Multi-level Monte Carlo for Reliability Theory

Louis J. M. Aslett (aslett@stats.ox.ac.uk), Tigran Nagapetyan, and Sebastian P. Vollmer
Department of Statistics, University of Oxford

i-like.org.uk

UNIVERSITY OF OXFORD

1. Introduction

Analysing the reliability of large and complex engineered systems can be computationally challenging, especially in the context of many differing types of component.

In this poster we demonstrate a natural mapping of MLMC onto the simplest possible reliability problem of estimating a functional of expected system lifetime, providing orders of magnitude speedup compared to textbook brute-force approaches.

This points to the potential for wider use of multi-level methods throughout reliability theory.

3. MLMC (see Giles, 2015)

Consider a sequence of estimators $T_0, T_1, \dots,$ which approximates T_L with increasing accuracy, but also increasing cost. By linearity of expectation,

$$\mathbb{E}[T_L] = \mathbb{E}[T_0] + \sum_{\ell=1}^{L} \mathbb{E}[T_\ell - T_{\ell-1}],$$

and therefore we can use the following unbiased estimator for $\mathbb{E}[T_L]$,

$$\frac{1}{N_0}\sum_{n=1}^{N_0} T_0^{(0,n)} + \sum_{\ell=1}^{L}\left\{\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left(T_\ell^{(\ell,n)} - T_{\ell-1}^{(\ell,n)}\right)\right\}$$

Level ℓ in the superscript (ℓ, n) indicates that samples used at each level are independent, but crucially the differences use common samples. Note, 'correction' since each T_ℓ generally *not* unbiased estimate.

Let $\sigma_0^2, \sigma_\ell^2$ and $\text{cost}_0, \text{cost}_\ell$ be the variance and expected cost of one sample of $T_0, T_\ell - T_{\ell-1}$ respectively. Then, overall for multi-level:

$$\text{cost}_{\text{MLMC}} = \sum_{\ell=0}^{L} N_\ell \cdot \text{cost}_\ell$$

$$\sigma^2_{\text{MLMC}} = \sum_{\ell=0}^{L} N_\ell^{-1} \cdot \sigma_\ell^2$$

$\therefore$ for accuracy $\varepsilon > 0$, $\text{cost}_{\text{MLMC}}$ minimised when $N_\ell \propto \sigma_\ell / \sqrt{\text{cost}_\ell}$

$$\implies \text{cost}_{\text{MLMC}} = \varepsilon^{-2}\left(\sum_{\ell=0}^{L} \sigma_\ell \sqrt{\text{cost}_\ell}\right)$$

Provided cost increases slower than variance decreases, can achieve savings.

4. `mlmc` R package

The `mlmc` R package (Aslett *et al.*, 2016) provides an easy to use interface which automates much of the MLMC estimation process.

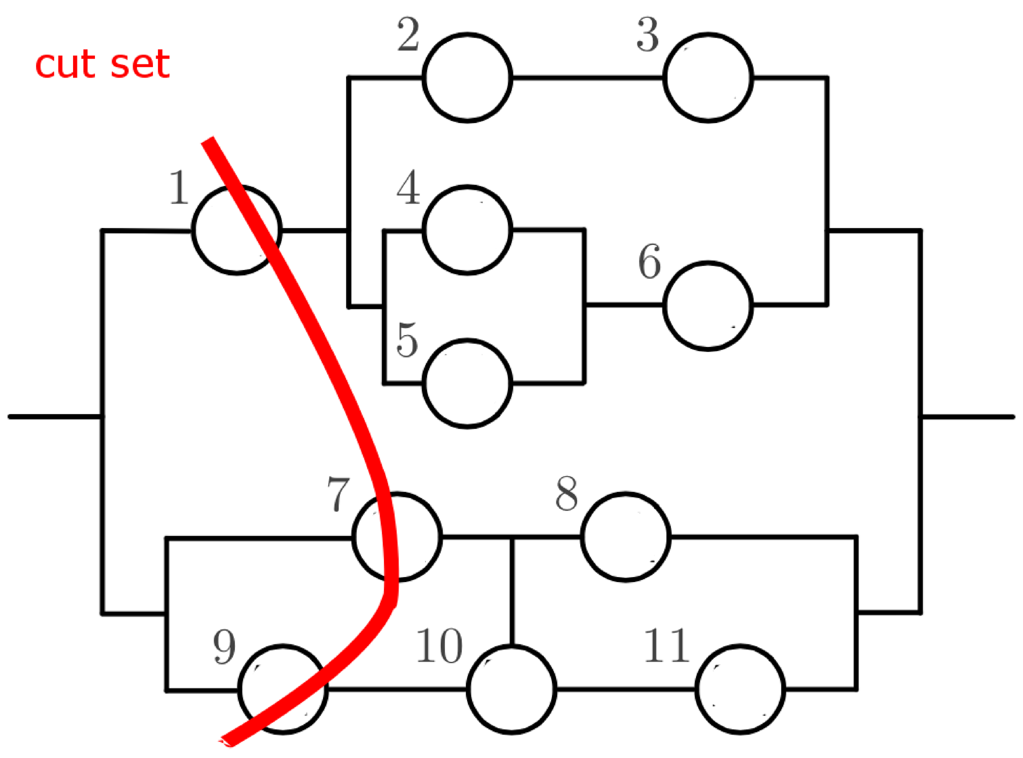
User simply needs to define a level sampler which is provided to the `mlmc()` function.

All standard graphical diagnostics easily plotted via overloaded `plot()` function on MLMC result object.

References

Aslett, L. J. M., Giles, M. B., Nagapetyan, T. and Vollmer, S. J. (2016), *mlmc: Tools for Multilevel Monte Carlo*. R package.
URL: *https://github.com/louisaslett/mlmc*
Barlow, R. E. and Proschan, F. (1981), *Statistical Theory of Reliability and Life Testing*, To Begin With Press.
Giles, M. (2015), \`Multilevel Monte Carlo methods', *Acta Numerica* **24**, 259--328.

2. System lifetime simulation



Cut sets: A set of components, C, is a *cut set* of the system if the system is failed whenever all the components in C are failed.

A cut set is said to be a *minimal cut set* if no subset of it is also a cut set.

Set of all minimal cut sets, $\mathcal{C}$, characterises the operational state of a system completely.

System lifetime: (Barlow and Proschan, 1981)

$$T_S = \min_{C \in \mathcal{C}} \left\{ \max_{c \in C}\{T_c\} \right\}.$$

Thus, the failure time for the system depends on the system structure (via $\mathcal{C}$) and the failure time distributions for each node.

Standard Monte Carlo:

$$\mathbb{E}[f(T_S)] \approx \hat{I}_n := \frac{1}{n}\sum_{i=1}^{n} f\left(\min_{C\in\mathcal{C}}\left\{\max_{c\in C}\{t_c^{(i)}\}\right\}\right)$$

where $t_c^{(i)} \sim F_c(\cdot)$. $\hat{I}_n \sim \text{N}(\mu, \sigma/\sqrt{n})$
$\therefore$ accuracy $\varepsilon > 0$ with α% confidence requires $n = z^2_{\alpha/2}\sigma^2\varepsilon^{-2} \implies \text{cost}_{\text{MC}} = \sigma^2 \cdot \varepsilon^{-2} \cdot |\mathcal{C}|$

$\therefore$ runtime depends on:

1. variance of the estimator;
2. target accuracy of the estimate;
3. number of cut sets.

Cheat? Use subset $\mathcal{C}' \subset \mathcal{C} \implies$

$$\min_{C\in\mathcal{C}'}\left\{\max_{i\in C}\{t_i\}\right\} = T'_S \geq T_S = \min_{C\in\mathcal{C}}\left\{\max_{i\in C}\{t_i\}\right\}.$$

But, $\hat{I}_n \to \eta \neq \mu$, and can only control variance

$$\mathbb{E}[(\hat{I}'_n - \mu)^2] = \frac{\sigma^2}{n} + (\eta - \mu)^2 = \text{var} + \text{bias}^2$$

5. Level selection in reliability problems

Can use coarse estimate idea in (2) with MLMC?

Sequence of estimators $T_0, \dots, T_L$ based on a nested sequence of minimal cutsets,

$$\mathcal{C}_0 \subset \cdots \subset \mathcal{C}_L = \mathcal{C}.$$

Note $T_L \equiv T_S$, which is not typically true in a general MLMC setting.

Need

1. geometric increase in cost;
2. geometric decrease in variance;
3. geometric decay in differences;
4. $\ell = 0$ should be cheap.

Cost: aim for doubling of min cutset collections ($\text{cost}_\ell = |\mathcal{C}_\ell|$), e.g.

$$|\mathcal{C}_0| = 8, \dots, |\mathcal{C}_5| = 250, |\mathcal{C}_6| = 500, |\mathcal{C}_7| = 1000$$

Prespecify these target sizes and select cutsets.

Level 0: Presimulate 100 component failures, take expectation and sort cutsets. Choose the $|\mathcal{C}_0|$ smallest.

Other levels: Continued selection based on sorted expectation works poorly.
We really want

$$\mathbb{E}[T_\ell - T_{\ell-1}] > \mathbb{E}[T_{\ell+1} - T_\ell]$$

i.e. given $\mathcal{C}_0, \dots, \mathcal{C}_{\ell-1}$ want level ℓ st $\mathbb{E}[T_{\ell-1} - T_\ell]$ is maximal. Note,

$$\mathbb{E}[T_{\ell-1} - T_\ell] \leq \mathbb{E}\left[T_{\ell-1} - \min\left\{T_{\ell-1}, \max_{C\in\mathcal{C}\setminus\mathcal{C}_{\ell-1}} C(\underline{T})\right\}\right]$$

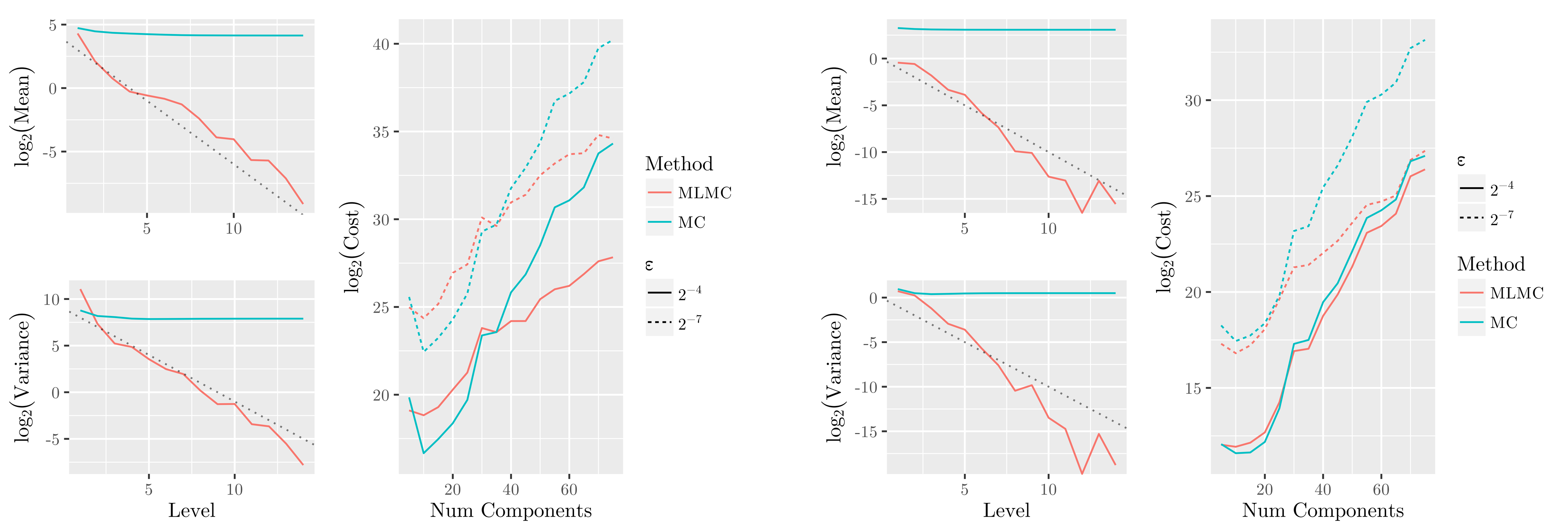
where $C(\underline{T})$ is cutset failure time.
$\therefore$ use the 100 presimulations to Monte Carlo estimate:

$$\mathbb{E}\left[T_{\ell-1} - \min\{T_{\ell-1}, C_i(\underline{T})\}\right] \quad \forall i \text{ st } C_i \in \mathcal{C}\setminus\mathcal{C}_{\ell-1}$$

and sort using this measure. Select smallest $|\mathcal{C}_\ell| - |\mathcal{C}_{\ell-1}|$.

6. Results

Left pairs: Diagnostic tests for largest system; Right: cost gains for nested randomly grown systems. Components Weibull with shape $\beta = 0.5$ (left), $\beta = 3$ (right) and uniformly distributed scale.



Funding

Louis Aslett is supported by EPSRC programme grant EP/K014463/1 (`www.i-like.org.uk`)
Tigran Nagapetyan and Sebastian Vollmer are supported by EPSRC grant EP/N000188/1