

Całkowanie Monte Carlo

Zagadnienia związane z wykorzystaniem MC do całkowania numerycznego

- podstawowa metoda całkowania
- testowanie jakości rozwiązania
- sposoby redukcji wariancji:
 - losowanie ważone
 - losowanie warstwowe
 - metoda zmiennej kontrolnej
 - metoda zmiennych antytetycznych
 - redukcja wymiarowości
 - estymator obciążony

Literatura:

- D.P. Kroese, T. Taimre, Z.I. Botev „Handbook of Monte Carlo Methods”
- M.H. Kalos, P.A. Withlock „Monte Carlo Methods”
- W.L. Dunn, J.K. Shultis „Exploring Monte Carlo Methods”

Naszym pierwszym „poważnym” celem stosowania Monte Carlo będzie **szacowanie** wartości całek oznaczonych. Załóżmy, że całka istnieje (ma skończoną wartość) i ma postać

$$C = \int_{-\infty}^{\infty} z g(z) dz$$

Jak znaleźć jej wartość?

Problemowi możemy nadać interpretację statystyczną:

- 1) zmienną z potraktujemy jako **zmienną losową**
- 2) funkcję $g(z)$ traktujemy jako **fgp rozkładu zmiennej z**

wówczas zagadnienie możemy przedstawić jako problem estymacji wartości oczekiwanej zmiennej losowej z

$$C = \langle z \rangle = \int_{-\infty}^{\infty} z g(z) dz \approx \bar{z} = \frac{1}{N} \sum_{i=1}^N z_i$$

z_i losowane z rozkładu danego $g(z)$

$$\int_{-\infty}^{\infty} g(z) dz = 1$$

Dość często jednak $g(z)$ jest nieznane

- może to być wynik działania skomplikowanego algorytmu
- rozkład możemy dopiero próbować tworząc łańcuch Markowa

Ale zmienna z najczęściej jest funkcją, której argumenty są zmiennymi losowymi

$$z = z(\vec{x}), \quad \vec{x} = [x_1, x_2, \dots, x_n]$$

o znanym rozkładzie

$$f(\vec{x}) = f(x_1, x_2, \dots, x_n)$$

Aby policzyć całkę dokonujemy transformacji zmiennych wykorzystując warunek ($dz > 0$) i

$$P\{Z \leq z\} = F(z) = \int_{-\infty}^z g(z') dz' = P\{\{X_i : X_i \leq x_i\}\} = \int d^n x f(\vec{x})$$

$$g(z) dz = f(\vec{x}) d^n x$$

co prowadzi do alternatywnego sformułowania problemu

$$\langle z \rangle = \int_{-\infty}^{\infty} z g(z) dz = \int_{a_1}^{b_1} dx_1 \dots \int_{a_n}^{b_n} dx_n z(\vec{x}) f(\vec{x})$$

wartość całki liczymy podobnie

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(\vec{x}_i), \quad \vec{x}_i \sim Dist\{f(\vec{x})\}$$

Losowanie możemy przeprowadzić na 2 sposoby

1) **rozkład nieskorelowany** – każda zmienna niezależna od innych, granice całek ustalone

$$f = f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n) \quad x_i \in [a_i, b_i], \quad i = 1, 2, \dots, n$$

przykład:
wielowymiarowy
rozkład normalny $N^n(\mu, \sigma)$

$$f(\vec{x}) = \left(\prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left[- \left(\frac{x_1^2}{2\sigma_1^2} + \frac{x_2^2}{2\sigma_2^2} + \dots + \frac{x_n^2}{2\sigma_n^2} \right) \right]$$

$$= \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{x_1^2}{2\sigma_1^2}} \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{x_2^2}{2\sigma_2^2}} \dots \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{x_n^2}{2\sigma_n^2}}$$

$$= f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n)$$

zmienne losujemy niezależnie
z indywidualnych rozkładów

2) **rozkład skorelowany** – wykorzystujemy ciąg rozkładów warunkowych, granice całek też są funkcjami zmiennych

$$f = f(x_1, x_2, \dots, x_n) = f_n(x_n | x_{n-1} \dots x_1) \dots f_3(x_3 | x_2 x_1) \cdot f_2(x_2 | x_1) \cdot f_1(x_1)$$

$$\begin{array}{l} \text{sekwencja losowania:} \\ x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_n \end{array} \quad \left\{ \begin{array}{l} x_1 \in [a_1, b_1] \\ x_2 \in [a_2(x_1), b_2(x_1)] \\ x_3 \in [a_3(x_1, x_2), b_3(x_1, x_2)] \\ \vdots \\ x_n \in [a_n(x_1, x_2, \dots, x_{n-1}), b_n(x_1, x_2, \dots, x_{n-1})] \end{array} \right.$$

$$\text{losowanie : } x_1 \leftarrow f_1(x_1) = \int_{a_n}^{b_n} dx_n \dots \int_{a_2}^{b_2} dx_2 f(x_1, \underbrace{x_2, x_3, \dots, x_n}) \quad - \text{ określa rozkład } x_1 \text{ niezależnie od innych}$$

$$\text{losowanie : } x_2 \leftarrow f_2(x_2 | x_1) = \int_{a_n}^{b_n} dx_n \dots \int_{a_3}^{b_3} dx_3 f(x_1, x_2, \underbrace{x_3, x_4, \dots, x_n}) \quad - \text{ rozkład } x_2 \text{ dla ustalonego } x_1$$

$$\text{losowanie : } x_3 \leftarrow f_3(x_3 | x_1, x_2) = \int_{a_n}^{b_n} dx_n \dots \int_{a_4}^{b_4} dx_4 f(x_1, x_2, x_3, \underbrace{x_4, x_5, \dots, x_n}) \quad - \text{ rozkład } x_3 \text{ dla ustalonego } x_1 \text{ i } x_2$$

$$\text{losowanie : } x_n \leftarrow f_n(x_n | x_1, x_2, \dots, x_{n-1}) = f(\underbrace{x_1, x_2, \dots, x_{n-1}}_{\text{ustalone}}, x_n) \quad - \text{ rozkład } x_n, \text{ reszta ustalona w poprzednich losowaniach}$$

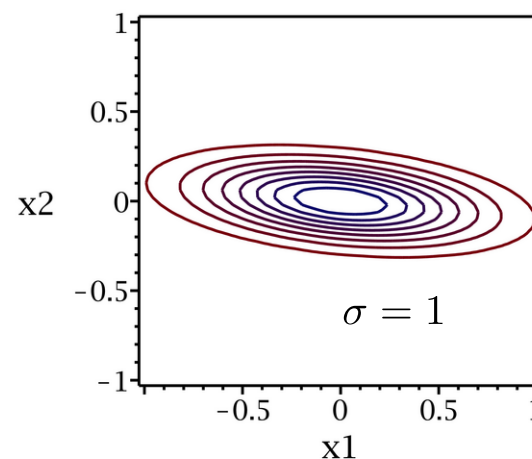
przykład rozkładu skorelowanego

$$\underbrace{f(y_1, y_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{y_1^2 + y_2^2}{2\sigma^2}}}_{\text{nieskorelowany}} \Rightarrow \left\{ \begin{array}{l} y_1 = x_1 + 7x_2 \\ y_2 = 2x_1 - x_2 \end{array} \right\} \Rightarrow \underbrace{f(x_1, x_2) = \frac{1}{\frac{2}{15}\pi\sigma^2} e^{-\frac{5x_1^2 + 10x_1x_2 + 50x_2^2}{2\sigma^2}}}_{\text{skorelowany}}$$

losowanie: $x_1 \leftarrow f_1(x_1) = \frac{1}{\sigma \frac{2}{3}\sqrt{\pi}} e^{-\frac{x_1^2}{\frac{4}{9}\sigma^2}}$

zmienił się rozkład (rozmycie): $\sigma_1 = \frac{\sqrt{2}}{3}\sigma$

Dla ustalonego x_1 , zmienną x_2 losujemy z $f(x_1, x_2)$.



Całkowanie bez jawnej postaci fgp argumentów funkcji – musimy ją wprowadzić

$$\langle z \rangle = \int_{\Omega} z(\vec{x}) d^n x = \int_{\Omega} \frac{z(\vec{x})}{f(\vec{x})} f(\vec{x}) d^n x$$

Musimy określić fgp wektora losowego x - z pierwotnej postaci wynika że jest to stała

$$f(\vec{x}) = C = \text{const} \quad \implies \quad \int_{\Omega} f(\vec{x}) d^n x = C \int_{\Omega} 1 d^n x = C \Omega = 1$$

$$f(\vec{x}) = \frac{1}{\Omega} \quad - \text{rozkład wektora losowego } x \text{ jest jednorodny w } \Omega$$

teraz możemy użyć jawnej postaci fgp

$$\langle z \rangle = \Omega \int_{\Omega} z(\vec{x}) f(\vec{x}) d^n x$$

teraz możemy szacować wartość całki licząc średnią (wektory x losujemy z rozkładu jednorodnego)

$$\langle z \rangle \approx \bar{z} = \frac{\Omega}{N} \sum_{i=1}^N z(\vec{x}_i), \quad \vec{x}_i \sim U^n(\Omega)$$

Wzór ten nazywany jest czasem „kwadraturą Monte Carlo” przez jego podobieństwo do kwadratur Newtona-Cotesa/Gaussa.

Ponieważ nadaliśmy interpretację statystyczną funkcji podcałkowej $z(x)$ (teraz: zmienna losowa) więc jej oszacowanie obarczone jest niepewnością („błędem”) – miarą niepewności jest wariancja.

Liczmy drugi moment

$$\langle z^2 \rangle = \int_{\Omega} [\Omega z(\vec{x})]^2 f(\vec{x}) d^n x$$

$$\langle z^2 \rangle \approx \overline{z^2} = \frac{\Omega^2}{N} \sum_{i=1}^N [z(\vec{x}_i)]^2$$

Mając 1 i 2 moment liczymy wariancję w standardowy sposób

$$\sigma_z^2 = \langle z^2 \rangle - \langle z \rangle^2 \approx \overline{z^2} - \bar{z}^2$$

$$\sigma_{\bar{z}} = \frac{\sigma_z}{\sqrt{N}} \approx \sqrt{\frac{\overline{z^2} - \bar{z}^2}{N}}$$

- precyzja/rozrzut wyniku jawnie zależy od liczby losowań (mianownik), ale również od wartości obu momentów
- w metodzie MC dążymy do redukcji wariancji, utrzymując stałe N (zakładamy, że losowanie jest kosztowne)
– zatem pozostaje nam manipulacja wartościami licznika

Test jakości rozwiązania

Ponieważ w obliczeniach używamy metody statystycznej otrzymany wynik nie jest dokładny.

Możemy zadać dwa pytania

- jak dokładny jest wynik?
- jaki jest rozrzut kolejnych wyników cząstkowych wokół otrzymanej wartości

Dokładność wyniku – to różnica pomiędzy wynikiem dokładnym (tego nie znamy) a wynikiem otrzymanym z obliczeń. Jeśli algorytm MC jest poprawny a obliczenia zostały wykonane prawidłowo to zgodnie z CTG i prawem wielkich liczb spodziewamy wzrostu dokładności przy wzroście liczby losowań N .

Uwaga: dokładność wyniku może zostać zaburzona przez wpływ błędu systematycznego
np. błędne oszacowanie wartości funkcji podcałkowej, użycie generatora o złych własnościach statystycznych

Precyzję wyniku określa szerokość przedziału ufności wokół uzyskanej wartości i wyrażamy ją za pomocą wariancji/odchylenia standardowego. Oczywiście oczekujemy jak najmniejszych wartości wariancji (wysokiej precyzji) a ta maleje jak odwrotność pierwiastka z liczby losowań.

Błąd względny

Daje informację o relacji błędu/niepewności do uzyskanej wartości

$$R = \frac{\sigma_{\bar{z}}}{\bar{z}} = \frac{\sqrt{\frac{z^2 - \bar{z}^2}{N}}}{\bar{z}}$$

przypadki szczególne: $z_i = \text{const}, \quad i = 1, 2, \dots, N \quad \implies \quad R = 0$ -brak rozrzutu wyinków losowań

$\text{sign}(z_i) = \text{const}, \quad i = 1, 2, \dots, N \quad \implies \quad \bar{z}^2 > \bar{z}^2 \quad \implies \quad R < 1$

Oczekujemy małego błędu względnego

$$R < 0.05$$

Wydajność algorytmu MC – *figure of merit*

Celem jest uzyskanie małej wariancji dla małej liczby losowań N .

Do określenia wydajności metody można użyć specjalnej funkcji (*figure of merit*)

$$FOM = \frac{1}{R^2 \cdot Time} \approx C = const$$

$$R^2 \sim \frac{1}{N}$$

$$Time \sim N$$

Porównując różne metody redukcji wariancji, wykonuje się krótkie serie losowań i wyznacza FOM. Metoda dająca największą wartość jest w danym zastosowaniu najwydajniejsza.

Znając FOM można określić przybliżony czas (liczbę losowań N) potrzebną do uzyskania założonej wartości błędu względnego R

$$Time \approx \frac{1}{R^2 \cdot FOM}$$

Wariancja wariancji (*variance of variance = VOV*)

Znajomość wszystkich momentów rozkładu daje pełną informację o rozkładzie – ale to nie jest w praktyce możliwe.

Możemy jednak wyznaczać wyższe momenty, które dostarczą dodatkowych informacji o wyniku.

Względny błąd wariancji (VOV)

$$VOV = \frac{\sigma_{\bar{z}}^4}{\sigma^2(\sigma_{\bar{z}}^2)} = \frac{\sum_{i=1}^N (z_i - \bar{z})^4}{\left[\sum_{i=1}^N (z_i - \bar{z})^2 \right]^2} - \frac{1}{N}$$

VOV jest miarą rozrzutu wariancji, z przeprowadzonych licznych testów statystycznych wynika, że wiarygodny przedział ufności dostaniemy dla warunku

$$VOV < 0.1$$

w praktyce wykorzystujemy sumy S_m będące składnikami 4 najniższych momentów

$$S_m = \sum_{i=1}^N z_i^m \quad \Rightarrow \quad VOV = \frac{S_4 - \frac{4}{N} S_1 S_3 + \frac{8}{N^2} S_2 S_1^2 - \frac{4}{N^3} S_1^4 - \frac{1}{N} S_2^2}{\left[S_2 - \frac{1}{N} S_1^2\right]^2}$$

Wartość VOV powinna maleć jak $1/N$ $VOV(N) \sim \frac{1}{N}$

Uwaga:

Dotychczasowe rozważania wskazują, że zwiększanie liczby losowań N prowadzi do uzyskania lepszego wyniku. Należy jednak pamiętać, że zmniejszenie błędu 10 krotnie wymusza zwiększenie liczby N 100 razy. Jeśli N losowań zajmuje nam 1 godzinę to $100N$ daje ...

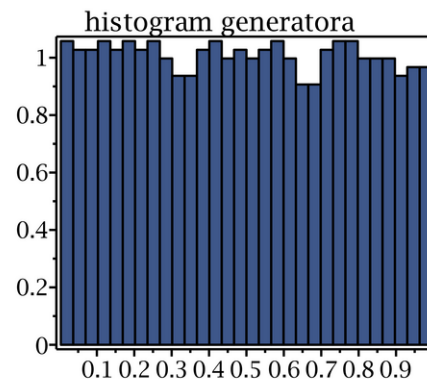
- wniosek jest oczywisty: zwiększanie N jest ostatnią deską ratunku w symulacjach MC, algorytmy MC należy optymalizować pod kątem wydajności.

Przykład. Liczymy całkę używając generatora U(0,1) o małym okresie

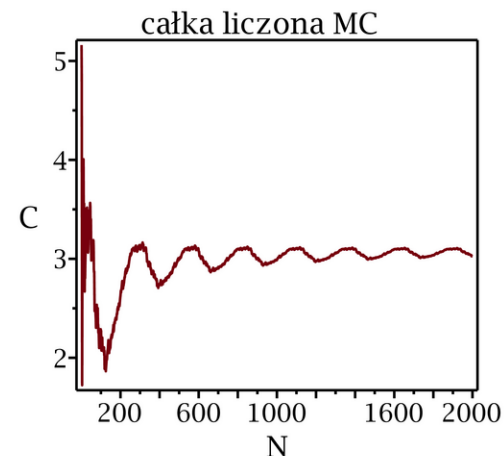
generator: $X_{i+1} = a \cdot X_i \mod m, \quad (X_0 = 11, a = 163, m = 269)$

$$C = \int_0^5 \ln(x) dx = 3.047$$

$$C \approx \frac{\Omega}{N} \sum_{i=1}^N \ln(x_i), \quad (\Omega = 5)$$



(taki sobie ten generator)



Oscylacje to wynik powtarzającego się ciągu liczb pseudolosowych. Amplituda oscylacji (błąd) maleje jak $1/\sqrt{N}$ - czyli wolno.

zbiega się do dobrej wartości

Problem może też stanowić sposób akumulowania wielkości w trakcie sumowania – problem dodawania małych i wielkich liczb.

Pojedyncza precyzja

– trzymamy 8 miejsc po przecinku/kropce

$$a = 1.123980980912390$$

$$b = 0.00000008908$$

$$a + b = 1.12398107$$

wszystkie cyfry powyżej 8 miejsca zostają skasowane

Podwójna precyzja

– trzymamy 15 miejsc po przecinku/kropce

$$a = 1.123980980912390$$

$$b = 0.00000008908$$

$$a + b = 1.123981069992390$$

- końcowe wyniki różnią się nieznacznie, ale wykonaliśmy tylko jedną operację dodawania, w MC wykonujemy ich miliony

Rozważmy przykład bardziej praktyczny: **sumowanie N liczb**

sumowanie identycznych liczb

$$S_1 = \sum_{i=1}^N \frac{1}{N} = 1$$

sumowanie liczb z generatora $U(0,1)$ - średnia daje wartość 0.5

$$S_2 = \frac{1}{N} \sum_{i=1}^N 2x_i, \quad x_i \sim U(0, 1)$$

N	$S_1(S)$	$S_2(S)$	$S_1(D)$	$S_2(D)$
10^2	0.99999934	0.98788345	1.00000000	0.98788378
10^3	0.99999070	1.05094624	1.00000000	1.05094643
10^4	1.00005352	1.00425220	1.00000000	1.00425086
10^5	1.00099015	0.99792010	1.00000000	0.99791809
10^6	1.00903893	1.00016904	1.00000000	1.00017082
10^7	1.06476748	0.99997652	1.00000000	1.00000394
10^8	0.25000000	0.33554432	1.00000000	1.00005262
10^9	0.03125000	0.03355443	0.99999999	1.00000956

S – single precision

D - double precision

po przekroczeniu 10^7 liczb, różnica jest zbyt duża i małe liczby są zerowane

a tu nie ma tego problemu dla $N < 10^{15}$

Losowanie ważne

Mamy oszacować wartość całki wielowymiarowej

$$\langle z \rangle = \int_{\Omega} z(\vec{x}) f(\vec{x}) d^n x \approx \bar{z} = \frac{1}{N} \sum_{i=1}^N z(\vec{x}_i)$$

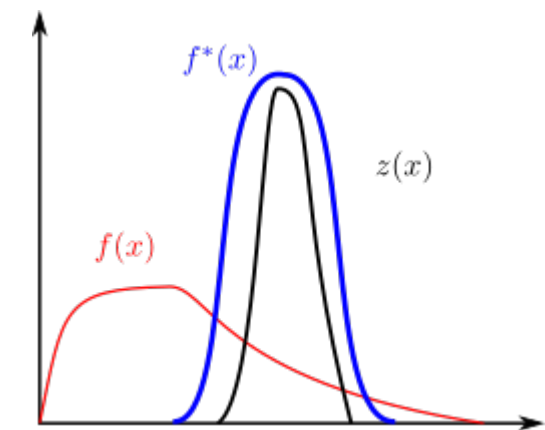
$f(\vec{x})$ - fgp rozkładu wektora losowego x

$$\vec{x}_i \sim \text{Dist}\{f(\vec{x})\}$$

Problem pojawia się, gdy obie funkcje podcałkowe przekrywają się w niewielkim stopniu tj. gdy $z(x)$ jest skupiona w pewnym obszarze

Wówczas dokonujemy modyfikacji fgp – wprowadzamy funkcję próbną $f^*(x)$

$$\langle z \rangle = \int_{\Omega} \frac{z(\vec{x}) f(\vec{x})}{f^*(\vec{x})} f^*(\vec{x}) d^n x$$



ma ona wszystkie cechy funkcji gęstości prawdopodobieństwa

$$f^*(\vec{x}) \geq 0, \quad \int_{\Omega} f^*(\vec{x}) d^n x = 1$$

oraz żądamy aby nowa funkcja podcałkowa była skończona

$$\frac{z(\vec{x}) f(\vec{x})}{f^*(\vec{x})} < \infty$$

Takie podstawienie zmienia sposób liczenia całki, ale nie jej wartość

$$\langle z \rangle = \langle z^* \rangle \approx \bar{z}^* = \frac{1}{N} \sum_{i=1}^N z^*(\vec{x}_i), \quad z^*(\vec{x}) = z(\vec{x}) W(\vec{x}) = z(\vec{x}) \frac{f(\vec{x})}{f^*(\vec{x})}$$

Jak zmiana wpływa na wariancję?

$$\sigma_z^2 = \langle z^2 \rangle - \langle z \rangle^2$$

$$\sigma_{z^*}^2 = \langle z^{*2} \rangle - \langle z^* \rangle^2 = \langle z^{*2} \rangle - \langle z \rangle^2$$

$$\langle z^* \rangle = \langle z \rangle$$

$$W(\vec{x}) = \frac{f(\vec{x})}{f^*(\vec{x})}$$

$$\begin{aligned} \langle z^{*2} \rangle &= \int_{\Omega} z^{*2}(\vec{x}) f^*(\vec{x}) d^n x = \int_{\Omega} z^2(\vec{x}) W^2(\vec{x}) f^*(\vec{x}) d^n x \\ &= \int_{\Omega} z^2(\vec{x}) W(\vec{x}) f(\vec{x}) d^n x \neq \int_{\Omega} z^2(\vec{x}) f(\vec{x}) d^n x = \langle z^2 \rangle \end{aligned}$$

Jeśli tak dobierzemy $f^*(x)$ że zachodzi warunek

$$W(\vec{x}) < 1 \quad \implies \quad \langle z^{*2} \rangle = \int_{\Omega} z^2(\vec{x}) W(\vec{x}) f(\vec{x}) d^n x < \int_{\Omega} z^2(\vec{x}) f(\vec{x}) d^n x$$

to wariancja ulegnie zmniejszeniu

$$\sigma_{z^*}^2 < \sigma_z^2$$

Jaka postać f^* jest optymalna? Znajdziemy ją stosując metodę **mnożników Lagrange'a**

$$\sigma_{z^*}^2 = \int_{\Omega} \frac{z^2(\vec{x}) f^2(\vec{x})}{f^{*2}(\vec{x})} f^*(\vec{x}) d^n x - \langle z \rangle^2$$

Szukamy minimum tego wyrażenia ze względu na f^* , drugi wyraz jest stały a więc nieistotny.
Wartość całki będzie małą jeśli f^* będzie duże, ale musimy pamiętać że $f^*=f_{gp}$ – jest unormowane

Konstruujemy funkcjonal z mnożnikiem Lagrange'a

$$L\{f^*\} = \int_{\Omega} \frac{z^2(\vec{x})f^2(\vec{x})}{f^*(\vec{x})} d^n x + \lambda \int_{\Omega} f^*(\vec{x}) d^n x$$

wariacja funkcjonału doprowadzi nas do minimum

$$\frac{\delta L}{\delta f^*} = -\frac{z^2(\vec{x})f^2(\vec{x})}{f^{*2}(\vec{x})} + \lambda = 0$$

$$f^*(\vec{x}) = \frac{|z(\vec{x})f(\vec{x})|}{\sqrt{\lambda}}$$

i rozważmy przypadek szczególny

$$\sqrt{\lambda} = \langle z \rangle, \quad z(\vec{x}) \geq 0$$

$$f^*(\vec{x}) = \frac{z(\vec{x})f(\vec{x})}{\langle z \rangle} \implies W(\vec{x}) = \frac{f(\vec{x})}{f^*(\vec{x})} = \frac{\langle z \rangle}{z(\vec{x})}$$

$$\langle z^{*2} \rangle = \int_{\Omega} z^2(\vec{x}) W(\vec{x}) f(\vec{x}) d^n x = \langle z \rangle \int_{\Omega} z(\vec{x}) f(\vec{x}) d^n x = \langle z \rangle^2 \implies \sigma_{z^*}^2 = \langle z^{*2} \rangle - \langle z \rangle^2 = 0$$

... ale taki sukces nam nie grozi – nie znamy $\langle z \rangle$ bo do jej wyznaczenia stosujemy MC

Niestety nie istnieje sposób doboru optymalnej postaci f^* - dobieramy ją tak aby dobrze opisywała wartości iloczynu $|z(x)f(x)|$ w obszarach, w których wnoszą duży wkład do całki.

Przykład transformacji funkcji podcałkowej (1)

$$\langle z \rangle = \int_0^1 \cos\left(\frac{\pi x}{2}\right) dx, \quad f(x) = \frac{1}{\Omega}, \quad \Omega = 1$$

$$\sigma_z^2 = \langle z^2 \rangle - \langle z \rangle^2 = \frac{1}{2} - \left(\frac{2}{\pi}\right)^2 = 0.0947152$$

$$\cos\left(\frac{\pi x}{2}\right) = 1 - \frac{\pi^2 x^2}{8} + \frac{\pi^4 x^4}{384} - \dots$$

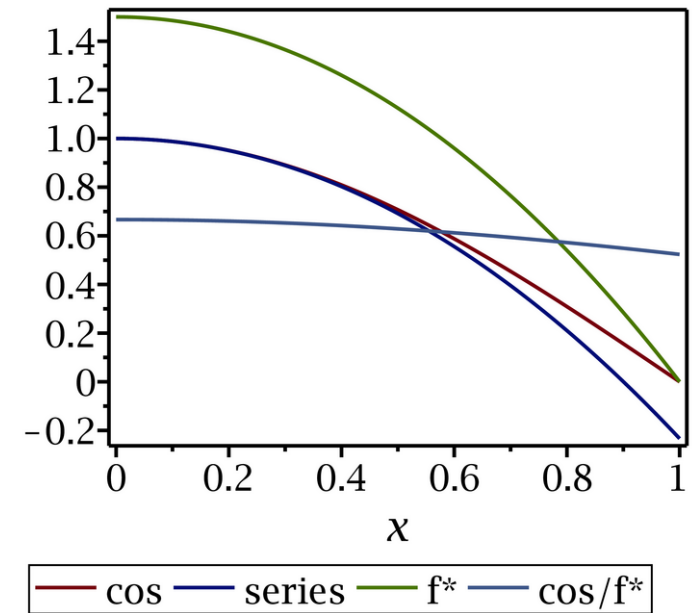
Na podstawie rozwinięcia proponujemy funkcję nieujemną i znormalizowaną

$$f^*(x) = \frac{3}{2}(1 - x^2), \quad \int_0^1 f^*(x) dx = 1$$

$$z^*(z) = \frac{z(x)}{f^*(x)} = \frac{2 \cos\left(\frac{\pi x}{2}\right)}{3(1 - x^2)}$$

$$\sigma_{z^*}^2 = \langle z^{*2} \rangle - \langle z^* \rangle^2 = 0.00099083$$

wskaźnik redukcji wariancji $\frac{\sigma_z^2}{\sigma_{z^*}^2} \approx 96$



Przykład transformacji funkcji podcałkowej (2) - minimalizacja wariancji względem parametru

$$\langle z \rangle = \int_0^1 \sqrt{1-x^2} dx = \frac{\pi}{4}$$

$$\sigma_z^2 = \langle z^2 \rangle - \langle z \rangle^2 = 0.0498$$

$$z(x) = \sqrt{1-x^2} = 1 - \frac{x^2}{2} + \dots$$

rozważamy funkcję f^* z parametrem wariacyjnym β

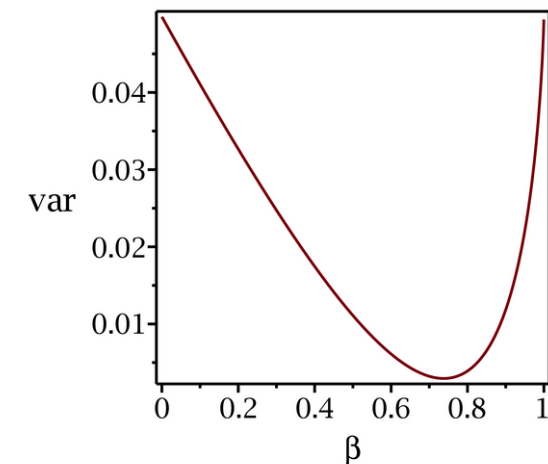
$$f^*(x) = \frac{1 - \beta x^2}{1 - \frac{\beta}{3}}, \quad \int_0^1 f^*(x) dx = 1$$

$$\sigma_{z^*}^2 = -1/16 \pi^2 - 1/3 \frac{\operatorname{arctanh}(\sqrt{\beta}) \beta^3 - 4 \operatorname{arctanh}(\sqrt{\beta}) \beta^2 + \beta^{5/2} - 3 \beta^{3/2} + 3 \operatorname{arctanh}(\sqrt{\beta}) \beta}{\beta^{5/2}}$$

$$\beta = 0.74 \quad \Rightarrow \quad \min \sigma_{z^*}^2 = 0.00294$$

wskaźnik redukcji wariancji $\frac{\sigma_z^2}{\sigma_{z^*}^2} = 16.9$

- rachunki wykonaliśmy analitycznie
- w praktyce konieczne byłoby wykonanie serii testów dla niewielkiego N i różnych β , z uzyskanego ciągu wybieramy optymalną wartość β



Losowanie ważne - całkowanie funkcji z osobliwością

Jeśli iloczyn funkcji podcałkowych $z(x)f(x)$ ma osobliwość wówczas wariancja może być nieograniczona.
Jak wówczas modyfikować metodę?

Szukamy takiej funkcji f^* aby spełniony był warunek

$$z^*(x) = \frac{z(x)f(x)}{f^*(x)} < \infty$$

Przykład

$$\langle z \rangle = \int_0^1 \frac{1}{x^{1/2}} dx \quad \langle z^2 \rangle = \int_0^1 \frac{1}{x} dx = \infty \quad \text{- wariancja nie istnieje}$$

Rozważamy funkcję f^* , która pozwoli określić wariancję – musimy zmodyfikować wykładnik x -a w mianowniku

$$f^*(x) = (1 - \beta)x^{-\beta} \quad \Rightarrow \quad \int_0^1 f^*(x) dx = 1 - \lim_{x \rightarrow 0^+} x^{1-\beta} \quad \Rightarrow \quad \beta < 1$$

$$z^*(x) = \frac{x^{\beta - \frac{1}{2}}}{1 - \beta} \quad \langle z^{*n} \rangle = (1 - \beta)^{-n+1} \int_0^1 x^{n\beta - \frac{n}{2}} x^{-\beta} dx$$

Aby całka istniała potęga x -a musi być większa od -1

$$(n - 1)\beta - \frac{n}{2} > -1$$

Istnienie skończonych momentów (i wariancji) zapewnia warunek

$$\frac{1}{2} \leq \beta < 1$$

Przykład 2 – całka z osobliwością

Należy oszacować metodą MC wartość całki

$$\langle z \rangle = \int_0^1 \frac{1}{\sqrt{x(1-x)}} dx = 2$$

$$\langle z^2 \rangle = \int_0^1 \frac{1}{x(1-x)} dx = \infty$$

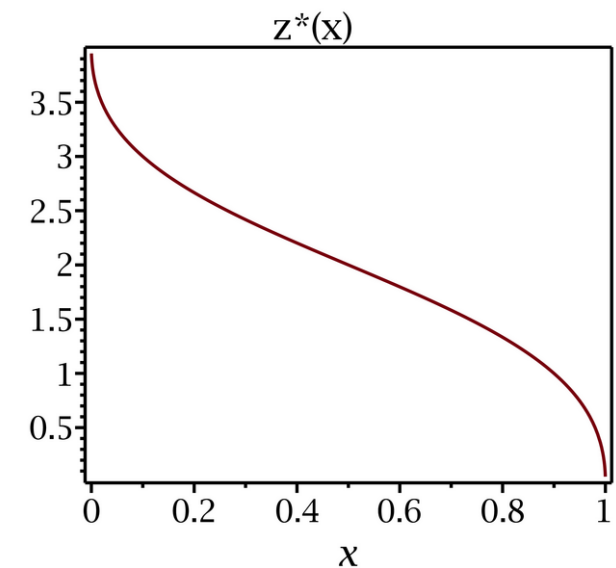
- funkcja ma dwie osobliwości w $x=0$ i w $x=1$
- funkcję f^* proponujemy w takiej postaci, aby miała też osobliwości w tych punktach

$$f^*(x) = \frac{1}{4\sqrt{x}} + \frac{1}{4\sqrt{1-x}}, \quad \int_0^1 f^*(x) dx = 1$$

zmodyfikowana funkcja podcałkowa nie będzie posiadać osobliwości – te usuwa f^*

$$z^*(x) = \frac{z(x)}{f^*(x)} = \frac{4}{\sqrt{x} + \sqrt{1-x}}$$

$$\langle z^{*2} \rangle = \int_0^1 z^{*2}(x) f^*(x) dx = 4.98$$



Przykład 3 – całka z osobliwością

Należy obliczyć całkę w trzech wymiarach, np.:

- wyznaczanie sygnału docierającego do detektora w transporcie promieniowania niejonizującego

$$\langle z \rangle = \int_V s(\vec{r}) \frac{\exp(-\mu|\vec{r} - \vec{r}_0|)}{|\vec{r} - \vec{r}_0|^2} d^3r$$

- $s(r)$ opisuje rozkład źródeł/gęstości ładunku
- μ opisuje wpływ absorpcji i rozpraszania (w elektrostatyce - ekranowanie)

Rozkład źródeł jest unormowany

$$\int_V s(\vec{r}) d^3r = 1 \quad - \text{traktujemy jako fgp}$$

identyfikujemy funkcję podcałkową

$$z(\vec{r}) = \frac{\exp(-\mu|\vec{r} - \vec{r}_0|)}{|\vec{r} - \vec{r}_0|^2}$$

- ze względu na postać mianownika wariancja jest nieskończona

Dokonujemy zamiany zmiennych ($\vec{r} \rightarrow \vec{r} + \vec{r}_0$) i szacujemy wartość 2 momentu całkując po kątach θ i ϕ

$$\langle z \rangle = \int_V s(\vec{r} + \vec{r}_0) \frac{\exp(-\mu|\vec{r}|)}{|\vec{r}|^2} d^3r$$

$$\begin{aligned} \langle z^2 \rangle &= \int_0^{2\pi} d\phi \int_0^\pi d\theta \int_0^{r_{max}} dr s(\vec{r} + \vec{r}_0) \frac{\exp(-2\mu r)}{r^4} r^2 = 4\pi \int_0^{r_{max}} s(\vec{r} + \vec{r}_0) \frac{\exp(-2\mu r)}{r^2} dr \\ &\geq 4\pi s_0 \underbrace{\int_0^{r_{max}} \frac{\exp(-2\mu r)}{r^2} dr}_{=\infty}, \quad 0 < s_0 \leq s(\vec{r}) \end{aligned}$$

- standardowe podejście MC do liczenia całki nie pozwoli oszacować błędu

Problemem jest wyraz $1/r^2$ i musimy się go pozbyć. Wprowadzamy pomocniczą funkcję f^*

$$f^*(\vec{r}) = \frac{1}{4\pi} \frac{\mu}{(1 - \exp[-\mu R_a])} \frac{\exp(-\mu r)}{r^2}$$

$$\int_0^{2\pi} d\phi \int_0^\pi d\theta \int_0^{R_a} dr f^*(r) r^2 \sin\theta = 1$$

- R_a to minimalny promień sfery zawierającej obszar V
- f^* jest funkcją radialną i unormowaną w 3D

$$f^*(\vec{r}) d^3r = f^*(r) r^2 \sin\theta d\phi d\theta dr = \underbrace{\frac{\mu \exp(-\mu r)}{1 - \exp(-\mu R_a)}}_{f_r(r)} \underbrace{\frac{\sin\theta}{2}}_{f_\theta(\theta)} \underbrace{\frac{1}{2\pi}}_{f_\phi(\phi)} d\phi = [f_r(r) dr] \cdot [f_\theta(\theta) d\theta] \cdot [f_\phi(\phi) d\phi]$$

- fgp wyraziliśmy jako iloczyn trzech niezależnych rozkładów
- jak wygenerować losowe wartości dla $\sin(\theta)$ oraz jednorodny dla ϕ wiemy
- losowanie zmiennej o rozkładzie eksponencjalnym może być wykonane na 2 sposoby

$$\begin{cases} R_a = \infty, & r = -\frac{1}{\mu} \ln U_1, & U_1 \sim U(0, 1) \\ R_a < \infty, & r = -\frac{1}{\mu} \ln [1 - U_1(1 - e^{-\mu R_a})], & U_1 \sim U(0, 1) \end{cases}$$

Po zmianie sposobu liczenia całki ma ona postać

$$\langle z \rangle = \int_V \frac{4\pi[1 - \exp(-\mu R_a)]}{\mu} s(\vec{r} + \vec{r}_0) f^*(\vec{r}) d^3r$$

oraz 2 moment

$$\langle z^2 \rangle = \left(\frac{4\pi[1 - \exp(-\mu R_a)]}{\mu} \right)^2 \int_V s^2(\vec{r} + \vec{r}_0) f^*(\vec{r}) d^3r < \infty$$

zatem wariancja staje się skończona i możemy jej użyć do szacowania niepewności szacowanej całki

Metoda losowania systematycznego

Redukcję wariancji możemy uzyskać rozbijając obszar całkowania na M podobszarów i wykonując całkowanie w tych podobszarach

$$V = \sum_{m=1}^M V_m$$

wylosowanie wektora x w każdym podobszarze ma określone prawdopodobieństwo

$$p_m = \frac{\int_{V_m} f(\vec{x}) d^n x}{\int_V f(\vec{x}) d^n x} = \int_{V_m} f(\vec{x}) d^n x$$

dla każdego obszaru definiujemy **unormowaną** fgp

$$f_m(\vec{x}) = \begin{cases} \frac{f(\vec{x})}{p_m} & \vec{x} \in V_m \\ 0 & \vec{x} \notin V_m \end{cases} \quad \int_{V_m} f_m(\vec{x}) d^n x = 1$$

Wykorzystajmy nowe fgp do obliczenia wartości oczekiwanej

$$\langle z \rangle = \int_V z(\vec{x}) f(\vec{x}) d^n x = \sum_{m=1}^M p_m \int_{V_m} z(\vec{x}) f_m(\vec{x}) d^n x$$

przyjmijmy oznaczenie

$$\langle z \rangle_m = \int_{V_m} z(\vec{x}) f_m(\vec{x}) d^n x$$

$$\langle z \rangle = \sum_{m=1}^M p_m \langle z \rangle_m$$

- **ważona suma wartości oczekiwanych w obszarach**
- **wagi to prawdopodobieństwa wylosowania wektora x z danego obszaru**

Wariancja zmiennej losowej $z(\vec{x})$ - definicja

$$\sigma_z^2 = \int_V [z(\vec{x}) - \langle z \rangle]^2 f(\vec{x}) d^n x = \int_V [z^2(\vec{x}) - 2z(\vec{x})\langle z \rangle + \langle z \rangle^2] f(\vec{x}) d^n x$$

i z rozbiem całki na sumę

$$\sigma_z^2 = \sum_{m=1}^M p_m \int_{V_m} [z^2(\vec{x}) - 2z(\vec{x})\langle z \rangle + \langle z \rangle^2] f_m(\vec{x}) d^n x$$

oznaczymy k-ty moment liczony w podobszarze (n-to liczba wymiarów)

$$\langle z^k \rangle_m = \int_{V_m} z^k(\vec{x}) f_m(\vec{x}) d^n x$$

$$\sigma_z^2 = \sum_{m=1}^M p_m (\langle z^2 \rangle_m - 2\langle z \rangle \langle z \rangle_m + \langle z \rangle^2)$$

- $\langle z \rangle$ liczone w obszarze V
- $\langle z \rangle_m$ liczone w podobszarze V_m

Przechodzimy do losowania, określamy całkowitą liczbę losowań N .
Wówczas w każdym z podprzedziałów wylosujemy N_m liczb

$$N_m = p_m \cdot N$$

i liczymy (globalną) średnią jako estymator wartości oczekiwanej

$$\langle z \rangle \approx \bar{z} = \sum_{m=1}^M p_m \bar{z}_m$$

średnia dla podobszaru

$$\langle z \rangle_m \approx \bar{z}_m = \frac{1}{N_m} \sum_{i_m=1}^{N_m} z(\vec{x}_{i_m}), \quad \vec{x}_{i_m} \sim \text{Dist}\{f_m(\vec{x})\}$$

Pamiętajmy że średnia to też zmienna losowa, policzmy jej wariancję

$$\sigma_{\bar{z}}^2 = \langle (\bar{z} - \langle z \rangle)^2 \rangle$$

$$\begin{aligned} \sigma_{\bar{z}}^2 &= \left\langle \left(\sum_m p_m \bar{z}_m - \sum_m p_m \langle z \rangle_m \right)^2 \right\rangle = \left\langle \left(\sum_m p_m (\bar{z}_m - \langle z \rangle_m) \right)^2 \right\rangle \\ &= \left\langle \sum_m p_m^2 (\bar{z}_m - \langle z \rangle_m)^2 \right\rangle + \sum_m \sum_{m' \neq m} p_m p_{m'} \underbrace{\langle \bar{z}_m - \langle z \rangle_m \rangle}_{\approx 0} \underbrace{\langle \bar{z}_{m'} - \langle z \rangle_{m'} \rangle}_{\approx 0} \\ &= \left\langle \sum_m p_m^2 \frac{1}{N_m^2} \sum_{i_m} (z_{i_m} - \langle z \rangle_m)^2 \right\rangle + \sum_m p_m^2 \frac{1}{N_m^2} \sum_{i_m} \sum_{i_{m'}} \underbrace{\langle z_{i_m} - \langle z \rangle_m \rangle}_{\approx 0} \underbrace{\langle z_{i_{m'}} - \langle z \rangle_{m'} \rangle}_{\approx 0} \\ &= \sum_m p_m^2 \frac{1}{N_m^2} \sum_{i_m} \langle (z_{i_m} - \langle z \rangle_m)^2 \rangle \end{aligned}$$

$$\sigma_{\bar{z}}^2 = \sum_{m=1}^M \frac{p_m^2}{N_m^2} \sum_{i_m=1}^{N_m} \langle (z(\vec{x}_{i_m}) - \langle z \rangle_m)^2 \rangle = \sum_{m=1}^M \frac{p_m^2}{N_m^2} \sum_{i_m=1}^{N_m} \sigma_m^2(z) = \sum_{m=1}^M \frac{p_m^2}{N_m} \sigma_m^2(z)$$

wykorzystajmy jeszcze związek pomiędzy p_m a N_m

$$p_m = \frac{N_m}{N}$$

wariancja
wartości
średniej



$$\sigma_{\bar{z}}^2 = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z)$$

skorzystamy z relacji

$$\bar{z} = \sum_{m=1}^M p_m \bar{z}_m$$

$$\bar{z}_m = \frac{1}{N_m} \sum_{i_m=1}^{N_m} z(\vec{x}_{i_m})$$

$$z(\vec{x}_{i_m}) = z_{i_m}$$

$$\begin{aligned} \sigma_m^2(z) &= \int_{V_m} [z(\vec{x}) - \langle z \rangle]^2 f(\vec{x}) d^n x \approx \bar{z}_m^2 - \bar{z}_m^2 \\ &\approx \frac{1}{N_m} \sum_{i_m=1}^M z^2(\vec{x}_{i_m}) - \left(\frac{1}{N_m} \sum_{i_m=1}^M z(\vec{x}_{i_m}) \right)^2 \end{aligned}$$

Policzmy wariancję dla metody podstawowej i porównajmy ją z wariancją losowania systematycznego.

wariancja wartości średniej
w losowaniu systematycznym

$$\langle z \rangle = \int_V z(\vec{x}) f(\vec{x}) d^n x \approx \bar{z}_p = \frac{1}{N} \sum_i^N z(\vec{x}_i)$$

$$\sigma_{\bar{z}}^2 = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z)$$

wariancja metody podstawowej

$$\begin{aligned} \sigma_{\bar{z}_p}^2 &= \frac{\sigma_z^2}{N} = \frac{1}{N} \langle [z(\vec{x}) - \langle z \rangle]^2 \rangle = \frac{1}{N} \int_V [z(\vec{x}) - \langle z \rangle]^2 f(\vec{x}) d^n x \\ &= \frac{1}{N} \sum_{m=1}^M p_m \int_{V_m} [z(\vec{x}) - \langle z \rangle]^2 f_m(\vec{x}) d^n x \\ &= \frac{1}{N} \sum_{m=1}^M p_m \int_{V_m} [(z(\vec{x}) - \langle z \rangle_m) + (\langle z \rangle_m - \langle z \rangle)]^2 f_m(\vec{x}) d^n x \\ &= \frac{1}{N} \sum_{m=1}^M p_m \left\{ \underbrace{\int_{V_m} [z(\vec{x}) - \langle z \rangle_m]^2 f_m(\vec{x}) d^n x}_{=\sigma_m^2(z)} + 2 \underbrace{\int_{V_m} [z(\vec{x}) - \langle z \rangle_m][\langle z \rangle_m - \langle z \rangle] f_m(\vec{x}) d^n x}_{\approx 0} \right. \\ &\quad \left. + \underbrace{\int_{V_m} [\langle z \rangle_m - \langle z \rangle]^2 f_m(\vec{x}) d^n x}_{>0} \right\} \end{aligned}$$

$$\sigma_{\bar{z}_p}^2 = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z) + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2$$

Wariancja średniej w metodzie podstawowej

$$\begin{aligned}\sigma_{\bar{z}_p}^2 &= \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z) + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 \\ &= \sigma_{\bar{z}}^2 + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 \\ &\geq \sigma_{\bar{z}}^2\end{aligned}$$

- drugi wyraz jest nieujemny więc spodziewamy się, że redukcja będzie tym większa im bardziej średnie w obszarach będą się różnić od średniej globalnej
- zerowanie drugiego wyrazu wystąpi w przypadku rozkładu jednorodnego, ale nie tylko - oscylacje przestrzenne funkcji $z(x)$ też mogą prowadzić do jego zerowania

wariancja wartości średniej
w losowaniu systematycznym

$$\sigma_{\bar{z}}^2 = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z)$$

Losowanie warstwowe

Jest to modyfikacja metody losowania systematycznego. Teraz nie zakładamy z góry wartości N_m , ale dobieramy je tak aby zminimalizować wartości wariancji w podobszarach

Punktem startowym są wyrażenia z losowania systematycznego

Wprowadzamy warunek normalizacji liczby losowań ($N=\text{const}$)

$$\sum_{m=1}^M N_m = N$$

i poszukujemy optymalnych N_m stosując metodę mnożników Lagrange'a

$$L = \sum_{m=1}^M \frac{p_m^2}{N_m} \sigma_m^2(z) + \lambda \sum_{m=1}^M N_m$$

- minimalizacja 1 wyrazu jest prosta – wymaga zwiększania N_m , ale drugi wyraz ogranicza dowolność N_m (normalizacja do N)

Liczymy pochodną wariacyjną i przyrównujemy do 0

$$\frac{\delta L}{\delta N_k} = 0 \quad \rightarrow \quad \frac{\delta}{\delta N_k} \sum_{m=1}^M \left[\frac{p_m^2 \sigma_m^2}{N_m} + \lambda N_m \right] = -\frac{p_k^2 \sigma_k^2}{N_k^2} + \lambda = 0$$

wyznaczamy optymalną liczbę losowań

$$N_k = \frac{p_k \sigma_k}{\sqrt{\lambda}}$$

korzystamy z warunku normalizacji N

$$N = \sum_{m=1}^M N_m = \sum_{m=1}^M \frac{p_m \sigma_m}{\sqrt{\lambda}} \quad \Rightarrow \quad \sqrt{\lambda} = \frac{1}{N} \sum_{m=1}^M p_m \sigma_m = \frac{\langle \sigma_m \rangle}{N}$$

$$\begin{aligned} \langle z \rangle &= \int_V z(\vec{x}) f(\vec{x}) d^n x = \sum_{m=1}^M \int_{V_m} z(\vec{x}) f(\vec{x}) d^n x \\ \bar{z}_w &= \sum_{m=1}^M p_m \bar{z}_m = \sum_{m=1}^M \frac{p_m}{N_m} \sum_{i_m=1}^{N_m} z(\vec{x}_{i_m}) \\ \sigma_{\bar{z}_w}^2 &= \sum_{m=1}^M \frac{p_m^2}{N_m} \sigma_m^2(z) \end{aligned}$$

znaleźliśmy wartość mnożnika Lagrange'a

$$\sqrt{\lambda} = \frac{\langle \sigma_m \rangle}{N}$$

dla której optymalna liczba losowań wynosi

$$N_m = \frac{p_m \sigma_m}{\langle \sigma_m \rangle} N = \frac{p_m \sigma_m(z)}{\sum_{m=1}^M p_m \sigma_m(z)} N$$

losowanie warstwowe

$$N_m = \frac{\sigma_m(z)}{\sum_{m=1}^M p_m \sigma_m(z)} p_m N$$

losowanie systematyczne

$$N_m = p_m N$$

- liczba losowań jest teraz zależna od dwóch czynników: wariancji podobszaru m-tego oraz średniej wariancji obszarów
- w obszarach o dużej wariancji σ_m wykonujemy więcej losowań niż wynika to tylko z wartości p_m
- obu wielkości nie znamy
- w praktyce wykonuje się serię krótkich testów w celu określenia wartości σ_m i użycia ich następnie do wyznaczenia N_m w docelowych obliczeniach MC
- przy większej liczbie wymiarów $n > 4$ podział V na podobszary V_m staje się problematyczny, liczba podobszarów szybko rośnie, dla k segmentów/wymiar $M = k^n$: $M = 5^4 = 625$ lub $M = 10^4$
- podział całego obszaru nie zawsze jest potrzebny: w przypadku wielowymiarowym identyfikujemy zmienną wnoszącą największy wkład do wariancji (zazwyczaj wynika to z silnych oscylacji) i dokonujemy podziału tylko w jednym wymiarze

Porównanie z MC podstawową (wzór na wariancję kilka slajdów wcześniej)

$$\begin{aligned}
 \sigma_{\bar{z}_p}^2 - \sigma_{\bar{z}_w}^2 &= \frac{1}{N} \sum_{m=1}^M p_m \sigma_m^2(z) + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 - \frac{1}{N} \langle \sigma_m \rangle^2 \\
 &= \frac{1}{N} [\langle \sigma_m^2 \rangle - \langle \sigma_m \rangle^2] + \frac{1}{N} \sum_{m=1}^M p_m [\langle z \rangle_m - \langle z \rangle]^2 \\
 &= \frac{1}{N} \left[\sum_{m=1}^M p_m (\sigma_m - \langle \sigma_m \rangle)^2 + \sum_{m=1}^M p_m (\langle z \rangle_m - \langle z \rangle)^2 \right] > 0
 \end{aligned}$$

- poza przypadkami szczególnymi, metoda losowania warstwowego prowadzi do obniżenia wartości wariancji

Przykład. całkowanie funkcji $f(x)$ metodami: podstawową, losowanie systematyczne, warstwowe

$$g(x) = 1 + \tanh(x), \quad x \in [-3, 3]$$

$$C = \int_{-3}^3 g(x) dx = 6$$

Obszar całkowania dzielimy na $M=10$ podobszarów, jako $f(x)$ wybieramy rozkład jednorodny.

$$f(x) = \frac{1}{b-a} = \text{const} \quad \Rightarrow \quad C = \int_a^b \underbrace{(b-a)g(x)}_{z(x)} f(x) dx = \int_a^b z(x) f(x) dx$$

metoda podstawowa:

$$U_i \sim U(0, 1)$$

$$X_i = a + (b-a)U_i \quad \Rightarrow \quad X_i \in [a, b]$$

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N z(X_i)$$

$$\overline{z^2} = \frac{1}{N} \sum_{i=1}^N z^2(X_i)$$

$$\sigma_{\bar{z}} = \sqrt{\frac{\overline{z^2} - (\bar{z})^2}{N}}$$

metoda losowania systematycznego

$$U_i \sim U(0, 1)$$

$$m = 1, 2, \dots, M$$

$$\Delta x = \frac{b-a}{M} \quad X_{i_m} \sim \text{Dist}\{f_m(x)\}$$

$$X_{i_m} = a + \Delta x \cdot (m-1) + U_i \cdot \Delta x$$

$$p_m = \frac{1}{M}$$

$$N_m = p_m M = \text{const}$$

$$\bar{z}_m^n = \frac{1}{N_m} \sum_{i_m=1}^{N_m} z^n(X_{i_m})$$

$$\bar{z} = \sum_{m=1}^M p_m \bar{z}_m$$

$$\sigma_m^2 = \overline{z_m^2} - \bar{z}_m^2$$

$$\sigma_{\bar{z}} = \sqrt{\frac{\sum_{m=1}^M p_m \sigma_m^2}{N}}$$

metoda losowania warstwowego

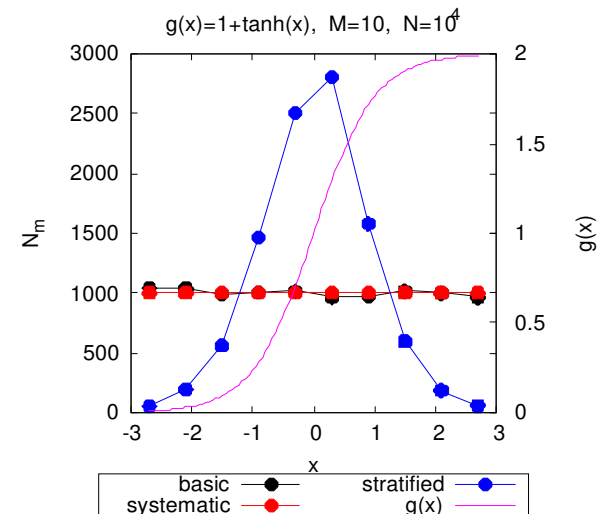
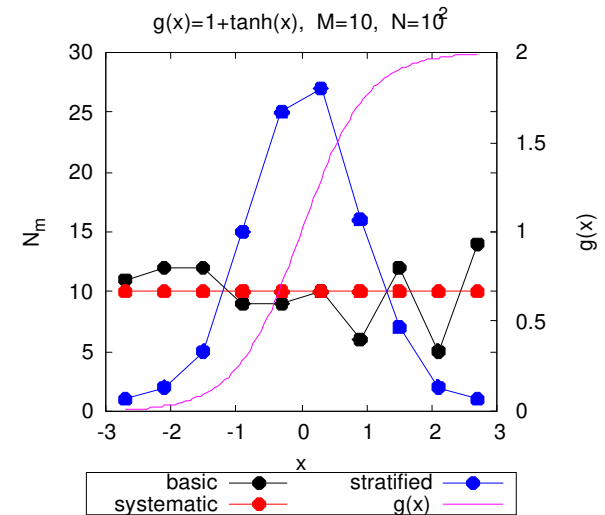
- pokazana tylko różnica
- resztę liczymy identycznie
- sigmy wyznaczamy nowe dla aktualnego rozkładu

$$N_m = \frac{p_m \hat{\sigma}_m(z)}{\sum_{m=1}^M p_m \hat{\sigma}_m} N, \quad \hat{\sigma}_m - \text{z losowania zwykłego}$$

Wyniki symulacji MC

metoda podstawowa			
N	C	$\sigma_{\bar{z}}$	$R = \frac{\sigma_{\bar{z}}}{\bar{z}} \cdot 100$
10^2	5.332	0.488	9.16
10^3	5.931	0.154	2.61
10^4	5.919	0.049	0.829
10^5	6.014	0.0155	0.258
losowanie systematyczne			
N	C	$\sigma_{\bar{z}}$	$R = \frac{\sigma_{\bar{z}}}{\bar{z}} \cdot 100$
10^2	5.950	0.051	0.850
10^3	6.010	0.015	0.249
10^4	5.993	0.0048	0.080
10^5	5.998	0.0015	0.026
losowanie warstwowe (optymalne)			
N	C	$\sigma_{\bar{z}}$	$R = \frac{\sigma_{\bar{z}}}{\bar{z}} \cdot 100$
10^2	6.040	0.031	0.523
10^3	6.000	0.011	0.183
10^4	5.992	0.0034	0.057
10^5	6.002	0.0011	0.018

- gęstość próbkowania w losowaniu systematycznym i standardowym jest podobna dla dużego N
- losowanie systematyczne daje ~10 razy mniejszy błąd względny
- w losowaniu warstwowym częściej próbkowany jest obszar o dużej zmienności funkcji podcałkowej



Metoda zmiennej kontrolnej

Wariancja zależy od fluktuacji zmiennej losowej, którą jest funkcja podcałkowa.

Fluktuacje zależą więc od wartości (amplitudy) funkcji podcałkowej.

Redukcję wariancji można uzyskać zmniejszając amplitudę poprzez odjęcie innej funkcji

$$\begin{aligned}\langle g \rangle &= \int g(x) f(x) dx = \int [g(x) + \underbrace{h(x) - h(x)}_{=0}] f(x) dx \\ &= \underbrace{\int [g(x) - h(x)] f(x) dx}_{\substack{=z(x) \\ \langle z \rangle}} + \underbrace{\int h(x) f(x) dx}_{C_2} \\ &= \langle z \rangle + C_2\end{aligned}$$

Warunek redukcji wymaga jednak dokładnej znajomości C_2 – ta całka powinna być liczona analitycznie.

Nowa zmienna losowa $z(x)$ ze względu na redukcję amplitudy powinna mieć mniejszą wariancję niż $g(x)$

$$\sigma_{\bar{g}}^2 > \sigma_{\bar{z}}^2$$

Oczekujemy że tak się stanie, gdy funkcja $h(x)$ będzie zbliżona do $g(x)$.

Generalnie zaleca się następujące postępowanie:

- jeśli $|g(x) - h(x)| \sim \text{const}$ to metoda zmiennej kontrolnej znacząco redukuje wariancję
- jeśli $|g(x) - h(x)| \sim |h(x)|$ wówczas metoda losowania ważonego będzie skuteczniejsza

Przykład zastosowania metody zmiennej kontrolnej

Należy oszacować wartość całki

$$\langle g \rangle = \int_1^2 \ln(x) dx = 2 \ln(2) - 1 = 0.386294$$

$$\langle g^2 \rangle = 0.1883 \quad \sigma_g = \langle g^2 \rangle - \langle g \rangle^2 = 0.03909$$

Wybieramy funkcję $h(x)=x$

$$\langle z \rangle = \int_1^2 [\ln(x) - x] dx = -1.1137$$

$$\langle z^2 \rangle = 1.24906$$

$$\sigma_z = \langle z^2 \rangle - \langle z \rangle^2 = 0.00872$$

krotność redukcji wariancji $\frac{\sigma_g}{\sigma_z} = 4.48$

Wyberzmy inną funkcję $h(x)=x-1$

$$\langle z \rangle = \int_1^2 [\ln(x) - (x - 1)] dx = -0.1137$$

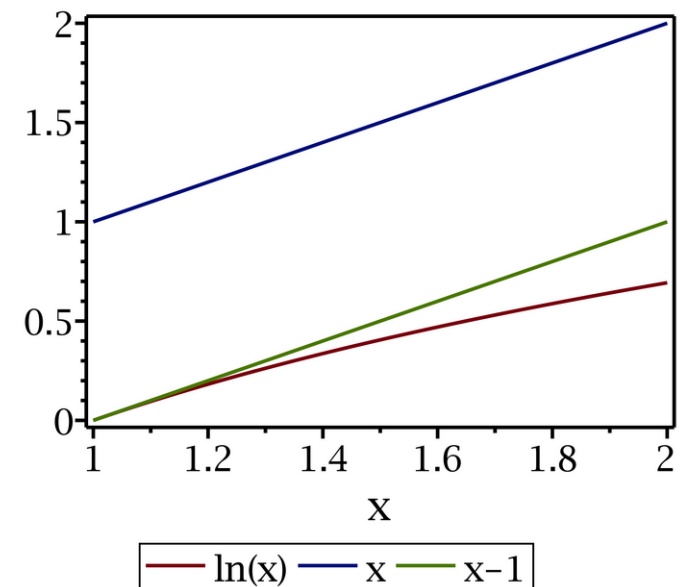
$$\langle z^2 \rangle = 0.02165$$

wynik nie uległ zmianie

$$\sigma_z = \langle z^2 \rangle - \langle z \rangle^2 = 0.00872$$

$$\frac{\sigma_g}{\sigma_z} = 4.48$$

- stopień redukcji wariancji zależy od lokalnych zmian wartości $|g(x)-h(x)|$ a dokładniej: jak dobrze $h(x)$ aproksymuje $g(x)$
- wartość $|g(x)-h(x)|$ jest mniej istotna (patrz rysunek)



Metoda zmiennych antytetycznych

Szacujemy wartość całki z jednorodnym rozkładem fgp zmiennej x

$$\langle g \rangle = \int_0^1 g(x) dx \approx \bar{g}_1 = \frac{1}{N} \sum_{i=1}^N g(X_i), \quad X_i \sim U(0, 1)$$

Ponieważ $f(x)=\text{const}$ więc symetryczna zmiana argumentu funkcji podcałkowej nie wpłynie na wynik

$$\langle g \rangle = \int_0^1 g(1-x) dx \approx \bar{g}_2 = \frac{1}{N} \sum_{i=1}^N g(1-X_i), \quad X_i \sim U(0, 1)$$

Całkę możemy obliczyć korzystając z obu wyników

$$\bar{g} = \frac{\bar{g}_1 + \bar{g}_2}{2} = \frac{1}{N} \sum_{i=1}^{N/2} [g(X_i) + g(1-X_i)] = \frac{1}{N} \sum_{i=1}^{N/2} [g_{1,i} + g_{2,i}]$$

Wartości g_1 i g_2 możemy liczyć jednocześnie korzystając z tego samego ciągu liczb pseudolosowych, ale to oznacza że te wartości mogą być skorelowane.

Para liczb losowych (g_1, g_2) stanowi parę antytetyczną, gdy obie zmienne mają identyczny rozkład, są skorelowane a współczynnik korelacji jest ujemny $\text{cov}(g_1, g_2) < 0$.

Policzmy wariancję całki (wykład pierwszy)

$$\begin{aligned} \sigma_{\bar{g}}^2 &= \frac{1}{N} \left(\frac{1}{N} \left[\sum_{i=1}^{N/2} (g_{1,i} - \bar{g})^2 + \sum_{i=1}^{N/2} (g_{2,i} - \bar{g})^2 + 2 \sum_{i=1}^{N/2} (g_{1,i} - \bar{g})(g_{2,i} - \bar{g}) \right] \right) \\ &= \frac{\frac{N}{2}}{N^2} [\sigma_{g_1}^2 + \sigma_{g_2}^2 + 2\text{cov}(g_1, g_2)] \\ &= \frac{\sigma_{g_k}^2}{N} [1 + \rho_{g_1, g_2}] \end{aligned}$$

$\sigma_{g_1}^2 = \sigma_{g_2}^2$ - wariancja pojedynczej zmiennej losowej

Wariancja wyniku gdy używane są zmienne antytetyczne

$$\sigma_{\bar{g}}^2 = \frac{\sigma_{g_k}^2}{N} [1 + \rho_{g_1, g_2}] \quad \rho_{g_1, g_2} = \frac{\text{cov}(g_1, g_2)}{\sigma_{g_k}^2}, \quad k = 1, 2$$

zatem wymagamy spełnienia warunku

$$\rho_{g_1, g_2} < 0$$

W ogólnym przypadku fgp nie będzie rozkładem jednorodnym

$$\langle g \rangle = \int_{-\infty}^{\infty} g(x) f(x) dx$$

wówczas nadal możemy wykorzystać metodę zmiennych antytetycznych

$$\bar{g} = \frac{1}{N} \sum_{i=1}^{N/2} [g(x_i) + g(y_i)]$$

ale zmienne X i Y generowane są przy użyciu tego samego ciągu liczb o rozkładzie jednorodnym
np. w metodzie odwracania dystrubuanty mamy

$$F(x) = \int_{-\infty}^x f(x') dx' = U_i, \quad U_i \sim U(0, 1)$$

$$X_i = F^{-1}(U_i), \quad Y_i = F^{-1}(1 - U_i)$$

- oczywiście dla różnych rozkładów stopień korelacji zmiennych będzie inny
- można też wybrać inny sposób korelacji zmiennych losowych

Przykład: całkowanie metodą zmiennych antytetycznych

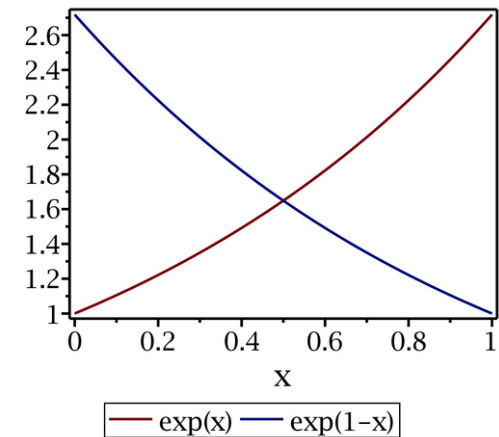
Należy oszacować wartość całki

$$\langle g \rangle = \int_0^1 g(x) dx = \int_0^1 e^x dx$$

wariancja zmiennej losowej g

$$\sigma_g^2 = \langle g^2 \rangle - \langle g \rangle^2 = 0.24203$$

całkowanie wykonujemy dla dwóch funkcji: pierwotnej i ze zmienionym argumentem (obie są symetryczne względem punktu $x=0.5$ - rysunek)



$$\langle g \rangle = \int_0^1 [g_1(x) + g_2(x)] dx = \int_0^1 \left[\frac{e^x}{2} + \frac{e^{1-x}}{2} \right] dx$$

wariancja liczona dla zmiennych antytetycznych

$$\sigma_{12}^2 = \sigma_{g_1}^2 + \sigma_{g_2}^2 + 2cov(g_1, g_2) = 0.003912$$

$$\sigma_{g_1}^2 = \sigma_{g_2}^2 = \frac{\sigma_g^2}{4} = 0.0605$$

$$cov(g_1, g_2) = -0.05855$$

współczynnik redukcji wariancji

$$\frac{\sigma_g^2}{\sigma_{12}^2} = 61.9$$

Przykład: całkowanie metodą zmiennych antytetycznych + metoda losowania ważonego

Modyfikujemy poprzedni wynik wprowadzając nową fgp pod całkę - $f^*(x)$

$$\langle g^* \rangle = \int_0^1 \underbrace{\left(\frac{g_1(x) + g_2(x)}{f^*(x)} \right)}_{=g^*(x)} f^*(x) dx$$

jako $f^*(x)$ weźmy ucięte rozwinięcie funkcji podcałkowej w szereg Taylora w punkcie $x=1/2$

$$g_1(x) + g_2(x) = \frac{e^x + e^{1-x}}{2} \approx 1.8548 - 0.82436x + 0.82436x^2 + \dots$$

po znormalizowaniu $f^*(x)$ ma ona postać

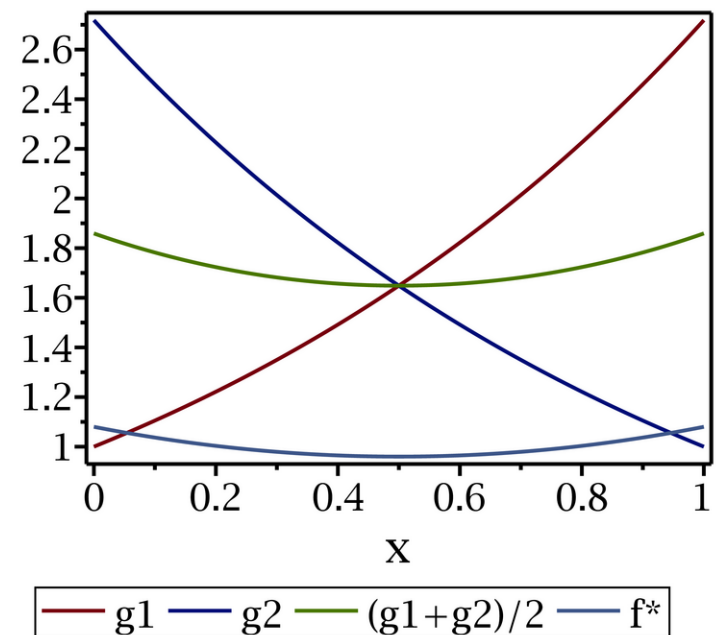
$$f^*(x) = 1.08 - 0.48x + 0.48x^2$$

wariancja g^* ma wartość

$$\sigma_{g^*}^2 = 0.000001218$$

współczynnik redukcji wariancji

$$\frac{\sigma_g^2}{\sigma_{g^*}^2} = 1.98 \cdot 10^5$$



Metoda redukcji liczby wymiarów

Szacujemy wartość całki wielowymiarowej

$$\langle z \rangle = \int_V z(\vec{x}) f(\vec{x}) dV \quad \vec{x} = [x_1, x_2, \dots, x_n]$$

$$dV = dx_1 dx_2 \dots dx_n$$

dokonujemy rozdziału zmiennych całkowania

$$\langle z \rangle = \int_{V_1} \int_{V_2} z(\vec{u}, \vec{v}) f(\vec{u}, \vec{v}) dV_1 dV_2$$

Podziału obszaru całkowania dokonujemy w taki sposób, aby możliwe było wykonanie całkowania fgp dla obszaru V_2

$$f_u(\vec{u}) = \int_{V_2} f(\vec{u}, \vec{v}) dV_2$$

funkcji $z(\mathbf{x})$ w tym obszarze z wykorzystaniem rozkładu warunkowego $f(\mathbf{v}|\mathbf{u})$

$$\langle z(\vec{v}|\vec{u}) \rangle = \int_{V_2} z(\vec{u}, \vec{v}) f(\vec{v}|\vec{u}) dV_2$$

rozkład warunkowy znajdujemy wykorzystując jego związek z $f(u,v)$ i rozkładem zredukowanym $f_u(u)$

$$f(\vec{u}, \vec{v}) = f(\vec{v}|\vec{u}) f_u(\vec{u}) \quad \implies \quad f(\vec{v}|\vec{u}) = \frac{f(\vec{u}, \vec{v})}{f_u(\vec{u})}$$

$$\vec{x} = [\vec{u}, \vec{v}]$$

$$\vec{u} = [x_1, x_2, \dots, x_m] \quad \vec{v} = [x_{m+1}, x_{m+2}, \dots, x_n]$$

$$dV_1 = \prod_{i=1}^m dx_i \quad dV_2 = \prod_{i=m+1}^n dx_i$$

Wykorzystując posiadane informacje możemy zapisać całkę w alternatywnej postaci

$$\begin{aligned}
 \langle z \rangle &= \int_{V_1} \int_{V_2} z(\vec{u}, \vec{v}) f(\vec{v}|\vec{u}) f_u(\vec{u}) dV_1 dV_2 \\
 &= \int_{V_1} f_u(\vec{u}) \left[\int_{V_2} z(\vec{u}, \vec{v}) f(\vec{v}|\vec{u}) dV_2 \right] dV_1 \\
 &= \int_{V_1} f_u(\vec{u}) \langle z(\vec{v}|\vec{u}) \rangle dV_1 \\
 &= \frac{1}{N} \sum_{i=1}^N \langle z(\vec{v}|\vec{u}_i) \rangle
 \end{aligned}$$

$$\langle z \rangle = \frac{1}{N} \sum_{i=1}^N \langle z(\vec{v}|\vec{u}_i) \rangle, \quad \vec{u}_i \sim \text{Dist}\{f_u(\vec{u})\}$$

czyli:

- składowe wektora u_i losujemy z rozkładu o fgp $f_u(u)$
- dla każdego wektora u_i wykonujemy całkowanie analityczne w obszarze V_2 – wynik to w zasadzie funkcja wektora u
- ponieważ część obliczeń jest wykonywana w sposób dokładny wariancja zawsze zostanie zmniejszona

Metoda estymatora obciążonego

Punktem wyjścia jest metoda losowania ważonego, czyli szacujemy wartość całki z nową fgp $f^*(x)$

$$\langle z \rangle = \int_V z(x) f(x) dx = \int_V \frac{z(x) f(x)}{f^*(x)} f^*(x) dx = \int_V \frac{g(x)}{f^*(x)} f^*(x) dx$$

Estymatorem nieobciążonym jest średnia wartość funkcji podcałkowej

$$\bar{z} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)}{f^*(x_i)}, \quad x_i \sim \text{Dist}\{f^*(x)\}$$

Metoda zapewnia dużą redukcję wariancji, gdy kształt funkcji $f^*(x)$ jest zbliżony do kształtu $g(x)$.

Problem pojawia się wówczas, gdy losowanie z fgp $f^*(x)$ jest trudne tj. mało wydajne

– bo nie potrafimy skonstruować wydajnego generatora i używamy np. metody eliminacji.

W takim przypadku możemy zastosować estymator obciążony (losujemy wektory w objętości V w sposób jednorodny)

$$\bar{z}_{bias} = \frac{\frac{1}{N} \sum_{i=1}^N g(\vec{x}_i)}{\frac{1}{N} \sum_{i=1}^N f^*(\vec{x}_i)}, \quad \vec{x}_i = [X_1, X_2, \dots, X_n] \sim \text{Dist}\{U^n(V)\}$$

czyli losowanie wektora x_i wykonujemy dla rozkładu jednorodnego w V i liczymy dwie sumy zamiast jednej.

Nowy estymator z_{bias} jest obciążony

$$\langle \bar{z}_{bias} \rangle = \left\langle \frac{\frac{1}{N} \sum_{i=1}^N g(\vec{x}_i)}{\frac{1}{N} \sum_{i=1}^N f^*(\vec{x}_i)} \right\rangle \neq \frac{V \langle \frac{1}{N} \sum_{i=1}^N g(\vec{x}_i) \rangle}{V \langle \frac{1}{N} \sum_{i=1}^N f^*(\vec{x}_i) \rangle}$$

Wartości graniczne licznika i mianownika

$$\lim_{N \rightarrow \infty} \frac{V}{N} \sum_{i=1}^N g(\vec{x}_i) = \lim_{N \rightarrow \infty} V \langle g(\vec{x}) \rangle = \langle z \rangle$$

$$\lim_{N \rightarrow \infty} \frac{V}{N} \sum_{i=1}^N f^*(\vec{x}_i) = \lim_{N \rightarrow \infty} V \langle f^*(\vec{x}) \rangle = 1$$

$$\lim_{N \rightarrow \infty} \bar{z}_{bias} = \frac{\lim_{N \rightarrow \infty} \frac{V}{N} \sum_{i=1}^N g(\vec{x}_i)}{\lim_{N \rightarrow \infty} \frac{V}{N} \sum_{i=1}^N f^*(\vec{x}_i)} = \langle z \rangle$$

Estymator z_{bias} jest spójnym przybliżeniem wartości oczekiwanej $\langle z \rangle$, bo dla $N \rightarrow \infty$ jego wartość dąży do $\langle z \rangle$. Różnice zauważalne będą dla skończonych wartości N .

Zalety stosowania tego podejścia

- można użyć $f^*(x)$ o dość skomplikowanej postaci np. z rozwinięcia w szereg Taylora lub przybliżenie Padego
- jeśli $f^*(x)$ będzie dobrze przybliżać postać $g(x)=z(x)f(x)$ to wartość ilorazu $g(x)/f^*(x)$ będzie w przybliżeniu stała
- wektory x_i losujemy z rozkładu jednorodnego, losowanie będzie szybkie jeśli obszar całkowania V będzie regularny

Całkowanie MC. Podsumowanie.

- całkowanie metodą MC wykonuje się tylko wówczas, gdy inne metody całkowania numerycznego zawodzą (najczęściej gdy liczba wymiarów przekracza 4/5)
- konieczność całkowania MC pojawia się i jest konieczne w zaawansowanych symulacjach MC – wówczas całkowanie to operacja gromadzenia wyników potrzebnych do zrozumienia zachowania układu fizycznego np. średnia energia układu N oddziałujących cząstek a fluktuacje wielkości fizycznych też mogą mieć interpretację fizyczną
- całkowanie obarczone jest zawsze błędem statystycznym, który należy minimalizować
- najprostsza metoda minimalizacji wariancji to zwiększenie liczby losowań (metoda prymitywna)
- zaawansowane metody pozwalają zredukować wariancję o dwa rzędy wielkości przy tej samej liczbie losowań
- generalnie przedstawione metody redukcji wariancji są rzadko stosowane w swej podstawowej postaci, dają one jednak ogólne wyobrażenie w jaki sposób można dokonać redukcji fluktuacji statystycznych wyniku, najczęściej redukcję wariancji uzyskujemy posługując się intuicją i doświadczeniem kombinując ze sobą kilka metod (mimo iż każda z osobna może nie dawać optymalnych wyników)

Inne metody całkowania MC nie omówione na wykładzie:

- ruletka - metoda podobna do losowania warstwowego, ale odpowiedniki prawdopodobieństw p_m określamy arbitralnie na podstawie analizowanego modelu MC (tę metodę wykorzystamy w symulacji oddziaływania fotonów z materią)
- kwadratura MC – odpowiednik klasycznych kwadratur, położenia węzłów są losowane, sortowane a następnie stosuje się kwadraturę trapezów, błąd metody zazwyczaj szybciej maleje niż $1/N$
- quasi-MC – wykorzystuje się w niej nie-losowe deterministyczne ciągi liczb o „zwiększonej” jednorodności (dla małych N bardziej równomierne pokrycie obszaru), zastosowanie głównie w modelach finansowych