

UN ALGORITM DE GENERARE A NUMERELOR PSEUDOALEATOARE

DE

A. KOVÁCS

(Cluj)

În lucrare se dă un algoritm de generare a numerelor pseudoaleatoare, precum și o prelucrare statistică a lor.

0. Pentru a modela procesele aleatoare care sînt legate de aplicațiile metodei Monte Carlo, este necesar să dispunem de un generator de numere aleatoare. Generarea cu ajutorul unui calculator poate fi făcută fie cu ajutorul unui dispozitiv special încadrat în calculator, fie pe baza unui algoritm. Numerele care se obțin pe baza acestui algoritm se numesc numere pseudoaleatoare. Acesta constituie doar o aproximație a numerelor aleatoare, însă avem posibilitatea de a le obține simplu cu ajutorul unui calculator.

Metodele clasice de generare a numerelor pseudoaleatoare — cum ar fi metoda lui J. von Neumann [3] sau metoda lui Lehmer [2] — sînt destul de anevoioase din punctul de vedere al programării pe un calculator cu virgulă fixă. Pe de altă parte șirul de numere pseudoaleatoare obținute pe baza algoritmului lui J. von Neumann are o perioadă foarte mică (perioada maximă fiind aproximativ 120). Șirul obținut pe baza algoritmului lui Lehmer prezintă o anumită structură cristalină, în sensul că numerele obținute sînt situate într-un număr relativ mic de hiperplane paralele.

În mod natural se pune problema elaborării unui algoritm de generare al acestor numere care să fie convenabil nu numai din punctul de vedere statistic ci și din punctul de vedere al programării.

1. După cum se știe calculatoarele electronice cu poziția fixă a virgulei lucrează cu numerele care satisfac condiția ca

$$2^{-p} \leq |x| \leq 1 - 2^{-p},$$

p fiind numărul biților într-un cuvânt. Prin urmare este avantajos un algoritm care să lucreze cu numerele reprezentate în sistemul de numerație binar, de forma

$$x = \sum_{i=1}^p x_i \cdot 2^{-i}, \text{ unde } x_i = 0 \text{ sau } x_i = 1.$$

Datorită particularității mașinii dacă $|x - y| < 2^{-p}$, atunci cele două numere se consideră ca fiind egale.

Fixînd o constantă arbitrară $k (0 < k < 1)$ și alegînd o valoare inițială $x_0 (0 < x_0 < 1)$ vom genera șirul $x_1, x_2, \dots, x_n, \dots$ printr-un procedeu inductiv.

Avînd pe

$$x_n = \sum_{j=1}^p x_j^{(n)} \cdot 2^{-j}, \quad k = \sum_{l=1}^p k_l \cdot 2^{-l}$$

termenul următor $x_{n+1} (n = 0, 1, \dots)$ se determină după cum urmează. Se formează produsul kx_n și se scrie acesta sub formă de fracție binară, obținînd

$$kx_n = \sum_{j=1}^p p_j \cdot 2^{-j} + \sum_{j=p+1}^{2p} p_j \cdot 2^{-j} = \sum_{j=1}^p p_j \cdot 2^{-j} + 2^{-p} \sum_{j=p+1}^{2p} p_j \cdot 2^{p-j}$$

și se ia în considerare în continuare coeficientul lui 2^{-p} :

$$\sum_{j=p+1}^{2p} p_j \cdot 2^{p-j} \equiv \sum_{j=1}^p r_j \cdot 2^{-j}.$$

Algoritmul de obținere al numărului $\sum_{j=1}^p r_j \cdot 2^{-j}$ este asemănător cu metoda, de generare propusă de Lehmer. Numărul astfel obținut se ridică la pătrat deci

$$\left(\sum_{j=1}^p r_j \cdot 2^{-j} \right)^2 = \sum_{j=1}^h y_j \cdot 2^{-j} + \sum_{j=h+1}^{h+p} y_j \cdot 2^{-j} + \sum_{j=h+p+1}^{2p} y_j \cdot 2^{-j},$$

unde $0 < h < p$. După acesta, înmulțind termenul din mijloc cu 2^h punem

$$2^h \cdot \sum_{j=h+1}^{h+p} y_j \cdot 2^{-j} \equiv \sum_{j=1}^p x_j^{(n+1)} \cdot 2^{-j}$$

și acest număr este termenul următor din șirul căutat, adică

$$x_{n+1} = \sum_{j=1}^p x_j^{(n+1)} \cdot 2^{-j}.$$

Partea a doua a acestui algoritm, după cum se vede, se aseamănă cu metoda lui Neumann.

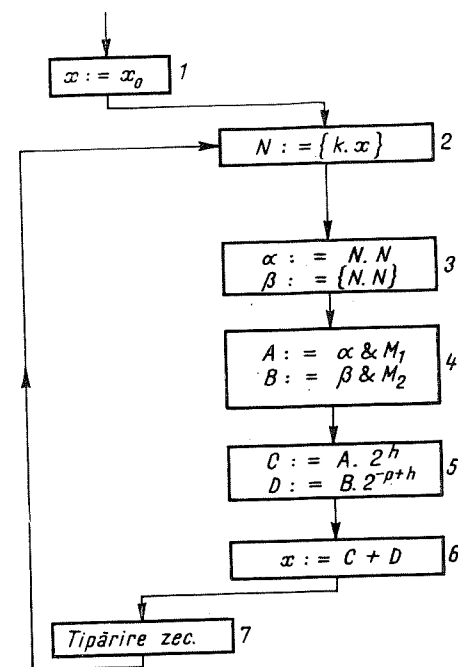


Fig. 1

Schema logică a algoritmului propus este prezentată în fig. 1, unde prin $\{z\}$ se înțelege numărul $\sum_{j=p+1}^{2p} z_j \cdot 2^{-j}$, iar numerele M_1 și M_2 cu ajutorul cărora se formează produsul logic sînt definite prin

$$M_1 = \sum_{j=h+1}^p 2^{-j} \text{ și } M_2 = \sum_{j=1}^h 2^{-j}.$$

Din punctul de vedere al programării metoda prezintă următoarele avantaje:

- a) $\{z\}$ se obține prin înmulțire de dublă lungime.
- b) Înmulțirile cu 2^l (respectiv 2^{-l}) revin la deplasările registrului spre stînga (dreapta) cu l biți.

Exemplu: Fie $h = 15$, $p = 33$ și să alegem

$$x_0 = 0,434294481_{(10)} = 0,33626754250_{(8)},$$

$$k = 0,314159265_{(10)} = 0,24066275660_{(8)},$$

ceea ce în sistemul binar se scrie

$$x_0 = 0, 011 011 110 010 110 111 101 100 010 101 000,$$

$$k = 0, 010 100 000 110 110 010 111 101 110 110 000.$$

Cu aceste valori inițiale se obține șirul următor:

0,408997579; 0,359708492; 0,880373975; 0,670923183; 0,505171567;
0,357581184; 0,692001888; 0,364442853; 0,235043807; 0,357655921;
0,229525631; ...

2. Pentru studiul statistic al numerelor obținute se pune problema determinării caracteristicilor statistice precum și cercetarea normalității [1], repartiției și a periodicității șirului obținut.

Luind $n = 10^3$ termeni din șirul $x_1, x_2, \dots$, pentru valorile fixate ale lui k și x_0 din exemplul precedent se obține:

— media $M(X) = 0,5034$,

— momentul inițial de ordin doi $M(X^2) = 0,3344$,

— abaterea mediei pătratică $\sigma(X) = 0,2846$.

Valorile teoretice sînt respectiv 0,5; 0,3333; 0,2888.

Pe baza definiției lui É. Borel un șir de cifre

$$a_1, a_2, a_3, \dots \quad (1)$$

este simplu normal, dacă frecvența de apariție a unei cifre oarecare tinde spre $\frac{1}{g}$, g fiind baza sistemului de numerație, adică limita nu depinde de cifra aleasă. Să asociem șirului considerat k — întinderea corespunzătoare:

$$(a_1, a_2, \dots, a_k); (a_2, a_3, \dots, a_{k+1}); (a_3, a_4, \dots, a_{k+2}); \dots$$

Se spune că șirul (1) este un șir normal dacă pentru orice număr natural k , k — întinderea șirului (1) este un șir simplu normal. Aici prin „cifră” se înțelege punctul $(a_{j+1}, \dots, a_{j+k})$ din spațiul aritmetic k — dimensional.

Prelucrarea statistică a șirului $\{x_n\}$ ne arată că proprietatea de normalitate este verificată. Programul corespunzător revine la efectuarea a $N - k + 1$ comparații pentru fiecare din cele g^k posibilități, unde $N = \frac{N - k + 1}{g^k}$. Frecvențele corespunzătoare sînt apropiate de cele teoretice:

De exemplu pentru $n = 100$, $p = 33$, $g = 2$ se obține:

$k = 2$ 00 ... 803
 01 ... 840
 10 ... 839
 11 ... 817

 $k = 3$ 000 ... 392
 011 ... 411

Frecvențele teoretice sînt 824,75 respectiv 412.

În vederea studiului repartiției șirului $\{x_n\}$ să împărțim intervalul $[0,1]$ în m părți egale cu ajutorul punctelor de coordonate $\xi_j = \frac{j}{m}$ ($j = 1, 2, \dots, m-1$). Să notăm cu v_j numărul termenilor x_k din șirul $x_1, x_2, \dots, x_n$ pentru care

$$x_k \in I_j = \left(\frac{j-1}{m}, \frac{j}{m} \right),$$

iar cu $n \cdot p_j$ valoarea medie a lui v_j în repartiția ipotetică. În cazul de față repartiția ipotetică fiind repartiția uniformă în intervalul $[0, 1]$, avem:

$$p_1 = p_2 = \dots = p_m = \frac{1}{m}.$$

Utilizînd testul χ^2

$$\chi_{m-1}^2 = \sum_{i=1}^m \frac{\left(v_i - \frac{n}{m} \right)^2}{\frac{n}{m}} = \frac{1}{m \cdot n} \sum_{i=1}^m (v_i - n)^2$$

și fixînd pe $m = 10$ ajungem la următoarele rezultate

n	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	χ_9^2	ε
600	51	61	63	61	67	62	54	58	60	55	2,17	0,99
700	68	66	69	74	80	74	67	69	69	64	2,85	0,96
800	80	76	79	87	90	83	75	75	77	78	2,97	0,98
900	89	85	87	98	100	93	83	85	84	96	3,93	0,95
1000	96	94	98	106	107	108	95	97	94	105	3	0,98

unde ε este coeficientul de încredere.

Lungimea medie a părții neperiodice din șirul de numere pseudoaleatoare generat cu algoritmul propus mai sus este egală aproximativ cu 10^3 .

Algoritmul a fost verificat cu mașina electronică de calcul DACICC-1, pentru care $p = 33$. Menționăm că programul în codul calculatorului DACICC-1 conține doar 13 cuvinte în timp ce programele corespunzătoare metodelor clasice sînt esențial mai lungi. Numerele generate pe baza algoritmului propus au fost utilizate la simularea mersului aleator pe o axă respectiv în plan.

Primit la redacție la 5.02.1970

EIN ALGORITHMUS FÜR DIE ERZEUGUNG DER PSEUDOZUFALLSVARIABLEN

(ZUSAMMENFASSUNG)

Die klassischen Methoden für die Erzeugung der Pseudozufallsvariablen — wie z.B. die Methode von Lehmer und J. von Neumann — sind vom Standpunkt ihrer Programmierung auf einem Rechenautomaten mit fester Lage des Kommas betrachtet, ziemlich schwierig.

In der Arbeit wird ein neuer Algorithmus für die Erzeugung der Pseudozufallsvariablen gegeben, der den Vorteil hat, daßer sehr leicht auf einem elektronischen Rechenautomaten programmiert werden kann, indem man die Charakteristiken des Rechenautomaten ausnützt.

Im zweiten Teil der Arbeit wird die erhaltene Folge bezüglich ihrer Normalität und Verteilung untersucht.

BIBLIOGRAFIE

1. BERTI ȘT. N., *Probleme legate de studiul sirurilor normale de cifre*. St. cerc. mat., 17, 9, 1425 — 1441 (1965).
 2. LEHMER D. H., *Mathematical methods in large-scale computing units*. Proc. sympos. on large-scale digital calcul. machinery, 141 — 145, Harvard Univ. press, 1949.
 3. NEUMANN J. VON, *Various techniques used in connection with random digits*. NBS appl. math. series, 12, 36 — 38, New York, 1951.
-