Convergence results for Tree methods in finance

Claude Martini

March 16, 1999

Premia 22

Contents

This follows Introduction to Tree methods in finance

1 Donsker theorem

1.1 Theorem statement

Theorem 1. *Let $(\xi_n)_{n\geq 1}$ a sequence of i.i.d. random variables with $E[\xi_n] = 0$ and $E[\xi_n^2] = \sigma^2$. Let $S_n = \sum_{k=0}^{n} \xi_k$. Let $h > 0$. Then the family of continuous processes*

$$X_t(h) = \frac{\sqrt{h}}{\sigma}\left(S_{n_t} + (t - n_t h)\,\xi_{n_t+1}\right)$$

where n_t is the integer such that $n_t h \leq t < (n_t + 1)\,h$, converges in law on $C(\mathbb{R}_+, \mathbb{R})$ to the standard Brownian motion as $h \to 0$.

1.2 Applying the theorem

1.2.1 Standard options

As a direct apllication this gives the convergence of the European Put prices in the Random Walk tree. The application to other trees is not straightforward since-take the CRR tree for instance-the distribution itself of the sequence $(\xi_n)_{n\geq 1}$ depends on h : either a more elaborate version of the theorem is needed, or some work is required-like a measure change within the tree-to fit the hypotheses, in this case the functional at hand will certainly be modified and will eventually depend on h so that the situation is more involved: in a few words, a direct application of this theorem is seldom feasible.

1.2.2 Path-dependent options

Of course the interesting feature of this theorem is its ability to deal with path-dependant functionals which are continuous and bounded for the uniform topology on $C(\mathbb{R}_+, \mathbb{R})$. To what kind of schemes may this theorem be applied?

A first idea is barrier options pricing. Indeed the payoff of the option, after a Girsanov transform and a suitable change of variables, may be expressed in terms of the terminal values of a Brownian motion B_T and its running maximum $\sup_{0\leq t\leq T}(B_t)$. Let us take for instance the quantity

$$E\left[f(B_T)\,1\left(\sup_{0\leq t\leq T}(B_t)<l\right)\right] \tag{1}$$

with $l>0$ and f continuous and bounded. Unfortunately in case of barrier options the supremum comes into play through an expression like the latter, $1\left(\sup_{0\leq t\leq T}B_t<l\right)$, which is not a continuous function over $C(\mathbb{R}_+,\mathbb{R})$: clearly two trajectories may be uniformly close to each other, the one breaching the barrier and not the other. At first glance the basic property of weak convergence on $C(\mathbb{R}_+,\mathbb{R})$ is not enough.

Nevertheless remember that for any measurable set A of $C(\mathbb{R}_+,\mathbb{R})$ such that $P(\partial A)=0$, ie a $P-$continuity set, then $P_n(A)\to P(A)$ if $P_n\to P$ in law. We are exactly in this situation here with

$$A=\left\{\omega,\ \sup_{0\leq t\leq T}\omega_t<l\right\}$$

It is easily seen that $\partial A=A\setminus intA\,\{\omega\in A,\omega_s=l\,for\ some\ s\}$. Now the law of $\sup_{0\leq t\leq T}B_t$ is absolutely continuous with respect to the lebesgue measure, so it doesn't weight points, this entails $P(\partial A)=0$ with P the law of B.

Observe that it is easy to build a three-dimensional tree which approximates the dynamic of the pair $\left(B_t,\sup_{0\leq s\leq t}B_s\right)$: first draw in the $z=0$ plane the standard Random Walk approximation tree of B, which corresponds to Bernoulli random variables in Donsker theorem. Then draw above each node $(n,X_{nh}(h))$ a line and sample along this line the possible values of $\sup_{0\leq k\leq n}X_{kh}(h)$. It is seen that there are at most $N+1$ possible values of $M_{Nh}(h)=\sup_{0\leq k\leq N}X_{kh}(h)$, so that there is no combinatoric blow up. The two sons of a point (n,x,M) (with weigths $\frac{1}{2}$) are seen to be $\left(n+1,x+\sqrt{h},\max\left(M,x+\sqrt{h}\right)\right)$ and $\left(n+1,x-\sqrt{h},M\right)$.

The algorithm to compute (1) by backward induction follows, the convergence is given by Donsker theorem.

2 Kushner theorem

By far the most powerful tool for dealing with the convergence of Markov chain based algorithms is the work of Harold Kushner in [2] and [6]. It deals with mixed jump-diffusions processes, for clarity we consider only continuous processes. Kushner's theorem says that the local consistency conditions, that is the matching of the first and second conditional moments of the increments of the approximating chain with those of the continuous-time limit with accuracy $o(h)$ grants the convergence of the expectations of usual functionals. In fact the setting is a stochastic control setting, Kushner's result deals with the convergence of the optimal controlled chain to the optimal controlled process. Here we simplify things in a crude manner by dropping the control. We also restrict ourselves to very particular stopping times especially pertaining to option pricing-this avoids the use of highly technical objects in the theorem statement, more precisely Skorokhod space and topology.

2.1 Notations

2.1.1 The continuous time problem

Consider the s.d.e. with values in $\mathbb{R}^d$

$$X_{t+s} = x + \int_t^{t+s} b(u, X_u)\, du + \int_t^{t+s} \sigma(u, X_u)\, .dW_u$$

where W is a $k-$dimensional Brownian motion. The problem at stake is to design an approximation of the quantity

$$V(t,x) = E_{t,x}\left[g(\tau, X_\tau)\right]$$

wher τ is the first exit time of (s, X_s) an open set G of $\mathbb{R}^{1+d}$, to be specified later, before a fixed time horizon $T > t$, ie

$$\tau = \inf\{u > t,\, (u, X_u) \notin G\} \wedge T$$

Assume:
$(A1)$ b and σ are continuous and bounded.
$(A2)$ g is continuous and bounded.
$(A3)$ Either:
(a) $G = \mathbb{R}^d$ or:
(b) For some index i

$$G = \{\, t < u < T,\, L_i(u) < x_i < U_i(u)\} \tag{2}$$

where L_i, U_i are continuous functions on $[t, T]$ with values in $\overline{\mathbb{R}}$. For the same index i, $\sum_{j=1}^k \sigma_{i,j}^2 (u, X_u) > \alpha$ for some $\alpha > 0$, uniformly in u.

Note that we don't make any non degeneracy assumptions on σ in $(A1)$. It is easily seen that in case of a deterministic process X, there is no hope to approximate $V(t,x)$ for an exit time corresponding to (2) under quite general and natural continuity requirements for the approximate chain-this is exactly the same situation as in conitnuity sets for the limiting distribution in a weak convergence statement. This is the explanation of the additional "wildness" assumption on σ in that case.

2.1.2 The Markov chain approximation

Let for N a positive integer $h = \frac{T}{N}$ and let $\left(\xi_n^h\right)_{n\geq 0}$ denote the value at time nh of a discrete-time Markov chain which satisfies with $\Delta\xi_n^h = \xi_{n+1}^h - \xi_n^h$ the following local consistency conditions:

$$\begin{aligned} E_{x,n}^h\left[\Delta\xi_n^h\right] &= b_h\left(nh,x\right)h = b\left(nh,x\right)h + o\left(h\right) \\ E_{x,n}^h\left[\left(\Delta\xi_n^h - E_{x,n}^h\left[\Delta\xi_n^h\right]\right).\left(\Delta\xi_n^h - E_{x,n}^h\left[\Delta\xi_n^h\right]\right)'\right] &= a_h\left(nh,x\right)h = a\left(nh,x\right)h + o\left(h\right) \end{aligned}$$

which also defines the functions $x \mapsto b_h\left(nh,x\right)$ and $x \mapsto a_h\left(nh,x\right)$.

Here $E_{x,n}^h$ denote the conditionnal expectation at time n knowing $\xi_n^h = x$, also $a\left(s,x\right) = \sigma\left(s,x\right)\sigma\left(s,x\right)'$. Notice that these conditions means that locally the chain has the conditional mean and variance of the continuous process since

$$\begin{aligned} E_{x,s}\left[X_{s+h}\right] &= x + b\left(s,x\right)h + o\left(h\right) \\ E_{x,s}\left[\left(X_{s+h}-x\right).\left(X_{s+h}-x\right)'\right] &= a\left(s,x\right)h + o\left(h\right) \end{aligned}$$

Assume also

$$\sup_{n,\omega}\left|\Delta\xi_n^h\right| \to 0\text{as } h \to 0 \tag{3}$$

Introduce now the approximating stopping time $\widehat{\tau}_h$ which is defined as the first exit time of G of the process $\left(t,\widehat{\xi}^h\left(t\right)\right)$ where $\widehat{\xi}^h\left(t\right)$ is the piecewise constant CADLAG continous time extension of $\left(\xi_n^h\right)$, ie the process

$$\widehat{\xi}^h\left(t\right) = \xi_{n_t h}^h$$

where n_t is the integer such that $n_t h \leq t < n_t h + h$.

Remark 1. *Kushner works with this non-continous interpolation for two purposes: firstly, he deals with mixed jump-diffusion processes. Secondly he makes use of weak convergence on the Skorokhod space which is easier to get than the weak convergence on* $C\left(\mathbb{R}_+,\mathbb{R}\right)$.

2.2 Theorem statement

Theorem 2. *([6], theorem 5.1) Assume* $(A1), (A2), (A3)$. *Then*

$$V(t,x) = E[g(\tau, X_\tau)] = \lim_{h\to 0} E\left[g\left(\widehat{\tau}_h, \widehat{\xi}^h(\widehat{\tau}_h)\right)\right]$$

Notice that an interesting fact in the local consistency condition is that all quantities of relevant interest need only to be known with an $o(h)$ certainty, at least as far as convergence is concerned. This maybe especially noteworthy when the "exact" approximating quantities look like $\exp(\varepsilon(h))$ with $\varepsilon(h)$ going to 0 suitably with h whereas they need to be recomputed at each step.

In the sequel we set

$$V_h(t,x) = E\left[g\left(\widehat{\tau}_h, \widehat{\xi}^h(\widehat{\tau}_h)\right)\right]$$

2.3 Applications

2.3.1 Checking the hypothesis

The first step in most commonly used models for financial markets is to look at the Markov chain approximation at hand not in terms of the underlying of the option S, but rather in $\ln(S)$, that is to set $X_t = \ln(S_t)$. Then $(A1)$ has a chance to be in force, think at the Black-Scholes model for instance.

What about $(A2)$? In a derivative context g stands for the discounted payoff of the option (as a function of the logarithm of S now). It's always positive, but may be unbounded (Call options!) and discontinuous (Digit options). A sound way to proceed is to replace g by $g \wedge a$ for big enough a and if necessary to regularize $g \wedge a$ by some mollifier ψ_ε. The convergence:

$$E\left[e^{-\rho\tau}\psi_\varepsilon * g \wedge a(X_\tau)\right] \to E\left[e^{-\rho\tau} g(X_\tau)\right]$$

is in practical situations always in force.

As already discussed in order to deal with unbounded payoffs some tricks like the Call-Put parity may be used.

2.3.2 Barrier options

The convergence of the Derman-Kani algorithm, or the Kamrad-Ritchken algorithms for say Down and Out European options is easily proved. Notice that the convergence of the crude CRR scheme is proved in the same way, this shows that the main interest of such a convergence theorem is theoretical: in practice, the CRR scheme for barrier options is definitely not usable.

2.3.3 Bidim trees

Here r stands for the instantaneous interest rate in the Black-Scholes model.

Assume (S^1, S^2) follows the Black-Scholes dynamic under the risk-neutral probability and set:

$Y^1 = \ln(S^1)$

$Y^2 = \ln(S^2)$

Then:

$$\begin{aligned} dY_t^1 &= \sigma_1 dB_t^1 + \left(r - \frac{\sigma_1^2}{2}\right) dt \\ dY_t^2 &= \rho\sigma_2 dB_t^1 + \sqrt{1-\rho^2}\sigma_2 dB_t^2 + \left(r - \frac{\sigma_2^2}{2}\right) dt \end{aligned}$$

Choose:

$$\begin{aligned} \left(\Delta\xi_n^h\right)_1 &= \left(r - \frac{\sigma_1^2}{2}\right) h + \sigma_1\sqrt{h}\varepsilon_n^1 \\ \left(\Delta\xi_n^h\right)_2 &= \left(r - \frac{\sigma_2^2}{2}\right) h + \sigma_2\sqrt{h}\varepsilon_n^2 \end{aligned}$$

where: $(\varepsilon_n^1, \varepsilon_n^2)$ is a sequence of iid random variable with

$$\begin{aligned} P\left(\varepsilon_0^1 = 1, \varepsilon_0^2 = 1\right) &= P\left(\varepsilon_0^1 = -1, \varepsilon_0^2 = -1\right) = \frac{1+\rho}{4} \\ P\left(\varepsilon_0^1 = 1, \varepsilon_0^2 = -1\right) &= P\left(\varepsilon_0^1 = -1, \varepsilon_0^2 = 1\right) = \frac{1-\rho}{4} \end{aligned}$$

Then

$$\begin{aligned} b_h(x)_1 h &= \left(r - \frac{\sigma_1^2}{2}\right) h \\ b_h(x)_2 h &= \left(r - \frac{\sigma_2^2}{2}\right) h \end{aligned}$$

and

$$\begin{array}{ll} a_h(x)_{11} h = \sigma_1^2 h & a_h(x)_{12} h = \rho\sigma_1\sigma_2 h \\ & a_h(x)_{22} h = \sigma_2^2 h \end{array}$$

Notice that the corresponding algorithm is of complexity N^3. It seems that an N^4 complexity-a three-dimensionnal problem plus the time component, for instance- is a practical limit.

2.3.4 Trees for markovian volatility

Here r stands for the instantaneous interest rate in the Black-Scholes model.

Assume S follows a one-dimensional Black-Scholes type model with stochastic volatility σ depending on S and t under the risk-neutral probability.

Assume also that

$$0 < \underline{\sigma} < \sigma\left(\ ,\ \right) < \overline{\sigma} < \infty$$

with $\underline{\sigma} = \inf \sigma, \overline{\sigma} = \sup \sigma$ and set:

$Y = \ln\left(S\right)$

Then with $\widehat{\sigma}\left(t, x\right) = \sigma\left(t, \exp\left(x\right)\right)$:

$$dY_t = \widehat{\sigma}\left(t, Y_t\right) dB_t + \left(r - \frac{\widehat{\sigma}\left(t, Y_t\right)^2}{2}\right) dt$$

Choose:

$$\left(\Delta \xi_n^\eta\right) = \left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h + \overline{\sigma} \alpha \sqrt{h} \varepsilon_n$$

where: $\left(\varepsilon_n\right)$ is a sequence of iid random variable with

$$\begin{aligned} P\left(\varepsilon_0 = -1\right) &= \beta + \gamma\sqrt{h},\ P\left(\varepsilon_0 = -1\right) = \beta - \gamma\sqrt{h} \\ P\left(\varepsilon_0 = 0\right) &= 1 - 2\beta \end{aligned}$$

Then

$$\begin{aligned} b_h\left(x\right) h &= \left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h + 2\gamma\overline{\sigma}\alpha h \\ a_h\left(x\right) h &= \left(\beta + \gamma\sqrt{h}\right) \left(\left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h + \alpha\overline{\sigma}\sqrt{h}\right)^2 \\ &\quad + \left(1 - 2\beta\right) \left(\left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h\right)^2 \\ &\quad + \left(\beta - \gamma\sqrt{h}\right) \left(\left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h - \alpha\overline{\sigma}\sqrt{h}\right)^2 \\ &\quad - \left[\left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h + 2\gamma\overline{\sigma}\alpha h\right]^2 \\ &= \left(\beta + \gamma\sqrt{h}\right) \overline{\sigma}^2 \alpha^2 h + \left(\beta - \gamma\sqrt{h}\right) \overline{\sigma}^2 \alpha^2 h - 4\gamma^2 \overline{\sigma}^2 \alpha^2 h \\ &= \left(2\beta - 4\gamma^2\right) \overline{\sigma}^2 \alpha^2 h \end{aligned}$$

So we're led to the choice:

$$\begin{aligned} \left(r - \frac{\overline{\sigma}^2 \delta}{2}\right) h + 2\gamma\overline{\sigma}\alpha h &= \left(r - \frac{\widehat{\sigma}\left(nh, \xi_n^\eta\right)^2}{2}\right) h \\ \left(2\beta - 4\gamma^2\right) \overline{\sigma}^2 \alpha^2 h &= \widehat{\sigma}\left(nh, \xi_n^\eta\right)^2 h \end{aligned}$$

ie

$$
\begin{aligned}
\gamma\overline{\sigma}\alpha &= \frac{1}{4}\left(\overline{\sigma}^2\delta - \widehat{\sigma}\left(nh,\xi_n^\eta\right)^2\right)\\
\beta\overline{\sigma}^2\alpha^2 &= \frac{1}{2}\left[\widehat{\sigma}\left(nh,\xi_n^\eta\right)^2 + \frac{1}{4}\left(\overline{\sigma}^2\delta - \widehat{\sigma}\left(nh,\xi_n^\eta\right)^2\right)^2\right]
\end{aligned}
$$

A natural choice which yields a symmetric scheme for constant volatility is $\delta = 1$, whence:

$$
\begin{aligned}
\gamma\overline{\sigma}\alpha &= \frac{1}{4}\left(\overline{\sigma}^2 - \widehat{\sigma}\left(nh,\xi_n^\eta\right)^2\right)\\
\beta\overline{\sigma}^2\alpha^2 &= \frac{1}{2}\left[\widehat{\sigma}\left(nh,\xi_n^\eta\right)^2 + \frac{1}{4}\left(\overline{\sigma}^2 - \widehat{\sigma}\left(nh,\xi_n^\eta\right)^2\right)^2\right]
\end{aligned}
$$

Considering α as a constant parameter we must check $1 - 2\beta \geq 0$ ie:

$$
\left[\widehat{\sigma}\left(nh,\xi_n^\eta\right)^2 + \frac{1}{4}\left(\overline{\sigma}^2 - \widehat{\sigma}\left(nh,\xi_n^\eta\right)^2\right)^2\right] \leq \overline{\sigma}^2\alpha^2
$$

Therefore it is enough to choose α^2 such that

$$
\alpha^2 \geq \sup_{u\in\left[\underline{\sigma}^2,\overline{\sigma}^2\right]} \frac{\left[u + \frac{1}{4}\left(\overline{\sigma}^2 - u\right)^2\right]}{\overline{\sigma}^2}
$$

Note that the best choice of α is questionable.

3 Analytical approach

An alternative route to the convergence of Tree methods is provided by the classical theory of approximation of semigroups of linear operators. An authority book on the subject is [5] (The 3^{rd} chapter is devoted to approximation issues). A survey may be found in [3].

3.1 From CRR to the Black-Scholes PDE

This corresponds to the well-known elementary calculation which relates the backward recursion scheme in the CRR model and the Black-Scholes PDE. The corresponding relations for other schemes may be found in [7].

One can re-write

$$
C\left(nh,x\right) = e^{-\rho h}\left[p^*\left(h\right)C\left(\left(n+1\right)h, xu\left(h\right)\right) + \left(1-p^*\left(h\right)\right)C\left(\left(n+1\right)h, xd\left(h\right)\right)\right]
$$

as

$$\begin{aligned} & C(nh,x) - C((n+1)h,x) \\ = \; & e^{-\rho h}\left[p^*(h)\left(C((n+1)h,xu(h)) - C((n+1)h,x)\right)\right. \\ & \left. + (1-p^*(h))\left(C((n+1)h,xd(h)) - C((n+1)h,x)\right)\right] \\ & + \left(1-e^{-\rho h}\right) C((n+1)h,x) \end{aligned}$$

Now by Taylor expansion

$$\begin{aligned} F(xu(h)) &= F(x) + (u(h)-1)\, xF'(x) + \frac{1}{2}(u(h)-1)^2 x^2 F''(x) + o\left((u(h)-1)^2\right) \\ F(xd(h)) &= F(x) + (d(h)-1)\, xF'(x) + \frac{1}{2}(d(h)-1)^2 x^2 F''(x) + o\left((d(h)-1)^2\right) \end{aligned}$$

Since $p^*(h)$ is the risk-neutral probability

$$p^*(h)(u(h)-1) + (1-p^*(h))(d(h)-1) = e^{\rho h} - 1$$

It is easily proved (this is the second moment condition from a probabilistic point of view) that

$$p^*(h)(u(h)-1)^2 + (1-p^*(h))(d(h)-1)^2 = \sigma^2 h + o(h)$$

so that putting all things together yields

$$\begin{aligned} & \frac{C(nh,x) - C((n+1)h,x)}{h} \\ = \; & e^{-\rho h}\left[\frac{\left(e^{\rho h}-1\right)}{h} x \frac{\partial C((n+1)h,x)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C((n+1)h,x)}{\partial x^2}\right] \\ & + \frac{\left(1-e^{-\rho h}\right)}{h} C((n+1)h,x) + o(h) \end{aligned}$$

Therefore by taking h to zero we get that C solves the PDE

$$-\frac{\partial C(t,x)}{\partial t} = \rho x \frac{\partial C(t,x)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 C(t,x)}{\partial x^2} - \rho C(t,x)$$

which is the Black-Scholes PDE.

3.2 Formal analytic point of view on Tree methods

A more synthetic way to write the above calculation is the following: a Tree method is specified by a backward functional operator which maps the price at time $(n+1)h$ to the price at time nh, formally

$$C_{nh} = F_h C_{(n+1)h}$$

Here the price should be seen as a function which gives the price of the option for any value of the underlying, that is

$$C_{nh} : x \mapsto C(nh, x)$$

Define now an operator A_h by the relation

$$F_h = Id + hA_h$$

Then the above computation says $A_h C_t \to AC_t$ where A is in our case the Black-Scholes differential operator, ie

$$Af = \rho x f'(x) + \frac{1}{2}\sigma^2 x^2 f''(x) - \rho f(x)$$

Now the price of the option at time 0 is given by

$$C_0(N) = F_{\frac{T}{N}}^N \varphi = \left(Id + \frac{T}{N} A_{\frac{T}{N}}\right)^N \varphi$$

We may therefore hope that

$$\lim_{N\to\infty} C_0(N) = \lim_{N\to\infty} \left(Id + \frac{T}{N} A\right)^N \varphi = \exp(T A)\varphi$$

where $t \mapsto \exp(tA)\varphi$ is defined as the solution to the equation

$$\frac{du}{dt} = Au \tag{4}$$

with initial condition $u(0) = \varphi$.

3.3 Semigroups of linear operators

Assume that the above linear equation has a unique solution. Then by unicity the map $u(0) \overset{Q_t}{\to} u(t)$ satisfies

$$Q_{t+s} = Q_t Q_s$$

ie we face a semigroup of linear operators. The theory of such objects has been extensively developed in relations to various PDE problems.

A good framework for these operator semigroups is that of Banach spaces.

From now on we consider a semigroup Q of contraction operators on a Banach space G, which is continuous, ie for any g in G

$$Q_t g \to g$$

as $t \to 0$.

It may be shown that there is a dense subspace of G on with $t \mapsto Q_t g$ is differentiable. If we denote A the linear operator (with domain) such defined, $Q_t g$ formally solves

$$\frac{du}{dt} = Au$$

with initial condition $u(0) = g$.

The operator A is called thye infinitesimal generator of the semigroup.

As an exemple:

Lemma 1. *Let* $G = \left\{ g : \mathbb{R}_+^* \to \mathbb{R},\ g\, C^0,\ \exists \lim_{x \to 0^+} \frac{g(x)}{1+x},\ \exists \lim_{x \to \infty} \frac{g(x)}{1+x} \right\}$*. Then* G *is a Banach space for the norm*

$$\|g\| = \sup_x \frac{|g(x)|}{1+x}$$

Moreover, if $Q_t g(x)$ *is the price of an European option with maturity* $t \geq 0$ *and payoff* g *in the Black-Scholes model, for a value* x *of the underlying, then* Q *is a continuous semigroup of contraction operators on* G*. Moreover if* g *is a* C^∞ *function then* $t \mapsto Q_t g$ *is differentiable and*

$$\frac{dQ_t g}{dt} = AQ_t g = \rho x \frac{\partial Q_t g(x)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 Q_t g(x)}{\partial x^2} - \rho Q_t g(x)$$

Lastly the space C_0^∞ *is dense in* G*.*

3.4 Approximation theory

The first ingredient is a quantitative estimate:

Lemma 2. *([5], section 3.5, corollary 5.2) Let* R *a linear contraction operator. Then for every integer* $n \geq 0$ *and* $g \in G$

$$\|\exp(n(R - Id))\, g - R^n g\| \leq \sqrt{n}\, \|g - Rg\|$$

The second ingredient is:

Theorem 3. *Let* $(A_h)_{h>0}$ *a family of bounded operators such that*

$$\sup_h \|\exp(tA_h)\|$$

is uniformly bounded for t *in a compact set.*

Let on a dense subset H *of* G

$$A_h g \to Ag$$

as $h \to 0$ Then for any $g \in G$

$$\exp(tA_h)\, g \to Q_t g$$

as $h \to 0$ and the convergence is uniform for t in a compact set.

We give the proof only in the case where A and A_h commute. Then for $g \in H$

$$\begin{aligned}
& \exp(tA_h)\, g - Q_t g \\
= & \left[\exp(sA_h)\, Q_{t-s} g\right]_{s=0}^{s=t} \\
= & \int_0^t \frac{d}{ds} \exp(sA_h)\, Q_{t-s} g ds \\
= & \int_0^t \left(A_h \exp(sA_h)\, Q_{t-s} g - \exp(sA_h)\, A Q_{t-s} g\right) ds \\
= & \int_0^t \left(\exp(sA_h)\, Q_{t-s} A_h g - \exp(sA_h)\, Q_{t-s} A g\right) ds \\
= & \int_0^t \exp(sA_h)\, Q_{t-s} \left(A_h - A\right) g ds
\end{aligned}$$

Whence

$$\begin{aligned}
\left\|\exp(tA_h)\, g - Q_t g\right\| & \leq \int_0^t \exp(sA_h)\, Q_{t-s} \left(A_h - A\right) g ds \\
& \leq \int_0^t \left\|\exp(sA_h)\right\| \left\|Q_{t-s}\right\| \left\|\left(A_h - A\right) g\right\| ds \\
& = \left\|\left(A_h - A\right) g\right\| \int_0^t \left\|\exp(sA_h)\right\| ds \\
& \leq t \left\|\left(A_h - A\right) g\right\| \sup_{h,[0,t]} \left\|\exp(sA_h)\right\|
\end{aligned}$$

which gives the result for $g \in H$.

The result for $g \in G$ follows by density since both $g \mapsto \exp(tA_h)$ and $g \mapsto Q_t g$ are continuous on G.

Here is now the main result:

Theorem 4. *Let $(F_h)_{h>0}$ a family of bounded linear operators such that for some M and every h, $k \in \mathbb{N}$*

$$\left\|F_h^k\right\| \leq M \tag{5}$$

Let for a dense subset H of G

$$\frac{F_h g - g}{h} \to Ag \tag{6}$$

as $h \to 0$ *Then for any* $t \geq 0$ *and* $g \in G$

$$F^n_{\frac{t}{n}} g \to Q_t g$$

as $n \to \infty$

Indeed

$$Q_t g - F^n_{\frac{t}{n}} g = Q_t g - \exp\left(t\frac{F_{\frac{t}{n}} - Id}{\frac{t}{n}}\right) g + \exp\left(t\frac{F_{\frac{t}{n}} - Id}{\frac{t}{n}}\right) g - F^n_{\frac{t}{n}} g$$

Setting now $A_h = \frac{F_h - Id}{h}$ let us notice

$$\begin{aligned}
\left\|\exp\left(t\frac{F_h - Id}{h}\right)\right\| &= \exp\left(-\frac{t}{h}\right)\sum_k \frac{\left(\frac{t}{h}\right)^k \left\|F_h^k\right\|}{k!} \\
&\leq M\exp\left(-\frac{t}{h}\right)\sum_k \frac{\left(\frac{t}{h}\right)^k}{k!} \\
&= M
\end{aligned}$$

From the previous theorem it follows that

$$\exp\left(t\frac{F_{\frac{t}{n}} - Id}{\frac{t}{n}}\right) g \to Q_t g$$

as $n \to \infty$ For the other term if $g \in H$

$$\begin{aligned}
&\left\|\exp\left(t\frac{F_{\frac{t}{n}} - Id}{\frac{t}{n}}\right) g - F^n_{\frac{t}{n}} g\right\| \\
= &\left\|\exp\left(n\left(F_{\frac{t}{n}} - Id\right)\right) g - F^n_{\frac{t}{n}} g\right\| \\
\leq &\sqrt{n}\left\|F_{\frac{t}{n}} g - g\right\| \\
\leq &\sqrt{n}\frac{t}{n} 2\left\|Ag\right\|
\end{aligned}$$

for big enough n.

The result is proved for $g \in H$, the theorem follows from the density of H in G since for $(g, h) \in G \times H$

$$\begin{aligned}
\left\|F^n_{\frac{t}{n}} g - F^n_{\frac{t}{n}} h\right\| &= \left\|F^n_{\frac{t}{n}}(g - h)\right\| \leq M\left\|g - h\right\| \\
\left\|Q_t g - Q_t h\right\| &= \left\|Q_t (g - h)\right\| \leq \left\|g - h\right\|
\end{aligned}$$

Remark 2. *The first condition in the theorem is a stability condition in the numerical analysis language, the second one a consistency condition. Thus this theorem may be seen as a version of the statement: consistency and stability yield convergence.*

Remark 3. *In practice, the tough part is to prove the consistency condition (6 on page 13). Indeed it must be shown that the limit holds in the Banach space sense, which is a much stronger assertion than a pointwise limit. On the other hand, the convergence of the approximation also holds in the Banach space sense, in this direction it is a stronger result than a statement in a probabilistic theorem which holds given the starting point of the diffusion, for instance.*

3.5 Application

As an application, let us prove the convergence of the Call option price in the CRR model to the Black-Scholes price.

First observe that the Call option payoff belongs to the above Banach space G. Now by the lemma 1 on page 12 and the approximation theorem it is enough to show that, if F_h denotes the backward transition operator in the CRR model, (5 on page 13) and (6 on page 13) hold.

For (5 on page 13):

$$F_h g(x) = e^{-\rho h}\left[p^*(h)\, g(xu(h)) + (1-p^*(h))\, g(xd(h))\right]$$

whence

$$\begin{aligned}
|F_h g(x)| &\leq \|g\|\, e^{-\rho h}\left[p^*(h)(1+xu(h)) + (1-p^*(h))(1+xd(h))\right] \\
&= \|g\|\, e^{-\rho h}\left(1+xe^{\rho h}\right) \\
&\leq \|g\|(1+x)
\end{aligned}$$

Thus

$$\|F_h g\| \leq \|g\|$$

and F_h is a contraction, in particular (5 on page 13) holds.

For (6 on page 13), for g in C_0^∞

$$\begin{aligned}
\frac{F_h g(x) - g(x)}{h} &= e^{-\rho h}\left[p^*(h)\left(g(xu(h)) - g(x)\right) + (1-p^*(h))\, g(xd(h) - g(x))\right] \\
&= \rho x g'(x) + \frac{1}{2}\sigma^2 x^2 g''(x) - \rho g(x) + \sqrt{h} R(g,x)
\end{aligned}$$

where $R(g,x)$ vanishes outside the support of g, by the computation in the beginning and Taylor's formula. This entails

$$\frac{F_h g - g}{h} \to Ag$$

in G, where A is the infinitesimal generator of the Black-Scholes semigroup, whence the result.

4 Some words about the rate of convergence

The question of the rate of convergence of a given tree for a given payoff is crucial: in practice a behavior in $O\left(\frac{1}{N}\right)$ (ie $V_h(t,x) - V(t,x) = O_{t,x,g}\left(\frac{1}{N}\right)$) of the algorithm is satisfactory, whereas $O\left(\frac{1}{\sqrt{N}}\right)$ is not.

Unfortunately this is a very tough question. Only a few tree algorithms have been studied in detail from that point of view until now, mostly for standard options. The main point is that the rate of convergence depends heavily not only on the scheme but also on the smoothness of the payoff.

4.1 Results for i.i.d. trees for standard options

By i.i.d. tree we mean a Markov chain approximation such that, with the notations of the Kushner theorem, the sequence $\left(\Delta\xi_n^h\right)_{n\geq 0}$ is i.i.d. for every fixed h. The first idea in this case is to make use of the characteristic function to study the convergence rate. Indeed for suitable f

$$f(x) = \int e^{i\lambda x}\widehat{f}(\lambda)\,d\lambda \tag{7}$$

where $\widehat{f}$ is the Fourier Transform of f. Therefore

$$\begin{aligned}
&E_{t,x}\left[f(X_T)\right] - E_{t,x}\left[f\left(\xi_N^h\right)\right] \\
= &\int\left(E_{t,x}\left[e^{i\lambda X_T}\right] - E_{t,x}\left[e^{i\lambda\xi_N^h}\right]\right)\widehat{f}(\lambda)\,d\lambda \\
= &\int\left(E_{t,x}\left[e^{i\lambda(X_T - x)}\right] - \left(E\left[e^{i\lambda\Delta\xi_0^h}\right]\right)^N\right)e^{i\lambda x}\widehat{f}(\lambda)\,d\lambda
\end{aligned}$$

where we made use of the i.i.d. assumption. Assume for instance $X_t = x + B_t$, then $E_{t,x}\left[e^{i\lambda X_T}\right] = e^{-\frac{\lambda^2}{2}T}$, and denoting by $\psi_h(\lambda)$ the characteristic function of $\Delta\xi_0^h$:

$$\psi_h(\lambda) = E\left[e^{i\lambda\Delta\xi_0^h}\right]$$

we get

$$
\begin{aligned}
& E_{t,x}\left[f\left(X_T\right)\right]-E_{t,x}\left[f\left(\xi_N^h\right)\right] \\
= & \int\left(e^{-\frac{\lambda^2}{2}T}-\psi_h\left(\lambda\right)^N\right)e^{i\lambda x}\widehat{f}\left(\lambda\right)d\lambda
\end{aligned}
\tag{8}
$$

Of course for these formal calculations to be meaningful we require that the inversion formula (7 on the previous page) holds and also to justify Funbini's theorem

$$
\begin{aligned}
\int e^{-\frac{\lambda^2}{2}T}\left|\widehat{f}\left(\lambda\right)\right|d\lambda &< \infty \\
\textit{For every } N,\ \int \psi_h\left(\lambda\right)^N\left|\widehat{f}\left(\lambda\right)\right|d\lambda &< \infty
\end{aligned}
$$

By standard Fourier transform theory it is seen that these requirements involve the smoothness of f. Expression (8) clearly shows that the rate of convergence depends on the control of $e^{-\frac{\lambda^2}{2}T}-\psi_h\left(\lambda\right)^N$ in terms of λ altogether with integrability properties of $\widehat{f}\left(\lambda\right)$, that is smoothness of f.

4.1.1 Random Walk approximations

A particular case which has been investigated in great detail is the Random Walk case, which means that we look at the approximation of say $E\left[f\left(B_1\right)\right]$ by $E\left[f\left(\xi_N^h\right)\right]$ where

$$
\begin{aligned}
& \left(\Delta\xi_n^h\right)_{n\geq 0}\text{ i.i.d.with }\Delta\xi_0^h=\sqrt{h}\varepsilon \\
& E\left[\varepsilon\right]=0,\ E\left[\varepsilon^2\right]=1
\end{aligned}
$$

where ε does not depend on h. Note that we relax here the boundedness assumption (3 on page 5). This subsection is taken from [1]

The above route leads to the following statements:

Theorem 5. *Assume f is C_b^5. Then*

$$
\left|E\left[f\left(X_T\right)\right]-E\left[f\left(\xi_N^h\right)\right]\right|=\frac{E\left[\varepsilon^3\right]}{\sqrt{N}}a\left(f\right)+\frac{E\left[\varepsilon^4\right]-3}{N}b\left(f\right)+o\left(\frac{1}{N}\right)
$$

with $a\left(f\right)=\frac{1}{6}E\left[f'''\left(B_1\right)\right]$ and $b\left(f\right)=\frac{1}{24}E\left[f^{(4)}\left(B_1\right)\right]$.

For f ressembling a Call or a Put payoff we have:

Theorem 6. *Assume f is Lipschitz and that f'' (taken in the sense of distributions) is a bounded measure. Then if $X = B$ and where and also*

$$E\left[\varepsilon^4\right] < \infty,\ E\left[\varepsilon^3\right] = 0$$

then

$$\left|E\left[f\left(B_1\right)\right] - E\left[f\left(\xi_N^h\right)\right]\right| \leq \frac{C}{N}\left(\|f''\|_{M_b(\mathbb{R})} + \|f'\|_{L^\infty}\right)$$

where C depends only on the law of ε.

By a suitable change of variables this result applies to European Call and Put options in the Random Walk tree.

It is interesting to note that in contrary to the C_b^5 case with $E\left[\varepsilon^3\right] = 0$, there is no expansion in $\frac{1}{N}$ in general under the assumptions of the theorem: indeed for $f(x) = |x|$ it may be proved when $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$ that

$$\begin{aligned} E\left[f\left(B_1\right)\right] - E\left[f\left(\xi_N^h\right)\right] &= \frac{1}{2N\sqrt{2\pi}} + o\left(\frac{1}{N}\right) \text{ if } N \text{is odd} \\ &= -\frac{1}{2N\sqrt{2\pi}} + o\left(\frac{1}{N}\right) \text{ if } N \text{is even} \end{aligned}$$

The corresponding result for the CRR tree for Call and Put option has been obtained recently by M&F Diener in [4] by a direct calculation:

Proposition 1. *Let a Call ATM option in the Black-Scholesmodel with no interest rate. Then if N is even*

$$CRR(N) = BS - \frac{1}{192\sqrt{\pi}}\sigma\sqrt{2}e^{-\frac{1}{8\sigma^2}}\left(\sigma^2 + 12\right) \times \frac{1}{m} + o\left(\frac{1}{m}\right)$$

where $N = 2m$.
If N is odd

$$CRR(N) = BS - \frac{1}{192\sqrt{\pi}}\sigma\sqrt{2}e^{-\frac{1}{8\sigma^2}}\left(\sigma^2 - 12\right) \times \frac{1}{m} + o\left(\frac{1}{m}\right)$$

where $N = 2m + 1$

Along this line one can wonder about Digit options. Not very surprisingly the following holds-this can be proved by a direct caclulation:

Proposition 2. *Assume $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$ (in particular, $E\left[\varepsilon^3\right] = 0$). Then*

$$E\left[1\left(B_1 < 0\right)\right] - E\left[1\left(\xi_N^h < 0\right)\right] = \frac{1}{\sqrt{2\pi N}} + o\left(\frac{1}{\sqrt{N}}\right)$$

4.2 Results for i.i.d. trees for path-dependant options

Very few is known, even for barrier options. Empirical studies suggest a convergence with a $\frac{1}{N}$ rate for improved trees like the Derman-Kani or the Kamrad-Ritchken tree. Fore more general situations, the general feeling is that a convergence rate of $O\left(\frac{1}{\sqrt{N}}\right)$ is in some sense the worst that can be achieved.

4.3 Order of Accuracy and Order of convergence

A widely-used concept in numerical analysis of a finite-difference scheme is that of order of accuracy: in the calculation above (cf 4 on page 10), this is the order in power of h at which the Black-Scholes PDE is satisfied. In a heuristic manner this amounts to look how far, locally, is the numerical approximation scheme from its continuous limit. A detailed study of the order of accuracy of tree methods may be found in [7], where also classical finite-difference schemes are studied (see also [8]).

The main thing to say about this concept is that it is not directly related in full generality to the good notion which is that of order of convergence as studied before. In analytical terms the point is that the order of accuracy is a local (event pointwise) measurement of the convergence of the infinitesimal generator of the semigroup, whereas because of the diffusive feature of the semigroups at hand a kind of uniform control is required to give something about convergence. This is well-explained by the proof of the abstract theorem (cf theorem 4 on page 13). Moreover, it was shown that for non-smooth payoffs, even for a scheme with high degree of accuracy the order of convergence is much lower. This is perfectly in tune with the fact that there is no smoothing effect of the discrete semigroup of the approximating scheme, so that the above computation can not be performed in the non-smooth case.

Nevertheless it seems it is possible in many cases to go from the order of accuracy to the order of convergence for sufficiently smooth intial condition (or payoff functions for us)-unfortunately, this is almost never the case in finance. An interesting exemple of this is the analytical reading of the previous Lamberton's theorem 5 on page 17. Let us look for a trinomial tree (in logarithm) such that the two first moments are 0 and 1. Let us denote by A the upper node, B the lower node, a and b the corresponding weights. Suppose also the middle node is centered, and let us compute the order of accuracy. Let f a smooth space-time function, $c = 1.0 - a - b$

$$
\begin{aligned}
& af\left(x + A\sqrt{h}, t+h\right) + cf\left(x, t+h\right) + bf\left(x - B\sqrt{h}, t+h\right) - f\left(t, x\right) \\
= \ & \frac{\partial f}{\partial t} h + \frac{\partial f}{\partial x}\left(aA\sqrt{h} - bB\sqrt{h}\right) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\left(aA^2 h + bB^2 h\right) \\
& + \frac{\partial^2 f}{\partial x \partial t}\left(aAh\sqrt{h} - bBh\sqrt{h}\right) \\
& + \frac{1}{6}\frac{\partial^3 f}{\partial x^3}\left(aA^3 h\sqrt{h} - bB^3 h\sqrt{h}\right) \\
& + \frac{1}{24}\frac{\partial^4 f}{\partial x^4}\left(aA^4 h^2 + bB^4 h^2\right) + \frac{1}{2}\frac{\partial^2 f}{\partial t^2} h^2 \\
& + \frac{\partial^3 f}{\partial x^2 \partial t}\left(aA^2 h^2 + bB^2 h^2\right)\frac{1}{2} + o\left(h^2\right)
\end{aligned}
$$

Now by the first and second moment conditions:

$$
\begin{aligned}
aA\sqrt{h} - bB\sqrt{h} &= 0 \\
aA^2 h + bB^2 h &= 1
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \frac{1}{h}\left[af\left(x + A\sqrt{h}, t+h\right) + cf\left(x, t+h\right) + bf\left(x - B\sqrt{h}, t+h\right) - f\left(t, x\right)\right] - \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\right) \\
= \ & \frac{1}{6}\frac{\partial^3 f}{\partial x^3}\left(aA^3 - bB^3\right) h\sqrt{h} \\
& + \left(\frac{1}{2}\frac{\partial^2 f}{\partial t^2} + \frac{1}{24}\frac{\partial^4 f}{\partial x^4}\left(aA^4 + bB^4\right) + \frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial t}\left(aA^2 + bB^2\right)\right) h^2
\end{aligned}
$$

Now the condition $E\left[\varepsilon^3\right] = 0$ in Lamberton's theorem reads

$$aA^3 h\sqrt{h} - bB^3 h\sqrt{h} = 0$$

which cancels the term of order $h\sqrt{h}$.

Let us now turn to the condition $E\left[\varepsilon^4\right] = 3$ in the theorem. This amounts here to

$$aA^4 + bB^4 = 3$$

The coefficient of the h^2 term is then

$$
\begin{aligned}
& \frac{1}{2}\frac{\partial^2 f}{\partial t^2} + \frac{3}{24}\frac{\partial^4 f}{\partial x^4} + \frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial t} \\
= \ & \frac{1}{2}\frac{\partial^2 f}{\partial t^2} + \frac{1}{8}\frac{\partial^4 f}{\partial x^4} + \frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial t}
\end{aligned}
$$

which does not vanish for an arbitrary function f. Nevertheless, if f satisfies the limiting PDE

$$\frac{\partial f}{\partial t}+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}=0$$

then $\frac{1}{2}\frac{\partial^4 f}{\partial x^4}=-\frac{\partial^3 f}{\partial x^2 \partial t}$ whence

$$\begin{aligned}
& \frac{1}{2}\frac{\partial^2 f}{\partial t^2}+\frac{1}{8}\frac{\partial^4 f}{\partial x^4}+\frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial t} \\
= & \frac{1}{2}\frac{\partial^2 f}{\partial t^2}+\frac{1}{4}\frac{\partial^3 f}{\partial x^2 \partial t} \\
= & \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial f}{\partial t}+\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\right) \\
= & 0
\end{aligned}$$

Thus a kind of reverse accuracy matching property is satisfied at one order further: if f is the solution to the PDE, then the backward scheme of the discrete algorithm is matched at order $o\left(h^2\right)$.
Nevertheless a general result about equivalence for smooth intial conditions between the order of accuracy and the order of convergence does not seem to be available yet.

References

[1] D.LAMBERTON. Random walk approximation and option prices. *Proceedings of the 5th CAP Workshop on Mathematical Finance, Columbia University, November 1998*, page Unknown, 1999. 17

[2] H.J.KUSHNER. *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations.* Academic Press, New York, 1977. 3

[3] G.RUIZ-GOLDSTEIN J.A.GOLDSTEIN, editor. *Semigroups of linear and nonlinear operators and applications.* Kluwer Academic Publisher, 1992. 9

[4] F.DIENER M.DIENER. Asymptotics of the price oscillations of a european call option in a tree model. *Mathematical finance*, 14(2), 2004. 18

[5] A. PAZY. *Semigroups of Linear Operators and Applications to partial Differential Equations.* Springer, 1983. 9, 12

[6] H.KUSHNER P.G.DUPUIS. *Numerical Methods for Stochastic Control Problems in Continous Time*. Springer-Verlag, 1992. 3, 6

[7] Y.W.KWOK. *Mathematical models of financial derivatives.* Springer Finance, 1998. 9, 19

[8] Y.W.KWOK. Some theoretical aspects of option pricing algorithms. *Preprint Hong-Kong University of Science and Technology*, 1999. 19