Metoda Monte Carlo i jej zastosowania

D. Szeliga A. Cichoń M. Kielian

Kraków, 2016

Plan

1. Liczby pseudolosowe
2. Metoda Monte Carlo

Liczby losowe a pseudolosowe

Liczba losowa (ang. random number) jest liczbą r należącą do pewnego zbioru wartości r1, ..., rn wybieranych z pewnym prawdopodobieństwem. Jeśli jako r może pojawić się każda z liczb zbioru z tym samym prawdopodobieństwem $p(r) = 1/n$, to mówimy o równomiernym rozkładzie prawdopodobieństwa liczb losowych z tego zbioru. Na przykład rozważmy rzut kostką. Każdy rzut daje liczbę losową r ze zbioru 1,2,3,4,5,6. Jeśli kostka nie jest oszukana, to każda z możliwych wartości r pojawia się w rzucie kostką z prawdopodobieństwem $p(r) = 1/6$. Liczby losowe r posiadają zatem równomierny rozkład prawdopodobieństwa.
Ponieważ nie możemy w prosty sposób mieć prawdziwych liczb losowych, musimy się zadowolić ich sztucznym odpowiednikiem - liczbami pseudolosowymi (ang. pseudorandom numbers). Liczby pseudolosowe wyglądają jak losowe, lecz tworzy się je algorytmicznie. Oznacza to, iż znając wzór generacyjny oraz kilka kolejnych liczb pseudolosowych możemy bez problemu wygenerować wszystkie dalsze - tej cechy nie posiadają liczby losowe, w przeciwnym razie totolotek straciłby sens.

- Komputer nie jest w stanie tak naprawdę losować liczb.
- Zamiast tego stosuje się ciągi liczbowe o ustalonych z góry wartościach, które bardzo dobrze imitują ciągi liczb losowych.
- Dlatego zamiast mówić o liczbach losowych mówi się o liczbach pseudolosowych.
- Komputery tak naprawdę generują tylko liczby z rozkładu jednostajnego.

Generowanie rozkładu normalnego

- Dany rozkład można generować poprzez funkcję odwrotną do dystrybuanty.
- Inną popularną metodą jest skorzystanie z centralnego twierdzenia granicznego.
- Wystarczy zsumować 12 liczb pochodzących z rozkładu jednostajnego, aby uzyskać rozkład normalny.
- 12 bierze się również z tego, że wariancja w rozkładzie jednostajnym dana jest jako:

  $$\frac{(b-a)^2}{12}$$

  Na odcinku (0,1) wynosi 1/12.

Przykład

```
#include <iostream>
#include <iomanip>
#include <cstdlib>
#include <time.h>

using namespace std;

int main()
{
    srand(time(NULL));

    for(int i = 1; i <= 20; i++) cout << setw(5) << rand() << endl;

    return 0;
}
```

Przykład

```cpp
#include <iostream>
#include <iomanip>
#include <cstdlib>
#include <time.h>

using namespace std;

int main()
{
    int a,b;

    cout << "a = "; cin >> a;
    cout << "b = "; cin >> b;

    cout << endl;

    srand(time(NULL));

    for(int i = 1; i <= 20; i++) cout << setw(5) << a + rand() % (b - a + 1) << endl;

    return 0;
}
```

Funkcja odwrotna do dystrybuanty

- Dystrybuanta jest dana na odcinku [0,1].
- Aby znaleźć wartość zmiennej z szukanego rozkładu należy znaleźć takie x, żeby dystrybuanta w tym punkcie miała wartość z rozkładu jednostajnego.
- Dystrybuanta rozkładu normalnego

$$\Phi(X) = \frac{1}{2}(1 + erf\frac{x-\mu}{\sigma\sqrt{2}})$$

- Dla u_i z rozkładu jednostajnego, aby uzyskać zmienną z rozkładu normalnego należy znaleźć kwantyl rozkładu normalnego rzędu u_i.

Najważniejsze rozkłady

ROZKŁADY DYSKRETNE
-rozkład jednopunktowy
-rozkład dwupunktowy
-rozkład dwumianowy Bernoulliego
-rozkład Poissona
ROZKŁADY CIĄGŁE
-jednostajny
-normalny
-t-Studenta
-Chi-kwadrat
-F-Snedecora

Geometryczny ruch Browna

- Wiele zjawisk w ekonomii modelowanych jest przez geometryczny ruch Browna (Geometric Brownian Motion).
- Logarytm losowej wartości odpowiada ruchowi Browna (procesowi Wienera).
- W ekonomii GBM wykorzystywany jest np. do modelowania cen akcji.
- Był on stosowany w wycenie cen akcji modelu Blacka-Scholesa.
- Wiele współczesnych badań wskazuje, że ceny akcji nie podlegają GBM.

Rozkłady wielowymiarowe

- Wiele zjawisk wielowymiarowych.
- Aby móc je modelować musimy umieć generować rozkłady wielowymiarowe.
- Generowanie zmiennych niezależnych jest proste- wystarczy osobno wylosować zmienne i złożyć wektor.
- Generowanie zmiennych zależnych jest trudniejsze i wymaga uwzględnienia zależności pomiędzy zmiennymi.
- COV(X,Y)=E(XY)-E(X)E(Y)

Generowanie zmiennych z rozkładu normalnego

- Aby generować zmienne niezależne z rozkładu normalnego można posłużyć się przekształceniem Boxa-Mullera.
- U_1 i U_2 są niezależne i pochodzą z rozkładu jednostajnego (0,1).
- $\mathrm{Cov}(X_1, X_2) = 0$

Metoda Monte Carlo (MC) jest stosowana do modelowania matematycznego procesów zbyt złożonych (obliczania całek, łańcuchów procesów statystycznych), aby można było przewidzieć ich wyniki za pomocą podejścia analitycznego. Istotną rolę w metodzie MC odgrywa losowanie (wybór przypadkowy) wielkości charakteryzujących proces, przy czym losowanie dokonywane jest zgodnie z rozkładem, który musi być znany.

- Metoda Monte Carlo jest jedną z najczęściej używanych metod symulacyjnych.
- Metoda została wymyślona i po raz pierwszy zastosowana przez Stanisława Ulama.
- Jedną z najczęstszych aplikacji metody Monte Carlo jest całkowanie numeryczne.
- Jak w każdej metodzie numerycznej należy pamiętać o zasadzie garbage in garbage out.

Metoda Crude Monte Carlo

- Chcemy policzyć całkę z funkcji f(x)

$$I = \int_a^b f(x)dx$$

- W przedziale [a,b] losujemy N punktów i przybliżamy

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

- Wygląda to tak samo jak całka Riemanna.

Metoda akceptacji- odrzuceń

- Najczęściej stosowana metoda Monte Carlo.
- Losujemy punkty z prostokąta (może być wielowymiarowy): [a,b]*[0,d].
- Liczymy proporcję punktów leżących nad wykresem do punktów leżących pod wykresem.
- $$I = \int_a^b f(x)dx \approx \frac{k}{N}(b-a)d$$
- k-liczba punktów pod wykresem, N-liczba losowanych punktów.

Przykład

- Obliczanie wartości
$$\pi$$
jest często podawanym przykładem zastosowania całkowania metodą MC
- Wiemy, że pole koła wynosi
$$\pi r^2$$
- Możemy zatem obliczyć pole koła metodą MC i wyliczyć stąd
$$\pi$$

```
#include <iostream>
#include <cmath>
#include <ctime>
#include <cstdlib>

using namespace std;

int main()
{
      srand(time(NULL)); //zainicjalizowanie maszyny generujacej liczby
pseudolosowe
      int n;
      int nk = 0;
      double x,y;
      float s;

      cout << "Podaj liczbe losowanych pkt:" << endl;
      cin >> n;

      for(int i = 1; i <= n; i++)
      {
          x = ((double)rand() / (RAND_MAX)) * 2 - 1;
          y = ((double)rand() / (RAND_MAX)) * 2 - 1;
          if(x*x + y*y <= 1)
          {
              nk++;
          }
      }

      cout << "Liczba pkt. w kole wynosi: " << nk << endl;
      cout << "Liczba pkt. w kwadracie wynosi: " << n << endl;
      s = 4. * nk / n;
      cout << "Liczba pi wynosi: " << s;
```

Zalety metody MC

- Możliwość rozwiązania trudnych prblemów.
- Posta forma zastąpienia rozwiązań analitycznych.
- Rosnąca moc obliczeniowa komputerów.
- Uwalniają użytkownika od skomplikowanej teorii i wzorów, pozwalając skupić się na istocie pytania, na które statystyka ma odpowiedzieć.

Wady metody MC

- Eksperymenty skończonej liczby prób.
- Wyniki zawsze będą przybliżeniem.
- Wyniki zależą od jakości generatora liczb pseudolosowych.