

Generátory pseudonáhodných čísel

Jakub Galgonek
Michal Heppler
Petr Janata
Alexandr Kazda
Tomáš Urban

Za skutečně náhodnou posloupnost délky n považujeme posloupnost, k jejímuž vyjádření je potřeba všech n čísel. Nelze ji tedy vyjádřit pomocí několika parametrů a podmínek.

Náhodná čísla využíváme v simulacích, kde chceme, aby určitá situace nastala s danou pravděpodobností. Např. simulujeme-li růst stromu, chování zvířat, vývoj trhu nebo integrujeme-li složitou funkci, která se nedá vyjádřit primitivní funkcí (metoda Monte Carlo¹).

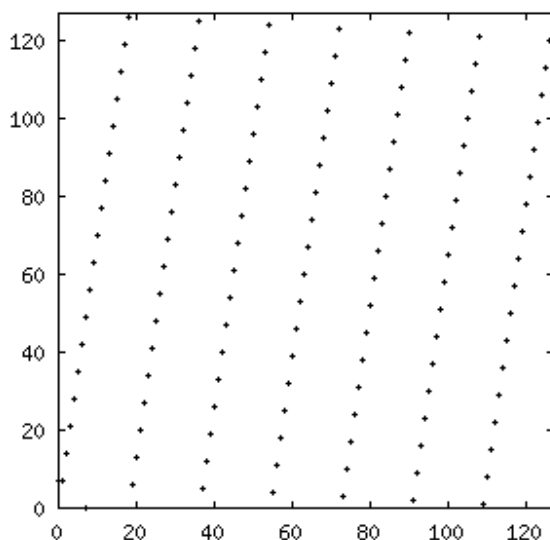
Generátory náhodných čísel užívají deterministický algoritmus jímž se snaží přiblížit skutečné náhodě. Kromě algoritmu potřebuje každý generátor inicializační hodnotu (často získané z nedeterministického systému, např. ze systémového času), což je počáteční hodnota, ze které se určují opakováním nějakého postupu další. Pro stejné inicializační hodnoty dostaneme i stejné posloupnosti. Většina generátorů vytváří rovnoměrné rozdělení pravděpodobnosti, které se dá upravit podle požadavků např. na normální rozdělení.

Lineární kongruentní generátory (LCG)

Čísla x se generují podle vzorce: $x_{i+1} = (a \cdot x_i + b) \bmod m$, kde a , b , m jsou přirozené konstanty (na jejich vhodné volbě závisí kvalita generátoru).

- Výsledkem je řada celých čísel o velikosti od 0 až do $m-1$.
- Výsledná řada je periodická, perioda řady je nejvýše m .
- I u dobrého LCG generátoru však body $[x_i; x_{i+1}]$ nevyplňují plochu hustě.

Např. pro $x_{i+1} = (7 \cdot x_i + 0) \bmod 127$



Aby nebylo nutné brát neustále nové inicializační hodnoty (a také kvůli zvýšení tzv. kryptografické bezpečnosti), je délka periody důležitý parametr generátoru. Zvýšení periody se dá dosáhnout buď zvyšováním m (což způsobuje pro velká m snížení rychlosti generátoru), nebo spojením několika generátorů.

Nejjednodušší takové spojení je prosté střídání dvou generátorů, liché prvky generuje první generátor, sudé druhý. Tato metoda zvyšuje maximální periodu na $m_1 \cdot m_2$ (pokud jsou m_1, m_2 nesoudělná). Tato metoda je rychlejší než vzít jeden generátor s $m' = m_1 \cdot m_2$.

Složitější je spojení tří LCG (každý má své parametry) tak, že v každém kroku vygeneruje každý z nich jednu hodnotu x_i, y_i, z_i a výsledná hodnota w_i se spočítá jako

$$w_i = \left(\frac{x_i}{m_x} + \frac{y_i}{m_y} + \frac{z_i}{m_z} \right) \bmod 1 \quad (x \bmod 1 \text{ vrací desetinnou část čísla } x, \text{ čili výsledkem jsou}$$

racionalní čísla v intervalu $\langle 0, 1 \rangle$). Pro tento generátor jsou zvlášť výhodná čísla objevená

Whichmanem a Hillem jsou to:

$$a_1=171, a_2=172, a_3=170$$

$$m_1=30269, m_2=30307, m_3=30323$$

$$b_1=b_2=b_3=0$$

Perioda je nejvýše $m_1 \cdot m_2 \cdot m_3$, rozdělení hodnot je rovnoměrné.

Kromě toho existují i neperiodické generátory. Například APRNG (Aperiodic Random Number Generator) používá dva LCG a nekonečný dvouprvkový aperiodický řetězec², který určuje, ze kterého generátoru brát hodnotu x_i (pokud je na i -tém místě první prvek, generuje x_i první generátor, pokud druhý, generuje druhý). Protože aperiodickou posloupnost (či aspoň její část) je třeba udržovat v paměti, má tento generátor vyšší požadavky na paměť, než prostý LCG (časová složitost je o také o trochu vyšší, kvůli generování aperiodické posloupnosti), ale zato vytváří skutečně aperiodickou posloupnost a stačí tedy zadat inicializační hodnotu pouze jednou.

Pozn.: 1) Monte Carlo je algoritmus pro přibližný výpočet určitého integrálu nějaké funkce $f(x)$. Bez újmny na obecnosti lze předpokládat, že integrujeme kladnou funkci. Náhodně se určí N bodů $[x, y]$, kde $x \in \langle a, b \rangle$ (a, b jsou meze integrace) a y je omezeno shora a zdola (necht' např. $y \in \langle c, d \rangle$). Poté se spočítají body ležící pod grafem funkce $f(x)$ a výsledek n_0 se vydělí N a dostaneme tak část obsahu obdelníku se stranami $b-a, d-c$, čili

$$\int_a^b f(x) = \frac{n_0}{N} (b-a)(d-c).$$

2) Takovýmto řetězcem je např. fibonacciho řetězec získaný opakovanou substitucí $A \rightarrow ABA, B \rightarrow AB$, přičemž začínáme písmenem A .

$A \rightarrow ABA \rightarrow ABAABABA \rightarrow ABAABABAABAABABAABABA \rightarrow \dots$