Eötvös Loránd Tudományegyetem
Matematika Intézet

Indruck Balázs

Pszeudovéletlen sorozatok és Legendre-összegek becslése

BSc szakdolgozat

Témavezetők:
Gyarmati Katalin, egyetemi docens
Sárközy András, professor emeritus

ELTE Algebra és Számelmélet Tanszék

Budapest 2015.

Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőimnek, Gyarmati Katalin tanárnőnek és Sárközy András tanár úrnak a témajavaslatért, valamint a sok tanácsért, észrevételért és segítségért, amit a dolgozat írása közben kaptam.

Hálás köszönettel tartozom a családomnak és barátaimnak, akik az évek során mellettem álltak és mindenben támogattak. Nélkülük ez a szakdolgozat nem jöhetett volna létre.

Tartalomjegyzék

1. Bevezetés

A véletlen fogalma már ősi idők óta foglalkoztatja az embereket. Elég csak a jóslásra vagy a szerencsejátékokra gondolni, de kezdve a sportfogadástól egészen a pénzügyi folyamatokig rengeteg helyen megjelenik a véletlen az életünkben. Nem meglepő tehát, hogy a véletlenség matematikai leírása igen fontos téma.

David Hilbert 1900-ban megjelent híres 23 problémája között szerepel a valószínűségelmélet axiomatizálása. Ma általánosan *A. N. Kolmogorov* axiómarendszere az elfogadott.

A XX. és XXI. században a számítógépek elterjedésével előtérbe kerültek véletlen számok, számsorozatok előállításának problémái. Például:

1. Amikor természeti jelenségeket akarunk vizsgálni számítógépes szimuláció segítségével, akkor elengedhetetlen, hogy véletlen számokat használjunk. Ilyenek például a részecskefizikában különböző részecskék ütközéseinek modellezése; az operációkutatásban, ha azt szeretnénk modellezni, hogy az emberek véletlen időközönként érkeznek meg egy buszállomásra; a populációdinamikában az egyedek szaporodásának és halálozásának szimulációjára.

2. Véletlen adatok jó tesztelési lehetőséget biztosítanak, ha egy algoritmust akarunk tesztelni helyesség vagy hatékonyság szempontjából. Továbbá igen fontosak a véletlen algoritmusok, amikkel sokszor jobb átlagos futásidőt érhetünk el, mint egy determinisztikussal. Érdekes alkalmazás még a számítógépes grafikában különböző mintázatok, textúrák létrehozása is.

3. Az internet térnyerésével egyre fontosabbak az olyan kódolási technikák, amikkel a jelszavainkat védhetjük meg az illetéktelenektől. A bonyolult kódolási eljárásokhoz pedig elengedhetetlen, hogy véletlen számokat tudjunk előállítani.

Igény merült fel, hogy az említett sorozatokat el tudjuk készíteni. Ehhez azonban új módszereket kellett találni és azt is le kellett írni valahogy, hogy ezek a sorozatok „mennyire is véletlenek”.

A szakdolgozatban megadjuk bináris sorozatok pszeudovéletlenségének fogalmát,

majd a kapott eredményekből kiindulva a Legendre-szimbólum segítségével elkészített pszeudorandom sorozatokat vizsgálunk.

A dolgozat felépítése a következő:

1. Ez a jelen bevezetés, ahol motivációt adtunk a vizsgálataink tárgyához.

2. Ebben fejezetben egy rövid történeti áttekintés után bevezetjük a pszeudovéletlenség intuitív fogalmát, majd leírjuk, hogy miért is hasznosak a pszeudovéletlen sorozatok.

3. Ebben a szakaszban bevezetjük bináris sorozatok pszeudovéletlen mértékeit, melyekkel kielégítő definíciót kapunk a sorozatok véletlenségére.

4. A negyedik fejezetben kitérőt teszünk a karakterek világába, hogy aztán az itt kapott eredményeket fel tudjuk használni a szakdolgozat további részeiben.

5. Az ötödik fejezetben megmutatjuk, hogy a Legendre-szimbólum segítségével elkészített bináris sorozatok a mértékek értelmében „jó" pszeudovéletlen sorozatokat alkotnak.

6. A hatodik szakaszban további vizsgálatoknak vetjük alá a Legendre-szimbólummal generált sorozatokat. A véges intervallumokon vett összegektől indulva eljutunk a Pólya–Vinogradov egyenlőtlenségig, majd ismertetjük a legújabb eredményeket az egyenlőtlenségben szereplő konstanssal kapcsolatban. Ezután összehasonlítjuk a konstansra adott becsléseket a valósággal, amit a *Sage* programcsomag segítségével teszünk meg.

7. Az utolsó fejezetben bővítjük az egyenletes eloszlás mértékét egy bizonyos α-súlyozással, majd a mérték valós értékeit vizsgáljuk ugyancsak a *Sage* segítségével.

A dolgozatban használt néhány jelölés és definíció

Legyen f és g két, természetes számokon értelmezett, valós értékű függvény. Ekkor azt írjuk, hogy

$$f = O(g),$$

ha létezik olyan $c > 0$ konstans és olyan $n_0 \in \mathbb{Z}_+$ küszöbszám, hogy minden $n > n_0$ esetén $|f(n)| \leq c|g(n)|$. Ezt néha úgy is fogjuk jelölni (a számelméletben használatosan), hogy $f \ll g$. Azt írjuk, hogy

$$f = o(g),$$

ha $f(n)/g(n) \to 0$, amint $n \to \infty$.

Az *Euler-konstans* értéke:

$$\gamma \approx 0.5772.$$

2. Pszeudovéletlen sorozatok

A számítógépen megjelenő számokat kettes számrendszerben ábrázoljuk, így alkalmazásainkhoz szükséges, hogy elő tudjunk állítani véletlen biteket, bitsorozatokat. Hogyan kaphatunk ilyen biteket? Régen, ha valakinek véletlen számokra volt szüksége, például számozott golyókat húzhatott ki egy urnából, esetleg kártyák közül oszthatott lapokat. Valódi véletlen számokat kaphatunk a radioaktív bomlásból.

Korábban sok ötlet született, hogy milyen gépeket is lehetne készíteni, amivel véletlen számokat generálhatunk. Ilyen volt például az 1951-ben megépített *Ferranti Mark I* számítógép, ami 20 véletlen bitet állított elő egy zaj-generátor segítségével. Az ötletet *A. M. Turing* javasolta.

Nem sokkal később, a számítógépek megjelenésével egyre hatékonyabb módszereket próbáltak keresni, hogyan is lehetne véletlen számsorozatokat előállítani. Egy előre elkészített táblázatot is lehetett használni, ez azonban memóriafelhasználás szempontjából nem volt előnyös megoldás. *Neumann Jánostól* származik az a javaslat, hogy használjuk a számítógép általános aritmetikai műveleteit véletlen számok előállításához.

Az ötlete, a „középső négyzet módszer" a következőképpen működik: ha például 10-jegyű számokat akarunk előállítani és az előző érték 5772156649 volt, akkor emeljük ezt négyzetre, hogy a következő számot kapjuk:

$$33317792380594909201,$$

majd ennek vegyük a középső tíz jegyét: 7923805949. Ez lesz a sorozat következő eleme. Hatékonyság szempontjából azonban ezt az algoritmust nem nevezhetjük túl jónak. A legtöbb esetben a csupa 0 sorozatba megy át vagy túl kicsi periódushosszt kapunk, mint ahogy az a következő példából is látható:

2.1. Példa. A Neumann-generátor kétjegyű tízes számrendszerbeli számokra futtatva:

Periódus hossza:	Hányszor szerepel:
1	7822
3	44
6	234

A 7822 darab 1 hosszúságú periódusból ráadásul 7449 megy át a csupa 0 sorozatba. (A táblázatban szereplő értékek kiszámolásához készítettem egy rövid C++ nyelven írt példaprogramot, amely megtalálható a szakdolgozat mellékleteként.)

A fenti példa több problémára is rávilágít. Hogyan tudjuk eldönteni egy véletlen számokat előállító algoritmusról, hogy az mennyire hatékony? Hogyan dönthető el egy számsorozatról, hogy véletlen számsorozat-e? Sőt, ha tekintjük a fenti algoritmust, akkor az általa generált sorozatról hogyan tudjuk azt állítani, hogy „véletlen", amikor az előző elem ismeretében ki tudjuk számolni a következőt? Egy ilyen sorozat nem véletlen, csak annak tűnik egy kívülálló számára, így adhatunk egy intuitív „definíciót" is a pszeudovéletlenségre: egy sorozatot *pszeudovéletlennek* nevezünk, ha egy determinisztikus algoritmus generálja, de egy kívülálló még számítógépek segítségével sem tudja megkülönböztetni egy fizikai módszerrel előállított „véletlen" sorozattól.

Ez persze nem egy pontos matematikai definíció, hiszen mi az, hogy „véletlennek tűnik"? A szakdolgozat további részében ezt próbáljuk megválaszolni.

Mielőtt azonban erre rátérnénk, megemlítjük, hogy miért is előnyösek a pszeudovéletlen számsorozatok. Gyakran meg akarunk ismételni valamilyen számítást, például ha egy számítógépes programot akarunk tesztelni helyesség szempontjából véletlen értékekkel. Ha a számok forrása valódi véletlen volt, akkor el kell ezeket tárolni, ami sok helyet igényel. Egy pszeudovéletlen sorozat előállításánál viszont elég csak a kezdeti értéket, értékeket eltárolni, amikkel később pontosan ugyanazt a számítást tudjuk elvégezni a determinisztikusság miatt.

3. Pszeudovéletlen sorozatok mértékei

A pszeudovéletlenség egyik klasszikus megközelítése bonyolultságelméleti, de ez a megközelítés nem egy adott sorozat pszeudovéletlen tulajdonságait minősíti, hanem bináris sorozatokat generáló algoritmusokét, tehát egy adott sorozatról semmit nem mond. Továbbá ebben a megközelítésben általában végtelen sorozatokat vizsgálnak, míg az alkalmazásainkban többnyire véges sorozatokra van szükség, valamint sok eredmény bizonyítatlan feltevéseken alapszik, mint például az egész számok prímfaktorizációjának nehézsége.

Ahhoz, hogy pszeudovéletlen sorozatokat tudjunk jellemezni a bonyolultságelmélet nélkül, újabb eszközöket kell bevezetnünk. Ezek az eszközök lesznek segítségünkre abban, hogy minőségileg jellemezni tudjunk egy véges sorozatot a véletlenség szempontjából. Bár bináris sorozatokkal foglalkozunk, ezek elemei most nem a 0 és 1 számok közül kerülnek ki (technikai okok miatt), hanem a -1 és $+1$ számok közül. (Ez nem jelent problémát, mert ezen sorozatok között kölcsönösen egyértelmű megfeleltetés létesíthető.)

Christian Mauduit és Sárközy András [3]-ban véges pszeudovéletlen sorozatok új mértékeit vezették be. Bár már korábban léteztek bizonyos statisztikai tesztek, melyekkel sorozatokat lehetett vizsgálni, ezek csak arra szolgáltak, hogy eldöntsük, egy sorozat átmegy-e a teszten vagy sem. További hátrány volt, hogy ezek „a posteriori" módon működtek, vagyis egy sorozatot csak a generálás után lehetett velük tesztelni. Ez azért is probléma, mert ellenőrizni, hogy egy sorozat átmegy-e a teszten, elég sok időt vehet igénybe. Ugyanakkor egy ilyen teszt mindig egyetlen konkrét pszeudovéletlen tulajdonságot vizsgál, míg a többi tulajdonsággal nem foglalkozik.

Az új mértékek használatával viszont lehetőség nyílik „a priori" tesztelésre is, vagyis egy sorozatról megmondhatjuk, hogy mindig teljesíti-e a feltételeket bizonyos elméleti tételek alapján, melyek a konstrukcióból származnak. Továbbá, egyszerre több pszeudovéletlen tulajdonságot is vizsgálni tudunk és így minőségileg is mérhetjük a sorozatokat.

A mértékektől elvárjuk, hogy a legfontosabb véletlen tulajdonságok jellemzésére legyenek alkalmasak, mint a következők:

1. normalitás

2. egyenletes eloszlás a számtani sorozatok mentén

3. kis rendű többszörös korreláció

3.1. *Megjegyzés.* A 2. számú tulajdonság azt próbálja megragadni, hogy milyen közeli a gyakorisága a $+1$ és -1 számoknak a számtani sorozatok mentén. Egy véletlen sorozattól azt várhatjuk el, hogy ezek nagyon közel legyenek egymáshoz. A másik két tulajdonsággal pedig azt próbáljuk leírni, hogy bizonyos tagok, részsorozatok között mi a kapcsolat. Például, ha a $(+1,+1)$ részsorozat sokkal gyakrabban szerepel, mint a $(-1,-1)$ részsorozat, akkor nyilvánvalóan nem mondhatjuk, hogy sorozatunk erős pszeudovéletlen tulajdonságokkal rendelkezik, sőt, ez még az alkalmazásainkban is problémát jelenthet (egy kódot könnyebb feltörni, stb.).

Legyen először $E = (e_1, e_2, \ldots) \in \{-1,+1\}^\infty$ egy végtelen bináris sorozat és értelmezzük az alábbi függvényeket. A jelölések Mauduittól és Sárközytől származnak [3].

3.1. Definíció (Mauduit–Sárközy)**.** $k \in \mathbb{N}$, $M \in \mathbb{N}$, $X = (x_1, \ldots, x_k) \in \{-1,+1\}^k$, $a \in \mathbb{Z}$, $b \in \mathbb{N}$, $D = (d_1, \ldots, d_k) \in \mathbb{N}^k$, $d_1 < \ldots < d_k$ esetén legyen

$$T(E,M,X) = |\{n : 0 \le n < M,\ (e_{n+1}, e_{n+2}, \ldots, e_{n+k}) = X\}|, \tag{3.1}$$

$$U(E,M,a,b) = \sum_{j=1}^{M} e_{a+jb}, \tag{3.2}$$

$$V(E,M,D) = \sum_{n=0}^{M-1} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k}. \tag{3.3}$$

Ezen függvények segítségével a fenti 1–3 tulajdonságokat a következőképpen definiáljuk:

3.2. Definíció (Mauduit–Sárközy)**.** Az E-ről azt mondjuk, hogy *normális*, ha

$$|T(E,M,X) - M/2^k| = o(M),$$

minden fix k-ra és X-re, amint $M \to \infty$.

A 2. és 3. tulajdonság teljesülését az jelenti, ha:

$$U(E, M, a, b) = o(M)$$

és

$$V(E, M, D) = o(M),$$

minden fix a-ra, b-re és D-re, amint $M \to \infty$.

Mivel az alkalmazásokban (főként a számítógépes számábrázolás végessége miatt) legtöbbször *véges* bináris sorozatokra van szükségünk, így szükséges a fenti fogalmak analóg jellemzése erre az esetre. Ehhez tekintsük az alábbi definíciót, amely Knuthtól származik:

3.3. Definíció (Knuth)**.** Egy véges $E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N$ sorozatról akkor mondjuk, hogy pszeudovéletlen, ha minden $k \in \mathbb{N}$ esetén, amire

$$k \leq \frac{\log N}{\log 2}, \tag{3.4}$$

és minden $X \in \{-1, +1\}^k$-ra

$$\left| T(E_N, N + 1 - k, X) - \frac{N + 1 - k}{2^k} \right| \leq \sqrt{N}. \tag{3.5}$$

Itt (3.5)-ben T-t a várható értékével hasonlítjuk össze. Ezen definíció értelmében azonban egy sorozatról csak két dolgot mondhatunk: a sorozat vagy véletlen vagy nem. Emiatt egy kicsit rugalmasabb mérőszámot szeretnénk bevezetni. Például egy sorozatot akkor is szeretnénk „jónak" tekinteni, ha nem elégíti ki a (3.5) feltételt, de kielégíti mondjuk $2\sqrt{N}$-es gyengébb esetre, miközben bizonyos más véletlen tulajdonságok érvényben maradnak, melyek az alkalmazásunkban különösen fontosak. Így megköveteljük az alábbiakat is:

4. A pszeudovéletlen bináris sorozatok jellemzésére használt mérték egy valós értékű függvény legyen, ami az összes véges bináris sorozatok halmazán van értelmezve, így bármely két, azonos hosszúságú sorozatot össze tudunk hasonlítani.

5. Ez a mérték legalább bizonyos „szép” sorozatokra könnyen meghatározható legyen.

6. Mivel egy univerzális mértéket lehetetlen megadni, ami minden sorozatra jól működne az alkalmazásoktól függően, ezért a mértéknek legyenek különböző rendjei és legalább az alacsony rendű értékeket könnyen meg lehessen határozni.

Legyen tehát $E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N$ egy véges sorozat és tekintsük a következő definíciókat:

3.4. Definíció (Mauduit–Sárközy). k-adrendű normalitás mérték:

$$N_k(E_N) = \max_{X \in \{-1,+1\}^k} \max_{0 < M \le N+1-k} |T(E_N, M, X) - M/2^k|.$$

3.5. Definíció (Mauduit–Sárközy). Normalitás mértéke:

$$N(E_N) = \max_{k \le (\log N)/\log 2} N_k(E_N).$$

3.6. Definíció (Mauduit–Sárközy). Egyenletes eloszlás mértéke:

$$W(E_N) = \max_{a,b,t} |U(E_N, t, a, b)| = \max_{a,b,t} \left| \sum_{j=1}^{t} e_{a+jb} \right|,$$

ahol a maximumot minden olyan a-ra, b-re és t-re határozzuk meg, ahol $a \in \mathbb{Z}$, $b, t \in \mathbb{N}$ és $1 \le a + b \le a + tb \le N$.

3.7. Definíció (Mauduit–Sárközy). k-adrendű korreláció mérték:

$$C_k(E_N) = \max_{M,D} |V(E_N, M, D)| = \max_{M,D} \left| \sum_{n=0}^{M-1} e_{n+d_1} e_{n+d_2} \ldots e_{n+d_k} \right|,$$

ahol a maximumot minden olyan $D = (d_1, \ldots, d_k)$-ra és M-re tekintjük, ahol $M - 1 + d_k \le N$.

3.8. Definíció (Mauduit–Sárközy). Korreláció mértéke:

$$C(E_N) = \max_{k \le (\log N)/\log 2} C_k(E_N).$$

(Egy másik lehetőség lehetne a $C^*(E_N) = \sum_{k=1}^{\infty} C_k(E_N)/2^k$.)

A fenti mértékek kapcsolatáról a következőket mondhatjuk (ezekből bővebben [3]-ban lehet olvasni):

3.1. Állítás. *Minden N-re, E_N-re és $k < N$-re*

$$N_k(E_N) \le \max_{1\le t\le k} |C_t(E_N)|. \tag{3.6}$$

Bizonyítás. Minden $k, N \in \mathbb{N}$, $X = (x_1, \dots, x_k) \in \{-1, +1\}^k$ és $1 \le M \le N+1-k$ esetén

$$\begin{aligned}
&|T(E_N, M, X) - M/2^k| = \\
&= \left| |\{n : 0 \le n < M, (e_{n+1}, e_{n+2}, \dots, e_{n+k}) = X\}| - \frac{M}{2^k} \right| = \\
&= \left| \sum_{n=0}^{M-1} \frac{x_1 \dots x_k}{2^k} \prod_{j=1}^{k} (e_{n+j} + x_j) - \frac{M}{2^k} \right| = \\
&= \left| \frac{x_1 \dots x_k}{2^k} \sum_{1\le d_1<\dots<d_t\le k} \left(\prod_{j\in\{1,\dots,k\}\setminus\{d_1,\dots,d_t\}} x_j \right) \sum_{n=0}^{M-1} e_{n+d_1} \dots e_{n+d_t} \right| \le \\
&\le \frac{1}{2^k} \sum_{\emptyset\ne D\subset\{1,2,\dots,k\}} |V(E_N, M, D)| \le \frac{1}{2^k} \sum_{t=1}^{k} \binom{k}{t} C_t(E_N) \le \\
&\le \max_{1\le t\le k} |C_t(E_N)|.
\end{aligned}$$

□

A korreláció mértéke és az egyenletes eloszlás mértéke közötti kapcsolatról pedig a következőt lehet mondani [16] alapján:

3.1. Tétel (Mauduit–Sárközy)**.** *Minden $N \in \mathbb{N}$ és $E_N = (e_1, \dots, e_N) \in \{-1, +1\}^N$ sorozat esetén*

$$W(E_N) \le 3(NC_2(E_N))^{1/2}.$$

Továbbá, a fenti egyenlőtlenség konstansszorostól eltekintve éles (lásd [16]-ban a 2. Tételt).

A fordított irányban azonban nem mondhatunk semmit. Előfordulhat, hogy mind a normalitás mértéke és az egyenletes eloszlás mértéke kicsi, azonban a korreláció mértéke nagyon nagy, mint ahogy azt a következő példa is mutatja:

3.1. Példa. Legyen $E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N$ olyan sorozat, melyről feltesszük, hogy a normalitás mértéke és az egyenletes eloszlás mértéke is meglehetősen kicsi és legyen $E'_{2N} = (e'_1, e'_2, \ldots, e'_{2N}) \in \{-1, +1\}^{2N}$ olyan, hogy

$$e'_n = \begin{cases} e_n & \text{ha} \quad 1 \le n \le N, \\ e_{n-N} & \text{ha} \quad N < n \le 2N. \end{cases}$$

(Tehát az E_N sorozatot megismételjük egymás után kétszer.)

Ekkor be lehet látni, hogy E'_{2N} normalitás mértéke és egyenletes eloszlás mértéke legfeljebb valamilyen konstansszorosa E_N ezen mértékeinek, de

$$C_2(E'_N) \ge \left| \sum_{n=1}^{N} e'_n e'_{n+N} \right| = N.$$

A fentiekből következik, hogy ahhoz, hogy egy véges bináris sorozatról azt mondhassuk, hogy pszeudovéletlen, (vagyis, hogy rendelkezik az 1–3 tulajdonságokkal), elegendő megmutatni, hogy mind az egyenletes eloszlás mértéke, mind a korreláció mértéke „kicsi" abban az értelemben, hogy $W(E_N)$ és $C_k(E_N)$ (legalább kis k-ra) $o(N)$ nagyságrendű, amint $N \to \infty$. Továbbá, ezeket mindenképpen ellenőrizni is kell. A két mértéket egyesíteni lehet, így egy „kombinált pszeudovéletlen mértéket" kapunk:

3.9. Definíció (Mauduit–Sárközy)**.** k-adrendű kombinált mérték:

$$Q_k(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \ldots e_{a+jb+d_k} \right| = \max_{a,b,t,D} |Z(a,b,t,D)|, \tag{3.7}$$

ahol

$$Z(a,b,t,D) = \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \ldots e_{a+jb+d_k}, \tag{3.8}$$

amit minden olyan a-ra, b-re, t-re és $D = (d_1, d_2, \ldots, d_k)$-ra definiálunk, ahol minden $a + jb + d_l$ részindex eleme az $\{1, \ldots, N\}$ halmaznak és a maximumot (3.7)-ben minden, a fentieknek megfelelő D k-dimenziós vektorra nézzük. (Más szavakkal $Q_k(E_N)$ a k-rendű korrelációt méri számtani sorozatok mentén.)

3.10. Definíció (Mauduit–Sárközy)**.** Kombinált pszeudovéletlen mérték:

$$Q(E_N) = \max_{k \le (\log N)/\log 2} Q_k(E_N), \tag{3.9}$$

vagy hasonlóan egy másik lehetőség, mint a korrelációs mértéknél:

$$Q^*(E_N) = \sum_{k=1}^{\infty} Q_k(E_N)/2^k.$$

Nézzük meg ezek alkalmazását a következő példán:

3.2. Példa. Tekintsük az $E_N = (e_1, \ldots, e_N) \in \{-1, +1\}^N$ sorozatot, amiről feltesszük, hogy mind a korrelációs mértéke, mind az egyenletes eloszlás mértéke (vagy akár a kombinált mértéke) kicsi és legyen $E'_{2N} = (e'_1, \ldots, e'_{2N}) \in \{-1, +1\}^{2N}$ olyan, hogy

$$e'_n = \begin{cases} e_n & \text{ha} \quad 1 \le n \le N, \\ e_{2N-n+1} & \text{ha} \quad N < n \le 2N. \end{cases}$$

(Tehát az E_N sorozat mögé a saját megfordítottját ragasztjuk.)

Ekkor belátható, hogy E'_{2N} korrelációs mértéke és egyenletes eloszlás mértéke legfeljebb E_N megfelelő mértékeinek konstansszorosa (tehát a mértékeink szerint a sorozatot pszeudovéletlennek kell tekintenünk), azonban világos, hogy egy „igazi" véletlen sorozat nem lehet annyira szimmetrikus, mint E'_{2N}.

A fenti példából látszik, hogy nincs tökéletes univerzális pszeudovéletlen mérték. Ha szeretnénk, egyre több és több feltételt vezethetünk be a véletlenség eldöntésére (amire az alkalmazásainkban szükség is lehet, pl. szimmetria mértéke [5]), azonban lehetséges, hogy az új mértékeket bonyolultabb lesz kezelni. Sőt, akár az is előfordulhat, hogy a megkövetelt feltételeket egyetlen (adott hosszúságú) sorozat sem elégíti ki. Így valahol meg kell húzni egy határt, hogy aztán a bevezetett mértékeinkkel dolgozni tudjunk. Például vizsgálatainkat leszűkíthetjük arra az esetre, ha kizárólag az első három pszeudovéletlen tulajdonságot követeljük meg, azaz a normalitást, egyenletes eloszlást és a kis rendű korrelációt.

4. Karakterek

A szakdolgozat későbbi tételeihez szükséges bevezetni néhány klasszikus definíciót a karakterek elméletéből (ld. [8], [9]):

4.1. Definíció. Egy $\chi : \mathbb{Z} \to \mathbb{C}$ függvényt *Dirichlet-karakternek* nevezünk, ha teljesülnek rá az alábbiak:

1. $\exists k > 0$ egész, hogy $\chi(n) = \chi(n+k)$ minden $n \in \mathbb{Z}$-re.
2. $\chi(n) \neq 0$ akkor és csak akkor, ha $(n, k) = 1$.
3. $\chi(mn) = \chi(m)\chi(n)$ minden $n, m \in \mathbb{Z}$-re.

A 1-es számú tulajdonság szerint χ periodikus k periódussal. Ezt a k számot a karakter *modulusának* hívjuk. Ez azzal ekvivalens, hogy ha $a \equiv b \pmod{k}$, akkor $\chi(a) = \chi(b)$. Ha $k = 1$, akkor a karakter a *triviális karakter*.

4.2. Definíció. Egy karaktert *főkarakternek* hívunk, ha a modulushoz relatív prím helyeken 1-et vesz fel, egyébként pedig 0-t. A főkarakter jelölésére a χ_1-et fogjuk használni.

A χ karaktert valósnak nevezzük, ha csak valós értékeket vesz fel, egyébként pedig komplexnek.

A 3-as számú tulajdonságból következik, hogy $\chi(1) = \chi(1 \cdot 1) = \chi(1)\chi(1)$. Mivel $(1, k) = 1$, ezért $\chi(1) \neq 0$, vagyis $\chi(1) = 1$. Ha $(a, k) = 1$, akkor az Euler-Fermat-tétel szerint $a^{\varphi(k)} \equiv 1 \pmod{k}$. Így $\chi(a^{\varphi(k)}) = \chi(1) = 1$ és a multiplikatív tulajdonság miatt $\chi(a^{\varphi(k)}) = \chi(a)^{\varphi(k)}$. Vagyis minden a-ra, amire $(a, k) = 1$, teljesül, hogy a egy $\varphi(k)$-adik komplex egységgyök.

4.3. Definíció. A χ karakter *rendjének* azt a legkisebb pozitív egész d számot nevezzük, melyre minden $(a, k) = 1$ esetén teljesül, hogy $\chi(a)^d = 1$.

Megjegyezzük, hogy a fent tárgyaltak következtében egy karakternek mindig létezik a rendje, hiszen $(a, k) = 1$ esetén $\chi(a)^{\varphi(k)} = 1$ és így a rend legfeljebb akkora, mint $\varphi(k)$.

Egy karakter előjelét a -1 helyen felvett értéke határozza meg. Ha $\chi(-1) = -1$, akkor a karakter páratlan, ha $\chi(-1) = 1$, akkor a karakter páros.

4.4. Definíció. Legyen χ egy Dirichlet-karakter a k modulus szerint és legyen d tetszőleges pozitív osztója k-nak. A d számot *indukált modulusnak* nevezzük a χ-re nézve, ha $\chi(a) = 1$ minden olyan a-ra, amire $(a, k) = 1$ és $a \equiv 1 \pmod{d}$.

Megfigyelhető, hogy k maga mindig indukált modulus.

4.1. Tétel. *Legyen χ egy Dirichlet-karakter* mod k*. Ekkor az* 1 *szám indukált modulus χ-re nézve akkor és csak akkor, ha $\chi = \chi_1$.*

Bizonyítás. Ha $\chi = \chi_1$, akkor $\chi(a) = 1$ minden olyan a-ra, ami relatív prím k-hoz. De mivel minden a-ra $a \equiv 1 \pmod{1}$, ezért az 1 indukált modulus.

Megfordítva, ha az 1 indukált modulus, akkor $\chi(a) = 1$, ha $(a, k) = 1$. Emiatt viszont $\chi = \chi_1$, hiszen $\chi = 0$ minden olyan helyen, ami nem relatív prím k-hoz. □

Mint említettük, bármely χ Dirichlet-karakterre mod k a k indukált modulus. Ha a k-n kívül nincs más indukált modulus, akkor a karakter *primitív*.

4.5. Definíció. Egy χ Dirichlet-karakterről (modulo k) azt mondjuk, hogy primitív, ha nincs $d < k$ indukált modulusa. Más szavakkal, χ primitív mod k akkor és csak akkor, ha minden $0 < d < k$, $d \mid k$ esetén létezik olyan $a \equiv 1 \pmod{d}$, $(a, k) = 1$, hogy $\chi(a) \neq 1$.

4.1. Példa. Ha $k > 1$, akkor a χ_1 főkarakter nem primitív, mert az 1 indukált modulusa.

4.2. Tétel. *Legyen p prímszám. Minden χ Dirichlet-karakterre* mod p*, amely nem főkarakter, teljesül, hogy χ primitív* mod p*.*

Bizonyítás. Mivel p osztói csak az 1 és a p, így ezekre kell ellenőrizni, hogy indukált modulusok-e. De ha $\chi \neq \chi_1$, akkor az 1 osztó nem indukált modulus, így χ-nek nincs olyan indukált modulusa, ami kisebb lenne, mint p. Tehát χ primitív. □

A Legendre-szimbólumot a megszokott módon értelmezzük:

4.6. Definíció (Legendre-szimbólum)**.** $\left(\frac{n}{p}\right) = 1$, ha n kvadratikus maradék mod p és $\left(\frac{n}{p}\right) = -1$, ha n kvadratikus nemmaradék mod p.

4.1. Állítás. *Legyen $p > 2$ prímszám, valamint definiáljuk $\chi(n)$-et a Legendre-szimbólummal, azaz*

$$\chi(n) = \begin{cases} \left(\frac{n}{p}\right) & \text{ha} \quad (n,p) = 1, \\ 0 & \text{ha} \quad p|n. \end{cases}$$

Ekkor χ primitív karakter és nem főkarakter.

Bizonyítás. Első lépésben belátjuk, hogy $\chi(n) = \left(\frac{n}{p}\right)$ valóban karakter. Ehhez felhasználjuk a számelméletből ismert Euler-lemmát:

$$\left(\frac{n}{p}\right) \equiv n^{\frac{p-1}{2}} \pmod{p}. \tag{4.1}$$

(Ez könnyen látható a Legendre-szimbólum definíciójából, valamint alkalmazva a binom kongruenciák megoldhatóságára vonatkozó tételt speciálisan a másodfokú kongruenciákra.)

Először ellenőrizni kell a p szerinti periodicitást:

$$\chi(n+p) = \left(\frac{n+p}{p}\right) \equiv (n+p)^{\frac{p-1}{2}} = \sum_{k=0}^{\frac{p-1}{2}} \binom{\frac{p-1}{2}}{k} n^k p^{\frac{p-1}{2}-k} \equiv n^{\frac{p-1}{2}} \equiv \left(\frac{n}{p}\right) = \chi(n).$$

Itt a binomiális tétel szerint kifejtve a $k = (p-1)/2$-höz tartozó tag kivételével mindegyikben szerepel a $p \equiv 0 \pmod{p}$. Így tehát $\chi(n+p) \equiv \chi(n)$, viszont χ értékei a $\{0, \pm 1\}$ halmazból kerülhetnek ki és $p > 2$, így $\chi(n+p) = \chi(n)$ is igaz.

Ismét felhasználva a (4.1) azonosságot a $\chi(n) \neq 0 \Leftrightarrow (n,p) = 1$ igaz lesz, hiszen p prím, és így a hatvány csak akkor lehet p-nek többszöröse (vagyis $0 \bmod p$), ha $n \equiv p \equiv 0 \pmod{p}$ teljesül.

Továbbá igaz az is, hogy χ teljesen multiplikatív:

$$\chi(nm) = \left(\frac{nm}{p}\right) \equiv (nm)^{\frac{p-1}{2}} = n^{\frac{p-1}{2}} m^{\frac{p-1}{2}} \equiv \left(\frac{n}{p}\right)\left(\frac{m}{p}\right) = \chi(n)\chi(m).$$

Itt a fent elmondottak szerint ugyancsak írhatunk egyenlőséget: $\chi(nm) = \chi(n)\chi(m)$.

Az nyilvánvalóan igaz, hogy χ nem főkarakter. Ugyanis, mivel pontosan ugyanannyi kvadratikus maradék van egy p modulus szerint, mint nemmaradék, így $\chi(n)$ szükségszerűen felveszi a -1 értéket is.

Már csak az maradt hátra, hogy χ primitív is. De p prím és egy prím modulusú karakter a 4.2 Tétel szerint primitív. □

5. A Legendre-szimbólum pszeudovéletlen tulajdonságai

A következőkben megmutatjuk, hogy a bevezetett kombinált mértékek - Q és Q_k - a gyakorlatban is jól használhatók bizonyos „szép” sorozatok pszeudovéletlenségének tesztelésére. A 3.1 Állítás értelmében elegendő megmutatni, hogy Q, illetve Q_k értékére jó felső korlátot tudunk adni, mert ekkor ez már maga után vonja a normalitást is. Erre a célra a Legendre-szimbólumot fogjuk használni.

5.1. Tétel (Mauduit–Sárközy)**.** *Létezik olyan p_0 küszöbszám, hogy minden $p > p_0$ prímszám és $k \in \mathbb{N}$, $k < p$ esetén, ha*

$$E_{p-1} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots, \left(\frac{p-1}{p}\right)\right),$$

akkor

$$Q_k(E_{p-1}) \leq 9kp^{1/2} \log p, \tag{5.1}$$

így az $N = p - 1$ jelöléssel:

$$Q(E_N) = \max_{k \leq (\log N)/\log 2} Q_k(E_N) \leq 27N^{1/2}(\log N)^2, \tag{5.2}$$

valamint

$$Q^*(E_N) = \sum_{k=1}^{\infty} Q_k(E_N)/2^k \leq 33N^{1/2} \log N. \tag{5.3}$$

Az 5.1 Tétel bizonyításához szükség lesz az alábbi eredményekre. A [3]-ban szereplő 5.1 és 5.2 Lemmát (ld. alább) itt nem látjuk be, de megjegyezzük, hogy ezek A. Weil tételének következményei. Erről bővebben [3]-ban lehet olvasni.

5.2. Tétel (Weil)**.** *Tegyük fel, hogy p egy adott prímszám, χ egy d rendű nem főkarakter modulo p (vagyis $d \mid p-1$) és $f(x) \in \mathbb{F}_p[x]$ egy k-adfokú polinom $\mathbb{F}_p$ felett (ahol $\mathbb{F}_p$ jelöli a modulo p maradékosztályok által alkotott p elemű testet). Továbbá tegyük fel, hogy f nem $cg(x)^d$ alakú, ahol $c \in \mathbb{F}_p$, $g(x) \in \mathbb{F}_p[x]$. Legyenek továbbá X és Y valós számok úgy, hogy $0 < Y \leq p$. Ekkor a következő egyenlőtlenség teljesül:*

$$\left| \sum_{X < n \leq X+Y} \chi(f(n)) \right| < 9kp^{1/2} \log p. \tag{5.4}$$

5.1. Lemma. *Legyenek p, χ, d, $f(x)$ és k úgy definiálva, mint az 5.2 Tételben, valamint legyen $a \in \mathbb{Z}$. Ekkor*

$$\left| \sum_{x \in \mathbb{F}_p} \chi(f(x)) e(ax/p) \right| \leq k p^{1/2}. \tag{5.5}$$

5.2. Lemma. *Tegyük fel, hogy $m \in \mathbb{N}$, $g(x) : \mathbb{Z} \to \mathbb{C}$ periodikus függvény m periódussal, valamint legyenek X és Y valós számok úgy, hogy $Y > 0$. Ekkor*

$$\left| \sum_{X < n \leq X+Y} g(n) \right| \leq \frac{Y+1}{m} \left| \sum_{n=1}^{m} g(n) \right| + \sum_{1 \leq |h| \leq m/2} |h|^{-1} \left| \sum_{n=1}^{m} g(n) e(hn/m) \right|. \tag{5.6}$$

Az 5.2 Tétel bizonyítása. Először alkalmazzuk az 5.2 Lemmát úgy, hogy p-t és $\chi(f(n))$-t helyettesítünk rendre m-be és $g(n)$-be, majd használjuk az 5.1 Lemmát. Ekkor a következőt kapjuk:

$$\begin{aligned}
&\left| \sum_{X < n \leq X+Y} \chi(f(n)) \right| \\
&\qquad \leq \frac{Y+1}{p} \left| \sum_{n=1}^{p} \chi(f(n)) \right| + \sum_{1 \leq |h| \leq p/2} |h|^{-1} \left| \sum_{n=1}^{p} \chi(f(n)) e(hn/p) \right| \\
&\qquad < 2kp^{1/2} + 2 \sum_{1 \leq h \leq p/2} h^{-1} k p^{1/2} \\
&\qquad < 2kp^{1/2}(1 + (1 + \log(p/2))) < 2kp^{1/2}(2 + \log p) \\
&\qquad \leq 2kp^{1/2} \left(2\frac{\log p}{\log 2} + \log p \right) < 9kp^{1/2} \log p.
\end{aligned}$$

□

5.1. Következmény. *Legyen p prím, $f(x) \in \mathbb{F}_p[x]$ k-adfokú polinom, ami nem $b(g(x))^2$ alakú, ahol $b \in \mathbb{F}_p$, $g(x) \in \mathbb{F}_p[x]$, valamint legyenek X és Y valós számok úgy, hogy $0 < Y \leq p$. Ekkor a*

$$\chi_p^*(n) = \begin{cases} \left(\frac{n}{p}\right) & \text{ha} \quad (n,p) = 1, \\ 0 & \text{ha} \quad p|n \end{cases}$$

jelölést használva a következő teljesül:

$$\left| \sum_{X < n \leq X+Y} \chi_p^*(f(n)) \right| < 9kp^{1/2} \log p.$$

Bizonyítás. Ha az 5.2 Tételben $\chi(n) = \chi_p^*(n)$, akkor $d = 2$ és így a tétel akkor és csak akkor igaz, ha $f(x)$ nem $b(g(x))^2$ alakú. □

A fentieket felhasználva már be tudjuk látni az 5.1 Tételt:

Az 5.1 Tétel bizonyítása. Legyen $Z(a, b, t, D)$ a (3.8) szerint definiálva, ahol most $e_n = \left(\frac{n}{p}\right)$. Akkor $k < p$ esetén

$$|Z(a,b,t,D)| = \left|\sum_{n=0}^{t}\left(\frac{a+nb+d_1}{p}\right)\left(\frac{a+nb+d_2}{p}\right)\ldots\left(\frac{a+nb+d_k}{p}\right)\right| \tag{5.7}$$

minden a, b, t, és $D = (d_1, \ldots, d_k)$-ra, amire $d_1 < d_2 < \ldots < d_k$, valamint

$$a + nb + d_l \in \{1, \ldots, p-1\}, \qquad n = 0,1,\ldots,t \quad l = 1,\ldots,k. \tag{5.8}$$

Ekkor az (5.8) feltételei alapján feltehetjük, hogy $(b, p) = 1$. Legyen $\overline{b}$ egy olyan egész szám, amire $b\overline{b} \equiv 1 \pmod{p}$, valamint jelöljön h_j egy olyan egész számot, amire

$$h_j \equiv (a + d_j)\overline{b} \pmod{p},$$

így

$$h_i \not\equiv h_j \pmod{p} \qquad (\forall 1 \le i < j \le k). \tag{5.9}$$

Legyen továbbá $f(n) = (n + h_1)(n + h_2)\ldots(n + h_k)$. Ekkor (5.7)-ből következik, hogy

$$\begin{aligned}
|Z(a,b,t,D)| &= \left|\sum_{n=0}^{t}\left(\frac{a\overline{b}+n+d_1\overline{b}}{p}\right)\left(\frac{a\overline{b}+n+d_2\overline{b}}{p}\right)\ldots\left(\frac{a\overline{b}+n+d_k\overline{b}}{p}\right)\right| \\
&= \left|\sum_{n=0}^{t}\left(\frac{n+h_1}{p}\right)\left(\frac{n+h_2}{p}\right)\ldots\left(\frac{n+h_k}{p}\right)\right| \\
&= \left|\sum_{n=0}^{t}\left(\frac{f(n)}{p}\right)\right| = \left|\sum_{n=0}^{t}\chi_p^*(f(n))\right|,
\end{aligned}$$

ahol a χ_p^*-ot az 5.1 Következményben definiáltuk.

Itt az első lépésben azt használtuk fel, hogy $(b, p) = 1$ és a Legendre-szimbólum multiplikatív tulajdonságából tetszőleges i esetén:

$$\left|\left(\frac{a+nb+d_i}{p}\right)\right| = \left|\left(\frac{(a\overline{b}+n+d_i\overline{b})b}{p}\right)\right| = \left|\left(\frac{a\overline{b}+n+d_i\overline{b}}{p}\right)\left(\frac{b}{p}\right)\right| =$$
$$\left|\left(\frac{a\overline{b}+n+d_i\overline{b}}{p}\right)\right|\left|\left(\frac{b}{p}\right)\right| = \left|\left(\frac{a\overline{b}+n+d_i\overline{b}}{p}\right)\right| \cdot 1.$$

Így felbontás után az összegből $\left(\frac{b}{p}\right)^k$ valóban kiemelhető.

Ha most $X = -1$, $Y = t+1$, akkor feltehetjük, hogy $0 < Y = t+1 \leq N+1 = p$. Mivel $f(x)$-nek nincs többszörös gyöke (hiszen úgy definiáltuk), így alkalmazhatjuk az 5.1 Következményt. Ebből azt kapjuk, hogy

$$|Z(a, b, t, D)| < 9kp^{1/2} \log p,$$

amivel beláttuk az (5.1) egyenlőtlenséget. (5.2) és (5.3) pedig egyszerűen adódik (5.1)-ből, ha Q_k helyébe a kapott felső becslést helyettesítjük. □

A korábban elmondottak értelmében tehát az $E_{p-1} = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots, \left(\frac{p-1}{p}\right)\right)$ szerint definiált sorozat valóban egy jó pszeudovéletlen sorozat, hiszen $cp^{1/2} \log p$ nagyságrendje $o(p)$ (ahol c konstans).

Megjegyezzük még, hogy Sárközy, Mauduit és Goubin [6]-ban olyan algoritmust adtak, amivel „jó” pszeudovéletlen sorozatok nagy családját lehet előállítani a Legendre-szimbólum segítségével. Ehhez nem egyszerűen az $\left(\frac{n}{p}\right)$ alakot használták, hanem az általánosabb $\left(\frac{f(n)}{p}\right)$-t, ahol f egy polinom.

6. Rövid intervallumokon vett összegek

A szakdolgozat hátralevő részében a Legendre-szimbólum segítségével előállított E_N sorozatot vizsgáljuk.

Legyen tehát E_N a következő:

$$E_N = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots, \left(\frac{p-1}{p}\right) \right), \tag{6.1}$$

ahol p egy rögzített prímszám és $N = p - 1$. Tekintsük az alábbi maximumot:

$$\max_{1 \le X < Y \le p-1} \left| \sum_{n=X}^{Y} \left(\frac{n}{p}\right) \right|. \tag{6.2}$$

A célunk, hogy ezt a maximumot a lehető legpontosabban meghatározzuk E_{p-1} függvényében, vagy legalább egy jó felső becslést adjunk rá. Megjegyezzük, hogy (6.2)-t a már korábban bevezetett $U(E, M, a, b)$-vel (lásd (3.2)-t) is kifejezhetjük, ha b helyére a rögzített 1 számot helyettesítjük. Ebből viszont következik, hogy a maximum legfeljebb akkora lehet, mint az egyenletes eloszlás $W(E_N)$ mértéke.

Térjünk át a karakteres jelölésekre. χ segítségével a fenti feladatot a következő formában írhatjuk fel:

$$S_\chi = \max_{1 \le X < Y \le p-1} \left| \sum_{n=X}^{Y} \chi(n) \right|.$$

1918-ban Pólya és Vinogradov egymástól függetlenül belátták, hogy bármely olyan Dirichlet-karakterre, amely nem főkarakter, teljesül a következő egyenlőtlenség.

6.1. Tétel (Pólya–Vinogradov–egyenlőtlenség)**.** *Ha χ nem főkarakter egy q modulussal és $S(\chi)$-t a következő módon definiáljuk:*

$$S(\chi) = \max_{0 \le M < N \le q} \left| \sum_{n=M}^{N} \chi(n) \right|,$$

akkor létezik egy olyan c univerzális konstans, hogy

$$S(\chi) \le c q^{1/2} \log q.$$

Ezen tétel szerint tehát $S_\chi \leq cp^{1/2} \log p$. Mekkora c értéke? A Pólya–Vinogradov egyenlőtlenség széleskörű alkalmazhatósága miatt ez egy igen népszerű és sokat vizsgált kérdés.

6.1. *Megjegyzés.* Megemlítjük például, hogy a konstans javításával jobb becsléseket lehet adni a legkisebb kvadratikus nemmaradék értékére is. Erről bővebben lehet olvasni [12]-ben és [13]-ban.

A legújabb eredmények Pomerance és Frolenkov nevéhez fűződnek. Frolenkov [11]-ben összefoglalta a konstansra adott eddigi becsléseket, majd javította Pomerance 2010-ben adott [10] értékét. Az alábbi egyenlőtlenségek [11]-ből származnak.

A c értékét kétféle alakban szokták kifejezni. Az első alak nem határozza meg explicit módon a maradéktagokat, vagyis ezek leginkább aszimptotikus eredmények. Ugyanakkor ebben az esetben kapunk jobb konstans értékeket a főtagban, mint ahogy az a következő tételben is látszik.

6.2. Tétel (Landau)**.** *Ha $\chi(-1) = 1$, akkor*

$$S(\chi) \leq \left(\frac{1}{\pi\sqrt{2}} + o(1) \right) \sqrt{q} \log q,$$

illetve, ha $\chi(-1) = -1$, akkor

$$S(\chi) \leq \left(\frac{1}{2\pi} + o(1) \right) \sqrt{q} \log q.$$

A másik eset az, amikor minden egyes értéket explicit módon meghatározunk. Ilyen Frolenkov numerikus becslése is.

6.3. Tétel (A. D. Frolenkov)**.** *Legyen χ egy primitív Dirichlet-karakter modulo q, valamint legyen $S(\chi)$ olyan, mint fent. Ekkor a következő igaz:*

1. *Ha $\chi(-1) = 1$ (azaz χ páros karakter), akkor*

$$S(\chi) \leq \frac{2}{\pi^2} \sqrt{q} \log q + \frac{4}{\pi^2} \sqrt{q}(1 + \gamma + \log C_0) + \psi_1(q).$$

2. *Ha $\chi(-1) = -1$ (azaz χ páratlan karakter), akkor*

$$S(\chi) \leq \frac{1}{2\pi} \sqrt{q} \log q + \frac{1}{\pi} \sqrt{q} \left(1 + \gamma + \log \frac{2C_0}{\pi} \right) + \psi_2(q).$$

Itt γ az Euler-konstans, $C_0 = 4\pi^{5/2} + 5$, valamint

$$\psi_1(q) = 1 + \frac{24}{\pi^2 C_0} + \frac{8}{\pi^2} \frac{\sqrt{q}}{\exp\left(\frac{2\sqrt{q}}{C_0}\right) - 1},$$

$$\psi_2(q) = 1 + \frac{3}{C_0} + \frac{2}{\pi} \frac{\sqrt{q}}{\exp\left(\frac{\pi\sqrt{q}}{C_0}\right) - 1}.$$

Amint az látható, ezek a formulák valóban bonyolultabbak, mint az előbb megadott aszimptotikus értékek. Ugyanakkor jó kiindulási alapot adnak ahhoz, hogy például a számítógép segítségével kiszámoljunk konkrét értékeket, majd összehasonlítsuk a valós eredményeket a fenti becslésekkel.

6.2. *Megjegyzés.* Érdemes még megjegyezni szintén [11] alapján a következő két egyenlőtlenséget. Feltételezve, hogy a Riemann-sejtés igaz, Montgomery és Vaughan belátták, hogy

$$S(\chi) \ll \sqrt{q} \log\log q. \tag{6.3}$$

Sőt, valójában ez a legjobb eredmény, ami kapható. Paley bebizonyította, hogy létezik végtelen sok olyan kvadratikus $\chi_n \pmod{q_n}$ karakter, amire

$$S(\chi_n) \gg \sqrt{q_n} \log\log q_n.$$

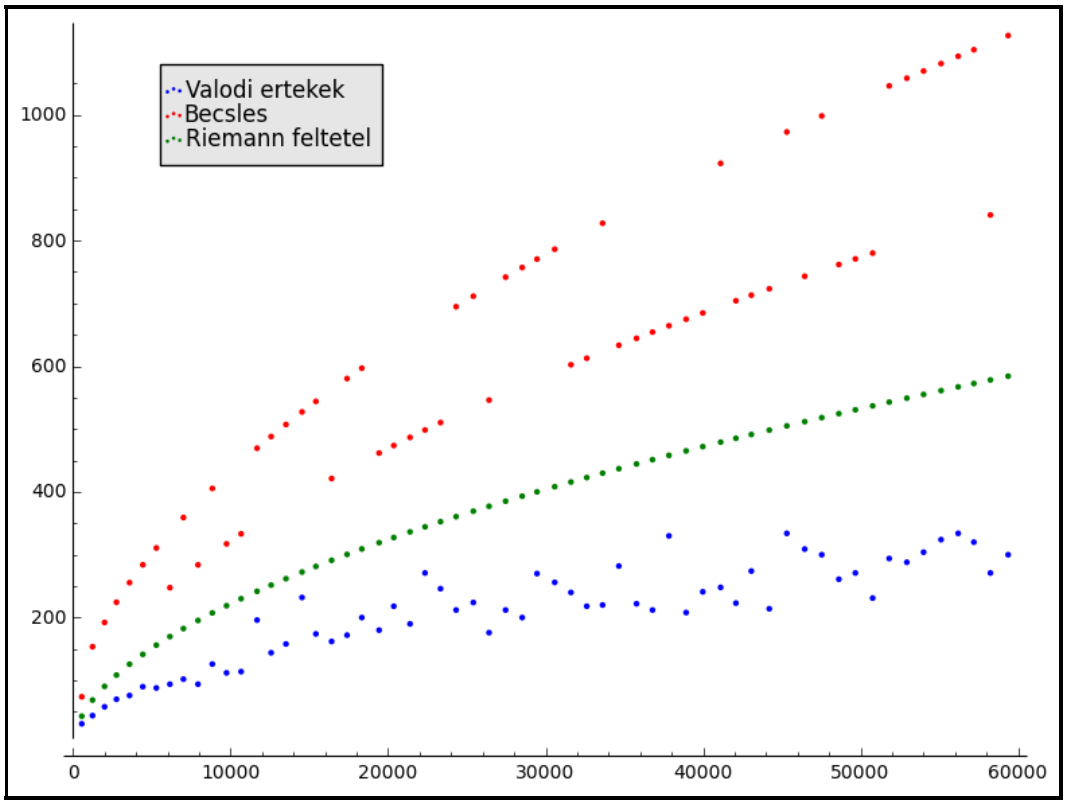
Számolási eredmények

S_χ értékeinek meghatározásához a Sage matematikai programcsomagot használtam. Különböző p prímek esetén elkészítettem a (6.1) sorozatot, majd elvégeztem az összegzést. Az algoritmus és a kiszámolt értékek az *szd_shortsum* nevű munkafüzetben találhatók.

A maximumok meghatározása után definiáltam még két függvényt, melyek segítségével a 6.3 Tétel becsléseit, illetve a (6.3)-ban szereplő $\sqrt{q}\log\log q$-t tudjuk grafikusan ábrázolni. Az alábbi 1. ábrán kék színnel vannak feltüntetve az S_χ értékei, míg rendre piros és zöld színnel a Frolenkov-féle becslés, illetve a Riemann-sejtésből származó $\sqrt{q}\log\log q$ értékei.

A számítások futásidejénél legfeljebb egy órás kereten belül maradtam az idő végessége miatt, így a legnagyobb p prím, amire kiszámoltam a maximumot 59369.

Sajnos a számítógépem nem igazán gyors, így százezres nagyságrendű prímeket már nem tudtam vizsgálni. A p értékeinél mindig körülbelül ezresével léptem előre, illetve $p < 10000$ esetén közelebbi számokat is vettem.



1. ábra. Legendre-szimbólummal elkészített sorozatok véges intervallumokon vett összegeinek maximuma.

Az ábráról azonban jól látható, hogy mindegyik maximum a $\sqrt{q}\log\log q$ grafikonja alatt marad. A 6.3 Tétel függvényei csak kis prímekre adnak közelítőleg éles becslést, azonban nagyobb számokra túl gyorsan növekszik és így a valódi értékektől vett távolság egyre nagyobb lesz. Ugyanakkor érdemes megfigyelni, hogy a maximumok növekedése nagyon lassú. Még 59369 esetén is csak legfeljebb 300 darab egymást követő $+1$ vagy -1 szerepel az elkészített sorozatban, ami p értékének kb. 1/200-ad része. Tehát a legjobb becslések a valódi értékektől még messze állnak, így elképzelhető jövőbeli javítás.

7. Súlyozott α-egyenletes eloszlás

A rövid intervallumokon vett összegek után ebben az fejezetben áttérünk számtani sorozatok mentén vett összegekre.

Cécile Dartyge, Sárközy András és Gyarmati Katalin egy éppen készülő cikkben [14] bináris sorozatok eloszlásait vizsgálták számtani sorozatok mentén. E célból bevezették a pszeudovéletlen sorozatok súlyozott α-egyenletes eloszlásának mértékét, majd ennek minimumára és maximumára becsléseket, illetve korlátokat próbáltak adni.

A probléma a következő: tegyük fel, hogy szükségünk van egy ismeretlen L hosszúságú pszeudovéletlen sorozatra. Ha L értékét körülbelül jól meg tudjuk határozni, például létezik olyan U, hogy $U < L < 2U$, akkor a feladat megoldható. Készítünk egy $2U < N < 4U$ hosszúságú $E_N = (e_1, \dots, e_N)$ pszeudovéletlen sorozatot, majd ennek vesszük az első L elemét. Ekkor $E_L = (e_1, \dots, e_L)$ szintén „jó" pszeudovéletlen sorozat lesz minden $U < L < 2U$ esetén. Ugyanis a korábban bevezetett W és C_k mértékek definíciójából következik, hogy ha E_N „jó" sorozat, akkor $0 \le n < n+M \le N$ és $M \gg N$ vagy $M > N^{1-\varepsilon}$ esetén $(e_{n+1}, e_{n+2}, \dots, e_{n+M})$ szintén „jó". Azonban, ha például csak annyit tudunk mondani, hogy $U < L < U^{100}$, akkor ez a módszer nem működik. A probléma az, hogy $M < N^{1/2}$ és $1 \le n < n+M \le N$ esetén $E_N = (e_1, \dots, e_N)$ „jó" pszeudovéletlen tulajdonságai nem garantálják, hogy $(e_{n+1}, e_{n+2}, \dots, e_{n+M})$ szintén pszeudovéletlen a mértékek értelmében.

A problémát kezelhetnénk, ha mondjuk képesek lennénk olyan $E_N \in \{-1, +1\}^N$ sorozatokat készíteni, melyeknél minden $M > N^{\varepsilon}$ esetén $(e_{n+1}, e_{n+2}, \dots, e_{n+M})$ szintén „jó", a W és C_k mértékeik például kisebbek, mint $M^{1/2+\varepsilon}$ vagy M^{1-c}.

A továbbiakban a W egyenletes eloszlás mértékét vizsgáljuk ebből a szempontból. Szükség lesz a következő jelölés bevezetésére: ha $E_N = (e_1, e_2, \dots, e_N) \in \{-1, +1\}^N$, $n \in \{0,1, \dots, N-1\}$, $M \in \mathbb{N}$ és $0 \le n < n+M \le N$, akkor legyen

$$E_N(n, M) = (e_{n+1}, e_{n+2}, \dots, e_{n+M}).$$

7.1. Definíció (Dartyge-Gyarmati-Sárközy)**.** Ha E_N olyan bináris sorozat, mint fent és $0 \le \alpha \le 1/2$, akkor E_N súlyozott α-egyenletes eloszlását úgy definiáljuk,

mint

$$W_\alpha(E_N) = \max_{0\le n<n+M\le N} M^{-\alpha} W(E_N(n,M)),$$

Itt W a korábban bevezetett egyenletes eloszlás mértéke:

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb} \right|,$$

ahol $a, b, t \in \mathbb{N}$ és $1 \le a \le a + (t-1)b \le N$. Látható, hogy $\alpha = 0$ esetén $W_0(E_N) = = W(E_N)$, vagyis az egyenletes eloszlás mértékét kapjuk vissza.

Vezessük még be a következő jelölést: $0 \le \alpha \le 1/2$ esetén legyen

$$m_\alpha(N) = \min_{E_N \in \{-1,+1\}^N} W_\alpha(E_N).$$

A fent tárgyaltak értelmében célunk ennek a minimumnak a vizsgálata.

$W_\alpha(E_N)$ tipikus értékéről az alábbiakat mondhatjuk el, aminek a bizonyítása [14]-ben szerepel:

7.1. Tétel (Dartyge-Gyarmati-Sárközy)**.** *Tegyük fel, hogy* $0 \le \alpha \le 1/2$. *Ekkor bármely* $\varepsilon > 0$*-hoz léteznek olyan* $N_0 = N_0(\varepsilon)$ *és* $\delta = \delta(\varepsilon)$ *számok, hogy minden* $N > N_0$ *esetén egy véletlenül választott* $E_N \in \{-1,+1\}^N$ *sorozatra (vagyis bármely* E_N *sorozatot* $\frac{1}{2^N}$ *valószínűséggel választva):*

$$P(W_\alpha(E_N) > \delta N^{(1/2)-\alpha}) > 1 - \varepsilon \tag{7.1}$$

és

$$P(W_\alpha(E_N) > 6(N \log N)^{1/2} N^{-\alpha}) < \varepsilon. \tag{7.2}$$

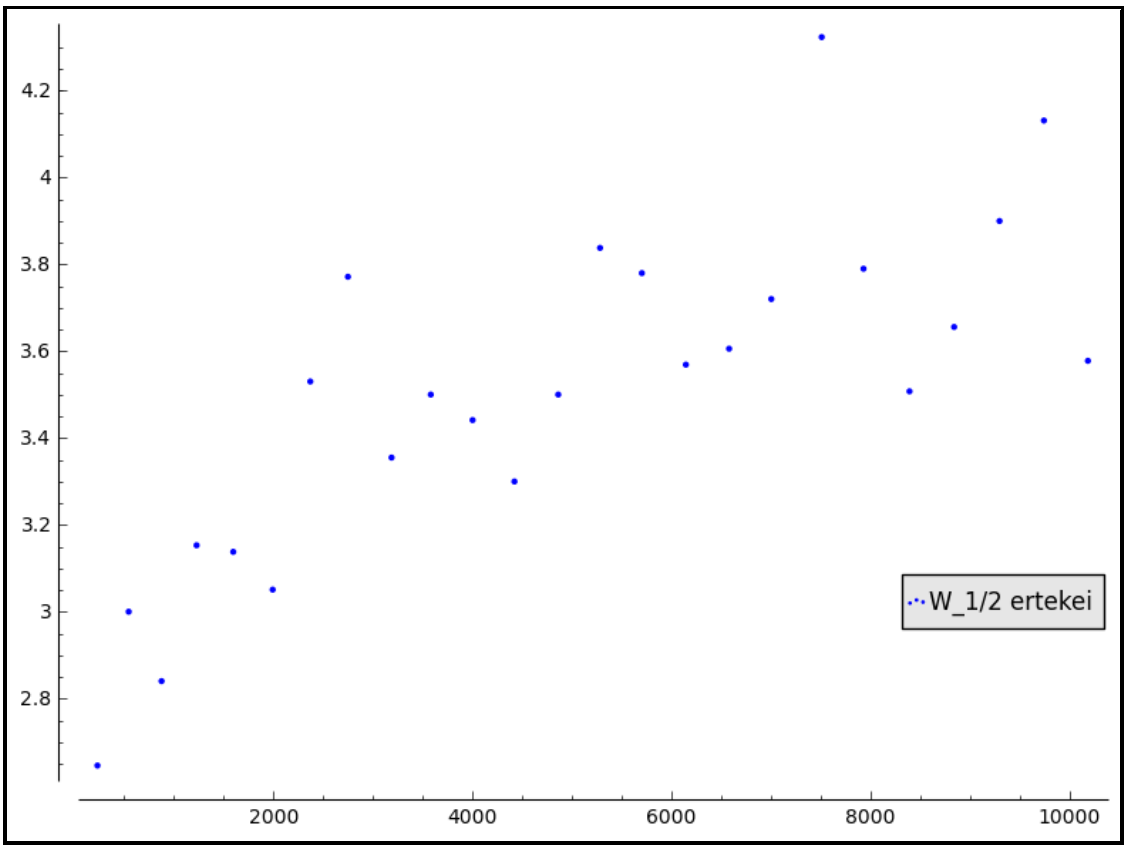
Azaz W_α értéke az esetek többségében $N^{(1/2)-\alpha}$ körül van. Ugyanakkor [14]-ben azt is sikerült belátni, hogy

$$cN^{\frac{1}{4}-\alpha} < m_\alpha(N).$$

Láthatjuk, hogy ismét megjelent egy c abszolút konstans. Ezt szintén vizsgálhatjuk a számítógép segítségével, hogy legalább körülbelüli becslést kapjunk rá, illetve később pontosan meghatározhassuk az értékét. Teszteléshez a Legendre-szimbólum által generált sorozatot használjuk és több α esetén is kiszámoljuk W_α-t.

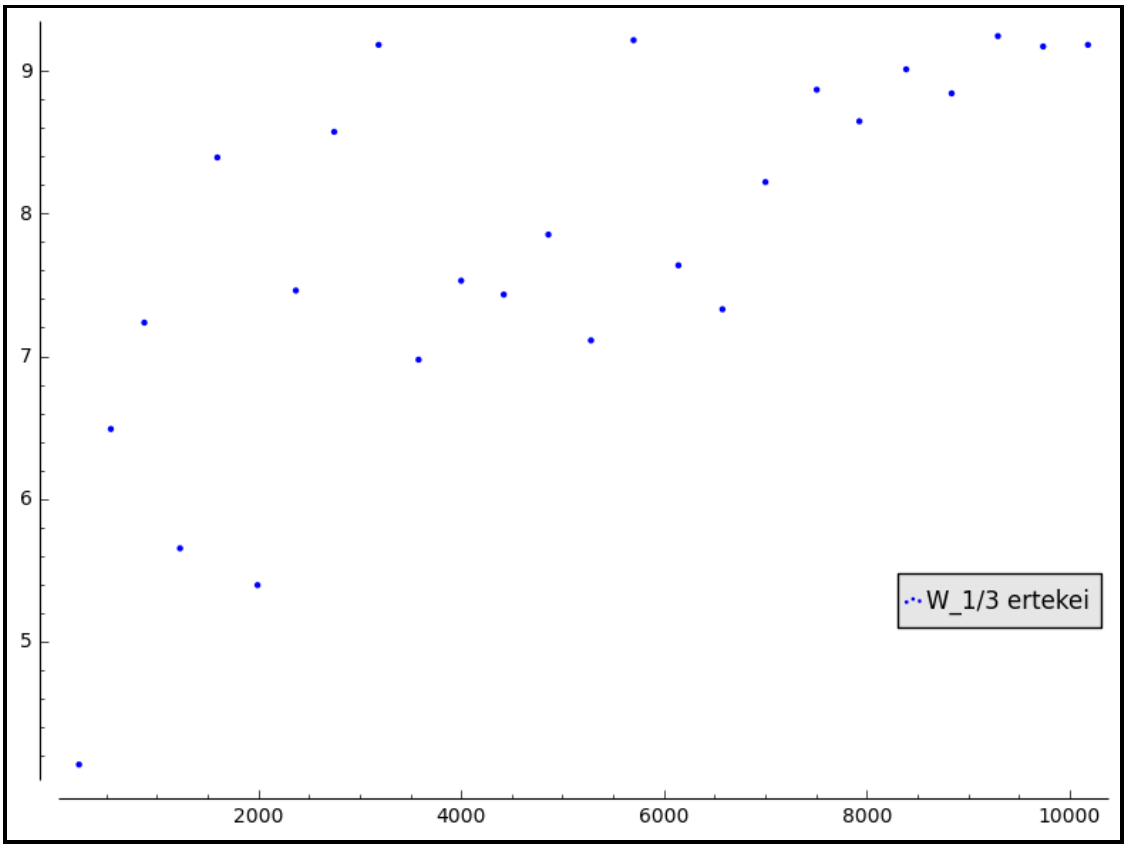
Számolási eredmények

W_α értékeinek meghatározásához ugyancsak a Sage programcsomagot használtam. Mivel ennek kiszámítása jóval bonyolultabb és több ideig tart, mint a rövid összegeknél vett számolások, ezért a sorozatok hosszát csak ezres nagyságrendben tudtam vizsgálni. A Sage munkafüzet *szd_sulyozottW* néven található meg a szakdolgozat mellékleteként. Az eredmények a 2-7. ábrán láthatók.



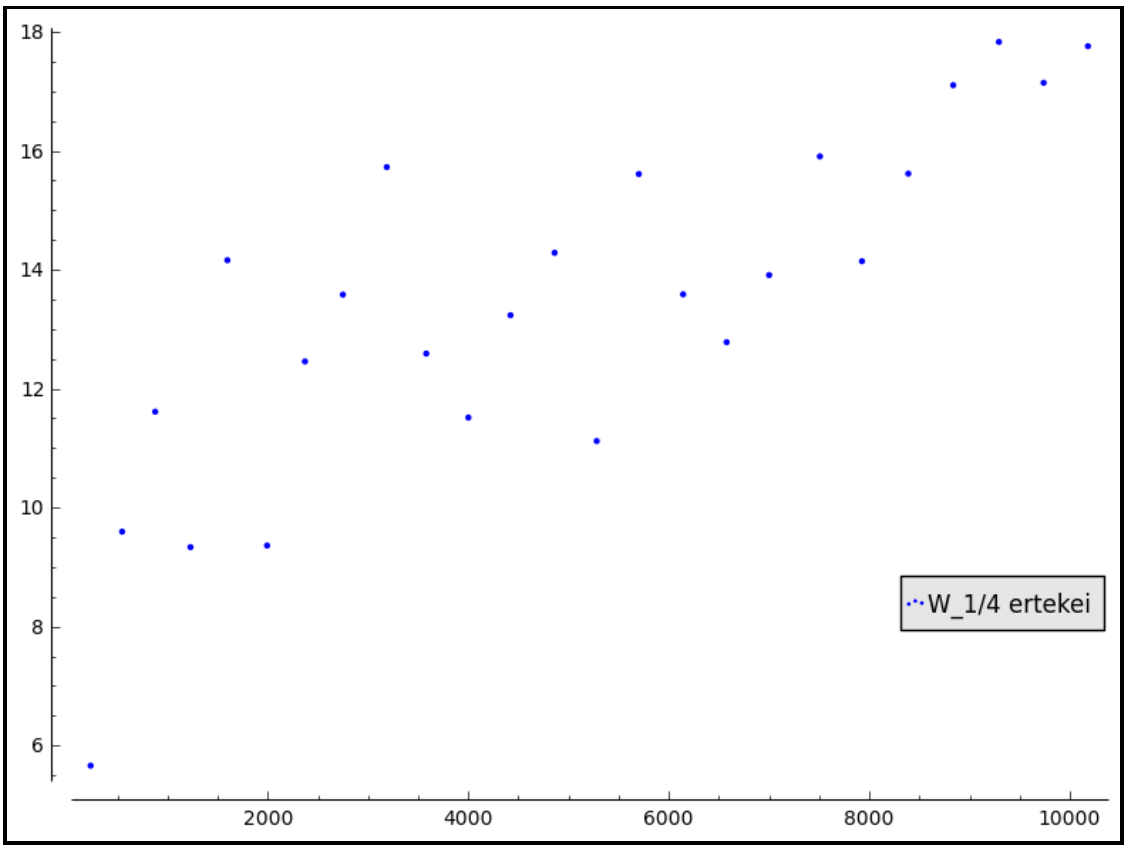
2. ábra. Legendre-szimbólummal elkészített sorozatok $W_{1/2}$ értékei.

Az x tengelyen a p prímek szerepelnek, míg az y tengelyen a p prímekhez tartozó $W_\alpha(E_{p-1})$ maximumok $\alpha = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}$ és 0 esetén. A számítások alapján azt látjuk, hogy $W_\alpha(E_{p-1})$ értékei (ahol E_{p-1} a Legendre–sorozat) nagyon lassan nőnek, amint $p \to \infty$-be. Természetesen itt még további vizsgálatokra van szükség, a grafikonok segítségével csak egy körülbelüli képet kapunk arról, hogy a maximumok hogyan

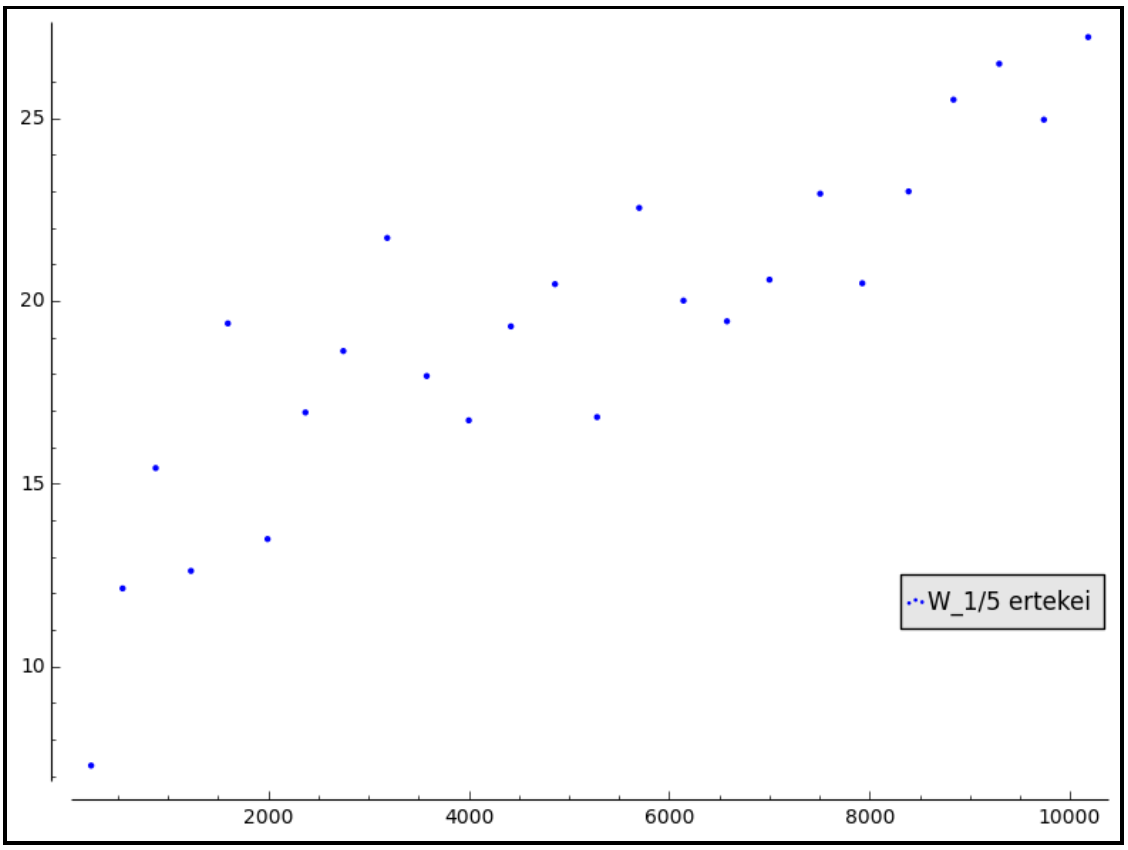


3. ábra. Legendre-szimbólummal elkészített sorozatok $W_{1/3}$ értékei.

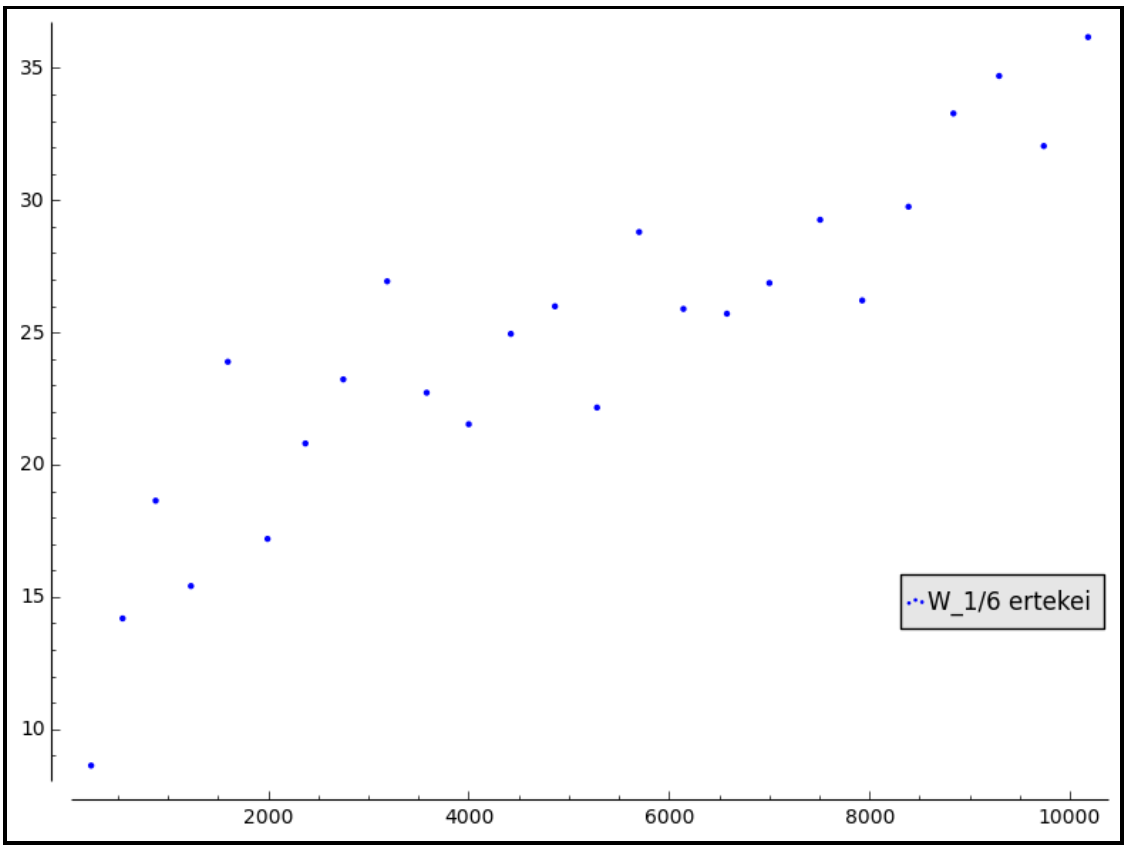
helyezkednek el egy konkrét sorozat esetén α függvényében.



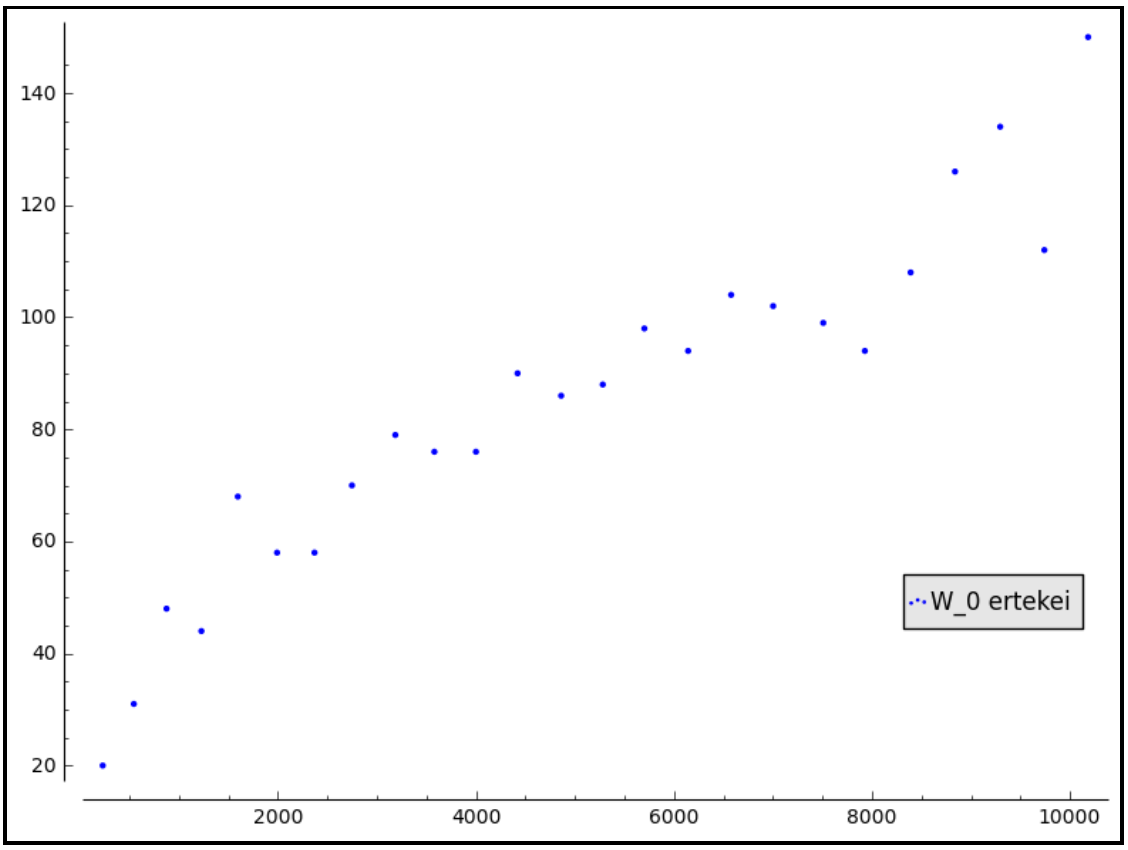
4. ábra. Legendre-szimbólummal elkészített sorozatok $W_{1/4}$ értékei.



5. ábra. Legendre-szimbólummal elkészített sorozatok $W_{1/5}$ értékei.



6. ábra. Legendre-szimbólummal elkészített sorozatok $W_{1/6}$ értékei.



7. ábra. Legendre-szimbólummal elkészített sorozatok W_0 értékei.

A lemezmelléklet tartalma

A lemezmelléklet a következő könyvtárakat tartalmazza:

Cikkek

Ebben a könyvtárban található meg azon cikkek és könyvek nagy része, melyeket a dolgozat írása közben felhasználtam.

Dolgozat

A „Dolgozat” könyvtárban szerepel a szakdolgozat .pdf formátumban.

Munkafüzetek

Ebben a könyvtárban található a két Sage munkafüzet, a *szd_shortsum.sws* és *szd_sulyozottW.sws*.

Neumann generátor

Ebben a könyvtárban a Neumann pszeudovéletlen generátort megvalósító rövid program, a C++ forráskód, valamint a kimenet és a számolási eredmények találhatók. Továbbá a *neumann* nevű alkönyvtárban a Visual Studio Solution projektfájl foglal helyet.

Felhasznált irodalom

Hivatkozások

[1] Donald Knuth, *Art of Computer Programming, Volume 2: Seminumerical Algorithms (3rd Edition)*. Addison-Wesley, 1997., ISBN-13: 978-0-201-89684-8, ISBN-10: 0-201-89684-2

[2] Lovász László, *Algoritmusok bonyolultsága*. Typotex kiadó, 2014., ISBN 978-963-279-253-8

[3] Christian Mauduit, András Sárközy, *On Finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol*. Acta Arithmetica 82 (4) (1997), 365-377

[4] Katalin Gyarmati, *Measures of pseudorandomness*. P. Charpin, A. Pott, A. Winterhof (eds.), Radon Series in Computational and Applied Mathematics, de Gruyter 2013., 43-64.

[5] Katalin Gyarmati, *On a pseudorandom property of binary sequences*. Ramanujan J. 8 (2004), 289-302.

[6] Louis Goubin, Christian Mauduit, András Sárközy, *Construction of large families of pseudorandom binary sequences*. Number Theory 106 (2004), 56-69.

[7] Freud Róbert, Gyarmati Edit, *Számelmélet - Második, javított és bővített kiadás*. Nemzeti Tankönyvkiadó, Budapest, 2006., ISBN 693-19-5888-4

[8] Tom M. Apostol, *Introduction to Analytic Number Theory*. Springer-Verlag, New York, 1976., ISBN 0-387-90163-9

[9] A. J. Hildebrand, *Introduction to Analytic Number Theory*. Math 531 Lecture Notes, Fall 2005. http://www.math.uiuc.edu/∼hildebr/ant

[10] Carl Pomerance, *Remarks on the Pólya–Vinogradov inequality*. Integers (Proceedings of the Integers Conference, October 2009), 11A (2011), Article 19, 11pp.

[11] D. A. Frolenkov, *Numerically explicit version of the Pólya–Vinogradov inequality*. Mosc. J. Comb. Number Theory, 1:3 (2011), 25–41

[12] Jonathan W. Bober, Leo Goldmakher, *Pólya–Vinogradov and the least quadratic nonresidue*. arXiv preprint arXiv:1311.7556, 2013.

[13] Terence Tao, *The least quadratic nonresidue, and the square root barrier*. https://terrytao.wordpress.com/2009/08/18/the-least-quadratic-nonresidue-and-the-square-root-barrier/, 2009.

[14] Cécile Dartyge, Katalin Gyarmati, András Sárközy, *On regularities of distribution of binary sequances relative to arithmetic progressions*. In preparation.

[15] SageMath, *Sage Standard Documentation v6.5*. http://www.sagemath.org/

[16] Christian Mauduit, András Sárközy, *On the measures of pseudorandomness of binary sequences*. Discrete Mathematics 271 (2003), 195-207