RELIABLE QUASI-MONTE CARLO WITH CONTROL VARIATES

BY

DA LI

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Applied Mathematics
in the Graduate College of the
Illinois Institute of Technology

Approved ____________________________
Advisor

Chicago, Illinois
Aug 01

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

With Quasi-Monte Carlo(QMC) methods being implemented in a guaranteed adaptive algorithm recently, possibility of combining that with traditional proficiency improvement techniques for Monte Carlo(MC) such as control variates is brought to table.

The problem for adding control variates to QMC is that optimal control variate coefficient for QMC is generally not the same as MC. Here we propose a method for computing the optimal control variate coefficients with guaranteed adaptive QMC algorithm. One merit of control variate is that there is always a good solution as good as using no control variates. And our method is implemented in an efficient way that even in that case the extra cost for control variates is not significant.

As for examples we will include two 16 dimensional integration on financial problem. One is pricing arithmetic mean Asian option with geometric mean Asian option as control variates and the other is barrier option with European option as control variates. Our results show that with good control variates, the cost of adaptive QMC is greatly reduced compared with no control variates cases.

CHAPTER 1

BACKGROUND

1.1 QMC

1.1.1 Digital Sequence.

Talk about the whole idea briefly. [1]

1.1.2 QMC.

Introduce QMC.

1.2 Control Variates

1.2.1 A Brief Review. Control variates has been know as variance reduction technique used in Monte Carlo methods.In this part we will brief review some crucial idea of this methods so we can see what's the problem for using it with QMC. [2]

Suppose we have the following integration approximation problem

$$I = \int_{[0,1]^d} f(x)dx$$

If we use Monte-Carlo method, the estimator should be

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i), X_i \sim \mathcal{U}[0,1)^d$$

Now suppose we have a known function h and its value on the interval $\int h(x)dx = \theta$. We construct a new estimator as the following

$$\hat{I}_{cv}(f) = \frac{1}{n}\sum_{i=1}^{n} \Big[f(X_i) - \beta_{mc}[h(X_i) - \theta] \Big] \quad s.t.\ X_i \sim \mathcal{U}[0,1),\ i.i.d.$$

We can easily see it's an unbiased estimator, i.e. $\mathbb{E}(\hat{I}_{cv}) = I$. Now we want to pick the right β_{mc} such that make the estimation more efficient. Base on previous MC error estimating formula (??), we know it's achievable by minimizing the variance of the estimator, which is

$$
\begin{aligned}
\mathrm{Var}_{mc}(\hat{I}_{cv}) =& \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}\big[f(X_i) - \beta_{mc}[h(X_i) - \theta]\big]\Big) \\
=& \frac{1}{n}\mathrm{Var}\Big(f(X_i) - \beta_{mc}[h(X_i) - \theta]\Big) \quad \text{by } X_i \text{ i.i.d} \\
=& \frac{1}{n}\mathbb{E}\Big(\big[f(X_i) - \beta_{mc}[h(X_i) - \theta] - I\big]^2\Big) \\
=& \frac{1}{n}\mathbb{E}\Big(\big[[f(X_i) - I] - \beta_{mc}[h(X_i) - \theta]\big]^2\Big) \\
=& \frac{1}{n}\mathbb{E}\big([f(X_i) - I]^2 - 2\beta_{mc}[f(X_i) - I][h(X_i) - \theta] + \beta_{mc}^2[h(X_i) - \theta]^2\big) \\
=& \frac{1}{n}\Big(\mathrm{Var}[f(X_i)] - 2\beta_{mc}\mathrm{Cov}[f(X_i), h(X_i)] + \beta_{mc}^2\mathrm{Var}[h(X_i)]\Big) \\
=& \frac{1}{n}\Big(\mathrm{Var}[h(X_i)](\beta_{mc} - \frac{\mathrm{Cov}[f(X_i), h(X_i)]}{\mathrm{Var}[h(X_i)]})^2 + \mathrm{Var}[f(X_i] - \frac{\mathrm{Cov}^2[f(X_i), h(X_i)]}{\mathrm{Var}[h(X_i)]}\Big)
\end{aligned}
$$

Then the optimal β_{mc} is given by

$$
\beta_{mc}^* = \frac{\mathrm{Cov}[f(X_i), h(X_i)]}{\mathrm{Var}[h(X_i)]} \tag{1.1}
$$

In which case the variance become

$$
\mathrm{Var}_{mc}(\hat{I}_{cv}) = \frac{\mathrm{Var}[f(X_i)]}{n}\big(1 - \mathrm{corr}^2[f(X_i), h(X_i)]\big)
$$

and we always have

$$
\mathrm{Var}_{mc}(\hat{I}_{cv}) \leq \frac{\mathrm{Var}[f(X_i)]}{n} = \mathrm{Var}_{mc}(\hat{I})
$$

Now we can see the merit of control variates as a variance reduction method.In worst case, we get a completely uncorrelated g that leads correlation to zero, and we have variance exactly the same as not using control variates. On the other hand, the more correlated our control variates is to the target function, the more variance we can get rid of by using the method.

1.2.2 Control Variates with QMC. As we pointed out earlier, QMC use a different way for generating X_i, they are still identical(from same distribution) but not independent anymore, which caused the problem for control variates.

Suppose $X_1, \ldots, X_n$ are generated by QMC rule, the estimator stays the same

$$\hat{I}_{cv}(f) = \frac{1}{n}\sum_{i=1}^{n}\Big[f(X_i) - \beta_{qmc}[h(X_i) - \theta]\Big] \quad X_i \in \mathcal{U}(0,1)$$

We can easily prove it is still unbiased

$$\mathbb{E}(\hat{I}_{cv}) = \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\Big[f(X_i) - \beta_{mc}[h(X_i) - \theta]\Big]\Big) = I$$

However, it's not the same case as MC like we presented before, because we do not have i.i.d for X_i this time

$$\mathrm{Var}_{qmc}(\hat{I}_{cv}) \neq \frac{1}{n}\mathrm{Var}\Big(f(X_i) - \beta_{mc}[h(X_i) - \theta]\Big)$$

Instead the variance become

$$\begin{aligned}
\mathrm{Var}_{qmc}(\hat{I}_{cv}) &= \mathrm{Var}\Big(\hat{I} - \beta_{qmc}\hat{H}\Big) \quad s.t.\ \hat{I} = \sum_{i=1}^{n} f(X_i),\ \hat{H} = \sum_{i=1}^{n}[h(X_i) - \theta] \\
&= \mathrm{Var}(\hat{I}) - 2\beta_{qmc}\mathrm{Cov}(\hat{I},\hat{H}) + \beta_{qmc}^2\mathrm{Var}(\hat{H}) \\
&= \mathrm{Var}(\hat{H})\Big(\beta_{qmc} - \frac{\mathrm{Cov}(\hat{I},\hat{H})}{\mathrm{Var}(\hat{H})}\Big)^2 + \mathrm{Var}(\hat{I}) - \frac{\mathrm{Cov}(\hat{I},\hat{H})}{\mathrm{Var}(\hat{H})}
\end{aligned}$$

The optimal β_{qmc} is

$$\beta_{qmc}^* = \mathrm{Var}(\hat{H})^{-1}\mathrm{Cov}(\hat{I},\hat{H}) \tag{1.2}$$

which leave the variance to be

$$\mathrm{Var}_{qmc}(\hat{I}_{cv}) = \mathrm{Var}(\hat{I})\big(1 - \mathrm{corr}^2[\hat{I},\hat{H}]\big)$$

Now we are interested that if our previous formula for $\hat{\beta}_{mc}$ could be an estimation for $\hat{\beta}_{qmc}$. The fact is that they are generally not the same. Let's take the covariance part of formula (1.2) and (1.1) to see the difference.

$$
\begin{aligned}
\mathrm{Cov}(\hat{I},\hat{H}) &= \int [f(X_1)+\cdots+f(X_n)][h(X_1)+\cdots+h(X_n)]\, d\mathbf{X} \\
&= \int [\sum_{i=1}^{n} f(X_i)h(X_i) + \sum_{i,j=1}^{i\neq j} f(X_i)h(X_j)]\, d\mathbf{X} \\
&\neq \int f(X_i)h(X_i)dX_i \\
&= \mathrm{Cov}x[(f(X_i),h(X_i)]
\end{aligned}
$$

There is also a very good example from Hicknell and Lemieux(2005) [2]'s paper, showing that β_{mc} and β_{qmc} can make a quite different results in some cases.

1.3 Reliable Adaptive QMC with digital sequence

1.3.1 Idea of adaptive cubature algorithm.

One practical problem for QMC method is that how to get the sample size big enough for a required error tolerance. The idea in work of Hickernell and Jimnez Rugama(2014) [1] is to construct a QMC algorithm with reliable error estimation on digital sequence. Here we briefly summarize their results.

The error of QMC method on digital sequence can be expressed in terms of Walsh coefficients of the integrand on certain cone conditions.

$$
\begin{aligned}
&\Big|\int_{[0,1)^d} f(x)dx - \hat{I}_m(f)\Big| \leq a(r,m) \sum_{\lfloor 2^{m-r-1}\rfloor}^{2^{m-r}-1} |\hat{f}_{m,k}| \\
&\hat{I}_m(f) := \frac{1}{b^m}\sum_{i=0}^{b^m-1} f(z_i \oplus \Delta) \\
&\hat{f}_{m,k} = \text{ discrete Fourier coefficients of } f \\
&a(r,m) = \text{ inflation factor that depends on } \mathcal{C}
\end{aligned}
$$

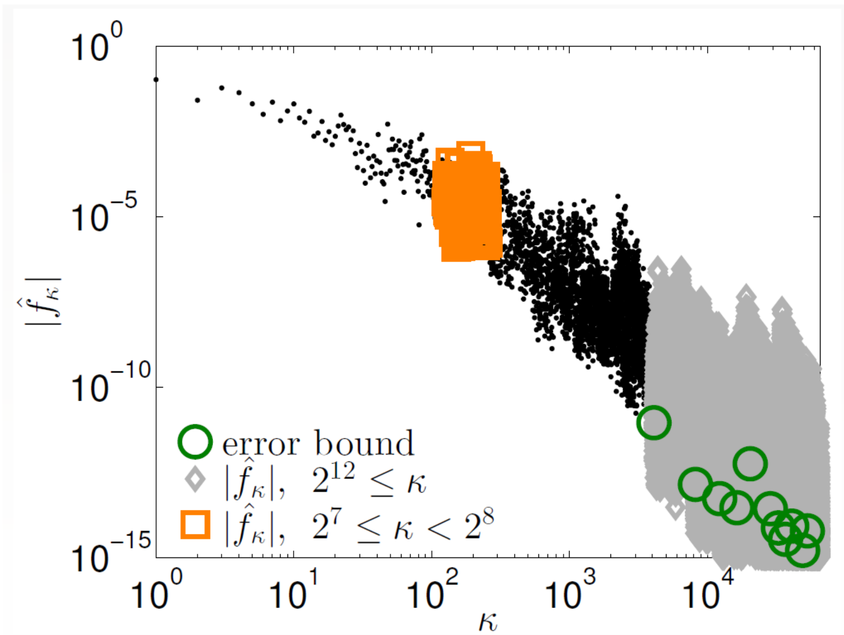
Here is the defination of the cone condition.

$$\left|\int_{[0,1)^d} f(x)dx - \hat{I}_m(f)\right| \leq \sum \bigcirc \leq \sum \square \leq a(r,m) \sum_{\lfloor b^{m-r-1}\rfloor}^{b^{m-r}-1} |\hat{f}_{m,k}|$$

$$\bigcirc := \sum_{\lambda=1}^{\infty} |\hat{f}_{\lambda b^m}|, \quad \square := \sum_{\kappa=b^{l-1}}^{b^l-1} |\hat{f}_\kappa|, \quad \Diamond := \sum_{\kappa=b^m}^{\infty} |\hat{f}_\kappa|$$

$$\mathcal{C} := \left\{ \sum \bigcirc \leq \sum \Diamond \leq \sum \square \right\}$$

Figure 1.1. Cone condition for reliable adpative QMC algorithm



CHAPTER 2

RELIABLE ADAPTIVE QMC SOBOL WITH CV

2.1 Theory

2.1.1 Idea to add control variates to cubSobol.

The idea is similar to traditional control variates technique for Monte-Carlo. If we know the integration of a function $\boldsymbol{h} = (h_1, \dots, h_J)$ on the interval same as our f, say $\int_{[0,1)^d} h_j dx = \theta_j$, then we can define a new function g

$$g := f - (\boldsymbol{h} - \boldsymbol{\theta})\boldsymbol{\beta} \tag{2.1}$$

$$\text{s.t. } \boldsymbol{\theta} = (\theta_1, \dots, \theta_J), \boldsymbol{\beta} = (\beta_1, \dots, \beta_J)^T$$

Then easily we can find that if we replace f with g, the integration stays the same.

$$\int_{[0,1)^d} g dx = \int_{[0,1)^d} f - (\boldsymbol{h} - \boldsymbol{\theta})\boldsymbol{\beta} dx = \int_{[0,1)^d} f dx$$

However, as we stated in chapter 3, we can not find optimal β by minimizing variance of estimator like Monte-Carlo. Instead we have a different way to estimate error with adaptive QMC method introduced in chapter 2.

Recall equation (**??**) our error bound for the new estimator of g still holds

$$\Big| \int_{[0,1)^d} g dx - \hat{I}_m(g) \Big| \leq a(r,m) \sum_{\lfloor 2^{m-r-1} \rfloor}^{2^{m-r}-1} |\hat{g}_{m,k}| \tag{2.2}$$

Naturally, the new estimator become

$$\hat{I}_m(g) := \frac{1}{b^m} \sum_{i=0}^{b^m-1} g(z_i + \Delta) \tag{2.3}$$

2.1.2 Optimizing β.

From (2.2) it is clear that the optimal β is the one that minimize the error term.

$$\boldsymbol{\beta}^* = \min_{\boldsymbol{\beta}} \sum_{\kappa=b^{m-r-1}}^{b^{m-r}-1} |\hat{g}_\kappa| \tag{2.4}$$

$$= \min_{\boldsymbol{\beta}} \sum_{\kappa=b^{m-r-1}}^{b^{m-r}-1} |\hat{f}_\kappa - (\hat{\boldsymbol{h}}_\kappa - \hat{\boldsymbol{\theta}})\boldsymbol{\beta}| \quad \hat{\boldsymbol{h}}_\kappa = (\hat{h}_{\kappa,1}, \dots, \hat{h}_{\kappa,J}), \hat{\boldsymbol{\theta}} = (\hat{\theta}_{\kappa,1}, \dots, \hat{\theta}_{\kappa,J}) \tag{2.5}$$

$$= \min_{\boldsymbol{\beta}} \|\hat{\boldsymbol{f}} - \hat{\boldsymbol{H}}\boldsymbol{\beta}\|_1 \quad \hat{\boldsymbol{f}} = (\hat{f}_{b^{m-r-1}}, \dots, \hat{f}_{b^{m-r}-1)})^T \tag{2.6}$$

$$\approx \min_{\boldsymbol{\beta}} \|\hat{\boldsymbol{f}} - \hat{\boldsymbol{H}}\boldsymbol{\beta}\|_2 \quad \hat{\boldsymbol{H}} = (\hat{\boldsymbol{H}}_1, \dots, \hat{\boldsymbol{H}}_J) \tag{2.7}$$

$$\hat{\boldsymbol{H}}_j = (\hat{h}_{b^{m-r-1},j} - \hat{\theta}_j, \dots, \hat{h}_{b^{m-r}-1,j} - \hat{\theta}_j)^T$$

The second equivalence is not hard to get, but the third one may not be so obvious. Let's consider it backwards. Suppose we have a vector A and it's $\mathcal{L}_1$-norm.

$$A = \begin{pmatrix} f_1 - z_1 \\ f_2 - z_2 \\ \dots \\ f_n - z_n \end{pmatrix}, \quad \|A\|_1 = \sum_{i=1}^{n} |f_i - z_i|, \quad z_i := (\boldsymbol{h}_i - \boldsymbol{\theta})$$

If we replace the index,A is exactly what's inside the $\mathcal{L}_1$-norm in (2.6). Hence we justified the third equivalence. The reason we use an approximation instead, i.e. the $\mathcal{L}_1$-norm, is because there is no efficient way to solve it compared to existing least square methods.

2.1.3 The problem with θ.

We noticed a problem in solution for optimal β, which is we do a lot of subtractions with θ. This could be a large cost when we have difficult functions which means b^{m-r} could be very large number. Therefore we present a way to avoid that part.

The idea is form a observation that Walsh transform of θ in (**??**) is actually zero, since $\hat{h}_\kappa = \hat{g}_\kappa - \theta\delta_{\kappa,0}$ and the summation is not start from $\kappa = 0$.

This simplifies (2.6) to the following. Note that we only need the information of function f and h to calculate β^*, θ has been get rid of the optimization process.

$$\boldsymbol{\beta}^* = \min_{\boldsymbol{\beta}} \|\hat{\boldsymbol{f}} - \hat{\boldsymbol{H}}\boldsymbol{\beta}\|_1 \qquad \hat{\boldsymbol{f}} = (\hat{f}_{b^{m-r-1}}, \ldots, \hat{f}_{b^{m-r}-1)})^T \tag{2.8}$$

$$\approx \min_{\boldsymbol{\beta}} \|\hat{\boldsymbol{f}} - \hat{\boldsymbol{H}}\boldsymbol{\beta}\|_2 \qquad \hat{\boldsymbol{H}} = (\hat{\boldsymbol{H}}_1, \ldots, \hat{\boldsymbol{H}}_J) \tag{2.9}$$

$$\hat{\boldsymbol{H}}_j = (\hat{h}_{b^{m-r-1},j}, \ldots, \hat{h}_{b^{m-r}-1,j})^T$$

The same problem happened with the estimator 2.3.We have the similar solution for that.

$$\begin{aligned}
\hat{I}_m(g) &= \frac{1}{b^m}\sum_{i=0}^{b^m-1} g(z_i + \Delta) \\
&= \frac{1}{b^m}\sum_{i=0}^{b^m-1} f(z_i + \Delta) - (\boldsymbol{h}(z_i + \Delta) - \boldsymbol{\theta})\boldsymbol{\beta} \\
&= \frac{1}{b^m}\sum_{i=0}^{b^m-1} [f(z_i + \Delta) - \boldsymbol{h}(z_i + \Delta)\boldsymbol{\beta}] + \boldsymbol{\theta}\boldsymbol{\beta}
\end{aligned} \tag{2.10}$$

After organize it the in format of (2.10), θ is eliminated from the summation part. From these two parts of work on θ we managed to save $\frac{b-1}{b}b^{m-r} + b^m$ operations of subtraction.

2.2 Algorithm

Now combining the work from **??** and our previous work, we have the following algorithm for reliable adaptive QMC with control variates.

2.3 Theorem

To be added.

Algorithm 1: Reliable Adaptive QMC with control variates

Data: function f and $\boldsymbol{H}$; value of $\int_{[0,1)^d} h_j dx = \theta_j$; tolerance ε

Result: estimate of $\int_{[0,1)^d} f dx$; samples size; optimal β

begin

$m, r \longleftarrow$ start numbers

$x \longleftarrow$ grab 2^m sobolset points

1 $\tilde{\kappa} \longleftarrow$ get kappa map using **??**

2 $\hat{f}, \hat{\boldsymbol{H}} \longleftarrow$ get Walsh transform of $f, \boldsymbol{h}$ using **??**

3 $a \longleftarrow 2^{m-r-1}, b \longleftarrow 2^{m-r} - 1$

4 $\boldsymbol{\beta} \longleftarrow \hat{H}\left\{\tilde{\kappa}[x(a:b)]\right\} \backslash \hat{f}\left\{\tilde{\kappa}[x(a:b)]\right\}, \hat{g} \longleftarrow \hat{f} - \hat{\boldsymbol{H}}\boldsymbol{\beta}$

5 $\tilde{S}_{m-r,m}(g) \longleftarrow \sum_a^b \left|\hat{g}\left\{\tilde{\kappa}[x(a:b)]\right\}\right|,$

6 **if** $a(m,r)\tilde{S}_{m-r,m}(g) \leq \varepsilon$ **then**

return $\hat{I}_m(g) = \sum_{i=0}^{2^m-1} f[x(i)] + \boldsymbol{\theta}\boldsymbol{\beta}$

return $\boldsymbol{\beta}, n = 2^m$

7 **for** $m = m+1 : mmax$ **do**

$xnext \longleftarrow$ grab the next 2^{m-1} points in sobolset

repeat step 3 with new m

update kappa map(repeat step 1 with $(x : xnext)$)

repeat step 2 on $(x : xnext)$

repeat stet 4 with new $\tilde{\kappa}, a, b$

repeat step 5 with updated $\tilde{\kappa}$

repeat step 6 with new $\tilde{\kappa}, \boldsymbol{\beta}$

CHAPTER 3
NUMERICAL EXPERIMENT

3.1 When beta is not accurate?

Note that in algorthm 1, we didn't recalculate β for every iteration. The reason is that for most functions this is not neccesary, but in certain case β need to be updated to get the right answer. Here is an example showing that for certain strange functions beta needs to be updated.

3.2 Option Pricing

Option Pricing has always been a challenging topic in financial mathematics. Here we are going to demonstrate several examples of pricing different options with ccontrol variates.

3.2.1 Asian Option.

There are two types of asian options, depends on which types of mean you want to use. For this example we take arithematic mean asian call option as our target function, whose payoff function is

$$C_T^{Amean} = \max\left(\frac{1}{d}\sum_{j=1}^{d} S(jT/d) - K, 0\right)$$

S0	K	TimeVector	r	volatility	abstol	reltol
120	130	1/52:1/52:16/52	0.01	0.5	1e-3	0

Table 3.1. Parameter Setup for Up and In Barrier Call Option

Figure **??** shows decrease rate the walsh coefficients for the target function and control variates in this example.

Sample Size			Time Cost		
cubSobol	cv_old	cv_new	cubSobol	cv_old	cv_new
65535	8192	9011	0.2783	0.1034	0.0673

Table 3.2. Results of cubSobol, cv_old and cv_new with Asian Option

3.2.2 Barrier Option.

We will take up and in barrier call option as an example. Here is the payoff function for up and in barrier call option.

$$C_T^{U\&I} = (S_T - K)^+ 1_{\{\max S_t \geq Barrier\}}$$

From the payoff function it is naturally to consider european call option as control variates. Since if we take the barrier same as strike price, then this is just an european call option. Table **??** shows our setup for the barrier option.

Table 3.3. Parameter Setup for Up and In Barrier Call Option

S0	K	TimeVector	r	volatility	abstol	reltol
120	130	1/52:1/52:16/52	0.01	0.5	1e-3	0

We then took three different barrier as listed in table **??**, then we compared both oringinal cubSobol algorithm and the one with our modification as described in Chapter 4. We can see from the results in table **??** that new CV method takes less time than the old one, and both of them are much faster than QMC without CV.

Barrier	Sample Size			Time Cost		
	cubSobol	cv_old	cv_new	cubSobol	cv_old	cv_new
140	524288	78643	65535	1.874	0.5016	0.2743
135	524288	5802	6963	1.959	0.0781	0.0519
130	524288	1024	1024	1.876	0.0270	0.0199

Table 3.4. Results of cubSobol, cv_old and cv_new with Barrier Option

BIBLIOGRAPHY

[1] F. J. Hickernell and L. A. J. Rugama, "Reliable adaptive cubature using digital sequences," *arXiv preprint arXiv:1410.8615*, 2014.

[2] F. J. Hickernell, C. Lemieux, A. B. Owen, *et al.*, "Control variates for quasi-monte carlo," *Statistical Science*, vol. 20, no. 1, pp. 1–31, 2005.