SDES

1. Examples/Applications

Some convincing

1.1. Noisey LRC Circuit. The charge $Q(t)$ at time t in a capacitor an LRC circuit satisfies

$$LQ''(t) + RQ'(t) + \frac{1}{C}Q(t) = F(t), \; Q(0) = Q_0, \; Q'(0) = I_0 \tag{1.1}$$

where $F(t)$ emf driving the circuit at time t. Suppose $F(t) = G(t) + \text{"noise"}$. How to solve for $Q(t)$?

1.2. Noisey Measurement. Suppose that you take measurements of $Q(t)$ but there's inherent noise in those measurements as well, so you're actually measuring

$$Z(s) = Q(s) + \text{"noise"}$$

What is the best estimate of $Q(t)$ satisfying eqn 1.1 based on $Z(s)$? Filter the noise away using a Kalman-Bucy filter.

1.3. Optimal Stopping. Suppose an asset (stock, commodities, real estate) is such that its price X_t on the open market varies according to a stochastic differential equation

$$\begin{aligned} \frac{dX_t}{dt} &= X_t(r+\alpha)X_t \cdot \text{"noise"} \\ &= X_t(r+\alpha \cdot \text{"noise"}) \end{aligned}$$

where r, α are known constants and the discount rate ρ is also a known constant (present value of future cash flows - money today is worth more than money tomorrow because of inflation and etc.). Given that you know X_s up to present time t and taking into consideration that the rate of change is noisey, at what time should you sell this asset? Since noisey we can only hope for optimal stopping strategy that maximizes expected returns (when inflation is taken into account). Can be solved by solving a corresponding semielliptic second order PDE with Dirichlet boundary conditions (function is specified on boundary, as opposed to derivatives specified on boundary [Von Neumann bvp])

1.4. Stochastic Control. Consider the risky investment from problem 3

$$\frac{dp_1}{dt} = (a + \alpha \cdot \text{"noise"})\, p_1$$

and consider a safe investment whose values grows exponentially

$$\frac{dp_2}{dt} = bp_2$$

Zero-sum capital: at each t you can invest a fraction u_t of total capital X_t in p_1 and $(1-u_t)X_t$ in p_2. Given a utitlity function U and stopping time T how to find the optimal portfolio $u_t \in [0,1]$ such that

$$\max_{u_t \in [0,1]} \left\{ E\left[U\left(X_T^{(u)} \right) \right] \right\}$$

1.5. Options Pricing. Suppose at some time t the person in problem 4 is offered the right (without obligation) to buy one unit of the risky asset at a specified price K at a specified future date $t = T$. Such a right/asset is called a *European call option*. How much should they be willing to pay for the option? Problem solved by Fischer Black and Myron Scholes - called the Black-Scholes equation for option pricing

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

where V is the price of the option as a function of the price of the asset, r is the risk-free interest rate (free money - tbills for example), and σ is the volatility of the stock.

2. Measure Theory Refresher

Definition 1. Ω is a set called the sample/event space. A σ-algebra $\mathcal{F}$ is a family $\mathcal{F}$ of subsets of Ω with the following properties

(1) $\emptyset \in \mathcal{F}$
(2) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$ i.e. $\mathcal{F}$ is closed under complementation.
(3) $A_i \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ i.e. $\mathcal{F}$ is closed under countable unions.

Definition 2. The pair $(\Omega, \mathcal{F})$ is called a measurable space and the subsets F of Ω which belond to $\mathcal{F}$ are called $\mathcal{F}$-measurable sets.

Definition 3. A probability measure P on a measurable space $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that

(1) $P(\emptyset) = 0, \ P(\Omega) = 1$
(2) For A_i such that $A_i \cap A_j = \emptyset$ for $i \neq j$ $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} P(A_i)$, called countable additivity.

Definition 4. The triple $(\Omega, \mathcal{F}, P)$ is appropriately called a measure space.

Definition 5. Given any family $\mathcal{U}$ of subsets of Ω there is a smallest σ-algebra $\mathcal{H}_{\mathcal{U}}$ containing $\mathcal{U}$

$$\mathcal{H}_{\mathcal{U}} = \bigcap \{\mathcal{H} : \mathcal{H}\, \sigma\text{-algebra of } \Omega, \mathcal{U} \subset \mathcal{H}\}$$

(intersection of all sigma algebras containing $\mathcal{U}$)

Definition 6. If $\mathcal{U}$ is the collection of all open subsets of a topological space Ω then $\mathcal{B} = \mathcal{H}_{\mathcal{U}}$ is called the *Borel σ-algebra* or topological *σ-algebra* and $B \in \mathcal{B}$ are called the *Borel sets.*

Definition 7. Let $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$ be measure spaces. Then a function $X : \Omega \to \Omega'$ is $\mathcal{F} - \mathcal{F}'$*-measurable* if for every $U \in \mathcal{F}'$

$$X^{-1}(U) = \{\omega \in \Omega; X(\omega) \in U\} \in \mathcal{F}$$

I.e. pre-images of measurable sets in the *σ-algebra* associated with the codomain are measurable sets in the *σ-algebra* associated with the domain. For a measure space $(\Omega, \mathcal{F}, P)$ an $\mathcal{F} - \mathcal{B}$*-measurable* function $X : \Omega \to \mathbb{R}^n$ is such that

$$X^{-1}(U) = \{\omega \in \Omega; X(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbb{R}^n$(this is because the open sets are the elements of the Borel *σ-algebra*).

Definition 8. If $X : \Omega \to \mathbb{R}^n$ is **any** function, then the σ-algebra $\mathcal{H}_X$ *generated* by X is the smallest σ-algebra that makes X *measurable*, i.e. containing all inverse images of open sets in $\mathbb{R}^n$.

It's actually even smaller

$$\mathcal{H}_X = \left\{X^{-1}(B); B \in \mathcal{B}\right\}$$

and X becomes $\mathcal{H}_X$-measurable.

Definition 9. A random variable is an $\mathcal{F}$*-measurable* function $X : \Omega \to \mathbb{R}^n$.

Every random variable induces a probability measure μ_X on $\mathbb{R}^n$, defined byin

$$\mu_X(B) = P\left(X^{-1}(B)\right)$$

where $B \in \mathcal{B}$ and $\mathcal{B}$ is the σ-algebra over $\mathbb{R}^n$. μ_X is called the *distribution* or *law of X.*

2.1. Lebesgue Integration aside.

Fact 10. *If for a function $X : \Omega \to \Omega'$*

$$X^{-1}(U) \in \mathcal{F}$$

for all $U \in \mathcal{G}$ where $\mathcal{G}$ is the collection of sets that generate $\mathcal{F}'$ then X is $\mathcal{F} - \mathcal{F}'$-measurable. I.e. measurability only needs to be verified on generating sets.

One way to generate the Borel σ-algebra over $\mathbb{R}$ to take all intervals of form (c, ∞) with $c \in \mathbb{R}$. Therefore if

$$X^{-1}((c, \infty)) = \{\omega \in \Omega; X(\Omega) > c\} \in \mathcal{F}$$

Now you can think of this as whatever you want (counting how much a function gets thrown into a bin but it's purely a result of the fact that measurability of a function need only be verified on the generating set of sets of codomain σ-algebra, i.e. there are alternative characterizations due to the same fact like that any of

$$\{\omega \in \Omega; X(\Omega) \geq c\}, \{\omega \in \Omega; X(\Omega) < c\}, \{\omega \in \Omega; X(\Omega) \leq c\}$$

should be measurable.

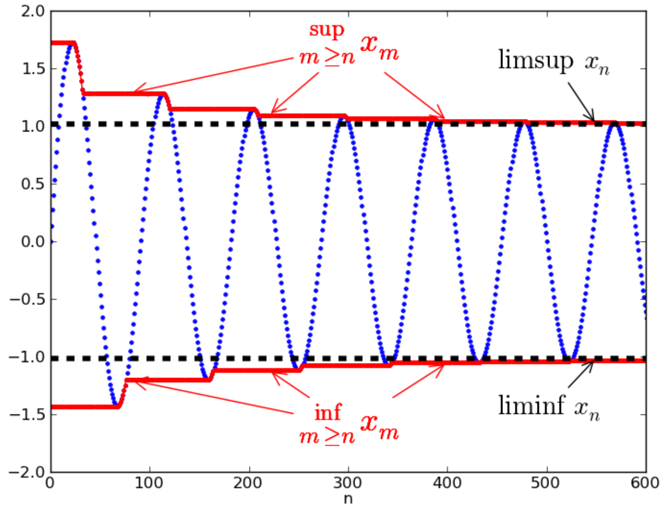

FIGURE 2.1. lim inf and lim sup

2.1.1. lim inf *and* lim sup. For real numbers x_n

$$\liminf_{k\to\infty} x_k := \sup_{k\geq 0}\left(\inf_{m\geq k} x_m\right)$$

$$\limsup_{k\to\infty} x_k := \inf_{k\to\infty}\left(\sup_{m\geq k} x_m\right)$$

Refer to fig 2.1. For functions $f_k(y)$ let $x_k = f_k(y_0)$.

Fact 11. *The set of measurable functions is closed under algebraic operations and pointwise limits, i.e. if f_k is measurable for every k then*

$$\sup_{k\in\mathbb{N}} f_k, \limsup_{k\in\mathbb{N}} f_k, \liminf_{k\in\mathbb{N}} f_k$$

are all measurable. And if the last two coincide then

$$\lim_{k\to\infty} f_k = f$$

exists and is measurable.

We "build up" the Lebesgue integral piece by piece. Let $(\Omega, \mathcal{F}, P)$ be a measure space with (probability) measure finite positive measure P

(1) Indicator functions. For $S \in \mathcal{F}$

$$\int 1_S(\omega)\, dP := P(S)$$

Note the argument is purely symoblic - the indicator isn't evaluted.

(2) Simple functions. Let $S(\omega) = \sum_k a_k 1_{S_k}$ with $a_k \geq 0$. Then

$$\int S(\omega)\, dP := \sum_k \int 1_{S_k}(\omega)\, dP = \sum_k a_k P(S_k)$$

Note this defintion is independent of representation of f (can be proven using additive property of measures).

(3) Non-negative functions. Let X be a non-negative measurable function defined on Ω. Then

$$\int_\Omega X(\omega)\, dP := \sup_s \left\{ \int_\Omega S(\omega)\, dP;\ 0 \leq S \leq X,\ S \text{ simple} \right\}$$

Why is this reasonable?

Lemma. *Any positive measurable function X is the pointwise limit of simple functions $S_j(\omega)$*

Proof. Fix $j \in \mathbb{N}$ and let

$$A_k^j = \begin{cases} \left\{\omega; k2^{-j} \leq X(\omega) \leq (k+1)\, 2^{-j}\right\} & k = 0, 1, \ldots, j2^j - 1 \\ \{\omega; X(\omega) \geq j\} & k = j2^j \end{cases}$$

I.e. slice up the graph of X. Then

$$S_j(\omega) = \sum_{k=1}^{j2^j} k2^{-j} 1_{A_k^j}(\omega)$$

Then ☐

- $|S_j(\omega) - X(\omega)| \leq 2^{-j}$ for fixed ω and $X(\omega) < j$
- $A_k^j = \{\omega; k2^{-j} \leq X(\omega)\} \cap \{\omega; X(\omega) < (k+1)2^{-j}\} \in \mathcal{F}$
- $\{\omega; X(\omega) \geq j\} \in \mathcal{F}$
- $0 \leq S_j(\omega) \leq X(\omega)$ and $S_j(\omega) \uparrow X(\omega)$

(4) Signed functions. Any signed function X can be written as

$$X(\omega) = X^+(\omega) - X^-(\omega)$$

where

$$X^+(\omega) = \begin{cases} X(\omega) & \text{if } X(\omega) > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$X^-(\omega) = \begin{cases} -X(\omega) & \text{if } X(\omega) < 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that both X^+ and X^- are non-negative measurable functions and that

$$|X| = X^+ + X^-$$

If

$$\min\left\{\int_\Omega X^+ dP, \int_\Omega X^+ dP\right\} < \infty$$

then

$$\int X dP := \int X^+ dP - \int X^- dP$$

and if

$$\int |X| \, dP = \int X^+ dP + \int X^- dP < \infty$$

then X is Lebesgue integrable.

3. Probability

Definition 12. If $\int_\Omega |X(\omega)| \, dP(\omega) < \infty$ then the *expectation* is defined as

$$E[X] := \int_\Omega X(\omega) \, dP(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x)$$

The notation dP is defined by

$$\int 1_S dP = P(S).$$

and note that the first integral is over the sample/event space (hence both X and dP are parameterized by ω). And generally if $f : \mathbb{R}^n \to \mathbb{R}$ is $(\mathcal{B}^n, \mathcal{B})$ measurable and $\int_\Omega |f(X(\omega))| \, dP(\omega) < \infty$ then

$$E[f(X)] := \int_\Omega f(X(\omega)) \, dP(\omega) = \int_{\mathbb{R}^n} f(x) \, d\mu_X(x)$$

Independence is defined in terms of measurable sets.

Definition 13. Two subsets $A, B \in \mathcal{F}$ are called *independent* if

$$P(A \cap B) = P(A) \cdot P(B)$$

A collection $\mathcal{A} = \{\mathcal{H}_i; i \in I\}$(where I is some index set [of any cardinality]) of families $\mathcal{H}_i$ of measurable sets is independent if for all choices of the index tuple $(i_j, \ldots, i_k)$ and for all choices of H in each $\mathcal{H}_{i_j}$

$$P\left(H_{i_j} \cap \cdots \cap H_{i_k}\right) = P\left(H_{i_j}\right) \cdots P\left(H_{i_k}\right)$$

A collection of random variables $\{X_i; i \in I\}$ **is independent** if the collection of generated σ-algebras $\mathcal{H}_{X_i}$ (generated! not induced! domain σ-algebra!) is independent.

Exercise 14. If two random variables $X, Y : \Omega \to \mathbb{R}$ are independent and $(E[|X|], E[|Y|], E[|XY|]) \prec (\infty, \infty\infty)$ then

$$E[XY] = E[X] E[Y]$$

Definition 15. A stochastic process is a parameterized collection of random variables $\{X_t\}_{t \in T}$ with each X_t defined on the same measure space (probability space) $(\Omega, \mathcal{F}, P)$ and $X_t : \Omega \to \mathbb{R}^n$.

The parameter space for T and the range of X_t determine taxonomical classifiction of the process with each possibly being either continuous or discrete.

For fixed t it's the case that $X_t(\omega)$ is just a plain random variable

$$X_t(\omega) : \Omega \to \mathbb{R}^n$$

and for fixed ω it's the case that $X_\omega(t)$ is just a function on $\mathbb{R}^n$

$$X_\omega(t) : T \to \mathbb{R}^n$$

called a *path*. Note that this gives a way to identify a function with each $\omega \in \Omega$ and therefore we may treat Ω as a subset of $\tilde{\Omega} \subset (\mathbb{R}^n)^T$ the space of *all* functions from T to $\mathbb{R}^n$.

Warning: it becomes a little muddled here when I talk about the standard topology on $\mathbb{R}^n$, which I call $\mathcal{B}^n$, and the topological σ-algebra on $\mathbb{R}^n$, which I also call $\mathcal{B}^n$. I will have to come back to this and clear it up.

3.1. Product Topology aside.

Definition 16. Let X be a set and Y a topological space. Given $x \in X$ and any open set $U \subset Y$, define

$$S(x, U) = \left\{ f \in Y^X ; f(x) \in U \right\}$$

This is a set of functions! And the universal quantifier is over x and U. The *product topology* on Y^X is the topology generated by the subbasis consisting of the sets S.

Note that $\{S\}$ are subbasis not a basis. To form basic sets you take finite intersections of subbasic sets (and then to form sets in the topology you take arbitrary unions of basic sets).

Fact 17. *It is sufficient to only consider basis sets U. I.e.*

$$\{S(x, U) ; x \in X, U \in \tau\}$$

where τ is a basis for the topology on Y, is a subbasis for the product topology on Y^X.

This definition coincides with the product topology on the Cartesian product of two topological spaces (hence the name)

Example 18. If Y is a topological space, then the product $Y \times Y$ can be viewed as a function fome Y^X, where $X = \{1, 2\}$. If $U \subset Y$ is open, then $S(1, U)$ is the set of all functions from X to Y that map 1 to an element of U and map 2 to some element of Y. Clearly then $S(1, U)$ can identified with all tuples (a, b) where $a \in U$ and $b \in Y$. Hence

$$S(1, U) \cong U \times Y$$

Similarly

$$S(2, U) \cong Y \times U$$

These sets form a subbasis for the product topology on the Cartesian product $Y \times Y$.

3.2. Stochastic Processes.

Given that we've identified Ω with $\tilde{\Omega}$ can we identify $\mathcal{F}$ with some σ-algebra $\tilde{\mathcal{F}}$ on $\tilde{\Omega}$? Recall that the topological σ-algebra on a set is just the collection of all open subsets of that set. Hence the sets

$$\{\omega_X ; \omega_X(t_1) \in B_1, \ldots, \omega_X(t_k) \in B_k\}, \ B \in \mathcal{B}^n$$

where $\omega_X(t) = X_\omega(t)$, generate the σ-algebra $\tilde{\mathcal{F}}$ on $\tilde{\Omega}$. Why? The sets

$$S(t, B) = \left\{ \omega_X \in \tilde{\Omega} = (\mathbb{R}^n)^T ; \omega_X(t) \in B \right\}, \ B \in \mathcal{B}^n$$

form a subbasis for the product topology on $\tilde{\Omega}$ (by fact 17 since $B \in \mathcal{B}^n$ are basic sets for the topology [topological σ-algebra is generated by basic sets in the topology]) and so finite intersections

$$S(t_1, B_1) \cap \cdots \cap S(t_k, B_k) = \left\{ \omega_X \in \tilde{\Omega} = (\mathbb{R}^n)^T ; \omega_X(t_1) \in B_1, \ldots, \omega_X(t_k) \in B_k \right\}$$

form a basis for the product topology on $\tilde{\Omega}$ and unions of those then form the actual elements of the product topology and hence generate the σ-algebra $\tilde{\mathcal{F}}$ on $\tilde{\Omega}$ (is the basis for a topology also the generator for the σ-algebra for the topological σ-algebra?). **Hence** it's reasonable to say that a stochastic process is a probability measure $\tilde{P}$ on the measurable space$\left(\tilde{\Omega} = (\mathbb{R}^n)^T, \tilde{\mathcal{F}}\right)$, called the *law of the process.* If $\Phi_X : \Omega \to \tilde{\Omega}$, where X is the process, defined by currying $X(t, \omega) = X_t(\omega)$

$$(\Phi_X(\cdot))(t) = X(t, \cdot) = X_t(\cdot)$$

then the law$\mathcal{L}_X$ is the *pushforward* of the measure on Ω

$$\mathcal{L}_X := (\Phi_X)_* (P) = P \circ \Phi_X^{-1}$$

The process also induces induces a set of measures on its range

Definition 19. The *finite-dimensional distributions* of the process $X = \{X_t\}_{t \in T}$ are the measures

$$\mu_{t_1, \ldots, t_k}(F_1 \times F_2 \times \cdots \times F_k) = P(X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k)$$

on $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k}$ (cartesian product), with $F_i \in \mathcal{B}^n$ being the σ-algebras for each $\mathbb{R}^n$.

Note that these are measures on the product ranges of X_t, not on the domains (different from $\tilde{\mathcal{F}}$ discussed above).

These measures determine many, but not all (and in some important cases crucially), properties the process X. Conversely given a family of measures $\{\mu_{t_1,\dots,t_k}; k \in \mathbb{N}, t_i \in T\}$ (all finite product measures indexed by all finite subsets of T) on $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{k}$ one can construct a stochastic process $X = \{X_t\}_{t\in T}$ which will have $\{\mu_{t_1,\dots,t_k}; k \in \mathbb{N}, t_i \in T\}$ as its finite-dimensional distributions given some "natural" consistency conditions.

Theorem 20. *Kolmogorov's extension theorem. For all finite tuples $(t_1, \dots, t_k)$ with $t_i \in T$ let $\mu_{t_1,\dots,t_k}$ be the probability measures defined above. If $\mu_{t_1,\dots,t_k}$ satisfy the consistency conditions*

(1) *Invariance under permutation: for all permutations p on $\{1, 2, \dots, k\}$*

$$\mu_{p(t_1),\dots,p(t_k)} (F_1 \times \cdots \times F_k) = \mu_{t_1,\dots,t_k} \left(F_{p^{-1}(1)} \times \cdots \times F_{p^{-1}(k)}\right)$$

(2) *Invariance under marginalization over probability 1 events:*

$$\mu_{t_1,\dots,t_k} (F_1 \times \cdots \times F_k) = \mu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}} \left(F_1 \times \cdots \times F_k \times \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{m} \right)$$

then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $X = \{X_t\}_{t\in T}$ on Ω with $\mu_{t_1,\dots,t_k}$ as its finite-dimensional distributions.

Kolmogorov's extension is actually an if and only if: any stochastic process trivially has these properties. The power of Kolmogorov extension is that the typical workflow is to define a stochastic process by its finite-dimensional distributions and that is completely sufficient[1].

4. Canonical Brownian Motion

4.1. **Heuristic PDE.** In 1828 Robert Brown observed pollen in a suspension exhibiting irregular motion. Einstein later modeled the motion as random collisions with the molecules of water. Here is a refined version of an argument that leads to the same model (1 dimensional model): imagine a 2 dimensional lattice indexed by $m = 0, \pm 1, \pm 2, \dots$ and $n = 0, 1, 2, \dots$ with spacings Δx in the m direction and Δt in the n direction (for some small fixed Δx and Δt). The m direction is spatial (the pollen particle moves either left or right on a line) and the n direction is time (time moves "forward"). Consider the pollen particle starting at $(m, n) = (0, 0)$. At each $n\Delta t$ time m increases by 1 with probability 1/2 or decreases by 1 with probability 1/2 (moves left or right on the m axis). Let $p(m\Delta x, n\Delta t)$ be the probability that the particle is at position $m\Delta x$ at time $n\Delta t$. Then

$$p(m\Delta x, 0) = \begin{cases} 0 & \text{if } m\Delta x \neq 0 \\ 1 & \text{if } m\Delta x = 0 \end{cases}$$

This type of process is called a random walk on $\mathbb{Z}$ or just a one dimensional random walk. Recursively

$$p(m\Delta x, (n+1)\Delta t) = \frac{1}{2} p((m+1)\Delta x, n\Delta t) + \frac{1}{2} p((m-1)\Delta x, n\Delta t)$$

and therefore

$$\begin{aligned} p(m\Delta x, (n+1)\Delta t) - p(m\Delta x, n\Delta t) &= \frac{1}{2} p((m+1)\Delta x, n\Delta t) + \frac{1}{2} p((m-1)\Delta x, n\Delta t) - p(m\Delta x, n\Delta t) \\ &= \frac{1}{2} \left(p((m+1)\Delta x, n\Delta t) - 2p(m\Delta x, n\Delta t) + p((m-1)\Delta x, n\Delta t)\right) \end{aligned}$$

Now let $D := (\Delta x)^2 / \Delta t$ and then

$$\frac{p(m\Delta x, (n+1)\Delta t) - p(m\Delta x, n\Delta t)}{\Delta t} = \frac{D}{2} \left(\frac{p((m+1)\Delta x, n\Delta t) - 2p(m\Delta x, n\Delta t) + p((m-1)\Delta x, n\Delta t)}{(\Delta x)^2} \right)$$

As $(\Delta t, \Delta x, m\Delta x, n\Delta t) \to (0, 0, x, t)$ at a rate such that D is preserved then

$$\begin{aligned} \frac{p(m\Delta x, (n+1)\Delta t) - p(m\Delta x, n\Delta t)}{\Delta t} &\to \frac{\partial}{\partial t} p(x,t) \\ \frac{p((m+1)\Delta x, n\Delta t) - 2p(m\Delta x, n\Delta t) + p((m-1)\Delta x, n\Delta t)}{(\Delta x)^2} &\to \frac{\partial^2}{\partial x^2} p(x,t) \\ &\Rightarrow \\ \frac{\partial}{\partial t} p(x,t) &= \frac{D}{2} \frac{\partial^2}{\partial x^2} p(x,t) \end{aligned}$$

This is a well studied Partial Differential Equation called the diffusion equation (or heat equation) with diffusion constant D and solution

$$p(x,t) = \frac{1}{\sqrt{2\pi D t}} e^{-\frac{1}{2}\frac{x^2}{Dt}}$$

[1]Kolmogorov extension guarnatees there is a correspondent probability space and process.

Therefore the interpretation is that for a particle starting $x=0$ at $t=0$ the distribution on positions at time $t=t_0$ is Normal $\mathrm{n}\left(0,Dt_0\right)$ distributed. Einstein calculated

$$D=\frac{RT}{N_A f}$$

where

- R is the ideal, or universal, gas constant, equal to the product of the Boltzmann constant and the Avogadro constant.
- T is the temperature of gas/fluid in Kelvin
- f is the friction coefficient between particle and fluid
- N_A is Avogadro's number

4.2. **Rigorous Distribution.** Envisage the same one dimensional random walk on $\mathbb{Z}$, with spatial increments Δx and time increments Δt. Let X_i be Bernoulli random variables iid with

$$P\left(X_i=0\right)=\frac{1}{2} \text{ and } P\left(X_i=1\right)=\frac{1}{2}$$

Then $\operatorname{Var}\left(X_i\right)=\frac{1}{4}$. Define

$$S_n \coloneqq \sum_{i=1}^{n} X_i$$

S_n is the number of steps to the right (not the distance but just the number of steps). Note that by iid

$$ES_n=\sum_{i=1}^{n} EX_i=\frac{n}{2}$$

Then let

$$X\left(n\Delta t\right) \coloneqq \underbrace{S_n\Delta x}_{\text{positive dist}} \underbrace{-\left(n-S_n\right)\left(-\Delta x\right)}_{\text{negative dist}}=\left(2S_n-n\right)\Delta x$$

i.e. the position of the particle at time $n\Delta t$. Then

$$\begin{aligned}
\operatorname{Var}\left(X\left(n\Delta t\right)\right) &= \operatorname{Var}\left(\left(2S_n-n\right)\Delta x\right)\\
&= \left(\Delta x\right)^2\operatorname{Var}\left(\left(2S_n-n\right)\right)\\
&= 4\left(\Delta x\right)^2\operatorname{Var}\left(S_n\right)\\
&= 4\left(\Delta x\right)^2\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right)\\
& \quad X_i \text{ iid } \Rightarrow\\
&= 4\left(\Delta x\right)^2 n\operatorname{Var}\left(X_i\right)\\
&= \left(\Delta x\right)^2 n=\frac{\left(\Delta x\right)^2}{\Delta t}t=Dt
\end{aligned}$$

and

$$\begin{aligned}
X\left(n\Delta t\right) &= \left(2S_n-n\right)\Delta x\\
&= \sqrt{n}\Delta x\left(\frac{S_n-\frac{n}{2}}{\sqrt{n/4}}\right)\\
&= \sqrt{Dt}\left(\frac{S_n-\frac{n}{2}}{\sqrt{n/4}}\right)\\
&= \sqrt{Dt}\left(\frac{\left(\sum_{i=1}^{n} X_i\right)-\frac{n}{2}}{\sqrt{n/4}}\right)\\
&= \sqrt{Dtn}\left(\frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)-\frac{1}{2}}{\sqrt{n/4}}\right)\\
&= \sqrt{Dt}\left[\left(\frac{\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)-\frac{1}{2}}{\sqrt{1/4}/\sqrt{n}}\right)\right]
\end{aligned}$$

The CLT[2] with

$$\sqrt{Dt}\left[\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i-\frac{1}{2}}{\sqrt{1/4}/\sqrt{n}}\right)\right]=\sqrt{Dt}\left[\left(\frac{\bar{X}-EX_i}{\sigma_X/\sqrt{n}}\right)\right]$$

[2]Actually a weaker theorem called Laplace-De Moivre.

implies that

$$\begin{aligned}
\lim_{n\to\infty} P\left(a \le \sqrt{Dt}X(t) \le b\right) &= \lim_{n\to\infty} P\left(a \le \sqrt{Dt}X(t) \le b\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{Dt}}}^{\frac{b}{\sqrt{Dt}}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi Dt}} \int_a^b e^{-\frac{x^2}{2Dt}} dx
\end{aligned}$$

So again $X(t) \sim \mathrm{n}(0, Dt)$.

4.3. **Defining Brownian motion.** Motivated by all this a reasonable definition of Brownian motion might be

Definition 21. A real-valued stochastic process $B(\cdot)$ is called a *Brownian motion* or *Wiener process* if

(1) $B(0) = 0$ almost surely, i.e. $P(B(0) = 0) = 1$.
(2) $B(t) - B(s) \sim \mathrm{n}(0, t-s)$ for $0 \le s \le t$.
(3) for all $0 < t_1 < t_2 < \cdots < t_n$ the random variables $B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent, i.e. increments of the process are independent.

Note that $B(t) = B(t) - B(0) \sim \mathrm{n}(0, t-0) \sim \mathrm{n}(0, t)$ so for a Wiener process $B(t)$

$$E(B(t)) = 0$$

and

$$\begin{aligned}
E\left((B(t))^2\right) &= \mathrm{Var}(B(t)) + E(B(t))^2 \\
&= \mathrm{Var}(B(t)) + 0 \\
&= \mathrm{Var}(\mathrm{n}(0,t)) \\
&= t
\end{aligned}$$

But the real test of Brownian motion is whether we can use Kolmogorov's extension theorem to specify a family of probability measures satisfying properties 1 and 2 of thm. 20. The measures will be chosen in particular to agree with the above results

Fix $x \in \mathbb{R}$ and define

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$$

for $y \in \mathbb{R}$. Then for $0 < t_1 < t_2 < \cdots < t_n$ define a measure $\mu_{t_1,\ldots,t_k}$ on $\mathbb{R}^k$ by

$$\begin{aligned}
\mu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) &= \int_{F_1} \cdots \int_{F_k} p(t_1, x, x_1)\, p(t_2 - t_1, x_1, x_2) \cdots \\
&\quad p(t_k - t_{k-1}, x_{k-1}, x_k)\, dx_1 dx_2 \cdots dx_k
\end{aligned}$$

for $F_i \subset \mathbb{R}$ with the convention that $p(0, x, y)\, dx = \delta_x(y)$ a unit point mass at x. Extend to all finite sequences $t_1, \ldots, t_n$ and since $\int_{F_i} p(t, x, y)\, dx = 1$ Kolmogorov's Extension theorem says there exists $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $B = \{B_t\}_{t\ge 0}$ on Ω such that the finite-dimensional distributions of B_t are

$$\begin{aligned}
P^x(B_{t_1} \in F_1 \times \cdots \times B_{t_k} \in F_k) &= \int_{F_1} \cdots \int_{F_k} p(t_1, x, x_1)\, p(t_2 - t_1, x_1, x_2) \cdots \\
&\quad p(t_k - t_{k-1}, x_{k-1}, x_k)\, dx_1 dx_2 \cdots dx_k
\end{aligned} \tag{4.1}$$

Such a process[3] is called a Brownian motion. The paths of a Brownian motion happen to be almost surely continuous[4] (but nowhere differentiable!) so with each $\omega \in \Omega$ you can identify a continuous function $t \to B_t(\omega)$ from $[0, \infty)$ to $\mathbb{R}$. Thus a perspective on Brownian motion is that it's the space of continuous functions $C([0,\infty), \mathbb{R})$ equipped with the measures P^x above[5].

4.3.1. *Normal Random Variables Interlude.*

Definition 22. Let $(\Omega, \mathcal{F}, P)$ be a probability space. An rv $X : \Omega \to \mathbb{R}$ is $\mathrm{n}\left(m, \sigma^2\right)$ if the distribution of X has a density of the form

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

In other words

$$\begin{aligned}
P[X \in G] &= \int_G p_X(x)\, dx \\
&= \int_G I_G dP
\end{aligned}$$

---

[3]One that has these finite-dimensional distributions.

[4]Wiener's theorem.

[5]Which as discussed are product measures on a function space.

for all Borel sets $G \subset \mathbb{R}$.

Then

$$E[X] = \int_\Omega X dP = \int_{\mathbb{R}} x p_X(x)\, dx = m$$

and

$$\operatorname{Var}(X) = E\left[(X-m)^2\right] = \int_{\mathbb{R}} (x-m)^2 p_X(x)\, dx = \sigma^2$$

In general an rv $X : \Omega \to \mathbb{R}^n$ is called *multinormal* $\mathrm{n}(\mathbf{m}, \Sigma)$ if the distribution of X has the form

$$p_X(x_1, \ldots, x_n) = \frac{1}{\sqrt[n]{2\pi |\Sigma^{-1}|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \Sigma^{-1}(\mathbf{x}-\mathbf{m})}$$

where $E[X] = \mathbf{m}$ and $\Sigma^{-1}_{jk} = E[(X_j - m_j)(X_k - m_k)]$ is the covariance matrix of X.

Theorem 23. *The characteristic function of $X \sim n(\mathbf{m}, \Sigma)$ for $X : \Omega \to \mathbb{R}^n$*

$$\phi_X(u_1, \ldots, u_n) = \exp\left(-\frac{1}{2}\left(\mathbf{u}^T \Sigma^{-1} \mathbf{u}\right) + i\mathbf{u}\cdot\mathbf{m}\right)$$

Theorem 24. *Let $X_j : \Omega \to \mathbb{R}^n$ then $X = (X_1, \ldots, X_n)$ is normal iff*

$$Y = \sum_{j=1}^{l} \lambda_j X_j$$

is normal for all $\boldsymbol{\lambda}$.

Theorem 25. *Let $Y_0, Y_1, \ldots, Y_l$ be real, rvs on Ω. Assume $X = (Y_0, \ldots, Y_l)$ is normal and Y_0 and Y_j are uncorrelated for each $j \geq 1$, i.e.*

$$E[(Y_0 - EY_0)(Y_j - EY_j)] = 0$$

then Y_0 is independent of $\{Y_1, \ldots, Y_n\}$.

Theorem 26. *Suppose $X_k : \Omega \to \mathbb{R}^n$ is normal and $\lim_k X_k = X \in L^2(\Omega)$, i.e.*

$$\lim_{k\to\infty} E\left[|X_k - X|^2\right] \to 0$$

Then X is normal.

Proof. Since $\left|e^{i\mathbf{u}\cdot\mathbf{x}} - e^{i\mathbf{u}\cdot\mathbf{y}}\right| < |\mathbf{u}| \cdot |\mathbf{x} - \mathbf{y}|$ we have that

$$E\left[\left(e^{i\mathbf{u}\cdot X_k} - e^{i\mathbf{u}\cdot X}\right)^2\right] \leq |\mathbf{u}|^2 E\left[|X_k - X|^2\right] \to 0$$

and hence

$$\lim_{k\to\infty} E\left[e^{i\mathbf{u}\cdot X_k}\right] \to e^{i\mathbf{u}\cdot X}$$

with $E[X] = \lim_{k\to\infty} E[X_k]$ and $\Sigma^{-1} = \lim_{k\to\infty} \Sigma_k^{-1}$. $\square$

4.4. **Brownian motion basic properties.** Here are some of the basic properties of Brownian motion as defined by defn. 4.1:

(1) Fix the range of B_t to be $\mathbb{R}$. For $x = x_0$ B_t is a *Gaussian process,* i.e. for all $0 < t_1 < t_2 < \cdots < t_k$ the random vector $Z = (B_{t_1}, \ldots, B_{t_k}) \in (\mathbb{R})^k$ is distributed multinormal. That means there exists a mean vector $\mathbf{m} \in (\mathbb{R})^k$ and covariance matrix $\Sigma = [c_{jm}] \in \mathbb{R}^{k\times k}$ such that

$$\begin{aligned}
E^x\left[\exp\left[i\sum_{j=1}^{k} u_j B_{t_j}\right]\right] &:= \int \exp\left[i\sum_{j=1}^{k} u_j B_{t_j}\right] dP^x \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \exp\left[i\sum_{j=1}^{k} u_j B_{t_j}\right] p(t_1, x_0, x_1)\, p(t_2 - t_1, x_1, x_2) \cdots \\
&\quad p(t_k - t_{k-1}, x_{k-1}, x_k)\, dx_1 dx_2 \cdots dx_k \\
&= \frac{1}{\sqrt[k]{2\pi t_1 (t_2 - t_1) \cdots (t_k - t_{k-1})}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \exp\left[i\sum_{j=1}^{k} u_j B_{t_j}\right] e^{-\frac{(x_1 - x_0)^2}{2t_1}} e^{-\frac{(x_1 - x_2)^2}{2(t_2 - t_1)}} \cdots \\
&\quad e^{-\frac{(x_{k-1} - x_k)^2}{2(t_k - t_{k-1})}}\, dx_1 dx_2 \cdots dx_k
\end{aligned}$$

Let's try this for $k=2$

$$E^x\left[\exp\left[i\sum_{j=1}^k u_j B_{t_j}\right]\right] = \frac{1}{2\pi\sqrt{t_1(t_2-t_1)}}\int_{\mathbb{R}}\int_{\mathbb{R}}\exp\left[i\sum_{j=1}^k u_j B_{t_j}\right]e^{-\frac{(x_1-x_0)^2}{2t_1}}e^{-\frac{(x_1-x_2)^2}{2(t_2-t_1)}}dx_1dx_2$$
$$= \frac{1}{2\pi\sqrt{t_1(t_2-t_1)}}\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\frac{1}{2}\left[\begin{pmatrix}x_1-x_0\\x_2-x_0\end{pmatrix}^T\begin{pmatrix}t_1&t_1\\t_1&t_2\end{pmatrix}^{-1}\begin{pmatrix}x_1-x_0\\x_2-x_0\end{pmatrix}-2i(u_1,u_2)\cdot(x_1,x_2)^T\right]}dx_1dx_2$$
$$= \frac{1}{2\pi\sqrt{t_1(t_2-t_1)}}\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2i\mathbf{u}^T\cdot\mathbf{x}\right]}dx_1dx_2$$

Transitioning wholly to vector notation we seek $\mathbf{a}$ and b such that

$$(\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2i\mathbf{u}^T\cdot\mathbf{x} = (\mathbf{x}-\mathbf{x_0}-\mathbf{a})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0}-\mathbf{a})+b$$

Expanding the right side

$$(\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2i\mathbf{u}^T\cdot\mathbf{x} = (\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2\mathbf{a}^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})+\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{a}+b$$
$$= (\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{x}+\left[2\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{x_0}+\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{a}+b\right]$$

Hence

$$\mathbf{a}^T\mathbf{\Sigma}^{-1} = i\mathbf{u}^T$$
$$\Rightarrow$$
$$\mathbf{a} = i\mathbf{\Sigma}\mathbf{u}$$

and

$$2\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{x_0}-\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{a}+b = 0$$
$$\Rightarrow$$
$$b = -\left(2\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{x_0}+\mathbf{a}^T\mathbf{\Sigma}^{-1}\mathbf{a}\right)$$
$$= -\left(2i\mathbf{u}^T\mathbf{x_0}+\left(i\mathbf{u}^T\right)(i\mathbf{\Sigma}\mathbf{u})\right)$$

Hence

$$E^x\left[\exp\left[i\mathbf{u}\cdot\mathbf{B}\right]\right] = \frac{1}{2\pi\sqrt{t_1(t_2-t_1)}}\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{x_0})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0})-2i\mathbf{u}^T\cdot\mathbf{x}\right]}dx_1dx_2$$
$$= \frac{\exp\left[-\frac{1}{2}\left(-\left(2i\mathbf{u}^T\mathbf{x_0}+\left(i\mathbf{u}^T\right)(i\mathbf{\Sigma}\mathbf{u})\right)\right)\right]}{2\pi\sqrt{t_1(t_2-t_1)}}\int_{\mathbb{R}}\int_{\mathbb{R}}e^{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{x_0}-\mathbf{a})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{x_0}-\mathbf{a})\right]}dx_1dx_2$$
$$= \exp\left[i\mathbf{u}^T\mathbf{x_0}-\frac{1}{2}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}\right]$$

which has the form of the characteristic function of a multivariate Normal

$$\exp\left[i\mathbf{u}^T\mathbf{m}-\frac{1}{2}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}\right]$$

with means $\mathbf{m}$ and covariances $\mathbf{\Sigma}$. Therefore in general[6] the finite dimensional distributions of a Brownian motion are multivariate Gaussians with mean vector $\mathbf{m}=\mathbf{x_0}$ and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} t_1I_n & t_1I_n & \cdots & t_1I_n \\ t_1I_n & t_2I_n & \cdots & t_2I_n \\ \vdots & \vdots & & \vdots \\ t_1I_n & t_2I_n & \cdots & t_kI_n \end{pmatrix}$$

where n is the dimension of each B_{t_j}. Therefore, since each component of a multivariate Gaussian is itself Gaussian with means and covariances being a function of $\mathbf{m}$ and $\mathbf{\Sigma}$ it's the case that

$$E^x\left[B_{t_j}\right] = \mathbf{m}_{t_j} = x_0$$

and

$$\mathrm{Var}\left(B_{t_j}\right) = E^x\left[\left(B_{t_j}-x_0\right)^2\right] = t_j\mathrm{Tr}\left(I_n\right) = nt_j$$

and because of the cascading[7] shape of the covariance matrix

$$\mathrm{Cov}\left(B_{t_j},B_{t_i}\right) = E^x\left[\left(B_{t_j}-x_0\right)\left(B_{t_i}-x_0\right)\right] = \mathrm{Tr}\left(I_n\right)\min\left(t_j,t_i\right) = n\min\left(t_j,t_i\right)$$

[6]With $B_{t_j}=\mathbf{B}_{t_j}\in\mathbb{R}^n$

[7]Just think about running your finger down and across the covariance matrix.

and

$$
\begin{aligned}
E^x\left[\left(B_{t_j}-B_{t_i}\right)^2\right] &= E^x\left[\left(B_{t_j}-x_0\right)^2-2\left(B_{t_j}-x_0\right)\left(B_{t_i}-x_0\right)+\left(B_{t_i}-x_0\right)^2\right]\\
&= \operatorname{Tr}\left(I_n\right)\left(t_j-2\min\left(t_j,t_i\right)+t_i\right)\\
&= \begin{cases} n\left(t_j-t_i\right) & \text{if } t_j\geq t_i\\ n\left(t_i-t_j\right) & \text{otherwise}\end{cases} \qquad (4.2)\\
&= n\left|t_i-t_j\right| \qquad (4.3)
\end{aligned}
$$

(2) B_t has independent increments, i.e. for $0\leq t_1\leq t_2\leq\cdots\leq t_k$

$$B_{t_1},B_{t_2}-B_{t_1},\ldots,B_{t_k}-B_{t_{k-1}}$$

are independent random variables. This follows from the fact that normal RVs are independent if their covariance is zero and for $t_i<t_j$

$$
\begin{aligned}
E^x\left[\left(B_{t_j}-B_{t_{j-1}}\right)\left(B_{t_i}-B_{t_{i-1}}\right)\right] &= E^x\left[\left(B_{t_j}-B_{t_{j-1}}\right)\left(B_{t_i}-B_{t_{i-1}}\right)\right]\\
&= E^x\left[B_{t_j}B_{t_i}-B_{t_j}B_{t_{i-1}}-B_{t_{j-1}}B_{t_i}+B_{t_{j-1}}B_{t_{i-1}}\right]\\
&= E^x\left[\left(B_{t_j}-x_0\right)\left(B_{t_i}-x_0\right)-\left(B_{t_j}-x_0\right)\left(B_{t_{i-1}}-x_0\right)-\left(B_{t_{j-1}}-x_0\right)\left(B_{t_i}-x_0\right)+\left(B_{t_{j-1}}-x_0\right)\left(\right.\right. \qquad (4.4)\\
&= n\left(t_i-t_{i-1}-t_i+t_{i-1}\right)=0
\end{aligned}
$$

(3) B_t *is* almost surely continuous but we need a little theory to prove it:

Definition. Suppose X_t and Y_t are stochastic processes on $(\Omega,\mathcal{F},P)$ then X_t is a *modification* of Y_t if for all t

$$P\left(\left\{\omega;X_t\left(\omega\right)=Y_t\left(\omega\right)\right\}\right)=1$$

Note that X_t and Y_t have the same law[8] and as such are essentially the same, but might have different path properties.

Theorem 27. ***Kolmogorov's continuity theorem****. Suppse for the process X_t there exist α,β,D for all T*

$$E\left[\left|X_t-X_s\right|^{\alpha}\right]\leq D\left|t-s\right|^{1+\beta}\quad \textit{for all } 0\leq s,t\leq T$$

Corollary 28. *With $\alpha=4,\beta=1,D=n\left(n+2\right)$ Brownian motion X_t satisfies the criterion for Kolmogorov's continuity theorem and therefore there exists a version Y_t with continuous paths.*

5. Ito Integral

5.1. **Motivation.** Return to the original problem of finding a reasonable interpretation of

$$\frac{dx}{dt}=b\left(t,x\right)+\sigma\left(t,x\right)\cdot\text{"noise"}$$

It turns out it's reasonable to model "noise" as some stochastic process W_t and so

$$\frac{dX_t}{dt}=b\left(t,X_t\right)+\sigma\left(t,X_t\right)\cdot W_t \qquad (5.1)$$

The response x has become an X_t, connoting stochastic process, because, as it will turn out, this is short hand for a relationship between the response and the input that necessitates that the reponse be a stochastic process itself. Based on empricial fact (experience) one expects that W_t have these properties:

(1) $t_1\neq t_2\Rightarrow W_{t_1}\perp W_{t_2}$
(2) $\{W_t\}$ is stationary, i.e. the joint distribution of $\{W_{t_1+\tau},\ldots,W_{t_k+\tau}\}$ does not depend on τ.
(3) $E\left[W_t\right]=0$ for all t.

Turns out, unfortunately, there's no continuous[9] process that satisfies 1 and 2 that works in a model like eqn. 5.1. The important property unsatisfiable by a continuous process on $\mathbb{R}$ is property 1: for W_t to be independent at arbitrary times would require the process to have infinite variance[10]. It is possible to represent W_t as a *generalized* process, meaning it can be constructed as a measure on the space of *tempered distributions*[11] on $[0,\infty)$. That notwithstanding let's rewrite eqn. 5.1 that lends itself to having W_t replaced by a proper stochastic process: for $0=t_0<t_1<\cdots<t_m=t$ discretize

$$X_{k+1}-X_k=b\left(t_k,X_k\right)\Delta t_k+\sigma\left(t_k,X_k\right)W_k\Delta t_k$$

where $X_k\coloneqq X_{t_k}$. Restated the question is: does there exist some V_t such that for $\Delta V_k\coloneqq V_{k+1}-V_k\coloneqq V_{t_{k+1}}-V_{t_k}$

$$
\begin{aligned}
X_{k+1}-X_k &= b\left(t_k,X_k\right)\Delta t_k+\sigma\left(t_k,X_k\right)\left(V_{k+1}-V_k\right)\\
&= b\left(t_k,X_k\right)\Delta t_k+\sigma\left(t_k,X_k\right)\Delta V_k
\end{aligned}
$$

[8]The same finite-dimensional distributions
[9]A strict requirement.
[10]Citation needed.
[11]**Todo**... but this makes sense given the previous line.

Assumptions 1,2,3 above suggest that stationary, independent, and mean 0 **increments**. Why? Because what appears in the discretized model are the increments. Turns out the only such process with continuous paths is Brownian motion B_t. Thus putting $V_t = B_t$ and taking sums we get

$$\sum_{j=0}^{k-1} (X_{k+1} - X_k) = X_k - X_0 = \sum_{j=0}^{k-1} \left(b\left(t_j, X_j\right) \Delta t_j + \sigma\left(t_j, X_j\right) \Delta B_j\right)$$

Given that B_t satisfies assumptions 1,2,3 does a limit (in some sense) as $\Delta t_j \to 0$ exist? Does it produce something like

$$X_k = X_0 + \int_0^t b\left(s, X_s\right) ds + \text{"} \int_0^t \sigma\left(s, X_s\right) dB_s \text{"} \tag{5.2}$$

such that we can adopt the convention that what eqn. 5.1 really means is that the response X_t satisfies eqn. 5.2. Thus it remains to prove existence for a wide class of function $f : [0, \infty) \times \Omega \to \mathbb{R}$, in some sense, of

$$\text{"} \int_0^t f\left(s, \omega\right) dB_s\left(\omega\right) \text{"}$$

where $B_t(\omega)$ is a Brownian motion (1-dimensional for the time being).

5.2. **A first attempt at the construction of the Ito integral.** Let $0 \le Q < T$ and start by defining

$$\int_Q^T \{\cdot\} \, dB_s\left(\omega\right)$$

for simple processes

$$S_n\left(t, \omega\right) = \sum_{j=0}^{\infty} a_j\left(\omega\right) 1_{[j\cdot 2^{-n}, (j+1)\cdot 2^{-n})}\left(t\right)$$

You can imagine that S_n is defined for all $t \in [0, \infty)$ by defining the $a_j = 0$ appropriately. So basically chop the real line into intervals of length 2^{-n} and define $S_t(\omega)$ piecewise constant on that mesh. Then define

$$\int_Q^T S_n\left(s, \omega\right) dB_s := \sum_{j=0}^{\infty} a_j\left(\omega\right) \left[B_{s_{j+1}^{(n)}}\left(\omega\right) - B_{s_j^{(n)}}\left(\omega\right)\right]$$

where

$$s_j^{(n)} := \begin{cases} \frac{j}{2^n} & \text{if } Q \le j \cdot 2^{-n} \le T \\ Q & \text{if } j \cdot 2^{-n} < Q \\ T & \text{if } j \cdot 2^{-n} > T \end{cases}$$

which just truncates the sum outside of $[Q, T]$ since for $s_j^{(n)} > T$

$$B_{s_{j+1}^{(n)}}\left(\omega\right) - B_{s_j^{(n)}}\left(\omega\right) = B_T\left(\omega\right) - B_T\left(\omega\right) = 0$$

and $T - Q = m \cdot 2^{-n}$ for some m and "around the edges" the error becomes neglibile as $n \to \infty$. This is manifestly reasonable but the next natural step, approximating arbitrary processes by these simple processes, when done carelessly leads to problems.

Example 29. Let $B_{j\cdot 2^{-n}}(\omega)$ be samples of a B_t process and define $a_j(\omega) = B_{j\cdot 2^{-n}}(\omega)$. Then construct the simple processes

$$\begin{aligned}
S_n\left(t, \omega\right) &= \sum_{j=0}^{\infty} a_j\left(\omega\right) 1_{[j\cdot 2^{-n}, (j+1)\cdot 2^{-n})}\left(t\right) \\
&= \sum_{j=0}^{\infty} B_{j\cdot 2^{-n}}\left(\omega\right) 1_{[j\cdot 2^{-n}, (j+1)\cdot 2^{-n})}\left(t\right) \\
S_n^{'}\left(t, \omega\right) &= \sum_{j=0}^{\infty} a_{j+1}\left(\omega\right) 1_{[j\cdot 2^{-n}, (j+1)\cdot 2^{-n})}\left(t\right) \\
&= \sum_{j=0}^{\infty} B_{(j+1)\cdot 2^{-n}}\left(\omega\right) 1_{[j\cdot 2^{-n}, (j+1)\cdot 2^{-n})}\left(t\right)
\end{aligned}$$

Then

$$\begin{aligned}
E\left[\int_0^T S_n\left(s, \omega\right) dB_s\left(\omega\right)\right] &= E\left[\sum_{j=0}^{\infty} B_{s_j^{(n)}} \left(B_{s_{j+1}^{(n)}}\left(\omega\right) - B_{s_j^{(n)}}\left(\omega\right)\right)\right] \\
&= \sum_{j=0}^{\infty} E\left[B_{s_j^{(n)}} \left(B_{s_{j+1}^{(n)}}\left(\omega\right) - B_{s_j^{(n)}}\left(\omega\right)\right)\right] \\
&= 0
\end{aligned}$$

Since $\left(B_{s_j^{(n)}}-0\right)$ and $\left(B_{s_{j+1}^{(n)}}-B_{s_j^{(n)}}\right)$ are independent increments and $E\left[B_{s_j^{(n)}}\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)\right]=0$ is the definition of independent RVs. But

$$\begin{aligned}
E\left[\int_0^T S_n'(s,\omega)\,dB_s(\omega)\right] &= \sum_{j=0}^{\infty} E\left[B_{s_{j+1}^{(n)}}\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)\right] \\
&= \sum_{j=0}^{\infty} E\left[\left(B_{s_{j+1}^{(n)}}(\omega)\right)^2-2B_{s_{j+1}^{(n)}}(\omega)B_{s_j^{(n)}}(\omega)+B_{s_{j+1}^{(n)}}(\omega)B_{s_j^{(n)}}(\omega)+\left(B_{s_j^{(n)}}(\omega)\right)^2-\left(B_{s_j^{(n)}}(\omega)\right)^2\right] \\
&= \sum_{j=0}^{\infty} E\left[\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)^2+B_{s_{j+1}^{(n)}}(\omega)B_{s_j^{(n)}}(\omega)-\left(B_{s_j^{(n)}}(\omega)\right)^2\right] \\
&= \sum_{j=0}^{\infty} E\left[\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)^2+B_{s_j^{(n)}}(\omega)\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)\right] \\
&= \sum_{j=0}^{\infty} E\left[\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)^2\right]+E\left[B_{s_j^{(n)}}(\omega)\left(B_{s_{j+1}^{(n)}}(\omega)-B_{s_j^{(n)}}(\omega)\right)\right] \\
&= \sum_{j=0}^{\infty}\left(\frac{(j+1)}{2^n}-\frac{j}{2^n}\right) \\
&= T\sum_{j=0}^{m-1}\frac{1}{2^n}+\sum_{j=m}^{\infty}(T-T) \\
&= T
\end{aligned}$$

This shows that despite both S and S' being reasonable approximations for $f(s,\omega)=B_s(\omega)$ their integrals are not even remotely close no matter how fine the mesh (differing by T).

5.2.1. *Total variation - technical aside.* Intuitively the integrals of S and S' don't agree is because $f(t)=f(t,\omega)=B_t(\omega)$ has infinite total variation;

$$TV(f):=\lim_{n\to\infty}\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|$$

over some $[Q,T]$[12]. Here's a short proof of this: first define quadratic variation

$$QV(f):=\lim_{n\to\infty}\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|^2$$

and notice that if f is continuous then

$$\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|^2\le\left(\max_{1\le j\le m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|\right)\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|$$

and so

$$\frac{\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|^2}{\max_{1\le j\le m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|}\le\sum_{j=1}^{m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|$$

and hence any continuous f that has non-zero quadratic variation has infinite total variation[13]. So all we need to prove is that B_s has non-zero quadratic variation. First some lemmas.

Lemma 30. *If*

$$\lim_{n\to\infty}Var\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}}-B_{t_{j-1}^{(n)}}\right)^2\right]=0$$

then$\lim_{n\to\infty}QV(f)=T-Q$ *in* L^2.

---

[12]Recall that $T-Q=m\cdot 2^{-n}$.

[13]Since $\max_{1\le j\le m}\left|f\left(t_j^{(n)}\right)-f\left(t_{j-1}^{(n)}\right)\right|\to 0$ as $|\Pi|\to\infty$ for any continuous f and your only hope for the left side of the inequality not blowing up is if the numerator, $QV(f)$, is 0.

Proof. Let $\Delta B_j^2 = \left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2$. Then if the variance goes to 0 [14]

$$\begin{aligned}
\lim_{n\to\infty} E\left[\left(\sum_{j=1}^{m} \Delta B_j^2\right)^2\right] &= \lim_{n\to\infty}\left(E\left[\sum_{j=1}^{m} \Delta B_j^2\right]\right)^2 \\
&= \lim_{n\to\infty}\left(\sum_{j=1}^{m} E\left[\Delta B_j^2\right]\right)^2 \\
&= \lim_{n\to\infty}\left(\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right)\right)^2 \quad \text{by eqn. 4.2} \\
&= \lim_{n\to\infty}(T-Q)^2
\end{aligned}$$

and so

$$\begin{aligned}
0 &= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m} \Delta B_j^2\right)^2\right] - (T-Q)^2\right) \\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m} \Delta B_j^2\right)^2\right] - 2(T-Q)^2 + (T-Q)^2\right) \\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m} \Delta B_j^2\right)^2\right] - 2(T-Q)\,E\left[\left(\sum_{j=1}^{m} \Delta B_j^2\right)\right] + (T-Q)^2\right) \\
&= \lim_{n\to\infty}\left(E\left[\left(\sum_{j=1}^{m} \Delta B_j^2 - (T-Q)\right)^2\right]\right)
\end{aligned}$$

which is the definition of convergence in L^2. □

Lemma 31. *On refinement of the mesh*

$$\lim_{n\to\infty} Var\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] = 0$$

Proof. The proof

$$\begin{aligned}
\operatorname{Var}\left[\sum_{j=1}^{m}\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] &= \sum_{j=1}^{m} \operatorname{Var}\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] \\
&= \sum_{j=1}^{m}\left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - \left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right]\right)^2\right) \\
&= \sum_{j=1}^{m}\left(E\left[\left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2\right) \\
&\quad \text{by Kolmogorov continuity} \\
&= \sum_{j=1}^{m}\left(1\,(1+2)\left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2\right) \\
&= 2\sum_{j=1}^{m}\left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2
\end{aligned}$$

which goes to 0 as the mesh is refined. □

Theorem 32. *For $f = B_t$ it's the case that $\lim_{n\to\infty} QV(f) = T - Q$ almost surely.*

[14] Since $\operatorname{Var}(X) = EX^2 - (EX)^2$.

Proof. Let

$$X_i^{(n)} = \Delta B_j^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)$$

and

$$Y_n := \sum_{j=1}^{m} X_i^{(n)} = \sum_{j=1}^{m} \left(\Delta B_j^2 - \left(t_j^{(n)} - t_{j-1}^{(n)}\right)\right) = \sum_{j=1}^{m} \Delta B_j^2 - (T - Q)$$

Then

$$\begin{aligned} EY_n &= E\left[\sum_{j=1}^{m} \left(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}\right)^2\right] - E(T-Q) \\ &= 0 \end{aligned}$$

and

$$EY_n^2 = E\left(\sum_{j=1}^{m} \left(X_i^{(n)}\right)^2 + \sum_{i<j} X_i^{(n)} X_j^{(n)}\right) = \sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right] + \sum_{i<j} E\left[X_i^{(n)} X_j^{(n)}\right]$$

but by eqn. 4.4 $E\left[X_i^{(n)} X_j^{(n)}\right] = 0$ so

$$EY_n^2 = \sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right]$$

and so by Chebyshev's inequality[15]

$$\begin{aligned} P(|Y_n| \geq \epsilon) &\leq \frac{E\left[(Y_n)^2\right]}{\epsilon^2} \\ &= \frac{1}{\epsilon^2} \sum_{j=1}^{m} E\left[\left(X_i^{(n)}\right)^2\right] \\ &= \frac{1}{\epsilon^2} \sum_{j=1}^{m} \left(t_j^{(n)} - t_{j-1}^{(n)}\right)^2 \\ &\leq \frac{1}{\epsilon^2} \frac{1}{2^n} \sum_{j=1}^{m} \left(t_j^{(n)} - t_{j-1}^{(n)}\right) \\ &= \frac{T-Q}{2^n \epsilon^2} \end{aligned}$$

and finally using Borel-Cantelli[16] with

$$\sum_{n=1}^{\infty} P(|Y_n| \geq \epsilon) \leq \sum_{n=1}^{\infty} \frac{T-Q}{2^n \epsilon^2} = \frac{T-Q}{\epsilon^2}$$

which implies almost sure convergence[17] of $Y_n \to 0$. □

5.2.2. *Back to the attempt.* Recall that the difference between S and S' was that for S the samples of B_t were taken at the left side of the interval $\left[j \cdot 2^{-n}, (j+1) \cdot 2^{-n}\right)$ and for S' the samples were taken at the right side. In general[18] you approximate a process $f(t, \omega)$ by

$$\sum_j f\left(t_j^*, \omega\right) \cdot 1_{[t_j, t_{j+1})}$$

where the only requirement on $t_j^* \in [t_j, t_{j+1})$ but it's apparent that for defining an integral of a stochastic process the choice matters. There are two standard choices:

(1) $t_j^* = t_j$ the left endpoint characterizes the Ito integral denoted

$$\int_Q^T f(t, \omega)\, dB_t(\omega)$$

(2) $t_j^* = (t_j + t_{j+1})/2$ the midpoint characterizes the *Stratonovich* integral denoted

$$\int_Q^T f(t, \omega) \circ dB_t(\omega)$$

---

[15] $P(|X - \mu| \geq \epsilon) \leq \frac{E\left[(X-\mu)^2\right]}{\epsilon^2}$

[16] If $\sum_{n=1}^{\infty} P(E_n) < \infty$ for some sequence of events E_n then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

[17] $P\left(\liminf_{n\to\infty} |X_n - X| < \epsilon\right) = 1$ for all ϵ. Naturally this is to equivalent $P\left(\liminf_{n\to\infty} |X_n - X| > \epsilon\right) = 0$ for all ϵ. Why? $\liminf$ is the set of points ω that is ultimately in all of the sets and $\limsup$ is the set of points ω appear infinitely often. So if the set of ω for which $|Y_n| \geq \epsilon$ occur infinitely often has measure 0 then set of ω for which $|Y_n| \leq \epsilon$ eventually always is almost all of them (otherwise $|Y_n| \geq \epsilon$ would keep happening once in a while).

[18] At least the Riemann-Stieltjes integral.

Choosing one resolves the problem of unbounded variation but still, in either case, one must restrict the class of integrand processes to a special class. Defining this class necessitates defining some properties. Heuristically these properties capture the quality that the special processes *only depend on the behavior of $B_t(\omega)$ up to time t_j.*

Definition 33. A *filtration* is a sequence of σ-algebras $\mathcal{F}_t$ of a σ-algebra $\mathcal{F}$ over a sample sample space Ω such that

$$t_1 \leq t_2 \Rightarrow \mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$$

Then a *stochastic basis* is a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, written $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\right)$.

Definition 34. Let $\{\mathcal{F}_t\}_{t\geq 0}$ be filtration. A process $g : \mathbb{R} \times \Omega \to \mathbb{R}$ is called $\mathcal{F}_t$*-adapted* if for each $t \geq 0$ the curried function

$$\omega \to g(t, \omega)$$

is $\mathcal{N}_t$-measureable.

The filtration that a process is adapted to, heuristically, contains all the "history of the process up to time t". Another name for adapted process is non-anticipating (because it doesn't depend on future).

5.3. **The construction of the Ito integral in earnest.**

Definition 35. Let $\mathcal{F}_t$ be the σ-algebra that B_t is adapted to and $\mathcal{F} = \cup_{t\geq 0}\mathcal{F}_t$. Define $\mathcal{V} = \mathcal{V}(Q, T)$ to be the class of processes

$$g(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}$$

such that

(1) $(t, \omega) \to g(t, \omega)$ is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B} = \sigma([0, \infty))$.
(2) $g(t, \omega)$ is $\mathcal{F}_t$-adapted.
(3) $E\left[\int_Q^T (g(s, \omega))^2 \, ds\right] < \infty$

Parts 1 and 2 of defn. are fairly obvious but part 3 is mysterious. Firstly what does it even mean? The integral in argument of the expectation is a mundane Lebesgue integral and it integrates out the t dependence in $g(t, \omega)$. So

$$X(\omega) = \int_Q^T (g(s, \omega))^2 \, ds$$

and hence the integral is just a mundane random variable and $E[X]$ is the expectation of that random variable. In fact

$$\|g\|_{\mathcal{V}} = \sqrt{\left|E\left[\int_Q^T (g(s, \omega))^2 \, ds\right]\right|} \tag{5.3}$$

is a semi-norm and is complete[19]. Let's call this norm the *ito norm.* This property is necessary for the *Ito isometry* which will be discussed shortly.

We will define the integral on simple processes and then use the next lemma to extend it.

Lemma 36. *(Ito isometry) If S is simple and bounded then*

$$E\left[\left(\int_Q^T S(s, \omega) \, dB_s\right)^2\right] = E\left[\int_Q^T (S(s, \omega))^2 \, ds\right]$$

Proof. First for $i < j$ by independence of increments

$$\begin{aligned} E\left[\left(a_i\left(B_{s_{i+1}} - B_{s_i}\right)\right)\left(a_j\left(B_{s_{j+1}} - B_{s_j}\right)\right)\right] &= E\left[a_i a_j\right] E\left[\left(B_{s_{i+1}} - B_{s_i}\right)\left(B_{s_{j+1}} - B_{s_j}\right)\right] \\ &= \begin{cases} 0 & \text{if } i \neq j \\ E\left[a_j^2\right] E\left[\left(B_{s_{j+1}} - B_{s_j}\right)^2\right] & \text{if } i = j \end{cases} \\ & \quad \begin{cases} 0 & \text{if } i \neq j \\ E\left[a_j^2\right]\left(s_{j+1} - s_j\right) & \text{if } i = j \end{cases} \end{aligned}$$

[19]Cauche sequences converge.

Thus

$$\begin{aligned}
E\left[\left(\int_Q^T S(s,\omega)\,dB_s\right)^2\right] &= E\left[\sum_{i=1}^{\infty}\left(a_i\left(B_{s_{i+1}}-B_{s_i}\right)\right)\sum_{j=1}^{\infty}\left(a_j\left(B_{s_{j+1}}-B_{s_j}\right)\right)\right] \\
&= \sum_{i,j}^{\infty} E\left[\left(a_i\left(B_{s_{i+1}}-B_{s_i}\right)\right)\left(a_j\left(B_{s_{j+1}}-B_{s_j}\right)\right)\right] \\
&= \sum_{j=1}^{\infty} E\left[a_j^2\right]\left(s_{j+1}-s_j\right) \\
&= E\left[\sum_{j=1}^{\infty} a_j^2\left(s_{j+1}-s_j\right)\right] \\
&= E\left[\int_Q^T \left(S(s,\omega)\right)^2 ds\right]
\end{aligned}$$

□

The rest of the construction proceeds in 3 steps:

(1) First we define the integral for simple processes

$$S_n(t,\omega) = \sum_{j=0}^{\infty} a_j(\omega)\, 1_{[t_j,t_{j+1})}(t)$$

(which are in $\mathcal{V}$ if $a_j(\omega)$ are $\mathcal{F}_{t_j}$-measurable, so that the whole simple process is adapted to B_t) to be

$$\int_Q^T S(s,\omega)\,dB_s := \sum_{j=0}^{\infty} a_j(\omega)\left[B_{s_{j+1}}(\omega) - B_{s_j}(\omega)\right]$$

(2) Let $g \in \mathcal{V}$ be uniformly[20] bounded and $g(\cdot,\omega)$ continuous for each ω. Then there exist simple processes $S_n \in \mathcal{V}$ such that

$$\lim_{n\to\infty} E\left[\int_Q^T (g - S_n)^2\, ds\right] = \lim_{n\to\infty} \|g - S_n\| = 0$$

Proof. Continuous processes can be arbitrarily approximating by simple processes hence

$$\lim_{n\to\infty} \int_Q^T (g - S_n)^2\, ds = 0$$

and by the Bounded convergence theorem[21]we get the result. Note the sequence of functions that we are applying bounded convergence to is $\left\{\int_Q^T (g-S_n)^2\right\}$. □

(3) Let $h \in \mathcal{V}$ be uniformly bounded. Then there exist uniformly bounded processes $g_n \in \mathcal{V}$ such that g_n is continuous for all n and ω and

$$\lim_{n\to\infty} E\left[\int_Q^T (h - g_n)^2\, ds\right] = \lim_{n\to\infty} \|h - g_n\| = 0$$

Proof. Suppose $|h(t,\omega)| \le M$. For each n let ψ_n be nonnegative, continuous processes on $\mathbb{R}$ such that □

(a) $\psi_n(x) = 0$ for $x \le -\frac{1}{n}$ and $x \ge 0$
(b) $\int_{-\infty}^{\infty} \psi_n(x)\, dx = 1$

So ψ_n converge to a $\delta(x)$. Define

$$g_n(t,\omega) = \int_0^t \psi_n(s-t)\, h(s,\omega)\, ds$$

So "smooth" h. Then $g_n(\cdot,\omega)$ is continuous for each ω and $|g_n| \le M$, and since $h \in \mathcal{V}$ it's the case that $g_n(t,\cdot)$ is $\mathcal{F}_t$ measurable for all t (use sums to approximate the integral defining g_n, meaning go back to the definition of the Lebesgue integral in terms of simple processes). Finally

$$\lim_{n\to\infty} \int_Q^T \left(g_n(s,\omega) - h(s,\omega)\right)^2 ds = 0$$

---

[20] $|g| \le M$

[21] If f_n is a sequence of uniformly bounded real-valued measurable functions which converges pointwise on a bounded measure space (i.e. one in which $\mu(\Omega)$ is finite) to a function f, then the limit f is an integrable function and$\lim_{n\to\infty} \int_S f_n d\mu = \int_S f d\mu$.

since $\{\psi_n\}$ constitute an approximate identity, which probably just means that convolving with the Dirac delta is some sort of identity transformation[22]. Then finally by Bounded convergence again

$$\lim_{n\to\infty} E\left[\int_Q^T (h-g_n)^2\, ds\right] = 0$$

(4) Let $f \in \mathcal{V}$. Then there exist uniformly bounded processes $h_n \in \mathcal{V}$ such that

$$\lim_{n\to\infty} E\left[\int_Q^T (f-h_n)^2\, ds\right] = \lim_{n\to\infty} \|f-h_n\| = 0$$

Proof. Let

$$h_n = \begin{cases} -n & \text{if } f < -n \\ f & \text{if } -n \leq f \leq n \\ n & \text{if } f > n \end{cases}$$

Then by DCT

$$\lim_{n\to\infty} E\left[\int_Q^T (f-h_n)^2\, ds\right] = 0$$

□

That concludes the approximation procedure. We now use this approximation by Finally the definition

Definition 37. Let $f \in \mathcal{V}(Q,T)$. Choose $S_n \in \mathcal{V}$ according to the approximation such that

$$\lim_{n\to\infty} E\left[\int_Q^T (f-S_n)^2\, ds\right] = \lim_{n\to\infty} \|f-S_n\| = 0$$

Then the *Ito integral* is defined

$$\mathcal{I}[f](\omega) := \int_Q^T f(s,\omega)\, dB_s := \lim_{n\to\infty} \int_Q^T S_n(s,\omega)\, dB_s$$

The limit in defn. 37 exists as an element of $L^2(P)$[23] **since** $\left\{\int_Q^T S_n dB_s\right\}$ **forms a Cauchy-sequence in** $L^2(P)$. Why? Firstly by the hypothesis of the definition S_n converge to f under the ito norm. Therefore, since convergent sequences are also cauchy convergent, we have for all k

$$\lim_{n\to\infty} E\left[\int_Q^T (S_{n+k}-S_n)^2\, ds\right] = 0$$

But by ito isometry

$$\begin{aligned}
\lim_{n\to\infty} E\left[\int_Q^T (S_{n+k}-S_n)^2\, ds\right] &= \lim_{n\to\infty} E\left[\left(\int_Q^T (S_{n+k}-S_n)\, dB_s\right)^2\right] \\
0 &= \lim_{n\to\infty} E\left[\left(\int_Q^T S_{n+k} dB_s - \int_Q^T S_n dB_s\right)^2\right] \\
&= \lim_{n\to\infty} E\left[\left(\sum_{j=0}^{\infty} a_j(\omega)\left[B_{s_{j+1}^{(n+k)}} - B_{s_j^{(n+k)}}\right] - \sum_{j=0}^{\infty} a_j(\omega)\left[B_{s_{j+1}^{(n)}} - B_{s_j^{(n)}}\right]\right)^2\right]
\end{aligned}$$

but this is a mundane $L^2(P)$ limit and the 0 on the left implies the limit converges. There are two consequences of the definition of the integral as the limit of a sequence of simple processes:

Corollary 38. *The ito isometry holds for all* $f \in \mathcal{V}$, *i.e.*

$$E\left[\left(\int_Q^T f(s,\omega)\, dB_s\right)^2\right] = E\left[\int_Q^T f^2(s,\omega)\, ds\right]$$

Corollary 39. *If* f_n *converges to* f *in the ito norm then* $\mathcal{I}[f_n]$ *converges to* $\mathcal{I}[f]$ *in* $L^2(P)$ *as well, i.e.*

$$E\left[\int_Q^T (f_n - f)\, ds\right] \to 0 \Rightarrow \int_Q^T f_n(s,\omega)\, dB_s \to \int_Q^T f(s,\omega)\, dB_s$$

This is mostly a reiteration of the fact that $\mathcal{I}[f]$ *is defined to be the limit of* $\mathcal{I}[f_n]$.

[22] Hoffman 1962 p 22.

[23] Square Lebesgue integrable functions over the measure P.

Alright finally we can actually do an ito integral!

Example 40. Assume $B_0 = 0$. Then

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{t}{2}$$

Proof. Let $t_j^{(n)} = j \cdot 2^{-n}$ and put $S_n(t,\omega) = \sum_{j=0}^{\infty} B_{t_j^{(n)}}(\omega) 1_{\left[t_j^{(n)}, t_{j+1}^{(n)}\right)}(t)$. Then we need to prove convergence in ito norm:

$$E\left[\int_0^t (S_n - B_s)^2 ds\right] = E\left[\sum_{j=0}^{\infty}\int_{t_j^{(n)}}^{t_{j+1}^{(n)}} \left(B_{t_j^{(n)}} - B_s\right)^2 ds\right] \tag{5.4}$$

$$= \sum_{j=0}^{\infty}\int_{t_j^{(n)}}^{t_{j+1}^{(n)}} E\left[\left(B_{t_j^{(n)}} - B_s\right)^2\right] ds \tag{5.5}$$

$$= \sum_{j=0}^{\infty}\int_{t_j^{(n)}}^{t_{j+1}^{(n)}} \left(s - t_j^{(n)}\right) ds = \frac{1}{2}\sum_{j=0}^{\infty}\left(t_{j+1}^{(n)} - t_j^{(n)}\right)^2 \tag{5.6}$$

where in line 5.4 we use the indicator in the definiton of S_n and in line 5.6 we use that since the limits of integration for s are $t_j^{(n)}$ and $t_{j+1}^{(n)}$ it's the case that $s > t_j^{(n)}$. Finally as the mesh is refined $\frac{1}{2}\sum_{j=0}^{\infty}\left(t_{j+1}^{(n)} - t_j^{(n)}\right)^2 \to 0$. So

$$\int_0^t B_s dB_s = \lim_{n\to\infty}\int_0^t S_n dB_s = \lim_{n\to\infty}\sum_{j=0}^{\infty} B_{t_j^{(n)}}\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)$$

Let $\Delta B_{t_j^{(n)}} := B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}$ and $\Delta\left(B_{t_j^{(n)}}^2\right) = B_{t_{j+1}^{(n)}}^2 - B_{t_j^{(n)}}^2$. Then

$$\int_0^t B_s dB_s = \lim_{n\to\infty}\sum_{j=0}^{\infty} B_{t_j^{(n)}}\Delta B_{t_j^{(n)}}$$

So we have to evaluate this limit. First

$$\begin{aligned}\Delta\left(B_{t_j^{(n)}}^2\right) &= \left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)^2 + 2B_{t_j^{(n)}}\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)\\ &= \left(\Delta B_{t_j^{(n)}}\right)^2 + 2B_{t_j^{(n)}}\Delta B_{t_j^{(n)}}\end{aligned}$$

The last term in this derivation is what we're interested in because it appears in out limit. Now since $B_0 = 0$ with probability 1

$$B_t^2 = \sum_{j=0}^{\infty}\Delta\left(B_{t_j^{(n)}}^2\right) = \sum_{j=0}^{\infty}\left(\Delta B_{t_j^{(n)}}\right)^2 + 2\sum_{j=0}^{\infty} B_{t_j^{(n)}}\Delta B_{t_j^{(n)}}$$

or

$$\sum_{j=0}^{\infty} B_{t_j^{(n)}}\Delta B_{t_j^{(n)}} = \frac{1}{2}B_t^2 - \frac{1}{2}\sum_{j=0}^{\infty}\left(\Delta B_{t_j^{(n)}}\right)^2$$

Finally by unbounded variation (thm 32)

$$\lim_{n\to\infty}\sum_{j=0}^{\infty}\left(\Delta B_{t_j^{(n)}}\right)^2 = t$$

and so

$$\lim_{n\to\infty}\sum_{j=0}^{\infty} B_{t_j^{(n)}}\Delta B_{t_j^{(n)}} = \frac{1}{2}B_t^2 - \frac{1}{2}t$$

□

Note the extra $\frac{1}{2}t$ term, exhibiting the difference from the standard integration rule of $\int x = \frac{1}{2}x^2$.

5.4. **Properties of the Ito Integral and the Ito antiderivative.**

Theorem 41. *Some properties of the Ito integral: let $f, g \in \mathcal{V}(0,T)$ and $0 \le Q < U < T$. Then*

1. $P\left(\int_Q^T f dB = \int_Q^U f dB + \int_U^T f dB\right) = 1$
2. *For* $c \in \mathbb{R}$: $P\left(\int_Q^T (cf + g)\, dB = \int_Q^T cf dB + \int_Q^T g dB\right) = 1$
3. $E\left[\int_Q^T f dB\right] = 0$
4. $\int_Q^T f dB$ *if* $\mathcal{F}_T$*-measurable*

Proof. All follow from the correspondent property holding for Ito integral of elementary functions. □

The Ito integral of a process turns out to be a *martingale*, whatever that is. Understanding this requires a short discussion of conditional expectation.

5.4.1. *Conditonal expectation.*

Definition 42. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{H} \subset \mathcal{F}$. The *conditional expectation* of the random variable X is the unique function $E[X|\mathcal{H}]$ that satisfies

(1) $E[X|\mathcal{H}]$ is $\mathcal{H}$-measurable
(2) $\int_H E[X|\mathcal{H}]\, dP = \int_H X dP$ for all $H \in \mathcal{H}$

In fact $E[X|\mathcal{H}]$ is the *Radon-Nikodym derivative* of $E[X]$ with respect to the measure "P restricted to $\mathcal{H}$": a consequence of the Radon-Nikodym theorem[24] is that for $A \in \mathcal{F}$

$$E[X]|_A = \int_A X dP$$

Then $E[X]|_A$ is absolutely continuous[25] w.r.t. $P|_{\mathcal{H}}$, i.e. there exists an $\mathcal{H}$-measurable function F such that

$$E[X]|_{\mathcal{H}}(H) = \int_H F\, dP|_{\mathcal{H}} \text{ for } H \in \mathcal{H}$$

and thus $E[X|\mathcal{H}] \coloneqq F$.

A small fact about conditional expectation

Lemma 43. *The law of the iterated expectation: Let $\mathcal{G}, \mathcal{H}$ such that $\mathcal{G} \subset \mathcal{H}$. Then*

$$E[X|\mathcal{G}] = E[E[X|\mathcal{H}]|\mathcal{G}]$$

Proof. If $G \in \mathcal{G}$ then $G \in \mathcal{H}$ and therefore

$$\int_G E[X|\mathcal{H}]\, dP = \int_G X dP$$

and so $E[X|\mathcal{G}] = E[E[X|\mathcal{H}]|\mathcal{G}]$ by uniqueness. □

5.4.2. *Martingales and stuff.*

Definition 44. A stochastic process M_t on $(\Omega, \mathcal{F}, P)$ is martingale with respect to a filtration $\{\mathcal{M}_t\}_{t\geq 0}$ if for all t

(1) M_t is $\mathcal{M}_t$-measureable
(2) $E[|M_t|] < \infty$
(3) $E[M_t|\mathcal{M}_s] = M_s$ for all $s \leq t$

Another way to express property 3 is $E[M_t - M_s|\mathcal{M}_s] = 0$ which, if M_t is interpreted as the winnings from a game of chance, means that the average winnings from time s to time t is zero.

Proposition 45. *Brownian motion B_t is a martingale w.r.t. the natural filtration $\mathcal{F}_s$ that $\{B_s\}_{s\leq t}$ is adapted to.*

Proof. Firstly since

$$\begin{aligned}
(E[|B_t|])^2 &\leq E\left[|B_t|^2\right] \\
&= E\left[|(B_t - B_0) + B_0|^2\right] \\
&= E\left[|B_t - B_0|^2\right] + 2E[|(B_t - B_0) B_0|] + E\left[|B_0|^2\right] \\
&= nt + 0 + E\left[|B_0|^2\right] < \infty
\end{aligned}$$

it's the case that $E[|B_t|] < \infty$. So property 1 is satisfied. For property 2

$$\begin{aligned}
E[B_t|\mathcal{F}_s] &= E[B_t - B_s + B_s|\mathcal{F}_s] \\
&= E[B_t - B_s|\mathcal{F}_s] + E[B_s|\mathcal{F}_s] \\
&= 0 + B_s
\end{aligned}$$

□

So Brownian motion is a martingale. Who cares? We'll use this property and *Doob's martingale inequality*[26] to prove that there exists a version of

$$\int_0^t f(s,\omega)\, dB_s$$

that is continuous w.r.t. to t.

[24] If a finite measure ν is absolutely continuous with respect to a μ then there is a measurable function f such that for any measurable subset A it's the case that $\nu(A) = \int_A f d\mu$.

[25] A measure μ absolutely continuous with respect to measure ν if for every measurable set A, $\nu(A) = 0$ implies $\mu(A) = 0$.

[26] If M_t is a continuous martingale then for all $p \geq 1, T \geq 0, \lambda > 0$ it's the case that $P\left[\sup_{0\leq t\leq T}|M_t| \geq \lambda\right] \leq E[|M_T|^p]/\lambda^p$. This is reminiscent of Chebyshev's inequality $P[|X| \geq \lambda] \leq E[|X|^p]/\lambda^p$.

Theorem 46. *Let $f \in \mathcal{V}(0,T)$. Then there exists a t continuous stochastic process J_t such that*

$$P\left[J_t = \int_0^t f dB\right] = 1$$

for all t such that $0 \leq t \leq T$, i.e. $\int_0^t f dB$ is continuous function of its upper limit of integration.

Proof. Let $\phi_n = \phi_n(t,\omega) = \sum_j a_j(\omega) 1_{\left[t_j^{(n)}, t_{j+1}^{(n)}\right)}(t)$ be simple process that converge to f in Ito norm. Put

$$I_n(t,\omega) = \int_0^t \phi_n dB$$

and

$$I_t(\omega) = I(t,\omega) = \int_0^t f dB$$

Then I_n is continuous w.r.t t and martingale w.r.t $\mathcal{F}_t$ the filtration that B_t is adapted to: for $s \leq t$

$$\begin{aligned}
E\left[I_n(t,\omega)|\mathcal{F}_s\right] &= E\left[\left(\int_0^s \phi_n dB + \int_s^t \phi_n dB\right)|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + E\left[\left(\int_s^t \phi_n dB\right)|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + E\left[\left(\sum_j \int_{t_j^{(n)}}^{t_{j+1}^{(n)}} \phi_n dB\right)|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + E\left[\left(\sum_j a_j\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)\right)|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + \sum_j E\left[a_j\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + \sum_j E\left[E\left[a_j\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)|\mathcal{F}_{t_j}\right]|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + \sum_j E\left[a_j E\left[\left(B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}\right)|\mathcal{F}_{t_j}\right]|\mathcal{F}_s\right] \\
&= \int_0^s \phi_n dB + 0 = I_n(s,\omega)
\end{aligned}$$

Hence $I_n - I_m$ is also an $\mathcal{F}_t$-martingale and by Doob's inequality: for $p = 2$

$$\begin{aligned}
\lim_{n\to\infty} P\left[\sup_{0\leq t\leq T}\left|I_n(t,\omega) - I_m(t,\omega)\right| \geq \lambda\right] &\leq \lim_{n\to\infty}\frac{1}{\lambda^2}E\left[\left|I_n(T,\omega) - I_m(T,\omega)\right|^2\right] \\
&= \frac{1}{\lambda^2}\lim_{n\to\infty} E\left[\int_0^T \left(\phi_n(T,\omega) - \phi_m(T,\omega)\right)^2 ds\right] \\
&= 0
\end{aligned}$$

So the sequence of probabilities on the left is cauchy convergent for any λ. Choose a subsequence n_k such that for $\lambda = 2^{-k}$

$$P\left[\sup_{0\leq t\leq T}\left|I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)\right| > 2^{-k}\right] < 2^{-k}$$

Then by the Borel-Cantelli lemma[27]

$$P\left[\sup_{0\leq t\leq T}\left|I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)\right| > 2^{-k} \text{ infinitely often}\right] = 0$$

Therefore there exists $k_1(\omega)$ with probability 1 such that for $k \geq k_1(\omega)$

$$\sup_{0\leq t\leq T}\left|I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)\right| \leq 2^{-k}$$

[27] If the sum of the probabilities of the events A_n is finite $\sum_{n=1}^{\infty} P(A_n) < \infty$, then the probability that infinitely many of them occur is 0, that is, $P(\limsup_{n\to\infty} A_n) = 0$ since $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$.

and hence the subsequence I_{n_k} is uniformly convergent for $t \in [0,T]$ with probability 1 and the limit of I_{n_k}, denoted by $J_t(\omega)$ is t-continuous for $t \in [0,T]$ with probability one[28]. Finally since $I_{n_k}(t,\cdot) \to I_t(\cdot)$ in $L^2(P)$ by uniqueness[29]

$$P(I_t = J_t) = 1$$

for all $t \in [0,T]$. $\square$

From now on assume that we're always working with the t-continuous version of the Ito integral.

Corollary 47. *Let $f \in \mathcal{V}(0,T)$. Then $I_t = I(t,\omega)$ is a $\mathcal{F}_t$-martingale and therefore*

$$P\left[\sup_{0\le t\le T} |I_t| \ge \lambda\right] \le E\left[\int_0^T f^2 ds\right]/\lambda^2$$

Proof. Follows from continuity of I_t, Doob's inequality, and Ito isometry. $\square$

6. Ito's Formula and the Martingale Representation Theorem

To derive a calculus of stochastic integrals we take a counterintuitive approach. Note that

$$B_t = \int_0^t dB_s$$

and recall that example 40 shows

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t$$

or

$$\frac{1}{2}B_t^2 = \int_0^t B_s dB_s - \frac{1}{2}t$$

So the function $g(x) = \frac{1}{2}x^2$ does not map the Ito integral $x = B_t = \int_0^t dB_s$ into another Ito integral; in fact it's the combination of two integrals

$$\frac{1}{2}B_t^2 = \int_0^t B_s dB_s - \int_0^t \frac{1}{2} ds \tag{6.1}$$

Hence define the class of *Ito processes*

Definition 48. Let B_t be a 1-D Brownian motion on $(\Omega, \mathcal{F}, P)$. An Ito process (or *stochastic integral*) is a stochastic process X_t on $(\Omega, \mathcal{F}, P)$ of the form

$$X_t = X_0 + \int_0^t u(s,\omega)\, ds + \int_0^t v(s,\omega)\, dB_s \tag{6.2}$$

where $v \in \mathcal{V}(0,T)$ and

$$P\left(\int_0^t v^2 ds < \infty \text{ for } t \ge 0\right) = 1$$

and

$$P\left(\int_0^t |u|\, ds < \infty \text{ for } t \ge 0\right) = 1$$

This class of processes is closed under smooth maps. Eqn. 6.2 is more typically written in "differential form"

$$dX_t = u dt + v dB_t \tag{6.3}$$

For example eqn. 6.1 is written

$$d\left(\frac{1}{2}B_t^2\right) = \frac{1}{2}dt + B_t dB_t$$

The main tool in stochastic calculus is the Ito formula

Theorem 49. *(Ito formula) Let X_t be an ito process and $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$ then*

$$Y_t = g(t, X_t)$$

is again an ito process and

$$dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2} \cdot (dX_t)^2 \tag{6.4}$$

where $(dX_t)^2 = (dX_t) \cdot (dX_t)$ is computed according to

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0 \quad dB_t \cdot dB_t = dt$$

---

[28] Uniform limit theorem.

[29] Of limits in $L^2(P)$

Proof. We'll prove this by showing equality of the two sides of eqn. 6.4. First substitute eqn. 6.3

$$\begin{aligned}\int_0^t dY_t &= g(t,X_t) - g(0,X_0) = \int_0^t \left(\frac{\partial g}{\partial s} ds + \frac{\partial g}{\partial x}(uds + vdB_s) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\cdot (uds+vdB_s)^2\right) \\ g(t,X_t) &= g(0,X_0) + \int_0^t \left(\frac{\partial g}{\partial s} + u\frac{\partial g}{\partial x} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\cdot v^2\right) ds + \int_0^t v\frac{\partial g}{\partial x} dB_s\end{aligned}$$

Then assume "all the things"[30] are bounded and continuously differentiable[31] and that u, v are simple processes. Using Taylor's theorem to expand $\Delta g(t_j, X_j)$

$$\begin{aligned}g(t,X_t) &= g(0,X_0) + \sum_j \Delta g(t_j,X_j) \\ &= g(0,X_0) + \sum_j \left(\frac{\partial g}{\partial t}\Delta t_j + \frac{\partial g}{\partial x}\Delta X_j + \frac{1}{2}\left[\frac{\partial^2 g}{\partial t^2}\cdot (\Delta t_j)^2 + \frac{\partial^2 g}{\partial t \partial x}(\Delta t_j)(\Delta X_j) + \frac{\partial^2 g}{\partial x^2}\cdot (\Delta X_j)^2\right]\right) \\ &\quad + \sum_j R_j\end{aligned}$$

where $R_j = O\left(|\Delta t_j|^2 + |\Delta X_j|^2\right)$. If $\Delta t_j \to 0$ then

$$\sum_j \left(\frac{\partial g}{\partial t}\Delta t_j + \frac{\partial g}{\partial x}\Delta X_j\right) \to \int_0^t \frac{\partial g}{\partial t} ds + \int_0^t \frac{\partial g}{\partial x} dX_s$$

Since u, v are elementary

$$\begin{aligned}\sum_j \frac{\partial^2 g}{\partial x^2}\cdot (\Delta X_j)^2 &= \sum_j \frac{\partial^2 g}{\partial x^2}\cdot (u\Delta t_j + v\Delta B_j)^2 \\ &= \sum_j \frac{\partial^2 g}{\partial x^2}\cdot \left(u^2(\Delta t_j)^2 + uv(\Delta t_j)(\Delta B_j) + (v)^2 \Delta B_j\right) \qquad (6.5)\end{aligned}$$

The terms that include Δt_j go to zero, for example since $E\left(\Delta B_j^2\right) \to \Delta t_j$

$$E\left[\left(\sum_j \frac{\partial^2 g}{\partial x^2} u_j v_j \Delta t_j \Delta B_j\right)^2\right] = \sum_j E\left[\left(\frac{\partial^2 g}{\partial x^2} u_j v_j\right)^2\right](\Delta t_j)^3$$

which goes to 0 as $\Delta t_j \to 0$. The last term in eqn. 6.5 tends to

$$\int_0^t \frac{\partial^2 g}{\partial x^2} v^2 ds$$

in $L^2(P)$ as $\Delta t_j \to 0$. Why? Let $a_j = a(t_j) = \frac{\partial^2 g(t_j, X_{t_j})}{\partial x^2} v^2$ and consider the $L^2(P)$ norm quantity

$$E\left[\left(\sum_j \left(a_j(\Delta B_j)^2 - a_j \Delta t_j\right)\right)^2\right] = \sum_{i,j} E\left[a_i a_j \left((\Delta B_i)^2 - \Delta t_i\right)\left((\Delta B_j)^2 - \Delta t_j\right)\right]$$

If $i < j$ then $E\left[a_i a_j \left((\Delta B_i)^2 - \Delta t_i\right)\left((\Delta B_j)^2 - \Delta t_j\right)\right]$ vanishes by independence of increments. Similarly for $i > j$. Hence

$$\begin{aligned}E\left[\left(\sum_j \left(a_j(\Delta B_j)^2 - a_j\Delta t_j\right)\right)^2\right] &= \sum_j E\left[a_j^2\left((\Delta B_j)^2 - \Delta t_j\right)^2\right] \\ &= \sum_j E\left[a_j^2\right] E\left[(\Delta B_j)^4 + (\Delta t_j)^2 - (\Delta B_j)^2(\Delta t_j)\right] \\ &= \sum_j E\left[a_j^2\right]\left(3(\Delta t_j)^2 + (\Delta t_j)^2 - 2(\Delta t_j)^2\right) \\ &= 2\sum_j E\left[a_j^2\right](\Delta t_j)^2\end{aligned}$$

which goes to 0 as $\Delta t_j \to 0$. This establishes

$$\sum_j a_j (\Delta B_j)^2 \to \int_0^t a(s)\, ds$$

---

[30] $g, \partial_t g, \partial_x g, \partial_x^2 g$

[31] Or at least that there exists a sequence of functions that converge g and are as such.

which is written

$$(dB_t)^2 = dt$$

and the theorem because it establishes that $\sum_j R_j \to 0$ as $\Delta t_j \to 0$. $\square$

Example 50. As a curiousity what is

$$\int_0^t s dB_s$$

Using $g(t, X) = tX$ and $Y_t = g(t, B_t)$ we get a sort of integration by parts

$$dY_t = B_t dt + t dB_t + 0$$

i.e.

$$d(tB_t) = B_t dt + t dB_t$$

and so

$$\int d(tB_t) = tB_t = \int_0^t B_s ds + \int_0^t s dB_s$$

i.e.

$$\int_0^t s dB_s = tB_t - \int_0^t B_s ds$$

This is true in general

Theorem 51. *For $f(s, \omega) = f(s)$*

$$\int_0^t f(s) dB_s = f(t) B_t - \int_0^t B_s df_s$$

6.1. **Martingale Representation Theorem.** Every ito integral/process is a martingale (47). But in fact every $\mathcal{F}_t^{(n)}$-martingale[32] can be represented as an ito integral/process. First an aside on uniform integrability and martingale convergence.

6.1.1. *Uniform integrability and Martingale convergence.*

Definition 52. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A family $\{f_j\}$ of real, measurable functions, for some index set $j \in J$, is uniformly integrable if

$$\lim_{M \to \infty} \left(\sup_{j \in J} \left\{ \int_{\{|f_j| > M\}} |f_j| \, dP \right\} \right) = 0$$

A test for uniform integrability follows from the following defintion

Definition 53. $\psi : [0, \infty) \to [0, \infty)$ is called a uniform integrability test function if ψ is convex[33] and

$$\lim_{x \to \infty} \frac{\psi(x)}{x} = \infty$$

The following theorem is obvious given the names:

Theorem 54. *A family $\{f_j\}$ of real, measurable functions, for some index set $j \in J$, is uniformly integrable iff there exists a uniform integrability test function such that*

$$\sup_{j \in J} \left\{ \int \psi(|f_j|) \, dP \right\} < \infty$$

The use of the uniform integrability characterization is the following theorem, a sort of generalization of various convergence theorems:

Theorem 55. *Suppose $f_j \to f$ pointwise, i.e. for all ω*

$$\lim_{j \to \infty} f_j(\omega) = f(\omega)$$

Then the following are equivalent:

(1) *$\{f_j\}$ is uniformly integrable.*
(2) *$f \in L^1(P)$, i.e $\int |f| \, dP < \infty$, and $f_j \to f$ in $L^1(P)$*

The germaine application of uniform integrability is a convergence theorem for martingales:

Theorem 56. *(Doob's martingale convergence theorem) Let N_t be a right continuous[34] supermartingale[35] such that $\sup_{t>0} E[N_t^-] < \infty$, where $N_t^- := \max(-N_t, 0)$, i.e only the negative parts but reflected about 0. Then $N_t(\omega)$ converges pointwise almost surely and the limit N is such that $E[N^-] < \infty$. Furthermore if $\{N_t\}$ is uniformly integrable then the convergence is almost everywhere and in $L^1(P)$ (actually an iff).*

---

[32] The natural filtration that B_t is adapted to.

[33] $\psi(\lambda x + (1 - \lambda) y) \le \lambda \psi(x) + (1 - \lambda) \psi(y)$

[34] $t \to N_t$ is right continuous.

[35] Given that the filtration $\mathcal{N}_t$ contains all the null sets of $\mathcal{N}$ and $\mathcal{N}_t$ is right continuous, i.e. $\mathcal{N}_t = \cap_{s>t} \mathcal{N}_s$ then N_t is a *supermartingale* if $N_t \ge E[N_s | \mathcal{N}_t]$. Similarly for *submargtingale* but with $\le$.

Using Doob's martingale convergence theorem with the uniform integrability test function $\psi(x) = x^p$ we

Corollary 57. *If M_t is a continuous martingale such that $\sup_{t>0} E[|M_t|^p] < \infty$ for some $p > 1$ then $M_t \to M$ almost everywhere and in $L^1(P)$.*

Note M is simply whatever M_t converges to. Further note that similar results can be obtained for discrete time super/submartingales and as a corollary to that we have this result:

Corollary 58. *Let $X \in L^1(P)$ and $\{\mathcal{N}_k\}_{k=1}^{\infty}$ be a filtration in $\mathcal{F}$ and define $\mathcal{N}_\infty := \sigma\left(\{\mathcal{N}_k\}_{k=1}^{\infty}\right)$. Then*

$$E[X|\mathcal{N}_k] \to E[X|\mathcal{N}_\infty]$$

almost everywhere and in $L^1(P)$.

Proof. Set $M_k := E[X|\mathcal{N}_k]$. Then $\{M_k\}$ is a uniformly integrable martingale (since $X \in L^1(P)$), so there exists M such that $M_k \to M$ almost everywhere and in $L^1(P)$. Consider that

$$\begin{aligned} \|M_k - E[M|\mathcal{N}_k]\|_{L^1(P)} &= \|E[M_k|\mathcal{N}_k] - E[M|\mathcal{N}_k]\|_{L^1(P)} \\ &= \|E[M_k - M|\mathcal{N}_k]\|_{L^1(P)} \\ &\leq \|M_k - M\|_{L^1(P)} \to 0 \end{aligned}$$

as $k \to \infty$ because conditional expectation "contracts L^p norms"[36]. Hence if $F \in \mathcal{N}_{k_0}$ and $k > k_0$

$$\int_F (X - M)\, dP = \int_F E[X - M|\mathcal{N}_k]\, dP = \int_F (M_k - E[M|\mathcal{N}_k])\, dP \to 0$$

as $k \to \infty$. The first equality follows from the definition of conditonal expectation[37], the second is definition, and the third by just above. Therefore $\int_F (X - M)\, dP = 0$ for all $F \in \cup_{k=1}^{\infty}\mathcal{N}_k = \mathcal{N}_\infty$ and hence $E[X|\mathcal{N}_\infty] = E[M|\mathcal{N}_\infty] = M$. □

6.2. Back to Representation theorems.

Lemma 59. *Fix $T > 0$. For $t_i \in [0, T]$ and $\phi \in C_0^\infty(\mathbb{R}^n)$ (The set of smooth functions, with compact support, over $\mathbb{R}^n$) the set of random variables $\phi(B_{t_1}, \ldots, B_{t_n})$ is dense in $L^2(\mathcal{F}_T, P)$ (square Lebesgue integrable functions again the measure P defined on the sigma algebra $\mathcal{F}_T$, the filtration that B_t is adapted to).*

Proof. Let $\{t_i\}_{i=1}^{\infty}$ be a dense subset of $[0, T]$ and for $n = 1, 2, \ldots$ let $\mathcal{H}_n$ be the σ-algebra generated by $B_{t_1}, \ldots, B_{t_n}$. Then $\mathcal{H}_n \subset \mathcal{H}_{n+1}$ and $\mathcal{F}_T$ is the smallest σ-algebra containing all the $\mathcal{H}_n$. Then by corollary 58 to Doob's Martingale convergence theorem □

7. SDEs

Finally let's solve some god damn stochastic differential equations!

Example 60. The population growth model from Ch. 1 is

$$\frac{dN_t}{dt} = a_t N_t = (r_t + \alpha W_t)\, N_t$$

where W_t is white-noise. The Ito interpretation of this model is

$$dN_t = (r_t N_t dt + \alpha N_t dB_t)$$

or

$$\frac{dN_t}{N_t} = r_t dt + \alpha dB_t \tag{7.1}$$

or

$$\int_0^t \frac{dN_s}{N_s} = rt + \alpha B_t$$

To evaluate the left-hand side use the Ito formula with $g(t, X_t) = \ln(X_t)$ to get

$$\begin{aligned} d(\log N_t) &= \frac{1}{N_t} dN_t + \frac{1}{2}\left(-\frac{1}{N_t^2}\right)(dN_t)^2 \\ &= \frac{dN_t}{N_t} - \frac{1}{2N_t^2}\alpha^2 N_t dt = \frac{dN_t}{N_t} - \frac{1}{2}\alpha^2 dt \end{aligned}$$

36

Proof. $\|E[X|\mathcal{N}]\|_{L^p} \leq \|X\|_{L^p}$ iff $E[|E[X|\mathcal{N}]|^p] \leq E[|X|^p]$. By Jensen's and the fact that $|\cdot|^p$ is convex for $p \geq 1$ it's the case that

$$E[|E[X|\mathcal{N}]|^p] \leq E[E[|X|^p|\mathcal{N}]] = E[|X|^p]$$

Note as a corollary conditional expectation is L^p continuous, i.e. $X_n \xrightarrow{L^p} X \Rightarrow E[X_n|\mathcal{N}] \xrightarrow{L^p} E[X|\mathcal{N}]$. □

[37] $\int_{F \in \mathcal{N}} X dP = \int E[X|\mathcal{N}]\, dP$

and hence

$$\frac{dN_t}{N_t} = d(\ln N_t) + \frac{1}{2}\alpha^2 dt$$

and using eqn. 7.1

$$\begin{aligned} \int_0^t \left(d(\ln N_s) + \frac{1}{2}\alpha^2 ds \right) &= rt + \alpha B_t \\ \ln N_t - \ln N_0 + \frac{1}{2}\alpha^2 t &= rt + \alpha B_t \end{aligned}$$

or

$$N_t = N_0 e^{\left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t}$$

which is called *geometric Brownian motion*. Sanity check: on average the noise should filter out and this result should agree with the deterministic population growth ODE[38], i.e.

$$E[N_t] = E[N_0] e^{rt}$$

as if there were no noise term in the driving force $r_t dt + \alpha dB_t$. Indeed this is true. Let $Y_t = e^{\alpha B_t}$ and apply Ito's formula

$$dY_t = \frac{1}{2}\alpha^2 e^{\alpha B_t} dt + \alpha e^{\alpha B_t} dB_t$$

or

$$Y_t - Y_0 = \frac{1}{2}\alpha^2 \int_0^t e^{\alpha B_s} ds + \alpha \int_0^t e^{\alpha B_s} dB_s$$

and since $E\left[\int_0^t e^{\alpha B_s} dB_s\right] = 0$ by property 3 or Ito integral we have that

$$E[Y_t - Y_0] = \frac{1}{2}\alpha^2 \int_0^t E\left[e^{\alpha B_s}\right] ds$$

i.e.

$$\frac{d}{dt} E[Y_t] = \frac{1}{2}\alpha^2 E[Y_t], E[Y_0] = 1$$

which necessarily implies

$$E[Y_t] = e^{\frac{1}{2}\alpha^2 t}$$

which itself implies that

$$E[N_t] = E[N_0] e^{rt}$$

since

$$N_t = N_0 e^{\left(r - \frac{1}{2}\alpha^2\right)t + \alpha B_t} = N_0 e^{\left(r - \frac{1}{2}\alpha^2\right)t} Y_t$$

and by independence

$$\begin{aligned} E[N_t] &= E[N_0] e^{\left(r - \frac{1}{2}\alpha^2\right)t} E[Y_t] \\ &= E[N_0] e^{\left(r - \frac{1}{2}\alpha^2\right)t} e^{\frac{1}{2}\alpha^2 t} \\ &= E[N_0] e^{rt} \end{aligned}$$

Example 61. The noisy filter problem from Ch. 1 is

$$LQ_t^{''} + RQ_t^{'} + \frac{1}{C}Q_t = F_t = G_t + \alpha W_t$$

Reduce this second order SDE to a system of first order SDEs by introduced

$$X_t = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Q_t \\ Q_t^{'} \end{pmatrix}$$

Then the system is

$$\begin{aligned} X_1^{'} &= X_2 \\ LX_2^{'} &= -RX_2 - \frac{1}{C}X_1 + G_t + \alpha W_t \end{aligned}$$

or using the standard Ito interpretation

$$\begin{aligned} dX_1 &= X_2 dt \\ LdX_2 &= -RX_2 dt - \frac{1}{C}X_1 dt + G_t dt + \alpha dB_t \end{aligned}$$

or in matrix form

$$\begin{pmatrix} dX_1 \\ dX_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{1}{CL} & \frac{R}{L} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} dt + \begin{pmatrix} 0 \\ \frac{1}{L}G_t \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{L}\alpha \end{pmatrix} dBt$$

[38] $\frac{dn}{dt} = an$

This is a *2-dimensional stochastic differential equation*. Rewrite using the matrix exponential[39]

$$e^{-At}dX_t - e^{-At}AX_tdt = e^{-At}\left(H_tdt + KdB_t\right) \tag{7.2}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -\frac{1}{CL} & \frac{R}{L} \end{pmatrix} \quad H_t = \begin{pmatrix} 0 \\ \frac{1}{L}G_t \end{pmatrix} \quad K = \begin{pmatrix} 0 \\ \frac{1}{L}\alpha \end{pmatrix}$$

Given this "algebraic representation of the system it's tempting to rewrite the left side of eqn. 7.2 as

$$d\left(e^{-At}X_t\right) = e^{-At}dX_t - e^{-At}AX_tdt$$

But to be able to do this you need something like a multidimensional form of the Ito formula; good thing we have one!

Theorem 62. *Let*

$$dX_t = udt + vdB_t$$

be an n-dimensional Ito process and $g(t,X) = (g_1(t,X), \dots, g_p(t,X))$ *be* C^2 *map from* $[0,\infty) \times \mathbb{R}^n \to \mathbb{R}^p$. *Then*

$$Y_t = g(t, X_t)$$

is an Ito process whose components abide by the Ito formula

$$dY_k = \frac{\partial g_k}{\partial t}dt + (\nabla g)\cdot(dX) + \frac{1}{2}dX \cdot H_g \cdot dX$$

where $dB_{t,i}dB_{t,j} = \delta_{i,j}dt$ *and* $dB_{t,i}dt = dtdB_{t,i} = 0$.

And now back to our regularly scheduled program: with

$$g(t, X_1, X_2) = e^{-At}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

we indeed get that

$$d\left(e^{-At}X_t\right) = e^{-At}dX_t - e^{-At}AX_tdt$$

and substituting into eqn. 7.2 and using integration by parts

$$\begin{aligned} \int_0^t d\left(e^{-At}X_t\right) &= \int_0^t e^{-At}\left(H_tdt + KdB_t\right) \\ e^{-At}X_t - X_0 &= e^{At}\left[X_0 + e^{-At}KB_t + \int_0^t e^{-As}\left(H_s + AKB_s\right)ds\right] \end{aligned}$$

[39] $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$