



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

Politechnika
Warszawska

Unia Europejska
Europejski Fundusz Społeczny



WSTĘP DO MATEMATYKI FINANSOWEJ

BARTOSZ KOŁODZIEJEK





WYDZIAŁ MATEMATYKI I NAUK INFORMACYJNYCH

Laboratoria 4

Projekt „NERW 2 PW. Nauka - Edukacja - Rozwój - Współpraca” współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

Zadanie 10 pn. „Modyfikacja programów studiów na kierunkach prowadzonych przez Wydział Matematyki i Nauk Informacyjnych”, realizowane w ramach projektu „NERW 2 PW. Nauka - Edukacja - Rozwój - Współpraca”, współfinansowanego jest ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego.

4 maja 2022

Legenda:  – Definicja, **TW.** – Twierdzenie,  – Przykład,  – Uwaga, **LEM.** – Lemat,  – Oznaczenie

1. L4 - REDUKCJA WARIANCJI W METODACH MC - TEORIA

- (1) Metody Monte Carlo. Podstawą teoretyczną metod MC są prawa wielkich liczb, np. MPWL (Kołmogorowa):
Jeśli $(X_n)_n$ są i.i.d. oraz $\mathbb{E}[|X_1|] < \infty$, to

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{1} \mathbb{E}[X_1].$$

Powyżej X 'y mogą być wielowymiarowe. Oznaczmy $\mu = \mathbb{E}[X_1]$. Błędem przybliżenia (nieznanej wartości) μ nazywamy różnicę

$$\varepsilon_n := \frac{X_1 + \dots + X_n}{n} - \mu.$$

Wiemy już, że ε_n zbiega z prawdopodobieństwem 1 do 0. Dla skończonego n , ε_n jest zmienną losową. Zauważmy, że na mocy CTG (o ile $\mathbb{E}[X_1^2] < \infty$) mamy

$$\frac{\sqrt{n}}{\sigma_X} \varepsilon_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma_X^2}} \xrightarrow{d} Z \sim N(0, 1),$$

gdzie $\sigma_X^2 = \text{Var}(X_1)$. Oznaczając $Z_n = \sqrt{n}\varepsilon_n/\sigma_X$ mamy

$$\frac{X_1 + \dots + X_n}{n} = \mu + \frac{\sigma_X}{\sqrt{n}} Z_n,$$

gdzie $Z_n \xrightarrow{d} Z \sim N(0, 1)$.

- (2) Widzimy, że błąd przybliżenia jest tym mniejszy im

- liczba n jest większa,
- wariancją $\text{Var}[X_1]$ jest mniejsza.

Wykonywanie dużej liczby symulacji może być bardzo kosztowne. Żeby zmniejszyć 2-krotnie błąd przybliżenia, powinniśmy zwiększyć n 4-krotnie.

- (3) Metody redukcji wariancji opierają się na następującym pomysśle: wymyślamy schemat, który przybliży to samo, ale szybciej. Innymi słowy, proponujemy nowe X 'y, które będą miały mniejszą wariancję. Dobór metod zależy od rozważanego problemu. Można stosować mix metod.

Jeśli $\mathbb{E}[|f(X)|] < \infty$, to dla ciągu i.i.d. mamy

$$\frac{f(X_1) + \dots + f(X_n)}{n} \xrightarrow{1} \mathbb{E}[f(X_1)] := \mu_f,$$

gdzie $\mu_f = \int_{\mathbb{R}} f(x) \mathbb{P}_X(dx)$. Ogólnie, metody MC służą do znajdowania całek względem miar na podstawie symulacji komputerowych.

- (4) **Importance sampling.** Załóżmy chwilowo, że X ma gęstość, przy czym metoda w oczywisty sposób przepisuje się na rozkłady dyskretne. Szukamy

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx = \int_{\mathbb{R}} \frac{g(x) f_X(x)}{f_Y(x)} f_Y(x) dx = \mathbb{E} \left[\frac{g(Y) f_X(Y)}{f_Y(Y)} \right].$$

By druga z powyższych równości zachodziła konieczne jest, by $\text{supp}(X) \subset \text{supp}(Y)$.

Schemat wygląda następująco: niech $(Y_i)_i$ będzie ciągiem i.i.d. o gęstości f_Y . Wtedy

$$\frac{\sum_{k=1}^n \frac{g(Y_k) f_X(Y_k)}{f_Y(Y_k)}}{n} \xrightarrow{1} \mathbb{E}[g(X)].$$

Działa to dla dowolnej gęstości Y takiej, że $f_X(x) > 0$ implikuje $f_Y(x) > 0$. W jaki sposób wybrać tę gęstość? Tak, by wariancja zmiennej losowej $\frac{g(Y) f_X(Y)}{f_Y(Y)}$ była możliwie mała, w ekstremalnym (i nierealnym) przypadku, gdy $f_Y(y) \propto g(y) f_X(y)$ widzimy, że $\text{Var}[\frac{g(Y) f_X(Y)}{f_Y(Y)}] = 0$. Widzimy, że $\text{Var}[g(X)] > \text{Var}[\frac{g(Y) f_X(Y)}{f_Y(Y)}]$ wtedy i tylko wtedy, gdy

$$\mathbb{E}[g(X)^2] > \mathbb{E} \left[\frac{g(Y)^2 f_X(Y)^2}{f_Y(Y)^2} \right],$$

czyli

$$\int_{\mathbb{R}} g(x)^2 f_X(x) dx > \int_{\mathbb{R}} \frac{g(x)^2 f_X(x)^2}{f_Y(x)} dx,$$

czyli

$$\int_{\mathbb{R}} g(x)^2 \left(1 - \frac{f_X(x)}{f_Y(x)}\right) f_X(x) dx > 0.$$

Słownomuzycznie: chcemy by dla “ważnych” argumentów x (czyli tam, gdzie $g(x)^2 f_X(x)$ jest duże) zachodziła nierówność $f_Y(x) > f_X(x)$. Czyli, regiony “ważne” dla naszego problemu, muszą być bardziej prawdopodobne dla Y .

⚙️ Powiedzmy, że chcemy przybliżyć wartość $\mu = 1 - \Phi(3)$. Skądinąd wiemy, że $\mu \approx 10^{-3}$. Spróbujmy najpierw naiwnie, $f(x) = \mathbb{1}_{(3,\infty)}(x)$, $(X_i)_i$ są i.i.d. z rozkładu $N(0, 1)$. Wtedy

$$\frac{\sum_{k=1}^n f(X_k)}{n} = \frac{\sum_{k=1}^n \mathbb{1}_{(3,\infty)}(X_k)}{n} \xrightarrow{1} \mu.$$

Zauważmy jedn, że dokładność będzie słaba, średnio raz na 10^3 kroków, otrzymamy jedną obserwację X_i przekraczającą 3. Mamy $\text{Var}[f(X_1)] = \mu(1 - \mu) \approx 10^{-3}$.

Dużo lepszym pomysłem jest tutaj zastosowanie importance sampling dla $Y \sim N(\alpha, 1)$. Wtedy

$$\frac{f_X(y)}{f_Y(y)} = e^{-\frac{y^2}{2} + \frac{(y-\alpha)^2}{2}} = e^{\alpha^2/2 - \alpha y}$$

Zatem, dla każdej $\alpha \in \mathbb{R}$ mamy

$$\frac{\sum_{i=1}^n \mathbb{1}_{(3,\infty)}(Y_i) e^{\alpha^2/2 - \alpha Y_i}}{n} \xrightarrow{1} 1 - \Phi(3)$$

lub równoważnie,

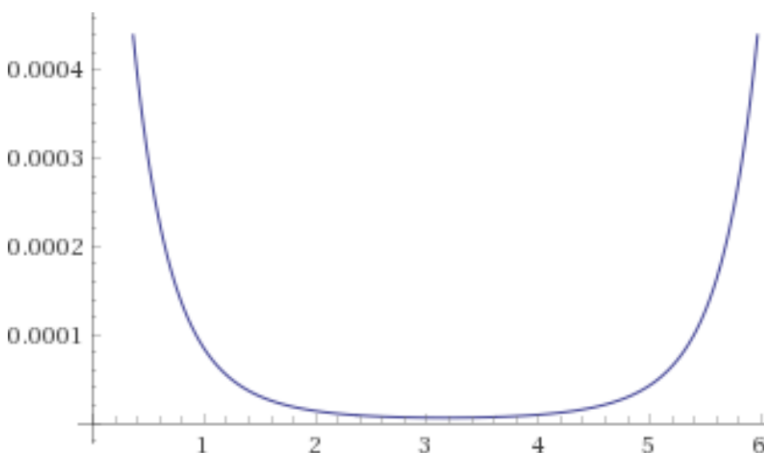
$$\frac{\sum_{i=1}^n \mathbb{1}_{(3,\infty)}(X_i + \alpha) e^{\alpha^2/2 - \alpha(X_i + \alpha)}}{n} = e^{-\alpha^2/2} \frac{\sum_{i=1}^n \mathbb{1}_{(3-\alpha,\infty)}(X_i) e^{-\alpha X_i}}{n} \xrightarrow{1} 1 - \Phi(3)$$

Jak wybrać α ? Oczywiście patrzymy na wariancję,

$$\text{Var} \left[\frac{f(Y)f_X(Y)}{f_Y(Y)} \right] = \text{Var}[\mathbb{1}_{(3,\infty)}(Y) e^{\alpha^2/2 - \alpha Y}] = \text{Var}[\mathbb{1}_{(3,\infty)}(\alpha + X) e^{\alpha^2/2 - \alpha(\alpha + X)}] = \text{Var}[\mathbb{1}_{(3-\alpha,\infty)}(X) e^{-\alpha^2/2 - \alpha X}]$$

Wykres funkcji $\alpha \mapsto \text{Var}[\mathbb{1}_{(3-\alpha,\infty)}(X) e^{-\alpha^2/2 - \alpha X}]$ wygląda następująco:

Plot:



Dla $\alpha = 3$, wariancja jest rzędu 10^{-6} .

(5) **Metoda zmiennych antytetycznych.** Załóżmy, że mamy dane dwa ciągi $(X_n)_n$ oraz $(X'_n)_n$, gdzie

- $(X_n)_n$ jest ciągiem i.i.d.,
- $(X'_n)_n$ jest ciągiem i.i.d.,
- $X_1 \stackrel{d}{=} X'_1$,
- X_n oraz X'_m są niezależne dla $n \neq m$, ale X_n oraz X'_n mogą być zależne.

Szukamy nieznanej wartości $\mathbb{E}[f(X)] = \mathbb{E}[f(X')]$. Rozważmy następujący schemat, który odpowiada wygenerowaniu ciągu długości $2n$:

$$\frac{\sum_{k=1}^n \frac{f(X_k) + f(X'_k)}{2}}{n} \xrightarrow{1} \mu_f.$$

Poprzez odpowiednie sparowanie (coupling) zmiennych X_n i X'_n możemy zmniejszyć wariancję. Zauważmy, że

$$\text{Var}\left[\frac{f(X_1) + f(X'_1)}{2}\right] = \frac{1}{4} (\text{Var}[f(X_1)] + \text{Var}[f(X'_1)] + 2\text{Cov}[f(X_1), f(X'_1)]) = \frac{1}{2} (\text{Var}[f(X_1)] + \text{Cov}[f(X_1), f(X'_1)])$$

Zatem, jeśli $\text{Cov}[f(X_1), f(X'_1)] < 0$, to wariancja zaproponowanego estymatora jest mniejsza niż wariancja estymatora dla $2n$ obserwacji.

Powstaje pytanie, w jaki sposób (możliwie tanio) generować pary (X, X') . Okazuje się, że (zwykle) optymalny X' wyznacza się na podstawie X w bardzo prosty sposób. Jeden z pomysłów opiera się na prostej obserwacji, że jeśli U ma rozkład jednostajny $U([0, 1])$, to $U \stackrel{d}{=} 1 - U$ oraz poniższym lemacie:

LEM. Niech $h: [0, 1] \rightarrow \mathbb{R}$ będzie funkcją monotoniczną, $\int_0^1 h(u)^2 du < \infty$ oraz $U \sim U([0, 1])$. Wtedy

$$\text{Cov}[h(U), h(1 - U)] < 0.$$

Korzystając z powyższego lematu, widzimy, że możemy wziąć np. $(X, X') = (F_X^{-1}(U), F_X^{-1}(1 - U))$. Można pokazać, że to jest najlepszy możliwy wybór, tzn. ten wybór minimalizuje odpowiednią kowariancję (a przynajmniej w sytuacji, gdy $f = \text{id}$). W szczególności, jeśli X ma rozkład symetryczny (np. $X \sim N(0, 1)$), to $X' = -X$.

⚙️ Powiedzmy, że chcemy obliczyć wartość całki

$$I = \int_0^1 \frac{1}{1+x} dx.$$

Jasne jest, że jeśli $(U_i)_i$ są i.i.d. z rozkładu $U([0, 1])$, to

$$\frac{1}{n} \sum_{k=1}^n \frac{1}{1+U_k} \xrightarrow{1} I$$

oraz wariancja $\text{Var}\left[\frac{1}{1+U_k}\right] = 1/2 - \log(2)^2 \approx 0.02$, czyli $\sigma = \sqrt{\text{Var}} \approx 0.14$. Jeśli z kolei rozważymy

$$\frac{1}{n} \sum_{k=1}^n \frac{\frac{1}{1+U_k} + \frac{1}{2-U_k}}{2} \xrightarrow{1} I,$$

to mamy $\text{Var}\left[\frac{\frac{1}{1+U_k} + \frac{1}{2-U_k}}{2}\right] \approx 0.0006$, czyli $\sigma \approx 0.0244$.

- (6) **Metoda warstwowania.** (Stratified sampling) Niech $(A_i)_{i=1}^k$ będzie skończonym (znanym) rozbiciem \mathbb{R} . Wtedy,

$$\mathbb{E}[f(X)] = \sum_{i=1}^k \mathbb{E}[f(X) \mid X \in A_i] \mathbb{P}(X \in A_i) =: \sum_{i=1}^n \mu_i p_i.$$

Zakładamy, że $(p_i)_i$ znamy. Zamiast generować n obserwacji z rozkładu X , możemy wygenerować n_i obserwacji z rozkładu $X \mid X \in A_i$ tak, by $n_1 + \dots + n_k = n$, a następnie wykorzystać je odpowiednio do estymowania wartości μ_i . Takie rozwiązanie może mieć szczególny sens, jeśli funkcja f dla niektórych argumentów jest prawie stała; dla ustalenia uwagi, rozważmy sytuację, gdy $f(x) \approx c$ dla $x \in A_1$. Żeby dokładnie wyestymować μ_1 nie potrzebujemy wielu symulacji, ponieważ wariancja $f(X) \mid X \in A_1$ będzie mała. "Zaoszczędzone" w ten sposób symulacje, możemy spożytkować na dokładniejszą estymację średniej f po innych obszarach.

Niech $q_i = n_i/n$ oraz

$$\frac{f(X_1^{(i)}) + \dots + f(X_{n_i}^{(i)})}{n_i}$$

będzie estymatorem, $\mu_i (X^{(i)} \stackrel{d}{=} X \mid X \in A_i)$. Wtedy estymatorem $\mathbb{E}[f(X)]$ jest

$$\hat{\mu} = \sum_{i=1}^k \left(\frac{f(X_1^{(i)}) + \dots + f(X_{n_i}^{(i)})}{n_i} \right) p_i = \frac{1}{n} \sum_{i=1}^k \frac{p_i}{q_i} \left(f(X_1^{(i)}) + \dots + f(X_{n_i}^{(i)}) \right)$$

Wariancja tego estymatora wynosi

$$\text{Var}[\hat{\mu}] = \frac{1}{n^2} \sum_{i=1}^k \frac{p_i^2}{q_i^2} n_i \text{Var}[f(X^{(i)})] = \frac{1}{n} \sum_{i=1}^k \frac{p_i^2}{q_i} \sigma^{(i)^2}.$$

Możemy minimalizować tę wariancję poprzez odpowiedni dobór liczności próbek, czyli ciągu $(q_i)_i$, $q_i \geq 0$, $q_1 + \dots + q_k = 1$. Można pokazać, że wybór $q_i = p_i$ (alokacja proporcjonalna) zawsze redukuje wariancję, jednak można lepiej. Za pomocą mnożników Lagrange'a pokazujemy, że optymalna alokacja zadana jest przez

$$q_i = \frac{p_i \sigma^{(i)}}{\sum_{j=1}^k p_j \sigma^{(j)}}.$$

Niestety, żeby z niej skorzystać, musielibyśmy znać ogólnie nieznane wartości $\sigma^{(i)}$, $i = 1, \dots, k$. Czasami w pierwszym kroku eksperymentu estymuje się te wartości, również metodami MC.

Pozostaje jeszcze problem jak generować zmienne losowe $X^{(i)}$ z rozkładu $X \mid X \in A_i$. Załóżmy, że zbiory A_i są postaci $(a_i, a_{i+1}]$. Ogólnie, wiemy, że $X \stackrel{d}{=} F_X^{-1}(U)$, gdzie $U \sim U([0, 1])$. W taki sam sposób można pokazać, że

$$X^{(i)} \stackrel{d}{=} F_X^{-1}(s_{i-1} + U p_i),$$

gdzie $p_i = \mathbb{P}(X \in A_i)$ oraz $s_i = p_1 + \dots + p_i$, $s_0 = 0$.

- (7) **Metoda zmiennych kontrolnych.** Zakładamy, że $\mathbb{E}[|k(X)|] < \infty$ oraz, że znamy wartość $\mathbb{E}[k(X)]$. Wtedy

$$\mu_f = \mu_k + (\mu_f - \mu_k) = \mu_k + \mu_{f-k}.$$

Żeby przybliżyć μ_f wystarczy przybliżyć μ_{f-k} . Jeśli funkcja $f - k$ jest bliska stałej (ma małą wariancję), to mamy szansę na duży uzysk. Istotnie,

$$\mu_k + \frac{\sum_{i=1}^n (f(X_i) - k(X_i))}{n} \xrightarrow{1} \mu_f$$

oraz szybkość zbieżności zależy od

$$\text{Var}[f(X) - k(X)] = \text{Var}[f(X)] + \text{Var}[k(X)] - 2\text{Cov}[f(X), k(X)].$$

Warunek $\text{Var}[f(X) - k(X)] < \text{Var}[f(X)]$ jest równoważny warunkowi $\text{Var}[k(X)] < 2\text{Cov}[f(X), k(X)]$, który w ogólnym przypadku może być sprawdzony symulacyjnie.

Często rozważa się wiele zmiennych kontrolnych $(k_i)_{i=1}^m$, tzn. gdy

$$k = \sum_{i=1}^m \alpha_i k_i,$$

gdzie $\alpha_i \in \mathbb{R}$, $i = 1, \dots, m$. Współczynniki α_i (również, gdy $m = 1$) można dobrać optymalnie z kryterium minimalizacji $\text{Var}[f(X) - k(X)]$.

⚙️ **Zastosowanie dla opcji azjatyckich.** Opcję azjatycką nazywamy (europejską) opcję o wypłacie $X = \left(\frac{1}{T} \sum_{k=1}^T S_{t_k} - K\right)^+$. „Azjatyckość” tej opcji polega na fakcie, że wypłata zależy od średniej cen aktywa w wybranych momentach. Wiemy, że jej wartość w chwili 0 wynosi

$$\Pi_0(X) = \mathbb{E}_{\mathbb{P}^*} \left[\frac{X}{B_T} \right],$$

Jako zmienną kontrolną dobrze jest wybrać

$$Y = \frac{1}{B_T} \left(\sqrt[T]{\prod_{k=1}^T S_{t_k}} - K \right)^+,$$

którego cenę w chwili 0 możemy jawnie znaleźć w modelu ciągłym Blacka-Scholesa (w swoim czasie...).

- (8) **Redukcja wariancji poprzez zamianę zmiennych.** Powiedzmy, że chcemy oszacować wartość całki

$$\int_3^\infty \frac{1}{\pi(1+x^2)} dx.$$

Żeby to zrobić moglibyśmy wygenerować ciąg $(X_n)_n$ i.i.d. z rozkładu Cauchy'ego $C(1)$ oraz rozważyć estymator

$$\frac{\sum_{i=1}^n \mathbb{1}_{(3,\infty)}(X_i)}{n}.$$

Mamy $\text{Var}[\mathbb{1}_{(3,\infty)}(X_i)] \approx 0.1 \cdot 0.9 = 0.09$.

Możemy również postąpić inaczej. W naszej całce zamieńmy zmienną $x = 3/y$. Wtedy

$$\int_3^\infty \frac{1}{\pi(1+x^2)} dx = \int_0^1 \frac{3}{\pi(9+y^2)} dy.$$

Zatem tę samą wartość możemy przybliżać za pomocą ciągu $(Y_n)_n$ i.i.d. z rozkładu $U([0, 1])$ i estymatora

$$\frac{\sum_{i=1}^n \frac{3}{\pi(9+Y_i^2)}}{n}.$$

Można sprawdzić, że $\text{Var}[\frac{3}{\pi(9+Y_i^2)}] \approx 10^{-5}$.

(9) **Quasi-Monte Carlo.** Powiedzmy, że mamy do przybliżenia wartość całki

$$I := \int_{(0,1)^d} f(u_1, \dots, u_d) du_1 \dots du_d$$

gdzie d jest bardzo duże. Standardowe MC mówi nam, że

$$\frac{\sum_{i=1}^n f(\underline{U}_i)}{n} \xrightarrow{1} I,$$

gdzie $(\underline{U}_n)_n$ jest ciągiem wektorów i.i.d. rozkładu $U([0, 1]^d) = U([0, 1])^{\otimes d}$. Wiemy, że błąd przybliżenia maleje jak $\frac{c}{\sqrt{n}}$.

Okazuje się, że jeśli zamienimy (pseudo)-losowy ciąg $(\underline{U}_n)_n$ na ciąg, który wypełnia przestrzeń “lepiej” (pełniej?) niż ciąg losowy, to możemy otrzymać zbieżność rzędu $c \frac{\log(n)^{d-1}}{n}$. Metoda QMC jest metodą w pełni deterministyczną (podobnie jak całkowanie numeryczne).

Przykładem takiego ciągu jest tzw. ciąg Haltona, który opiera się na podziale odcinka $[0, 1]$ na coraz większą liczbę części. W \mathbb{R}^1 wygląda następująco:

$$(baza = 2) : \quad 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, \dots$$

$$(baza = 3) : \quad 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9, 1/27, \dots$$

Żeby otrzymać ciąg w \mathbb{R}^2 można skleić oba powyższe ciągi.