Stochastic Calculus, Summer 2022, May 25, Lecture 1 Construction of Brownian motion

Reading for this lecture (for references see the end of the lecture):

- [1] pp. 72-108

Scaled Random Walks. Our main object of study in this course will be *stochastic processes*. Loosely speaking, stochastic process is a collection of random variables indexed by "time" variable t: $X(t)$. Variable t could take values in a discreet set, for instance in a set of positive integers or it could take values in a continuous set, for instance in $[0, \infty)$. An example of stochastic process is a stock price that changes in time.

Our main objective for today is to build a very important stochastic process called **Brownian Motion**. Brownian Motion is a continuous-time stochastic process usually indexed by $[0, \infty)$. One dimensional Brownian Motion satisfies the following properties:

Independence of increments.: If $t_0 < t_1 < \cdots < t_n$ then random variables $B(t_0), B(t_1) - B(t_0), \ldots, B(t_n) - B(t_{n-1})$ are independent.

Normal distribution of increments.: If $s, t \geq 0$ then

$$\mathbb{P}\left[B(t+s) - B(s) \in A)\right] = \int\limits_A \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx. \tag{1}$$

Continuity of trajectories.: With probability one, $B_0 = 0$ and $t \to B_t$ is continuous.

It is not obvious from the above definition that Brownian Motion exists. We show that it does exist by essentially constructing Brownian Motion.

We start with a simpler stochastic process that is called **Symmetric Random Walk** (SRW). To construct a SRW we repeatedly toss a fair coin: $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$. We define the successive outcomes of the tosses by $w_1, w_2, w_3 \cdots \in \{H, T\}$ and define $w = w_1 w_2 w_3 \ldots$, in other words, w is the infinite sequence of tosses and w_n is the outcome of the n^{th} toss. Next, let X_j be a random variable defined as

$$X_j = \begin{cases} 1 & \text{if } w_j = \text{H}, \\ -1 & \text{if } w_j = \text{T}. \end{cases}$$

Define discrete-time stochastic process $M(n)$ in the following way:

$$M(0) = 0, M(n) = \sum_{j=1}^{n} X_j.$$

The process $M(n)$ is called a symmetric random walk. With each step it either steps up one unit or down one unit, and each of the two possibilities is equally likely.

Exercise 1. For positive integers k, n what is the probability $\mathbb{P}(M(n) = k)$?

[1]this version May 25, 2022

Increments of the SRW. A random walk has independent increments. This means that if we choose nonnegative integers $0 = n_0 < n_1 < \cdots < n_k$, then random variables

$$M(n_1) = M(n_1) - M(n_0), M(n_2) - M(n_1), \ldots, M(n_k) - M(n_{k-1})$$

are independent. Each of these random variables

$$M(n_{i+1}) - M(n_i) = \sum_{j=n_i+1}^{n_{i+1}} X_j$$

is called an increment of the random walk. It is the change in the position of the random walk between times n_i and n_{i+1}. Increments over non-overlapping time intervals are independent because they depend on different coin tosses. Moreover,

$$\mathbb{E}(M(n_{i+1}) - M(n_i)) = \sum_{j=n_i+1}^{n_{i+1}} \mathbb{E}X_j = 0$$

and

$$\mathrm{Var}(M(n_{i+1}) - M(n_i)) = \sum_{j=n_i+1}^{n_{i+1}} \mathrm{Var}X_j = n_{i+1} - n_i \text{ since } \mathrm{Var}X_j = 1.$$

Martingale property for the SRW. In order to define martingale we need the notion of σ-algebra. There is a very important, nontechnical reason to include σ-algebras in the study of stochastic processes, and that is to keep track of information. The temporal feature of a stochastic process suggests a flow of time, in which, at every moment $t \geq 0$, we can talk about a past, present and future and can ask how much an observer of the process knows about it at present, as compared to how much he knew at some point in the past or will know at some point in the future.

Definition 1. *A filtration on $(\Omega, \mathcal{F})$ is a family $\{\mathcal{F}_t\}_{t\geq 0}$ of σ-algebras $\mathcal{F}_t \subset \mathcal{F}$ such that*

$$0 \leq s \leq t \Longrightarrow \mathcal{F}_s \subset \mathcal{F}_t, \textit{ i.e., } \{\mathcal{F}_t\}_{t\geq 0} \textit{ is increasing.}$$

Thus, informally, one can think about the σ-algebra $\mathcal{F}_t$ in the above definition as a "knowledge" available at time t. The fact that $\{\mathcal{F}_t\}_{t\geq 0}$ is increasing just reflects the fact that as time goes on the amount of available information increases.

Definition 2. *A stochastic process $\{Z_t\}_{t\geq 0}$ is called a martingale with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if*

- **a):** *Z_t is $\mathcal{F}_t$-measurable for all t*
- **b):** *$\mathbb{E}|Z_t| < \infty$ for all t*
- **c):** *$\mathbb{E}(Z_t|\mathcal{F}_s) = Z_s$ for all $s \leq t$.*

The most important property in the above definition is property *c*). Simply put, it states that "the best" guess for X_t given information available at moment s is X_s, and thus process X_t has neither upward nor downward drift.

Now let's check that the SRW is a martingale. First, we need to specify a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Let $\mathcal{F}_n$ be the a σ-algebra(read information) generated by the set of random variables $X_1, X_2, \cdots X_n$.

a): $M(n) = \sum_{j=1}^{n} X_j$ is $\mathcal{F}_n$-measurable since it depends only on $X_j, j \leq n$, i.e. on the information available at time n.

b): From the fact that $|M(t)| \leq t$

$$(\mathbb{E}|M(t)|)^2 \leq t < \infty.$$

c): using the fact that $\mathbb{E}(M(s)|\mathcal{F}_s) = M(s)$ and the fact that random variable $M(t) - M(s)$ is independent of σ-algebra $\mathcal{F}_s$ we can write:

$$\begin{aligned}\mathbb{E}(M(t)|\mathcal{F}_s) &= \mathbb{E}(M(t) - M(s) + M(s)|\mathcal{F}_s) = \mathbb{E}(M(t) - M(s)|\mathcal{F}_s) + \mathbb{E}(M(s)|\mathcal{F}_s) \\ &= \mathbb{E}(M(t) - M(s)) + M(s) = 0 + M(s) = M(s). \qquad (2)\end{aligned}$$

Quadratic variation of the SRW. Quadratic variation of a discrete stochastic process up to time t is defined as

$$\langle M, M \rangle_t = \sum_{j=1}^{t} (M(j) - M(j-1))^2. \qquad (3)$$

The quadratic variation up to time t along a path is computed by taking all the one-step increments $M(j) - M(j-1)$ along that path, squaring these increments and then summing them. Clearly, for the symmetric random walk increments can take values ± 1 and thus

$$\langle M, M \rangle_t = t. \qquad (4)$$

Note that this is computed path by path and the fact that quadratic variation is the same along any path is a special feature of the symmetric random walk we consider. The consequence of the fact that quadratic variation is computed path by path is that for a general stochastic process quadratic variation is a random quantity, depending on the trajectory.

Scaled SRW. To approximate a Brownian motion we speed up time and scale down the step size of a SRW. More precisely, we fix a positive integer n and define the scaled SRW at rational points $\frac{k}{n}$ as

$$B^{(n)}\left(\frac{k}{n}\right) = \frac{1}{\sqrt{n}} M(k). \qquad (5)$$

At all other points we define $B^{(n)}(t)$ be linear interpolation between its values at the nearest points of the form $\frac{k}{n}$.

The following properties of the scaled SRW could be easily proved using the corresponding properties of the SRW and we leave their proof as an exercise:

a): independence of increments, i.e., for all rational numbers $0 = t_0 < t_1 < \cdots < t_n$ of the form $\frac{k}{n}$ random variables

$$B^{(n)}(t_1) - B^{(n)}(t_0), B^{(n)}(t_2) - B^{(n)}(t_1), \ldots, B^{(n)}(t_n) - B^{(n)}(t_{n-1}) \qquad (6)$$

are independent(because they depend on different coin tosses).

b): $\mathbb{E}(B^{(n)}(t) - B^{(n)}(s)) = 0, \mathrm{Var}(B^{(n)}(t) - B^{(n)}(s)) = t - s.$

c): $\mathbb{E}(B^{(n)}(t)|\mathcal{F}_s) = B^{(n)}(s).$

d): quadratic variation of the scaled SRW:

$$\langle B^{(n)} B^{(n)} \rangle_t = t.$$

Limiting distribution of the scaled SRW. Now the idea is to let $n \to \infty$ in (5) and in the limit obtain Brownian motion. In the previous section we studied a single path(trajectory) of the scaled SRM. In other words, we have fixed a sequence of coin tosses $w = w_1 w_2 \dots$ and drawn the path of the resulting process as time t varies. Another way to think about the scaled symmetric random walk is to fix time t and consider the set of all possible paths evaluated at that time t, i.e., consider the distribution of random variable $B^{(n)}(t)$. By definition $B^{(n)}(t) = \frac{1}{\sqrt{n}} M_{nt}$ or if $t = \frac{k}{n}$ then $B^{(n)}(t) = \frac{1}{\sqrt{n}} M_k$. Since $M_k = \sum_{j=1}^{k} X_j$ has a binomial distribution with parameters k and $1/2$ one can easily calculate the distribution of $B^{(n)}(t)$. In particular, one can draw the histogram of $B^{(n)}(t)$.

We know that random variable $B^{(n)}(t)$ has mean 0 and variance t. If we draw on top of the histogram of $B^{(n)}(t)$ the graph of normal density with mean zero and variance t we will see that the distribution of $B^{(n)}(t)$ is nearly normal. Actually, the Central Limit Theorem asserts that $B^{(n)}(t)$ converges in distribution to the normal random variable $N(0, t)$. But let us not use this theorem and prove the convergence of $B^{(n)}(t)$ from scratch. The whole point of this exercise is learning a very powerful mathematical tool known as "characteristic functions".

Theorem 3. *For any fixed $t \geq 0$, the distribution of the scaled random walk $B^{(n)}(t)$ evaluated at time t converges to the normal distribution with mean zero and variance t.*

Proof. To prove this theorem we use a tool from probability theory - characteristic functions.

Definition 4. *Let X be a random variable with distribution $\mathbb{P}$. Characteristic function of X is defined as*

$$\varphi_X(u) = \mathbb{E} e^{iuX} = \int e^{iux} dP(x) \tag{7}$$

for every $u \in \mathbb{R}$.

Example.
1) $X = Be(p)$. Then

$$\varphi_X(u) = p(e^{iu} - 1) + 1. \tag{8}$$

2) $X = Bi(n, p)$. Then $\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$ and

$$\varphi_X(u) = (p(e^{iu} - 1) + 1)^n. \tag{9}$$

3) $X = \text{Poisson}(\lambda)$, then $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ and

$$\varphi_X(u) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{iuk} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{iu})^k}{k!} = e^{\lambda(e^{it} - 1)}. \tag{10}$$

4) X is a Gaussian random variable $N(m, \sigma^2)$. Then

$$\varphi_X(u) = e^{ium - u^2 \sigma^2 / 2}. \tag{11}$$

The importance of the characteristic functions consists could be summarized in the following two theorems:

1. If two random variables have the same characteristic function then they have the same distribution.(of course, if two random variables have the same distributions then they have the same characteristic functions).

2. **Paul Lévy Theorem.** Let X_n be a sequence of random variables on $\mathbb{R}$. Then

a) If X_n converges to X is distribution then $\varphi_{X_n}(u) \to \varphi_X(u)$ for every u and in fact this convergence is uniform.

b) If $\varphi_{X_n}(u) \to \varphi_X(u)$ then X_n converges to X is distribution.

c) If $\varphi_{X_n}(u)$ converges to some function $\varphi(u)$ and $\varphi(u)$ is continuous at 0 then there exists a unique random variable X such that φ is the characteristic function of X and X_n converges to X in distribution.

We use part (b) of the theorem to prove that $B^{(n)}(t)$ converges to normal distribution with expectation zero and variance t.

$$\varphi_{B^{(n)}}(u) = \mathbb{E}e^{iuB^{(n)}(t)} = \mathbb{E}e^{iu\frac{1}{\sqrt{n}}M_{nt}} = \mathbb{E}e^{iu\frac{1}{\sqrt{n}}\sum_{j=1}^{nt} X_j} = \left(\mathbb{E}e^{iu\frac{1}{\sqrt{n}}X_j}\right)^{nt}$$
$$= \left(\frac{1}{2}e^{iu/\sqrt{n}} + \frac{1}{2}e^{-iu/\sqrt{n}}\right)^{nt}. \tag{12}$$

We need to show that as $n \to \infty$

$$\varphi_n(u) = \left(\frac{1}{2}e^{iu/\sqrt{n}} + \frac{1}{2}e^{-iu/\sqrt{n}}\right)^{nt}$$

converges to the characteristic function of the normal random variable with mean 0 and variance t, i.e., to $e^{-u^2t/2}$. Expanding in Taylor series

$$\frac{1}{2}e^{iu/\sqrt{n}} + \frac{1}{2}e^{-iu/\sqrt{n}} = 1 - \frac{u^2}{2n} + O(\frac{1}{n^{3/2}}) \tag{13}$$

and thus

$$\varphi_n(u) = \left(1 - \frac{u^2}{2n} + O(\frac{1}{n^{3/2}})\right)^{nt} \to e^{-\frac{u^2}{2n}\cdot nt} = e^{-u^2t/2}.$$

Thus we have proved that $B^{(n)}(t)$ converges in distribution to a normal random variable with expectation zero and variance t.

Remark 5. Truly speaking, we have just demonstrated that for a fixed moment t distribution of $B^{(n)}(t)$ converges to $N(0,t)$ but we did not prove the convergence of the whole stochastic process to Brownian motion. To do this one has to prove the convergence of the finite dimensional distributions and use Prokhorov's theorem.

□

References

[1] Steven Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*

[2] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*

[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

[4] Bernt Oksendal, Stochastic Differential Equations

Stochastic Calculus, Summer 2022,
Lecture 2
Properties of the Brownian motion

Reading for this lecture (for references see the end of the lecture):

- [1] §3.4, 3.4

Let us summarize what we did during the last lecture. We proved that the symmetric random walk, properly scaled, converges to some stochastic process that has the following properties:

Independence of increments: For every sequence $t_0, t_1, \ldots, t_n$ such that $t_0 < t_1 < \cdots < t_n$ random variables $B(t_0), B(t_1) - B(t_0), \ldots, B(t_n) - B(t_{n-1})$ are independent

Normally distributed increments: $\forall s, t > 0$ random variable $B(s+t) - B(s)$ distributed as a normal random variable with expectation zero and variance t, i.e. for any set $A \subset \mathbb{R}$

$$\mathbb{P}(B(s+t) - B(s) \in A) = \frac{1}{\sqrt{2\pi t}} \int\limits_A e^{-\frac{x^2}{2t}} dx.$$

Continuity of trajectories: with probability one trajectories of Brownian motion are continuous, in other words $B(t)$ as a function of time t is a continuous function

The limiting object described above is called **standard Brownian motion**. It will be our main object of study in this course and main building block in many applications. Today we will explore some simple properties of the Brownian motion.

Covariance of BM. Let $0 \le s \le t$. Then

$$\mathbb{E}B(s)B(t) = \mathbb{E}B(s)(B(t) - B(s) + B(s)) = \mathbb{E}B(s)(B(t) - B(s)) + \mathbb{E}B(s)^2.$$

From the independence of increments property one has $\mathbb{E}B(s)(B(t) - B(s)) = \mathbb{E}B(s)\mathbb{E}(B(t) - B(s)) = 0$. From the fact that $B(s)$ is a normally distributed random variable with expectation zero and variance s one has $\mathbb{E}B(s)^2 = s$ and thus

$$\mathbb{E}B(s)B(t) = s.$$

Non-differentiability of Brownian paths. In the following theorem we prove that trajectories of Brownian motion are very special – they are nondifferentiable [2] at every point. For you to appreciate this statement, try to come up with such an example, i.e. example of a function that is nondifferentiable at every point. To get an intuition what the above statement implies about Brownian paths, think about function $f(x) = |x|$. It is nondifferentiable at point 0 where it has a cusp, thus Brownian paths have cusp at every point, or yet in other words, Brownian motion changes direction every single moment.

Theorem 1. *With probability one Brownian paths are not Lipschitz continuous(and hence not differentiable) at any point.*

Proof. First of all, let us recollect the difference between continuity, Lipschitz continuity and differentiability.

- function f is continuous at point x_0 if $f(x) \to f(x_0)$ as $x \to x_0$

[1] this version June 1, 2022

[2] nondifferentiable = derivative does not exist

- function f is Lipschitz continuous at point x_0 if $|f(x) - f(x_0)| \le C|x - x_0|$ for some constant $C > 0$.
- function f is differentiable at point x_0 if $\lim\limits_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ exists.

Clearly, being Lipschitz continuity is a stronger property than continuity, it not just tells you that $f(x) \to f(x_0)$ as $x \to x_0$, it also tells you that the speed of convergence cannot be slower than $C|x - x_0|$. For instance, function $f(x) = \sqrt{|x|}$ is continuous at point 0 but not Lipschitz continuous. In it own turn, differentiability is a stronger property than Lipschitz continuity, an example of a function that is Lipschitz continuous but nondifferentiable is $f(x) = |x|$.

Fix some constant C and define event

$$A_n = \{w : \text{there is an } s \in [0,1] \text{ s.t. } |B(t) - B(s)| \le C|t - s| \text{ for } |t - s| \le\ 3/n\}.$$

Clearly, if a path is Lipschitz continuous with constant C at some point s then it belongs to set A_n for some large n. Therefore to prove that BM is not Lipschitz continuous it is enough to prove that $\mathbb{P}(A_n) = 0$. Let us think about the structure of sets A_n. Clearly $A_1 \subset A_2 \subset \dots \subset A_n \subset \dots$ Therefore, $\mathbb{P}(A_1) \le \mathbb{P}(A_2) \le \dots \le \mathbb{P}(A_n) \le \dots$

For $1 \le k \le n - 2$ let

$$\begin{aligned} Y_{k,n} &= \max\Big\{\Big|B\Big(\frac{k+j}{n}\Big) - B\Big(\frac{k+j-1}{n}\Big)\Big|, j = 0, 1, 2\Big\} \qquad (1)\\ &= \max\Big\{\Big|B\Big(\frac{k}{n}\Big) - B\Big(\frac{k-1}{n}\Big)\Big|, \Big|B\Big(\frac{k+1}{n}\Big) - B\Big(\frac{k}{n}\Big)\Big|, \Big|B\Big(\frac{k+2}{n}\Big) - B\Big(\frac{k+1}{n}\Big)\Big|\Big\}. \end{aligned}$$

and define event $G_n = \{w : \text{at least one } Y_{k,n} \text{ is } \le 5C/n\}$. We claim that $A_n \subset G_n$ and thus it is enough to prove that $\mathbb{P}(G_n) = 0$. Let's explain why $A_n \subset G_n$. Assume that point s is between points $\frac{k}{n}$ and $\frac{k+1}{n}$ and $|B(t) - B(s)| \le C|t - s|$ for $|t - s| \le \frac{3}{n}$. Then clearly

$$\begin{aligned} \Big|B\Big(\frac{k+1}{n}\Big) - B\Big(\frac{k}{n}\Big)\Big| &\le \Big|B(s) - B\Big(\frac{k}{n}\Big)\Big| + \Big|B\Big(\frac{k+1}{n}\Big) - B(s)\Big| \\ &\le C\Big(s - \frac{k}{n}\Big) + C\Big(\frac{k+1}{n} - s\Big) = C\frac{1}{n}. \qquad (2) \end{aligned}$$

$$\begin{aligned} \Big|B\Big(\frac{k+2}{n}\Big) - B\Big(\frac{k+1}{n}\Big)\Big| &\le \Big|B\Big(\frac{k+1}{n}\Big) - B(s)\Big| + \Big|B\Big(\frac{k+2}{n}\Big) - B(s)\Big| \\ &\le C\Big(\frac{k+1}{n} - s\Big) + C\Big(\frac{k+2}{n} - s\Big) = C\frac{3}{n}. \qquad (3) \end{aligned}$$

$$\begin{aligned} \Big|B\Big(\frac{k+3}{n}\Big) - B\Big(\frac{k+2}{n}\Big)\Big| &\le \Big|B\Big(\frac{k+2}{n}\Big) - B(s)\Big| + \Big|B\Big(\frac{k+3}{n}\Big) - B(s)\Big| \\ &\le C\Big(\frac{k+2}{n} - s\Big) + C\Big(\frac{k+3}{n} - s\Big) = C\frac{5}{n}. \qquad (4) \end{aligned}$$

Thus we conclude that $A_n \subset G_n$ and $\mathbb{P}(A_n) \le \mathbb{P}(G_n) \le n\mathbb{P}(Y_{k,n} \le 5C/n)$. The last inequality follows from the fact that

$$G_n = \bigcup_{1 \le k \le n-2} \{Y_{k,n} \le 5C/n\}.$$

From the independence of increments property of Brownian motion it follows that random variables $B(\frac{k}{n}) - B(\frac{k-1}{n}), B(\frac{k+1}{n}) - B(\frac{k}{n}), B(\frac{k+2}{n}) - B(\frac{k+1}{n})$ are independent and thus

$$n\mathbb{P}(Y_{n,k} \leq 5C/n) = n\mathbb{P}\Big(\Big|B\Big(\frac{1}{n}\Big)\Big| \leq 5C/n\Big)^3$$
$$= n\mathbb{P}(|B(1)| \leq 5C/\sqrt{n})^3 \leq n\left(\frac{10C}{\sqrt{n}}\frac{1}{\sqrt{2\pi}}\right)^3. \quad (5)$$

Letting $n \to \infty$ shows that $\mathbb{P}(A_n) \to 0$. Notice that $n \to A_n$ is increasing and thus $\mathbb{P}(A_n) = 0$ for all n. $\square$

Scaling properties of the Brownian Motion. If we fix time t then the distribution of Brownian motion $B(t)$ is normal with zero expectation and variance t. Actually, for any fixed set of times $t_1 < t_2 < \cdots < t_n$ random variables $B(t_1), B(t_2), \ldots, B(t_n)$ are jointly normal with zero mean and covariance given by $\mathbb{E}B(s)B(t) = \min(s,t)$.

Theorem 2. *If $B(0) = 0$ then for any $\lambda > 0$ stochastic process $\frac{1}{\sqrt{\lambda}}B(\lambda t), t \geq 0$ is a Brownian motion.*

Proof. First, let us notice that $X(t) = \frac{1}{\sqrt{\lambda}}B(\lambda t)$ is a Gaussian process, i.e. for any set $t_1 < t_2 < \cdots < t_n$ the joint distribution of $X(t_1), X(t_2), \ldots, X(t_n)$ is a multivariate Gaussian distribution. This property is clearly inherited from Brownian motion properties.

Since normal distribution is characterized by its mean and covariance we have to check that the mean and covariance of the process $X(t)$ coincide with those of Brownian motion. It is easy:

$$\mathbb{E}X(t) = \mathbb{E}\frac{1}{\sqrt{\lambda}}B(\lambda t) = \frac{1}{\sqrt{\lambda}}\mathbb{E}B(\lambda t) = 0, \quad (6)$$

and for $s < t$

$$\mathbb{E}X(s)X(t) = \mathbb{E}\frac{1}{\sqrt{\lambda}}B(\lambda s)\frac{1}{\sqrt{\lambda}}B(\lambda t) = \frac{1}{\lambda}\mathbb{E}B(\lambda s)B(\lambda t) = \frac{1}{\lambda}\min(\lambda s, \lambda t) = s. \quad (7)$$

Thus $\frac{1}{\sqrt{\lambda}}B(\lambda t)$ has the same distribution as a Brownian motion. $\square$

Theorem 3. *If $B(t)$ is a Brownian Motion starting at 0 then so is the process defined by $X(0) = 0$ and $X(t) = tB(1/t)$ for $t > 0$.*

Proof. Fix $t_1 < t_2 < \cdots < t_n$. Then clearly $X(t_1) = t_1 B(1/t_1), \ldots, X(t_n) = t_n B(1/t_n)$ has a multivariate Gaussian distribution. We just have to check that it has the same mean and covariance structure. First of all,

$$\mathbb{E}X(t) = t\mathbb{E}B(1/t) = t \cdot 0 = 0.$$

Also, for $0 < s < t$

$$\mathbb{E}X(s)X(t) = st\mathbb{E}B(1/s)B(1/t) = st\min(1/s, 1/t) = st \cdot 1/t = s$$

Thus $X(t)$ is a Brownian motion. $\square$

Quadratic Variation of Brownian Motion. Last time we computed quadratic variation of the scaled symmetric random walk and it turned out to be t. In the following theorem we prove that the quadratic variation of the BM is also t.

Let me again emphasize that the paths of Brownian motion are unusual in that their quadratic variation is not zero. This makes stochastic calculus different from ordinary calculus. In fact, non-zero quadratic variation is the source of volatility term in the Black-Scholes equation.

To see how Brownian Motion is different from functions we are used to in ordinary calculus let us consider a function $f : \mathbb{R} \to \mathbb{R}$ that is continuously differentiable and compute its quadratic variation up to time T. Let us introduce the norm of the partition $0 = t_0 < t_1 < \cdots < t_n = T$

$$||\Pi|| = \max_{j=0,1,\ldots,n-1} (t_{j+1} - t_j).$$

Then quadratic variation is

$$\langle f, f \rangle_T = \lim_{\max_j |t_{j+1}-t_j| \to 0} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2 = \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2. \quad (8)$$

We will use the mean-value theorem which says that in each interval (x_1, x_2) there is a point x_0 such that

$$f(x_2) - f(x_1) = f'(x_0)(x_2 - x_1).$$

In other words, somewhere between x_1 and x_2 the tangent line is parallel to the chord connecting points $(x_1, f(x_1))$ and $(x_2, f(x_2))$.

Applying mean-value theorem to the interval (t_j, t_{j+1}) we get that there is a point t_j^* such that

$$f(t_{j+1}) - f(t_j) = f'(t_j^*)(t_{j+1} - t_j).$$

Therefore

$$\begin{aligned} \sum_{j=0}^{n-1} (f(t_{j+1}) - f(t_j))^2 &= \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j)^2 \\ &\le \max_j |t_{j+1} - t_j| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j) \\ &= ||\Pi|| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j). \qquad (9) \end{aligned}$$

Sum $\sum\limits_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j)$ is a Riemann sum for the integral $\int\limits_0^T |f'(t)|^2 dt$.

$$\begin{aligned} \langle f, f \rangle_T &\le \lim_{||\Pi|| \to 0} ||\Pi|| \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j) \\ &= \lim_{||\Pi|| \to 0} ||\Pi|| \lim_{||\Pi|| \to 0} \sum_{j=0}^{n-1} |f'(t_j^*)|^2 (t_{j+1} - t_j). \qquad (10) \end{aligned}$$

The first limit is zero while the second has some finite value. As a result quadratic variation $\langle f, f \rangle$ is zero for functions f with finite $\int |f'(t)|^2 dt$. In particular, functions with continuous derivative have zero quadratic variation. For this reason we rarely consider quadratic variation in ordinary calculus.

Theorem 4. *Let $B(t)$ be a Brownian motion. Then $\langle B, B\rangle_T = T$.*

For the proof of this theorem see [2].

The fact that $\langle B, B\rangle_t = t$ is informally written as

$$dB(t)dB(t) = dt.$$

Remark 5. In addition to computing the quadratic variation of Brownian motion

$$\lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} (B(t_{j+1}) - B(t_j))^2 = T$$

we can compute the cross variation of $B(t)$ with t and the quadratic variation of t with itself, which are

$$\lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} (B(t_{j+1}) - B(t_j))(t_{j+1} - t_j) = 0$$

and

$$\lim_{||\Pi||\to 0} \sum_{j=0}^{n-1} (t_{j+1} - t_j)^2 = 0$$

Volatility of the Brownian Motion. Let α and σ be constants and define the geometric Brownian motion

$$S(t) = s(0)e^{\sigma B(t) + (\alpha - \sigma^2/2)t}.$$

This is the asset-price model used in the Black-Scholes formula. We show how to use the quadratic variation of BM to identify the volatility σ from a path of the process.

Let $0 \le T_1 < T_2$ be give and suppose we observe geometric BM $S(t)$ on time interval $T_1 \le tT_2$. We may then choose a partition of this interval $T_1 = t_0 < t_1 < \cdots < t_m = T_2$ and observe "log-returns"

$$\log \frac{S(t_{j+1})}{S(t_j)} = \sigma(B(t_{j+1}) - B(t_j)) + (\alpha - \sigma^2/2)(t_{j+1} - t_j)$$

over each of the subintervals $[t_j, t_{j+1}]$. The sum of the squares of the log-returns, sometimes called the realized volatility is

$$\begin{aligned} \sum_{j=0}^{m-1} \left(\log \frac{S(t_{j+1})}{S(t_j)}\right)^2 &= \sigma^2 \sum_{j=0}^{m-1} (B(t_{j+1}) - B(t_j))^2 + (\alpha - \sigma^2/2) \sum_{j=0}^{m-1} (t_{j+1} - t_j)^2 \\ &+ 2\sigma(\alpha - \sigma^2/2) \sum_{j=0}^{m-1} (B(t_{j+1}) - B(t_j))(t_{j+1} - t_j) \\ &\to \sigma^2(T_2 - T_1). \end{aligned} \tag{11}$$

References

[1] Steven Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*

[2] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*

[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

[4] Bernt Oksendal, *Stochastic Differential Equations*

Reading for this lecture:

(1) [1] pp. 108-115
(2) [2] pp. 25-31
(3) [3] pp. 95-97

Background: Vanilla Barrier Options. In finance, a barrier option is a type of contract where option to exercise at maturity depends on some observable (also called underlying of the option) crossing or reaching a given level, the so-called barrier. There are several types of vanilla barrier options ("vanilla" stands for simple and liquid on the market instrument). Some "knock-out" when the underlying asset price crosses a barrier (i.e., they become worthless). If the underlying asset price begins below the barrier and must cross the barrier above it to cause the knock-out, the option is said to be *up-and-out*. A *down-and-out* option has the barrier below the initial asset price and knocks out if the asset price falls below the barrier. Other options "knock-in" at a barrier (i.e., payoff zero unless they cross a barrier). Knock-in options also fall in two categories: *up-and-in* and *down-and-in*. The payoff at expiration is often a fixed amount, a call or a put. There are also more complicated barrier options, for instance, range accrual options which for a specified financial observable, for instance 3-month Libor rate or 2-year CMS rate, pays a fixed amount multiplied by the fraction of observations when the index is inside a specified range.

Later in this lecture we treat the simplest case. We price an up-and-out option whose payoff at expiration is a call and we assume that the stock price is modelled by Brownian motion. Assumption that the stock price can be modelled by Brownian motion is rather unrealistic as Brownian motion (even with large drift) can be negative with positive probability. Later in the course we will be able to apply similar calculations for the more realistic case when the asset price follows a geometric Brownian motion.

Running maximum and first passage time. For a stochastic process $X_t, t \geq 0$ we define the running maximum as

$$M_t = \max_{0 \leq s \leq t} X_s. \tag{1}$$

For this process to be well defined we assume that the process X_t is has continuous trajectories(in fact, we need only with probability one). Closely related to running maximum is the first passage time which is defined as

$$T_a = \inf\{t > 0 : X_t = a\}, \tag{2}$$

that is for a fixed a time T_a is the first time when X_t reaches level a. From the definitions it is clear that events $\{T_a < t\}$ and $\{M_t \geq a\}$ coincide, i.e.,

$$w \in \{T_a < t\} \Longleftrightarrow w \in \{M_t > a\}. \tag{3}$$

From the examples in the beginning of the lecture one can conclude that the value of many barrier options depends on the behavior of the maximum asset price prior to option expiration, or equivalently, on the distribution of the running maximum. For instance, the knowledge of the distribution of the running maximum (or equivalently of the first passage time) is enough to compute the value of the simple knock-in/out options.

But it turns out that for a general stochastic process X_t the distribution of the running maximum is rather complicated and often cannot be written out in a closed form.

Brownian motion. In the case when stochastic process X_t is a Brownian motion we can explicitly calculate the distribution of the running maximum and hitting time.

Theorem 1 (Reflection Principle.)**.** *Let* $a > 0$*. Then* $\mathbb{P}(T_a < t) = 2\mathbb{P}(B_t > a)$.

Remark 2. Before the start of the proof let us rewrite the above equation as

$$\mathbb{P}(T_a < t) = 2\int_a^\infty \frac{e^{-x^2/(2t)}}{\sqrt{2\pi t}}dx. \tag{4}$$

To find the probability density of T_a we change variables $x = \frac{\sqrt{t}a}{\sqrt{s}}$. Then $dx = -\frac{\sqrt{t}a}{2s^{3/2}}$ and

$$\mathbb{P}(T_a < t) = \int_0^t \frac{a}{\sqrt{2\pi s^3}}e^{-a^2/2s}ds, \tag{5}$$

and thus the density of T_a is $\frac{a}{\sqrt{2\pi s^3}}e^{-a^2/2s}$.

Proof. Assume that B_t hits level a at some time $s < t$. From the independence of increments property it follows that $B_t - B_{T_a}$ is independent of what happened before time T_a. Moreover, the increment $B_t - B_{T_a}$ is normally distributed. Since normal distribution is symmetric and probability of B_t being equal to a is zero we obtain

$$\mathbb{P}(T_a < t, B_t > a) = \frac{1}{2}\mathbb{P}(T_a < t). \tag{6}$$

We multiply by 2 and notice that event $\{B_t > a\}$ is a subset of the event $\{T_a < t\}$, thus

$$\mathbb{P}(T_a < t) = 2\mathbb{P}(T_a < t, B_t > a) = 2\mathbb{P}(B_t > a). \tag{7}$$

□

Remark 3. Truly speaking, the fact that $B_t - B_{T_a}$ is independent of $\mathcal{F}_{T_a}$ ("information" available at time T_a) needs to be proved using the strong Markov property for Brownian motion. We refer to [2] for the proof.

One can go further and generalize the result of Theorem 1. Let $u < v \le a$, then using the reflection principle we can easily show that

$$\mathbb{P}(T_a < t, u < B_t < v) = \mathbb{P}(2a - v < B_t < 2a - u). \tag{8}$$

Since events $\{T_a < t\}$ and $\{M_t > a\}$ are equivalent we obtain

$$\mathbb{P}(M_t > a, u < B_t < v) = \mathbb{P}(2a - v < B_t < 2a - u). \tag{9}$$

Now let the interval (u, v) shrink to x, so that

$$\mathbb{P}(M_t > a, B_t = x) = \mathbb{P}(B_t = 2a - x) = \frac{1}{\sqrt{2\pi t}}e^{-(2a-x)^2/(2t)}. \tag{10}$$

Differentiating with respect to a we get the joint density

$$\mathbb{P}(M_t = a, B_t = x) = \frac{2(2a - x)}{\sqrt{2\pi t^3}}e^{-(2a-x)^2/(2t)}. \tag{11}$$

Knock-In/Out Option on Brownian motion. Let us consider the option which is knocked-out if before the expiration time t asset price crosses barrier M. The payoff at expiration is a call with strike K. Let us assume for simplicity that the interest rate is equal to zero so that the value of the option is equal to

$$\mathbb{E}\,\mathbf{1}_{M_t<M}(B_t-K)_+. \tag{12}$$

In order to calculate the above expectation we use our knowledge of the joint density of M_t and B_t given by (11):

$$\begin{aligned}
\mathbb{E}\,\mathbf{1}_{M_t<M}(B_t-K)_+ &= \iint \mathbf{1}_{a<M}(x-K)_+\frac{2(2a-x)}{\sqrt{2\pi t^3}}e^{-(2a-x)^2/(2t)}dadx \\
&= \int_K^M\int_x^M (x-K)\frac{2(2a-x)}{\sqrt{2\pi t^3}}e^{-(2a-x)^2/(2t)}dadx \\
&= \int_K^M (x-K)\frac{e^{-x^2/2t}}{\sqrt{2\pi t}}dx - \int_K^M (x-K)\frac{e^{-(x-2M)^2/2t}}{\sqrt{2\pi t}}dx. \qquad (13)
\end{aligned}$$

The above difference could be easily expressed in terms of the cumulative normal distribution. Indeed, the first term is a difference of the following two integrals

$$\begin{aligned}
\int_K^M x\frac{e^{-x^2/2t}}{\sqrt{2\pi t}}dx &= \sqrt{t}\int_{K/\sqrt{t}}^{M/\sqrt{t}} x\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = \sqrt{\frac{t}{2\pi}}\left(e^{-K^2/2t}-e^{-M^2/2t}\right) \\
\int_K^M K\frac{e^{-x^2/2t}}{\sqrt{2\pi t}} &= K\int_{K/\sqrt{t}}^{M/\sqrt{t}} \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx = K\left(N\Big(\frac{M}{\sqrt{t}}\Big)-N\Big(\frac{K}{\sqrt{t}}\Big)\right), \qquad (14)
\end{aligned}$$

where $N(x)=\int\limits_{-\infty}^{x} e^{-z^2/2}/\sqrt{2\pi}dz$ is the cumulative normal distribution.

Let us remark that the same method applies for the case of general payoff and knock-out condition. If the payoff at expiration is given by function $f(x)$ and knock-out condition is given by $M_t\in A$ then the value of the option is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{1}_{a\in A}f(x)\frac{2(2a-x)}{\sqrt{2\pi t^3}}e^{-(2a-x)^2/(2t)}dadx. \tag{15}$$

One has to compute the above integral numerically, for instance, by using 2D Simpson's rule.

Let us say a few words why having a closed form solution for vanilla options is so important. The reason is that the greeks of the closed form solution could be calculated analytically while otherwise the greeks are computed numerically. Calculating derivatives numerically could be very noisy and may lead to unstable hedges which is quite undesirable taking into consideration the transaction cost.

The first passage time. Finally, let us compute the distribution of the first passage time using approach different from the presented above. In particular, we use the martingale approach.

Let us start with the proof that for any positive constant σ stochastic process $Z_t = e^{\sigma B_t - \sigma^2 t/2}$ is a martingale. Indeed, for $s < t$

$$\begin{aligned}\mathbb{E}\left(e^{\sigma B_t - \sigma^2 t/2} | \mathcal{F}_s\right) &= \mathbb{E}\left(e^{\sigma(B_t - B_s) + \sigma B_s - \sigma^2 t/2} | \mathcal{F}_s\right) \\ &= e^{\sigma B_s - \sigma^2 t/2}\mathbb{E}\left(e^{\sigma(B_t - B_s)} | \mathcal{F}_s\right) \qquad (16) \\ &= e^{\sigma B_s - \sigma^2 t/2} e^{\sigma^2 (t-s)/2}. \qquad (17)\end{aligned}$$

Equation (16) follows from the fact that B_s is measurable in σ-algebra $\mathcal{F}_s$, or in other words, with information up to time s at your disposal you know B_s. Equation (17) follows from the independence of increments property and the fact that for a normally distributed random variable $X = N(a, \sigma^2)$

$$\mathbb{E}e^{uX} = e^{ua + \frac{1}{2}u^2\sigma^2}. \tag{18}$$

Let us use the fact that Z_t is a martingale to find the moment generating function of the first passage time T_a. Let us apply the martingale property at moment T_a :

$$\mathbb{E}\left(Z_{T_a} | \mathcal{F}_0\right) = Z_0, \tag{19}$$

or equivalently

$$\mathbb{E}e^{\sigma B_{T_a} - \sigma^2 T_a/2} = 1. \tag{20}$$

But $B_{T_a} = a$ by the definition of the first passage time. Thus

$$\mathbb{E}e^{\sigma a - \sigma^2 T_a/2} = 1 \Longleftrightarrow \mathbb{E}e^{-\sigma^2 T_a/2} = e^{-\sigma a}. \tag{21}$$

Changing variable $\sigma^2 = 2u$

$$\mathbb{E}e^{-uT_a} = e^{-a\sqrt{2u}}. \tag{22}$$

Thus we obtained the moment generating function of the first passage time T_a. Why is it important? Because by the moment generating function we can unambiguously recover the distribution function of T_a.

References

[1] Steven Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*
[2] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*
[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

Stochastic Calculus, Summer 2022, June 15,
Lecture 4
Construction of the Ito Integral

Reading for this lecture (for references see the end of the lecture):

- [1] pp. 125-137
- [2] pp. 33-67
- [3] pp. 128-148
- [4] pp. 21-42

Motivation. Consider an asset whose price per share is equal to $X_t, t \geq 0$ and a portfolio that initially consists of Δ_0 shares. Consider the following trading strategy: keep an initial position Δ_0 up to time $t_1 \geq t_0 = 0$ and then re-balance the portfolio by taking position Δ_1 in the asset. Keep it up to time $t_2 \geq t_1$ and then re-balance the portfolio again by taking position Δ_2 is the asset. In general, we re-balance the portfolio at trading date t_i by taking position Δ_i in the asset and keeping it till the next trading date t_{i+1}. What is the the profit $I_T(\Delta)$ of the above trading strategy at time T? Clearly it should be

$$I_T(\Delta) = \Delta_0(X_{t_1} - X_{t_0}) + \Delta_1(X_{t_2} - X_{t_1}) + \cdots + \Delta_{n-1}(X_{t_n} - X_{t_{n-1}}) \qquad (1)$$

and by analogy with the Riemann integral we write symbolically

$$I_T(\Delta) = \int_0^T \Delta(t) dX(t), \qquad (2)$$

where $\Delta(t)$ is a piecewise constant function which is equal to Δ_i on $[t_i, t_{i+1}]$.

We fix an interval $[S, T]$ and try to make sense of

$$\int_S^T f(t, w) dX_t(w), \qquad (3)$$

where $f(t, w)$ is a random function and $dX_t(w)$ refers to the increments of a stochastic process X_t. Before we proceed we have to clarify a few things.

- First, we restrict attention to such functions f that for any fixed t random variable $f(t, w)$ is $\mathcal{F}_t$-measurable. To explain this restriction let's come back to our example (2). Position Δ_i we take in the asset at time $t_i, i \geq 1$, may depend on the price history of the asset, $\mathcal{F}_t$, but it must be independent of the future behavior of the process X_t.
- Second, we restrict our consideration to the case when X_t is a Brownian motion. The case of general stochastic process X_t is quite similar.

If X_t is a differentiable function, then we can define

$$\int_S^T f(t, w) dX_t = \int_S^T f(t, w) X_t' dt, \qquad (4)$$

where the right-hand side is an ordinary Riemann integral with respect to time, that is for every trajectory of X_t we can define $\int_S^T f(t, w) dX_t$. This approach does

[1]this version June 16, 2022

not work for Brownian motion as we saw that the trajectories of Brownian motion are not differentiable.

Stochastic integral of simple function $f(t,w)$**.** Just like in the construction of the Riemann integral $\int_S^T f(t)dt$, where $f(t)$ is a deterministic function, we start with a construction of stochastic integral for a simple class of functions f and then extend by some approximation procedure.

Assume that $\Pi = \{t_0, t_1, \dots, t_n\}$ is a partition of $[S,T]$, i.e.

$$S = t_0 \leq t_1 \leq \cdots \leq t_n = T, \tag{5}$$

and that $f(t,w)$ is constant in t on each subinterval $(t_j, t_{j+1}]$. Such a process $f(t,w)$ called a *simple process*. We start with defining integral (3) for simple processes. Consider interval $[t_0, t_1]$, on this interval $f(t,\omega) = e_1(\omega)$ is a random quantity, but independent of t and thus it is natural to define

$$\int_{t_0}^{t_1} f\left(t,\omega\right) dB_t\left(\omega\right) = \int_{t_0}^{t_1} e_1\left(\omega\right) dB_t\left(\omega\right) = e_1\left(\omega\right)\left(B\left(t_1\right) - B\left(t_0\right)\right).$$

Applying this procedure to intervals $[t_1,t_2],[t_2,t_3],\dots,[t_{n-1},t_n]$ we get

$$\begin{aligned}\int_S^T f\left(t,w\right) dB_t\left(w\right) = e_1\left(w\right)\left(B_{t_1} - B_{t_0}\right) + e_2\left(w\right)\left(B_{t_2} - B_{t_1}\right)\\ + e_3\left(w\right)\left(B_{t_3} - B_{t_2}\right) + \dots + e_n\left(w\right)\left(B_{t_n} - B_{t_{n-1}}\right).\end{aligned} \tag{6}$$

Naturally, to define stochastic integral (3) for general process $f(t,w)$ we approximate it with simple processes similarly to approximation of continuous functions by stepwise constant functions in the theory of Riemann integration. But without any further assumption on approximating functions $e_i(w)$, our definition of the integral leads to difficulties. Here is an example of what kind of difficulties we can expect. Consider

$$\int_0^T B_t dB_t. \tag{7}$$

Riemann integral is a limit of Riemann sums:

$$\int_S^T f(t)dt \approx \sum_{i=0} f(t_i^*)(t_{i+1} - t_i), \tag{8}$$

where t_i^* is ANY point on the interval $[t_i, t_{i+1}]$. When the length of the longest interval in the partition tends to zero the limit is $\int_S^T f(t)dt$. Let us point out that it was not important what point t_i^* we took inside the interval $[t_i, t_{i+1}]$. For example, it could be t_i(left point approximation) or t_{i+1}(right point approximation). Let us try to do the same for integral (7).

Left point approximation:

$$I_1 \cong \sum_i B\left(t_i\right)\left(B\left(t_{i+1}\right) - B\left(t_i\right)\right). \tag{9}$$

Right point approximation:

$$I_2 \cong \sum_i B(t_{i+1})(B(t_{i+1}) - B(t_i)). \tag{10}$$

From the independence of increments of Brownian motion and the fact that $\mathbb{E}[(B(t_{i+1}) - B(t_i))] = \mathbb{E}[B(t_1)] = 0$ we have

$$\begin{aligned}
\mathbb{E}(I_1) &= \sum_i \mathbb{E}[B(t_i)(B(t_{i+1}) - B(t_i))] \\
&= \sum_i \mathbb{E}[B(t_i)]\,\mathbb{E}[(B(t_{i+1}) - B(t_i))] = 0. \\
\mathbb{E}(I_2) &= \sum_i \mathbb{E}[B(t_{i+1})(B(t_{i+1}) - B(t_i))] \\
&= \sum_i \mathbb{E}\left[B(t_{i+1})^2 - B(t_{i+1})B(t_i)\right] \\
&= \sum_i [t_{i+1} - t_i] = T, \qquad (11)
\end{aligned}$$

since $\mathbb{E}\left[B(t_{i+1})^2\right] = t_{i+1}$ and $\mathbb{E}[B(t_{i+1})B(t_i)] = t_i$ as follows from

$$\mathbb{E}(B(t)B(s)) = \min(s,t). \tag{12}$$

Thus we see that depending on the choice of the point t_i^* in the approximation we can get very different results. Function $f(t,w)$ is $\mathcal{F}_t$-measurable and thus it is reasonable to choose the approximating simple function to be $\mathcal{F}_t$-measurable as well. We therefore have to choose the left end point approximation. In what follows we choose

$$t_i^* = t_i \text{ (left end point approximation)} \tag{13}$$

which leads to the Itô integral.

Remark 1. If for each $t \geq 0$ random variable $f(t,w)$ is $\mathcal{F}_t$ measurable we say that the process $f(t,w)$ is $\mathcal{F}_t$-adapted. For example, if $\mathcal{F}_t$ is filtration of Brownian motion then the process $f_t(t,w) = B(t/2)$ is $\mathcal{F}_t$-adapted, while $f_t(t,w) = B(2t)$ is not.

Properties of the Itô integral for simple processes. The Itô integral (3) is defined as the gain from treading in the martingale B_t. A martingale has no tendency to rise or fall and hence it is to be expected that

$$I_t(f) = \int_0^t f(t,w)dB_t \tag{14}$$

also has no tendency to rise or fall.

Theorem 2. *Itô integral is a martingale.*

Proof. see [1], pages 128-129. □

Because $I_t(f)$ is a martingale and $I_0 = 0$ we have $\mathbb{E}I_t(f) = 0$ for all $t \geq 0$. It follows that $\mathrm{Var}I_t(f) = \mathbb{E}I_t^2(f)$ can be evaluated by the formula in the following theorem.

Theorem 3 (Itô's Isometry)**.** *The Itô integral satisfies*

$$\mathbb{E}I_t^2(f) = \mathbb{E}\int\limits_0^t f(s,w)^2 ds \tag{15}$$

Proof. For the simplicity of notation we introduce $\Delta B_i = B(t_{i+1}) - B_t(t_i), e_i = e_i(w)$. Then by definition

$$I_t(f) = \int_0^t f(s,\omega)\, dB_s = \sum_i e_i \Delta B_i \tag{16}$$

and

$$\left(\int_S^T f(t,w)\, dB_t\right)^2 = \left(\sum_i e_i \Delta B_i\right)^2 = \sum_{i,j} e_i e_j \Delta B_i \Delta B_j. \tag{17}$$

Taking expectation

$$\mathbb{E}\left(\int_S^T f(t,\omega)\, dB_t\right)^2 = \sum_{i,j} \mathbb{E}\left(e_i e_j \Delta B_i \Delta B_j\right). \tag{18}$$

$$\mathbb{E}\left(e_i e_j \Delta B_i \Delta B_j\right) = \begin{cases} 0 = \mathbb{E}\left(\Delta B_j\right), \text{ if } i < j \\ \mathbb{E}(e_i^2 \Delta B_i^2), \text{ if } i = j \end{cases}$$

For $i = j$ we use independence of increments property to conclude

$$\mathbb{E}\left(e_i^2 \left(\Delta B_i\right)^2\right) = \mathbb{E}\left(e_i^2\right)\mathbb{E}\left(\Delta B_i\right)^2 = \mathbb{E}\left(e_i^2\right)\left(t_{i+1} - t_i\right) = \mathbb{E}\left(e_i^2\right)\Delta t_i. \tag{19}$$

$$\sum_i \mathbb{E}\left(e_i^2\right)\Delta t_i = \mathbb{E}\sum_i \left(e_i^2\right)\Delta t_i = \mathbb{E}\int_0^t \phi\left(t,\omega\right)^2 dt \tag{20}$$

□

Theorem 4. *Quadratic variation of the stochastic integral* (3) *is equal to*

$$\int\limits_0^T f^2(t,w)dt = \sum_i e_i^2 \Delta t_i. \tag{21}$$

For this purpose we consider the quantity

$$\mathbb{E}\left(\sum_i (e_i \Delta B_i)^2 - \sum_i e_i^2 \Delta t_i\right)^2 \tag{22}$$

and prove that it approaches 0 as $||\Pi|| \to 0$. We first rewrite it as

$$\begin{aligned}\mathbb{E}\Big(\sum_i e_i^2 (\Delta B_i)^2 - \sum_i e_i^2 \Delta t_i\Big)^2 &= \mathbb{E}\left(\sum_i e_i^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big]\right)^2 \\ &= \mathbb{E}\sum_{i,j} e_i^2 e_j^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big]\Big[(\Delta B_j)^2 - \Delta t_j\Big]. \end{aligned} \tag{23}$$

Just as in the calculation of the quadratic variation of the Brownian motion let us split the above sum in two sums: in the first one we keep the terms with $i \neq j$ and in the second one we keep terms with $i = j$.

Let us first look at terms with $i \neq j$, for instance $i < j$. Then $\Big[(\Delta B_j)^2 - \Delta t_j\Big]$ is independent of $e_i^2 e_j^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big]$ an thus

$$\begin{aligned}\mathbb{E}e_i^2 e_j^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big]\Big[(\Delta B_j)^2 - \Delta t_j\Big] &= \mathbb{E}e_i^2 e_j^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big]\mathbb{E}\Big[(\Delta B_j)^2 - \Delta t_j\Big] \\ &= \mathbb{E}e_i^2 e_j^2 \Big[(\Delta B_i)^2 - \Delta t_i\Big] \cdot 0 = 0. \end{aligned} \tag{24}$$

Let us now consider the case of $i = j$. Then

$$\mathbb{E}e_i^4 \Big[(\Delta B_i)^2 - \Delta t_i\Big]^2 = \mathbb{E}e_i^4 \mathbb{E}\Big[(\Delta B_i)^2 - \Delta t_i\Big]^2, \tag{25}$$

since random variables e_i^4 and $\Big[(\Delta B_i)^2 - \Delta t_i\Big]^2$ are independent. It follows from the fact that e_i is $\mathcal{F}_t$-measurable and thus ΔB_i is independent of e_i. But $\mathbb{E}\Big[(\Delta B_i)^2 - \Delta t_i\Big]^2 = 2(\Delta t_i)^2$ and thus

$$\begin{aligned}\mathbb{E}\left(\sum_i (e_i \Delta B_i)^2 - \sum_i e_i^2 \Delta t_i\right)^2 &= \sum_i 2\mathbb{E}e_i^4 (\Delta t_i)^2 \\ &\leq ||\Pi|| \sum_i 2\mathbb{E}e_i^4 \Delta t_i. \end{aligned} \tag{26}$$

Since $\sum_i 2\mathbb{E}e_i^4 \Delta t_i$ converges to $\int_0^T \mathbb{E}f^4 dt < \infty$. Thus

$$\mathbb{E}\left(\sum_i (e_i \Delta B_i)^2 - \sum_i e_i^2 \Delta t_i\right)^2 \to 0 \tag{27}$$

and quadratic variation of $I(f)$ is proved to be

$$\int_0^T f^2(t, w) dt. \tag{28}$$

Itô integral for general functions. We now describe the class of functions $f(t, w)$ for which the Itô integral will be defined.

Definition 5. *Let $V = V(S, T)$ be the class of functions $f(t, w) : [0, \infty) \times \Omega \to \mathbb{R}$, such that:*

(1) *$f(t, w)$ is $\mathcal{F}_t$ – adapted, that is $f(t, w)$ is $\mathcal{F}_t$ measurable for all $t \leq T$*
(2) *$\int_S^T f(t, w)^2 dt < \infty$.*

We claim that each function $f \in V(S, T)$ can be approximated by a sequence $\{\varphi_n\}_{n=1,2,...}$ of simple functions (or equivalently, by a sequence of simple processes) in the sense that as $n \to \infty$

$$\mathbb{E}\int_S^T (f - \varphi_n)^2 \to 0. \tag{29}$$

The approximation is done is three steps:
Step 1 (Approximate bounded continuous functions with simple functions)

Let $g \in V$ be bounded, i. e., every trajectory $g(.,w)$ (w is fixed and t changes) is continuous. Then, there exists a sequence of simple functions $\varphi_n \in V$, such that as $n \to \infty$

$$\mathbb{E} \int_S^T (g - \varphi_n)^2 \, dt \to 0. \tag{30}$$

Step 2 (Approximate bounded functions with bounded continuous functions)
Let $h \in V$ be bounded, then there exists a sequence of bounded continuous functions g_n, such that

$$\mathbb{E} \int_S^T (h - g_n)^2 \, dt \to 0. \tag{31}$$

Step 3 (Approximate general functions with bounded functions)
Let $f \in V$, then there exists a sequence of bounded functions h_n, such that

$$\mathbb{E} \int_S^T (f - h_n)^2 \, dt \to 0. \tag{32}$$

Putting together steps 1,2 and 3 we get that for any function $f(t,w) \in V$ there exists a sequence of simple functions $\varphi_n(t,w)$ such that (30) is true. We define then the Itô integral of function $f(t,w)$ as

$$I(f) = \int_S^T f(t,\omega) \, dB_t = \lim_{n\to\infty} I(\varphi_n). \tag{33}$$

Question: Why does the limit exist and in what sense?

Answer: By Theorem 3 we have that

$$\begin{aligned} \mathbb{E}(I(\varphi_n) - I(\varphi_m))^2 &= \mathbb{E} \int\limits_S^T (\varphi_n - \varphi_m)^2 dt \\ &\le \mathbb{E} \int\limits_S^T (f - \varphi_m)^2 dt + \mathbb{E} \int\limits_S^T (\varphi_n - f)^2 dt \to 0. \end{aligned} \tag{34}$$

Thus the sequence of random variables $\left\{ \int_S^T \varphi_n(t,\omega) \, dB_t \right\}$ forms a Cauchy sequence in $L_2(\Omega, \mathcal{F}, \mathbb{P})$. Since $L_2(\Omega, \mathcal{F}, \mathbb{P})$ is a complete space then there exists a limit of $I(\varphi_n)$ as an element of $L_2(\Omega, \mathcal{F}, \mathbb{P})$. This limit is by definition the Itô integral $I(f)$.

Example: Compute $\int\limits_0^T B_t dB_t$.

By definition

$$\int_0^T B_t dB_t = \lim_{n\to\infty} \int_0^T \varphi_n(t,\omega) \, dB_t, \tag{35}$$

where φ_n is such that $\mathbb{E} \int_0^T (\phi_n - B_t)^2 \, dt \to 0$ and φ_n is $\mathcal{F}_t$-adapted.

As we already saw in the beginning of the lecture we can approximate $f(t,w) = B_t$ by partitioning $[0,T]$ into $[0,t_1], [t_1,t_2], \ldots, [t_n-1, t_n = T]$ and defining $\varphi_n(t,w) = B(t_i)$ for $t \in [t_i, t_{i+1}]$. Let us first check that φ_n indeed approximated f in the sense of (29):

$$\mathbb{E}\int_0^T (\phi_n - B_t)^2\,dt = \mathbb{E}\sum_i \int_{t_i}^{t_{i+1}} (\phi_n - B_t)^2\,dt = \sum_i \int_{t_i}^{t_{i+1}} \mathbb{E}\,(B\,(t_i) - B\,(t))^2\,dt$$

$$= \sum_i \int_{t_i}^{t_{i+1}} (t - t_i)\,dt = \sum_i \frac{(t_{i+1} - t_i)^2}{2}. \tag{36}$$

If we define $\max(t_{i+1} - t_i) = M_n$ then

$$\sum_i \frac{(t_{i+1} - t_i)^2}{2} \le \sum_i \frac{t_{i+1} - t_i}{2} M_n = \frac{M_n}{2}\sum_i (t_{i+1} - t_i) = \frac{M_n}{2} T \to 0. \tag{37}$$

Thus we have to compute $\int_0^T \phi_n dB_t = \sum_i B\,(t_i)\,\Delta B_i$, where $\Delta B_i = B(t_{i+1}) - B(t_i)$.

We use the following identity

$$\Delta B_i^2 = B(t_{i+1})^2 - B(t_i)^2 = (B(t_{i+1}) - B(t_i))^2 + 2B(t_{i+1})B(t_i) - 2B(t_i)^2$$

$$= (B(t_{i+1}) - B(t_i))^2 + 2B(t_i)(B(t_{i+1}) - B(t_i)). \tag{38}$$

Summing both parts over i we get

$$B_T^2 = \sum_i (B(t_{i+1}) - B(t_i))^2 + 2I\,(\phi_n)\,. \tag{39}$$

Therefore

$$I\,(\varphi_n) = \frac{B_T^2}{2} - \frac{1}{2}\sum_i (B_{i+1} - B_i)^2\,,\ \text{but}\ \sum_i (B_{i+1} - B_i)^2 \xrightarrow[L_2]{} T. \tag{40}$$

Finally,

$$\int_0^T B_t dB_t = \frac{B_T^2}{2} - \frac{T}{2}. \tag{41}$$

References

[1] Steven Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*

[2] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*

[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

[4] Bernt Oksendal, Stochastic Differential Equations

Stochastic Calculus, Summer 2022, June 22, Lecture 5 Itô Formula

Reading for this lecture (for references see the end of the lecture):

- [1] pp. 137-153
- [2] pp. 68-70
- [3] pp. 148-156
- [4] pp. 21-42

During the last lecture we defined Itô integral

$$\int_0^t f(s,\omega)\, dB_s = I(f) \tag{1}$$

for stochastic processes $f(t,w)$ that are $\mathcal{F}_t$-adapted and square-integrable. We defined $I(f)$ to be the limit in $L^2(\Omega,\mathcal{F},\mathbb{P})$ of $I(\varphi_n)$ for any sequence of approximating simple stochastic processes φ_n. Approximating sequence was defined as satisfying

$$\mathbb{E}\int_0^t (\phi_n - f)^2\, dt \to 0 \text{ when } n\to\infty. \tag{2}$$

Thus the procedure of calculating Itô integral (1) from the definition is rather work consuming.

Let us step back for a second and take a look at the Riemann integral. Even though it is defined as the limit of Riemann sums in practice one never does this. Instead, one uses the fundamental theorem of calculus and the chain rule. For instance, to compute $I(t)=\int\limits_0^t se^{-s^2/2}ds$ we notice that $(-e^{-s^2/2})' = se^{-s^2/2}$ and thus

$$I(t) = \int\limits_0^t se^{-s^2/2}ds = -e^{-s^2/2}\Big|_0^t = 1-e^{-t^2/2}. \tag{3}$$

In general, it is desirable to have some analog of the chain rule in the case of Itô integral (to avoid taking the limit of $I(\varphi_n)$. The analog for the chain rule is the Itô formula.

Itô formula for Brownian motion. Let us first analyze the difficulties that arise when we want to "differentiate" expressions of the form $f(B_t)$, where f is a differentiable function and B_t is a Brownian motion. If Brownian motion were differentiable then the chain rule would give

$$\frac{df(B_t)}{dt} = \frac{df}{dx}\Big|_{x=B_t}\frac{dB_t}{dt}. \tag{4}$$

The problem that we face is the fact that Brownian motion B_t is non-differentiable and thus formula (4) does not apply.

[1]this version June 22, 2022

Theorem 1. *Let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x)$, $f_x(t,x)$ and $f_{xx}(t,x)$ are defined and continuous, and let B_t be a Brownian motion. Then for every $T \geq 0$*

$$f(T,B_T) = f(0,B_0) + \int_0^T f_t(t,B_t)dt + \int_0^T f_x(t,B_t)dB_t + \frac{1}{2}\int_0^T f_{xx}(t,B_t)dt. \tag{5}$$

Proof. see [1], pp. 138-139. Briefly, from the Taylor series expansion we obtain

$$\begin{aligned} f(t_{j+1},x_{j+1}) - f(t_j,x_j) &= f_t(t_j,x_j)(t_{j+1}-t_j) + f_x(t_j,x_j)(x_{j+1}-x_j) \\ &+ \frac{1}{2}f_{xx}(t_j,x_j)(x_{j+1}-x_j)^2 + f_{tx}(t_j,x_j)(x_{j+1}-x_j)(t_{j+1}-t_j) \\ &+ \frac{1}{2}f_{tt}(t_j,x_j)(t_{j+1}-t_j)^2 + \text{ higher order terms.} \end{aligned} \tag{6}$$

Apply this formula with $x_{j+1} = B(t_{j+1}), x_j = B(t_j)$ and sum over j :

$$\begin{aligned} f(T,B_T) - f(0,B_0) &= \sum_j \Big(f(t_{j+1},B(t_{j+1})) - f(t_j,B(t_j)) \Big) \\ &= \sum_j f_t(t_j,B(t_j))(t_{j+1}-t_j) \\ &+ \sum_j f_x(t_j,B(t_j))(B(t_{j+1})-B(t_j)) \\ &+ \sum_j \frac{1}{2} f_{xx}(t_j,B(t_j))(B(t_{j+1})-B(t_j))^2 \\ &+ \sum_j f_{tx}(t_j,B(t_j))(B(t_{j+1})-B(t_j))(t_{j+1}-t_j) \\ &+ \sum_j \frac{1}{2} f_{tt}(t_j,B(t_j))(t_{j+1}-t_j)^2 + \text{ higher order terms.} \end{aligned} \tag{7}$$

As we take the limit $||\Pi|| \to 0$ then the first term on the right-hand side converges to an ordinary Riemann integral

$$\sum_j f_t(t_j,B(t_j))(t_{j+1}-t_j) \to \int_0^T f_t(t,B_t)dt. \tag{8}$$

As $||\Pi|| \to 0$ the second term converges to an Itô integral

$$\sum_j f_x(t_j,B(t_j))(B(t_{j+1})-B(t_j)) \to \int_0^T f_x(t,B_t)dB_t. \tag{9}$$

Let us study the third sum. To simplify notation put $a_j = f_{xx}(t_j,B_j), \Delta B_j = B(t_{j+1}) - B(t_j)$. Then

$$\sum_j \frac{1}{2} f_{xx}(t_j,B(t_j))(B(t_{j+1})-B(t_j))^2 = \sum_j a_j \left(\Delta B_j\right)^2. \tag{10}$$

Consider

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

If $i < j$ then $a_i a_j((\Delta B_i)^2 - \Delta t_i)$ and $(\Delta B_j)^2 - \Delta t_j$ are independent and so the terms vanish in this case. Similarly for $i > j$. So we are left with

$$\begin{aligned}\sum_j \mathbb{E}\left[a_j^2((\Delta B_j)^2 - \Delta t_j)^2\right] &= \sum_j \mathbb{E}\left[a_j^2\right]\mathbb{E}\left[(\Delta B_j)^4 - 2(\Delta B_j)^2 \Delta t_j + (\Delta t_j)^2\right] \\ &= \sum_j \mathbb{E}\left[a_j^2\right]\left[3(\Delta t_j)^2 - 2\Delta t_j \Delta t_j + (\Delta t_j)^2\right] \\ &= \sum_j \mathbb{E}\left[a_j^2\right](\Delta t_j)^2 \to 0 \text{ as } \Delta t_j \to 0. \qquad (11)\end{aligned}$$

Thus the third term converges to

$$\frac{1}{2}\int_0^T f_{xx}(t, B_t)dt. \tag{12}$$

□

Remark 2. One often writes Itô formula in the differential form:

$$df(t, B_t) = f_t(t, B_t)dt + f_x(t, B_t)dB_t + \frac{1}{2}f_{xx}(t, B_t)dt. \tag{13}$$

The mathematically meaningful form of the Itô formula is (5). This is because we have precise definitions for all terms appearing in the right-hand side: the first and the third terms are ordinary Riemann integrals, while the second term is an Itô integral we defined last time.

Example 1. Last lecture we applied the definition of the Itô integral to find that

$$\int_0^T B_t dB_t = \frac{B_T^2}{2} - \frac{T}{2}. \tag{14}$$

Let us show how Itô formula simplifies the computation of Itô integrals. For example, with $f(t, x) = \frac{1}{2}x^2$ this formula says that

$$df(t, B_t) = f_t dt + f_x dB_t + \frac{1}{2}f_{xx}dt = 0 \cdot dt + B_t dB_t + \frac{1}{2} \cdot 1 \cdot dt = B_t dt + \frac{1}{2}dt. \tag{15}$$

Integrating we further obtain

$$\begin{aligned}\int_0^t d\left(\frac{1}{2}B_s^2\right) &= \int_0^t \left(B_s dB_s + \frac{1}{2}ds\right), \\ \frac{1}{2}B_t^2 - \frac{1}{2}B_0^2 &= \int_0^t B_s dB_s + \frac{t}{2}, \\ \frac{1}{2}B_t^2 - \frac{t}{2} &= \int_0^t B_s dB_s. \qquad (16)\end{aligned}$$

Example 2. see p. 55, [4]. Let $\beta_k(t) = E\left[B_t^k\right]$. We will use Itô formula to prove recursive relation

$$\beta_k(t) = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)\,ds \text{ for } k \geq 2. \tag{17}$$

Note that we can choose: $f(t, B_t) = B_t^k$. Then

$$df(t, B_t) = 0 \cdot dt + kB_t^{k-1}dB_t + \frac{1}{2}k(k-1)B_t^{k-2}dt. \tag{18}$$

Integrating

$$\int_0^t df_(s, B_s) = f(t, B_t) - f(0, B_0) = B_t^k = \int_0^t kB_t^{k-1}dB_t + \frac{1}{2}k(k-1)\int_0^t B_t^{k-2}dt. \tag{19}$$

Note that for $\mathcal{F}_t$-adapted functions f

$$\mathbb{E}\int_0^t f(s, B_s)\,dB_s = 0, \tag{20}$$

hence,

$$\mathbb{E}\int_0^t kB_t^{k-1}dB_t = 0. \tag{21}$$

We are left with $\mathbb{E}\left[\frac{1}{2}k(k-1)\int_0^t B_s^{k-2}ds\right] = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)dt$.

We can simply obtain that

$$\mathbb{E}\left(B_t^4\right) = 3t^2. \tag{22}$$

Indeed

$$\mathbb{E}\left(B_t^4\right) = \beta_t^4 = \frac{1}{2} \cdot 4 \cdot 3 \int_0^t sds = 3t^2, \tag{23}$$

because $\int_0^t sds = \frac{t^2}{2}$.

$$\mathbb{E}\left(B_t^{2k+1}\right) = 0, \forall k, \mathbb{E}\left(B_t^{2k}\right) = \frac{(2k)!t^k}{2^k k!}, \mathbb{E}\left(B_t^6\right) = 15t^3. \tag{24}$$

Itô formula for general diffusion. We extend the Itô formula to stochastic processes more general than Brownian Motion.

Definition 3. *An Itô process is a stochastic process X_t on $(\Omega, \mathcal{F}, \mathbb{P})$ of the form:*

$$X_t = X_0 + \int_0^t \mu(s, \omega)\,ds + \int_0^t \nu(s, \omega)\,dB_s, \tag{25}$$

where $\nu \in V$, $\mathbb{P}\left(\int_0^t \nu^2(s, \omega)\,ds < \infty, \forall t \geq 0\right) = 1$. We also assume that μ is $\mathcal{F}_t$-adapted and $\mathbb{P}\left(\int_0^t |\mu(s, \omega)|\,ds < \infty, \forall t \geq 0\right) = 1$

Equation (25) is often written as

$$dX_t = \mu dt + \nu dB_t, \tag{26}$$

and μdt is called drift, νdB_t is called a volatile part.

Theorem 4 (Itô's Lemma)**.** *Let X_t be an Itô process given by* (25)*. Let $g(t, X) \in C^2([0, \infty) \times \mathbb{R})$, i. e., g is twice continuously differentiable on $[0, \infty) \times \mathbb{R}$. Then $Y_t = g(t, X_t)$ is again an Itô process and*

$$dY_t = g_t'(t, X_t)\,dt + g_x'(t, X_t)\,dX_t + \frac{1}{2}g_{xx}''(t, X_t)\,(dX_t)^2\,, \tag{27}$$

where $(dX_t)^2$ is calculated according to the following rules

	dt	dB_t
dt	0	0
dB_t	0	dt

Thus $(dX_t)^2 = (\mu dt + \nu dB_t)^2 = \mu^2 (dt)^2 + \nu^2 (dB_t)^2 + 2\mu\nu dt dB_t = \nu^2 dt$.

Notation: $g_t' = \frac{\partial g}{\partial t}$, $g_x' = \frac{\partial g}{\partial X}$.
Substituting dX_t

$$\begin{aligned} dY_t &= g_t' dt + g_x'(\mu dt + \nu dB_t) + \frac{1}{2}g_{xx}''\nu^2 dt \\ &= \left(g_t' + g_x'\mu + \frac{1}{2}g_{xx}''\nu^2\right)dt + g_x'\nu dB_t, \end{aligned} \tag{28}$$

with the first term representing new drift and the second – new volatility term.

Example 3.[Variance Swaps] An investor that recognizes uncertainty about future realized variance as a risk factor distinct from the risk associated to the level of asset price might want to modify his exposure with respect to variance risk. This is possible by trading in a simple contract that depends exclusively on variance risk: the variance swap. In this contract, two parties agree at $t = 0$ to exchange a dollar amount proportional to the variance realized by a reference asset between 0 and T, against a fixed sum set at inception. The valuation problem associated to a variance swap is the computation of the fixed payment that makes the contract worthless at inception. More formally, variance swap payoff at maturity T given by the normalized empirical variance

$$\frac{251}{T}\sum_{i=0}^{n}(S_{i+1} - S_i)^2.$$

In order to price variance swap in the case when S_t is a traded security one applies Itô lemma

$$d(S_t - S_0)^2 = 2(S_t - S_0)dS_t + (dS_t)^2.$$

Integrating from 0 to T one gets

$$(S_T - S_0)^2 = 2\int_0^T (S_t - S_0)dS_t + \int_0^T (dS_t)^2.$$

Thus realized variance is represented in as the terminal gain of a portfolio composed by two pieces: a static position on a quadratic contract with final payoff $(S_T - S_0)^2$, and the gains of a dynamic position in S, with notional $-2(S_t - S_0)$.

Stochastic Differential Equations. General SDE:

$$X_t = X_0 + \int_0^t b(s,w)\,ds + \int_0^t \sigma(s,w)\,dB_s \tag{29}$$

Why stochastic? Because $\int_0^t \sigma(s,w)\,dB_s$ is a random component, also $b = b(s,w)$ can be a random function. The solution to an SDE is not a function, but a family of functions. Short form:

$$dX_t = bdt + \sigma dB_t, X(0) = X_0.$$

Black-Scholes SDE.

$$dS_t = rS_t dt + \sigma S_t dB_t, \tag{30}$$

initial condition $S(0) = S_0$, where the term $\sigma S_t dB_t$ represents uncertainty.

- r – risk-free interest rate (assumed constant)
- σ – volatility (also assumed constant)
- B_t – standard Brownian motion
- S_t – stock price

Consider the ordinary stochastic differential equation without uncertainty:

$$dS_t = rS_t dt, \text{initial condition } S(0) = S_0.$$

Dividing both sides by S_t yields:

$$\frac{dS_t}{S_t} = rdt \Longrightarrow d(\log S_t) = \frac{dS_t}{S_t} \Longrightarrow d(\log S_t) = rdt.$$

Integrating both sides from 0 to t, we have

$$\log S_t - \log S_0 = rt, \Longrightarrow S_t = S_0 \exp(rt).$$

In other words, investing \$S_0 at time 0 results in $\$S_0 \exp(rt)$at time t.

Thus, Black-Scholes is a formula for return on risk-free investment plus uncertainty.

It formula: $dg(t, S_t) = g_t' dt + g_s' ds + \frac{1}{2} g_{ss}'' (dt)^2$

$$d\log S_t = \frac{dS_t}{S_t} - \frac{1}{2}\frac{(dS_t)^2}{S_t^2}$$

$$(dS_t)^2 = r^2 S_t^2 (dt)^2 + 2r\sigma S_t^2 dt dB_t + \sigma^2 S_t^2 (dB_t)^2 = \sigma^2 S_t^2 (dB_t)^2$$

$$d\log S_t = rdt + \sigma dB_t - \frac{1}{2}\sigma^2 dt = \left(r - \frac{\sigma^2}{2}\right) dt + \sigma dB_t$$

$$\log S_t - \log S_0 = \left(r - \frac{\sigma^2}{2}\right) t + \sigma B_t$$

$$S_t = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right) t + \sigma B_t\right) \tag{31}$$

The solution to Black-Scholes implies that $S_t \geq 0$, as long as $S_0 \geq 0$. Result (31) allows to calculate the prices for vanilla European options.

In particular, the call price $= e^{-rt}\mathbb{E}(S_t - K)^+ = S_0 N(d_1) - K\exp(-rt) N(d_2)$

$$d_1 = \frac{\left(\log \frac{S_0}{K} + rt\right) + \frac{\sigma^2 t}{2}}{\sigma\sqrt{t}}, d_2 = d_1 - \sigma\sqrt{t}, N\left(x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

Ornstein-Uhlenbeck SDE.

$$dX_t = \left(\beta - X_t\right) dt + \sigma dB_t, X\left(0\right) = X_0$$

$$dX_t = \beta dt - X_t dt + \sigma dB_t$$

$$dX_t + X_t dt = \beta dt + \sigma dB_t$$

Note that, for$g\left(t,x\right) = e^t X_t$, $d\left(e^t X_t\right) = e^t X_t dt + e^t dX_t + 0$,$d\left(e^t X_t\right) = e^t\left(X_t dt + dX_t\right)$ Thus, $d\left(e^t X_t\right) = \beta e^t dt + \sigma e^t dB_t$

$$e^t X_t - e^0 X_0 = \beta \int_0^t e^s ds + \sigma \int_0^t e^s dB_s$$

$$e^t X_t - X_0 = \beta\left(e^t - 1\right) + \sigma \int_0^t e^s dB_s$$

$$X_t = X_0 e^{-t} + \beta\left(1 - e^{-t}\right) + \sigma e^{-t} \int_0^t e^s dB_s$$

It implies that

$$\mathbb{E}\left[X_t\right] = X_0 e^{-t} + \beta\left(1 - e^{-t}\right)$$

because $\mathbb{E} \int_0^t e^s dB_s = 0$. Clearly, X_t is normally distributed, as long as the function in the integrand is deterministic, because that makes X_t the sum of normally distributed random variables

$$\int_0^t f\left(s\right) dB_s \approx \sum_{i=0}^{n-1} f\left(S_i\right)\left(B\left(S_{i+1}\right) - B\left(S_i\right)\right)$$

As $t \to \infty$ $\mathbb{E}[X_t] \to \beta$. Therefore, this process is called mean reverting.

References

[1] Steven Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*

[2] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*

[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

[4] Bernt Oksendal, Stochastic Differential Equations

Stochastic Calculus, Summer 2022
Lecture 7
Girsanov's Theorem and Applications

Reading for this lecture:

(1) [3] pp. 190-198

Motivation. To motivate the results of this lecture let us consider independent normal random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that $\mathbb{E}Z_i = 0, \mathbb{E}Z_i^2 = 1$. Given a vector $(\mu_1, \mu_2, \ldots, \mu_n) \in \mathbb{R}^n$ we consider a new probability measure $\widetilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ given by

$$\widetilde{\mathbb{P}}(dw) = e^{\sum\limits_{i=1}^{n} \mu_i Z_i(w) - \frac{1}{2}\sum\limits_{i=1}^{n} \mu_i^2} \mathbb{P}(dw). \tag{1}$$

Then $\widetilde{\mathbb{P}}(Z_1 \in dz_1, \ldots, Z_n \in dz_n)$ is given by

$$\begin{aligned}
&e^{\sum\limits_{i=1}^{n} \mu_i z_i - \frac{1}{2}\sum\limits_{i=1}^{n} \mu_i^2} \mathbb{P}(Z_1 \in dz_1, \ldots, Z_n \in dz_n) \\
&= e^{\sum\limits_{i=1}^{n} \mu_i z_i - \frac{1}{2}\sum\limits_{i=1}^{n} \mu_i^2} \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} z_i^2} dz_1 \ldots dz_n \\
&= \frac{1}{(2\pi)^{n/2}} e^{\sum\limits_{i=1}^{n} \mu_i z_i - \frac{1}{2}\mu_i^2 - \frac{1}{2} z_i^2} dz_1 \ldots dz_n \\
&= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} (z_i - \mu_i)^2} dz_1 \ldots dz_n.
\end{aligned} \tag{2}$$

Therefore under the new measure $\widetilde{\mathbb{P}}$ random variables $Z_1, \ldots, Z_n$ are independent normally distributed with $\widetilde{\mathbb{E}}Z_i = \mu_i$ and $\widetilde{\mathbb{E}}(Z_i - \mu_i)^2 = 1$. In other words, $\widetilde{Z}_i = Z_i - \mu_i, 1 \leq i \leq n$ are independent standard normal random variables on $(\Omega, \mathcal{F}, \widetilde{\mathbb{P}})$.

Girsanov's Theorem. Girsanov's Theorem extends this idea from the discrete to the continuous setting. Rather than beginning with an n-dimensional vector $(Z_1, \ldots, Z_n)$ of independent standard normal random variables, we begin with a Brownian motion under measure $\mathbb{P}$(one dimensional for simplicity) and then construct a new measure $\widetilde{\mathbb{P}}$ under which a "translated" process is a Brownian motion.

So assume that $B_t, t \geq 0$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is a Brownian motion starting from zero. Now define for a given constant γ

$$Z_t = e^{\int\limits_0^t \gamma dB_s - \frac{1}{2}\int\limits_0^t \gamma^2 ds} = e^{\gamma B_t - \frac{1}{2}\gamma^2 t} \tag{3}$$

and for each $T > 0$ define a new probability measure $\widetilde{\mathbb{P}}_T$ on $\mathcal{F}_T$(i.e. on the σ-algebra of the events known at time T) by

$$\widetilde{\mathbb{P}}_T = \mathbb{E}(\mathbf{1}_A Z_T) \quad \text{for} \quad A \in \mathcal{F}_T. \tag{4}$$

Theorem 1 (Girsanov, Cameron-Martin)**.** *Define a process $\widetilde{B}_t = B_t - \gamma t$. Then for each fixed $T < \infty$ the process $\widetilde{B}_t$ is a Brownian motion on $(\Omega, \mathcal{F}_T, \widetilde{\mathbb{P}}_T)$.*

Proof. To prove the theorem we need to prove the following two lemmas.

Lemma 2. *Z_t is a martingale under measure $\mathbb{P}$.*

Proof of Lemma 2. For every $s < t$ we clearly have

$$\begin{aligned}
\mathbb{E}(Z_t|\mathcal{F}_s) &= \mathbb{E}(e^{\gamma B_t - \gamma^2 t/2}|\mathcal{F}_s) \\
&= \mathbb{E}(e^{\gamma(B_t - B_s) - \gamma^2(t-s)/2 + \gamma B_s - \gamma^2 s/2}|\mathcal{F}_s) \\
&= e^{\gamma B_s - \gamma^2 s/2}\mathbb{E}(e^{\gamma(B_t - B_s) - \gamma^2(t-s)/2}|\mathcal{F}_s) \\
&= Z_s \mathbb{E}(e^{\gamma(B_t - B_s) - \gamma^2(t-s)/2}) = Z_s.
\end{aligned}$$

$\square$

Lemma 3. *Measures $\widetilde{\mathbb{P}}_T$ are consistent, that is for $t < T$ and $A \in \mathcal{F}_t$*

$$\widetilde{\mathbb{P}}_T(A) = \widetilde{\mathbb{P}}_t(A). \tag{5}$$

Proof of Lemma 3. Indeed, using the tower property of conditional expectation

$$\begin{aligned}
\widetilde{\mathbb{P}}_T(A) &= \widetilde{\mathbb{E}}_T(\mathbf{1}_A) = \mathbb{E}(Z_T \mathbf{1}_A) = \mathbb{E}(\mathbb{E}(Z_T \mathbf{1}_A|\mathcal{F}_t)) \\
&= \mathbb{E}(\mathbf{1}_A \mathbb{E}(Z_T|\mathcal{F}_t)) = \mathbb{E}(\mathbf{1}_A Z_t) = \widetilde{\mathbb{P}}_t(A).
\end{aligned} \tag{6}$$

$\square$

Fix a sequence $t_1 \le t_2 \le \cdots \le t_n$ and consider Brownian increments $B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$. By the definition of $\widetilde{\mathbb{P}}_T$

$$\begin{aligned}
&\widetilde{\mathbb{P}}_T\left[B(t_1) \in dz_1, B(t_2) - B(t_1) \in dz_2, \ldots, B(t_n) - B(t_{n-1}) \in dz_n\right] \\
&= e^{\gamma B(t_n) - \gamma^2 t_n/2}\mathbb{P}_T\left[B(t_1) \in dz_1, B(t_2) - B(t_1) \in dz_2, \ldots, B(t_n) - B(t_{n-1}) \in dz_n\right] \\
&= e^{\gamma(z_1 + \cdots + z_n) - \gamma^2 t_n/2}\frac{1}{(\sqrt{2\pi})^n}e^{-z_1^2/(2t_1) - z_2^2/(2(t_2 - t_1)) - \cdots - z_n^2/(2(t_n - t_{n-1}))} \\
&= \frac{1}{(\sqrt{2\pi})^n}e^{(\gamma z_1 - \gamma^2 t_1/2 - z_1^2/(2t_1)) + \cdots + (\gamma z_n - \gamma^2(t_n - t_{n-1})/2 - z_n^2/(2(t_n - t_{n-1})))} \\
&= \frac{1}{(\sqrt{2\pi})^n}e^{-(z_1 - \gamma t_1)^2/(2t_1) - \cdots - (z_n - \gamma(t_n - t_{n-1}))^2/(2(t_n - t_{n-1}))}.
\end{aligned} \tag{7}$$

Thus under $\widetilde{\mathbb{P}}$ random variables $B(t_1) - \gamma t_1, B(t_2) - B(t_1) - \gamma(t_2 - t_1), \ldots, B(t_n) - B(t_{n-1}) - \gamma(t_n - t_{n-1})$ are independent normal random variables with expectation zero and variances $t_1, t_2 - t_1, \ldots, t_n - t_{n-1}$. So we proved that the finite dimensional distributions of $\widetilde{B}_t = B_t - \gamma t$ under $\widetilde{\mathbb{P}}$ coincide with the finite dimensional distributions of Brownian motion. $\square$

In fact, a more general theorem is true. We may assume that $\gamma = \gamma(t, w)$ is an adapted process. Define

$$Z_t = e^{\int_0^t \gamma(s,w)dB_s(w) - 1/2\int_0^t \gamma^2(s,w)ds}. \tag{8}$$

Then $\widetilde{B}_t = B_t - \int_0^t \gamma(s, w)ds$ is a Brownian motion in the new measure.

Importance Sampling. The first application of the Girsanov theorem ideas is the importance sampling method in the Monte Carlo scheme. As a simple example, consider the problem of estimating $\mathbb{P}(X > a)$ when $X \sim N(0,1)$ is a standard normal random variable and a is large. The naive Monte Carlo method would be to generate N sample standard normals, $X_k, k = 1, \dots, N$, and take

$$\begin{cases} X_k \sim N(0,1), k = 1, \dots, N \\ A = \mathbb{P}(X > a) \approx \frac{1}{N} \# X_k > a = \frac{1}{N} \sum\limits_{k=0}^{N} \mathbf{1}_{X_k > a}. \end{cases} \tag{9}$$

For large a the hits, $X_k > a$, would be a small fraction f the sample with the rest being wasted. Importance sampling strategy consists in changing the measure to $\widetilde{\mathbb{P}}$ such that under this new measure $\widetilde{\mathbb{P}}$ random variable X has mean a, i.e. $X = N(a,1)$. Actually, we know that for

$$\widetilde{\mathbb{P}}(X \in dx) = e^{ax - a^2/2} \mathbb{P}(X \in dx) \tag{10}$$

we have $X \sim N(a,1)$ under $\widetilde{\mathbb{P}}$. Therefore,

$$\mathbb{P}(X \in dx) = e^{-ax + a^2/2} \widetilde{\mathbb{P}}(X \in dx) \tag{11}$$

and the importance sampling estimate is

$$\begin{cases} X_k \sim N(a,1), k = 1, \dots, N \\ A \approx \frac{1}{N} e^{a^2/2} \sum\limits_{k=0}^{N} e^{-aX_k} \mathbf{1}_{X_k > a}. \end{cases} \tag{12}$$

A simple way to generate $N(a,1)$ random variable is to start with $Y_k \sim N(0,1)$ and add $a : X_k = Y_k + a$. In this form, $e^{a^2/2} e^{-aX_k} = e^{-a^2/2} e^{-aY_k}$ and $X_k > a$ is equivalent to $Y_k > 0$ so the variance reduction estimator becomes

$$\begin{cases} Y_k \sim N(0,1), k = 1, \dots, N \\ A \approx \frac{1}{N} e^{-a^2/2} \sum\limits_{k=0}^{N} e^{-aY_k} \mathbf{1}_{Y_k > 0}. \end{cases} \tag{13}$$

Thus, the naive Monte Carlo method produces a small A be getting a small number of hits in many trials. The importance sampling gets roughly 50 % hits but discounts each hit by a factor of at least $e^{-a^2/2}$ to get the same expected value as the naive estimator.

Brownian Motion with drift. Consider a Brownian motion $B_t, t \geq 0$ and recall that the passage time T_b to the level $b \neq 0$ has density

$$\mathbb{P}(T_b \in dt) = \frac{|b|}{\sqrt{2\pi t^3}} e^{-b^2/(2t)}, t > 0. \tag{14}$$

From Girsanov's theorem the process $\widetilde{B}_t = B_t - \mu t$ is a Brownian motion under the unique probability measure $\widetilde{\mathbb{P}}$ which satisfies

$$\widetilde{\mathbb{P}}(A) = \mathbb{E}(\mathbf{1}_A Z_t) \quad \text{for} \quad A \in \mathcal{F}_t. \tag{15}$$

We therefore say that $B_t = \mu t + \widetilde{B}_t$ is a Brownian motion with drift μ under $\widetilde{\mathbb{P}}$. By the definition of $\widetilde{\mathbb{P}}$ and the tower property of conditional expectation

$$\begin{aligned}\widetilde{\mathbb{P}}(T_b < t) &= \mathbb{E}(\mathbf{1}_{T_b \le t} Z_t) = \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_{T_b \le t} Z_t | \mathcal{F}_{T_b}\right)\right) \\ &= \mathbb{E}\left(\mathbf{1}_{T_b \le t} \mathbb{E}\left(Z_t | \mathcal{F}_{T_b}\right)\right) = \mathbb{E}\left(\mathbf{1}_{T_b \le t} Z_{T_b}\right) \\ &= \mathbb{E}(\mathbf{1}_{T_b \le t} e^{\mu b - \frac{1}{2}\mu^2 T_b}) = \int_0^t e^{\mu b - \frac{1}{2}\mu^2 s} \mathbb{P}(T_b \in ds). \end{aligned} \tag{16}$$

It implies that the density of T_b under $\widetilde{\mathbb{P}}$ is

$$\widetilde{\mathbb{P}}(T_b \in dt) = \frac{|b|}{\sqrt{2\pi t^3}} e^{-(b-\mu t)^2/(2t)}, t > 0. \tag{17}$$

References

[3] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

Stochastic Calculus, Summer 2022, July 20,
Lecture 8
Connection of Stochastic Calculus and Partial Differential Equations

Reading for this lecture:

(1) [1] pp. 125-175
(2) [2] pp. 239-280
(3) Professor R. Kohn's lecture notes PDE for Finance, in particular Lecture 1
`http://www.math.nyu.edu/faculty/kohn/pde_finance.html`

Today throughout the lecture we will be using the following lemma.

Lemma 1. *Assume we are given a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t\geq 0}$. Then $\mathbb{E}(X|\mathcal{F}_t)$ is a martingale with respect to filtration $(\mathcal{F}_t)_{t\geq 0}$.*

Proof. The proof is very easy and follows from the tower property of the conditional expectation. ☐

Corollary 2. *Let X_t be a Markov process and $\mathcal{F}_t$ be the natural filtration associated with this process. Then according to the above lemma for any function V process $\mathbb{E}(V(X_T)|\mathcal{F}_t)$ is a martingale and applying Markov property we get that $\mathbb{E}(V(X_T)|X_t)$ is a martingale. In the following we often write $\mathbb{E}(V(X_T)|X_t)$ as $\mathbb{E}_{X_t=x}V(X_T)$.*

As we will see this corollary together with Itô's formula yield some powerful results on the connection of partial differential equations and stochastic calculus.

Expected value of payoff $V(X_T)$. Assume that X_t is a stochastic process satisfying the following stochastic differential equation

$$dX_t = a(t, X_t)dt + \sigma(t, X_t)dB_t, \tag{1}$$

or in the integral form

$$X_t - X_0 = \int_0^t a(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s. \tag{2}$$

Let

$$u(t, x) = \mathbb{E}_{X_t=x}V(X_T) \tag{3}$$

be the expected value of some payoff V at maturity $T > t$ given that $X_t = x$. Then $u(t, x)$ solves

$$u_t + a(t, x)u_x + \frac{1}{2}(\sigma(t, x))^2 u_{xx} = 0 \text{ for } t < T, \text{ with } u(T, x) = V(x). \tag{4}$$

By Corollary 2 we conclude that $u(t, x)$ defined by (3) is a martingale. Applying Itô's lemma we obtain

$$\begin{aligned} du(t, X_t) &= u_t dt + u_x dX_t + \frac{1}{2}u_{xx}(dX_t)^2 \\ &= u_t dt + u_x(adt + \sigma dB_t) + \frac{1}{2}u_{xx}\sigma^2 dt \\ &= (u_t + au_x + \frac{1}{2}\sigma^2 u_{xx})dt + \sigma u_x dB_t, \end{aligned} \tag{5}$$

[1]this version July 20, 2022

Since $u(t,x)$ is a martingale the drift term must be zero and thus $u(t,x)$ solves

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0.$$

Substituting $t = T$ is (3) we get that $u(T,x) = \mathbb{E}_{X_T=x}(V(X_T)) = V(x)$.

Feynman-Kac formula. Suppose that we are interested in a suitably "discounted" final-time payoff of the form

$$u(t,x) = \mathbb{E}_{X_t=x}\Big(e^{-\int_t^T b(s,X_s)ds} V(X_T)\Big) \tag{6}$$

for some specified function $b(t,X_t)$. We will show that u then solves

$$u_t + a(t,x)u_x + \frac{1}{2}\sigma^2 u_{xx} - b(t,x)u = 0 \tag{7}$$

and final-time condition $u(T,x) = V(x)$.

The fact that $u(T,x) = V(x)$ is clear from the definition of function u. Therefore let us concentrate on the proof of (7). Our strategy is to apply Corollary 2 and thus we have to find some martingale involving $u(t,x)$. For this reason let us consider

$$\begin{aligned} e^{-\int_0^t b(s,X_s)ds} u(t,x) &= e^{-\int_0^t b(s,X_s)ds} \mathbb{E}_{X_t=x}\Big(e^{-\int_t^T b(s,X_s)ds} V(X_T)\Big) \\ &= \mathbb{E}_{X_t=x}\Big(e^{-\int_0^T b(s,X_s)ds} V(X_T)\Big). \end{aligned} \tag{8}$$

According to Corollary 2

$$\mathbb{E}_{X_t=x}\Big(e^{-\int_0^T b(s,X_s)ds} V(X_T)\Big)$$

is a martingale and thus $e^{-\int_0^t b(s,X_s)ds} u(t,x)$ is a martingale. Applying Itô's lemma we get

$$\begin{aligned} d\Big(e^{-\int_0^t b(s,X_s)ds} u(t,x)\Big) &= (u_t + a(t,x)u_x + \frac{1}{2}\sigma^2 u_{xx} - b(t,x)u)e^{-\int_0^t b(s,X_s)ds} dt \\ &\quad + e^{-\int_0^t b(s,X_s)ds} u_x dB_t. \end{aligned} \tag{9}$$

Since the drift must be equal to zero we obtain that $u(t,x)$ satisfies (7).

Example. [Black-Scholes PDE] We assume that the underlying (stock, for instance) follows a geometric Brownian motion. That is in the risk-neutral measure it satisfies SDE

$$dS_t = rS_t dt + \sigma S_t dB_t, \tag{10}$$

where r is the risk-free rate which we assume to be constant. The payoff of a European option at maturity T is known and is equal to $V(S_T)$. Then to find the value of the option at some earlier time $t < T$ we have to compute

$$\mathbb{E}_{S_t=x}\left(e^{-rt}V(S_T)\right). \tag{11}$$

From the Feynman-Kac formula we conclude that $u(t,x)$ solves partial differential equation

$$u_t + rxu_x + \frac{1}{2}\sigma^2 x^2 u_{xx} - ru = 0 \tag{12}$$

with $u(T,x) = V(x)$. Equation (12) is the famous known Black-Scholes PDE.

Running payoff. Now suppose that we are interested in

$$u(t,x) = \mathbb{E}_{X_t=x}\Big(\int\limits_t^T b(s,X_s)ds\Big) \tag{13}$$

for some specified function $b(t,x)$. First of all, let us find the final-time condition for $u(t,x)$. Clearly, $u(T,x) = 0$. Our next step is to find a martingale involving $u(t,x)$ and then use Corollary 2. Add to both parts of (13) $\int\limits_0^t b(s,X_s)ds$. Then

$$u(t,x) + \int\limits_0^t b(s,X_s)ds = \mathbb{E}_{X_t=x}\Big(\int\limits_0^T b(s,X_s)ds\Big) \tag{14}$$

is a martingale. Applying Itô's lemma we obtain that $u(t,x)$ satisfies

$$u_t + a(t,x)u_x + \frac{1}{2}\sigma^2 u_{xx} + c(t,x) = 0. \tag{15}$$

Boundary value problems and exit times. In previous examples we were interested in the expectation of form (3), that the expectation of some payoff at specified maturity T.

Now let us assume that we are given a region $D \subset \mathbb{R}$ and process X_t starts from some point $x \in D$. Let

$$\tau(x) = \min\{T, \inf\{t : X_t \notin D\}\}.$$

That is $\tau(x)$ is the first time X_t exits from region D if prior to T, otherwise $\tau = T$. Assume at exit time τ the payoff of an option is given by function V. We are interested in the fair price of such an option at some earlier time t, i.e., in the following quantity

$$u(t,x) = \mathbb{E}_{X_t=x}V(\tau, X_\tau). \tag{16}$$

We will see that just like in the previous examples $u(t,x)$ solves partial differential equation, but in contrast the PDE must be solved inside the region D with suitable boundary data. The key to derivation of the PDE is the following lemma.

Lemma 3. *Let $\mathbb{E}\tau < \infty$. Then $\mathbb{E}(V(X_\tau)|\mathcal{F}_t)$ is a martingale with respect to filtration $(\mathcal{F}_{t\wedge\tau})_{t\geq 0}$.*

Applying Itô's lemma we get

$$du(t,x) = (u_t + au_x + \frac{1}{2}\sigma^2 u_{xx})dt + u_x\sigma dB_t. \tag{17}$$

By lemma 3 function $u(t,x)$ from (16) is a martingales and therefore there is no drift term in (17). Thus $u(t,x)$ solves the following PDE

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0 \tag{18}$$

with boundary conditions $u(t,x) = V(t,x)$ for $x \in \partial D$ and $u(t,x) = V(T,x)$ for $x \in D$.

Application: distribution of the first arrivals. As the first application let us consider $\mathbb{E}_{X_t=x}\mathbf{1}_{\tau<T} = \mathbb{P}_{X_t=x}(\tau < T)$. According to the above we have to solve PDE

$$u_t + au_x + \frac{1}{2}\sigma^2 u_{xx} = 0 \tag{19}$$

with boundary condition $u = 1$ at $x \in \partial D$.

References

[1] Richard Durrett, *Stochastic Calculus: A Practical Introduction.*
[2] Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus.*

PDE for Finance Notes – Section 1.

Notes by Robert V. Kohn, Courant Institute of Mathematical Sciences. For use in connection with the NYU course PDE for Finance, MATH-GA2706. Prepared in 2003, minor updates made in 2011 and 2014.

Links between stochastic differential equations and PDE. A stochastic differential equation, together with its initial condition, determines a diffusion process. We can use it to define a deterministic function of space and time in two fundamentally different ways:

(a) by considering the expected value of some "payoff," as a function of the initial position and time; or

(b) by considering the probability of being in a certain state at a given time, given knowledge of the initial state and time.

Students of finance will be familiar with the Black-Scholes PDE, which amounts to an example of (a). Thus in studying topic (a) we will be exploring among other things the origin of the Black-Scholes PDE. The basic mathematical ideas here are the *backward Kolmogorov equation* and the *Feynman-Kac formula*.

Viewpoint (b) is different from (a), but not unrelated. It is in fact *dual* to viewpoint (a), in a sense that we will make precise. The evolving probability density solves a different PDE, the *forward Kolmogorov equation* – which is actually the adjoint of the backward Kolmogorov equation.

It is of interest to consider how and when a diffusion process crosses a barrier. This arises in thinking subjectively about stock prices (e.g. what is the probability that IBM will reach 200 at least once in the coming year?). It is also crucial for pricing barrier options. Probabilistically, thinking about barriers means considering *exit times*. On the PDE side this will lead us to consider *boundary value problems* for the backward and forward Kolmogorov equations.

For a fairly accessible treatment of much of this material see Gardiner (the chapter on the Fokker-Planck Equation). Parts of my notes draw from Oksendal, however the treatment there is much more general and sophisticated so not easy to read.

Our main tool will be Ito's formula, coupled with the fact that any Ito integral of the form $\int_a^b f\,dw$ has expected value zero. (Equivalently: $m(t) = \int_a^t f\,dw$ is a martingale.) Here w is Brownian motion and f is non-anticipating. The stochastic integral is defined as the limit of Ito sums $\sum_i f(t_i)(w(t_{i+1} - w(t_i))$ as $\Delta t \to 0$. The sum has expected value zero because each of its terms does: $E[f(t_i)(w(t_{i+1}) - w(t_i))] = E[f(t_i)]E[w(t_{i+1}) - w(t_i)] = 0$.

****

Expected values and the backward Kolmogorov equation. Here's the most basic version of the story. Suppose $y(t)$ solves the scalar stochastic differential equation

$$dy = f(y, s)ds + g(y, s)dw,$$

and let

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(T))\right]$$

be the expected value of some payoff Φ at maturity time $T > t$, given that $y(t) = x$. Then u solves

$$u_t + f(x,t)u_x + \tfrac{1}{2}g^2(x,t)u_{xx} = 0 \text{ for } t < T, \text{ with } u(x,T) = \Phi(x). \tag{1}$$

The proof is easy: for any function $\phi(y,t)$, Ito's lemma gives

$$\begin{aligned} d(\phi(y(s),s)) &= \phi_y dy + \tfrac{1}{2}\phi_{yy} dy dy + \phi_s ds \\ &= (\phi_s + f\phi_y + \tfrac{1}{2}g^2\phi_{yy})dt + g\phi_y dw. \end{aligned}$$

Choosing $\phi = u$, the solution of (1), we get

$$u(y(T),T) - u(y(t),t) = \int_t^T (u_t + fu_y + \tfrac{1}{2}g^2 u_{yy})ds + \int_t^T g u_y dw.$$

Taking the expected value and using the PDE gives

$$E_{y(t)=x}\left[\Phi(y(T))\right] - u(x,t) = 0$$

which is precisely our assertion.

That was the simplest case. It can be jazzed up in many ways. We discuss some of them:

Vector-valued diffusion. Suppose y solves a vector-valued stochastic differential equation

$$dy_i = f_i(y,s)ds + \sum_j g_{ij}(y,s)dw_j,$$

where each component of w is an independent Brownian motion. Then

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(T))\right]$$

solves

$$u_t + \mathcal{L}u = 0 \text{ for } t < T, \text{ with } u(x,T) = \Phi(x),$$

where $\mathcal{L}$ is the differential operator

$$\mathcal{L}u(x,t) = \sum_i f_i \frac{\partial u}{\partial x_i} + \tfrac{1}{2}\sum_{i,j,k} g_{ik}g_{jk}\frac{\partial^2 u}{\partial x_i \partial x_j}.$$

The justification is just as in the scalar case, using the multidimensional version of Ito's lemma. The operator $\mathcal{L}$ is called the "infinitesimal generator" of the diffusion process $y(s)$.

The Feynman-Kac formula. We discuss the scalar case first, for clarity. Consider as above the solution of

$$dy = f(y,s)ds + g(y,s)dw$$

but suppose we are interested in a suitably "discounted" final-time payoff of the form:

$$u(x,t) = E_{y(t)=x}\left[e^{-\int_t^T b(y(s),s)ds}\Phi(y(T))\right] \tag{2}$$

for some specified function $b(y)$. Then u solves

$$u_t + f(x,t)u_x + \tfrac{1}{2}g^2(x,t)u_{xx} - b(x,t)u = 0 \tag{3}$$

instead of (1). (Its final-time condition is unchanged: $u(x,T) = \Phi(x)$.) If you know some finance you'll recognize that when y is log-normal and b is the interest rate, (3) is precisely the Black-Scholes partial differential equation. Also: if $b(y(s),s)$ is the spot interest rate, then (3) with $\Phi = 1$ gives the time-t value of a zero-coupon bond with maturity T, given that the spot interest rate at time t is $b(x,t)$.

To explain (3), we must calculate the stochastic differential $d[z_1(s)\phi(y(s),s)]$ where $z_1(s) = e^{-\int_t^s b(y(r),r)dr}$. The multidimensional version of Ito's lemma gives

$$d[z_1(s)z_2(s)] = z_1 dz_2 + z_2 dz_1 + dz_1 dz_2.$$

We apply this with z_1 as defined above and $z_2(s) = \phi(y(s),s)$. Ito's lemma (or ordinary differentiation) gives

$$dz_1(s) = -z_1 b(y(s),s)ds$$

and we're already familiar with the fact that

$$\begin{aligned} dz_2(s) &= (\phi_s + f\phi_y + \tfrac{1}{2}g^2\phi_{yy})ds + g\phi_y dw \\ &= (\phi_s + \mathcal{L}\phi)ds + g\phi_y dw. \end{aligned}$$

Notice that $dz_1 dz_2 = 0$. Applying the above with $\phi = u$, the solution of the PDE (3), gives

$$\begin{aligned} d\left(e^{-\int_t^s b(y(r),r)dr}u(y(s),s)\right) &= z_1 dz_2 + z_2 dz_1 \\ &= z_1\left[(u_s + \mathcal{L}u)ds + gu_y dw\right] - z_1 ubds \\ &= z_1 gu_y dw. \end{aligned}$$

The right hand side has expected value 0, so

$$E_{y(t)=x}[z_1(T)z_2(T)] = z_1(t)z_2(t) = u(x,t)$$

as asserted.

A moment's thought reveals that vector-valued case is no different. The discounted expected payoff (2) solves the PDE

$$u_t + \mathcal{L}u - bu = 0$$

where $\mathcal{L}$ is the infinitesimal generator of the diffusion y.

Running payoff. Suppose we are interested in

$$u(x,t) = E_{y(t)=x}\left[\int_t^T \Psi(y(s),s)ds\right]$$

for some specified function Ψ. Then u solves

$$u_t + \mathcal{L}u + \Psi(x,t) = 0.$$

The final-time condition is $u(x,T) = 0$, since we have included no final-time term in the "payoff." The proof is hardly different from before: by Ito's lemma,

$$\begin{aligned} d[u(y(s),s)] &= (u_t + \mathcal{L}u)ds + \nabla u \cdot g \cdot dw \\ &= -\Psi(y(s),s)ds + \nabla u \cdot g \cdot dw. \end{aligned}$$

Integrating and taking the expectation gives

$$E_{y(t)=x}\left[u(y(T),T)\right] - u(x,t) = E_{y(t)=x}\left[-\int_t^T \Psi(y(s),s)ds\right].$$

This gives the desired assertion, since $u(y(T),T) = 0$.

In valuing options, "running payoffs" are relatively rare. However terms of this type will be very common later in the course, when discuss optimal control problems.

****

Boundary value problems and exit times. The preceding examples use stochastic integration from time t to a fixed time T, and they give PDE's that must be solved for all $x \in R^n$. It is also interesting to consider integration from time t to the first time y exits from some specified region. The resulting PDE must be solved on this region, with suitable boundary data.

Let D be a region in R^n. Suppose y is an R^n-valued diffusion solving

$$dy = f(y,s)ds + g(y,s)dw \text{ for } s > t, \text{ with } y(t) = x$$

with $x \in D$. Let

$$\begin{aligned} \tau(x) &= \text{the first time } y(s) \text{ exits from } D, \text{ if} \\ & \text{prior to } T; \text{ otherwise } \tau(x) = T. \end{aligned}$$

This is an example of a *stopping time*. (Defining feature of a stopping time: the statement "$\tau(x) < t$" is $\mathcal{F}_t$-measurable; in other words, the decision whether to stop or not at time t depends only on knowledge of the process up to time t. This is clearly true of the exit time defined above.)

Here is the basic result: the function

$$u(x,t) = E_{y(t)=x}\left[\int_t^{\tau(x)} \Psi(y(s),s)ds + \Phi(y(\tau(x)),\tau(x))\right]$$

solves

$$u_t + \mathcal{L}u + \Psi = 0 \text{ for } x \in D$$

with boundary condition

$$u(x,t) = \Phi(x,t) \text{ for } x \in \partial D \tag{4}$$

and final-time condition

$$u(x,T) = \Phi(x,T) \text{ for all } x \in D. \tag{5}$$

The justification is entirely parallel to our earlier examples. The only change is that we integrate, in the final step, to the stopping time τ rather than the final time T. (This is permissible for any stopping time satisfying $E[\tau] < \infty$. The statement that $E\left[\int_t^\tau f\, dw\right] = 0$ when $E[\tau] < \infty$ is known as Dynkin's theorem.)

A subtlety is hiding here: the hypothesis that $E[\tau] < \infty$ is not a mere technicality. Rather, there are simple and interesting examples where it is false and $E\left[\int_t^\tau f\, dw\right] \neq 0$. One such example is related to the "gambler's ruin" paradox. Consider the standard Brownian process $w(s)$, starting at $w(0) = 0$. Let τ_* be the first time $w(s)$ reaches 1. Then $w(\tau_*) - w(0) = 1 - 0 = 1$ certainly does not have mean 0, so $E\left[\int_0^{\tau_*} dw\right] \neq 0$ in this case. This doesn't contradict Dynkin's theorem; it just shows that $E[\tau_*] = \infty$. To understand the situation better, consider τ_n= the time of first exit from $[-n, 1]$. You'll show on HW1 that $E[\tau_n] < \infty$ for each n, but $E[\tau_n] \to \infty$ as $n \to \infty$. The Brownian motion process eventually reaches 1 with probability one, but it may make extremely large negative excursions before doing so. Here's the coin-flipping version of this situation: consider a gambler who decides to bet by flipping coins and never quitting till he's ahead by a fixed amount. If there is no limit on the amount he is permitted to lose along the way, then he'll eventually win with probability one. But if there is a threshold of losses beyond which he must stop then there is a nonzero probability of ruin and his expected outcome is 0.

There's something slightly misleading about our notation in (4)-(5). We use the same notation Φ for both the boundary condition (4) and the final-time condition (5) because they come from the same term in the payoff: $\Phi(y(\tau), \tau)$ where τ is the time the curve $(y(s), s)$ exits from the cylinder $D \times [0, T]$. But Φ should be thought of as representing two *distinct* functions – one at the spatial boundary $\partial D \times [0, T]$, the other at the final time boundary $D \times \{T\}$ (see the figure). These two functions need have nothing to do with one

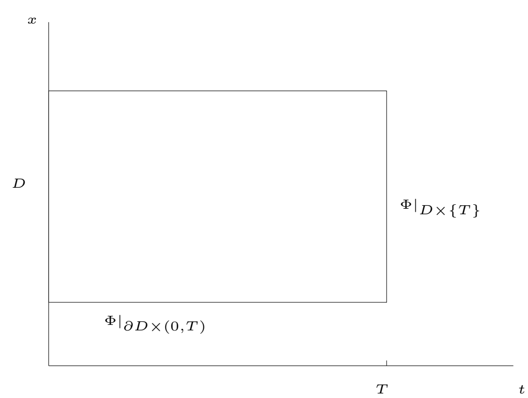

Figure 1: *Distinguishing between the two different parts of* Φ*.*

another. Often one is chosen to be zero, while the other is nontrivial. [A financial example:

when one values a barrier option using the risk-neutral expectation of the payoff, Φ is zero at the knock-out price, and it equals the payoff at the maturity time.]

Elliptic boundary-value problems. Now suppose f and g in the stochastic differential equation don't depend on t, and for $x \in D$ let

$$\tau(x) = \text{the first time } y(s) \text{ exits from } D.$$

(Unlike the previous example, we do not impose a final time T. Clearly this amounts to taking $T = \infty$ in the previous definition.) Suppose furthermore the process does eventually exit from D, (and more: assume $E[\tau(x)] < \infty$, for all $x \in D$). Then

$$u(x) = E_{y(0)=x}\left[\int_0^{\tau(x)} \Psi(y(s))ds + \Phi(y(\tau(x)))\right]$$

solves

$$\mathcal{L}u + \Psi = 0 \text{ for } x \in D,$$

with boundary condition

$$u = \Phi \text{ for } x \in \partial D.$$

The justification is again entirely parallel to our earlier examples.

****

Applications: some properties of the Brownian motion process. Let us use these results to deduce – by solving appropriate PDE's – some properties of the Brownian motion process. (This discussion is taken from Oksendal's example 7.4.2. Related material on exit times will be explored in HW1.)

QUESTION 1. Consider n-dimensional Brownian motion starting at x. What is the mean time it takes to exit from a ball of radius R, for $R > |x|$? Answer: apply the last example with $f = 0$, $g =$ identity matrix, $\Psi = 1$, $\Phi = 0$. It tells us the mean exit time is the solution $u(x)$ of

$$\tfrac{1}{2}\Delta u + 1 = 0$$

in the ball $|x| < R$, with $u = 0$ at $|x| = R$. The (unique) solution is

$$u(x) = \frac{1}{n}(R^2 - |x|^2).$$

(To do this calculation we must know in advance that the expected exit time is finite. We'll justify this as Question 3 below.)

QUESTION 2. Consider the scalar lognormal process

$$dy = \mu y dt + \sigma y dw$$

with μ and σ constant. Starting from $y(0) = x$, what is the mean exit time from a specified interval (a, b) with $a < x < b$? Answer: the mean exit time $u(x)$ solves

$$\mu x u_x + \tfrac{1}{2}\sigma^2 x^2 u_{xx} + 1 = 0 \text{ for } a < x < b$$

with boundary conditions $u(a) = u(b) = 0$. The solution is

$$u(x) = \frac{1}{\frac{1}{2}\sigma^2 - \mu}\left(\log(x/a) - \frac{1 - (x/a)^{1-2\mu/\sigma^2}}{1 - (b/a)^{1-2\mu/\sigma^2}}\log(b/a)\right)$$

(readily verified by checking the equation and boundary conditions). This answer applies only if $\mu \neq \frac{1}{2}\sigma^2$. See HW1 for the case $\mu = \frac{1}{2}\sigma^2$.

QUESTION 3: Returning to the setting of Question 1, how do we know the mean exit time is finite? Answer: assume D is a bounded domain in R^n, and $y(s)$ is multidimensional Brownian motion starting at $x \in D$. Recall that by Ito's lemma, $t \to \phi(y(t))$ satisfies

$$d\phi = \nabla\phi \, dw + \tfrac{1}{2}\Delta\phi \, dt \tag{6}$$

for any function ϕ. Let's apply this with $\phi(y) = |y|^2$, integrating in time up to the stopping time

$$\tau_T(x) = \min\{\tau(x), T\} = \begin{cases} \text{first time } y(s) \text{ exits from } D & \text{if less than } T \\ T & \text{otherwise.} \end{cases}$$

We get

$$\begin{aligned} E\left[|y(\tau_T(x))|^2\right] - |x|^2 &= \frac{1}{2}E\int_0^{\tau_T(x)} \Delta\phi(y(s))ds \\ &= nE\left[\tau_T(x)\right] \end{aligned} \tag{7}$$

since $\Delta\phi = 2n$. Now let $T \to \infty$. The left hand side of (7) stays finite, since we're considering a *bounded* domain, and by definition $y(\tau_T(x))$ is either in D or on the boundary of D. Thus we conclude that

$$\lim_{T\to\infty} E\left[\tau_T(x)\right] < \infty.$$

It follows (using the monotone convergence theorem, from real variables) that the exit time $\tau = \lim_{T\to\infty} \tau_T$ is almost surely finite, and $E[\tau] < \infty$, for any starting point $x \in D$.

QUESTION 4: Consider Brownian motion in R^n, starting at a point x with $|x| = b$. Given $r < b$, what is the probability that the path ever enters the ball of radius r centered at 0? Answer: for $n = 1, 2$ this probability is 1. (Interpretation: Brownian motion is "recurrent" in dimensions 1 and 2 – it comes arbitrarily close to any point, infinitely often, regardless of where it starts.) In higher dimensions the situation is different: in dimension $n \geq 3$ the probability of entering the ball of radius r is $(b/r)^{2-n}$. (Interpretation: Brownian motion is "transient" in dimension $n \geq 3$.)

Consider first the case $n \geq 3$. We use the stopping time τ_k = first exit time from the annulus

$$D_k = \{r < |x| < 2^k r\}.$$

Since D_k is bounded, $E[\tau_k] < \infty$ and we can integrate the stochastic differential equation (6) up to time τ_k. Let's do this with the special choice

$$\phi(y) = |y|^{2-n}.$$

This ϕ solves Laplace's equation $\Delta\phi = 0$ away from its singularity at $y = 0$. (The singularity does not bother us, since we only evaluate ϕ at points $y(s) \in D_k$ and 0 does not belong to D_k.) The analogue of (7) is

$$E\left[|y(\tau_k)|^{2-n}\right] - b^{2-n} = \frac{1}{2}\int_0^{\tau_k} \Delta\phi(y(s))ds = 0.$$

If p_k is the probability that y leaves the annulus D_k at radius r, and $q_k = 1 - p_k$ is the probability that it leaves the annulus at radius $2^k r$, we have

$$r^{2-n}p_k + (2^k r)^{2-n}q_k = b^{2-n}.$$

As $k \to \infty$ this gives $p_k \to (b/r)^{2-n}$, as asserted.

The case $n = 2$ is treated similarly, using

$$\phi(y) = \log y,$$

which solves $\Delta\phi = 0$ in the plane, away from $y = 0$. Arguing as before we get

$$p_k \log r + q_k \log(2^k r) = \log b.$$

As $k \to \infty$ this gives $q_k \to 0$. So $p_k \to 1$, as asserted.

The case $n = 1$ is similar to $n = 2$, using $\phi(y) = |y|$.

****

Another application: distribution of first arrivals. Consider a scalar diffusion whose drift and volatility are functions of y alone, independent of t:

$$dy = f(y(s))ds + g(y(s))dw.$$

The initial condition is $y(t) = x$. We are interested in the first arrival of $y(s)$ at a given threshold, say $y = 0$. Assume to fix ideas that $x > 0$.

What is the distribution of arrival times? Let the density of arrival times be $\rho(s)$. Its cumulative distribution function $\int_0^T \rho(s)\, ds$ is the probability that the first arrival occurs by time T. According to our discussion of the backward Kolmogorov equation, this is $u(x, 0)$ where u solves

$$u_t + fu_x + \tfrac{1}{2}g^2 u_{xx} = 0 \quad \text{for } x > 0, \quad 0 < t < T \tag{8}$$

with boundary condition

$$u = 1 \quad \text{at } x = 0$$

and final-time condition

$$u = 0 \quad \text{at } t = T.$$

Clearly u depends also on the final time T; let's make that dependence explicit by writing $u = u(x, t; T)$. Evidently $u(x, 0; T) = \int_0^T \rho(s)\, ds$, so by differentiation we get

$$\rho(s) = \frac{\partial u}{\partial T}(x, 0; s).$$

For special choices of f and g (e.g. Brownian motion with drift, or lognormal) the PDE (8) can be solved explicitly (we'll discuss how to do this later), yielding an explicit formula for the distribution of first arrivals.

Suppose the mean arrival time is finite.[1] Then we know it should be given by $v(x)$ where $fv_x + \frac{1}{2}g^2 v_{xx} = -1$ for $x > 0$ with $v = 0$ at $x = 0$. On the other hand, the mean arrival time is

$$\int_0^\infty s\rho(s)\, ds = \int_0^\infty s\partial_s u(x, 0; s)\, ds.$$

Are these apparently different expressions consistent? Yes indeed! To show this, we observe (writing u_s for $\partial_s u(x, 0; s)$, for simplicity) that

$$\int_0^\infty s u_s\, ds = -\int_0^\infty s(1-u)_s\, ds = \int_0^\infty (1-u)\, ds$$

by integration by parts, since[2] $\int_0^\infty \partial_s[s(1-u)]\, ds = s(1-u)|_0^\infty = 0$. Moreover the function $v(x) = \int_0^\infty (1-u)\, ds$ clearly satisfies $v = 0$ at $x = 0$, and

$$fv_x + \tfrac{1}{2}g^2 v_{xx} = -\int_0^\infty fu_x + \tfrac{1}{2}g^2 u_{xx}\, ds.$$

But since f and g are independent of time, $u(x, t; T)$ depends on t and T only through $T - t$, so $\partial u/\partial T = -\partial u/\partial t$. Therefore, using the backward Kolmogorov equation,

$$-\int_0^\infty fu_x + \tfrac{1}{2}g^2 u_{xx}\, ds = \int_0^\infty u_t\, ds = -\int_0^\infty \partial_s u(x, 0; s)\, ds.$$

The last expression is clearly $u(x, 0; 0) - u(x, 0; \infty) = 0 - 1 = -1$. Thus v solves the anticipated PDE.

[1] This is *not* true when $f = 0$ and $g = 1$ – see the discussion on page 5, and the problem related to it on HW1. Specifying conditions on f and g that make the mean arrival time finite is a nontrivial task, which we won't attempt to address here.

[2] We prefer to work with $1 - u$ rather than u, since it *vanishes* at the spatial boundary $x = 0$. Since f and g depend only on position, $u(x, t; T)$ depends only on the "elapsed time" $T - t$. As $T - t \to \infty$, we expect $w = 1 - u$ to approach a solution of the "stationary" (time-independent) problem $fw_x + \frac{1}{2}g^2 w_{xx} = 0$ for $x > 0$, $w = 0$ at $x = 0$. The obvious choice is $w = 0$; with some modest hypotheses on f and g, $w = 0$ is the only solution of this boundary value problem on the half-line $x > 0$ that remains uniformly bounded as $x \to \infty$. The "integration by parts" calculation that follows uses this fact, and a little more: it uses (more precisely, it assumes) that $s[(1 - u(x, 0, s)] \to 0$ as $s \to \infty$. This is closely related to our hypothesis that the mean arrival time is finite. We will simply assume this property holds, without attempting to specify conditions on f and g that guarantee it.

****

Transition probabilities and the forward Kolmogorov equation. We've shown that when the state evolves according to a stochastic differential equation

$$dy_i = f_i(y,s)ds + \sum_j g_{ij}(y,s)dw_j$$

the expected final position

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(T))\right]$$

solves the backward Kolmogorov equation

$$u_t + \sum_i f_i \frac{\partial u}{\partial x_i} + \tfrac{1}{2}\sum_{i,j,k} g_{ik}g_{jk}\frac{\partial^2 u}{\partial x_i \partial x_j} = 0 \text{ for } t < T, \text{ with } u = \Phi \text{ at } t = T. \tag{9}$$

We can write the backward Kolmogorov equation as

$$u_t + \mathcal{L}u = 0 \tag{10}$$

with

$$\mathcal{L}u = \sum_i f_i \frac{\partial u}{\partial x_i} + \sum_{i,j} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}, \tag{11}$$

where $a_{ij} = \frac{1}{2}\sum_k g_{ik}g_{jk} = \frac{1}{2}(gg^T)_{ij}$.

The solution of the stochastic differential equation is a *Markov process*, so it has a well-defined *transition probability*

$p(z,s;x,t) =$ probability of being at z at time s, given that it started at x at time t.

More precisely: $p(\cdot,s;x,t)$ is the probability density of the state at time s, given that it started at x at time t. Of course p is only defined for $s > t$. To describe a Markov process, p must satisfy the Chapman-Kolmogorov equation

$$p(z,s;x,t) = \int_{R^n} p(z_1,s_1;x,t)p(z,s;z_1,s_1)\,dz_1$$

for any s_1 satisfying $t < s_1 < s$. Intuitively: the state can get from (x,t) to (z,s) by way of being at various intermediate states z_1 at a chosen intermediate time s_1. The Chapman-Kolmogorov equation calculates $p(z,s;x,t)$ by adding up (integrating) the probabilities of getting from (x,t) to (z,s) via (z_1,s_1), for all possible intermediate positions z_1.

How should we visualize p? Consider first the case when y is multidimensional Brownian motion. Then $p(\cdot,s;x,t)$ is the density of a Gaussian random variable with mean x and variance $s-t$. The graph of $z \to p(z,s;x,t)$ always has volume 1 below it (since p is a probability density); as $s \to \infty$ its maximum value tends to 0 (a Brownian particle diffuses further and further away, on average, as time increases); as $s \to t$ it becomes infinitely tall and thin (at time $s \approx t$ the Brownian particle is very close to its initial position x). The situation for a general stochastic differential equation is similar: p becomes infinitely tall and thin, concentrating at $z = x$, as $s \to t$; and if $gg^T > 0$ then the graph of p keeps spreading as $s \to \infty$. Of course in the general case p does not describe a Gaussian distribution, and there is no simple formula for the mean or variance – they are simply the mean and variance of $y(s)$.

If the stochastic differential equation does not involve time explicitly, then the transition probability depends only on the "elapsed time":

if $dy = f(y)dt + g(y)dw$ with f, g depending only on y, then $p(z, s; x, t) = p(z, s - t; x, 0)$.

If the stochastic differential equation does not involve space explicitly, then the transition probability depends only on the "relative position":

if $dy = f(t)dt + g(t)dw$ with f, g depending only on t, then $p(z, s; x, t) = p(z - x, s; 0, t)$.

The initial position of a Markov process need not be deterministic. Even if it is (e.g. if $y(0) = x$ is fixed), we may wish to consider a later time as the "initial time." The transition probability determines the evolution of the spatial distribution, no matter what its initial value: if $\rho_0(x)$ is the probability density of the state at time t then

$$\rho(z, s) = \int_{R^n} p(z, s; x, t)\rho_0(x)\, dx \tag{12}$$

gives the probability density (as a function of z) at any time $s > t$.

The crucial fact about the transition probability is this: it solves the *forward Kolmogorov equation* in s and z:

$$-p_s - \sum_i \frac{\partial}{\partial z_i}(f_i(z, s)p) + \tfrac{1}{2}\sum_{i,j,k} \frac{\partial^2}{\partial z_i \partial z_j}(g_{ik}(z, s)g_{jk}(z, s)p) = 0 \text{ for } s > t, \tag{13}$$

with initial condition

$$p = \delta_x(z) \text{ at } s = t.$$

We can write the forward Kolmogorov equation as

$$-p_s + \mathcal{L}^* p = 0 \tag{14}$$

with

$$\mathcal{L}^* p = -\sum_i \frac{\partial}{\partial z_i}(f_i p) + \sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}(a_{ij} p). \tag{15}$$

Here $a_{ij} = \frac{1}{2}(gg^T)_{ij}$ just as before. The initial condition $p = \delta_x(z)$ encapsulates the fact, already noted, that the graph of $p(\cdot, s; x, t)$ becomes infinitely tall and thin at x as s decreases to t. The technical meaning is that

$$\int_{R^n} p(z, s; x, t)f(z)\, dz \to f(x) \text{ as } s \text{ decreases to } t \tag{16}$$

for any continuous f.

Recall that if the initial state distribution is ρ_0 then the evolving distribution is $\rho(z,s) = \int p(z,s;x,t)\rho_0(x)\,dx$. This function $\rho(z,s)$ automatically solves the forward equation (just bring the derivatives under the integral, and use that p solves it). The initial condition on p is just what we need to have $\rho(z,s) \to \rho_0(z)$ as $s \to t$. (Demonstration: multiply (16) by $\rho_0(x)$ and integrate in x to see that

$$\int \rho(z,s)f(z)\,dz = \int p(z,s;x,t)f(z)\rho_0(x)\,dzdx \to \int f(x)\rho_0(x)\,dx$$

as $s \to t$. Since this is true for every continuous f, we conclude that $\rho(z,s)$ converges [weakly] to $\rho_0(z)$ as $s \to t$.)

Please note that the forward Kolmogorov equation describes the probability distribution by solving an initial-value problem, while the backward Kolmogorov equation describes the expected final payoff by solving a final-value problem. Students familiar with pricing options via binomial trees will find this familiar. The stock prices at various nodes of a tree are determined by working forward in time; the option values at various nodes of a tree are determined by working backward in time.

Notice that the forward and backward Kolmogorov equations are, in general, completely different. There is one case, however, when they are closely related: for Brownian motion the forward equation starting at $t = 0$ is

$$p_s - \tfrac{1}{2}\Delta p = 0 \text{ for } s > 0$$

while the backward equation with final time T is

$$u_t + \tfrac{1}{2}\Delta u = 0 \text{ for } t < T.$$

In this special case the backward equation is simply the forward equation with time reversed. More careful statement: if $u(x,t)$ solves the backward equation then $\tilde{u}(z,s) = u(z, T-s)$ solves the forward equation, and conversely. This is an *accident,* associated with the the self-adjointness of the Laplacian. The situation is different even for Brownian motion with constant drift f: then the forward equation is $p_s + f \cdot \nabla p - \frac{1}{2}\Delta p = 0$, while the backward equation is $u_t + f \cdot \nabla u + \frac{1}{2}\Delta u = 0$, and the two are not equivalent under time-reversal.

Students with a background in physical modeling will be accustomed to equations of the form $v_t = \text{div }(a(x)\nabla v)$. Neither the forward nor the backward Kolmogorov equation has this form. Such equations are natural in physics, but not in problems from control theory and stochastic differential equations.

Application: steady-state distributions. The backward Kolmogorov equation comes up more often than the forward one in finance. But one important application involves the large-time behavior of a diffusion. If $\rho(z,s)$ is the probability density of a diffusion, then evidently $\rho_\infty(z) = \lim_{s\to\infty} \rho(z,s)$ represents (if it exists) the large-time statistics of the process. For Brownian motion $\rho_\infty = 0$, reflecting the fact that Brownian particles wander a lot. The situation is quite different however for the Ornstein-Uhlenbeck process

$dy = -kyds + \sigma dw$. We expect y to remain near 0 due to the deterministic term $-ky$, which constantly pushes it toward 0. And indeed the steady-state distribution is

$$\rho_\infty(z) = Ce^{-kz^2/\sigma^2}$$

where C is chosen so that ρ_∞ has integral 1. (It's easy to check that this gives a steady-state solution of the forward Kolmogorov equation. In fact this gives the long-time asymptotics of a fairly general initial condition, but this is *not* so obvious.)

This application can be generalized. Consider the stochastic PDE $dy = -V'(y)dt + \sigma dw$. Its deterministic part pushes y toward a local minima of V. If V grows rapidly enough at infinity (so the diffusion is succesfully confined, and does not wander off to infinity) then the long-time statistics are described by the steady-state distribution

$$\rho_\infty(z) = Ce^{-2V(z)/\sigma^2}.$$

Testing the plausibility of the forward equation. We will explain presently why the forward equation holds. But first let's get used to it by examining some consequences and checking some special cases. Let $\rho_0(x)$ be the probability density of the state at time 0, and consider

$$\rho(z, s) = \int p(z, s; x, 0)\rho_0(x)\, dx$$

for $s > 0$. It gives the probability density of the state at time s.

Checking the integral. Since ρ is a probability density we expect that $\int \rho(z, s)\, dz = 1$ for all s. In fact, from the forward equation

$$\begin{aligned} \frac{d}{ds}\int \rho\, dz &= \int \rho_s\, dz \\ &= \int \mathcal{L}^*\rho\, dz \\ &= 0 \end{aligned}$$

since each term of $\mathcal{L}^*\rho$ is a perfect derivative. (Here and below, we repeatedly integrate by parts, with no "boundary terms" at $\pm\infty$. We are implicitly assuming that ρ and its derivatives decay rapidly as $z \to \pm\infty$. This is true, provided the initial distribution ρ_0 has this property.)

If the stochastic differential equation has no drift then the expected position is independent of time. In general, $E[y(s)] - E[y(0)] = E\int_0^s f(y(r), r)\, dr$ since the expected value of the integral dw vanishes. Thus when $f = 0$ the expected position $E[y(s)]$ is constant. Let's prove this again using the forward equation:

$$\begin{aligned} \frac{d}{ds}(\text{expected position}) &= \frac{d}{ds}\int z\rho(z, s)\, dz \\ &= \int z\rho_s(z, s)\, dz \\ &= \int z\mathcal{L}^*\rho(z, s)\, dz \\ &= 0 \quad \text{when } f = 0. \end{aligned}$$

The last step is the result of integration by parts; for example, if y is scalar valued ($dy = g(y,t)dw$) we have

$$\begin{aligned}\int z\mathcal{L}^*\rho\,dz &= \tfrac{1}{2}\int z\left(g^2\rho\right)_{zz}\,dz\\ &= -\tfrac{1}{2}\int\left(g^2\rho\right)_z\,dz\\ &= 0.\end{aligned}$$

(As noted above, to justify the integrations by parts one must know that ρ vanishes rapidly enough at spatial infinity.)

The special case $f = constant$, $g = 0$. If $g = 0$ then we're studying a deterministic motion. If in addition $f =$ constant then the solution is explicit and very simple: $y(t) = y(0) + ft$. Clearly

$$\text{Prob of being at } z \text{ at time } s = \text{Prob of being at } z - fs \text{ at time } 0,$$

whence

$$\rho(z,s) = \rho_0(z - fs).$$

In particular, $\rho_s + f\cdot\nabla\rho = 0$, which agrees with the forward equation (since f is constant).

Biting the bullet. Enough playing around; let's explain why the forward equation holds. The first main ingredient is the observation that

$$E_{y(t)=x}\left[\Phi(y(T))\right] = \int\Phi(z)p(z,T;x,t)\,dz. \tag{17}$$

We know how to determine the left hand side (by solving the backward equation, with final value Φ at $t = T$). This relation determines the integral of $p(\cdot,T;x,t)$ against any function Φ, for any value of x,t,T. This is a lot of information about p – in fact, it fully determines p. Our task is to make this algorithmic, i.e. to explain how p can actually be computed. (The answer, of course, will be to solve the forward equation in z and s.)

The second main ingredient is the relation between $\mathcal{L}$ and $\mathcal{L}^*$. Briefly: $\mathcal{L}^*$ is the *adjoint* of $\mathcal{L}$ in the L^2 inner product. Explaining this: recall from linear algebra that if A is a linear operator on an inner-product space, then its adjoint A^* is defined by

$$\langle Ax,y\rangle = \langle x,A^*y\rangle.$$

When working in R^n we can represent A by a matrix, and A^* is represented by the transpose A^T. The situation is similar here, but our inner product space consists of all (square-integrable, scalar-valued) functions on R^n, with inner product

$$\langle v,w\rangle = \int_{R^n} v(x)w(x)\,dx.$$

We claim that

$$\langle\mathcal{L}v,w\rangle = \langle v,\mathcal{L}^*w\rangle. \tag{18}$$

When y is scalar-valued our claim says that

$$\int_R \left(f v_x + \tfrac{1}{2} g^2 v_{xx} \right) w \, dx = \int_R v \left(-(fw)_x + \tfrac{1}{2}(g^2 w)_{xx} \right) \, dx.$$

This is a consequence of integration by parts. For example, the first term on the left equals the first term on the right since

$$\int_R [fw] v_x \, dx = - \int_R [fw]_x v \, dx.$$

The second term on each side matches similarly, integrating by parts twice. Notice that f and g can depend on time as well as space; it doesn't change the argument. The proof of (18) when y is vector valued is essentially the same as the scalar case.

The third main ingredient is hiding in our derivation of the backward equation. We know from this derivation that

$$E_{y(t)=x} \left[\phi(y(T), T) \right] - \phi(x,t) = E_{y(t)=x} \left[\int_t^T (\phi_s + \mathcal{L}\phi)(y(s), s) \, ds \right] \tag{19}$$

for any function $\phi(y,s)$. Our main use of this relation up to now was to choose ϕ so that the right hand side vanished, i.e. to choose ϕ to solve the backward equation. But we don't have to make such a restrictive choice: relation (19) holds for *any* ϕ.

Let's put these ingredients together. Rewriting (19) using the transition probabilities gives

$$\int_{R^n} \phi(z,T) p(z,T;x,t) \, dz - \phi(x,t) = \int_t^T \int_{R^n} (\phi_s + \mathcal{L}\phi)(z,s) p(z,s;x,t) \, dz ds. \tag{20}$$

Using (18) and doing the obvious integration by parts in time, the right hand side becomes

$$\int_t^T \int_{R^n} -\phi p_s + \phi \mathcal{L}^* p \, dz ds + \int_{R^n} \phi(z,s) p(z,s;x,t) \, dz \Big|_{s=t}^{s=T} . \tag{21}$$

This is true for *all* ϕ. Since the left hand side of (20) involves only the initial and final times (t and T) we conclude that

$$-p_s + \mathcal{L}^* p = 0.$$

Therefore (20)-(21) reduce to

$$\int_{R^n} \phi(z,t) p(z,t;x,t) \, dz = \phi(x,t)$$

for all ϕ, which is what we mean by the initial condition "$p = \delta_x$ when $s = t$". Done!

The argument is simple; but maybe it's hard to encompass. To recapitulate its essence, let's give a new proof (using the forward equation) of the fact (known via Ito calculus) that

$$u \text{ solves the backward equation} \Longrightarrow \frac{d}{ds} E\left[u(y(s), s) \right] = 0.$$

In fact: if $\rho(z,s)$ is the probability distribution of the state at time s,

$$\begin{aligned}
\frac{d}{ds}E\left[u(y(s),s)\right] &= \frac{d}{ds}\int u(z,s)\rho(z,s)\,dz \\
&= \int u_s\rho + u\rho_s\,dz \\
&= \int u_s\rho + u\mathcal{L}^*\rho\,dz \\
&= \int u_s\rho + (\mathcal{L}u)\rho\,dz \\
&= 0
\end{aligned}$$

using in the last step our hypothesis that u solves the backward equation.

****

Boundary value problems. The preceding discussion concerned the backward and forward Kolmogorov equations in all space. When working in a bounded domain, the boundary conditions for the forward Kolmogorov equation depend on what the random walk does when it reaches the boundary.

We discuss here just the case of most interest for financial applications: the *absorbing* boundary condition, i.e. a random walk that we track only till it hits the boundary for the first time. (After that time we think of the random walker as disappearing, i.e. being "absorbed" by the boundary.) The corresponding boundary condition for the forward Kolmogorov equation is that the probability density vanish there (since it represents the density of not-yet-absorbed walkers).

Let's explain briefly why this choice is right. Consider the backward Kolmogorov equation in a bounded domain, with boundary condition $u = 0$:

$$\begin{gathered}
u_t + \mathcal{L}u = 0 \text{ for } x \in D, t < T \\
u(x,T) = \phi(x) \text{ at } t = T \\
u(x,t) = 0 \text{ for } x \in \partial D.
\end{gathered}$$

We know that

$$u(x,t) = E_{y(t)=x}\left[\Phi(y(\tau),\tau)\right]$$

where $\tau = \tau(x)$ is the exit time from D (or T, if the path doesn't exit by time T) and

$$\Phi = 0 \text{ for } x \in \partial D;\ \Phi = \phi \text{ at the final time } T.$$

This formula for u can be written as

$$u(x,t) = \int_{R^n} \phi(z)q(z,T;x,t)\,dz$$

where

$q(z,s;x,t)$ = probability that the diffusion arrives at z at time s, starting from x at time t, without hitting ∂D first.

Our assertion is that $q(z, s; x, t)$ *solves the forward Kolmogorov equation for* $z \in D$ *and* $s > t$*, with boundary condition* $q = 0$ *for* $z \in \partial D$*, and initial condition* $q = \delta_x$. The justification is very much like the argument given above for R^n.

One thing changes significantly when we work in a bounded domain: $\int_D q(z, s; x, t)\, dz < 1$. The reason is that q gives the probability of arriving at z at time s *without hitting the boundary first*. Thus

$$1 - \int_D q(z, s; x, t)\, dz = \text{prob of hitting } \partial D \text{ by time } s, \text{ starting from } x \text{ at time } t.$$

Evidently $\int q(z, s; x, t)\, dz$ is decreasing in time. Let's check this for Brownian motion, for which $q_s - \frac{1}{2}\Delta q = 0$. We have

$$\begin{aligned} \frac{d}{ds}\int_D q(z, s; x, t)\, dz &= \int_D q_s\, dz \\ &= \tfrac{1}{2}\int_D \Delta q\, dz \\ &= \tfrac{1}{2}\int_{\partial D} \frac{\partial q}{\partial n} \\ &\le 0. \end{aligned}$$

The inequality in the last step is elementary: since $q = 0$ at ∂D and $q \geq 0$ in D we have $\partial q / \partial n \leq 0$ at ∂D, where n is the outward unit normal.

Application to the exit time distribution. We used the backward Kolmogorov equation to express the probability that a diffusion reaches a certain threshold before time T (see (8)). The forward Kolmogorov equation gives a very convenient alternative expression for the same quantity. Indeed, if ρ solves the forward Kolmogorov equation in the domain D of interest, with $\rho = 0$ at the boundary and $\rho = \delta_x$ at time 0, then $\int_D \rho(x, T)\, dx$ gives the probability of surviving till time T. So $1 - \int_D \rho(x, T)\, dx$ is the probability of hitting the boundary by time T, given that you started at x at time 0. When D is a half-space, this is an alternative expression for $u(x, 0; T)$ defined by (8).