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

28 lutego 2024

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 arytmetyczna

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

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

$$\langle \overline{X} \rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N} x_i \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\langle x_i \rangle = \frac{N\langle X \rangle}{N} = \langle X \rangle \tag{2}$$

czyli o wartości równej wartości oczekiwanej pojedynczej zmiennej losowej X, natomiast wariancja średniej jest równa

$$\sigma^2_{\overline{X}} = \frac{\overline{X^2} - \overline{X}^2}{N} \tag{3}$$

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 podzieloną przez ich liczbę N

$$Z = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{8}$$

to wartość oczekiwana nowej zmiennej losowej określa wyrażenie

$$\langle Z\rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N} X_i \right\rangle = \frac{1}{N}\sum_{i=1}^{N}\langle X_i\rangle = \frac{1}{N}\sum_{i=1}^{N} p = \frac{Np}{N} = p \tag{9}$$

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

$$\sigma_Z^2 = \frac{\sum\limits_{i=1}^{N}\sigma_{X_i}^2}{N^2} = \frac{\sum\limits_{i=1}^{N}(p-p^2)}{N^2} = \frac{p-p^2}{N} \tag{10}$$

Porównując teraz wzory (1, 3) z wzorami (9, 10) możemy zauważyć że wykonując bezpośredxnio eksperyment komputerowy, jego wyniki możemy bezpośrednio porównać z wynikiem analitycznym. Najlepiej jest policzyć wartość bezwzględną błędu względnego ϵ

$$\epsilon_{\overline{X}} = \left|\frac{\overline{X}-\langle Z\rangle}{\langle Z\rangle}\right| \tag{11}$$

$$\epsilon_{\sigma^2_{\overline{X}}} = \left|\frac{\sigma^2_{\overline{X}}-\sigma^2_Z}{\sigma^2_Z}\right| \tag{12}$$

Zgodnie z CTG błąd $\epsilon_{\overline{X}}$ powinien dążyć do zera dla $N \to \infty$, podobnej zależności spodziewamy się też dla błędu wariancji. Celem tego projektu jest numeryczne potwierdzenie słuszności CTG.

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 $(0, 1)$. Wykonując wielokrotnie takie losowanie, otrzymamy ciąg 0 i 1, przy czym 1 będą pojawiać się z załozonym prawdopodobieństwem p.

W celu uzyskania ciągu liczb o rozkładzie jednorodnym w przedziale $(0, 1)$ użyjemy standardowego generatora $rand()$ dostępnego w $C/C++$ bibliotece *stdlib*/*cstdlib*

```
#include<stdlib.h>

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

1.4 pseudokod programu

Zaletą Monte Carlo jest możliwość monitorowania wyniku w trakcie wykonywanych obliczeń, możliwość tę wykorzystamy aby obliczyć błędy względne dla $N = 10^2$, 10^3, ..., wóczas wyliczamy potrzebne wielkości i zapisujemy do pliku.

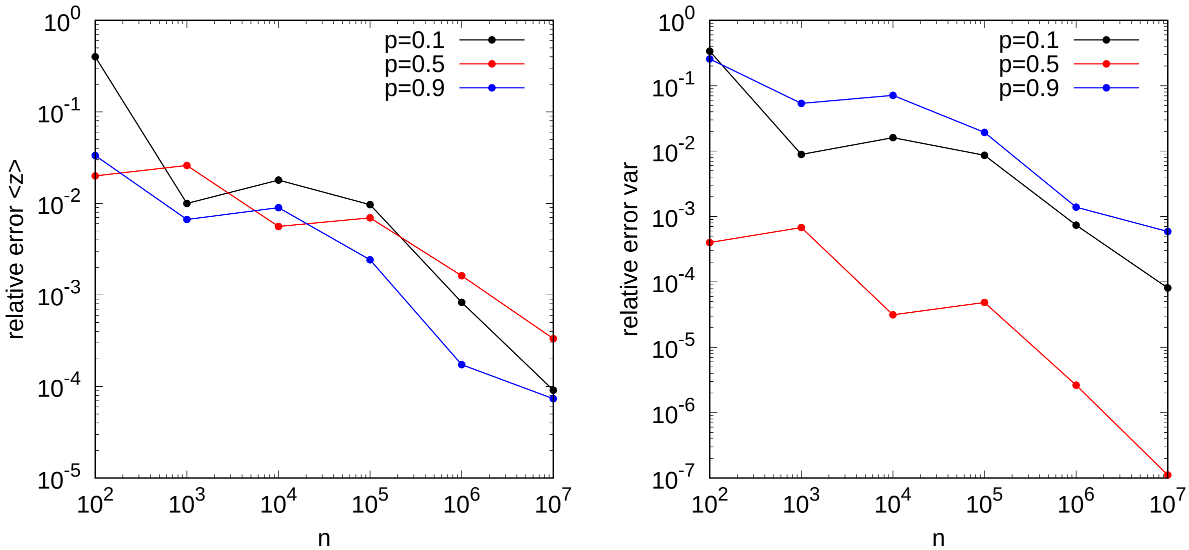
```
inicjalizacja:  N, p
                k = 2
                sum_X = 0
                sum_X2 = 0
for(i=1; i⩽N; i++ ){
        U_1=uniform()
        if(U_1 ⩽ p) X = 1
        else X = 0
        sum_X = sum_X+X
        sum_X2 = sum_X2+X^2
        if(i == 10^k){
                k++
                X̄ = sum_X / i
                X̄² = sum_X2 / i
                err_X = |(X̄ - p) / p|
                var_num = (X̄² - (X̄)^2) / i
                var_teo = (p - p^2) / i
                err_var = |(var_num - var_teo) / var_teo|
                zapis do pliku: p,i,X̄,X̄²,err_X,err_var
        }

}
```

2 Zadania do wykonania

1. zaimplementować algorytm generowania liczb o rozkładzie jednorodnym i Bernoulliego
2. wygenerować ciąg $N = 10^7$ liczb z rozkładu Bernoulliego dla $p = 0.1$, 0.5, 0.9
3. dla $i = 10^k$, $k = 2, 3, 4, 5, 6, 7$ obliczyć i zapisać do pliku: ilość liczb n, wartości numeryczne $\overline{X}$ i $\sigma^2_{\overline{X}}$ oraz analityczne $\langle Z \rangle$ i σ^2_Z , błąd względny wartości średniej i wariancji; sprządzić wykresy błędów względnych w funkcji aktualnej liczby losowań i - zastosować skalę logarytmiczną na obu osiach
4. w raporcie przedyskutować uzyskane wyniki

3 przykładowe wyniki



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