

Metody Monte Carlo - laboratorium 4

Temat: Całkowanie z wykorzystaniem metod Monte Carlo

Zadanie 1. Całkowanie metodą podstawową Monte Carlo

Stosując **metodę podstawową** Monte Carlo obliczyć wartość całki wraz z odchyleniem standardowym:

$$I = \int_{-4}^4 e^{-x^2} dx$$

Uwaga:

$$\int_a^b f(x) dx = (b-a) \int_a^b f(x) dP \quad \text{gdzie} \quad dP = \frac{dx}{b-a}$$

Wobec powyższego wartość całki I na przedziale (a, b) można interpretować jako pole prostokąta o długości podstawy $a - b$ i wysokości równej wartości oczekiwanej zmiennej losowej $Y = f(X)$, gdzie X oznacza zmienną losową o rozkładzie równomiernym na przedziale (a, b) . Odchylenie standardowe estymaty całki wyznaczamy z zależności:

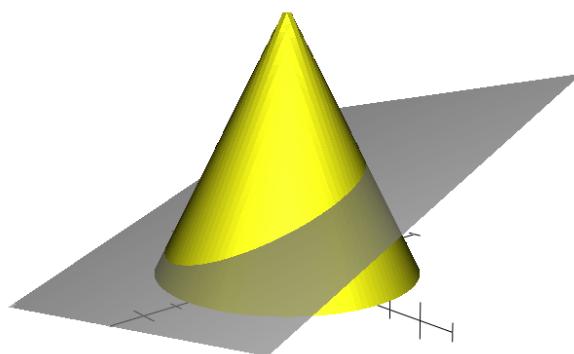
$$\hat{\sigma}(\hat{I}) = V \sigma(\hat{J}), \quad \text{gdzie}$$

$$\sigma(\hat{J}) = \frac{1}{\sqrt{N(N-1)}} \sqrt{\sum_{i=1}^N f^2(x^{(i)}) - \frac{1}{N} \left[\sum_{i=1}^N f(x^{(i)}) \right]^2}$$

a V – objętość obszaru, po którym dokonujemy całkowania (w tym przypadku jednowymiarowego).

Zadanie 2. Całkowanie metodą orzeł-reszka

Dany jest stożek o wysokości 4 i promieniu podstawy 2, przy czym podstawa ta leży na płaszczyźnie $z = 0$. Stosując metodę **orzeł-reszka** (ang. **hit and miss**) obliczyć objętość tego fragmentu stożka, który znajduje się ponad płaszczyzną opisaną równaniem $z = 0,2x + 0,4y + 1,2$.



Metody Monte Carlo - laboratorium 4

Proszę prześledzić zbieżność wyniku oraz jego niepewności, rozumianej jako podwojona wartość estymaty jego odchylenia standardowego $2\hat{\sigma}(\hat{V}_{stożka}) = 2V\hat{\sigma}(\hat{I})$, gdzie V jest objętością obszaru

próbkiowania zaś $\hat{\sigma}(\hat{I}) = \sqrt{\frac{1}{N} \frac{M}{N} \left(1 - \frac{M}{N}\right)}$, przy czym M oznacza liczbę punktów znajdujących się wewnątrz interesującej nas części stożka, zaś N – liczbę wszystkich wylosowanych punktów w przestrzeni R^3 .

Zadanie 3. Ograniczenie wariancji estymaty całki – zamiana zmiennych

Niech $f(x) = \sin x$. Obliczyć analitycznie wartość całki

$$I = \int_0^{\pi/2} \sin x \, dx.$$

Wartość tej całki należy również oszacować metodą podstawową MC. Następnie dokonać zamiany

zmiennych $x = g(z) = \sqrt{z}$ i scałkować metodą podstawową MC wyrażenie $f[g(z)] \frac{dg}{dz}$

(patrz *Przypomnienie* poniżej). Jako początkową wartość liczności próby MC dla obu przypadków (całkowanie z zamianą zmiennych i bez niej) przyjąć np. 100, a następnie zwiększać ją co 10 do momentu, w którym obliczenia staną się nieakceptowalnie długie. Dla każdej wartości N zaobserwować \hat{I} (estymatę całki), jej niepewność oraz wartość bezwzględną błędu estymaty całki $\Delta_I = |\hat{I} - I|$. Przedstawić na wspólnym wykresie zależność niepewności oraz błędu od liczności próby dla „zwykłego” całkowania MC oraz całkowania z użyciem zamiany zmiennych.

Przypomnienie

Błąd całkowania MC można ograniczyć stosując przekształcenie, dzięki któremu funkcja podcałkowa stanie się możliwie bliska stałej. Jedną z metod jest zamiana zmiennych. Jeśli wprowadzimy nową zmienną z , taką że $x = g(z)$, całka przyjmie postać

$$I = \int_{x_{\min}}^{x_{\max}} f(x) \, dx = \int_{z_{\min}}^{z_{\max}} f[g(z)] \frac{dg}{dz} \, dz$$

gdzie $z_{\min} = g^{-1}(x_{\min})$, $z_{\max} = g^{-1}(x_{\max})$. Możliwe jest takie dobranie funkcji $g(z)$, by wyrażenie podcałkowe miało wariancję mniejszą niż $f(x)$. W przypadku naszego zadania można się o tym przekonać, wykreślając wyrażenie podcałkowe na przedziale $[z_{\min}, z_{\max}]$.

Metody Monte Carlo - laboratorium 4

Zadanie 4. Ograniczenie wariancji estymaty całki – zmienne kontrolne

Wyświetlić na jednym wykresie funkcje: $f(x) = \sin \pi x$, $g(x) = -4x(x-1)$ oraz ich różnicę w przedziale $x \in [0, 1]$. Obliczyć analitycznie wartość całki

$$I = \int_0^1 \sin \pi x \, dx.$$

Następnie oszacować wartość tej całki metodą podstawową MC z zastosowaniem zmiennej kontrolnej $g(x)$ oraz bez niej (patrz *Przypomnienie* poniżej). Obie estymaty można wyznaczyć w tym samym eksperymencie, gdyż metoda zmiennej kontrolnej sprowadza się do użycia poprawki addytywnej. Jako początkową wartość liczności próby MC przyjąć np. 100, a następnie zwiększać ją co 10 do momentu, w którym obliczenia staną się nieakceptowalnie długie. Dla każdej wartości N zaobserwować \hat{I} (estymatę całki), jej niepewność oraz wartość bezwzględną błędu estymaty całki $\Delta_I = |\hat{I} - I|$. Przedstawić na wspólnym wykresie zależność niepewności oraz błędu od liczności próby dla „zwykłego” całkowania MC oraz całkowania z użyciem zmiennej kontrolnej.

Przypomnienie

Metoda zmiennych kontrolnych polega na wykorzystaniu przy całkowaniu zależności

$$\int f(x) \, dx = \int [f(x) - c g(x)] \, dx + c \int g(x) \, dx,$$

gdzie c jest pewną stałą. Jeżeli wartość całki $g(x)$ potrafimy wyznaczyć analitycznie, metodą MC szacujemy jedynie składnik pierwszy, czyli całkę różnicy obu funkcji. Jeśli funkcje te mają na przedziale całkowania podobny przebieg, wariancja ich różnicy jest znacznie mniejsza od wariancji samej $f(x)$. Pozwala to na znaczące ograniczenie niepewności szacowania.

W powyższym przykładzie całkowanie odbywa się na przedziale $[0, 1]$, więc sprowadza się do wyznaczania wartości średniej funkcji podcałkowej. Pozwala to na uproszczenie zapisu powyższego wyrażenia:

$$E^*[f(x)] = E[f(x) - c g(x)] + c \bar{g} = E[f(x)] - c E[g(x) - \bar{g}], \quad (1)$$

gdzie \bar{g} jest (znaną) wartością oczekiwaną $g(x)$ a gwiazdka sugeruje, że mamy do czynienia z „ulepszoną” postacią estymaty. Wariancja powyższego wyrażenia to po prostu wariancja różnicy dwóch skorelowanych zmiennych losowych:

$$\text{Var}(E^*[f(x)]) = \text{Var}(E[f(x)] - 2c \text{Cov}(f(x), g(x)) + c^2 \text{Var}(E[g(x)]),$$

Metody Monte Carlo - laboratorium 4

gdzie $\text{Var}(\cdot)$ i $\text{Cov}(\cdot, \cdot)$ to odpowiednio operatory wariancji i kowariancji. Wariancja ta jest minimalna, jeśli

$$c = \frac{\text{Cov}(E[f(x)], E[g(x)])}{\text{Var}(E[g(x)])} = \frac{\text{Cov}(f(x), g(x))}{\text{Var}(g(x))}. \quad (2)$$

O ile wariancję $g(x)$ można często wyznaczyć analitycznie, o tyle wartość kowariancji $f(x)$ i $g(x)$ musi być oszacowana numerycznie – zwykle w tym samym eksperymencie MC, w którym wyznaczana jest $E^*[f(x)]$. Może do tego posłużyć wzór:

$$\text{Cov}(f(x), g(x)) = \frac{1}{N} \sum_{n=1}^N f(x_n) g(x_n) - \frac{1}{N^2} \left(\sum_{n=1}^N f(x_n) \right) \left(\sum_{n=1}^N g(x_n) \right) \quad (3)$$

W przypadku naszego zadania należy postąpić następująco:

1. Dla funkcji $g(x)$ obliczyć (analitycznie) wartość oczekiwaną \bar{g} oraz wariancję $\text{Var}(g(x))$.
2. Dla liczności próby N zmieniającej się w szerokim zakresie (np. od 10^2 do 10^8 lub więcej):
 - 2.1. Dla funkcji $f(x)$ i $g(x)$ wyznaczyć estymaty wartości średnich $E[f(x)]$ i $E[g(x)]$ oraz – za pomocą wzoru (3) – kowariancji obu funkcji.
 - 2.2. Wyznaczyć optymalną wartość współczynnika c za pomocą wzoru (2).
 - 2.3. Skorygować wartość estymaty wartości średniej $f(x)$ za pomocą wzoru (1).

Zadanie 5. Ograniczenie wariancji estymaty całki – próbkowanie ważone

Niech $f(x) = \sin x$. Obliczyć analitycznie wartość całki

$$I = \int_0^{\pi/2} \sin x \, dx.$$

Wartość tej całki należy również oszacować metodą podstawową MC. W kolejnym całkowaniu MC zastosować próbkowanie z rozkładu $p(x) = 8x/\pi^2$ na przedziale $(0, \pi/2)$ (patrz *Przypomnienie* poniżej). Zmienne z tego rozkładu można uzyskać dowolną znaną metodą, np. odwracania dystrybuanty lub eliminacji. Jako początkową wartość liczności próby MC dla obu przypadków (całkowanie z zamianą zmiennych i bez niej) przyjąć np. 100, a następnie zwiększać ją co 10 do momentu, w którym obliczenia staną się nieakceptowalnie długie. Dla każdej wartości N zaobserwować \hat{I} (estymatę całki), jej niepewność oraz wartość bezwzględną błędu estymaty całki $\Delta_I = |\hat{I} - I|$. Przedstawić na wspólnym wykresie zależność niepewności oraz błędu od liczności próby dla „zwykłego” całkowania MC oraz całkowania z próbkowaniem ważonym.

Metody Monte Carlo - laboratorium 4

Przypomnienie

Próbkowanie ważone polega na uprzywilejowaniu tych zakresów argumentu, dla których funkcja podcałkowa $f(x)$ przyjmuje wartości najbardziej różne od zera, a zatem wnosi największy wkład do wartości całki. Wartości argumentu losuje się więc z takiego rozkładu prawdopodobieństwa $p(x)$, który możliwie dobrze przybliży zmienność $f(x)$. Korzysta się przy tym z następującej tożsamości

$$I = \int_{x_{\min}}^{x_{\max}} f(x) dx = \int_{x_{\min}}^{x_{\max}} \left[\frac{f(x)}{p(x)} \right] p(x) dx.$$

Całkowaniu podlega więc wyrażenie $f(x)/p(x)$, natomiast drugi czynnik $p(x)$ jest funkcją gęstości prawdopodobieństwa argumentu.

Zadanie 6. Adaptacyjne próbkowanie ważone i warstwowe

Zapoznać się z zasadą działania algorytmów całkowania *plain*, *vegas* i *miser*, zaimplementowanych w bibliotece *gsl*. Następnie za pomocą każdego z nich wyznaczyć wartości całek:

a) $I_{1D} = \int_{-\pi}^{\pi} 1 + \cos x \, dx$

b) $I_{3D} = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} (1 + \cos x)(1 + \cos y)(1 + \cos z) \, dx \, dy \, dz$

Jako początkową wartość parametru N (liczność próby MC) przyjąć np. 100, a następnie zwiększać ją co 10 do momentu, w którym obliczenia staną się nieakceptowalnie długie. Dla każdej wartości N zaobserwować wynik \hat{I} (estymatę całki), jego niepewność oraz wartość bezwzględnej błęd estymaty całki $\Delta_I = |\hat{I} - I|$. Wykreślić zależność niepewności oraz błęd od liczności próby dla wszystkich trzech metod. Porównać wyniki dla funkcji a) i b).

Poniżej przykład dla algorytmu *plain* i funkcji b).

```
#include <gsl/gsl_rng.h>
#include <stdlib.h>
#include <gsl/gsl_math.h>
#include <gsl/gsl_monte.h>
#include <gsl/gsl_monte_plain.h>
using namespace std;
```

Metody Monte Carlo - laboratorium 4

```
const int DIM = 3; // Liczba wymiarow dziedziny
const double wynik_dokladny_3D = . . . ; // Tu wpisujemy prawdziwa wartosc calki
double funkcja_przykladowa (double *x, size_t dim, void *p) {
    return (1+cos(x[0]))*(1+cos(x[1]))*(1+cos(x[2]));
}
int main(void) {
    gsl_rng *generatorek;
    double wynik_plain, niepewnosc_plain;
    double xmin[DIMENSION] = {-M_PI, -M_PI, -M_PI };
    double xmax[DIMENSION] = { M_PI,  M_PI,  M_PI };
    size_t N = 1000000;
    gsl_monte_function F = { &funkcja_3D, DIM, NULL };
    generatorek = gsl_rng_alloc(gsl_rng_default);
    gsl_monte_plain_state *stan = gsl_monte_plain_alloc (DIM);
    gsl_monte_plain_integrate (&F, xmin, xmax, DIM, N, generatorek,
                                stan, &wynik_plain, &niepewnosc_plain);
    printf("%d\t%.2e\t%.2e\n",
           N, niepewnosc_plain, fabs(wynik_plain - wynik_dokladny_3D)
    );
    gsl_monte_plain_free(stan);
    gsl_rng_free(generatorek);
}
```

Zadanie 7. Porównanie próbkowania pseudolosowego i quasi-losowego w całkowaniu

Obliczyć analitycznie wartość całki

$$I = \int_0^{\pi} \sin x \, dx$$

Następnie:

- Oszacować wartość tej całki metodą **orzeł-reszka** dla liczności próby 1000 lub większej. Obliczenia powtórzyć w pętli kilka lub kilkanaście razy. Co można powiedzieć o wartościach błędu (nie niepewności!) wnoszonego przez tę metodę? Na ile są one powtarzalne?
- Przyjąć postępowanie podobne jak w punkcie (a), lecz zamiast generatora pseudolosowego użyć wybranego generatora liczb **quasi-losowych** (punkt *Quasi-Random Sequences* instrukcji do biblioteki *gsl*). Porównać błąd metody quasi-Monte Carlo z błędem metody użytej w punkcie (a).

Poniżej przedstawiono przykład użycia generatora quasi-losowego ciągu Sobola – można również wykorzystać ciąg Niederreitera lub Haltona:

Metody Monte Carlo - laboratorium 4

```
#include <stdio.h>
#include <gsl/gsl_qrng.h>
int main (void) {
    double punkt[2];
    gsl_qrng *qgenerator = gsl_qrng_alloc (gsl_qrng_sobol, 2);
    for (int i = 0; i < 1024; i++) {
        gsl_qrng_get (qgenerator, punkt);
        printf ( "%.5f %.5f\n", punkt[0], punkt[1]);
    }
    gsl_qrng_free (qgenerator);
    return 0;
}
```