Metoda Monte Carlo i jej zastosowania

Tomasz Mostowski

Zajęcia 31.03.2008

Plan

Plan

1. Prawa wielkich liczb
   - PWL
2. Liczby pseudolosowe
3. Metoda Monte Carlo
   - Wstęp
   - Metoda Crude Monte Carlo
   - Metoda Akceptacji - Odrzuceń
   - Wady i zalety metod Monte Carlo

Przypomnienie

Istnieje wiele wariantów praw wielkich liczb. Wspólną ich cechą jest asymptotyczne zachowanie wyrażeń typu

$$\frac{X_1 + X_2 + \cdots + X_n - a_n}{b_n}$$

Przypomnimy najważniejsze dwa z nich

- Prawo Wielkich Liczb Bernoulliego
- Mocne Prawo Wielkich Liczb Kołmogorowa

Prawo Wielkich Liczb Bernoulliego

Jeśli S_n jest liczbą sukcesów w schemacie Bernoulliego z prawdopodobieństwem sukcesu p, to dla każdego $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) \longrightarrow_{n\to\infty} 1$$

Intuicyjne przykłady

- Liczba orłów przy dużej liczbie rzutów uczciwą stabilizuje się wokół 0.5
- liczba jedynek w rzutach uczciwą kostką powinna wynosić $\frac{1}{6}$ wszystkich rzutów.

Mocne Prawo Wielkich Liczb Kołmogorowa

Jeśli $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie z wartością oczekiwaną μ, to dla każdego $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\sup_{m\geq n}\left|\frac{X_1 + X_2 + \cdots + X_n}{m} - \mu\right| \leq \varepsilon\right) = 1$$

Intuicyjnie:

- Losując coraz więcej liczb z tego samego rozkładu, średni wynik nie będzie się różnił od wartości oczekiwanej o więcej niż dowolnie mała liczba.

Wprowadzenie

Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.

John von Neumann 1951

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

Przykład

- Mamy ciąg liczb 13, 21, 57, 69, 73. Jaka liczba będzie następna?.
x_n	
13	
21	
57	
69	
73	
?	

Przykład

Algorytm

- Algorytm jest następujący $x_{n+1} = \mod(x_n * B, M)$
- Tutaj $x_{n+1} = \mod(x_n * 17, 100)$
-

x_n	$17x_n$
13	221
21	357
57	969
69	1173
73	1241
41	697

  Modulo to reszta z dzielenia.
- Czasami przy obliczeniach dodaje się jeszcze stałą $x_{n+1} = \mod(x_n * B + c, M)$

Uwagi

- Punkt startowy to ziarno losowania.
- Złe dobranie ziarna, albo modulo może powodować szybką okresowość algorytmu.
- Ręczne ustawienie ziarna jest niezbędne do replikowania symulacji i obliczeń.
- Można „losować” ziarno – brane jest wtedy z czasu systemowego.

generowanie innych rozkładów

- Komputer potrafi tylko generować ciągi liczb naturalnych.
- Stworzenie z nich rozkładu jednostajnego jest proste – dzielimy przez modulo.
- Rozkłady dyskretne zazwyczaj określa się przez odpowiedni podział odcinka.
- Uzyskiwanie innych rozkładów ciągłych jest bardziej skomplikowane.
- Stosuje się różne przekształcenia.
- Można stosować funkcję odwrotną do dystrybuanty.

generowanie rozkładu normalnego

- Jednym z ważniejszych rozkładów jest rozkład normalny.
- Możemy generować go poprzez funkcję odwrotną do dystrybuanty.
- Inną popularną metodą jest skorzystanie z centralnego twierdzenia granicznego.
- $$U_i \sim U(0,1) \qquad X = \sum_{i=1}^{12} U_i - 6 \sim N(0,1)$$
- Wystarczy zsumować 12 liczb pochodzących z rozkładu jednostajnego, aby uzyskać rozkład normalny.
- 12 bierze się też stąd, że wariancja w rozkładzie jednostajnym dana jest jako $\frac{(b-a)^2}{12}$. Na odcinku $(0,1)$ wynosi $1/12$

Funkcja odwrotna do dystrybuanty

Zastosowanie do generowania zmiennych z rozkładów ciągłych

- 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 $\Phi(x) = \frac{1}{2}\left(1 + erf\frac{x-\mu}{\sigma\sqrt{2}}\right)$
- Dla u_i z rozkładu jednostajnego, aby uzyskać zmienną z rozkładu normalnego należy znaleźć kwantyl rozkładu normalnego rzędu u_i

Przypomnienie najważniejszych rozkładów

ROZKŁADY DYSKRETNE

- Rozkład jednopunktowy
- Rozkład dwupunktowy
- Rozkład dwumianowy Bernoulliego
- Rozkład Poissona

Przypomnienie najważniejszych rozkładów

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!

Zapis formalny

- Równanie stochastyczne $dS_t = \mu S_t dt + \sigma S_t dW_t$
- Rozwiązanie analityczne $S_t = S_0 \exp((\mu - \frac{\sigma^2}{2})t + \sigma W_t)$
- Implikuje to, że $\ln\left(\frac{S_t}{S_0}\right) \sim N\left((\mu - \frac{\sigma^2}{2})t, \sigma^2 t\right)$
- SAS nie ma oprogramwanego generowania zmiennych z GBM, ale można samemu przekształcać.

Rozkłady wielowymiarowe

- Wiele zjawisk jest wielowymiarowych.
- Aby móc je modelować musimy umieć generować rozkłady wielowymiarowe.
- Generowanie zmiennych niezależnych jest proste – wystarczy osobno wylosować zmienne i złożyć w 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-Müllera
- U_1 i U_2 są niezależne i pochodzą z rozkładu jednostajnego (0,1)
- $X_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$
- $X_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$
- $Cov(X_1, X_2) = 0$

Plan

Wstęp

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

Plan

Metoda Crude Monte Carlo

- Chcemy policzyć całkę z funkcji $f(x)$

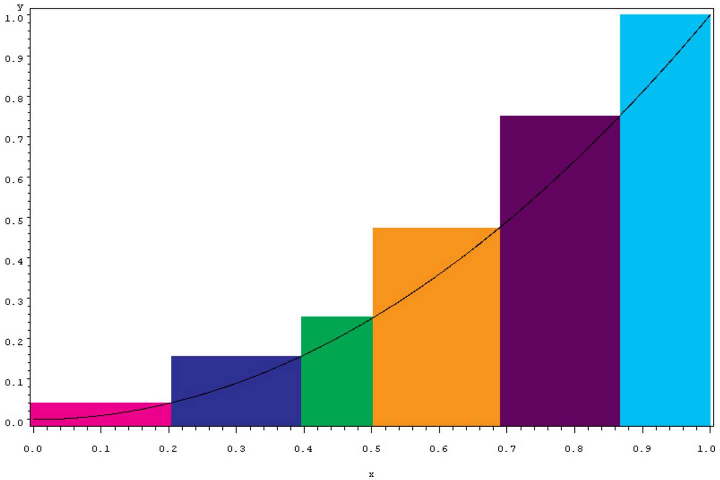
$$I = \int_a^b f(x)dx$$

- 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.

Przykład



Plan

Metoda Akceptacji - Odrzuceń

- Jest to najczęściej stosowana metoda Monte Carlo
- Czasami nazywana jest `Typowym 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.

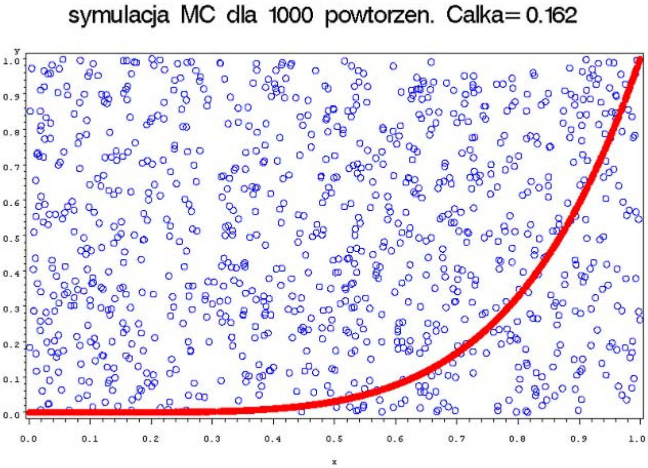
Metoda Akceptacji - Odrzuceń – przykład

- Chcemy policzyć całkę $y = \int_0^1 x^5 dx$
- W tym przypadku możemy policzyć analitycznie

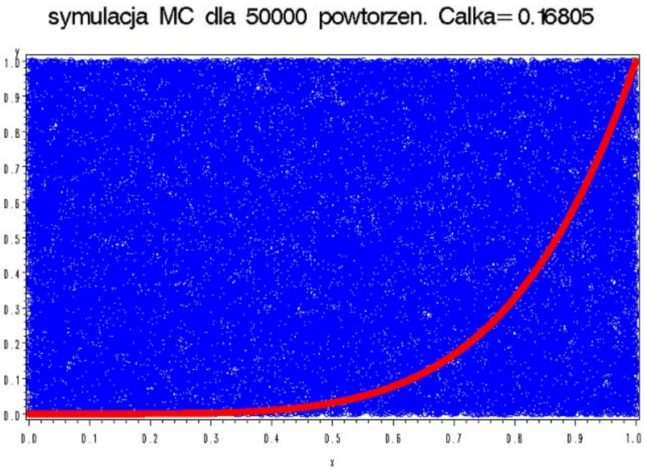
$$y = \int_0^1 x^5 dx = \frac{1}{6} x^6 \Big|_0^1 = \frac{1}{6} = 0.16(6)$$

- Ale możemy policzyć też metodą MC.
- Losujemy N punktów z kwadrattu $[0, 1]$ i patrzymy ile z nich leży pod krzywą.

Przykład


symulacja MC dla 1000 powtorzen. Calka= 0.162
y
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
x

Przykład


symulacja MC dla 50000 powtorzen. Calka= 0.16805
y
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
x

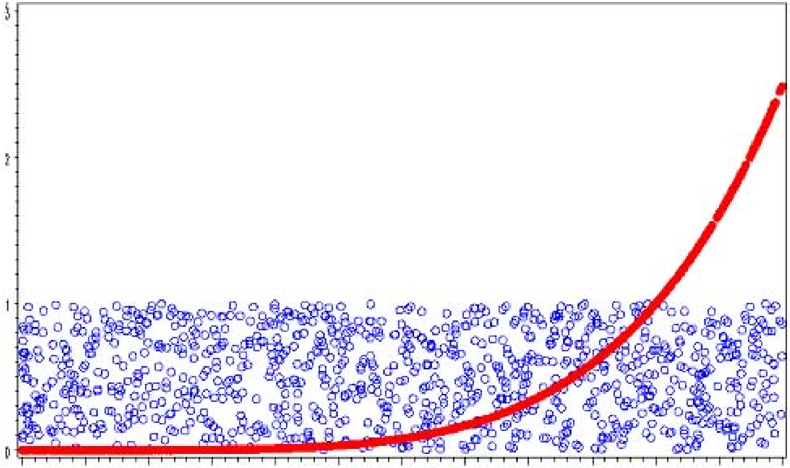
Algorytm

1. Ustaw $k = 0$ i $n = 0$.
2. Wylosuj liczby x i y z rozkładu jednostajnego $[0, 1]$
3. Sprawdź, czy $y < f(x)$. Jeżeli tak zwiększ k o 1.
4. Powtórz punkty 1. i 2. n razy. n powinno być dużą liczbą (np. 10.000).
5. Podziel licznik przez liczbę powtórzeń $s = k/n$
6. s jest przybliżeniem całki z funkcji na odcinku [0.1].

Algorytm w przypadku ogólnym.

- W metodzie tej tak naprawdę liczymy prawdopodobieństwo znalezienia się pod wykresem.
- Jeżeli losujemy punkty z prostokąta o polu innym niż 1, to aby obliczyć całkę należy przemnożyć prawdopodobieństwo przez pole prostokąta.
- $\int = \frac{\text{l. punktów pod wykresem}}{\text{liczba wylosowanych punktów}} \text{pole prostokąta}$
- Należy pamiętać, żeby prostokąt, z którego losujemy obejmował maksimum z $f(x)$

Przykład

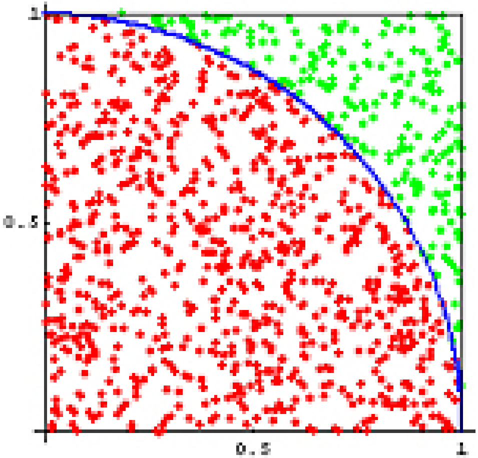


Przykład - obliczenie π

How I want a drink, alcoholic of course, after the heavy lectures involving quantum mechanics!

- Obliczanie wartości liczby π jest często podawanym przykładem zastosowania całkowania metodą MC
- Wiemy, że pole koła wynosi πr^2
- Możemy zatem obliczyć pole koła metodą MC i wyliczyć stąd π
- Dla łatwiejszych obliczeń możemy losować punkty z pierwszej ćwiartki koła, a następnie przemnożyć wynik przez 4.
- Podobno w mniej więcej taki sposób obliczano π w XIX wieku, rzucając w kwadratową tarczę z narysowanym kołem.

Przykład



Plan

Zalety metody MC

- możliwość rozwiązania trudnych problemów
- prosta 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 metod MC

- eksperymenty dla jednak skończonej liczby prób
- wyniki zawsze będą przybliżeniem
- wyniki zależą od jakości generatora liczb pseudolosowych