Monte Carlo: dyskretny rozkład Bernoulliego, rozkład normalny, centralne twierdzenie graniczne

28 lutego 2022

1 Wstęp

1.1 CTG

Jeśli zmienne losowe $x_1, x_2, \ldots, x_N$ są opisywane są rozkładami f_{x_i} o wartościach oczekiwanych μ_i oraz wariancjach σ_i^2 (zmienne mogą pochodzić z różnych rozkładów) to ich średnia

$$Z = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1}$$

też jest zmienną losową i zgodnie z *centralnym twierdzeniem granicznym* jej wartość oczekiwana dla $N \to \infty$ ma rozkład normalny

$$\langle Z \rangle = \langle \frac{1}{N}\sum_{i=1}^{N} x_i \rangle = \frac{1}{N}\sum_{i=1}^{N} \langle x_i \rangle \tag{2}$$

o wariancji równej

$$\sigma^2_{\langle Z \rangle} = \langle Z^2 \rangle - \langle Z \rangle^2 = \frac{1}{N^2}\sum_{i=1}^{N} \sigma^2_{X_i} \tag{3}$$

1.2 rozkład Bernouliego

Rozkład Bernoulliego zmiennej losowej X definiujemy następująco

$$X \in \{0,1\}, \qquad P\{X=0\} = q, \qquad P\{X=1\} = p, \qquad p+q=1 \tag{4}$$

Momenty rozkładu

$$\langle X \rangle = p \tag{5}$$

$$\langle X^2 \rangle = p \tag{6}$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2 = p - p^2 \tag{7}$$

Jeśli nowa zmienna losowa będzie sumą zmiennych z rozkładu Bernoulliego to wartość oczekiwana rozkładu normalnego będzie równa

$$\langle Z \rangle = \frac{1}{N}\sum_{i=1}^{N} \mu_i = \frac{1}{N}\sum_{i=1}^{N} p = \frac{Np}{N} = p \tag{8}$$

czyli identyczna jak dla rozkładu Bernoulliego, natomiast wariancja maleje ze wzrostem N

$$\sigma^2_{\langle Z\rangle} = \frac{\sum_{i=1}^{N}\sigma^2_{X_i}}{N^2} = \frac{\sum_{i=1}^{N}(p-p^2)}{N^2} = \frac{p-p^2}{N} \tag{9}$$

1.3 generowanie zmiennej o rozkładzie Bernoulliego

Pseudokod

```
U1 = uniform()
if U1 ⩽ p then
    X=1
else
    X=0
end if
```

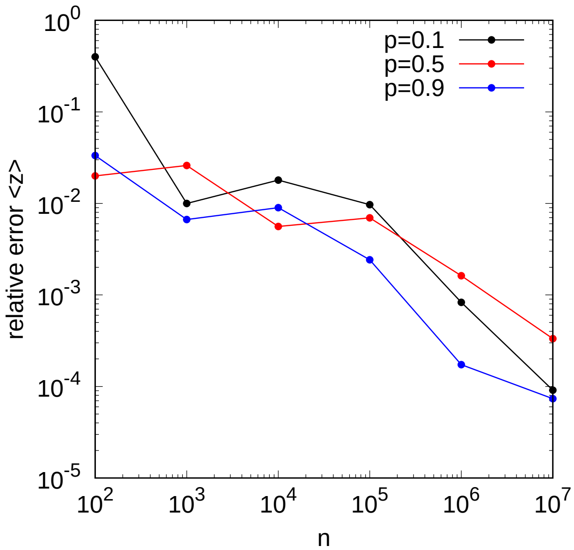
Pierwsza linijka oznacza losowanie zmiennej U_1 przy użyciu generatora o rozkładzie równomiernym w przedziale $[0,1]$. W projekcie użyjemy standardowego generatora $rand()$ dostępnego w $C/C++$ bibliotece *stdlib*

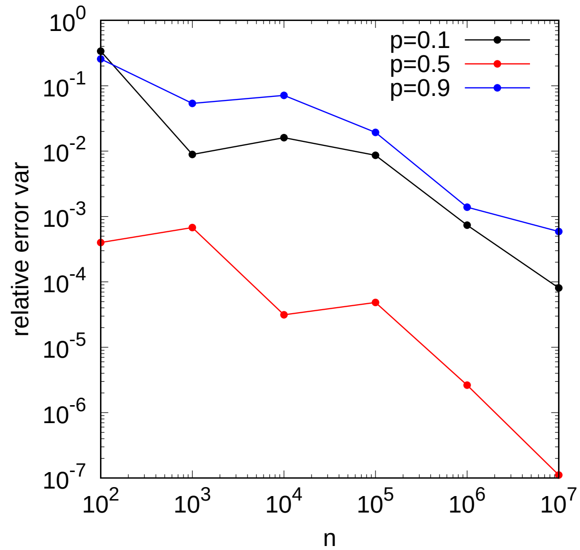
```
#include<stdlib.h>

double uniform(){
  return (rand()/(double)RAND_MAX);
}
```

2 Zadania do wykonania

1. zaimplementować algorytm generowania liczb o rozkładzie jednorodnym i Bernulliego
2. wygenerować ciąg $N = 10^7$ liczb z rozkładu Bernoulliego dla $p = 0.1$, 0.5, 0.9
3. w trakcie symulacji na bieżąco wyznaczać aktualne wartości: $\langle Z\rangle$,$\langle Z^2\rangle$ tj. dla aktualnej ilości wylosowanych już liczb $n \leqslant N$
4. dla $n = 10^k$, $k = 2, 3, 4, 5, 6, 7$ obliczyć i zapisać do pliku: ilość liczb n, wartość oczekiwaną $\langle Z\rangle$ oraz wariancję $\sigma^2_{\langle Z\rangle}$ obliczone numerycznie i analitycznie, błąd względny wartości oczekiwanej i wariancji; sprządzić wykresy błędów względnych w funkcji n - zastosować skalę logarytmiczną na obu osiach
5. w raporcie przedyskutować uzyskane wyniki

3 przykładowe wyniki





Rysunek 1: Błąd względny wartości oczekiwanej i wariancji dla sumy n liczb o rozkładzie Bernoulliego.