Clustered Multilevel Monte Carlo (C-MLMC)

June 16, 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 proportionally to $\sqrt{P_{\ell,c} V_{\ell,c}}$.

Algorithm 1 Stratified Clustered MLMC Estimator

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

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

$$P_{\ell,c} = \frac{\sum_{i=1}^{N_0} \mathbb{1}_{\{\phi_i \in c\}}}{N_0}, \quad V_{\ell,c} = \mathrm{Var}\left(Y_i \mid \phi_i \in c\right)$$

  else
    Estimate global variance: $V_0 = \mathrm{Var}(Y_i)$
  end if
end for
while not converged **do**
  for each level $\ell > 0$ **do**
    Compute per-cluster sample allocations:

$$N_{\ell,c} \propto P_{\ell,c}\sqrt{V_{\ell,c}}, \quad \text{and normalize such that } \sum_{c=1}^{n_c} N_{\ell,c} = N_\ell$$

    Estimate cost per sample $C_\ell \leftarrow$ `estimated from runtime or set a priori`
    Compute aggregated variance: $V_\ell = \sum_{c=1}^{n_c} \frac{P_{\ell,c}^2 V_{\ell,c}}{N_{\ell,c}}$
  end for
  Compute total variance target:

$$\sum_{\ell=0}^{L} V_\ell \le \theta^2 \varepsilon^2$$

  For level $\ell = 0$: increase N_0 if needed and update V_0
  Estimate total bias from finest levels (e.g., regression fit $|\mathbb{E}[\Delta P_\ell]| \sim \mathcal{O}(2^{-\alpha\ell})$)
  if estimated bias $> (1-\theta)^2\varepsilon^2$ **then**
    Increase level count: $L \leftarrow L + 1$, and perform pilot on new level
  else
    break
  end if
end while
for each level ℓ **do**
  if $\ell > 0$ **then**
    for each cluster c **do**
      Sample $N_{\ell,c}$ new paths with cluster label c
      Compute sample mean: $\hat{\mu}_{\ell,c} = \frac{1}{N_{\ell,c}} \sum Y_i^{(c)}$
    end for
    Compute level mean:

$$\hat{\mu}_\ell = \sum_{c=1}^{n_c} P_{\ell,c}\hat{\mu}_{\ell,c}$$

  else
    Estimate level mean: $\hat{\mu}_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} Y_i$
  end if
end for
Return: Estimator $\hat{P} = \sum_{\ell=0}^{L} \hat{\mu}_\ell$