Elementary stochastic calculus for finance with infinitesimals

Jiří Witzany

Abstract. The concept of an equivalent martingale measure is of key importance for pricing of financial derivative contracts. The goal of the paper is to apply infinitesimals in the non-standard analysis set-up to provide an elementary construction of the equivalent martingale measure built on hyperfinite binomial trees with infinitesimal time steps.

Keywords: equivalent martingale measure; option pricing; stochastic processes; non-standard analysis

Classification: Primary 60H05, 60J65

1. Introduction

Stochastic calculus has become a key mathematical tool for derivatives pricing. The basic ideas can be quite easily explained in a discrete setup (Shreve, 2005) but the full continuous time theoretical foundations (see, e.g. Shreve, 2004) present a challenge for non-mathematical students and practitioners. In fact, many lecturers and popular textbooks on derivatives (see e.g. Hull, 2011, or Rebonato, 2004) explain the principles in detail only in the context of finite binomial trees and then pass to the continuous models in a heuristic way avoiding a rigorous treatment of the stochastic analysis. The goal of this survey paper is to outline key theoretical tools using the concept of infinitesimals that allows us to work in a discrete setup with infinitesimally small time steps and so use the results more or less directly in continuous time. We believe that the heart of the derivative pricing argument really lies in a very simple one-step binomial tree model while everything else, i.e. passage to multi-step binomial trees and continuous models, is just a "technical stuff". Therefore, our aim is to make the derivatives pricing technique more accessible and applicable for financial engineers and practitioners that do not have special financial calculus training.

Let us start with a few classical definitions. A *martingale* is a zero drift stochastic process $M(t)$ such that $M(t) = E[M(T) \,|\, t]$ for every $t < T$ where $E[\cdot \,|\, t]$ denotes the conditional expectation given the information at time t. If $M(t) = \frac{f(t)}{g(t)}$

DOI 10.14712/1213-7243.2015.192

The author was supported in part by the Czech Science Foundation grant P402/12/G097 "Dynamical Models in Economics" and by the Institutional Support grant VSE IP100040.

is the ratio of prices of derivative security prices $f(t)$ and $g(t)$ where $g(t)$, called *numeraire*, is always positive, and if $M(t)$ is a martingale, then $f(t)$ can be expressed as $g(t) \times E\left[\frac{f(T)}{g(T)} \mid t\right]$ as long as we are able to calculate analytically or numerically the conditional expected value. The time T is typically the expiration time of a European type derivative. The most popular numeraires are the money market account or the zero coupon bond. The *money market account* starts with a unit value $g(0) = 1$ and accrues the market interest rate on continuous basis, i.e. $g(t) = \exp\left(\int_0^t r(s)ds\right)$. The *zero coupon bond* $P(t,T)$ is specified as the market value a unit payment discounted from time T to t. The main advantage of the zero coupon bond numeraire is that if $M(t) = \frac{f(t)}{P(t,T)}$ is a martingale then $f(t) = P(t,T) \times E_t[f(T)]$ since $P(T,T) = 1$. The numeraire is in particular useful for valuation of interest rate derivatives where we must simply work with the fact that interest rates are stochastic (not constant as in the basic Black-Scholes model). Nevertheless, it is clear that the fraction will rarely be a martingale in the real financial world where investors require a higher return for higher risk. If we were able to adjust the measure under which the expected values are taken so that the process becomes a martingale then there is a chance that the valuation could be realized.

The equivalent martingale measure theorem provides exactly what we need. It says, roughly speaking, that if $g(t)$ is a numeraire then the original measure P can be adjusted to a new measure Q so that $M(t) = \frac{f(t)}{g(t)}$ becomes a martingale with respect to Q. In the context of the pricing method explained above, this is the key result for valuation of various derivative contracts (see Hull (2011)).

The theorem is often applied and even heuristically proved without mentioning the rigorous concepts related to stochastic processes like σ-algebras, filtrations, etc. The standard mathematical theory behind the technique is indeed quite technically difficult. It should be noted that the famous Black-Scholes formula has been discovered using heuristic arguments without the rigorous modern theory of stochastic processes. The stochastic theory has been developed ex post, partially in order to give a precise foundation to the Black-Scholes theory. We are going to build equivalent martingale measures in an elementary set-up on hyperfinite binomial trees applying the methods of Cox, Ross, and Rubinstein (1979) but with infinitesimal time steps. We follow the paper of Cutland, Kopp, and Willinger (1991) but generalize the approach to a general numeraire with multiple sources of uncertainty and also simplify the construction eliminating the notion of filtration building the processes step-by-step only on binomial trees. We believe that the approach provides an intuitive framework on which further modeling concepts like processes with stochastic volatility could be easily developed.

Besides the traditional intuitive approach, infinitesimals can be introduced in several logically consistent frameworks. The fact that the set of real numbers can be extended within ZFC (Zermelo-Fraenkel set theory with the Axiom of Choice) with infinitesimals satisfying most of their intuitive properties was discovered by A. Robinson (1961). The field of mathematics that has subsequently developed is called Non-Standard Analysis (NSA). In the Robinsonian approach, the set of classical (standard) numbers R is extended to ${}^*R \supset R$ and more generally all functions or structures $X \in V(R)$ are extended (mapped) to ${}^*X \in V({}^*R)$ so that a transfer principle holds. This approach is appealing since we start from the known world and enrich it with many new useful objects, in particular with infinitesimals. It turns out that concepts like continuity, derivative, integral, etc., can be suddenly intuitively and very quickly defined with infinitesimals in hand. However, there is a danger that contradicts the original goal to make things more simple. The existence of the standard and non-standard universe leads to the development of new non-standard concepts in parallel to the standard ones and to the investigation of relationships between the two worlds. At the end, the non-standard mathematics often becomes much more complex and less accessible than the standard approach. For example, looking at the book Albeverio et al. (1986) that provides a very good overview of non-standard methods in stochastic analysis, one realizes that the reader must capture not only the classical stochastic analysis concepts, but also the non-standard counterparts, and the complex interplay between them in the form of various lifting theorems. There is a number of other publications developing non-standard probability theory and non-standard stochastic analysis with applications in finance, in particular Loeb (1975), Anderson (1976), Keisler (1984), Cutland et al. (1991, 1993, 1995), Kopp (1997), Cutland (2000), or Lindstrøm (2008). In spite of the effort, the NSA approach has not become widespread within the applied financial engineering or more theoretical mathematical finance literature.

Another NSA framework called Internal Set Theory (IST), aiming to be more elementary, has been proposed by Nelson (1977). The idea is to start from the classical universe and extend the language by a new predicate *standard* postulating existence of a non-standard natural number and an analogy of the Robinsonian transfer principle. This axiomatic approach offers a new perspective: looking at the classical universe, e.g. a set of natural numbers, we suddenly realize that some numbers are really finite (standard) while other numbers become in a sense infinite (non-standard or hyperfinite in the Robinsonian terminology). We can say that the class of standard sets in IST corresponds to $V(R)$ while the full universe corresponds to the class of all internal sets ${}^*V(R) = \bigcup_n {}^*V_n(R)$. In other words, the IST universe contains only internal sets corresponding to the well behaved internal sets in the Robinsonian framework but not external sets that can be defined only as classes by external formulas using the predicate *standard*. Note that external sets can be formed in the Robinsonian framework. For example, the class of all standard real numbers is not a set in IST. Therefore, we cannot work with classical functions and other structures built on the set of standard real numbers. This fact imposes a self-discipline on mathematicians who are forced to work essentially only with internal functions and objects using the infinitesimals in the non-standard universe, i.e. eliminating the temptation to investigate the relationship between the standard and nonstandard objects via various lifting techniques. The disadvantage of the approach, in the view of the author, lies in the confusing change of the perspective where objects that used to be finite suddenly become "infinite". For example, natural numbers need to be called as bounded and unbounded in the IST instead of the more intuitive finite and hyperfinite in order to avoid a possible confusion. The key references where the IST infinitesimal stochastic calculus is well developed are Nelson (1987), Berg (2000), or Herzberg (2013).

The third NSA set theoretical framework we want to mention is the Alternative Set Theory of P. Vopěnka (1979). The theory can be compared to IST but it goes much deeper in the effort to study the phenomena of vagueness, indiscernibility, and natural infinity. This means that one has to leave the universe of classical sets and enter a much more restricted universe of sets and classes in AST. Indeed, the theory is weaker than ZFC (it has the same strength as the 3rd order Peano arithmetics — see Pudlák and Sochor, 1984). There are some foundations of the probability theory laid down in AST (Kalina, 1989) but not much else has been done in further development of the stochastic analysis within AST.

Our approach will be a compromise between the Robinsonian NSA and the Nelsonian IST frameworks. Starting from the known world, our key (axiomatic) assumption is that the set of reals R (and the structures built on R) can be consistently extended with infinitesimals. At the same time we restrict ourselves from the analysis of relationships between $V(R)$ and $V({}^*R)$. We are going to use the language of IST and focus our attention to internal sets and infinitesimal structures. Our motivation is instrumental: we want to build useful financial models with infinitesimals, i.e. models that allow us to price and analyze financial instruments. All we need to take care of is that the models are based on plausible assumptions and that the final results are unique (up to an infinitesimal error). It turns out that it is quite easy to prove existence of stochastic processes with certain properties simply performing a "Monte Carlo" simulation with infinitesimal time steps. But it is more difficult to verify their uniqueness in terms of the distributional properties or results we are interested in.

In the next section, we are going to explain the elementary one-step binomial tree argument that we claim to be (almost) sufficient to develop the derivatives pricing theory. In Section 3, we provide a brief overview of the non-standard analysis foundations. The stochastic processes on hyperfinite binomial trees will be introduced in Section 4. The main results, change of measure and Itô's lemma in one and multidimensional case will be given in Sections 5 and 6.

2. One step binomial tree

Let us consider a security with price S, a derivative price f contingent on S, and another positively valued security used as a general discount factor $g > 0$ (also called a *numeraire*). We know the initial prices S_0 and g_0 but not the derivative price f_0. In order to value the derivative we need to model the future behavior of the underlying asset price and of the discount factor. The basic model is very simple: we consider only two scenarios, "up" and "down", over the time horizon $[0, T]$. We assume to know the values of S_T and g_T in the two scenarios, i.e.

S_u, g_u and S_d, g_d but also the derivative values f_u and f_d, (see Figure 1). We assume, for simplicity, that the securities do not pay any income (e.g. dividends or coupons) during the time interval $[0, T]$. The values f_T are equal to the known payoff if the derivative is of European type with maturity T. For example, in case of a European call option, $f_T = \max(S_T - K, 0)$ where K is the strike price. Regarding the discount factor, in the simplest case assuming constant interest rates, it can be defined as $g_0 = e^{-rT}$ and $g_u = g_d = 1$. However, this set-up allows to relax the assumption of constant interest rates and use, for example, the zero coupon bond $P(t, T_M)$ with $T_M > T$ as a discount factor, i.e. $g_0 = P(0, T_M)$ and $g_T = P(T, T_M)$. Note that in this case it is not necessary that $g_u = g_d$.

The core of the derivative valuation arguments lies in the idea of *replication*. It is based on the assumption that the market is arbitrage-free. An *arbitrage*, generally speaking, is a combination of transactions, or a strategy, so that the arbitrageur starts with zero and ends up with a positive profit with a non-zero probability and with no possibility of loss. We also implicitly assume that the securities can be shorted without any limitation and that there are no transaction costs and bid-ask spreads.

Lemma 2.1. *Let f be the price process of a derivative security contingent on an underlying asset S defined on the one-step binomial tree and let g be a numeraire so that S does not linearly depend on g. Then, assuming that the market is arbitrage-free, there is a portfolio Π combining the underlying asset and the discount security with weights α and β defined at time 0 so that it replicates the values of the contingent derivative at time T, i.e. $\Pi_T = f_T$.*

Proof: We just need to solve a set of two equations with two unknowns:

$$\begin{aligned} f_u &= \alpha S_u + \beta g_u \\ f_d &= \alpha S_d + \beta g_d. \end{aligned}$$

This equation has a solution if (S_u, S_d) and (g_u, g_d) are not co-linear which is our assumption. Now, the value of the replication portfolio Π_0 must be equal to f_0, i.e. the initial derivative value can be calculated as

$$f_0 = \alpha S_0 + \beta g_0.$$

This is not a pure mathematical statement but a consequence of the assumption that the market is arbitrage-free. By contradiction, if $\Pi_0 < f_0$ then the arbitrageur could short f and invest into Π with the risk-less profit $f_0 - \Pi_0$ plus accrued interested collected at T. Similarly, we get a contradiction if $\Pi_0 > f_0$. $\square$

The replication argument can be used to define certain artificial probabilities of the up and down scenarios known as the *risk-neutral probabilities* that are of paramount importance in derivative pricing.

Lemma 2.2. *Under the assumptions of Lemma 2.1 there is a unique up-branching probability q that makes the discounted underlying value $Z = S/g$ a martingale.*

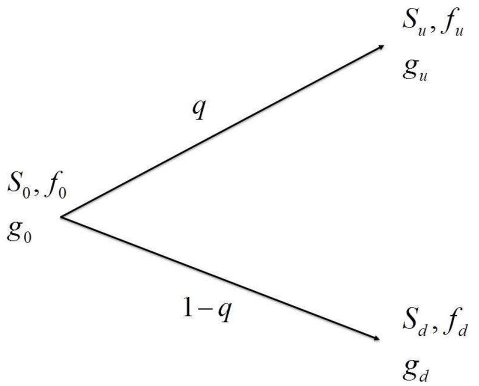

FIGURE 1. One step binomial tree

Under this probability measure, the discounted derivative value f/g will be a martingale as well.

PROOF: In the one-step binomial tree set-up we simply require that $Z_0 = E_Q[Z_T]$ where the expectation is with respect to the probability measure given by q. This means to solve one equation with one unknown

$$Z_0 = qZ_u + (1-q)Z_d$$

which has a unique solution again if S and g are not co-linear. The key result is that this probability also makes the ratio f/g a martingale. Indeed, let us consider the replication portfolio $\Pi_T = f_t$, then

$$E_Q\left[\frac{f_T}{g_T}\right] = E_Q\left[\alpha\frac{S_T}{g_T} + \beta\frac{g_T}{g_T}\right] = \alpha\frac{S_0}{g_0} + \beta\frac{g_0}{g_0} = \frac{f_0}{g_0}.$$

□

Therefore, given the risk-neutral probability q obtained from Lemma 2.2, we can express $f_0 = g_0 E_Q[f_T/g_T]$ without explicitly calculating the coefficients α and β.

Moreover, the risk neutral probabilities allow us to generalize the arguments for a general N-step binomial tree with a time step $\Delta t = T/N$. It is true that one can repeat the replication argument on the individual one-step binomial trees going backward from the known payoff value at time T to the time 0. However, it is much more efficient, given a numeraire g, to define (calculate) the branching risk-neutral probabilities for each one-step tree, and then multiply through the branching probabilities to obtain the probability $q(\omega)$ for every path ω of the tree assuming independence of the branchings. Therefore, according to the result above and the principle of iterated expectations, $Z = S/g$ as well as f/g become martingales with respect to this probability measure Q, i.e. $Z_t = E_Q[Z_T]$ for any $0 \leq t < T$ and the derivative can be valued using the risk-neutral expectation $f_0 = g_0 E_Q[f_T/g_T]$.

Letting N go into infinity, or simply being infinite and Δt infinitesimal, we can pass, at least intuitively, to a continuous time model where the martingale property holds. Hence, the first task is to make this step more rigorous. Another issue is the problem of calibration, i.e., how to define the up and down values in order to reflect the empirical reality and in particular, how to change the calibration when N is becoming large. Empirical research shows that the expected (log) return and variance of returns over a small time period of length Δt are both proportionate to Δt. This evidence leads to the class of diffusion type processes where $E[\ln S_{t+\Delta t}/S_t] = \mu \Delta t$ and $\mathrm{var}[\ln S_{t+\Delta t}/S_t] = \sigma^2 \Delta t$. An important result is that when we start from a binomial tree with an objective (real-world) measure then by changing the probability based on a diffusion type numeraire we do change the expected return (μ) but do not asymptotically change the variance, i.e. the change of $\mathrm{var}[\ln S_{t+\Delta t}/S_t]$ is negligible with respect to Δt. Equivalently, the volatility σ remains the same (in continuous time) when we change a (diffusion type) numeraire to another one. This result has a practical consequence: in order to value a derivative, all we need are volatilities estimated from the real-world empirical data. The real-world volatilities do not change with respect to the risk-neutral probability measure and with appropriate distributional assumptions we have a good chance to evaluate the risk-neutral expectation. The volatility asymptotic invariance result could be obtained in the discrete set-up, however, the tedious calculations can be quite simplified when we work with infinitesimals allowing us to neglect terms of infinitely smaller orders. Last but not least we need to check that our results are unique, in particular independent on infinitesimal changes of the model parameters.

3. Nonstandard analysis — an overview

For a detailed treatment of NSA see, for example, Hurd and Loeb (1985) or the elementary textbook by Keisler (2000). However, let us review the basic principles.

There is (in the universe of sets with the Axiom of Choice) an extension ${}^*R$ of the real line R with a number of properties outlined below. The extension ${}^*R$ includes elements defined as non-zero "infinitesimals" ($x \in^* R$ satisfies $|x| < \epsilon$ for all $\epsilon > 0$ in R) and their "infinite" multiplicative inverses.

Let us give a few useful basic definitions. We say that x and y in ${}^*R$ are infinitesimally close $x \approx y$ if $x - y$ is infinitesimal; x is larger but not infinitesimally close to y, i.e. $x \gg y$, if $x > y$ and $x \not\approx y$. We say that $x \in {}^*R$ is finite if $|x| < n$ for some $n \in N$. It can be easily shown that for every finite $x \in {}^*R$ there is unique number $r \in R$ such that $x \approx r$. This number, which can be viewed as a result of infinitesimal rounding to the nearest standard number, is called the *standard part* of x, and denoted ${}^o x = st(x)$. We say that a number x is of order $o(y)$ if $y \neq 0$ and $\frac{x}{y}$ is infinitesimal. We say that x is of order $O(y)$ if $y \neq 0$ and $\frac{x}{y}$ is finite.

The extension itself is not unique: ${}^*R$ can be, for example, defined as an ultrapower $R^N/\mathcal{U}$ of the reals by a non-principal ultrafilter $\mathcal{U}$ on N.[1] What is important here is that the arithmetical operations and order relations valid in R are extended to ${}^*R$, i.e., the tuple $({}^*R; +; \times; <)$ is an ordered field.

For any set S, the superstructure $V(S)$ over S is defined as $V(S) = \bigcup_{n=0}^{\infty} V_n(S)$ where $V_0(S) = S$ and

$$V_{n+1}(S) = V_n(S) \cup P(V_n(S)),\ (n \in N).$$

This construction can be applied to R as well as to ${}^*R$. The superstructures $V = V(R)$ and $V({}^*R)$ are connected by a mapping $* : V(R) \to V({}^*R)$ which associates to each standard object $A \in V(R)$ its extension ${}^*A \in V({}^*R)$. The nonstandard universe (whose members are known as internal sets) is given by

$${}^*V = \{x : x \in {}^*A \text{ for some } A \in V\} = \bigcup_{n=0}^{\infty} {}^*V_n(R).$$

The Transfer Principle states that any bounded quantifier statement[2] holds in V iff it holds in ${}^*V$. This principle enables us to "switch" from the "standard world" V to *internal* objects (elements of ${}^*V$) in "the non-standard world" and back again. In a proof, we can therefore "translate" a statement into the language of internal sets, manipulate it within ${}^*V$ and then translate the results back into the context of V. Note that not all sets in ${}^*V(R)$ are internal. For example, R itself cannot be internal since it is a bounded subset of ${}^*R$, yet it does not have any supremum. The sets that are not internal are called external. The transfer principle cannot be applied to external sets.

For a development of nonstandard probability theory the ω_1-saturation principle is also often required to hold: if $(A_n)_{n\in N}$ is a decreasing sequence of nonempty internal sets then their intersection is nonempty as well, $\bigcap_{n\in N} A_n \neq \emptyset$. Regarding infinitesimal analysis we will follow the basic notation used by Robinson (1961) but will try to recall the definitions and basic theorems whenever needed.

The advantage of NSA is that objects can be built using infinitesimally small or infinitely large numbers. The objects can then be lifted back to the "normal" mathematical world if needed. Many papers have been devoted to various lifting techniques. However, as discussed in the introduction, this effort might be considered as an anachronism from the perspective of applications or our understanding of the real world. If our objects constructed in ${}^*V(R)$ provide a reasonable and applicable model of the investigated reality then why should we lift it back to the standard universe $V(R)$? Hence, we will focus more on understanding and

---

[1]A non-principal ultrafilter on N is a collection of subsets of N closed under intersections and supersets, containing no finite sets, and such that for any $A \subseteq N$, either A or $N - A$ belongs to U. The existence of such ultrafilters is a consequence of the Axiom of Choice.

[2]A bounded quantifier statement is a mathematical statement which can be written so that all quantifiers range over prescribed sets. This includes most statements used in practice.

applicability of our constructions and will not care too much whether the results and objects can be translated back into the fully standard set-up.

4. Infinitesimal stochastic calculus

We are going to build stochastic processes on *hyperfinite binomial trees*. Let $T > 0$ be a given (maturity) time and $H \in {}^*N - N$ an infinite integer. We set $\delta t = \frac{T}{H}$ to be our *elementary infinitesimal time step* and $\mathbf{T}_0 = \{0, \delta t, 2\,\delta t, \ldots, H\,\delta t = T\}$ or $\mathbf{T} = \mathbf{T}_0 - \{0\}$ to be our *hyperfinite time line*. We reserve the notation δt for this particular elementary infinitesimal time step. On the other hand, by dt will always denote a general infinitesimal time step, an integer multiple of δt, possibly δt, but sometimes its infinite multiple, yet still infinitesimally small. While in NSA literature the stochastic processes are built on the space $\{-1,+1\}^{\mathbf{T}}$ of sequences of ± 1 of lengths H throughout the hyperfinite binomial tree, we will rather use the set

$$\mathcal{T} = \{-1,+1\}^{\leq T} = \bigcup_{t \in \mathbf{T}_0} \{-1,+1\}^{\mathbf{T} \cap (0,t]}$$

corresponding to individual nodes of the tree including the root $\emptyset$. For $\omega \in \mathcal{T}$ we denote $t(\omega)$ to be the length of ω, i.e. $t(\omega) = \max(\mathrm{dom}(\omega)) \in \mathbf{T}$ if ω is non-empty, and 0 if ω is empty (i.e., if it is the root). The nodes with length $t(\omega) < T$ always branch into two subsequent nodes[3] $\omega^\frown\{+1\}$ and $\omega^\frown\{-1\}$. For $t \in \mathbf{T}$ it will be useful to set

$$\mathcal{T}(t) = \{\omega \in \mathcal{T} : t(\omega) = t\} \text{ and}$$
$$\mathcal{T}(< t) = \{\omega \in \mathcal{T} : t(\omega) < t\}.$$

If $\tau \leq t(\omega)$ then we denote $\omega \restriction \tau = \omega \restriction (\mathbf{T} \cap (0,\tau])$ to be the truncated path of length $\tau \in \mathbf{T}$. This notation allows us to eliminate completely the notion of filtration used, for example, in the "radically elementary" treatment of Nelson (1987) or Herzberg (2013). Since functions can be generally built recursively on finite trees we can do the same, applying the transfer principle, on hyperfinite trees but restricting ourselves only to internal sets and functions throughout the construction.

In general, we are going to study *stochastic processes* represented by internal functions $X : \mathcal{T} \to {}^*R$, i.e. in the form of hyperfinite binomial trees. For $\omega \in \mathcal{T}$ the value $X(\omega)$ naturally depends only on the information encoded in ω up to the time $t(\omega)$ and we do not have to bother with filtrations to take care of the issue of anticipation. Sometimes, for $\tau \leq t(\omega)$, we will use the notation $X(\omega,\tau)$ instead of $X(\omega \restriction \tau)$. If $\omega \in \mathcal{T}(T)$ then $\langle X(\omega,t) \,|\, t \in \mathbf{T}\rangle$ is the realization of the stochastic process X corresponding to the path ω. On the other hand, for $t \in \mathbf{T}$ we denote $X(t)$ or X_t to be the random variable $X(\omega,t)$ with $\omega \in \mathcal{T}(t)$.

---

[3]Formally, we define the two subsequent nodes of length $t + \delta t$ as follows: $\omega^\frown\{j\}(s) = \omega(s)$ for $s \leq t, s \in \mathbf{T}$ and $\omega^\frown\{j\}(t + \delta t) = j$ with $j = +1, -1$.

Of course, we are interested mainly in processes that are generated by a law applied step-by-step throughout a binomial tree. Building a stochastic process we start with an initial value assigned to the root $X(\emptyset) = X_0$. Then, for each already calculated $X(\omega)$ for a path ω of length smaller than T, we define $X(\omega^\frown\{+1\}) = X_u(\omega)$ and $X(\omega^\frown\{-1\}) = X_d(\omega)$ on the subsequent nodes. Hence, the two values $+1$ and -1 correspond to a coin tossing random element. The "up" and "down" values $X_u(\omega)$ and $X_d(\omega)$ are (internal) functions of the values of X along ω or of values of other stochastic processes up to time $t = t(\omega)$.

The most well-known process is the Brownian motion first constructed using NSA infinitesimals by Anderson (1976). The original idea of Brown (1828) as well as the first applications of Bachelier (1900) or Einstein (1956) were in fact based on the intuitive notion of infinitesimals. Set $Z(\emptyset) = 0$ and

$$Z(\omega^\frown\{j\}) = Z(\omega) + j\sqrt{\delta t} \tag{4.1}$$

where $j = +1, -1$. Therefore,

$$Z(\omega) = \sum_{s\in\mathbf{T}\cap(0,t(\omega)]} \omega(s)\sqrt{\delta t}.$$

It will be useful to denote δZ to be the two-value function $\delta Z(j) = j\sqrt{\delta t}$ for $j = +1, -1$, or a binomial random variable if the up and down branching probabilities are given. Then any stochastic process $X : \mathcal{T} \to {}^*R$ can be, in fact, expressed by the equation

$$\begin{gathered} \delta X(\omega) = a(\omega)\delta t + b(\omega)\delta Z \quad \text{where} \\ \delta X = X(\omega^\frown\{j\}) - X(\omega) \end{gathered}$$

describing how to generate its values from the initial one $X(\emptyset) = X_0$. The (internal) functions $a, b : \mathcal{T}(< T) \to {}^*R$ can be obtained solving for every ω the following two equations with the two unknowns:

$$\begin{aligned} X(\omega^\frown\{+1\}) - X(\omega) &= a(\omega)\delta t + b(\omega)\sqrt{\delta t} \quad \text{and} \\ X(\omega^\frown\{-1\}) - X(\omega) &= a(\omega)\delta t - b(\omega)\sqrt{\delta t}. \end{aligned} \tag{4.2}$$

So far we did not need any probability measure, but in order to characterize the distribution of the processes Z or X we have to introduce one. The advantage of the hyperfinite probability theory is that any *hyperfinite probability measure* P on a hyperfinite set Ω is given by elementary probabilities $p(\omega)$ on the individual elementary events $\omega \in \Omega$. The probability measure is then in a straightforward manner defined on all internal subsets of Ω and can be extended using Caratheodory theorem and the ω_1-saturation principle (Loeb, 1975) to a standard measure denoted $L(P)$ which is σ-complete and takes standard values in R_0^+.

In the case of hyperfinite binomial trees, in line with our intuition, it is enough to specify the branching probabilities on $\mathcal{T}(< T)$. That is, if an internal function specifies for every $\omega \in \mathcal{T}(< T)$ the probability $p_{+1}(\omega) \in (0,1)$ of the process to go up, and if we set $p_{-1}(\omega) = 1 - p_{+1}(\omega)$ to be the complementary down-branching probability (analogously to Figure 1), then we can define the elementary probability for every $\omega \in \mathcal{T}$ with $t = t(\omega)$ setting

$$p(\omega) = \prod_{s \in \mathbf{T} \cap (0,t)} p_{\omega(s)}(\omega \restriction s).$$

Equivalently, the measure can be specified by an internal function $p : \mathcal{T}(T) \to {}^*R^+$ such that $\sum_{\omega \in \mathcal{T}(T)} p(\omega) = 1$. Then, for $t < T$, $t \in \mathbf{T}$ we can extend the probability function on any $\omega \in \mathcal{T}(t)$ setting

$$p(\omega) = \sum_{\omega' \in \mathcal{T}(T),\ \omega' \restriction t = \omega} p(\omega').$$

The corresponding branching probabilities can be then defined by the equation

$$p_{+1}(\omega) = \frac{p(\omega^\frown\{+1\}}{p(\omega)}$$

for $\omega \in \mathcal{T}(< T)$. Consequently, any measure on $\mathcal{T}$ is given by a branching probability function.

This approach allows us to define easily the concepts of *conditional expectation* and of a *martingale*. If $t < T$ and $\omega \in \mathcal{T}(t)$ then the conditional expectation is defined as

$$E[X(T) \,|\, \omega] = \frac{1}{p(\omega)} \sum_{\omega' \in \mathcal{T}(T),\ \omega' \restriction t = \omega} p(\omega') X(\omega').$$

By the conditional expectation $E_t[X(T)] = E[X(T) \,|\, t]$ we mean the random variable defined on $\mathcal{T}(t)$ that assigns $E[X(T) \,|\, \omega]$ to every $\omega \in \mathcal{T}(t)$. We say that X is a martingale if $X(\omega) \approx E[X(T) \,|\, \omega]$ for every $t < T$ and $\omega \in \mathcal{T}(t)$.

The basic measure on $\mathcal{T}$ is the *uniform counting measure* corresponding to constant branching probabilities all equal to $\frac{1}{2}$. Then $p(\omega) = 2^{-H} = \frac{1}{|\mathcal{T}(T)|}$ for every $\omega \in \mathcal{T}(T)$. It can be shown (Anderson, 1975) that with respect to this measure the process Z defined by (4.1) has the properties of a Brownian motion, or more precisely that the standardized process $z(\omega, {}^o t) = {}^o Z(\omega \restriction t)$, $\omega \in \mathcal{T}(T)$ satisfies the classical distributional conditions put on the Brownian motion[4]. If $a(\omega)$ is defined by (4.2) then it is easy to show that a sufficient condition for X to be a martingale with respect to the uniform measure is that $a(\omega) \approx 0$ for all $\omega \in \mathcal{T}(< T)$.

The key property we need to verify is that the increments $Z(\omega_1) - Z(\omega_0)$ are normally distributed, where ω_0 is fixed and ω_1 is such that $\omega_1 \restriction t(\omega_0) = \omega_0$. This

[4]We say that a function $F : \mathbf{T} \to {}^*R$ is $\mathcal{S}$-continuous if $F(x) \approx F(y)$ whenever $x \approx y$. It can be shown that for almost all $\omega \in \mathcal{T}(T)$ with respect to the Loeb measure the functions $Z(\omega, \cdot)$ are $\mathcal{S}$-continuous. Hence for almost all ω the definition of $z(\omega, t)$ is correct.

easily follows from a hyperfinite version of the Central Limit Theorem (see Cutland (1991)) which will also serve as our key tool. The theorem is obtained simply applying the transfer principle to the classical Central Limit Theorem (CLT).

Theorem 4.1. *Let $\{X_n : n \leq M\}$ with $M \in {}^*N - N$ be an internal sequence of $*$-independent random variables on a hyperfinite probability space $\{\Omega, P\}$ with a common standard distribution function F that have mean 0 and the standard deviation 1. Then the variable*

$$X = \frac{1}{\sqrt{M}} \sum_{n=1}^{M} X_n$$

is normally distributed in the sense that $P[X \leq x] \approx {}^*\psi(x)$ for any $x \in {}^*R$ where

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx$$

is the cumulative standardized normal distribution function.

In case of the (Anderson's) Brownian motion the theorem can be directly applied to the independent incremental random variables $X_n = +1$ or -1 with probabilities $\frac{1}{2}$. It follows that the variable Z on $\mathcal{T}(T)$ which can be expressed as

$$Z = \sum_{n=1}^{H} \sqrt{\delta t} X_n = \sqrt{T} \cdot \frac{1}{\sqrt{H}} \sum_{n=1}^{H} X_n$$

has the normal distribution (up to an infinitesimal error) with mean 0 and variance T. Moreover, for a given infinitesimal $dt = N \cdot \delta t$, being an infinite multiple of δt, and for a given $\omega \in \mathcal{T}(< T)$ the increment dZ, where we set $dZ = dZ(\omega') = Z(\omega^\frown\omega') - Z(\omega)$ for ω' of length dt, is also normally distributed with mean 0 and variance dt but in this case up to an infinitesimal error of order $o(dt)$. This allows us to interpret, for a, b finite, the *classical stochastic differential equation* (SDE) in the form

$$dX = a\, dt + b\, dZ$$

as an equivalent of saying that dX is a normally distributed random variable with mean $a\, dt$ and variance $b^2\, dt$ up to an infinitesimal error of order $o(dt)$.

We say that X is an *Itô's process* if it is defined on the elementary time step level by the equation $\delta X = a\, \delta t + b\, \delta Z$, where a and b are internal functions on $\mathcal{T}(< T)$. In order to obtain in this way reasonable processes we need to set certain additional conditions on a and b, e.g. requiring that a is finite almost surely (a.s.) with respect to the uniform Loeb measure, and b is square S-integrable. We say that $b : \mathcal{T}(< T) \to {}^*R$ is *square S-integrable* if $E[\sum_{\mathbf{T}} b^2 \delta t]$ is finite and if $E[\sum_{\mathbf{T}} b^2 \delta t \cdot 1_A] \approx 0$ for every $A \subseteq \mathcal{T}(T)$ with $P(A) \approx 0$. It is useful to prove the following lemma which says that when building a stochastic process step by step, errors of order $o(\delta t)$ do not matter.

Lemma 4.2. *Let $X_i : \mathcal{T} \to {}^*R$, $i = 1, 2$ be two Itô's stochastic processes defined by $\delta X_i = a_i(\omega)\delta t + b_i(\omega)\delta Z$ such that $X_1(0) \approx X_2(0)$, $a_1 \approx a_2$ a.s., $b = b_1 - b_2$ is square S-integrable, and $b \approx 0$ a.s. with respect to the uniform Loeb measure. Then $X_1 \approx X_2$ a.s. Moreover, if a_i is finite a.s. and if b_i is square S-integrable then for every $u \le T$ the value $X_i(u)$ is finite a.s.*

Remark 4.3. If a_i and b_i are as above, i.e. $a_1 \approx a_2$ and $b_1 \approx b_2$ a.s. with respect to the uniform Loeb measure, we say that the right hand side of $\delta X_1 = a_1(\omega)\delta t + b_1(\omega)\delta Z$ differs from the right hand side of $\delta X_2 = a_2(\omega)\delta t + b_2(\omega)\delta Z$ by an error of order $o(\delta t)$.

Proof: Let $u \le T$ and

$$\begin{aligned} Y(u) &= X_2(u) - X_1(u) \\ &= (X_2(0) - X_1(0)) + \sum_{t<u}(a_2 - a_1)\delta t + \sum_{t<u}(b_2 - b_1)\delta Z. \end{aligned} \tag{4.3}$$

In order to prove that $Y(u) \approx 0$ a.s., it is enough to show that the first sum and the second sum are both infinitesimal a.s. Let $\epsilon \gg 0$ and $N \subseteq \mathcal{T}(T)$ be internal such that $P(N) < \epsilon$ and $a_1(\omega, t) \approx a_2(\omega, t)$ for all $\omega \in \mathcal{T}(T)\backslash N$ and $t \le T$. Then

$$c_1 = \max_{\omega \in \mathcal{T}(T)\backslash N, t \le T} |a_2(\omega, t) - a_1(\omega, t)| \approx 0,$$

and so $|\sum_{t<u}(a_2 - a_1)\delta t| \le c_1 u \approx 0$ on $\mathcal{T}(T)\backslash N$. Since $\epsilon \gg 0$ is arbitrary this proves that the first sum in (4.3) is infinitesimal a.s.

Regarding the second sum in (4.3), let $M(u) = \sum_{t<u} b\delta Z$. Note that M is a martingale, the increments $b(t_1)\delta Z(t_1)$ and $b(t_2)\delta Z(t_2)$ are independent for $t_1 < t_2$, and $\delta Z^2 = \delta t$. Therefore, we have

$$\begin{aligned} \operatorname{var}[M(u)] &= E\left[M(u)^2\right] = E\left[\left(\sum_{t<u} b\delta Z\right)^2\right] \\ &= \sum_{t_1 \le t_2 < u} E\left[b(t_1)\delta Z(t_1) b(t_2)\delta Z(t_2)\right] = E\left[\sum_{t<u} b^2\delta t\right]. \end{aligned} \tag{4.4}$$

Let $\epsilon \gg 0$ and $A = \{\omega \in \mathcal{T}(T); \max_{t<T} b(\omega, t) > \epsilon\}$ then $P(A) \approx 0$ as $b \approx 0$ a.s., and so $E[\sum_{t<u} b^2\delta t \cdot 1_A] \approx 0$ by square S-integrability of b. It follows from the definition of A that $E\left[\sum_{t<u} b^2\delta t \cdot 1_{\mathcal{T}(T)\backslash A}\right] \le \epsilon^2 u$. Therefore $\operatorname{var}[M(u)] \le 2\epsilon^2 u$ for any ϵ, and so $\operatorname{var}[M(u)] \approx 0$ proving that $M(u) \approx 0$ a.s. on $\mathcal{T}(T)$.

An analogous argument can be used to show that X_i is finite a.s. if a_i is finite a.s. and b_i is square S-integrable. Let $\epsilon \gg 0$ and find $N \subseteq \mathcal{T}(T)$ internal such that $P(N) < \epsilon$, and a_i is finite on $\mathcal{T}(T)\backslash N$. Set $c = \max_{\omega \in \mathcal{T}(T)\backslash N, t \le T} |a_i(\omega, t)| < \infty$. Then the first sum $|\sum_{t<u} a_i\delta t| \le cu$ is finite on $\mathcal{T}(T)\backslash N$, and so it is finite on $\mathcal{T}(T)$ a.s. since $\epsilon \gg 0$ was arbitrary. Regarding the second sum we again use

the identity $\mathrm{var}[\sum_{t<u} b_i \delta Z] = E[\sum_{t<u} b_i^2 \delta t] < \infty$ to see that $\sum_{t<u} b_i \delta Z$ must be finite a.s. $\square$

As a corollary, two stochastic processes X_1 and X_2 satisfying the assumptions in the first part of the lemma are *nearly equivalent* in the sense of Nelson (1987) and Herzberg (2013). The definition may appear a little bit abstract: the processes are called *nearly equivalent* if $E[F(X_1)] \approx E[F(X_2)]$ for all continuous functionals on the set of paths of X_1 and X_2. A continuous functional is a map $F : \Lambda \to *R$ taking finite values where $\Lambda \subseteq *R^{\mathbf{T}}$ and $F(\lambda) \approx F(\xi)$ whenever $\lambda(t) \approx \xi(t)$ for all $t \in \mathbf{T}$. However, this requirement has a very practical financial interpretation since our aim is to value derivatives by evaluation of expected payoffs and a payoff function is generally a path dependent functional satisfying reasonable properties like continuity and finiteness. Therefore, two nearly equivalent processes yield the same expectations and derivative values in line with our instrumental philosophy.

5. Change of measure

Let us firstly investigate what happens with the mean, variance, and the probability distribution of the Brownian motion Z when the uniform branching probability is changed by an infinitesimal quantity.

Lemma 5.1. *If $p_{+1} = \frac{1}{2}$ is changed to $\frac{1}{2} + \delta p$ for a $\delta p \approx 0$ then the expected value and variance of δZ under the new measure Q are*

$$E_Q[\delta Z] = 2\delta p\sqrt{\delta t},$$

and $\mathrm{var}_Q[\delta Z] = \delta t(1 - 4\delta p^2)$.

Proof: The result is obtained by elementary calculations using the changed probabilities:

$$E_Q[\delta Z] = \left(\frac{1}{2} + \delta p\right)\sqrt{\delta t} + \left(\frac{1}{2} - \delta p\right)\left(-\sqrt{\delta t}\right) = 2\delta p\sqrt{\delta t},$$

and

$$\begin{aligned}\mathrm{var}_Q[\delta Z] &= E_Q[\delta Z^2] - E_Q[\delta Z]^2 \\ &= \left(\frac{1}{2} + \delta p\right)\delta t + \left(\frac{1}{2} - \delta p\right)\delta t - 4\delta p^2\,\delta t = \delta t\left(1 - 4\delta p^2\right).\end{aligned}$$

$\square$

If δp is constant then the random variable $Z(T) - Z(0)$ can be expressed as $\sum_{n=0}^{H-1} \delta Z_n$ where δZ_n are independent and all have the same distribution with the up-value $+\sqrt{\delta t}$ attained with the probability $\frac{1}{2} + \delta p$ and the down-value $-\sqrt{\delta t}$ attained with the probability $\frac{1}{2} - \delta p$. Consequently,

$$E_Q[Z(T) - Z(0)] = H\,2\,\delta p\sqrt{\delta t} = T\,\frac{2\delta p}{\sqrt{\delta t}}$$

and

$$\mathrm{var}_Q[Z(T) - Z(0)] = T(1 - 4\delta p^2).$$

Since we restrict ourselves only to processes with finite mean we need $\frac{2\delta p}{\sqrt{\delta t}}$ to be finite, i.e. $\delta p = \frac{\alpha}{2}\sqrt{\delta t}$ for a finite α. In other words, only changes of order $\sqrt{\delta t}$ are admissible.

Corollary 5.2. *If α is finite and the up-branching probability is changed throughout the tree $\mathcal{T}$ by the constant $\delta p = \frac{\alpha}{2}\sqrt{\delta t}$ defining a new measure Q, then $E_Q[Z(T) - Z(0)] = \alpha T$ and $\mathrm{var}_Q[Z(T) - Z(0)] = T(1 - \alpha^2 \delta t) \approx T$. It means that the expected value of $Z(T)$ can be changed to an arbitrary finite number while the variance remains unchanged up to an infinitesimal error. Moreover, the distribution of $Z(T)$ remains normal up to an infinitesimal error.*

Proof: The first part of the corollary follows immediately from Lemma 5.1 and the argument above. Preservation of the normal distribution property follows from the CLT. Let

$$X_n = \frac{\delta Z_n - \alpha \delta t}{\sqrt{\delta t}},$$

then X_n are independent with the same distribution with mean 0 and variance 1. Consequently, we may apply the hyperfinite central limit theorem to

$$\frac{1}{\sqrt{H}} \sum_{n=1}^{H} X_n = \frac{Z(T) - Z(0) - \alpha T}{\sqrt{T}}.$$

Therefore, we have proved that $Z(T) - Z(0)$ is a normally distributed variable (up to an infinitesimal error) with mean αT and variance T. □

The same analysis applies to the differential dZ for an infinitesimal dt that is at the same time an infinite multiple of δt. The Wiener process under the new measure Q is defined by adjusting the mean of Z, i.e. $\delta \tilde{Z} = \delta Z - \alpha \delta t$ or equivalently $\tilde{Z} = Z - \alpha T$. Then Z, representing the Wiener process under the original uniform measure, with respect to the changed measure satisfies the classical stochastic differential equation $dZ = \alpha\, dt + d\tilde{Z}$ in the sense that dZ is normally distributed with mean $\alpha\, dt$ and variance dt up to an error of order $o(dt)$. Note that we have got at least two different representations of the (real-world) process satisfying this equation (having infinitesimal increments with mean $\alpha\, dt$ and variance dt): first, the process $\tilde{Z}$ with drift α defined on $\mathcal{T}$ with the uniform measure P, and then the original Wiener process Z on $\mathcal{T}$ with the changed measure Q.

On the other hand, given a process X following the SDE $\delta X = a\, \delta t + b\, \delta Z$ with a constant a and b on $\mathcal{T}$ with the uniform measure, we may change the measure (as above with the coefficient $\alpha = -a/b$) so that X becomes a martingale (has zero drift) with respect to the changed measure. To show that this is a special case of an equivalent martingale measure we yet need to show that the changed measure is equivalent to the uniform measure on $\mathcal{T}$.[5]

The density of the changed measure Q with respect to the uniform measure P can be for an $\omega \in \mathcal{T}(t)$ expressed as

$$\frac{Q(\omega)}{P(\omega)} = \prod_{s \in \mathrm{dom}(\omega)} \left(1 + \alpha\omega(s)\sqrt{\delta t}\right).$$

To understand the expression on the right hand side we need the following (see also Cutland at al. (1991)).

Lemma 5.3. *For any $\omega \in \mathcal{T}(t)$ such that $|Z(\omega)|$ is finite and for any finite constants α and β*

$$\prod_{s \in dom(\omega)} (1 + \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t) \approx \exp\left(\left(\beta - \frac{\alpha^2}{2}\right) t + \alpha Z(\omega)\right). \tag{5.1}$$

Proof: If we apply the logarithm to the left-hand side of the equation then

$$\begin{aligned} &\ln\left(\prod_{s \in \mathrm{dom}(\omega)} \left(1 + \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t\right)\right) \\ &= \sum_{s \in \mathrm{dom}(\omega)} \ln\left(1 + \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t\right). \end{aligned} \tag{5.2}$$

The Taylor expansion of $\ln(1+x)$ is $\ln(1+x) = x - \frac{x^2}{2} + O(x^3)$ for a sufficiently small x, hence specifically for $x = \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t$ it follows that

$$\begin{aligned} &\ln(1 + \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t) = \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t \\ &\quad - \frac{1}{2}(\alpha^2\delta t + 2\alpha\beta\omega(s)\delta t^{3/2} + \beta^2\delta t^2) + O(\delta t^{3/2}) \\ &= \alpha\,\omega(s)\sqrt{\delta t} + (\beta - \frac{1}{2}\alpha^2)\delta t + o(\delta t). \end{aligned} \tag{5.3}$$

Since the $o(\delta t)$ terms can be neglected in the sum over $\mathrm{dom}(\omega)$ we get

$$\sum_{s \in \mathrm{dom}(\omega)} \ln\left(1 + \alpha\,\omega(s)\sqrt{\delta t} + \beta\,\delta t\right) \approx \alpha\,Z(\omega) + (\beta - \frac{1}{2}\alpha^2)t.$$

Because the exponential function is $\mathcal{S}$-continuous we have proved (5.1). □

Theorem 5.4. *The changed measure Q is equivalent to the uniform measure P on $\Omega = \mathcal{T}(T)$.*

[5] Generally we say that two measures P and Q on a hyperfinite space Ω are equivalent if $P(A) \approx 0$ whenever $Q(A) \approx 0$ for any internal $A \subseteq \Omega$. This is the same as saying that the Loeb measures $L(P)$ and $L(Q)$ are equivalent in the classical sense.

PROOF: Let us assume $A \subseteq \Omega$ is such that $P(A) \gg 0$. Since $Z(\omega)$ is finite $L(P)$-almost surely there is an $n \in N$ such that $A_n = \{\omega \in A : |Z(\omega)| < n\}$ is also positive, i.e. $P(A_n) \gg 0$. It follows from Lemma 5.3 that there is a finite constant K such that $\exp(-K) \le \frac{Q(\omega)}{P(\omega)} \le \exp(K)$ for every $\omega \in A_n$. Consequently $Q(A) \ge Q(A_n) \ge \exp(-K) \cdot P(A_n) \gg 0$. Similarly, we can prove that if $Q(A) \gg 0$ then $P(A) \gg 0$ as well. □

Generally, we wish to consider the change of measure determined by varying probability adjustments on the binomial branches, i.e. by changes of $p_{+1} = \frac{1}{2}$ to $\frac{1}{2}(1 + \alpha(\omega)\sqrt{\delta t})$ where $\alpha : \mathcal{T}(< T) \to {}^*R$ is an internal function taking only (or at least $L(P)$ — almost surely) finite values. Based on the analysis above, the mean of the elementary increment δZ is adjusted to $\alpha(\omega)\delta t$ while the variance remains unchanged, i.e. δt up to an error of order $o(\delta t)$. If α is a continuous deterministic function of $t = t(\omega)$ then the analysis above can be easily generalized: $X(t) - X(0)$ is normal (up to an infinitesimal error) with variance t and the mean equal to

$$\int_0^t \alpha(s)ds = \sum_{s \in \mathbf{T} \cap [0,t)} \alpha(s)\delta t.$$

The changed measure Q is in this case again equivalent to the uniform measure P. Nevertheless, if α depends on $\omega \in \mathcal{T}(t)$ then we cannot generally conclude that the process $\delta X(\omega) = \alpha(\omega)\,\delta t + \delta Z$ leads to a normal distribution. Note that the process is well defined for any internal function α. However, we will restrict our attention only to those drift functions α resulting in realistic stochastic processes that are almost surely finite and continuous.

Let us consider the geometric Brownian motion frequently used for stock price modelling. It is described formally by the classical stochastic differential equation $dS = \mu\, S\, dt + \sigma\, S\, dZ$ where μ is the drift and σ the volatility. Note that the coefficients of dt and dZ are now stochastic (depending on S). The process can be constructed inductively on $\mathcal{T}$ with the uniform measure.

Theorem 5.5. *Let us define the stochastic process S recursively setting $S(\emptyset) = S_0 > 0$ and*

$$S(\omega^\frown\{j\}) = S(\omega) + \mu S(\omega)\delta t + \sigma S(\omega) j \sqrt{\delta t},$$

where $\mu, \sigma \ge 0$ are finite constants. Then, with respect to the uniform measure, $S(T)$ is lognormally distributed, i.e. $\ln S(T)$ is normally distributed with mean $\ln S_0 + (\mu - \frac{\sigma^2}{2})T$ and standard deviation $\sigma\sqrt{T}$, up to an infinitesimal error.

PROOF: Note that the process S remains positive as $S_0 > 0$,

$$S(\omega^\frown\{j\}) = S(\omega)(1 + \mu\,\delta t + \sigma\, j \sqrt{\delta t})$$

and $1 + \mu\,\delta t + \sigma\, j\sqrt{\delta t} > 0$ is always positive for an infinitesimal δt. To show that $S(T)$ is lognormal set $X = \ln S$. Then using the technique of the proof of

Lemma 5.3 we get that

$$X(\omega^\frown\{j\}) - X(\omega) = \left(\mu - \frac{\sigma^2}{2}\right)\delta t + \sigma j\sqrt{\delta t} + o(\delta t).$$

According to Lemma 4.2 the process X has (up to an infinitesimal error) the same distribution as the process given by $\delta\tilde{X} = (\mu - \frac{\sigma^2}{2})\delta t + \sigma\delta Z$. Consequently, $X(T) - X(0) = \ln\frac{S(T)}{S(0)}$ has the normal distribution with mean $(\mu - \frac{\sigma^2}{2})T$ and standard deviation $\sigma\sqrt{T}$. In other words, $S(T)$ is (up to an infinitesimal error) lognormally distributed. $\square$

When the uniform probability is changed with a finite constant α as above, i.e. the uniform branching probabilities are changed from $\frac{1}{2}$ to $\frac{1}{2} + \alpha\frac{\sqrt{\delta t}}{2}$, then the drift of S is adjusted to $\mu + \alpha$ but the volatility σ remains unchanged (up to an infinitesimal error). In particular, we may start with a process without any drift, $S(\emptyset) = S_0$ and $S(\omega^\frown\{j\}) = S(\omega) + \sigma S(\omega)j\sqrt{\delta t}$, and change the measure on $\mathcal{T}$ to achieve any desired finite drift μ of S by setting $\alpha = \frac{\mu}{\sigma}$ and changing the measure accordingly.

6. Equivalent martingale measure

Let us consider a more general derivative security F depending on n underlying assets corresponding to n primary sources of uncertainty $Z_1, \ldots, Z_n$ modelled as independent Brownian motions. Let us assume that F satisfies the classical stochastic differential equation of the form

$$dF = \mu F dt + \sigma_1 F dZ_1 + \cdots + \sigma_n F dZ_n. \tag{6.1}$$

We will firstly assume that the drift μ and volatility σ_i coefficients are constant and later relax the assumption to general Itô's processes.

To model n-dimensional stochastic processes, let $\mathcal{T}_1, \ldots, \mathcal{T}_n$ be n copies of the hyperfinite binomial tree on $\mathbf{T}$ and let $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n = \{\langle\omega_1, \ldots, \omega_n\rangle : \omega_1 \in \mathcal{T}_1, \ldots, \omega_n \in \mathcal{T}_n, \text{ and } t(\omega_1) = \cdots = t(\omega_n)\}$.

Given measures $P_1, \ldots, P_n$ on $\mathcal{T}_1, \ldots, \mathcal{T}_n$ we define the product measure $P = P_1 \otimes \cdots \otimes P_n$ by $P(\langle\omega_1, \ldots, \omega_n\rangle) = P_1(\omega_1)\cdots P_n(\omega_n)$. We again start with the uniform product measure corresponding to $\frac{1}{2}$ branching probabilities throughout each of the hyperfinite binomial trees.

Based on (6.1) we can build the process as follows: $F(\emptyset) = F_0 > 0$ and

$$\begin{aligned} (6.2)\quad & F(\langle\omega_1^\frown\{j_1\}, \ldots, \omega_n^\frown\{j_n\}\rangle) \\ &\qquad = F(\langle\omega_1, \ldots, \omega_n\rangle)\,(1 + \mu\delta t + \sigma_1 j_1\sqrt{\delta t} + \cdots + \sigma_n j_n\sqrt{\delta t}). \end{aligned}$$

Note that process F remains positive since multiplier on the right hand side of 6.2 is positive. Using the same technique as in Theorem 5.5 we can show that the log process $X = \ln F$ follows the SDE (up to an error of order $o(\delta t)$)

$$\delta X = \left(\mu - \frac{1}{2}\sum_i \sigma_i^2\right)\delta t + \sum_i \sigma_i \delta Z_i. \tag{6.3}$$

In fact, we have been using a specific case of Itô's lemma that can be proved in a nonstandard version (see also Anderson, 1976, Albeverio, 1986, or Herzberg, 2013).

Lemma 6.1 (Itô). *Let $G(x,t)$ be a real function of two variables with partial derivatives of all orders. Let X be an Itô's process on $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n$ following with respect to the uniform measure the stochastic differential equation $\delta X = a\delta t + \sum_{i=1}^n b_i \delta Z_i$. Then the stochastic process $G = G(X,t)$ is an Itô's process following up to an error of order $o(\delta t)$ the equation*

$$\delta G = \left(\frac{\partial G}{\partial X} a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial X^2}\sum_{n=1}^n b_i^2\right)\delta t + \frac{\partial G}{\partial X}\sum_{i=1}^n b_i \delta Z_i. \tag{6.4}$$

PROOF: Let us apply the Taylor's expansion to $\delta G = G(x+\delta x, t+\delta t) - G(x,t)$ for infinitesimal increments δx and δt :

$$\delta G = \frac{\partial G}{\partial x}\delta x + \frac{\partial G}{\partial t}\delta t + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\delta x^2 + o(|\delta x|^2 + |\delta t|).$$

Next, let us plug in the equation $\delta X = a(X,t)\delta t + \sum_{i=1}^n b_i(X,t)\delta Z_i$ defining the process X with respect to the uniform measure into the Taylor's expansion. We use the facts that $\delta Z_i^2 = \delta t$ and $\delta Z_i \delta Z_j$ for $i \neq j$ take only the two values δt and $-\delta t$ with equal probabilities, hence $\delta Z_i \delta Z_j$ has the mean 0 and variance δt^2. According to Lemma 4.2 we may neglect the deterministic elements of order less than δt and the stochastic elements with variance of order less than δt (collected within the term $o(\delta t)$). Therefore,

$$\begin{aligned}
\delta G &= a\frac{\partial G}{\partial x}\delta t + \sum_{i=1}^n \frac{\partial G}{\partial x} b_i(X,t)\delta Z_i + \frac{\partial G}{\partial t}\delta t \\
&\quad + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}(a^2\delta t + \sum_{i=1}^n b_i^2 \delta t + \sum_{i\neq j} b_i b_j \delta Z_i \delta Z_j) + o(\delta t) \\
&= \left(a\frac{\partial G}{\partial x} + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sum_{i=1}^n b_i^2\right)\delta t + \sum_{i=1}^n \frac{\partial G}{\partial x} b_i \delta Z_i + o(\delta t).
\end{aligned}$$

Consequently, in the sense of Lemma 4.2, G follows the equation (6.4). □

Now, we are going to change the up-branching probabilities along $\mathcal{T}_i$ to $\frac{1}{2} + \alpha_i \frac{\sqrt{\delta t}}{2}$ for some finite α_i. In other words we change P_i to Q_i and set $Q = Q_1 \otimes \cdots \otimes Q_n$. Then the means of the Wiener process increments are changed to

$E_Q[\delta Z_i] = \alpha_i \delta t$, while variance remains unchanged $\mathrm{var}_Q[\delta Z_i] = \delta t + o(\delta t)$ up to an infinitesimal error. If X is the log-process following 6.3 then, as δZ_i are independent,

$$E_Q[\delta X] = (\mu - \frac{1}{2}\sum_i \sigma_i^2 + \alpha_1\sigma_1 + \cdots + \alpha_n\sigma_n)\delta t, \text{ and}$$
$$\mathrm{var}_Q[\delta X] = (\sigma_1^2 + \cdots + \sigma_n^2)\delta t + o(\delta t).$$

Hence, applying Itô's lemma again $F \approx \exp(X)$ satisfies with respect to the changed measure the stochastic differential equation

$$dF = (\mu + \alpha_1\sigma_1 + \cdots + \alpha_n\sigma_n)F\,dt + \sigma_1 F\,dZ_1 + \cdots + \sigma_n F\,dZ_n.$$

We are almost ready to build up a general equivalent martingale measure. Let G be a numeraire, i.e. an (almost surely) positive Itô's process with the n sources of uncertainty satisfying the equation:

$$dG = \mu_g G\,dt + \sigma_{g,1} G\,dZ_1 + \cdots + \sigma_{g,n} G\,dZ_n.$$

Let us start with its canonical representation on $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n$ with the uniform counting measure given by: $G(\emptyset) = G_0 > 0$ and

$$G(\langle \omega_1 ^\frown \{j_1\}, \ldots, \omega_n ^\frown \{j_n\}\rangle) = G(\langle \omega_1, \ldots, \omega_n\rangle)\,(1 + \mu_g \delta t + \sigma_{g,1} j_1 \sqrt{\delta t} + \cdots + \sigma_{g,n} j_n \sqrt{\delta t}). \tag{6.5}$$

We would like to change the uniform measure to a measure Q so that for any other derivative security F with the same sources of uncertainty the ratio $\frac{F}{G}$ becomes a martingale, i.e.

$$\frac{F(t)}{G(t)} \approx E_Q\left[\frac{F(T)}{G(T)}|t\right]$$

for any $t < T$. At this point we need to use the notion of a risk-free interest rate r earned by risk-free bonds or any risk-free portfolio due to the assumption of arbitrage-free markets discussed already in Section 2. Note that the (instantaneous) risk-free interest rate r itself can be in general stochastic.

The arbitrage-free principle is used to prove that for each source of uncertainty dZ_i there is a price of risk λ_i increasing the expected return of a security with dZ_i-volatility σ_i with respect to the risk-free interest rate by $\lambda_i\sigma_i$ (see e.g. Hull, 2011).

Lemma 6.2. *Let us assume that there are n underlying securities $\theta_1, \ldots, \theta_n$ representing the n sources of uncertainty following the Itô's SDEs*

$$d\theta_i = \mu_i \theta_i dt + \sigma_i \theta_i dZ_i$$

with $\sigma_i \neq 0$. Let us define the price of risk of the i-th source of uncertainty as $\lambda_i = \frac{\mu_i - r}{\sigma_i}$. If the derivative security price F follows the Itô's SDE

$$dF = \mu_f F\, dt + \sigma_{f,1} F\, dZ_1 + \cdots + \sigma_{f,n} F\, dZ_n$$

then $\mu = r + \lambda_1 \sigma_{f,1} + \cdots + \lambda_n \sigma_{f,n}$ up to an infinitesimal error where r is the risk-free interest rate.

Proof: Let us consider the canonical representation of the processes θ_i on $\mathcal{T}_i$ with the uniform measure, i.e. $\delta\theta_i = \mu_i \theta_i \delta t + \sigma_i \theta_i \delta Z_i$ for $i = 1, \ldots, n$. Similarly,

$$\delta F = \mu_f F\, \delta t + \sigma_{f,1} F\, \delta dZ_1 + \cdots + \sigma_{f,n} F\, \delta Z_n$$

on $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n$ with the uniform measure.

Let $\Pi = F - \sum_{i=1}^n k_i \theta_i$, where $k_i = \frac{\sigma_{f,i} F(\omega)}{\sigma_i \theta_i(\omega)}$ is the value of the portfolio (set at time $t = t(\omega)$ at the state ω) consisting of a linear combination of the security F and the underlying securities θ_i so that all the δZ_i's are eliminated, i.e. its values at all states at $t + \delta t$ are the same. Consequently, the portfolio is risk-free and according to the non-arbitrage assumption $\delta\Pi = r\Pi\delta t + o(\delta t)$. Expanding the left-hand side and the right-hand side we get the equation

$$\left(\mu_f - \sum_{i=0}^{n} \mu_i \frac{\sigma_{f,i}}{\sigma_i}\right) F\delta t = \left(r - \sum_{i=0}^{n} r \frac{\sigma_{f,i}}{\sigma_i}\right) F\delta t + o(\delta t),$$

and so

$$\mu_f - r \approx \sum_{i=0}^{n} (\mu_i - r) \frac{\sigma_{f,i}}{\sigma_i} = \sum_{i=0}^{n} \lambda_i \sigma_{f,i}.$$

□

It is important to note that when we change the uniform measure on $\mathcal{T}_i$ by adjusting the branching probability from $\frac{1}{2}$ to $\frac{1}{2} + \alpha_i \frac{\sqrt{\delta t}}{2}$, we are in fact changing the price of risk from the original λ_i to $\lambda_i + \alpha_i$. Now, we are ready to prove the equivalent martingale measure theorem.

Theorem 6.3 (Equivalent martingale measure)**.** *Let G be a numeraire represented by (6.5) on $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n$ with the uniform measure. Let Q be the changed measure on $\mathcal{T}_1 \otimes \cdots \otimes \mathcal{T}_n$ obtained by adjusting the splitting probabilities on $\mathcal{T}_i$ from $\frac{1}{2}$ to $\frac{1}{2} + \alpha_i \frac{\sqrt{\delta t}}{2}$ with $\alpha_i = \sigma_{g,i} - \lambda_i$, i.e. changing the price of risk to $\sigma_{g,i}$. Then, if F is any Itô's process with the same sources of uncertainty represented by the equation (6.2), $\frac{F}{G}$ is a martingale with respect to Q, i.e. $\frac{F(t)}{G(t)} \approx E_Q[\frac{F(T)}{G(T)} \,|\, t]$ for any $t < T$.*

Proof: We could show that $\frac{F}{G}$ is a martingale with respect to Q just applying an elementary but tedious arithmetics of infinitesimals, however, it is more elegant to employ Itô's lemma.

To prove that $\frac{F}{G}$ is a martingale with respect to Q, we will use the logarithmic transformation. Let us first derive the SDE for $\ln\frac{F}{G}$ and $\frac{F}{G}$ under the uniform measure. If λ_i denote price of risk of the i-the source of uncertainty, then according to Lemma 6.2

$$\delta F = \left(r + \sum_{i=1}^{n} \lambda_i\,\sigma_{f,i}\right) F\,\delta t + \sum_{i=1}^{n} \sigma_{f,i}\,F\,\delta Z_i + o(\delta t) \quad \text{and}$$

$$\delta G = \left(r + \sum_{i=1}^{n} \lambda_i\,\sigma_{g,i}\right) G\,\delta t + \sum_{i=1}^{n} \sigma_{g,i}\,G\,dZ_i + o(\delta t).$$

According to Itô's lemma the processes $\ln F$ and $\ln G$ satisfy

$$\delta(\ln F) = \left(r + \sum_{i=1}^{n} \lambda_i\,\sigma_{f,i} - \frac{1}{2}\sum_{i=1}^{n} \sigma_{f,i}^2\right)\delta t + \sum_{i=1}^{n} \sigma_{f,i}\,\delta Z_i + o(\delta) \quad \text{and}$$

$$\delta(\ln G) = \left(r + \sum_{i=1}^{n} \lambda_i\,\sigma_{g,i} - \frac{1}{2}\sum_{i=1}^{n} \sigma_{g,i}^2\right) dt + \sum_{i=1}^{n} \sigma_{g,i}\,dZ_i + o(\delta t).$$

Hence $\ln\frac{F}{G} = \ln F - \ln G$ follows

$$\begin{aligned}\delta\left(\ln\frac{F}{G}\right) &= \left(\sum_{i=1}^{n}\left(\lambda_i(\sigma_{f,i}-\sigma_{g,i}) - \frac{1}{2}\sigma_{f,i}^2 - \frac{1}{2}\sigma_{g,i}^2\right)\right) dt \\ &\quad + \sum_{i=1}^{n}(\sigma_{f,i}-\sigma_{g,i})dZ_i + o(\delta t) \\ &= \sum_{i=1}^{n}\left((\lambda_i - \sigma_{g,i})(\sigma_{f,i}-\sigma_{g,i}) - \frac{1}{2}(\sigma_{f,i}-\sigma_{g,i})^2\right)\delta t \\ &\quad + \sum_{i=1}^{n}(\sigma_{f,i}-\sigma_{g,i})\delta Z_i + o(\delta t).\end{aligned}$$

Let us apply the exponential function to $\ln\frac{F}{G}$ and Itô's lemma again to get

$$\begin{aligned}\delta\left(\exp\ln\frac{F}{G}\right) &= \delta\left(\frac{F}{G}\right) \\ &= \sum_{i=1}^{n}\left((\lambda_i - \sigma_{g,i})(\sigma_{f,i}-\sigma_{g,i})\right)\frac{F}{G}\delta t + \sum_{i=1}^{n}(\sigma_{f,i}-\sigma_{g,i})\frac{F}{G}dZ_i + o(\delta t).\end{aligned}$$

Finally, let us change the uniform measure with $\alpha_i = \sigma_{g,i} - \lambda_i$, i.e. changing the i-th price of risk to $\sigma_{g,i}$. Then, according to the equation above, the mean of $\delta(\frac{F}{G})$ becomes zero up to an error of order $o(\delta t)$ and so according to Lemma 4.2

$\frac{F}{G}$ becomes a martingale with respect to the new measure Q, i.e. for every $t < T$:

$$\frac{F(t)}{G(t)} \approx E_Q \left(\frac{F(T)}{G(T)} \,|\, t \right).$$

$\square$

7. Conclusion

We have given a brief overview of the three main approaches to the building of infinitesimal stochastic analysis. Our approach has been instrumental in focusing on financial modeling applications. We believe that our presentation based on a compromise between the Robinsonian and the Nelsonian IST approach provides a simple and efficient way of transferring the elementary discrete time arguments into continuous time stochastic models needed to understand and further develop financial derivatives valuation techniques.

References

[1] Albeverio S., Fenstad J.E., Hoegh-Krohn R., Lindstrom T., *Nonstadard Methods in Stochastic Analysis and Mathematical Physics*, Dover Publications, 1986.

[2] Anderson R.M., *A nonstandard representation for Brownian motion and Itô integration*, Israel Math. J. **25** (1976), 15–46.

[3] Bacheier L., *La Theorie de la Speculation*, Annales de l'Ecole Normale Superieure, 3, Gauthier-Villars, Paris, 1900.

[4] Berg I., *Principles of Infinitesimal Stochastic and Financial Analysis*, World Scientific, Singapore, 2000.

[5] Brown R., *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies*, Phil. Mag. **4** (1828), 161–173.

[6] Cox J., Ross S., Rubinstein M., *Option pricing: a simplified approach*, J. Financial Econom. **7** (1979), 229–263.

[7] Cutland N., Kopp E., Willinger W., *A nonstandard approach to option pricing*, Mathematical Finance **1** (1991), no. 4, 1–38.

[8] Cutland N., Kopp E., Willinger W., *A nonstandard treatment of options driven by Poisson processes*, Stochastics Stochastics Rep. **42** (1993), 115–133.

[9] Cutland N., Kopp E., Willinger W., *From discrete to continuous stochastic calculus*, Stochastics Stochastics Rep. **52** (1995), 173–192.

[10] Cutland N., Kopp E., Willinger W., Wyman M.C., *Convergence of Snell envelopess and critical prices in the American put*, Mathematics of Derivative Securitities (M.A.H. Dempster and S.R. Pliska, eds.), Cambridge University Press, Cambridge, 1997, pp. 126–140.

[11] Cutland N., *Loeb Measures in Practice: Recent Advances*, Lecture Notes in Mathematics, 1751, Springer, Berlin, 2000.

[12] Einstein A., *Investigations on the theory of the Brownian movement*, R. Fürth (Ed.), Dover Publications, New York, 1956.

[13] Herzberg F.S., *Stochastic Calculus with Infinitesimals*, Springer, Heidelberg, 2013.

[14] Hull J., *Options, Futures, and Other Derivatives*, 8th edition, Prentice Hall, 2011.

[15] Hurd A.E., Loeb P.A., *An Introduction to Nonstandard Real Analysis*, Academic Press, Orlando, FL, 1985.

[16] Kalina M., *Probability in the alternative set theory*, Comment. Math. Univ. Carolin. **30** (1989), no. 2, 347–356.

[17] Keisler J.H., *An infinitesimal approach to stochastic analysis*, Mem. Amer. Math. Soc. **297**, 1984.

[18] Keisler J.H., *Elementary Calculus - an Infinitesimal Approach*, 2nd edition, University of Wisconsin - Creative Commons., 2000.

[19] Kopp P.E., *Hyperfinite Mathematical Finance*, in Arkeryd et al. Nonstandard Analysis: Theory and Applications, Kluwer, Dordrecht, 1997.

[20] Lindstrom T., *Nonlinear stochastic integrals for hyperfinite Lévy processes*, Logic and Analysis **1** (2008), 91–129.

[21] Loeb P.A., *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), 113–122.

[22] Nelson A., *Internal set theory: a new approach to nonstandard analysis*, Bull. Amer. Math. Soc. **83** (1977), no. 6, 1165–1198.

[23] Nelson E., *Radically Elementary Probability Theory*, Princeton University Press, Princeton, NJ, 1987.

[24] Pudlák P., Sochor A., *Models of the alternative set theory*, J. Symbolic Logic **49** (1984), 570–585.

[25] Rebonato R., *Volatility and Correlation: the Perfect Hedger and the Fox*, 2nd edition, Wiley, 2004.

[26] Robinson A., *Nonstandard analysis*, Proc. Roy. Acad. Amsterdam Ser. A **64** (1961), 432–440.

[27] Shreve S., *Stochastic Calculus for Finance II. Continuous Time Models*, Springer, New York, 2004.

[28] Shreve S., *Stochastic Calculus for Finance I. The Binomial Asset Pricing Model*, Springer, New York, 2005.

[29] Vopěnka P., *Mathematics in the Alternative Set Theory*, Teubner, Leipzig, 1979.

Faculty of Finance and Accounting, University of Economics,
Winston Churchill Sq. 4, Prague 3, Czech Republic

E-mail: witzanyj@vse.cz

(*Received* February 21, 2015, *revised* December 2, 2016)