Clustered Multilevel Monte Carlo (C-MLMC)

May 31, 2025

1 Problem Setup and Notation

We aim to estimate $\mathbb{E}[P]$ for a path-dependent payoff P under an SDE model. Using the Multilevel Monte Carlo (MLMC) framework, we define level-wise differences:

$$\Delta P_\ell = \begin{cases} P_0, & \text{if } \ell = 0, \\ P_\ell - P_{\ell-1}, & \text{if } \ell \geq 1. \end{cases}$$

Let $V_\ell = \mathrm{Var}[\Delta P_\ell]$, and C_ℓ be the average cost per sample on level ℓ. Classical MLMC chooses:

$$N_\ell \propto \sqrt{\frac{V_\ell}{C_\ell}}$$

to meet the MSE constraint:

$$\sum_{\ell=0}^{L} \frac{V_\ell}{N_\ell} \leq (1-\theta)\varepsilon^2.$$

2 Motivation for Clustering

When the variance is concentrated in rare modes of the input space, uniform sampling becomes inefficient. C-MLMC improves efficiency by:

- Extracting features from simulation details,
- Clustering samples via K-means into n_c variance modes,
- Estimating per-cluster statistics $(P_{\ell,c}, V_{\ell,c})$,
- Allocating samples proportional to $\sqrt{P_{\ell,c} V_{\ell,c}}$.

3 C-MLMC Algorithm

Algorithm 1 Clustered MLMC Estimator

Require: Simulator $f(\ell, N, \texttt{return\_details})$, feature extractor φ, tolerance ε, pilot size N_0, clusters n_c

Initialize level count $L \leftarrow L_{\min}$, precision control $\theta \in (0, 1)$
for each level $\ell \leq L$ **do**
  Run pilot: simulate $\{(Y_i, d_i)\}_{i=1}^{N_0} \leftarrow f(\ell, N_0, \texttt{true})$
  Extract features: $\phi_i \leftarrow \varphi(d_i)$
  Cluster $\{\phi_i\}$ into n_c clusters using K-means
  Compute:

$$P_{\ell,c} = \frac{\#\{\text{samples in cluster } c\}}{N_0}, \quad V_{\ell,c} = \mathrm{Var}\left(Y_i \mid \phi_i \in \text{cluster } c\right)$$

  Store cluster-wise sums $S_{\ell,c} = \sum Y_i$, and counts $N_{\ell,c}$
end for
while not converged **do**
  Aggregate level stats: $V_\ell = \sum_{c=1}^{n_c} P_{\ell,c} V_{\ell,c}$, $C_\ell = \frac{\text{total cost}}{N_\ell}$
  Allocate level-wise samples: $N_\ell \propto \sqrt{V_\ell / C_\ell}$
  Allocate cluster-wise samples:

$$N_{\ell,c} \propto \sqrt{P_{\ell,c} V_{\ell,c}} \quad \text{s.t.} \quad \sum_c N_{\ell,c} = N_\ell$$

  for each level $\ell > 0$, cluster c **do**
    Sample until $N_{\ell,c}$ accepted samples with cluster label c
    Accumulate cluster sum $S_{\ell,c} \leftarrow S_{\ell,c} + \sum Y_i$
    Update counts $N_{\ell,c} \leftarrow N_{\ell,c} + \#\text{new samples}$
  end for
  Bias check: extrapolate tail bias using $\widehat{\mathbb{E}}[\Delta P_\ell] \sim \mathcal{O}(2^{-\alpha\ell})$; if large, increment L
end while
Final estimator:

$$\hat{P} = \frac{1}{N_0} \sum_{i=1}^{N_0} Y_{0,i} + \sum_{\ell=1}^{L} \sum_{c=1}^{n_c} \frac{S_{\ell,c}}{N_{\ell,c}}$$

4 Optimal Allocation via $\sqrt{P_c V_c}$

To minimize the estimator variance:

$$\mathrm{Var}(\hat{Y}) = \sum_{c=1}^{n_c} \frac{P_c^2 V_c}{N_c}, \quad \text{subject to} \quad \sum_{c=1}^{n_c} N_c = N,$$

the optimal allocation is:

$$N_c \propto P_c \sqrt{V_c} \quad \Rightarrow \quad \text{weights} \propto \sqrt{P_c V_c}.$$

5 Estimator Properties

Each cluster-wise estimator

$$\hat{Y}_{\ell,c} = \frac{S_{\ell,c}}{N_{\ell,c}}$$

is unbiased for its conditional mean, and the full estimator

$$\hat{P} = \sum_{\ell=0}^{L} \hat{Y}_\ell, \quad \hat{Y}_\ell = \begin{cases} \frac{1}{N_0} \sum_{i=1}^{N_0} Y_{0,i}, & \ell = 0, \\ \sum_{c=1}^{n_c} \frac{S_{\ell,c}}{N_{\ell,c}}, & \ell \geq 1 \end{cases}$$

is an unbiased estimator of $\mathbb{E}[P_L]$. Bias $\mathbb{E}[P - P_L]$ is controlled via extrapolation on the last levels.

6 Conclusion

C-MLMC enhances standard MLMC by stratifying variance-heavy simulations via clustering and allocating samples proportionally to $\sqrt{P_c V_c}$. This stratified estimator preserves unbiasedness while improving computational efficiency in high-variance regimes.