Multi-level Monte Carlo for System Reliability Simulation

Tigran Nagapetyan, **Louis J. M. Aslett**
and Sebastian J. Vollmer

Department of Statistics, University of Oxford

London Mathematical Society
Mathematical Methods in Reliability
Durham University, 13 April 2016

i-like.org.uk

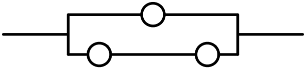
Outline

1. Introduction
    - Simulating system lifetimes
    - Recap standard Monte Carlo methods
2. Multi-level Monte Carlo (MLMC)
    - Introduction to MLMC for reliability audience
3. MLMC for reliability
    - Level grouping criterion
    - Mean and variance decay properties
    - Computational advantage
4. Future work

Introduction

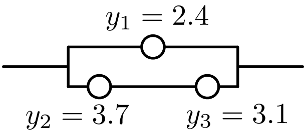
Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...



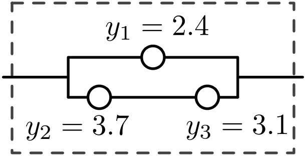
Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...



Simulating system lifetimes

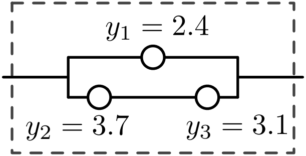
Simple approach to simulating system lifetimes in a system with $K = 3$ components ...



$t_S = 3.1$

Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...

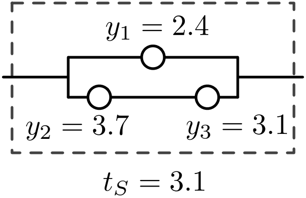


$$t_S = 3.1$$

$$\mathcal{C} = \{\{1, 2\}, \{1, 3\}\}$$

Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...

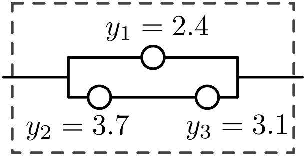


$$\mathcal{C} = \{\{1,2\},\{1,3\}\}$$

$$t_S = \min_{C \in \mathcal{C}} \{ \max_{i \in C} \{y_i\}\}$$

Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...



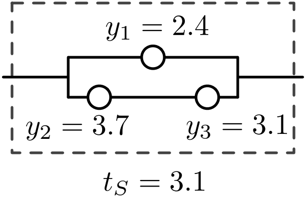
$t_S = 3.1$

$$\mathcal{C} = \{\{1, 2\}, \{1, 3\}\} \qquad \mathcal{P} = \{\{1\}, \{2, 3\}\}$$

$$t_S = \min_{C \in \mathcal{C}} \left\{ \max_{i \in C} \{y_i\} \right\}$$

Simulating system lifetimes

Simple approach to simulating system lifetimes in a system with $K = 3$ components ...



$$\mathcal{C} = \{\{1,2\},\{1,3\}\} \qquad \mathcal{P} = \{\{1\},\{2,3\}\}$$

$$t_S = \min_{C \in \mathcal{C}} \left\{ \max_{i \in C} \{y_i\} \right\} \qquad t_S = \max_{P \in \mathcal{P}} \left\{ \min_{i \in P} \{y_i\} \right\}$$

Monte Carlo simulation (I)

To estimate the expectation of some functional of the system lifetime, $\mu = \mathbb{E}[f(T_S)]$, simply perform Monte Carlo simulation:

$$\mathbb{E}[f(T_S)] \approx \hat{I}_n \triangleq \frac{1}{n}\sum_{j=1}^{n} f(T_S^{(j)})$$
$$= \frac{1}{n}\sum_{j=1}^{n} f\Big(\min_{C\in\mathcal{C}}\{\max_{i\in C}\{y_i^{(j)}\}\}\Big)$$

Monte Carlo simulation (I)

To estimate the expectation of some functional of the system lifetime, $\mu = \mathbb{E}[f(T_S)]$, simply perform Monte Carlo simulation:

$$\mathbb{E}[f(T_S)] \approx \hat{I}_n \triangleq \frac{1}{n}\sum_{j=1}^{n} f(T_S^{(j)})$$
$$= \frac{1}{n}\sum_{j=1}^{n} f\Big(\min_{C\in\mathcal{C}}\{\max_{i\in C}\{y_i^{(j)}\}\}\Big)$$

We know that,

$$\mathbb{P}\left(|\hat{I}_n - \mu| > z\frac{\sigma}{\sqrt{n}}\right) \approx \mathbb{P}(|Z| > z)$$

for $Z \sim \mathrm{N}(0,1)$, with $\hat{I}_n$ an unbiased estimate of μ.

Monte Carlo simulation (II)

Thus, for a desired level of accuracy $\varepsilon > 0$ with $\alpha\%$ confidence, we require

$$n = z_{\alpha/2}^2 \sigma^2 \varepsilon^{-2}$$

Monte Carlo simulations, where $z_{\alpha/2}$ is such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$.

Monte Carlo simulation (II)

Thus, for a desired level of accuracy $\varepsilon > 0$ with $\alpha\%$ confidence, we require

$$n = z_{\alpha/2}^2 \sigma^2 \varepsilon^{-2}$$

Monte Carlo simulations, where $z_{\alpha/2}$ is such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$.

$z_{\alpha/2}$ constant for a given confidence, so the variable compute cost in simulation can be defined as

$$\text{Cost}_{MC} = \sigma^2 \varepsilon^{-2} |\mathcal{C}|$$

Costly for:

- High accuracy (small ε)
- Large systems (many cutsets)
- Large system lifetime variance

Trying to cheat ...

May want to try getting a coarser estimate.

$$\mathcal{C}' \subset \mathcal{C} \implies \min_{C \in \mathcal{C}'} \{ \max_{i \in C} \{y_i\} \} = t'_S \geq t_S = \min_{C \in \mathcal{C}} \{ \max_{i \in C} \{y_i\} \}$$

But now, $\hat{I}'_n$ is a biased estimate of μ. Let $\hat{I}'_n \to \eta$, then:

$$\begin{aligned} \mathbb{E}\left[(\hat{I}'_n - \mu)^2\right] &= \mathbb{E}\left[(\hat{I}'_n - \eta + \eta - \mu)^2\right] \\ &= \mathbb{E}\left[(\hat{I}'_n - \eta)^2\right] + (\eta - \mu)^2 \\ &= \frac{\sigma^2}{n} + (\eta - \mu)^2 \end{aligned}$$

so that the error is composed of contributions from both the coarse approximation variance and the bias in the estimate.

Multi-level Monte Carlo (MLMC)

Multi-level Monte Carlo (MLMC)

MLMC grew out of stochastic partial differential equation applications where simulations are so expensive that performing enough to reduce Monte Carlo variance to acceptable levels was impractical.

MLMC, pioneered by Stefan Heinrich (TU Kaiserslautern) (Heinrich 1998) and Mike Giles (Oxford) (Giles 2008), combines simulations at different *levels* of approximation to achieve the same accuracy ε with far lower computational cost.

Multi-level Monte Carlo (MLMC)

MLMC grew out of stochastic partial differential equation applications where simulations are so expensive that performing enough to reduce Monte Carlo variance to acceptable levels was impractical.

MLMC, pioneered by Stefan Heinrich (TU Kaiserslautern) (Heinrich 1998) and Mike Giles (Oxford) (Giles 2008), combines simulations at different *levels* of approximation to achieve the same accuracy ε with far lower computational cost.

Can we use MLMC in combination with an approximation arising from evaluation of subsets of the minimal cutsets?

Lightning introduction to MLMC (I)

Assume have $L+1$ *levels* of accuracy with which we can simulate system failure time, $T_0, \dots, T_L \equiv T_S$, with level L being equivalent to standard Monte Carlo[1].

Here, $T_0, \dots, T_L$ is the estimate based on a nested sequence of cutsets, $\mathcal{C}_0 \subset \cdots \subset \mathcal{C}_L = \mathcal{C}$.

[1] It is typically not the case that the highest accuracy level is unbiased in most MLMC applications

Lightning introduction to MLMC (I)

Assume have $L+1$ *levels* of accuracy with which we can simulate system failure time, $T_0, \dots, T_L \equiv T_S$, with level L being equivalent to standard Monte Carlo[1].

Here, $T_0, \dots, T_L$ is the estimate based on a nested sequence of cutsets, $\mathcal{C}_0 \subset \cdots \subset \mathcal{C}_L = \mathcal{C}$.

Then, a telescoping sum can be formed:

$$\mathbb{E}[T_S] \equiv \mathbb{E}[T_0] + \sum_{l=1}^{L} \mathbb{E}[T_l - T_{l-1}]$$

[1] It is typically not the case that the highest accuracy level is unbiased in most MLMC applications

Lightning introduction to MLMC (I)

Assume have $L+1$ *levels* of accuracy with which we can simulate system failure time, $T_0, \ldots, T_L \equiv T_S$, with level L being equivalent to standard Monte Carlo[1].

Here, $T_0, \ldots, T_L$ is the estimate based on a nested sequence of cutsets, $\mathcal{C}_0 \subset \cdots \subset \mathcal{C}_L = \mathcal{C}$.

Then, a telescoping sum can be formed:

$$\mathbb{E}[T_S] \equiv \mathbb{E}[T_0] + \sum_{l=1}^{L} \mathbb{E}[T_l - T_{l-1}]$$

In a nut shell, this identity provides an estimator with same expected value as standard Monte Carlo ... but, for a fixed variance (fixed accuracy) has much lower computational cost.

[1] It is typically not the case that the highest accuracy level is unbiased in most MLMC applications

Lightning introduction to MLMC (II)

Independently estimate each term. Crucially, within each term, T_l and T_{l-1} use the same random component simulations:

$$\mathbb{E}[T_l - T_{l-1}] \approx N_l^{-1} \sum_{j=1}^{N_l} \left(t_l^{(j)} - t_{l-1}^{(j)} \right)$$

with each level having cost $\propto \sigma_l^2 |\mathcal{C}_l|$.

Lightning introduction to MLMC (II)

Independently estimate each term. Crucially, within each term, T_l and T_{l-1} use the same random component simulations:

$$\mathbb{E}[T_l - T_{l-1}] \approx N_l^{-1} \sum_{j=1}^{N_l} \left(t_l^{(j)} - t_{l-1}^{(j)} \right)$$

with each level having cost $\propto \sigma_l^2 |\mathcal{C}_l|$.
The overall MLMC variance is then

$$\text{Var}\left(\sum_{l=0}^{L} \mathbb{E}[T_l - T_{l-1}] \right) = \sum_{l=0}^{L} N_l^{-1} \sigma_l^2$$

at a cost of

$$\text{Cost}_{ML} = \sum_{l=0}^{L} N_l |\mathcal{C}_l|$$

Lightning introduction to MLMC (II)

Independently estimate each term. Crucially, within each term, T_l and T_{l-1} use the same random component simulations:

$$\mathbb{E}[T_l - T_{l-1}] \approx N_l^{-1} \sum_{j=1}^{N_l} \left(t_l^{(j)} - t_{l-1}^{(j)} \right)$$

with each level having cost $\propto \sigma_l^2 |\mathcal{C}_l|$.
The overall MLMC variance is then

$$\mathrm{Var}\left(\sum_{l=0}^{L} \mathbb{E}[T_l - T_{l-1}] \right) = \sum_{l=0}^{L} N_l^{-1} \sigma_l^2$$

at a cost of

$$\mathrm{Cost}_{ML} = \sum_{l=0}^{L} N_l |\mathcal{C}_l|$$

Hence, given a target fixed variance (accuracy), taking for each level $N_l \propto \frac{\sigma_l}{\sqrt{|\mathcal{C}_l|}}$ will minimise the computational cost.

Lightning introduction to MLMC (III)

For a desired accuracy $\varepsilon > 0$, this leads to an overall cost:

$$\begin{aligned}\text{Cost}_{ML} &= \sum_{l=0}^{L} N_l |\mathcal{C}_l| \\ &= \varepsilon^{-2} \left(\sum_{l=0}^{L} \sigma_l \sqrt{|\mathcal{C}_l|} \right)^2\end{aligned}$$

Lightning introduction to MLMC (III)

For a desired accuracy $\varepsilon > 0$, this leads to an overall cost:

$$\begin{aligned}\text{Cost}_{ML} &= \sum_{l=0}^{L} N_l |\mathcal{C}_l| \\ &= \varepsilon^{-2} \left(\sum_{l=0}^{L} \sigma_l \sqrt{|\mathcal{C}_l|} \right)^2\end{aligned}$$

Recall,

$$\text{Cost}_{MC} = \sigma^2 \varepsilon^{-2} |\mathcal{C}| \triangleq \sigma^2 \varepsilon^{-2} |\mathcal{C}_L|$$

So, σ_l must decay in order for MLMC to beat MC.

See Giles (2015) for deeper review.

MLMC for reliability

MLMC for reliability

Recall, our proposed application to reliability involves a nested sequence of cutsets providing improving accuracy:

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

How do we choose?

MLMC for reliability

Recall, our proposed application to reliability involves a nested sequence of cutsets providing improving accuracy:

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

How do we choose?

- It has been proven that anything less than a geometric increase in cost moving up the levels is suboptimal.

MLMC for reliability

Recall, our proposed application to reliability involves a nested sequence of cutsets providing improving accuracy:

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

How do we choose?

- It has been proven that anything less than a geometric increase in cost moving up the levels is suboptimal.
- The variance must likewise decay rapidly (geometrically) to ensure that the overall MLMC cost beats MC.

MLMC for reliability

Recall, our proposed application to reliability involves a nested sequence of cutsets providing improving accuracy:

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

How do we choose?

- It has been proven that anything less than a geometric increase in cost moving up the levels is suboptimal.
- The variance must likewise decay rapidly (geometrically) to ensure that the overall MLMC cost beats MC.
- The mean of each level decaying rapidly is desirable so that each additional term has little influence.

MLMC for reliability

Recall, our proposed application to reliability involves a nested sequence of cutsets providing improving accuracy:

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

How do we choose?

- It has been proven that anything less than a geometric increase in cost moving up the levels is suboptimal.
- The variance must likewise decay rapidly (geometrically) to ensure that the overall MLMC cost beats MC.
- The mean of each level decaying rapidly is desirable so that each additional term has little influence.
- By the same argument, we want level 0 to be the best estimate possible since it will have smallest cost and be repeated most.

Setup

Some notation:

- Let $\mathcal{C} \triangleq \{C_1, \ldots, C_M\}$.
- Let there be K components, with lifetimes $\underline{\tau} = (\tau_1, \ldots, \tau_K)$.
- Let $C_i(\underline{\tau}) \triangleq \max_{j \in C_i}\{\tau_j\}$.
- Bracket subscript for some specified order statistics, $C_{(1)}$.

Setup

Some notation:

- Let $\mathcal{C} \triangleq \{C_1, \dots, C_M\}$.
- Let there be K components, with lifetimes $\underline{\tau} = (\tau_1, \dots, \tau_K)$.
- Let $C_i(\underline{\tau}) \triangleq \max_{j \in C_i} \{\tau_j\}$.
- Bracket subscript for some specified order statistics, $C_{(1)}$.

To achieve geometric growth in cost we will aim for approx doubling of the number of cutsets in each level. For example, $M = 1000$, take:

$$
\begin{aligned}
|\mathcal{C}_7| &= 1000 \\
|\mathcal{C}_6| &= 500 \\
|\mathcal{C}_5| &= 250 \\
&\vdots \\
|\mathcal{C}_0| &= 8
\end{aligned}
$$

Selecting the levels, $l = 0$

1. Set $L = \max\{\lfloor \log_2 |\mathcal{C}| - 2 \rfloor, 3\}$

Selecting the levels, $l = 0$

1. Set $L = \max\{\lfloor \log_2 |\mathcal{C}| - 2 \rfloor, 3\}$
2. Simulate 100 failures for each of K components, $\underline{\tau}^{(1)}, \ldots, \underline{\tau}^{(100)}$.
3. Estimate

$$\mathbb{E}[C_i(\underline{\tau})] \approx 0.01 \sum_{j=1}^{100} C_i(\underline{\tau}^{(j)})$$

Selecting the levels, $l = 0$

1. Set $L = \max\{\lfloor \log_2 |\mathcal{C}| - 2 \rfloor, 3\}$
2. Simulate 100 failures for each of K components, $\underline{\mathcal{T}}^{(1)}, \dots, \underline{\mathcal{T}}^{(100)}$.
3. Estimate
$$\mathbb{E}[C_i(\underline{\mathcal{T}})] \approx 0.01 \sum_{j=1}^{100} C_i(\underline{\mathcal{T}}^{(j)})$$
4. Sort cutsets by order statistic of expected failure time, $(C_{(1)}, \dots, C_{(M)})$

Selecting the levels, $l = 0$

1. Set $L = \max\{\lfloor \log_2 |\mathcal{C}| - 2 \rfloor, 3\}$
2. Simulate 100 failures for each of K components, $\underline{\tau}^{(1)}, \dots, \underline{\tau}^{(100)}$.
3. Estimate
$$\mathbb{E}[C_i(\underline{\tau})] \approx 0.01 \sum_{j=1}^{100} C_i(\underline{\tau}^{(j)})$$
4. Sort cutsets by order statistic of expected failure time, $(C_{(1)}, \dots, C_{(M)})$
5. Set $\mathcal{C}_0 = \{C_{(1)}, \dots, C_{(m_0)}\}$ where $m_0 = \lceil 2^{-L}(|\mathcal{C}| - 1) + 1 \rceil$.

This provides a crude estimate of the most common failure cause cutsets.

Selecting the levels, $l > 0$ (I)

For the remaining levels, note that we want

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

Therefore, having chosen level $l - 1$, $\mathcal{C}_{l-1}$, want level l st

$$\mathbb{E}[T_{l-1} - T_l] \to \max$$

Selecting the levels, $l > 0$ (I)

For the remaining levels, note that we want

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

Therefore, having chosen level $l-1$, $\mathcal{C}_{l-1}$, want level l st

$$\mathbb{E}[T_{l-1} - T_l] \to \max$$

But,

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

We will attempt to crudely achieve this ordering using the 100 simulations already done.

Selecting the levels, $l > 0$ (II)

For remaining $l \in \{1, \ldots, L\}$

1. Reindex remaining cutsets in $\mathcal{C} \backslash \mathcal{C}_{l-1}$ from 1 to $M - m_{l-1}$.

Selecting the levels, $l > 0$ (II)

For remaining $l \in \{1, \dots, L\}$

1. Reindex remaining cutsets in $\mathcal{C} \backslash \mathcal{C}_{l-1}$ from 1 to $M - m_{l-1}$.
2. Compute

$$\delta_i = 0.01 \sum_{j=1}^{100} \left[T_{l-1} - \min \left\{ C_i(\underline{\tau}^{(j)}), T_{l-1} \right\} \right] \quad \forall\, i = 1, \dots, M - m_{l-1}$$

Selecting the levels, $l > 0$ (II)

For remaining $l \in \{1, \ldots, L\}$

1. Reindex remaining cutsets in $\mathcal{C} \backslash \mathcal{C}_{l-1}$ from 1 to $M - m_{l-1}$.
2. Compute

$$\delta_i = 0.01 \sum_{j=1}^{100} \left[T_{l-1} - \min \left\{ C_i(\underline{\tau}^{(j)}), T_{l-1} \right\} \right] \quad \forall\, i = 1, \ldots, M - m_{l-1}$$

3. Sort cutsets by order statistic of δ_i, $(C_{(1)}, \ldots, C_{(M-m_{l-1})})$

Selecting the levels, $l > 0$ (II)

For remaining $l \in \{1, \ldots, L\}$

1. Reindex remaining cutsets in $\mathcal{C} \backslash \mathcal{C}_{l-1}$ from 1 to $M - m_{l-1}$.
2. Compute

$$\delta_i = 0.01 \sum_{j=1}^{100} \left[T_{l-1} - \min \left\{ C_i(\underline{\tau}^{(j)}), T_{l-1} \right\} \right] \quad \forall\, i = 1, \ldots, M - m_{l-1}$$

3. Sort cutsets by order statistic of δ_i, $(C_{(1)}, \ldots, C_{(M-m_{l-1})})$
4. Set $\mathcal{C}_l = \mathcal{C}_{l-1} \cup \{C_{(1)}, \ldots, C_{(m_l)}\}$ where $m_l = \lceil 2^{-L+l}(|\mathcal{C}| - 1) + 1 \rceil - m_{l-1}$.

Running the MLMC (summary)

Finally, with the levels all selected, set a desired precision $\varepsilon > 0$ and proceed:

1. Initially set $N_l = 100$ for $l = 0, \ldots, 3$ and simulate these levels

Running the MLMC (summary)

Finally, with the levels all selected, set a desired precision $\varepsilon > 0$ and proceed:

1. Initially set $N_l = 100$ for $l = 0, \ldots, 3$ and simulate these levels
2. Estimate level variance

Running the MLMC (summary)

Finally, with the levels all selected, set a desired precision $\varepsilon > 0$ and proceed:

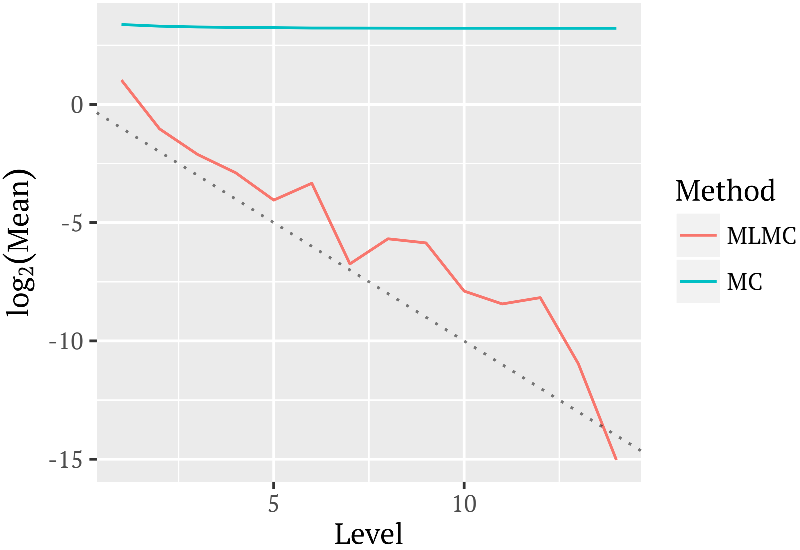
1. Initially set $N_l = 100$ for $l = 0, \ldots, 3$ and simulate these levels
2. Estimate level variance
3. Compute additional number of iterations, N_l, at each level to achieve ε precision. Repeat 2 until less than 1% growth in $N_l \; \forall \; l$.

Running the MLMC (summary)

Finally, with the levels all selected, set a desired precision $\varepsilon > 0$ and proceed:

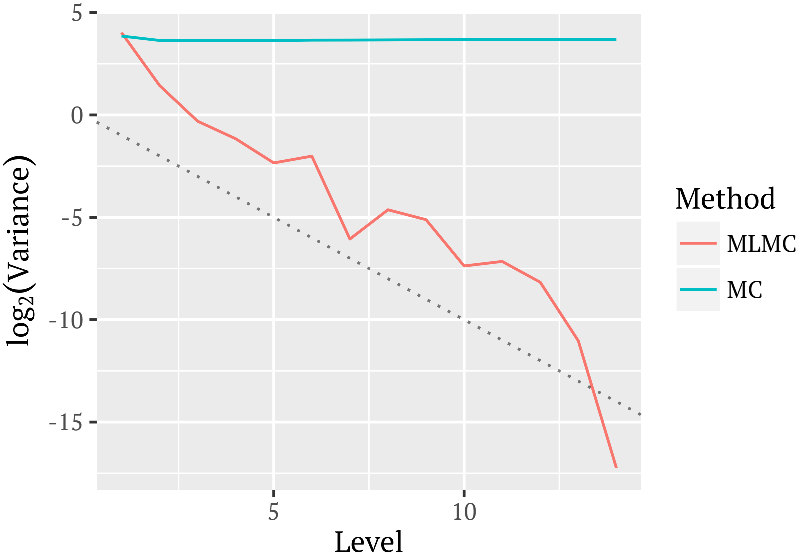
1. Initially set $N_l = 100$ for $l = 0, \ldots, 3$ and simulate these levels
2. Estimate level variance
3. Compute additional number of iterations, N_l, at each level to achieve ε precision. Repeat 2 until less than 1% growth in $N_l \; \forall \; l$.
4. Variance has converged, now test bias. If bias within tolerance end, otherwise add a new level and return to 2.

Mean plot — 75 different Weibull components



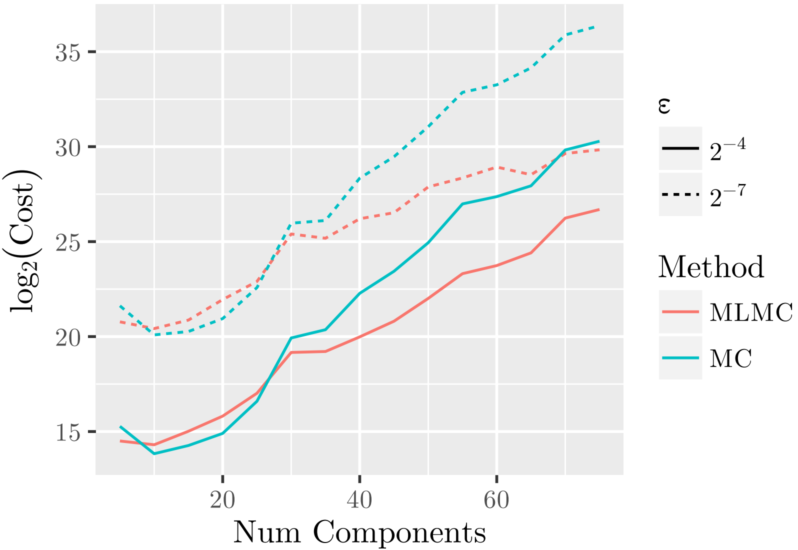
$|\mathcal{C}| = 293,101$

Variance plot — 75 different Weibull components



$|\mathcal{C}| = 293,101$

Cost plot — all different Weibull components



$K = 65 : |\mathcal{C}| = 58,564; K = 70 : |\mathcal{C}| = 224,365; K = 75 : |\mathcal{C}| = 293,101$

Future work

Open questions ...

Many possible directions:

- Estimation of the full system lifetime distribution (Giles 2015)
- Use MLMC with a survival signature based simulation in the independent/exchangeable case (?)
- Preserving privacy of component lifetimes — can MLMC provide enough efficiency for private simulation (?)

References I

Giles, M. B. (2008). Multilevel monte carlo path simulation. *Operations Research*, 56/3: 607–17. INFORMS.

Giles, M. B. (2015). Multilevel monte carlo methods. *Acta Numerica*, 24: 259–328. Cambridge Univ Press.

Giles, M. B., Nagapetyan, T., & Ritter, K. (2015). Multilevel monte carlo approximation of distribution functions and densities. *SIAM/ASA Journal on Uncertainty Quantification*, 3/1: 267–95. SIAM.

Heinrich, S. (1998). Monte carlo complexity of global solution of integral equations. *Journal of Complexity*, 14/2: 151–75. Elsevier.