The Rate of Convergence of the Binomial Tree Scheme

John B. Walsh[1]

Department of Mathematics, University of British Columbia, Vancouver B.C. V6T 1Y4, Canada
(e-mail: walsh@math.ubc.ca)

Abstract. We study the detailed convergence of the binomial tree scheme. It is known that the scheme is first order. We find the exact constants, and show it is possible to modify Richardson extrapolation to get a method of order three-halves. We see that the delta, used in hedging, converges at the same rate. We analyze this by first embedding the tree scheme in the Black-Scholes diffusion model by means of Skorokhod embedding. We remark that this technique applies to much more general cases.

Key words: Tree scheme, options, rate of convergence, Skorokhod embedding

Mathematics Subject Classification (1991): 91B24, 60G40, 60G44

JEL Classification: G13

1 Introduction

The binomial tree scheme was introduced by Cox, Ross, and Rubinstein [1] as a simplification of the Black-Scholes model for valuing options, and it is a popular and practical way to evaluate various contingent claims. Much of its usefulness stems from the fact that it mimics the real-time development of the stock price, making it easy to adapt it to the computation of American and other options. From another point of view, however, it is simply a numerical method for solving initial-value problems for a certain partial differential equation. As such, it is known to be of first order [6], [7], [2], [3], at least for standard options. That is, the error varies inversely with the number of time steps.

A key point in typical financial problems is that the data is not smooth. For instance, if the stock value at term is x, the payoff for the European call option is of the form $f(x) = (x - K)^+$, which has a discontinuous derivative. Others, such as digital and barrier options, have discontinuous payoffs. This leads to

[1] I would like to thank O. Walsh for suggesting this problem and for many helpful conversations.

an apparent irregularity of convergence. It is possible, for example, to halve the step size and actually increase the error. This phenomenon comes from the discontinuity in the derivative, and makes it quite delicate to apply things such as Richardson extrapolation and other higher-order methods which depend on the existence of higher order derivatives in the data.

The aim of this paper is to study the convergence closely. We will determine the exact rate of convergence and we will even find an expression for the constants of this rate.

Merely knowing the form of the error allows us to modify the Richardson extrapolation method to get a scheme of order 3/2.

We will also see that the delta, which determines the hedging strategy, can also be determined from the tree scheme, and converges at exactly the same rate.

The argument is purely probabilistic. The Black-Scholes model treats the stock price as a diffusion process, while the binomial scheme treats it as a Markov chain. We use a procedure called Skorokhod embedding to embed the Markov chain in the diffusion process. This allows a close comparison of the two, and an accurate evaluation of the error. This was done in a slightly different way by C.R. Rogers and E.J. Stapleton, [9], who used it to speed up the convergence of the binomial tree scheme.

This embedding lets us split the error into two relatively easily analyzed parts, one which depends on the global behavior of the data, and the other which depends on its local properties.

2 Embeddings

The stock price (S_t) in the Black-Scholes model is a logarithmic Brownian motion, and their famous hedging argument tells us that in order to calculate option prices, the *discounted* stock price $\tilde{S}_t \stackrel{\text{def}}{=} e^{-rt}S_t$ should be a martingale. This hedging argument does not depend on the fact that the stock price is a logarithmic Brownian motion, but only on the fact that the market is complete: the stock prices in other complete-market models should also be martingales, at least for the purposes of pricing options. Even in incomplete markets, it is common to use a martingale measure to calculate option prices, at least as a first approximation.

It is a general fact [8] that any martingale can be embedded in a Brownian motion with the same initial value by Skorokhod embedding, and a strictly positive martingale can be embedded in a logarithmic Brownian motion. That means that one can embed the discounted stock price from other single-stock models in the discounted Black-Scholes stock price. Suppose for example, that $Y_k,\ k = 0, 1, 2, \dots$ is the stock price in a discrete model, and that $Y_0 = S_0$. Under the martingale measure, the discounted stock price $\tilde{Y}_k \stackrel{\text{def}}{=} e^{-kr\delta}Y_n$ is a martingale. Then there are (possibly randomized) stopping times $0 = \tau_0 < \tau_1 < \dots$ for S_t such that the processes $\{\tilde{Y}_k,\ k = 0, 1, 2, \dots\}$ and $\{\tilde{S}_{\tau_k},\ k = 0, 1, 2, \dots\}$ have exactly the same distribution. Thus the process $(\tilde{Y}_k)$ is embedded in $\tilde{S}_t$: $\tilde{Y}_k$

is just the process $\tilde{S}_t$ sampled at discrete times. However, the times are random, not fixed. This is what we mean by embedding.

We note that this embedding works for a single-stock market, but not in general for a multi-stock market, unless the stocks evolve independently, or nearly so.

Let f be a positive function. Suppose there is a contingent claim, such as a European option, which pays off an amount $f(S_T)$ at time T if the stock price at time T is S_T. If $S_0 = s_0$, its value at time zero is $V(s_0, 0) \equiv e^{-rT} E\{f(S_T)\}$. On the other hand, if $T = n\delta$, the same contingent claim for the discrete model pays $f(Y_n)$ at maturity and has a value at time zero of $U(s_0, 0) \stackrel{\text{def}}{=} e^{-rT} E\{f(Y_n)\}$. But $Y_n = e^{rT}\tilde{Y}_n$ has the same distribution as $e^{rT}\tilde{S}_{\tau_n}$, while $S_T = e^{rT}\tilde{S}_T$. Thus $U(s_0, 0) = e^{-rT} E\{f(e^{rT}\tilde{S}_{\tau_n})\}$, and the difference between the two values is

$$U(s_0, 0) - V(s_0, 0) = e^{-rT} E\{f(e^{rT}\tilde{S}_{\tau_n}) - f(e^{rT}\tilde{S}_T)\}. \qquad (1)$$

This involves the same process at two different times, the fixed time T and the random time τ_n. In cases such as the binomial tree, we have a good hold on the embedding times τ_n and can use this to get quite accurate estimates of the error. Although we will only embed discrete parameter martingales here, the theorem is quite general: it is used for the trinomial tree in [10]; one can even embed continuous martingales, so this could apply to models in which the stock price has discontinuities.

We should note that it is the discounted stock prices which are embedded, not the stock prices themselves, although there is a simple relation between the two. Rogers and Stapleton [9] have suggested modifying the binomial tree slightly in order to embed the stock prices directly.

3 The Tree Scheme

Let r be the interest rate. We will consider the Cox-Ross-Rubinstein binomial tree model for the *discounted* stock price. Let $\delta > 0$, and let the stock price at time $t = k\delta$ be Y_k; the discounted price is $\tilde{Y}_k = e^{-rk\delta} Y_k$. We will assume the probability measure is the martingale measure, so that $(\tilde{Y}_k)$ is a martingale, and we assume it takes values in the discrete set of values a^j, $j = 0, 1, 2, \ldots$. At each step, Y_k can jump to one of two possible values: either $Y_{k+1} = aY_k$ or $Y_{k+1} = a^{-1}Y_k$, where $a > 1$ is a real number. The martingale property assures us that

$$P\{\tilde{Y}_{j+1} = a\tilde{Y}_j \mid \tilde{Y}_j\} = \frac{1}{a+1} \stackrel{\text{def}}{=} q, \quad P\{\tilde{Y}_{j+1} = a^{-1}\tilde{Y}_j \mid \tilde{Y}_j\} \stackrel{\text{def}}{=} 1 - q\,.$$

so $(\tilde{Y}_k)$ is a Markov chain with these transition probabilities.

Let $f(x)$ be a positive function, and consider the contingent claim which pays $f(Y_T)$ at time T, for some given function f. Fix an integer n and let $\delta = T/n$. If $Y_0 = s_0$, then the value of the claim at time zero is $U(s_0, 0)$, and its value at some intermediate time $t = k\delta$ is

$$U(\tilde{Y}_k, k) \stackrel{\text{def}}{=} e^{-r(T-k\delta)} E\{f(Y_n) \mid Y_k\} = e^{-r(T-k)\delta)} E\{f(e^{rT}\tilde{Y}_n) \mid Y_k\}\,. \tag{2}$$

Let $u(j,k) = U(a^j, k)$. Then u is the solution of the difference scheme

$$\begin{cases} u(j,k) &= e^{-r\delta}\big(q\,u(j+1,k+1) + (1-q)\,u(j-1,k+1)\big),\ j \in \mathbb{Z},\ k = 0,\dots,n-1\,, \\ u(j,n) &= f(e^{rT}a^j),\ j \in \mathbb{Z}\,. \end{cases} \tag{3}$$

Under its own martingale measure, the corresponding Black-Scholes model will have a stock price given by

$$S_t = S_0 e^{\sigma W_t + (r - \frac{1}{2}\sigma^2)t},\ t \geq 0\,, \tag{4}$$

where W_t is a standard Brownian motion and $\sigma > 0$ is the volatility. The discounted stock price is the martingale $\tilde{S}_t = e^{\sigma W_t - \frac{1}{2}\sigma^2 t}$. In this model, the above contingent claim pays $f(S_T)$ at time T, its value at time zero is $V(s_0, 0)$, and its value at an intermediate time $0 < t < T$ is

$$V(S_t, t) \stackrel{\text{def}}{=} e^{-r(T-t)} E\{f(S_T) \mid S_t\}\,. \tag{5}$$

There is a relation between a, δ, n, T, and σ which connects these models:

$$\delta = \frac{T}{n}\,, \quad \log a = \sigma\sqrt{\frac{T}{n}}\,.$$

If we let n, j, and k tend to infinity in such a way that $kT/n \longrightarrow t$ and $e^{j\sigma\sqrt{T/n} + krT/n} \to x$, then $u(j,k)$ will converge to $V(x,t)$. The question we will answer is "How fast?"

4 Results

We say that a function f is *piecewise* $C^{(k)}$ if $f, f', \dots, f^{(k)}$ have at most finitely many discontinuities and no oscillatory discontinuities. We will treat the following class of possible payoff functions.

Definition 4.1 *Let $\mathcal{K}$ be the class of real-valued functions f on $\mathbb{R}$ which satisfy*
(i) f is piecewise $C^{(2)}$;
(ii) at each x, $f(x) = \frac{1}{2}(f(x+) + f(x-))$.
(iii) f, f', and f'' are polynomially bounded: i.e. there exist $K > 0$ and $p > 0$ such that $|f(x)| + |f'(x)| + |f''(x)| \leq K(1 + |x|^p)$ for all x.

Let us introduce some notation which will be in force for the remainder of the paper. Let $f \in \mathcal{K}$ and consider a contingent claim which pays an amount $f(s)$ at a fixed time $T > 0$ if the stock price at time T is s. Let n be the number of time steps in the discrete model, so that the time-step is $\delta = T/n$. The space step h is then $h = \sigma\sqrt{T/n}$.

The error depends on the discontinuities of f and f', and on the relation of these discontinuities to the lattice points.

$$\begin{aligned}\Delta f(s) &= f(s^+) - f(s^-);\\ \Delta f'(s) &= f'(s^+) - f'(s^-);\\ \theta(s) &= \mathrm{frac}\Big(\frac{\log s}{2h}\Big)\end{aligned}$$

where $\mathrm{frac}(x)$ is the fractional part of x.

Let the initial price be $S_0 = s_0$. The value in the Black-Scholes model is given by $V(s_0, 0)$, and its value in the binomial tree scheme is given by $U(s_0, 0)$, so the error of the tree scheme is defined to be

$$\mathcal{E}_{\mathrm{tot}}(f) \stackrel{\mathrm{def}}{=} U(s_0, 0) - V(s_0, 0)\,. \tag{6}$$

Let $h\mathbb{Z}$ be the set of all multiples of h, $\mathbb{N}_e^h \stackrel{\mathrm{def}}{=} 2h\mathbb{Z}$ the set of all *even* multiples of h, and $\mathbb{N}_o^h \stackrel{\mathrm{def}}{=} h + \mathbb{N}_e^h$ the set of all *odd* multiples of h. The density of $X_T \stackrel{\mathrm{def}}{=} \log(\tilde{S}_T/s_0) \equiv \sigma W_T - \sigma^2 T/2$ is

$$\hat{p}(x) \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{-\frac{(x+\frac{1}{2}\sigma^2 T)^2}{2\sigma^2 T}}\,.$$

The main result is Theorem 4.2 below, but we will first give a simple and easily-used corollary.

Corollary 4.1 *Suppose $f \in \mathcal{K}$ and that n is an even integer. If f is discontinuous, and if the discontinuity is not on a lattice point, then $\mathcal{E}_{tot}(f) = O(n^{-1/2})$. If all discontinuities are on lattice points, then $\mathcal{E}_{tot}(f)$ is $O(n^{-1})$. Moreover, if f is a European call or put of strike price K, if $\tilde{K} = Ke^{-rT}$ is the discounted strike price, and if s_0 is the initial stock price, then $\mathcal{E}_{tot}(f)$ is of the form*

$$\mathcal{E}_{tot}(f) = \big(A + B\theta(1-\theta)\big)\frac{1}{n} + O(n^{-3/2})\,, \tag{7}$$

where $\theta = \theta(\tilde{K}/s_0)$, A is a constant which depends on f, and

$$B = 2\sigma^2 T K \Delta f'(K)\, \hat{p}(\log \tilde{K}/s_0)\,.$$

This is a special case of Theorem 4.2 below, so there is no need for a separate proof. We collect (10), and Propositions 9.5, 9.6 and 9.7, and use (46) to express them in terms of f instead of g. We get:

Theorem 4.2 *Suppose that $f \in \mathcal{K}$. Let $s_1, s_2, \ldots, s_k$ be the set of discontinuity points of f and f', and let s_0 be the initial stock price. For any real s, let $\tilde{s} = se^{-rT}$. Let n be an even integer. Then the error in the tree scheme is*

$$
\begin{aligned}
\mathcal{E}_{tot}(f) = \frac{e^{-rT}}{n}\Bigg[&\Big(\frac{5}{12}+\frac{\sigma^2 T}{6}+\frac{\sigma^4 T^2}{192}\Big)E\Big\{f(S_T)\Big\}-\frac{1}{6\sigma^2 T}\,E\Big\{(\log(\tilde S_T/s_0))^2 f(S_T)\Big\}\\
&-\frac{1}{12\sigma^4 T^2}\,E\Big\{(\log(\tilde S_T/s_0))^4 f(S_T)\Big\}+\frac{2}{3}\sigma^2 T\,E\Big\{S_T^2 f''(S_T)\Big\}\\
+\sigma^2 T\sum_i&\Big(s_i\,\Delta f'(s_i)-\frac{1}{2}\Delta f(s_i)\Big)\Big(\frac{1}{3}+2\theta(\tilde s_i/s_0)(1-\theta(\tilde s_i/s_0))\Big)\hat p(\log(\tilde s_i/s_0))\\
&-\frac{1}{3}\sum_{i:\log(\tilde s_i/s_0)\in\mathbb{N}_e^h}\log(\tilde s_i/s_0)\,\Delta f(s_i)\,\hat p(\log(\tilde s_i/s_0))\\
&+\frac{1}{6}\sum_{i:\log(\tilde s_i/s_0)\in\mathbb{N}_o^h}\log(\tilde s_i/s_0)\,\Delta f(s_i)\,\hat p(\log(\tilde s_i/s_0))\Bigg]\\
+e^{-rT}\frac{\sigma\sqrt{T}}{\sqrt{n}}&\sum_{i:\log(\tilde s_i/s_0)\notin h\mathbb{Z}}(2\theta(\tilde s_i/s_0)-1)\,\Delta f(s_i)\hat p(\log(\tilde s_i/s_0))+O\Big(\frac{1}{n^{3/2}}\Big) \qquad (8)
\end{aligned}
$$

where the expectations are taken with respect to the martingale measure.

Remark 4.3 We have expressed the errors in terms of $E\{f(S_T)\}$. However, we can also express them in terms of $e^{rT}\tilde S_{\tau_n}$, and it might be better to do so, since this is exactly what the binomial scheme computes. Indeed, the theorem tells us that the expectations of $f(S_T)$ and $f(e^{rT}\tilde S_{\tau_n})$ only differ by $O(1/n)$, and they occur as coefficients multiplying $1/n$ in (8) so one can replace S_T by $e^{rT}\tilde S_{\tau_n}$ in (8) and the result will only change by $O(n^{-2})$, so these formulas remain correct. So in fact S_T and $e^{rT}\tilde S_{\tau_n}$ are interchangeable in (8); and, for the same reason, both are interchangeable with S_{τ_n}.

The delta, which determines the hedging strategy in the Black-Scholes model, can also be estimated in the tree scheme, and its estimate also converges with order one. (See Section 10.) Let $\check\theta(s) = \text{frac}\,(\frac{h+\log s}{2h})$

Corollary 4.4 *Suppose that f is continuous and both f and f' are in $\mathcal{K}$. The symmetric estimate (35) of the delta converges with order one. For a call or put option with strike price K, there are constants A and B such that the error at time 0 is of the form*

$$
\big(A+B\check\theta(\tilde K)(1-\check\theta(\tilde K))\big)\frac{1}{n}+o(n^{-1})\,. \qquad (9)
$$

5 Remarks and Extensions

1. The random walk $\tilde Y_k$ is periodic, and alternates between even and odd lattice points. This leads to a well-understood even/odd fluctuation in the tree scheme. To avoid this, we work exclusively with even values of n. We could as well have worked exclusively with odd values, but not with both.

2. The striking fact about the tree scheme's convergence is that, even when restricted to even values of n, the error goes to zero at the general rate of $O(1/n)$, but "with a wobble:" there are constants $c_1 < c_2$ for which $c_1/n < \mathcal{E}_{\text{tot}}(f) < c_2/n$, and the error fluctuates quasi-periodically between these bounds.[2]

The reason is clear from (8). For example, a typical European call with strike price K pays off $f(x) = (x - K)^+$ and (8) simplifies: the last three series vanish, and the first reduces to the single term

$$\sigma^2 TK\Big(\frac{1}{3} + 2\theta(1-\theta)\Big)\hat{p}(\log(\tilde{s}/s_0))\,.$$

The quantity to focus on is θ. It is in effect the fractional distance (in log scale) from $\tilde{K}$ to the nearest even lattice point. In log scale, the lattice points are multiples of $\sigma\sqrt{T/n}$, so the whole lattice changes as n changes. This means that θ changes with n too. It can vary from 0 to 1, so this term can vary by a factor of nearly three. It is not the only error term, but it is important, and it is why there are cases where one can double the number of steps and more than double the error at the same time.

3. The coefficients in Theorem 4.2 are rather complex, and Corollary 4.1 is handier for vanilla options. It shows that one can make a Richardson-like extrapolation to increase the order of convergence. If we run the tree for three values of n which give different values of θ, we can then write down (7) for the three, solve for the coefficients A and B, and subtract off the first order error terms, giving us potentially a scheme of order 3/2. In fact, one could do this cheaply: use two runs at roughly the square root of n, and then one at n. This might be of interest when using the scheme to value American options.

4. It is usually the raw stock price, not the discounted price, which evolves on the lattice. However, our numerical studies have shown that the behavior of the two schemes is virtually identical: to adapt Corollary 4.1 to the evolution of the raw price, just replace the discounted strike price $\tilde{K}$ by the raw strike price K in the definition of θ. We have therefore used the discounted price for its convenience in the embedding.

5. From a purely probabilistic point of view, Theorem 4.2 is a rate-of-convergence result for a central limit theorem for Bernoulli random variables. If we take f to be the indicator function of $(-\infty, z]$, we recover the Berry-Esseen bound. (We thank the referee for pointing this out.)

6 Embedding the Markov Chain in the Diffusion

The argument in Section 2 showed that the tree scheme could be embedded in a logarithmic Brownian motion, but didn't say how. In fact the embedding times

[2] F. and M. Diener [2], [3] have investigated this wobble from a quite different point of view, based on an asymptotic expansion of the binomial coefficients derived by a modification of Laplace's method.

can be defined explicitly. Define stopping times $\tau_0, \tau_1, \tau_2 \dots$ by induction:

$$\tau_0 = 0, \quad \tau_{k+1} = \inf\{t > \tau_k : \tilde{S}_t = a\tilde{S}_{\tau_k} \text{ or } a^{-1}\tilde{S}_{\tau_k}\}\,.$$

As $\tilde{S}_t$ is a martingale, so is $\tilde{S}_{\tau_0}, \tilde{S}_{\tau_1}, \dots$ Since $\tilde{S}_{\tau_{k+1}}$ can only equal $a\tilde{S}_{\tau_k}$ or $a^{-1}\tilde{S}_{\tau_k}$, we must have

$$P\{\tilde{S}_{\tau_{k+1}} = a\tilde{S}_{\tau_k} \mid \tilde{S}_{\tau_k}\} = \frac{1}{a+1} \quad P\{\tilde{S}_{\tau_{k+1}} = a^{-1}\tilde{S}_{\tau_k} \mid \tilde{S}_{\tau_k}\} = \frac{a}{a+1}\,.$$

It follows that $(\tilde{S}_{\tau_k})$ is a Markov chain with the same transition probabilities as $(\tilde{Y}_k)$; since $\tilde{S}_{\tau_0} = Y_0 = 1$, the two are *identical processes*. It follows that the error in the binomial scheme (considered as an approximation to the Black-Scholes model) is given by

$$\mathcal{E}_{\text{tot}}(f) \stackrel{\text{def}}{=} u(1,0) - v(1,0) = e^{-rT} E\{f(e^{rT}\tilde{S}_{\tau_n}) - f(e^{rT}\tilde{S}_T)\}\,. \tag{10}$$

Here is a quick heuristic argument to show that the convergence is first order. Expand $E\{f(S_{T+s})\}$ in a Taylor series. It is

$$E\{f(X_{T+s})\} = E\{f(S_T)\} + a_1 s + a_2 s^2 + O(s^3)\,.$$

for some a_1 and a_2. Now $\tau_n = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1})$ is a sum of i.i.d. random variables, so it has mean T and variance $n\,\text{var}(\tau_1) = c/n$, (see Prop 11.1) so that $E\{\tau_n - T\} = 0$ and $E\{(\tau_n - T)^2\} = c/n$. Moreover, τ_n is independent of the sequence (S_{τ_j}), so if we stretch things a bit and assume it is independent of (S_t), and set $s = \tau_n - T$, we would have

$$\begin{aligned} E\{f(S_{\tau_n}) - f(S_T)\} &\sim E\{a_1(\tau_n - T) + a_2(\tau_n - T)^2\} \\ &= a_2 c/n \end{aligned} \tag{11}$$

implying that the error is indeed $O(1/n)$.

This argument is not rigorous, since τ_n is a function of the process (S_t), so it can't be independent of it. Nevertheless, the dependence is essentially a local property, and we can isolate it by breaking the error into a global part, on which this argument is rigorous, and a local part, which can be handled directly.

One lesson to draw from the above is that it is important that $E\{\tau_n\} = T$. If it were not, the lowest order error term above would not drop out, and would in fact make a contribution of $O(1/\sqrt{n})$.

7 Splitting the Error

We will make two simplifying transformations. First, we take logarithms of the stock price: set $X_t \stackrel{\text{def}}{=} \log(\tilde{S}_t/s_0) = \sigma W_t - \frac{1}{2}\sigma^2 t$. From the form of the times τ_j we see that (X_{τ_j}) is a random walk on $h\mathbb{Z}$, the integer multiples of h, while (X_t) is a Brownian motion with drift.

Next, we make a Girsanov transformation to remove the drift of X_t. Let ξ be the maximum of T, τ_n, and τ_J, where τ_J is defined below—the value of ξ is not important, so long as it is larger than the values of t we work with—and set

$$dQ = e^{\frac{1}{2}X_\xi + \frac{1}{8}\sigma^2\xi}\, dP\,.$$

By Girsanov's Theorem [4], $\{\frac{1}{\sigma}X_t,\ 0 \le t \le \xi\}$ is a standard Brownian motion on $(\Omega, \mathcal{F}, Q)$. We will call Q the *Brownian measure* to distinguish it from the martingale measure P. We will do all our calculations in terms of Q, and then translate the results back to P at the very end. Under the measure Q, X_t is a Brownian motion, and (X_{τ_j}) is a simple symmetric random walk on $h\mathbb{Z}$. It alternates between even and odd multiples of h. To smooth this out, we will restrict ourselves to even values of j and n.

Thus let $n = 2m$ for some integer m and define

$$J = \inf\{2j : \tau_{2j} > T\}\,.$$

Then J is a stopping time for the sequence $\tau_0,\ \tau_1,\ \dots$ with even-integer values, and τ_J is a stopping time for (X_t). Notice that $\tau_J > T$. We expect that $\tau_J \sim T$. In terms of the martingale measure P, the error is

$$\begin{aligned}
\mathcal{E}_{\text{tot}}(f) &= e^{-rT}E\{f(S_{\tau_n}) - f(S_{\tau_J})\} + e^{-rT}E\{f(S_{\tau_J}) - f(S_T)\} \\
&\stackrel{\text{def}}{=} \mathcal{E}_{\text{glob}}(f) + \mathcal{E}_{\text{loc}}(f)\,.
\end{aligned}$$

As the notation suggests, $\mathcal{E}_{\text{glob}}(f)$ depends on global properties of f, such as its integrability, while $\mathcal{E}_{\text{loc}}(f)$ depends on local properties, such as its continuity and smoothness. Notice that these concern the signed error, not the absolute error.

Define a function g by

$$g(x) \stackrel{\text{def}}{=} f(s_0 e^{x+rT}) e^{-\frac{x}{2} - \frac{\sigma^2 T}{8}}\,. \tag{12}$$

In terms of Q, the error in (10) is

$$\mathcal{E}_{\text{tot}}(f) = e^{-rT}E^Q\left\{\big(f(s_0e^{X_{\tau_n}+rT}) - f(s_0e^{X_T+rT})\big)e^{-\frac{1}{2}X_\xi - \frac{1}{8}\sigma^2\xi}\right\}. \tag{13}$$

Now $e^{-\frac{1}{2}X_t - \sigma^2 t/8}$ is a Q-martingale, so as $\tau_n \le \xi$,

$$\begin{aligned}
E^P\{f(S_{\tau_n})\} &= E^Q\big\{f(S_{\tau_n})e^{-\frac{1}{2}X_\xi - \sigma^2\xi/8}\big\} \\
&= E^Q\big\{f(S_{\tau_n})e^{-\frac{1}{2}X_{\tau_n} - \sigma^2\tau_n/8}\big\} = E^Q\big\{g(X_{\tau_n})e^{-\sigma^2(\tau_n - T)/8}\big\}
\end{aligned}$$

since $S_{\tau_n} = s_0 e^{X_{\tau_n}}$. Similarly

$$E^P\{f(S_T)\} = E^Q\{g(X_T)\}\,.$$

Thus

$$\mathcal{E}_{\text{tot}}(f) = e^{-rT} E^Q\{g(X_{\tau_n}) - g(X_T)\} + e^{-rT} E^Q\big\{g(X_{\tau_n})\left(e^{-\frac{\sigma^2}{8}(\tau_n - T)} - 1\right)\big\}\,.$$

Now X_t is a Q-Brownian motion, so that the times $\tau_1,\ \tau_2, \dots$ are independent of $X_{\tau_1},\ X_{\tau_2},\ \dots$, (see Proposition 11.1) and the last term above is

$$e^{-rT} E^Q\{g(X_{\tau_n})\} E^Q\big\{e^{-\frac{\sigma^2}{8}(\tau_n - T)} - 1\big\}\,.$$

But now $\tau_n - T = (\tau_1 - T/n) + (\tau_2 - \tau_1 - T/n) + \cdots + (\tau_n - \tau_{n-1} - T/n)$; the summands are i.i.d., so the last expectation is

$$E^Q\big\{e^{-\frac{\sigma^2}{8}(\tau_1 - \frac{T}{n})}\big\}^n = \left(\frac{e^{\frac{\sigma^2 T}{8n}}}{\cosh\sqrt{\frac{\sigma^2 T}{4n}}}\right)^n = 1 + \frac{\sigma^4 T^2}{192n} + O(1/n^2)\,.$$

where we have used Proposition 11.1 and expanded in powers of $1/n$. Thus

$$\mathcal{E}_{\text{tot}}(f) = e^{-rT} E^Q\{g(X_{\tau_n}) - g(X_T)\} + e^{-rT}\frac{\sigma^4 T^2}{192n} E^Q\{g(X_{\tau_n})\} + O(\frac{1}{n^2})$$

$$= e^{-rT} E^Q\{g(X_{\tau_n}) - g(X_{\tau_J})\} + e^{-rT} E^Q\{g(X_{\tau_J}) - g(X_T)\} + e^{-rT}\frac{\sigma^4 T^2}{192n} E^Q\{g(X_{\tau_n})\} + O(\frac{1}{n^2})$$

$$\stackrel{\text{def}}{=} \hat{\mathcal{E}}_{\text{glob}}(g) + \hat{\mathcal{E}}_{\text{loc}}(g) + e^{-rT}\frac{\sigma^4 T^2}{192n} E^Q\{g(X_{\tau_n})\} + O(1/n^2) \quad (14)$$

which defines $\hat{\mathcal{E}}_{\text{glob}}(g)$ and $\hat{\mathcal{E}}_{\text{loc}}(g)$. The final term comes from the fact that we defined g with a fixed time T instead of the random time ξ when we changed the probability measure.

This splits the error into two parts. The global error $\hat{\mathcal{E}}_{\text{glob}}(g)$ can be handled with a suitable modification of the Taylor series argument of (11). The local error $\hat{\mathcal{E}}_{\text{loc}}(g)$ can be computed explicitly, and it is here that the local properties such as the continuity and differentiability of g come into play.

8 The Global Error

Let us first look at the global error in (14).

Theorem 8.1 *Let g be measurable and exponentially bounded. Then*

$$\hat{\mathcal{E}}_{glob}(g) = \frac{1}{6n}\Big[\frac{5}{2} E^Q\{g(X_{\tau_n})\} - \frac{1}{\sigma^2 T} E^Q\{X_{\tau_n}^2\, g(X_{\tau_n})\} - \frac{1}{12\sigma^4 T^2} E^Q\{X_{\tau_n}^4 g(X_{\tau_n})\}\Big] + O(n^{-\frac{3}{2}})\,. \quad (15)$$

Proof. Let $P_n(x)$ be the transition probabilities of a simple symmetric random walk on the integers, so that $P_j(x) = P^Q\{X_{\tau_j} = hx\}$. Let us remark that J is independent of (X_{τ_j}) so that

$$P^Q\{X_{\tau_J} = hx\} = \sum_{k=-n}^{\infty} P^Q\{J - n = k\} P_{n+k}(x)\,,$$

and, for integers p, q, and r,

$$\sum_{k=-n}^{\infty} \sum_{x=-n}^{n} P^Q\{J-n = k\} P_n(x) \frac{k^p x^q}{n^r} g(xh) = E^Q\Big\{\Big(\frac{J-n}{\sqrt{n}}\Big)^p\Big\}\, E^Q\big\{X_{\tau_n}^q\, g(X_{\tau_n})\big\} \frac{n^{\frac{p+q}{2}-r}}{(\sigma\sqrt{T})^q}\,. \tag{16}$$

By Proposition 11.2 of the Appendix, the two expectations are bounded, so if $p \neq 1$ this term has order $\frac{p+q}{2} - r$, which is the *effective order* of $\frac{k^p x^q}{n^r}$. By Corollary 11.4, the contributions to this integral for $|x| > n^{\frac{3}{5}}$ and/or $|k| > n^{\frac{3}{5}}$ go to zero faster than any power of n. Thus we can restrict ourselves to the sum over the values $\max|x|, |k| \leq n^{\frac{3}{5}}$, in which case $P_{n+k}(x)$ and $P_n(x)$ are both defined, and

$$\begin{aligned} E^Q\{g(X_{\tau_n}) - g(X_{\tau_J})\} &= \sum_k \sum_x P^Q\{J - n = k\}\big(P_n(x) - P_{n+k}(x)\big)\, g(hx) \\ &= \sum_k \sum_x \Big(\frac{k}{2n} - \frac{3k^2 + 4kx^2}{8n^2} + \frac{3k^2x^2}{4n^3} - \frac{k^2x^4}{8n^4} + Q_3\Big) P_n(x)\; g(hx)\,, \end{aligned}$$

by Proposition 11.5, where Q_3 is a sum of terms of effective order at most $-\frac{3}{2}$. By (16), we identify this as

$$\begin{aligned} \frac{1}{n}\Big[&\Big(\frac{1}{2}E^Q\{J-n\} - \frac{3}{8n}E^Q\{(J-n)^2\}\Big)\; E\{g(X_{\tau_n})\} \\ &- \frac{1}{\sigma^2 T}\Big(\frac{1}{2}E^Q\{J-n\} - \frac{3}{4n}E^Q\{(J-n)^2\}\Big)\; E^Q\{X_{\tau_n}^2\; g(X_{\tau_n})\} \\ &- \frac{1}{8\sigma^4 T^2 n} E^Q\{(J-n)^2\} E^Q\{X_{\tau_n}^4 g(X_{\tau_n})\Big] + O(n^{-3/2})\,. \end{aligned} \tag{17}$$

Proposition 11.2 gives the values of $E\{J-n\} = 4/3 + O(h)$ and $E\{(J-n)^2\} = 2n/3 + O(1)$. Substituting, we get (15). ∎

9 The Local Error

The local error, $\mathcal{E}_{\text{loc}}$ comes from the interval of time between T and τ_J. This is short, but it is where the local properties of the payoff function f come in.

We will express this in terms of g rather than f. Now g inherits the continuity and differentiability properties of f, and the polynomial boundedness of f

translates into exponential boundedness of g: there exist $A > 0$ and $a > 0$ such that $|g(x)| \leq Ae^{a|x|}$ for all x.

Let $\mathbb{N}_e^h \stackrel{\text{def}}{=} 2h\mathbb{Z}$ and $\mathbb{N}_o^h \stackrel{\text{def}}{=} h + \mathbb{N}_e^h$ be the sets of even and odd multiples of h respectively. Recall that J was the first *even* integer j such that $\tau_j > T$. Let us define

$$L \stackrel{\text{def}}{=} \sup\{j : X_{\tau_j} < T\}\,, \tag{18}$$

so that τ_L is the last stopping time before T.

There are two cases. Either L is an odd integer, in which case $X_{\tau_L} \in \mathbb{N}_o^h$, $L = J - 1$, and $\tau_L = \tau_{J-1} < T < \tau_J$; or L is an even integer, in which case $X_{\tau_L} \in \mathbb{N}_e^h$, $L = J - 2$, $\tau_L = \tau_{J-2} < T < \tau_{J-1}$. Note that in either case, $\tau_L \leq t \leq T \Longrightarrow |X_t - X_{\tau_L}| < h$.

Define two operators, Π_e and Π_o on functions $u(x),\ x \in \mathbb{R}$ by:

- $\Pi_e u(x) = u(x)$ if $x \in \mathbb{N}_e^h$, and $x \mapsto \Pi_e u(x)$ is linear in each interval $[2kh, (2k+2)h]$, $k \in \mathbb{N}$.
- $\Pi_o u(x) = u(x)$ if $x \in \mathbb{N}_o^h$, and $x \mapsto \Pi_o u(x)$ is linear in each interval $[(2k-1)h, (2k+1)h]$, $k \in \mathbb{N}$.

Thus $\Pi_e u$ and $\Pi_o u$ are linear interpolations of u in between the even (respectively odd) multiples of h.

Apply the Markov property at T. X_t is a Brownian motion from T on, and if L is odd, then τ_J is the first time after T that X_t hits $\mathbb{N}_e^h$, so, using the known hitting probabilities of Brownian motion,

$$E\{g(X_{\tau_J}) \mid X_T,\ L \text{ is odd }\} = \Pi_e g(X_T)\,. \tag{19}$$

On the other hand, if L is even, $X_{\tau_L} \in \mathbb{N}_e^h$, so that τ_{J-1} is the first time after T when $X_t \in \mathbb{N}_o^h$, and τ_J is the first time after τ_{J-1} that $X_t \in \mathbb{N}_e^h$. Moreover, τ_{J-1} coincides with a stopping time when L is even, so we can apply the Markov property at τ_{J-1} to see

$$E\{g(X_{\tau_J}) \mid X_{\tau_{J-1}},\ L \text{ is even }\} = \Pi_e g(X_{\tau_{J-1}})\,.$$

But if L is even, τ_{J-1} is the first hit of $\mathbb{N}_o^h$, so

$$\begin{aligned} E\{g(X_{\tau_J}) \mid X_T,\ L \text{ is even }\} &= E\{\Pi_e g(X_{\tau_{J-1}}) \mid X_T,\ L \text{ is even }\} \\ &= \Pi_o \Pi_e g(X_T)\,. \end{aligned}$$

Let

$$q(x) \stackrel{\text{def}}{=} P\{L \text{ is even } \mid X_T = x\}\,.$$

Note that L is even if and only if $X_{\tau_L} \in \mathbb{N}_e^h$, and if this is so, X_t does not hit $\mathbb{N}_o^h$ between τ_L and T. Reverse X_t from time T: let $\hat{X}_t \stackrel{\text{def}}{=} X_{T-t}$, $0 \leq t \leq T$. Then L is even if and only if $\hat{X}_t$ hits $\mathbb{N}_e^h$ before it hits $\mathbb{N}_o^h$. But now, if we condition on $X_T = x$, or equivalently on $\hat{X}_0 = x$, then $\{\hat{X}_t,\ 0 \leq t \leq T\}$ is a Brownian bridge, and $\hat{X}_t - \hat{X}_0$ is a Brownian motion. Thus we can calculate the exact probability that it hits $\mathbb{N}_e^h$ before $\mathbb{N}_o^h$. More simply, we can just note that if h is small, the probability of hitting $\mathbb{N}_e^h$ before $\mathbb{N}_o^h$ is not much influenced by the drift, and so it is approximately that of unconditional Brownian motion. Thus, if $\hat{X}_0 = x \in (2kh, (2k+1)h)$, $q(x) = P\{\hat{X}_t$ reaches $2kh$ before $(2k+1)h\} \sim \frac{(2k+1)h-x}{h} = \mathrm{dist}(x, \mathbb{N}_o^h)/h$, where $\mathrm{dist}(x, \Lambda)$ is the distance from x to the set Λ.

Thus

$$q(x) = \frac{1}{h}\mathrm{dist}(x, \mathbb{N}_o^h) + O(h)\,. \tag{20}$$

Proposition 9.1

$$E\{g(X_{\tau_J}) - g(X_T)\} = E\{\Pi_e g(X_T) - g(X_T)\} + E\{(\Pi_o\Pi_e g(X_T) - \Pi_e g(X_T))q(X_T)\}.$$

Proof. Let us write $E\{g(X_{\tau_J})\} = E\{g(X_{\tau_J}), L \text{ even }\} + E\{g(X_{\tau_J}), L \text{ odd }\}$. Note that $\{L \text{ odd }\} \in \mathcal{F}_T$, so it is conditionally independent of $\{X_{T+t},\ t \geq 0\}$ given X_T. Thus

$$\begin{aligned} E\{g(X_{\tau_J})\} = E\Big\{ & E\{g(X_{\tau_J}) \mid X_T,\ L \text{ even }\}P\{L \text{ even } \mid X_T\} \\ & + E\{g(X_{\tau_J}) \mid X_T,\ L \text{ odd }\}P\{L \text{ odd } \mid X_T\}\Big\}. \end{aligned}$$

By the definition of q and the above relations, we see this is

$$\begin{aligned} &= E\{\Pi_o\Pi_e g(X_T) q(X_T)\} + E\{\Pi_e g(X_T)(1 - q(X_T))\} \\ &= E\big\{\big(\Pi_o\Pi_e g(X_T) - \Pi_e g(X_T)\big)q(X_T)\big\} + E\{\Pi_e g(X_T)\}\,, \end{aligned}$$

which proves the proposition. ∎

Definition 9.1 Let $p(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{-\frac{x^2}{2\sigma^2 T}}$ be the density of X_T under the Brownian measure.

Corollary 9.2 *Let $\Delta_k = (\Pi_e g)'(2kh+) - (\Pi_e g)'(2kh-)$ be the jump in the derivative of $\Pi_e g$ at the point $2kh$. Then*

$$\mathcal{E}_{loc}(g) = \int_{-\infty}^{\infty} (\Pi_e g(x) - g(x))p(x)\,dx + \frac{h^2}{3}\sum_k (\Delta_k)\,p(2kh) + O(h^3)\,. \tag{21}$$

Proof. The first integral equals the first expectation on the right hand side of Proposition 9.1. The second expectation can be written

$$\int_{-\infty}^{\infty} (\Pi_o\Pi_e g(x) - \Pi_e g(x))q(x)p(x)\,dx\,. \tag{22}$$

To simplify the second term, let $\xi(x) = \Pi_e g(x)$. Then ξ is piecewise linear with vertices on $\mathbb{N}_e^h$, so we can write it in the form

$$\xi(x) = ax + b + \sum_k \frac{1}{2}\Delta_k |x - 2kh|$$

for some a and b. Since Π_o is a linear operator and $\Pi_o(ax+b) \equiv ax+b$, we see this is

$$= \frac{1}{2}\sum_k \Delta_k \int_{-\infty}^{\infty} (\Pi_o|x-2kh| - |x-2kh|)q(x)p(x)\,dx\,.$$

Now $|x-2kh|$ is linear on both $(-\infty, 2kh)$ and $(2kh, \infty)$, so that $\Pi_o|x-2kh| = |x-2kh|$ except on the interval $[(2k-1)h, (2k+1)h]$. On that interval, $\Pi_o|x-2kh| \equiv h$ and $q(x) = (h-|x-2kh|)/h$, for $q(x)$ is approximately $1/h$ times the distance to the nearest odd multiple of h. Write $p(x) = p(2kh) + O(h)$ there. Then

$$\begin{aligned}\int_{-\infty}^{\infty} (\Pi_o|x-2kh| - |x-2kh|)q(x)p(x)\,dx &= \frac{1}{h}\int_{(2k-1)h}^{(2k+1)h} (h-|x-2kh|)^2(p(2kh)+O(h))\,dx \\ &= \frac{2}{3}h^2 p(2kh) + O(h^3)\,. \qquad (23)\end{aligned}$$

If we remember that if $g = |x-2kh|$, $\Delta_k = 2$, the corollary follows. ∎

We can decompose g as follows. Define the modified Heaviside function $\tilde{H}(x)$ by

$$\tilde{H}(x) = \begin{cases} 1 & \text{if } x > 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Lemma 9.3 *Let g be piecewise $C^{(2)}$. Then we can write $g = g_1 + g_2 + g_3$, where g_1 is a step function with at most finitely many discontinuities, g_2 is continuous and piecewise linear, and $g_3 \in C^{(1)}$ with g_3'' piecewise continuous. Moreover, we have*

$$g_1(x) = \sum_y \Delta g(y)\tilde{H}(x-y)\,, \quad g_2(x) = \sum_y \frac{1}{2}(\Delta g')(y)|x-y|\,. \qquad (24)$$

Proof. The sums in (24) are finite. By the definition of $\mathcal{K}$, $g(x) = \frac{1}{2}(g(x+) + g(x-))$ at any discontinuity of g. It is easy to check that if we define g_1 by (24), then $g - g_1$ is continuous. (This is the reason we modified the Heaviside function.) However, it may still have a finite number of discontinuities in its derivative. We remove these by subtracting g_2: $g_3 \stackrel{\text{def}}{=} g - g_1 - g_2$. Then it is easy to see that g_3 is continuous, has a continuous first derivative, and that g_3'' is piecewise continuous. ∎

Remark 9.4 The local error is not hard to calculate, but it will have to be handled separately for each of the functions g_1, g_2, and g_3.

9.1 The Smooth Case

Proposition 9.5 *Suppose g is in $C^{(2)}$ and that g and its first two derivitaves are exponentially bounded. Then*

$$\hat{\mathcal{E}}_{loc}(g) = \frac{2h^2}{3}\int_{-\infty}^{\infty} g''(x)p(x)\,dx + o(h^2). \tag{25}$$

If $g \in C^{(4)}$ and g''' and $g^{(iv)}$ are exponentially bounded, the error is $O(h^4)$.

Proof. We will calculate the right hand side of (21). Let I_k be the interval $[2kh, (2k+2)h]$ and let $y_k = (2k+1)h$ be its midpoint. Write $\int_{-\infty}^{\infty}(\Pi_e g(x) - g(x))p(x)\,dx = \sum_k \int_{I_k}(\Pi_e g(x) - g(x))p(x)\,dx$. Expand g around y_k: $g(x) = g(y_k) + g'(y_k)(x - y_k) + \frac{1}{2}g''(y_k)(x-y_k)^2 + o(h^2)$. Notice that on I_k, $\Pi_e g(x) - g(x) = \frac{1}{2}g''(y_k)(h^2 - (x-y_k)^2) + o(h^2)$. (Indeed, $\Pi_e g = g$ if g is a linear function of x, while $g \mapsto \Pi_e g$ is linear, so that the first order terms drop out. The rest follows since $x \mapsto \Pi_e g(x)$ is linear on I_k and equals g at the endpoints.) Write $p(x) = p(y_k) + O(h)$ on I_k; $\int_{I_k}(\Pi_e g(x) - g(x))p(x)\,dx = \frac{2}{3}g''(y_k)p(y_k)(h^3 + o(h^3))$. Summing, we get

$$\int_{-\infty}^{\infty}(\Pi_e g(x) - g(x))p(x)\,dx = \sum_k \frac{1}{3}g''(y_k)p(y_k)(h^2 + o(h^2))2h\,.$$

This is a Riemann sum for the integral $(h^2/3)\int g''(x)p(x)\,dx$. (One has to be slightly careful here: the $o(h^2)$ term is uniform, so it doesn't cause trouble in the improper integral. There is an $o(1)$ error in approximating the integral by the sum, but as it multiplies the coefficent of h^2, the error is $o(h^2)$ in any case. If $g \in C^{(4)}$, it is $O(h^4)$.) Thus

$$= \frac{h^2}{3}\int g''(x)p(x)\,dx \; + o(h^2)\,. \tag{26}$$

The second contribution to the error in (21) is $\frac{h^2}{3}\sum_k \Delta_k\, p(2kh) + O(h^3)$, where Δ_k is the discontinuity of the derivative of $\Pi_e g$ at $x_k \stackrel{\text{def}}{=} 2kh$:

$$\begin{aligned}
\Delta_k &= (g(x_{k+1}) - 2g(x_k) + g(x_{k-1}))/2h \\
&= (1/2h)\int_0^{2h}(g'(x_k + x) - g'(x_k - x))\,dx \\
&= 2g''(x_k)(h + o(h))\,.
\end{aligned}$$

This gives $\frac{h^2}{3}\sum_k g''(x_k)p(x_k)(2h + o(h)) = \frac{h^2}{3}\int g''(x)p(x)\,dx + o(h^2)$. Add this to (26) to finish the proof. ∎

9.2 The Piecewise Linear Case

For $x \in \mathbb{R}$, let $\hat{\theta}(x) = \text{frac}(x/2h)$ be the fractional part of $x/2h$. Then $x = 2kh + 2\hat{\theta}(x)h$ for some integer k. Let $\Delta g'(x) = g'(x+) - g'(x-)$.

Proposition 9.6 *Suppose g is continuous and piecewise linear. Then*

$$\hat{\mathcal{E}}_{loc}(g) = h^2 \sum_y \Delta g'(y)\Big(\frac{1}{3} + 2\hat{\theta}(y)(1-\hat{\theta}(y))\Big)p(y) + O(h^3)\,. \tag{27}$$

PROOF. Write $g(x) = ax + b + \frac{1}{2}\sum_y \Delta g'(y)|x-y|$, which we can do by Lemma 9.3.

Now Π_e is a linear operator, and $\Pi_e f = f$ if f is affine, so it is enough to prove this for $g(x) = |x-y|$ for some fixed y. Let us compute the two terms in (21).

Let I_k be the interval $[2kh, (2k+2)h]$. If $y \in I_k$, $x \mapsto g(x)$ is linear on the semi-infinite intervals on both sides of I_k, so $\Pi_e g(x) = g(x)$ except on I_k. Now $x \mapsto \Pi_e g(x)$ is linear in I_k and equals $g(x)$ at the endpoints. As $g(2kh) = 2\hat{\theta}(y)h$ and $g((2k+2)h) = 2(1-\hat{\theta}(y))h$, we can write $\Pi_e g - g$ explicitly. Let $p(x) = p(y) + O(h)$ for $x \in I_k$. Then the first term in (21) is

$$\begin{aligned}\int_{-\infty}^{\infty} (\Pi_e g(x) - g(x))p(x)\,dx &= (p(y)+O(h))\int_{I_k} (\Pi_e g(x) - g(x))\,dx \\ &= 4h^2\,\hat{\theta}(y)(1-\hat{\theta}(y))p(y) + O(h^3)\,. \end{aligned} \tag{28}$$

Turning to the final term in (21), notice that as $\Pi_e g = g$ outside of I_k,

$$\sum_y \Delta(\Pi_e g)'(z)'(y) = \sum_z \Delta g'(z) = \Delta g'(y) = 2\,,$$

so the final term is just $2h^2/3$. Adding this to (28), we get (27). ∎

9.3 The Step Function Case

Proposition 9.7 *Suppose that g is a step function. Then*

$$\begin{aligned}\hat{\mathcal{E}}_{loc}(g) = h\sum_{y\notin h\mathbb{Z}} (2\hat{\theta}(y)-1)\Delta g(y)p(y) - \frac{h^2}{3\sigma^2 T}\sum_{y\in\mathbb{N}_e^h} y\,\Delta g(y)\,p(y) \\ + \frac{h^2}{6\sigma^2 T}\sum_{y\in\mathbb{N}_o^h} y\,\Delta g(y)\,p(y)\,. \end{aligned} \tag{29}$$

PROOF. By Lemma 9.3 we can write $g(x) = \sum_y \Delta g(y)\tilde{H}(x-y)$. By linearity, it is enough to consider the case where $g(x) = \tilde{H}(x-y)$. Once again, we compute the integrals in (21). Let $I_k = [2kh, (2k+2)h]$. If $y \in I_k$ and $0 < \hat{\theta}(y) < 1$, we note that $\Pi_e g(x) = 0$ if $x < 2kh$, $\Pi_e g(x) = 1$ if $x > 2(k+2)h$, and $\Pi_e g$ is linear in I_k. Write $p(x) = p(y) + O(h)$ on I_k and note that the only contribution to the integral comes from I_k:

$$
\begin{aligned}
\int_{-\infty}^{\infty} (\Pi_e g(x) - g(x)) p(x)\, dx &= \Big[\int_{2kh}^{(2k+2)h} \frac{(x-2kh)}{2h}\, dx - \int_{2(k+\hat\theta(y))h}^{(2k+2)h} dx\Big] (p(y) + O(h)) \\
&= (2\hat\theta(y) - 1) p(y) h + O(h^2)\,. \qquad (30)
\end{aligned}
$$

The second term in (21) is easily handled. We note that since $\tilde H'(x-y) = 0$ for all $x \neq y$, $\sum_y \Delta(\tilde H')(x-y) = 0$, so that by (21), that integral is

$$
\int_{-\infty}^{\infty} (\Pi_o \Pi_e g(x) - \Pi_e g(x)) p(x)\, dx = O(h^3)\,.
$$

The cases $\hat\theta(y) = \frac{1}{2}$ and $\hat\theta(y) = 0$ are special. In both cases we need to expand p up to a linear term, since the constant term cancels out. So write $p(x) = p(y) + p'(y)(x-y) + O(h^2)$, $x \in I_k$. If $g(x) = \tilde H(x-y)$, $y \in I_k$, then $\hat\theta(y) = \frac{1}{2}$, means $y = (2k+1)h$. Noting that the contribution from $p(y)$ vanishes, the first error term will be

$$
\begin{aligned}
\int_{-\infty}^{\infty} (\Pi_e g(x) - g(x)) p(x)\, dx &= p'(y) \left[\int_{2kh}^{(2k+2)h} \frac{(x-2kh)^2}{2h}\, dx - \int_{2(k+\hat\theta(y))h}^{(2k+2)h} dx\right] \\
&= -\frac{1}{6} h^2\, p'(y) \qquad (31) \\
&= \frac{h^2 y}{6\sigma^2 T} p(y)\,,
\end{aligned}
$$

where we have used the fact that $p'(x) = -\frac{x}{\sigma^2 T} p(x)$.

If $\hat\theta = 0$, then $g = \tilde H(x)$, so $g(2kh) = \frac{1}{2}$. Thus $\Pi_e g(x) = (x - (2k-2)h)/4h$ if $(2k-2)h < x < (2k+2)h$. It is zero for $x \leq (2k-2)h$ and one for $x > (2k+2)h$, so that

$$
\begin{aligned}
\int_{-\infty}^{\infty} (\Pi_e g(x) - g(x)) p(x)\, dx = p(2kh) \Big[\int_{(2k-2)h}^{(2k+2)h} \frac{x-(2k-2)h}{4h} - \int_{2kh}^{(2k+2)h} dx\Big] \\
+ p'(2kh) \Big[\int_{(2k-2)h}^{(2k+2)h} \frac{(x-(2k-2)h)^2}{4h} - \int_{2kh}^{(2k+2)h} (x-2kh)\, dx\Big]\,. \qquad (32)
\end{aligned}
$$

The first term in square brackets vanishes, so this is

$$
= \frac{1}{3} p'(2kh) h^2 + O(h^3) = -\frac{2kh}{3\sigma^2 T} p(2kh) + O(h^3)\,. \qquad (33)
$$

The proposition follows upon adding (30), (31), and (33). ∎

10 Convergence of the Delta: Proof of Cor. 4.4

The price of our derivative at time $t < T$ is

$$V(s,t) = e^{r(T-t)} E\{f(S_T) \mid S_t = s\}\,. \tag{34}$$

The hedging strategy depends on the space derivative $\frac{\partial V}{\partial s}$, which is called the *delta*. It is of interest to know how well the tree scheme estimates this. From (34)

$$\frac{\partial V}{\partial s}(s,t) = e^{-r(T-t)} E\{f'(S_T) \mid S_t = s\} = \lim_{h\to 0} \frac{V(e^h s,t) - V(e^{-h}s,t)}{s(e^h - e^{-h})}$$

If $t = k\delta$ and $s = e^{jh+rt}$, we approximate $\partial V/\partial s$ by the symmetric discrete derivative

$$\frac{u(j+1,k) - u(j-1,k)}{s(e^h - e^{-h})}\,. \tag{35}$$

where u is the solution of the tree scheme (3).

Remark 10.1 Estimating the delta is essentially equivalent to running the scheme on f', not f. If f' is continuous, the result follows from Theorem 4.2. However, if f' is discontinuous—as it is for a call or a put—and if the discontinuity falls on a non-lattice point, Theorem 4.2 would give order 1/2, not order 1, which does not imply Corollary 4.4. In fact it depends on some uniform bounds which come from Theorem 4.2 and the fact we use the symmetric estimate of the derivative. Thus there is something to prove.

Proof. By the Markov property, it is enough to prove the result for $t = 0$ and $S_0 = 1$. We will also assume that $r = 0$ to simplify notation.

The key remark is that if S_t is a logarithmic Brownian motion from s, then $e^h S_t$ and $e^{-h} S_t$ are logarithmic Brownian motions from $e^h s$ and $e^{-h} s$ respectively, so that

$$\frac{\partial V}{\partial s}(1,0) = \lim_{h\to 0} \frac{B(e^h s, 0) - V(e^{-h}s, 0)}{e^h - e^{-h}} = \lim_{h\to 0} E\{\hat f(S_T, h)\}\,,$$

where

$$\hat f(s,h) = \frac{f(e^h s) - f(e^{-h}s)}{e^h - e^{-h}}\,.$$

Now $f' \in \mathcal{K}$ so that f and its first three derivatives are polynomially bounded, hence there is a polynomial $Q(s)$ which bounds $\hat f$, $\partial \hat f/\partial s$ and $\partial^2 \hat f/\partial s^2$, *uniformly for* $h < 1$. This will justify passages to the limit, so that, for instance, $\frac{\partial V}{\partial s}(1,0) = E\{S_T f'(S_T)\}$.

$$\begin{aligned}
\mathcal{E}(h) &\stackrel{\text{def}}{=} E\{\hat f(S_{\tau_n}, h) - S_T f'(S_T)\} \\
&= E\{\hat f(S_{\tau_n}, h) - \hat f(S_T, h)\} + E\{\hat f(S_T, h) - S_T f'(S_T)\} \\
&\stackrel{\text{def}}{=} \mathcal{E}_1(h) + \mathcal{E}_2(h)\,.
\end{aligned}$$

Now $f \in \mathcal{K}$, hence so is $\hat{f}(\cdot, h)$. Thus $\mathcal{E}_1(h)$ is the error for a payoff function $\hat{f}(\cdot, h)$, and Theorem 4.2 applies. By the uniform polynomial bound of $\hat{f}$ and related functions, these coefficients are uniformly bounded in h for $h < 1$, and we can conclude that there is a constant A such that $\mathcal{E}_1(h) \leq Ah^2$ for small h.

The bound on $\mathcal{E}_2(h)$ is straight analysis. We can write

$$\mathcal{E}_2(h) = \int_{-\infty}^{\infty} (\hat{f}(s,h) - sf'(s))p(s)\, ds\,,$$

where p is the density of S_T. Now $\hat{f}(s,h) - sf'(s) = \frac{1}{e^h - e^{-h}} \int_{se^{-h}}^{se^h} (f'(u) - f'(s))\, du$. If $f \in C^{(2)}$ on the interval, expand f' to first order in a Taylor series and integrate to see this is $s^2h^2 f''(s) + o(h^2)$. In any case, if $|f''| \leq C$ on the interval, it is bounded by Cs^2h. There are only finitely many points where $f \notin C^{(3)}$, each contributes at most Cs^2h^2 to $\mathcal{E}_2$ we see that $|\mathcal{E}_2(h)| \leq Bh^2$ for some other constant B.

To prove (9), let $f = (s-K)^+$ and evaluate $\mathcal{E}_1(h)$ by Theorem 4.2. Note that $\hat{f}$ will have discontinuities of approximately $1/2h$ and $-1/2h$ at $s = k - h$ and $s = K + h$ respectively. Note also that $\theta(\log s)$ (see section 3) is periodic with period $2h$, so that $\theta(e^h K) = \theta(e^{-h}K) \stackrel{\text{def}}{=} \check{\theta}$, so that we can write (8) in the form

$$\mathcal{E}_2(h) = h^2 \Big[C + D\check{\theta}(1-\check{\theta}) \frac{\hat{p}(\log(K-h)) - \hat{p}(\log(K+}{2h} \Big]\,.$$

the ratio converges to $-\hat{p}'(logK)$ and (9) follows. This completes the proof, except to remark that $\check{\theta}$ corresponds to $\hat{\theta}$, for the *odd* h instead of even multiples of h. ∎

11 Appendix

11.1 Moments of τ_n and J

The (very complicated!) coefficients in Theorem 4.2 come from moments of τ_n and J. We will derive them here. We will assume that P is the Brownian measure, i.e. that X_t is a Brownian motion. Thus we will not write E^Q and P^Q to indicate that we are using the Brownian measure. We can write $X_t = \sigma W_t$, where $\{W_t,\ t \geq 0\}$ is a standard Brownian motion.

Proposition 11.1 *(i) τ_1 has the same distribution as $\frac{T}{n}\nu$, where $\nu = \inf\{t > 0 : |W_t| = 1\}$, so it has the moment generating function*

$$F_1(\lambda) \stackrel{def}{=} E\{e^{\lambda \tau_1}\} = \Big(\cos\sqrt{2\lambda \frac{T}{n}}\ \Big)^{-1}, \ -\infty < \lambda < \frac{n\pi^2}{8T}\,. \qquad (36)$$

(ii) $\tau_1,\ \tau_2 - \tau_1,\ \tau_3 - \tau_2, \dots$ are i.i.d. , independent of $X_{\tau_1},\ X_{\tau_2}, \dots$.

(iii) $E\{\tau_1\} = \frac{T}{n}$, $\quad \mathrm{var}\{\tau_1\} = \frac{2T^2}{3n^2} \quad E\{\tau_n\} = T, \quad \mathrm{var}\{\tau_n\} = \frac{2T^2}{3n}$.

(iv) For each $k \geq 1$ there are constants $c_k > 0$, $C_k > 0$ such that

$$E\{\tau_1^k\} = \frac{c_k T^k}{n^k}, \quad E\{|\tau_n - T|^k\} \le C_k \frac{T^k}{n^{k/2}}\,.$$

PROOF. (*i*) follows by Brownian scaling, and the moment generating function is well-known for $\lambda < 0$ [4]; it is not difficult to extend to $\lambda > 0$. Then (*ii*) is well-known, see e.g. [9], and (*iii*) is an easy consequence of (*i*).

For (*iv*), notice that τ_1 has finite exponential moments, so that the moments in question are finite. The kth moment of τ_1 is determined by Brownian scaling: $c_k = E\{\nu^k\}$. To get the kth central moment, note that by (ii) we can write $\tau_n - T$ as a sum of n i.i.d. copies of $\tau_1 - T/n$, say $\tau_n - T = \eta_1, + \cdots + \eta_n$. The η_j have mean zero, so by Burkholder's and Hölder's inequalities in that order,

$$E\{(\tau_n - T)^k\} \le C_k E\Big\{\Big(\sum_{j=1}^n \eta_j^2\Big)^{\frac{k}{2}}\Big\} \le C_k n^{\frac{k}{2}} \sum_{j=1}^n E\{\eta_j^k\}\,.$$

But $E\{\eta_j^k\} = E\{(\tau_1 - T/n)^k\} = C_k' T^k/n^k$, which implies (iv). ∎

Proposition 11.2 *Suppose n is an even integer. Then*

(i) $E\{\tau_J\} = T + \frac{4}{3}\frac{T}{n} + O(h^3)$;

(ii) $E\{J\} = n + \frac{4}{3} + O(h)$;

(iii) $E\{(J-n)^2\} = \frac{2}{3}n + O(1)$.

(iv) For $k > 1$, there exists a constant C_k such that $E\{(J-n)^k\} \le C_k n^{k/2}$

PROOF. Set $\eta_j = T_j - T_{j-1} - \frac{T}{n}$, $j = 1,\ 2, \dots$ and put

$$M_j = \sum_{i=1}^j \eta_j = \tau_j - jE\{\tau_1\}\,.$$

Then (M_j) is a martingale. Apply the stopping theorem to the bounded stopping time $J \wedge N$ and let $N \to \infty$ to see that $0 = E\{M_J\} = E\{\tau_J\} - E\{J\}E\{\tau_1\}$, so that

$$E\{J\} = \frac{E\{\tau_J\}}{E\{\tau_1\}} = \frac{n}{T} E\{\tau_J\}\,. \tag{37}$$

Now $\tau_J > T$, so to find its expectation, notice that, as in Proposition 9.1, T_J will either be the first hit of $\mathbb{N}_e^h$ after T—if L is odd (see (18)—or it will be the first hit of $\mathbb{N}_e^h$ *after* the first hit of $\mathbb{N}_o^h$ after T, if L is even. The expected time for Brownian motion to reach the endpoints of an interval is well known: if $X_0 = x \in (a,b)$, the expected time for X to leave (a,b) is $\sigma^{-2}(x-a)(b-x)$. Let $\mathrm{dist}(x,A)$ be the shortest distance from x to the set A. If $X_T = x$, the expected additional time to reach $\mathbb{N}_e^h$ is $\sigma^{-2}(h^2 - \mathrm{dist}^2(x,\mathbb{N}_o^h))$, while the expectated additional time to reach $\mathbb{N}_o^h$ is $\sigma^{-2}(h^2 - \mathrm{dist}^2(x,\mathbb{N}_e^h))$. Once at $\mathbb{N}_o^h$, the expected time to reach $\mathbb{N}_e^h$ from there is $T/n = h^2/\sigma^2$. Now by (20), $P\{L \text{ is even} \mid X_T = x\} = q(x) = \mathrm{dist}\,(x,\mathbb{N}_o^h)/h + O(h)$, and $P\{L \text{ is odd} \mid X_T = x\} = \mathrm{dist}\,(x,\mathbb{N}_e^h)/h + O(h)$.

Thus, as L is conditionally independent of $\{X_{T+t},\ t \geq 0\}$ given X_T, we have $E\{\tau_J - T \mid X_T = x\} = P\{L \text{ is odd} \mid X_T = x\}E\{\tau_J - T \mid X_T = x, L \text{ is odd}\} + P\{L \text{ is even} \mid X_T = x\}E\{\tau_J - T \mid X_T = x, L \text{ is even}\}$, so that if $p(x)$ is the density of X_T,

$$E\{\tau_J - T\} = \int_{-\infty}^{\infty} \sigma^{-2} p(x)\Big(h^2 - \mathrm{dist}^2(x, \mathbb{N}_o^h)\Big)\Big(h^{-1}\mathrm{dist}\,(x, \mathbb{N}_e^h) + O(h)\Big)\,dx$$
$$+ \int_{-\infty}^{\infty} \sigma^{-2} p(x)\Big(2h^2 - \mathrm{dist}^2(x, \mathbb{N}_e^h)\Big)\Big(h^{-1}\mathrm{dist}\,(x, \mathbb{N}_o^h) + O(h)\Big)\,dx\,. \quad (38)$$

Now let $x_k = (2k+1)h$ and write $\int_{-\infty}^{\infty} = \sum_k \big[(1/2h)\int_{x_k-h}^{x_k+h}\big]2h$. Write $p(x) = p(x_k)(1 + O(h))$ on the interval $(x_k - h, x_k + h)$. We can then do the integrals explicitly:

$$\sum_k \frac{p(x_k)}{2h}\int_{x_k-h}^{x_k+h}(1 + O(h))\sigma^{-2}\Big(h^2 - \mathrm{dist}^2(x, \mathbb{N}_o^h)\Big)(\mathrm{dist}\,(x, \mathbb{N}_e^h)/h + O(h))\,dx$$
$$= \frac{5}{12}\sigma^{-2}h^2\sum_k p(x_k)\,2h + O(h^3) \sim \frac{5}{12}\frac{T}{n} \quad (39)$$

since the Riemann sum approximates $\int p(x)\,dx = 1$. The other integral is similar, and gives $\frac{11}{12}\frac{T}{n}$, so we see $E\{\tau_J - T\} = \frac{4}{3}\frac{T}{n} + O(h^3)$, which implies (i) and (ii).

To see (iii), note that $M_j^2 - j\,\mathrm{var}(\tau_1)$ is also a martingale. As with M, we can stop it at time J to see that that $E\{M_J^2\} = E\{J\}\mathrm{var}\{\tau_1\}$. Now $E\{M_J^2\} = E\{\tau_J^2\} - 2E\{J\tau_J\}E\{\tau_1\} + E\{\tau_1\}^2E\{J^2\}$.

But $E\{J\tau_J\} = TE\{J\} + E\{J(\tau_j - T)\}$, and J is $\mathcal{F}_T$-measurable, while the value of $E\{\tau_j - T \mid \mathcal{F}_T\}$ depends only on the value of X_T and on the parity of L, the index of the last stopping time before T. The joint distribution of X_T and the parity of L was determined in Section 9 by reversing X_t from T. It did not depend on J. In short, $E\{J\,\tau_J\} = E\{J\}E\{\tau_J\}$. Thus

$$E\{T_J^2\} - 2E\{J\}E\{\tau_J\}E\{\tau_1\} + E\{\tau_1\}E\{J^2\} = E\{J\}\mathrm{var}(\tau_1)\,.$$

We have found the values of all the quantities except $E\{J^2\}$ and $E\{(\tau_J - T)^2\}$, but we know the latter will be of the form $\gamma\frac{T^2}{n^2}$ for some $\gamma > 0$; this turns out to be negligeable, so we can solve for $E\{J^2\}$ to find that

$$E\{J^2\} = n^2 + \frac{10}{3}n + O(1)\,,$$

and (iii) follows.

To see (iv), notice that from (42), $P\big\{|J - n| > y\sqrt{n}\big\} \leq 4e^{-\frac{3y}{4}}$, so that

$$E\Big\{\Big(\frac{|J-n|}{\sqrt{n}}\Big)^k\Big\} \leq 4\int_{-\infty}^{\infty} y^k\,d(e^{-\frac{3y}{4}}) \stackrel{\mathrm{def}}{=} C_k\,,$$

which proves the assertion. ∎

We will need to control the tails of the distributions of τ_n and J. The following proposition is a key.

Proposition 11.3 *Let (ξ_n) be a sequence of reals. Suppose $m = n\xi_n$, where $\sqrt{n}\xi_n \to \infty$ as $n \to \infty$. Then, as $n \to \infty$,*

$$P\Big\{\sqrt{n}\,|\tau_m - \frac{m}{n}T| > \rho\Big\} \le 2e^{-\frac{3\rho^2}{4T^2\xi_n}}\Big(1 + O(\frac{1}{\xi_n\sqrt{n}})\Big). \tag{40}$$

PROOF. By Chebyshev's inequality and (36),

$$\begin{aligned} P_n \quad &\stackrel{\text{def}}{=} \quad P\left\{\sqrt{n}\Big(\tau_m - \frac{m}{n}T\Big) > \rho\right\} \;\le e^{-\lambda\rho}E\{e^{\lambda\sqrt{n}(\tau_m - \frac{m}{n}T)}\} \\ &= \quad e^{-\lambda\rho}\left(\frac{e^{-\frac{\lambda T}{\sqrt{n}}}}{\cos\sqrt{\frac{2\lambda T}{\sqrt{n}}}}\right)^{n\xi_n} = e^{-\lambda\rho}\left(\frac{e^{-\frac{x^2}{2}}}{\cos x}\right)^{\frac{4\lambda^2T^2\xi_n}{x^4}}. \end{aligned}$$

where $x = \sqrt{\frac{2\lambda T}{\sqrt{n}}}$. Take logs and choose $\lambda = \frac{3\rho}{2\xi_n T^2}$ to see that

$$\log P_n \le \frac{\rho^2}{\xi_n T^2}\Big(-\frac{3}{2} - \frac{9}{x^4}\Big(\frac{x^2}{2} + \log\cos x\Big)\Big).$$

Expand $\log\cos x$ near $x = 0$ and note that $x^2 = O(1/\xi_n\sqrt{n}) = o(1)$, so this is

$$= -\frac{3\rho^2}{4\xi_n T^2} + O(\frac{1}{\xi_n\sqrt{n}}). \tag{41}$$

The other direction is similar. For $\lambda > 0$, let

$$\begin{aligned} P_n \stackrel{\text{def}}{=} P\big\{\tau_m - \frac{m}{n}T < -\rho\big\} \quad &\le \quad e^{-\lambda\rho}E\big\{e^{-\lambda(\tau_m - \frac{m}{n}T)}\big\} \\ &= \quad e^{-\lambda\rho}\left(\frac{e^{\frac{\lambda T}{\sqrt{n}}}}{\cosh\sqrt{\frac{2\lambda T}{\sqrt{n}}}}\right)^{n\xi_n} \end{aligned}$$

which differs from the above only in that cosh replaces the cosine. Exactly the same manipulations show that P_n is again bounded by (41), and the conclusion follows. ∎

Corollary 11.4 *Let $y > 0$ and let g be exponentially bounded. Then*

$$P\Big\{\frac{|J - n|}{\sqrt{n}} \ge y\Big\} \le 2e^{-\frac{3}{4}\frac{y^2}{(1+y/\sqrt{n})}} \tag{42}$$

and for all strictly positive p and ϵ,

$$\lim_{n\to\infty} n^p E\big\{g(X_{\tau_n});\ |\tau_n - T| > n^{-\frac{1}{4}+\epsilon}\big\} = 0\,; \tag{43}$$

$$\lim_{n\to\infty} n^p E\big\{g(X_{\tau_J});\ |J - n| > n^{\frac{1}{2}+\epsilon}\big\} = 0\,. \tag{44}$$

Proof. Let $\xi > 1$. $P\{J > n\xi\} = P\{\tau_{n\xi} < T\} \le P\{\sqrt{n}\,|\tau_{n\xi} - \xi T| > \sqrt{n}\,(\xi - 1)T\}$. By (40), $P\{J - n > n(\xi - 1)\} \le e^{-\frac{3n(\xi-1)^2}{4\xi}}$. Take $y = (\xi - 1)\sqrt{n}$, to see that $P\{J - n > y\sqrt{n}\} \le 2e^{-\frac{3}{4}\frac{y^2}{(1+y/\sqrt{n})}}$. Similarly, for $\xi < 1$, $P\{J < n\xi\} = P\{\tau_{n\xi} > T\} \le P\{\sqrt{n}\,|\tau_{n\xi} - \xi T| > \sqrt{n}\,(1-\xi)T\}$. Use (40) to get the same bound for $P\{J - n < -y\sqrt{n}\}$, and add to get (42).

Next, X_{τ_n} and τ_n are independent and $|g(x)| \le Ae^{a|x|}$ for some A and a, so

$$|E\{g(X_{\tau_n});\ |\tau_n - T| > n^{-\frac{1}{4}+\epsilon}\}| \le AE\{e^{a|X_{\tau_n}|}\}P\{|\tau_n - T| > n^{-\frac{1}{4}+\epsilon}\}\,.$$

X_{τ_n} is binomial so we use its moment generating function to see that $E\{e^{a|X_{\tau_n}|}\} \le 2e^{\sigma\sqrt{nT}}$. Combine this with the bound (40) on the tails of τ_n to see (43).

The second assertion follows from Corollary 11.4, once we notice that in any case, $|X_{\tau_J} - X_T| \le 4h$, so that, $|E\{g(X_{\tau_J})\}| \le Ae^{a|X_{\tau_J}|} \le Ae^{a|X_T|+4h}$. ∎

11.2 Transition Probabilities

Let $P_n(x)$ be the transition probability of a simple symmetric random walk on the integers:

$$P_n(x) = \frac{n!\ 2^{-n}}{\left(\frac{n+x}{2}\right)!\ \left(\frac{n-x}{2}\right)!} \tag{45}$$

which is the probability of taking $\frac{1}{2}(n + x)$ positive steps and $\frac{1}{2}(n - x)$ negative steps out of n total steps. Now let

$$R(n, k, x) \stackrel{\text{def}}{=} \frac{P_{n+k}(x)}{P_n(x)}\,.$$

Define the *effective order* $\hat{O}$ of a monomial $\frac{k^p x^q}{n^r}$ to be $\hat{O}(\frac{k^p x^q}{n^r}) \stackrel{\text{def}}{=} \frac{1}{2}(p+q)-r$.

Proposition 11.5 *Let n, k, and x be even integers with* $\max(|k|, |x|) \le n^{\frac{3}{5}}$. *Then*

$$R(n, k, x) = 1 - \frac{k}{2n} + \frac{3k^2 + 4kx^2}{8n^2} - \frac{3k^2x^2}{4n^3} + \frac{k^2x^4}{8n^4} + Q_3 + O(n^{-3/2})\,,$$

where Q_3 is a sum of monomials of effective order at most $-\frac{3}{2}$.

Proof. We need Stirling's formula in the form $n! = \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}+O(n^{-3})}$. The $O(n^{-3})$ term is uniform in the sense that there is an a such that for all n it is between $-a/n^3$ and a/n^3.

Write $R(n,k,x)$ in terms of factorials and use Stirling's formula on each. We can write $R(n,k,x) = R_1 R_2$, where R_2 comes from the factors $e^{\frac{1}{12n}+O(n^{-3})}$. Let $\xi = k/n$, $\eta = x/n$, and $m = n/2$, and take logarithms. We find

$$\log R_1 = (2m(1+\xi)+\frac{1}{2})\log(1+\xi)+(m(1+\eta)+\frac{1}{2})\log(1+\eta)+(m(1-\eta+\frac{1}{2})\log(1-\eta)$$
$$-(m(1+\xi+\eta)+\frac{1}{2})\log(1+\xi+\eta)-(m(1+\xi-\eta)+\frac{1}{2})\log(1+\xi-\eta)$$

$$\log R_2 = \frac{1}{24m}\Big(\frac{1}{1+\xi}+\frac{2}{1+\eta}+\frac{2}{1-\eta}-1-\frac{2}{1+\xi+\eta}-\frac{2}{1+\xi-\eta}\Big).$$

Notice that errors in $\log R$ are of the same order as those of R; that is, if $r_n \to r > 0$ and $|\log r_n - \log r| < cn^{-p}$, then $|r_n - r| < 2rcn^{-p}$ for large n. Thus it is enough to determine $\log R$ up to terms of order $1/n$.

Since $|k|$ and $|x|$ are smaller than $n^{3/5}$, ξ and η are smaller than $n^{-2/5}$, and an easy calculation shows that

$$\log R_2 = \frac{\xi}{4n} + O(n^{-3/2}).$$

Expand $\log R_1$ in a power series in ξ and η. To see how many terms we need to keep, note that the coefficients may be $O(n)$. If we include terms up to order 6 in ξ and η, the remainder will be $o(n^{-3/2})$; those making an $O(1/n)$ or larger contribution will be of the form $\xi^p\eta^q$ with $p+q \le 2$, and $m\xi^p\eta^q$ with $p+q \le 5$, so that

$$\log R_1 = -\frac{1}{2}\xi + \frac{1}{4}\xi^2 + m\eta^2\xi - m\eta^2\xi^2 + m\eta^2\xi^3 + \frac{1}{2}m\eta^4\xi + \hat{S}(m,\xi,\eta) + o(n^{-\frac{3}{2}}),$$

where $\hat{S}(m,\xi,\eta)$ is a polynomial whose terms are all $o(1/n)$. Notice that $\log R_2$ is also of $o(1/n)$, so we can include it as part of $\hat{S}$. In terms of n, k, and x,

$$\begin{aligned}\log R &= -\frac{k}{2n} + \frac{k^2+2kx^2}{4n^2} - \frac{k^2x^2}{2n^3} + \frac{2k^3x^2+kx^4}{4n^4} + S(n,k,x) + o(n^{-\frac{3}{2}})\\ &\stackrel{\text{def}}{=} Q(n,k,x) + S(n,k,x) + o(n^{-\frac{3}{2}}).\end{aligned}$$

where $S(n,k,x)$ is a sum of monomials, each of which is $o(1/n)$. The largest term in Q is $h^2x^2/n^2 \le n^{-1/5}$, so that we can write

$$\begin{aligned}R &= e^{Q+S}(1+o(n^{-\frac{3}{2}})) = \left(\sum_{p=1}^{8}\frac{(Q+S)^p}{p!}\right)(1+o(n^{-1}))\\ &\stackrel{\text{def}}{=} (1+o(n^{-\frac{3}{2}}))\,Q_1.\end{aligned}$$

Check the effective order of the terms in the polynomial Q_1: we see that

$$\hat{O}\Big(\frac{k}{2n}\Big) = \hat{O}\Big(\frac{kx^2}{2n^2}\Big) = -\frac{1}{2}, \quad \hat{O}\Big(\frac{k^2}{4n^2}\Big) = \hat{O}\Big(\frac{k^2x^2}{n^3}\Big) = -1\,,$$

and the other two terms have effective order $-3/2$. All terms in S have effective order less than -1, and hence less than or equal to $-3/2$, since all effective orders are multiples of $1/2$. Now the effective order of the product of monomials is the sum of the effective orders, so that the only terms in Q_1 of effective order at least -1 are those in Q and the three terms from $\frac{1}{2}Q^2$ which come from the squares and products of the terms of effective order $-1/2$, namely

$$\frac{k^2}{8n^2} + \frac{k^2x^4}{8n^4} - \frac{k^2x^2}{4n^3}.$$

All other terms have effective orders at most $-3/2$. Thus define

$$Q_2 \stackrel{\text{def}}{=} 1 - \frac{k}{2n} + \frac{3k^2 + 4kx^2}{8n^2} - \frac{3k^2x^2}{4n^3} + \frac{k^2x^4}{8n^4}\,,$$

and let Q_3 be all the other terms in Q_1. Then $R = (Q_2+Q_3)(1+o(n^{-3/2}))$ where all terms in Q_3 have effective order at most $-3/2$. This proves the proposition. ∎

12 Summary and Translation

We have done all our work in terms of the function g, but we would like the final results in terms of the original function f. The translation is straightforward.

Let $\rho(x) = e^{-\frac{x}{2}-\frac{\sigma^2 T}{8}}$. Then $g(x) = f(s_0e^{x+rT})\rho(x)$. Moreover, $\hat{p}(x) = \rho(x)p(x)$, where p and $\hat{p}$ are the densities of X_T under the Brownian and martingale measures respectively. Moreover, the martingale measure P and the Brownian measure Q are connected by $dP = \rho(X_T)dQ$ on $\mathcal{F}_t$. Thus $E^Q\{g(X_T)\} = E^P\{f(S_T)\}$. We also have formulas involving the derivatives of g. Note first that $\rho'(x) = -\frac{1}{2}\rho(x)$. Let $s = s_0e^{x+rT}$, $S_T = s_0e^{X_T+rT}$, so $X_T = \log(\tilde{S}_T/s_0)$ and

$$\begin{aligned}
E^Q\{g(X_T)\} &= E^P\{f(S_T)\} \\
E^Q\{g''(X_T)\} &= E^P\Big\{S_T^2 f''(S_T) + \frac{1}{4}f(S_T)\Big\} \\
E^Q\{X_T^k g(X_T)\} &= E^P\{(\log(\tilde{S}_T/s_0))^k f(S_T)\}\,.
\end{aligned} \tag{46}$$

References

[1] Cox, J.C., S.A. Ross, and M. Rubinstein, Option pricing, a simplified approach, Journal of Financial Economics 7, (1979) 229–263.

[2] Diener, F. and M. Diener, Asymptotics of the price oscillations of a vanilla option in a tree model, (preprint)

[3] Diener, F. and M. Diener, Asymptotics of the price of a barrier option in a tree model (preprint)

[4] Karatzas, I. and S. Shreve, Brownian Motion and Stochastic Calculus, Springer-Verlag, 1988.

[5] Karatzas, I. and S. Shreve, Methods of Mathematical Finance, Springer-Verlag, 1998.

[6] Leisen, D.P. and M. Reimer, Binomial models for option valuation-examining and improving convergence, Applied Math. Finance 3, (1996) 319–346

[7] Leisen, D.P., Pricing the American put option, a detailed convergence analysis for binomial models. Journal of economic dynamics and control 22, (1998) 1419–1444.

[8] Monroe, Itrel, On embedding right-continuous martingales in Brownian motion, Ann. Math. Stat 43, 1293–1316.

[9] Rogers, L.C.G. and E.J. Stapleton, Fast accurate binomial pricing, Finance and Stochastics 2 (1998), 3–17.

[10] Walsh, Owen D. and John B., Embedding and the convergence of the binomial and trinomial tree schemes, Numerical Methods and Stochastics, Fields Communications Series 34, pp. 101-121 (in press).