Metoda Monte Carlo - Cvičení III ZS 2012/13 a domácí úlohy

1 Výpočet určitého integrálu $J = \int_a^b f(x)dx$ pomocí metody Monte Carlo

Některé potřebné vzorečky

- Tato kapitolka není úplná, obsahuje pouze některé vzorečky.
- Detaily, odvození a notace viz poznámky ze cvičení nebo skripta k přednášce.

1. Metoda vzorkování plochy
   zvolit vhodnou mez c
   $\alpha_i \sim R(< a, b > \times < 0, c >)$, $\xi_i = 1$ pk. $f(\alpha_{i1}) \leq \alpha_{i2}$
   $\theta_1 = \frac{c(b-a)}{N} \sum_{i=1}^{N} \xi_i$, N je počet realizací ξ
   $E\theta_1 = J$,
   $D\theta_1 = \frac{c^2(b-a)^2}{N^2} \sum_{i=1}^{N} D\xi_i = \frac{c^2(b-a)^2}{N} D\xi$, $D\xi$ lze odhadnout např. výběrovým rozptylem s^2 nebo podle definice $D\xi = E\xi^2 - (E\xi)^2$ odhadem $E\xi^2$ a $E\xi$
   $s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\xi - E\xi)^2 = \frac{1}{N-1}(S_2 - \frac{(S_1)^2}{N}), S_2 = \sum_i \xi_i^2, S_i = \sum_i \xi_i$.

2. Metoda vzorkování funkce
   $\theta_2 = \frac{b-a}{N} \sum_{i=1}^{N} f(\xi_i), \xi \sim R(a, b)$,
   $E\theta_2 = J$,
   $D\theta_2 = \frac{(b-a)^2}{N^2} \sum_{i=1}^{N} Df(\xi_i) = \frac{(b-a)^2}{N} Df(\xi) = \frac{(b-a)^2}{N}(Ef^2(\xi) - (Ef(\xi))^2)$, $Df(\xi)$ lze odhadnout např. výběrovým rozptylem nebo odhadem $Ef^2(\xi)$.

3. Metoda vzorkování funkce s vyjmutím hlavní části
   zvolit vhodnou funkci g
   $\theta_3 = \frac{b-a}{N} \sum_{i=1}^{N} (f(\xi_i) - g(\xi_i)) + I, \xi \sim R(a, b)$, $I = \int_a^b g(x)dx$ známe
   $E\theta_3 = J$,
   $D\theta_3 = \frac{(b-a)^2}{N^2} \sum_{i=1}^{N} D(f(\xi_i) - g(\xi_i))$

4. Metoda vzorkování funkce se skupinovým výběrem
   zvolit vhodný počet intervalů m, vhodné l_k (délka intervalu) a N_k (počet realizací)
   $\theta_4 = \sum_{k=1}^{m} \frac{l_k}{N_K} \sum_{i=1}^{N_k} f(\xi_i^k), \xi^k \sim R(a_{k-1}, a_k)$, $l_k = (a_k - a_{k-1})$, N_k zvolíme, např. $N_k = \frac{l_k N}{b-a}$
   $D\theta_4 = \sum_k \frac{l_k^2}{N_K} Df(\xi^k)$

5. Metoda vzorkování plochy s vyjmutím hlavní části – obdobně

6. Metoda vzorkování plochy se skupinovým výběrem – obdobně

7. Metody vzorkování funkce a plochy s relevantním výběrem

8. Integrály funkcí více proměnných - např. metoda vzorkování funkce
   $\theta_5 = \frac{\prod_j (b_j - a_j)}{N} \sum_{i=1}^{N} f(\xi_i)$,
   $E\theta_5 = J$,
   $D\theta_5 = \frac{\prod_j (b_j - a_j)^2}{N^2} \sum_{i=1}^{N} Df(\xi_i) = \frac{\prod_j (b_j - a_j)^2}{N} Df(\xi) = \frac{\prod_j (b_j - a_j)^2}{N}(Ef^2(\xi) - (Ef(\xi))^2)$, $Df(\xi)$ lze odhadnout např. výběrovým rozptylem nebo odhadem $Ef^2(\xi)$.

9. Integrály funkcí více proměnných - metoda vzorkování plochy – obdobně

2 Domácí úloha: Programy

Implementujte čtyři z následujících metod pro výpočet integrálu $J = \int_a^b f(x)dx$. Funkce vždy spočte odhad integrálu a směrodatnou odchylku odhadu. Alespoň jednu z metod implementuje obecně (pro integrál o n proměnných).

1. Vzorkování plochy
2. Vzorkování plochy s vyjmutím hlavní části
3. Vzorkování plochy se skupinovým výběrem
4. Vzorkování plochy s relevantním výběrem
5. Vzorkování funkce
6. Vzorkování funkce s vyjmutím hlavní části
7. Vzorkování funkce se skupinovým výběrem
8. Vzorkování funkce s relevantním výběrem

Příklady

1. $J = \int_0^4 (\sqrt{x} - 1)dx$.
2. $J = \int_0^1 (e^x - sin(x))dx$.
3. $J = \int_0^1 \int_0^1 (8 - x^2 - y^2)dydx$ (stačí pro metodu, kterou jste implementovali obecně).

Pomocí implementovaných funkcí spočtěte odhady (a jejich směrodatné odchylky) integrálů θ pro různé hodnoty N (např. $10, 100, 10000, 1000000, 10000000$). Pokud je možné analyticky spočítat hodnotu integrálu J, spočtětě i chybu $\delta = |\theta - J|$. Porovnejte přesnost a rozptyl odhadu a časové nároky vašich implementací jednotlivých metod. Pokuste se zvolit co nejvhodnější parametry metod (omezující obdélník, funkce g, intervaly...)

Přiložený dokument by měl obsahovat především:

- Popis, které metody jste implementovali a **jak jste volili jejich parametry** (např. omezující obdélník u metody vzorkování plochy, funkci g u metod s vyjmutím hlavní části atd.).
- Tabulku (tabulky), kde budou hodnoty odhadu a odhadu jeho směrodatné odchylky (spočtené za využití výběrového rozptylu) a čas výpočtu pro různé hodnoty N (např. $N = 10^2$, 10^4, 10^6, 10^8). Výsledky jednotlivých experimentů mohou být dohromady v jedné tabulce nebo ve zvláštních tabulkách.
- Slovní zhodnocení výsledků (hlavně přesnost odhadu v závislosti na zvolené metodě a na N a časové nároky metody)
- Případně zajímavé implementační detaily.

Pošlete zprávu i zdrojové soubory.