Chapter 1 Modeling Tools for Financial Options

1.1 Options

What do we mean by option? An option is the right (but not the obligation) to buy or sell one unit of a risky asset at a prespecified fixed price within a specified period. An option is a financial instrument that allows —amongst other things— to make a bet on rising or falling values of an underlying asset. The **underlying** asset typically is a stock, or a parcel of shares of a company. Other examples of underlyings include stock indices (as the Dow Jones Industrial Average), currencies, or commodities. Since the value of an option depends on the value of the underlying asset, options and other related financial instruments are called *derivatives* ($\longrightarrow$ Appendix A2). An option is a contract between two parties about trading the asset at a certain future time. One party is the *writer*, often a bank, who fixes the terms of the option contract and sells the option. The other party is the *holder*, who purchases the option, paying the market price, which is called *premium*. How to calculate a fair value of the premium is a central theme of this book. The holder of the option must decide what to do with the rights the option contract grants. The decision will depend on the market situation, and on the type of option. There are numerous different types of options, which are not all of interest to this book. In Chapter 1 we concentrate on standard options, also known as *vanilla options*. This Section 1.1 introduces important terms.

Options have a limited life time. The *maturity date* T fixes the time horizon. At this date the rights of the holder expire, and for later times ($t > T$) the option is worthless. There are two basic types of option: The **call** option gives the holder the right to *buy* the underlying for an agreed price K by the date T. The **put** option gives the holder the right to *sell* the underlying for the price K by the date T. The previously agreed price K of the contract is called **strike** or **exercise price**[1]. It is important to note that the holder is not obligated to *exercise* —that is, to buy or sell the underlying according to the terms of the contract. The holder may wish to close his position by selling the option. In summary, at time t the holder of the option can choose to

[1] The price K as well as other prices are meant as the price of one unit of an asset, say, in $.

R.U. Seydel, *Tools for Computational Finance*, Universitext,
DOI 10.1007/978-1-4471-2993-6_1,

- sell the option at its current market price on some options exchange (at $t < T$),
- retain the option and do nothing,
- exercise the option ($t \leq T$), or
- let the option expire worthless ($t \geq T$).

In contrast, the writer of the option has the obligation to deliver or buy the underlying for the strike price K, in case the holder chooses to exercise. The risk situation of the writer differs strongly from that of the holder. The writer receives the premium when he issues the option and somebody buys it. This up-front premium payment compensates for the writer's potential liabilities in the future. The asymmetry between writing and owning options is evident. This book mostly takes the standpoint of the holder (long position in the option).

Not every option can be exercised at any time $t \leq T$. For **European options**, exercise is only permitted at expiration T. **American options** can be exercised at any time up to and including the expiration date. For options the labels American or European have no geographical meaning; both types are traded in each continent. Options on stocks are mostly American style.

The value of the option will be denoted by V. The value V depends on the price per share of the underlying, which is denoted S. This letter S symbolizes stocks, which are the most prominent examples of underlying assets. The variation of the asset price S with time t is expressed by S_t or $S(t)$. The value of the option also depends on the remaining time to expiry $T - t$. That is, V depends on time t. The dependence of V on S and t is written $V(S, t)$. As we shall see later, it is not easy to define and to calculate the fair value V of an option for $t < T$. But it is an easy task to determine the terminal value of V at expiration time $t = T$. In what follows, we shall discuss this topic, and start with European options as seen with the eyes of the holder.

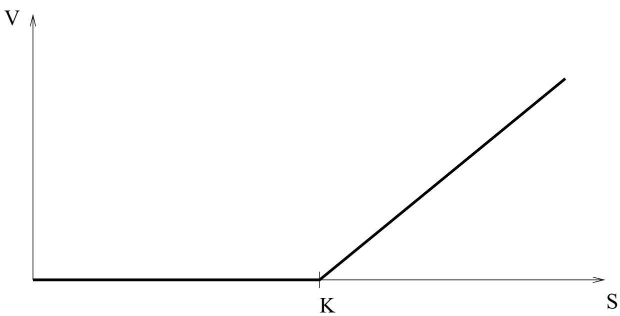

Fig. 1.1. Intrinsic value of a call with exercise price K (payoff function)

The Payoff Function

At time $t = T$, the holder of a European call option will check the current price $S = S_T$ of the underlying asset. The holder has two alternatives to acquire the underlying asset: either buying the asset on the spot market (costs S), or buying the asset by exercising the call option (costs K). For a rational investor, the decision is easy: the costs are to be minimal. The holder will exercise the call if and only if $S > K$. For then the holder can immediately sell the asset for the spot price S and makes a gain of $S - K$ per share. In this situation the value of the option is $V = S - K$. (This reasoning ignores transaction costs.) In case $S < K$ the holder will not exercise, since then the asset can be purchased on the market for the cheaper price S. In this case the option is worthless, $V = 0$. In summary, the value $V(S,T)$ of a call option at expiration date T is given by

$$V(S_T,T) = \begin{cases} 0 & \text{in case } S_T \le K \text{ (option expires worthless)} \\ S_T - K & \text{in case } S_T > K \text{ (option is exercised)} \end{cases}$$

Hence

$$V(S_T,T) = \max\{S_T - K,\, 0\}\,.$$

Considered for all possible prices $S_t > 0$, $\max\{S_t - K, 0\}$ is a function of S_t, in general for $0 \le t \le T$.[2] This **payoff function** is shown in Figure 1.1. Using the notation $f^+ := \max\{f,\, 0\}$, this payoff can be written in the compact form $(S_t - K)^+$. Accordingly, the value $V(S_T,T)$ of a call at maturity date T is

$$V(S_T,T) = (S_T - K)^+\,. \tag{1.1C}$$

For a European put, exercising only makes sense in case $S < K$. The payoff $V(S,T)$ of a put at expiration time T is

$$V(S_T,T) = \begin{cases} K - S_T & \text{in case } S_T < K \text{ (option is exercised)} \\ 0 & \text{in case } S_T \ge K \text{ (option is worthless)} \end{cases}$$

Hence

$$V(S_T,T) = \max\{K - S_T, 0\}\,,$$

or

$$V(S_T,T) = (K - S_T)^+\,, \tag{1.1P}$$

compare Figure 1.2.

---

[2] In this chapter, the payoff evaluated at t only depends on the current value S_t. Payoffs that depend on the *entire path* S_t for all $0 \le t \le T$ occur for exotic options, see Chapter 6.

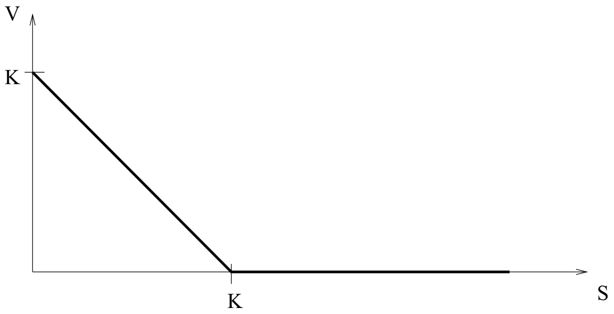

Fig. 1.2. Intrinsic value of a put with exercise price K (payoff function)

The curves in the payoff diagrams of Figures 1.1 and 1.2 show the option values from the perspective of the holder. The profit is not shown. For an illustration of the profit, the initial costs for buying the option at $t = t_0$ must be subtracted. The initial costs basically consist of the premium and the transaction costs. Since both are paid upfront, they are multiplied by $\mathrm{e}^{r(T-t_0)}$ to take account of the time value; r is the continuously compounded interest rate. Subtracting the costs leads to shifting down the curves in Figures 1.1 and 1.2. The resulting *profit diagram* shows a negative profit for some range of S-values, which of course means a loss (see Figure 1.3).

The payoff function for an American call is $(S_t - K)^+$ and for an American put $(K - S_t)^+$ for any $t \leq T$. The Figures 1.1 and 1.2 as well as the equations (1.1C), (1.1P) remain valid for American type options.

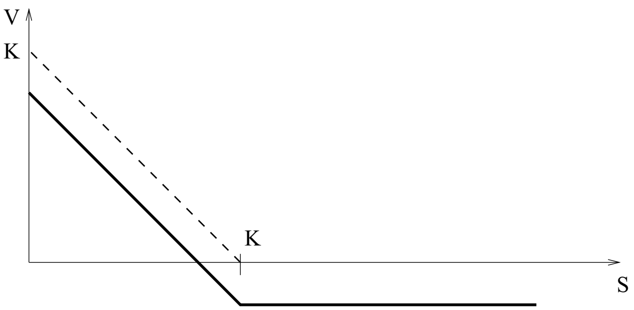

Fig. 1.3. Profit diagram of a put

The payoff diagrams of Figures 1.1, 1.2 and the corresponding profit diagrams show that a potential loss for the purchaser of an option (long position) is limited by the initial costs, no matter how bad things get. The situation for the writer (short position) is reverse. For him the payoff curves of Figures 1.1, 1.2 as well as the profit curves must be reflected on the S-axis. The writer's profit or loss is the reverse of that of the holder. Multiplying the payoff of a call in Figure 1.1 by (-1) illustrates the potentially unlimited risk of a short call. Hence the writer of a call must carefully design a strategy to compensate for his risks. We will come back to this issue in Section 1.5.

A Priori Bounds

No matter what the terms of a specific option are and no matter how the market behaves, the values V of the options satisfy certain bounds. These bounds are known a priori. For example, the value $V(S,t)$ of an American option can never fall below the payoff, for all S and all t. These bounds follow from the *no-arbitrage principle* ($\longrightarrow$ Appendices A2, A3).

To illustrate the strength of no-arbitrage arguments, we assume for an American put that its value $V_{\mathrm{P}}^{\mathrm{Am}}$ is below the payoff. $V<0$ contradicts the definition of the option. Hence $V\geq 0$, and S and V would be in the triangle seen in Figure 1.2. That is, $S<K$ and $0\leq V<K-S$. This scenario would allow an arbitrage strategy as follows: Borrow the cash amount of $S+V$, and buy both the underlying and the put. Then immediately exercise the put, selling the underlying for the strike price K. The profit of this arbitrage strategy is $K-S-V>0$. This is in conflict with the no-arbitrage principle. Hence the assumption that the value of an American put is below the payoff must be wrong. We conclude for the put

$$V_{\mathrm{P}}^{\mathrm{Am}}(S,t)\geq (K-S)^{+}\quad \text{for all } S,t\,.$$

Similarly, for the call

$$V_{\mathrm{C}}^{\mathrm{Am}}(S,t)\geq (S-K)^{+}\quad \text{for all } S,t\,.$$

(The meaning of the notations $V_{\mathrm{C}}^{\mathrm{Am}}$, $V_{\mathrm{P}}^{\mathrm{Am}}$, $V_{\mathrm{C}}^{\mathrm{Eur}}$, $V_{\mathrm{P}}^{\mathrm{Eur}}$ is evident.)

Other bounds are listed in Appendix D1. For example, a European put on an asset that pays no dividends until T may also take values below the payoff, but is always above the lower bound $K\mathrm{e}^{-r(T-t)}-S$. The value of an American option should never be smaller than that of a European option because the American type includes the European type exercise at $t=T$ and in addition *early exercise* for $t<T$. That is

$$V^{\mathrm{Am}}\geq V^{\mathrm{Eur}}$$

as long as all other terms of the contract are identical. When no dividends are paid until T, the values of put and call for European options are related by the *put-call parity*

$$S+V_{\mathrm{P}}^{\mathrm{Eur}}-V_{\mathrm{C}}^{\mathrm{Eur}}=K\mathrm{e}^{-r(T-t)}\,,$$

which can be shown by applying arguments of arbitrage ($\longrightarrow$ Exercise 1.1).

Options in the Market

The features of the options imply that an investor purchases puts when the price of the underlying is expected to fall, and buys calls when the prices are about to rise. This mechanism inspires speculators. An important application of options is hedging (⟶ Appendix A2).

The value of $V(S,t)$ also depends on other factors. Dependence on the strike K and the maturity T is evident. Market parameters affecting the price are the interest rate r, the **volatility** σ of the price S_t, and dividends in case of a dividend-paying asset. The interest rate r is the risk-free rate, which applies to zero bonds or to other investments that are considered free of risks (⟶ Appendices A1, A2). The important volatility parameter σ can be defined as standard deviation of the fluctuations in S_t, for scaling divided by the square root of the observed time period. The larger the fluctuations, represented by large values of σ, the harder is to predict a future value of the asset. Hence the volatility is a standard measure of risk. The dependence of V on σ is highly sensitive. On occasion we write $V(S,t;T,K,r,\sigma)$ when the focus is on the dependence of V on market parameters.

Time is measured in years. The units of r and σ^2 are per year. Writing $\sigma = 0.2$ means a volatility of 20%, and $r = 0.05$ represents an interest rate of 5%. Table 1.1 summarizes the key notations of option pricing. The notation is standard except for the strike price K, which is sometimes denoted X, or E.

The time period of interest is $t_0 \le t \le T$. One might think of t_0 denoting the date when the option is issued and t as a symbol for "today." But this book mostly sets $t_0 = 0$ in the role of "today," without loss of generality. Then the interval $0 \le t \le T$ represents the remaining life time of the option. The price S_t is a stochastic process, compare Section 1.6. In real markets, the interest rate r and the volatility σ vary with time. To keep the models and the analysis simple, we mostly assume r and σ to be constant on $0 \le t \le T$. Further we suppose that all variables are arbitrarily divisible and consequently can vary continuously —that is, all variables vary in the set $\mathbb{R}$ of real numbers.

Table 1.1. List of important variables

t	current time, $0 \le t \le T$
T	expiration time, date of maturity, terminal time
r	risk-free interest rate, continuously compounded
$S,\ S_t$	spot price, current price per share of stock/asset/underlying
σ	annual volatility
K	strike, exercise price per share
$V(S,t)$	value of an option at time t and underlying price S

The Geometry of Options

As mentioned, our aim is to calculate $V(S,t)$ for fixed values of K,T,r,σ. The values $V(S,t)$ can be interpreted as a surface over the subset

$$S > 0\ ,\ \ 0 \le t \le T$$

of the (S,t)-plane. Figure 1.4 illustrates the character of such a surface for the case of an American put. For the illustration assume $T = 1$. The figure depicts six curves obtained by cutting the *option surface* with the planes $t = 0, 0.2, \ldots, 1.0$. For $t = T$ the payoff function $(K - S)^+$ of Figure 1.2 is clearly visible.

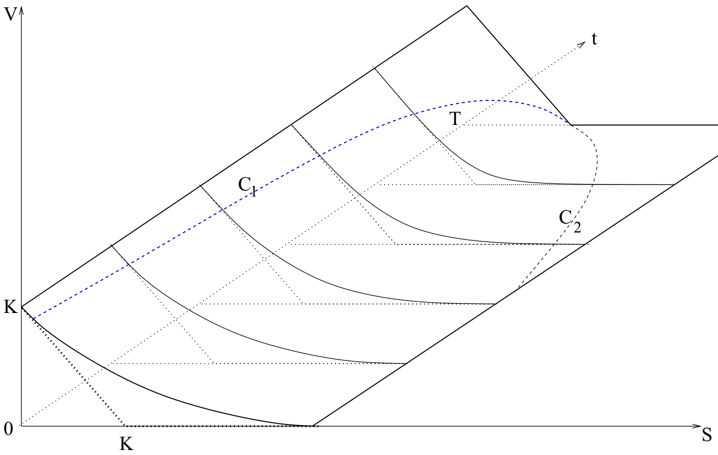

Fig. 1.4. Value $V(S,t)$ of an American put (schematically)

Shifting this payoff curve parallel for all $0 \le t < T$ creates another surface, which consists of the two planar pieces $V = 0$ (for $S \ge K$) and $V = K - S$ (for $S < K$). This *payoff surface* $(K - S)^+$ is a lower bound to the option surface, $V(S,t) \ge (K - S)^+$. Figure 1.4 shows two curves C_1 and C_2 on the option surface. The curve C_1 is the *early-exercise curve*, because on the planar part with $V(S,t) = K - S$ holding the option is not optimal. (This will be explained in Section 4.5.) The curve C_2 has a technical meaning explained below. Within the area limited by these two curves C_1, C_2, the option surface is clearly above the payoff surface, $V(S,t) > (K - S)^+$. Outside that area, both surfaces coincide. This is strict "above" C_1, where $V(S,t) = K - S$, and holds approximately for S beyond C_2, where $V(S,t) \approx 0$ or $V(S,t) < \varepsilon$ for a small value of $\varepsilon > 0$. The location of C_1 and C_2 is not known, these curves are calculated along with the calculation of $V(S,t)$. Of special interest is $V(S,0)$, the value of the option "today." This curve is seen in Figure 1.4 for $t = 0$ as the front edge of the option surface. This front curve may be seen as smoothing the corner in the payoff function. The schematic illustration of Figure 1.4 is completed by a concrete example of a calculated put surface in Figure 1.5. An approximation of the curve C_1 is shown.

The above was explained for an American put. For other options the bounds are different ($\longrightarrow$ Appendix D1). As mentioned before, a European put takes values above the lower bound $K\mathrm{e}^{-r(T-t)} - S$, compare Figure 1.6 and Exercise 1.1b.

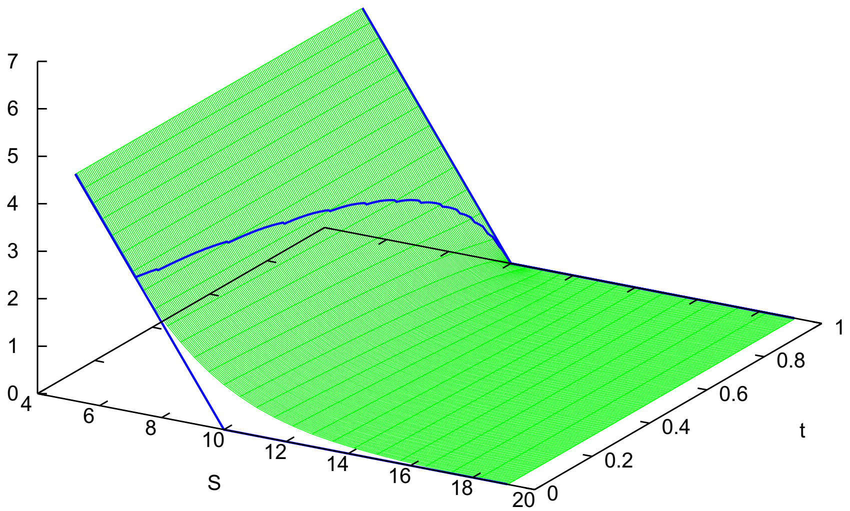

Fig. 1.5. Value $V(S,t)$ of an American put with $r = 0.06$, $\sigma = 0.30$, $K = 10$, $T = 1$

In summary, this Section 1.1 has introduced an option with the following features: it depends on *one* underlying, and its payoff is $(K - S)^+$ or $(S - K)^+$, with S evaluated at the current time instant. This is the standard option called vanilla option. All other options are called *exotic*. To clarify the distinction between vanilla options and exotic options, we hint at ways how an option can be "exotic." For example, an option may depend on a basket of several underlying assets, or the payoff may be different, or the option may be *path-dependent* in that V no longer depends solely on the current (S_t, t) but on the entire path S_t for $0 \leq t \leq T$. To give an example of the latter, we mention an *Asian option*, where the payoff depends on the average value of the asset for all times until expiry. Or for a *barrier option* the value also depends on whether the price S_t hits a prescribed barrier during its life time. We come back to exotic options later in the book.

1.2 Model of the Financial Market

Ultimately it is the market that decides on the value of an option. Above, we have been anticipating "surfaces" $V(S,t)$, pretending a value V for any S, t. In the reality of markets, prices V^{mar} of options are only known for a selection of discrete values of the underlying's prices, times, or parameters. Geometrically, the available data form only relatively few points on the anticipated "surfaces" V. If we try to *calculate* a reasonable value of the option, we need a mathematical model of the market. Mathematical models can serve as approximations and idealizations of the complex reality of the financial world. The

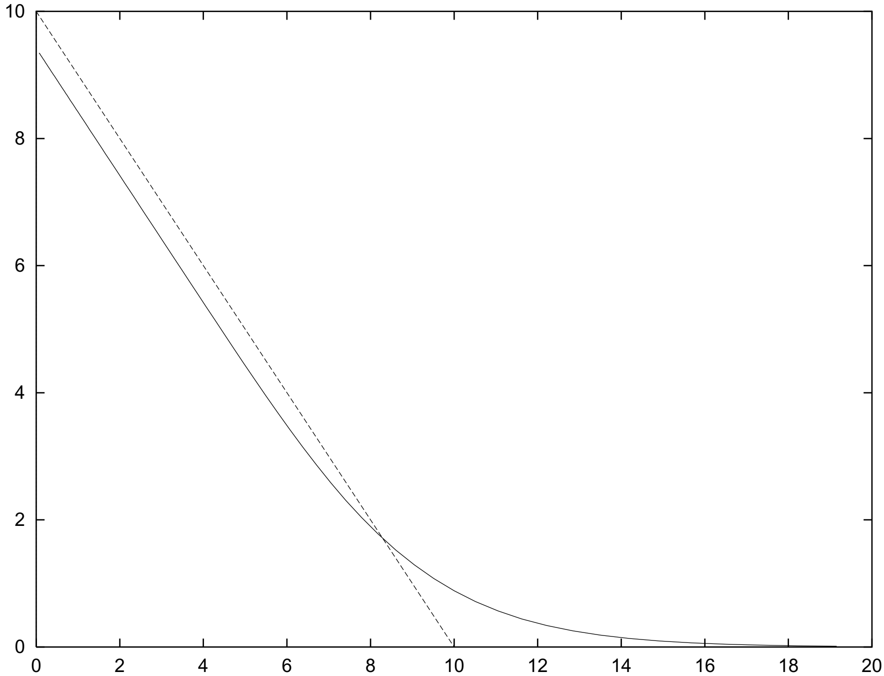

Fig. 1.6. Value of a European put $V(S,0)$ for $T = 1$, $K = 10$, $r = 0.06$, $\sigma = 0.3$. The payoff $V(S,T)$ is drawn with a dashed line. For small values of S the value V approaches its lower bound, here $9.4 - S$.

most prominent example of a model is the model named after the pioneers Black, Merton and Scholes. Their approaches have been both successful and widely accepted. This Section 1.2 introduces some key elements of market models. Based on a chosen mathematical model, the value and the potential of an option is assessed. This includes both the calculation of $V(S,t)$, and an analysis of how sensitive V reacts on changes in S, t, or on variations in the parameters. Of course, the results are subject to the uncertainty of the model.

It is attractive to define the option surfaces $V(S,t)$ on the *half strip* $S > 0$, $0 \leq t \leq T$ as solutions of suitable equations. Then calculating V amounts to solving the equations. In fact, a series of assumptions allows to characterize *value functions* $V(S,t)$ as solutions of certain partial differential equations or partial differential inequalities. The model of Black, Merton and Scholes is represented by the famous Black–Scholes equation, which was suggested in 1973.

Definition 1.1 (Black–Scholes equation)

$$\boxed{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0} \tag{1.2}$$

Equation (1.2) is a partial differential equation (PDE) for the value function $V(S,t)$ of options. This equation may serve as symbol of the classical market model. But what are the assumptions leading to the Black–Scholes equation?

Assumptions 1.2 (Black–Merton–Scholes model of the market)
(a) *There are no arbitrage opportunities.*
(b) *The market is frictionless.*
This means that there are no transaction costs (fees or taxes), the interest rates for borrowing and lending money are equal, all parties have immediate access to any information, and all securities and credits are available at any time and in any size.[3] Consequently, all variables are perfectly divisible —that is, can take any real number. Further, individual trading will not influence the price.
(c) *The asset price follows a geometric Brownian motion.*
(This stochastic motion will be discussed in Sections 1.6–1.8.)
(d) r and σ are constant for $0 \le t \le T$. No dividends are paid in that time period. The option is European.

These are the assumptions that lead to the Black–Scholes equation (1.2). The assumptions are rather strong, in particular, the volatility σ being constant. Some of the assumptions can be weakened. We come to more complex models later in the text. For brevity, we call the restricted model represented by Assumptions 1.2 Black–Scholes model, because Merton has also extended it to include jumps, which are ruled out by (c). A derivation of the Black–Scholes partial differential equation (1.2) is given in Appendix A4. Admitting all real numbers t within the interval $0 \le t \le T$ leads to characterize the model as *continuous-time model.* In view of allowing also arbitrary values of $S > 0$, $V > 0$, we speak of a continuous model.

A value function $V(S,t)$ is not fully defined by merely requesting that it solves (1.2) for all S and t out of the half strip. In addition to solving this PDE, the function $V(S,t)$ must satisfy a **terminal condition**. The terminal condition for $t = T$ is

$$V(S,T) = \Psi(S)\,,$$

where Ψ denotes the payoff function (1.1C) or (1.1P), depending on the type of option. This terminal condition is no artificial requirement. It is a prime statement and naturally represents the definition of an option. In theory, (1.2)

[3] In particular, this holds for trading the underlying.

with $V(S,T) = \Psi(S)$ is a Cauchy problem and completes one possibility of defining a value function $V(S,t)$.

For computational purposes, the full half strip with $S > 0$ is typically truncated, say, to $S_{\min} \le S \le S_{\max}$. Then **boundary conditions** for $S_{\min}$ and $S_{\max}$ are needed in addition. Sometimes they are given by the financial terms of the option, for example, for barrier options. But often boundary conditions are secondary and artificial, and not immediately provided by the financial construction. Rather, boundary conditions are required to make a solution of the partial differential equation meaningful. In Chapter 4 we will come back to the Black–Scholes equation and to boundary conditions.

For (1.2) an analytic solution is known [equation (A4.10) in Appendix A4]. Note that the partial differential equation (1.2) is linear in the value function V.[4] The partial differential equation is no longer linear when Assumptions 1.2(b) are relaxed. For example, for considering trading intervals Δt and transaction costs as k per unit, one could add the nonlinear term

$$-\sqrt{\frac{2}{\pi}}\,\frac{k\sigma S^2}{\sqrt{\Delta t}}\,\left|\frac{\partial^2 V}{\partial S^2}\right|$$

to (1.2), see [WiDH96], and Section 7.1. Also finite liquidity (feedback of trading to the price of the underlying) leads to nonlinear terms in the PDE. In the general case, closed-form solutions do not exist, and a solution is calculated numerically, especially for American options. For the American-style option a further nonlinearity stems from the early-exercise feature ($\longrightarrow$ Chapter 4). For solving (1.2) numerically, a variant with dimensionless variables can be used ($\longrightarrow$ Exercise 1.2).

Of course, the calculated value V of an option depends on the chosen market model. Writing $V(S,t;\,T,K,r,\sigma)$ suggests a focus on the Black–Scholes equation. This could be made definite by writing V^{BS}, for example. Other market models may involve more parameters. Then, in general, the corresponding value of the value function V is different from V^{BS}. Since we mostly stick to the market model of Assumptions 1.2, we drop the superscript. All our prices V are model prices, not market prices. For the relation between our model prices V and market prices V^{mar}, see Section 1.10.

Based on the chosen mathematical model, a **sensitivity analysis** is possible. We ask, for example, how does the price V change to a value $V + \mathrm{d}V$, when the price S of the underlying changes to $S + \mathrm{d}S$? Similarly, what is the effect of a change $\mathrm{d}\sigma$ in the parameter σ? When the value function $V(S,t;\ldots)$ is smooth, the Taylor expansion

$$\mathrm{d}S = \frac{\partial V}{\partial S}\mathrm{d}S + \frac{\partial V}{\partial t}\mathrm{d}t + \frac{\partial V}{\partial \sigma}\mathrm{d}\sigma + \frac{\partial V}{\partial r}\mathrm{d}r + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}(\mathrm{d}S)^2 + \ldots \quad (1.3)$$

[4] The function V is not linear in S or t. Also the payoff is nonlinear; the vanilla functions $\Psi(S) = (K-S)^+$ and $\Psi(S) = (S-K)^+$ are convex.

suggest an answer. The proper partial derivative of V is an amplification factor. For small enough $\mathrm{d}t$ it provides a first-order guess on how sensitive V may react to changes in the corresponding variable or parameter. In the finance context, these partial derivatives of V are called "Greeks." For example, "delta" is the name for

$$\Delta := \frac{\partial V}{\partial S}\,.$$

The second-order derivative "gamma" $\frac{\partial^2 V}{\partial S^2}$ is important too, and is included in the list of first-order terms in (1.3) by reasons that will become clear in Sections 1.6 and 1.8. Several of these *sensitivity parameters* or *hedge parameters* need to be approximated as well.

At this point, a word on the notation is appropriate. The symbol S for the asset price is used in different roles: First it comes without subscript in the role of an independent real variable $S > 0$ on which the value function $V(S,t)$ depends, say as solution of the partial differential equation (1.2). Second it is used as S_t with subscript t to emphasize its random character as stochastic process. When the subscript t is omitted, the current role of S becomes clear from the context.

1.3 Numerical Methods

Applying numerical methods is inevitable in all fields of technology including financial engineering. Often the important role of numerical algorithms is not noticed. For example, an analytic formula at hand [such as the Black–Scholes formula (A4.10)] might suggest that no numerical procedure is needed. But closed-form solutions may include evaluating the logarithm or the computation of the distribution function of the normal distribution. Such elementary tasks are performed using sophisticated numerical algorithms. In pocket calculators one merely presses a button without being aware of the numerics. The robustness of those elementary numerical methods is so reliable and the efficiency so high that underlying algorithms almost appear not to exist. But even for apparently simple tasks the methods are quite demanding ($\longrightarrow$ Exercise 1.3). The methods must be carefully designed because inadequate strategies might produce inaccurate results ($\longrightarrow$ Exercise 1.4).

Spoilt by generally available black-box software and graphics packages we take the support and the success of numerical workhorses for granted. We make use of the numerical tools with great respect but without further comments, and we just assume an education in elementary numerical methods. An introduction to important methods and hints on the literature are given in Appendix C1.

Since financial markets undergo apparently stochastic fluctuations, stochastic approaches provide natural tools to simulate prices. These methods are based on formulating and simulating stochastic differential equations. This leads to Monte Carlo methods (⟶ Chapter 3). In computers, related simulations of options are performed in a deterministic manner. It will be decisive how to simulate randomness (⟶ Chapter 2). Chapters 2 and 3 are devoted to tools for simulation. These methods can be applied easily even in case the Assumptions 1.2 are not satisfied.

More efficient methods will be preferred provided their use can be justified by the validity of the underlying models. For example it may be advisable to solve the partial differential equations of the Black–Scholes type. Then one has to choose among several methods. The most elementary ones are finite-difference methods (⟶ Chapter 4). A somewhat higher flexibility concerning error control is possible with finite-element methods (⟶ Chapter 5). The numerical treatment of exotic options requires a more careful consideration of stability issues (⟶ Chapter 6). The methods based on differential equations will be described in the larger part of this book. And beyond Black and Scholes, even more tools are needed (⟶ Chapter 7).

The various methods are discussed in terms of accuracy and speed. Ultimately the methods must give quick and accurate answers to real-time problems posed in financial markets. Efficiency and reliability are key demands. Internally the numerical methods must deal with diverse problems such as convergence order or stability. So the numerical analyst is concerned in error estimates and error bounds. Technical criteria such as complexity or storage requirements are relevant for the implementation.

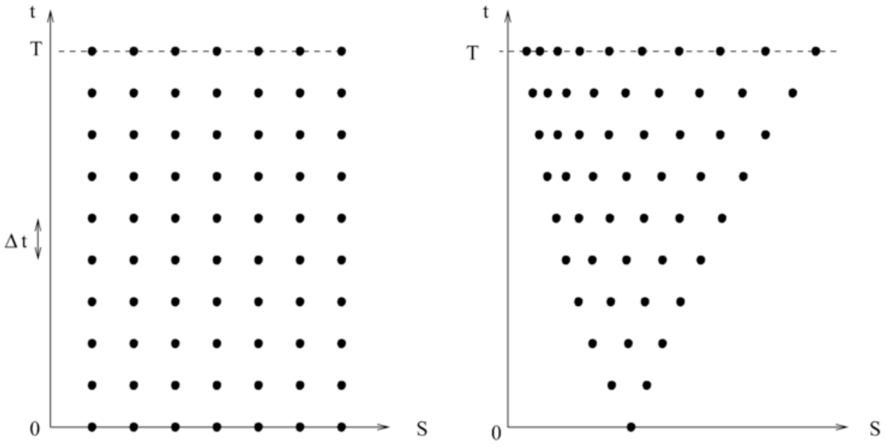

Fig. 1.7. Grid points in the (S, t)-domain

The mathematical formulation benefits from the assumption that all variables take values in the continuum $\mathbb{R}$. This idealization is practical since it avoids initial restrictions of technical nature, and it gives us freedom to impose *artificial* discretizations convenient for the numerical methods. The hypothesis of a continuum applies to the (S,t)-domain of the half strip $0 \leq t \leq T$, $S > 0$, and to the differential equations. In contrast to the hypothesis of a continuum, the financial reality is rather discrete: Neither the price S nor the trading times t can take any real value. The artificial discretization introduced by numerical methods is at least twofold:

1.) The (S,t)-domain is replaced by a **grid** of a finite number of (S,t)-points, illustrated in Figure 1.7.
2.) The differential equations are adapted to the grid and replaced by a finite number of algebraic equations.

The restriction of the differential equations to the grid causes **discretization errors**. The errors depend on the coarseness of the grid. In Figure 1.7, the distance between two consecutive t-values of the grid is denoted Δt.[5] So the errors will depend on Δt and on ΔS. It is one of the aims of numerical algorithms to control the errors. The left-hand figure in Figure 1.7 shows a simple rectangle grid, whereas the right-hand figure shows a tree-type grid as used in Section 1.4. The type of the grid matches the kind of underlying equations. The values of $V(S,t)$ are primarily approximated at the grid points. Intermediate values can be obtained by interpolation.

The continuous model is an idealization of the discrete reality. But the numerical discretization does not reproduce the original discretization. For example, it would be a rare coincidence when Δt represents a day. The derivations that go along with the twofold transition

$$\text{discrete} \quad \longrightarrow \quad \text{continuous} \quad \longrightarrow \quad \text{discrete}$$

do not compensate.

Another kind of discretization is that computers replace the real numbers by a finite number of rational numbers, namely, the floating-point numbers. The resulting rounding error will not be relevant for much of our analysis, except for investigations of stability.

1.4 The Binomial Method

The major part of the book is devoted to continuous models and their discretizations. With much less effort a discrete approach provides us with a short way to establish a first algorithm for calculating options. The resulting *binomial method* is robust and widely applicable.

In practice one is often interested in the one value $V(S_0, 0)$ of an option at the current spot price S_0. Then it can be unnecessarily costly to calculate

[5] The symbol Δt denotes a small increment in t (analogously $\Delta S, \Delta W$). In case Δ would be a number, the product with u would be denoted $\Delta \cdot u$ or $u\Delta$.

the surface $V(S,t)$ for the entire domain to extract the required information $V(S_0,0)$. The relatively small task of calculating $V(S_0,0)$ can be comfortably solved using the binomial method. This method is based on a tree-type grid applying appropriate binary rules at each grid point. The grid is not predefined but is constructed by the method. For illustration see the right-hand grid in Figure 1.7, and Figure 1.10.

1.4.1 A Discrete Model

We begin with discretizing the continuous time t, replacing t by equidistant time instances t_i. Let us use the notations

M: number of time steps
$\Delta t := \frac{T}{M}$
$t_i := i \cdot \Delta t, \quad i = 0, ..., M$
$S_i := S(t_i)$

So far the domain of the (S,t) half strip is *semidiscretized* in that it is replaced by parallel straight lines with distance Δt apart, leading to a discrete-time model. The next step of discretization replaces the continuous values S_i along the parallel $t = t_i$ by discrete values $S_{j,i}$, for all i and appropriate j. For a better understanding of the S-discretization compare Figure 1.8. This figure shows a mesh of the grid, namely, the transition from t to $t + \Delta t$, or from t_i to t_{i+1}.

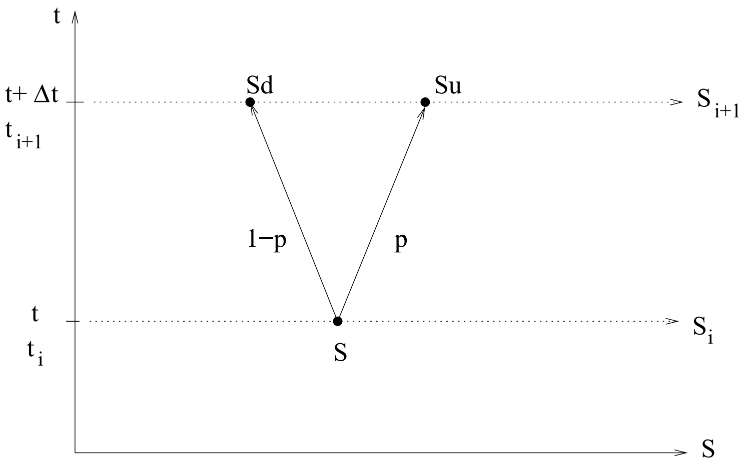

Fig. 1.8. The principle setup of the binomial method

Assumptions 1.3 (binomial method)

(Bi1) The price S over each period of time Δt can only have two possible outcomes: An initial value S either evolves "up" to Su, or "down" to Sd, with $0 < d < u$. Here u is the factor of an upward movement and d is the factor of a downward movement.

(Bi2) The probability of an up movement is p, $\mathsf{P}(\text{up}) = p$.

The rules (Bi1) and (Bi2) represent the framework of a binomial process. Such a process behaves like tossing a biased coin where the outcome "head" (up) occurs with probability p. At this stage of the modeling, the values of the three parameters u, d and p are undetermined. They are fixed in a way such that the model is consistent with the continuous model in case $\Delta t \to 0$. This aim leads to further assumptions. The basic idea of the approach is to equate the expectation and the variance of the discrete model with the corresponding values of the continuous model. This amounts to require

(Bi3) Expectation and variance of S refer to their continuous counterparts, evaluated for the risk-free interest rate r.

This assumption leads to equations for the parameters u, d, p. The resulting probability P of (Bi2) does not reflect the expectations of an individual in the market. Rather P is an artificial risk-neutral probability that matches (Bi3).[6] The expectation E below in (1.4) refers to this probability; this is sometimes written E_P. (We shall return to the assumptions (Bi1), (Bi2), and (Bi3) in the subsequent Section 1.5.) Let us further assume that no dividend is paid within the time period of interest. This assumption simplifies the derivation of the method and can be removed later.

1.4.2 Derivation of Equations

Recall the definition of the expectation for the discrete case, Appendix B1, equation (B1.13), and conclude

$$\mathsf{E}(S_{i+1}) = pS_iu + (1-p)S_id\,.$$

Here S_i is an arbitrary value, which develops randomly to S_{i+1}, when t_i proceeds to t_{i+1}, following the assumptions (Bi1) and (Bi2). In this sense, E is a conditional expectation. As will be seen in Section 1.7.2, the expectation of the continuous model is

$$\mathsf{E}(S_{i+1}) = S_i\,\mathrm{e}^{r\Delta t} \tag{1.4}$$

Equating gives

[6] To distinguish this specific "money market measure" P from other probabilities, one gives it a specific notation. In later sections we shall use the symbol Q.

$$e^{r\Delta t} = pu + (1-p)d\,. \tag{1.5}$$

This is the first of three equations required to fix u, d, p. Solved for the risk-neutral probability p we obtain

$$p = \frac{e^{r\Delta t} - d}{u - d}\,. \tag{1.6}$$

To be a valid model of probability, $0 \le p \le 1$ must hold. This is equivalent to

$$d \le e^{r\Delta t} \le u\,. \tag{1.7}$$

These inequalities relate the upward and downward movements of the asset price to the riskless interest rate r. The inequalities (1.7) are no new assumption but follow from the no-arbitrage principle. The assumption $0 < d < u$ is sustained.

Next we equate variances. Via the variance the volatility σ enters the model. From the continuous model we apply the relation

$$\mathsf{E}(S_{i+1}^2) = S_i^2\, e^{(2r+\sigma^2)\Delta t}\,. \tag{1.8}$$

For the relations (1.4) and (1.8) we refer to Section 1.8 ($\longrightarrow$ Exercise 1.12). Recall that the variance satisfies $\mathsf{Var}(S) = \mathsf{E}(S^2) - (\mathsf{E}(S))^2$ ($\longrightarrow$ Appendix B1). Equations (1.4) and (1.8) combine to

$$\mathsf{Var}(S_{i+1}) = S_i^2 e^{2r\Delta t}(e^{\sigma^2\Delta t} - 1)\,.$$

On the other hand the discrete model satisfies

$$\begin{aligned}\mathsf{Var}(S_{i+1}) &= \mathsf{E}(S_{i+1}^2) - (\mathsf{E}(S_{i+1}))^2 \\ &= p(S_i u)^2 + (1-p)(S_i d)^2 - S_i^2(pu + (1-p)d)^2\,.\end{aligned}$$

Equating variances of the continuous and the discrete model, and applying (1.5) leads to

$$\begin{aligned}e^{2r\Delta t}(e^{\sigma^2\Delta t} - 1) &= pu^2 + (1-p)d^2 - (e^{r\Delta t})^2 \\ e^{2r\Delta t + \sigma^2\Delta t} &= pu^2 + (1-p)d^2\end{aligned} \tag{1.9}$$

The equations (1.5), (1.9) constitute two relations for the three unknowns u, d, p. We are free to impose an arbitrary third equation. One example is the plausible assumption

$$u \cdot d = 1\,, \tag{1.10}$$

which reflects a symmetry between upward and downward movement of the asset price. Now the parameters u, d and p are fixed. They depend on r, σ and Δt. So does the grid, which is analyzed next (Figure 1.9).

The above rules are applied to each grid line $i = 0, \ldots, M$, starting at $t_0 = 0$ with the specific value $S = S_0$. Attaching meshes of the kind depicted in Figure 1.8 for subsequent values of t_i builds a tree with node values $Su^j d^k$

and $j+k=i$. In this way, specific discrete values $S_{j,i}$ of S_i and the nodes of the tree are defined. Since the same constant factors u and d underlie all meshes and since $Sud = Sdu$ holds, after the time period $2\Delta t$ the asset price can only take three values rather than four: The tree is recombining. It does not matter which of the two possible paths we take to reach Sud. This property extends to more than two time periods. Consequently the binomial process defined by Assumption 1.3 is *path independent*. Accordingly at expiration time $T = M\Delta t$ the price S can take only the $(M+1)$ discrete values $Su^j d^{M-j}$, $j = 0, 1, ..., M$. By (1.10) these are the values $Su^{2j-M} =: S_{j,M}$. The number of nodes in the tree grows quadratically in M. (Why?)

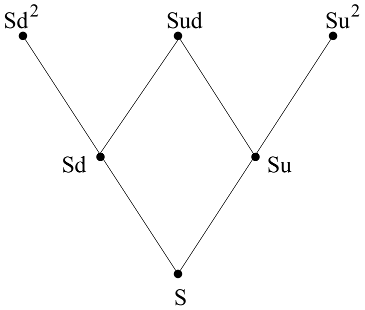

Fig. 1.9. Sequence of several meshes (schematically)

The symmetry of the choice $ud = 1$ becomes apparent in that after two time steps the asset value S repeats. (Compare also Figure 1.10.) For $ud = 1$, the central line of the tree grows vertically. The vertical arrangement is advantageous for matching a tree to barriers. But to smooth the convergence, it may be advisable to bend the tree such that its central line ends up at the strike. (We return to such improvements below.) In a (t, S)-plane the tree can be interpreted as a grid of exponential-like curves. The binomial approach defined by (Bi1) with the proportionality between S_i and S_{i+1} reflects exponential growth or decay of S. Since the tree extends from $S_0 d^M$ to $S_0 u^M$, all grid points have the desirable property $S > 0$, but for large M the tree becomes unrealistically wide.

1.4.3 Solution of the Equations

Using the abbreviation $\alpha := \mathrm{e}^{r\Delta t}$ we obtain by elimination (which the reader may check in more generality in Exercise 1.14b) the quadratic equation

$$0 = u^2 - u(\underbrace{\alpha^{-1} + \alpha \mathrm{e}^{\sigma^2 \Delta t}}_{=:2\beta}) + 1\,,$$

with solutions $u = \beta \pm \sqrt{\beta^2 - 1}$. By virtue of $ud = 1$ and Vieta's Theorem, d is the solution with the minus sign. In summary the three parameters u, d, p are given by

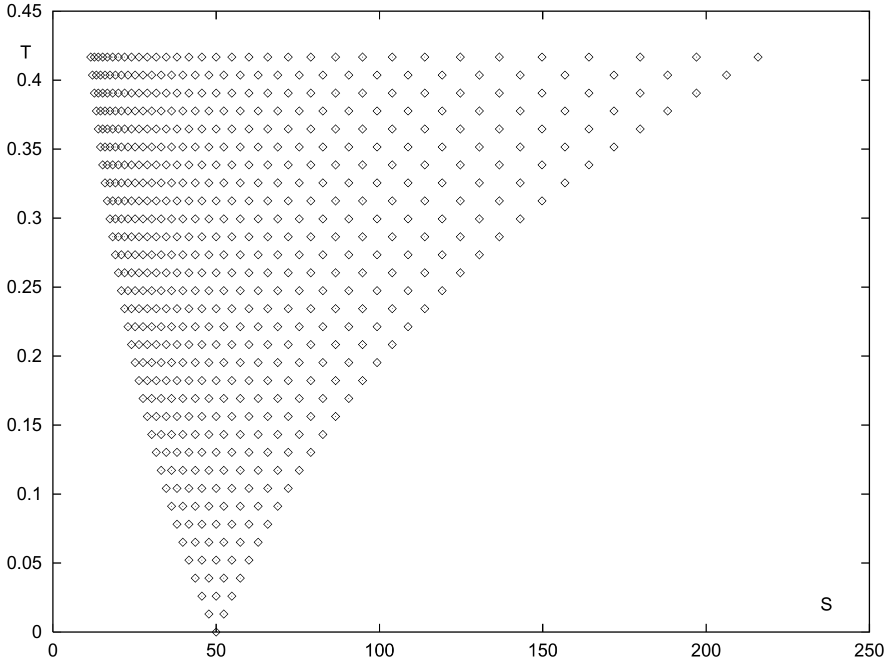

Fig. 1.10. Tree in the (S, t)-plane for $M = 32$ (data of Example 1.6)

$$\boxed{\begin{aligned} \beta &:= \frac{1}{2}(\mathrm{e}^{-r\Delta t} + \mathrm{e}^{(r+\sigma^2)\Delta t}) \\ u &= \beta + \sqrt{\beta^2 - 1} \\ d &= 1/u = \beta - \sqrt{\beta^2 - 1} \\ p &= \frac{\mathrm{e}^{r\Delta t} - d}{u - d} \end{aligned}} \tag{1.11}$$

A consequence of this approach is that up to terms of higher order the relation $u = \mathrm{e}^{\sigma\sqrt{\Delta t}}$ holds ($\longrightarrow$ Exercise 1.6). Therefore the extension of the tree in S-direction matches the volatility of the asset. So the tree is scaled well and will cover a relevant range of S-values.

1.4.4 A Basic Algorithm

Next we transform the binomial method into an algorithm.

Forward Phase: Initializing the Tree

Now the factors u and d can be considered as known, and the node values of S for each t_i until $t_M = T$ can be calculated. The current spot price $S = S_0$

for $t_0 = 0$ is the root of the tree. (To adapt the matrix-like notation to the two-dimensional grid of the tree, this initial price will be also denoted $S_{0,0}$.) Each initial price S_0 leads to another tree of node values $S_{j,i}$.

For $i = 1, 2, ..., M$ *calculate* :
$S_{j,i} := S_0 u^j d^{i-j}, \quad j = 0, 1, ..., i$

Now the grid points $(S_{j,i}, t_i)$ are fixed, on which approximations to the option values $V_{j,i} := V(S_{j,i}, t_i)$ are to be calculated.

Calculating the Option Value, Valuation on the Tree

For t_M and vanilla options, the payoff $V(S, t_M)$ is known from (1.1C), (1.1P). The payoff is valid for each S, including $S_{j,M} = Su^j d^{M-j}$, $j = 0, ..., M$. This defines the values $V_{j,M}$:

Call: $V(S(t_M), t_M) = \max\{S(t_M) - K, 0\}$, hence:

$$V_{j,M} := (S_{j,M} - K)^+ \tag{1.12C}$$

Put: $V(S(t_M), t_M) = \max\{K - S(t_M), 0\}$, hence:

$$V_{j,M} := (K - S_{j,M})^+ \tag{1.12P}$$

The **backward phase** recursively calculates for $t_{M-1},\ t_{M-2}, ...$ the option values V for all t_i, starting from $V_{j,M}$. The recursion is based on Assumption 1.3, (Bi3). Repeating the equation that corresponds to (1.5) with double index leads to

$$S_{j,i} \mathrm{e}^{r\Delta t} = pS_{j,i}u + (1-p)S_{j,i}d\,,$$

and

$$S_{j,i} \mathrm{e}^{r\Delta t} = pS_{j+1,i+1} + (1-p)S_{j,i+1}\,.$$

Relating the Assumption 1.3, (Bi3) of risk neutrality to V, $V_i = \mathrm{e}^{-r\Delta t}\mathsf{E}(V_{i+1})$, we obtain in double-index notation the recursion

$$V_{j,i} = \mathrm{e}^{-r\Delta t}\ (pV_{j+1,i+1} + (1-p)V_{j,i+1})\,. \tag{1.13}$$

So far, this recursion for $V_{j,i}$ is merely an analogy, which might be seen as a further assumption. But the following Section 1.5 will give a justification for (1.13), which turns out to be a consequence of the no-arbitrage principle and the risk-neutral valuation.

For **European options**, (1.13) is a recursion for $i = M-1, \ldots, 0$, starting from (1.12), and terminating with $V_{0,0}$. (For an illustration see Figure 1.11.) The obtained value $V_{0,0}$ is an approximation to the value $V(S_0, 0)$ of the continuous model, which results in the limit $M \to \infty$ ($\Delta t \to 0$). The accuracy of the approximation $V_{0,0}$ depends on M. This is reflected by writing $V_0^{(M)}$

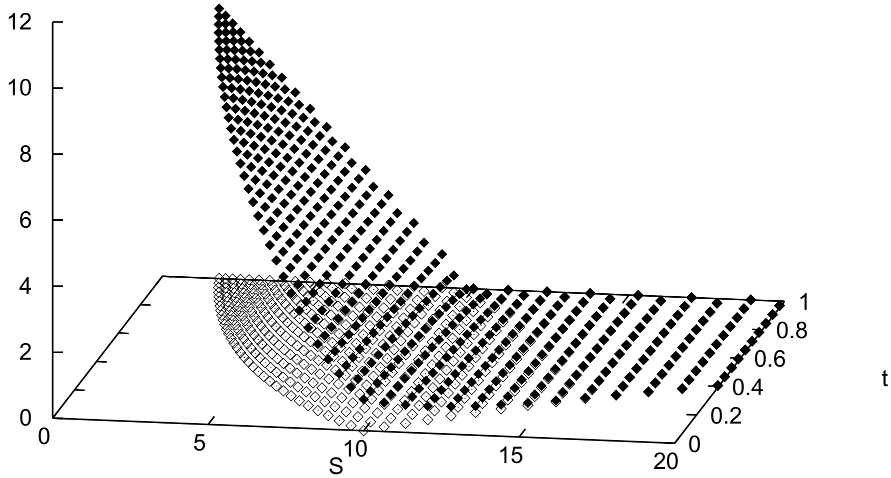

Fig. 1.11. Tree in the (S,t)-plane with (S,t,V)-points for $M = 32$ (data as in Figure 1.5)

($\longrightarrow$ Exercise 1.7). The basic idea of the approach implies that the limit of $V_0^{(M)}$ for $M \to \infty$ is the Black–Scholes value $V(S_0, 0)$ ($\longrightarrow$ Exercise 1.8).

For **American options**, the above recursion must be modified by adding a test whether early exercise is to be preferred. To this end the value of (1.13) is compared with the value of the payoff $\Psi(S)$. In this context, the value (1.13) is the "continuation value," denoted $V_{j,i}^{\text{cont}}$. And at any time t_i the holder optimizes the position and decides which of the two choices

$$\{\,\text{exercise, continue to hold}\,\}$$

is preferable. So the holder chooses the maximum

$$\max\{\,\Psi(S_{j,i}),\ V_{j,i}^{\text{cont}}\,\}\,.$$

This amounts to the *dynamic programming* principle: The optimality of the decision policy must be optimal also for the remaining time period. In summary, the dynamic-programming procedure, based on the equations (1.12) for i rather than M, combined with (1.13), reads as follows:

$$\begin{aligned}
V_{j,i}^{\text{cont}} &:= \mathrm{e}^{-r\Delta t}\cdot(pV_{j+1,i+1} + (1-p)V_{j,i+1}) \\
V_{j,i} &= \max\left\{(S_{j,i}-K)^+,\ V_{j,i}^{\text{cont}}\right\} \text{ for a call} \\
V_{j,i} &= \max\left\{(K-S_{j,i})^+,\ V_{j,i}^{\text{cont}}\right\} \text{ for a put}
\end{aligned} \tag{1.14}$$

The resulting algorithm is

Algorithm 1.4 (binomial method, basic version)

input: $r,\ \sigma,\ S = S_0,\ T,\ K,$ choice of put or call, European or American, M

calculate: $\Delta t := T/M,\ u,\ d,\ p$ from (1.11)

$S_{0,0} := S_0$

$S_{j,M} = S_{0,0} u^j d^{M-j},\ j = 0, 1, ..., M$

(for American options, also $S_{j,i} = S_{0,0} u^j d^{i-j}$ for $0 < i < M$, $j = 0, 1, ..., i$)

valuation: $V_{j,M}$ from (1.12)

$$V_{j,i} \text{ for } i < M \begin{cases} \text{from (1.13) for European options} \\ \text{from (1.14) for American options} \end{cases}$$

output: $V_{0,0}$ is the approximation $V_0^{(M)}$ to $V(S_0, 0)$

1.4.5 Improving the Convergence

The convergence order of the binomial method should be one. Then, ideally, extrapolation would make sense ($\longrightarrow$ Exercise 1.15). But the basic version of Algorithm 1.4 suffers from the fact that the payoff is not smooth at the strike K. This affects the accuracy at nodes near the kink $(S, t, V) = (K, T, 0)$. The convergence of Algorithm 1.4 can be easily improved in one of two ways.

For $S_0 \neq K$ the accuracy of the above basic version of Algorithm 1.4 also depends on how the strike K is grasped by the tree and its grid points. The error depending on M may oscillate, which is mainly caused by the erratic way how the point $(S, t) = (K, T)$ takes its place among the nodes $S_{j,M}$. This can be cured in an easy way. The tree can be bent such that for $i = M$ the medium grid point falls on the strike value K, no matter what (even) value of M is chosen. This is possible by generalizing (1.10) to $ud = \gamma$ for a suitable value of γ ($\longrightarrow$ Exercise 1.14). Corresponding special choices of u and d smooth the error significantly. This improvement of Algorithm 1.4 is straightforward to implement. With this version, extrapolation does make sense [LeR96].

Alternatively, certain critical intermediate results can be smoothed. Note that, even when the option is of the American style, the continuation values $V^{\text{cont}}_{j,M-1}$ in the last line $i = M-1$ are European style. As suggested by [BrD96], the linear combinations (1.13) for $i = M - 1$ can be replaced by the Black–Scholes formula (A4.10) or (A4.11). This only makes sense for a few nodes

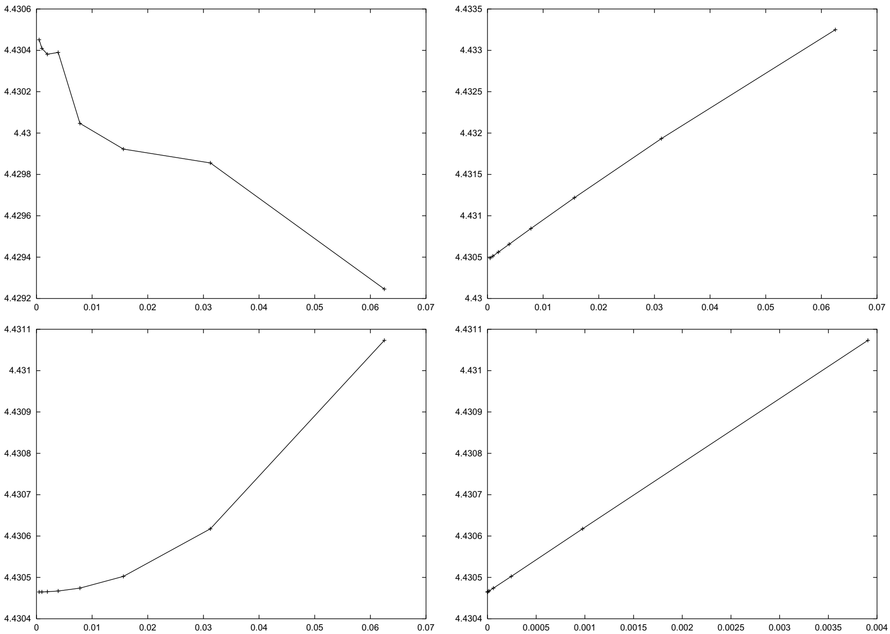

Fig. 1.12. Example 1.5: European-style option. Approximations $V^{(M)}$ over $\Delta t = 1/M$. top left: the basic Algorithm 1.4, linear convergence is hardly visible; top right: the improved algorithm with $ud = \gamma$ and γ from Exercise 1.14, linear convergence is clearly visible; bottom left: extrapolated values $V^{(M,\text{extr})}$ based on two approximations with M and $M/2$, $V^{(M,\text{extr})} := 2V^{(M)} - V^{(M/2)}$; bottom right: $V^{(M,\text{extr})}$ over Δt^2 shows quadratic convergence.

around the strike K, since for other j the improvement is not noticeable. Equation (1.14) must be adapted ($\longrightarrow$ Exercise 1.23).

Example 1.5 (European put)

Choose $K = 10,\ S = S_0 = 5,\ r = 0.06,\ \sigma = 0.3,\ T = 1$.

Recall that for European-style vanilla options an analytic solution exists, and Algorithm 1.4 is not needed. Hence, applying Algorithm 1.4 to Example 1.5 is only to create an ideal setting for the purpose of investigating accuracy and convergence. — The Table 1.2 lists approximations $V^{(M)}$ to $V(5, 0)$, both for $ud = 1$ and for $ud = \gamma$. The two main columns of Table 1.2 are graphed in the top two illustrations of Figure 1.12. The convergence towards the Black–Scholes value $V(S, 0)$ is visible; the latter was calculated by evaluating the analytic solution (A4.10). (In this book the number of printed decimals illustrates at best the attainable accuracy and does not reflect economic practice.)

The convergence rate of Algorithm 1.4 is visible in the results of Table 1.2, and in Figure 1.12. The rate is linear, $O(\Delta t) = O(M^{-1})$. For $S_0 \neq K$

and $ud = 1$ this rate is corrupted and hard to observe. The reader may wish to investigate more closely how the error of the basic version with $ud = 1$ decays with M ($\longrightarrow$ Exercises 1.7). It turns out that for the described basic version of the binomial method the convergence in M is not monotonic. It will not be recommendable to extrapolate these $V^{(M)}$-data to the limit $M \to \infty$, at least not the data of Table 1.2 ($ud = 1$). But the linear convergence rate can be seen well from the much better results obtained for $ud = \gamma$. The linear rate is reflected by the plots $V^{(M)}$ over M^{-1}, where the values of $V^{(M)}$ lie close to a straight line, which in this figure represents the linear error decay. Here extrapolation works well (lower illustrations in Figure 1.12). The convergence rate can also be calculated from the data ($\longrightarrow$ Exercises 1.15). This can be seen from Table 1.2 in a perfect way.

In case the function $V(S, 0)$ is to be approximated for several S out of an interval of S-values, other methods should be applied. The Figure 1.6 shows related results obtained by using the methods of Chapter 4.

Table 1.2. Results of Example 1.5, for γ see Exercise 1.14

M	$V^{(M)}(5,0)$ for $ud = 1$	$V^{(M)}(5,0)$ for $ud = \gamma$	with order
8	4.42507	4.43542	
16	4.42925	4.43325	0.833
32	4.429855	4.431933	0.923
64	4.429923	4.431218	0.963
128	4.430047	4.430846	0.982
256	4.430390	4.430657	0.991
2048	4.430451	4.430489	0.999
Black–Scholes	4.43046477621		

Example 1.6 (American put)

Choose $K = 50$, $S = 50$, $r = 0.1$, $\sigma = 0.4$, $T = 0.41666...$ ($\frac{5}{12}$ for 5 months), $M = 32$.

Here the pricing is at the money, so $\gamma = 1$. Figure 1.10 shows the tree for $M = 32$. The corresponding approximation to V_0 is $V^{(32)} = 4.2719$, calculated with Algorithm 1.4; almost three digits are correct. With $M = 2048$ and extrapolation we obtain 4.2842. At the early-exercise curve the surface $V(S, t)$ is not C^2-smooth. As a consequence the convergence order is not as close to $q = 1$ as in Example 1.5. — Note again that the function $V(S, 0)$ can be approximated with the methods of Chapter 4, compare Figure 4.11.

1.4.6 Sensitivities

The sensitivity parameters at $(S, t) = (S_{0,0}, 0)$

$$\text{delta} = \frac{\partial V}{\partial S}, \quad \text{gamma} = \frac{\partial^2 V}{\partial S^2}, \quad \text{theta} = \frac{\partial V}{\partial t},$$

can be approximated by difference quotients. The variations of V with S and t are expressed by the tree, and therefore information on derivatives can be obtained as by-product. For example, $\frac{V_{1,1}-V_{0,1}}{S_{1,1}-S_{0,1}}$ serves as a rough approximation for delta. But this quotient is evaluated at $t_1 = \Delta t$ rather than at $t = 0$. And a corresponding approximation of gamma requires three node values, which are available for t_2. To improve the accuracy, the difference quotients should be evaluated at the root node $(S, t) = (S_{0,0}, 0)$. This can be accomplished with a nice idea [PeV94]. The tree can be extended by starting it with a root at $t = -2\Delta t$ rather than at $t = 0$, with an S-value $S_{-1,-2}$. The extended tree follows the rules of Assumptions 1.3 and embeds the core tree. In this way, two additional lines of nodes are created, one at each side of the core tree. In particular, this creates two additional nodes at $t = 0$, with S-values $S_{-1,0}$ and $S_{1,0}$, and corresponding V-values $V_{-1,0}$ and $V_{1,0}$. Figure 1.9 may serve as illustration, when Sud stands for $S_{0,0}$. The approximations are

$$\text{delta:} \quad \frac{V_{1,0} - V_{-1,0}}{S_{1,0} - S_{-1,0}}$$

$$\text{gamma:} \quad \frac{\frac{V_{1,0}-V_{0,0}}{S_{1,0}-S_{0,0}} - \frac{V_{0,0}-V_{-1,0}}{S_{0,0}-S_{-1,0}}}{(S_{1,0} - S_{-1,0})/2}$$

$$\text{theta:} \quad \frac{V_{0,0} - V_{-1,-2}}{2\Delta t} \quad \text{(for example, when } ud = 1)$$

The costs of calculating these difference quotients can be neglected, because essentially the tree is not recalculated. This also holds for the extended tree: Compared with the overall costs of $O(M^2)$, the costs of the $2M+5$ additional nodes of the improved version are relatively small as long as M is large. Algorithm 1.4 needs to be adapted ($\longrightarrow$ Exercise 1.23).

Since the above sensitivities with respect to S and t are revealed by one calculated tree, they can be considered as bargain Greeks. In contrast, the sensitivities with respect to the parameters σ and r are more costly to approximate; these are the expensive Greeks because the entire tree must be recalculated. For example, to set up a difference quotient for the Greek vega$= \frac{\partial V}{\partial \sigma}$ requires to recalculate the tree for a parameter value σ_1 close to σ. If the corresponding value of the option obtained by the σ_1-tree is denoted V_1, then we have a difference-quotient approximation

$$\text{vega} \approx \frac{V - V_1}{\sigma - \sigma_1}.$$

In case one wishes an improved accuracy, one might apply a symmetric difference quotient, and recalculate the tree again on the other side, for $\sigma_2 := 2\sigma - \sigma_1$.

1.4.7 Extensions

The paying of dividends can be incorporated into the binomial algorithm. If a dividend D is paid at t_D the price of the asset drops by the same amount D. To take this jump into account, the tree is cut at t_D and the S-node values for $t < t_D$ are modified appropriately, see the remarks in Chapter 4, and [Hull00]. To allow for a constant dividend yield δ, replace r in (1.11) by $r - \delta$, but not in the discounting in (1.13), (1.14). ($\longrightarrow$ Exercise 1.22)

An extension of the binomial model is the *trinomial model*. Here each mesh offers three outcomes, with probabilities p_1, p_2, p_3 and $p_1 + p_2 + p_3 = 1$. The trinomial model allows for higher accuracy. The reader may wish to derive the trinomial method. For further hints, see Notes and Comments at the end of Chapter 1.

1.5 Risk-Neutral Valuation

In the previous Section 1.4 we have used the Assumptions 1.3 to derive an algorithm for valuation of options. This Section 1.5 discusses the assumptions again, leading to a different interpretation.

The situation of a path-independent binomial process with the two factors u and d continues to be the basis of the argumentation. The scenario is illustrated in Figure 1.13. Here the time period is the time to expiration T, which replaces Δt in the local mesh of Figure 1.8. Accordingly, this global model is called *one-period model*. The one-period model with only two possible values of S_T has two clearly defined values of the payoff, namely, $V^{(d)}$ (corresponds to $S_T = S_0 d$) and $V^{(u)}$ (corresponds to $S_T = S_0 u$). In contrast to the Assumptions 1.3 we neither assume the risk-neutral world (Bi3) nor the corresponding probability $\mathsf{P}(\text{up}) = p$ from (Bi2). Instead we derive the probability using the no-arbitrage argument. In this section the factors u and d are assumed to be given.

Let us construct a portfolio of an investor with a short position in one option and a long position consisting of Δ shares of an asset, where the asset is the underlying of the option. The portfolio manager must **choose the number Δ of shares such that the portfolio is riskless.** That is, a hedging strategy is needed. To discuss the hedging properly assume that no funds are added or withdrawn.

By Π_t we denote the wealth of this portfolio at time t. Initially the value is

$$\Pi_0 = S_0 \cdot \Delta - V_0 \,, \tag{1.15}$$

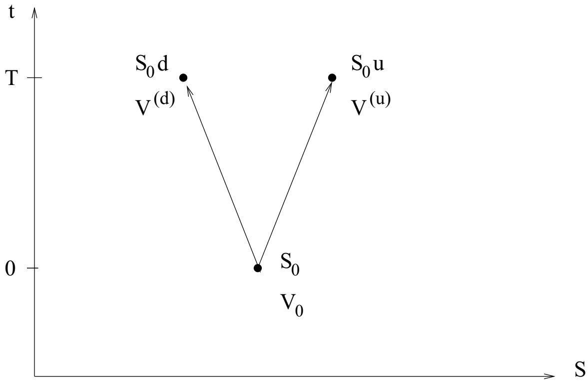

Fig. 1.13. One-period binomial model

where the value V_0 of the written option is not yet determined. At the end of the period the value V_T either takes the value $V^{(u)}$ or the value $V^{(d)}$. So the value of the portfolio Π_T at the end of the life of the option is either

$$\Pi^{(u)} = S_0 u \cdot \Delta - V^{(u)}$$

or

$$\Pi^{(d)} = S_0 d \cdot \Delta - V^{(d)} \, .$$

In the no-arbitrage world, Δ is chosen such that the value Π_T is riskless. Then all uncertainty is removed and $\Pi^{(u)} = \Pi^{(d)}$ must hold. This is equivalent to

$$(S_0 u - S_0 d) \cdot \Delta = V^{(u)} - V^{(d)} \, ,$$

which defines the strategy

$$\Delta = \frac{V^{(u)} - V^{(d)}}{S_0(u-d)} \, . \tag{1.16}$$

With this value of Δ the portfolio with initial value Π_0 evolves to the final value $\Pi_T = \Pi^{(u)} = \Pi^{(d)}$, regardless of whether the stock price moves up or down. Consequently the portfolio is riskless.

If we rule out early exercise, the final value Π_T is reached with certainty. The value Π_T must be compared to the alternative risk-free investment of an amount of money that equals the initial wealth Π_0, which after the time period T reaches the value $\mathrm{e}^{rT}\Pi_0$. Both the assumptions $\Pi_0 \mathrm{e}^{rT} < \Pi_T$ and $\Pi_0 \mathrm{e}^{rT} > \Pi_T$ would allow a strategy of earning a risk-free profit. This is in contrast to the assumed arbitrage-free world. Hence both $\Pi_0 \mathrm{e}^{rT} \geq \Pi_T$ and $\Pi_0 \mathrm{e}^{rT} \leq \Pi_T$ and equality must hold.[7] Accordingly the initial value Π_0 of

[7] For an American option it is not certain that Π_T can be reached because the holder may choose early exercise. In this situation we have only the inequality $\Pi_0 \mathrm{e}^{rT} \leq \Pi_T$.

the portfolio equals the discounted final value Π_T, discounted at the interest rate r,

$$\Pi_0 = \mathrm{e}^{-rT}\Pi_T\,.$$

This means

$$S_0 \cdot \Delta - V_0 = \mathrm{e}^{-rT}(S_0 u \cdot \Delta - V^{(u)})\,,$$

which upon substituting (1.16) leads to the value V_0 of the option:

$$\begin{aligned}
V_0 &= S_0 \cdot \Delta - \mathrm{e}^{-rT}(S_0 u \Delta - V^{(u)}) \\
&= \mathrm{e}^{-rT}\{\Delta \cdot [S_0 \mathrm{e}^{rT} - S_0 u] + V^{(u)}\} \\
&= \tfrac{\mathrm{e}^{-rT}}{u-d}\{(V^{(u)} - V^{(d)})(\mathrm{e}^{rT} - u) + V^{(u)}(u-d)\} \\
&= \tfrac{\mathrm{e}^{-rT}}{u-d}\{V^{(u)}(\mathrm{e}^{rT} - d) + V^{(d)}(u - \mathrm{e}^{rT})\} \\
&= \mathrm{e}^{-rT}\{V^{(u)}\tfrac{\mathrm{e}^{rT}-d}{u-d} + V^{(d)}\tfrac{u-\mathrm{e}^{rT}}{u-d}\} \\
&= \mathrm{e}^{-rT}\{V^{(u)}q + V^{(d)} \cdot (1-q)\}
\end{aligned}$$

with

$$q := \frac{\mathrm{e}^{rT} - d}{u - d}\,. \tag{1.17}$$

We have shown that with q from (1.17) the value of the option is given by

$$V_0 = \mathrm{e}^{-rT}\{V^{(u)}q + V^{(d)} \cdot (1-q)\}\,. \tag{1.18}$$

The expression for q in (1.17) is identical to the formula for p in (1.6), which was derived in the previous section. Again we have

$$0 < q < 1 \quad \Longleftrightarrow \quad d < \mathrm{e}^{rT} < u\,.$$

Presuming these bounds for u and d, q can be interpreted as a probability Q. Then $qV^{(u)} + (1-q)V^{(d)}$ is the expected value of the payoff with respect to this probability (1.17),

$$\mathsf{E}_{\mathsf{Q}}(V_T) = qV^{(u)} + (1-q)V^{(d)}\,.$$

Now (1.18) can be written

$$V_0 = \mathrm{e}^{-rT}\mathsf{E}_{\mathsf{Q}}(V_T)\,. \tag{1.19}$$

That is, the value of the option is obtained by discounting the expected payoff [with respect to q from (1.17)] at the risk-free interest rate r. An analogous calculation shows

$$\mathsf{E}_{\mathsf{Q}}(S_T) = qS_0 u + (1-q)S_0 d = S_0 \mathrm{e}^{rT}\,.$$

The probabilities p of Section 1.4 and q from (1.17) are defined by identical formulas (with T corresponding to Δt). Hence $p = q$, and $\mathsf{E}_{\mathsf{P}} = \mathsf{E}_{\mathsf{Q}}$. But the underlying arguments are different. Recall that in Section 1.4 we showed the implication

$$\mathsf{E}(S_T) = S_0 \mathrm{e}^{rT} \quad \Longrightarrow \quad p = \mathsf{P}(\mathrm{up}) = \frac{\mathrm{e}^{rT} - d}{u - d}\,,$$

whereas in this section we arrive at the implication

$$p = \mathsf{P}(\mathrm{up}) = \frac{\mathrm{e}^{rT} - d}{u - d} \quad \Longrightarrow \quad \mathsf{E}(S_T) = S_0 \mathrm{e}^{rT}\,.$$

So both statements must be equivalent. Setting the probability of the up movement equal to p is equivalent to assuming that the expected return on the asset equals the risk-free rate. This can be rewritten as

$$\mathrm{e}^{-rT} \mathsf{E}_{\mathsf{P}}(S_T) = S_0\,. \tag{1.20}$$

The important property expressed by equation (1.20) is that of a *martingale*: The random variable $\mathrm{e}^{-rT} S_T$ of the left-hand side has the tendency to remain at the same level. That is why a martingale is also called "fair game." A martingale displays no trend, where the trend is measured with respect to E_{P}. In the martingale property of (1.20) the discounting at the risk-free interest rate r exactly matches the risk-neutral probability P of (1.6)/(1.17). The specific probability for which (1.20) holds is also called *martingale measure.*

Summary of results for the one-period model: Under the Assumptions 1.2 of the market model, the choice Δ of (1.16) eliminates the random-dependence of the payoff and makes the portfolio riskless. There is a specific probability Q (P in Section 1.4) with $\mathsf{Q}(\mathrm{up}) = q$, q from (1.17), such that the value V_0 satisfies (1.19), and S_0 the analogous property (1.20). These properties involve the risk-neutral interest rate r. That is, the option is valued in a risk-neutral world, and the corresponding Assumption 1.3 (Bi3) is meaningful.

In the real-world economy, growth rates in general are different from r, and individual subjective probabilities differ from our Q. But the assumption of a risk-neutral world leads to a fair valuation of options. The obtained value V_0 can be seen as a *rational* price. In this sense the resulting value V_0 applies to the real world. The risk-neutral valuation can be seen as a technical tool. The assumption of risk neutrality is just required to define and calculate a rational price or fair value of V_0. For this specific purpose we do not need actual growth rates of prices, and individual probabilities are not relevant. But note that we do not really assume that financial markets are actually free of risk.

The general principle outlined for the one-period model is also valid for the multiperiod binomial model and for the continuous model of Black and Scholes ($\longrightarrow$ Exercise 1.8).

The Δ of (1.16) is the hedge parameter *delta*, which eliminates the risk exposure of our portfolio caused by the written option. In multiperiod models and continuous models Δ must be adapted dynamically. The expression (1.16) can be seen as a discretized version of the continuous-case definition

$$\Delta = \Delta(S,t) = \frac{\partial V(S,t)}{\partial S}\,.$$

1.6 Stochastic Processes

Brownian motion originally meant the erratic motion of a particle (pollen) on the surface of a fluid, caused by tiny impulses of molecules. Wiener suggested a mathematical model for this motion, the *Wiener process*. But earlier Bachelier had applied Brownian motion to model the motion of stock prices, which instantly respond to the numerous upcoming information similar as pollen react to the impacts of molecules (Figure 1.14). To model such behavior, we use stochastic processes.

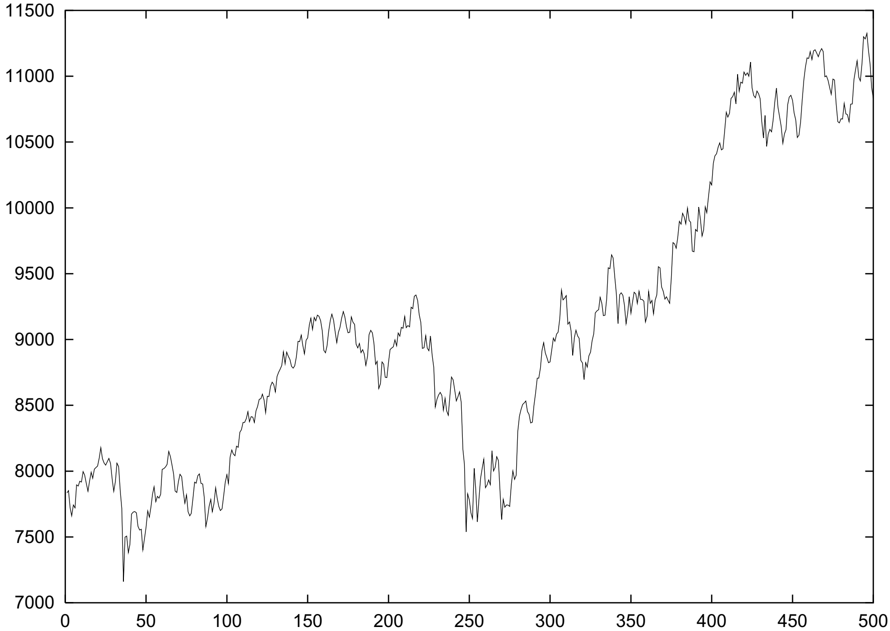

Fig. 1.14. The Dow at 500 trading days from September 8, 1997 through August 31, 1999

A *stochastic process* is a family of random variables X_t, which are defined for a set of parameters t ($\longrightarrow$ Appendix B1). Here we consider the continuous-time situation. That is, $t \in \mathbb{R}$ varies continuously in a time interval I, which typically represents $0 \le t \le T$. A more complete notation for a stochastic process is $\{X_t, t \in I\}$, or $(X_t)_{0 \le t \le T}$. Let the chance "play," then the resulting function X_t is called *realization* or *path* of the stochastic process.

Special properties of stochastic processes have lead to the following names:

Gaussian process: All finite-dimensional distributions $(X_{t_1}, \ldots, X_{t_k})$ are Gaussian. Hence specifically X_t is distributed normally for all t.

Markov process: Only the present value of X_t is relevant for its future motion. That is, the past history is fully reflected in the present value.[8]

An example of a process that is both Gaussian and Markov, is the Wiener process. Wiener processes are important building blocks for models of financial markets, and are the main theme of this section.

1.6.1 Wiener Process

Definition 1.7 (Wiener process, standard Brownian motion)

A Wiener process (or standard Brownian motion; notation W_t or W) is a time-continuous process for $t \geq 0$ with the properties

(a) $W_0 = 0$
(b) $W_t \sim \mathcal{N}(0, t)$ for all $t \geq 0$. That is, for each t the random variable W_t is *distributed normally*, with mean $\mathsf{E}(W_t) = 0$ and variance $\mathsf{Var}(W_t) = \mathsf{E}(W_t^2) = t$.
(c) All increments $\Delta W_t := W_{t+\Delta t} - W_t$ on non overlapping time intervals are *independent*: That is, the displacements $W_{t_2} - W_{t_1}$ and $W_{t_4} - W_{t_3}$ are independent for all $0 \leq t_1 < t_2 \leq t_3 < t_4$.
(d) W_t depends *continuously* on t.

Generally for $0 \leq s < t$ the property $W_t - W_s \sim \mathcal{N}(0, t-s)$ holds, in particular

$$\mathsf{E}(W_t - W_s) = 0 \,, \tag{1.21a}$$

$$\mathsf{Var}(W_t - W_s) = \mathsf{E}((W_t - W_s)^2) = t - s \,. \tag{1.21b}$$

The relations (1.21a,b) can be derived from Definition 1.7 ($\longrightarrow$ Exercise 1.9). The relation (1.21b) is also known as

$$\mathsf{E}((\Delta W_t)^2) = \Delta t \,. \tag{1.21c}$$

The independence of the increments according to Definition 1.7(c) implies for $t_{j+1} > t_j$ the independence of W_{t_j} and $(W_{t_{j+1}} - W_{t_j})$, but not of $W_{t_{j+1}}$ and $(W_{t_{j+1}} - W_{t_j})$. Wiener processes are examples of martingales —there is no drift. This process is an integral element of more involved models. For example, $X_t := \alpha + \mu t + W_t$ is a general Brownian motion with drift μ.

Discrete-Time Model

Let $\Delta t > 0$ be a constant time increment. For the discrete instances $t_j := j\Delta t$ the value W_t can be written as a sum of increments ΔW_k,

---

[8] This assumption together with the assumption of an immediate reaction of the market to arriving information are called *hypothesis of the efficient market* [Bou98].

$$W_{j\Delta t} = \sum_{k=1}^{j} \underbrace{\left(W_{k\Delta t} - W_{(k-1)\Delta t}\right)}_{=:\Delta W_k} .$$

The ΔW_k are independent and because of (1.21) normally distributed with $\mathsf{Var}(\Delta W_k) = \Delta t$. Increments ΔW with such a distribution can be calculated from standard normally distributed random numbers Z. The implication

$$Z \sim \mathcal{N}(0,1) \implies Z \cdot \sqrt{\Delta t} \sim \mathcal{N}(0, \Delta t)$$

leads to the discrete model of a Wiener process

$$\Delta W_k = Z\sqrt{\Delta t} \quad \text{for} \quad Z \sim \mathcal{N}(0,1) \quad \text{for each } k\,. \tag{1.22}$$

We summarize the numerical simulation of a Wiener process as follows:

Algorithm 1.8 (simulation of a Wiener process)

Start: $t_0 = 0,\ W_0 = 0;\ \Delta t$
loop $j = 1, 2, ...:$
$t_j = t_{j-1} + \Delta t$
draw $Z \sim \mathcal{N}(0,1)$
$W_j = W_{j-1} + Z\sqrt{\Delta t}$

The drawing of Z —that is, the calculation of $Z \sim \mathcal{N}(0,1)$— will be explained in Chapter 2. The values W_j are realizations of W_t at the discrete points t_j. The Figure 1.15 shows a realization of a Wiener process; 5000 calculated points (t_j, W_j) are joined by linear interpolation.

Almost all realizations of Wiener processes are nowhere differentiable. This becomes intuitively clear when the difference quotient

$$\frac{\Delta W_t}{\Delta t} = \frac{W_{t+\Delta t} - W_t}{\Delta t}$$

is considered. Because of relation (1.21b) the standard deviation of the numerator is $\sqrt{\Delta t}$. Hence for $\Delta t \to 0$ the normal distribution of the difference quotient disperses and no convergence can be expected.

1.6.2 Stochastic Integral

For motivation, let us suppose that the price development of an asset is described by a Wiener process W_t. Let $b(t)$ be the number of units of the asset held in a portfolio at time t. We start with the simplifying assumption that trading is only possible at discrete time instances t_j, which define a

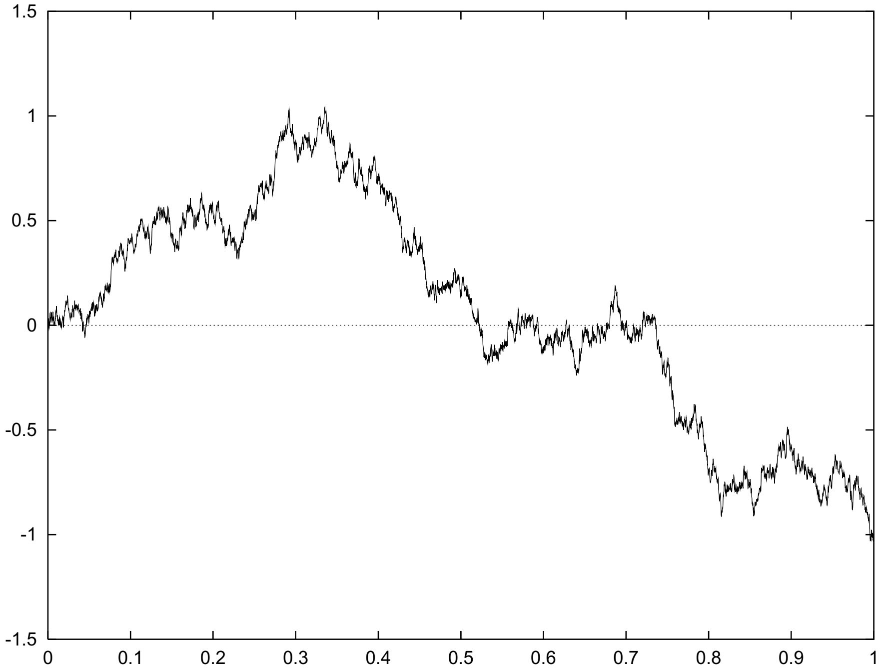

Fig. 1.15. Realization of a Wiener process, with $\Delta t = 0.0002$

partition of the interval $0 \le t \le T$. Then the trading strategy b is piecewise constant,

$$\begin{aligned} &b(t) = b(t_{j-1}) \quad \text{for} \quad t_{j-1} \le t < t_j \\ &\text{and} \quad 0 = t_0 < t_1 < \ldots < t_N = T\,. \end{aligned} \tag{1.23}$$

Such a function $b(t)$ is called *step function*. The trading gain for the subinterval $t_{j-1} \le t < t_j$ is given by $b(t_{j-1})(W_{t_j} - W_{t_{j-1}})$, and

$$\sum_{j=1}^{N} b(t_{j-1})(W_{t_j} - W_{t_{j-1}}) \tag{1.24}$$

represents the trading gain over the time period $0 \le t \le T$. The trading gain (possibly < 0) is determined by the strategy $b(t)$ and the price process W_t.

We now drop the assumption of fixed trading times t_j and allow b to be arbitrary continuous functions. This leads to the question whether (1.24) has a limit when with $N \to \infty$ the size of all subintervals tends to 0. If W_t would be of bounded variation than the limit exists and is called *Riemann–Stieltjes integral*

$$\int_0^T b(t)\,\mathrm{d}W_t\,.$$

In our situation this integral generally does not exist because almost all Wiener processes are not of bounded variation. That is, the *first variation* of W_t, which is the limit of

$$\sum_{j=1}^{N} |W_{t_j} - W_{t_{j-1}}| \,,$$

is unbounded even in case the lengths of the subintervals vanish for $N \to \infty$.

Although this statement is not of primary concern for the theme of this book[9], we digress for a discussion because it introduces the important rule $(\mathrm{d}W_t)^2 = \mathrm{d}t$. For an arbitrary partition of the interval $[0, T]$ into N subintervals the inequality

$$\sum_{j=1}^{N} |W_{t_j} - W_{t_{j-1}}|^2 \leq \max_j (|W_{t_j} - W_{t_{j-1}}|) \sum_{j=1}^{N} |W_{t_j} - W_{t_{j-1}}| \tag{1.25}$$

holds. The left-hand sum in (1.25) is the *second variation* and the right-hand sum the first variation of W for a given partition into subintervals. The expectation of the left-hand sum can be calculated using (1.21),

$$\sum_{j=1}^{N} \mathsf{E}(W_{t_j} - W_{t_{j-1}})^2 = \sum_{j=1}^{N} (t_j - t_{j-1}) = t_N - t_0 = T \,.$$

But even convergence in the mean holds:

Lemma 1.9 (second variation: convergence in the mean)

Let $t_0 = t_0^{(N)} < t_1^{(N)} < \ldots < t_N^{(N)} = T$ be a sequence of partitions of the interval $t_0 \leq t \leq T$ with

$$\delta_N := \max_j (t_j^{(N)} - t_{j-1}^{(N)}) \,. \tag{1.26}$$

Then (dropping the ${}^{(N)}$)

$$\underset{\delta_N \to 0}{\mathrm{l.i.m.}} \sum_{j=1}^{N} (W_{t_j} - W_{t_{j-1}})^2 = T - t_0 \tag{1.27}$$

Proof: The statement (1.27) means convergence in the mean ($\longrightarrow$ Appendix B1). Because of $\sum \Delta t_j = T - t_0$ we must show

$$\mathsf{E}\left(\sum_j ((\Delta W_j)^2 - \Delta t_j) \right)^2 \to 0 \quad \text{for} \quad \delta_N \to 0 \,.$$

Carrying out the multiplications and taking the mean gives

---

[9] The less mathematically oriented reader may like to skip the rest of this subsection.

$$2\sum_j (\Delta t_j)^2$$

($\longrightarrow$ Exercise 1.10). This can be bounded by $2(T-t_0)\delta_N$, which completes the proof.

Part of the derivation can be summarized to

$$\mathsf{E}((\Delta W_t)^2 - \Delta t) = 0\,,\ \mathsf{Var}((\Delta W_t)^2 - \Delta t) = 2(\Delta t)^2\,.$$

Symbolically, this property of a Wiener process is written

$$\boxed{(\mathrm{d}W_t)^2 = \mathrm{d}t} \tag{1.28}$$

It will be needed in subsequent sections.

Now we know enough about the convergence of the left-hand sum of (1.25) and turn to the right-hand side of this inequality. The continuity of W_t implies

$$\max_j |W_{t_j} - W_{t_{j-1}}| \to 0 \qquad \text{for} \quad \delta_N \to 0\,.$$

Convergence in the mean applied to (1.25) shows that the vanishing of this factor must be compensated by an unbounded growth of the other factor, to make (1.27) happen. So

$$\sum_{j=1}^{N} |W_{t_j} - W_{t_{j-1}}| \to \infty \quad \text{for} \quad \delta_N \to 0\,.$$

In summary, Wiener processes are not of bounded variation, and the integration with respect to W_t can not be defined as an elementary limit of (1.24).

The aim is to construct a stochastic integral

$$\int_{t_0}^{t} f(s)\,\mathrm{d}W_s$$

for general stochastic integrands $f(t)$. For our purposes it suffices to briefly sketch the Itô integral, which is the prototype of a stochastic integral.

For a step function b from (1.23) an integral can be defined via the sum (1.24),

$$\int_{t_0}^{t} b(s)\mathrm{d}W_s := \sum_{j=1}^{N} b(t_{j-1})(W_{t_j} - W_{t_{j-1}})\,. \tag{1.29}$$

This is the Itô integral over a step function b. In case the $b(t_{j-1})$ are random variables, b is called a *simple process*. Then the Itô integral is again defined by (1.29). Stochastically integrable functions f can be obtained as limits of simple processes b_n in the sense

$$\mathsf{E}\left[\int_{t_0}^{t} (f(s) - b_n(s))^2 \mathrm{d}s\right] \to 0 \quad \text{for} \quad n \to \infty\,. \tag{1.30}$$

Convergence in terms of integrals $\int \mathrm{d}s$ carries over to integrals $\int \mathrm{d}W_t$. This is achieved by applying Cauchy convergence $\mathsf{E}\int (b_n - b_m)^2 \mathrm{d}s \to 0$ and the *isometry*

$$\mathsf{E}\left[\left(\int_{t_0}^{t} b(s)\,\mathrm{d}W_s\right)^2\right] = \mathsf{E}\left[\int_{t_0}^{t} b(s)^2\,\mathrm{d}s\right].$$

Hence the integrals $\int b_n(s)\mathrm{d}W_s$ form a Cauchy sequence with respect to convergence in the mean. Accordingly the Itô integral of f is defined as

$$\int_{t_0}^{t} f(s)\,\mathrm{d}W_s := \text{l.i.m.}_{n\to\infty} \int_{t_0}^{t} b_n(s)\,\mathrm{d}W_s\,,$$

for simple processes b_n defined by (1.30). The value of the integral is independent of the choice of the b_n in (1.30). The Itô integral as function in t is a stochastic process with the martingale property.

If an integrand $a(x,t)$ depends on a stochastic process X_t, the function f is given by $f(t) = a(X_t, t)$. For the simplest case of a constant integrand $a(X_t, t) = a_0$ the Itô integral can be reduced via (1.29) to

$$\int_{t_0}^{t} \mathrm{d}W_s = W_t - W_{t_0}\,.$$

For the "first" nontrivial Itô integral consider $X_t = W_t$ and $a(W_t, t) = W_t$. Its solution will be presented in Section 3.2.

Wiener processes are the driving machines for diffusion models (next section). There are other stochastic processes that can be used for modeling financial markets. For several models jump processes are considered. We turn to jump processes in Section 1.9.

1.7 Diffusion Models

Many fundamental models of financial markets use Wiener processes as driving process. These are the diffusion models discussed in this section. We discuss the main representative geometric Brownian motion, and explain the risk-neutral valuation in this context. Then we turn to more general processes, such as mean reversion.

1.7.1 Itô Process

Phenomena in nature, technology and economy are often modeled by means of deterministic differential equations $\dot{x} = \frac{\mathrm{d}}{\mathrm{d}t}x = a(x,t)$. This kind of modeling neglects stochastic fluctuations and is not appropriate for stock prices. If processes x are to include Wiener processes as special case, the derivative $\frac{\mathrm{d}}{\mathrm{d}t}x$ is meaningless. To circumvent non-differentiability, *integral equations* are used to define a general class of stochastic processes. Randomness is inserted additively,

$$x(t) = x_0 + \int_{t_0}^{t} a(x(s), s)\mathrm{d}s \quad + \text{ randomness},$$

with an Itô integral with respect to the Wiener process W_t. The first integral in the resulting integral equation is an ordinary (Lebesgue- or Riemann-) integral. The final integral equation is symbolically written as a “stochastic differential equation” (SDE) and named after Itô.

Definition 1.10 (Itô stochastic differential equation)

An Itô stochastic differential equation is

$$\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\,\mathrm{d}W_t\,; \tag{1.31a}$$

this together with $X_{t_0} = X_0$ is a symbolic short form of the integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_s, s)\,\mathrm{d}s + \int_{t_0}^{t} b(X_s, s)\,\mathrm{d}W_s\,. \tag{1.31b}$$

The terms in (1.31) are named as follows:

$a(X_t, t)$: drift term or drift coefficient
$b(X_t, t)$: diffusion coefficient

The integral equation (1.31b) defines a large class of stochastic processes X_t; solutions X_t of (1.31b) are called Itô process, or stochastic diffusion.

As intended, the Wiener process is a special case of an Itô process, because from $X_t = W_t$ the trivial SDE $\mathrm{d}X_t = \mathrm{d}W_t$ follows, hence the drift vanishes, $a = 0$, and $b = 1$ in (1.31). If $b \equiv 0$ and X_0 is constant, then the SDE becomes deterministic.

An experimental approach may help to develop an intuitive understanding of Itô processes. The simplest numerical method combines the discretized version of the Itô SDE

$$\Delta X_t = a(X_t, t)\Delta t + b(X_t, t)\Delta W_t \tag{1.32}$$

with the Algorithm 1.8 for approximating a Wiener process, using the same Δt for both discretizations. The result is

Algorithm 1.11 (Euler discretization of an SDE)

Approximations y_j to X_{t_j} are calculated by

Start: $t_0,\ y_0 = X_0,\ \Delta t,\ W_0 = 0$
loop $j = 0, 1, 2, ...$
$t_{j+1} = t_j + \Delta t$
$\Delta W = Z\sqrt{\Delta t}$ with $Z \sim \mathcal{N}(0,1)$
$y_{j+1} = y_j + a(y_j, t_j)\Delta t + b(y_j, t_j)\Delta W$

In the simplest setting, the *step length* Δt is chosen equidistant, $\Delta t = T/m$ for a suitable integer m. Of course the accuracy of the approximation depends on the choice of Δt ($\longrightarrow$ Chapter 3). The evaluation is straightforward. In case the functions a and b are easy to calculate, the greatest effort may be to calculate random numbers $Z \sim \mathcal{N}(0,1)$ ($\longrightarrow$ Section 2.3). Solutions to the SDE or to its discretized version for a given realization of the Wiener process are called *trajectories* or paths. By *simulation* of the SDE we understand the calculation of one or more trajectories. For the purpose of visualization, the discrete data are mostly joined by straight lines.

Example 1.12 $\mathrm{d}X_t = 0.05X_t\,\mathrm{d}t + 0.3X_t\,\mathrm{d}W_t$

Without the diffusion term the exact solution would be $X_t = X_0\mathrm{e}^{0.05t}$. For $X_0 = 50$, $t_0 = 0$ and a time increment $\Delta t = 1/250$ the Figure 1.16 depicts a trajectory X_t of the SDE for $0 \leq t \leq 1$. For another realization of a Wiener process W_t the solution looks different. This is demonstrated for a similar SDE in Figure 1.17.

1.7.2 Geometric Brownian Motion

Next we discuss one of the most important continuous models for the motion of stock prices S_t. This standard model assumes that the relative change (return) $\mathrm{d}S/S$ of a security in the time interval $\mathrm{d}t$ is composed of a deterministic drift $\mu\,\mathrm{d}t$ plus stochastic fluctuations in the form $\sigma\mathrm{d}W_t$:

Model 1.13 (geometric Brownian motion, GBM)

$$\boxed{\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t} \qquad (1.33,\ \text{GBM})$$

This SDE is linear in $X_t = S_t$, and $a(S_t, t) = \mu S_t$ is the drift rate with the expected *rate of return* μ, $b(S_t, t) = \sigma S_t$, σ is the volatility. (Compare Example 1.12 and Figure 1.16.) The geometric Brownian motion of (1.33) is

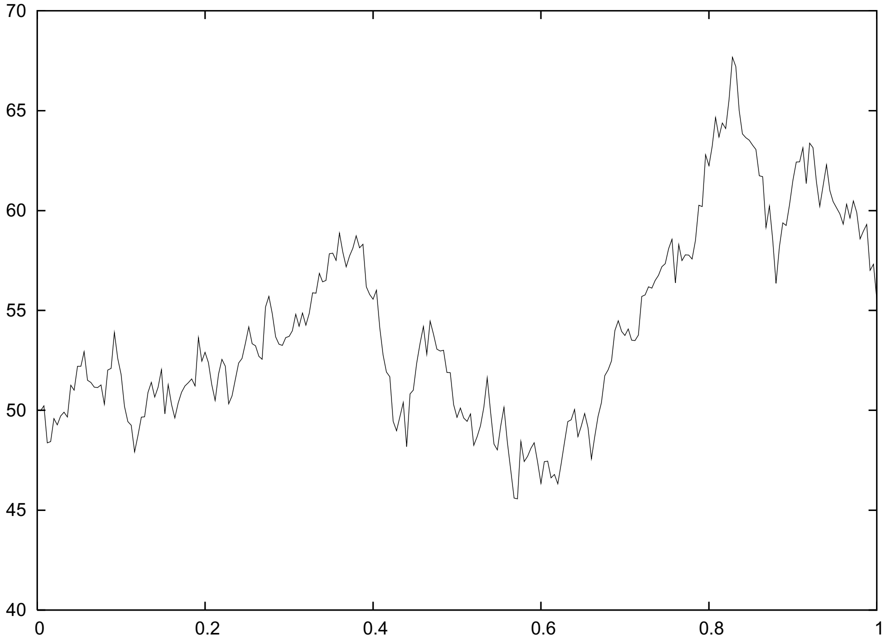

Fig. 1.16. Numerically approximated trajectory of Example 1.12 with $a = 0.05X_t$, $b = 0.3X_t$, $\Delta t = 1/250$, $X_0 = 50$

the reference model on which, for example, the Black–Scholes model is based. To match Assumption 1.2 assume that μ and σ are constant.

A theoretical solution of (1.33) will be given in (1.54). The deterministic part of (1.33) is the ordinary differential equation

$$\dot{S} = \mu S$$

with solution $S_t = S_0 e^{\mu(t-t_0)}$. For the linear SDE of (1.33) the expectation $\mathsf{E}(S_t)$ solves $\dot{S} = \mu S$. Hence

$$S_0 e^{\mu(t-t_0)} = \mathsf{E}(S_t \mid S_{t_0} = S_0)$$

is the expectation of the stochastic process and μ is the expected continuously compounded return earned by an investor per year, conditional on starting at S_0. The rate of return μ is also called *growth rate*. The function $S_0 e^{\mu(t-t_0)}$ can be seen as a core about which the process fluctuates. Accordingly the simulated values S_1 of the ten trajectories in Figure 1.17 group around the value $50 \cdot e^{0.1} \approx 55.26$.

Let us test empirically how the values S_1 distribute about their expected value. To this end calculate, for example, 10000 trajectories and count how many of the terminal values S_1 fall into the subintervals $k5 \leq t < (k+1)5$,

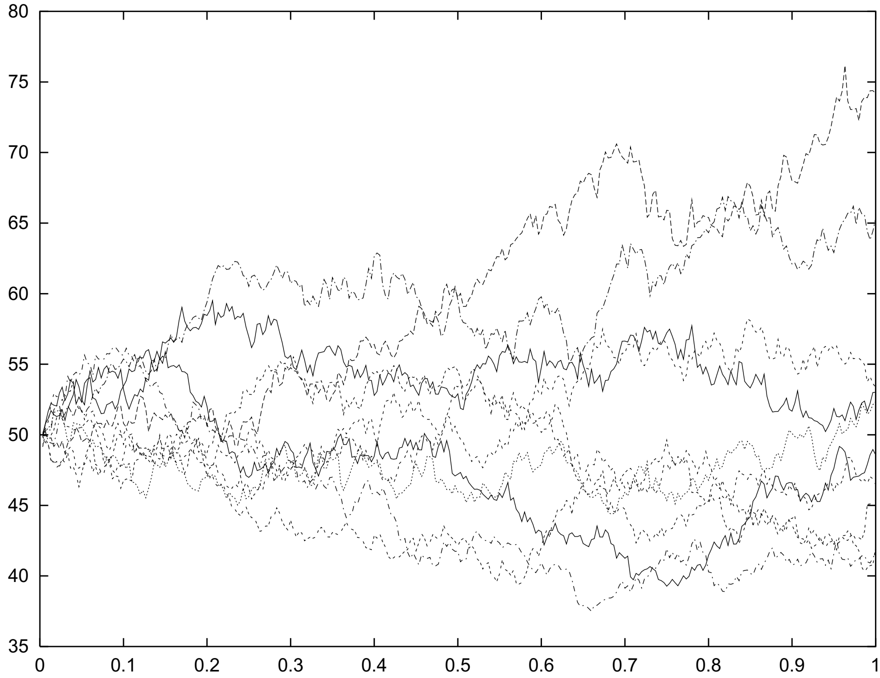

Fig. 1.17. 10 paths of SDE (1.33) with $S_0 = 50$, $\mu = 0.1$ and $\sigma = 0.2$

for $k = 0, 1, 2 \ldots$. Figure 1.18 shows the resulting histogram. Apparently the distribution is skewed. We revisit this distribution in the next section.

A discrete version of (1.33) is

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma Z \sqrt{\Delta t}\,, \tag{1.34a}$$

known from Algorithm 1.11. This approximation is valid as long as Δt is small and $S > 0$ ($\longrightarrow$ Exercise 1.24). The relative return reflected by the ratio $\frac{\Delta S}{S}$ is called one-period *simple return*, where we interpret Δt as one period. According to (1.34a) this return satisfies

$$\frac{\Delta S}{S} \sim \mathcal{N}(\mu \Delta t, \sigma^2 \Delta t)\,. \tag{1.34b}$$

The distribution of the simple return matches actual market data in a crude approximation, see for instance Figure 1.21. This allows to calculate estimates of historical values of the volatility σ.[10] Of course this assumes the market data to be correctly described by GBM. We will return to this in Section 1.8.

[10] For the *implied volatility* see Exercise 1.5.

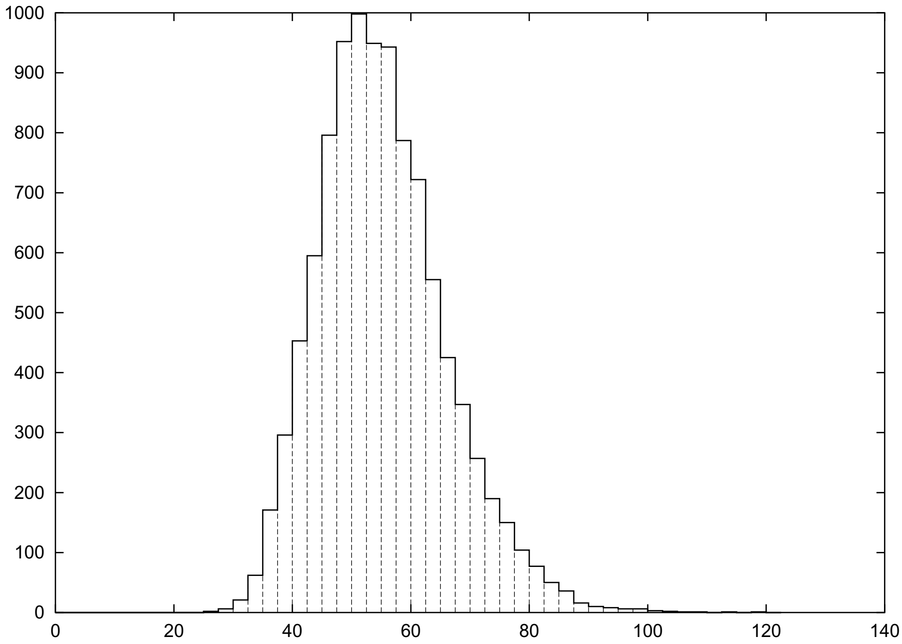

Fig. 1.18. Histogram of 10000 calculated values S_1 corresponding to (1.33), with $S_0 = 50$, $\mu = 0.1$, $\sigma = 0.2$

1.7.3 Risk-Neutral Valuation

We digress for the length of this subsection and again turn to the topic of a risk-neutral valuation, now for the continuous-time setting. In Section 1.5 we have shown

$$V_0 = \mathrm{e}^{-rT}\mathsf{E}_{\mathsf{Q}}(V_T)$$

for the one-period model. Formally, the same holds true for the market model based on GBM. But now the understanding of the risk-neutral probability Q is more involved. This subsection sketches the framework for GBM.

Let us rewrite GBM from (1.33) to get

$$\begin{aligned} \mathrm{d}S_t &= rS_t\,\mathrm{d}t + (\mu - r)S_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t \\ &= rS_t\,\mathrm{d}t + \sigma S_t\left[\frac{\mu - r}{\sigma}\,\mathrm{d}t + \mathrm{d}W_t\right], \end{aligned} \tag{1.35}$$

where W is Wiener process under the probability measure P. In the reality of the market, an investor expects $\mu > r$ as compensation for the risk that is higher for stocks than for bonds. In this sense, the quotient γ of the *excess return* $\mu - r$ to the risk σ,

$$\gamma := \frac{\mu - r}{\sigma}, \tag{1.36}$$

is called *market price of risk*. With this variable γ, (1.35) is written

$$dS_t = rS_t\,dt + \sigma S_t[\gamma\,dt + dW_t]\,. \tag{1.37}$$

For $\gamma \neq 0$ the drifted Brownian motion W_t^γ defined by

$$dW_t^\gamma = \gamma\,dt + dW_t \tag{1.38}$$

is no Wiener process under P. But under certain assumptions on γ there is another probability measure Q such that the process W_t^γ is a (standard) Wiener process under Q.[11] Equation (1.37) becomes

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t^\gamma\,. \tag{1.39}$$

Comparing this SDE to (1.33), notice that the growth rate μ is replaced by the risk-free rate r. Together the transition consists of

μ	$\rightarrow$	r
P	$\rightarrow$	Q
W	$\rightarrow$	W^γ

which is named **risk-neutral valuation principle** for GBM. To simulate (1.39) under Q, just apply the standard Algorithm 1.8 for the Wiener process W_t^γ. Then the rate r in (1.39) and W_t^γ correspond to the "risk-neutral measure" Q.

What is the reason for adjusting the probability measure $\mathsf{P} \rightarrow \mathsf{Q}$? The advantage of the risk-neutral measure Q is that the discounted process $e^{-rt}S_t$ is a martingale under Q,

$$d(e^{-rt}S_t) = \sigma e^{-rt}S_t\,dW_t^\gamma\,.$$

The **fundamental theorem of asset pricing** states that a market model is free of arbitrage if and only if there exists a probability measure Q such that the discounted asset prices are martingales with respect to Q [HaP81]. Hence the property of $e^{-rt}S_t$ having no drift is an essential ingredient of a no-arbitrage market and a prerequisite to modeling options. For a thorough discussion of the continuous model, martingale theory is used. (Some more background and explanation is provided by Appendix B3.) Let us summarize the situation in a remark:

Remark 1.14 (risk-neutral valuation principle)

For modeling options with underlying GBM, the original probability is adjusted to the risk-neutral probability Q. To simulate the process under Q, the return rate μ is replaced by the risk-free interest rate r, and W_t^γ is approximated as Wiener process.

[11] Girsanov's theorem, see Appendix B2. Q and P are equivalent.

1.7.4 Mean Reversion

The assumptions of a constant interest rate r and a constant volatility σ are quite restrictive. To overcome this simplification, SDEs for r_t and σ_t have been constructed that control r_t or σ_t stochastically. One class of models is based on the SDE

$$\mathrm{d}r_t = \alpha(R - r_t)\,\mathrm{d}t + \sigma_\mathrm{r} r_t^\beta\,\mathrm{d}W_t\,, \quad \alpha > 0\,, \tag{1.40}$$

again with driving force W_t as Wiener process. The drift term in (1.40) is positive for $r_t < R$ and negative for $r_t > R$, which causes a pull to R. This effect is called *mean reversion.* A *frequency* parameter α influences the strength of the reversion. The parameter R, which may depend on t, corresponds to a long-run mean of the interest rate over time. SDE (1.40) defines a general class of models, including several interesting special cases known under special names:

$\beta = 0,\ R = 0:$ Ornstein–Uhlenbeck process (OU)
$\beta = 0,\ R > 0:$ Vasicek model
$\beta = \frac{1}{2},\ R > 0:$ Cox–Ingersoll–Ross process (CIR)

Hull and White have extended the Vasicek model incorporating time dependence in the parameters. The CIR model [CoxIR85] is also called *square-root process*. Its volatility $\sigma_\mathrm{r}\sqrt{r_t}$ and with it the stochastic part vanish when r_t tends to zero. An illustration of the mean reversion is provided by Figure 1.19. In a transient phase (until $t \approx 1$ in the run documented in the figure) the relatively large deterministic term dominates, and the range $r \approx R$ is reached quickly. Thereafter the stochastic term dominates, and r dances about the mean value R. Figure 1.19 shows this for a Cox–Ingersoll–Ross model. For a discussion of related models we refer to [LaL96], [Hull00], [Kwok98]. The *calibration* of the models (that is, the adaption of the parameters to the data) is a formidable task ($\longrightarrow$ Section 1.10).

The SDE (1.40) is of a different kind as the GBM in (1.33). Coupling the SDE for r_t to that for S_t leads to a system of two SDEs. Even larger systems are obtained when further SDEs are coupled to define a stochastic process R_t or to calculate stochastic volatilities. Related examples are given by Examples 1.15 and 1.16 below. In particular for modeling options, stochastic volatilities have shown great potential. We come back to this in the Examples 1.15 and 1.16 below.

1.7.5 Vector-Valued SDEs

The Itô equation (1.31) is formulated as scalar equation; accordingly the SDE (1.33) represents a *one-factor model.* The general *multifactor* version can be written in the same notation. Then $X_t = (X_t^{(1)}, \ldots, X_t^{(n)})$ and $a(X_t, t)$ are n-dimensional vectors. The Wiener processes of each component SDE need

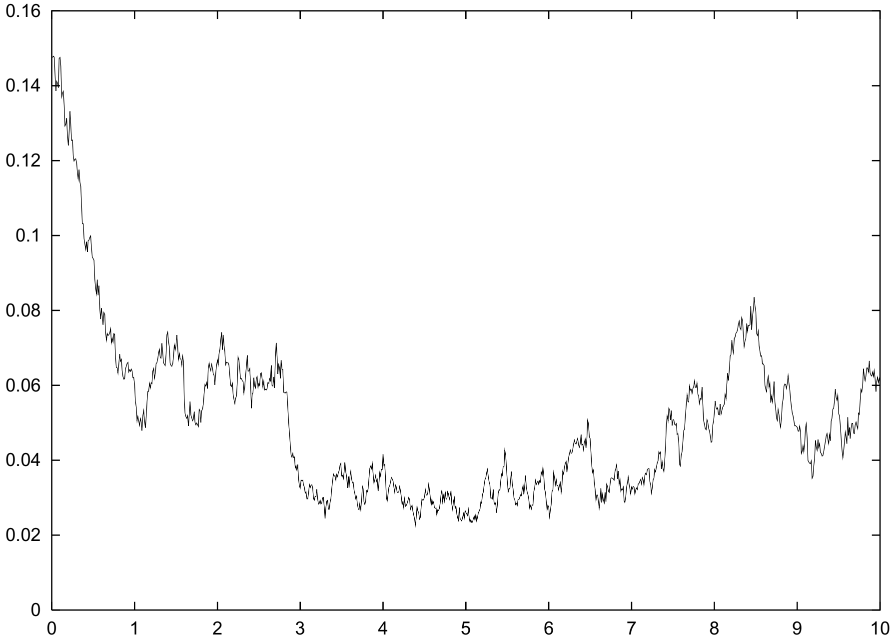

Fig. 1.19. Simulation r_t of the Cox–Ingersoll–Ross model (1.40) with $\beta = 0.5$ for $R = 0.05$, $\alpha = 1$, $\sigma_{\mathrm{r}} = 0.1$, $r_0 = 0.15$, $\Delta t = 0.01$

not be correlated. In the general situation, the Wiener process can be m-dimensional, with components $W_t^{(1)}, ..., W_t^{(m)}$. Then $b(X_t, t)$ is an $(n \times m)$-matrix, with elements b_{ik}. The interpretation of the SDE systems is componentwise. The scalar stochastic integrals are sums of m stochastic integrals,

$$X_t^{(i)} = X_0^{(i)} + \int_0^t a_i(X_s, s)\,\mathrm{d}s + \sum_{k=1}^{m} \int_0^t b_{ik}(X_s, s)\,\mathrm{d}W_s^{(k)}\,, \tag{1.41a}$$

for $i = 1, ..., n$, and $t_0 = 0$ for convenience. Or in the symbolic SDE notation, this system reads

$$\mathrm{d}X_t = a(X_t, t)\,\mathrm{d}t + b(X_t, t)\,\mathrm{d}W_t\,, \tag{1.41b}$$

where $b\,\mathrm{d}W$ is a matrix multiplication. When we take the components of the vector $\mathrm{d}W$ as uncorrelated,

$$\mathsf{E}\ (\mathrm{d}W^{(k)}\mathrm{d}W^{(j)}) = \begin{cases} 0 & \text{for } k \neq j \\ \mathrm{d}t & \text{for } k = j \end{cases} \tag{1.42}$$

then possible correlations between the components of $\mathrm{d}X$ must be carried by b.[12]

[12] We come back to this issue in Sections 2.3.3, and 3.5.5, and in Exercise 3.14.

Example 1.15 (mean-reverting volatility tandem)

We consider a three-factor model [HoPS92] with stock price S_t, instantaneous spot volatility σ_t and an averaged volatility ζ_t serving as mean-reverting parameter:

$$\begin{aligned} \mathrm{d}S &= \sigma S \, \mathrm{d}W^{(1)} \\ \mathrm{d}\sigma &= -(\sigma - \zeta)\mathrm{d}t + \alpha\sigma \, \mathrm{d}W^{(2)} \\ \mathrm{d}\zeta &= \beta(\sigma - \zeta)\mathrm{d}t \end{aligned}$$

Here and sometimes later on, we suppress the subscript t, which is possible when the role of the variables as stochastic processes is clear from the context. The rate of return μ of S is zero; $\mathrm{d}W^{(1)}$ and $\mathrm{d}W^{(2)}$ may be correlated. As seen from the SDE, the stochastic volatility σ follows the mean volatility ζ and is simultaneously perturbed by a Wiener process. Both σ and ζ provide mutual mean reversion, and stick together. Accordingly the two SDEs for σ and ζ may be seen as a tandem controlling the dynamics of the volatility. We recommend numerical tests. For motivation see Figure 3.2.

Example 1.16 (Heston's model)

Heston [Hes93] uses an Ornstein–Uhlenbeck process to model a stochastic volatility σ_t. Then the variance $v_t := \sigma_t^2$ follows a Cox–Ingersoll–Ross process (1.40). ($\longrightarrow$ Exercise 1.20) The system of Heston's model is

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \, \mathrm{d}t + \sqrt{v_t} \, S_t \, \mathrm{d}W_t^{(1)} \\ \mathrm{d}v_t &= \kappa(\theta - v_t) \, \mathrm{d}t + \sigma_\mathrm{v} \sqrt{v_t} \, \mathrm{d}W_t^{(2)} \end{aligned} \tag{1.43}$$

with two correlated Wiener processes $W_t^{(1)}, W_t^{(2)}$ and suitable parameters μ, κ, θ, σ_v, ρ, where ρ is the correlation between $W_t^{(1)}, W_t^{(2)}$. Hidden parameters might be the initial values S_0, v_0, if not available. This model establishes a correlation between price and volatility.

Computational Matters

Stochastic differential equations are simulated in the context of Monte Carlo methods. Thereby, the SDE is integrated N times, with N large ($N = 10000$ or much larger). Then the weight of any single trajectory is almost negligible. Expectation and variance are calculated over the N trajectories. Generally this costs an enormous amount of computing time. The required instruments are:

1.) Generating $\mathcal{N}(0, 1)$-distributed random numbers ($\longrightarrow$ Chapter 2)
2.) Integration methods for SDEs ($\longrightarrow$ Chapter 3)

1.8 Itô Lemma and Applications

Itô's lemma is most fundamental for stochastic processes. It may help, for example, to derive solutions of SDEs ($\longrightarrow$ Exercise 1.11). Suppose a "chain" of two functions X_t and $g(X_t, t)$. When a differential equation for X_t is given, what is the differential equation for $g(X_t, t)$?

1.8.1 Itô Lemma

Itô's lemma is the stochastic counterpart of the chain rule for deterministic functions $x(t)$ and $y(t) := g(x(t), t)$, which is

$$\frac{\mathrm{d}}{\mathrm{d}t} g(x(t), t) = \frac{\partial g}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial g}{\partial t} \,,$$

and can be written

$$\mathrm{d}x = a(x(t), t)\,\mathrm{d}t \;\Rightarrow\; \mathrm{d}g = \left(\frac{\partial g}{\partial x} a + \frac{\partial g}{\partial t} \right) \mathrm{d}t \;.$$

Here we state the one-dimensional version of the Itô lemma; for the multidimensional version see the Appendix B2.

Lemma 1.17 (Itô)

Suppose X_t follows an Itô process (1.31), $\mathrm{d}X_t = a(X_t, t)\mathrm{d}t + b(X_t, t)\mathrm{d}W_t$, and let $g(x, t)$ be a $\mathcal{C}^{2,1}$-smooth function (continuous $\frac{\partial g}{\partial x}$, $\frac{\partial^2 g}{\partial x^2}$, $\frac{\partial g}{\partial t}$). Then $Y_t := g(X_t, t)$ follows an Itô process with the *same* Wiener process W_t:

$$\mathrm{d}Y_t = \left(\frac{\partial g}{\partial x} a + \frac{\partial g}{\partial t} + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} b^2 \right) \mathrm{d}t + \frac{\partial g}{\partial x} b \; \mathrm{d}W_t \tag{1.44}$$

where the derivatives of g as well as the coefficient functions a and b in general depend on the arguments (X_t, t).

For a proof we refer to [Arn74], [Øk98], [Ste01], [Pro04]. Here we confine ourselves to the basic idea. When t varies by Δt, then X by $\Delta X = a \cdot \Delta t + b \cdot \Delta W$ and Y by $\Delta Y = g(X + \Delta X, t + \Delta t) - g(X, t)$. The Taylor expansion of ΔY begins with the linear part $\frac{\partial g}{\partial x} \Delta X + \frac{\partial g}{\partial t} \Delta t$, in which $\Delta X = a\Delta t + b\Delta W$ is substituted. The additional term with the derivative $\frac{\partial^2 g}{\partial x^2}$ is new and is introduced via the $O(\Delta x^2)$-term of the Taylor expansion,

$$\frac{1}{2} \frac{\partial^2 g}{\partial x^2} (\Delta X)^2 = \frac{1}{2} \frac{\partial^2 g}{\partial x^2} b^2 (\Delta W)^2 + \text{ t.h.o.}$$

Because of (1.28), $(\Delta W)^2 \approx \Delta t$, the leading term is also of the order $O(\Delta t)$ and belongs to the linear terms. Taking correct limits (similar as in Lemma 1.9) one obtains the integral equation represented by (1.44).

1.8.2 Consequences for Geometric Brownian Motion

Suppose the stock price follows a geometric Brownian motion, hence $X_t = S_t,\ a = \mu S_t,\ b = \sigma S_t$, for constant μ, σ. The value V_t of an option depends on S_t, $V_t = V(S_t, t)$. Assuming a $\mathcal{C}^2$-smooth value function V depending on S and t, we apply Itô's lemma. For $V(S,t)$ in the place of $g(x,t)$ the result is

$$\mathrm{d}V_t = \left(\frac{\partial V}{\partial S}\mu S_t + \frac{\partial V}{\partial t} + \frac{1}{2}\frac{\partial^2 V}{\partial S^2}\sigma^2 S_t^2\right)\mathrm{d}t + \frac{\partial V}{\partial S}\sigma S_t\,\mathrm{d}W_t\,. \tag{1.45}$$

This SDE is used to derive the Black–Scholes equation, see Appendix A4.

As second application of Itô's lemma consider $Y_t = \log(S_t)$, viz $g(x,t) := \log(x)$, for S_t solving GBM with constant μ, σ. Itô's lemma leads to the linear SDE

$$\mathrm{d}\log S_t = (\mu - \frac{1}{2}\sigma^2)\,\mathrm{d}t + \sigma \mathrm{d}W_t\,. \tag{1.46}$$

In view of (1.31) the solution is straightforward:

$$\begin{aligned} Y_t &= Y_{t_0} + (\mu - \frac{1}{2}\sigma^2)\int_{t_0}^t \mathrm{d}s + \sigma\int_{t_0}^t \mathrm{d}W_s \\ &= Y_{t_0} + (\mu - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0}) \end{aligned} \tag{1.47}$$

From the properties of the Wiener process W_t we conclude that Y_t is distributed normally. To write down the density function $\hat{f}(Y_t)$, the mean $\hat{\mu} := \mathsf{E}(Y_t)$ and the variance $\hat{\sigma}$ are needed. For this linear SDE (1.46) the expectation $\mathsf{E}(Y_t)$ satisfies the deterministic part

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(Y_t) = \mu - \frac{\sigma^2}{2}\,.$$

The solution of $\dot{y} = \mu - \frac{\sigma^2}{2}$ with initial condition $y(t_0) = y_0$ is

$$y(t) = y_0 + (\mu - \frac{\sigma^2}{2})(t - t_0)\,.$$

In other words, the expectation of the Itô process Y_t is

$$\hat{\mu} := \mathsf{E}(\log S_t) = \log S_0 + (\mu - \frac{\sigma^2}{2})(t - t_0)\,.$$

Analogously, we see from the differential equation for $\mathsf{E}(Y_t^2)$ (or from the analytic solution of the SDE for Y_t) that the variance of Y_t is $\sigma^2(t - t_0)$. In view of (1.46) the simple SDE for Y_t implies that the stochastic fluctuation of Y_t is that of σW_t, namely, $\hat{\sigma}^2 := \sigma^2(t - t_0)$. So, from (B1.9) with $\hat{\mu}$ and $\hat{\sigma}$, the density of Y_t is

$$\widehat{f}(Y_t) := \frac{1}{\sigma\sqrt{2\pi(t-t_0)}} \exp\left\{ -\frac{\left(Y_t - y_0 - \left(\mu - \frac{\sigma^2}{2}\right)(t-t_0)\right)^2}{2\sigma^2(t-t_0)} \right\}.$$

Back transformation using $Y = \log(S)$ and considering $\mathrm{d}Y = \frac{1}{S}\mathrm{d}S$ and $\widehat{f}(Y)\mathrm{d}Y = \frac{1}{S}\widehat{f}(\log S)\mathrm{d}S = f(S)\mathrm{d}S$ yields the density of $S_t > 0$:

$$f_{\mathrm{GBM}}(S, t-t_0;\, S_0, \mu, \sigma) := \frac{1}{S\sigma\sqrt{2\pi(t-t_0)}} \exp\left\{ -\frac{\left(\log(S/S_0) - \left(\mu - \frac{\sigma^2}{2}\right)(t-t_0)\right)^2}{2\sigma^2(t-t_0)} \right\} \tag{1.48}$$

This is the density of the *lognormal* distribution, conditional on $S_{t_0} = S_0$. It describes the probability of a transition

$$(S_0, t_0) \longrightarrow (S, t)$$

under the basic assumption that the stock price S_t follows a geometric Brownian motion (1.33). The distribution is skewed, see Figure 1.20. Now the skewed behavior coming out of the experiment reported in Figure 1.18 is clear. Notice that the parameters in Figures 1.18 and 1.20 match. Figure 1.18 is an approximation of the solid curve in Figure 1.20.

In summary, the assumption of GBM amounts to

$$S_t = S_0 \exp(Y_t)\,, \tag{1.49}$$

where the log-price Y_t is a Brownian motion with drift, $Y_t = (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$. — Having derived the density (1.48), we now can prove equation (1.8), with $\mu = r$ according to Remark 1.14 ($\longrightarrow$ Exercise 1.12). For vector-valued SDEs an appropriate version of the Itô lemma is (B2.1).

1.8.3 Integral Representation

An important application of a known density function is that it allows for an integral representation of European options. This will be revisited in Subsection 3.5.1, where we show for a European put under GBM

$$V(S_0, 0) = \mathrm{e}^{-rT} \int_0^\infty (K - S_T)^+ \, f_{\mathrm{GBM}}(S_T, T;\, S_0, r, \sigma)\, \mathrm{d}S_T\,. \tag{1.50}$$

Note the risk-free interest rate r as argument in the density. This reflects that the integral is the conditional expectation of the payoff under the assumed risk-neutral measure,

$$\mathsf{E}_{\mathsf{Q}} = \int_0^\infty \text{payoff} \cdot \text{density } \mathrm{d}S_T\,.$$

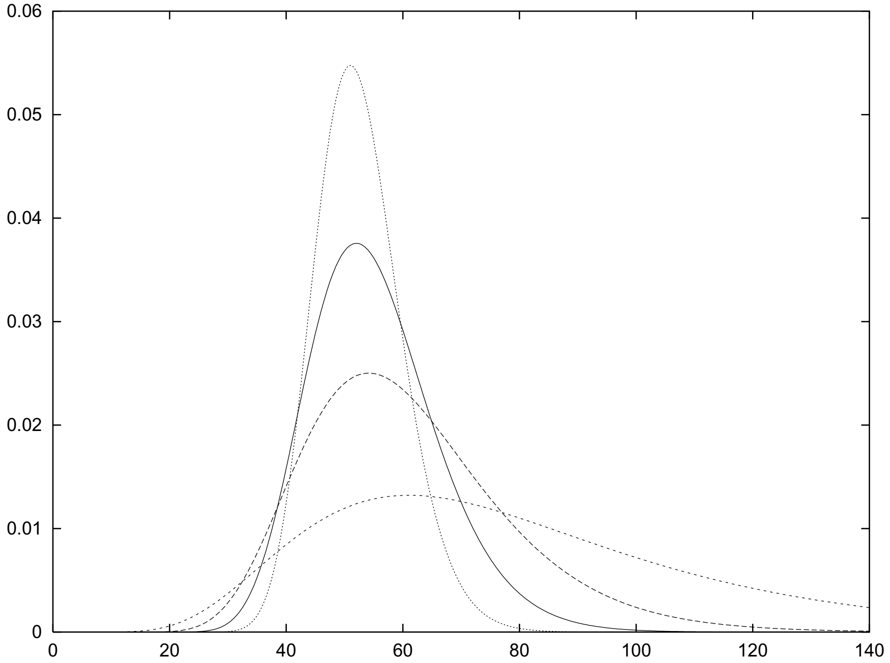

Fig. 1.20. Density (1.48) over S for $\mu = 0.1,\ \sigma = 0.2,\ S_0 = 50,\ t_0 = 0$ and $t = 0.5$ (dotted curve with steep gradient), $t = 1$ (solid curve), $t = 2$ (dashed) and $t = 5$ (dotted with flat gradient)

The integral representation for European-style options

$$V(S_0, 0) = \mathrm{e}^{-rT}\,\mathsf{E}_{\mathsf{Q}}(V(S_T, T) \mid S_t \text{ starting from } (S_0, 0))\,. \qquad (1.51)$$

holds for arbitrary payoff functions and density functions of a general class of valuation models.

1.8.4 Bermudan Options

The integral representation (1.50)/(1.51) for European options can be applied to approximate American options. To this end, discretize the time interval $0 \le t \le T$ into an equidistant grid of time instances t_i, similar as done for the binomial method of Section 1.4:

$$\Delta t := \frac{T}{M} \quad , \quad t_i := i\,\Delta t \quad (i = 0, \ldots, M)\,.$$

This defines lines in the (S, t)-domain, and cuts it into M slices. An option that restricts early exercise to specified discrete dates during its life is called a **Bermudan option**. The above slicing defines an artificial Bermudan option, constructed for the purpose of approximating the corresponding American option.

Let $V^{\mathrm{Ber}(M)}$ denote the value of the Bermudan option in the above setting of M slices of equal size. Clearly,

$$V^{\mathrm{Eur}} \le V^{\mathrm{Ber}(M)} \le V^{\mathrm{Am}} \quad \text{for all } M\,,$$

because of the additional exercise possibilities of an otherwise identical option. Note that the Bermudan options serve as lower bounds for the American option, and $V^{\mathrm{Eur}} = V^{\mathrm{Ber}(1)}$. One can show

$$\lim_{M\to\infty} V^{\mathrm{Ber}(M)} = V^{\mathrm{Am}}\,.$$

Hence, for suitable M the value $V^{\mathrm{Ber}(M)}$ can be used as approximation to V^{Am}.

Let us consider the time slice $t_i \le t \le t_{i+1}$ for any i. For the valuation of the option's value at t_i, the "inner payoff" is $V(S, t_{i+1})$ along the line $t = t_{i+1}$. Since a Bermudan option can not be exercised for $t_i < t < t_{i+1}$, its continuation value for t_i is given by the integral representation of a European option. This continuation value is

$$V^{\mathrm{cont}}(x, t_i) = \mathrm{e}^{-r(t_{i+1}-t_i)} \int_{-\infty}^{\infty} V(\xi, t_{i+1})\, f(\xi,\, t_{i+1} - t_i;\ x, \ldots)\, \mathrm{d}\xi \qquad (1.52\mathrm{a})$$

for arbitrary x. Here a value S at line $t = t_i$ is represented by x, and the price at t_{i+1} by ξ. The dots stand for the parameters of the risk-neutral evaluation of the chosen model, and f is its density conditional on $S_{t_i} = x$. For an n-factor model, the domain of integration is $\mathbb{R}^n$.

Since the Bermudan option can be exercised at t_i, its value is again given by the dynamic programming principle,

$$V(x, t_i) = \max\{\, \Psi(x),\ V^{\mathrm{cont}}(x, t_i)\,\}\,, \qquad (1.52\mathrm{b})$$

where Ψ denotes the payoff. Equations (1.52) define for $i = M-1, \ldots, 0$ a backward recursive algorithm. It starts from the given payoff at T, which provides $V(S, t_M)$. That is, only for the first time level $i = M-1$, the option is "vanilla," whereas for $i < M-1$ the inner payoffs are given by (1.52b).

In the algorithm, the evaluation of the integral in (1.52a) is done by numerical quadrature ($\longrightarrow$ Appendix C1), and the continuation value functions V^{cont} are approximated by interpolating functions $C(x)$ based on m nodes in x-space [Que07]. In the simplest case of a one-factor model ($n = 1$), the nodes may represent equidistantly chosen S_j $(1 \le j \le m)$. The inner payoffs are denoted g_i, and the Bermudan option is to be evaluated at $(x, 0) := (S, 0)$.

Algorithm 1.18 (Bermudan option)

set m nodes $x_1, \ldots, x_m \in \mathbb{R}^n$.

$g_M(x) := V(x, t_M) = V(x, T) = \Psi(x)$.

recursively backwards ($i = M-1, \ldots, 0$):

(1) input: g_{i+1}
loop ($j = 1, \ldots, m$): calculate by **quadrature**

$$q_j := \mathrm{e}^{-r(t_{i+1}-t_i)} \int g_{i+1}(\xi)\, f(\xi,\, t_{i+1}-t_i;\; x_j, \ldots)\, \mathrm{d}\xi$$

output: $q_1, \ldots, q_m$

(2) **interpolate** $(x_1, q_1), \ldots, (x_m, q_m)$. output: $C(x)$

(3) $g_i(x) := \max\{\Psi(x),\, C(x)\}$

The final $g_0(x)$ is the approximation of $V^{\mathrm{Ber}(M)}(x, 0)$, which in turn approximates $V^{\mathrm{Am}}(x, 0)$. The integral (1.52a) is taken over a suitably truncated interval $\xi_{\min} \le \xi \le \xi_{\max}$. The method works also for general non-GBM models, as long as they are not path-dependent. The order of convergence in Δt is linear. If necessary, the nodes x_j can be readjusted after each i; extrapolation is possible. For example, when two values $V^{\mathrm{Ber}(M)}(x, 0)$, $V^{\mathrm{Ber}(2M)}(x, 0)$ are available, an improved approximation is

$$\bar{V} = 2\, V^{\mathrm{Ber}(2M)}(x, 0) - V^{\mathrm{Ber}(M)}(x, 0)\,.$$

For details see [Que07].

1.8.5 Empirical Tests

It is inspiring to test the idealized Model 1.13 of a geometric Brownian motion against actual empirical data. Suppose the time series $S_1, ..., S_M$ represents consecutive quotations of a stock price. To test the data, histograms of the returns are helpful ($\longrightarrow$ Figure 1.21). The transformation $y = \log(S)$ is most practical. It leads to the notion of the *log return*, defined by[13]

$$R_{i,i-1} := \log \frac{S_i}{S_{i-1}}\,. \tag{1.53}$$

Let Δt be the equally spaced sampling time interval between the quotations S_{i-1} and S_i, measured in years. Then (1.48) leads to

$$R_{i,i-1} \sim \mathcal{N}((\mu - \frac{\sigma^2}{2})\Delta t\;,\; \sigma^2 \Delta t)\,.$$

Comparing with (1.34) we realize that the variances of the simple return and of the log return are identical. The sample variance $\sigma^2 \Delta t$ of the data allows to calculate estimates of the historical volatility σ ($\longrightarrow$ Exercise 1.13). But the shape of actual market histograms is usually not in good agreement with the well-known bell shape of the Gaussian density. The symmetry may

[13] Since $S_i = S_{i-1} \exp(R_{i,i-1})$, the log return is also called *continuously compounded return* in the ith time interval [Tsay02].

be perturbed, and in particular the tails of the data are not well modeled by the hypothesis of a geometric Brownian motion: The exponential decay expressed by (1.48) amounts to *thin tails*. This underestimates extreme events and hence hardly matches the reality of stock prices.

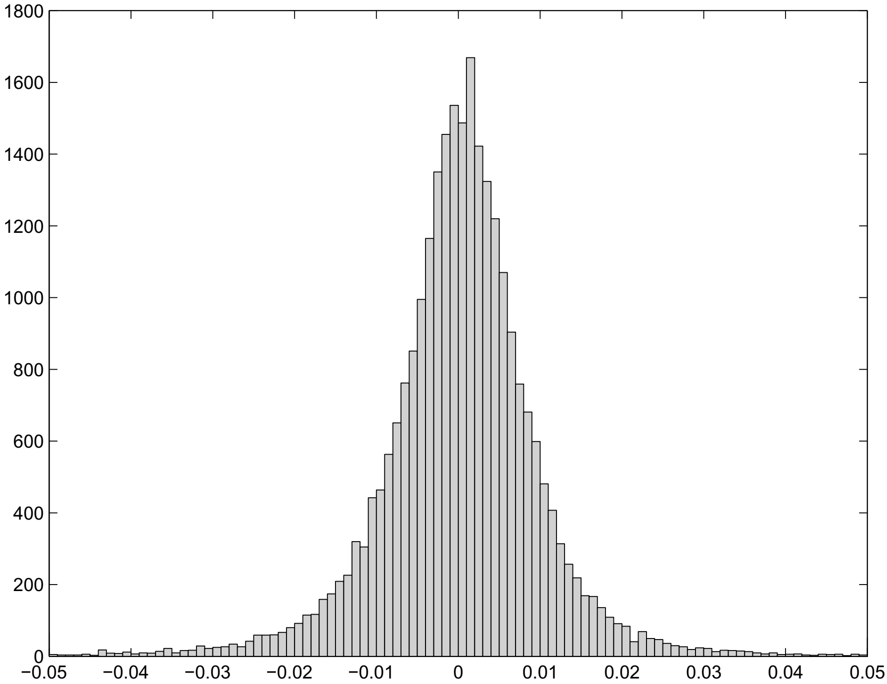

Fig. 1.21. Histogram (compare Exercise 1.13): frequency of daily log returns $R_{i,i-1}$ of the Dow in the time period 1901-1999.

We conclude this section by listing again the analytic solution of the basic linear constant-coefficient SDE (1.33)

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}W_t$$

of GBM. From (1.47) or (1.49), the process

$$S_t := S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right) \tag{1.54}$$

solves the linear constant-coefficient SDE (1.33). Equation (1.54) generalizes to the case of nonconstant coefficients ($\longrightarrow$ Exercise 1.18). As a consequence we note that $S_t > 0$ for all t, provided $S_0 > 0$.

1.9 Jump Models

The geometric Brownian motion Model 1.13 has continuous paths S_t. As noted before, the continuity is at variance with those rapid asset price movements that can be considered almost instantaneous. Such rapid changes can be modeled as jumps. This section introduces a basic building block of a jump process, namely, the Poisson process. Related simulations (like that of Figure 1.22) may look more authentic than continuous paths. But one has to pay a price: With a jump process the risk of an option in general can not be hedged away to zero. And calibration becomes more involved.

To define a Poisson process, denote the time instances for which a jump arrives τ_j, with

$$\tau_1 < \tau_2 < \tau_3 < \dots$$

Let the number of jumps be counted by the counting variable J_t, where

$$\tau_j = \inf\{t \geq 0 \ , \ J_t = j\} \, .$$

A Bernoulli experiment describes the probability that a jump occurs. For this local discussion and an arbitrary time instant t, consider n subintervals of length $\Delta t := \frac{t}{n}$ and allow for only two outcomes, jump *yes* or *no*, with the probabilities

$$\begin{array}{rcl} \mathsf{P}(J_t - J_{t-\Delta t} = 1) & = & \lambda \Delta t \\ \mathsf{P}(J_t - J_{t-\Delta t} = 0) & = & 1 - \lambda \Delta t \end{array} \tag{1.55}$$

for some λ such that $0 < \lambda \Delta t < 1$. The parameter λ is referred to as the *intensity* of this jump process. Consequently k jumps in $0 \leq \tau \leq t$ have the probability

$$\mathsf{P}(J_t - J_0 = k) = \binom{n}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} \, ,$$

where the trials in each subinterval are considered independent. A little reasoning reveals that for $n \to \infty$ this probability converges to

$$\frac{(\lambda t)^k}{k!} \mathrm{e}^{-\lambda t} \, ,$$

which is known as the Poisson distribution with parameter $\lambda > 0$ ($\longrightarrow$ Appendix B1). This leads to the Poisson process.

Definition 1.19 (Poisson process)

The stochastic process $\{J_t \ , \ t \geq 0\}$ is called Poisson process if the following conditions hold:
(a) $J_0 = 0$
(b) $J_t - J_s$ are integer-valued for $0 \leq s < t < \infty$ and

$$\mathsf{P}(J_t - J_s = k) = \frac{\lambda^k (t-s)^k}{k!} \, \mathrm{e}^{-\lambda(t-s)} \quad \text{for } k = 0, 1, 2 \dots$$

(c) The increments $J_{t_2} - J_{t_1}$ and $J_{t_4} - J_{t_3}$ are independent for all $0 \le t_1 < t_2 < t_3 < t_4$.

Several properties hold as consequence of this definition:

Properties 1.20 (Poisson process)

(d) J_t is right-continuous and nondecreasing.
(e) The times between successive jumps are independent and exponentially distributed with parameter λ. Thus,

$$\mathsf{P}(\tau_{j+1} - \tau_j > \Delta\tau) = \mathrm{e}^{-\lambda\Delta\tau} \quad \text{for each } \Delta\tau\,.$$

(f) J_t is a Markov process.
(g) $\mathsf{E}(J_t) = \lambda t$, $\mathsf{Var}(J_t) = \lambda t$

Simulating Jumps

Following the above introduction of Poisson processes, there are two possibilities to calculate jump instances τ_j such that the above probabilities are met. First, the equation (1.55) may be used together with uniform deviates ($\longrightarrow$ Chapter 2). In this way a Δt-discretization of a t-grid can be easily exploited by drawing a random number to decide whether a jump occurs in a subinterval. The other alternative is to calculate exponentially distributed random numbers $h_1, h_2, \ldots$ ($\longrightarrow$ Section 2.2.2) to simulate the intervals $\Delta\tau$ between consecutive jump instances, and set

$$\tau_{j+1} := \tau_j + h_j\,.$$

The expectation of the h_j is $\frac{1}{\lambda}$.

The unit amplitudes of the jumps of the Poisson counting process J_t are not relevant for the purpose of establishing a market model. The jump *sizes* of the price of a financial asset should be considered random. This requires —in addition to the arrival times τ_j— another random variable.

Let the random variable S_t jump at τ_j, and denote τ^+ the (infinitesimal) instant immediately after the jump, and τ^- the moment before. Then the absolute size of the jump is

$$\Delta S = S_{\tau^+} - S_{\tau^-}\,,$$

which we model as a *proportional jump*,

$$S_{\tau^+} = qS_{\tau^-} \quad \text{with } q > 0\,. \tag{1.56}$$

So, $\Delta S = qS_{\tau^-} - S_{\tau^-} = (q-1)S_{\tau^-}$. The jump sizes equal $q - 1$ times the current asset price. Accordingly, this model of a jump process depends on a random variable q_t and is written

$$\mathrm{d}S_t = (q_t - 1)S_{t^-}\,\mathrm{d}J_t\,, \text{ where } J_t \text{ is a Poisson process.}$$

We assume that $q_{\tau_1}, q_{\tau_2}, ...$ are i.i.d. The resulting process with the two involved processes J_t, q_t is called **compound Poisson process**.

Jump Diffusion

Next we superimpose the jump process to stochastic diffusion, here to GBM. The combined geometric Brownian and compound Poisson process is given by

$$\boxed{\mathrm{d}S_t = S_{t^-}\,(\,\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t + (q_t - 1)\,\mathrm{d}J_t\,)\ .} \tag{1.57}$$

Here σ is the same as for the GBM, hence conditional on no jump. Such a combined model represented by (1.57) is called **jump-diffusion process**. It involves three different stochastic driving processes, namely, W_t, J_t, and q_t. We assume that J, q, W are independent of one another. Figure 1.22 shows a simulation of the SDE (1.57).

An analytic solution of (1.57) can be calculated on each of the jump-free subintervals $\tau_j < t < \tau_{j+1}$ where the SDE is just the GBM diffusion $\mathrm{d}S = S(\mu\mathrm{d}t + \sigma\mathrm{d}W)$. For example, in the first subinterval until τ_1, the solution is given by (1.54). At τ_1 a jump of the size

$$(\Delta S)_1 := (q_{\tau_1} - 1)S_{\tau_1^-}$$

occurs, and thereafter the solution continues with

$$S_t = S_0 \cdot \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) + (q_{\tau_1} - 1)S_{\tau_1^-}\ ,$$

until τ_2. The interchange of continuous parts and jumps proceeds in this way, all jumps are added. So the SDE can be written as

$$S_t = S_0 + \int_0^t S_s(\mu\mathrm{d}s + \sigma\mathrm{d}W_s) + \sum_{j=1}^{J_t} S_{\tau_j^-}(q_{\tau_j} - 1)\,, \tag{1.58}$$

or

$$S_t = S_0 \exp\left(\left((\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) \cdot \prod_{j=1}^{J_t} q_j\ .$$

This is the model based on Merton's paper [Mer76]. The equation (1.58) can be rewritten in the log-framework, with $Y_t := \log S_t$. The log-jump sizes according to model (1.56) are

$$\begin{aligned}(\Delta Y)_\tau :&= Y_{\tau^+} - Y_{\tau^-} = \log(qS_{\tau^-}) - \log S_{\tau^-} \\ &= \log q_\tau\ .\end{aligned}$$

Following (1.54), the model can be written

$$Y_t = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t + \sum_{j=1}^{J_t} (\Delta Y)_{\tau_j} \tag{1.59}$$

—that is the sum of a drift term, a Brownian motion, and a jump process. The summation term $\sum(\Delta Y)$ in (1.59) is the compound process. Merton assumes normally distributed ΔY, which amounts to lognormal q. In summary we emphasize again that the jump-diffusion process has three driving processes, namely, W, J, and q. As in the GBM case, see (1.49)/(1.54), the price process is of the form $S_t = S_0 \exp(Y_t)$.

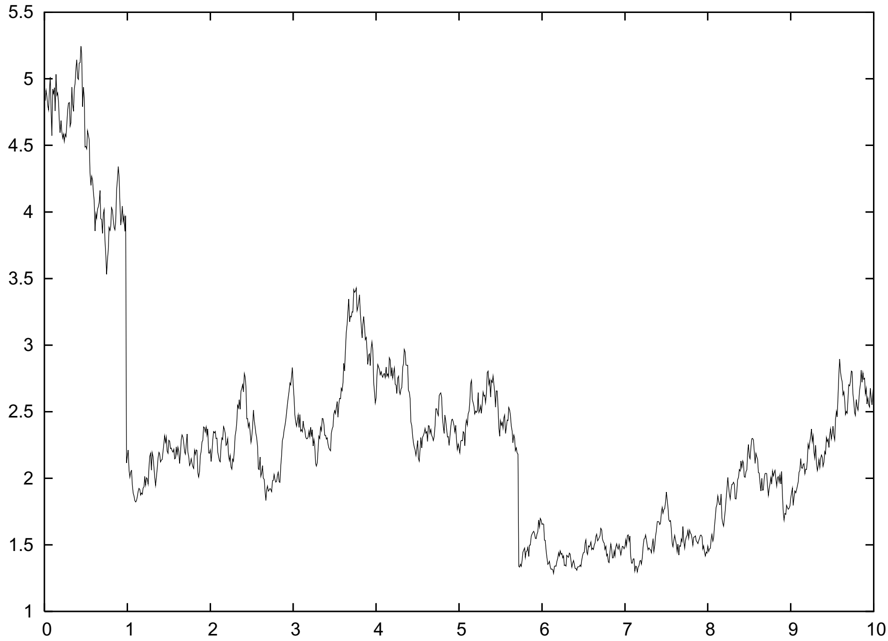

Fig. 1.22. Example 1.21: Sample path S_t of (1.57); jump report in Table 1.3

Example 1.21 (jump-diffusion)

Here we assume an interest rate $r = 0.06$, and a process S_t following (1.57) with diffusion volatility $\sigma = 0.3$. For a hypothetical crash modeling, let us assume Poisson jumps with an intensity rate $\lambda = 0.2$, which means that on the average one jump occurs every 5 years. Following Merton's model, we take $\log(q) \sim \mathcal{N}(\mu_\mathrm{J}, \sigma_\mathrm{J}^2)$, and choose $\mu_\mathrm{J} = -0.3$ and $\sigma_\mathrm{J} = 0.4$. To get random numbers with distribution $\sim \mathcal{N}(\mu_\mathrm{J}, \sigma_\mathrm{J}^2)$, we calculate random numbers $Z \sim \mathcal{N}(0, 1)$ (Chapter 2), and set $\log q = \sigma_\mathrm{J} Z + \mu_\mathrm{J}$. The chosen value of μ_J corresponds to a mean $q = \exp(\mu_\mathrm{J}) = 0.7408$, which amounts to an average 26% drop in S_τ at a jump instant τ. For the integration of

(1.57), a growth rate is chosen such that risk neutrality is achieved. As will be explained in Section 7.3, the martingale property is satisfied with

$$\mu = r - \lambda\,(\exp[\mu_{\mathrm{J}} + \tfrac{1}{2}\sigma_{\mathrm{J}}^2] - 1)\,,$$

which for our numbers gives the growth rate 0.0995. This rate μ is larger than r, and —roughly speaking— compensates for the tendency that in case $\mu_{\mathrm{J}} < 0$ down jumps are more likely than up jumps. Now we are ready to solve (1.57) numerically. In Figure 1.22 we show one calculated trajectory. We see three jumps, with data in Table 1.3. In this particular simulation, there are two heavy down jumps within the time interval $0 \le t \le 10$, which are clearly visible in Figure 1.22.

Table 1.3 Jumps in Figure 1.22

τ	$\log(q)$	q	jump
0.99	−0.642	0.526	47% down
4.76	0.0495	1.05	5% up
5.72	−0.534	0.586	41% down

The task of valuing options leads to a partial integro-differential equation (A4.14), shown in Appendix A4, and in Section 7.3.

The above jump-diffusion process is not the only jump process used in finance. There are also processes with an infinite number of jumps in finite time intervals. To model such processes, building blocks are provided by a more general class of jump processes, namely, the Lévy processes. Simply speaking, think of relaxing the properties (b), (d) of Definition 1.7 of a Wiener process such that non-normal distributions and jumps are permitted. Consult Section 7.3 for some basics on Lévy processes.

1.10 Calibration

Which model should be chosen for a particular application?

This is a truly fundamental question. The question involves two views, namely, a qualitative and a quantitative aspect.

When one speaks of a “model,” the focus is on its quality. This refers to the structure and the type of equation. Important ingredients of a model are, for example, a diffusion term, a jump feature, a specific nonlinearity, or whether the volatility is considered as a constant or a stochastic process. Ideally, the model and its equations represent economical laws. On the other hand, the quantitative aspect of the model consists in the choice of specific numbers for the coefficients or parameters of the model. “Modeling” refers to the setup of a chosen equation, and "calibration" is the process of matching the parameters of the chosen model to the data that represent reality.

The distinction between modeling and calibration is not always obvious. For example, consider the class of mean-reversion models represented by (1.40). There is the exponent β in the factor r_t^β. This exponent β can be regarded either as parameter, or as a structural element of the model. The three cases

$$\begin{array}{rcl} \beta = 0 & : & \text{the factor is unity, } r^\beta = 1 \text{, it "disappears,"} \\ \beta = 1 & : & \text{the factor is linear, it represents a proportionality,} \\ \beta = 1/2 & : & \text{the factor } \sqrt{r} \text{ is a specific nonlinearity,} \end{array}$$

point at the qualitative aspect of this specific parameter. Typically, modeling sets forth some argument why a certain parameter is preset in a specific way, and not subjected to calibration. Modeling places emphasis on capturing market behavior rather than the peculiarities of a given data set.

Let us denote N parameters to be calibrated by $c_1, \ldots, c_N$. Examples are the volatility σ in GBM (1.33), or α, R for the mean-reversion term in (1.40), or the jump intensity λ of a jump-diffusion process. For the mean-reverting volatility tandem of Example 1.15, the vector to calibrate consists of five parameters,

$$c = (\alpha, \beta, \rho, \sigma_0, \zeta_0)\,.$$

Here ρ is the correlation between the two Wiener processes $W^{(1)}, W^{(2)}$, and σ_0, ζ_0 are the initial values for the processes σ_t, ζ_t. For the volatility tandem it makes sense to assume $\zeta_0 = \sigma_0$, which cuts down the calibration dimension N from five to four. The initial stock price S_0 is known. The interest rates r that match a maturity T are obtained, for example, from EURIBOR, and are not object of the calibration. Any attempt to cut down the calibration dimension N is welcome because the costs of calibration are significant.

Suppose an initial guess of the calibration vector c. Then the calibration procedure is based on the three steps

(1) simulate the model —that is, solve it numerically,
(2) compare the calculated results with the market data —that is, calculate the defect, and
(3) adapt c such that the model better matches the data —that is, the defect should decrease.

These three steps are repeated iteratively. How to perform step (3) is not obvious; there is no unique way how to decrease the defect. A standard approach is to minimize the defect in a least-squares fashion.

In our context of calibrating models for finance, data of vanilla options are available as follows: The price S of the underlying is known as well as market prices V^{mar} for several strikes K and maturities T. Let the option prices V^{mar} be observed for M pairs $(T_1, K_1), \ldots, (T_M, K_M)$. That is, the available data are

$$S, (T_k, K_k, V_k^{\text{mar}}),\ k = 1, \ldots, M\,.$$

For definiteness of the calibration require sufficiently many data in the sense $M \geq N$. Raw data may be subjected to a smoothing process [GlH10].

First, a model is specified. Then, in step (1), the chosen model is evaluated for each of the M data (S, T_k, K_k), which gives model prices $V(S; 0;\, T_k, K_k;\, c)$. In general, this valuation process is expensive. An excellent approach for the simultaneous valuation of a large number of European options is the FFT method of Carr and Madan [CaM99], see Section 7.4. In step (2), the result of the valuation is compared to the market prices. There will be a defect. Therefore, in step (3), an iteration is set up to improve the current fit c. The least-squares approach is to minimize the sum of the squares of all defects, over all c,

$$\min_c \sum_{k=1}^{M} (V_k^{\text{mar}} - V(S, 0; T_k, K_k; c))^2\,. \tag{1.60}$$

The sum in (1.60) is a function of c and can be visualized as a surface over the parameter c-space. It can be modified by weighting the terms appropriately. Finally, the calibration results in a minimizing c ($\longrightarrow$ Appendix C4). In view of the data error, it hardly makes sense to calculate the minimizing parameter vector c with high accuracy.

A simple example is provided by the implied volatility, see Exercise 1.5. Here $N = 1$, $M = 1$, $c := \sigma$, and it is possible to make the defect vanish — the minimum in (1.60) becomes zero. But in general the minimum of (1.60) will be a positive value. It is tempting to regard this value as a measure of the discrepancy or defect of the chosen model. But this would be misleading; we come back to this below.

As a numerical example, we calibrate two models on the same data set of standard European calls on the DAX index observed in the time period January 2002 through September 2005. For this example, the calibration of Heston's model (1.43) results in the five parameters

$$\kappa = 1.63\,,\ \theta = 0.0934\,,\ \sigma_{\text{v}} = 0.473\,,\ v_0 = 0.0821\,,\ \rho = -0.8021\,,$$

with $\mu = r$ for the risk-neutrality. This parameter set matches the criterion $2\kappa\theta \geq \sigma_{\text{v}}$ which guarantees $v > 0$. — The same data are applied to calibrate the Black-Scholes model: The data are matched by GBM with the constant $\sigma = 0.239$ (from [End08]). This is comparable to the calibration of the Heston model with its $\sqrt{v_0} \approx 0.28$.

So far, we have not come close to an answer to the initial question on the "best" choice of an appropriate model. An attempt to decide on the quality of a model might be to compare the defects. For instance, compare the values of the sums in (1.60). In the above experiment, Heston's model has the smaller defect; the defect of the Black–Scholes model is five times as large.

One might think that one model is better than another one, when the discrepancy is smaller. But this is a wrong conclusion! Admitting a large enough number of parameters enables to reach a seemingly best fit with a small discrepancy. The danger with a large number of parameters is *overfitting*. Overfitting can be detected as follows: Divide the data into halves, fit the model on the one half (*in-sample fit*), and then test the quality of the fit on the other half of the data (*out-of-sample fit*). In case the out-of-sample fit matches the data much worse than the in-sample fit, we have a strong clue on overfitting. Then any predictive power of the model may be lost. A vanishing defect might be seen as hint of the model being useless. Overfitting is related to the *stability of parameters*. If the parameters c change drastically when exchanging one data set by a similar data set, then the model is considered unstable. In order to obtain information on the parameter uncertainty, the discrepancy must be analyzed more closely around the calculated best fit c. The defect function (1.60) can exhibit a large flat region. Then significantly different values of c yield a similar error. In this sense, a calibration problem can be ill-posed [He06].

There is another test of the quality of a model, namely, how well hedging works. A hedging strategy based on the model is compared to the reality of the data. Empirical tests and comparisons in [Dah10], [End08] suggest that in the context of option pricing, a stochastic volatility may be a more basic ingredient of a good model than jump processes are. In terms of stability, out-of-sample fitting, and hedging of options, Heston's model (Example 1.16) is recommendable — these conclusions have been based on the prices of European options on the DAX 2002–2005. In terms of hedging capabilities, the classical Black–Scholes model is competitive.

To summarize, it is obvious that calibration is a formidable task, in particular if several parameters are to be fitted. The attainable level of calibration quality depends on the chosen model. In case the structure of the equation is not designed properly, an attempt to improve parameters may be futile. For a given model, it might well happen that a perfect calibration is never found. It is unlikely that some model eventually might emerge as generally "most recommendable." Calibration does not remove the risk of having chosen the wrong model. With our focus on computational tools, it does make sense to consider the classical Black–Scholes model as a benchmark. It captures a significant part of the essence of option markets.

Notes and Comments

on Section 1.1:

This section presents a brief introduction to standard options. For more comprehensive studies of financial derivatives we refer, for example, to [CoR85], [WiDH96], [Hull00]. Mathematical detail can be found in [LaL96], [MuR97], [KaS98], [Shi99], [Epps00], [Ste01]. Other books on financial markets include [ElK99], [Gem00], [MeVN02], [DaJ03]. (All hints on the literature are examples; an extensive overview on the many good books in this rapidly developing field is hardly possible.)

on Section 1.2:

Black, Merton and Scholes developed their approaches concurrently, with basic papers in 1973 ([BlS73], [Mer73]; compare also [Mer90]). Merton and Scholes were awarded the Nobel Prize in economics in 1997. (Black had died in 1995.) One of the results of these authors is the so-called Black–Scholes equation (1.2) with its analytic solution formula (A4.10). For reference on discrete-time models, consult [Pli97], [FöS02]. Transaction costs and market illiquidity or feedback effects are discussed in Section 7.1.

on Section 1.3:

References on specific numerical methods are given where appropriate. As computational finance is concerned, most quotations refer to research papers. Other general text books discussing computational issues include [WiDH96], [Hig04], [AcP05]; further hints can be found in [RoT97]. For the calculation of the sample variance (Exercise 1.4) see [ChGL83], [Hig96].

on Section 1.4:

Binomial or trinomial methods are sometimes found under the heading *tree methods* or *lattice methods*. Basic versions of the binomial method were introduced in 1979 by [CoRR79][14] and [ReB79]. [CoRR79] suggested

$$u := \mathrm{e}^{\sigma\sqrt{\Delta t}},\ d := \mathrm{e}^{-\sigma\sqrt{\Delta t}},\ \tilde{p} := \frac{1}{2}(1 + \frac{r}{\sigma}\sqrt{\Delta t}), \qquad \text{(CRR)}$$

where $\tilde{p}$ is a first-order approximation to the p of (1.6) (the reader may check). The influential paper by Cox, Ross and Rubinstein has coined the name CRR for their approach. [HuW88] pointed out that (1.11) is slightly more correct than the CRR choice. [ReB79] suggested the choice $p = \frac{1}{2}$, which leads to values of u and d ($\longrightarrow$ Exercise 1.21). Of course, another set of parameters u, d, p leads to a different approximation. Example 1.6, which is from [Hull00], and $M = 100$ yields $V = 4.28041$ with the parameter set

[14] William Sharpe has been credited for suggesting the advantages of the discrete-time approach.

(1.11), and $V = 4.27806$ with u, d from (CRR). But for $M \to \infty$ convergence is maintained in either case. — The dynamic programming principle is due to [Bel57]. In the literature, the result of the dynamic programming procedure is often listed under the name Snell envelope.

The Table 1.2 might suggest that it is easy to obtain high accuracy with binomial methods. This is not the case; flaws were observed in particular close to the early-exercise curve [CoLV02]. As illustrated by Figure 1.10, the described standard version wastes many nodes $S_{j,i}$ close to zero and far away from the strike region even for small M.

For advanced binomial methods and for speeding up convergence, consult also [Bre91], [LeR96], [Lei99], [Kla01]. [FiG99] insert a patch of higher resolution close to $(S, t) = (K, T)$ into the trinomial tree. The resulting adaptive mesh model exhibits higher accuracy. In order to maintain accuracy for barrier options one takes care that layers coincide with the barrier, see for instance [DaL10]. For a detailed account of the binomial method see also [CoR85]. By correcting the terminal probabilities, which come out of the binomial distribution ($\longrightarrow$ Exercise 1.8), it is possible to adjust the tree to actual market data [Rub94a], see also the *implied tree* of [DeK94], outlined also in [Sey12]. [HoP02] explains how to implement the binomial method in spreadsheets. Many applications of binomial trees are found in [Lyuu02].

on Section 1.5:

When we expect Δ to be positive, then we should assume the option is a call. But the argumentation is the same for a put, then $\Delta < 0$. As shown in Section 1.5, a valuation of options based on a hedging strategy is equivalent to the risk-neutral valuation described in Section 1.4. Another equivalent valuation is obtained by a *replication* portfolio. This basically amounts to including the risk-free investment, to which the hedged portfolio of Section 1.5 was compared, into the portfolio. To this end, the replication portfolio includes a bond with the initial value $B_0 := -(\Delta \cdot S_0 - V_0) = -\Pi_0$ and interest rate r. The portfolio consists of the bond and Δ shares of the asset. At the end of the period T the final value of the portfolio is $\Delta \cdot S_T + \mathrm{e}^{rT}(V_0 - \Delta \cdot S_0)$. The hedge parameter Δ and V_0 are determined such that the value of the portfolio is V_T, independent of the price evolution. By adjusting B_0 and Δ in the right proportion we are able to replicate the option position. This strategy is *self-financing*: No initial net investment is required. The result of the self-financing strategy with the replicating portfolio is the same as what was derived in Section 1.5. The reader may like to check this. For the continuous-time case, see Appendix A4.

Frequently discounting is done with the factor $(1 + r \cdot \Delta t)^{-1}$. This r would not be a continuously compounding interest rate. Our $\mathrm{e}^{-r\Delta t}$ or e^{-rT} is consistent with the approach of Black, Merton and Scholes. For references on risk-neutral valuation we mention [Hull00], [MuR97], [Kwok98] and [Shr04].

on Section 1.6:

Introductions into stochastic processes and further hints on advanced literature can be found in [Doob53], [Fre71], [Arn74], [Bil79], [ReY91], [KlP92], [Shi99], [Sato99], [Shr04]. In the literature, the terms Wiener process and Brownian motion are often used as synonyms, and the modifier "standard" is used to specialize on the drift-free case. Here we follow the convention as in Definition 1.7, where the term Wiener process is mostly reserved for the "standard" scalar drift-free Brownian motion. The definition of a Wiener process depends on the underlying probability measure P, which enters through the definition of independence, and by its distribution being Gaussian, see (B1.1). For more hints on martingales, see Appendix B2. Algorithm 1.8 is also called "Gaussian random walk."

For a proof of the nondifferentiability of Wiener processes, see [HuK00]. In contrast to the results for Wiener processes, differentiable functions W_t satisfy for $\delta_N \to 0$

$$\sum |W_{t_j} - W_{t_{j-1}}| \longrightarrow \int |W'_s| \,\mathrm{d}s\,, \quad \sum (W_{t_j} - W_{t_{j-1}})^2 \longrightarrow 0\,.$$

The Itô integral and the alternative Stratonovich integral are explained in [Doob53], [Arn74], [ChW83], [ReY91], [KaS91], [KlP92], [Mik98], [Øk98], [Sch80], [Shr04]. The class of (Itô-)stochastically integrable functions is characterized by the properties $f(t)$ is $\mathcal{F}_t$ adapted and $\mathsf{E}\int f(s)^2 \mathrm{d}s < \infty$. We assume that all integrals occurring in the text exist. The integrator W_t needs not be a Wiener process. The stochastic integral can be extended to semimartingales [HuK00].

on Section 1.7:

The Algorithm 1.11 is sometimes named after Euler and Maruyama.

The general linear SDE is of the form

$$\mathrm{d}X_t = (a_1(t)X_t + a_2(t))\,\mathrm{d}t + (b_1(t)X_t + b_2(t))\,\mathrm{d}W_t\,.$$

The expectation $\mathsf{E}(X_t)$ of a solution process X_t of a linear SDE satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(X_t) = a_1 \mathsf{E}(X_t) + a_2\,,$$

and for $\mathsf{E}(X_t^2)$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{E}(X_t^2) = (2a_1 + b_1^2)\mathsf{E}(X_t^2) + 2(a_2 + b_1 b_2)\mathsf{E}(X_t) + b_2^2\,.$$

This is obtained by taking the expectation of the SDEs for X_t and X_t^2, the latter one derived by Itô's lemma [KlP92], [Mik98]. Combining both differential equations allows to calculate the variance. [KlP92] in Section 4.4 gives a list of SDEs that are analytically solvable or reducible.

A process (1.33) with variable $\mu(t), \sigma(t)$ is called generalized GBM [Shr04]. For CIR of Example 1.16, provided $r_0 > 0$, $R > 0$, and a strong enough upward drift in the sense

$$\alpha R \geq \tfrac{1}{2}\sigma_{\mathrm{r}}^2 \,,$$

the solution of (1.40) satisfies $r_t > 0$ for all t; this criterion is attributed to Feller. For a PDE, the Feller condition is replaced by a boundary condition at $r = 0$ [EkLT09]. Based on the CIR system and a dependent variable $u(S, v, t)$ a two-dimensional PDE is presented in [Hes93], see Example 5.7.

The model of a geometric Brownian motion of equation (1.33) is the classical model describing the dynamics of stock prices. It goes back to Samuelson (1965; Nobel Prize in economics in 1970). Already in 1900 Bachelier had suggested to model stock prices with Brownian motion. Bachelier used the arithmetic version, which can be characterized by replacing the left-hand side of (1.33) by the absolute change $\mathrm{d}S$. This amounts to the process of the drifting Brownian motion $S_t = S_0 + \mu t + \sigma W_t$. Here even the theoretical stock price can become negative. Main advantages of the geometric Brownian motion are its exponential growth or decay, the success of the approaches of Black, Merton and Scholes, which is based on that motion, and the existence of moments (as the expectation). For positive S, the form (1.33) of GBM is not as restrictive as it might seem, see Exercise 1.18. A variable volatility $\sigma(S, t)$ is called *local volatility.* Such a volatility can be used to make the Black–Scholes model compatible with observed market prices [Dup94].

on Section 1.8:

The Itô lemma is also called Doeblin-Itô formula, after the early manuscript [Doe40] was disclosed. The Algorithm 1.18 was suggested by [Que07], including the use of radial basis functions, a tricky control of truncation errors, and a convergence analysis. The approximation quality of American options is quite satisfactory even for small values of M.

In view of their continuity, GBM processes are not appropriate to model jumps, which are characteristic for the evolution of stock prices. Jumps lead to relatively *heavy tails* in the distribution of empirical returns (see Figure 1.21)[15]. As already mentioned, the tails of the lognormal distribution are too thin. Other distributions match empirical data better. One example is the Pareto distribution, which has tails behaving like $x^{-\alpha}$ for large x and a constant $\alpha > 0$. A correct modeling of the tails is an integral basis for *value at risk* (VaR) calculations. For the risk aspect consult [EmKM97], [BaN97], [Dowd98], [ArDEH99], and the survey [EbFKO07]. For distributions that match empirical data see [EbK95], [Shi99], [BoP00], [MaRGS00], [BrTT00]. Estimates of future values of the volatility are obtained by (G)ARCH methods, which work with different weights of the returns [Shi99], [Hull00],

[15] The thickness is measured by the *kurtosis* $\mathsf{E}((X-\mu)^4)/\sigma^4$. The normal distribution has kurtosis 3. So the *excess kurtosis* is the difference to 3. Frequently, data of returns are characterized by large values of excess kurtosis.

[Tsay02], [FrHH04], [Rup04]. Promising are models of behavioral finance that consider the market as *dynamical system* [Lux98], [BrH98], [ChDG00], [BiV00], [MaCFR00], [Sta01], [DiBG01], [BiS06]. These systems experience the nonlinear phenomena *bifurcation* and *chaos*, which require again numerical methods. Such methods exist, and are explained elsewhere [Sey10].

on Section 1.9:

Section 1.9 concentrates on Merton's jump-diffusion process. For building Lévy models we refer to [Sato99], [ConT04], and Section 7.3.

on Section 1.10:

The CIR-based Heston model can be extended to jump-diffusion. This can be applied to both processes S_t and v_t in (1.43), which defines a general class of models with 10 parameters [DuPS00]. But applying jumps only for S_t , one obtains the same quality with eight parameters [Bat96]. Also the OU-based Schöbel–Zhu model is recommendable [ScZ99]. Another FFT based valuation approach is [FeO08]. Artificial smoothing of the least-squares function (1.60) allows to apply gradient-based methods. This is discussed in [KaMS09]. For hedging issues and practical aspects, consult [Jos03].

Exercises

Exercise 1.1 Put-Call Parity

Consider a portfolio consisting of three positions related to the same asset, namely, one share (price S), one European put (value V_{P}), plus a short position of one European call (value V_{C}). Put and call have the same expiration date T, and no dividends are paid.

a) Assume a no-arbitrage market without transaction costs. Show

$$S + V_{\mathrm{P}} - V_{\mathrm{C}} = K\mathrm{e}^{-r(T-t)}$$

for all t, where K is the strike and r the risk-free interest rate.

b) Use the put-call parity to realize

$$V_{\mathrm{C}}(S,t) \geq S - K\mathrm{e}^{-r(T-t)}$$
$$V_{\mathrm{P}}(S,t) \geq K\mathrm{e}^{-r(T-t)} - S\,.$$

Exercise 1.2 Transforming the Black–Scholes Equation

Show that the Black–Scholes equation (1.2)

$$\frac{\partial V}{\partial t} + \frac{\sigma^2}{2}S^2\frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

for $V(S,t)$ with constant σ and r is equivalent to the equation

$$\frac{\partial y}{\partial \tau} = \frac{\partial^2 y}{\partial x^2}$$

for $y(x,\tau)$. For proving this, you may proceed as follows:

a) Use the transformation $S = K\mathrm{e}^x$ and a suitable transformation $t \leftrightarrow \tau$ to show that (1.2) is equivalent to

$$-\dot{V} + V'' + \alpha V' + \beta V = 0$$

with $\dot{V} = \frac{\partial V}{\partial \tau}$, $V' = \frac{\partial V}{\partial x}$, α, β depending on r and σ.

b) The next step is to apply a transformation of the type

$$V = K \exp(\gamma x + \delta \tau)\, y(x,\tau)$$

for suitable γ, δ.

c) Transform the terminal condition of the Black–Scholes equation accordingly.

Exercise 1.3 Standard Normal Distribution Function

Establish an algorithm to calculate

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{t^2}{2})\,\mathrm{d}t\,.$$

Hint: Construct an algorithm to calculate the *error function*

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\,\mathrm{d}t$$

and use $\mathrm{erf}(x)$ to calculate $F(x)$. Use quadrature methods ($\longrightarrow$ Appendix C1).

Exercise 1.4 Calculating the Sample Variance

An estimate of the variance of M numbers $x_1, ..., x_M$ is

$$s_M^2 := \frac{1}{M-1} \sum_{i=1}^{M} (x_i - \bar{x})^2, \quad \text{with} \quad \bar{x} := \frac{1}{M} \sum_{i=1}^{M} x_i$$

The alternative formula

$$s_M^2 = \frac{1}{M-1} \left(\sum_{i=1}^{M} x_i^2 - \frac{1}{M} \left(\sum_{i=1}^{M} x_i \right)^2 \right) \qquad (\diamond)$$

can be evaluated with only one loop $i = 1, ..., M$, but should be avoided because of the danger of cancellation. The following single-loop algorithm is recommended instead of $(\diamond)$:

$$\begin{aligned}
&\alpha_1 := x_1,\ \beta_1 := 0\\
&\textit{for}\ \ i = 2, ..., M:\\
&\alpha_i := \alpha_{i-1} + \frac{x_i - \alpha_{i-1}}{i}\\
&\beta_i := \beta_{i-1} + \frac{(i-1)(x_i - \alpha_{i-1})^2}{i}
\end{aligned}$$

a) Show $\bar{x} = \alpha_M,\ s_M^2 = \frac{\beta_M}{M-1}$.
b) For the ith *update* in the algorithm carry out a rounding error analysis. What is your judgment on the algorithm?

Exercise 1.5 Implied Volatility

For European options we take the valuation formula of Black and Scholes of the type $V = v(S, \tau, K, r, \sigma)$, where τ denotes the time to maturity, $\tau := T - t$. For the definition of the function v see Appendix A4, equation (A4.10). If actual market data V^{mar} of the price are known, then one of the parameters considered known so far can be viewed as unknown and fixed via the implicit equation

$$V^{\text{mar}} - v(S, \tau, K, r, \sigma) = 0\,. \qquad (*)$$

In this calibration approach the unknown parameter is calculated iteratively as solution of equation $(*)$. Consider σ to be in the role of the unknown parameter. The volatility σ determined in this way is called *implied volatility* and is zero of $f(\sigma) := V^{\text{mar}} - v(S, \tau, K, r, \sigma)$.

Assignment:
a) Implement the evaluation of V_{C} and V_{P} according to (A4.10).
b) Design, implement and test an algorithm to calculate the implied volatility of a call. Use Newton's method to construct a sequence $x_k \to \sigma$. The derivative $f'(x_k)$ can be approximated by the difference quotient

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}\,.$$

For the resulting *secant iteration* invent a stopping criterion that requires smallness of both $|f(x_k)|$ and $|x_k - x_{k-1}|$.
c) Calculate the implied volatilities for the data

$$T - t = 0.211\ ,\ S_0 = 5290.36\,,\ r = 0.0328$$

and the pairs K, V from Table 1.4 (for more data see `www.compfin.de`). For each calculated value of σ enter the point (K, σ) into a figure, joining the points with straight lines. (You will notice a convex shape of the curve. This shape has lead to call this phenomenon *volatility smile*.)

Table 1.4. Calls on the DAX on Jan 4th 1999

K	6000	6200	6300	6350	6400	6600	6800
V	80.2	47.1	35.9	31.3	27.7	16.6	11.4

Exercise 1.6 Price Evolution for the Binomial Method

For β from (1.11) and $u = \beta + \sqrt{\beta^2 - 1}$ show

$$u = \exp\left(\sigma\sqrt{\Delta t}\right) + O\left(\sqrt{(\Delta t)^3}\right) .$$

Exercise 1.7 Implementing the Binomial Method

Design and implement an algorithm for calculating the value $V^{(M)}$ of a European or American option. Use the basic version of Algorithm 1.4.

INPUT: r (interest rate), σ (volatility), T (time to expiration in years), K (strike price), S (price of asset), and the choices *put* or *call*, and *European* or *American*.

Control the mesh size $\Delta t = T/M$ adaptively. For example, calculate V for $M = 8$ and $M = 16$ and in case of a significant change in V use $M = 32$ and possibly $M = 64$.

Test examples:
a) put, European, $r = 0.06$, $\sigma = 0.3$, $T = 1$, $K = 10$, $S = 5$
b) put, American, $S = 9$, otherwise as in a)
c) call, otherwise as in a)
d) The mesh size control must be done carefully and has little relevance to error control. To make this evident, calculate for the test numbers a) a sequence of $V^{(M)}$ values, say for $M = 100, 101, 102, \ldots, 150$, and plot the error $|V^{(M)} - 4.430465|$.

Exercise 1.8 Limiting Case of the Binomial Model

Consider a European Call in the binomial model of Section 1.4. Suppose the calculated value is $V_0^{(M)}$. In the limit $M \to \infty$ the sequence $V_0^{(M)}$ converges to the value $V_C(S_0, 0)$ of the continuous Black–Scholes model given by (A4.10) ($\longrightarrow$ Appendix A4). To prove this, proceed as follows:
a) Let j_K be the smallest index j with $S_{jM} \geq K$. Find an argument why

$$\sum_{j=j_K}^{M} \binom{M}{j} p^j \, (1-p)^{M-j} \, (S_0 u^j d^{M-j} - K)$$

is the expectation $\mathsf{E}(V_T)$ of the payoff. (For an illustration see Figure 1.23.)

b) The value of the option is obtained by discounting, $V_0^{(M)} = \mathrm{e}^{-rT}\mathsf{E}(V_T)$. Show

$$V_0^{(M)} = S_0 B_{M,\tilde{p}}(j_K) - \mathrm{e}^{-rT} K B_{M,p}(j_K)\,.$$

Here $B_{M,p}(j)$ is defined by the binomial distribution ($\longrightarrow$ Appendix B1), and $\tilde{p} := pu\mathrm{e}^{-r\Delta t}$.

c) For large M the binomial distribution is approximated by the normal distribution with distribution $F(x)$. Show that $V_0^{(M)}$ is approximated by

$$S_0 F\left(\frac{M\tilde{p}-\alpha}{\sqrt{M\tilde{p}(1-\tilde{p})}}\right) - \mathrm{e}^{-rT} K F\left(\frac{Mp-\alpha}{\sqrt{Mp(1-p)}}\right),$$

where

$$\alpha := -\frac{\log\frac{S_0}{K} + M\log d}{\log u - \log d}\,.$$

d) Substitute the p, u, d by their expressions from (1.11) to show

$$\frac{Mp-\alpha}{\sqrt{Mp(1-p)}} \longrightarrow \frac{\log\frac{S_0}{K} + (r-\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$$

for $M \to \infty$. Hint: Use Exercise 1.6: Up to terms of high order the approximations $u = \mathrm{e}^{\sigma\sqrt{\Delta t}}$, $d = \mathrm{e}^{-\sigma\sqrt{\Delta t}}$ hold. (In an analogous way the other argument of F can be analyzed.)

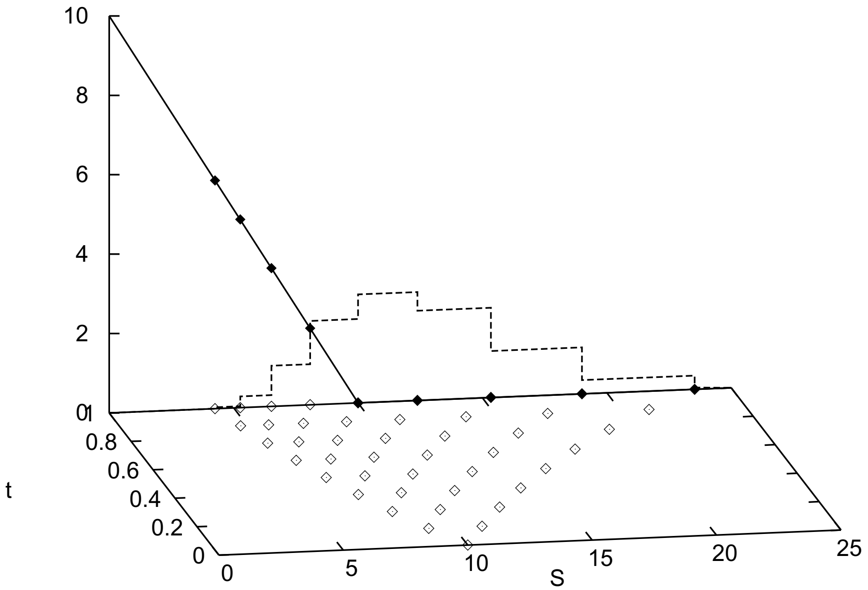

Fig. 1.23. Illustration of a binomial tree and payoff for Exercise 1.8, here for a put, (S, t) points for $M = 8$, $K = S_0 = 10$. The binomial density of the risk-free probability is shown, scaled with factor 10.

Exercise 1.9

In Definition 1.7 the requirement (a) $W_0 = 0$ is dispensable. Then the requirement (b) reads

$$\mathsf{E}(W_t - W_0) = 0\,, \quad \mathsf{E}((W_t - W_0)^2) = t\,.$$

Use these relations to deduce (1.21).
Hint: $(W_t - W_s)^2 = (W_t - W_0)^2 + (W_s - W_0)^2 - 2(W_t - W_0)(W_s - W_0)$

Exercise 1.10

a) Suppose that a random variable X_t satisfies $X_t \sim \mathcal{N}(0, \sigma^2)$. Use (B1.4) to show

$$\mathsf{E}(X_t^4) = 3\sigma^4\,.$$

b) Apply a) to show the assertion in Lemma 1.9,

$$\mathsf{E}\left(\sum_j ((\varDelta W_j)^2 - \varDelta t_j)\right)^2 = 2\sum_j (\varDelta t_j)^2$$

Exercise 1.11 Analytical Solution of Special SDEs

Apply Itô's lemma to show

a) $X_t = \exp\left(\lambda W_t - \frac{1}{2}\lambda^2 t\right)$ solves $\mathrm{d}X_t = \lambda X_t\,\mathrm{d}W_t$
b) $X_t = \exp\left(2W_t - t\right)$ solves $\mathrm{d}X_t = X_t\,\mathrm{d}t + 2X_t\,\mathrm{d}W_t$

Hint: Use suitable functions g with $Y_t = g(X_t, t)$. In (a) start with $X_t = W_t$ and $g(x,t) = \exp(\lambda x - \frac{1}{2}\lambda^2 t)$.

Exercise 1.12 Moments of the Lognormal Distribution

For the density function $f(S; t - t_0, S_0)$ from (1.48) show

a) $\int_0^\infty S f(S; t - t_0, S_0)\,\mathrm{d}S = S_0 \mathrm{e}^{\mu(t-t_0)}$
b) $\int_0^\infty S^2 f(S; t - t_0, S_0)\,\mathrm{d}S = S_0^2 \mathrm{e}^{(\sigma^2+2\mu)(t-t_0)}$

Hint: Set $y = \log(S/S_0)$ and transform the argument of the exponential function to a squared term.
In case you still have strength afterwards, calculate the value of S for which f is maximal.

Exercise 1.13 Return of the Underlying

Let a time series $S_1, ..., S_M$ of a stock price be given (for example data in the domain `www.compfin.de`).

The simple return

$$\hat{R}_{i,j} := \frac{S_i - S_j}{S_j}\,,$$

an index number of the success of the underlying, lacks the desirable property of additivity

$$R_{M,1} = \sum_{i=2}^{M} R_{i,i-1} \, . \qquad (*)$$

The log return

$$R_{i,j} := \log S_i - \log S_j \, .$$

has better properties.

a) Show $R_{i,i-1} \approx \hat{R}_{i,i-1}$, and
b) $R_{i,j}$ satisfies $(*)$.
c) For empirical data calculate the $R_{i,i-1}$ and set up histograms. Calculate sample mean and sample variance.
d) Suppose S is lognormally distributed. How can a value of the volatility be obtained from an estimate of the variance?
e) The mean of the 26866 log returns of the time period of 98.66 years of Figure 1.21 is 0.000199 and the standard deviation is 0.01069. Calculate an estimate of the historical volatility σ.

Exercise 1.14 Anchoring the Binomial Grid at K

The equation (1.10) has established a kind of symmetry for the grid. As an alternative, one may anchor the grid by requiring (for even M)

$$S_0 u^{M/2} d^{M/2} = K \, .$$

a) Give a geometrical interpretation.
b) Derive from equations (1.5), (1.9) and $ud = \gamma$ for some constant γ (not necessarily $\gamma = 1$ as in (1.10)) the relation

$$u = \beta + \sqrt{\beta^2 - \gamma} \quad \text{for} \quad \beta := \frac{1}{2}(\gamma \mathrm{e}^{-r\Delta t} + \mathrm{e}^{(r+\sigma^2)\Delta t}) \, .$$

c) Show that the solution is given by

$$ud = \gamma := \exp\left[\frac{2}{M} \log \frac{K}{S_0}\right] .$$

Exercise 1.15 Extrapolation

Let $\eta^* \in \mathbb{R}$ denote the exact solution of an equation, Δ denotes the grid size of a numerical approximation scheme, and $\eta(\Delta)$ the approximating solution. Further assume an error model

$$\eta(\Delta) - \eta^* = c\, \Delta^q ,$$

with $c, q \in \mathbb{R}$. q is the *order* of the approximation scheme. Suppose that for two grid sizes

$$\Delta_1,\ \Delta_2 = \frac{1}{2}\Delta_1$$

approximations $\eta_1 := \eta(\Delta_1)$, $\eta_2 := \eta(\Delta_2)$ are calculated.

a) For the case of a known η^* (or η^* approximated with very high accuracy) establish a formula for the order q out of η^*, η_1, η_2.

b) For a known order q show that

$$\eta^* = \frac{1}{2^q - 1}(2^q\eta_2 - \eta_1).$$

In general, the error model holds only approximately. Hence this formula for η^* is only an approximation to the exact η^* ("extrapolation").

Exercise 1.16 Portfolios

Figure 1.24 sketches some payoffs over S: (a) bull spread, (b) bear spread, (c) strangle, (d) butterfly spread. For each of these payoffs, construct portfolios out of two or three vanilla options such that the portfolio meets the payoff.

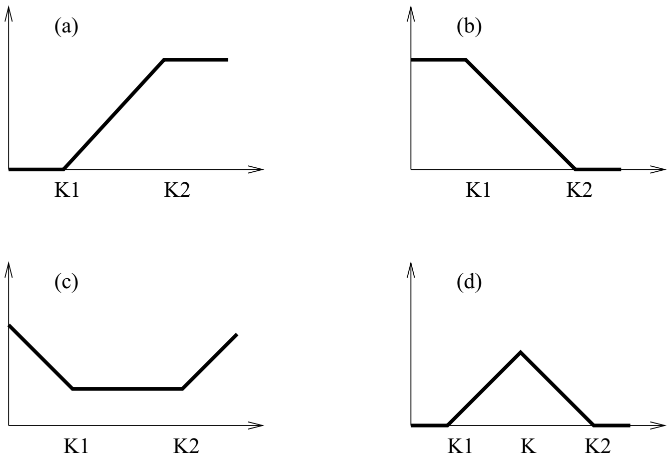

Fig. 1.24. Four payoffs, value over S; see Exercise 1.16

Exercise 1.17 Bounds and Arbitrage

Using arbitrage arguments, show the following bounds for the values V_C of vanilla call options:

a) $0 \le V_C$

b) $(S - K)^+ \le V_C^{Am} \le S$

Exercise 1.18 Positive Itô Process

Let X_t be a positive one-dimensional Itô process for $t \ge 0$.
Show that there exist functions α and β such that

$$\mathrm{d}X_t = X_t(\alpha_t \,\mathrm{d}t + \beta_t \,\mathrm{d}W_t)$$

and

$$X_t = X_0 \exp\left\{ \int_0^t (\alpha_s - \frac{1}{2}\beta_s^2)\,\mathrm{d}s + \int_0^t \beta_s \,\mathrm{d}W_s \right\}$$

Exercise 1.19 General Black–Scholes Equation

Assume a portfolio

$$\Pi_t = \alpha_t S_t + \beta_t B_t$$

consisting of α_t units of a stock S_t and β_t units of a bond B_t, which obey

$$\mathrm{d}S_t = \mu(S_t, t)\,\mathrm{d}t + \sigma(S_t, t)\,\mathrm{d}W_t$$
$$\mathrm{d}B_t = r(t)B_t \,\mathrm{d}t$$

The functions μ, σ, and r are assumed to be known, and $\sigma > 0$. Further assume the portfolio is *self-financing* in the sense

$$\mathrm{d}\Pi_t = \alpha_t \,\mathrm{d}S_t + \beta_t \,\mathrm{d}B_t \,,$$

and *replicating* such that Π_T equals the payoff of a European option. (Then Π_t equals the price of the option for all t.) Derive the Black–Scholes equation for this scenario, assuming $\Pi_t = g(S_t, t)$ with g sufficiently often differentiable.
Hint: coefficient matching of two versions of $\mathrm{d}\Pi_t$

Exercise 1.20 Ornstein–Uhlenbeck Process

An Ornstein–Uhlenbeck process is defined as solution of the SDE

$$\mathrm{d}X_t = -\alpha X_t \,\mathrm{d}t + \gamma \,\mathrm{d}W_t \quad , \quad \alpha > 0$$

for a Wiener process W.

a) Show

$$X_t = \mathrm{e}^{-\alpha t}\left(X_0 + \gamma \int_0^t \mathrm{e}^{\alpha s}\mathrm{d}W_s \right)$$

b) Suppose the volatility σ_t is an Ornstein–Uhlenbeck process. Show that the variance $v_t := \sigma_t^2$ follows a Cox–Ingersoll–Ross process, namely,

$$\mathrm{d}v_t = \kappa(\theta - v_t)\,\mathrm{d}t + \sigma_{\mathrm{v}}\sqrt{v_t}\,\mathrm{d}W_t \ .$$

Exercise 1.21 Binomial Method with $\mathbf{p = 0.5}$

Use the equations (1.5), (1.9) and $p = 1/2$ to show

$$u = \mathrm{e}^{r\Delta t}\,(1 + \sqrt{\mathrm{e}^{\sigma^2 \Delta t} - 1})$$
$$d = \mathrm{e}^{r\Delta t}\,(1 - \sqrt{\mathrm{e}^{\sigma^2 \Delta t} - 1})\,.$$

Exercise 1.22 Dividend Payment and the Binomial Method

A dividend yield δ can be calculated by annualizing a known dividend payment D per year by setting $\delta = D/S$. For a binomial tree, the effects of paying either

a) a fixed amount D or

b) a proportional amount δS

are different.

Assume a dividend payment at time $t_D < T$ and a node of the tree at $t_\nu = t_D$. For a share value of S at $t_{\nu-1}$ discuss the tree evolution at $t_{\nu+1}$ with focus on recombination, comparing the two scenarios a) and b).

Exercise 1.23 Improved Binomial Tree

The Algorithm 1.4 is to be improved as follows:

a) Apply the anchoring of Exercise 1.14.

b) Extend the tree by starting at $-2\Delta t$ as discussed in Section 1.4.6, and calculate approximations for the Greeks delta and gamma by using difference quotients.

Use Example 1.5 to compare these approximations with those from the analytic values from Appendix A4. Implement this in a computer program.

Exercise 1.24 Negative Prices

Assume $Z \sim \mathcal{N}(0,1)$, $S > 0$, $\sigma > 0$, and a step $(t, S) \rightarrow (t + \Delta t, S + \Delta S)$ of the discretized GBM

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma Z \sqrt{\Delta t}\,.$$

What is the probability that the resulting price $S + \Delta S$ is negative? Discuss the result and think about remedy.