Multi-Level Monte Carlo

Devon Troester, Eric Hallman

NCSU

July 10, 2020

Introduction

Goal: Estimate $\mathbb{E}_\theta[q(\theta)]$ using Monte Carlo.

Procedure: For i.i.d. samples $\Theta = \{\theta_1, \dots, \theta_N\}$,

$$\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} q(\theta_i)$$

$$\mathbb{E}_\Theta[\bar{X}_N] = \mathbb{E}_\theta[q(\theta)]$$

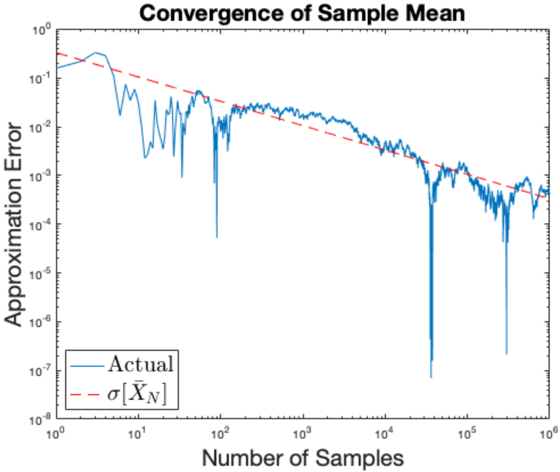
$$\mathbb{V}_\Theta[\bar{X}_N] = \frac{1}{N}\mathbb{V}_\theta[q(\theta)]$$

Introduction

The Mean Squared Error (MSE) is

$$\begin{aligned}\mathbb{E}_\Theta[(\bar{X}_N - \mathbb{E}[q])^2] &= \underbrace{\mathbb{V}_\Theta[\bar{X}_N]}_{\text{Variance}} + \underbrace{(\mathbb{E}[\bar{X}_N] - \mathbb{E}[q])^2}_{\text{Bias}^2} \\ &= \frac{1}{N}\mathbb{V}_\theta[q(\theta)] + 0 \\ &\propto 1/N.\end{aligned}$$

Example: $q(\theta) = \frac{4}{3}\theta$, where $\theta \sim U(0, 1)$



Introduction

Common Problem: We can approximate $q(\theta)$, but not compute it directly

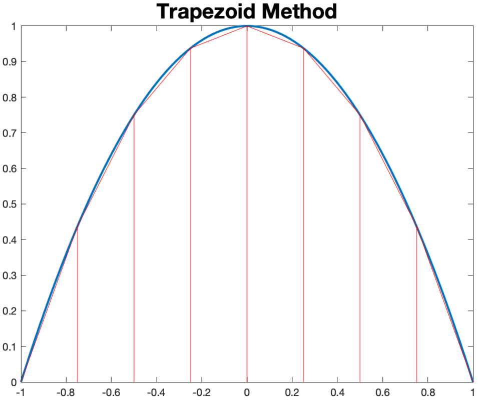
Examples:

- $q(\theta) = y(T, \theta)$ where $y'(t;\theta) = f(t, y; \theta)$
- $q(\theta) = \theta^T A^{-1} \theta$ where A is large/sparse
- $q(\theta) = \int_x f(x;\theta)\, dx$

Define $q_k(\theta)$ for $k \in \{0, 1, 2, \ldots, K\}$. The **cost** of evaluating $q_k(\theta)$ and the **accuracy** of $q_k(\theta)$ both increase with k.

Single Level Method

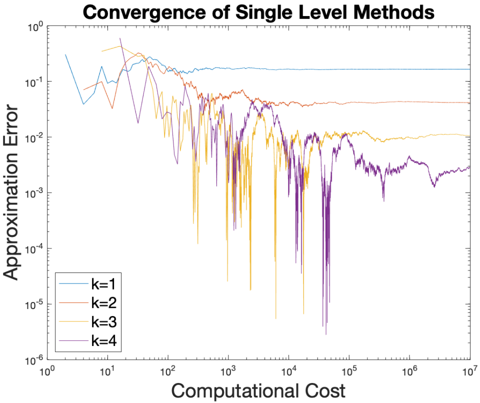
Example: $f(x;\theta) = \theta(1 - x^2)$, $\quad q(\theta) = \int_{-1}^{1} f(x;\theta)\, dx = \frac{4}{3}\theta$



With 2^k trapezoids, $q_k(\theta) = \frac{4}{3}\theta(1 - 1/4^k)$

Single Level Method

Idea 1: To get MSE ϵ^2, choose k so that $\underbrace{(\mathbb{E}[q_k(\theta)] - \mathbb{E}[q(\theta)])^2}_{\text{Bias}^2} \leq \epsilon^2/2$



Single Level Method

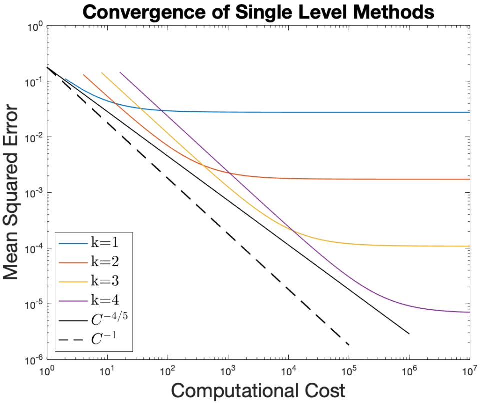
Idea 1: To get MSE ϵ^2, choose k so that $\underbrace{(\mathbb{E}[q_k(\theta)] - \mathbb{E}[q(\theta)])^2}_{\text{Bias}^2} \leq \epsilon^2/2$

- Our trapezoid method problem: Cost $\propto 2^k$ and Bias$^2 \propto 2^{-4k}$
  - Thus: choose k so $C_k \propto 1/\sqrt{\epsilon}$
- Need $N_k \propto 1/\epsilon^2$ evaluations, since Var $\propto 1/N$
- Total cost: $C = C_k N_k \propto \epsilon^{-\frac{5}{2}}$
- Alternately: $\epsilon^2 \propto C^{-\frac{4}{5}}$

This is not asymptotically optimal! We want $\epsilon^2 \propto C^{-1}$

Single Level Method

Idea 1: To get MSE ϵ^2, choose k so that $\underbrace{(\mathbb{E}[q_k(\theta)] - \mathbb{E}[q(\theta)])^2}_{\text{Bias}^2} \leq \epsilon^2/2$



Multi-Level Methods

Idea 2: Write $q_K(\theta)$ as a telescoping series. Define

$$\Delta_k = \begin{cases} q_0 & k = 0, \\ q_k - q_{k-1} & k \geq 1, \end{cases}$$

get

$$\mathbb{E}[q_K(\theta)] = \sum_{k=0}^{K} \mathbb{E}[\Delta_k(\theta)]$$

Multi-Level Methods

Idea 2: Write $q_K(\theta)$ as a telescoping series. Define

$$\Delta_k = \begin{cases} q_0 & k = 0, \\ q_k - q_{k-1} & k \geq 1, \end{cases}$$

get

$$\mathbb{E}[q_K(\theta)] = \sum_{k=0}^{K} \mathbb{E}[\Delta_k(\theta)]$$

- When k is small, Δ_k is **cheap to evaluate**

Multi-Level Methods

Idea 2: Write $q_K(\theta)$ as a telescoping series. Define

$$\Delta_k = \begin{cases} q_0 & k = 0, \\ q_k - q_{k-1} & k \geq 1, \end{cases}$$

get

$$\mathbb{E}[q_K(\theta)] = \sum_{k=0}^{K} \mathbb{E}[\Delta_k(\theta)]$$

- When k is small, Δ_k is **cheap to evaluate**
- When k is large, Δ_k has **low variance**

Multi-Level Methods

Idea 2: Write $q_K(\theta)$ as a telescoping series. Define

$$\Delta_k = \begin{cases} q_0 & k = 0, \\ q_k - q_{k-1} & k \geq 1, \end{cases}$$

get

$$\mathbb{E}[q_K(\theta)] = \sum_{k=0}^{K} \mathbb{E}[\Delta_k(\theta)]$$

- When k is small, Δ_k is **cheap to evaluate**
- When k is large, Δ_k has **low variance**

Can we combine these to our advantage??

Cost Analysis

- C_k: The cost of 1 evaluation of $q_k(\theta)$ (increases with k)
- V_k: The variance $\mathbb{V}[\Delta_k(\theta)]$ (decreases with k)
- N_k: The number of evaluations at level k (how to choose?)

- Total cost: $\sum_{k=0}^{K} N_k C_k$
- Total variance: $\sum_{k=0}^{K} N_k^{-1} V_k$

Cost Analysis

Optimizing, find

$$N_k = \mu \sqrt{V_k / C_k},$$
$$\mu = \epsilon^{-2} \sum_{k=0}^{k} \sqrt{V_k / C_k},$$

yielding variance $V = \epsilon^2$ and total cost

$$C = \epsilon^{-2} \left(\sum_{k=0}^{K} \sqrt{V_k C_k} \right)^2 .$$

Important: does $V_k C_k$ increase or decrease with k?

- Increase: **finest** level dominates the cost (bad)
- Decrease: **coarsest** level dominates the cost (good)

Cost Analysis

Theorem (Giles 2008)

Suppose we have α, β, γ satisfying $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and

1. $|\mathbb{E}[q_K - q]| = \mathcal{O}(2^{-\alpha k})$,
2. $V_k = \mathcal{O}(2^{-\beta k})$,
3. $C_k = \mathcal{O}(2^{\gamma k})$.

Then for any $\epsilon > 0$ there is a multilevel estimator with MSE $< \epsilon^2$ and cost

$$C = \mathcal{O}\begin{cases} \epsilon^{-2} & \beta > \gamma, \quad (\textit{optimal}) \\ \epsilon^{-2}(\log \epsilon)^2 & \beta = \gamma, \\ \epsilon^{-2-(\gamma-\beta)/\alpha} & \beta < \gamma. \end{cases}$$

If $\beta > \gamma$, we can attain the optimal rate!

Cost Analysis

Applied to our trapezoid method problem:

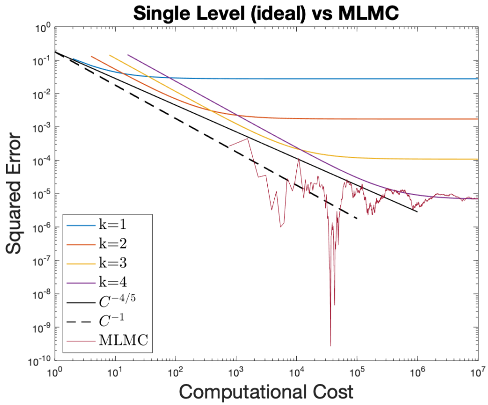
$$q_k(\theta) = \frac{4}{3}\theta(1 - 1/4^k)$$

1. $|\mathbb{E}[q_k - q]| = \frac{2}{3} \cdot 4^{-k} \propto 2^{-2k}$ $\quad (\alpha = 2)$
2. $V_k = \frac{4}{3} \cdot 4^{-2k} \propto 2^{-4k}$ $\quad (\beta = 4)$
3. $C_k = 2^k + 1 \propto 2^k$ $\quad (\gamma = 1)$

Since $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and $\beta > \gamma$, we can attain the optimal rate.

MLMC: Example

Try: $N_k \propto \sqrt{V_k / C_k} \propto 2^{-\frac{5}{2}k}$
Sample $k = 1, 2, 3, 4$ in approximate ratio $181 : 32 : 6 : 1$



Same limiting accuracy as the lowest level, but gets there faster

Summary

- Exact measurement $q(\theta)$ is inaccessible/expensive
- Multiple levels $q_k(\theta)$ offer cost/bias tradeoff
- Combine many coarse samples with a few fine samples
- Optimal rate MSE $\propto C^{-1}$ often attainable
- **Not** an unbiased method, but we can make the bias arbitrarily small by adding more levels

Summary(?)

But wait, there's more!

Unbiased Estimator

Idea 3: Write $q(\theta)$ as a telescoping series

$$\mathbb{E}[q(\theta)] = \sum_{k=0}^{\infty} \mathbb{E}[\Delta_k(\theta)]$$

Unbiased Estimator

Idea 3: Write $q(\theta)$ as a telescoping series

$$\mathbb{E}[q(\theta)] = \sum_{k=0}^{\infty} \mathbb{E}[\Delta_k(\theta)]$$

- Cannot directly estimate infinitely many terms
- Use a probabilistic version!

Unbiased Estimator

Idea 3: Write $q(\theta)$ as a telescoping series

$$\mathbb{E}[q(\theta)] = \sum_{k=0}^{\infty} \mathbb{E}[\Delta_k(\theta)]$$

- Cannot directly estimate infinitely many terms
- Use a probabilistic version!

Several variants; simplest is

$$Z = \frac{\Delta_N}{p_N}$$

where $N \geq 0$ is an integer-valued random variable.

Unbiased Estimator

Justification for estimator Z: for any k, $\mathbb{E}_N\left[\frac{\mathrm{I}[N=k]}{\Pr[N=k]}\right] = 1$. Thus,

$$\begin{aligned}
\mathbb{E}_\theta[q(\theta)] &= \sum_{k=0}^{\infty} \mathbb{E}_\theta[\Delta_k(\theta)] \\
&= \sum_{k=0}^{\infty} \mathbb{E}_\theta[\Delta_k(\theta)] \mathbb{E}_N\left[\frac{\mathrm{I}[N=k]}{\Pr[N=k]}\right] \\
&= \mathbb{E}_{\theta,N}\left[\sum_{k=0}^{\infty} \Delta_k(\theta) \frac{\mathrm{I}[N=k]}{\Pr[N=k]}\right] \\
&= \mathbb{E}_{\theta,N}\left[\frac{\Delta_N(\theta)}{p_N}\right] \\
&= \mathbb{E}_{\theta,N}[Z].
\end{aligned}$$

Unbiased Estimator

Theorem (Rhee/Glynn 2015)

The second moment of Z, if finite, is given by

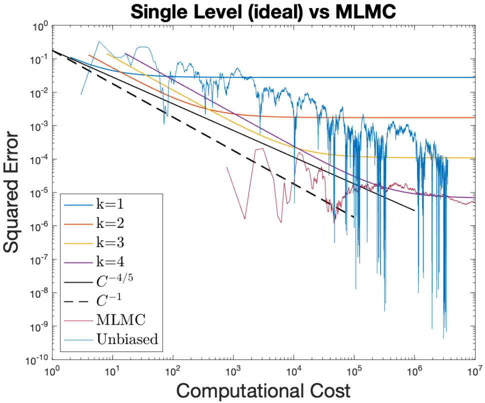
$$\mathbb{E}[Z^2] = \sum_{k=0}^{\infty} \frac{\mathbb{E}[\Delta_k(\theta)]^2}{p_k}$$

The expected cost of one evaluation of Z is approximately $\sum_{k=0}^{\infty} C_k p_k$.

- Finding optimal $\{p_k\}_{k=0}^{\infty}$ is hard
- Rough heuristic: set $p_k \propto 2^{-rk}$, where $1 < r < 2\alpha$
- Geometric distribution—p_k is cheap to sample.

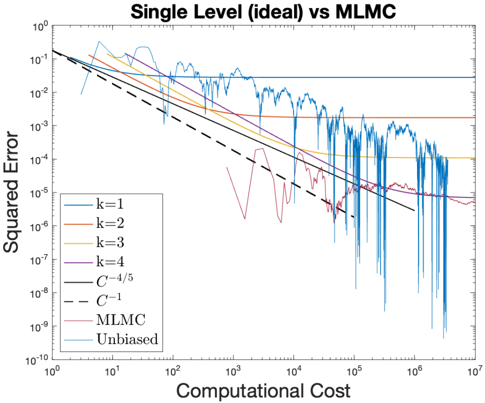
Unbiased Estimator

Implementation with $r = 2.5$:



Unbiased Estimator

Implementation with $r = 2.5$:



...possible implementation issues, or choice of parameter? Needs closer examination.

Future Considerations

- How to choose the number of levels K and sample sizes N_k in practice?
  - We could analyze our toy problem with pencil and paper, but real problems are much more complicated.

Conclusion

Thanks for listening!

For Further Reading I

Michael B. Giles
Multilevel Monte Carlo Methods
https://people.maths.ox.ac.uk/gilesm/files/acta15.pdf

Chang-Han Rhee and Peter W. Glynn
Unbiased Estimation with Square Root Convergence for SDE Models
https://chrhee.github.io/papers/RheeGlynn13a.pdf

Michael B. Giles
MLMC for Nested Expectations
https://people.maths.ox.ac.uk/gilesm/files/SLOAN80-056.pdf