

AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

WYDZIAŁ FIZYKI I INFORMATYKI STOSOWANEJ
KIERUNEK STUDIÓW: FIZYKA TECHNICZNA



METODY MONTE CARLO

Laboratorium 1

Dyskretny rozkład Bernoulliego, rozkład normalny, centralne twierdzenie graniczne

zrealizował
Przemysław Ryś

Kraków, 26 Luty 2024

1 Opis zagadnienia

Centralne Twierdzenie Graniczne (CTG)

Centralne Twierdzenie Graniczne (CTG) stanowi fundamentalne twierdzenie statystyki, które opisuje zachowanie się średniej arytmetycznej próbek z populacji. Mówi ono, że jeśli zmienne losowe x_1, x_2, \dots, x_N są opisywane rozkładami f_{x_i} o wartościach oczekiwanych μ_i oraz wariancjach σ_i^2 , to ich średnia arytmetyczna:

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

jest również zmienną losową. Dla $N \rightarrow \infty$, wartość oczekiwana \bar{X} ma rozkład normalny:

$$\langle \bar{X} \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N x_i \right\rangle \xrightarrow{N \rightarrow \infty} \langle X_i \rangle = \mu_i$$

Natomiast wariancja średniej jest równa:

$$\sigma_{\bar{X}}^2 = \frac{\bar{X}^2 - \langle \bar{X} \rangle^2}{N}$$

Rozkład Bernoulliego

Rozkład Bernoulliego opisuje eksperymenty z dwoma możliwymi wynikami: sukcesem (1) lub porażką (0). Definiujemy go następująco:

$$X \in \{0, 1\}, \quad P\{X = 0\} = q, \quad P\{X = 1\} = p, \quad p + q = 1$$

Momenty rozkładu Bernoulliego są określone jako:

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

Porównanie wyników

Wartość oczekiwana nowej zmiennej losowej Z określana jest jako:

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

Natomiast wariancja maleje ze wzrostem N :

$$\sigma_Z^2 = \left\langle \frac{p - p^2}{N} \right\rangle$$

Porównując teraz wyniki eksperymentalne z analitycznymi, możemy obliczyć błąd względny ϵ_X oraz $\epsilon_{\sigma_{\bar{X}}^2}$.

$$\epsilon_{\bar{X}} = \left| \frac{\bar{X} - \langle Z \rangle}{\langle Z \rangle} \right| \quad (11)$$

$$\epsilon_{\sigma_{\bar{X}}^2} = \left| \frac{\sigma_{\bar{X}}^2 - \sigma_Z^2}{\sigma_Z^2} \right| \quad (12)$$

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

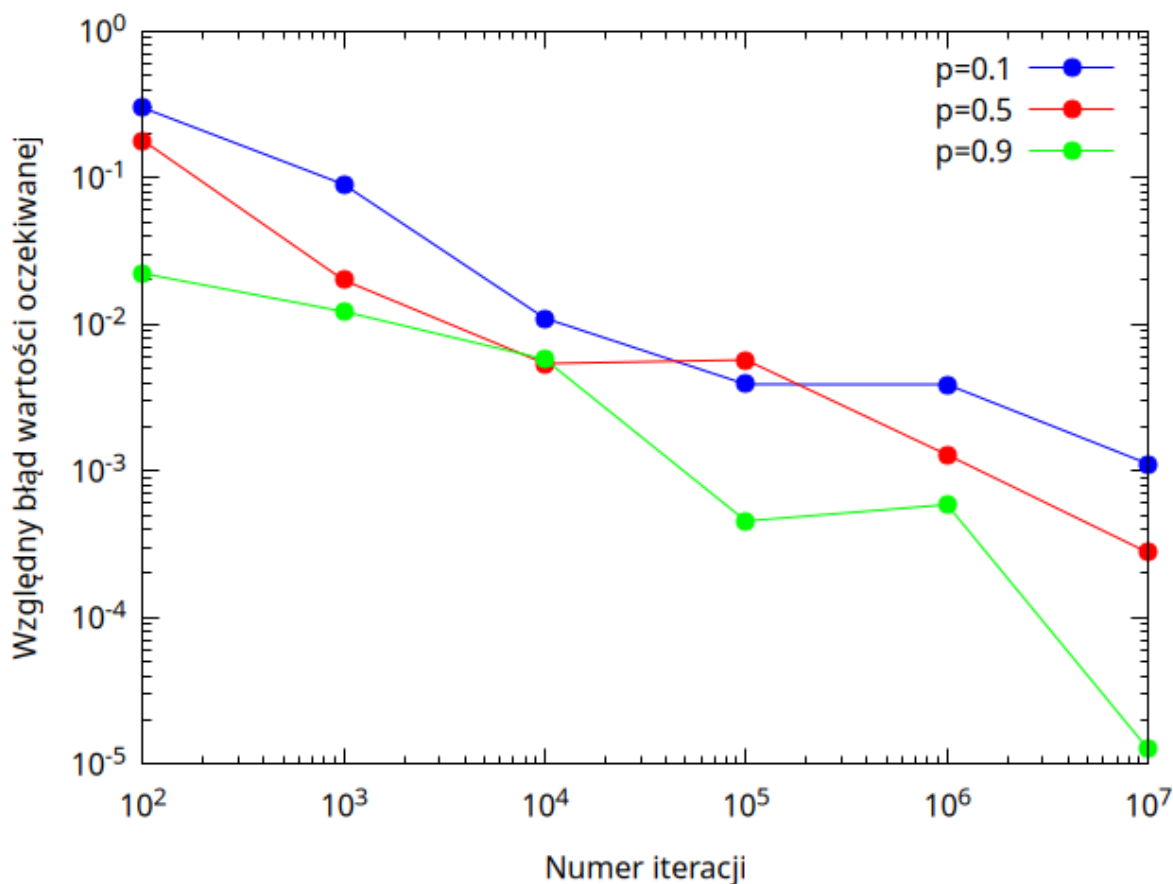
2 Wyniki

Przyjmując liczbę iteracji wynoszącą 10^7 wygenerowałem wyniki dla trzech różnych prawdopodobieństw $p = 0.1, p = 0.5, p = 0.9$. Wyniki były zapisywane do pliku, a ich zawartość odpowiednio została zestawiona w poniższych tabelach:

Tab. 1: Błędy względne wartości oczekiwanej oraz wariancji dla sumy $i = 10^k$ liczb. z rozkładu Bernoulliego w numerze iteracji

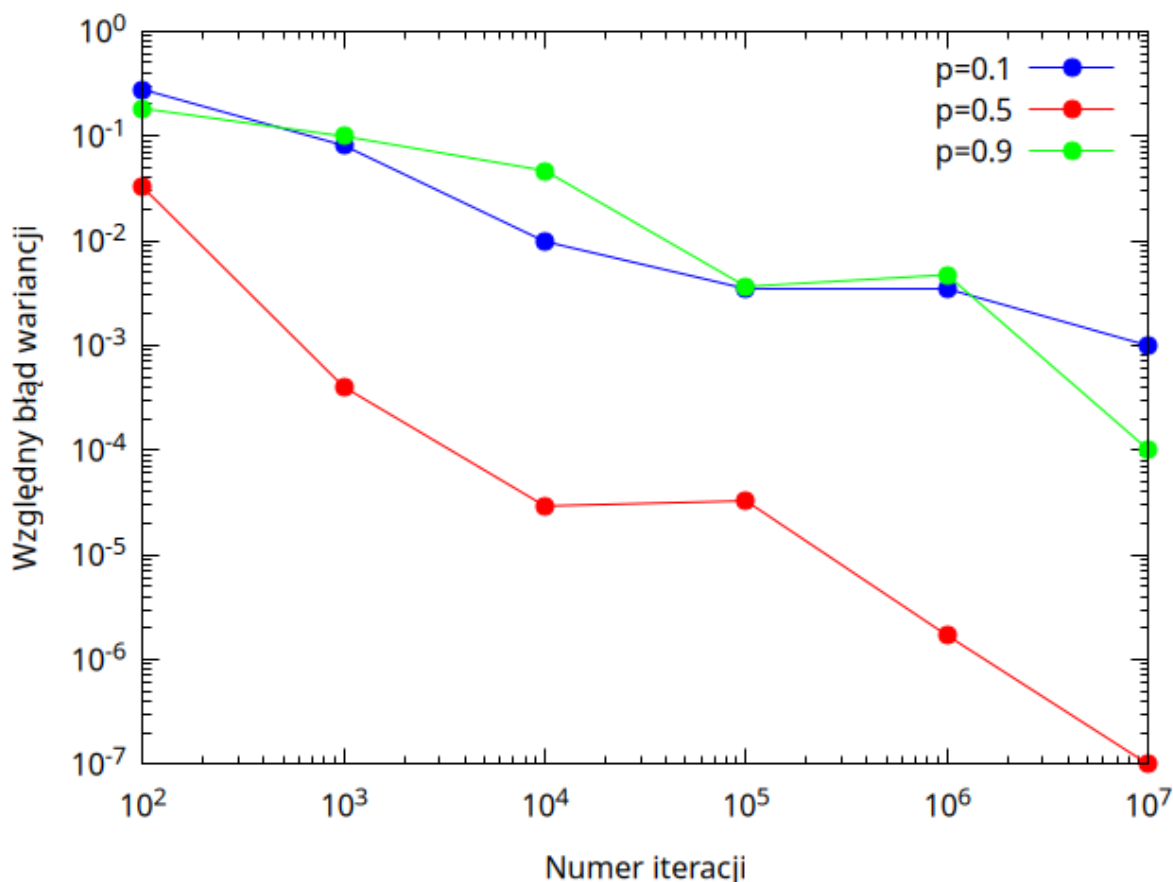
1.1: $p = 0.1$			1.2: $p = 0.5$			1.3: $p = 0.9$		
k	$\frac{\Delta EX}{EX}$	$\frac{\Delta var}{var}$	k	$\frac{\Delta EX}{EX}$	$\frac{\Delta var}{var}$	k	$\frac{\Delta EX}{EX}$	$\frac{\Delta var}{var}$
2	0.3000000	0.2766667	2	0.1800000	0.0324000	2	0.0222222	0.1822222
3	0.0900000	0.0809000	3	0.0200000	0.0004000	3	0.0122222	0.0991222
4	0.0110000	0.0097912	4	0.0054000	0.0000292	4	0.0057778	0.0465227
5	0.0039000	0.0034684	5	0.0057200	0.0000327	5	0.0004556	0.0036426
6	0.0038900	0.0034595	6	0.0012900	0.0000017	6	0.0005900	0.0047169
7	0.0011180	0.0009936	7	0.0002774	0.0000001	7	0.0000127	0.0001013

W dalszej części wygenerowałem wykresy przy użyciu programu gnuplot w odpowiednio napisanym skrypcie, który również został załączony wraz z kodem źródłowym. Każdy pogrubiony punkt odpowiada naniesionym danym z pliku. Zostały one połączone w celu zobrazowania monotoniczności danych, która to efektywnie jest malejąca, czego oczekiwaliśmy, ponieważ błąd względny w idealnym przypadku dąży do zera. Wartości wydają się nie maleć szybko, co spowodowane jest zastosowaniem osi logarytmicznej zarówno w osi X jak i Y w celu zapewnienia lepszego wglądu w analizowane dane.



Rys. 1: Wartości względnych błędów wartości oczekiwanej, obie osie są w skali logarytmicznej

Wartości dla względnych błędów wartości oczekiwanej zestawione zostały na powyższym rysunku 1. Błąd względny w jej przypadku dąży do wartości rzędu 10^{-5} . Maleje on ze względu na zwiększającą się liczebność danych, z których wyciągnięta średnia dąży do wartości teoretycznej, czyli prawdopodobieństwa p , dla zmiennej z rozkładu Bernoulliego.



Rys. 2: Wartości względnych błędów wariancji, obie osie są w skali logarytmicznej

Wartości dla względnych błędów wariancji zestawione zostały na powyższym rysunku 2. Błąd względny w jej przypadku dąży do wartości rzędu 10^{-7} . Maleje on ze względu na zwiększającą się liczebność danych, ponieważ uśredniony rozrzut danych dąży do wartości teoretycznej, czyli wartości $\sigma_Z^2 = \frac{p*(1-p)}{N}$. Jak można zauważyć średnia wartość wariancji maleje odwrotnie proporcjonalnie do liczebności danych.

3 Podsumowanie

Podsumowując, przeprowadzona analiza błędów względnych wartości oczekiwanej oraz wariancji na wykresach o skalach logarytmicznych pozwala na lepsze zrozumienie zachowania się tych wartości w kontekście wzrostu liczby próbek. Wyniki potwierdzają oczekiwane zachowanie się błędów względnych, które maleją wraz ze zwiększaniem liczby próbek.

W przypadku błędu względnego wartości oczekiwanej, malejąca tendencja wynika z faktu, że uśrednione wartości danych z rozkładu Bernoulliego zbliżają się do teoretycznej wartości prawdopodobieństwa p wraz z większą liczbą próbek.

Natomiast w przypadku błędu względnego wariancji, malejący trend jest wynikiem uśredniania rozrzutu danych, który maleje odwrotnie proporcjonalnie do liczby próbek, zgodnie z teoretycznym wzorem na wariancję dla rozkładu Bernoulliego.

Analiza ta potwierdza poprawność przeprowadzanych obliczeń oraz wykazuje, że wraz z większą ilością danych, obserwowane błędy względne maleją, co świadczy o lepszej dokładności oszacowań wartości oczekiwanej i wariancji.