Generatori de numere aleatoare

Curs 2

Bibliografie suplimentară:

- Knuth, D. E.(1983) *Tratat de programare a calculatoarelor, Vol. 2 - Algoritmi seminumerici*, Editura Tehnică.
- Knuth D.E. (1974) *Tratat de programare a calculatoarelor, Vol 1 - Algoritmi fundamentali*, Editura Tehnică.
- Văduva, I (1977) *Modele de simulare cu calculatorul*, Ed. Tehnică.

Recapitularea unor noţiuni probabiliste

Experiment aleator= un experiment al cărui rezultat nu este cunoscut înainte.
Spaţiu de selecţie (S)= spatiul tuturor rezultatelor posibile.
Eveniment (A)= orice submulţime a spaţiului de selecţie. Dacă rezultatul unui experiment aparţine lui A, atunci se spune că a avut loc A.
Probabilitate= probabilitatea de apariţie a evenimentului A $P(A)$ este un număr care:

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$, $\forall n$, $A_i \cap A_j = \phi$

$\Rightarrow 1 = P(S) = P(A \cup A^c) = P(A) + P(A^c)$

Probabilitate condiţionată= $P(A|B) = \frac{P(A \cap B)}{P\{B\}}$
Evenimente independente= $P(A \cap B) = P(A)P(B)$

Recapitularea unor noţiuni probabiliste

Variabilă aleatoare $X : S \to \mathbb{R}$ (discrete şi continue)
Funcţie de repartiţie $F(x) = P\{X \leq x\}$
Funcţie şi densitate de probabilitate $p(x) = P\{X = x\}$, $f(x) = F'(x)$,
$P\{X \in C\} = \int_C f(x)dx$, $F(a) = \int_{-\infty}^{a} f(x)dx$
Două variabile aleatoare $F(x, y) = P\{X \leq x, Y \leq y\}$,
$p(x, y) = P\{X = x, Y = y\}$,
$P\{X \in C, Y \in D\} = \int\int_{x \in C, y \in D} f(x, y)dxdy$
Variabile aleatoare independente
$P\{X \in C, Y \in D\} = P\{X \in C\}P\{Y \in D\}$,
$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\}$, $f(x, y) = f_X(x)f_Y(y)$

Numere aleatoare

Şir de numere aleatoare: (definiţie intuitivă) un sir de numere alese la întâmplare astfel încât se cunoaşte probabilitatea de apariţie a fiecărui număr într-o succesiune de valori dată.

Şirurile de numere aleatoare au aplicaţii în: criptografie, simulare, etc.

În general pentru şirurile de numere aleatoare probabilitatea de apariţie a unei valori corespunde **repartiţiei uniforme**.

Repartiţia uniformă: (intuitiv) Toate valorile sunt egal probabile.

În **simulare:** Şiruri de numere aleatoare $\Rightarrow$ Valori ale variabilelor aleatoare $\Rightarrow$ Model de simulare

Numere aleatoare: Valori ale unor variabile aleatoare uniforme pe $[0, 1]$.

Şirurile de numere aleatoare pot fi obţinute din:

- tabele - greu de implementat;
- fenomene fizice (cea mai buna sursă de aleator, de exemplu: intervalele de timp dintre apăsarea unei taste şi mişcarea mouse-ului) - nu se pot refolosi;
- algoritmi.

Algoritmi de numere aleatoare

Se bazează pe utilizarea unei **valori iniţiale (sămâmnţă)** şi a unei **relaţii de recurenţă** cu ajutorul căreia se obţin celelalte valori ale şirului.

Numere **pseudo-aleatoare**: numerele obţinute nu sunt chiar aleatoare pentru că se bazează pe o relaţie de recurenţă. Numerele produse trebuie să aibă două proprietăţi statistice importante:

- uniformitate;
- independenţă.

Majoritatea algoritmilor generează X_n numere întregi între 0 şi $m-1$ (de obicei $m-1$ este valoarea maximă a tipului întreg memorat în calculator) şi apoi se iau:

$$U_n = \frac{X_n}{m}$$

uniforme pe $[0, 1]$.

Algoritmi de numere aleatoare

Trebuie să aibă următoarele **proprietăţi**:

- Rapiditate;
- Portabilitate de diverse calculatoare;
- Şirul de numere produs
  - să aibă o perioadă mare;
  - să fie cât mai aproape de independenţă şi uniformitate;
  - să fie reproductibil.

Metoda părţii din mijloc a pătratului

- Prima metodă propusă pentru a fi implementată pe calculatoare. A fost descrisă de John von Neumann în 1946.
- Are doar interes istoric.
- Presupunem că vrem să generăm numere aleatoare cu cel mult i cifre. Atunci:

$$X_{n+1} = \text{cele } i \text{ cifre din mijloc ale lui } X_n^2$$

Algoritmi de numere aleatoare

- Şirul tinde să se stabilizeze în scurte cicluri de elemente.
- De exemplu: 43, 84, 05, 02, 00, 00,.... (se pun 0-uri în faţa numerelor care nu au patru sau două cifre).

Metoda Fibonacci

- Doar interes istoric;
- Se bazează pe relaţia

$$X_{n+1} = (X_n + X_{n-1}) \mod m;$$

- numerele produse nu sunt destul de aleatoare.

Alte metode:

- metoda regiştrilor de translaţie (shift register), metode combinate.

Metoda congruenţială liniară

- Este folosit cel mai frecvent.
- Se bazează pe relaţia:

$$X_{n+1} = (aX_n + c) \mod m \tag{1}$$

  unde
  - $m > 0$ = modulul;
  - a = multiplicatorul;
  - c = incrementul;
  - X_0 = termenul iniţial.
- Fie $b = a - 1$. Se presupune $a \geq 2$ şi $b \geq 1$, pentru că pentru $a = 0$ şi $a = 1$ nu se obţin şiruri aleatoare.
- Din (1), pentru $k \geq 0$ şi $n \geq 0$, rezultă că:

$$X_{n+k} = (aX_{n+k-1} + c) \mod m$$

  şi prin urmare relaţia dintre termenii şirului aflaţi la distanta k este:

$$
\begin{aligned}
X_{n+k} &= [a(aX_{n+k-2}+c) \mod m + c] \mod m \\
&= [a^2 X_{n+k-2} + (a+1)c] \mod m \\
&= [a^3 X_{n+k-3} + (a^2+a+1)c] \mod m \\
&= .... \\
&= [a^k X_n + (a^{k-1} + a^{k-2} + ... + a + 1)c] \mod m \\
&= \left[a^k X_n + \frac{a^k - 1}{a-1} c\right] \mod m
\end{aligned}
$$

relaţie utila pentru alegerea valorilor care caracterizează şirul.

Alegerea modulului

- m: trebuie să fie suficient de mare şi să asigure o complexitate scăzută a calculului.
- o alegere convenabilă a lui m ar fi w = dimensiunea cuvântului calculatorului $(= 2^x$ unde x este numărul de biti pe care sunt construiţi regiştrii calculatorului$)$.
- alte alegeri: $m = w \pm 1$ sau m = cel mai mare număr prim mai mic decât w.

Alegerea multiplicatorului

Se alege a pentru un m oarecare astfel încât pentru orice valoare a lui X_0 să rezulte un generator de perioadă maximă.

Teorema 1. *Şirul congruenţial liniar definit de m, a, c, şi X_0 are perioada de lungime maximă m dacă şi numai dacă:*

1. *c şi m sunt două numere întregi prime între ele;*
2. *$b = a - 1$ este un multiplu de p, pentru orice număr prim p care-l divide pe m.*
3. *b este multiplu de 4 dacă m este multiplu de 4.*

Datorită următoarei leme este suficientă demonstrarea teoremei pentru m putere a unui număr prim.

Lemă 1. *Fie descompunerea lui m în factori primi:*

$$m = p_1^{e_1} ... p_t^{e_t}. \tag{2}$$

Lungimea λ a perioadei şirului congruenţial liniar definit de (X_0, a, c, m) este cel mai mic multiplu comun al lungimilor λ_j ale perioadelor şirurilor congruenţiale liniare $(X_0 \mod p_j^{e_j}, a \mod p_j^{e_j}, c \mod p_j^{e_j}, p_j^{e_j})$, $1 \leq j \leq t$.

De aici rezultă că este suficientă demonstrarea teoremei pentru m, putere a unui numar prim:

$$p_1^{e_1} ... p_t^{e_t} = \lambda = \text{c.m.m.m.c}\{\lambda_1, ..., \lambda_t\} \leq p_1^{e_1} ... p_t^{e_t} \tag{3}$$

iar această relaţie poate avea loc dacă şi numai dacă $\lambda_j = p_j^{e_j}$ pentru $\forall j$, $1 \leq j \leq t$. De aceea se poate presupune $m = p^e$, unde p este un număr prim iar e este un număr întreg pozitiv.

Perioada poate avea lungime m dacă şi numai dacă orice număr întreg din $[0, m)$ apare în cadrul perioadei o singură dată. Dacă luăm $X_0 = 0$, atunci:

$$X_n = \left(\frac{a^n - 1}{a - 1}\right) c \mod m$$

Dacă c şi m nu sunt prime între ele, atunci în acest şir nu poate exista 1. Prin urmare condiţia 1. din teoremă este necesară. Demonstrarea teoremei se reduce la demonstrarea următoarei leme:

Lemă 2. *Presupunem că* $1 < a < p^e$*, cu* p *număr prim. Dacă* λ *este cel mai mic număr întreg pozitiv pentru care*

$$\frac{a^\lambda - 1}{a - 1} \equiv 0 \mod p^e$$

atunci

$$\lambda = p^e$$

dacă şi numai dacă:

- *pentru* $p = 2$ $a \equiv 1 \mod 4$*;*
- *pentru* $p > 2$ $a \equiv 1 \mod p$*.*

Această lemă se demonstrează aplicând de mai multe ori următoarea lemă:

Lemă 3. *Fie* p *un număr prim şi fie* e *un număr întreg pozitiv cu* $p^e > 2$. *Dacă*

$$x \equiv 1 \pmod{p^e}, \;\; x \not\equiv 1 \pmod{p^{e+1}} \tag{4}$$

atunci

$$x^p \equiv 1 \pmod{p^{e+1}}, \;\; x^p \not\equiv 1 \pmod{p^{e+2}} \tag{5}$$

Generatorul multiplicativ congruenţial

Este un generator liniar congruenţial cu $c = 0$:

$$X_{n+1} = aX_n \mod m \tag{6}$$

Observăm că X_n şi m trebuie să fie prime între ele, pentru că altfel generatorul ar deveni un şir de 0. Prin urmare lungimea perioadei poate fi maxim $\varphi(m)$, numărul numerelor întregi cuprinse între 0 şi m, prime cu m.

Putem să presupunem din nou că $m = p^e$ cu p = nr. prim şi e întreg pozitiv. Avem:

$$X_n = a^n X_0 \mod p^e$$

Dacă a este multiplu de p, atunci perioada are lungime 1 $\Rightarrow$ a trebuie să fie prim cu p.

Testarea şirurilor de numere aleatoare se face cu **teste statistice**:

- Teste de frecvenţă (testează repartiţia uniformă pe care trebuie să o aibă numerele): testul χ^2, testul Kolmogorov-Smirnov.
- Teste de independenţă.