Gyarmati Katalin
"Pseudorandomness of finite binary sequences and lattices" című értekezésének bírálata

A XX. század második felétől kezdődő számítástechnikai fejlődés a kriptográfiai kutatások gyors fellendülését eredményezte. A véletlen sorozatok létrehozása fontos kihívást jelentett mind a matematikusoknak, mind a mérnököknek. Számos fizikai megvalósulás (pl. Zener diódák termális zaját, vagy radioaktív bomlást használó eljárás) mellett a matematikusok is több módszert fejlesztettek ki véletlen sorozatok előállítására. Itt egy determinisztikus algoritmussal olyan sorozatot generálunk, ami ugyan a determinisztikussága miatt ténylegesen nem véletlenszerű, de egy "igazi" véletlen sorozat főbb tulajdonságaival rendelkezik. Igy jutunk el egy pszeudovéletlen sorozathoz. A pszeudovéletlenséget először bonyolultságelméleti módon értelmezték. Ebben az esetben a problémát az okozza, hogy itt végtelen sorozatot kell vizsgálni, míg a gyakorlatban mindig véges sorozatok fordulnak elő.

A véges pszeudovéletlen sorozatok vizsgálatában mérföldkő volt Mauduit és Sárközy 1997-es, Acta Arithmeticában megjelent dolgozata, mely egy hétrészes cikksorozat első része volt. Ebben először is a hagyományos 0 vagy 1 számokat tartalmazó sorozatok helyett a ± 1 számokból álló sorozatokat vettek (nyilván a kettő kölcsönösen egyértelműen megfeleltethető egymásnak). Ebben a dolgozatban dolgozták ki a pszeudovéletlenség legfontosabb kvantitativ mértékeit, amik azóta is a véges pszeudovéletlen sorozatok vizsgálatának alapfogalmai. A továbbiakban E_N-nel jelöljük egy N hosszúságú ± 1 számokat tartalmazó sorozatot: legyen $E_N = (e_1, e_2, \dots, e_N)$, ahol $e_n \in \{1, -1\}$. Először is bevezették az ún. eloszlási mértéket, ami azt méri, hogy számtani sorozatokban a -1 és $+1$ számok mennyire egyenletesen fordulnak elő:

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

Legyen $D = d_1, d_2, \dots, d_l$, $0 \le d_1 < d_2 < \cdots < d_l$. Ekkor az l-edrendű korreláció mértéke:

$$C_l(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_l} \right|.$$

Az ún. k-adrendű normalitás mértéke azt fejezi ki, hogy egy sorozat hányszor fordul elő mint részsorozat a várható értékhez képest: Legyen $X = (x_1, x_2, \dots, x_l)$, $x_i \in \{-1, 1\}$,

$$N_l(E_N) = \max_{M,X} \left| |\{n : 0 \le n \le M-1, (e_{n+1}, e_{n+2}, \dots, e_{n+l}) = X\}| - \frac{M}{2^l} \right|.$$

Mauduit és Sárközy 1997-ben igazolta, hogy a normalitás mértéke felülről becsülhető a korrelációs mértékkel, emiatt a normalitás mértékét nem szokták külön vizsgálni. Az általános korrelációs mérték definíciója a következő:

$$Q_l(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t-1} e_{a+jb+d_1} e_{a+jb+d_2} \dots e_{a+jb+d_l} \right|.$$

Ezzel lehet például azt kiszűrni, hogy a sorozatunk nem olyan, ahol a páratlan indexű tagok véletlen sorozatot tartalmaznak, de a páros indexű az előtte lévő páratlan szám (-1)-szerese. Cassaigne, Mauduit és Sárközy igazoltak, hogy majdnem minden E_N sorozatra $W(E_N) = N^{1/2}\log^c N$, valamint $C_l(E_N) = N^{1/2}\log^{c_l} N$. Ez alapján egy E_N sorozatot erős pszeudovéletlen tulajdonsággal rendelkezőnek mondunk, ha $W(E_N) = o(N)$ és $C_l(E_N) = o(N)$. Sárközy Andrásék cikke óta a pszeudovéletlen sorozatok elemzése a számelmélet egyik gyorsan fejlődő ágává vált, ami egyrészt a benne lévő matematika szépségének másrészt az alkalmazhatóságának köszönhető. Számos magyar kutató mellett több francia, kínai és egyéb nemzetiségű matematikus is dolgozik ezen a területen. A kutatás egyik iránya bevezett fogalmak közötti kapcsolatok feltérképezésére irányulnak, másrészt speciális sorozatok esetén vizsgálják azok fenti (és további) mértékekre vonatkozó tulajdonságait.

Gyarmati Katalin értekezésében az alkalmazások szempontjából is fontos több sorozatra bizonyítja az erős pszeudovéletlen tulajdonságot illetve a bevezetett mértékek között mély kapcsolatot bizonyít.

Az értekezés a bevezető fejezettel kezdődik, ahol a szerző áttekinti a pszeudovéletlen sorozatok történetét, a kvantitatív vizsgálatokhoz szükséges legfontosabb fogalmakat, az ezen a területen elért legfontosabb eredményeket illetve ennek sorába illesztve a doktori értekezés eredményeit ismerteti.

Az értekezés 2. fejezet a Blum-Blum-Shub-féle hatványgenerátorral foglalkozik. Itt egy alkalmasan választott kiindulási szám és modulus esetén az egymás utáni hatványozás során kapott szám bináris felírásának utolsó számjegye határozza meg az E_N sorozatot. Ez az egyik legtöbbet használt és vizsgált véletlen sorozat generátor. A szerző az eloszlás mértékére, a normalitás mértékére valamint korreláció mértékére bizonyítja a sorozat véletlen tulajdonságát erős formában. A bizonyítás során exponenciális összegekre vonatkozó modern eredményeket (Bourgain-tétele, Friedlander-Hansen-Shparlinski-tétele) használ illetve fejleszt tovább. A bizonyítás alapgondolata Mauduit, Rivat és Sárközy 2004-ben megjelent cikkében szerepel: az általuk bevezetett súlyfüggvényt szerepelteti az exponenciális összegben. A bizonyításhoz kidolgozott exponenciális összegekre vonatkozó lemmák más problémáknál is jelentősek lehetnek.

Az értekezés 3. fejezetében korrelációs mértékek közötti kapcsolatokat vizsgál a szerző. Amint azt az $E_N = \{1, -1, 1, -1 \dots\}$ sorozat mutatja, hogy a $C_{2k+1}(E_N) = O(1)$ lehetséges. Ekkor azonban $C_{2l}(E_N) \gg N$. Alon, Kohayakawa, Mauduit és Rödl 2007-ben bizonyította be, hogy minden páros l esetén $C_l(E_N) \geq \sqrt{\frac{1}{2}[\frac{N}{l+1}]}$. Az alapkérdés az volt, hogy vajon van-e olyan E_N sorozat, melyre $C_{2k+1}(E_N)$ és $C_{2l}(E_N)$ is kicsi, például $C_2(E_N) = O(\sqrt{N})$ és $C_3(E_N) = O(1)$ egyszerre lehetséges-e. A szerző a $C_{2k+1}(E_N)C_{2l} \gg N^{c(k,l)}$, ahol $c(k,l) = 1$, ha $k \geq l$ és $c(k,l) = \frac{1}{2} + \frac{2k+1}{4l}$, ha $k < l$ egyenlőtlenségeket igazolja. Speciálisan a $C_2(E_N)C_3(E_N) \gg N^{2/3}$ egyenlőtlenség adódik. A bizonyítás egy általánosabb korrelációs mértékeket tartalmazó egyenlőtlenség következménye, amiről a szerző megmutatja, hogy optimális. A bizonyítás során a szerző bevezet egy polinomot, melynek tulajdonságait felhasználva egy szép négyoldalas bizonyítását adja a becslésnek.

A 4. fejezetben a szerző egy Legendre-szimbólumokkal definiált $\mathcal{F}$ család komplexitását vizsgálja. Legyen $\mathcal{F}$ bizonyos E_N sorozatokból álló halmaz. Ekkor $C(\mathcal{F})$-fel jelöljük a legnagyobb olyan j számot, hogy minden ± 1 számokból álló j hosszú sorozat előfordul tetszőleges $1 \leq i_1 < i_2 < \dots < i_j \leq N$ esetén valamelyik $E_N \in \mathcal{F}$ sorozatban mint $(e_{i_1}, e_{i_2}, \dots, e_{i_j})$. Sárközy és társai bizonyították be, hogy bizonyos feltételeknek eleget tevő legfeljebb K-adfokú polinomok esetén a Legendre-szimbolum segítségével az általuk generált sorozatok komplexitása legalább K. A másik oldalról megmutatták, hogy $C(\mathcal{F}) \leq \frac{\log|\mathcal{F}|}{\log 2}$. Gyarmati Katalin megmutatja, hogy nem túl nagy K-k esetén a felső becslés nagyságrendileg pontos: $C(\mathcal{F}) \geq \frac{K-1}{2\log 2} - O(K\log(K\log p))$. A bizonyításban a polinom értékeken vett karakterösszegekre vonatkozó Weil-tételt használja

egy átlagolásos eljárást alkalmazva.

A pszeudovéletlen sorozatok alkalmazásánál gyakran fontos, hogy a pszeudovéletlenség lokálisan is teljesüljön. Az 5. fejezetben olyan sorozatokat konstruál a szerző, amelyek rövidebb részsorozatai is rendelkeznek jól becsült pszeudovéletlen tulajdonságokkal. A bizonyítás során a szerző kétdimenziós pszeudovéletlen rácsot generál, amiből vetítéssel kapja az E_{N^2} sorozatot. A bizonyítás során megmutatja azt az önmagában is érdekes eredményt, hogy ha a rács korrelációs mértéke kicsi, akkor a vetítéssel kapott sorozat korrelációs mértéke is kicsi.

A 6. fejezetben Legendre-szimbólumok segítségével kapott rácsok pszeudovéletlen tulajdonságát vizsgálja a szerző. Mauduit és Sárközy bizonyították be, hogy ha $f(x) \in \mathbb{F}_p[x]$ egy "szép" tulajdonsággal rendelkező polinom, akkor az $e_n = \left(\frac{f(n)}{p}\right)$, $1 \le n \le N$ sorozat eloszlás mértéke jól kontrollálható. Gyarmati, Sárközy és Stewart ennek a tételnek kétváltozós változatát bizonyítja: olyan kétváltozós $f(x_1, x_2) \in \mathbb{F}_p[x_1, x_2]$ polinomokat definiál, melyekből származó alkalmasan definiált korrelációs mérték kicsi. A szerzők két esetet különböztetnek meg aszerint, hogy a polinom $f(x_1, x_2) = \left(\prod_{j=1}^{r} f_j(\alpha_j x_1 + \beta_j x_2)\right) g(x_1, x_2)^2$ alakú vagy sem. A bizonyítás itt is karakterösszegeket és Weil-tételt használ. Az 5., 6. és 7. fejezetben a bizonyítások a korábbi szintén összetett bizonyításokhoz képest is sok - nem csak technikai - nehézség leküzdését igénylik.

Az értekezésben szereplő bizonyítások részletesen kidolgozottak, jól követhetők, hibát nem találtam bennük. Az értekezés felépítése logikus, talán a 2. és 3. fejezetet lehetett volna felcserélni, hogy a speciális sorozatokra vonatkozó eredmények egy tömbben legyenek. Az értekezés tipográfialag gondosan kidolgozott, a terjedelméhez képest elenyésző számú sajtóhibát tartalmaz. A tételeknél egyértelműen vannak jelezve a szerzők; itt még érdemes lett volna a megjelenés évszámát is feltüntetni. Zárójelben jegyzem meg, hogy a Weil-tétel kétszer is ki lett mondva: ez a Lemma 4.1 és a Lemma 5.2 is.

Összefoglalva: A szerző a pszeudovéletlen sorozatok elméletének terén ért el a legújabb nemzetközi kutatások élvonalába tartozó eredményeket. A bizonyításokban a szerző az exponenciális összegek és az algebra mély eszközeit kreatívan használja, néhol továbbfejleszti. A kimondott eredmények és a bizonyítások részletei is jelentős érdeklődésre tarthatnak számot, ami az eddigi hivatkozásokon is látszik. Ezek alapján az értekezést alkalmasnak tartom a nyilvános védésre és javaslom Gyarmati Katalin számára az MTA doktori cím odaítélését.

Kérdések:

1. Milyen más mértékei vannak a ± 1 számokat tartalmazó véges sorozatok pszeudovéletlenségnek?

2. Kapcsolódik-e szabadalom az értekezésben szerreplő konstrukciókhoz? Van-e olyan szoftver, ami a szerző konstrukcióit használja?

Budapest, 2014. április 25. Sándor Csaba