Pszeudovéletlenség

A számítógépes számelmélet tárgy

Neal Koblitznek A Course in Number Theory and Cryptography

című könyvére épülő anyagának kiegészítése

Gyarmati Katalin, Sárközy András

Eötvös Loránd Tudományegyetem

2012

A számelmélet talán legfontosabb alkalmazása a véletlenség szimulációja. Megemlítjük, hogy gyakorlatilag minden számítógépes szoftvernek része egy véletlen szám táblázat. Ezek a táblázatok, számsorozatok rendszerint számelméleti eszközök felhasználásával készülnek. A pszeudovéletlenség legfontosabb alkalmazásai a numerikus analízishez, illetve a kriptográfiához kapcsolódnak. A továbbiakban viszonylag részletesen fogunk foglalkozni ezzel a két területtel.

1. Pszeudovéletlen $[0, 1)$ sorozatok alkalmazásai a numerikus analízisben

A pszeudovéletlenségnek számos megközelítése van (e jegyzetben a függelékben tárgyalunk néhányat). A sok megközelítés közül kiindulópontként az alábbi definíciót tekintjük:

1.1. Definíció. *Ha $(x_1, x_2, \ldots, x_n)$ olyan valós számokból álló (véges) sorozat, hogy $i = 1, 2, \ldots, n$ esetén $0 \leq x_i < 1$, és az*

$$N(\alpha, \beta) = |\{i : \ 1 \leq i \leq n, \ \alpha \leq x_i < \beta\}|$$

jelölést bevezetve, bármely $0 \leq \alpha < \beta \leq 1$ esetén

$$\left| \frac{N(\alpha, \beta)}{n} - (\beta - \alpha) \right|$$

„kicsi" akkor azt mondjuk, hogy a sorozat egyenletes eloszlású $[0,1)$*-ben.*

Alkalmazástól függő, hogy a definícióban a „kicsi" alatt pontosan milyen korlátot értünk. Végtelen sorozat esetén a definíció egyszerűen tehető pontosabbá:

1.2. Definíció. *Az* $(x_1, x_2, \dots)$ *végtelen sorozat egyenletes eloszlású* $[0,1)$*-ben, ha* $i = 1, 2, \dots$ *esetén* $0 \le x_i < 1$*, és az*

$$N(\alpha, \beta, n) = |\{i : \ 1 \le i \le n, \ \alpha \le x_i < \beta\}|$$

jelölést bevezetve, bármely $0 \le \alpha < \beta \le 1$ *esetén*

$$\lim_{n\to\infty} \left| \frac{N(\alpha, \beta, n)}{n} - (\beta - \alpha) \right| = 0.$$

A többdimenziós egyenletes eloszlás fontos szerepet játszik a pszeudovéletlenség kritériumai között. Ezt ismertetjük most.

1.3. Definíció. *Ha* $\alpha, \beta \in [0,1]^s$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$ *és* $\beta = (\beta_1, \beta_2, \dots, \beta_s)$*, akkor azt mondjuk, hogy*

$$\alpha \le \beta,$$

ha α *és* β *minden koordinátájára*

$$\alpha_i \le \beta_i \qquad (i = 1, 2, \dots, s)$$

teljesül.

Az $(\mathbf{x}_1, \mathbf{x}_2, \dots)$ *végtelen sorozat egyenletes eloszlású* $[0,1)^s$*-ben, ha* $i = 1, 2, \dots$ *esetén* $\mathbf{x_i} \in [0,1)^s$ *és* $\alpha, \beta \in [0,1]^s$ *vektorokra az*

$$N(\alpha, \beta, n) = |\{i : \ 1 \le i \le n, \ \alpha \le \mathbf{x_i} < \beta\}|$$

jelölést bevezetve, bármely $\alpha, \beta \in [0,1]^s$, $(0, \dots, 0) \le \alpha = (\alpha_1, \dots, \alpha_s) \le \beta = (\beta_1, \dots, \beta_s) \le (1, \dots, 1)$ *esetén*

$$\lim_{n\to\infty} \left| \frac{N(\alpha, \beta, n)}{n} - \prod_{j=1}^{s} (\beta_j - \alpha_j) \right| = 0.$$

A $[0,1)$ sorozatok pszeudovéletlenségének egyik legelfogadottabb kritériuma a többdimenziós egyenletes eloszlást használja. E szerint egy $(x_1, x_2, \dots)$ végtelen $[0,1)$ sorozat pszeudovéletlen, ha minden $k \ge 1$ természetes számra az $\left((x_n, x_{n+1}, \dots, x_{n+k-1}) \left(\in [0,1)^k\right) : n = 1, 2, \dots\right)$ sorozat egyenletes eloszlású a $[0,1)^k$ kockában, ez az ún. „**serial teszt**".

A pszeudovéletlen $[0,1)$ sorozatokhoz irodalom: I. M. Szobol: A Monte-Carlo módszerek alapjai [18] vagy D. E. Knuth: A Számítógépprogramozás művészete [8].

A pszeudovéletlen $[0,1)$ sorozatok legfontosabb alkalmazása a **Monte-Carlo módszerhez** kapcsolódik. E fogalomnak nincsen szigorú matematikai definíciója, de nagyjából elfogadható a következő definíció:

1.4. Definíció. ***Monte Carlo módszerről*** *akkor beszélünk, ha egy kiszámítandó mennyiség értékét az alábbi úton becsüljük: keresünk olyan* ξ *va-*

lószínűségi változót, amelynek várható értéke a: $M(\xi) = a$. *Ezután* ξ *értékére nagyszámú független megfigyelést végzünk, a kapott értékek legyenek* $\xi_1, \xi_2, \ldots, \xi_N$. *Ekkor a-t* $\frac{1}{N}\sum_{i=1}^{N}\xi_i$*-vel közelítjük:*

$$a \approx \frac{1}{N}\sum_{i=1}^{N}\xi_i.$$

(Ezt a tényt indokolja, hogy a jobb oldal várható értéke a.)

Fontos speciális eset az $[a, b]$ zárt intervallumon Riemann integrálható $f(x)$ függvény integráljának becslése:

$$I = \int_a^b f(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} f(\xi_i)(b-a).$$

(A függelékben részletesen vizsgáljuk a többváltozós integrálokat is.)

Tegyük fel az egyszerűség kedvéért, hogy $[a, b] = [0, 1]$. Ekkor a gyakorlatban ezt a módszert úgy alkalmazzuk, hogy tekintünk egy pszeudovéletlen $(x_1, x_2, \ldots, x_n)$ sorozatot, ahol $0 \leq x_i \leq 1$, kiszámoljuk $f(x_1), f(x_2), \ldots, f(x_n)$-et, és I-t

$$I \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

-vel közelítjük.

Mint már említettük a pszeudovéletlen struktúrák általában valamilyen **számelméleti konstrukcióval** készülnek; néhány kivételes esetben előre tudjuk, hogy a megkonstruált sorozat rendelkezik valamilyen véletlen tulajdonsággal vagy tulajdonságokkal (ilyenkor „apriori” vagy „elméleti” tesztről

beszélünk), az esetek többségében azonban az elkészült sorozatot utólag tesztelik valamilyen valószínűségszámítási módszerrel („aposzteriori" vagy „empirikus teszt").

A legfontosabb módszer pszeudovéletlen $[0,1)$ sorozat készítésére a **lineáris kongruencia módszer**, mely D. H. Lehmertől (1949) és Neumann Jánostól ered.

1.5. Definíció. ***Lineáris kongruencia módszer*** *pszeudovéletlen* $[0,1)$ *sorozat készítésére az alábbi: mondjuk* N *hosszúságú pszeudovéletlen* $[0,1)$ *sorozatot akarunk készíteni. Ehhez célszerű először valamely* N*-nél „kicsit nagyobb" m számot választani, majd még további három* $(\in \mathbb{Z})$ *paramétert:* A, $a(\not\equiv 0 \pmod m)$, b. *Ezután képezzük az* $y_1, y_2, \ldots, y_N$ *sorozatot a következő rekurzióval:*

$$y_1 \equiv A \pmod m \quad 0 \leq y_1 < m$$

és

$$y_{n+1} \equiv ay_n + b \pmod m, \quad 0 \leq y_{n+1} < m \tag{1}$$

$$n = 1, 2, \ldots, N-1 \text{ esetén.}$$

Legyen továbbá $i = 1, 2, \ldots, N$ *esetén* $x_i = \frac{y_i}{m}$. *H. Niederreiter [14], [15], [16] igazolta, hogy bizonyos paraméter választások mellett az így konstruált* $(x_1, x_2, \ldots, x_N)$ *sorozat kielégíti a serial tesztet (tehát ez apriori teszt), és így tekinthető pszeudovéletlen* $[0,1)$ *sorozatnak.*

1.1. Példa. *Ha* $y_1 = A = m - 1,\ a = -1,\ b = 0$, *akkor indukcióval igazolható, hogy*

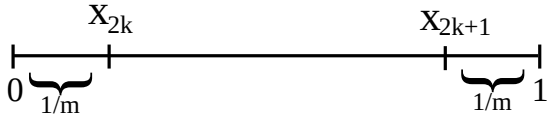
$$y_n \equiv (-1)^n \pmod{m} \quad (n = 1, 2, \dots),$$

ahonnan következik, hogy

$$y_{2k} = 1, \quad y_{2k+1} = m - 1,$$

és így

$$x_{2k} = \frac{y_{2k}}{m} = \frac{1}{m}, \quad x_{2k+1} = \frac{y_{2k+1}}{m} = 1 - \frac{1}{m} :$$



(minden $k \in \mathbb{N}$*-re),*
tehát a sorozat nincs egyenletesen eloszolva $[0, 1)$*-ben.*

Könnyű belátni a skatulyaelvvel (HF), hogy az $(x_1, x_2, \dots)$ sorozat mindig periodikus; ha a periódus kicsi (mint a példában) a sorozat nem jó, ha a periódus nagy („nem sokkal kisebb" m-nél), akkor a sorozat Niederreiter [14], [15], [16] eredménye szerint jó, abban az értelemben, hogy kielégíti a serial tesztet.

A lineáris kongruencia módszert tovább lehet fejleszteni a következőképpen: annak definíciójában szerepelt az (1), a rekurziót definiáló kongruencia, ami $y_{n+1} \equiv f(y_n) \pmod{m}$ alakban írható, ahol $f(x)$ egy lineáris polinom.

Ez a konstrukció általánosítható úgy, hogy $f(x)$ gyanánt magasabbfokú polinomokat is vehetünk. Ez természetesen némi többletmunkával jár, és nehezebben bizonyítható, hogy a megkonstruált sorozat erős pszeudovéletlen tulajdonságokkal rendelkezik, de ezért bőven kárpótol, hogy így „még véletlenszerűbb" sorozatokat konstruálunk.

2. Pszeudovéletlen bináris sorozatok, alkalmazásuk a kriptográfiában

Irodalom: A Menezes, P van Oorschot, S. Vanstone: Handbook of Applied Cryptography, (internetről letölthető) [13].

Talán a legfontosabb, ténylegesen használt titkosítási rendszer az un. „**Vernam cipher**", melyet a XX. század második évtizedében fejlesztett ki Vernam amerikai matematikus.

2.1. Definíció. ((Vernam cipher)) *: először a továbbítandó információt kifejezzük („címkézzük") egy bitsorozat formájában:* $(a_1, a_2, \ldots, a_N)$ *(*$a_i \in \{0,1\},\ i = 1, 2, \ldots, N$*-re). Ezután tekintünk egy* $(e_1, e_2, \ldots, e_N)$ ***véletlen*** *vagy esetleg* ***pszeudovéletlen*** N *hosszú bitsorozatot (akkor véletlen, ha* $\{0,1\}^N$*-ből bármely sorozatot* $\frac{1}{2^N}$ *valószínűséggel választunk: ez az un. „one time pad" (egyszer használatos kulcs)). Ekkor az* f *titkosítási transzformáció*

$(e_1, e_2, \ldots, e_n)$-*nek* $(a_1, a_2, \ldots, a_N)$-*hez való bitenkénti modulo 2 hozzáadása:*

$$\begin{aligned} & (a_1, \ldots, a_N) \\ \oplus\ & \underline{(e_1, \ldots, e_N)} \\ =\ & (b_1, \ldots, b_N). \end{aligned}$$

2.1. Példa.

$$\begin{aligned} & (0111001) && \leftarrow \textit{ titkosítandó információ} \\ \oplus\ & \underline{(1110010)} && \leftarrow \textit{ kulcs} \\ =\ & (1001011) && \leftarrow \textit{ titkosított szöveg ("ciphertext")} \end{aligned}$$

Az elolvasási transzformáció: még egyszer alkalmazni f-et.

Ismeretes, hogy ez a titkosítási rendszer - $(e_1, e_2, \ldots, e_N)$-et valóban véletlenül választva - feltörhetetlen.

Követelmény, hogy az $(e_1, e_2, \ldots, e_N)$ kulcs bitsorozat valóban **véletlen** jellegű legyen.

2.2. Definíció. *Egy* ***véletlen bit generátor*** *egy olyan készülék vagy algoritmus, mely statisztikusan független és torzítatlan („unbiased") biteket állít elő.*

Régen tipikusan hardware bázisú generátorokat (pl. dióda) használtak, de software bázisú generátorok (gépidő, memórianagyság, s í.t.) is lehetségesek;

valamelyik módszerrel előállítanak egy bitsorozatot, és utána ezt tesztelik bizonyos statisztikai tesztekkel (tehát **aposzteriori tesztelést** végeznek). Ez igen komplikált, nehézkes és nem kielégítő technika. Ezért ma már a véletlen bit generátorokat **pszeudovéletlen** bit generátorokkal helyettesítik.

2.3. Definíció. *Egy* ***Pszeudovéletlen bit generátor*** *egy olyan determinisztikus algoritmus, mely egy valóban véletlen k hosszú bináris sorozatot megadva, abból egy ℓ (mely k-nál sokkal nagyobb) hosszú „véletlenszerűnek látszó" bináris sorozatot készít. Az „input" k hosszú véletlen bináris sorozatot „**magnak**" („seed"), a készült ℓ hosszú „output" sorozatot* ***pszeudovéletlen bit sorozatnak*** *nevezzük.*

(Természetesen a bit sorozatok, azaz $0, 1$-ből álló sorozatok kölcsönösen egyértelműen megfeleltethetők bármely más két szimbólumból álló sorozatnak, ezért pl. vizsgálhatunk bit sorozat helyett $-1, +1$-ből álló sorozatot; ez sokszor előnyös.)

De mitől „véletlenszerűnek látszó", azaz, „jó" pszeudovéletlen sorozat egy bizonyos bináris sorozat? Az alkalmazástól (is) függ, milyen véletlen tulajdonságot díjazunk.

Egy kriptográfiában gyakran használt követelmény a megjósolhatatlanság („unpredictability"):

2.4. Definíció. *Azt mondjuk, hogy egy pszeudovéletlen generátor kielégíti*

a ***következő bit tesztet****, ha nincs olyan „polinomiális idejű" algoritmus, mellyel az első k jegy ismeretében a $k+1$-edik $1/2$-nél „lényegesen nagyobb" valószínűséggel megjósolható.*

Kritikája: lényegében csak rekurzív konstrukciók minősítésére alkalmas, ezért lehet, hogy egy generátort elfogadunk, de pl. az első és utolsó bit ismeretében esetleg az egész sorozat egyértelműen meghatározott; polinomiális idejű algoritmus nem-létezése csak feltételesen igazolható; hogyan definiálható az, hogy $1/2$-nél „lényegesen nagyobb"?

A legfontosabb, a következő bit tesztet (feltételesen) kielégítő konstrukció:

2.5. Definíció. *A* ***Blum-Blum-Shub*** *pszeudovéletlen generátor:*

1. Tekintsünk két „nagy", véletlen, $4k+3$ alakú p, q prímet, legyen $n = pq$.

2. Tekintsünk egy véletlen a számot („mag") $0 < a < n$, $(a, n) = 1$-gyel. Legyen x_0 az a^2-nek modulo n legkisebb nem negatív maradéka.

3. Ha $x_0, x_1, \ldots, x_k$ már ismert, legyen x_{k+1} az x_k^2-nek a modulo n vett maradéka.

4. Legyen z_i az x_i szám utolsó számjegye a 2-es alapú számrendszerben ($i = 1, 2, \ldots, t$ esetén).

2.1. Állítás. *A $(z_0, z_1, \ldots, z_t)$ bit sorozat (ahol t az n függvényében nem túl nagy) kielégíti a következő bit tesztet, feltéve, hogy n $(= pq)$ nagyságú számok faktorizálása számítástechnikailag nem lehetséges. (Nem bizonyítjuk.*

Megjegyezzük, hogy a bizonyítás során van szükség arra a feltételre, hogy p és q két $4k+3$ alakú prím.)

Láttuk: a pszeudovéletlenség fenti definíciója („következő bit teszt") a gyakorlatban nem igazán kielégítő. Ezért 1997-től kezdődően kifejlesztésre került bináris sorozatok pszeudovéletlenségének egy új, a korábbinál konstruktívabb, gyakorlati célokra alkalmasabb elmélete. Bevezetésre kerültek bináris sorozatok bizonyos tulajdonságainak véletlen jellegét mérő mértékek (az „eloszlási mérték", „ k-adrendű korreláció", ld. függelék); ha e mértékek „kicsik", akkor a sorozatot „jó" pszeudovéletlen sorozatnak tekintjük. (Ezt azért tehetjük meg, mert majdnem minden sorozatra ezek a mértékek kicsik.) Sikerült megkonstruálni néhány nagy családját olyan bináris sorozatoknak, melyekre e pszeudovéletlen mértékek „kicsik", vagyis e sorozatok **bizonyítottan** (tehát „apriori tesztelten") „jó" pszeudovéletlen sorozatok. A következőkben a legfontosabb ilyen konstrukciók közül említek néhányat:

2.1. Konstrukció. (Goubin, Mauduit, Sárközy [3]) *Tegyük fel, hogy N hosszúságú pszeudovéletlen bináris sorozatot szeretnénk konstruálni. Legyen p olyan prím, amelyre $2N < p$ és a 2 primitív gyök modulo p.(ilyen p prím várhatóan nem sokkal $2N$ után, például általában $4N$-ig van.) Legyen $f(x) \in \mathbb{Z}[x]$ olyan polinom, melynek foka „nem túl nagy" p függvényében (például a fok $< p^{1/10}$), és melynek nincs többszörös gyöke. Ekkor*

$E_N = (e_1, e_2, \ldots, e_N)$ sorozat elemeit az

$$e_n = \begin{cases} \left(\frac{f(n)}{p}\right), & \text{ha } p \nmid f(n), \\ 1, & \text{ha } p \mid f(n) \end{cases}$$

képlettel definiáljuk, ahol $\left(\frac{x}{p}\right)$ a Legendre szimbólum. $E_N = (e_1, e_2, \ldots, e_N)$ ***bizonyítottan „jó"*** *pszeudovéletlen sorozat. (Annak mély oka van, hogy miért kell 2-nek primitív gyöknek lennie modulo p.)*

2.2. Konstrukció. (Mauduit, Sárközy [12]) *Legyen p prím, $f(x) \in \mathbb{Z}[x]$ olyan polinom, amelynek a foka „nem túl nagy" p függvényében, és amelynek nincs többszörös gyöke. Jelöljük n legkisebb nemnegatív maradékát modulo p $r_p(n)$-nel, és legyen $(a, p) = 1$ esetén a multiplikatív inverze modulo p az a^{-1}: $aa^{-1} \equiv 1 \pmod{p}$ (fix p-re). Ekkor e_n-et*

$$e_n = \begin{cases} +1, & \text{ha } (f(n), p) = 1 \text{ és } r_p(f(n)^{-1}) \leq p/2, \\ -1, & \text{ha vagy } (f(n), p) = 1 \text{ és } r_p(f(n)^{-1}) > p/2 \text{ vagy } p \mid f(n) \end{cases}$$

képlettel definiálva, $E_p = (e_1, e_2, \ldots, e_p)$ ***bizonyítottan „jó"*** *pszeudovéletlen sorozat.*

E két konstrukció nagy előnye, hogy egyrészt gyorsan implementálhatók, másrészt a konstruált sorozatok pszeudovéletlen mértékei jól becsülhetők. (A két konstrukció közül talán a Legendre szimbólumra épülő valamivel jobb.) E két konstrukción kívül ma már számos más konstrukció ismert, amelyek

viszonylag jól implementálhatók, és a pszeudovéletlen mértékek is viszonylag jól becsülhetők. Ezek közül ismertetünk két további konstrukciót.

2.3. Konstrukció. (Gyarmati [5]) *Legyen p prím, g primitív gyök modulo p. Amennyiben $(n,p)=1$ definiáljuk* ind n*-et azzal a legkisebb pozitív egésszel, amelyre*

$$n \equiv g^{\operatorname{ind} n} \pmod{p}.$$

Ekkor $1 \le \operatorname{ind} n \le p-1$. Legyen $f(x) \in \mathbb{Z}[x]$ olyan polinom, amelynek a foka „nem túl nagy" p függvényében, és amelynek nincs többszörös gyöke. Ekkor $E_p = (e_1, e_2, \dots, e_p)$ sorozat n-edik elemét

$$e_n = \begin{cases} +1 & \text{ha } (f(n),p)=1 \text{ és } 1 \le \operatorname{ind} f(n) \le p/2, \\ -1 & \text{ha vagy } (f(n),p)=1 \text{ és } p/2 < \operatorname{ind} f(n) < p \text{ vagy } p \mid f(n) \end{cases}$$

képlettel definiálva, azt kapjuk, hogy az $E_p = (e_1, e_2, \dots, e_p)$ sorozat ***bizonyítottan „jó"*** *pszeudovéletlen sorozat.*

Megjegyezzük, hogy ez a konstrukció lassabban implementálható mint az előző kettő, de ezen némi munkával lehet javítani:

2.4. Konstrukció. (Gyarmati [4], [6]) *Definiáljuk p-t, g-t,* ind n*-et és $f(x) \in \mathbb{Z}[x]$-et úgy, ahogy a 2.3. Konstrukcióban. Legyen továbbá m páros szám $p-1$-nek egy kicsi pozitív osztója. Jelölje* $\operatorname{ind}^* n$ *azt a számot, amelyre*

$$\operatorname{ind}^* n \equiv \operatorname{ind} n \pmod{m} \quad \text{és } 1 \le \operatorname{ind}^* n \le m.$$

Ekkor $E_p = (e_1, e_2, \ldots, e_p)$ *sorozat* n*-edik elemét*

$$e_n = \begin{cases} +1 & \textit{ha } (f(n), p) = 1 \textit{ és } 1 \leq \mathrm{ind}^* f(n) \leq m/2, \\ -1 & \textit{ha vagy } (f(n), p) = 1 \textit{ és } m/2 < \mathrm{ind}^* f(n) \leq m \textit{ vagy } p \mid f(n) \end{cases}$$

képlettel definiálva, azt kapjuk, hogy az E_p *sorozat* ***bizonyítottan „jó"*** *pszeudovéletlen sorozat.*

F. Függelék

Az alábbiakban az elmélyedni kívánóknak ismertetek néhány további részletet. Az F.1. és F.2. fejezetben Niederreiter, *Quasi-Monte Carlo methods and pseudorandom numbers* című cikke alapján [17] adunk rövid ismertetőt a többdimenziós Monte-Carlo módszerekről és a pszeudovéletlen $[0, 1)$ sorozatokról. Az F.3. fejezetben a bináris sorozatok pszeudovéletlenségét tanulmányozzuk.

F.1. Monte Carlo módszerek több dimenzióban

Legyen f egy s változós függvény, amelyet $I^s = [0, 1)^s$-en szeretnénk integrálni. Tegyük fel, hogy $x_1.x_2, \ldots, x_N$ véletlen pontok I^s-ben, amelyeket egymástól függetlenül választunk. Ekkor az

$$\int_{I^s} f(t)dt \tag{2}$$

integrált az

$$\frac{1}{N}\sum_{n=1}^{N} f(x_n) \tag{3}$$

értékkel közelítjük. Azt várjuk, hogy amint az x_i pontok számát növeljük, a (3) kifejezés értéke (2)-höz tart, azaz

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) = \int_{I^s} f(t)dt. \tag{4}$$

A következőkben szeretnénk megadni az f függvényeknek és az $(x_1, x_2, \dots)$ sorozatoknak egy olyan nagy családját, amelyre (4) ténylegesen fenn áll. Viszonylag könnyen igazolható, hogy ilyen nagy családot ad például az I^s-en folytonos f függvények halmaza és azon $(x_1, x_2, \dots)$ sorozatok halmaza, amelyek egyenletes eloszlásúak I^s-ben. Valójában azonban több is igaz: ha $(x_1, x_2 \dots)$ egyenletes eloszlású I^s-ben, akkor (4) az összes Riemann integrálható f függvényre fenn áll. (Másrészt (4) nem áll fenn az összes Lebesgue integrálható függvényre, legyen pl. $f(x) = 1$ ha $x \in \{x_1, x_2, \dots\}$ és $f(x) = 0$ különben.)

A Monte-Carlo módszerek tovább általánosíthatóak I^s részhalmazaira. Azaz nagyon általános f-re és $E \subseteq I^s$-re vonatkozó feltételek mellett teljesül, hogy ha $(x_1, x_2 \dots)$ sorozat egyenletes eloszlású I^s-ben, akkor

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_E(x_n) f(x_n) = \frac{1}{|E|}\int_E f(t)dt,$$

ahol $\chi_E(x_n) = 1$ ha $x_n \in E$ és $\chi_E(x_n) = 0$ ha $x_n \notin E$, továbbá $|E|$ az E halmaz Lebesgue mértékét jelöli.

Általában - a gyakorlati alkalmazások során - azonban csak véges sok x_i pontunk van. Ezért szükség van egy mértékre amely azt méri, hogy adott véges sok pont „mennyire egyenletes eloszolású" I^s-ben.

F.1. Definíció. *Jelölje $\mathcal{J}$ az I^s részintervallumainak halmazát, azaz $\mathcal{J} = \{J : \ J = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_s, b_s],$ ahol $0 \leq a_i \leq b_i < 1$ minden i-re$\}$. Legyen*

$$A(J, N) \stackrel{def}{=} \sum_{n=1}^{N} \chi_E(x_n),$$

ahol $\chi_J(x_n) = 1$ ha $x_n \in J$ és $\chi_J(x_n) = 0$ ha $x_n \notin J$. Ekkor a D_N diszkrepancia:

$$D_N = \sup_{J \in \mathcal{J}} \left| \frac{A(J, N)}{N} - |J| \right|,$$

ahol $|J|$ a J halmaz Lebesgue mértékét jelöli, azaz $J = [a_1, b_1] \times \cdots \times [a_s, b_s]$ esetén $|J| = \prod_{i=1}^{s} (b_i - a_i)$.

A diszkrepancia segítségével jól karakterizálható az egyenletes eloszlás véges sok pont esetén. Így a diszkrepancia jól használható a Monte-Carlo módszerekben is, ahol a diszkrepancia segítségével becsülhető a $\int\limits_{I^s} f(t)dt$ és a $\sum_{n=1}^{N} f(x_n)$ kifejezések eltérése. Minél kisebb az $(x_1, x_2, \ldots, x_n)$ sorozat diszkrepanciája, annál élesebb becslés adható a

$$\left| \int\limits_{I^s} f(t)dt - \sum_{n=1}^{N} f(x_n) \right|$$

abszolút értékre. Ezért fontos, hogy minél több olyan $(x_1, x_2, \dots)$ sorozatot találjunk, melynek a diszkrepanciája minél alacsonyabb. Ezeket a véges - vagy esetleg végtelen - sorozatokat alacsony diszkrepanciájú sorozatoknak nevezzük. Így az alkalmazások kapcsán felmerülő fontos kérdés, hogy mennyire lehet alacsony egy sorozat diszkrepanciája. Chung [2] és Kiefer [7] eredménye szerint majdnem minden $(x_1, x_2, \dots, x_N)$ sorozatra

$$D_N = O(N^{-1/2}(\log\log N)^{1/2}). \tag{5}$$

A következőkben példákat mutatunk I-ben olyan $(x_1, x_2, \dots, x_N)$ sorozatokra, amelyekre

$$D_N = O(N^{-1/2}\log N). \tag{6}$$

Legyen α egy irracionális szám, és legyen $x_i = \{i\alpha\}$. Ekkor (6) valóban fenn áll. A másik példa az úgynevezett van der Corput sorozat [19], amely két ok miatt is nagyon fontos. Egyrészt a diszkrepanciája alacsonyabb mint az előző sorozaté, másrészt ez a sorozat csupán n számjegyeit használja, így a módszer könnyen adaptálható bináris számítógépekre. A van der Corput sorozatot [19] a következőképp definiálják: Legyen $g \geq 2$ egy természetes szám. Ekkor minden pozitív egész n egyértelműen felírható

$$n = \sum_{i=0}^{k} a_i g^i$$

alakban, ahol $a_i \in \{0, 1, 2, \ldots, g-1\}$. Legyen

$$\phi_g(n) = \sum_{i=0}^{k} a_i g^{-i-1}.$$

(Ekkor $0 \leq \phi_g(n) < 1$.) A van der Corput sorozat: $(\phi_g(0), \phi_g(1), \ldots)$. Jelenleg nem ismert olyan sorozat, amelynek a diszkrepanciája alacsonyabb lenne, mint a van der Corput sorozaté. (Erre a sorozatra is bizonyított (6), ahol az ordóban szereplő implicit konstans a lehető legalacsonyabb a jelenleg ismert konstrukciók diszkrepanciájára adott felső becslések között.)

Számos erős konstrukció ismert a többdimenziós esetben is (azaz, amikor $(x_1, x_2, \ldots)$ sorozat elemei I^s-ből valók), azonban ezeknek a konstrukcióknak a definíciója komplikáltabb, mint a fenn említett két egydimenziós konstrukcióé. Ezért ezeket a konstrukciókat itt nem ismertetjük.

F.2. Pszeudovéletlen $[0, 1)$ sorozatok

A Monte-Carlo módszerek során többnyire nem olyan - nagyon specifikus - sorozatot alkalmazunk, amelynek a diszkrepanciája a lehető legalacsonyabb, hanem úgynevezett pszeudovéletlen sorozatokat. Ezt - többek közt - az is alátámasztja, hogy majdnem minden N hosszú sorozatra fenn áll (5), míg konkrét konstrukciókra bizonyítani csak a jóval gyengébb (6) állítást tudjuk. Ebben a fejezetben a pszeudovéletlen $[0, 1)$ sorozatokat vizsgáljuk meg közelebbről. Ezeknek a sorozatoknak számos matematikai, fizikai és számí-

tástechnikai alkalmazása van.

Annak, hogy egy $[0,1)$ sorozat véletlennek tűnjön az alkalmazások szempontjából, két fő kritériuma van:

(i) A sorozatnak eleget kell tennie bizonyos eloszlási kritériumoknak.

(ii) Ezeknek az eloszlási tulajdonságoknak invariánsnak kell maradnia akkor is, amikor csak bizonyos részsorozatokat vizsgálunk.

(7)

Knuth: A Számítógépprogramozás művészete című könyvében [8] részletesen tárgyalja ezeket a feltételeket. Minimális elvárás az egyenletes eloszlás I^s-ben, de ez nem feltétlen elegendő. Gondoljunk például a van der Corput sorozatra, ahol az alkalmazott α irracionális szám 0-hoz közeli pozitív szám. Ennél a sorozatnál az egymást követő bitek nagy valószínűséggel a $[0,\frac{1}{2})$ és $[\frac{1}{2},1)$ intervallumok közül ugyanabba esnek. Ennek alapján azt mondjuk, hogy egy $(x_1, x_2, \dots)$ végtelen sorozat *teljesen* egyenletes eloszlású $[0,1)$-ben, ha az

$$((x_n, x_{n+1}, \dots, x_{n+s-1}) : \ n = 1, 2, \dots)$$

sorozat egyenletes eloszlású I^s-ben minden $s \geq 1$ egész számra (ld. az 1. fejezetben ismertetett serial teszt). Eddig még nem tárgyaltuk a pszeudovéletlenség (7)-ben említett (ii) elvét. Amennyiben minden részsorozatra megkövetelnénk a teljesen egyenletes eloszlást, úgy egyetlen pszeudovéletlen sorozat sem létezne, hiszen tekinthetjük például azt a részsorozatot, amely

$[0, \frac{1}{2})$ intervallumba eső elemeket tartalmazza; ez a sorozat nyilvánvalóan nem egyenletes eloszlású. Knuth: A Számítógépprogramozás művészete című könyvében [8] részletesen vizsgálja, hogy milyen részsorozatokra érdemes megkövetelni a teljesen egyenletes eloszlást.

Véges sorozatokra vonatkozó pszeudovéletlen vizsgálatok Kolmogorov: Three approaches to the definition of the concept „quality of information" című munkájára [9] épülnek. Ebben a megközelítésben akkor tekintünk egy sorozatot pszeudovéletlennek, ha csak „hosszú" programmal tudjuk leprogramozni egy Turing gépen.

Véges pszeudovéletlen sorozatokat generálnak fizikai módszerekkel is, azonban ezzel kapcsolatban felmerül néhány probléma: ilyen például az adatok tárolása vagy az, hogy szükséges a véletlenség gyakori ellenőrzése, mivel ezeket a sorozatokat nem számítógéppel generálták, hanem egy külső eszközzel.

A pszeudovéletlenség axiomatikusan bevezetett fogalmának a gyakorlati alkalmazások során derül ki egy gyenge pontja: nevezetesen egyetlen egy konkrét sorozat sem ismert, amely eleget tenne a kritériumoknak. Ezért úgy gondolják, hogy valójában elegendő azon sorozatokat vizsgálni, amelyek a gyakorlatban - számítógépek segítségével sem - nem különböztethetőek meg egy valódi véletlen sorozattól. Létezik néhány statisztikus teszt, amely ilyen szempontból vizsgálja a pszeudovéletlenséget. Ezek közül a tesztek közül

alább ismertetünk néhányat (több-kevesebb részletességgel):

Egyenletes eloszlás teszt. Minden N-re tekintjük a sorozat első N elemét: $x_1, x_2, \ldots, x_N$. A sorozat állja a tesztet, ha ezen sorozatok D_N diszkrepanciája minden N-re kicsi. Egy alternatív módszerben I-t részintervallumokra osztjuk, és azt nézzük, hogy hány elem esik a különböző részintervallumokba. Az így kapott adatokra alkalmazunk egy χ^2 próbát. Ez az úgynevezett *gyakoriság teszt.*

Hézag teszt. Legyen J az I intervallumnak egy fix részintervalluma. Ha egy $n \geq 1$-ra $x_{n-1} \notin J$ (vagy $n = 1$) és $x_n, x_{n+1}, \ldots, x_{n+k-1} \notin J$ de $x_{n+k} \in J$, akkor k hosszú hézagról beszélünk. Rögzített h-ra összeszámoljuk a h-nál rövidebb és a h vagy annál hosszabb hézagok számát. Ezekre az eredményekre alkalmazunk egy χ^2 próbát.

A *Futás teszt* a monoton részsorozatokat vizsgálja. Hasonlóan definiálható még a *permutáció teszt*, a *„serial-correlation" teszt*, a *spektrál teszt* és az 1. fejezetben ismertetett *serial teszt.*

F.3. Bináris sorozatok

A 0-1 sorozatoknak kiemelt jelentősége van a kriptográfiai alkalmazások során. Például - a már ismertett - Vernam féle titkosító eljárás használ 0-1 sorozatokat. Amennyiben a kulcsként alkalmazott 0-1 sorozat valódi vélet-

lensorozat (azaz a sorozat bitjeit egymástól függetlenül választjuk és minden bit $\frac{1}{2}-\frac{1}{2}$ valószínűséggel 0 vagy 1) úgy az úgynevezett *„one-time pad"* titkosító eljárásról beszélünk. A one-time pad óriási előnye, hogy további feltétel nélküli, teljes biztonságot garantál. Így bizonyos értelemben tökéletes titkosító algoritmus. Az eljárás egyetlen hátránya, hogy a kulcsnak azonos hosszúnak kell lennie a titkosítandó üzenettel, viszont nagyon nehéz hosszú valódi véletlen sorozatot generálni.

Eleinte -a hidegháború idején- fizikai módszerekkel, például diódával generáltak bináris sorozatokat. S bár a tapasztalat azt mutatja, hogy az ily módon generált sorozat valóban megfelelhet a matematikai modellnek, ezt nem tudjuk bizonyítani. A fizikai módszerrel történő véletlengenerálás további hátránya, hogy lassú és költséges folyamat. Ezért manapság számítógépekkel generálnak egy valódi véletlen kisebb titkos kulcsból egy hosszú „véletlennek tűnő" kulcsot, amely azonban várhatóan -még számítógépek segítségével sem- különböztethető meg egy valódi véletlen sorozattól. Azonban fontos megfogalmazni, hogy milyen véletlentulajdonságokat várunk egy ily módon konstruált pszeudovéletlen sorozattól.

1996-ban Christian Mauduit és Sárközy András [11] új, kvantitatív mértékeket vezetett be bináris sorozatok pszeudovéletlen tulajdonságainak vizsgálatára. A továbbiakban technikai okokból áttérünk a 0-1 sorozatokról a ± 1 sorozatok vizsgálatára. (A 0-1 és ± 1 sorozatok között egy-egy értelmű

megfeleltetés adható.)

F.2. Definíció. *Legyen* $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$ *egy* N *hosszú* ± 1 *sorozat. Ekkor az* eloszlási mértéket *a*

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

képlettel definiáljuk, ahol a maximum az összes olyan a, b, t*-n fut, ahol* $1 \leq a \leq a + tb \leq N$.

Az eloszlási mérték azt vizsgálja, hogy a számtani sorozatokban a $+1$ és -1-ek száma mennyire közel van. Azonban előfordulhat, hogy a sorozatban több bit között áll fenn összefüggés, például $(+1, +1)$ részsorozat lényegesen többször fordul elő, mint a $(-1, -1)$ részsorozat. A több bit egymás közötti függetlenségének vizsgálatára vezette be Christian Mauduit és Sárközy András a normalitás és a korreláció fogalmát.

F.3. Definíció. *Legyen* $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$ *egy* N *hosszú* ± 1 *sorozat.* $X = (x_1, x_2, \ldots, x_k) \in \{-1, +1\}^k$ *és* $1 \leq M \leq N - k + 1$ *esetén jelölje* $T(E_N, M, X)$ *a következő mennyiséget:*

$$T(E_N, M, X) = |n : \ 1 \leq n \leq M, \ (e_n, e_{n+1}, \ldots, e_{n+k-1}) = (x_1, x_2, \ldots, x_k)| .$$

Ekkor a k-adrendű normalitás mértéket *a*

$$N_k(E_N) = \max_{M,x} \left| T(E_N, M, X) - \frac{M}{2^k} \right|$$

kifejezéssel definiáljuk, ahol a maximum az összes olyan pozitív egész M-en és X szám k-ason fut, ahol $1 \leq M \leq N-k+1$ *és* $X = (x_1, x_2, \ldots, x_k) \in \{-1, +1\}^k$.

Christian Mauduit és Sárközy András [11] igazolta, hogy a normalitás mérték felülről becsülhető a korreláció mértékekkel. Ezeket a mértékeket a következőképpen definiáljuk.

F.4. Definíció. *Legyen* $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$ *egy* N *hosszú* ± 1 *sorozat. Ekkor a* k-adrendű korreláció mértéket *a*

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

képlettel definiáljuk, ahol a maximum az összes olyan M *egész számon és* $D = (d_1, d_2, \ldots, d_k)$ *szám* k*-ason fut, ahol* $0 \leq d_1 < d_2 < \cdots < d_k < M + d_k \leq N$.

Ekkor - Christian Mauduit és Sárközy András eredménye szerint -

$$N_k(E_N) \leq \max_{1 \leq t \leq k} C_t(E_N),$$

vagyis a korreláció segítségével a normalitás mérték jól becsülhető. (Ezért a legtöbb cikkben nem kezelik külön a normalitás mértéket.)

A kombinált mérték a k-adrendű korreláció és az eloszlási mérték általánosítása:

F.5. Definíció. *Legyen* $E_N = (e_1, e_2, \ldots, e_N) \in \{-1, +1\}^N$ *egy* N *hosszú*

± 1 *sorozat. Ekkor a* k-adrendű kombinált mértéket *a*

$$Q_k(E_N) = \max_{a,b,t,D} \left| \sum_{j=1}^{t} e_{a+bj+d_1} e_{a+bj+d_2} \cdots e_{a+bj+d_k} \right|$$

képlettel definiáljuk, ahol a maximum az összes olyan a, b, t *pozitív egész számon és* $D = (d_1, d_2, \ldots, d_k)$ *szám* k*-ason fut, ahol az összes előforduló* $a + bj + d_i$ *index az* $\{1, 2, \ldots, N\}$ *halmazba esik.*

A fenti mértékeken túl is lehet definiálni pszeudovéletlen mértékeket, azonban a túl sok mérték egyszerre nehezen kezelhető. Így a legtöbb alkalmazás során az eloszlási és korreláció mértékre szorítkoznak.

Az eloszlási és korreláció mérték maximuma N-hez közeli érték. J. Cassaigne, C. Mauduit és Sárközy A. [1] bebizonyította, hogy majdnem minden N hosszú sorozatra

$$W(E_N) < N^{1/2}(\log N)^c, \qquad C_k(E_N) < N^{1/2}(\log N)^{c_k}$$

fenn áll. Ez alapján egy E_N sorozatot erős pszeudovéletlen tulajdonságokkal rendelkezőnek nevezünk, ha

$$W(E_N) = o(N), \qquad C_k(E_N) = o(N).$$

A 2. fejezetben ismertetett konstrukciók erős pszeudovéletlen tulajdonságokkal rendelkeznek, ezekre az E_N sorozatokra

$$W(E_N) < N^{1/2}(\log N)^{c'}, \qquad C_k(E_N) < N^{1/2}(\log N)^{c'_k}$$

valóban fenn áll.

A véges bináris sorozatok elméletéről számos cikk született, rohamosan fejlődik a terület. Számos hazai és külföldi kutató kutat jelenleg is. Mostanában a pszeudovéletlen sorozatok mellett a többdimenziós pszeudovéletlen rácsok is bekerültek a kutatás középpontjába.

Hivatkozások

[1] J. Cassaigne, C. Mauduit és A. Sárközy, *On finite pseudorandom binary sequences VII: The measures of pseudorandomness*, Acta Arith. 103 (2002), 97-118.

[2] K. L. Chung, *An estimate concerning the Kolmogoroff limit distribution*, Trans. Amer. Math. Soc. 67 (1949), 36-50.

[3] L. Goubin, C. Mauduit és A. Sárközy, *Construction of large families of pseudorandom binary sequences*, J. Number Theory 106, (2004), 56-69, 2004.

[4] K. Gyarmati, *A note to the paper "On a fast version of a pseudorandom generator"*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 49 (2006), 143-149.

[5] K. Gyarmati, *On a family of pseudorandom binary sequences,* Period. Math. Hungar. 49 (2004), 45-63.

[6] K. Gyarmati, *On a fast version of a pseudorandom generator*, General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, Vol. 4123, Springer Berlin / Heidelberg 2006, 326-342,

[7] J. Kiefer, *On large deviations of the empiric d.f. of vector chance variables and a low of the iterated logarithm*, Pacific J. Math. 11 (1961), 649-660.

[8] D. E. Knuth, *A Számítógépprogramozás művészete*, 2. kötet, 2. kiadás, Műszaki Könyvkiadó, Budapest 1994.

[9] Kolmogorov, *Three approaches to the definition of the concept „quality of information"*, Problemy Peredaci Informacii 1 (1965), 3-11 (oroszul).

[10] D. H. Lehmer, *Mathematical methods in large-scale computing units*, Proc. 2nd Sympos. on Large-Scale Digital Calculating Machinery (Camebridge, MA, 1949), Harvard Univ. Press, Cambridge, MA, 1951, 141-146.

[11] C. Mauduit és A. Sárközy, *On finite pseudorandom binary sequence I: Measures of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365-377.

[12] C. Mauduit és A. Sárközy, *Construction of pseudorandom binary sequences by using the multiplicative inverse*, Acta Math. Hungar. 109 (2005), 75-107.

[13] A Menezes, P van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, 1997, CRC Press Inc. (internetről letölthető).

[14] H. Niederreiter, *On the distribution of pseudo-random numbers generated by the linear congruential method.*, Math. Comp. 26 (1972), 793-795.

[15] H. Niederreiter, *On the distribution of pseudo-random numbers generated by the linear congruential method. II*, Math. Comp. 28 (1974), 1117-1132.

[16] H. Niederreiter, *On the distribution of pseudo-random numbers generated by the linear congruential method. III*, Math. Comp. 30 (1976), 571-597.

[17] H. Niederreiter, *Quasi-Monte Carlo methods and pseudorandom numbers*, Bull. Amer. Math. Soc. 84 (1978), 957-1041.

[18] I. M. Szobol, *A Monte-Carlo módszerek alapjai*, Műszaki Könyvkiadó, Budapest, 1981.

[19] J. G. van der Corput, *Verteilungsfunktionen, I,II*, Nederl. Akad. Wetensch. Proc. 38 (1935), 813-821, 1058-1066.