MS455/555 Simulation for Finance

Denis Patterson

January 18, 2018

Acknowledgements

These notes are based on a course taught at Dublin City University to final year Actuarial and Financial Maths students. Earlier iterations of this course were developed by Dr. Olaf Menkens and these notes are partially derived from course notes due to Dr. Eberhard Mayerhofer, particularly the first three chapters.

All errors, omissions, and failings of judgement are my own. I would appreciate if errata were sent to denis.patterson2@mail.dcu.ie.

Denis Patterson, September 4th, 2017.

Contents

Chapter 1

Probability Theory Review

1.1 Fundamentals

We work in $\mathbb{R}$ for ease of exposition but all results and statements which follow have natural (and mostly obvious) analogues in $\mathbb{R}^n$.

If Ω is an uncountable set, such as $\mathbb{R}$, it turns out that we cannot hope to assign measures (resp. probabilities) to all subsets of Ω and retain a mathematically (or physically) reasonable theory. Hence we restrict ourselves to "sensible" classes of subsets of Ω and the following definition supplies precisely the correct notion of "sensible".

Definition 1.1.1 (σ–algebra)**.** *If Ω is any non–empty set, then Σ is a σ–algebra on Ω if it has the following properties:*

(i.) $\emptyset \in \Sigma$ *and* $\Omega \in \Sigma$,

(ii.) if $E \in \Sigma$, *then* $E^c \in \Sigma$ *(where E^c is the complement of E),*

(iii.) if $\{E_n\}_{n\geq 1}$ *are all in* Σ, *then* $\cup_{n\geq 1} E_n \in \Sigma$.

The Borel σ–algebra on $\mathbb{R}$ is the smallest σ–algebra on $\mathbb{R}$ which contains all the open sets and is henceforth denoted by $\mathcal{B}(\mathbb{R})$. If $E \subset \mathbb{R}$ is in $\mathcal{B}(\mathbb{R})$, then it is called a Borel set.

Definition 1.1.2 (Borel measurable function)**.** *If (Ω, Σ) is a measure space, i.e. Σ is a σ–algebra on the set Ω, and $f : \Omega \mapsto \mathbb{R}$, then f is Borel measurable if*

$$f^{-1}(E) = \{x : f(x) \in E\} \in \Sigma \quad \text{for each } E \in \mathcal{B}(\mathbb{R}).$$

Intuitively, measurable functions map measurable sets to measurable sets. When we say that a real–valued function is measurable we will always mean Borel measurable.

Definition 1.1.3 (Random variable)**.** *Let (Ω, Σ) be a measure space. A random variable (r.v. for short) is a map*

$$X: \quad \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega)$$

which is measurable.

Example 1.1.4. *We model a cointoss. $\Omega = \{H, T\}$, where H represents the event "head", and T represents "tail". The σ–algebra Σ is the one generated by all events, i.e. the power set of Ω,*

$$\Sigma = \{\emptyset = \{\}, \{H, T\} = \Omega, \{H\}, \{T\}\}.$$

A cointoss could be defined as the random variable

$$X : H \mapsto 1, \quad T \mapsto 0. \tag{1.1.1}$$

But any other choice could be reasonable as well. For example,

$$X : H \mapsto 1, \quad T \mapsto -1.$$

Definition 1.1.5 (Probability Measure)**.** *If (Ω, Σ) is a measure space, then $\mathbb{P} : \Sigma \mapsto [0, 1]$ is a probability measure on this space if it satisfies:*

(i.) $\mathbb{P}[\Omega] = 1$,

(ii.) if $\{E_n\}_{n\geq 1}$ are all in Σ and pairwise disjoint, then $\mathbb{P}\left[\cup_{n\geq 1} E_n\right] = \sum_{n\geq 1} \mathbb{P}[E_n]$.

Suppose we have a probability measure $\mathbb{P}$ defined on (Ω, Σ), which means we make the latter a probability space $(\Omega, \Sigma, \mathbb{P})$. The distribution of a random variable X can be defined as follows:

Definition 1.1.6. *The cumulative distribution function (c.d.f for short) of X is defined as the function*

$$F_X: \quad \mathbb{R} \to [0, 1]$$

given by

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}).$$

Example 1.1.7. *Let us continue with the cointoss described by X in (1.1.1). The c.d.f. of X is given by*

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

because, for $x < 0$

$$F_X(x) = \mathbb{P}(\emptyset) = 0,$$

and for $x < 1$,

$$F_X(x) = \mathbb{P}(\{T\}) = \frac{1}{2}.$$

We now introduce the uniform distribution on the unit interval $I = [0, 1]$.

Definition 1.1.8. *A random variable U is uniformly distributed on $[0, 1]$, if for any $0 \leq a \leq b \leq 1$ we have*

$$\mathbb{P}\left[U \in [a, b]\right] = b - a.$$

We write $U \sim U([0, 1])$.

Lemma 1.1.9. *The distribution function of $U \sim U([0, 1])$ is given by*

$$F_U(u) = \begin{cases} 0, & u \leq 0, \\ u, & u \in (0, 1), \\ 1, & u \geq 1. \end{cases}$$

Proof. For $x \leq 0$ or $x \geq 1$, the claim follows directly from the fact, that U only takes values in the unit interval. For $0 \leq x \leq 1$, by the very definition of the uniform distribution, we have

$$F_U(u) = \mathbb{P}[U \leq u] = \mathbb{P}[U \in [0, u]] = u - 0 = u.$$

□

Definition 1.1.10. *The distribution $F_X(x)$ of X has a density if there exists a continuous function $\xi \mapsto f_X(\xi)$ which satisfies*

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi.$$

We call f_X the probability density function or p.d.f of X.

Lemma 1.1.11. *If $x \mapsto F_X(x)$ is continuously differentiable, then it has a density which is given by*

$$f_X(\xi) = \frac{dF_X}{dx}(\xi).$$

Proof. By the fundamental theorem of calculus,

$$\int_{-\infty}^{x} \frac{dF_X}{dx}(u) du = F_X(x) - F_X(-\infty) = F_X(x) - 0 = F_X(x).$$

□

Remark 1.1.12. *If X is a random variable on $\mathbb{R}$ with continuous density f, then*

$$\mathbb{P}[X = x] = \int_{x}^{x} f(s)\, ds = 0, \quad x \in \mathbb{R}.$$

Clearly, a direct interpretation of the p.d.f as the analogue of the probability mass function for a discrete random variable breaks down immediately. However, there is some valuable intuition which can still be salvaged from this seemingly flawed analogy. Note that

$$\frac{1}{2\epsilon}\mathbb{P}[X \in (x-\epsilon, x+\epsilon)] = \frac{1}{2\epsilon}\int_{x-\epsilon}^{x+\epsilon} f(s)\,ds \longrightarrow f(x) \quad \text{as } \epsilon \to 0^+, \quad x \in \mathbb{R}.$$

In other words, the probability that X lies in a neighbourhood of x is proportional to $f(x)$ (modulo the appropriate scaling). Motivated by the intuition given above, we will calculate the "empirical density of a sample" (of random numbers) by creating histograms with very small intervals in the hope of approximating the true density (see Figure 1.1.12).

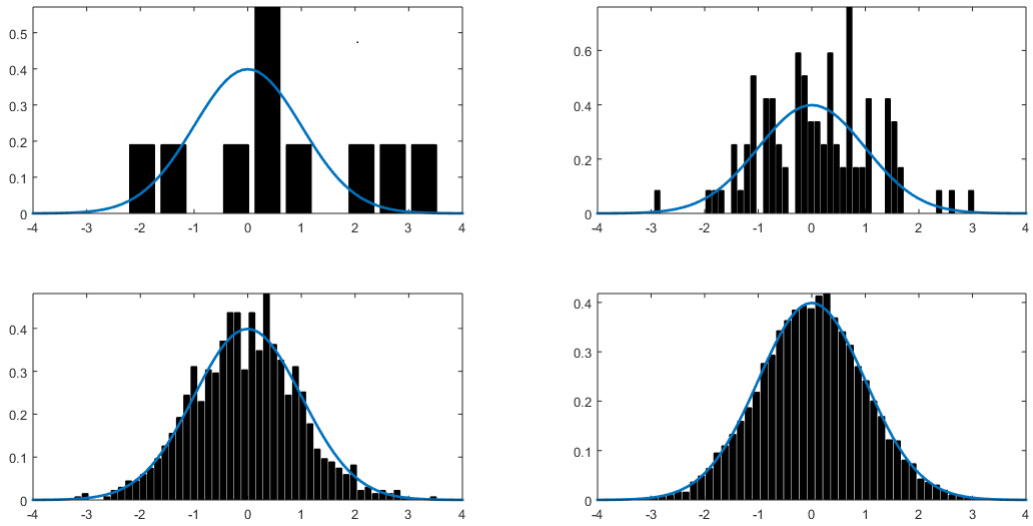


Figure 1.1: Each plot is a histogram (with very small width bins) based on an $N \times 1$ vector of i.i.d. $\mathcal{N}(0,1)$ random variables overlaid with the true p.d.f for an $\mathcal{N}(0,1)$ distribution ($N = 10, 100, 1000, 10000$). We must normalise the histogram by the size of the sample so that the area under the empirical p.d.f is (approximately) 1.

In general, we are interested in constructing random variables with a given distribution $F(x)$ from uniformly distributed random variables; the following lemma is invaluable in this regard.

Lemma 1.1.13. *Let U be a uniformly distributed random variable, and let $F(x)$ be a c.d.f. If F is invertible, then $X = F^{-1}(U)$ is distributed according to F.*

Proof. Elementary and given below, but be sure you can justify each equality:

$$\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x), \quad x \in \mathbb{R}.$$

□

Definition 1.1.14 (Expectation)**.** *We choose to define the expectation $\mathbb{E}[X]$ of a real–valued random variable X which has continuous density f as follows:*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

When is a function of a random variable still a well–defined random variable?

Lemma 1.1.15. *Suppose X and Y are independent, identically distributed random variables (i.d.d. for short), and let g be a real–valued measurable function. Then $g(X)$, $g(Y)$ are also independent, identically distributed random variables.*

Proof. Since X and Y are measurable (by definition) so are the compositions $g(X)$ and $g(Y)$. Therefore $X_1 = g(X), Y_1 = g(Y)$ are random variables as well.

If A, B are Borel sets, then

$$\begin{aligned}\mathbb{P}[g(X) \in A, g(Y) \in B] &= \mathbb{P}[X \in g^{-1}(A), Y \in g^{-1}(B)] \\ &= \mathbb{P}[X \in g^{-1}(A)]\mathbb{P}[Y \in g^{-1}(B)] = \mathbb{P}[g(X) \in A]\mathbb{P}[g(Y) \in B]\end{aligned}$$

and hence $g(X)$ and $g(Y)$ are independent.

Next show that $g(X), g(Y)$ have the same distribution. For any Borel set A,

$$\mathbb{P}[g(X) \in A] = \mathbb{P}[X \in g^{-1}(A)] = \mathbb{P}[Y \in g^{-1}(A)] = \mathbb{P}[g(Y) \in A],$$

completing the proof. □

Lemma 1.1.16 (Law of the Unconscious Statistician)**.** *If X is a random variable with density f and g is a measurable function, then*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, ds.$$

Definition 1.1.17. *(Sample) Let X be a random variable with distribution function $F = F_X$. A sample from X with sample size $n \geq 1$ is a sequence of independent random variables $X_1, X_2, \ldots X_n$ which are i.i.d (independent, identically distributed) random variables with distribution F (probabilistic viewpoint). Statisticians would rather consider a sample as a single realization of this sequence of random variables, say*

$$X_1(\omega), \ldots, X_n(\omega)$$

where $\omega \in \Omega$. We will unashamedly use sample for both statements since we are now alert to any possible confusion.

Example 1.1.18. *The sequence of outcomes from tossing a coin* 5 *times could be*

$$H, H, T, T, H$$

which is a sample from the Binomial distribution with parameters $n = 1, p = 1/2$ *(if the coin is fair). The probabilist interprets this sample as a single realization of an i.i.d sequence of random variables*

$$X_1, X_2, \ldots, X_5.$$

1.2 Key Theorems

The theorems from this section are among the most important in probability theory and form the theoretical foundation of much of this course. Before we can state these famous results we need to recall some elementary definitions.

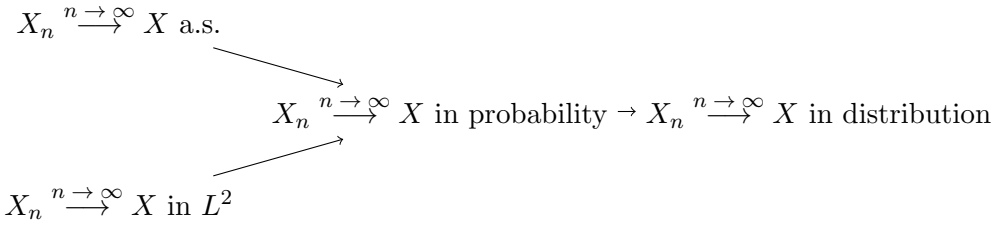
Definition 1.2.1 (Sample)**.** *Let* $X_1, \ldots, X_n$ *be a sample from a distribution* F*. The sample mean is defined as the random variable*

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Definition 1.2.2 (Modes of Convergence)**.** *Let* $\{X_n\}_{n\geq 1}$ *be a sequence of random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* X *a random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$*. We say that* $X_n \to X$ *as* $n \to \infty$*,*

Almost surely if $\mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] = 1$ *(a.s. for short);*
In L^2 *if* $\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0$;
In probability if, for each $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}[|X_n - X| < \epsilon] = 1$;
In distribution if $\mathbb{P}[X_n \leq x] = F_{X_n}(x) \overset{n\to\infty}{\longrightarrow} F_X(x) = \mathbb{P}[X \leq x]$.

The following implications hold in general:



Theorem 1.2.3. *(Strong Law of Large Numbers, SLLN for short) Let $\{X_n\}_{n\geq 1}$ be an i.i.d sequence of random variables with finite mean and variance, i.e,*

$$\mu = \mathbb{E}[X_i] < \infty, \quad \sigma^2 = \mathbb{E}[(X_i - \mu)^2] < \infty, \quad i \in \mathbb{N}.$$

Then

$$\lim_{n\to\infty} S_n = \mu \quad a.s.$$

Shorthand: for a pair of random variable variables X, Y, we write $X \sim Y$ to indicate that X and Y have the same distribution.

If X has a given distribution F, for each $n \geq 1$, one could ask: what is the distribution of the sample average S_n? In general, even for nice distributions F, the sample average has a complicated distribution which is not well–known (hence not even given a particular name, and thus not tabulated). But there are several important examples where the sample average follows a "known" distribution.

Example 1.2.4. *If $X \sim \mathcal{N}(\mu, \sigma^2)$, i.e. X is normally distributed with mean μ and standard deviation σ, then*

$$S_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Proof. The sum of normally distributed random variables is normal (see Lemma 1.2.5 below). Hence we only need to calculate mean and variance of the sample $X_1, \ldots, X_n$. Since the sample is i.d.d.,

$$\mathbb{E}[S_n] = \frac{n \times \mathbb{E}[X_1]}{n} = \mu.$$

Furthermore, independence implies that $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$. Therefore

$$\mathrm{Var}(S_n) = \frac{1}{n^2}(n\mathrm{Var}(X_1)) = \frac{\sigma^2}{n}.$$

□

Lemma 1.2.5. *If X_1 and X_2 are independent, normally distributed random variables with means μ_1, μ_2 and variances σ_1^2, σ_2^2, then*

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Proof. The characteristic functions of X_i are given by

$$\mathbb{E}[e^{iuX_i}] = e^{iu\mu_i - \frac{\sigma^2 u^2}{2}}.$$

For independent random variables, the expectations of products is the product of expectations. Thus

$$\mathbb{E}[e^{iuX}] = \mathbb{E}[e^{iu(X_1+X_2)}] = \mathbb{E}[e^{iuX_1}e^{iuX_2}] = \mathbb{E}[e^{iuX_1}]\mathbb{E}[e^{iuX_2}] = e^{iu(\mu_1+\mu_2)-\frac{u^2(\sigma_1^2+\sigma_2^2)}{2}},$$

and the last term is just the characteristic function of a $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ distributed random variable. Since characteristic functions of random variables uniquely determine their distribution, we are done. □

An extension of the law of large numbers is the central limit theorem (CLT for short), which states that for any distribution, not just the normal-distribution, sample averages converge in distribution to a normal distribution.

Theorem 1.2.6 (Central Limit Theorem)**.** *Let $\{X_n\}_{n\geq 1}$ be an i.i.d sequence of random variables with finite mean and variance, i.e.*

$$\mu = \mathbb{E}[X_i] < \infty, \quad \sigma^2 = \mathbb{E}[(X_i - \mu)^2] < \infty, \quad i \in \mathbb{N}.$$

Informally, for large sample sizes n, the sample average S_n is approximately normally distributed

$$S_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

More precisely,

$$\frac{\sqrt{n}}{\sigma}(S_n - \mu) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

in distribution.

Example 1.2.7. *If $X \sim Bin(k = 1, p = 1/2)$, i.e. a coin toss, then*

$$nS_n = X_1 + \ldots X_n \sim B(n, p).$$

The expectation and variance of a cointoss X are finite and given by

$$\mu = 1/2 * 0 + 1/2 * 1 = 1/2, \quad \sigma^2 = (0 - 1/2)^2 * 1/2 + (1 - 1/2)^2 * 1/2 = 1/4.$$

The strong of large numbers implies that

$$S_n \overset{n\to\infty}{\longrightarrow} \mu = 1/2, \quad a.s.$$

By the central limit theorem,

$$\frac{\sqrt{n}}{\sigma}(S_n - \mu) \approx \mathcal{N}(0,1).$$

Hence

$$S_n \approx \mathcal{N}\left(\mu = 1/2,\, \sigma^2/n = \frac{1}{4n}\right).$$

Definition 1.2.8. *Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution with finite mean μ and variance σ^2. Let further Z be a standard normally distributed random variable, and let z be the two-sided α quantile, that is*

$$\mathbb{P}[-z \leq Z \leq z] = \alpha.$$

The asymptotic confidence interval for the sample average S_n is defined as

$$I_n = \left[\mu - z\frac{\sigma}{\sqrt{n}}, \mu + z\frac{\sigma}{\sqrt{n}}\right].$$

By the Central Limit Theorem,

$$\mathbb{P}[S_n \in I_n] \approx \alpha$$

when n is sufficiently large.

Remark 1.2.9. *The Central Limit Theorem applies to most commonly used distributions but not the Cauchy distribution. Even worse, the conclusion of the CLT is wrong for Cauchy distributed random variables.*

Justification (sketch). Part 1: If X is Cauchy distributed it has a density given by

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Since

$$\mathbb{E}[X] = \int_{x\in\mathbb{R}} x f_X(x) dx = \infty$$

the finite mean assumption of the CLT is not satisfied.

Part 2: If $X_1, X_2, \ldots X_n$ is an i.i.d sample from the Cauchy distribution, then S_n is Cauchy distributed. Hence the distribution S_n is independent of n and S_n converges in distribution to a Cauchy distribution, not to a normal distribution. □

1.3 Computer Laboratory 1

For the moment, use MATLAB built–in routines to generate random numbers from the requisite distributions.

Exercise 1: Generating Random Numbers in MATLAB

(a.) Generate samples of N uniform random numbers in [0,1] for $N = 10, 100, 1000, 10000$ and plot the empirical density function for each N. Calculate the mean and variance for each sample.

(b.) Repeat part (*a.*) for the standard Normal distribution ($\mu = 0$ and $\sigma = 1$).

(c.) Plot the true density functions for the Uniform and Normal distributions on the same axes as the empirical density functions from parts (*a.*) and (*b.*). HINT: use the *trapz* routine.

Exercise 2: Central Limit Theorem and Asymptotic Confidence Intervals

(a.) Plot the empirical density function of the sample mean for samples of sizes $10, 50, 100, 250$ using exponential random variables with parameter $\lambda = 1$ (compute the sample mean 1,000 times for each sample size).

(b.) Compute the 95% and 99% confidence intervals from the empirical density function of the sample mean.

(c.) Compare your answers from part (*b.*) with the exact asymptotic confidence intervals predicted by the Central Limit Theorem.

Exercise 3: Frequency & Severity Pricing in Insurance

Suppose an insurance company knows that for a particular line of business the number of claims occurring each year has a Poisson distribution with mean 5. It happens that they also know that the size of each claim is independent of all other claims (and of the claim frequency) and has a Pareto distribution (Type–1) with scale parameter 2 and shape parameter $10,000$. Plot the empirical density function of the claims for this line of business and verify that the simulated mean claim amount agrees with the theoretical answer.

Chapter 2

Generating Random Numbers

2.1 Uniform Random Numbers

If we know how to create a sample from a uniform distribution, then we can (according to Lemma 1.1.13) obtain a sample with a given distribution F. Therefore it is often enough to have samples from the uniform distribution and hence this distribution plays a central role in the theory of random number generation.

2.1.1 On creating uniform samples

In some sense, there are no random number generators. Computers can only execute algorithms, which are deterministic instructions, and thus they can only yield samples which appear random. We call these numbers pseudorandom numbers, and the algorithms which produce these numbers are called pseudorandom number generators.

The theoretical wish

A generator of **genuine** random numbers is an algorithm that produces a sequence of random variables U_1, U_2, ... which satisfies

(i.) each U_i is uniformly distributed between 0 and 1.

(ii.) the U_i are mutually independent.

Property (ii.) is the more important one since the normalisation in $(i.)$ is convenient but not crucial. Property (ii.) implies that all pairs of values are uncorrelated and, more generally, that the value of U_i should not be predictable from $U_1, \ldots, U_{i-1}$. Of course, the properties listed above are those of authentically random numbers; the goal is to come as close as possible to these properties with our artificially generated pseudorandom numbers.

2.1.2 Linear Congruential Generators

An important and simple class of generators are the linear congruential generators, abbreviated as LCGs. We need the modulo operation in order to define this class of generators.

Definition 2.1.1. *For nonnegative integers x and m, we call $y = x \bmod m$ the integer remainder from the division x/m; we will write this as*

$$y = x(m),$$

or more usually

$$y = x \bmod m.$$

Definition 2.1.2. *A linear congruential generator (LCG) is an iteration of the form*

$$\begin{aligned} x_{i+1} &= (ax_i + c) \mod m \\ u_{i+1} &= \frac{x_{i+1}}{m} \in (0, 1), \end{aligned}$$

where a, c, and m are integers.

Example 2.1.3. *Choose $a = 6$, $c = 0$, and $m = 11$. Starting from $x_0 = 1$, which is called the* seed*, gives*

$$1, 6, 3, 7, 9, 10, 5, 8, 4, 2, 1, 6, \ldots$$

Choosing $a = 3$ yields the sequence

$$1, 3, 9, 5, 4, 1, \ldots$$

whereas the seed $x_0 = 2$ results in

$$2, 6, 7, 10, 8, 2, \ldots$$

Conditions for a full period I

Theorem 2.1.4. *Suppose $c \neq 0$. The generator has full period (that is the number of distinct values generated from any seed x_0 is $m - 1$) if and only if the following conditions hold:*

(i) *c and m are relatively prime (their only common divisor is 1),*

(ii) *every prime number that divides m divides $a - 1$ as well,*

(iii) *$a - 1$ is divisible by 4 if m is.*

Corollary 2.1.5. *If m is a power of 2, the generator has full period if c is odd and $a = 4n + 1$ for some integer n.*

Example 2.1.6. *The Borland C++ LCG has parameters*

$$m = 2^{32}, \quad a = 22695477 = 1 + 4 * 5673869, \quad c = 1.$$

Hence, by Corollary 2.1.5, the LCG has full period.

Conditions for a Full Period II

If $c = 0$ and m is a prime, then full period is achieved from any $x_0 \neq 0$ when

(i) $a^{m-1} - 1$ is a multiple of m,

(ii) $a^j - 1$ is not a multiple of m for $j = 1, \ldots, m - 2$.

If a satisfies these two properties it is called a *primitive root* of m. In this situation, the sequence $\{x_i\}_{i \geq 1}$ is of the form

$$x_0, ax_0, a^2x_0, a^3x_0, \ldots \mod m,$$

given that $c = 0$. The sequence returns to x_0 for the first time for the smallest k which satisfies $a^k x_0 \mod m = x_0$. This is the smallest k for which $a^k \mod m = 1$, that is $a^k - 1$ is a multiple of m.

Hence, the definition of a prime root coincides with the requirement that the series does not return to x_0 until $a^{m-1}x_0$.

Examples of LCG Parameters

Modulus m	Multiplier a	Reference
$2^{31} - 1$	16807	Lewis, Goodman, and Miller,
$(= 2147483647)$		Park and Miller
	39373	L'Ecuyer
	742938285	Fishman and Moore
	950706376	Fishman and Moore
	1226874159	Fishman and Moore
2147483399	40692	L'Ecuyer
2147483563	40014	L'Ecuyer

Example 2.1.7. *Define an LCG with parameters*

$$c = 0, \quad m = 2^3 - 1 = 7, \quad a = 3.$$

Does this LCG have full period?

Remark 2.1.8. *Linear congruential generators are no longer (and have not been for some time) used practically. The Mersenne Twister family of algorithms is among the most popular in practice and modern implementations are appropriate for most applications.*

2.2 Testing uniformity

Previously, we have only checked for a full period, to see whether a pseudorandom number generator is reasonable or not. In this section, we demonstrate more elaborate means to test the quality of a pseudorandom number generator.

Given a sample from a supposedly uniform distribution, one can use statistical tests to reject the hypothesis of uniformity. The samples provided by a computer are fake since they are totally deterministic, and therefore not random (as such they cannot be uniformly distributed). However, they are so "well chosen", that they might appear random. Hence we require statistical tests of randomness in order to judge the quality of a candidate PRNG.

2.2.1 Chi-Squared Test

Definition 2.2.1. *Let $k \geq 1$, and let $X_1, \ldots, X_k$ be a sequence of i.i.d. standard normally distributed random variable. The distribution of the sum of squares*

$$S = X_1^2 + \cdots + X_k^2$$

is called chi-square with k degrees of freedom.

The probability density function of $\chi(k)$ is given by

$$f(x,k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2},$$

where Γ denotes the gamma function, which is defined by

$$\Gamma(\xi) = \int_0^\infty x^{t-1} e^{-x} dx.$$

For integers $n \in \mathbb{N}$, we can write the Gamma function in terms of the factorial function as follows:

$$\Gamma(n) = (n-1)! = (n-1)(n-2) \cdots \cdot 2 \cdot 1.$$

Definition 2.2.2. *Let*

$$X_1, X_2, \ldots, X_n$$

be a sample.

The chi-squared test for uniformity is constituted by the following:

- *the null hypothesis H_0: The sample is from a uniform distribution against the alternative hypothesis H_a that it is not.*
- *the test statistic*

$$T = \frac{k}{n} \sum_{j=1}^{k} \left(n_j - \frac{n}{k} \right)^2,$$

where k is the number of equidistant partitions (so-called " bins", to be chosen) of the unit interval, given by

$$[0, 1/k), \quad [1/k, 2/k), \ldots, [(k-1)/k, 1].$$

and n_j is the number of observations in the jth bin.

- *The confidence level α (to be chosen).*

The following is given without proof.

Lemma 2.2.3. *As $n \to \infty$, T converges, in distribution, to the chi-square distribution χ_{k-1} with $k-1$ degrees of freedom.*

2.2.2 The Kolmogorov–Smirnov Test

Another simple test uses the empirical distribution function of the sample. The Kolmogorov Smirnov test bases on the following intuition: If the sample is uniformly distributed, the deviation of the empirical distribution function from the theoretical distribution function, as given in Lemma 1.1.9 should be small.

Definition 2.2.4. *If $x = (x_1, \ldots, x_n)$ is a sample, then the empirical distribution function of x is given by*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x)}(x_i), \quad x \in \mathbb{R}.$$

Definition 2.2.5. *The Kolmogorov Smirnov test for uniformity is constituted by the following:*

- *the null hypothesis H_0: The sample is from a given distribution F against the alternative hypothesis H_a that it is not.*
- *the test statistic*

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|,$$

where n is the sample size.

- *The confidence level α (to be chosen).*

As $n \to \infty$, $\sqrt{n}D_n$ converges to

$$\sup_{t \in \mathbb{R}} |B(F(t))|, \tag{2.2.1}$$

in distribution, where B is a Brownian bridge (i.e the quantity $\sup_{t\in\mathbb{R}} |B(F(t))|$) is a random variable.

For $F(t) = t$, the uniform distribution, the critical values of the Kolmogorov statistics D_n are known. For large n, the statistics converge in distribution to the so-called Kolmogorov distribution, which satisfies

$$K = \sup_{t \in [0,1]} |B(t)|.$$

In fact, it can be shown that

$$\mathbb{P}[K \leq x] = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}, \quad x \in \mathbb{R}^+.$$

2.3 Practical Considerations

In the preceding sections we outlined some simple statistical tests to illustrate how one may put a candidate PRNG through its paces. In practice, more stringent tests are used, namely the DIEHARD test suite or the more modern TestU01 test suite.

In applications, the following considerations are typically the most important when considering the appropriateness of a random number generating scheme:

- Reproducibility,
- Speed,
- Portability,
- Period Length (if any),
- Randomness.

2.4 Computer Laboratory 2

Exercise 1: Linear Congruential Generators

Write a function **LCG** which takes as input five parameters N, a, c, m and x_0 and outputs a vector of N uniform pseudorandom numbers on $[0, 1]$ based on a linear congruential generator with parameters a, c and m, and a seed of x_0. Perform some simple sense checks on the output, e.g. calculate the mean and variance of your samples and plot the empirical density function, for *good* choices of the parameters.

Exercise 2: Chi-Square Test for Uniformity

(a.) Write a function **ChiTest** which takes as input two parameters N and *sample*, where N is the number of bins and *sample* is a vector of uniform pseudorandom numbers on $[0, 1]$. The output of **ChiTest** should be the Chi-Square statistic for *sample* with N bins under the null hypothesis that *sample* is in fact a vector of i.i.d. uniform random numbers on $[0, 1]$.

(b.) Using your **ChiTest** function test the following random number generators for uniformity at the 90%, 95% and 99% confidence levels (choose any *sensible* seed):

(i.) the **LCG** routine with $a = 16807$, $c = 1$, $m = 2^{32}$,

(ii.) the **LCG** routine with $a = 48271$, $c = 0$, $m = 2147483644$,

(iii.) the MATLAB implementation 'Mersenne' of the Mersenne Twister algorithm.

Exercise 3: The Empirical Distribution Function

(a.) Compute the empirical distribution function for samples of uniform random numbers ($N = 10, 100, 500, 1000$) and compare them with the true distribution function. HINT: Use the MATLAB routine *stairs*.

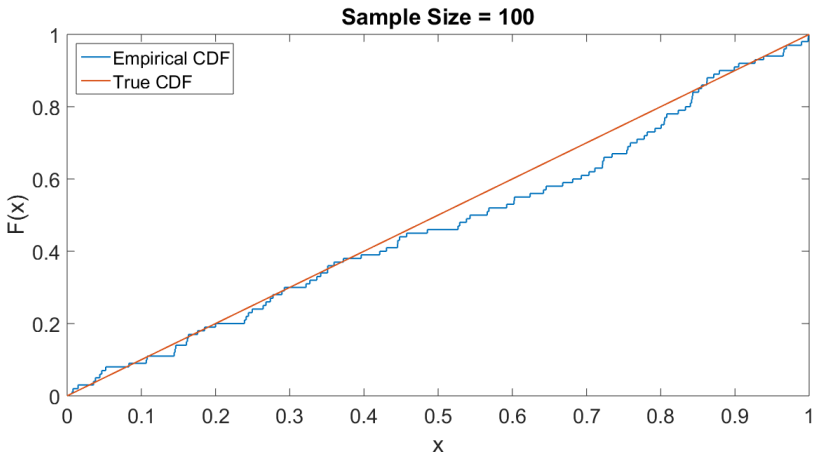


Figure 2.1: Plot comparing the empirical distribution function and the true distribution function for the uniform distribution.

(b.) Using your calculation from parts (a.), calculate $\sup_{x\in\mathbb{R}} |F_N(x) - F(x)|$ **numerically** for the uniform distribution (for each value of N), where $F(x)$ is the true distribution function.

(c.) Comment on your results to parts (a.) and (b.). Are the results as expected? Your comments should make clear reference to relevant theoretical results discussed during lectures.

The following table should be useful in parts (b.) and (c.).

Table 2.1: Quantile table for the Kolmogorov-Smirnov Distribution.

N	90%	95%	99%
100	1.207	1.340	1.608
1,000	1.219	1.353	1.622
10,000	1.222	1.356	1.626

2.5 The Inverse Transform Method

The following is an immediate consequence of Lemma 1.1.13.

Lemma 2.5.1. *Let $U_1, U_2, \ldots, U_n$ be a sequence of i.i.d. random variables with uniform distribution on $[0, 1]$. Then*

$$X_1 = F^{-1}(U_1), \ldots, X_n = F^{-1}(U_n)$$

is a sequence of i.i.d random variables with distribution $F_{X_1} = F$.

2.6 The Acceptance–Rejection Method

Suppose f and g are densities on $\mathbb{R}$. Moreover, assume that the following relationship holds for some positive constant c:

$$0 \leq f(x) \leq c \cdot g(x), \qquad \text{for each } x \in \mathbb{R}. \tag{2.6.1}$$

If we know how to generate samples with density g, the **Acceptance–Rejection method** is a way to generate samples with density f by taking a sample X with density g and accepting it as a sample Y with density f with probability $f(X)/cg(X)$ – we can achieve this by sampling from the uniform distribution since if $U \sim U([0,1])$, then

$$\mathbb{P}\left[U \leq \frac{f(x)}{c\,g(x)}\right] = \frac{f(x)}{c\,g(x)}, \quad x \in \mathbb{R}.$$

Equation (2.6.1) guarantees that $f(x)/cg(x) \in [0,1]$ for each $x \in \mathbb{R}$. Therefore we accept X as a random sample with density f if $U \leq f(X)/cg(X)$ – we will prove that this acceptance rule does indeed generate the desired distribution.

The algorithm to generate a single random variable with density f is of the form:

```
generate X from the distribution with density g;
generate U from U([0,1]);    (independent of X)
while U > f(X)/(cg(X))
       generate X from the distribution with density g;
       generate U from U([0,1]);
end;
return Y = X;
```

Proposition 2.6.1 (Law of Total Probability)**.** *For any random variables X and Y and any subset $A \subset \mathbb{R}$, one has that*

$$P\left[Y \in A\right] = \int\limits_{x \in \mathbb{R}} P\left[Y \in A \mid X = x\right] g(x)\, dx\,, \tag{2.6.2}$$

where X is assumed to be a continuous random variable with density g.

Remark 2.6.2. *The above result is the continuous version of the discrete Law of Total Probability. Suppose X and Y are discrete random variables and $A \subset \mathbb{R}$. Since X is discrete it can only take on countably many values, say $\{x_1, x_2, \ldots, x_n, \ldots\}$. In this case, the Law of Total Probability reads*

$$\mathbb{P}\left[Y \in A\right] = \sum_{j=1}^{\infty} \mathbb{P}\left[Y \in A \mid X = x_j\right] \mathbb{P}\left[X = x_j\right].$$

Now compare the formula above with equation (2.6.2).

Theorem 2.6.3. *The sample created by the Acceptance–Rejection Method (ARM) has c.d.f* $F(y) = \int_{-\infty}^{y} f(u)du$ *for each* $y \in \mathbb{R}$. *Furthermore, the most efficient sampling scheme is achieved with the smallest* $c \in [1, \infty)$.

Proof. Let X have density g and $U \sim U([0,1])$. $Y = X \mid U \leq f(X)/cg(X)$ is the random variable generated by the ARM.

In order to show that the most efficient scheme is achieved with the smallest value of c we must calculate the rate at which random numbers are accepted. By the Law of Total Probability, the probability of acceptance is

$$\begin{aligned} P\left[U \leq \frac{f(X)}{c \cdot g(X)}\right] &= \int\limits_{x \in \mathbb{R}} \mathbb{P}\left[U \leq \frac{f(x)}{c \cdot g(x)} \,\middle|\, X = x\right] g(x)\, dx \\ &= \int\limits_{x \in \mathbb{R}} \frac{f(x)}{c \cdot g(x)} \cdot g(x)\, dx = \frac{1}{c}\,, \end{aligned} \tag{2.6.3}$$

Thus the smaller the value of c, the more random variables are accepted and hence the more efficiently the ARM performs; this proves the second assertion in the statement of the theorem.

It remains to show that the sample from the ARM has c.d.f given by F, i.e. if Y is the ARM sample, then $\mathbb{P}[Y \leq y] = \int_{-\infty}^{y} f(u)\, du$ for each $y \in \mathbb{R}$. Proceed by direct calculation as follows:

$$\begin{aligned} P[Y \leq y] = P\left[X \leq y \,\middle|\, U \leq \frac{f(X)}{c\, g(X)}\right] &= \frac{\mathbb{P}\left[X \leq y\,,\, U \leq \frac{f(X)}{c\, g(X)}\right]}{P\left[U \leq \frac{f(X)}{c\, g(X)}\right]} \quad \text{(by definition)} \\ &= c\,\mathbb{P}\left[X \leq y\,,\, U \leq \frac{f(X)}{c\, g(X)}\right] \qquad \text{(using equation (2.6.3))} \\ &= c \int\limits_{x \in \mathbb{R}} \mathbb{P}\left[x \leq y\,,\, U \leq \frac{f(x)}{c\, g(x)} \,\middle|\, X = x\right] g(x)\, dx \quad \text{(Prop. 2.6.1)} \\ &= \int_{-\infty}^{y} \mathbb{P}\left[U \leq \frac{f(x)}{c\, g(x)} \,\middle|\, X = x\right] g(x)\, dx \\ &= c \int_{-\infty}^{y} \frac{f(x)}{c\, g(x)} g(x)\, dx = \int_{-\infty}^{y} f(x)\, dx, \end{aligned}$$

as required. □

The most important point in the proof above is that $1/c$ is the acceptance rate and hence we want to choose c as small as possible for efficiency. It follows that we should always try to take $c = \sup_{x \in \mathbb{R}} f(x)/g(x)$. Another useful interpretation is that, on average, we need to generate c random numbers with density g to obtain one random number with density f.

Example 2.6.4. *Suppose that we wish to generate samples from a probability distribution with density*

$$f(x) = \begin{cases} \frac{1}{2}x^2 e^{-x}, & x \geq 0, \\ 0, & \textit{otherwise}, \end{cases}$$

using the ARM. Choose the majorising density to be an exponential distribution with parameter λ. For maximal efficiency, we then determine the λ that minimizes the average number of trial samples needed to generate a sample with density f.
By diktat,

$$g(x) = \lambda e^{-\lambda x}, \quad x \geq 0,$$

and hence we introduce the so–called likelihood ratio

$$m(x) = \frac{f(x)}{g(x)} = \frac{1}{2\lambda} x^2 e^{-x(1-\lambda)}.$$

We minimize m with respect to x as follows:

$$2\lambda m'(x) = e^{-x(1-\lambda)} \left(2x - (1-\lambda)x^2\right) = 0.$$

For $x > 0$, the right–hand side is zero if and only if

$$2 - (1-\lambda)x = 0.$$

Clearly, $\lambda < 1$ (otherwise m goes to infinity for $x \to \infty$) and thus we conclude that

$$x_* = \frac{2}{1-\lambda}$$

is the minimiser. Furthermore,

$$c = m(x_*) = \frac{1}{2\lambda} \left(\frac{2}{1-\lambda}\right)^2 e^{-2}$$

The value of $\lambda \in (0,1)$ which minimizes c is that which minimizes

$$r(\lambda) := \frac{1}{\lambda} \frac{1}{(1-\lambda)^2}.$$

Differentiation yields

$$r' = \frac{-1}{\lambda^2} \frac{1}{(1-\lambda)^2} + \frac{2}{\lambda} \frac{1}{(1-\lambda)^3} = 0$$

and hence

$$\lambda = \frac{1}{3}.$$

Therefore the optimal density for the ARM is given by

$$g(x) = \frac{1}{3} e^{-x/3}, \quad x \geq 0.$$

2.7 Computer Laboratory 3

Exercise 1: Inverse Transform Method for the Cauchy Distribution

(a.) Compute the inverse distribution function for the Cauchy distribution.

(b.) Generate random samples from the Cauchy distribution via the inverse transform method and empirically demonstrate that the Central Limit Theorem does not hold for Cauchy random variables.
HINT: Plot the empirical density function of the sample mean distribution (using Cauchy random variables) and overlay it with the best fitting normal density – the fit should be terrible!

Exercise 2: Inverse Transform Method for the Pareto Distribution

(a.) Compute the inverse distribution function for the Pareto distribution.
HINT: the p.d.f of a Pareto random variable is given by

$$f(x) = \begin{cases} \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, & \text{if } x \geq \beta > 0 \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$.

(b.) Using the inverse transform method generate samples from the Pareto distribution for different parameter values, compare the empirical and analytic density functions and compare the empirical and analytic moments.

Exercise 3: The Acceptance–Rejection Method

(a.) Write a MATLAB function which uses the ARM method to sample from the distribution studied in Example 2.6.4. Your function should take as input a natural number N and output an $N \times 1$ vector of random variables with density f.

(b.) Demonstrate that your function from part $(a.)$ faithfully reproduces random variables with density f by plotting empirical p.d.f's from samples of various sizes and overlaying them with the true p.d.f.

Chapter 3

The Monte Carlo Method

3.1 The Main Idea

Suppose we want to calculate an expectation of a random variable $Y = A(X)$, for a measurable function $x \mapsto A(x)$, and X has density $f_X(x)$. One can attempt to calculate this expectation directly using Lemma 1.1.16, i.e.

$$\mu = \mathbb{E}[A(X)] = \int_{\mathbb{R}} A(x) f_X(x) dx.$$

However, often the function A is sufficiently complicated that the integral cannot be derived in closed form. For example, A might be the payoff function of an exotic option or be related to the claims policy on an insurance contract. The Monte-Carlo method (MCM) uses simulations to derive an approximation of μ as follows:

Suppose we have a sample from the same distribution as X, namely

$$X_1, \ldots, X_n.$$

Then

$$A(X_1), \ldots, A(X_n)$$

is a sample with the distribution of $A(X)$ (see Lemma 1.1.15). The strong law of large numbers says that the sample average converges almost surely to the expectation μ, i.e.

$$\lim_{n\to\infty} \frac{A(X_1) + \cdots + A(X_n)}{n} = \mathbb{E}[A(X)] = \mu, \quad a.s.$$

In practice, one runs a simulation and obtains a sample, say

$$x_1, \ldots, x_n,$$

and thus a sample from the same distribution as $A(X)$,

$$A(x_1), \ldots, A(x_n).$$

For large n, the SLLN therefore allows to use the sample average as proxy for μ, that is,

$$\hat{\mu}_n := \frac{A(x_1) + \cdots + A(x_n)}{n} \approx \mu.$$

However, each time we run a new simulation the sample average delivers a new value. This begs the question:

3.2 How good is the approximation?

The Monte Carlo estimator has several desirable properties as a statistical estimator. In particular, the Monte Carlo estimator is:

- Unbiased, i.e. $\mathbb{E}[\hat{\mu}_n] = \mu$, or "the sample mean is equal to the true mean".
- Strongly consistent, i.e. $\lim_{n\to\infty} \hat{\mu}_n = \mu$ a.s.

Furthermore, we can quantify the error of the Monte Carlo estimator. If $\sigma^2 = \mathbb{E}[(A(X) - \mu)^2]$, the variance of $A(X)$, was known, then we could use the CLT to quantify the quality of approximation of the MC estimator. In fact,

$$\frac{\sqrt{n}}{\sigma}(\hat{\mu}_n - \mu) \approx \mathcal{N}(0, 1).$$

Therefore, the α confidence interval of $\hat{\mu}_n$ is approximately equal to

$$\mu \pm z_{\frac{1-\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \tag{3.2.1}$$

where z_β is the β-quantile of $\mathcal{N}(0, 1)$ distribution (see Definition 1.2.8). However, there are some other possible sources of error inherent in the MC approach, for example:

- "Payoff" discretisation error, i.e. the function A must be approximated numerically.
- "Model" discretisation error, i.e. when we generate sample from $A(X)$ we typically do so imperfectly, such as when discretising the solution to an SDE.

In well–constructed numerical methods, these types of errors vanish as we discretise more and more finely but this increased precision often comes at a considerable computational cost.

3.3 The practical approach

Recall that for a sample $y_1, \ldots, y_n$, the sample mean and variance are given by

$$\bar{y} = \frac{y_1 + \ldots y_n}{n}, \quad \bar{s} = \frac{\sum_{k=1}^{n}(y_k - \bar{y})^2}{n-1}.$$

Often, one does not know the precise values of μ or σ^2 (especially μ, otherwise we don't need to run Monte-Carlo simulations at all!). One thus replaces μ and σ^2 in (3.2.1) by their empirical estimates as follows:

$$\hat{\mu}_n \pm z_{\frac{1-\alpha}{2}} \frac{\hat{\sigma}_n}{\sqrt{n}}, \tag{3.3.1}$$

where $\hat{\mu}_n, \hat{\sigma}_n^2$ are the empirical mean and variance of the given sample

$$A(x_1), \ldots, A(x_n).$$

The quantity $\hat{\sigma}_n/\sqrt{n}$ is called standard error. Summarizing, we have

$$S.E. = \sqrt{\frac{\sum_{i=1}^{n}(A(x_i) - \hat{\mu}_n)^2}{n(n-1)}}. \tag{3.3.2}$$

The standard error expression in (3.3.2) tells us that the error scales like $n^{-1/2}$. For example, if we wish to make our estimate 10 times more accurate we need to increase n by a factor of 100! **This is a key feature of the Monte Carlo method and persists even in a higher dimensional setting.** In one–dimensional the simple trapezoidal rule has an error which scales like n^{-2} and is hence far superior to the MCM. However, in d–dimensions the trapezoidal rule error scales like $n^{-2/d}$ while the MCM retains its $n^{-1/2}$ error scaling. Therefore MCMs are attractive methods for quickly and accurately evaluating higher dimensional integrals whose closed form expressions are not available.

Remark 3.3.1. *We will see how to represent option prices as expectations (i.e. integrals) which can be evaluated via MCMs.*

3.4 Examples

Example 3.4.1. *Let X be a uniformly distributed r.v. on the unit interval. Let us study the Monte Carlo estimation of*

$$\mu = \mathbb{E}[X^2].$$

For size n, let us sample from the uniform distribution, which gives us

$$x_1, x_2, \ldots, x_n$$

independent copies of X. The Monte Carlo estimator for μ is just the sample average

$$\mu \approx \frac{x_1^2 + \cdots + x_n^2}{n} =: \hat{\mu}_n.$$

The theoretical value of μ can actually be calculated explicitly, and it is given by

$$\mu = \int_0^1 x^2 f_X(x) dx = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}.$$

where we recall that the uniform distribution has density $f_X \equiv 1$.

The variance σ^2 of the Monte Carlo estimator is the Variance of X^2 divided by n, σ^2/n, where

$$\sigma^2 = \mathbb{E}\left[\left(X^2 - \mu\right)^2\right] = \mathbb{E}\left[X^4\right] - \mathbb{E}\left([X^2]\right)^2 = 1/5 - (1/3)^2 = \frac{9-5}{45} = \frac{4}{45}.$$

We conclude that the α confidence interval of $\hat{\mu}_n$ is approximately

$$\mu \pm z_{\frac{1-\alpha}{2}} \frac{2}{\sqrt{9n}}. \tag{3.4.1}$$

Example 3.4.2. *Suppose we want to calculate the integral*

$$\int_0^1 e^{x^3} dx,$$

which has no closed–form expression. We can write this integral as the expectation of a random variable as follows:

$$\mu = \mathbb{E}[e^{X^3}],$$

where X is uniformly distributed on the unit interval. Then, we run a simulation which gives us a large sample of X,

$$x_1, \ldots, x_n.$$

Thus

$$\mu = \int_0^1 e^{x^3} dx \approx \frac{e^{x_1^3} + \cdots + e^{x_n^3}}{n} = \hat{\mu}_n.$$

3.5 Computer Laboratory 4

Exercise 1:

(a.) Interpret the integral

$$I_1 = \int_0^\infty x e^{-x} dx$$

as an expectation of a r.v. Y. HINT: Use the exponential distribution.

(b.) Calculate μ, σ^2 of Y (i.e. the mean and variance).

(c.) Give an approximate 95% confidence interval for the MC estimate of I_1 using the CLT.

Exercise 2:

(a.) Interpret the integral

$$I_2 = \int_0^5 e^{-x} dx$$

as an expectation of a r.v. Y.
HINT: Use the uniform distribution on $[0, 5]$.

(b.) Calculate μ, σ^2 of Y.

(c.) Give an approximate 99% confidence interval for the MC estimate of I_2 using the CLT.

Exercise 3:

(a.) Using MATLAB simulations, empirically check the quality of the confidence intervals for the MC estimates of I_1 and I_2 obtained in Exercises 1 and 2 (try $N = 10, 25, 100, 1000$).

Chapter 4

Stochastic Differential Equations

4.1 Theoretical Background

4.1.1 Stochastic Processes

Definition 4.1.1 (Stochastic Process)**.** *A stochastic process is a collection of random variables* $\{X_t\}_{t\in T}$ *on a common probability space* $(\Omega, \Sigma, \mathbb{P})$ *where we interpret the set* T *as time.*

The two classes of stochastic processes which feature in this course are continuous processes, i.e. $T = [0, \infty)$, and discrete processes, i.e. $T = \mathbb{N}$, since theses classes are typically the most useful in applications.

If we consider a process $\{X_t\}_{t\in T}$ and fix $\omega \in \Omega$, then we can think of the function

$$t \mapsto X_t(\omega), \quad t \in T,$$

as a single realisation or "experiment" of the process X. For example, if X models stock prices then when we look at the stock price chart we are observing a single realisation of the stock price process; we often refer to this as a path of the process.

Example 4.1.2 (Simple Random Walk)**.** *Suppose* $\{X_n\}_{n\geq 1}$ *is a sequence of i.d.d random variables on* $(\Omega, \Sigma, \mathbb{P})$ *with mean zero and unit variance. Note that we could always translate and scale the* X_n*'s to have zero mean and unit variance. Construct another stochastic process from this sequence by taking*

$$Y_n = \sum_{j=1}^{n} X_j, \quad n \geq 1, \quad Y_0 = 0.$$

Clearly, $\mathbb{E}[Y_n] = 0$ *for all* $n \geq 0$ *and* $Var[Y_n] = n$ *for all* $n \geq 1$*. By construction,* Y_n *depends on* $Y_{n-1}, \ldots, Y_1$ *but the increment* $Y_n - Y_{n-1}$ *is* independent *of* $Y_{n-1}, \ldots, Y_1$*. Hence*

$$\mathbb{P}[Y_n \in B \,|\, Y_{n-1} = y_{n-1}, \ldots, Y_1 = y_1] = \mathbb{P}[Y_n \in B \,|\, Y_{n-1} = y_{n-1}], \quad B \in \mathcal{B}(\mathbb{R}) \tag{4.1.1}$$

and Y *is a Markov ("memoryless") process. In general,* Y_n *will not have a simple "known" distribution. However, regardless of the exact distribution of the* X_n*'s, the CLT tells us that the* Y_n*'s will converge in distribution to normal random variables with mean zero and variance* n*. Thus the "normal random walk", in which each* $X_n \sim \mathcal{N}(0,1)$*, is essentially the canonical random walk in discrete time — we presently discuss the continuous time analogue of this process.*

We need a way to rigorously capture the information that is accumulated as we observe the evolution of a stochastic process over time; the following definitions provide the correct notions. Note that we omit the constant reminder that we are dealing with processes on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 4.1.3 (Generated σ–algebra)**.** *If* $X_n \mapsto \Omega$ *is a random variable for each* $n \in \mathbb{N}$*, then*

$$\Sigma := \sigma\left(X_n : n \in \mathbb{N}\right)$$

is the smallest σ*–algebra such that each* X_n *is* Σ*–measurable.*

Definition 4.1.4 (Filtration)**.** *A filtration* $\{\mathcal{F}_n : n \in \mathbb{N}\}$ *on* (Ω, Σ) *is an increasing family of* σ*–algebras, i.e.*

$$\mathcal{F}_m \subset \mathcal{F}_n, \quad m < n, \quad m, n \in \mathbb{N},$$

such that $\mathcal{F}_n \subset \Sigma$ *for all* $n \in \mathbb{N}$*.*

Definition 4.1.5 (Adaptedness)**.** *Let* $\{X_n\}_{n \geq 0}$ *be a stochastic process on* $(\Omega, \mathcal{F})$*.* X *is adapted to a filtration* $\{\mathcal{F}_n : n \in \mathbb{N}\}$ *if*

$$X_n \text{ is } \mathcal{F}_n\text{–measurable for each } n \in \mathbb{N}.$$

Definition 4.1.6 (Natural Filtration)**.** *Let* $\{X_n\}_{n \geq 0}$ *be a stochastic process on* $(\Omega, \mathcal{F})$*. The natural filtration for the process* X *is the family of* σ*–algebras given by*

$$\mathcal{F}_n = \sigma\left(X_m : m \in \{0, 1, \ldots, n\}\right), \quad n \in \mathbb{N}. \tag{4.1.2}$$

By definition, a process is adapted to its natural filtration and this is the minimal filtration to which the process could be adapted.

If we observe the process X from time 0 to time n, then we know the values of $X_0, \ldots, X_n$ and (intuitively) we should be able to decide whether an event of

the form $\{\omega : X_n(\omega) \in B\}$ occurred or not (for $B \in \mathcal{B}(\mathbb{R})$); this is exactly the information contained in the σ–algebra $\mathcal{F}_n$ from (4.1.2)! Furthermore, we should not be able to decide whether events of the form $\{X_{n+1} \in B\}$ have occurred at time n if our intuitions regarding causality and clairvoyance are to be respected; this is prevented by asking that the process be adapted – adapted processes are often called non–anticipative.

Remark 4.1.7. *In light of the newly introduced formality above, we could restate* (4.1.1) *as*

$$\mathbb{P}\left[Y_n \in B \,|\, \mathcal{F}_n\right] = \mathbb{P}[Y_n \in B \,|\, Y_{n-1}], \quad B \in \mathcal{B}(\mathbb{R}),$$

where $\mathcal{F}_n = \sigma\left(Y_m : m \in \{0, 1, \ldots, n\}\right)$ *is* Y*'s natural filtration.*

All of the definitions above extend naturally to the continuous–time setting (modulo technical considerations) and we need only retain the intuition of these concepts going forward.

4.1.2 Brownian Motion

Definition 4.1.8. *A standard Brownian motion* $B = \{B_t, \mathcal{F}_t : t \geq 0\}$ *is a continuous adapted process defined on some probability space* $(\Omega, \Sigma, \mathbb{P})$ *such that*

(i.) $B_0 = 0$ *a.s.*

(ii.) $B_t - B_s$ *is independent of* $\mathcal{F}_s$ *for all* $0 \leq s < t$,

(iii.) $B_t - B_s \sim \mathcal{N}(0, t - s)$ *for all* $0 \leq s < t$.

Remark 4.1.9. *In the definition above, think of* $\mathcal{F}_t$ *as simply being the natural filtration of the process.*

Brownian motion is the most important continuous time stochastic process and is ubiquitous in models from finance, engineering, physics, and a variety of other areas. Mathematically, our first question is: does such a process actually exist? There are a number of ways to construct standard Brownian motion and the interested reader can consult [5] for a thorough exposition. From the point of view of applications, the key properties of Brownian motion are:

- Brownian paths are a.s. continuous, i.e. the function

  $$t \mapsto B_t(\omega), \quad t \in [0, \infty),$$

  is continuous for all $\omega \in A \subset \Sigma$ with $\mathbb{P}[A] = 1$.

- **Brownian paths are a.s. nowhere differentiable (nonsmooth paths)**.

- $\mathbb{E}[B_t] = 0$ and $\mathrm{Var}[B_t] = t$ for each $t \geq 0$.
- $\mathrm{Cov}(B_t, B_s) = \min(t, s)$ for all $t, s \in [0, \infty)$.
- Brownian motion is a martingale, i.e.

$$\mathbb{E}[B_t \,|\, \mathcal{F}_s] = B_s, \quad 0 \leq s < t.$$

Remark 4.1.10. *The non–differentiability property of Brownian paths may seem strange, especially given the continuity of the paths, but this is actually crucial for financial modelling — C^1 paths lead to arbitrage and nonsensical models.*

We close this section with the definition of an important class of processes of which Brownian motion is the canonical example.

Definition 4.1.11 (Square–integrable Martingales)**.** *Let $X = \{X_t, \mathcal{F}_t : t \geq 0\}$ be a continuous process. X is a martingale if*

$$\mathbb{E}\left[X_t \,|\, \mathcal{F}_s\right] = X_s, \quad 0 \leq s < t,$$

and X is square integrable if $\mathbb{E}\left[X_t^2\right] < \infty$ for each $t \geq 0$. The class of continuous square integrable martingales is denoted by $\mathcal{M}_2^c$.

4.1.3 Itô Integrals and Itô's Lemma

The goal of this section is to give a sensible meaning to "differential" equations of the form

$$\frac{d}{dt}x(t) = a(t, x(t)) + b(t, x(t)) \times \text{"random noise"}. \tag{4.1.3}$$

where the "random noise" is provided by a Brownian motion (in an appropriate sense). It turns out that to rigorously formulate (4.1.3) we must write it as an integral equation as follows:

$$x(t) = x(0) + \int_0^t a(s, x(s))\, ds + \int_0^t b(s, x(s)) dB_s, \quad t \geq 0, \tag{4.1.4}$$

where in the second integral we "integrate with respect to the Brownian motion". Equation (4.1.4) is called a *stochastic differential equation*, or SDE for short. By defining the second integral term in (4.1.4) rigorously we can give a precise meaning to (4.1.3) – formulating this "dB_s" integral rigorously is the basis of what is called Itô calculus.

Remark 4.1.12. *We can think of "dB_t" as an increment of Brownian motion over an infinitesimally small interval, just as we might think of the "dt" in a Riemann integral as being an infinitesimally small increment in the variable t.*

The two main steps in the construction of the Itô integral are as follows (see [5]):

(1.) **Construct the integral for simple processes:**

Definition 4.1.13. *A process X on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is called simple if there exists a strictly increasing sequence of real numbers $\{t_n\}_{n\geq 0}$ with $t_0 = 0$ and $\lim_{n\to\infty} t_n = \infty$, and a sequence of (surely) uniformly bounded random variables $\{\xi_n\}_{n\geq 0}$ where each ξ_n is $\mathcal{F}_{t_n}$–measurable such that*

$$X_t(\omega) = \xi_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{j=0}^{\infty} \xi_j(\omega)\mathbb{1}_{(t_j, t_{j+1}]}(t), \quad t \geq 0, \quad \omega \in \Omega.$$

If X is a simple process, define the Itô integral as

$$\int_0^t X_s \, dB_s := \sum_{j=0}^{\infty} \xi_j \left(B_{t\wedge t_{j+1}} - B_{t\wedge t_j} \right), \quad t \geq 0.$$

N.B. The definition above justifies thinking of dB_t as an infinitesimal increment of Brownian motion.

Lemma 4.1.14. *Let X and Y be simple processes. The Itô integral of a simple process enjoys the following properties:*

- $\int_0^t X_s \, dB_s$ *is $\mathcal{F}_t$–adapted,*
- $\int_0^0 X_s \, dB_s = 0$ *a.s.,*
- $\mathbb{E}\left[\int_0^t X_s \, dB_s \,|\, \mathcal{F}_u \right] = \int_0^u X_s \, dB_s$ *for $0 \leq u < t$, i.e. the Itô integral is a martingale,*
- $\mathbb{E}\left[\left(\int_0^t X_s \, dB_s \right)^2 \right] = \int_0^t X_s^2 \, ds$ *(the so–called Itô Isometry),*
- $\int_0^t (\alpha X_s + \beta Y_s) \, dB_s = \alpha \int_0^t X_s \, dB_s + \beta \int_0^t Y_s \, dB_s$*, i.e. the Itô integral defines a linear operator.*

(2.) **Approximate non–simple processes by simple processes:**

Lemma 4.1.15. *If X is a bounded, measurable, $\mathcal{F}_t$–adapted process, then there exists a sequence $\{X^n\}_{n\geq 0}$ of simple processes that approximate X arbitrarily well. More precisely,*

$$\sup_{T>0} \lim_{n\to\infty} \mathbb{E}\left[\int_0^T |X_s - X_s^n|^2 \, ds \right] = 0. \tag{4.1.5}$$

The lemma above is representative and more general classes of processes can be handled similarly. Finally, we define the Itô integral for a general class of processes as the limit of the discrete approximations discussed above.

Definition 4.1.16. *Let X be a progressively measurable process such that $\sup_{T>0} \mathbb{E}[|X_T|] < \infty$. Define $\left\{\int_0^t X_s dB_s, \mathcal{F}_t : t \geq 0\right\}$ as the unique square–integrable martingale which satisfies*

$$\lim_{n\to\infty} \left\| \int_0^{\cdot} X_s \, dB_s - \int_0^{\cdot} X_s^n \, dB_s \right\| = 0,$$

for some appropriate norm $\|\cdot\|$ and for every sequence $\{X^n\}_{n\geq 0}$ of simple processes such that (4.1.5) holds.

N.B. It turns out that Itô integral given by Definition 4.1.16 retains all of the important properties it had for simple processes (see Lemma 4.1.14).

4.1.4 Solutions to Stochastic Differential Equations

We call a process X of the form

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dB_t, \quad t \geq 0, \tag{4.1.6}$$

an Itô process and when we do so we tacitly assume that a and b are such that X is well–defined. The function $a : (t, X_t) \mapsto a(t, X_t)$ is called the **drift coefficient** and the function $b : (t, X_t) \mapsto b(t, X_t)$ is referred to as the **diffusion coefficient**. Equation (4.1.6) is written in informal "differential" notation, but is understood to refer to the integral equation

$$X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t b(s, X_s)\, dB_s, \quad t \geq 0, \tag{4.1.7}$$

where X_0 is deterministic and given as part of the problem data.

When we write down expressions such as (4.1.4) or (4.1.6), we must ask: does there exist a process which satisfies our equation (and is adapted, continuous, etc.)? In fact, (4.1.4) and (4.1.6) are initial value problems whose solutions only exist under certain restrictions on a and b. Furthermore, such solutions are only unique (in a certain sense) under even stricter conditions. First, we must be clear by what we mean when we say that a process is a solution to an SDE.

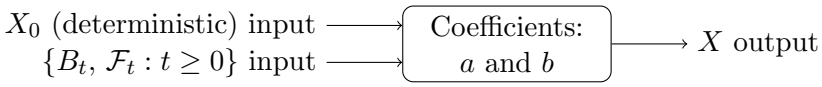
Definition 4.1.17 (Solution to an SDE)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space endowed with a Brownian motion $\{B_t, \mathcal{F}_t : t \geq 0\}$ where the filtration $\mathcal{F}_t$ has been suitably augmented. A process X is a solution to (4.1.6) if*

(i.) X *is adapted to* $\mathcal{F} = \{\mathcal{F}_t : t \geq 0\}$,

(ii.) $\mathbb{P}\left[\int_0^t |a(s, X_s)| + b^2(s, X_s)\, ds < \infty\right] = 1$ *for each* $t \geq 0$,

(iii.) X *obeys* (4.1.7) *almost surely.*

Figure 4.1: We can think of the initial value problem (4.1.6) in terms of input ("problem data") and output (the solution process) as in the diagram below.



Definition 4.1.18 (Uniqueness)**.** *If for any two solutions* X *and* $\tilde{X}$ *to* (4.1.6),

$$\mathbb{P}\left[X_t = \tilde{X}_t,\, t \geq 0\right] = 1,$$

then the solution to (4.1.6) *is unique.*

The following result gives sufficient conditions under which (4.1.6) has a unique strong solution.

Theorem 4.1.19 (Existence and Uniqueness Criteria)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space endowed with a Brownian motion* $\{B_t, \mathcal{F}_t : t \geq 0\}$ *where the filtration* $\mathcal{F}_t$ *has been suitably augmented. If*

- a *and* b *are measurable functions from* $[0, \infty) \times \mathbb{R}$ *to* $\mathbb{R}$,
- a *and* b *are globally Lipschitz continuous, i.e. there exists a* $K > 0$ *such that*

  $$|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leq K|x - y|,$$

  for each $t \geq 0$ *and all* $x, y \in \mathbb{R}$,
- a *and* b *grow no faster than linearly, i.e. there exists a* $K > 0$ *such that*

  $$|a(t, x)|^2 + |b(t, x)|^2 \leq K(1 + |x|)^2,$$

  for each $t \geq 0$ *and all* $x \in \mathbb{R}$,

then (4.1.6) *has a unique solution.*

Exercise 4.1.20. *The most important case for applications is when* a *and* b *are linear and do not depend on time, i.e.* $a(t, x) = \alpha x$ *for some* $\alpha \in \mathbb{R}$ *and* $b(t, x) = \beta x$ *for some* $\beta \in \mathbb{R}$. *Convince yourself that all the hypotheses of Theorem 4.1.19 hold in this special case.*

Given an Itô process satisfying a particular SDE we often need to derive the dynamics of a related process. For ordinary differential equations we use the chain rule to derive associated ODEs; for SDEs we need a "stochastic" chain rule to take into account that the processes we are interested in do not have derivatives in the usual sense.

Theorem 4.1.21 (Itô's Rule, a.k.a Itô's Lemma)**.** *If $f : (t,x) \mapsto f(t,x)$ is in $C^{1,2}([0,\infty)\times\mathbb{R};\mathbb{R})$ and X is an Itô process obeying (4.1.6), then*

$$\begin{aligned} f(t,X_t) &= f(0,X_0) + \int_0^t \frac{\partial}{\partial t} f(s,X_s)\,ds + \int_0^t \frac{\partial}{\partial x} f(s,X_s)\,dX_s \qquad \text{(Ito)}\\ &\quad + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2} f(s,X_s) b^2(s,X_s)\,ds \\ &= f(0,X_0) + \int_0^t \frac{\partial}{\partial t} f(s,X_s)\,ds + \int_0^t \frac{\partial}{\partial x} f(s,X_s) a(s,X_s)\,ds \\ &\quad + \int_0^t \frac{\partial}{\partial x} f(s,X_s) b(s,X_s)\,dB_s + \frac{1}{2}\int_0^t \frac{\partial^2}{\partial x^2} f(s,X_s) b^2(s,X_s)\,ds, \quad \text{a.s.} \end{aligned}$$

for each $t \geq 0$.

In differential form (Ito) reads

$$df(t,X_t) = \frac{\partial}{\partial t} f(t,X_t)dt + \frac{\partial}{\partial x} f(t,X_t)\,dX_t + \frac{1}{2}\frac{\partial^2}{\partial x^2} f(t,X_t) b^2(t,X_t)\,dt.$$

The first two terms on the right–hand side match the usual chain rule and the final term is essentially a correction for the "irregularity" of the random processes under consideration. When using Itô's rule in differential form the following formal rules are often useful:

- $dt \times dt = 0$,
- $dt \times dB_t = 0$,
- $dB_t \times dB_t = 0$.

Corollary 4.1.22. *Suppose $f : (x,y) \mapsto f(x,y)$ is in $C^{2,2}(\mathbb{R}\times\mathbb{R};\mathbb{R})$, X is an Itô process obeying (4.1.6), and Y is an Itô process obeying*

$$dY_t = \alpha(t,Y_t)\,dt + \beta(t,Y_t)\,dW_t, \quad t \geq 0$$

where W is another Brownian motion with $\mathrm{Corr}(B,W) = \rho$. Then

$$\begin{aligned} df(X_t,Y_t) &= \frac{\partial}{\partial x} f(X_t,Y_t)\,dX_t + \frac{\partial}{\partial y} f(X_t,Y_t)\,dY_t + \rho \times \frac{\partial^2}{\partial x \partial y} b(t,X_t)\beta(t,Y_t)\,dt \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2} f(X_t,Y_t) b^2(X_t,Y_t)\,dt + \frac{1}{2}\frac{\partial^2}{\partial y^2} f(X_t,Y_t)\beta^2(X_t,Y_t)\,dt \quad \text{a.s.} \end{aligned}$$

for each $t \geq 0$.

4.2 Computer Laboratory 5

Exercise 1: Simulating Brownian Motion

It can be proven that a standard Brownian motion is actually the limit of a scaled random walk as the step size tends to zero. Hence, by choosing our step size $h > 0$ sufficiently small, we can approximate Brownian motion on an interval $[0, T]$ by a discrete time process $\{B_t : t \in h\mathbb{N}\}$ as follows:

$$B_{ih} = B_{(i-1)h} + \sqrt{h}\, Z_i, \quad i = 1, ..., N; \quad W_0 = 0,$$

where $Z_i \sim N(0, 1)$ for $i = 1, ..., N$ and $Nh = T$.

By simulating sample paths of Brownian motion, denoted by B_t, on some interval $[0, T]$ demonstrate the following properties empirically:

(a.) $\mathbb{E}[B_t] = 0$ and $\text{Var}[B_t] = t$ for all $t \geq 0$;

(b.) $\text{Cov}(B_t, B_s) = \min(t, s)$ for all $t, s \geq 0$;

(c.) $M_t := \max_{0 \leq s \leq t} B_s$ obeys $\mathbb{E}[M_t] = \sqrt{2t/\pi}$ for all $t \geq 0$.

Exercise 2: Itô's Lemma

Let $\{B_t : t \geq 0\}$ be a standard Brownian motion throughout.

(a.) Use Itô's Lemma to show that $Y_t = \frac{1}{2}B_t^2$ obeys an SDE of the form $dY_t = \mu(t)\, dt + \sigma(t)\, dB_t$ and hence calculate the integral $\int_0^t B_s dB_s$.

(b.) Following the approach you used in part (a.), calculate $\int_0^t s\, dB_s$.
HINT: Choose $Y_t = tB_t$.

(c.) Consider the system of SDEs given by

$$\begin{aligned} dS_t &= rS_t\, dt + \sqrt{V_t} S_t\, dB_t^S, && t > 0, \\ dV_t &= \kappa(\theta - V_t)dt + \xi\sqrt{V_t}\, dB_t^V, && t > 0, \end{aligned}$$

where B_t^S and B_t^V are independent Brownian motions. Use Itô's Lemma to show that the process $Y_t = \log(S_t)$ obeys the SDE

$$dY_t = \left(r - \frac{V_t}{t}\right) dt + \sqrt{V_t}\, dB_t^S, \quad t > 0.$$

4.3 Discretising Stochastic Differential Equations

4.3.1 Overview and Scheme Performance

We now discuss how to numerically approximate solutions to stochastic differential equations of the form

$$dX_t = a(t, X_t)\,dt + b(t, X_t)\,dB_t, \quad t \in [0, T], \tag{4.3.1}$$

for some $T > 0$. We focus on schemes which approximate the true solution to (4.3.1) at uniformly spaced points on $[0, T]$. To wit, we choose a parameter $h \in (0, 1]$ (typically very small) and approximate the true solution at the "grid points" $h, 2h, \ldots Nh$ where $N = 1 + \lfloor T/h \rfloor$. If we denote the approximate solution at the grid point nh by x_n, then we can define a piecewise continuous approximation X^h to X by linear interpolation as follows:

$$X_t^h = x_n + \frac{t - nh}{h}\left(x_{n+1} - x_n\right), \quad t \in [nh, (n+1)h), \quad n \in \{0, 1, \ldots, N+1\}.$$

Note that we always choose $x_0 = X_0$. Naturally, we need a way to measure whether or not our approximations are any good; the following definitions outline the standard metrics for this purpose.

Definition 4.3.1 (Strong Convergence)**.** *A numerical scheme is said to be strongly convergent if*

$$\lim_{h \to 0^+} \mathbb{E}\left[\left|X_T - X_T^h\right|\right] = 0.$$

Definition 4.3.2 (Weak Convergence)**.** *A numerical scheme is said to be weakly convergent if*

$$\lim_{h \to 0^+} \mathbb{E}\left[g\left(X_T\right) - g\left(X_T^h\right)\right] = 0,$$

for every polynomial function g.

Remark 4.3.3. *Every continuous function can be approximated arbitrarily well by a polynomial on a compact interval (Stone–Weierstrass Theorem); this motivates the use of polynomials in the definition of weak convergence as a "representative" subclass of continuous functions.*

In general, strong convergence criteria (e.g. Definition 4.3.1) are concerned with the true and approximate solutions being close in a pathwise sense, i.e. for each outcome ω (which we think of as a single realisation of the solution process). By contrast, weak convergence criteria (e.g. Definition 4.3.2) ask that the true and approximate solutions are close in distribution. Hence

- strongly convergent schemes are more appropriate when the pathwise behaviour of the solution is relevant for our desired application, e.g. pricing a **path–dependent** derivative (American option, Asian option, Barrier option, etc.),
- weakly convergent schemes are fit for purpose if our intended application only requires that we capture the distribution or behaviour of moments of the true solution, e.g. pricing a European option (the payoff only depends on the price of the underlying at maturity).

Suppose our numerical scheme converges to the true solution to (4.3.1) (in some appropriate sense). From both a computation and practical perspective, the next natural question is: how quickly does the scheme converge?

Definition 4.3.4 (Convergence Order)**.** *Let $h > 0$ denote the step size of the discretisation scheme. A numerical scheme is strongly convergent with order γ if there exists a positive constant K such that*

$$\mathbb{E}\left[|X_T - X_T^h|\right] \leq K\, h^\gamma.$$

A numerical scheme is weakly convergent with order γ if there exists a positive constant K such that

$$\mathbb{E}\left[g\left(X_T\right) - g\left(X_T^h\right)\right] \leq K\, h^\gamma,$$

for every polynomial function g.

Remark 4.3.5. *The constants K in Definition 4.3.4 may depend on T and, in the case of weak convergence, on the polynomial function g.*

If a scheme has convergence order γ, then making the discretisation parameter, h, k times smaller will decrease the approximation error by a factor of k^γ. Thus, if a scheme has order of convergence of 1, then making the step size 100 times smaller means that the error will be 100 times smaller. However, if our scheme is only convergent with order $1/2$, then making the step size 100 times smaller will only decrease the error by a factor of 10. Therefore **the higher the convergence order the better!**

4.3.2 The Euler–Maruyama Scheme

The Euler–Maruyama (EM for short) scheme is the simplest method for approximating the solutions of SDEs. Despite its simplicity, it is widely used and perfectly adequate in many applications. For SDEs of the form (4.3.1) the EM scheme has the form

$$x_{n+1} = x_n + h \cdot a(nh, x_n) + b(nh, x_n)\,\Delta B_h, \quad n \geq 0, \tag{EM}$$

where $\Delta B_h = B_{(n+1)h} - B_{nh}$ is an increment of Brownian motion. We can *informally* motivate (EM) as follows: integrate (4.3.1) from nh to $(n+1)h$ to obtain

$$X_{(n+1)h} - X_{nh} = \int_{nh}^{(n+1)h} a\left(u, X_u\right)\, du + \int_{nh}^{(n+1)h} b\left(u, X_u\right)\, dB_u,$$

where X is the **true solution** to (4.3.1) so that $X_{nh} =: x_n$. Now approximate both integrals on the right–hand side by taking the value of the integrands at the lower limit of integration; this yields

$$X_{(n+1)h} \approx X_{nh} + h \cdot a(nh, X_{nh}) + b(nh, X_{nh}) \left(B_{(n+1)h} - B_{nh}\right),$$

and we thus recover (EM).

As the following result confirms, the EM scheme is both strongly and weakly convergent under some technical conditions on the coefficients a and b.

Theorem 4.3.6 (Sufficient conditions for Euler–Maruyama scheme convergence)**.**

(A1.) ***Measurability:*** *a and b are jointly measurable,*

(A2.) ***Lipschitz:*** *a and b are Lipschitz continuous, i.e.*

$$|a(t,x) - a(t,y)| \leq K_1|x-y|, \quad |b(t,x) - b(t,y)| \leq K_2|x-y|,$$

for some $K_1, K_2 > 0$ for all $t \in [0,T]$ and all $x, y \in \mathbb{R}$,

(A3.) ***Linear Growth:*** *There exists a constant $K > 0$ such that*

$$|a(t,x)|^2 \leq K^2\left(1 + |x|^2\right), \quad |b(t,x)|^2 \leq K^2\left(1 + |x|^2\right),$$

for all $t \in [0,T]$ and all $x \in \mathbb{R}$.

Under conditions (A1.–A3.), the Euler–Maruyama scheme (EM) *is strongly convergent to the solution of* (4.3.1)*.*

Conditions (A1.–A3.) in the theorem above highlight the fact that the EM scheme (and indeed most other "standard" schemes) are not appropriate for highly nonlinear problems. In this course, we will mainly be interested in problems which are linear (or almost linear) and hence obey the hypotheses above. However, it is worth noting that highly nonlinear problems often require special schemes (e.g. adaptive time stepping, implicit schemes, etc.).

From the results above, we see that the EM scheme converges under somewhat restrictive assumptions on the coefficients a and b. Furthermore, the following results give criteria for the EM scheme to achieve

- strong convergence of order $1/2$,

- weak convergence of order 1.

Theorem 4.3.7. *Suppose conditions (A1.–A3.) from Theorem 4.3.6 hold, and that there exists a constant $K > 0$ such that for all $x \in \mathbb{R}$ and all $t, s \in [0, T]$,*

$$|a(t,x) - a(s,x)| + |b(t,x) - b(s,x)| \leq K(1 + |x|)\sqrt{|t-s|}.$$

Then the Euler–Maruyama scheme (EM) converges strongly with order 1/2 to the solution of (4.3.1).

Theorem 4.3.8. *Consider the autonomous SDE given by*

$$dX_t = a\left(X_t\right) dt + b\left(X_t\right) dB_t, \quad t > 0. \tag{4.3.2}$$

If both a and b are 4 times continuously differentiable, grow no faster than polynomially and have uniformly bounded derivatives, then the EM scheme converges weakly with order 1 to the solution of (4.3.2).

Remark 4.3.9. *A function $g : \mathbb{R} \mapsto \mathbb{R}$ is polynomially bounded if there exist positive constants K and q such that*

$$|g(x)| \leq K\left(1 + |x|^q\right), \quad x \in \mathbb{R}.$$

One might argue that the EM scheme can still be useful for applications in which both high accuracy and pathwise precision are required if we just make the step size sufficiently small. However, making the step size smaller and smaller can result in round off error (or underflow) due to the finite precision arithmetic inherent in working on a computer. Hence we seek a discretisation scheme which has strong convergence of order 1.

4.3.3 The Milstein Scheme

The Milstein discretisation scheme for (4.3.2) is given by

$$x_{n+1} = x_n + h \cdot a(x_n) + b(x_n)\Delta B_h + \frac{1}{2} b'(x_n) b(x_n)\left((\Delta B_h)^2 - h\right), \quad n \geq 0. \tag{Mil}$$

Comparing (Mil) to (EM), the final term on the right–hand side of (Mil) is new – it is a higher quality approximation of the stochastic integral term in (4.3.2). The EM scheme approximates the integrals in (4.3.2) as follows:

$$\int_t^{t+h} a(X_s)\, ds \approx h \cdot a(X_t), \quad h > 0. \tag{4.3.3}$$

$$\int_t^{t+h} b(X_s) dB_s \approx b(X_t)\left(B_{t+h} - B_t\right) = b(X_t)\Delta B_h, \quad h > 0, \tag{4.3.4}$$

where $\Delta B_h = B_{t+h} - B_t$.

In terms of a Taylor series expansion, the approximation of the drift term in (4.3.3) is $O(h)$ as $h \to 0^+$. However, in probability, the approximation in (4.3.4) is only $O\left(\sqrt{h}\right)$ accurate since the ΔB_h term is $O\left(\sqrt{h}\right)$. Therefore we should improve the approximation in (4.3.4) to $O(h)$ in order to obtain a scheme which is strongly convergent with order 1. It is actually remarkable that the EM scheme achieves weak order 1 with an estimate as rough as (4.3.4)!

Below we give a heuristic derivation of the Milstein scheme; it is useful and instructive to keep the following formal relations in mind:

- $h \times h \approx O\left(h^2\right)$ as $h \to 0^+$,
- $h \times \Delta B_h \approx O\left(h^{3/2}\right)$ as $h \to 0^+$,
- $\Delta B_h \times \Delta B_h \approx O\left(h\right)$ as $h \to 0^+$.

Since the Euler scheme approximates the drift coefficient to $O(h)$ already, we focus on the diffusion term. By Itô's Lemma,

$$\begin{aligned} d\left(b(X_t)\right) &= b'(X_t)\, dX_t + \frac{1}{2} b''(X_t) b^2(X_t)\, dt \\ &= a(X_t) b'(X_t)\, dt + b(X_t) b'(X_t)\, dB_t + \frac{1}{2} b''(X_t) b^2(X_t)\, dt, \end{aligned}$$

or in integral form

$$\begin{aligned} b(X_u) = b(X_t) &+ \int_t^u a(X_s) b'(X_s)\, ds + \int_t^u b(X_s) b'(X_s)\, dB_s \\ &+ \frac{1}{2} \int_t^u b''(X_s) b^2(X_s)\, ds, \quad u \in [t, t+h]. \end{aligned}$$

Now use the standard Euler approximation to approximate $b(X_u)$ for $u \in [t, t+h]$ as follows:

$$\begin{aligned} b(X_u) \approx b(X_t) &+ a(X_t) b'(X_t)(u-t) + b(X_t) b'(X_t)\left(B_u - B_t\right) \\ &+ \frac{1}{2}(u-t) b''(X_t) b^2(X_t), \quad u \in [t, t+h] \end{aligned}$$

Neglect the $O(u-t)$ terms to obtain the local approximation

$$b(X_u) \approx b(X_t) + b'(X_t) b(X_t)\left(B_u - B_t\right), \quad u \in [t, t+h].$$

Hence

$$\begin{aligned}\int_t^{t+h} b(X_u)\,dB_u &\approx \int_t^{t+h} \left\{b(X_t) + b'(X_t)b(X_t)\,(B_u - B_t)\right\} dB_u \\ &= b(X_t)\,(B_{t+h} - B_t) \\ &\quad + b(X_t)b'(X_t)\left\{\int_t^{t+h} B_u\,dB_u - B_t\,(B_{t+h} - B_t)\right\}. \qquad (4.3.5)\end{aligned}$$

The final term on the right–hand side above requires further simplification for practical purposes. As we showed in an earlier exercise,

$$\int_0^t B_u dB_u = \frac{1}{2}B_t^2 - \frac{t}{2}.$$

Thus

$$\int_t^{t+h} B_u\,dB_u = \frac{1}{2}B_{t+h}^2 - \frac{1}{2}B_t^2 - \frac{h}{2},$$

and plugging the identity above into (4.3.5) yields

$$\begin{aligned}\int_t^{t+h} b(X_u)\,dB_u &\approx b(X_t)\,(B_{t+h} - B_t) + \frac{1}{2}b(X_t)b'(X_t)\left\{(B_{t+h} - B_t)^2 - h\right\} \\ &= b(X_t)\Delta B_h + \frac{1}{2}b(X_t)b'(X_t)\left\{(\Delta B_h)^2 - h\right\}.\end{aligned}$$

Finally, applying the approximation above on a uniformly spaced mesh gives the Milstein iterative scheme from (Mil).

Theorem 4.3.10 (Milstein Convergence Criteria)**.** *Suppose a and b obey condition* $(A1.)$ *from Theorem 4.3.6, and that a, a', b and b' all obey conditions* $(A2.)$ *and* $(A3.)$ *from Theorem 4.3.6. Then the Milstein scheme given by* (Mil) *converges strongly with order* 1 *to the solution of* (4.3.2)*.*

4.4 Computer Laboratory 6

Exercise 1:

Consider the stochastic differential equation

$$dX_t = \mu X_t\,dt + \sigma X_t\,dB_t, \quad t \in [0, T], \qquad (4.4.1)$$

where $T > 0$ and $\mu, \sigma \in \mathbb{R}$.

(a.) Does (4.4.1) have a unique solution? What kind of performance would you expect from the EM scheme if you applied it to (4.4.1)?
HINT: Check the relevant results and conditions from these notes.

(b.) Use Itô's Lemma to compute $\log(X_t)$ and hence show that the solution to (4.4.1) is given by

$$X_t = X_0 \exp\left((\mu - \tfrac{\sigma^2}{2})t + \sigma B_t\right), \quad t \in [0, T].$$

(c.) Numerically demonstrate the strong convergence of the EM scheme for (4.4.1).
HINT: Let $X_0 = \mu = \sigma = T = 1$ if you wish (since these parameters should not affect the convergence). Use the formula from part (b.) and the Monte Carlo method to estimate the quantity $\mathbb{E}\left[|X_T - X_T^h|\right]$ for successively smaller values of $h > 0$.

(d.) Repeat part (c.) using the Milstein scheme. Compare the performance of the EM and Milstein schemes – are your results as expected?

Chapter 5

Applications in Finance

5.1 Options

5.1.1 What are Options?

An asset is any financial object whose value is known at present but may change over time. For example, shares in a company, commodities, real–estate, and currencies are all assets.

Definition 5.1.1 (European Call)**.** *A European call option gives the holder the right (but not the obligation) to purchase from the writer a prescribed asset for a prescribed price at a prescribed time in the future.*

The prescribed price in the definition above is called the **strike price**, the prescribed asset is called the **underlying** and the prescribed time is called the **expiry or maturity date**.

Example 5.1.2. *Suppose I hold a European call option for 100 shares of Apple stock with a strike price of €20,000. If the actual market price of 100 Apple shares were more than €20,000 on the expiry date, then I should exercise my option since I could instantly sell my shares at a profit. However, if the market price of the shares is below the strike price, then I should choose not to exercise my option.*

From the example above, we observe that the holder of an option cannot lose money (regardless of how the price of the underlying asset moves) and the writer of the option can lose an unlimited amount. Hence the writer will require compensation in order to sell a European call option – our job is to figure out how much this compensation should be, i.e. what is the fair price of an option?

The "opposite" of a European call option is a European put option.

Definition 5.1.3 (European Put Option)**.** *A European put option gives the holder the right (but not the obligation) to sell to the writer a prescribed asset for a prescribed price at a prescribed time in the future.*

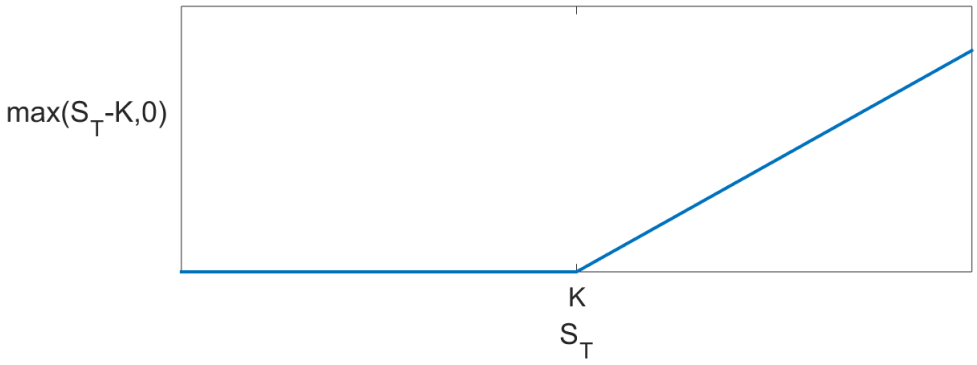
Like a European call, the holder of a European put will never exercise unless conditions are favourable. The question of deriving the fair value of an option is equally pertinent in this case.

For future reference, the payoff of a European call option on an underlying asset S and with strike price K and expiry time T is given by

$$\max(S_T - K,\, 0) = \begin{cases} S_T - K, & S_T > K, \\ 0, & S_T \leq K. \end{cases}$$

where S_T is the value of the underlying asset at time T.

Figure 5.1: The payoff function of a European call option.



Similarly, the payoff function of a European put option is given by

$$\max(K - S_T,\, 0).$$

Exercise 5.1.4. *Why is the current price of the underlying asset an upper bound for the price of a European Call option?*

5.1.2 Why do Options exist?

There are two primary reasons why options are traded in such huge volume:

(1.) Options provide an inexpensive way to hedge risk or to speculate on market movements;

(2.) There is a systematic way to determine the fair value of options and hence they can be traded with some confidence (relative to other asset classes).

We will discuss point (2.) in some detail in the following lectures and computer labs so it is not worthy of further elaboration here. Point (1.) is illustrated by the following simple example.

Example 5.1.5. *Nick fervently believes that Apple shares which are currently worth €200 each will be worth €250 each in 3 months time. One way of "putting his money where his mouth is" would be to buy Apple shares. If Nick's budget for speculation were €1,000 he could buy 5 shares in Apple and would expect to be €250 richer. However, options are vastly cheaper than the underlying shares and a 3 month call option on Apple shares with strike price of €250 might only set Nick back, say, €20. Nick could afford 50 such call options and if his predictions are correct his profit will be €1500 (assuming he borrows the strike price and pays it back instantaneously).*

Exercise 5.1.6. *Compare the outcomes of Nick's two trading strategies in the case that he is wrong and Apples share price is less than €250 in three months time.*

Table 5.1: The table below shows the share price of Royal Bank of Scotland (1634.00) and the associated option prices from September 19th, 2003. Note how inexpensive the options are relative to purchasing the underlying shares directly [3].

		Calls			Puts		
Option	Strike	Oct.	Nov.	Dec.	Oct.	Nov.	Dec.
RBS	1600	67.00	92.50	109.50	29.00	49.00	62.5
(1634.00)	1700	19.50	43.50	59.00	82.00	100.00	112.50

Many companies operate on a multinational basis and use options to hedge currency risk cheaply. For example, an American company may need to meet a **fixed** financial obligation in Euro's in a few months time (e.g. an American technology firm paying an EU regulatory fine) and could take out a call option to hedge the risk of a disadvantageous currency move in the meantime.

By combining different types of options investors can take very complicated positions which would be difficult or impossible to achieve by only trading in the underlying assets. Of course, this flexibility can be used both for solving complex risk management problems and for sophisticated speculation.

5.1.3 Pricing Principles

This section outlines the core ideas behind risk neutral derivative pricing using a mixture of nonrigorous (but plausible) mathematics and intuition. A fully rigorous

treatment of this topic is beyond the scope of this course (and in fact beyond the needs of most practitioners) but can be found in, for example, [1].

Definition 5.1.7. *We say that a portfolio of assets replicates an option if the portfolio has the same cash flows as the option.*

There are three core "principles" which are essential in explaining how option prices can be arrived at using the Monte Carlo Method.

(1.) **Replication:** If an option can be perfectly replicated through trading in the other assets available in the market, then the fair price of the option is the cost of the replicating portfolio. This is essentially another way of saying that a market does not contain *arbitrage opportunities*.

(2.) **Martingale Measure:** Discounted asset prices are martingales (see Definition 4.1.11) under a probability measure associated with the choice of numeraire (reference asset, e.g. cash in a bank account). Option prices are thus expectations of discounted payoffs under such a martingale measure.

(3.) **Market Completeness:** A complete market is one in which any payoff can be replicated through a trading strategy involving the available assets. The martingale measure associated with a particular numeraire is unique in a complete market and option prices are uniquely determined by the prices of the other assets in the market.

Principle (1.) describes how most financial institutions who write options operate. The option writer will typically eliminate the risk of the option they have issued by taking appropriate positions in other assets and hence the cost of this offsetting position largely determines the price they charge the option buyer. Of course, the issuer must also eat and duly charges a small additional premium for their services. In a market with no trading costs (or other frictions), the cost of an option and its replicating portfolio must be the same to eliminate arbitrage opportunities. If these prices do not agree an arbitrageur will simply buy the cheaper instrument and sell the more expensive one, generating a guaranteed risk–free profit. The existence of arbitrage is incompatible with a consistent pricing model and with the basic notions of economic equilibrium – it must be ruled out.

Principle (2.) is technical in nature but is in fact the key link between the Monte Carlo Method we discussed in Chapter 3 and the problem of determining the fair value of an option. This principle asserts that option prices can be represented as expectations (i.e. integrals) with respect to an appropriate probability measure. Crucially, this appropriate probability measure is not the "real–world" probability measure with respect to which we observe prices but rather a special risk–adjusted probability measure – more on this to come!

Principle (3.) essentially tells us under what situations we can expect principles (1.) and (2.) to hold true and hence when we can expect to be able to determine fair prices for options. In a complete market there is (in a sense) no need for options since their payoffs can be exactly replicated by trading in the other available assets. However, there are many practical reasons to want to "package" trading strategies as options even if they are redundant in a complete market (mathematically speaking).

5.1.4 Pricing via Replication

In order to develop the principles above, consider a market with d assets. Suppose that the dynamics of the i–th asset are given by

$$\frac{dS_t^i}{S_t^i} = \mu_i\left(S_t^i, t\right)\, dt + \sigma_i\left(S_t^i, t\right) dB_t^i, \quad t \geq 0, \tag{5.1.1}$$

where $\{B_t, \mathcal{F}_t : t \geq 0\}$ is a vector of (possibly correlated) Brownian motions.

Definition 5.1.8 (Portfolio)**.** *A vector $\theta = \left(\theta^1, \ldots, \theta^d\right)^T \in \mathbb{R}^d$ such that θ^i is the number of units of asset i held is called a portfolio.*

It follows that the value of a portfolio at time t in this market is given by

$$\theta^1 S_t^1 + \theta^2 S_t^2 + \ldots + \theta^d S_t^d = \theta^T S_t.$$

Definition 5.1.9. *A trading strategy is a stochastic process $\{\theta_t, \mathcal{F}_t : t \geq 0\}$ which is predictable ("only depends on past market information") such that for each fixed $t \geq 0$, θ_t is a portfolio.*

Over the interval $[t, t+h]$ the gain (or loss) in portfolio value due to market movements is given by $\theta_t^T\left[S_{t+h} - S_t\right] \approx \theta_t^T dS_t$ if we fix θ_t. Thus "summing" these gains suggests that the change in portfolio value over $[0, t]$ due to trading is given by

$$\int_0^t \theta_u^T dS_u.$$

In our analysis we only allow trading strategies which satisfy the self–financing condition:

$$\theta_t^T S_t - \theta_0^T S_0 = \int_0^t \theta_u^T dS_u, \quad t \geq 0. \tag{5.1.2}$$

Condition (5.1.2) simply says that the change in portfolio value over $[0, t]$ is equal to the change in value due to trading, i.e. no funds are withdrawn from or added to the portfolio. The self–financing condition can be rewritten as

$$\theta_t^T S_t = \theta_0^T S_0 + \int_0^t \theta_u^T dS_u, \quad t \geq 0, \tag{5.1.3}$$

and this reformulation has an important interpretation: starting from an initial investment $\theta_0^T S_0$ we can achieve a portfolio value of $\theta_t^T S_t$ by following the strategy θ over $[0,t]$.

Now consider an option with payoff $P(S_T)$ at expiry time T. Note that we are allowing the payoff to depend on the entire market of assets $S = \{S^1, \dots, S^d\}$. More typically, this option might be a European call option on a single share or an option on a handful of underlying assets at most. Suppose further that the value of this option at time $t \in [0,T]$ is given by some function $V : (S_t, t) \in \mathbb{R}^d \times \mathbb{R}^+ \mapsto V(S_t, t)$, i.e. the option price at time t only depends on the share prices at time t – this is an assumption, albeit a plausible one in light of (5.1.1). If V only depended on S^i and was sufficiently smooth, then we could consider $V : (x,t) \in \mathbb{R} \times \mathbb{R}^+ \mapsto V(x,t)$ and Itô's Lemma would yield

$$\begin{aligned} V(S_t^i, t) = V(S_0^i, 0) + \int_0^t \frac{\partial}{\partial u} V(S_u^i, u)\, du + \int_0^t \frac{\partial}{\partial x} V(S_u^i, u)\, dS_u^i \\ + \frac{1}{2} \int_0^t \frac{\partial^2}{\partial x^2} V(S_u^i, u) \left(S_u^i\right)^2 \sigma_i^2(S_u^i, u)\, du. \end{aligned}$$

In this case our replicating portfolio would takes on the simple form

$$\theta_t = \left(0, \dots, \theta_t^i, 0, \dots, 0\right).$$

In other words, the option only depends on S^i so the replicating portfolio should only depend on S^i as well. If V depends on the whole market and is sufficiently smooth we can apply Itô's Lemma (see Corollary 4.1.22) to show that the dynamics of V are given by

$$\begin{aligned} V(S_t, t) = V(S_0, 0) + \sum_{i=1}^d \int_0^t \frac{\partial V(S_u, u)}{\partial S^i}\, dS_u^i \\ + \int_0^t \left[\frac{\partial V(S_u, u)}{\partial u} + \frac{\rho_{i,j}}{2} \sum_{i,j=1}^d S_u^i S_u^j \sigma_i(S_u^i, u) \sigma_j(S_u^j, u) \frac{\partial^2 V(S_u, u)}{\partial S^i \partial S^j} \right] du, \end{aligned} \tag{5.1.4}$$

for each $t \geq 0$ and where $\rho_{i,j}$ denotes the correlation coefficient between B^i and B^j. Principle (3.) asserts that all payoffs can be replicated so there must be a self–financing trading strategy θ such that

$$V(S_t, t) = V(S_0, 0) + \sum_{i=1}^d \int_0^t \theta_u^i dS_u^i, \quad t \geq 0 \tag{5.1.5}$$

(compare equation (5.1.3)). Principle (1.) asserts that an option and its replicating portfolio must have the same value. Thus we can equate (5.1.4) and (5.1.5), and try to force equality by matching terms. We then conclude that

$$\theta_t^i = \frac{\partial V(S_t,t)}{\partial S^i}, \quad t \geq 0, \quad i \in \{1,\ldots,d\}, \tag{5.1.6}$$

and

$$\frac{\partial V(S_t,t)}{\partial t} + \frac{\rho_{i,j}}{2}\sum_{i,j=1}^{d} S_u^i S_u^j \sigma_i(S_u^i,u)\sigma_j(S_u^j,u)\frac{\partial^2 V(S_t,t)}{\partial S^i \partial S^j} = 0, \quad t \geq 0. \tag{5.1.7}$$

Since θ is a replicating portfolio for V, the value of the option must always equal the value of the portfolio, that is $V(S_t,t) = \sum_{i=1}^{d} \theta_t^i S_t^i$ for each $t \geq 0$. Combining this observation with (5.1.6) shows that

$$V(S_t,t) = \sum_{i=1}^{d} \frac{\partial V(S_t,t)}{\partial S^i} S_t^i. \tag{5.1.8}$$

Furthermore, at expiry the option value is simply the payoff, i.e.

$$V(S_T,T) = P(S_T) \tag{5.1.9}$$

Equations (5.1.8) and (5.1.7) describe the option value through a partial differential equation with (5.1.9) as boundary condition. If we could solve the PDE problem posed above:

- From (5.1.5), we would have a self–financing trading strategy θ which replicates the option payoff $V(S_t,t)$;
- By investing the initial capital $V(S_0,0)$ we could implement the strategy θ to replicate the option exactly;
- **By principle (1.), $V(S_0,0)$ would be the fair price of the option!**

There is at least one remarkable feature of the equations (5.1.6), (5.1.7) and (5.1.8): none of these equations depend explicitly on the μ_i's, the drift coefficients of the assets! In a certain sense, the drift coefficients of the assets reflect investor attitudes towards risk. In particular, risk averse investors would demand higher return for higher risk ("larger diffusion means larger drift"). However, if investors were indifferent to risk levels, all the μ_i's would be equal – there would be no risk premium. Hence we could price options without knowing anything about investors risk preferences! The apparent indifference to risk inherent within our formulae makes sense because we can perfectly replicate the payoff of the option – there is no risk since any risk can be perfectly hedged by adopting the replicating strategy θ and by Principle (1.) the price of the option is simply the cost of implementing the hedging strategy.

5.1.5 Risk–Neutral Pricing

We now demonstrate how the theory above combines with Principle (2.) to make option pricing amenable to Monte Carlo Methods. If the payoff of an option is path dependent (e.g. Asian option), then the assumption that the option price at time t depends only on S_t generally fails to hold. Similarly, if the option depends on numerous underlying assets, solving the PDE (5.1.7) quickly becomes impractical. In these scenarios in particular, the Monte Carlo approach is favoured.

Suppose we are still in the market described by (5.1.1) and that the probability measure with respect to which this market is defined is called $\mathbb{P}_0$; henceforth $\mathbb{P}_0$ is referred to as the real–world probability measure.

Definition 5.1.10 (Attainable price process)**.** *A stochastic process V_t is called an attainable price process if there exists a self–financing strategy θ such that $V_t = \theta_t^T S_t$ a.s. for each $t \geq 0$.*

We now introduce a seemingly mysterious object whose necessity will soon become clear.

Definition 5.1.11 (Stochastic Discount Factor)**.** *A positive process $\{Z_t, \mathcal{F}_t : t \geq 0\}$ is called a stochastic discount factor (SDF) if for every attainable price process V, the process V/Z is a martingale. In other words,*

$$\frac{V_s}{Z_s} = \mathbb{E}_0\left[\frac{V_t}{Z_t}\,\middle|\,\mathcal{F}_s\right], \quad 0 \leq s \leq t, \quad \textit{for each attainable price process } V. \tag{5.1.10}$$

From the definition above, it is clear that if Z is an SDF, then so is aZ for any positive constant a. Hence we can fix $Z_0 = 1$ without loss of generality.

Now suppose that V_t is the value at time t of an option (which we are assuming is attainable of course). Suppose that this particular option expires at time T. Letting $s = 0$ and $t = T$ in (5.1.10) shows that

$$V_0 = \mathbb{E}_0\left[\frac{V_T}{Z_T}\middle|\mathcal{F}_0\right] = \mathbb{E}_0\left[\frac{V_T}{Z_T}\right]. \tag{5.1.11}$$

Thus **the option value at time zero is the expected option payoff modified by the SDF**. We might have expected the option price to simply be the expected payoff, i.e. $\mathbb{E}_0[V_t]$, but formula (5.1.11) differs from the expected payoff in two important ways:

(i.) Time Value of Money: the option payoff will only be paid at time T but we must pay the option premium at time zero. Hence the payoff must be discounted from time T back to time zero.

(ii.) Risk Adjustment: in the real–world investors are typically risk averse and demand higher return in exchange for bearing higher risk. If we calculate the price under the real–world measure we must take this into account and $\mathbb{E}_0[V_t]$ cannot be the price since it does not factor in risk preferences.

Formula (5.1.11) tells us that adjusting the payoff by the SDF allows us to price. Therefore the SDF must (somehow) take into account both the time value of money and investors preferences towards risk! Furthermore, the **existence of an SDF rules out arbitrage opportunities** in the market – this is a technical point which we do not elaborate upon further (see [1]) other than to reiterate that the absence of arbitrage is a necessary property for pricing.

Finally, we suppose that one of the assets in our market is riskless – by this we mean that its diffusion coefficient is zero and it produces a predictable rate of return which is known a priori. For the sake of simplicity we assume that this riskless asset has dynamics

$$d\beta_t = r\beta_t\, dt, \quad r \in \mathbb{R}, \quad t \geq 0.$$

Hence an investment of 1 unit in the asset β will be worth e^{rt} at time t. β is certainly an attainable price process and hence β/Z must be a martingale. Moreover, since β/Z is a strictly positive process we can use it to define a new probability measure (based on $\mathbb{P}_0$). To this end let

$$\mathbb{P}_\beta[A] = \mathbb{E}_0\left[\mathbb{1}_{\{A\}}\,\frac{\beta_t}{Z_t}\right], \quad A \in \mathcal{F}_t, \quad t \geq 0.$$

We call $\mathbb{P}_\beta$ the **risk–neutral measure** – this name will shortly be justified.

Exercise 5.1.12. *Use Definition 1.1.5 to check that $\mathbb{P}_\beta$ is a probability measure.*

Naturally, our risk–neutral measure $\mathbb{P}_\beta$ generates a corresponding expectation operator which is given by

$$\mathbb{E}_\beta[X] = \mathbb{E}_0\left[X\,\frac{\beta_t}{Z_t}\right], \quad X \text{ an } \mathcal{F}_t\text{–measurable random variable.}$$

Now consider equation (5.1.11) once more and write it in terms of the risk–neutral measure as follows:

$$V_0 = \mathbb{E}_0\left[\frac{V_T}{Z_T}\right] = e^{-rT}\mathbb{E}_0\left[\frac{\beta_T}{Z_T}\,V_T\right] = e^{-rT}\mathbb{E}_\beta\left[V_T\right]. \tag{5.1.12}$$

It is no exaggeration to say that the formula above is easily the single most important line of text in these notes.

Equation (5.1.12) says that **the fair time zero price of an option is the discounted expected payoff calculated under the risk–neutral measure.** The discount factor e^{-rT} is taking into account the time value of money so the risk–neutral measure must be accounting for investors risk preferences. Since investors will not demand increased returns for increased risk in a risk–neutral market all assets can be assumed to deliver the riskless rate of return under $\mathbb{P}_\beta$ — we don't need to estimate asset returns to price options! Furthermore, in simple markets, equation (5.1.12) also allows us to price options without explicitly finding the stochastic discount factor or replicating portfolios!

Practically, our pricing schema will be as follows:

(i) Replace the drift coefficients of the assets (the μ_i's) with the risk–free rate (i.e. work under $\mathbb{P}_\beta$) and simulate the asset paths;

(ii) Calculate the payoff of the option along each asset path;

(iii) Discount the simulated option payoffs at the risk–free rate;

(iv) Average simulated discounted payoffs to obtain the discounted expected payoff under the risk–neutral measure (which is the option price by (5.1.12)).

5.2 Computer Laboratory 7

Exercise 1: Option Pricing & the Black–Scholes Model

In the Black–Scholes market model the stock price process $\{S_t : t \geq 0\}$ is assumed to be a geometric Brownian motion, i.e. its dynamics are given by

$$dS_t = \mu S_t \, dt + \sigma S_t \, dB_t, \quad t > 0. \tag{5.2.1}$$

Complete the following tasks:

(a.) Using the Euler–Maruyama scheme to disretise (5.2.1), price a European call option with maturity of 1 year ($T = 1$) in the Black–Scholes market. Assume that there are 252 business days per year ($h = 1/252$). Take as parameters $S_0 = 100$, $K = 105$ (strike price), $\mu = r = 6\%$ (r is the risk free rate) and $\sigma = 15\%$.

(b.) Compare your answers from part (a.) to the exact closed form solution (see Appendix A).

(c.) Calculate 95% confidence intervals for the price of the option from part (a.) for different sample sizes.

Exercise 2: The Heston Model

One long-standing criticism of the Black-Scholes model is that observed market volatility is not constant. In response many authors have suggested so–called stochastic volatility models where the volatility itself is randomly distributed. One such model is that of Heston and is given by the following system of SDEs:

$$dS_t = rS_t dt + \sqrt{V_t} S_t dB_t^S, \qquad t > 0, \tag{5.2.2a}$$

$$dV_t = \kappa(\theta - V_t)dt + \xi\sqrt{V_t} dB_t^V, \qquad t > 0, \tag{5.2.2b}$$

where $\{V_t : t \geq 0\}$ is the volatility process, r is the risk–free rate, and B_t^S and B_t^V are Brownian motions which may be correlated; denote their correlation coefficient by ρ.

(a.) Price a European call option in the Heston model by discretising (5.2.2) using the Euler–Maruyama scheme. Model parameters: $S_0 = 100$, $V_0 = 20\%$, $K = 105$, $T = 1$, $h = 1/252$, $\mu = r = 6\%$, $\theta = 15\%$, $\kappa = 5$, $\xi = 1$. Compute prices for both $\rho = 0$ and $\rho = 0.5$

(b.) How does the price of the European call under the Heston model compare with the price in the Black–Scholes model? Are your results as expected? Explain.

Bibliography

[1] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford university press, 2009.

[2] P. Glasserman. *Monte Carlo Methods in Financial Engineering*, volume 53. Springer Science & Business Media, 2013.

[3] D. J. Higham. *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*, volume 13. Cambridge University Press, 2004.

[4] J. Jacod and P. Protter. *Probability Essentials*. Springer Science & Business Media, 2004.

[5] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer Science & Business Media, 2012.

[6] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1999.

[7] B. Øksendal. *Stochastic Differential Equations*. Springer, 2003.

Appendix A

The Black–Scholes Model

In the simplest version of this ubiquitous market model we have one (non–dividend paying) risky asset $S = \{S_t, \mathcal{F}_t : t \geq 0\}$ and one riskless asset β. The asset dynamics (under a "real–world" probability measure $\mathbb{P}_0$) are given by

$$dS_t = \mu S_t\, dt + \sigma S_t\, dB_t, \quad t \geq 0, \tag{S}$$
$$d\beta_t = r\beta_t\, dt, \quad t \geq 0, \tag{β}$$

where the initial values S_0 and β_0 are deterministic and assumed to be known a priori. The solution to the SDE (S) is given by

$$S_t = S_0 \exp\left((\mu - \sigma^2/2)t + \sigma B_t\right), \quad t \geq 0.$$

The formula above is easily verified using Itô's Lemma. Suppose V_t is the time t price for $t \in [0, T]$ of a European Call option with expiry time T and strike K. We must have

$$V_0 = e^{-rT}\mathbb{E}_0\left[\max\{S_0 \exp\left((r - \sigma^2/2)T + \sigma B_T\right) - K,\, 0\}\right], \tag{A.0.1}$$

where we have used formula (5.1.12). Since $B_T \sim N(0, T)$, the expectation in (A.0.1) can be computed in closed form. In fact, it can be shown that

$$V_0 = \Phi(d_1)S_0 - \Phi(d_2)Ke^{-rT}, \tag{A.0.2}$$

where:

- K is the strike price;
- T is the time to maturity;
- r is the risk–free rate;

- Φ is the cumulative distribution function of the standard Normal distribution;
- $$d_1 = \frac{1}{\sigma\sqrt{T}}\left[\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T\right];$$
- $$d_2 = d_1 - \sigma\sqrt{T}.$$

It follows from (A.0.2) and Put–Call parity that the price of a European Put option in the Black–Scholes market is given by

$$\Phi(-d_2)Ke^{-rT} - \Phi(-d_1)S_0.$$