Crash course in probability

Samuel N. Cohen
InFoMM

Hilary Term 2020

University of Oxford
Mathematical Institute

Oxford Mathematics

Introduction

Probabilistic setup

The basic setup on which we define our random objects is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Random variables are measurable functions $X : \Omega \to \mathbb{R}$.

We are interested in modelling random phenomena which evolve through time. We call them stochastic processes. Formally, a stochastic process is simply a collection of random variables (X_t) indexed by $t \in \mathcal{T}$. Basic examples are

- $\mathcal{T} = \{t_1, \ldots, t_n\}$ or $\mathcal{T} = \mathbb{N}$ – discrete time,
- $\mathcal{T} = \mathbb{R}_+$ or $\mathcal{T} = [0, T]$ – continuous time.

Probabilistic setup

UNIVERSITY OF OXFORD
Mathematical Institute

The basic setup on which we define our random objects is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Random variables are measurable functions $X : \Omega \to \mathbb{R}$.

We are interested in modelling random phenomena which evolve through time. We call them stochastic processes. Formally, a stochastic process is simply a collection of random variables (X_t) indexed by $t \in \mathcal{T}$. Basic examples are

- $\mathcal{T} = \{t_1, \ldots, t_n\}$ or $\mathcal{T} = \mathbb{N}$ – discrete time,
- $\mathcal{T} = \mathbb{R}_+$ or $\mathcal{T} = [0, T]$ – continuous time.

To encode the idea of temporal evolution we need to describe

- the flow of information,
- transition mechanism for X_t evolving into X_{t+1}.

Mathematical Institute

A key question is when two random variables are the same:

- We can say they agree in every state of the world
- We can say the agree with probability one (or *almost surely* – often written a.s. or w.p.1)

In general the latter is more useful, but it gets tricky when you have (infinite) families of random variables floating around.

Flow of Information – Filtrations

Filtrations

UNIVERSITY OF OXFORD
Mathematical Institute

The flow of information is encoded by a filtration.
A filtration $(\mathcal{F}_t)$ is an increasing sequence of σ-algebras,
$\mathcal{F}_u \subset \mathcal{F}_t \subset \mathcal{F}$, $u \leq t$. At time t only events in $\mathcal{F}_t$ are known.

The flow of information is encoded by a filtration.
A filtration $(\mathcal{F}_t)$ is an increasing sequence of σ-algebras, $\mathcal{F}_u \subset \mathcal{F}_t \subset \mathcal{F}$, $u \leq t$. At time t only events in $\mathcal{F}_t$ are known.

The basic example is $\mathcal{F}_t = \sigma(X_u : u \leq t)$, the natural filtration of X i.e. the minimal flow of information needed to know the value of X_t at time t.

However we may want to consider different filtrations.
From now on assume that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. All stochastic process will be adapted, i.e. X_t is $\mathcal{F}_t$–measurable.

Stopping times

Mathematical Institute

In practice we speak of various times

- 03 March 2017, Easter Sunday 2018 – deterministic dates,
- *the next time I see you*, *when this stock rises by at least* 20%, *when I either double or lose half of my money* etc. – well defined but random times.

Mathematical Institute

In practice we speak of various times

- 03 March 2017, Easter Sunday 2018 – deterministic dates,
- *the next time I see you*, *when this stock rises by at least* 20%, *when I either double or lose half of my money* etc. – well defined but random times.

Definition

A stopping time is a random variable $\tau : \Omega \to \mathcal{T} \cup \{\infty\}$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathcal{T}$.

The idea is that at any given time we can say if τ has occurred or not.

The above examples are stopping times. The following one is NOT: *when this stock achieves its maximum price in the next year.*

Naturally, stopping times are relative to a given filtration.
We fix $\mathcal{T} = \mathbb{N}$ and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$.

Lemma

Let (X_n) be an adapted stochastic process taking values in $\mathbb{R}^d$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$. Then

$$\inf\{n \geq 0 : X_n \in \Gamma\}$$

is a stopping time (the first hitting time).

For example, if S_t is the stock price process then stopping times of the form $\inf\{t : (t, S_t) \in \Gamma\}$ yield optimal exercise policies for American options.

UNIVERSITY OF OXFORD
Mathematical Institute

Lemma

A deterministic time is a stopping time. If τ, ρ are stopping times then

$$\tau \wedge \rho \quad \textit{and} \quad \tau \vee \rho$$

are also stopping times.

A random time τ is a stopping time iff

$$X_t = 1_{\{t \leq \tau\}} = \begin{cases} 1, & t \leq \tau, \\ 0, & t > \tau. \end{cases}$$

is adapted.

Stopping times - examples

We can look a the value of process at a random (stopping) time τ:

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega).$$

This is measurable with respect to $\mathcal{F}_\tau$, the σ-algebra of events 'known at time τ'. Formally, this is defined by

$$\mathcal{F}_\tau = \Big\{A \in \mathcal{F} \text{ such that } A \cap \{\tau \le n\} \in \mathcal{F}_n \text{ for all } n \ge 0\Big\}.$$

Note that if $\rho \le \tau$ are two stopping times then $\mathcal{F}_\rho \subset \mathcal{F}_\tau$.
The stopped process, often denoted X^τ, is given by $X^\tau_n = X_{\tau \wedge n}$.

UNIVERSITY OF OXFORD
Mathematical Institute

A coin-tossing game has the following rules. At each time we toss a coin if it is heads we win one pound, if it is tails, we lose one pound. The payoff is a random variable ξ with

$$\xi = \begin{cases} 1 & \text{if} \quad \mathtt{H} \\ -1 & \text{if} \quad \mathtt{T} \end{cases}$$

Let ξ_i be the outcome from game i, then $S_n = \sum_{i=1}^{n} \xi_i$ is the total winnings from the game after n rounds. This can be regarded as the position of a random walker after n steps, who starts at the origin and at each timestep moves one unit either to the left, if the toss was a $\mathtt{T}$, or to the right if the toss was a $\mathtt{H}$.

Simple Random Walk – cont.

It is easy to see that, by independence of ξ_i, if the coin is fair,

$$\mathbb{E}[S_n] = 0 \text{ and } var(S_n) = n.$$

Crucially, $S_n = S_{n-1} + \xi_n$. If we know today's value S_{n-1} then other information about past is irrelevant for S_n. This is the Markov property and S_n is a Markov process: *a stochastic process in which the future evolution depends on the past only through the present value:*

$$\mathbb{E}[g(X_n)|\,\mathcal{F}_{n-1}] = \mathbb{E}[g(X_n)|\sigma(X_{n-1})].$$

In order to describe the process completely we just need to determine its transition mechanism: probabilities of going from state x to state y.

Here, if $p_n(k) = \mathbb{P}(S_n = k)$, the probability that the walker is at position k after n steps, then

$$p_{n+1}(k) = \frac{1}{2}\left(p_n(k-1) + p_n(k+1)\right)$$

from which we can find $p_n(k)$ recursively. In fact S_n has a Binomial distribution. If the coin is biased we lose symmetry but we retain the Markov property. In continuous time we refer to (continuous) Markov processes as *diffusions*.

Conditional Expectation (recall)

UNIVERSITY OF OXFORD
Mathematical Institute

Definition

Conditional expectation of X given σ-algebra $\mathcal{F}$ is an $\mathcal{F}$-measurable random variable Y, often written as $Y = \mathbb{E}[X|\,\mathcal{F}]$, such that

$$\mathbb{E}[X1_A] = \mathbb{E}[Y1_A], \quad \forall_{A\in\mathcal{F}}.$$

Lemma

The conditional expectation $\mathbb{E}[\cdot|\,\mathcal{F}]$ is a linear operator which satisfies

- *if X is $\mathcal{F}$ measurable then $\mathbb{E}[X|\,\mathcal{F}] = X$ and if X is independent of $\mathcal{F}$ then $\mathbb{E}[X|\,\mathcal{F}] = \mathbb{E}[X]$;*
- *if $\mathcal{F} \subset \mathcal{G}$ then $\mathbb{E}[X|\,\mathcal{F}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\,\mathcal{F}]$;*

Definition

A real-valued process (X_t) with $\mathbb{E}|X_t| < \infty$, $t \geq 0$ is called

- an $(\mathcal{F}_t)$–martingale if for any $u < t$ we have $\mathbb{E}[X_t | \mathcal{F}_u] = X_u$,
- an $(\mathcal{F}_t)$–submartingale if for any $u < t$ we have $\mathbb{E}[X_t | \mathcal{F}_u] \geq X_u$,
- an $(\mathcal{F}_t)$–supermartingale if for any $u < t$ we have $\mathbb{E}[X_t | \mathcal{F}_u] \leq X_u$.

A martingale is a mathematical model for an equitable game: we have $\mathbb{E}[X_n] = X_0$ and, more generally, my expected gain in the next round is always zero: $\mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = 0$.

A submartingale models a profitable game.

Mathematical Institute

Example (Example 1)

- Consider two players A and B with some capitals a and b.
- They play a simple repetitive game, after each round one of them wins $1 from the other.
- The game stops when one of the players is ruined.
- We are interested in the $\mathbb{P}(\mathrm{A\ wins})$.

Properties

- Martingales have constant mean. For $t > s > 0$ we have

$$\mathbb{E}(M_s) = \mathbb{E}(\mathbb{E}(M_t | \mathcal{F}_s)) = \mathbb{E}(M_t).$$

  Thus $\mathbb{E}(M_t) = \mathbb{E}(M_0)$ for all $t \geq 0$.

- For submartingales we have

$$\mathbb{E}(M_t) \geq \mathbb{E}(M_s) \geq \mathbb{E}(M_0), \text{ for all } 0 \leq s \leq t.$$

- For supermartingales we have

$$\mathbb{E}(M_t) \leq \mathbb{E}(M_s) \leq \mathbb{E}(M_0), \text{ for all } 0 \leq s \leq t.$$

Mathematical Institute

Example (Example 1 – cont.)

- Let p be the probability that (in each round) the player A eventually wins.
- We denote the outcome of the n^{th} round by ξ_n and $S_n = \xi_1 + \ldots + \xi_n$.
- The capital of player A after n rounds is thus $a + S_n$ and similarly for player B is $b - S_n$.
- We will use the stopping times

$$\tau^A = \inf\{n : a+S_n = 0\}, \ \tau^B = \inf\{n : b-S_n = 0\}, \ \text{ and } \tau = \tau^A \wedge \tau^B.$$

- Observe that $\tau < \infty$ a.s. while τ^A, τ^B may have a positive probability of being infinite. $\mathbb{P}(\text{A wins}) = \mathbb{P}(\tau^A > \tau^B)$.

University of Oxford Mathematical Institute

Example (Example 1 – cont.)

We know that ξ_n are i.i.d. random variables with $\mathbb{P}(\xi_n = 1)$, $\mathbb{P}(\xi_n = -1) = q = 1 - p$. Let $\mathcal{F}_n = \sigma(\xi_i : i \leq n)$.

- If $p = 1/2$ then $S_n = \sum_{i=1}^n \xi_n$ is an $(\mathcal{F}_n)$–martingale.
- If $p \neq 1/2$ then $\left(\frac{q}{p}\right)^{S_n}$ is an $(\mathcal{F}_n)$–martingale (and S_n is a submartingale or a supermartingale).
- $\frac{\exp S_n}{(\mathbb{E}[\exp(\xi_1)])^n}$ is a martingale,
- $S_n^2 - n$ is a martingale if $p = 1/2$.

Martingales – examples (cont.)

Mathematical Institute

Example

Let (X_n) be an $(\mathcal{F}_n)$–martingale and V_n a sequence of random variables s.t. V_n is $\mathcal{F}_{n-1}$ measurable. Then

$$Z_n = V_0 X_0 + V_1(X_1 - X_0) + \ldots + V_n(X_n - X_{n-1})$$

is an $(\mathcal{F}_n)$–martingale. It is the discrete version of stochastic integral.

Example

Let (X_t) be a martingale and τ a stopping time. Then $X_t^\tau := X_{t\wedge\tau}$ is a martingale. It is called the stopped process.

Example

Let (X_t) be a martingale and $f \geq 0$ a convex function with $\mathbb{E}[f(X_t)] < \infty$. Then $Y_t = f(X_t)$ is a submartingale.

Convergence and Stopping

University of Oxford Mathematical Institute

Theorem

A supermartingale with $\sup_n \mathbb{E}[X_n^-] < \infty$ *converges a.s. to an integrable random variable denoted* X_∞.

Doob's idea for the proof was to show that a martingale can only cross an interval, however small, finitely often. Let $a < b$ and

$$T_0 = \inf\{n : X_n \leq a\},\ T_1 = \inf\{n \geq T_0 : X_n \geq b\},\ T_{2k} = \inf\{n \geq T_{2k-1} : X_n \leq a\}..$$

Let $U_a^b(n)$ be the number of upcrossings of $[a, b]$ before time n:

$$\{U_a^b(n) = k\} = \{T_{2k-1} \leq n < T_{2k+1}\}.$$

Lemma

$$\mathbb{E}\left[U_a^b(n)\right] \leq \frac{1}{b-a}\mathbb{E}[(X_n - a)^-]$$

Theorem

For a martingale (X_n), the following three conditions are equivalent

- $\lim_{n\to\infty} X_n$ *exists in the L^1–sense;*
- *there exists a r.v. X_∞ in L^1 such that $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$;*
- *the family $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable.*

The convergence is then also almost sure.

By the Dominated Convergence Thm, the above holds if $|X_n| \leq Y \in L^1$. In particular, a bounded martingale converges a.s. and in L^1.

- Let Y_n be the result of n^{th} toss of a coin: $Y_n = 1$ for heads and 0 for tails.
- We suspect the coin is biased as we would like to decide whether $\mathbb{P}(Y = 1)$ is p or r, for some $p \neq r$.
- Define

$$\rho(1) = r, \;\; \rho(0) = 1 - r, \quad \pi(1) = p, \;\; \pi(0) = 1 - p,$$

so that we have the ratio of likelihoods

$$X_n := \frac{\rho(Y_1) \cdot \ldots \cdot \rho(Y_n)}{\pi(Y_1) \cdot \ldots \cdot \pi(Y_n)}.$$

Application to a statistical experiment

Suppose $\mathbb{P}(Y_n = 1) = p$. Then (X_n) is a martingale in its natural filtration, $\mathbb{E}\, X_n = X_0 = 1$. Further $X_n \geq 0$ and hence X_n converges a.s.

$$X_{n+1} = X_n \cdot \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})}, \quad \text{with } \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})} \in \left\{ \frac{r}{p}, \frac{1-r}{1-p} \right\}$$

and hence $\lim X_n = 0$ a.s.

Application to a statistical experiment

Suppose $\mathbb{P}(Y_n = 1) = p$. Then (X_n) is a martingale in its natural filtration, $\mathbb{E}\, X_n = X_0 = 1$. Further $X_n \geq 0$ and hence X_n converges a.s.

$$X_{n+1} = X_n \cdot \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})}, \quad \text{with } \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})} \in \left\{ \frac{r}{p}, \frac{1-r}{1-p} \right\}$$

and hence $\lim X_n = 0$ a.s.

Similarly, if $\mathbb{P}(Y_n = 1) = r$ then $1/X_n$ is a martingale and $\lim_n X_n = \infty$ a.s.

Application to a statistical experiment

Suppose $\mathbb{P}(Y_n = 1) = p$. Then (X_n) is a martingale in its natural filtration, $\mathbb{E}\, X_n = X_0 = 1$. Further $X_n \geq 0$ and hence X_n converges a.s.

$$X_{n+1} = X_n \cdot \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})}, \quad \text{with } \frac{\rho(Y_{n+1})}{\pi(Y_{n+1})} \in \left\{ \frac{r}{p}, \frac{1-r}{1-p} \right\}$$

and hence $\lim X_n = 0$ a.s.

Similarly, if $\mathbb{P}(Y_n = 1) = r$ then $1/X_n$ is a martingale and $\lim_n X_n = \infty$ a.s.

We conclude that observation of X_n may help us decide whether $\mathbb{P}(Y_n = 1) = p$ or $\mathbb{P}(Y_n = 1) = r$.

Application to a statistical experiment – cont.

University of Oxford
Mathematical Institute

The above experiment was trying to decide which probability measure is the "true" one: $\mathbb{P}$ with $\mathbb{P}(Y_n = 1) = p$ or $\mathbb{Q}$ with $\mathbb{Q}(Y_n = 1) = r$?
Let us investigate $\mathbb{E}^{\mathbb{P}}[X_1 f(Y_1)]$ for a function $f(0) = f_0$, $f(1) = f_1$. We have

$$\mathbb{E}^{\mathbb{P}}[X_1 f(Y_1)] = p\frac{r}{p}f_1 + (1-p)\frac{1-r}{1-p}f_0 = rf_1 + (1-r)f_0 = \mathbb{E}^{\mathbb{Q}}[f(Y_1)]$$

that is X_1 allows to change the measure from $\mathbb{P}$ to $\mathbb{Q}$ (at least on $\mathcal{F}_1$). In fact we have $\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}|_{\mathcal{F}_n} = X_n$ (we investigate such changes of measures later).

Remark. As $n \to \infty$, $X_n \to 0$ and the two measures become singular – asymptotically they have a.s. different behaviour and hence we cannot switch between them.

Optional Sampling (stopping) Thm

Note that if τ is a bounded stopping time, e.g. $\tau = N$, then X_n^τ satisfies the conditions of the L^1 convergence theorem and $X_\infty^\tau = X_\tau$, $X_{n\wedge\tau} = \mathbb{E}[X_\tau | \mathcal{F}_n] = \mathbb{E}[X_\tau | \mathcal{F}_{\tau\wedge n}]$, $n \leq N$. In fact we have:

University of Oxford
Mathematical Institute

Note that if τ is a bounded stopping time, e.g. $\tau = N$, then X_n^τ satisfies the conditions of the L^1 convergence theorem and $X_\infty^\tau = X_\tau$, $X_{n\wedge\tau} = \mathbb{E}[X_\tau | \mathcal{F}_n] = \mathbb{E}[X_\tau | \mathcal{F}_{\tau\wedge n}]$, $n \leq N$. In fact we have:

Theorem (Optional Stopping Theorem)

If X is a martingale and τ, ρ are two bounded stopping times, $\rho \leq \tau$ then

$$X_\rho = \mathbb{E}[X_\tau | \mathcal{F}_\rho], \text{ a.s.}$$

If X is uniformly integrable, the above holds for any stopping times $\rho \leq \tau$.

UNIVERSITY OF OXFORD
Mathematical Institute

Note that if τ is a bounded stopping time, e.g. $\tau = N$, then X_n^τ satisfies the conditions of the L^1 convergence theorem and $X_\infty^\tau = X_\tau$, $X_{n\wedge\tau} = \mathbb{E}[X_\tau | \mathcal{F}_n] = \mathbb{E}[X_\tau | \mathcal{F}_{\tau\wedge n}]$, $n \leq N$. In fact we have:

Theorem (Optional Stopping Theorem)

If X is a martingale and τ, ρ are two bounded stopping times, $\rho \leq \tau$ then

$$X_\rho = \mathbb{E}[X_\tau | \mathcal{F}_\rho], \; a.s.$$

If X is uniformly integrable, the above holds for any stopping times $\rho \leq \tau$.

In particular a stopped martingale is a martingale.

A converse result is possible:

Theorem

Suppose M is an adapted (right continuous) process such that $\mathbb{E}[|M_\tau|] < \infty$ for every bounded stopping time τ. Then M is a martingale if and only if $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$ for every bounded stopping time.

Note: These results hold equally in continuous and discrete time.

Example 1 – cont

Recall our simple example with two players A and B with initial capitals a and b who play a simple repetitive game, after each round one of them wins \$1 from the other. We are interested in $\mathbb{P}(\text{A wins}) = \mathbb{P}(\tau^A > \tau^B) =: p^A = 1 - p^B$.

Example 1 – cont

Recall our simple example with two players A and B with initial capitals a and b who play a simple repetitive game, after each round one of them wins \$1 from the other. We are interested in $\mathbb{P}(\text{A wins}) = \mathbb{P}(\tau^A > \tau^B) =: p^A = 1 - p^B$.

If $p = 1/2$ then S_n is a martingale and hence by the optional stopping theorem

$$\mathbb{E}\, S_{\tau\wedge n} = 0, \quad \forall n > 0.$$

Further $|S_{\tau\wedge n}| \leq a \vee b$ and $S_{\tau\wedge n} \to S_\tau$ a.s. since $\tau < \infty$ a.s. By the dominated convergence theorem we conclude that

$$\mathbb{E}\, S_\tau = 0, \ \ i.e. \quad (-a)(1 - p^A) + bp^A = 0, \ \text{and hence } p^A = \frac{a}{a+b}.$$

Example 1 – cont

University of Oxford Mathematical Institute

If $p \neq 1/2$ then $(q/p)^{S_n}$ is a martingale and hence, using similar arguments as previously,

$$\mathbb{E}\left(\frac{q}{p}\right)^{S_\tau} = 1, \; i.e. \quad \left(\frac{q}{p}\right)^{-a}(1 - p^A) + \left(\frac{q}{p}\right)^{b} p^A = 1,$$

which implies $p^A = \dfrac{\left(\frac{q}{p}\right)^{a} - 1}{\left(\frac{q}{p}\right)^{a+b} - 1}$.

UNIVERSITY OF OXFORD
Mathematical Institute

What happens if the player B is a bank with ∞ capital? We take limit as $b \to \infty$ to get (but this needs a proof!)

$$p^A = \begin{cases} 0, & \text{for } p \leq 1/2, \\ 1 - (q/p)^a, & \text{for } p \geq 1/2. \end{cases}$$

Exercise

We start with an urn with black and white balls.

- *We have initially b_0 black balls and w_0 white balls.*
- *Each time we draw a ball we put it back into the urn together with an additional $m \geq 1$ balls of its colour.*
- *We denote by $X_n = b_n/(b_n + w_n)$ the fraction of black balls in the urn when the nth draw is completed.*

a) *Show that X is a martingale (in its own filtration).*

b) *Deduce the probability that in the nth draw we pick a black ball.*

c) *Using an appropriate martingale convergence theorem describe the asymptotic behaviour of the fraction of black balls in the urn. Can we specify its expectation?*

Changes of Measure

Finite space

Let us start with a simple coin tossing.

- Players A and B play the following game.
- A machine tosses a coin twice and shows $X = 1$ or $X = 0$ accordingly for outcomes HH,HT or TH,TT.
- Player A wins if $X = 1$

Finite space

Suppose player A has some control over the machine and is thinking how to increase his odds. He can do two things:

- He can redefine the variable X. The machine could now flash $X = 1$ for HH,HT,TT giving him a chance of $\frac{3}{4}$ instead of $\frac{1}{2}$.

Finite space

Suppose player A has some control over the machine and is thinking how to increase his odds. He can do two things:

- He can redefine the variable X. The machine could now flash $X = 1$ for HH,HT,TT giving him a chance of $\frac{3}{4}$ instead of $\frac{1}{2}$.
- He can change the fair coin to a baised coin which has H with probability $p > \frac{1}{2}$. He then increases his chances of winning to $p^2 + p(1-p) = p$.

Suppose player A has some control over the machine and is thinking how to increase his odds. He can do two things:

- He can redefine the variable X. The machine could now flash $X = 1$ for HH,HT,TT giving him a chance of $\frac{3}{4}$ instead of $\frac{1}{2}$.
- He can change the fair coin to a baised coin which has H with probability $p > \frac{1}{2}$. He then increases his chances of winning to $p^2 + p(1-p) = p$.

The first option keeps $(\Omega, \mathbb{P})$ fixed and changes the values of random variables.

The second option keeps Ω and X fixed and changes the measure $\mathbb{P}$.

Change of measure

UNIVERSITY OF OXFORD
Mathematical Institute

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $D \geq 0$ be a random variable with $\mathbb{E}^{\mathbb{P}}[D] = 1$. We can define a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by

$$\mathbb{Q}(\Gamma) = \int_{\Gamma} D(\omega)\, \mathbb{P}(\mathrm{d}\,\omega) = \mathbb{E}^{\mathbb{P}}[D 1_{\Gamma}], \quad \Gamma \in \mathcal{F}\,.$$

We denote this $\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} = D$ and D is called the (Radon-Nikodym) density of $\mathbb{Q}$ with respect to $\mathbb{P}$.

Change of measure

Note that when $D > 0$ the reverse density $\frac{\mathrm{d}\,\mathbb{P}}{\mathrm{d}\,\mathbb{Q}} = \frac{1}{D}$ is well defined and allows us to change measure back from $\mathbb{Q}$ to $\mathbb{P}$. We say that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent: $\mathbb{P} \sim \mathbb{Q}$.

Note, if $\mathbb{P}(D = 0) > 0$ then the change of measure is singular – some events of positive $\mathbb{P}$-probability become $\mathbb{Q}$-negligible. It is not possible to start with $\mathbb{Q}$ and define $\mathbb{P}$.

UNIVERSITY OF OXFORD
Mathematical Institute

Note that when $D > 0$ the reverse density $\frac{\mathrm{d}\,\mathbb{P}}{\mathrm{d}\,\mathbb{Q}} = \frac{1}{D}$ is well defined and allows us to change measure back from $\mathbb{Q}$ to $\mathbb{P}$. We say that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent: $\mathbb{P} \sim \mathbb{Q}$.

Note, if $\mathbb{P}(D = 0) > 0$ then the change of measure is singular – some events of positive $\mathbb{P}$-probability become $\mathbb{Q}$-negligible. It is not possible to start with $\mathbb{Q}$ and define $\mathbb{P}$.

Definition

We say that $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ if, for any $\Gamma \in \mathcal{F}$, $\mathbb{P}(\Gamma) = 0$ implies $\mathbb{Q}(\Gamma) = 0$. We write $\mathbb{Q} \ll \mathbb{P}$.
We say that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if $\mathbb{P} \gg \mathbb{Q}$ and $\mathbb{Q} \gg \mathbb{P}$.

Theorem (Radon–Nikodym)

Let $\mathbb{P}, \mathbb{Q}$ be two (probability) measures. If $\mathbb{Q} \ll \mathbb{P}$ then there exists a r.v. $D \geq 0$ (denoted $\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}$) such that

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{Q}}[1_A] = \mathbb{E}^{\mathbb{P}}[D1_A], \quad \forall\, A \in \mathcal{F}\,.$$

In particular $\mathbb{E}^{\mathbb{P}}[D] = 1$. $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if and only if $D > 0$ $\mathbb{P}$-a.s.

Theorem (Radon–Nikodym)

Let $\mathbb{P}, \mathbb{Q}$ be two (probability) measures. If $\mathbb{Q} \ll \mathbb{P}$ then there exists a r.v. $D \geq 0$ (denoted $\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}$) such that

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{Q}}[1_A] = \mathbb{E}^{\mathbb{P}}[D 1_A], \quad \forall\, A \in \mathcal{F}\,.$$

In particular $\mathbb{E}^{\mathbb{P}}[D] = 1$. $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if and only if $D > 0$ $\mathbb{P}$-a.s.

Remark. Classic example is when $\mathbb{P}$ is Lebesgue measure and $\mathbb{Q}$ is a continuous probability distribution, with density $\frac{d\,\mathbb{Q}}{d\,\mathbb{P}} = f(x)$. In statistics, this is also called the likelihood.

So far we have not paid any attention to information or σ-algebras.
How does conditional expectation transform under change of measure?

So far we have not paid any attention to information or σ-algebras. How does conditional expectation transform under change of measure?
Suppose we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subset \mathcal{F}$. Let $D > 0$ be an $\mathcal{F}$–measurable r.v., $\mathbb{E}^{\mathbb{P}}[D] = 1$, and define $\mathbb{Q}$ via $\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} = D$. Let ξ be an $\mathcal{F}$–measurable r.v. Then

$$\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}] = \frac{1}{\mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]}\,\mathbb{E}^{\mathbb{P}}[\xi D|\,\mathcal{G}]$$

Proof.

By definition $\mathbb{E}^{\mathbb{Q}}[\xi|\mathcal{G}]$ is a $\mathcal{G}$–measurable random variable such that

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\xi|\mathcal{G}]1_A\right] = \mathbb{E}^{\mathbb{Q}}[\xi 1_A], \quad \forall A \in \mathcal{G}\,.$$

We have

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{1}{\mathbb{E}^{\mathbb{P}}[D|\mathcal{G}]}\,\mathbb{E}^{\mathbb{P}}[\xi D|\mathcal{G}]1_A\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{D}{\mathbb{E}^{\mathbb{P}}[D|\mathcal{G}]}\,\mathbb{E}^{\mathbb{P}}[\xi D|\mathcal{G}]1_A\right]$$
$$= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[\frac{D}{\mathbb{E}^{\mathbb{P}}[D|\mathcal{G}]}\,\mathbb{E}^{\mathbb{P}}[\xi D|\mathcal{G}]1_A\middle|\mathcal{G}\right]\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{\mathbb{E}^{\mathbb{P}}[\xi D|\mathcal{G}]1_A}{\mathbb{E}^{\mathbb{P}}[D|\mathcal{G}]}\,\mathbb{E}^{\mathbb{P}}[D|\mathcal{G}]\right]$$
$$= \mathbb{E}^{\mathbb{P}}[\xi D 1_A] = \mathbb{E}^{\mathbb{Q}}[\xi 1_A]$$

□

We can approach the problem from a different angle. We may ask: how to define a measure $\tilde{\mathbb{Q}}$ on $(\Omega, \mathcal{G})$ in such a way that $\tilde{\mathbb{Q}} = \mathbb{Q}$ on $\mathcal{G}$?

Change of measure and conditional expectation (2)

University of Oxford
Mathematical Institute

We can approach the problem from a different angle. We may ask: how to define a measure $\tilde{\mathbb{Q}}$ on $(\Omega, \mathcal{G})$ in such a way that $\tilde{\mathbb{Q}} = \mathbb{Q}$ on $\mathcal{G}$?

Naturally we need $\frac{\mathrm{d}\,\tilde{\mathbb{Q}}}{\mathrm{d}\,\mathbb{P}} = D_0 := \mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]$. Since if η is a $\mathcal{G}$-measurable r.v. then

$$\mathbb{E}^{\tilde{\mathbb{Q}}}[\eta] = \mathbb{E}^{\mathbb{P}}[\eta D_0] = \mathbb{E}^{\mathbb{P}}[\eta\,\mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\eta D|\,\mathcal{G}]] = \mathbb{E}^{\mathbb{P}}[\eta D] = \mathbb{E}^{\mathbb{Q}}[\eta].$$

We write this as

$$\left.\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right|_{\mathcal{G}} = D_0$$

Change of measure and conditional expectation (2)

University of Oxford
Mathematical Institute

We can approach the problem from a different angle. We may ask: how to define a measure $\tilde{\mathbb{Q}}$ on $(\Omega, \mathcal{G})$ in such a way that $\tilde{\mathbb{Q}} = \mathbb{Q}$ on $\mathcal{G}$?

Naturally we need $\frac{\mathrm{d}\,\tilde{\mathbb{Q}}}{\mathrm{d}\,\mathbb{P}} = D_0 := \mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]$. Since if η is a $\mathcal{G}$-measurable r.v. then

$$\mathbb{E}^{\tilde{\mathbb{Q}}}[\eta] = \mathbb{E}^{\mathbb{P}}[\eta D_0] = \mathbb{E}^{\mathbb{P}}[\eta\,\mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\eta D|\,\mathcal{G}]] = \mathbb{E}^{\mathbb{P}}[\eta D] = \mathbb{E}^{\mathbb{Q}}[\eta].$$

We write this as

$$\left.\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right|_{\mathcal{G}} = D_0 = \mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\middle|\,\mathcal{G}\right]$$

Change of measure and conditional expectation (2)

Going back to $\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}]$, this is a random variable such that

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}]1_A\right] = \mathbb{E}^{\mathbb{Q}}[\xi 1_A] = \mathbb{E}^{\mathbb{P}}[\xi D 1_A], \quad \forall A \in \mathcal{G},$$

but

$$\mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}]1_A\right] = \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}]1_A\,\mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}]\right]$$

and comparing both sides it follows that

$$\mathbb{E}^{\mathbb{Q}}[\xi|\,\mathcal{G}]\,\mathbb{E}^{\mathbb{P}}[D|\,\mathcal{G}] = \mathbb{E}^{\mathbb{P}}[\xi D|\,\mathcal{G}].$$

UNIVERSITY OF OXFORD
Mathematical Institute

Suppose now that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. Fix some time $T > 0$ and let $D > 0$ be an $\mathcal{F}_T$-measurable r.v., $\mathbb{E}^{\mathbb{P}}[D] = 1$. Then we can define a measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\left.\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right|_{\mathcal{F}_T} = D.$$

UNIVERSITY OF OXFORD
Mathematical Institute

Suppose now that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. Fix some time $T > 0$ and let $D > 0$ be an $\mathcal{F}_T$-measurable r.v., $\mathbb{E}^{\mathbb{P}}[D] = 1$. Then we can define a measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by

$$\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\bigg|_{\mathcal{F}_T} = D.$$

It follows that on $\mathcal{F}_t$ for $t < T$ the measure $\mathbb{Q}$ has density given by

$$\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\bigg|_{\mathcal{F}_t} = \mathbb{E}^{\mathbb{P}}[D|\mathcal{F}_t] =: D_t.$$

By definition, $(D_t : t \leq T)$ is a non-negative $\mathbb{P}$-martingale. We conclude that relative to a given filtration, the change of measure density is a non-negative martingale of expectation 1.

Suppose now that we have a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with $\mathcal{F} = \mathcal{F}_\infty$ is generated by $\cup_t \mathcal{F}_t$. From a non-negative martingale (D_t), $\mathbb{E}^{\mathbb{P}}[D_t] = 1$, we may hope to define a measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ via

$$\left.\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}}\right|_{\mathcal{F}_t} = D_t, \quad t \geq 0.$$

Then $\mathbb{Q} \ll \mathbb{P}$ on any $\mathcal{F}_t$ and if $D_t > 0$ $\mathbb{P}$-a.s. then in fact $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}_t$. However this does not tell us anything about the relation of $\mathbb{Q}$ and $\mathbb{P}$ on $\mathcal{F}$!

If (D_t) is a uniformly integrable martingale then it converges $D_t \to D$ and $\mathbb{E}^{\mathbb{P}}[D|\mathcal{F}_t] = D_t$, so we can extend the equivalence to $\mathcal{F}$ by putting

$$\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} = D.$$

If (D_t) is a uniformly integrable martingale then it converges $D_t \to D$ and $\mathbb{E}^{\mathbb{P}}[D|\mathcal{F}_t] = D_t$, so we can extend the equivalence to $\mathcal{F}$ by putting

$$\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} = D.$$

In contrast, in the statistical experiment example above, $D_t \to 0$ a.s. but not in L^1 and the above argument fails.

Continuous time

Definition

A stochastic process (W_t) is called a Brownian motion (relative to $(\mathcal{F}_t)$) if

- $W_0 = 0$, $\mathbb{E}[W_t] = 0$, $W_t \neq 0$ *a.s.*,
- W_t *has continuous paths,*
- (W_t) *is adapted and for any* $u < t$, $W_t - W_u$ *is independent of* $\mathcal{F}_u$.
- $W_t - W_u$ *is distributed as* $N(0, t - u)$.

- An alternative definition is that Brownian motion is a continuous Gaussian process (all finite dimensional marginal distributions have a multivariate normal distribution) with mean 0 and covariance $\mathbb{E}(W_t W_s) = \min(t, s)$.
- Brownian motion is also a martingale and our focus today will be developing ideas of stochastic calculus for general continuous martingales.

Some simple martingales based on Brownian motion

- Brownian motion is a martingale.
- $M_t = W_t^2 - t$ is a martingale
- $M_t = \exp(\theta W_t - \frac{1}{2}\theta^2 t)$ is a martingale

The Poisson process N_t is the number of random events that happen by time t where events happen at rate λ. Natural Poisson process martingales are

- $M_t = N_t - \lambda t$
- $M_t = \exp(\theta N_t - \lambda t e^{\theta})$

Many interesting properties can be shown to hold for (sub-)martingales. This is one which is very useful to prove theorems, and also for getting some intuition:

Lemma

A submartingale X can be decomposed as $X_t = M_t + A_t$, where M is a martingale and A is a predictable increasing process.

In discrete time, predictable means that A_t is $\mathcal{F}_{t-1}$-measurable for all t. In continuous time, it's that it is determined by the values of (i.e. measurable with respect to the σ-algebra generated by) continuous adapted processes. Usually A will be continuous.

Convergence of Martingales

UNIVERSITY OF OXFORD
Mathematical Institute

- We have already seen some results on the convergence of (sub-)martingales in discrete time.
- We have also seen optional stopping results.
- In continuous time, the same results hold, provided our martingales are (right-)continuous
- For all reasonable settings, martingales can be assumed to be right continuous (this is proven using discrete time convergence results)
- The proof usually goes: check the result in discrete time, apply this on subsets of the rationals, take limits using continuity.

Quadratic variation

UNIVERSITY OF OXFORD
Mathematical Institute

- We wish to know about the *volatility* of a (continuous) Martingale through time. This is formally done by considering the *quadratic variation*.
- For Brownian motion, we consider $Q_n(t) = \sum_{i=1}^{n}(W_{t_i} - W_{t_{i-1}})^2$, where $0 = t_0 < t_1 < \cdots < t_n = t$ is a partition of the time interval $[0, t]$.
- Problem: we can show that, taking the mesh to zero, this quantity will explode with probability one.

Quadratic Variation

- Let's take limits in a different way...
- We can check that for all t

$$\mathbb{E}\, Q_n(t) = t, \;\; \text{var}(Q_n(t)) = 2t^2/n.$$

- Thus as $n \to \infty$ we have a miracle

$$Q_n(t) \to t \text{ in } L^2.$$

That is the random variable $Q_n(t)$ becomes deterministic as the mesh of the partition goes to 0!

- For a general definition of quadratic variation we can set

$$[M]_t = [M, M]_t = \lim \sum_{i=0}^{n-1} (M_{t_{i+1}} - M_{t_i})^2,$$

where the limit is taken in probability, as the mesh size of the partition $0 = t_0 < \ldots < t_n = t$ goes to zero.
- An alternative way to categorise the quadratic variation is to consider a martingale associated with Brownian motion.
- Recall that

$$W_t^2 - t \text{ is a martingale.}$$

Thus W^2 minus the quadratic variation is a martingale.

Theorem

If M is a continuous martingale, then there exists a unique increasing continuous process $\langle M, M\rangle_t = \langle M\rangle_t$, zero at zero, such that $M_t^2 - \langle M, M\rangle_t$ is a continuous martingale.

Moreover this process is the quadratic variation as defined through the limit of the sum of squared increments as the mesh size goes to zero.

Proof: Show M^2 is a submartingale, apply the Doob–Meyer decomposition, check continuity.

For a general (not necessarily continuous) martingale X_t we have

$$[X, X]_t := \lim \sum (X_{t_{i+1}} - X_{t_i})^2 = \langle X^c, X^c \rangle_t + \sum_{u \leq t} (\Delta X_u)^2,$$

where X^c is the continuous martingale part of X and ΔX_u is the jump at time u.

Quadratic Co-variation

- If we have two martingales M, N that depend on each other we can define their quadratic co-variation $\langle M, N\rangle_t$.
- This can be defined in two ways – via sums

$$\langle M, N\rangle_t = \lim \sum (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}),$$

  where the limit is probability as the mesh of the partition goes to zero.
- Or via the fact that the quadratic co-variation is the unique process such that $M_t N_t - \langle M, N\rangle_t$ is a martingale.

Various useful results can be obtained using this concept.

- For a stopping time τ we have $\langle M^\tau, N\rangle_t = \langle M, N^\tau\rangle_t = \langle M, N\rangle_{t\wedge\tau}$,
- If M, N are independent, then $\langle M, N\rangle_t = 0$
- A continuous (local) martingale with 0 quadratic variation must be a constant.
- If M is a continuous (local) martingale then M and $\langle M\rangle$ are constant on the same intervals.
- $\mathbb{E}[M_t^2 - \langle M\rangle_t] = 0$, so if $\sup_t \mathbb{E}[\langle M\rangle_t] < \infty$, we know that M is uniformly integrable and hence converges.
- In general, M converges precisely for those paths ω where $\lim_{t\to\infty}\langle M\rangle_t < \infty$

In fact, the quadratic variation gives us an alternative definition for Brownian motion.

Theorem

For an $(\mathcal{F}_t)_{t\geq 0}$ adapted process M, with $M_0 = 0$, the following are equivalent

- *M is an $(\mathcal{F}_t)_{t\geq 0}$ – Brownian motion;*
- *M is a continuous (local) martingale and $\langle M \rangle_t = t$.*

The assumption of continuity is important. If $M_t = N_t - t$ is the compensated Poisson process of rate 1 then M_t is a martingale and $\langle M \rangle_t = t$.

Stochastic Integration

Stochastic integrals – intro

- If X is differentiable, we can define the integral against X by

$$\int K_s dX_s = \int K_s \frac{dX_s}{ds} ds.$$

- Suppose $X \notin C^1$ is an increasing process. We can then define the integral against X by using Riemann sums:

$$\int K_s dX_s = \lim \sum_i K_{u_i}(X_{t_{i+1}} - X_{t_i}), \quad \text{for some } u_i \in [t_i, t_{i+1}].$$

  Provided K is continuous, this will converge.
- If X is of finite variation, then it is the difference of two increasing processes, so the integral can be extended linearly.

Stochastic integrals – intro

- We want to integrate a process K_t against a martingale M_t

$$\int_0^t K_s \,\mathrm{d}\, M_s.$$

- The natural idea is to consider Riemann sums

$$\sum_i K_{u_i}(M_{t_{i+1}} - M_{t_i}), \quad \text{for some } u_i \in [t_i, t_{i+1}].$$

- However these sums do not usually converge pathwise because M_t has paths of infinite variation.
- We have to proceed more carefully...

We want the 'integral' to obey some properties

- the integral is linear in the integrand,
- if $Y_t = \int_0^t K_u \,\mathrm{d}\, X_u$ then $\int_0^t L_u \,\mathrm{d}\, Y_u = \int_0^t L_u K_u \,\mathrm{d}\, X_u$,
- given a 'simple' function $K_t = \sum_i K_i 1_{t<t_i}$, where K_i is $\mathcal{F}_{t_i}$-measurable, we have the Riemann sum

$$\int_0^t K_u \,\mathrm{d}\, X_u = \sum_{i=0}^{n-1} K_i (X_{t_{i+1}} - X_{t_i}) + K_n (X_t - X_{t_n}), \quad t_n \leq t < t_{n+1}$$

- integrals of *nice* converging integrands converge.

Furthermore, it would be nice if

- The integral with respect to a martingale is a (local) martingale.

- The idea of Itô was to take a particular choice of time point for the integrand in the approximating sum – the left end point.
- Let $I_n(t) = \sum_{i=0}^{n-1} K_{t_i}(M_{t_{i+1}} - M_{t_i}) + K_{t_n}(M_t - M_{t_n})$
- We then want to define $\int_0^t K_u \,\mathrm{d}\, M_u = \lim_{n\to\infty} I_n(t)$ for a suitable class of martingales M and processes K, with the limit in the right sense.

- Recalling some discrete martingale theory we have

$$\mathbb{E}(I_n(t)|\,\mathcal{F}_s) = I_n(s).$$

Thus the discrete version of the integral is a martingale.

- We can also compute the variance

$$\mathbb{E}(I_n(t)^2) = \mathbb{E}\left(\sum_{i=0}^{n-1} K_{t_i}^2(\langle M\rangle_{t_{i+1}} - \langle M\rangle_{t_i})\right)$$

- This is a discrete version of the Itô isometry and gives a natural way to define the integral.

Stochastic Integrals

- Let H^2 be the space of continuous L^2-bounded martingales

$$H^2 = \{M : \sup_{t \geq 0} \mathbb{E}[M_t^2] < \infty\}.$$

- This is a nice space (it's a Hilbert space with inner product $(M, N) = \mathbb{E}[\langle M, N\rangle_\infty])$
- For $M \in H^2$, let $L^2(M)$ be the (Hilbert) space of adapted processes K s.t.

$$||K||_M^2 = \mathbb{E}\left[\int_0^\infty K_s^2 \,\mathrm{d}\langle M\rangle_t\right] < \infty.$$

University of Oxford
Mathematical Institute

The key result for the construction of stochastic integrals is

Theorem (Itô)

Let $M \in H^2$ *and* $K \in L^2(M)$*. Then the integral* $\int_0^t K_s dM_s = \lim I_n(t)$ *is a martingale in* H^2*, the limit being taken in* H^2*.*

We also have the Itô isometry

$$\mathbb{E}\Big[\Big(\int_0^t K_s \,\mathrm{d}\, M_s\Big)^2\Big] = \mathbb{E}\Big[\int_0^t K_s^2 \,\mathrm{d}\langle M\rangle_s\Big].$$

The isometry is the key, as it means that a sequence of integrands converging in $L^2(M)$ gives a sequence of integrals converging in H^2.

UNIVERSITY OF OXFORD
Mathematical Institute

Theorem

The stochastic integral satisfies

- *it is a linear operator*
- $\int_0^t H_u \,\mathrm{d}\left(\int_0^u K_s dM_s\right) = \int_0^t H_u K_u \,\mathrm{d}\, M_u$ *for any locally bounded* K, H;
- $\int_0^{t\wedge\tau} K_u \,\mathrm{d}\, M_u = \int_0^t K_u 1_{[0,\tau]}(u) \,\mathrm{d}\, M_u = \int_0^t K_u \,\mathrm{d}\, M_u^\tau$, *for any stopping time* τ;
- *If* K^n, K, L *are locally bounded processes s.t.* $K_t^n \to K_t$ *a.s.* $\forall$ $t \le T$, $|K_t^n - K_t| \le L_t$ *for all* n, $t \le T$, *then*

$$\int_0^T K_u^n \,\mathrm{d}\, M_u \to \int_0^T K_u \,\mathrm{d}\, M_u, \ a.s.$$

Rules for stochastic calculus

The key rules for stochastic calculus are

- $\int_0^t K_u \,\mathrm{d}\, M_u$ is a martingale in H^2 if $K \in L^2(M), M \in H^2$.
- $\langle \int_0^\cdot K_u \,\mathrm{d}\, M_u \rangle_t = \int_0^t K_u^2 \,\mathrm{d} \langle M \rangle_u$
- Itô isometry

We can summarize the first two rules and write them in differential form

- If $\mathrm{d}\, X = K \,\mathrm{d}\, M$, then X is a martingale and $\mathrm{d}\langle X \rangle = K^2 \,\mathrm{d}\langle M \rangle$.
- $\mathrm{d}\, M \;\mathrm{d}\, t = 0$, $(\mathrm{d}\, M)^2 = d\langle M \rangle$
- If $\mathrm{d}\, X = KdM$ and $\mathrm{d}\, Y = H \,\mathrm{d}\, N$, then

$$\mathrm{d}\langle X, Y \rangle = KH \;\mathrm{d}\langle M, N \rangle.$$

Side Remark:

- The construction of the stochastic integral, as a limit of discrete approximations, immediately opens the door to numerical methods.
- Unlike in classical integration, higher order methods are often very tricky, as the *left* limits are used for the integrand.
- Using other endpoints (so using a different approximation) naturally leads to a different limiting process, which will not have the martingale property.

Local martingales

UNIVERSITY OF OXFORD
Mathematical Institute

If we allow an unbounded time set, then Brownian motion is not in H^2! We need a more general theory.

Definition

A process (M_t) is called a local martingale if there exists a sequence of stopping times $\tau_n \to \infty$ a.s. such that $M^{\tau_n} = (M_{t\wedge\tau_n})$ is a martingale for every n.

A martingale is a local martingale but not vice-versa. In fact a local martingale may possess strong integrability properties (e.g. uniformly integrability) and not be a true martingale.

Lemma

Let M be a local martingale. Suppose that $\{M_\tau\}$ is uniformly integrable, where τ is any stopping time. Then M is a (uniformly integrable) martingale.

One can show that this is equivalent to $\mathbb{E}[\sup_t |M_t|] < \infty$, which is guaranteed by $\sup_t \mathbb{E}[\langle M \rangle_t] < \infty$

A canonical example of a local martingale which is not a martingale is given by

$$M_t = \frac{1}{\sqrt{(W_t^1)^2 + (W_t^2)^2 + (W_t^3)^2}},$$

where W_t^1, W_t^2, W_t^3 are independent std BM.

UNIVERSITY OF OXFORD
Mathematical Institute

Definition

A continuous $(\mathcal{F}_t)$–adapted process (X_t) is a semimartingale if it can be written as $X_t = M_t + A_t$, where (M_t) is a continuous local martingale and (A_t), $A_0 = 0$, is a continuous adapted process of finite variation.

Note that the above decomposition is unique and that $\langle X \rangle_t = \langle M \rangle_t$.

The Doob–Meyer decomposition guarantees that every continuous submartingale is a semimartingale. (Discontinuous case also true, but we need to require A to be predictable, not continuous)

The general case

- This allows us to extend the previous construction by localisation procedure to (M_t) which is a local martingale and K which is in $L^2_{loc}(M)$.
- For an adapted process of finite variation A_t, the integral $\int_0^t K_u \, \mathrm{d} A_u$ can be defined in a more standard way (the Steltjes integral). This allows us to define stochastic integrals w.r.t. semimartingales.
- For a continuous semimartingale (X_t) and an adapted locally bounded process (K_t) we define the stochastic integral of K w.r.t. X by

$$\int_0^t K_u \, \mathrm{d} X_u = \int_0^t K_u \, \mathrm{d} M_u + \int_0^t K_u \, \mathrm{d} A_u.$$

UNIVERSITY OF OXFORD
Mathematical Institute

- The notation $Y_t = \int_0^t K_u \, \mathrm{d} X_u$ is often abbreviated to $\mathrm{d} Y_t = K_t \, \mathrm{d} X_t$ and is referred to as an SDE (*Stochastic Differential Equation*).
- The power of stochastic calculus comes from the fact that we can have rules to differentiate and integrate, as we do for the classical calculus.
- The basic operation is integration by parts.

Lemma

If X, Y are two continuous semimartingales then

$$X_t Y_t = X_0 Y_0 + \int_0^t X_u \, \mathrm{d} Y_u + \int_0^t Y_u \, \mathrm{d} X_u + \langle X, Y \rangle_t .$$

University of Oxford
Mathematical Institute

Theorem

Let $X^1, \dots, X^d$ be continuous semimartingales and $F \in C^2(\mathbb{R}^d)$. Then $F(\mathrm{X}_t) = F(X_t^1, \dots, X_t^d)$ is a continuous semimartingale and

$$F(\mathrm{X}_t) = F(\mathrm{X}_0) + \sum_{i=1}^{d} \int_0^t \frac{\partial F}{\partial x_i}(\mathrm{X}_u)\,\mathrm{d}\,X_u^i + \frac{1}{2}\sum_{i,j} \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(\mathrm{X}_u)\,\mathrm{d}\langle X^i, X^j\rangle_u.$$

In particular,

$$\begin{aligned} F(t, X_t) = & F(0, X_0) + \int_0^t \frac{\partial F}{\partial u}(u, X_u)\,\mathrm{d}\,u + \int_0^t \frac{\partial F}{\partial x}(u, X_u)\,\mathrm{d}\,X_u \\ & + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial x^2}(u, X_u)\,\mathrm{d}\langle X\rangle_u. \end{aligned}$$

Mathematical Institute

Theorem

For a continuous local martingale (M_t), and $\lambda \in \mathbb{R}$, the process

$$\mathcal{E}(M)_t = \exp\left\{\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\right\},$$

is a continuous local martingale. It is called the exponential martingale of M.

Proof.

Apply Itō's formula to $f(M_t, \langle M\rangle_t)$ with $f(x, y) = \exp(\lambda x + \lambda^2/2y)$. We obtain that

$$\mathcal{E}(M)_t = \int_0^t \mathcal{E}(M)_u \,\mathrm{d}\, M_u, \quad \text{i.e.} \quad \mathrm{d}\,\mathcal{E}(M)_t = \mathcal{E}(M)_t \,\mathrm{d}\, M_t.$$

□

A common example is obtained as follows.

Example

Consider $M_t = \int_0^t K_u \, \mathrm{d}\, W_u$ and $\lambda = 1$. We see that

$$\exp\left(\int_0^t K_u \, \mathrm{d}\, W_u - \frac{1}{2}\int_0^t K_u^2 \, \mathrm{d}\, u\right)$$

is a local martingale.

For $K_t \equiv \sigma$ we obtain a geometric Brownian motion.

Exercise

a) *Show directly from the definitions that the following processes are martingales:*

i. $M_t = e^{\sigma W_t - \sigma^2 t/2}$

ii. $N_t = W_t^3 - 3tW_t$

b) *By using the optional stopping theorem show that, if* $\tau_a = \inf\{t : |W_t| \geq a\}$, *then*

$$\mathbb{E}[\exp(-\lambda \tau_a)] = 1/\cosh(a\sqrt{2\lambda}).$$

If I have a martingale, can I write it as a stochastic integral against another martingale? (We just consider the case of Brownian motion.)

Theorem

Let (W_t) be a Brownian motion and $(\mathcal{F}_t)$ its natural filtration. Every $(\mathcal{F}_t)$–local martingale (M_t) may be written as

$$M_t = c + \int_0^t H_u \, \mathrm{d} W_u, \quad t \geq 0,$$

for some constant c and a predictable process (H_t) (locally in $L^2(W)$). In particular, any $(\mathcal{F}_t)$–local martingale is continuous.

- Applying this result to a random variable ξ, we can write

$$\xi = \mathbb{E}[\xi] + \int_0^T H_u dW_u$$

for some H.

- The theorem generalises to a multi-dimensional setting.
- Note however that it may no longer hold for a general filtration (without major changes).

Examples

We note that we have already seen some simple examples

$$W_t^2 - t = \int_0^t 2W_u dW_u.$$

and if

$$M_t = \exp(-\lambda W_t - \frac{1}{2}\lambda^2 t),$$

then

$$M_t = 1 - \int_0^t \lambda M_u dW_u.$$

In general however the representation theorem is an existence theorem.

Exercise

Find the process giving the representation for the random variables

- $W_{T/2}$
- W_T^4 *(Hint: you may find it helpful to consider* $X_t = E[W_T^4|\mathcal{F}_t]$*)*
- $\sin(W_T)$ *(Hint: Itô's formula on* $e^{\alpha}\sin(W_t)$ *might be good to calculate)*

Stochastic Differential Equations

- We will always assume that there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which our processes are defined.
- A one dimensional diffusion process can be described by a stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t,$$

where W is a standard n-dimensional Brownian motion,
- $b(t, x)$ is called the drift and $\sigma(t, x)$ is the volatility or diffusion coefficient and the initial state $X_0 = x_0$.

- We interpret this as an integral equation

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s.$$

- The coefficients $b(t, x), \sigma(t, x)$ are given functions and we assume that ($\mathbb{P}$-almost surely)

$$\int_0^t \sigma^2(s, X_s) + |b(s, X_s)|ds < \infty.$$

Multidimensional SDEs

UNIVERSITY OF OXFORD
Mathematical Institute

- A multi-dimensional diffusion process can be described by a set of stochastic differential equations

$$dX_t^i = b^i(t, X_t)dt + \sum_{j=1}^{n} \sigma_j^i(t, X_t)dW_t^j, i = 1, \dots, d$$

- We interpret this, for each $i = 1, \dots, d$, as

$$X_t^i = X_0^i + \int_0^t b^i(s, X_s)ds + \sum_{j=1}^{n} \int_0^t \sigma_j^i(s, X_s)dW_s^j.$$

Multidimensional SDEs

UNIVERSITY OF OXFORD
Mathematical Institute

$$dX_t^i = b^i(t, X_t)dt + \sum_{j=1}^{n} \sigma_j^i(t, X_t)dW_t^j, i = 1, \ldots, d$$

- W is a standard n-dimensional Brownian motion, $\{b^i(t,x)\}_{i=1}^d$ is called the drift vector, $\{\sigma_j^i(t,x)\}_{i,j}$ is the $d \times n$ volatility matrix, $a(t,x) = \sigma(t,x)\sigma(t,x)^T$ is the $d \times d$ diffusion matrix and $X_0^i = x_0^i$.
- As before $b^i(t,x), \sigma_j^i(t,x)$ are given functions, and we assume that if $a^{ij} = \sum_k \sigma_k^i \sigma_k^j$, then for all i, j

$$\int_0^t |a^{ij}(s, X_s)| + |b^i(s, X_s)|ds < \infty, \ \ \mathbb{P}-a.s.$$

There are different notions for existence and uniqueness of solutions for SDEs. We usually think of them as a pair of processes (X, W).

1. Strong solutions: Here we assume a probability space is given along with a Brownian motion W and we say X is a strong solution if there is an $(\mathcal{F}_t)$-adapted process X starting from x_0 satisfying the equation. Uniqueness means that the paths are the same with probability one.

2. Weak solutions: Here we are only given the starting point x_0 and the solution consists of finding the probability space with a Brownian motion W and the adapted process X that starts from x_0 and satisfies the equation. Here uniqueness means that the probability distribution is unique

Note that we could also start the process off from an initial probability distribution.

The key difference between strong and weak solutions is *information*. A weak solution can depend on a world with more information in it than a strong solution.

- The theory of stochastic differential equations allows for the x-coordinate of the coefficients to be a function of the whole past history of the process.
- That is, if we write

$$\sigma(t, X_{(\cdot)}) = \sigma(t, \{X_s; 0 \leq s \leq t\}),$$

$$b(t, X_{(\cdot)}) = b(t, \{X_s; 0 \leq s \leq t\}),$$

for the drift and diffusion coefficients which depend on the path, then we can consider equations

$$dX_t^i = b^i(t, X_{(\cdot)})dt + \sum_j \sigma_j^i(t, X_{(\cdot)})dW_t^j, i = 1, \ldots, d.$$

- We will just focus on the simple case where the coefficients are

$$b^i(t,x) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}, \qquad \sigma^i_j(t,x) : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$$

  for $i = 1, \ldots, d$ and $j = 1, \ldots, n$.
- Just as for ODEs, continuity assumptions on b and σ will be key to obtaining existence and uniqueness results.

Strong Solutions

UNIVERSITY OF OXFORD
Mathematical Institute

We assume

- Local Lipschitz: For each $N \in \mathbb{R}_+$ there is a K_N such that

$$|\sigma(t,x) - \sigma(t,x')| + |b(t,x) - b(t,x')| \leq K_N |x - x'|,$$

  for $|x|, |x'| \leq N$ and $0 \leq t \leq N$. (here $|.|$ denotes the usual vector or matrix norm.)

- Linear growth: To ensure the solution exists for all time: For each $T > 0$ there is a C_T such that for $0 \leq t \leq T$,

$$|\sigma(t,x)| + |b(t,x)| \leq C_T(1 + |x|).$$

Theorem

Under these conditions there is a unique strong solution to the SDE defined for all time.

- The process X satisfying the SDE is called a diffusion process and it has a strong solution under the local Lipschitz and linear growth conditions.
- Two key properties:
  1. The diffusion process has continuous sample paths almost surely.
  2. They are strong Markov processes.
  3. The solution is continuous with respect to all the inputs (in appropriate metrics)

UNIVERSITY OF OXFORD
Mathematical Institute

A weak solution ensures that there is a process with the required probability distribution. Here we consider the case where the matrix $a(x)$ and the vector $b(x)$ are independent of time.

Theorem

If a, b are measurable, the matrix a is continuous and $a(x)$ is strictly positive definite for each x and the growth condition for all i, j

$$|a^{ij}(x)| \leq K(1 + |x|^2), \quad |b^i(x)| \leq K(1 + |x|).$$

Then there is a unique weak solution to the SDE.

Note: A strong solution implies a weak solution

In the one dimensional case it is possible to obtain a much sharper result.

Example (Brownian motion with drift in $\mathbb{R}^d$)

Here b is a constant vector and we take $\sigma(t,x)$ to be the identity matrix. In vector form we have

$$dX = bdt + dW_t, \quad X_0 = x.$$

We can solve the SDE by integrating

$$X_t = x + bt + W_t.$$

Example (Exponential Brownian motion in $\mathbb{R}$)

Here $b(t,x) = \mu x$ and $\sigma(t,x) = \sigma x$, where μ and σ are constants. Then

$$dX_t = \mu X_t dt + \sigma X_t dW_t.$$

Given that $X_0 = x$ we have

$$X_t = x \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

Example (Ornstein–Uhlenbeck process)

This is a mean reverting linear diffusion with

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t.$$

In order to solve this we consider $Y_t = e^{\kappa t} X_t$ and use Ito

$$X_t = \theta + (x - \theta)e^{-\kappa t} + \sigma \int_0^t e^{-\kappa(t-s)} dW_s.$$

Example (Brownian bridge)

This is a diffusion conditioned to be at y at time 1

$$dX = \frac{y - X_t}{1 - t} dt + dW_t, \quad X_0 = x.$$

Solving this we have

$$X_t = yt + (1 - t)\left(x + \int_0^t \frac{dW_s}{1 - s}\right).$$

Mathematical Institute

Example (CIR model)

This is a standard interest rate model

$$dX = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t.$$

It can be transformed by $Y = \sqrt{X}$.

Example (Bessel process)

This is the radial part of an n-dimensional Brownian motion. In the case $n = 3$ we have

$$dX_t = \frac{1}{X_t}dt + dW_t.$$

- Just as in the classical case, there are may SDEs which don't have explicit solutions.
- The Euler(–Maruyama) scheme will converge, but higher order methods are a little tricky (but possible!)
- Path dependence can be included without too much trouble
- Error analysis depends on what sort of error you want

Connections with PDEs

Transition densities

University of Oxford
Mathematical Institute

- The transition density for a diffusion process is denoted by $p(t, x; T, y)$.
- This is the density of the probability distribution function for the stochastic process in that for any set $A \subset \mathbb{R}^d$

$$P(X_T \in A | X_t = x) = \int_A p(t, x; T, y) dy.$$

- Thus it is the density of the probability for the diffusion to go from x at time t to y at time T.

- Once we have this quantity we can determine many properties of the diffusion process.
- In particular we can find for any integrable function $h : \mathbb{R}^d \to \mathbb{R}$

$$\mathbb{E}[h(X_T) | \mathcal{F}_t] = \int_{\mathbb{R}^d} h(y) p(t, x; T, y) dy.$$

Example (Examples in one dimension)

Mathematical Institute

1. Brownian motion: For standard Brownian motion $dX_t = dW_t$ and as W_t has a $N(0,t)$ distribution we have

$$X_T = x + W_T - W_t$$

and thus

$$p(t,x;T,y) = \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(x-y)^2}{2(T-t)}\right).$$

2. Brownian motion with drift b: This is a simple change of variable for Brownian motion as X_T has a $N(x+b(T-t), T-t)$ distribution
3. Exponential Brownian motion:
This has a log normal distribution.
4. The OU process is also Gaussian.

How do we find equations for the transition density in general?

We discuss two approaches

1. Using Ito and the Feynman-Kac formula relating SDEs and PDEs.
2. Approximation via discrete random walk.

We give the general result which can be used to solve the Black-Scholes equation.

Theorem

Let

$$Lu := \frac{\partial u}{\partial t} + b(t,x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 u}{\partial x^2} + r(t,x)u.$$

If $u(t,x)$ *satisfies the PDE* $Lu = 0$ *subject to a terminal condition* $u(T,x) = \Psi(x)$, *then we can write*

$$u(t,x) = \mathbb{E}\left[\exp\left(\int_t^T r(s,X_s)ds\right)\Psi(X_T)\middle| X_t = x\right],$$

where X *satisfies the SDE*

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t.$$

Proof.

Let $M_s = U_s R_s$ where

$$U_s = u(s, X_s), \qquad R_s = e^{\int_t^s r(v,X_v)dv},$$

for $t \le s \le T$. Thus we have $M_t = u(t, X_t)$. If we can show that M_s is a martingale, then

$$u(t, X_t) = M_t = \mathbb{E}_t\, M_T = \mathbb{E}_t \left(\exp\left(\int_t^T r(s, X_s)ds \right) \Psi(X_T) \right),$$

as required.

ctd.

Proof ctd.

We now use Itô to show the result: first

$$dU_s = u_t(s, X_s)ds + u_x(s, X_s)dX_s + \frac{1}{2}u_{xx}(s, X_s)d\langle X\rangle_s,$$

and

$$dR_s = r(s, X_s)R_s ds.$$

Using the Itô Product rule we have

$$\begin{aligned} dM_s &= R_s dU_s + U_s dR_s + d[U_s, R_s] \\ &= R_s(Lu)ds + R_s u_x(s, X_s)\sigma(s, X_s)dW_s \end{aligned}$$

ctd.

Proof ctd.

As $u(s, X_s)$ satisfies the PDE, we have no drift term and

$$dM_s = R_s u_x(s, X_s)\sigma(s, X_s)dW_s.$$

Thus it is a (local) martingale and we have the result under reasonable conditions. □

Corollary

In the case $r \equiv 0$, if $u(t, x)$ satisfies the PDE $Lu = 0$ subject to a terminal condition $u(T, x) = \Psi(x)$, then

$$u(x, t) = \mathbb{E}(\Psi(X_T)|X_t = x),$$

where X satisfies the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t.$$

UNIVERSITY OF OXFORD
Mathematical Institute

This last Feynman–Kac formula can be used to study the transition density of the SDE. For the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t,$$

we have that, if $Lu = 0$ (where we set $r = 0$) with $u(T, x) = h(x)$, then

$$u(t, x) = \mathbb{E}(h(X_T)|X_t = x) = \int h(z)p(t, x; T, z)dz$$

Thus, taking h to be a delta function at y we have

$$u(t, x) = \mathbb{E}(\delta_y(X_T)|X_t = x) = p(t, x; T, y).$$

Lemma (Kolmogorov Backward Equation)

The transition density for the SDE satisfies the PDE

$$p_t(t,x;T,y) + b(t,x)p_x(t,x;T,y) + \frac{1}{2}\sigma^2(t,x)p_{xx}(t,x;T,y) = 0,$$

with $p(T,x;T,y) = \delta_y(x)$.

The Kolmogorov Backward equation

- This is the backward equation because the variables are t, x.
- If the SDE is autonomous in that b and σ do not depend on time, then, writing $\tau = T - t$, we have

$$-p_\tau(\tau, x, y) + b(x)p_x(\tau, x, y) + \frac{1}{2}\sigma^2(x)p_{xx}(\tau, x, y) = 0.$$

- We can write this as

$$p_\tau(\tau, x, y) = \mathcal{A}p(\tau, x, y),$$

where the quantity

$$\mathcal{A} = b(t, x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2}{\partial x^2}$$

is called the infinitesimal generator of the diffusion.

We can recover the coefficients of the SDE from the transition density p

- The infinitesimal drift is

$$\begin{aligned} b(t,x) &= \lim_{\Delta\to 0} E(X_{t+\Delta} - x | X_t = x) \\ &= \lim_{\Delta\to 0} \int_{-\infty}^{\infty} (y-x) p(t,x;t+\Delta,y) dy, \end{aligned}$$

  The volatility is

$$\begin{aligned} \sigma^2(t,x) &= \lim_{\Delta\to 0} E((X_{t+\Delta} - x)^2 | X_t = x) \\ &= \lim_{\Delta\to 0} \int_{-\infty}^{\infty} (y-x)^2 p(t,x;t+\Delta,y) dy. \end{aligned}$$

- In higher dimensions we can recover the diffusion matrix as

$$a^{ij}(t,x) = \lim_{\Delta\to 0} E((X^i_{t+\Delta} - x^i)(X^j_{t+\Delta} - x^j) | X_t = x).$$

Alternative approach

- Consider Brownian motion. Suppose we are at (T, y).
- At the previous step we must have been at one of $(T - \delta T, y - \delta y)$ or $(T - \delta T, y + \delta y)$.
- Omitting the dependence on (t, x) as they will not change, we have

$$p(T, y) = \frac{1}{2}p(T - \delta T, y - \delta y) + \frac{1}{2}p(T - \delta T, y + \delta y).$$

Alternative approach

- Now use Taylor's theorem,

$$
\begin{aligned}
p + p_T \delta T + \ldots = p(T, y) \\
&= \frac{1}{2}\left(p - \delta y\, p_y + \frac{1}{2}\delta y^2 p_{yy} + \ldots\right) \\
&\quad + \frac{1}{2}\left(p + \delta y\, p_y + \frac{1}{2}\delta y^2 p_{yy} + \ldots\right) \\
&= p + \frac{1}{2}\delta y^2 p_{yy} + \ldots
\end{aligned}
$$

where p and all the derivatives are evaluated at $(T - \delta T, y)$.

Now substitute $\delta y = \sqrt{\delta T}$ and take the limit to obtain

$$\frac{\partial p}{\partial T} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}.$$

- This is the forward Kolmogorov (Fokker-Planck, Feynman-Kac, Fisher) equation for Brownian motion.
- It shows how the probability density of future states evolves, starting from (t, x).
- As the Brownian motion started from x has a normal distribution mean x variance $T - t$, we recover the transition density

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

- At $T = t$ this is equal to $\delta_x(y)$ (the delta function at x).

- A similar analysis shows that the dependence on (t, x) is

$$-\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$$

- This is the backward equation again and, thanks to symmetries of the Brownian motion, tells us the probability that we *were* at (t, x) given that we are now at (T, y) just as given by the Feynman-Kac formula.

- An alternative derivation of the general Kolmogorov forward equation is to use the adjoint equation:
- By Itô we have for any suitable function $v(t,x)$ that

$$
\begin{aligned}
& v(T, X_T) \\
&= v(t, X_t) + \int_t^T \left(v_s(s, X_s) + b(s, X_s) v_x(s, X_s) + \frac{1}{2}\sigma^2(s, X_s) v_{xx}(s, X_s) \right) ds \\
&\qquad + \int_t^T \sigma(s, X_s) v_x(s, X_s) dW_s.
\end{aligned}
$$

- Taking expectations conditional on $X_t = x$ we have

$$
\begin{aligned}
& \mathbb{E}_t(v(T, X_T)) - v(t, x) \\
&= \int_t^T \mathbb{E}_t \left(v_s(s, X_s) + b(s, X_s) v_x(s, X_s) + \frac{1}{2}\sigma^2(s, X_s) v_{xx}(s, X_s) \right) ds.
\end{aligned}
$$

- The function v is a general test function.
- Let $v(t,x) = 0$ and $v(s,x) \to 0$ as $s \to T$, uniformly in x.
- Then

$$0 = \int_t^T \mathbb{E}_t \Big(v_s(s,X_s) + b(s,X_s)v_x(s,X_s) \\ + \frac{1}{2}\sigma^2(s,X_s)v_{xx}(s,X_s)\Big)\, ds.$$

- Now write in terms of the transition density and integrate by parts twice

- This shows that the density $p(t, x; T, y)$ satisfies

$$\frac{\partial p}{\partial T} + \frac{\partial (bp)}{\partial y} - \frac{1}{2}\frac{\partial^2(\sigma^2 p)}{\partial y^2} = 0.$$

- This can be written as

$$\frac{\partial p}{\partial T} = \mathcal{A}^* p,$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$.

Example (Brownian motion)

The KBE and KFE are the backward and forward heat equations;

$$-\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}, \qquad \frac{\partial p}{\partial T} = \frac{1}{2}\frac{\partial^2 p}{\partial y^2}.$$

Example (Exponential BM)

We have the KBE

$$\frac{\partial p}{\partial t} + rx\frac{\partial p}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 p}{\partial x^2} = 0,$$

and the KFE

$$\frac{\partial p}{\partial T} + r\frac{\partial}{\partial y}(yp) - \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial y^2}(y^2 p) = 0.$$

Stationary distributions

UNIVERSITY OF OXFORD
Mathematical Institute

- Some diffusions have an equilibrium or stationary distribution.
- In the limit as $T \to \infty$ we expect the probability distribution of the diffusion to converge to this stationary distribution in the sense that

$$\lim_{T\to\infty} p(t,x;T,y) = \psi(y),$$

where ψ is a probability density function.
- By considering the KFE we see that the stationary distribution should be given by ψ such that $\mathcal{A}^*\psi = 0$.

Example (O–U Process)

The process

$$dX_t = \kappa(\theta - X_t)dt + \sigma dW_t,$$

has a stationary distribution which is normal mean θ, variance $\sigma^2/(2\kappa)$.

Example (Langevin dynamics)

For a (smooth) scalar potential V, the (multidimensional) process

$$dX_t = -\nabla V(X_t)dt + \sigma dW_t$$

(where W is a vector of independent Brownian motions) usually(!) has a stationary distribution with density

$$f(x) \propto \exp(-2V(x)/\sigma^2).$$

Exit times

First exit times

- The first exit time is the random time at which a stochastic process first reaches a given boundary.
- We will find the probability of exiting at one boundary rather than another and the mean time for the process to do this.

- We consider the SDE for the diffusion X,

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t.$$

- Consider two boundaries given by differentiable functions $U(t), L(t)$ with $L(t) \leq U(t)$ for $0 \leq t \leq T$.
- What is the probability that X remains within the boundaries for $0 \leq t \leq T$?

- We can compute this by solving the backward Kolmogorov equation (with (T, y) fixed)

$$-p_t = \mathcal{A}p = b(t,x)p_x + \frac{1}{2}\sigma^2(x,t)p_{xx},$$

with boundary conditions

$$\begin{cases} p(t, L(t); T, y) = p(t, U(t); T, y) = 0 & \text{for } t \leq T \\ p(T, x; T, y) = \delta_y(x) & \text{for } L(T) < x < U(T) \end{cases}$$

- Then

$$P(X_s \in [L, U], t \leq s \leq T | X_t = x) = \int_{L(T)}^{U(T)} p(t, x; T, y) dy$$

Exercise

Consider a model in which the probability of an event occuring at a given time is modelled as a diffusion process satisyfing the SDE

$$dX_t = (b(1-X_t)-aX_t))dt+\sqrt{X_t(1-X_t)}dW_t; \qquad X_0 = x_0 \in [0,1]$$

where $a > 0$ and $b > 0$ are constants.

1. *Explain why this should be a diffusion on $[0,1]$*
2. *Find the forward and backward Kolmogorov equations for its transition density*
3. *(Fiddly) Show that if $a, b \geq 1/2$, then the stationary distribution is a beta distribution (note that $\int_0^1 x^{-2a}(1-x)^{-2b}dx = \infty$ if either of $a, b \geq 1/2$).*

Mathematical Institute

Hitting probabilities

Autonomous SDEs

- We now focus on time homogeneous or autonomous SDEs

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t.$$

- We can change the diffusion to its 'natural scale' using a scale function. That is we look for a function s such that $Y_t = s(X_t)$ is a martingale.

- This is an application of Itô:

$$dY_t = s'(x)dX + \frac{1}{2}s''(x)d\langle X\rangle = s'\sigma dW_t + (\frac{1}{2}s''\sigma^2 + bs')dt$$

- Thus if $bs' + \frac{1}{2}\sigma^2 s'' = 0$, then Y is a martingale.
- Solving, we have for any c,

$$s(x) = \int_c^x \exp\left(-2\int_c^y \frac{b(z)}{\sigma^2(z)}dz\right)dy.$$

(changing c just changes s to $\alpha s + \beta$ so Y is still a martingale).

Example

1. For Brownian motion $dX = dW$ we take $s(x) = x$, as W is already a martingale.
2. Brownian motion with drift: $b(x) = b, \sigma(x) = \sigma$, we can compute with $c = 0$,

$$s(x) = \frac{\sigma^2}{2b}\left(1 - \exp\left(-\frac{2bx}{\sigma^2}\right)\right).$$

3. Exponential BM: If $2\mu \neq \sigma^2$, then with $c = 1$,

$$s(x) = \frac{x^{1-2\mu/\sigma^2} - 1}{1 - 2\mu/\sigma^2},$$

otherwise $s(x) = \log(x)$.

Hitting probabilities

UNIVERSITY OF OXFORD
Mathematical Institute

- We find $H(x)$, the probability of X hitting the upper boundary U before the lower L started from $L \leq x \leq U$.
- We use the martingale Y as this probability is the same as the probability of Y hitting $s(U)$ before $s(L)$ started from $y = s(x)$.
- Let $\tau = \inf\{s : Y_s \notin (s(L), s(U)\}$. As Y is a martingale we can apply the optional stopping theorem

$$y = \mathbb{E}Y_\tau = s(U)H(x) + s(L)(1 - H(x)).$$

- Thus

$$H(x) = \frac{s(x) - s(L)}{s(U) - s(L)}.$$

Example (CIR model)

Suppose

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t.$$

- The scale function is given by

$$s(x) = \int_c^x \left(\frac{c}{y}\right)^{2\kappa\theta/\sigma^2} \exp\left(\frac{2\kappa(y-c)}{\sigma^2}\right)dy.$$

- From this we can show that $s(x) - s(0) = \infty$ if $2\kappa\theta > \sigma^2$.
- Thus, under this condition, X never hits 0.

Expected exit times

University of Oxford Mathematical Institute

- We now compute an ODE to find the mean exit time from a region.
- We use martingale methods again. Let $\tau = \inf\{t : X_t \notin (L, U)\}$.
- Suppose that we can find a function f such that $Y_t = f(X_t) + t$ is a martingale and $f(L) = f(U) = 0$.
- Then applying the optional stopping theorem we have (as $f(X_\tau) = 0$)

$$f(x) = E(f(X_\tau) + \tau | X_0 = x) = E(\tau | X_0 = x).$$

- Thus this function f will determine the expected exit time from the interval for all starting points x.

- To find f we just use Itô to get

$$\begin{aligned}dY_t &= dt + f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t \\ &= \left(1 + b(X_t)f'(X_t) + \frac{1}{2}\sigma^2(X_t)f''(X_t)\right)dt \\ &\quad + f'(X_t)\sigma(X_t)dW_t.\end{aligned}$$

- In order for Y to be a martingale we must set the dt term to be 0, so we must have f satisfying

$$1 + b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) = 0,$$

with $f(L) = f(U) = 0$.

Mathematical Institute

Example (Brownian motion)

Just solve $f''(x) = -2$ with the boundary conditions to get

$$f(x) = (x - L)(U - x).$$

Exercise

Consider the process Y satisfying the SDE

$$dY_t = \frac{\alpha}{Y_t} dt + \sigma dW_t, \qquad Y_0 = y > 0.$$

a) *Describe the qualitative difference in behaviour when $\alpha > 0$ or $\alpha < 0$*

b) *Let $0 < a < y < b$ and $T_{a,b} = \inf\{t : Y_t \notin (a, b)\}$. Find the scale function s for Y and hence calculate $P(Y_{T_{a.b}} = a | Y_0 = y)$.*

c) *Show that $f(y) = \mathbb{E}(Y_{a,b} | Y_0 = y)$ satisfies the boundary value problem*

$$0 = 1 + \alpha f'(y)/y + \sigma^2 f''(y)/2, \qquad f(a) = f(b) = 0.$$

Exercise

d) *Take $a \to 0$, $b \to \infty$, to show that this process will hit 0 with probability 1 if $\alpha < 0$, and find the mean time to do so.*

e) *Describe qualitatively the behaviour of the process $X_t = 1/Y_t$.*

Exercise

Let W be a Brownian motion and X be the solution to the SDE

$$dX_t = h(W_t)X_t dW_t, \qquad X_0 = 1$$

where h is a bounded function.

a) *Find the quadratic variation of X_t, and so argue why X is a martingale.*

b) *Find the dynamics of $\log(X_t)$, and so explain why $X_t > 0$ a.s. for all t.*

c) *Explain why we can define a new probability measure $\mathbb{Q}$ by $\mathbb{E}_{\mathbb{Q}}[\xi] = \mathbb{E}[\xi X_T]$, for each t*

Exercise

d) *Show that* $\mathbb{E}_{\mathbb{Q}}[W_\tau - \int_0^\tau h(W_t)dt] = 0$ *for all bounded stopping times* $\tau \leq T$.

e) *Conclude that* $\tilde{W}_t = W_t - \int_0^t h(W_s)ds$ *is a Brownian motion under the measure* $\mathbb{Q}$, *and so the SDE*

$$dY_t = h(Y_t)dt + dW_t$$

admits a weak solution on the time horizon $[0, T]$

This result is known as 'Girsanov's theorem'.

Note that in part (e), we have no assumption of continuity on h.