Expected Value of Partial Perfect Information

Mike Giles [1], Takashi Goda [2], Howard Thom [3]
Wei Fang [1], Zhenru Wang [1]

[1]Mathematical Institute, University of Oxford

[2]School of Engineering, University of Tokyo

[3]School of Social and Community Medicine, University of Bristol

MCM International Conference on Monte Carlo Methods

July 4, 2017

Outline

- EVPI and EVPPI
- MLMC for nested simulation
- numerical analysis
- some numerical results
- using MCMC inputs
- conclusions

Decision making under uncertainty

The motivation comes from decision-making in funding medical research.

Models of the cost-effectiveness of medical treatments depend on various parameters.

The parameter values are not known precisely, and are instead modelled as random variables coming from some prior distribution.

The research question is whether it is worth conducting some medical research to eliminate uncertainty in some of the parameters.

Similar applications arise in oil reservoir recovery – is it worth sinking an additional exploratory oil well to learn more about the oilfield?

Decision making under uncertainty

Given no knowledge of independent uncertain parameters X, Y, best treatment out of some finite set D corresponds to

$$\max_{d \in D} \mathbb{E}\left[f_d(X,Y)\right]$$

while with perfect knowledge we would have

$$\mathbb{E}\left[\max_{d \in D} f_d(X,Y)\right].$$

However, if X is known but not Y, this is reduced to

$$\mathbb{E}\left[\max_d \mathbb{E}\left[f_d(X,Y) \,|\, X\right]\right]$$

giving a nested simulation problem.

EVPI & EVPPI

EVPI, the expected value of perfect information, is the difference

$$\text{EVPI} = \mathbb{E}\left[\max_d f_d(X,Y)\right] - \max_d \mathbb{E}[f_d(X,Y)]$$

which can be estimated with $O(\varepsilon^{-2})$ complexity by standard methods, assuming an $O(1)$ cost per sample $f_d(X,Y)$.

EVPPI, the expected value of partial perfect information, is the difference

$$\text{EVPPI} = \mathbb{E}\left[\max_d \mathbb{E}\left[f_d(X,Y)\,|\,X\right]\right] - \max_d \mathbb{E}[f_d(X,Y)].$$

In practice, we choose to estimate

$$\text{EVPI} - \text{EVPPI} = \mathbb{E}\left[\max_d f_d(X,Y)\right] - \mathbb{E}\left[\max_d \mathbb{E}\left[f_d(X,Y)\,|\,X\right]\right]$$

MLMC treatment

Based on work by Oxford colleagues (Bujok, Hambly, Reisinger, 2015) Takashi Goda (arXiv, April 2016) proposed an efficient MLMC estimator using 2^ℓ samples on level ℓ for conditional expectation.

For given sample X, define

$$Z_\ell \;=\; \tfrac{1}{2}\left(\max_d \overline{f_d}^{(a)} + \max_d \overline{f_d}^{(b)} \right) \;-\; \max_d \overline{f_d}$$

where

- $\overline{f_d}^{(a)}$ is an average of $f_d(X,Y)$ over $2^{\ell-1}$ independent samples for Y;
- $\overline{f_d}^{(b)}$ is an average over a second independent set of $2^{\ell-1}$ samples;
- $\overline{f_d}$ is an average over the combined set of 2^ℓ inner samples.

MLMC treatment

The expected value of this estimator is

$$\mathbb{E}[Z_\ell] = \mathbb{E}[\max_d \overline{f_d}_{,2^{\ell-1}}] - \mathbb{E}[\max_d \overline{f_d}_{,2^\ell}]$$

where $\overline{f_d}_{,2^\ell}$ is an average of 2^ℓ inner samples, and hence

$$\begin{aligned} \sum_{\ell=1}^{L} \mathbb{E}[Z_\ell] &= \mathbb{E}[\max_d f] - \mathbb{E}[\max_d \overline{f_d}_{,2^L}] \\ &\to \mathbb{E}[\max_d f] - \mathbb{E}[\max_d \mathbb{E}[f(X,Y)\,|\,X]] \end{aligned}$$

as $L \to \infty$, giving us the desired estimate.

MLMC treatment

How good is the estimator?

With reference to standard MLMC theorem, $\gamma = 1$, but what are α and β?

Define

$$F_d(X) = \mathbb{E}\left[f_d(X,Y) \,|\, X \right], \qquad d_{opt}(X) = \arg\max_d F_d(X)$$

so $d_{opt}(x)$ is piecewise constant, with a lower-dimensional manifold K on which it is not uniquely-defined.

Note that for any d, $\frac{1}{2}(\overline{f_d}^{(a)} + \overline{f_d}^{(b)}) - \overline{f_d} = 0$, so $Z_\ell = 0$ if the same d maximises each term in Z_ℓ.

Numerical analysis

Heuristic analysis:

- $\overline{f_d}^{(a)} - \overline{f_d}^{(b)} = O(2^{-\ell/2})$, due to CLT
- $O(2^{-\ell/2})$ probability of both being within distance $O(2^{-\ell/2})$ of K
- under this condition, $Z_\ell = O(2^{-\ell/2})$
- hence $\mathbb{E}[Z_\ell] = O(2^{-\ell})$ and $\mathbb{E}[Z_\ell^2] = O(2^{-3\ell/2})$, so $\alpha=1, \beta=3/2$.

It is possible to make this rigorous given some assumptions.

Numerical analysis

Assumptions

- $\mathbb{E}\left[\,|f_d(X,Y)|^p\right]$ *is finite for all* $p \geq 2$.

  Comment: helps to bound the difference between $\overline{f_d}$ *and* $F_d(X)$.

- *There exists a constant* $c_0 > 0$ *such that for all* $0 < \epsilon < 1$

$$\mathbb{P}(\min_{y\in K} \|X - y\| \leq \epsilon) \leq c_0\, \epsilon.$$

  Comment: bounds the probability of X *being close to* K.

- *There exist constants* $c_1, c_2 > 0$ *such that if* $X \notin K$, *then*

$$\max_d F_d(X) - \max_{d \neq d_{opt}(X)} F_d(X) \;>\; \min(c_1,\; c_2 \min_{y\in K} \|X - y\|).$$

  Comment: ensure linear separation of the optimal F_d *away from* K.

Numerical analysis

Building on the heuristic analysis, and past analyses (Giles & Szpruch, 2015), we obtain the following theorem:

Theorem

If the Assumptions are satisfied, and $\overline{f_d}^{(a)}$, $\overline{f_d}^{(b)}$, $\overline{f_d}$ are as defined previously for level ℓ, with 2^ℓ inner samples being used for $\overline{f_d}$, then for any $\delta > 0$

$$\mathbb{E}\left[\tfrac{1}{2}(\max_d \overline{f_d}^{(a)} + \max_d \overline{f_d}^{(b)}) - \max_d \overline{f_d}\right] = o(2^{-(1-\delta)\ell}).$$

and

$$\mathbb{E}\left[\left(\tfrac{1}{2}(\max_d \overline{f_d}^{(a)} + \max_d \overline{f_d}^{(b)}) - \max_d \overline{f_d}\right)^2\right] = o(2^{-(3/2-\delta)\ell}).$$

Supporting theory

Lemma

If X_i are iid scalar samples with zero mean, and $\overline{X}_N$ is average of N samples, then for $p \geq 2$, if $\mathbb{E}[|X|^p]$ is finite then there exists a constant C_p, depending only on p, such that

$$\begin{aligned} \mathbb{E}[\,|\overline{X}_N|^p] &\leq C_p\, N^{-p/2}\, \mathbb{E}[\,|X|^p] \\ \mathbb{P}[\,|\overline{X}_N| > c\,] &\leq C_p\, \mathbb{E}[\,|X|^p]\,/\,(c^2 N)^{p/2}. \end{aligned}$$

Proof.

Use discrete Burkholder-Davis-Gundy inequality applied to $\sum_i X_i$ and then Markov inequality. □

Supporting theory

Applying the result to the EVPPI problem, we get:

Lemma

If all moments of f_d are finite, then for $p \geq 2$, there exists a constant C_p, depending only on p, such that

$$\begin{aligned} \mathbb{E}[\,|\overline{f_d} - F_d|^p] &\leq C_p\, 2^{-p\ell/2}\, \mathbb{E}[\,|f_d - F_d|^p] \\ \mathbb{P}[\,|\overline{f_d} - F_d| > c\,] &\leq C_p\, \mathbb{E}[\,|f_d - F_d|^p] \,/\, (c^2 2^\ell)^{p/2}. \end{aligned}$$

Proof.

Conditional on X, we have $\mathbb{E}[\,|\overline{f_d} - F_d|^p \,|\, X] \leq C_p\, 2^{-p\ell/2}\, \mathbb{E}[\,|f_d - F_d|^p \,|\, X]$ then taking an outer expectation w.r.t. X gives the desired result.
A similar argument works for $\mathbb{P}[\,|\overline{f_d} - F_d| > c\,] \equiv \mathbb{E}[\,\mathbf{1}_{|\overline{f_d} - F_d| > c}\,]$ □

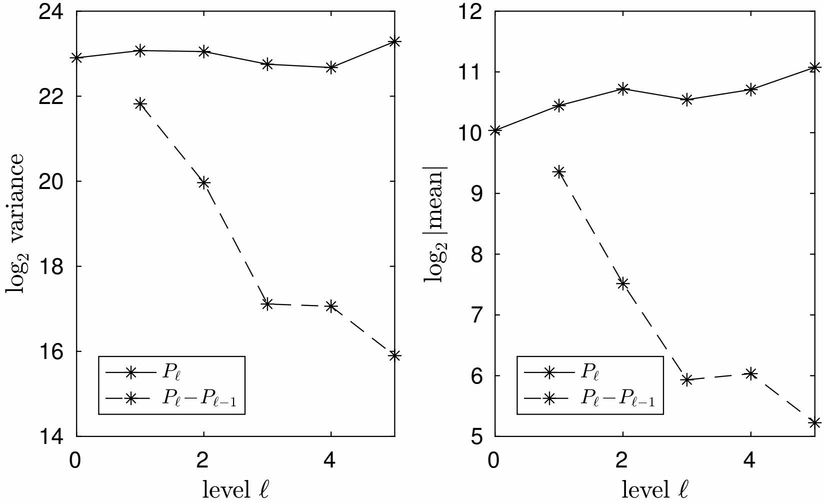
Numerical results

Real application: Cost-effectiveness of Noval Oral AntiCoagulants in Atrial Fibrillation

A choice of 6 different treatments, and a total of 99 input random variables:

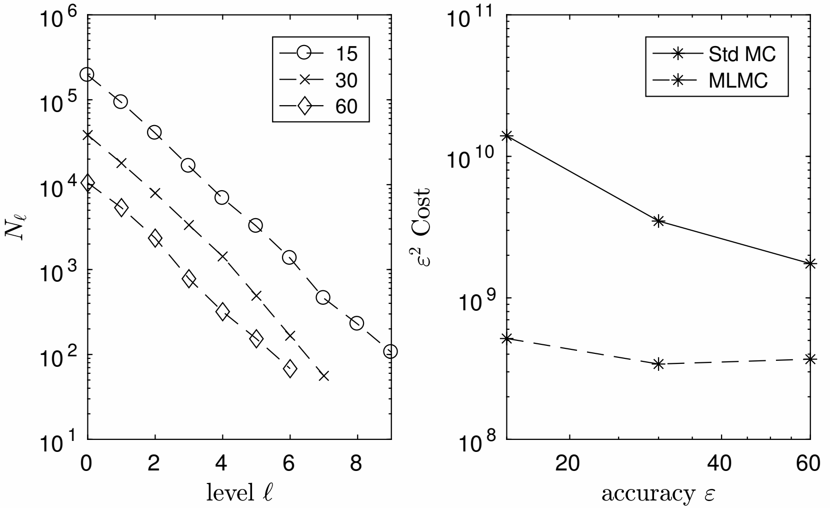
- 34 Normal
- 8 uniform
- 15 Beta
- 3 sets of MCMC inputs approximated by multivariate Normals of dimension 7, 7 and 28 corresponding to an additional 42 Normals

Numerical results



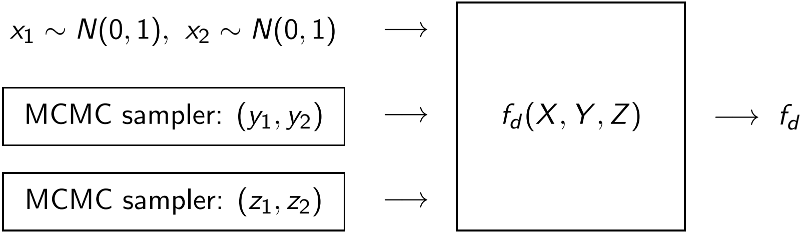
$\beta \approx 1.5$ $\alpha \approx 1.0$

Numerical results



MCMC inputs

A key new challenge is that some of the random inputs come from a Bayesian posterior distribution, with samples generated by MCMC methods.



The standard procedure (?) would be to generate MCMC samples on demand, and then use them immediately.

But how can this be extended to MLMC?

MCMC within MLMC

An alternative idea is to first pre-generate and store a large collection of MCMC samples:

$$\{Y_1, Y_2, Y_3, \ldots, Y_N\}$$

Then, random samples for Y can be obtained from this collection by uniformly distributed selection.

This avoids the problem of strong correlation between successive MCMC samples, and also works fine for the MLMC calculation.

This approach also leads to ideas on QMC for empirical datasets, and dataset thinning, reducing the number of samples which are stored.

Conclusions

- nested simulation is an important new direction for MLMC research
- splitting a large set of samples into 2 or more subsets is key to a good coupling
- complete numerical analysis proves $O(\varepsilon^{-2})$ complexity
- numerical tests show good benefits in some cases, but in others QMC is more effective – MLQMC might be best?
- MCMC inputs can be introduced through random selection from a large pre-generated empirical dataset

Webpages:

http://people.maths.ox.ac.uk/gilesm/slides.html

http://people.maths.ox.ac.uk/gilesm/mlmc_community.html