SPECTRAL PROPERTIES OF TRINOMIAL TREES

ALEKSANDAR MIJATOVIĆ
IMPERIAL COLLEGE LONDON

Abstract. In this paper we prove that the probability kernel of a random walk on a trinomial tree converges to the density of a Brownian motion with drift at the rate $O(h^4)$, where h is the distance between the nodes of the tree. We also show that this convergence estimate is optimal in that the density of the random walk cannot converge at a faster rate. The proof is based on an application of spectral theory to the transition density of the random walk. This yields an integral representation of the discrete probability kernel that allows us to determine the convergence rate.

Key Words: trinomial trees, Brownian motion with drift, convergence estimates for probability densities, spectral theory

1. Introduction

Binomial models, introduced in (Cox, Ross & Rubinstein 1979), play a central role in the theory and practice of derivatives pricing. They are widely used to obtain algorithms for determining the numerical value of options with general payoffs and provide a flexible framework that can accommodate a variety of models (see for example (Hull & White 1988), (Madan, Milne & Shefrin 1991), (Rubinstein 1994) and (Derman, Kani & Chriss 1996)). The equivalence between binomial and trinomial trees has been established in (Rubinstein 2000). The present paper will focus on the convergence properties of the latter.

It is known that the prices obtained from binomial and trinomial models converge to the prices given by continuous-time and -space models (see e.g. (He 1990) and (Amin & Khanna 1994)). The question of the *rate* at which the discrete option price converges to its continuous limit has been examined in (Heston & Zhou 2000). The authors show that the convergence rate of a call option is at least of the order $O(1/\sqrt{n})$, where n is the number of time-steps in the binomial model. When expressed in terms of the lattice spacing h, this convergence rate is equivalent to $O(h)$ (the relationship between the number of steps n and the distance h between the state nodes is given by $n = \frac{3\sigma^2 T}{h^2}$, see equation (4) in section 2). In addition the authors show that for twice continuously differentiable functions on a bounded interval in $\mathbb{R}$ the prices converge at the rate $O(h^2)$. They propose to improve the convergence rate for non-differentiable payoffs by using smoothing strategies and applying the convergence result for smooth functions. This approach is well-suited to general vanilla payoffs since they are continuous and non-differentiable only at a single point. It is less applicable to functions characterized by a higher degree of irregularity such as the payoffs of European double digitals and butterfly spreads.

I would like to thank Prof. Eckhard Platen for useful comments.

The probability density function of a random process can be viewed, in terms of pricing theory, as a forward value of an option whose payoff is the Dirac delta function. The singular nature of this function has hindered efforts to address the question of the convergence rate of the probability density function of a discrete model to the probability density of a continuous model, which is an interesting problem both theoretically and in practice.

In this paper we address the question of the convergence of the probability density function of a random walk on a commonly used trinomial tree (see section 2 for definition) to its continuous counterpart, a Brownian motion with drift. The main result, contained in theorem 3.1, states that the probability density of the random walk on a trinomial tree converges to the density of the Brownian motion with drift at the rate $O(h^4)$, where the parameter h represents the distance between the nodes of the tree at any time-step. Furthermore we show that the convergence of the probability density functions is uniform in the state-space (i.e. the same rate applies to any pair of starting and terminal points in the state-space) and that it is of the order $O(h^4)$ and no faster. This result can be compared with the rate of convergence of the probability density of a continuous-time Markov chain to a Brownian motion with drift which has been shown to be of the order $O(h^2)$ (see (Albanese & Mijatović 2006)).

Related results dealing with the convergence rate of density functions of a sum of independent identically distributed random variables to the density of the limiting variable, known as local limit theorems, have been developed in mathematical statistics (see for example (Kolassa 2006) and chapter XVI in (Feller 1971)). The proofs are based on Edgeworth expansions which use Hermite polynomials to approximate the density of the sum. See also (Ibragimov & Linnik 1971) for applications of Fourier methods to the same problem.

The paper is organised as follows. In section 2 we give a definition of a widely used random walk on the trinomial tree using natural moment matching conditions and find the spectral representation of its probability kernel. Section 3 contains the statement and proof of our main result (theorem 3.1). Section 4 concludes the paper.

2. Probability kernel of the random walk

A general idea behind multinomial lattice models is to approximate a stochastic process on a continuous state-space by a random walk on a discrete state-space. A very natural question for such an approximation is that of the convergence rate of the probability kernel of the random walk to the probability density function of the initial stochastic process. In this section we shall first recall a well-known approximation of the Brownian motion with drift[1] by a discrete-time Markov chain (i.e. a random walk) on a trinomial tree. Once we have established the probability kernel of the chain, we will find its spectral representation as defined in appendix A. This representation will allow us to prove our main result, contained in theorem 3.1.

Let $Y_t := x + \mu t + \sigma W_t$ denote a Brownian motion with a real drift μ and a positive volatility σ, which started at some point x in $\mathbb{R}$. As usual W_t denotes the standard Brownian motion. Let

[1]This process is at the heart of the Black-Scholes model (see (Black & Scholes 1973)) since an exponential of it is a geometric Brownian motion.

us fix a time horizon T. The basic idea of a trinomial tree is to divide the time interval $[0, T]$ into n small time-steps Δt, i.e. $n = \frac{T}{\Delta t}$, and approximate Y_T with a sum

$$Y_T^h := x + \sum_{i=1}^{n} X_i, \tag{1}$$

where the random variables X_i, for $i = 1, \ldots, n$, are independent identically distributed with a domain $\{-h+\mu\Delta t, \mu\Delta t, h+\mu\Delta t\}$. At any fixed time-step $i \in \{1, \ldots, n\}$ the parameter h denotes the distance between the consecutive nodes in the tree. In other words the variable X_i is used to approximate the evolution of the process $\mu t + \sigma W_t$ in the short time interval of length Δt. The local geometry of the trinomial tree is described in figure 1.

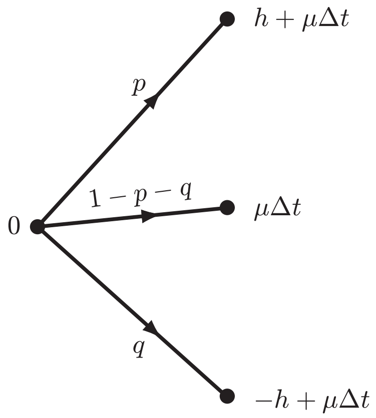

Figure 1. The local geometry of the trinomial tree. Each of the random variables X_i, for $i \in \{1, \ldots, n\}$, takes three possible values. The middle node is the mean, the other two are a distance h away from the mean. It is clear that the size of h is directly related to the variance of X_i. We will see that if the spacing h and the time-step Δt are related by the formula in (4) and the probabilities p and q are equal to $\frac{1}{6}$, then the first five moments of X_i coincide with those of the normal distribution $N(\mu\Delta t, \sigma^2\Delta t)$.

It is intuitively clear that the convergence rate of the probability kernel of Y_T^h to the probability density of the process Y_T improves as we satisfy the moment matching conditions

$$\mathbb{E}_t\left[(Y_{t+\Delta t} - Y_t - \mu\Delta t)^k\right] = \mathbb{E}[(X_i - \mathbb{E}[X_i])^k], \quad \text{where} \quad t = i\frac{\Delta t}{T} \quad \text{for} \quad i \in \{1, \ldots, n\}, \tag{2}$$

for more integers k. The probabilities p and q from figure 1 can be chosen in such a way that the first five moments of the random variable X_i coincide with the first five moments of the normal distribution $N(\mu\Delta t, \sigma^2\Delta t)$ of the increment $Y := Y_{t+\Delta t} - Y_t$. Recall that the moments of the normal random variable $Y - \mu\Delta t$ are of the form

$$\mathbb{E}_t[(Y - \mu\Delta t)^k] = \begin{cases} \frac{(2n)!}{2^n n!}(\sigma^2\Delta t)^{2n} & \text{if } k = 2n, \\ 0 & \text{if } k = 2n-1, \end{cases} \tag{3}$$

for any k in $\mathbb{N}$. Satisfying condition (2) for all odd integers k reduces to the equation $p = q$, which guarantees that the variables X_i are symmetric around their mean.

The moment matching conditions in (2) for k equals 2 and 4, together with formula (3) yield the following system of equations: $2ph^2 = \sigma^2\Delta t$ and $2ph^4 = 3(\sigma^2\Delta t)^2$. This system uniquely determines the local geometry of our trinomial tree as it implies that the relationship between the time-step Δt and the spacing h is given by the following key formula:

$$h^2 = 3\sigma^2 \Delta t. \tag{4}$$

We also find that the probabilities p and q in figure 1 have to be equal to $\frac{1}{6}$. The local geometry of our trinomial tree is now completely determined.

Assuming that we have fixed the spacing h and the corresponding time-step Δt, we can define a discrete-time Markov chain Y_t^h as in (1), for every time t which is an integer multiple of Δt. It is clear that the domain of the chain Y_t^h, at time t, is a subset of $\mu t + h\mathbb{Z}$ if and only if the starting point x of the random walk Y_t^h is an element of $h\mathbb{Z}$ (see figure 2).

Let us now assume that x and y are arbitrary elements in $h\mathbb{Z}$. The Markov chain Y_t^h defined above is time homogeneous[2] and its evolution is therefore determined uniquely by the bounded operator $\mathbf{P}_{\Delta t}^h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$, which can be expressed using the orthonormal basis[3] $(\delta_y)_{y \in h\mathbb{Z}}$ as follows

$$\mathbf{P}_{\Delta t}^h(\delta_y)(x) := \mathbb{P}(Y_{\Delta t}^h = \mu\Delta t + y | Y_0^h = x). \tag{5}$$

To simplify notation we introduce the coordinate expression $\mathbf{P}_{\Delta t}^h(x,y) := \mathbf{P}_{\Delta t}^h(\delta_y)(x)$. In other words definition (5) provides a way of expressing the transition probability kernel of the Markov chain Y_t^h as a bounded linear operator $\mathbf{P}_{\Delta t}^h$ on the Hilbert space $l^2(h\mathbb{Z})$. Furthermore, since the transition probabilities $\mathbb{P}(Y_{\Delta t}^h = \mu\Delta t + y | Y_0^h = x)$ have been uniquely determined by the local geometry of the trinomial tree, we obtain the following coordinate expression for our operator:

$$\mathbf{P}_{\Delta t}^h(x,y) = \frac{1}{6}\begin{cases} 1 & \text{if } |y - x| = h, \\ 4 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Recall that the *discrete Laplace operator* $\Delta_h : l^2(h\mathbb{Z}) \to l^2(h\mathbb{Z})$ can be defined in the following way for every sequence f in $l^2(h\mathbb{Z})$:

$$\Delta_h f(x) := \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}, \quad \text{where } x \in h\mathbb{Z}.$$

A simple yet important observation, based on relation (4), is that the operator $\mathbf{P}_{\Delta t}^h$ can be expressed as a linear combination of the identity operator I on the Hilbert space $l^2(h\mathbb{Z})$ and the discrete Laplace operator Δ_h in the following way:

$$\mathbf{P}_{\Delta t}^h = I + \frac{\sigma^2 \Delta t}{2}\Delta_h. \tag{7}$$

[2]By definition this means that the following equality holds $\mathbb{P}(Y_{t+\Delta t}^h = \mu(t+\Delta t) + y | Y_t^h = \mu t + x) = \mathbb{P}(Y_{\Delta t}^h = \mu\Delta t + y | Y_0^h = x)$ for all elements $x, y \in h\mathbb{Z}$. In our case this is obvious since all summands in definition (1) are identically distributed and the left-hand side equals $\mathbb{P}(X_i = \mu\Delta t + y - x)$, for some index i, while the right-hand side equals $\mathbb{P}(X_1 = \mu\Delta t + y - x)$.

[3] The sequence δ_y in $l^2(h\mathbb{Z})$ takes value 1 at y and value 0 everywhere else. It is clear that such sequences for all $y \in h\mathbb{Z}$ form a basis of the Hilbert space $l^2(h\mathbb{Z})$ with the inner product given by (22) in appendix A. The measure μ in that definition assigns a weight of 1 to each element in $h\mathbb{Z}$.

The easiest way to show that (7) holds is by expressing the discrete Laplace operator in coordinates, substituting it into the formula on the right-hand side, applying equation (4) and noticing that the result matches coordinate expression (6) for the probability kernel $\mathbf{P}^h_{\Delta t}$. Equality (7) will play a central role in obtaining the spectral representation of the probability kernel of the random walk Y^h_t, which will be established in proposition 2.1 and applied in the proof of theorem 3.1.

So far we have concerned ourselves with the evolution of the chain Y^h_t over the short time period Δt. Our next task is to understand the probability kernel

$$p^h_T(x,y) := \mathbb{P}(Y^h_T = \mu T + y | Y^h_0 = x), \quad \text{where} \quad x, y \in h\mathbb{Z}. \tag{8}$$

The global geometry of the trinomial tree, for fixed values of the drift μ and volatility σ, is shown in figure 2.

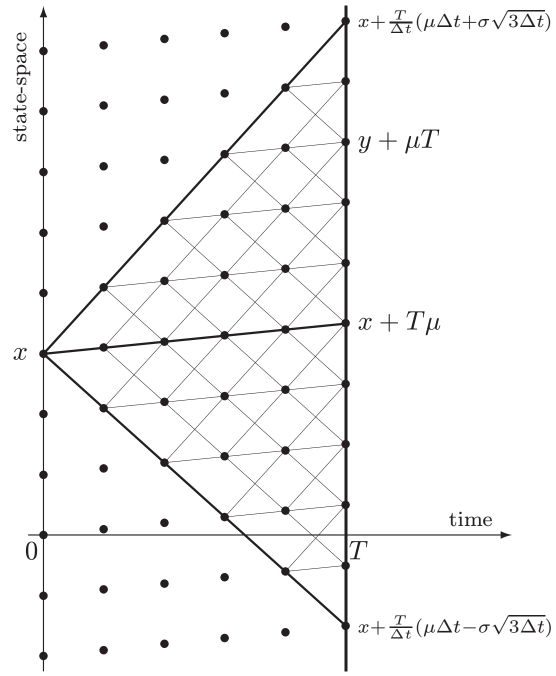

Figure 2. The global geometry of the trinomial tree. The figure corresponds to a drift μ of 0.1 and volatility sigma σ of 0.57. For convenience h (resp. Δt) is chosen to equal one unit of length (resp. time). Notice that the random walk Y^h_t, which started at $x \in h\mathbb{Z}$, can only reach with non-zero probability (at time T) states in the interval $[x + n(\mu\Delta t - h), x + n(\mu\Delta t + h)]$. The integer n equals $\frac{T}{\Delta t}$. In section 3 we shall investigate the convergence properties of the probability kernel $p^h_T(x,y)$, defined in (8), for an arbitrary pair of points $x, y \in h\mathbb{Z}$ as the spacing h goes to zero. It is clear from this figure that in the limit the set of reachable states in the tree sweeps out the entire real line.

We now need to express probability kernel (8) in terms of the operator $\mathbf{P}^h_{\Delta t}$ and find its spectral representation. This is done in the following proposition.

Proposition 2.1. *Let Y^h_t be the Markov chain defined above and let $p^h_T(x,y)$ be its probability kernel given by (8). Let T be a time horizon and Δt a time-step. Then the following holds:*

(a) *The transition probability* $p^h_T(x,y)$ *equals* $(\mathbf{P}^h_{\Delta t})^n(\delta_y)(x)$*, where* δ_y *is the element of the orthonormal basis of* $l^2(h\mathbb{Z})$ *that corresponds to the singleton* $y \in h\mathbb{Z}$ *(see equation (5)) and the number of time-steps* n *is* $\frac{T}{\Delta t}$*.*

(b) *The spectral representation of the probability kernel* $p^h_T(x,y)$ *is given by the integral*

$$p^h_T(x,y) = \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \left(\frac{2}{3}+\frac{1}{3}\cos(hp)\right)^{\frac{3\sigma^2 T}{h^2}} e^{ip(x-y)}dp,$$

for any pair of elements x, y *in* $h\mathbb{Z}$*.*

Before proceeding to the proof of proposition 2.1, we should note that the above representation for the probability kernel $p^h_T(x,y)$ of the random walk Y^h_t has the property that the time parameter t and the space parameters x, y feature independently. That is, the kernel of the integral in proposition 2.1 (b) is a product of two functions, one depending on time T and the other on the dislocation $(x-y)$. This feature of the spectral representation will play a crucial role in the proof of theorem 3.1.

Proof. Part (a) of the proposition follows from the representation in (5), which specifies the evolution of our chain over the time-step Δt, and an iterative application of the Chapman-Kolmogorov equations (see chapter 6 in (Grimmett & Stirzaker 2001)).

In order to prove part (b), we need to find a spectral representation of the operator $(\mathbf{P}^h_{\Delta t})^n$ in the sense of appendix A. This may be achieved in the following way: find a diagonal representation for the discrete Laplace operator. The corresponding unitary transformations in this representation (see appendix A for the role of unitary transformations in the definition of the spectral representation) can also be used to define a (trivial) representation for the identity operator I. Thus by (7) we obtain the spectral representation of $\mathbf{P}^h_{\Delta t}$.

The spectral representation, as defined in appendix A, for the discrete Laplace operator will be given by a unitary transformation $\mathcal{F}_h : l^2(h\mathbb{Z}) \to L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$, also known as the *semidiscrete Fourier transform*, defined by

$$\mathcal{F}_h(f)(p) := \sqrt{\frac{h}{2\pi}}\sum_{m\in\mathbb{Z}} f(hm)e^{-imhp},$$

where an arbitrary element f of the Hilbert space $l^2(h\mathbb{Z})$ is viewed as a function from $h\mathbb{Z}$ to $\mathbb{C}$.

The usual inner product on $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ with respect to the Lebesgue measure on the domain $[-\frac{\pi}{h}, \frac{\pi}{h}]$ (see appendix A for definition) makes the family of functions $p \mapsto \sqrt{\frac{h}{2\pi}}e^{-imhp}$, where $m \in \mathbb{Z}$, into an orthonormal basis of $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ (for the proof of this fundamental fact see theorem II.9 in (Reed & Simon 1980)). In particular this implies that the mapping $\mathcal{F}_h^{-1}$: $L^2([-\frac{\pi}{h}, \frac{\pi}{h}]) \to l^2(h\mathbb{Z})$ given by

$$\mathcal{F}_h^{-1}(\phi)(hm) := \sqrt{\frac{h}{2\pi}}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} \phi(p)e^{imhp}dp$$

is well-defined and is an inverse of the semidiscrete Fourier transform (both of these facts follow directly from the definition of the inner product on $L^2([-\frac{\pi}{h}, \frac{\pi}{h}])$ and the fact that the functions $p \mapsto \sqrt{\frac{h}{2\pi}}e^{-imhp}$ form an orthonormal basis). It follows from the definitions of the respective inner products on Hilbert spaces $l^2(h\mathbb{Z})$ and $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ that $\mathcal{F}_h$ is a unitary transformation (for definition see appendix A). Since $\mathcal{F}_h^{-1}$ is an inverse of $\mathcal{F}_h$, it must also be a unitary operator.

The following calculation in the Hilbert space $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ yields a spectral representation of the discrete Laplace operator Δ_h:

$$\begin{aligned}
\mathcal{F}_h\Delta_h(f)(p) &= \sum_{m\in\mathbb{Z}} \frac{f(h(m+1))+f(h(m-1))-2f(hm)}{h^2}\sqrt{\frac{h}{2\pi}}e^{-imhp} \\
&= \sqrt{\frac{h}{2\pi}}\sum_{m\in\mathbb{Z}} f(hm)e^{-imhp}\frac{e^{ihp}+e^{-ihp}-2}{h^2} \\
&= \frac{2(\cos(hp)-1)}{h^2}\mathcal{F}_h(f)(p).
\end{aligned}$$

All infinite sums that feature in this calculation are well-defined elements of $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ since the coefficients are the elements of the convergent sequence $f\in l^2(h\mathbb{Z})$.

This calculation, together with equality (7), implies that the operator $(\mathbf{P}^h_{\Delta t})^n$ has a spectral representation of the form

$$\mathcal{F}_h(\mathbf{P}^h_{\Delta t})^n\mathcal{F}_h^{-1}(\phi)(p) = \left(1+\sigma^2\Delta t\frac{\cos(hp)-1}{h^2}\right)^n\phi(p), \tag{9}$$

where ϕ is any element of $L^2([-\frac{\pi}{h},\frac{\pi}{h}])$ and n equals the number of time-steps $\frac{T}{\Delta t}$ in the trinomial tree.

We are now in the position to obtain the spectral representation of the probability kernel $p^h_T(x,y)$ of the random walk Y^h_t at a given time horizon T, by combining equality (9) with part (a) of the proposition. By substituting ϕ with $\mathcal{F}_h(\delta_y)(p)=\sqrt{\frac{h}{2\pi}}e^{-ipy}$ in (9), we find the following:

$$\begin{aligned}
p^h_T(x,y) &= (\mathbf{P}^h_{\Delta t})^n(\delta_y)(x) = \mathcal{F}_h^{-1}\left(\left(1+\sigma^2\Delta t\frac{\cos(hp)-1}{h^2}\right)^n\sqrt{\frac{h}{2\pi}}e^{-ipy}\right)(x) \\
&= \frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{\frac{\pi}{h}}\left(1+\sigma^2\Delta t\frac{\cos(hp)-1}{h^2}\right)^{\frac{3\sigma^2 T}{h^2}} e^{ip(x-y)}dp.
\end{aligned}$$

The proof of part (b) in the proposition can now be concluded by substituting equality (4) into the formula we have just obtained. □

3. The main theorem

In this section we shall examine the convergence of the discrete probability kernel $p^h_T(x,y)$ to the conditional probability density function $p_T(x,y)$ of the Brownian motion with drift. We will establish the precise convergence rate of $p^h_T(x,y)$ to $p_T(x,y)$ and prove that it is uniform in the state variables x and y (see theorem 3.1). Before stating this result we should recall that the probability density function $p_T(x,y)=\mathbb{P}(Y_T=\mu T+y|Y_0=x)$ of the Brownian motion with drift $Y_t=x+\mu t+\sigma W_t$ at a given time horizon T has a spectral representation of the form:

$$p_T(x,y)=\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{p^2}{2}\sigma^2 T}e^{ip(x-y)}dp. \tag{10}$$

For the proof of this standard fact see for example section 4 in (Albanese & Mijatović 2006).

Before proceeding to the statement of theorem 3.1, let us clarify the limiting procedure as the parameter h goes to zero, which involves a passage from a discrete state-space $h\mathbb{Z}$ to the continuum of $\mathbb{R}$. First of all we need to fix a strictly decreasing sequence of positive real numbers $(h_n)_{n\in\mathbb{N}}$ with the following two properties:

$$\lim_{n\to\infty} h_n = 0 \quad \text{and} \quad \frac{h_i}{h_j} \in \mathbb{N} \ \text{ for all } \ j \geq i.$$

The first property enables us to study the behaviour of transition probabilities as the lattice spacing goes to zero. The second property ensures that the lattice $h_j\mathbb{Z}$ contains the lattice $h_i\mathbb{Z}$ for all $j \geq i$. It is clear that there are many sequences satisfying these requirements. A very simple example is given by $h_n := (\frac{1}{2})^n$.

Let us now fix a sequence $(h_n)_{n\in\mathbb{N}}$ which satisfies the above conditions. In all that follows we shall assume that the distance h between two consecutive points in the lattice is equal to one of the elements in the sequence $(h_n)_{n\in\mathbb{N}}$. A limit when the spacing h goes to zero is defined as the limit where the parameter h visits all the elements of the sequence $(h_n)_{n\in\mathbb{N}}$ from some index $N \in \mathbb{N}$ onwards. It is clear that the limit as h goes to zero yields the same result for any choice of sequence $(h_n)_{n\in\mathbb{N}}$ which satisfies the above conditions and is therefore a well-defined concept. We may now state our main theorem.

Theorem 3.1. *Let a positive real number T be a time horizon. Let p_T^h be the probability kernel of the random walk Y_t^h, given by (8) in section 2. For any pair of elements $x, y \in h\mathbb{Z}$ let $p_T(x,y)$ denote the conditional probability density function of the Brownian motion with drift, given in (10). Then the following holds*

$$p_T(x,y) = \frac{1}{h}p_T^h(x,y) + O(h^4)$$

and the error term $O(h^4)$ is independent of x and y. Equivalently this means that there exist positive constants C and δ such that the inequality $|p_T(x,y) - \frac{1}{h}p_T^h(x,y)| \leq Ch^4$ holds for all $h < \delta$ and all $x, y \in h\mathbb{Z}$. Furthermore the probability kernel $\frac{1}{h}p_T^h(x,y)$ does NOT converge to the density $p_T(x,y)$ at the rate $O(h^4 f(h))$, for any non-constant function f such that $\lim_{h\to 0} f(h) = 0$.

In this theorem we are comparing the probability density functions of the family of the discrete state-space processes Y_t^h, with the density of Y_t whose state-space is clearly continuous. Since the conditional probability density of the Brownian motion with drift (starting at $x \in h\mathbb{Z}$), can be defined by the limit $p_T(x,y) = \lim_{h\to 0}\frac{1}{h}\mathbb{P}(y - h < Y_T - \mu T \leq y | Y_0 = x)$, it is clear that the density of the discrete state-space process Y_t^h should be given by the quotient $\frac{1}{h}p_T^h(x,y)$ which is equal to $\frac{1}{h}\mathbb{P}(y-h < Y_T^h - \mu T \leq y | Y_0^h = x)$. Moreover this argument shows that the "discrete" density $\frac{1}{h}p_T^h(x,y)$ converges to the "continuous" density $p_T(x,y)$.

Before proceeding to the proof of theorem 3.1 recall that by definition, a function $k(h)$ is of type $O(g(h))$ if and only if it is bounded above by $Mg(h)$, for a positive constant M, on some interval $[0, \epsilon)$ where $\epsilon > 0$, i.e. $\limsup_{h\searrow 0}\left|\frac{k(h)}{g(h)}\right| \leq M$.

Proof. Let us start by choosing an arbitrary starting point $x \in h\mathbb{Z}$ and an arbitrary terminal point $y \in h\mathbb{Z}$ (see figure 2). Our goal is to find an upper bound on the difference

$$D(h) := |p_T(x, y) - \frac{1}{h} p_T^h(x, y)|$$

of the densities for Brownian motion with drift Y_t and the discrete-time Markov chain Y_t^h at a given time horizon T. This will be achieved by comparing the spectral representation for $p_T(x, y)$, given by (10), with the representation of the discrete probability kernel $p_T^h(x, y)$, specified in (b) of proposition 2.1.

The basic idea that allows this comparison is as follows. We will start by defining a positive function $K(h)$ which goes to infinity "very slowly" as the distance h between the nodes of the tree approaches 0. This will enable us to decompose the integration domains of the spectral representations mentioned above into disjoint sets $[-K(h), K(h)]$ and $\mathbb{R} - [-K(h), K(h)]$ and study the difference $D(h)$ by expressing it as Riemann integrals over the components of this decomposition. This will yield the desired upper bound. A similar strategy was used in the proof of theorem 5.1 in (Albanese & Mijatović 2006) (for more on this comparison see figure 3).

Before specifying the function $K(h)$ let us define the function $f_h(p)$, which contains all the essential information about the spectrum of the operator $\mathbf{P}_{\Delta t}^h$ from (5) (see figure 3), using the formula

$$f_h(p) := \frac{3}{h^2} \left(\log(2 + \cos(hp)) - \log(3)\right). \tag{11}$$

Using the function $f_h(p)$ we can rewrite the representation given in proposition 2.1 (b) of the probability kernel $p_T^h(x, y)$ in the following way

$$p_T^h(x, y) = \frac{h}{2\pi} \int_{-\frac{\pi}{h}}^{\frac{\pi}{h}} e^{f_h(p)\sigma^2 T} e^{ip(x-y)} dp. \tag{12}$$

Let $K(h) := \frac{1}{\sigma\sqrt{T}} \sqrt{10 \log(1/h)}$ be the positive function which we will use to decompose the domain of integration. Note that since the inequality $K(h) \leq \frac{\pi}{h}$ holds for small h (i.e. the interval $[-K(h), K(h)]$ is contained in the domain $[-\frac{\pi}{h}, \frac{\pi}{h}]$ for small h), we can bound the difference $D(h)$ using the integrals

$$\begin{aligned}
A_1(h) &:= \int_{-K(h)}^{K(h)} \left| e^{f_h(p)\sigma^2 T} - e^{-\frac{p^2}{2}\sigma^2 T} \right| dp, \\
A_2(h) &:= \int_{K(h)}^{\frac{\pi}{h}} e^{f_h(p)\sigma^2 T} dp, \\
A_3(h) &:= \int_{K(h)}^{\infty} e^{-\frac{p^2}{2}\sigma^2 T} dp
\end{aligned}$$

in the following way:

$$2\pi D(h) \leq A_1(h) + 2(A_2(h) + A_3(h)). \tag{13}$$

The factor of 2 in front of $A_2(h)$ and $A_3(h)$ arises because the integrands in $D(h)$ are symmetric functions on the components of the set $\mathbb{R} - [-K(h), K(h)]$. Note that in obtaining (13) we used the well-known inequality $|\int_U f(p)dp| \leq \int_U |f(p)|dp$ for any integrable function f and any Borel measurable set U in $\mathbb{R}$. Observe also that the right-hand side of (13) is independent of

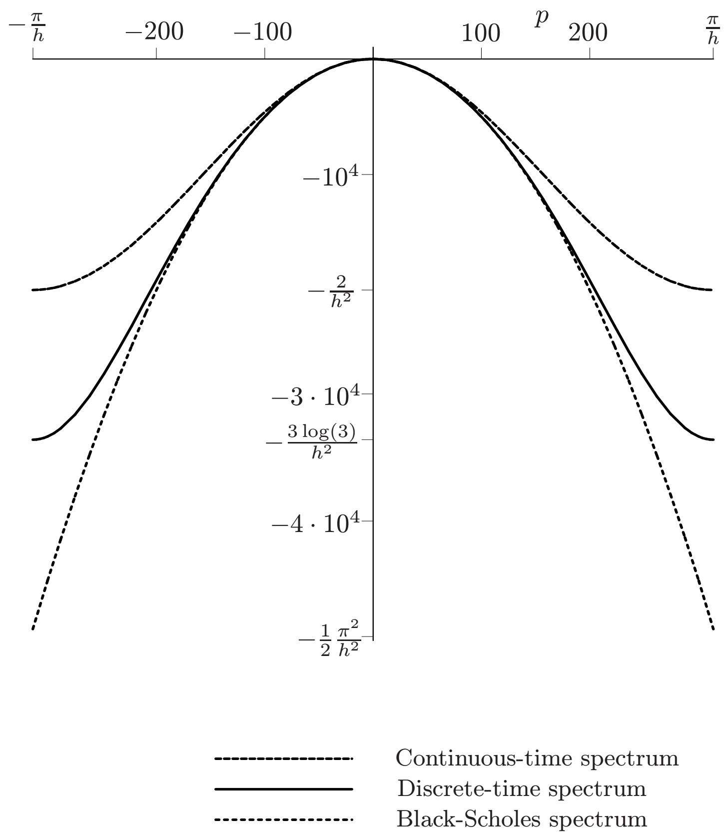

FIGURE 3. Why does the discrete-time Markov chain converge to the Brownian motion two orders of magnitude faster than the continuous-time Markov chain? The answer lies in this figure. The spectrum of the Laplace operator $\mathcal{L} = \frac{1}{2}\Delta$ (i.e. the Black-Scholes spectrum $p \mapsto -\frac{p^2}{2}$), which is the generator for Brownian motion, is approximated in a much better way by the discrete-time spectrum $f_h(p)$ defined in (11) than by the spectrum of the generator $\mathcal{L}_h = \frac{1}{2}\Delta_h$ for the continuous-time Markov chain given by the function $p \mapsto \frac{\cos(hp)-1}{h^2}$. Using the backward Kolmogorov equation, these spectral representations can be transformed to their respective probability densities by applying functional calculus: (i) formula (10) for the PDF of the Brownian motion is obtained from $e^{T\mathcal{L}} = \mathcal{F}^{-1}e^{-\frac{p^2}{2}\sigma^2 T}\mathcal{F}$ where $\mathcal{F}$ denotes the Fourier transform; (ii) proposition 2.1 implies that the probability density of the discrete-time process is given by $\mathcal{F}_h^{-1}e^{f_h(p)\sigma^2 T}\mathcal{F}_h$; (iii) the density of the continuous-time Markov chain is specified by $e^{T\mathcal{L}_h} = \mathcal{F}_h^{-1}e^{\frac{\cos(hp)-1}{h^2}\sigma^2 T}\mathcal{F}_h$ (see sections 2 and 4 in (Albanese & Mijatović 2006) for a rigorous treatment of the spectral properties of the generators $\mathcal{L}$ and $\mathcal{L}_h$ and the respective probability kernels). These spectral representations of the probability densities give an intuitive explanation for the difference in convergence rates. For simplicity in this figure the volatility σ is normalized to one and time to maturity T is equal to one year. We plot the spectral representation functions for the discrete- and continuous-time Markov chains for $h = 100^{-1}$. Notice that the value of the function $K(h)$, defined at the beginning of the proof of theorem 3.1, at $h = 100^{-1}$ is less than 6.8. It is obvious that on the interval $[-6.8, 6.8]$ the spectrum of the trinomial tree virtually coincides with the spectrum of the Brownian motion. This observation offers some justification to our strategy for the proof of theorem 3.1, since it is intuitively clear from the figure that the difference of the two spectral representations remains negligible on the interval $[-K(h), K(h)]$ for all small values of h and that on the complement $\mathbb{R} - [-K(h), K(h)]$ the spectra are mapped by the exponential to extremely small positive values.

the coordinates x and y since they only appear in the expression $e^{ip(x-y)}$ whose absolute value equals one.

Our task now is to prove the inequalities $A_j(h) \leq A_j^0 h^4$ for some positive constants A_j^0, where the index j runs from 1 to 3. Let us start by estimating the integral $A_1(h)$ for small values of the spacing h:

$$\begin{aligned} A_1(h) &\leq \int_{-K(h)}^{K(h)} e^{-\frac{p^2}{2}\sigma^2 T} \left| e^{\left(f_h(p)+\frac{p^2}{2}\right)\sigma^2 T} - 1 \right| dp \\ &\leq \int_{-K(h)}^{K(h)} e^{-\frac{p^2}{2}\sigma^2 T} \left(e^{\left|f_h(p)+\frac{p^2}{2}\right|\sigma^2 T} - 1 \right) dp. \end{aligned} \tag{14}$$

The second inequality follows from the elementary relationship $|e^z - 1| \leq e^{|z|} - 1$ which is valid for all z in the complex plane (take the absolute value of the Taylor series for the function $z \mapsto e^z - 1$ and apply to it the triangle inequality: $|e^z - 1| = |\sum_{n=1}^\infty z^n| \leq \sum_{n=1}^\infty |z|^n = e^{|z|} - 1$). It is clear from (14) that in order to bound $A_1(h)$ we first need to understand the asymptotic behaviour of the expression $|f_h(p) + \frac{p^2}{2}|$. The next claim constitutes a central step in the proof of the theorem.

Claim. The following inequality holds

$$\left| f_h(p) + \frac{p^2}{2} \right| \leq C h^4 p^6, \tag{15}$$

for some positive constant C and all p in the interval $[-K(h), K(h)]$.

To see this we need to recall that the function $x \mapsto \log(x)$ has a Taylor expansion of the form

$$\log(x) = \log(3) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{3^n n}(x-3)^n,$$

which converges uniformly on compact subsets of the interval $(0, 6)$. Using this fact we can express the function $f_h(p)$ as an infinite series

$$f_h(p) = -\frac{3}{h^2} \sum_{n=1}^{\infty} \frac{1}{3^n n}(1 - \cos(hp))^n,$$

since $|\cos(hp) - 1|$ is always strictly smaller than 3. This expression leads to the following inequality

$$\left| f_h(p) + \frac{p^2}{2} \right| \leq \left| \frac{p^2}{2} - \frac{1-\cos(hp)}{h^2} - \frac{(1-\cos(hp))^2}{6h^2} \right| + \frac{3}{h^2} \sum_{n=3}^{\infty} \frac{1}{3^n n} |1 - \cos(hp)|^n. \tag{16}$$

In order to obtain a bound for the infinite series in (16) we should recall that $\lim_{h \to 0} hK(h)^m = 0$ for any $m \in \mathbb{N}$, which implies that $hp < 1$ for $p \in [-K(h), K(h)]$ and all small h. Therefore we find that $|1 - \cos(hp)| \leq \frac{h^2p^2}{2}$, since the error of a partial sum of an alternating series whose elements form a monotonically decreasing sequence is bounded above by the absolute value of the first element in the series after the partial sum. This implies the following inequalities

$$\frac{3}{h^2} \sum_{n=3}^{\infty} \frac{1}{3^n n} |1 - \cos(hp)|^n \;\leq\; 3h^4p^6 \sum_{n=0}^{\infty} \left(\frac{h^2p^2}{6} \right)^n \;\leq\; 6h^4p^6. \tag{17}$$

The last inequality in (17) is valid for small values of the spacing h, which guarantees that the geometric series is bounded above 2.

We are now left with the task of estimating the first summand on the right-hand side of (16), i.e. $G(h) := \left|\frac{p^2}{2} - \frac{1-\cos(hp)}{h^2} - \frac{(1-\cos(hp))^2}{6h^2}\right|$. To that end we recall the following decomposition of the cosine:

$$1 - \cos(hp) = \frac{h^2p^2}{2} + (hp)^4 S_h(p) \quad \text{where} \quad S_h(p) := \sum_{n=2}^{\infty} (-1)^{n-1}\frac{(hp)^{2n-4}}{(2n)!}.$$

Substituting this expression into the definition of $G(h)$ we find

$$\begin{aligned} G(h) &= \left|\frac{h^2p^4}{4!} + h^4p^6 \sum_{n=3}^{\infty}(-1)^{n-1}\frac{(hp)^{2n-6}}{(2n)!} - \frac{1}{6h^2}\left(\frac{h^4p^4}{4} + (hp)^6 S_h(p) + (hp)^8 S_h(p)^2\right)\right| \\ &\leq h^4p^6\left(\sum_{n=3}^{\infty}\frac{|hp|^{2n-6}}{(2n)!} + \frac{1}{6}\sum_{n=2}^{\infty}\frac{|hp|^{2n-4}}{(2n)!} + h^2p^2\left(\sum_{n=2}^{\infty}\frac{|hp|^{2n-4}}{(2n)!}\right)^2\right) \\ &\leq G_0 h^4 p^6, \end{aligned} \tag{18}$$

for some constant G_0 and all small h. Inequality (18) holds because the convergent geometric series $\sum_{n=0}^{\infty}(hp)^{2n}$, whose value is below 2 for small h, dominates each of the infinite series above. Substituting inequalities (17) and (18) into (16) and defining $C := G_0 + 6$ proves the claim.

We can now estimate the integral $A_1(h)$. The elementary fact $e^x - 1 \leq 2x$, which is valid for small non-negative x, tells us that, for small values of h, (14) implies the following estimate

$$A_1(h) \leq h^4 2\sigma^2 TC \int_{-K(h)}^{K(h)} p^6 e^{-\frac{p^2}{2}\sigma^2 T} dp \leq A_1^0 h^4, \tag{19}$$

for some positive constant A_1^0.

It is clear from figure 3 that the function $f_h(p)$ is decreasing on the interval $[0, \frac{\pi}{h}]$. The quantity $A_2(h)$ is therefore bounded above by the expression $\frac{\pi}{h}e^{f_h(K(h))\sigma^2 T}$. Recall that by definition $K(h)$ was set to $\frac{1}{\sigma\sqrt{T}}\sqrt{10\log(1/h)}$. Since our claim implies that the following holds $f_h(p) \leq \left|f_h(p) + \frac{p^2}{2}\right| - \frac{p^2}{2} \leq Ch^4p^6 - \frac{p^2}{2}$ for $p \in [-K(h), K(h)]$, we can conclude that

$$A_2(h) \leq \frac{\pi}{h} e^{C\sigma^2 T h^4 K(h)^6} e^{-\frac{K(h)^2}{2}\sigma^2 T} \leq \frac{\pi}{h} 2h^5 = 2\pi h^4. \tag{20}$$

The second inequality in (20) follows directly from the definition of $K(h)$ and the property $\lim_{h\to 0} h^4 K(h)^6 = 0$.

We are now left with the easy task of estimating the decay rate for $A_3(h)$. A straightforward application of L'Hospital's rule implies that $A_3(h)$ is bounded above by $e^{-\frac{K(h)^2}{2}\sigma^2 T}$. By substituting the definition of $K(h)$ we conclude that

$$A_3(h) \leq h^5. \tag{21}$$

Inequality (13), together with (19), (20) and (21), now yields the necessary convergence estimate and thus proves the first part of the theorem.

We are now left with the task of showing that the rate of convergence cannot be faster than $O(h^4)$. We will do this by assuming that the difference $D(h)$ converges to zero at the rate $O(h^4 f(h))$ for some nonconstant function $f(h)$, such that $\lim_{x\to 0} f(h) = 0$, and then showing that this assumption leads to contradiction.

Note that if we redefine $K(h) := \frac{1}{\sigma\sqrt{T}}\sqrt{12\log(1/h)}$, the same calculation as above implies that the integrals $A_2(h)$ and $A_3(h)$ converge to zero at least at the rate $O(h^5)$. Under our current assumptions this implies that the integral

$$A_1(h) = 2\int_0^{K(h)} \left| e^{\left(f_h(p)+\frac{p^2}{2}\right)\sigma^2 T} - 1 \right| e^{-\frac{p^2}{2}\sigma^2 T} dp$$

tends to zero at the rate of $O(h^4 g(h))$, where $g(h) = \max\{f(h), h\}$, since it can be bounded above by a linear combination of $D(h)$, $A_2(h)$ and $A_3(h)$. On the other hand it is not hard to see that the following holds

$$f_h(p) + \frac{p^2}{2} = \frac{h^4 p^6}{1080} + h^6 p^8 k(hp),$$

where $k : (-\epsilon, \epsilon) \to \mathbb{R}$ is an infinitely differentiable function defined on an interval containing zero (i.e. $\epsilon > 0$) and has the key property $\lim_{x\to 0} k(x) = 0$. The elementary inequality $e^x - 1 \geq x$, valid for all real numbers x, implies the following

$$\frac{1}{h^4}\left(e^{\left(f_h(p)+\frac{p^2}{2}\right)\sigma^2 T} - 1\right) \geq \sigma^2 T\left(\frac{p^6}{1080} + h^2 p^8 k(hp)\right) \geq \sigma^2 T \frac{p^6}{2\cdot 1080}$$

for all small h and $p \in [0, K(h)]$. The following calculation yields a contradiction

$$\frac{1}{h^4 g(h)} A_1(h) \geq \frac{\sigma^2 T}{g(h)} \int_0^{K(h)} \frac{p^6}{2\cdot 1080} e^{-\frac{p^2}{2}\sigma^2 T} dp,$$

since the last integral clearly converges to a finite positive value, while the limit $\lim_{h\to 0} g(h)$ equals 0. This concludes the proof of the theorem. □

4. Conclusion

In this paper we examined the convergence rate of the probability density function of a random walk on a trinomial tree to the density of a Brownian motion with drift. We found that the rate of convergence is precisely of the order $O(h^4)$ and no faster, the parameter h being the distance between the neighbouring nodes of the tree at each time-step.

From the point of view of pricing theory, this result can be interpreted in terms of the convergence of the prices of Arrow-Debreu securities. An *Arrow-Debreu security* is an option whose payoff at some maturity T is the Dirac delta function concentrated at a given point of the domain of the underlying process. It is clear that the probability density function can be viewed as a non-discounted value of such an option and that theorem 3.1 can therefore be viewed as a convergence result for securities with a very singular payoff. This should be contrasted with the well-known fact (see for example (Heston & Zhou 2000)) that the smoothness of the payoff function is crucial for proving the convergence estimates for discretization schemes such as trinomial trees and PDE methods.

Appendix A. Spectral representations for bounded linear operators

The general references for this section are (Reed & Simon 1980) and (Huston & Pym 1980). See also appendices in (Albanese & Mijatović 2006). Let us start with some basic definitions.

Definition. Let $\mathcal{H}$ be a vector space equipped with an inner product, $\langle\cdot,\cdot\rangle_{\mathcal{H}} : \mathcal{H}\times\mathcal{H}\to\mathbb{C}$, which gives rise to the norm on $\mathcal{H}$ defined as $\|x\| := \sqrt{\langle x,x\rangle_{\mathcal{H}}}$, for every $x\in\mathcal{H}$. The vector space $\mathcal{H}$ is a *Hilbert space* if every Cauchy sequence in $\mathcal{H}$, with respect to this norm, has a limit in $\mathcal{H}$.

In other words this definition says that a vector space is a Hilbert space if and only if it is complete with respect to the norm induced by the inner product. An example of a Hilbert space, denoted by $L^2(\mathbb{R},\mu)$, is the set of equivalence classes of measurable functions $f:\mathbb{R}\to\mathbb{C}$ such that

$$\int_{\mathbb{R}} f(x)\overline{f(x)}d\mu(x) < \infty,$$

where the integral is a Lebesgue integral over the real line with respect to the positive measure μ (for the proof of this fundamental fact see for example section 2.5 in (Huston & Pym 1980)). If the measure μ is concentrated on a discrete set M in $\mathbb{R}$ then, since the set M must be countable (a discrete subspace of a second countable topological space cannot have cardinality larger than the set of natural numbers $\mathbb{N}$) we can represent the above integral condition as

$$\sum_{m\in M} f(m)\overline{f(m)}\mu(m) < \infty. \tag{22}$$

The Hilbert space of all sequences $(f(m))_{m\in M}$, satisfying this condition, is in this case denoted by $l^2(M,\mu)$ (or simply by $l^2(M)$, if it is clear what the measure μ is).

We will now introduce another basic concept which is of use to us in sections 2 and 3. A subset S of the Hilbert space $\mathcal{H}$ is called *orthonormal* if the following holds $\langle x,y\rangle_{\mathcal{H}} = \delta_{xy}$ for all $x,y\in S$. Here the function δ_{xy} denotes the familiar Kronecker delta. A subset S is an *orthonormal basis* for $\mathcal{H}$ if it is not contained, as a proper subset, in any other orthonormal set in $\mathcal{H}$.

The reason orthonormal bases are important is because they allow us to express any element $y\in\mathcal{H}$ as sum of a convergent series

$$y = \sum_{x\in S}\langle y,x\rangle_{\mathcal{H}} y. \tag{23}$$

See theorem II.6 in (Reed & Simon 1980) for the proof of this fundamental fact. It should be noted that the convergence of the series in (23) is in the topology induced by the inner product on $\mathcal{H}$. This point is of particular relevance if $\mathcal{H}$ is a function space, because in that case it would be tempting to conclude that the series converges, say, point wise, which is not necessarily true.

A function $A:\mathcal{N}\to\mathcal{M}$, mapping a Hilbert space $\mathcal{N}$ into a Hilbert space $\mathcal{M}$, is called a *linear operator* if it is additive and invariant under scalar multiplication. If, in addition, the operator A preserves the inner product (i.e. $\langle Ax,Ay\rangle_{\mathcal{M}} = \langle x,y\rangle_{\mathcal{N}}$ for all $x,y\in\mathcal{N}$), then A is called a *unitary operator.*

Definition. Let $\mathcal{N}$ and $\mathcal{M}$ be Hilbert spaces. A linear operator $A:\mathcal{N}\to\mathcal{M}$ is *bounded* if the inequality $\|Ax\|_{\mathcal{M}}\le C\|x\|_{\mathcal{N}}$ holds for all $x\in\mathcal{N}$ and some positive real constant C.

Bounded operators come in ample supply. For example it follows directly form the definition that any unitary operator is a bounded linear operator. It should also be noted that, if the

Hilbert space $\mathcal{N}$ is finite-dimensional, all linear operators on $\mathcal{N}$ are necessarily bounded and therefore continuous.

Definition. Let A be a linear operator on a Hilbert space $\mathcal{H}$. The *resolvent set* of A is the set of all complex numbers λ, such that the inverse $(A-\lambda I_{\mathcal{H}})^{-1}$ exists and is a bounded operator on $\mathcal{H}$ (the operator $I_{\mathcal{H}}$ is the identity on $\mathcal{H}$). The *spectrum* of A, denoted by $\sigma(A)$, is the complement (in $\mathbb{C}$) of the resolvent set.

If the Hilbert space $\mathcal{H}$ is finite-dimensional, then the spectrum of A consists solely of the set of eigenvalues of A. If, on the other hand, $\mathcal{H}$ is infinite-dimensional, then the spectrum $\sigma(A)$ is no longer necessarily discrete but is still a closed subset of the complex plane. In the case of a unitary operator it is not difficult to see directly from the definition that every element of the spectrum must have modulus equal to 1. An effective way of understanding any operator A on a Hilbert space $\mathcal{H}$ is to understand its spectrum $\sigma(A)$. In order to achieve the latter one usually resorts to some sort of spectral representation.

Definition. Let A be a linear operator on a Hilbert space $\mathcal{H}$. The *spectral representation* of A consists of a pair (M, μ), where M is a measure space and μ is a positive measure on M, together with a unitary operator $U : \mathcal{H} \to L^2(M, \mu)$ and a measurable function $F_A : M \to \mathbb{C}$ such that the following holds

$$(UAU^{-1}f)(x) = F_A(x)f(x)$$

for all functions f in the Hilbert space $L^2(M, \mu)$.

It follows from this definition that the spectral representation of an operator A, defined on a finite-dimensional space, is equivalent to diagonalizing a matrix that represents the operator A in some basis. If we assume that the matrix for A can be diagonalized and that its eigenvalues are all distinct, then the ingredients of the spectral decomposition can be described easily as follows: the space M consists of a discrete set of eigenvalues (i.e. $M = \sigma(A)$), μ is a positive measure which assigns a non-zero weight to every point in M and the function F_A is a natural inclusion of M into $\mathbb{C}$. The eigenvector corresponding to an element λ of M is simply the indicator function of the set $\{\lambda\} \subset M$.

It comes as no surprise that the spectral representation of an operator, as described in the previous definition, might not exist, since many matrices are not equivalent to diagonal transformations. Nevertheless in spectral theory, much like in linear algebra, there are sufficient conditions which guarantee that a bounded linear operator possesses a spectral representation (see chapters VII and VIII in (Reed & Simon 1980) and chapters 9 and 10 in (Huston & Pym 1980)). We make no use of these important results in this paper, because we are able to find an explicit spectral representation of the operator $\mathbf{P}^h_{\Delta t}$ given in (5) of section 2, which defines the probability kernel of our random walk.

REFERENCES

Albanese, C. & A. Mijatović (2006), 'Convergence rates for diffusions on continuous-time lattices', *preprint* .
URL: *http://www.imperial.ac.uk/people/a.mijatovic*

Amin, K. & A. Khanna (1994), 'Convergence of American option values from discrete to continuous-time financial models', *Math. Finance* **4**, 289–304.

Black, F. & M. Scholes (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Cox, J.C., S.A. Ross & M. Rubinstein (1979), 'Option pricing: A simple approach', *Journal of Financial Economics* **7**, 229–263.

Derman, E., G. Kani & N. Chriss (1996), 'Implied trinomial trees of the volatility smile', *Goldman Sachs Quantitative Strategies Research Notes* .

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn, John Wiley & Sons.

Grimmett, G. & D. Stirzaker (2001), *Probability and Random Processes*, 3rd edn, Oxford University Press.

He, H. (1990), 'Convergence from discrete to continuous time contingent claims prices', *Rev. Financial Stud.* **3**, 523–546.

Heston, S. & G. Zhou (2000), 'On the rate of convergence of discrete-time contingent claims', *Mathematical Finance* **10**(1), 53–75.

Hull, J. & A. White (1988), 'The use of the control variate technique in option pricing', *J. Financial Quant. Anal.* **23**, 237–251.

Huston, V. & J. S. Pym (1980), *Applications of Functional Analysis and Operator Theory*, Academic Press.

Ibragimov, I.A. & Y.V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff.

Kolassa, J.E. (2006), *Series Approximation Methods in Statistics*, number 88 *in* 'Lecture notes in statistics', Springer, New York.

Madan, D., F. Milne & H. Shefrin (1991), 'The multinomial option pricing model and its Brownian and Poisson limits', *Review of Financial Studies* **2**, 251–265.

Reed, M. & B. Simon (1980), *Methods of Modern Mathematical Physics 1: Functional Analysis*, Academic Press.

Rubinstein, M. (1994), 'Implied binomial trees', *Journal of Finance* **49**, 771–818.

Rubinstein, M. (2000), 'On the relation between binomial and trinomial option pricing models', *working paper* .

ALEKSANDAR MIJATOVIĆ, INSTITUTE FOR MATHEMATICAL SCIENCES, IMPERIAL COLLEGE LONDON

E-mail address: `a.mijatovic@imperial.ac.uk`