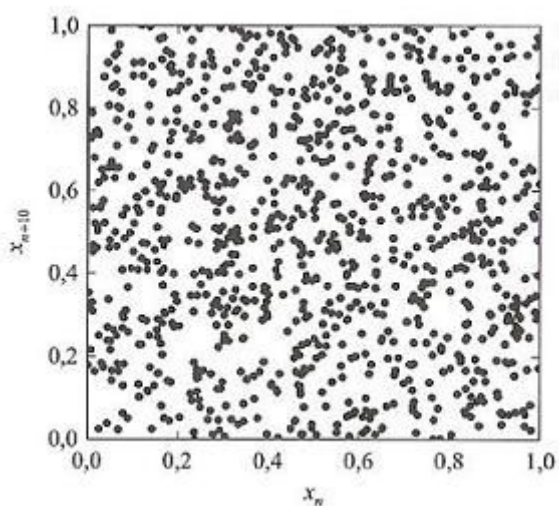


Przedmiot specjalizacyjny II

„Generatory liczb losowych”



Rozkład punktów (x_n, x_{n+10}) dla $x_i \in [0,1]$.
Na rysunku przedstawiono 1000 punktów.

Ogólnie o liczbach losowych

Wiele procesów obserwowanych przez nas w przyrodzie, technice, ekonomii czy życiu społecznym sprawia wrażenie zjawisk losowych, a więc takich, dla których nie potrafimy przewidzieć ich przyszłego przebiegu ani nie potrafimy ustalić przyczyn, które je wywołały.

Powodem tego może być brak informacji dotyczących danego zjawiska, nieznaną stopień precyzji dostępnych informacji lub też błędy obserwacji uniemożliwiające precyzyjną jego identyfikację. Na przeszkodzie może stać tu brak technicznych możliwości uzyskania dostępnych informacji lub niemożność wykonania jakichś istotnych pomiarów. Przyczyna losowości zjawiska może być też inna: jego specyficzne cechy fizyczne albo niezmierna komplikacja niemożliwa do objęcia żadnym zdeterminowanym modelem. W każdej z przedstawionych sytuacji, niezależnie od tego, czy udaje się nam ustalić, jakie są przyczyny losowości zjawiska, możemy spróbować opisać to zjawisko ilościowo, wykorzystując do tego celu pojęcie prawdopodobieństwa rozumianego jako ilościowa miara niepewności (losowości).

Podobnie jak w fizyce i przyrodzie, efekty losowe mogą występować również w świecie liczb. Wśród liczb naturalnych mamy do czynienia z ciągami liczb losowych, czyli takich, których pojawienie się nie może być z pewnością przewidziane, a struktura nie może być opisane żadnym określonym wzorcem.

W ogólnej sytuacji, ciągiem liczb losowych (ogólniej: losowym ciągiem znaków) nazwiemy taki ciąg, którego nie można zapisać za pomocą algorytmu w postaci krótszej od samego ciągu. Na podstawie takiego ciągu nie można stworzyć żadnych reguł, które pozwalałyby odtworzyć ten ciąg bez znajomości wszystkich jego wyrazów. Nie można również podać żadnego wyrazu tego ciągu na podstawie znajomości innych (wszystkich pozostałych) wyrazów tego ciągu.

Analogicznie jak w przyrodzie, możliwa jest jednak i taka sytuacja, że istnieją reguły (krótkie w porównaniu z długością ciągu) opisujące sposób wypisania wszystkich jego

wyrazów, jednak jedynie nasza niewiedza nie pozwala ich zidentyfikować. Ciąg taki nazwiemy pseudolosowym i w wielu sytuacjach będziemy go mogli traktować tak jak losowy ciąg liczb.

Wykorzystanie liczb losowych

Istnieją losowe ciągi liczb. Powstaje pytanie: jak można je wykorzystać dla celów praktycznych? Liczby losowe przydają się wszędzie tam, gdzie efekt przypadkowy, losowy, jest lepszy niż działanie zdeterminowane przez człowieka, a więc obciążone jego subiektywną oceną sytuacji. Liczby losowe można wykorzystać w reprezentatywnych badaniach statystycznych (przy losowaniu próby z populacji generalnej lub, szerzej, planowaniu schematu losowania), a zatem w zagadnieniach statystycznej kontroli jakości produktów, wszelkich badaniach ekonomicznych, społecznych, marketingowych itp. W naukach eksperymentalnych liczb losowych możemy użyć w zagadnieniach planowania eksperymentu.

Innym sposobem wykorzystania liczb losowych są wszelkie badania symulacyjne. Są to metody wykorzystujące techniki obliczeniowe, w których przedstawiamy przebieg realnego zjawiska, opisanego odpowiednimi równaniami, uwzględniając wpływające na nie czynniki losowe. Badania takie są często jedynym sposobem ilościowej analizy skomplikowanego procesu technologicznego lub zjawiska przyrodniczego. Podobnie, liczby losowe są źródłem losowości stwarzającej złudzenie realizmu we wszelkich popularnych grach komputerowych, inteligentnych automatach do gier zręcznościowych, trenerach oraz profesjonalnych grach strategicznych, to znaczy w programach umożliwiających wirtualny udział gracza w operacjach wojennych, ekonomicznych lub społecznych. Liczby losowe mogą ponadto służyć do budowania modeli skomplikowanych obiektów geometrycznych, na przykład fraktali losowych, wzorów powierzchni itp. Również badania symulacyjne statystyki matematycznej, tak zwane metody bootstrapowe lub sznurowadłowe, wymagają zastosowania liczb losowych.

Liczby losowe są także niezbędnym elementem metod Monte Carlo, czyli wykorzystania metod probabilistycznych w obliczeniach przeprowadzanych zwykle metodami deterministycznymi, na przykład w przybliżonym obliczaniu całek

wielowymiarowych, rozwiązywaniu równań różniczkowych i algebraicznych, optymalizacji (często sprowadzającej się do szukania minimum funkcji), algorytmach genetycznych itd.

Zastosowaniem liczb losowych, które w ostatnim czasie zyskało ogromnie na znaczeniu, jest kryptografia, czy też szerzej rozumiana ochrona informacji.

Rodzaje generatorów

W wielu zastosowaniach z zakresu symulacji numerycznych potrzebne jest użycie generatorów liczb losowych albo do wybrania konfiguracji wyjściowej, albo do samego przeprowadzenia symulacji. W programie komputerowym nic nie jest *losowe*. Komputer wykona program w dokładnie taki sam sposób jak poprzednio, jeśli zostanie dokładnie tak samo uruchomiony. Jak wcześniej wspomniałem generatory liczb losowych są więc w istocie rzeczy generatorami liczb pseudolosowych, które generują szeregi liczb o bardzo długim okresie powtarzalności i rozkładzie zbliżonym do zadanego.

Poniżej przedstawię niektóre podstawowe generatory liczb losowych używane w obliczeniach i symulacjach komputerowych.

Generatory liczb losowych o rozkładzie jednostajnym

Najbardziej użyteczne są generatory dające jednostajny rozkład liczb losowych z przedziału $[0,1]$. Trzy najważniejsze kryteria, jakie powinien spełniać dobry generator liczb losowych o rozkładzie jednostajnym, są następujące: (Park i Miller, 1998).

- I. Dobry generator powinien mieć bardzo długi okres powtarzalności, najlepiej bliski zakresu liczb całkowitych danego komputera. Na przykład dla komputerów 32-bitowych dobry generator powinien mieć okres bliski liczby $2^{31}-1=2\ 147\ 483\ 647$, gdyż zakres liczb całkowitych dla komputera 32-bitowego to przedział $[-2^{31}, 2^{31}-1]$ (jeden bit służy do zapisu znaku liczby całkowitej).

II. Dobry generator powinien charakteryzować się dobrą *losowością*. Korelacja pomiędzy wszystkimi, kolejno generowanymi liczbami powinna być możliwie mała. Innymi słowy, funkcje korelacji $\langle x_{n_1} x_{n_2} \dots x_{n_l} \rangle (X_{n_1} X_{n_2} \dots X_{n_l})$ dla $l = 2, 3, 4, \dots$, gdzie $\langle A \rangle$ oznacza średnią statystyczną zmiennej A , powinny być bardzo małe. Dobrym sposobem przekonania się, jaki jest charakter funkcji korelacji $\langle x_n x_{n+l} \rangle$, jest przedstawienie liczb X_n i X_{n+l} jako współrzędnych punktów na płaszczyźnie. Dobry generator liczb losowych powinien dawać punkty o bardzo dużej równomierności rozkładu dla dowolnego $l \neq 0$. Dla złego generatora punkty mogą układać się w paski czy siatki lub wykazywać inne niejednorodności rozkładu.

III. Dobry generator powinien być szybki. W praktyce potrzeba bardzo wielu liczb losowych, by otrzymać wyniki o dobrej jakości statystycznej. W związku z tym szybkość generatora może być bardzo istotna w pewnych rodzajach zastosowań.

W najprostszym generatorze liczb losowych o rozkładzie jednostajnym wykorzystuje się tzw. liniowy algorytm kongruentny. Kolejne liczby losowe otrzymuje się z równania

$$x_{n+1} = (ax_n + b) \bmod c$$

gdzie liczby a , b i c nazywa się liczbami *magicznymi*. Od ich wyboru zależy jakość generatora. Jeden z takich zestawów liczb, mianowicie $a = 7^5 = 16807$, $b = 0$, $c = 2^{31} - 1 = 2\,147\,483\,647$, został wielokrotnie sprawdzony i okazał się doskonałym wyborem dla komputerów 32-bitowych. Daje największy możliwy okres $2^{31} - 1$ i jest bardzo szybki.

Implementacja takiego generatora liczb losowych na komputerze nie zawsze jest trywialna, ponieważ różne komputery mają różne zakresy liczb całkowitych. Na przykład dla większości komputerów 32-bitowych zakresem liczb całkowitych jest przedział $[-2^{31}, 2^{31}-1]$. Gdy w wyniku obliczeń uzyskuje się liczbę spoza tego zakresu, komputer zamienia ją na zero.

Poniższa funkcja zewnętrzna spełnia funkcję omówionego wyżej generatora liczb losowych o rozkładzie jednostajnym.

```
FUNCTION RANF()  
DATA IA/16807/, IC/2147483647/, IQ/127773/, IR/2836/  
COMMON /CSEED/ ISEED  
IH = ISEED/IQ  
IL = MOD(ISEED, IQ)  
IT = IA*IL-IR*IH  
IF(IT.GT.O) THEN  
    ISEED = IT  
ELSE  
    ISEED = IC+IT  
END IF  
RANF = ISEED/FLOAT(IC)  
RETURN  
END
```

W tym programie $IQ = IC/IA$, $IR = \text{MOD}(IC, IA)$, a $ISEED \in [1, 2^{31} - 1]$. Park i Miller (1988) podali program w języku Pascal, który jest bardzo podobny do powyższej funkcji. Można łatwo stwierdzić, że powyższa funkcja przesuwaa liczby spoza zakresu o $c = 2^{31} - 1$ na dowolnym komputerze o słowie 32-bitowym lub dłuższym. Aby móc użyć tej funkcji zewnętrznej, należy w programie, z którego jest ona wywoływana, umieścić taki sam blok COMMON, tak by przekazywana była nowa wartość zmiennej ISEED. Aby przekonać się, że podana funkcja realizuje algorytm w sposób poprawny, można położyć na początek $ISEED = 1$, a po 10 000 kroków uzyska się wartość $ISEED = 1\ 043\ 618\ 065$ (Park i Miller, 1988).

Aby generator liczb losowych był uruchamiany za każdym razem w inny sposób, potrzebna jest jakaś systematyczna metoda uzyskiwania różnych wartości zmiennej wyjściowej, tzn. wartości początkowej zmiennej ISEED. W przeciwnym razie, tzn. jeśli program byłby uruchamiany za każdym razem z tą samą wartością zmiennej wyjściowej, dawałby za każdym razem dokładnie taki sam wynik.

Inne rozkłady

Gdy już ma się dobry generator liczb losowych o rozkładzie jednostajnym, można go wykorzystać do budowy innych rodzajów generatorów liczb losowych. Na przykład można użyć generatora liczb losowych o rozkładzie jednostajnym do wytworzenia generatorów liczb o rozkładzie wykładniczym lub gaussowskim. Każdy rozkład wykładniczy można sprowadzić do jego najprostszej postaci:

$$p(x) = e^{-x}$$

przez odpowiedni dobór jednostek i współrzędnych. Na przykład dla układu o poziomach energetycznych E_0, E_1, \dots, E_n prawdopodobieństwo, że układ ma w temperaturze T energię E_i jest dane wzorem

$$p(E_i, T) \propto e^{-(E_i - E_0) / k_B T},$$

gdzie k_B jest stałą Boltzmana. Powyższe równanie przybiera postać $p(x) = e^{-x}$, jeśli wybierze się $k_B T$ jako jednostkę energii, a E_0 jako jej wartość zerową.

Aby wytworzyć rozkład wykładniczy, musimy znaleźć jego związek z rozkładem jednostajnym. Na przykład rozkład jednostajny $f(y) = 1$ dla $y \in [0, 1]$ można powiązać z rozkładem wykładniczym za pomocą związku

$$f(y)dy = dy = p(x)dx = e^{-x} dx,$$

co daje

$$y(x) - y(0) = 1 - e^{-x}.$$

Kładąc $y(0) = 0$, z równania tego otrzymuje się

$$x = -\ln(1 - y),$$

co stanowi związek rozkładu wykładniczego zmiennej x z rozkładem jednostajnym zmiennej $y \in [0,1]$. Poniższa funkcja zewnętrzna służy do generowania liczb losowych o rozkładzie wykładniczym na podstawie generatora liczb losowych o rozkładzie jednostajnym przy wykorzystaniu powyższego równania.

```

FUNCTION ERNFO
COMMON /CSEED/ ISEED
  ERNF = -ALOG(1.0-RANF())
  RETURN
END

```

Wykorzystuje się tu generator liczb losowych o rozkładzie jednostajnym. Innym ważnym rozkładem jest rozkład Gaussa (rozkład normalny)

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2},$$

gdzie σ jest wariancją rozkładu, którą możemy chwilową przyjąć za równą 1. Rozkład o $\sigma \neq 1$ można otrzymać z rozkładu o $\sigma = 1$ przez zmianę jednostki dla zmiennej x o czynnik σ . Wykorzystując rozkład jednostajny $f(\phi) = 1$ dla $\phi \in [0, 2\pi]$ oraz rozkład wykładniczy $p(t) = \exp(-t)$ dla $t \in [0, \infty]$, można otrzymać dwa rozkłady gaussowskie $g(x)$ i $g(y)$.

Iloczyn rozkładów jednostajnego i wykładniczego można powiązać z iloczynem dwóch rozkładów gaussowskich za pomocą związku:

$$\frac{1}{2\pi} \int f(\Phi) d\Phi p(t) dt = \int g(x) dx g(y) dy,$$

skąd wynika

$$e^{-t} dt d\Phi = e^{-(x^2 + y^2)/2} dx dy$$

Powyższe równanie można potraktować jako równanie zmiany współrzędnych z układu biegunowego

(p, Φ) , gdzie $p = \sqrt{2t}$, do układu prostokątnego (x, y) , tzn.

$$x = \sqrt{2t} \cos \Phi,$$

$$y = \sqrt{2t} \sin \Phi,$$

Obie zmienne x i y mają rozkład gaussowski, jeśli zmienna t ma rozkład wykładniczy, a Φ ma rozkład jednostajny w przedziale $[0, 2\pi]$. Dysponując już generatorem liczb losowych o rozkładzie wykładniczym oraz generatorem liczb losowych o rozkładzie jednostajnym, możemy natychmiast skonstruować dwie liczby losowe o rozkładzie gaussowskim. Generator liczb losowych o rozkładzie wykładniczym otrzymujemy z generatora liczb losowych o rozkładzie jednostajnym. Poniższy podprogram wytwarza dwie liczby o rozkładzie gaussowskim na podstawie dwóch liczb o rozkładzie jednostajnym w przedziale $[0,1]$.

```
SUBROUTINE GRNF(X,Y)
COMMON /CSEED/ ISEED
PI = 4.0*ATAN(1.0)
R1 = -ALOG(1.0-RANF()) R2 = 2.0*PI*RANF()
R1 = SQRT(2.0*R1)
X = R1*COS(R2)
Y = R1*SIN(R2)
RETURN
END
```

W zasadzie każdy rozkład może być wytworzony numerycznie. Korzystając z generatorów liczb o rozkładach gaussowskim i wykładniczym, tworzymy nowe generatory, wykorzystując przekształcenia całkowite do powiązania rozkładów poszukiwanych ze znanymi. Można stworzyć ogólną procedurę numeryczną do analizy numerycznej przekształceń całkowitych.

Wykorzystana literatura:

Zbigniew Kotulski

„Generatory liczb losowych: algorytmy, testowanie, zastosowania”
Warszawa 2001

Tao Pang

„Metody obliczeniowe w fizyce”
Wydawnictwa Naukowe PWN Warszawa 2001