LECTURE 1: CONTINUOUS TIME STOCHASTIC PROCESSES

PATRICK S. HAGAN
PHAGAN1@BLOOMBERG.NET
GORILLA SCIENCE
11 PALISADE PLAZA
EDGEWATER, NJ 07020

Abstract. Here we introduce the mathematical basis needed for the rest of the year. We construct Brownian motion, show that $dW^2 = dt$, derive Ito's lemma, discuss Martingales, derive Kolmogorov's backward equation and the Feynman-Kac formula. We then briefly introduce multi-dimensional and non-Markovian extensions.

1. Gaussian (normal) density. The Gaussian normal density for a variable with mean μ and variance V is

$$p(x)dx = \frac{1}{\sqrt{2\pi V}} e^{-(x-\mu)^2/2V}. \tag{1.1}$$

If a random variable x has this density, we speak of it being $G(\mu, V)$. Working out the first few moments:

$$\begin{array}{rlr} \langle x \rangle \equiv \int x p(x) ds = \mu & & \text{(mean)}, \\ \left\langle (x-\mu)^2 \right\rangle \equiv \int (x-\mu)^2 p(x) ds = V & & \text{(variance)}, \\ \left\langle (x-\mu)^3 \right\rangle \equiv \int (x-\mu)^3 p(x) ds = 0 & & (3^{\text{rd}} \text{ moment}), \\ \left\langle (x-\mu)^4 \right\rangle \equiv \int (x-\mu)^4 p(x) ds = 3V^2 & & (4^{\text{th}} \text{ moment}). \end{array} \tag{1.2}$$

We note that if x is $G(\mu, \Delta t)$, then

$$x = \mu + \sqrt{\Delta t}\xi, \tag{1.3}$$

where ξ is a standard Gaussian variable with mean zero and variance 1; i.e., ξ is $G(0,1)$.

2. Construction of Brownian motion. Previously we used lattices to model random processes, like asset prices. One naturally wonders whether these processes go to a limit as the step size is taken to be finer and finer. Or more to the point, if we use lattices to model asset prices, does the model make sense in the limit that the step size goes to zero? To answer this question we now advance to continuous random processes.

We are aiming to develop models based on stochastic differential equations,

$$dS = a(t,S)dt + b(t,S)dW \tag{2.1}$$

for the asset price $S(t)$, where the $a(t,S)dt$ term accounts for "deterministic motions," whatever those are, and the other term $b(t,S)dW$ accounts for "random motions," whatever those are.

The first step in developing such models is to decide what we should use for $W(t)$, the random part of the model No matter how we sub-divide it, the curve $W(t)$ should still be random and composed of pieces with identical statistical properties. This, because asset prices appear random on even very fine time scales.

More formally, we wish $W(t)$ to have the following properties:

a) independent increments. For any date τ and for any $\Delta\tau > 0$, the value of $\Delta W = W(\tau + \Delta\tau) - W(\tau)$ is independent of $W(t)$ for all $t \le \tau$. So *increments of Brownian motion are independent of everything that has happened on or before the current date* τ. In particular, the increments $\Delta W \equiv W(t_2) - W(t_1)$ and $\Delta W \equiv W(t_3) - W(t_4)$ are independent *whenever* the two intervals $t_1 \le t \le t_2$ and $t_3 \le t \le t_4$ don't overlap.

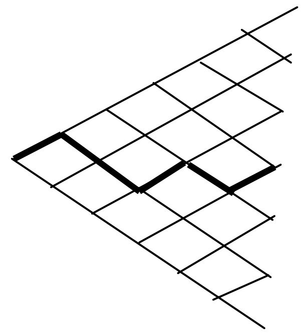

FIG. 2.1. *Lattice path*

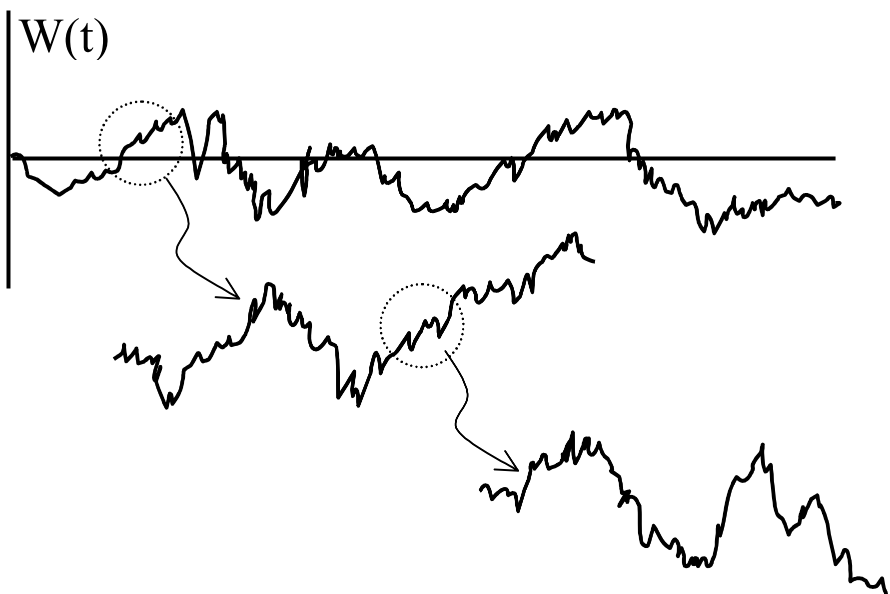

FIG. 2.2. *Subdivision of Brownian motion (from an early photograph)*

Said another way, $W(t) - W(\tau)$ for $t > \tau$ does not depend on how one got to $W(\tau)$. As we shall see, this is a very powerful simplifying assumption. .

b) Increments $\Delta W \equiv W(t_2) - W(t_1)$ are Gaussian random variables with mean 0 and variance $\Delta t \equiv t_2 - t_1$. That is,

$$\Delta W \equiv W(t_2) - W(t_1) = \sqrt{t_2 - t_1}\xi \tag{2.2}$$

FIG. 2.3. *The change $\Delta W(t) = W(t_{end}) - W(t_{st})$ over distinct intervals is independent*

where ξ is $G(0,1)$; i.e., ξ is a Gaussian random variable with mean zero and variance 1.

The reason we want ΔW to have mean zero is because we want it to represent the random part of the asset price movements; any non-zero mean would represent a deterministic piece which we could put in the drift term $a(t,S)dt$.

The fact that ΔW is Gaussian with variance Δt follows directly from our desire to have $W(t)$ to be subdivideable into finer and finer intervals, each with identical statistical properties. To show this, consider

$$\Delta W = W(t_1) - W(t_0) \equiv \sum_{k=0}^{n-1} \left[W(\tau_{k+1}) - W(\tau_k) \right], \qquad \text{with } \tau_k = t_0 + \frac{k}{n}(t_1 - t_0) \tag{2.3a}$$

where each $\delta W_k \equiv W(\tau_{k+1}) - W(\tau_k)$ are independent random variables (by the independent increment assumption) with identical distributions. Since the variables are independent, the variances add,

$$\text{Var}\left\{\Delta W\right\} = \sum_{k=0}^{n-1} \text{Var}\left\{W(\tau_{k+1}) - W(\tau_k)\right\} = n \,\text{Var}\left\{W(t_1) - W(t_0)\right\}. \tag{2.3b}$$

Let us abbreviate $v(y) = \text{Var}\left\{W(t+y) - W(t)\right\}$. We have shown that

$$v(t_1 - t_0) = nv\left(\frac{t_1 - t_0}{n}\right) \tag{2.4}$$

for any t_1, t_0 and n. I.e., for any $\Delta t > 0$ and any n, we have

$$v(n\Delta t) = nv(\Delta t). \tag{2.5}$$

This is a functional equation, and it shouldn't be surprising that the only reasonable solutions are linear:

$$v(\Delta t) = \alpha \Delta t \tag{2.6}$$

for some constant α. Brownian motion is normalized so that this constant is 1:

$$\text{Var}\left\{\Delta W\right\} \equiv \text{Var}\left\{W(t_1) - W(t_0)\right\} = t_1 - t_0 \qquad \text{for all } t_0, t_1. \tag{2.7}$$

Thus ΔW is the sum of n independent, identically disributed variales with mean 0 and variance $\frac{t_1 - t_0}{n}$,

$$\Delta W = W(t_1) - W(t_0) \equiv \sum_{k=0}^{n-1} \left[W(\tau_{k+1}) - W(\tau_k) \right], \tag{2.8}$$

As we take $n \to \infty$, the central limit theorem guarantees that ΔW is Gaussian with mean zero and variance $t_1 - t_0$.

3. Properties of Brownian motion. Brownian motion has the following properties. Of these, the first two are part of the defintion of Brownian motion, and the other three are derived below:

a1) *the increments ΔW are independent of the present and past values of $W(t)$.* In particular, increments of non-overlapping intervals are independent:

$$\Delta W = W(t_2) - W(t_1) \text{ is independent of } W(t) \text{ for all } t \le t_1 \tag{3.1a}$$

$$\Delta W = W(t_2) - W(t_1) \text{ is independent of } \Delta \tilde{W} = W(t_4) - W(t_3) \text{ if } (t_1, t_2) \cap (t_3, t_4) = \varnothing. \tag{3.1b}$$

.

a2) $\Delta W = W(t+\Delta t) - W(t)$ *is Gaussian with mean zero and variance* $\sqrt{\Delta t}$. Said another way,

$$\Delta W = W(t+\Delta t) - W(t) = \sqrt{\Delta t}\xi, \tag{3.2}$$

where ξ is $G(0,1)$, a Gaussian varible with mean 0 and variance 1.

b1) $W(t)$ *is a continuous random process.* This is easily proven, since for any $\delta > 0$,

$$\text{prob}\left\{|W(t+\Delta t) - W(t)| > \delta\right\} = \text{prob}\left\{|\xi| > \frac{\delta}{\sqrt{\Delta t}}\right\} \longrightarrow 0 \qquad \text{as } \Delta t \longrightarrow 0. \tag{3.3a}$$

This is the definition of a continuous stochastic processes.

b2) $W(t)$ *is almost surely nowhere differentiable.* This is again easily shown. For any $K > 0$, we argue that

$$\text{prob}\left\{\left|\frac{W(t+\Delta t) - W(t)}{\Delta t}\right| < K\right\} = \text{prob}\left\{|\xi| > K\sqrt{\Delta t}\right\} \longrightarrow 0 \qquad \text{as } \Delta t \longrightarrow 0. \tag{3.3b}$$

So the probability that the slope is bounded is zero as $\Delta t \longrightarrow 0$.

b3) The continuity and non-differentiability followed directly from the scaling of ΔW. Since $\Delta W = \sqrt{\Delta t}\xi$, where ξ is $G(0,1)$, we can write $\Delta W \sim O(\sqrt{\Delta t})$, or more succinctly

$$dW \sim O\left(\sqrt{dt}\right). \tag{3.4a}$$

Later we shall prove a much more stunning result, *the quadratic property of Brownian motion,*

$$(dW)^2 = dt. \tag{3.4b}$$

Note that the right side dt is not stochastic, which means that dW^2 is dt with certainty. This property is the key to deriving Ito's lemma, the Backward's Kolmogorov equation, Feynman-Kac equation and many of the other day-to-day tools used in pricing. Before we can show this result, we need to define what we mean by differentials like dW and dt.

4. Integration. It would seem natural to try to derive differential equations of the form

$$\frac{dX}{dt} = a(t,X) + b(t,X)\frac{dW}{dt}. \tag{4.1}$$

Since we've shown that $W(t)$ is nowhere differentiable, this would give mathematicians a heart attack. So, mercifully, we write our differential equations as

$$dX = a(t,X)dt + b(t,X)dW. \tag{4.2}$$

Processes $X(t)$ which follow such equations are known as *Ito processes.*

The differential 4.2 is interpretted to mean that for any t, $X(t)$ is given by the following the integral

$$X(t) = X(t_0) + \int_{t_0}^{t} a(\tau, X(\tau)d\tau + \int_{W(t_0)}^{W(t)} b(\tau, X(\tau))dW(\tau). \tag{4.3}$$

In turn, the integral is interpretted to mean that $X(t)$ is the limit as $n \to \infty$ of $X(\tau_n)$, where

$$X(\tau_j) = X(t_0) + \lim_{n\to\infty}\left\{\sum_{k=0}^{j-1} a(\tau_k, X(\tau_k))\left[\tau_{k+1} - \tau_k\right] + \sum_{k=0}^{j-1} b(\tau_k, X(\tau_k))\left[W(\tau_{k+1}) - W(\tau_k)\right]\right\}, \tag{4.4a}$$

and

$$\tau_k = t_0 + \frac{k}{n}(t - t_0). \tag{4.4b}$$

Observe that

$$W(\tau_{k+1}) - W(\tau_k) = \sqrt{\tau_{k+1} - \tau_k}\,\xi_k = \sqrt{\frac{t - t_0}{n}}\,\xi_k \tag{4.4c}$$

where $\xi_0, \xi_1, \ldots \xi_{n-1}$ are independent $G(0,1)$ variables (Gaussian variables, mean 0, variance 1). To construct a *realization* (path) for an Ito process from t_0 to t, one first picks n, and subdivides the interval t_0 to t into n parts. One then picks n independent $G(0,1)$ variables for the ξ_k and uses these to construct the $X(t)$ at all the node points τ_k. This is directly analogous to setting up a lattice, and then creating a path by choosing up and down at random at each node.

One can prove that these sums converge (in probability distribution) as $n \to \infty$. This material will be covered in a later course. We emphasize that the differential in equation 4.2 means nothing more (and nothing less) than the Ito integral in 4.3; and that the Ito integral in 4.3 means nothing more and nothing less than the limiting sum in 4.4a - 4.4c.

4.1. Stratonovich calculus. It is very important to note that in the sum 4.4a defining the Ito integral, the coefficients $a(\tau, X(\tau))$ and $b(\tau, X(\tau))$ are evaluated at the beginning point τ_k of each interval (τ_k, τ_{k+1}). This turns out to be the simplest case to analyze mathematically, and turns out to be appropriate for modeling financial asset prices. However it is not the only reasonable choice. An alternative definition evaluates $a(\tau, X(\tau))$ and $b(\tau, X(\tau))$ at the mid-point

$$\tau_{k+1/2} = t_0 + \frac{k + \frac{1}{2}}{n}(t - t_0) \tag{4.5a}$$

of each interval:
(4.5b)

$$X(\tau_j) = X(t_0) + \lim_{n\to\infty} \left\{ \sum_{k=0}^{j-1} a(\tau_{k+1/2}, X(\tau_{k+1/2}))\,[\tau_{k+1} - \tau_k] + \sum_{k=0}^{j-1} b(\tau_{k+1/2}, X(\tau_{k+1/2}))\,[W(\tau_{k+1}) - W(\tau_k)] \right\}.$$

This is the *Statonovich integral*, and defines the Stratonovich calculus. Although these are not use in finance, in many of the sciences, the Stratonovich calculus provides a better description of physical noise processes.

It is easy to demonstrate that the two integrals give numerically different answers. The differences are

$$d_1 = \sum_{k=0}^{n-1} \left[a(\tau_{k+1/2}, X(\tau_{k+1/2})) - a(\tau_k, X(\tau_k)) \right] [\tau_{k+1} - \tau_k], \tag{4.6a}$$

$$d_2 = \sum_{k=0}^{n-1} \left[b(\tau_{k+1/2}, X(\tau_{k+1/2})) - b(\tau_k, X(\tau_k)) \right] [W(\tau_{k+1}) - W(\tau_k)] \tag{4.6b}$$

Since $W(\tau_{k+1}) - W(\tau_k) = O(\sqrt{\tau_{k+1} - \tau_k}) = O(1/\sqrt{n})$, clearly we also have $W(\tau_{k+1/2}) - W(\tau_k) = O(1/\sqrt{n})$, whilst $\tau_{k+1/2} - \tau_k = O(1/n)$. Assuming that $a(\tau, X)$ and $b(\tau, X)$ are smooth, say bounded with bounded first derivatives, then the first sum gives

$$d_1 = \sum_{k=0}^{n-1} \left[\frac{\partial a}{\partial \tau} O(1/n) + \frac{\partial a}{\partial X} O(1/\sqrt{n}) \right] O(1/n). \tag{4.7a}$$

Since there are n terms in the sum, and the biggest term is $O(1/n^{3/2})$, one expects this sum to converge to zero as $n \longrightarrow \infty$. The second sum is a different matter. Because it is multiplied by ΔW instead of Δt, we obtain the scaling

$$d_2 = \sum_{k=0}^{n-1} \left[\frac{\partial b}{\partial \tau} O(1/n) + \frac{\partial b}{\partial X} O(1/\sqrt{n}) \right] O(1/\sqrt{n}). \tag{4.7b}$$

We see that this sum has n terms of size $O(1/n)$, so as $n \longrightarrow \infty$ we expect this sum to converge to a finite quantity.

5. The quadratic propery. We now derive the property $dW^2 = dt$. It is this property that gives stochastic processes their outre character. Like any other diffential, this differential is defined in terms of its integral:

$$\int_{t_0}^{t_1} (dW)^2 \equiv \lim_{n\to\infty} \sum_{k=0}^{n-1} \left[W(\tau_{k+1}) - W(\tau_k) \right]^2, \tag{5.1a}$$

where

$$\tau_k = t_0 + \frac{k}{n}(t_1 - t_0). \tag{5.1b}$$

But

$$W(\tau_{k+1}) - W(\tau_k) = \sqrt{\tau_{k+1} - \tau_k}\,\xi_k = \sqrt{\frac{t_1 - t_0}{n}}\,\xi_k, \tag{5.2}$$

so this is just

$$\int_{t_0}^{t_1} (dW)^2 = \lim_{n\to\infty} \frac{t_1 - t_0}{n} \sum_{k=0}^{n-1} \xi_k^2, \tag{5.3}$$

where $\xi_0, \xi_1, \ldots \xi_{n-1}$ are independent $G(0,1)$ variables. Clearly the mean of the sum is

$$\mathrm{E}\left\{ \frac{t_1 - t_0}{n} \sum_{k=0}^{n-1} \xi_k^2 \right\} = \frac{t_1 - t_0}{n} \sum_{k=0}^{n-1} \mathrm{E}\left\{ \xi_k^2 \right\} = t_1 - t_0. \tag{5.4}$$

Since the ξ's are independent, the variance of the sum is

$$\begin{aligned} \mathrm{Var}\left\{ \frac{t_1 - t_0}{n} \sum_{k=0}^{n-1} \xi_k^2 \right\} &= \frac{(t_1 - t_0)^2}{n^2} \mathrm{Var}\left\{ \sum_{k=0}^{n-1} \xi_k^2 \right\}, \\ &= \frac{(t_1 - t_0)^2}{n^2} \sum_{k=0}^{n-1} \mathrm{Var}\left\{ \xi_k^2 \right\} \\ &= \frac{(t_1 - t_0)^2}{n^2} \sum_{k=0}^{n-1} \mathrm{E}\left\{ \left(\xi_k^2 - 1 \right)^2 \right\}. \end{aligned} \tag{5.5}$$

For unit Gaussian variables, $\mathrm{E}\{(\xi_k^2-1)^2\}=2$, so the variance of the sum works out to

$$\operatorname{Var}\left\{\frac{t_1-t_0}{n}\sum_{k=0}^{n-1}\xi_k^2\right\}=\frac{2}{n}(t_1-t_0)^2. \tag{5.6}$$

Thus

$$\int_{t_0}^{t_1}(dW)^2\equiv\lim_{n\to\infty}S_n, \tag{5.7}$$

where the sum S_n has mean t_1-t_0 and variance $O(1/n)$. We conclude that in the limit $n\to\infty$, this integral is t_1-t_0 with certainty. Thus,

$$\int_{t_0}^{t_1}(dW)^2=t_1-t_0 \qquad \text{for any } t_0 \text{ and } t_1. \tag{5.8}$$

Since differentials are defined only in terms of their integrals, we can re-write this as

$$(dW)^2=dt \tag{5.9}$$

For your homework, you will be asked to strengthen this result by showing that if $X(t)$ is an Ito process, then

$$\int_{t_0}^{t_1}c(t,X(t)\,(dW)^2=\int_{t_0}^{t_1}c(t,X(t))dt \tag{5.10a}$$

for sufficiently nice functions $c(t,X)$. Thus,

$$c(t,X(t)\,(dW)^2=c(t,X(t)dt, \tag{5.10b}$$

as suggested by the notation.

5.1. Box algebra. The other quadratic differentials are zero: $(dt)^2=0$ and $dW\,dt=0$. To show this, let us write out their integrals. First,

$$\int_{t_0}^{t_1}(dt)^2=\lim_{n\to\infty}\sum_{k=0}^{n-1}(\tau_{k+1}-\tau_k)^2=\lim_{n\to\infty}\frac{1}{n}(t_1-t_0)^2=0. \tag{5.11}$$

Just as easy,

$$\begin{aligned}\int_{t_0}^{t_1}dtdW&=\lim_{n\to\infty}\sum_{k=0}^{n-1}(\tau_{k+1}-\tau_k)\,[W(\tau_{k+1})-W(\tau_k)]=\lim_{n\to\infty}\sum_{k=0}^{n-1}(\tau_{k+1}-\tau_k)^{3/2}\,\xi_k\\&=\lim_{n\to\infty}\frac{(t_1-t_0)^{3/2}}{n^{3/2}}\sum_{k=0}^{n-1}\xi_k..\end{aligned} \tag{5.12}$$

Since means and variances of independent variables are additive, clearly the sum of the ξ_k gives a Gaussian variable with mean 0 and variance n, so

$$\int_{t_0}^{t_1}dtdW=\lim_{n\to\infty}\frac{(t_1-t_0)^{3/2}}{n}\xi. \tag{5.13}$$

where ξ is a $G(0,1)$ variable. Clearly this is zero in the limit.

Putting this together with our preceding results gives the so-called box algebra:

$$(dW)^2 = dt, \qquad dW\,dt = 0, \qquad (dt)^2 = 0. \tag{5.14a}$$

Of course, all higher powers are aslo zero:

$$\begin{aligned} (dW)^k &= 0 \qquad \text{for } k > 2, \\ (dW)^k\, dt &= 0 \qquad \text{for } k > 1, \\ (dt)^k &= 0 \qquad \text{for } k > 1. \end{aligned} \tag{5.14b}$$

In your homework, you will be asked to strengthen these results by showing that

$$C(t, X(t))dW dt = 0, \qquad C(t, X(t))\,(dt)^2 = 0. \tag{5.15a}$$

Of course, all higher powers are aslo zero:

$$\begin{aligned} C(t, X(t))\,(dW)^k &= 0 \qquad \text{for } k > 2, \\ C(t, X(t))\,(dW)^k\, dt &= 0 \qquad \text{for } k > 1, \\ C(t, X(t))\,(dt)^k &= 0 \qquad \text{for } k > 1. \end{aligned} \tag{5.15b}$$

where $C(t, X)$ is a sufficiently nice function and $X(t)$ is an Ito process.

6. Monte Carlo (MC) simulations. Suppose we have some variable, an asset price perhaps, which we model by an Ito process:

$$dX = a(t, X)dt + b(t, X)dW. \tag{6.1a}$$

Commonly the value of a financial instrument will turn out to be the expected value of some payoff at the expiry date T,

$$V = \mathrm{E}\left\{P(X(T)\right\} \tag{6.1b}$$

The Monte Carlo method is the most direct method of calculating such expected values. Recall that the Ito process 6.1a is equivalent to stating that $X(t)$ is the limit as $n \to \infty$ of $X(\tau_n)$, where

$$X(\tau_j) = X(t_0) + \lim_{n\to\infty} \left\{ \sum_{k=0}^{j-1} a(\tau_k, X(\tau_k))\,[\tau_{k+1} - \tau_k] + \sum_{k=0}^{j-1} b(\tau_k, X(\tau_k))\sqrt{\tau_{k+1} - \tau_k}\xi_k \right\}, \tag{6.2}$$

for $j = 0, 1, \ldots, n-1$. Here $\tau_k = t_0 + \frac{k}{n}(T - t_0)$ and $\xi_0, \xi_1, \ldots, \xi_{n-1}$ are independent $G(0,1)$ variables.

For the MC method, we first discretize in time, picking $\tau_0, \tau_1, \ldots, \tau_n = T$. We then pick the n independent $G(0,1)$ variables $\xi_0, \xi_1, \ldots, \xi_{n-1}$. Substituting these into 6.2 then gives a possible path (also called a *realization*) or the asset price $X(t)$. Were the asset to follow this path, the financial instrument would yield $P(X(T)) = P_{path\ 1}$. Repeating this procedure for newly selected random variables $\xi_0, \ldots, \xi_{n-1}$ yields a second possible path and payoff $P_{path\ 2}$. Repeating this many times and averaging over the outcomes then gives the option price

$$V \equiv \mathrm{E}\left\{P(X(T)\right\} = \frac{1}{N_{path}} \sum_k P_{path\ k} \tag{6.3}$$

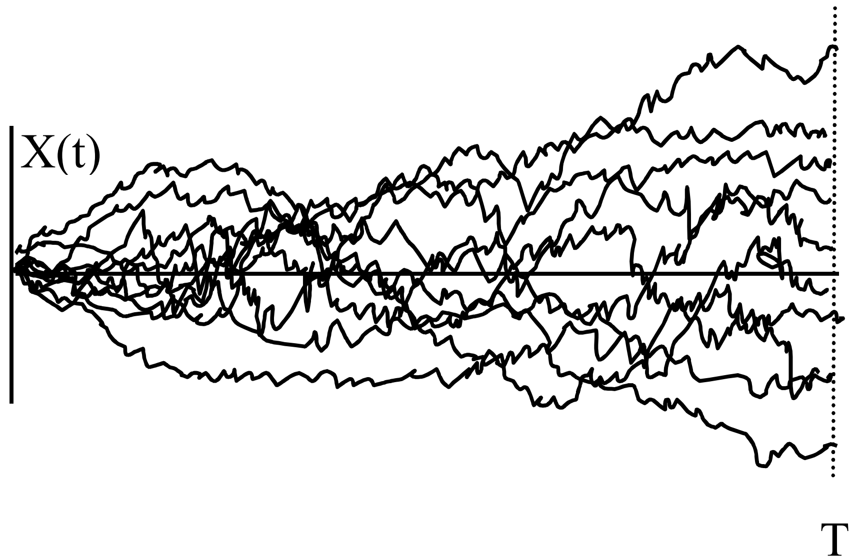

FIG. 6.1. *Realizations in a MC simulation*

The Monte Carlo method's key advantage is that it is very flexible. For example,suppoe we have a path dependent financial instruments, whose payoffs depends on, say the maximum, minimum or average value of $X(t)$ between t_0 and T. Since MC simulates the entire path, the value of these options is no harder to determine that the value of a European option.

Weaknesses of MC is that it is slow and computationally expensive. In fact, let σ be the standard deviation of the payoff::

$$\sigma^2 = \mathrm{Var}\left\{P(X(T)\right\}. \approx \mathrm{Var}\left\{P_{path\ k}\right\} \tag{6.4}$$

(The last equality is only approximate due to our time discretization in the MC method). Since the random variables on each path are chosen independently, the variances on different paths add. So we have

$$\mathrm{Var}\left\{\frac{1}{N_{path}}\sum_k P_{path\ k}\right\} = \frac{1}{N_{path}^2}\mathrm{Var}\left\{\sum_k P_{path\ k}\right\} = \frac{1}{N_{path}}\mathrm{Var}\left\{P_{path\ k}\right\} = \frac{\sigma^2}{N_{path}}. \tag{6.5}$$

The typical error in the MC evaluation is the standard deviation $\sigma/\sqrt{N_{path}}$. So quite generally the error in the MC method goes down like $1/\sqrt{N_{path}}$. In other words, to reduce the error by a factor of 10, one needs to 100 times as many paths.

7. Martingales. Recall that a stochastic process $M(t)$ is a Martingale if and only if at each time t, the expected value of $M(T)$ at any date $T > t$ is the current value $M(t)$. I.e., for any t,

$$m = \mathrm{E}\left\{M(T) \mid M(t) = m\right\} \tag{7.1}$$

Let us once again consider the Ito process

$$dX = a(t, X)dt + b(t, X)dW, \tag{7.2}$$

and ask which of these processes are Martingales. Let us choose an arbitrary time t_0, and let $X(t_0) = x$. Writing the Ito process as the limit as $n \to \infty$ of $X(\tau_n)$, where

$$X(\tau_j) = x + \lim_{n\to\infty} \left\{ \sum_{k=0}^{j-1} a(\tau_k, X(\tau_k)) \left[\tau_{k+1} - \tau_k\right] + \sum_{k=0}^{j-1} b(\tau_k, X(\tau_k)) \left[W(\tau_{k+1}) - W(\tau_k)\right] \right\}. \tag{7.3}$$

Taking the expected value yields

$$\begin{aligned} \mathrm{E}\left\{X(\tau_j) \mid X(t_0) = x\right\} = x + \lim_{n\to\infty} \Bigg\{ &\sum_{k=0}^{j-1} \mathrm{E}\left\{a(\tau_k, X(\tau_k)) \mid X(t_0) = x\right\} \left[\tau_{k+1} - \tau_k\right] \\ &+ \sum_{k=0}^{j-1} \mathrm{E}\left\{b(\tau_k, X(\tau_k)) \left[W(\tau_{k+1}) - W(\tau_k)\right] \mid X(t_0) = x\right\} \Bigg\} \end{aligned} \tag{7.4}$$

By construction, it is clear that $X(\tau_k)$ depends only on events happening at times $t < \tau_k$. Recall that the increments $\Delta W = W(\tau_{k+1}) - W(\tau_k)$ are independent of all events happening on or before τ_k, and hence must be independent of $X(\tau_k)$. Since ΔW is Gaussian with mean zero, clearly

$$\begin{aligned} &\mathrm{E}\left\{b(\tau_k, X(\tau_k)) \left[W(\tau_{k+1}) - W(\tau_k)\right] \mid X(t_0) = x\right\} \\ &\qquad = \mathrm{E}\left\{b(\tau_k, X(\tau_k)) \mid X(t_0) = x\right\} \mathrm{E}\left\{W(\tau_{k+1}) - W(\tau_k)\right\} = 0. \end{aligned} \tag{7.5}$$

Therefore

$$\mathrm{E}\left\{X(\tau_j) \mid X(t_0) = x\right\} = x + \lim_{n\to\infty} \left\{ \sum_{k=0}^{j-1} \mathrm{E}\left\{a(\tau_k, X(\tau_k)) \mid X(t_0) = x\right\} \left[\tau_{k+1} - \tau_k\right] \right\}. \tag{7.6a}$$

More succinctly, we can write this as

$$d\,\mathrm{E}\left\{X(t)\right\} = \mathrm{E}\left\{a(t, X(t))\right\} dt. \tag{7.6b}$$

If $a(t, X) \equiv 0$ for all t and X, then clearly

$$\mathrm{E}\left\{X(t) \mid X(t_0) = x\right\} = \mathrm{E}\left\{X(\tau_n) \mid X(t_0) = x\right\} = x. \tag{7.7}$$

So $X(t)$ is a Martingale if $a(t, X) \equiv 0$ for all t and X. On the other hand, if $a(t, X)$ is non-zero at any point, we can let this point be $t = t_0, X = x$ and argue that the first step has a non-zero mean,

$$\mathrm{E}\left\{X(\tau_1) \mid X(t_0) = x\right\} = x + a(t_0, z) \left[\tau_1 - t_0\right] \tag{7.8}$$

By making more careful estimates, one can show that

$$\mathrm{E}\left\{X(\tau_1) \mid X(t_0) = x\right\} \neq x \tag{7.9}$$

for τ_1 larger than, but sufficiently near, t_0. Therefore, *an Ito process*

$$dX = a(t, X)dt + b(t, X)dW \tag{7.10}$$

is a Martingale if and only if $a(t, X) \equiv 0$ *for all* t, X.

The $a(t, X)dt$ term is known as the *drift term*; at each t, X is determines the mean value of the change in X for small Δt. The $b(t, X)dW$ term is known as the *diffusion term*; it governs the increase in the variance of X for small Δt.

8. Ito's lemma. THEOREM 8.1. *(Ito's lemma). Suppose $X(t)$ is an Ito process:*

$$dX = a(t, X)dt + b(t, X)dW. \tag{8.1a}$$

Consider the stochastic process

$$F(t) \equiv f(t, X(t)) \tag{8.1b}$$

for some smooth function $f(t, x)$, say $C_t^2 \cap C_x^3$. Then $F(t)$ is an Ito process with

$$dF = \left(\frac{\partial f}{\partial t} + a \frac{\partial f}{\partial X} + \tfrac{1}{2} b^2 \frac{\partial^2 f}{\partial X^2} \right) dt + b \frac{\partial f}{\partial X} dW. \tag{8.1c}$$

Here, $a = a(t, X(t)), b = b(t, X(t))$, and $f = f(t, X(t))$

This theorem is very, very important. To derive this result, suppose that

$$dX = a(t, X)dt + b(t, X)dW. \tag{8.2}$$

We first show that

$$(dX)^2 = b^2(t, X)dt \tag{8.3}$$

If we blithely follow the *box algebra*, we argue that

$$(dX)^2 = a^2(t, X)\,(dt)^2 + 2a(t, X)b(t, X)dtdW + b^2(t, X)\,(dW)^2, \tag{8.4}$$

and use the results

$$(dW)^2 = dt, \qquad dW dt = 0, \qquad (dt)^2 = 0, \tag{8.5}$$

to write this as

$$(dX)^2 = b^2(t, X)dt. \tag{8.6}$$

In your homework, you will be asked to provide a sounder derivation by recalling that a differential is defined in terms of the integral, and arguing along the lines

$$\begin{aligned} \int_{t_0}^{t} (dX)^2 &= \lim_{n\to\infty} \sum_{k=0}^{n-1} \{X(\tau_{k+1}) - X(\tau_k)\}^2 \\ &= \lim_{n\to\infty} \sum_{k=0}^{n-1} \{a(\tau_k, X(\tau_k))\,[\tau_{k+1} - \tau_k] + b(\tau_k, X(\tau_k))\,[W(\tau_{k+1}) - W(\tau_k)]\}^2 \\ &= \sum_{k=0}^{n-1} a_k^2\,[\tau_{k+1} - \tau_k]^2 + 2a_k b_k\,[\tau_{k+1} - \tau_k]\,[W(\tau_{k+1}) - W(\tau_k)] + b_k^2\,[W(\tau_{k+1}) - W(\tau_k)]^2 \end{aligned} \tag{8.7}$$

By extending the arguments used to show the box alegebra results 8.5, one can then show that the first two terms are negligible and the last one can be taken to be

$$\int_{t_0}^{t} (dX)^2 = \int_{t_0}^{t} b^2(\tau_k, X(\tau_k))(dW)^2 = . \int_{t_0}^{t} b^2(\tau_k, X(\tau_k))dt, \tag{8.8}$$

which establishes the result.

Similarly,

$$dXdt = \{a(t,X)dt + b(t,X)dW\}\, dt = a(t,X)dt^2 + b(t,X)dtdW \\ = 0 + 0 = 0,. \tag{8.9a}$$

and

$$\begin{aligned} (dX)^k &= 0 \qquad \text{for } k > 2, \\ (dX)^k\, dt &= 0 \qquad \text{for } k > 1, \\ (dt)^k &= 0 \qquad \text{for } k > 1. \end{aligned} \tag{8.9b}$$

Let us now examine $F(t) = f(t, X(t))$. Clearly,

$$dF \equiv F(t+dt) - F(t) = f(t+dt, X+dX) - f(t,X). \tag{8.10}$$

Since f is smooth, we can expand this as

$$\begin{aligned} dF = {} & \frac{\partial f(t,X)}{\partial t}dt + \frac{\partial f(t,X)}{\partial X}dX \\ & + \tfrac{1}{2}\frac{\partial^2 f(t,X)}{\partial t^2}dt^2 + \frac{\partial^2 f(t,X)}{\partial t \partial X}dtdX + \tfrac{1}{2}\frac{\partial^2 f(t,X)}{\partial X^2}dX^2 \\ & + \tfrac{1}{6}\frac{\partial^3 f(t,X)}{\partial X^3}dX^3 + \cdots . \end{aligned} \tag{8.11}$$

Using the above results to replace

$$\begin{aligned} dF &= \frac{\partial f(t,X)}{\partial t}dt + \frac{\partial f(t,X)}{\partial X}dX + \tfrac{1}{2}\frac{\partial^2 f(t,X)}{\partial X^2}dX^2 \\ &= \frac{\partial f(t,X)}{\partial t}dt + \frac{\partial f(t,X)}{\partial X}dX + \tfrac{1}{2}b^2(t,X)\frac{\partial^2 f(t,X)}{\partial X^2}dt,. \end{aligned} \tag{8.12}$$

and substituting $dX = a(t,X)dt + b(t,X)dW$ and collecting terms yields

$$dF = \left\{ \frac{\partial f(t,X)}{\partial t} + a(t,X)\frac{\partial f(t,X)}{\partial X} + \tfrac{1}{2}b^2(t,X)\frac{\partial^2 f(t,X)}{\partial X^2} \right\} dt + b(t,X)\frac{\partial f(t,X)}{\partial X}dW, \tag{8.13}$$

as was to be shown.

9. Working with SDE's (stochastic differential equations). Operationally, working with SDE's is straightforward. One

(a) expands in a Taylor series;

(b) treats $dt = O(dt)$ and $dW = O(\sqrt{dt})$, and then neglects all termsof order $(dt)^{3/2}$ and higher order; and

(c) replaces dW^2 by dt.

Using this procedure immediately gives us Ito's lemma.

9.0.1. Examples. *Example 1.* Show that

$$M(t) \equiv W^2(t) - t \tag{9.1}$$

is a Martingale.

From Ito's lemma, $dM = 2WdW - dt + \frac{1}{2} \cdot 2 \cdot dW^2 = 2WdW$ since $dW^2 = dt$. Since this doesn't have a drift term, it is a Martingale.

Example 2. Show that

$$M(t) \equiv e^{\alpha W(t) - \frac{1}{2}\alpha^2 t} \tag{9.2}$$

is a Martingale.
From Ito's lemma,

$$dM = e^{\alpha W(t) - \frac{1}{2}\alpha^2 t} \left\{ \alpha dW - \frac{1}{2}\alpha^2 dt + \frac{1}{2} \cdot \alpha^2 dW^2 \right\} = \alpha M dW \tag{9.3}$$

since $dW^2 = dt$. Since this process doesn't have a drift term, it is a Martingale.

Example 3. Define I by

$$dI \equiv WdW, \tag{9.4a}$$

so

$$I = I(0) + \int_0^t W(t')dW(t'). \tag{9.4b}$$

Write I in simpler form, and find it's mean and variance.
We try $d\left(\frac{1}{2}W^2\right)$ and find that

$$d\left(\frac{1}{2}W^2\right) = WdW + \frac{1}{2}dW^2 = WdW + \frac{1}{2}dt. \tag{9.5a}$$

So clearly, $d\left(\frac{1}{2}W^2 - \frac{1}{2}t\right) = WdW$, so clearly $dI = d\left(\frac{1}{2}W^2 - \frac{1}{2}t\right)$,and

$$I(t) = \tfrac{1}{2}W^2(t) - \tfrac{1}{2}t - \tfrac{1}{2}W^2(0). \tag{9.5b}$$

Since $W(t)$ is Gaussian with mean $W(0)$ and variance t, we have

$$\left\langle [W(t) - W(0)]^2 \right\rangle = t, \qquad \left\langle [W(t) - W(0)]^4 \right\rangle = 3t^2, \tag{9.6a}$$

so working out the algebra yields

$$\langle I(t) \rangle = I(0), \qquad \left\langle [I(t) - I(0)]^2 \right\rangle = \tfrac{1}{2}t^2 - W^2(0)t. \tag{9.6b}$$

Note: Inspection shows $dI = WdW$ has no drift term, so I is a Martingale and the result $\langle I(t) \rangle = I(0)$ was pre-ordained. Also note that authors often take Brownian motion to have $W(0) = 0$ without comment, simplifying the above results.

Example 4. Find the mean, variance and distribution of $I(t)$, defined by

$$dI(t) \equiv \alpha(t)dW, \qquad I(0) = 0. \tag{9.7}$$

Writing

$$I(t) = \int_0^t \alpha(t')dW(t') = \lim_{n\to\infty} \sum_{i=0}^{n-1} \alpha(t_i) \left[W(t_{i+1}) - W(t_i) \right], \tag{9.8a}$$

we see that the i^{th} term in the sum is a Gaussian random variable with mean 0 and variance $\alpha^2(t_i)\,[t_{i+1}-t_i]$. Since the Gaussians are independent, the sum is also Gaussian with mean zero and variance

$$\left\langle I^2(t)\right\rangle = \sum_{i=0}^{n-1}\alpha^2(t_i)\,[t_{i+1}-t_i] \to \int_0^t \alpha^2(t')dt' \tag{9.8b}$$

Alternatively, one could recognize that the integral is just the sum over independent Gaussians, and conclude that the integral is also Gaussian. Since dI does not have a drift term, one knows it is a Martingale, so the mean value of $I(t)$ must be $I(0)=0$. Finally, one could derive the variance by using Ito's formula:

$$dI^2 = 2IdI + \tfrac{1}{2}\left(dI\right)^2 = 2\alpha(t)IdW + \alpha^2(t)dt. \tag{9.9a}$$

Taking the expected value yields

$$\left\langle dI^2(t)\right\rangle = d\left\langle I^2(t)\right\rangle = \left\langle 2\alpha(t)IdW\right\rangle + \left\langle \alpha^2(t)dt\right\rangle = \alpha^2(t)dt, \tag{9.9b}$$

since the expected value of "anything$\cdot dW$" is zero by the independent increment property of dW combined with the mean value of dW being 0. Thus,

$$\left\langle I^2(t)\right\rangle = \int_0^t \alpha^2(t')dt'. \tag{9.9c}$$

Example 5. Log normal prices. Suppose

$$\frac{dS}{S} = \mu dt + \sigma dW. \tag{9.10}$$

(This is often written as $dS = \mu S dt + \sigma S dW$.). Define $X = \log S$, and solve for X. Find the mean, variance, and distribution of X. Use this to find the mean and variance of $S = e^X$.
Using Ito's formula,

$$\begin{aligned} dX &= \frac{1}{S}dS + \tfrac{1}{2}\left(-\frac{1}{S^2}\right)(dS)^2 \\ &= \mu dt + \sigma dW + \tfrac{1}{2}\left(-\frac{1}{S^2}\right)\sigma^2 S^2 dW^2 \\ &= \left(\mu - \tfrac{1}{2}\sigma^2\right)dt + \sigma dW. \end{aligned} \tag{9.11a}$$

Therefore

$$X(t) = X(0) + \left(\mu - \tfrac{1}{2}\sigma^2\right)t + \sigma\left[W(t) - W(0)\right]. \tag{9.11b}$$

Since $W(t)-W(0)$ is Gaussian with mean zero and variance t, clearly $X(t)$ is Gaussian with mean $X(0) + \left(\mu - \tfrac{1}{2}\sigma^2\right)t$ and variance $\sigma^2 t$. Since $S(0) = e^{X(0)}$, clearly

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma[W(t)-W(0)]}. \tag{9.12a}$$

Let $x = W(t) - W(0)$. Since x is Gaussian with mean 0 and variance t, the n^{th} moment of $S(t)$ is

$$\begin{aligned} \left\langle S^n(t)\right\rangle &= \frac{1}{\sqrt{2\pi}}\int e^{-x^2/2}S^n(0)e^{n(\mu-\frac{1}{2}\sigma^2)t + n\sigma x}dx \\ &= S^n(0)e^{n(\mu-\frac{1}{2}\sigma^2)t}\frac{1}{\sqrt{2\pi}}\int e^{n\sigma x - x^2/2}dx \\ &= S^n(0)e^{n(\mu-\frac{1}{2}\sigma^2)t + \frac{1}{2}n^2\sigma^2 t}. \end{aligned} \tag{9.12b}$$

The mean value is

$$\bar{S}(t) \equiv \langle S(t) \rangle = S(0) e^{(\mu - \frac{1}{2}\sigma^2)t + \frac{1}{2}\sigma^2 t}, \tag{9.13a}$$

so we can simplify

$$\langle S^n(t) \rangle = \left[\bar{S}(t)\right]^n e^{\frac{1}{2}n(n-1)\sigma^2 t}. \tag{9.13b}$$

In particular,

$$\mathrm{Var}\left\{S(t)\right\} = \left\langle S^2 \right\rangle - \bar{S}^2 = \bar{S}^2 \left(e^{\sigma^2 t} - 1\right) = S^2(0) e^{(2\mu - \sigma^2)t + \sigma^2 t} \left(e^{\sigma^2 t} - 1\right). \tag{9.13c}$$

Example 6. Normal prices. Suppose we switch the model to

$$dS = \mu S dt + \sigma dW. \tag{9.14}$$

(Note that there is no S multiplying dW. This is often called the *normal* model for an asset price). Find the mean, variance, and distribution of $S(t)$. Hint: Define $A(t) = S(t)e^{-\mu t}$.
Clearly,

$$dA = e^{-\mu t} dS - \mu S e^{-\mu t} dt, \tag{9.15a}$$

and substituting for dS yields

$$dA = e^{-\mu t} \sigma dW. \tag{9.15b}$$

From example 4, $A(t)$ is Gaussian variable with mean $A(0) = S(0)$ and variance

$$\mathrm{Var}\left\{A(t)\right\} = \int_0^t \sigma^2 e^{-2\mu t} dt = \sigma^2 \frac{1 - e^{-2\mu t}}{2\mu}. \tag{9.15c}$$

Therefore $S(t) = A(t)e^{\mu t}$ is a Gaussian variable with

$$\mathrm{E}\left\{S(t)\right\} = S(0) e^{\mu t}, \tag{9.16a}$$

$$\mathrm{Var}\left\{S(t)\right\} = e^{2\mu t} \sigma^2 \frac{1 - e^{-2\mu t}}{2\mu} = \sigma^2 \frac{e^{2\mu t} - 1}{2\mu}. \tag{9.16b}$$

10. Backward's Kolmogorov equation. Consider a discrete lattice model in which S_j is the value of S on the j^{th} time step. Let s_j^k be the values that S_j can take on the lattice. Suppose we are trying to price a European option with the payoff $F(S_N)$ at the N^{th} time step. Recall that calculating the option price usually required a backward induction (rollback) step in which one had to find the expected value of the payoff at step N, given that S is at the value s_n^k at step n. That is,

$$V(s_n^k) = \mathrm{E}\left\{F(S_N) \mid S_n = s_n^k\right\}. \tag{10.1}$$

Once we found $V(s_n^k)$ for all states k, then we would average to find $V(s_{n-1}^k)$ at the previous step.

The backward's Kolmogorov equation, and its cousin, the Feynman-Kac equation, are powerful methods for carrying out this procedure in continuous time. Suppose that $X(t)$ evolves according to the Ito process

$$dX = a(t, X)dt + b(t, X)dW, \tag{10.2a}$$

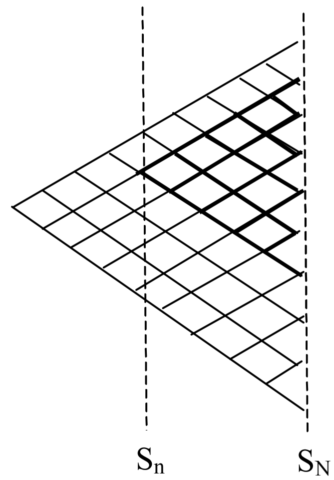

FIG. 10.1.

and suppose we are given the task of finding the expected value of some payoff $F(X(T))$ at date T. Clearly the expected value depends on the current value of $X(t)$, so

$$V(t,x) = \mathrm{E}\left\{F(X(T)) \mid X(t) = x\right\}. \tag{10.2b}$$

One method of evaluating $V(t,x)$ is to use a Monte Carlo procedure. One discretizes time,

$$t_0, t_1, t_2 \ldots, t_K \tag{10.3a}$$

where $t_0 = t$ and $t_K = T$, and writes

$$X(t_0) = x, \tag{10.3b}$$

$$X(t_{k+1}) = X(t_k) + a(t_k, x_k)\Delta t + b(t_k, X_k)\sqrt{\Delta t_k}\xi_k. \tag{10.3c}$$

One creates each path by selecting K independent Gaussian variables $\xi_0, \xi_1, \ldots$ with mean zero and variance 1, and calculates the payoff $F(X(T))$ for that path. One then averages the payoff over all the paths j

$$V(t,x) = \frac{1}{n_{paths}} \sum_{path\ j} F(X_j(T)) \tag{10.3d}$$

This converges to the answer as $n_{paths} \to \infty$. Unfortunately, the convergence is slow, with errors $O(1/\sqrt{n_{paths}})$. $\blacksquare$

A much more effective method is provided by the following theorem.

THEOREM 10.1. *(Backwards Kolmogorov equation). Suppose that*

$$V(t,x) = \mathrm{E}\left\{F(X(T)) \mid X(t) = x\right\}, \tag{10.4a}$$

where X(t) is the Ito process

$$dX = a(t,X)dt + b(t,X)dW. \tag{10.4b}$$

Then V(t,x) is the (unique) solution of the equation

$$\frac{\partial V}{\partial t} + a(t,x)\frac{\partial V}{\partial x} + \tfrac{1}{2}b^2(t,x)\frac{\partial^2 V}{\partial x^2} = 0 \qquad \textit{for } t < T \tag{10.5a}$$

satisfying the condition

$$V(t,x) = F(x) \qquad \text{at } t = T \tag{10.5b}$$

The backward's Kolmogorov equation is useful because there is a century of experience in solving PDEs. Partial differential equation can be solved by finite difference methods or finite element methods, and there is enough experience to be assured of selecting the best methods. In the right circumstances, one can also use Green's functions, Laplace transforms, Fourier transforms, singular perturbation techniques, ... to great effect. Lattice methods are a special case of (explicit) finite difference methods, and are rarely competitive with standard finite difference techniques. In most circumstances Monte Carlo methods are inferior to finite difference or finite element methds.

To establish the Kolmogorov equation, define the *transition density* $p(t_1, x_1; t_2, x_2)$ as the probability density

$$p(t_1, x_1; t_2, x_2)dx_2 = \text{prob}\{x_2 < X(t_2) < x_2 + dx_2 \mid X(t_1) = x_1\} \tag{10.6}$$

Then for any $t < T$ we have

$$V(t,x) = \int_{-\infty}^{\infty} p(t,x;T,X)F(X)dX. \tag{10.7}$$

Clearly for any s in $t < s < T$, we have

$$p(t,x;T,X) \equiv \int_{-\infty}^{\infty} p(t,x;s,y)p(s,y;T,X)dy \tag{10.8}$$

(This is called the Chapman-Kolmogorov equation for Markovian processes). Consequently,

$$V(t,x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(t,x;s,y)p(s,y;T,X)F(X)dydX, \tag{10.9}$$

and inverting the order of integration yields

$$V(t,x) = \int_{-\infty}^{\infty} p(t,x;s,y)V(s,y)dy. \tag{10.10}$$

Aside: Observe that this equation is the continuous analog of the procedure used for discrete lattices, where we averaged over the values $V(t_n, y)$ to get the values at the preceding time t_{n-1}.

Let us now set s very close to t :

$$s = t + dt, \qquad y = x + dX. \tag{10.11a}$$

Then

$$y - x = dX = \int_t^{t+dt} a(t', X(t'))dt' + \int_t^{t+dt} b(t', x')dW(t'). \tag{10.11b}$$

Clearly $dX = O(dt^{1/2})$. The expected value

$$\begin{aligned}\int (y-x)p(t,x;s,y)dy &= \langle y - x\rangle \\ &= \int_t^{t+dt} \langle a(t', X(t'))\rangle\, dt' = a(t,x)dt + O(dt^{3/2}).\end{aligned} \tag{10.12a}$$

$$\int (y-x)^2 p(t,x;s,y)dy = \left\langle (y-x)^2 \right\rangle$$
$$= \left\langle \left(\int_t^{t+dt} b(t',x')dW(t') \right)^2 \right\rangle + O(dt^2)$$
$$= b^2(t,x) \left(\int_t^{t+dt} dW(t') \right)^2 + O(dt^{3/2})$$
$$= b^2(t,x)dt + O(dt^{3/2}). \tag{10.12b}$$

Expanding,

$$V(s,y) = V(t+dt, x+[y-x])$$
$$= V(t,x) + \frac{\partial V(t,x)}{\partial t}dt + (y-x)\frac{\partial V(t,x)}{\partial x} + \tfrac{1}{2}(y-x)^2 \frac{\partial^2 V(t,x)}{\partial x^2} + O(dt^{3/2}). \tag{10.13}$$

Substituting this into eq. 10.10 yields

$$V(t,x) = \int_{-\infty}^{\infty} p(t,x;s,y) \left\{ V(t,x) + \frac{\partial V(t,x)}{\partial t}dt \right.$$
$$\left. + (y-x)\frac{\partial V(t,x)}{\partial x} + \tfrac{1}{2}(y-x)^2 \frac{\partial^2 V(t,x)}{\partial x^2} + O(dt^{3/2}) \right\} dy$$
$$= V(t,x) + \frac{\partial V(t,x)}{\partial t}dt + \langle y-x \rangle \frac{\partial V(t,x)}{\partial x} + \tfrac{1}{2} \left\langle (y-x)^2 \right\rangle \frac{\partial^2 V(t,x)}{\partial x^2} + O(dt^{3/2}). \tag{10.14}$$

Subtracting $V(t,x)$ off both sides, dividing by dt and letting $dt \to 0$ then yields

$$\frac{\partial V}{\partial t} + a(t,x)\frac{\partial V}{\partial x} + \tfrac{1}{2}b^2(t,x)\frac{\partial^2 V}{\partial x^2} = 0 \qquad \text{for } t < T. \tag{10.15a}$$

Clearly at $t = T$ we have $X = x$ with certainty, so

$$V(t,x) = F(x) \qquad \text{for } t < T. \tag{10.15b}$$

The theory of PDEs guarantees that there is only one solution to this boundary value problem.

10.1. Feynman-Kac. Commonly one ends up solving

$$V(t,x) = \mathrm{E}\left\{ e^{-\int_t^T r(t',X(t'))dt'} F(X(T)) \mid X(t) = x \right\},$$

in finance, since one usually has to discount the payoffs. This is handled by a variant of the backwards Kolmogorov equation, the Feynman-Kac equation:

Theorem 10.2. *(Feynman-Kac equation). Suppose that*

$$V(t,x) = \mathrm{E}\left\{ e^{-\int_t^T r(t',X(t'))dt'} F(X(T)) \mid X(t) = x \right\}, \tag{10.16a}$$

where X(t) is the Ito process

$$dX = a(t,X)dt + b(t,X)dW. \tag{10.16b}$$

Then V(t,x) is the (unique) solution of the equation

$$\frac{\partial V}{\partial t} + a(t,x)\frac{\partial V}{\partial x} + \tfrac{1}{2}b^2(t,x)\frac{\partial^2 V}{\partial x^2} = r(t,x)V \qquad \textit{for } t < T \tag{10.17a}$$

satisfying the condition

$$V(t,x) = F(x) \qquad \textit{at } t = T \tag{10.17b}$$

Apart from the $r(t,x)V$ on the right hand side of the equation, this is identical to the backward's Kolmogorov equation.

To derive this formula, we write

$$V(t,x) = \mathrm{E}\left\{e^{-\int_t^{t+dt} r(t',X(t'))dt'}e^{-\int_{t+dt}^T r(t',X(t'))dt'}F(X(T)) \mid X(t) = x\right\}. \tag{10.18}$$

But

$$e^{-\int_t^{t+dt} r(t',X(t'))dt'} = [1 - r(t,x)dt] + O(dt^{3/2}). \tag{10.19}$$

Since the right hand side is deterministic (apart from the $O(dt^{3/2})$ error term), we can pull it outside the expected value::

$$V(t,x) = [1 - r(t,x)dt]\,\mathrm{E}\left\{e^{-\int_{t+dt}^T r(t',X(t'))dt'}F(X(T)) \mid X(t) = x\right\} + O(dt^{3/2}). \tag{10.20}$$

We now use iterated expected value (telescope rule for expected value). Consider the expected value

$$U(t+dt,y) = \mathrm{E}\{A \mid X(t+dt) = y\} \tag{10.21a}$$

for some function A, given that we are at $X(t+dt) = y$. If we take the expected value of this expected value over y, given that $X(t) = x$,

$$\mathrm{E}\{U(t+dt,y) \mid X(t) = x\} = \mathrm{E}\{\mathrm{E}\{A \mid X(t+dt) = y\} \mid X(t) = t\}, \tag{10.21b}$$

the law of the iterated expected value states that this is identical to taking the expected value given $X(t) = x$

$$\mathrm{E}\{\mathrm{E}\{A \mid X(t+dt) = y\} \mid X(t) = t\} = \mathrm{E}\{A \mid X(t) = x\}. \tag{10.22}$$

I.e, one can dispense with the intermediate point y.

For our problem, we write

$$\begin{aligned} &\mathrm{E}\left\{e^{-\int_{t+dt}^T r(t',X(t'))dt'}F(X(T)) \mid X(t) = x\right\} \\ &= \mathrm{E}\left\{\mathrm{E}\left\{e^{-\int_{t+dt}^T r(t',X(t'))dt'}F(X(T)) \mid X(t+dt) = y\right\} \mid X(t) = x\right\} \\ &= \mathrm{E}\{V(t+dt,y) \mid X(t) = x\} \end{aligned} \tag{10.23}$$

Substituting this into eq. 10.20, we obtain

$$V(t,x) = [1 - r(t,x)dt]\,\mathrm{E}\{V(t+dt,y) \mid X(t) = x\} + O(dt^{3/2}) \tag{10.24}$$

Following the derivation of the backward's Kolmogorov equation, we can now easily show that the same steps as in the derivation of

$$\mathrm{E}\left\{V(t+dt,y) \mid X(t)=x\right\} = V + \left\{\frac{\partial V}{\partial t} + a(t,x)\frac{\partial V}{\partial x} + \tfrac{1}{2}b^2(t,x)\frac{\partial^2 V}{\partial x^2}\right\}dt + O(dt^{3/2}) \tag{10.25}$$

where the argument of V is t, x. Substituting this into 10.24, subtracting off the $V(t,x)$ term, dividing by dt, and letting $dt \to 0$ then establishes the result:

$$\frac{\partial V}{\partial t} + a(t,x)\frac{\partial V}{\partial x} + \tfrac{1}{2}b^2(t,x)\frac{\partial^2 V}{\partial x^2} = r(t,x)V \tag{10.26}$$

11. Final notes: Non-Markovian processes. Throughout this lecture we have considered Ito processes of the form

$$dX = a(t,X)dt + b(t,X)dW, \tag{11.1}$$

in which the coefficients $a(t,X)$ and $b(t,X)$ depend only on the current value of $X(t)$. These are known as *Markovian* Ito processes. Although most models are Markovian, this is needlessly restrictive. The coefficients a and b can depend on averages of X, on historical values of X, on other Ito processes, ... Apart from innocuous measurability restrictions, the only real restriction is that at date t, the coefficients a and b cannot depend on *any future events*. Often these general coefficents are written as $a(t,\omega)$ and $b(t,\omega)$, where ω represents, roughly speaking, measureable events resolvable by time t.

We note that the all-important Ito's formula remains correct for non-Markovian Ito processes, and the box formulas remain valid. However, for non-Markovian processes, the expected value $\mathrm{E}\left\{F(X(T)) \mid X(t)=x\right\}$ may depend not only on the value x at t, but may depend on the history of $X(t)$ before date t. (Consider an Ito process with b depending on the three year running average of $X(t)$, for example). This means that the backward's Kolmogorov equation and the Feynman-Kac formula *do not* remain valid in general.