

# Ročníkový projekt - letný semester

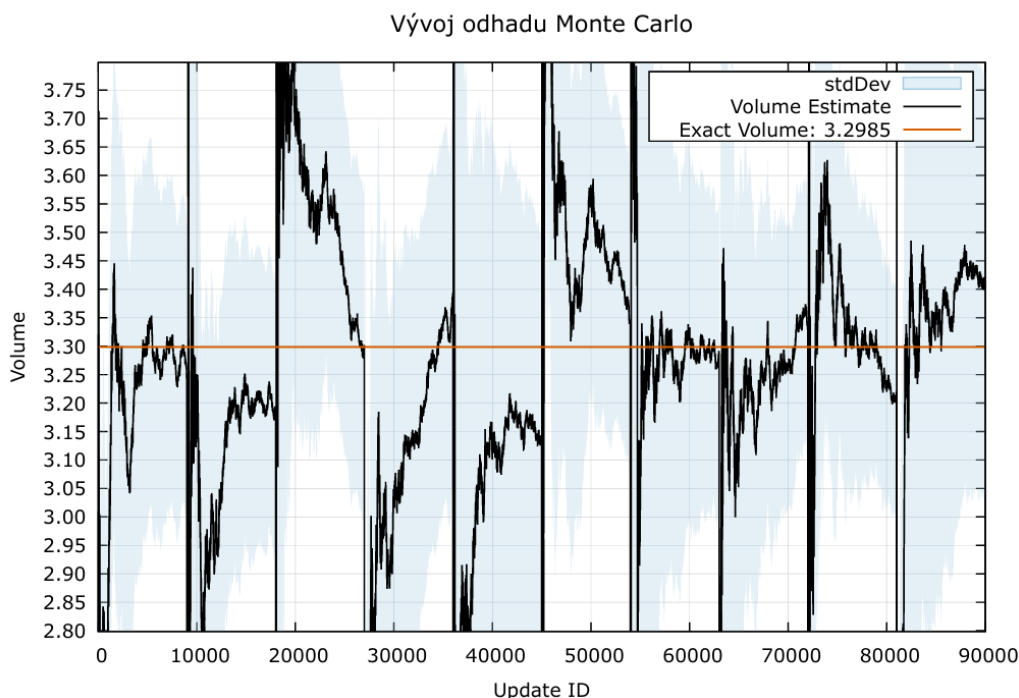
Cieľom tohto ročníkového projektu v letnom semestri bola analýza a detailné debuggovanie metód Monte Carlo integrácie. Projekt bol zameraný na randomizovanú optimalizačnú metódu (rOpt), ktorá bola navrhnutá a implementovaná počas zimného semestra. Kým zimný semester bol venovaný zabezpečeniu opakovania pokusov pomocou správy seedu pre generátor a samotnému návrhu metódy rOpt, tak práca v letnom semestri spočívala v pochopení vnútorného správania sa tohto algoritmu.

Hlavným problémom, ktorý má deterministická metóda opt, je jej náchylnosť k neoptimálnemu zameraniu sa na jedinú dimenziu v prípade, ak v nej počiatkové vzorky náhodne vykázali vysoký relatívny rozptyl (CVar2). Metóda rOpt tento problém rieši priradením pravdepodobnostných váh jednotlivým dimenziám, čo zabezpečuje vyváženejšie vzorkovanie. Na to aby sme overili tento predpoklad a pre analýzu vývoja odhadov bolo treba implementovať logovanie údajov a následne ich vizualizovať vo forme grafov.

## Metodika experimentu a logovanie

Pre účely analýzy bol zvolený primárne experiment s dimenziou nastavenou na  $d = 9$ . Výpočet celkového objemu bol rozložený na osem samostatných integrátorov. Počas simulácie boli po každom kroku aktualizácie zaznamenávané hodnoty pre každú z týchto ôsmich dimenzií. Sledovali sme vývoj lokálnych odhadov integrálu, ich rozptyly a hodnoty relatívneho rozptylu. Zároveň s lokálnymi štatistikami bol zaznamenávaný vývoj globálneho odhadu celkového objemu integrálu. Tieto záznamy nám umožnili vizualizovať proces algoritmu v čase a identifikovať kľúčové anomálie v správaní odhadov.

## Vývoj celkového odhadu

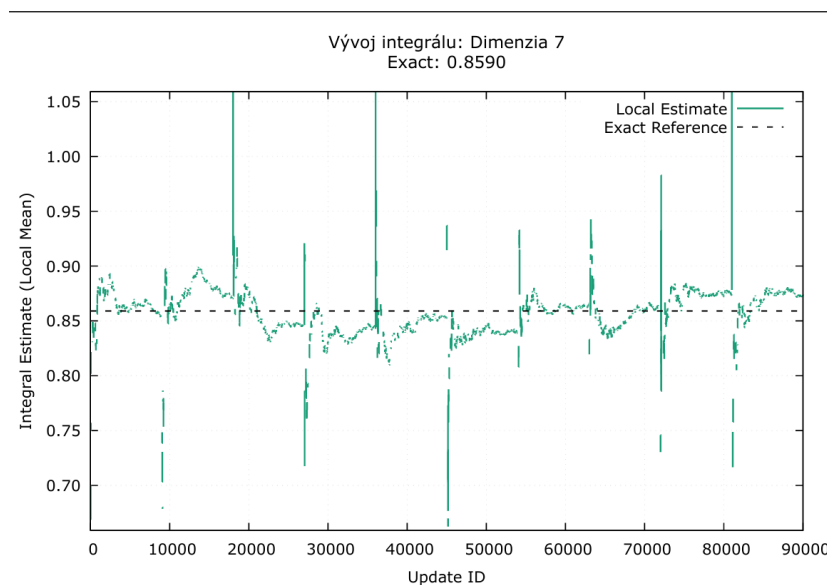


Obr. č.1: Odhad celého integrálu pre 10 repetícií

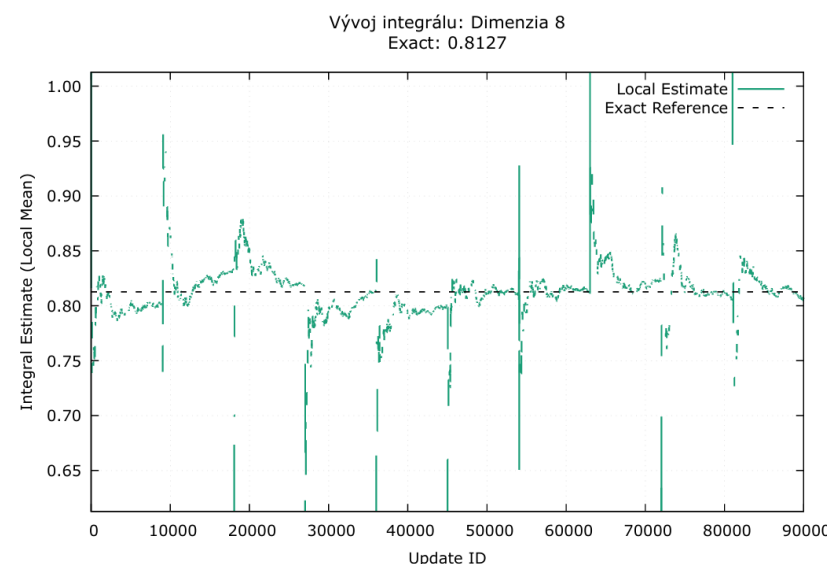
Hlavným výstupom analýzy je graf znázorňujúci vývoj celkového odhadu objemu pomocou metódy Monte Carlo v porovnaní s exaktnou referenčnou hodnotou (3.2985). Z grafu je vidieť, že v počiatkových fázach odhad výrazné oscilácie, ktoré však postupne s rastúcim počtom vzoriek klesajú, a algoritmus stabilne konverguje k presnej hodnote.

V grafe je tiež jasne vidieť veľké lokálne maximá a minimá (skoky). Tieto skoky sú spôsobené samotnou podstatou opakovaných experimentov (repetícií). Na začiatku každého nového opakovania algoritmus nemá dostatočné množstvo vzoriek, čo logicky vedie k vysokej miere nepresnosti lokálnych odhadov a k následným prudkým výkyvom. S pribúdajúcim počtom aktualizácií sa vplyv týchto počiatkových extrémov znižuje, výpočet sa stabilizuje a blíži k referenčnej hodnote.

## Rozdelenie na dimenzie a hľadanie chýb



Obr č.2: Odhady integrálu v dimenzii 7



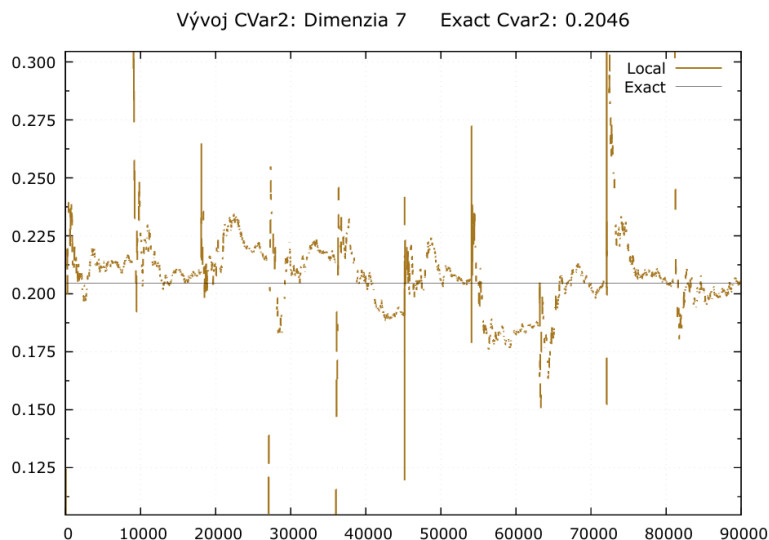
Obr č.3: Odhady integrálu v dimenzii 8

Analýza čiastkových grafov pre jednotlivé dimenzie poskytla pohľad na to, ako lokálne odchýlky ovplyvňujú celkový výsledok. Keďže výsledný integrál je tvorený súčinom dimenzionálnych odhadov, odchýlky v jednotlivých dimenziách môžu viesť k zosilneniu chyby vo výslednom odhade.

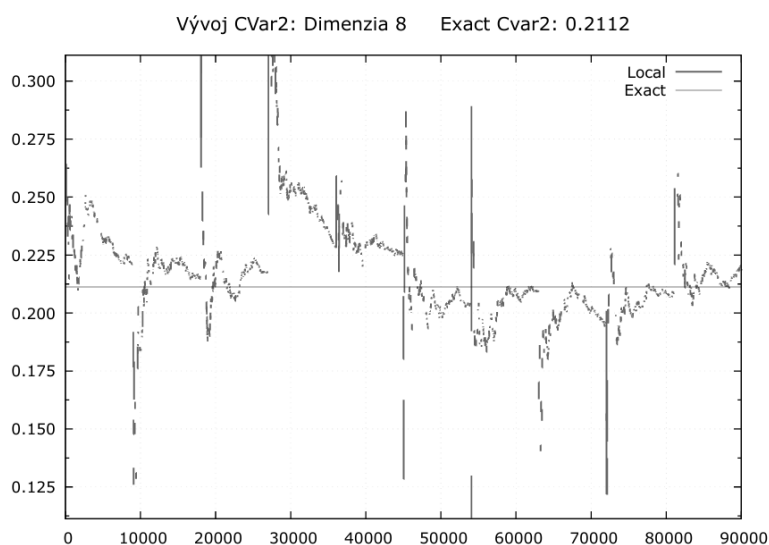
Z porovnania grafov vyplynulo (intuitívne) zistenie. V situácii, kedy algoritmus na základe počiatkovej vzorky podhodnotí reálnu hodnotu integrálu viacerými dimenziami súčasne, výsledný súčin týchto

podhodnotených odhadov vedie k veľkej odchýlke celkového odhadu objemu smerom nadol. Rovnaký princíp platí aj pri nadhodnotení. Toto hovorí o dôležitosti znižovania rozptylu vo všetkých dimenziách, pretože zanedbanie čo i len jednej dimenzie môže mať veľký vplyv na celkovú chybu.

## Opodstatnenie rOpt na základe relatívneho rozptylu



Obr č.4: Vývoj Cvar2 v dimenzii 7



Obr č.5: Vývoj Cvar2 v dimenzii 8

Grafy znázorňujúce vývoj CVar2 v jednotlivých dimenziách odôvodňujú použitie randomizovaného prístupu metódy rOpt v porovnaní s deterministickou metódou opt. V skorých fázach opakovania algoritmu sa v grafoch CVar2 objavujú extrémne výkyvy spôsobené príchodom vzorky s vysokou odchýlkou.

Ak by v takejto situácii bola použitá metóda opt, algoritmus by bral danú dimenziu ako najviac chybovú a deterministicky by do nej vkladal väčšinu ďalších vzoriek. Tým by znemožnil upresňovanie ostatných dimenzií. V dôsledku chýb by to viedlo k zníženiu presnosti celkového odhadu. Naproti tomu, metóda rOpt pomocou váženého pravdepodobnostného výberu prideli takejto problémovej dimenzii vyššiu prioritu. Zároveň zachová nenulovú pravdepodobnosť vzorkovania aj pre ostatné dimenzie. Tým efektívne predchádza stagnácii a zabezpečuje lepšiu konvergenciu aj napriek prítomnosti počiatočného štatistického šumu.

## Lokálny odhad a relatívny rozptyl

Pri analýze niektorých experimentov boli ale spozorované anomálie: v niektorých repetíciách (napr. v 5. repetícii z obrázkov) zostával celkový odhad integrálu dlhodobo podhodnotený napriek tomu, že hodnoty (CVar2) nemali výkyvy a stabilne klesali.

CVar2 vyjadruje iba mieru rozptýlenia vzoriek okolo ich vlastného lokálneho priemeru. Neposkytuje žiadnu informáciu o tom, ako ďaleko je tento lokálny priemer vychýlený od skutočnej analytickej hodnoty integrálu. Ak v počiatkových fázach repetície padne na základe seedu taká postupnosť vzoriek, ktorá posunie lokálny priemer nadol, ale samotné vzorky zostanú v relatívnej blízkosti tohto nového (aj keď nesprávneho) stredu, hodnota CVar2 zostane nízka. Výsledkom je, že presnosť odhadu v počiatkových krokoch danej repetície silne závisí od počiatkovej distribúcie. Skreslenie sa následne eliminuje až pribúdajúcimi vzorkami. Vplyvom multiplikatívneho charakteru chýb ale môže tento počiatkový „bias“ dočasne spôsobiť výrazný prepád globálneho odhadu.