MTH771P Foundations of Mathematical Modelling in Finance

Christian Beck

Chapter 1

Presessional Probability

We start by recalling some basic concepts from probability, which are essential for the development of mathematical models in finance. Much of this presessional material should be familiar to you. If not, here is the opportunity to learn it again.

1.1 Basic probability theory

Consider an experiment such as tossing a coin or throwing a die. The set of all possible outcomes is called the **sample space**. We shall often, but not always, use S to denote sample spaces.

Example 1.1. Flipping a coin. The outcome is either H (for heads) or T (for tails). The sample space is

$$S = \{H, T\} \, .$$

Example 1.2. Flipping a coin twice. The outcome is either H followed by H, H followed by T, T followed by H, or T followed by T. The sample space is

$$S = \{HH, HT, TH, TT\} \, .$$

Let S be a sample space with n elements, say $S = \{1, \ldots, n\}$. Suppose we are given n real numbers p_i, $i = 1, \ldots, n$, such that

(i) $p_i \geq 0$ for every $i \in S$;

(ii) $\sum_{i=1}^{n} p_i = 1$.

We can then interpret p_i to be the likelihood of the outcome i, for any $i \in S$, and we shall say that the p_i's define a **probability measure** on S.

Example 1.3. In Example 1.1 above, we could have chosen $p_H = p_T = \frac{1}{2}$, or $p_H = \frac{1}{3}$, $p_T = \frac{2}{3}$.
In Example 1.2 above, we could have chosen $p_{HH} = p_{HT} = p_{TH} = p_{TT} = \frac{1}{4}$.

A subset of the sample space S is called an **event**. If A is an event, then we define the probability of A occurring, denoted by $P(A)$, by

$$P(A) = \sum_{i \in A} p_i \, .$$

Note that, in particular, $P(S) = \sum_{i \in S} p_i = 1$.

Example 1.4. Let a, b, c be three companies. Let (i, j, k) be the outcome that in 2010 company i makes more profit than company j and that company j makes more profit than company k. Then the sample space is

$$S = \{(a,b,c), (a,c,b), (b,a,c), (b,c,a), (c,a,b), (c,b,a)\} .$$

Define a probability measure on S by letting

$$p_{(i,j,k)} = \frac{1}{6} \quad \text{for every } i,\ j,\ k.$$

Let A be the event that a makes most profit in 2010. Then $A = \{(a,b,c), (a,c,b)\}$ and $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Suppose that S is a sample space and that $A \subset S$ and $B \subset S$ are two events. Let us recall the following definitions.

- The **complement** of the event A, denoted by A' or $\overline{A}$, is given by

$$A' = \{\, s \in S \mid s \notin A \,\} .$$

- The **union** $A \cup B$ of the events A and B is given by

$$A \cup B = \{\, s \in S \mid s \in A \text{ or } s \in B \text{ or both} \,\} .$$

- The **intersection** $A \cap B$ of the events A and B is given by

$$A \cap B = \{\, s \in S \mid s \in A \text{ and } s \in B \,\} .$$

Note that

- $P(A \cup B)$ is the probability that at least one of A or B occurs;
- $P(A \cap B)$ is the probability that both A and B occur.

Example 1.5. Profits of companies (Example 1.4) continued. Let A be the event that company a makes most profit and let B be the event that company c makes least profit. Then

$$A = \{(a,b,c), (a,c,b)\} \quad \text{and} \quad B = \{(a,b,c), (b,a,c)\} .$$

We have $A \cup B = \{(a,b,c), (a,c,b), (b,a,c)\}$ and $A \cap B = \{(a,b,c)\}$. Moreover $P(A) = P(B) = \frac{2}{6} = \frac{1}{3}$, $P(A \cup B) = \frac{3}{6} = \frac{1}{2}$, and $P(A \cap B) = \frac{1}{6}$.

We now recall a useful result that relates $P(A \cap B)$ and $P(A \cup B)$.

Theorem 1.6. *For two events A and B*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) .$$

Example 1.7. In Example 1.5 above we have

$$P(A \cup B) = \frac{1}{2} = \frac{1}{3} + \frac{1}{3} - \frac{1}{6} = P(A) + P(B) - P(A \cap B) .$$

Example 1.8. Let the probability that the FTSE100 increases today be 0.52 and the probability that it increases tomorrow be 0.52 as well. Suppose that the probability it increases both today and tomorrow is 0.28. What is the probability that the FTSE100 increases neither today nor tomorrow?

Solution. Let A be the event that the FTSE100 increases today and let B be the event that the FTSE100 increases tomorrow. We know that $P(A) = P(B) = 0.52$, that $P(A \cap B) = 0.28$ and we want to find $P((A \cup B)')$. Now $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.76$, so $P((A \cup B)') = 1 - 0.76 = 0.24$ is the desired probability that the FTSE100 increases neither today nor tomorrow. □

Recall that if A and B are events, then the **conditional probability** of A given B, denoted $P(A|B)$, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,.$$

Example 1.9. Profits of companies (Example 1.5) continued. If A is the event that a makes most profit and B the event that c makes least profit, what is the probability that a makes most profit given that c makes least profit?

Solution.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/3} = \frac{1}{2}$$

is the desired probability that a makes most profit given that c makes least profit. □

We say that two events A and B are **independent** if

$$P(A|B) = P(A)\,,$$

or, equivalently, if

$$P(A \cap B) = P(A)P(B)\,.$$

Example 1.10. In Example 1.9 above, the two events A and B are dependent (that is, not independent), since $P(A|B) = \frac{1}{2} > P(A)$.

1.2 Random variables

Definition 1.11. A **random variable** is a quantity X determined by the outcome of an experiment. It is given by the following data:

(i) possible values $x_1, \ldots, x_n$ it can take on;

(ii) probabilities $p_1, \ldots, p_n$.

We interpret $p_i = P(X = x_i)$ to be the likelihood with which X takes the value x_i. The collection of the p_i's is referred to as the **probability distribution** of the random variable X.

Recall that if X is a random variable as defined above, then its **expectation**, denoted by $E(X)$, is given by

$$E(X) = \sum_{i=1}^{n} x_i p_i = \sum_{i=1}^{n} x_i P(X = x_i)\,.$$

Example 1.12. Suppose that a certain company

- makes £1,000,000 with probability $\frac{1}{4}$;
- loses £500,000 with probability $\frac{1}{4}$;
- makes £2,000,000 with probability $\frac{1}{2}$.

If X denotes the profit of the company in pounds, then X is a random variable with

$$x_1 = 1000000 \qquad p_1 = \frac{1}{4}$$
$$x_2 = -500000 \qquad p_2 = \frac{1}{4}$$
$$x_3 = 2000000 \qquad p_3 = \frac{1}{2}$$

In particular, the expected profit of the company in pounds is given by

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 = 1125000\,.$$

Definition 1.13. Let $p \in [0,1]$. A random variable X is said to **Bernoulli(p)** (or simply Bernoulli) distributed if its possible values are 0 and 1, and if $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

Note that if X is Bernoulli(p) distributed then $E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$.
Expectation is linear in the following sense:

Lemma 1.14. *Let X be a random variable and let a and b be constants. Then*

$$E(aX + b) = aE(X) + b\,.$$

Proof. Suppose that X takes on the value x_i with probability p_i. Then $aX + b$ is a random variable which takes on the values $ax_i + b$ with probability p_i. Thus

$$E(aX + b) = \sum_{i=1}^{n}(ax_i + b)p_i = a\sum_{i=1}^{n} x_ip_i + b\sum_{i=1}^{n} p_i = aE(X) + b\,.$$

□

Repeated application of the lemma above yields the following important result.

Proposition 1.15. *If $X_1, \ldots, X_m$ are random variables and $\alpha_1, \ldots, \alpha_m$ are constants, then*

$$E(\sum_{j=1}^{m} \alpha_jX_j) = \sum_{j=1}^{m} \alpha_jE(X_j)\,.$$

Definition 1.16. The **variance** of a random variable X is given by

$$\mathrm{Var}(X) = E((X - E(X))^2)\,.$$

The **standard deviation** of a random variable is given by

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}\,.$$

The following result is often useful.

Lemma 1.17. *If X is a random variable, then*

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 \,.$$

Proof. Using the linearity of the expectation we see that

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \,.\end{aligned}$$

□

Example 1.18. Let X be Bernoulli(p) distributed. Then $E(X) = p$ and

$$E(X^2) = 1^2 p + 0^2(1-p) = p \,,$$

so

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) \,.$$

Unlike expectation, variance is not linear as the following result shows.

Lemma 1.19. *If X is a random variable and a and b are constants, then*

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X) \,.$$

Proof. This is a short calculation using the definition of variance and is left as an exercise. □

While variance is not linear in general, it is sometimes possible to conclude that the variance of a sum of two random variables is the sum of the variances of the random variables. Before stating this result we recall the following important concept.

Definition 1.20. Two random variables X and Y are said to be **independent** if

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i) \,.$$

Furthermore, a sequence of random variables $X_1, X_2, \ldots$ is said to be independent if X_i and X_j are independent whenever $i \neq j$.

The result alluded to earlier can now be formulated as follows.

Proposition 1.21. *If $X_1, X_2, \ldots, X_m$ is a sequence of independent random variables, then*

$$\mathrm{Var}\left(\sum_{j=1}^{m} X_j\right) = \sum_{j=1}^{m} \mathrm{Var}(X_j) \,.$$

Proof. This will follow from a more general result to be discussed shortly (see Proposition 1.27). □

Another concept that you have already encountered and that will play an important role later on is the following. A **random walk** is a sum of independent random variables. To be precise, suppose that $X_1, X_2, \ldots$ is a sequence of random variables with the following properties:

- $P(X_j = 1) = P(X_j = -1) = \frac{1}{2}$ for all j;
- the random variables $X_1, X_2, \ldots$ are independent.

We can think of X_j as the j-th step of the random walk. Now define:

$$S_n = \sum_{j=1}^{n} X_j \,.$$

Then S_n is a random walk.

Note that

$$E(X_j) = 1\frac{1}{2} + (-1)\frac{1}{2} = 0\,,$$

so

$$E(S_n) = E(X_1) + \cdots E(X_n) = 0\,.$$

In order to determine the variance of S_n note that

$$E(X_j^2) = 1^2\frac{1}{2} + (-1)^2\frac{1}{2} = 1\,,$$

so

$$\mathrm{Var}(X_j) = E(X_j^2) - E(X_j)^2 = 1 - 0^2 = 1\,,$$

and thus, by Proposition 1.21

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \cdots \mathrm{Var}(X_n) = 1 + \cdots + 1 = n\,.$$

Recall that the **covariance** of two random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y)))\,.$$

Note that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ and that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. The following is a useful reformulation, the simple proof which is left as an exercise.

Lemma 1.22. *Let X and Y be two random variables. Then*

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)\,.$$

Covariance turns out to be linear in each of its arguments. We shall consider a special case first.

Lemma 1.23. *Let X_1, X_2 and Y be three random variables. Then*

$$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)\,.$$

Proof. Using the previous lemma and the linearity of expectation we see that

$$\begin{aligned}
\mathrm{Cov}(X_1 + X_2, Y) &= E((X_1 + X_2)Y) - E(X_1 + X_2)E(Y) \\
&= E(X_1Y + X_2Y) - (E(X_1) + E(X_2))E(Y) \\
&= E(X_1Y) + E(X_2Y) - E(X_1)E(Y) - E(X_2)E(Y) \\
&= E(X_1Y) - E(X_1)E(Y) + E(X_2Y) - E(X_2)E(Y) \\
&= \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)\,.
\end{aligned}$$

□

Repeated application of the previous lemma, together with the fact that $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ yields the following result.

Proposition 1.24. *Let X_i, $i = 1, 2, \ldots, m$ and Y_j, $j = 1, 2, \ldots n$ be two sequences of random variables. Then*

$$\mathrm{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j).$$

Proof. Try it as an exercise. □

The **correlation** of two random variables X and Y is given by

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}\,.$$

It is a non-trivial fact that

$$-1 \le \mathrm{Cor}(X,Y) \le 1\,.$$

We now recall that two random variables X and Y are said to **uncorrelated** if

$$\mathrm{Cov}(X,Y) = 0\,.$$

Lemma 1.25. *If two random variables X and Y are independent, then they are uncorrelated.*

Proof. Suppose that X and Y are independent and that X takes on values $x_1, \ldots, x_m$ while Y takes on values $y_1, \ldots, y_n$. Then

$$\begin{aligned} E(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i, Y = y_j) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i) P(Y = y_j) \\ &= \sum_{i=1}^{m} x_i P(X = x_i) \sum_{j=1}^{n} y_j P(Y = y_j) \\ &= E(X)E(Y)\,. \end{aligned}$$

Thus $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$. □

Note that the converse of the lemma is false, that is, X and Y uncorrelated does not imply that X and Y are independent, as the following example shows.

Example 1.26. Let X be a random variable taking on values $1, 0, -1$ and let Y be a random variable taking on values $1, 0$. Suppose that

$$P(X = 1, Y = 1) = P(X = -1, Y = 1) = P(X = 0, Y = 0) = \frac{1}{3}$$

and that the remaining joint probabilities are all 0. Thus

$$P(X = 1) = P(X = 0) = P(X = -1) = \frac{1}{3}$$

and

$$P(Y = 1) = \frac{2}{3}, \quad P(Y = 0) = \frac{1}{3}\,.$$

Now

$$E(X) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3} = 0\,,$$

while

$$E(XY) = 1 \cdot 1 \cdot \frac{1}{3} + (-1) \cdot 1 \cdot \frac{1}{3} + 0 \cdot 0 \cdot \frac{1}{3} = 0\,.$$

Thus $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$, so X and Y are uncorrelated. However, X and Y are not independent since

$$P(X = 0, Y = 0) = \frac{1}{3} \neq \frac{1}{9} = P(X = 0)P(Y = 0)\,.$$

Proposition 1.27. *Given a sequence $X_1, X_2, \ldots, X_n$ of random variables we have*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)\,.$$

If the random variables $X_1, X_2, \ldots, X_n$ are mutually uncorrelated (that is X_i and X_j are uncorrelated whenever $i \neq j$), then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)\,.$$

Proof. Follows from Proposition 1.24. □

1.3 Continuous random variables

Recall that a **continuous random variable** X takes values in $\mathbb{R}$ and is associated with a **probability density function** (abbreviated 'pdf'), that is, an integrable function f_X on $\mathbb{R}$ such that

(i) $f_X(x) \geq 0$, for every $x \in \mathbb{R}$;

(ii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

If a and b are real numbers with $a \leq b$ we interpret $\int_a^b f_X(x)\,dx$ to be the likelihood that X takes on values between a and b, that is,

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx\,.$$

Recall that if X is a continuous random variable the **expectation** $E(X)$ of X is given by

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx\,,$$

while the **variance** of X is given by

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)^2\,.$$

whenever these quantities are finite.

If X is a random variable and n is a positive integer, we say that the n-th moment of X exists if

$$E(|X|^n) < \infty\,.$$

If the n-th moment exists we call $E(X^n)$ the n-th **moment** of X. In particular, the first moment of a random variable equals its expectation (if it exists).

Example 1.28. Let $A > 0$ and let X be a random variable such that

$$f_X(t) = \frac{A}{\pi^2 A^2 + t^2}\,.$$

If this is the case we say that X has a **Cauchy distribution**. Note that $f_X(x) \geq 0$ for any real x. Moreover, using the substitution $x = \pi A \tan\theta$, we see that

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{A}{\pi^2A^2 + x^2}\,dx &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{A}{\pi^2A^2 + \pi^2A^2\tan^2\theta}\pi A\sec^2\theta\,d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\pi A^2\sec^2\theta}{\pi^2A^2\sec^2\theta}\,d\theta = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\pi}\,d\theta = \pi\frac{1}{\pi} = 1\,,\end{aligned}$$

so f_X is a proper pdf.

It turns out that the second moment of the Cauchy distribution does not exist. To see this note that

$$\lim_{x\to\infty} x^2 f_X(x) = \lim_{x\to\infty} \frac{Ax^2}{\pi^2A + x^2} = \lim_{x\to\infty} \frac{A}{\frac{\pi^2A}{x^2} + 1} = \frac{A}{1} = A\,,$$

so

$$E(|X|^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \infty\,,$$

that is, the second moment of X does not exist. In fact, the first moment does not exist either. In order to see this note that

$$\lim_{x\to\infty} \frac{x\frac{A}{\pi^2A^2+x^2}}{\frac{A}{x}} = 1\,,$$

so

$$E(|X|) = \int_{-\infty}^{\infty} |x|\,f_X(x)\,dx = 2\int_0^{\infty} x\frac{A}{\pi^2A^2 + x^2}\,dx \approx \int_0^{\infty} \frac{A}{x}\,dx = \infty\,,$$

because $\int \frac{1}{x}\,dx = \log x \to \infty$ as $x \to \infty$.

We now recall the following important concept.

Definition 1.29. The **characteristic function** G_X of a random variable X is defined for any $\alpha \in \mathbb{R}$ by

$$G_X(\alpha) = E(e^{i\alpha X}) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx\,.$$

Knowing the characteristic function of a random variable makes it possible to calculate its moments, as the following result shows.

Lemma 1.30. *Given a random variable X with characteristic function G_X we have for any $\alpha \in \mathbb{R}$*

$$G_X(\alpha) = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} E(X^k)\,.$$

Proof. Observe that

$$G_X(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(i\alpha x)^k}{k!} f_X(x)\,dx$$
$$= \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} \int_{-\infty}^{\infty} x^k f_X(x)\,dx = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} E(X^k)\,.$$

□

Example 1.31. Suppose that $X \sim \mathrm{Uniform}(0,1)$, that is, X is a random variable with

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$G_X(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx = \int_0^1 e^{i\alpha x} = \left[\frac{e^{i\alpha x}}{i\alpha}\right]_{x=0}^{x=1} = \frac{e^{i\alpha} - 1}{i\alpha}\,.$$

Now,

$$G_X(\alpha) = \frac{1}{i\alpha}\left(\sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} - 1\right) = \frac{1}{i\alpha} \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} = \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{(k+1)!} = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} \frac{1}{k+1}\,.$$

Using the previous lemma we see that

$$E(X^k) = \frac{1}{k+1}\,.$$

Note that we could also have calculated the moments directly:

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \int_0^1 x^k\,dx = \left[\frac{1}{k+1} x^{k+1}\right]_{x=0}^{x=1} = \frac{1}{k+1}\,.$$

Recall that the **joint probability distribution function** of two continuous random variables X and Y is denoted by $f_{X,Y}$ and has the property that for every $a \le b$ and every $c \le d$

$$P(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy\,.$$

In particular

$$P(a \le X \le b, -\infty \le Y \le \infty) = \int_{-\infty}^{\infty} \int_a^b f_{X,Y}(x,y)\,dx\,dy\,,$$

so

$$P(a \le X \le b) = \int_a^b \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx\,,$$

hence

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,.$$

Similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,.$$

It turns out that two continuous random variables are independent if and only if their joint probability distribution function factorises. To be precise:

Lemma 1.32. *Two continuous random variables X and Y are independent if and only if*

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, y \in \mathbb{R}\,.$$

1.4 Gaussian or normal random variables

A random variable X is said to be **Gaussian** or **normal** with mean μ and variance σ^2 where $\mu \in \mathbb{R}$ and $\sigma > 0$ if X has pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

If X is normal with parameters μ and σ^2 we write $X \sim N(\mu, \sigma)$. It turns out that

$$E(X) = \mu \quad \text{and} \quad \mathrm{Var}(X) = \sigma^2.$$

Furthermore, it turns out that the sum of two independent Gaussian random variables is again a Gaussian random variable. To be precise we have the following important result.

Lemma 1.33. *Suppose that $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$. If X_1 and X_2 are independent then*

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

If $\mu = 0$ and $\sigma = 1$, we say that X is **standard normal**, in which case

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Note that every normal random variable can be transformed to a standard normal random variable as follows.

Lemma 1.34. *If $X \sim \mathrm{N}(\mu, \sigma^2)$ then*

$$\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1).$$

Proof. See Remark 1.40. □

The **cumulative distribution function** of a standard normal random variable is defined to be

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

Note that if $X \sim \mathrm{N}(0,1)$, then $\Phi(x) = P(X \le x)$. Note also that

$$\Phi(x) = P(X \le x) = P(X \ge -x) = 1 - P(X \le -x) = 1 - \Phi(-x). \tag{1.1}$$

Moreover, if $a \le b$ then

$$P(a \le X \le b) = P(X \le b) - P(X \le a) = \Phi(b) - \Phi(a).$$

The function Φ cannot be expressed in terms of elementary functions. In this module we will use tables to determine the values of Φ. Using relation (1.1) we see that $\Phi(x)$ only needs to be tabulated for $x > 0$. In the table distributed in the lectures, $\Phi(x)$ is tabulated for arguments x, correct to 2 decimal places. If higher precision is required **linear interpolation** can be used. This is done as follows.

Suppose that $\underline{x} < x < \overline{x}$, where $\underline{x}$ and $\overline{x}$ are the nearest tabulated arguments of Φ. Then a good approximation to $\Phi(x)$ is given by

$$\Phi(x) \approx \frac{\overline{x} - x}{\overline{x} - \underline{x}} \Phi(\underline{x}) + \frac{x - \underline{x}}{\overline{x} - \underline{x}} \Phi(\overline{x}).$$

Example 1.35. Let $x = 1.116$. Then $\underline{x} = 1.11$ and $\overline{x} = 1.12$ and a good approximation of $\Phi(x)$ is

$$\Phi(x) \approx 0.4\Phi(1.11) + 0.6\Phi(1.12) = 0.4 \cdot 0.8665 + 0.6 \cdot 0.8686 = 0.8678\,.$$

Example 1.36. IQ scores of 11 year olds are normally distributed with mean value 100 and standard deviation 14.2. What is the probability that a randomly chosen 11 year has IQ more than 130?

Solution. Let X be the IQ of a randomly chosen 11 year old. We know that $X \sim \mathrm{N}(\mu, \sigma^2)$, where $\mu = 100$ and $\sigma^2 = (14.2)^2$. Now

$$P(X > 130) = P\left(\frac{X-\mu}{\sigma} > \frac{130-\mu}{\sigma}\right) = P\left(\frac{X-\mu}{\sigma} > 2.113\right) = 1 - \Phi(2.113)\,.$$

In order to determine $\Phi(2.113)$ we use linear interpolation. The nearest tabulated arguments are $\underline{x} = 2.11$ and $\overline{x} = 2.12$ and a good approximation to $\Phi(2.113)$ is

$$\Phi(2.113) \approx 0.7\Phi(2.11) + 0.3\Phi(2.12) = 0.7 \cdot 0.9826 + 0.3 \cdot 0.9830 = 0.9827\,.$$

Thus, the desired probability is

$$P(X > 130) = 1 - \Phi(2.113) = 0.017\,.$$

□

Note 1.37. You only need to use linear interpolation if you are specifically asked to do so.

Our next task is to derive the transformation formula for the probability distribution functions of continuous random variables. Before doing so recall that if X is a continuous random variable with pdf f_X, its **cumulative distribution function**, denoted F_X and abbreviated cdf, is defined to be

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt\,.$$

Note that by the fundamental theorem of calculus

$$\frac{d}{dx}F_X(x) = f_X(x)\,.$$

Suppose now that X is a continuous random variable and g a real valued function. We are now going to answer the question how the pdf of the random variable $Y = g(X)$ is related to that of X.

Theorem 1.38 (Transformation Formula)**.** *Let X be a continuous random variable and let $Y = g(X)$, where g is a differentiable function which is*

(i) either strictly monotonically increasing (so $g'(x) > 0\ \forall x \in \mathbb{R}$)

(ii) or strictly monotonically decreasing (so $g'(x) < 0\ \forall x \in \mathbb{R}$).

Then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| & \text{for all } y \text{ for which } g^{-1}(y) \text{ exists} \\ 0 & \text{for all other } y \end{cases}$$

Proof. We give the proof for strictly monotonically increasing g only. The other case is similar. We start by calculating the cdf of Y:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)\,dy = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))\,.$$

Now

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = F_X'(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\,,$$

by the chain rule. Observe that since g is monotonically increasing, so is g^{-1}. Thus, for all y for which $g^{-1}(y)$ exists

$$f_Y(y) = f_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|\,,$$

since

$$\frac{d}{dy}g^{-1}(y) > 0\,.$$

On the other hand, $F_Y(y)$ is either 0 or 1 for all y for which $g^{-1}(y)$ does not exist, so $f_Y(y) = 0$ in this case. □

Example 1.39. Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and let a and b be real constants with $a \neq 0$. Let $Y = aX + b$. What is the pdf of Y?

Solution. Write $g(x) = ax + b$. Then $Y = g(X)$. Note that $g'(x) = a$, so g is strictly monotonically increasing if $a > 0$ and strictly monotonically decreasing if $a < 0$. Now, if $y = ax + b$, then $x = (y - b)/a$, so

$$g^{-1}(y) = \frac{y-b}{a}\,,$$

and

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{a}\,.$$

Moreover, since $X \sim \mathrm{N}(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,.$$

Using the Transformation Formula we see that

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\right)\left|\frac{1}{a}\right| = \frac{1}{\sqrt{2\pi}\,|a|\,\sigma}\exp\left(-\frac{((y-a\mu-b)^2}{2a^2\sigma^2}\right)\,.$$

Thus

$$Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)\,.$$

□

Remark 1.40. The above example also shows that if $X \sim \mathrm{N}(\mu, \sigma^2)$, then

$$\frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right) \sim \mathrm{N}\left(\frac{1}{\sigma}\mu + \left(-\frac{\mu}{\sigma}\right), \frac{1}{\sigma^2}\sigma^2\right)\,,$$

that is

$$\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)\,.$$

1.5 Lognormal random variables

Lognormal random variables are a particular type of continuous random variables that occur in a number of practical applications.

Definition 1.41. A random variable Y is said to be **lognormal** with parameters μ and σ^2 where $\mu \in \mathbb{R}$ and $\sigma > 0$, if

$$\log Y \sim \mathrm{N}(\mu, \sigma^2)\,.$$

We write

$$Y \sim \mathrm{LogNormal}(\mu, \sigma^2)\,.$$

Note that if $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$, then $Y = \exp(X)$, where $X \sim \mathrm{N}(\mu, \sigma^2)$.

We will now use the Transformation Formula to determine the pdf of a lognormal random variable.

Proposition 1.42. *If $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$ then*

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right) & \text{if } y > 0; \\ 0 & \text{if } y \leq 0. \end{cases}$$

Proof. Write $g(x) = e^x$. Then $Y = g(X)$. Now, if $y = e^x$, then $x = \log y$, so

$$g^{-1}(y) = \log y \quad \text{for } y > 0\,,$$

and

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{y} \quad \text{for } y > 0\,.$$

Moreover, since $X \sim \mathrm{N}(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,.$$

Using the Transformation Formula we see that if $y > 0$ then

$$f_Y(y) = f_X(\log y)\frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)\,,$$

while $f_Y(y) = 0$ if $y \leq 0$. □

.

Proposition 1.43. *Let $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$. Then*

$$E(Y) = \exp(\mu + \frac{1}{2}\sigma^2) \quad \textit{and} \quad \mathrm{Var}(Y) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1)\,.$$

Proof. Try it as an exercise. □

Note that we can calculate the cdf of a lognormally distributed random variable using the table for Φ.

Example 1.44. Suppose $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$ with $\mu = 0.20$ and $\sigma = 0.50$. Determine y such that $P(Y \leq y) = 0.95$.

Solution. Note that $P(Y \leq y) = P(\log Y \leq \log y)$ where $\log Y \sim \mathrm{N}(\mu, \sigma^2)$. Thus

$$0.95 = P(Y \leq y) = P(\log Y \leq \log y) = P\left(\frac{\log Y - \mu}{\sigma} \leq \frac{\log y - \mu}{\sigma}\right) = \Phi\left(\frac{\log y - \mu}{\sigma}\right).$$

From the table for Φ we see that

$$\frac{\log y - \mu}{\sigma} = 1.645\,,$$

so

$$y = \exp(\mu + 1.645\sigma) = 2.78\,.$$

□

1.6 The Central Limit Theorem

One of the reasons that Gaussian random variables occur so often is the Central Limit Theorem (CLT), one of the gems of Probability Theory. To motivate the formulation of the CLT suppose for the moment that $X_1, X_2, \ldots, X_n$ are independent $\mathrm{N}(\mu, \sigma^2)$-distributed random variables. Then, by Lemma 1.33,

$$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)\,,$$

and so

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \sim \mathrm{N}(0,1)\,.$$

The CLT generalises this fact to sequences of independent random variables that need not have Gaussian distributions, but it only holds in the limit as $n \to \infty$. Here is a precise formulation.

Theorem 1.45 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be independent identically distributed random variables with mean $E(X_i) = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Define $S_n = \sum_{i=1}^n X_i$. Then*

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x) \quad (\forall x \in \mathbb{R})\,.$$

The conclusion of the CLT is often informally expressed as '$\frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to $\mathrm{N}(0,1)$'. The proof of the CLT relies on the following three auxiliary results.

Lemma 1.46. *Let Y_n be a sequence of random variables and let Z be a random variable. Suppose that their characteristic functions satisfy*

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = G_Z(\alpha) \quad (\forall \alpha \in \mathbb{R})\,.$$

Then

$$\lim_{n\to\infty} P(Y_n \leq x) = P(Z \leq x) \quad (\forall x \in \mathbb{R})\,.$$

Lemma 1.47. *If $X_1, X_2, \ldots, X_n$ are independent random variables and $S_n = X_1 + \cdots + X_n$, then*

$$G_{S_n}(\alpha) = G_{X_1}(\alpha) G_{X_2}(\alpha) \cdots G_{X_n}(\alpha) \quad (\forall \alpha \in \mathbb{R})\,.$$

Lemma 1.48. *Let Z be a random variable. Then*

$$Z \sim \mathrm{N}(0,1) \quad \textit{if and only if} \quad G_Z(\alpha) = \exp(-\frac{\alpha^2}{2})\,.$$

Proof of the CLT. We shall first prove a special case. Suppose for the moment that $\mu = 0$ and $\sigma = 1$. In order to prove the CLT in this case we need to show that

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) \leq x\right) = \Phi(x)\,.$$

Let

$$Y_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)\,.$$

We have

$$\begin{aligned} G_{Y_n}(\alpha) &= E(e^{i\alpha Y_n}) \\ &= E(e^{i\alpha\frac{1}{\sqrt{n}}(X_1+\cdots+X_n)}) \\ &= G_{S_n}\left(\frac{\alpha}{\sqrt{n}}\right) && \text{where } S_n = X_1 + \cdots + X_n \\ &= G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right) G_{X_2}\left(\frac{\alpha}{\sqrt{n}}\right) \cdots G_{X_n}\left(\frac{\alpha}{\sqrt{n}}\right) && \text{by Lemma 1.47} \\ &= G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right)^n, \end{aligned}$$

where the last equality holds since the X_i's are identically distributed. Moreover,

$$\begin{aligned} G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right) &= 1 + \sum_{k=1}^{\infty} \frac{\left(\frac{i\alpha}{\sqrt{n}}\right)^k}{k!} E(X_1^k) \\ &= 1 + \frac{i\alpha}{\sqrt{n}} E(X_1) + \frac{1}{2}\left(\frac{i\alpha}{\sqrt{n}}\right)^2 E(X_1^2) + \cdots \\ &= 1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots \end{aligned}$$

since, by hypothesis, $E(X_1) = 0$ and $E(X_1^2) = \mathrm{Var}(X_1) + E(X_1)^2 = 1$. Thus

$$G_{Y_n}(\alpha) = G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right)^n = \left(1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots\right)^n = \exp\left(n \log\left(1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots\right)\right)\,. \quad (1.2)$$

In order to proceed we need the Taylor-Maclaurin expansion of $\log(1-x)$, which we now quickly derive. Note that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

thus

$$\underbrace{\int_0^x \frac{1}{1-t}\,dt}_{=-\log(1-x)} = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \cdots$$

so

$$\log(1-x) = -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{1}{4}x^4 - \cdots$$

Using the first term of the above expansion of $\log(1-x)$ in equation 1.2 we see that

$$G_{Y_n}(\alpha) = \exp\left(n\left(-\frac{1}{2}\frac{\alpha^2}{n} - \cdots\right)\right) = \exp\left(-\frac{\alpha^2}{2} - \cdots\right)\,,$$

so

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = \exp\left(-\frac{\alpha^2}{2}\right) .$$

Let now $Z \sim \mathrm{N}(0,1)$. By Lemma 1.48,

$$\exp\left(-\frac{\alpha^2}{2}\right) = G_Z(\alpha)\,,$$

so combining the last two equations gives

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = G_Z(\alpha)\,,$$

and Lemma 1.46 now implies

$$\lim_{n\to\infty} P(Y_n \le x) = P(Z \le x) = \Phi(x)\,,$$

that is, we have proved the CLT in the special case where $\mu = 0$ and $\sigma = 1$.

Now we prove the CLT when μ and $\sigma > 0$ are arbitrary, using the fact the CLT is true for $\mu = 0$ and $\sigma = 1$. For $i \in \mathbb{N}$ call

$$W_i = \frac{X_i - \mu}{\sigma}\,.$$

Then

$$E(W_i) = \frac{\mu - \mu}{\sigma} = 0$$

and

$$\mathrm{Var}(W_i) = \frac{1}{\sigma^2}\mathrm{Var}(X_i) = \frac{1}{\sigma^2}\sigma^2 = 1\,.$$

Since the W_i's are independent random variables with mean 0 and variance 1 we can now apply the special case of the CLT we have just established to conclude that

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}(W_1 + \cdots + W_n) \le x\right) = \Phi(x)\,,$$

so

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}\left((X_1 - \mu) + \cdots + (X_n - \mu)\right) \le x\right) = \Phi(x)\,,$$

hence

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu) \le x\right) = \Phi(x)\,,$$

thus

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x)\,,$$

and the proof of the general case of the CLT is finished. □

Chapter 2

Introduction to Stochastic Processes

2.1 Basic definition

In mathematical finance, asset price dynamics is modelled by suitable stochastic processes. Let us start by defining what a stochastic process is from a mathematical point of view.

A **stochastic process** is a function of time t and a random variable X. We write $Y(t) = f(t, X)$. In other words, $Y(t)$ is a family of random variables indexed by a parameter t usually interpreted as time. If X is fixed, then $Y(t)$ becomes an ordinary function of time t, which we call a **realisation** of the stochastic process.

A stochastic process is completely determined by its n-**point probability density functions**

$$f_n(y_1, t_1; y_2, t_2; \ldots; y_n, t_n) \quad \text{where } 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n,$$

which have the property that

$$P(a_1 \leq Y(t_1) \leq b_1, a_2 \leq Y(t_2) \leq b_2, \ldots, a_n \leq Y(t_n) \leq b_n) = \int_{a_n}^{b_n} \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_1}^{b_1} f_n(y_1, t_1; y_2, t_2; \ldots; y_n, t_n)\, dy_1\, dy_2 \cdots dy_n\,,$$

for all $a_1 \leq b_1, a_2 \leq b_2, \ldots, a_n \leq b_n$.

Note that, in general, $Y(t_1), Y(t_2), \ldots Y(t_n)$ are dependent so, in general,

$$f_n(y_1, t_1; y_2, t_2; \ldots; y_n, t_n) \neq f_1(y_1, t_1) f_1(y_2, t_2) \cdots f_1(y_n, t_n)\,.$$

While a stochastic process is completely determined by its n-point probability density functions, it turns out to be useful to introduce n-**point conditional probability density functions** defined by

$$f_n(y_n, t_n | y_{n-1}, t_{n-1}; y_{n-2}, t_{n-2}; \ldots; y_1, t_1) = \frac{f_n(y_1, t_1; y_2, t_2; \ldots; y_n, t_n)}{f_{n-1}(y_1, t_1; y_2, t_2; \ldots; y_{n-1}, t_{n-1})}\,.$$

We think of $y_1, y_2, \ldots, y_{n-1}$ as being given and interpret $f_n(y_n, t_n | y_{n-1}, t_{n-1}; y_{n-2}, t_{n-2}; \ldots; y_1, t_1)$ as the pdf of $Y(t_n)$ conditional on the event $Y(t_1) = y_1, Y(t_2) = y_2, \ldots, Y(t_{n-1}) = y_{n-1}$.

We are now going to focus on a particularly simple type of stochastic process. A stochastic process is called a **Markov process** if

$$f_n(y_n, t_n | y_{n-1}, t_{n-1}; y_{n-2}, t_{n-2}; \ldots; y_1, t_1) = f_2(y_n, t_n | y_{n-1}, t_{n-1})\,.$$

This means that for a Markov process $Y(t_n)$ does not depend on the complete history $Y(t_1) = y_1, Y(t_2) = y_2, \ldots, Y(t_{n-1}) = y_{n-1}$, but only depends on the most recent value $Y(t_{n-1})$. We call

$$f_2(y_n, t_n | y_{n-1}, t_{n-1})$$

the **transition probability density function** of the Markov process. Note that

$$f_2(y_n, t_n | y_{n-1}, t_{n-1}) = \frac{f_2(y_{n-1}, t_{n-1}; y_n, t_n)}{f_1(y_{n-1}, t_{n-1})}$$

so

$$f_2(y_{n-1}, t_{n-1}; y_n, t_n) = f_2(y_n, t_n | y_{n-1}, t_{n-1}) f_1(y_{n-1}, t_{n-1}) \,.$$

Moreover,

$$f_3(y_n, t_n | y_{n-1}, t_{n-1}; y_{n-2}, t_{n-2}) = \frac{f_3(y_{n-2}, t_{n-2}; y_{n-1}, t_{n-1}; y_n, t_n)}{f_2(y_{n-2}, t_{n-2}; y_{n-1}, t_{n-1})} \,,$$

so, for a Markov process,

$$\begin{aligned} f_3(y_{n-2}, t_{n-2}; y_{n-1}, t_{n-1}; y_n, t_n) &= f_3(y_n, t_n | y_{n-1}, t_{n-1}; y_{n-2}, t_{n-2}) f_2(y_{n-2}, t_{n-2}; y_{n-1}, t_{n-1}) \\ &= f_2(y_n, t_n | y_{n-1}, t_{n-1}) f_2(y_{n-2}, t_{n-2}; y_{n-1}, t_{n-1}) \\ &= f_2(y_n, t_n | y_{n-1}, t_{n-1}) f_2(y_{n-1}, t_{n-1} | y_{n-2}, t_{n-2}) f_1(y_{n-2}, t_{n-2}) \end{aligned}$$

Similarly, one can show that for a Markov process

$$f_n(y_1, t_1; \ldots; y_n, t_n) = f_2(y_n, t_n | y_{n-1}, t_{n-1}) f_2(y_{n-1}, t_{n-1} | y_{n-2}, t_{n-2}) \cdots f_2(y_2, t_2 | y_1 . t_1) f_1(y_1, t_1) \,.$$

In other words, a Markov process is completely determined by its 1-point probability density function and its transition probability density function.

2.2 The Wiener process

The Wiener process, sometimes also called Brownian motion, is an extremely important example of a stochastic process. It is a Gaussian Markov process, and has applications in many different areas of science.

Definition 2.1. The **Wiener process** or **Brownian motion**, usually denoted $W(t)$, is a Markov process with 1-point pdf

$$f_1(y_1, t_1) = \frac{1}{\sqrt{2\pi t_1}} \exp\left(-\frac{y_1^2}{2t_1}\right)$$

and transition pdf

$$f_2(y_2, t_2 | y_1, t_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \exp\left(-\frac{(y_2 - y_1)^2}{2(t_2 - t_1)}\right) \,.$$

Note that the special form of the 1-point pdf of the Wiener process implies that

$$W(t) \sim \mathrm{N}(0, t) \quad (\forall t > 0) \,,$$

and one also has

$$W(t) - W(s) \sim \mathrm{N}(0, t - s) \quad (\forall t > s \geq 0) \,.$$

It turns out that these two properties provide a characterisation of the Wiener process contained in the following theorem:

Theorem 2.2. *The Wiener process $W(t)$ is the unique Markov process such that*

(a) $W(0) = 0$

(b) $W(t) \sim \mathrm{N}(0, t)$ *for every* $t > 0$

(c) $W(t) - W(s) \sim \mathrm{N}(0, t - s)$ *for every* $t > s \geq 0$

(d) for all time points $0 \leq t_1 < t_2 < \cdots < t_n$ *the random variables*

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1})$$

are independent.

For a stochastic process $S(t)$, the difference $S(t) - S(s)$ where $t > s$ is known as an 'increment' of the stochastic process. Property (d) above can then be rephrased as saying that the Wiener process has **independent increments**.

Note that the theorem above implies the following. If $W(t)$ is the Wiener process then for every $t > 0$

$$E(W(t)) = 0 \quad \text{and} \quad \mathrm{Var}(W(t)) = t$$

since $W(t) \sim \mathrm{N}(0, t)$. Slightly deeper is the following result.

Lemma 2.3. *If $W(t)$ is the Wiener process and $0 \leq s < t$ then*

$$\mathrm{Cov}(W(s), W(t)) = \min(s, t) = s\,.$$

Proof. We have

$$\begin{aligned}
\mathrm{Cov}(W(s), W(t)) &= E(W(s)W(t)) - E(W(s))E(W(t)) \\
&= E(W(s)W(t)) \\
&= E(W(s)[W(s) + W(t) - W(s)]) \\
&= E(W(s)^2) + E(W(s)[W(t) - W(s)]) \\
&= \mathrm{Var}(W(s)) + E([W(s) - W(0)][W(t) - W(s)]) \\
&\overset{(*)}{=} s + E(W(s) - W(0))E(W(t) - W(s)) \\
&= s\,.
\end{aligned}$$

Here equation (*) follows from the fact that $W(s) - W(0)$ and $W(t) - W(s)$ are independent by the independent increments property of the Wiener process (Theorem 2.2 (d)).

□

When performing concrete calculations with the Wiener process it is useful to remember that, since $W(t) \sim \mathrm{N}(0, t)$, we have

$$\frac{1}{\sqrt{t}} W(t) \sim \mathrm{N}(0, 1)\,.$$

Example 2.4. Let $W(t)$ be the Wiener process. What is $P(W(4) > 5)$?

Solution.

$$P(W(4) > 5) = P\left(\frac{W(4)}{\sqrt{4}} > \frac{5}{\sqrt{4}}\right) = 1 - \Phi\left(\frac{5}{2}\right) = 1 - 0.9938 = 0.0062\,.$$

□

In a more intuitive picture, the Wiener process can be regarded as a rescaled version of a random walk. Recall that S_n is a random walk if

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

where the X_i's are independent and identically distributed random variables with

$$P(X_i = 1) = P(X_i = -1) = \frac{1}{2} \,.$$

Before formulating this result let us just recall the following definition. Given a real number x, the **floor** of x, denoted by $\lfloor x \rfloor$, is the largest integer less than or equal to x. Thus, for example,

$$\lfloor 1 \rfloor = 1 \,, \quad \lfloor 3.2 \rfloor = 3 \,, \quad \lfloor 5.9 \rfloor = 5 \,.$$

We are now able to formulate the following interesting fact:

Theorem 2.5 (Convergence of the random walk to the Wiener process)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common distribution given by $P(X_i = 1) = P(X_i = -1) = 1/2$ for all i. For $t > 0$ let*

$$Y(t) = \lim_{h \to 0} \sqrt{h} \sum_{i=1}^{\lfloor t/h \rfloor} X_i \,.$$

Then $Y(t)$ is the Wiener process.

The above formula for $Y(t) = W(t)$ is often used as a definition of the Wiener process (= Brownian motion) in the more physically orientated literature.

A typical Brownian path (a realization of the Wiener process) is continuous, but nowhere differentiable. This can be seen from the theorem above, by noting that a typical random walk of infinite length has infinitely many points where it changes direction.

2.3 Brownian motion with drift

For finance models one often has to consider a systematic drift in addition to random-walk like behaviour.

Definition 2.6. Brownian motion with drift having **drift parameter** $\mu \in \mathbb{R}$ and **volatility parameter** $\sigma > 0$ is the stochastic process $Y(t)$ given by

$$Y(t) = \mu t + \sigma W(t) \,,$$

where $W(t)$ is the Wiener process.

Note that if $Y(t)$ is Brownian motion with drift parameter μ and volatility parameter σ, then for every $t > 0$

$$E(Y(t)) = E(\mu t + \sigma W(t)) = \mu t + \sigma E(W(t)) = \mu t \,,$$

while

$$\mathrm{Var}(Y(t)) = \mathrm{Var}(\mu t + \sigma W(t)) = \mathrm{Var}(\sigma W(t)) = \sigma^2 \mathrm{Var}(W(t)) = \sigma^2 t \,.$$

But since for every $t > 0$ the random variable $Y(t)$ is normal we have

$$Y(t) \sim \mathrm{N}(\mu t, \sigma^2 t) \quad (\forall t > 0) \,.$$

This allows us to perform calculations for Brownian motion with drift, just as for the Wiener process itself.

Example 2.7. Let $Y(t)$ be Brownian motion with drift parameter $\mu = 0.2$ and volatility parameter $\sigma = 0.1$. Find $E(Y(9))$, $\mathrm{Var}(Y(9))$ and $P(Y(9) > 1)$.

Solution. Using the above results, we obtain

$$E(Y(9)) = 9\mu = 1.8\,,$$

and

$$\mathrm{Var}(Y(9)) = 9\sigma^2 = 0.09\,.$$

Finally, we have

$$\begin{aligned} P(Y(9) > 1) &= P(9\mu + \sigma W(9) > 1) \\ &= P\left(W(9) > \frac{1 - 9\mu}{\sigma}\right) \\ &= P(W(9) > -8) \\ &= P\left(\frac{W(9)}{\sqrt{9}} > -\frac{8}{\sqrt{9}}\right) \\ &= \Phi\left(\frac{8}{3}\right) \\ &= 0.9962\,. \end{aligned}$$

□

2.4 Geometric Brownian motion

Brownian motion with drift can become negative, a feature that rules it out as a suitable model for let's say a share price in the stock market. But the following variant does not suffer from this shortcoming, and provides a decent model of share prices.

Definition 2.8. Geometric Brownian motion having **drift parameter** $\mu \in \mathbb{R}$, **volatility parameter** $\sigma > 0$, and **starting parameter** S_0 is the stochastic process $S(t)$ given by

$$S(t) = S_0 \exp(\mu t + \sigma W(t))\,,$$

where $W(t)$ is the Wiener process.

Note that

$$S(0) = S_0 \exp(0 + 0) = S_0$$

hence S_0 is just denoting the initial value. Moreover,

$$\frac{S(t)}{S_0} = \exp(\mu t + \sigma W(t))$$

so

$$\log \frac{S(t)}{S_0} = \mu t + \sigma W(t)$$

where

$$\mu t + \sigma W(t) \sim \mathrm{N}(\mu t, \sigma^2 t)\,.$$

Thus

$$\frac{S(t)}{S_0} \sim \text{LogNormal}(\mu t, \sigma^2 t)\,.$$

Observe also that for any $0 \leq t_1 < t_2 < \cdots < t_n$ the successive ratios

$$\begin{aligned}
\frac{S(t_2)}{S(t_1)} &= \exp(\mu(t_2 - t_1) + \sigma(W(t_2) - W(t_1))) \\
\frac{S(t_3)}{S(t_2)} &= \exp(\mu(t_3 - t_2) + \sigma(W(t_3) - W(t_2))) \\
&\cdots \\
\frac{S(t_n)}{S(t_{n-1})} &= \exp(\mu(t_n - t_{n-1}) + \sigma(W(t_n) - W(t_{n-1})))
\end{aligned}$$

are all independent since Brownian motion has independent increments. Furthermore, the ratios are all lognormally distributed. The above is indeed a continuous-time version of so-called IID lognormal models.

Assuming that a share price evolves according to geometric Brownian motion $S(t)$, one can calculate probabilities for all kinds of events.

Example 2.9. Suppose a particular share price evolves according to geometric Brownian motion with drift parameter $\mu = 0.004$ and volatility parameter $\sigma = 0.1$, where time t is measured in weeks.

What is the probability that the price of this share rises without interruption from one day to the next for the next 3 days?

Solution. We shall first calculate the probability p that the price of the share is higher at the end of a day compared to the beginning. Now

$$p = P(S(1/7) > S(0)) = P\left(\log \frac{S(1/7)}{S(0)} > 0\right)\,.$$

But

$$\log \frac{S(1/7)}{S(0)} \sim \text{N}(\frac{1}{7}\mu, \frac{1}{7}\sigma^2)\,,$$

so

$$\begin{aligned}
p = P\left(\log \frac{S(1/7)}{S(0)} > 0\right) &= P\left(\frac{\log \frac{S(1/7)}{S(0)} - \frac{1}{7}\mu}{\sqrt{\frac{1}{7}}\sigma} > -\frac{\frac{1}{7}\mu}{\sqrt{\frac{1}{7}}\sigma}\right) \\
&= 1 - \Phi\left(-\sqrt{\frac{1}{7}}\frac{\mu}{\sigma}\right) = \Phi\left(\sqrt{\frac{1}{7}}\frac{\mu}{\sigma}\right) = \Phi(0.02) = 0.5080\,.
\end{aligned}$$

But since the the random variables

$$\frac{S(1/7)}{S(0)}, \frac{S(2/7)}{S(1/7)}, \frac{S(3/7)}{S(2/7)}\,,$$

are all independent and have the same distribution, the probability that the price of the share rises from one day to the next for the next 3 days is $p^3 = 0.13 = 13\%$.

□

We saw earlier on that the Wiener process can obtained from a random walk by a limiting procedure (see Theorem 2.5). It turns out that a similar result holds for geometric Brownian motion. We shall now describe the multiplicative random walk-like process that will yield geometric Brownian motion in a suitable limit.

Let h denote a small increment of time. After every time interval of length h the previous price S_0 of some financial product either goes up or down.

- If it goes up, it goes up by multiplication with the factor $u = e^{\sigma\sqrt{h}}$;
- If it goes down, it goes down by multiplication with the factor $d = e^{-\sigma\sqrt{h}}$.

Note that $ud = 1$. The price changes are random, where

- the probability that it goes up is $p = \frac{1}{2}(1 + \frac{\mu}{\sigma}\sqrt{h})$;
- and the probability that it goes down is $1 - p = \frac{1}{2}(1 - \frac{\mu}{\sigma}\sqrt{h})$.

Here, μ and σ are the drift and volatility parameters of the geometric Brownian motion which we are approximating.

Theorem 2.10. *As $h \to 0$ the random walk model we have just described tends to geometric Brownian motion with drift parameter μ, volatility parameter σ, and starting parameter S_0.*

Proof. Let $S(t)$ denote the limit of the above model as $h \to 0$. We shall only show that $\frac{S(t)}{S(0)}$ approaches a lognormal distribution with parameters μt and $\sigma^2 t$ as $h \to 0$.

Let $t = nh$ for some integer n. Note that the number of jumps the price has made so far is n. Next define random variables $X_1, X_2, \ldots, X_n$ by

$$X_i = \begin{cases} 1 & \text{if the price goes up at time } ih, \\ 0 & \text{if the price goes down at time } ih, \end{cases}$$

where $i = 1, 2, \ldots, n$. Observe that the total number of jumps up is $\sum_{i=1}^n X_i$ and the total number of jumps down is $n - \sum_{i=1}^n X_i$. Now

$$S(nh) = S_0 u^{\sum_{i=1}^n X_i} d^{n - \sum_{i=1}^n X_i} = S_0 d^n \Big(\frac{u}{d}\Big)^{\sum_{i=1}^n X_i},$$

hence, since $t = nh$,

$$\frac{S(t)}{S(0)} = d^{t/h} \Big(\frac{u}{d}\Big)^{\sum_{i=1}^{t/h} X_i},$$

thus

$$\log \frac{S(t)}{S(0)} = \frac{t}{h} \log d + \sum_{i=1}^{t/h} X_i \log\left(\frac{u}{d}\right).$$

Recalling that $u = e^{\sigma\sqrt{h}}$ and $d = e^{-\sigma\sqrt{h}}$ we see that

$$\log d = -\sigma\sqrt{h} \quad \text{and} \quad \log\left(\frac{u}{d}\right) = 2\sigma\sqrt{h}\,.$$

Therefore

$$\log \frac{S(t)}{S(0)} = -\frac{\sigma t}{\sqrt{h}} + 2\sigma\sqrt{h} \sum_{i=1}^{t/h} X_i\,.$$

Since the X_i's are all independent, the CLT (Theorem 1.45) now implies that the distribution of the random variable $\sum_{i=1}^{t/h} X_i$ becomes more and more normal as $h \to 0$. Hence $\log \frac{S(t)}{S(0)}$ becomes more and more normal as $h \to 0$. But

$$E\left(\log \frac{S(t)}{S(0)}\right) = -\frac{\sigma t}{\sqrt{h}} + 2\sigma\sqrt{h}\sum_{i=1}^{t/h} E(X_i) = -\frac{\sigma t}{\sqrt{h}} + 2\sigma\sqrt{h}\sum_{i=1}^{t/h} p\,,$$

since $E(X_i) = p$, so

$$E\left(\log \frac{S(t)}{S(0)}\right) = -\frac{\sigma t}{\sqrt{h}} + 2\sigma\sqrt{h}\frac{t}{h}\frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{h}\right) = \mu t\,.$$

Moreover, since the X_i's are all independent,

$$\mathrm{Var}\left(\log \frac{S(t)}{S(0)}\right) = \left(2\sigma\sqrt{h}\right)^2 \sum_{i=1}^{t/h} \mathrm{Var}(X_i) = 4\sigma^2 h\frac{t}{h}\mathrm{Var}(X_i)\,.$$

But $\mathrm{Var}(X_i) = p(1-p)$, so

$$\mathrm{Var}\left(\log \frac{S(t)}{S(0)}\right) = 4\sigma^2 h\frac{t}{h}\frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{h}\right)\frac{1}{2}\left(1 - \frac{\mu}{\sigma}\sqrt{h}\right) \to 4\sigma^2 t \cdot \frac{1}{2}\cdot 1 \cdot \frac{1}{2}\cdot 1 = \sigma^2 t$$

as $h \to 0$. Therefore the distribution of $\log \frac{S(t)}{S(0)}$ approaches $\mathrm{N}(\mu t, \sigma^2 t)$ as $h \to 0$, and hence, as $h \to 0$, the distribution of $\frac{S(t)}{S(0)}$ approaches $\mathrm{LogNormal}(\mu t, \sigma^2 t)$, as required. □

Chapter 3

Basic facts on interest rates and present value analysis

A central ingredient in finance models is the fact that interest is being paid. Interest rates can be either constant or time dependent. We recall the basic facts from a mathematical point of view.

3.1 Constant interest rates

Suppose we borrow an amount P, called the **principal**, with interest at rate r per time period T.[1] This means that after time period T we need to repay the principal and the interest, that is

$$P + rP = P(1 + r)\,.$$

By definition, if interest is **compounded** every $\frac{1}{n}$-th of a year, then after one year we owe

$$P\Big(1 + \frac{r}{n}\Big)^n\,. \tag{3.1}$$

Example 3.1. On your credit card you borrow money at interest rate 18% compounded monthly. How much do you owe after one year?

Solution. Let P be the amount borrowed. After one year you owe

$$P\left(1 + \frac{0.18}{12}\right)^{12} = 1.195P\,.$$

□

When interest is compounded, the quantity r is called the **nominal rate**. In this case, the **effective interest rate** r_{eff} is defined to be

$$r_{eff} = \frac{\text{amount paid after time period } T - P}{P}\,.$$

For compound interest we have

$$r_{eff} = \frac{P(1 + r/n)^n - P}{P} = \Big(1 + \frac{r}{n}\Big)^n - 1\,.$$

[1] Usually T is equal to one year, and we shall adopt the convention that if no time period is explicitly specified then T equals a year.

Imagine now that interest is compounded at ever shorter time intervals. In the limiting case as $n \to \infty$ we shall say that interest is **compounded continuously**. To be precise, when interest is compounded continuously at a nominal rate r per time period, the amount to be repaid after one period is

$$\lim_{n\to\infty} P\Big(1+\frac{r}{n}\Big)^n = Pe^r\,.$$

Note that for continuously compounded interest the effective interest rate is

$$r_{eff} = \frac{Pe^r - P}{P} = e^r - 1\,.$$

More generally, if interest is charged at nominal rate r compounded every $\frac{1}{n}$-th of a year, then after t years you owe

$$\underbrace{P\Big(1+\frac{r}{n}\Big)^n\Big(1+\frac{r}{n}\Big)^n\cdots\Big(1+\frac{r}{n}\Big)^n}_{t \text{ times}} = P\Big(1+\frac{r}{n}\Big)^{nt};$$

and if interest is compounded continuously you owe

$$\underbrace{Pe^re^r\cdots e^r}_{t \text{ times}} = Pe^{rt}$$

after t years.

3.2 Present value analysis

The question we want to explore in this section is the following. How much is a payment received in the future worth today?

To get a feeling for what is going on here, imagine the following scenario. Suppose the current interest rate is 10%. If somebody gives you 100 pounds today, you can put it in your bank and after a year you will have 110 pounds. Thus 100 pounds received today have a value of 110 pounds in a year's time, and conversely, 110 pounds received in a year's time have a present value of 100 pounds.

More generally, suppose you can borrow and lend money at nominal rate r per time period. What is the value of V pounds received at the end of period i? If you borrow the amount $(1+r)^{-i}V$ now, then you can put it in the bank and at the end of period i you will have $(1+r)^{-i}V(1+r)^i = V$ pounds. Thus, V pounds received in i periods of time at interest rate r is worth $(1+r)^{-i}V$ in today's money. Thus, we define the **present value** of V pounds at the end of period i to be $(1+r)^{-i}V$.

We can generalise the notion of present value to a **cash flow stream** $\mathbf{a} = (a_1, a_2, \ldots, a_n)$, that is, a sequence of payments, where a_i is the payment made at the end of period i. The present value of the cash flow stream $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is denoted as $PV(\mathbf{a})$ and defined as

$$PV(\mathbf{a}) = \sum_{i=1}^{n} \frac{a_i}{(1+r)^i}\,.$$

This concept helps to compare the value of payments obtained at different times.

Example 3.2. You are offered three different jobs. The salary paid at the end of each year (in multiples of 1000 pounds) is

Job	1	2	3	4	5
A	12	14	16	18	20
B	16	16	15	15	15
C	20	16	14	12	10

Which is the best job for $r = 0.1$, $r = 0.2$, and $r = 0.3$?

Solution. We shall compare the present values of the cash flow streams. Now, the present value for job A is

$$\frac{12}{1+r} + \frac{14}{(1+r)^2} + \frac{16}{(1+r)^3} + \frac{18}{(1+r)^4} + \frac{20}{(1+r)^5},$$

while the present value for job B is

$$\frac{16}{1+r} + \frac{16}{(1+r)^2} + \frac{15}{(1+r)^3} + \frac{15}{(1+r)^4} + \frac{15}{(1+r)^5}.$$

Finally, the present value for job C is

$$\frac{20}{1+r} + \frac{16}{(1+r)^2} + \frac{14}{(1+r)^3} + \frac{12}{(1+r)^4} + \frac{10}{(1+r)^5}.$$

Taking the different interest rates into account we get the following present values.

r	PV A	PV B	PV C
0.1	59.21	58.60	56.33
0.2	45.70	46.39	45.69
0.3	36.49	37.89	38.12

Thus,

- if $r = 10\%$, then A is best, then B, then C;
- if $r = 20\%$, then B is best, then A, then C;
- if $r = 30\%$, then C is best, then B, then A.

□

Clearly, present values of constant payments received in the future or past are related to geometric series. Let us recall some facts about geometric series.

Theorem 3.3. *Let $b \neq 1$ and let n be a positive integer. Then*

$$1 + b + b^2 + \cdots + b^n = \frac{1 - b^{n+1}}{1 - b}.$$

Proof. Let

$$x = 1 + b + b^2 + \cdots + b^n.$$

Then

$$bx = b + b^2 + \cdots + b^n + b^{n+1},$$

so

$$x - bx = 1 - b^{n+1}.$$

Thus

$$x = \frac{1 - b^{n+1}}{1 - b},$$

as claimed. □

Note that if $|b| < 1$, then

$$\lim_{n\to\infty} b^n = 0\,.$$

Thus

$$\sum_{i=0}^{\infty} b^i = \lim_{n\to\infty} \sum_{i=0}^{n} b^i = \frac{1-0}{1-b} = \frac{1}{1-b}\,,$$

Example 3.4. Prof X wants to save for his retirement in 20 years. From the time of his retirement onwards he wants to withdraw $£1000$ every month at the beginning of each month for 30 years. What amount of money does he have to save every month at the beginning of each month for the next 20 years to fund his retirement if the nominal interest rate is 6% compounded monthly?

Solution. Let A denote the as yet unknown amount in pounds he deposits every month. We will first calculate the present value PV_A of all his deposits. Write

$$\beta = \frac{1}{1+\frac{0.06}{12}} = \frac{1}{1.005}\,.$$

Then, since there are 240 months in 20 years,

$$PV_A = A + A\beta + A\beta^2 + \cdots + A\beta^{239} = A(1+\beta+\beta^2+\cdots+\beta^{239}) = A\frac{1-\beta^{240}}{1-\beta}\,.$$

Let $W = 1000$ denote his monthly withdrawals in pounds. Since his first withdrawal will take place in 240 months time, and since there are 360 months in 30 years, the present value PV_W of all his withdrawals is

$$PV_W = W\beta^{240} + W\beta^{241} + \cdots + W\beta^{240+359} = W\beta^{240}(1+\beta+\cdots+\beta^{359}) = W\beta^{240}\frac{1-\beta^{360}}{1-\beta}\,.$$

Clearly, the present value of his deposits must equal the present value of his withdrawals. Thus

$$PV_A = PV_W\,,$$

so

$$A\frac{1-\beta^{240}}{1-\beta} = W\beta^{240}\frac{1-\beta^{360}}{1-\beta}\,,$$

hence

$$A = W\beta^{240}\frac{1-\beta^{360}}{1-\beta^{240}} = 360.9\,,$$

that is, he has to deposit $£361$ every month to fund his retirement. □

Strictly speaking, for examples like the above one one should also take into account inflation: The buying power of $£1000$ obtained in 20 or 50 years time will be much smaller than now.

Let r_i denote the rate of inflation and let r_{off} denote the official interest rate. If you put P in the bank now, then after one time period you will have $(1+r_{off})P$. But the purchasing power of the amount $(1+r_{off})P$ one time period from now is the same as the amount $(1+r_{off})P(1+r_i)^{-1}$ now. So the 'real' interest rate — real in the sense that it reflects the increase of purchasing power — denoted r_a, also known as **inflation adjusted interest rate**, is given by

$$r_a = \frac{\frac{1+r_{off}}{1+r_i}P - P}{P} = \frac{1+r_{off}}{1+r_i} - 1 = \frac{r_{off}-r_i}{1+r_i} \approx r_{off} - r_i\,.$$

For example, if a bank offers you to borrow at interest rate 4.5% and the rate of inflation is 2.5%, then the inflation adjusted interest rate is $4.5\% - 2.5\% = 2.0\%$. If you have a lot of debt, then inflation is a good thing!

3.3 Variable interest rates

So far we have assumed that interest rates remain constant throughout our calculations. In this section we shall briefly consider the case where interest is compounded continuously and rates are allowed to vary with time. We shall write $r = r(t)$ and call $r(t)$ the **instantaneous interest rate**.

We are interested in the amount $D(t)$ that has accumulated in your bank account at time t, given that you deposited the amount $D(0)$ at time 0. It turns out that the as yet unknown function $D(t)$ satisfies a differential equation which we shall now derive. In order to do this, let h be a small time period. We assume that the interest rate $r(t)$ does not change too drastically between times t and $t + h$, so

$$D(t+h) \approx D(t)e^{r(t)h}\,.$$

Recalling that $e^x \approx 1 + x$ if x is small, the above becomes

$$D(t+h) \approx D(t)(1 + r(t)h)\,,$$

so

$$D(t+h) - D(t) \approx D(t)r(t)h\,,$$

thus

$$\frac{D(t+h) - D(t)}{h} \approx D(t)r(t)\,.$$

By definition

$$\lim_{h\to 0} \frac{D(t+h) - D(t)}{h} = D'(t).$$

Thus

$$D'(t) = D(t)r(t)\,,$$

which is the desired differential equation. This is a separable first order differential equation which can be solved as follows. Writing

$$\frac{D'(t)}{D(t)} = r(t)$$

and integrating both sides yields

$$\int_0^t \frac{D'(u)}{D(u)}\,du = \int_0^t r(u)\,du\,.$$

Thus

$$[\log D(u)]_{u=0}^{u=t} = \int_0^t r(u)\,du\,,$$

hence

$$\log D(t) = \log D(0) + \int_0^t r(u)\,du\,,$$

and so

$$D(t) = D(0)\exp\left(\int_0^t r(u)\,du\right), \tag{3.2}$$

which is the desired relation between the amount $D(t)$ in your bank at time t and the time-varying interest rate function $r(t)$. Note that if $r(t) = r_0$, that is, the interest rate is constant,

then $D(t) = D(0)e^{r_0 t}$. Thus, the general formula (3.2) reduces to the familiar formula for continuously compounded interest if the interest rate is constant, just as it should.

A simple consequence of the above formula is the following observation. The present value of an amount $D(t)$ received at time t is

$$D(0) = D(t) \exp\left(- \int_0^t r(u)\, du \right).$$

In the context of a time-varying interest rate $r(t)$, the concept of the **yield curve** $\bar{r}(t)$, defined by

$$\bar{r}(t) = \frac{1}{t} \int_0^t r(u)\, du\,,$$

turns out to be useful. Note that $\bar{r}(t)$ is just the average interest rate on the interval $(0, t)$. We have

$$D(t) = D(0) \exp\left(\int_0^t r(u)\, du \right) = D(0) \exp(\bar{r}(t)t)\,.$$

Example 3.5. Let the time-varying interest rate be given by

$$r(t) = r_2 - \frac{r_2 - r_1}{1+t}\,,$$

with $0 < r_1 < r_2 < 1$.

The yield curve is

$$\bar{r}(t) = \frac{1}{t} \int_0^t \left(r_2 - \frac{r_2 - r_1}{1+u} \right) du = \frac{1}{t}(r_2 t - (r_2 - r_1)\log(1+t)) = r_2 - \frac{r_2 - r_1}{t} \log(1+t)\,.$$

Note that

$$\lim_{t \to \infty} \bar{r}(t) = r_2\,.$$

Moreover, the present value of $D(t)$ pounds received at time t is

$$D(0) = D(t) \exp(-\bar{r}(t)t) = D(t) \exp\left((r_2 - r_1)\log(1+t) - r_2 t \right) = D(t)(1+t)^{r_2 - r_1} e^{-r_2 t}\,.$$