Topic 5: Finite difference methods in derivative pricing

Jaroslav Borovička

Computational Dynamics (Spring 2023)

New York University

Economic problem

- We have a financial market with known price dynamics for a set of assets.
- We are interested in pricing securities whose payoffs are derived from the price of these assets.
- Valuation of these derivative securities must not lead to arbitrages.

Tools

- Brownian motion and Itô processes
- Black–Scholes formula
- Finite difference approximation of partial differential equations

Textbook

- *Brownian motion and Itô calculus*: Duffie (2001), Chapters 5.A–5.D. Øksendal (2007), Chapters 1–6.
- *Black–Scholes model*: Duffie (2001), Chapters 5.E–5.H, 6.G–6.I. Øksendal (2007), Chapter 12.3.
- *Numerical methods*: Judd (1998), Chapter 10, Holmes (2007), Thomas (1995), Candler (2001).

Applications

- Merton (1973), Black and Scholes (1973), Cox et al. (1979)

Brownian motion and Itô calculus

We study the problem of pricing of derivative securities in a continuous-time environment.

- Two assets: a stock with price Q_t that follows a given process, and a risk-free investment at interest rate r
- A derivative security that generates a one-time cash flow at time T in the amount $G(Q_T)$
- We are interested in the price of this derivative security at time $t \leq T$.

Black and Scholes (1973) and Merton (1973) provided a path-breaking solution to this problem.

- An application of the arbitrage pricing theory (APT) of Ross (1976)
- Two assets or portfolios that provide identical payoffs also must have the same price.

The derivative pricing result was formulated in a continuous-time model.

- Uncertainty is driven by a special process called the Brownian motion
- Characterization of the solution takes the form of a partial differential equation (PDE).

Discrete-time deterministic model of capital accumulation

$$k_{t+1} = (1 - \delta_t)\, k_t + i_t, \tag{5.1}$$

- δ_t is the depreciation rate and i_t is the investment rate

Now assume a time period of length Δt. Then

$$k_{t+\Delta t} - k_t = i_t \Delta t - \delta_t k_t \Delta t$$

- terms involving Δt represent investment and depreciation flows

Dividing by Δt and taking the limit as $\Delta t \to 0$ yields

$$\frac{dk_t}{dt} = i_t - \delta_t k_t.$$

- denote the investment rate $\iota_t = i_t / k_t$, then

$$\frac{dk_t}{dt}\frac{1}{k_t} = \frac{d \log k_t}{dt} = \iota_t - \delta_t.$$

The differential equation

$$\frac{dk_t}{dt}\frac{1}{k_t} = \frac{d\log k_t}{dt} = \iota_t - \delta_t.$$

with initial condition k_0 has the solution

$$k_t = k_0 \exp\left(\int_0^t (\iota_s - \delta_s)\, ds\right).$$

- We have solved for the stock of capital k_t by integrating up net investment $\iota_s - \delta_s$ along the trajectory of the economy over time on $s \in [0, t]$.
- This continuous-time limit expressed in the form of an integral is valid even in situations when functions ι and δ are stochastic, as long as the integral exists for each stochastic path.
- In this stochastic case, ι_t and δ_t are adapted to filtration $\{\mathcal{F}_t\}$, $t \in \mathcal{T} = \{0, 1, \ldots, T\}$.
- For a given path, the integral is a standard Riemann–Stieltjes integral, since the law of motion (5.1) implies that k_{t+1} is so-called ‘predictable’, i.e., k_{t+1} is $\mathcal{F}_t$-measurable.

The predictability assumption used in the preceding example is rather restrictive.

Consider the joint evolution of two security prices

$$\begin{aligned} Q_{t+1} &= Q_t + \mu_t + \sigma_t \left(W_{t+1} - W_t\right) \\ B_{t+1} &= B_t + r_t B_t. \end{aligned} \tag{5.2}$$

- Q_t is the price of a non-dividend paying stock
- B_t is the cumulative value of investment into a sequence of one-period risk-free bond contracts with one-period interest rate r_t
- $W_{t+1} - W_t \sim N\left(0, I\right)$ is a normally distributed shock
- the joint dynamics of the two processes generate a filtration $\{\mathcal{F}_t\}$, $t \in \mathcal{T} = \{0, 1, \ldots, T\}$

The expected return on the stock is

$$E\left[\frac{Q_{t+1} - Q_t}{Q_t} \mid \mathcal{F}_t\right] = \frac{\mu_t}{Q_t}$$

and σ_t is the one-period volatility of the stock return.

At any date t, the investor chooses to invest the current wealth J_t

- purchase θ_t^f units of the risk-free asset at price B_t, and θ_t^r units of the risky asset at price Q_t
- the budget constraint is

$$J_t = \theta_t^f B_t + \theta_t^r Q_t.$$

- the value of this portfolio at time $t+1$ is

$$J_{t+1} = \theta_t^f B_{t+1} + \theta_t^r Q_{t+1},$$

which can be subsequently reinvested again.

Manipulating this expression yields

$$J_{t+1} = \theta_t^f \left(B_{t+1} - B_t\right) + \theta_t^r \left(Q_{t+1} - Q_t\right) + \underbrace{\theta_t^f B_t + \theta_t^r Q_t}_{J_t}.$$

Summing up wealth gains $J_{t+1} - J_t$ over time, we have

$$\sum_{t=0}^{T-1} (J_{t+1} - J_t) = J_T - J_0 = \sum_{t=0}^{T-1} \left[\theta_t^f (B_{t+1} - B_t) + \theta_t^r (Q_{t+1} - Q_t)\right].$$

The intertemporal portfolio choice is determined as a solution to the problem of maximizing expected utility from time-T wealth J_T,

$$E[u(J_T)]$$

subject to the intertemporal budget constraint and initial condition J_0, with

$$J_T = J_0 + \sum_{t=0}^{T-1} \left[\theta_t^f r_t B_t + \theta_t^r \mu_t + \theta_t^r \sigma_t (W_{t+1} - W_t)\right].$$

- the fact that the investor chooses the portfolio at time t and has to hold it fixed until returns in period $t+1$ are realized is also called the self-financing property

Repeating the continuous-time approximation, we have the dynamics on periods with interval Δt

$$\begin{aligned} Q_{t+\Delta t} - Q_t &= \mu_t \Delta t + \sigma_t \left(W_{t+\Delta t} - W_t \right) \\ B_{t+\Delta t} - B_t &= r_t B_t \Delta t, \end{aligned}$$

with $W_{t+\Delta} - W_t \sim N\left(0, \Delta t\right)$. We would like to take the continuous-time limit that should lead to

$$\begin{aligned} dQ_t &\approx \mu_t + \sigma_t \text{“} dW_t \text{”} \\ dB_t &= r_t B_t dt. \end{aligned} \tag{5.3}$$

- The question is how to construct the limiting approximation of the stochastic component “dW_t” on the first line rigorously.
- The limit will lead to a so-called stochastic differential equation, which cannot be characterized by a Riemann–Stieltjes or Lebesgue integral.
- Uncertainty in Q_t will be driven by innovations to a Brownian motion that could be interpreted as a limiting sequence of normally distributed increments.

Similarly, the time refinement of the wealth accumulation process is given by

$$\begin{aligned} J_T &= J_0 + \sum_{i=0}^{I-1} \left[\theta^f_{i\Delta t} \left(B_{(i+1)\Delta t} - B_{i\Delta t}\right) + \theta^r_{i\Delta t} \left(Q_{(i+1)\Delta t} - Q_{i\Delta t}\right)\right] \\ &= J_0 + \sum_{i=0}^{I-1} \left[\left(\theta^f_{i\Delta t} r_{i\Delta t} B_{i\Delta t} + \theta^r_{i\Delta t} \mu_{i\Delta t}\right) \Delta t + \theta^r_{i\Delta t} \sigma_{i\Delta t} \left(W_{(i+1)\Delta t} - W_{i\Delta t}\right)\right] \end{aligned}$$

with $I = T/\Delta t$.

- We are interested in the continuous-time limit of the portfolio strategy $\left\{\theta^f_t, \theta^r_t\right\}$ that leads to

$$J_T = J_0 + \int_0^T \left[\left(\theta^f_t r_t B_t + \theta^r_t \mu_t\right) dt + \theta^r_t \sigma_t \text{``} dW_t \text{''}\right]. \tag{5.4}$$

In the discrete-time model, the investor chooses the portfolio shares θ_t^f, θ_t^r at discrete times $t = 0, 1, \ldots, T - 1$, where each pair θ_t^f, θ_t^r is $\mathcal{F}_t$-measurable.

In the continuous-time limit, the investor will adjust the portfolio continuously in a sense that needs to be made precise so that it preserves the self-financing property.

- Wealth accumulation needs to satisfy the principle that the investor chooses a portfolio and then must hold it fixed 'over the next instant'.

This strategy will be represented by a pair of stochastic processes θ_t^f, θ_t^r, $t \in [0, T]$ that will depend on the observed histories of the shocks, and satisfy certain measurability restrictions.

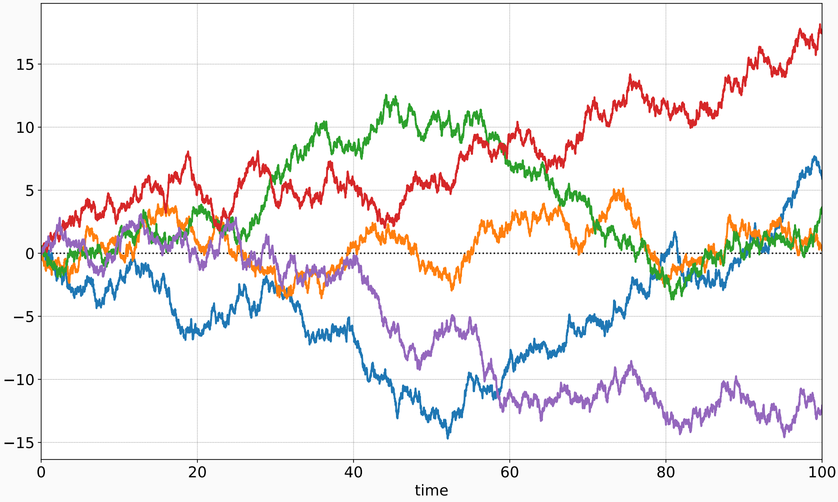
Definition 5.1

A k-**dimensional Brownian motion** is a stochastic process W on $\mathbb{R}^k$ such that

1. $W_0 = 0$,
2. $\forall s, t \in \mathcal{T}$ for which $s \leq t$, the difference $W_t - W_s \sim N\left(0, (t-s)\, I_k\right)$,
3. for all $t_0 < t_1 < t_2 < \ldots t_n \in \mathcal{T}$, the random variables $W_{t_j} - W_{t_{j-1}}, j \in \{1, \ldots, n\}$ are independent.

Said simply, the Brownian motion is a process with independent, normally distributed increments.

This definition characterizes a unique process, as long as we restrict our attention to processes with continuous sample paths.



Sample paths of a Brownian motion are nowhere differentiable, and have ‘infinite length’.

Filtered probability space for the Brownian motion

Formally, the Brownian motion is defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$.

- Ω is the sample space, or the set of all paths of the Brownian motion, with elements $\omega \in \Omega$
- $W(\omega)$ represents one particular path of the Brownian motion, and $W_t(\omega)$ the associated value along that path at time t
- the σ-algebra $\mathcal{F}$ represents the set of all sets of paths to which probabilities can be assigned
- the Brownian motion generates a filtration $\{\mathcal{F}_t\}$ where, somewhat informally, $\mathcal{F}_t$ is the information set that contains all information about the realized path of the Brownian motion up to time t.
- P is the probability measure over the paths implied by the Gaussian assumption

The Brownian motion satisfies the Markov property: $\forall t, s \geq 0$ and for every (Borel) set $H \in \mathcal{B}$ on $\mathbb{R}^k$

$$P\left(W_{t+s} \in H \mid \mathcal{F}_t\right) = P\left(W_{t+s} \in H \mid W_t\right).$$

- distribution of W_{t+s} conditional on time-t information set is the same as the distribution conditional only on the value W_t

The Brownian motion is also a martingale with respect to filtration $\{\mathcal{F}_t\}$. For $s \leq t$,

$$E\left[W_t \mid \mathcal{F}_s\right] = E\left[W_t - W_s \mid \mathcal{F}_s\right] + W_s = W_s$$

and, at the same time,

$$\left(E\left[|W_t|\right]\right)^2 \leq E\left[|W_t|^2\right] = nt < +\infty.$$

We want to establish a formal definition of how 'variable' the paths of a stochastic process are.

- partition a particular time interval $\mathcal{T}$
- define a discrete-time version of variability of the paths by measuring changes in the value of the stochastic process along the path between the nodes of the partition
- take a continuous-time limit as the partition is refined and the distance between the nodes of the partition vanishes to zero.

Definition 5.2

The set of points $\mathcal{P} = \{t_0, \ldots, t_n\}$ with $0 = t_0 < t_1 < \ldots < t_n = t$ is a **partition of the interval** $[0, t]$. Define

$$l(\mathcal{P}) = \max |t_j - t_{j-1}| \, .$$

to be the **norm of the partition**.

Denote $l(\mathcal{P}) \to 0$ to be the limit of an arbitrary sequence of partitions $\mathcal{P}$ such that the norm of the partitions in the sequence converges to zero.

Definition 5.3

Let $X : \Omega \times \mathcal{T} \to \mathbb{R}$ be a continuous stochastic process. Then for $p > 0$ define the **p-th variation process** of X_t as

$$\langle X, X \rangle_t^p (\omega) = \lim_{l(\mathcal{P}) \to 0} \sum_{j=0}^{n-1} \left| X_{t_{j+1}} (\omega) - X_{t_j} (\omega) \right|^p$$

where the limit is in probability.

For $p = 1$, the variation process is called the **total variation process**, and for $p = 2$, it is called the **quadratic variation process**.

Lemma 5.4

For the univariate Brownian motion W,

$$\langle W, W \rangle_t^2 (\omega) = t \quad a.s..$$

To show this, start with a partition $\mathcal{P}$ of the time interval $[0, t]$, and consider

$$E\left[\left(\sum_{t_j \le t} \left(W_{t_{j+1}} - W_{t_j}\right)^2 - t\right)^2\right] = E\left[\left(\sum_{t_j \le t} \left(W_{t_{j+1}} - W_{t_j}\right)^2\right)^2\right] - 2t \sum_{t_j \le t} E\left[\left(W_{t_{j+1}} - W_{t_j}\right)^2\right] + t^2$$

$$= \sum_{t_j \le t} 3\left(t_{j+1} - t_j\right)^2 + \sum_{\substack{t_j, t_k \le t \\ j \ne k}} \left(t_{j+1} - t_j\right)\left(t_{k+1} - t_k\right) - 2t^2 + t^2 =$$

$$= 2\sum_{t_j \le t} \left(t_{j+1} - t_j\right)^2 + t^2 - 2t^2 + t^2 \to 0$$

as $l(\mathcal{P}) \to 0$. Therefore $\sum_{t_j \le t} \left(W_{t_{j+1}} - W_{t_j}\right)^2 \to t$.

Quadratic variation of a Brownian motion

For the univariate Brownian motion W,

$$\langle W, W\rangle_t^2(\omega) = t \quad \text{a.s..}$$

- this shows that every individual path of the Brownian motion on $[0, t]$ has the same 'length' t when measured using the quadratic variation.
- hence $\langle W, W\rangle_{\Delta t}^2(\omega) = \Delta t$ for an arbitrarily short interval Δt
- this provides heuristic intuition why we can write "$(dW_t)^2 = dt$", which is a central insight of Itô calculus, as manifested in Itô's lemma

Since the quadratic variation is finite, it can be shown that the total variation of a Brownian motion must be infinite,

$$\forall t > 0 : \langle W, W\rangle_t^1(\omega) = +\infty.$$

This conclusion also implies that the paths of a Brownian motion are nowhere differentiable.

We now use the Brownian motion to build more complicated processes called stochastic integrals.

- the geometric argument underlying the construction is conceptually similar to that of the Riemann–Stieltjes integral
- technical complications associated with the irregularity of paths of the Brownian motion are substantial and require a careful treatment

Stochastic integrals are incredibly versatile

- martingale representation theorem: any martingale in an environment with uncertainty generated by a Brownian motion can be represented as a stochastic integral

In order to construct the Riemann integral of a piecewise continuous function $f(t)$ on $\mathcal{T} = [0, T]$ we choose a partition $\mathcal{P}$ of $\mathcal{T}$, and then define the integral through the limit

$$\int_0^T f(t)\, dt \doteq \lim_{l(\mathcal{P})\to 0} \sum_{j=0}^{n-1} f(\tau_j)\,(t_{j+1} - t_j)\,, \tag{5.5}$$

- τ_j are arbitrary values within the intervals of the partition, $\tau_j \in [t_j, t_{j+1}]$.
- the integral is well defined only if the limit does not depend on a particular choice of the sequence of partitions, nor on the choices of the points $\tau_j \in [t_j, t_{j+1}]$.
- feometrically, the construction approximates the area under the curve $f(t)$ using the sum of rectangular areas $f(\tau_j)\,(t_{j+1} - t_j)$.

The Stieltjes integral integrates along the path of a sufficiently smooth function $g(t)$:

$$\int_0^T f(t)\, dg(t) \doteq \lim_{l(\mathcal{P})\to 0} \sum_{j=0}^{n-1} f(\tau_j)\,(g(t_{j+1}) - g(t_j))\,. \tag{5.6}$$

The idea underlying the construction of the stochastic integral is similar

- integration along the path of a smooth function g is replaced with integration along the path of the Brownian motion $W(\omega)$

We desire to form the discrete-time approximation using a partition $\mathcal{P}$,

$$\sum_{j=0}^{n-1} f_{\tau_j}(\omega)\left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right), \tag{5.7}$$

and ask how to construct a well-defined limit as $l(\mathcal{P}) \to 0$, in the same way the Riemann–Stieltjes integral is formulated in (5.6).

- the sum in (5.7) depends on the particular path ω of the Brownian motion
- integrand $f_{\tau_j}(\omega)$ can also be a stochastic process
- integral defined pathwise, for each ω

The stochastic integral that is the desired outcome of this construction is therefore also a stochastic process.

Due to the infinite total variation of W, we need to choose the points τ_j in a specific way to make the limit well defined.

- our particular interest in financial applications leads us to choose τ_j to be the initial point of the interval, $\tau_j = t_j$
- this yields the so-called Itô integral

The construction proceeds in several steps.

- first providing the definition of the integral for a class of so-called elementary processes
- then extend this definition to larger classes of processes through limits

Let the share price evolve as a Brownian motion W. Consider an investor who can trade shares only at a finite number of dates $t_j \in [0, T]$ which define a partition $\mathcal{P}$.

Denote $\theta_{t_j}(\omega)$ the number of shares bought at time t_j.

- we assume that the choice $\theta_{t_j}(\omega)$ can depend on information available up to time t_j
- evolution of wealth J_t is given by

$$J_t(\omega) = J_0 + \sum_{j=0}^{n(t)-1} \theta_{t_j}(\omega)\left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right) + \theta_{n(t)}(\omega)\left(W_t(\omega) - W_{t_{n(t)}}(\omega)\right) \tag{5.8}$$

  where $n(t)$ is the index of the interval in the partition such that $t \in [t_{n(t)}, t_{n(t)+1})$, and $n(T) = n$.
- the wealth process represents cumulative gains from investments between trading dates
- the process θ_t viewed as a continuous-time process is constant on the intervals $[t_j, t_{j+1})$, and is called a dynamic strategy

Processes that have piecewise constant trajectories that are allowed to jump only at a finite number of times are called elementary processes.

Definition 5.5

An **elementary** (also called simple) **process** ϕ on $[0, T]$ is a process for which there exists a partition $\mathcal{P}$ of $[0, T]$ such that $\phi_t = \phi_{t_j}$ for $t \in [t_j, t_{j+1})$.

Definition 5.6

For the class of elementary processes ϕ, the **Itô integral** of ϕ is defined as

$$\int_0^t \phi_s(\omega)\, dW_s(\omega) \doteq \sum_{j=0}^{n(t)-1} \phi_{t_j}(\omega)\left(W_{t_{j+1}}(\omega) - W_{t_j}(\omega)\right) + \phi_{t_{n(t)}}(\omega)\left(W_t(\omega) - W_{t_{n(t)}}(\omega)\right), \tag{5.9}$$

where the last term reflects the interrupted last interval of the partition and $n(t)$ is such that $t \in [t_{n(t)}, t_{n(t)+1})$.

We use elementary processes to approximate more complicated processes. The goal is to extend the definition of Itô integral as a limiting approximation using Itô integral of elementary processes.

To illustrate the challenge, consider two candidate approximations of the path of a Brownian motion W on the interval $\mathcal{T} = [0, T]$, on a given partition $\mathcal{P}$ of $\mathcal{T}$:

$$\begin{aligned} \phi_t(\omega) &= \sum_{j=0}^{n-1} W_{t_j}(\omega)\,\mathbf{1}_{[t_j, t_{j+1})}(t) \\ \psi_t(\omega) &= \sum_{j=0}^{n-1} W_{t_{j+1}}(\omega)\,\mathbf{1}_{[t_j, t_{j+1})}(t). \end{aligned}$$

- ϕ_t approximates path of $W_t(\omega)$ on $[t_j, t_{j+1})$ with the initial value $W_{t_j}(\omega)$ on each interval
- ψ_t approximates path of $W_t(\omega)$ on $[t_j, t_{j+1})$ with the terminal value $W_{t_{j+1}}(\omega)$ on each interval
- as we refine the partition, the approximations approach each other in a loose sense, so it would seem that choosing ϕ_t or ψ_t will lead to the same conclusions $l(\mathcal{P}) \to 0$.

In the case of a Riemann integral, the limits of integrals of the two approximations as we refine the partition indeed coincide, and define the Riemann integral of the Brownian motion:

$$\lim_{l(\mathcal{P})\to 0} \int_0^T \phi_t(\omega)\, dt = \lim_{l(\mathcal{P})\to 0} \int_0^T \psi_t(\omega)\, dt \doteq \int_0^T W_t(\omega)\, dt.$$

This is not surprising because this construction is in line with the definition of Riemann integral.

- since the path of a Brownian motion is continuous, the path is Riemann integrable
- for a Riemann integral, the choice of approximation points $\tau_j \in [t_j, t_{j+1}]$ is inconsequential
- geometrically, the areas under the two curves given by ϕ_t and ψ_t converge to each other

The situation is markedly different in the case of the stochastic integral. Since ϕ and ψ are elementary processes, their Itô integrals on $[0, T]$ are defined, omitting the path arguments ω, as

$$\begin{aligned} \int_0^t \phi_s dW_s &\doteq \sum_{j=0}^{n(t)-1} W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) + W_{t_{n(t)}}\left(W_t - W_{t_{n(t)}}\right) \\ \int_0^t \psi_s dW_s &\doteq \sum_{j=0}^{n(t)-1} W_{t_{j+1}}\left(W_{t_{j+1}} - W_{t_j}\right) + W_{t_{n(t)+1}}\left(W_t - W_{t_{n(t)}}\right). \end{aligned}$$

To see the distinction between the two constructions, compute the expectations

$$\begin{aligned} E\left[\int_0^T \phi_t dW_t \mid \mathcal{F}_0\right] &= E\left[\sum_{j=0}^{n-1} W_{t_j}\left(W_{t_{j+1}} - W_{t_j}\right) \mid \mathcal{F}_0\right] = 0, \\ E\left[\int_0^T \psi_t dW_t \mid \mathcal{F}_0\right] &= E\left[\sum_{j=0}^{n-1} W_{t_{j+1}}\left(W_{t_{j+1}} - W_{t_j}\right) \mid \mathcal{F}_0\right] = T. \end{aligned} \tag{5.10}$$

These expectations will therefore remain distinct even if we take the limit $l(\mathcal{P}) \to 0$.

Despite the fact that both ϕ and ψ seem to be reasonable approximations of W, they yield different answers for the stochastic integral of W.

- the intuitive reason is that there is too much variation in W over time (W is a process of infinite total variation)
- the approximations ϕ and ψ are too distinct once we integrated along a path with infinite total variation
- in contrast, the Riemann–Stieltjes integral (5.6) integrates against a function g that has finite total variation

Itô integrals as adapted processes

From the perspective of financial applications, the approximation via ϕ that uses the initial points of the intervals is the desirable choice.

- this choice aligns with the wealth accumulation process generated by self-financing dynamic portfolio strategies
- mathematically, the Itô integral is adapted to filtration $\{\mathcal{F}_t\}$ generated by the Brownian motion, meaning that the portfolio choice at the t cannot depend on information at future dates $u > t$
- on the other hand, the integral constructed using ψ is not adapted to $\{\mathcal{F}_t\}$ because the integral up to time $t \in [t_j, t_{j+1})$ uses the value of ψ at $t_{j+1} > t$
- while this may seem innocuous in the limit as $l(\mathcal{P}) \to 0$, it is a key difference
- in the wealth accumulation example, ψ corresponds to choosing the portfolio θ_{t_j} at time t_j according to the realized return $W_{t_{j+1}}$ at the future date t_{j+1}

The Itô integral is also a martingale with respect to the filtration generated by the Brownian motion, consistently with the expected value in (5.10). For $u < t$,

$$\begin{aligned} E\left[\int_0^t \phi_s dW_s \mid \mathcal{F}_u\right] &= E\left[\int_0^u \phi_s dW_s + \int_u^t \phi_s dW_s \mid \mathcal{F}_u\right] \\ &= \int_0^u \phi_s dW_s + E\left[\int_u^t \phi_s dW_s \mid \mathcal{F}_u\right] = \int_0^u \phi_s dW_s. \end{aligned} \tag{5.11}$$

- the first equality follows from the linearity of the integral
- the second equality from the fact that $\int_0^u \phi_s dW_s$ is measurable with respect to $\mathcal{F}_u$
- the third equality utilizes the same calculation as in (5.8).

Having defined the Itô integral on the class of elementary processes in Definition 5.6, the goal now is to extend this definition to a larger class of stochastic processes f, giving a meaning to the expression

$$\int_0^t f_s(\omega)\, dW_s(\omega)\,.$$

- construction is based on limiting approximations of the stochastic process f using elementary processes (for details, see, for example, Øksendal (2007), Chapter 3)

We need to restrict attention to a suitable class of processes that can be meaningfully approximated.

- Let $\mathcal{L}$ be the set of all processes adapted to the filtration generated by the Brownian motion. Define

$$\mathcal{H}^2 = \left\{ f \in \mathcal{L} : E\left[\int_0^T (f_t)^2\, dt\right] < \infty \right\}.$$

Definition 5.7

Let $f \in \mathcal{H}^2$. Then the **Itô integral** of f is defined by

$$\int_0^T f_t(\omega)\, dW_t(\omega) = \lim_{k \to \infty} \int_0^T \phi_t^k(\omega)\, dW_t(\omega) \tag{5.12}$$

where $\left\{\phi^k\right\}$ is a sequence of elementary functions in $\mathcal{H}^2$ such that

$$E\left[\int_0^T \left(f_t(\omega) - \phi_t^k(\omega)\right)^2 dt\right] \to 0 \quad \text{as } k \to \infty. \tag{5.13}$$

The idea of the definition is to construct a sequence of elementary processes $\left\{\phi^k\right\}$ with piecewise constant paths such that ϕ^k approximates the process f successively better as $k \to \infty$.

The definition is only meaningful if every such sequence $\left\{\phi^k\right\}$ that satisfies (5.13) yields the same value of the limit on the right-hand side of (5.12), which is a result that needs to be proven. Then this common value defines the Itô integral of f.

Itô integrals of processes $f \in \mathcal{H}^2$ preserve the martingale property, just like in the case of elementary processes (5.11):

$$E\left[\int_0^t f_s(\omega)\, dW_s(\omega) \mid \mathcal{F}_u\right] = \int_0^u f_s(\omega)\, dW_s(\omega)\,.$$

We now define a class of processes called Itô processes that additively combine Riemann integrals and Itô integrals. This class is sufficiently general to cover many interesting applications.

Definition 5.8

An *n*-**dimensional Itô process** is a process $X : \Omega \times \mathcal{T} \to \mathbb{R}^n$ such that

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s \tag{5.14}$$

where W is a k-dimensional Brownian motion. We assume that μ and σ are $\mathcal{F}_t$-adapted where $\{\mathcal{F}_t\}$ is a filtration with respect to which W is a martingale.

An **Itô diffusion** is an Itô process for which the coefficients satisfy $\mu_s = \mu(X_s)$ and $\sigma_s = \sigma(X_s)$ for all $s \in \mathcal{T}$.

Often, equation (5.14) is written in the 'differential' form

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

- The process μ is called drift, and σ is called the volatility of the Itô process.
- The Itô process X_t defined above is n-dimensional, with uncertainty generated by a k-dimensional Brownian motion
- μ is an $n \times 1$-dimensional vector process, and σ is an $n \times k$ dimensional

When the processes $\mu, \sigma \in \mathcal{H}^2$, then the Itô integral is a martingale and the argument of the Itô integral is square integrable.

It then follows that for $t, u \geq 0$,

$$
\begin{aligned}
E\left[X_{t+u} \mid \mathcal{F}_t\right] &= X_t + \int_t^{t+u} \mu_s ds \\
Var\left[X_{t+u} \mid \mathcal{F}_t\right] &= E\left[\left(\int_t^{t+u} \sigma_s dW_s\right)^2 \mid \mathcal{F}_t\right] = E\left[\int_t^{t+u} |\sigma_s|^2 \, dt \mid \mathcal{F}_t\right]
\end{aligned}
$$

where the last equality follows from a result known as Itô isometry.

Then we can localize the mean and variance by constructing the infinitesimal expected growth rate and variance of the process:

$$\begin{aligned} \frac{d}{du} E\left[X_{t+u} \mid \mathcal{F}_t\right]\Big|_{u=0} &= \mu_t \quad \text{a.s.} \\ \frac{d}{du} Var\left[X_{t+u} \mid \mathcal{F}_t\right]\Big|_{u=0} &= |\sigma_s|^2 = \sigma_t \sigma_t' \quad \text{a.s.} \end{aligned}$$

which justifies calling the two coefficients the drift and volatility of the Itô process.

Informally, we will write these results in the shorthand differential form

$$\begin{aligned} E_t\left[dX_t\right] &= \mu_t dt \\ Var_t\left[dX_t\right] &= \sigma_t \sigma_t' dt, \end{aligned}$$

where $E_t[\cdot] = E[\cdot \mid \mathcal{F}_t]$.

The definition of an Itô process X in (5.14) seems to be restrictive, since it involves a linear combination of a Riemann integral and an Itô integral.

- It would then seem that nonlinear transformations of X would no longer be Itô processes.
- For example, in the case of the discrete-time linear vector-autoregression

$$x_{t+1} = A_o x_t + C w_{t+1} \qquad w_{t+1} \sim N(0, I_p) \quad \text{iid},$$

a transformation $y_t = f(x_t)$ for some nonlinear function f would no longer yield a linear vector-autoregression for y_t.

Starting from a given Itô process X, we want to characterize its nonlinear transformation $Y_t = f(t, X_t)$ where f is a given, sufficiently differentiable function.

- It turns out that Y_t is again an Itô process.

The characterization is provided by a key result of stochastic calculus, Itô's lemma, due to Itô (1951).

We provide its scalar version, with only a heuristic proof.

Theorem 5.9 (Itô's lemma)

Let X be a univariate Itô process

$$dX_t = \mu_t dt + \sigma_t dW_t$$

where W is a univariate Brownian motion. Let $f : \mathbb{R}^2 \to \mathbb{R}$ with $f \in C^2(\mathcal{T} \times \mathbb{R})$ (twice continuously differentiable). Then $Y_t = f(t, X_t)$ is an Itô process and

$$dY_t = f_t(t, X_t)\, dt + f_x(t, X_t)\, \mu_t dt + \frac{1}{2} f_{xx}(t, X_t)\, \sigma_t^2 dt + f_x(t, X_t)\, \sigma_t dW_t.$$

Proof. The heuristic proof goes as follows. First consider a 'second-order' Taylor approximation

$$dY_t = f_t dt + f_x dX_t + \frac{1}{2} f_{tt} (dt)^2 + f_{tx} dt dX_t + \frac{1}{2} f_{xx} (dX_t)^2$$

Now observe

$$\begin{aligned} dt dX_t &= dt\left(\mu_t dt + \sigma_t dW_t\right) = \mu_t (dt)^2 + \sigma_t dt dW_t \\ (dX_t)^2 &= \mu_t^2 (dt)^2 + 2\mu_t \sigma_t dt dW_t + \sigma_t^2 (dW_t)^2 \end{aligned}$$

- when we computed the quadratic variation of an Itô process, we argued that $(dW_t)^2 = dt$
- hence $(dW_t)^2$ is a first-order term in dt
- since dW_t can be argued to have mean zero and variance dt, the term $dt dW_t$ will be mean zero and variance $(dt)^2$, which is a higher-order stochastic term than dW_t
- therefore, the only term left in the two expressions above is $\sigma_t^2 (dW_t)^2 = \sigma_t^2 dt$.
- combining these results yields the statement of Itô's lemma

■

A key observation obtained from Itô's lemma is that the process Y_t also follows an Itô diffusion:

$$Y_t = Y_0 + \int_0^t \left[f_t\left(s, X_s\right) + f_x\left(s, X_s\right)\mu_s + \frac{1}{2} f_{xx}\left(s, X_s\right)\sigma_s^2 \right] ds + \int_0^t f_x\left(s, X_s\right)\sigma_s dW_s.$$

- The linearity of the Itô process and additivity of its two integrals is therefore without loss of generality, and preserved under the nonlinear transformation $Y_t = f(t, X_t)$.
- The nonlinearity is embedded in the transformation of the drift and volatility coefficients of the Itô process.

Itô's lemma can be directly extended to multivariate Brownian motions when we note that for two independent Brownian motions W^j and W^k, we have $\left(dW_t^j\right)\left(dW_t^k\right) = 0$.

Theorem 5.10 (Multivariate Itô's lemma)

Let W be a k-dimensional Brownian motion, X an n-dimensional Itô process

$$dX_t = \mu_t dt + \sigma_t dW_t$$

and $f : \mathcal{T} \times \mathbb{R}^n \to \mathbb{R}^m$ be from C^2. Then for $Y_t = f(t, X_t)$, we have for the k-th component Y_t^k

$$dY_t^k = \left[f_t^k + f_x^k \mu_t + \frac{1}{2} \mathrm{tr}\left[\sigma_t \sigma_t' f_{xx}^k\right]\right] dt + f_x^k \sigma_t dW_t.$$

Let X be an Itô process characterized in differential form by

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

with a given initial condition X_0. We can proceed by integrating

$$\begin{aligned} \int_0^t dX_s &= X_t - X_0 \\ &= \int_0^t \mu ds + \int_0^t \sigma dW_s = \mu \int_0^t ds + \sigma \int_0^t dW_s = \mu t + \sigma \left(W_t - W_0\right). \end{aligned}$$

Since $W_0 = 0$, we obtain the explicit solution for X_t in the form

$$X_t = X_0 + \mu t + \sigma W_t,$$

which is a process called the arithmetic Brownian motion. In particular, since $W_t \sim N\left(0, t\right)$, the distribution of X_t conditional on X_0 is

$$X_t \sim N\left(X_0 + \mu t, \sigma^2 t\right).$$

Let X be an Itô process characterized in differential form by

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

with a given initial condition X_0. To find an explicit solution for X_t, we cannot integrate both sides of the above formulas since the right-hand side also depends on X_t.

Let us therefore first define $Y_t = \log X_t$ and apply Itô's lemma

$$\begin{aligned} dY_t &= d\log X_t = \frac{1}{X_t}dX_t - \frac{1}{2}\frac{1}{X_t^2}(dX_t)^2 = \frac{1}{X_t}(\mu X_t dt + \sigma X_t dW_t) - \frac{1}{2}\frac{1}{X_t^2}\sigma^2 X_t^2 dt \\ &= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t. \end{aligned}$$

We can now integrate both sides of the equation

$$\int_0^t dY_s = Y_t - Y_0 = \int_0^t \left(\mu - \frac{1}{2}\sigma^2\right)ds + \int_0^t \sigma dW_s = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t.$$

Hence we obtain

$$Y_t = Y_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t.$$

Exponentiating and noticing that $X_t = \exp(Y_t)$, we have the solution

$$X_t = X_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W_t\right),$$

which is a process called the geometric Brownian motion.

X_t is therefore a random variable that is log-normally distributed conditional on X_0,

$$X_t \sim N\left(\log X_0 + \left(\mu - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right),$$

and

$$E[X_t | X_0] = X_0 \exp(\mu t),$$

which follows from the formula for the mean of a log-normally distributed variable.

The Black–Scholes model

The Black–Scholes model for option pricing has been developed in Black and Scholes (1973), with a central insight based on dynamic hedging provided by Robert Merton.

The model was extended to the pricing of more complicated derivative securities in Merton (1973), and to more complex environments in the subsequent literature.

While research on the pricing of derivative securities has been active before, the central contribution of Black and Scholes (1973) and Merton (1973) is that they were able to derive the valuation formulas in terms of relatively easy-to-measure parameters.

Time is continuous and given by a finite interval $\mathcal{T} = [0, T]$. Two securities are traded.

Risk-free bond provides a constant risk-free return r over each infinitesimal horizon.

- An initial investment $B_0 = 1$ into this security accumulates over time as

$$dB_t = rB_t dt \tag{5.15}$$

so that the value of such an investment at time t is

$$B_t = \exp\left(\int_0^t r ds\right) = \exp(rt). \tag{5.16}$$

Risky non-dividend yielding stock has price Q_t that follows a geometric Brownian motion

$$dQ_t = \mu Q_t dt + \sigma Q_t dW_t \tag{5.17}$$

- constant scalar parameters μ and σ and a given initial price Q_0.

The security market is hence characterized by three parameters, the risk-free rate r, the expected return on the risky investment μ and the volatility of the risky investment σ.

The stock price process has an explicit solution

$$Q_t = Q_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right). \tag{5.18}$$

The expected price conditional on Q_0 then is

$$E[Q_t|Q_0] = Q_0 \exp(\mu t).$$

We can compute the annualized expected returns over an infinitesimal horizon t.

For the investment into the risk-free security

$$\lim_{t\to 0} \frac{1}{t}\frac{E[B_t] - B_0}{B_0} = \lim_{t\to 0} \frac{1}{t}(\exp(rt) - 1) = r,$$

and for the risky stock

$$\lim_{t\to 0} \frac{1}{t}\frac{E[Q_t] - Q_0}{Q_0} = \lim_{t\to 0} \frac{1}{t}(\exp(\mu t) - 1) = \mu.$$

The infinitesimal risk premium on the stock therefore is $\mu - r$.

At any time t, an investor can choose to purchase

- θ_t^f units of the risk-free asset at price B_t
- θ_t^r units of the stock at price Q_t.

The financial gain over an infinitesimal horizon from this investment is

$$\theta_t^f dB_t + \theta_t^r dQ_t$$

and a given portfolio strategy θ^f, θ^r yields terminal wealth at time T

$$J_T = J_0 + \int_0^T \left[\left(\theta_t^f r B_t + \theta_t^r \mu \right) dt + \theta_t^r \sigma dW_t \right].$$

The value J_T is the terminal payoff from the self-financing portfolio strategy.

We have a market with two assets and uncertainty driven by a univariate Brownian motion

- one risk-free and one risky with a nontrivial volatility of the price, with portfolio strategy θ^f, θ^r

This market is so-called complete.

- consider an arbitrary time-T payoff G_T that is $\mathcal{F}_T$-measurable
- market completeness means that any such payoff G_T can be replicated as an outcome of a suitably chosen dynamic portfolio strategy θ^f, θ^r, with a particular amount of initial wealth.

Hence every other security with a given time-T payoff is so-called redundant.

We are interested in pricing a security with terminal payoff at time T equal to $G(Q_T)$.

Since the payoff is a function of the underlying stock price, such a security is called derivative.

- Typical examples of derivative securities are options. A call option with strike price K has payoff

$$G(Q_T) = \max(Q_T - K, 0) \equiv (Q_T - K)_+ \,, \tag{5.19}$$

  while a put option with strike price K has payoff

$$G(Q_T) = \max(K - Q_T, 0) \equiv (K - Q_T)_+ \,. \tag{5.20}$$

- The term option comes from the fact that, for example in the case of a call option, its payoff is equivalent to the right to buy the underlying stock at time T for the price K.

We want to infer the price of the derivative security at time $t \leq T$.

- the time-T payoff $G(Q_T)$ of the derivative security only depends on Q_T
- the interest rate r is constant
- the distribution of the future stock price conditional on time-t information only depends on Q_t
- we also need to explicitly encode time, to measure time remaining to maturity

We can therefore conjecture that the time-t price can be written as $g(Q_t, t)$, where g is a pricing function we need to derive.

Pricing function for the derivative security

Since the market is complete, the derivative security is redundant.

- its payoff can be achieved by a suitable dynamic portfolio strategy in the bond and stock

It follows from absence of arbitrage that the price $g(Q_t, t)$ must be equal to the value of the portfolio needed to replicate the same terminal payoff $G(Q_T)$.

- if it were not, then a strategy that would purchase the cheaper asset or portfolio while selling the more expensive one would generate immediate positive payoff without any future financial consequences

The central argument is to characterize the replicating portfolio.

Replication argument

We need to determine the portfolio positions that generate the replicating portfolio.

- based on a dynamic hedging argument (pointed out to Black and Scholes by Robert Merton, see Footnote 3 in Black and Scholes (1973))
- the idea is to find a particular combination of the bond and stock such that the infinitesimal return is the same as the infinitesimal return on the derivative security
- extending the infinitesimal argument to finite horizon T yields the desired answer

If the portfolio replicates the returns 'step-by-step', it also has to replicate the time T payoff.

Replication argument

We develop the idea in an equivalent way, from a slightly different angle.

- construct a portfolio consisting of a particular combination of the stock and the derivative that makes the return on this portfolio risk-free, over an infinitesimal horizon
- since the portfolio is risk-free, it must earn the risk-free rate r, otherwise an arbitrage opportunity would emerge

Let such a self-financing portfolio consist of

- one option with current price $g(Q_t, t)$
- a position of θ_t^r units of the risky stock with price Q_t

The value of this portfolio is

$$1 \cdot g(Q_t, t) + \theta_t^r Q_t.$$

By the self-financing assumption, the infinitesimal financial gain is

- $\theta_t^r dQ_t$ on stock portion of this portfolio
- $1 \cdot dg\left(Q_t, t\right)$ on the option portion.

An application of Itô's lemma implies that

$$\begin{aligned} dg\left(Q_t, t\right) &= g_Q\left(Q_t, t\right) dQ_t + \frac{1}{2} g_{QQ}\left(Q_t, t\right)\left(dQ_t\right)^2 + g_t\left(Q_t, t\right) dt \\ &= \left[g_Q\left(Q_t, t\right) \mu Q_t + \frac{1}{2} g_{QQ}\left(Q_t, t\right) \sigma^2 + g_t\left(Q_t, t\right)\right] dt + g_Q\left(Q_t, t\right) \sigma Q_t dW_t. \end{aligned}$$

The evolution of the value of the portfolio is therefore given by

$$\begin{aligned} dg\left(Q_t, t\right) + \theta_t^r dQ_t &= \left[\left(g_Q\left(Q_t, t\right) + \theta_t^r\right) \mu Q_t + \frac{1}{2} g_{QQ}\left(Q_t, t\right) \sigma^2 Q_t^2 + g_t\left(Q_t, t\right)\right] dt \\ &\quad + \left[g_Q\left(Q_t, t\right) + \theta_t^r\right] \sigma Q_t dW_t. \end{aligned}$$

The evolution of the value of the portfolio:

$$
\begin{aligned}
dg\left(Q_t, t\right) + \theta_t^r dQ_t \quad = \quad & \left[\left(g_Q\left(Q_t, t\right) + \theta_t^r\right) \mu Q_t + \frac{1}{2} g_{QQ}\left(Q_t, t\right) \sigma^2 Q_t^2 + g_t\left(Q_t, t\right)\right] dt \\
& + \underbrace{\left[g_Q\left(Q_t, t\right) + \theta_t^r\right]}_{\text{risk exposure}} \sigma Q_t dW_t .
\end{aligned}
$$

We want to choose θ_t^r to make the gain on the portfolio locally risk-free

- this corresponds to a zero risk exposure of the financial gain

This implies we must choose

$$\theta_t^r = -g_Q\left(Q_t, t\right).$$

With the choice $\theta_t^r = -g_Q(Q_t, t)$, the financial gain on the portfolio is equal to

$$dg(Q_t, t) - g_Q(Q_t, t)\, dQ_t = \left[\frac{1}{2} g_{QQ}(Q_t, t)\, \sigma^2 Q_t^2 + g_t(Q_t, t)\right] dt. \tag{5.21}$$

Absence of arbitrage implies that this portfolio then must earn the risk-free rate r, and hence also

$$dg(Q_t, t) - g_Q(Q_t, t)\, dQ_t = r\left[g(Q_t, t) - g_Q(Q_t, t)\, Q_t\right] dt. \tag{5.22}$$

Equalizing the drift terms on the right-hand sides of (5.21) and (5.22), and writing Q instead of Q_t, we obtain the equation

$$rg(Q, t) = g_t(Q, t) + g_Q(Q, t)\, rQ + \frac{1}{2} g_{QQ}(Q, t)\, \sigma^2 Q^2.$$

This is a second-order partial differential equation for the price of the derivative security $g(Q, t)$.

A second-order PDE for the value of the derivative

Second-order PDE for $g(Q,t)$:

$$rg(Q,t) = g_t(Q,t) + g_Q(Q,t)\,rQ + \frac{1}{2}g_{QQ}(Q,t)\,\sigma^2 Q^2. \tag{5.23}$$

This second-order PDE has a terminal boundary condition $g(Q,T) = G(Q)$.

- the price of the derivative security at maturity time T is equal to the payoff $G(Q)$

The PDE does not depend on the expected return on the stock μ.

- this is a path-breaking result shown by Black and Scholes (1973)
- the risk-free rate r is directly observable and the volatility of risky asset returns σ can be reasonably inferred from high-frequency data
- measuring the expected return on a risky asset μ is an inherently difficult task

Independence of μ is the result of the replication argument combined with absence of arbitrage.

- this argument carries over to a variety of extensions as well

The prices of the call and put options with payoffs (5.19)–(5.20) can be determined as closed-form expressions which only depend on quantiles of the normal distribution.

Proposition 5.11

Time-t prices of European call and put options with payoffs (5.19) and (5.20), respectively, with strike price K and maturity T, are given by

$$\begin{aligned} C(Q,t) &= QN(z_1) - \exp(-r(T-t))KN(z_2) \\ P(Q,t) &= \exp(-r(T-t))KN(-z_2) - QN(-z_1) \end{aligned}$$

where $N(\cdot)$ is the cumulative standard normal distribution function, and

$$\begin{aligned} z_1 &= \frac{\log\left(\frac{Q}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} \\ z_2 &= z_1 - \sigma\sqrt{T-t}. \end{aligned}$$

It can be verified that $C(Q,t)$ and $P(Q,t)$ satisfy the partial differential equation (5.23).

Put-call parity

Given a strike price K, it is sufficient to compute only the price of one of the options because the call and put option price are related through the so-called put-call parity

$$P(Q_t, t) + Q_t = C(Q_t, t) + K \exp(-r(T - t)) . \tag{5.24}$$

The put-call parity result is based on a replication argument.

- the left-hand side of (5.24) is the value of a portfolio consisting of a put option and the stock, with payoff $\max(K - Q_T, 0) + Q_T = \max(K, Q_T)$
- the right-hand side is the value of a portfolio invested in a call option and a risk-free investment with face value K, with total payoff $\max(Q_T - K, 0) + K = \max(Q_T, K)$

Portfolios on both sides of the equation have identical payoffs at time T.

- by the no-arbitrage argument, they must also have the same time-t valuation.
- $P(Q_t, t)$ and $C(Q_t, t)$ are the prices of the options, Q_t is the stock price, and $K \exp(-r(T - t))$ is the time-t value of the risk-free investment.

For computational purposes, it may be useful to transform the state variable in the PDE

- use $q = \log Q$ instead of Q
- more suitable for an equidistant grid when Q follows a geometric Brownian motion

Define the transformation $f(q,t) = f(\log Q, t) = g(Q,t) = g(\exp(q), t)$:

$$
\begin{aligned}
f_q(q,t) &= \frac{d}{dq} g(\exp(q), t) = g_Q(\exp(q), t)\exp(q) = g_Q(Q,t)\,Q \\
f_{qq}(q,t) &= \frac{d}{dq}\left(g_Q(\exp(q), t)\exp(q)\right) = g_{QQ}(Q,t)\,Q^2 + g_Q(Q,t)\,Q.
\end{aligned}
$$

- the partial differential equation is transformed to

$$
rf(q,t) = f_t(q,t) + \left(r - \frac{1}{2}\sigma^2\right) f_q(q,t) + \frac{1}{2}\sigma^2 f_{qq}(q,t)
$$

with the terminal boundary condition $f(q,t) = G(\exp(q))$.

Finite-difference method

We study numerical solutions to a general class of PDEs

$$h(x,t) \underbrace{-v(x,t)\,r(x,t) + v_x(x,t)\,\mu(x,t) + \frac{1}{2} v_{xx}(x,t)\,\sigma(x,t)^2}_{\doteq \mathcal{D}v(x,t)} + v_t(x,t) = 0 \tag{5.25}$$

- state-time space $\mathcal{X} \times \mathcal{T} = [l, r] \times [0, T]$
- $v(x,t)$ unknown function, $h(x,t)$, $r(x,t)$, $\mu(x,t)$, $\sigma(x,t)$ known parameters
- terminal condition $v(x,T) = H(x,T)$
- boundary condition

$$\alpha(x,t)\,v_x(x,t) + \beta(x,t)\,v(x,t) = \gamma(x,t) \qquad x \in \{l, r\}, t \in [0, T] \tag{5.26}$$

Relationship to the Black–Scholes problem

The Black–Scholes PDE is (almost) a special case of (5.25):

$$rg(Q,t) = g_t(Q,t) + g_Q(Q,t)\,rQ + \frac{1}{2}g_{QQ}(Q,t)\,\sigma^2 Q^2 \tag{5.27}$$

so we have

$$r(x,t) = r,\ \mu(x,t) = rx,\ \sigma(x,t) = \sigma x,\ h(x,t) = 0,\ H(x,T) = G(x).$$

The issue is the state space: $\mathcal{X} = (0,\infty)$ in the Black–Scholes model

- boundary conditions need to be determined using other considerations
- a heuristic approach: specify a sufficiently 'wide' interval $[l,r]$ and approximate the boundary condition (5.26) with (5.27), setting $g_{QQ}(Q,t) = g_t(Q,t) = 0$
- then we have $\beta(x,t) = r$, $\alpha(x,t) = -rx$, and $\gamma(x,t) = 0$
- heuristic: nonlinearity vanishes in the tails + solution interior to choice of 'distant' boundaries

The Feynman–Kac formula relates the solution of the PDE for $v(x,t)$

$$h(x,t) \underbrace{- v(x,t)\, r(x,t) + v_x(x,t)\, \mu(x,t) + \frac{1}{2} v_{xx}(x,t)\, \sigma(x,t)^2}_{\doteq \mathcal{D}v(x,t)} + v_t(x,t) = 0 \tag{5.28}$$

with terminal condition $v(x,T) = H(x,T)$ to the present value problem

$$\begin{aligned} v(x,t) &= E\left[\int_t^T \phi(t,s)\, h(X_s,s)\, ds + \phi(t,T)\, H(X_T,T) \mid X_t = x\right] \\ \phi(t,s) &= \exp\left(-\int_t^s r(X_\tau,\tau)\, d\tau\right) \\ dX_t &= \mu(X_t,t)\, dt + \sigma(X_t,t)\, dW_t \qquad \text{on } \mathcal{X} = (l,r) \subseteq \mathbb{R}. \end{aligned} \tag{5.29}$$

We can map the Black–Scholes PDE

$$rg(Q,t) = g_t(Q,t) + g_Q(Q,t)\,rQ + \frac{1}{2}g_{QQ}(Q,t)\,\sigma^2 Q^2$$

with terminal condition $g(Q,T) = G(T)$ to the Feynman–Kac formula.

This implies that the solution to the Black–Scholes problem can be equivalently written as

$$g(Q_t,t) = E\left[e^{-r(T-t)}G(Q_T)\,|Q_t\right] \tag{5.30}$$

with Q_t following the dynamics

$$dQ_t = rQ_t dt + \sigma Q_t dW_t^*. \tag{5.31}$$

Equation (5.30) is the present value of $G(Q_T)$, discounted by a hypothetical SDF

$$\frac{S_T^*}{S_t^*} = e^{-r(T-t)}$$

with dynamics of the stock price modified to (5.31)

- expected return on the stock equal to r instead of $\mu \implies$ risk-neutral pricing

We have the PDE in the form

$$-v_t(x,t) = \mathcal{D}v(x,t) + h(x,t)$$

with $\mathcal{D}v(x,t)$ given in (5.25), terminal condition $v(x,T) = H(x,T)$, and a general boundary condition

$$\alpha(x,t)\, v_x(x,t) + \beta(x,t)\, v(x,t) = \gamma(x,t) \qquad x \in \{l, r\}, t \in [0,T]$$

- allows to incorporate various types of boundary behavior

The idea is to overlay a grid of points over the rectangle $[l, r] \times [0, T]$, approximate derivatives with differences, and turn the problem to an algebraic system of linear equations.

- The PDE is of the so-called 'parabolic' type $\implies$ allows solving the problem in 'time layers'.
  - classification based on coefficients on the second order terms

$$Av_{xx} + Bv_{xt} + Cv_{tt} + \text{lower order terms} = 0$$

  $B^2 - AC < 0$ elliptic (models of static equilibria), $= 0$ parabolic (heat dissipation), > 0 hyperbolic (wave propagation)

Equidistant grid

Construct grids

- space dimension $i = \{0, \dots, I\}$, step size $\Delta x = (r - l)/I$
- time dimension $j = \{0, \dots, J\}$, step size $\Delta t = T/J$
- denote $v_{i,j} = v(i\Delta x, j\Delta t)$

Considerations

- What if l, r are infinite? Need to choose a suitable truncation.
- Equidistant grids are not the only choice. More complicated notation.
- Often a change of variable (logs vs levels) is a better adjustment than choosing non-equidistant grids.

$$
\begin{array}{ccccccc}
\boxed{x = l} & v_{0,0} & \cdots & v_{0,j-1} & v_{0,j} & \cdots & v_{0,J} \\
 & \vdots & & \vdots & \vdots & & \vdots \\
 & v_{i-1,0} & & v_{i-1,j-1} & v_{i-1,j} & & v_{i-1,J} \\
\Delta x \updownarrow & v_{i,0} & & v_{i,j-1} & v_{i,j} & & v_{i,J} \\
 & v_{i+1,0} & & v_{i+1,j-1} & v_{i+1,j} & & v_{i+1,J} \\
 & \vdots & & \vdots & \vdots & & \vdots \\
\boxed{x = r} & v_{I,0} & \cdots & v_{I,j-1} & v_{I,j} & \cdots & v_{I,J} \\
 & & & \longleftarrow & \Delta t \longrightarrow & & \\
 & \boxed{t = 0} & & & & & \boxed{t = T}
\end{array}
$$

- terminal condition $v(x, T) = G(x, T)$ for $x \in [l, r]$
- boundary conditions $v(l, t)$ and $v(r, T)$ for $t \in [0, T]$

We replace derivatives v_x, v_{xx} and v_t at an interior node (i, j) with differences.

Approximation of first-order derivative v_x at $x = i\Delta x$ and $t = j\Delta t$:

$$\begin{aligned}
\text{forward difference} \quad &: \quad v_x(i\Delta x, j\Delta t) \approx v^{\bar{x}}_{i,j} \doteq \frac{1}{\Delta x}\left(v_{i+1,j} - v_{i,j}\right) \\
\text{central difference} \quad &: \quad v_x(i\Delta x, j\Delta t) \approx v^{x_c}_{i,j} \doteq \frac{1}{2\Delta x}\left(v_{i+1,j} - v_{i-1,j}\right) \\
\text{backward difference} \quad &: \quad v_x(i\Delta x, j\Delta t) \approx v^{\underline{x}}_{i,j} \doteq \frac{1}{\Delta x}\left(v_{i,j} - v_{i-1,j}\right)
\end{aligned}$$

- which difference is used sometimes matters a lot (see upwind scheme)

Approximation of second-order derivative v_{xx} at $x = i\Delta x$ and $t = j\Delta t$:

$$v_{xx}(i\Delta x, j\Delta t) \approx v^{\bar{x}\underline{x}}_{i,j} = \frac{1}{\Delta x}\left(v^{\bar{x}}_{i,j} - v^{\bar{x}}_{i,j-1}\right) = \frac{1}{(\Delta x)^2}\left(v_{i+1,j} - 2v_{i,j} + v_{i,j-1}\right)$$

Collecting terms, we replace

$$\mathcal{D}v\left(x,t\right) = -v\left(x,t\right)r\left(x,t\right) + v_x\left(x,t\right)\mu\left(x,t\right) + \frac{1}{2}v_{xx}\left(x,t\right)\sigma\left(x,t\right)^2$$

at $(x,t) = (i\Delta x, j\Delta t)$

$$\left(Dv\right)_{i,j} = -v_{i,j}r_{i,j} + v_{i,j}^{\bar{x}}\mu_{i,j} + \frac{1}{2}v_{i,j}^{\bar{x}x}\sigma_{i,j}^2$$

- here, we used forward difference $v_{i,j}^{\bar{x}}$ as an example

At the boundaries, proceed in the same way.

- use forward difference at $x = l$ and backward difference at $x = r$

$$
\begin{aligned}
\alpha_{0,j} v_{0,j} + \beta_{0,j} v^{\bar{x}}_{0,j} &= \gamma_{0,j} \\
\alpha_{I,j} v_{I,j} + \beta_{I,j} v^{\underline{x}}_{I,j} &= \gamma_{I,j}
\end{aligned}
$$

- solve for $v_{0,j}$ and $v_{I,j}$ and use it to substitute out $v_{0,j}$ at node $i = 1$ and $v_{I,j}$ at node $i = I - 1$.

Recall that the (known) coefficients α, β, γ will depend on the economics of the problem and boundary behavior of X_t.

If the boundary condition also contains time derivative v_t, then treat the boundary points $i \in \{0, I\}$ in the same way as interior points.

In the time dimension, we proceed in the same way

$$
\begin{aligned}
\text{forward difference:} \qquad & v_t\,(i\Delta x, j\Delta t) \approx v^{\bar{t}}_{i,j} \doteq \frac{1}{\Delta t}\,(v_{i,j+1} - v_{i,j}) \\
\text{backward difference:} \qquad & v_t\,(i\Delta x, j\Delta t) \approx v^{\underline{t}}_{i,j} \doteq \frac{1}{\Delta t}\,(v_{i,j} - v_{i,j-1})
\end{aligned}
$$

The choice will determine two difference solution methods

- forward difference $\Longrightarrow$ implicit scheme
- backward difference $\Longrightarrow$ explicit scheme

We characterize stability properties of these choices in some simple cases.

The PDE approximation at node (i, j) using the backward time difference is

$$-v_{i,j}^{t} = (Dv)_{i,j} + h_{i,j}$$

which we can write as

$$v_{i,j-1} = (\Delta t)\,(Dv)_{i,j} + v_{i,j} + (\Delta t)\,h_{i,j}$$

If we know the solution at nodes

$$(i,j) \;\text{ for all } i \in \{1, \ldots, I-1\} \qquad \text{[i.e., at time } t = j\Delta t\text{]}$$

we can explicitly compute the solution at nodes

$$(i,j-1) \;\text{ for all } i \in \{1, \ldots, I-1\} \qquad \text{[i.e., at time } t = (j-1)\,\Delta t\text{]}.$$

$$
\begin{array}{c|cccccc}
\boxed{x = l} & v_{0,0} & \cdots & v_{0,j-1} & v_{0,j} & \cdots & v_{0,J} \\
 & \vdots & & \vdots & \vdots & & \vdots \\
 & v_{i-1,0} & & v_{i-1,j-1} & v_{i-1,j} & & v_{i-1,J} \\
 & v_{i,0} & & v_{i,j-1} & v_{i,j} & & v_{i,J} \\
 & v_{i+1,0} & & v_{i+1,j-1} & v_{i+1,j} & & v_{i+1,J} \\
 & \vdots & & \vdots & \vdots & & \vdots \\
\boxed{x = r} & v_{I,0} & \cdots & v_{I,j-1} & v_{I,j} & \cdots & v_{I,J} \\
 & \boxed{t = 0} & & & & & \boxed{t = T}
\end{array}
$$

- equation at node (i, j) involves known values $v_{i-1,j}$, $v_{i,j}$, $v_{i+1,j}$ and an unknown value $v_{i,j-1}$

The system of equation is linear in $v_{i,j}$, and we can write it in matrix form.

$$v^{int}_{\cdot,j-1} = A_j v^{int}_{\cdot,j} + \widetilde{h}_j$$

- A_j in an $(I-1) \times (I-1)$ matrix
- $v^{in}_{\cdot,j} = (v_{1,j}, \dots v_{I-1,j})'$ is the vector for the solution at interior nodes
- $\widetilde{h}_j$ is an $(I-1) \times 1$ vector

For notational simplicity, restrict ourselves to the heat equation case

$$r(x,t) = \mu(x,t) = 0, \sigma(x,t) = \sigma$$

with boundary conditions

$$\alpha(x,t) = 1, \beta(x,t) = 0, \gamma(l,t) = \gamma_l, \gamma(r,t) = \gamma_r$$

In this simple case, we have

$$A = \begin{pmatrix} 1 - \frac{\Delta t}{(\Delta x)^2}\sigma^2 & \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 0 & 0 & \cdots \\ \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 1 - \frac{\Delta t}{(\Delta x)^2}\sigma^2 & \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 0 & \cdots \\ 0 & \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 1 - \frac{\Delta t}{(\Delta x)^2}\sigma^2 & \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & \cdots \\ 0 & 0 & \ddots & \ddots & \ddots \end{pmatrix}$$

$$\widetilde{h}_j = \begin{pmatrix} (\Delta t)\, h_{i,j} + \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2}\gamma_l \\ (\Delta t)\, h_{i,j} \\ \vdots \\ (\Delta t)\, h_{i,j} \\ (\Delta t)\, h_{i,j} + \frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2}\gamma_h \end{pmatrix}$$

Even in the general case, A remains tri-diagonal.

In the explicit scheme (with $A_j = A$), solving the problem corresponds to iterating backward

$$v^{int}_{\cdot,0} = Av^{int}_{\cdot,1} + \widetilde{h}_1 = A^2 v^{int}_{\cdot,2} + A\widetilde{h}_2 v^{int}_{\cdot,1} + \widetilde{h}_1 = \dots$$

Stability of the explicit scheme depend on the eigenvalues of A.

- eigenvalues of this matrix are given by

$$\lambda_i = 1 - 2\frac{\Delta t}{(\Delta x)^2}\sigma^2 \left(\sin\frac{i\pi}{2I}\right)^2 \qquad i \in \{1, \dots, I-1\}$$

- we require $|\lambda_i| < 1$, $\forall i$, i.e.,

$$-1 < 1 - 2\frac{\Delta t}{(\Delta x)^2}\sigma^2 \left(\sin\frac{i\pi}{2I}\right)^2 < 1$$

- it is sufficient to choose Δt and Δx so that they satisfy

$$\sigma^2 \Delta t < (\Delta x)^2$$

Explicit scheme is therefore conditionally stable

- we should choose the space and time grids suitably

$$\sigma^2 \Delta t < (\Delta x)^2$$

- stability assures that small inaccuracies in the solution do not explode as we iterate

In the more general case with non-constant coefficients, it is not easy to characterize sufficient conditions explicitly, but the general intuition (stability of matrices A_j) still holds.

The PDE approximation at node $(i, j-1)$ using the forward time difference is

$$-v^{\bar{t}}_{i,j-1} = (Dv)_{i,j-1} + h_{i,j-1}$$

which we can write as

$$v_{i,j-1} - (\Delta t)\,(Dv)_{i,j-1} = v_{i,j} + (\Delta t)\,h_{i,j-1}$$

We again apply the same principle. If we know the solution at nodes

$$(i,j) \;\text{ for all } i \in \{1,\ldots,I-1\} \qquad [\text{i.e., at time } t = j\Delta t]$$

we can explicitly compute the solution at nodes

$$(i,j-1) \;\text{ for all } i \in \{1,\ldots,I-1\} \qquad [\text{i.e., at time } t = (j-1)\,\Delta t].$$

- equation in each node $(i, j-1)$ now involves three unknowns $v_{i-1,j-1}$, $v_{i,j-1}$, $v_{i+1,j-1}$

Implicit scheme

$x = l$	$v_{0,0}$	$\cdots$	$v_{0,j-1}$	$v_{0,j}$	$\cdots$	$v_{0,J}$
	$\vdots$		$\vdots$	$\vdots$		$\vdots$
	$v_{i-1,0}$		$v_{i-1,j-1}$	$v_{i-1,j}$		$v_{i-1,J}$
	$v_{i,0}$		$v_{i,j-1}$	$v_{i,j}$		$v_{i,J}$
	$v_{i+1,0}$		$v_{i+1,j-1}$	$v_{i+1,j}$		$v_{i+1,J}$
	$\vdots$		$\vdots$	$\vdots$		$\vdots$
$x = r$	$v_{I,0}$	$\cdots$	$v_{I,j-1}$	$v_{I,j}$	$\cdots$	$v_{I,J}$
	$t = 0$					$t = T$

- equation at node $(i, j-1)$ involves a known value $v_{i,j}$ and three unknown known values $v_{i-1,j-1}$, $v_{i,j-1}$, $v_{i+1,j-1}$

In matrix form, we now have

$$A_j v^{int}_{\cdot,j-1} = v^{int}_{\cdot,j} + \widetilde{h}_j$$

where, for the simple heat equation case,

$$A_j = A = \begin{pmatrix} 1 + \frac{\Delta t}{(\Delta x)^2}\sigma^2 & -\frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 0 & 0 & \cdots \\ -\frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 1 + \frac{\Delta t}{(\Delta x)^2}\sigma^2 & -\frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 0 & \cdots \\ & -\frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & 1 + \frac{\Delta t}{(\Delta x)^2}\sigma^2 & -\frac{\Delta t}{(\Delta x)^2}\frac{\sigma^2}{2} & \cdots \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

Iterating on the implicit scheme now requires a matrix inversion

$$v^{int}_{\cdot,j-1} = (A_j)^{-1} \left(v^{int}_{\cdot,j} + \widetilde{h}_j \right)$$

- computationally cheap when A_j does not depend on j
- even in the general case, A_j is still only a tri-diagonal matrix

It can be shown that for the simple heat equation case, eigenvalues of A^{-1} lie in the unit circle.

- scheme is unconditionally stable
- this does not automatically generalize but intuition still holds

Another concern is the stability of the difference scheme with respect to the behavior of the first derivative v_x.

- this is an issue well known in computational fluid dynamics
- (conditional) stability of the explicit scheme depends on the relationship between the sign of $\mu(x, t)$ and the choice of forward or backward derivative in the approximation of v_x

See notes for details.

1. How does the accuracy of the method depend on the grid choice?
   - get insights from Taylor series approximation of $v(x, t)$
   - more accurate schemes involve approximations using more than just the adjacent nodes
2. What if the PDE is nonlinear in v_x or v_{xx}?
   - this often happens in optimization problems
   - explicit scheme will still work, but the implicit scheme would involve inverting a nonlinear operator

Summary

Summary

We have developed a model for the pricing of derivative securities.

- the argument is based on the combination of dynamic hedging and absence of arbitrage
- Black and Scholes (1973) provided a characterization in continuous time but the substance of the problem carries over to other environments as well

In the continuous-time Brownian information setup, the characterization leads to a second-order PDE for the price of the security.

- there is a variety of methods for solving such PDEs
- we analyzed a method based on the discretization of time and state space using finite differences
- this is a versatile method, even though it suffers from the same curse of dimensionality as other grid methods

Appendix

Literature I

Black, Fischer and Myron Scholes (1973) "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81 (3), 637–654.

Candler, Graham V. (2001) "Finite-Difference Methods for Continuous-Time Dynamic Programming," in *Computational Methods for the Study of Dynamic Economies*, Chap. 8, 172–194: Oxford University Press.

Cox, John C., Stephen A. Ross, and Mark Rubinstein (1979) "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, 7 (3), 229–263.

Duffie, Darrell (2001) *Dynamic Asset Pricing Theory*: Princeton University Press, 3rd edition.

Holmes, Mark H. (2007) *Introduction to Numerical Methods in Differential Equations*: Springer.

Itô, Kiyosi (1951) "On a Formula Concerning Stochastic Differentials," *Nagoya Mathematical Journal*, 3, 55–65.

Judd, Kenneth L. (1998) *Numerical Methods in Economics*: The MIT Press, Cambridge, MA.

Merton, Robert C. (1973) "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4 (1), 141–183.

Øksendal, Bernt (2007) *Stochastic Differential Equations*: Springer Verlag, 6th edition.

Ross, Stephen A. (1976) "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13 (3), 341–360.

Thomas, James W. (1995) *Numerical Partial Differential Equations: Finite Difference Methods*: Springer Verlag New York.