Financial Mathematics I

Lecture Notes
Universität Ulm

VERSION: February 6, 2004

THIS IS A PRELIMINARY VERSION
THERE WILL BE UPDATES DURING THE COURSE

Prof.Dr. Rüdiger Kiesel
Abteilung Finanzmathematik
Universität Ulm

email:kiesel@mathematik.uni-ulm.de

Short Description.

Times and Location: Lectures will be Monday 10-12; Tuesday 8-10 in He120.
First Lecture Tuesday, 14.10.2003

Content.
This course covers the fundamental principles and techniques of financial mathematics in discrete- and continuous-time models. The focus will be on probabilistic techniques which will be discussed in some detail. Specific topics are

- Classical Asset Pricing: Mean-Variance Analysis, CAPM, Arbitrage.

- Martingale-based stochastic market models: Fundamental Theorems of Asset Pricing.

- Contingent Claim Analysis: European, American and Exotic Options.

- Interest Rate Theory: Term Structure Models, Interest Rate Derivatives.

Pre-requisites. Probability Theory, Calculus, Linear Algebra

Literature.

- N.H.Bingham & R.Kiesel, Risk Neutral Valuation, Springer 1998.

- H.Föllmer & A.Schied: Stochastic Finance: An Introduction in Discrete Time, De Gruyter 2002.

- J.Hull: Options, Futures & Other Derivatives, 4th edition, Prentice Hall, 1999.

- R.Jarrow & S.Turnbull, Derivative Securities, 2nd edition, 2000.

Office Hours. Tuesday 10-11. He 230
course webpage: www.mathematik.uni-ulm.de/finmath
email: kiesel@mathematik.uni-ulm.de.

Contents

Chapter 1

Arbitrage Theory

1.1 Derivative Background

Definition 1.1.1. *A derivative security, or contingent claim, is a financial contract whose value at expiration date T (more briefly, expiry) is determined exactly by the price (or prices within a prespecified time-interval) of the underlying financial assets (or instruments) at time T (within the time interval $[0, T]$).*

This section provides the institutional background on derivative securities, the main groups of underlying assets, the markets where derivative securities are traded and the financial agents involved in these activities. As our focus is on (probabilistic) models and not institutional considerations we refer the reader to the references for excellent sources describing institutions such as Davis (1994), Edwards and Ma (1992) and Kolb (1991).

1.1.1 Derivative Instruments

Derivative securities can be grouped under three general headings: *Options, Forwards and Futures* and *Swaps*. During this text we will mainly deal with options although our pricing techniques may be readily applied to forwards, futures and swaps as well.

Options.

An option is a financial instrument giving one the *right but not the obligation* to make a specified transaction at (or by) a specified date at a specified price. *Call* options give one the right to buy. *Put* options give one the right to sell. *European* options give one the right to buy/sell on the specified date, the expiry date, on which the option expires or matures. *American* options give one the right to buy/sell at any time prior to or at expiry.

Over-the-counter (OTC) options were long ago negotiated by a broker between a buyer and a seller. In 1973 (the year of the Black-Scholes formula, perhaps the central result of the subject), the Chicago Board Options Exchange (CBOE) began trading in options on some stocks. Since then, the growth of options has been explosive. Options are now traded on all the major world exchanges, in enormous volumes. *Risk* magazine (12/97) estimated \$35 *trillion* as the gross figure for worldwide derivatives markets in 1996. By contrast, the Financial Times of 7 October 2002 (Special Report on Derivatives) gives the interest rate and currency derivatives volume as \$ 83 *trillion* - an indication of the rate of growth in recent years! The simplest call and put options are now so standard they are called *vanilla* options. Many kinds of options now exist, including so-called *exotic* options. Types include: *Asian* options, which depend on the *average* price over a period, *lookback* options, which depend on the *maximum* or *minimum* price over a period and *barrier* options, which depend on some price level being attained or not.

Terminology. The asset to which the option refers is called the *underlying asset* or the *underlying.* The price at which the transaction to buy/sell the underlying, on/by the expiry date

(if exercised), is made, is called the *exercise price* or *strike price*. We shall usually use K for the strike price, time $t = 0$ for the initial time (when the contract between the buyer and the seller of the option is struck), time $t = T$ for the expiry or final time.

Consider, say, a European call option, with strike price K; write $S(t)$ for the value (or price) of the underlying at time t. If $S(t) > K$, the option is *in the money*, if $S(t) = K$, the option is said to be *at the money* and if $S(t) < K$, the option is *out of the money*.

The payoff from the option, which is

$$S(T) - K \text{ if } S(T) > K \quad \text{and} \quad 0 \text{ otherwise}$$

(more briefly written as $(S(T) - K)^+$). Taking into account the initial payment of an investor one obtains the profit diagram below.
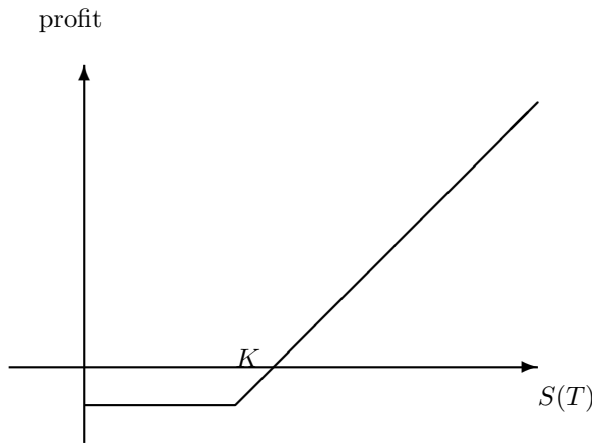
Figure 1.1: Profit diagram for a European call

Forwards

A *forward contract* is an agreement to buy or sell an asset S at a certain future date T for a certain price K. The agent who agrees to buy the underlying asset is said to have a *long* position, the other agent assumes a *short* position. The settlement date is called *delivery date* and the specified price is referred to as *delivery price*. The *forward price* $f(t, T)$ is the delivery price which would make the contract have zero value at time t. At the time the contract is set up, $t = 0$, the forward price therefore equals the delivery price, hence $f(0, T) = K$. The forward prices $f(t, T)$ need not (and will not) necessarily be equal to the delivery price K during the life-time of the contract.

The payoff from a long position in a forward contract on one unit of an asset with price $S(T)$ at the maturity of the contract is

$$S(T) - K.$$

Compared with a call option with the same maturity and strike price K we see that the investor now faces a downside risk, too. He has the obligation to buy the asset for price K.

Swaps

A *swap* is an agreement whereby two parties undertake to exchange, at known dates in the future, various financial assets (or cash flows) according to a prearranged formula that depends on the

value of one or more underlying assets. Examples are currency swaps (exchange currencies) and interest-rate swaps (exchange of fixed for floating set of interest payments).

1.1.2 Underlying securities

Stocks.

The basis of modern economic life – or of the capitalist system - is the *limited liability company* (UK: & Co. Ltd, now plc - public limited company), the corporation (US: Inc.), 'die Aktiengesellschaft' (Germany: AG). Such companies are owned by their shareholders; the shares

- provide partial ownership of the company, pro rata with investment,

- have value, reflecting both the value of the company's (real) assets and the earning power of the company's dividends.

With publicly quoted companies, shares are quoted and traded on the Stock Exchange. Stock is the generic term for assets held in the form of shares.

Interest Rates.

The value of some financial assets depends solely on the level of interest rates (or yields), e.g. Treasury (T-) notes, T-bills, T-bonds, municipal and corporate bonds. These are fixed-income securities by which national, state and local governments and large companies partially finance their economic activity. Fixed-income securities require the payment of interest in the form of a fixed amount of money at predetermined points in time, as well as repayment of the principal at maturity of the security. Interest rates themselves are notional assets, which cannot be delivered. Hedging exposure to interest rates is more complicated than hedging exposure to the price movements of a certain stock. A whole term structure is necessary for a full description of the level of interest rates, and for hedging purposes one must clarify the nature of the exposure carefully. We will discuss the subject of modelling the term structure of interest rates in Chapter 8.

Currencies.

A currency is the denomination of the national units of payment (money) and as such is a financial asset. The end of fixed exchange rates and the adoption of floating exchange rates resulted in a sharp increase in exchange rate volatility. International trade, and economic activity involving it, such as most manufacturing industry, involves dealing with more than one currency. A company may wish to hedge adverse movements of foreign currencies and in doing so use derivative instruments (see for example the exposure of the hedging problems British Steel faced as a result of the sharp increase in the pound sterling in 96/97 Rennocks (1997)).

Indexes.

An index tracks the value of a (hypothetical) basket of stocks (FT-SE100, S&P-500, DAX), bonds (REX), and so on. Again, these are not assets themselves. Derivative instruments on indexes may be used for hedging if no derivative instruments on a particular asset (a stock, a bond, a commodity) in question are available and if the correlation in movement between the index and the asset is significant. Furthermore, institutional funds (such as pension funds, mutual funds etc.), which manage large diversified stock portfolios, try to mimic particular stock indexes and use derivatives on stock indexes as a portfolio management tool. On the other hand, a speculator may wish to bet on a certain overall development in a market without exposing him/herself to a particular asset.

A new kind of index was generated with the Index of Catastrophe Losses (CAT-Index) by the Chicago Board of Trade (CBOT) lately. The growing number of huge natural disasters (such as hurricane Andrew 1992, the Kobe earthquake 1995) has led the insurance industry to try to find

new ways of increasing its capacity to carry risks. The CBOT tried to capitalise on this problem by launching a market in insurance derivatives. Currently investors are offered options on the CAT-Index, thereby taking in effect the position of traditional reinsurance.

Derivatives are themselves assets – they are traded, have value etc. - and so can be used as underlying assets for new contingent claims: options on futures, options on baskets of options, etc. These developments give rise to so-called *exotic options*, demanding a sophisticated mathematical machinery to handle them.

1.1.3 Markets

Financial derivatives are basically traded in two ways: on organized exchanges and over-the-counter (OTC). Organised exchanges are subject to regulatory rules, require a certain degree of standardisation of the traded instruments (strike price, maturity dates, size of contract etc.) and have a physical location at which trade takes place. Examples are the Chicago Board Options Exchange (CBOE), which coincidentally opened in April 1973, the same year the seminal contributions on option prices by Black and Scholes Black and Scholes (1973) and Merton Merton (1973) were published, the London International Financial Futures Exchange (LIFFE) and the Deutsche Terminbörse (DTB).

OTC trading takes place via computers and phones between various commercial and investment banks (leading players include institutions such as Bankers Trust, Goldman Sachs – where Fischer Black worked, Citibank, Chase Manhattan and Deutsche Bank).

Due to the growing sophistication of investors boosting demand for increasingly complicated, made-to-measure products, the OTC market volume is currently (as of 1998) growing at a much faster pace than trade on most exchanges.

1.1.4 Types of Traders

We can classify the traders of derivative securities in three different classes:

Hedgers.

Successful companies concentrate on economic activities in which they do best. They use the market to insure themselves against adverse movements of prices, currencies, interest rates etc. Hedging is an attempt to reduce exposure to risk a company already faces. *Shorter Oxford English Dictionary* (OED): Hedge: '*trans.* To cover oneself against loss on (a bet etc.) by betting, etc., on the other side. Also *fig.* 1672.'

Speculators.

Speculators want to take a position in the market – they take the opposite position to hedgers. Indeed, speculation is needed to make hedging possible, in that a hedger, wishing to lay off risk, cannot do so unless someone is willing to take it on.

In speculation, available funds are invested opportunistically in the hope of making a profit: the underlying itself is irrelevant to the investor (speculator), who is only interested in the potential for possible profit that trade involving it may present. Hedging, by contrast, is typically engaged in by companies who have to deal habitually in intrinsically risky assets such as foreign exchange next year, commodities next year, etc. They may prefer to forgo the chance to make exceptional windfall profits when future uncertainty works to their advantage by protecting themselves against exceptional loss. This would serve to protect their economic base (trade in commodities, or manufacture of products using these as raw materials), and also enable them to focus their effort in their chosen area of trade or manufacture. For speculators, on the other hand, it is the market (forex, commodities or whatever) itself which is their main forum of economic activity.

Arbitrageurs.

Arbitrageurs try to lock in riskless profit by simultaneously entering into transactions in two or more markets. The very existence of arbitrageurs means that there can only be very small arbitrage opportunities in the prices quoted in most financial markets. The underlying concept of this book is the absence of arbitrage opportunities (cf. §1.2).

1.1.5   Modelling Assumptions

Contingent Claim Pricing.

The fundamental problem in the mathematics of financial derivatives is that of *pricing*. The modern theory began in 1973 with the seminal *Black-Scholes* theory of option pricing, Black and Scholes (1973), and Merton's extensions of this theory, Merton (1973).

To expose the relevant features, we start by discussing contingent claim pricing in the simplest (idealised) case and impose the following set of assumptions on the financial markets (We will relax these assumptions subsequently):

No market frictions	No transaction costs, no bid/ask spread, no taxes, no margin requirements, no restrictions on short sales
No default risk	Implying same interest for borrowing and lending
Competitive markets	Market participants act as price takers
Rational agents	Market participants prefer more to less
No arbitrage	

Table 1.1: General assumptions

All real markets involve frictions; this assumption is made purely for simplicity. We develop the theory of an ideal – frictionless – market so as to focus on the irreducible essentials of the theory and as a first-order approximation to reality. Understanding frictionless markets is also a necessary step to understand markets with frictions.

The risk of failure of a company – bankruptcy – is inescapably present in its economic activity: death is part of life, for companies as for individuals. Those risks also appear at the national level: quite apart from war, or economic collapse resulting from war, recent decades have seen default of interest payments of international debt, or the threat of it. We ignore default risk for simplicity while developing understanding of the principal aspects (for recent overviews on the subject we refer the reader to Jameson (1995), Madan (1998)).

We assume financial agents to be price takers, not price makers. This implies that even large amounts of trading in a security by one agent does not influence the security's price. Hence agents can buy or sell as much of any security as they wish without changing the security's price.

To assume that market participants prefer more to less is a very weak assumption on the preferences of market participants. Apart from this we will develop a preference-free theory.

The relaxation of these assumptions is subject to ongoing research and we will include comments on this in the text.

We want to mention the special character of the no-arbitrage assumption. If we developed a theoretical price of a financial derivative under our assumptions and this price did not coincide with the price observed, we would take this as an arbitrage opportunity in our model and go on to explore the consequences. This might lead to a relaxation of one of the other assumptions and a restart of the procedure again with no-arbitrage assumed. The no-arbitrage assumption thus has a special status that the others do not. It is the basis for the *arbitrage pricing technique* that we shall develop, and we discuss it in more detail below.

1.2 Arbitrage

We now turn in detail to the concept of arbitrage, which lies at the centre of the relative pricing theory. This approach works under very weak assumptions. We do not have to impose any assumptions on the tastes (preferences) and beliefs of market participants. The economic agents may be heterogeneous with respect to their preferences for consumption over time and with respect to their expectations about future states of the world. All we assume is that they prefer more to less, or more precisely, an increase in consumption without any costs will always be accepted.

The principle of arbitrage in its broadest sense is given by the following quotation from OED: '3 [Comm.]. The traffic in Bills of Exchange drawn on sundry places, and bought or sold in sight of the daily quotations of rates in the several markets. Also, the similar traffic in Stocks. 1881.'

Used in this broad sense, the term covers financial activity of many kinds, including trade in options, futures and foreign exchange. However, the term arbitrage is nowadays also used in a narrower and more technical sense. Financial markets involve both riskless (bank account) and risky (stocks, etc.) assets. To the investor, the only point of exposing oneself to risk is the opportunity, or possibility, of realising a greater profit than the riskless procedure of putting all one's money in the bank (the mathematics of which – compound interest – does not require a textbook treatment at this level). Generally speaking, the greater the risk, the greater the return required to make investment an attractive enough prospect to attract funds. Thus, for instance, a clearing bank lends to companies at higher rates than it pays to its account holders. The companies' trading activities involve risk; the bank tries to spread the risk over a range of different loans, and makes its money on the difference between high/risky and low/riskless interest rates.

The essence of the technical sense of arbitrage is that it should not be possible to guarantee a profit without exposure to risk. Were it possible to do so, arbitrageurs (we use the French spelling, as is customary) would do so, in unlimited quantity, using the market as a 'money-pump' to extract arbitrarily large quantities of riskless profit. This would, for instance, make it impossible for the market to be in equilibrium. We shall restrict ourselves to markets in equilibrium for simplicity – so we must restrict ourselves to markets without arbitrage opportunities.

The above makes it clear that a market with arbitrage opportunities would be a disorderly market – too disorderly to model. The remarkable thing is the converse. It turns out that the minimal requirement of absence of arbitrage opportunities is enough to allow one to build a model of a financial market which – while admittedly idealised – is realistic enough both to provide real insight and to handle the mathematics necessary to price standard contingent claims. We shall see that arbitrage arguments suffice to determine prices - the *arbitrage pricing technique*. For an accessible treatment rather different to ours, see e.g. Allingham (1991).

To explain the fundamental arguments of the arbitrage pricing technique we use the following:

Example.

Consider an investor who acts in a market in which only three financial assets are traded: (riskless) bonds B (bank account), stocks S and European Call options C with strike $K = 1$ on the stock. The investor may invest today, time $t = 0$, in all three assets, leave his investment until time $t = T$ and get his returns back then (we assume the option expires at $t = T$, also). We assume the current £ prices of the financial assets are given by

$$B(0) = 1, \ S(0) = 1, \ C(0) = 0.2$$

and that at $t = T$ there can be only two states of the world: an up-state with £ prices

$$B(T, u) = 1.25, \ S(T, u) = 1.75, \ \text{and therefore } C(T, u) = 0.75,$$

and a down-state with £ prices

$$B(T, d) = 1.25, \ S(T, d) = 0.75, \ \text{and therefore } C(T, d) = 0.$$

Financial asset	Number of	Total amount in £
Bond	10	10
Stock	10	10
Call	25	5

Table 1.2: Original portfolio

Now our investor has a starting capital of £25, and divides it as in Table 1.2 below (we call such a division a portfolio). Depending of the state of the world at time $t = T$ this portfolio will give the £ return shown in Table 1.3. Can the investor do better? Let us consider the restructured

State of the world	Bond	Stock	Call	Total
Up	12.5	17.5	18.75	48.75
Down	12.5	7.5	0	20.

Table 1.3: Return of original portfolio

portfolio of Table 1.4. This portfolio requires only an investment of £24.6. We compute its return

Financial asset	Number of	Total amount in £
Bond	11.8	11.8
Stock	7	7
Call	29	5.8

Table 1.4: Restructured portfolio

in the different possible future states (Table 1.5). We see that this portfolio generates the same time $t = T$ return while costing only £24.6 now, a saving of £0.4 against the first portfolio. So the investor should use the second portfolio and have a free lunch today!

In the above example the investor was able to restructure his portfolio, reducing the current (time $t = 0$) expenses without changing the return at the future date $t = T$ in both possible states of the world. So there is an arbitrage possibility in the above market situation, and the prices quoted are not arbitrage (or market) prices. If we regard (as we shall do) the prices of the bond and the stock (our underlying) as given, the option must be mispriced. We will develop in this book models of financial market (with different degrees of sophistication) which will allow us to find methods to avoid (or to spot) such pricing errors. For the time being, let us have a closer look at the differences between portfolio 1, consisting of 10 bonds, 10 stocks and 25 call options, in short $(10, 10, 25)$, and portfolio 2, of the form $(11.8, 7, 29)$. The difference (from the point of view of portfolio 1, say) is the following portfolio, D: $(-1.8, 3, -4)$. Sell short three stocks (see below), buy four options and put £1.8 in your bank account. The left-over is exactly the £0.4 of the example. But what is the effect of doing that? Let us consider the consequences in the possible states of the world. From Table 1.6 below, we see in both cases that the effects of the different positions of the portfolio offset themselves. But clearly the portfolio generates an income at $t = 0$ and is therefore itself an arbitrage opportunity.

If we only look at the position in bonds and stocks, we can say that this position covers us against possible price movements of the option, i.e. having £1.8 in your bank account and being three stocks short has the same time $t = T$ effects of having four call options outstanding against us. We say that the bond/stock position is a *hedge* against the position in options.

Let us emphasise that the above arguments were independent of the preferences and plans of

State of the world	Bond	Stock	Call	Total
Up	14.75	12.25	21.75	48.75
Down	14.75	5.25	0	20.

Table 1.5: Return of the restructured portfolio

world is in state **up**		world is in state **down**	
exercise option	3	option is worthless	0
buy 3 stocks at 1.75	−5.25	buy 3 stocks at 0.75	−2.25
sell bond	2.25	sell bond	2.25
Balance	0	**Balance**	0

Table 1.6: Difference portfolio

the investor. They were also independent of the interpretation of $t = T$: it could be a fixed time, maybe a year from now, but it could refer to the happening of a certain event, e.g. a stock hitting a certain level, exchange rates at a certain level, etc.

1.3 Arbitrage Relationships

We will in this section use arbitrage-based arguments (arbitrage pricing technique) to develop general bounds on the value of options. Such bounds, deduced from the underlying assumption that no arbitrage should be possible, allow one to test the plausibility of sophisticated financial market models.

In our analysis here we use stocks as the underlying.

1.3.1 Fundamental Determinants of Option Values

We consider the determinants of the option value in table 1.7 below. Since we restrict ourselves to non-dividend paying stocks we don't have to consider cash dividends as another natural determinant.

Current stock price	$S(t)$
Strike price	K
Stock volatility	σ
Time to expiry	$T - t$
Interest rates	r

Table 1.7: Determinants affecting option value

We now examine the effects of the single determinants on the option prices (all other factors remaining unchanged).

We saw that at expiry the only variables that mattered were the stock price $S(T)$ and strike price K: remember the payoffs $C = (S(T) - K)^+, P = (S(T) - K)^- (:= \max\{K - S(T), 0\})$. Looking at the payoffs, we see that an increase in the stock price will increase (decrease) the value

of a call (put) option (recall all other factors remain unchanged). The opposite happens if the strike price is increased: the price of a call (put) option will go down (up).

When we buy an option, we bet on a favourable outcome. The actual outcome is uncertain; its uncertainty is represented by a probability density; favourable outcomes are governed by the tails of the density (right or left tail for a call or a put option). An increase in volatility flattens out the density and thickens the tails, so increases the value of both call and put options. Of course, this argument again relies on the fact that we don't suffer from (with the increase of volatility more likely) more severe unfavourable outcomes – we have the right, but not the obligation, to exercise the option.

A heuristic statement of the effects of time to expiry or interest rates is not so easy to make. In the simplest of models (no dividends, interest rates remain fixed during the period under consideration), one might argue that the longer the time to expiry the more can happen to the price of a stock. Therefore a longer period increases the possibility of movements of the stock price and hence the value of a call (put) should be higher the more time remains before expiry. But only the owner of an American-type option can react immediately to favourable price movements, whereas the owner of a European option has to wait until expiry, and only the stock price then is relevant. Observe the contrast with volatility: an increase in volatility increases the likelihood of favourable outcomes at expiry, whereas the stock price movements before expiry may cancel out themselves. A longer time until expiry might also increase the possibility of adverse effects from which the stock price has to recover before expiry. We see that by using purely heuristic arguments we are not able to make precise statements. One can, however, show by explicit arbitrage arguments that an increase in time to expiry leads to an increase in the value of call options as well as put options. (We should point out that in case of a dividend-paying stock the statement is not true in general for European-type options.)

To qualify the effects of the interest rate we have to consider two aspects. An increase in the interest rate tends to increase the expected growth rate in an economy and hence the stock price tends to increase. On the other hand, the present value of any future cash flows decreases. These two effects both decrease the value of a put option, while the first effect increases the value of a call option. However, it can be shown that the first effect always dominates the second effect, so the value of a call option will increase with increasing interest rates.

The above heuristic statements, in particular the last, will be verified again in appropriate models of financial markets, see §4.5.4 and §6.2.3.

We summarise in table 1.8 the effect of an increase of one of the parameters on the value of options on non-dividend paying stocks while keeping all others fixed:

Parameter (increase)	Call	Put
Stock price	Positive	Negative
Strike price	Negative	Positive
Volatility	Positive	Positive
Interest rates	Positive	Negative
Time to expiry	Positive	Positive

Table 1.8: Effects of parameters

We would like to emphasise again that these results all assume that all other variables remain fixed, which of course is not true in practice. For example stock prices tend to fall (rise) when interest rates rise (fall) and the observable effect on option prices may well be different from the effects deduced under our assumptions.

Cox and Rubinstein (1985), p. 37–39, discuss other possible determining factors of option value, such as expected rate of growth of the stock price, additional properties of stock price movements, investors' attitudes toward risk, characteristics of other assets and institutional environment (tax rules, margin requirements, transaction costs, market structure). They show that

in many important circumstances the influence of these variables is marginal or even vanishing.

1.3.2 Arbitrage bounds

We now use the principle of no-arbitrage to obtain bounds for option prices. Such bounds, deduced from the underlying assumption that no arbitrage should be possible, allow one to test the plausibility of sophisticated financial market models. We focus on European options (puts and calls) with identical underlying (say a stock S), strike K and expiry date T. Furthermore we assume the existence of a risk-free bank account (bond) with constant interest rate r (continuously compounded) during the time interval $[0, T]$. We start with a fundamental relationship:

Proposition 1.3.1. *We have the following put-call parity between the prices of the underlying asset S and European call and put options on stocks that pay no dividends:*

$$S + P - C = Ke^{-r(T-t)}. \tag{1.1}$$

Proof. Consider a portfolio consisting of one stock, one put and a short position in one call (the holder of the portfolio has written the call); write $V(t)$ for the value of this portfolio. Then

$$V(t) = S(t) + P(t) - C(t)$$

for all $t \in [0, T]$. At expiry we have

$$V(T) = S(T) + (S(T) - K)^- - (S(T) - K)^+ = S(T) + K - S(T) = K.$$

This portfolio thus guarantees a payoff K at time T. Using the principle of no-arbitrage, the value of the portfolio must at any time t correspond to the value of a sure payoff K at T, that is $V(t) = Ke^{-r(T-t)}$. ∎

Having established (1.1), we concentrate on European calls in the following.

Proposition 1.3.2. *The following bounds hold for European call options:*

$$\max\left\{S(t) - e^{-r(T-t)}K, 0\right\} = \left(S(t) - e^{-r(T-t)}K\right)^+ \leq C(t) \leq S(t).$$

Proof. That $C \geq 0$ is obvious, otherwise 'buying' the call would give a riskless profit now and no obligation later.

Similarly the upper bound $C \leq S$ must hold, since violation would mean that the right to buy the stock has a higher value than owning the stock. This must be false, since a stock offers additional benefits.

Now from put-call parity (1.1) and the fact that $P \geq 0$ (use the same argument as above), we have

$$S(t) - Ke^{-r(T-t)} = C(t) - P(t) \leq C(t),$$

which proves the last assertion. ∎

It is immediately clear that an American call option can never be worth less than the corresponding European call option, for the American option has the added feature of being able to be exercised at any time until the maturity date. Hence (with the obvious notation): $C_A(t) \geq C_E(t)$. The striking result we are going to show (due to R.C. Merton in 1973, (Merton 1990), Theorem 8.2) is:

Proposition 1.3.3. *For a non-dividend paying stock we have*

$$C_A(t) = C_E(t). \tag{1.2}$$

Proof. Exercising the American call at time $t < T$ generates the cash-flow $S(t) - K$. From Proposition 1.3.2 we know that the value of the call must be greater or equal to $S(t) - Ke^{-r(T-t)}$, which is greater than $S(t) - K$. Hence selling the call would have realised a higher cash-flow and the early exercise of the call was suboptimal. ∎

Remark 1.3.1. *Qualitatively, there are two reasons why an American call should not be exercised early:*

(i) *Insurance. An investor who holds a call option instead of the underlying stock is 'insured against a fall in stock price below K, and if he exercises early, he loses this insurance.*

(ii) *Interest on the strike price. When the holder exercises the option, he buys the stock and pays the strike price, K. Early exercise at $t < T$ deprives the holder of the interest on K between times t and T: the later he pays out K, the better.*

We remark that an American put offers additional value compared to a European put.

1.4 Single-Period Market Models

1.4.1 A fundamental example

We consider a one-period model, i.e. we allow trading only at $t = 0$ and $t = T = 1$(say). Our aim is to value at $t = 0$ a European derivative on a stock S with maturity T.

First idea. Model S_T as a random variable on a probability space $(\Omega, \mathcal{F}, I\!P)$. The derivative is given by $H = f(S_T)$, i.e. it is a random variable (for a suitable function $f(.)$). We could then price the derivative using some discount factor β by using the expected value of the discounted future payoff:

$$H_0 = I\!E(\beta H). \tag{1.3}$$

Problem. How should we pick the probability measure $I\!P$? According to their preferences investors will have different opinions about the distribution of the price S_T.

Black-Scholes-Merton (Ross) approach. Use the no-arbitrage principle and construct a hedging portfolio using only known (and already priced) securities to duplicate the payoff H. We assume

1. Investors are non-satiable, i.e. they always prefer more to less.

2. Markets do not allow arbitrage , i.e. the possibility of risk-free profits.

From the no-arbitrage principle we see:

If it is possible to duplicate the payoff H of a derivative using a portfolio V of underlying (basic) securities, i.e. $H(\omega) = V(\omega)$, $\forall \omega$, the price of the portfolio at $t = 0$ must equal the price of the derivative at $t = 0$.

Let us assume there are two tradeable assets

- a riskfree bond (bank account) with $B(0) = 1$ and $B(T) = 1$, that is the interest rate $r = 0$ and the discount factor $\beta(t) = 1$. (In this context we use $\beta(t) = 1/B(t)$ as the discount factor).

- a risky stock S with $S(0) = 10$ and two possible values at $t = T$

$$S(T) = \begin{cases} 20 & \text{with probability } p \\ 7.5 & \text{with probability } 1 - p. \end{cases}$$

We call this setting a $(B, S)-$ market. The problem is to price a European call at $t = 0$ with strike $K = 15$ and maturity T, i.e. the random payoff $H = (S(T) - K)^+$. We can evaluate the call in every possible state at $t = T$ and see $H = 5$ (if $S(T) = 20$) with probability p and $H = 0$ (if $S(T) = 7.5$) with probability $1 - p$. This is illustrated in figure (1.4.1)

The key idea now is to try to find a portfolio combining bond and stock, which synthesizes the cash flow of the option. If such a portfolio exists, holding this portfolio today would be equivalent to holding the option – they would produce the same cash flow in the future. Therefore the price

today   one period

$$\begin{cases} S_1 & = & 20 \\ B_1 & = & 1 \\ H_1^+ & = & \max\{20-15,0\} \\ & = & 5 \end{cases}$$ up-state

$S_0 = 10$
$B_0 = 1$
$H_0 = ?$

$$\begin{cases} S_1 & = & 7.5 \\ B_1 & = & 1 \\ H_1^- & = & \max\{7.5-15,0\} \\ & = & 0. \end{cases}$$ down-state
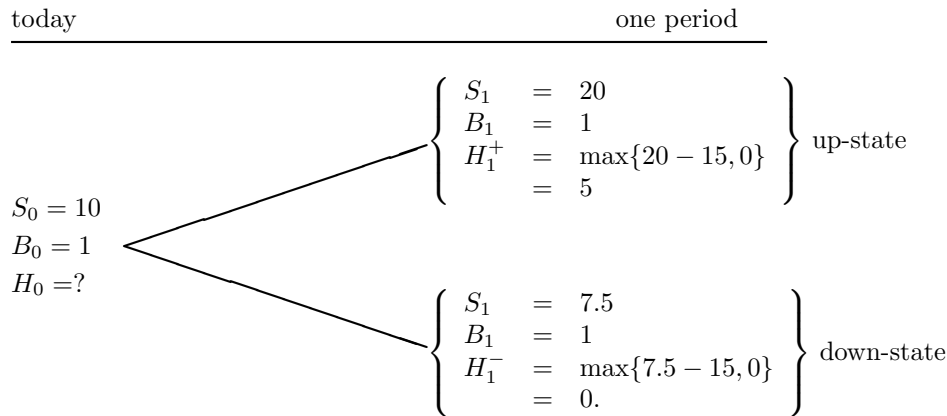
Figure 1.2: One-period example

of the option should be the same as the price of constructing the portfolio, otherwise investors could just restructure their holdings in the assets and obtain a riskfree profit today.

We briefly present the constructing of the portfolio $\boldsymbol{\theta} = (\theta_0, \theta_1)$, which in the current setting is just a simple exercise in linear algebra. If we buy θ_1 stocks and invest θ_0 £ in the bank account, then today's value of the portfolio is

$$V(0) = \theta_0 + \theta_1 \cdot S(0).$$

In state 1 the stock price is 20 £ and the value of the option 5 £, so

$$\theta_0 + \theta_1 \cdot 20 = 5.$$

In state 2 the stock price is 7.5 £ and the value of the option 0 £, so

$$\theta_0 + \theta_1 \cdot 7.5 = 0.$$

We solve this and get $\theta_0 = -3 \quad$ and $\quad \theta_1 = 0.4$.
So the value of our portfolio at time 0 in £ is

$$V(0) = -3B(0) + 0.4S(0) = -3 + 0.4 \times 10 = 1$$

$V(0)$ is called the no-arbitrage price. Every other price allows a riskless profit, since if the option is too cheap, buy it and finance yourself by selling short the above portfolio (i.e. sell the portfolio without possessing it and promise to deliver it at time $T = 1$ – this is riskfree because you own the option). If on the other hand the option is too dear, write it (i.e. sell it in the market) and cover yourself by setting up the above portfolio.

We see that the no-arbitrage price is independent of the individual preferences of the investor (given by certain probability assumptions about the future, i.e. a probability measure $I\!P$). But one can identify a special, so called risk-neutral, probability measure $I\!P^*$, such that

$$H_0 = I\!E^* \left(\beta H \right) = (p^* \cdot \beta(S_1 - K) + (1 - p^*) \cdot 0) = 1.$$

In the above example we get from $1 = p^* 5 + (1 - p^*)0$ that $p^* = 0.2$ This probability measure $I\!P^*$ is equivalent to $I\!P$, and the discounted stock price process, i.e. $\beta_t S_t$, $t = 0, 1$ follows a $I\!P^*$-martingale. In the above example this corresponds to $S(0) = p^* S(T)^{up} + (1 - p^*)S(T)^{down}$, that is $S(0) = I\!E^* \left(\beta S(T) \right)$.

We will show that the above generalizes. Indeed, we will find that *the no-arbitrage condition is equivalent to the existence of an equivalent martingale measure* (first fundamental theorem of

asset pricing) and that the property that we can *price assets using the expectation operator is equivalent to the uniqueness of the equivalent martingale measure.*

Let us consider the construction of hedging strategies from a different perspective. Consider a one-period (B, S)−market setting with discount factor $\beta = 1$. Assume we want to replicate a derivative H (that is a random variable on some probability space $(\Omega, \mathcal{F}, I\!\!P)$). For each hedging strategy $\boldsymbol{\theta} = (\theta_0, \theta_1)$ we have an initial value of the portfolio $V(0) = \theta_0 + \theta_1 S(0)$ and a time $t = T$ value of $V(T) = \theta_0 + \theta_1 S(T)$. We can write $V(T) = V(0) + (V(T) - V(0))$ with $G(T) = V(T) - V(0) = \theta_1(S(T) - S(0))$ the gains from trading. So the costs $C(0)$ of setting up this portfolio at time $t = 0$ are given by $C(0) = V(0)$, while maintaining (or achieving) a perfect hedge at $t = T$ requires an additional capital of $C(T) = H - V(T)$. Thus we have two possibilities for finding 'optimal' hedging strategies:

- *Mean-variance hedging.* Find θ_0 (or alternatively $V(0)$) and θ_1 such that

$$I\!\!E\left((H - V(T))^2\right) = I\!\!E\left((H - (V(0) + \theta_1(S(T) - S(0))))^2\right) \to \min$$

- *Risk-minimal hedging.* Minimize the cost from trading, i.e. an appropriate functional involving the costs $C(t)$, $t = 0, T$.

In our example mean-variance hedging corresponds to the standard linear regression problem, and so

$$\theta_1 = \frac{\mathbb{C}ov(H, (S(T) - S(0)))}{\mathbb{W}ar(S(T) - S(0))} \quad \text{and} \quad V_0 = I\!\!E(H) - \theta_1 I\!\!E(S(T) - S(0)).$$

We can also calculate the optimal value of the risk functional

$$R_{\min} = \mathbb{W}ar(H) - \theta_1^2 \mathbb{W}ar(S(T) - S(0)) = \mathbb{W}ar(H)(1 - \rho^2),$$

where ρ is the correlation coefficient of H and $S(T)$. Therefore we can't expect a perfect hedge in general. If however $|\rho| = 1$, i.e. H is a linear function of $S(T)$, a perfect hedge is possible. We call a market complete if a perfect hedge is possible for all contingent claims.

1.4.2   A single-period model

We proceed to formalise and extend the above example and present in detail a simple model of a financial market. Despite its simplicity it already has all the key features needed in the sequel (and the reader should not hesitate to come back here from more advanced chapters to see the bare concepts again).

We introduce in passing a little of the terminology and notation of Chapter 4; see also Harrison and Kreps (1979). We use some elementary vocabulary from probability theory, which is explained in detail in Chapter 2.

We consider a single period model, i.e. we have two time-indices, say $t = 0$, which is the current time (date), and $t = T$, which is the terminal date for all economic activities considered.

The financial market contains $d + 1$ traded financial assets, whose prices at time $t = 0$ are denoted by the vector $S(0) \in I\!\!R^{d+1}$,

$$S(0) = (S_0(0), S_1(0), \ldots, S_d(0))'$$

(where $'$ denotes the transpose of a vector or matrix). At time T, the owner of financial asset number i receives a random payment depending on the state of the world. We model this randomness by introducing a *finite* probability space $(\Omega, \mathcal{F}, I\!\!P)$, with a finite number $|\Omega| = N$ of points (each corresponding to a certain state of the world) $\omega_1, \ldots, \omega_j, \ldots, \omega_N$, each with positive probability: $I\!\!P(\{\omega\}) > 0$, which means that every state of the world is possible. $\mathcal{F}$ is the set of subsets of Ω (events that can happen in the world) on which $I\!\!P(.)$ is defined (we can quantify how probable these events are), here $\mathcal{F} = \mathcal{P}(\Omega)$ the set of all subsets of Ω. (In more complicated models it is

not possible to define a probability measure on all subsets of the state space Ω, see §2.1.) We can now write the random payment arising from financial asset i as

$$S_i(T) = (S_i(T, \omega_1), \ldots, S_i(T, \omega_j), \ldots, S_i(T, \omega_N))'.$$

At time $t = 0$ the agents can buy and sell financial assets. The portfolio position of an individual agent is given by a *trading strategy* φ, which is an $I\!\!R^{d+1}$ vector,

$$\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_d)'.$$

Here φ_i denotes the quantity of the ith asset bought at time $t = 0$, which may be negative as well as positive (recall we allow short positions).

The dynamics of our model using the trading strategy φ are as follows: at time $t = 0$ we invest the amount $S(0)'\varphi = \sum_{i=0}^{d} \varphi_i S_i(0)$ and at time $t = T$ we receive the random payment $S(T, \omega)'\varphi = \sum_{i=0}^{d} \varphi_i S_i(T, \omega)$ depending on the realised state ω of the world. Using the $(d+1) \times N$-matrix $\vec{S}$, whose columns are the vectors $S(T, \omega)$, we can write the possible payments more compactly as $\vec{S}'\varphi$.

What does an *arbitrage opportunity* mean in our model? As arbitrage is 'making something out of nothing'; an arbitrage strategy is a vector $\varphi \in I\!\!R^{d+1}$ such that $S(0)'\varphi = 0$, our net investment at time $t = 0$ is zero, and

$$S(T, \omega)'\varphi \geq 0, \ \forall \omega \in \Omega \ \text{ and there exists a } \ \omega \in \Omega \ \text{ such that } S(T, \omega)'\varphi > 0.$$

We can equivalently formulate this as: $S(0)'\varphi < 0$, we borrow money for consumption at time $t = 0$, and

$$S(T, \omega)'\varphi \geq 0, \ \forall \omega \in \Omega,$$

i.e we don't have to repay anything at $t = T$. Now this means we had a 'free lunch' at $t = 0$ at the market's expense.

We agreed that we should not have arbitrage opportunities in our model. The consequences of this assumption are surprisingly far-reaching.

So assume that there are no arbitrage opportunities. If we analyse the structure of our model above, we see that every statement can be formulated in terms of Euclidean geometry or linear algebra. For instance, absence of arbitrage means that the space

$$\Gamma = \left\{ \begin{pmatrix} x \\ y \end{pmatrix}, x \in I\!\!R, y \in I\!\!R^N : x = -S(0)'\varphi, y = \vec{S}'\varphi, \varphi \in I\!\!R^{d+1} \right\}$$

and the space

$$I\!\!R_+^{N+1} = \left\{ z \in I\!\!R^{N+1} : z_i \geq 0 \ \ \forall \ 0 \leq i \leq N \ \ \exists i \ \text{ such that } z_i > 0 \right\}$$

have no common points. A statement like that naturally points to the use of a separation theorem for convex subsets, the separating hyperplane theorem (see e.g. Rockafellar (1970) for an account of such results, or Appendix A). Using such a theorem we come to the following characterisation of no arbitrage.

Theorem 1.4.1. *There is no arbitrage if and only if there exists a vector*

$$\psi \in I\!\!R^N, \quad \psi_i > 0, \quad \forall \ 1 \leq i \leq N$$

such that

$$\vec{S}\psi = S(0). \tag{1.4}$$

Proof. The implication '$\Leftarrow$' follows straightforwardly: assume that $S(T, \omega)'\varphi \geq 0$, $\omega \in \Omega$ for a vector $\varphi \in I\!\!R^{d+1}$.Then

$$S(0)'\varphi = (\vec{S}\psi)'\varphi = \psi'\vec{S}'\varphi \geq 0,$$

since $\psi_i > 0$, $\forall 1 \leq i \leq N$. So no arbitrage opportunities exist.

To show the implication '⇒' we use a variant of the separating hyperplane theorem. Absence of arbitrage means the Γ and $I\!\!R_+^{N+1}$ have no common points. This means that $K \subset I\!\!R_+^{N+1}$ defined by

$$K = \left\{ z \in I\!\!R_+^{N+1} : \sum_{i=0}^{N} z_i = 1 \right\}$$

and Γ do not meet. But K is a compact and convex set, and by the separating hyperplane theorem (Appendix C), there is a vector $\lambda \in I\!\!R^{N+1}$ such that for all $z \in K$

$$\lambda' z > 0$$

but for all $(x, y)' \in \Gamma$

$$\lambda_0 x + \lambda_1 y_1 + \ldots + \lambda_N y_N = 0.$$

Now choosing $z_i = 1$ successively we see that $\lambda_i > 0$, $i = 0, \ldots N$, and hence by normalising we get $\psi = \lambda/\lambda_0$ with $\psi_0 = 1$. Now set $x = -S(0)'\varphi$ and $y = \vec{S}'\varphi$ and the claim follows. ∎

The vector ψ is called a *state-price vector*. We can think of ψ_j as the marginal cost of obtaining an additional unit of account in state ω_j. We can now reformulate the above statement to:

There is no arbitrage if and only if there exists a state-price vector.

Using a further normalisation, we can clarify the link to our probabilistic setting. Given a state-price vector $\psi = (\psi_1, \ldots, \psi_N)$, we set $\psi_0 = \psi_1 + \ldots + \psi_N$ and for any state ω_j write $q_j = \psi_j/\psi_0$. We can now view $(q_1, \ldots, q_N)$ as probabilities and define a new probability measure on Ω by $\mathbb{Q}(\{\omega_j\}) = q_j$, $j = 1, \ldots, N$. Using this probability measure, we see that for each asset i we have the relation

$$\frac{S_i(0)}{\psi_0} = \sum_{j=1}^{N} q_j S_i(T, \omega_j) = I\!\!E_{\mathbb{Q}}(S_i(T)).$$

Hence the normalized price of the financial security i is just its expected payoff under some specially chosen 'risk-neutral' probabilities. Observe that we didn't make any use of the specific probability measure $I\!\!P$ in our given probability space.

So far we have not specified anything about the denomination of prices. From a technical point of view we could choose any asset i as long as its price vector $(S_i(0), S_i(T, \omega_1), \ldots, S_i(T, \omega_N))'$ only contains positive entries, and express all other prices in units of this asset. We say that we use this asset as *numéraire*. Let us emphasise again that arbitrage opportunities do not depend on the chosen numéraire. It turns out that appropriate choice of the numéraire facilitates the probability-theoretic analysis in complex settings, and we will discuss the choice of the numéraire in detail later on.

For simplicity, let us assume that asset 0 is a riskless bond paying one unit in all states $\omega \in \Omega$ at time T. This means that $S_0(T, \omega) = 1$ in all states of the world $\omega \in \Omega$. By the above analysis we must have

$$\frac{S_0(0)}{\psi_0} = \sum_{j=1}^{N} q_j S_0(T, \omega_j) = \sum_{j=1}^{N} q_j 1 = 1,$$

and ψ_0 is the discount on riskless borrowing. Introducing an interest rate r, we must have $S_0(0) = \psi_0 = (1 + r)^{-T}$. We can now express the price of asset i at time $t = 0$ as

$$S_i(0) = \sum_{j=1}^{N} q_j \frac{S_i(T, \omega_j)}{(1 + r)^T} = I\!\!E_{\mathbb{Q}} \left(\frac{S_i(T)}{(1 + r)^T} \right).$$

We rewrite this as

$$\frac{S_i(T)}{(1 + r)^0} = I\!\!E_{\mathbb{Q}} \left(\frac{S_i(T)}{(1 + r)^T} \right).$$

In the language of probability theory we just have shown that the processes $S_i(t)/(1+r)^t$, $t = 0, T$ are Q-martingales. (Martingales are the probabilists' way of describing fair games: see Chapter 3.) It is important to notice that under the given probability measure $I\!\!P$ (which reflects an individual agent's belief or the markets' belief) the processes $S_i(t)/(1+r)^t$, $t = 0, T$ generally do not form $I\!\!P$-martingales.

We use this to shed light on the relationship of the probability measures $I\!\!P$ and Q. Since $Q(\{\omega\}) > 0$ for all $\omega \in \Omega$ the probability measures $I\!\!P$ and Q are equivalent and (see Chapters 2 and 3) because of the argument above we call Q an *equivalent martingale measure*. So we arrived at yet another characterisation of arbitrage:

There is no arbitrage if and only if there exists an equivalent martingale measure.

We also see that risk-neutral pricing corresponds to using the expectation operator with respect to an equivalent martingale measure. This concept lies at the heart of stochastic (mathematical) finance and will be the *golden thread* (or *roter Faden*) throughout this book.

We now know how the given prices of our $(d + 1)$ financial assets should be related in order to exclude arbitrage opportunities, but how should we price a newly introduced financial instrument? We can represent this financial instrument by its random payments $\delta(T) = (\delta(T, \omega_1), \dots, \delta(T, \omega_j), \dots, \delta(T, \omega_N))'$ (observe that $\delta(T)$ is a vector in $I\!\!R^N$) at time $t = T$ and ask for its price $\delta(0)$ at time $t = 0$. The natural idea is to use an equivalent probability measure Q and set

$$\delta(0) = I\!\!E_Q(\delta(T)/(1+r)^T)$$

(recall that all time $t = 0$ and time $t = T$ prices are related in this way). Unfortunately, as we don't have a unique martingale measure in general, we cannot guarantee the uniqueness of the $t = 0$ price. Put another way, we know every equivalent martingale measure leads to a reasonable relative price for our newly created financial instrument, but which measure should one choose?

The easiest way out would be if there were only one equivalent martingale measure at our disposal – and surprisingly enough the classical economic pricing theory puts us exactly in this situation! Given a set of financial assets on a market the underlying question is whether we are able to price any new financial asset which might be introduced in the market, or equivalently whether we can replicate the cash-flow of the new asset by means of a portfolio of our original assets. If this is the case and we can replicate every new asset, the market is called *complete*.

In our financial market situation the question can be restated mathematically in terms of Euclidean geometry: do the vectors $S_i(T)$ span the whole $I\!\!R^N$? This leads to:

Theorem 1.4.2. *Suppose there are no arbitrage opportunities. Then the model is complete if and only if the matrix equation*

$$\vec{S}'\varphi = \delta$$

has a solution $\varphi \in I\!\!R^{d+1}$ for any vector $\delta \in I\!\!R^N$.

Linear algebra immediately tells us that the above theorem means that the number of independent vectors in $\vec{S}'$ must equal the number of states in Ω. In an informal way we can say that if the financial market model contains 2 (N) states of the world at time T it allows for 1 ($N - 1$) sources of randomness (if there is only one state we know the outcome). Likewise we can view the numéraire asset as risk-free and all other assets as risky. We can now restate the above characterisation of completeness in an informal (but intuitive) way as:

A financial market model is complete if it contains at least as many independent risky assets as sources of randomness.

The question of completeness can be expressed equivalently in probabilistic language (to be introduced in Chapter 3), as a question of representability of the relevant random variables or whether the σ-algebra they generate is the full σ-algebra.

If a financial market model is complete, traditional economic theory shows that there exists a unique system of prices. If there exists only one system of prices, and every equivalent martingale measure gives rise to a price system, we can only have a unique equivalent martingale measure. (We will come back to this important question in Chapters 4 and 6).

The (arbitrage-free) market is complete if and only if there exists a unique equivalent martingale measure.

Example (continued).

We formalise the above example of a binary single-period model. We have $d + 1 = 2$ assets and $|\Omega| = 2$ states of the world $\Omega = \{\omega_1, \omega_2\}$. Keeping the interest rate $r = 0$ we obtain the following vectors (and matrices):

$$S(0) = \begin{bmatrix} S_0(0) \\ S_1(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 150 \end{bmatrix}, \ S_0(T) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ S_1(T) = \begin{bmatrix} 180 \\ 90 \end{bmatrix}, \ \vec{S} = \begin{bmatrix} 1 & 1 \\ 180 & 90 \end{bmatrix}.$$

We try to solve (1.4) for state prices, i.e. we try to find a vector $\psi = (\psi_1, \psi_2)'$, $\psi_i > 0$, $i = 1, 2$ such that

$$\begin{bmatrix} 1 \\ 150 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 180 & 90 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}.$$

Now this has a solution

$$\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix},$$

hence showing that there are no arbitrage opportunities in our market model. Furthermore, since $\psi_1 + \psi_2 = 1$ we see that we already have computed risk-neutral probabilities, and so we have found an equivalent martingale measure $\boldsymbol{Q}$ with

$$\boldsymbol{Q}(\omega_1) = \frac{2}{3}, \ \boldsymbol{Q}(\omega_2) = \frac{1}{3}.$$

We now want to find out if the market is complete (or equivalently if there is a unique equivalent martingale measure). For that we introduce a new financial asset δ with random payments $\delta(T) = (\delta(T, \omega_1), \delta(T, \omega_2))'$. For the market to be complete, each such $\delta(T)$ must be in the linear span of $S_0(T)$ and $S_1(T)$. Now since $S_0(T)$ and $S_1(T)$ are linearly independent, their linear span is the whole $\mathbb{R}^2 (= \mathbb{R}^{|\Omega|})$ and $\delta(T)$ is indeed in the linear span. Hence we can find a replicating portfolio by solving

$$\begin{bmatrix} \delta(T, \omega_1) \\ \delta(T, \omega_2) \end{bmatrix} = \begin{bmatrix} 1 & 180 \\ 1 & 90 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \end{bmatrix}.$$

Let us consider the example of a European option above. There $\delta(T, \omega_1) = 30, \delta(T, \omega_2) = 0$ and the above becomes

$$\begin{bmatrix} 30 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 180 \\ 1 & 90 \end{bmatrix} \begin{bmatrix} \varphi_0 \\ \varphi_1 \end{bmatrix},$$

with solution $\varphi_0 = -30$ and $\varphi_1 = \frac{1}{3}$, telling us to borrow 30 units and buy $\frac{1}{3}$ stocks, which is exactly what we did to set up our portfolio above. Of course an alternative way of showing market completeness is to recognise that (1.4) above admits only one solution for risk-neutral probabilities, showing the uniqueness of the martingale measure.

Example. Change of numéraire.

We choose a situation similar to the above example, i.e. we have $d + 1 = 2$ assets and $|\Omega| = 2$ states of the world $\Omega = \{\omega_1, \omega_2\}$. But now we assume two risky assets (and no bond) as the financial instruments at our disposal. The price vectors are given by

$$S(0) = \begin{bmatrix} S_0(0) \\ S_1(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ S_0(T) = \begin{bmatrix} 3/4 \\ 5/4 \end{bmatrix}, \ S_1(T) = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}, \ \vec{S} = \begin{bmatrix} 3/4 & 5/4 \\ 1/2 & 2 \end{bmatrix}.$$

We solve (1.4) and get state prices

$$\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} 6/7 \\ 2/7 \end{bmatrix},$$

showing that there are no arbitrage opportunities in our market model. So we find an equivalent martingale measure $\boldsymbol{Q}$ with

$$\boldsymbol{Q}(\omega_1) = \frac{3}{4}, \ \boldsymbol{Q}(\omega_2) = \frac{1}{4}.$$

Since we don't have a risk-free asset in our model, this normalisation (this numéraire) is very artificial, and we shall use one of the modelled assets as a numéraire, say S_0. Under this normalisation, we have the following asset prices (in terms of $S_0(t, \omega)$!):

$$\tilde{S}(0) = \left[\begin{array}{c} \tilde{S}_0(0) \\ \tilde{S}_1(0) \end{array} \right] = \left[\begin{array}{c} S_0(0)/S_0(0) \\ S_1(0)/S_0(0) \end{array} \right] = \left[\begin{array}{c} 1 \\ 1 \end{array} \right],$$

$$\tilde{S}_0(T) = \left[\begin{array}{c} S_0(T, \omega_1)/S_0(T, \omega_1) \\ S_0(T, \omega_2)/S_0(T, \omega_2) \end{array} \right] = \left[\begin{array}{c} 1 \\ 1 \end{array} \right],$$

$$\tilde{S}_1(T) = \left[\begin{array}{c} S_1(T, \omega_1)/S_0(T, \omega_1) \\ S_1(T, \omega_2)/S_0(T, \omega_2) \end{array} \right] = \left[\begin{array}{c} 2/3 \\ 8/5 \end{array} \right].$$

Since the model is arbitrage-free (recall that a change of numéraire doesn't affect the no-arbitrage property), we are able to find risk-neutral probabilities $\tilde{q}_1 = \frac{9}{14}$ and $\tilde{q}_2 = \frac{5}{14}$.

We now compute the prices for a call option to exchange S_0 for S_1. Define $Z(T) = \max\{S_1(T) - S_0(T), 0\}$ and the cash flow is given by

$$\left[\begin{array}{c} Z(T, \omega_1) \\ Z(T, \omega_2) \end{array} \right] = \left[\begin{array}{c} 0 \\ 3/4 \end{array} \right].$$

There are no difficulties in finding the hedge portfolio $\varphi_0 = -\frac{3}{7}$ and $\varphi_1 = \frac{9}{14}$ and pricing the option as $Z_0 = \frac{3}{14}$.

We want to point out the following observation. Using S_0 as numéraire we naturally write

$$\frac{Z(T, \omega)}{S_0(T, \omega)} = \max\left\{ \frac{S_1(T, \omega)}{S_0(T, \omega)} - 1, 0 \right\}$$

and see that this seemingly complicated option is (by use of the appropriate numéraire) equivalent to

$$\tilde{Z}(T, \omega) = \max\left\{ \tilde{S}_1(T, \omega) - 1, 0 \right\},$$

a European call on the asset $\tilde{S}_1$.

1.4.3   A few financial-economic considerations

The underlying principle for modelling economic behaviour of investors (or economic agents in general) is the maximisation of **expected utility**, that is one assumes that agents have a utility function $U(.)$ and base economic decisions on expected utility considerations. For instance, assuming a one-period model, an economic agent might have a utility function over current($t = 0$) and future ($t = T$) values of consumption

$$U(c_0, c_T) = u(c_0) + I\!\!E(\beta u(c_T)), \tag{1.5}$$

where c_t is consumption at time t.

$u(.)$ is a standard utility function expressing

- non-satiation - investors prefer more to less; u is increasing;

- risk aversion - investors reject an actuarially fair gamble; u is concave;

- and (maybe) decreasing absolute risk aversion and constant relative risk aversion.

Typical examples are power utility $u(x) = (x^\gamma - 1)/\gamma$, log utility $u(x) = \log(x)$ or quadratic utility $u(x) = x^2 + dx$ (for which only the first two properties are true.)

Assume such an investor is offered at $t = 0$ at a price p a random payoff X at $t = T$. How much will she buy? Denote with ξ the amount of the asset she chooses to buy and with e_τ, $\tau = 0, T$ her original consumption. Thus, her problem is

$$\max_\xi \left[u(c_0) + I\!\!E[\beta u(c_T)] \right]$$

such that

$$c_0 = e_0 - p\xi \quad \text{and } c_T = e_T + X\xi.$$

Substituting the constraints into the objective function and differentiating with respect to ξ, we get the first-order condition

$$pu'(c_0) = I\!E\left[\beta u'(c_T)X\right] \quad \Leftrightarrow \quad p = I\!E\left[\beta \frac{u'(c_T)}{u'(c_0)}X\right].$$

The investor buys or sells more of the asset until this first-order condition is satisfied.

If we use the stochastic discount factor

$$m = \beta \frac{u'(c_T)}{u'(c_0)},$$

we obtain the central equation

$$p = I\!E(mX). \tag{1.6}$$

We can use (under regularity conditions) the random variable m to perform a change of measure, i.e. define a probability measure $I\!P^*$ using

$$I\!P^*(A) = I\!E^*(\mathbf{1}_A) = I\!E(m\mathbf{1}_A).$$

We write (1.6) under measure $I\!P^*$ and get

$$p = I\!E^*(X)$$

Returning to the initial pricing equation, we see that under $I\!P^*$ the investor has the utility function $u(x) = x$. Such an investor is called risk-neutral, and consequently one often calls the corresponding measure a risk-neutral measure. An excellent discussion of these issues (and further much deeper results) are given in Cochrane (2001).

Chapter 2

Financial Market Theory

2.1 Choice under Uncertainty

In a complete financial market model prices of derivative securities can be obtained by arbitrage arguments. In incomplete market situations, these financial instruments carry an intrinsic risk which cannot be hedged away. Thus in order to price these instruments further assumptions on the investors, especially regarding their preferences towards the risks involved, are needed.

2.1.1 Preferences and the Expected Utility Theorem

Let $\mathcal{X}$ be some non-empty set. An element $x \in \mathcal{X}$ will be interpreted as a possible choice of an economic agent. If presented with two choices $x, y \in \mathcal{X}$ the agent might prefer one over the other. This can be formalized

Definition 2.1.1. *A binary relation $\succeq$ defined on $\mathcal{X} \times \mathcal{X}$ is called a preference relation, if it is*

- *transitive: $x \succeq y, y \succeq z \Rightarrow x \succeq z$.*

- *complete: for all $x, y \in \mathcal{X}$ either $x \succeq y$ or $y \succeq x$.*

If $x \succeq y$ and $y \succeq x$ we write $x \sim y$ (indifference relation). x is said to be strictly preferred to y, denoted by $x \succ y$, if $x \succeq y$ and $y \not\succeq x$.

Definition 2.1.2. *A numerical representation of a preference order $\succeq$ is a function $U : \mathcal{X} \to \mathbb{R}$ such that*

$$y \succeq x \Leftrightarrow U(y) \geq U(x). \tag{2.1}$$

In order to characterize existence of a numerical representation we need

Definition 2.1.3. *Let $\succeq$ be a preference relation on $\mathcal{X}$. A subset $\mathcal{Z}$ of $\mathcal{X}$ is called order dense if for any pair $x, y \in \mathcal{X}$ such that $x \succ y$ there exists some $z \in \mathcal{Z}$ such that $x \succeq z \succeq y$.*

Theorem 2.1.1. *For the existence of a numerical representation of a preference relation $\succeq$ it is necessary and sufficient that $\mathcal{X}$ contains a countable order dense subset $\mathcal{Z}$. In particular, any preference relation admits a numerical representation if $\mathcal{X}$ is countable.*

Suppose that each possible choice of our economic agent corresponds to a probability distribution on a sample space $(\Omega, \mathcal{F})$. Thus the set $\mathcal{X}$ can be identified with a subset $\mathcal{M}$ of the set $\mathcal{M}_1(\Omega, \mathcal{F})$ of all probability distributions on $(\Omega, \mathcal{F})$. In the context of the theory of choice the elements of $\mathcal{M}$ are sometimes called lotteries. We assume in the sequel that $\mathcal{M}$ is convex. The aim in the following is to characterize the preference orders $\succeq$ that allow a numerical representation of the form

$$\mu \succeq \nu \quad \Leftrightarrow \quad \int\limits_{\Omega} u(\omega)\mu(d\omega) \geq \int\limits_{\Omega} u(\omega)\nu(d\omega) \tag{2.2}$$

Definition 2.1.4. *A numerical representation U of a preference order $\succeq$ on $\mathcal{M}$ is called a von Neumann-Morgenstern representation if it is of form (2.2).*

Any von Neumann-Morgenstern representation of U is affine on $\mathcal{M}$ in the sense that

$$U(\alpha\mu + (1 - \alpha)\nu) = \alpha U(\mu) + (1 - \alpha)U(\nu)$$

for all $\mu, \nu \in \mathcal{M}$ and $\alpha \in [0, 1]$.

Affinity implies the tow following properties or axioms

- Independence (substitution) (I):
  Let $\mu, \nu, \lambda \in \mathcal{M}$ and $\alpha \in (0, 1]$ then

$$\mu \succ \nu \quad \Rightarrow \alpha\mu + (1 - \alpha)\lambda \succ \alpha\nu + (1 - \alpha)\lambda$$

- Archimedean (continuity) (A):
  If $\mu \succ \lambda \succ \nu \in \mathcal{M}$ then there exist $\alpha, \beta \in (0, 1)$ such that

$$\alpha\mu + (1 - \alpha)\nu \succ \lambda \succ \beta\mu + (1 - \beta)\nu$$

Theorem 2.1.2. *Suppose the preference relation $\succeq$ on $\mathcal{M}$ satisfies the axioms (A) and (I). Then there exists an affine numerical representation U of $\succeq$. Moreover, U is unique up to positive affine transformations, i.e. any other affine numerical representation $\tilde{U}$ with these properties is of the form $\tilde{U} = aU + b$ for some $a > 0$ and $b \in \mathbb{R}$.*

In case of a discrete (finite) probability distribution existence of an affine numerical representation is equivalent to existence of a von Neumann-Morgenstern representation.

In the general case one needs to introduce a further axiom, the so-called sure thing principle: For $\mu, \nu \in \mathcal{M}$ and $A \in \mathcal{F}$ such that $\mu(A) = 1$:

$$\delta_x \succ \nu \quad \text{for all} \quad x \in A \quad \Rightarrow \quad \mu \succ \nu$$

and

$$\nu \succ \delta_x \quad \text{for all} \quad x \in A \quad \Rightarrow \quad \nu \succ \mu.$$

From now on we will work within the framework of the expected utility representation.

2.1.2 Risk Aversion

We focus now on individual financial assets under the assumption that their payoff distributions at a fixed time are known. We can view these asset distributions as probability distributions on some interval $S \subseteq \mathbb{R}$. Thus we take $\mathcal{M}$ as a fixed set of Borel probability measures on S. We also assume that $\mathcal{M}$ is convex and contains all point masses δ_x for $x \in S$. Also, we assume that each $\mu \in \mathcal{M}$ has a well defined expectation

$$m(\mu) = \int x\mu(dx) \in \mathbb{R}.$$

For assets with random payoff μ resp. insurance contracts with random damage $m(\mu)$ is often called fair price resp. fair premium. However, actual prices resp. premia will typically be different due to risk premia, which can be explained within our conceptual framework. We assume in the sequel that preference relations have a von Neumann-Morgenstern representation.

Definition 2.1.5. *A preference relation $\succ$ on $\mathcal{M}$ is called monotone if*

$$x > y \quad implies \quad \delta_x \succ \delta_y.$$

The preference relation is called risk averse if for $\mu \in \mathcal{M}$

$$\delta_{m(\mu)} \succ \mu \quad unless \quad \mu = \delta_{m(\mu)}. \tag{2.3}$$

Remark 2.1.1. *An economic agent is called risk-averse if his preference relation is risk averse. A risk-averse economic agent is unwilling to accept or indifferent to every actuarially fair gamble. An economic agent is strictly risk averse if he is unwilling to accept every actuarially fair gamble.*

Proposition 2.1.1. *A preference relation $\succ$ is*

(i) *monotone, iff u is strictly increasing.*

(ii) *risk averse, iff u is concave.*

Proof.(i) Monotonicity is equivalent to

$$u(x) = \int u(s)\delta_x(ds) = U(\delta_x) > U(\delta_y) = u(y)$$

for $x > y$.
(ii) For $\mu = \alpha\delta_x + (1-\alpha)\delta_y$ we have

$$m(\mu) = \int s(\alpha\delta_x + (1-\alpha)\delta_y)(ds) = \alpha x + (1-\alpha)y.$$

So if $\succ$ is risk-averse, then

$$\delta_{\alpha x + (1-\alpha)y} \succ \alpha\delta_x + (1-\alpha)\delta_y$$

holds for all distinct $x, y \in S$ and $\alpha \in (0,1)$. Hence

$$u(\alpha x + (1-\alpha)y) > \alpha u(x) + (1-\alpha)u(y),$$

so u is strictly concave. Conversely, if u is strictly concave, then Jensen's inequality implies risk aversion, since

$$U(\delta_{m(\mu)}) = u\left(\int x\mu(dx)\right) \geq intu(x)\mu(dx) = U(\mu)$$

with equality iff $\mu = \delta_{m(\mu)}$. ∎

Definition 2.1.6. *A function $u : S \to \mathbb{R}$ is called utility function if it is strictly concave, strictly increasing and continuous on S.*

By the intermediate value theorem there is for any $\mu \in \mathcal{M}$ a unique number $c(\mu)$ such that

$$u(c(\mu)) = U(\mu) = \int u d\mu. \tag{2.4}$$

So $\delta_{c(\mu)} \sim \mu$, i.e. there is indifference between μ and the sure amount $c(\mu)$.

Definition 2.1.7. *The certainty equivalent of μ, denoted by $c(\mu)$ is the number defined in (2.4). It is the amount of money for which the individual is indifferent between the lottery μ and the certain amount $c(\mu)$ The number*

$$\rho(\mu) = m(\mu) - c(\mu) \tag{2.5}$$

is called the risk premium.

The certainty equivalent can be viewed as the upper price an investor would pay for the asset distribution μ. Thus the fair price must be reduced by the risk premium if one wants an agent to buy the asset.

Consider now an investor who has the choice to invest a fraction of his wealth in a riskfree and the remaining fraction of his wealth in a risky asset. We want to find conditions on the distribution of the risky asset and the preferences (utility function) of the investor in order to determine his willingness for a risky investment. Formally, we consider the following optimisation problem

$$f(\lambda) = U(\mu_\lambda) = \int u \, d\mu_\lambda \to \max, \tag{2.6}$$

where μ_λ is the distribution of

$$X_\lambda = (1-\lambda)X + \lambda c$$

with X an integrable random variable with non-degenerate distribution μ and $c \in S$ is a certain amount.

Proposition 2.1.2. *We have $\lambda^* = 1$ if $I\!\!E(X) \le c$ and $\lambda^* > 0$ if $c \ge c(\mu)$.*

Proof. By Jensen's inequality

$$f(\lambda) \le u\left(I\!\!E\left[X_\lambda\right]\right) = u\left((1-\lambda)I\!\!E(X) + \lambda c\right)$$

with equality iff $\lambda = 1$. It follows that $\lambda^* = 1$ if the right-hand side is increasing in λ, i.e. $I\!\!E(X) \le c$.

Now, strict concavity of u implies

$$f(\lambda) \ge I\!\!E\left((1-\lambda)u(X) + \lambda u(c)\right) = (1-\lambda)u(c(\mu)) + \lambda u(c)$$

with equality iff $\lambda \in \{0,1\}$. The right-hand side is increasing in λ if $c \ge c(\mu)$, and this implies $\lambda^* > 0$. ∎

Remark 2.1.2. *(i) (Demand for a risky asset.) The price of a risky asset must be below the expected discounted payoff in order to attract any risk-averse investor.*
(ii) (Demand for insurance.) Risk aversion can create a demand for insurance even if the insurance premium lies above the fair price.
(iii) If $u \in C^1(I\!\!R)$ then

$$\lambda^* = 1 \qquad \Leftrightarrow \qquad I\!\!E(X) \le c$$

$$\lambda^* = 0 \qquad \Leftrightarrow \qquad c \le \frac{I\!\!E\left(Xu'(X)\right)}{I\!\!E\left(U'(X)\right)}.$$

We assume now that μ has a finite variance $\mathbb{V}ar(\mu)$. We consider the Taylor expansion of a sufficiently smooth utility function $u(x)$ at $x = c(\mu)$ around $m = m(\mu)$. We have

$$u(c(\mu)) \approx u(m) + u'(m)\left(c(\mu) - m\right) = u(m) + u'(m)\rho(m).$$

On the other hand,

$$
\begin{aligned}
u(c(\mu)) &= \int u(x)\mu(dx) + \int \left[u(m) + u'(m)\left(c(\mu) - m\right) + \frac{1}{2}u''(m)\left(c(\mu) - m\right)^2 + r(x)\right]\mu(dx) \\
&\approx u(m) + \frac{1}{2}u''(m)\mathbb{V}ar(\mu)
\end{aligned}
$$

(where $r(x)$ denotes the remainder term in the Taylor expansion). So

$$\rho(\mu) \approx -\frac{u''(m)}{2u'(m)}\mathbb{V}ar(\mu) = \frac{1}{2}\alpha(m)\mathbb{V}ar(\mu). \tag{2.7}$$

Thus $\alpha(m(\mu))$ is the factor by which an economic agent with utility function u weighs the risk (measured by $\mathbb{V}ar(\mu)$) in order to determine the risk premium.

Definition 2.1.8. *Suppose that u is a twice continuously differentiable utility function on S. Then*

$$\alpha(x) = -\frac{u''(x)}{u'(x)} \tag{2.8}$$

is called the Arrow-Pratt coefficient of absolute risk aversion of u at level x.

Example 2.1.1. *(i) Constant absolute risk aversion (CARA). Here $\alpha(x) \equiv \alpha$ for some constant. This implies a (normalized) utility function*

$$u(x) = 1 - e^{-\alpha x}.$$

(ii) Hyperbolic absolute risk aversion (HARA). Here $\alpha(x) = (1-\gamma)/x$ for some $\gamma \in [0,1)$. This implies

$$
\begin{aligned}
u(x) &= \log x \quad for \quad \gamma = 0 \\
u(x) &= x^{\gamma}/\gamma \quad for \quad 0 < \gamma < 1.
\end{aligned}
$$

Remark 2.1.3. *For u and $\tilde{u}$ two utility functions on S with α and $\tilde{\alpha}$ the corresponding Arrow-Pratt coefficients we have*

$$\alpha(x) \geq \tilde{\alpha}(x) \quad \forall x \in S \quad \Leftrightarrow \quad \rho(\mu) \geq \tilde{\rho}(\mu)$$

with $\rho, \tilde{\rho}$ the corresponding risk premia.

It is also useful to analyse risk aversion in terms of the proportion of wealth that is invested in a risky asset.

Definition 2.1.9. *Suppose that u is a twice continuously differentiable utility function on S. Then*

$$\alpha_R(x) = \alpha(x)x = -x\frac{u''(x)}{u'(x)} \tag{2.9}$$

is called the Arrow-Pratt coefficient of relative risk aversion of u at level x.

Remark 2.1.4. *(i) An individuals utility function displays decreasing (constant, increasing) absolute risk aversion if $\alpha(x)$ is decreasing (constant, increasing).*

(ii) An individuals utility function displays decreasing (constant, increasing) relative risk aversion if $\alpha_R(x)$ is decreasing (constant, increasing).

2.1.3 Further measures of risk

In this section we focus on the question whether one (risky asset) distribution is preferred to another, regardless of the choice of a particular utility function. Assume $S = I\!R$ and $\mathcal{M} = cM_1$ (expectation exists).

Definition 2.1.10. *Let μ, ν be elements of $\mathcal{M}$. We say that μ is uniformly preferred over ν, notation $\mu \succ_{uni} \nu$ if for all utility functions u*

$$\int u d\mu \geq \int u d\nu. \tag{2.10}$$

$\succ_{uni}$ *is also called second order stochastic dominance.*

Theorem 2.1.3. *For any pair $\mu, \nu \in \mathcal{M}$ the following conditions are equivalent:*

(i) $u \succ_{uni} \nu$.

(ii) $\int f d\mu \geq \int f d\nu$ *for all increasing concave functions* f.

(iii) For all $c \in \mathbb{R}$ *we have*

$$\int (c-x)^+ \mu(dx) \leq \int (c-x)^+ \nu(dx). \tag{2.11}$$

(iv) If F *and* G *denote the respective distribution functions of* μ *and* ν, *then*

$$\int_{-\infty}^{c} F(x)dx \leq \int_{-\infty}^{c} G(x)dx \quad \forall c \in \mathbb{R}. \tag{2.12}$$

Proof. Recall u is a utility function iff it is strictly concave and strictly increasing.

$(\mathbf{a}) \Rightarrow (\mathbf{b})((b) \Rightarrow (a)$ is clear). Choose any utility function u_0 with $\int u_0 d\mu$ and $\int u_0 d\nu$ finite. For instance,

$$u_0(x) = \begin{cases} x - e^{x/2} + 1 & \text{if} \quad x \leq 0 \\ \sqrt{x+1} - 1 & \text{if} \quad x \geq 0 \end{cases}$$

Then for f concave and increasing and for $\alpha \in [0,1)$

$$u_\alpha(x) = \alpha f(x) + (1-\alpha)u_0(x)$$

is a utility function. Hence

$$\int f d\mu = \lim_{\alpha \uparrow 1} \int u_\alpha d\mu \geq \lim_{\alpha \uparrow 1} \int u_\alpha d\nu = \int f d\nu$$

$(\mathbf{b}) \Leftrightarrow (\mathbf{c})$

"$\Rightarrow$": Follows, because $f(x) = -(c-x)^+$ is concave and increasing.

"$\Leftarrow$": Let f be an increasing concave function and $h = -f$. Then h is convex and decreasing and its increasing right-hand derivative $h := h'_+$ can be regarded as a distribution function of a non-negative (Radon) measure γ on $\mathbb{R}$. Thus

$$h'(b) = h'(a) + \gamma([a,b]) \quad \text{for} \quad a < b.$$

For $x < b$ we find

$$h(x) = h(b) - h'(b)(b-x) + \int_{[-\infty,b]} (z-x)^+ \gamma(dz) \quad \text{for} \quad x < b.$$

Using (c), the fact that, $h'(b) \leq 0$ and Fubini's theorem we obtain

$$\int_{(-\infty,b]} h d\mu$$
$$= \quad h(b)(\mu(-\infty,b]) - h'(b)\int(b-x)^+ \mu(dx) + \int_{(-\infty,b]} \int (z-x)^+ \mu(dx)\gamma(dz)$$
$$\leq \quad h(b)(\nu(-\infty,b]) - h'(b)\int(b-x)^+ \nu(dx) + \int_{(-\infty,b]} \int (z-x)^+ \nu(dx)\gamma(dz)$$
$$= \quad \int_{(-\infty,b]} h d\nu.$$

Now letting $b \uparrow \infty$ yields $\int f d\mu \geq \int f d\nu$.

$(\mathbf{c}) \Leftrightarrow (\mathbf{d})$. By Fubini's theorem

$$\int_{-\infty}^{c} F(y)dy \quad = \quad \int_{-\infty}^{c} \int_{(-\infty,y)} \mu(dz)dy$$
$$= \quad \int \int \mathbf{1}(z \leq y \leq c)dy\mu(dz)$$
$$= \quad \int (c-z)^+ \mu(dz).$$

Remark 2.1.5. *(i) Also:* $\mu \geq_{uni} \nu \Leftrightarrow \int_0^\lambda F^{-1}(y)dy \geq \int_0^\lambda G^{-1}(y)dy$ *for all* $\lambda \in (0,1]$, *where*
F^{-1}, G^{-1} *are the inverses, or quartile functions of the distribution.*

(ii) Taking $f(x) = x$ *in* (b), *we see* $\mu \geq_{uni} \nu \Rightarrow m(\mu) \geq m(\nu)$.

(iii) For normal distributions, we have

$$\mathcal{N}(m, \sigma^2) \geq_{uni} \mathcal{N}$$

iff $m \geq \tilde{m}$ *and* $\sigma^2 \leq \tilde{\sigma}^2$

If $\mu, \nu \subset \mathcal{M}$ such that $m(\mu) = m(\nu)$ and $\mu \geq_{uni} \nu$ then $var(\mu) \leq var(\nu)$. Here

$$var(\mu) = \int (x - m(\mu))\mu(dx) = \int x^2 \mu(dx) - m(\mu)^2$$

is the variance of μ (use condition (b) with $f(x) = -x^2$) which holds under $m(\nu) = m(\nu)$ for **all**
concave functions.

In the financial context, comparison of portfolios with known payoff distributions often use a
mean-variance approach with

$$\mu \geq \nu \Leftrightarrow m(\nu) \geq m(\nu) \quad \text{and} \quad var(\mu) \leq var(\nu).$$

For normal distributions $\mu \geq \nu$ is equivalent to $\mu \geq_{uni} \nu$, but not in general.

Example 2.1.2. $\mu = U[-1,1]$, *so* $m(\mu) = 0, var(\mu) = \frac{1}{2} \int_{-1}^{1} x^2 dx = \frac{1}{3}\left(\frac{x^3}{3}\big|_{-1}^1\right) = \frac{1}{3}$ *for* $\nu = $
$p\,\delta_{-1/2} + (1-p)\delta_2$, *with* $p = \frac{4}{5}$ *we have*

$$m(\nu) = \frac{4}{5}\left(-\frac{1}{2}\right) + \frac{1}{5} \cdot 2 = 0, \quad var(\nu) = \frac{4}{5} \cdot \frac{1}{4} + \frac{1}{5} \cdot 4 = 1$$

Thus $var(\nu) > var(\mu)$, *but*

$$\int(-\tfrac{1}{2} - x)^+ \mu(dx) = \tfrac{1}{2} \int_{-1}^{-1/2}(-\tfrac{1}{2} - x)dx = \tfrac{1}{2}(-\tfrac{1}{2}x - \tfrac{1}{2}x)\big|_{-1}^{-1/2}$$

$$= \quad \tfrac{1}{2}\left(\tfrac{1}{4} - \tfrac{1}{8} - \tfrac{1}{2} + \tfrac{1}{2}\right) = 1/16$$

and

$$\int\left(-\frac{1}{2} - x\right)^+ \nu(dx) = 0.$$

So $\mu \geq_{uni} \nu$ *does not hold.*

A further important class of distributions is discussed in the following example.

Example 2.1.3. *A real-valued random variable* Y *on some probability space* $(\Omega, \mathcal{F}, I\!P)$ *is called*
log-normally distributed *with parameters* $\alpha \in \mathbb{R}$ *and* $\sigma \in \mathbb{R}_+$ *if it can be written as*

$$Y = \exp(\alpha + \sigma X)$$

where X *has a standard normal law* $\mathcal{N}(0,1)$. *For log-normal distributions* $\mu \geq_{uni} \breve{\mu} \Leftrightarrow \sigma^2 \leq \tilde{\sigma}^2$
and $\alpha + \frac{1}{2}\sigma^2 \geq \tilde{\alpha} + \frac{1}{2}\tilde{\sigma}^2$

Definition 2.1.11. *Let* μ *and* ν *be two arbitrary probability measures on* $\mathbb{R}$. *We say that* μ
stochastically dominates ν, *notation* $\mu \geq_{mon} \nu$ *if*

$$\int f d\mu \geq \int f d\nu$$

for all bounded increasing functions $f \in C(\mathbb{R})$. *Stochastic dominance is also called first-order
stochastic dominance.*

Theorem 2.1.4. *For $\mu, v \in \mathcal{M}_1(\mathbb{R})$ the following conditions are equivalent*

(a) $\mu \geq_{mon} \nu$;

(b) *for all $x, F(x) \leq G(x)$ where F, G are respectively the distribution functions of μ, ν;*

(c) *there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with random variables X and Y having respective distributions μ and ν such that $X \geq Y \quad \mathbb{P}-a.s.$*

2.2 Optimal Portfolios

2.2.1 The mean-variance approach

Recall our one-period model with securities $S_0, S_1, \ldots, S_d$ and security prices $S_i(T)$ at the final time $t = T$. Here S_0 is the risk-free bond and $S_1, \ldots, S_d$ are random variables on some prob. space $(\Omega, \mathcal{F}, \mathbb{P})$. For the purpose of this section we disregard the risk free asset and invest only in the risky assets. We consider their returns

$$R_i(T) = \frac{S_i(T)}{S_i(0)} \quad i = 1, \ldots, d$$

and assume we know (or have estimated) their means and covariance matrix

$$\mathbb{E}(R_i(T)) = m_i \quad i = 1, \ldots, d$$

and

$$\mathbb{C}ov(R_i(T), R_j(T)) = \sigma_{ij} \quad i, j = 1, \ldots, d$$

(Observe that $\Sigma = (\sigma_{ij})$ is positive semi-definite).

We consider portfolio vectors $\varphi_i \in \mathbb{R}^d$ with $\varphi_i \geq 0$ (in order to avoid the possibility of negative final wealth).

Definition 2.2.1. *An investor with initial wealth $x > 0$ is assumed to hold $\varphi_i \geq 0$ shares of security $i, i = 1, \ldots, d$ with*

$$\sum_{i=1}^{d} \varphi_1 S_i(0) = x \quad "budget\ equation".$$

Then the portfolio vector $\pi = (\pi_1, \ldots, \pi_d)$ is defined as

$$\pi_i = \frac{\varphi_i \cdot S_i(0)}{x} \quad i = 1, \ldots, d$$

and

$$R^{\pi} = \sum_{i=1}^{d} \pi_i R_i(T)$$

is called the corresponding portfolio return.

Remark 2.2.1. *(1) The components of the portfolio vector represent the fractions of total wealth invested in the corresponding securities. In particular, we have*

$$\sum_{i=1}^{d} \pi_i = \frac{1}{x} \sum_{i=1}^{d} \varphi_i S_i(0) = \frac{x}{x} = 1$$

(2) Let $V^\pi(T)$ denote the final wealth corresponding to an initial wealth of x and a portfolio vector φ, i.e.

$$V^\pi(T) = \sum_{i=1}^d \varphi_i S_i(T)$$

then we find

$$R^\pi = \sum_{i=1}^d \pi_i R_i(T) = \sum_{i=1}^d \frac{\varphi_i S_i(0)}{x} \cdot \frac{S_i(T)}{S_i(0)} = \frac{V^\pi(T)}{x}$$

(3) The mean and the variance of the portfolio return are given by

$$\mathbb{E}(R^\pi) = \sum_{i=1}^d \pi_i m_i, \quad \mathbb{V}ar(R^\pi) = \sum_{i=1}^d \sum_{j=1}^d \pi_i \sigma_{ij} \pi_j$$

We now need to consider criteria for selecting a portfolio. The basic (by now classical) idea of Markowitz was to look for a balance between risk (i.e. portfolio variance) and return (i.e. portfolio mean). He considered the problem of requiring a lower bound for the portfolio return (minimum return) and then choosing from the corresponding set the portfolio vector with the minimal variance. Alternatively, set an upper bound for the variance and determine the portfolio vector with the highest possible mean return. We consider

Definition 2.2.2. *A portfolio is a frontier portfolio if it has the minimum variance among portfolios that have the same expected rate of return. The set of all frontier portfolios is called the portfolio frontier.*

We now discuss briefly the assumptions of the mean-variance approach.

(1) A preference for expected return and an aversion to variance is implied by monotonicity and strict concavity of a utility function. However, for arbitrary utility functions, expected utility cannot be defined over just the expected return and variances. For $\mu \in \mathcal{M}$ assume

$$
\begin{aligned}
U(\mu) &= \int u(x)\mu(dx) = \int \sum_{k=0}^{\infty} \frac{1}{k} u^{(k)}(m)(x-m)^k \mu(dx) \\
&= u(m) + \frac{1}{2} u''(m) \mathbb{V}ar(\mu) + R_3(\mu)
\end{aligned}
$$

(i.e. convergence of Taylor series and interchangeability of integral). Thus, the remainder term needs (for the general case) to be considered as well.

(2) Assuming quadratic utility, i.e.

$$u(x) = x - \frac{b}{2} x^2 \quad, b > 0$$

we find

$$U(\mu) = m - \frac{b}{2} m_2 = m - \frac{b}{2}(\mathbb{V}ar(\mu) + m^2).$$

Unfortunately, quadratic utility displays satiation (negative utility for increasing wealth) and increasing absolute risk aversion.

(3) For μ_i normal distributions, $R^\pi \sim \mathcal{N}$, thus preferences can be expressed solely from mean and variance.

Proposition 2.2.1. *A portfolio p is a frontier portfolio if and only if the portfolio weight vector π^p is a solution to the optimisation problem*

$$\min_{\boldsymbol{\pi}} \frac{1}{2} \boldsymbol{\pi}' \boldsymbol{\Sigma} \boldsymbol{\pi} = \min_{\boldsymbol{\pi}} \frac{1}{2} \sum_i \sum_j \pi_i \pi_j \sigma_{ij}^2 \quad (= 2\mathbb{V}ar(R^\pi)) \tag{2.13}$$

subject to

$$\boldsymbol{\pi}'\boldsymbol{\mu} = \sum_i \pi_i \mu_i \quad = \quad \mathbb{E}(R^\pi) := m_p$$

$$\boldsymbol{\pi}'\mathbf{1} = \sum_i \pi_i \quad = \quad 1.$$

$\mathbf{1}$ *is an N-vector of ones,* $\boldsymbol{m} = (m_1, \ldots, m_d)$ *is the vector of expected returns of the assets and* m_p *is a fixed rate of portfolio-return.*

We can solve (2.13) and write the solution as

$$\boldsymbol{\pi}_p = \boldsymbol{g} + \boldsymbol{h}m_p \tag{2.14}$$

with

$$\boldsymbol{g} \quad = \quad \frac{1}{D}\left[B(\boldsymbol{\Sigma}^{-1}\mathbf{1}) - A(\boldsymbol{\Sigma}^{-1}\boldsymbol{m})\right]$$

$$\boldsymbol{h} \quad = \quad \frac{1}{D}\left[C(\boldsymbol{\Sigma}^{-1}\boldsymbol{m}) - A(\boldsymbol{\Sigma}^{-1}\mathbf{1})\right],$$

where $A = \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m} = \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\mathbf{1}, B = \boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}, C = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}$ and $D = BC - A^2 > 0$.

For $m_p = 0$ we find from (2.14) that the optimal portfolio is g. Also, for $m_p = 1$ (2.14) implies that the optimal portfolio is $g + h$. Now for any given expected return m_q we find

$$\boldsymbol{\pi}^q = \boldsymbol{g} + \boldsymbol{h}m_q = (1 - m_q)\boldsymbol{g} + m_q(\boldsymbol{g} + \boldsymbol{h}).$$

This generalizes to

Proposition 2.2.2 (Two-fund separation). *The portfolio frontier can be generated by any two distinct frontier portfolios.*

The covariance between the rates of return of any frontier portfolios p and q is

$$\mathbb{C}ov\left(R^p, R^q\right) = \frac{C}{D}\left(m_p - \frac{A}{C}\right)\left(m_q - \frac{A}{C}\right) + \frac{1}{C}. \tag{2.15}$$

Thus, for the variances

$$\frac{\sigma^2(R^p)}{1/C} - \frac{(m_p - A/C)^2}{D/C^2} = 1,$$

which is a hyperbola in the standard-deviation – expected return $(\sigma - \mu)$space.

(i) The portfolio having the minimum variance of all feasible portfolios is called minimum variance portfolio and denoted as **mvp**.

(ii) A frontier portfolio is **efficient** if it has a strictly higher expected return than the mvp.

(iii) Frontier portfolios that are neither mvp nor efficient are called **inefficient**.

(iv) The **efficient frontier** is the part of the curve lying above the point of global minimum of standard deviation.

We can use (2.15) to show that for any frontier portfolio p (except the mvp) there exists a unique frontier portfolio $zc(p)$ which has zero covariance with p.

Proposition 2.2.3. *If* $\boldsymbol{\pi}$ *is any envelope portfolio, then for any other portfolio (envelope or not)* $\tilde{\boldsymbol{\pi}}$ *we have the relation*

$$m_\pi = c + \beta_\pi(m_{\tilde{\pi}} - c) \tag{2.16}$$

where

$$\beta_\pi = \frac{\mathbb{C}ov(\boldsymbol{\pi}, \tilde{\boldsymbol{\pi}})}{\sigma_{\tilde{\pi}}^2}.$$

Furthermore, c is the expected return of a portfolio $\boldsymbol{\pi}^*$ *whose covariance with* $\tilde{\boldsymbol{\pi}}$ *is zero.*

Existence of a risk-free asset has the effect of making the efficient frontier a straight line extending from the rate to the point where the line is tangential to the original efficient frontier for the risky assets. This leads to the *one-fund theorem*: There is a single fund (portfolio) such that any efficient portfolio can be constructed as a combination of the fund and the risk-free asset.

The implication of Proposition 2.2.3 in the presence of the a risk-free asset is that there exists a linear relationship between any security and portfolios on the efficient frontiers involving so-called 'beta factors'.

$$m_i - r = \beta_{i,p}(m_p - r), \tag{2.17}$$

where $\beta_{i,p}$ is a linear factor defined as

$$\beta_{i,p} = \mathbb{C}ov(R_i, R^p)/\sigma_p^2.$$

This extends to any portfolio q, i.e.

$$m_q - r = \beta_{q,p}(m_p - r),$$

with $\beta_{q,p} = \mathbb{C}ov(R^q, R^p)/\sigma_p^2$.

2.2.2 Capital asset pricing model

We now consider a so-called **Equilibrium Model.** The focus of attention is turned from the individual investor to the aggregate market for securities (and all investors) as a whole.

We need assumptions on the investors' behaviour and the market as a whole.

- All investors have the same one-period horizon.

- All investors can borrow or lend at the same risk-free rate.

- The markets for risky assets are perfect. Information is freely and instantly available to all investors and no investor believes that they can affect the price of a security by their own action.

- Investors have the same estimates of the expected returns, standard deviations and covariances over the one-period horizon.

- All investors measure in the same numéraire.

Under the assumptions of mean-variance theory we have in equilibrium

1. If investors have homogeneous expectations, then they are all faced by the same efficient frontier of risky securities.

2. If there is a risk-free asset the efficient frontier collapses for all investors to a straight line which passes through the risk-free rate of return on the m axis and is tangential to the efficient frontier.

All investors face the same efficient frontier because they have the same views on the available securities. Thus:

1. All rational investors will hold a combination of the risk-free asset and the portfolio of assets where the straight line through the risk-free return touches the original efficient frontier.

2. Because investors share the same efficient frontier they all hold the same diversified portfolio. Because this portfolio is held in different quantities by all investors it must consist of all risky assets in proportion to their market capitalization. It is commonly called the 'market portfolio' .

3. Other strategies are non-optimal.

The line denoting the efficient frontier is called the **capital market line** and its equation is

$$m_p - r = (m_M - r)\frac{\sigma_p}{\sigma_M} \tag{2.18}$$

where m_p is the expected return of any portfolio p on the efficient frontier; σ_p is the standard deviation of the return on portfolio p; m_M is the expected return on the market portfolio; σ_M is the standard deviation of the market portfolio and r is the risk-free rate of return.

Thus the expected return on any portfolio is a linear function of its standard deviation. The factor $\frac{m_M - r}{\sigma_M}$ is called the **market price of risk**.

We can also develop an equation relating the expected return of any asset to the return of the market

$$m_i - r = (m_M - r)\beta_i \tag{2.19}$$

where m_i is the expected return on security i; β_i is the beta factor of security i, defined as $\mathbb{C}ov(R_i, R^M)/\mathbb{V}ar(R^M)$; m_M is the expected return on the market portfolio and r is the risk-free rate of return. Equation (2.19) is called the **security market line**. It shows that the expected return of any security (and portfolio) can be expressed as a linear function of the securities covariance with the market as a whole.

2.2.3 Portfolio optimisation and the absence of arbitrage

Consider the standard one-period model with assets $(S_0, \ldots, S_d)$, S_0 the risk-free bond with interest rate $r > 0$. Denote by $\varphi = (\varphi_0, \ldots, \varphi_d)$ portfolio vectors specifying the amount of shares of the assets in the portfolio.

We consider an investor with utility function $\tilde{u}$. A rational choice of the investor's portfolio will be based on expected utility

$$\mathbb{E}(\tilde{u}(\varphi'S(T)))$$

of the payoff $\varphi'S(T)$ at time T, where the portfolio φ satisfies the budget constraint

$$\varphi'S(0) \leq \omega$$

with ω the initial of the investor. We consider the discounted net gain.

$$\frac{\varphi'S(T)}{1+r} - \varphi'S(0) = \varphi'Y$$

with $Y = (Y_0, Y_1, \ldots, Y_d)$ and $Y_i = \frac{S_i(T)}{1+r} - S_i(0)$. (Here $Y_0 = 0$!).

For any portfolio φ with $\varphi'S(0) < \omega$, adding the risk-free investment $\omega - \varphi'S(0)$ would lead to the strictly better portfolio $(\varphi_0 + \omega - \varphi'S(0), \varphi_1, \ldots, \varphi_d)$. Thus we can focus on φ with $\varphi'S(0) = \omega$. Then the payoff is an affine function of the discounted net gain

$$\varphi'S(T) = (1+r)(\varphi'Y + \omega).$$

Since $Y_0 = 0$ we only need the focus on the risky assets. Define the following transformation of the original utility function $\tilde{u}$

$$u(y) = \tilde{u}((1+r)(y+\omega)).$$

So the optimization problem is equivalent to maximizing the expected utility of

$$\mathbb{E}(u(\varphi'Y)) \tag{2.20}$$

among all $\varphi \in \mathbb{R}^d$ such that $\varphi'Y$ is contained in the domain D of u.

Assumption A1. Either

(a) $D = \mathbb{R}$, then we admit all $\varphi' \in \mathbb{R}^d$, but assume the u is bounded from above (Example: $u(x) = 1 - e^{-\alpha x}$.)

or

(b) $D = [a, \infty)$ for some $a < 0$. In this case we only consider portfolios with $\varphi'Y \geq a$ and assume that the expected utility generated by such portfolios is finite, i.e.

$$I\!\!E[u(\varphi'Y)] < \infty \quad \text{for all} \quad \varphi \in I\!\!R^d \quad \text{with} \quad \varphi'Y \geq a.$$

(Example $u(x) = \frac{1}{\gamma}(x - c)^\gamma$).

Denote by

$$\mathcal{S}(D) = \{\varphi \in I\!\!R^d | \varphi'Y \in D\}$$

the set of admissible portfolios for D. Our aim now is to find a $\varphi^* \in \mathcal{S}(D)$ which maximizes the expected utility $I\!\!E(u(\varphi'Y))$ among $\varphi \in \mathcal{S}(D)$.

Theorem 2.2.1. *Let the above assumption A1 hold true. Then there exits a maximizer of the expected utility*

$$I\!\!E[n(\varphi'Y)] \quad \varphi \in \mathcal{S}(D)$$

if the only if the market model is arbitrage-free. Moreover, there exists at most one maximizer if the market model is complete.

Proof. Uniqueness follows from the strict concavity of the function $\varphi \to I\!\!E(u(\varphi'Y))$ for complete market models.
In case the model is incomplete we can find a complete submodel and apply the result to this submodel. So we may assume completeness. (Recall completeness is equivalent to $\eta'Y = 0 \Rightarrow \eta = 0$). If the model admits arbitrage, we find a vector $\xi' \cdot Y \geq 0 I\!\!P$-a.s. and $I\!\!P(\xi'Y) > 0$ with no initial investment. So for φ^* optimal

$$I\!\!E(u(\varphi^{*'}Y)) < I\!\!E(u((\varphi^* + \eta)'Y))$$

a contradiction.
Assume now that the market is arbitrage-free. We consider the case $D = [a, \infty)$ for some $a \in (-\infty, 0)$. We show

(i) $\mathcal{S}(D)$ is compact;

(ii) $\varphi \to I\!\!E[u(\varphi'Y)]$ is continuous.

Clearly, (i) and (ii) imply existence of a maximizer.
To show **(i)**, assume that (φ_n) is a diverging sequence in $\mathcal{S}(D)$. By choosing a subsequence if necessary, we may assume that $\eta_n = \varphi_n/|\varphi_n|$ converges to some unit vector $\eta \in I\!\!R^d$. Then

$$\eta'Y = \lim_{n \to \infty} \frac{\varphi_n'Y}{|\varphi_n|} \geq \frac{a}{|\varphi_n|} = 0 \quad I\!\!P a.s.$$

and so by completeness $\tilde{\varphi} = (-S(0)'\eta, \eta)$ is an arbitrage opportunity. However, under completeness $\eta'Y = 0$ $I\!\!P$a.s. implies $\eta = 0$.
(ii) To show continuity it suffices to construct an integrable random variable which dominates $u(\varphi'Y)$ for all $\varphi \in \mathcal{S}(D)$. Define $\eta \in I\!\!R^d$ by

$$\eta_i = 0 \vee \max_{\varphi \in \mathcal{S}(D)} \varphi^i < \infty.$$

Then, $\eta'S(T) \geq \varphi'S(T)$ for $\varphi \in \mathcal{S}(D)$ and hence

$$\varphi'Y = \frac{\varphi'S(T)}{1+r} - \varphi'S(0) \leq \frac{\eta' \cdot S(T)}{1+r} - \left(0 \wedge \min_{\xi \in \mathcal{S}(D)} \xi'S(0)\right).$$

Now $\eta'Y$ is bounded below by $-\eta'S(0)$ and there exists some $\alpha \in (0, 1]$ such that $\alpha\eta'S(0) < |a|$. Hence $\alpha\eta \in \mathcal{S}(D)$ and by our assumption $I\!\!E(u(\alpha\eta'Y)) < \infty$.

Applying Lemma 2.2.1 first with $b := -\alpha S(0)'\eta$ and then with $b := -0 \wedge \min\limits_{\varphi \in \mathcal{S}(D)} \varphi' S(0)$ shows that

$$\mathbb{E}\left[u\left(\frac{\eta' S(0)}{1+r} - \sigma \wedge \min_{\varphi \in \mathcal{S}(D)} \varphi' S(0)\right)\right] < \infty.$$

■

Lemma 2.2.1. *If $D = [a, \infty), b < |a|, 0 < \alpha \leq 1$, and X is a non-negative random variable, then*

$$\mathbb{E}[u(\alpha X - b)] < \infty \Rightarrow \mathbb{E}[u(X)] < \infty.$$

Proof. The concavity of n implies that u has a decreasing right-continuous derivative u'. Hence

$$u(aX - b) = u(-b) + \int_{-b}^{\alpha X - b} u'(x)dx \geq u(-b) + \int_{0}^{\alpha X} u'(y)dy$$

$$= u(-b) + \alpha \int_{0}^{X} u'(\alpha z)dz \geq u(-b) + \alpha(u(X) - u(0)).$$

This shows that $u(X)$ can be dominated by a multiple of $u(\alpha X - b)$ plus some constant. ■

We now turn to a characterization of the solution φ^* of the utility maximization problem for continuously differentiable utility functions.

Proposition 2.2.4. *Let n be a twice continuously differentiable utility function on D such that $\mathbb{E}(u(\varphi'Y))$ is finite for all $\varphi \in \mathcal{S}(D)$. Suppose that φ^* is a solution of the utility maximization problem, and that one of the following conditions is satisfied. Either*

- *u is defined on $D = \mathbb{R}$ and bounded from above or*

- *u is defined on $D = [a, \infty)$ and φ^* is an interior point of $\mathcal{S}(D)$.*

Then

$$u'(\varphi^{*'}Y)|Y| \in \mathcal{L}^1(\mathbb{P})$$

and the following first-order condition holds

$$\mathbb{E}[u'(\varphi^{*'}Y) \cdot Y] = 0$$

Proof. For $\varphi \in \mathcal{S}(D)$ and $\varepsilon \in [0, 1]$ let $\varphi_\varepsilon = \varepsilon\varphi + (1 - \varepsilon)\varphi^*$ and define

$$\Delta_\varepsilon = \frac{u(\varphi_\varepsilon'Y) - u(\varphi^{*'}Y)}{\varepsilon}$$

The concavity of u implies that $\Delta_\varepsilon \geq \Delta_\delta$ for $\varepsilon \leq \delta$, and so

$$\Delta_\varepsilon \uparrow u'(\varepsilon *' Y)(\varphi - \varphi^*)'Y \quad \text{as} \quad \varepsilon \downarrow 0. \tag{2.21}$$

Since $\Delta_1 = u(\varepsilon'Y) \in \mathcal{L}(\mathbb{P})$ monotone convergence and the optimality of φ^* yield

$$0 \geq \mathbb{E}(\Delta_\varepsilon) \uparrow \mathbb{E}\left(u'(\varphi^{*'}Y)(\varphi - \varphi^*)'Y\right) \quad \text{as} \quad \varepsilon \downarrow 0. \tag{2.22}$$

In particular, the expectation on the right-hand side is finite.

Both sets of assumptions imply that φ^* is an interior point of $\mathcal{S}(D)$. Hence from (2.22) we find by letting $\eta = \varphi - \varphi^*$ that

$$\mathbb{E}(u'(\varphi *' Y)\eta'Y) \leq 0$$

for all η in a small ball centered in the origin of $\mathbb{R}^d$. Replacing η by $-\eta$ shows that the expectation must vanish; i.e. $\mathbb{E}(u'(\varphi^{*'}Y)\eta'Y) = 0 \quad \forall \eta$ in a small ball around the origin, so $\mathbb{E}(u'(\varphi^{*'}Y)\eta'Y) = 0$. ■

We can now give a characterisation of an equivalent risk-neutral measure.

Corollary 2.2.1. *Suppose that the market model is arbitrage-free and that the assumptions of Proposition 2.2.4 are satisfied for a utility function $u : D \to \mathbb{R}$. Let ϕ^* be a maximizer of the expected utility. Then*

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = \frac{u'(\varphi^{*'}Y)}{\mathbb{E}(u'(\varphi^{*'}Y))} \qquad (2.23)$$

defines an equivalent risk-neutral measure.

Proof. Recall that $\mathbb{E}^*(Y) = 0$ is the criterion for risk-neutrality which is satisfied by Proposition 2.2.4. Hence $\mathbb{P}^*$ is a risk-neutral measure if it is well-defined, i.e.

$$\mathbb{E}(u'(\varphi^{*'}Y)) \in \mathcal{L}^1(\mathbb{P}).$$

Let

$$c = \sup\left\{u'(x)|x \in D \text{ and } |x| \le |\varphi^*|\right\} \le \begin{cases} u'(a) & D \in [a, \infty) \\ u'(-|\varphi^*|) & D = \mathbb{R} \end{cases}$$

which is finite by our assumption that u is continuously differentiable on D. So

$$0 \le u'(\varphi^{*'}Y) \le c + u'(\varphi^{*'}Y)|Y|\mathbf{1}_{\{|Y|\ge1\}}$$

and the right-hand side has finite expectation.∎

Remark 2.2.2. *We can now give a constructive proof of the first fundamental theorem of asset pricing. Suppose the model is arbitrage-free.*
(i) If Y is a.s. $\mathbb{P}$-a.s. bounded, then so is $u'(\varphi^{*'}Y)$ and the measure $\mathbb{P}^*$ is an equivalent martingale measure with a bounded density.
(ii) If Y is unbounded we may consider the bounded random vector

$$\tilde{Y} = \frac{Y}{1 + |Y|}$$

which also satisfies the no-arbitrage condition.

Let $\tilde{\varphi}^*$ be a maximiser of expected utility $\mathbb{E}(u'(\varphi'\tilde{Y}))$. Then an equivalent martingale measure is given by $\mathbb{P}^*$ defined via the bounded density

$$\frac{d\mathbb{P}^*}{d\mathbb{P}} = c\frac{u'(\tilde{\varphi}^{*'}\tilde{Y})}{1 + |Y|}$$

where c is an appropriate constant.

Chapter 3

Discrete-time models of financial markets

3.1 The model

We will study so-called finite markets – i.e. discrete-time models of financial markets in which all relevant quantities take a finite number of values. Following the approach of Harrison and Pliska (1981) and Taqqu and Willinger (1987), it suffices, to illustrate the ideas, to work with a finite probability space $(\Omega, \mathcal{F}, I\!\!P)$, with a finite number $|\Omega|$ of points ω, each with positive probability: $I\!\!P(\{\omega\}) > 0$.

We specify a time horizon T, which is the terminal date for all economic activities considered. (For a simple option pricing model the time horizon typically corresponds to the expiry date of the option.)

As before, we use a filtration $I\!\!F = \{\mathcal{F}_t\}_{t=0}^T$ consisting of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_T$: we take $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial σ-field, $\mathcal{F}_T = \mathcal{F} = \mathcal{P}(\Omega)$ (here $\mathcal{P}(\Omega)$ is the power-set of Ω, the class of all $2^{|\Omega|}$ subsets of Ω: we need every possible subset, as they all – apart from the empty set – carry positive probability).

The financial market contains $d + 1$ financial assets. The usual interpretation is to assume one risk-free asset (bond, bank account) labelled 0, and d risky assets (stocks, say) labelled 1 to d. While the reader may keep this interpretation as a mental picture, we prefer not to use it directly. The prices of the assets at time t are random variables, $S_0(t, \omega), S_1(t, \omega), \ldots, S_d(t, \omega)$ say, non-negative and $\mathcal{F}_t$-measurable (i.e. adapted: at time t, we know the prices $S_i(t)$). We write $S(t) = (S_0(t), S_1(t), \ldots, S_d(t))'$ for the vector of prices at time t. Hereafter we refer to the probability space $(\Omega, \mathcal{F}, I\!\!P)$, the set of trading dates, the price process S and the information structure $I\!\!F$, which is typically generated by the price process S, together as a securities market model.

It will be essential to assume that the price process of at least one asset follows a strictly positive process.

Definition 3.1.1. *A numéraire is a price process $(X(t))_{t=0}^T$ (a sequence of random variables), which is strictly positive for all $t \in \{0, 1, \ldots, T\}$.*

For the standard approach the risk-free bank account process is used as numéraire. In some applications, however, it is more convenient to use a security other than the bank account and we therefore just use S_0 without further specification as a numéraire. We furthermore take $S_0(0) = 1$ (that is, we reckon in units of the initial value of our numéraire), and define $\beta(t) := 1/S_0(t)$ as a discount factor.

A *trading strategy* (or *dynamic portfolio*) φ is a $I\!\!R^{d+1}$ vector stochastic process $\varphi = (\varphi(t))_{t=1}^T = ((\varphi_0(t, \omega), \varphi_1(t, \omega), \ldots, \varphi_d(t, \omega))')_{t=1}^T$ which is predictable (or previsible): each $\varphi_i(t)$ is $\mathcal{F}_{t-1}$-measurable for $t \geq 1$. Here $\varphi_i(t)$ denotes the number of shares of asset i held in the portfolio at time t – to

be determined on the basis of information available *before* time t; i.e. the investor selects his time t portfolio after observing the prices $S(t-1)$. However, the portfolio $\varphi(t)$ must be established before, and held until after, announcement of the prices $S(t)$. The components $\varphi_i(t)$ may assume negative as well as positive values, reflecting the fact that we allow short sales and assume that the assets are perfectly divisible.

Definition 3.1.2. *The value of the portfolio at time t is the scalar product*

$$V_\varphi(t) = \varphi(t) \cdot S(t) := \sum_{i=0}^{d} \varphi_i(t) S_i(t), \ (t = 1, 2, \ldots, T) \ \ and \ V_\varphi(0) = \varphi(1) \cdot S(0).$$

The process $V_\varphi(t, \omega)$ is called the wealth or value process of the trading strategy φ.

The initial wealth $V_\varphi(0)$ is called the *initial investment* or *endowment* of the investor.

Now $\varphi(t) \cdot S(t-1)$ reflects the market value of the portfolio just after it has been established at time $t-1$, whereas $\varphi(t) \cdot S(t)$ is the value just after time t prices are observed, but before changes are made in the portfolio. Hence

$$\varphi(t) \cdot (S(t) - S(t-1)) = \varphi(t) \cdot \Delta S(t)$$

is the change in the market value due to changes in security prices which occur between time $t-1$ and t. This motivates:

Definition 3.1.3. *The gains process G_φ of a trading strategy φ is given by*

$$G_\varphi(t) := \sum_{\tau=1}^{t} \varphi(\tau) \cdot (S(\tau) - S(\tau-1)) = \sum_{\tau=1}^{t} \varphi(\tau) \cdot \Delta S(\tau), \ \ (t = 1, 2, \ldots, T).$$

Observe the – for now – formal similarity of the gains process G_φ from trading in S following a trading strategy φ to the martingale transform of S by φ.

Define $\tilde{S}(t) = (1, \beta(t) S_1(t), \ldots, \beta(t) S_d(t))'$, the vector of discounted prices, and consider the *discounted value process*

$$\tilde{V}_\varphi(t) = \beta(t)(\varphi(t) \cdot S(t)) = \varphi(t) \cdot \tilde{S}(t), \ \ (t = 1, 2, \ldots, T)$$

and the *discounted gains process*

$$\tilde{G}_\varphi(t) := \sum_{\tau=1}^{t} \varphi(\tau) \cdot (\tilde{S}(\tau) - \tilde{S}(\tau-1)) = \sum_{\tau=1}^{t} \varphi(\tau) \cdot \Delta\tilde{S}(\tau), \ \ (t = 1, 2, \ldots, T).$$

Observe that the discounted gains process reflects the gains from trading with assets 1 to d only, which in case of the standard model (a bank account and d stocks) are the risky assets.

We will only consider special classes of trading strategies.

Definition 3.1.4. *The strategy φ is self-financing, $\varphi \in \Phi$, if*

$$\varphi(t) \cdot S(t) = \varphi(t+1) \cdot S(t) \ \ (t = 1, 2, \ldots, T-1). \tag{3.1}$$

Interpretation.

When new prices $S(t)$ are quoted at time t, the investor adjusts his portfolio from $\varphi(t)$ to $\varphi(t+1)$, without bringing in or consuming any wealth. The following result (which is trivial in our current setting, but requires a little argument in continuous time) shows that renormalising security prices (i.e. changing the numéraire) has essentially no economic effects.

Proposition 3.1.1 (Numéraire Invariance). *Let $X(t)$ be a numéraire. A trading strategy φ is self-financing with respect to $S(t)$ if and only if φ is self-financing with respect to $X(t)^{-1} S(t)$.*

Proof. Since $X(t)$ is strictly positive for all $t = 0, 1, \ldots, T$ we have the following equivalence, which implies the claim:

$$\varphi(t) \cdot S(t) = \varphi(t+1) \cdot S(t) \quad (t = 1, 2, \ldots, T-1)$$
$$\Leftrightarrow$$
$$\varphi(t) \cdot X(t)^{-1} S(t) = \varphi(t+1) \cdot X(t)^{-1} S(t) \quad (t = 1, 2, \ldots, T-1).$$

$\blacksquare$

Corollary 3.1.1. *A trading strategy φ is self-financing with respect to $S(t)$ if and only if φ is self-financing with respect to $\tilde{S}(t)$.*

We now give a characterisation of self-financing strategies in terms of the discounted processes.

Proposition 3.1.2. *A trading strategy φ belongs to Φ if and only if*

$$\tilde{V}_\varphi(t) = V_\varphi(0) + \tilde{G}_\varphi(t), \quad (t = 0, 1, \ldots, T). \tag{3.2}$$

Proof. Assume $\varphi \in \Phi$. Then using the defining relation (3.1), the numéraire invariance theorem and the fact that $S_0(0) = 1$

$$\begin{aligned}
V_\varphi(0) + \tilde{G}_\varphi(t) &= \varphi(1) \cdot S(0) + \sum_{\tau=1}^{t} \varphi(\tau) \cdot (\tilde{S}(\tau) - \tilde{S}(\tau-1)) \\
&= \varphi(1) \cdot \tilde{S}(0) + \varphi(t) \cdot \tilde{S}(t) \\
&\quad - \sum_{\tau=1}^{t-1} (\varphi(\tau) - \varphi(\tau+1)) \cdot \tilde{S}(\tau) - \varphi(1) \cdot \tilde{S}(0) \\
&= \varphi(t) \cdot \tilde{S}(t) = \tilde{V}_\varphi(t).
\end{aligned}$$

Assume now that (3.2) holds true. By the numéraire invariance theorem it is enough to show the discounted version of relation (3.1). Summing up to $t = 2$ (3.2) is

$$\varphi(2) \cdot \tilde{S}(2) = \varphi(1) \cdot \tilde{S}(0) + \varphi(1) \cdot (\tilde{S}(1) - \tilde{S}(0)) + \varphi(2) \cdot (\tilde{S}(2) - \tilde{S}(1)).$$

Subtracting $\varphi(2) \cdot \tilde{S}(2)$ on both sides gives $\varphi(2) \cdot \tilde{S}(1) = \varphi(1) \cdot \tilde{S}(1)$, which is (3.1) for $t = 1$. Proceeding similarly – or by induction – we can show $\varphi(t) \cdot \tilde{S}(t) = \varphi(t+1) \cdot \tilde{S}(t)$ for $t = 2, \ldots, T-1$ as required. $\blacksquare$

We are allowed to borrow (so $\varphi_0(t)$ may be negative) and sell short (so $\varphi_i(t)$ may be negative for $i = 1, \ldots, d$). So it is hardly surprising that if we decide what to do about the risky assets and fix an initial endowment, the numéraire will take care of itself, in the following sense.

Proposition 3.1.3. *If $(\varphi_1(t), \ldots, \varphi_d(t))'$ is predictable and V_0 is $\mathcal{F}_0$-measurable, there is a unique predictable process $(\varphi_0(t))_{t=1}^{T}$ such that $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_d)'$ is self-financing with initial value of the corresponding portfolio $V_\varphi(0) = V_0$.*

Proof. If φ is self-financing, then by Proposition 3.1.2,

$$\tilde{V}_\varphi(t) = V_0 + \tilde{G}_\varphi(t) = V_0 + \sum_{\tau=1}^{t} (\varphi_1(\tau)\Delta\tilde{S}_1(\tau) + \ldots + \varphi_d(\tau)\Delta\tilde{S}_d(\tau)).$$

On the other hand,

$$\tilde{V}_\varphi(t) = \varphi(t) \cdot \tilde{S}(t) = \varphi_0(t) + \varphi_1(t)\tilde{S}_1(t) + \ldots + \varphi_d(t)\tilde{S}_d(t).$$

Equate these:

$$\begin{aligned}
\varphi_0(t) &= V_0 + \sum_{\tau=1}^{t} (\varphi_1(\tau)\Delta\tilde{S}_1(\tau) + \ldots + \varphi_d(\tau)\Delta\tilde{S}_d(\tau)) \\
&\quad - (\varphi_1(t)\tilde{S}_1(t) + \ldots + \varphi_d(t)\tilde{S}_d(t)),
\end{aligned}$$

which defines $\varphi_0(t)$ uniquely. The terms in $\tilde{S}_i(t)$ are

$$\varphi_i(t)\Delta\tilde{S}_i(t) - \varphi_i(t)\tilde{S}_i(t) = -\varphi_i(t)\tilde{S}_i(t-1),$$

which is $\mathcal{F}_{t-1}$-measurable. So

$$
\begin{aligned}
\varphi_0(t) &= V_0 + \sum_{\tau=1}^{t-1}(\varphi_1(\tau)\Delta\tilde{S}_1(\tau) + \ldots + \varphi_d(\tau)\Delta\tilde{S}_d(\tau)) \\
&\quad -(\varphi_1(t)S_1(t-1) + \ldots + \varphi_d(t)\tilde{S}_d(t-1)),
\end{aligned}
$$

where as $\varphi_1, \ldots, \varphi_d$ are predictable, all terms on the right-hand side are $\mathcal{F}_{t-1}$-measurable, so φ_0 is predictable. ∎

Remark 3.1.1. *Proposition 3.1.3 has a further important consequence: for defining a gains process $\tilde{G}_\varphi$ only the components $(\varphi_1(t), \ldots, \varphi_d(t))'$ are needed. If we require them to be predictable they correspond in a unique way (after fixing initial endowment) to a self-financing trading strategy. Thus for the discounted world predictable strategies and final cash-flows generated by them are all that matters.*

We now turn to the modelling of derivative instruments in our current framework. This is done in the following fashion.

Definition 3.1.5. *A contingent claim X with maturity date T is an arbitrary $\mathcal{F}_T = \mathcal{F}$-measurable random variable (which is by the finiteness of the probability space bounded). We denote the class of all contingent claims by $L^0 = L^0(\Omega, \mathcal{F}, I\!P)$.*

The notation L^0 for contingent claims is motivated by the them being simply random variables in our context (and the functional-analytic spaces used later on).

A typical example of a contingent claim X is an option on some underlying asset S, then (e.g. for the case of a European call option with maturity date T and strike K) we have a functional relation $X = f(S)$ with some function f (e.g. $X = (S(T) - K)^+$). The general definition allows for more complicated relationships which are captured by the $\mathcal{F}_T$-measurability of X (recall that $\mathcal{F}_T$ is typically generated by the process S).

3.2   Existence of Equivalent Martingale Measures

3.2.1   The No-Arbitrage Condition

The central principle in the single period example was the absence of arbitrage opportunities, i.e. the absence investment strategies for making profits without exposure to risk. As mentioned there this principle is central for any market model, and we now define the mathematical counterpart of this economic principle in our current setting.

Definition 3.2.1. *Let $\tilde{\Phi} \subset \Phi$ be a set of self-financing strategies. A strategy $\varphi \in \tilde{\Phi}$ is called an arbitrage opportunity or arbitrage strategy with respect to $\tilde{\Phi}$ if $I\!P\{V_\varphi(0) = 0\} = 1$, and the terminal wealth of φ satisfies*

$$I\!P\{V_\varphi(T) \geq 0\} = 1 \quad and \quad I\!P\{V_\varphi(T) > 0\} > 0.$$

So an arbitrage opportunity is a self-financing strategy with zero initial value, which produces a non-negative final value with probability one and has a positive probability of a positive final value. Observe that arbitrage opportunities are always defined with respect to a certain class of trading strategies.

Definition 3.2.2. *We say that a security market $\mathcal{M}$ is arbitrage-free if there are no arbitrage opportunities in the class Φ of trading strategies.*

We will allow ourselves to use 'no-arbitrage' in place of 'arbitrage-free' when convenient.

The fundamental insight in the single-period example was the equivalence of the no-arbitrage condition and the existence of risk-neutral probabilities. For the multi-period case we now use probabilistic machinery to establish the corresponding result.

Definition 3.2.3. *A probability measure $\mathbb{P}^*$ on $(\Omega, \mathcal{F}_T)$ equivalent to $\mathbb{P}$ is called a martingale measure for $\tilde{S}$ if the process $\tilde{S}$ follows a $\mathbb{P}^*$-martingale with respect to the filtration $\mathbb{F}$. We denote by $\mathcal{P}(\tilde{S})$ the class of equivalent martingale measures.*

Proposition 3.2.1. *Let $\mathbb{P}^*$ be an equivalent martingale measure ($\mathbb{P}^* \in \mathcal{P}(\tilde{S})$) and $\varphi \in \Phi$ any self-financing strategy. Then the wealth process $\tilde{V}_\varphi(t)$ is a $\mathbb{P}^*$-martingale with respect to the filtration $\mathbb{F}$.*

Proof. By the self-financing property of φ (compare Proposition 3.1.2, (3.2)), we have

$$\tilde{V}_\varphi(t) = V_\varphi(0) + \tilde{G}_\varphi(t) \quad (t = 0, 1, \ldots, T).$$

So

$$\tilde{V}_\varphi(t+1) - \tilde{V}_\varphi(t) = \tilde{G}_\varphi(t+1) - \tilde{G}_\varphi(t) = \varphi(t+1) \cdot (\tilde{S}(t+1) - \tilde{S}(t)).$$

So for $\varphi \in \Phi$, $\tilde{V}_\varphi(t)$ is the martingale transform of the $\mathbb{P}^*$ martingale $\tilde{S}$ by φ (see Theorem C.4.1) and hence a $\mathbb{P}^*$ martingale itself. ∎

Observe that in our setting all processes are bounded, i.e. the martingale transform theorem is applicable without further restrictions. The next result is the key for the further development.

Proposition 3.2.2. *If an equivalent martingale measure exists - that is, if $\mathcal{P}(\tilde{S}) \neq \emptyset$ – then the market $\mathcal{M}$ is arbitrage-free.*

Proof. Assume such a $\mathbb{P}^*$ exists. For any self-financing strategy φ, we have as before

$$\tilde{V}_\varphi(t) = V_\varphi(0) + \sum_{\tau=1}^{t} \varphi(\tau) \cdot \Delta\tilde{S}(\tau).$$

By Proposition 3.2.1, $\tilde{S}(t)$ a (vector) $\mathbb{P}^*$-martingale implies $\tilde{V}_\varphi(t)$ is a P^*-martingale. So the initial and final $\mathbb{P}^*$-expectations are the same,

$$\mathbb{E}^*(\tilde{V}_\varphi(T)) = \mathbb{E}^*(\tilde{V}_\varphi(0)).$$

If the strategy is an arbitrage opportunity its initial value – the right-hand side above – is zero. Therefore the left-hand side $\mathbb{E}^*(\tilde{V}_\varphi(T))$ is zero, but $\tilde{V}_\varphi(T) \geq 0$ (by definition). Also each $\mathbb{P}^*(\{\omega\}) > 0$ (by assumption, each $\mathbb{P}(\{\omega\}) > 0$, so by equivalence each $\mathbb{P}^*(\{\omega\}) > 0$). This and $\tilde{V}_\varphi(T) \geq 0$ force $\tilde{V}_\varphi(T) = 0$. So no arbitrage is possible. ∎

Proposition 3.2.3. *If the market $\mathcal{M}$ is arbitrage-free, then the class $\mathcal{P}(\tilde{S})$ of equivalent martingale measures is non-empty.*

For the proof (for which we follow Schachermayer (2000) we need some auxiliary observations.

Recall the definition of arbitrage, i.e. Definition 3.2.1, in our finite-dimensional setting: a self-financing trading strategy $\varphi \in \Phi$ is an arbitrage opportunity if $V_\varphi(0) = 0$, $V_\varphi(T, \omega) \geq 0 \; \forall \omega \in \Omega$ and there exists a $\omega \in \Omega$ with $V_\varphi(T, \omega) > 0$.

Now call $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ the set of random variables on $(\Omega, \mathcal{F})$ and

$$L^0_{++}(\Omega, \mathcal{F}, \mathbb{P}) := \{X \in L^0 : X(\omega) \geq 0 \; \forall \omega \in \Omega \text{ and } \exists \, \omega \in \Omega \text{ such that } X(\omega) > 0\}.$$

(Observe that L^0_{++} is a *cone* -closed under vector addition and multiplication by positive scalars.) Using L^0_{++} we can write the arbitrage condition more compactly as

$$V_\varphi(0) = \tilde{V}_\varphi(0) = 0 \quad \Rightarrow \quad \tilde{V}_\varphi(T) \notin L^0_{++}(\Omega, \mathcal{F}, \mathbb{P}).$$

for any self-financing strategy φ.

The next lemma formulates the arbitrage condition in terms of discounted gains processes. The important advantage in using this setting (rather than a setting in terms of value processes) is that we only have to assume predictability of a vector process $(\varphi_1, \ldots, \varphi_d)$. Recall Remark 3.1.1 and Proposition 3.1.3 here: we can choose a process φ_0 in such a way that the strategy $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_d)$ has zero initial value and is self-financing.

Lemma 3.2.1. *In an arbitrage-free market any predictable vector process $\varphi' = (\varphi_1, \ldots, \varphi_d)$ satisfies*

$$\tilde{G}_{\varphi'}(T) \notin L^0_{++}(\Omega, \mathcal{F}, I\!P).$$

(Observe the slight abuse of notation: for the value of the discounted gains process the zeroth component of a trading strategy doesn't matter. Hence we use the operator $\tilde{G}$ for d-dimensional vectors as well.)

Proof. By Proposition 3.1.3 there exists a unique predictable process $(\varphi_0(t))$ such that $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_d)$ has zero initial value and is self-financing. Assume $\tilde{G}_{\varphi'}(T) \in L^0_{++}(\Omega, \mathcal{F}, I\!P)$. Then using Proposition 3.1.2,

$$V_\varphi(T) = \beta(T)^{-1}\tilde{V}_\varphi(T) = \beta(T)^{-1}(V_\varphi(0) + \tilde{G}_\varphi(T)) = \beta(T)^{-1}\tilde{G}_{\varphi'}(T) \geq 0,$$

and is positive somewhere (i.e. with positive probability) by definition of L^0_{++}. Hence φ is an arbitrage opportunity with respect to Φ. This contradicts the assumption that the market is arbitrage-free. ∎

We now define the space of contingent claims, i.e. random variables on $(\Omega, \mathcal{F})$, which an economic agent may replicate with zero initial investment by pursuing some predictable trading strategy φ.

Definition 3.2.4. *We call the subspace K of $L^0(\Omega, \mathcal{F}, I\!P)$ defined by*

$$K = \{X \in L^0(\Omega, \mathcal{F}, I\!P) : X = \tilde{G}_\varphi(T), \ \varphi \ predictable\}$$

the set of contingent claims attainable at price 0.

We can now restate Lemma 3.2.1 in terms of spaces
A market is arbitrage-free if and only if

$$K \cap L^0_{++}(\Omega, \mathcal{F}, I\!P) = \emptyset. \tag{3.3}$$

Proof of Proposition 3.2.3. Since our market model is finite we can use results from Euclidean geometry, in particular we can identify L^0 with $I\!R^{|\Omega|}$). By assumption we have (3.3), i.e. K and L^0_{++} do not intersect. So K does not meet the subset

$$D := \{X \in L^0_{++} : \sum_{\omega \in \Omega} X(\omega) = 1\}.$$

Now D is a compact convex set. By the separating hyperplane theorem, there is a vector $\lambda = (\lambda(\omega) : \omega \in \Omega)$ such that for all $X \in D$

$$\lambda \cdot X := \sum_{\omega \in \Omega} \lambda(\omega)X(\omega) > 0, \tag{3.4}$$

but for all $\tilde{G}_\varphi(T)$ in K,

$$\lambda \cdot \tilde{G}_\varphi(T) = \sum_{\omega \in \Omega} \lambda(\omega)\tilde{G}_\varphi(T)(\omega) = 0. \tag{3.5}$$

Choosing each $\omega \in \Omega$ successively and taking X to be 1 on this ω and zero elsewhere, (3.4) tells us that each $\lambda(\omega) > 0$. So

$$\mathbb{P}^*(\{\omega\}) := \frac{\lambda(\omega)}{\sum_{\omega' \in \Omega} \lambda(\omega')}$$

defines a probability measure equivalent to $\mathbb{P}$ (no non-empty null sets). With $\mathbb{E}^*$ as $\mathbb{P}^*$-expectation, (3.5) says that

$$\mathbb{E}^* \left(\tilde{G}_\varphi(T) \right) = 0,$$

i.e.

$$\mathbb{E}^* \left(\sum_{\tau=1}^{T} \varphi(\tau) \cdot \Delta \tilde{S}(\tau) \right) = 0.$$

In particular, choosing for each i to hold only stock i,

$$\mathbb{E}^* \left(\sum_{\tau=1}^{T} \varphi_i(\tau) \Delta \tilde{S}_i(\tau) \right) = 0 \quad (i = 1, \ldots, d).$$

Since this holds for any predictable φ (boundedness holds automatically as Ω is finite), the martingale transform lemma tells us that the discounted price processes $(\tilde{S}_i(t))$ are $\mathbb{P}^*$-martingales. ∎

Note. Our situation is finite-dimensional, so all we have used here is Euclidean geometry. We have a subspace, and a cone not meeting the subspace except at the origin. Take λ orthogonal to the subspace on the same side of the subspace as the cone. The separating hyperplane theorem holds also in infinite-dimensional situations, where it is a form of the Hahn-Banach theorem of functional analysis.

We now combine Propositions 3.2.2 and 3.2.3 as a first central theorem in this chapter.

Theorem 3.2.1 (No-Arbitrage Theorem). *The market $\mathcal{M}$ is arbitrage-free if and only if there exists a probability measure $\mathbb{P}^*$ equivalent to $\mathbb{P}$ under which the discounted d-dimensional asset price process $\tilde{S}$ is a $\mathbb{P}^*$-martingale.*

3.2.2 Risk-Neutral Pricing

We now turn to the main underlying question of this text, namely the pricing of contingent claims (i.e. financial derivatives). As in chapter 1 the basic idea is to reproduce the cash flow of a contingent claim in terms of a portfolio of the underlying assets. On the other hand, the equivalence of the no-arbitrage condition and the existence of risk-neutral probability measures imply the possibility of using risk-neutral measures for pricing purposes. We will explore the relation of these tow approaches in this subsection.

We say that a contingent claim is *attainable* if there exists a *replicating strategy* $\varphi \in \Phi$ such that

$$V_\varphi(T) = X.$$

So the replicating strategy generates the same time T cash-flow as does X. Working with discounted values (recall we use β as the discount factor) we find

$$\beta(T)X = \tilde{V}_\varphi(T) = V(0) + \tilde{G}_\varphi(T). \tag{3.6}$$

So the discounted value of a contingent claim is given by the initial cost of setting up a replication strategy and the gains from trading. In a highly efficient security market we expect that the law of one price holds true, that is for a specified cash-flow there exists only one price at any time instant. Otherwise arbitrageurs would use the opportunity to cash in a riskless profit. So the no-arbitrage condition implies that for an attainable contingent claim its time t price must be given by the value (inital cost) of any replicating strategy (we say the claim is uniquely replicated in that case). This is the basic idea of the *arbitrage pricing theory*.

Let us investigate replicating strategies a bit further. The idea is to replicate a given cash-flow at a given point in time. Using a self-financing trading strategy the investor's wealth may go negative at time $t < T$, but he must be able to cover his debt at the final date. To avoid negative wealth the concept of admissible strategies is introduced. A self-financing trading strategy $\varphi \in \Phi$ is called *admissible* if $V_\varphi(t) \geq 0$ for each $t = 0, 1, \ldots, T$. We write Φ_a for the class of admissible trading strategies. The modelling assumption of admissible strategies reflects the economic fact that the broker should be protected from unbounded short sales. In our current setting all processes are bounded anyway, so this distinction is not really needed and we use self-financing strategies when addressing the mathematical aspects of the theory. (In fact one can show that a security market which is arbitrage-free with respect to Φ_a is also arbitrage-free with respect to Φ; see Exercises.)

We now return to the main question of the section: given a contingent claim X, i.e. a cash-flow at time T, how can we determine its value (price) at time $t < T$? For an attainable contingent claim this value should be given by the value of any replicating strategy at time t, i.e. there should be a unique value process (say $V_X(t)$) representing the time t value of the simple contingent claim X. The following proposition ensures that the value processes of replicating trading strategies coincide, thus proving the uniqueness of the value process.

Proposition 3.2.4. *Suppose the market $\mathcal{M}$ is arbitrage-free. Then any attainable contingent claim X is uniquely replicated in $\mathcal{M}$.*

Proof. Suppose there is an attainable contingent claim X and strategies φ and ψ such that

$$V_\varphi(T) = V_\psi(T) = X,$$

but there exists a $\tau < T$ such that

$$V_\varphi(u) = V_\psi(u) \text{ for every } u < \tau \text{ and } V_\varphi(\tau) \neq V_\psi(\tau).$$

Define $A := \{\omega \in \Omega : V_\varphi(\tau, \omega) > V_\psi(\tau, \omega)\}$, then $A \in \mathcal{F}_\tau$ and $I\!P(A) > 0$ (otherwise just rename the strategies). Define the $\mathcal{F}_\tau$-measurable random variable $Y := V_\varphi(\tau) - V_\psi(\tau)$ and consider the trading strategy ξ defined by

$$\xi(u) = \begin{cases} \varphi(u) - \psi(u), & u \leq \tau \\ \mathbf{1}_{A^c}(\varphi(u) - \psi(u)) + \mathbf{1}_A(Y\beta(\tau), 0, \ldots, 0), & \tau < u \leq T. \end{cases}$$

The idea here is to use φ and ψ to construct a self-financing strategy with zero initial investment (hence use their difference ξ) and put any gains at time τ in the savings account (i.e. invest them riskfree) up to time T.

We need to show formally that ξ satisfies the conditions of an arbitrage opportunity. By construction ξ is predictable and the self-financing condition (3.1) is clearly true for $t \neq \tau$, and for $t = \tau$ we have using that $\varphi, \psi \in \Phi$

$$\begin{aligned} \xi(\tau) \cdot S(\tau) &= (\varphi(\tau) - \psi(\tau)) \cdot S(\tau) = V_\varphi(\tau) - V_\psi(\tau), \\ \xi(\tau+1) \cdot S(\tau) &= \mathbf{1}_{A^c}(\varphi(\tau+1) - \psi(\tau+1)) \cdot S(\tau) + \mathbf{1}_A Y\beta(\tau)S_0(\tau) \\ &= \mathbf{1}_{A^c}(\varphi(\tau) - \psi(\tau)) \cdot S(\tau) + \mathbf{1}_A(V_\varphi(\tau) - V_\psi(\tau))\beta(\tau)\beta^{-1}(\tau) \\ &= V_\varphi(\tau) - V_\psi(\tau). \end{aligned}$$

Hence ξ is a self-financing strategy with initial value equal to zero. Furthermore

$$\begin{aligned} V_\xi(T) &= \mathbf{1}_{A^c}(\varphi(T) - \psi(T)) \cdot S(T) + \mathbf{1}_A(Y\beta(\tau), 0, \ldots, 0) \cdot S(T) \\ &= \mathbf{1}_A Y\beta(\tau)S_0(T) \geq 0 \end{aligned}$$

and

$$I\!P\{V_\xi(T) > 0\} = I\!P\{A\} > 0.$$

Hence the market contains an arbitrage opportunity with respect to the class Φ of self-financing strategies. But this contradicts the assumption that the market $\mathcal{M}$ is arbitrage-free. ∎

This uniqueness property allows us now to define the important concept of an arbitrage price process.

Definition 3.2.5. *Suppose the market is arbitrage-free. Let X be any attainable contingent claim with time T maturity. Then the arbitrage price process $\pi_X(t)$, $0 \leq t \leq T$ or simply arbitrage price of X is given by the value process of any replicating strategy φ for X.*

The construction of hedging strategies that replicate the outcome of a contingent claim (for example a European option) is an important problem in both practical and theoretical applications. Hedging is central to the theory of option pricing. The classical arbitrage valuation models, such as the Black-Scholes model ((Black and Scholes 1973), depend on the idea that an option can be perfectly hedged using the underlying asset (in our case the assets of the market model), so making it possible to create a portfolio that replicates the option exactly. Hedging is also widely used to reduce risk, and the kinds of delta-hedging strategies implicit in the Black-Scholes model are used by participants in option markets. We will come back to hedging problems subsequently.

Analysing the arbitrage-pricing approach we observe that the derivation of the price of a contingent claim doesn't require any specific preferences of the agents other than nonsatiation, i.e. agents prefer more to less, which rules out arbitrage. So, the pricing formula for any attainable contingent claim must be independent of all preferences that do not admit arbitrage. In particular, an economy of risk-neutral investors must price a contingent claim in the same manner. This fundamental insight, due to Cox and Ross (Cox and Ross 1976) in the case of a simple economy – a riskless asset and one risky asset - and in its general form due to Harrison and Kreps (Harrison and Kreps 1979), simplifies the pricing formula enormously. In its general form the price of an attainable simple contingent claim is just the expected value of the discounted payoff with respect to an equivalent martingale measure.

Proposition 3.2.5. *The arbitrage price process of any attainable contingent claim X is given by the **risk-neutral valuation formula***

$$\pi_X(t) = \beta(t)^{-1} I\!\!E^* \left(X\beta(T)|\mathcal{F}_t \right) \quad \forall t = 0, 1, \ldots, T, \tag{3.7}$$

where $I\!\!E^*$ is the expectation operator with respect to an equivalent martingale measure $I\!\!P^*$.

Proof. Since we assume the the market is arbitrage-free there exists (at least) an equivalent martingale measure $I\!\!P^*$. By Proposition 3.2.1 the discounted value process $\tilde{V}_\varphi$ of any self-financing strategy φ is a $I\!\!P^*$-martingale. So for any contingent claim X with maturity T and any replicating trading strategy $\varphi \in \Phi$ we have for each $t = 0, 1, \ldots, T$

$$
\begin{aligned}
\pi_X(t) &= V_\varphi(t) = \beta(t)^{-1} \tilde{V}_\varphi(t) \\[1em]
&= \beta(t)^{-1} E^*(\tilde{V}_\varphi(T)|\mathcal{F}_t) && \text{(as } \tilde{V}_\varphi(t) \text{ is a } I\!\!P^*\text{-martingale)} \\[1em]
&= \beta(t)^{-1} E^*(\beta(T) V_\varphi(T)|\mathcal{F}_t) && \text{(undoing the discounting)} \\[1em]
&= \beta(t)^{-1} E^*(\beta(T) X|\mathcal{F}_t) && \text{(as } \varphi \text{ is a replicating strategy for } X\text{).}
\end{aligned}
$$

$\blacksquare$

3.3 Complete Markets: Uniqueness of Equivalent Martingale Measures

The last section made clear that attainable contingent claims can be priced using an equivalent martingale measure. In this section we will discuss the question of the circumstances under which all contingent claims are attainable. This would be a very desirable property of the market $\mathcal{M}$, because we would then have solved the pricing question (at least for contingent claims) completely. Since contingent claims are merely $\mathcal{F}_T$-measurable random variables in our setting, it should be no surprise that we can give a criterion in terms of probability measures. We start with:

Definition 3.3.1. *A market $\mathcal{M}$ is complete if every contingent claim is attainable, i.e. for every $\mathcal{F}_T$-measurable random variable $X \in L^0$ there exists a replicating self-financing strategy $\varphi \in \Phi$ such that $V_\varphi(T) = X$.*

In the case of an arbitrage-free market $\mathcal{M}$ one can even insist on replicating nonnegative contingent claims by an admissible strategy $\varphi \in \Phi_a$. Indeed, if φ is self-financing and $\mathbb{P}^*$ is an equivalent martingale measure under which discounted prices $\tilde{S}$ are $\mathbb{P}^*$-martingales (such $\mathbb{P}^*$ exist since $\mathcal{M}$ is arbitrage-free and we can hence use the no-arbitrage theorem (Theorem 3.2.1)), $\tilde{V}_\varphi(t)$ is also a $\mathbb{P}^*$-martingale, being the martingale transform of the martingale $\tilde{S}$ by φ (see Proposition 3.2.1). So

$$\tilde{V}_\varphi(t) = E^*(\tilde{V}_\varphi(T)|\mathcal{F}_t) \quad (t = 0, 1, \ldots, T).$$

If φ replicates X, $V_\varphi(T) = X \geq 0$, so discounting, $\tilde{V}_\varphi(T) \geq 0$, so the above equation gives $\tilde{V}_\varphi(t) \geq 0$ for each t. Thus all the values at each time t are non-negative – not just the final value at time T – so φ is admissible.

Theorem 3.3.1 (Completeness Theorem). *An arbitrage-free market $\mathcal{M}$ is complete if and only if there exists a unique probability measure $\mathbb{P}^*$ equivalent to $\mathbb{P}$ under which discounted asset prices are martingales.*

Proof. '$\Rightarrow$': Assume that the arbitrage-free market $\mathcal{M}$ is complete. Then for any $\mathcal{F}_T$-measurable random variable X (contingent claim), there exists an admissible (so self-financing) strategy φ replicating X: $X = V_\varphi(T)$. As φ is self-financing, by Proposition 3.1.2,

$$\beta(T)X = \tilde{V}_\varphi(T) = V_\varphi(0) + \sum_{\tau=1}^{T} \varphi(\tau) \cdot \Delta\tilde{S}(\tau).$$

We know by the no-arbitrage theorem (Theorem 3.2.1) that an equivalent martingale measure $\mathbb{P}^*$ exists; we have to prove uniqueness. So, let $\mathbb{P}_1, \mathbb{P}_2$ be two such equivalent martingale measures. For $i = 1, 2$, $(\tilde{V}_\varphi(t))_{t=0}^{T}$ is a $\mathbb{P}_i$-martingale. So,

$$\mathbb{E}_i(\tilde{V}_\varphi(T)) = \mathbb{E}_i(\tilde{V}_\varphi(0)) = V_\varphi(0),$$

since the value at time zero is non-random ($\mathcal{F}_0 = \{\emptyset, \Omega\}$) and $\beta(0) = 1$. So

$$\mathbb{E}_1(\beta(T)X) = \mathbb{E}_2(\beta(T)X).$$

Since X is arbitrary, $\mathbb{E}_1, \mathbb{E}_2$ have to agree on integrating all integrands. Now $\mathbb{E}_i$ is expectation (i.e. integration) with respect to the measure $\mathbb{P}_i$, and measures that agree on integrating all integrands must coincide. So $\mathbb{P}_1 = \mathbb{P}_2$, giving uniqueness as required.

'$\Leftarrow$': Assume that the arbitrage-free market $\mathcal{M}$ is incomplete: then there exists a non-attainable $\mathcal{F}_T$-measurable random variable X (a contingent claim). By Proposition 3.1.3, we may confine attention to the risky assets $S_1, \ldots, S_d$, as these suffice to tell us how to handle the numéraire S_0.

Consider the following set of random variables:

$$\tilde{K} := \left\{ Y \in L^0 : Y = Y_0 + \sum_{t=1}^{T} \varphi(t) \cdot \Delta\tilde{S}(t), \ Y_0 \in \mathbb{R}, \ \varphi \ \text{predictable} \right\}.$$

(Recall that Y_0 is $\mathcal{F}_0$-measurable and set $\varphi = ((\varphi_1(t), \ldots, \varphi_d(t))')_{t=1}^{T}$ with predictable components.) Then by the above reasoning, the discounted value $\beta(T)X$ does not belong to $\tilde{K}$, so $\tilde{K}$ is a *proper* subset of the set L^0 of all random variables on Ω (which may be identified with $\mathbb{R}^{|\Omega|}$). Let $\mathbb{P}^*$ be a probability measure equivalent to $\mathbb{P}$ under which discounted prices are martingales (such $\mathbb{P}^*$ exist by the no-arbitrage theorem (Theorem 3.2.1)). Define the scalar product

$$(Z, Y) \to \mathbb{E}^*(ZY)$$

on random variables on Ω. Since $\tilde{K}$ is a proper subset, there exists a non-zero random variable Z orthogonal to $\tilde{K}$ (since Ω is finite, $I\!\!R^{|\Omega|}$ is Euclidean: this is just Euclidean geometry). That is,

$$I\!\!E^*(ZY) = 0, \quad \forall \, Y \in \tilde{K}.$$

Choosing the special $Y = 1 \in \tilde{K}$ given by $\varphi_i(t) = 0, t = 1, 2, \ldots, T; \ i = 1, \ldots, d$ and $Y_0 = 1$ we find

$$I\!\!E^*(Z) = 0.$$

Write $\|X\|_\infty := \sup\{|X(\omega)| : \omega \in \Omega\}$, and define $I\!\!P^{**}$ by

$$I\!\!P^{**}(\{\omega\}) = \left(1 + \frac{Z(\omega)}{2\,\|Z\|_\infty}\right) I\!\!P^*(\{\omega\}).$$

By construction, $I\!\!P^{**}$ is equivalent to $I\!\!P^*$ (same null sets - actually, as $I\!\!P^* \sim I\!\!P$ and $I\!\!P$ has no non-empty null sets, neither do $I\!\!P^*, I\!\!P^{**}$). From $I\!\!E^*(Z) = 0$, we see that $\sum I\!\!P^{**}(\omega) = 1$, i.e. is a probability measure. As Z is non-zero, $I\!\!P^{**}$ and $I\!\!P^*$ are *different*. Now

$$
\begin{aligned}
I\!\!E^{**}\left(\sum_{t=1}^T \varphi(t) \cdot \Delta\tilde{S}(t)\right) &= \sum_{\omega \in \Omega} I\!\!P^{**}(\omega)\left(\sum_{t=1}^T \varphi(t, \omega) \cdot \Delta\tilde{S}(t, \omega)\right) \\
&= \sum_{\omega \in \Omega}\left(1 + \frac{Z(\omega)}{2\,\|Z\|_\infty}\right) I\!\!P^*(\omega)\left(\sum_{t=1}^T \varphi(t, \omega) \cdot \Delta\tilde{S}(t, \omega)\right).
\end{aligned}
$$

The '1' term on the right gives

$$I\!\!E^*\left(\sum_{t=1}^T \varphi(t) \cdot \Delta\tilde{S}(t)\right),$$

which is zero since this is a martingale transform of the $I\!\!P^*$-martingale $\tilde{S}(t)$ (recall martingale transforms are by definition null at zero). The 'Z' term gives a multiple of the inner product

$$(Z, \sum_{t=1}^T \varphi(t) \cdot \Delta\tilde{S}(t)),$$

which is zero as Z is orthogonal to $\tilde{K}$ and $\sum_{t=1}^T \varphi(t) \cdot \Delta\tilde{S}(t) \in \tilde{K}$. By the martingale transform lemma (Lemma C.4.1), $\tilde{S}(t)$ is a $I\!\!P^{**}$-martingale since φ is an arbitrary predictable process. Thus $I\!\!P^{**}$ is a second equivalent martingale measure, different from $I\!\!P^*$. So incompleteness implies non-uniqueness of equivalent martingale measures, as required. ∎

Martingale Representation.

To say that every contingent claim can be replicated means that every $I\!\!P^*$-martingale (where $I\!\!P^*$ is the risk-neutral measure, which is unique) can be written, or *represented*, as a martingale transform (of the discounted prices) by a replicating (perfect-hedge) trading strategy φ. In stochastic-process language, this says that all $I\!\!P^*$-martingales can be *represented* as martingale transforms of discounted prices. Such martingale representation theorems hold much more generally, and are very important. For background, see (Revuz and Yor 1991, Yor 1978).

3.4   The Cox-Ross-Rubinstein Model

In this section we consider simple discrete-time financial market models. The development of the risk-neutral pricing formula is particularly clear in this setting since we require only elementary mathematical methods. The link to the fundamental economic principles of the arbitrage pricing method can be obtained equally straightforwardly. Moreover binomial models, by their very construction, give rise to simple and efficient numerical procedures. We start with the paradigm of all binomial models - the celebrated Cox-Ross-Rubinstein model (Cox, Ross, and Rubinstein 1979).

3.4.1 Model Structure

We take $d = 1$, that is, our model consists of two basic securities. Recall that the essence of the relative pricing theory is to take the price processes of these basic securities as given and price secondary securities in such a way that no arbitrage is possible.

Our time horizon is T and the set of dates in our financial market model is $t = 0, 1, \ldots, T$. Assume that the first of our given basic securities is a (riskless) bond or bank account B, which yields a riskless rate of return $r > 0$ in each time interval $[t, t+1]$, i.e.

$$B(t+1) = (1+r)B(t), \quad B(0) = 1.$$

So its price process is $B(t) = (1+r)^t, \quad t = 0, 1, \ldots, T$. Furthermore, we have a risky asset (stock) S with price process

$$S(t+1) = \begin{cases} (1+u)S(t) & \text{with probability} & p, \\ (1+d)S(t) & \text{with probability} & 1-p, \end{cases} \quad t = 0, 1, \ldots, T-1$$

with $-1 < d < u, S_0 \in I\!\!R_0^+$ (see Fig. 3.4.1 below).

$$p$$
$$S(1) = (1+u)S(0)$$
$$S(0)$$
$$1-p$$
$$S(1) = (1+d)S(0)$$

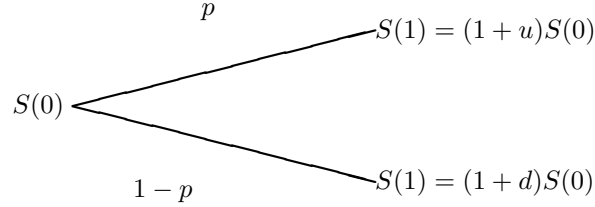Figure 3.1: One-step tree diagram

Alternatively we write this as

$$Z(t+1) := \frac{S(t+1)}{S(t)} - 1, \quad t = 0, 1, \ldots, T-1.$$

We set up a probabilistic model by considering the $Z(t), \; t = 1, \ldots, T$ as random variables defined on probability spaces $(\tilde{\Omega}_t, \tilde{\mathcal{F}}_t, \tilde{I\!\!P}_t)$ with

$$\begin{aligned} \tilde{\Omega}_t \;\; &= \tilde{\Omega} \;\; = \{d, u\}, \\ \tilde{\mathcal{F}}_t \;\; &= \tilde{\mathcal{F}} \;\; = \mathcal{P}(\tilde{\Omega}) = \{\emptyset, \{d\}, \{u\}, \tilde{\Omega}\}, \\ \tilde{I\!\!P}_t \;\; &= \tilde{I\!\!P} \;\;\; \text{with} \;\; \tilde{I\!\!P}(\{u\}) = p, \; \tilde{I\!\!P}(\{d\}) = 1-p, \; p \in (0,1). \end{aligned}$$

On these probability spaces we define

$$Z(t, u) = u \;\; \text{and} \;\; Z(t, d) = d, \quad t = 1, 2, \ldots, T.$$

Our aim, of course, is to define a probability space on which we can model the basic securities (B, S). Since we can write the stock price as

$$S(t) = S(0) \prod_{\tau=1}^{t} (1 + Z(\tau)), \quad t = 1, 2, \ldots, T,$$

the above definitions suggest using as the underlying probabilistic model of the financial market the *product space* $(\Omega, \mathcal{F}, I\!\!P)$ (see e.g. (Williams 1991) ch. 8), i.e.

$$\Omega = \tilde{\Omega}_1 \times \ldots \times \tilde{\Omega}_T = \tilde{\Omega}^T = \{d, u\}^T,$$

with each $\omega \in \Omega$ representing the successive values of $Z(t)$, $t = 1, 2, \ldots, T$. Hence each $\omega \in \Omega$ is a T-tuple $\omega = (\tilde{\omega}_1, \ldots, \tilde{\omega}_T)$ and $\tilde{\omega}_t \in \tilde{\Omega} = \{d, u\}$. For the σ-algebra we use $\mathcal{F} = \mathcal{P}(\Omega)$ and the probability measure is given by

$$I\!\!P(\{\omega\}) = \tilde{I\!\!P}_1(\{\omega_1\}) \times \ldots \times \tilde{I\!\!P}_T(\{\omega_T\}) = \tilde{I\!\!P}(\{\omega_1\}) \times \ldots \times \tilde{I\!\!P}(\{\omega_T\}).$$

The role of a product space is to model independent replication of a random experiment. The $Z(t)$ above are two-valued random variables, so can be thought of as tosses of a biased coin; we need to build a probability space on which we can model a succession of such independent tosses.

Now we redefine (with a slight abuse of notation) the $Z(t)$, $t = 1, \ldots, T$ as random variables on $(\Omega, \mathcal{F}, I\!\!P)$ as (the tth projection)

$$Z(t, \omega) = Z(t, \omega_t).$$

Observe that by this definition (and the above construction) $Z(1), \ldots, Z(T)$ are independent and identically distributed with

$$I\!\!P(Z(t) = u) = p = 1 - I\!\!P(Z(t) = d).$$

To model the flow of information in the market we use the obvious filtration

$$\begin{aligned}
\mathcal{F}_0 &= \{\emptyset, \Omega\} & \text{(trivial } \sigma\text{-field)}, \\
\mathcal{F}_t &= \sigma(Z(1), \ldots, Z(t)) = \sigma(S(1), \ldots, S(t)), \\
\mathcal{F}_T = \mathcal{F} &= \mathcal{P}(\Omega) & \text{(class of all subsets of } \Omega).
\end{aligned}$$

This construction emphasises again that a multi-period model can be viewed as a sequence of single-period models. Indeed, in the Cox-Ross-Rubinstein case we use identical and independent single-period models. As we will see in the sequel this will make the construction of equivalent martingale measures relatively easy. Unfortunately we can hardly defend the assumption of independent and identically distributed price movements at each time period in practical applications.

Remark 3.4.1. *We used this example to show explicitly how to construct the underlying probability space. Having done this in full once, we will from now on feel free to take for granted the existence of an appropriate probability space on which all relevant random variables can be defined.*

3.4.2 Risk-Neutral Pricing

We now turn to the pricing of derivative assets in the Cox-Ross-Rubinstein market model. To do so we first have to discuss whether the Cox-Ross-Rubinstein model is arbitrage-free and complete.

To answer these questions we have, according to our fundamental theorems (Theorems 3.2.1 and 3.3.1), to understand the structure of equivalent martingale measures in the Cox-Ross-Rubinstein model. In trying to do this we use (as is quite natural and customary) the bond price process $B(t)$ as numéraire.

Our first task is to find an equivalent martingale measure $\boldsymbol{Q}$ such that the $Z(1), \ldots, Z(T)$ remain independent and identically distributed, i.e. a probability measure $\boldsymbol{Q}$ defined as a product measure via a measure $\tilde{\boldsymbol{Q}}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ such that $\tilde{\boldsymbol{Q}}(\{u\}) = q$ and $\tilde{\boldsymbol{Q}}(\{d\}) = 1 - q$. We have:

Proposition 3.4.1. *(i) A martingale measure $\boldsymbol{Q}$ for the discounted stock price $\tilde{S}$ exists if and only if*

$$d < r < u. \tag{3.8}$$

(ii) If equation (3.8) holds true, then there is a unique such measure in $\mathcal{P}$ characterised by

$$q = \frac{r - d}{u - d}. \tag{3.9}$$

Proof. Since $S(t) = \tilde{S}(t)B(t) = \tilde{S}(t)(1+r)^t$, we have $Z(t+1) = S(t+1)/S(t) - 1 = (\tilde{S}(t+1)/\tilde{S}(t))(1+r) - 1$. So, the discounted price $(\tilde{S}(t))$ is a $\boldsymbol{Q}$-martingale if and only if for $t = 0, 1, \ldots, T-1$

$$\mathbb{E}^{\boldsymbol{Q}}[\tilde{S}(t+1)|\mathcal{F}_t] = \tilde{S}(t) \Leftrightarrow \mathbb{E}^{\boldsymbol{Q}}[(\tilde{S}(t+1)/\tilde{S}(t))|\mathcal{F}_t] = 1$$
$$\Leftrightarrow \quad \mathbb{E}^{\boldsymbol{Q}}[Z(t+1)|\mathcal{F}_t] = r.$$

But $Z(1), \ldots, Z(T)$ are mutually independent and hence $Z(t+1)$ is independent of $\mathcal{F}_t = \sigma(Z(1), \ldots, Z(t))$. So

$$r = \mathbb{E}^{\boldsymbol{Q}}(Z(t+1)|\mathcal{F}_t) = \mathbb{E}^{\boldsymbol{Q}}(Z(t+1)) = uq + d(1-q)$$

is a weighted average of u and d; this can be r if and only if $r \in [d, u]$. As $\boldsymbol{Q}$ is to be *equivalent* to $\boldsymbol{P}$ and $\boldsymbol{P}$ has no non-empty null sets, $r = d, u$ are excluded and (3.8) is proved.

To prove uniqueness and to find the value of q we simply observe that under (3.8)

$$u \times q + d \times (1-q) = r$$

has a unique solution. Solving it for q leads to the above formula. ∎

From now on we assume that (3.8) holds true. Using the above Proposition we immediately get:

Corollary 3.4.1. *The Cox-Ross-Rubinstein model is arbitrage-free.*

Proof. By Proposition 3.4.1 there exists an equivalent martingale measure and this is by the no-arbitrage theorem (Theorem 3.2.1) enough to guarantee that the Cox-Ross-Rubinstein model is free of arbitrage. ∎

Uniqueness of the solution of the linear equation (4.7) under (3.8) gives completeness of the model, by the completeness theorem (Theorem 3.3.1):

Proposition 3.4.2. *The Cox-Ross-Rubinstein model is complete.*

One can translate this result – on uniqueness of the equivalent martingale measure – into financial language. Completeness means that all contingent claims can be replicated. If we do this in the large, we can do it in the small by restriction, and conversely, we can build up our full model from its constituent components. To summarize:

Corollary 3.4.2. *The multi-period model is complete if and only if every underlying single-period model is complete.*

We can now use the risk-neutral valuation formula to price *every* contingent claim in the Cox-Ross-Rubinstein model.

Proposition 3.4.3. *The arbitrage price process of a contingent claim X in the Cox-Ross-Rubinstein model is given by*

$$\pi_X(t) = B(t)\mathbb{E}^*\left(X/B(T)|\mathcal{F}_t\right) \quad \forall t = 0, 1, \ldots, T,$$

where $\mathbb{E}^*$ is the expectation operator with respect to the unique equivalent martingale measure $\boldsymbol{P}^*$ characterised by $p^* = (r-d)/(u-d)$.

Proof. This follows directly from Proposition 3.2.4 since the Cox-Ross-Rubinstein model is arbitrage-free and complete. ∎

We now give simple formulas for pricing (and hedging) of European contingent claims $X = f(S_T)$ for suitable functions f (in this simple framework all functions $f : \mathbb{R} \to \mathbb{R}$). We use the notation

$$F_\tau(x, p) := \sum_{j=0}^{\tau} \binom{\tau}{j} p^j (1-p)^{\tau-j} f\left(x(1+u)^j(1+d)^{\tau-j}\right) \tag{3.10}$$

Observe that this is just an evaluation of $f(S(j))$ along the probability-weighted paths of the price process. Accordingly, $j, \tau - j$ are the numbers of times $Z(i)$ takes the two possible values d, u.

Corollary 3.4.3. *Consider a European contingent claim with expiry T given by $X = f(S_T)$. The arbitrage price process $\pi_X(t)$, $t = 0, 1, \ldots, T$ of the contingent claim is given by (set $\tau = T - t$)*

$$\pi_X(t) = (1 + r)^{-\tau} F_\tau(S_t, p^*). \tag{3.11}$$

Proof. Recall that

$$S(t) = S(0) \prod_{j=1}^{t} (1 + Z(j)), \quad t = 1, 2, \ldots, T.$$

By Proposition 3.4.3 the price $\Pi_X(t)$ of a contingent claim $X = f(S_T)$ at time t is

$$
\begin{aligned}
\pi_X(t) &= (1 + r)^{-(T-t)} I\!E^*[f(S(T))|\mathcal{F}_t] \\[2mm]
&= (1 + r)^{-(T-t)} I\!E^*\left[f\left(S(t) \prod_{i=t+1}^{T} (1 + Z(i)) \right) \middle| \mathcal{F}_t \right] \\[2mm]
&= (1 + r)^{-(T-t)} I\!E^*\left[f\left(S(t) \prod_{i=t+1}^{T} (1 + Z(i)) \right) \right] \\[2mm]
&= (1 + r)^{-\tau} F_\tau(S(t), p^*).
\end{aligned}
$$

We used the role of independence property of conditional expectations from Proposition B.5.1 in the next-to-last equality. It is applicable since $S(t)$ is $\mathcal{F}_t$-measurable and $Z(t + 1), \ldots, Z(T)$ are independent of $\mathcal{F}_t$. ∎

An immediate consequence is the pricing formula for the European call option, i.e. $X = f(S_T)$ with $f(x) = (x - K)^+$.

Corollary 3.4.4. *Consider a European call option with expiry T and strike price K written on (one share of) the stock S. The arbitrage price process $\Pi_C(t)$, $t = 0, 1, \ldots, T$ of the option is given by (set $\tau = T - t$)*

$$\Pi_C(t) = (1 + r)^{-\tau} \sum_{j=0}^{\tau} \binom{\tau}{j} p^{*j} (1 - p^*)^{\tau - j} (S(t)(1 + u)^j (1 + d)^{\tau - j} - K)^+. \tag{3.12}$$

For a European put option, we can either argue similarly or use put-call parity.

3.4.3 Hedging

Since the Cox-Ross-Rubinstein model is complete we can find unique hedging strategies for replicating contingent claims. Recall that this means we can find a self-financing portfolio $\varphi(t) = (\varphi_0(t), \varphi_1(t))$, φ predictable, such that the value process $V_\varphi(t) = \varphi_0(t)B(t) + \varphi_1(t)S(t)$ satisfies

$$\Pi_X(t) = V_\varphi(t), \quad \text{for all } t = 0, 1, \ldots, T.$$

Using the bond as numéraire we get the discounted equation

$$\tilde{\Pi}_X(t) = \tilde{V}_\varphi(t) = \varphi_0(t) + \varphi_1(t)\tilde{S}(t), \quad \text{for all } t = 0, 1, \ldots, T.$$

By the pricing formula, Proposition 3.4.3, we know the arbitrage price process and using the restriction of predictability of φ, this leads to a unique replicating portfolio process φ. We can compute this portfolio process at any point in time as follows. The equation $\tilde{\Pi}_X(t) = \varphi_0(t) + \varphi_1(t)\tilde{S}(t)$ has to be true for each $\omega \in \Omega$ and each $t = 1, \ldots, T$. Given such a t we only can use information up to (and including) time $t - 1$ to ensure that φ is predictable. Therefore we know $S(t - 1)$, but we only know that $S(t) = (1 + Z(t))S(t - 1)$. However, the fact that $Z(t) \in \{d, u\}$

leads to the following system of equations, which can be solved for $\varphi_0(t)$ and $\varphi_1(t)$ uniquely. Making the dependence of $\tilde{\Pi}_X$ on $\tilde{S}$ explicit, we have

$$\begin{aligned}
\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+u)) &= \varphi_0(t) + \varphi_1(t)\tilde{S}_{t-1}(1+u), \\
\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+d)) &= \varphi_0(t) + \varphi_1(t)\tilde{S}_{t-1}(1+d).
\end{aligned}$$

The solution is given by

$$\begin{aligned}
\varphi_0(t) &= \frac{\tilde{S}_{t-1}(1+u)\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+d)) - \tilde{S}_{t-1}(1+d)\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+u))}{\tilde{S}_{t-1}(1+u) - \tilde{S}_{t-1}(1+d)} \\
&= \frac{(1+u)\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+d)) - (1+d)\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+u))}{(u-d)} \\
\varphi_1(t) &= \frac{\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+u)) - \tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+d))}{\tilde{S}_{t-1}(1+u) - \tilde{S}_{t-1}(1+d)} \\
&= \frac{\tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+u)) - \tilde{\Pi}_X(t, \tilde{S}_{t-1}(1+d))}{\tilde{S}_{t-1}(u-d)}.
\end{aligned}$$

Observe that we only need to have information up to time $t-1$ to compute $\varphi(t)$, hence φ is predictable. We make this rather abstract construction more transparent by constructing the hedge portfolio for the European contingent claims.

Proposition 3.4.4. *The perfect hedging strategy $\varphi = (\varphi_0, \varphi_1)$ replicating the European contingent claim $f(S_T)$ with time of expiry T is given by (again using $\tau = T - t$)*

$$\begin{aligned}
\varphi_1(t) &= \frac{(1+r)^{-\tau}\left(F_\tau(S_{t-1}(1+u), p^*) - F_\tau(S_{t-1}(1+d), p^*)\right)}{S_{t-1}(u-d)}, \\
\varphi_0(t) &= \frac{(1+u)F_\tau(S_{t-1}(1+d), p^*) - (1+d)F_\tau(S_{t-1}(1+u), p^*)}{(u-d)(1+r)^T}.
\end{aligned}$$

Proof. $(1+r)^{-\tau}F_\tau(S_t, p^*)$ must be the value of the portfolio at time t if the strategy $\varphi = (\varphi(t))$ replicates the claim:
$$\varphi_0(t)(1+r)^t + \varphi_1(t)S(t) = (1+r)^{-\tau}F_\tau(S_t, p^*).$$

Now $S(t) = S(t-1)(1+Z(t)) = S(t-1)(1+u)$ or $S(t-1)(1+d)$, so:

$$\varphi_0(t)(1+r)^t + \varphi_1(t)S(t-1)(1+u) = (1+r)^{-\tau}F_\tau(S_{t-1}(1+u), p^*),$$

$$\varphi_0(t)(1+r)^t + \varphi_1(t)S(t-1)(1+d) = (1+r)^{-\tau}F_\tau(S_{t-1}(1+d), p^*).$$

Subtract:

$$\varphi_1(t)S(t-1)(u-d) = (1+r)^{-\tau}\left(F_\tau(S_{t-1}(1+u), p^*) - F_\tau(S_{t-1}(1+d), p^*)\right).$$

So $\varphi_1(t)$ in fact depends only on $S(t-1)$, thus yielding the predictability of φ, and

$$\varphi_1(t) = \frac{(1+r)^{-\tau}\left(F_\tau(S_{t-1}(1+u), p^*) - F_\tau(S_{t-1}(1+d), p^*)\right)}{S(t-1)(u-d)}.$$

Using any of the equations in the above system and solving for $\varphi_0(t)$ completes the proof. ∎

To write the corresponding result for the European call, we use the following notation.

$$C(\tau, x) := \sum_{j=0}^{\tau} \binom{\tau}{j} p^{*j}(1-p^*)^{\tau-j}(x(1+u)^j(1+d)^{\tau-j} - K)^+.$$

Then $(1+r)^{-\tau}C(\tau, x)$ is value of the call at time t (with time to expiry τ) given that $S(t) = x$.

Corollary 3.4.5. *The perfect hedging strategy $\varphi = (\varphi_0, \varphi_1)$ replicating the European call option with time of expiry T and strike price K is given by*

$$\varphi_1(t) = \frac{(1+r)^{-\tau}\left(C(\tau, S_{t-1}(1+u)) - C(\tau, S_{t-1}(1+d))\right)}{S_{t-1}(u-d)},$$

$$\varphi_0(t) = \frac{(1+u)C(\tau, S_{t-1}(1+d)) - (1+d)C(\tau, S_{t-1}(1+u))}{(u-d)(1+r)^T}.$$

Notice that the numerator in the equation for $\varphi_1(t)$ is the difference of two values of $C(\tau, x)$, with the larger value of x in the first term (recall $u > d$). When the payoff function $C(\tau, x)$ is an increasing function of x, as for the European call option considered here, this is non-negative. In this case, the Proposition gives $\varphi_1(t) \geq 0$: the replicating strategy does not involve short-selling. We record this as:

Corollary 3.4.6. *When the payoff function is a non-decreasing function of the asset price $S(t)$, the perfect-hedging strategy replicating the claim does not involve short-selling of the risky asset.*

If we do not use the pricing formula from Proposition 3.4.3 (i.e. the information on the price process), but only the final values of the option (or more generally of a contingent claim) we are still able to compute the arbitrage price and to construct the hedging portfolio by backward induction. In essence this is again only applying the one-period calculations for each time interval and each state of the world. We outline this procedure for the European call starting with the last period $[T-1, T]$. We have to choose a replicating portfolio $\varphi(T) = (\varphi_0(T), \varphi_1(T))$ based on the information available at time $T-1$ (and so $\mathcal{F}_{T-1}$-measurable). So for each $\omega \in \Omega$ the following equation has to hold:

$$\pi_X(T, \omega) = \varphi_0(T, \omega)B(T, \omega) + \varphi_1(T, \omega)S(T, \omega).$$

Given the information $\mathcal{F}_{T-1}$ we know all but the last coordinate of ω, and this gives rise to two equations (with the same notation as above):

$$\pi_X(T, S_{T-1}(1+u)) = \varphi_0(T)(1+r)^T + \varphi_1(T)S_{T-1}(1+u),$$

$$\pi_X(T, S_{T-1}(1+d)) = \varphi_0(T)(1+r)^T + \varphi_1(T)S_{T-1}(1+d).$$

Since we know the payoff structure of the contingent claim time T, for example in case of a European call. $\pi_X(T, S_{T-1}(1+u)) = ((1+u)S_{T-1} - K)^+$ and $\pi_X(T, S_{T-1}(1+d)) = ((1+d)S_{T-1} - K)^+$, we can solve the above system and obtain

$$\varphi_0(T) = \frac{(1+u)\Pi_X(T, S_{T-1}(1+d)) - (1+d)\Pi_X(T, S_{T-1}(1+u))}{(u-d)(1+r)}$$

$$\varphi_1(t) = \frac{\Pi_X(T, S_{T-1}(1+u)) - \Pi_X(T, S_{T-1}(1+d))}{S_{T-1}(u-d)}.$$

Using this portfolio one can compute the arbitrage price of the contingent claim at time $T-1$ given that the current asset price is S_{T-1} as

$$\pi_X(T-1, S_{T-1}) = \varphi_0(T, S_{T-1})(1+r)^{T-1} + \varphi_1(T, S_{T-1})S(T-1).$$

Now the arbitrage prices at time $T-1$ are known and one can repeat the procedure to successively compute the prices at $T-2, \ldots, 1, 0$.

The advantage of our risk-neutral pricing procedure over this approach is that we have a single formula for the price of the contingent claim at all times t at once, and don't have to go to a backwards induction only to compute a price at a special time t.

3.5   Binomial Approximations

Suppose we observe financial assets during a continuous time period $[0, T]$. To construct a stochastic model of the price processes of these assets (to, e.g. value contingent claims) one basically has two choices: one could model the processes as continuous-time stochastic processes (for which the theory of stochastic calculus is needed) or one could construct a sequence of discrete-time models in which the continuous-time price processes are approximated by discrete-time stochastic processes in a suitable sense. We describe the the second approach now by examining the asymptotic properties of a sequence of Cox-Ross-Rubinstein models.

3.5.1   Model Structure

We assume that all random variables subsequently introduced are defined on a suitable probability space $(\Omega, \mathcal{F}, I\!P)$. We want to model two assets, a riskless bond B and a risky stock S, which we now observe in a continuous-time interval $[0, T]$. To transfer the continuous-time framework into a binomial structure we make the following adjustments. Looking at the nth Cox-Ross-Rubinstein model in our sequence, there is a prespecified number k_n of trading dates. We set $\Delta_n = T/k_n$ and divide $[0, T]$ in k_n subintervals of length Δ_n, namely $I_j = [j\Delta_n, (j+1)\Delta_n]$, $j = 0, \ldots, k_n - 1$. We suppose that trading occurs only at the equidistant time points $t_{n,j} = j\Delta_n$, $j = 0, \ldots, k_n - 1$. We fix r_n as the riskless interest rate over each interval I_j, and hence the bond process (in the nth model) is given by

$$B(t_{n,j}) = (1 + r_n)^j, \quad j = 0, \ldots, k_n.$$

In the continuous-time model we compound continuously with spot rate $r \geq 0$ and hence the bond price process $B(t)$ is given by $B(t) = e^{rt}$. In order to approximate this process in the discrete-time framework, we choose r_n such that

$$1 + r_n = e^{r\Delta_n}. \tag{3.13}$$

With this choice we have for any $j = 0, \ldots, k_n$ that $(1 + r_n)^j = \exp(rj\Delta_n) = \exp(rt_{n,j})$. Thus we have approximated the bond process exactly at the time points of the discrete model.

Next we model the one-period returns $S(t_{n,j+1})/S(t_{n,j})$ of the stock by a family of random variables $Z_{n,i}; i = 1, \ldots, k_n$ taking values $\{d_n, u_n\}$ with

$$I\!P(Z_{n,i} = u_n) = p_n = 1 - I\!P(Z_{n,i} = d_n)$$

for some $p_n \in (0, 1)$ (which we specify later). With these $Z_{n,j}$ we model the stock price process S_n in the nth Cox-Ross-Rubinstein model as

$$S_n(t_{n,j}) = S_n(0) \prod_{i=1}^{j} (1 + Z_{n,i}), \quad j = 0, 1, \ldots, k_n.$$

With the specification of the one-period returns we get a complete description of the discrete dynamics of the stock price process in each Cox-Ross-Rubinstein model. We call such a finite sequence $Z_n = (Z_{n,i})_{i=1}^{k_n}$ a *lattice* or *tree*. The parameters u_n, d_n, p_n, k_n differ from lattice to lattice, but remain constant throughout a specific lattice. In the triangular array $(Z_{n,i})$, $i = 1, \ldots, k_n$; $n = 1, 2, \ldots$ we assume that the random variables are row-wise independent (but we allow dependence between rows). The approximation of a continuous-time setting by a sequence of lattices is called the lattice approach.

It is important to stress that for each n we get a different discrete stock price process $S_n(t)$ and that in general these processes do not coincide on common time points (and are also different from the price process $S(t)$).

Turning back to a specific Cox-Ross-Rubinstein model, we now have as in §3.4 a discrete-time bond and stock price process. We want arbitrage-free financial market models and therefore have to choose the parameters u_n, d_n, p_n accordingly. An arbitrage-free financial market model is

guaranteed by the existence of an equivalent martingale measure, and by Proposition 3.4.1 (i) the (necessary and) sufficient condition for that is

$$d_n < r_n < u_n.$$

The risk-neutrality approach implies that the expected (under an equivalent martingale measure) one-period return must equal the one-period return of the riskless bond and hence we get (see Proposition 3.4.1(ii))

$$p_n^* = \frac{r_n - d_n}{u_n - d_n}. \tag{3.14}$$

So the only parameters to choose freely in the model are u_n and d_n. In the next sections we consider some special choices.

3.5.2 The Black-Scholes Option Pricing Formula

We now choose the parameters in the above lattice approach in a special way. Assuming the risk-free rate of interest r as given, we have by (3.13) $1 + r_n = e^{r \Delta_n}$ and the remaining degrees of freedom are resolved by choosing u_n and d_n. We use the following choice:

$$1 + u_n = e^{\sigma \sqrt{\Delta_n}}, \quad \text{and} \quad 1 + d_n = (1 + u_n)^{-1} = e^{-\sigma \sqrt{\Delta_n}}.$$

By condition (3.14) the risk-neutral probabilities for the corresponding single period models are given by

$$p_n^* = \frac{r_n - d_n}{u_n - d_n} = \frac{e^{r \Delta_n} - e^{-\sigma \sqrt{\Delta_n}}}{e^{\sigma \sqrt{\Delta_n}} - e^{-\sigma \sqrt{\Delta_n}}}.$$

We can now price contingent claims in each Cox-Ross-Rubinstein model using the expectation operator with respect to the (unique) equivalent martingale measure characterised by the probabilities p_n^* (compare §3.4.2). In particular we can compute the price $\Pi_C(t)$ at time t of a European call on the stock S with strike K and expiry T by formula (3.12) of Corollary 3.4.4. Let us reformulate this formula slightly. We define

$$a_n = \min \left\{ j \in I\!N_0 | S(0)(1 + u_n)^j (1 + d_n)^{k_n - j} > K \right\}. \tag{3.15}$$

Then we can rewrite the pricing formula (3.12) for $t = 0$ in the setting of the nth Cox-Ross-Rubinstein model as

$$
\begin{aligned}
\Pi_C(0) &= (1 + r_n)^{-k_n} \sum_{j=a_n}^{k_n} \binom{k_n}{j} p_n^{*\,j} (1 - p_n^*)^{k_n - j} (S(0)(1 + u_n)^j (1 + d_n)^{k_n - j} - K) \\
&= S(0) \left[\sum_{j=a_n}^{k_n} \binom{k_n}{j} \left(\frac{p_n^*(1 + u_n)}{1 + r_n} \right)^j \left(\frac{(1 - p_n^*)(1 + d_n)}{1 + r_n} \right)^{k_n - j} \right] \\
&\quad - (1 + r_n)^{-k_n} K \left[\sum_{j=a_n}^{k_n} \binom{k_n}{j} p_n^{*\,j} (1 - p_n^*)^{k_n - j} \right].
\end{aligned}
$$

Denoting the binomial cumulative distribution function with parameters (n, p) as $B^{n,p}(.)$ we see that the second bracketed expression is just

$$\bar{B}^{k_n, p_n^*}(a_n) = 1 - B^{k_n, p_n^*}(a_n).$$

Also the first bracketed expression is $\bar{B}^{k_n, \hat{p}_n}(a_n)$ with

$$\hat{p}_n = \frac{p_n^*(1 + u_n)}{1 + r_n}.$$

That $\hat{p}_n$ is indeed a probability can be shown straightforwardly. Using this notation we have in the nth Cox-Ross-Rubinstein model for the price of a European call at time $t = 0$ the following formula:

$$\Pi_C^{(n)}(0) = S_n(0)\bar{B}^{k_n,\hat{p}_n}(a_n) - K(1 + r_n)^{-k_n}\bar{B}^{k_n,p_n^*}(a_n). \tag{3.16}$$

(We stress again that the underlying is $S_n(t)$, dependent on n, but $S_n(0) = S(0)$ for all n.) We now look at the limit of this expression.

Proposition 3.5.1. *We have the following limit relation:*

$$\lim_{n\to\infty} \Pi_C^{(n)}(0) = \Pi_C^{BS}(0)$$

with $\Pi_C^{BS}(0)$ given by the Black-Scholes formula (we use $S = S(0)$ to ease the notation)

$$\Pi_C^{BS}(0) = SN(d_1(S,T)) - Ke^{-rT}N(d_2(S,T)). \tag{3.17}$$

The functions $d_1(s,t)$ and $d_2(s,t)$ are given by

$$d_1(s,t) = \frac{\log(s/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}},$$

$$d_2(s,t) = d_1(s,t) - \sigma\sqrt{t} = \frac{\log(s/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$$

and $N(.)$ is the standard normal cumulative distribution function.

Proof. Since $S_n(0) = S$ (say) all we have to do to prove the proposition is to show

$$
\begin{array}{lll}
\text{(i)} & \lim_{n\to\infty} \bar{B}^{k_n,\hat{p}_n}(a_n) & = & N(d_1(S,T)), \\
\text{(ii)} & \lim_{n\to\infty} \bar{B}^{k_n,p_n^*}(a_n) & = & N(d_2(S,T)).
\end{array}
$$

These statements involve the convergence of distribution functions.

To show (i) we interpret

$$\bar{B}^{k_n,\hat{p}_n}(a_n) = I\!P(a_n \leq Y_n \leq k_n)$$

with (Y_n) a sequence of random variables distributed according to the binomial law with parameters $(k_n, \hat{p}_n)$. We normalise Y_n to

$$\tilde{Y}_n = \frac{Y_n - I\!E(Y_n)}{\sqrt{Var(Y_n)}} = \frac{Y_n - k_n\hat{p}_n}{\sqrt{k_n\hat{p}_n(1-\hat{p}_n)}} = \frac{\sum_{j=1}^{k_n}(B_{j,n} - \hat{p}_n)}{\sqrt{k_n\hat{p}_n(1-\hat{p}_n)}},$$

where $B_{j,n}$, $j = 1, \ldots, k_n$; $n = 1, 2, \ldots$ are row-wise independent Bernoulli random variables with parameter $\hat{p}_n$. Now using the central limit theorem we know that for $\alpha_n \to \alpha$, $\beta_n \to \beta$ we have

$$\lim_{n\to\infty} I\!P(\alpha_n \leq \tilde{Y}_n \leq \beta_n) = N(\beta) - N(\alpha).$$

By definition we have

$$I\!P(a_n \leq Y_n \leq k_n) = I\!P\left(\alpha_n \leq \tilde{Y}_n \leq \beta_n\right)$$

with

$$\alpha_n = \frac{a_n - k_n\hat{p}_n}{\sqrt{k_n\hat{p}_n(1-\hat{p}_n)}} \quad \text{and} \quad \beta_n = \frac{k_n(1-\hat{p}_n)}{\sqrt{k_n\hat{p}_n(1-\hat{p}_n)}}.$$

Using the following limiting relations:

$$\lim_{n\to\infty} \hat{p}_n = \frac{1}{2}, \quad \lim_{n\to\infty} k_n(1 - 2\hat{p}_n)\sqrt{\Delta_n} = -T\left(\frac{r}{\sigma} + \frac{\sigma}{2}\right),$$

and the defining relation for a_n, formula (3.15), we get

$$
\begin{aligned}
\lim_{n\to\infty} \alpha_n &= \lim_{n\to\infty} \frac{\dfrac{\log(K/S) + k_n\sigma\sqrt{\Delta_n}}{2\sigma\sqrt{\Delta_n}} - k_n\hat{p}_n}{\sqrt{k_n\hat{p}_n(1-\hat{p}_n)}} \\[2mm]
&= \lim_{n\to\infty} \frac{\log(K/S) + \sigma k_n\sqrt{\Delta_n}(1-2\hat{p}_n)}{2\sigma\sqrt{k_n\Delta_n\hat{p}_n(1-\hat{p}_n)}} \\[2mm]
&= \frac{\log(K/S) - (r + \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} = -d_1(S,T).
\end{aligned}
$$

Furthermore we have

$$
\lim_{n\to\infty} \beta_n = \lim_{n\to\infty} \sqrt{k_n\hat{p}_n^{-1}(1-\hat{p}_n)} = +\infty.
$$

So $N(\beta_n) \to 1, N(\alpha_n) \to N(-d_1) = 1 - N(d_1)$, completing the proof of (i).

To prove (ii) we can argue in very much the same way and arrive at parameters α_n^* and β_n^* with $\hat{p}_n$ replaced by p_n^*. Using the following limiting relations:

$$
\lim_{n\to\infty} p_n^* = \frac{1}{2}, \quad \lim_{n\to\infty} k_n(1-2p_n^*)\sqrt{\Delta_n} = T\left(\frac{\sigma}{2} - \frac{r}{\sigma}\right),
$$

we get

$$
\begin{aligned}
\lim_{n\to\infty} \alpha_n^* &= \lim_{n\to\infty} \frac{\log(K/S) + \sigma n\sqrt{\Delta_n}(1-2p_n^*)}{2\sigma\sqrt{n\Delta_n p_n^*(1-p_n^*)}} \\[2mm]
&= \frac{\log(K/S) - (r - \frac{\sigma^2}{2})T}{\sigma\sqrt{T}} = -d_2(s,T).
\end{aligned}
$$

For the upper limit we get

$$
\lim_{n\to\infty} \beta_n^* = \lim_{n\to\infty} \sqrt{k_n(p_n^*)^{-1}(1-p_n^*)} = +\infty,
$$

whence (ii) follows similarly. $\blacksquare$

By the above proposition we have derived the classical Black-Scholes European call option valuation formula as an asymptotic limit of option prices in a sequence of Cox-Ross-Rubinstein type models with a special choice of parameters. We will therefore call these models discrete Black-Scholes models. Let us mention here that in the continuous-time Black-Scholes model the dynamics of the (stochastic) stock price process $S(t)$ are modelled by a geometric Brownian motion (or exponential Wiener process). The sample paths of this stochastic price process are almost all continuous and the probability law of $S(t)$ at any time t is lognormal. In particular the time T distribution of $\log\{S(T)/S(0)\}$ is $N(T\mu, T\sigma^2)$. Looking back at the construction of our sequence of Cox-Ross-Rubinstein models we see that

$$
\log\frac{S_n(T)}{S(0)} = \sum_{i=1}^{k_n} \log(1 + Z_{n,i}),
$$

with $\log(Z_{n,i})$ Bernoulli random variables with

$$
I\!P(\log(1 + Z_{n,i}) = \sigma\sqrt{\Delta_n}) = p_n = 1 - I\!P(\log(1 + Z_{n,i}) = -\sigma\sqrt{\Delta_n}).
$$

By the (triangular array version) of the central limit theorem we know that $\log\frac{S_n(T)}{S(0)}$ properly normalised converges in distribution to a random variable with standard normal distribution. Doing similar calculations as in the above proposition we can compute the normalising constants and get

$$
\lim_{n\to\infty} \log\frac{S_n(T)}{S(0)} \sim N(T(r - \sigma^2/2), T\sigma^2),
$$

i.e. $\frac{S_n(T)}{S(0)}$ is in the limit lognormally distributed.

3.6 American Options

Consider a general multi-period framework. The holder of an American derivative security can 'exercise' in any period t and receive payment $f(S_t)$ (or more general a non-negative payment f_t). In order to hedge such an option, we want to construct a self-financing trading strategy φ_t such that for the corresponding value process $V_\varphi(t)$

$$\begin{aligned} V_\varphi(0) &= x \quad \text{initial capital} \\ V_\varphi(t) &\geq f(S_t), \quad \forall t. \end{aligned} \tag{3.18}$$

Such a hedging portfolio is minimal, if for a stopping time τ

$$V_\varphi(\tau) = f(S_\tau).$$

Our aim in the following will be to discuss existence and construction of such a stopping time.

3.6.1 Stopping Times, Optional Stopping and Snell Envelopes

A random variable τ taking values in $\{0, 1, 2, \ldots; +\infty\}$ is called a *stopping time* (or optional time) if

$$\{\tau \leq n\} = \{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n \quad \forall \, n \leq \infty.$$

From $\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n - 1\}$ and $\{\tau \leq n\} = \bigcup_{k \leq n}\{\tau = k\}$, we see the equivalent characterisation

$$\{\tau = n\} \in \mathcal{F}_n \quad \forall \, n \leq \infty.$$

Call a stopping time τ bounded, if there is a constant K such that $I\!P(\tau \leq K) = 1$. (Since $\tau(\omega) \leq K$ for some constant K and all $\omega \in \Omega \setminus N$ with $I\!P(N) = 0$ all identities hold true except on a null set, i.e. almost surely.)

Example. Suppose (X_n) is an adapted process and we are interested in the time of first entry of X into a Borel set B (typically one might have $B = [c, \infty)$):

$$\tau = \inf\{n \geq 0 : X_n \in B\}.$$

Now $\{\tau \leq n\} = \bigcup_{k \leq n}\{X_k \in B\} \in \mathcal{F}_n$ and $\tau = \infty$ if X never enters B. Thus τ is a stopping time.

Intuitively, think of τ as a time at which you decide to quit a gambling game: whether or not you quit at time n depends only on the history up to and including time n – NOT the future. Thus stopping times model gambling and other situations where there is no foreknowledge, or prescience of the future; in particular, in the financial context, where there is no insider trading. Furthermore since a gambler cannot cheat the system the expectation of his hypothetical fortune (playing with unit stake) should equal his initial fortune.

Theorem 3.6.1 (Doob's Stopping-Time Principle (STP)). *Let τ be a bounded stopping time and $X = (X_n)$ a martingale. Then X_τ is integrable, and*

$$I\!E(X_\tau) = I\!E(X_0).$$

Proof. Assume $\tau(\omega) \leq K$ for all ω, where we can take K to be an integer and write

$$X_{\tau(\omega)}(\omega) = \sum_{k=0}^{\infty} X_k(\omega)\mathbf{1}_{\{\tau(\omega)=k\}} = \sum_{k=0}^{K} X_k(\omega)\mathbf{1}_{\{\tau(\omega)=k\}}$$

Thus using successively the linearity of the expectation operator, the martingale property of X, the $\mathcal{F}_k$-measurability of $\{\tau = k\}$ and finally the definition of conditional expectation, we get

$$
\begin{aligned}
I\!\!E(X_\tau) &= I\!\!E\left[\sum_{k=0}^{K} X_k \mathbf{1}_{\{\tau=k\}}\right] = \sum_{k=0}^{K} I\!\!E\left[X_k \mathbf{1}_{\{\tau=k\}}\right] \\
&= \sum_{k=0}^{K} I\!\!E\left[I\!\!E(X_K|\mathcal{F}_k)\mathbf{1}_{\{\tau=k\}}\right] = \sum_{k=0}^{K} I\!\!E\left[X_K \mathbf{1}_{\{\tau=k\}}\right] \\
&= I\!\!E\left[X_K \sum_{k=0}^{K} \mathbf{1}_{\{\tau=k\}}\right] = I\!\!E(X_K) = I\!\!E(X_0).
\end{aligned}
$$

∎

The stopping time principle holds also true if $X = (X_n)$ is a supermartingale; then the conclusion is

$$ I\!\!E X_\tau \leq I\!\!E X_0. $$

Also, alternative conditions such as
(i) $X = (X_n)$ is bounded ($|X_n(\omega)| \leq L$ for some L and all n, ω);
(ii) $I\!\!E\tau < \infty$ and $(X_n - X_{n-1})$ is bounded;
suffice for the proof of the stopping time principle.

The stopping time principle is important in many areas, such as sequential analysis in statistics. We turn in the next section to related ideas specific to the gambling/financial context.

We now wish to create the concept of the σ-algebra of events observable up to a stopping time τ, in analogy to the σ-algebra $\mathcal{F}_n$ which represents the events observable up to time n.

Definition 3.6.1. *Let τ be a stopping time. The stopping time $\sigma-$algebra $\mathcal{F}_\tau$ is defined to be*

$$ \mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n, \text{ for all } n\}. $$

Proposition 3.6.1. *For τ a stopping time, $\mathcal{F}_\tau$ is a $\sigma-$algebra.*

Proof. We simply have to check the defining properties. Clearly $\Omega, \emptyset$ are in $\mathcal{F}_\tau$. Also for $A \in \mathcal{F}_\tau$ we find

$$ A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus (A \cap \{\tau \leq n\}) \in \mathcal{F}_n, $$

thus $A^c \in \mathcal{F}_\tau$. Finally, for a family $A_i \in \mathcal{F}_\tau$, $i = 1, 2, \ldots$ we have

$$ \left(\bigcup_{i=1}^{\infty} A_i\right) \cap \{\tau \leq n\} = \bigcup_{i=1}^{\infty} (A_i \cap \{\tau \leq n\}) \in \mathcal{F}_n, $$

showing $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\tau$. ∎

Proposition 3.6.2. *Let σ, τ be stopping times with $\sigma \leq \tau$. Then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.*

Proof. Since $\sigma \leq \tau$ we have $\{\tau \leq n\} \subseteq \{\sigma \leq n\}$. So for $A \in \mathcal{F}_\sigma$ we get

$$ A \cap \{\tau \leq n\} = (A \cap \{\sigma \leq n\}) \cap \{\tau \leq n\} \in \mathcal{F}_n, $$

since $(.) \in \mathcal{F}_n$ as $A \in \mathcal{F}_\sigma$. So $A \in \mathcal{F}_\tau$. ∎

Proposition 3.6.3. *For any adapted sequence of random variables $X = (X_n)$ and a.s. finite stopping time τ, define*

$$ X_\tau = \sum_{n=0}^{\infty} X_n \mathbf{1}_{\{\tau=n\}}. $$

Then X_τ is $\mathcal{F}_\tau$-measurable.

Proof. Let B be a Borel set. We need to show $\{X_\tau \in B\} \in \mathcal{F}_\tau$. Now using the fact that on the set $\{\tau = k\}$ we have $X_\tau = X_k$, we find

$$\{X_\tau \in B\} \cap \{\tau \le n\} = \bigcup_{k=1}^{n} \{X_\tau \in B\} \cap \{\tau = k\} = \bigcup_{k=1}^{n} \{X_k \in B\} \cap \{\tau = k\}.$$

Now sets $\{X_k \in B\} \cap \{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$, and the result follows. ∎

We are now in position to obtain an important extension of the Stopping-Time Principle, Theorem 3.6.1.

Theorem 3.6.2 (Doob's Optional-Sampling Theorem, OST). *Let $X = (X_n)$ be a martingale and let σ, τ be bounded stopping times with $\sigma \le \tau$. Then*

$$\mathbb{E}\left[X_\tau | \mathcal{F}_\sigma\right] = X_\sigma$$

and thus $\mathbb{E}(X_\tau) = \mathbb{E}(X_\sigma)$.

Proof. First observe that X_τ and X_σ are integrable (use the sum representation and the fact that τ is bounded by an integer K) and X_σ is $\mathcal{F}_\sigma$-measurable by Proposition 3.6.3. So it only remains to prove that

$$\mathbb{E}(\mathbf{1}_A X_\tau) = \mathbb{E}(\mathbf{1}_A X_\sigma) \quad \forall A \in \mathcal{F}_\sigma. \tag{3.19}$$

For any such fixed $A \in \mathcal{F}_\sigma$, define ρ by

$$\rho(\omega) = \sigma(\omega)\mathbf{1}_A(\omega) + \tau(\omega)\mathbf{1}_{A^c}(\omega).$$

Since

$$\{\rho \le n\} = (A \cap \{\sigma \le n\}) \cup (A^c \cap \{\tau \le n\}) \in \mathcal{F}_n$$

ρ is a stopping time, and from $\rho \le \tau$ we see that ρ is bounded. So the STP (Theorem 3.6.1) implies $\mathbb{E}(X_\rho) = \mathbb{E}(X_0) = \mathbb{E}(X_\tau)$. But

$$\mathbb{E}(X_\rho) = \mathbb{E}\left(X_\sigma \mathbf{1}_A + X_\tau \mathbf{1}_{A^c}\right),$$

$$\mathbb{E}(X_\tau) = \mathbb{E}\left(X_\tau \mathbf{1}_A + X_\tau \mathbf{1}_{A^c}\right).$$

So subtracting yields (3.19). ∎

We can establish a further characterisation of the martingale property.

Proposition 3.6.4. *Let $X = (X_n)$ be an adapted sequence of random variables with $\mathbb{E}(|X_n|) < \infty$ for all n and $\mathbb{E}(X_\tau) = 0$ for all bounded stopping times τ. Then X is a martingale.*

Proof. Let $0 \le m < n, \infty$ and $A \in \mathcal{F}_m$. Define a stopping time τ by $\tau = n\mathbf{1}_A + m\mathbf{1}_{A^c}$. Then

$$0 = \mathbb{E}(X_\tau) = \mathbb{E}\left(X_n \mathbf{1}_A + X_m \mathbf{1}_{A^c}\right),$$

$$0 = \mathbb{E}(X_m) = \mathbb{E}\left(X_m \mathbf{1}_A + X_m \mathbf{1}_{A^c}\right),$$

and by subtraction we obtain $\mathbb{E}(X_m \mathbf{1}_A) = \mathbb{E}(X_n \mathbf{1}_A)$. Since this holds for all $A \in \mathcal{F}_m$, $\mathbb{E}(X_n | \mathcal{F}_m) = X_m$ by definition of conditional expectation. This says that (X_n) is a martingale, as required. ∎

Write $X^\tau = (X_n^\tau)$ for the sequence $X = (X_n)$ stopped at time τ, where we define $X_n^\tau(\omega) := X_{\tau(\omega) \wedge n}(\omega)$.

Proposition 3.6.5. *(i) If X is adapted and τ is a stopping time, then the stopped sequence X^τ is adapted.*
(ii) If X is a martingale (super-, submartingale) and τ is a stopping time, X^τ is a martingale (super-, submartingale).

Proof. Let $C_j := \mathbf{1}_{\{j \leq \tau\}}$; then

$$X_{\tau \wedge n} = X_0 + \sum_{j=1}^{n} C_j (X_j - X_{j-1})$$

(as the right is $X_0 + \sum_{j=1}^{\tau \wedge n}(X_j - X_{j-1})$, which telescopes to $X_{\tau \wedge n}$). Since $\{j \leq \tau\}$ is the complement of $\{\tau < j\} = \{\tau \leq j - 1\} \in \mathcal{F}_{j-1}$, (C_n) is predictable. So (X_n^τ) is adapted.

If X is a martingale, so is X^τ as it is the martingale transform of (X_n) by (C_n) (use Theorem C.4.1). The right-hand side above is $X_{\tau \wedge (n-1)} + C_n(X_n - X_{n-1})$. So taking conditional expectation given $\mathcal{F}_{n-1}$ and using predictability of (C_n),

$$\mathbb{E}(X_{\tau \wedge n}|\mathcal{F}_{n-1}) = X_{\tau \wedge (n-1)} + C_n(\mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}).$$

Then $C_n \geq 0$ shows that if X is a supermartingale (submartingale), so is X^τ. ∎

We now discuss the Snell envelope, which will be an important tool for the valuation of American options. The idea is due to Snell (1952); for a textbook account, see e.g. Neveu (1975), VI.

Definition 3.6.2. *If $X = (X_n)_{n=0}^N$ is a sequence adapted to a filtration $\mathcal{F}_n$ with $\mathbb{E}(|X_n|) < \infty$, the sequence $Z = (Z_n)_{n=0}^N$ defined by*

$$\begin{cases} Z_N := X_N, \\ Z_n := \max\{X_n, \mathbb{E}(Z_{n+1}|\mathcal{F}_n)\} & (n \leq N - 1) \end{cases}$$

is called the Snell envelope of X.

Theorem 3.6.3. *The Snell envelope Z of X is a supermartingale, and is the smallest supermartingale dominating X (that is, with $Z_n \geq X_n$ for all n).*

Proof. First, $Z_n \geq \mathbb{E}(Z_{n+1}|\mathcal{F}_n)$, so Z is a supermartingale, and $Z_n \geq X_n$, so Z dominates X.

Next, let $Y = (Y_n)$ be any other supermartingale dominating X; we must show Y dominates Z also. First, since $Z_N = X_N$ and Y dominates X, we must have $Y_N \geq Z_N$. Assume inductively that $Y_n \geq Z_n$. Then as Y is a supermartingale,

$$Y_{n-1} \geq \mathbb{E}(Y_n|\mathcal{F}_{n-1}) \geq \mathbb{E}(Z_n|\mathcal{F}_{n-1}),$$

and as Y dominates X,

$$Y_{n-1} \geq X_{n-1}.$$

Combining,

$$Y_{n-1} \geq \max\{X_{n-1}, \mathbb{E}(Z_n|\mathcal{F}_{n-1})\} = Z_{n-1}.$$

By repeating this argument (or more formally, by backward induction), $Y_n \geq Z_n$ for all n, as required. ∎

Proposition 3.6.6. $\tau^* := \inf\{n \geq 0 : Z_n = X_n\}$ *is a stopping time, and the stopped process Z^{τ^*} is a martingale.*

Proof. Since $Z_N = X_N$, $\tau^* \in \{0, 1, \ldots, N\}$ is well-defined and clearly bounded. For $k = 0$, $\{\tau^* = 0\} = \{Z_0 = X_0\} \in \mathcal{F}_0$; for $k \geq 1$,

$$\{\tau^* = k\} = \{Z_0 > X_0\} \cap \cdots \cap \{Z_{k-1} > X_{k-1}\} \cap \{Z_k = X_k\} \in \mathcal{F}_k.$$

So τ^* is a stopping time.

As in the proof of Proposition 3.6.5,

$$Z_n^{\tau^*} = Z_{n \wedge \tau^*} = Z_0 + \sum_{j=1}^{n} C_j \Delta Z_j,$$

where $C_j = \mathbf{1}_{\{j \leq \tau^*\}}$ is predictable. For $n \leq N-1$,

$$Z_{n+1}^{\tau^*} - Z_n^{\tau^*} = C_{n+1}(Z_{n+1} - Z_n) = \mathbf{1}_{\{n+1 \leq \tau^*\}}(Z_{n+1} - Z_n).$$

Now $Z_n := \max\{X_n, E(Z_{n+1}|\mathcal{F}_n)\}$, and by definition of τ^*,

$$Z_n > X_n \quad \text{on} \quad \{n+1 \leq \tau^*\}.$$

So from the definition of Z_n,

$$Z_n = E(Z_{n+1}|\mathcal{F}_n) \quad \text{on} \quad \{n+1 \leq \tau^*\}.$$

We next prove

$$Z_{n+1}^{\tau^*} - Z_n^{\tau^*} = \mathbf{1}_{\{n+1 \leq \tau^*\}}(Z_{n+1} - E(Z_{n+1}|\mathcal{F}_n)). \tag{3.20}$$

For, suppose first that $\tau^* \geq n+1$. Then the left of (3.20) is $Z_{n+1} - Z_n$, the right is $Z_{n+1} - E(Z_{n+1}|\mathcal{F}_n)$, and these agree on $\{n+1 \leq \tau^*\}$ by above. The other possibility is that $\tau^* < n+1$, i.e. $\tau^* \leq n$. Then the left of (3.20) is $Z_{\tau^*} - Z_{\tau^*} = 0$, while the right is zero because the indicator is zero, completing the proof of (3.20). Now apply $E(.|\mathcal{F}_n)$ to (3.20): since $\{n+1 \leq \tau^*\} = \{\tau^* \leq n\}^c \in \mathcal{F}_n$,

$$E\left[(Z_{n+1}^{\tau^*} - Z_n^{\tau^*})|\mathcal{F}_n\right] = \mathbf{1}_{\{n+1 \leq \tau^*\}} E\left[(Z_{n+1} - E(Z_{n+1}|\mathcal{F}_n))|\mathcal{F}_n\right]$$

$$= \mathbf{1}_{\{n+1 \leq \tau^*\}}\left[E(Z_{n+1}|\mathcal{F}_n) - E(Z_{n+1}|\mathcal{F}_n)\right] = 0.$$

So $E(Z_{n+1}^{\tau^*}|\mathcal{F}_n) = Z_n^{\tau^*}$. This says that Z^{τ^*} is a martingale, as required. ∎

Write $\mathcal{T}_{n,N}$ for the set of stopping times taking values in $\{n, n+1, \dots, N\}$ (a finite set, as Ω is finite). Call a stopping time $\sigma \in \mathcal{T}_{n,N}$ *optimal* for (X_n) if

$$E(X_\sigma|\mathcal{F}_n) = \sup\{E(X_\tau|\mathcal{F}_n) : \tau \in \mathcal{T}_{n,N}\}.$$

We next see that the Snell envelope can be used to solve the optimal stopping problem for (X_n) in $\mathcal{T}_{0,N}$. Recall that $\mathcal{F}_0 = \{\emptyset, \Omega\}$ so $E(Y|\mathcal{F}_0) = E(Y)$ for any integrable random variable Y.

Proposition 3.6.7. τ^* *solves the optimal stopping problem for X:*

$$Z_0 = E(X_{\tau^*}) = \sup\{E(X_\tau) : \tau \in \mathcal{T}_{0,N}\}.$$

Proof. To prove the first statement we use that $(Z_n^{\tau^*})$ is a martingale and $Z_{\tau^*} = X_{\tau^*}$; then

$$Z_0 = Z_0^{\tau^*} = E\left(Z_N^{\tau^*}\right) = E(Z_{\tau^*}) = E(X_\tau^*). \tag{3.21}$$

Now for any stopping time $\tau \in \mathcal{T}_{0,N}$, since Z is a supermartingale (above), so is the stopped process Z^τ (see Proposition 3.6.5). Together with the property that Z dominates X this yields

$$Z_0 = Z_0^\tau \geq E(Z_N^\tau) = E(Z_\tau) \geq E(X_\tau). \tag{3.22}$$

Combining (3.21) and (3.22) and taking the supremum on τ gives the result. ∎

The same argument, starting at time n rather than time 0, gives

Corollary 3.6.1. *If $\tau_n^* := \inf\{j \geq n : Z_j = X_j\}$,*

$$Z_n = E(X_{\tau_n^*}|\mathcal{F}_n) = \sup\{E(X_\tau|\mathcal{F}_n) : \tau \in \mathcal{T}_{n,N}\}.$$

As we are attempting to maximise our payoff by stopping $X = (X_n)$ at the most advantageous time, the Corollary shows that τ_n^* gives the best stopping time that is realistic: it maximises our expected payoff given only information currently available.

We proceed by analysing optimal stopping times. One can characterize optimality by establishing a martingale property:

Proposition 3.6.8. *The stopping time $\sigma \in \mathcal{T}$ is optimal for (X_t) if and only if the following two conditions hold.*

(i) $Z_\sigma = X_\sigma$;

(ii) Z^σ is a martingale.

Proof. We start showing that (i) and (ii) imply optimality. If Z^σ is a martingale then

$$Z_0 = I\!E(Z_0^\sigma) = I\!E(Z_N^\sigma) = I\!E(Z_\sigma) = I\!E(X_\sigma),$$

where we used (i) for the last identity. Since Z is a supermartingale Proposition 3.6.5 implies that Z^τ is a supermartingale for any $\tau \in \mathcal{T}_{0,N}$. Now Z dominates X, and so

$$Z_0 = I\!E(Z_0^\tau) \geq I\!E(Z_N^\tau) = I\!E(Z_\tau) \geq I\!E(X_\tau).$$

Combining, σ is optimal.

Now assume that σ is optimal. Thus

$$Z_0 = \max\{I\!E(X_\tau);\ \tau \in \mathcal{T}_{0,N}\} = I\!E(X_\sigma) \leq I\!E(Z_\sigma),$$

since Z dominates X. Since Z^σ is a supermartingale, we also have $Z_0 \geq I\!E(Z_\sigma)$. Combining,

$$I\!E(X_\sigma) = Z_0 = I\!E(Z_\sigma).$$

But $X \leq Z$, so $X_\sigma \leq Z_\sigma$, while by the above X_σ and Z_σ have the same expectation. So they must be a.s. equal: $X_\sigma = Z_\sigma$ a.s., showing (i). To see (ii), observe that for any $n \leq N$

$$I\!E(Z_\sigma) = Z_0 \geq I\!E(Z_{\sigma \wedge n}) \geq I\!E(Z_\sigma) = I\!E(I\!E(Z_\sigma|\mathcal{F}_n)),$$

where the second inequality follows from Doob's OST (Theorem 3.6.2) with the bounded stopping times $(\sigma \wedge n) \leq \sigma$ and the supermartingale Z. Using that Z is a supermartingale again, we also find

$$Z_{\sigma \wedge n} \geq I\!E(Z_\sigma|\mathcal{F}_n). \tag{3.23}$$

As above, this inequality between random variables with equal expectations forces a.s. equality: $Z_{\sigma \wedge n} = I\!E(Z_\sigma|\mathcal{F}_n)$ a.s.. Apply $I\!E(.|\mathcal{F}_{n-1})$:

$$I\!E\left(Z_{\sigma \wedge n}|\mathcal{F}_{n-1}\right) = I\!E\left(I\!E(Z_\sigma|\mathcal{F}_n)|\mathcal{F}_{n-1}\right) = I\!E(Z_\sigma|\mathcal{F}_{n-1}) = Z_{\sigma \wedge (n-1)},$$

by (3.23) with $n-1$ for n. This says

$$I\!E\left(Z_n^\sigma|\mathcal{F}_{n-1}\right) = Z_{n-1}^\sigma,$$

so Z^σ is a martingale. ■

From Proposition 3.6.6 and its definition (first time when Z and X are equal) it follows that τ^* is the smallest optimal stopping time . To find the largest optimal stopping time we try to find the time when Z 'ceases to be a martingale'. In order to do so we need a structural result of genuine interest and importance

Theorem 3.6.4 (Doob Decomposition). *Let $X = (X_n)$ be an adapted process with each $X_n \in \mathcal{L}^1$. Then X has an (essentially unique) Doob decomposition*

$$X = X_0 + M + A : \quad X_n = X_0 + M_n + A_n \quad \forall n \tag{3.24}$$

with M a martingale null at zero, A a predictable process null at zero. If also X is a submartingale ('increasing on average'), A is increasing: $A_n \leq A_{n+1}$ for all n, a.s..

Proof. If X has a Doob decomposition (3.24),

$$I\!E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = I\!E[M_n - M_{n-1}|\mathcal{F}_{n-1}] + I\!E[A_n - A_{n-1}|\mathcal{F}_{n-1}].$$

The first term on the right is zero, as M is a martingale. The second is $A_n - A_{n-1}$, since A_n (and A_{n-1}) is $\mathcal{F}_{n-1}$-measurable by predictability. So

$$I\!E[X_n - X_{n-1}|\mathcal{F}_{n-1}] = A_n - A_{n-1}, \tag{3.25}$$

and summation gives

$$A_n = \sum_{k=1}^{n} I\!E[X_k - X_{k-1}|\mathcal{F}_{k-1}], \quad a.s.$$

So set $A_0 = 0$ and use this formula to *define* (A_n), clearly predictable. We then use (3.24) to *define* (M_n), then a martingale, giving the Doob decomposition (3.24). To see uniqueness, assume two decompositions, i.e. $X_n = X_0 + M_n + A_n = X_0 + \tilde{M}_n + \tilde{A}_n$, then $M_n - \tilde{M}_n = A_n - \tilde{A}_n$. Thus the martingale $M_n - \tilde{M}_n$ is predictable and so must be constant a.s..

If X is a submartingale, the LHS of (3.25) is ≥ 0, so the RHS of (3.25) is ≥ 0, i.e. (A_n) is increasing. ∎

Although the Doob decomposition is a simple result in discrete time, the analogue in continuous time – the Doob-Meyer decomposition – is deep. This illustrates the contrasts that may arise between the theories of stochastic processes in discrete and continuous time.

Equipped with the Doob-decomposition we return to the above setting and can write

$$Z = Z_0 + L + B$$

with L a martingale and B predictable and decreasing. Then $M = Z_0 + L$ is a martingale and $A = (-B)$ is increasing and we have $Z = M - A$.

Definition 3.6.3. *Define a random variable $\nu : \Omega \to I\!N_0$ by setting*

$$\nu(\omega) = \begin{cases} N & \text{if } A_N(\omega) = 0 \\ \min\{n \geq 0 : A_{n+1} > 0\} & \text{if } A_N(\omega) > 0. \end{cases}$$

Observe that ν (bounded by N) is a stopping time, since

$$\{\nu = n\} = \bigcup_{k \leq n} \{A_k = 0\} \cap \{A_{n+1} > 0\} \in \mathcal{F}_n$$

as A is predictable.

Proposition 3.6.9. *ν is optimal for (X_t), and it is the largest optimal stopping time for (X_t).*

Proof. We use Proposition 3.6.8. Since for $k \leq \nu(\omega), Z_k(\omega) = M_k(\omega) - A_k(\omega) = M_k(\omega)$, Z^ν is a martingale and thus we have (ii) of Proposition 3.6.8. To see (i) we write

$$\begin{aligned} Z_\nu &= \sum_{k=0}^{N-1} \mathbf{1}_{\{\nu=k\}} Z_k + \mathbf{1}_{\{\nu=N\}} Z_N \\ &= \sum_{k=0}^{N-1} \mathbf{1}_{\{\nu=k\}} \max\{X_k, I\!E(Z_{k+1}|\mathcal{F}_k)\} + \mathbf{1}_{\{\nu=N\}} X_N. \end{aligned}$$

Now $I\!E(Z_{k+1}|\mathcal{F}_k) = I\!E(M_{k+1} - A_{k+1}|\mathcal{F}_k) = M_k - A_{k+1}$. On $\{\nu = k\}$ we have $A_k = 0$ and $A_{k+1} > 0$, so $I\!E(Z_{k+1}|\mathcal{F}_k) < Z_k$. Hence $Z_k = \max\{X_k, I\!E(Z_{k+1}|\mathcal{F}_k)\} = X_k$ on the set $\{\nu = k\}$. So

$$Z_\nu = \sum_{k=0}^{N-1} \mathbf{1}_{\{\nu=k\}} X_k + \mathbf{1}_{\{\nu=N\}} X_N = X_\nu,$$

which is (i) of Proposition 3.6.8. Now take $\tau \in \{\mathcal{T}\}_{0,N}$ with $\tau \geq \nu$ and $I\!P(\tau > \nu) > 0$. From the definition of ν and the fact that A is increasing, $A_\tau > 0$ with positive probability. So $I\!E(A_\tau) > 0$, and

$$I\!E(Z_\tau) = I\!E(M_\tau) - I\!E(A_\tau) = I\!E(Z_0) - I\!E(A_\tau) < I\!E(Z_0).$$

So τ cannot be optimal. ∎

3.6.2 The Financial Model

We assume now that we work in a market model $(\Omega, \mathcal{F}, I\!F, I\!P)$, which is complete with $I\!P^*$ the unique martingale measure.

Then for any hedging strategy φ we have that under $I\!P^*$

$$M(t) = \tilde{V}_\varphi(t) = \beta(t) V_\varphi(t) \tag{3.26}$$

is a martingale. Thus we can use the STP (Theorem 3.6.1) to find for any stopping time τ

$$V_\varphi(0) = M_0 = I\!E^*(\tilde{V}_\varphi(\tau)). \tag{3.27}$$

Since we require $V_\varphi(\tau) \geq f_\tau(S)$ for any stopping time we find for the required initial capital

$$x \geq \sup_{\tau \in \mathcal{T}} I\!E^*(\beta(\tau) f_\tau(S)). \tag{3.28}$$

Suppose now that τ^* is such that $V_\varphi(\tau^*) = f_{\tau^*}(S)$ then the strategy φ is minimal and since $V_\varphi(t) \geq f_t(S)$ for all t we have

$$x = I\!E^*(\beta(\tau^*) f_{\tau^*}(S)) = \sup_{\tau \in \mathcal{T}} I\!E^*(\beta(\tau) f_\tau(S)) \tag{3.29}$$

Thus (3.29) is a necessary condition for the existence of a minimal strategy φ. We will show that it is also sufficient and call the price in (3.29) the rational price of an American contingent claim.

Now consider the problem of the option writer to construct such a strategy φ. At time T the hedging strategy needs to cover f_T, i.e. $V_\varphi(T) \geq f_T$ is required (We write short f_t for $f_t(S)$). At time $T - 1$ the option holder can either exercise and receive f_{T-1} or hold the option to expiry, in which case $B(T-1) I\!E^*(\beta(T) f_T | F_{T-1})$ needs to be covered. Thus the hedging strategy of the writer has to satisfy

$$V_\varphi(T - 1) = \max\{f_{T-1}, B(T-1) I\!E^*(\beta(T) f_T | \mathcal{F}_{T-1})\} \tag{3.30}$$

Using a backwards induction argument we can show that

$$V_\varphi(t - 1) = \max\{f_{t-1}, B(t-1) I\!E^*(\beta(t) V_\varphi(t) | \mathcal{F}_{t-1})\}. \tag{3.31}$$

Considering only discounted values this leads to

$$\tilde{V}_\varphi(t - 1) = \max\{\tilde{f}_{t-1}, I\!E^*(\tilde{V}_\varphi(t) | \mathcal{F}_{t-1})\}. \tag{3.32}$$

Thus we see that $\tilde{V}_\varphi(t)$ is the Snell envelope Z_t of $\tilde{f}_t$.

In particular we know that

$$Z_t = \sup_{\tau \in \mathcal{T}_t} I\!E^*(\tilde{f}_\tau | \mathcal{F}_t) \tag{3.33}$$

and the stopping time $\tau^* = \min\{s \geq t : Z_s = \tilde{f}_s\}$ is optimal. So

$$Z_t = I\!E^*(\tilde{f}_{\tau^*} | \mathcal{F}_t) \tag{3.34}$$

In case $t = 0$ we can use $\tau_0^* = \min\{s \geq 0 : Z_s = \tilde{f}_s\}$ and then

$$x = Z_0 = I\!E^*(\tilde{f}_{\tau_0^*}) = \sup_{\tau \in \mathcal{T}_0} I\!E^*(\tilde{f}_\tau) \tag{3.35}$$

is the rational option price.

We still need to construct the strategy φ. To do this recall that Z is a supermartingale and so the Doob decomposition yields

$$Z = \tilde{M} - \tilde{A} \tag{3.36}$$

with a martingale $\tilde{M}$ and a predictable, increasing process $\tilde{A}$. We write $M_t = \tilde{M}_t B_t$ and $A_t = \tilde{A}_t B_t$. Since the market is complete we know that there exists a self-financing strategy $\bar{\varphi}$ such that

$$\tilde{M}_t = \tilde{V}_{\bar{\varphi}}(t). \tag{3.37}$$

Also using (3.36) we find $Z_t B_t = V_{\bar{\varphi}}(t) - A_t$. Now on $C = \{(t,\omega) : 0 \leq t < \tau^*(\omega)\}$ we have that Z is a martingale and thus $A_t(\omega) = 0$. Thus we obtain from $\tilde{V}_{\bar{\varphi}}(t) = Z_t$ that

$$\tilde{V}_{\bar{\varphi}}(t) = \sup_{t \leq \tau \leq T} I\!E^*(\tilde{f}_\tau | \mathcal{F}_t) \quad \forall\, (t,\omega) \in C. \tag{3.38}$$

Now τ^* is the smallest exercise time and $\tilde{A}_{\tau^*(\omega)} = 0$. Thus

$$\tilde{V}_{\bar{\varphi}}(\tau^*(\omega), \omega) = Z_{\tau^*(\omega)}(\omega) = \tilde{f}_{\tau^*(\omega)}(\omega) \tag{3.39}$$

Undoing the discounting we find

$$V_{\bar{\varphi}}(\tau^*) = f_{\tau^*} \tag{3.40}$$

and therefore $\bar{\phi}$ is a minimal hedge.

Now consider the problem of the option holder, how to find the optimal exercise time. We observe that the optimal exercise time must be an optimal stopping time, since for any other stopping time σ (use Proposition 3.6.8)

$$\tilde{V}_{\varphi}(\sigma) = Z_\sigma > \tilde{f}_\sigma \tag{3.41}$$

and holding the asset longer would generate a larger payoff. Thus the holder needs to wait until $Z_\sigma = \tilde{f}_\sigma$ i.e. (i) of Proposition 3.6.8 is true. On the other hand with ν the largest stopping time (compare Definition 3.6.3) we see that $\sigma \leq \nu$. This follows since using $\bar{\phi}$ after ν with initial capital from exercising will always yield a higher portfolio value than the strategy of exercising later. To see this recall that $V_{\bar{\varphi}} = Z_t B_t + A_t$ with $A_t > 0$ for $t > \nu$. So we must have $\sigma \leq \nu$ and since $A_t = 0$ for $t \leq \nu$ we see that Z^σ is a martingale. Now criterion (ii) of Proposition 3.6.8 is true and σ is thus optimal. So

Proposition 3.6.10. *A stopping time $\sigma \in \mathcal{T}_t$ is an optimal exercise time for the American option (f_t) if and only if*

$$I\!E^*(\beta(\sigma)f_\sigma) = \sup_{\tau \in \mathcal{T}_t} I\!E^*(\beta(\tau)f_\tau) \tag{3.42}$$

3.6.3 American Options in the Cox-Ross-Rubinstein model

We now consider how to evaluate an American put option in a standard CRR model. We assume that the time interval $[0, T]$ is divided into N equal subintervals of length Δ say. Assuming the risk-free rate of interest r (over $[0,\text{T}]$) as given, we have $1 + \rho = e^{r\Delta}$ (where we denote the risk-free rate of interest in each subinterval by ρ). The remaining degrees of freedom are resolved by choosing u and d as follows:

$$1 + u = e^{\sigma\sqrt{\Delta}}, \quad \text{and} \quad 1 + d = (1 + u)^{-1} = e^{-\sigma\sqrt{\Delta}}.$$

By condition (3.9) the risk-neutral probabilities for the corresponding single period models are given by

$$p^* = \frac{\rho - d}{u - d} = \frac{e^{r\Delta} - e^{-\sigma\sqrt{\Delta}}}{e^{\sigma\sqrt{\Delta}} - e^{-\sigma\sqrt{\Delta}}}.$$

Thus the stock with initial value $S = S(0)$ is worth $S(1 + u)^i(1 + d)^j$ after i steps up and j steps down. Consequently, after N steps, there are $N + 1$ possible prices, $S(1 + u)^i(1 + d)^{N-i}$ ($i = 0, \ldots, N$). There are 2^N possible paths through the tree. It is common to take N of the order of 30, for two reasons:

(i) typical lengths of time to expiry of options are measured in months (9 months, say); this gives a time step around the corresponding number of days,
(ii) 2^{30} paths is about the order of magnitude that can be comfortably handled by computers (recall that $2^{10} = 1,024$, so 2^{30} is somewhat over a billion).

We can now calculate both the value of an American put option and the optimal exercise strategy by working backwards through the tree (this method of backward recursion in time is a form of the dynamic programming (DP) technique, due to Richard Bellman, which is important in many areas of optimisation and Operational Research).

1. Draw a binary tree showing the initial stock value and having the right number, N, of time intervals.
2. Fill in the stock prices: after one time interval, these are $S(1 + u)$ (upper) and $S(1 + d)$ (lower); after two time intervals, $S(1 + u)^2$, S and $S(1 + d)^2 = S/(1 + u)^2$; after i time intervals, these are $S(1 + u)^j(1 + d)^{i-j} = S(1 + u)^{2j-i}$ at the node with j 'up' steps and $i - j$ 'down' steps (the '(i, j)' node).
3. Using the strike price K and the prices at the terminal nodes, fill in the payoffs $f_{N,j}^A = \max\{K - S(1 + u)^j(1 + d)^{N-j}, 0\}$ from the option at the terminal nodes underneath the terminal prices.
4. Work back down the tree, from right to left. The no-exercise values f_{ij} of the option at the (i, j) node are given in terms of those of its upper and lower right neighbours in the usual way, as discounted expected values under the risk-neutral measure:

$$f_{ij} = e^{-r\Delta}[p^* f_{i+1,j+1}^A + (1 - p^*)f_{i+1,j}^A].$$

The intrinsic (or early-exercise) value of the American put at the (i, j) node – the value there if it is exercised early – is

$$K - S(1 + u)^j(1 + d)^{i-j}$$

(when this is non-negative, and so has any value). The value of the American put is the higher of these:

$$
\begin{aligned}
f_{ij}^A &= \max\{f_{ij}, K - S(1 + u)^j(1 + d)^{i-j}\} \\
&= \max\left\{e^{-r\Delta}(p^* f_{i+1,j+1}^A + (1 - p^*)f_{i+1,j}^A), K - S(1 + u)^j(1 + d)^{i-j}\right\}.
\end{aligned}
$$

5. The initial value of the option is the value f_0^A filled in at the root of the tree.
6. At each node, it is optimal to exercise early if the early-exercise value there exceeds the value f_{ij} there of expected discounted future payoff.

3.6.4 A Three-period Example

Assume we have two basic securities: a risk-free bond and a risky stock. The one-year risk-free interest rate (continuously compounded) is $r = 0.06$ and the volatility of the stock is 20%. We price calls and puts in three-period Cox-Ross-Rubinstein model. The up and down movements of the stock price are given by

$$1 + u = e^{\sigma\sqrt{\Delta}} = 1.1224 \quad \text{and} \quad 1 + d = (1 + u)^{-1} = e^{-\sigma\sqrt{\Delta}} = 0.8910,$$

with $\sigma = 0.2$ and $\Delta = 1/3$. We obtain risk-neutral probabilities by (3.9)

$$p^* = \frac{e^{r\Delta} - d}{u - d} = 0.5584.$$

We assume that the price of the stock at time $t = 0$ is $S(0) = 100$. To price a European call option with maturity one year ($N = 3$) and strike $K = 10$) we can either use the valuation formula (3.12) or work our way backwards through the tree. Prices of the stock and the call are given in Figure 4.2 below. One can implement the simple evaluation formulae for the CRR- and the BS-models
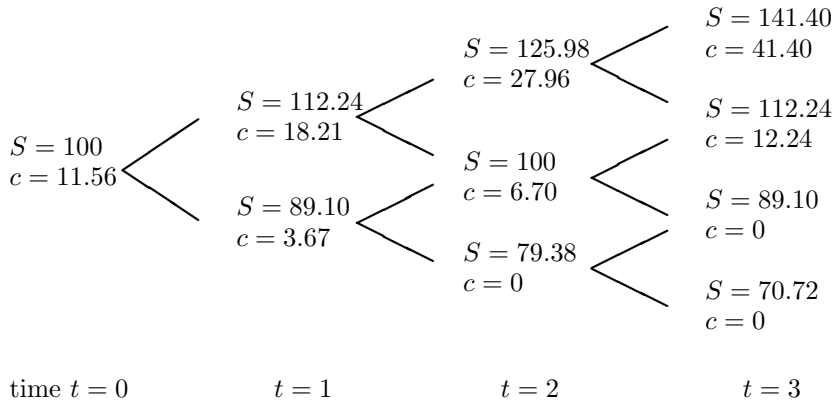
Figure 3.2: Stock and European call prices

and compare the values. Figure 3.3 is for $S = 100, K = 90, r = 0.06, \sigma = 0.2, T = 1$.
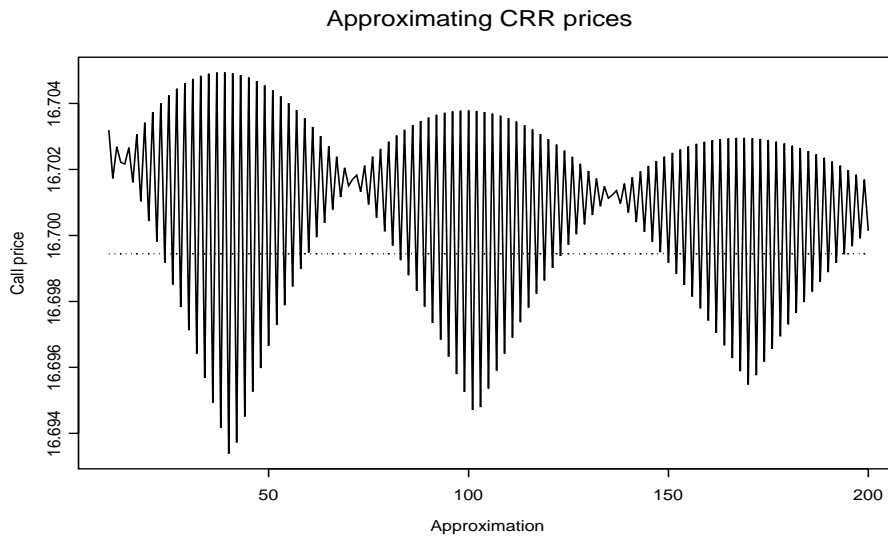
Figure 3.3: Approximation of Black-Scholes price by Binomial models

To price a European put, with price process denoted by $p(t)$, and an American put, $P(t)$, (maturity $N = 3$, strike 100), we can for the European put either use the put-call parity (1.1), the risk-neutral pricing formula, or work backwards through the tree. For the prices of the American put we use the technique outlined in §4.8.1. Prices of the two puts are given in Figure 4.4. We indicate the early exercise times of the American put in bold type. Recall that the discrete-time rule is to exercise if the intrinsic value $K - S(t)$ is larger than the value of the corresponding European put.

$p = 5.82$
$P = 6.18$

$p = 2.08$
$P = 2.08$

$p = 10.65$
$\boldsymbol{P} = 11.59$

$p = 0$
$P = 0$

$p = 4.76$
$P = 4.76$

$p = 18.71$
$\boldsymbol{P} = 20.62$

$p = 0$
$P = 0$

$p = 0$
$P = 0$

$p = 10.90$
$P = 10.90$

$p = 29.28$
$P = 29.28$

time $t = 0$ $t = 1$ $t = 2$ $t = 3$

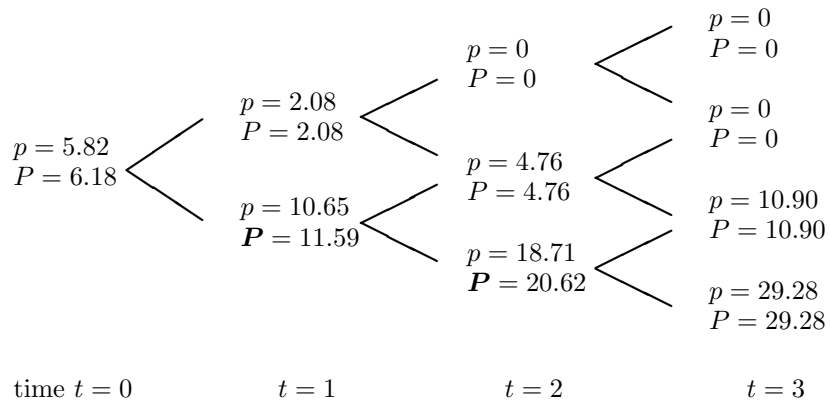Figure 3.4: European $p(.)$ and American $P(.)$ put prices

Chapter 4

Continuous-time Financial Market Models

4.1 The Stock Price Process and its Stochastic Calculus

4.1.1 Continuous-time Stochastic Processes

A *stochastic process* $X = (X(t))_{t \geq 0}$ is a family of random variables defined on $(\Omega, \mathcal{F}, I\!\!P, I\!\!F)$. We say X is *adapted* if $X(t) \in \mathcal{F}_t$ (i.e. $X(t)$ is $\mathcal{F}_t$-measurable) for each t: thus $X(t)$ is known when $\mathcal{F}_t$ is known, at time t.

The martingale property in continuous time is just that suggested by the discrete-time case:

Definition 4.1.1. *A stochastic process* $X = (X(t))_{0 \leq t < \infty}$ *is a martingale relative to* $(I\!\!F, I\!\!P)$ *if*

(i) *X is adapted, and* $I\!\!E |X(t)| < \infty$ *for all* $\leq t < \infty$;

(ii) $I\!\!E[X(t)|\mathcal{F}_s] = X(s)$ $I\!\!P - a.s.$ $(0 \leq s \leq t)$,

and similarly for sub- and supermartingales.

There are regularisation results, under which one can take $X(t)$ RCLL in t (basically $t \to I\!\!E X(t)$ has to be right-continuous). Then the analogues of the results for discrete-time martingales hold true.

Interpretation. Martingales model fair games. Submartingales model favourable games. Supermartingales model unfavourable games.

Brownian motion originates in work of the botanist Robert Brown in 1828. It was introduced into finance by Louis Bachelier in 1900, and developed in physics by Albert Einstein in 1905.

Definition 4.1.2. *A stochastic process* $X = (X(t))_{t \geq 0}$ *is a standard (one-dimensional) Brownian motion, BM or* $BM(I\!\!R)$, *on some probability space* $(\Omega, \mathcal{F}, I\!\!P)$, *if*

(i) $X(0) = 0$ *a.s.,*

(ii) *X has independent increments:* $X(t + u) - X(t)$ *is independent of* $\sigma(X(s) : s \leq t)$ *for* $u \geq 0$,

(iii) *X has stationary increments: the law of* $X(t + u) - X(t)$ *depends only on* u,

(iv) *X has Gaussian increments:* $X(t + u) - X(t)$ *is normally distributed with mean 0 and variance* u, $X(t + u) - X(t) \sim N(0, u)$,

(v) *X has continuous paths:* $X(t)$ *is a continuous function of* t, *i.e.* $t \to X(t, \omega)$ *is continuous in* t *for all* $\omega \in \Omega$.

We shall henceforth denote standard Brownian motion $BM(\mathbb{R})$ by $W = (W(t))$ (W for Wiener), though $B = (B(t))$ (B for Brown) is also common. Standard Brownian motion $BM(\mathbb{R}^d)$ in d dimensions is defined by $W(t) := (W_1(t), \ldots, W_d(t))$, where $W_1, \ldots, W_d$ are independent standard Brownian motions in one dimension (independent copies of $BM(\mathbb{R})$).

We have Wiener's theorem:

Theorem 4.1.1 (Wiener). *Brownian motion exists.*

For further background, see any measure-theoretic text on stochastic processes. A treatment starting directly from our main reference of measure-theoretic results, Williams (Williams 1991), is Rogers and Williams (Rogers and Williams 1994), Chapter 1. The classic is Doob's book, (Doob 1953), VIII.2. Excellent modern texts include (Karatzas and Shreve 1991, Revuz and Yor 1991) (see particularly (Karatzas and Shreve 1991), §2.2-4 for construction).

Geometric Brownian Motion

Now that we have both Brownian motion W and Itô's Lemma to hand, we can introduce the most important stochastic process for us, a relative of Brownian motion - *geometric* (or *exponential*, or *economic*) Brownian motion.

Suppose we wish to model the time evolution of a stock price $S(t)$ (as we will, in the Black-Scholes theory). Consider how S will change in some small time-interval from the present time t to a time $t + dt$ in the near future. Writing $dS(t)$ for the change $S(t + dt) - S(t)$ in S, the *return* on S in this interval is $dS(t)/S(t)$. It is economically reasonable to expect this return to decompose into two components, a *systematic* part and a *random* part. The systematic part could plausibly be modelled by μdt, where μ is some parameter representing the mean rate of return of the stock. The random part could plausibly be modelled by $\sigma dW(t)$, where $dW(t)$ represents the noise term driving the stock price dynamics, and σ is a second parameter describing how much effect this noise has - how much the stock price fluctuates. Thus σ governs how volatile the price is, and is called the *volatility* of the stock. The role of the driving noise term is to represent the random buffeting effect of the multiplicity of factors at work in the economic environment in which the stock price is determined by supply and demand.

Putting this together, we have the stochastic differential equation

$$dS(t) = S(t)(\mu dt + \sigma dW(t)), \quad S(0) > 0, \qquad (4.1)$$

due to Itô in 1944. This corrects Bachelier's earlier attempt of 1900 (he did not have the factor $S(t)$ on the right - missing the interpretation in terms of returns, and leading to negative stock prices!) Incidentally, Bachelier's work served as Itô's motivation in introducing Itô calculus. The mathematical importance of Itô's work was recognised early, and led on to the work of (Doob 1953), (Meyer 1976) and many others (see the memorial volume (Ikeda, Watanabe, M., and Kunita 1996) in honour of Itô's eightieth birthday in 1995). The economic importance of geometric Brownian motion was recognised by Paul A. Samuelson in his work from 1965 on ((Samuelson 1965)), for which Samuelson received the Nobel Prize in Economics in 1970, and by Robert Merton (see (Merton 1990) for a full bibliography), in work for which he was similarly honoured in 1997.

4.1.2 Stochastic Analysis

Stochastic integration was introduced by K. Itô in 1944, hence its name Itô calculus. It gives a meaning to

$$\int\limits_0^t X dY = \int\limits_0^t X(s, \omega) dY(s, \omega),$$

for suitable stochastic processes X and Y, the integrand and the integrator. We shall confine our attention here mainly to the basic case with integrator Brownian motion: $Y = W$. Much greater

generality is possible: for Y a continuous martingale, see (Karatzas and Shreve 1991) or (Revuz and Yor 1991); for a systematic general treatment, see (Protter 2004).

The first thing to note is that stochastic integrals with respect to Brownian motion, if they exist, must be quite different from the measure-theoretic integral. For, the Lebesgue-Stieltjes integrals described there have as integrators the difference of two monotone (increasing) functions, which are locally of bounded variation. But we know that Brownian motion is of infinite (unbounded) variation on every interval. So Lebesgue-Stieltjes and Itô integrals must be fundamentally different.

In view of the above, it is quite surprising that Itô integrals can be defined at all. But if we take for granted Itô's fundamental insight that they can be, it is obvious how to begin and clear enough how to proceed. We begin with the simplest possible integrands X, and extend successively in much the same way that we extended the measure-theoretic integral.

Indicators.

If $X(t, \omega) = \mathbf{1}_{[a,b]}(t)$, there is exactly one plausible way to define $\int X dW$:

$$\int\limits_0^t X(s, \omega) dW(s, \omega) := \begin{cases} 0 & \text{if } t \leq a, \\ W(t) - W(a) & \text{if } a \leq t \leq b, \\ W(b) - W(a) & \text{if } t \geq b. \end{cases}$$

Simple Functions.

Extend by linearity: if X is a linear combination of indicators, $X = \sum_{i=1}^n c_i \mathbf{1}_{[a_i, b_i]}$, we should define

$$\int\limits_0^t X dW := \sum_{i=1}^n c_i \int\limits_0^t \mathbf{1}_{[a_i, b_i]} dW.$$

Already one wonders how to extend this from constants c_i to suitable random variables, and one seeks to simplify the obvious but clumsy three-line expressions above.

We begin again, this time calling a stochastic process X *simple* if there is a partition $0 = t_0 < t_1 < \ldots < t_n = T < \infty$ and uniformly bounded $\mathcal{F}_{t_n}$-measurable random variables ξ_k ($|\xi_k| \leq C$ for all $k = 0, \ldots, n$ and ω, for some C) and if $X(t, \omega)$ can be written in the form

$$X(t, \omega) = \xi_0(\omega) \mathbf{1}_{\{0\}}(t) + \sum_{i=0}^n \xi_i(\omega) \mathbf{1}_{(t_i, t_{i+1}]}(t) \quad (0 \leq t \leq T, \omega \in \Omega).$$

Then if $t_k \leq t < t_{k+1}$,

$$I_t(X) := \int\limits_0^t X dW = \sum_{i=0}^{k-1} \xi_i(W(t_{i+1}) - W(t_i)) + \xi_k(W(t) - W(t_k))$$

$$= \sum_{i=0}^n \xi_i(W(t \wedge t_{i+1}) - W(t \wedge t_i)).$$

Note that by definition $I_0(X) = 0 \quad I\!\!P - a.s.$. We collect some properties of the stochastic integral defined so far:

Lemma 4.1.1. *(i)* $I_t(aX + bY) = aI_t(X) + bI_t(Y)$.
(ii) $I\!\!E(I_t(X)|\mathcal{F}_s) = I_s(X) \quad I\!\!P - a.s. \quad (0 \leq s < t < \infty)$, *hence* $I_t(X)$ *is a continuous martingale.*

The stochastic integral for simple integrands is essentially a martingale transform, and the above is essentially the proof that martingale transforms are martingales.

We pause to note a property of square-integrable martingales which we shall need below. Call $M(t) - M(s)$ the increment of M over $(s, t]$. Then for a martingale M, the product of the increments over disjoint intervals has zero mean. For, if $s < t \leq u < v$,

$$\mathbb{E}\left[(M(v) - M(u))(M(t) - M(s))\right]$$
$$= \mathbb{E}\left[\mathbb{E}((M(v) - M(u))(M(t) - M(s))|\mathcal{F}_u)\right]$$
$$= \mathbb{E}\left[(M(t) - M(s))\mathbb{E}((M(v) - M(u))|\mathcal{F}_u)\right],$$

taking out what is known (as $s, t \leq u$). The inner expectation is zero by the martingale property, so the left-hand side is zero, as required.

We now can add further properties of the stochastic integral for simple functions.

Lemma 4.1.2. *(i) We have the Itô isometry*

$$\mathbb{E}\left((I_t(X))^2\right) = \mathbb{E}\left(\int_0^t X(s)^2 ds\right).$$

(ii) $\mathbb{E}\left((I_t(X) - I_s(X))^2|\mathcal{F}_s\right) = \mathbb{E}\left(\int_s^t X(u)^2 du\right)$ *$\mathbb{P} - a.s.$*

The Itô isometry above suggests that $\int_0^t X dW$ should be defined only for processes with

$$\int_0^t \mathbb{E}\left(X(u)^2\right) du < \infty \quad \text{for all} \quad t.$$

We then can transfer convergence on a suitable L^2-space of stochastic processes to a suitable L^2-space of martingales. This gives us an L^2-theory of stochastic integration, for which Hilbert-space methods are available.

For the financial applications we have in mind, there is a fixed time-interval - $[0, T]$ say - on which we work (e.g., an option is written at time $t = 0$, with expiry time $t = T$). Then the above becomes

$$\int_0^T \mathbb{E}(X(u)^2) du < \infty.$$

Approximation.

We seek a class of integrands suitably approximable by simple integrands. It turns out that:

(i) The suitable class of integrands is the class of $(\mathcal{B}([0, \infty)) \otimes \mathcal{F})$-measurable, $(\mathcal{F}_t)$- adapted processes X with $\int_0^t \mathbb{E}\left(X(u)^2\right) du < \infty$ for all $t > 0$.
(ii) Each such X may be approximated by a sequence of simple integrands X_n so that the stochastic integral $I_t(X) = \int_0^t X dW$ may be defined as the limit of $I_t(X_n) = \int_0^t X_n dW$.
(iii) The properties from both lemmas above remain true for the stochastic integral $\int_0^t X dW$ defined by (i) and (ii).

Example.

We calculate $\int W(u) dW(u)$. We start by approximating the integrand by a sequence of simple functions.

$$X_n(u) = \begin{cases} W(0) = 0 & \text{if} \quad 0 \leq u \leq t/n, \\ W(t/n) & \text{if} \quad t/n < u \leq 2t/n, \\ \vdots & \vdots \\ W\left(\frac{(n-1)t}{n}\right) & \text{if} \quad (n-1)t/n < u \leq t. \end{cases}$$

By definition,

$$\int\limits_0^t W(u)dW(u) = \lim_{n\to\infty} \sum_{k=0}^{n-1} W\left(\frac{kt}{n}\right)\left(W\left(\frac{(k+1)t}{n}\right) - W\left(\frac{kt}{n}\right)\right).$$

Rearranging terms, we obtain for the sum on the right

$$\sum_{k=0}^{n-1} W\left(\frac{kt}{n}\right)\left(W\left(\frac{(k+1)t}{n}\right) - W\left(\frac{kt}{n}\right)\right)$$
$$= \frac{1}{2}W(t)^2 - \frac{1}{2}\left[\sum_{k=0}^{n-1}\left(W\left(\frac{(k+1)t}{n}\right) - W\left(\frac{kt}{n}\right)\right)^2\right].$$

Since the second term approximates the quadratic variation of W and hence tends to t for $n \to \infty$, we find

$$\int\limits_0^t W(u)dW(u) = \frac{1}{2}W(t)^2 - \frac{1}{2}t. \qquad (4.2)$$

Note the contrast with ordinary (Newton-Leibniz) calculus! Itô calculus requires the second term on the right – the Itô correction term – which arises from the quadratic variation of W.

One can construct a closely analogous theory for stochastic integrals with the Brownian integrator W above replaced by a square-integrable martingale integrator M. The properties above hold, with (i) in Lemma replaced by

$$\mathbb{E}\left[\left(\int\limits_0^t X(u)dM(u)\right)^2\right] = \mathbb{E}\left[\int\limits_0^t X(u)^2 d\langle M\rangle(u)\right].$$

Quadratic Variation, Quadratic Covariation.

We shall need to extend quadratic variation and quadratic covariation to stochastic integrals. The quadratic variation of $I_t(X) = \int_0^t X(u)dW(u)$ is $\int_0^t X(u)^2 du$. This is proved in the same way as the case $X \equiv 1$, that W has quadratic variation process t. More generally, if $Z(t) = \int_0^t X(u)dM(u)$ for a continuous martingale integrator M, then $\langle Z\rangle(t) = \int_0^t X^2(u)d\langle M\rangle(u)$. Similarly (or by polarisation), if $Z_i(t) = \int_0^t X_i(u)dM_i(u)$ $(i = 1, 2)$, $\langle Z_1, Z_2\rangle(t) = \int_0^t X_1(u)X_2(u)d\langle M_1, M_2\rangle(u)$.

4.1.3 Itô's Lemma

Suppose that b is adapted and locally integrable (so $\int_0^t b(s)ds$ is defined as an ordinary integral), and σ is adapted and measurable with $\int_0^t \mathbb{E}\left(\sigma(u)^2\right)du < \infty$ for all t (so $\int_0^t \sigma(s)dW(s)$ is defined as a stochastic integral). Then

$$X(t) := x_0 + \int\limits_0^t b(s)ds + \int\limits_0^t \sigma(s)dW(s)$$

defines a stochastic process X with $X(0) = x_0$. It is customary, and convenient, to express such an equation symbolically in differential form, in terms of the stochastic differential equation

$$dX(t) = b(t)dt + \sigma(t)dW(t), \quad X(0) = x_0. \qquad (4.3)$$

Now suppose $f : \mathbb{R} \to \mathbb{R}$ is of class C^2. The question arises of giving a meaning to the stochastic differential $df(X(t))$ of the process $f(X(t))$, and finding it. Given a partition $\mathcal{P}$ of $[0, t]$, i.e. $0 = t_0 < t_1 < \ldots < t_n = t$, we can use Taylor's formula to obtain

$$
\begin{aligned}
f(X(t)) - f(X(0)) &= \sum_{k=0}^{n-1} f(X(t_{k+1})) - f(X(t_k)) \\
&= \sum_{k=0}^{n-1} f'(X(t_k))\Delta X(t_k) \\
&\quad + \frac{1}{2} \sum_{k=0}^{n-1} f''(X(t_k) + \theta_k \Delta X(t_k))(\Delta X(t_k))^2
\end{aligned}
$$

with $0 < \theta_k < 1$. We know that $\sum(\Delta X(t_k))^2 \to \langle X \rangle(t)$ in probability (so, taking a subsequence, with probability one), and with a little more effort one can prove

$$
\sum_{k=0}^{n-1} f''(X(t_k) + \theta_k \Delta X(t_k))(\Delta X(t_k))^2 \to \int_0^t f''(X(u))d\langle X \rangle(u).
$$

The first sum is easily recognized as an approximating sequence of a stochastic integral; indeed, we find

$$
\sum_{k=0}^{n-1} f'(X(t_k))\Delta X(t_k) \to \int_0^t f'(X(u))dX(u).
$$

So we have

Theorem 4.1.2 (Basic Itô formula). *If X has stochastic differential given by 4.3 and $f \in C^2$, then $f(X)$ has stochastic differential*

$$
df(X(t)) = f'(X(t))dX(t) + \frac{1}{2}f''(X(t))d\langle X \rangle(t),
$$

or writing out the integrals,

$$
f(X(t)) = f(x_0) + \int_0^t f'(X(u))dX(u) + \frac{1}{2}\int_0^t f''(X(u))d\langle X \rangle(u).
$$

More generally, suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is a function, continuously differentiable once in its first argument (which will denote time), and twice in its second argument (space): $f \in C^{1,2}$. By the Taylor expansion of a smooth function of several variables we get for t close to t_0 (we use subscripts to denote partial derivatives: $f_t := \partial f/\partial t$, $f_{tx} := \partial^2 f/\partial t\partial x$):

$$
\begin{aligned}
f(t, X(t)) &= f(t_0, X(t_0)) \\
&\quad + (t - t_0)f_t(t_0, X(t_0)) + (X(t) - X(t_0))f_x(t_0, X(t_0)) \\
&\quad + \frac{1}{2}(t - t_0)^2 f_{tt}(t_0, X(t_0)) + \frac{1}{2}(X(t) - X(t_0))^2 f_{xx}(t_0, X(t_0)) \\
&\quad + (t - t_0)(X(t) - X(t_0))f_{tx}(t_0, X(t_0)) + \ldots,
\end{aligned}
$$

which may be written symbolically as

$$
df = f_t dt + f_x dX + \frac{1}{2}f_{tt}(dt)^2 + f_{tx}dtdX + \frac{1}{2}f_{xx}(dX)^2 + \ldots.
$$

In this, we substitute $dX(t) = b(t)dt + \sigma(t)dW(t)$ from above, to obtain

$$
\begin{aligned}
df \quad = \quad & f_t dt + f_x(bdt + \sigma dW) \\
& + \frac{1}{2} f_{tt}(dt)^2 + f_{tx}dt(bdt + \sigma dW) + \frac{1}{2}f_{xx}(bdt + \sigma dW)^2 + \ldots
\end{aligned}
$$

Now using the formal multiplication rules $dt \cdot dt = 0, dt \cdot dW = 0, dW \cdot dW = dt$ (which are just shorthand for the corresponding properties of the quadratic variations, we expand

$$(bdt + \sigma dW)^2 = \sigma^2 dt + 2b\sigma dt dW + b^2(dt)^2 = \sigma^2 dt + \text{higher-order terms}$$

to get finally

$$df = \left(f_t + bf_x + \frac{1}{2}\sigma^2 f_{xx} \right) dt + \sigma f_x dW + \text{higher-order terms}.$$

As above the higher-order terms are irrelevant, and summarising, we obtain *Itô's lemma*, the analogue for the Itô or stochastic calculus of the chain rule for ordinary (Newton-Leibniz) calculus:

Theorem 4.1.3 (Itô's Lemma). *If $X(t)$ has stochastic differential given by 4.3, then $f = f(t, X(t))$ has stochastic differential*

$$df = \left(f_t + bf_x + \frac{1}{2}\sigma^2 f_{xx} \right) dt + \sigma f_x dW.$$

That is, writing f_0 for $f(0, x_0)$, the initial value of f,

$$f = f_0 + \int_0^t (f_t + bf_x + \frac{1}{2}\sigma^2 f_{xx})dt + \int_0^t \sigma f_x dW.$$

We will make good use of:

Corollary 4.1.1. $I\!\!E\left(f(t, X(t))\right) = f_0 + \int_0^t I\!\!E\left(f_t + bf_x + \frac{1}{2}\sigma^2 f_{xx}\right) dt.$

Proof. $\int_0^t \sigma f_2 dW$ is a stochastic integral, so a martingale, so its expectation is constant ($= 0$, as it starts at 0). ∎

The differential equation (4.1) above has the unique solution

$$S(t) = S(0)\exp\left\{ \left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma dW(t) \right\}.$$

For, writing

$$f(t, x) := \exp\left\{ \left(\mu - \frac{1}{2}\sigma^2 \right) t + \sigma x \right\},$$

we have

$$f_t = \left(\mu - \frac{1}{2}\sigma^2 \right) f, \quad f_x = \sigma f, \quad f_{xx} = \sigma^2 f,$$

and with $x = W(t)$, one has

$$dx = dW(t), \quad (dx)^2 = dt.$$

Thus Itô's lemma gives

$$
\begin{aligned}
df(t, W(t)) \quad = \quad & f_t dt + f_x dW(t) + \frac{1}{2} f_{xx}(dW(t))^2 \\
= \quad & f\left(\left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW(t) + \frac{1}{2}\sigma^2 dt \right) \\
= \quad & f(\mu dt + \sigma dW(t)),
\end{aligned}
$$

so $f(t, W(t))$ is a solution of the stochastic differential equation, and the initial condition $f(0, W(0)) = S(0)$ as $W(0) = 0$, giving existence.

4.1.4   Girsanov's Theorem

Consider first independent $N(0,1)$ random variables $Z_1, \ldots, Z_n$ on a probability space $(\Omega, \mathcal{F}, I\!\!P)$. Given a vector $\gamma = (\gamma_1, \ldots, \gamma_n)$, consider a new probability measure $\tilde{I\!\!P}$ on $(\Omega, \mathcal{F})$ defined by

$$\tilde{I\!\!P}(d\omega) = \exp\left\{\sum_{i=1}^{n} \gamma_i Z_i(\omega) - \frac{1}{2}\sum_{i=1}^{n}\gamma_i^2\right\} I\!\!P(d\omega).$$

As $\exp\{.\} > 0$ and integrates to 1, as $\int \exp\{\gamma_i Z_i\} dI\!\!P = \exp\{\frac{1}{2}\gamma_i^2\}$, this is a probability measure. It is also equivalent to $I\!\!P$ (has the same null sets), again as the exponential term is positive. Also

$$\tilde{I\!\!P}(Z_i \in dz_i, \quad i = 1, \ldots, n)$$

$$= \exp\left\{\sum_{i=1}^{n}\gamma_i Z_i - \frac{1}{2}\sum_{i=1}^{n}\gamma_i^2\right\} I\!\!P(Z_i \in dz_i, \quad i = 1, \ldots, n)$$

$$= (2\pi)^{-\frac{n}{2}}\exp\left\{\sum_{i=1}^{n}\gamma_i z_i - \frac{1}{2}\sum_{i=1}^{n}\gamma_i^2 - \frac{1}{2}\sum_{i=1}^{n}z_i^2\right\}\prod_{i=1}^{n}dz_i$$

$$= (2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(z_i - \gamma_i)^2\right\}dz_1 \ldots dz_n.$$

This says that if the Z_i are independent $N(0,1)$ under $I\!\!P$, they are independent $N(\gamma_i, 1)$ under $\tilde{I\!\!P}$. Thus the effect of the *change of measure* $I\!\!P \to \tilde{I\!\!P}$, from the original measure $I\!\!P$ to the *equivalent* measure $\tilde{I\!\!P}$, is to *change the mean*, from $0 = (0, \ldots, 0)$ to $\gamma = (\gamma_1, \ldots, \gamma_n)$.

This result extends to infinitely many dimensions - i.e., from random vectors to stochastic processes, indeed with random rather than deterministic means. Let $W = (W_1, \ldots W_d)$ be a d-dimensional Brownian motion defined on a filtered probability space $(\Omega, \mathcal{F}, I\!\!P, I\!\!F)$ with the filtration $I\!\!F$ satisfying the usual conditions. Let $(\gamma(t) : 0 \leq t \leq T)$ be a measurable, adapted d-dimensional process with $\int_0^T \gamma_i(t)^2 dt < \infty \quad a.s., \ i = 1, \ldots, d$, and define the process $(L(t) : 0 \leq t \leq T)$ by

$$L(t) = \exp\left\{-\int_0^t \gamma(s)' dW(s) - \frac{1}{2}\int_0^t \|\gamma(s)\|^2 \, ds\right\}. \tag{4.4}$$

Then L is continuous, and, being the stochastic exponential of $-\int_0^t \gamma(s)' dW(s)$, is a local martingale. Given sufficient integrability on the process γ, L will in fact be a (continuous) martingale. For this, *Novikov's condition* suffices:

$$I\!\!E\left(\exp\left\{\frac{1}{2}\int_0^T \|\gamma(s)\|^2 \, ds\right\}\right) < \infty. \tag{4.5}$$

We are now in the position to state a version of Girsanov's theorem, which will be one of our main tools in studying continuous-time financial market models.

Theorem 4.1.4 (Girsanov). *Let γ be as above and satisfy Novikov's condition; let L be the corresponding continuous martingale. Define the processes $\tilde{W}_i$, $i = 1, \ldots, d$ by*

$$\tilde{W}_i(t) := W_i(t) + \int_0^t \gamma_i(s) ds, \quad (0 \leq t \leq T), \quad i = 1, \ldots, d.$$

Then under the equivalent probability measure $\tilde{I\!\!P}$ (defined on $(\Omega, \mathcal{F}_T)$) with Radon-Nikodým derivative

$$\frac{d\tilde{I\!\!P}}{dI\!\!P} = L(T),$$

the process $\tilde{W} = (\tilde{W}_1, \ldots, \tilde{W}_d)$ is d-dimensional Brownian motion.

In particular, for $\gamma(t)$ constant $(= \gamma)$, change of measure by introducing the Radon-Nikodým derivative $\exp\left\{-\gamma W(t) - \frac{1}{2}\gamma^2 t\right\}$ corresponds to a change of drift from c to $c - \gamma$. If $\mathbb{F} = (\mathcal{F}_t)$ is the Brownian filtration (basically $\mathcal{F}_t = \sigma(W(s), 0 \le s \le t)$ slightly enlarged to satisfy the usual conditions) any pair of equivalent probability measures $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F} = \mathcal{F}_T$ is a Girsanov pair, i.e.

$$\left. \frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = L(t)$$

with L defined as above. Girsanov's theorem (or the Cameron-Martin-Girsanov theorem) is formulated in varying degrees of generality, discussed and proved, e.g. in (Karatzas and Shreve 1991), §3.5, (Protter 2004), III.6, (Revuz and Yor 1991), VIII, (Dothan 1990), §5.4 (discrete time), §11.6 (continuous time).

4.2 Financial Market Models

4.2.1 The Financial Market Model

We start with a general model of a frictionless (compare Chapter 1) security market where investors are allowed to trade continuously up to some fixed finite planning horizon T. Uncertainty in the financial market is modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}$ satisfying the usual conditions of right-continuity and completeness. We assume that the σ-field $\mathcal{F}_0$ is trivial, i.e. for every $A \in \mathcal{F}_0$ either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, and that $\mathcal{F}_T = \mathcal{F}$.

There are $d + 1$ primary traded assets, whose price processes are given by stochastic processes $S_0, \ldots, S_d$. We assume that the processes $S_0, \ldots, S_d$ represent the prices of some traded assets (stocks, bonds, or options).

We have not emphasised so far that there was an implicit numéraire behind the prices $S_0, \ldots, S_d$; it is the numéraire relevant for domestic transactions at time t. The formal definition of a numéraire is very much as in the discrete setting.

Definition 4.2.1. *A numéraire is a price process $X(t)$ almost surely strictly positive for each $t \in [0, T]$.*

We assume now that $S_0(t)$ is a non-dividend paying asset, which is (almost surely) strictly positive and use S_0 as numéraire. 'Historically' (see (Harrison and Pliska 1981)) the money market account $B(t)$, given by $B(t) = e^{r(t)}$ with a positive deterministic process $r(t)$ and $r(0) = 0$, was used as a numéraire, and the reader may think of $S_0(t)$ as being $B(t)$.

Our principal task will be the pricing and hedging of contingent claims, which we model as $\mathcal{F}_T$-measurable random variables. This implies that the contingent claims specify a stochastic cash-flow at time T and that they may depend on the whole path of the underlying in $[0, T]$ - because $\mathcal{F}_T$ contains all that information. We will often have to impose further integrability conditions on the contingent claims under consideration. The fundamental concept in (arbitrage) pricing and hedging contingent claims is the interplay of self-financing replicating portfolios and risk-neutral probabilities. Although the current setting is on a much higher level of sophistication, the key ideas remain the same.

We call an $\mathbb{R}^{d+1}$-valued predictable process

$$\varphi(t) = (\varphi_0(t), \ldots, \varphi_d(t)), \quad t \in [0, T]$$

with $\int_0^T \mathbb{E}(\varphi_0(t))dt < \infty, \sum_{i=0}^d \int_0^T \mathbb{E}(\varphi_i^2(t))dt < \infty$ a trading strategy (or dynamic portfolio process). Here $\varphi_i(t)$ denotes the number of shares of asset i held in the portfolio at time t - to be determined on the basis of information available *before* time t; i.e. the investor selects his time t portfolio after observing the prices $S(t-)$. The components $\varphi_i(t)$ may assume negative as well as positive values, reflecting the fact that we allow short sales and assume that the assets are perfectly divisible.

Definition 4.2.2. *(i) The value of the portfolio φ at time t is given by the scalar product*

$$V_\varphi(t) := \varphi(t) \cdot S(t) = \sum_{i=0}^{d} \varphi_i(t) S_i(t), \quad t \in [0, T].$$

The process $V_\varphi(t)$ is called the value process, or wealth process, of the trading strategy φ.
(ii) The gains process $G_\varphi(t)$ is defined by

$$G_\varphi(t) := \int_0^t \varphi(u) dS(u) = \sum_{i=0}^{d} \int_0^t \varphi_i(u) dS_i(u).$$

(iii) A trading strategy φ is called self-financing if the wealth process $V_\varphi(t)$ satisfies

$$V_\varphi(t) = V_\varphi(0) + G_\varphi(t) \quad \text{for all } t \in [0, T].$$

Remark 4.2.1. *(i) The financial implications of the above equations are that all changes in the wealth of the portfolio are due to capital gains, as opposed to withdrawals of cash or injections of new funds.*
(ii) The definition of a trading strategy includes regularity assumptions in order to ensure the existence of stochastic integrals.

Using the special numéraire $S_0(t)$ we consider the discounted price process

$$\tilde{S}(t) := \frac{S(t)}{S_0(t)} = (1, \tilde{S}_1(t), \ldots \tilde{S}_d(t))$$

with $\tilde{S}_i(t) = S_i(t)/S_0(t)$, $i = 1, 2, \ldots, d$. Furthermore, the discounted wealth process $\tilde{V}_\varphi(t)$ is given by

$$\tilde{V}_\varphi(t) := \frac{V_\varphi(t)}{S_0(t)} = \varphi_0(t) + \sum_{i=1}^{d} \varphi_i(t) \tilde{S}_i(t)$$

and the discounted gains process $\tilde{G}_\varphi(t)$ is

$$\tilde{G}_\varphi(t) := \sum_{i=1}^{d} \int_0^t \varphi_i(t) d\tilde{S}_i(t).$$

Observe that $\tilde{G}_\varphi(t)$ does not depend on the numéraire component φ_0.

It is convenient to reformulate the self-financing condition in terms of the discounted processes:

Proposition 4.2.1. *Let φ be a trading strategy. Then φ if self-financing if and only if*

$$\tilde{V}_\varphi(t) = \tilde{V}_\varphi(0) + \tilde{G}_\varphi(t).$$

Of course, $V_\varphi(t) \geq 0$ if and only if $\tilde{V}_\varphi(t) \geq 0$.

The *proof* follows by the numéraire invariance theorem using S_0 as numéraire. ∎

Remark 4.2.2. *The above result shows that a self-financing strategy is completely determined by its initial value and the components $\varphi_1, \ldots, \varphi_d$. In other words, any set of predictable processes $\varphi_1, \ldots, \varphi_d$ such that the stochastic integrals $\int \varphi_i d\tilde{S}_i$, $i = 1, \ldots, d$ exist can be uniquely extended to a self-financing strategy φ with specified initial value $\tilde{V}_\varphi(0) = v$ by setting the cash holding as*

$$\varphi_0(t) = v + \sum_{i=1}^{d} \int_0^t \varphi_i(u) d\tilde{S}_i(u) - \sum_{i=1}^{d} \varphi_i(t) \tilde{S}_i(t), \ t \in [0, T].$$

4.2.2   Equivalent Martingale Measures

We develop a relative pricing theory for contingent claims. Again the underlying concept is the link between the no-arbitrage condition and certain probability measures. We begin with:

Definition 4.2.3. *A self-financing trading strategy φ is called an arbitrage opportunity if the wealth process V_φ satisfies the following set of conditions:*

$$V_\varphi(0) = 0, \quad I\!\!P(V_\varphi(T) \geq 0) = 1, \quad and \quad I\!\!P(V_\varphi(T) > 0) > 0.$$

Arbitrage opportunities represent the limitless creation of wealth through risk-free profit and thus should not be present in a well-functioning market.

The main tool in investigating arbitrage opportunities is the concept of equivalent martingale measures:

Definition 4.2.4. *We say that a probability measure $\boldsymbol{Q}$ defined on $(\Omega, \mathcal{F})$ is an equivalent martingale measure if:*

(i) $\boldsymbol{Q}$ is equivalent to $I\!\!P$,
(ii) the discounted price process $\tilde{S}$ is a $\boldsymbol{Q}$ martingale.

We denote the set of martingale measures by $\mathcal{P}$.

A useful criterion in determining whether a given equivalent measure is indeed a martingale measure is the observation that the growth rates relative to the numéraire of all given primary assets under the measure in question must coincide. For example, in the case $S_0(t) = B(t)$ we have:

Lemma 4.2.1. *Assume $S_0(t) = B(t) = e^{r(t)}$, then $\boldsymbol{Q} \sim I\!\!P$ is a martingale measure if and only if every asset price process S_i has price dynamics under $\boldsymbol{Q}$ of the form*

$$dS_i(t) = r(t)S_i(t)dt + dM_i(t),$$

where M_i is a $\boldsymbol{Q}$-martingale.

The *proof* is an application of Itô's formula.

In order to proceed we have to impose further restrictions on the set of trading strategies.

Definition 4.2.5. *A self-financing trading strategy φ is called tame (relative to the numéraire S_0) if*

$$\tilde{V}_\varphi(t) \geq 0 \quad for \ all \ t \in [0, T].$$

We use the notation Φ for the set of tame trading strategies.

We next analyse the value process under equivalent martingale measures for such strategies.

Proposition 4.2.2. *For $\varphi \in \Phi$ $\tilde{V}_\varphi(t)$ is a martingale under each $\boldsymbol{Q} \in \mathcal{P}$.*

This observation is the key to our first central result:

Theorem 4.2.1. *Assume $\mathcal{P} \neq \emptyset$. Then the market model contains no arbitrage opportunities in Φ.*

Proof. For any $\varphi \in \Phi$ and under any $\boldsymbol{Q} \in \mathcal{P}$ $\tilde{V}_\varphi(t)$ is a martingale. That is,

$$I\!\!E_{\boldsymbol{Q}}\left(\tilde{V}_\varphi(t)|\mathcal{F}_u\right) = \tilde{V}_\varphi(u), \quad \text{for all} \quad u \leq t \leq T.$$

For $\varphi \in \Phi$ to be an arbitrage opportunity we must have $\tilde{V}_\varphi(0) = V_\varphi(0) = 0$. Now

$$I\!\!E_{\boldsymbol{Q}}\left(\tilde{V}_\varphi(t)\right) = 0, \quad \text{for all} \quad 0 \leq t \leq T.$$

Now φ is tame, so $\tilde{V}_\varphi(t) \geq 0$, $0 \leq t \leq T$, implying $\mathbb{E}_{\boldsymbol{Q}}\left(\tilde{V}_\varphi(t)\right) = 0$, $0 \leq t \leq T$, and in particular $\mathbb{E}_{\boldsymbol{Q}}\left(\tilde{V}_\varphi(T)\right) = 0$. But an arbitrage opportunity φ also has to satisfy $\mathbb{P}\left(V_\varphi(T) \geq 0\right) = 1$, and since $\boldsymbol{Q} \sim \mathbb{P}$, this means $\boldsymbol{Q}\left(V_\varphi(T) \geq 0\right) = 1$. Both together yield

$$\boldsymbol{Q}\left(V_\varphi(T) > 0\right) = \mathbb{P}\left(V_\varphi(T) > 0\right) = 0,$$

and hence the result follows. ∎

4.2.3 Risk-neutral Pricing

We now assume that there exists an equivalent martingale measure $\mathbb{P}^*$ which implies that there are no arbitrage opportunities with respect to Φ in the financial market model. Until further notice we use $\mathbb{P}^*$ as our reference measure, and when using the term martingale we always assume that the underlying probability measure is $\mathbb{P}^*$. In particular, we restrict our attention to contingent claims X such that $X/S_0(T) \in L^1(\mathcal{F}, \mathbb{P}^*)$.

We now define a further subclass of trading strategies:

Definition 4.2.6. *A self-financing trading strategy φ is called ($\mathbb{P}^*$-) admissible if the relative gains process*

$$\tilde{G}_\varphi(t) = \int\limits_0^t \varphi(u)d\tilde{S}(u)$$

is a ($\mathbb{P}^*$-) martingale. The class of all ($\mathbb{P}^*$-) admissible trading strategies is denoted $\Phi(\mathbb{P}^*)$.

By definition $\tilde{S}$ is a martingale, and $\tilde{G}$ is the stochastic integral with respect to $\tilde{S}$. We see that any sufficiently integrable processes $\varphi_1, \ldots, \varphi_d$ give rise to $\mathbb{P}^*$-admissible trading strategies.

We can repeat the above argument to obtain

Theorem 4.2.2. *The financial market model $\mathcal{M}$ contains no arbitrage opportunities in $\Phi(\mathbb{P}^*)$.*

Under the assumption that no arbitrage opportunities exist, the question of pricing and hedging a contingent claim reduces to the existence of replicating self-financing trading strategies. Formally:

Definition 4.2.7. *(i) A contingent claim X is called attainable if there exists at least one admissible trading strategy such that*

$$V_\varphi(T) = X.$$

We call such a trading strategy φ a replicating strategy for X.
(ii) The financial market model $\mathcal{M}$ is said to be complete if any contingent claim is attainable.

Again we emphasise that this depends on the class of trading strategies. On the other hand, it does not depend on the numéraire: it is an easy exercise in the continuous asset-price process case to show that if a contingent claim is attainable in a given numéraire it is also attainable in any other numéraire and the replicating strategies are the same.

If a contingent claim X is attainable, X can be replicated by a portfolio $\varphi \in \Phi(\mathbb{P}^*)$. This means that holding the portfolio and holding the contingent claim are equivalent from a financial point of view. In the absence of arbitrage the (arbitrage) price process $\Pi_X(t)$ of the contingent claim must therefore satisfy

$$\Pi_X(t) = V_\varphi(t).$$

Of course the questions arise of what will happen if X can be replicated by more than one portfolio, and what the relation of the price process to the equivalent martingale measure(s) is. The following central theorem is the key to answering these questions:

Theorem 4.2.3 (Risk-Neutral Valuation Formula). *The arbitrage price process of any attainable claim is given by the risk-neutral valuation formula*

$$\Pi_X(t) = S_0(t) \mathbb{E}_{\mathbb{P}^*} \left[\left. \frac{X}{S_0(T)} \right| \mathcal{F}_t \right].\tag{4.6}$$

The uniqueness question is immediate from the above theorem:

Corollary 4.2.1. *For any two replicating portfolios $\varphi, \psi \in \Phi(\mathbb{P}^*)$ we have*

$$V_\varphi(t) = V_\psi(t).$$

Proof of Theorem 4.2.3 Since X is attainable, there exists a replicating strategy $\varphi \in \Phi(\mathbb{P}^*)$ such that $V_\varphi(T) = X$ and $\Pi_X(t) = V_\varphi(t)$. Since $\varphi \in \Phi(\mathbb{P}^*)$ the discounted value process $\tilde{V}_\varphi(t)$ is a martingale, and hence

$$
\begin{aligned}
\Pi_X(t) &= V_\varphi(t) = S_0(t)\tilde{V}_\varphi(t) \\[2mm]
&= S_0(t)\mathbb{E}_{\mathbb{P}^*}\left[\left. \tilde{V}_\varphi(T) \right| \mathcal{F}_t \right] = S_0(t)\mathbb{E}_{\mathbb{P}^*}\left[\left. \frac{V_\varphi(T)}{S_0(T)} \right| \mathcal{F}_t \right] \\[2mm]
&= S_0(t)\mathbb{E}_{\mathbb{P}^*}\left[\left. \frac{X}{S_0(T)} \right| \mathcal{F}_t \right].
\end{aligned}
$$

$\blacksquare$

4.2.4 The Black-Scholes Model

The Model

We concentrate on the classical Black-Scholes model

$$
\begin{aligned}
dB(t) &= rB(t)dt, & B(0) &= 1, \\
dS(t) &= S(t)\left(bdt + \sigma dW(t)\right), & S(0) &= p \in (0, \infty),
\end{aligned}
$$

with constant coefficients $b \in \mathbb{R}$, $r, \sigma \in \mathbb{R}_+$. We write as usual $\tilde{S}(t) = S(t)/B(t)$ for the discounted stock price process (with the bank account being the natural numéraire), and get from Itô's formula

$$d\tilde{S}(t) = \tilde{S}(t)\left\{(b - r)dt + \sigma dW(t)\right\}.$$

Equivalent Martingale Measure

Because we use the Brownian filtration any pair of equivalent probability measures $\mathbb{P} \sim \mathbb{Q}$ on $\mathcal{F}_T$ is a Girsanov pair, i.e.

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = L(t)$$

with

$$L(t) = \exp\left\{ -\int_0^t \gamma(s)dW(s) - \frac{1}{2}\int_0^t \gamma(s)^2 ds \right\},$$

and $(\gamma(t) : 0 \leq t \leq T)$ a measurable, adapted d-dimensional process with $\int_0^T \gamma(t)^2 dt < \infty \quad a.s..$ By Girsanov's theorem 4.1.4 we have

$$dW(t) = d\tilde{W}(t) - \gamma(t)dt,$$

where $\tilde{W}$ is a $\mathbb{Q}$-Wiener process. Thus the $\mathbb{Q}$-dynamics for $\tilde{S}$ are

$$d\tilde{S}(t) = \tilde{S}(t)\left\{(b - r - \sigma\gamma(t))dt + \sigma d\tilde{W}(t)\right\}.$$

Since $\tilde{S}$ has to be a martingale under $\mathbb{Q}$ we must have

$$b - r - \sigma\gamma(t) = 0 \quad t \in [0, T],$$

and so we must choose

$$\gamma(t) \equiv \gamma = \frac{b - r}{\sigma},$$

(the 'market price of risk'). Indeed, this argument leads to a unique martingale measure, and we will make use of this fact later on. Using the product rule, we find the $\mathbb{Q}$-dynamics of S as

$$dS(t) = S(t)\left\{rdt + \sigma d\tilde{W}\right\}.$$

We see that the appreciation rate b is replaced by the interest rate r, hence the terminology risk-neutral (or yield-equating) martingale measure.

We also know that we have a unique martingale measure $\mathbb{P}^*$ (recall $\gamma = (b-r)/\sigma$ in Girsanov's transformation).

Pricing and Hedging Contingent Claims

Recall that a contingent claim X is a $\mathcal{F}_T$-measurable random variable such that $X/B(T) \in L^1(\Omega, \mathcal{F}_T, \mathbb{P}^*)$. (We write $\mathbb{E}^*$ for $\mathbb{E}_{\mathbb{P}^*}$ in this section.) By the risk-neutral valuation principle the price of a contingent claim X is given by

$$\Pi_X(t) = e^{\{-r(T-t)\}}\mathbb{E}^*\left[X|\mathcal{F}_t\right],$$

with $\mathbb{E}^*$ given via the Girsanov density

$$L(t) = \exp\left\{-\left(\frac{b-r}{\sigma}\right)W(t) - \frac{1}{2}\left(\frac{b-r}{\sigma}\right)^2 t\right\}.$$

Now consider a European call with strike K and maturity T on the stock S (so $\Phi(T) = (S(T) - K)^+$), we can evaluate the above expected value (which is easier than solving the Black-Scholes partial differential equation) and obtain:

Proposition 4.2.3 (Black-Scholes Formula). *The Black-Scholes price process of a European call is given by*

$$C(t) = S(t)N(d_1(S(t), T - t)) - Ke^{-r(T-t)}N(d_2(S(t), T - t)). \tag{4.7}$$

The functions $d_1(s, t)$ and $d_2(s, t)$ are given by

$$d_1(s, t) = \frac{\log(s/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}},$$

$$d_2(s, t) = d_1(s, t) - \sigma\sqrt{t} = \frac{\log(s/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$$

Observe that we have already deduced this formula as a limit of a discrete-time setting.

To obtain a replicating portfolio we start in the discounted setting. We have from the risk-neutral valuation principle

$$M(t) = \exp\{-rT\}\mathbb{E}^*\left[\Phi(S(T))|\mathcal{F}_t\right].$$

Now we use Itô's lemma to find the dynamics of the $\mathbb{P}^*$-martingale $M(t) = G(t, S(t))$:

$$dM(t) = \sigma S(t)G_s(t, S(t))d\tilde{W}(t).$$

Using this representation, we get for the stock component of the replicating portfolio

$$h(t) = \sigma S(t) G_s(t, S(t)).$$

Now for the discounted assets the stock component is

$$\varphi_1(t) = G_s(t, S(t)) B(t),$$

and using the self-financing condition the cash component is

$$\varphi_0(t) = G(t, S(t)) - G_s(t, S(t)) S(t).$$

To transfer this portfolio to undiscounted values we multiply it by the discount factor, i.e $F(t, S(t))$- $= B(t) G(t, S(t))$ and get:

Proposition 4.2.4. *The replicating strategy in the classical Black-Scholes model is given by*

$$\varphi_0 = \frac{F(t, S(t)) - F_s(t, S(t)) S(t)}{B(t)},$$

$$\varphi_1 = F_s(t, S(t)).$$

We can also use an arbitrage approach to derive the Black-Scholes formula. For this consider a self-financing portfolio which has dynamics

$$dV_\varphi(t) = \varphi_0(t) dB(t) + \varphi_1(t) dS(t) = (\varphi_0(t) r B(t) + \varphi_1(t) \mu S(t)) dt + \varphi_1(t) \sigma S(t) dW(t).$$

Assume that the portfolio value can be written as

$$V_\varphi(t) = V(t) = f(t, S(t))$$

for a suitable function $f \in C^{1,2}$. Then by Itô's formula

$$dV(t) = (f_t(t, S_t) + f_x(t, S_t) S_t \mu + \frac{1}{2} S_t^2 \sigma^2 f_{xx}(t, S_t)) dt + f_x(t, S_t) \sigma S_t dW_t.$$

Now we match the coefficients and find

$$\varphi_1(t) = f_x(t, S_t)$$

and

$$\varphi_0(t) = \frac{1}{r B(t)} (f_t(t, S_t) + \frac{1}{2} \sigma^2 S_t^2 f_{xx}(t, S_t)).$$

Then looking at the total portfolio value we find that $f(t, x)$ must satisfy the following PDE

$$f_t(t, x) + r x f_x(t, x) + \frac{1}{2} \sigma^2 x^2 f_{xx}(t, x) - r f(t, x) = 0 \tag{4.8}$$

and initial condition $f(T, x) = (x - K)^+$.

In their original paper Black and Scholes (1973), Black and Scholes used an arbitrage pricing approach (rather than our risk-neutral valuation approach) to deduce the price of a European call as the solution of a partial differential equation (we call this the PDE approach). The idea is as follows: start by assuming that the option price $C(t)$ is given by $C(t) = f(t, S(t))$ for some sufficiently smooth function $f : I\!R_+ \times [0, T] \to I\!R$. By Itô's formula (Theorem 4.1.3) we find for the dynamics of the option price process (observe that we work under $I\!P$ so $dS = S(b dt + \sigma dW)$)

$$dC = \left\{ f_t(t, S) + f_s(t, S) S b + \frac{1}{2} f_{ss}(t, S) S^2 \sigma^2 \right\} dt + f_s S \sigma dW. \tag{4.9}$$

Consider a portfolio ψ consisting of a short position in $\psi_1(t) = f_s(t, S(t))$ stocks and a long position in $\psi_2(t) = 1$ call and assume the portfolio is self-financing. Then its value process is

$$V_\psi(t) = -\psi_1(t)S(t) + C(t),$$

and by the self-financing condition we have (to ease the notation we omit the arguments)

$$
\begin{aligned}
dV_\psi &= -\psi_1 dS + dC \\[2mm]
&= -f_s(Sbdt + S\sigma dW) + \left(f_t + f_s Sb + \frac{1}{2} f_{ss} S^2 \sigma^2 \right) dt + f_s S\sigma dW \\[2mm]
&= \left(f_t + \frac{1}{2} f_{ss} S^2 \sigma^2 \right) dt.
\end{aligned}
$$

So the dynamics of the value process of the portfolio do not have any exposure to the driving Brownian motion, and its appreciation rate in an arbitrage-free world must therefore equal the risk-free rate, i.e.

$$dV_\psi(t) = rV_\psi(t)dt = (-rf_s S + rC)\, dt.$$

Comparing the coefficients and using $C(t) = f(t, S(t))$, we must have

$$-rSf_s + rf = f_t + \frac{1}{2}\sigma^2 S^2 f_{ss}.$$

This leads again to the Black-Scholes partial differential equation (4.8) for f, i.e.

$$f_t + rsf_s + \frac{1}{2}\sigma^2 s^2 f_{ss} - rf = 0.$$

Since $C(T) = (S(T) - K)^+$ we need to impose the terminal condition $f(s, T) = (s - K)^+$ for all $s \in \mathbb{R}^+$.

Note.

One point in the justification of the above argument is missing: we have to show that the trading strategy short ψ_1 stocks and long one call is self-financing. In fact, this is not true, since $\psi_1 = \psi_1(t, S(t))$ is dependent on the stock price process. Formally, for the self-financing condition to be true we must have

$$dV_\psi(t) = d(\psi_1(t)S(t)) + dC(t) = \psi_1(t)dS(t) + dC(t).$$

Now $\psi(t) = \psi(t, S(t))$ depends on the stock price and so we have

$$d(\psi_1(t, S(t))S(t)) = \psi_1(t)dS(t) + S(t)d\psi_1(t, S(t)) + d\langle \psi_1, S \rangle(t).$$

We see that the portfolio ψ is self-financing, if

$$S(t)d\psi_1(t, S(t)) + d\langle \psi_1, S \rangle(t) = 0.$$

It is an exercise in Itô calculus to show that this is not the case.

4.2.5 The Greeks

We will now analyse the impact of the underlying parameters in the standard Black-Scholes model on the prices of call and put options. The Black-Scholes option values depend on the (current) stock price, the volatility, the time to maturity, the interest rate and the strike price. The sensitivities of the option price with respect to the first four parameters are called the *Greeks* and are widely

used for hedging purposes. We can determine the impact of these parameters by taking partial derivatives. Recall the Black-Scholes formula for a European call (4.7):

$$C(0) = C(S, T, K, r, \sigma) = SN(d_1(S, T)) - Ke^{-rT}N(d_2(S, T)),$$

with the functions $d_1(s, t)$ and $d_2(s, t)$ given by

$$d_1(s, t) \quad = \quad \frac{\log(s/K) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}},$$

$$d_2(s, t) \quad = \quad d_1(s, t) - \sigma\sqrt{t} = \frac{\log(s/K) + (r - \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}.$$

One obtains

$$\Delta \quad := \quad \frac{\partial C}{\partial S} \quad = \quad N(d_1) > 0,$$

$$\mathcal{V} \quad := \quad \frac{\partial C}{\partial \sigma} \quad = \quad S\sqrt{T}n(d_1) > 0,$$

$$\Theta \quad := \quad \frac{\partial C}{\partial T} \quad = \quad \frac{S\sigma}{2\sqrt{T}}n(d_1) + Kre^{-rT}N(d_2) > 0,$$

$$\rho \quad := \quad \frac{\partial C}{\partial r} \quad = \quad TKe^{-rT}N(d_2) > 0,$$

$$\Gamma \quad := \quad \frac{\partial^2 C}{\partial S^2} \quad = \quad \frac{n(d_1)}{S\sigma\sqrt{T}} > 0.$$

(As usual N is the cumulative normal distribution function and n is its density.) From the definitions it is clear that Δ – delta – measures the change in the value of the option compared with the change in the value of the underlying asset, $\mathcal{V}$ – vega – measures the change of the option compared with the change in the volatility of the underlying, and similar statements hold for Θ – theta – and ρ – rho (observe that these derivatives are in line with our arbitrage-based considerations in §1.3). Furthermore, Δ gives the number of shares in the replication portfolio for a call option (see Proposition 4.2.4), so Γ measures the sensitivity of our portfolio to the change in the stock price.

The Black-Scholes partial differential equation (4.8) can be used to obtain the relation between the Greeks, i.e. (observe that Θ is the derivative of C, the price of a European call, with respect to the time to expiry $T - t$, while in the Black-Scholes PDE the partial derivative with respect to the current time t appears)

$$rC = \frac{1}{2}s^2\sigma^2\Gamma + rs\Delta - \Theta.$$

Let us now compute the dynamics of the call option's price $C(t)$ under the risk-neutral martingale measure $I\!\!P^*$. Using formula (4.9) we find

$$dC(t) = rC(t)dt + \sigma N(d_1(S(t), T - t))S(t)d\tilde{W}(t).$$

Defining the *elasticity coefficient* of the option's price as

$$\eta^c(t) = \frac{\Delta(S(t), T - t)S(t)}{C(t)} = \frac{N(d_1(S(t), T - t))}{C(t)}$$

we can rewrite the dynamics as

$$dC(t) = rC(t)dt + \sigma\eta^c(t)C(t)d\tilde{W}(t).$$

So, as expected in the risk-neutral world, the appreciation rate of the call option equals the risk-free rate r. The volatility coefficient is $\sigma\eta^c$, and hence stochastic. It is precisely this feature that causes difficulties when assessing the impact of options in a portfolio.

4.2.6 Barrier Options

The question of whether or not a particular stock will attain a particular level within a specified period has long been an important one for risk managers. From at least 1967 – predating both CBOE and Black-Scholes in 1973 – practitioners have sought to reduce their exposure to specific risks of this kind by buying options designed with such barrier-crossing events in mind. As usual, the motivation is that buying specific options – that is, taking out specific insurance – is a cheaper way of covering oneself against a specific danger than buying a more general one.

One-barrier options specify a stock-price level, H say, such that the option pays ('knocks in') or not ('knocks out') according to whether or not level H is attained, from below ('up') or above ('down'). There are thus four possibilities: 'up and in', 'up and out', 'down and in' and 'down and out'. Since barrier options are path-dependent (they involve the behaviour of the path, rather than just the current price or price at expiry), they may be classified as exotic; alternatively, the four basic one-barrier types above may be regarded as 'vanilla barrier' options, with their more complicated variants, described below, as 'exotic barrier' options. Note that holding both a knock-in option and the corresponding knock-out is equivalent to the corresponding vanilla option with the barrier removed. The sum of the prices of the knock-in and the knock-out is thus the price of the vanilla – again showing the attractiveness of barrier options as being cheaper than their vanilla counterparts.

A barrier option is often designed to pay a *rebate* – a sum specified in advance – to compensate the holder if the option is rendered otherwise worthless by hitting/not hitting the barrier. We restrict attention to zero rebate here for simplicity.

Consider, to be specific, a down-and-out call option with strike K and barrier H (the other possibilities may be handled similarly). The payoff is (unless otherwise stated min and max are over $[0, T]$)

$$(S(T) - K)^+ \mathbf{1}_{\{\min S(.) \geq H\}} = (S(T) - K)\mathbf{1}_{\{S(T) \geq K, \min S(.) \geq H\}},$$

so by risk-neutral pricing the value of the option is

$$DOC_{K,H} := \mathbb{E}\left[e^{-rT}(S(T) - K)\mathbf{1}_{\{S(T) \geq K, \min S(.) \geq H\}} \right],$$

where S is geometric Brownian motion, $S(t) = p_0 \exp\{(\mu - \frac{1}{2}\sigma^2 t) + \sigma W(t)\}$. Write $c := \mu - \frac{1}{2}\sigma^2/\sigma$; then $\min S(.) \geq H$ iff $\min(ct + W(t)) \geq \sigma^{-1}\log(H/p_0)$. Writing X for $X(t) := ct + W(t)$ – drifting Brownian motion with drift c, m, M for its minimum and maximum processes

$$m(t) := \min\{X(s) : s \in [0, t]\}, \quad M(t) := \max\{X(s) : s \in [0, t]\},$$

the payoff function involves the bivariate process (X, m), and the option price involves the joint law of this process.

Consider first the case $c = 0$: we require the joint law of standard Brownian motion and its maximum or minimum, (W, M) or (W, m). Taking (W, M) for definiteness, we start the Brownian motion W at the origin at time zero, choose a level $b > 0$, and run the process until the *first-passage time* (see Exercise 5.2)

$$\tau(b) := \inf\{t \geq 0 : W(t) \geq b\}$$

at which the level b is first attained. This is a stopping time, and we may use the strong Markov property for W at time $\tau(b)$. The process now begins afresh at level b, and by symmetry the probabilistic properties of its further evolution are invariant under *reflection* in the level b (thought of as a mirror). This *reflection principle* leads to the joint density of $(W(t), M(t))$ as

$$\mathbb{P}_0\left(W(t) \in dx, M(t) \in dy\right)$$

$$= \frac{2(2y - x)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{1}{2}(2y - x)^2/t\right\} \quad (0 \leq x \leq y),$$

a formula due to Lévy. (Lévy also obtained the identity in law of the bivariate processes $(M(t) - W(t), M(t))$ and $(|W(t)|, L(t))$, where L is the local time process of W at zero: see e.g. Revuz

and Yor (1991), VI.2). The idea behind the reflection principle goes back to work of Désiré André in 1887, and indeed further, to the *method of images* of Lord Kelvin (1824-1907), then Sir William Thomson, of 1848 on electrostatics. For background on this, see any good book on electromagnetism, e.g. Jeans (1925), Chapter VIII.

Lévy's formula for the joint density of $(W(t), M(t))$ may be extended to the case of general drift c by the usual method for changing drift, Girsanov's theorem. The general result is

$$\mathbb{P}_0\left(X(t) \in dx, M(t) \in dy\right)$$

$$= \frac{2(2y-x)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(2y-x)^2}{2t} + cx - \frac{1}{2}c^2 t\right\} \quad (0 \le x \le y).$$

See e.g. Rogers and Williams (1994), I, (13.10), or Harrison (1985), §1.8. As an alternative to the probabilistic approach above, a second approach to this formula makes explicit use of Kelvin's language – mirrors, sources, sinks; see e.g. Cox and Miller (1972), §5.7.

Given such an explicit formula for the joint density of $(X(t), M(t))$ – or equivalently, $(X(t), m(t))$ – we can calculate the option price by integration. The factor $S(T) - K$, or $S - K$, gives rise to two terms, in S and K, while the integrals, involving relatives of the normal density function n, may be obtained explicitly in terms of the normal distribution function N – both features familiar from the Black-Scholes formula. Indeed, this resemblance makes it convenient to decompose the price $DOC_{K,H}$ of the down-and-out call into the (Black-Scholes) price of the corresponding vanilla call, C_K say, and the *knockout discount*, $KOD_{K,H}$ say, by which the knockout barrier at H lowers the price:

$$DOC_{K,H} = C_K - KOD_{K,H}.$$

The details of the integration, which are tedious, are omitted; the result is, writing $\lambda := r - \frac{1}{2}\sigma^2$,

$$KOD_{K,H} = p_0(H/p_0)^{2+2\lambda/\sigma^2} N(c_1) - Ke^{-rT}(H/p_0)^{2\lambda/\sigma^2} N(c_2),$$

where p_0 is the initial stock price as usual and c_1, c_2 are functions of the price $p = p_0$ and time $t = T$ given by

$$c_{1,2}(p,t) = \frac{\log(H^2/pK) + (r \pm \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}$$

(the notation is that of the excellent text Musiela and Rutkowski (1997), §9.6, to which we refer for further detail). The other cases of vanilla barrier options, and their sensitivity analysis, are given in detail in Zhang (1997), Chapter 10.

Chapter 5

Interest Rate Theory

5.1 The Bond Market

5.1.1 The Term Structure of Interest Rates

We start with a heuristic discussion, which we will formalize in the following section. The main traded objects we consider are zero-coupon bonds. A zero-coupon bond is a bond that has no coupon payments. The price of a zero-coupon bond at time t that pays, say, a sure £ at time $T \geq t$ is denoted $p(t, T)$. All zero-coupon bonds are assumed to be default-free and have strictly positive prices. Various different interest rates are defined in connection with zero-coupon bonds, but we will only consider continuously compounded interest rates (which facilitates theoretical considerations).

Using the arbitrage pricing technique, we easily obtain pricing formulas for coupon bonds. Coupon bonds are bonds with regular interest payments, called coupons, plus a principal repayment at maturity. Let c_j be the payments at times t_j, $j = 1, \ldots, n$, F be the face value paid at time t_n. Then the price of the coupon bond B_c must satisfy

$$B_c = \sum_{j=1}^{n} c_j p(0, t_j) + F p(0, t_n). \tag{5.1}$$

Hence, we see that a coupon bond is equivalent to a portfolio of zero-coupon bonds.

The yield-to-maturity is defined as an interest rate per annum that equates the present value of future cash flows to the current market value. Using continuous compounding, the yield-to-maturity y_c is defined by the relation

$$B_c = \sum_{j=1}^{n} c_j \exp\{-y_c t_j\} + F \exp\{-y_c t_n\}.$$

If for instance the t_j, $j = 1, \ldots, n$ are expressed in years, then y_c is an annual continuously compounded yield-to-maturity (with the continuously compounded annual interest rate, $r(T)$, defined by the relation $p(0, T) = \exp\{-r(T)(T/365)\}$).

The term structure of interest rates is defined as the relationship between the yield-to-maturity on a zero-coupon bond and the bond's maturity. Normally, this yields an upward sloping curve (as in figure 5.1), but flat and downward sloping curves have also been observed.

In constructing the term structure of interest rates, we face the additional problem that in most economies no zero-coupon bonds with maturity greater than one year are traded (in the USA, Treasury bills are only traded with maturity up to one year). We can, however, use prices of coupon bonds and invert formula (5.1) for zero-coupon prices. In practice, additional complications arise, since the maturities of coupon bonds are not equally spaced and trading in bonds with some maturities may be too thin to give reliable prices. We refer the reader to Jarrow and Turnbull (2000) for further discussion of these issues.
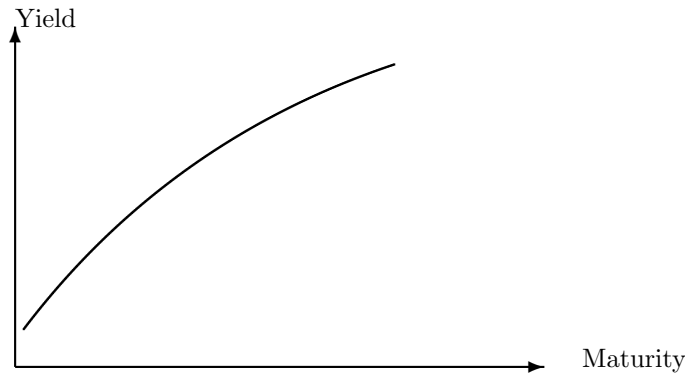
Figure 5.1: Yield curve

5.1.2 Mathematical Modelling

Let $(\Omega, \mathcal{F}, I\!\!P, I\!\!F)$ be a filtered probability space with a filtration $I\!\!F = (\mathcal{F}_t)_{t \leq T^*}$ satisfying the usual conditions (used to model the flow of information) and fix a terminal time horizon T^*. We assume that all processes are defined on this probability space. The basic building blocks for our relative pricing approach, zero-coupon bonds, are defined as follows.

Definition 5.1.1. *A zero-coupon bond with maturity date T, also called a T-bond, is a contract that guarantees the holder a cash payment of one unit on the date T. The price at time t of a bond with maturity date T is denoted by $p(t,T)$.*

Obviously we have $p(t,t) = 1$ for all t. We shall assume that the price process $p(t,T)$, $t \in [0,T]$ is adapted and strictly positive and that for every fixed t, $p(t,T)$ is continuously differentiable in the T variable.

Based on arbitrage considerations (recall our basic aim is to construct a market model that is free of arbitrage), we now define several risk-free interest rates. Given three dates $t < T_1 < T_2$ the basic question is: what is the risk-free rate of return, determined at the contract time t, over the interval $[T_1, T_2]$ of an investment of 1 at time T_1? To answer this question we consider the arbitrage Table 5.1 below (compare §1.3 for the use of arbitrage tables).

Time	t	T_1	T_2
	Sell T_1 bond Buy $\frac{p(t,T_1)}{p(t,T_2)}$ T_2 bonds	Pay out 1	Receive $\frac{p(t,T_1)}{p(t,T_2)}$
Net investment	0	-1	$+\frac{p(t,T_1)}{p(t,T_2)}$

Table 5.1: Arbitrage table for forward rates

To exclude arbitrage opportunities, the equivalent constant rate of interest R over this period (we pay out 1 at time T_1 and receive $e^{R(T_2 - T_1)}$ at T_2) has thus to be given by

$$e^{R(T_2 - T_1)} = \frac{p(t,T_1)}{p(t,T_2)}.$$

We formalize this in:

Definition 5.1.2. *(i) The forward rate for the period $[T_1, T_2]$ as seen at time at t is defined as*

$$R(t; T_1, T_2) = -\frac{\log p(t, T_2) - \log p(t, T_1)}{T_2 - T_1}.$$

(ii) The spot rate $R(T_1, T_2)$, for the period $[T_1, T_2]$ is defined as

$$R(T_1, T_2) = R(T_1; T_1, T_2).$$

(iii) The instantaneous forward rate with maturity T, at time t, is defined by

$$f(t, T) = -\frac{\partial \log p(t, T)}{\partial T}.$$

(iv) The instantaneous short rate at time t is defined by

$$r(t) = f(t, t).$$

Definition 5.1.3. *The money account process is defined by*

$$B(t) = \exp\left\{\int_0^t r(s)ds\right\}.$$

The interpretation of the money market account is a strategy of instantaneously reinvesting at the current short rate.

Lemma 5.1.1. *For $t \leq s \leq T$ we have*

$$p(t, T) = p(t, s) \exp\left\{-\int_s^T f(t, u)du\right\},$$

and in particular

$$p(t, T) = \exp\left\{-\int_t^T f(t, s)ds\right\}.$$

In what follows, we model the above processes in a generalized Black-Scholes framework. That is, we assume that $W = (W_1, \ldots, W_d)$ is a standard d-dimensional Brownian motion and the filtration $I\!\!F$ is the augmentation of the filtration generated by $W(t)$. The dynamics of the various processes are given as follows:

Short-rate Dynamics:

$$dr(t) = a(t)dt + b(t)dW(t), \tag{5.2}$$

Bond-price Dynamics:

$$dp(t, T) = p(t, T)\{m(t, T)dt + v(t, T)dW(t)\}, \tag{5.3}$$

Forward-rate Dynamics:

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW(t). \tag{5.4}$$

We assume that in the above formulas, the coefficients meet standard conditions required to guarantee the existence of the various processes – that is, existence of solutions of the various stochastic differential equations. Furthermore, we assume that the processes are smooth enough to allow differentiation and certain operations involving changing of order of integration and differentiation. Since the actual conditions are rather technical, we refer the reader to Björk (1997), Heath, Jarrow, and Morton (1992) and Protter (2004) (the latter reference for the stochastic Fubini theorem) for these conditions.

Following Björk (1997) for formulation and proof, we now give a small toolbox for the relationships between the processes specified above.

Proposition 5.1.1. *(i) If $p(t,T)$ satisfies (5.3), then for the forward-rate dynamics we have*

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW(t),$$

where α and σ are given by

$$\begin{cases} \alpha(t,T) & = v_T(t,T)v(t,T) - m_T(t,T), \\ \sigma(t,T) & = -v_T(t,T). \end{cases}$$

(ii) If $f(t,T)$ satisfies (5.4), then the short rate satisfies

$$dr(t) = a(t)dt + b(t)dW(t),$$

where a and b are given by

$$\begin{cases} a(t) & = f_T(t,t) + \alpha(t,t), \\ b(t) & = \sigma(t,t). \end{cases} \tag{5.5}$$

(iii) If $f(t,T)$ satisfies (5.4), then $p(t,T)$ satisfies

$$dp(t,T) = p(t,T)\left\{ r(t) + A(t,T) + \frac{1}{2}\|S(t,T)\|^2 \right\}dt + p(t,T)S(t,T)dW(t),$$

where

$$A(t,T) = -\int_t^T \alpha(t,s)ds, \quad S(t,T) = -\int_t^T \sigma(t,s)ds. \tag{5.6}$$

Proof. To prove (i) we only have to apply Itô's formula to the defining equation for the forward rates.

To prove (ii) we start by integrating the forward-rate dynamics. This leads to

$$f(t,t) = r(t) = f(0,t) + \int_0^t \alpha(s,t)ds + \int_0^t \sigma(s,t)dW(s). \tag{5.7}$$

Writing also α and σ in integrated form

$$\alpha(s,t) \quad = \quad \alpha(s,s) + \int_s^t \alpha_T(s,u)du,$$

$$\sigma(s,t) \quad = \quad \sigma(s,s) + \int_s^t \sigma_T(s,u)du,$$

and inserting this into (5.7), we find

$$
\begin{aligned}
r(t) \quad = \quad & f(0,t) + \int_0^t \alpha(s,s)ds + \int_0^t \int_s^t \alpha_T(s,u)duds \\
& + \int_0^t \sigma(s,s)dW(s) + \int_0^t \int_s^t \sigma_T(s,u)dudW(s).
\end{aligned}
$$

After changing the order of integration we can identify terms to establish (ii).

For (iii) we use a technique from Heath, Jarrow, and Morton (1992); compare Björk (1997). By the definition of the forward rates we may write the bond-price process as

$$
p(t,T) = \exp\{Z(t,T)\},
$$

where Z is given by

$$
Z(t,T) = -\int_t^T f(t,s)ds. \tag{5.8}
$$

Again we write (5.4) in integrated form:

$$
f(t,s) = f(0,s) + \int_0^t \alpha(u,s)dt + \int_0^t \sigma(u,s)dW(u).
$$

We insert the integrated form in (5.8) to get

$$
Z(t,T) = -\int_t^T f(0,s)ds - \int_0^t \int_t^T \alpha(u,s)dsdu - \int_0^t \int_t^T \sigma(u,s)dsdW(u).
$$

Now, splitting the integrals and changing the order of integration gives us

$$
\begin{aligned}
Z(t,T) \quad = \quad & -\int_0^T f(0,s)ds - \int_0^t \int_u^T \alpha(u,s)dsdu - \int_0^t \int_u^T \sigma(u,s)dsdW(u) \\
& + \int_0^t f(0,s)ds + \int_0^t \int_u^t \alpha(u,s)dsdu + \int_0^t \int_u^t \sigma(u,s)dsdW(u) \\
= \quad & Z(0,T) - \int_0^t \int_u^T \alpha(u,s)dsdu - \int_0^t \int_u^T \sigma(u,s)dsdW(u) \\
& + \int_0^t f(0,s)ds + \int_0^t \int_0^s \alpha(u,s)duds + \int_0^t \int_0^s \sigma(u,s)dW(u)ds.
\end{aligned}
$$

The last line is just the integrated form of the forward-rate dynamics (5.4) over the interval $[0,s]$. Since $r(s) = f(s,s)$, this last line above equals $\int_0^t r(s)ds$. So we obtain

$$
Z(t,T) = Z(0,T) + \int_0^t r(s)ds - \int_0^t \int_u^T \alpha(u,s)dsdu - \int_0^t \int_u^T \sigma(u,s)dsdW(u).
$$

Using A and S from (5.6), the stochastic differential of Z is given by

$$
dZ(t,T) = (r(t) + A(t,T))dt + S(t,T)dW(t).
$$

Now we can apply Itô's lemma to the process $p(t,T) = \exp\{Z(t,T)\}$ to complete the proof. $\blacksquare$

5.1.3 Bond Pricing, Martingale Measures and Trading Strategies

We will now examine the mathematical structure of our bond-market model in more detail. As usual, our first task is to find a convenient characterization of the no-arbitrage assumption. By Theorem 4.2.1, absence of arbitrage is guaranteed by the existence of an equivalent martingale measure $\mathbb{Q}$. Recall that by definition an equivalent martingale measure has to satisfy $\mathbb{Q} \sim \mathbb{P}$ and the discounted price processes (with respect to a suitable numéraire) of the basic securities have to be $\mathbb{Q}$-martingales. For the bond market this implies that all zero-coupon bonds with maturities $0 \le T \le T^*$ have to be martingales.

More precisely, taking the risk-free bank account $B(t)$ as numéraire we have

Definition 5.1.4. *A measure $\mathbb{Q} \sim \mathbb{P}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is an equivalent martingale measure for the bond market, if for every fixed $0 \le T \le T^*$ the process*

$$\frac{p(t,T)}{B(t)}, \quad 0 \le t \le T$$

is a $\mathbb{Q}$-martingale.

Assume now that there exists at least one equivalent martingale measure, say $\mathbb{Q}$. Defining contingent claims as $\mathcal{F}_T$-measurable random variables such that $X/B(T) \in L^1(\mathcal{F}_T, \mathbb{Q})$ with some $0 \le T \le T^*$ (notation: T-contingent claims), we can use the risk-neutral valuation principle (4.6) to obtain:

Proposition 5.1.2. *Consider a T-contingent claim X. Then the price process $\Pi_X(t)$, $0 \le t \le T$ of the contingent claim is given by*

$$\Pi_X(t) = B(t)\mathbb{E}_{\mathbb{Q}}\left[\left.\frac{X}{B(T)}\right| \mathcal{F}_t\right] = \mathbb{E}_{\mathbb{Q}}\left[\left.Xe^{-\int_t^T r(s)ds}\right| \mathcal{F}_t\right].$$

In particular, the price process of a zero-coupon bond with maturity T is given by

$$p(t,T) = \mathbb{E}_{\mathbb{Q}}\left[\left.e^{-\int_t^T r(s)ds}\right| \mathcal{F}_t\right].$$

Proof. We just have to apply Theorem 4.2.3.

We thus see that the relevant dynamics of the price processes are those given under a martingale measure $\mathbb{Q}$. The implication for model building is that it is natural to model all objects directly under a martingale measure $\mathbb{Q}$. This approach is called martingale modelling. The price one has to pay in using this approach lies in the statistical problems associated with parameter estimation.

5.2 Short-rate Models

Following our introductory remarks, we now look at models of the short rate of the type

$$dr(t) = a(t, r(t))dt + b(t, r(t))dW(t), \tag{5.9}$$

with functions a, b sufficiently regular and W a real-valued Brownian motion.

The crucial point in this setting is the assumption on the probability measure under which the short rate is modelled.

If we model under the objective probability measure $\mathbb{P}$ and assume that a locally risk-free asset B (the money market) exists, we face the question whether in an arbitrage-free market bond prices – quite naturally viewed as derivatives with the short rate as underlying – are uniquely determined. In contrast to the equity market setting, with a risky asset and a risk-free asset available for trading, the short rate r is not the price of a traded asset, and hence we only can

set up portfolios consisting of putting money in the bank account. We thus face an incomplete market situation, and the best we can hope for is to find consistency requirements for bonds of different maturity. Given a single 'benchmark' bond, we should then be able to price all other bonds relative to this given bond.

On the other hand, if we assume that the short rate is modelled under an equivalent martingale measure, we can immediately price all contingent claims via the risk-neutral valuation formula. The drawback in this case is the question of calibrating the model (we do not observe the parameters of the process under an equivalent martingale measure, but rather under the objective probability measure!).

5.2.1 The Term-structure Equation

Let us assume that the short-rate dynamics satisfy (5.9) under the historical probability measure $I\!P$. In our Brownian setting, we know that each equivalent martingale measure Q is given in terms of a Girsanov density

$$L(t) = \exp\left\{ -\int_0^t \gamma(u)dW(u) - \frac{1}{2}\int_0^t \gamma(u)^2 du \right\}, \quad 0 \le t \le T.$$

Assume now that γ is given as $\gamma(t) = \lambda(t, r(t))$, with a sufficiently smooth function λ. (We will use the notation $Q(\lambda)$ to emphasize the dependence of the equivalent martingale measure on λ). By Girsanov's Theorem 4.1.4, we know that $\tilde{W} = W + \int \lambda dt$ is a $Q(\lambda)$-Brownian motion. So the $Q(\lambda)$-dynamics of r are given by

$$dr(t) = \{a(t, r(t)) - b(t, r(t))\lambda(t, r(t))\}dt + b(t, r(t))d\tilde{W}(t).$$

Now consider a T-contingent claim $X = \Phi(r(T))$, for some sufficiently smooth function $\Phi : I\!R \to I\!R^+$. We know that using the risk-neutral valuation formula we obtain arbitrage-free prices for any contingent claim by applying the expectation operator under an equivalent martingale measure (to the discounted time T value). An slight modification of the argument used to find the Black-Scholes PDE yields, for any $Q(\lambda)$,

$$I\!E_{Q(\lambda)}\left[e^{-\int_t^T r(u)du}\Phi(r(T)) \Big| \mathcal{F}_t \right] = F(t, r(t)),$$

where $F : [0, T^*] \times I\!R \to I\!R$ satisfies the partial differential equation (we omit the arguments (t, r))

$$F_t + (a - b\lambda)F_r + \frac{1}{2}b^2 F_{rr} - rF = 0 \tag{5.10}$$

and terminal condition $F(T, r) = \Phi(r)$, for all $r \in I\!R$. Suppose now that the price process $p(t, T)$ of a T-bond is determined by the assessment, at time t, of the segment $\{r(\tau), t \le \tau \le T\}$ of the short rate process over the term of the bond. So we assume

$$p(t, T) = F(t, r(t); T),$$

with F a sufficiently smooth function.

Since we know that the value of a zero-coupon bond is one unit at maturity, we have the terminal condition $F(T, r; T) = 1$. Thus we have

Proposition 5.2.1 (Term-structure Equation). *If there exists an equivalent martingale measure of type $Q(\lambda)$ for the bond market (implying that the no-arbitrage condition holds) and the price processes $p(t, T)$ of T-bonds are given are given by a sufficiently smooth function F as above, then F will satisfy the partial differential equation (5.10) with terminal condition $F(T, r; T) = 1$.*

5.2.2 Martingale Modelling

We now fix an equivalent martingale measure $\mathbf{Q}$ (which we assume to exist), and return to modelling the short-rate dynamics directly under $\mathbf{Q}$. Thus we assume that r has $\mathbf{Q}$-dynamics

$$dr(t) = a(t, r(t))dt + b(t, r(t))dW(t) \tag{5.9}$$

with W a (real-valued) $\mathbf{Q}$-Wiener process. We can immediately apply the risk-neutral valuation technique to obtain the price process $\Pi_X(t)$ of any sufficiently integrable T-contingent claim X by computing the $\mathbf{Q}$-expectation, i.e.

$$\Pi_X(t) = \mathbb{E}_{\mathbf{Q}}\left[e^{-\int_t^T r(u)du}X|\mathcal{F}_t\right]. \tag{5.11}$$

If, additionally, the contingent claim is of the form $X = \Phi(r(T))$ with a sufficiently smooth function Φ, we obtain

Proposition 5.2.2 (Term-structure Equation). *Consider T-contingent claims of the form $X = \Phi(r(T))$. Then arbitrage-free price processes are given by $\Pi_X(t) = F(t, r(t))$, where F is the solution of the partial differential equation*

$$F_t + aF_r + \frac{b^2}{2}F_{rr} - rF = 0 \tag{5.12}$$

with terminal condition $F(T, r) = \Phi(r)$ for all $r \in \mathbb{R}$. In particular, T-bond prices are given by $p(t, T) = F(t, r(t); T)$, with F solving (5.12) and terminal condition $F(T, r; T) = 1$.

Suppose we want to evaluate the price of a European call option with maturity S and strike K on an underlying T-bond. This means we have to price the S-contingent claim

$$X = \max\{p(S, T) - K, 0\}.$$

We first have to find the price process $p(t, T) = F(t, r; T)$ by solving (5.12) with terminal condition $F(T, r; T) = 1$. Secondly, we use the risk-neutral valuation principle to obtain $\Pi_X(t) = G(t, r)$, with G solving

$$G_t + aG_r + \frac{b^2}{2}G_{rr} - rG = 0 \quad \text{and} \quad G(S, r) = \max\{F(S, r; T) - K, 0\}, \ \forall r \in \mathbb{R}.$$

So we are clearly in need of efficient methods of solving the above partial differential equations, or from a modelling point of view, we need short-rate models that facilate this computational task. Fortunately, there is a class of models, exhibiting an *affine term structure (ATS)*, which allows for simplification.

Definition 5.2.1. *If bond prices are given as*

$$p(t, T) = \exp\{A(t, T) - B(t, T)r\}, \quad 0 \le t \le T,$$

with $A(t, T)$ and $B(t, T)$ are deterministic functions, we say that the model possesses an affine term structure.

Assuming that we have such a model in which both a and b^2 are affine in r, say $a(t, r) = \alpha(t) - \beta(t)r$ and $b(t, r) = \sqrt{\gamma(t) + \delta(t)r}$, we find that A and B are given as solutions of ordinary differential equations,

$$A_t(t, T) - \alpha(t)B(t, T) + \frac{\gamma(t)}{2}B^2(t, T) = 0, \quad A(T, T) = 0,$$

$$(1 + B_t(t, T)) - \beta(t)B(t, T) - \frac{\delta(t)}{2}B^2(t, T) = 0, \quad B(T, T) = 0.$$

The equation for B is a Riccati equation, which can be solved analytically, see Ince (1944), §2.15, 12.51, A.21. Using the solution for B we get A by integrating.

Examples of short-rate models exhibiting an affine term structure include the following.

1. *Vasicek model:* $dr = (\alpha - \beta r)dt + \gamma dW$;

2. *Cox-Ingersoll-Ross (CIR) model:* $dr = (\alpha - \beta r)dt + \delta\sqrt{r}dW$;

3. *Ho-Lee model:* $dr = \alpha(t)dt + \gamma dW$;

4. *Hull-White (extended Vasicek) model:* $dr = (\alpha(t) - \beta(t)r)dt + \gamma(t)dW$;

5. *Hull-White (extended CIR) model:* $dr = (\alpha(t) - \beta(t)r)dt + \delta(t)\sqrt{r}dW$.

5.3 Heath-Jarrow-Morton Methodology

5.3.1 The Heath-Jarrow-Morton Model Class

Modelling the term structure with only one explanatory variable leads to various undesirable properties of the model (to say the least). Various authors have proposed models with more than one state variable, e.g. the short rate and a long rate and/or intermediate rates. The Heath-Jarrow-Morton method (compare Heath, Jarrow, and Morton (1992)) is at the far end of this spectrum – they propose using the entire forward rate curve as their (infinite-dimensional) state variable. More precisely, for any fixed $T \leq T^*$ the dynamics of instantaneous, continuously compounded forward rates $f(t,T)$ are *exogenously* given by

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW(t), \tag{5.4}$$

where W is a d-dimensional Brownian motion with respect to the underlying (objective) probability measure $I\!P$ and $\alpha(t,T)$ resp. $\sigma(t,T)$ are adapted $I\!R$ resp. $I\!R^d$-valued processes. For any fixed maturity T, the initial condition of the stochastic differential equation (5.4) is determined by the current value of the empirical (observed) forward rate for the future date T which prevails at time 0. Observe that we have defined an infinite-dimensional stochastic system, and that by construction we obtain a perfect fit to the observed term structure (thus avoiding the problem of inverting the yield curve).

The exogenous specification of the family of forward rates $\{f(t,T); T > 0\}$ is equivalent to a specification of the entire family of bond prices $\{p(t,T); T > 0\}$. Furthermore, by Proposition 5.1.1 we obtain the dynamics of the bond-price processes as

$$dp(t,T) = p(t,T)\{m(t,T)dt + S(t,T)dW(t)\}, \tag{5.13}$$

where

$$m(t,T) = r(t) + A(t,T) + \frac{1}{2}\|S(t,T)\|^2, \tag{5.14}$$

$A(t,T) = -\int_t^T \alpha(t,s)ds$ and $S(t,T) = -\int_t^T \sigma(t,s)ds$ (compare (5.6)). We now explore what conditions we must impose on the coefficients in order to ensure the existence of an equivalent martingale measure with respect to a suitable numéraire. By Theorem 4.2.1, we then could conclude that our bond market model is free of arbitrage.

As a first possible choice of numéraire, we use the money-market account B (assuming that there exists a measurable version of $f(t,t)$ in $[0,T^*]$), given by

$$B(t) = \exp\left\{\int_0^t f(u,u)du\right\} = \exp\left\{\int_0^t r(u)du\right\}.$$

So we allow investments in a savings account too. We must find an equivalent measure such that

$$Z(t,T) = \frac{p(t,T)}{B(t)}$$

is a martingale for every $0 \leq T \leq T^*$. We will call such a measure *risk-neutral martingale measure* to emphasize the dependence on the numéraire.

Theorem 5.3.1. *Assume that the family of forward rates is given by (5.4). Then there exists a risk-neutral martingale measure if and only if there exists an adapted process $\lambda(t)$, with the properties that*

(i) $\int_0^T \|\lambda(t)\|^2 \, dt < \infty,$ *a.s. and* $I\!E(L(T)) = 1,$ *with*

$$L(t) = \exp\left\{ -\int_0^t \lambda(u)' dW(u) - \frac{1}{2}\int_0^t \|\lambda(u)\|^2 \, du \right\}. \tag{5.15}$$

(ii) *For all* $0 \le T \le T^*$ *and for all* $t \le T$, *we have*

$$\alpha(t, T) = \sigma(t, T)\int_t^T \sigma(t, s) ds + \sigma(t, T)\lambda(t). \tag{5.16}$$

Proof. Since we are working in a Brownian framework we know that any equivalent measure $\boldsymbol{Q} \sim I\!P$ is given via a Girsanov density (5.15). Using Itô's formula and Girsanov's Theorem 4.1.4, we find the $\boldsymbol{Q}$-dynamics of $Z(t, T)$ (omitting the arguments) as:

$$dZ = Z\left(A + \frac{1}{2}\|S\|^2 - S\lambda\right)dt + ZSd\tilde{W}, \tag{5.17}$$

with $\tilde{W}$ a $\boldsymbol{Q}$-Brownian motion. In order for Z to be a $\boldsymbol{Q}$-martingale, the drift coefficient in (5.17) has to be zero, so we obtain

$$A(t, T) + \frac{1}{2}\|S(t, T)\|^2 = S(t, T)\lambda(t). \tag{5.18}$$

Taking the derivative with respect to T, we get

$$-\alpha(t, T) - \sigma(t, T)S(t, T) = -\sigma(t, T)\lambda(t),$$

which after rearranging is (5.16). $\blacksquare$

It is possible to interpret λ as a risk premium, which has to be exogenously specified to allow the choice of a particular risk-neutral martingale measure. In view of (5.16) this leads to a restriction on drift and volatility coefficients in the specification of the forward rate dynamics (5.4). The particular choice $\lambda \equiv 0$ means that we assume we model directly under a risk-neutral martingale measure $\boldsymbol{Q}$. In that case the relations between the various infinitesimal characteristics for the forward rate are known as the 'Heath-Jarrow-Morton drift condition'.

Theorem 5.3.2 (Heath-Jarrow-Morton). *Assume that $\boldsymbol{Q}$ is a risk-neutral martingale measure for the bond market and that the forward-rate dynamics under $\boldsymbol{Q}$ are given by*

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)d\tilde{W}(t), \tag{5.19}$$

with $\tilde{W}$ a $\boldsymbol{Q}$-Brownian motion. Then we have:

(i) *the Heath-Jarrow-Morton drift condition*

$$\alpha(t, T) = \sigma(t, T)\int_t^T \sigma(t, s) ds, \quad 0 \le t \le T \le T^*, \ \boldsymbol{Q} - a.s., \tag{5.20}$$

(ii) *and bond-price dynamics under $\boldsymbol{Q}$ are given by*

$$dp(t, T) = p(t, T)r(t)dt + p(t, T)S(t, T)d\tilde{W}(t),$$

with S as in (5.6).

5.3.2 Forward Risk-neutral Martingale Measures

For many valuation problems in the bond market it is more suitable to use the bond price process $p(t, T^*)$ as numéraire. We then have to find an equivalent probability measure $\boldsymbol{Q}^*$ such that the auxiliary process

$$Z^*(t, T) = \frac{p(t, T)}{p(t, T^*)}, \quad \forall t \in [0, T],$$

is a martingale under $\boldsymbol{Q}^*$ for all $T \leq T^*$. We will call such a measure *forward risk-neutral martingale measure*. In this setting, a savings account is not used and the existence of a martingale measure $\boldsymbol{Q}^*$ guarantees that there are no arbitrage opportunities between bonds of different maturities.

In order to find sufficient conditions for the existence of such a martingale measure, we follow the same programme as above. By (5.13) bond price dynamics under the original probability measure $\boldsymbol{P}$ are given as

$$dp(t, T) = p(t, T) \{m(t, T)dt + S(t, T)dW(t)\},$$

with $m(t, T)$ as in (5.14). Now applying Itô's formula to the quotient $p(t, T)/p(t, T^*)$ we find

$$dZ^*(t, T) = Z^*(t, T) \{\tilde{m}(t, T)dt + (S(t, T) - S(t, T^*))dW(t)\}, \tag{5.21}$$

with $\tilde{m}(t, T) = m(t, T) - m(t, T^*) - S(t, T^*)(S(t, T) - S(t, T^*))$. Again, any equivalent martingale measure $\boldsymbol{Q}^*$ is given via a Girsanov density $L(t)$ defined by a function $\gamma(t)$ as in Theorem 5.3.1. So the drift coefficient of $Z^*(t, T)$ under $\boldsymbol{Q}^*$ is given as

$$\tilde{m}(t, T) - (S(t, T) - S(t, T^*))\gamma(t).$$

Now for $Z^*(t, T)$ to be a $\boldsymbol{Q}^*$-martingale this coefficient has to be zero, and replacing $\tilde{m}$ with its definition we get

$$(A(t, T) - A(t, T^*)) + \frac{1}{2}\left(\|S(t, T)\|^2 - \|S(t, T^*)\|^2\right)$$
$$= (S(t, T^*) + \gamma(t))(S(t, T) - S(t, T^*)).$$

Written in terms of the coefficients of the forward-rate dynamics, this identity simplifies to

$$\int_T^{T^*} \alpha(t, s)ds + \frac{1}{2}\left\|\int_T^{T^*} \sigma(t, s)ds\right\|^2 = \gamma(t)\int_T^{T^*} \sigma(t, s)ds.$$

Taking the derivative with respect to T, we obtain

$$\alpha(t, T) + \sigma(t, T)\int_T^{T^*} \sigma(t, s)ds = \gamma(t)\sigma(t, T).$$

We have thus proved:

Theorem 5.3.3. *Assume that the family of forward rates is given by (5.4). Then there exists a forward risk-neutral martingale measure if and only if there exists an adapted process $\gamma(t)$, with the properties (i) of Theorem 5.3.1, such that*

$$\alpha(t, T) = \sigma(t, T)(S(T, T^*) + \gamma(t)), \quad 0 \leq t \leq T \leq T^*.$$

5.4 Pricing and Hedging Contingent Claims

5.4.1 Gaussian HJM Framework

Assume that the dynamics of the forward rate are given under a risk-neutral martingale measure $\boldsymbol{Q}$ by

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)d\tilde{W}(t), \quad f(0,T) = \hat{f}(0,T),$$

with all processes real-valued. We restrict the class of models by assuming that the forward rate's volatility is deterministic. The HJM-drift condition (5.20) and the integrated form of the forward rate (compare (5.7)) lead to

$$f(t,t) = r(t) = f(0,t) + \int_0^t (-\sigma(u,t)S(u,t))du + \int_0^t \sigma(u,t)d\tilde{W}(u),$$

which implies that the short-rate (as well as the forward rates $f(t,T)$) have Gaussian probability laws (hence the terminology). By Theorem 5.3.2, bond-price dynamics under $\boldsymbol{Q}$ are given by

$$dp(t,T) = p(t,T)\left\{r(t)dt + S(t,T)d\tilde{W}(t)\right\},$$

which we can solve in special cases explicitly for $p(t,T)$.

To price options on zero-coupon bonds, we use the change of numéraire technique. Consider a European call C on a T^*-bond with maturity $T \leq T^*$ and strike K. So we consider the T-contingent claim

$$X = (p(T,T^*) - K)^+. \tag{5.22}$$

The price of this call at time $t = 0$ is given as

$$C(0) = p(0,T^*)\boldsymbol{Q}^*(A) - Kp(0,T)\boldsymbol{Q}^T(A),$$

with $A = \{\omega : p(T,T^*) > K\}$ and $\boldsymbol{Q}^T$ resp. $\boldsymbol{Q}^*$ the T- resp. T^*-forward risk-neutral measure. Now

$$\tilde{Z}(t,T) = \frac{p(t,T^*)}{p(t,T)}$$

has $\boldsymbol{Q}$-dynamics (omitting the arguments and writing S^* for $S(t,T^*)$)

$$d\tilde{Z} = \tilde{Z}\left\{S(S-S^*)dt - (S-S^*)d\tilde{W}(t)\right\},$$

so a deterministic variance coefficient. Now

$$\boldsymbol{Q}^*(p(T,T^*) \geq K) = \boldsymbol{Q}^*\left(\frac{p(T,T^*)}{p(T,T)} \geq K\right) = \boldsymbol{Q}^*(\tilde{Z}(T,T) \geq K).$$

Since $\tilde{Z}(t,T)$ is a $\boldsymbol{Q}^T$-martingale with $\boldsymbol{Q}^T$-dynamics

$$d\tilde{Z}(t,T) = -\tilde{Z}(t,T)(S(t,T) - S(t,T^*))dW^T(t),$$

we find that under $\boldsymbol{Q}^T$ (again $S = S(t,T), S^* = S(t,T^*)$)

$$\tilde{Z}(T,T) = \frac{p(0,T^*)}{p(0,T)}\exp\left\{-\int_0^T (S-S^*)dW^T(t) - \frac{1}{2}\int_0^T (S-S^*)^2 dt\right\}$$

(with W^T a $\boldsymbol{Q}^T$-Brownian motion). The stochastic integral in the exponential is Gaussian with zero mean and variance

$$\Sigma^2(T) = \int_0^T (S(t,T) - S(t,T^*))^2 dt.$$

So

$$\boldsymbol{Q}^T(p(T,T^*) \geq K) = \boldsymbol{Q}^T(\tilde{Z}(T,T) \geq K) = N(d_2)$$

with

$$d_2 = \frac{\log\left(\frac{p(0,T)}{Kp(0,T^*)}\right) - \frac{1}{2}\Sigma^2(T)}{\sqrt{\Sigma^2(T)}}.$$

Similarly, for the first term

$$Z^*(t,T) = \frac{p(t,T)}{p(t,T^*)}$$

has $\boldsymbol{Q}$-dynamics (compare (5.21))

$$dZ^* = Z^*\left\{S^*(S^* - S)dt + (S - S^*)d\tilde{W}(t)\right\},$$

and also a deterministic variance coefficient. Now

$$\boldsymbol{Q}^*(p(T,T^*) \geq K) = \boldsymbol{Q}^*\left(\frac{1}{p(T,T^*)} \leq \frac{1}{K}\right) = \boldsymbol{Q}^*(Z^*(T,T) \leq \frac{1}{K}).$$

Under $\boldsymbol{Q}^*$ $Z^*(t,T)$ is a martingale with

$$dZ^*(t,T) = Z^*(t,T)(S(t,T) - S(t,T^*))dW^*(t),$$

so

$$dZ^*(T,T) = \frac{p(0,T)}{p(0,T^*)} \exp\left\{\int_0^T (S - S^*)dW^*(t) - \frac{1}{2}\int_0^T (S - S^*)^2 dt\right\}.$$

Again we have a Gaussian variable with the (same) variance $\Sigma^2(T)$ in the exponential. Using this fact it follows (after some computations) that:

$$\boldsymbol{Q}^*(p(T,T^*) \geq K) = N(d_1),$$

with

$$d_1 = d_2 + \sqrt{\Sigma^2(T)}.$$

So we obtain:

Proposition 5.4.1. *The price of the call option defined in (5.22) is given by*

$$C(0) = p(0,T^*)N(d_2) - Kp(0,T)N(d_1), \tag{5.23}$$

with parameters given as above.

5.4.2 Swaps

This section is devoted to the pricing of swaps. We consider the case of a *forward swap settled in arrears*. Such a contingent claim is characterized by:

- a fixed time t, the contract time,
- dates $T_0 < T_1, \ldots < T_n$, equally distanced $T_{i+1} - T_i = \delta$,
- R, a prespecified fixed rate of interest,
- K, a nominal amount.

A swap contract S with K and R fixed for the period $T_0, \ldots T_n$ is a sequence of payments, where the amount of money paid out at T_{i+1}, $i = 0, \ldots, n-1$ is defined by

$$X_{i+1} = K\delta(L(T_i, T_i) - R).$$

The floating rate over $[T_i, T_{i+1}]$ observed at T_i is a simple rate defined as

$$p(T_i, T_{i+1}) = \frac{1}{1 + \delta L(T_i, T_i)}.$$

We do not need to specify a particular interest-rate model here, all we need is the existence of a risk-neutral martingale measure. Using the risk-neutral pricing formula we obtain (we may use $K = 1$),

$$
\begin{aligned}
\Pi(t, S) &= \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{Q}} \left[e^{-\int_t^T r(s)ds} \delta(L(T_i, T_i) - R) \Big| \mathcal{F}_t \right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{Q}} \left[\mathbb{E}_{\boldsymbol{Q}} \left[e^{-\int_{T_{i-1}}^{T_i} r(s)ds} \Big| \mathcal{F}_{T_{i-1}} \right] \right. \\
&\qquad \left. \times\, e^{-\int_t^T r(s)ds} \left(\frac{1}{p(T_{i-1}, T_i)} - (1 + \delta R) \right) \Big| \mathcal{F}_t \right] \\
&= \sum_{i=1}^{n} \left(p(t, T_{i-1}) - (1 + \delta R) p(t, T_i) \right) = p(t, T_0) - \sum_{i=1}^{n} c_i p(t, T_i),
\end{aligned}
$$

with $c_i = \delta R, i = 1, \ldots, n-1$ and $c_n = 1 + \delta R$. So a swap is a linear combination of zero-coupon bonds, and we obtain its price accordingly. This again shows the power of risk-neutral pricing. Using the linearity of the expectation operator we can reduce complicated claims to sums of simpler ones.

5.4.3 Caps

An interest cap is a contract where the seller of the contract promises to pay a certain amount of cash to the holder of the contract if the interest rate exceeds a certain predetermined level (the cap rate) at some future date. A cap can be broken down in a series of caplets. A caplet is a contract written at t, in force between $[T_0, T_1]$, $\delta = T_1 - T_0$, the nominal amount is K, the cap rate is denoted by R. The relevant interest rate (LIBOR, for instance) is observed in T_0 and defined by

$$p(T_0, T_1) = \frac{1}{1 + \delta L(T_0, T_0)}.$$

A caplet C is a T_1-contingent claim with payoff $X = K\delta(L(T_0, T_0) - R)^+$. Writing $L = L(T_0, T_0), p = p(T_0, T_1), R^* = 1 + \delta R$, we have $L = (1-p)/(\delta p)$, (assuming $K = 1$) and

$$
\begin{aligned}
X &= \delta(L - R)^+ = \delta \left(\frac{1-p}{\delta p} - R \right)^+ \\
&= \left(\frac{1}{p} - (1 + \delta R) \right)^+ = \left(\frac{1}{p} - R^* \right)^+.
\end{aligned}
$$

The risk-neutral pricing formula leads to

$$
\begin{aligned}
\Pi_C(t) &= \mathbb{E}_{\boldsymbol{Q}}\left[e^{-\int_t^{T_1} r(s)ds}\left(\frac{1}{p}-R^*\right)^+ \Big| \mathcal{F}_t\right] \\[2mm]
&= \mathbb{E}_{\boldsymbol{Q}}\left[\mathbb{E}_{\boldsymbol{Q}}\left[e^{-\int_{T_0}^{T_1} r(s)ds}\Big|\mathcal{F}_{T_0}\right]e^{-\int_t^{T_0} r(s)ds}\left(\frac{1}{p}-R^*\right)^+\Big|\mathcal{F}_t\right] \\[2mm]
&= \mathbb{E}_{\boldsymbol{Q}}\left[p(T_0,T_1)\,e^{-\int_t^{T_0} r(s)ds}\left(\frac{1}{p}-R^*\right)^+\Big|\mathcal{F}_t\right] \\[2mm]
&= \mathbb{E}_{\boldsymbol{Q}}\left[e^{-\int_t^{T_0} r(s)ds}\left(1-pR^*\right)^+\Big|\mathcal{F}_t\right] \\[2mm]
&= R^*\,\mathbb{E}_{\boldsymbol{Q}}\left[e^{-\int_t^{T_0} r(s)ds}\left(\frac{1}{R^*}-p\right)^+\Big|\mathcal{F}_t\right].
\end{aligned}
$$

So a caplet is equivalent to R^* put options on a T_1-bond with maturity T_0 and strike $1/R^*$.

Appendix A

Basic Probability Background

A.1 Fundamentals

To describe a random experiment we use a *sample space* Ω, the set of all possible outcomes. Each point ω of Ω, or *sample point*, represents a possible random outcome of performing the random experiment.

Examples. Write down Ω for experiments such as flip a coin three times, roll two dice.

For a set $A \subseteq \Omega$ we want to know the probability $I\!P(A)$. The class $\mathcal{F}$ of subsets of Ω whose probabilities $I\!P(A)$ are defined (call such A *events*) should be a σ-algebra , i.e.

(i) $\emptyset, \Omega \in \mathcal{F}$.

(ii) $F \in \mathcal{F}$ implies $F^c \in \mathcal{F}$.

(iii) $F_1, F_2, \ldots \in \mathcal{F}$ then $\bigcup_n F_n \in \mathcal{F}$.

We want a probability measure defined on $\mathcal{F}$

(i) $I\!P(\emptyset) = 0, I\!P(\Omega) = 1$,

(ii) $I\!P(A) \geq 0$ for all A,

(iii) If $A_1, A_2, \ldots,$ are disjoint, $I\!P(\bigcup_i A_i) = \sum_i I\!P(A_i)$ countable additivity.

Definition A.1.1. *A probability space, or Kolmogorov triple, is a triple $(\Omega, \mathcal{F}, I\!P)$ satisfying Kolmogorov axioms (i),(ii) and (iii) above.*

A probability space is a mathematical model of a random experiment.

Examples. Assign probabilities for the above experiments.

Definition A.1.2. *Let $(\Omega, \mathcal{F}, I\!P)$ be a probability space. A random variable (vector) X is a function $X : \Omega \to I\!R(I\!R^k)$ such that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}(\mathcal{B}(I\!R^k))$.*

For a random variable X
$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$
for all $x \in I\!R$. So define the *distribution function F_X* of X by

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

Recall: $\sigma(X)$, the σ-algebra *generated* by X.

Some important probability distributions

- Binomial distribution: Number of successes

$$IP(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- Geometric distribution: Waiting time

$$IP(N = n) = p(1-p)^{n-1}.$$

- Poisson distribution:

$$IP(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

- Density of Uniform distribution:

$$f(x) = \frac{1}{b-a} \mathbf{1}_{\{(a,b)\}}.$$

- Density of Exponential distribution:

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\{[0,\infty)\}}.$$

- Density of standard Normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Definition A.1.3. *The expectation $I\!E$ of a random variable X on $(\Omega, \mathcal{F}, I\!P)$ is defined by*

$$I\!E X := \int_{\Omega} X dI\!P, \quad or \quad \int_{\Omega} X(\omega) dI\!P(\omega).$$

The variance of a random variable is defined as

$$\mathbb{V}ar(X) := I\!E \left[(X - I\!E(X))^2 \right] = I\!E \left(X^2 \right) - (I\!E X)^2.$$

If X is real-valued with density f (i.e $f(x) \geq 0 : \int f(x)dx = 1$),

$$I\!E X := \int x f(x) dx$$

or if X is discrete, taking values $x_n (n = 1, 2, \ldots)$ with probability function $f(x_n)(\geq 0)$,

$$I\!E X := \sum x_n f(x_n).$$

Examples. Calculate moments for some of the above distributions.

Definition A.1.4. *Random variables $X_1, \ldots, X_n$ are independent if whenever $A_i \in \mathcal{B}$ (the Borel σ-algebra) for $i = 1, \ldots n$ we have*

$$IP\left(\bigcap_{i=1}^{n}\{X_i \in A_i\}\right) = \prod_{i=1}^{n} IP(\{X_i \in A_i\}).$$

Lemma A.1.1. *In order for $X_1, \ldots, X_n$ to be independent it is necessary and sufficient that for all $x_1, \ldots x_n \in (-\infty, \infty]$,*

$$IP\left(\bigcap_{i=1}^{n}\{X_i \leq x_i\}\right) = \prod_{i=1}^{n} IP(\{X_i \leq x_i\}).$$

Theorem A.1.1 (Multiplication Theorem). *If $X_1, \ldots, X_n$ are independent and $IE|X_i| < \infty$, $i = 1, \ldots, n$, then*

$$IE\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} IE(X_i).$$

If X, Y are independent, with distribution functions F, G, define $Z := X + Y$ with distribution function H. We call H the *convolution* of F and G, written $H = F * G$.

Suppose X, Y have densities f, g, then Z has a density h with

$$h(z) = \int_{-\infty}^{\infty} f(z-y)g(y)dy = \int_{-\infty}^{\infty} f(x)g(z-x)dx.$$

Example. Assume $t_1, \ldots, t_n$ are independent random variables that have an exponential distribution with parameter λ. Then $T = t_1 + \ldots + t_n$ has the Gamma(n, λ) density function

$$f(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}.$$

Definition A.1.5. *If X is a random variable with distribution function F, its moment generating function ϕ_X is*

$$\phi(t) := IE(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} dF(x).$$

The mgf *takes convolution into multiplication*: if X, Y are independent,

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Observe $\phi^{(k)}(t) = IE(X^k e^{tX})$ and $\phi(0) = IE(X^k)$.

For X on nonnegative integers use the *generating function*

$$\gamma_X(z) = IE(z^X) = \sum_{k=0}^{\infty} z^k IP(Z = k).$$

A.2 Convolution and Characteristic Functions

The most basic operation on numbers is addition; the most basic operation on random variables is addition of independent random variables. If X, Y are independent, with distribution functions F, G, and

$$Z := X + Y,$$

let Z have distribution function H. Then since $X + Y = Y + X$ (addition is commutative), H depends on F and G symmetrically. We call H the *convolution* (German: *Faltung*) of F and G, written

$$H = F * G.$$

Suppose first that X, Y have densities f, g. Then

$$H(z) = \mathbb{P}(Z \le z) = \mathbb{P}(X + Y \le z) = \int\limits_{\{(x,y):x+y \le z\}} f(x)g(y)dxdy,$$

since by independence of X and Y the joint density of X and Y is the product $f(x)g(y)$ of their separate (marginal) densities, and to find probabilities in the density case we integrate the joint density over the relevant region. Thus

$$H(z) = \int_{-\infty}^{\infty} f(x) \left\{ \int_{-\infty}^{z-x} g(y)dy \right\} dx = \int_{-\infty}^{\infty} f(x)G(z-x)dx.$$

If

$$h(z) := \int_{-\infty}^{\infty} f(x)g(z-x)dx,$$

(and of course symmetrically with f and g interchanged), then integrating we recover the equation above (after interchanging the order of integration. This is legitimate, as the integrals are non-negative, by Fubini's theorem, which we quote from measure theory, see e.g. (Williams 1991), §8.2). This shows that if X, Y are independent with densities f, g, and $Z = X + Y$, then Z has density h, where

$$h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy.$$

We write

$$h = f * g,$$

and call the density h the *convolution* of the densities f and g.

If X, Y do not have densities, the argument above may still be taken as far as

$$H(z) = \mathbb{P}(Z \le z) = \mathbb{P}(X + Y \le z) = \int_{-\infty}^{\infty} F(x-y)dG(y)$$

(and, again, symmetrically with F and G interchanged), where the integral on the right is the *Lebesgue-Stieltjes integral* of §2.2. We again write

$$H = F * G,$$

and call the distribution function H the *convolution* of the distribution functions F and G.

In sum: *addition* of *independent* random variables corresponds to *convolution* of distribution functions or densities.

Now we frequently need to add (or average) lots of independent random variables: for example, when forming sample means in statistics – when the bigger the sample size is, the better. But convolution involves integration, so adding n independent random variables involves $n-1$ integrations, and this is awkward to do for large n. One thus seeks a way to transform distributions so as to make the awkward operation of convolution as easy to handle as the operation of addition of independent random variables that gives rise to it.

Definition A.2.1. *If X is a random variable with distribution function F, its characteristic function ϕ (or ϕ_X if we need to emphasise X) is*

$$\phi(t) := I\!\!E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} dF(x), \quad (t \in I\!\!R).$$

Note.

Here $i := \sqrt{-1}$. All other numbers – t, x etc. – are real; all expressions involving i such as e^{itx}, $\phi(t) = I\!\!E(e^{itx})$ are complex numbers.

The characteristic function *takes convolution into multiplication*: if X, Y are independent,

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

For, as X, Y are independent, so are e^{itX} and e^{itY} for any t, so by the multiplication theorem (Theorem B.3.1),

$$I\!\!E(e^{it(X+Y)}) = I\!\!E(e^{itX} \cdot e^{itY}) = I\!\!E(e^{itX}) \cdot I\!\!E(e^{itY}),$$

as required.

We list some properties of characteristic functions that we shall need.

1. $\phi(0) = 1$. For, $\phi(0) = I\!\!E(e^{i \cdot 0 \cdot X}) = I\!\!E(e^0) = I\!\!E(1) = 1$.
2. $|\phi(t)| \leq 1$ for all $t \in I\!\!R$.
Proof. $|\phi(t)| = \left| \int_{-\infty}^{\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{\infty} |e^{itx}| \, dF(x) = \int_{-\infty}^{\infty} 1 dF(x) = 1.$
Thus in particular the characteristic function *always exists* (the integral defining it is always absolutely convergent). This is a crucial advantage, far outweighing the disadvantage of having to work with complex rather than real numbers (the nuisance value of which is in fact slight).
3. ϕ is continuous (indeed, ϕ is uniformly continuous).
Proof.

$$
\begin{aligned}
|\phi(t+u) - \phi(t)| &= \left| \int_{-\infty}^{\infty} \{e^{i(t+u)x} - e^{itx}\} dF(x) \right| \\
&= \left| \int_{-\infty}^{\infty} e^{itx}(e^{iux} - 1) dF(x) \right| \leq \int_{-\infty}^{\infty} |e^{iux} - 1| \, dF(x),
\end{aligned}
$$

for all t. Now as $u \to 0$, $|e^{iux} - 1| \to 0$, and $|e^{iux} - 1| \leq 2$. The bound on the right tends to zero as $u \to 0$ by Lebesgue's dominated convergence theorem (which we quote from measure theory: see e.g. (Williams 1991), §5.9), giving continuity; the uniformity follows as the bound holds uniformly in t.
4. (*Uniqueness theorem*): ϕ determines the distribution function F uniquely.
Technically, ϕ is the *Fourier-Stieltjes transform* of F, and here we are quoting the uniqueness property of this transform. Were uniqueness not to hold, we would lose information on taking characteristic functions, and so ϕ would not be useful.
5. (*Continuity theorem*): If X_n, X are random variables with distribution functions F_n, F and characteristic functions ϕ_n, ϕ, then convergence of ϕ_n to ϕ,

$$\phi_n(t) \to \phi(t) \quad (n \to \infty) \quad \text{for all} \quad t \in I\!\!R$$

is equivalent to convergence in distribution of X_n to X. This result is due to Lévy; see e.g. (Williams 1991), §18.1.
6. *Moments.* Suppose X has kth moment: $I\!\!E|X|^k < \infty$. Take the Taylor (power-series) expansion of e^{itx} as far as the kth power term:

$$e^{itx} = 1 + itx + \cdots + (itx)^k/k! + \boldsymbol{o}\left(t^k\right),$$

where '$\boldsymbol{o}\left(t^{k}\right)$' denotes an error term of smaller order than t^{k} for small k. Now replace x by X, and take expectations. By linearity, we obtain

$$\phi(t) = I\!\!E(e^{itX}) = 1 + it I\!\!E X + \cdots + \frac{(it)^{k}}{k!} I\!\!E(X^{k}) + e(t),$$

where the error term $e(t)$ is the expectation of the error terms (now random, as X is random) obtained above (one for each value $X(\omega)$ of X). It is not obvious, but it is true, that $e(t)$ is still of smaller order than t^{k} for $t \to 0$:

$$\text{if} \quad I\!\!E\left(|X|^{k}\right) < \infty, \quad \phi(t) = 1 + it I\!\!E(X) + \ldots + \frac{(it)^{k}}{k!} I\!\!E\left(X^{k}\right) + \boldsymbol{o}\left(t^{k}\right) \quad (t \to 0).$$

We shall need the case $k = 2$ in dealing with the central limit theorem below.

Examples

1. Standard Normal Distribution,

$N(0,1)$. For the standard normal density $f(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^{2}\}$, one has, by the process of 'completing the square' (familiar from when one first learns to solve quadratic equations!),

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{tx} f(x) dx &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{tx - \frac{1}{2}x^{2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x-t)^{2} + \frac{1}{2}t^{2}\right\} dx \\
&= \exp\left\{\frac{1}{2}t^{2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x-t)^{2}\right\} dx.
\end{aligned}
$$

The second factor on the right is 1 (it has the form of a normal integral). This gives the integral on the left as $\exp\{\frac{1}{2}t^{2}\}$.

Now replace t by it (legitimate by analytic continuation, which we quote from complex analysis, see e.g. (Burkill and Burkill 1970)). The right becomes $\exp\{-\frac{1}{2}t^{2}\}$. The integral on the left becomes the characteristic function of the standard normal density – which we have thus now identified (and will need below in §2.8).

2. General Normal Distribution,

$N(\mu, \sigma)$. Consider the transformation $x \mapsto \mu + \sigma x$. Applied to a random variable X, this adds μ to the mean (a change of *location*), and multiplies the variance by σ^{2} (a change of *scale*). One can check that if X has the standard normal density above, then $\mu + \sigma X$ has density

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

and characteristic function

$$
\begin{aligned}
I\!\!E e^{it(\mu + \sigma X)} &= \exp\{i\mu t\} I\!\!E\left(e^{(i\sigma t)X}\right) = \exp\{i\mu t\} \exp\left\{-\frac{1}{2}(\sigma t)^{2}\right\} \\
&= \exp\left\{i\mu t - \frac{1}{2}\sigma^{2}t^{2}\right\}.
\end{aligned}
$$

Thus the general normal density and its characteristic function are

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu)^{2}/\sigma^{2}\right\}, \quad \phi(t) = \exp\left\{i\mu t - \frac{1}{2}\sigma^{2}t^{2}\right\}.$$

3. Poisson Distribution,

$P(\lambda)$. Here, the probability mass function is

$$f(k) := I\!\!P(X = k) = e^{-\lambda}\lambda^k/k!, \quad (k = 0, 1, 2, \ldots).$$

The characteristic function is thus

$$
\begin{aligned}
\phi(t) &= I\!\!E\left(e^{itX}\right) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \cdot e^{itk} \\
&= e^{-\lambda}\sum_{k=0}^{\infty}(\lambda e^{it})^k/k! = e^{-\lambda}\exp\{\lambda e^{it}\} = \exp\{-\lambda(1 - e^{it})\}.
\end{aligned}
$$

A.3 The Central Limit Theorem

Readers of this book will be well aware that

$$\left(1 + \frac{x}{n}\right)^n \to e^x \quad (n \to \infty) \quad \forall x \in I\!\!R.$$

This is the formula governing the passage from discrete to continuous compound interest. Invest one pound (or dollar) for one year at $100x\%$ p.a.; with interest compounded n times p.a., our capital after one year is $(1 + \frac{x}{n})^n$. With continuous compounding, our capital after one year is the exponential e^x: *exponential growth corresponds to continuously compounded interest.*

We need two extensions: the formula still holds with $x \in I\!\!R$ replaced by a complex number $z \in \mathbb{C}$:

$$\left(1 + \frac{z}{n}\right)^n \to e^z \quad (n \to \infty) \quad \forall z \in \mathbb{C},$$

and if $z_n \in \mathbb{C}$, $z_n \to z$,

$$\left(1 + \frac{z_n}{n}\right)^n \to e^z \quad (n \to \infty) \quad (z_n \to z \in \mathbb{C}).$$

As a first illustration of the power of transform methods, we prove the weak law of large numbers:

Theorem A.3.1 (Weak Law of Large Numbers). *If $X_1, X_2, \ldots$ are independent and identically distributed with mean μ, then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad (n \to \infty) \quad \text{in probability.}$$

Proof. If the X_i have characteristic function ϕ, then by the moment property of §2.8 with $k = 1$,

$$\phi(t) = 1 + i\mu t + \boldsymbol{o}(t) \quad (t \to 0).$$

Now using the i.i.d. assumption, $\frac{1}{n}\sum_{i=1}^{n} X_i$ has characteristic function

$$
\begin{aligned}
I\!\!E\left(\exp\left\{it \cdot \frac{1}{n}\sum_{1}^{n} X_j\right\}\right) &= I\!\!E\left(\prod_{i=1}^{n}\exp\left\{it \cdot \frac{1}{n}X_j\right\}\right) \\
&= \prod_{i=1}^{n} I\!\!E\left(\exp\left\{\frac{it}{n}X_j\right\}\right) = (\phi(t/n))^n \\
&= \left(1 + \frac{i\mu t}{n} + \boldsymbol{o}(1/n)\right)^n \to e^{i\mu t} \quad (n \to \infty),
\end{aligned}
$$

and $e^{i\mu t}$ is the characteristic function of the constant μ (for fixed t, $\boldsymbol{o}\left(1/n\right)$ is an error term of smaller order than $1/n$ as $n \to \infty$). By the continuity theorem,

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to \mu \quad \text{in distribution,}$$

and as μ is constant, this says (see §2.6) that

$$\frac{1}{n} \sum_{1}^{n} X_i \to \mu \quad \text{in probability.}$$

$\blacksquare$

The main result of this section is the same argument carried one stage further.

Theorem A.3.2 (Central Limit Theorem). *If $X_1, X_2, \ldots$ are independent and identically distributed with mean μ and variance σ^2, then with $N(0,1)$ the standard normal distribution,*

$$\frac{\sqrt{n}}{\sigma} \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)/\sigma \to N(0,1) \quad (n \to \infty) \quad \text{in distribution.}$$

That is, for all $x \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)/\sigma \leq x \right) \to \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2} y^2} dy \quad (n \to \infty).$$

Proof. We first centre at the mean. If X_i has characteristic function ϕ, let $X_i - \mu$ have characteristic function ϕ_0. Since $X_i - \mu$ has mean 0 and second moment $\sigma^2 = \mathbb{V}ar(X_i) = \mathbb{E}[(X_i - \mathbb{E}X_i)^2] = \mathbb{E}[(X_i - \mu)^2]$, the case $k = 2$ of the moment property of §2.7 gives

$$\phi_0(t) = 1 - \frac{1}{2} \sigma^2 t^2 + \boldsymbol{o}\left(t^2\right) \quad (t \to 0).$$

Now $\sqrt{n}(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu)/\sigma$ has characteristic function

$$\mathbb{E}\left(\exp\left\{ it \cdot \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{j=1}^{n} X_j - \mu \right) \right\} \right)$$

$$= \mathbb{E}\left(\prod_{j=1}^{n} \exp\left\{ \frac{it(X_j - \mu)}{\sigma\sqrt{n}} \right\} \right) = \prod_{j=1}^{n} \mathbb{E}\left(\exp\left\{ \frac{it}{\sigma\sqrt{n}}(X_j - \mu) \right\} \right)$$

$$= \left(\phi_0 \left(\frac{t}{\sigma\sqrt{n}} \right) \right)^n = \left(1 - \frac{\frac{1}{2}\sigma^2 t^2}{\sigma^2 n} + \boldsymbol{o}\left(\frac{1}{n} \right) \right)^n \to e^{-\frac{1}{2} t^2} \quad (n \to \infty),$$

and $e^{-\frac{1}{2} t^2}$ is the characteristic function of the standard normal distribution $N(0,1)$. The result follows by the continuity theorem. $\blacksquare$

Note.

In Theorem A.3.2, we:

(i) *centre* the X_i by subtracting the mean (to get mean 0);
(ii) *scale* the resulting $X_i - \mu$ by dividing by the standard deviation σ (to get variance 1). Then if $Y_i := (X_i - \mu)/\sigma$ are the resulting *standardised* variables, $\frac{1}{\sqrt{n}} \sum_{1}^{n} Y_i$ converges in distribution to standard normal.

Example: the Binomial Case.

If each X_i is Bernoulli distributed with parameter $p \in (0, 1)$,

$$I\!P(X_i = 1) = p, \quad I\!P(X_i = 0) = q := 1 - p$$

– so X_i has mean p and variance pq - $S_n := \sum_{i=1}^n X_i$ is binomially distributed with parameters n and p:

$$I\!P\left(\sum_{i=1}^n X_i = k\right) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!k!} p^k q^{n-k}.$$

A direct attack on the distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - p)/\sqrt{pq}$ can be made via

$$I\!P\left(a \le \sum_{i=1}^n X_i \le b\right) = \sum_{k: np + a\sqrt{npq} \le k \le np + b\sqrt{npq}} \frac{n!}{(n-k)!k!} p^k q^{n-k}.$$

Since n, k and $n - k$ will all be large here, one needs an approximation to the factorials. The required result is *Stirling's formula* of 1730:

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}} \quad (n \to \infty)$$

(the symbol $\sim$ indicates that the ratio of the two sides tends to 1). The argument can be carried through to obtain the sum on the right as a Riemann sum (in the sense of the Riemann integral: §2.2) for $\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$, whence the result. This, the earliest form of the central limit theorem, is the de Moivre-Laplace limit theorem (Abraham de Moivre, 1667–1754; P.S. de Laplace, 1749–1827). The proof of the de-Moivre-Laplace limit theorem sketched above is closely analogous to the passage from the discrete to the continuous Black-Scholes formula: see §4.6 and §6.4.

Local Limit Theorems.

The central limit theorem as proved above is a *global limit theorem*: it relates to *distributions* and convergence thereof. The de Moivre-Laplace limit theorem above, however, deals directly with individual probabilities in the discrete case (the sum of a large number of which is shown to approximate an integral). A limit theorem dealing with *densities* and convergence thereof in the density case, or with the discrete analogues of densities – such as the individual probabilities $I\!P(S_n = k)$ in the binomial case above – is called a *local limit theorem*.

Poisson Limit Theorem.

The de Moivre-Laplace limit theorem – convergence of binomial to normal – is only one possible limiting regime for binomial models. The next most important one has a *Poisson* limit in place of a normal one.

Suppose we have a sequence of binomial models $B(n, p)$, where the success probability $p = p_n$ varies with n, in such a way that

$$np_n \to \lambda \quad > 0, \quad (n \to \infty). \tag{A.1}$$

Thus $p_n \to 0$ – indeed, $p_n \sim \lambda/n$. This models a situation where we have a large number n of Bernoulli trials, each with small probability p_n of success, but such that np_n, the expected total number of successes, is 'neither large nor small, but intermediate'. Binomial models satisfying condition (A.1) converge to the Poisson model $P(\lambda)$ with parameter $\lambda > 0$.

This result is sometimes called the law of small numbers. The Poisson distribution is widely used to model statistics of accidents, insurance claims and the like, where one has a large number n of individuals at risk, each with a small probability p_n of generating an accident, insurance claim etc. ('success probability' seems a strange usage here!).

Appendix B

Facts form Probability and Measure Theory

We will assume that most readers will be familiar with such things from an elementary course in probability and statistics; for a clear introduction see, e.g. Grimmett and Welsh (1986), or the first few chapters of **?**; Ross (1997), Resnick (2001), Durrett (1999), Ross (1997), Rosenthal (2000) are also useful.

B.1 Measure

The language of modelling financial markets involves that of probability, which in turn involves that of *measure theory*. This originated with Henri Lebesgue (1875-1941), in his thesis, 'Intégrale, longueur, aire' Lebesgue (1902). We begin with defining a measure on $I\!R$ generalising the intuitive notion of length.

The length $\mu(I)$ of an interval $I = (a, b), [a, b], [a, b)$ or $(a, b]$ should be $b - a$: $\mu(I) = b - a$. The length of the disjoint union $I = \bigcup_{r=1}^{n} I_r$ of intervals I_r should be the sum of their lengths:

$$\mu\left(\bigcup_{r=1}^{n} I_r\right) = \sum_{r=1}^{n} \mu(I_r) \qquad \text{(finite additivity)}.$$

Consider now an infinite sequence $I_1, I_2, \ldots(ad\ infinitum)$ of disjoint intervals. Letting n tend to ∞ suggests that length should again be additive over disjoint intervals:

$$\mu\left(\bigcup_{r=1}^{\infty} I_r\right) = \sum_{r=1}^{\infty} \mu(I_r) \qquad \text{(countable additivity)}.$$

For I an interval, A a subset of length $\mu(A)$, the length of the complement $I \setminus A := I \cap A^c$ of A in I should be

$$\mu(I \setminus A) = \mu(I) - \mu(A) \qquad \text{(complementation)}.$$

If $A \subseteq B$ and B has length $\mu(B) = 0$, then A should have length 0 also:

$$A \subseteq B \quad \text{and} \quad \mu(B) = 0 \ \Rightarrow \ \mu(A) = 0 \qquad \text{(completeness)}.$$

The term 'countable' here requires comment. We must distinguish first between finite and infinite sets; then countable sets (like $I\!N = \{1, 2, 3, \ldots\}$) are the 'smallest', or 'simplest', infinite sets, as distinct from uncountable sets such as $I\!R = (-\infty, \infty)$.

Let $\mathcal{F}$ be the smallest class of sets $A \subset I\!R$ containing the intervals, closed under countable disjoint unions and complements, and complete (containing all subsets of sets of length 0 as sets of length 0). The above suggests – what Lebesgue showed – that length can be sensibly defined on the

sets $\mathcal{F}$ on the line, but on no others. There are others – but they are hard to construct (in technical language: the axiom of choice, or some variant of it such as Zorn's lemma, is needed to demonstrate the existence of non-measurable sets – but all such proofs are highly non-constructive). So: some but not all subsets of the line have a length. These are called the *Lebesgue-measurable sets*, and form the class $\mathcal{F}$ described above; length, defined on $\mathcal{F}$, is called *Lebesgue measure* μ (on the real line, $I\!\!R$). Turning now to the general case, we make the above rigorous. Let Ω be a set.

Definition B.1.1. *A collection $\mathcal{A}_0$ of subsets of Ω is called an algebra on Ω if:*

(i) $\Omega \in \mathcal{A}_0$,
(ii) $A \in \mathcal{A}_0 \Rightarrow A^c = \Omega \setminus A \in \mathcal{A}_0$,
(iii) $A, B \in \mathcal{A}_0 \Rightarrow A \cup B \in \mathcal{A}_0$.

Using this definition and induction, we can show that an algebra on Ω is a family of subsets of Ω closed under finitely many set operations.

Definition B.1.2. *An algebra $\mathcal{A}$ of subsets of Ω is called a σ-algebra on Ω if for any sequence $A_n \in \mathcal{A}$, $(n \in I\!\!N)$, we have*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

Such a pair $(\Omega, \mathcal{A})$ is called a measurable space.

Thus a σ-algebra on Ω is a family of subsets of Ω closed under any countable collection of set operations.

The main examples of σ-algebras are σ-algebras generated by a class $\mathcal{C}$ of subsets of Ω, i.e. $\sigma(\mathcal{C})$ is the smallest σ-algebra on Ω containing $\mathcal{C}$.

The Borel σ-algebra $\mathcal{B} = \mathcal{B}(I\!\!R)$ is the σ-algebra of subsets of $I\!\!R$ generated by the open intervals (equivalently, by half-lines such as $(-\infty, x]$ as x varies in $I\!\!R$.

As our aim is to define measures on collection of sets we now turn to set functions.

Definition B.1.3. *Let Ω be a set, $\mathcal{A}_0$ an algebra on Ω and μ_0 a non-negative set function $\mu_0 : \mathcal{A}_0 \to [0, \infty]$ such that $\mu_0(\emptyset) = 0$. μ_0 is called:*

(i) additive, if $A, B \in \mathcal{A}_0, A \cap B = \emptyset \Rightarrow \mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$,
(ii) countably additive, if whenever $(A_n)_{n \in I\!\!N}$ is a sequence of disjoint sets in $\mathcal{A}_0$ with $\bigcup A_n \in \mathcal{A}_0$ then

$$\mu_0 \left(\bigcup_{n=0}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

Definition B.1.4. *Let $(\Omega, \mathcal{A})$ be a measurable space. A countably additive map*

$$\mu : \mathcal{A} \to [0, \infty]$$

is called a measure on $(\Omega, \mathcal{A})$. The triple $(\Omega, \mathcal{A}, \mu)$ is called a measure space.

Recall that our motivating example was to define a measure on $I\!\!R$ consistent with our geometrical knowledge of length of an interval. That means we have a suitable definition of measure on a family of subsets of $I\!\!R$ and want to extend it to the generated σ-algebra. The measure-theoretic tool to do so is the Carathéodory extension theorem, for which the following lemma is an inevitable prerequisite.

Lemma B.1.1. *Let Ω be a set. Let $\mathcal{I}$ be a π-system on Ω, that is, a family of subsets of Ω closed under finite intersections: $I_1, I_2 \in \mathcal{I} \Rightarrow I_1 \cap I_2 \in \mathcal{I}$. Let $\mathcal{A} = \sigma(\mathcal{I})$ and suppose that μ_1 and μ_2 are finite measures on $(\Omega, \mathcal{A})$ (i.e. $\mu_1(\Omega) = \mu_2(\Omega) < \infty$) and $\mu_1 = \mu_2$ on $\mathcal{I}$. Then*

$$\mu_1 = \mu_2 \quad on \ \ \mathcal{A}.$$

Theorem B.1.1 (Carathéodory Extension Theorem). *Let Ω be a set, $\mathcal{A}_0$ an algebra on Ω and $\mathcal{A} = \sigma(\mathcal{A}_0)$. If μ_0 is a countably additive set function on $\mathcal{A}_0$, then there exists a measure μ on $(\Omega, \mathcal{A})$ such that*

$$\mu = \mu_0 \quad on \ \ \mathcal{A}_0.$$

If μ_0 is finite, then the extension is unique.

For proofs of the above and further discussion, we refer the reader to Chapter 1 and Appendix 1 of Williams (1991) and the appendix in Durrett (1996).

Returning to the motivating example $\Omega = I\!R$, we say that $A \subset I\!R$ belongs to the collection of sets $\mathcal{A}_0$ if A can be written as

$$A = (a_1, b_1] \cup \ldots \cup (a_r, b_r],$$

where $r \in I\!N$, $-\infty \leq a_1 < b_1 \leq \ldots \leq a_r < b_r \leq \infty$. It can be shown that $\mathcal{A}_0$ is an algebra and $\sigma(\mathcal{A}_0) = \mathcal{B}$. For A as above define

$$\mu_0(A) = \sum_{k=1}^{r} (b_k - a_k).$$

μ_0 is well-defined and countably additive on $\mathcal{A}_0$. As intervals belong to $\mathcal{A}_0$ our geometric intuition of length is preserved. Now by Carathéodory's extension theorem there exists a measure μ on $(\Omega, \mathcal{B})$ extending μ_0 on $(\Omega, \mathcal{A}_0)$. This μ is called Lebesgue measure.

With the same approach we can generalise:

(i) the area of rectangles $R = (a_1, b_1) \times (a_2, b_2)$ – with or without any of its perimeter included – given by $\mu(R) = (b_1 - a_1) \times (b_2 - a_2)$ to Lebesgue measure on Borel sets in $I\!R^2$;
(ii) the volume of cuboids $C = (a_1, b_1) \times (a_2, b_2) \times (a_3, b_3)$ given by

$$\mu(C) = (b_1 - a_1) \cdot (b_2 - a_2) \cdot (b_3 - a_3)$$

to Lebesgue measure on Borel sets in $I\!R^3$;
(iii) and similarly in k-dimensional Euclidean space $I\!R^k$. We start with the formula for a k-dimensional box,

$$\mu \left(\prod_{i=1}^{k} (a_i, b_i) \right) = \prod_{i=1}^{k} (b_i - a_i),$$

and obtain Lebesgue measure μ, defined on $\mathcal{B}$, in $I\!R^k$.

We are mostly concerned with a special class of measures:

Definition B.1.5. *A measure $I\!P$ on a measurable space $(\Omega, \mathcal{A})$ is called a probability measure if*

$$I\!P(\Omega) = 1.$$

The triple $(\Omega, \mathcal{A}, I\!P)$ is called a probability space.

Observe that the above lemma and Carathéodory's extension theorem guarantee uniqueness if we construct a probability measure using the above procedure. For example the unit cube $[0, 1]^k$ in $I\!R^k$ has (Lebesgue) measure 1. Using $\Omega = [0, 1]^k$ as the underlying set in the above construction we find a unique probability (which equals length/area/volume if $k = 1/2/3$).

If a property holds everywhere except on a set of measure zero, we say it holds *almost everywhere* (a.e.). If it holds everywhere except on a set of probability zero, we say it holds *almost surely* (a.s.) (or, with probability one).

Roughly speaking, one uses addition in countable (or finite) situations, integration in uncountable ones. As the key measure-theoretic axiom of countable additivity above concerns addition, countably infinite situations (such as we meet in discrete time) fit well with measure theory. By contrast, uncountable situations (such as we meet in continuous time) do not – or at least, are considerably harder to handle. This is why the discrete-time setting of Chapters 3, 4 is easier than,

and precedes, the continuous-time setting of Chapters 5, 6. Our strategy is to do as much as possible to introduce the key ideas – economic, financial and mathematical – in discrete time (which, because we work with a finite time-horizon, the expiry time T, is actually a finite situation), before treating the harder case of continuous time.

B.2 Integral

Let $(\Omega, \mathcal{A})$ be a measurable space. We want to define integration for a suitable class of real-valued functions.

Definition B.2.1. *Let $f : \Omega \to \mathbb{R}$. For $A \subset \mathbb{R}$ define $f^{-1}(A) = \{\omega \in \Omega : f(\omega) \in A\}$. f is called ($\mathcal{A}$-) measurable if*

$$f^{-1}(B) \in \mathcal{A} \;\; for \; all \;\; B \in \mathcal{B}.$$

Let μ be a measure on $(\Omega, \mathcal{A})$. Our aim now is to define, for suitable measurable functions, the (Lebesgue) integral with respect to μ. We will denote this integral by

$$\mu(f) = \int_\Omega f \, d\mu = \int_\Omega f(\omega) \mu(d\omega).$$

We start with the simplest functions. If $A \in \mathcal{A}$ the indicator function $\mathbf{1}_A(\omega)$ is defined by

$$\mathbf{1}_A(\omega) = \left\{ \begin{array}{ll} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A. \end{array} \right.$$

Then define $\mu(\mathbf{1}_A) = \mu(A)$.

The next step extends the definition to simple functions. A function f is called *simple* if it is a finite linear combination of indicators: $f = \sum_{i=1}^n c_i \mathbf{1}_{A_i}$ for constants c_i and indicator functions $\mathbf{1}_{A_i}$ of measurable sets A_i. One then extends the definition of the integral from indicator functions to simple functions by linearity:

$$\mu \left(\sum_{i=1}^n c_i \mathbf{1}_{A_i} \right) := \sum_{i=1}^n c_i \mu(\mathbf{1}_{A_i}) = \sum_{i=1}^n c_i \mu(A_i),$$

for constants c_i and indicators of measurable sets A_i.

If f is a non-negative measurable function, we define

$$\mu(f) := \sup\{\mu(f_0) : f_0 \text{ simple}, f_0 \le f\}.$$

The key result in integration theory, which we must use here to guarantee that the integral for non-negative measurable functions is well-defined is:

Theorem B.2.1 (Monotone Convergence Theorem). *If (f_n) is a sequence of non-negative measurable functions such that f_n is strictly monotonic increasing to a function f (which is then also measurable), then $\mu(f_n) \to \mu(f) \le \infty$.*

We quote that we can construct each non-negative measurable f as the increasing limit of a sequence of simple functions f_n:

$$f_n(\omega) \uparrow f(\omega) \quad \text{for all } \omega \in \Omega \quad (n \to \infty), \;\; f_n \;\; \text{simple}.$$

Using the monotone convergence theorem we can thus obtain the integral of f as

$$\mu(f) := \lim_{n \to \infty} \mu(f_n).$$

Since f_n increases in n, so does $\mu(f_n)$ (the integral is order-preserving), so either $\mu(f_n)$ increases to a finite limit, or diverges to ∞. In the first case, we say f is *(Lebesgue-) integrable* with *(Lebesgue-) integral* $\mu(f) = \lim \mu(f_n)$.

Finally if f is a measurable function that may change sign, we split it into its positive and negative parts, $f_{\pm}$:

$$f_+(\omega) := \max(f(\omega), 0), \quad f_-(\omega) := -\min(f(\omega), 0),$$
$$f(\omega) = f_+(\omega) - f_-(\omega), \quad |f(\omega)| = f_+(\omega) + f_-(\omega).$$

If both f_+ and f_- are integrable, we say that f is too, and define

$$\mu(f) := \mu(f_+) - \mu(f_-).$$

Thus, in particular, $|f|$ is also integrable, and

$$\mu(|f|) = \mu(f_+) + \mu(f_-).$$

The Lebesgue integral thus defined is, by construction, an absolute integral: f is integrable iff $|f|$ is integrable. Thus, for instance, the well-known formula

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}$$

has no meaning for Lebesgue integrals, since $\int_1^\infty \frac{|\sin x|}{x} dx$ diverges to $+\infty$ like $\int_1^\infty \frac{1}{x} dx$. It has to be replaced by the limit relation

$$\int_0^X \frac{\sin x}{x} dx \to \frac{\pi}{2} \quad (X \to \infty).$$

The class of (Lebesgue-) integrable functions f on Ω is written $\mathcal{L}(\Omega)$ or (for reasons explained below) $\mathcal{L}^1(\Omega)$ – abbreviated to $\mathcal{L}^1$ or $\mathcal{L}$.

For $p \geq 1$, the $\mathcal{L}^p$ space $\mathcal{L}^p(\Omega)$ on Ω is the space of measurable functions f with $\mathcal{L}^p$-norm

$$\|f\|_p := \left(\int_\Omega |f|^p \, d\mu \right)^{\frac{1}{p}} < \infty.$$

The case $p = 2$ gives $\mathcal{L}^2$, which is particular important as it is a *Hilbert space* (Appendix A).

Turning now to the special case $\Omega = I\!\!R^k$ we recall the well-known *Riemann integral*. Mathematics undergraduates are taught the Riemann integral (G.B. Riemann (1826–1866)) as their first rigorous treatment of integration theory – essentially this is just a rigorisation of the school integral. It is much easier to set up than the Lebesgue integral, but much harder to manipulate.

For finite intervals $[a, b]$,we quote:

(i) for any function f Riemann-integrable on $[a, b]$, it is Lebesgue-integrable to the same value (but many more functions are Lebesgue integrable);

(ii) f is Riemann-integrable on $[a, b]$ iff it is continuous a.e. on $[a, b]$. Thus the question, 'Which functions are Riemann-integrable?' cannot be answered without the language of measure theory – which gives one the technically superior Lebesgue integral anyway.

Suppose that $F(x)$ is a non-decreasing function on $I\!\!R$:

$$F(x) \leq F(y) \quad \text{if} \quad x \leq y.$$

Such functions can have at most countably many discontinuities, which are at worst jumps. We may without loss redefine F at jumps so as to be right-continuous. We now generalise the starting points above:

- Measure. We take $\mu((a, b]) := F(b) - F(a)$.

- Integral. We have $\mu(\mathbf{1}_{(a,b]}) = \mu((a,b]) = F(b) - F(a)$.

We may now follow through the successive extension procedures used above. We obtain:

- Lebesgue-Stieltjes measure μ_F,

- Lebesgue-Stieltjes integral $\mu_F(f) = \int f d\mu_F$, or even $\int f dF$.

The approach generalises to higher dimensions; we omit further details.

If instead of being monotone non-decreasing, F is the difference of two such functions, $F = F_1 - F_2$, we can define the integrals $\int f dF_1$, $\int f dF_2$ as above, and then define

$$\int f dF = \int f d(F_1 - F_2) := \int f dF_1 - \int f dF_2.$$

If $[a, b]$ is a finite interval and F is defined on $[a, b]$, a finite collection of points, $x_0, x_1, \ldots, x_n$ with $a = x_0 < x_1 < \cdots < x_n = b$, is called a *partition* of $[a, b]$, $\mathcal{P}$ say. The sum $\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|$ is called the variation of F over the partition. The least upper bound of this over all partitions $\mathcal{P}$ is called the *variation* of F over the interval $[a, b]$, $V_a^b(F)$:

$$V_a^b(F) := \sup_{\mathcal{P}} \sum |F(x_i) - F(x_{i-1})|.$$

This may be $+\infty$; but if $V_a^b(F) < \infty$, F is said to be of *bounded variation* on $[a, b]$, $F \in BV_a^b$. If F is of bounded variation on all finite intervals, F is said to be locally of bounded variation, $F \in BV_{\text{loc}}$; if F is of bounded variation on the real line $I\!R$, F is of bounded variation, $F \in BV$.

We quote that the following two properties are equivalent:

(i) F is locally of bounded variation,
(ii) F can be written as the difference $F = F_1 - F_2$ of two monotone functions.

So the above procedure defines the integral $\int f dF$ when the integrator F is of bounded variation.

Remark B.2.1. *(i) When we pass from discrete to continuous time, we will need to handle both 'smooth' paths and paths that vary by jumps – of bounded variation – and 'rough' ones – of unbounded variation but bounded quadratic variation;*
(ii) The Lebesgue-Stieltjes integral $\int g(x) dF(x)$ is needed to express the expectation $I\!Eg(X)$, where X is random variable with distribution function F and g a suitable function.

B.3 Probability

As we remarked in the introduction of this chapter, the mathematical theory of probability can be traced to 1654, to correspondence between Pascal (1623–1662) and Fermat (1601–1665). However, the theory remained both incomplete and non-rigorous until the 20th century. It turns out that the Lebesgue theory of measure and integral sketched above is exactly the machinery needed to construct a rigorous theory of probability adequate for modelling reality (option pricing, etc.) for us. This was realised by Kolmogorov (1903-1987), whose classic book of 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundations of Probability Theory), Kolmogorov (1933), inaugurated the modern era in probability.

Recall from your first course on probability that, to describe a random experiment mathematically, we begin with the *sample space* Ω, the set of all possible outcomes. Each point ω of Ω, or *sample point*, represents a possible – random – outcome of performing the random experiment. For a set $A \subseteq \Omega$ of points ω we want to know the probability $I\!P(A)$ (or $\Pr(A), \text{pr}(A)$). We clearly want

(i) $I\!P(\emptyset) = 0, I\!P(\Omega) = 1$,

(ii) $I\!P(A) \geq 0$ for all A,

(iii) If $A_1, A_2, \ldots, A_n$ are disjoint, $I\!P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n I\!P(A_i)$ (finite additivity), which, as above we will strengthen to

(iii)* If $A_1, A_2 \ldots$ (*ad inf.*) are disjoint,

$$I\!P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} I\!P(A_i) \qquad \text{(countable additivity).}$$

(iv) If $B \subseteq A$ and $I\!P(A) = 0$, then $I\!P(B) = 0$ (completeness).

Then by (i) and (iii) (with $A = A_1$, $\Omega \setminus A = A_2$),

$$I\!P(A^c) = I\!P(\Omega \setminus A) = 1 - I\!P(A).$$

So the class $\mathcal{F}$ of subsets of Ω whose probabilities $I\!P(A)$ are defined (call such A *events*) should be closed under countable, disjoint unions and complements, and contain the empty set $\emptyset$ and the whole space Ω. Therefore $\mathcal{F}$ should be a σ-algebra and $I\!P$ should be defined on $\mathcal{F}$ according to Definition 2.1.5. Repeating this:

Definition B.3.1. *A probability space, or Kolmogorov triple, is a triple $(\Omega, \mathcal{F}, I\!P)$ satisfying Kolmogorov axioms (i),(ii),(iii)*, (iv) above.*

A probability space is a mathematical model of a random experiment.

Often we quantify outcomes ω of random experiments by defining a real-valued function X on Ω, i.e. $X : \Omega \to I\!R$. If such a function is measurable it is called a random variable.

Definition B.3.2. *Let $(\Omega, \mathcal{F}, I\!P)$ be a probability space. A random variable (vector) X is a function $X : \Omega \to I\!R$ $(X : \Omega \to I\!R^k)$ such that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}(I\!R)$ $(B \in \mathcal{B}(I\!R^k))$.*

In particular we have for a random variable X that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in I\!R$. Hence we can define the *distribution function* F_X of X by

$$F_X(x) := I\!P\left(\{\omega : X(\omega) \leq x\}\right).$$

The smallest σ-algebra containing all the sets $\{\omega : X(\omega) \leq x\}$ for all real x (equivalently, $\{X < x\}$, $\{X \geq x\}$, $\{X > x\}$) is called the σ-algebra *generated* by X, written $\sigma(X)$. Thus,

$$X \text{ is } \mathcal{F} - \text{measurable (is a random variable) iff } \sigma(X) \subseteq \mathcal{F}.$$

The events in the σ-algebra generated by X are the events $\{\omega : X(\omega) \in B\}$, where B runs through the Borel σ-algebra on the line. When the (random) value $X(\omega)$ is *known*, we know *which* of these events have happened.

Interpretation.

Think of $\sigma(X)$ as representing *what we know when we know X*, or in other words *the information contained in X* (or in knowledge of X). This is reflected in the following result, due to J.L. Doob, which we quote:

$$\sigma(X) \subseteq \sigma(Y) \qquad \text{if and only if} \qquad X = g(Y)$$

for some measurable function g. For, knowing Y means we know $X := g(Y)$ – but not vice versa, unless the function g is one-to-one (injective), when the inverse function g^{-1} exists, and we can go back via $Y = g^{-1}(X)$.

Note.

An extended discussion of generated σ-algebras in the finite case is given in Dothan's book Dothan (1990), Chapter 3. Although technically avoidable, this is useful preparation for the general case, needed for continuous time.

A measure (§2.1) determines an integral (§2.2). A probability measure $\mathbb{P}$, being a special kind of measure (a measure of total mass one) determines a special kind of integral, called an expectation.

Definition B.3.3. *The expectation $\mathbb{E}$ of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is defined by*

$$\mathbb{E}X := \int_\Omega X d\mathbb{P}, \quad or \quad \int_\Omega X(\omega) d\mathbb{P}(\omega).$$

The expectation – also called the mean – describes the location of a distribution (and so is called a location parameter). Information about the scale of a distribution (the corresponding scale parameter) is obtained by considering the variance

$$\mathbb{V}ar(X) := \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}\left(X^2\right) - (\mathbb{E}X)^2.$$

If X is real-valued, say, with distribution function F, recall that $\mathbb{E}X$ is defined in your first course on probability by

$$\mathbb{E}X := \int x f(x) dx \quad \text{if } X \text{ has a density } f$$

or if X is discrete, taking values $x_n (n = 1, 2, \ldots)$ with probability function $f(x_n)(\geq 0)$ $(\sum_{x_n} f(x_n) = 1)$,

$$\mathbb{E}X := \sum x_n f(x_n).$$

These two formulae are the special cases (for the density and discrete cases) of the general formula

$$\mathbb{E}X := \int_{-\infty}^{\infty} x dF(x)$$

where the integral on the right is a Lebesgue-Stieltjes integral. This in turn agrees with the definition above, since if F is the distribution function of X,

$$\int_\Omega X d\mathbb{P} = \int_{-\infty}^{\infty} x dF(x)$$

follows by the change of variable formula for the measure-theoretic integral, on applying the map $X : \Omega \to \mathbb{R}$ (we quote this: see any book on measure theory, e.g. Dudley (1989)).

Clearly the expectation operator $\mathbb{E}$ is linear. It even becomes multiplicative if we consider independent random variables.

Definition B.3.4. *Random variables $X_1, \ldots, X_n$ are independent if whenever $A_i \in \mathcal{B}$ for $i = 1, \ldots n$ we have*

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \in A_i\}).$$

Using Lemma B.1.1 we can give a more tractable condition for independence:

Lemma B.3.1. *In order for $X_1, \ldots, X_n$ to be independent it is necessary and sufficient that for all $x_1, \ldots x_n \in (-\infty, \infty]$,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \leq x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \leq x_i\}).$$

Now using the usual measure-theoretic steps (going from simple to integrable functions) it is easy to show:

Theorem B.3.1 (Multiplication Theorem). *If $X_1, \ldots, X_n$ are independent and $\mathbb{E}\,|X_i| < \infty$, $i = 1, \ldots, n$, then*

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

We now review the distributions we will mainly use in our models of financial markets.

Examples.

(i) *Bernoulli* distribution. Recall our arbitrage-pricing example from §1.4. There we were given a stock with price $S(0)$ at time $t = 0$. We assumed that after a period of time Δt the stock price could have only one of two values, either $S(\Delta t) = e^u S(0)$ with probability p or $S(\Delta t) = e^d S(0)$ with probability $1 - p$ $(u, d \in \mathbb{R})$. Let $R(\Delta t) = r(1)$ be a random variable modelling the logarithm of the stock return over the period $[0, \Delta t]$; then

$$\mathbb{P}(r(1) = u) = p \quad \text{and} \quad \mathbb{P}(r(1) = d) = 1 - p.$$

We say that $r(1)$ is distributed according to a Bernoulli distribution. Clearly $\mathbb{E}(r(1)) = up + d(1 - p)$ and $\mathbb{V}ar(r(1)) = u^2 p + d^2(1 - p) - (\mathbb{E}X)^2$.

The standard case of a Bernoulli distribution is given by choosing $u = 1, d = 0$ (which is not a very useful choice in financial modelling).

(ii) *Binomial* distribution. If we consider the logarithm of the stock return over n periods (of equal length), say over $[0, T]$, then subdividing into the periods $1, \ldots, n$ we have

$$
\begin{aligned}
R(T) &= \log\left[\frac{S(T)}{S(0)}\right] = \log\left[\frac{S(T)}{S(T - \Delta t)} \cdots \frac{S(\Delta t)}{S(0)}\right] \\
&= \log\left[\frac{S(T)}{S(T - \Delta t)}\right] + \ldots + \log\left[\frac{S(\Delta t)}{S(0)}\right] = r(n) + \ldots + r(1).
\end{aligned}
$$

Assuming that $r(i)$, $i = 1, \ldots, n$ are independent and each $r(i)$ is Bernoulli distributed as above we have that $R(T) = \sum_{i=1}^{n} r(i)$ is binomially distributed. Linearity of the expectation operator and independence yield $\mathbb{E}(R(T)) = \sum_{i=1}^{n} \mathbb{E}(r(i))$ and $\mathbb{V}ar(R(T)) = \sum_{i=1}^{n} \mathbb{V}ar(r(i))$.

Again for the standard case one would use $u = 1, d = 0$. The shorthand notation for a binomial random variable X is then $X \sim B(n, p)$ and we can compute

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{(n-k)}, \quad \mathbb{E}(X) = np, \quad \mathbb{V}ar(X) = np(1 - p).$$

(iii) *Normal* distribution. As we will show in the sequel the limit of a sequence of appropriate normalised binomial distributions is the (standard) normal distribution. We say a random variable X is normally distributed with parameters μ, σ^2, in short $X \sim N(\mu, \sigma^2)$, if X has density function

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

One can show that $\mathbb{E}(X) = \mu$ and $\mathbb{V}ar(X) = \sigma^2$, and thus a normally distributed random variable is fully described by knowledge of its mean and variance.

Returning to the above example, one of the key results of this text will be that the limiting model of a sequence of financial markets with one-period asset returns modelled by a Bernoulli distribution is a model where the distribution of the logarithms of instantaneous asset returns is normal. That means $S(t + \Delta t)/S(t)$ is lognormally distributed (i.e. $\log(S(t + \Delta t)/S(t))$ is normally distributed). Although rejected by many empirical studies (see Eberlein and Keller (1995) for a

recent overview), such a model seems to be the standard in use among financial practitioners (and we will call it the standard model in the following). The main arguments against using normally distributed random variables for modelling log-returns (i.e. log-normal distributions for returns) are asymmetry and (semi-) heavy tails. We know that distributions of financial asset returns are generally rather close to being symmetric around zero, but there is a definite tendency towards asymmetry. This may be explained by the fact that the markets react differently to positive as opposed to negative information (see Shephard (1996) §1.3.4). Since the normal distribution is symmetric it is not possible to incorporate this empirical fact in the standard model. Financial time series also suggest modelling by probability distributions whose densities behave for $x \to \pm\infty$ as

$$|x|^\rho \exp\{-\sigma |x|\}$$

with $\rho \in I\!R, \sigma > 0$. This means that we should replace the normal distribution with a distribution with heavier tails. Such a model like this would exhibit higher probabilities of extreme events and the passage from ordinary observations (around the mean) to extreme observations would be more sudden. Among suggested (classes of) distributions to be used to address these facts is the class of *hyperbolic* distributions (see Eberlein and Keller (1995) and §2.12 below), and more general distributions of normal inverse Gaussian type (see Barndorff-Nielsen (1998), Rydberg (1996), Rydberg (1997)) appear to be very promising.

(iv) *Poisson* distribution. Sometimes we want to incorporate in our model of financial markets the possibility of sudden jumps. Using the standard model we model the asset price process by a continuous stochastic process, so we need an additional process generating the jumps. To do this we use point processes in general and the *Poisson process* in particular. For a Poisson process the probability of a jump (and no jump respectively) during a small interval Δt are approximately

$$I\!P(\nu(1) = 1) \approx \lambda\Delta t \quad \text{and} \quad I\!P(\nu(1) = 0) \approx 1 - \lambda\Delta t,$$

where λ is a positive constant called the rate or intensity. Modelling small intervals in such a way we get for the number of jumps $N(T) = \nu(1) + \ldots + \nu(n)$ in the interval $[0, T]$ the probability function

$$I\!P(N(T) = k) = \frac{e^{-\lambda T}(\lambda T)^k}{k!}, \quad k = 0, 1, \ldots$$

and we say the process $N(T)$ has a Poisson distribution with parameter λT. We can show $I\!E(N(T)) = \lambda T$ and $W\!ar(N(T)) = \lambda T$.

Glossary.

Table B.1 summarises the two parallel languages, measure-theoretic and probabilistic, which we have established.

Measure	Probability
Integral	Expectation
Measurable set	Event
Measurable function	Random variable
Almost-everywhere (a.e.)	Almost-surely (a.s.)

Table B.1: Measure-theoretic and probabilistic languages

B.4 Equivalent Measures and Radon-Nikodým Derivatives

Given two measures $I\!P$ and Q defined on the same σ-algebra $\mathcal{F}$, we say that $I\!P$ is *absolutely continuous* with respect to Q, written

$$I\!P << Q$$

if $I\!\!P(A) = 0$, whenever $\mathbf{Q}(A) = 0$, $A \in \mathcal{F}$. We quote from measure theory the vitally important *Radon-Nikodým theorem*:

Theorem B.4.1 (Radon-Nikodým). *$I\!\!P << \mathbf{Q}$ iff there exists a ($\mathcal{F}$-) measurable function f such that*

$$I\!\!P(A) = \int_A f d\mathbf{Q} \quad \forall A \in \mathcal{F}.$$

(Note that since the integral of anything over a null set is zero, any $I\!\!P$ so representable is certainly absolutely continuous with respect to $\mathbf{Q}$ – the point is that the converse holds.)

Since $I\!\!P(A) = \int_A dI\!\!P$, this says that $\int_A dI\!\!P = \int_A f d\mathbf{Q}$ for all $A \in \mathcal{F}$. By analogy with the chain rule of ordinary calculus, we write $dI\!\!P/d\mathbf{Q}$ for f; then

$$\int_A dI\!\!P = \int_A \frac{dI\!\!P}{d\mathbf{Q}} d\mathbf{Q} \quad \forall A \in \mathcal{F}.$$

Symbolically,

$$\text{if} \quad I\!\!P << \mathbf{Q}, \quad dI\!\!P = \frac{dI\!\!P}{d\mathbf{Q}} d\mathbf{Q}.$$

The measurable function (random variable) $dI\!\!P/d\mathbf{Q}$ is called the *Radon-Nikodým derivative* (RN-derivative) of $I\!\!P$ with respect to $\mathbf{Q}$.

If $I\!\!P << \mathbf{Q}$ and also $\mathbf{Q} << I\!\!P$, we call $I\!\!P$ and $\mathbf{Q}$ *equivalent* measures, written $I\!\!P \sim \mathbf{Q}$. Then $dI\!\!P/d\mathbf{Q}$ and $d\mathbf{Q}/dI\!\!P$ both exist, and

$$\frac{dI\!\!P}{d\mathbf{Q}} = 1 / \frac{d\mathbf{Q}}{dI\!\!P}.$$

For $I\!\!P \sim \mathbf{Q}$, $I\!\!P(A) = 0$ iff $\mathbf{Q}(A) = 0$: $I\!\!P$ and $\mathbf{Q}$ *have the same null sets*. Taking negations: $I\!\!P \sim \mathbf{Q}$ iff $I\!\!P, \mathbf{Q}$ have the same sets of positive measure. Taking complements: $I\!\!P \sim \mathbf{Q}$ iff $I\!\!P, \mathbf{Q}$ have the same sets of probability one (the same a.s. sets). Thus the following are equivalent:

$$\begin{aligned} I\!\!P \sim \mathbf{Q} \quad &\text{iff} \quad I\!\!P, \mathbf{Q} \text{ have the same null sets,} \\ &\text{iff} \quad I\!\!P, \mathbf{Q} \text{ have the same a.s. sets,} \\ &\text{iff} \quad I\!\!P, \mathbf{Q} \text{ have the same sets of positive measure.} \end{aligned}$$

Far from being an abstract theoretical result, the Radon-Nikodým theorem is of key practical importance, in two ways:

(a) It is the key to the concept of conditioning (§2.5, §2.6 below), which is of central importance throughout,

(b) The concept of equivalent measures is central to the key idea of mathematical finance, *risk-neutrality*, and hence to its main results, the Black-Scholes formula, fundamental theorem of asset pricing, etc. The key to all this is that prices should be the discounted expected values under an equivalent martingale measure. Thus equivalent measures, and the operation of change of measure, are of central economic and financial importance. We shall return to this later in connection with the main mathematical result on change of measure, Girsanov's theorem (see §5.7).

B.5 Conditional expectation

For basic **events** define

$$I\!\!P(A|B) := I\!\!P(A \cap B)/I\!\!P(B) \quad \text{if} \quad I\!\!P(B) > 0. \tag{B.1}$$

From this definition, we get the **multiplication rule**

$$I\!\!P(A \cap B) = I\!\!P(A|B)I\!\!P(B).$$

Using the partition equation $I\!\!P(B) = \sum_n I\!\!P(B|A_n)I\!\!P(A_n)$ with (A_n) a finite or countable partition of Ω, we get the **Bayes rule**

$$I\!\!P(A_i|B) = \frac{I\!\!P(A_i)I\!\!P(B|A_i)}{\sum_j I\!\!P(A_j)I\!\!P(B|A_j)}.$$

We can always write $I\!\!P(A) = I\!\!E(\mathbf{1}_A)$ with $\mathbf{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbf{1}_A(\omega) = 0$ otherwise. Then the above can be written

$$I\!\!E(\mathbf{1}_A|B) = \frac{I\!\!E(\mathbf{1}_A\mathbf{1}_B)}{I\!\!P(B)} \tag{B.2}$$

This suggest defining, for suitable random variables X, the $I\!\!P$-average of X over B as

$$I\!\!E(X|B) = \frac{I\!\!E(X\mathbf{1}_B)}{I\!\!P(B)}. \tag{B.3}$$

Consider now **discrete** random variables X and Y. Assume X takes values $x_1, \ldots, x_m$ with probabilities $f_1(x_i) > 0$, Y takes values $y_1, \ldots, y_n$ with probabilities $f_2(y_j) > 0$, while the vector (X, Y) takes values (x_i, y_j) with probabilities $f(x_i, y_j) > 0$. Then the **marginal distributions** are

$$f_1(x_i) = \sum_{j=1}^n f(x_i, y_j) \quad \text{and} \quad f_2(y_j) = \sum_{i=1}^m f(x_i, y_j).$$

We can use the standard definition above for the events $\{Y = y_j\}$ and $\{X = x_i\}$ to get

$$I\!\!P(Y = y_j|X = x_i) = \frac{I\!\!P(X = x_i, Y = y_j)}{I\!\!P(X = x_i)} = \frac{f(x_i, y_j)}{f_1(x_i)}.$$

Thus conditional on $X = x_i$ (given the information $X = x_i$), Y takes on the values $y_1, \ldots, y_n$ with (conditional) probabilities

$$f_{Y|X}(y_j|x_i) = \frac{f(x_i, y_j)}{f_1(x_i)}.$$

So we can compute its expectation as usual:

$$I\!\!E(Y|X = x_i) = \sum_j y_j f_{Y|X}(y_j|x_i) = \frac{\sum_j y_j f(x_i, y_j)}{f_1(x_i)}.$$

Now define the random variable $Z = I\!\!E(Y|X)$, the conditional expectation of Y given X, as follows:

$$\text{if } X(\omega) = x_i, \text{ then } Z(\omega) = I\!\!E(Y|X = x_i) = z_i \text{ (say)}$$

Observe that in this case Z is given by a 'nice' function of X. However, a more abstract property also holds true. Since Z is constant on the the sets $\{X = x_i\}$ it is $\sigma(X)$-measurable (these sets generate the σ-algebra). Furthermore

$$\int_{\{X=x_i\}} Z dI\!\!P = z_i I\!\!P(X = x_i) = \sum_j y_j f_{Y|X}(y_j|x_i)I\!\!P(X = x_i)$$

$$= \sum_j y_j I\!\!P(Y = y_j; X = x_i) = \int_{\{X=x_i\}} Y dI\!\!P.$$

Since the $\{X = x_i\}$ generate $\sigma(X)$, this implies

$$\int_G Z dI\!\!P = \int_G Y dI\!\!P \quad \forall\, G \in \sigma(X).$$

Density case. If the random vector (X, Y) has density $f(x, y)$, then X has (marginal) density $f_1(x) := \int_{-\infty}^{\infty} f(x, y)dy$, Y has (marginal) density $f_2(y) := \int_{-\infty}^{\infty} f(x, y)dx$. The conditional density of Y given $X = x$ is:

$$f_{Y|X}(y|x) := \frac{f(x, y)}{f_1(x)}.$$

Its expectation is

$$I\!E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \frac{\int_{-\infty}^{\infty} y f(x, y)dy}{f_1(x)}.$$

So we define

$$c(x) = \begin{cases} I\!E(Y|X = x) & \text{if } f_1(x) > 0 \\ 0 & \text{if } f_1(x) = 0, \end{cases}$$

and call $c(X)$ the conditional expectation of Y given X, denoted by $I\!E(Y|X)$. Observe that on sets with probability zero (i.e $\{\omega : X(\omega) = x; f_1(x) = 0\}$) the choice of $c(x)$ is arbitrary, hence $I\!E(Y|X)$ is only defined up to a set of probability zero; we speak of different versions in such cases. With this definition we again find

$$\int_G c(X)d I\!P = \int_G Y d I\!P \quad \forall \, G \in \sigma(X).$$

Indeed, for sets G with $G = \{\omega : X(\omega) \in B\}$ with B a Borel set, we find by Fubini's theorem

$$\begin{aligned} \int_G c(X)d I\!P &= \int_{-\infty}^{\infty} \mathbf{1}_B(x)c(x)f_1(x)dx \\ &= \int_{-\infty}^{\infty} \mathbf{1}_B(x)f_1(x) \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_B(x)y f(x, y)dy dx = \int_G Y d I\!P. \end{aligned}$$

Now these sets G generate $\sigma(X)$ and by a standard technique (the π-systems lemma, see Williams (2001), §2.3) the claim is true for all $G \in \sigma(X)$.

Example. Bivariate Normal Distribution,

$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

$$I\!E(Y|X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),$$

the familiar regression line of statistics (linear model) – see Exercise 2.6.

General case. Here, we follow Kolmogorov's construction using the Radon-Nikodým theorem. Suppose that $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, $\mathcal{G} \subset \mathcal{F}$. If Y is a non-negative random variable with $I\!E Y < \infty$, then

$$\mathbf{Q}(G) := \int_G Y d I\!P \qquad (G \in \mathcal{G})$$

is non-negative, σ-additive – because

$$\int_G Y d I\!P = \sum_n \int_{G_n} Y d I\!P$$

if $G = \cup_n G_n$, G_n disjoint – and defined on the σ-algebra $\mathcal{G}$, so it is a measure on $\mathcal{G}$.

If $I\!P(G) = 0$, then $\mathbf{Q}(G) = 0$ also (the integral of anything over a null set is zero), so $\mathbf{Q} << I\!P$.

By the Radon-Nikodým theorem, there exists a Radon-Nikodým derivative of $\mathbf{Q}$ with respect to $I\!P$ on $\mathcal{G}$, which is $\mathcal{G}$-measurable. Following Kolmogorov, we call this Radon-Nikodým derivative the conditional expectation of Y given (or conditional on) $\mathcal{G}, I\!E(Y|\mathcal{G})$, whose existence we now have established. For Y that changes sign, split into $Y = Y^+ - Y^-$, and define $I\!E(Y|\mathcal{G}) := I\!E(Y^+|\mathcal{G}) - I\!E(Y^-|\mathcal{G})$. We summarize:

Definition B.5.1. *Let Y be a random variable with $I\!E(|Y|) < \infty$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. We call a random variable Z a version of the conditional expectation $I\!E(Y|\mathcal{G})$ of Y given $\mathcal{G}$, and write $Z = I\!E(Y|\mathcal{G})$, a.s., if*

(i) Z is $\mathcal{G}$-measurable;

(ii) $I\!E(|Z|) < \infty$;

(iii) for every set G in $\mathcal{G}$, we have

$$\int_G Y dI\!P = \int_G Z dI\!P \qquad \forall G \in \mathcal{G}. \tag{B.4}$$

Notation. Suppose $\mathcal{G} = \sigma(X_1, \ldots, X_n)$. Then

$$I\!E(Y|\mathcal{G}) = I\!E(Y|\sigma(X_1, \ldots, X_n)) =: I\!E(Y|X_1, \ldots, X_n),$$

and one can compare the general case with the motivating examples above.

To see the intuition behind conditional expectation, consider the following situation. Assume an experiment has been performed, i.e. $\omega \in \Omega$ has been realized. However, the only information we have is the set of values $X(\omega)$ for every $\mathcal{G}$-measurable random variable X. Then $Z(\omega) = I\!E(Y|\mathcal{G})(\omega)$ is the expected value of $Y(\omega)$ given this information.

We used the traditional approach to define conditional expectation via the Radon-Nikodým theorem. Alternatively, one can use Hilbert space projection theory (Neveu (1975) and Jacod and Protter (2000) follow this route). Indeed, for $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, I\!P)$ one can show that the conditional expectation $Z = I\!E(Y|\mathcal{G})$ is the least-squares-best $\mathcal{G}$-measurable predictor of Y: amongst all $\mathcal{G}$-measurable random variables it minimises the quadratic distance, i.e.

$$I\!E[(Y - I\!E(Y|\mathcal{G}))^2] = \min\{I\!E[(Y - X)^2] : X \quad \mathcal{G} - \text{measurable}\}.$$

Note.

1. To check that something is a conditional expectation: we have to check that it integrates the right way over the right sets (i.e., as in (B.4)).

2. From (B.4): if two things integrate the same way over all sets $B \in \mathcal{G}$, they have the same conditional expectation given $\mathcal{G}$.

3. For notational convenience, we shall pass between $I\!E(Y|\mathcal{G})$ and $I\!E_{\mathcal{G}}Y$ at will.

4. The conditional expectation thus defined coincides with any we may have already encountered – in regression or multivariate analysis, for example. However, this may not be immediately obvious. The conditional expectation defined above – via σ-algebras and the Radon-Nikodým theorem – is rightly called by Williams ((Williams 1991), p.84) 'the central definition of modern probability'. It may take a little getting used to. As with all important but non-obvious definitions, it proves its worth in action: see §2.6 below for properties of conditional expectations, and Chapter 3 for its use in studying stochastic processes, particularly martingales (which are defined in terms of conditional expectations).

We now discuss the fundamental properties of conditional expectation. From the definition linearity of conditional expectation follows from the linearity of the integral. Further properties are given by

Proposition B.5.1. *1. $\mathcal{G} = \{\emptyset, \Omega\}$, $I\!E(Y|\{\emptyset, \Omega\}) = I\!EY$.*

2. *If $\mathcal{G} = \mathcal{F}$, $I\!\!E(Y|\mathcal{F}) = Y$ $I\!\!P - a.s.$.*

3. *If Y is $\mathcal{G}$-measurable, $I\!\!E(Y|\mathcal{G}) = Y$ $I\!\!P - a.s.$.*

4. *Positivity. If $X \geq 0$, then $I\!\!E(X|\mathcal{G}) \geq 0$ $I\!\!P - a.s.$.*

5. *Taking out what is known. If Y is $\mathcal{G}$-measurable and bounded, $I\!\!E(YZ|\mathcal{G}) = Y I\!\!E(Z|\mathcal{G})$ $I\!\!P - a.s.$.*

6. *Tower property. If $\mathcal{G}_0 \subset \mathcal{G}$, $I\!\!E[I\!\!E(Y|\mathcal{G})|\mathcal{G}_0] = I\!\!E[Y|\mathcal{G}_0]$ $a.s.$.*

7. *Conditional mean formula. $I\!\!E[I\!\!E(Y|\mathcal{G})] = I\!\!E Y$ $I\!\!P - a.s.$*

8. *Role of independence. If Y is independent of $\mathcal{G}$, $I\!\!E(Y|\mathcal{G}) = I\!\!E Y$ $a.s.$*

9. *Conditional Jensen formula. If $c : I\!\!R \rightarrow I\!\!R$ is convex, and $I\!\!E|c(X)| < \infty$, then*

$$I\!\!E(c(X)|\mathcal{G}) \geq c\left(I\!\!E(X|\mathcal{G})\right).$$

Proof. 1. Here $\mathcal{G} = \{\emptyset, \Omega\}$ is the *smallest* possible σ-algebra (*any* σ-algebra of subsets of Ω contains $\emptyset$ and Ω), and represents 'knowing nothing'. We have to check (B.4) for $G = \emptyset$ and $G = \Omega$. For $G = \emptyset$ both sides are zero; for $G = \Omega$ both sides are $I\!\!E Y$.

2. Here $\mathcal{G} = \mathcal{F}$ is the *largest* possible σ-algebra, and represents 'knowing everything'. We have to check (B.4) for *all* sets $G \in \mathcal{F}$. The only integrand that integrates like Y over *all* sets is Y itself, or a function agreeing with Y except on a set of measure zero.

Note. When we condition on $\mathcal{F}$ ('knowing everything'), we *know* Y (because we know everything). There is thus no uncertainty left in Y to average out, so taking the conditional expectation (averaging out remaining randomness) has no effect, and leaves Y unaltered.

3. Recall that Y is *always* $\mathcal{F}$-measurable (this is the definition of Y being a random variable). For $\mathcal{G} \subset \mathcal{F}$, Y may not be $\mathcal{G}$-measurable, but if it is, the proof above applies with $\mathcal{G}$ in place of $\mathcal{F}$.

Note. To say that Y is $\mathcal{G}$-measurable is to say that Y is known given $\mathcal{G}$ – that is, when we are conditioning on $\mathcal{G}$. Then Y is no longer random (being known when $\mathcal{G}$ is given), and so counts as a constant when the conditioning is performed.

4. Let Z be a version of $I\!\!E(X|\mathcal{G})$. If $I\!\!P(Z < 0) > 0$, then for some n, the set

$$G := \{Z < -n^{-1}\} \in \mathcal{G} \quad \text{and} \quad I\!\!P(\{Z < -n^{-1}\}) > 0.$$

Thus

$$0 \leq I\!\!E(X\mathbf{1}_G) = I\!\!E(Z\mathbf{1}_G) < -n^{-1} I\!\!P(G) < 0,$$

which contradicts the positivity of X.

5. First, consider the case when Y is discrete. Then Y can be written as

$$Y = \sum_{n=1}^{N} b_n \mathbf{1}_{B_n},$$

for constants b_n and events $B_n \in \mathcal{G}$. Then for any $B \in \mathcal{G}$, $B \cap B_n \in \mathcal{G}$ also (as $\mathcal{G}$ is a σ-algebra), and using linearity and (B.4):

$$
\begin{aligned}
\int_B Y I\!\!E(Z|\mathcal{G}) dI\!\!P &= \int_B \left(\sum_{n=1}^{N} b_n \mathbf{1}_{B_n}\right) I\!\!E(Z|\mathcal{G}) dI\!\!P = \sum_{n=1}^{N} b_n \int_{B \cap B_n} I\!\!E(Z|\mathcal{G}) dI\!\!P \\
&= \sum_{n=1}^{N} b_n \int_{B \cap B_n} Z dI\!\!P = \int_B \sum_{n=1}^{N} b_n \mathbf{1}_{B_n} Z dI\!\!P \\
&= \int_B Y Z dI\!\!P.
\end{aligned}
$$

Since this holds for all $B \in \mathcal{G}$, the result holds by $(B.4)$.

For the general case, we approximate to a general random variable Y by a sequence of discrete random variables Y_n, for each of which the result holds as just proved. We omit details of the proof here, which involves the standard approximation steps based on the monotone convergence theorem from measure theory (see e.g. (Williams 1991), p.90, proof of (j)). We are thus left to show the $I\!\!E(|ZY|) < \infty$, which follows from the assumption that Y is bounded and $Z \in \mathcal{L}^1$.

6. $I\!\!E_{\mathcal{G}_0} I\!\!E_{\mathcal{G}} Y$ is $\mathcal{G}_0$-measurable, and for $C \in \mathcal{G}_0 \subset \mathcal{G}$, using the definition of $I\!\!E_{\mathcal{G}_0}, I\!\!E_{\mathcal{G}}$:

$$\int\limits_C I\!\!E_{\mathcal{G}_0}[I\!\!E_{\mathcal{G}} Y] d I\!\!P = \int\limits_C I\!\!E_{\mathcal{G}} Y d I\!\!P = \int\limits_C Y d I\!\!P.$$

So $I\!\!E_{\mathcal{G}_0}[I\!\!E_{\mathcal{G}} Y]$ satisfies the defining relation for $I\!\!E_{\mathcal{G}_0} Y$. Being also $\mathcal{G}_0$-measurable, it *is* $I\!\!E_{\mathcal{G}_0} Y$ (a.s.).

We also have:

6'. If $\mathcal{G}_0 \subset \mathcal{G}$, $I\!\!E[I\!\!E(Y|\mathcal{G}_0)|\mathcal{G}] = I\!\!E[Y|\mathcal{G}_0]$ *a.s.*.

Proof. $I\!\!E[Y|\mathcal{G}_0]$ is $\mathcal{G}_0$-measurable, so $\mathcal{G}$-measurable as $\mathcal{G}_0 \subset \mathcal{G}$, so $I\!\!E[.|\mathcal{G}]$ has no effect on it, by 3.

Note.

6, 6' are the two forms of the *iterated conditional expectations property*. When conditioning on two σ-algebras, one larger (finer), one smaller (coarser), the coarser rubs out the effect of the finer, either way round. This may be thought of as the *coarse-averaging property*: we shall use this term interchangeably with the iterated conditional expectations property (Williams (1991) uses the term *tower property*).

7. Take $\mathcal{G}_0 = \{\emptyset, \Omega\}$ in 6. and use 1.

8. If Y is independent of $\mathcal{G}$, Y is independent of $\mathbf{1}_B$ for every $B \in \mathcal{G}$. So by (B.4) and linearity,

$$\begin{aligned} \int\limits_B I\!\!E(Y|\mathcal{G}) d I\!\!P &= \int\limits_B Y d I\!\!P = \int\limits_\Omega \mathbf{1}_B Y d I\!\!P \\ &= I\!\!E(\mathbf{1}_B Y) = I\!\!E(\mathbf{1}_B) I\!\!E(Y) = \int\limits_B I\!\!E Y d I\!\!P, \end{aligned}$$

using the multiplication theorem for independent random variables. Since this holds for all $B \in \mathcal{G}$, the result follows by $(B.4)$.

9. Recall (see e.g. Williams (1991), §6.6a, §9.7h, §9.8h), that for every convex function there exists a countable sequence $((a_n, b_n))$ of points in $I\!\!R^2$ such that

$$c(x) = \sup_n (a_n x + b_n), \quad x \in I\!\!R.$$

For each fixed n we use 4. to see from $c(X) \geq a_n X + b_n$ that

$$I\!\!E[c(X)|\mathcal{G}] \geq a_n I\!\!E(X|\mathcal{G}) + b_n.$$

So,

$$I\!\!E[c(X)|\mathcal{G}] \geq \sup_n (a_n I\!\!E(X|\mathcal{G}) + b_n) = c(I\!\!E(X|\mathcal{G})).$$

$\blacksquare$

Remark B.5.1. *If in* $6, 6'$ *we take* $\mathcal{G} = \mathcal{G}_0$, *we obtain:*

$$I\!\!E[I\!\!E(X|\mathcal{G})|\mathcal{G}] = I\!\!E(X|\mathcal{G}).$$

Thus the map $X \to I\!\!E(X|\mathcal{G})$ *is idempotent: applying it twice is the same as applying it once. Hence we may identify the conditional expectation operator as a projection.*

B.6   Modes of Convergence

So far, we have dealt with one probability measure – or its expectation operator – at a time. We shall, however, have many occasions to consider a whole sequence of them, converging (in a suitable sense) to some limiting probability measure. Such situations arise, for example, whenever we approximate a financial model in continuous time (such as the continuous-time Black-Scholes model of §6.2) by a sequence of models in discrete time (such as the discrete-time Black-Scholes model of §4.6).

In the stochastic-process setting – such as the passage from discrete to continuous Black-Scholes models mentioned above – we need concepts beyond those we have to hand, which we develop later. We confine ourselves here to setting out what we need to discuss convergence of random variables, in the various senses that are useful.

The first idea that occurs to one is to use the ordinary convergence concept in this new setting, of random variables: then if X_n, X are random variables,

$$X_n \to X \quad (n \to \infty)$$

would be taken literally – as if the X_n, X were non-random. For instance, if X_n is the observed frequency of heads in a long series of n independent tosses of a fair coin, $X = 1/2$ the expected frequency, then the above in this case would be the man-in-the-street's idea of the 'law of averages'. It turns out that the above statement is *false* in this case, taken literally: some qualification is needed. However, the qualification needed is absolutely the minimal one imaginable: one merely needs to exclude a set of probability zero – that is, to assert convergence on a set of probability one ('almost surely'), rather than everywhere.

Definition B.6.1. *If X_n, X are random variables, we say X_n converges to X almost surely –*

$$X_n \to X \quad (n \to \infty) \quad a.s.$$

– if $X_n \to X$ with probability one – that is, if

$$I\!P(\{\omega : X_n(\omega) \to X(\omega) \quad as \quad n \to \infty\}) = 1.$$

The loose idea of the 'law of averages' has as its precise form a statement on convergence almost surely. This is Kolmogorov's *strong law of large numbers* (see e.g. (Williams 1991), §12.10), which is quite difficult to prove.

Weaker convergence concepts are also useful: they may hold under weaker conditions, or they may be easier to prove.

Definition B.6.2. *If X_n, X are random variables, we say that X_n converges to X in probability –*

$$X_n \to X \quad (n \to \infty) \quad in\ probability$$

- if, for all $\epsilon > 0$,
$$I\!P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \to 0 \quad (n \to \infty).$$

It turns out that convergence almost surely implies convergence in probability, but not in general conversely. Thus almost-sure convergence is a stronger convergence concept than convergence in probability. This comparison is reflected in the form the 'law of averages' takes for convergence in probability: this is called the *weak law of large numbers*, which as its name implies is a weaker form of the strong law of large numbers. It is correspondingly much easier to prove: indeed, we shall prove it in §2.8 below.

Recall the L^p-spaces of pth-power integrable functions (§2.2). We similarly define the L^p-spaces of pth-power integrable random variables: if $p \geq 1$ and X is a random variable with

$$\|X\|_p := (I\!\!E|X|^p)^{1/p} < \infty,$$

we say that $X \in L^p$ (or $L^p(\Omega, \mathcal{F}, I\!\!P)$ to be precise). For X_n, $X \in L^p$, there is a natural convergence concept: we say that X_n *converges to X in L^p*, or *in pth mean*,

$$X_n \to X \quad \text{in} \quad L^p,$$

if

$$\|X_n - X\|_p \to 0 \quad (n \to \infty),$$

that is, if

$$I\!\!E(|X_n - X|^p) \to 0 \quad (n \to \infty).$$

The cases $p = 1, 2$ are particularly important: if $X_n \to X$ in L^1, we say that $X_n \to X$ *in mean*; if $X_n \to X$ in L^2 we say that $X_n \to X$ *in mean square*. Convergence in pth mean is not directly comparable with convergence almost surely (of course, we have to restrict to random variables in L^p for the comparison even to be meaningful): neither implies the other. Both, however, imply convergence in probability.

All the modes of convergence discussed so far involve the *values* of random variables. Often, however, it is only the *distributions* of random variables that matter. In such cases, the natural mode of convergence is the following:

Definition B.6.3. *We say that random variables X_n converge to X in distribution if the distribution functions of X_n converge to that of X at all points of continuity of the latter:*

$$X_n \to X \quad \text{in distribution}$$

if

$$I\!\!P(\{X_n \le x\}) \to I\!\!P(\{X \le x\}) \quad (n \to \infty)$$

for all points x at which the right-hand side is continuous.

The restriction to continuity points x of the limit seems awkward at first, but it is both natural and necessary. It is also quite weak: note that the function $x \mapsto I\!\!P(\{X \le x\})$, being monotone in x, is continuous except for at most countably many jumps. The set of continuity points is thus uncountable: 'most' points are continuity points.

Convergence in distribution is (by far) the weakest of the modes of convergence introduced so far: convergence in probability implies convergence in distribution, but not conversely. There is, however, a partial converse (which we shall need in §2.8): if the limit X is *constant* (non-random), convergence in probability and in distribution are equivalent.

Weak Convergence.

If $I\!\!P_n$, $I\!\!P$ are probability measures, we say that

$$I\!\!P_n \to I\!\!P \quad (n \to \infty) \quad \text{weakly}$$

if

$$\int f d I\!\!P_n \to \int f d I\!\!P \quad (n \to \infty) \tag{B.5}$$

for all bounded continuous functions f. This definition is given a full-length book treatment in (Billingsley 1968), and we refer to this for background and details. For ordinary (real-valued) random variables, weak convergence of their probability measures is the same as convergence in distribution of their distribution functions. However, the weak-convergence definition above applies equally, not just to this one-dimensional case, or to the finite-dimensional (vector-valued) setting, but also to infinite-dimensional settings such as arise in convergence of stochastic processes. We shall need such a framework in the passage from discrete- to continuous-time Black-Scholes models.

Appendix C

Stochastic Processes in Discrete Time

C.1 Information and Filtrations

Access to full, accurate, up-to-date information is clearly essential to anyone actively engaged in financial activity or trading. Indeed, information is arguably the most important determinant of success in financial life. Partly for simplicity, partly to reflect the legislation and regulations against insider trading, we shall confine ourselves to the situation where agents take decisions on the basis of information in the public domain, and available to all. We shall further assume that information once known remains known – is not forgotten – and can be accessed in real time.

In reality, of course, matters are more complicated. Information overload is as much of a danger as information scarcity. The ability to retain information, organise it, and access it quickly, is one of the main factors which will discriminate between the abilities of different economic agents to react to changing market conditions. However, we restrict ourselves here to the simplest possible situation and do not differentiate between agents on the basis of their information-processing abilities. Thus as time passes, new information becomes available to all agents, who continually update their information. What we need is a mathematical language to model this information flow, unfolding with time. This is provided by the idea of a *filtration*; we outline below the elements of this theory that we shall need.

The Kolmogorov triples $(\Omega, \mathcal{F}, P)$, and the Kolmogorov conditional expectations $I\!\!E(X|\mathcal{B})$, give us all the machinery we need to handle static situations involving randomness. To handle dynamic situations, involving randomness which unfolds with time, we need further structure.

We may take the initial, or starting, time as $t = 0$. Time may evolve discretely, or continuously. We postpone the continuous case to Chapter 5; in the discrete case, we may suppose time evolves in integer steps, $t = 0, 1, 2, \ldots$ (say, stock-market quotations daily, or tick data by the second). There may be a final time T, or time horizon, or we may have an infinite time horizon (in the context of option pricing, the time horizon T is the expiry time).

We wish to model a situation involving randomness unfolding with time. As above, we suppose, for simplicity, that information is never lost (or forgotten): thus, as time increases we learn more. We recall from Chapter 2 that σ-algebras represent information or knowledge. We thus need a sequence of σ-algebras $\{\mathcal{F}_n : n = 0, 1, 2, \ldots\}$, which are increasing:

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \quad (n = 0, 1, 2, \ldots),$$

with $\mathcal{F}_n$ representing the information, or knowledge, available to us at time n. We shall always suppose all σ-algebras to be *complete* (this can be avoided, and is not always appropriate, but it simplifies matters and suffices for our purposes). Thus $\mathcal{F}_0$ represents the initial information (if

there is none, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial σ-algebra). On the other hand,

$$\mathcal{F}_\infty := \lim_{n \to \infty} \mathcal{F}_n = \sigma \left(\bigcup_n \mathcal{F}_n \right)$$

represents all we ever will know (the 'Doomsday σ-algebra'). Often, $\mathcal{F}_\infty$ will be $\mathcal{F}$ (the σ-algebra from Chapter 2, representing 'knowing everything'). But this will not always be so; see e.g. Williams (1991), §15.8 for an interesting example. Such a family $I\!F := \{\mathcal{F}_n : n = 0, 1, 2, \ldots\}$ is called a *filtration*; a probability space endowed with such a filtration, $\{\Omega, I\!F, \mathcal{F}, P\}$ is called a *stochastic basis* or *filtered probability space*. These definitions are due to P. A. Meyer of Strasbourg; Meyer and the Strasbourg (and more generally, French) school of probabilists have been responsible for the 'general theory of (stochastic) processes', and for much of the progress in stochastic integration, since the 1960s; see e.g. Dellacherie and Meyer (1978), Dellacherie and Meyer (1982), Meyer (1966), Meyer (1976).

For the special case of a finite state space $\Omega = \{\omega_1, \ldots, \omega_n\}$ and a given σ-algebra $\mathcal{F}$ on Ω (which in this case is just an algebra) we can always find a unique finite partition $\mathcal{P} = \{A_1, \ldots, A_l\}$ of Ω, i.e. the sets A_i are disjoint and $\bigcup_{i=1}^l A_i = \Omega$, corresponding to $\mathcal{F}$. A filtration $I\!F$ therefore corresponds to a sequence of finer and finer partitions $\mathcal{P}_n$. At time $t = 0$ the agents only know that some event $\omega \in \Omega$ will happen, at time $T < \infty$ they know which specific event ω^* has happened. During the flow of time the agents learn the specific structure of the (σ-) algebras $\mathcal{F}_n$, which means they learn the corresponding partitions $\mathcal{P}$. Having the information in $\mathcal{F}_n$ revealed is equivalent to knowing in which $A_i^{(n)} \in \mathcal{P}_n$ the event ω^* is. Since the partitions become finer the information on ω^* becomes more detailed with each step.

Unfortunately this nice interpretation breaks down as soon as Ω becomes infinite. It turns out that the concept of filtrations rather than that of partitions is relevant for the more general situations of infinite Ω, infinite T and continuous-time processes.

C.2 Discrete-Parameter Stochastic Processes

The word 'stochastic' (derived from the Greek) is roughly synonymous with 'random'. It is perhaps unfortunate that usage favours 'stochastic process' rather than the simpler 'random process', but as it does, we shall follow it.

We need a framework which can handle dynamic situations, in which time evolves, and in which new information unfolds with time. In particular, we need to be able to speak in terms of 'the information available at time n', or, 'what we know at time n'. Further, we need to be able to increase n – thereby increasing the information available as new information (typically, new price information) comes in, and talk about the information flow over time. One has a clear mental picture of what is meant by this – there is no conceptual difficulty. However, what is needed is a precise mathematical construct, which can be conveniently manipulated - perhaps in quite complicated ways – and yet which bears the above heuristic meaning. Now 'information' is not only an ordinary word, but even a technical term in mathematics – many books have been written on the subject of information theory. However, information theory in this sense is not what we need: for us, the emphasis is on the flow of information, and how to model and describe it. With this by way of motivation, we proceed to give some of the necessary definitions.

A *stochastic process* $X = \{X_n : n \in I\}$ is a family of random variables, defined on some common probability space, indexed by an index-set I. Usually (always in this book), I represents time (sometimes I represents space, and one calls X a spatial process). Here, $I = \{0, 1, 2, \ldots, T\}$ (finite horizon) or $I = \{0, 1, 2, \ldots\}$ (infinite horizon). The (stochastic) process $X = (X_n)_{n=0}^\infty$ is said to be *adapted* to the filtration $I\!F = (\mathcal{F}_n)_{n=0}^\infty$ if

$$X_n \text{ is } \mathcal{F}_n - \text{measurable for all } n.$$

So if X is adapted, we will know the value of X_n at time n. If

$$\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$$

we call $(\mathcal{F}_n)$ the natural filtration of X. Thus a process is always adapted to its natural filtration. A typical situation is that

$$\mathcal{F}_n = \sigma(W_0, W_1, \ldots, W_n)$$

is the natural filtration of some process $W = (W_n)$. Then X is adapted to $I\!F = (\mathcal{F}_n)$, i.e. each X_n is $\mathcal{F}_n$- (or $\sigma(W_0, \cdots, W_n)$-) measurable, iff

$$X_n = f_n(W_0, W_1, \ldots, W_n)$$

for some measurable function f_n (non-random) of $n + 1$ variables.

Notation.

For a random variable X on $(\Omega, \mathcal{F}, I\!P)$, $X(\omega)$ is the value X takes on ω (ω represents the randomness). For a stochastic process $X = (X_n)$, it is convenient (e.g., if using suffixes, n_i say) to use X_n, $X(n)$ interchangeably, and we shall feel free to do this. With ω displayed, these become $X_n(\omega)$, $X(n, \omega)$, etc.

The concept of a stochastic process is very general – and so very flexible – but it is too general for useful progress to be made without specifying further structure or further restrictions. There are two main types of stochastic process which are both general enough to be sufficiently flexible to model many commonly encountered situations, and sufficiently specific and structured to have a rich and powerful theory. These two types are Markov processes and martingales. A *Markov process* models a situation in which where one is, is all one needs to know when wishing to predict the future – how one got there provides no further information. Such a 'lack of memory' property, though an idealisation of reality, is very useful for modelling purposes. We shall encounter Markov processes more in continuous time (see Chapter 5) than in discrete time, where usage dictates that they are called Markov chains. For an excellent and accessible recent treatment of Markov chains, see e.g. Norris (1997). *Martingales*, on the other hand (see §3.3 below) model fair gambling games – situations where there may be lots of randomness (or unpredictability), but no tendency to drift one way or another: rather, there is a tendency towards stability, in that the chance influences tend to cancel each other out on average.

C.3 Definition and basic properties of martingales

Excellent accounts of discrete-parameter martingales are Neveu (1975), Williams (1991) and Williams (2001) to which we refer the reader for detailed discussions. We will summarise what we need to use martingales for modelling in finance.

Definition C.3.1. *A process $X = (X_n)$ is called a martingale relative to $(\{\mathcal{F}_n\}, I\!P)$ if*

(i) X is adapted (to $\{\mathcal{F}_n\}$);
(ii) $I\!E\,|X_n| < \infty$ for all n;
(iii) $I\!E[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ $I\!P - a.s.$ $(n \geq 1)$.

X is a supermartingale if in place of (iii)

$$I\!E[X_n|\mathcal{F}_{n-1}] \leq X_{n-1} \quad I\!P - a.s. \quad (n \geq 1);$$

X is a submartingale if in place of (iii)

$$I\!E[X_n|\mathcal{F}_{n-1}] \geq X_{n-1} \quad I\!P - a.s. \quad (n \geq 1).$$

Martingales have a useful interpretation in terms of dynamic games: a martingale is 'constant on average', and models a fair game; a supermartingale is 'decreasing on average', and models an unfavourable game; a submartingale is 'increasing on average', and models a favourable game.

Note.

1. Martingales have many connections with harmonic functions in probabilistic potential theory. The terminology in the inequalities above comes from this: supermartingales correspond to superharmonic functions, submartingales to subharmonic functions.

2. X is a submartingale (supermartingale) if and only if $-X$ is a supermartingale (submartingale); X is a martingale if and only if it is both a submartingale and a supermartingale.

3. (X_n) is a martingale if and only if $(X_n - X_0)$ is a martingale. So we may without loss of generality take $X_0 = 0$ when convenient.

4. If X is a martingale, then for $m < n$ using the iterated conditional expectation and the martingale property repeatedly (all equalities are in the a.s.-sense)

$$
\begin{aligned}
I\!\!E[X_n|\mathcal{F}_m] &= I\!\!E[I\!\!E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m] = I\!\!E[X_{n-1}|\mathcal{F}_m] \\
&= \ldots = I\!\!E[X_m|\mathcal{F}_m] = X_m,
\end{aligned}
$$

and similarly for submartingales, supermartingales.

From the *Oxford English Dictionary: martingale* (etymology unknown)

1. 1589. An article of harness, to control a horse's head.
2. Naut. A rope for guying down the jib-boom to the dolphin-striker.
3. A system of gambling which consists in doubling the stake when losing in order to recoup oneself (1815).
Thackeray: 'You have not played as yet? Do not do so; above all avoid a martingale if you do.'

Gambling games have been studied since time immemorial – indeed, the Pascal-Fermat correspondence of 1654 which started the subject was on a problem (de Méré's problem) related to gambling. The doubling strategy above has been known at least since 1815.

The term 'martingale' in our sense is due to J. Ville (1939). Martingales were studied by Paul Lévy (1886–1971) from 1934 on (see obituary (Loêve 1973)) and by J.L. Doob (1910–) from 1940 on. The first systematic exposition was (Doob 1953). This classic book, though hard going, is still a valuable source of information.

Examples.

1. Mean zero random walk: $S_n = \sum X_i$, with X_i independent with $I\!\!E(X_i) = 0$ is a martingale (submartingales: positive mean; supermartingale: negative mean).
2. Stock prices: $S_n = S_0\zeta_1 \cdots \zeta_n$ with ζ_i independent positive r.vs with existing first moment.
3. Accumulating data about a random variable (Williams (1991), pp. 96, 166–167). If $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, I\!\!P)$, $M_n := I\!\!E(\xi|\mathcal{F}_n)$ (so M_n represents our best estimate of ξ based on knowledge at time n), then using iterated conditional expectations

$$
I\!\!E[M_n|\mathcal{F}_{n-1}] = I\!\!E[I\!\!E(\xi|\mathcal{F}_n)|\mathcal{F}_{n-1}] = I\!\!E[\xi|\mathcal{F}_{n-1}] = M_{n-1},
$$

so (M_n) is a martingale. One has the convergence

$$
M_n \to M_\infty := I\!\!E[\xi|\mathcal{F}_\infty] \quad a.s. \text{ and in } \mathcal{L}^1.
$$

C.4 Martingale Transforms

Now think of a gambling game, or series of speculative investments, in discrete time. There is no play at time 0; there are plays at times $n = 1, 2, \ldots$, and

$$
\Delta X_n := X_n - X_{n-1}
$$

represents our net winnings per unit stake at play n. Thus if X_n is a martingale, the game is 'fair on average'.

Call a process $C = (C_n)_{n=1}^\infty$ *predictable* if C_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$. Think of C_n as your stake on play n (C_0 is not defined, as there is no play at time 0). Predictability says that you have to decide how much to stake on play n based on the history *before* time n (i.e., up to and including play $n-1$). Your winnings on game n are $C_n \Delta X_n = C_n(X_n - X_{n-1})$. Your total (net) winnings up to time n are

$$Y_n = \sum_{k=1}^n C_k \Delta X_k = \sum_{k=1}^n C_k(X_k - X_{k-1}).$$

We write

$$Y = C \bullet X, \quad Y_n = (C \bullet X)_n, \quad \Delta Y_n = C_n \Delta X_n$$

($(C \bullet X)_0 = 0$ as $\sum_{k=1}^0$ is empty), and call $C \bullet X$ the *martingale transform* of X by C.

Theorem C.4.1. *(i) If C is a bounded non-negative predictable process and X is a supermartingale, $C \bullet X$ is a supermartingale null at zero.*
(ii) If C is bounded and predictable and X is a martingale, $C \bullet X$ is a martingale null at zero.

Proof. $Y = C \bullet X$ is integrable, since C is bounded and X integrable. Now

$$\mathbb{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}[C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}]$$

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

(as C_n is bounded, so integrable, and $\mathcal{F}_{n-1}$-measurable, so can be taken out)

$$\leq 0$$

in case (i), as $C \geq 0$ and X is a supermartingale,

$$= 0$$

in case (ii), as X is a martingale. ∎

Interpretation. You can't beat the system! In the martingale case, predictability of C means we can't foresee the future (which is realistic and fair). So we expect to gain nothing – as we should.

Note.

1. Martingale transforms were introduced and studied by Burkholder (1966). For a textbook account, see e.g. Neveu (1975), VIII.4.
2. Martingale transforms are the discrete analogues of stochastic integrals. They dominate the mathematical theory of finance in discrete time, just as stochastic integrals dominate the theory in continuous time.

Lemma C.4.1 (Martingale Transform Lemma). *An adapted sequence of real integrable random variables (X_n) is a martingale iff for any bounded predictable sequence (C_n),*

$$\mathbb{E}\left(\sum_{k=1}^n C_k \Delta X_k\right) = 0 \quad (n = 1, 2, \ldots).$$

Proof. If (X_n) is a martingale, Y defined by $Y_0 = 0$,

$$Y_n = \sum_{k=1}^n C_k \Delta X_k \quad (n \geq 1)$$

is the martingale transform $C \bullet X$, so is a martingale. Now $I\!E(Y_1) = I\!E(C_1 I\!E(X_1 - X_0)) = 0$ and we see by induction that

$$I\!E(Y_{n+1}) = I\!E(C_{n+1}(X_{n+1} - X_n)) + I\!E(Y_n) = 0.$$

Conversely, if the condition of the proposition holds, choose j, and for any $\mathcal{F}_j$-measurable set A write $C_n = 0$ for $n \neq j + 1$, $C_{j+1} = \mathbf{1}_A$. Then (C_n) is predictable, so the condition of the proposition, $I\!E(\sum_{k=1}^{n} C_k \Delta X_k) = 0$, becomes

$$I\!E[\mathbf{1}_A(X_{j+1} - X_j)] = 0.$$

Since this holds for every set $A \in \mathcal{F}_j$, the definition of conditional expectation gives

$$I\!E(X_{j+1}|\mathcal{F}_j) = X_j.$$

Since this holds for every j, (X_n) is a martingale. ∎

Remark C.4.1. *The proof above is a good example of the value of Kolmogorov's definition of conditional expectation – which reveals itself, not in immediate transparency, but in its ease of handling in proofs. We shall see in Chapter 4 the financial significance of martingale transforms $H \bullet M$.*

Bibliography

Allingham, M., 1991, *Arbitrage. Elements of financial economics.* (MacMillan).

Barndorff-Nielsen, O.E., 1998, Processes of normal inverse Gaussian type, *Finance and Stochastics* 2, 41–68.

Billingsley, P., 1968, *Convergence of probability measures.* (Wiley, New York).

Björk, T., 1995, Arbitrage theory in continuous time, Notes from Ascona meeting.

Björk, T., 1997, Interest rate theory, in *Financial Mathematics*, ed. by W.J. Runggaldier Lecture Notes in Mathematics pp. 53–122. Springer, Berlin New York London.

Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 72, 637–659.

Brown, R.H., and S.M. Schaefer, 1995, Interest rate volatility and the shape of the term structure, in S.D. Howison, F.P. Kelly, and P. Wilmott, eds.: *Mathematical models in finance* (Chapman & Hall,).

Burkholder, D.L., 1966, Martingale transforms, *Ann. Math. Statist.* 37, 1494–1504.

Burkill, J.C., and H. Burkill, 1970, *A second course in mathematical analysis.* (Cambridge University Press).

Cochrane, J.H., 2001, *Asset pricing.* (Princeton University Press).

Cox, D.R., and H.D. Miller, 1972, *The theory of stochastic processes.* (Chapman and Hall, London and New York) First published 1965 by Methuen & Co Ltd.

Cox, J.C., and S.A. Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145–166.

Cox, J.C., S. A. Ross, and M. Rubinstein, 1979, Option pricing: a simplified approach, *J. Financial Economics* 7, 229–263.

Cox, J.C., and M. Rubinstein, 1985, *Options markets.* (Prentice-Hall).

Davis, M.H.A., 1994, A general option pricing formula, Preprint, Imperial College.

Dellacherie, C., and P.-A. Meyer, 1978, *Probabilities and potential* vol. A. (Hermann, Paris).

Dellacherie, C., and P.-A. Meyer, 1982, *Probabilities and potential* vol. B. (North Holland, Amsterdam New York).

Doob, J. L., 1953, *Stochastic processes.* (Wiley).

Dothan, M. U., 1990, *Prices in financial markets.* (Oxford University Press).

Dudley, R.M., 1989, *Real Analysis and Probability.* (Wadsworth, Pacific Grove).

Duffie, D., 1992, *Dynamic Asset Pricing Theory.* (Princton University Press).

Duffie, D., and R. Kan, 1995, Multi-factor term structure models, in S.D. Howison, F.P. Kelly, and P. Wilmott, eds.: *Mathematical models in finance* (Chapman & Hall,).

Durrett, R., 1996, *Probability: Theory and Examples.* (Duxbury Press at Wadsworth Publishing Company) 2nd edn.

Durrett, R., 1999, *Essentials of stochastic processes.* (Springer).

Eberlein, E., and U. Keller, 1995, Hyperbolic distributions in finance, *Bernoulli* 1, 281–299.

Edwards, F.R., and C.W. Ma, 1992, *Futures and Options.* (McGraw-Hill, New York).

El Karoui, N., R. Myneni, and R. Viswanathan, 1992, Arbitrage pricing and hedging of interest rate claims with state variables: I. Theory, Université de Paris VI and Stanford University.

Grimmett, G.R., and D.J.A. Welsh, 1986, *Probability: An introduction.* (Oxford University Press).

Harrison, J.M., 1985, *Brownian motion and stochastic flow systems.* (John Wiley and Sons, New York).

Harrison, J.M., and D. M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *J. Econ. Th.* 20, 381–408.

Harrison, J.M., and S.R. Pliska, 1981, Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Processes and their Applications* 11, 215–260.

Heath, D., R. Jarrow, and A. Morton, 1992, Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuation, *Econometrica* 60, 77–105.

Ikeda, N., S. Watanabe, Fukushima M., and H. Kunita (eds.), 1996, *Itô stochastic calculus and probability theory.* (Springer, Tokyo Berlin New York) Festschrift for Kiyosi Itô's eightieth birthday, 1995.

Ince, E.L., 1944, *Ordinary differential equations.* (Dover New York).

Jacod, J., and P. Protter, 2000, *Probability essentials.* (Springer).

Jameson, R. (ed.), 1995, *Derivative credit risk.* (Risk Publications London).

Jarrow, R.A., and S.M. Turnbull, 2000, *Derivative Securities.* (South-Western College Publishing, Cincinnati) 2nd edn. 1st ed. 1996.

Jeans, Sir James, 1925, *The mathematical theory of electricity and magnetism.* (Cambridge University Press) 5th. edn.

Karatzas, I., and S. Shreve, 1991, *Brownian motion and stochastic calculus.* (Springer, New York) 2nd edn.

Kolb, R.W., 1991, *Understanding Futures Markets.* (Kolb Publishing, Miami) 3rd edn.

Kolmogorov, A.N., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung.* (Springer) English translation: *Foundations of probability theory*, Chelsea, New York, (1965).

Lebesgue, H., 1902, Intégrale, longueur, aire, *Annali di Mat.* 7, 231–259.

Loêve, M., 1973, Paul Lévy (1886-1971), obituary, *Annals of Probability* 1, 1–18.

Madan, D., 1998, Default risk, in D.J. Hand, and S.D. Jacka, eds.: *Statistics in finance* (Arnold, London,).

Merton, R.C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141–183.

Merton, R.C., 1990, *Continuous-time finance.* (Blackwell, Oxford).

Meyer, P.-A., 1966, *Probability and potential.* (Blaisdell, Waltham, Mass.).

Meyer, P.-A., 1976, Un cours sur les intégrales stochastiques, in *Séminaire de Probabilités X* no. 511 in Lecture Notes in Mathematics pp. 245–400. Springer, Berlin Heidelberg New York.

Musiela, M., and M. Rutkowski, 1997, *Martingale methods in financial modelling* vol. 36 of *Applications of Mathematics: Stochastic Modelling and Applied Probability.* (Springer, New York).

Neveu, J., 1975, *Discrete-parameter martingales.* (North-Holland).

Norris, J.R., 1997, *Markov chains.* (Cambridge University Press).

Protter, P., 2004, *Stochastic Integration and Differential Equations.* (Springer, New York) 2nd edn. 1rst edition, 1992.

Rennocks, John, 1997, Hedging can only defer currency volatility impact for British Steel, Financial Times 08, Letter to the editor.

Resnick, S., 2001, *A probability path.* (Birkhäuser) 2nd printing.

Revuz, D., and M. Yor, 1991, *Continuous martingales and Brownian motion.* (Springer, New York).

Rockafellar, R.T., 1970, *Convex Analysis.* (Princton University Press, Princton NJ).

Rogers, L.C.G., and D. Williams, 1994, *Diffusions, Markov processes and martingales, Volume 1: Foundation.* (Wiley) 2nd edn. 1st ed. D. Williams, 1970.

Rosenthal, J.S., 2000, *A first look at rigorous probability theory.* (World Scientific).

Ross, S.M., 1997, *Probability models.* (Academic Press) 6th edn.

Rydberg, T.H., 1996, Generalized hyperbolic diffusions with applications towards finance, Research Report 342, Department of Theoretical Statistics, Institute of Mathematics, University of Århus University.

Rydberg, T.H., 1997, The normal inverse Gaussian Lévy process: Simulation and approximation, Research Report, Department of Theoretical Statistics, Institute of Mathematics, University of Århus University.

Samuelson, P.A., 1965, Rational theory of warrant pricing, *Industrial Management Review* 6, 13–39.

Schachermayer, W., 2000, Introduction to the mathematics of financial markets, to appear as Springer Lecture Notes.

Shephard, N., 1996, Statistical aspects of ARCH and stochastic volatility, in D.R. Cox, D.V. Hinkley, and O.E. Barndorff-Nielsen, eds.: *Time Series Models - in econometrics, finance and other fields* (Chapman & Hall,).

Snell, J. L., 1952, Applications of martingale systems theorems, *Trans. Amer. Math. Soc.* 73, 293–312.

Taqqu, M.S., and W. Willinger, 1987, The analysis of finite security markets using martingales, *Adv. Appl. Prob.* 19, 1–25.

Williams, D., 1991, *Probability with martingales.* (Cambridge University Press).

Williams, D., 2001, *Weighing the odds.* (Cambridge University Press).

Yor, M., 1978, Sous-espaces denses dans L^1 et H^1 et représentation des martingales, in *Séminaire de Probabilités, XII* no. 649 in Lecture Notes in Mathematics pp. 265–309. Springer.

Zhang, P.G., 1997, *Exotic Options.* (World Scientific, Singapore).