Journal Pre-proof

Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation

Zhenghang Xu, Zhijian He, Xiaoqun Wang

PII: S0377-0427(23)00687-8
DOI: https://doi.org/10.1016/j.cam.2023.115745
Reference: CAM 115745

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 16 April 2021
Revised date : 27 November 2023

Efficient risk estimation via nested multilevel quasi-Monte Carlo simulation*

Zhenghang Xu[a], Zhijian He[b,*], Xiaoqun Wang[a]

[a] *Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, People's Republic of China*
[b] *School of Mathematics, South China University of Technology, Guangzhou, 510641, People's Republic of China*

Abstract

We consider the problem of estimating the probability of a large loss from a financial portfolio, where the future loss is expressed as a conditional expectation. Since the conditional expectation is intractable in most cases, one may resort to nested simulation. To reduce the complexity of nested simulation, we present an improved multilevel Monte Carlo (MLMC) method by using quasi-Monte Carlo (QMC) to estimate the portfolio loss in each financial scenario generated via Monte Carlo. We prove that using QMC can accelerate the convergence rates in both the crude nested simulation and the multilevel nested simulation. Under certain conditions, the complexity of the proposed MLMC method can be reduced to $O(\epsilon^{-2}(\log \epsilon)^2)$. On the other hand, we find that using QMC in MLMC encounters a high-kurtosis phenomenon due to the existence of indicator functions. To remedy this, we propose a smoothed method which uses logistic sigmoid functions to approximate indicator functions. Numerical results show that the optimal MLMC complexity $O(\epsilon^{-2})$ is almost attained even in moderate high dimensions.

Keywords: nested simulation, quasi-Monte Carlo, multilevel Monte Carlo, risk estimation
2010 MSC: 65C05, 62P05

1. Introduction

We consider the problem of estimating

$$\theta = \mathbb{P}(g(\omega) > c) = \mathbb{E}[\mathbb{I}\{g(\omega) > c\}] = \mathbb{E}[\mathbb{I}\{\mathbb{E}[X|\omega] > c\}] \tag{1}$$

*This work of the second author was funded by the National Science Foundation of China (No. 12071154), Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515010275). And the third author was funded by the National Science Foundation of China (No. 720711119).

*Corresponding author
Email addresses: `xzh17@mails.tsinghua.edu.cn` (Zhenghang Xu), `hezhijian@scut.edu.cn` (Zhijian He), `wangxiaoqun@mail.tsinghua.edu.cn` (Xiaoqun Wang)

via simulation for a given constant c, where $g(\omega) := \mathbb{E}[X|\omega]$. The inner expectation of the one-dimensional random variable X is conditioned on the value of the outer multidimensional random variable ω. This nested expectation appears, for instance, when estimating the probability of a large loss from a financial portfolio. If the portfolio consists of some complex financial derivatives (such as path-dependent options and exotic options) or the underlying model is complicated, the future loss $g(\omega)$ of the portfolio on a fixed period is only known as a conditional expectation with respect to the risk factor ω, which does not have an analytical form of ω. A typical approach to estimate such an expectation of a function of a conditional expectation is to use nested estimation. Specifically, nested simulation refers to a two-level simulation procedure. In the outer level, one generates a number of scenarios of ω. Then, in the inner level, a number of samples of X are generated for each simulated ω to estimate the conditional expectation $\mathbb{E}[X|\omega]$; see [1] and [2].

Nested simulation has been widely studied in the literature due to its broad applicability, particularly in portfolio risk measurement. Gordy and Juneja [2] considered and analyzed uniform nested simulation estimators which employ a constant number of inner samples across all scenarios in the outer level. They showed that to achieve a root mean squared error (RMSE) of ϵ, allocating proper computational effort in each level results in a total cost of $O(\epsilon^{-3})$. Nested simulation can be more efficient by allocating computational effort nonuniformly across scenarios. Particularly, Broadie et al. [1] proposed an adaptive procedure to allocate computational effort to inner simulations. The resulting nonuniform nested simulation estimator enjoys a reduced complexity of $O(\epsilon^{-5/2})$ under certain conditions. Recently, Giles and Haji-Ali [3] proposed to use multilevel Monte Carlo (MLMC) method in nested simulation. They showed that in the original MLMC, the complexity is $O(\epsilon^{-5/2})$. By incorporating the adaptive allocations procedure of [1], Giles and Haji-Ali [3] showed that the complexity of MLMC can be reduced to $O(\epsilon^{-2}(\log \epsilon)^2)$ under certain conditions. For some applications of nested MLMC, we refer to [4–8] and references therein.

MLMC is a sophisticated variance reduction technique introduced by Heinrich [9] for parametric integration and by Giles [10] for the estimation of the expectations arising from stochastic differential equations. Nowadays the MLMC method has been extended extensively. For a thorough review of MLMC methods, we refer to [11]. On the other hand, quasi-Monte Carlo (QMC) and randomized QMC (RQMC) methods are powerful tools to improve the efficiency of traditional Monte Carlo; see [12–14] for details. QMC methods have achieved great success in finance applications, such as option pricing and hedging [15]. It is natural to incorporate (R)QMC in the MLMC framework. Giles and Waterhouse [16] first attempted to apply QMC method into multilevel path simulation in financial problems. In recent years, multilevel QMC methods have received increasing attention amongst researchers due to its broad applicability, particularly in problems of partial differential equations with random coefficients [17–19] and in uncertainty quantification [20, 21].

In this paper, we focus on the combination of MLMC and QMC in nested simulation. Specifically, in the outer simulation, we use Monte Carlo to generate a number of scenarios of ω, whereas in the inner simulation, we use RQMC to estimate the portfolio loss in each scenario. Our work is closely related to Goda et al. [8], who used nested multilevel RQMC method to deal with the expected value of partial perfect information (EVPPI) problem. A central problem of EVPPI is to estimate an expectation of the form $\mathbb{E}[f(g(\omega))]$, where $g(\omega) = \mathbb{E}[X|\omega]$ and $f(\cdot)$ is a continuous function. As Goda et al. [8] pointed out, the multilevel RQMC estimator can achieve the optimal complexity of $O(\epsilon^{-2})$ under some mild conditions. However, for the problem (1) considered in this paper, the performance function $f(\cdot)$ becomes an indicator function $\mathbb{I}(\cdot > c)$, making it much harder than EVPPI for MLMC algorithms [5]. Moreover, the antithetic MLMC estimator for EVPPI used in [8] does not help to reduce the variance convergence rate in our setting because the discontinuity in the indicator function violates the differentiability requirements of the antithetic estimator [3]. As a result, the efficiency of using the antithetic form in multilevel RQMC is subtle for estimating (1). The gain of using RQMC in nested MLMC is also unclear. Can the multilevel RQMC estimator achieve the optimal complexity of $O(\epsilon^{-2})$ or the sub-optimal complexity of $O(\epsilon^{-2}(\log\epsilon)^2)$? It is well-known that the performance of (R)QMC integration depends on the smoothness of the integrand and the dimension of the problem [22–25]. So it is natural to ask how these factors affect the efficiency of the multilevel RQMC algorithm.

The main contribution of this work is to reduce the complexity of the MLMC method for financial risk management by using QMC methods in the multilevel scheme. We call the proposed method the multilevel QMC (MLQMC) as in [16–18] although QMC is used in estimating the inner expectation rather than the outer expectation. In Section 2, we review QMC methods and nested simulation. The complexity of the uniform nested simulation combined with QMC is analyzed. After introducing the basic MLMC method, we develop a multilevel QMC procedure for the problem (1). The effects of using the QMC method and the indicator function on the nested simulation are discussed. In Section 3, we develop a new smoothed coupling method, which aims to overcome the high-kurtosis phenomenon caused by the discontinuity of the indicator function. Numerical experiments are performed in Section 4. Finally, we present some remarks in Section 5.

2. Nested simulation and multilevel Monte Carlo

2.1. Nested simulation and quasi-Monte Carlo

Our goal is to estimate an expectation of a function of a conditional expectation via nested simulation, with the problem (1) of interest. In the uniform nested simulation, one takes

$$\hat{\theta}_{n,m} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}\{\hat{g}_m(\omega_i) > c\}, \tag{2}$$

and

$$\hat{g}_m(y) = \frac{1}{m}\sum_{j=1}^{m} X_j(y), \tag{3}$$

where ω_i are independent and identically distributed (iid) replications of ω, and $X_j(y)$ are iid replications of X given $\omega = y$. The quadrature rule $\hat{g}_m(y)$ in (3) is used to estimate the conditional expectation $g(y) = \mathbb{E}[X|\omega = y]$ in the inner simulation.

In this paper, we incorporate QMC methods within the inner simulation. Let Ω be the possible scenarios of the random variable ω. To fit MC or QMC framework, we assume that given $\omega = y$, X can be generated via the mapping

$$X(y) = \psi(\boldsymbol{U}; y),\ \ y \in \Omega, \tag{4}$$

for some function ψ, where $\boldsymbol{U} \sim \mathbb{U}([0,1)^d)$. As a result, the inner estimator (3) is replaced by

$$\hat{g}_m(y) = \frac{1}{m}\sum_{j=1}^{m} \psi(\boldsymbol{u}_j; y), \tag{5}$$

and $g(y) = \int_{[0,1)^d} \psi(\boldsymbol{u}; y)\mathrm{d}\boldsymbol{u}$. If $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ are iid uniform points over $[0,1)^d$, the approximation (5) refers to the MC method. If $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ are QMC points, which are deterministic points chosen from $[0,1)^d$ and are more uniformly distributed than random points, the approximation (5) refers to the QMC method.

The error of the QMC quadrature can be bounded by the Koksma-Hlawka inequality (see [14])

$$|g(y) - \hat{g}_m(y)| \le V_{\mathrm{HK}}\big(\psi(\cdot; y)\big) D^*(\boldsymbol{u}_1, \dots, \boldsymbol{u}_m), \tag{6}$$

where $V_{\mathrm{HK}}\big(\psi(\cdot; y)\big)$ is the variation of the integrand $\psi(\cdot; y)$ for given y in the sense of Hardy and Krause which measures the smoothness of $\psi(\cdot; y)$, and $D^*(\boldsymbol{u}_1, \dots, \boldsymbol{u}_m)$ is the star discrepancy which measures the uniformity of the points set $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$. The Koksma-Hlawka inequality (6) implies that for functions of finite variation, the convergence rate of the QMC approximation is determined by the factor $D^*(\boldsymbol{u}_1, \dots, \boldsymbol{u}_m)$, which is of order $O(m^{-1}(\log m)^d)$ for low discrepancy points.

There are various constructions of QMC point sets in the literature [20], such as digital nets and lattice rule point sets. In this paper, we use (t,d)-sequences in base $b \ge 2$. In practice, one uses RQMC points for ease of evaluating quadrature errors. RQMC retains the essential equi-distribution structure, while allowing a statistical error estimation based on independent replications. Moreover, RQMC is able to improve the rate of convergence for smooth integrands, such as the Owen's scrambling technique [26]. For a survey on RQMC, we refer to [13]. From now on, in the RQMC setting, the quadrature points $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ in (5) are from a digit scrambled (t,d)-sequence in base b so that $\mathbb{E}[\hat{g}_m(y)] = g(y)$.

It is well known that when $\psi(\cdot; y)$ in (5) is square integrable and $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ are iid uniform points over $[0,1)^d$, then the variance of the estimate $\hat{g}_m(y)$

$$\mathrm{Var}(\hat{g}_m(y)) = \frac{\mathrm{Var}(X|\omega = y)}{m}.$$

In the RQMC setting, the variance depends on the smoothness of the integrands. Some upper bounds for the variance in RQMC have been established under certain conditions.

Theorem 1. *If $\psi(\cdot;y)$ in (5) is of finite variation in the sense of Hardy and Krause for any $y \in \Omega$ and $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ are from a digit scrambled (t,d)-sequence in base b, then*

$$\mathrm{Var}(\hat{g}_m(y)) = O(m^{-2}(\log m)^{2d}).$$

The proof is provided in [26]. There is an improved result with a variance of $O(m^{-3}(\log m)^{d-1})$ for smooth enough integrands [27]. Furthermore, for any square integrable integrand, the scrambled net variance $\mathrm{Var}(\hat{g}_m(y))$ has a conservative upper bound $M\mathrm{Var}(X|\omega = y)/m$, where the constant M depends on t, d and b [28].

In summary, in the MC and RQMC settings, the results can be expressed in a unified form by

$$\mathrm{Var}(\hat{g}_m(y)) \leq \frac{\sigma^2(y)}{m^\eta}, \tag{7}$$

where the constant $\eta \geq 1$, the function $\sigma^2(\cdot) \geq 0$ and the logarithmic term is hidden for simplicity. If the integrand $\psi(\cdot;y)$ in (5) is sufficiently smooth and RQMC is applied, then $\eta \approx 2$ or even larger. For instance, Lemma 1 in Section 4 presents an example with $\eta = 2$ and $\sigma^2(y)$ being equal to a constant for any $y \in \Omega$.

The next theorem is a generalization of Proposition 1 in [2], and its proof is adapted from [2].

Theorem 2. *Suppose that for any $y \in \Omega$, the variance of $\hat{g}_m(y)$ satisfies (7). Let $f(\cdot)$ be the probability density function of $g(\omega)$. Assume the following:*

- *The joint density $p_m(x,y)$ of $g(\omega)$ and $m^{\eta/2}[\hat{g}_m(\omega) - g(\omega)]$ and partial derivatives $(\partial/\partial x)p_m(x,y)$ and $(\partial^2/\partial x^2)p_m(x,y)$ exist for each m and (x,y).*
- *For each $m \geq 1$, there exist functions $f_{i,m}(\cdot)$ such that*

$$\left|\frac{\partial^i}{\partial x^i} p_m(x,y)\right| \leq f_{i,m}(y), i = 0,1,2.$$

In addition,

$$\sup_{m\geq 1} \int |y|^r f_{i,m}(y)\mathrm{d}y < \infty, \;\; \text{for } i = 0,1,2 \text{ and } 0 \leq r \leq 4.$$

Then the bias of the nested estimator $\hat{\theta}_{n,m}$ in (2) asymptotically satisfies

$$|\mathbb{E}[\hat{\theta}_{n,m} - \theta]| \leq \frac{|\Theta'(c)|}{m^\eta} + O(m^{-3\eta/2}), \tag{8}$$

where c is a constant given in (1),

$$\Theta(c) = \frac{1}{2} f(c)\mathbb{E}[\sigma^2(\omega)|g(\omega) = c],$$

and $\sigma^2(y)$ is given in (7).

Proof. Let

$$\theta_m := \mathbb{E}[\hat{\theta}_{n,m}] = \mathbb{P}(\hat{g}_m(\omega) > c).$$

Note that

$$\theta_m = \int_{\mathbb{R}} \int_{c-ym^{-\eta/2}}^{\infty} p_m(x, y)dxdy,$$

so that

$$\theta_m - \theta = \int_{\mathbb{R}} \int_{c-ym^{-\eta/2}}^{c} p_m(x, y)dxdy.$$

Consider the Taylor series expansion of the density function $p_m(x, y)$ at $x = c$,

$$p_m(x, y) = p_m(c, y) + (x - c)\frac{\partial}{\partial x} p_m(c, y) + \frac{(x - c)^2}{2} \frac{\partial^2}{\partial x^2} p_m(\tilde{c}, y),$$

where $\tilde{c}$ is a number between c and x. From this expansion and the assumptions in the theorem, it follows that

$$\theta_m - \theta = \int_{\mathbb{R}} \frac{y}{m^{\eta/2}} p_m(c, y)dy - \int_{\mathbb{R}} \frac{y^2}{2m^\eta} \frac{\partial}{\partial x} p_m(c, y)dy + O(m^{-3\eta/2}). \quad (9)$$

For the first term on the right hand side of (9), we have

$$\begin{aligned} \int_{\mathbb{R}} \frac{y}{m^{\eta/2}} p_m(c, y)dy &= \frac{f(c)}{m^{\eta/2}} \int_{\mathbb{R}} y \cdot \frac{p_m(c, y)}{f(c)} dy \\ &= \frac{f(c)}{m^{\eta/2}} \mathbb{E}[m^{\eta/2}(\hat{g}_m(\omega) - g(\omega))|g(\omega) = c]. \end{aligned}$$

This term is zero because

$$\mathbb{E}[m^{\eta/2}(\hat{g}_m(\omega) - g(\omega))|g(\omega) = c] = \mathbb{E}[\mathbb{E}[m^{\eta/2}(\hat{g}_m(\omega) - g(\omega))|g(\omega) = c, \omega]] = 0.$$

For the second term on the right hand side of (9), we have

$$\begin{aligned} \int_{\mathbb{R}} \frac{y^2}{2m^\eta} \frac{\partial}{\partial x} p_m(c, y)dy &= \frac{1}{2m^\eta} \frac{d}{dc} \int_{\mathbb{R}} y^2 p_m(c, y)dy \\ &= \frac{1}{2m^\eta} \frac{d}{dc} \left[f(c) \int_{\mathbb{R}} y^2 \cdot \frac{p_m(c, y)}{f(c)} dy \right] \\ &= \frac{1}{2m^\eta} \frac{d}{dc} \left[f(c)\mathbb{E}[m^\eta(\hat{g}_m(\omega) - g(\omega))^2|g(\omega) = c] \right] \\ &= \frac{1}{2m^\eta} \frac{d}{dc} \left[f(c)\mathbb{E}[\mathbb{E}[m^\eta(\hat{g}_m(\omega) - g(\omega))^2|\omega]|g(\omega) = c] \right] \\ &= \frac{1}{2m^\eta} \frac{d}{dc} \left[f(c)\mathbb{E}[\sigma^2(\omega)|g(\omega) = c] \right] \\ &= \frac{\Theta'(c)}{m^\eta}. \end{aligned}$$

By (9), the proof is completed. □

We now use Theorem 2 to analyze the mean squared error (MSE) of the nested estimator. By (8), it is easy to see that

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_{n,m}) &= \frac{1}{n}\mathrm{Var}(\mathbb{I}(\hat{g}_m(\omega_1) > c)) = \frac{\theta_m(1-\theta_m)}{n} \\ &= \frac{\theta(1-\theta)}{n} + \frac{(\theta_m-\theta)(1-\theta_m)}{n} + \frac{\theta(\theta-\theta_m)}{n} \\ &\le \frac{\theta(1-\theta)}{n} + O(n^{-1}m^{-\eta}),\end{aligned}$$

where $\theta_m = \mathbb{P}(\hat{g}_m(\omega) > c)$. So we have

$$\begin{aligned}\mathbb{E}[(\hat{\theta}_{n,m}-\theta)^2] &= \mathrm{Var}(\hat{\theta}_{n,m}) + (\mathbb{E}[\hat{\theta}_{n,m}-\theta])^2 \\ &\le \frac{\theta(1-\theta)}{n} + \frac{\Theta'(c)^2}{m^{2\eta}} + O(m^{-5\eta/2}) + O(n^{-1}m^{-\eta}).\end{aligned}$$

To achieve an RMSE of ϵ, this suggests the optimal allocations $n = O(\epsilon^{-2})$ and $m = O(\epsilon^{-1/\eta})$. The total computational cost is then $O(\epsilon^{-2-1/\eta})$. In the MC setting where $\eta = 1$, the complexity is $O(\epsilon^{-3})$, see [2]. In the RQMC setting, it is possible to achieve $\eta \approx 2$, so that the complexity is around $O(\epsilon^{-5/2})$. Broadie et al. [1] used adaptive allocations in the inner simulation to reduce the nested MC down to $O(\epsilon^{-5/2})$. Giles and Haji-Ali [3] showed that in the original MLMC, the complexity is also $O(\epsilon^{-5/2})$. By incorporating the adaptive allocations idea of [1], Giles and Haji-Ali [3] showed that the complexity of MLMC can be reduced to $O(\epsilon^{-2}(\log\epsilon)^2)$. We next focus on using the framework of multilevel method to reduce the complexity of RQMC-based nested simulation.

2.2. MLMC estimators

We review briefly the basic idea of MLMC. Denote

$$P = \mathbb{I}\{g(\omega) > c\} = \mathbb{I}\{\mathbb{E}[X|\omega] > c\}.$$

Since $g(\omega) = \mathbb{E}[X|\omega]$ is intractable, we are not able to simulate the random variable P directly. Instead of dealing with P directly, we consider a sequence of random variables $P_0, P_1, \ldots$ with increasing approximation accuracy to P but with increasing cost per sample. The ℓth level approximation of P is defined as $P_\ell = \mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\}$, where $m_\ell = 2^{\ell+\ell_0}$ in our setting and $\ell_0 \ge 0$. Due to the linearity of expectation, we have the following telescoping representation

$$\mathbb{E}[P_L] = E[P_0] + \sum_{\ell=1}^{L}\mathbb{E}[P_\ell - P_{\ell-1}]. \tag{10}$$

Let Y_ℓ be a random variable satisfying $\mathbb{E}[Y_\ell] = \mathbb{E}[P_\ell - P_{\ell-1}]$ for $\ell \ge 1$ and $Y_0 = P_0$, we further have

$$\mathbb{E}[P_L] = \sum_{\ell=0}^{L}\mathbb{E}[Y_\ell]. \tag{11}$$

The MLMC method uses the above equality and estimates each term on the right hand side of (11) independently. The resulting MLMC estimator is given by

$$\hat{\theta}_{\mathrm{MLMC}} = \sum_{\ell=0}^{L} Z_\ell$$

with

$$Z_\ell = \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} Y_\ell^{(i)},$$

where $Y_\ell^{(1)}, \dots, Y_\ell^{(N_\ell)}$ are iid replications of Y_ℓ for $\ell = 0, \dots, L$. Denote the variance and the computational cost of Y_ℓ by V_ℓ and C_ℓ, respectively. The MSE of $\hat{\theta}_{\mathrm{MLMC}}$ is then given by

$$\mathbb{E}[(\hat{\theta}_{\mathrm{MLMC}} - \theta)^2] = \sum_{\ell=0}^{L} \frac{V_\ell}{N_\ell} + (\mathbb{E}[P_L - \theta])^2.$$

The total cost of $\hat{\theta}_{\mathrm{MLMC}}$ is $C = \sum_{\ell=0}^{L} N_\ell C_\ell$.

The pioneering work by Giles [11] established the following theorem for MLMC.

Theorem 3. *If there are constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \min(\beta, \gamma)/2$ and*

- $|\mathbb{E}[P_\ell - \theta]| \leq c_1 m_\ell^{-\alpha}$,
- $V_\ell \leq c_2 m_\ell^{-\beta}$,
- $C_\ell \leq c_3 m_\ell^{\gamma}$,

then there exists a constant $c_4 > 0$ such that for any $\epsilon < 1/e$, MLMC estimator $\hat{\theta}_{\mathrm{MLMC}}$ has an MSE bound $\mathbb{E}[(\hat{\theta}_{\mathrm{MLMC}} - \theta)^2] \leq \epsilon^2$ with a total computational cost C bounded by

$$C \leq \begin{cases} c_4\epsilon^{-2}, & \beta > \gamma, \\ c_4\epsilon^{-2}(\log\epsilon)^2, & \beta = \gamma, \\ c_4\epsilon^{-2-(\gamma-\beta)/\alpha}, & \beta < \gamma. \end{cases}$$

The constants α and $\beta/2$ in Theorem 3 describe the rates of bias and variance decreasing, respectively, which are usually called *weak convergence* and *strong convergence rates*. The constant γ controls the increasing of computing budget for each sample.

Now we develop our nested multilevel QMC estimator. Here we use the following form for coupling the two consecutive levels: $Y_0 = P_0 = \mathbb{I}\{\hat{g}_{m_0}(\omega) > c\}$, and for $\ell \geq 1$,

$$Y_\ell = P_\ell - P_{\ell-1} = \mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \mathbb{I}\{\hat{g}_{m_{\ell-1}}(\omega) > c\}, \tag{12}$$

where

$$\hat{g}_{m_\ell}(y) = \frac{1}{m_\ell}\sum_{j=1}^{m_\ell} \psi(\boldsymbol{u}_j; y),$$

and ψ is given by (4).

By (12), we have $\mathbb{E}[Y_\ell] = \mathbb{E}[P_\ell - P_{\ell-1}]$ for $\ell \geq 1$. Also, $C_\ell = O(m_\ell)$, i.e., $\gamma = 1$. Under the conditions in Theorem 2, we obtain $\alpha = \eta \geq 1/2$ in Theorem 3.

We next study the variance of Y_ℓ. Note that for $\ell \geq 1$, we have

$$\mathrm{Var}(Y_\ell) \leq 2\mathrm{Var}(P_\ell - P) + 2\mathrm{Var}(P_{\ell-1} - P), \tag{13}$$

since $\mathrm{Var}(A+B) \leq 2[\mathrm{Var}(A) + \mathrm{Var}(B)]$ for any random variables A, B with finite variances. It thus suffices to study the decay rate of $\mathrm{Var}(P_\ell - P)$, which can be described by the value of the constant β.

Assumption 1. *Assume that the variance of $\hat{g}_m(\omega)$ is bounded as in (7) and the random variable $|g(\omega) - c|/\sigma(\omega)$ admits a density denoted by $\rho(x)$, where $\sigma(\omega)$ is the square root of $\sigma^2(\omega)$ in (7). Moreover, assume that there exist constants $\rho_0 > 0$ and $x_0 > 0$ such that $\rho(x) \leq \rho_0$ for all $x \in [0, x_0]$.*

The next theorem is a generalization of Proposition 2.2 in [3], and its proof is adapted from [3].

Theorem 4. *Suppose that Assumption 1 is satisfied. Then*

$$\mathrm{Var}(P_\ell - P) \leq \mathbb{E}[(P_\ell - P)^2] = O(m_\ell^{-\eta/2}).$$

Proof. Let us start from

$$\begin{aligned}\mathbb{E}\left[(\mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \mathbb{I}\{g(\omega) > c\})^2 \middle| \omega\right] &= \mathbb{P}\left[|\mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \mathbb{I}\{g(\omega) > c\}| = 1 \middle| \omega\right]\\ &\leq \mathbb{P}\left[|\hat{g}_{m_\ell}(\omega) - g(\omega)| \geq |g(\omega) - c| \middle| \omega\right].\end{aligned}$$

By Chebyshev's inequality, we have

$$\begin{aligned}\mathbb{P}\left(|\hat{g}_{m_\ell}(\omega) - g(\omega)| \geq |g(\omega) - c| \middle| \omega\right) &\leq \min\left(1, |g(\omega) - c|^{-2}\mathrm{Var}(\hat{g}_{m_\ell}(\omega)|\omega)\right)\\ &\leq \min\left(1, |g(\omega) - c|^{-2}\frac{\sigma^2(\omega)}{m_\ell^\eta}\right).\end{aligned}$$

Taking expectation over ω yields

$$\begin{aligned}\mathbb{E}\left[(\mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \mathbb{I}\{g(\omega) > c\})^2\right] &\leq \int_0^\infty \min(1, x^{-2}m_\ell^{-\eta})\rho(x)dx\\ &\leq \rho_0 \int_0^\infty \min(1, x^{-2}m_\ell^{-\eta})dx + \min(1, x_0^{-2}m_\ell^{-\eta})\\ &\leq 2\rho_0 m_\ell^{-\eta/2} + x_0^{-2}m_\ell^{-\eta}.\end{aligned}$$

Hence, the variance is $O(m_\ell^{-\eta/2})$. $\square$

Based on (13) and Theorem 4, we have $\mathrm{Var}(Y_\ell) = O(m_\ell^{-\beta})$ with $\beta = \eta/2$. Since $\gamma = 1$, by Theorem 3, the total cost of MLQMC thus becomes

$$C = \begin{cases} O(\epsilon^{-2}), & \eta > 2, \\ O(\epsilon^{-2}(\log \epsilon)^2), & \eta = 2, \\ O(\epsilon^{-3/2-1/\eta}), & \eta \in [1, 2). \end{cases} \tag{14}$$

It is a common strategy to use an antithetic form for coupling the consecutive levels, i.e.,

$$Y_\ell = \mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \frac{1}{2}\left(\mathbb{I}\{\hat{g}^{(1)}_{m_{\ell-1}}(\omega) > c\} + \mathbb{I}\{\hat{g}^{(2)}_{m_{\ell-1}}(\omega) > c\}\right),$$

where

$$\hat{g}^{(i)}_{m_{\ell-1}}(\omega) = \frac{1}{m_{\ell-1}} \sum_{j=1+(i-1)m_{\ell-1}}^{im_{\ell-1}} \psi(\boldsymbol{u}_j; \omega), \ i = 1, 2.$$

Antithetic sampling can sometimes reduce variance apparently, see [6, 8]. But one should be cautious to adapt antithetic sampling when the underlying function is discontinuous. It is required that

$$\mathbb{P}(\hat{g}^{(1)}_{m_{\ell-1}}(\omega) > c|\omega) = \mathbb{P}(\hat{g}^{(2)}_{m_{\ell-1}}(\omega) > c|\omega)$$

in order to ensure the telescoping representation (10) under the RQMC scheme. This can be achieved if the first half RQMC points $\boldsymbol{u}_1, \dots, \boldsymbol{u}_{m_{\ell-1}}$ have the same joint distribution as that of the second half RQMC points $\boldsymbol{u}_{m_{\ell-1}+1}, \dots, \boldsymbol{u}_{m_\ell}$. As pointed out by Goda et al. [8], the well-known explicitly constructed (t, d)-sequences in base $b = 2$ satisfy this condition if using Owen's scrambling method [26]. However, antithetic sampling does not change the variance convergence rate in this problem because the discontinuity in the indicator function violates the differentiability requirements of the antithetic estimator.

3. Smoothed MLMC method

The numerical results in [8] show that the MLMC method performs remarkably when the outer function $f(\cdot)$ of the nested expectation $\mathbb{E}[f(g(\omega))]$ is continuous. This is verified by some numerical experiments with some continuous functions, such as $f(x) = \max(0, x)$ and $f(x) = x^2$, under some financial settings. So it can be argued that the difficulties in problem (1) are caused by the indicator function which is discontinuous.

The indicator function also leads to other effects. Recall that

$$Y_\ell = P_\ell - P_{\ell-1} = \mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\} - \mathbb{I}\{\hat{g}_{m_{\ell-1}}(\omega) > c\}.$$

The "fine" estimator $\hat{g}_{m_\ell}(\omega)$ and the "coarse" estimator $\hat{g}_{m_{\ell-1}}(\omega)$ conditioned on the same scenario ω have small differences in most situations, especially for large ℓ, making Y_ℓ different from zero only in a tiny proportion of the scenarios, which leads to a high-kurtosis phenomenon.

The kurtosis of Y_ℓ is defined by

$$\kappa_\ell = \frac{\mathbb{E}[(Y_\ell - \mathbb{E}[Y_\ell])^4]}{(\mathrm{Var}(Y_\ell))^2}.$$

High-kurtosis is a challenge for both MLMC and MLQMC. As shown in Hammouda et al. [29], the sample variance of $Y_\ell^{(1)}, \dots, Y_\ell^{(N_\ell)}$ has a standard error of

$$\sigma_\ell := \frac{\mathrm{Var}(Y_\ell)}{\sqrt{N_\ell}} \sqrt{\kappa_\ell - 1 + \frac{2}{N_\ell - 1}}.$$

And the high kurtosis makes it challenging to estimate V_ℓ accurately in deeper levels, since there are less samples in deeper levels under the MLMC scheme [29].

It should be noted that the high-kurtosis phenomenon is due to the structure of the indicator function. We thus propose to replace the indicator function with a smoother function. The smoothing strategy was used in [30], in which substantial analysis on the effect of smoothing is conducted. In this paper, we use the logistic sigmoid function

$$S(x) = \frac{1}{1 + \exp(-x)}, \ x \in \mathbb{R}.$$

The logistic sigmoid function is widely applied in traditional classification tasks, acting as an activation function in deep learning, see e.g. [31, 32]. Based on this function, we construct a family of sigmoid-like functions

$$S(x;k) = \frac{1}{1 + \exp(-kx)}, \ x \in \mathbb{R},$$

where $k > 0$. It is obvious that $S(x;k)$ has the following properties:

- $S(x;k)$ takes value in $(0,1)$ for any k;
- $S(x;k)$ is centrosymmetric about $S(0;k) = \frac{1}{2}$ for any k;
- $S(x;k)$ converges to $\mathbb{I}\{x > 0\}$ as $k \to \infty$ for any fixed x.

Figure 1 presents several examples of $S(x;k)$. It can be seen that the sigmoid-like functions approximate the indicator function. When the value of k gets larger, the approximation of $S(x;k)$ to $\mathbb{I}(x > 0)$ gets better.

Now we are ready to develop a smoothed MLMC method. Define the ℓth level approximation of $P = \mathbb{I}\{g(\omega) > c\}$ as

$$\tilde{P}_\ell = S_\ell^{(k_0,r)}(\hat{g}_{m_\ell}(\omega) - c),$$

instead of $P_\ell = \mathbb{I}\{\hat{g}_{m_\ell}(\omega) > c\}$ used in Subsection 2.2, where $S_\ell^{(k_0,r)}(x) = S(x; k_0 r^\ell)$, $k_0 > 0$ is a constant and $r > 1$ controls the rate of $S_\ell^{(k_0,r)}(x)$ converging to $\mathbb{I}\{x > 0\}$ as $\ell \to \infty$.

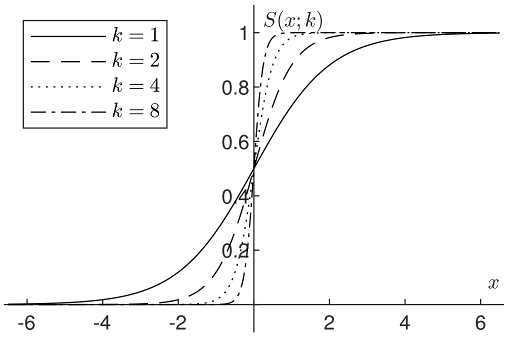

Figure 1: $S(x;k)$ with different values of k.

Similar to (10), we have the following telescoping representation

$$\mathbb{E}[\tilde{P}_L] = E[\tilde{P}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\tilde{P}_\ell - \tilde{P}_{\ell-1}].$$

Let $\tilde{Y}_\ell = \tilde{P}_\ell - \tilde{P}_{\ell-1}$ for $\ell \geq 1$ and $\tilde{Y}_0 = \tilde{P}_0$. It is obvious that

$$\tilde{P}_L = \sum_{\ell=0}^{L} \tilde{Y}_\ell,$$

and

$$\mathbb{E}[\tilde{P}_L] = \sum_{\ell=0}^{L} \mathbb{E}[\tilde{Y}_\ell].$$

This yields a new MLMC estimator

$$\tilde{\theta}_{\mathrm{MLMC}} = \sum_{\ell=0}^{L} \tilde{Z}_\ell$$

with

$$\tilde{Z}_\ell = \frac{1}{N_\ell} \sum_{j=1}^{N_\ell} \tilde{Y}_\ell^{(i)},$$

where $\tilde{Y}_\ell^{(1)}, \ldots, \tilde{Y}_\ell^{(N_\ell)}$ are iid replications of $\tilde{Y}_\ell$ for $\ell = 0, \ldots, L$. Notice that $\tilde{P}_L$ converges to P as L goes to infinity, just like P_L does. Denoting the variance and cost of each sample of $\tilde{Y}_\ell$ by $\tilde{V}_\ell$ and $\tilde{C}_\ell$, the MSE of $\tilde{\theta}_{\mathrm{MLMC}}$ is given by

$$\mathbb{E}[(\tilde{\theta}_{\mathrm{MLMC}} - \theta)^2] = \sum_{\ell=0}^{L} \frac{\tilde{V}_\ell}{N_\ell} + (\mathbb{E}[\tilde{P}_L - \theta])^2.$$

The total cost of $\tilde{\theta}_{\text{MLMC}}$ is $\tilde{C} = \sum_{\ell=0}^{L} N_\ell \tilde{C}_\ell$. It is critical to verify the strong convergence. To this end, we make the following assumption.

Assumption 2. *Let $\Pi_\ell(\cdot)$ denote the cumulative distribution function of the random variable $\hat{g}_{m_\ell}(\omega) - c$. Assume that the corresponding density functions, denoted by $\pi_\ell(\cdot)$ for $\ell = 0, 1, \dots,$ exist in a common neighborhood of the origin, say $(-\delta_0, \delta_0)$. Assume also that there exists a maximum*

$$\pi_\ell^* = \sup_{x \in (-\delta_0, \delta_0)} \pi_\ell(x) < +\infty$$

for all ℓ and there is a constant C_π such that $\sup_\ell \pi_\ell^* < C_\pi$.

The first part of Assumption 2 is similar to the assumptions in Theorem 2, but here we only require the densities to exist in a neighborhood of the origin. For the later part of Assumption 2, heuristically, we have $\lim_{\ell\to\infty} \pi_\ell(x) = \pi(x)$, where $\pi(x)$ is density of $g(\omega)$. Then we can expect that π_ℓ^* can be controlled by a constant C_π.

Theorem 5. *Suppose that Assumptions 1 and 2 are satisfied. Then*

$$\mathrm{Var}(\tilde{P}_\ell - P) = O(\max(r^{-\ell}, m_\ell^{-\eta/2})).$$

Proof. First, we have

$$\begin{aligned}
\mathbb{E}[(\tilde{P}_\ell - P_\ell)^2] &= \int_{-\infty}^{-\delta_0} [S_\ell^{(k_0,r)}(x)]^2 d\Pi_\ell(x) + \int_{-\delta_0}^{0} [S_\ell^{(k_0,r)}(x)]^2 \pi_\ell(x) dx \\
&\quad + \int_0^{\delta_0} [S_\ell^{(k_0,r)}(-x)]^2 \pi_\ell(x) dx + \int_{\delta_0}^{+\infty} [S_\ell^{(k_0,r)}(-x)]^2 d\Pi_\ell(x) \\
&\le 2[S_\ell^{(k_0,r)}(-\delta_0)]^2 + 2\pi_\ell^* \int_{-\delta_0}^{0} [S_\ell^{(k_0,r)}(x)]^2 dx \\
&= \frac{2}{[1+\exp(k_0 r^\ell \delta_0)]^2} + \left[\ln \frac{2}{1+e^{-k_0 r^\ell \delta_0}} + \frac{1}{2} - \frac{1}{1+e^{-k_0 r^\ell \delta_0}}\right] \cdot \frac{2C_\pi}{k_0 r^\ell} \\
&= O(r^{-\ell}),
\end{aligned}$$

where we used the facts that $\exp(r^\ell) \ge 1 + r^\ell$ and

$$\ln \frac{2}{1+e^{-k_0 r^\ell \delta_0}} + \frac{1}{2} - \frac{1}{1+e^{-k_0 r^\ell \delta_0}} = O(1)$$

for any given k_0, δ_0 and r.

By Theorem 4, we have $\mathbb{E}[(P_\ell - P)^2] = O(m_\ell^{-\eta/2})$. It then follows that

$$\mathbb{E}[(\tilde{P}_\ell - P)^2] \le 2\mathbb{E}[(\tilde{P}_\ell - P_\ell)^2] + 2\mathbb{E}[(P_\ell - P)^2] = O(\max(r^{-\ell}, m_\ell^{-\eta/2})).$$

□

Similarly, we can verify the weak convergence of the smoothed MLMC method.

Theorem 6. *Suppose that the conditions of Theorem 2 and Assumption 2 are satisfied. Then*

$$|\mathbb{E}[\tilde{P}_\ell - \theta]| = O(\max(r^{-\ell}, m_\ell^{-\eta})). \tag{15}$$

Proof. Note that

$$|\mathbb{E}[\tilde{P}_\ell - \theta]| \le \mathbb{E}[|\tilde{P}_\ell - P_\ell|] + |\mathbb{E}[P_\ell - \theta]|. \tag{16}$$

By Theorem 2, $|\mathbb{E}[P_\ell - \theta]| = O(m_\ell^{-\eta})$. Under Assumption 2, we have

$$\begin{aligned}
\mathbb{E}[|\tilde{P}_\ell - P_\ell|] &= \int_{-\infty}^{-\delta_0} S_\ell^{(k_0,r)}(x) d\Pi_\ell(x) + \int_{-\delta_0}^{0} S_\ell^{(k_0,r)}(x)\pi_\ell(x)dx \\
&\quad + \int_0^{\delta_0} S_\ell^{(k_0,r)}(-x)\pi_\ell(x)dx + \int_{\delta_0}^{+\infty} S_\ell^{(k_0,r)}(-x)d\Pi_\ell(x) \\
&\le 2S_\ell^{(k_0,r)}(-\delta_0) + 2\pi_\ell^* \int_{-\delta_0}^{0} S_\ell^{(k_0,r)}(x)dx \\
&= \frac{2}{1+\exp(k_0 r^\ell \delta_0)} + \ln\frac{2}{1+\exp(-k_0 r^\ell \delta_0)} \cdot \frac{2C_\pi}{k_0 r^\ell} \\
&= O(r^{-\ell}).
\end{aligned}$$

The result (15) follows from (16). □

For a given η, it is appropriate to take $r = 2^{\eta/2}$, and then $O(r^{-\ell}) \le O(m_\ell^{-\eta/2})$, since $m_\ell = O(2^\ell)$. As a result, the strong convergence of the smoothed method is the same as the original MLMC and MLQMC methods while the weak convergence may become $O(m_\ell^{-\eta/2})$ because of smoothing. When $\eta = 2$, we can take $r = 2$ such that there are constants c_2, c_3 such that $\tilde{V}_\ell \le c_2 m_\ell^{-1}$ and $\tilde{C}_\ell \le c_3 m_\ell$, which means $\beta = \gamma = 1$, then the cost is in the second regime in Theorem 3, where weak convergence makes little difference. So $r = 2$ is an alternative for this situation to get the expected improvement.

The next theorem can be obtained immediately from Theorem 3.

Corollary 1. *If for any $y \in \Omega$,*

$$\mathrm{Var}(\hat{g}_m(y)) \le \frac{\sigma^2(y)}{m^2}$$

and $r = 2$, then for any $\epsilon < 1/e$, the smoothed MLMC estimator $\tilde{\theta}_{\mathrm{MLMC}}$ has an MSE bound $\mathbb{E}[(\tilde{\theta}_{\mathrm{MLMC}} - \theta)^2] \le \epsilon^2$ with a total computational cost of $O(\epsilon^{-2}(\log \epsilon)^2)$.

As suggested by Theorem 1, QMC may suffer from the curse of dimensionality. When the dimension is relatively high, η may be smaller than 2 for RQMC. As a result, we should take a smaller r. Particularly, in the following numerical studies, we take $r = 2$ for one-dimensional problems and a conservative value $r = \sqrt{2}$ for multiple-dimensional problems. We take $k_0 = 8$ in

$\tilde{P}_\ell = S_\ell^{(k_0,r)}(\hat{g}_{m_\ell}(\omega) - c)$ so that the sigmoid function is closed to the indicator function at the coarse levels. The choice of r is rather heuristic, and how to determine an optimal value of r is open and left for future research.

With regard to the high-kurtosis phenomenon, it is considered to be caused by the noise in estimation of $\hat{g}_{m_\ell}(\omega)$ and $\hat{g}_{m_{\ell-1}}(\omega)$, where indicator function makes the estimation of Y_ℓ more unstable and then magnifies further the noise. On the contrary, the sigmoid functions are continuous in the neighborhood of the origin, then the estimations of $\tilde{Y}_\ell$ become less sensitive to the noise and more reliable.

4. Numerical study

In the numerical study we consider a portfolio that consists of European options written on d stocks whose price dynamics follow the Black-Scholes (B-S) model. For simplicity, we assume that the stock returns are the same, denoted by μ, while the risk-free interest rate is μ_0. Price dynamics of the stocks $\boldsymbol{S}_t = (S_t^1, \ldots, S_t^d)$ evolve according to

$$\frac{dS_t^i}{S_t^i} = \mu' dt + \sum_{j=1}^{d} \sigma_{ij} dW_t^i,\ i = 1, \ldots, d,$$

where $\mu' = \mu$ under the real-world probability measure, and $\mu' = \mu_0$ under the risk-neutral probability measure. Here, $\boldsymbol{W}_t = (W_t^1, \ldots, W_t^d)$ is a d-dimensional standard Brownian motion which represents d risk factors in the model and σ_{ij} are the corresponding volatility effects. We have,

$$S_t^i = S_0^i \exp\left\{ \left(\mu' - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2 \right) t + \sum_{j=1}^{d} \sigma_{ij} W_t^j \right\},\ i = 1, \ldots, d,$$

where $\boldsymbol{S}_0 = (S_0^1, \ldots, S_0^d)$ are the initial prices of stocks.

We assume that the maturities of all the European options in the portfolio are the same, denoted by T. We want to measure the portfolio risk at a future time τ ($\tau < T$). In the simulation, we first simulate the random variable $\omega = \boldsymbol{S}_\tau = (S_\tau^1, \ldots, S_\tau^d)$ under the real-world probability measure, which denotes the prices of stocks at the risk horizon τ as the outer sample. And then simulate $\boldsymbol{S}_T = (S_T^1, \ldots, S_T^d)$ under the risk-neutral probability measure which presents the prices of stocks at maturity T given ω as the inner samples. Denote by $V_0 = \sum_{i=1}^{d} v_0^i$ the initial value of the portfolio, which can be observed from the financial market. Then the portfolio value change is

$$g(\omega) := V_0 - \mathbb{E}[V_T(\boldsymbol{S}_T)|\omega],$$

where $V_T(\boldsymbol{S}_T)$ is the discounted payoff of the portfolio at time T which is a known function of $\boldsymbol{S}_T$.

Our target is to estimate the loss probability $\theta = \mathbb{P}(g(\omega) > c)$ for a given threshold c. We take $m_\ell = 32 \times 2^\ell$ in all experiments and $k_0 = 8$ for the smoothed methods. We compare MLMC, MLQMC, and smoothed MLQMC (denoted by SMLQMC) in each example.

4.1. Single asset

Firstly, we consider a portfolio consists of a single put option, i.e., $d = 1$. This example was studied in [1].

In the simulation, the outer random variable is generated according to

$$\omega = S_\tau = S_0 \exp\{(\mu - \sigma^2/2)\tau + \sigma\sqrt{\tau}Z\},$$

where the real-valued risk factor Z is a standard normal random variable. The portfolio value change is

$$g(\omega) = v_0 - \mathbb{E}[e^{-\mu_0(T-\tau)}(K - S_T(\omega, W))^+|\omega],$$

where the expectation is taken over a standard normal random variable W independently of Z, and $S_T(\omega, W)$ is given by

$$S_T(\omega, W) = \omega \exp\{(\mu_0 - \sigma^2/2)(T - \tau) + \sigma\sqrt{T-\tau}W\}.$$

Now let $X = v_0 - e^{-\mu_0(T-\tau)}(K - S_T(\omega, W))^+$. For $\omega = y$, X can be generated via

$$X(y) = \psi(u; y) = v_0 - e^{-\mu_0(T-\tau)}(K - y\exp\{(\mu_0 - \sigma^2/2)(T-\tau) + \sigma\sqrt{T-\tau}\Phi^{-1}(u)\})^+.$$

For $d = 1$, the variation of Hardy and Krause is the familiar total variation [33]. Since $\psi(u; y)$ is an increasing function with respect to u, we have

$$V_{\mathrm{HK}}(\psi(\cdot; y)) = \psi(1; y) - \psi(0; y) = e^{-\mu_0(T-\tau)}K.$$

As shown in [26], scrambling retains the net properties with probability one. When using a scrambled $(t, \ell, 1)$-net in base $b = 2$ in the inner simulation, by (6), we have

$$\mathrm{Var}(\hat{g}_m(y)) \le V_{\mathrm{HK}}(\psi(\cdot; y))^2 \mathbb{E}[D^*(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)^2] \le \frac{C}{m^2},$$

where C is a constant that does not depend on y, and the sample size has the form $m = 2^\ell$. Then we have the following lemma.

Lemma 1. *In the single asset situation under the B-S model, for any $y \in \Omega$, if a scrambled $(t, 1)$-sequence in base 2 is applied in the simulation, then*

$$\mathrm{Var}(\hat{g}_m(y)) \le \frac{\sigma^2(y)}{m^\eta}$$

is satisfied with $\eta = 2$ and $\sigma(y) \equiv C$.

Note that $g(y)$ is a strictly continuous function. So the cumulative distribution function of $g(\omega)$ can be computed easily, namely,

$$\begin{aligned}\mathbb{P}(g(\omega) \leq t) &= \mathbb{P}(\omega \leq g^{-1}(t)) \\ &= \mathbb{P}(S_0 \exp\{(\mu - \sigma^2/2)\tau + \sigma\sqrt{\tau}Z\} \leq g^{-1}(t)) \\ &= \Phi([\log(g^{-1}(t)/S_0) - (\mu - \sigma^2/2)\tau]/(\sigma\sqrt{\tau})).\end{aligned}$$

The density of $g(\omega)$ is then given by

$$f(t) = \frac{\phi([\log(g^{-1}(t)/S_0) - (\mu - \sigma^2/2)\tau]/(\sigma\sqrt{\tau}))}{\sigma\sqrt{\tau}g^{-1}(t)g'(g^{-1}(t))}.$$

It is easy to see that $f(0) < \infty$ so that Assumption 1 is satisfied. Then, by (14) we have the following result.

Corollary 2. *The MLQMC gives a complexity of $O(\epsilon^{-2}(\log \epsilon)^2)$ under the B-S model.*

The parameters in this case are set as follows: $S_0 = 100$, $\mu = 8\%$, $\mu_0 = 3\%$, the annualized volatility $\sigma = 20\%$, the strike of the put option $K = 95$, the maturity $T = 0.25$ years (i.e., three months), and the risk horizon $\tau = 1/52$ years (i.e., one week). With these parameters, the initial value of the put option is $v_0 = 1.669$ by the Black–Scholes formula. The inner expectation $g(\omega)$ can be computed explicitly by the Black-Scholes formula. We choose $c = 0.476887$ so that the loss probability $\theta = 0.3$. For this one-dimensional problem, we take $r = 2$ in the smoothed MLQMC. The results on testing the weak convergence and the strong convergence in Figure 2 are based on 500,000 outer samples in each level.

Figure 2(a) displays the weak convergence of the three methods, from which we can see that both crude MLQMC and smoothed MLQMC method attain larger α than the MLMC method. This implies that QMC methods enjoy a better weak convergence and converge to the true value faster.

Figure 2(b) shows the comparison of strong convergence, which plays a central role in MLMC methods. It is obvious that QMC methods work much better than MLMC, yielding the slopes of the lines lower than -1. Thus, QMC methods make the constant β exceed $\gamma = 1$, which means the variance V_ℓ decreases faster than the cost C_ℓ increases. And the smoothed MLQMC method shares the same rate of variance decreasing with the crude MLQMC method but achieves an even smaller variance. As a result, the costs in each level are reduced accordingly, which is confirmed by Figure 2(c).

Figure 2(d) displays the tendency of kurtosis. As we explained, the kurtosis increases with the variance decrease as ℓ gets larger. This phenomenon is inevitable due to the structure of the indicator function. But combined with the smoothed method, the increase of kurtosis slows down significantly, which means the smoothed MLQMC method takes effect on the high-kurtosis phenomenon.

As we stated earlier, QMC methods are able to accelerate the strong convergence, leading to $\beta > \gamma = 1$ in Theorem 3. Then the total cost falls in

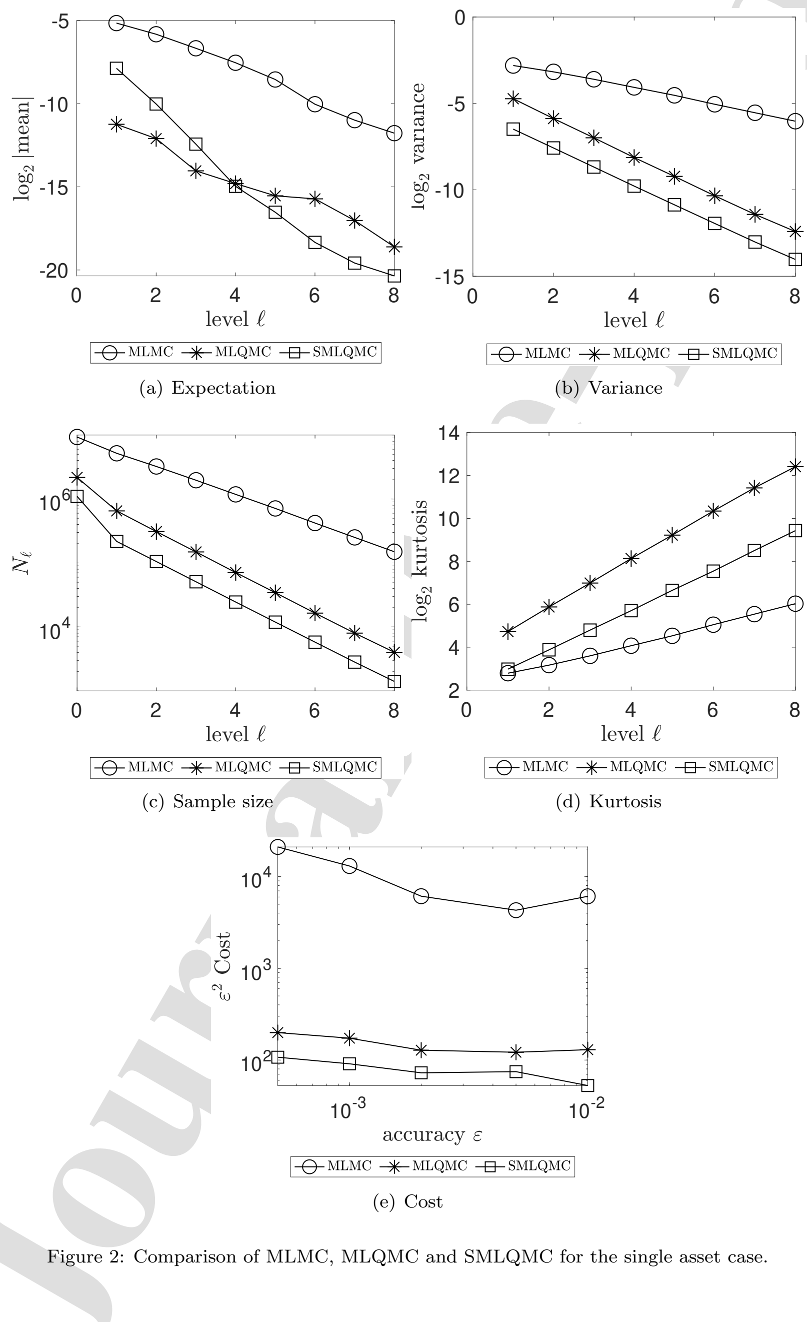

Figure 2: Comparison of MLMC, MLQMC and SMLQMC for the single asset case.

the third regime in Theorem 3, which is the optimal complexity bound $O(\epsilon^{-2})$ for MLMC methods. The numerical results support this in Figure 2(e). Both QMC methods are of the best complexity, while the smoothed MLQMC method reduces further the computational burden.

4.2. Multiple assets

Now we consider a portfolio consisting of d European call options, which was studied in [34]. Given the outer sample $\omega = \boldsymbol{S}_\tau = (S_\tau^1, \dots, S_\tau^d)$ which denotes the prices of stocks at the risk horizon τ under real-world measure, and $\boldsymbol{K} = (K^1, \dots, K^d)$, the strike prices for call options, the portfolio value change is

$$g(\omega) = V_0 - \mathbb{E}[e^{-\mu_0(T-\tau)} \sum_{i=1}^{d} (S_T^i(\omega^i, \boldsymbol{W}) - K^i)^+ | \omega],$$

where ω^i is the ith element of the vector ω and the expectation is taken over the random variable $\boldsymbol{W} = (W^1, \dots, W^d) \sim N(\boldsymbol{0}, \boldsymbol{I}_d)$, and samples of

$$S_T^i(\omega^i, \boldsymbol{W}) = \omega^i \exp\{(\mu_0 - \frac{1}{2}\sum_{j=1}^{d} \sigma_{ij}^2)(T-\tau) + \sum_{j=1}^{d} \sigma_{ij}\sqrt{T-\tau}W^j\}$$

are simulated under the risk-neutral measure.

To handle the effect of high dimensionality, we combine the gradient PCA (GPCA) method proposed by Xiao and Wang [35] within SMLQMC (smoothed MLQMC), which is abbreviated as GMLQMC hereafter. GPCA is a general method to reduce the effective dimension of functions for improving the efficiency of the QMC method. Here we apply GPCA in the inner simulation. We also examine the antithetic sampling of GMLQMC (abbreviated as AMLQMC). As we know, using an antithetic estimator does not change the variance convergence rate in the crude MLMC because of the indicator function. It is of interest to test whether antithetic sampling can get more benefit in the smoothed method.

The parameters in our experiments are set as follows: $S_0^1 = \cdots = S_0^d = 100$, $\mu = 8\%$, $\mu_0 = 5\%$, the strikes $K^1 = \cdots = K^d = 95$, the maturity $T = 0.1$, and the risk horizon $\tau = 0.02$. Without loss of generality, we let $\Sigma = (\sigma_{ij})$ be a sub-triangular matrix satisfying $C = \Sigma\Sigma^T$ which corresponds with Cholesky decomposition of C, where $C_{ij} = 0.3 \cdot 0.98^{|i-j|}$. The decomposition of C is not unique, and have an impact on the efficiency of QMC methods. The crude MLQMC method takes the Cholesky decomposition of C. To handle the high-dimensionality, the GMLQMC method takes a specific decomposition of C due to Xiao and Wang [35]. We set the threshold $c = 20\% V_0$, and take $r = \sqrt{2}$ in the smoothed MLQMC methods.

Firstly, we perform the convergence tests for estimating α and β in Theorem 3. The results are based on 300,000 outer samples in each level. Figure 3 shows the behavior of the absolute values of the expectations of Y_ℓ for crude methods and $\tilde{Y}_\ell$ for smoothed methods. It can be seen that QMC methods

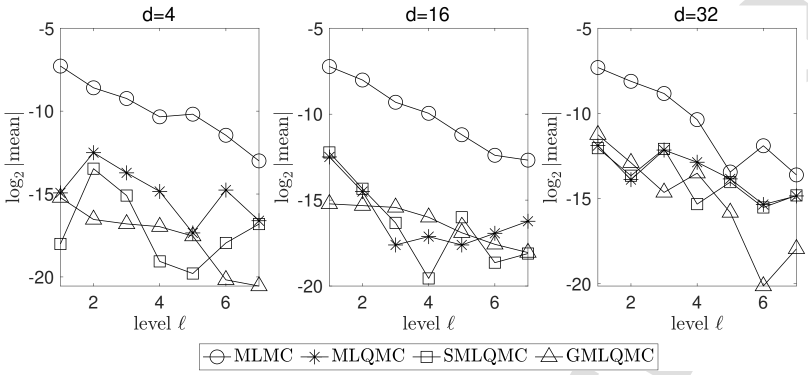

Figure 3: Estimations of $|\mathbb{E}[Y_\ell]|$ (MLMC and MLQMC) and $|\mathbb{E}[\tilde{Y}_\ell]|$ (SMLQMC, GMLQMC and AMLQMC) for $\ell = 1, \ldots, 7$.

have smaller values but with more volatility. The high volatility is due to the high-kurtosis phenomenon. There is only a tiny proportion of samples different from zero and the high sensitivity makes it difficult to predict the value of α, even with such a size of outer samples.

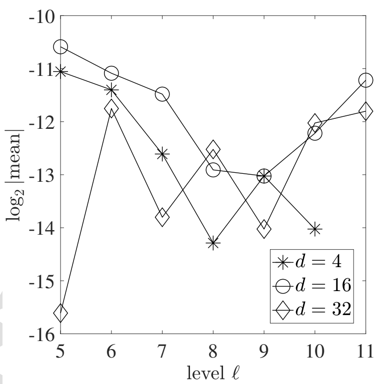

Figure 4: Estimations of $|\mathbb{E}[Y_\ell]|$ for MLMC when $\ell = 5, \ldots, 11$.

Not only MLQMC methods are faced with this difficulty, but also the MLMC method is. Figure 4 shows the results of the MLMC method in deep levels. Under the same magnitude, both MLMC and MLQMC have large volatility. But it can be observed in Figure 3 that the rates of the MLQMC methods are not worse than the MLMC method and the value of α does not matter here actually for the strong convergence being good enough. Nevertheless, it should be noted that the smoothed MLQMC methods are affected less by the discontinuous structure and look more stable.

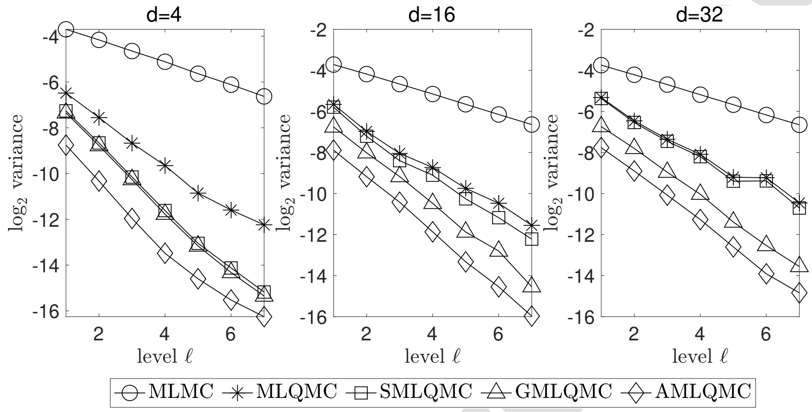

Figure 5: Estimations of $\mathrm{Var}(Y_\ell)$ for original methods (MLMC and MLQMC) and $\mathrm{Var}(\tilde{Y}_\ell)$ for smoothing methods (SMLQMC, GMLQMC and AMLQMC).

Figure 5 shows the empirical variances of Y_ℓ and $\tilde{Y}_\ell$ for different levels, which can be used to predict the strong convergence rate by the usual linear regression. For the plain MLMC, we observe $\beta \approx 0.5$ for all dimensions. For the crude MLQMC method, the dimension d has an impact on β, and $\beta \approx 1$ for moderately large $d = 32$. When combined with GPCA method in the MLQMC method (i.e., the GMLQMC method), we observe a larger β, usually 1.1. When antithetic sampling method is applied in smoothed MLQMC method, β is further improved to 1.2. These insights suggest that antithetic sampling can benefit from the smooth coupling, achieving a larger β. The strong convergence gets apparent improvement with QMC methods, leading to $\beta = \gamma$ or $\beta > \gamma$. The total cost can be improved to the sub-optimal case $O(\epsilon^{-2}(\log \epsilon)^2)$ or the optimal case $O(\epsilon^{-2})$.

Figure 6 shows the total computation cost of each method. We can see that the complexity bound of MLQMC methods is $O(\epsilon^{-2})$ or $O(\epsilon^{-2}\log(\epsilon)^2)$ as expected. The smooth coupling indeed helps to reduce the total cost.

As shown in Figures 5 and 6, the crude MLQMC method is good enough, almost attaining the optimal complexity, even for the large dimension $d = 32$. The effect of the dimension d on the performance of the crude MLQMC method seems minor. A possible explanation is that the integrand in the inner simulation may enjoy low effective dimension, which is beneficial for QMC method [36]. It is clear that the choice of the matrix Σ has an impact on the effective dimension of the integrand.

Now we examine another risk factors structure, with Σ satisfying $\tilde{C} = \Sigma\Sigma^T$, where $\tilde{C}_{ij} = 0.3(d - |i - j|)/d$. Figure 7 shows the decays of the variances of Y_ℓ and $\tilde{Y}_\ell$ with the new matrix Σ for $d = 32$ and $d = 64$. For these cases, the strong convergence rate β of the crude MLQMC methods is not much better

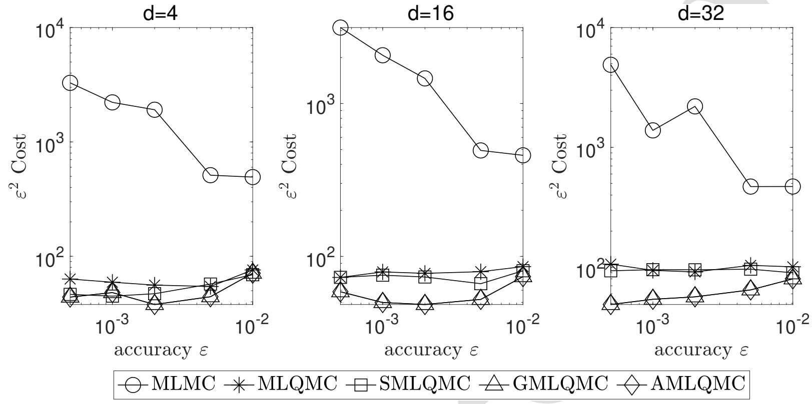

Figure 6: Tests of total cost.

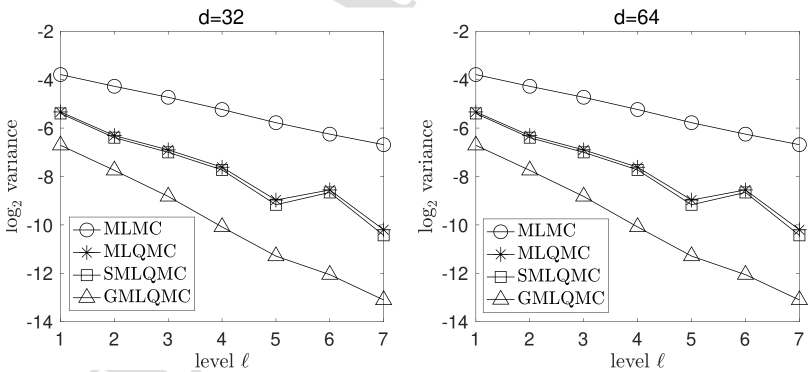

Figure 7: Tests of dimension effects.

than the MLMC method. But when it is combined with the dimension reduction technique GPCA, the strong convergence is improved dramatically, in which β exceeds 1. We thus conclude that the dimension of the integrand in the inner simulation has a significant impact on the performance of MLQMC methods. As the dimension gets higher, plain QMC methods may render a small $\eta < 2$ in (7). When combined with dimension reduction methods, such as the GPCA method, the efficiency of QMC methods can be reclaimed, achieving a better strong convergence. This highlights the importance of using dimension reduction methods in nested MLQMC.

5. Conclusion

In this paper, we have incorporated randomized QMC methods into MLMC to deal with financial risk estimation via nested simulation. We have proved that the new MLQMC estimator can achieve a better complexity bound under some assumptions. At the meantime, we discussed the high-kurtosis phenomenon caused by the character of RQMC points and discontinuity of the indicator function in the problem. Then we developed a new smoothed MLMC method and we proved that the smoothed MLMC method still reserves the advantages of MLMC without extra requirement. The smoothed method can also take advantage of antithetic sampling, which does not work for the original method. The superiority of the (smoothed) MLQMC methods over the original MLMC method has been empirically shown in numerical studies.

Further improvements for the MLQMC methods can be expected in the follow respects. On the one hand, we only applied RQMC methods in the inner simulation. If RQMC methods are used not only in the inner sampling but also in the outer sampling, the computational cost may be further reduced. And the smooth coupling estimator is more suitable for RQMC methods and an improved efficiency can be expected. On the other hand, we take a fixed number of inner samples in an identical level in this paper. If we are able to allocate the sample size according to the outer samples adaptively under the QMC scheme, like the way in the Monte Carlo scheme [1, 3], a further improvement can be also expected.

Acknowledgments

The authors gratefully thank anonymous reviewers, whose suggestions and comments helped to improve the paper greatly. ZH would like to appreciate Prof. Micheal B. Giles and Dr. Abdul-Lateef Haji-Ali for useful discussions and comments at an early stage of this research.

References

[1] M. Broadie, Y. Du, C. C. Moallemi, Efficient risk estimation via nested sequential simulation, Manag. Sci. 57 (2011) 1172–1194. URL: https://www.jstor.org/stable/25835766.

[2] M. B. Gordy, S. Juneja, Nested simulation in portfolio risk measurement, Manag. Sci. 56 (2010) 1833–1848. URL: https://www.jstor.org/stable/40864742.

[3] M. B. Giles, A.-L. Haji-Ali, Multilevel nested simulation for efficient risk estimation, SIAM/ASA J. Uncertain. Quantif. 7 (2019) 497–525. doi:10.1137/18M1173186.

[4] K. Bujok, B. M. Hambly, C. Reisinger, Multilevel simulation of functionals of Bernoulli random bariables with application to basket credit derivatives, Methodol. Comput. Appl. Probab. 17 (2015) 579–604. doi:10.1007/s11009-013-9380-5.

[5] M. B. Giles, MLMC for nested expectations, in: Contemporary Computational Mathematics-A Celebration of the 80th Birthday of Ian Sloan, Springer, 2018, pp. 425–442.

[6] M. B. Giles, T. Goda, Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI, Stat. Comput. 29 (2019) 739–751. doi:10.1007/s11222-018-9835-1.

[7] T. Goda, T. Hironaka, T. Iwamoto, Multilevel Monte Carlo estimation of expected information gains, Stoch. Anal. Appl. 38 (2020) 581–600. doi:10.1080/07362994.2019.1705168.

[8] T. Goda, D. Murakami, K. Tanaka, K. Sato, Decision-theoretic sensitivity analysis for reservoir development under uncertainty using multilevel quasi-Monte Carlo methods, Comput. Geosci. 22 (2018) 1009–1020. doi:10.1007/s10596-018-9735-7.

[9] S. Heinrich, Monte Carlo complexity of global solution of integral equations, J. Complexity 14 (1998) 151–175. doi:10.1006/jcom.1998.0471.

[10] M. B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (2008) 607–617. URL: https://www.jstor.org/stable/25147215.

[11] M. B. Giles, Multilevel Monte Carlo methods, Acta Numer. 24 (2015) 259–328. doi:10.1017/S096249291500001X.

[12] J. Dick, F. Y. Kuo, I. H. Sloan, High-dimensional integration: The quasi-Monte Carlo way, Acta Numer. 22 (2013) 133–288. doi:10.1017/S0962492913000044.

[13] P. L'Ecuyer, C. Lemieux, Recent advances in randomized quasi-Monte Carlo methods, in: Modeling Uncertainty, Springer, 2002, pp. 419–474.

[14] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, Society for Industrial and Applied Mathematics, 1992.

[15] P. L'Ecuyer, Quasi-Monte Carlo methods with applications in finance, Finance Stoch. 13 (2009) 307–349. doi:10.1007/s00780-009-0095-y.

[16] M. B. Giles, B. J. Waterhouse, Multilevel quasi-Monte Carlo path simulation, Advanced Financial Modelling (2009) 165–181.

[17] F. Y. Kuo, R. Scheichl, C. Schwab, I. H. Sloan, E. Ullmann, Multilevel quasi-Monte Carlo methods for lognormal diffusion problems, Math. Comp. 86 (2017) 2827–2860. doi:10.1090/mcom/3207.

[18] F. Y. Kuo, C. Schwab, I. H. Sloan, Multi-level quasi-Monte Carlo finite element methods for a class of elliptic PDEs with random coefficients, Found. Comput. Math. 15 (2015) 411–449. doi:10.1007/s10208-014-9237-5.

[19] P. Robbe, D. Nuyens, S. Vandewalle, Recycling samples in the multigrid multilevel (quasi-)Monte Carlo method, SIAM J. Sci. Comput. 41 (2019) S37–S60.

[20] J. Dick, R. N. Gantner, Q. T. Le Gia, C. Schwab, Multilevel higher-order quasi-Monte Carlo Bayesian estimation, Math. Models Methods Appl. Sci. 27 (2017) 953–995. doi:10.1142/S021820251750021X.

[21] R. Scheichl, A. M. Stuart, A. L. Teckentrup, Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems, SIAM/ASA J. Uncertain. Quantif. 5 (2017) 493–518. doi:10.1137/16M1061692.

[22] Z. He, Quasi-Monte Carlo for discontinuous integrands with singularities along the boundary of the unit cube, Math. Comp. 87 (2018) 2857–2870. doi:10.1090/mcom/3324.

[23] Z. He, On the error rate of conditional quasi-Monte Carlo for discontinuous functions, SIAM J. Numer. Anal. 57 (2019) 854–874. doi:10.1137/18M118270X.

[24] Z. He, X. Wang, On the convergence rate of randomized quasi-Monte Carlo for discontinuous functions, SIAM J. Numer. Anal. 53 (2015) 2488–2503. doi:10.1137/15M1007963.

[25] F. Xie, Z. He, X. Wang, An importance sampling-based smoothing approach for quasi-Monte Carlo simulation of discrete barrier options, European J. Oper. Res. 274 (2019) 759–772. doi:10.1016/j.ejor.2018.10.030.

[26] A. B. Owen, Randomly permuted (t, m, s)-nets and (t, s)-sequences. in: Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (H. Niederreiter and P. J.-S. Shiue, Eds.), J. Complexity (1995) 299–317. doi:10.1007/978-1-4612-2552-2_19.

[27] A. B. Owen, Scrambling Sobol' and Niederreiter-Xing points, J. Complexity 14 (1998) 259–328. doi:10.1006/jcom.1998.0487.

[28] A. B. Owen, Scrambled net variance for integrals of smooth functions, Ann. Statist. 25 (1997) 1541–1562. URL: https://www.jstor.org/stable/2959062.

[29] C. Ben Hammouda, N. Ben Rached, R. Tempone, Importance sampling for a robust and efficient multilevel Monte Carlo estimator for stochastic reaction networks, Stat. Comput. 30 (2020) 1665–1689. doi:10.1007/s11222-020-09965-3.

[30] M. B. Giles, T. Nagapetyan, K. Ritter, Adaptive multilevel Monte Carlo approximation of distribution functions, arXiv preprint arXiv:1706.06869 (2017).

[31] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data mining, Inference, and Prediction, 2nd ed., New York: Springer, 2016.

[32] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature (London) 521 (2015) 436–444. doi:10.1038/nature14539.

[33] A. B. Owen, Multidimensional variation for quasi-Monte Carlo, in: Contemporary Multivariate Analysis And Design Of Experiments: In Celebration of Professor Kai-Tai Fang's 65th Birthday, World Scientific, 2005, pp. 49–74.

[34] L. J. Hong, S. Juneja, G. Liu, Kernel smoothing for nested estimation with application to portfolio risk measurement, Oper. Res. 65 (2017) 657–673. doi:10.1287/opre.2017.1591.

[35] Y. Xiao, X. Wang, Enhancing quasi-Monte Carlo simulation by minimizing effective dimension for derivative pricing, Comput. Econom. 54 (2017) 343–366. doi:10.1007/s10614-017-9732-2.

[36] X. Wang, On the effects of dimension reduction techniques on some high-dimensional problems in finance, Oper. Res. 54 (2006) 1063–1078. doi:10.1287/opre.1060.0334.