Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

CONVERGENCE OF THE STOCHASTIC MESH ESTIMATOR FOR PRICING AMERICAN OPTIONS

Athanassios N. Avramidis

Département d' Informatique
et de Recherche Opérationnelle
Université de Montréal, C.P. 6128, Succ. Centre-Ville
Montréal, H3C 3J7, CANADA

Heinrich Matzinger

Faculty of Mathematics
Universtity of Bielefeld
Postfach 100131, 33501 Bielefeld, GERMANY

ABSTRACT

Broadie and Glasserman proposed a simulation-based method they named *stochastic mesh* for pricing high-dimensional American options. Based on simulated states of the assets underlying the option at each exercise opportunity, the method produces an estimator of the option value at each sampled state. Under the mild assumption of the finiteness of certain moments, we derive an asymptotic upper bound on the probability of error of the mesh estimator, where both the error size and the probability bound vanish as the sample size increases. We include the empirical performance for the test problems used by Broadie and Glasserman in a recent unpublished manuscript. We find that the mesh estimator has large bias that decays very slowly with the sample size, suggesting that in applications it will most likely be necessary to employ bias and/or variance reduction techniques.

1 INTRODUCTION

In the financial markets, sophisticated, complex products are continuously offered and traded. There are many financial products whose values depend on more than one underlying asset. Examples include basket options (options on the average of several underlying assets), out-performance options (options on the maximum of several assets), spread options (options on the difference between two assets), and quantos (options whose payoff is adjusted by some stochastic variable, typically an exchange rate). Even when there is a single underlying asset, there is trend towards models with multiple stochastic factors (sources of uncertainty), e.g., single-asset model with stochastic volatility. In addition, multi-factor models are gaining more acceptance and use for modeling interest rates, where models with two to four factors are common and models with up to ten factors are being tested (Broadie and Glasserman 1997a). As computing power is steadily increasing, multi-factor option-pricing models are likely to become more prevalent.

Pricing and hedging options (European or American) using multi-factor models is a difficult task. Especially for American options, which allow early exercise, analytical formulas for pricing are rarely available. Various deterministic numerical techniques are used, for example the numerical solution of an appropriate partial differential equation. However, such methods require work that grows exponentially in the number of state variables. This work requirement renders these methods ineffective when the state space dimension is higher than three or four.

Monte Carlo simulation techniques are conceptually simple, yet powerful in addressing option pricing problems of great complexity, whether the complexity arises from the stochastic process driving the assets, the structure of the payoff (path-dependent), or the early exercise features (American). Until recently, the prevailing opinion was that American options could not be handled using Monte Carlo simulation. Recent developments, however, have started to pave the way for estimating American option prices via Monte Carlo methods.

Barraquand and Martineau (1995) proposed an algorithm that only approximately solves the American option pricing problem. They partition the state space of stochastic factors into a tractable number of cells and compute an approximately optimal exercise policy that is constant over each cell. Although this method is fast, it yields an estimate that does not necessarily converge to the true price as work increases. Broadie and Glasserman (1997b) were the first to develop a simulation procedure that yields provably convergent estimates for American option prices. Their method is based on a simulated tree of the state variables. The main drawback of their method is that the work is exponential in the number of exercise opportunities. For a comprehensive review of the literature in Monte Carlo methods for Pricing American Options, see Broadie and Glasserman (1997a).

An important method developed recently for valuing American options via simulation is the stochastic mesh method (Broadie and Glasserman 1997c). The stochastic mesh method begins by generating a number b of randomly sampled states of the stochastic factors underlying the option at each exercise opportunity. Based on this sample, the *mesh estimator* of the option value at each sampled state is computed (a full description is deferred until Section 2.2). The authors also propose a *path estimator*, obtained by simulating paths of the stochastic factors underlying the options and estimating an approximate exercise policy based on the mesh values; see Broadie and Glasserman (1997c) for more details. It is shown that the mesh and path estimators are biased high and low, respectively. In addition, under certain technical assumptions, it is shown that both estimators converge (in norm) to the true option value as the sample size (the number of sampled states per stage) b goes to infinity.

In this paper we derive an asymptotic upper bound on the probability of error of the mesh estimator with respect to b. Both the error size and the upper bound on the probability of error are functions of b that vanish as $b \to \infty$. Our assumptions are mild–namely the finiteness of certain moments. We also present empirical results on the estimator's behavior on the test problems in Broadie and Glasserman (1997c).

This paper is organized as follows. Section 2 contains brief background on the problem of pricing American options and a description of the stochastic mesh method. Section 3 contains our main theoretical result, namely an asymptotic bound on the probability of error of the mesh estimator with respect to the number b of states sampled at each stage. In Section 4 we present computational results on the test problems in Broadie and Glasserman (1997c) and in Section 5 we offer conclusions.

2 BACKGROUND

2.1 American Option Pricing

Let S_t denote the vector of stochastic factors underlying the option, modeled as a Markov process on $\mathbf{R}^d$ with discrete-time parameter $t = 0, 1, 2, ..., T$. The argument t indexes the set of times when the option is exerciseable, also called *exercise opportunities* or simply *stages*. Let $h(t, x)$ denote the payoff to the option holder from exercise at time t in state x, discounted to time 0 with the possibly stochastic discount factor recorded in x. This view of $h(t, x)$ as the discounted-to-time-0 payoff is adopted to simplify the notation and does not reduce the generality of the method.

By the dynamic programming principle, the option value can be written as follows:

$$q(t, x) = \begin{cases} h(t, x) & t = T, \text{ all } x \\ \max\{h(t, x), c(t, x)\} & 0 \le t \le T - 1, \text{ all } x \end{cases}$$

where

$$c(t, x) = E[q(t + 1, S_{t+1})|S_t = x] \tag{1}$$

is called the *continuation value* at (t, x), equal to the value of the option (discounted to time 0) when it is not exercised at (time, state) pair (t, x). It is well-known from arbitrage pricing theory that the arbitrage-free price of the option is obtained when the conditional expectation in (1) is taken with respect to the risk-neutral measure, defined as the measure that makes the value of any tradeable security, discounted to time 0, a martingale. For a rigorous treatment of arbitrage pricing theory, see Duffie (1996) and Harrison and Pliska (1981); for an excellent and mathematically lighter treatment, see Baxter and Rennie (1996). Given the known state of S_0 at time 0, say x_0, the option-pricing problem is to compute $q(0, x_0)$.

2.2 The Stochastic Mesh Method

In reviewing the method, we follow Broadie and Glasserman (1997c). The mesh method generates a stochastic *mesh* of sample states $\{S_t^j\}$, $j = 1, 2, ..., b$ for each $t = 1, ..., T$. For notational convenience, we define b nonrandom mesh points at stage 0, $S_0^j = x_0$, $j = 1, 2, ..., b$. For $t = 1, 2, ..., T$, let $g_t(\cdot)$ denote the probability density from which the points $\{S_t^j\}_{j=1}^b$ are sampled (to be specified later), and let $f_t(x, \cdot)$ denote the conditional risk-neutral density of S_{t+1} given $S_t = x$. (We assume throughout the paper the existence of such densities.) Let $\mathcal{E} = \{0, 1, \ldots, T - 1\}$ denote the index set of *early-exercise opportunities* and let $\mathcal{I} = \{1, 2, \ldots, b\}$ denote the index set of sampled points per stage. The Broadie-Glasserman mesh estimator is calculated as a backward recursion for $t = T, T - 1, ..., 0$:

$$\widehat{q}_H(T, S_T^j) = \begin{cases} h(T, S_T^j) & j \in \mathcal{I} \\ \max\{h(t, S_t^j), \widehat{c}(t, S_t^j)\} & j \in \mathcal{I}, t \in \mathcal{E} \end{cases}$$

where the estimate of the continuation value function $\widehat{c}(t, x)$ is

$$\widehat{c}(t, x) := \sum_{j=1}^{b} \frac{\widehat{q}_H(t + 1, S_{t+1}^j) f_t(x, S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)}. \tag{2}$$

Note that the point S_{t+1}^j is weighed by the likelihood ratio $f_t(x, S_{t+1}^j)/g_{t+1}(S_{t+1}^j)$.

In Broadie and Glasserman (1997c), it is argued that the choice of sampling densities $g_{t+1}(\cdot)$ is crucial to the success of the method; and the choice recommended by the authors is as follows. We simulate independently b paths of S_t starting from x_0 at time 0 and let S_t^j denote the state of the j-th path at time t; and then we "forget" the path to which a point belongs. This is called by the authors the *stratified implementation*. For any t, j, we call the ordered pair (S_t^j, S_{t+1}^j) a *parent* and *child*, respectively.

We clarify some properties of the stratified implementation. Let π be a random permutation of the integers in $\{1, 2, ..., b\}$ chosen with equal probability from all possible such permutations, and let $\mathcal{F}_t$ be the σ-field $\mathcal{F}_t = \sigma(S_t^1, S_t^2, ..., S_t^b)$. Then

$$\begin{gathered}\text{Conditional on } \mathcal{F}_t,\\ \{S_{t+1}^{\pi(1)}, S_{t+1}^{\pi(2)}, ..., S_{t+1}^{\pi(b)}\} \overset{\text{i.d.}}{\sim} g_{t+1}(\cdot) := \tfrac{1}{b}\textstyle\sum_{i=1}^{b} f_t(S_t^i, \cdot)\end{gathered} \tag{3}$$

where $\overset{\text{i.d.}}{\sim}$ means "are identically distributed with density ...". Note that the density $g_{t+1}(\cdot)$ is defined conditionally on $\mathcal{F}_t$. Also note that $\{S_{t+1}^{\pi(1)},S_{t+1}^{\pi(2)},...,S_{t+1}^{\pi(b)}\}$ are conditionally dependent random vectors. On the other hand,

$$\begin{gathered}\text{Conditional on } \mathcal{F}_t,\\ \{S_{t+1}^1, S_{t+1}^2, ..., S_{t+1}^b\} \text{ are independent.}\end{gathered} \tag{4}$$

Also note that $\{S_{t+1}^1, S_{t+1}^2, ..., S_{t+1}^b\}$ are conditionally *not* identically distributed; they are unconditionally independent and identically distributed.

3 CONVERGENCE IN PROBABILITY

Under an assumption on the finiteness of certain moments, we will show that the estimator $\widehat{q}_H(0, x_0)$ with the stratified implementation converges in probability to $q(0, x_0)$ as $b \to \infty$; in fact, we derive an asymptotic upper bound on the probabilty of error, where both the error size and the probability upper bound vanish as $b \to \infty$.

We require the following moment assumptions, where S_t^1,S_t^2,S_t^3 denote paths which are independent of each other and have the distribution of S_t conditioned under $S_0 = x_0$, and where C is a constant that will appear on the probability bound.

$$\max_{t\in\mathcal{E}} \mathrm{E}\left[\max_{t+1\le r\le T}\{h^4(r, S_r^1)\}\right] \le C/8 \tag{5}$$

$$\max_{t\in\mathcal{E}} \mathrm{E}\left[\max_{t+1\le r\le T}\{h^4(r, S_r^2)\}\frac{f_t^4(S_t^1, S_{t+1}^2)}{f_t^4(S_t^3, S_{t+1}^2)}\right] \le C/8 \tag{6}$$

$$\max_{t\in\mathcal{E}} \mathrm{E}\left[\max_{t+1\le r\le T}\{h^4(r, S_r^1)\}\frac{f_t^4(S_t^1, S_{t+1}^1)}{f_t^4(S_t^3, S_{t+1}^1)}\right] < \infty \tag{7}$$

$$\max_{t\in\mathcal{E}} \mathrm{E}\left[\frac{f_t^4(S_t^1, S_{t+1}^2)}{f_t^4(S_t^3, S_{t+1}^2)}\right] \le C/8 \tag{8}$$

$$\max_{t\in\mathcal{E}} \mathrm{E}\left[\frac{f_t^4(S_t^1, S_{t+1}^1)}{f_t^4(S_t^3, S_{t+1}^1)}\right] < \infty \tag{9}$$

Theorem 1. *Suppose b mesh paths $\{(S_t^j : t = 0, 1, \dots, T)\}_{j=1}^b$ are generated independently with $S_0^j = x_0$ for all $j \in \{1, 2, ..., b\}$, where $x_0 \in \mathbf{R}^d$ is known at time 0. Under assumptions (5)-(9),*

$$P\left\{|\widehat{q}_H(0, x_0) - q(0, x_0)| > (1 + \frac{\delta}{b^\gamma})^T - 1\right\}$$
$$\le \tfrac{6CT}{\delta^4 b^{1-4\gamma}} + O(b^{-2+4\gamma}) \quad \textit{for any } \delta > 0,\ 0 < \gamma < 1/4.$$

Proof. We start with a few definitions. Unless explicitly stated, the time index $t \in \mathcal{E}$. Let

$$\bar{c}(t, x) := \frac{1}{b}\sum_{j=1}^{b}\frac{q(t+1, S_{t+1}^j) f(x, S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)}.$$

In other words $\bar{c}(t, x)$ is the natural estimate we would make of $c(t, x)$ if $q(t+1, \cdot)$ were known (which of course is not the case). Fix $\delta > 0$ and $0 < \gamma < 1/4$, and define the events

$$A_1(t) = \left\{\omega : |\bar{c}(t, S_t^j) - c(t, S_t^j)| \le \frac{\delta}{b^\gamma},\ \forall j \in \mathcal{I}\right\}$$

and

$$A_2(t) = \left\{\omega : \left|\frac{1}{b}\sum_{j=1}^{b}\frac{f_t(S_t^i, S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)} - 1\right| \le \frac{\delta}{b^\gamma},\ \forall j \in \mathcal{I}\right\}$$

where ω denotes a generic point in the sample space, and where for notational simplicity we suppress the dependence of all random variables on ω. Let A_1 be the event that $A_1(t)$ holds for each $t \in \mathcal{E}$, i.e.,

$$A_1 := \cap_{t\in\mathcal{E}} A_1(t).$$

Similarly, define $A_2 = \cap_{t\in\mathcal{E}} A_2(t)$. Finally, define the event of direct interest

$$A = \left\{\omega : |\widehat{q}_H(0, x_0) - q(0, x_0)| \le (1 + \frac{\delta}{b^\gamma})^T - 1\right\}.$$

Claim 1. $A \supset A_1 \cap A_2$.

Proof. We assume that events A_1 and A_2 hold and show by a recursive argument going backwards in time that event A must hold. We start by showing that an error bound that holds uniformly over all estimates at time $t+1$ can be iterated backwards in time. Fix $\varepsilon > 0$ and suppose that for some t $(0 < t \le T-1)$ the error of the estimates at the forward points satisfies

$$|\widehat{q}_H(t+1, S^j_{t+1}) - q(t+1, S^j_{t+1})| \le \varepsilon \quad \text{for all } j \in \mathcal{I}. \quad (10)$$

Then

$$\begin{aligned}
&|\widehat{c}(t,x) - \bar{c}(t,x)| \\
&= \frac{1}{b}\left|\sum_{j=1}^{b} \frac{\widehat{q}_H(t+1, S^j_{t+1}) f_t(x, S^j_{t+1})}{g_{t+1}(S^j_{t+1})} - \sum_{j=1}^{b} \frac{q(t+1, S^j_{t+1}) f_t(x, S^j_{t+1})}{g_{t+1}(S^j_{t+1})}\right| \\
&= \frac{1}{b}\left|\sum_{j=1}^{b} \left(\widehat{q}_H(t+1, S^j_{t+1}) - q(t+1, S^j_{t+1})\right) \times \frac{f_t(x, S^j_{t+1})}{g_{t+1}(S^j_{t+1})}\right| \\
&\le \frac{\varepsilon}{b} \sum_{j=1}^{b} \frac{f_t(x, S^j_{t+1})}{g_{t+1}(S^j_{t+1})} \\
&\le \varepsilon\left(1 + \frac{\delta}{b^\gamma}\right) \quad \text{for all } x \in \{S^1_t, S^2_t, .., S^b_t\}, \qquad (11)
\end{aligned}$$

where the last inequality follows since A_2 holds. So if (10) holds, then the error of $\widehat{q}_H$ at stage t $(0 \le t \le T-1)$ can be bound uniformly on j as follows:

$$\begin{aligned}
&\left|\widehat{q}_H(t, S^j_t) - q(t, S^j_t)\right| \\
&= \left|\max\{h(t, S^j_t), \widehat{c}(t, S^j_t)\} - \max\{h(t, S^j_t), c(t, S^j_t)\}\right| \\
&\le \left|\widehat{c}(t, S^j_t) - c(t, S^j_t)\right| \\
&\le \left|\widehat{c}(t, S^j_t) - \bar{c}(t, S^j_t)\right| + \left|\bar{c}(t, S^j_t) - c(t, S^j_t)\right| \\
&\le \varepsilon\left(1 + \frac{\delta}{b^\gamma}\right) + \frac{\delta}{b^\gamma} \quad \text{for all } j \in \mathcal{I} \qquad (12)
\end{aligned}$$

where in the last inequality we used (11) and that event A_1 holds.

Now the recursive bounding is as follows. We start the error bounding with the special case $t = T-1$, where we observe that $\widehat{c}(T-1, S^j_{T-1}) - \bar{c}(T-1, S^j_{T-1}) = 0$ for all j, and so the definition of the event $A_1(T-1)$ implies that (12) holds for $t = T-1$ with $\varepsilon = 0$. Iterating the bounding argument in (12) with $t = T-2, T-3, ..., 0$, we get

$$\begin{aligned}
|\widehat{q}_H(0, x_0) - q(0, x_0)| &\le \frac{\delta}{b^\gamma} \sum_{j=0}^{T-1} \left(1 + \frac{\delta}{b^\gamma}\right)^j \\
&= \frac{\delta}{b^\gamma} \frac{(1 + \frac{\delta}{b^\gamma})^T - 1}{1 + \frac{\delta}{b^\gamma} - 1} \\
&= \left(1 + \frac{\delta}{b^\gamma}\right)^T - 1
\end{aligned}$$

which completes the proof of Claim 1. □

Letting A^c denote the complement of the event A, we have $P(A^c) \le P(A^c_1) + P(A^c_2)$. To complete the proof, we will show that $P(A^c_1) \le \frac{3CT}{\delta^4 b^{1-4\gamma}} + O(b^{-2+4\gamma})$ and $P(A^c_2) \le \frac{3CT}{\delta^4 b^{1-4\gamma}} + O(b^{-2+4\gamma})$.

We first obtain the upper bound for $P(A^c_1)$. Define the event

$$A_1(t,i) = \left\{\omega : |\bar{c}(t, S^i_t)(\omega) - c(t, S^i_t)(\omega)| \le \frac{\delta}{b^\gamma}\right\}.$$

Recall that $A_1 = \cap_{t=0}^{T-1} A_1(t) = \cap_{t=0}^{T-1} \cap_{i=1}^{b} A_1(t,i)$, so

$$P(A^c_1) \le \Sigma_{t=0}^{T-1} \Sigma_{i=1}^{b} P(A^c_1(t,i)) = b\Sigma_{t=0}^{T-1} P(A^c_1(t,1)), \quad (13)$$

since $\{\{S^i_t, \{S^j_t\}^b_{j=1}\}^b_{i=1}$ are identically distributed. We will show that

$$P(A^c_1(t,1)) \le \frac{3C}{\delta^4 b^{2-4\gamma}} + O(b^{-3+4\gamma}) \text{ for all } t \in \mathcal{E}, \quad (14)$$

which, in view of (13), proves that $P(A^c_1) \le \frac{3CT}{\delta^4 b^{1-4\gamma}} + O(b^{-2+4\gamma})$.

The key for proving that $\bar{c}(t, S^1_t) - c(t, S^1_t)$ is small with high probability as $b \to \infty$ is that it can be written as the sum of b random variables which conditionally have mean 0 and are independent.

Claim 2. *$\bar{c}(t, S^1_t) - c(t, S^1_t) = \frac{1}{b}\sum_{j=1}^{b} Z^j(t)$, where*

$$\begin{aligned}
Z^j(t) \;:=\; &\frac{q(t+1, S^j_{t+1}) f(S^1_t, S^j_{t+1})}{g_{t+1}(S^j_{t+1})} \\
&- \mathrm{E}\left[\frac{q(t+1, S^j_{t+1}) f(S^1_t, S^j_{t+1})}{g_{t+1}(S^j_{t+1})}\middle|\mathcal{F}_t\right], \quad j \in \mathcal{I},
\end{aligned}$$

where we recall that $\mathcal{F}_t$ is the σ-field $\mathcal{F}_t = \sigma(S^1_t, S^2_t, ..., S^b_t)$.

Proof.

$$\frac{1}{b}\sum_{j=1}^{b} Z^j(t)$$
$$= \frac{1}{b}\sum_{j=1}^{b}\left(\frac{q(t+1,S_{t+1}^j)f(S_t^1,S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)} - \mathrm{E}\left[\frac{q(t+1,S_{t+1}^j)f(S_t^1,S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)}\middle|\mathcal{F}_t\right]\right)$$
$$= \bar{c}(t,S_t^1) - \mathrm{E}\left[\frac{1}{b}\sum_{j=1}^{b}\frac{q(t+1,S_{t+1}^j)f(S_t^1,S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)}\middle|\mathcal{F}_t\right]$$

$$= \bar{c}(t,S_t^1) - \mathrm{E}\left[\frac{1}{b}\sum_{j=1}^{b}\frac{q(t+1,S_{t+1}^{\pi(j)})f(S_t^1,S_{t+1}^{\pi(j)})}{g_{t+1}(S_{t+1}^{\pi(j)})}\middle|\mathcal{F}_t\right]$$
$$= \bar{c}(t,S_t^1) - \mathrm{E}\left[\frac{q(t+1,X)f(S_t^1,X)}{g_{t+1}(X)}\middle|\mathcal{F}_t\right]$$

where X represents a random variable which is obtained by choosing one of the points $S_{t+1}^1, S_{t+1}^2, .., S_{t+1}^b$ at random with equal probability. The key behind the third step is the invariance of the sum inside the expectation with respect to permutations of the $\{S_{t+1}^j\}_{j=1}^b$. The conditional distribution of X when conditioned under $\mathcal{F}_t$ has the density $g_{t+1}(\cdot)$ in (3), so

$$\mathrm{E}\left[\frac{q(t+1,X)f(S_t^1,X)}{g_{t+1}(X)}\middle|\mathcal{F}_t\right] = \mathrm{E}[q(t+1,S_{t+1}^1))|\mathcal{F}_t] = c(t,S_t^1)$$

which completes the proof of Claim 2. □

Conditional on $\mathcal{F}_t$, each of the variables $\{Z_t^1, Z_t^2, ..., Z_t^b\}$ is a function of the single random variable $\{S_{t+1}^1, S_{t+1}^2, ..., S_{t+1}^b\}$, respectively. As such, the $\{Z^j(t)\}_{j=1}^b$ have two key properties: (a) they have conditional mean 0; and (b) they are conditionally independent, in view of (4). Our upper bound for the probability of $P(A_1^c(t,1))$ will use Markov's inequality with the 4th moment of the deviation $\bar{c}(t,S_t^1) - c(t,S_t^1)$. We will show that this 4th moment goes to zero sufficiently fast with b via the following two lemmas.

Lemma 1. *Suppose Y is a nonnegative random variable with* $\mathrm{E}[Y^4] < \infty$. *Then* $\mathrm{E}[(Y - \mathrm{E}[Y|\mathcal{F}])^4] \le 8\mathrm{E}[Y^4]$, *where* $\mathcal{F}$ *is an arbitrary* σ*-field.*

Proof.

$$\begin{aligned}
&\mathrm{E}[(Y-\mathrm{E}[Y|\mathcal{F}])^4]\\
&= \mathrm{E}\Big(Y^4 - 4Y^3\mathrm{E}[Y|\mathcal{F}] + 6Y^2\mathrm{E}^2[Y|\mathcal{F}]\\
&\quad -4Y\mathrm{E}^3[Y|\mathcal{F}] + \mathrm{E}^4[Y|\mathcal{F}]\Big)\\
&\le \mathrm{E}[Y^4] + 6\mathrm{E}(Y^2\mathrm{E}^2[Y|\mathcal{F}]) + \mathrm{E}(\mathrm{E}^4[Y|\mathcal{F}])\\
&\le \mathrm{E}[Y^4] + 6\sqrt{\mathrm{E}[Y^4]}\sqrt{\mathrm{E}(\mathrm{E}^4[Y|\mathcal{F}])} + \mathrm{E}(\mathrm{E}[Y^4|\mathcal{F}])\\
&\le 2\mathrm{E}[Y^4] + 6\sqrt{\mathrm{E}[Y^4]}\sqrt{\mathrm{E}(Y^4)}\\
&= 8\mathrm{E}[Y^4].
\end{aligned}$$

In the second step, we dropped nonpositive random variables from the expectation. In the third step, we used the Cauchy-Schwartz inequality for the secod term and Jensen's inequality for the third term, and in the fourth step we used again Jensen's inequality inside the second square root. □

Lemma 2. *Let* $\mathcal{F}$ *denote an arbitrary* σ*-field, and let* $Z_1, Z_2, ..., Z_b$ *be random variables which, conditional on* $\mathcal{F}$ *have mean 0, are conditionally independent of each other, and such that* $\mathrm{E}[Z_1^4] < \infty$ *and* $\mathrm{E}[Z_j^4] \le C$ *for each* $j \ne 1$, *where the expectations are unconditional, and Cis a constant. Then*

$$\mathrm{E}\left[\left(\tfrac{1}{b}\textstyle\sum_{j=1}^b Z_j\right)^4\right] \le \frac{3C}{b^2} + O(b^{-3}).$$

Proof. $\mathrm{E}[(\sum_{j=1}^b Z_j)^4] = \Sigma\mathrm{E}[\mathrm{E}[Z_{j_1}Z_{j_2}Z_{j_3}Z_{j_4}|\mathcal{F}]]$, where the four indices are ranging independently from 1 to b. Since $\mathrm{E}[Z_{j_1}|\mathcal{F}] = 0$, the conditional independence of the Z's implies that the summand vanishes if there is one index different from the three others. This leaves terms of the form $\mathrm{E}[\mathrm{E}[Z_{j_1}^4|\mathcal{F}]]$, of which there are b, and terms of the form $\mathrm{E}[\mathrm{E}[Z_{j_1}^2 Z_{j_2}^2|\mathcal{F}]]$ for $j_1 \ne j_2$, of which there are $3b(b-1)$. For each of the two different forms, the number of terms with any index equal to 1 is $O(b^{-1})$ of the total number of such terms, and so the finiteness of $\mathrm{E}[Z_1^4]$ implies that the relative contribution of these terms to the total is $O(b^{-1})$. Now focusing on terms where all indices are different than 1, we have $\mathrm{E}[\mathrm{E}[Z_{j_1}^4|\mathcal{F}]] = \mathrm{E}[Z_{j_1}^4] \le C$, and $\mathrm{E}[\mathrm{E}[Z_{j_1}^2 Z_{j_2}^2|\mathcal{F}]] = \mathrm{E}[Z_{j_1}^2 Z_{j_2}^2] \le \sqrt{\mathrm{E}[Z_{j_1}^4]}\sqrt{\mathrm{E}[Z_{j_2}^4]} \le C$. Hence

$$\mathrm{E}\left[\left(\textstyle\sum_{j=1}^b Z_j\right)^4\right] \le bC\left(1+O(b^{-1})\right) + 3b(b-1)C\left(1+O(b^{-1})\right)$$

which completes the proof of Lemma 2. □

Claim 3. *The* $Z^j(t)$ *satisfy the conditions of Lemma 2 for the* σ*-field* $\mathcal{F} = \mathcal{F}_t$.

Proof. Applying Lemma 1 with $Y = \frac{q(t+1,S_{t+1}^j)f(S_t^1,S_{t+1}^j)}{g_{t+1}(S_{t+1}^j)}$ and $\mathcal{F} = \mathcal{F}_t$, we get

$$\begin{aligned}
&\mathrm{E}[(Z^j(t))^4] \\
&\quad\leq\quad 8\mathrm{E}\left[\frac{q^4(t+1,S_{t+1}^j)f^4(S_t^1,S_{t+1}^j)}{g_{t+1}^4(S_{t+1}^j)}\right] \\
&\quad\leq\quad 8\mathrm{E}\left[\frac{\max_{t+1\leq r\leq T}\{h^4(r,S_r^j)\}f^4(S_t^1,S_{t+1}^j)}{g_{t+1}^4(S_{t+1}^j)}\right]\ \forall j\in\mathcal{I}.
\end{aligned}$$

The $\{Z^j(t)\}_{j=2}^b$ are unconditionally identically distributed, and we have

$$\begin{aligned}
&\mathrm{E}[(Z^2(t))^4] \\
&\quad\leq\quad 8\mathrm{E}\left[\max_{t+1\leq r\leq T}\{h^4(r,S_r^2)\}\right. \\
&\qquad\qquad \left.\times\frac{1}{b}\left(\frac{f^4(S_t^1,S_{t+1}^2)}{f^4(S_t^1,S_{t+1}^2)}+\sum_{s\neq 1}\frac{f^4(S_t^1,S_{t+1}^2)}{f^4(S_t^s,S_{t+1}^2)}\right)\right] \\
&\quad=\quad 8\left\{\frac{1}{b}\mathrm{E}\left[\max_{t+1\leq r\leq T}\{h^4(r,S_r^2)\}\right]\right. \\
&\qquad\qquad \left.+\frac{b-1}{b}\mathrm{E}\left[\max_{t+1\leq r\leq T}\{h^4(r,S_r^j)\}\left(\frac{f^4(S_t^1,S_{t+1}^2)}{f^4(S_t^3,S_{t+1}^2)}\right)\right]\right\} \\
&\quad\leq\quad 8\left\{\frac{1}{b}\frac{C}{8}+\frac{b-1}{b}\frac{C}{8}\right\} \\
&\quad=\quad C\quad\text{for all } t\in\mathcal{E},
\end{aligned}$$

where for the first step we recall the definition of g_{t+1} in (3) and we use the fact (Jensen's inequality) that for any $x_1, x_2, ..., x_b > 0$,

$$\frac{1}{\left(\frac{x_1+...+x_b}{b}\right)^4}\leq\frac{1}{b}\left(\frac{1}{x_1^4}+...+\frac{1}{x_b^4}\right).$$

An analogous argument combined with assumption (7) shows that $\mathrm{E}[(Z^1(t))^4] < \infty$ for all t. □

Now we have

$$\begin{aligned}
P(A_1^c(t,1)) &= P\left(|\bar{c}(t,S_t^1)-c(t,S_t^1)|\geq\frac{\delta}{b^\gamma}\right) \\
&= P\left(\frac{1}{b}\left|\sum_{j=1}^b Z^j(t)\right|\geq\frac{\delta}{b^\gamma}\right) \\
&\leq \frac{\mathrm{E}\left[\left(\frac{1}{b}\sum_{j=1}^b Z^j(t)\right)^4\right]b^{4\gamma}}{\delta^4} \qquad (15) \\
&\leq \frac{3C}{\delta^4 b^{2-4\gamma}}+O(b^{-3+4\gamma}) \qquad (16)
\end{aligned}$$

for each $t\in\mathcal{E}$. In step three, we used Markov's inequality with power 4, and in step four we used Lemma 2 with $Z_j = Z^j(t)$ and $\mathcal{F} = \mathcal{F}_t$. This is precisely what was required in (14), and completes the proof that $P(A_1^c)\leq\frac{3CT}{\delta^4 b^{1-4\gamma}}+O(b^{-2+4\gamma})$.

The probability bound $P(A_2^c)\leq\frac{3CT}{\delta^4 b^{1-4\gamma}}+O(b^{-2+4\gamma})$ is proved by noting that A_2^c can be written as an event of the form A_1^c for the function $q(\cdot,\cdot)=1$, and then assumptions (8) and (9) will serve in place of (6) and (7), respectively. This completes the proof of Theorem 1. □

The following result shows that the rate of convergence may be sharpened using moments of order higher than 4 as we did in assumptions (5)-(9).

Theorem 2. *Suppose the mesh paths $\{S_t^j\}_{j=1}^b$ are generated independently with $S_0^j = x_0$ for all $j\in\{1,2,...,b\}$, where $x_0\in\mathbf{R}^d$ is the known state at time 0. Under assumptions (5)-(9) where we replace the power 4 by the power 8 and let C_1 be the corresponding constant,*

$$\begin{aligned}
&P\left\{|\widehat{q}_H(0,x_0)-q(0,x_0)|>(1+\frac{\delta}{b^\gamma})^T-1\right\} \\
&\quad\leq\frac{2520C_1T}{\delta^8 b^{5-8\gamma}}+O(b^{-6+8\gamma})\ \textit{for any}\ \delta>0,\ 0<\gamma<5/8.
\end{aligned}$$

Sketch of Proof. One can show that $P(A_1^c(t,1))\leq\frac{1260C_1}{\delta^8 b^{5-8\gamma}}+O(b^{-6+8\gamma})$ by arguing analogously to (15)-(16), using Markov's inequality with power 8 instead of power 4 and a result analogous to Lemma 2 for the 8th moment. The other steps in the proof are as in Theorem 1. □

4 EMPIRICAL PERFORMANCE

We report empirical results on the performance of the mesh estimator on the test problems in Broadie and Glasserman (1997c) . Under the risk-neutral measure, the d assets are independent, and each follows a geometric Brownian motion process:

$$dS_\tau(k) = S_\tau(k)[(r-\delta)d\tau+\sigma dW_\tau(k)],\quad k=1,\ldots,d,$$

where $W_\tau(k), k=1,\ldots,d$ are independent Brownian motions, r is the riskless interest rate, δ is the divident rate, and σ is a volatility parameter. Exercise opportunities occur at the set of calendar times $\tau_t = t\mathcal{T}/T, t=0,1,\ldots,T$, where $\mathcal{T}$ is the calendar option expiration time. Under the risk-neutral measure, the random variables $\log(S_{\tau_t}(k)/S_{\tau_{t-1}}(k))$ for $k=1,\ldots,d$ are independent and normally distributed with mean $(r-\delta-\sigma^2/2)(\tau_t-\tau_{t-1})$ and variance $\sigma^2(\tau_t-\tau_{t-1})$.

Tables 1-3 contain results for a *maximum option,* which is a call option on the maximum of the assets with payoff equal to

$$h(t,(S(k),k=1,\ldots,d)) = e^{-r\tau_t}\left(\max_{1\leq k\leq d}S(k)-K\right)^+$$

where $(x)^+ := max(x, 0)$. The parameters are $d = 5$, $r = 0.05$, $\delta = 0.1$, $\sigma = 0.2$, $K = 100$, $\mathcal{T} = 3$, and $T = 3$, 6, and 9, respectively. Tables 4-5 contain results for a *goemetric average option*, which is a call option on the geometric average of the assets with payoff equal to

$$h\left(t, (S(k), k = 1, \ldots, d)\right) = e^{-r\tau_t} \left(\left(\prod_{k=1}^{d} S(k)\right)^{\frac{1}{d}} - K \right)^+$$

and parameters $d = 5$ and 7 assets respectively, $r = 0.03$, $\delta = 0.05$, $\sigma = 0.4$, $K = 100$, $\mathcal{T} = 1$, and $T = 10$. Within each table, the two panels contain results for out-of-the-money and in-the money cases, specifically with $S_0(k) = x_0, k = 1, \ldots, d$, where $x_0 = 90$ and 110, respectively. Within each panel, we set the mesh size b to the values 200, 400, 800, and 1600. The column labeled "CPU" measures CPU time in seconds per replication of $\widehat{q}_H$ on a SUN Ultra 5 workstation. Our performance measures are the relative bias (RB), relative standard error (RSE), and relative root mean square error (RRMSE) of $\widehat{q}_H$, defined as the bias, standard error, and root mean square error (RMSE) divided by the true option value, respectively. We approximated the true option values using the results in Broadie and Glasserman (1997c) as follows. For the max option, we used the most accurate estimates in that paper, which have a relative error less than 0.35% with 99% confidence. For the geometric average option, the values are calculated from a single-asset binomial tree, presumably with negligible error. For completeness, these approximated "true" option values are listed here in the order in which they appear in the tables, i.e., Table 1, panel 1; Table 1, panel 2; Table 2, panel, 1; etc. The values are: 16.006, 35.695, 16.474, 36.497, 16.659, 36.782, 1.362, 10.211, 0.761, and 10. The estimates $\widehat{\text{RB}}$, $\widehat{\text{RSE}}$, and $\widehat{\text{RRMSE}}$ in these tables are based on 64 independent replications of $\widehat{q}_H$.

Table 1: Maximum Option on Five Assets, $T = 3$.

x_0	b	CPU	$\widehat{\text{RB}}$	$\widehat{\text{RSE}}$	$\widehat{\text{RRMSE}}$
90	200	3.3	0.175	0.093	0.198
	400	8.4	0.127	0.052	0.137
	800	24.1	0.089	0.038	0.097
	1600	78.1	0.064	0.023	0.068
110	200	3.3	0.149	0.044	0.155
	400	8.4	0.115	0.036	0.121
	800	24.3	0.074	0.021	0.077
	1600	78.0	0.054	0.015	0.056

It is obvious that the mesh estimator is highly positively biased, with (relative) bias being the dominant factor in the estimator's overall error, as measured by RRMSE. The number of exercise opportunities T is an important factor, with relative bias and overall error increasing fast with T. This is expected in view of Theorem 1, which shows a geometric growth of the estimator's error bound with the number of exercise opportunities. In all cases, the bias decays slowly with b, and this appears to be the general pattern over further experiments not reported here. In view of the quadratic growth of work with b, the obvious extrapolation from these tables suggests that the large bias will persist for most feasible sample sizes.

Table 2: Maximum Option on Five Assets, $T = 6$.

x_0	b	CPU	$\widehat{\text{RB}}$	$\widehat{\text{RSE}}$	$\widehat{\text{RRMSE}}$
90	200	6.6	0.402	0.098	0.414
	400	17.0	0.337	0.066	0.343
	800	49.0	0.288	0.043	0.291
	1600	158.5	0.231	0.029	0.233
110	200	6.6	0.370	0.066	0.376
	400	16.9	0.331	0.038	0.333
	800	48.7	0.256	0.023	0.257
	1600	158.5	0.203	0.018	0.204

Table 3: Maximum Option on Five Assets, $T = 9$.

x_0	b	CPU	$\widehat{\text{RB}}$	$\widehat{\text{RSE}}$	$\widehat{\text{RRMSE}}$
90	200	9.9	0.557	0.096	0.566
	400	25.6	0.521	0.064	0.525
	800	73.2	0.466	0.042	0.468
	1600	238.4	0.402	0.032	0.403
110	200	9.8	0.556	0.061	0.559
	400	25.5	0.503	0.040	0.505
	800	73.2	0.445	0.026	0.446
	1600	239.4	0.368	0.021	0.368

Table 4: Geometric Average Option on Five Assets, $T = 10$.

x_0	b	CPU	$\widehat{\text{RB}}$	$\widehat{\text{RSE}}$	$\widehat{\text{RRMSE}}$
90	200	10.9	0.621	0.320	0.699
	400	28.4	0.610	0.218	0.647
	800	80.7	0.584	0.139	0.601
	1600	260.3	0.493	0.090	0.502
110	200	11.0	0.533	0.101	0.542
	400	28.6	0.460	0.061	0.464
	800	81.7	0.367	0.042	0.370
	1600	260.4	0.277	0.032	0.279

5 CONCLUSION

We have derived an asymptotic upper bound on the probability of error of the mesh estimator for pricing American options with respect to the number b of states sampled at each stage. Both the error size and the upper bound on the probability of error are functions of b that vanish as $b \to \infty$. The constant C appearing on the probability bound involves the fourth moment of the likelihood ratio of 1-step transition densities between a parent and a non-child

Table 5: Geometric Average Option on Seven Assets, $T = 10$.

x_0	b	CPU	$\widehat{\text{RB}}$	$\widehat{\text{RSE}}$	$\widehat{\text{RRMSE}}$
90	200	15.4	0.628	0.336	0.712
	400	39.5	0.635	0.269	0.690
	800	112.9	0.605	0.198	0.636
	1600	362.9	0.610	0.141	0.626
110	200	15.4	0.477	0.100	0.488
	400	39.3	0.455	0.061	0.459
	800	112.6	0.396	0.041	0.398
	1600	365.3	0.338	0.029	0.340

to another non-parent and the same child multiplied by the maximum future payoff over a path that starts at the child.

Despite the demonstrated guaranteed convergence of the mesh estimator under our mild required assumptions, our computational experience shows very poor behavior, specifically large positive bias. The bias is present even for small number of exercise opportunities, and decays slowly with b. In view of our theoretical result, we conclude that for the specific problems studied, the constant C is very large. This observation is consistent with the experience of many researchers that likelihood ratios are often highly variable random variables. We expect that the constant C grows rapidly with the problem dimension d and the number of exercise opportunities T, making the practical viability of the method questionnable. From an application perspective, we conclude that caution should be exercised when using this method.

REFERENCES

Barraquand, J., and D. Martineau. 1995. Numerical valuation of high dimensional multivariate american securities. *Journal of Financial and Quantitative Analysis* 30 (2):383–405.

Baxter, M., and A. Rennie. 1996. *Financial calculus*. Cambridge: Cambridge University Press.

Broadie, M., and P. Glasserman. 1997a. Monte Carlo methods for pricing high-dimensional american options: An overview. *The Electronic Journal of Financial Risk* 3 (December):15–37.

Broadie, M., and P. Glasserman. 1997b. Pricing american-style securities using simulation. *Journal of Economic Dynamics and Control* 21 (8-9):1323–1352.

Broadie, M., and P. Glasserman. 1997c, November. *A stochastic mesh nethod for pricing high-dimensional american options*. Working paper, Graduate School of Business, Columbia University, New York, NY 10027.

Duffie, D. 1996. *Dynamic asset pricing theory*. 2d ed. Princeton, New Jersey: Princeton University Press.

Harrison, M., and S. Pliska. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11:215–260.

AUTHOR BIOGRAPHIES

ATHANASSIOS (THANOS) N. AVRAMIDIS is an Invited Researcher at the Département d' Informatique et de Recherche Opérationnelle at the Universite de Montréal in Québec, Canada. He has been in the faculty at Cornell University and a consultant with SABRE Decision Technologies. His recent research interests are in Monte Carlo simulation methods with applications to derivative valuation, hedging, and risk management. He holds M.S. and Ph.D. degrees from the School of Industrial Engineering at Purdue University and a diploma in Mechanical Engineering from the University of Thessaloniki in Greece. He can be reached via e-mail at <avramidi@iro.umontreal.ca>.

HEINRICH MATZINGER is an Invited Researcher at the Department of Mathematics of the University of Bielefeld in Germany. He has been lecturer in the Department of Mathematics at MIT and a postdoctoral fellow at Eurandom in Holland. His recent research interests are in discrete probability. He holds a Ph.D. degree from the School of Industrial Engineering and Operations Research of Cornell University and a diploma of Mathematics from the Swiss Federal Institute of Technology in Zurich. He can be reached via e-mail at <kikimatz@yahoo.com>.