arXiv:2304.07797v1 [math.ST] 16 Apr 2023

无限时域下随机化无偏估计量的最优分布与自适应算法

Chao Zheng、Jiangtao Pan 和 Qun Wang

中国杭州，浙江财经大学数据科学学院 *

摘要

Rhee 和 Glynn（《运筹学》：63(5)，1026–1043，2015年）提出的随机化无偏估计量，在近似随机微分方程（SDE）相关路径泛函的期望时，可能具有很高的效率。然而，目前缺乏用于计算无限时域下最优分布的算法。本文基于 Cui 等人（《运筹学快报》：477–484，2021年）的方法，证明在温和的假设条件下，最优分布具有一种简单的表示形式。随后，我们开发了一种自适应算法，用于计算无限时域下的最优分布；该算法在先验估计中只需少量计算时间。最后，我们给出数值结果，以展示该自适应算法的效率。

关键词：随机化无偏估计量、最优分布、随机微分方程、自适应算法

AMS 学科分类（2000）：60H35、65C30、90C34、91G60

1 引言

如果没有解析解，蒙特卡罗方法可用于近似随机过程泛函的期望。对于标准蒙特卡罗

*电子邮箱：chao.zheng12@gmail.com；211510011019@zufe.edu.cn；wangqun@zufe.edu.cn

方法，当对底层随机过程的泛函进行精确抽样时，均方误差（MSE）的收敛速率为 $O(c^{-1})$，其中 c是计算成本。本文考虑遵循某些随机微分方程（SDE）的底层随机过程，这些过程难以精确生成。在这种情况下，我们可以采用时间离散格式（例如欧拉格式、米尔斯坦格式）来获得近似值。经典的时间离散格式见 Kloeden 和 Platen [13]。然而，这些与标准蒙特卡罗方法结合使用的时间离散格式通常会导致有偏的蒙特卡罗估计量。对于此类有偏估计量，收敛速率低于 $O(c^{-1})$；不过，通过根据计算成本 c仔细选择时间离散格式的步长和样本数，可以平衡蒙特卡罗估计量的偏差和方差（参见 Duffie 和 Glynn[4]）。

Rhee 和 Glynn [15] 通过构造若干无偏蒙特卡洛估计量，取得了突破；这些估计量适用于底层随机过程由时间离散格式近似的情形。McLeish [14]）也研究了类似的思路。这些估计量恢复了收敛速率 $O(c^{-1})$，可视为 Giles 提出的多层蒙特卡罗估计量的无偏版本（参见 Giles [7]）。它们可以轻松结合任何在 L^2 范数下以足够高阶收敛的时间离散格式，因此有许多应用。此后，无偏估计量的思路已扩展到更复杂的情形。Glynn 和 Rhee [6] 研究了无偏估计量在马尔可夫链平衡期望中的应用。Blanchet 和 Glynn [1] 探讨了对期望函数进行精确模拟。Vihola [16] 提出了一类更一般的无偏估计量。Zheng 和 Glynn [18] 为无限分层无偏估计量建立了中心极限定理。

Rhee 和 Glynn [15] 构造无偏估计量时，需要确定随机变量 N 的最优分布。这是一个受某些约束条件限制的无限时域优化问题。Rhee 和 Glynn [15] 提出了一种 m截断动态规划算法，可在 $O(m^3)$ 次运算内求得 N 的最优分布。Cui 等人 [3] 通过提供一个复杂度为 $O(m)$ 的算法改进了这一结果。然而，上述算法仅适用于m截断最优分布，而非无限时域下的最优分布。在实际应用中，人们可以启发式地选择一个 m 值，运行算法求得 m截断最优分布，最后再启发式地选择尾部分布（或者干脆忽略它）。显然，这种方法在无限时域意义下可能并非最优。另一方面，Rhee 和 Glynn [15] 的算法以及 Cui 等人 [3] 的算法都依赖于对 m 样本方差的先验估计——

尽管选择较大的 m 值通常会使 m截断最优分布接近无限时域下的最优分布，但当 m 较大时，先验估计的计算成本可能很高。

本文针对这两个问题提出了解决方案。基于许多时间离散格式能够以足够高的阶在 L^2 范数下收敛这一事实，我们对这些格式的收敛行为提出了较弱的假设，并证明最优分布，尤其是最优尾部分布，具有简单的表示形式。证明基于 Cui 等人 [3] 提出的优化方法，并仔细分析了最优分布的一些结构。为了实际应用，我们开发了一种自适应算法，它是在 Cui 等人 [3] 的算法基础上扩展而来，通过加入 m 的自适应取值和最优尾部分布实现。在数值实验中，我们发现，较小的 m 值通常就足以得到无限时域下最优分布的高精度近似，从而大幅减少先验估计所需的计算时间。

本文其余部分安排如下：第 2 节回顾 Rhee 和 Glynn [15] 中的无偏估计量。第 3 节回顾 Cui 等人 [3] 提出的方法。第 4 节推导无限时域下的最优分布，并提出一种自适应算法。第 5 节报告数值结果，以说明我们算法的效率。最后，我们在第 6 节总结全文。

2 随机无偏估计量

设 $(X(t) : t \geq 0)$ 是以下随机微分方程的唯一解

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t),$$

其中 μ: $\mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ ，且 $(B(t) : t \geq 0)$ 是一个 m 维标准布朗运动。在许多应用中，需要计算期望$E[f(X)],$ ，其中 f 是 X 的泛函。

一般而言，难以精确生成 X 和 $f(X)$。因此，可以使用时间离散格式进行近似。最简单的近似是欧拉格式

$$X_h((j+1)h) = X_h(jh) + \mu(X_h(jh)) + \sigma(X_h(jh))(B((j+1)h) - B(jh)),$$

其中， h 是步长，且 $X_h(0) = X(0)$。这里， X_h 和 $f(X_h)$分别是 X 和 $f(X)$的近似。然而，典型的时间离散格式存在偏差，即$E(f(X_h)) \neq E(f(X))$，尽管当 h趋于 0 时， $E(f(X_h))$可能收敛到 $E(f(X))$。

在这种情况下，与时间离散格式相关的标准蒙特卡洛估计量的均方误差收敛速度低于 $O(c^{-1})$（参见 Duffie 和 Glynn[4]），其中 c 是计算成本。本节回顾 Rhee 和 Glynn [15], 提出的无偏估计量；在上述设定下，这些估计量能够达到标准收敛速度 $O(c^{-1})$。

设 L^2 为平方可积随机变量的希尔伯特空间，并令 $Y \in L^2$（即 $E(Y^2) < \infty$）。有限时间内可能难以生成 Y，但我们假设存在一个近似序列 $(Y_n : n \geq 0)$，可以在有限时间内生成，并且满足 $\lim_{x \to +\infty} E[(Y_n - Y)^2] = 0$。设 N 为一个取非负整数值的随机变量，它与 $Y_n, n \geq 0$ 相互独立，并令

$$Z = \sum_{n=0}^{N} \frac{\Delta_n}{P(N \geq n)}$$

其中 $\Delta_n = Y_n - Y_{n-1}$ 满足 $Y_{-1} = 0$，且 P 是 N 的概率。我们称 Z 为耦合和估计量。定理 2.1（引自 Rhee 和 Glynn 的定理 1 [15]）给出了 Z 成为无偏估计量的条件：

定理 2.1. 若

$$\sum_{n=1}^{\infty} \frac{E[(Y_n - Y)^2]}{P(N \geq n)} < \infty \tag{1}$$

则 $Z \in L^2$ 是 $E(Y)$ 的无偏估计量，并且

$$E(Z^2) = \sum_{n=0}^{\infty} \frac{v_n}{P(N \geq n)}$$

其中 $v_n = E[(Y_{n-1} - Y)^2] - E[(Y_n - Y)^2]$。

Rhee 和 Glynn [15] 提出了另一种无偏估计量 Z，其定义为 $E(Y)$

$$\widetilde{Z} = \sum_{n=0}^{N} \frac{\widetilde{\Delta}_n}{P(N \geq n)},$$

其中 $\Delta_n = Y_n - Y_{n-1}$。这里，(Y_n, Y_{n-1}) 与 (Y_n, Y_{n-1}) 具有相同的边缘分布，但对于每个 n，Δ_n 彼此独立。这个估计量称为独立和估计量。以下定理引自 Rhee 和 Glynn [15]，保证 Z 是无偏的。

~定理 2.2. 若式 (1) 成立，则 $Z \in L^2$ 是 $E(Y)$ 的无偏估计量。此外，

$$E(\widetilde{Z}^2) = \sum_{n=0}^{\infty} \frac{\widetilde{v}_n}{P(N \geq n)}$$

其中 $\widetilde{v}_n = var(Y_n - Y_{n-1}) + (E(Y) - E(Y_{n-1}))^2 - (E(Y) - E(Y_n))^2$。

此外，还引入了第三种无偏估计量，称为单项估计量；有关更多讨论，请参见 Rhee 和 Glynn [15]。本文重点讨论耦合和估计量。对独立和估计量的分析非常相似。

注意，必须指定 N 的分布。最低要求是选择满足式（1）的分布。为了最大化无偏估计量 Z 的效率，Rhee 和 Glynn [15] 提出寻找一种分布，使乘积 $E(\tau) \times var(Z)$ 最小。这里，τ 表示生成一个 Z 样本所需的计算时间，并且

$$E(\tau) = E\left(\sum_{n=0}^{N} t_n\right) = \sum_{n=0}^{\infty} t_n P(N \geq n),$$

其中，t_n 是计算 Y_n 所需的期望计算时间；对于方差 $var(Z)$，根据定理 2.1 可得

$$var(Z) = E(Z^2) - (E(Z))^2 = \sum_{n=0}^{\infty} \frac{v_n}{P(N \geq n)} - (E(Y))^2.$$

因此，最小化 $E(\tau) \times var(Z)$ 的问题可以写成如下优化问题：

$$\begin{aligned} \min_{F} \quad & g(F) := \left(\sum_{n=0}^{\infty} \frac{\beta_n}{F_n}\right)\left(\sum_{n=0}^{\infty} t_n F_n\right) \\ s.t. \quad & F_i \geq F_{i+1}, \forall i \geq 0 \\ & F_i > 0, \forall i \geq 0 \\ & F_0 = 1, \end{aligned} \tag{2}$$

其中 $\beta_0 = v_0 - (E(Y))^2$, $\beta_n = v_n$, $n \geq 1$ 和 $F_n = P(N \geq n)$。我们假设 $(\beta_i, i \geq 0)$ 是非负序列，且 $t_i, i \geq 0$ 有正的下界，因此该优化问题存在解；参见 Rhee 和 Glynn [15] 中的命题 2。

由于一般而言，没有简单的公式可以计算所有$n \geq 0$对应的最优 F_n ， Rhee 和 Glynn [15] 考虑了如下的有限版本优化问题：

$$\begin{aligned} \min_{F} \quad & g(F) := \left(\sum_{n=0}^{m} \frac{\beta_n}{F_n}\right)\left(\sum_{n=0}^{m} t_n F_n\right) \\ s.t. \quad & F_i \geq F_{i+1}, \forall i \geq 0 \\ & F_i > 0, \forall i \geq 0 \\ & F_0 = 1 \end{aligned} \tag{3}$$

对于任意 $m \in \mathbb{N}$，并随后提供了一种动态规划算法，用于计算阶数为 $O(m^3)$ 的最优 F 。我们将优化问题（3）的解称为“m-截断”最优分布。

然而，如果使用“m-截断”最优分布，而非无限时域下的最优分布，无偏估计量 Z 是否可能会损失效率，目前尚不明确。另一方面，尽管从启发式角度看，增大 m 的值可以使“m-截断”最优分布足够接近最优分布，但计算“m-截断”最优分布需要对 β_n、 $n = 0$、1、...、m 进行先验估计；当 m 较大时，这可能耗费大量时间。此外，在蒙特卡罗样本数量相同的情况下，以较高的 n 估计 β_n ，比以较低的 n 估计它耗时多得多。这带来了一个挑战：如何开发一种算法，以较低的先验估计计算成本处理最优分布。

3 M 截断最优分布

本节回顾 Cui 等人提出的方法[3] ，用于计算优化问题（3）中的“m-截断”最优分布。我们对无限时域最优分布的分析以及自适应算法的开发，均以他们的研究成果为基础。

Cui 等人 [3] 的主要方法是将原始优化问题转化为其对偶问题，其形式如下：

$$\begin{aligned} \min_{F} \quad & \sum_{n=0}^{m} \left(\frac{\beta_n}{F_n} + \mu t_n F_n\right) \\ s.t. \quad & F_i \geq F_{i+1}, \forall i \geq 0 \\ & F_i > 0, \forall i \geq 0 \end{aligned}, \tag{4}$$

~ ~ ~ ~其中 $\mu > 0$。设 $F = (F_0, F_1, ..., F_m)$ 为该问题的解。原始优化问题与其对偶问题之间的联系如下方命题 3.1 所示：

~命题 3.1. 若对于所有 n， β_n 和 t_n 均为正，则对偶问题（4）的解 F 存在，且对于任意 $\mu > 0$,

$$F^* = \frac{\widetilde{F}}{\widetilde{F}_0}$$

都是优化问题（3）的解。

证明. 参见 Cui 等人 [3] 中的定理 2.2 和定理 2.3，以及补充材料中的相关证明。 □

命题 3.1 表明，在~问题（3）中计算 m-截断最优分布，实际上等同于在问题（4）中计算 F。Cui 等人~[3] 提出了一种计算 F 的算法。该算法的计算复杂度为 $O(m)$，优于 Rhee 和 Glynn [15] 提出的动态规划算法，其计算复杂度为 $O(m^3)$。

本文始终假设定理 3.1 中的假设成立（即对于所有 n， β_n 和 t_n 均为正）。令

$$f_n(F_n) = \frac{\beta_n}{F_n} + \mu t_n F_n.$$

容易证明， f_n 是凸函数，且 f_n 的最小值可通过求解方程 $f'(x_n) = 0$ 在$x_n :$ $= \sqrt{\frac{\beta_n}{\mu t_n}}$处取得。接下来考虑一个稍微复杂一些的双变量问题：

$$\begin{aligned} \min_{F_n, F_{n+1}} \quad & [f_n(F_n) + f_{n+1}(F_{n+1})] \\ s.t. \quad & F_n \geq F_{n+1} > 0, \end{aligned}$$

并令其解为 $(\widetilde{F}_n, \widetilde{F}_{n+1})$。令 $(x_n, x_{n+1}) = \left(\sqrt{\frac{\beta_n}{\mu t_n}}, \sqrt{\frac{\beta_{n+1}}{\mu t_{n+1}}}\right)$。若 $x_n > x_{n+1}$, ~ ~ ~ ~, 则 $(F_n, F_{n+1}) = (x_n, x_{n+1})$；否则，我们得到 $(F_n, F_{n+1}) = (x^*, x^*)$，其中 x^* 是以下问题的解

$$\min_x [f_n(x) + f_{n+1}(x)],$$

且 $x_n \le x^* \le x_{n+1}$。具体而言， x^* 满足方程 $f_n'(x^*) + f_{n+1}'(x^*) = 0$，因此

$$x^* = \sqrt{(\beta_n + \beta_{n+1}) / (\mu t_n + \mu t_{n+1})}.$$

$\sim$ $\sim$上述计算解（F_n, F_{n+1}）的步骤表明，当 $x_n \le x_{n+1}$ 时，我们会将两个函数 f_n 和 f_{n+1} “合并”为一个便于分析的单一函数（即令 $f_{n,\ n+1} := f_n + f_{n+1}$，并重点关注 $\min_x \ f_{n,\ n+1}(x)$）。命题 3.2 为上述步骤及其推广提供了理论依据，该命题改编自 Cui 等人的命题 2.1 [3]。

命题 3.2. 设 $f_n(F_n) = \frac{\beta_n}{F_n} + \mu t_n F_n$。令 $(\widetilde{F}_s, .., \widetilde{F}_t)$ 为 $\min_{F_s,..,F_t} \sum_{n=s}^{t} f_n(F_n)$ 的解，满足 $F_s \ge ... \ge F_t > 0$；令 $(\widetilde{F}_{t+1}, .., \widetilde{F}_l)$ 为 $\min_{F_{t+1},..,F_l} \sum_{n=t+1}^{l} f_n(F_n)$ 的解，并满足类似的约束，其中 $s < l$。假设 $\sim$ $\sim$ $\sim$ $\sim$ $F_s = ... = F_t = V_1$ 且 $F_{t+1} = ... = F_l = V_2$，其中

$$V_1 = \sqrt{\frac{\sum_{n=s}^{t} \beta_n}{\mu \sum_{n=s}^{t} t_n}}, \quad V_2 = \sqrt{\frac{\sum_{n=t+1}^{l} \beta_n}{\mu \sum_{n=t+1}^{l} t_n}}.$$

如果 $V_1 \le V_2$，那么我们有

$$\bar{V} = \sqrt{\frac{\sum_{n=s}^{l} \beta_n}{\mu \sum_{n=l}^{l} t_n}} \in [V_1, V_2],$$

使得 $(F_s^*, .., F_l^*) = (\bar{V}, ..., \bar{V})$ 是满足约束条件 $F_s \ge ... \ge F_l > 0$ 的 $\min_{F_s, .., F_l} \sum_{n=s}^{l} f_n(F_n)$ 的解。

基于命题 3.2，Cui 等人 [3] 提出了一种计算问题（4）解的算法，然后利用命题 3.1 计算问题（3）中的 m-截断最优分布。

4 具有无限时域的最优分布

Rhee 和 Glynn 中的算法 [15] 以及 Cui 等人[3] 提出的算法都只能生成 m 截断的最优分布。本节着重推导具有无限时域的最优分布的简单表示，前提是 β_n 和 t_n 满足某些温和的假设。

在随机微分方程模型无偏估计量的背景下， Y_n 是使用步长通常为 $T/2^n$ 的时间离散格式对Y进行的近似，其中 $T > 0$ 为时间范围。由于 t_n 是生成Y_n所需计算时间的期望，因此显然 t_n 与 2^n成正比。因此，假设$t_n = 2^n$, $n = 0, 1, ...$ 是合理的。另一方面，对于一个以 $p > 1/2$ 阶强收敛的时间离散格式，即，

$$E[(Y_n - Y)^2] = O(2^{-2pn}),$$

将其与具有有限方差且计算时间有限的无偏估计量结合起来很容易。然后，根据 β_n 的定义，我们有 $\beta_n = O(2^{-2pn})$，这通常会在 n 足够大时得到 $\beta_n/\beta_{n+1} \approx 4^p$ 。现在，我们可以提出以下假设 4.1：

假设 4.1。假设对于所有 n， $t_n = 2^n$ 且 $\beta_n > 0$ 。假设存在 $m \geq 1$，使得对于所有 $n \geq m$， $4^p - \epsilon < \beta_n/\beta_{n+1} < 4^p + \epsilon$ ，其中 $p > 1/2$ 且 $\epsilon \in (0, 1)$。

假设 4.1 意味着对于所有 $n \geq m$， $\beta_n > \beta_{n+1} > \beta_{n+2} > ...$ 。本节中，我们始终假设假设 4.1 成立。回顾一下，无限时域优化问题（即问题（2））为

$$\begin{aligned} \min_{F} \quad & g(F) := \left(\sum_{n=0}^{\infty} \frac{\beta_n}{F_n}\right)\left(\sum_{n=0}^{\infty} t_n F_n\right) \\ s.t. \quad & F_i \geq F_{i+1}, \forall i \geq 0 \\ & F_i > 0, \forall i \geq 0 \\ & F_0 = 1. \end{aligned} \tag{5}$$

我们的目标是在假设 4.1 下推导出解的一个简单表示。设 $J := (L_i, i \geq 0)$ 为一个严格递增的整数值序列，使得$L_0 = 0$ 且 $\beta_i(J)/t_i(J)$ 是关于 i 的严格递减函数，其中

$$\beta_i(J) := \sum_{k=L_i}^{L_{i+1}-1} \beta_k, \quad t_i(J) := \sum_{k=L_i}^{L_{i+1}-1} t_k.$$

请注意， J 可能并不唯一，我们令 $\mathcal{J}$ 为所有 J组成的集合。我们考虑

问题（5）的以下子问题:

$$\begin{aligned} \min_{\overline{F}} \quad & g_J(\overline{F}) := \left(\sum_{n=0}^{\infty} \frac{\beta_n(J)}{\overline{F}_n}\right)\left(\sum_{n=0}^{\infty} t_n(J)\overline{F}_n\right) \\ s.t. \quad & \overline{F}_i \geq \overline{F}_{i+1}, \forall i \geq 0 \\ & \overline{F}_i > 0, \forall i \geq 0 \\ & \overline{F}_0 = 1. \end{aligned} \tag{6}$$

问题（6）可视为在附加约束下的问题（5）

$$F_{L_i} = F_{L_i+1} = ... = F_{L_{i+1}-1}, \quad \forall i \geq 0$$

与 J 相对应。对于任意 $J \in \mathcal{J}$，Rhee 和 Glynn [15] 中的命题1表明，问题（6）的解为

$$\overline{F}_i = \sqrt{\frac{\beta_i(J)/t_i(J)}{\beta_0(J)/t_0(J)}}, \quad \forall i \geq 0.$$

此外，设 F^*:=($F_i^*, i \geq 0$) 为问题（5）的解。Rhee 和 Glynn [15] 中的命题3和定理3表明，存在一个最优的 $J^* \in \mathcal{J}$，使得问题（5）的解就是与 $J = J^*$:=(L_i^*, $i \geq 0$) 对应的问题（6）的解，即该解为

$$F_{L_i^*}^* = F_{L_i^*+1}^* = ... = F_{L_{i+1}^*-1}^* = \overline{F}_i^* = \sqrt{\frac{\beta_i(J^*)/t_i(J^*)}{\beta_0(J^*)/t_0(J^*)}}, \quad \forall i \geq 0. \tag{7}$$

因此，寻找问题（5）的解等价于寻找最优的 J^*。

设 $\mathcal{J}_{(L_1,L_2,...,L_l)} \subseteq \mathcal{J}$，使得所有 $J \in \mathcal{J}_{(L_1,L_2,...,L_l)}$ 的前 $l+1$ 个元素分别为 $L_0, L_1, ..., L_l$，其中 $L_l \geq m$ 且 $\frac{\beta_{l-1}(J)}{t_{l-1}(J)} > \frac{\beta_{L_l}}{t_{L_l}}$。下面的命题 4.1 给出了 $\mathcal{J}_{(L_1,L_2,...,L_l)}$ 的一种表示。

命题 4.1. 假设满足假设 4.1。若 $\mathcal{J}^{(L_1,L_2,...,L_l)}$ 非空，则我们得到

$$\mathcal{J}^{(L_1,L_2,...,L_l)} = \{(L_0, L_1, L_2, ..., L_l, L_l + k_1, L_{l+1} + k_2, ...), k_i \in \mathbb{N}^+\}.$$

证明。根据定义，我们有

$$\mathcal{J}^{(L_1,L_2,...,L_l)} \subseteq \{(L_0, L_1, L_2, ..., L_l, L_l + k_1, L_{l+1} + k_2, ...), k_i \in \mathbb{N}^+\}.$$

因此，只需证明逆命题也成立。对于任意 $i \geq m$，我们有$\beta_i > \beta_{i+1} > ... > \beta_{i+k}$ 和 $t_i < t_{i+1} < ... < t_{i+k}$，其中 $k \in \mathbb{N}^+$，因此有

$$\frac{\beta_i}{t_i} > \frac{\sum_{j=i}^{i+k} \beta_j}{\sum_{j=i}^{i+k} t_j} > \frac{\beta_{i+k}}{t_{i+k}}.$$

那么，对于任意 $J = (L_0, L_1, L_2, ..., L_l, L_l + k_1, L_{l+1} + k_2, ...)$ $k_i \in \mathbb{N}^+$，我们得到

$$\frac{\beta_0(J)}{t_0(J)} > ... > \frac{\beta_{l-1}(J)}{t_{l-1}(J)} > \frac{\beta_{L_l}}{t_{L_l}} > \frac{\beta_l(J)}{t_l(J)} > \frac{\beta_{l+1}(J)}{t_{l+1}(J)} > ...$$

这意味着 $J \in \mathcal{J}^{(L_1,L_2,...,L_l)}$。证明完毕。 □

在所有 $J \in \mathcal{J}^{(L_1,L_2,...,L_l)}$ 中，最重要的是

$$J_*^{(L_1,L_2,...,L_l)} := (L_0, L_1, L_2, ..., L_l, L_l + 1, L_l + 2, ...).$$

为方便记号，我们将 $J^{(L_1,L_2,...,L_l,L_l+1,L_l+2,...)_*}$ 记作 J_*^L。

引理 4.1. 假设假设 4.1 成立，且 $\mathcal{J}_{(L_1,L_2,...,L_l)}$ 非空。那么对于所有 $J \in \mathcal{J}^{(L_1,L_2,...,L_l)}$，有

$$\min_{\overline{F}} g_J(\overline{F}) \geq \min_{\overline{F}} g_{J_*^L}(\overline{F}).$$

证明。对于 $J_*^L \in \mathcal{J}_{(L_1,L_2,...,L_l)} \subseteq \mathcal{J}$，将（7）代入（6）可得

$$\min_{\overline{F}} g_{J_*^L}(\overline{F}) = \left(\sum_{n=0}^{\infty} \sqrt{\beta_n(J_*^L) t_n(J_*^L)} \right)^2.$$

令 $J = (L_0, L_1, L_2, ..., L_l, L_l + 2, L_l + 3, ...)$。由于

$$\sqrt{(a_1 + a_2)(b_1 + b_2)} \geq \sqrt{a_1 b_1} + \sqrt{a_2 b_2}$$

对于任意 a_1、a_2、b_1、$b_2 \geq 0$，我们可以很容易地证明

$$\min_{\overline{F}} g_J(\overline{F}) = \left(\sum_{n=0}^{l-1} \sqrt{\beta_n(J_*^L) t_n(J_*^L)} + \sqrt{(\beta_l(J_*^L) + \beta_{l+1}(J_*^L))(t_l(J_*^L) + t_{l+1}(J_*^L))} \right.$$
$$\left. + \sum_{n=l+2}^{\infty} \sqrt{\beta_n(J_*^L) t_n(J_*^L)} \right)^2 \geq \min_{\overline{F}} g_{J_*^L}(\overline{F}).$$

上述不等式可以理解为：当我们合并 J_*^L 的第 l 项和第（$l+1$）项时，$\min_{\overline{F}} g_J(\overline{F})$ 的值会增大。

更一般地，对于任意 $k_0, k_1, ..., k_M \in \mathbb{N}^+$，令

$$J_1 = (L_0, L_1, ..., L_l, L_l + k_0, ..., L_{l+M} + k_M, L_{l+M} + k_M + 1, L_{l+M} + k_M + 2, ...),$$

and

$$J_2 = (L_0, L_1, ..., L_l, L_l + k_0, ..., L_{l+M} + k_M, L_{l+M} + k_M + 2, L_{l+M} + k_M + 3, ...),$$

其中 J_2 可视为将J_1的第（$l+M$）项和第（$l+M+1$）项合并而成。类似地，我们可以证明

$$\min_{\overline{F}} g_{J_2}(\overline{F}) \geq \min_{\overline{F}} g_{J_1}(\overline{F}).$$

注意，任意的 $J \in \mathcal{J}_{(L_1,\ L_2,\ ...L_l)}$ 都可以通过对相关项反复进行类似的合并（最多可数无限次）从 J_*^L得到，并且每次合并时，$\min_{\overline{F}} g_J(\overline{F})$ 的值都会增大。因此，我们可以得出，对于所有 $J \in \mathcal{J}_{(L_1,\ L_2,\ ...L_l)}$，$\min_{\overline{F}} g_J(\overline{F}) \geq \min_{\overline{F}} g_{J_*^L}(\overline{F})$，证明完毕。 □

注 4.1. 引理 4.1 与公式（7）表明，当 n 足够大时，最优分布 F_n^*的尾部与 $\sqrt{\frac{\beta_n}{t_n}}$ 成正比。由于 $\beta_n = O(2^{-2pn})$ 且 $t_n = O(2^n)$，我们得到最优分布 $F_n^* = O(2^{-(2p+1)n/2})$。特别地，当 $p = 1$ 时，我们得到 $F_n^* = O(2^{-3n/2})$。

下一个问题是，在已知 F_n 的最优“m-截断”分布的情况下，如何指定 F_n 的最优分布。答案见定理 4.1。

定理 4.1。设 F^* 是无限时域下问题（5）的解。设 F^*（m）是以下问题的解

$$\begin{aligned} \min_F \quad & g(F) := \left(\sum_{n=0}^{m} \frac{\beta_n}{F_n}\right)\left(\sum_{n=0}^{m} t_n F_n\right) \\ s.t. \quad & F_i \geq F_{i+1}, \forall i \geq 0 \\ & F_i > 0, \forall i \geq 0 \\ & F_0 = 1. \end{aligned} \tag{8}$$

假设假设 4.1 成立且 $F_{m-1}^*(m) \neq F_m^*(m)$。那么我们有

$$F_n^* = F_n^*(m), \quad n = 0, 1, 2, ..., m$$

$$F_{n+1}^* = F_n^* \sqrt{\frac{\beta_{n+1}}{2\beta_n}}, \quad n = m, m+1, ...$$

证明。设 $J^* = (L_1^*, ..., L_k^*, L_k^* + 1, L_k^* + 2, ...)$，其中 $(L_1^*, ..., L_k^*)$ 是问题（8）的最优解。条件 $F_{m-1}^*(m) \neq F_m^*(m)$ 意味着 $L_k^* = m$ 和 $\frac{\beta_{L_{k-1}}}{t_{L_{k-1}}} > \frac{\beta_m}{t_m}$。我们将证明 J^* 在无限意义下也同样是最优的。

假设 4.1 意味着，对于任意 $i \geq m$，我们有 $\beta_i > \beta_{i+1} > ... > \beta_{i+n}$ and $t_i < t_{i+1} < ... < t_{i+n}$，其中 $n \in \mathbb{N}^+$，因此成立

$$\frac{\beta_i}{t_i} > \frac{\sum_{j=i}^{i+n} \beta_j}{\sum_{j=i}^{i+n} t_j} > \frac{\beta_{i+n}}{t_{i+n}}.$$

因此，对于任意 $J = (L_i, i \geq 0) \in \mathcal{J}$，存在 $l \in \mathbb{N}^+$，使得 $L_l \geq m$ and $\frac{\beta_{l-1}(J)}{t_{l-1}(J)} > \frac{\beta_{L_l}}{t_{L_l}}$，即 $J \in \mathcal{J}_{(L_1, L_2, ..., L_l)}$。令 $J_*^L = (L_1, L_2, ..., L_l, L_l + 1, L_l + 2, ...)$。由命题 4.1 和引理 4.1 可得 $J_*^L \in \mathcal{J}_{(L_1, L_2, ..., L_l)}$ and

$$\min_{\overline{F}} g_J(\overline{F}) \geq \min_{\overline{F}} g_{J_*^L}(\overline{F}). \tag{9}$$

因此，只需比较 $\min_{\overline{F}} g_{J_*^L}(F)$ 和 $\min_{\overline{F}} \overline{g}_{J^*}(F)$。

回顾一下，对于该优化问题

$$\min_{\overline{F}} g_{J_*^L}(\overline{F}) = \left(\sum_{n=0}^{\infty} \frac{\beta_n(J_*^L)}{\overline{F}_n} \right) \left(\sum_{n=0}^{\infty} t_n(J_*^L) \overline{F}_n \right),$$

其解为

$$\overline{F}_i(J_*^L) := \sqrt{\frac{\beta_i(J_*^L)/t_i(J_*^L)}{\beta_0(J_*^L)/t_0(J_*^L)}}.$$

由于 $J_*^L = (L_1, L_2, ..., L_l, L_l + 1, L_l + 2, ...)$，我们得到递推公式

$$\overline{F}_n(J_*^L) = \overline{F}_l(J_*^L) \sqrt{\frac{\beta_n t_{L_l}}{t_n \beta_{L_l}}}$$

对于任意 $n > l$。令

$$F_{L_i}(J_*^L) = F_{L_i+1}(J_*^L) = ... = F_{L_{i+1}-1}(J_*^L) = \overline{F}_l(J_*^L), \quad \forall i \geq 0,$$

对应于 J_*^L，由此可得

$$F_n(J_*^L) = F_{L_i}(J_*^L) \sqrt{\frac{\beta_n t_{L_l}}{t_n \beta_{L_l}}}$$

对于任意 $n > L_i$，因此我们有

$$\begin{aligned}
&\min_{\overline{F}} g_{J_*^L}(\overline{F}) \\
&= \left(\sum_{n=0}^{L_l} \frac{\beta_n}{F_n(J_*^L)} + \sum_{n=L_l+1}^{\infty} \frac{\beta_n}{F_{L_l}(J_*^L)\sqrt{\frac{\beta_n t_{L_l}}{t_n \beta_{L_l}}}} \right) \\
&\quad \times \left(\sum_{n=0}^{L_l} t_n F_n(J_*^L) + \sum_{n=L_l+1}^{\infty} t_n F_{L_l}(J_*^L) \sqrt{\frac{\beta_n t_{L_l}}{t_n \beta_{L_l}}} \right) \\
&= \left(\sum_{n=0}^{L_l-1} \frac{\beta_n}{F_n(J_*^L)} + \frac{(1+c)\beta_{L_l}}{F_{L_l}(J_*^L)} \right) \left(\sum_{n=0}^{L_l-1} t_n F_n(J_*^L) + (1+c) t_{L_l} F_{L_i}(J_*^L) \right),
\end{aligned}$$

其中

$$c = \frac{\sum_{n=L_l+1}^{\infty} \sqrt{\beta_n t_n}}{\sqrt{\beta_{L_l} t_{L_l}}} < \infty.$$

对于 $J^* = (L_1^*, ..., L_k^*, L_k^* + 1, L_k^* + 2, ...)$，类似地可得

$$\begin{aligned}
&\min_{\overline{F}} g_{J^*}(\bar{F}) \\
&= \left(\sum_{n=0}^{L_l-1} \frac{\beta_n}{F_n(J^*)} + \frac{(1+c)\beta_{L_l}}{F_{L_l}(J^*)} \right) \left(\sum_{n=0}^{L_l-1} t_n F_n(J^*) + (1+c) t_{L_l} F_{L_i}(J^*) \right).
\end{aligned}$$

由于 $(L_1^*, ..., L_k^*)$ 是问题（8）的最优序列，且 $F_{m-1}^*(m) \neq F_m^*(m)$，因此有

$$\frac{\beta_{k-1}(J^*)}{t_{k-1}(J^*)} > \frac{\beta_m}{t_m} > ... > \frac{\beta_{L_l}}{t_{L_l}} = \frac{(1+c)\beta_{L_l}}{(1+c)t_{L_l}}.$$

然后，由命题 3.2 可知，$(L_1^*, ..., L_k^*, L_k^* + 1, ..., L_l)$ 是该问题的最优序列

$$\min_{F} \left(\sum_{n=0}^{L_l-1} \frac{\beta_n}{F_n} + \frac{(1+c)\beta_{L_l}}{F_{L_l}} \right) \left(\sum_{n=0}^{L_l-1} t_n F_n + (1+c) t_{L_l} F_{L_l} \right).$$

因此，我们有

$$\min_{\overline{F}} g_{J_*^L}(\overline{F}) \geq \min_{\overline{F}} g_{J^*}(\overline{F}). \tag{10}$$

因此，结合（9）和（10），我们得出，对于任意 $J \in \mathcal{J}$，都存在相应的 J_*^L，使得

$$\min_{\overline{F}} g_J(\overline{F}) \geq \min_{\overline{F}} g_{J_*^L}(\overline{F}) \geq \min_{\overline{F}} g_{J^*}(\overline{F}).$$

这表明，对于具有无限时域的问题（5），$J^* = (L_1^*, ..., L_k^*, L_k^* + 1, L_k^* + 2, ...)$ 是最优的。由于 $(L_1^*, ..., L_k^*)$ 对 m-截断问题（8）是最优的，因此我们得到，对于 $n = 0, 1, 2, ..., m$，有 $F_n^* = F_n^*(m)$，而对于 $m + 2, ...$，有 $F_{n+1}^* = F_n^* \sqrt{\frac{\beta_{n+1}}{2\beta_n}}$，其中 $n = m + 1$，由此完成证明。 □

定理 4.1 表明，我们可以利用一个 m 截断算法来求得最优分布 $F_n^*, n = 0, 1, .., m$，然后使用以下公式递归计算 $F_n^*, n = m + 1, m + 2, ...$

$$F_{n+1}^* = F_n^* \sqrt{\frac{\beta_{n+1}}{2\beta_n}} \approx 2^{\frac{-2p-1}{2}} F_n^*. \tag{11}$$

此处，m 的值可以从 $m = 1$ 开始自适应地确定，直至满足条件 $4^p - \epsilon \leqslant \beta_m/\beta_{m+1} \leqslant 4^p + \epsilon$ 和 $F_{m-1}^*(m) \neq F_m^*(m)$。在我们的设定中，有 $\beta = O(2^{-2pn})$，$p > 1/2$，因此当 n 足够大时，$\beta_n/\beta_{n+1} \approx 4^p$ 成立，这意味着 $4^p - \epsilon \leqslant \beta_n/\beta_{n+1} \leqslant 4^p + \epsilon$。此处，$\epsilon$ 控制近似（11）的精度。显然，ϵ 越小，近似就越精确。然而，如果 ϵ 太小，为满足条件 $4^p - \epsilon \leqslant \beta_n/\beta_{n+1} \leqslant 4^p + \epsilon, n \geq m$，就需要较大的 m，这会增加先验估计的计算成本。在实际应用中，我们发现，对于具有 $p = 1$ 的时间离散格式，$\epsilon = 0.5$ 通常已足够。另一个条件 $F_{m-1}^*(m) \neq F_m^*(m)$ 也很重要，它意味着

$$\frac{\beta_{k-1}(J^*)}{t_{k-1}(J^*)} > \frac{\beta_k(J^*)}{t_k(J^*)} = \frac{\beta_m}{t_m}$$

其中，$J^* = (L_1^*, ..., L_k^*)$ 是 m 截断分布的最优解。由于下面的引理，该条件很容易满足：

引理 4.2. 如果满足假设 4.1，且 $F_{m-1}^*(m) = F_m^*(m)$，则有

$$F_m^*(m + 1) \neq F_{m+1}^*(m + 1),$$

where $F^*(m + 1)$ 是 $(m + 1)$ 截断问题的最优分布 m.

证明。设 $J^* = (L_1^*, ..., L_k^*)$ 是 m 截断分布的最优解，其中 $L_k^* < m$ 由 $F_{m-1}^*(m) = F_m^*(m)$ 得出。由于 $\beta_k(J^*) = \sum_{i=L_k^*}^m \beta_i > \beta_m$ 并且

$$t_k(J^*) = \sum_{i=L_k^*}^m t_i = \sum_{i=L_k^*}^m 2^{-(m-i)} t_m < 2t_m,$$

我们有

$$\frac{\beta_k(J^*)}{t_k(J^*)} > \frac{\beta_m}{2t_m} > \frac{\beta_{m+1}}{t_{m+1}},$$

其中，$\beta_m > \beta_{m+1}$，这是由假设 4.1 得出的。因此，由命题 3.2 可知，$(L_1^*, ..., L_k^*, m+1)$ 对于具有 $F_m^*(m+1) \neq F_{m+1}^*(m+1)$ 的 $(m+1)$ 截断分布是最优的。

□

注 4.2。引理 4.2 表明，条件 $F_{m-1}^*(m) \neq F_m^*(m)$ 比条件 $4^p - \epsilon \leqslant \beta_n/\beta_{n+1} \leqslant 4^p + \epsilon, n \geq m$ 容易满足得多。在我们遇到的大多数情况下，如果后一个条件成立，前一个条件也成立。

基于以上讨论，我们现在可以给出一种自适应算法（算法 1），用于计算具有无限时域的最优分布 F^*。该算法扩展了 Cui 等人[3] 提出的算法，并采用自适应的 m 值。在算法 1 中，$BetaEstn(m)$ 是一个自定义函数，用于估计 β_m。我们只需估计 $\beta_n, n = 0, 1, ..., m, m+1$，而无需估计所有 β_n，从而节省大量计算时间。这里，输出 P_n 是当 $n =$ 为 0, 1, 2, ..., m 时的最优分布 F_n^*。条件 $L_k = R_k$（其中 $k = Index$）等价于 $F_{m-1}^*(m) \neq F_m^*(m)$。请注意，使用 Cui 等人[3], 提出的原始算法计算 m 截断的最优分布时，需要依次计算第 2、3、...、$(m-1)$ 个截断的最优分布。因此，对于我们的自适应算法，在 m 值相同的情况下，计算最优分布的计算成本与计算 m 截断的最优分布完全相同。

5 数值实验

本节评估自适应算法的效率。该算法与若干应用于随机微分方程模型的时间离散格式相结合，包括布莱克-斯科尔斯模型、Heston模型和Heston-Hull-White模型。这些模型在金融领域得到广泛应用。

我们考虑两种无偏估计量来近似期望值 $E(Y)$：耦合和估计量

$$Z = \sum_{n=0}^{N} \frac{\Delta_n}{P(N \geq n)},$$

其中 $\Delta_n = Y_n - Y_{n-1}$，以及独立和估计量

$$\widetilde{Z} = \sum_{n=0}^{N} \frac{\widetilde{\Delta}_n}{P(N \geq n)},$$

算法 1 最优分布的自适应算法

$k \leftarrow 0, Index \leftarrow -1$

对于 $m = 0$: 执行 10 次

$L_m \leftarrow m, R_m \leftarrow m, t_m \leftarrow 2^m$

$V_m \leftarrow (\beta_m / t_m)^{1/2}$

如果 $m = 0$, 则

$\beta_m = BetaEstn(m), \quad \beta_{m+1} = BetaEstn(m+1)$

else

$\beta_{m+1} = BetaEstn(m+1)$

结束如果

$Index \leftarrow Index + 1$

当 $k < Index$ 时执行

如果 $V_k \leqslant V_{k+1}$, 则

$R_k \leftarrow R_{k+1}$

$V_k \leftarrow [(\sum_{n=L_k}^{R_k} \beta_n)/(\sum_{n=L_k}^{R_k} t_n)]^{1/2}$

$L_{k+1}, R_{k+1}, V_{k+1} \leftarrow \emptyset$

如果 $k \neq 0$ 则

$k \leftarrow k - 1$

结束如果

else

$k \leftarrow k + 1$

结束如果

结束循环

如果 $m > 0$ 且 $\mid \beta_m / \beta_{m+1} - 4^p \mid < \epsilon$ 且 $L_k = R_k$ 则

跳出

结束 if

结束 for

$F_n \leftarrow V_k, n = L_k, ..., R_k, k = 0, 1, ..., Index$

$P_n = F_n / F_0, n = 0, 1, ..., m$

返回 m , $P_n, n = 0, 1, ..., m$

~ ~ ~其中 $\Delta_n = Y_n - Y_{n-1}$。如前所述，上述两个估计量的效率通过乘积 $E(\tau) \times var(Z)$ 来衡量，其中 $E(\tau)$ 是平均计算时间， $var(Z)$ 是无偏估计量 Z 的方差。此外，该乘积取决于概率 $F_n = P(N \geq n)$。我们比较 F_n 的三种选择，分别记为 $Dist1$、 $Dist2$ 和 $Dist3$。第一种选择为

$$Dist1 := (2^{-n(2p+1)/2} : n \geq 0),$$

它表示 Rhee 和 Glynn [15] 在第 4 节中引入的 N 的次典范分布。第二种选择为

$$Dist2 := (F_n^*(m) : m = 7, n \geq 0),$$

其中 $F_n^*(m)$ 是 N 在 $m = 7$ 条件下的 m-截断最优分布，通过 Cui 等人 [3] 中的原始算法计算得到，尾部分布遵循 $F_{n+1}^*(m) = F_n^*(m) \times 2^{-(2p+1)/2}$、 $n = m$、 $m + 1$， .....第三种选择为

$$Dist3 := (F_0^*, F_1^*, ..., F_m^*, F_m^* \times 2^{-(2p+1)/2}, ...),$$

其中， F_n^* 是无限时域下的最优分布，通过我们采用自适应 m的算法计算得到，并设 $\epsilon = 0.5$。

注意，对于无偏估计量 Z 和$\sim Z$ ，计算 $Dist2$ 和 $Dist3$ 需要对前 m 项进行先验估计

$$v_n = E[(Y_{n-1} - Y)^2] - E[(Y_n - Y)^2]$$

and

$$\widetilde{v}_n = var(Y_n - Y_{n-1}) + (E(Y) - E(Y_{n-1}))^2 - (E(Y) - E(Y_n))^2,$$

分别地，其中 Y_n 是基于步长为 T/n的时间离散格式对 Y所作的近似。由于通常难以精确生成 Y ，我们对$Y \approx Y_{10}$进行近似。对于量 v_n的估计，通常只需基于 5×10^5 个(Y_{n-1}, Y_n, Y)样本获得高精度估计；对于量 $\widetilde{v}_n$的估计，我们至少生成 10^6 个(Y_{n-1}, Y_n, Y)样本。在数值测试中，我们令 $Y = e^{-r}(S(1) - 1, 0)^+$，这对应于标准欧式看涨期权的收益。这里， S 是随机微分方程模型的路径， r 是利率。

我们的目标是研究哪种分布 F_n 对无偏估计量最有效，且在先验估计中的计算时间最短。所有数值实验均在 MATLAB 中完成。

5.1 布莱克-斯科尔斯模型

布莱克-斯科尔斯模型（Black 和 Scholes [2]）定义如下：

$$dS(t) = rS(t)dt + \sigma S(t)dB(t),$$

其中，r、$\sigma \in \mathbb{R}^+$ 和 $B = (B(t) : t \geq 0)$ 是标准布朗运动。我们采用米尔斯坦格式来近似该路径：

$$\begin{aligned} S_h((j+1)h) =& S_h(jh) + rS_h(jh)h + \sigma S_h(jh)(B((j+1)h) - B(jh)) \\ &+ \frac{1}{2}\sigma^2 S_h(jh)((B((j+1)h) - B(jh))^2 - h), \end{aligned}$$

其中，$j = 1$、2、...、$T/h - 1$，且 $S_h(0) = S(0)$。

表 1 比较了基于三种不同分布$Dist1$、$Dist2$ 和 $Dist3$ 的 N所得到的耦合和估计量 Z。其中，Var 表示样本量为 $n = 10^6$ 时 $\frac{1}{n}\sum_{i=1}^{n} Z_i$ 的方差，$time$ 则是计算这些量所需的实际计算时间（秒）。从表 1 可以看出，与 $Dist1$相比，我们的自适应算法得到的 $Dist3$ 具有更小的 $Var \times time$，因此计算效率更高。此外，由于分布非常相似，$Dist2$ 和 $Dist3$ 的 $Var \times time$也相近。然而，对于 $Dist3$，先验估计所需的计算时间要短得多，因为其对应的m 小得多。因此，我们更倾向于选择 $Dist3$，而非其他分布。

此外，我们看到 $\beta_n = O(2^{-2pn})$ 与 $p = 1$。对于至少 $m = 1$，满足停止条件 $F^*_{m-1}(m) \neq F^*_m(m)$ 和 $3.5 < \beta_m/\beta_{m+1} < 4.5$，因此 m 在 $Dist3$ 中的自适应值为 1。对于 $n > m$，我们可以验证条件 $3.5 < \beta_n/\beta_{n+1} < 4.5$ 也满足；因此，近似 $F^*_{n+1} = F^*_n\sqrt{\frac{\beta_{n+1}}{2\beta_n}} \approx 2^{-3/2}F^*_n$ 是合理的，这意味着上述停止条件与 $m = 1$ 足以保持我们自适应算法的稳健性。对于独立的 $\tilde{}$ 和估计量 Z，也有类似结论，如表 2 所示。

表 1 和表 2 中的数值示例并未涉及某些 n 的 $F^*_n = F^*_{n+1}$，因此我们的自适应算法所需的 m 值相当小。表 3 和表 4 展示了相反的情况，此时可能出现 $F^*_n = F^*_{n+1}$。我们看到，对于无偏估计量 $\tilde{}$ Z 和 Z，结论与先前的数值研究类似。尤其是，以 $var \times time$ 衡量的 $Dist3$ 的计算效率优于 $Dist1$。然而，在这种情况下，所需的 m 值可能很大（例如表 4 中的 $m = 5$）。

表1：耦合和估计量 Z；布莱克-斯科尔斯模型； $r = 0.05$, $\sigma = 0.20$, $T = 1$,$S(0) = 1$；样本量 10^6。

Z	m	var		$time$		$var \times time$	
分布1	$\diagdown$	2.21×10^{-8}		26.98		5.98×10^{-7}	
分布2	7	2.70×10^{-8}		11.91		3.21×10^{-7}	
Dist3	1	2.68×10^{-8}		11.92		3.19×10^{-7}	
n	0	1	2	3	4	5	6
β_n	0.0032	5.51×10^{-5}	1.48×10^{-5}	3.74×10^{-6}	1.08×10^{-6}	2.49×10^{-7}	5.73×10^{-8}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.0293	0.0108	0.0038	0.0015	0.0005	0.0002
Dist3	1.000	0.0293	0.0104	0.0037	0.0013	0.0005	0.0002

$\widetilde{\ }$表2：独立和估计量 Z；布莱克-斯科尔斯模型； $r = 0.05$, $\sigma = 0.20$,$T = 1$ $S(0) = 1$；样本量 10^6。

$\widetilde{Z}$	m	var		$time$		$var \times time$	
分布1	$\diagdown$	1.99×10^{-8}		24.92		4.96×10^{-7}	
分布2	7	2.40×10^{-8}		2.29		5.51×10^{-8}	
Dist3	1	2.41×10^{-8}		2.01		4.84×10^{-8}	
n	0	1	2	3	4	5	6
β_n	0.00305	2.62×10^{-5}	7.46×10^{-6}	2.12×10^{-6}	5.27×10^{-7}	1.43×10^{-7}	3.43×10^{-8}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.0207	0.0078	0.0029	0.0010	0.0004	0.0001
Dist3	1.000	0.0207	0.0073	0.0026	0.0009	0.0003	0.0001

表3：耦合和估计量 Z；布莱克-斯科尔斯模型； $r = 0.05, \sigma = 2, T = 1, S(0) = 1$；样本量 10^6。

Z	m	var		$time$		$var \times time$	
分布1	$\diagdown$	6.13×10^{-4}		27.28		0.0167	
分布2	7	1.25×10^{-4}		88.67		0.0111	
Dist3	2	1.37×10^{-4}		80.27		0.0110	
n	0	1	2	3	4	5	6
β_n	12.03	10.25	37.99	8.97	2.55	0.71	0.20
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.8175	0.8175	0.3053	0.1151	0.0430	0.0162
Dist3	1.000	0.8175	0.8175	0.2890	0.1022	0.0361	0.0128

~表 4：独立和估计量 Z；布莱克-斯科尔斯模型； $r = 0.05, \sigma = 2, T = 1 S(0) = 1$；样本量 10^6。

$\widetilde{Z}$	m	var		$time$		$var \times time$	
分布1	$\diagdown$	1.68×10^{-2}		25.05		0.4218	
分布2	7	4.53×10^{-4}		400.71		0.1815	
Dist3	5	5.25×10^{-4}		421.96		0.2215	
n	0	1	2	3	4	5	6
β_n	13.82	26.01	64.98	87.02	35.10	19.69	5.44
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	1.000	1.000	0.8523	0.3823	0.2027	0.0753
Dist3	1.000	1.000	1.000	0.8523	0.3823	0.2027	0.0717

5.2 Heston模型

Heston模型（Heston [11]）由下式给出

$$dS(t) = rS(t) + \sqrt{V(t)}S(t)dB_1(t)$$
$$dV(t) = k(\theta - V(t))dt + \sigma\sqrt{V(t)}dB_2(t)$$

其中 $r, k, \theta, \sigma \in \mathbb{R}^+$。这里， $B_1 = (B_1(t) : t \geq 0)$ 和 $B_2 = (B_2(t) : t \geq 0)$ 是相互独立的标准布朗运动。

如果模型参数满足 Feller边界条件 $2k\theta \geq \sigma^2$，则漂移隐式米尔斯坦格式可确保对 V 的近似非负。它可以与 Giles 和 Szpruch [8], 中的反变量截断Milstein方法结合，从而得到如下时间离散格式:

$$\ln(S_h((j+1)h)) = \ln(S_h(jh)) + (r - \frac{1}{2}V_h(jh))h + \sqrt{V_h(jh)}\Delta B_{1,j} + \frac{\sigma}{4}\Delta B_{1,j}\Delta B_{2,j}$$
$$V_h((j+1)h) = V_h(jh) + k(\theta - V_h((j+1)h))h + \sigma\sqrt{V_h(jh)}\Delta B_{2,j} + \frac{\sigma^2}{4}(\Delta B_{2,j}^2 - h)$$

其中 $j = 0$, 1, .., T/h。

相关结果见表5和表6。可以看出，对于~耦合估计量 Z 和独立和估计量 Z，分布 $Dist2$ 和 $Dist3$ 所得的 $var \times time$值相近，因为它们的分布非常相似，且其 $var \times time$ 小于 $Dist1$所得的结果。此外， $Dist3$ 先验估计的计算成本低于 $Dist2$。因此，我们更倾向于采用 $Dist3$，而不是 $Dist1$和 $Dist2$。

5.3 Heston-Hull-White模型

Heston-Hull-White模型定义如下:

$$dS(t) = r(t)S(t)dt + \sqrt{V(t)}S(t)(\rho dB_1(t) + \sqrt{1-\rho^2}dB_2(t))$$
$$dV(t) = k(\theta - V(t))dt + \sigma\sqrt{V(t)}dB_1(t)$$
$$dr(t) = \alpha(\beta - r(t))dt + \gamma dB_3(t)$$

其中， B_1、 B_2 和 B_3 是相互独立的布朗运动。这里，参数为 k、 θ、 σ、 α、 β、 $\gamma > 0$ 和 $\rho \in [-1, 1]$。该模型是在Heston模型基础上扩展而来，其中利率遵循Hull-White模型（Hull和White [12]）；更多详情参见Grzelak和Oosterlee [10] 。我们考虑一种半精确格式

表5：耦合和估计量 Z；Heston模型； $r = 0.05$, $\sigma = 0.25$, $k = 1$,$\theta = 0.04$, $T = 1$, $S(0) = 1, V(0) = 0.04$；样本量 10^6。

Z	m	var		$time$		$var \times time$	
分布1	\	2.60×10^{-8}		54.59		1.42×10^{-6}	
分布2	7	4.05×10^{-8}		28.70		1.16×10^{-6}	
Dist3	1	4.48×10^{-8}		26.90		1.20×10^{-6}	
n	0	1	2	3	4	5	6
β_n	0.0306	6.19×10^{-4}	1.55×10^{-4}	4.07×10^{-5}	1.09×10^{-5}	2.97×10^{-6}	8.23×10^{-7}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.1006	0.0356	0.0129	0.0047	0.0017	0.0006
Dist3	1.000	0.1006	0.0355	0.0126	0.0044	0.0016	0.0006

表6：独立和估计量 $\tilde{Z}$；Heston模型； $r = 0.05$, $\sigma = 0.25$, $k = 1$,$\theta = 0.04$, $S(0) = 1, V(0) = 0.04$；样本量 10^6。

$\tilde{Z}$	m	var		$time$		$var \times time$	
分布1	\	2.99×10^{-8}		67.63		2.02×10^{-6}	
分布2	7	4.47×10^{-8}		31.77		1.42×10^{-6}	
Dist3	1	4.75×10^{-8}		30.30		1.44×10^{-6}	
n	0	1	2	3	4	5	6
β_n	0.0367	3.15×10^{-4}	8.18×10^{-5}	2.20×10^{-5}	6.19×10^{-6}	1.77×10^{-6}	5.31×10^{-7}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.0655	0.0236	0.0087	0.0032	0.0012	0.0005
Dist3	1.000	0.0655	0.0232	0.0082	0.0029	0.0010	0.0004

由 Zheng 和 Pan [17] 提出。令 $X(T) = \ln(S(T))$，则解可写为

$$\begin{aligned} X(T) =& X(0) + \int_0^T r(t)dt + (\frac{\rho k}{\sigma} - \frac{1}{2}) \int_0^T V(t)dt \\ &+ \frac{\rho}{\sigma}(V(T) - V(0) - k\theta T) + \sqrt{1-\rho^2}\sqrt{\int_0^T V(t)dt}N \end{aligned}$$

其中，N 是一个与 V 和 r相互独立的标准正态随机变量。这里，$V(t)$服从与 Heston 模型中相同的非中心卡方分布，而$r(t)$服从正态分布（参见 Glasserman [5]）。接下来，可以方便地近似

$$\int_0^T r_t dt \approx \sum_{i=0}^{T/h-1} r_{ih}h, \int_0^T V_t dt \approx \sum_{i=0}^{T/h-1} V_{ih}h$$

其中，h是时间步长。此外，标准欧式看涨期权的价格可表示为

$$E\left[e^{-\int_0^T r(t)dt} \max(S(T) - 1, 0)\right]$$

其中，积分可用类似的方法进行近似。对于该格式，理论收敛率为 $p = 1$；参见 Zheng 和 Pan [17]。

数值实验的结果见表 7 和表 8。对于这两种$\widetilde{\text{估}}$计量 Z 和 Z，分布 $Dist3$ 最令人满意。此外，我们发现，对于所有 $n = m$、$m + 1$.....，均有 $3.5 < \beta_n/\beta_{n+1} < 4.5$，这再次展现了自适应算法的稳健性。

6 结论

本文研究了 Rhee 和 Glynn [15] 针对随机微分方程模型提出的无偏估计量相关的优化问题。具体而言，这些无偏估计量的计算效率取决于随机变量的累积分布函数的选择，而最优分布应在满足特定约束条件的前提下，使无偏估计量的方差与计算时间的乘积最小化。基于 Rhee 和 Glynn [15] 以及 Cui 等人[3], 的结果，我们证明，在关于 β_n收敛性的一个温和假设下，无限时域下的最优分布可以用一种简单的形式表示。这

表 7：耦合和估计量；Heston-Hull-White； $k = 3, \theta = 0.04, \sigma = 0.25, \alpha =$**1**, $\beta = 0.06$, $\gamma = 0.5, S$**(0)** $= 1, V$**(0)** $= 0.04, r$**(0)** $= 0.05$；样本量 10^6。

Z	m	var		$time$		$var \times time$	
分布1	＼	2.07×10^{-7}		18.44		3.82×10^{-6}	
分布2	7	1.33×10^{-7}		22.42		2.98×10^{-6}	
Dist3	2	1.32×10^{-7}		22.46		2.98×10^{-6}	
n	0	1	2	3	4	5	6
β_n	0.0314	1.37×10^{-2}	4.71×10^{-3}	1.30×10^{-4}	3.44×10^{-5}	8.69×10^{-5}	2.20×10^{-5}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.4671	0.1936	0.0719	0.0262	0.0093	0.0033
Dist3	1.000	0.4671	0.1936	0.0685	0.0242	0.0086	0.0030

表 8：独立和估计量；Heston-Hull-White； $k = 3, \theta = 0.04, \sigma =$**0.25**, $\alpha = 1, \beta = 0.06$, $\gamma = 0.5, S$**(0)** $= 1, V$**(0)** $= 0.04, r$**(0)** $= 0.05$；样本量 10^6。

Z	m	var		$time$		$var \times time$	
分布1	＼	1.56×10^{-7}		21.50		3.35×10^{-6}	
分布2	7	1.17×10^{-7}		26.23		3.07×10^{-6}	
Dist3	2	1.19×10^{-7}		27.12		3.23×10^{-6}	
n	0	1	2	3	4	5	6
β_n	0.0322	1.23×10^{-2}	3.67×10^{-3}	9.85×10^{-4}	2.53×10^{-4}	6.40×10^{-5}	1.61×10^{-5}
分布1	1.000	0.3536	0.1250	0.0442	0.0156	0.0055	0.0020
分布2	1.000	0.4368	0.1690	0.0619	0.0222	0.0079	0.0028
Dist3	1.000	0.4368	0.1690	0.0597	0.0211	0.0075	0.0026

该结果建立了 m 截断最优分布与无限时域最优分布之间的联系，使我们能够针对 m 的自适应取值构造最优分布的自适应算法。与 Rhee 和 Glynn [15] 以及 Cui 等人[3], 提出的 m 截断算法相比，我们的自适应算法有以下优点：首先，它能够处理具有无限时域的最优分布；其次，通常只需较小的 m 值，就能准确估计最优分布，从而大幅缩短 β_n 的先验估计所需的计算时间。通过若干金融领域中经典随机微分方程模型的数值示例，展示了该自适应算法的效率。

致谢

本研究得到国家自然科学基金（项目编号：11801504、11801502）资助。

参考文献

[1] Blanchet J.H 和 Glynn W.P（2015）。通过多层随机化实现优化问题及期望函数的无偏蒙特卡罗方法。《冬季仿真会议论文集》。3656-3667。[2] Black F 和 Scholes M（1973）。期权与公司负债的定价。《政治经济学杂志》。81(3)，637-654。[3] Cui Z、Lee C、Zhu L 和 Zhu Y（2021）。论随机化无偏蒙特卡罗估计量的最优设计。《运筹学快报》。49，477-484。[4]Duffie D 和 Glynn W.P（1995）。证券价格的高效蒙特卡罗模拟。《应用概率年刊》。5(4)，897-905。[5] Glasserman, P（2003）。《金融工程中的蒙特卡罗方法》。Springer Sciences and Business Media，纽约。[6] Glynn W.P 和 Rhee CH（2014）。马尔可夫链平衡期望的精确估计。《应用概率杂志》。51，377-389。[7] Giles, M（2008）。多层蒙特卡罗路径模拟。《运筹学》。56(3)，607-617。

[8] Giles M.B 和 Szpruch L（2014）。无需列维面积模拟的多维随机微分方程反变量多层蒙特卡罗估计。《应用概率年刊》。24(4)，1585-1620。[9] Glynn P.W 和 White W（1992）。模拟估计量的渐近效率。《运筹学》。40(3)，505-520。[10] Grzelak L.A 和 Oosterlee C.W（2011）。随机利率下的 Heston模型。《SIAM金融数学杂志》。2(1)，255-286。[11] Heston, S（1993）。随机波动率期权的闭式解及其在债券和货币期权中的应用。《金融研究评论》6(2)，327-343。[12] Hull J.C 和 White, A（1990）。利率衍生证券定价。《金融研究评论》。3(4)，573-592。[13] Kloeden P 和 Platen E（1999）。《随机微分方程的数值解》，第3版，Springer Verlag，纽约。[14] Mcleish D（2011）。蒙特卡罗估计量去偏的一般方法。《应用蒙特卡洛方法》。17(2011)，301-315。[15] RheeCH 和 Glynn W.P（2015）。随机微分方程模型的无偏估计与平方根收敛。《运筹学》。63(5)，1026-1043。[16] Vihola M（2017）。无偏估计量与多层蒙特卡罗。《运筹学》。66(2)，448-462。[17] Zheng C 和 Pan J（2023）。随机利率下 Heston模型的无偏估计量。arxiv.org/abs/2301.12072。[18] Zheng Z 和 Glynn W.P（2017）。无穷分层估计量的中心极限定理及其在去偏多层蒙特卡罗中的应用。ESAIM Proc.Surv. 59(2017)，104-114。