Dyskretny rozkład Bernoulliego, rozkład normalny, centralne twierdzenie graniczne

Metody Monte Carlo w Fizyce

Julia Potempa (411073)

Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

5. marca 2025r.

1 Wstęp

Centralne twierdzenie graniczne mówi o tym, że średnia arytmetyczna z N zmiennych losowych opisanych rozkładami f_{x_i}, które mogą być różne, również jest zmienną losową i jej wartość oczekiwana dla $N \to \infty$ ma rozkład normalny - to znaczy, że jest równa wartości oczekiwanej z pojedynczej, pierwotnej zmiennej losowej:

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

$$\langle \overline{X} \rangle = \langle X \rangle$$

Wariancja wtedy jest dana równaniem:

$$\sigma_{\overline{X}}^2 = \frac{\overline{X^2} - \overline{X}^2}{N}$$

Rozkład Bernoulliego lub też rozkład zero-jedynkowy, to rodzaj dyskretnego rozkładu prawdopodobieństwa, w którym zmienna losowa przyjmuje tylko wartości 0 albo 1. Posiada on szereg własności:

$$X \in \{0, 1\},\ P(X = 1) = p,\ P(X = 0) = 1 - p$$

$$\langle X \rangle = p$$

$$\langle X^2 \rangle = p$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2 = p - p^2$$

Tworzymy nową zmienną losową, która jest sumą zmiennych z rozkładu Bernoulliego podzieloną przez ich sumę N:

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

Następnie obliczamy jej wartość oczekiwaną:

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

oraz wariancję:

$$\sigma_Z^2 = \frac{\sum_{i=1}^{N} \sigma_{X_i}^2}{N^2} = \frac{\sum_{i=1}^{N} (p - p^2)}{N^2} = \frac{p - p^2}{N}$$

Widzimy, że wartość oczekiwana jest identyczna jak dla rozkładu Bernoulliego oraz wariancja jest odwrotnie proporcjonalna do N.

2 Problematyka

Niniejsze ćwiczenie miało na celu przeprowadzenie eksperymentu numerycznego, który miał sprawdzić, czy błąd względny wartości oczekiwanej oraz wariancji nowej zmiennej losowej wraz z N dążącym do nieskończoności dążą do zera zgodnie z centralnym twierdzeniem granicznym.

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

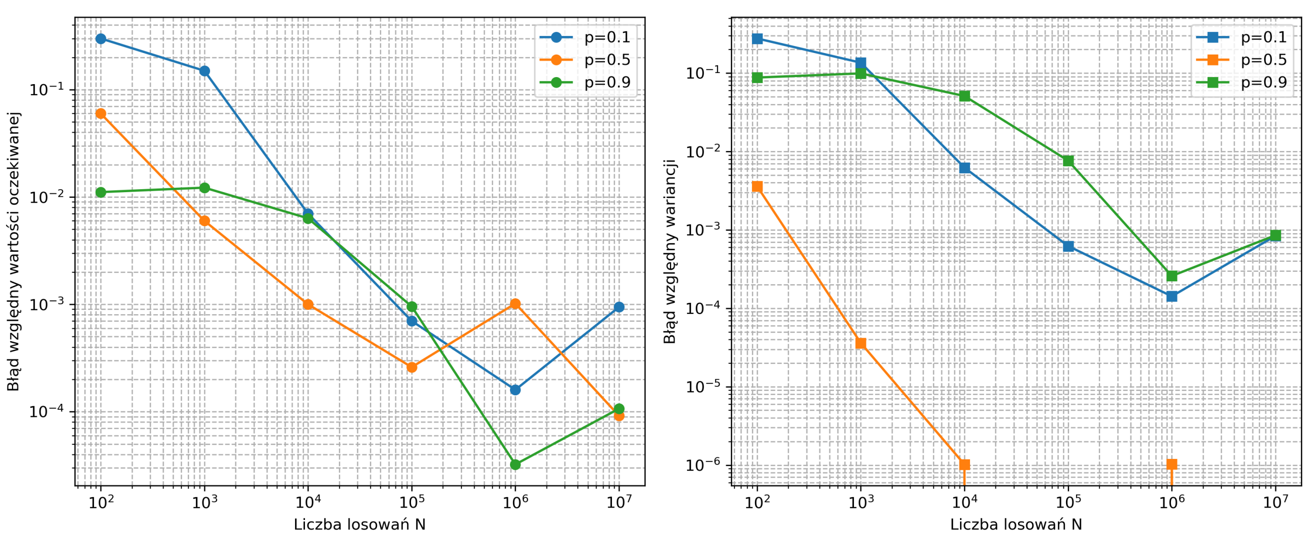
$$\epsilon_{\sigma^2_{\overline{X}}} = \left| \frac{\sigma^2_{\overline{X}} - \sigma^2_Z}{\sigma^2_Z} \right|$$

W tym celu został napisany program w języku C++, który wykorzystywał generator liczb pseudolosowych z biblioteki `cstdlib` zaimplementowany za pomocą poniższej funkcji:

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

Za jego pomocą były losowane liczby z zakresu od 0 do 1, które następnie były porównywane z zadanym p ($p \in \{0.1, 0.5, 0.9\}$) i w ten sposób otrzymany został rozkład Bernoulliego. Następnie w każdym punkcie kontrolnym ($N = 10^2, 10^3, \ldots, 10^7$) były zapisywane błędy względne, które następnie zostały przedstawione na wykresie wygenerowanym za pomocą kodu przygotowanego w języku Python.

3 Wyniki



Rysunek 1: Wygenerowane wykresy a) błędu względnego wartości oczekiwanej oraz b) błędu względnego wariancji.

Otrzymane wykresy przedstawiają przewidziany wcześniej spadek zarówno błędu wartości oczekiwanej, jak i błędu wariancji. W przypadku wariancji widzimy, że błąd spada dla $p = 0.5$ w dwóch punktach do 0. Może to być spowodowane tym, że wariancja analityczna równa się wariancji numerycznej.

4 Podsumowanie

W niniejszym ćwiczeniu został zrealizowany program, w którym za pomocą generatora liczb pseudolosowych został otrzymany rozkład Bernoulliego dla różnych prawdopodobieństw sukcesu ($p \in \{0.1, 0.5, 0.9\}$. W każdym przypadku zostały obliczone zmienne losowe będące sumą tych wylosowanych. Następnie została sprawdzona zasadność centralnego twierdzenia granicznego obliczając błędy względne zarówno wartości oczekiwanej jak i wariancji otrzymanej sumy. Zestawienie wartości na wykresie pozwoliło w łatwy sposób przeanalizować otrzymane wyniki i potwierdzić założoną zależność.