Taught Course Centre Short Course

"Computational Methods for Uncertainty Quantification"

Robert Scheichl, University of Bath

Model Solutions for Exercise Sheet 1

1. (a) It follows from the Berry-Esseen Inequality that

$$\Phi(x) - \frac{\rho}{2\sigma^3\sqrt{N}} \le \mathbf{P}\{S_N^* \le x\} \le \Phi(x) + \frac{\rho}{2\sigma^3\sqrt{N}}$$

and consequently

$$\begin{aligned}\mathbf{P}\{|S_N^*| \le x\} = \mathbf{P}\{S_N^* \le x\} - \mathbf{P}\{S_N^* \le -x\} &\le \Phi(x) + \frac{\rho}{2\sigma^3\sqrt{N}} - \Phi(-x) + \frac{\rho}{2\sigma^3\sqrt{N}} \\ &= \underbrace{\Phi(x) - \Phi(-x)}_{=:\gamma_x} + \frac{\rho}{\sigma^3\sqrt{N}} \qquad (1)\end{aligned}$$

Similarly, we can show $\mathbf{P}\{|S_N^*| \le x\} \ge \gamma_x - \frac{\rho}{\sigma^3\sqrt{N}}$. Since $S_N^* = \frac{S_N - N\mu}{\sqrt{N}\sigma}$ this implies

$$\gamma_x - \frac{\rho}{\sigma^3\sqrt{N}} \le \mathbf{P}\left(\mu \in \left[\frac{S_N}{N} - \frac{\sigma x}{\sqrt{N}}, \frac{S_N}{N} + \frac{\sigma x}{\sqrt{N}}\right]\right) \le \gamma_x + \frac{\rho}{\sigma^3\sqrt{N}}\,.$$

As an example, choosing $x = 1.96$ we get $\phi(x) = 0.95$ and so

$$0.95 - \frac{\rho}{\sigma^3\sqrt{N}} \le \mathbf{P}\left(\mu \in \left[\frac{S_N}{N} - \frac{1.96\sigma}{\sqrt{N}}, \frac{S_N}{N} + \frac{1.96\sigma}{\sqrt{N}}\right]\right) \le 0.95 + \frac{\rho}{\sigma^3\sqrt{N}}\,. \qquad (2)$$

(b) In the Buffon needle problem, we have

$$\mu = p,\ \sigma^2 = p(1-p),\ \rho = p(1-p)(1-2p+2p^2).$$

and in Lazzarini's experiment $N = 3408$ and $p = \frac{2\ell}{\pi d} = \frac{5}{3\pi}$. Therefore, from (**??**) (neglecting the correction $\frac{\rho}{\sigma^3\sqrt{N}}$ for finite N), we get an (asymptotic) 95% confidence interval for p of

$$\left[\frac{1808}{3408} - \frac{1.96\sigma}{\sqrt{3408}}, \frac{1808}{3408} - \frac{1.96\sigma}{\sqrt{3408}}\right] = [0.51376, 0.54727]$$

or equivalently, multiplying by the number of throws, the (asymptotic) 95% confidence interval for the number of intersections S_{3408} in 3408 throws is $[1751, 1865]$. Strictly speaking, since $\frac{\rho}{\sigma^3\sqrt{N}} = 0.0172$, the probability that S_{3408} is in that interval is bigger than 93.3% and smaller than 96.7%.

Also, using the exact value for $p = \frac{5}{3\pi}$, we see from (**??**) that the probability that

$$|S_N^*| = \left|\frac{S_N - Np}{\sqrt{Np(1-p)}}\right| = \sqrt{\frac{N}{p(1-p)}}\left|\frac{S_N}{N} - p\right|$$

is less than $x = \sqrt{\frac{3408}{p(1-p)}}\left|\frac{1808}{3408} - p\right| = 5.27 \cdot 10^{-6}$ is less than $\gamma_x + \frac{\rho}{\sigma^3\sqrt{3408}} = 4.2 \cdot 10^{-6} + 0.01722564 = 0.01723$. So the probability that Lazzarini's machine would produce exactly 1808 intersections in 3408 throws is less than 1.7%.

2. Recalling from Slide 9 in Lecture 2 that $\mathbf{E}\left[\widehat{Q}_M\right] - \mathbf{E}\left[Q_M\right] = 0$ we get

$$
\begin{aligned}
\mathbf{E}\left[\left(\mathbf{E}\left[Q\right] - \widehat{Q}_M\right)^2\right] &= \mathbf{E}\left[\left(\underbrace{\mathbf{E}\left[Q\right] - \mathbf{E}\left[Q_M\right]}_{=\mathbf{E}[Q-Q_M]} + \mathbf{E}\left[\widehat{Q}_M\right] - \widehat{Q}_M\right)^2\right] \\
&= \mathbf{E}\left[\left(\mathbf{E}\left[Q - Q_M\right]\right)^2 + \left(\mathbf{E}\left[\widehat{Q}_M\right] - \widehat{Q}_M\right)^2 + 2\mathbf{E}\left[Q - Q_M\right]\left(\mathbf{E}\left[\widehat{Q}_M\right] - \widehat{Q}_M\right)\right]
\end{aligned}
$$

Using linearity of the expected value and the fact that most of the terms under the expected value are not actually random, we can simplify this to

$$
\begin{aligned}
\mathbf{E}\left[\left(\mathbf{E}\left[Q\right] - \widehat{Q}_M\right)^2\right] &= \left(\mathbf{E}\left[Q - Q_M\right]\right)^2 + \mathbf{Var}[\widehat{Q}_M] + \ 2\mathbf{E}\left[Q - Q_M\right]\underbrace{\left(\mathbf{E}\left[\widehat{Q}_M\right] - \mathbf{E}\left[\widehat{Q}_M\right]\right)}_{=0} \\
&= \left(\mathbf{E}\left[Q - Q_M\right]\right)^2 + \frac{\mathbf{Var}[Q_M]}{N}.
\end{aligned}
$$

3. (a) Expanding the definition of the variance we get

$$
\begin{aligned}
\mathbf{Var}\left[\frac{1}{2}(\widehat{Q}_{M,N} + \widehat{\widetilde{Q}}_{M,N})\right] &= \mathbf{E}\left[\left(\frac{1}{2}(\widehat{Q}_{M,N} + \widehat{\widetilde{Q}}_{M,N}) - \frac{1}{2}(\mathbf{E}\left[Q\right] + \mathbf{E}\left[Q\right])\right)^2\right] \\
&= \frac{1}{4}\mathbf{E}\left[(\widehat{Q}_{M,N} - \mathbf{E}\left[Q\right])^2 + (\widehat{\widetilde{Q}}_{M,N} - \mathbf{E}\left[Q\right])^2 + 2(\widehat{Q}_{M,N} - \mathbf{E}\left[Q\right])(\widehat{\widetilde{Q}}_{M,N} - \mathbf{E}\left[Q\right])\right] \\
&= \frac{1}{4}\left(\mathbf{Var}\left[\widehat{Q}_{M,N}\right] + \mathbf{Var}\left[\widehat{\widetilde{Q}}_{M,N}\right] + 2\,\mathbf{Cov}\left(\widehat{Q}_{M,N}, \widehat{\widetilde{Q}}_{M,N}\right)\right)
\end{aligned}
$$

Using the definition of the sample variances and sample covariances of $\{Q_M^{(k)}\}$ and $\{\widetilde{Q}_M^{(k)}\}$ from lectures and expanding we get

$$
\begin{aligned}
s_Q^2 &:= \frac{1}{N-1}\sum_{k=1}^{N}(Q_M^{(k)} - \widehat{Q}_{M,N})^2 \ = \ \frac{1}{N-1}\left(\sum_{k=1}^{N}\left(Q_M^{(k)}\right)^2 - \frac{1}{N}\left(\sum_{k=1}^{N}Q_M^{(k)}\right)^2\right) \\
s_{\widetilde{Q}}^2 &:= \frac{1}{N-1}\sum_{k=1}^{N}(\widetilde{Q}_M^{(k)} - \widehat{\widetilde{Q}}_{M,N})^2 \ = \ \frac{1}{N-1}\left(\sum_{k=1}^{N}\left(\widetilde{Q}_M^{(k)}\right)^2 - \frac{1}{N}\left(\sum_{k=1}^{N}\widetilde{Q}_M^{(k)}\right)^2\right) \\
c_{Q,\widetilde{Q}} &:= \frac{1}{N-1}\sum_{k=1}^{N}(Q_M^{(k)} - \widehat{Q}_{M,N})(\widetilde{Q}_M^{(k)} - \widehat{\widetilde{Q}}_{M,N}) \\
&\qquad = \frac{1}{N-1}\left(\sum_{k=1}^{N}Q_M^{(k)}\widetilde{Q}_M^{(k)} - \frac{1}{N}\left(\sum_{k=1}^{N}Q_M^{(k)}\right)\left(\sum_{k=1}^{N}\widetilde{Q}_M^{(k)}\right)\right)
\end{aligned}
$$

Hence, we can estimate

$$
\mathbf{Var}\left[\frac{1}{2}(\widehat{Q}_{M,N} + \widehat{\widetilde{Q}}_{M,N})\right] \quad \text{by} \quad \frac{s_Q^2 + s_{\widetilde{Q}}^2 + 2c_{Q,\widetilde{Q}}}{4N}.
$$

Within the iteration over the samples in the code we only have to keep track of the sums

$$
\sum_{k=1}^{N}Q_M^{(k)}, \ \ \sum_{k=1}^{N}\widetilde{Q}_M^{(k)}, \ \ \sum_{k=1}^{N}\left(Q_M^{(k)}\right)^2, \ \ \sum_{k=1}^{N}\left(\widetilde{Q}_M^{(k)}\right)^2 \ \ \text{and} \ \ \sum_{k=1}^{N}Q_M^{(k)}\widetilde{Q}_M^{(k)}.
$$

(b) See my model code.

(c) See my model code. In my model code the variance is reduced by almost a factor 5, but this reduction does not get bigger for smaller tolerances TOL.

4. (a) Let us define the following cost functional (including the constraint on the variance via a Lagrange multiplier):

$$\mathcal{L}(N_0, \dots, N_L, \lambda) = \sum_{\ell=0}^{L} \mathcal{C}_\ell N_\ell + \lambda \left(\sum_{\ell=0}^{L} \frac{\mathbf{Var}[Y_\ell]}{N_\ell} - \frac{\mathrm{TOL}^2}{2} \right).$$

The first order optimality conditions are to set to zero all the first-order partial derivatives of $\mathcal{L}$ with respect to its arguments. This leads to

$$0 = \frac{\partial \mathcal{L}}{\lambda} = \sum_{\ell=0}^{L} \frac{\mathbf{Var}[Y_\ell]}{N_\ell} - \frac{\mathrm{TOL}^2}{2} \tag{3}$$

$$0 = \frac{\partial \mathcal{L}}{N_\ell} = \mathcal{C}_\ell - \lambda \frac{\mathbf{Var}[Y_\ell]}{N_\ell^2}, \qquad \ell = 0, \dots, L \tag{4}$$

Equations (**??**) imply

$$N_\ell = \sqrt{\lambda} \sqrt{\frac{\mathbf{Var}[Y_\ell]}{\mathcal{C}_\ell}}, \qquad \ell = 0, \dots, L, \tag{5}$$

as claimed in the notes. To find the constant $\sqrt{\lambda}$ (i.e. the square root of the Lagrange multiplier), we substitute into (**??**) and get

$$\sum_{\ell=0}^{L} \mathbf{Var}[Y_\ell] \sqrt{\frac{\mathcal{C}_\ell}{\lambda\, \mathbf{Var}[Y_\ell]}} = \frac{\mathrm{TOL}^2}{2} \qquad \Rightarrow \qquad \sqrt{\lambda} = \frac{2}{\mathrm{TOL}^2} \sum_{\ell=0}^{L} \sqrt{\mathcal{C}_\ell\, \mathbf{Var}[Y_\ell]}.$$

(b) See either the paper `https://people.maths.ox.ac.uk/gilesm/files/OPRE_2008.pdf` or my paper `http://www.maths.bath.ac.uk/~masrs/cgst_mlmc_cvs2010.pdf` for proofs of this theorem that essentially use the argument in (a).

5. (a) See my model code.

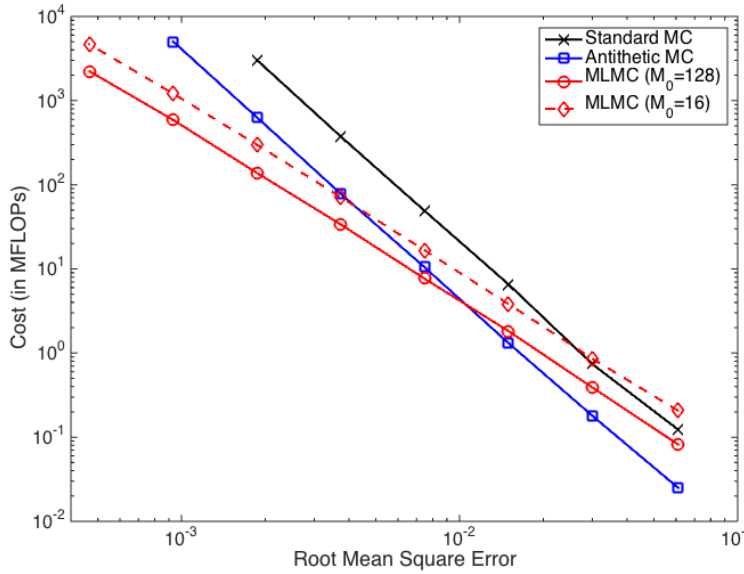
I did not implement the fully adaptive algorithm in the lecture notes. Instead I pass N_0, the number of samples on the coarsest level, as an argument and then derive N_ℓ from (**??**). By taking the ratio N_ℓ/N_0 we do not need to know (or estimate) the constant $\sqrt{\lambda}$. Instead, with the choice $s = 2$, we get

$$N_\ell = N_0 \sqrt{\frac{\mathbf{Var}[Y_\ell]\mathcal{C}_0}{\mathbf{Var}[Y_0]\mathcal{C}_\ell}} = \frac{2}{3} N_0 2^{-\ell/2} \sqrt{\frac{\mathbf{Var}[Y_\ell]}{\mathbf{Var}[Y_0]}}$$

where I have used that $M_\ell = 2^\ell M_0$ and $\mathcal{C}(Q_\ell^{(k)}) = 8M_\ell$, since in each step of the Euler method my code carries out 8 floating point operations. This implies that $\mathcal{C}(Y_\ell^{(k)}) = 8(M_\ell + M_{\ell-1}) = 12M_\ell$, for $\ell > 0$. The total number of floating point operations is

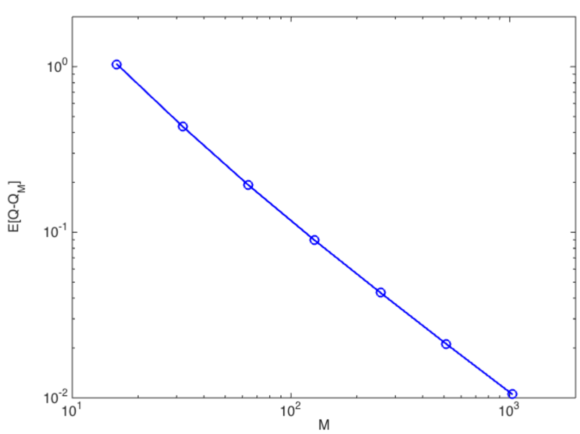
$$\mathcal{C}(\widehat{Q}^{\mathrm{ML}}_{L,\{N_\ell\}}) = 8M_0N_0 + 12 \sum_{\ell=1}^{L} M_\ell N_\ell\,.$$

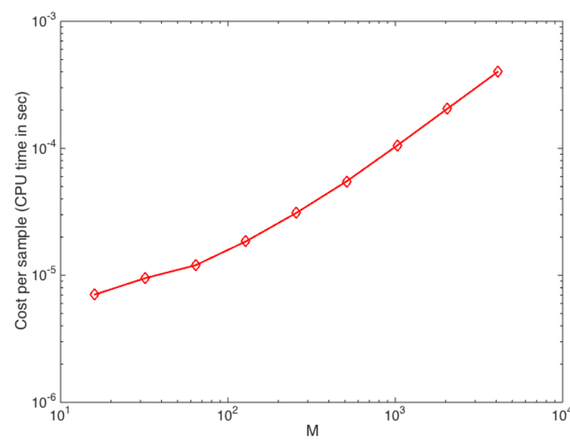
Here is a plot of cost against tolerance with the 3 codes (standard MC, anithetic MC, MLMC):



As predicted, the cost for standard and antithetic MC grows like TOL^{-3} and the cost for MLMC grows like TOL^{-2}. The actual cost depends on the choice of coarsest grid.

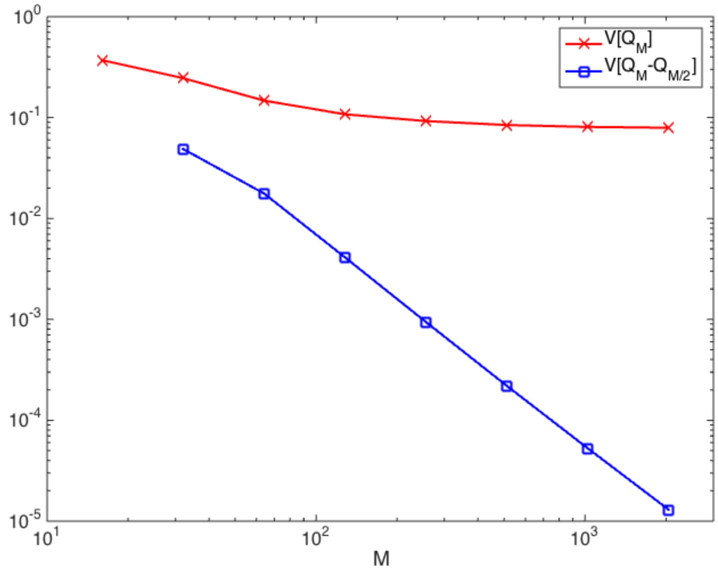
(b) To estimate α, I use my MLMC code with only two levels, i.e. $L = 1$ and $s = M_1/M_0$ and N both sufficiently large, so that essentially the finer calculation is exact and the sampling error is negligible. In the following figure (left) we see a log-log plot of $|\widehat{Y}_1| \approx |\mathbf{E}[Q_{M_1} - Q_{M_0}]| \approx |\mathbf{E}[Q - Q_{M_0}]|$. Clearly the error decays like M_0^{-1}.





To estimate γ (above figure, right), I simply measured the CPU-time (with tic and toc in Matlab) averaged over N samples. We see that $\gamma \approx 1$ for M sufficiently large.

Finally, in the last figure below, we see a plot of $\mathbf{Var}[\widehat{Y}_\ell]$ and $\mathbf{Var}[\widehat{Q}_{M_\ell}]$ for a range of values of ℓ. We see that the numerically observed rate $\beta \approx 2$. To prove this, use the bound on the Euler discretisation error on Slide 18 from Lecture 2:



$$
\begin{aligned}
\mathbf{Var}[\widehat{Y}_\ell] &= \frac{1}{N_\ell}\,\mathbf{Var}[Q_{M_\ell} - Q_{M_{\ell-1}}] \\
&\le \frac{1}{N_\ell}\,\mathbf{E}\left[\left(Q_{M_\ell} - Q_{M_{\ell-1}}\right)^2\right] \\
&\le \frac{2}{N_\ell}\left(\mathbf{E}\left[\left(Q - Q_{M_{\ell-1}}\right)^2\right] + \mathbf{E}\left[\left(Q - Q_{M_\ell}\right)^2\right]\right) \\
&\le \frac{2}{N_\ell}\left(KLM_{\ell-1}^{-2} + KLM_\ell^{-2}\right) \le \underbrace{2KL(1+s^2)}_{\text{constant}}\,N_\ell^{-1}M_\ell^{-2}\,.
\end{aligned}
$$

(c) For example, you could combine antithetic sampling and MLMC, or use a quasi-Monte Carlo method (see Wednesday).