Sprawozdanie z ćwiczeń laboratoryjnych z Metod Numerycznych

Gabriel Naleźnik

Wydział Fizyki i Informatyki Stosowanej
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

15.06.2020 r.

Laboratoria XV : *Całkowanie metodą Monte Carlo w 2D z rozkładem eksponencjalnym.*

1. Wstęp Teoretyczny.

Podstawowa metoda całkowania Monte Carlo - Interesuje nas wyznaczenie, a raczej estymacja, wartości oczekiwanej zmiennej losowej $\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{x})$, która jest funkcją wektora zmiennych (losowych): $\boldsymbol{x} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_m}]$. Rozkład prawdopodobieństwa zmiennej losowej z opisuje funkcja gęstości g(z): $\int_{-\infty}^{\infty} g\big(z(x)\big)\, dz = 1$, a rozkład prawdopodobieństwa wektora **x** opisuje funkcja gęstości f(x): $\int_V \; f(x)dx = 1$.

Przy takich założeniach, zgodnie z centralnym twierdzeniem granicznym, dystrybuanta wartości oczekiwanej ma rozkład normalny:

$$\lim_{N\to\infty} P\left\{\frac{|\bar{z} - \langle z\rangle|}{\frac{\sigma(z)}{\sqrt{N}}} \leq \lambda\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-\frac{u^2}{2}} du$$

Metodę Monte Carlo szacowania wartości całek w wersji podstawowej definiują wzory na:

- wartość całki:

$$I = \int_V z(x) f(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} z(x)$$

- błąd oszacowania:

$$\sigma(I) = \sqrt{\int_V (z - \langle z \rangle)^2 f(x) dx} \approx \frac{\sigma(z)}{\sqrt{N}}$$

Zazwyczaj obszarem całkowania jest określony podzbiór przestrzeni RM. W takim przypadku obliczaną całkę trzeba zapisać w nieco zmienionej postaci:

$$I = \int_V z(x) f(x) dx = \int_\Omega 1_v(x) z(x) f(x) dx$$

gdzie funkcja przynależności do zbioru $1_v(x) = \begin{cases} 1 \Leftrightarrow x \,\epsilon\, V \\ 0 \Leftrightarrow x \,\epsilon\, \Omega / V \end{cases}$

Test chi-kwadrat - Jest najczęściej stosowanym testem. Badamy w nim hipotezę że generowana zmienna losowa X ma rozkład prawdopodobieństwa o dystrybuancie F.

Jeżeli F(a) = 0 i F(b) = 1 to możemy dokonać następującego podziału zbioru wartości zmiennej X: $a < a_1 < a_2 < \cdots < a_k < b \ \ zatem \ \ p_i = P\{a_{i-1} < X \leq a_i\}, i = 1,2,3 \ldots$

Generujemy ciąg n liczb: X_1, X_2, X_3 i sprawdzamy ile z nich spełnia warunek: $a_{i-1} < X \leq a_i$ ich liczbę oznaczamy n_i.

Statystyką testu jest:

$$\chi^2_{\mathrm{k-1}} = \sum_{\mathrm{i=1}}^{\mathrm{k}} \frac{(\mathrm{n_i} - \mathrm{np_i})^2}{np_i}$$

Dla dużego n statystyka ta ma rozkład χ^2 o (k-1) stopniach swobody. Możemy tak dobrać szerokości przedziałów aby otrzymać zależność $p_i = \frac{1}{k}$ wówczas statystyka przyjmuje prostszą postać:

$$\chi^2_{\mathrm{k-1}} = \frac{\mathrm{k}}{\mathrm{n}} \sum_{\mathrm{i=1}}^{\mathrm{k}} \mathrm{n_i^2} - \mathrm{n}$$

2. Opis problemu.

Problemem w tym zadaniu z laboratoriów było oszacowanie wartości całki I, metodą Monte Carlo. Wartość dokładna całki została przyjęta jako: $\boldsymbol{I_{dok}} = \mathbf{0.2557840245}$.

$$I = \int_0^1 \int_0^1 g(x,y) * f(x) * f(y) dx\, dy$$

gdzie: $f(x) = e^{-x}$ oraz $g(x,y) = \frac{\sin(x+y)}{\ln(2+x+y)}$

Oszacowanie należało wyznaczyć używając bezpośrednio kwadratury Monte Carlo dla funkcji podcałkowej :

$$I \approx \frac{1}{n} \sum_{i=1}^{n} g(x_i, y_i) * f(x_i) * f(y_i)$$

gdzie: $x_i, y_i \in U(0,1)$

Oraz wzoru wykorzystującego zmienne z rozkładu eksponencjalnego:

$$I \approx \frac{1}{n} \sum_{i=1}^{n} g(x_i, y_i) \frac{1}{C_x C_y}$$

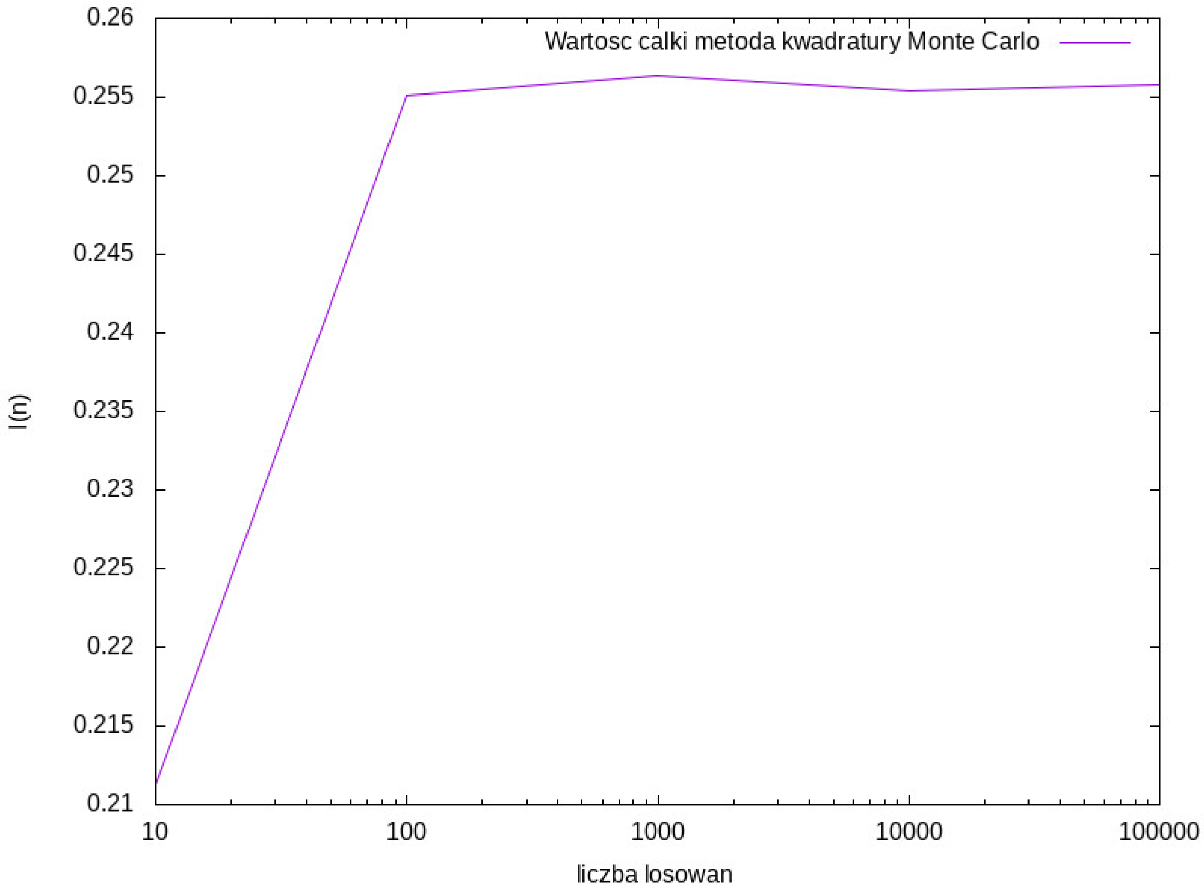
gdzie: $x_i, y_i \in Q$ (rozkład eksponencjalny)

3. Wyniki

Napisałem program w języku C, w którym zaimplementowałem wyznaczanie wartości całki metodą Monte Carlo. Poniżej zamieszczam uzyskane wyniki:

Aktualna ilość wykonanych losowań	Wyznaczona wartość całki	Odchylenie standardowe wartości średniej:
10	0.211296	2.30926e-05
100	0.25507	1.29756e-05
1000	0.256346	1.20911e-05
10000	0.255385	1.17934e-05
100000	0.255732	1.1607e-05

Tab. 3.1 Oszacowanie wartości całki w zależności od liczby wykonanych losowań, dla metody bezpośrednich kwadratur Monte Carlo .



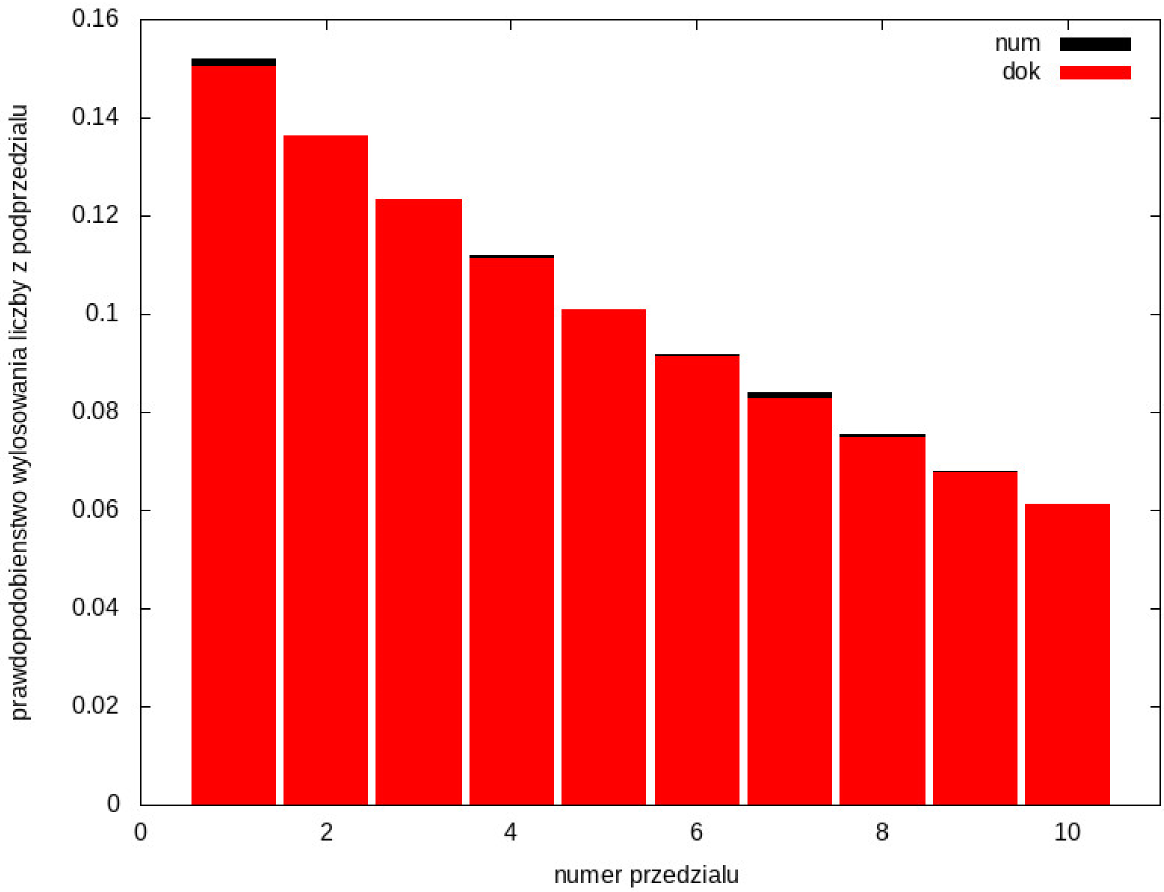
Rys. 3.1 Oszacowanie wartości całki w zależności od liczby wykonanych losowań, dla metody bezpośrednich kwadratur Monte Carlo .

Aktualna ilość wykonanych losowań	Wyznaczona wartość całki	Odchylenie standardowe wartości średniej:
10	0.257924	1.45711e-05
100	0.254227	1.31928e-05
1000	0.250227	1.69301e-05
10000	0.255817	1.54934e-05
100000	0.255756	1.54628e-05

Tab. 3.2 Oszacowanie wartości całki w zależności od liczby wykonanych losowań, dla metody bezpośrednich kwadratur Monte Carlo z rozkładem eksponencjalnym .

Rys. 3.2 Oszacowanie wartości całki w zależności od liczby wykonanych losowań, dla metody bezpośrednich kwadratur Monte Carlo z rozkładem eksponencjalnym.

Jak można zauważyć, wartość całki oszacowana przy pomocy wzoru wykorzystującego rozkład eksponencjalny, jest lepszym oszacowaniem wartości teoretycznej mimo że posiada większe odchylenie standardowe wartości średniej przybliżenia.



Rys. 3.3 Histogram prawdopodobieństwa wylosowania liczby z danego podprzedziału (0, 1) dla rozkładu Q wyznaczonego teoretycznie oraz za pomocą wygenerowanych danych

Rozkład pozyskanych z generatora Q, 10^5punktów z przedziału (0, 1) prawie pokrywa się z teoretycznym prawdopodobieństwem wylosowania liczby z danego podprzedziału, wyznaczonego za pomocą wzoru:

$$\mathrm{P_i}(\mathrm{x_{i-1}} < x \leq \mathrm{x_i}) = \mathrm{F}(\mathrm{x_i}) - \mathrm{F}(\mathrm{x_{i-1}})$$

Świadczy to o dobrej jakości generatora Q, jako generatora dla zmiennej o rozkładzie eksponencjalnym.

Wartość statystyki χ^2_{k-1} dla tego generatora wynosi: $\mathbf{12.1081}$

4. Wnioski

Metoda Monte Carlo okazała się być skuteczną metodą do oszacowania wartości całki wielokrotnej. Wykorzystanie większej ilości wylosowanych liczb zwiększyło dokładność uzyskanego przybliżenia. Ponadto, wzór wykorzystujący rozkład eksponencjalny również dał dokładniejsze przybliżenie.