)$ ,  $R_a(p)$  ábrázolása az átskálázott  $N(p,q)$  függvényében, ha  $K=30$

Látható, hogy  $K=30$  esetén az a pontok illeszkednek görbéhez.

## 7 Összegzés és további lehetőségek

Dolgozatomban a Zingg-féle osztályozási rendszer segítségével ( $a \geq b \geq c$ ,  $\frac{b}{a} = x$ ,  $\frac{c}{a} = y$ ) megmutattam, hogy adott kavicspopuláció tengelyarányait egy Descartes-féle koordináta rendszerben vizsgálva a pontokból több egyenes kirajzolódik. Ezek az egyenesek természetes módon jelen vannak, ha egy kavicspopulációt Zingg-féle osztályozás szerint ábrázolunk. Ha ezek az egyenesek nincsenek jelen, akkor lehet okunk gyanakodni arra, hogy a vizsgált adathalmaz manipulált.

Megmutattam, hogy az előbb említett egyenesek láthatósága annál jobb, minél kisebb a  $(p_1, p_2)$  vektor normája, ahol  $\frac{p_1}{p_2}$  a vizsgált egyenes meredeksége és  $p_1, p_2$  számok relatív prímek.

Az előbbieket demonstrálásához saját készítésű Excel és Matlab programokat használtam eltérő számosságú adathalmazokon. A vizsgálat során több fajta modellt, ezekhez kapcsolódó ábrázolásokat és módszereket használtam mire megtaláltam azt, amivel a legjobban le lehet írni egy egyenes vizuális detektálhatóságát.

Ennek a témának a kutatását tervezem folytatni. Jelen pillanatban két lehetőséget látok erre. A determinisztikus esetben a vizsgálatunk során a függőleges és vízszintes egyenesekkel nem foglalkoztunk, mert ezeket nem lehetett jól jellemezni az általunk használt módszerekkel. Egy jó megoldás lehet erre a koordináta rendszerünk elforgatása. Ezt egy elforgatás mátrixával való szorzással lehet kivitelezni, viszont figyelni kell arra, hogy a forgatás során a koordinátáink racionálisak maradjanak. Erre egy jó megoldás, ha mátrixban szereplő  $\sin(\alpha)$  és  $\cos(\alpha)$  értékek racionális számok. Ezt úgy érhetjük el, ha ezeket pitagoraszi számhármakból képezzük.

A másik lehetőség pedig a véletlen modell vizsgálata. Szeretnénk adni egy becslést arra, hogy a véletlen modell szerint sorsolva hány kísérletet kell elvégeznünk ahhoz, hogy a determinisztikus modellben az egyeneseken lévő pontok 90%-a 90%-os valószínűséggel megjelenjen az ábránkon. Erre egyelőre nincs megoldásunk, azonban, ha csak egy egyenest vizsgálunk a determinisztikus modellben, arra van egy elméletünk, hogy hogyan lehet becsülni véletlen modellben a kísérletek számát.



## Köszönetnyilvánítás

Köszönöm a dolgozatom elkészítésében nyújtott elméleti és programozási segítséget Dr. Sipos András Árpádnak, valamint az elméleti segítséget és támogatást Dr. Domokos Gábornak.

## Irodalomjegyzék

- [1] Zingg T.. 1935 Beitrag zur Schotteranalyse. *Mineral. Petrol. Mitte.* **15**, 39–140.
- [2] Sneed E., Folk R. L. (1958) Pebbles in the lower Colorado River, Texas, a study in particle morphogenesis. *Journal of Geology* **66**, 114–150
- [3] Carr A. P.. 1969 Size grading along a pebble beach: Chesil beach, England. *J. Sedimentary Petrol.* **39**, 297–311. doi:10.1306/74D71C3A-2B21-11D7-8648000102C1865D (doi:10.1306/74D71C3A-2B21-11D7-8648000102C1865D).
- [4] G. Domokos and G. W. Gibbons: 2012, The evolution of pebble size and shape in space and time, The Royal Society Publishing, <https://doi.org/10.1098/rspa.2011.0562>

## Ábrajegyzék

Ábra sorszáma	Forrás, saját készítésű kép esetén felhasznált program
1.1.	Excel
3.1.	Szabó Tímea, Domokos Gábor: <a href="http://doktori.bme.hu/bme_palyazat/2011/hallgato/szabo_timea_en.htm">http://doktori.bme.hu/bme_palyazat/2011/hallgato/szabo_timea_en.htm</a> letöltve: 2021.10.2
3.2.	<a href="https://www.researchgate.net/figure/Particle-shape-Sneed-and-Folk-triangular-diagram-Letters-a-b-and-c-represent-the_fig3_242136059">https://www.researchgate.net/figure/Particle-shape-Sneed-and-Folk-triangular-diagram-Letters-a-b-and-c-represent-the_fig3_242136059</a> letöltve:2021.10.26
4.1.	Excel és Matlab
4.2.	Matlab
6.1.	Matlab
6.2.	Matlab
6.3.	Matlab
6.4.	Matlab
6.5.	Matlab
6.6.	Matlab
6.7.	Matlab
6.8.	Matlab